FP7 Support Action - European Exascale Software Initiative

DG Information Society and the unit e-Infrastructures

EESI Final Conference

EESI WG4.1: Hardware roadmap, links and vendors

Herbert Huber, Chair

Riccardo Brunino, Vice Chair

STRATOS

WG4.1: Goal

- Compile an international cartography of multi-Peta- to Exascale vendor HPC R&D activities and provide a set of recommendations to the EC
- Identify gaps, risks and deficiencies in the vendor HPC R&D research plan
- Judge their severity and conclude a set of recommendations

WG4.1 Experts

Leif Nordlund	AMD	SE	Business Development Manager
Jean-Pierre Panziera	Bull	FR	Director of Performance Engineering
Wilfried Oed	Cray	DE	HPC Engineer, EMEA
Luigi Brochard	IBM	FR	Distinguished HPC Engineer
Andrey Semin	Intel	RUS	HPC Technology Manager, EMEA
Frank Baetke	HP	DE	Worldwide Director of High Performance Computing
Michael Kagan	Mellanox	IL	CTO of Mellanox Technologies
Dev Tyagi	Supermicro	UK	General Manager at Supermicro, UK
Giampetro Tecchiolli	Eurotech	IT	CEO/CTO of Eurotech
Salvatore Rinaudo	ST Microelectronics	IT	ST Microelectronics
Rüdiger Wolff	SGI	DE	Principal Systems Engineer & Systems Engineer Manager, EMEA
Dmitry Tkachev	T-Platforms	RUS	Chief Architect
Pierre Lagier	Fujitsu	FR	Technical Director
Ulrich Brüning	ExTOLL	DE	Computer Architecture Research Manager
Aad van der Steen	NCF	NL	HPC architectures & benchmarking
Alex Ramirez	BSC	ES	Computer Architecture Research Manager
Dominik Ulmer	CSCS	CH	HPC Center
Jean Gonnord	CEA	FR	Numerical Simulation & Computer Sciences

Questions for the experts

- **What are the major challenges/issues in HPC in the next 5 to 10 years?**
- **What efforts will your institution or company undertake to address these challenges/issues?**
  - **a) If applicable, please mention all R&D on promising future multi-Peta to Exascale hardware technologies such as hybrid many-core CPUs, novel memory technologies (3D-stacked memory, memristor, Spin Torque Transfer Magnetic RAM, graphene based memory, etc.), novel background storage technologies, photonic interconnects, highly cooling efficient system packaging technologies, etc.**
    b) If applicable, please mention all R&D on promising future multi-Peta to Exascale software technologies such as highly scalable programming languages, fault tolerant parallel programming environments, parallel file systems with end-to-end data integrity, highly scalable and energy efficient system management software, hierarchical storage solutions, OS with coherent inter-node scheduling mechanisms, etc.
    c) If applicable, please mention all R&D on promising highly scalable numerical algorithm for multi-Peta to Exascale applications

International HPC R&D(*) cartography

NOVEL PROCESSING ELEMENTS													
System on a Chip	X	X		X	X	X	X	X	X	X	X	X	11
Heterogeneous many-core	X	X			X	X	X	X	X	X	X	X	10
Homogeneous many-core	X	X	X		X	X						X	6
Company	AMD	ARM	Fujitsu	HP	IBM	Intel	LSI	nVIDIA	Supermicro	STMicroelectronics	Tilera	T-Platforms	Counts

(*) An "X" in a field of the table means the vendor is performing active research and development in this field

International HPC R&D(*) cartography

SEMICONDUCTOR & STORAGE TECHNOLOGY												
3D Stacking	X	X	X	X	X	X		X		X	X	9
Novel Memory Technologies			X	X	X	X		X			X	6
Novel Background Storage Technologies			X	X	X	X	X		X	X		7
Novel Semiconductor Materials			X	X	X	X					X	5
Company	AMD	ARM	Fujitsu	HP	IBM	Intel	LSI	NVIDIA	SGI	Supermicro	STMicroelectronics	Counts

(*) An "X" in a field of the table means the vendor is performing active research and development in this field

International HPC R&D(*) cartography

INTER-NODE COMMUNICATION NETWORK & INTRA-NODE INTERCONNECTS															
In-interconnect processing elements	X	X	X					X		X	X				6
Silicon Photonics	X		X		X	X	X	X	X		X	X	X		11
Inter Node Communication Network & Network Routing Algorithms	X	X	X	X	X	X	X	X	X	X	X			X	13
Intra Node Interconnect	X	X			X	X	X	X	X		X		X	X	10
Company	AMD	BULL	CRAY	Eurotech	ExTOLL	Fujitsu	HP	IBM	Intel	Mellanox	SGI	Supermicro	STMicro-electronics	Tilera	T-Platforms	Counts

(*) An "X" in a field of the table means the vendor is performing active research and development in this field

International HPC R&D(*) cartography

RESILIENCE AND FAULT TOLERANCE															
HW Level	X	X	X	X	X	X	X	X	X	X	X	X	X	X	15
SW Level				X		X	X	X	X		X	X			X	8
Company	AMD	ARM	BULL	CRAY	Fujitsu	HP	IBM	Intel	Mellanox	nVIDIA	Panasas	SGI	Supermicro	Tilera	T-Platforms	Counts

(*) An "X" in a field of the table means the vendor is performing active research and development in this field

International HPC R&D(*) cartography

MULTI-PETA- TO EXASCALE SOFTWARE STACK																	
Parallel Runtime Systems with**/without Fault Tolerance			X	X		X	X	X	X	X	X			X		X			8 2
Parallel File Systems			X		X			X		X			X				X	X	7
Compiler, Debugger, Performance Tools	X	X		X				X		X	X	X			X				8
Novel Programming Languages	X	X		X				X	X	X	X	X			X				9
Numerical Libraries	X			X				X		X	X	X							6
OS with coherent inter-node scheduling mechanisms			X	X				X		X				X		X			6
Co-Design	X	X	X	X		X	X	X	X	X	X	X		X	X	X			14
Company	AMD	ARM	BULL	CRAY	DDN	Eurotech	ExTOLL	Fujitsu	HP	IBM	Intel	NVIDIA	Panasas	SGI	Tilera	T-Platforms	Whamcloud	Xyratex	Counts

(*) An "X" in a field of the table means the vendor is performing active research and development in this field

(**) A fault tolerant runtime system will be capable to run a user job to completion even in the case of a compute node failure

International HPC R&D(*) cartography

ENERGY EFFICIENCY																		
Power Efficient Hardware Technologies	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	18
Power Distribution and Management			X	X	X	X		X	X	X	X		X	X			X	11
Power Efficient System Packaging and Cooling Technologies	X		X	X	X	X		X	X	X	X		X	X			X	12
Power-aware Software Technologies	X	X	X	X	X	X		X	X	X	X		X				X	12
Company	AMD	ARM	BULL	CRAY	DDN	Eurotech	ExTOLL	Fujitsu	HP	IBM	Intel	NVIDIA	SGI	Supermicro	STMicroelectronics	Tilera	T-Platforms	Xyratex	Counts

(^) An "X" in a field of the table means the vendor is performing active research and development in this field

Major deficiencies in current HPC technology provider R&D plans

1. **Future many-Peta to Exascale file systems**
   - Metadata performance
   - Scalability and reliability
     - Number of concurrent clients
     - Number of files
     - File system capacity
   - Data protection and recovery mechanisms
     - Distributed RAID mechanisms
     - Analysis and diagnostics database
   - Data mining and visualization tools
2. **Future I/O**
   - Mechanisms for end-to-end data integrity
   - High speed SERDES technology
3. **Resilience and fault tolerance**
   - Fault tolerant parallel runtime systems
   - Mechanisms for error prediction, automatic error detection and recovery

Major deficiencies in current HPC technology provider R&D plans

4. **Fundamental Research**
   - Novel semiconducting materials
   - Novel memory technologies
   - Novel background storage technologies
   - Integration of optical technology into the standard CMOS process

5. **Communication networks with in-interconnect processing capabilities**

6. **Operating system enhancements**
   - Coherent inter-node scheduling mechanisms
   - Micro OS
   - Data locality

7. **Numerical libraries and highly scalable performance analysis tools**

Opportunities for European HPC Technology Providers

- Green IT:
  - Cooling efficient system packaging technologies
  - Energy efficient data centre infrastructures
  - New technologies for reuse of system waste heat
  - Software for data centre energy monitoring and control
- Data handling and file I/O: Intelligent parallel multi-Tier file storage and parallel file systems
- Fault tolerant parallel runtime environments

Important findings

- HPC vendors are willing to collaborate with the EC and the European HPC community
- At least one European HPC system vendor has recently successfully demonstrated it's expertise to build HPC system at the multi-Petascale performance level
- Europe still has all the necessary knowledge at hand to develop its own HPC technology stack if deemed necessary