MANGO: implications and contributions to Extreme-Scale Demonstrators

EsD roundtable@ European HPC Summit Week

May 18th, 2017, Barcelona, Catalunya

Alessandro Cilardo

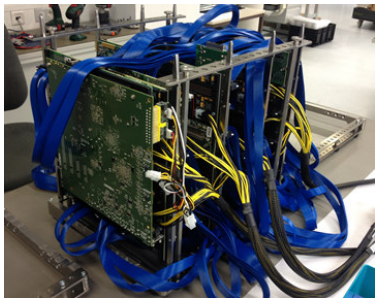
acilardo@unina.it

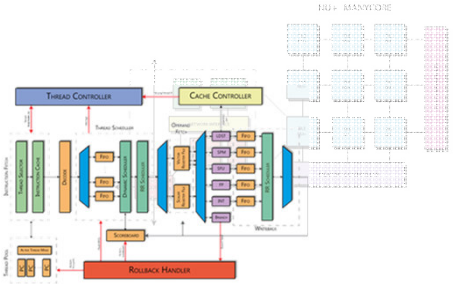
MANGO: exploring Manycore Architectures for Next-GeneratiOn HPC systems

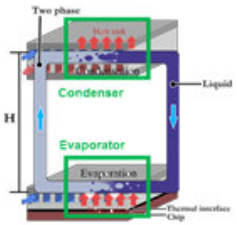
This project has received funding from the European Union's H2020 research and innovation programme under grant agreement No 671668

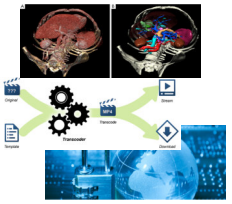
HIGHLIGHTS

- MANGO **FETHPC**-2014 project:
  - is about ***manycore architecture exploration in HPC***
- General-purpose nodes (Xeon+GPGPU) coupled with Heterogeneous nodes, HNs:
  - A large-scale cluster of high-capacity FPGAs
  - A robust, scalable interconnect for a **multi-FPGA manycore** system
  - Will enable FPGA acceleration *at scale*:
    → a key ingredient for the EsD roadmap
  - A continuum from FPGA emulation to the final physical platform (might be an ASIC manycore, FPGA, mixed…)
    → ***under a stable software environment***
  - Native isolation and partitioning mechanisms for **QoS-aware capacity computing** HPC applications
  - Highly customizable GPU-like / vector cores
- Two-phase passive **energy-efficient cooling**
- Demonstrated applications with stringent high-performance and QoS requirements









MANGO AND EsD

- What will MANGO bring to the EsD roadmap?

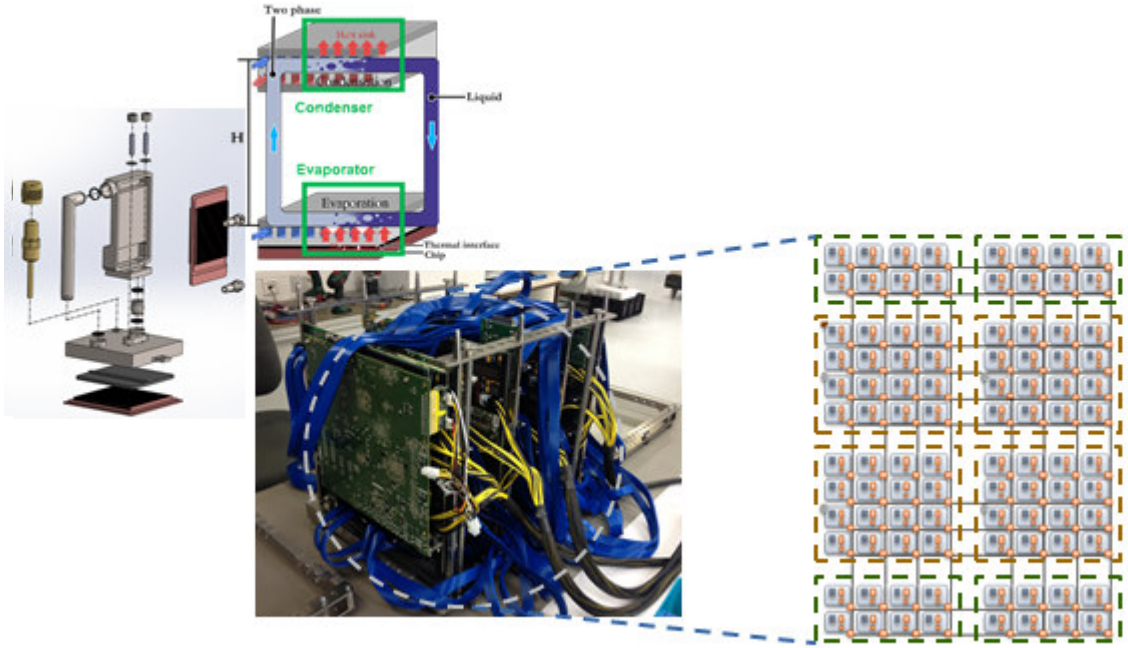
→ answer three important questions:

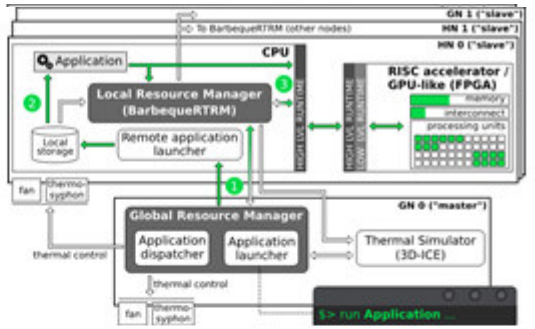
  – How to shape custom hardware acceleration in HPC ***?***

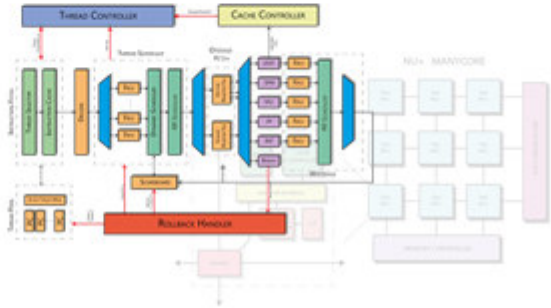
  – How to organize and exploit FPGA devices *at scale* ***?***

  – How to substantially reduce cooling cost in heterogeneous nodes ***?***

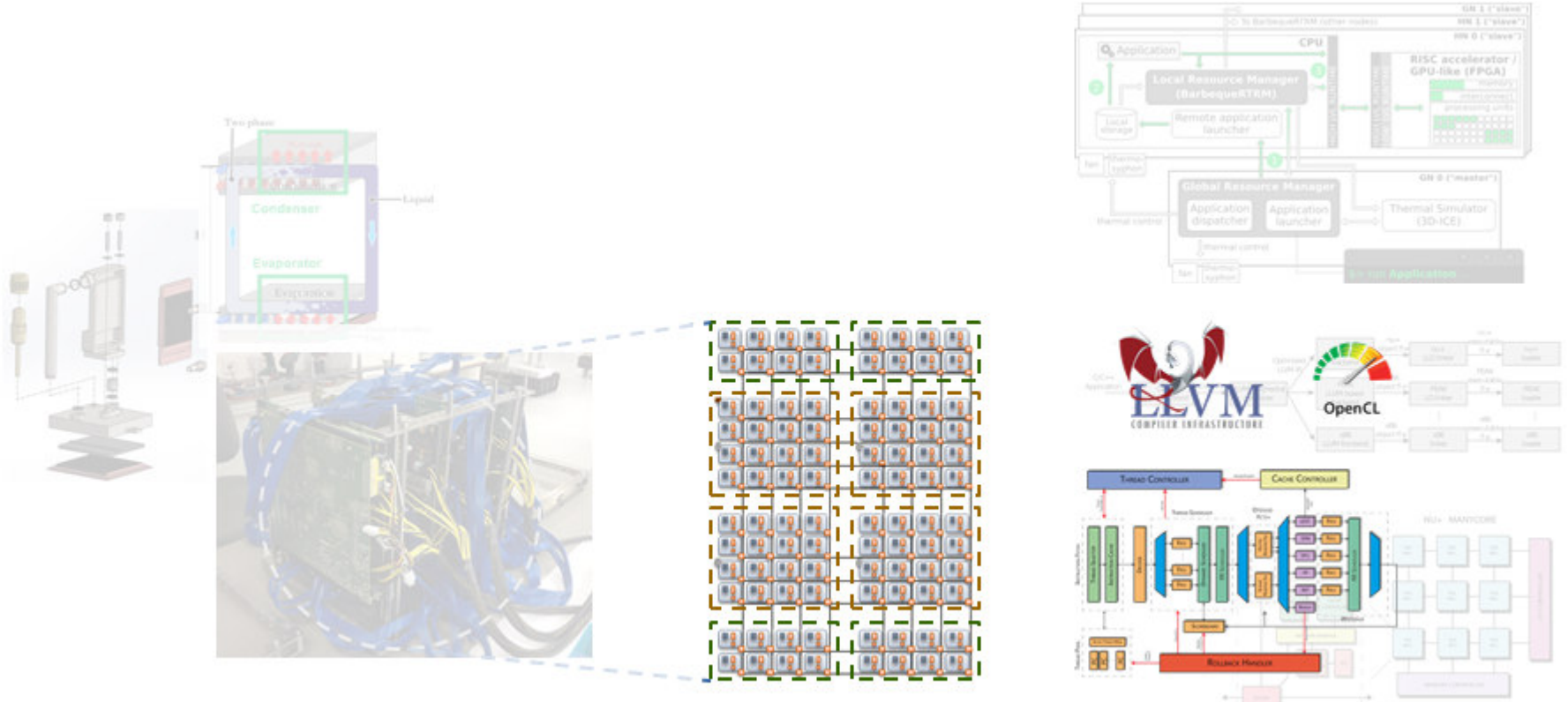
THE MANGO HW/SW ECOSYSTEM







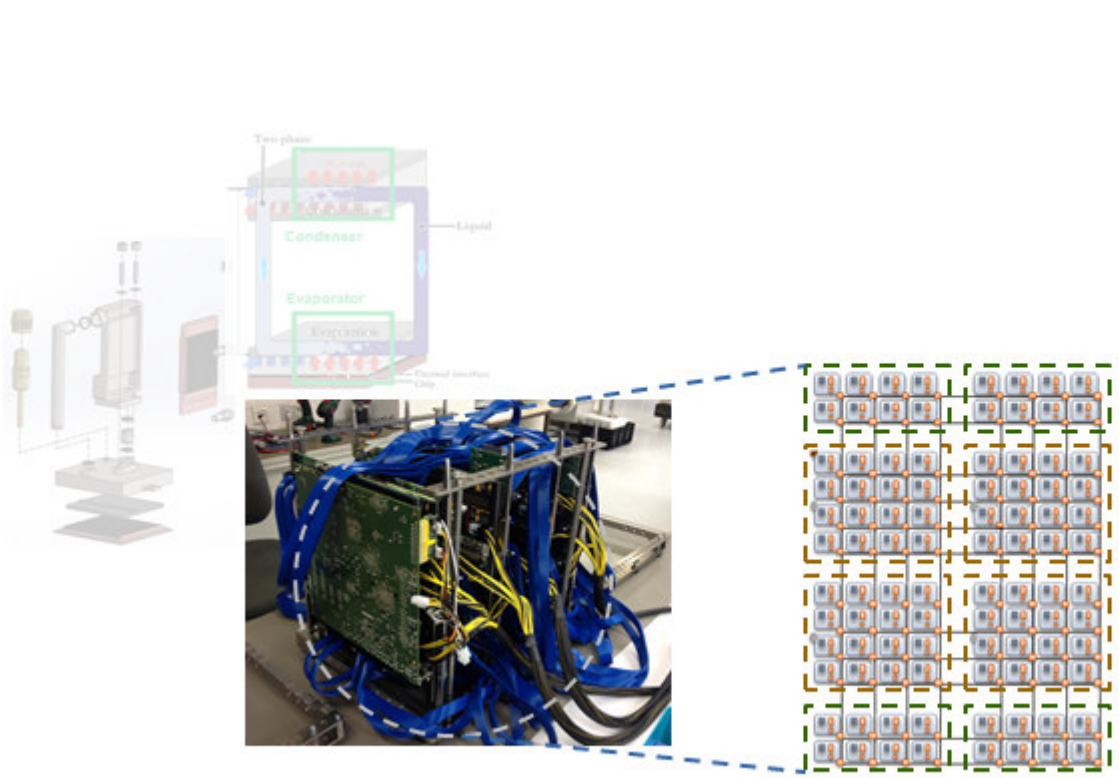
MANGO: CUSTOM COMPUTE UNITS



Configurable vector/GPU-like accelerators enabling application-driven customization

- Vector/GPU-like units (***nu+*** core) within a multi-level manycore system
- Fully customizable hardware features: FP precision, lanes, hw threads etc…
- Stable software environment (LLVM compiler, OpenCL support, API)
- Coupled with specialized algorithm accelerators, possibly generated through HLS

MANGO: MANYCORE INFRASTRUCTURE



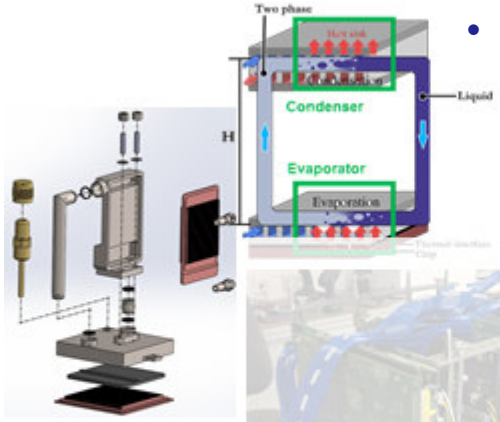
Multi-FPGA infrastructure and interconnect

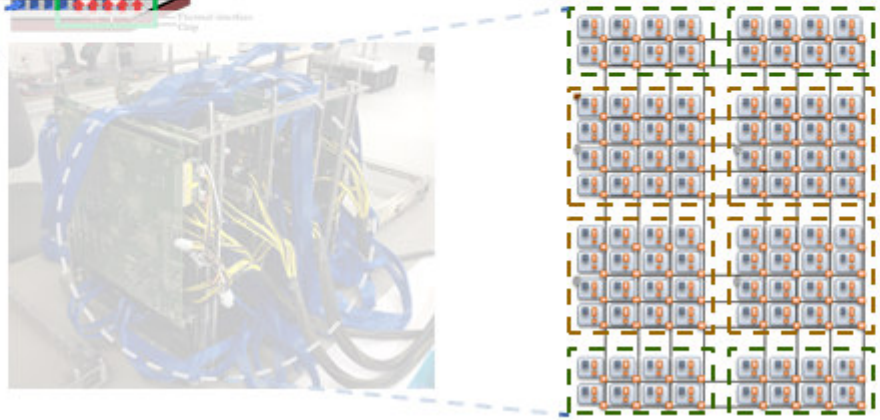
- Board design, advanced **multi-FPGA** manycore, interfacing, …
- Scalable interconnect ("off-chip" NoC)
- Partitioning/isolation mechanisms for QoS-aware resource management

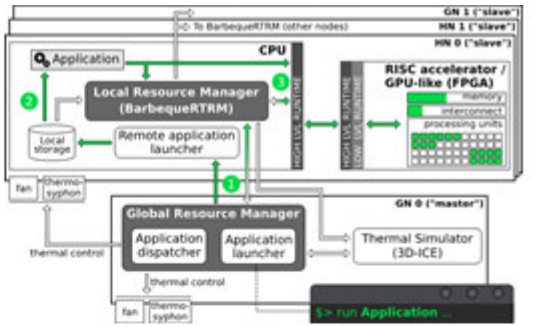
MANGO: COOLING SYSTEM

Energy-efficient passive cooling

- **Thermosyphon** concept: two-phase passive cooling
- PUE = **1.02** (*vs*. **1.60** of air cooling or **1.10** of liquid cooling)







KEY CONTRIBUTIONS

So, to recap…

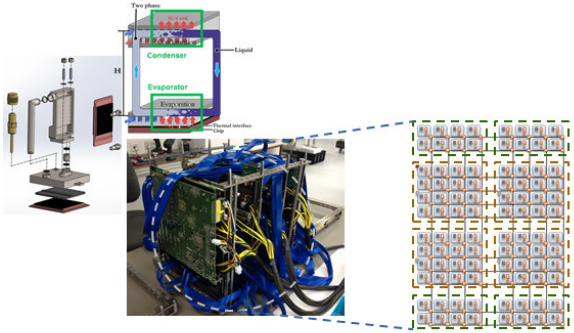
- MANGO key contributions to EsD:
  - Customizable, ***software-programmable*** large accelerators (possibly coupled with specific custom hardware blocks)
    - Vector units, custom precisions, customized non-coherent memory…
    - Compute unit architecture can be mapped to various hw technologies relying on a *stable software ecosystem*
  - Infrastructure for interconnecting FPGAs in a manycore system
    - Advanced network with QoS/isolation mechanisms embracing clusters of FPGAs (enables architecture-wide customization, memory partitioning, some form of *close-to-data* computing)
    - Makes HPC ready for ***FPGA acceleration at scale***
  - Innovative concept for 2-phase passive ***energy-efficient cooling***.

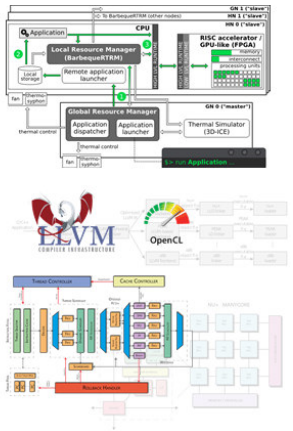
USING/INTEGRATING MANGO TECHNOLOGIES

- Vector/GPU-like ***nu+*** core
  - **LLVM** backend available
  - **OpenCL** support to be provided soon
  - can be coupled with commercial OpenCL-based **HLS** flows
  - Possible technology remapping (with *no change* at the SW level)
- Multi-FPGA / manycore **infrastructure**
  - Custom interconnect hidden to applications and software
  - Non-proprietary interfaces: PCIe, Gigabit Ethernet, DDR3
  - Integration with general-purpose nodes already demonstrated
  - Configuration knobs (mapping, partitioning,…) exposed to RTMS
- RunTime Management System (RTMS) implementation
  - Global RunTime Management System based on **SLURM**
  - Policies as plugins: no need to modify the SLURM core
  - Local RTMS based on the Barbeque open-source project
- Cooling system:
  - involves the mechanical design at the board/rack level
  - MANGO developed a general methodology for **cooling design**
  - can be readily applied to next-generation HPC systems

MATURITY AND ESD ROADMAP

- Key innovations have been demonstrated
  - Intermediate Review Meeting held May 10th, 2017



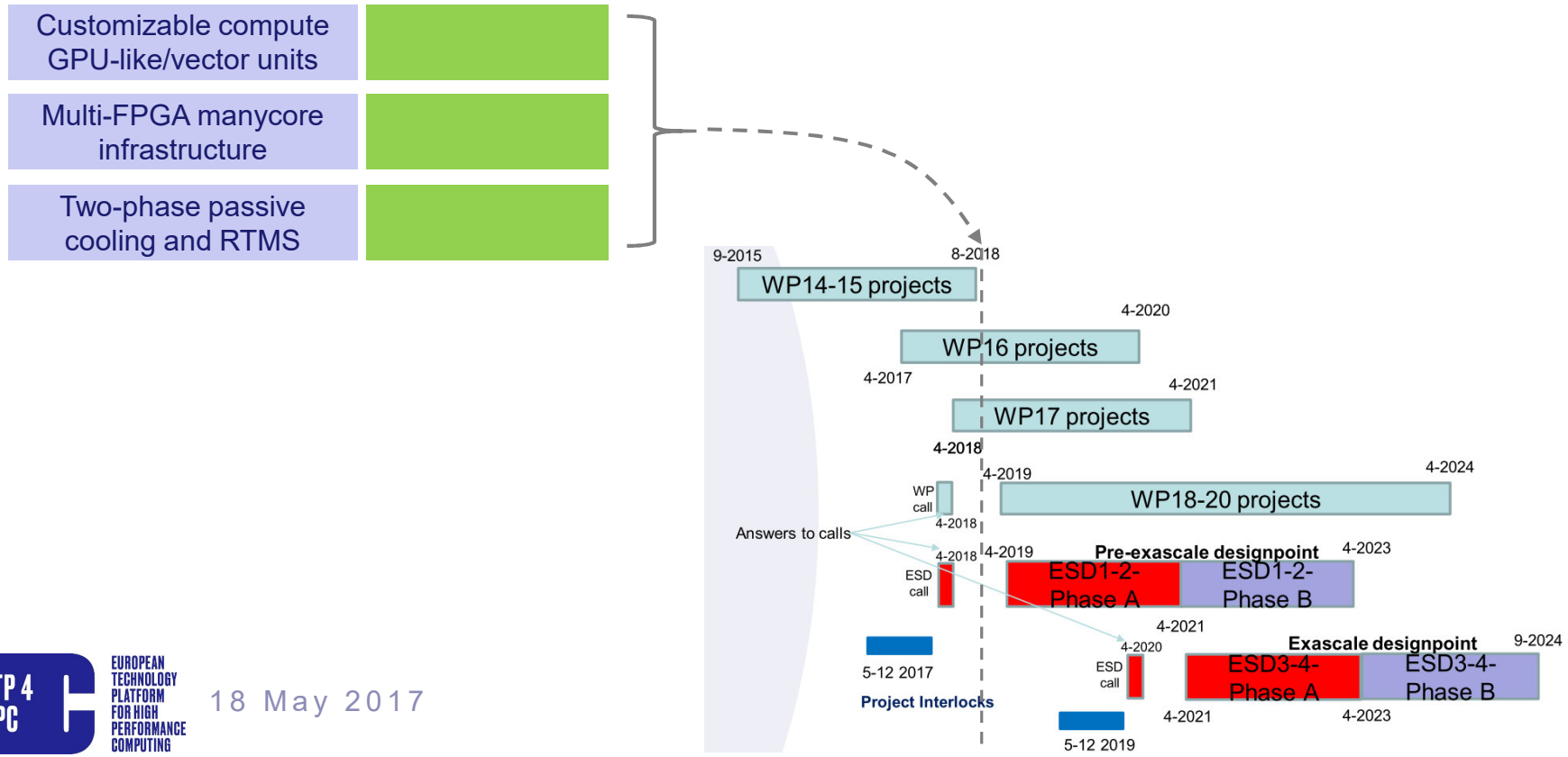


9/2018

Customizable compute GPU-like/vector units	75%
Multi-FPGA manycore infrastructure	70%
Two-phase passive cooling and RTMS	65%

MATURITY AND ESD ROADMAP

- Key innovations have been demonstrated
  - Intermediate Review Meeting held May 10th, 2017
- **TRL6** / **TRL7** expected by Oct 2018 (pre EsD1-2 Phase A)



WHAT'S NEXT?

- Timing and maturity fit the EsD roadmap
- Relevant real-world applications are being fully ported
  - Videotranscoding, medical imaging, DSP and real-time crypto-processing
- MANGO has **complementarities** with other FETHPC projects
  - Potential synergies:
  - Standard CPUs/accelerators, storage and new memory technologies, advanced programming models
- MANGO will provide a few ***key missing pieces*** for EsD
  - Customizable compute units that can be specified in an application-driven fashion
  - Comprehensive, scalable, future-proof infrastructure support for hardware acceleration in HPC
  - Innovative passive cooling enabling unprecedented values of PUE
- Next step on top of the current MANGO roadmap
  - Launch a **Pilot** to demonstrate factual interplay with other projects
  - We will soon solicit focused exchange actions