HERCULES

High-Performance Real-time Architectures for Low-Power Embedded Systems

INTRODUCING HERCULES

MARKO BERTOGNA
UNIVERSITY OF MODENA, ITALY

High-Performance Real-Time Lab

H2020, ICT-04-2015, G.A.:688860

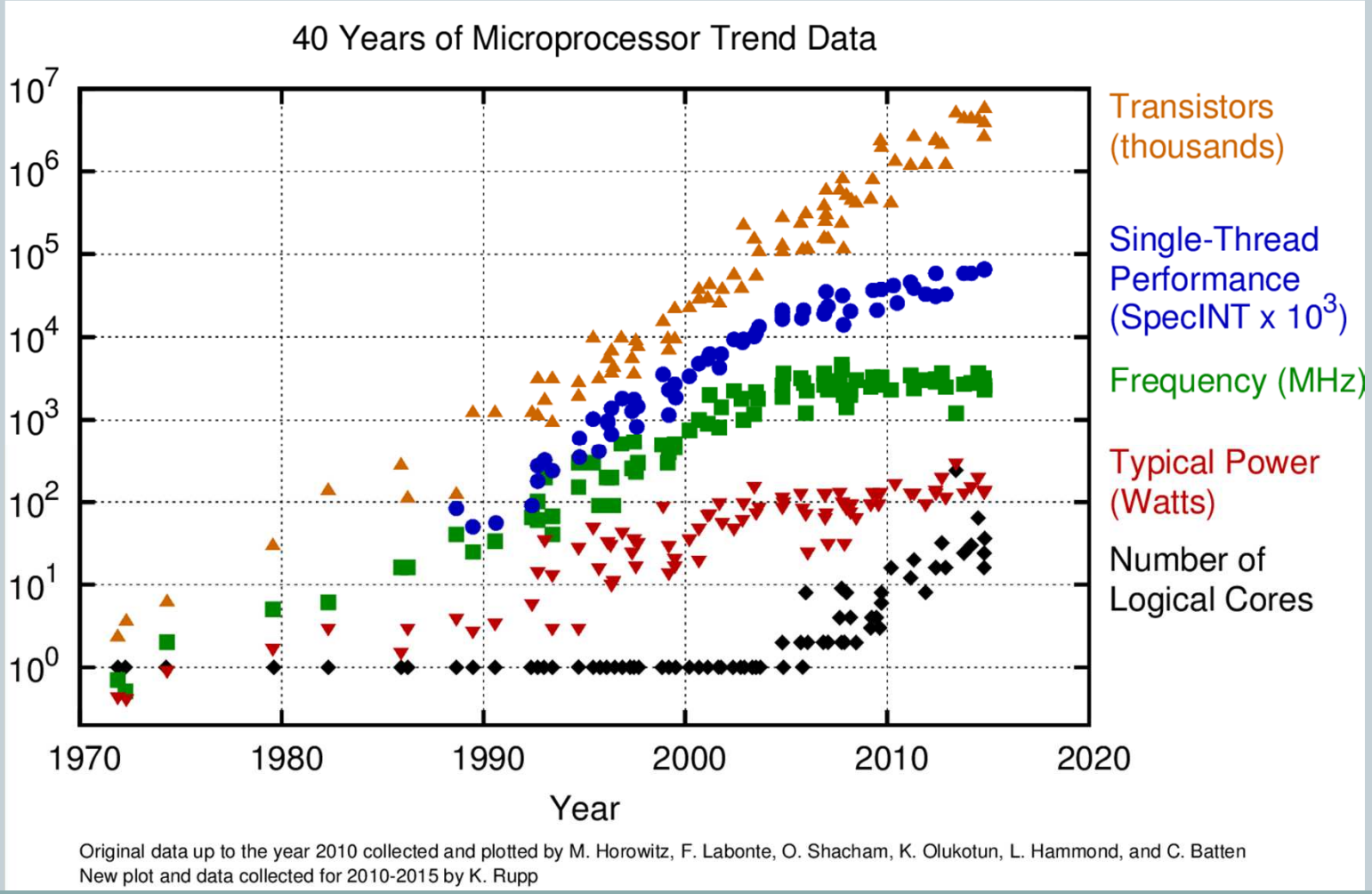
Overview

- Partners

1 (Coordinator)	University of Modena	UNIMORE	Italy
2	Czech Technical University in Prague	CTU	Czech Republic
3	ETH Zurich	ETHZ	Switzerland
4	Evidence Srl	EVI	Italy
5	Pitom snc	PIT	Italy
6	Airbus Gmbh	AB	Germany
7	Magneti Marelli	MM	Italy

- Timespan
  - January 2016 – December 2018
- Budget: ~3.3 M
  - 2.1M EU, 700k Switzerland, 500k industrial co-funding

Technological trend



Observations

- You will be using multi/many-core systems
- Performance will keep growing only for properly designed parallel applications
- Mastering parallelism is not so easy
- Achieving a predictable behavior is harder
  - Parallel concurrency: inter-core dependencies
  - Multiple contention sources: bus, caches, memory, I/O, etc.
- Existing solutions either sacrifice performance (overprovisioning) or predictability

Applications Trend

- New applications requiring a prompt interaction with the environment
- Replace human activities
  - Driving, flying, sailing, farming, tracking, manufacturing, building, checking, testing, etc.
- Higher workload
  - E.g., from multiple cameras and sensing devices
  - Require parallel computing platforms/accelerators
- Real-time guarantees
  - What if a self-driving car “misses” a deadline?
- Higher criticality/safety requirements

HERCULES target

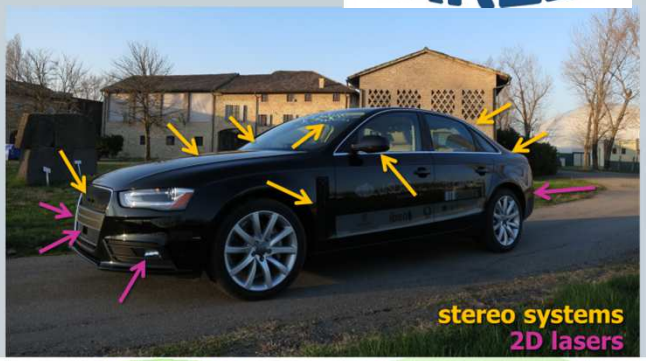
- Real-Time Embedded Super-Computing Platforms
- Integrated framework to achieve predictable performance on top of cutting-edge heterogeneous COTS multi-core platforms
- Technological baseline
  - Real-time scheduling techniques and execution models recently proposed in the research community
  - High-performance/Low-power embedded COTS platforms
  - Next generation real-time applications

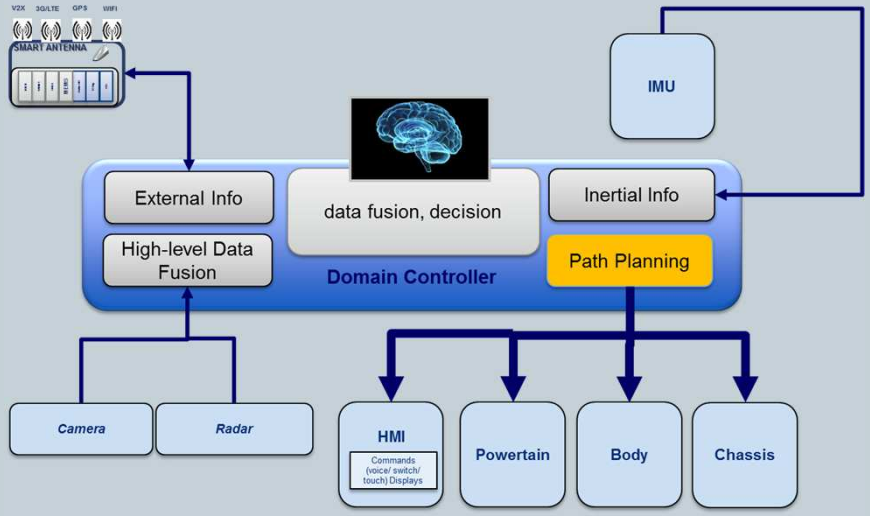
Main Goals

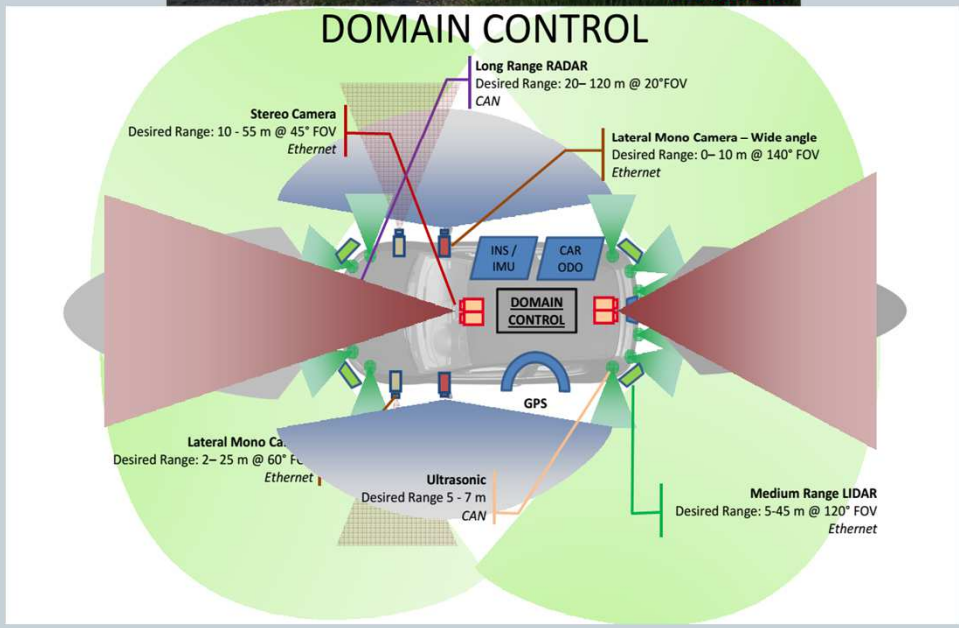
- Goal G1
  - Demonstrate and implement the first **industrial-grade framework** to provide **real-time guarantees** on top of **cutting-edge heterogeneous COTS platforms** for the embedded domain
- Goal G2
  - Obtain an **order-of-magnitude improvement** in the **energy efficiency** and **cost** of **next generation real-time systems**
- Goal G3
  - Provide a **homogeneous programming interface** to simplify the development of future real-time application on top of heterogeneous COTS platforms

Use Case 1: Autonomous Driving

- Domain controller
  - Multi-sensory data fusion
  - Situation awareness
  - Trajectory planning

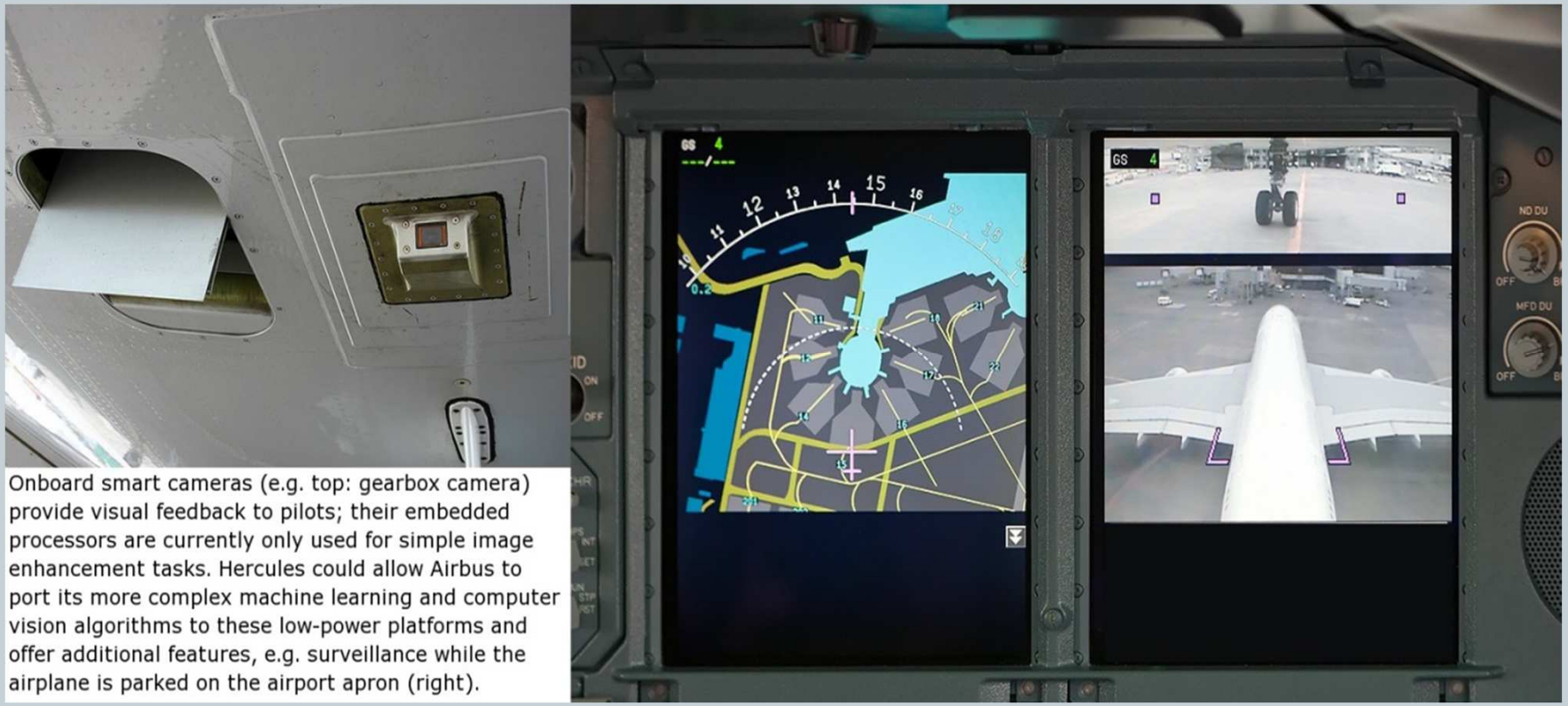






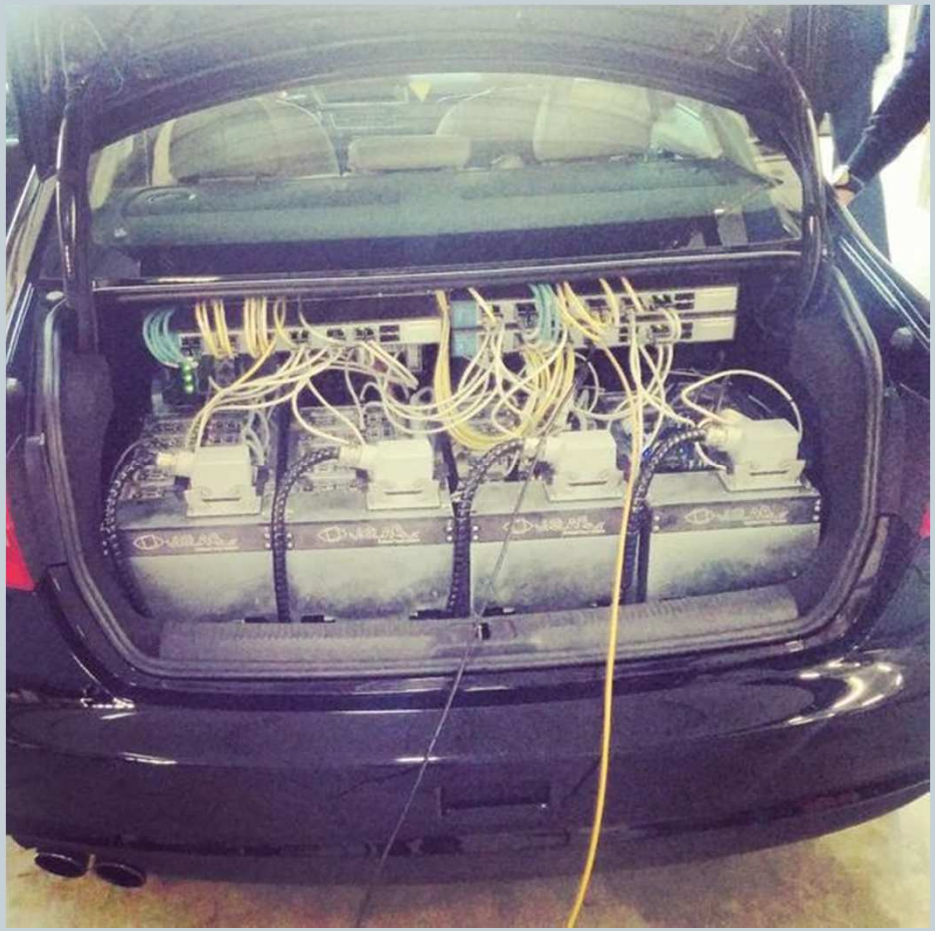
Use Case 2: Avionics

- Machine vision
  - Online computer learning for object detection and tracking

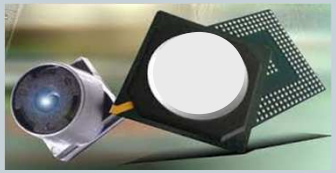


Onboard smart cameras (e.g. top: gearbox camera) provide visual feedback to pilots; their embedded processors are currently only used for simple image enhancement tasks. Hercules could allow Airbus to port its more complex machine learning and computer vision algorithms to these low-power platforms and offer additional features, e.g. surveillance while the airplane is parked on the airport apron (right).

HERCULES at a glance

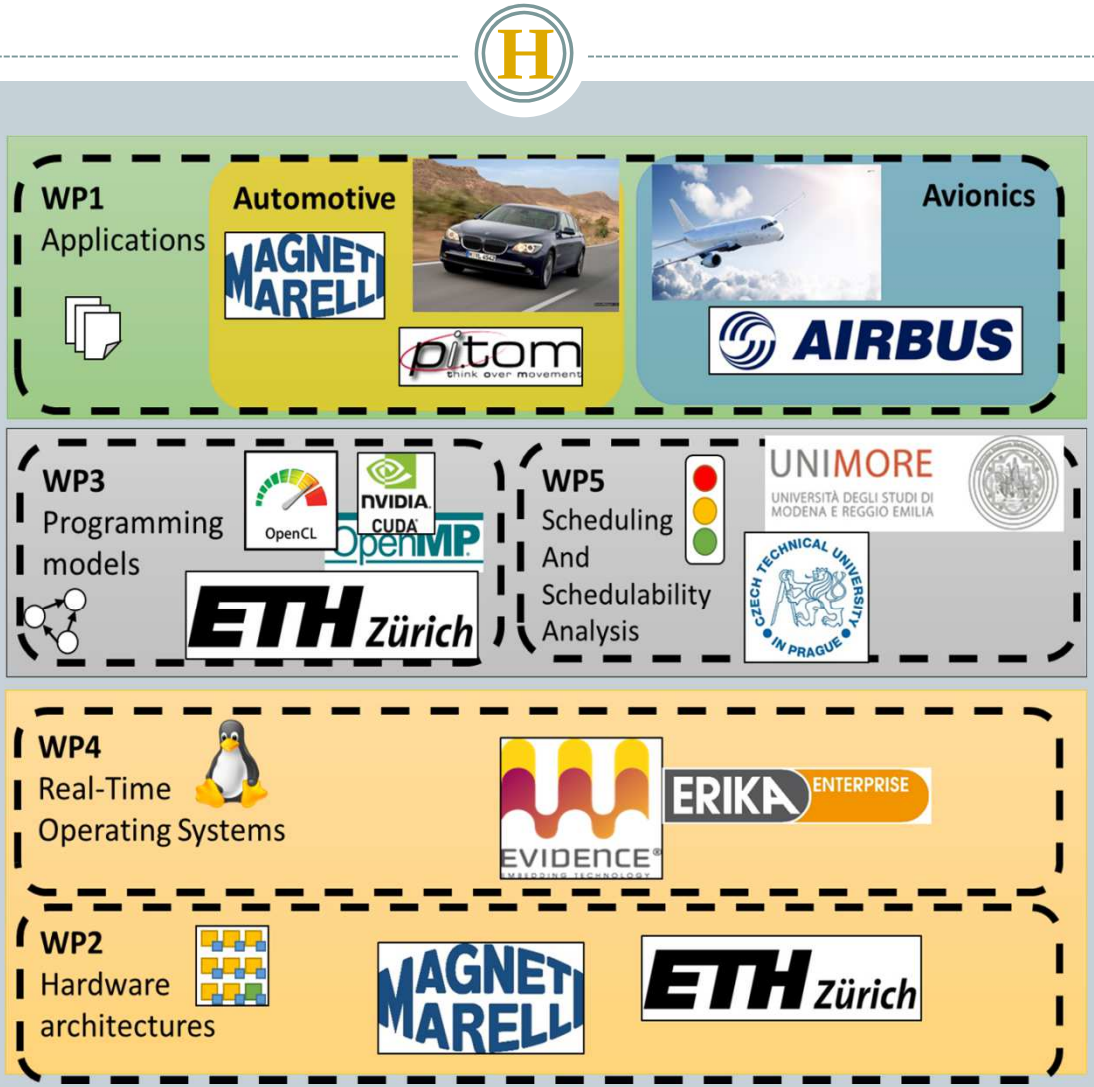


- Cost
- Power
- Size



+Isolation
+Predictability
+Programmability
+Safety
+Openness

Who Does What



Ecosystem Building

- Industrial partners customer base and supply chain
  - Magneti Marelli, Airbus, Evidence, Pitom
- Academic dissemination
  - UNIMORE, ETH, CTU + Real-Time research community
- Software development
  - Open source community, ERIKA, Linux, Jailhouse, etc.
- Industrial Advisory Board
  - Automotive: BMW, Porsche, Continental, Autoliv, Codeplay, …
  - Avionics: Finmeccanica, Selex ES, Honeywell, MBDA, …
  - Farming, Construction: Topcon, Yanmar, …
  - Industrial Automation: SACMI, IMA, Tetra Pak, Datalogic, …
  - Multi/many-core IP: Nvidia, ARM, …
  - General audience: Tom's Hardware

Hardware Platform

- Multi-core host + accelerator(s)
  - ARM big.LITTLE or similar power-efficient multi-core host
  - GPU, DSP cluster, many-core fabric or FPGA acceleration
- Two representative platforms selected @ month 6
  - One for each use-case domain (automotive and avionics)
  - Decision based on **predictability**, **programmability**, **performance/cost**, **power efficiency**
  - Cutting edge technologies
    - Nvidia Tegra X1/Parker architecture (20/16nm), Xilinx Zynq Ultrascale (16nm), Renesas R-Car H3 (16nm), Samsung Exynos 7 Octa (14nm), Qualcomm Snapdragon 820 (16nm), Intel 5th gen Core (14nm), Kalray MPPA (28 nm), TI KeystoneII (28nm)

Software Platform

- Predictable Hypervisor
- Lightweight RTOS
  - ERIKA Kernel on LITTLE cores
- Linux with real-time patch
  - `sched_deadline`
- Lightweight OpenMP runtime
- Predictable host-to-accelerator offloading routines
- Predictable execution models
- Compiler support

Conclusions

- HERCULES will provide a **software framework** to simplify the development of **next-generation real-time applications** on **heterogeneous COTS platforms**
  - TRL 5-6: *validation in representative environments*
- Multiple targets:
  - Performance with real-time guarantees
  - Low power/Low cost
- Mostly open-source
  - Linux, ERIKA, OpenMP
  - May protect some IP for market opportunities

Thank you!

marko.bertogna@unimore.it