Computing and Storage

Chapter 9

Ronald Begeer, Marc Duranton &
Arnaud Samama

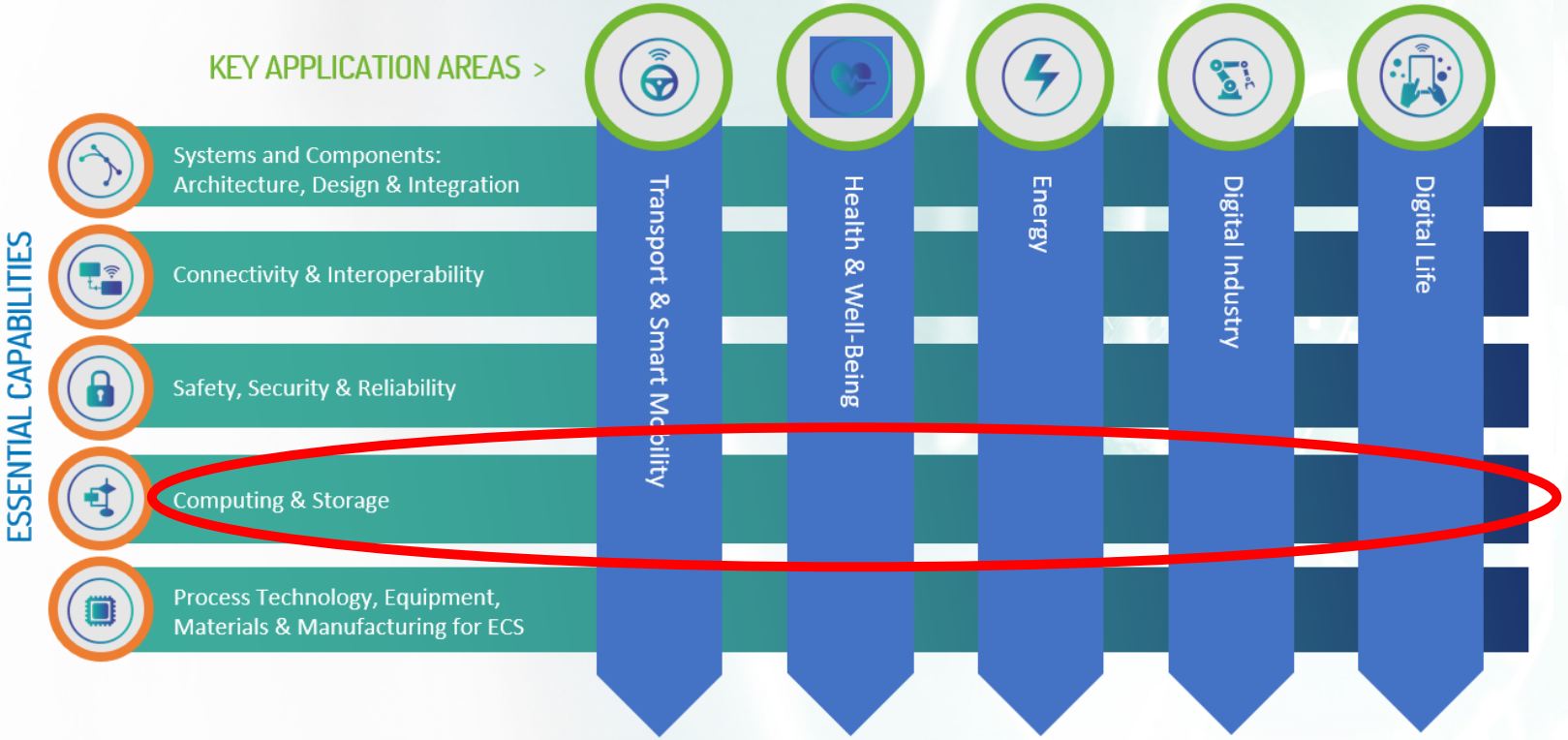
Organised by:
AENEAS, ARTEMIS-IA, EPoSS, ECSEL-JU &
European Commission

Associated organiser:
EUREKA

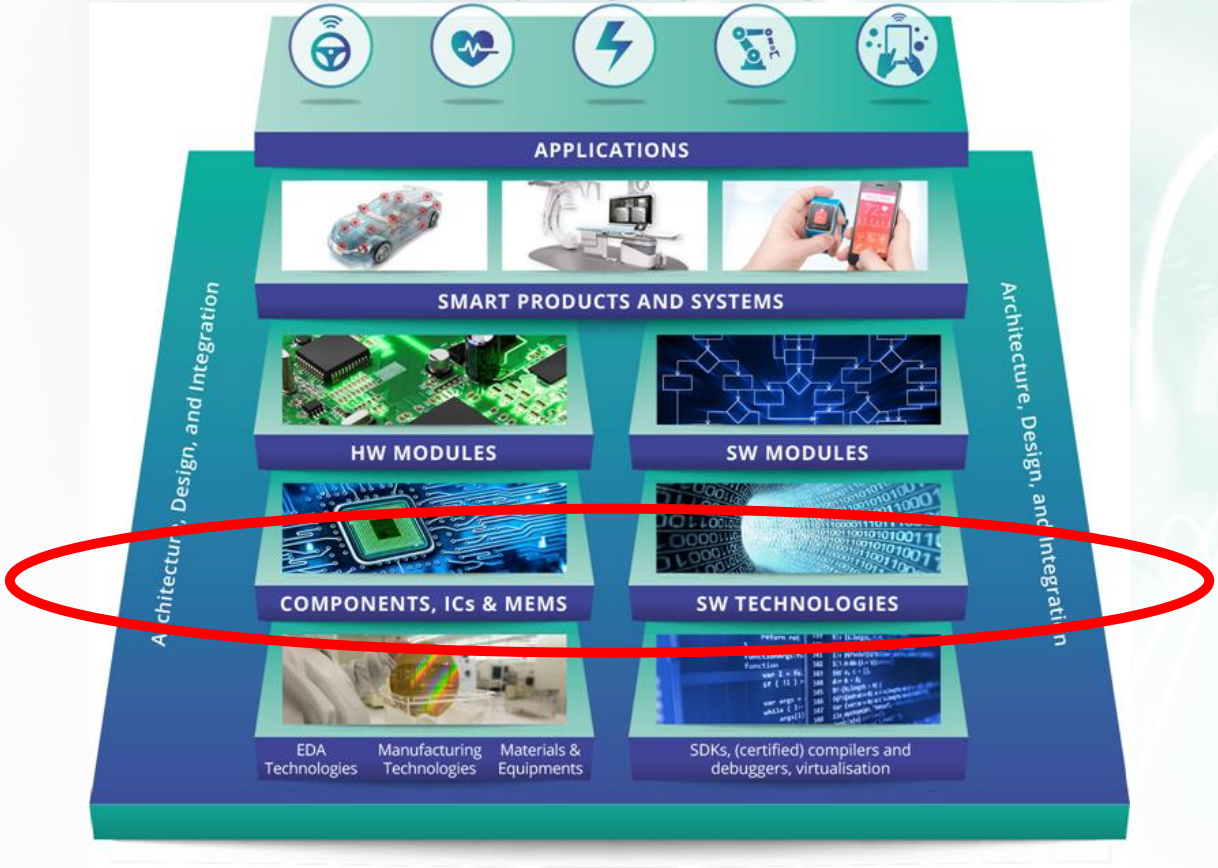
Computing & storage

Computing and storage are the fuel of the digital revolution in providing ever increasing performance for existing and new applications at a constant or decreasing cost.

Overall SRA structure



Fit in overall SRA



NEW

Open Computing Platforms

HECS 2018 Our digital future

- Coprocessors, GPU and Machine Learning accelerators are becoming more and more important often controlled by the USA.
- A rising approach to change this situation and prevent dependence on foreign closed processing technologies, relies on Open Hardware initiatives (Open Compute Project, RISC-V, OpenCores, OpenCAPI, etc.).
- The adoption of a shared open ecosystem approach, with a globally and incrementally built know-how by multiple actors, prevents that a single entity can be purchased or cease to exist for other reasons.
- The very low up-front cost of open hardware/silicon IP lowers the barrier of innovation for small players to create, customize, integrate or improve Open IP to their specific needs (*but the cost of foundry is still a bottleneck*).

NEW

Energy Usage/5G

HECS 2018
Our digital future

- Energy cost of communication is not only at the chip/board and system levels: it is also between systems.
- The energy usage challenge is not only at the component level, but also at the system and even infrastructure level
- The emergence of 5G will provide new way of distributing and optimizing energy usage, but will also require the design of the computing elements required to ensure efficient 5G systems and infrastructure.
- The guaranteed low latency and quality of service of 5G will open the path to more reliable distributed systems.

High Performance Computing

- The unprecedented level of computing power offered by Exascale and post exascale is expected to significantly enhance our knowledge for the benefit of a large spectrum of industries
- The HPC community also see for the future a change in the nature of the loads: convergence between simulations, applications related to big data processing and Artificial Intelligence loads will be mixed with simulations like digital twins and ”*real-time*” requirements of *HPC in the loop*

- The ETP4HPC, HiPEAC and BDVA are working together in order to precise the future requirements for post exascale machines.

Data Analytics

- This trend leads to the following major challenges for computing technologies:
- Increasing performance (including efficiency) at acceptable costs
  - For High Performance Computing (HPC) and servers
  - For low power and ultra-low power computing (edge and deep edge computing)
  - For Data Analytics (data intensive systems)
- Making computing systems more integrated with the real world and
- Making "intelligent" machi
- M
re systems

The advance in artificial intelligence impacts devices like mobile Phones, self-driving car etc, allowing them to better interact with the physical world.

Growing Application Complexity

- The application complexity is increasing dramatically (parallelism, heterogeneity, distributed) and optimisation represents a challenge.
- The development of such complex applications also requires new automated tools to support the debugging, validation and certification tasks.
- Software should also contribute to the solution of hardware-critical challenges by providing robust, energy-aware and fault-tolerant, self-healing applications.

Software Challenges in Growing Complexity

- To support system design and realization, further development of software technologies that enable system specification at higher abstraction levels (o.a. model-driven SW realization, Domain-Specific Languages) are required.
- New Computing technologies, applied to largely distributed and dynamic systems needs the creation of software technologies to enable viable development and implementation.
- Without efficient SW technologies and tools, the benefits of the compute technologies and architectures will remain theoretical rather than practical. (link with Chapter 6)

Qualified / Explainable AI

- Deep learning systems are essentially *black-boxes* and their results are often difficult to confront with the initial requirements (due to unforeseen *bias* in the selection of the learning databases), therefore a current challenge is to be able to qualify/certify Deep Learning based systems.
- A subsequent step is in Explainable (for human) Artificial Intelligence
- A more general challenge is determining whether a complex system, composed of white, grey, and black, boxes, will ensure the objectives and quality of service for which it has been designed.

Release Notes

- Challenges related to energy consumption are discussed, as they require new approaches not only at the component level, but also at the system and even infrastructure level.
- Consequences of the on-going revolution of how we now interact with machines, mainly due to the advance in Artificial intelligence, are developed.
- Open Hardware initiatives (Open Compute Project, RISC-V, OpenCores, OpenCAPI, etc.) are addressed in this new edition of the SRA. Indeed, while Coprocessors, GPU and Deep Learning accelerators (and other accelerators) are becoming more and more important in computing architectures, those Open Hardware initiatives (Open Compute Project, RISC-V, OpenCores, OpenCAPI, etc.) are gaining momentum as a way to prevent dependence on non-European companies for the provisioning of these devices.
- In the “Developing new disruptive technologies” section, some content have been moved into the new Long Term Vision chapter (chapter 11), while other topics have been explicitly linked with the technology development that they require (e.g., quantum computing requirements were linked to quantum technology developments addressed in chapter 10).