Addressing the Challenge of Exascale
Towards Exascale roadmap implementation

European Exascale Software Initiative
EESI
EESI2 - 2014 Recommendations

Philippe RICOUX (TOTAL) , EESI2 Coordinator
EESI Motivations and Organization

Origins of EESI: **IESP** (Jack Dongarra)
www.exascale.org

**Build an international plan for developing the next generation open source software for scientific high-performance computing**

IESP meetings
- Nov 2008
- Nov 2009
- Apr 2010
- Oct 2011
- Sept 2012
- May 2013

www.eesi-project.eu
Build and consolidate a vision and roadmap at the European level, including applications, both from academia and industry to address the challenge of performing scientific computing on the new generation of super-computers, hundreds of Petaflop/PBytes in 2015 and Exaflop/ExaBytes in 2020.
What is at stake?

- Velocity Model ~ Snap Shot
- Seismic interpretation ~ Geological Study
- C – Interprétation structurale
- B – Seismic Depth Imaging

Human Brain Project

EXASCALE ON CRITICAL PATH
14 countries covering Europe including Russia, International Links, Participation of US, Israel, Lithuania, ...
European Strengths

- HPC intensively used in academy thanks to PRACE and in industry: Energy (TOTAL, EDF, Schlumberger), Aeronautic (AIRBUS), Finance, Defense&Security, ...
- Applications: astrophysics/cosmology, fusion research, materials sciences, life sciences, Engineering
- Hardware: Europe has all the competences to build an Exascale computer
- Applied math.: Europe has one of the best scientists in applied mathematics …
- Software: Europe has leading position in some software areas: system design (mobile, network, energy efficiency), simulation frameworks/coupling tools, meshing tools

European weaknesses

- Not clear how European ISV are preparing the coming of Exascale
- Simulation not as strongly supported as observations and experiments (ITER)

From a technical point of view:

- Lack of coordination in the development of HPC software, i.e. often few isolated centres
- Not enough participation in the definition of new standards for programming (MPI, OpenMP, C++, Fortran, …)
- Lack of critical mass on critical software domains: Fault tolerance, OS, compilers, File system, hybrid core design
**EESI recommendations – general statement**

- **Fund** strategic projects where Europe is **strong** and able to federate significant **critical mass**
- Insure the **coordination** of European efforts with the **rest of the world**, in particular:
  - Reinforce Multi-disciplinary HPC Centers providing **support** in terms of code development, porting and optimization as well as algorithm development.
  - **co-design centers/centers of excellence** should conciliates scientific multi-disciplinarity, international dimension,
- Europe need for a **sustainable, long term and coordinated** effort: A 2,5 to 3,5 billions euros total budget over 10 years, supported by EC, National European funding agencies, industry,
- Scientific Computing at Exascale, from a computing and data intensive point of view are **strategic** for maintaining and developing both **European Scientific Excellence and Industry Competitiveness**
- **International** collaboration is required
- Beside legacy codes, Europe should encourage the development of **Open Source solutions** to foster international collaborations and the emergence of international **de facto** standards, enabling commercial exploitation
Towards Exascale: Main issues to be addressed (EESI1)

- **At the level of simulation environment:**
  - **Unified Simulation Framework** and associated services: CAD, mesh generation, data setting tools, computational scheme editing aids, visualization, etc.
  - **Multi-physics simulation**: establishment of standard coupling interfaces and software tools, mixing legacy and new generation codes
  - Common (jointly developed) **mesh-generation tool**, automatic and adaptive meshing, highly parallel
  - Standardized efficient parallel IO and data management (sorting memory for fast access, allocating new memory as needed in smaller chunks, treat parts of memory that are rarely/never needed based on heuristic algorithms, …)

- **At the level of codes/applications:**
  - **New numerical methods**, algorithms, solvers/libraries, improved efficiency
  - Coupling between stochastic and deterministic methods: Numerical scheme involving Stochastic HPC computing for uncertainty and risk quantification
  - Meshless methods and **particle simulation**
  - Scalable program, strong and weak scalability, load balancing, fault-tolerance techniques, multi-level parallelism (issues identified with multi-core with reduced memory bandwidth per core, Collective communications, Efficient parallel IO)
  - Development of standards programming models (MPI, OpenMP, C++, Fortran, …) handling multi-level parallelism and heterogeneous architecture

- **Human resources, training (what level?)**

50% Computer Power for Data movement
Synchronization and Communication reducing algorithms
From EESI to EESI2

EESI roadmaps, vision and recommendations need to be monitored, updated, on a dynamical way

Key issues to be addressed are pointed out in EESI1 ... Now EESI2 must recommend R&D actions how to tackle them

- Extend, refine, and update Exascale cartography (directly in the dedicated WG for better analysis of each WG) and roadmaps from HPC community, on software, tools, methods, R&D and industrial applications, ...
  
  With a Gap Analysis.

  Including WG on disruptive technologies

Address “Cross Cutting issues”: Data management and exploration, Uncertainties - UQ&VQ, Power & Performance, Resilience, Disruptive technologies

- Investigation on funding scheme and opportunities, education, co-design centres, international coordination

- Operational Software maturity level methodology, evaluation
Contractual partners: TOTAL (coordinator), PRACE AISBL (acting for third parties LZR, GENCI, BSC, CINECA, EPCC, SARA…)
Contributing partners, involved in the management of EESI2 tasks but not associated to PRACE AISBL: INTEL, DLR, EDF, ANR, CERFACS, …
Supporting partners: more than 50 letters of Support
EESI2 proposal submitted in November to INFRA-2012-3.3: Coordination actions, conferences and studies supporting policy development, including international cooperation, for e-Infrastructures.
Requested funding: 1.5 M€ → 1.36 M€ accepted by EC
Duration: 30 months, Start 1st September 2012 - kick off 18th September 2012 (extended to 34 months)
2013/2014 EESI2 recommendations

Enabling technologies for Exaflop computing
- Hardware roadmap, links with vendors
- Numerical Analysis
- Scientific software engineering, software eco-system and programmability
- Disruptive technologies

Application Grand Challenges
- Industrial and Engineering Applications
- Weather, Climatology and Earth Sciences
- Fundamental Sciences
- Life science & Health
- Disruptive technologies

Cross cutting issues
- Data management and exploration
- Uncertainties - UQ&VQ
- Power & Performance
- Resilience
- Disruptive technologies

Contractual partners: TOTAL (Coordinator), PRACE AISBL
INFRA-2012-3.3: Coordination actions, conferences and studies supporting policy development, including international cooperation, for e-Infrastructures.
Funding: 1.36 M€
Duration: 30 months
Start: 1st September 2012

ORAP-Paris-20150402-EESI-PhR
2014 EESI2 recommendations

D7.2
2014 Update
Vision & Recommendations

All documents on EESI: WP, WG Reports, All Final Reports, Roadmap, Vision & Recommendations, .....

http://www.eesi-project.eu
Industries in Top500: Insight into HPC Performance

Projected Performance Development November 2014

TOTAL: 2 YEARS

TOTAL
2.3 PF 2013
6.9 PF 2015
Next:
30 PF 2017

Energy, Aeronautics, Automotive: Eur. Companies

Quite all Oil & Gas Companies
ENI: 5PF

ORAP-Paris-20150402-EESI-PhR
Exascale, a breakthrough

Technical Exascale issues:

- Concurrency/load balancing
- Data locality/Memory management
- Resilience/fault tolerance
- Energy efficiency

20 MW for an EXAFLOP also means a Petaflop in a box … and 20 KW

Huge impact for those, academic, industrial, including SMEs, that will be able to take advantage of “Exascale” technology, not just for a few heroes.

Software layer and applications need to exploit these new hardware trends that cannot be handled by existing software stack.

Holistic approach is needed:

From hardware, software to applications.

Best Coupling Between Architecture, Algorithms, Applications (AAA)

EXASCALE, a methodological, technological and scientific breakthrough
14 countries covering Europe including Russia, International Links, Participation of US, Israel, Lithuania, ...

Exascale, a technological breakthrough, imposes
To think different and differently

Exascale cannot be justified only if we are just planning to do the usual thing but bigger

Exascale needs breakthroughs in several domains (Algorithms, Algebra, Uncertainties, Couplers, Meshing ...)

ORAP-Paris-20150402-EESI-PhR 15
New thinking is required to develop new programming models, new algorithms, new tools, new data processing methods ... not only bigger than the present ones but far beyond the required innovation, disruptive improvements are needed.

Extreme computing and Extreme Data should be tackled simultaneously.

Assuring the efficiency and productivity of tools and applications at Exascale.
2014 EESI2 recommendations

PM & Tools

Ultra Scalable Algorithms

Efficient Exascale Applications

Data Centric Approaches
Data Centric Approaches pillar
A vision “Software for Data Centric Approaches to Extreme Computing”

• Software for Data Centric Approaches to Extreme Computing (2013 update)

  • Toward real time Extreme Data Processing and better science through I/O avoidance in High-Performance Computing systems
    • Identification of turbulent flow features from massively parallel Exaflops and Exabytes simulations

  • Towards flexible and efficient Exascale software couplers (direct or not, exchange of big data) (2013 update)

  • Declarative processing frameworks for big data analytics, extreme data fusion
Motivations: foster synergies between data-intensive science and extreme-scale computing, understanding the full cycle of data

Proposal:

- Data/compute research topics:
  - data transformational algorithms (extreme concurrency, asynchronous parallel data movement and access patterns, ...),
  - disruptive I/O and scalable multi-tiered data storage and parallel data management system
  - data analytics algorithms and techniques

- End-to-end simulation cycle:
  - simplify human-in-the-loop workflows,
  - libraries of scalable data analytics and data mining algorithms,
  - full-service data lifecycle management systems, in-situ vs. ex-situ data processing

- Task force: for the creation of a set of multi-disciplinary proxy applications (or mini applications) and scientific use cases

- Bridge the gap between scientists trained in both extreme-scale and data-intensive computing
Motivations

- Exascale challenges related to extreme data, energy and I/O constraints will make impossible for scientists to save a sufficient amount of raw simulation data to persistent storage for subsequent processing.
- Post-process centric approaches -> real time in situ/in transit processing (reduction, topological features extraction, quality assessment) for benefiting from data locality, save energy and perform visualisation and computational steering:
  - In situ = use of the primary compute resources
  - In transit = data processing is offloaded to a set of secondary resources using asynchronous data transfers

Proposal: Fund R&D programs in order to explore

- Real Time (In situ) data-related energy/performance trade-offs for end-to-end simulation workflows running at scale on current high-end computing systems
- Take into account future multi tiered storage architectures and asynchronous data transfers
- Design and implement new analysis techniques typically performed on large-scale scientific simulations: topological analysis, descriptive statistics, surrogate data model, filtering, compression, active learning, pattern/feature discovery, error analysis …
- Explore disruptive approaches like sub-linear algorithms addressing the fundamental mathematical problem of understanding global features of a data set using limited resources.
- The use of in situ data analysis for tracking and checking fault or error propagation into the simulation, associating a resilience aspect with the execution of workflows on parallel systems.
Industrial support for Data Management

In Situ Extreme Data Processing in massively parallel numerical simulations: A necessary approach for Exascale (Cf Paper from EESI)

Large Scale Reservoir Modelling
Finite Volume on 1 Billion Cells
Generation of 350 TB by Run Storage?? Post processing??

Even VISUALIZATION!

Turbulent combustion

Generation of 1PB Every 30 minutes
Towards flexible and efficient Exascale software couplers (2013 update)

Motivations

- Multiscale and multiphysics simulations as one of the main driver for Exascale systems will require flexible and efficient couplers
- Current coupling technologies include direct coupling or coupling via top-level interfaces
- Both approaches need to be supported and redesigned in order to address same challenges: use more parallelism, less communications and less memory

Proposal

- Assess pro and cons of direct coupling vs. top level API approached in the context of Exascale
- design standard coupling API in order to enable interoperability, ease the integration of new models and cross disciplinary exchanges and sharing of performance analysis
- Implement asynchronism into localization process, geometrical or mesh changes process or communications between coupled models.
- Support of heterogeneous resources and new programming models (component based model) at the level of the coupler and the coupled models
- Integrate coupling tools into an optimised pre/post processing stack including mesh connection between model, quick verifications of conformity, evaluation of physical quantities during computations, joint exploitation of massive results…

Europe very well positioned couplers with SALOME (EDF/CEA), OpenPALM (CERFACS) and mpCCI (Fraunhofer SCAI)
Declarative processing frameworks for big data analytics, extreme data fusion

Motivations:

- There is a convergence in data challenges between large scale instruments generating acquired data and (pre)Exascale systems delivering massive amount of synthetic data
- The Variety, Volume and Velocity, Value, Veracity of data imply a synergy and collaboration between different fields of science in order to extract full intelligence and knowledge from these data in close to real time.
- Current declarative language concepts need to be expanded in order to support data locality and volume and complexity while keeping abstraction and portability

Proposal: develop novel methods in order to address the challenges of:

- Declarative specification and automatic deployment of complex data analysis programs
- Develop declarative scalable data analysis libraries
- Continuous, workload-aware optimization and execution of data analysis programs over evolving data
- Adaptive, seamless deployment (automatically optimize for specific properties of the underlying hardware)
- Explore trading-off virtualization
- First results fast
- Better account of models
- Provide automatic mathematical based procedure to recognize in raw data features that will help in better interpreting and/or visualizing and/or interact with data

Europe is well positioned with the Stratosphere/Flink Eu funded project
Motivations

- The rise of multi Petascale and Exascale systems will allow turbulent simulations based on LES and DNS methods to address high fidelity complex problems in climate, combustion, astrophysics or fusion.
- These billion-cells simulations will produce such deluge of data which is difficult and inefficient to post process asynchronously later after by a single researcher.
- It becomes mandatory to propose tools able to extract and follow automatically pertinent flow structures.

Proposal: Develop a complete toolbox based on

- Massively parallel high-order low-pass and band-pass filters
- Conservative high-order interpolation kernels for the interpolation of fine grids to coarser grids
- Massively parallel Singular Value Decomposition algorithms for Dynamic Mode Decomposition of large sets of data
- Highly efficient linear solvers for symmetric matrices as those encountered in implicit filters
- Work in collaboration with scientific communities in order to make this toolbox an European standard used for addressing turbulent structures used in climate, combustion, traditional CFD, astrophysics, fusion, …

European researchers in France, Germany, Spain, UK or Sweden are worldwide leaders in LES and DNS methods.
2014 EESI2 recommendations

Ultra Scalable Algorithms pillar
extremely efficient scalable solvers for a wide range of applications

- Algorithms for Communication and Data-Movement Avoidance

- Parallel-in-Time: a fundamental step forward in Exascale Simulations (disruptive approach)

ORAP-Paris-20150402-EESI-PhR
Motivations:

- **Data Movement** as a key pitfall of Exascale systems for performance and energy reasons
- Tuning and communications overlap as a technique but will reach its limits with Exascale
- New techniques for communication and data movement avoidance far beyond the well-known communication hiding techniques, including the design of Hierarchical algorithms (based on compression techniques) have to be developed.

Proposal:

- Explore the design of a new generation of algorithms, for both dense and sparse linear algebra and beyond, that are able to drastically reduce communication costs, and even provably minimize it in some cases;
- Focus on operations that are at the intersection with the data mining community, as for example computing the low rank approximation of a very large matrix, an important tool in different areas;
- Focus on a comprehensive treatment of data movement within & between computing nodes;
- Enable the development of sustainable software that implements this new generation of communication avoiding numerical algorithms;
- Enable leadership of European researchers in selected areas;
- Enable the coordination and federation of multiple efforts to reach a critical mass.

Target: Dense and sparse linear algebra operations, mesh generation algorithms, parallel in time methods,…

**European researchers co-invented communication avoiding algorithms for dense and sparse linear algebra**
Parallel-in-Time: a fundamental step forward in Exascale Simulations

Motivations:
• As Exascale systems will consist on millions of computing cores, their efficient use will rely on extracting sufficient parallelism from applications to support parallel execution across millions of threads of execution.
• Today most of the applications contain time-stepping in some form and present-day codes make little or no use of parallelism in the time domain ➔ time stepping is currently treated as a serial process
• Potential application areas include: climate research, computational fluid dynamics, life sciences, materials science, nuclear engineering, etc

Proposal: Fund R&D programs in order to
• Establish multi-disciplinary consortia to work on the deployment of Parallel-in-Time methods and applications to new fields, combining the expertise of applied mathematicians, application scientists, computational scientists and HPC technology specialists, following a co-design approach.
• Develop a series of benchmarks and test cases in order to have a clear view of the advantages, disadvantages and quality of the different Parallel-in-Time methods. The test cases should be at the same time close to the real applications and simple enough to allow both the integration and test of Parallel-in-Time methods.
• Parallel-in-Time software should be encapsulated in reusable scalable libraries

Disruptive approaches delivering 10 to 100x performance speedups
European researchers are leading Parallel-in-Time developments
2014 EESI2 recommendations

Tools & Programming Models pillar

• High productivity programming models for Extreme Computing (2013 update)

• Holistic approach for extreme heterogeneity management of Exascale supercomputers

• Software Engineering Methods for High-Performance Computing (2013 update)

• Holistic approach to resilience for simulations and data analytics (2013 update)

• Verification Validation and Uncertainties Quantifications tools evolution for a for better exploitation of Exascale capacities (2013 update)
High productivity programming models (2013 update)

Motivations:
• Ease the development of applications
• Decoupling the logic of the applications from the actual resources
• Enabling rapid prototyping, performance analysis, simulation
• Enable exploitation of heterogeneous and parallel platform from runtimes: exploit locality or load balancing, minimize global communications, promote asynchronous communication, minimize energy conception, support resilience, …

Proposal:
• High level of abstraction
• New approaches in task-based asynchronous execution model, hiding the details of the HW platform
• Automatic exploitation of parallelism enabling scalability at large number of nodes
• Communication hiding programming in heterogeneous architectures
• Embedded Domain Specific Languages for high efficient implementations in heterogeneous environments (prototyping programming languages or scripting languages)
• Tools for automatized detection of data- and task-dependencies for multi-threaded task based programming
• Smart & efficient runtime (efficient resource management, exploiting data locality, automatic load balancing, energy aware)
• Tools for the development and analysis of the applications

Europe very well positioned in programming models, right moment to fund initiatives which can unify good ideas from Europe, and get a strong European programming model
Holistic approach for extreme heterogeneity management of Exascale supercomputers

Motivations:

- Exascale systems require a 50x gain in energy efficiency w.r.t. the today top500
- Disruptive technologies are needed to achieve at the same time a significant energy and performance gain
- The entire software stack is needed to evolve to hide the complexity of this wide design space supporting the programmer as well as the system administrator

Proposal: design and develop new efficient HW/SW APIs for the integrated management of heterogeneous systems, near-data technologies and energy-aware devices, to enable exascale-ready applications

- Design strategies for scalable and efficient heterogeneous-aware exascale applications;
- Scalable and efficient community scientific applications for exascale;
- System software to support efficient usage of exascale heterogeneous supercomputers in production.
- Coordinated design of the SW stack: programming models, run-time, OS and system support software
- Development of novel management and control HW/SW APIs which can be exploited by the programmer and by the runtime to deploy at Exascale the potential energy-efficiency of novel architectures in the different application domains.
Motivations:
• Need methods and tools: design, quality management, and reuse aspects of HPC software
• Because of the complexity and lifespan (x10 decades) of HPC applications

Proposal: productivity-enhancing methods and tools development and maintenance of Exascale software
• Develop predictive methods and tools to assist software re-design and co-design
• Support the transformation of codes into resilient software (see resilient recommendation)
• Carry out a survey of current software engineering practices and processes, define and validate optimal software new engineering processes for exascale applications
• Develop scalable debugging and performance analysis tools

Challenges at exascale:
• Performance
• Scalability
• Fault tolerance (hard and soft errors)
• Energy efficiency
• Change of algorithmic paradigms
• Pre- and post-processing
Holistic approach to resilience for simulations and data analytics (2013 update)

Motivations:
• dealing the resilience challenge,
• HW not able to detect and correct all sources of errors&failures
• Checkpoint technique is becoming a serious bottleneck
• Multiples sparse techniques need to be coordinated: multi-level checkpointing, fault tolerant protocols, alternative to checkpointing, failure prediction, resilient algorithms, detection of silent data corruptions

Proposal:
• resilience API global solution from hardware/OS, algorithms, software
  • understanding and modeling of fault propagation,
  • error detection,
  • failure prediction,
  • roll-back and roll-forward recovery.
• An international effort: take a complementary position to USA and Japan

Sources of faults
• fail stop errors (process, node, network, file system, OS, … failures)
• Silent Data Corruptions

State of the art:
• fault model - No clear leader
• Resilience for tasks based programming models - Europe is leading (OMPSS)
• Fault notification/management back plane - No clear leader (CIFTS is no longer funded)
• Resilient OS (make sure that OS confine faults) - No leader for OS
• Resilient Algorithms - Avoid overlap with USA and Japan, focus on complementarities (may need to establish a forum/working group)
Verification Validation and Uncertainty Quantification tools (2013 update)

Motivations: enhance the scientific recognition and societal acceptance of Exascale simulation
- Defining the domain of predictability of a Exascale simulation tool
- Enabling, improving decision making of the Exascale simulation process

Proposal: unified framework for usability VVUQ and Optimization on Exascale Systems
- Ultrascalable tools for VVUQ and optimisation
- Accessibility of the software: an unified European-wide package for VVUQ
- Methodological progresses (model errors in the validation process, surrogate models and reduced basis models)

State of the art:
Strong European expertise: OpenTurns (AIRBUS, EdF, CEA …), URANIE (CEA) platforms
US involvement: DAKOTA (Sandia Lab/DOE), PSUADE (Lawrence Livermore Lab/DOE)

Sources of uncertainties:
- Lack of knowledge on a physical parameter (epistemic uncertainty),
- Parameter with a random nature (aleatory uncertainty),
- Uncertainty related to the model (model error),
- Uncertainty related to the numerical errors (numerical errors)
Multiscale/multiphysics + big data → UUVQ
- Multiple sources of uncertainties due to
  - lack of knowledge on a physical parameter (epistemic uncertainty)
  - parameter with a random nature (aleatory uncertainty)
  - uncertainty related to the model (model error, too simplified model)
  - uncertainty related to the numerical errors (numerical errors of the model, to the input and output data, …).

Understanding uncertainties essential for acceptance of numerical simulation for decision making
- Strong impact in industry (automotive, oil & gas, aeronautics, nuclear,…) and academia (climate)
- EU well positionned in UUVQ : Uranie and OpenTurns

→ Toward an unified UUVQ env. for Exascale
- Dev of ultrascalable UUVQ tools (scheduling, optimisation, …)
- Embeeding UQ loops at the lowest levels of the simulation code
- Use of surrogate models and reduced basis models
Coupler: a major software component

- Multiphysics simulations / legacy and new codes
- Crucial for industry (aeronautics, automotive, …) and academia (climate, astrophysics, life sciences, …)
- Coupling of 100k cores applications is a good driver for Exascale

Europe owns multiple coupler tools

- OpenPALM, OASIS, MpCCI, …

2 approaches

- Direct coupling (multiple binaries) vs coupling via top-level interfaces (one unified binary)

But strong challenges to face for Exascale

- Standard coupling API
- Memory footprint, use of asynchronous (reduced) communications, …
- Interpolation methods, smart search algorithms, …
EESI2 recommandation on Mesh

Mesh Generation

- Automatic, Adaptive, Intelligent
- From billions « regular » meshes ... to millions!

With numerical methods for solving differential equations such as
Discontinuous Galerkin: combining features of the finite element and the finite volume framework
DG methods have a huge interest for electrodynamics, fluid mechanics and plasma physics equations.
Conclusions

IDC 2014 Study for EESI: **HPC must produce ROI**
There is the only way for Exascale!
Industrial applications are the key factors of potential Exascale viability

-> **Exascale applications must be efficient for ROI!**
   largely than they are today!

So, need of R&D programs, need of innovation, disruptive methods on all scientific domains ...

Please Submit EU Calls on concrete R&D for Exascale

**Thank you for your future active contribution**
Extreme Data and Computing Initiative

Following EESI, and gathering the EC pillars of HPC

With an economic vision of HPC and Exascale
And Integration of the “fruits” of EESI recommendations

Accepted by EC On March 2015,
Start: Sept 2015 (?)
Leader: Sergi Girona (Prace)
Scientific Director: François Bodin (Irisa)
EESI is also friendly and largely open

Open to new disruptive
Ex: Non Linear Algebra ($2^k$, zero divisors, ...) breaking Vectors & Matrices structure,

THANK YOU

Le Tremblay 2013

Bologna 2014