D3.3 Final report on applications working groups

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

Due date of deliverable: 31/06/2014
Actual submission date: 30/09/2015
Publication date: 30/09/2015

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), Lee Margetts (NAFEMS, Univ Manchester), 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 final 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	X
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)	

Table of Contents

Glossary

Abbreviation / acronym	Description
ACARE	Advisory Council for Aeronautics Research in Europe
API	Application Programming Interface
CAD	Computer Assisted Design
CAGR	Compound annuel growth rate
CECAM	Centre Européen de Calcul Atomique et Moléculaire
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
cPPP	Contractual Public Private Partnership between ETP4HPC and EC in the field of HPC
CSE	Computational Science and Engineering
CSM	Computational Structural Mechanics
DEEP	Dynamical Exascale Entry Platform
DNS	Direct Numerical Simulation
DoF	Degree of Freedom
DSL	Domain Specific Langage
EC	European Commission
ECMWF	European Center for Medium range Weather Forecasting
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
LES	Large Eddy Simulation
LFT	Lattice field theory
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
NCSI	National Strategic Computing Initiative
ISV	Independent Software Vendors
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
PRACE-4IP	Forth Implementation Project of PRACE, a H2020 funded project
RANS	Reynolds-Averaged Navier–Stokes equations
SRA	Strategic Research Agenda (ETP4HPC)
SPH	Smooth Particle Hydrodynamics
TRL	Technology Readiness Level
UQ	Uncertainty Quantification
VERIFI	Virtual Engine Research Institute and Fuels Initiative

VVUQ	Verification, Validation and Uncertainty Quantification
WCES	EESI-2 WP3 Weather, Climatology and Solid Earth Sciences working group
WG	Working Group (in EESI2)
WP3	EESI2 workpackage on Applications
WP4	EESI2 workpackage on Enabling technologies
WP5	EESI2 workpackage on Cross cutting issues
WP6	EESI2 workpackage on Operational software maturity-level methodology

1. Executive Summary

This document is the last and final EESI2 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).

As a complement of the 2 previous WP3 reports issued during the EESI2 project (D3.1 and D3.2), this report resulting from the work of 43 international experts in the field of scientific and industrial applications highlights the latest developments of the European Centers of Excellence and the first FET HPC projects related to application development, an update of the scientific challenges of each application domains and a gap analysis with regard to the situation at the beginning of the project.

Exascale is seen for all the experts who participated into the different working group as a critical path for providing in an integrated way both Exaflops in terms of computing capability/capacity and Exabytes (exabytes will come faster than exaflops) in terms of storage bandwidth and capacity. This is the only way of addressing scientific and industrial challenges who are already facing this convergence between compute and data (generated from instruments or from large scale simulations).

Applications are THE MEANING of Exascale, so no applications-enabled when the Exascale will come will be a disaster for all the HPC Vendors, the HPC infrastructure and more globally all the HPC ecosystem. Petascale has been a smooth transition with Terascale (which was a smooth transition with Gigascale) but Exascale in the timeframe of 2020-2024 will introduce for energy consumption reasons major constraints (heterogeneous manycore architectures, deep levels of storage/memories, extreme level of parallelism to express, energy movement to limit, fault tolerance to handle, asynchronism, reproducibility error propagation, …) which need to be anticipated NOW.

The reports produced by all the applications experts during EESI2 are stating that applications-enabled mean developing the best coupling between application, algorithms and architectures, leading to :

- the use of **multi-physics and multi-scale coupled simulations** with a strong emphasis on efficient software couplers for coupling new and legacy applications in heterogeneous environments. Exascale systems will be used by some “hero” applications able to scale out to more than 30% of a single machine but most of the needs will rely on capacity (or farming) simulations where coupled (or ensemble) simulations running on >100k threads each will be executed simultaneously;
- the development of energy-efficient **ultrascalable solvers** based on hierarchical algorithms which are able to reduce dramatically communications and synchronisations operations;
- **Mesh generation** is considered as **a major bottleneck to unlock** with need of robust and optimal mesh adaptation methods, curved mesh element generation for higher-order discretization or newer strategies like cut cells, strand grids, “meshless”;
- instruments and large scale hi-fidelity simulations will generate huge volume of data which become, for energy-saving and human time, impossible to store and then analyse off-line. That means that it becomes mandatory to rely on the new HPC and data intensive architectures for addressing the development of innovative data analytics techniques based on **in-situ or in-transit post-processing**, smart tools for for automatic pattern/shape/feature recognition/tracking/reduction/extraction and compression on massive amount of data. Such tools will also allow to implement real time **computational steering or large-scale instruments steering** which could save a lot of time and money for scientists using them. In this domain disruptive approaches in the field of active learning, sub-linear algorithms for features data reduction are mandatory to explore;
- such huge volume of data and such simulations will introduce sources of uncertainties so the issues of **VVUQ will become critical** in all the scientific and industrial domains, leading to the crucial need to develop European opensource framework for dealing with uncertainties. Europe hold already a good position in such expertise, it is time to consolidate it and develop next generation VVUQ tools;

- Exascale architectures will be used if they **remain programmable**, that means that programming models based on standards will need to be improved or developed. Applications are like supertankers, their development time is very long, longer than technologies so they can't take every new software trend appearing every year but much more relay on proven and perene standards. Here a Russian doll approach with portable abstraction layers for hiding the complexity of the underlying hardware could be developed while more advanced layers with pragmas or smart runtime system could be used for fine tuning with the target architectures.

In a situation where Europe owns the development of many applications used in the world (70% of the applications in chemistry are developed in Europe for example), its important to notice that:

- The experts reported during the project strong efforts of scaling out or rewriting of existing scientific and industrial applications to **scalability levels > 100 000 cores**, giving to these applications the potential to scale at least on Pre Exascale architectures to appear in the 2018-2020 timeframe.
  In the case of scaling out of existing applications, these results have been achieved very often thanks to specific optimisations regarding the management of the memory, the optimization of collective communications, the use of more scalable solvers or numerical methods or I/O improvements relying on tiered-storage;
- Such developments and the use now of these applications at this scale require a mandatory effort to **maintain a European HPC research infrastructure like PRACE up to date** and available for supporting science and industry at the same level as what is available in US (thought Incite) or in Japan or China. The example of Renault who performed a world-premiere in crash optimization using PRACE resources, giving to Renault a major competitive advantage in forecasting and assessing advanced numerical methods is a clear example about the benefits for European companies to access to such infrastructures;
- At the same time scientists are acceding to large scale HPC infrastructure like PRACE, it is also important for them to **access to hardware and software prototypes** developed in the field of the FET HPC calls, with good levels of TRL, in order to co design future applications;
- Some **European industrial companies** in Oil & Gas, automotive, energy and automotive **are one of the most actives** in the development and the use of large-scale engineering applications and for some case in the acquisition of multi petascale internal HPC facilities with strong roadmaps toward Exascale.
- There is a challenge for Europe in arming industrial companies with HPC access and best fitted software. That means that an effort need to be made in the industrialization, the user support and the long term maintenance of open source software developed by EU research projects in order to reach a good level of readiness (TRL>5) and be used in confidence by industry. In that sense the EESI2 recommendation issued by WP6 toward the establishment of an **European Extreme-scale Software Centre (EESC)** is fully endorsed by applications communities.
- But a survey launched jointly end of 2014 by NAFEMS and EESI2 WP3 showed that **a lot of ISV are not aware about the challenges related to Exascale** (heterogeneous manycore computing, resilience, energy, …) and a lot of companies are running scientific applications on very low core numbers. This argues to the development of initiatives in Europe toward the use of HPC by industry, PRACE and some national agencies started to deploy enabling initiatives for accessing HPC resources (PRACE OpenR&D) or enabling services (PRACE SHAPE) but this need now to **be more amplified in order to sustain European competitiveness**.
- The start of the structuration of the efforts of multiple teams across Europe inside Center of Excellence is a good signal, fostering synergies and critical mass but the experts are expecting to see higher level of funding (the EESI1 report was proposing a level of funding 8 to 10x higher per CoE), longer duration for the supported projects (its difficult to commit into long term development, support and maintenance of applications and tools without longer visibility) and advise to factorize the material based CoEs and launch a new one related to industrial applications.
- As Exascale is expected to hit the market in the 2020-2024 timeframe, the experts are again pointing the **strong shortage in HPC skills** and urge EC and national agencies to fund specific programmes toward the development of double hat (HPC/big data and domain-science) experts in Europe.

Finally this deliverable also reviews the process of elaboration of five of the EESI2 global recommendations where WP3 has been involved directly:

- European and international efforts on **mini applications**, lighter versions of complex HPC applications, capturing code's essential features for facilitating the co design process by using early prototypes or simulators. Some initiatives are already existing in Europe but a wider effort, potentially supported by the newly created CoE, could be necessary in order to insure a good representation of European applications and needs among the HPC industry;
- Towards flexible and efficient **Exascale software couplers**: the use of such tools is <u>a critical issue</u> for aeronautics, automotive, oil & gas, energy, … almost all the fields of engineering as well as in climate, weather forecast, materials, fusion, life sciences, chemistry, … where multiscale and multiphysics will be used. In this field Europe already owns strong assets and could lead the domain if a specific action is launched toward the development of next-gen scalable and efficient couplers.
- Enhanced **unified framework for model verification and validation and uncertainty quantification**: with the huge volume of data generated as input (from instruments) or output (from simulations) and the complexity of HPC applications and architectures, uncertainties in numerical simulations process could now arise from many sources:
    - Lack of knowledge on a physical parameter (epistemic uncertainty),
    - Parameter with a random nature (aleatory uncertainty),
    - Uncertainty related to the model (model error),
    - Uncertainty related to the numerical errors (numerical errors).

  Taking into account these uncertainties is essential for the acceptance of numerical simulation for decision making. These uncertainties must be integrated in the verification and validation process of the simulation codes. This process is now commonly called VVUQ (Verification, Validation and Uncertainty Quantification). Again in this field Europe owns a strong position but next generation of VVUQ tools need to be developed by using ultrascalable approaches like surrogate models and reduced basis models and by making these tools known (training like PRACE started to do) and disseminated across HPC infrastrucures.
- **Parallel-in-Time**: a major step forward in Parallel simulations: this disruptive recommendation issued by WG3.5 aims to extract parallelism from a new dimension: the time, and represent a huge potential for a better exploration of future Exascale architectures. 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 and the recommendation aimed to assess the impact of such methods, develop benchmarks and an European software library.
- Declarative **processing frameworks for big data analytics, extreme data fusion** e.g. identification of turbulent flow features from massively parallel Exaflops and Exabytes simulations.

  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.

2. WP3 overall activity

2.1 WP3 working groups goals and composition

This is the final 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. The list of all the experts is provided as annex of this report.

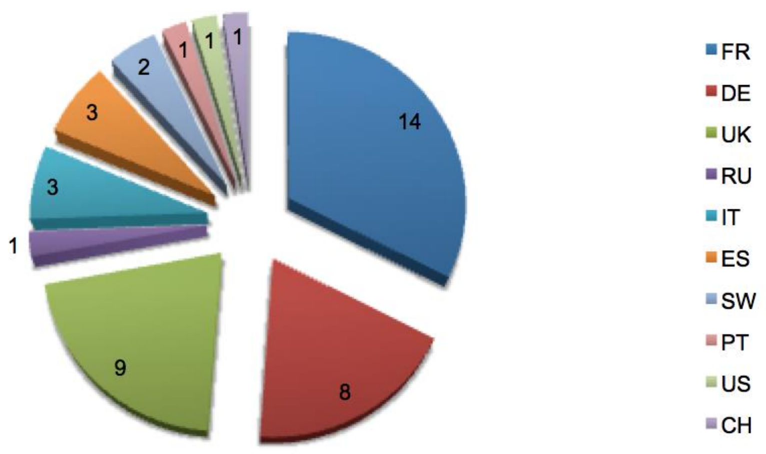


Figure 1 - Repartition per country of the 43 WP3 experts

2.2 Assessment of Centers of Excellence

Beyond the assessment of the scientific roadmaps of the different communities performed inside the different working groups, WP3 had also the role of following the rise of the Centers of Excellence and the outcome of H2020 calls related to applications.

As proposed in the final recommendations of the EESI-1 project in 2012 and stated into the Communication “High Performance Computing (HPC): Europe’s place in a global race”[1] of the European Commission in 2012 and the conclusions of the European Council of Competitiveness[2] on May 2013, the EC has launched on November 2013 a call for proposal (called H2020-EINFRA-2014-2015 – Centres of excellence for computing applications) for selecting a first set of 8 Centers of Excellence (CoE).

A total of 23 proposals were received by the European Commission and at the end of the evaluation on February 2015 the following 8 proposals were selected:

Coordinator		Proposal	
Organisation	**Country**	**Acronym**	**Title**
CEA	FR	EoCoE	Energy oriented Centre of Excellence for computer applications
KTH	SE	BioExcel	Centre of Excellence for Biomolecular Research
MPG	DE	NoMaD	The Novel Materials Discovery Laboratory
CNR	IT	MaX	Materials design at the eXascale
DKRZ	DE	ESiWACE	Excellence in Simulation of Weather and Climate in Europe
UCD	IE	E-CAM	An e-infrastructure for software, training and consultancy in simulation and modelling
BSC	ES	POP	Performance Optimisation and Productivity
POTSDAM	DE	COEGSS	Center of Excellence for Global Systems Science

The following first analysis (the projects are supposed to start for 3 to 4 years on October 1st 2015) has been compiled with the support of experts of the WGs and experts from the PRACE-4IP project.

EoCoE: The aim of the present proposal is to establish an Energy Oriented Centre of Excellence for computing applications. EoCoE (pronounce “Echo”) 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

[1] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0045:FIN:EN:PDF

[2] http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/intm/137344.pdf

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.

Key applications areas: energy, meteorology, materials, water, fusion

BioExcel: Life Science research has become increasingly digital, and this development is accelerating rapidly. Biomolecular modelling techniques such as homology modelling, docking, and molecular simulation have advanced tremendously due to world-leading European research, resulting in extreme demands for better computational performance and throughput as these tools are used in applied research and industrial development. This research has direct influence on our daily life in areas such as health and medical applications, the development of new drugs, efficient drug delivery, biotechnology, environment, agriculture and food industry. Life Science is one of the largest and fastest growing communities in need of high-end computing, and it is a critically important industrial sector for Europe. However, compared to some other disciplines, the use of e-Infrastructure is still relatively new – many advanced techniques are not applied commercially due to limited experience. It requires significant support to:

- Make e-Infrastructure useable by researchers who are not computing experts;
- Improve the performance and applicability of key life science applications;
- Handle large amounts of data in computational workflows.

BioExcel proposes to tackle these challenges by establishing a dedicated CoE for Biomolecular Research, covering structural and functional studies of the building blocks of living organisms – proteins, DNA, saccharides, membranes, solvents and small molecules like drug compounds – all areas where with large academic and industrial users bases in Europe. Specifically, BioExcel will:

- Improve the efficiency and scalability of important software packages for biomolecular research;
- Improve the usability of ICT technologies for biomolecular researchers in academia and industry;
- Promote best practices and train end users in making good use of both software and e-Infrastructure.
- Develop appropriate governance structures and business plans for a sustainable CoE.

Key applications areas: biomolecular modelling, data analytics

Other areas: bioinformatics, systems biology, health

NoMaD: Essentially every new commercial product, be they smart phones, solar cells, batteries, transport technology, artificial hips, etc., depends on improved or even novel materials. Computational materials science is increasingly influential as a method to identify such critical materials for both R&D. Enormous amounts of data, precious but heterogeneous and difficult to access or utilise, are already stored in repositories scattered across Europe. The NoMaD CoE will open new HPC opportunities by enabling access to this data and delivering powerful new tools to search, retrieve and manage it. NoMaD will foster sharing of all relevant data, building on the unique CECAM, Psi-k and ETSF communities, putting Europe ahead of materials science in other continents. Unprecedented, already initialised networking with researchers, with industry, with students and with other stakeholders will guarantee relevance and end-user value. NoMaD will become a crucial tool for atomistic simulations and multi-scale modelling in the physical, materials, and quantum-chemical sciences. This field is characterised by a healthy but heterogeneous eco-system of many different codes that are used at all HPC centers worldwide, with millions of CPU hours spent every day, some of them at petascale performance. NoMaD will integrate the leading codes and make their results comparable by converting (and compressing) existing inputs and outputs into a common format, thus making these valuable data accessible to academia and industry: NoMaD will develop “big-data analytics” for materials science. This will require novel algorithms, e.g., for statistical learning based on the created materials encyclopedia, offering complex searches and novel visualisations. These challenges exploit the essential resources of our HPC partners. Without the infrastructure and services provided by the NoMaD CoE, much of the information created with the above mentioned petascale (towards exascale) computations would be wasted.

Key applications areas: material science, data analytics

MAX: Materials are crucial to scientific and technological advances and industrial competitiveness, and to tackle key societal challenges – from energy and environment to health care, information and communications, manufacturing, safety and transportation. The current accuracy and predictive power of materials’ simulations allow a paradigm shift for computational design and discovery, in which massive computing efforts can be launched to identify novel materials with improved properties and performance; behaviour of ever-increasing complexity can be addressed; sharing of data and work-flows accelerates synergies and empowers the science of big-data; and services can be provided in the form of data, codes, expertise, turnkey solutions, and a liquid market of computational resources. Europe has the human resources, track record and infrastructure to be worldwide leader in this field, and we want to create a CoE in materials’ modelling, simulations, and design to endow our researchers and innovators with powerful new instruments to address the key scientific, industrial and societal challenges that require novel materials. This CoE will be a user-focused, thematic effort supporting the needs and the vision of all our core communities: domain scientists, software scientists and vendors, end-users in industry and in academic research, and high-performance computing centres. The proposal is structured along two core actions: (1) Community codes, their capabilities and reliability; provenance, preservation and sharing of data and workflows; the ecosystem that integrates capabilities; and hardware support and transition to exascale architectures. (2) Integrating, training, and providing services to our core communities, while developing and implementing a model for sustainability, with the core benefit of propelling materials simulations in the practice of scientific research and industrial innovation.

Key applications areas: materials science, solid state physics

Other areas: surface science, quantum chemistry, data analytics

Application codes: Quantum Espresso, Siesta, Yambo and Fleur

ESiWACE: will substantially improve efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms by supporting the end-to-end workflow of global Earth system modelling in HPC environment. This will be obtained by improving and supporting (1) scalability of models, tools and data management on state-of-the-art supercomputer systems (2) Usability of models and tools throughout the European HPC eco-system, and (3) the Exploitability of the huge amount of resulting data. We will develop solutions for crosscutting HPC challenges particular to the weather and climate domain. This will range from the development of specific software products to the deployment of user facing services for both, computing and storage. ESiWACE leverages two established European networks, namely (1) the European Network for Earth System modelling, representing the European climate modelling community and (2) the world leading European Centre for Medium-Range Weather Forecasts. The governance structure that defines the services to be provided will be driven by the European weather and climate science community. Weather and climate computing have always been one of the key drivers for HPC development, with domain specific scientific and technical requirements that stretch the capability and capacity of existing software and hardware to the limits. By developing solutions for Europe and at European scale, ESiWACE will directly impact on the competitiveness of the European HPC industry by engendering new products, providing opportunities for exploitation beyond the project itself, and by enhancing the skills base of staff in both industry and academia. ESiWACE will be at once thematic, as it focuses on the HPC application domain of climate and weather modelling, transversal, as it covers several aspects of computational science, and challenge-driven, as climate and weather predictability represents a major societal issue.

Key applications areas: climate modelling, weather forecasting

Other areas: ocean modelling

Application codes: ICON

E-CAM: will create, develop and sustain a European infrastructure for computational science applied to simulation and modelling of materials and of biological processes of industrial and societal importance. Building on the already significant network of 15 CECAM centres across Europe and the PRACE research infrastructure, it will create a distributed, sustainable centre for simulation and modelling at and across the atomic, molecular and continuum scales. The ambitious goals of E-CAM will be achieved through three complementary instruments:

1. development, testing, maintenance, and dissemination of robust software modules targeted at end-user needs;

   advanced training of current and future academic and industrial researchers able to exploit these capabilities;
2. multidisciplinary, coordinated, top-level applied consultancy to industrial end-users (both large multinationals and SMEs).

The creation and development of this infrastructure will also impact academic research by creating a training opportunity for over 300 researchers in computational science as applied to their domain expertise. It will also provide a structure for the optimisation and long-term maintenance of important codes and provide a route for their exploitation. Based on the requests from its industrial end-users, E-CAM will deliver new software in a broad field by creating over 200 new, robust software modules. The modules will be written to run with maximum efficiency on hardware with different architectures, available at four PRACE centres and at the Hartree Centre for HPC in Industry. The modules will form the core of a software library (the E-CAM library) that will continue to grow and provide benefit well beyond the funding period of the project.

Key applications areas: classical MD, electronic structure, quantum dynamics, meso and multiscale modelling
Applications codes: LAMMPS, DL_POLY, Charmm, NAMD, GROMACS, OPENMM, AMBER, Quantum Espresso, Electronic Structure Library

POP: High performance Computing is becoming a fundamental tool for the progress of science and engineering and as such for economic competitiveness. The growing complexity of parallel computers is leading to a situation where code owners and users are not aware of the detailed issues affecting the performance of their applications. The result is often an inefficient use of the infrastructures. Even when the need to get further performance and efficiency is perceived, code developers may not have sufficient insight on its detailed causes for addressing the problem properly. This may lead to blind attempts to restructure codes and consequent lack of efficiency. The objective of POP is to operate a Centre of Excellence in Computing Applications in the area of Performance Optimisation and Productivity. POP will offer the service of precisely assessing the performance of computing application of any sort, from a few hundred to many thousand processors. Also, POP will show its Customers the issues affecting the performance of their code and the most optimal way to alleviate them. POP will target code owners and users from all domains, including infrastructure operators, academic and industrial users. The estimated population of such applications in Europe is 1500 and within the project lifetime POP has the ambition of serving over 150 such codes. The Added Value of POP's services is the savings generated in the operation and use of a code, which will result in a significant Return on Investment (fixing a code costs less than running it below its optimal levels) by employing best-in-class services and release capacity for resolving other priority issues. POP will be a best-in-class centre. By bringing together the European world-class expertise in the area and combining excellent academic resources with a practical, hand-on approach, it will improve the access to computing applications, thus allowing European researchers and industry to be more competitive.

Applications Areas	Centres of Excellence							
	EoCoE	BioExcel	NoMaD	MaX	ESiWACE	E-CAM	POP	COEGSS
Biosciences:								
+ Biomolecular Modelling		■						
+ Bioinformatics		□						
+ Systems Biology								
+ Medical Science		□						□
Chemical Sciences:								
+ Chemistry				□		■		
+ Materials Science	■		■	■		■		
+ Surface Science				□				
Physics and Astronomy:								
+ Astrophysics & Cosmology								
+ Nuclear Physics								
+ Particle Physics								
+ Plasma Physics	■							
+ Soft Matter Physics	□							
+ Solid State Physics				■				
Environmental Sciences:								
+ Climate Science					■			
+ Geophysics & Seismology								
+ Weather Forecasting	□				■			
+ Ocean Science					□			
Engineering:								
+ Combustion								
+ CFD								
+ Mesoscale Engineering								
+ Structural Engineering								
Others:								
+ CS/Informatics							□	
+ Data Analytics		■	■	□				■
+ Economics/Finance								□
+ HPC Research							■	
+ Mathematics								
+ Humanities								
+ Social Science								■

■ Key applications area

□ Other area

Key applications areas: performance optimisation, HPC research

Other areas: informatics

CoEGGS: Global Systems Science – GSS – is an emerging research field focused on the risks and opportunities involved in global coordination problems. Examples of global systems include the internet, financial markets, intellectual property rights, global energy use and others. Developing evidence and understanding in view of such systems and of related policies is rapidly becoming a vital challenge for modern societies. It requires capabilities for transdisciplinary work that cannot be mastered without massive use of ICT. By the nature of the problem, the relevant datasets are mostly very big, including data streams from social media. To make things more complicated, the relevant algorithms do require the power of high-performance computing. High Performance Data Analysis (HPDA) is the key to success for GSS! A key contribution of the Centre of Excellence for Global Systems Science – COEGSS – will be the development of an HPC-based framework to generate customized synthetic populations for GSS applications. By blending GSS and HPC, we will be able to provide decision makers and civil society with real-time assessments of global risks and opportunities as well as with essential background knowledge about them. This will enable the HPC industry to supply hardware and software for applications well beyond the issues to which HPC has been dedicated so far.

Key application areas: synthetic populations, data analytics, social habits

Other areas: epidemiology, economics

Applications Areas

The table below is based on an exhaustive, but probably not fully comprehensive, list of applications areas. For each of the CoEs, we mark the key applications areas and other areas involved. This allows an analysis to identify areas that are not well represented in the current CoEs.

Conclusions of WP3

These first selected CoE will have between 3 to 4 years to structure a European research community, start the development of new applications, numerical methods and tools, perform training and make these developments available to wide communities including industry and SMEs. The total budget of around 40M€ available for the 8 CoE is by far close to 10 times lower than the budget proposed by the experts of EESI1 in their recommendations for Co Design Centres in 2012.

Also the first selected CoE show a strong number of CoE targeted to materials and no CoE dedicated to industrial applications where the needs of developing open source common software components like multiscale and multiphysics environments, scalable code couplers, scalable grid/meshing tools, in situ post processing, or uncertainties quantification toolboxes have been strongly recommended in order to sustain European industrial competitiveness.

It is expected in 3 years a new call for proposals from the European Commission, the experts of WP3 are recommending to EC to try to factorise/aggregate the materials-oriented CoE and to launch a new CoE in the field of industrial applications as:

1. Europe is holding a strong background of applications which could be used by industry but;
2. these applications not enough known or not enough industrialized (low level of TRL) or long-term supported for being adopted by industry.

2.3 Outcome of the first FET HPC calls

In the field of the cPPP between ETPH4HPC and the European Commission, EC launched a first set of call for proposals for the FET (Future and Emerging Technologies) HPC.

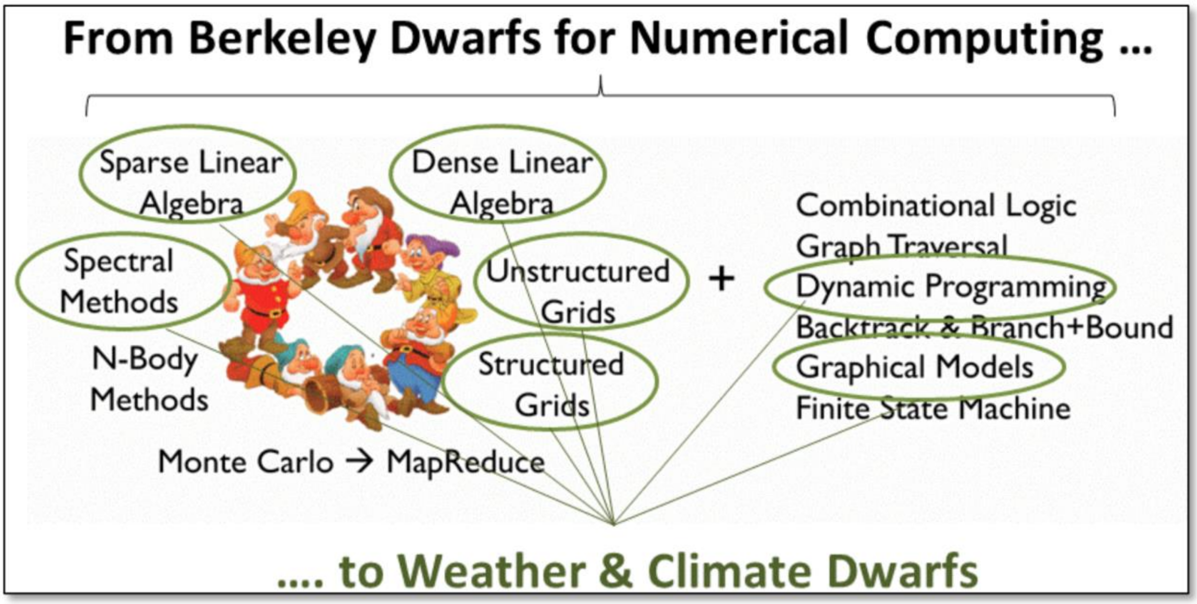
The first one called FETHPC-1-2014 – HPC Core Technologies, Programming Environments and Algorithms for Extreme Parallelism and Extreme Data Applications, with a total funding of 97M€ and addressing the development of the following technologies :

- HPC core technologies and architectures, addressing one or more of the HPC core technologies (processors, memory, interconnect and storage) and their optimal integration into extreme scale HPC systems, platforms and prototypes. Proposals should have a co-design approach driven by ambitious applications and in close cooperation with the scientific disciplines and stakeholders concerned, aiming at radical overall system performance improvement while at the same time addressing issues such as: a holistic understanding of energy efficiency across the full HPC system architecture; I/O, storage and data-throughput capabilities especially for big-data applications; radical scalability, concurrency, locality and resilience in the presence of millions of cores. A minimum of 60% of the available budget for this topic will be allocated to research under this part.
- Programming methodologies, environments, languages and tools: development of new programming models, domain-specific languages, programming paradigms, visualisation and data-analysis tools to facilitate the effective exploitation of the full system capabilities (including energy management) of the emerging large- and extreme scale systems, in particular for extreme parallelism and extreme data applications.
- APIs and system software for future extreme scale systems: New APIs and the corresponding efficient, flexible and scalable exascale system software for managing extreme scale systems, taking into account extreme parallelism, extreme data, energy consumption and resilience. Proposals are expected to include communication and dissemination activities towards relevant standards bodies and research programmes. It is expected that proposals on this point have the critical mass, if necessary beyond Europe, to strategically coordinate the API work in the exascale stack.
- New mathematical and algorithmic approaches for existing or emerging applications on extreme scale systems. Work proposed should include energy-aware algorithms and maximally exploit the projected characteristics of exascale-class systems. Specific issues are quantification of uncertainty and noise, multiscale and extreme data. Software engineering for extreme parallelism should be addressed. Open source development is privileged.

Has leaded to the selection of 18 research projects in which 3 of them are directly related to the co-design of new technologies for addressing specific scientific applications and challenges:

- The Compat[3] (Computing Patterns for High Performance Multiscale Computing) project leaded by University of Amsterdam together with 9 others partners has the objective to develop generic and reusable High Performance Multiscale Computing algorithms that will address the exascale challenges posed by heterogeneous architectures and will enable to run multiscale applications with extreme data requirements while achieving scalability, robustness, resiliency, and energy efficiency. The approach is based on generic multiscale computing patterns that allow us to implement customized algorithms to optimise load balancing, data handling, fault tolerance and energy consumption under generic exascale application scenarios;
- The ESCAPE[4] (Energy-efficient Scalable Algorithms for weather Prediction at Exascale) project leaded by ECMWF (UK) together with 11 partners will develop world-class, extreme-scale computing capabilities for European operational numerical weather prediction (NWP) and future climate models. The biggest challenge for state-of-the-art NWP arises from the need to simulate complex physical phenomena within tight production schedules. Existing extreme-scale application software of weather and climate services is ill equipped to adapt to the rapidly evolving hardware.

[3] http://cordis.europa.eu/project/rcn/197534_en.html
[4] http://cordis.europa.eu/project/rcn/197542_en.html



- The Exaflow[5] project leaded by KUNGLIGA TEKNISKA HOEGSKOLAN (Sweden) together with 7 partners aims to is to address algorithmic challenges to enable the use of accurate simulation models in exascale environments. Driven by problems of practical engineering interest we focus on important simulation aspects including:
    - error control and adaptive mesh refinement in complex computational domains,
    - resilience and fault tolerance in complex simulations
    - heterogeneous modelling
    - evaluation of energy efficiency in solver design
    - parallel input/output and in-situ compression for extreme data.

It's also important to note that 2 others proposals are related to the development of ultra-scalable solvers (ExaHype and NLAFET) and one is related to the development of Exascale Programming, Multi-objective Optimisation and Resilience Management Environment Based on Nested Recursive Parallelism (AllScal). These topics are very relevant towards the recommendations of the WP3 experts.

In the second call for proposals (FETHPC-2-2014 – HPC Ecosystem Development) of the first FET HPC call, a total of 2 proposals of CSA (Coordination & support action) projects were selected:

- EXDCI (European eXtreme Data and Computing Initiative) leaded by PRACE aisbl together with ETP4HPC and some partners from the previous EESI(1,2) projects. The three most significant HPC bodies in Europe, PRACE, ETP4HPC and EESI, have come together within EXDCI to coordinate the strategy of the European HPC Ecosystem in order to deliver its objectives. In particular, the project will harmonize the road-mapping and performance monitoring activities of the ecosystem to produce tools for coherent strategy-making and its implementation by:
    - Producing and aligning roadmaps for HPC Technology and HPC Applications
    - Measuring the implementation of the European HPC strategy
    - Building and maintaining relations with other international HPC activities and regions
    - Supporting the generation of young talent as a crucial element of the development of European HPC

[5] http://cordis.europa.eu/project/rcn/197537_en.html

EXDCI has been launched in early September 2015 with a dedicated workpackage called "Applications roadmaps toward Exascale" which will continue the work on EESI WP3 by involving newly created CoE and working in synergy with the roadmap exercise performed at the same time by ETP4HPC.

- Eurolab-4-HPC – Foundations of a European Research Center of Excellence in High Performance Computing Systems coordinated by CHALMERS TEKNISKA HOEGSKOLA AB together with 12 others partners in Europe which aims to build connected and sustainable leadership in high-performance computing systems by bringing together the different and leading performance orientated communities in Europe, working across all layers of the system stack and, at the same time, fuelling new industries in HPC.

2.4 Involvement of WP3 on EESI2 global recommendations

During its lifetime the EESI2 project issued a series of global recommendations in which WP3 experts have been either leading the process or deeply involved.

2.4.1 Project recommendation "European and international efforts on mini applications"

Experts from WP3 have been involved into the elaboration of one of the first set of EESI2 recommendation issued in 2014 towards the development of European mini applications in an international effort.

Mini applications as lighter versions of complex HPC applications attract growing interest as a flexible test bed to facilitate software development. They aim to project how future more complex applications will interact with future computing hardware and are highly valuable to system designers and architects as long as they represent the behaviour of the complete workload with sufficient fidelity. They should capture a code's essential features - core data structures, algorithms, parallel constructs- while demonstrating state-of-the-art implementations relying on modern programming paradigms.

Some initiatives in US and Asia have already started in order to develop such mini applications for easing the co design of Exascale (capable) architectures and applications.

In Europe some efforts already started in the Intel European Exascale Laboratories of Paris and Leuven, inside some communities (climate, astrophysics for example) or inside the FP7 funded project EXA2CT.

The recommendation proposed to launch a call for proposal targeted to the development of 15 to 20 multi disciplinary exascale mini applications. This effort could be handled with the support of the newly created Centers of Excellence and could lead to a stronger representation of European mini applications derivated from real applications in the worldwide landscape (instead of having the floor occupied for the moment by US mini apps).

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

Experts from WG3.1 and WG3.2 have been involved into the recommendation toward the development of flexible and efficient software couplers. Coupling of multi physics at multi scales codes is <u>a critical issue</u> for aeronautics, automotive, oil & gas, energy, … almost all the fields of engineering. Like in climate and wheatear forecast, Europe owns the development of many couplers used by engineering like OpenPALM for the coupling of CFD codes in turbo-machines (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.

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[6] 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", 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 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.
- 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 …

2.4.3 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

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

- 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, assess uncertainties and reduce time and cost to solution.

2.4.4 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 15M€ 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.

2.4.5 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 Wuppertal 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 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.

It is envisaged a number of world-leading projects utilizing the Parallel-in-Time method but focusing on different approaches and on different application areas. With 2 to 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.

3. Update on key challenges

3.1 Challenges in 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, increasing the safety and reducing the cost of development of the products.

3.1.1 In Oil and Gas

Despite the drop of the price of the crude oil barrel in 2014, 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. Oil & Gas became the second largest profitable market for HPC (just after Finance) with according to IDC an increase of the CAGR of 9.2% between 2012 and 2017. Most of the Oil & Gas major companies or contractors owns multi petascale HPC resources including TOTAL (with Pangea a 6.7 PFlops system), ENI, BP, Exxon, Shell and Chevron for the majors or Petrobras, Woodside Energy (Australia), Petrochina (China) for the NOC or PGS (Petroleum Geo-Services with a 5 PFlops system), CGG-Veritas or Schlumberger for the major contractors have also distributed HPC datacenters with cumulated performed beyond 10 PFlops.

Such companies developed a strong roadmap towards Exascale primarily for the development of efficient and accurate seismic processing methods for exploration and production (4D seismic coupled to reservoir modelling). Also, some companies expanded the use of HPC to scale out basin and reservoir modelling or more recently to molecular dynamics and chemistry applied to processes of refining or global optimization of plants including multi-disciplinary interactions, e.g. simulating full plant lifetime and control systems, emulation of operations and risk analysis.

For addressing such challenges companies which use internal software as well as ISV software will need to work on the scalability of numerical solvers, the availability of multi-scale and multi-physics efficient couplers, the management using in-situ techniques of massive amount of data as well as the need to develop tools for handling uncertainties and verification and validation.

3.1.2 In the Automotive domain

For the automotive domain the classical CAE for crash modelling and CFD studies for aerodynamics continue to drive the demand of HPC facilities in most of the European car companies as well as their subcontractors.

The access to general large scale HPC infrastructures like PRACE or more domain oriented platforms like the ASC-S (Automotive Simulation Centre in Stuttgart) allowed a major jump both in terms of number of meshes and number of unknowns assessed during crash modelling simulations.

As an example 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 PRACE allocation of 42 million core hours on CURIE (at GENCI@TGCC) the biggest crash modelling 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.

Traditional external CFD simulations become also more complex including co-simulation of airflow with FEM for temperature, convection, thermal radiation and heat conduction. Internal CFD simulation like combustion is also gaining a lot of interest due to the challenges of pollutants reduction. Companies start to move from traditional averaged RANS methods to more precise LES (or small DNS) simulations for a more accurate description of the turbulent flows. Companies in Europe like PSA, Continental or Renault as well as GM in the US showed recent results of the use of such new technologies by using HPC resources provided by PRACE or by DoE Incite.

In the domain of combustion applied to automotive but also to aeronautics, aerospace or burners/furnaces, Europe holds strong assets in terms of expertise in advanced methods (LES and DNS) codes developed. During the period all of these applications have been scaled out extremely to more than 100k cores on PRACE and Incite platforms, and now combustion communities are facing challenges in portability of the codes (how to support efficiently complex HPC architectures through pragmas, OpenMP 4.x, external runtime system, DSL, …), meshing tools, asynchronous execution of the models, in-situ analytics, …

But the automotive domain is now facing by 2020 five new major challenges: safety, environment, driving pleasure/life on board, affordability and integration of mobility systems with 5 leading technologies for reaching these challenges: energy efficiency, connected cars and autonomous driving.

Addressing such new challenges will require to continue to reduce the pollutants (CO_2, NO_X) emissions, reduce the weight of the cars by using composite materials, optimise aerodynamics, integrate more detailed virtual dummies in crash modelling, develop real time algorithms for autonomous driving, …

For sure these challenges will impact highly the use of HPC resources by the automotive industry and will require the development of multi-scale multi physics applications, uncertainties quantification frameworks, reduced models, ultra scalable solvers for dealing with large scale CFD or structure mechanics simulations, as well as smart big data management due to the huge deluge of data. As an example Daimler stated that today a crash study over one year is generating around 40 EB of data which is reduced on the fly to a volume of 400 TB where data analytics is performed. The company is expecting a big increase of the size of the rough data generated and tries also to use the temporal data during the on the fly reduction phase to enrich its data analytics.

The issue of developing in-situ data analytics with smart tools for automatic structure (or pattern) detection will become a key issue in the future for supporting engineers when diving into so huge amount of data.

3.1.3 In Aeronautics and aerospace

The impact of computer simulation in aircraft design has been significant and it continues to grow with the expectations of regulators (Acare and Flightpath) and the need to provide optimized designs and reduced development risks and costs.

The latest ACARE (Advisory Council for Aeronautics Research in Europe) (and the Flightpath 2050 Europe's Vision for Aviation) reports are expecting for 2050 a CO_2 reduction by 75%, a NO_x reduction by 90%, by 65% the reduction in perceived aircraft noise (regarding a 2000 typical aircraft), by 80% the accident rate, while air companies are expecting reliable and less kerosene hungry planes with a 3x increase of the traffic and 99% of flights within 15’ of schedule!

To meet the challenges of future aircraft transportation (Greening the Aircraft), it is indispensable to be able to flight-test a virtual aircraft with all its multi-disciplinary interactions in a computer environment and to compile all of the data required for development and certification with guaranteed accuracy in a reduced time frame.

Such Digital Aircraft vision will have an impact on the use of HPC by all the aeronautics value chain, requiring the availability of leading edge HPC resources in both capacity (or farming mode) as well as capability. This will also require strong software development efforts for:

- Increasing the scalability of individual CFD, structure, combustion, acoustics, … simulation codes by working on new scalable numerical solvers on advanced numerical methods (like moving from RANS to LES or using Lattice Boltzmann for improve predictions of complex flow phenomena around full aircraft configurations)
- Introduction of non HPC technologies (CAD) in multi-disciplinary optimisation
- Developing automatic grid generation tools for handling complex geometries
- Next generation of couplers for handling multi scale and multi physics heterogeneous applications (aerodynamics, structural mechanics, aero-elastics, flight mechanics, aero-acoustics, engines, ...). Exascale systems used in capacity mode will allow simulating several manoeuvres simultaneously (flying the aircraft by equations)
- Next generation of uncertainties/optimisation framework for multidisciplinary aircraft design based on high-fidelity methods
- Handling and visualisation of big data with the development of in-situ data analysis

3.1.4 In Power generation and nuclear plants

In this industrial domain, the objectives are multiple: first Improvement of safety and efficiency of the facilities (especially nuclear plants), and second optimization of maintenance operation and extended life span. This is one field in which physical experimentation, for example with nuclear plants, can be both impractical and unsafe. Computer simulation, in both the design and operational stages, is therefore indispensable.

The recent events in Fukushima (in 2011) had a strong impact on the enforcement of the safety regulations making the use of HPC even more mandatory for performing multi -scale and multi-physics simulations and assessing the associated level of uncertainties. Such challenges will require the development of improved CFD and structure codes as well as the availability of next generation of couplers and uncertainties quantification platforms.

Energy companies also started to use HPC in the following domains:

- The development of renewable energies where studies are performed for optimizing the placement of wind-turbines or marine current turbines for an improved power production;
- The development of the smart grids towards the assessment of the impact of distributed and intermittent power generation on electricity networks (due to renewable sources for example) or smart cities for the optimization of the power resources, water, waste, human behaviour, … with issues on management of big amounts of data and smart data analytics methods.

3.2 Challenges in Weather, Climatology and solid Earth Sciences

It is known that the computational cost of weather and climatology models increases nonlinearly with higher resolution, and that high fidelity climate simulations at 1 km resolution will require extreme scale computers. However, I/O and memory-bound, multi-physics codes present particular challenges to computational performance. Indeed, most of the legacy and even some of the current advanced climate applications require code re-engineering, new parallel model design and new algorithms, to exploit new and emerging computational architectures and to increase spatial and time resolution at reasonable cost. It is necessary to design scalable computational kernels and algorithms, as well as considering new approaches and paradigms which are better suited to exploit high levels of concurrency found in parallel HPC systems. A co-design approach is suggested, allowing scientific experts from the application domains, including computational scientists and software engineers, as well as mathematicians to work together on the scientific problem tightly with the technology developers. This will speed-up the development of models and their use on future exascale computers, improve the efficiency of the modelling community and the dissemination of model results to a large community of users.

Such communities are facing the following challenges:

- Data management: Exascale challenges in the domains covered by this working group are equally a big data and HPC challenge. The experts emphasized some challenges connected with data production, data analysis, data storage, and data used by a larger community and mainly related to I/O issues and parallel files systems, scientific data discovery and visualization, parallelization strategies and new frameworks for parallel data analysis, novel programming models for big data, new data structures and storage models, data provenance, and metadata management.

- Uncertainty quantification: Future projections of climate change are uncertain for a number of reasons. The future forcing by greenhouse gases and aerosols are uncertain, and climate variations have both a natural and anthropogenic component and both need to be represented in climate models. The models are also inherently imperfect owing to physical processes that are either not completely understood or yet to be adequately represented because of limited computer power.
- Development of full Earth System Models: Today it is clear that models must also include more sophisticated representations of non-physical processes and subsystems, which are of major importance for long-term climate development, like the carbon cycle. Scientists are strongly interested to know the sensitivity of predictions not only to unresolved physical processes (as, *e.g.*, the cloud feedbacks mentioned above), but also to non-physical ones, like those related to biology and chemistry (including, for example, those involving the land surfaces, and -greenhouse gases reactions).
  It should be noted that including the representation of biogeochemical cycles using different biochemical tracers and aerosols typically increase computing time by a factor of 5 to 20 (depending on the complexity of the parameterizations and the number of tracers). An increase of computing power by a factor of 5 to 20 is then required to better account for the complexity of the system.
- Coupling Technologies: The flexible deployment of Earth System Models requires not only that they be loosely-coupled to their peer models, but also that the code comprising these models is loosely-coupled to the coupling technology and other communication libraries it employs. To facilitate this objective a further level of abstraction is required for the coupling technology. New coupling tools and libraries aim at enhancing the interoperability among the current coupling strategies and at hiding the underlying technology in use. This approach permits models to maintain a consistent interface to multiple coupling and communication toolkits, such as MCT and ESMF, and allows the interoperable use of either or both of these libraries within the models of a coupled model with only trivial configuration changes.

  One of the characteristics of this approach is a separation between the specification of the fields that have existed in a model and their subsequent use in the computational kernel of that model. This permits the particular coupling and halo-exchange facilities used by a model to be changed with only minimal alteration to the source code. This technique also permits the use of a “Configuration Module” in which each field can be configured to use either ESMF or MCT as its underlying communications technology. An additional benefit of this modular structure is the ability to compile model code separately from the code implementing the coupling interoperable framework. This greatly simplifies the task of distributing both models and framework and of porting their code to new execution environments.
- Efficient workflows: Looking to the future, as storing all data produced at high-resolution would not feasible, a typical workflow could be the following: (i) agree, with other scientists and impact people, on a number of spatio-temporal domains and number of variables of particular interest, (ii) run a first high-resolution simulation, store only a reduced number of parameters at global lower resolution (for global analysis of simulation) and store the variables of particular interest on the domains of particular interest at high resolution and (iii) if a further particular study needs additional high-resolution data, rerun the simulation to produce this additional data. Such a workflow could also include on-line statistical diagnostics (e.g., mean, min, max, variance, skew-ness, kurtosis, percentile) of ensemble experiments in order to avoid storing each member.

Currently most of the available software is based on sequential codes. Compression methods such as netcdf4 or jpeg2000 must also be investigated to reduce the amount of data but these methods are often weakly compatible with (strong) parallelism.
Scalable, parallel and robust software frameworks (runtime environment) to support the aforementioned workflows are needed especially in view of running large ensembles of model experiments. Additionally, the WCES community would strongly need:

- on-line statistics on ensemble simulations, and data reduction,
- efficient compression,
- parallel tools for analysis and visualization,
- efficient portable I/O library or server, including on-line reduction of data,
- user friendly and robust (error handling) workflow environment to run and re-run simulations easily changing the configuration,
- quality control of data,
- alternative data formats (e.g. HDF5) to reduce the number of files to be moved and/or generated on the HPC systems.

A plan for achieving interoperability relates to hardware (in terms of comparable / sufficient per node memory on different systems, fast network and sufficient temporary storage space for pre- and post-processing and re-arranging data), software (in terms of APIs for doing data processing, analysis, mining, visualization, etc.) and conventions for data, models and metadata.

3.3 Challenges in Fundamental Sciences

Categorizing the fundamental sciences may be done by disciplines and domains, or by the time- and spatial scales which characterize the fields. Picture a two-dimensional map, where time- and length scales are defined by the characteristic scales of the objects under study, such as elementary particles, charged particles, atoms, molecules, aggregates or large scale structures which exhibit a range in spatial scales from $<10^{-10}$ m to $>10^{+20}$ m, and time scales from $<10^{-15}$ s to $>10^{+15}$ s. This exercise produces fields with more than 30 orders of magnitude differences in space and time. From the domain side, this diagram is spanned by quantum chromodynamics/high energy physics, nuclear physics, laser-plasma physics, nuclear fusion research, quantum chemistry, soft matter research, materials sciences and astrophysics/cosmology. Although there is a large diversity in length and time scales and therefore also different methods and paradigms in the domains, various domains are interlinked by multiscale simulation approaches that aim, for example, to bridge the gap between a microscopic atomistic description and a macroscopic continuum description of matter.

Methods used in fundamental sciences often explore particle or mesh methods (or combinations of both). For example, in statistical physics, physical chemistry, plasma or astrophysics applications, particle methods often are used to explore molecular dynamics, Brownian dynamics, Monte Carlo or particle-in-cell (PIC) methods. The scalability of these methods strongly depends on the computational load and the locality of the algorithm. PIC codes often show a better scalability, due to their locality, than molecular dynamics methods, which include long-range interactions requiring information exchange over the whole system. Grid-based methods, as applied for example in hydrodynamics or magneto-hydrodynamics, include finite element or finite volume methods work on structured or non-structured adaptive meshes. Applying solvers such as multigrids introduces problems on large-scale platforms, e.g., load-balancing in terms of grid partitioning and maintaining a balanced workload over the different grid levels.

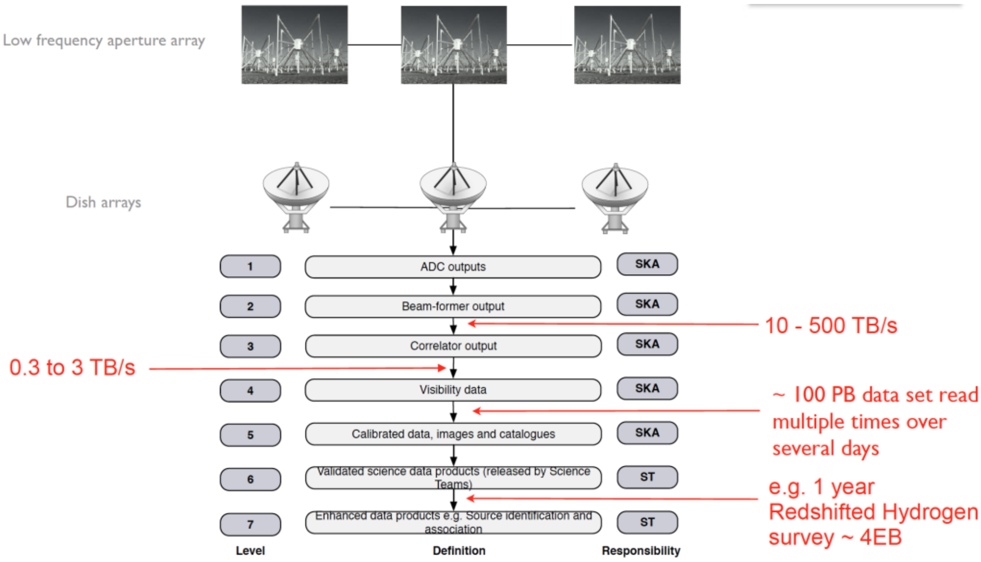
The availability of ultra scalable solvers has been highlighted as a strong requirements by the experts, already some promising results have been observed on the one hand optimal algorithms like Fast

Multipole methods, multigrid or H-matrix methods; and on the other hand, the development of local algorithms such as real space methods that may, for example, apply wavelets.

One of the main challenges in physics and chemistry is the accurate description of complex systems on large time and spatial scales. The most accurate description would rely purely on quantum *ab-initio* methods, but due to the computational complexity, these methods are usually restricted to a small number of atoms and short time scales. For instance, considering the dynamical evolution of polymer systems, consisting of chain-molecules of several thousands of monomers, where the relaxation time increases quadratically with the number of monomers, would lead to numerically unsolvable problems, even on exascale machines. Therefore both the number of degrees of freedom as well as the time scale involved, i.e., the number of time-integration steps, are intractable for most accurate methods. Interesting phenomena like self-aggregation, structure formation, charge transport through porous media or solvation of large molecules (to name a few) would be excluded due to the complex time evolution and system size. In fact, self aggregation phenomena occur on time scales of s-secs, which is a factor of $O(10^9\text{-}10^{15})$ longer than the time to resolve intra-molecular bond-vibrations in a simulation. Therefore the application of computationally expensive methods, like *ab-initio* methods, is not applicable. In addition, these methods usually do not show a strong scaling behavior up to $>10^4$ compute cores, so that even exascale machines cannot solve this problem.

To overcome these limitations, the usual way is to coarse grain the description of the systems, e.g. neglecting electrons, polarization effects, detailed atomistic properties or in making the transition from an atomistic description to a field description. Traditionally these levels of description were used in their own fields, e.g. a force field description in molecular dynamics or continuum field descriptions in fluid dynamics. However, it is also possible to combine these levels of descriptions in different multiscale descriptions (vertical or horizontal) in a simulation setup.

The next decade will see the advent of large-scale cosmological experiments, both from the ground (LSST) and from space (Euclid and WFIRST) or new radio telescopes like LOFAR or SKA. These instruments will generate a huge volume of data which will need to be post-processed by supercomputers stems resources leading to a convergence between instruments and HPC.
The example of the SKA is really representative of that convergence. The Square Kilometre Array (SKA) is a radio telescope project that will be built in Australia and South Africa. It will have a total collecting area of approximately one square kilometre. It will operate over a wide range of frequencies and its size will make it 50 times more sensitive than any other radio instrument. It will require very high performance central computing engines and long-haul links with a capacity greater than the current global Internet traffic. It will be able to survey the sky more than ten thousand times faster than ever before.
With receiving stations extending out to distance of at least 3,000 kilometres (1,900 mi) from a concentrated central core, it will exploit radio astronomy's ability to provide the highest resolution images in all astronomy.
Construction of the SKA is scheduled to begin in 2018 for initial observations by 2020, but the construction budget is not secured at this stage. The SKA will be built in two phases, with Phase 1 (2018-2023) representing about 10% of the capability of the whole telescope. Phase 1 of the SKA was cost-capped at 650 million euros in 2013, while Phase 2's cost has not yet been established (but estimated to 1200M€).
Roadmap in terms of compute and data processing needs for the first phase of SKA is the following:



This show the strong need of having architectures where there is a strong integration between storage, computing and networking, a more intelligent, hierarchical object storage and application driven networking and a dynamic integration of several memory and cache levels into intelligent data movement/pre-fetch agents.
Applications running on this architecture are based on workflows dealing with data reduction operators in order to ingest, edit, calibrate, image, find source, analyse and archive, running constantly (as the telescope observes).
Phase 2 of SKA planned for 2020-2024 will be even more exigent in terms of data and compute requirements:

# 2013 estimate by SKA South Africa	MeerKAT Pre-Cursor 2014-15	SKA Phase 1 2017-19	SKA Phase 2 Est. 2020-24
Data into CSP	2 Tbps	50 Tbps	up to 5 Pbps
Data into SDP	0.4 Tbps	20 Tbps	up to 500 Tbps
Into Storage	35 Gbps	300+ Gbps	up to 2 Tbps
Computing load	200 TFlops	30+ PFlops	3+ EFlops

In astrophysics again some 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 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.

In particle physics Lattice field theory (LFT) calculations can exploit a broad range of hardware platforms. Central to LFT are extreme scale capability machines. The generation of gauge field configurations is the single most CPU extensive task in LFT simulations. The calculations of physical quantities from these ensembles require smaller partitions and can be carried out on workstations clusters or farms GPUs. The substantial requirements of lattice ensemble generation are expected to be well suited to future exascale computers. These calculations require a challenging balance between floating point capability and communication bandwidth and latency. However, given the steep scaling of computation cost with lattice spacing, the total memory requirements grow slowly with computing performance. Current trends in petaflop systems appear efficiently matched to these LFT calculations. Even with economical mesh-like layers, a regular problem like LFT achieves a good performance, and this trend is expected to continue to the exaflop scale. Thus, we expect that LFT will continue to be one of the leading application areas demonstrating the transformative potential for extreme scale computing capability.

Atomic physics addresses the quantum dynamics of electron and positron collisions with atoms and molecules, plus field/multiphoton interactions with atoms and molecules in laser fields, including strong fields and ultrafast laser pulses.

Computationally, the grand challenges, apart from introduction of new science into the programs, include extension (and in certain cases, introduction) of parallelism at both multicore and peta/exa-scaling levels to cope with the complexity of the multi-channel, multi-electron wavefunctions and interactions: propagation, Hamiltonian diagonalization, (complex) operator matrix construction and extraction of scattering parameters, together with defining the accuracy of variational calculations that do not intrinsically involve a minimum principle.

3.4 Challenges in Life Sciences and Health

Since there is a wide range of projects requiring Exascale performance in Life Sciences, we have organized our panel of experts into four main areas:

- In SYSTEMS BIOLOGY we are now at the stage of collecting data to build models for complex simulations that will describe in the next future the dynamics of cells and organs that remain unknown. The models that are developed today are stored in databases. Progress is rapid and systems biology will allow to couple the simulations of the models with a biomedical problem (e.g. monitor mutations in a specific genome that can change the activity of a protein). This will require large computational resources and systems biology will benefit from Exaflop capabilities, but aspects related to data management are going to be as important as pure processing capability.
- In GENOMICS research we face problems (e.g. the sequencing of 2,500 genomes of cancer patients) involving the management of massive amounts of data in programs that can require hundreds of thousands of processors, but little inter-processor communication. However, the vast amount of data to be managed (and often confidentiality and privacy aspects) hampers the use of cloud or grid-computing initiatives as a general solution. Suitable and flexible access to computer resources is crucial in this area. The genomic subpanel asserts that currently known cornerstones for an Exascale system (number of computer nodes, I/O and memory capacities) are clearly driving the focus only to reach the Exaflop peak performance. For most of the genomics challenges an Exaflop computer that could be even less “balanced” than today‟s HPC systems would be a substantial barrier to use these machines efficiently. The genomics subpanel and by extension the entire Life Sciences panel wishes to stress their major concerns that their Exascale problem will be difficult to treat with the unbalanced architectures of anticipated Exaflop computers.
- In MOLECULAR SIMULATION projects, Exaflop capabilities will allow the use of more accurate formalisms and enable molecular simulation for high-throughput applications (e.g. study of larger number of systems). Unfortunately, Exaflop capabilities will not favour the possibility to study longer time scales, since it will not be possible to scale into systems based on hundreds of thousand cores (as the simulated systems typically have less than 1 million atoms). The needs of the molecular simulation field will be better served by a heterogeneous machine, with hierarchical capabilities in terms of number of cores, amount of memory, memory access bandwidth and inter-core communication. This should be contrasted with current ideas regarding a „flat‟ machine with peak Exaflop power. Exascale capability will however facilitate biased-sampling techniques, which require parallel computing, enabling *in silico* experiments unreachable today. Examples include the proteome-scale screening of

chemical libraries to find new drugs and the study of entire organelles, or even cells, at the molecular level. In most of these cases, parallelization is expected to be hierarchical (e.g. ensemble simulations, multi-scale modelling or a mix of parallelization and high throughput).

- In BIOMEDICINE SIMULATION we envision projects such as the simulation of the brain, organs and tissue modelling, and *in silico* toxicity prediction. In these areas, Exaflop capabilities will be a necessary, but not sufficient, requirement, since the integration of experimental information, human-interaction with calculations and the refinement of underlying physical models will also be instrumental for success. As in the case of Molecular Simulation, multi-scale modelling is one of the major challenges of this area and represents one of the major cross-cutting issues of Exascale systems for Life Sciences.

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*

3.5 Disruptive applications

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.

4. 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:

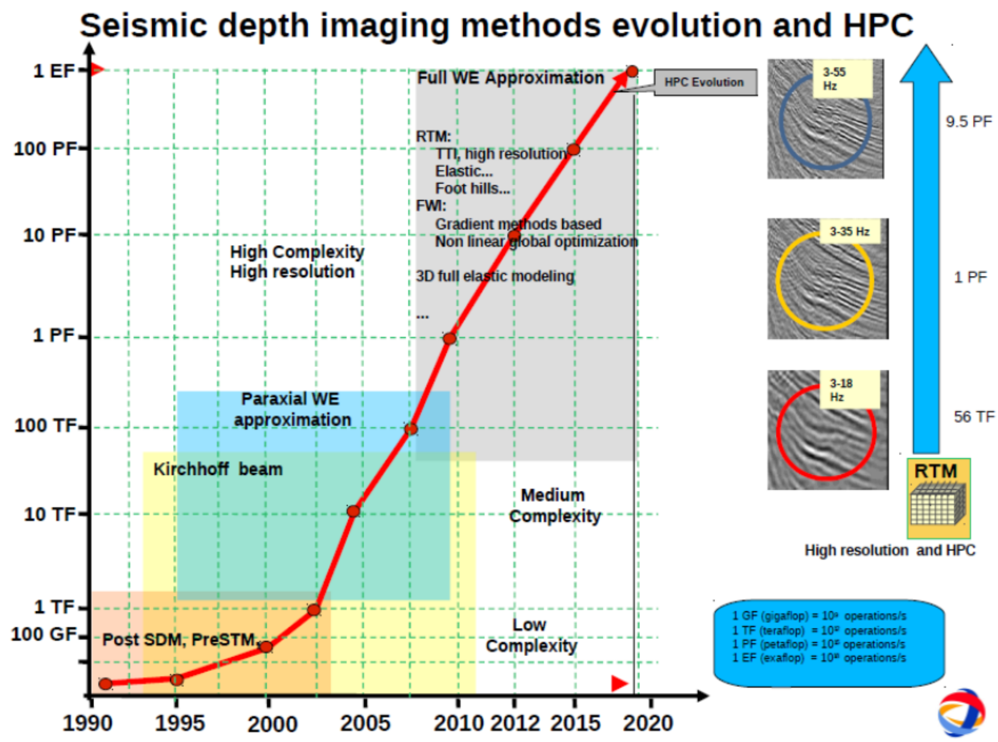


Figure 2 - Previous Oil & Gas Exascale Roadmap produced during EESI1

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:

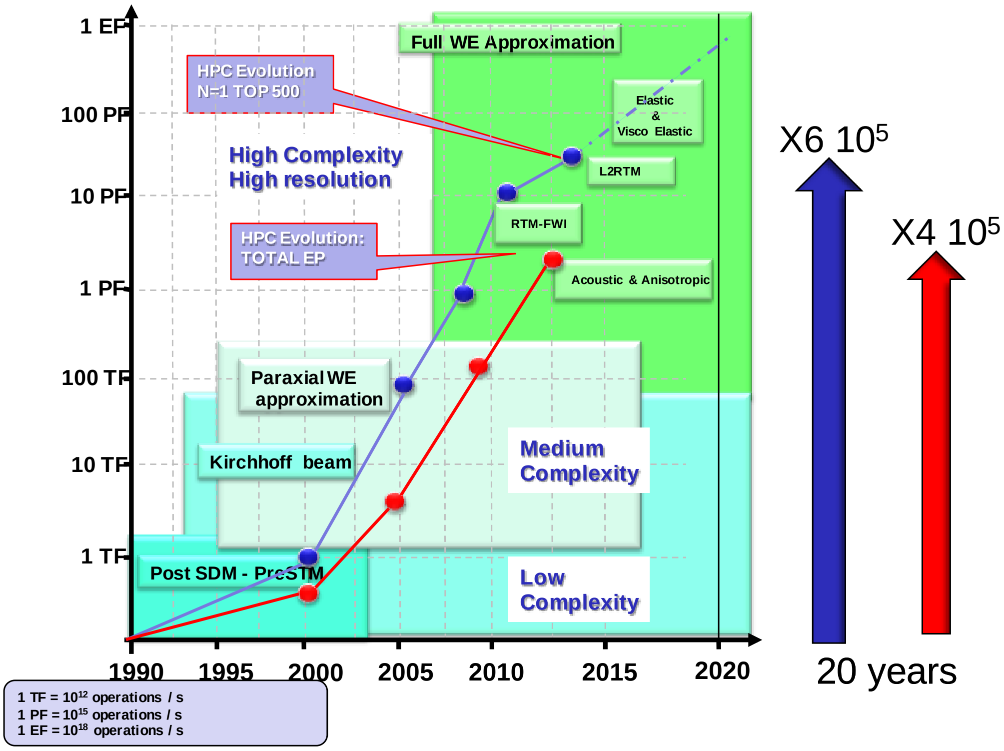


Figure 3 - Updated Oil & Gas Exascale Roadmap (from TOTAL)

This figure 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.

In the automotive domain a new roadmap has been provided in the field of combustion of IC Engines from GM using HPC and experimental facilities at Argone National Laboratories, USA (called VERIFI[7] for performing Hi Fidelity LES combustion studies).

This roadmap has been shared with European experts who consider as well as applicable as the European car industry.

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

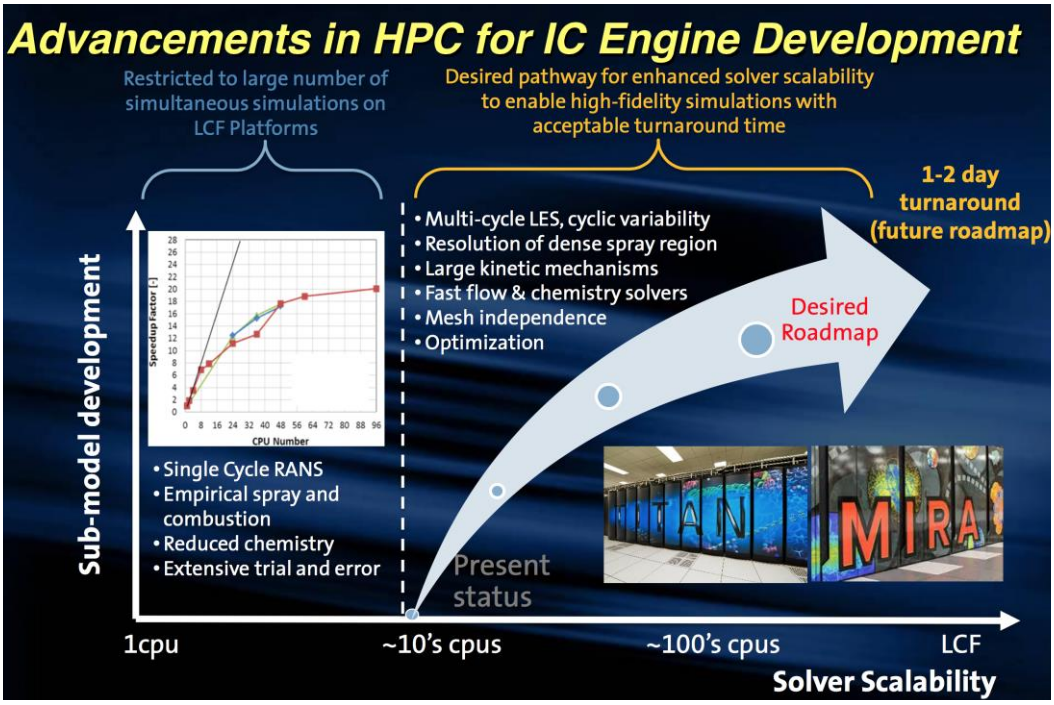


Figure 4 - Roadmap of development of models and use of HPC facilities at GM

This roadmap highlights the shift of the industry from traditional RANS simulation on one singe cycle (one iteration of the behaviour of the engine) to full LES with rich chemical models applied to multiple cycles and multi cylinders of an engine in order to capture more precisely all the turbulence, and assess cyclic variability of the combustion process.

Also companies like Renault-Nissan who used in 2013 PRACE resources for a massive crash optimisation study started to re engage into the use of HPC and expose now aggressive roadmap in both crash, aerodynamics and combustion :

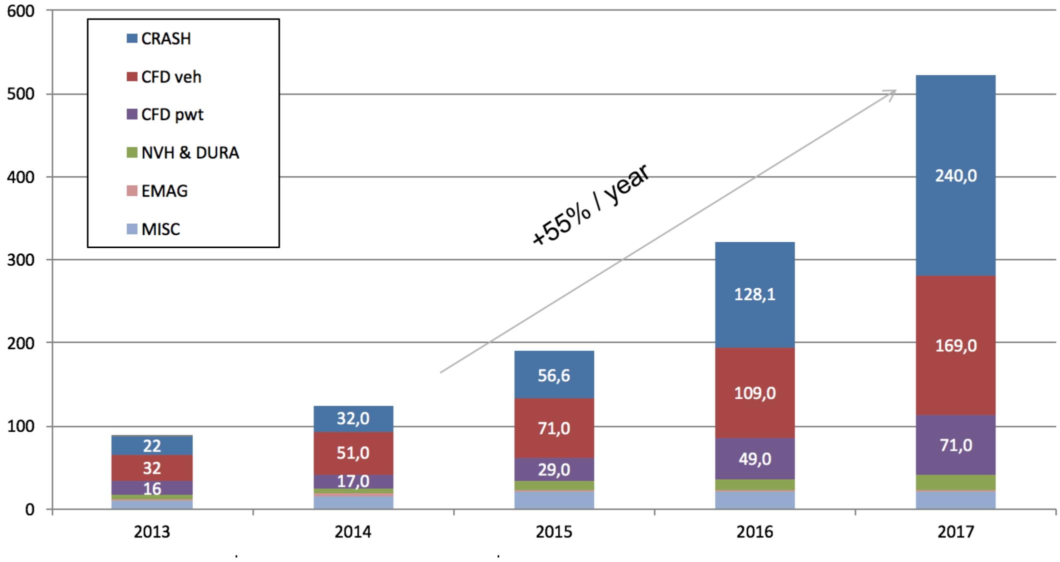
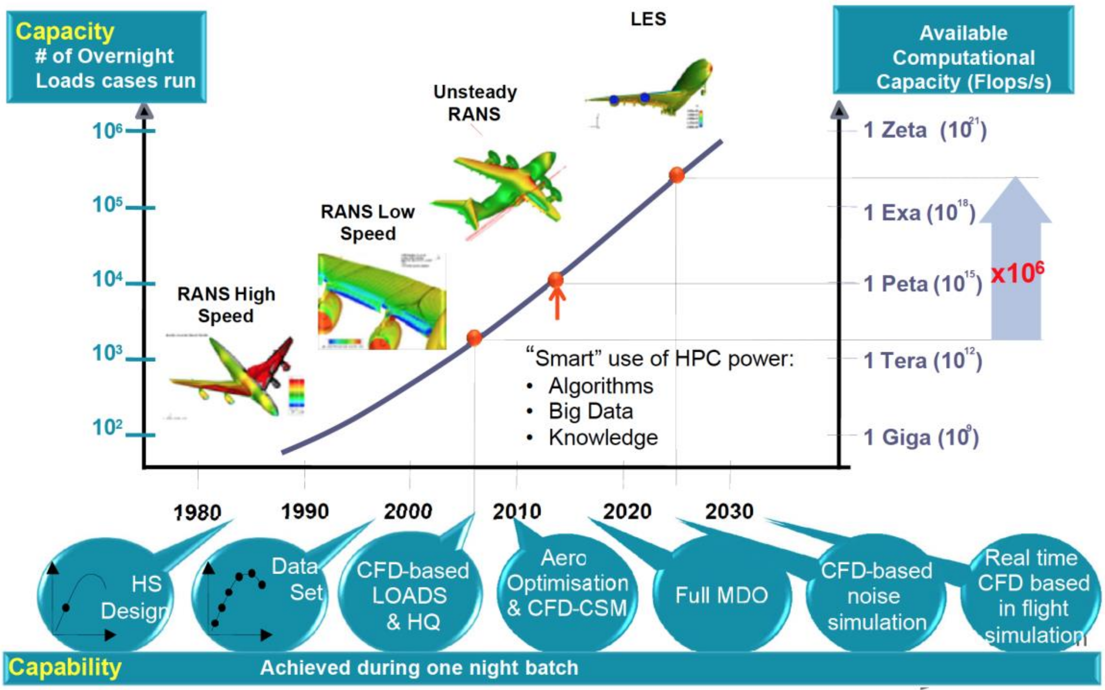


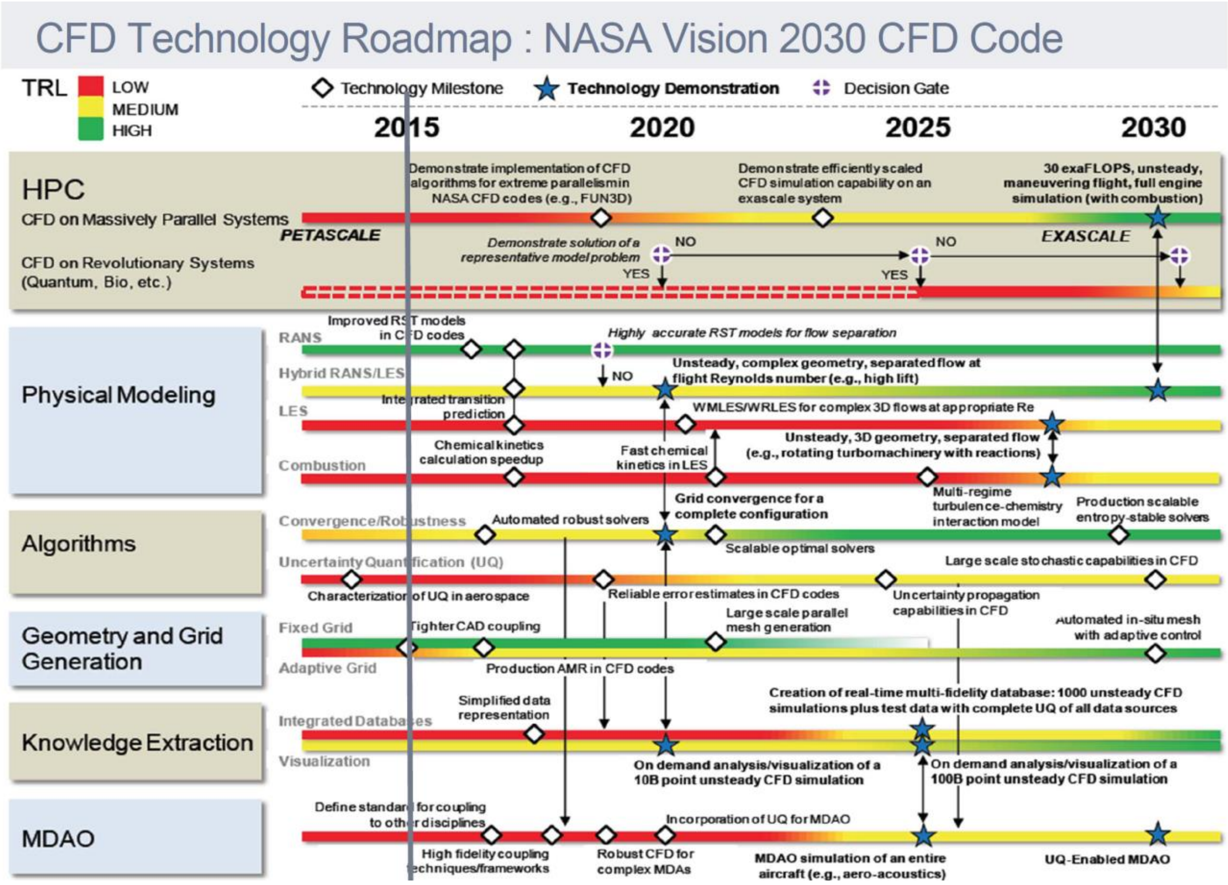
Figure 5 - HPC needs of Renault

In the field of aeronautics the following roadmap provided by Airbus:



has been confirmed again by the experts, leading nearly in 2030 to an integrated multi physics (aerodynamics, CSM, MDO, noise, aero elastic, …) simulation of a full plane. Exascale is then considered as a stepping stone toward a final requirement of resources close to one Zetaflops, used in a mix of capability (for a single applications) and capacity (for coupled multiphysics applications).

The experts also provided another roadmap towards CFD in 2030 coming from NASA which has been also endorsed by the whole aeronautics community:



This roadmap highlights more precisely some issues raised already in EESI2 by the experts of the working group toward the evolutions of the physical modelling (RANS to cover the area where the boundary layer (BL) is too thin for LES, hybrid RANS/LES for a good compromise in performance and quality, or full LES), the development of new algorithms and meshing tools, the growing need of analytics and in-situ technologies and finally the importance of the development of multi disciplinary analysis and optimisation (MDAO) with uncertainties quantification.

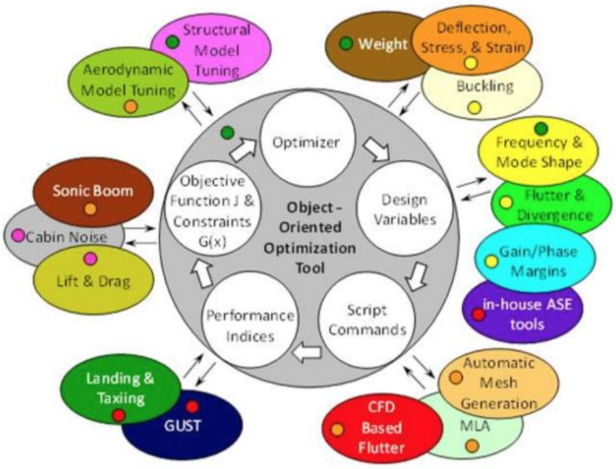


Figure 6 – NASA' MDAO requirements

For a wall-modelled and wall-resolved LES CFD simulation in 2030 on a simple geometry and Re=1e8 a machine close to 30 EFlops will be required for a single execution.

Mesh generation is considered as a major bottleneck to unlock with need of robust and optimal mesh adaptation methods, curved mesh element generation for higher-order discretization or newer strategies like cut cells, strand grids, “meshless”.

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, …

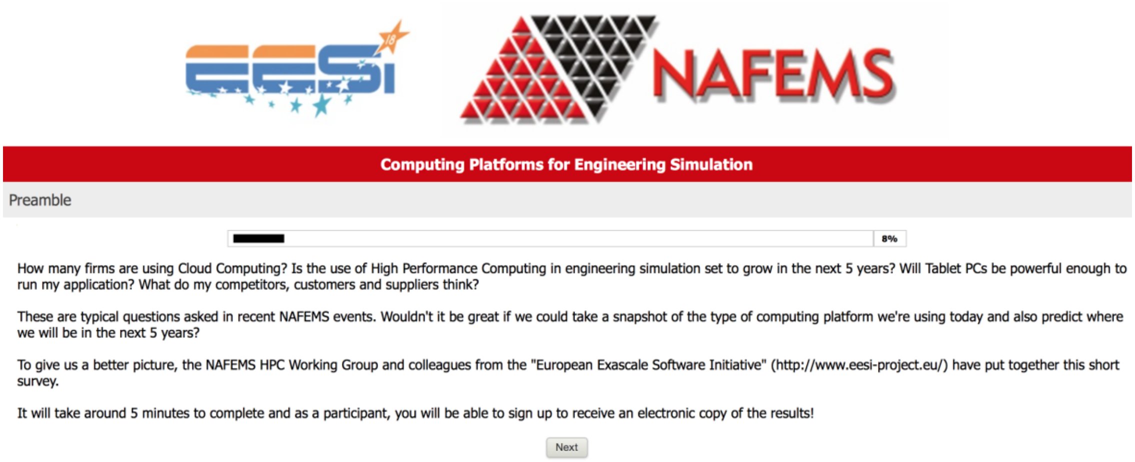
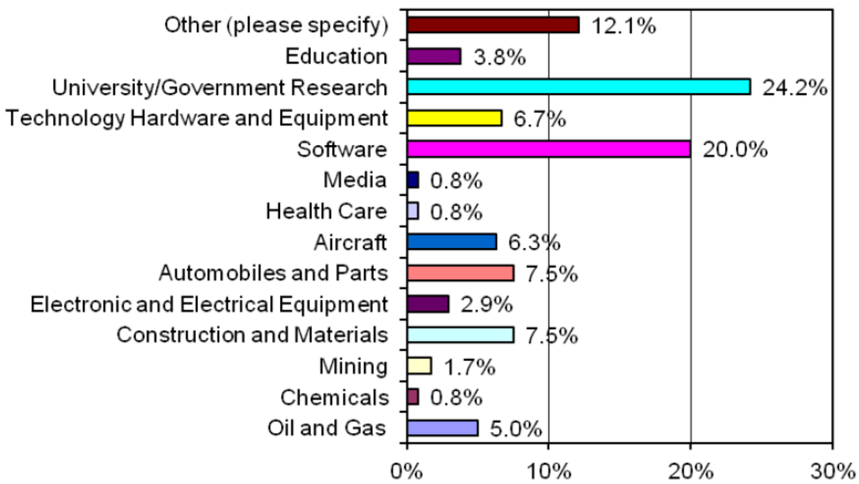


Figure 7 - Front Panel of the joint NAFEMS/EESI Survey

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 is proved in the Annex 2 of this document.

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.

A full analysis of this survey is provided as annex of this document.

The main indications to highlight in this report are the following:

- Industry tend to model smaller problems than 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, multi-body 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;
- In the field of potential Exascale hardware and software technologies (GPU, manycore, ARM processors, FPGA, fault tolerance, energy aware, …) GPU are starting the be adopted (because with workstations they represent an important market for software vendors), the others technologies/challenges for Exascale are not very considered/known by ISV. The situation is a bit better for Open Source Software (right) but there is clearly an issue for such communities to take the HPC wagon.

This study is 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.

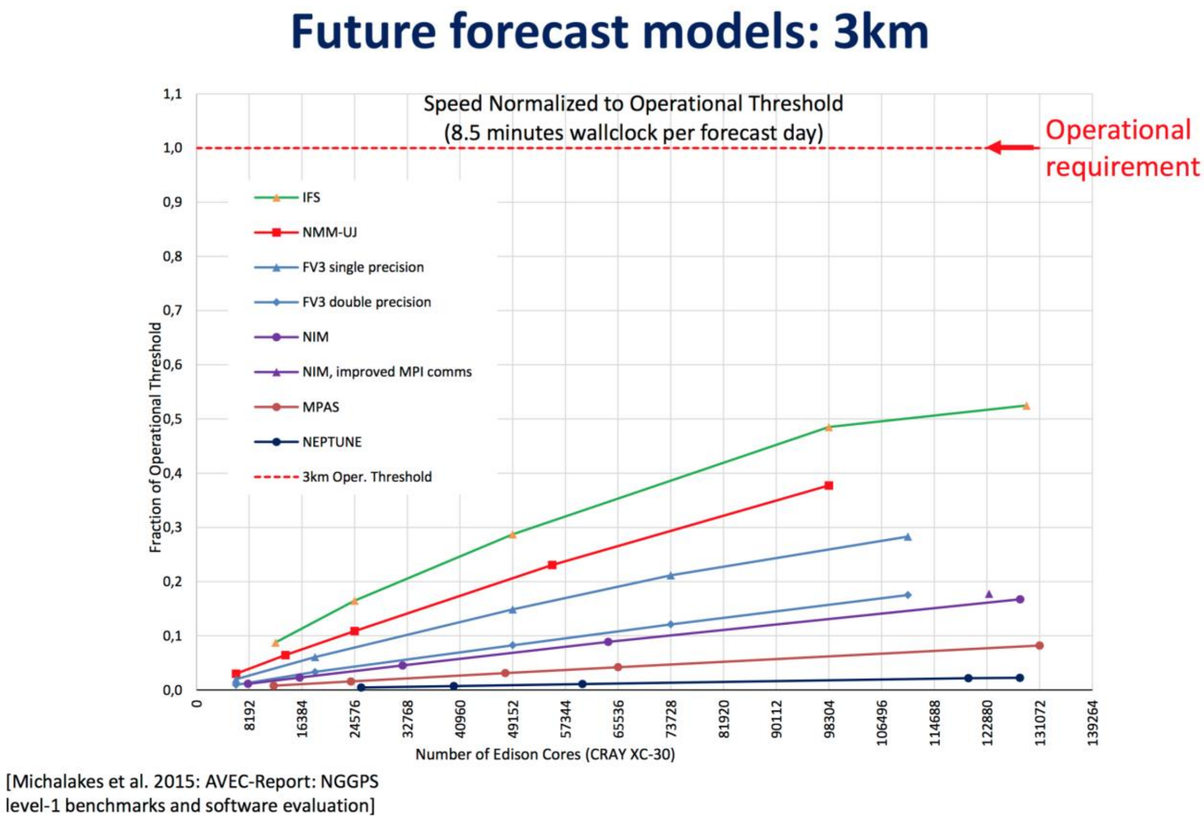
This is a different philosophy to pushing a single CFD simulation, for example, to Exascale.

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 identify 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.

In the field of weather forecast a new roadmap issued in 2015 by ECMWF (European Center for Medium range Weather Forecasting) in the field of its Scalability Programme[8] showed that a strong effort on the weather models is needed to reach the expected level of resolution (3km) and time to response (8.5 min per forecast day) for the next years.



In the field of Life Sciences the announced Human Brain Project FET European project started its activities in early 2014. The goal of the Human Brain Project (HBP) is to gather all existing knowledge about the human brain and to reconstruct the brain, piece by piece, in multi-scale models and supercomputer-based simulations of these models. The resulting “virtual brain” offers the prospect of a fundamentally new understanding of the human brain and its diseases, as well as novel, brain-like computing technologies. The HBP is developing six Information & Communication Technology (ICT) Platforms, dedicated respectively to Neuroinformatics, Brain Simulation, High-Performance Computing (HPC), Medical Informatics, Neuromorphic Computing and Neurorobotics.

---

[8] http://www.eesi-project.eu/wp-content/uploads/2015/05/Bauer_-_ECMWF_-_Accounts_and_expectations_towards_Exascale_-_EESI2_Final_Conference.pdf

In October 2013, the European Commission (EC) started supporting this vision through its Future & Emerging Technologies (FET) Flagship Initiative. The HBP's 21⁄2 year Ramp- Up Phase, which will last until March 2016, is funded by the EU's 7th Framework Programme (FP7). This phase should be followed by a partially overlapping Operational Phase, which will be supported under the next Framework Programme, Horizon 2020.

The HBP as a whole is planned to be implemented in three phases, spread over ten years, with an estimated total budget of more than EUR 1 billion. The project, which is coordinated by the Ecole Polytechnique Fédérale de Lausanne (EPFL) in Switzerland, already brings together 80 European and international research institutions. In 2014, more partners should join the consortium via the HBP's Competitive Call Programme.

The HPC Platform Subproject (SP 7) is one of the HBP's 12 operational subprojects. Its mission is to build and manage the hardware and software for the supercomputing and data infrastructure required to run cellular brain model simulations up to the size of a full human brain. SP 7 will make this infrastructure available to the consortium and the scientific community worldwide.

Central to the HPC Platform is the HBP Supercomputer, the project's main production system, to be located at Jülich Supercomputing Centre. The HBP Supercomputer will be built in stages, with an intermediate "pre-exascale" system of the order of 50 PFLOPs planned for the 2016-17 timeframe. Full brain simulations are expected to require exascale capabilities, which, according to most potential suppliers' roadmaps, are likely to be available in approximately 2021-22. As well as providing sufficient computing performance, the HBP Supercomputer will also need to support data-intensive interactive supercomputing and large memory footprints. Besides the HBP HPC main production system in Jülich, there will be Software Development System at CSCS, Switzerland, a Subcellular Computing System at BSC, Spain, and a Data Analytics System at Cineca, Italy.

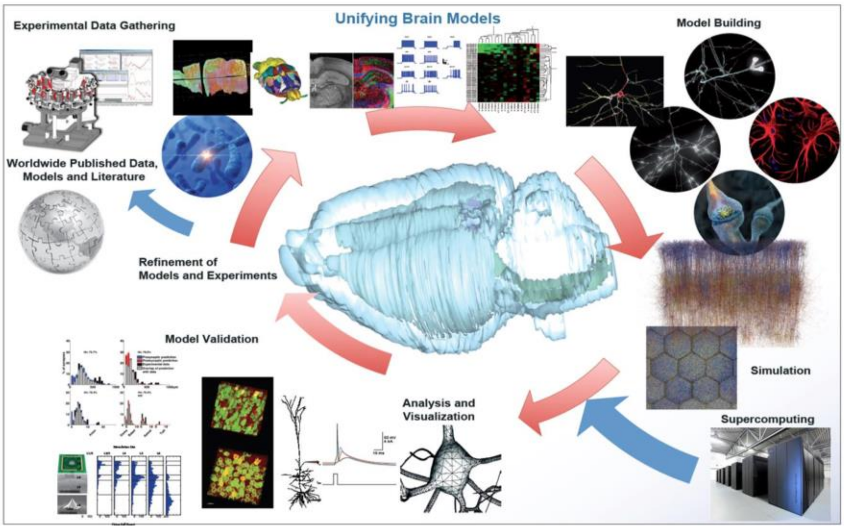


Figure 8 - The HBP Data integration strategy (from F. Schurmann, EPFL)

While exascale supercomputers will become available sooner or later without any HBP intervention, it is unlikely that these future systems will meet unique HBP requirements without additional research and development (R&D). This includes topics like tightly integrated visualization, analytics, and simulation capabilities, efficient European-wide data management, dynamic resource management providing co- scheduling of heterogeneous resources, and a significant enlargement of memory capacity based on power-efficient memory technologies. In line with the EC's established HPC strategy, Forschungszentrum Jülich will therefore drive R&D for innovative HPC technologies that meet the specific requirements of the HBP through a Pre Commercial Procurement (PCP) launched in April 2014[9].

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)[10]

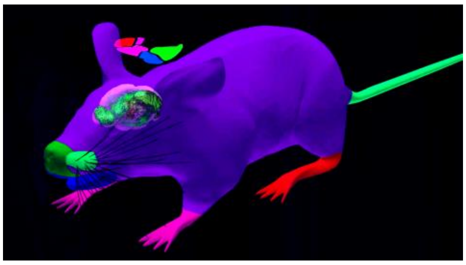


Figure 9 - Virtual mouse body (source HBP)

In the other hand, the Virtual Physiological Human (VPH[11]; 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.

As one of many examples of the data volumes to be handled by the HBP, consider BigBrain, an Ultrahigh-Resolution 3D Human Brain Model derived from experimental data. The volume of a human cerebral cortex is ~7500 times larger than a mouse cortex, and the amount of white matter is 53,000 times larger in humans than in mice. A recently published data set of the digitized mouse brain with 1-mm resolution has a total amount of uncompressed volume data of 8 TByte. The creation of a volume with similar spatial resolution for the human brain would result in ~21,000 TByte. The interactive exploration (as opposed to simple storage) of such a data set is beyond the capacities of current computing. Thus, among other methodological problems, data processing becomes a major challenge for any project aiming at the reconstruction of a human brain at cellular resolution.

In mid 2014, few month after the start of HBP, a working group of the US National Institutes of Health (NIH) published[12] a ten-year plan – Brain 2025: A Scientific Vision – for the agency's portion of

[9] http://apps.fz-juelich.de/hbp-pcp/index.php/Home

[10] 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

[11] http://www.vph-institute.org

[12] http://www.braininitiative.nih.gov/2025/index.htm

the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative. The plan's recommended budget would quadruple the current allocation, providing a $4.5 billion investment over 10 years, beginning in fiscal year 2016. The project, which aims to map all activity in the human brain, was initially given a $1 billion budget to be spent over 12 years.

US officials have compared the BRAIN initiative with the Human Genome Project, which had a similarly high price tag. The 10-year project that resulted in the first human genome being sequenced in 2003 cost $3 billion.

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[13]. 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."

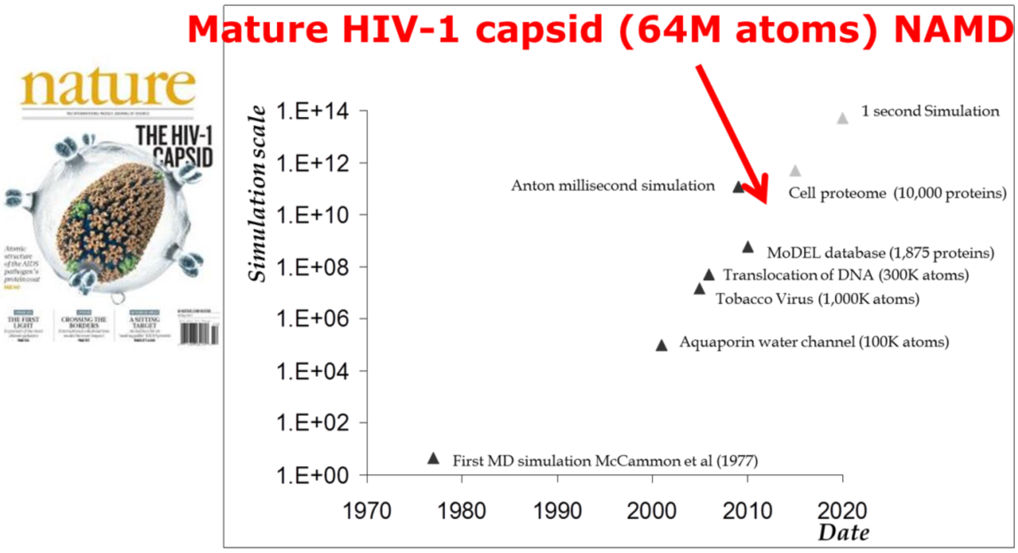
The simulations revealed that the HIV capsid contained 216 protein hexagons and 12 protein pentagons arranged just as the experimental data had indicated. The proteins that composed these pentagons and hexagons were all identical, and yet the angles of attachment between them varied from one region of 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[14]).

[13] Nature 497, 643–646

[14] 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 10 - 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

5. Conclusions

This final report, the synthesis of the activity of the applications working groups, is again emphazing the leading position of Europe in the development and the use of scientific applications used both by academia and industry.

During the period the experts reported major advances in the rewriting or the scaling out of applications (to more than 100k cores) and scientific breakthroughs using PRACE or Incite resources, the gap analysis showed that the roadmaps expressed at the end of EESI1 are still valid even if on some cases they have been shifted by 1 or 2 years (same applies to the Exascale roadmap shifted from 2020 to 2022). New roadmaps and needs have been reported especially in the field of automotive (combustion) and disruptive approaches (connected cars, composite materials, parallelisation in time, …).

Experts of WP3 have been also very active into the development of the 8 Centers of Excellence which have been granted by the European Commission mid 2015. With this third pillar now formed the Exascale HPC of the Commission is now in place alongside with the technology pillar (represented by ETP4HPC) and the infrastructure pillar (represented by PRACE). In order to stay competitive worldwide with the US (and the newly National Strategic Computing Initiative announced by Pres. Obama end of July 2015) or with Japan or Asia where Exascale strategies are also committed, these pillars need to receive significant, long-term and balanced funding from EC, member states and interested stakeholders.

This report highlighted also the convergence between HPC and big data generated from simulations or observations, resulting to a need to develop new techniques for in situ / in transit data analytics and smart tools for automatic pattern/shape/feature recognition/tracking on massive amount of data. This deluge of data coming from observations, computations from large-scale single or coupled multiscale/multiphysics ensemble simulations will stress the need to develop rapidly a framework for assessing and quantifying uncertainties. In this domain as well as the domain of scalable couplers, Europe owns already strong assets and with an appropriate level of funding it could be possible to develop frameworks which could become standards used worldwide;

Finally, the working groups start to think about disruptive approaches and launched a joint survey with NAFEMS across ISV which highlight the fact that efforts will need to be done rapidly for engaging this community into the porting and the scaling of their applications on (Pre)exascale platforms in order to stay competitive. At the same time Europe hold also a wide number of open source projects which start to raise appropriate level of TRL for being used on very important scientific and industrial projects.

All of these efforts will be continued into a more inclusive approach in the field of the EXDCI (European eXtreme Data and Computing Initiative) project, leaded by PRACE with ETP4HPC and the EESI community. This project started in September 2015 has the goal to develop and animate the European HPC Ecosystem. One of its main actions will allow a synchronised roadmapping exercise between technology and applications leading to the update of the Strategic Research Agenda (SRA) of ETP4HPC and the PRACE SSC Scientific Case. The EXDCI' workpackage dedicated to applications roadmap will incorporate heavily the contribution of the newly created CoE and will continue to monitor that the EESI2 recommendations will be incorporated into future calls of the European Commission.

The authors of this report want again to thank all the 43 experts of this workpackage for their time and their contribution during these 30 months.

6. Annex 1 – List of WP2 experts

6.1 List of experts of WG3.1

EESI2 - WG 3.1	**Industrial and Engineering applications**			
Name	Organization	Email	Country	Area of Expertise
Stéphane Requena (Chair)	PRACE/GENCI	stephane.requena@genci.fr	FR	HPC
Ange Caruso (Vice Chair)	EDF	Ange.caruso@edf.fr	FR	Nuclear Energy
Norbert Kroll (Vice Chair)	DLR	Norbert.kroll@dlr.de	DE	CFD, aeronautics
Philippe Ricoux	TOTAL	Philippe.ricoux@total.fr	FR	Oil & Gas
Thierry Poinsot	CERFACS	Thierry.poinsot@cerfacs.fr	FR	Combustion, CFD
Eric Chaput	Airbus	Eric.chaput@airbus.com	FR	Flight Physics, Aeronautics
Heinz Pitsch	University of Aachen	h.pitsch@itv.rwt-aachen.de	DE	Automotive, CFD, combustion
Lee Margetts	NAFEMS / University of Manchester	Lee.margretts@manchester.ac.uk	UK	ISV
Yves Tourbier	Renault	Yves.tourbier@renault.com	FR	Automotive
Vladimir Belsky	Dassault Systèmes	Vladimir.belsky@3ds.com	USA	ISV

6.2 List of experts of WG3.2

EESI2 - WG 3.2	**WCES (Weather, Climatology and solid Earth Sciences)**			
Name	Organization	Email	Country	Area of Expertise
Giovanni Aloisio (Chair)	UNIV. SALENTO & CMCC	giovanni.aloisio@unisalento.it	IT	Exascale Computing
Massimo Cocco	INGV	massimo.cocco@ingv.it	IT	Solid Earth, Seismology
Sandro Fiore	UNIV. SALENTO & CMCC	sandro.fiore@unisalento.it	IT	Scientific Data Management
Pierre-Philippe Mathieu	ESA/ESRIN	pierre.philippe.mathieu@esa.int	IT	Earth Observation Science and Appl.
Jean-Claude André (Co-Chair)	Jca Consultance & Analyse	jc_andre@sfr.fr	FR	Earth-system Modelling
Sylvie Joussaume	IPSL	sylvie.joussaume@lsce.ipsl.fr	FR	Earth-system Modelling
Sophie Valcke	CERFACS	valcke@cerfacs.fr	FR	Coupled Climate Models
Jean-Pierre Vilotte	IPGP	vilotte@ipgp.jussieu.fr	FR	Solid Earth, Seismology
Marie-Alice Foujols	IPSL	marie-alice.foujols@ipsl.jussieu.fr	FR	Earth-system Modelling
Sebastien Masson	LOCEAN - IPSL/UPMC	smasson@locean-ipsl.upmc.fr	FR	Oceanography
Bryan Lawrence	NCAS, STFC, UNIV.READING	bryan.lawrence@stfc.ac.uk	UK	Earth-system Modelling
Christopher Maynard	METOFFICE	christopher.maynard@metoffice.gov.uk	UK	Exascale Computing
Joachim Biercamp	DKRZ	biercamp@dkrz.de	DE	Exascale Computing
Heiner Igel	LMU	heiner.igel@geophysik.uni-muenchen.de	DE	Solid Earth, Geophysics
Colin Jones	SMHI	colin.jones@smhi.se	SE	Earth-system Modelling

6.3 List of experts of WG3.3

In June 2013, the Vice Chair of the working group 3.3 has changed from Romain Teyssier (CEA and ETHZ) to Alan Sacha Brun (CEA).

EESI2 - WG 3.3	**Fundamental Physics**			
Name	Organization	Email	Country	Area of Expertise
Godehard Sutmann (Chair)	JSC	g.sutmann@fz-juelich.de	DE	Quantum Chemistry
Alan Sacha Brun (Vice Chair)	CEA	sacha.brun@cea.fr	FR	Astrophysics
Thierry Deutsch	CEA	thierry.deutsch@cea.fr	FR	Quantum Chemistry
Nicola Marzari	Univ Oxford	nicola.marzari@materials.ox.ac.uk	UK	Material Sciences
Maurizio Ottaviani	CEA	Maurizio.Ottaviani@cea.fr	FR	Fusion
Volker Springel	Garching, MPI Astrophysik	volker@mpa-garching.mpg.de	DE	Astrophysics
Mike Payne	University of Cambridge	mcp1@cam.ac.uk	UK	Quantum Chemistry
Louis Silva	Universidade Tecnica de Lisboa	luis.silva@ist.utl.pt	PT	Laser Plasma Interaction
Ulf Meißner	University of Bonn /FZ Jülich	meissner@hiskp.uni-bonn.de	DE	Hadron/Nuclear Physics

6.4 List of experts of WG3.4

EESI2 - WG 3.4	**Life Sciences and Health**			
Name	Organization	Email	Country	Area of Expertise
Modesto Orozco (Chair)	IRB Barcelona	modesto@mmb.pcb.ub.es	SP	Modeling and Bioinformatics
Charles Laughton (Vice Chair)	U Nottingham	Charles.Laughton@nottingham.ac.uk	UK	Biochemistry & Drug discovery
Paolo Carloni	German Research School	p.carloni@grs-sim.de	DE	Computational Biophysics
Peter Coveney	University College London	P.V.Coveney@ucl.ac.uk	UK	Integrative medicine
Alfonso Valencia	CNIO	avalencia@cnio.es	SP	Biological Sciences
Richard Lavery	University of Lyon	richard.lavery@ibcp.fr	FR	Simulation
Erik Lindhal	Royal Institute of Technology (KTH)	lindahl@cbr.su.se	SE	Molecular Dynamics
Henry Markram	École Polytechnique Fédérale de Lausanne	henry.markram@epfl.ch	CH	Organ simulation
Darren Green	GSK	darren.vs.green@gsk.com	UK	Drug Discovery and HPC
Supporting Experts				
Shantenu Jha	Rutgers	shantenu.jha@rutgers.edu	US	
Martin Guest	Cardiff	GuestMF@cardiff.ac.uk	UK	

6.5 List of experts of WG3.5

EESI2 - WG 3.5	**Disruptive Applications**			
Name	Organization	Email	Country	Area of Expertise

Godehard Sutmann (Chair)	JSC	g.sutmann@fz-juelich.de	DE	Quantum Chemistry
Ulrich Rude	U. Erlangen/FAU	uli.ruede@fau.de	DE	Computer Science
Jean Claude André	JCA Consulance	jc_andre@sfr.fr	FR	Earth-system Modelling
Matthias Bolten	University of Wuppertal	bolten@math.uni-wuppertal.de	DE	Applied Computer Science
Yvon Maday	University of Paris 6	maday@ann.jussieu.fr	FR	Applied mathematics

7. Annex 2 – Report of the joint NAFEMS / EESI2 survey

