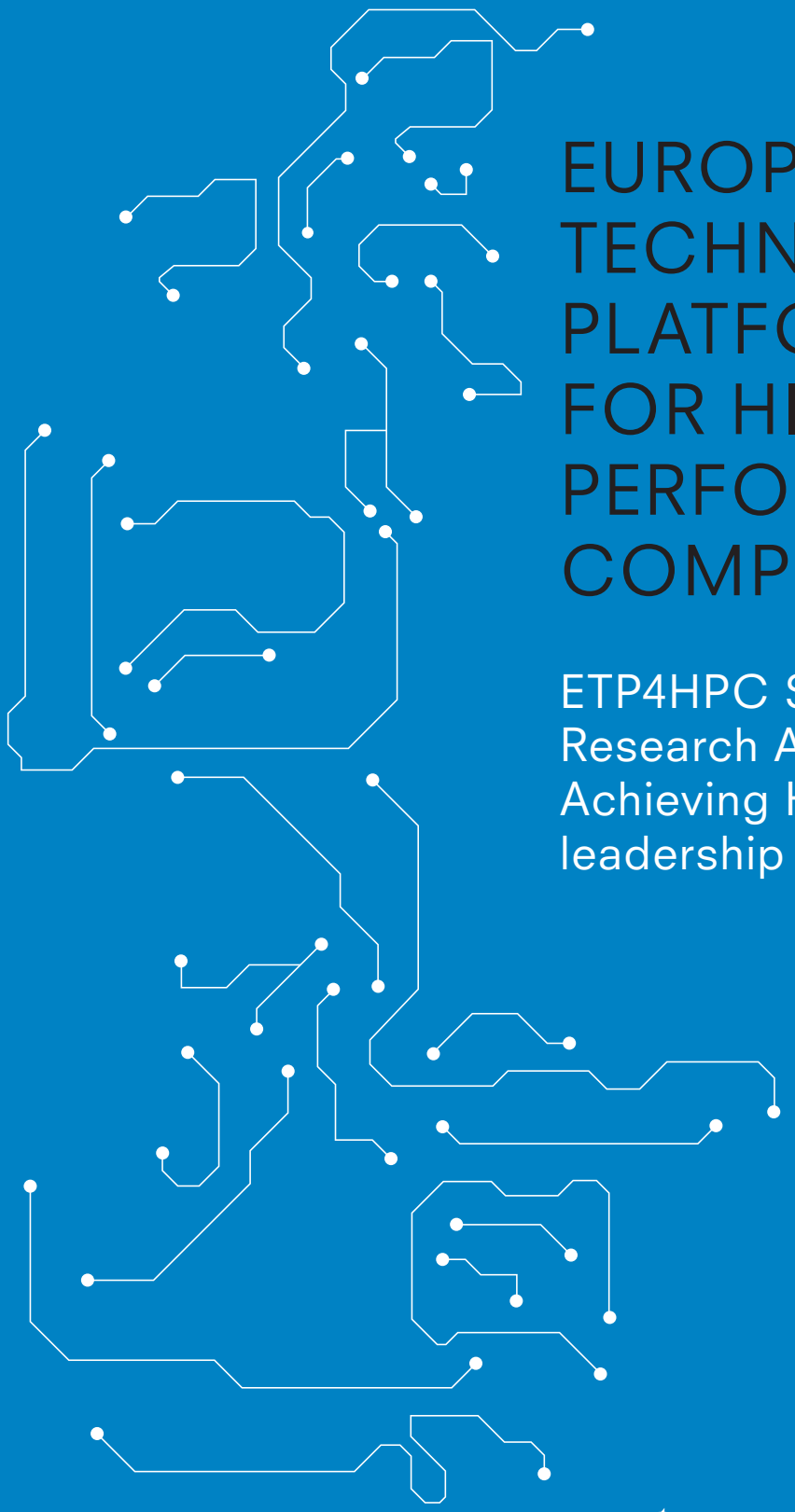


EUROPEAN TECHNOLOGY PLATFORM FOR HIGH PERFORMANCE COMPUTING

ETP4HPC Strategic Research Agenda Achieving HPC leadership in Europe

ETP 4 HPC

ETP 4 HPC

FOREWORD

High Performance Computing (HPC) is a pervasive technology that strongly contributes to the excellence of science and the competitiveness of industry. Questions as diverse as how to develop a safer and more efficient generation of aircraft, how to improve the green energy systems such as solar panels or wind turbines, what is the impact of some phenomena on the climate evolution, how to deliver customized treatments to patients… cannot be answered without using HPC.

As today "to out compute is to out compete", Europe could improve its position in almost all scientific fields and industrial sectors if we were able to apply HPC on a large scale to produce new knowledge, design innovative products and services and reduce cost and time to market.

To achieve this HPC leadership, a strong ecosystem of HPC technologies is mandatory. This expertise and competence will be key in order to wisely invest in the field, develop best practices, anticipate technical ruptures and also to create a solid European industry in the growing and strategic market of HPC solutions.

This Strategic Research Agenda (SRA) proposes a roadmap to develop this HPC technology ecosystem and European leadership that will benefit all HPC users - both academic and industrial R&D centres. It has been prepared by a broad team of experts taking into account various inputs and the European position. It sets ambitious objectives for the different technologies required to produce leading-edge HPC supercomputers up to the Exascale level.

If we are successful in achieving the targets addressed in this SRA, Europe will take the lead in HPC that drives many research, industrial and societal challenges. Yes, we can make it happen with a determined and coordinated action of all stakeholders.

On behalf of ETP4HPC Founding Members representatives:

David Lecomber, **Allinea**
Ian Phillips, **ARM**
Francesc Subirada, **BSC**
François Bodin, **CAPS**
Jean Gonnord, **CEA**
Sanzio Bassini, **CINECA**
Giampietro Tecchiolli, **Eurotech**
Guy Lonsdale, **Fraunhofer**
Andreas Pflieger, **IBM**
Bernadette Andrietti, **Intel**
Thomas Lippert, **FZJ**
Arndt Bode, **LRZ**
Hugo Falter, **ParTec**
Patrick Blouet, **ST**
Malcolm Muggeridge, **Xyratex**

Jean-François Lavignon, **Bull**
Chair of ETP4HPC

Contact ETP4HPC:
office@etp4hpc.eu

CONTENTS

Foreword 5
Executive summary 8
1. Introduction 10
2. The Added Value of HPC for Europe 12
2.1 The Added Value of the European HPC Supply Chain 14
2.2 The Availability of HPC End-User Solutions 14
2.3 Addressing Grand Challenges 15
2.4 Expected Return on Investment 15
3. Building the Strategic Research Agenda 16
3.1 The Approach 17
3.2 Other Initiatives 18
3.3 Input from PRACE 18
3.4 Input from Industrial HPC End-Users 19
3.5 Input from ISVs 20
3.6 ETP4HPC SWOT Analysis 22
4. A Multidimensional HPC Vision 24
5. Technical Research Priorities 28
5.1 HPC System Architecture and Components 31
5.2 System Software and Management 35
5.3 Programming Environment 39
5.4 Energy and Resiliency 43
5.5 Balance Compute, I/O and Storage Performance 49
5.6 Big Data and HPC Usage Models 54
6. Completing the Value Chain 60
6.1 HPC Services 61
6.2 Independent Software Vendors 64
6.3 SMEs as HPC Technology Providers 64
6.4 Education and Training 66
7. Links with other initiatives 70
8. Making it happen 72
9. Glossary of Abbreviations 76
10. References 78
11. Other references and links 79
12. Acknowledgments 80
13. Appendix – Research Topic Timelines 82

EXECUTIVE SUMMARY

High-Performance Computing (HPC) plays a pivotal role in stimulating Europe's economic growth. HPC is a pervasive tool, allowing industry and academia to develop world-class products, services and inventions in order to maintain and reinforce Europe's position in the competitive worldwide arena. HPC is also recognized as crucial in addressing grand societal challenges. "Today, to out-compute is to out-compete" best describes the role of HPC.

This document is the Strategic Research Agenda (SRA) of ETP4HPC, the European Technology Platform (ETP) in the area of HPC. ETP4HPC is an organisation whose members are key HPC technology providers and research centres involved in HPC research in Europe with the objective of strengthening the European HPC ecosystem and thus to contribute to the competitiveness of the European economy. The purpose of this document is to define a roadmap for the implementation of a European research programme for HPC technologies.

The ETP4HPC believes a research programme is needed in order to reinforce and grow the European HPC Ecosystem, crosscutting the value chain provided through European HPC technology. Europe can benefit from this programme through a return on investment in innovative HPC technologies, the availability of cutting edge HPC end-user solutions and the provision of tools to tackle Europe's grand societal and economic challenges.

A comprehensive research programme with an investment of 150 million Euros per year between 2014 and 2020 can produce tangible results, including the creation of jobs and ownership of technologies developed in Europe by a competitive HPC technology value chain.

The Strategic Research Agenda has been developed in consultation with other players of the European HPC environment, such as PRACE, HPC industrial end-users and Independent Software Vendors (ISVs). The recommendations of the document are based on a multidimensional HPC model suggested by ETP4HPC. This model includes the following:

- HPC stack elements
- Extreme scale requirements
- New HPC deployments
- HPC usage expansion

Applying this model, ETP4HPC proposes research in the following main areas of HPC:

- HPC system architecture
- System software and management
- Programming environment
- Energy and resiliency
- Balance compute, I/O and storage performance
- Big Data and HPC usage models

In addition, ETP4HPC also suggests actions in other, complementary areas. These recommendations facilitate the creation of a dynamic HPC environment in Europe:

- HPC services
- Independent software vendors
- SMEs as HPC technology providers
- Education and training

The document also provides links to other initiatives necessary to implement the suggested research programme. Furthermore, an outline of an action plan is presented.

A quick guide through the structure of this document:

Chapter 1 and 2 introduce the relevance of HPC in general and a research agenda in particular. Chapter 3 describes the foundation on which the agenda is built, the different sources of input, and the conclusion of analysis performed by the ETP4HPC. Chapter 4 provides a high-level view of the research dimensions, whereas Chapter 5 presents the core of the document and details on the research areas. Chapter 6 outlines other important areas influencing the success of an HPC research roadmap. Chapter 7 describes links with other initiatives, and Chapter 8 outlines the next steps and concludes the document.

1. INTRODUCTION

This Strategic Research Agenda has been prepared by **ETP4HPC**[1] (www.etp4hpc.eu), the European Technology Platform in the area of High-Performance Computing. The objective of this document is to outline a roadmap for the implementation of a research programme aiming at the development of European HPC technologies.

High-Performance Computing (HPC) refers to any form of computing where the density of processing or the size of the problems addressed require more than a standard or commodity computing system in order to achieve the expected result under the given constraints, and the application of advanced techniques such as the use of multiple processors (tens, hundreds, thousands or even more) connected together by some kind of network to achieve a performance well above that of a single processor.

Two traditional HPC categories are distinguished:

Capability computing refers to the use of a large and high-performing computing infrastructure to solve a single, highly complex problem in the shortest possible time. Systems used for this purpose are called supercomputers and consist of many tightly coupled compute nodes with distributed memory, all controlled by a single unit.

Capacity computing refers to optimising the efficiency of using a compute system to solve as many mid-sized or smaller problems as possible at the same time at the lowest possible cost.

However, HPC is an open field with evolutions that are able to deliver the computing power needed in areas such as Cloud Computing and Big Data. HPC continuously targets the most complex and most demanding computing tasks within the reach of emerging technologies.

As described in our Vision Document [ETP4HPC-1], a competitive European HPC ecosystem can be attained by achieving the following objectives:

· Build a globally competitive, world-class European HPC technology value chain by achieving a critical mass of convergent resources

· Leverage the transformative power of HPC to boost European competitiveness in science and business

· Expand the HPC user base, especially among SMEs, through the facilitation of access to HPC resources and technologies

· Open possibilities for SMEs to participate in the provision of competitive HPC technology solutions.

· Facilitate the provision of innovative solutions to tackle grand societal challenges in Europe

· Foster international cooperation in research and industry.

Our role in this process is the provision of expert advice in relation to research priorities and their implementation. ETP4HPC's ambition is to act as a catalyst, bridging the gap between science and industry in the implementation of research results.

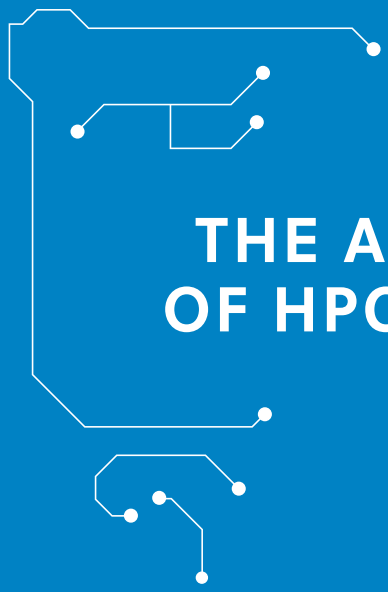
ETP4HPC will continue to update its SRA on an on-going basis, taking into consideration the evolution of technology and the ecosystem's requirements. Besides the definition of research priorities, this document also suggests an implementation path.

This SRA will be presented to the European Commission with the objective of its becoming the basis for future R&D programmes.

The implementation of the recommendations of the SRA will have the following impact:

- Strengthen the European HPC technology provision eco-system and increase its global market share
- Allow Europe to achieve global leadership in HPC-related technological areas, with the possibility of transferring such technologies to other industries
- Address some of the globally recognised grand challenges, such as energy efficiency and the handling of large data volumes
- Design HPC solutions required by European science and industry

[1] The creation of ETP4HPC was announced on 18 October 2011 in Barcelona. The organisation's founding members are industrial and academic organisations involved in research in HPC technologies in Europe. The founding members of ETP4HPC are (industry) Allinea, ARM, Bull, Caps Entreprise, Eurotech, IBM, Intel, ParTec, STMicroelectronics, Xyratex; (research) Barcelona Supercomputing Center (BSC), Commissariat a l'energie atomique et aux energies alternatives (CEA), Cineca, Fraunhofer-Gesellschaft (FHG), Forschungszentrum Julich (FZJ), and Leibniz-Rechenzentrum (LRZ).



2.
THE ADDED VALUE OF HPC FOR EUROPE

Global HPC development is driven by the need to address societal and economic challenges that require extreme computational resources and by the need of industry to innovate using simulation and digital prototyping.

HPC has been used for decades in meteorological predictions, in applied research in the car and aircraft industries, in high-energy physics, in energy, in material science, and in drug design.

Since the mid-1990s, HPC has been integrated into industrial design and engineering processes, playing a decisive role in improving the quality and efficiency of complex products, and significantly reducing the time to market. Its expanded capacities have enabled simulations at a higher level of precision and complexity, significantly impacting new areas, such as financial applications, biology, biomedical engineering, and medical imaging, by, for example, bridging across different scales, integrating more data from observation and experiment, and getting closer to real-time simulations.

Through the improved competitiveness of European science and research as well as European industry, HPC leads to significant added value for society.

The establishment of Partnership for Advanced Computing in Europe (PRACE) has enabled access to world-class HPC infrastructure for research. PRACE's Scientific Case [PRACE] includes further examples of research and industrial areas requiring advanced HPC solutions in the future.

The focus of ETP4HPC is on **innovation** in HPC technologies in Science and Industry. New HPC technologies will provide tools for linking Research and Innovation, and thus enable European industry to reap the benefits of HPC on par with Science. Europe's competitive advantage is its ability to **innovate**, owing to, among other things, its vibrant SME ecosystem. As SMEs are a key driver of technology and research-led innovation, therefore for the European HPC industry and supply chain to succeed, it must foster a thriving SME base.

The development of European HPC is recognised as a key element of European economic competitiveness [EC]. The value added by the existence of a HPC ecosystem in Europe is discussed in the next three sections:

2.1 THE ADDED VALUE OF THE EUROPEAN HPC SUPPLY CHAIN

The European HPC Supply Chain is formed by companies providing HPC technologies. Its value can be measured by its contribution to the European economy, i.e., the number of jobs created, the revenue generated, and the intellectual property created in Europe. Also, there is the intangible and indeterminately large value achieved by providing a complete HPC ecosystem—from suppliers to users. There is general consensus that Europe remains one of the leading consumers of HPC systems, while European system vendors hold a significantly lower share of the market (note that a significant portion of the hardware components are manufactured in Europe).

A strong EU HPC Supply Chain can lead to a significantly increased contribution to the European economy.

According to IDC [IDC-1] the broader HPC market (compute, storage, middleware, application and services) was worth $20.3 billion in 2011 with a compound annual growth rate (CAGR) of 7.6% between 2011 and 2016.

On the other hand, the server market was worth $10.3 billion in 2011, with a predicted annual growth of 7.3% until 2016, reaching $14.6 billion. 31% of this market in 2011 resided in Europe.

The share of European HPC system vendors in the global HPC system market is less than 5%.

2.2 THE AVAILABILITY OF HPC END-USER SOLUTIONS

Europe is home to a vibrant HPC end-user environment. Companies using HPC find HPC indispensable for their ability to compete on a global level [IDC-3]. The availability of competitive, leading-edge HPC end-users solutions (i.e., systems and applications) will generate economic value through the return on investment achieved through innovation, shorter time-to-market, and increased operational efficiency.

HPC AND EUROPEAN INDUSTRIAL COMPETITIVENESS

The competitiveness of a number of European industries depends on the availability of, and easy access to HPC resources:

· **Medicine and life sciences.**
Genomic therapy and personalized medicine are now recognised as powerful tools. The explosion of biomedical information (e.g. EBI's data volume increased from 6,000 TBytes in 2009 to 11,000 TBytes in 2010, with more than 4.6 million requests per day [EBI]) requires a huge increase in the processing capability to analyse these data. A typical drug discovery pipeline involves scanning more than 100,000 molecules per day to check their potential effect. Identification of potential drug candidates for disease targets will be fuelled by the next generation of supercomputers.

· **Materials science, Chemistry and Nanoscience.**
The incessant demand for new materials in domains, such as the consumer electronics industry (relying heavily on nano-electronics), medicine (enabling new diagnosis methods), chemistry (facilitating the development and use of products with lower environmental impact), is strongly dependant on HPC evolution. The objective of computational materials science, chemistry and nanoscience is to create new materials or chemical agents whose electronic reaction times ranging from the femtosecond range to geological periods that enter materials formation.

· **Aeronautics.**
HPC is already a key technology for aircraft manufacturers. It allows the design and modelling of aircraft components and subsystems with superior performance characteristics, energy efficiency, and greatly reduced environmental impact. The challenge for future aircraft desing is to test fly a virtual aircraft with all of its multi-disciplinary interactions in a computer environment, compiling all the data required for development and certification—with guaranteed accuracy in a reduced time frame. Such an objective clearly is beyond Exascale and possibly even zeta-scale systems.

· **Automotive.**
The automotive industry has profited strongly from their take-up of HPC: product development cycles and costs were significantly reduced by adopting virtual design, simulation and testing, and the efficiency, safety and eco-friendliness of

products could be significantly increased. This industry already foresees the need for EFlops-class systems to address the following challenges: long-range vehicle lifetimes without repairs, full-body crash tests, which also include potential soft-tissue damages, longer-lasting batteries (in particular for electrical and hybrid cars).

· **Energy.**
Here HPC demand is strongly driven by the need for improved safety and efficiency of facilities and also for optimising the overall energy infrastructure to reduce waste and continuously match offer with demand. HPC is also required in the development of new energies, such as wind power, solar energy, or nuclear fusion.

· **Oil and Gas.**
HPC demand in this industry is mainly driven by exploration for new resources and is typically used in the process of identifying underground oil and gas resources using seismic methods. Other areas, such as the design of facilities for the cultivation of hydrocarbon, the drilling of wells and construction of plant facilities, operations during the lifetime of an oil field, and eventually the decommissioning of facilities at the end of production, all require high-end computation to improve safety and efficiency.

2.3 ADDRESSING GRAND CHALLENGES

The growth of the European economy as well as the prosperity and security of the European citizens depend to a large extent on the resolution of the Grand Societal and Economic Challenges:

· Health, demographic shift and well-being
· Safety of the population in case of natural or industrial disasters
· Food security and the bio-based economy
· Secure, clean and efficient energy
· Smart, green and integrated transport
· Supply of raw materials, resource efficiency, and climate action
· Inclusive, innovative and secure societies

While HPC will enable end-user applications to tackle these problems, it will also be the source of new technologies that may, for instance, enable new levels of energy efficiency or smart, embedded systems.

In addition to the direct contribution of its supply chain, HPC is also a prime catalyst of growth on a macroeconomic level through its technological interaction with science and other industrial and technological areas. Because of the high levels of return on investment achieved, there is a correlation between investment in HPC and leadership in industry and science. HPC is also able to stimulate innovation in other ICT sectors by the provision of novel computing technologies that subsequently are likely to be re-used in other applications (as has been the case with parallel programming). In turn, the growing requirements of HPC system may be satisfied by solutions developed in other areas such as the consumer product domains.

2.4 EXPECTED RETURN ON INVESTMENT

Launching the research programme proposed below by the ETP4HPC will bring Europe the following benefits:

· Increase the economic value contributed by the European HPC eco-system, namely:
  - Additional market revenue
  - Job creation through the provision of new technologies, new use cases, and new products
  - Market leadership through intellectual property created in Europe

· Transfer technology to other industrial areas.

· Reduce the dependency on HPC technologies provided by other regions

In their report [IDC-2], IDC states that industries that leverage HPC could add up to 2-3% to Europe's GDP in 2020 (in 2011 Europe's GDP was 12,600 billion Euros according to the International Monetary Fund) by improving their products and services.

3.
BUILDING THE STRATEGIC RESEARCH AGENDA

3.1
THE APPROACH

The writing of this Strategic Research Agenda has involved representatives of the entire European HPC technology community. Most of the technical content has been created by technical experts representing the members of ETP4HPC. An internal ETP4HPC team coordinated the process of gathering input and organised a number of internal and external workshops.

ETP4HPC has also called upon external experts in the process of creating this agenda. The conclusions of this SRA are based on an analysis of input from the following members of the European HPC ecosystem:

· Analyses completed by **other initiatives** (including HiPEAC, EESI and PlanetHPC)
· Input obtained from **PRACE** (using the PRACE Scientific Case and the material obtained during a meeting with a representative of the PRACE Council and a representative of PRACE Scientific Steering Committee in Munich on 13 December 2012)
· Input obtained during a workshop (Barcelona, 4 October 2012) and interviews involving **industrial HPC end-users**
· Input from **ISVs** obtained through a survey and interviews (carried out in Nov./Dec. 2012)
· The conclusions of a **SWOT** (Strengths, Weaknesses, Opportunities and Threats) analysis completed collectively by the members of ETP4HPC (a separate workshop held on 25 October 2012)

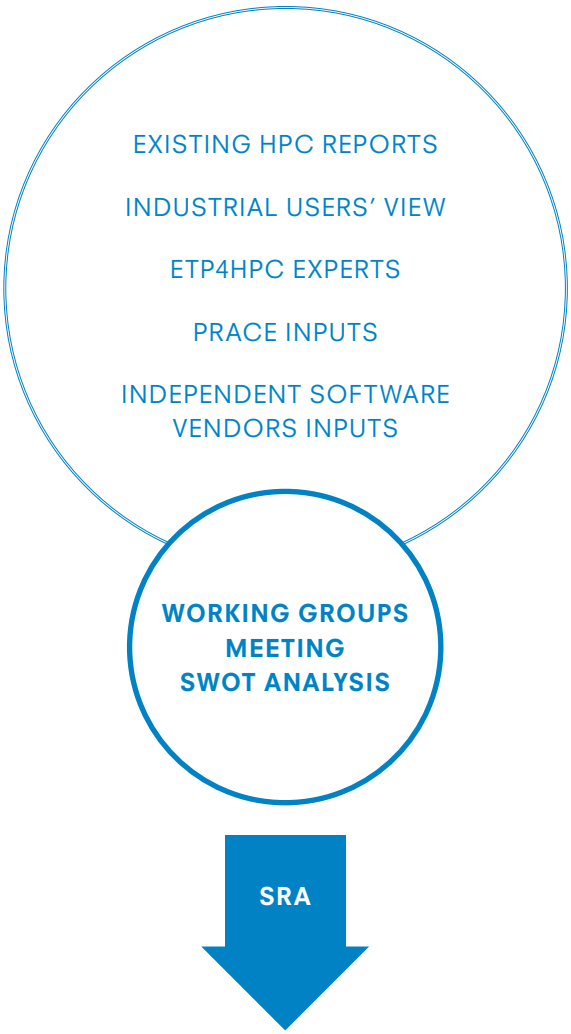


Figure 1
The process of collecting input for the Strategic Research Agenda

3.2 OTHER INITIATIVES

A number of other initiatives have presented an analysis of how HPC can add value to European competitiveness. This section outlines the main conclusions and recommendations of that work.

The High-Performance and Embedded Architecture and Compilation [HiPeac] initiative recommends to address the issues of efficiency (maximising power efficiency and performance through heterogeneous computer system design and catering for data locality), **system complexity** (tools and systems for building parallel and heterogeneous software; maintaining cross-component optimisation in component-based design; processor cores deigned for energy-efficiency, reliability and predictability), **dependability and applications** (bridging the gap between the growth of data and processing power; safety, predictability and ubiquitous availability of systems).

The European Exascale Software Initiative [EESI-1 and 2] states that Europe should exploit its world-leading position in HPC applications (83% of the HPC application software used in Europe is indigenous, 66% based on IP generated in Europe [IDC-4]) and software design (e.g. software libraries, programming models, performance tools, runtime design, system design, simulation frameworks, code coupling and meshing tools). EESI is convinced that all the expertise required to build an Exascale system is present in Europe, and it recommends funding a number of competitive projects to develop a European Exascale hardware platform. It also supports the establishment of a single European Exascale Software Centre.

Some other conclusions/areas to be addressed identified by EESI include the following:

- The need of European ISVs (and their commercial applications) to adapt to technological change
- Lack of coordination in the development of HPC software in Europe (while the standards are controlled and defined by non-European entities) and too limited participation in the definition of new global programming standards (e.g., OpenMP, C++, and Fortran)
- Lack of critical mass in some critical software domains, such as operating systems, compilers, message-passing libraries and file systems.

In a similar manner, the main technological challenges identified by **PlanetHPC** [PlanetHPC] are data locality (i.e., the proximity of data to the processing location), new programming models and tools (for massively parallel and heterogonous systems), portability of applications between different HPC architectures, and technologies supporting new and emerging applications that require robust HPC with real-time capability, data-intensive HPC, low-energy computing from both an architectural and application perspective.

3.3 INPUT FROM PRACE

PRACE (www.prace-ri.eu) is the pan-European infrastructure of high-end HPC systems with the objective to support the best computational science in the world. PRACE recently issued the report "The scientific case for high-performance computing in Europe 2012-2020" [PRACE], which outlines the research community's requirements regarding HPC technologies.

PRACE clearly identifies the research areas that will require **extreme-scale computing** in order to deliver relevant results. Most of the disciplines aim for more than 100 Pflop/s sustained simulations by 2020 (even though the adequate software might not be available today). There is a strong demand for the development of Exascale HPC technologies provided suitable algorithms are available.

According to PRACE, there is an enormous **skills gap** in academia and industry, hindering the development of European HPC. Europe does not have enough professionals with combined expertise in computing, numerical mathematics, and domain science. There is also a shortage of software engineers, caused by the poor career structure of that profession. An effort is needed to train more people, give software developers better career prospects and enable them to adapt to the rapid technological changes.

Another important concern of the PRACE community is the development of **new codes** that will be adapted to the extreme parallelism of the forthcoming HPC systems. Most algorithms will face re-writing challenges to work well on the fastest machines. Tools will be required for the management of complex workflows, performance evaluation and optimisation. Scientists will welcome any technology that could facilitate that process.

Other expectations of the research community regarding future HPC technologies include the following:

· **Simplicity and usability** – Researchers require technological solutions that are easy to work with. The technological complexity of a system should not affect its user interface and application performance.
· **Big data** – Handling large data volumes generated by research will be a major challenge and opportunity for future HPC systems.
· **Integrated environment for compute and data** – Another major challenge is the end-to-end management of, and fast access to, large and diverse data sets through the infrastructure hierarchy because most application areas foresee the need to run long jobs (for months/years) at sustained performances of around 100 Pflop/s to generate core data sets and very many shorter jobs (for hours or days) at lower performance for pre- and post-processing, model searches, and uncertainty qualification.
· **Co-design** – Science is in a position to identify the technology paths suitable for top-class systems by taking advantage of Europe's long-term experience in co-design (e.g., in the area of embedded microprocessors).
· **Stability of standards** – to develop long-term projects, researchers require technologies based on international standards.
· **Security and privacy** will be essential to realise health, social and economic benefits.
· **On-demand access** (instead of planned and batch access) will grow as some simulations will be relevant to urgent decision making.

3.4 INPUT FROM INDUSTRIAL HPC END-USERS

Fifteen companies and organizations spanning a wide range of sectors with focus on industrial applications and ranging from SMEs to very large companies participated in the ETP4HPC October workshop. The business and activity segments covered were energy, oil and gas, automotive, aeronautics, steel, consumer goods, weather forecast, biology, and services in the field of IT and computing.

The ETP4HPC presented its approach and proposed research topics at that time, whereas end-user representatives presented their expectations and visions related to HPC technologies and their usage, prepared in response to questionnaire with questions such as:

· How can HPC improve the competitiveness of their business – What are the missing features, limiting factors, and the most important HPC technical issues to be addressed?
· How can HPC improve the competitiveness of their industrial sector in Europe; potential impact of a strengthened European HPC technology delivery ecosystem

The main findings were as follows:

· Although a few flagship industrial applications are clearly requiring Exascale class systems for capability usage by the end of the decade, a lot of requirements are focused on capacity, with production application profiles ranging from task-farming and ensemble calculations to intermediate-size parallelism, and with a growing awareness and concern that big data will be a ubiquitous issue.
· Industrials want to focus on their core business, and ease of use is a major requirement. HPC technologies by themselves are mostly not a major focus of interest, but users demand solutions at the programming or system and resource management levels
    - for tuning and optimising applications (a critical bottleneck at all scales)
    - for modular and easy integration and configuration of HPC features into business-specific processes
· At extreme scale, the need to revisit algorithms and linear algebra is emphasized in order to accommodate more complex hardware by alleviating data movements. There already are very well identified industrial grand challenges that could stimulate, and benefit from, advanced research on these topics – especially in aeronautics, energy, oil and gas.
· Resiliency, error management, precision, and indeterminacy are highlighted as important matters for industry that so far are not sufficiently taken into consideration, but already are a concern today. This is also related to cost: such as the cost of not having enough reliability and thus wasting energy and compute cycles or the cost of reliability – these costs can escalate by orders of magnitude in terms of equipment cost.
· Affordability is a general concern, especially the cost of ISV licensing, which comes in addition to equipment cost, and can be a barrier to SME entry in HPC usage, as industry mainly uses commercial software.
· Altogether, there is a large demand for services, expertise and consultancy to support a more widespread industry usage, as well as for education. Industrial companies of

all sizes are not necessarily willing to hire HPC experts in large numbers in addition to their core business experts, but some multi-disciplinary characteristics are perceived as beneficial, as well as the presence of skilled experts in service and consultancy companies.

3.5 INPUT FROM ISV's

3.5.1 *Company profiles, application domains and uses*

Ten Independent Software Vendors took part in the survey. The company staff sizes ranged from 12 to approximately 1000; most ISVs deliver expertise and consultancy in their respective domains, as well as develop and distribute simulation software. The software portfolios represented by these companies encompass turnkey applications and fully integrated workflow environments as well as components and libraries meant for integration with customized solutions. The majority of software is proprietary, but some open-source packages, such as Scilab or OpenFOAM, are based on a service rather than a license business model.

The application domains mainly cover virtual prototyping, engineering design, and process optimisation, fabrication and performance analysis for a diversity of manufacturing and service sectors including, but not limited to, transport (automotive, aeronautics, train), energy, communications, and electronics. Pharmaceutical and bioinformatics applications as well as materials science and chemistry were also represented.

Models and algorithms found in the diversity of software packages and applications mentioned by the ISVs span a wide spectrum:

- Generic solvers for Finite Element Analysis (FEA), Computational Fluid Dynamics (CFD), electromagnetics, structural mechanics, materials simulation (quantum chemistry, molecular dynamics, etc.)
- Process-oriented modules, such as for metal forming, welding
- Pre- and post-processing of simulation programs and data, i.e., mesh generation, visualization and virtual-reality rendering and interaction. Moreover, many ISVs specifically indicated a growing need for and importance of visualisation in particular

The simulation software is usually ported to and available on multiple platforms. The average reported simulation use is often limited in scale to 8 to 16 cores, sometimes up to 64 cores. The main target for the majority of customers still is the entry-level commoditized desktop or desk-side or small-cluster configurations. Increasingly though, there are exceptions as some codes are already enabled and used at capability level on hundreds or thousands of cores (e.g., in the field of CFD, FEA, materials simulation).

3.5.2 *Usage trends and demands, potential impact*

The usage trends and needs are a mix of more capability and more capacity, often ensemble, computations. There are many domains, if not most, in which proven algorithms and models will persist, that do not require much higher scalability but rather adaptability to new technologies and platforms, and more capacity-type deployments and usages. That said, repeatedly, the request for availability of more highly-scaling solver libraries was voiced. Multi-physics and model coupling is very often cited as increasingly being a crucial need, as are parameter space exploration and design optimisation. These needs arise in particular from market pressures: evolving legal requirements put high pressure on the safety and performance of objects and systems designed; high competition on the market requires accelerated product development cycles. Note that multi-physics simulations can actually generate computing requirements at both capability and capacity levels. Accuracy and robustness of simulations are also a very frequent and highly sensitive concern, often even ranked higher in the priorities than scalability.

Data management is identified as an important and growing aspect, and as more and more intimately linked to computation all along the chain from pre- to post-processing.

The status of end users w.r.t. access to resources varies greatly, and seems mostly related to the size of industrial companies (large ones can afford bigger computers of their own or buy access to external resources). But the wider availability and easier use of larger and more powerful computing resources are considered a strong enabler: they would

unleash the potential of already existing software and probably themselves in turn promote the use of HPC for industrial simulation.

The same issue is observed regarding access to new technologies for experimentation and risk/opportunity assessment: small and medium-sized enterprises often do not have easy access to new technologies or the human resources to actively experiment with them.

3.5.3 *Technical expectations and needs*

Depending on the domains and solvers, distributed memory and/or shared memory, parallel versions of software are available and often hybrid (e.g., MPI+OpenMP) versions. Heterogeneous computing (use of accelerators, GPUs, and more recently, Intel® Xeon PhiTM many-core processors) has been explored by most actors, but for various reasons not often deployed: risk, difficulty or cost of finalising porting and optimisations, lack of demand and of easy access to relevant platforms and resources on the user side. The same holds for scaling up some algorithms and solvers to high levels of parallelism on homogenous large CPU clusters: there have been a number of experimentations and R&D projects, but the operational deployment has not always happened because of lack of demand and available computing resources or because the entire problem-solving chain must scale, rather than only a given step.

In the near or mid-term future, accelerators and heterogeneous computing are considered a major technical trend with high promise, but at the same time this represents a complex and risky move. Other important trends or needs mentioned are

- Remote visualisation,
- Efficient communication and data movements at all levels and on different flavours and levels of architectures (SMP, NUMA; between memory and processors, between CPU and accelerators, between compute nodes in DMP configuration, to and from storage…),
- Virtualisation techniques for cloud-like deployments,
- Efficient scheduling for large ensemble computations,
- Data security, integrity and end-to-end management of privacy,
- Revisited algorithms, better suited to parallel implementation and delivering speed-up, either for general-purpose numerical algorithms or more domain-specific processing, and
- Tools for easier installation, configuration and deployment of applications on complex and diverse system configurations.

Programming environments, not surprisingly, deserve a section by themselves. The call for more standardized APIs, libraries, and/or pragmas and directives for efficient and transparent control of parallelism is general. Tools for profiling, analysing and supporting optimisation are of almost equal importance. Both APIs and tools must become mature enough to accommodate the complexity of often huge legacy codes, and hide (as much as possible) the growing and changing complexity of the underlying hardware.

Cloud deployment is considered with interest by everybody, with high expected potential and impact, and the usual concerns expressed relate to issues such as ease of use, user friendliness; cost; security and confidentiality; availability, reliability and serviceability; ability to handle data management and post-processing as well as computation.

3.5.4 *Obstacles*

The major obstacles identified mostly relate to the interaction and complex equation between the fast and multi-fold evolution of hardware technologies and the weight and inertia of large bases of legacy codes that must be maintained and certified with strict operational criteria on the longer term, while remaining efficient and portable. The choices for the evolution of software are risk-prone in the face of diverse and changing architectural trends (accelerators etc.).

In this respect, mature and standard programming environments and tools as well as easier and affordable access to expertise and experimental platforms are considered key enabling factors.

3.6 ETP4HPC SWOT ANALYSIS

The ETP4HPC has undertaken a strategic SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis to determine how the RTD & Innovation plans developed by its **Working Groups** (reflecting the expertise of the ETP's members) **could add value to European competitiveness**. This analysis has been performed from the point of view of the **European HPC solution providers**. The resulting findings are briefly summarized here and influence the proposed research topics outlined in the next chapter:

Europe is home to advanced **core technologies** that can be used to develop world-class competitive HPC technologies. Examples are embedded processors, networking technology and photonics, as well as non-volatile memory, and hot-water cooling. In this way, a competitive advantage can be gained despite huge entry barriers such as initial investment levels (e.g., when introducing a new CPU product line) and lack of large component vendors.

System software solutions such as new APIs, batch scheduling, I/O workflow and provisioning techniques will be needed for a number of future initiatives, and Europe should strive to achieve leadership in this area by exploiting its current resources and expertise. Market opportunities also exist in improving the runtime support of heterogeneous architectures. High-reliability and high-resiliency solutions are areas with significant improvement potential, representing vast opportunities for European vendors.

As identified in a number of other studies [e.g. IDC-2], Europe's **software expertise** is world-class. Europe has developed a strong open-source community, a large number of vibrant ISVs, and an open research environment. These strengths should be leveraged to achieve global leadership in the area of HPC software design.

Europe's traditional focus on **energy efficiency** (demonstrated through, for example, "green policies", relevant public agencies, and advanced research in this technology) coupled with **expertise in low-energy solutions** across all the elements of the HPC system stack (e.g., low-energy CPUs and systems and highly efficient cooling technology) can facilitate the development of solutions meeting the demand for low-energy solutions (e.g., for data centres) and are a key enabler to extreme computing where power is the limiting factor in system scaling, spearheading innovation in this field on a worldwide level.

Europe is in a position to lead the global development based on **new technologies**, such as mobile device architectures (currently not used in HPC systems), embedded technology, and other advanced technologies (e.g., FDSOI, 3D integration, photonics). These competencies, in conjunction with the **existing CPU and System Architecture expertise** (e.g., represented by "global" companies with a European R&D footprint) can place Europe in the forefront of global efforts to develop, for example, **a highly efficient HPC computer** architecture and build cost-effective and energy-efficient systems in line with the objective of the provisioning of energy-efficient solutions.

Europe is in a unique position to excel in the area of **HPC Usage and Big Data** owing to the experience level of current and potential users (and the recognition of the importance of data by such users as CERN, ESA, and biological data banks) and the presence of leading ISVs for large-scale business applications. Europe should exploit that knowledge to create competitive solutions for big-data business applications, by providing easier access to data and to leading-edge HPC platforms, by broaden the user base (e.g., through Cloud Computing and Software as a Service (SaaS)), and by responding to new and challenging technologies.

Europe's **expertise in HPC applications and business software is world-class**. Europe should focus on application scalability, legacy-code adaptation and porting, and increasing the application lifetime. Europe could also lead in the provisioning of **security and privacy** solutions owing to its expertise in the area and the emphasis on the area due to societal concerns. Europe should take advantage of the opportunities created by **co-design** projects.

Because of its diversity and competence in many high-tech fields, Europe can achieve leadership in **cross-domain technology transfer** between adjacent fields (e.g., telecom and networking). Such efforts can promote Europe's influence the international community.

Although the EU business environment has a number of strengths (such as **a strong SME community** and programmes supporting that sector), the following issues need be addressed:

· The variability of research funding over time and across countries
· Lack of a strong body promoting HPC
· Slow HPC technology adoption processes, in particular by industrial users such as SMEs, because of the high "cost of entry"
· Widespread budget restrictions
· Limited access to/availability of funding for risky research & product development (such as venture capital)
· Lengthy regulatory processes, fragmented policy zones and decision-making procedures

Europe needs to invest in retaining critical talents and become more attractive to experts from other geographies in order to guarantee a **critical skilled workforce** as required by the HPC technology providers and ISVs, thus enabling future growth.

The global HPC eco-system is dominated by system vendors from other regions. Europe remains more of a leading user than a provider of HPC technologies. Emerging economies have superior financial and intellectual resources with lower labour costs. To face these challenges, Europe needs to **align the efforts of its HPC vendors and research centres**, and the creation of this ETP is meant to tackle this issue.

4.
MULTIDIMENSIONAL HPC VISION

The analysis of the above-mentioned inputs shows that there is a demand for R&D and innovation in both extreme performance systems and mid-range HPC systems. Almost all scientific domains and some industrial users want to achieve extreme-scale performance systems as soon as possible. At the same time, there is need particularly expressed by industrial users and ISVs for more flexible, easier-to-use, more productive and more cost-effective HPC systems delivering mid-range performance.

The ETP4HPC HPC technology providers are also convinced that to build a sustainable ecosystem, their R&D investments should target not only the Exascale objective. This market will be too narrow to yield a sufficient return on investment and support sustainable technology development, and that such a strategy will weaken the European players. On the contrary, an approach that aims at developing technologies able to serve both the extreme scale requirements and mid-market needs can be successful in strengthening Europe's position.

As a consequence, the SRA has two dimensions: one targeting the R&D aiming at developing the new technologies able to offer more competitive and innovative HPC systems for a broad HPC market, and one to enhance these technologies with the right characteristics to address the extreme scale requirements.

A third element coming out of the inputs is the trend for new HPC applications. Besides traditional HPC workloads, more and more big-data applications will need to be addressed with HPC solutions. There is also a request from some domains to use HPC systems for the control of complex systems, such as smart grids. The cloud delivery model is yet another trend that will impact the features of future HPC solutions. Accordingly, the SRA has a dimension to address all these new usages.

One last striking input is the concern of some of the stakeholders that HPC development could be limited by some barriers, namely, shortage of skills, insufficient availability of services to fill the gap between potential user demands and HPC solution offerings, usability of solutions or flexibility and vitality of the ecosystem. The SRA defines a fourth dimension addressing these concerns.

As a result the SRA focuses on four dimensions, as shown in Figure 2.

- The creation of new technologies within the entire HPC stack (**HPC Stack Elements**)
- The improvement of system characteristics (**Extreme Scale Requirements**)
- New deployment fields for HPC (**New HPC Deployments**)
- The expansion of the use of HPC (**HPC Usage Expansion**)

The broad field of HPC application software development is not in the scope of this SRA or of the ETP4HPC. Of course, a strong interaction and collaboration with developers, users and owners of HPC applications is mandatory for advancing the HPC system technology, but in the interest of a better focus, this SRA takes requirements from this domain rather than defining its roadmap.

Chapter 5 outlines in detail the first three areas, and Chapter 6 covers the HPC usage expansion and HPC services, the needs of ISVs, and the dedicated focus on industrial SMEs. "Usability" and "Affordability" are key requirements that arose in every dialogue with end-users to lower the "barrier of entry" into HPC deployment. Both are to be seen as comprehensive requirements affecting all research priorities outlined below, and are covered there.

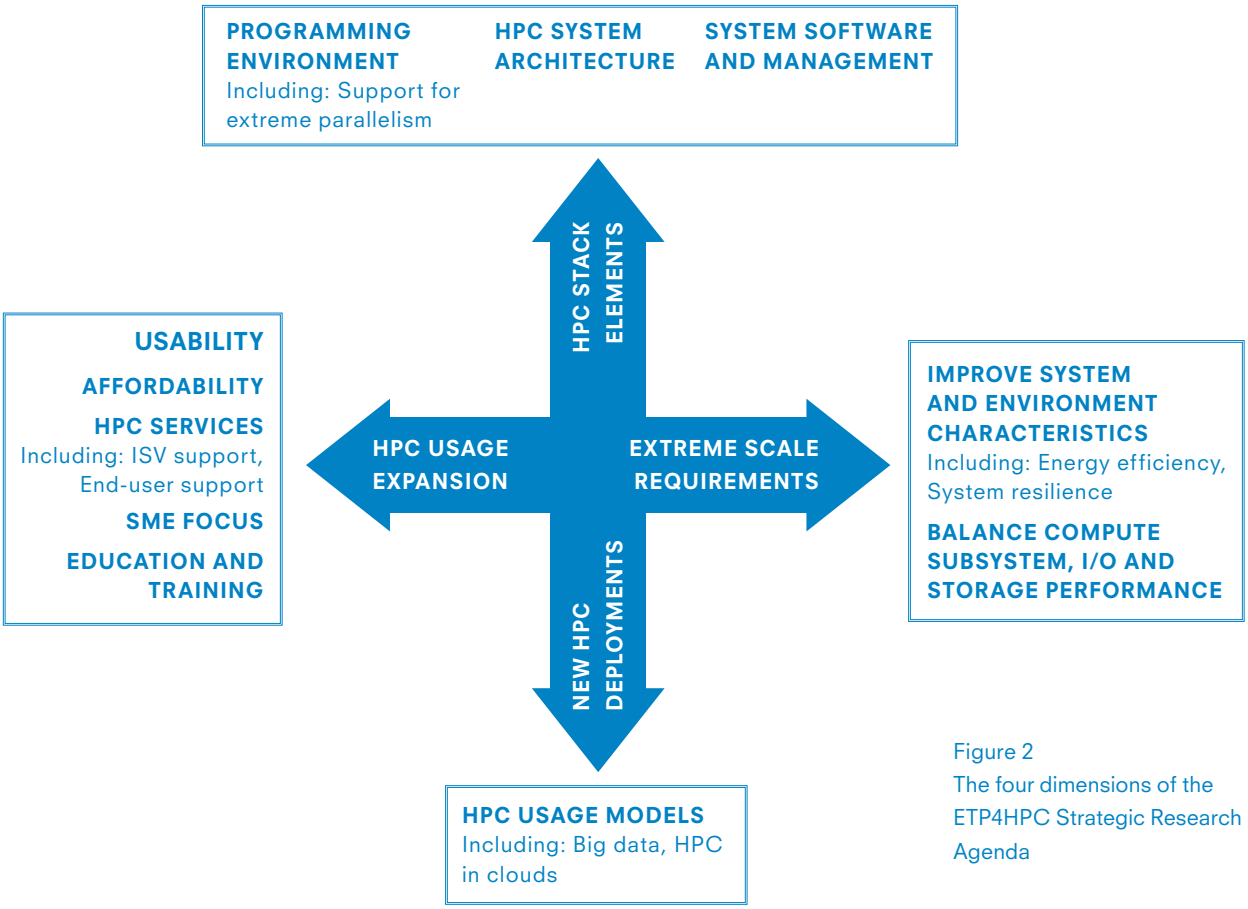


Figure 2
The four dimensions of the ETP4HPC Strategic Research Agenda

HPC STACK ELEMENTS

The HPC system architecture, system software and management, and programming environments are the core elements of a stack of HPC technologies that must be continually improved:

· HPC System Architecture
It is important to consider how different components and their underlying technologies (processors, memories, networks and storage) need to be further developed and integrated to create both extreme scale systems and extremely efficient, mid-size HPC platforms.
· System Software and Management
In the coming years, most system software building blocks will have to be revisited in terms of the node operating system and system run-time environment, interconnect and programming APIs, cluster management, and resource and task management. Heterogeneous workload support for the entire simulation chain from pre- to post-processing, and modularity and ease of integration of stack components will be of particular importance.
· Programming Environment (including support for Extreme Parallelism)
It is essential to address the "software gap" between complex hardware and complex applications for targets ranging from Exascale computing and general HPC use to the area where embedded computing and HPC meet.

The research priorities reflect that range: innovative parallel algorithms and programming approaches; enabling massive parallelism; support mechanisms for energy efficiency; tools and methodologies for software correctness, performance and fault tolerance enabling the right balance between ease of programming, portability and performance, static and dynamic techniques; definition of APIs that will standardise interactions between the various software tools and the system/architecture management layers.

EXTREME SCALE REQUIREMENTS

Extreme scale integration of these elements raises a number of cross-cutting and holistic issues, including energy efficiency, system resiliency, and the overall balance of the system design.

· Improving System Characteristics (Energy Efficiency and System Resiliency)
Energy efficiency and system resiliency are key concerns for existing systems. Both the acquisition and the operational costs of an HPC system are affected by energy efficiency and system resiliency. This is obvious for energy, but resiliency can also affect cost, e.g., by having to compensate for reduced hardware reliability by means of redundancy or by having to provide for lower system effectiveness due to poor reliability. Energy and resiliency are interdependent, as additional RAS (Reliability, Availability and Serviceability) features consume die area and energy, and the right balance between resilience and energy efficiency has to be found.

Energy efficiency is also a global challenge, which needs to be addressed using an integrated approach including the HPC system environment and hosting facility and infrastructure (i.e., by considering cooling, energy re-use, power monitoring and control, power-aware performance metrics and power-efficient algorithms).

· Balance Compute Subsystem, I/O and Storage Performance
Critical to future systems design is ensuring that the right balance of performance is maintained between the capabilities of the compute elements of the systems (including effects of changing workloads on processor and memory) and the data-storage systems such that overall system performance and capability meet or exceed the user needs. This concerns all elements of the I/O stack from middleware to data-storage devices and associated networks.

NEW HPC DEPLOYMENTS

· New HPC Usage Models (including Big Data and HPC in the Cloud)
Beyond the traditional usage model based on highly optimized workloads for a given platform, new HPC usage models are evolving that will impact the full HPC stack - tools, programming models, algorithms, user interfaces, and hardware: ranging from HPC resources as (dedicated) instruments for big science or large-scale data analysis, to HPC in the cloud, commoditisation of HPC for non-HPC experts, and embedded and real-time systems.

HPC computing is quickly becoming user-centric and no longer is machine-centric. Issues such as data integration, data security and privacy will have to be integrated with the underlying HPC computations. Thus, for many HPC users, the idea of having a user-friendly "front end" and no longer having to deal with the "nuts and bolts" of computing is a strong incentive, as long as the required level of performance can be achieved.

It is clear that highly parallel HPC systems cannot be used efficiently without considerable investments in administration and user support. The openness of cloud platforms can be the opportunity to share HPC expertise and facilitate the fast adoption of HPC by a much wider community of enterprises that are new to high-end computing.

HPC USAGE EXPANSION

· A panel of European industrial end-users confirmed that HPC needs to become much more pervasive across all industrial sectors, including Small and Medium-Sized Enterprises (SMEs). Technically, this may require other trade-offs to achieve good affordability and usability, while benefiting from research at extreme scales, such as energy efficiency and ease of programming.
· Another new focus area is HPC services as the link between vendors, ISVs and end users. While large research centres handle most of the service in-house, industrial end-users have a strong desire to focus on their core business and prefer to rely on external expertise and consultancy, especially when migrating to new platforms, tuning the systems for better performance and integrating the use of HPC into their business-process.
· The lack of a well-trained workforce in the area of HPC system design and development, deployment and operation is a known problem that needs to be tackled from several sides. Alongside with defining a Research agenda, the ETP4HPC proposes to set up a working group in collaboration with PRACE, universities and other bodies to help remedy this situation considerably (see also Chapter 6.4 Education and Training).

5.
TECHNICAL RESEARCH PRIORITIES

MAJOR CHALLENGES ADDRESSED

The SRA focuses on specific challenges within the four dimensions shown in Figure 2. They are discussed in detail in the subsequent sections of this chapter because they are the drivers underlying the research topics:

· At the System Architecture level:
  - Future HPC platform architectures
  - Energy cost and power consumption
  - I/O latency and bandwidth (memory, interconnect, and storage)
  - Concurrency and data locality
  - Extreme scale from sub-component to total system
  - Resiliency , Reliability, Availability, Serviceability (RAS)
  - The "storage gap" between storage and compute performance

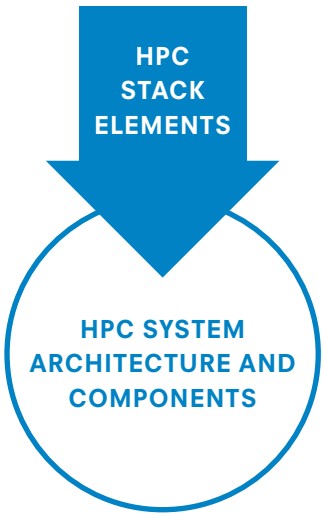
· At the System Software level:
  - Scalability, modularity, robustness
  - Capability for virtualisation
  - Extensive system monitoring
  - Increased system heterogeneity
  - Awareness of data-movement cost

· For the programming environment:
  - Hierarchical models
  - Data distribution and locality
  - Performance analytics
  - Emergence of new parallel algorithms
  - Awareness of data-movement cost
  - Application code migration and re-writing

· Related to new HPC usage models:
  - Explosion of data volumes ("Big Data")
  - Increasing heterogeneity of data
  - HPC workloads in cloud computing

5.1 HPC SYSTEM ARCHITECTURE AND COMPONENTS

5.1.1 *Area*



The goal is to ensure that HPC performance continues to improve in the foreseeable future; an important milestone being the ability to reach Exascale performance levels around 2020 with a performance that is about 1,000 times higher than today's Petascale systems with the same constraints as today in terms of power consumption, system resiliency and space. The evolution of components alone will not be sufficient to reach this target and it will be necessary to introduce new technologies for every component in HPC systems: processors, memory, interconnect and storage. It might be necessary to develop some of the core technologies specifically for HPC within the ETP4HPC. And even though a good part of the necessary core technologies will be developed independently, it will be necessary to influence, adapt and integrate them for the needs of HPC systems.

Table 1 summarizes in the 1st column, the characteristic ranges for today's production HPC Petascale systems taken from the Nov. 2012 Top500 list. The expected values for HPC systems of different capabilities (Exascale for large HPC data centers, departmental Petascale systems and embedded Terascale) in 2020 are then listed in columns 2 to 4. Compared with 2012's Petascale systems all performance characteristics will have to scale by 1000x to reach Exascale while staying within the same constraints as today in terms of system budget, space and power consumption which may not exceed 20 MW.

Table 1
Expected systems characteristics ranges in 2020

	Petascale system (2012)	Exascale / data center (2020)	Petascale / departmental (2020)	Terascale / embedded (2020)
Number of nodes	[3-8] x10^3	[50-200] x10^3 (20x)	[50-100]	1
Computation (Flop/s & Instructions)	10^15	10^18 (1000x)	10^15	10^12
Memory Capacity (B)	[1-2] x10^14	> 10 ^17 (1000x)	> 10^14	> 10^11
Global Memory bandwidth (B/s)	[2-5] x10^14	> 10^17 (1000x)	> 10^14	> 10^11
Interconnect bisection bandwidth (B/s)	[5-10] x10^13	~10^16 (1000x)	~10^13	N/A
Storage Capacity (B)	[1-10] x10^15	>10^18 (1000x)	> 10^15	> 10^12
Storage bandwidth (B/s)	[10-500] x10^9	> 10 x10^12 (1000x)	> 10 x10^9	> 10^6
IO operations/s	100 x10^3	> 100 x10^6 (1000x)	> 100 x10^3	> 100
Power Consumption (W)	[.5-1.] x10^6	< 20 x10^6 (20x)	< 20 x10^3	< 20

When scaled to the same power consumption and system size, we realize that each system architecture parameter must be multiplied by a factor of more than 50x, be it performance, capacity or bandwidth. In addition, we must pay special attention to achieve sufficient reliability at the system level without sacrificing power efficiency, for example by tight integration of more reliable hardware components.

For the drastic improvements sought, we will rely on the development of new technologies. ETP4HPC will drive the investigations necessary for the definition of future HPC platforms architecture. Closer integration of system components are expected to maximize gains in performance and energy efficiency, but it will also be necessary to identify the right trade-offs between components:

· Processors: Develop more energy-efficient processors and system-on-chips (SoCs) leveraging best in class advanced digital silicon technologies.
· Memory: Achieve higher bandwidth, lower power consumption; initially for example by using silicon inter-posers, followed later by full 3D stacking of processors and memory.
· Interconnect: At increasingly higher data rates, optical interconnects may become a better solution than copper if costs are coming down drastically. Optical interconnects could be used for chip-to-chip connections and might even be used advantageously within chips.
· Storage: Non-Volatile Random Access Memory (NVRAM) promises increasing capacity, with fast access at low power, supplementing other memory and storage devices and possibly simplifying the memory hierarchy thus reducing the data transfers and associated energy consumption.

ETP4HPC will drive the research on a future HPC platform architecture to build an exascale solution and beyond. The ETP4HPC will cooperate with other European initiatives which are dealing with key enabling technologies such as Catrene, Eniac and Photonics21 to ensure that the specific needs of HPC systems are taken into account in these programs.

5.1.2 *The outcomes of the SWOT analysis – Main Strengths and Opportunities*

The European HPC eco-system should focus its efforts on the following areas:

· Build on existing European innovation, expertise and research infrastructure
· Leverage the current CMOS technologies. Techniques like FinFET and FDSOI will allow shrinking manufacturing processes from today's 20-28 nm to 7-10 nm in 2020.
· Exploit the expertise in the use of extreme ultraviolet for lithography (now at the laboratory prototype stage in Europe). This technology will greatly facilitate the continuous reduction of structure sizes, which is at the heart of Moore's law.
· Over the past decade, Europe has spearheaded research and development in photonics and optical interconnect. Building on this research, Europe should develop and expand its efforts on HPC interconnect solutions that are cost-competitive at either the rack or the motherboard level.
· Europe has a strong presence in embedded and mobile computing, for example, in the area of CPUs and GPUs for mobile systems (ARM, Imagination Technologies, etc.) and complex efficient SoCs could be designed together with chip companies (STMicroelectronics, STE, Infineon, NXP, etc.). As these components were designed for energy efficiency, some may be reused for HPC, or evolved to serve the HPC market. Furthermore, global companies like AMD, IBM and Intel have significant R&D and manufacturing facilities in Europe.
· More recently, Europe has quite actively been researching Spintronics, one of the Non-Volatile Memory (NVM) variants. Europe now accounts for 1/3 of all publications on this topic. As NVM is much faster than current disks and consumes very little energy, it can enable a new class of storage devices for the IT industry, including HPC.
· Europe has a strong background in HPC applications and analysis tools development; this expertise should be applied in the definition of optimal architectures for Exascale and for mid-range HPC systems.

5.1.3 *Research topics and expected results*

The Platform architecture
The Platform architecture requires a holistic approach. A successful architecture will be the result of a complete Multi-Domain Optimization process that addresses the four design challenges mentioned below in a coherent vision. This is true for reaching Exascale but also for follow-on systems. Architecture definitions will need a continuous capacity to perform top-down and bottom-up iterations to understand the impact on hardware and applications. It is clear that this task will need important investments and tool support to be effective. This is a significant opportunity for Europe.

The Energy and Power Challenge
This challenge arises from the fact that no combination of currently available technologies will be able to deliver sufficiently powerful systems within the practical limitations for power consumption (which are estimated to be in the 20 MW range).

The Memory, Interconnect and Storage Challenge
This challenge addresses the lack of mature memory, storage and interconnect technologies which will be able to fulfil the I/O latency and bandwidth requirements of future Exascale systems within an acceptable power envelope.

The Concurrency and Locality Challenge
The continued trend towards many core processors means that the core count grows significantly faster than the per-core performance. This causes a major challenge for application developers who need to scale their applications to many more threads. In addition, as systems grow in size resources become more distributed and data locality becomes essential for application performance.

The Resilience Challenge
This concern addresses the explosive growth in component count for the top HPC systems as well as the need to use advanced technology at operating points such as low voltage and high temperature, where individual devices and circuits become more sensitive to operating environments and increased per-component failure rates can be anticipated.
To respond to these challenges, each component of the computing technology must evolve. This evolution cannot be studied without a full understanding of the impact of the new technology on the programming and execution practice; it must address the balance between processor performance, memory capacity and access as well as the data movement between compute nodes and to storage.

5.1.3.1 Energy and power

When designing a supercomputer, energy efficiency is the first constraint to be addressed. Apart from all the technical problems linked with thermal and power issues, it has a major impact on the total cost of ownership of such a system. Energy efficiency is a transversal problem, which needs to be addressed at every level, starting first with semiconductor and memory technologies, and then covering the chip and system design and implementation.

- Ultra-low power CMOS
- 2.5D/3D Heterogeneous integration
- Energy proportionality (CPU, interconnect, switches, etc.)
- Heterogeneous architectures and accelerators (GPU, vector, …)
- Integration of the different components into a SoC (CPU, GPU, memory controller, IO controller, network interface and switch) or system on a package.

These basic technology research topics have to be mastered to improve the efficiency level from 2 GFlop/s/W to above 100 GLOPS/W for the compute nodes.

5.1.3.2 Memory and storage

Memory technology is one of the cornerstones of system efficiency. The main issue of a computing system is to bring data to the computing node at the required speed and within the lowest energy budget possible. Today new memory technologies are emerging with non-volatile properties and good access performance. None are yet at par with the fastest classical DRAM and SRAM but these technologies open new opportunities for the architecture of the memory system.

Table 2
Expected memory characteristics ranges in 2020

	Traditional technologies			Emerging technologies			
Memory type	DRAM	SRAM	NAND	FeRAM	STT-MRAM	PCRAM	ReRAM
Physical effect at work	Silicon	Silicon	Tunnel effect	Ferroelectrics	advanced Spintronics	Phase-Change	Redox / Memristor
Technology status	Mature			product	products being introduced	1st Generation	Research
Read time (ns)	<1	<0.3	<50	<45	<20	<60	<50
Write erase time(ns)	<0.5	<0.3	10^6	10	10	60	<250
Retention time (years)	N/A	N/A	>10	>10	>10	>10	>10
Write endurance (nb of cycles)	10^16	10^16	10^5	10^14	10^16	10^9	10^15
Density (Gbit/cm2)	6.67	0.17	2.47	0.14	1	1.48	250
Technology potential scalability	Major technological barriers			Limited	Promising	Promising	Promising
Exascale applicability	No recognised alternative	Potential to scale suitably?	Tech barriers and endurance, may be overtaken by other NV devices	Unlikely	Promising, potential successor for DRAM	Very Promising	Unclear, potential as compatible SoC device

As seen in Table 2, the characteristics of these new memories (estimation as of mid-2012) can drastically impact the memory structure especially in the case of HPC, where checkpointing is an important function. The ability to include substantial non-volatile memory within nodes, within interconnects and networks or within I/O storage nodes is essential to scale up the system performance. Optimising the use of that technology, its placement in the system, its management and resultant data flow to ensure seamless data migration between the multiple tiers of memory and storage is the key area for HPC research.

Such a non-volatility of the memory can also have a huge impact on the energy policies of the system and hence on its energy efficiency. This requires research into the granularity of energy control (at chip or board level) and of data placement.

5.1.3.3 Interconnects

Interconnects are the backbone of HPC systems. It is essential to scale the communication bandwidth with the performance of nodes and components and minimize the remote data access latency while keeping within today's energy consumption limits. Interconnects must be tightly integrated with systems components and provide support for accelerating global communication operations. Such technologies are critical elements of Europe's HPC efforts.

The exponential increase in capacity, processing power, bandwidth, and density in future systems will partly be supported by conventional copper links for cost reasons and partly by high-density optical channels within HPC system enclosures first at the rack level, then at board and ultimately the chip level.

5.1.3.4 Concurrency and locality

Following current trends, it is expected that an Exascale system will have millions of cores and it will be a major challenge for applications to scale to that level. The optimal CPU design will be a trade-off between the number of cores, their power consumption and the global application performance. It is important to find the right size for the CPU cores and, to trade-off energy-efficiency, per-core performance, and achievable performance for key application classes.

As systems get bigger, access to remote data becomes a challenge and any feature that improves data locality will increase the application's performance and reduce the power consumption.

Finally, fully integrated interconnects will lower the impact of remote data access and improve the overall system efficiency and reliability.

5.1.3.5 Resilience

As HPC systems get larger and incorporate many more parts, the overall system resilience decreases. This could be partially counterbalanced if the reliability of individual components can be improved.

The integration of the various components of the system nodes into a single chip reduces the number of parts and connections thus improving reliability and lowering power consumption.

Other systems aspects must also be addressed such as built-in error correction, component redundancy and fail-over mechanisms, and in particular software resiliency and resource management mechanism. As error correction and redundancy will increase die area and power consumption, a good balance with the objective of energy efficiency has to be found.

Checkpoint/restart is an essential component of the global system resilience. If it is done locally and leverages new non-volatile memory technologies, the overhead for performing checkpoints and the energy used will be optimized. Further details can be found in Chapter 5.4 Energy and Resiliency.

5.1.3.6 Exascale system architecture

To achieve the ambitious goal of an Exascale class computer with applications able to take full benefit of it, the entire architecture process must be rethought. This is a multi-domain optimization problem requiring a collaborative design that integrates all aspects of the system architecture, hardware components, software and applications.

The Exascale system design will need to explore and model the potential of new technologies. It must address the balance between processor performance, memory capacity and access as well as data movement between compute nodes and to storage.

Moreover, it is important to remember that the development of such an Exascale system can enable efficient and competitive Petascale servers for departmental use and Terascale devices suitable for embedded systems, both of which will be very valuable for a wide range of industrial users.

5.1.4 *Milestones*

Deadline	Milestones
2015	**M-ARCH-1:** High-performance local data access (1GB/s)
	M-ARCH-2: NVRAM at DRAM capacity
	M-ARCH-3: Chip-to-chip photonics industrialization
	M-ARCH-4: ~100 kW/PFlop/s (End-2015)
	M-ARCH-5: Low dissipation high-performance compute module using silicon interposer (prototype)
	M-ARCH-6: 3D integration of high-throughput memories with high performance compute module (prototype - End-2015)
2018	**M-ARCH-7:** NVRAM available at a price and capacity that enable its integration into HPC systems
	M-ARCH-8: 45 kW/P/
	M-ARCH-9: High-speed, low-power processor-internal data transfer (prototype)
	M-ARCH-10: Silicon photonics for chip-to-chip interconnects
2020	**M-ARCH-11:** Exascale system at <=20 MW power consumption (20 kW/PFlop/s)

5.2 SYSTEM SOFTWARE AND MANAGEMENT

5.2.1 *Area*

In the coming years the system software will be a critical component that controls and supervises the productivity, power consumption, and the resilience of supercomputers of growing scale and complexity – up to Exascale and beyond. The hardware complexity and heterogeneity of foreseeable Exascale components involve a redesign of the software system solution to make it more flexible, robust, and scalable, as well as more modular to support a wider diversity of workloads. The power cost of data movement and ubiquitous Big Data will also have a critical impact on the software system solution, which will need to be more data and locality sensitive.

Most of the system software building blocks will need to be revisited:

The **operating system** (OS) will have to take into account new intra-node architectures with many cores, intensive and dynamic multi-threaded concurrent task execution environment, and a heterogeneous memory model.

Extremely scalable performance of applications will require the re-definition of the borderline between **Runtime and OS**.

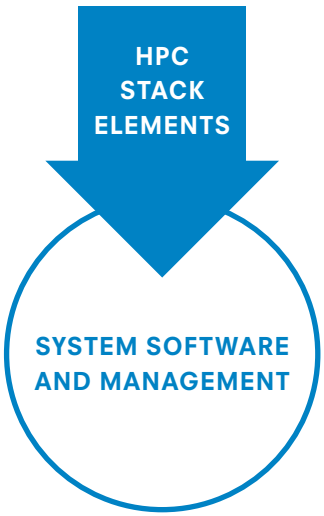
New requirements such as virtualization support, data centric intra-node scheduling, and real-time capabilities need to be added to current OS solutions.

Interconnect management needs to reflect affinity, congestion control, dynamic routing, topologies, and diagnostics, and to allow the control of power consumption of the fabric.

Cluster management will face a real monitoring challenge, because of the number of resources and events to be reflected in on-the-fly data analysis and post-mortem data mining. Event-driven health-checking and introspection are vital for stable operation, and for ensuring that the system stays within the power budget. Modularity and heterogeneity of the system configuration require the development of an advanced integration model.

Resource management and job scheduling have to enable extreme allocation flexibility, tightly coupled with applications and must take new allocation criteria into account (e.g. network topologies, interconnect bandwidth, I/O associated workflow, CPU architectures, power targets and caps).

Power is a critical cross cutting issue. All Power items mentioned in this chapter belong to a global cross cutting solution fully described in a dedicated section 5.4 Energy and resiliency.



5.2.2 *The outcomes of the SWOT analysis – Main issues to be addressed*

The European HPC eco-system should focus its efforts on addressing the following issues:

· There has been a long tradition of designing, administrating, and maintaining large computers in the European Community. Leveraging the skills of a full ecosystem (computer manufacturers and designers, large computer centres and specialized HPC service SMEs), Europe has a chance to tackle the global integration challenges to build a production Exascale system software stack.
· The heterogeneity of the hardware infrastructure and the required modularity of the target system will require a complex integration and validation process, on-site

integration, and check facilities that have not yet been specified.

· A hardware infrastructure without open specifications may inhibit the development of a software system solution.

· Increasing levels of parallelism in multi- and many-core chips and the emerging heterogeneity of computational resources coupled with energy and memory constraints force a re-engineering of operating systems and runtime environments. New interfaces must be standardized. New operating system identities and responsibilities must be defined to enable advanced applications development

5.2.3 *Research topics and expected results*

5.2.3.1 Operating system (OS)

The purpose of an operating system is to bridge the gap between the physical resources provided by a computing system and the runtime system needed to implement a programming model. Given the rapid change in resources and programming models, basic operating system APIs must be defined for the Exascale community. A common set of APIs will be defined, that can be used by a runtime system to support fully autonomic resource management, including adaptive management policies that identify, and react to, load imbalances and intermittent resource loss, and that supervise power consumption of all system components and make power management mechanisms at the hardware level available.

To achieve this goal, the operating system must expose low-level resource APIs and the runtime must be aware of the context in the application of a specific computation.
Operating systems have the root responsibility of hardware resource exposure. Therefore, without OS re-design for new Hardware and new requirements for OS interaction with applications run time, there is a great risk that the performance of applications will also be substantially below the peak performance of the parallel environment.

Research topics:

To address the above issues we envision, among other things, the following long-term solutions:

· Enhance the programming models so that they offer a simple but efficient abstraction of the hardware.
  - Improve hardware abstractions in compilers and compiler features, e.g. optimization and parallelism extraction.
  - Add ability to seamlessly integrate non-standard (e.g. power-optimized, high-throughput) computation cores. Provide programming model developers with efficient API for run-time.

· Extend runtime systems by developing new algorithms for managing data locality, scheduling, I/O, coherency, power use etc. better

· Produce APIs to support inter- and intra-node communication, thread management, and explicit management of the memory hierarchy provided by the entire system.

· Effective and efficient management of hybrid memory systems for HPC usage scenarios.

· APIs to support energy management and resilience will be essential.

· Fault tolerant/masking strategies for collective OS services are indispensable.

· Research on deconstructing the OS for extremely scalable applications by dedicating some cores to applications and others to the operating system to reduce OS jitter.

· Research into runtime systems that support emerging fault-tolerant and fault-resilient computational models.

Expected results

Capabilities to build different OS instances and runtime systems adapted to applications´ needs with respect to new emerging hardware specifications, e.g. heterogeneous compute elements, embedded cores, etc.

5.2.3.2 Interconnect management (IC)

Interconnect topology management and routing are expected to react to communication requests from applications or I/O in real time and to create an optimal traffic pattern. First, this challenge impacts the routing algorithms which will need to be involved. Secondly, optimization and scaling of application requests will require more and more OS bypassing. A new border between low level system control in kernel space and application execution in user space on top of interconnect interfaces needs to be specified.

Congestion and trouble-shooting diagnostics of such a network will be a real challenge, requiring the development of

new management solutions. A further area of improvement is the power management of the interconnect fabric and the NICs.

At the interconnect adapter level, driver and low-level interfaces will evolve depending on the hardware technology and on the needs of new programming models and applications. Here, tight cooperation between system software developers, system architects, and programming model designers will be necessary to integrate new low level protocol capabilities and define a well-adapted API for higher levels of programming software.

Research topics:

· Adaptive and dynamic routing algorithms
· OS bypass and hardware interface integrity protection
· Congestion control and concurrent access regulation
· Power management of interconnect fabric and NICs
· Scalable interconnect management tools
· Intra-node and inter-node network model

5.2.3.3 Cluster management software (CM)

Exascale supercomputers are expected to consist of hundreds of thousands to millions of components including compute and service nodes (I/O, gateways, and management), network and interconnect infrastructures and Petabytes of attached storage. Already today, it is difficult for Petaflop supercomputers offering a scalable framework to launch parallel commands, gather and aggregate outputs, monitor equipment, collect event data, identify faulty components, etc. – as all this requires sophisticated cluster management systems and health-check routines.

These features will be key factors for running Exascale supercomputers efficiently with the high productivity required by unprecedented application workflows. To achieve this critical target, the objective is to develop an innovative cluster management framework that will be able to scale to millions of components.

Research topics:

· On-the-fly analysis monitoring model
· Models based on simple event logs are not scalable; new event-collecting and filtering models must be prepared that can react within a few seconds even if the number of events is huge. It will also have to cope with new classes of events related to power management.
· This model will have to provide capabilities for sharing this data with programming tools. Graphical supervision for Exascale must be designed.
· Maintaining a clear view of the configuration and status of HPC systems is indispensable as they are exceedingly complex and susceptible to small perturbations having an extraordinary impact on performance, consistency, and usability.
· Data mining , data analysis — on-the-fly and post-mortem, introspection
· Data mining and data analytics tools to facilitate the management of large amounts of event data and to identify key factors that need to be addressed for diagnostics and corrective actions.
· Flexible system-image configuration and management
· The huge variety of hardware, programming environment, and applications requirements will necessitate new tools to compose, integrate and store system images.
· System resource maintenance and provisioning
· A clear view of the availability of resources is necessary for the provisioning of alternative resources in case of failure.
· System security
· System management also has to ensure system integrity, with system security being a major factor.
· Performance data

Performance analysis of Exascale applications depends on having access to up-to-date detailed data on CPU, memory and interconnects performance. An established method to provide this for CPUs is performance counters. For future HPC systems, such data has to be provided for all system components, with sufficient scalability and minimal overhead.

Expected results:

A strong operation and monitoring framework with tight integration and configuration capabilities as well as powerful supervision tools.

5.2.3.4 Resource management (RM) and job scheduling (RMJS)

The current centralized control of the RMJS will hardly be able to handle the continuous increase of the number of computing resources, the heterogeneous environments, the need for optimizing energy consumption, and the growing scheduler complexity. Hence the growing need of new approaches able to deal with the above issues.

Research topics:

· New distributed architecture and scheduling algorithms
· Centralized control with only one management node will certainly not be able to scale up and stand the complexity and multitude of the scheduling choices. So, multiple management nodes along with parallelized internal resource and job management processing have to be studied. In addition, scheduling algorithms will need to take into account power related objectives.
· Integration of I/O criteria and workflow
· The amount of data to move and save will be a new constraint for the job scheduler. It will need to integrate data locality into allocation criteria and to coordinate resource scheduling with the I/O workflow.
· Real-time data aggregation and analysis
· The use of HPC in the Big Data space will require the provisioning of real-time data analysis and reduction, its integration into HPC workflows and management/scheduling by the RMJS system.
· Adaptive scheduling and application interaction
· Multi-objective adaptive scheduling in a dynamic environment will probably be a way to deal with most of the above issues. However, the complexity of the scheduling will be enormous and its scalability will be an important barrier. Since energy usage and power control mechanisms are themselves highly dynamic and application dependent, adaptive scheduling has a high potential to help with the energy challenge.
· Resilient framework
· Exascale systems require a failure-tolerant application environment implementing at least effective checkpoint/restart features.

Expected results:

Dynamic resource allocation capabilities and powerful multi-criteria scheduling algorithms, power-aware scheduling, advances in adaptive scheduling.

5.2.4 *Milestones*

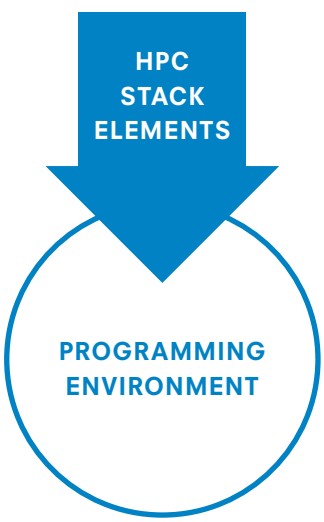
Deadline	Milestones
2014	**M-SYS-CL-1:** Model for on-the-fly analysis and control
	M-SYS-IC-1: Scalable interconnect management
	M-SYS-OS-1: Specification of an external standard for OS / Runtime and API for system-wide power management, explicit memory management
	M-SYS-RM-1: Power-aware resource management and job scheduling (RMJS)
2015	**M-SYS-CL-2:** Model for system resource maintenance and management
	M-SYS-IC-2: OS-bypass and hardware interface integrity protection
	M-SYS-IC-3: Interconnect adaptive and dynamic routing algorithm and congestion control
	M-SYS-OS-2: New lightweight base OS development with support of virtualization and HPC hypervisor
	M-SYS-OS-3: System security management
	M-SYS-OS-4: Prototype for system simulation environment
	M-SYS-RM-2: New Scalable scheduling enhancement with execution environment and data provisioning integration
2016	**M-SYS-CL-3:** On-the-fly data analysis, data mining
	M-SYS-CL-4: i) Modular generation and deployment of OS instances and runtime environment and ii) Flexible system image configuration and integration
	M-SYS-OS-5: Hybrid and heterogeneous memory and CPU OS support
2017	**M-SYS-OS-6:** System with real-time capabilities
	M-SYS-RM-3: New multi-criteria adaptive algorithms: heterogeneity-/memory- and locality-aware
2018	**M-SYS-CL-5:** Graphical supervision for Exascale
	M-SYS-OS-7 : Scalable system simulation environment
	M-SYS-RM-4: Fault-tolerant MPI and checkpointing
2020	**M-SYS-CL-6:** Load balance with tolerance to noise and temporary shortage of resources
	M-SYS-IC-4: Intra-/ inter-node network model implementation
	M-SYS-OS-8: Resilient OS with API
	M-SYS-OS-9: Hypervisor and deconstructed OS
	M-SYS-RM-5: Resilient framework

5.3 PROGRAMMING ENVIRONMENT

The development of efficient, massively parallel and energy-efficient applications is highly dependent on the capabilities of the programming environment. The current state of the art renders the development task tedious and error prone, and it does not lead to very portable codes. This strategic research agenda defines a set of directions that will lead to new programming environments that simplify massively parallel programming, debugging and performance analysis, while also taking into account such comprehensive issues as energy consumption and fault tolerance. To increase the chance of success, this agenda will have to be pursued in a spirit of co-design with application scientists and in a vertical integration manner to foster efficient interaction of the underlying software components in the programming environment, which will lead to efficient applications codes in terms of all vertical criteria.

Although revolutionary approaches should be encouraged, it is of paramount importance to also address legacy codes. This SRA wishes to promote the emergence of technologies that will foster an incremental migration of existing codes to massively parallel systems.

The lifespan of any large code, either for research or for commercial purpose, is far greater than that of any computer. The only practical way for a programmer to write code that survives the migration from one generation of computer to the next is to rely on standards. Yet, standards need to evolve to cope with the emergence of new technologies. It is expected that the research will support the emergence of new standardized APIs.



5.3.1 *Support for Extreme Parallelism*

Here the strategic research agenda targets a programming environment that handles the complete range of HPC systems, because portability between systems of different scales is just as important as portability between systems at the same scale. However, the extreme parallelism required of applications to be executed on the extreme-scale systems means that specific features and aspects of programming environment components need to be part of the research agenda for the programming environment, even if they do not impact applications running on smaller-scale systems. The optimal solution would be that this should be transparent to the application developer. Here are some of the main issues for supporting extreme parallelism, which will also be included in the detailed descriptions below:

· Hierarchical decomposition: at the extreme scale, a "flat" model of MPI ranks will be insufficient to achieve high performance. Future programming models must either expose the hierarchy to the programmer or manage it automatically.
· Data distribution and locality: at the extreme scale, the cost of moving data across the system will be enormous, increasing the need for intelligent data placement. In addition, algorithms will have to avoid global barriers and non-essential collective communication.
· Performance analytics: performance analysis at the extreme scale is intractable with current approaches because of the enormous amount of data involved. As system sizes increase, there will be a growing need for intelligent analysis tools that use techniques from data mining, clustering and structure detection to focus the programmer's attention on the most relevant aspects and actively avoid unnecessary details.
· New parallel algorithms: extreme-scale systems will tax existing parallel algorithms to the limit; hence for extreme-scale problems, new algorithms will be needed.

5.3.2 *Area*

In defining the research agenda for the programming environment, we aim to provide holistic solutions for the HPC technology software stack, that lie between those for the applications software layer and the system's software and hardware layers. The central problem addressed is the "software gap" between the potential performance of evolving advanced HPC systems and what "pay-load" applications achieve in practice. This software gap is caused by the long development times and the increasing system complexity.

There is a strong interrelation with those chapters of the SRA that address the comprehensive issues. The research themes explicitly addressing extreme parallelism play a key role, but the targets range from the issues of Exascale computing to the general HPC use and acceptance to the area where embedded computing and HPC meet. The research priorities reflect that range: innovative parallel algorithms and programming approaches for complex and evolving architectures; enabling massive parallelism; support for mechanisms for better energy efficiency; tools and methodologies for software correctness, and fault tolerance. Of great importance is the need to find the right balance between ease of programming, portability and performance, and between static and dynamic techniques. Furthermore, we target the definition of APIs that will standardize interactions between the various software tools and the system/architecture management layers.

5.3.3 *Outcome of the SWOT analysis - Main issues to be addressed*

In September 2010, representatives from major companies meeting at the HPC User Forum [HPC User Forum] concluded that "there is a growing apprehension within the HPC community that many codes may need to be fundamentally re-thought and re-written in the next five years or so — to take better advantage of larger, more highly parallel, increasingly heterogeneous HPC systems." While this reality is hitting the HPC community first because of its demand and need for the ultimate performance, the same issues will affect all other ICT markets in the near future as technology design decisions from the high-end start to reach the desktop and embedded domains. This was highlighted in the report from an Expert Workshop on "Towards a breakthrough in software for advanced computing systems" in July, 2012[2], which noted that the lessons learnt by the HPC community could and should be transferred to the embedded domain (although the real-time nature and reliability requirements of embedded applications pose additional challenges).

Billions of lines of codes will need to migrate over the next 10 years. Some of them will become obsolete and will need to be redeveloped, but a large majority of the software-developing business and research community cannot afford a full rewrite. The cost of development and deployment of applications on high-performance system is one of the main obstacles to generalize the use of HPC. Thus, mechanisms for "migration" to advanced computing systems are required.

The most influential parameter behind application development and deployment is the programming environment. The move to massively parallel processor architectures has redefined the landscape of runtime and programming technology, and the market has rarely been so open to new technologies and standards, particularly to allow for the development of maintainable applications software. Furthermore, without significantly improved tools, the complexity of extreme parallelism cannot be handled without extreme programming complexity. Programmers are waiting for solutions to abstract coding and allow application portability. The entire software industry will be impacted by this research, and the ability to provide new tools with a large market acceptance will be a tremendous asset for European ISVs, and will promote the use of HPC technology at all levels of the economy.

As identified in a number of other studies [IDC 2012], Europe's software expertise is world-class. Europe has developed a strong open-source community, a large number of vibrant ISVs and an open research environment. These strengths should be taken advantage of to achieve global leadership in changing the paradigms in HPC software design.

5.3.4 *Research topics and expected results*

The proposed agenda is expected to lead to the following:

- Enhanced code portability and quality
- Reduced software development and maintenance costs (also through availability of appropriate debugging tools)
- Contribution to standardization effort
- Efficient exploitation of hardware
- More efficient application software (through the algorithmic & programming APIs/language innovations, and performance-monitoring and analysis systems)
- Lower operational costs thanks to software-driven energy management
- Greater resilience arising from fault-tolerance features within the programming environment.

[2] http://cordis.europa.eu/fp7/ict/computing/documents/advanced_computing_ws_report.pdf

The remainder of this section expands on the specific topics.

5.3.4.1 Parallel Programming APIs and Languages (API)

Defining new programming APIs and languages for expressing heterogeneous massive parallelism in a way that provides an abstraction of the system architecture, promotes high performance and efficiency, and interoperates with energy and resiliency management systems.

The APIs and languages should interoperate with MPI, where the maximum benefit from the hybrid approach is obtained, including any MPI communication overlapping with local communication and computation.

Measurements of success for this topic encompass the following metrics:

1. Ease of programming
2. Ease of porting of legacy codes, with minimal source-code changes
3. Code efficiency
4. Standardization
5. Portability across the range of available heterogeneous systems
6. Benchmark and mini-apps coverage, reflecting the needs of full-scale production applications

5.3.4.2 Runtime supports/systems (RT)

Develop runtime systems to implement programming model abstractions, achieving high performance and low overhead, while abstracting system complexity.

Measurements of success for this topic encompass the following metrics:

1. Runtime support for auto-tuning and self-adapting systems and dynamic selection between alterative implementations
2. Management and monitoring of runtime systems in dynamic environments
3. Runtime support for communication optimization: data-locality management, caching, and prefetching
4. Runtime support for fault tolerance and power management

5.3.4.3 Debugging and correctness (DC)

Debugger technology that is able to support applications developed on and for heterogeneous computing systems using both current and non-conventional programming models, languages and APIs, and deployed on the full range of target computer systems up to Exascale.

Measurements of success for this topic encompass the following metrics:

1. Scalability
2. Automation/programmer efficiency
3. Handling of heterogeneity
4. Understanding programming model abstractions

5.3.4.4 High-performance libraries/components (LIB)

Libraries and components that exploit massively parallel heterogeneous computing resources and are capable of adapting to changing execution contexts (caused, for example, by faults).

Measurements of success for this topic encompass the following metrics:

1. Scalability and efficiency of the “solutions/services” delivered to applications
2. Usability for a broad range of application software
3. Ability to maintain efficiency by dynamic adaptation to execution environment changes
4. Portability/standardization
5. Ability to exploit the developments in other research topics, specifically RT, DC, and PT (see below).

5.3.4.5 Performance Tools (PT)

Develop scalable tools and techniques for practicable analysis of the entire program performance.

Measurements of success for this topic encompass the following metrics:

1. Scalability
2. Ergonomic aspects
3. Ability to handle heterogeneous systems
4. Portability/standardization
5. Ability to address ”umbrella” issues (e.g., energy, resilience)

5.3.6 *Milestones*

Deadline	Milestones
2014	**M-PROG-API-1:** Develop benchmarks and mini-apps for new programming models/languages
2015	**M-PROG-API-2:** APIs and annotations for legacy codes[3]
	M-PROG-API-3: Advancements of MPI+X approaches (beyond current realisations)
	M-PROG-DC-1: Data race detection tools with user support for problem resolution
	M-PROG-LIB-1: Self-/auto-tuning libraries and components
	M-PROG-PT-1: Scalable trace collection and storage: sampling and folding
	M-PROG-RT-1: Runtime and compiler support for auto-tuning and self-adapting systems
	M-PROG-RT-2: Management and monitoring of runtime systems in dynamic environments
	M-PROG-RT-3: Runtime support for communication optimization: data-locality management, caching, and pre-fetching
2016	**M-PROG-API-4:** APIs for auto-tuning performance or energy
	M-PROG-LIB-2: Components/library interoperability APIs
	M-PROG-PT-2: Performance tools using programming model abstractions
	M-PROG-PT-3: New metrics, analysis techniques and models
	M-PROG-RT-4: Enhanced interaction between runtime and OS or VM monitor (w.r.t. current practice)
2017	**M-PROG-API-5:** Domain-specific languages (specific languages and potentially also development frameworks)
	M-PROG-API-6: Efficient and standard implementation of PGAS
	M-PROG-DC-1: Debugger tool performance and overheads (in CPU and memory) optimised to allow scaling of code debugging at Peta- and Exascale
	M-PROG-DC-2: Techniques for automated support for debugging (static, dynamic, hybrid) and anomaly detection as well as or checking the programming model assumptions
	M-PROG-DC-3: Co-design of debugging and programming APIs to allow debugging to be presented in the application developer's original code and also to support applications developed through high-level model descriptions
	M-PROG-PT-4: Performance analytics tools
2018	**M-PROG-API-7:** Non-conventional parallel programming approaches (i.e., not MPI, not OpenMP/pthread/ PGAS, but targeting asynchronouzs models, data flow, functional programming, model-based)
	M-PROG-LIB-3: Template-, skeleton- or component-based approaches and languages
	M-PROG-RT-5: Scalable scheduling of million-way multi-threading
2019	**M-PROG-LIB-4:** New parallel algorithms parallelisation paradigms
	M-PROG-PT-5: Inspection of data locality at Exascale level

[2] Note that the migration of a major application to a new programming paradigm generally takes 5 to 7 years.

5.4
ENERGY AND RESILIENCY

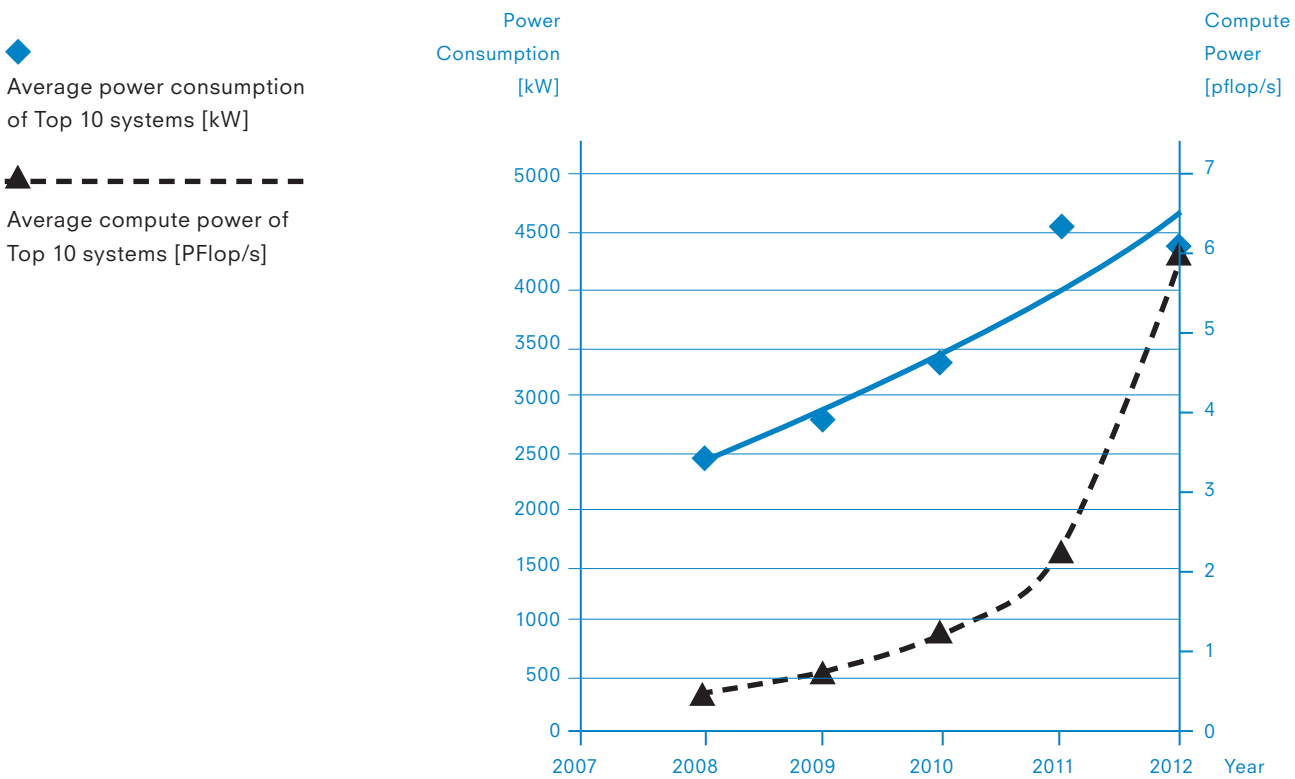
5.4.1
Area

Power, energy and reliability are widely recognized to be among the most important areas of improvement for present and future HPC systems. They represent the two faces of the same coin: both are significant challenges that need to be addressed for future large HPC systems (multi-Petaflop/s and Exascale systems) to be viable and sustainable, and both require holistic solutions at the system and, in the case of energy efficiency, the data-center level.

Figure 3 shows the average power consumption of the world's 10 most powerful supercomputers according to the Top500 list since the power values were started to be recorded in 2008. The power consumption of these top systems is currently in the range of 1 to 10 Megawatts and rising. In view of the increasing burden of the operating costs of current HPC systems and even more so by the projected costs of future (e.g., Exascale) systems, which will simply be no longer sustainable, energy costs are quickly becoming a subject of intense research in HPC.

Figure 3
Average power consumption and average compute power of the Top 10 systems of the Top 500 list

Average Power Consumption and Compute Power of the 10 most Powerful Systems Worlwide



System resiliency is the ability of a system to recover quickly from system faults so as to limit the impact of the faults on the system's operation and functionality and throughput. However, as keeping the application running at an unduly high cost in power or performance is unacceptable, resilience research aims to produce total system solutions that are balanced in their impact on power, performance and cost.

RAS is an acronym that indicates reliability, availability and serviceability. Reliability is the capability of a system to limit the number of faults in time. Availability is the amount of time a system actually operates as percentage of total time it should be operating. Serviceability is defined as the degree to which a system is easy to service, i.e., to install, configure, monitor, maintain and troubleshoot a system.

Resiliency and RAS are closely related. If a system does not fail, system resiliency becomes redundant. The opposite is also true: if the reliability of a system decreases, system resilience becomes much more important. A resilient system has a greater availability because it can lower the impact of faults over the uptime. A system that is more serviceable can be, by definition, more resilient and more available. Typical examples for this are hot swappable components that can be substituted without stopping system operation.

Clearly, the massive scale-out of HPC systems intensifies the importance of resiliency for HPC systems, in particular for large Petascale and future Exascale systems. These systems will typically gather from half a million to several millions of CPU cores running up to a billion of threads. Extrapolating from current knowledge and observations of existing large systems, it is anticipated that large systems will experience various kind of faults many times per day. It is also anticipated that the current approach to resilience, which relies on automatic or application-level checkpoint-restart, will not work because of energy consumption, performance, recovery time, and cost issues.

5.4.2 *Outcome of the SWOT analysis - Main issues to be addressed*

Europe has some advantages and strengths in the field of energy efficiency and somewhat less so in that of resilience. Indeed, the big opportunity for Europe is to become established as the clear leader in both of these fields and to spearhead innovation. While research in energy efficiency is strong, Europe lacks strong research efforts in resilience.

Energy efficiency is a key factor in modern HPC systems and poses important challenges. We have identified a number of important technological challenges that affect virtually all aspects of the modern HPC tool chains:

- Facility and system design for cooling
- System and processor architecture
- System software and operating systems
- Algorithms and performance metrics

The importance of resiliency lies in the fundamental fact that faults limit the possibility to improve performance even when increasing the number of nodes and assuming perfect linear scalability. Looking ahead, the expected growth in HPC system scale poses significant challenges for system and application fault resilience.

Key technical challenges in the area of resiliency include:

- Resiliency and RAS technologies and methodologies (including redundancy) that are able to prevent, avoid, and detect faults and restore operating conditions with only limited impact on performance, energy consumption, I/O and hardware cost
- Resiliency technologies and methodologies that support fault-oblivious results and help obtain information about the accuracy of the results
- Increased hardware reliability when it seems to have attained its economically feasible improvement
- The entire ecosystem of RAS and resiliency technologies (hardware, software, algorithms, tools) needs to be adapted to extreme HPC scales
- Holistic resiliency and RAS technologies and methodologies based on coordination throughout the HPC system stack

5.4.3 *Research topics and expected results*

5.4.3.1 Cooling and Energy Reuse

Efficient system, facility cooling and energy re-use all aim to drastically reduce the energy consumption of systems and installations. Specific research tasks here include HPC

facility optimization, improved cooling of racks and components, and embedded chip-level liquid cooling (see the **M-ER-DC** milestones).

Optimized facility-level power provisioning and air cooling. As part of this task, air-cooled data centres need to be improved to the best possible cooling efficiency and power delivery efficiency (best practises). The general focus is on accurate measurements of efficiency (PUE), air-flow management, adjustments of flow and thermostats, use of free cooling, and optimization of power distribution. Management of air flow is accomplished through optimized placement of racks in terms of hot- and cold-aisle configuration. To improve air flow at the top and the edges of aisles, cold- and/or hot-air containment will be used. Currently, container-based solutions allow the distribution of standardized units with optimized air flow and excellent cooling efficiencies.

Rack-level cold - and warm-water cooling. The real key to reducing the energy consumption of a computing facility is liquid cooling because the heat removal capacity of water is roughly 4000x larger than that of air. One approach to improve the cooling efficiency is to locate the removal device closer to the heat-generating source. Typically, the need for IRC (it arises when computer loads are greater than 10 kW per server cabinet. Rack- and row-mounted, close-coupled cooling devices are more efficient than computer-room cooling devices, and are appropriate for legacy data centre retrofit upgrades as well as for new data centres. Lead users include general data centre and HPC users and business analytics.

Chip-level hot water cooling and energy reuse. Addressing the cooling problem will require significant progress in a wide array of scales. A key point is cooling at the chip level. A possible solution is to use microchannel heat sinks. The goal is to reduce the thermal resistance between the die and the fluid so that cooling water temperatures of 60° to 70°C can be used. This will in turn reduce the need for chillers and will allow the direct utilisation of the collected thermal energy. This development will further help in reducing the total cost of ownership for HPC installations.

Chip-embedded hot water cooling. Another key focus point for addressing the cooling problem is to dramatically increase the compute and communication efficiency to keep communication energies low enough, since communication is expected to dramatically affect the overall power consumption. Obviously, this will drive volumetric power densities up and make such implementations very challenging in terms of power delivery and cooling. Possible solutions may include smaller main boards with closer proximity between processors and main memory and wider and faster memory busses. In addition, advanced designs that include 3D chip stacks with main memory included i.e. with hybrid memory cubes could be considered. Of course, these chip stacks will be much more challenging to cool since the power density is higher and the chip internal thermal resistances are higher. A possible solution can be to employ two sided cooling solutions where cooling is provided through the two layer silicon (embedded cooler) and through a cooled silicon interposer.

5.4.3.2 Energy-efficient design of computer systems

The second major technological challenge is an energy-efficient design of computer systems and their components (see the **M-ER-ARCH** milestones).

New compute main boards with better CPU memory proximity. Analysis has shown that a large percentage of the energy is currently spent on the memory subsystem, in particular on communication and power lines that connect the CPU with the memory subsystem. In Exascale systems, the percentage of energy spent on the memory subsystem is projected to be as high as 30%. Combined with the power spent on interconnects, the total projection for power spent on moving data to and from the CPU reaches a whooping 60% [Kogge, 2008]. Thus, bringing memory closer to the CPU is a well justified solution.

Efficient CPU architectures (microservers). The microserver paradigm is an emerging proposal with several appealing characteristics, among them also energy efficiency. The main motivation is once more to reduce distances and increase simplicity to minimize energy waste.

Heterogeneous architectures and accelerators. Here the main goal is to study accelerators and special compute units, such as signal-processing co-processors, and their integration in a heterogeneous design.

Dynamic resource engagement and disengagement. Currently, technology allows dynamic frequency and voltage scaling (DVFS) of CPU cores. Taking a step beyond this, and in view of the many-core architectures, this approach can be generalized to the switching on and off of complete cores.

Furthermore, we need to generalize this idea in a holistic way to the full HPC system in which not only cores, but also memory units and interconnect links may be dynamically engaged and disengaged.

5.4.3.3 System Software and OS Optimization

There is a lot of room for improvements in designing system software (including system libraries, operating systems, and runtime environments) with the goal of minimizing energy consumption. In addition, designing tools for monitoring the power performance is deemed particularly important (see the **M-ER-T** milestones).

Software to monitor power consumption at high resolution. Measuring power consumption at high temporal and magnitude resolution will be crucial to understand the behaviour of fine-grained applications. High temporal resolution is needed to integrate accurate total energy consumption measurements. Such tools need to be available as standalone applications (i.e., measuring the total system power consumption) and also as libraries, easily callable at the user level. Code instrumentation needs to be lightweight and transparent. It is essential to add power-monitoring tools for memory and I/O devices as well as interconnects.

Add power-monitoring functions to traditional performance-monitoring tools. Power measurement tools need to be directly usable (e.g., embedded as plugins) for already established traditional performance-monitoring tools.

Software control of power consumption: Policy- or user-driven. Current DVFS techniques need to be significantly expanded and extended. That is, easy-to-use user interfaces need to be built, and automatic power consumption control policies need to be developed as part of OS and scheduling programs (e.g., LoadLeveler and others). Compilers need to be equipped with power-control options in a manner similar to optimization flags.

System simulators for predicting and assessing power consumption at the system and the data centre level. Finally, simulator tools will be particularly crucial in understanding the power needs of future designs and extreme scale-out platforms.

5.4.3.4 Energy-Efficient Algorithms & Energy-Aware Performance Metrics

Improvements in algorithms and new energy-aware performance metrics hold great promise. Combined with advances in hardware, systems and tools, they can have a positive multiplicative effect in reducing energy and power consumption (see the **M-ER-ALG** milestones).

Energy-Efficient Basic Algorithmic Motifs. The vast majority of HPC applications can be broken down into a small number of basic algorithmic motifs (see for example [Asanovic et al., 2006] and the Mantevo[4] project). Examples include dense linear algebra kernels, such as for solving linear systems and least-squares problems (i.e., LU and QR decompositions), sparse matrix-vector multiplications, Fast Fourier transforms, sorting, random number generation and others. We propose to investigate the development of energy-efficient implementations of these motifs.

Energy-Aware Performance Metrics for Applications and HPC/Data Centres. It is clear that the Flop/s metric that has traditionally been used to measure HPC performance is not adequate to understand energy-efficient applications. The community has proposed the Flop/s/WATT metric (see for example the Green500 list); however research has shown that even that metric will still promote power-hungry algorithms (see for example [Bekas 2010]). Energy-aware performance metrics are needed not only to rate computing systems, but also to understand the qualitative characteristics of algorithms and applications with respect to their power and energy requirements, and thus to help design better methods and software as well as more energy-efficient computing platforms. Post-PUE performance metrics reflect the need to provide a holistic way of understanding power consumption in data and HPC centres across the full stack (from processors, systems and racks to full HPC facilities). We need a deep and comprehensive understanding of the full impact of HPC at the facility scale and at the societal scale.

5.4.3.5 Resilience and RAS

The challenges posed by the resilience and RAS topics have been grouped into three main research areas (see the **M-ER-T** milestones).

Fault tolerance, recovery and reporting, which groups all research topics dealing with post-fault management, that is,

[4] https://software.sandia.gov/mantevo/

fault detection, reporting, recovery and restoration of normal system conditions. This area of research involves the entire stack from hardware to software, tools, I/O, algorithms, and maintenance.

At the hardware level, it is concerned with improving the fault-level detection, reporting and recovery within and beyond techniques using error-correction coding (ECC).

At the software level, it groups topics that aim to render applications and software more robust by building resilient approaches at the algorithm, coding, and runtime levels.

Current tools need to be enhanced to so that they are capable of coordinating with fault-tolerance mechanisms and of handling the effects of faults.

Efforts in the area of algorithm-based fault tolerance should deliver ways to monitor computation, detect faults, and recover the original state. Another development sees algorithms (and applications) that are able to generate final results despite an uncompleted computation due to some part of it being lost in faulty nodes.

Middleware, libraries and APIs are often written with performance rather than reliability in mind. There are opportunities for developing resilience at the library level (MPI above all) as well as building or extending checkpointing libraries.

Post-fault research involves not only designing automatic recovery but also finding improved architectures (at the systems and data centre level) to permit rapid manual restart, including designing high serviceability systems and methodologies.

The need to maintain a high level of availability and to limit the impact of faults on critical applications/jobs is interrelated to energy, performance and cost. For example, there is a trade-off in the widely used "checkpoint and restart" that on the one hand is effective in protecting applications/jobs from crashing, but on the other hand costs a lot in terms of time and resources consumed. Indeed, global checkpointing techniques in large systems are predicted to be so demanding in terms of energy and network utilization to be almost unpractical.

Proposed research topics include:

- Local and global checkpointing over large systems
- Application/runtime-managed checkpointing
- Fault-tolerant libraries (MPI)
- Fault detection, propagation and understanding
- Algorithm-based Fault tolerance
- Detecting and reporting hardware and software errors

Fault prediction and avoidance groups all research topics that study new ways of predicting and avoiding faults and involves a holistic approach with hardware- and software-related research to increase system reliability, develop predictive capabilities, and design redundant system architecture.

On the reliability side, research should focus on system architectures that improve the robustness of an HPC system. This may involve hardware design and manufacturing, but also software design and development. It is a vast topic that involves areas such as materials, high-quality production, component and board design, system hardware architectures, software design and architectures, parallel software programming environments, debugging, etc.

Predicting faults is the proactive approach to resiliency. One way to achieve this is through adequate monitoring and control. High-end systems have grown to such a scale that the mere determination of the state of the system has become difficult. Detection of errors is another key point, and there the problem of identifying silent errors is one of the most critical areas to develop in order to guarantee reliable computational results. To ensure pre-emptive or corrective action, reporting becomes of crucial importance. Here the challenges are the collection, management and analysis of monitoring data as well as the management of alarms through the system stack. This field of research and development is like to extend the frontiers in health monitoring and system control, and means addressing the questions from all angles, for example, silent error detection and corrective action, error reporting, component level hardware interfaces, sensor networks, RAS communication protocols, RAS system software, etc.

Moreover, it must not be forgotten that although the primary goal of a RAS system is to enhance the fault tolerance and resilience of the platform it serves, the scale of RAS systems required to support future HPC platforms presents a resilience problem of its own. New fault tolerance

methodologies must be developed to enable RAS systems to meet their own resilience needs.

Least but not last in this area is research related to redundancy techniques at different levels in a system. The challenge is to provide the reliability of an N-modular redundancy scheme at only a fraction of the current energy and hardware costs involved.

Example of research topics are:

- High reliability of system architecture
- RAS analysis and RAS systems
- Hardware and software quality
- Fault prediction, proactive actions, and replication

Accuracy of results is the third area of research directly related to resiliency. Here, the challenge is to ensure the reliability of the computational results with a precise confidence interval. Undetected errors (silent errors) may cause a loss of precision that is not reported to the operator. On many occasions, silent errors are detected only because of inconsistent results.

Solutions for this problem involve improving the system's error-detection capability and designing algorithms (and applications) capable of producing results that are not or less strongly affected by errors (fault-oblivious algorithms).

Another challenge related to the accuracy of result is data integrity. Given the nature of HPC usage, it is imperative that the answer generated be underpinned by the correct data to make informed decisions. Silent data corruption, SDC, poses a threat to computational science, with several studies documenting SDC or related problems on real systems. Therefore it is imperative that research be undertaken to characterize the impact of SDC on users and that methods to protect computations from SDC be developed.

[5] Proposed in Colella, Phillip, Defining software requirements for scientific computing, included in David Patterson's 2005 talk: http://www.lanl.gov/orgs/hpc/salishan/salishan2005/davidpatterson.pdf

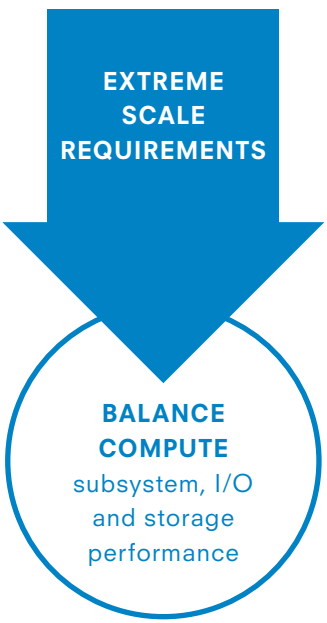
5.4.4 *Milestones*

Deadline	Milestones
2014	**M-ER-ARCH-1:** Improve hardware reliability by better redundancy techniques and system architectures (including cooling). Achieve at least the mean time to failure per core as that of current designs.
	M-ER-DC-1: Achieve PUE < 1.2 in any climate and implement energy re-use
	M-ER-T-1: Develop power/energy-measurement tools that are of similar ease of use and functionality as current performance-measurement tools
2015	**M-ER-ALG-1**[5]**:** Develop post Flop/s/WATT and PUE metrics. Energy/power-aware implementations/algorithms for the Berkeley Dwarfs
	M-ER-ARCH-2: A/N sensors for energy/power. New Designs: 3D, microservers, error collecting
	M-ER-T-2: RAS systems to scale to at least 5M threads
	M-ER-T-3: Design resilience-friendly MPI libraries so that applications do not get hung when a message is lost, or the core or network is down.
	M-ER-T-4: Develop energy/power-aware job dispatch tools. Develop energy/power-aware checkpointing tools
	M-ER-T-5: Create resiliency expressing/friendly programming environments
	M-ER-T-6: Maintain cost of checkpointing at constant levels wrt core count in terms of computation, energy, hardware resources and I/O.
	M-ER-T-7: Error reporting to scale to at least 1M threads for better situational awareness
2016	**M-ER-ALG-2:** Develop fully fault-tolerant Berkeley Dwarfs.
	M-ER-ALG-3: Develop fault-prediction algorithms, software and tools that enable a system to predict where/when a failure is likely to occur and to adapt accordingly to minimize impact or completely avoid the failure
	M-ER-DC-2: Achieve PUE 1.3 for hot climates, air-cooled 40 C, 60 C hot-water cooling, > 90% energy seasonal reuse
2017	**M-ER-DC-3:** Develop full hot-water-cooled infrastructure support (storage, UPS etc.)
	M-ER-DC-4: Develop hardware, software and data centre architectures with high reliability: Mean time to error to remain at least at the same levels as 2012, in spite of an at least 10x fold increase in core counts
2018	**M-ER-ARCH-3:** Compact systems (e.g., microservers) become the standard building block
	M-ER-ARCH-4: High-end systems to be developed by a co-design cycle
	M-ER-T-8: Power-efficient numerical libraries (e.g. LAPACK, FFTW)
2019	**M-ER-ALG-4:** Develop methods to determine the precision and reliability of results
	M-ER-ALG-5: Develop fault-oblivious algorithms that are able to resist, by construction, to high levels of undetected errors (wrt Berkeley Dwarfs)
	M-ER-ARCH-5: Chips with double-side cooling
	M-ER-T-8: Develop solutions for silent-error detection

5.5 BALANCE COMPUTE, I/O AND STORAGE PERFORMANCE

5.5.1 *Area*

Here the target is to ensure that in future systems an appropriate balance of performance is maintained between the capabilities of the compute elements of the systems, including the effects of changing applications and usage influencing the I/O system workloads, and the data storage systems such that overall system performance and capability are maintained or improved. This covers **all elements of the I/O stack — from middleware to data storage devices** — and associated networks.



5.5.2 *Data Storage and I/O: The need for a balanced system*

The key value created by any computation or analysis is data; this must therefore be secure, reliably retrievable, and non-volatile or be re-creatable. In some instances, it must be immutable (proven to be unchanged).

All current computer systems have evolved with the limited performance of the non-volatile data-storage system component that often operating orders of magnitude slower than the compute components do. With the ability to scale out the core counts and to deploy massive parallelism, the computer is expected to increase its aggregate performance by two to three orders of magnitude in the coming years, but the disk drive can only be expected to increase its performance by approx. 10% annually. Moreover, physically scaling the number of storage devices to match the computation needs is totally impractical. If not addressed, this will increase the imbalance of I/O and storage performance, causing unacceptable system bottlenecks. For checkpoint operations and for coping with the growth in big data, the capacity of non-volatile data storage capacity must also increase in a similar fashion as the system's (volatile) memory. However this is easier to achieve as the storage capacities continue to grow more rapidly than the performance.

In more classical HPC environments two particular, fundamental I/O patterns should be addressed when dealing with HPC user applications: autonomous checkpoint/restart and continuous saving of computed results. Many HPC applications, in fact, need to be stopped and started many times, as usually the time slots allotted to users are insufficient to complete a simulation. Users also often want to save snapshots of the fundamental quantities of their applications to enable a restart of the application in case of crashes. In large HPC systems, the time spent in checkpoint/restart activities may become critical. If such techniques cannot be avoided through improved system architectures, the checkpointing performance will be need to be improved by orders of magnitude beyond today's traditional storage system capabilities. Achieving this will require the use of new devices, storage architectures and will involve a more 'intelligent/predictive' use of the storage resources or alternative approaches to system resiliency.

Other critical patterns may have conflicting demands upon storage systems, creating either high-bandwidth streams or high rates of file creation, requiring file systems that are able to adapt and deliver guaranteed performance in multiple cases concurrently.

The increasing use of HPC with big data will also put heavy constraints to ensure the data feed to the compute clusters can be maintained: today's storage system hierarchy cannot achieve these goals and must incorporate new organisational

techniques and integrate higher-performance non-volatile storage media to achieve the necessary performance.

A key aspect of system performance is the 'real-world' performance. The raw capabilities of the systems will be wasted and left idle if the whole of the data transfer and storage process is used inefficiently. It is extremely difficult for an application programmer to understand how their method to access data can provide an efficient use of this infrastructure, and it is known that today many (or perhaps most) systems operate far below their optimal capabilities. These problems pervade all layers of the I/O stack interconnects and storage systems, often with orders of magnitude affecting performance, a situation that could possibly be addressed by changing the file-access methods.

The domain of HPC continues to expand. The huge growth of data created through both scientific and social media, in particular, continues to increase the volume of data-centric HPC applications. Examples, such as CERN or the SKA, are at the 'eye watering' end of the data scale. Increasingly, scientific, medical, industrial and social progress will come from the analysis of data created by sensors, social media, and scientific experiments. This is the subject of a separate WG, but is relevant here as the changes in workload seen in such systems can severely affect the balance of I/O and compute performance known from more traditional supercomputer environments. Future systems may have orders of magnitude increase in the number of users/jobs, have a tendency toward unstructured or semi-structured data organisation, or HPC jobs will tend to involve the analysis of pre-existing data, which will require careful staging, and often have extremely random access patterns.

Three main areas of this topic are covered by the scope of this WG, namely, I/O Middleware, Interconnects/Networks, and Data Storage, each having its own characteristics in terms of market, value, and impact.

5.5.3 *I/O Middleware*

The I/O middleware is an often overlooked, but essential element of any system, enabling applications to operate efficiently. It is currently dominated by open-source solutions, a situation that is expected to remain this way. An essential characteristic is the ability for application developers to use common methods and tools that are portable across systems. Features for scalability, reliability and performance optimization can be addressed very close to the applications at the middleware layer, regardless of backend storage architectures and file systems.

As highly scalable applications will most probably be based on a hybrid programming model combining MPI and threading, new I/O middleware and APIs must be able to coexist with MPI and provide the capability to overlap I/O communication with concurrent computation, e.g., by providing asynchronous I/O functionality. Europe has the ability to set standards in this area, and thus gain worldwide influence. However, it is crucial that the key knowledge and skills be maintained within Europe. Moreover, the application, operating systems, tools and libraries must be tightly linked together during the development as this will provide a critical performance/time-to-market advantage to any systems deployed in Europe and will improve the scientific use of Europe's supercomputers by a significant percentage.

A critical area which can be addressed here is that the multiplicity of applications, uses, delivery mechanisms etc., which result in very different workloads seen by a storage system, means that no system can work optimally at all times, particularly as in today's systems little or no information is provided to the storage or I/O systems for them to pre-emptively engage in optimisations to improve performance.

5.5.4 *Interconnects and Networks*

It is clear that the interconnect fabric is an important component of any parallel HPC machine; the same holds true for the network connecting the elements of a distributed system. For tightly coupled HPC systems, such as HPC clusters, the application performance they can deliver clearly depends on the interconnect fabric, which has to match the performance requirements of both the parallel application and the I/O subsystem. The former is quite well understood, as evidenced by the large body of performance analysis work of MPI applications (e.g., by the HPC Advisory Council). It is the I/O specific requirements for interconnect fabrics that this WG focusses on. For loosely coupled or distributed systems, the influence of the network tends to be larger, as these systems often use commodity networking technologies with lower absolute performance than HPC interconnects.

In contrast to, for instance, a MPI library, I/O middleware implementations can hide latency, combine several operations that work on small datasets, and predict future data accesses. This changes the spectrum of "message sizes" and renders raw interconnect latency less important, but emphasizes the need for high throughput. For large granular data accesses, highly efficient "streaming-mode" communication can greatly help performance, as can effective support for multicast. Applications assume that data read from or written to permanent I/O devices is safe from corruption, so mechanisms for fault detection and recovery are very important. This is true in particular for output data, which applications do not explicitly check once it has been written. Finally, integration of storage-class memory with interconnect components can enable highly efficient distributed data-caching schemes.

Highly efficient interconnects or networks and close integration of their functionalities with I/O middleware will immediately benefit operators and users of HPC systems – faster I/O performance means faster start-up of a parallel application (which requires the application image to be read by each node in addition to input datasets), and reduced time for storing intermediate and final results. It also lessens the performance impact of checkpointing. The latter is important as for very large configurations. The MTBF may drop below the runtime of many applications, leading to significant wastage for re-running complete applications.

A co-design scheme between I/O systems (middleware and storage systems) and the underlying interconnect/network would maximize the impact of the work, driving towards agreement on interconnect features and the performance characteristics needed by I/O and at the same time enhancing the I/O stacks (and maybe also the interfaces) to make best use of these interconnect features.

Europe is well positioned to research and develop solutions in this field, owing to its industry leaders in interconnect switches and NICs as well as in the design and production of storage systems.

In addition, Europe is strongly positioned in optical communications and photonics, which, we believe, will be essential to achieve significant increases in data rates and energy efficiency at the same time. This will require a shift in focus from telecommunications (WAN) to local interconnects.

5.5.5 *Data Storage*

The permanent storage of data is a key function of all systems. A number of characteristics of these storage systems are already reaching their limits when applied to the HPC field, and it is clear that a continued evolution of current techniques will not deliver the necessary improvements. In some areas, radically new approaches are required if the demands of the emerging extreme-scale systems are to be satisfied.

The most obvious characteristic is scalability: Here we must understand that it is not just sufficient that the system improves its ability to scale to huge capacity and performance, but concomitantly availability must be improved, especially in view of inherently worse reliability (due to higher component counts) and the dramatically more severe potential impact of failure. The ability to manage the vast systems must improve. Moreover, it is essential that we gain the ability to diagnose (and predict) failure or, more subtly, the impact of degradation in performance, something that is not yet possible today. Another core area of research is the variability in performance, particularly that inherent in the use of mechanical devices such as tape and disks, combined with changing workloads and wide workload variations in one installation, and the addition of solid-state non-volatile devices into storage.

Adaptive, pre-emptive, intelligent, controllable, scalable, dynamic, predictable, manageable are some keywords that could describe this area of research.

5.5.6 *Outcome of the SWOT analysis - Main issues to be addressed*

Key technical challenges in these areas include:

- The widening of the so-called Storage Gap, i.e., the gap between the performance of storage and compute devices.
- Inefficient performance utilisation of storage (applications unable to obtain the performance they require or expect)
- Low storage performance leas to low overall service levels
- The shared nature of storage (many customers demanding concurrent access)

· Data storage is extremely difficult to manage, optimise or diagnose
· Poor resilience of existing systems

"Soft Issues" that also need to be addressed:

· Storage has been overlooked, with insufficient investment in research and development.
· Tiny community of skilled technologists and researchers (especially in Europe), and hence a lack of skills in the technology supply industry and in the effective use of systems.
· Core storage technology (both solid-state and electro-mechanical) does not exist in Europe

5.5.7 *Research topics and expected results*

A number of quite detailed research areas have been identified and can be elaborated at a later stage. In this document, we will focus on the key topics and discuss some of the background and potential outcomes of such research.

To reach the required performance levels that have been identified to be commensurate with the processing systems of the future, it is clear that no single software improvement nor the use of any one technology can deliver the necessary boost in performance without exceeding physical, power or reliability limitations and that efficient use of storage to ensure best overall system utilisation through effective management tools and controls is essential.

5.5.7.1 I/O Interfaces

The I/O Interfaces of applications must be improved such that storage systems can deliver and receive data more efficiently as well as achieve better optimisation under wider operating and workload conditions. Particular areas of research include:

Guided I/O that provides information to the lower I/O layers and storage systems. Optimization of the low-level I/O layers often depends on a-priori knowledge about the application's needs or about how data will be accessed. Although in some cases, this information can be automatically extracted, frequently the application (or the programmer) has better knowledge of this kind of information. Finding the best method for optimisation is essential to achieving the improvements in system performance necessary in future systems because "brute force" performance increases in the hardware cannot be expected to deliver the dramatic changes that are necessary.

Transparent application pattern optimisation that is able to recognise data structures and workloads to initiate alternative data consolidation and optimisation solutions can also be applied.

Big data analytics tools and techniques suitable for huge and diverse unstructured and semi-structured datasets increase the overhead of data access and must be handled in specialised parallelised fashion. Also here is much more research required to achieve the improvements needed.

I/O system resiliency is a critical factor and must be solved at all layers, from the user interface to physical hardware. Research on resiliency and fault tolerance in the I/O subsystem is expected to encompass many areas, such as a deeper understanding of error-propagation paths, impact on MTBF (Mean Time Between Failures) by "tiering" newer classes of devices into the architecture, development of faster error detection/correction codes and the introduction of more application resiliency to I/O infrastructure faults. Application resiliency includes research in areas such as "concurrent" (performing checkpointing as computations are in progress) and "incremental" (checkpointing only the differences between the current and the preceding checkpoint) checkpointing approaches — which could drastically reduce the overhead of checkpointing and provide more flexibility on the MTBF of the infrastructure elements.

Metadata and Data in Parallel File Systems. Performance of both metadata and data (throughput) is clearly critical and will require new methods to ensure scalability to the requirements of Exascale systems and beyond. Changing the method of storing metadata such that it is embedded in the data structures may also enable dramatic improvements in scaling and performance. Use of object-based stores, and their allocation and performance optimisation can also relieve system-scaling limitations. In addition, the use of controls to enable Quality-of-Service provisioning of such services can bring about a huge change in the manageability and overall utilisation of HPC systems.

5.5.7.2 Information Lifecycle

The data explosion, a consequence of both increased computational power creating larger datasets and many more ensembles as well as the aggregation of data from sensors,

social media, and a vast array of sources, renders data management, control of overall system workflows, staging and de-staging, and efficient movement of vast volumes of data essential. Research on Information Lifecycle Management (ILM), Hierarchical Storage Management (HSM), scheduling and staging of data methods must continue.

5.5.7.3 Storage Hierarchy

To scale the performance of data storage to future needs, it is clear that the electro-mechanical devices, disk drives and tapes, although generally meeting the needs for growth in capacity cannot increase their performance at the rates needed, and the traditional method of using more of them in parallel cannot be sustained indefinitely on cost, power, space, scaling or other grounds.

The advent of solid-state media providing a much better cost vs. performance trade-off (but not capacity) and the expected arrival of other solid-state devices, such as PCM, ST-MRAM and others, can result in dramatic performance improvements if these solid-state media and devices are correctly integrated into the caching and data consolidation layers of HPC systems. Such technologies could be deployed in a number of ways between the processing elements and long-term storage media, and would require intelligent middleware to optimally exploit the various storage tiers, caching and data aggregation methods to best match the data-transfer and transactional performance to those of the access methods of lower-tier devices and of the interconnecting networks between devices.

5.5.7.4 Storage Services

A critically limiting factor in any system working with very large datasets is the data transfer between storage and compute elements. Often the compute elements will perform quite trivial tasks, performing data reduction through filtering and summarising data elements. It is feasible for some of those tasks to be performed much closer to storage, relieving the overhead of the data transfer and compute usage for more trivial tasks. Research into efficient ways to enable such services or run other more generic data-centric applications within embedded storage services is required. This also aligns with investing the potential to perform more complex metadata services in similar ways.

5.5.7.5 I/O System Simulation

Predicting the performance of any storage system is an extremely complex and difficult task. The interaction of a multiplicity of factors, from workload to hardware capability, may cause orders of magnitude variations in a given system or across similar implementations. Storage performance prediction and understanding the behaviour of massive scale storage subsystems, as will necessary for HPC and large-scale cloud infrastructures, through modelling and simulation methods still are an open problem today and in need of solutions.

Typical storage systems normally are over-provisioned so as not to be the bottleneck — which can work, but comes at a cost. The cost of implementing Tier-0 supercomputers with their associated storage is massive and continues to increase, as does the cost of identifying and correcting performance anomalies. The need to predict the performance of the I/O systems prior to deployment and subsequently the ability to diagnose (offline) variations in performance are critical to success.

The creation of an open, extensible modelling and simulation framework capable of predicting performance up to Exascale is required prior to any planned deployment of real systems.

5.5.7.6 Interconnects and Networks

The end-to-end performance of any parallel I/O stack depends to a large degree on the performance potential of the underlying interconnect as well as on optimizations in the I/O stack implementation that leverage this potential. R&D in the higher levels of the I/O stack will therefore require work in optimizing its use of a given interconnect, which is addressed in Section 5.2.

In contrast, our efforts here should focus on the potential for significant improvements in end-to-end performance (i.e., in the I/O speed) and reductions in energy use as offered by introducing new functionalities at the interconnect layer or by introducing new technologies for network implementations.

Areas of potential research would include:

Distributed caching and prefetching in networks, including integration of storage-class memory devices within networks. Leveraging high-level information on I/O patterns, flow-down of a-priori requirements from applications, and high-speed short-reach and affordable optical interconnects.

5.5.8 *Milestones*

Deadline	Milestones
2014	**M-BIO-1:** Guided I/O architectures defined; API for applications developed
2015	**M-BIO-2:** Resilient storage scalability techniques defined
	M-BIO-3: Architectures for extreme small-file performance defined
2016	**M-BIO-4:** I/O Quality-of-Service capability
	M-BIO-5: Common I/O system simulation framework established
2017	**M-BIO-6:** Exascale I/O simulation verified
	M-BIO-7: Tightly coupled SCM demo
	M-BIO-9: I/O middleware libraries adopted
2018	**M-BIO-8:** MD + QoS Exascale file I/O demo
2019	**M-BIO-10:** Hardware resilient Exascale storage demo
2020	**M-BIO-11:** Big Data + HPC integrated SW stack

5.6 BIG DATA AND HPC USAGE MODELS

5.6.1 *Area*

The use of HPC in the early history of computing was quite clear: Major users of HPC systems typically involved large national laboratories, often conducting classified research on weapons, universities with advanced computational research programs, and a few large companies. The problems solved were mostly driven by solving partial differential equations and numerical problems arising from computational materials science, and typically large dedicated supercomputers were used.

However, the tremendous advances in technology, modeling, mathematics and algorithms have dramatically extended the user base of HPC. A careful study of the evolution of the Top500 list as well as the research papers presented at key HPC conferences, such as SC and others, reveal an emerging highly diverse user base of HPC. Although national laboratories and universities continue to be key players and industries such as the chemical, automobile and aeronautics remain major clients, newcomers such as financial services, insurance, the pharmaceutical industry, and data and social analysis contribute strongly to the changing user landscape.

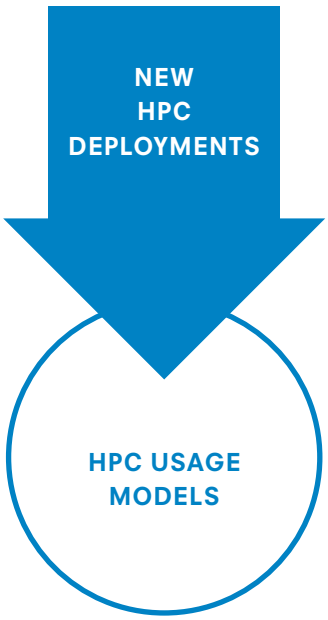
Thousands of highly educated scientists and engineers, with a strong background in HPC, nowadays, work for highly diverse companies, ranging from telecom behemoths to heap social network startups. This has triggered a major extension of how people use or envision to use HPC. Some ideas and uses can be potentially disrupting, causing major shifts in market trends, programming practices, and tools usage.

A major particular trend concerns Big Data, which is drastically influencing computing in general. In particular, we argue that HPC will need to adapt not only to the newly created needs but, more crucially, that HPC is becoming essential in realising the promise that Big Data holds. Traditionally, HPC tool chains were machine-centered, where expert users formulated their problem (together with any input/output) in such a way that they could get the most out of HPC resources. HPC was chiefly compute-intensive and not data-intensive: HPC programs typically had quite small inputs (i.e., a few Mbytes), whereas outputs could be larger, but

not significantly so (consider, for example, fluid dynamics simulations). However, dramatic increases in computational power meant that users could simulate ever more detailed models and perform compute-intensive tasks such as visualization. These trends created the need for very large storage systems attached to HPC resources. This was the first encounter of the traditional HPC world with the problem of large data. However, what is fundamentally different with the new world of Big Data we see emerging today is that the size of input is unprecedented. Big Data holds a tremendous wealth of information that potentially is extremely valuable. However, the value to become available, deep analysis of the ever expanding amount of data is required. The computational complexity of analyzing the data and its sheer size mean that in addition to compute-intensive problems we now also have data-intensive computing. HPC will therefore be crucial in unlocking the wealth contained in the data.

In addition, we are experiencing a major paradigm shift in how users consume the result of computations. Individuals, for example, social-network users using their smart phones to try to automatically find a friend in a large photo album, or a doctor consulting an intelligence system, such as Watson, to understand and improve a diagnosis, pose highly demanding computational problems to the underlying infrastructure without even realizing it. Computing is quickly becoming user-centric rather than machine-centric. Issues such as data integration, data security and privacy will have to be integrated with the underlying HPC computations. Thus, for many HPC users, the idea of having a user-friendly "front end" and no longer having to concern themselves with the back-end of computing is a strong incentive. In addition, for many new users or prospective users the complexity and the initial investment needed to succeed in the use of HPC are an inhibitor.

With the emergence of Cloud computing, virtualization and portals facilitate the migration of many users from the Grid computing paradigm to cloud platforms. These platforms offer a flexible execution environment and the capacity to standardize access to resources. New users can have fast and easy access to a well-managed and configured HPC resource. It is clear that an Exaflop computer cannot be efficient without a substantial investment in administration and user support. The openness of cloud platforms could be the opportunity to share HPC expertise and facilitate the fast adoption of HPC by companies that are new to high-end computing.



5.6.1.1 Definition

Here the investigation and categorization of the possible HPC usage and delivery models as these have recently evolved and are driven by major new applications such as analyzing Big Data are deemed particularly important. Understanding the new trends in these applications will help shape research agendas for the full HPC stack, such as tools, programming models, algorithms, user interfaces and H/W, because different usage and delivery models target different users with widely different backgrounds and competencies. There is a need for a focus on the complete tool chain. For example, we argue that although the use of HPC for Big Data may initially be interesting chiefly for users that work in data analysis, statistics and the new data-centric applications such as social networks, eventually most HPC users will benefit from research in this field (see above). In terms of new delivery models for HPC, such as the cloud, we stress that important advantages include the ability to deliver fully operational cluster instances (virtualized or not) on-demand, through a self-service API and Web UI (portal) interface. Moreover, the layered and modular design of the platform will allow seamless integration of highly varying workloads, such as Big Data, visualization, and multimedia processing.

5.6.1.2 Motivation and Opportunities

Changes in HPC usage models and delivery, the advent of the Petaflop and Big Data era and the immanent Exaflop era as well as the advent of desktop HPC performance clearly have a huge disrupting potential.

Understanding these trends has as clear goal the anticipation of change, and will increase the competitiveness of the HPC value chain in Europe. In addition, Big Data brings will generate new opportunities across various market segments — from healthcare to retail.

Our recommendations will affect a wide range of market segments as the disruption due to the emerging new HPC usage and delivery models can affect virtually all aspects of business and create new opportunities. SMEs in particular will acquire an important competitive advantage by gaining easy access to consumable HPC resources.

We are interested in understanding cross-usage trends and thus gain insights on how usages can be expanded by synergies and cross-breeding of practices.

Europe clearly has had a very strong advantage in basic research in algorithms and methods, as well as in software. The latter has a huge multiplicative potential. However, we see that understanding current HPC trends, coupled with deep software knowledge, can shape the design of HW to a large degree (see co-design efforts).

We believe that understanding HPC usage trends and new delivery models will help the exploitation of HPC and also help open up new usage cases, affecting virtually all aspects of modern society. In particular, European enterprises that target the new business of Big Data will have the opportunity to significantly benefit from these advances.

We also believe that a wide spectrum of lead actors will be involved, ranging from HPC and software vendors all the way to data/content suppliers and individual developers.

Major users include SMEs, individual software suppliers, healthcare, insurance and risk management companies, governments and finally also end-users that migrate from the Grid to the Cloud.

Understanding the new and emerging usage and delivery models for HPC will enable HPC providers to maintain their competitiveness and offer competitive solutions to an entire new user base. In addition, the introduction to new workloads will push innovation and is expected to trigger new research and the creation of new solutions. In particular, successfully addressing the Big Data challenge holds great promise to put European technology at the forefront of a very important major new trend.

5.6.2 *Outcome of the SWOT analysis - Main issues to be addressed*

Europe is in a unique position to excel in this area thanks to the potential of its internal market and the experience levels of current and potential users (and the recognition of the importance of data by such key users as CERN). Europe should exploit that knowledge to **create competitive solutions for Big Data** business applications, providing easier access to data, broadening the user base (e.g., through Cloud computing and SaaS), and responding to new and challenging technologies. On the other hand, there is the danger that Europe largely remains a user and not a leader in the Big Data revolution, as most of the hot innovations start elsewhere (mainly in the USA).

5.6.3 *Research topics and expected results*

5.6.3.1 HPC as an Instrument

This challenge considers the use of HPC resources as (dedicated) instruments. That is, by parting from the time/spatially shared usage model, HPC resources are increasingly used as key facilitators of large projects (consider, for example, the dedicated cluster used for data analysis at CERN), where unique and spontaneous usage is essential. Another prominent example refers to the use of HPC for on-line simulations and decision-support tools. Consider the use of HPC resources for the on-line modeling and prediction of the spreading of a wildfire or a hurricane. HPC can greatly extend the prediction horizon, help analyze potential consequences of certain decisions, and thus help the immediate response of the authorities (see **M-BDUM-7, M-BDUM-8**).

Understanding the Complete Data Flow. Dedicated HPC resources are typically embedded into a larger infrastructure, taking input data and producing output data. Thus, we need to understand the dataflow requirements of such uses, and what these mean in terms of I/O capabilities. It is key to understand and anticipate any important bottlenecks, such as in data acquisition, network bandwidth/latency, or visualization bandwidth. We anticipate that users of dedicated HPC resources will cover quite a large and diverse spectrum. For instance, obvious users such as CERN, that operate very long pipelines of experimental apparatuses, need to be supported. However, on the other hand, there could be SMEs that depend on real-time response times to deliver their products, such as on-line trading, multimedia providers on mobile apps, and the like. Finally, new adopters of HPC will also include users that can now afford HPC resources and see a potentially huge benefit in adopting HPC, such as a small architectural business or an architecture research group at a university that can use HPC to provide augmented-reality experiences to their customers (researchers).

Real-time and interactive use. Large-scale simulations (or data analysis) performed on dedicated HPC resources need to be interactive and data-driven. This means that users need to be able to demand on-line visualization without the data moving to other HW. Lead users of the technologies developed herein are anticipated to be the same as those in the preceding topic.

5.6.3.2 HPC for Big Data Workloads

A second key technological challenge is the design of HPC systems for large-scale data analysis. The tasks involved are quite different from traditional HPC applications, such as CFD and Molecular Dynamics, requiring a not so frequently used part of the instruction set such as integer arithmetic, and deep branching. Another important aspect concerns the pressing need for privacy and security.

Big Data workload characterization. Big data brings a different kind of workloads to HPC. While floating-point arithmetic remains important, integer operations and deep branching also become very important. We expect practitioners and researchers who need to handle large amounts of data to benefit from advances in this area.

Data migration and storage. It is clear that Big Data also poses great new challenges in storage and migration of huge volumes of information. Again, practitioners and researchers who need to deal with large amounts of data will benefit.

Dealing with highly varying/streaming data. A major important characteristic of Big Data is its dynamic and constantly changing nature. Several applications exist where the volume of the data is so large that storing all the data is simply impossible. This could apply to the complete spectrum of large data users.

Privacy & Security. HPC resources need to adopt high security standards as non-scientific, i.e., social and private data is handled. In addition, research in scalable algorithms for privacy is needed to maintain the privacy of data owners and creators.

5.6.3.3 Industrial Use of HPC as a Commodity

A very important technological challenge is the commoditization of HPC and its practices, which will enable a much larger user base that needs HPC but is does not necessarily have the traditionally required knowhow to make heavy use of HPC resources (see **M-BDUM-2, M-BDUM-4**).

Identifying HPC modules. The identification of HPC modules and the creation of complete and easy to deploy and use solutions are key to a wider adoption of HPC. Non-experts need to be able to quickly and reliably find HPC modules that can make a difference for their particular problem. General users are envisioned.

Productivity tools for non-traditional HPC users. Perhaps the single most important bottleneck in the wider adoption of HPC is the very steep learning curve in becoming proficient in HPC practices. That is, parallel programming is difficult to learn and, even more importantly, it is quite difficult to learn it well to achieve satisfactory performance. Thus tools that enhance productivity when using HPC systems are imperative — not only for the non-expert, but also for the expert user.

5.6.3.4 The Use of HPC in Cloud Environments

An emerging technological challenge concerns the use of HPC in Cloud environments. That is, the ability to access HPC resources in a Cloud fashion so that users can drastically reduce the cost of HW acquisition, maintenance and even SW development by adopting a "pay as you go" model (see **M-BDUM-1**).

5.6.3.5 Very Large Volume

"Big" means a very large volume, i.e., many (hundreds of) Terabytes is today's "large" end, and tens of petabytes are easily foreseeable for the near future (see **M-BDUM-3**).

Complete pipeline. A key research task concerns the analysis and determination of the complete pipeline that spans from data acquisition to consumption of analysis. Lead users include all users of Big Data applications; however the emphasis will mainly on the designers and builders of HPC solutions that deal with Big Data.

5.6.3.6 HPC Solutions for Distributed, Streaming Data and Noisy Data

Big Data is by nature distributed, which means that distributed algorithms are key and issues such as data migration are very important (see **M-BDUM-3, M-BDUM-6**).
Algorithms and H/W for distributed data. This effort aims at creating highly scalable algorithms particularly suited for Big Data applications on specially designed platforms. All users and developers will be beneficiaries.

Streaming and dynamically changing data. Big Data can be streaming or constantly changing (dynamic). Thus, analysis often needs to be able to run at real time, straining our current computational capabilities to the limit.

Dealing with noisy (uncertain) Big Data. Big Data is often noisy (uncertain). Data uncertainty quantification becomes imperative. It relies on computationally intensive statistical and machine-learning techniques.

5.6.3.7 New HPC Workloads

Big Data is creating a new kind of workloads for HPC. 5. For example, it is evident that relational databases alone are not adequate to represent complex relationships of high-dimensional sparse data. Thus, graphical databases are emerging (see RDFs) and are used to understand and represent context in data. However, although HPC has traditionally worked with large sparse graphs since the inception of the Finite-Element method in the 1950s, the characteristics of these new graphs differ completely from those of the graphs from discretized PDEs: A node, denoting an individual in a social network, may have 10 or 10,000 connections. Phenomena such as the small world phenomenon are evident. In addition, these graphs are very large (as are the Finite-Element graphs in solving PDEs).

Post-graph-partitioning techniques and implications for HW architectures. While classical techniques, such as graph partitioning, largely fail in partitioning graphs that originate from Big Data, massive HPC resources are still needed to analyze these graphs. Once more, the lead user base is expected to be very broad, with an emphasis on application developers (see **M-BDUM-5**).

New HW and cache architectures. Memory accesses in this space are much more chaotic, posing great difficulties to cache systems of modern processors. New cache architectures for Big Data applications are needed. Lead users here are system designers (and vendors) (see **M-BDUM-5**).

5.6.4
Milestones

Deadline	Milestones
2014	M-BDUM-1: Productivity tools on the Cloud
	M-BDUM--2: Data structures for big dynamic data. HPC graph DBs and HPC machine-learning algorithms on those DBs
2015	M-BDUM-3: HADOOP/MAP-REDUCE + PGAS + MPI
	M-BDUM-4: Concurrent computation and visualization on the same HW and accelerators
	M-BDUM-5: New instruction sets and active storage
2017	M-BDUM-6: Algorithms working on compressed data and on data in graph form
2018	M-BDUM-7: Compute-capable I/O, interconnect and memory
2020	M-BDUM-8: Problem-solving environments for large data

6.
COMPLETING THE VALUE CHAIN

As outlined at the beginning of the preceding chapter, besides identifying important research fields and priorities, the SRA also emphasizes complementary subjects and aspects that need to be focussed on to complete the value chain. Unless these areas are dealt with, the SRA will not achieve the impact expected. The areas will fill the gap between HPC systems per se and the needs of most users, particularly industrial end-users, who need to focus on their core business and may not have the resources to properly run, exploit, and manage their own HPC infrastructure: HPC-related services, ISV support, special attention to the HPC needs of SMEs, and education and training.

6.1 HPC SERVICES

Although the application profile in the industrial use of HPC does not differ greatly from scientific use if we focus on applied sciences such as engineering or life sciences, there is a significant difference in the way HPC is used.

In science, HPC often is a vertically integrated scientific discipline of its own, covering all aspects from the application to the hardware. In industry however, HPC separates into a process-integrated application layer and a service layer.

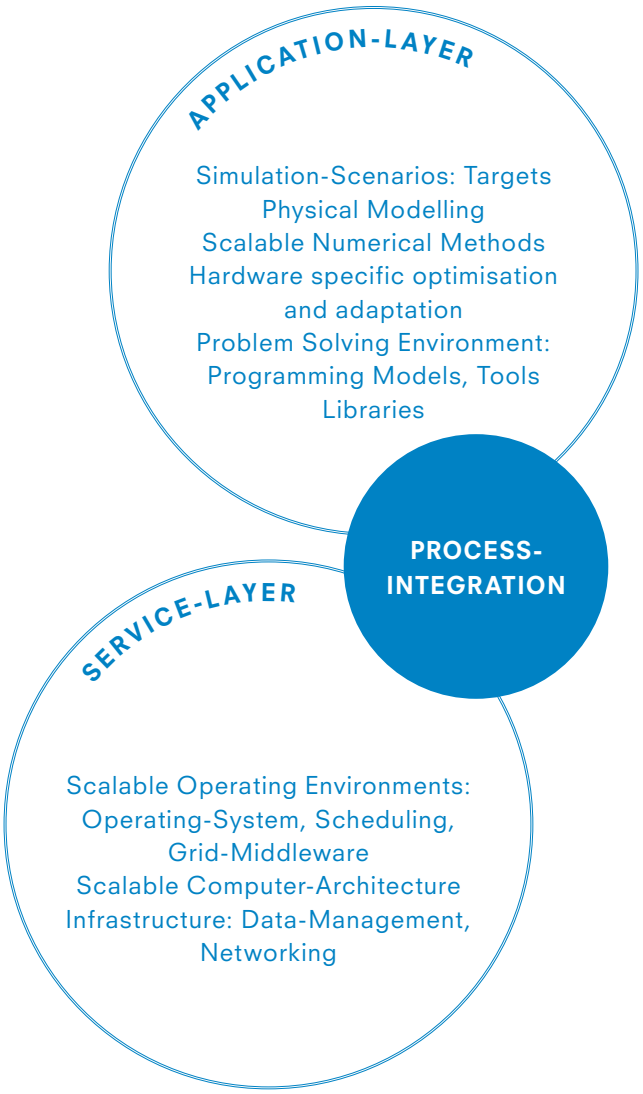


Figure 4
The layers of the HPC process in a typical industrial environment

The service layer in industrial use is subject to external provisioning, either from an internal IT organisation or from a service provider. Depending on the internal Key Performance Indicator (KPI) system of the particular enterprise, the service layer can be split between the pure provisioning of hardware (capital expenses, CAPEX) and the management of the entire infrastructure (operational expenses, OPEX), or a pure OPEX-based approach can be adopted by completely sourcing HPC as a service.

Concerning the application layer, there are significant differences depending on the industrial sector and organisation size. In areas in which the applications are either in-house or come from research institutions (as is typical, for example, in aerospace, oil & gas, or life sciences), application management and process integration are typically a core competence of the industrial end-user or sourced to specialised providers. However, in areas in which the application layer is governed by ISV code, the application management in smaller organisations is done mostly by the ISV and thus the ISV is the interface to the service (see Figure 5). This is true for most manufacturing industries and the automotive sector, especially for the suppliers. Only large corporations (e.g., the automotive OEMs) will have business relations with a service provider as well as with the ISVs.

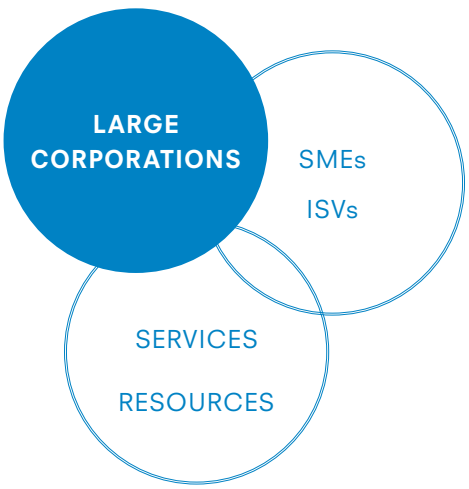


Figure 5
Engagement model for HPC service providers

Therefore professional HPC Services need to cover a variety of aspects to get accepted by industrial customers.

6.1.1 *HPC Infrastructure*

· Static Provisioning

For their everyday R&D business, most industrial enterprises have a quite static HPC base load that is part of the standard workflows and processes in R&D. To cover this load, the industrial customers need a service that is very closely integrated into the customer's security environment and dedicated for this specific customer to avoid compliance and confidentiality issues. For this service, industrial customers will accept a flat-rate-based service fee. Such a service has two major service items:

- Provisioning of dedicated HPC systems at the customer or in the data centre of the service provider.
- Management of the dedicated HPC systems owned by the customer or provided by a service provider.

· Dynamic Provisioning

Besides the relatively static business, there are always projects that come as an add-on. This can be the demand for additional capacity, or, in the case of special feasibility studies, the need to access a system with high-end capabilities. The latter may not be much in terms of volume, but critical for the competitiveness of the enterprise.

- Capacity Overflow
- Capability Overflow (Multi-Petascale systems)

Whereas handling Capacity Overflow is more a topic for commercial HPC Cloud providers, Capability Overflow could potentially be a task for national and international public HPC centres, like those of the PRACE organisation. However several legal issues would have to be solved to achieve a general availability on a commercial basis (and not only for the projects reviewed herein).

· Integration of all HPC components in a single environment

To render both the static and the dynamic components usable in customer workflows, these infrastructures have to be integrated into a seamless environment:

- Grid Middleware
- Distributed License Management
- Accounting and Security

6.1.2 *HPC Applications*

· Porting and Optimisation
  - Support for applications with source-code availability
  - Exploit new architectures
  - Ensure scalability
  - Validation
· Support for ISV codes such as CFD or CSM
· CAx workflow integration
· Integration into PDM, PLM
· Support for virtual reality (VR)
· Flexible and dynamic license provisioning

6.1.3 *IT-Management Processes*

Although HPC services require a certain degree of specialisation, these processes have to be established following the best-practices rules of the IT infrastructure library (ITIL). At least in most large corporations, HPC has reached a dimension that renders its management by the engineering or scientific departments ("shadow-IT") no longer acceptable or feasible; therefore it is usually under the governance of the CIO and the IT organisation. Specifically, the following SM processes are relevant for HPC:

· Incident Management
· Problem Management
· Change Management
· Request Management
· Escalation Management
· Business Continuity Management
Business continuity in commercial IT typically means high availability, but this term has no real meaning in HPC. In HPC, the goal is instead to achieve a high throughput and a low rate of aborted and rerun jobs. In industrial service contracts, the service provider typically has to guarantee turnaround times for jobs and that aborted jobs will be rerun under the provider's responsibility and outside the resources committed by the customer.
· Release Management
In the industrial design process, non-reproducible results are unacceptable, even if the underlying scientific or engineering problem is unstable by nature. Therefore software releases have to be pre-defined for the runtime of individual projects. This can create complicated interdependencies between otherwise independent projects and between the projects and the IT infrastructure.

To address these three service domains, the following research topics need be investigated (they have been covered in detail in the preceding chapter):

· **Virtualisation and Resilience**: Virtualisation today is mostly used as an instrument to partition resources into smaller units. HPC, however, is about federating resources to larger units. In any case, virtualisation is key for achieving more flexible and efficient HPC-services in the industrial sector:
  - Virtualisation allows a user-defined software-stack, and thus the undesired coupling of projects and infrastructures explained above under "release management" can be avoided.
  - Virtualisation allows the migration of virtual machines during runtime should a problem with a node occur. Therefore this will improve fault tolerance and avoids the need for additional capacities to rerun failed jobs.
· **Seamless Integration**: In a flexible scenario, in-house, cloud and external HPC resources as well as application software licenses will be used by the same users or even in the same workflow. This is only possible if these resources are integrated into a single virtual environment ("Simulation Cockpit").
· **Distributed Data Management**: Intelligent distributed data management needs to ensure that the data is where the jobs will run to minimize data transfer (caching etc.) and to use the distribution as an opportunity improving data safety.
· Cross-organisational workflows and security
· Scalability
· Fault tolerance of applications
· Real-time and embedded HPC

6.2 INDEPENDENT SOFTWARE VENDORS

ISVs are key players of an HPC ecosystem. Besides providing essential software application components and applications for different scientific and engineering areas, they often also act as expertise or service providers for industrial SMEs or larger enterprises or work closely with providers of services or computing resources (see the preceding chapter).

The following actions and approaches are recommended:

· Some form of co-design between hardware and software suppliers and ISVs (including, but not limited to, programming environments; this also means for instance I/O, memory subsystems and their software parts). For example, ISVs could work with HPC technology experts to provide testing environments that reflect the needs of their software systems.
· Easier access to new technology test-beds for earlier assessment. This represent an area of a joint interest with SMEs using HPC (see also the section on "SMEs as HPC Technology Providers" and "Education and Training")
· As mentioned in the section on "HPC Services", working with PRACE, which can provide access to prototypes on PRACE sites for experimentation as mentioned below. Moreover, this can help build experience on HPC systems of significant size with customised support and expertise from large computing sites that otherwise would not be feasible. Accessing large systems can have a cascading virtue and benefit, in the sense that even if ISVs and their customers do not always use the largest systems running at a given time, experimenting on them can help them gain experience and forecast the usages on systems that will be available more routinely 3, 5 or 10 years from now.
· Reflection on business models and licensing models for better SME access to scalable software on large HPC configurations, i.e., which mechanisms could lower the threshold for SME users to HPC usage, while preserving the interests of ISVs which often are SMEs themselves.
· Jointly define research and business development actions toward the delivery of ISV simulation solutions on HPC Cloud, in terms of ease of use, user friendliness; cost; security and confidentiality; availability, reliability and serviceability; ability to handle data management and post-processing as well as computation.

6.3 SMEs AS HPC TECHNOLOGY PROVIDERS

In HPC, SMEs have a large role to play — from outward-facing knowledge-based system integration, consultancy or software businesses to the hardware and software technology at the centre of the HPC supply chain.

That SMEs are able to succeed in HPC is evidenced by recent success stories such as Bright Computing, CAPS Enterprise and Allinea in both the domestic and export markets.

· In hardware, examples of SME success include storage and networking.
· In software, examples include development tools and libraries, cluster management, and industrial ISV applications such as CFD and FEM software.
· In services, there are a large number of - mostly national - cluster-integration and consultancy SMEs that add significant value for HPC end-users.

The objective is to increase the number of these success stories, and thereby increase the economic success of the EU.

6.3.1 *SME and Start-up Background*

More than 99% of all European businesses are SMEs, and they provide two out of three of jobs in the private sector: SMEs are said to be the backbone of Europe. In particular, SMEs are primarily responsible for the wealth and the economic growth of the EU economy, playing a key role in technology innovation and R&D.

The success of SMEs owes much to their simpler structure and receptivity. Surveys have revealed that enterprises which combine newness, smallness, and high R&D intensity are rare, but achieve significantly higher innovative sales that are new to the market than other innovative firms. As HPC is a field with very large R&D requirements, HPC should offer a proportionately larger opportunity and role than many other technology fields.

6.3.2
SME and Start-up Challenges

The challenges faced by SMEs and start-ups in HPC are identical to those facing SMEs in other technology fields. Shortcomings in European innovation and growth relative to the USA have been attributed to the larger capacity of the US economy to generate Young Innovative Companies (YICS). It is therefore vital to ensure such companies in the EU can be supported and nurtured. Financial constraints are clearly an important factor hampering innovation, and this has been described as a severe market failure as it prevents fair access to key inputs. Access to funding - either private investments or government/state funding - is therefore key to success.

Skills are also a major challenge for technology SMEs, especially recruiting and retaining highly skilled employees. Without the right individuals, SMEs cannot succeed. The size of the talent pool is critical: SMEs, with only "shallow" pockets, must compete with better endowed enterprises for the best talent.

Obtaining skills from across the EU is a second challenge. The talent pool is international in the EU, and in the 21st Century, remote working is easily managed. However, the widely differing employment practices render employing individuals in different member states expensive - and frequently require the founding of a subsidiary company in the remote employee's home state. This problem is not specific to HPC.

Entering the market as a supplier is a significant challenge to any SME as it requires access to the market "network". HPC is frequently considered as tightly knit community as substantial and strong collaboration and communication exist between many HPC centres and other classes of HPC users. The community also extends to technology providers, which can be a positive factor. Often, SMEs attempting to enter the market are expected to demonstrate a previous successful new product deployment, and exactly this requirement is difficult to achieve for most SMEs. Many HPC supply-side technologies require access to substantial infrastructure to develop or test fully a new product, and clearly such an infrastructure is often beyond the reach of SMEs.

Expanding into overseas markets is the ultimate challenge. The barriers to trade within the EU are fortunately low owing to the Single Market, but entering the rest of the world is a significant leap for most SMEs and requires careful execution. Nonetheless, the leading European HPC solutions should be able to compete on a global scale, and export opportunity should be a primary goal for an SME-focussed strategy.

6.3.3
Fostering and Nurturing Innovation

As the main limiting factor for SMEs is access to funding, policies for enabling the creation and sustainable development of SMEs are pursued by many governments in the EU and overseas.

One overseas example is the US Small Business Innovation Research (SBIR), which supports scientific excellence and technological innovation through investment of funds in critical priority areas to build a strong national economy. The program not only aims to meet Federal research and development needs, but also enables private-sector commercialization of research generated by Federal funding. Grants are given for each of the three phases of research: from prototyping and feasibility to commercialization. This program has been used to fund US HPC start-ups.

A policy that both enables creating and sustaining innovative enterprises in Europe is required to ensure that not only the initial phase is covered, but to enable enterprises to become self-sustaining. The ETP4HPC members believe the EC has a role in this, and this will help to maintain a level playing field.

6.3.4
Role of Universities and the Research Sector in SMEs

Funding is not the only opportunity for action: It is difficult to generalize on the scenarios that lead to the creation and success of an SME, but one agent that plays a key role is the university and research sector. Universities and other research institutions are able to bring some solutions to the issues of skills and market access identified above:

· Universities provide the highly skilled workforce and training required by both SMEs and larger enterprises (as

detailed in the "Education and Training" section of this document).
· Through the provisioning of internships, SMEs can contribute to employment and career improvement of students and researchers.
· Many technology SMEs in HPC as well as non-HPC markets are spin-offs of university research. Although it usually takes a long time to achieve the transfer from academic research to commercial product, university spin-offs have an above average survival rate.
· Universities are sources of research excellence in HPC — and can be used for consultancy and project collaboration with SMEs.
· Universities are consumers of HPC, providing a friendly market for SMEs, and have the potential to be a test-bed for innovation. Entering new markets and export opportunities is vital for creating wealth within the EU. Existing successful SMEs will be able to use Europe's HPC centres of excellence as reference customers.

6.3.5 *Proposed Actions on SMEs*

ETP4HPC proposes a level of support to HPC SMEs by establishing an HPC SME workgroup with the following objectives:

· Investigating and reporting on EU support initiatives for SMEs — to raise awareness among HPC SMEs. This could cover practical support opportunities, for example legal, IP, establishment overseas subsidiaries, etc.
· Showcasing successful European HPC start-ups to encourage the supply chain to include other European enterprises and to collaborate in overseas opportunities. This may include exhibition opportunities for the newly formed SMEs (e.g., at SC conferences).
· Help identify opportunities arising for European SMEs on the HPC provider side in emerging and growth areas of HPC technology.
· Provide information on EU funding opportunities and a central resource and advice for HPC-related projects
· Act as a facilitator for collaboration by bringing businesses and the education/research sectors together for specific proposals and enabling SMEs to have visibility on potential project groups.
· Encourage project proposals to include SMEs and micro-SMEs in responses to calls.
· Offer a forum for increasing the visibility of HPC technology-transfer opportunities from the university/research sector.
· Analyse how PRACE can facilitate HPC development and testing infrastructure for SME technology, with opportunities for both software and hardware providers. Systems must meet the security and performance requirements of SMEs.

6.4 EDUCATION AND TRAINING

Exascale HPC will be at the extreme end of the transition to parallelism in all forms of computing. In the EU today approximately 2 million people classify their job as "software development". Only a tiny fraction of them (less than 10,000) have direct skills in parallelism. Europe needs all programmers to have a basic understanding of parallelism and a much larger number of people skilled in its use.

The development of HPC education and training is a key success factor for Europe. There clearly is a high potential for creating economic value with HPC, but lack of skilled people could prevent Europe from achieving this potential.

A skilled workforce is the cornerstone of any successful skill-based society, and HPC is no exception. It is essential to remember however that HPC has its own special educational needs, and that even within this there are two distinct specific requirements, presented hereafter, that have to be met to create a vibrant ecosystem of technologies for HPC as well as to achieve wider economic and societal vitality. It is also important to keep in mind the "carry-over" effect to other industrial sectors. As HPC is used as enabling technology in many enterprises, well-trained HPC skills often migrate into the HPC-exploiting side of the business, so the benefit of investing into HPC skills carries over.

The chart below shows the growing gap between science and technology education at degree level[6] and the needs of the S&T labour market[7]. These underlying negative trends in core skills and lack of specialised education opportunities for the HPC technology and user communities must be addressed.

[6] http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=EDUC_THFLDS

[7] http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=TSC00025

Science and Technology in EU (27 countries)
Source: Eurostat

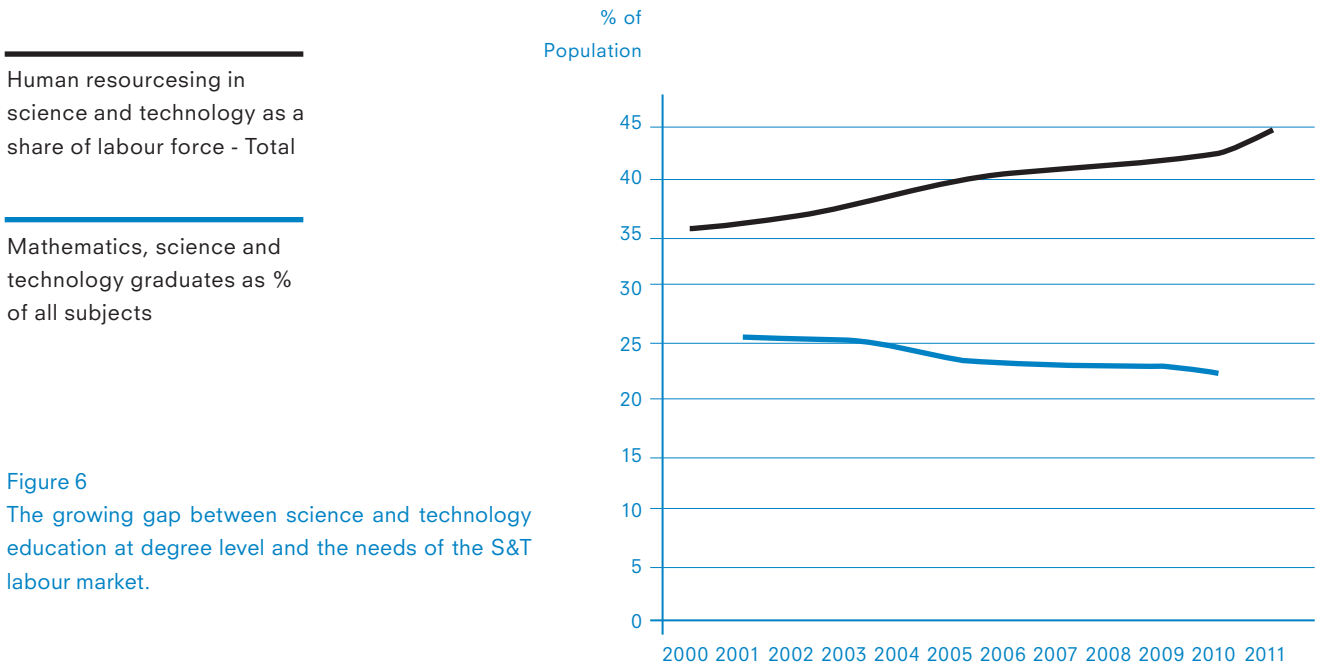


Figure 6
The growing gap between science and technology education at degree level and the needs of the S&T labour market.

6.4.1
HPC Technologists and Engineers

For HPC to be delivered by European technology suppliers, this ecosystem must include and be underpinned by a good academic infrastructure of education, training and research as well as the commercial development of such technology. To achieve this, Europe has to improve and increase the pool of educated and trained professionals with skills and knowledge of the core technologies, Computer Engineering, Computer Science as well as Electrical and Mechanical Engineering. We highlight here the term Engineering as a particular discipline focused on the delivery of computing, networking and storage software and hardware.

Even when looking only at the HPC technology and infrastructure provider side (there is also the application side), the spectrum of deep knowledge and expertise required to develop competitive HPC infrastructures is extremely broad and goes far beyond traditional computer science curricula:

· **Core technologies**: processors and micro-electronics, nano-technology, memory, non-volatile solid-state devices, photonics (especially short to medium reach opto-electronics), cooling technologies on component level, packaging, etc.

· **Systems Hardware** – high-speed data connections (electrical and optical), systems cooling and heat re-use, high-bandwidth and low-latency interconnects and networks, storage device integration, electronics, and mechanics, etc.

· **Systems Architecture**: processing nodes, interconnects, storage, system cooling, etc.

· **Systems Software**: operating systems, file systems, I/O middleware, libraries, etc.

· **Management and Tools**: debugging tools, system monitoring and management tools, scheduling and resource management, etc.

Today, many of the members of the ETP have extreme difficulties finding a suitably educated workforce, which limits growth and often delays developments because of the need for additional training over an extended period of time after staff has been hired. In other areas of the world, computer technology development is much stronger and universities provide much more relevant curricula.

6.4.2
HPC Application and Use

Equally important are the skills to apply this HPC technology to gain maximum advantage in the almost infinitely wide range of applications and the deployment of these codes to accelerate and improve progress in the respective industrial sectors.

To achieve this, a key challenge is the development of interdisciplinary dialogue. This does not mean that we must educate people with parallelisation competences, but that there must be a sufficient understanding of the use and value of applying HPC in the educational path of several fields to allow the trained people to interact effectively. Application-domain (such as engineering, biology, materials science, etc.), numerical-analysis and computer-science specialists should be encouraged to work together, and should receive the necessary education to enable interaction. Fundamentally, they must have a common language and understanding of HPC to allow them to communicate effectively.

In summary, efforts should be made to develop both Engineering and Application and domain-specific related integrated education and training structures all the way from the college level to PhD specialisation. This should be implemented at both a national and a European level to build a groundswell of skills and a connected academic, scientific and industrial community able to ensure the vitality of the industry and maximise the "return on investment" for the wider community.

ETP4HPC's members consider education and training as pivotal for a vital HPC value chain in Europe and will participate in this effort as outlined below. More specifically, they are ready to support the strengthening of HPC technology and engineering curricula for meeting their industry's needs and to collaborate with other stakeholders to encourage multi-disciplinary training. The presence of a strong HPC supply industry in Europe, together with a suitable HPC infrastructure and HPC applications, can foster a "virtuous circle", creating conditions that result in better training for people locally and are able to attract the most talented people to stay and create value in Europe.

There already are several initiatives and institutions promoting and conducting tutorials, lectures and a variety of workshops in the field of HPC such as 'HPC-Europa2' and 'CECAM' (http://www.hpc-europa.org/ and http://www.cecam.org/tutorials.html). The EIT ICT Labs offer a well-integrated professional education in key ICT areas (http://www.eitictlabs.eu/). However, HPC is not (yet) a focus area in their curricula, and therefore should be promoted as a new "innovation catalyst". PRACE has its own education programme and there are prominent examples of collaborations between industry and academic institutions in providing in-depth education in very specific computer-science disciplines important to HPC technology (e.g. http://www.scalus.eu/). A small number of Master programmes also exist in Europe (for example http://www.epcc.ed.ac.uk/msc).

Another good example is a collaboration between the Forschungszentrum Jülich and the University of Aachen (RWTH Aachen) called the German Research School for Simulation Sciences (http://www.grs-sim.de/), which offers a Masters course and PhD positions in Applied Supercomputing in Engineering, Materials Science, Biophysics etc.

6.4.3
Proposed Actions on Education and Training

Being industry-led, the ETP4HPC has a unique opportunity to help improve the situation with regard to the above-mentioned aspects of education and training by proposing the establishment of a workgroup with wide scope of interest across all areas of HPC technology and its usage (possibly through collaboration of ETP4HPC and PRACE-RI). The ETP4HPC will ensure that the skills and education requirements germane to the EMEA HPC industry are properly represented and taken into account in all actions of the working group. The objectives of that workgroup are:

Among its objectives are the consolidation of knowledge of the multitude of educational needs of the technology supplier industry, ISVs, service providers, industrial and scientific users of HPC, and of ways to map these to educational organisations and initiatives available through commercial, EU and national bodies. Also, it should map the existing facilities and analyse their distribution in terms of geography, scope, focus areas, funding models, etc.

· Work with partners in the scientific and education domain (e.g. PRACE) to Develop education and training directions for the vast range of skills required throughout

the entire stack of HPC technology, applications, deployment and industrial use. Review and support HPC Master programmes across Europe. Enterprises will benefit from this effort by being able to hire staff with the necessary skills straight from university.

· Under these directions, foster the establishment of specific HPC technology and application-focused education and training programmes, both short- and long-term.

· To co-operate closely with such educational organisations in all fields necessary for the success of HPC and its deployment

· Assess the feasibility of a Training & Evaluation Facility & Infrastructure focused on HPC technology and its use, with emphasis on industrial users and with the following objectives:

- Education and training of users (with focus on existing industrial users) in terms of the best ways to deploy the technology in their businesses.
- Again, together with PRACE, provide facilities for technology providers to investigate and optimise their new techniques and solutions to maximise the effectiveness of their products.
- Bring these together in an "infrastructure" of facilities across Europe, available to ALL (and targeted at industrial use).

· Enhance domain-specific application interaction to attract new industrial users, specifically SMEs, to the deployment of HPC. As shown in Figure 7, industrial SMEs have a very close relationship to ISVs, who in turn work with service providers for accessing HPC resources. Figure 7 shows the triangle of this relationship. Providing a platform and an environment in which this interaction between these entities can be developed for new usage scenarios, new applications, new SMEs starting to invest into HPC, or new ISVs starting their business can be crucial to the success of HPC in the industrial SME community. ISVs look for early access to HPC infrastructure, SMEs want to pay per use, Service Providers need to get an early understanding of emerging new needs and requirements.

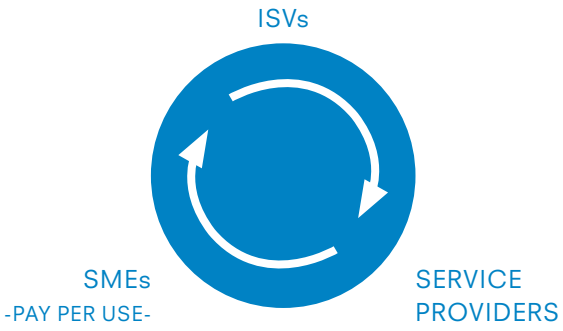


Figure 7
Relationship and benefits in a close collaboration environment

This should help increase the acceptance and knowledge of the benefits achievable through HPC-based simulation and modelling and should lower "the barrier of entrance" to newcomers.

Building on the EC initiatives in Objective FoF-ICT-2013.7.1 [EC-2], this approach is a next step. The platform could be provided as an extension to PRACE centres in Europe. A successful example is the collaboration model in place for many years now between HLRS and the automotive industry.

7.
LINKS WITH OTHER INITIATIVES

As outlined in Chapter 3, this SRA has been created based on a broad set of inputs, the experience of the ETP4HPC members, and an extensive analysis. The specific research topics have been chosen on the basis of their importance and impact on a competitive HPC technology ecosystem in Europe and the strengths available in Europe in terms of knowledge, experience, IP and skill assets.

Another aspect in the choice of research topics is the desire to create a meaningful interaction and collaboration with other ETPs in boundary technology areas to ensure that, on the one hand, any dependencies this SRA articulates are met and that, on the other hand, HPC-specific requirements are known and accepted by the other ETPs. Thus avoiding duplication of work, but also filling any gaps are an objective. Good examples of a close interaction are the areas of chip technology, photonics and hardware components, such as processors, memory, storage, etc. Moreover, also ETPs with a focus on ICT services, Cloud computing and Big Data are ideal collaborators. With this SRA, HPC is now well represented in the portfolio of ICT-focussed European research initiatives, closing a gap that has been the subject of much debate for many years.

Another important link will be with the KET (Key Enabling Technologies) initiative[8] the EC launched in 2012. Nanotechnology, micro- and nanoelectronics, including semiconductors, and photonics are among these KETs and are relevant to HPC. ETP4HPC can leverage a number of these technologies and possibly influence them and contribute to translate some of them into products and services (via the delivery of supercomputers and the development of their use). ETP4HPC could also complement these efforts regarding some limited and specific core technology development or extensions.

The following list shows the ETPs the ETP4HPC will collaborate with:

· ENIAC (http://www.eniac.eu): The ENIAC Joint Undertaking (JU) is a public-private partnership that coordinates European research on micro- and nano-electronics by organising calls for proposals and managing research projects. It consists of ENIAC Member/Associated States, the European Commission, and AENEAS (an association representing European R&D actors in this field), and also collaborates with the EUREKA Cluster for Application and Technology Research on NanoElectronics (CATRENE).
· ARTEMIS (http://www.artemis-ju.eu) The ARTEMIS European Technology Platform was established in June 2004. Its aim is to exploit the synergies among the key players in the Embedded Computing arena across the entire spectrum of industrial sectors. In 2008, The ARTEMIS Joint Undertaking (JU) was established to implement significant parts of the Strategic Research Agenda co-funded by industry, research organisations, participating Member States and the Commission's own ICT programme.
· EPoSS[9] (http://www.smart-systems-integration.org/public): EPoSS is the European Technology Platform on Smart Systems Integration and integrated Micro- and Nanosystems.
· Photonics21 (http://www.photonics21.org): Photonics21 is the European Technology Platform for photonics. Photonics21 aims to establish Europe as a leader in the development and deployment of Photonics in five industrial areas (Information and Communication, Lighting and Displays, Manufacturing, Life Sciences, and Security) as well as in Education and Training. Photonics 21 recently issued a Multiannual Strategic Roadmap[10] including optical data centre infrastructures for HPC centres.
· Other ETPs can be considered as sharing some objectives with ETP4HPC because they can either benefit from HPC technologies or cross-fertilisation can be relevant in certain areas (usages, services), for example, with NESSI (http://www.nessi-europe.com) active in Information and Communication Technologies. NESSI stands for the Networked European Software and Service Initiative.

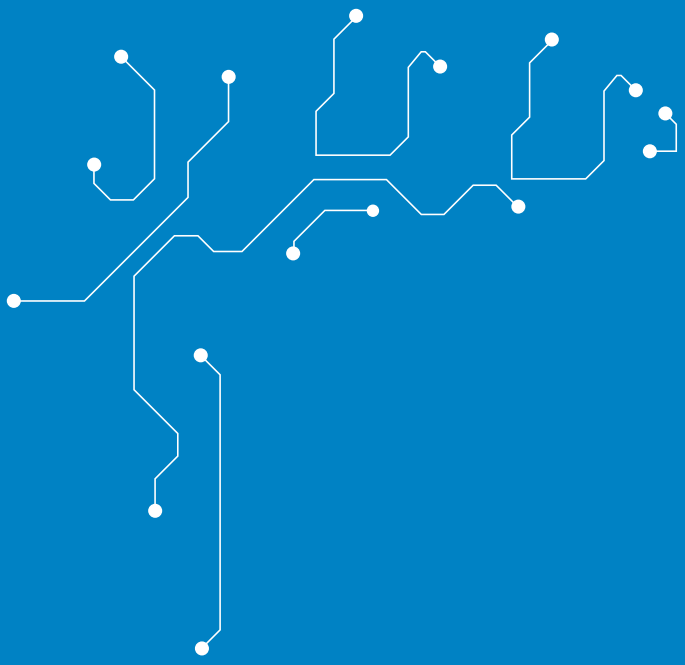
In addition, there are also several European initiatives active in the area of promoting High-Performance Computing with various goals, such as HiPEAC, Prospect, Teratec, and EESI2. ETP4HPC either has direct contacts into these organisations through its members or intends to interlock on a regular basis.

[8] http://ec.europa.eu/enterprise/sectors/ict/key_technologies/index_en.htm

[9] ENIAC, ARTEMIS, EPoSS will probably merge into a common JU, rendering coordination even easier.

[10] http://www.photonics21.org/download/PhotonicsRoadmap/PhotonicsMultiannualStrategicRoadmap_PublicConsultation.pdf

8.
MAKING IT HAPPEN



The preceding sections have presented research priorities and actions aimed at developing HPC in four dimensions (HPC stack elements, Extreme scale requirements, new HPC deployments, HPC usage expansion) identified as necessary to achieve a strong HPC technology ecosystem in Europe. There is a need to develop a research plan and undertake actions in a timely manner to maximise the impact and the expected return on investment.

There is clearly an opportunity for Europe to be a leader in HPC technology for Exascale. The disruptions that are needed to supply future systems open up the arena and as Europe has recognized skills in key domains – such as low-power embedded chip design, multi-criteria optimisation of hardware and software (co-design), managing the complexity arising from extreme parallelism, etc. – for tomorrow's solutions, Europe can achieve a much more dynamic position in the HPC technology landscape than what it has today.

There is also a favourable environment for the development of HPC. The awareness of the importance of HPC opens up competitive but highly attractive markets for European players. Delaying action will ruin this opportunity, as others will take gain ground in developing new technologies, and the HPC market expansion will be addressed by non-European players.

To be able to develop the right synergies (between technology, infrastructure and usage) and to maximise the impact on the ecosystem this action plan has to involve all HPC stakeholders. As explained in the document entitled 'Europe achieving leadership in HPC' issued by the ETP4HPC in November 2012 [ETP4HPC-2], a Public Private Partnership (PPP) could potentially be the right instrument to define the objective of the action plan and to align the actions of all the stakeholders.

Because of the urgency to maximise the impact, this potential PPP would need to be put in place in 2013. One objective of this PPP should be to implement the research priorities and actions presented in this document. The research plan has to cover the complete time line of Horizon 2020[11] if we want to deliver, as one of its outputs, the technologies for Exascale systems.

The research plan can be organized with a **first** phase of 3 years for the R&D actions targeting the milestones identified by the working groups in this document. This phase will deliver a first set of results that will be useful both for the extreme scale systems and for HPC solutions addressing a large market. This first phase will be helpful to refine the milestones of a **second** phase, from 2017 to 2020, that will aim at delivering technologies for Exascale solutions and for more efficient HPC systems.

For some of the outputs of the research programme, it could be important that towards the end of the first phase prototypes as hardware and/or software technology are built. These prototypes could be necessary to assess disruptive technologies and to build the ecosystem around these new solutions. If we want disruptive technologies to be integrated into Exascale solutions, the existence of prototypes around 2017 will be key to expose these technologies to a wide community so that (1) other components can be adapted to these technologies, (2) the base of applications that can benefit from them can be grown, and (3) users or other technology providers can provide feedback. The prototypes would be the basis for the development, in the second phase of the research plan, of a new generation of solutions based on these first results. The new generation should be competitive and able to come with the ecosystem necessary for an efficient use. It should be mature enough to be adopted by the market around 2020 without the need of a new prototype phase.

[11] Horizon 2020 is the European Commission's Framework Programme for Research and Innovation for 2014-2020, http://ec.europa.eu/research/horizon2020

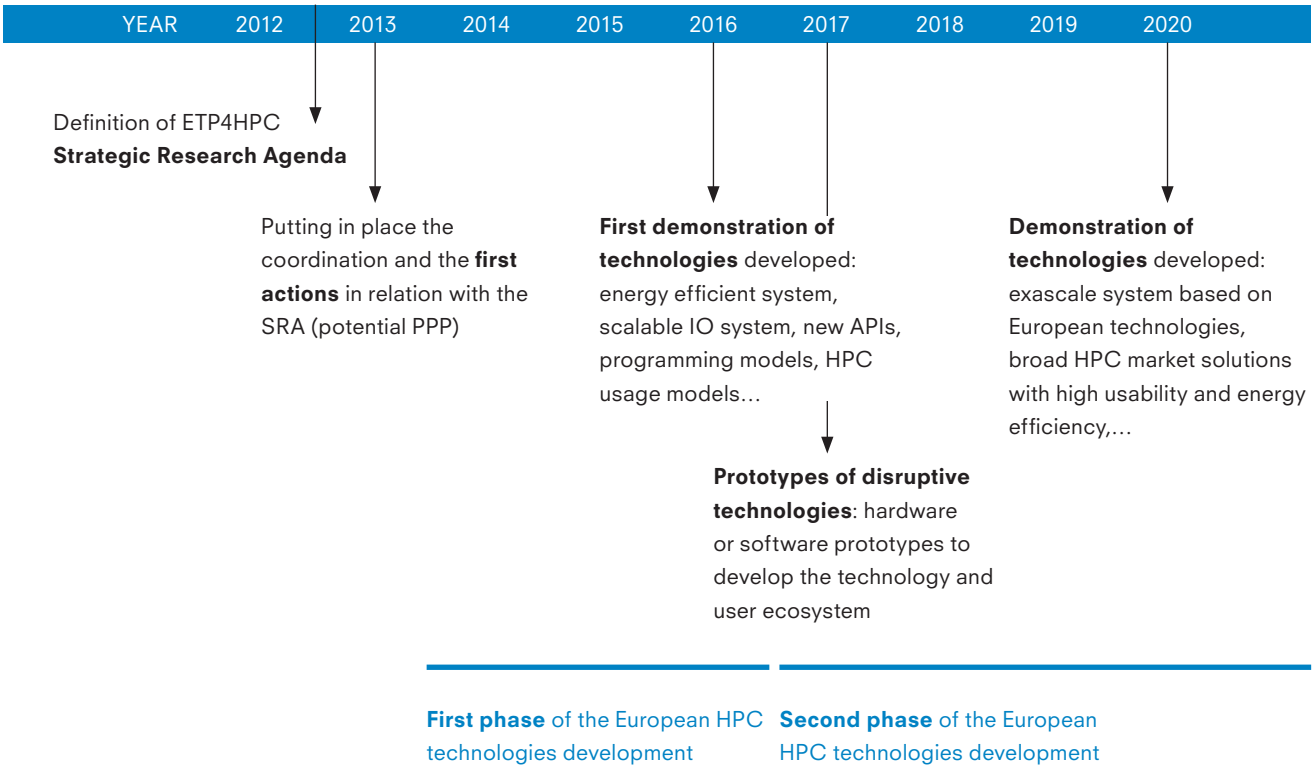


Figure 8
Implementing the Strategic Research Agenda - Timeline

To develop this plan and to achieve a leadership of Europe in key domains for tomorrow's HPC systems, the ETP4HPC has evaluated that a budget of 150M€ per year will be necessary for the research plan. This figure is for a research plan that will cover the topics presented in this document and that is focused on the domains where Europe can make a difference. The budget will cover the range of necessary technologies and will allow the development of a comprehensive set of results. Moreover, the plan will also provide sufficient resilience to prevent that, if an R&D programme fails to succeed completely, the entire effort is endangered and to ensure that the objective of putting Europe at the forefront of HPC technology will be safeguarded.

The budget does not include the cost of the prototypes as it is difficult and very uncertain and risky to decide already today which disruptive technologies could be successful and lead to the need for a prototype. In addition, the size and the cost of the prototype(s) could vary widely depending on the target outcome. Small investments will be sufficient if the objective is only to expose the technology to a wider community, and more expensive actions will be needed if the goal is to assess the ability to provide extreme-scale HPC solutions.

The proposed budget is also consistent with the conclusion of both the EESI and IDC studies. This total annual amount of 150M€ for the most part covers the cost of the additional skilled resources needed to carry out the research tasks presented in Chapter 5. A part of this budget will be invested by the companies and research institutions engaging in performing the research.

This program will have an impact on the economic value created in Europe thanks to HPC. The main expected results are:

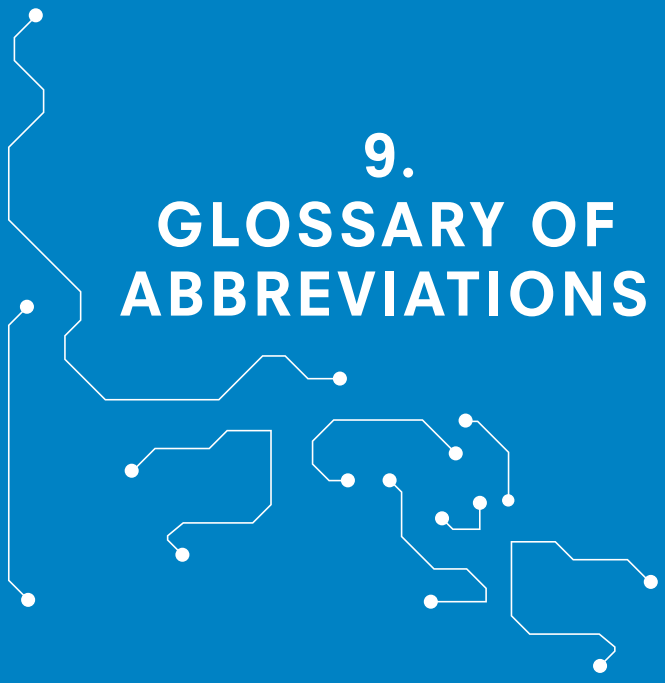
· An increased competitiveness of industrial HPC users by providing optimised products or services, by reducing the time-to-market, and by decreasing development costs. A strong HPC technology ecosystem will facilitate the creation of HPC added value by European HPC users.
· A competitive European HPC industry regaining market share in the worldwide market. The European players can successfully use the technology evolution to position

themselves as key actors in the supply chain of HPC systems and services. The economic value created in Europe in this field can increase in the range of some billion Euros.
· HPC technologies that have a positive impact on other IT markets and help European players be successful in these markets. HPC pushes the frontiers for hardware, software, tools, and methodologies. The progress made in HPC gives European industry a better position in other computer science domains.
· The creation of jobs in Europe by the industry using HPC and by the HPC supply industry. The increased competitiveness of the HPC users and the growth of the European HPC industry will lead to creation of high-skill jobs in both HPC engineering and HPC usage. The development of HPC services will also translate into additional staff requirements.

In a wider context, a successful implementation of the SRA will have an impact on both the European economy and its society. Developing HPC in Europe will help address the challenges identified in Horizon 2020 and will deliver practical benefits to the three pillars of this research-framework:

· Strengthening industrial leadership
· Help address societal challenges
· Strengthening the EU's position in science and research.

9. GLOSSARY OF ABBREVIATIONS



API	Application Programming Interface
ASIC	Application-Specific Integrated Circuit
CAGR	Compound Annual Growth Rate
CAPEX	Capital expenditure
CERN	European Organization for Nuclear Research
CFD	Computational fluid dynamics
CIO	Chief Information Officer
CMOS	Complementary Metal Oxide Semiconductor
CPU	Central Processing Unit
CSM	Computational Structural Mechanics
DOE	Department of Energy (USA)
DRAM	Dynamic Random Access Memory
DVFS	Dynamic frequency and Voltage Scaling
EC	European Commission
ECC	Error-Correcting Code
EESI	European Exascale Software Initiative
EOFS	European Open File System
ESA	European Space Agency
ETP	European Technology Platform
FDSOI	Fully-Depleted Silicon on Insulator
Flop/s	Floating-Point Operations per Second
GPU	Graphical Processing Unit
HLRS	Höchstleistungsrechenzentrum Stuttgart
HPC	High-Performance Computing
HSM	Hierarchical Storage Management
I/O	Input /Output
ICT	Information and Communication Technology
IDC	International Data Corporation
ILM	Information Lifecycle Management
ISV	Independent Software Vendor
ITIL	IT Infrastructure Library according to ISO 20001
KPI	Key Performance Indicator
MPI	Message-Passing Interface
MTBF	Mean Time between Failure, a reliability indicator
NSF	National Science Foundation
NVRAM	Non-volatile Random Access Memory
OEM	Original Equipment Manufacturer
OPEX	Operational expenditure
PCB	Printed Circuit Board
PCM	Phase Change Memory, a new memory technology
PDE	Partial Differential Equation
PGAS	Partitioned Global Address Space , a programming model
PPP	Public Private Partnership
PRACE	Partnership for Advanced Computing in Europe
PUE	Power Usage Effectiveness
QoS	Qualities of Service
R&D	Research and Development
RAS	Reliability, Availability and Serviceability
RDF	Resource Description Framework
RT	Run Time
SaaS	Software as a Service
SCM	Storage Class Memory
SDC	Silent Data Corruption
SKA	“Square Kilometre Array” program (http://www.astron.nl/r-d-laboratory/ska/ska)
SME	Small and Medium-Sized Enterprise
SoC	System on a Chip
SOI	Silicon on Insulator, a CMOS technology term
SRA	Strategic Research Agenda
ST-MRAM	Spin-Torque Magnetic Random Access Memory
SWOT	Strengths, Weaknesses, Opportunities and Threats – a recognized strategic analysis tool
UI	User Interface
WAN	Wide-Area Network

10. REFERENCES

Krste Asanovic, Ras Bodik, Bryan Christopher Catanzaro, Joseph James Gebis, Parry Husbands, Kurt Keutzer, David A. Patterson, William Lester Plishker, John Shalf, Samuel Webb Williams and Katherine A. Yelick, EECS Department, University of California, Berkeley, Technical Report No. UCB/EECS-2006-183, December 18, 2006, The Landscape of Parallel Computing Research: A View from Berkeley, http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.pdf

EBI, European Bioinformatics Institute, Annual Report 2010, http://www.ebi.ac.uk/information/brochures/pdf/Annual_report_2010_hi_res.pdf

EC-1, The European Community, High-Performance Computing: Europe's place in a Global Race, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels, 15 Feb 2012, COM(2012), http://ec.europa.eu/information_society/newsroom/cf/item-detail-dae.cfm?item_id=7826

EC-2, The European Commission, ICT Work Programme 2013, http://ec.europa.eu/research/participants/portal/download?docId=32767

EESI-1, European Exascale Software Initiative, Final Report, October 2011, http://www.eesi-project.eu/media/download_gallery/EESI_D5%206_FinalReport-R2.0.pdf

EESI-2, European Exascale Software Initiative, Final Conference – Presentation, October 2011, http://www.eesi-project.eu/media/BarcelonaConference/9-EESI_Final-Conference-recommendations.pdf

ETP4HPC-1, European HPC Technology Platform, ETP4HPC Vision Paper, June 2012, http://www.etp4hpc.eu/documents/Joint%20ETP%20Vision%20FV.pdf

ETP4HPC-2, European HPC Technology Platform, Europe achieving leadership in HPC, November 2012

HiPeac, High Performance and Embedded Architecture and Compilation, HiPEAC Roadmap - 2011, October 2011, http://www.hipeac.net/roadmap

HPC User Forum Meeting, April 2011, Event Material, http://www.hpcuserforum.com/presentations/houston_presentations/EarlMeetingslidesHPCUF.pptx

IDC-1, International Data Corporation, IDC HPC Market Update SC2012, presented at Supercomputing Conference 2012, November 2012, http://www.youtube.com/watch?v=vBgynpYZCoo

IDC-2, International Data Corporation, A Strategic Agenda for European Leadership in Supercomputing: HPC 2020, September 2010, http://www.euroris-net.eu/sites/www.euroris-net.eu/files/5847.pdf

IDC-3, International Data Corporation, Financing a Software Infrastructure for Highly Parallelised Codes, July 2011, http://www.hpcuserforum.com/EU/downloads/ParallelSoftwareFINALStudyWP27.17.2011.pdf

IDC, International Data Corporation, Applications (2011 survey), as quoted in http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/berthou.pdf

Planet HPC, Strategy for Research and Innovation through HPC, November 2011, http://www.planethpc.eu/images/stories/planethpc-strategy2.pdf

PRACE, Partnership for Advances Computing in Europe, PRACE Scientific Case 2012-2020, October 2012, http://www.prace-ri.eu/PRACE-The-Scientific-Case-for-HPC

11. OTHER REFERENCES AND LINKS

'A European strategy for Key Enabling Technologies – A bridge to growth and jobs', June 2012 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0341:FIN:EN:PDF

ARTEMIS http://www.artemis-ju.eu/

ENIAC http://www.eniac.eu:

EPoss http://www.smart-systems-integration.org/public

ETPs in ICT http://cordis.europa.eu/technology-platforms/ict_en.html

EU SMEs in 2012: Annual report on SME performance in the EU 2011/2, http://ec.europa.eu/enterprise/policies/sme

High-Level Expert Group on Key Enabling Technologies, http://ec.europa.eu/enterprise/sectors/ict/files/kets/hlg_report_final_en.pdf

HiPEAC http://www.hipeac.net/

http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/index_en.htm

http://ec.europa.eu/enterprise/sectors/ict/key_technologies/index_en.htm

NESSI http://www.nessi-europe.com/

Photonics21 http://www.photonics21.org/

Scalus, www.scalus.eu

12. ACKNOWLEDGMENTS

Chair of ETP4HPC
Jean-François Lavignon

SRA Editorial Board
Michael Malms
Jean-Philippe Nominé
Marcin Ostasz

Steering Board Members
Bernadette Andrietti
Sanzio Bassini
Patrick Blouet
Arndt Bode
Francois Bodin
Hugo Falter
Jean Gonnord
David Lecomber
Thomas Lippert
Guy Lonsdale
Malcolm Muggeridge
Andreas Pflieger
Ian Phillips
Francesc Subirada
Giampietro Tecchiolli

Working Group Chairs and Co-Chairs
Costas Bekas
François Bodin
Paul Carpenter
Alessandro Curioni
Hugo Falter
Jacques-Charles Lafoucrière
Guy Lonsdale
Giovanbattista Mattiussi
Malcolm Muggeridge
Jean-Pierre Panziera
Pascale Rossé-Laurent
Giampietro Tecchiolli

Independent Software Vendor Representatives
Laurent Anné
Philippe.Barabinot
Jean-Pierre Delsemme
Matt Dunbar
Rolf Fischer
Claude Gomez
Charles Hirsch
Koutaiba Kassem-Manthey
Dominique Lefebvre
Mark Loriot
Gino Perna
Antoine Petitet
Struan Robertson
Christian Saguez

End-User Representatives
Nicola Bienati
Serge Bogaerts
Ricard Borell
Norbert Bourneix
Ange Caruso
Thierry Chevalier
Alfred Geiger
Andy Jenkinson
Oscar Laborda Sanchez
Oriol Lehmkuhl
Felix Manchado Perdiguero
Gael Mathis
Alessandro Prandi
Hugues Prisker
Bernard Querleux
Michel Ravachol
Philippe Ricoux
Francisco Santisteve Puyuelo
Francesco Spada
Yves Tourbier
Mauro Varasi

Other Technical Contributors

Cédric Bastoul
Sergio Bernardi
Stéphane Bihan
Sven Breuner
Otto Buechner
Mark Bull
Carlo Cavazzoni
Albert Cohen
Guillaume Colin de Verdière
Toni Cortes
Tim Courtney
James Cownie
Marc Dollfus
Marc Duranton
Ton Engbersen
Thomas Fieseler
Hartmut Fischer
Uwe Fischer
Tony Ford
Alfred Geiger
Brent Gorda
Alan Gray
Wolfgang Guerich
Wilhelm Homberg
Hans-Christian Hoppe
Herbert Huber
Jean-Luc Leray
Kare Lochsen
Hervé Lozach
Luigi Brochard
Bruno Michel
Bernd Mohr
Raymond Namyst
Sai Narasimhamurthy
Michael Ott
Emre Ozer
Marco Paganoni
Mark Parsons
Oliver Pell
Mirko Rahn
Einar Rustad
Heiko Joerg Schick
Horst Schwichtenberg
Thomas Soddemann
Burkhard Steinmacher-Burow
John Taylor
Ralph Thesen
Thomas Weigold
Michele Weiland
Robert W. Wisniewski

ISV companies interviewed

Accelrys, DISTENE, EnginSoft, ESI Group, GNS-mbH, INTES, LMS Samtech, NUMECA, Scilab, SIMULIA

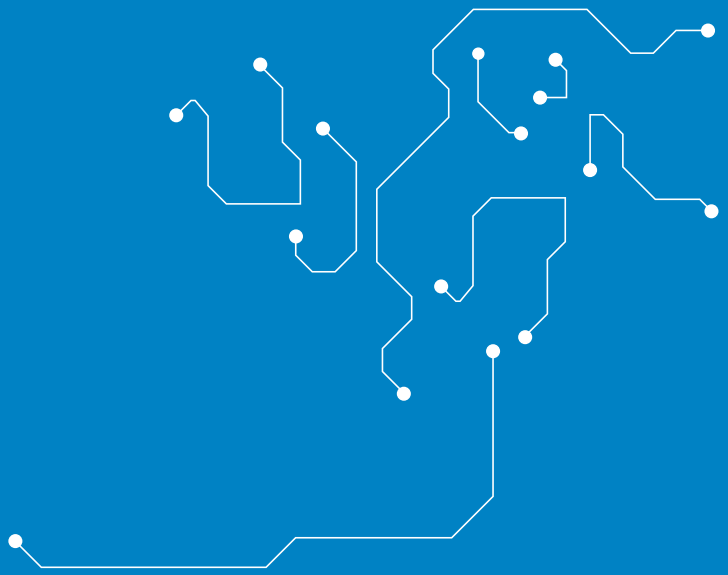
End-user organisations and companies interviewed

Airbus, Arcelor Mittal, CENAERO, Dassault Aviation, EBI, EDF, ENI, Epson Meteo, Finmeccanica, L'OREAL, Renault, SAFRAN, Termo Fluids, TOTAL, T-Systems

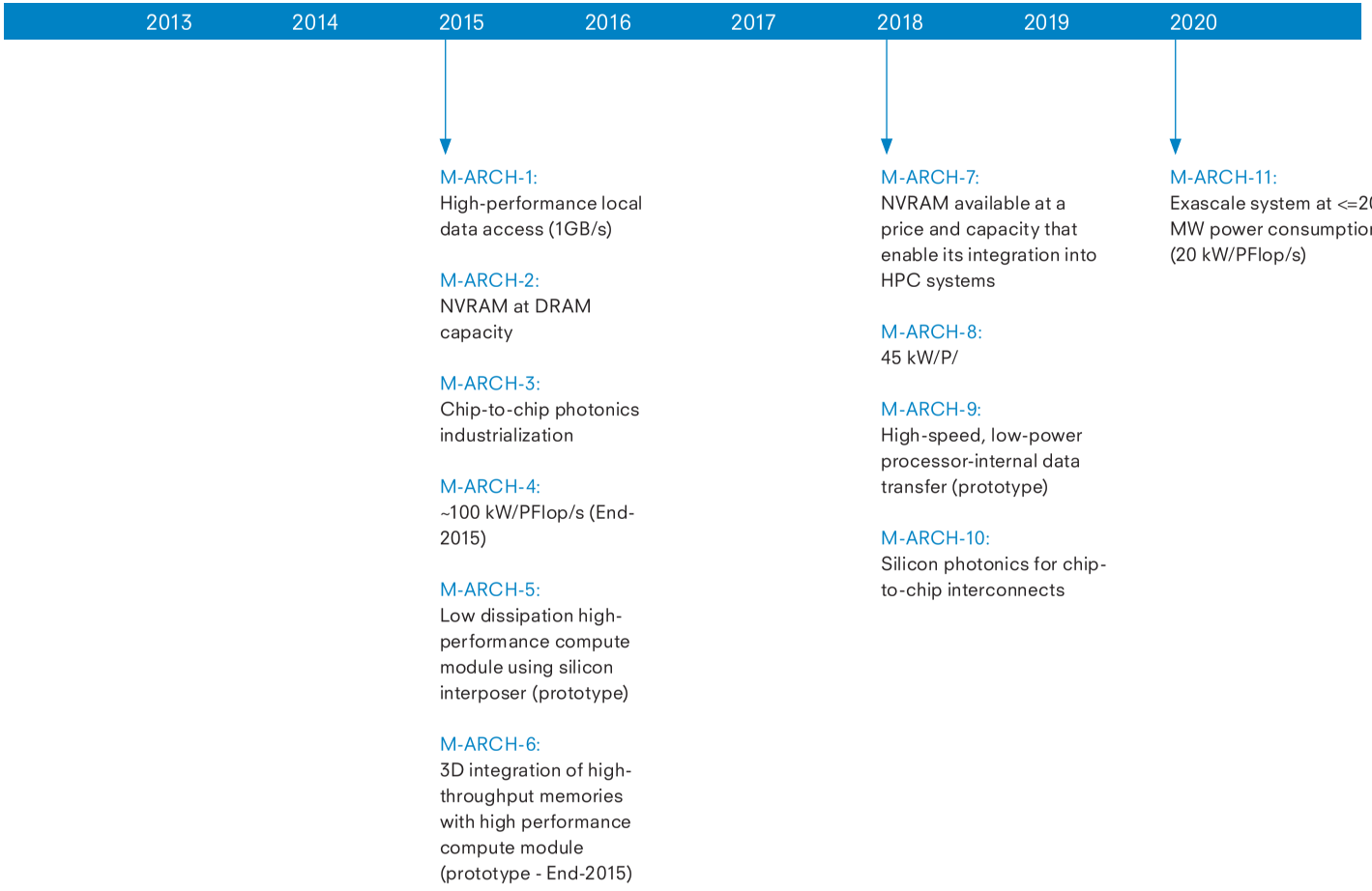
Document Editing

Charlotte Bolliger
Laura Bermúdez
Sonia Piou

13.
APPENDIX – RESEARCH TOPIC TIMELINES



Milestones for "HPC System Architecture and Components"



Milestones for System Software

2014

M-SYS-CL-1:
Model for on-the-fly analysis and control

M-SYS-IC-1:
Scalable interconnect management

M-SYS-OS-1:
Specification of an external standard for OS / Runtime and API for system-wide power management, explicit memory management

M-SYS-RM-1:
Power-aware resource management and job scheduling (RMJS)

2015

M-SYS-CL-2:
Model for system resource maintenance and management

M-SYS-IC-2:
OS-bypass and hardware interface integrity protection

M-SYS-IC-3: Interconnect adaptive and dynamic routing algorithm and congestion control

M-SYS-OS-2:
New lightweight base OS development with support of virtualization and HPC hypervisor

M-SYS-OS-3:
System security management

M-SYS-OS-4:
Prototype for system simulation environment

M-SYS-RM-2:
New Scalable scheduling enhancement with execution environment and data provisioning integration

2016

M-SYS-CL-3:
On-the-fly data analysis, data mining

M-SYS-CL-4:
I) Modular generation and deployment of OS instances and runtime environment and II) Flexible system image configuration and integration

M-SYS-OS-5:
Hybrid and heterogeneous memory and CPU OS support

2017

M-SYS-OS-6:
System with real-time capabilities

M-SYS-RM-3:
New multi-criteria adaptive algorithms: heterogeneity-/memory- and locality-aware

2018

M-SYS-CL-5:
Graphical supervision for Exascale

M-SYS-OS-7:
Scalable system simulation environment

M-SYS-RM-4:
Fault-tolerant MPI and checkpointing

2019

M-SYS-CL-6:
Load balance with tolerance to noise and temporary shortage of resources

M-SYS-IC-4:
Intra-/ inter-node network model implementation

M-SYS-OS-8:
Resilient OS with API

M-SYS-OS-9:
Hypervisor and deconstructed OS

M-SYS-RM-5:
Resilient framework

Milestones for "Programming Environment"

2014

M-PROG-API-1:
Develop benchmarks and mini-apps for new programming models/ languages

2015

M-PROG-API-2:
APIs and annotations for legacy codes[1]

M-PROG-API-3:
Advancements of MPI+X approaches (beyond current realisations)

M-PROG-DC-1:
Data race detection tools with user support for problem resolution

M-PROG-LIB-1:
Self-/auto-tuning libraries and components

M-PROG-PT-1:
Scalable trace collection and storage: sampling and folding

M-PROG-RT-1:
Runtime and compiler support for auto-tuning and self-adapting systems

M-PROG-RT-2:
Management and monitoring of runtime systems in dynamic environments

M-PROG-RT-3:
Runtime support for communication optimization: data-locality management, caching, and pre-fetching

2016

M-PEP-API-4:
APIs for auto-tuning performance or energy

M-PROG-LIB-2:
Components /library interoperability APIs

M-PROG-PT-2:
Performance tools using programming model abstractions

M-PROG-PT-3:
New metrics, analysis techniques and models

M-PROG-RT-4:
Enhanced interaction between runtime and OS or VM monitor (w.r.t. current practice)

2017

M-PROG-API-5:
Domain-specific languages (specific languages and potentially also development frameworks)

M-PROG-API-6:
Efficient and standard implementation of PGAS

M-PROG-DC-1:
Debugger tool performance and overheads (in CPU and memory) optimised to allow scaling of code debugging at Peta- and Exascale

M-PROG-DC-2:
Techniques for automated support for debugging (static, dynamic, hybrid) and anomaly detection as well as or checking the programming model assumptions

M-PROG-DC-3:
Co-design of debugging and programming APIs to allow debugging to be presented in the application developer's original code and also to support applications developed through high-level model descriptions

2018

M-PROG-API-7:
Non-conventional parallel programming approaches (i.e., not MPI, not OpenMP/pthread/ PGAS, but targeting asynchronous models, data flow, functional programming, model-based)

M-PROG-LIB-3:
Template-, skeleton- or component-based approaches and languages

M-PROG-RT-5:
Scalable scheduling of million-way multi-threading

2019

M-PROG-LIB-4:
New parallel algorithms parallelisation paradigms

M-PROG-PT-5:
Inspection of data locality at Exascale level

[1] *Note that the migration of a major application to a new programming paradigm generally takes 5 to 7 years.*

Milestones for "Energy and Resiliency"

2014

M-ER-ARCH-1:
Improve hardware reliability by better redundancy techniques and system architectures (including cooling). Achieve at least the mean time to failure per core as that of current designs

M-ER-DC-1:
Achieve PUE < 1.2 in any climate and implement energy re-use

M-ER-T-1:
Develop power/energy-measurement tools that are of similar ease of use and functionality as current performance-measurement tools

2015

M-ER-ALG-1:
Develop post Flop/s/WATT and PUE metrics. Energy/power-aware implementations / algorithms for the Berkeley Dwarfs[2]

M-ER-ARCH-2:
A/N sensors for energy/power. New Designs: 3D, microservers, error collecting

M-ER-T-2:
RAS systems to scale to at least 5M threads

M-ER-T-3:
Design resilience-friendly MPI libraries so that applications do not get hung when a message is lost, or the core or network is down

M-ER-T-4:
Develop energy/power-aware job dispatch tools. Develop energy/power-aware checkpointing tools

M-ER-T-5:
Create resiliency expressing/friendly programming environments

M-ER-T-6:
Maintain cost of checkpointing at constant levels wrt core count in terms of computation, energy, hardware resources and I/O

M-ER-T-7:
Error reporting to scale to at least 1M threads for better situational awareness

2016

M-ER-ALG-2:
Develop fully fault-tolerant Berkeley Dwarfs

M-ER-ALG-3:
Develop fault-prediction algorithms, software and tools that enable a system to predict where / when a failure is likely to occur and to adapt accordingly to minimize impact or completely avoid the failure

M-ER-DC-2:
Achieve PUE 1.3 for hot climates, air-cooled 40°C, 60°C hot-water cooling, > 90% energy seasonal reuse

2017

M-ER-DC-3:
Develop full hot-water-cooled infrastructure support (storage, UPS etc.)

M-ER-DC-4:
Develop hardware, software and data centre architectures with high reliability: Mean time to error to remain at least at the same levels as 2012, in spite of an at least 10x fold increase in core counts

2018

M-ER-ARCH-3:
Compact systems (e.g., microservers) become the standard building block

M-ER-ARCH-4:
High-end systems to be developed by a co-design cycle

M-ER-T-8:
Power-efficient numerical libraries (e.g. LAPACK, FFTW)

2019

M-ER-ALG-4:
Develop methods to determine the precision and reliability of results

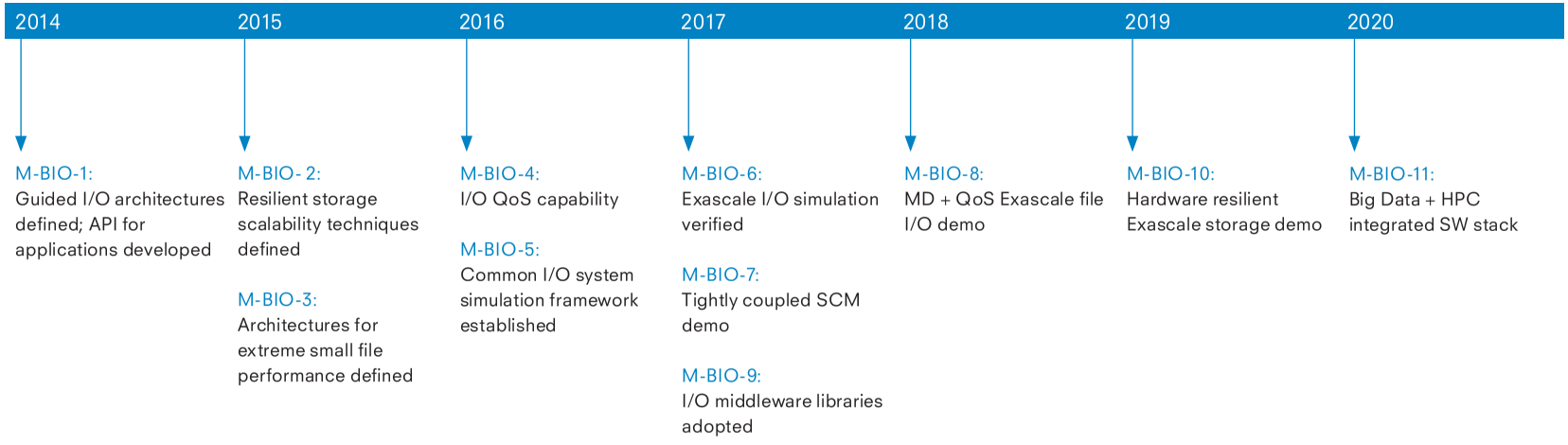
M-ER-ALG-5:
Develop fault-oblivious algorithms that are able to resist, by construction, to high levels of undetected errors (wrt Berkeley Dwarfs)

M-ER-ARCH-5:
Chips with double-side cooling

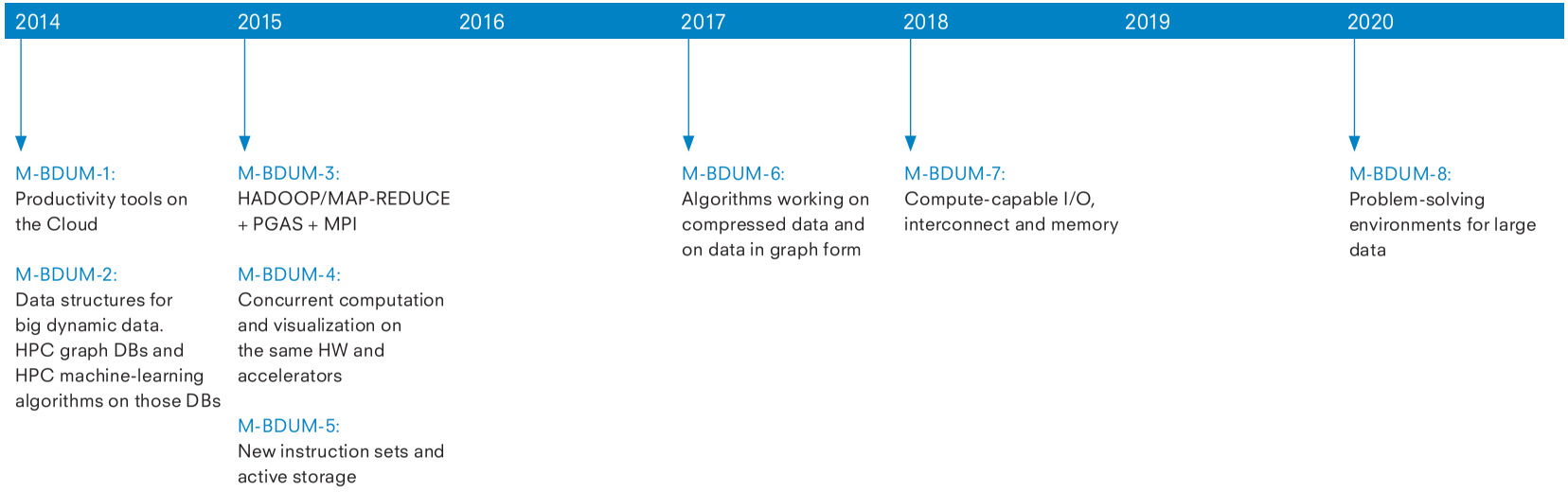
M-ER-T-8:
Develop solutions for silent-error detection

[2] *Proposed in Colella, Phillip, Defining software requirements for scientific computing, included in David Patterson's 2005 talk: http://www.lanl.gov/orgs/hpc/salishan/salishan2005/davidpatterson.pdf*

Milestones for "Balance Compute Subsystem, I/O and Storage Performance"



Milestones for "HPC usage, Big data, HPC in clouds"



IMPRINT

© ETP4HPC

Text:
ETP4HPC

Graphic design and layout:
www.pilarsola.com

Paper:
Bio top 250 g/m²
Bio top 120 g/m²

Printed and bound in Barcelona in May 2013:
www.grafiko.cat

Contact ETP4HPC:
office@etp4hpc.eu

www.etp4hpc.eu