Green HPC with Low-power Manycores

Márcio Castro

marcio.castro@ufsc.br - www.marciocastro.com

Graduate Program in Computer Science (PPGCC)
Department of Informatics and Statistics (INE)
Federal University of Santa Catarina (UFSC)

EnergySFE Workshop 2016
Grenoble, France

LAPESD

UFSC

About me

- **Associate Professor**
  - Federal Univ. of Santa Catarina (UFSC)
  - Florianópolis, SC, Brazil

- **Distributed Systems Research Laboratory (LaPeSD)**
  - 7 faculty members
  - 8 Ph.D. students
  - 17 M.Sc. students
  - 6 undergraduate students

Research Interests

- Parallel programming models
- Thread and data affinity
- HPC applications
- Parallel skeletons
- Scheduling and load balancing
- Energy-aware algorithms
- Multicores, manycores and accelerators
- Performance analysis

Students

- **Graduate students (M.Sc.)**

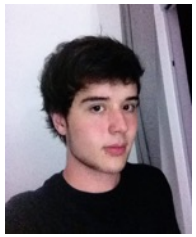
Pedro Penna

- Workload-aware loop scheduling on multicores

Alyson Deives

- Parallel skeletons for heterogeneous architectures

- **Undergraduate student in Computer Science**



Emmanuel Podestá Jr.

- Energy-efficient stencil computations on low-power manycore processors

Motivation

Motivation

- **Until the last decade**
  - Performance of HPC platforms has been quantified by their **processing power (Flops)**
- **Nowadays**
  - **Energy efficiency (Flops/Watt)** is as important as processing power
  - **Critical aspect** to the development of scalable systems

Motivation

- **Defense Advanced Research Projects Agency, EUA (DARPA) report**
  - Acceptable energy efficiency for Exascale systems → **50 GFlops/Watt**
  - Current HPC systems → **7 GFlops/Watt**

Motivation

- **New alternatives for low-power HPC**
  - **Low-power manycore processors**
  - Hundreds of cores in a single chip
  - Very low power consumption: few tens of watts
  - **Examples:**
    - Mellanox TILE-Gx
    - Sunway SW26010 (TaihuLight)
    - **Kalray MPPA-256**

Outline

- Overview of MPPA-256
- Current research efforts
- Results
- Conclusions and future works

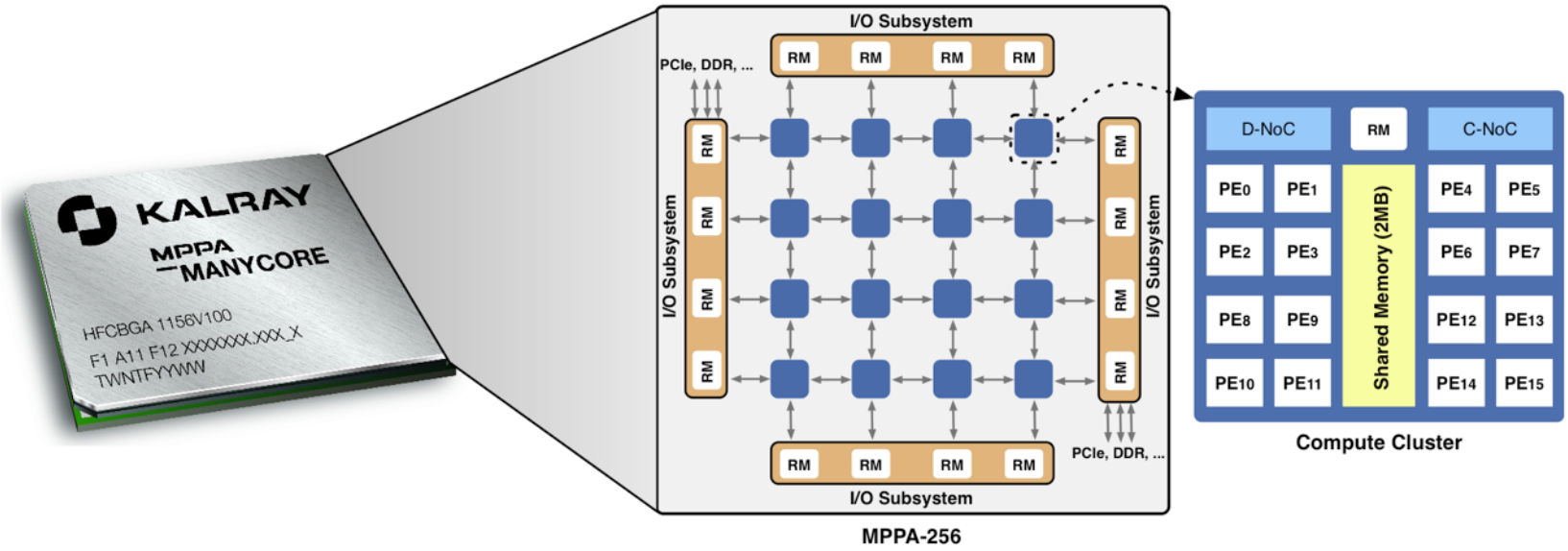
MPPA-256 overview

Kalray MPPA-256 overview

- **Kalray**
  - French semiconductor and software company (Grenoble and Paris) developing and selling a new generation of manycore processors

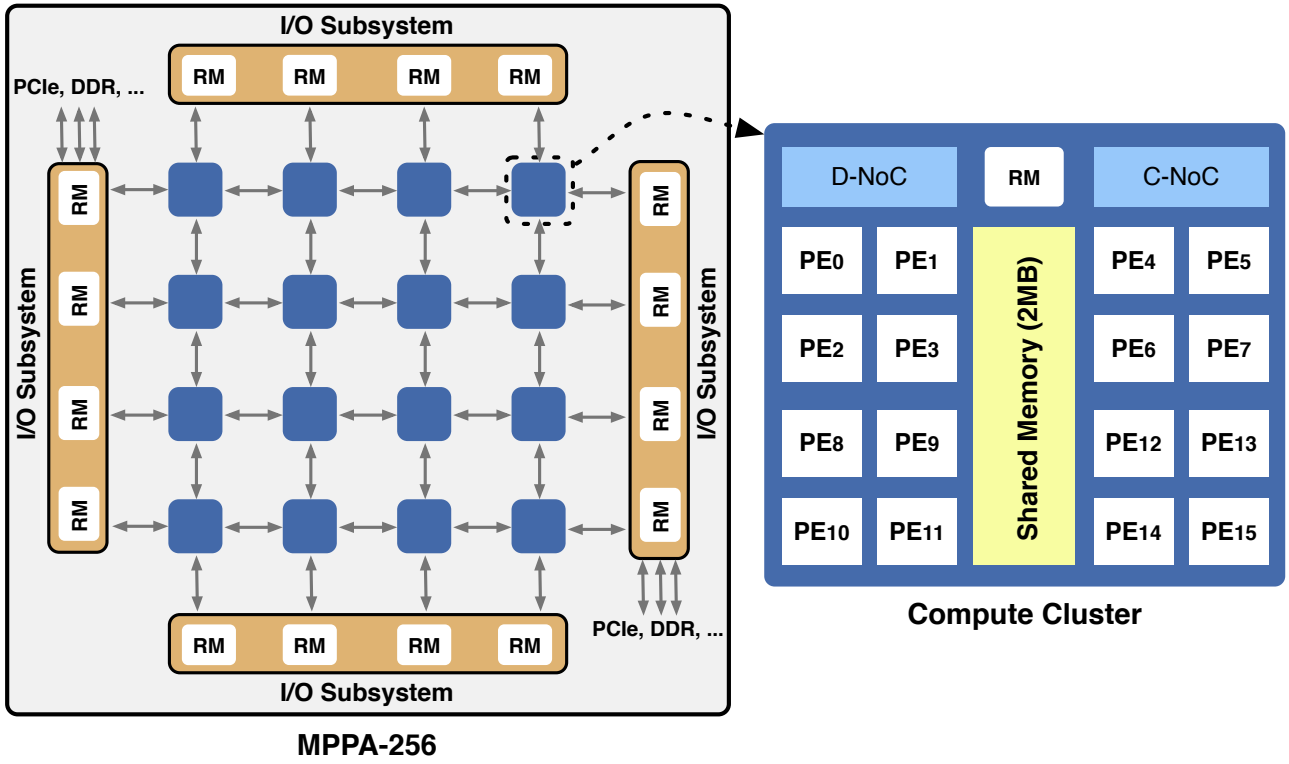
- **MPPA-256**
  - Multi-Purpose Processor Array (MPPA)
  - Manycore processor: 256 cores in a single chip
  - Low power consumption (less than 20W)

Kalray MPPA-256 overview

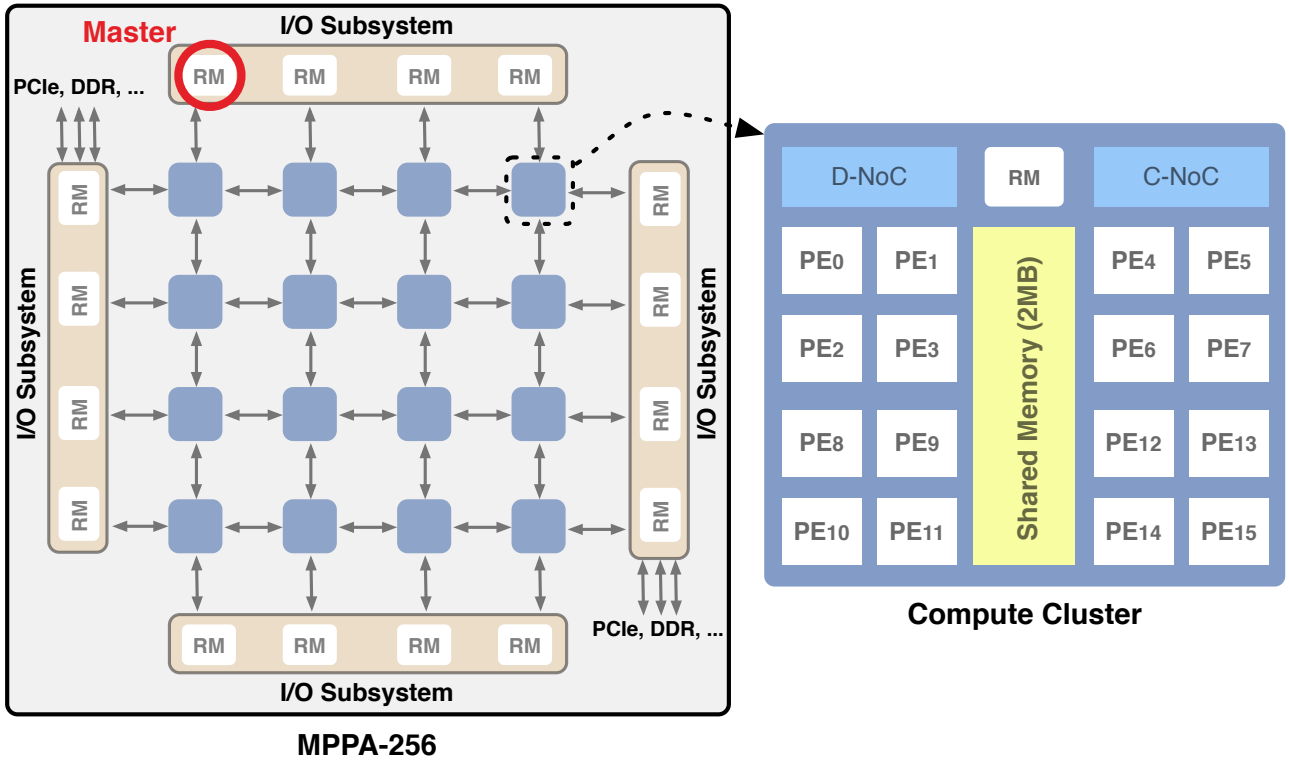


- **256 cores (PEs) @ 400 MHz**: 16 clusters, 16 PEs per cluster
- PEs **share 2 MB of memory**
- **Absence of cache coherence** protocol inside the cluster
- **Network-on-Chip (NoC)**: communication between clusters
- **4 I/O subsystems**: 2 connected to external memory

Kalray MPPA-256 overview

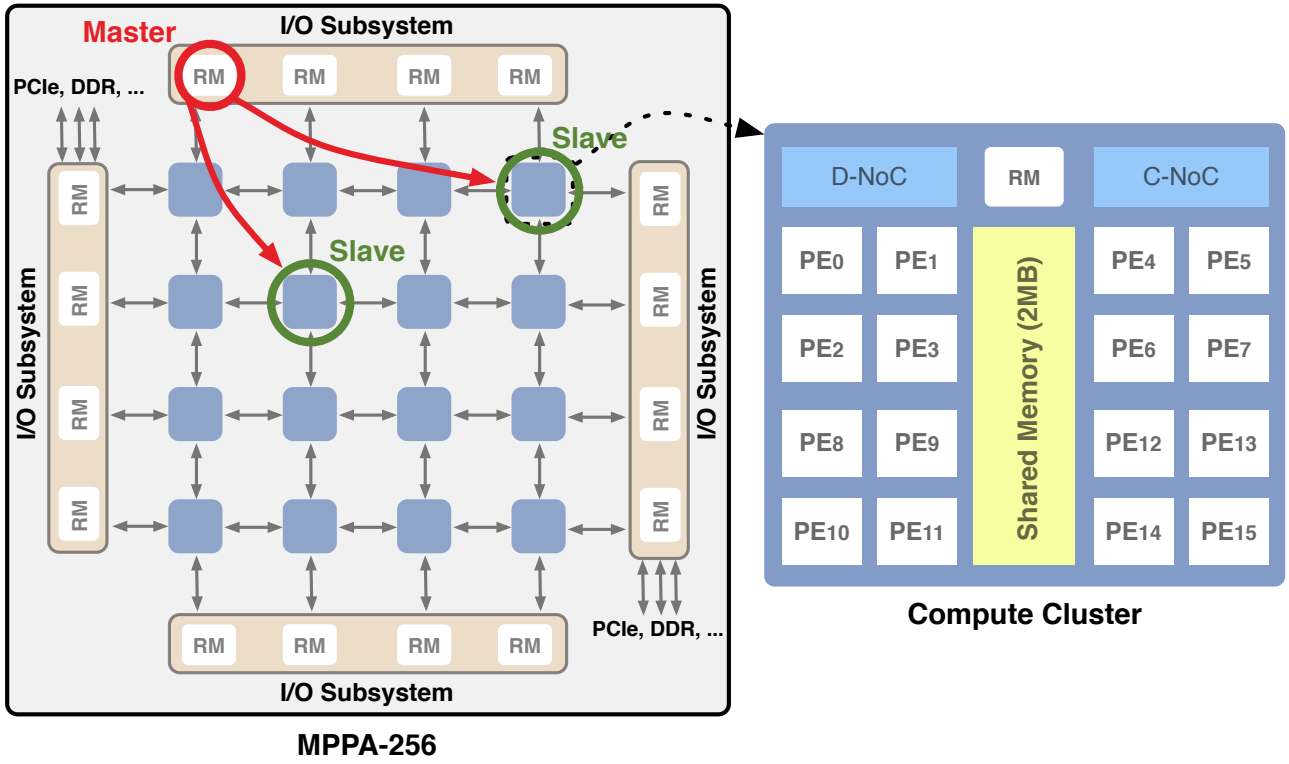


Kalray MPPA-256 overview



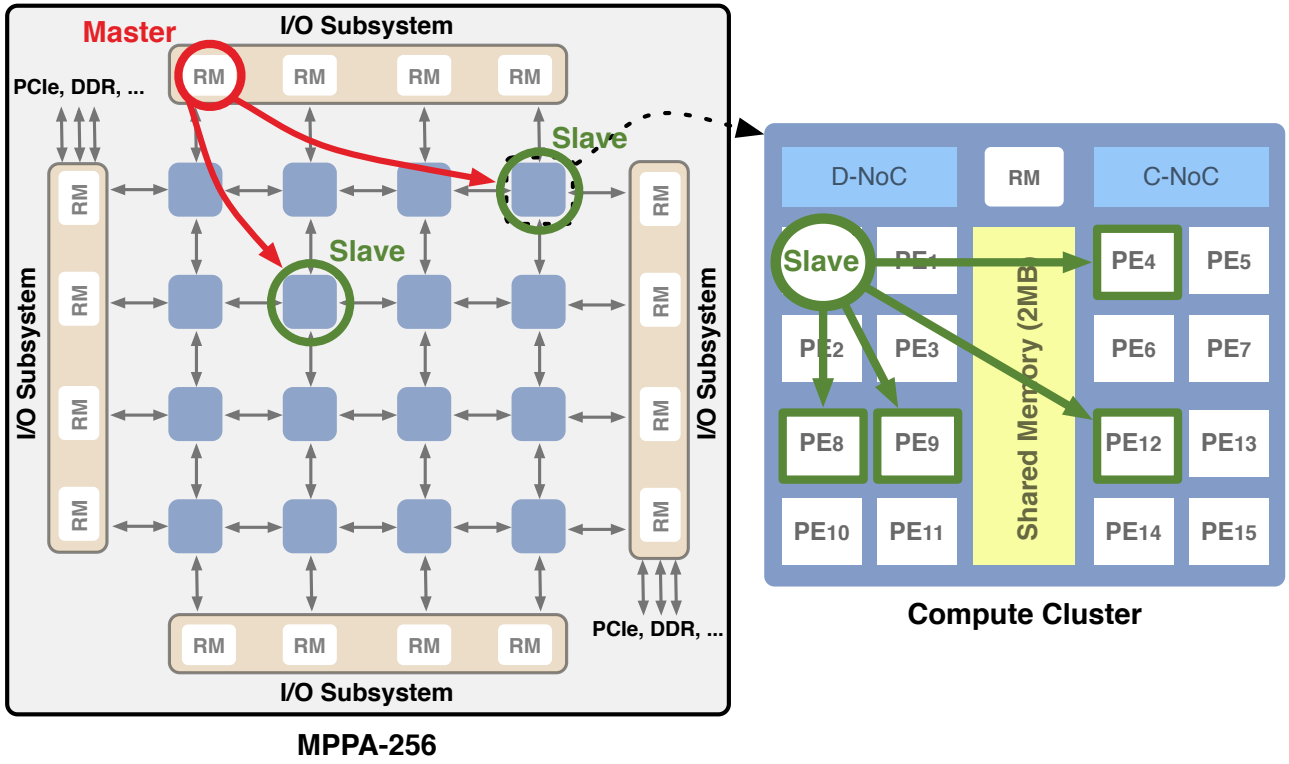
- A **master** process runs on an **RM** of one of the **I/O subsystems**

Kalray MPPA-256 overview



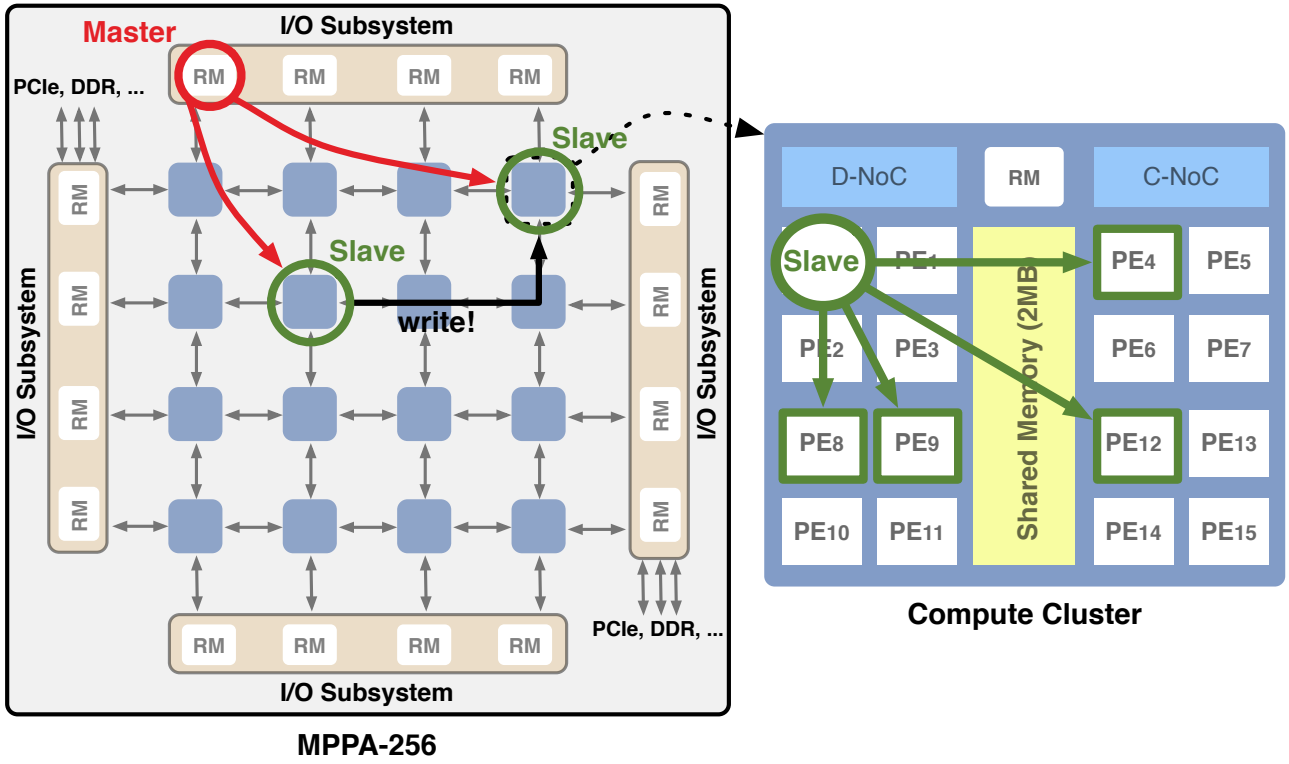
- The **master** process spawns **slave processes**
- One slave process per cluster

Kalray MPPA-256 overview



- The **slave process** runs on the **PE0** and may create up to **16 threads**, one for each PE
  - Pthreads or OpenMP
- Threads share 2 MB of memory

Kalray MPPA-256 overview



- Communications take the form of **remote writes**
- Data is sent through the **NoC**

Current Research Efforts

LAPESD

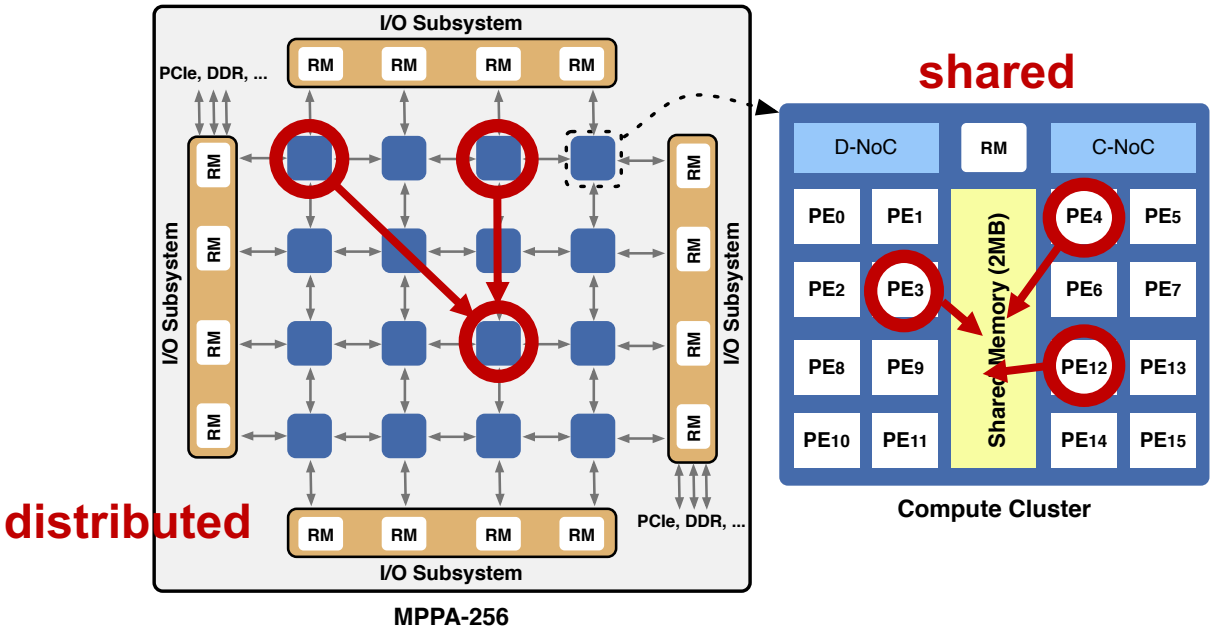
UFSC

Current Research Efforts

- **Many challenges must be faced when developping efficient parallel applications on MPPA-256**
  - Hybrid programming model
  - Memory constraints
  - NoC

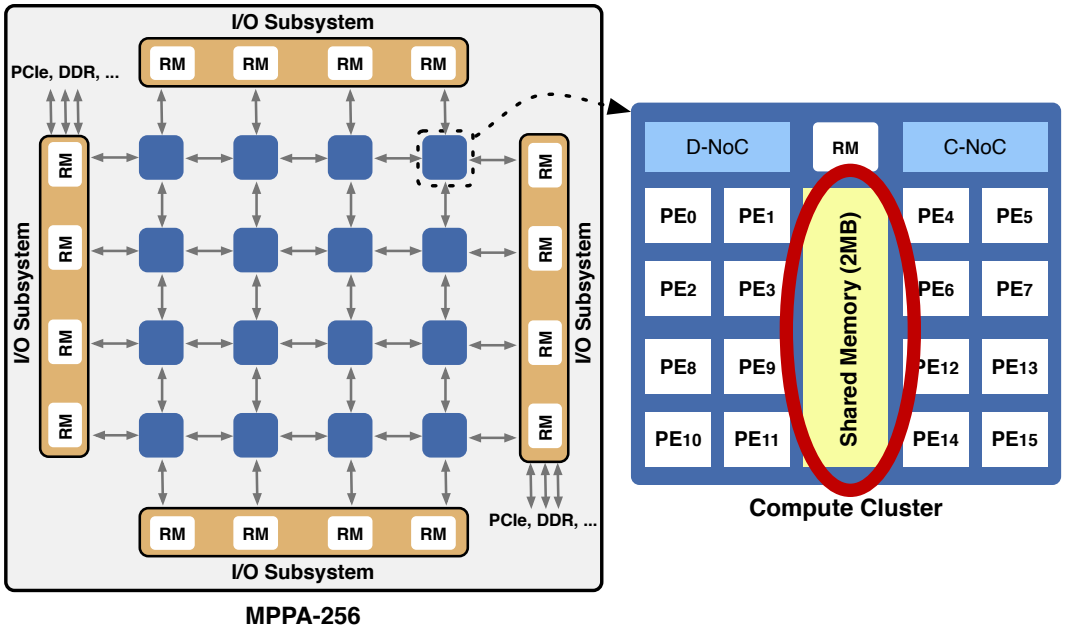
Current Research Efforts

- **Challenges: hybrid programming model**
  - Shared + distributed memory
  - OpenMP/Pthreads + Low-level comm. API



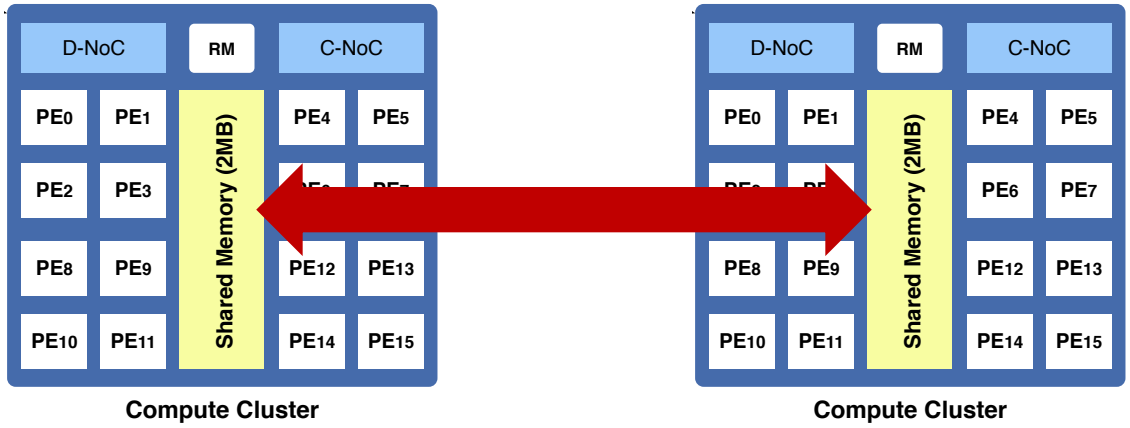
Current Research Efforts

- **Challenges: memory**
  - Scientific apps. don't fit into 32MB
  - Explicit data transfers between the I/O subsystem (DDR) and clusters' internal memory



Current Research Efforts

- **Challenges: NoC**
  - Low-level API to perform remote read/write operations
  - Asynchronous data transfers to overlap communications with computations



Current Research Efforts

- **Research goals**
  - Evaluate the use of MPPA-256 for **high performance computing**
  - Adapt parallel applications for MPPA-256
    - **Different workloads**: cpu-bound, memory-bound, communication-bound, …
    - TSP, **K-Means** and **Seismic Wave Propagation**
  - Propose **new programming models** to ease the development of parallel applications for MPPA-256

Current Research Efforts

- **Research goals (cont.)**
  - Compare the obtained results with other parallel processors
    - General-purpose multicores, embedded multicores and accelerators (GPUs and Xeon Phi)
  - Consider two main metrics:
    - Performance (**time-to-solution**, speedup, ...)
    - **Energy-to-solution**

Results

LAPESD

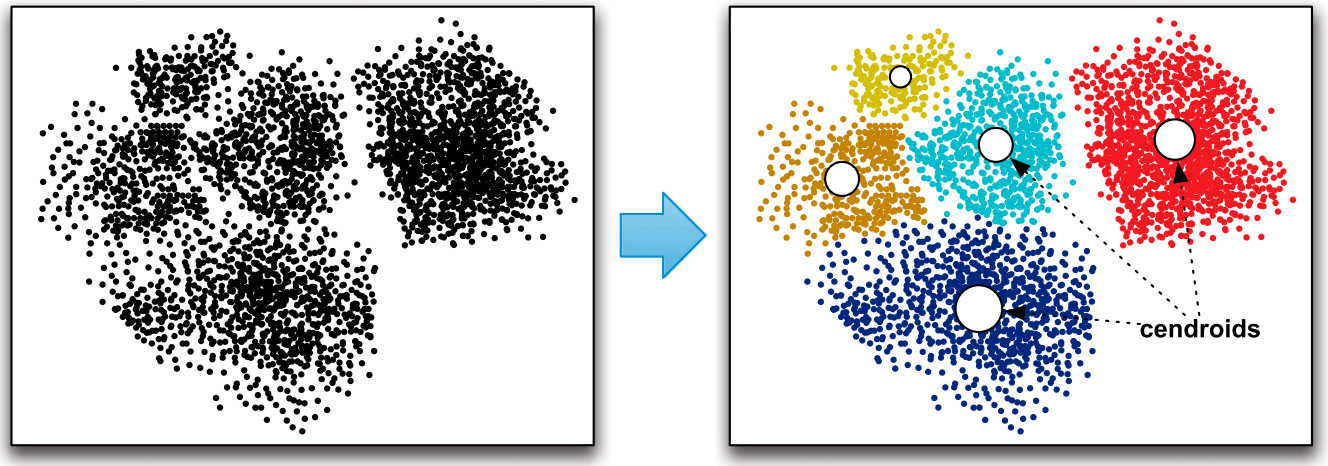
UFSC

Processors

- **General-purpose Processors**
  - **Intel Xeon E5**
    - 8-core Intel Xeon E5 at 2.4 GHz
  - **SGI Altix UV 2000**
    - NUMA with 24 8-core Intel Xeon E5 nodes (192 cores)
- **Accelerators**
  - **GPU NVIDIA**
    - Tesla K20 (2496 cores, 758 MHz)
  - **Intel Xeon Phi**
    - 57 cores (*4-way multithreaded*), 1.10 GHz
- **Embedded Processors**
  - **Carma:** NVIDIA Tegra 3, 1.3 GHz
  - **Exynos 5:** ARM Cortex-A15 com 4 cores, 1.6 GHz

Results

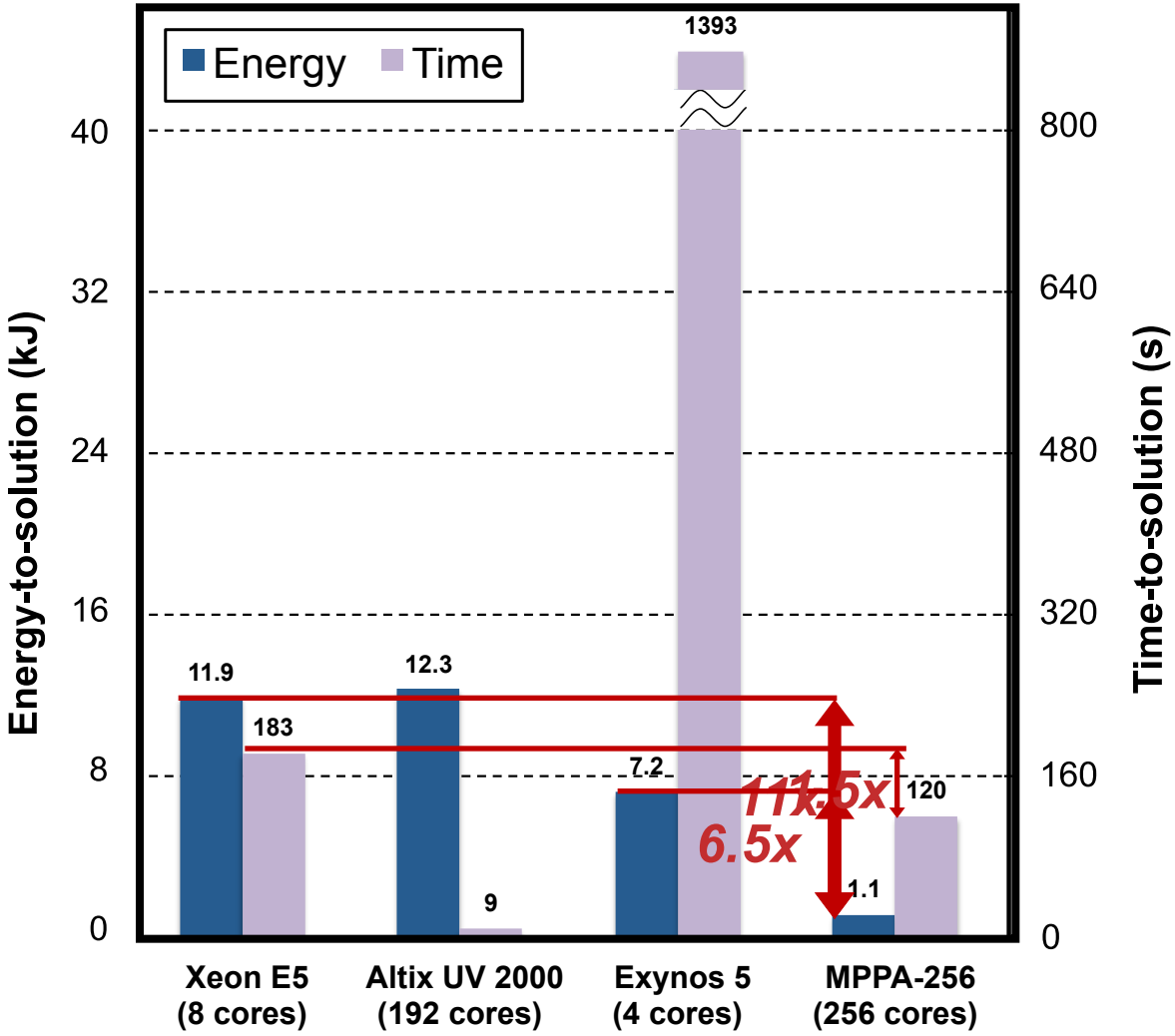
- **K-Means Clustering**
  - Given a set of *n* points in a real *d*-dimensional space, the problem is to partition these *n* points into *k* partitions, so as to minimize the mean squared distance from each point to the center of the partition (centroid) it belongs to
  - **Mixed workload (CPU/Memory)**



Original data

Clustered data

K-Means

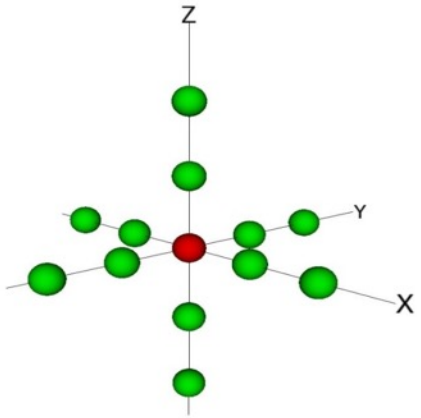


Emilio Francesquini, *et al*. **On the Energy Efficiency and Performance of Irregular Application Executions on Multicore, NUMA and Manycore Platforms**. In: International Journal of Parallel and Distributed Computing (JPDC), 2015.

Results

- **Seismic Wave Propagation**
  - Finite-differences method is used for solving the wave propagation problem

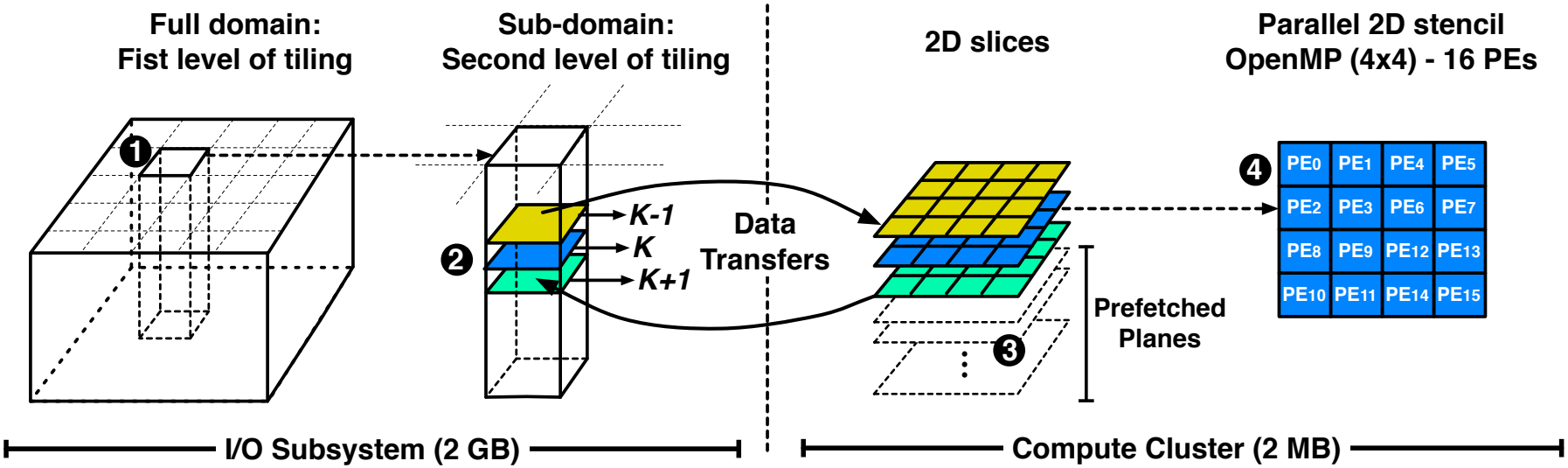
```
for x ← 1 to x_dimension
  do { for y ← 1 to y_dimension
       do { for z ← 1 to z_dimension
            do { COMPUTE_VELOCITY() }
for x ← 1 to x_dimension
  do { for y ← 1 to y_dimension
       do { for z ← 1 to z_dimension
            do { COMPUTE_STRESS() }
```



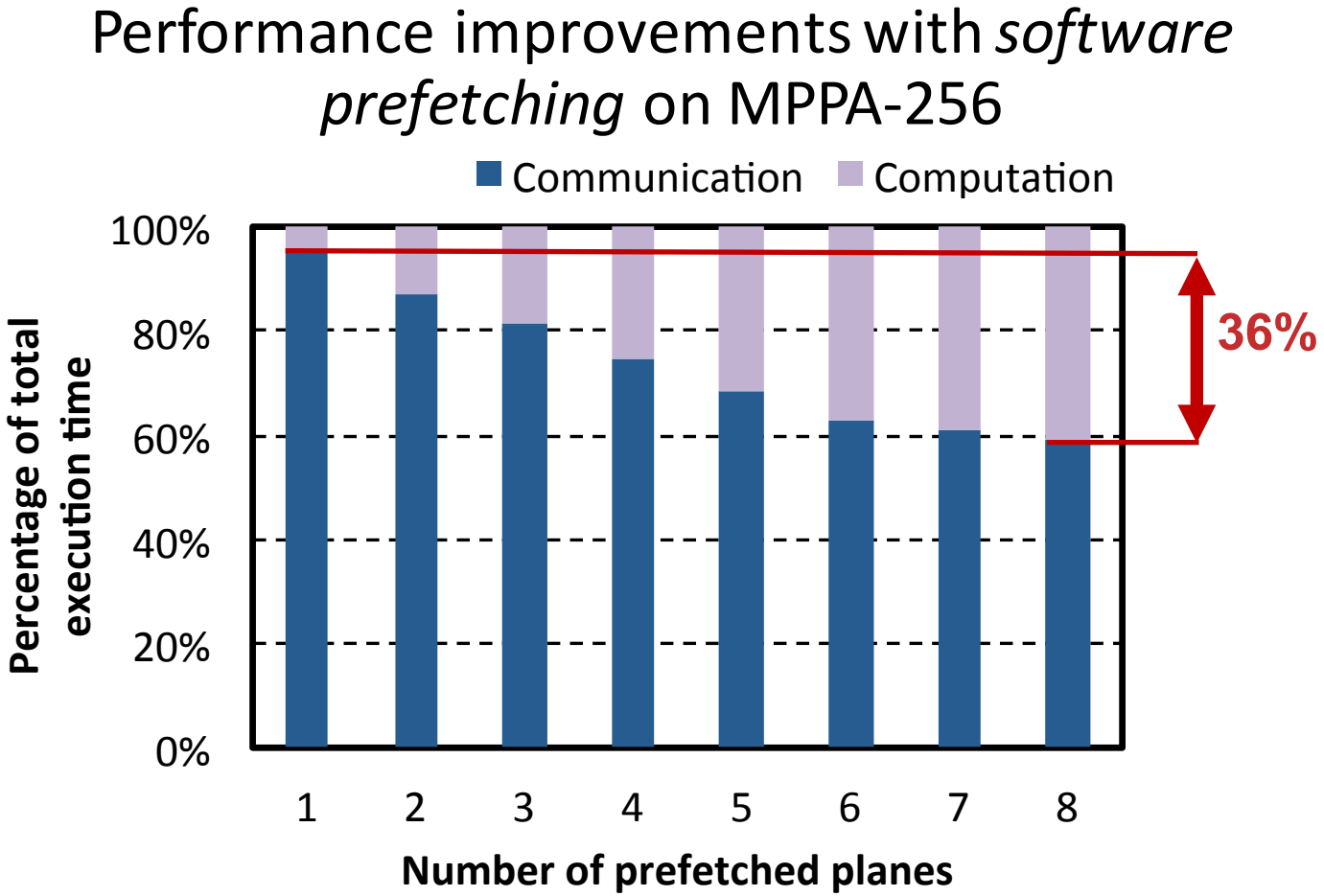
4th-order stencil

Seismic Wave Propagation

- **A two-level tiling scheme** to exploit the memory hierarchy of MPPA-256
- ***Software prefetching*** to overlap data transfers with computations

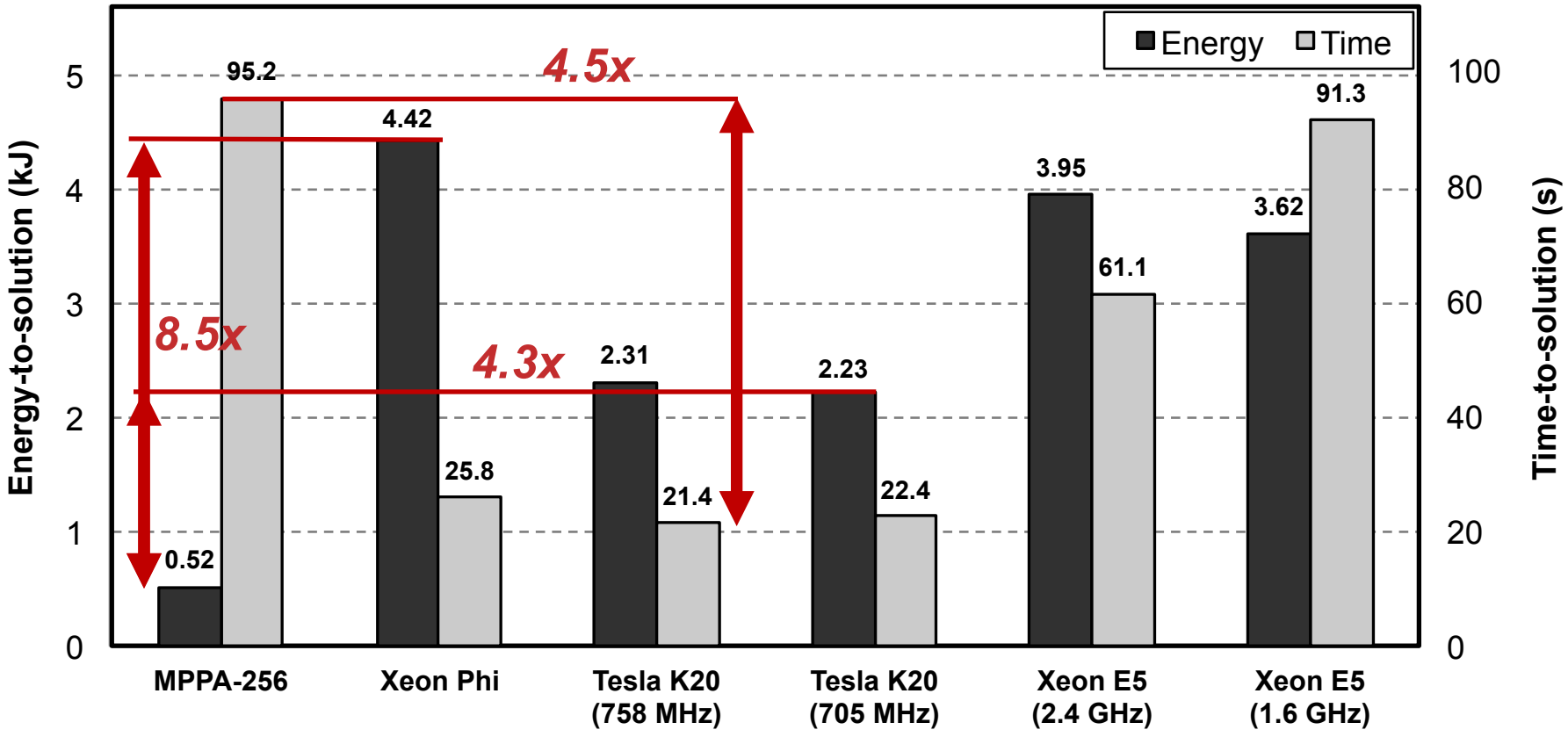


Seismic Wave Propagation



Márcio Castro, *et al*. **Seismic Wave Propagation Simulations on Low-power and Performance-centric Manycores**. Parallel Computing (PARCO), 2016

Seismic Wave Propagation



Márcio Castro, *et al*. **Seismic Wave Propagation Simulations on Low-power and Performance-centric Manycores**. Parallel Computing (PARCO), 2016

Conclusions and Future Works

LAPESD

UFSC

Conclusions and Future Works

- **Low-power manycores**
  - Opportunity to perform highly-parallel energy-efficient computations :-)
  - But... they are very difficult to program :-(
- **Current efforts**
  - High-level programming models for MPPA-256 via *parallel skeletons*

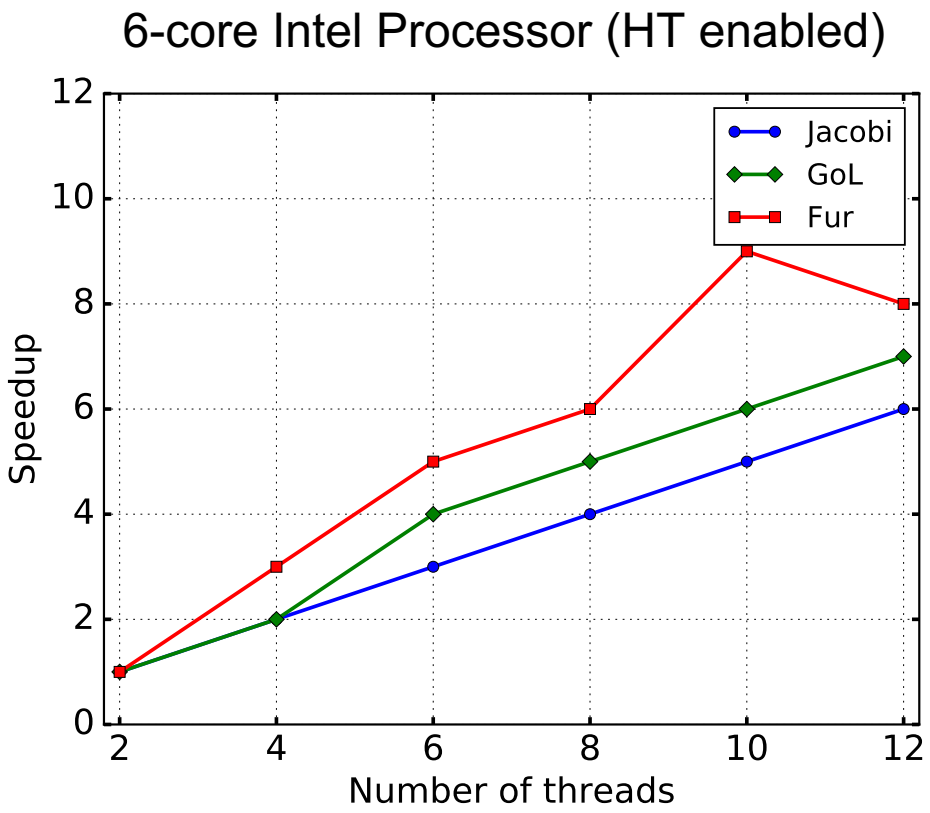
Conclusions and Future Works

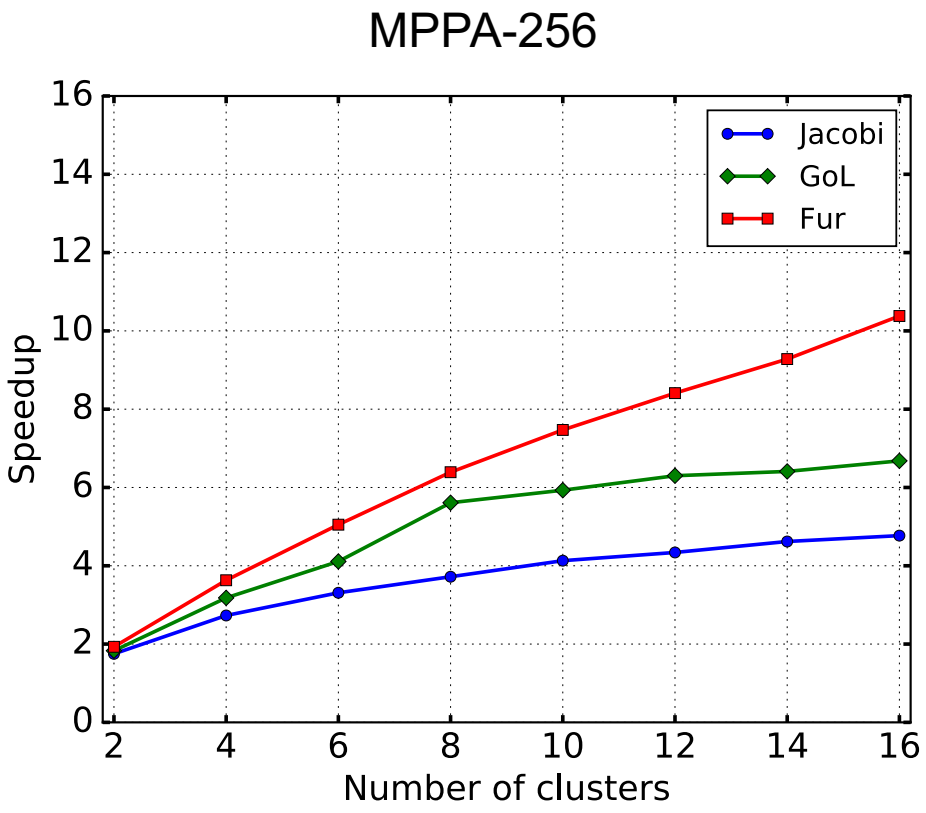
- **PSkelMPPA:** a back-end for the PSkel[1] stencil framework for the MPPA-256
  - Transparent data movements between the I/O subsystem (DDR) and clusters
  - Optimizations can be included at the runtime level
  - All applications implemented with PSkel can run on MPPA-256 without any source code modifications

[1] http://pskel.github.io

Conclusions and Future Works

- **PSkelMPPA: preliminary results**





Questions?