Deliverable D5.2

WP5 Second Intermediate Report

Cross Cutting Issues Working Groups

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

Due date of deliverable: 30/04/2015

Actual submission date: 31/03/2015

Publication date: 31/03/2015

Start date of project: 1 September 2012 Duration: 30 months

Name of lead contractor for this deliverable: PRACE-CINECA

Authors: Giovanni Erbacci (CINECA), Francois Bodin (CAPS), Anne-Laure Popelin (EDF), Simon McIntosh-Smith (Bristol Univerity), Thomas Ludwig (DKRZ), Franck Cappello (INRIA), Carlo Cavazzoni (CINECA), Marie-Christine Sawlay (Intel).

Name of reviewers for this deliverable:

Abstract: This is the second intermediate report of EESI2 WP5 Cross Cutting Issues Work Groups. The document reports the outcome of the second year of activity of the five working groups identified in EESI-2-WP5 and presents the recommendations.

Revision 1.0

Project co-funded by the European Commission within the Seventh Framework Programme (FP7/2007-2013)		
Dissemination Level		
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

List of Figures

Glossary

Abbreviation / acronym	Description
ABFT	Algorithm-Based Fault Tolerance for Linear Algebra
ADS	Adaptive Directional Stratification
CDF	Cumulative Distribution Function
CPU	Central Processing Unit
DOE	Designs of Experiments
DAKOTA	Design Analysis Kit for Optimization and Terascale Applications
DARPA	Defense Advanced Research Projects Agency
DDDC	Double Device Data Correction
DDDC+1	memory enhanced Double Device Data Correction
DDR	Double Data Rate
DIMM	Dual In-line Memory Module
DRAM	Dynamic Random Access Memory
ECC	Error Correcting Code
EEHPC	Energy efficient HPC
EESI	European Exascale Software Initiative
EIOW	Exascale I/O Workgroup Middleware
ESREDA	European Safety, REliability and Data Association
FEM	Finite Element Methods
FeRAM	Ferroelectric RAM
FFT	Fast Fourier Transform
FORM	First Order Reliability Method
GPU	Graphics Processing Unit
HPC	High Performance Computing
IESP	International Exascale Software Project
IO	Input Output
I/O	Input Output
LARS	Least-Angle Regression
MPI	Message Passing Interface
NTV	Near-Threshold Voltage
NVRAM	Non-Volatile RAM
PCE	Parametric Constrain Evaluation
PCM	Phase-Change Material
PCRAM	Phase-Change RAM
PDE	Partial Differential Equation
RAID	Redundant Array of Inexpensive Disks

RAM	Random Access Memory
RMA	Remote Memory Access
ReRAM	Resistive RAM
RBD	Random Balance Designs
RBM	Reduced Basis Models
R&D	Research & Development
SDCs	silent data corruptions
SPMD	Single Program, Multiple Data
ST-MRAM	Spin-Torque Magnetoresistive RAM
STTRAM	Spin Torque-Transfer RAM
TBB	Task Building Blocks
VVUQ	Verification, Validation, and Uncertainty of Quantification

1. Executive Summary

This document reports the second year of activity of EESI 2 WP5 “Cross cutting issues Work Groups” which had the objective to better focus and evolve the activity addressed during the first year of the EESI2 Project. The Cross cutting issues address themes transversal to the different activities from applications to technologies making the activity in WP5 synergic to the activity in WP3 “ Applications” and WP4.”Enabling Technologies”. The five working groups (WG) on cross cutting issues have addressed the following actions:

WG 5.1 Data Management and Exploration: One of the major challenge of Exascale applications addressing scientific discovery nowadays. The issue is central for the organization of the scientific discovery workflow and aims to set up actions to address end-to-end techniques for efficient disruptive I/O and data analysis, involving the full life-cycle of data.

WG 5.2 Uncertainties (UQ/V&V): Science in the Exascale era involves computational models and applications which are multidisciplinary and complex, involving a huge amount of parameters and variables, so the verification, validation and uncertainty quantification of computer models' results becomes fundamental both for industry and academia. Identify methodologies and enhance tools for the analysis of these uncertainty sources, is fundamental for the exploitation of Exascale applications.

WG 5.3 Power & Performance: The power monitoring and power management at all levels of the system architecture, addressing energy efficient performance of applications, is a crucial issue to address in the Exascale era. Guidelines have been addressed in this area underlining the need of standards in parallel with the urgency of formation of professional HPC developers experts in green programming methodologies.

WG 5.4 Resilience: Robust fault tolerance protocols as well as performing checkpoint/restart methods, to increase the efficiency of Exascale systems, are becoming urgent to manage the fore coming systems with millions of cores. The activity in this WG continued the gap analysis between existing reports and projection about the resilience challenge for Exascale simulation, started in the first year of the Project activity. The set of recommendations based on this gap analysis have been better focalized and an holistic approach for resilience has been recommended.

WG 5.5 Disruptive Technologies: The roadmap toward Exascale and beyond will be guided and modeled by disruption in semiconductor technologies, I/O and memory technologies, cooling technologies and facility management, networking and data transfer technologies. The disruptive technologies analysed last year have been further investigated and the software implication have been considered to address specific recommendations for supporting programmability, efficiency and productivity of tools and applications, energy aware, at Exascale.

This document represents an update of the deliverable D5.1, and focus on the activity addressed in each WG during the second year of EESI2 action. When strong technological issues are involved, significant progresses cannot be observed and evaluated just in the span of a year, however the activity registered some progresses, as reported in this deliverable, and was of paramount importance to define and formulate the recommendations presented by EESI2 in July 2014. Mainly in the Pillars *Tools & Programming Models* and *Data Centric Approaches.*

2. WG 5.1 Data Management and Exploration

2.1 Introduction

Workgroup 5.1 has been addressing "Data management and exploration" in Exascale applications viewed as the organization of the scientific discovery workflow.

Efficiency at Exascale level requires breaking with the traditional scientific workflow where simulation data are stored on disk for later analysis. This disruption comes in sync with new memory technologies, new photonic networks as well as the decreasing cost of transistors. For instance new non-volatile memories (e.g. Memristors) hold the promise of providing persistent memories close to the CPU that are fast, large, energy efficient and at a reasonable cost. On the software side, big data and other in memory computing technologies may be providing new solutions to help scientists facing the coming deluge of data. Holistic approaches considering all data cycles from sensors capture to visualization, encompassing simulation, code coupling, in-situ, pre and post analysis can guarantee that no bottlenecks are introduced in the scientific discovery process. In particular, it is strongly wished that new systems simplify human-in-the-loop workflows.

2.1 Data Centric View of Exascale Applications

Data management is at the core of the design of Exascale applications, as widely analysed during the first year of WG 5.1 activity and reported in Deliverable 5.1 [1] .

As illustrated in **Figure 1**, data may follow any paths (blue arrows) in the ecosystem, each component having its own performance profile, quality of service and cost. For instance while HPC technology optimizes writing in parallel the data, data mining techniques favour the reading. Choosing to use one or the other technology must be carefully planned according to a global view of the workflow.

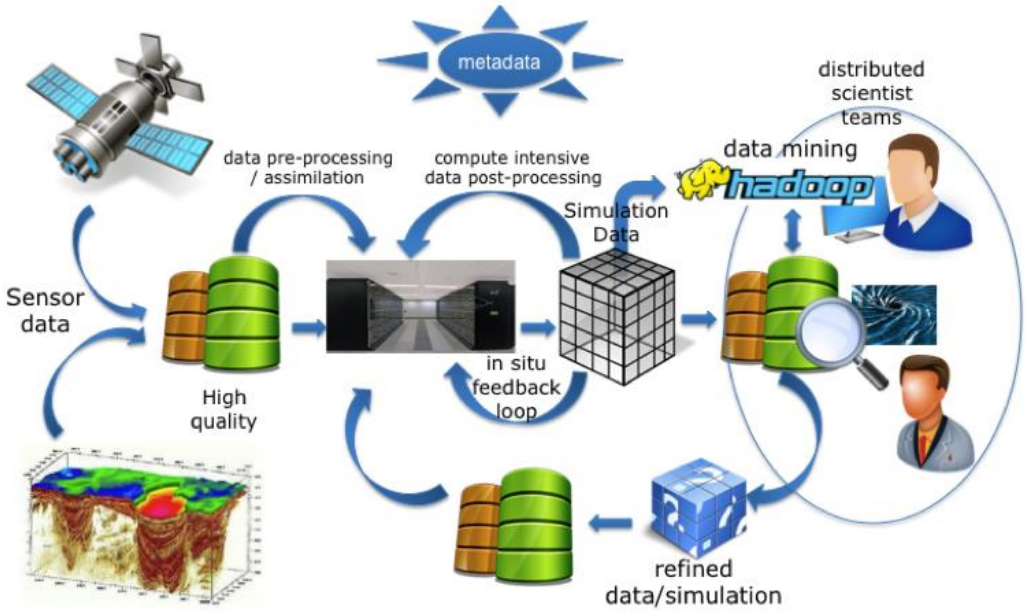

Figure 1: ***complex work flow of Exascale applications***

On one hand one must consider the rising price of IO systems. On the other hand as a deluge of data is to be expected synergies between big data and traditional HPC techniques has to be well thought-out. Data categories are also an important concern. For instance, data from sensors cannot be regenerated and must be stored safely while some data produced by simulation may be easier to re-compute when combine with in-situ data processing techniques. Each data must be stored and organized to use the proper resources. As well, metadata and provenance must be kept consistent all the way. This likely will strongly disrupt current practices.

Designing an Exascale application that make rational and efficient uses of communication, compute, storage resources requires engineer skills that are currently in shortage or just not available to scientists. New best practices will have to be defined and implemented. They will very likely require setting up interdisciplinary support team capable of addressing extreme parallelism, fault tolerance and IO issues.

The deluge of data requires new data analysis techniques. Big data technology may provide new disruptive methods for such task. These techniques need to be extended to take advantage of highly scalable parallel infrastructure. This may be a return contribution of HPC to the big data field. Behind this topic lies many complex and holistic issues such as: serialization/deserialization of data, design of data structures able to cope with highly asynchronous execution as well as compute / IO activities interleaving. More generally, data mining techniques must be extended to fit the file formats used in HPC (e.g. HDF5, netCFD) and bridges must be established between HPC and big data usual formats.

Metadata management and specification is also a critical challenge. They are keys elements in the science discovery process. Their design is particularly important to obtain a consistent end-to-end use of the data. Furthermore, they impact on sharing policy management implementation (e.g. at the core of the decision process concerning data to be set public, what storage migration, etc.).

Analysis and visualization of data produced by large-scale simulations are often sidelined in favor of pure computation performance. As we foresee Exascale systems in the next decade, the offline analysis approach shows its limits: more and more scientists see the scalability of their simulations dropping because of unmatched computation and I/O performance as well as higher I/O variability. However, in-situ[1] approaches (potentially more efficient) have difficulties in getting accepted, as scientists fear to dive into fundamental code changes in a simulation they have used for years. Defining the right tradeoff here is a challenge. Also related to the same limitation in I/O performance, HPC scientists predict fundamental changes in the way I/O and data management will be handled in the near future. In particular, the heterogeneous processor environment and memory hierarchy of the new platforms, together with the increasing use of GPU and accelerators, open new alternatives for data analysis.

This topic cannot be viewed only under the technology angle. Indeed, designing the applications requires finding tradeoffs between in-situ vs. ex-situ processing, selecting data format, access policy, data relocation, format changes, etc. These tradeoffs are not only driven by technology and performance but also by the ecosystem exposed to the researchers. Furthermore, It is important to note that a global efficient use of the Exascale resources can be contradictory with the objectives of individual research teams. Understanding the full cycle of data is probably the most important question to drive Exascale technology development.

2.2 New Technologies, New Challenges

The Exascale goal stimulates the seeking out of new solutions and requires the exploiting of new technologies being brought to market such as photonics communications, energy optimized processors, chip stacking, non-volatile memories (NVM), etc. In the following we discuss the case of NVM as this technology likely will introduce disruption in most of the elements of the software stack of an Exascale system and the data streams. NVM hold the promise of providing persistent memories close to the CPU that are fast, large and at a reasonable cost.

NVM have been announced since a long time (e.g. 2008) but still now the freely available information is not clear about the characteristics of such technologies. For instance, if we consider Resistive RAM (RRAM), also denominated Memristor by HP, different sources of information are still contradictory on the capabilities of such memories in terms of writing latency.

Nevertheless, despite the late arrival of this technology on the market, it cannot be ignored in the race to Exascale computers. HP that hopes to deliver operational systems with Memristor technology by 2018 has made the most aggressive announcement.

[1] See the EESI recommendation on this specific topic.

Figure 2 shows a potential instance of NVM based compute node. This node integrates a set of NVRAM stacks using a set of photonic connections. Current numbers from HP indicates 256 GB SoC as a start.

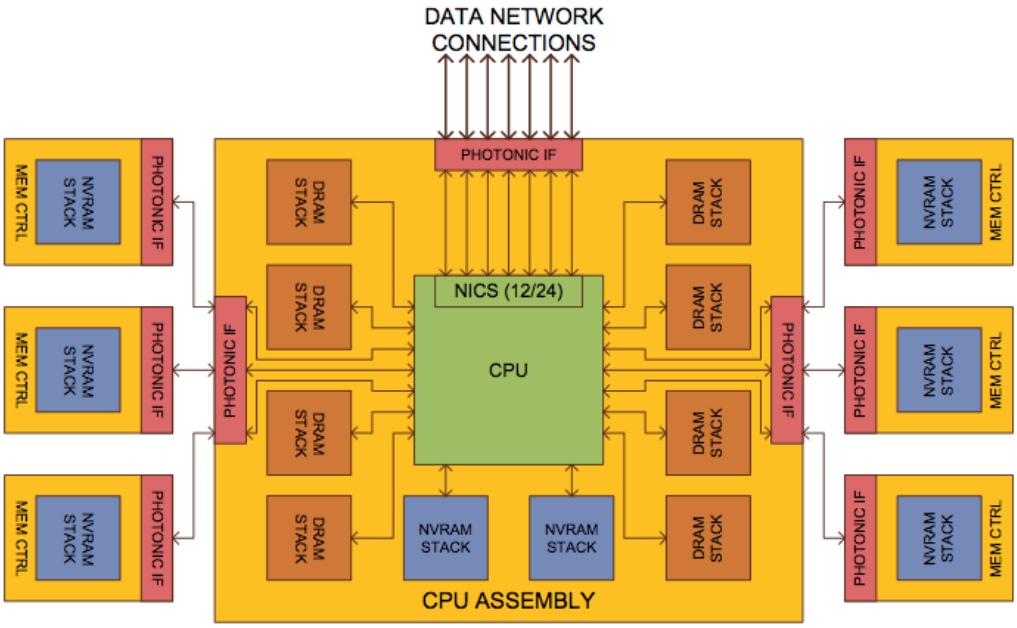

Figure 2: HP Strawman Exascale System.

The characteristics of such technology to be expected by HP are the following:

1) Node NVRAM capacity of many time large than DRAM capacity
2) Scaling down to less than 10 nm width per cell, ~ 32 Gbyte/cm2/layer by 2018
3) Scaling up to multiple (≥ 8) layers on chip, ~ 0.25 Tbyte/cm2/chip by 2018
4) Truly nonvolatile – many, many years
5) Random access at byte level
6) Fast cell write and erase (~ nanosec)
7) Low energy cell write and erase (~ picoJ)
8) Endurance > 10^{10} cycles, expectation is to exceed the 10^{18} cycles of professional DRAMs

It is important to note that the NVM technology is leveraged thanks to two other technologies: 3D stacking and on-chip photonic.

Here we have presented, as an instance of NVM based system, the one proposed by HP, but it is notable to underline that other manufacturers are engaged in similar technological efforts.

From the application point of view, having a large memory, with close to DRAM performance and persistence, is likely to introduce a revolution in application design.

As current roadmaps foresee that large NVMs attached to compute nodes will reach the market by 2018-2020, it is urgent to handle the software issues involved with this technology that very likely will be a disruption in the race to Exascale systems with wide implications on the design of Exascale applications.

2.3 Impact on Application Development

Application development for Exascale systems is of a rare complexity. Complexity in scalability and in roadmapping the software. On one hand it is best if legacy codes can be reused, on the other hand it is likely that many codes will have to be deeply re-designed / re-developed. *Figure 3* illustrates this tradeoff. Domain specific approaches may be able to hide complexity to users but as they are more

specific they address a smaller community. In the end, a tradeoff must be made between development cost (including the tools, API, maintenance, etc.) and the potential users base. As application software moves much slower than hardware technology we believe that anticipation is extremely crucial.

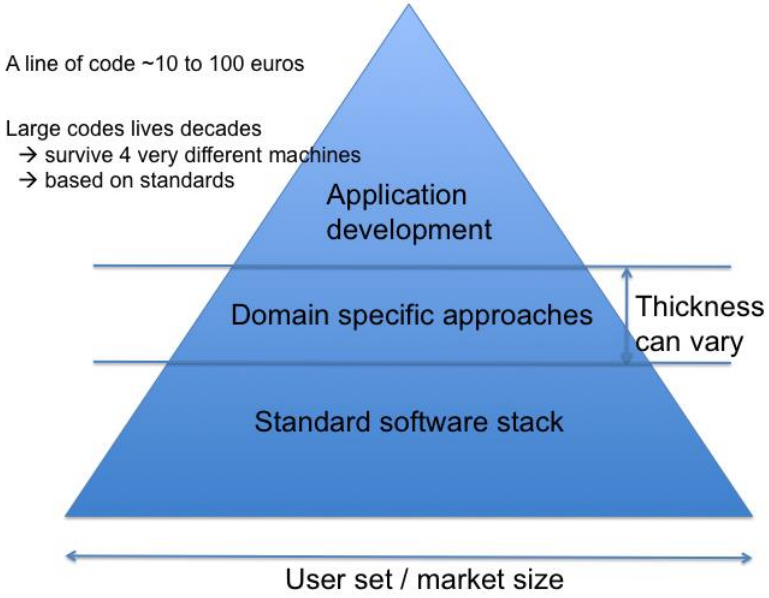

Figure 3: Domain specific approach trade-off.

The reminder of this section analyses the impact of the data centric approach on applications as a consequence of the new NVM technology.

2.3.1 Application Data Flow Design

A first tradeoff to deal with is *what data to output?* For instance, it is frequent that adding more in-situ computation will negatively impact the efficiency of the simulation part of the applications. However, if this later provides for a faster and simpler analysis of the data, it will be worthwhile to pay the corresponding penalty. It is important to remember that human time, even in Exascale environment remains the most expensive resources.

The data life cycle must be clearly understood to allow building an indexing and typology of the data that promote an efficient use of the different storage systems. The most reliable storage must only be used in a cost-effective manner. For instance, it is necessary to distinguish the needs in pre- and post-processing so that the right technology could be used. Typically, three cases can be distinguished: Post processing very large, out of memory data that requires powerful computing power (e.g. out of memory FFTs like); In-memory processing of mid-size chuncks of data (e.g. can benefit of Hadoop technology); Complex search with associative patterns over very large, out of memory data. These techniques, to be fully exploited in an HPC context, will require disruptive practices.

Applications must optimise the use of IO bandwidth thanks to interleaving compute and data transfers in a manner known/understood by the system. This requires new programming methods and tools.

2.3.2 Software stack design

The data management throughout the entire software stack is likely to be modified extensively when new memory technologies will be introduced. This will have to be taken into account in the development of the applications as well as the supporting software stack. In the following we list non-exhaustively some crucial aspects that will affect the software stack as a consequence of the new NVM technologies.

Compute vs storage: It has been anticipated that compute power will grow faster than storage capacity for HPC machine (number of core vs node main memory size). The availability of NVM may

change this trend (it is important to note that the low energy consumption of NVM allows to provide very large space). However, as more memory is available inside the node, getting the data out may stress out the IO system.

IO stack: Speed of expected NVM is so fast that it requires revisiting the landscape of storage software stack: current storage software's latency (and corresponding energy) will become the main cost. In the end the status of NVM may neither be the one devoted to RAM nor the one devoted to usual storage. New application development will define what is expected from this new hardware capability.

Resilience: Constraint on the implementation of resilience can be fully revisited. Memristor performance may be fast enough to allow saving the processor state frequently enough in such manner that transient errors are completely handled at system lowest level and transparent to users. This means redirecting the research effort at higher-level functionalities, closer to the application.

Programming API: If NVM are to be used for application development it must be exposed to programmer with a standard efficient API. The list of challenges in designing this API is quite long since it must address data persistence (e.g. dealing with issues such as pointer address), resource management (e.g. NVM allocation and IO organization), energy management (e.g. since the data is persistent some part of a system can be turned off while participating temporarily), data sharing between nodes (e.g. PGAS, peer-to-peer exchanges) and performance.

In-situ analysis, pre/post processing: NVM technology carries the potential to turn compute centric-machine into data-centric systems. As a consequence the data analysis and processing tasks that where envision for a different kind of system (e.g. big data machine) can be efficiently implemented close to the compute part. This opens many opportunities for developing a new kind of scientific applications more data oriented. It should be noted that thanks to photonic-based network, it is possible to aggregate node NVM to built a very large memory space dedicated to data mining and other analysis tasks. Visualizing the data in such context is still to be clarified.

Code coupling: Code coupling may be made more efficient as well as simpler to implement. Current libraries for code coupling will need to be revisited to take into account this new storage. Furthermore, the coupling "frequency" might also have to be reconsidered at numerical scheme level[2].

Compiler and runtime technology: Because of energy management, heterogeneous hardware and system configuration compiler research has been studying auto-tuning and runtime libraries to adapt to runtime changing context (e.g. because some core are out, data size is different, etc.). This technique usually relies on code versioning, specialization and run-time code generation. All these require run-time performance analysis and code tuning during a discovering phase. Keeping local data on node over application execution (thanks to persistence) would help to reduce the cost of this phase that in many cases impact significantly performance.

Debugging, performance tools: performance tools such as Tau, Vampire, Paraver, Scalasca are based on tracing events on each node. With the increase of parallel activity storing the huge amount of events is challenging at Exascale level. The ability to store and process locality as much as keep performance history on each node is likely to help redesigning this tool to handle the massive parallelism (e.g. post-mortem analysis). Similar studies are also needed for designing the next generation for debuggers.

2.4 Recommendations

Migrating or designing a new code for the Exascale is an extremely challenging task that requires making multiple algorithmic and technological choices in an uncertain environment. New memory technologies have the potential of combining in a single system the compute-centric tasks with the data-centric tasks of modern HPC application facing a deluge of data.

[2] See EESI recommendation on this specific topic.

End-to-end data life cycle poses many challenges due to combining the technology, human resources and the ecosystem economy. As a consequence of the previous considerations, system and programming environment designers should provide to application developers efficient and standard APIs (or other methods), and corresponding best practices, to drive the hierarchies of storage to use and to describe more about the exploitation of the data.

It is urgent to encourage the community to form multi-disciplinary research groups capable of handling the complete set of concepts necessary to design data centric approaches to Exascale computing. It is particularly important to also integrate ecosystem and economical issues. For instance, energy cost is a growing concern that may lead to move from "*charge by core-hours*" to "*charge by kilowatt-hours*" in order to capture the entire complexity of a data centric approach to Exascale.

3. WG 5.2 Uncertainties (UQ / V&V)

3.1 Introduction

The ability to simulate very complex and inter-disciplinarily phenomena, taking into account the effect of a wide number of input parameters it is a chance which will heavily grow in the Exascale era, but, at the same time, we must point out that the quantitative uncertainty assessment of the results becomes a fundamental issue for assuring the credibility of computer model based studies, and represents a challenge too.

The most challenging point is to bridge the cultural gap between a traditional scientific and engineering deterministic viewpoint and the probabilistic and statistical approach which considers the result of a model as an "uncertain" variable. The step forward is to develop and to spread in the scientific and engineering community an enhanced unified framework for model verification & validation and uncertainty propagation, what is commonly called *VVUQ*.

As underlined by the WG 5.2 activity last year, this unified framework shall need at the same time:

- multidisciplinary skilled teams (statistics & probability, numerical analysis, PDE, physicians),
- high computational power, as the statistical methods for calibration and validation need to evaluate several times a (possibly) costly numerical code.

HPC and uncertainty quantification have a two-sided relationship. On the one hand, the ever increasing size of the computational data leads to increasing sources of uncertainties, due to the accumulation of numerical errors. On the other hand, HPC gives access to computational power that can be used to tackle explicitly the evaluation of uncertainties, be it by embedded methods or by design of experiments. The activity in WG 5.2, aims at exploring these different aspects in the relationship between uncertainties and HPC.

3.2 Second year activity

Several workshops and dissemination activities were organized around a circle of identified experts:

- Stefano Tarantola, JRC-ISPRA, Italy
- Christophe Prud'homme, University of Strasbourg, France
- Olivier Le Maître, LIMSI, Duke University, US
- Renaud Barate, EDF R&D, France
- Bertrand Iooss, EDF R&D, France
- Fabrice Gaudier, CEA, France

In the first year of EESI2 program, the main outcomes, summarized in the Deliverable D 5.1 [1] were:

- an identification of the main methodologies for the analysis of these uncertainty sources;
- a presentation of several software tools related to uncertainty analysis;
- some guidelines for the evolutions required both in tools and in methodologies for exploitation of Exaflop machines.

In this second year of EESI2 project, the WG 5.2 activities have updated the feedback from its experts, adding some newly met specialists in this field:

- Laurence Viry (University of Grenoble, France)
- Bruno Sudret and Stefano Marelli (ETH Zürich)
- Eric Phipps (SANDIA National Lab, Albuquerque, New Mexico, US)

The main outcomes from these new contributions to the global work about HPC and VVUQ challenges are reported in the next sections. The synthesis of the activity done allowed the proposition of specific recommendations on this theme. Furthermore, the WG 5.2 EESI2 partners in 2014 continued to enhance dissemination activities, focusing on the need in VVUQ of multidisciplinary skills in addition to the expertise in the specific field of application of the numerical code.

3.2.1 Some UQ challenges in HPC

During the SIAM Conference on Uncertainty Quantification (March 31 – April 3, 2014, Savannah, USA, http://www.siam.org/meetings/uq14/), Eric Phipps has made a tutorial on the UQ challenges in HPC. The usual suspects in UQ can be the dimension input space, the non smoothness of the model responses and the need to accurately estimate some rare events probabilities. HPC bring some solutions to these problems. Two subjects have been emphasized by E. Phipps:

- A critical component of predictive simulation is computing derivatives of simulation responses with respect to model states and parameters. Such information is needed for steady-state nonlinear solves, implicit time integration, stability analysis. Moreover, derivatives can be useful for sensitivity analysis, optimization, meta-model approximation, Bayesian inference sampling, error estimation, and uncertainty quantification. However hand-coding derivatives is time-consuming, error prone, and difficult to verify, particularly for parameter derivatives, adjoints, and higher derivatives.

  Automatic differentiation (AD) is a technique for computing analytic derivatives in simulation codes without hand-coding the derivative computation itself, and is based on simple mathematical and computer science principles. A web site is devoted to this subject: http://www.autodiff.org/. A lot of AD software exists since a long time, as TAPËNADE, ADOL-C, ADIFOR, …etc., but are mainly restricted to Fortran and C codes. Interpreted languages as Matlab and python can also be considered in AD.

  AD for C++ code has been considered for a long time as a serious scientific challenge. E. Phipps has presented Sacado, a Trilinos package for automatic differentiation of C++ codes, being developed at Sandia. Sacado is designed for incorporation into large-scale C++ codes and leverages the C++ language itself to implement AD using operator overloading and templating. This approach has proven to be quite successful in the Charon application code by enabling rapid development of complex physics and providing accurate and efficient derivative calculations for advanced analysis techniques such as optimization and stability analysis. E. Phipps has explained the steps needed to incorporate these tools into large C++ simulation codes, using the Sacado::FEApp 1-D finite element example application as a demonstration example.
- The material aspects for HPC infrastructures become critical for UQ of petascale simulations (it leads to exascale computing needs). Schematically, the limits have been reached on the transistor sizes and the energy required for cooling. We then need to develop a parallelism hierarchy and organize the memory access (using tools as OpenMP and OpenACC). Improving linear algebra solvers will also allow to gain in memory.

3.2.2 An interesting formalization of HPC needs for UQ

We have met L. Viry during the Spring Research School «*Analyse de sensibilité, propagation d'incertitudes et exploration numérique de modèles en sciences de l'environnement*» (May 5 – May 7, 2014 Les Houches, France, http://aspen.forge.imag.fr/). L. Viry is an expert engineer on the HPC and UQ topic. She has provided a deep analysis about the trends in HPC and the subsequent problems for UQ. She has insisted on the meta-model solution in the context of HPC.

First, she proposes to define the following processes:

- P1: parameterization,
- P2: sampling,
- P3: distribution of model input data,
- P4: results recuperation,
- P5: estimation of quantity of interest,
- P6: metamodel building if necessary,
- P7: model evaluation,
- P8: metamodel evaluation.

Then, for each process, we have to describe the characteristics of the calculation:

- CPU time cost T_i of one evaluation of the numerical model,

- Storage volume V_i,
- Communication (data exchange) volume C_i,
- Environnement software, process complexity.

In the case of no meta-model, the problem can be formalized as described in*Figure 4* Figure 4

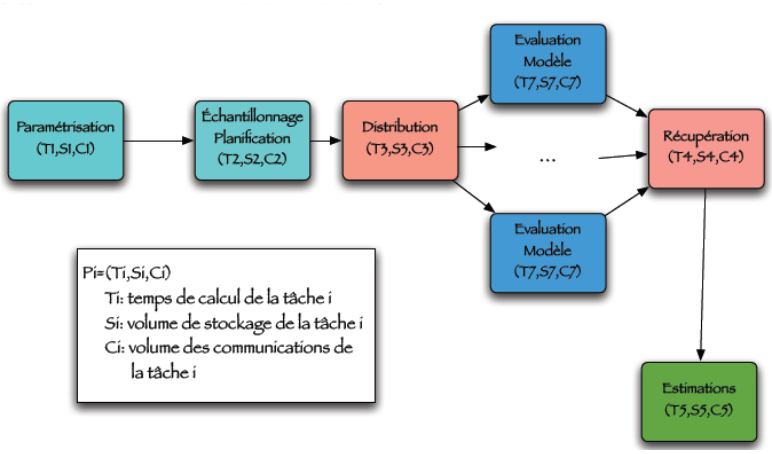

Figure 4: Formalization of HPC needs in UQ - Case without metamodel [Viry courtesy]

The analysis gives the following results (n is the number of model evaluations):

- Total cpu time cost: $T=T_1+T_2+T_3+nT_7+T_4+T_5$;
- Storage: S_1, S_2 and S_5 low bulky; S_7 depends on the model input (X) / output (Y) volume; S_4 depends on the volume of the quantities of interest ;
- Communication volume (for P_3, P_7, P_4): $C_3=n$ taille(X); $C_4=n$ taille(Y); C_7 depends on the numerical model.

In the case of a meta-model-based UQ, we have the problem formalization presented in Figure 5.

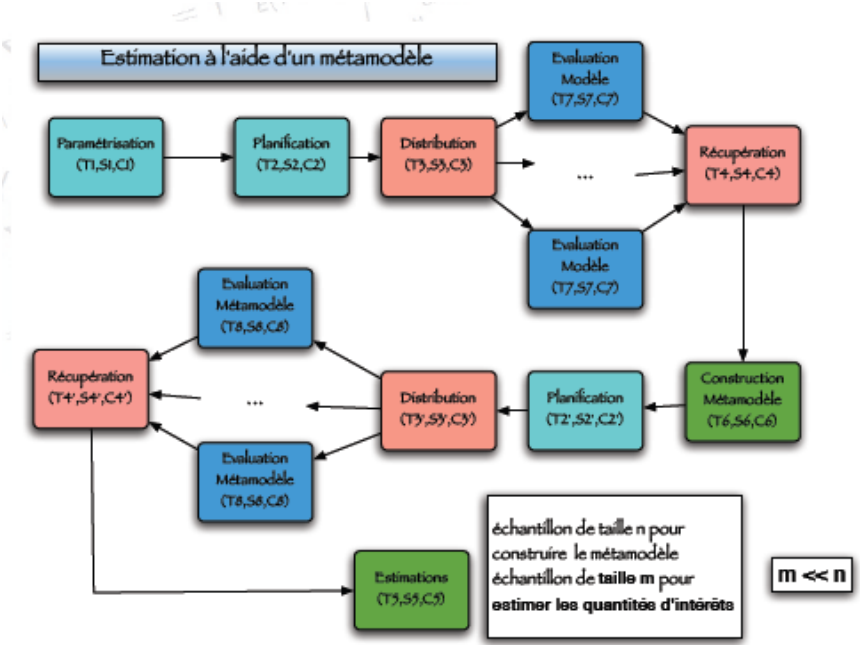

Figure 5: Formalization of HPC needs in metamodel-based UQ [Viry courtesy]

The analysis gives the following results (*n* is the number of model evaluations for building the meta-model and *m* is the number of meta-model evaluations):

- Total CPU time cost: $T=T_1+T_2+T_3+nT_7+T_4+T_6+T'_2+T'_3+mT_8+T'_4+T_5$;
- Storage: S_1, S_2, S'_2 and S_5 low bulky; S_7 depends on the model input (X) / output (Y) volume; S_4 and S'_4 depends on the volume of the quantities of interest;
- Communication volume (for P_3, P_7, P_4): $C_3=n$ taille(X); $C_4=n$ taille(Y); C_7 depends on the numerical model.

3.2.3 A new tool for uncertainty treatment in Matlab environment

B. Sudret and S. Marelli were met during the MASCOT-NUM 2014 meeting (April 23 – April 25, 2014, Zürich, Switzerland, http://www.ibk.ethz.ch/su/mascotnum2014/). They have developed a software, called UQLab (http://www.ibk.ethz.ch/su/research/uqlab_EN), for UQ in the Matlab environment, with some HPC objectives. We describe below this software.

UQ is an emerging field at the boundary of computer simulation-based engineering on the one hand, applied mathematics, statistics and probability theory on the other hand. This general formulation covers a vast field of approaches including, among others:

- structural reliability analysis, which aims at computing the probability of failure of a structure (or, more generally, of a system) w.r.t a given performance function and the probabilistic description of uncertain parameters;
- sensitivity analysis, which aims at determining the input parameters of a computational model whose uncertainty explains at best the system's performance variability;
- reliability-based design optimization, which aims at optimization systems under reliability constraints (e.g. minimize the system's mass/cost while ensuring a minimal – acceptable – probability of failure);
- Bayesian techniques for calibration and validating computer models w.r.t experiments.

In terms of research, contributions to this broad topic equally come from the engineering, statistics and applied mathematics communities, sometimes with their own vocabulary. In terms of computational tools, several tools are available (see http://www.gdr-mascotnum.fr/software.html). In Matlab, which is the most developed scientific platform in the industry, no tool covers the broad scope mentioned above.

The UQLab project aims at developing a Matlab-based architecture that allows one to use UQ algorithms in a distributed high-performance computing environment. The platform is designed as a central module providing low-level services in probabilistic modelling and parallel computing, onto which plug-in modules can be easily added. The foreseen modules include revisited algorithms for reliability analysis, meta-modelling techniques (polynomial chaos expansions, Kriging, support vector machines), sensitivity analysis (Sobol' indices), Markov Chain Monte Carlo methods, etc.

The platform will be used for different applications in uncertainty quantification and risk analysis in the fields of civil and mechanical engineering. When the project is mature it is intended to disseminate it at a larger scale by moving to an open-source development framework.

3.3 Recommendations

The activity in WG 5.2 has led to the recommendation *"Verification Validation and Uncertainty Quantification tools evolution for a better exploitation of Exascale capacities"* as part of the *Tools & Programming Models Pillar;* See [2]. The recommendation aims at preparing an unified European VVUQ package for Exascale computing by identifying and solving problems limiting usability of these tools on advanced HPC systems, and furthermore to facilitate the access to VVUQ techniques for the HPC community.

4. WG 5.3 Power & Performance

4.1 Introduction

In the quest to achieve Exascale systems in the 2020 timeframe, energy efficiency has become one of the primary challenges. Exascale designs projected unacceptably high power requirements in excess of 100 MW for each system, leading to a surge of research and development searching for breakthroughs in energy efficient hardware and software. Today, significant advances have been made in many areas, but there are many challenges still remaining that need to be addressed if we are to meet our goal of Exascale systems within a 20 MW power envelope. The activity in WG5.3 has the objective to address these challenges. The members of the Working Group are:

- Simon McIntosh-Smith (chair), University of Bristol, UK;
- Thomas Ludwig (vice-chair), University Hamburg/ German Climate Computing Centre Hamburg, Germany;
- Alex Ramirez, nVIDIA, US, former Barcelona Supercomputing Centre, Spain;
- Matthias Müller, RWTH Aachen University, Germany;
- Jean-Marc Pierson, Laboratoire IRIT, France;
- Laurent Lefevre, INRIA / University of Lyon, France;
- James Perry, EPCC, University of Edinburgh, UK.

In 2014 the following new experts were added:

- Paul Carpenter, Barcelona Supercomputing Centre, Spain;
- Daniel Hackenberg, TU Dresden, Germany;
- Manuel Dolz, University of Hamburg, Germany.

4.2 Second year activity

Starting from the WG5.3 report produced after the first year of EESI2, the WG5.3 technical experts have held a number of conference calls to further deepen the scope of this working group. Our interactions culminated in a face to face meeting organised to coincide with International SuperComputing (ISC) in Leipzig, Germany in June 2014, the second time the group had met at this conference. In addition, a number of other meetings were held, including a face to face at the University of Hamburg in May 2014. We also ran a birds of a feather session at ISC in 2014 on energy efficient HPC, asking the audience of 20-30 attendees what projects they were aware of tackling this issue, and include the information gather at this BOF in the report below. Finally, we organised a panel session at the Energy Aware HPC (Ena-HPC) conference in Hamburg in September 2014 where we gathered the most up to date information from this field. The majority of the new technical information in this report was gathered by the experts at the end of this process, after our activities at ISC and Ena-HPC. Our experts have also been participating in a number of important, energy efficient HPC projects, including LPGPU, Exa2Green and Mont Blanc, and some of the recommendations in this report come from the learnings gained from these projects.

4.2.1 Remaining key energy efficiency and power management challenges to achieve Exascale systems

Some of these challenges were identified in the report from the end of the first year. See Deliverable 5.1 [1]. New challenges from the second year are explicitly highlighted:

- Ability to profile applications for energy efficiency (critical).
- Fine resolution power mode manipulation mechanisms in all devices (critical).
- Improving scalability to improve energy efficiency (critical).
- Model power consumption (critical).

- Dynamic, energy aware load balancing across heterogeneous resources (important).
- Conduct overall benefit-cost-ratio analysis (important).
- Develop application benchmarks to measure energy efficiency (important).

(New challenge) No fundamentally new breakthroughs in trying to tackle energy efficient HPC**:** At the Energy Aware HPC (Ena-HPC) conference in September 2014, Thomas Ludwig noted at a panel session that, despite "diverse and successful research on energy efficiency" and "special conferences and workshops (on energy efficient HPC)", the momentum behind trying to tackle this challenge appears to be waning. Specifically he noted that "(In) Germany: no special funding for energy efficiency in HPC despite enormous costs", and that the Euro-Par conference has said that "we do not need a special Green IT workshop because now green is in everything".

4.2.2 Recent activities around energy efficient HPC (EEHPC)

We began the latest phase of the EESI2 project by analysing the energy efficiency related research being reported at the three main HPC conferences: IEEE/ACM SuperComputing (SC), International Supercomputing (ISC) and Energy Aware HPC (Ena-HPC).

At the 2013 SC conference, 4 of the 90 technical papers were related to energy efficiency (one of CPU-GPU load balancing, one of energy aware job schedulers, one on organising memory for energy efficiency, and one of in-situ data analysis to improve energy efficiency). This is a rate of about 5% of the total papers, not a very high ratio. Two of the Birds of a Feather (BoF) sessions were about EEHPC issues, out of a total of 74, a ratio of just under 3%. The rate was a little higher at ISC in 2013, with 3 talks (out of 33) and 2 BoFs (out of 20) related to EEHPC. Ena-HPC, being focused on this topic specifically, pulled together a program of 12 papers, all on the topic of EEHPC. These numbers are not very high, leading us to conclude that not enough is being done to encourage enough high quality research in this challenge area.

4.2.3 Current state of the art

The content of the report of the first year is still valid. There is important additional new work to report with respect to benchmarking.

Energy efficiency benchmarks: there are a few benchmarks available that address energy efficiency. The Linpack benchmark used to create the TOP500 and Green500 list has been extended to include power consumption. However, the metric is rather simple (MFlops/Watt) and the run rules for how to measure the power consumption lack precision. The SPECpower_ssj2008 benchmark was specifically created to measure energy efficiency, has very detailed and precise run rules, but is focused on Java workloads. SPEC OMP2012 and ACCEL are application benchmarks with scientific applications using OpenMP, OpenCL and OpenACC that have been extended with an energy efficiency metric and detailed run rules for energy measurement. These should be useful for performing direct measurements of energy efficiency on "like for like" workloads across different computer architectures, such as CPUs from workstations or embedded systems, GPUs, and many-core processors such as the Intel Xeon Phi.

4.3 Gap analysis

The gap analysis described in the previous deliverable is still up to date. There are important fields where we still find gaps that need to be filled with respect to research and development. As stated in the first reports this refers to the following fields:

- Hardware energy monitoring.
- Energy profiling of applications.
- Standard API for accessing energy information.
- Performance and operating states in future processors and systems.
- Modelling power and energy consumption in future architectures.
- Deploying and managing large scale numbers of energy sensors.
- Increased concurrency to offset decreased clock speeds.

- Addressing whole-system power consumption.

4.4 Recommendations

At the end of the second year, the WG 5.3 has the following set of focus areas that it would recommend should be a target for funding to stimulate high quality research and collaboration with industry:

- Modelling and prediction
- New hardware architectures
- Foster competition on hardware development
- Budgeting based on kWh
- Energy aware software (libraries, runtimes, …)
- Integrate all levels of hardware and software
- Examine power and * (resilience, …)
- APIs that are truly open and work with all Vendors

These recommendations integrate the previous suggestions provided at the end of the first year from WG5.3 which included:

- Create a standard interface for power monitoring and power management at all levels of the system architecture.
- Create a task force to look at the relevant software development tools from the embedded computing space.
- Training and education to prepare developers to face the power wall.

Many of these recommendations have been synthesized in the recommendation presented in Deliverable D 7.2 [2].

Energy efficiency is a crucial challenge that must be successfully addressed if the benefits of Exascale supercomputing are to be realised. Yet little research is being stimulated in this area, and the fundamental APIs required are not being developed in an open, standard format. These shortcomings must be addressed through a combination of targeted funding calls and industrial engagements. Without them, Exascale machines will remain an unreachable goal.

5. WG 5.4 Resilience

5.1 Introduction

Resilience addresses the increase of system failure rate due to the explosive growth in component count in supercomputers as well as the use of advanced technologies such as NTV (near threshold voltage). The main objective of WG 5.4 is to address how Exascale computers must dynamically compensate for failures:

- Understand the need for resilience BEFORE the system is built (some of the largest HPC systems on earth have not considered this issue…)
- Understand trends of resilience approaches, compare them qualitatively, quantitatively
- Understand how resilience, performance and energy impact each others
- Understand the need to involve the application developers.

The resilience challenge cannot be addressed in isolation looking at a single software or hardware component. Resilience needs to be addressed considering the whole system: all layers of the software stack, all hardware components constituting the Exascale system and all usages of this system.

Three different kinds of problems have been introduced during the first year of WG 5.4 activity

- Process crashes (fail stop errors)
- Transient errors (detected but not corrected)
- Data corruptions (silent soft errors); data corruption could ultimately lead to process crash.

The next section reports the activity done in the second year of activity of WG 5.4, from the group of experts, on the topics of resilience, The work, continued the activity issued during the first year of WG 5.4, aiming at providing: i) a gap analysis between existing reports and projection about the resilience challenge for exascale simulation; ii) a set of recommendations based on this gap analysis.

5.2 Second year activity

Resilience is becoming a very hot topic in HPC, as reported in [3] and addressed in different conferences and meetings. Just to mention a few:

- 2014 SIAM conference on parallel processing featured 17 talks covering many aspects of resilient algorithms.
- Dagshtull seminar on Resilience on September 2014
- Many papers on resilience presented at ACM HPDC2014 (Checkpointing intel MIC, RDMA message logging, etc.)
- two Tutorials on Resilience have been presented at SC2014

These and other meeting have been the occasion for meeting the WG 5.4 group of experts and addressing the second year activity.

The activity continued the analysis on RAS (Reliability, Availability Serviceability) system for Exascale, at different levels: node hardware level, node system software level, interconnection level, file system and storage level.

Then the activity focused on the critical issue of runtime investigating the few advances on how the runtime (and programming models) can enhance system resilience. As reported in the first year report, the research in this domain has just started and more efforts should be put on understanding how to leverage and control by the runtime hardware resilience features.

Another important item is represented by the high performance checkpointing and multi level checkpointing, crucial points for Exascale systems. More research is still needed to better understand how redundancy across multiple processes relates to data structures at application level in order to identify applications classes that can benefit from specific optimizations.

The requests for specific research activities in these fields, as indentified in the analysis done during the first year of activity, are still valid and important research activities still must be addressed.

Furthermore, the need for advanced fault tolerant protocols and resilient numerical algorithms is still critical and is urgent to address research in this topics when exascale system are approaching.

5.3 Gap analysis

The gap analysis done during the first year activity and described in Deliverable 5.1 [1] is still up to date. One of the big issues is that a fault model is still not available and the worst is the different views on failure rate and silent soft error could increase with aggressive power saving technologies.

Other serious issues are concerning silent data corruptions (SDCs): no quantification, no understanding of the propagation, and no clear injection model are still available.

5.4 Recommendations

Last year, the recommendation presented by WG 5.4 aimed at finding relevant effective and efficient solutions for Exascale resilience, addressing both fail stop errors and silent data corruptions, taking into account the multifaceted aspect of this problem. The recommendation was divided in 6 different tasks:

- SP1: Extend the applicability of Checkpoint/restart and migration
- SP2: Improve system efficiency and execution recovery in presence of fail stop errors through better fault tolerant protocols
- SP3: Investigate alternatives to checkpoint/restart: tasks based checkpoint/restart, migration and redundancy
- SP4: Fault aware software stack
- SP5: Develop failure prediction
- SP6: Resilient algorithms

The recommendation is still valid and specific programmes should be launched at Eurpoean level to give momentum to resilience in the field of HPC approaching Exascale.

In additions in this second year, a new recommendation "*Holistic approach to resilience for simulations and data analytics*" integrates the previous one. This new recommendation fits in the Pillar *Tools and Programming Models* and proposes the development of resilience API that will provide the required integration of resilience techniques and coordination of software resilience mechanisms and by improving critical resilience mechanisms:

- Understanding and modeling of fault propagation
- Push Checkpoint restart as far as possible
- Error detection
- Failure prediction
- Roll back and roll forward recovery
- Resilient Runtime, Resilient OS and Resilient Algorithms.

It is recommended not an integrated project on Resilience covering all layers from hardware to the applications, but a project on Integrated Resilience, adopting an holistic approach, covering from numerical algorithms to resilient supporting software (libraries, runtime, OS, etc.).

6. WG 5.5 Disruptive technologies

6.1 Introduction

The last decade has seen significant changes in processor architectures to improve computing performances and to overcome the physical limitations of increasing the clock frequency. This allowed processors to still scale according to Moore's law. However, a new challenge is becoming relevant today: the ever increasing power and energy requirements for operating the latest generation of HPC systems are key factors in limiting the peek performances of newer HPC systems and might make Exascale computing unsustainable for both technical and economic reasons.

A possible solution is to identify disruptive technologies in terms of new hardware architectures and energy aware system software which maximize the performance of HPC systems within a given power or energy budget.

WG 5.5 focus on the search of disruptive candidates technology/components that have good potential to create a discontinuity on the current architectural trends while reducing the demands on other components of the HPC environment, especially regarding system density and efficiency.

The activity analysis done in the first year of activity of the WG 5.5 was devoted to the analysis of potential disruption coming from the hardware component of HPC architectures: semiconductor technology, packaging, data transfer, memory, network, cooling ad I/O. The result was that three main area of innovation could determine a disruption with respect of current paradigms and approaches: hybrid systems and processors, new "high bandwidth" memories and I/O subsystem based on new NVRAM technology.

6.2 Second year activity

In the second year we have better investigated all the aspect of the HPC infrastructure but with the focus on the software components that may disrupt the HPC software stack.

To analyze these important aspects we involved three new experts that joined the group of WG 5.5 experts and actively contributed to the activity of the Working Group:

- Dr. Andrea Bartolini University of Bologna and ETH Zurich;
- Prof. Luca Benini, University of Bologna and ETH Zurich;
- Prof. Cristina Silvano (Polytechnic of Milano);

achieving so a total of 11 experts in WG 5.5. that worked remotely and by way of ad hoc teleconferences. Furthermore the participation to the main HPC events like Supercomputing (November 2013 in Denver and November 2014 in New Orleans) and the ISC series of conferences in Europe (Leipzig, June 2013 and 2014) was the occasion for have further updates and brainstorming on innovative and disruptive technologies for HPC.

We found that energy efficiency and awareness is the main driver in the evolution of the software components, at all levels: firmware, operating system, scheduler (low and high level), monitoring, and applications.

The analysis of our experts started from considering that the top 17 positions of the Green500 list are currently occupied by heterogeneous computing systems.

The average energy efficiency for these systems, when measured in MFlops/Watt, is 1,938 MFlops/Watt, whereas it is only 743 MFlops/Watt for the homogeneous systems. To conclude, heterogeneous systems currently dominate the top of the Green500 list and this dominance is expected to be a trend for the next coming years to reach the target of 20 MW Exascale supercomputers. However, to fulfill this target, energy-efficient heterogeneous supercomputers need to be coupled with a radically new software stack capable of exploiting the benefits offered by heterogeneity at all the different levels (supercomputer, job, node) to meet the scalability and energy efficiency required by the Exascale era.

In particular Energy-efficient heterogeneous supercomputing architectures need to be coupled with a radically new software stack capable of exploiting the benefits offered by the heterogeneity at all the

different levels (supercomputer, job, node) to meet the scalability and energy efficiency required by Exascale supercomputers.

An API for monitoring and measuring power consumption in HPC systems is being organized by the United States Department of Energy's Sandia National Laboratory and the National Renewable Energy Laboratory (effort led by Natalie Bates at LBNL, and presented at IEEE/ACM SuperComputing in November 2014). The vision presented by the group of WG 5.5 experts goes beyond the “power API” for monitoring power consumption: For the management of Exascale supercomputers, an holistic approach must be adopted involving the different layers of the HPC architecture and environment.

The group of experts in WG 5.5 investigated also some other potential disruptive technologies:

6.2.1 Advanced cooling technologies

While cold water cooling systems are now becoming mainstream in HPC, there are several more advanced alternatives starting to emerge which might lead to even more efficient cooling systems. The most promising appear to be “immersive cooling” and “hot liquid cooling”. In the former, the equipment is partially or completely submersed in a non-conductive coolant, such as mineral oil. This can conduct far more heat away from the equipment than other approaches, and can do so requiring less energy. The hot water cooling is a variant on today’s cold water cooling, where only a small delta in the inlet and outlet water temperatures is needed to cool the hardware. This small delta, typically 10 degrees or less, can usually be achieved with free outside cooling, i.e. it does away with the need to run energy-hungry compressors, and so leads to even cheaper and greener cooling systems than cold water cooling provides.

6.2.2 Optical computers

Optalysys is a UK company developing an optical supercomputer that can perform mathematical functions such as FFTs and matrix multiplications using light focused by liquid crystal patterns. This contrasts with the electronic approaches used by today’s mainstream computers. The optical computer approach has the promise of higher performance and much greater energy efficiency.

Future technologies such as Quantum computing offer potentially huge increases in processing power, but it is not clear yet exactly what functionality they will provide.

Optalysys is a spin out from the University of Cambridge and is at the prototype stage. For more information see: http://optalysys.com.

6.2.3 Quantum computing

The nature of quantum computers is totally different from the classical digital computers based on transistors, as they make direct use of quantum-mechanical phenomena such as superposition and entanglement to perform operations on data. The computation is based on qubits, object obeying to the rules of quantum-mechanics .

A classical computer has a memory made up of bits, where each bit represents either a one or a zero. A quantum computer maintains a sequence of qubits. A single qubit can represent a one, a zero, or any quantum superposition of those two qubit states; a pair of qubits can be in any quantum superposition of 4 states, and three qubits in any superposition of 8 states. In general, a quantum computer with n qubits can be in an arbitrary superposition of up to 2^n different states simultaneously (this compares to a normal computer that can only be in *one* of these 2^n states at any one time). A quantum computer operates by setting the qubits in a controlled initial state that represents the problem at hand and by manipulating those qubits with a fixed sequence of quantum logic gates.

From the hardware point of view, all the costs relative to maintaining operative a quantum computer is due to the cooling of the machine, whose temperature should be close enough to 0 K degrees.

From the algorithmic point of view, the challenge for the use of quantum computers, is how to map classical problems of interest of the HPC world to problems solvable with a quantum algorithm. This issue is the one that still makes difficult to consider quantum computing an approachable solution and probably will be the subject of the work in the next future.

D-Wave Systems, Inc. is a quantum computing company, based in Canada. On May 11, 2011, D-Wave Systems announced D-Wawe One, described as "the world's first commercially available quantum computer," operating on a 128-qbits chipset using quantum annealing (a general method for finding the global minimum of a function by a process using quantum fluctuations) to solve optimization problems. In May 2013 it was released the 512-qubit D-Wave Two system.
The last product of D-Wave, the D-Wave Two has been recently adopted by NASA, Lockheed-Martin, Google and USC to tackle several problems from machine learning to minima-finding problems.
In June 2015 D-Wave announced a new quantum processor with over 1000 qubits designed to deliver a 1,152 qubit region out of a complete 2,048 qubit fabric.
The new processor will run in an environment 40% colder than the previous generation, which operated at about 20 mK.

In May 2015 a private meeting has been organized in CINECA with Murray Thom and Andy Mason, representatives of the D-Wave company. During this meeting the approach of quantum computing was presented and the possible perspectives in the context of the HPC ecosystem have been investigated.
The promise of quantum computing is very exciting for some fields of applications.Once mature algorithms will be ready for the quantum computers, this technology will certainly be a solution of absolute interest for the high-performance computing industry.

6.2.4 Other potential disruptions

It is worth noting that there are other research topics which may disrupt the energy efficiency of HPC in the future. Many of these are materials topics. For example, **graphene** is being explored as a potential alternative or complement for silicon as the basis for microchips as we come to the end of Moore's Law scaling for traditional materials. Similarly, higher performance, more energy efficient photonic interconnects are now working in the laboratory, and these are expected to appear in high-end HPC systems in the next few years. Other research topics which may lead to other potential disruptions are represented by the EC Flagship **Human Brain Project** activity. Research on neuroscience at the human brain level can address new technologies in the field on brain inspired computing as well as neuromorphic computing.

6.3 Recommendations

Last year, the activity in WG 5.5 aimed at finding potential disruptions focusing on the hardware components and technologies for Exascale. The recommendations aimed to push research and investment in these new technologies, mainly, from one hand, in the field of new memory technologies and in general data movement integration and, from the other hand, efficiency of the architectures. Disruptive technologies appearing in these two fields may allow dramatic redesign in system architecture (for example with non volatile memory or optical interconnect to substitute PCB) and in new application paradigms.

These recommendations are still valid in the second year of activity but must be complemented by specific recommendations aimed to integrate the hardware architecture with the software aspects investigated during the second year of activity. A disruptive holistic approach spanning all the decision layers composing the supercomputer software stack and exploiting effectively the full system capabilities (including heterogeneity and energy management) is recommended. A breakthrough approach is needed to express application self-adaptability at design-time and at runtime to manage and auto-tune applications for green and heterogeneous HPC systems up to the Exascale level.

Key elements of this approach are:

- scale out monitoring based on big data techniques and low level protocols
- control over the hardware
- domain specific languages
- separation of concern

The applications then need to be re-factored to exploit the separation of concerns with respect of the energy consuming part, and domain specific languages need to be developed to be able to cope with the lower layers allocating the workload to the available resources in order to optimize the efficiency.

All these considerations have led to the recommendation *"Holistic approach for extreme heterogeneity management of Exascale supercomputers"* as part of the *Tools & Programming Models Pillar;* See [2]. The recommendation aims at foster the research and development of:

- Hardware/Software APIs to manage the complexity and the programmability gap inherent of extreme heterogeneous Exascale level supercomputers;
- Design strategies for scalable and efficient heterogeneous-aware exascale applications;
- Scalable and efficient community scientific applications for exascale;
- System software to support efficient usage of exascale heterogeneous supercomputers in production.

7. Conclusions

This document is the second intermediate report of the EESI2 WP5 *Cross Cutting Issues*. The WP is organized in five WGs: Data Management and exploration, Uncertainties (UQ/Verification & Validation), Power & Performance, Resilience, Disruptive Technologies.

The deliverable presents an update of the first report (Deliverable 5.1) focusing on the activity undertaken during the second year of the project, reviewing the process of the project recommendations and distilling new ones, deduced by the themes analysed in the second year of this WP.

Between the topics that the experts have highlighted as challenges, we find:

- Data management represents one of the major challenges of Exascale applications addressing scientific discovery nowadays. Actions must be issued to address end-to-end techniques for efficient disruptive I/O and data analysis, involving the full life-cycle of data.
- For Exascale applications, verification, validation and uncertainty quantification of computer models' results becomes fundamental both for industry and academia. Identify methodologies and enhance tools for the analysis of these uncertainty sources, is fundamental for the exploitation of Exascale applications.
- Power monitoring and power management at all levels of the system architecture, addressing energy efficient performance of applications, is a crucial issue to address in the Exascale era. The definition of standards is urgent as well as the formation of professional HPC developers experts in green programming methodologies.
- Robust fault tolerance protocols as well as performing checkpoint/restart methods, to increase the efficiency of Exascale systems, are becoming urgent to manage the fore coming systems with millions of cores.
- The roadmap toward Exascale and beyond will be guided and modeled by disruption in semiconductor technologies, I/O and memory technologies, networking and data transfer technologies, energy aware and advanced cooling technologies and facility management. The entire Software stack (programming models, run-time, OS and system support software) needs to be fully innovated to support programmability and efficient performance/energy usage of the different resources and the computational models.

8. Bibliography

[1] EESI2 Deliverable 5.1 "*First Intermediate report on cross cutting issues"*, September 2013.

[2] EESI2 Deliverable 7.2 "2014 Update Vision & Recommendations", July 2014.

[3] Franck Cappello, Al Geist, William Gropp, Sanjay Kale, Bill Kramer, Marc Snir, "*Toward Exascale Resilience: 2014 update* ", Supercomputing Frontiers and Innovations Journal, 2014.