The INFN COSA project

Daniele Cesini – INFN-CNAF
(On behalf of the COSA collaboration)

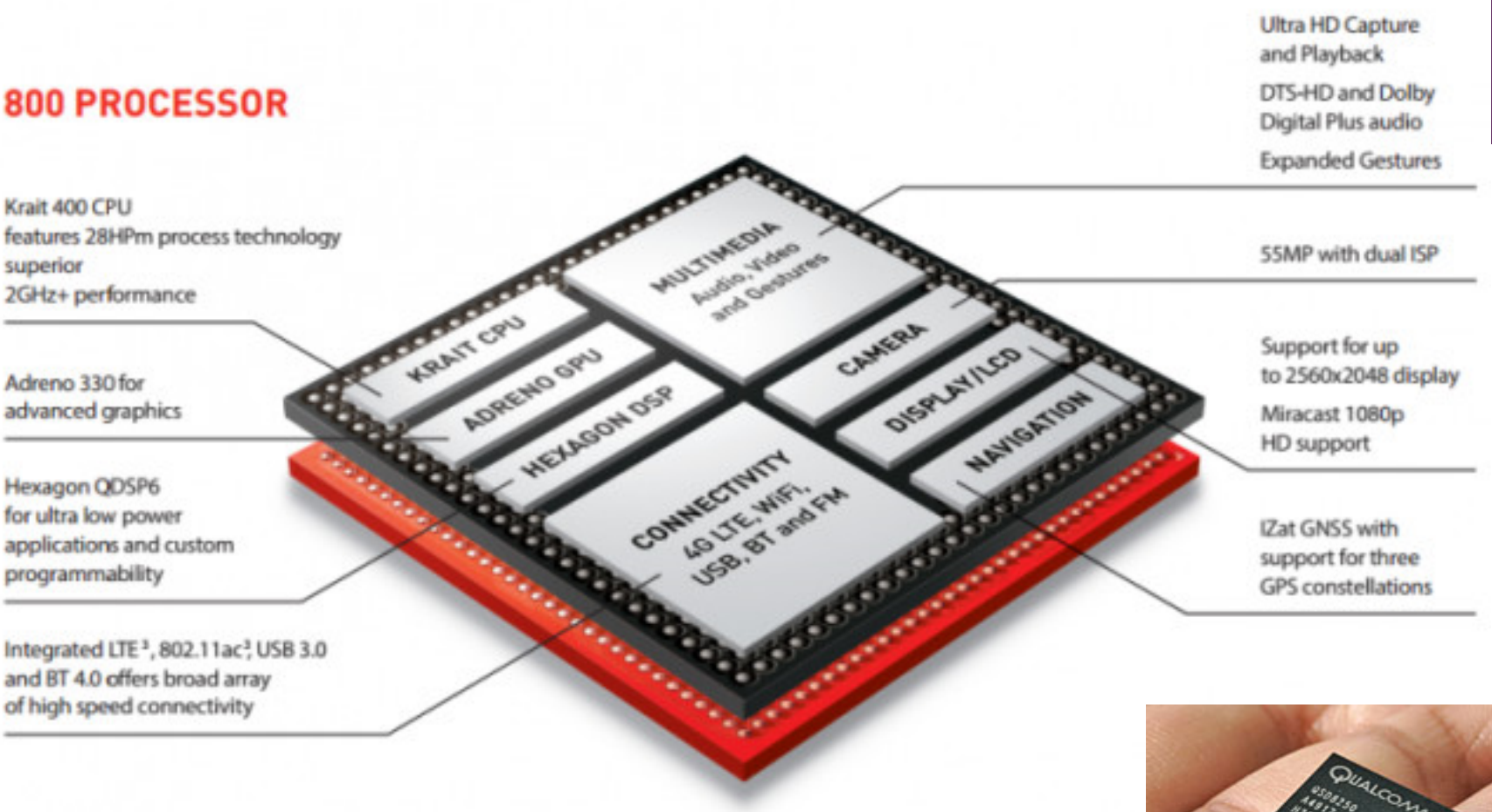
INFN-CSN5 COSA project

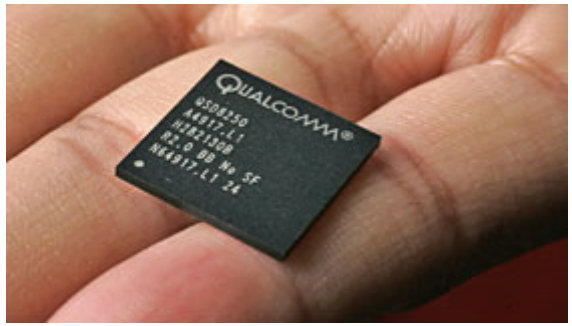
- COSA: Computing On SOC Architecture
- Duration: 2 years from January 2015
- Departments: 7 INFN
  - CNAF, PI, PD, ROMA1, FE, PR, LNL
- BUDGET :51.5 kEuro Year I

Objectives

- Acquire know-how
  - Porting e benchmarking of low power System on Chip
  - Operations of Linux system on SoCs
  - Benchmarking hybrid architectures
- Unification of INFN HW testing activities
  - Continuation of the COKA project
    - Porting on traditional accelerator (GPU/MIC)
  - Continuation of the HEPMARK projects
    - X86 benchmarking
- Study of dedicated low latency interconnection with ARM+FPGA devices
- Prepare H2020 proposal on LowPower computing calls

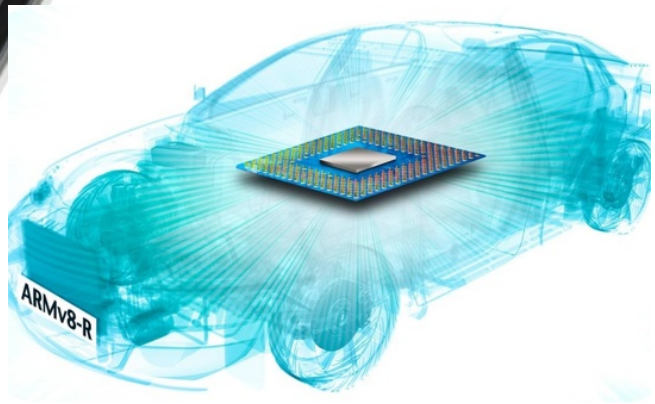
Low-Power System on Chip (SoCs)



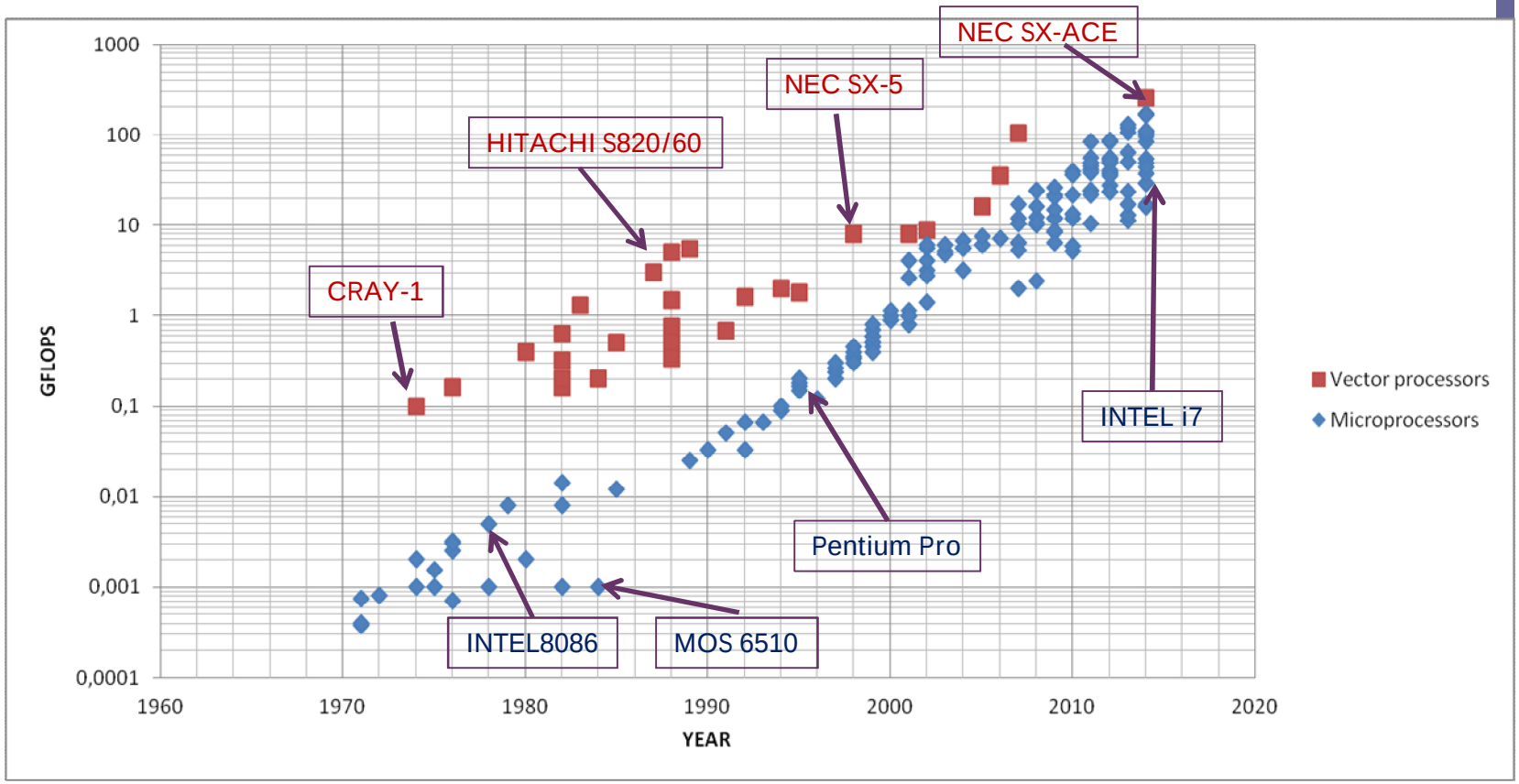


Where do I find a SoC?

- Mobile
- Embedded

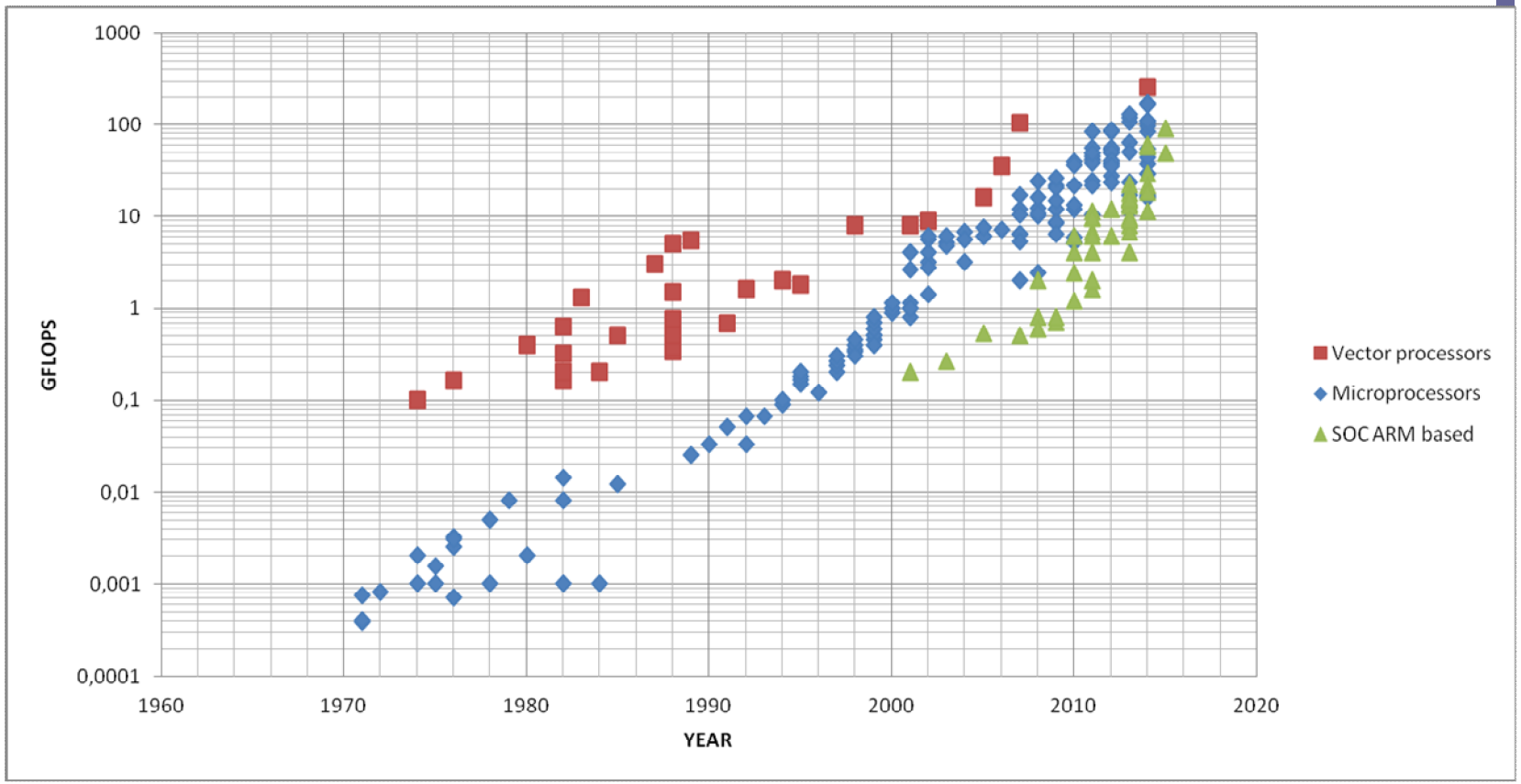


Vector vs Micro computing power



- Why did microprocessors take over?
  - They have never been more powerful…
  - ….but they were cheaper, highly available and less power demanding

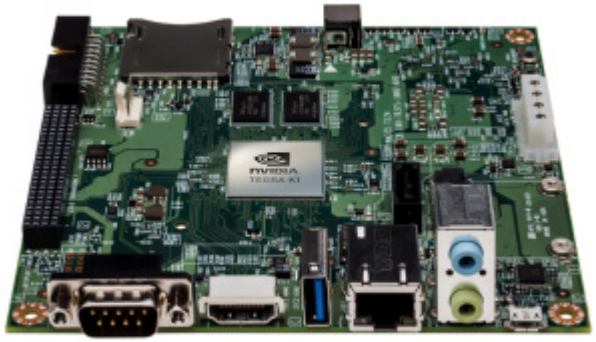
Vector vs Micro vs ARM based



- Is history repeating?

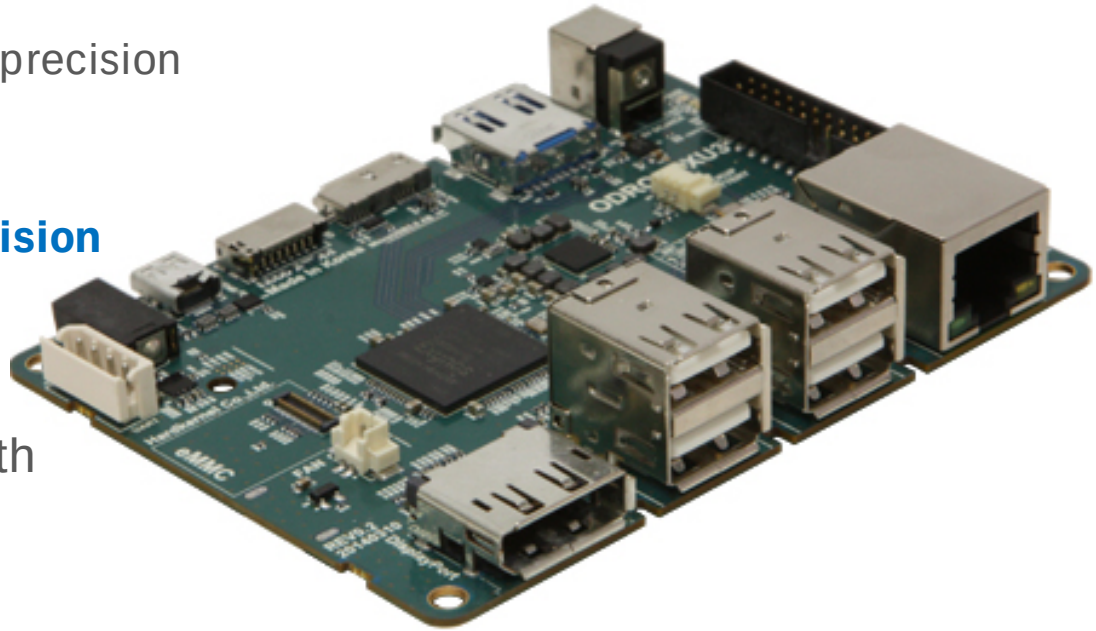
Ok, but then….an iPhone cluster?

- NO, we are not thinking to build an iPhone cluster
- We want to use these processors in a standard computing center configuration
  - Rack mounted
  - Linux powered
  - Running scientific application mostly in a batch environment
- ….. Use development board…



ODROID-XU3

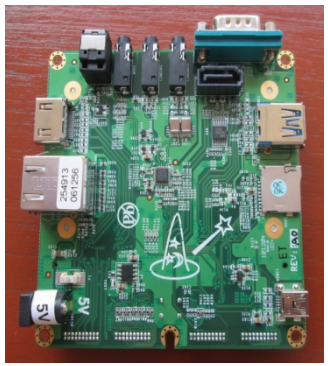
- Powered by ARM® big.LITTLE™ technology, with a **Heterogeneous Multi-Processing (HMP)** solution
  - 4 core ARM A15 + 4 cores ARM A7
- Exynos 5422 by Samsung
  - ~ 20 GFLOPS peak (32bit) single precision
- **Mali-T628 MP6 GPU**
  - **~ 110 GFLOPS peak single precision**
- 2 GB RAM
- 2xUSB3.0, 2xUSB2.0, 1x107100 eth
- Ubuntu 14.4
- HDMI 1.4 port
- 64 GB flash storage



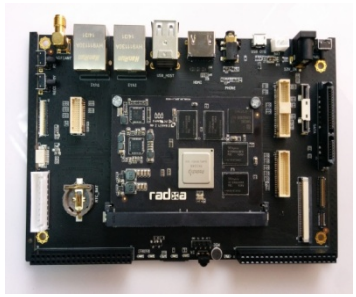
Power consumption max ~ 15 W

Costs 150 euro!

Other nice boards



WandBoard

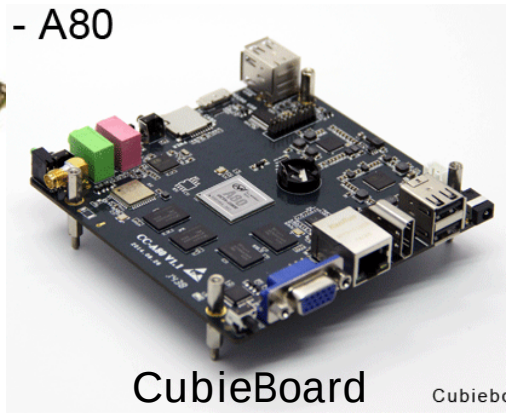


Rock2Board

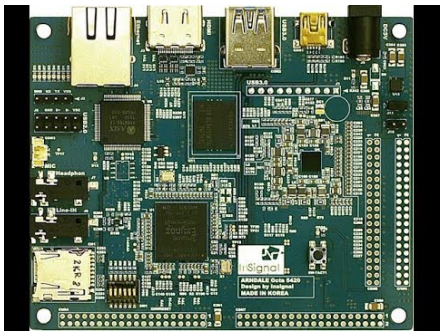
PandaBoard

DragonBoard

SabreBoard



CubieBoard



Arndale OCTA Board

Texas Instruments EVMK2H

http://elinux.org/Development_Platforms

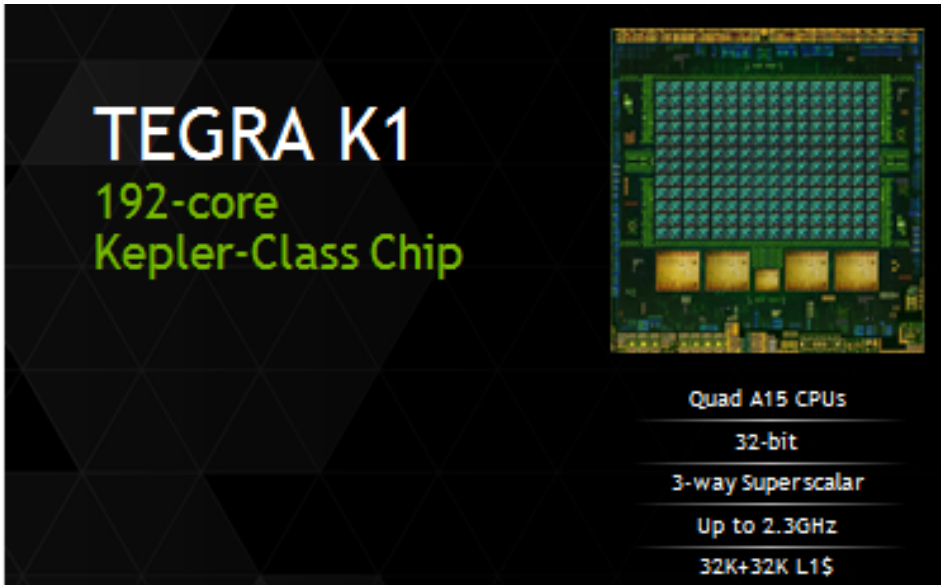
- ...and counting...

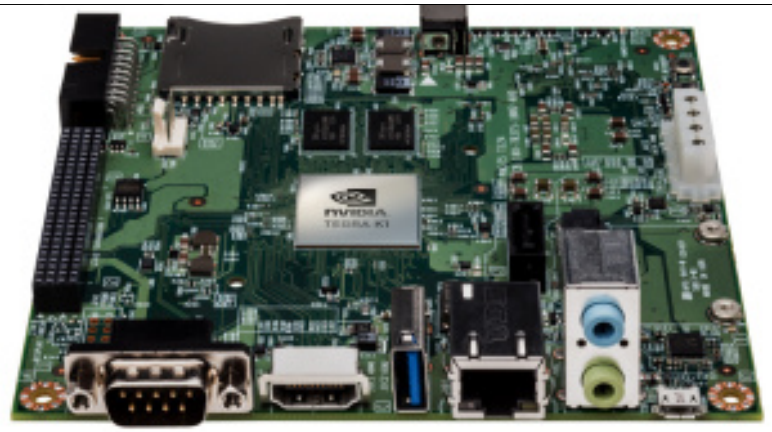
Some specs

BOARD	SOC				GFLOPS (CPU+GPU)	Eth
	Model	ARM IP	GPU IP	DSP IP		
FREESCALE (Embedded SoC) SABRE Board	**Freescale** i.MX6Q	**ARM** A9(4)	**Vivante** GC2100 (19.2GFlops)		25	1Gb
ARNDALE (Mobile SoC) Octa Board	**Samsung** Exynos 5420	**ARM** A15(4) A7(4)	**ARM** Mali-T628 MP6 (110Gflops)		115	10/100
HARDKERNEL (Mobile SoC) Odroid-XU-E	**Samsung** Exynos 5410	**ARM** A15(4) A7(4)	**Imagination Technologies** PowerVR SGX544MP3 (51.1 Gflops)		65	10/100
HARDKERNEL (Mobile SoC) Odroid-XU3	**Samsung** Exynos 5422	**ARM** A15(4) A7(4) **(HMP)**	**ARM Mali-T628 MP6 (110 Gflops)**		130	10/100
INTRINSIC (Mobile SoC) DragonBoard	**Qualcomm** Snapdragon 800	**Qualcomm** Krait(4)	**Qualcomm** Adreno 330 (130Gflops)		145	1Gb
TI (Embedded SoC) EVMK2H	**TI Keystone** 66AK2H14	**ARM** A15(2)		**TI** MS320C66x (189Gflops)	210	1Gb (10Gb)

TDP between 5W and 15W
(EVMK2H > 15W)

NVIDIA JETSON TK1





- First **ARM+CUDA programmable SoC based** Linux development board
- 4 cores ARM A15 CPU
- 192 cores NVIDIA GPU
  ➔ 300 GFLOPS (peak sp)
- ~ **21 GFLOPS/W (sp**
- ... for less than 200 Euros
- 32bit
  - 64bit version announced

How do you program them? (in a Linux environment)

- GCC+OpenMP+MPI available for ARM architectures
- OpenCL for the GPU
  - If you are lucky enough to find working drivers
- CUDA available only on the Jetson K1
  - Computing capability 3.2 (vs 3.5)
- Cross compilation
- GCC5+OpenMP4 tests ongoing…

Limitations of the tested boards

- Commodity SoCs and development boards have a number of limitations:
  - 32 bit
  - Small caches
  - Small RAM size in the boards (O(2GB))
    - However modern SoCs can address 40bit
  - No ECC memory
  - Frequent failures and system crashes
  - Slow connections (10/100Mb eth) in many cases
    - Ethernet via USB in same boards
  - HW bugs

COSA Clusters

CNAF

~25 board SoC Based
(GFLOPS nominali di **2 server** tradizionali con **GPU**) — Anno I

\+ ~10 nuove board — Anno II

ex **cluster COKA** (2 server)
+ nuove acquisizioni (1 server) — Anno I

+ nuove acquisizioni (1 server) — Anno II

(**server = cpu + acceleratori**)

ROMA1

4 **board ARM+FPGA** based
+ 1 server — Anno I

16 board ARM+FPGA based
+ 4 server — Anno I

PD

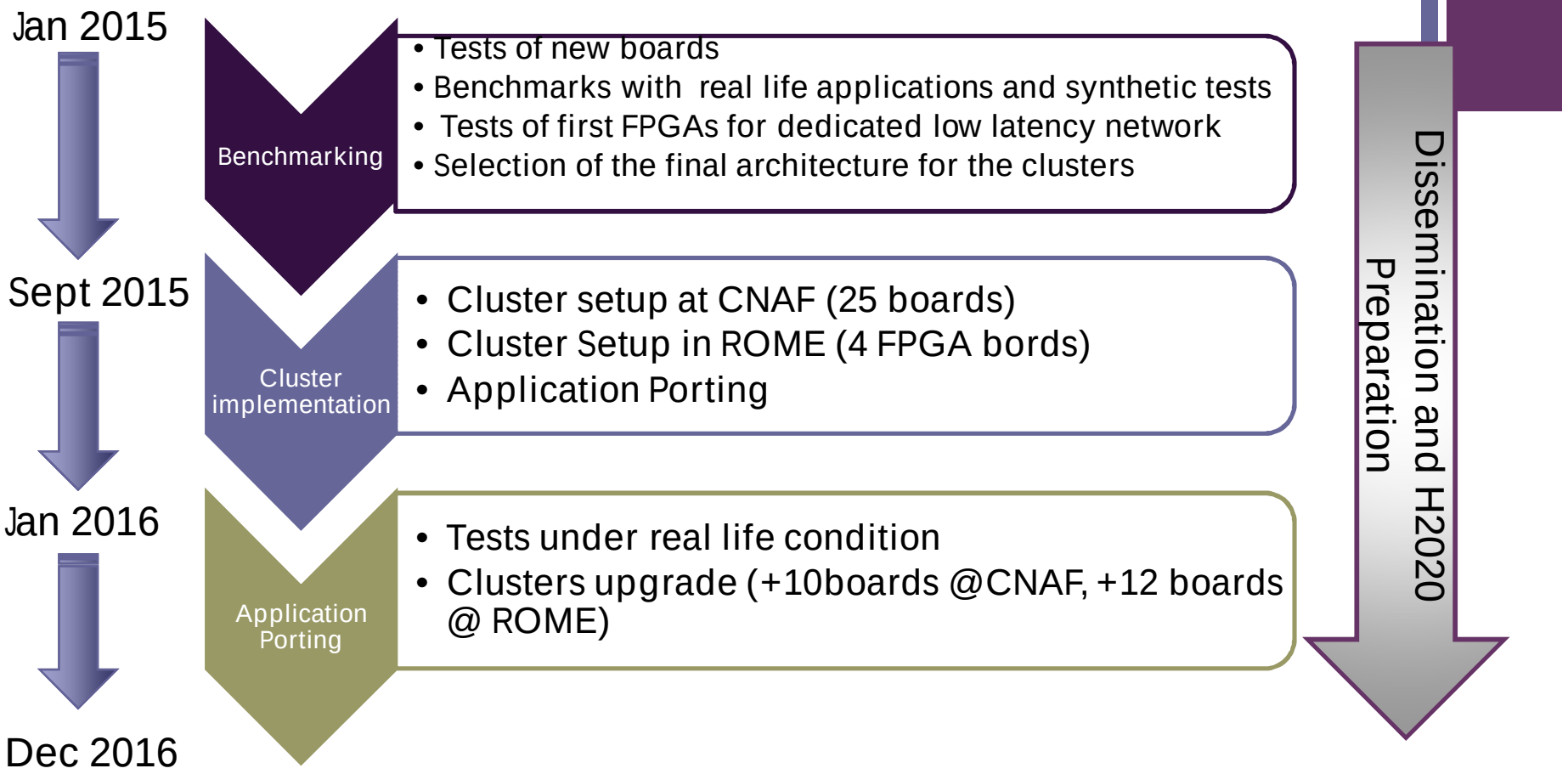
Ex cluster HEPMARK
+ nuove acquisizioni (~ 3 server) — Anno I

+ nuove acquisizione (~ 3 server) — Anno II

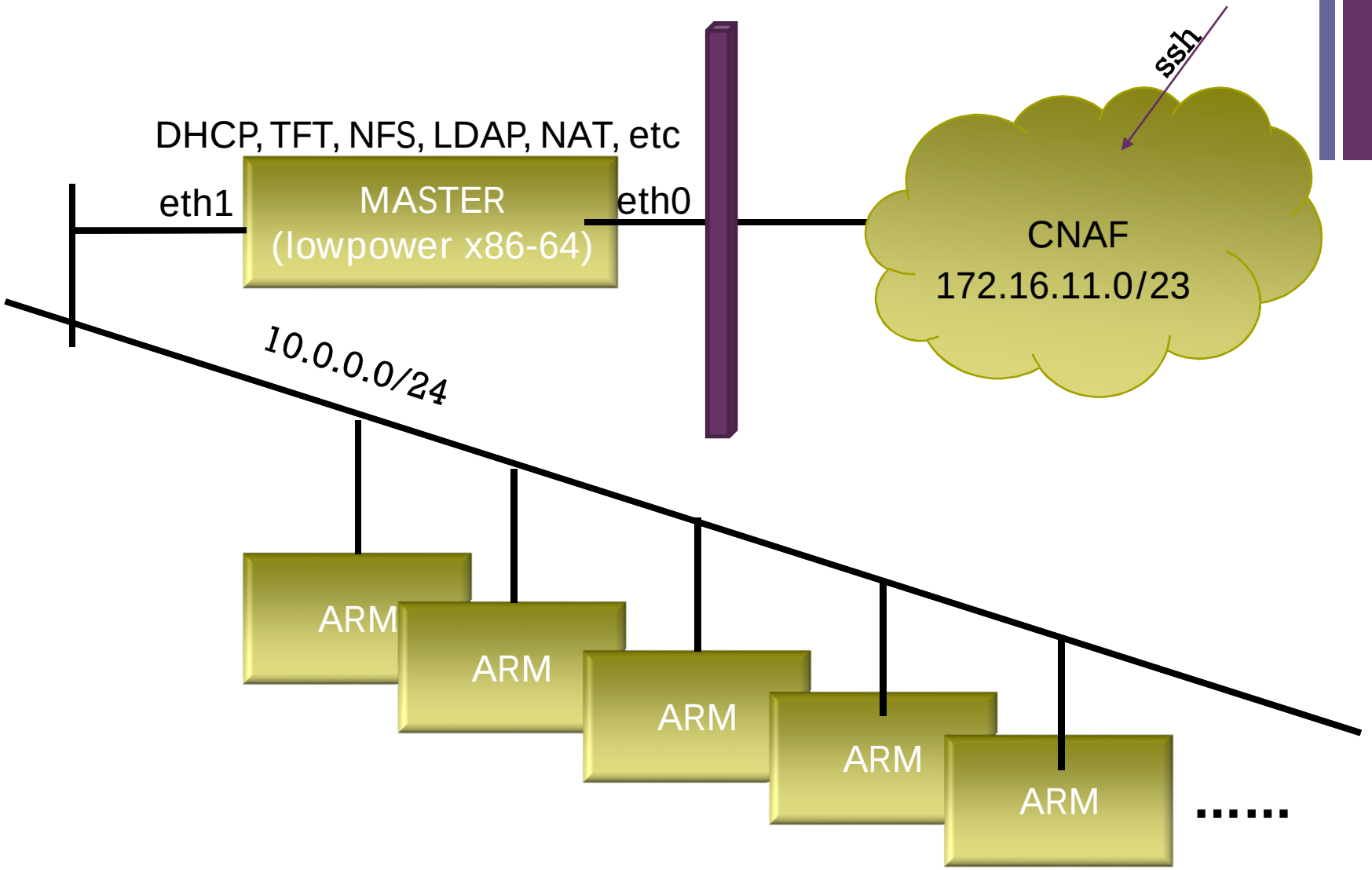
Activities & WPs

- WP1: Management
- WP2: Technology Tracking and Benchmarking
- WP3: Implementation of the CNAF prototype
- WP4: Development of dedicated network connections
- WP5: Application Porting
- WP6: Technology Transfer and Dissemination

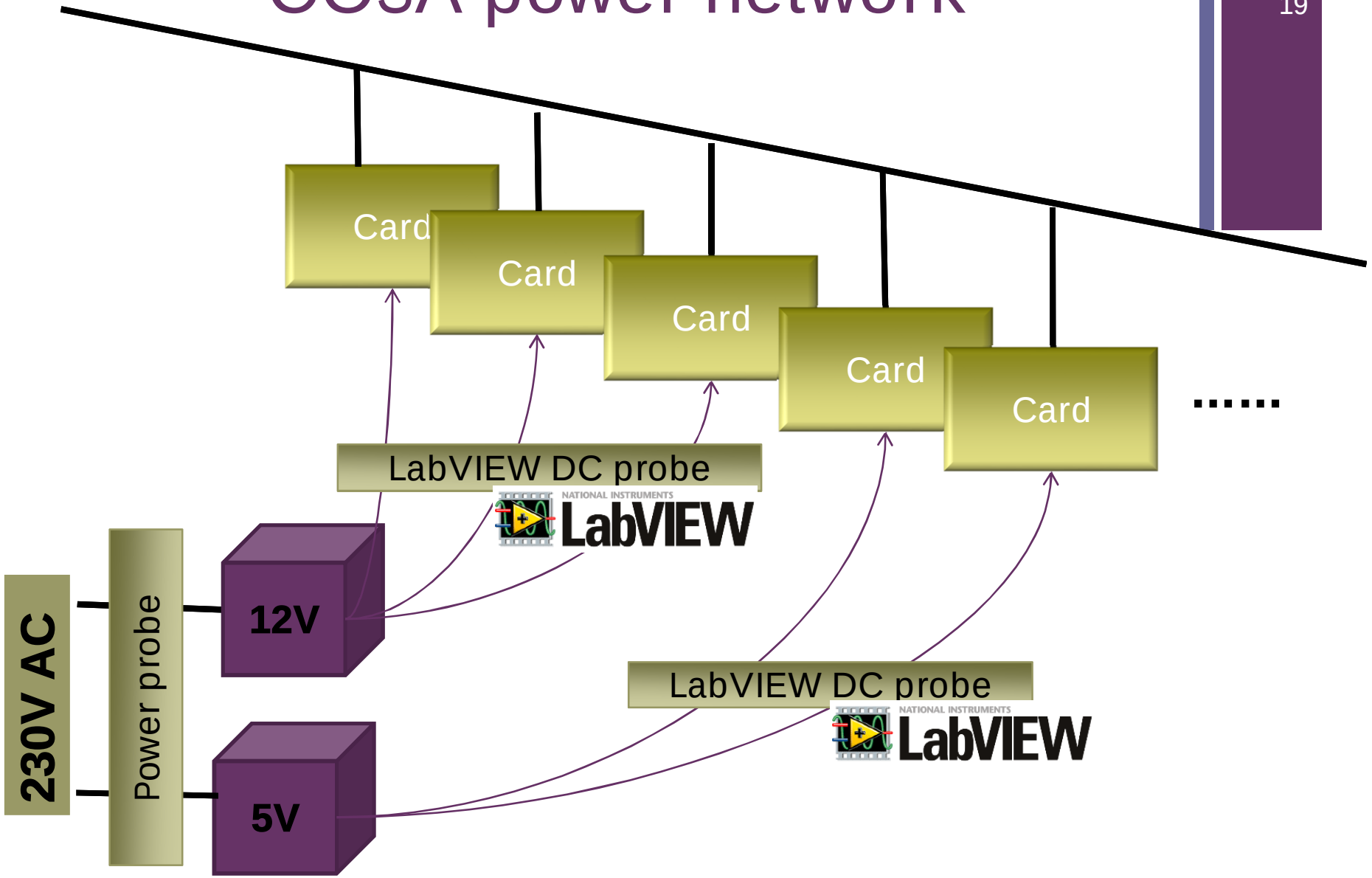
Project Timeline



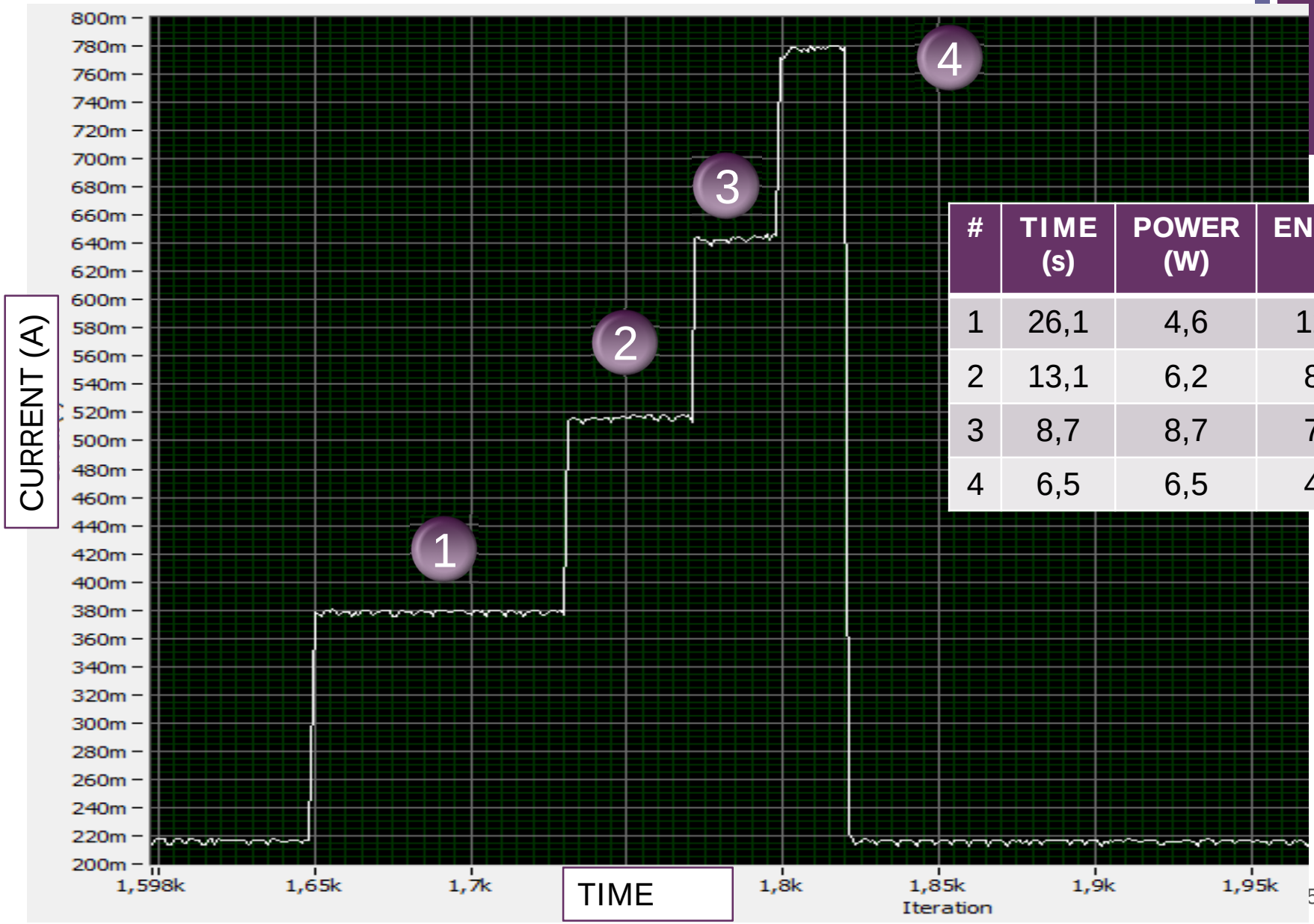
CNAF COSA cluster network



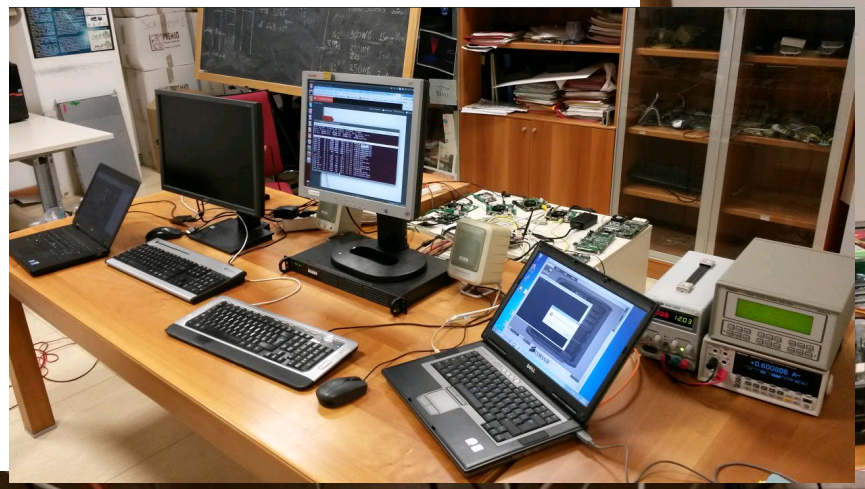
COSA power network

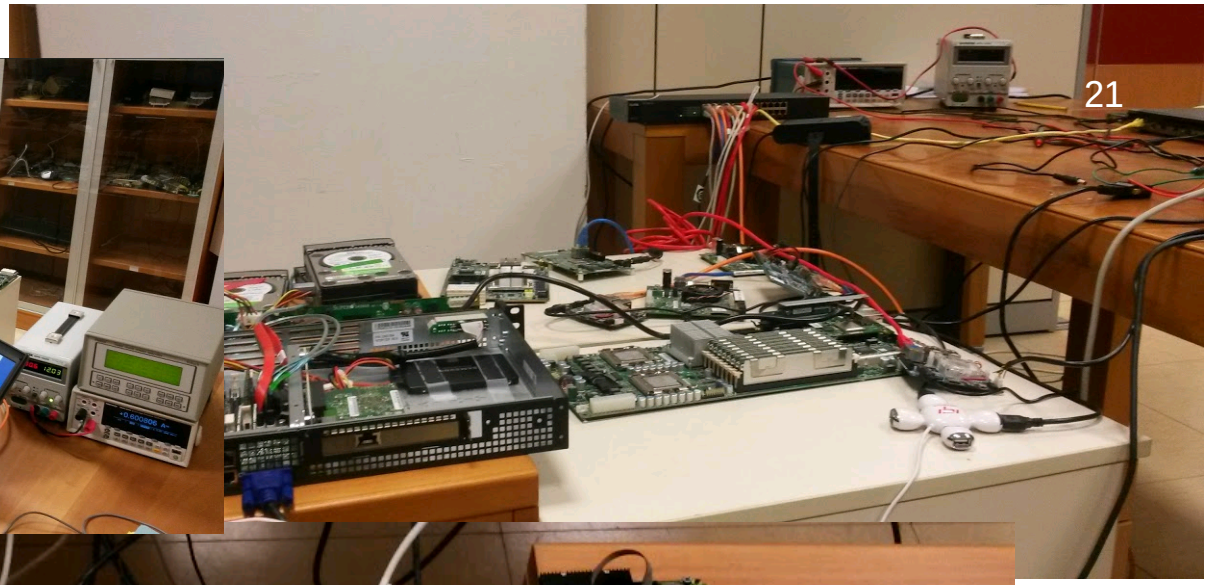


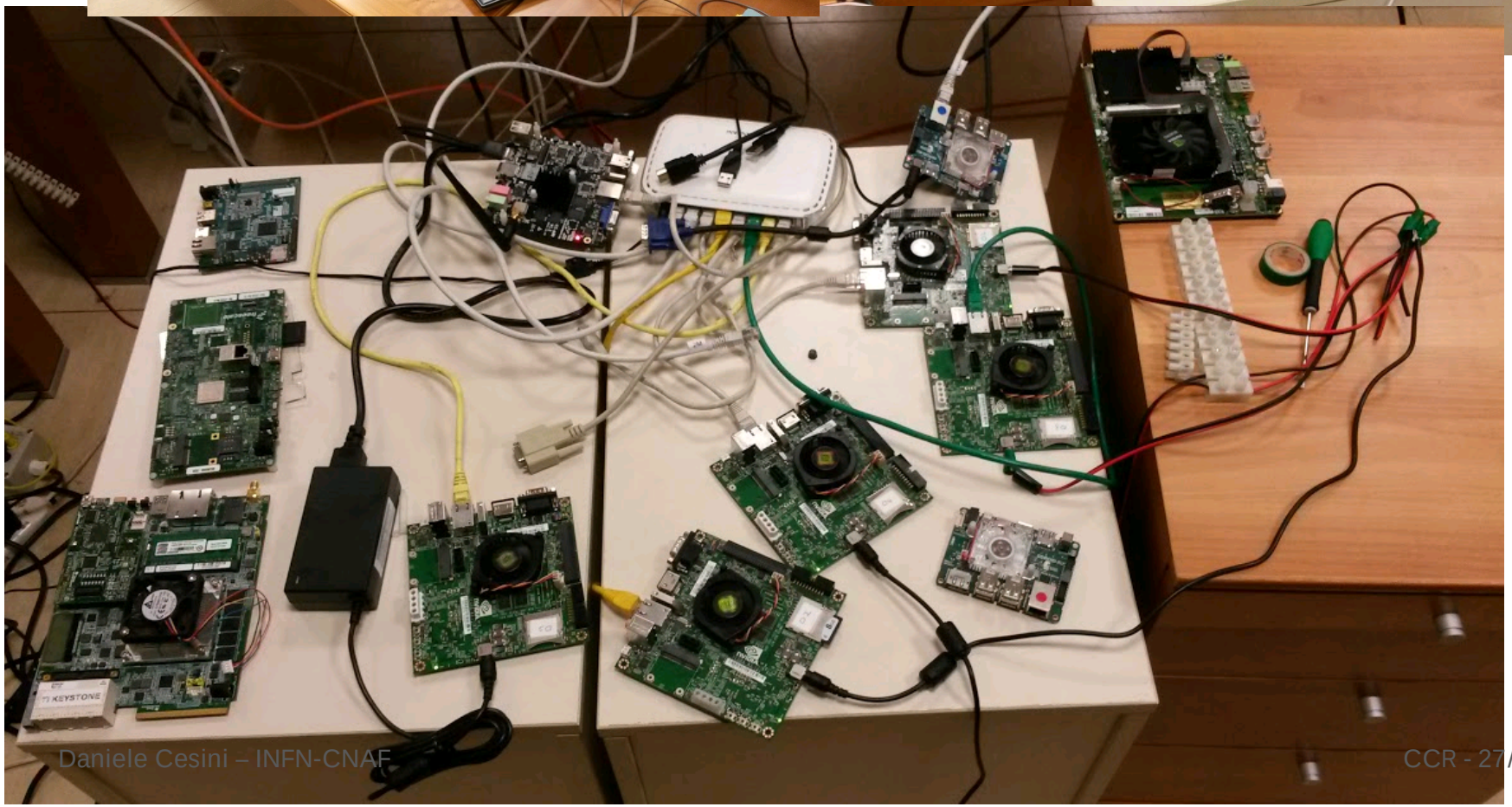
OPENMP π computation on Jetson



#	TIME (s)	POWER (W)	ENERGY (J)
1	26,1	4,6	120,1
2	13,1	6,2	81,2
3	8,7	8,7	75,7
4	6,5	6,5	42,2







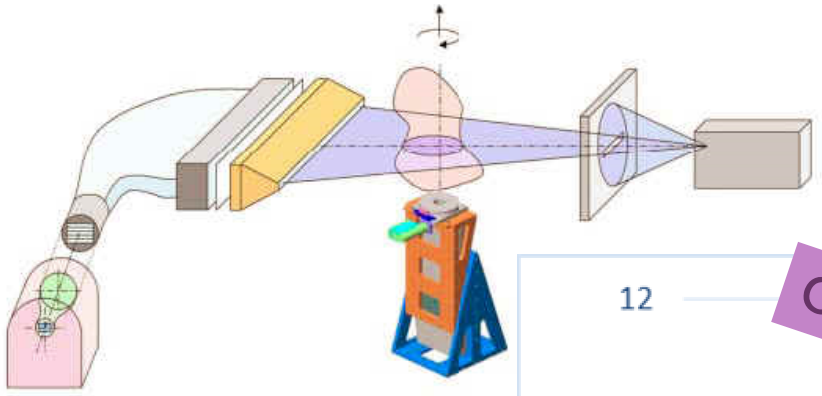
PSU&Cables

- PSU HX1000i
  - 12 linee 12V (Jetson)
  - 6 linee 5V
- Cavi GRIDSEED
  - Da 1 MOLEX a 3 BARREL

Applications

- Theoretical Physics (PR, FE)
  - Parallel applications usually run in HPC environments
    - Lattice Boltzmann fluid dynamics
    - Monte Carlo simulations of Spin-Glass systems
    - Lattice Quantum ChromoDynamics simulations
- Experimental Physics (PI, PD, CNAF)
  - HEP experiments High Level Trigger applications
  - Montecarlo and analysis of LHC experiments
  - Applications needing portable systems
    - Computer tomography
- Neural Networks (RM1)
  - DPSNN-STDP code

Filtered Backprojection (1/2)



(*) in collaboration with the physics department of the Bologna University and CHNET

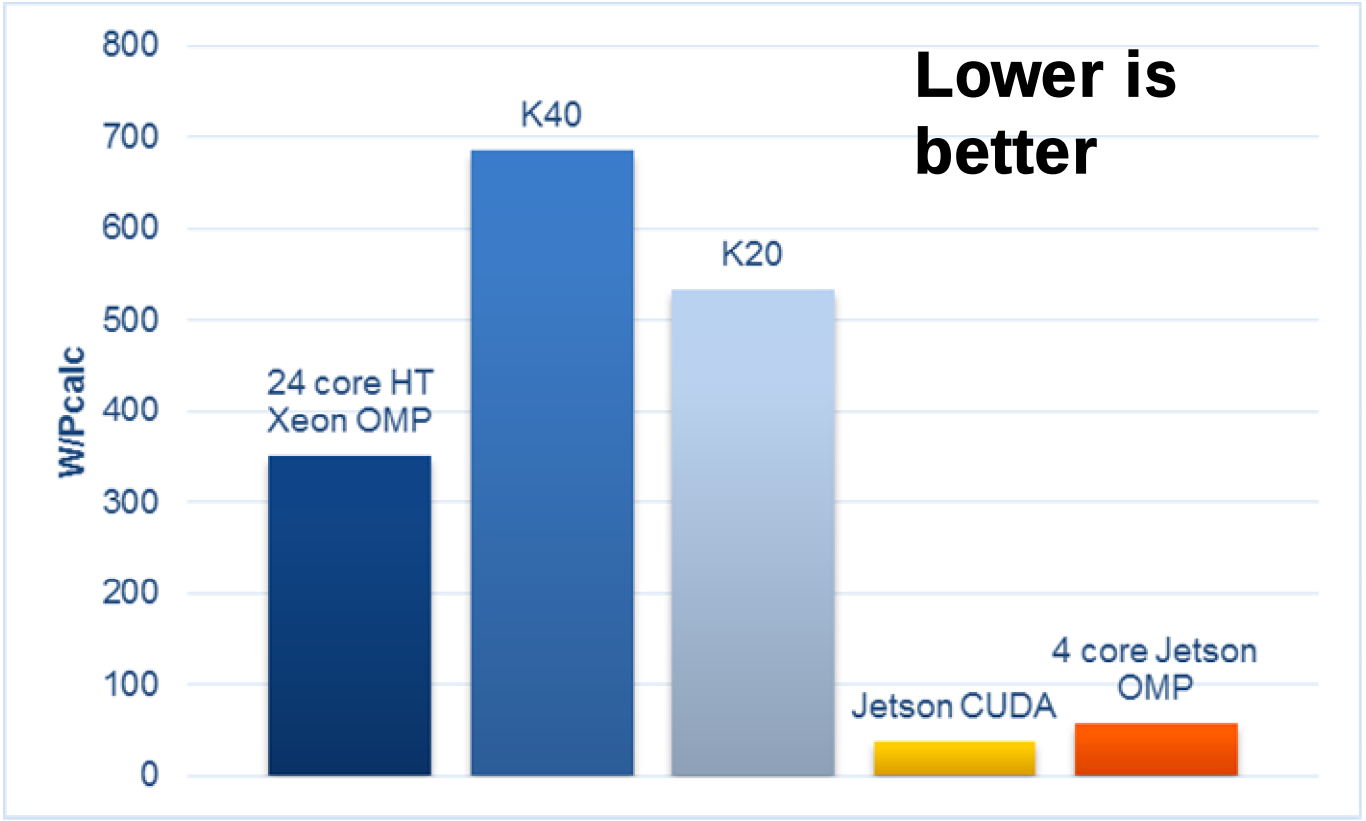
Master thesis of Elena Corni:
Implementazione dell'algoritmo Filtered Back-Projection (FBP) per architetture Low-Power di tipo Systems-On-Chip

Filtered Backprojection (2/2)

On (2xE5-2620+K20): 3241 slices in 1 h: **350** Wh

On **5** x Jetson-K1: 3840 slices in 1 h: **41** Wh

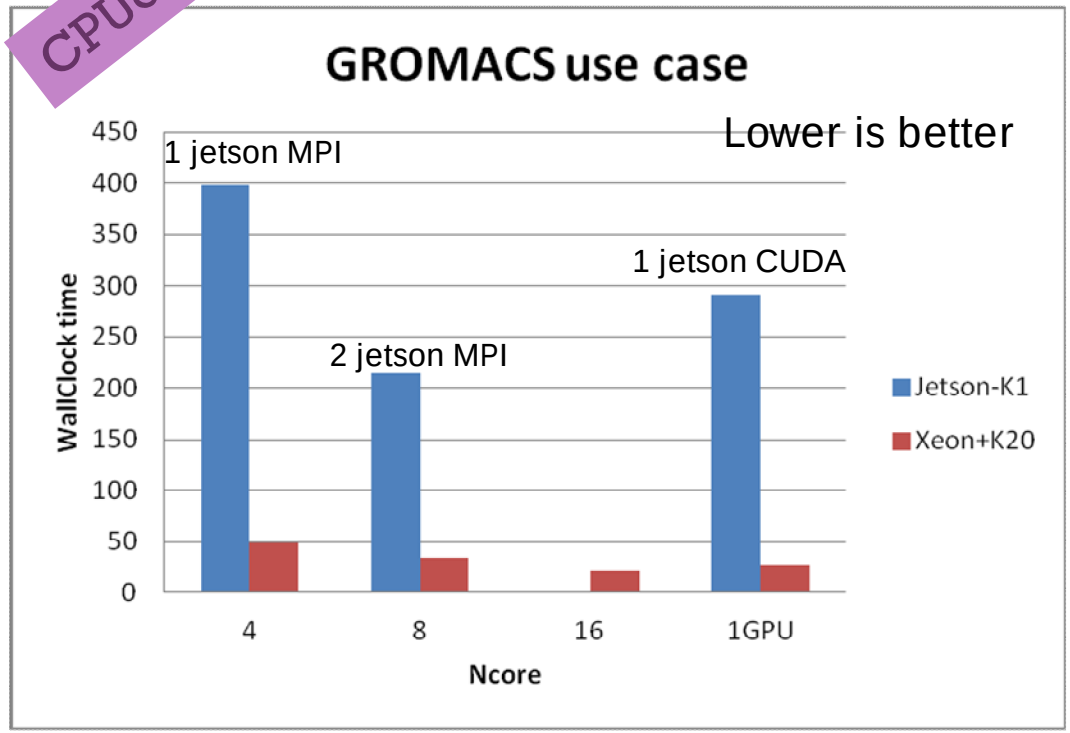
1024 x 1024 pixel

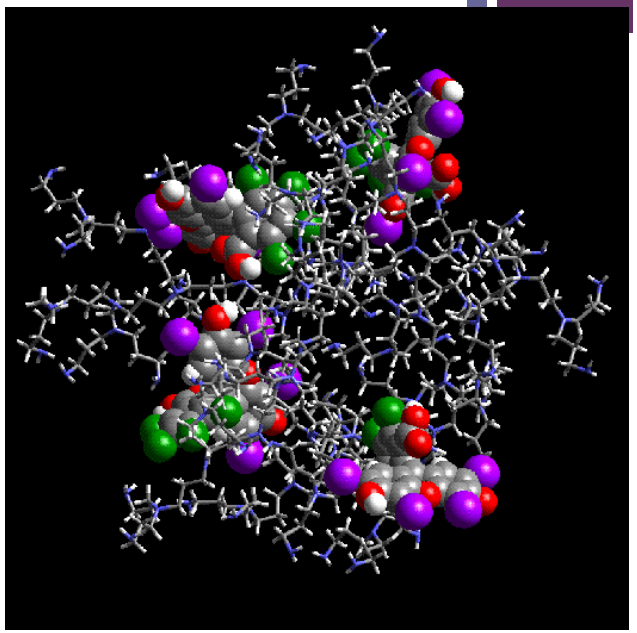


Molecular Dynamics on Jetson-K1

CPU&GPU

Parallel application for CPU and GPU:
real life use case with GROMACS



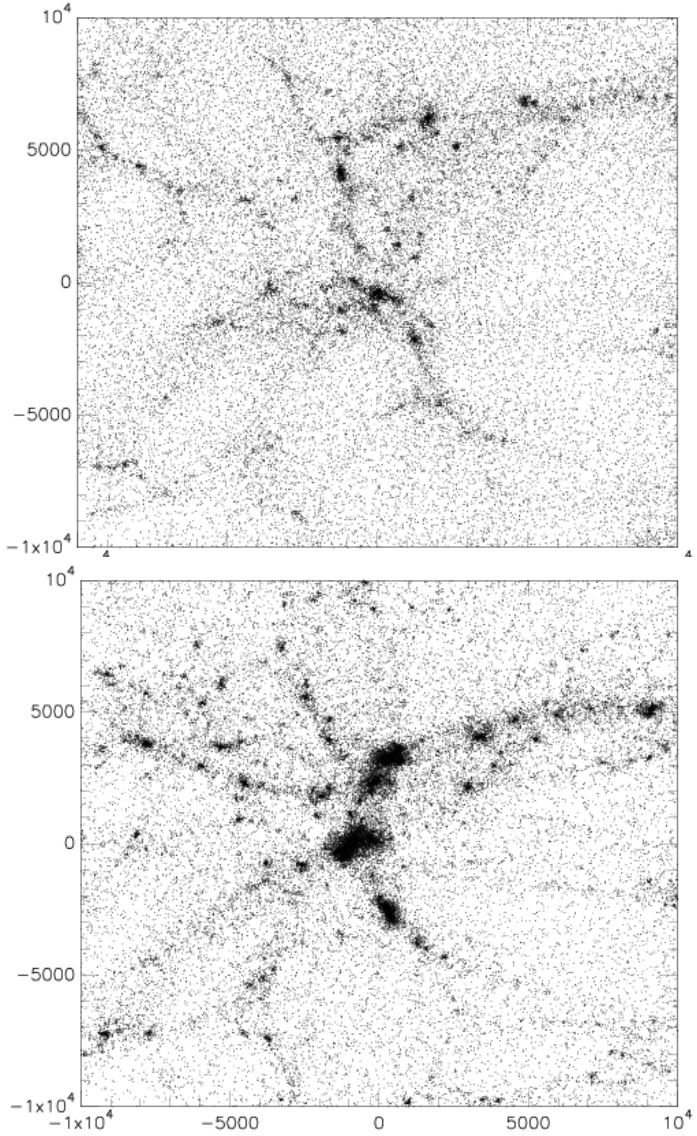


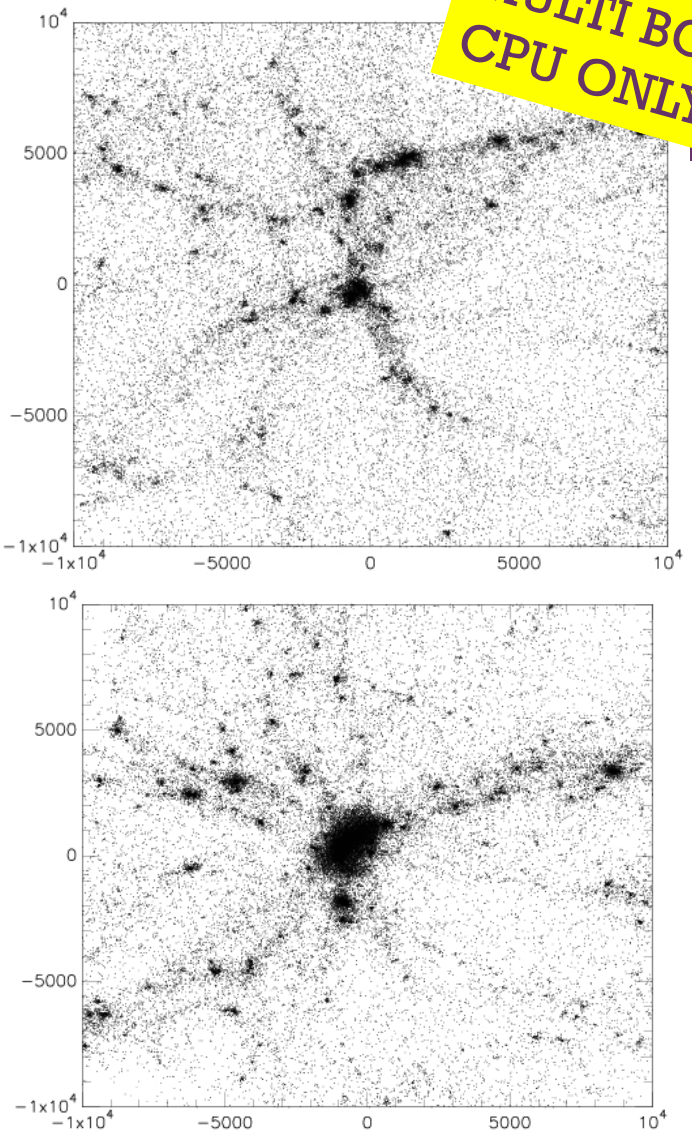
- Jetson-K1 about 10X slower using the same number of CPU cores
- Jetson-K1 about 10X slower using the GPU (vs. an NVIDIA Tesla K20)
  - Jetson-K1 **13.5W**
  - Xeon+K20 **~320W**

Formation of a galaxy cluster with Gadget2

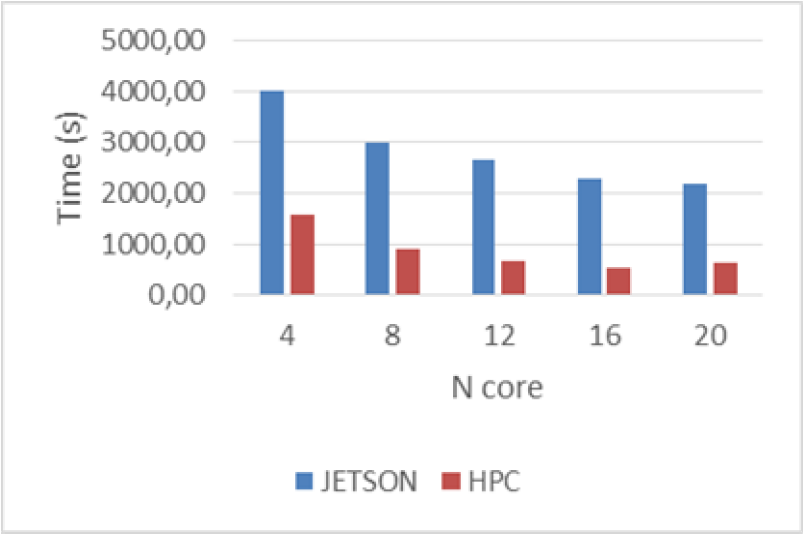
MULTI BOARD CPU ONLY

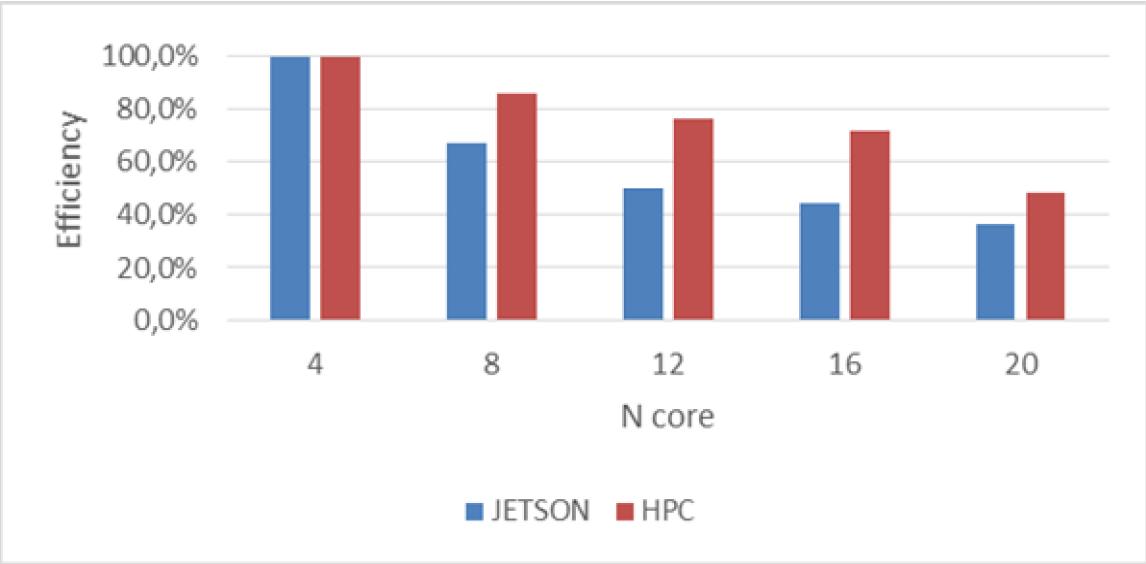
276498 particles, formation of a cluster of galaxies in an expanding Universe





Formation of a galaxy cluster





Neural Networks: DPSNN-STDP

CORTICONIC

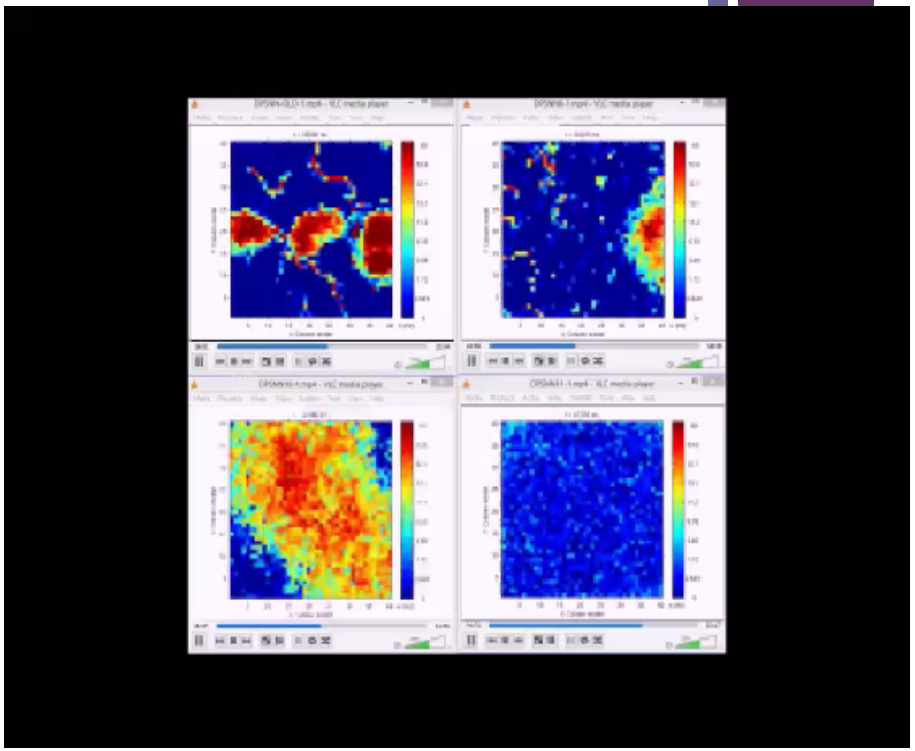
A Challenging Problem

The simulation of the cortical field activity can be accelerated using parallel/distributed many-processor computers. However, there are several challenges, including:

- Neural networks heavily interconnected at multiple distances, local activity rapidly produces effects at all distances → Prototype of non-trivial parallelization problem
- Each neural spike originates a cascade of synaptic events at multiple times: $t + \Delta t_s$ → Complex data structures and synchronization. Mixed time-driven (delivery of spiking messages

Energy to Solution, Speed and Power

- 2.2 micro-Joule per simulated synaptic event on the "embedded dual socket node"
  - 4.4 times better than spent by "server platform"
- instantaneous power consumption: "embedded "14.4 times better than "server"
- "server" platform 3.3 faster than "embedded"
- All inclusive, measured using amperometric clamp on 220V@50Hz power supply on:
- Details in arXiv:1505.03015 – May 2015

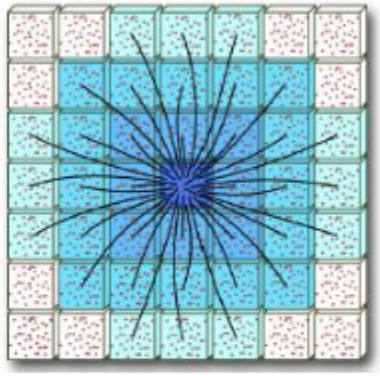


COSA Meeting - Roma - 2015 05 18

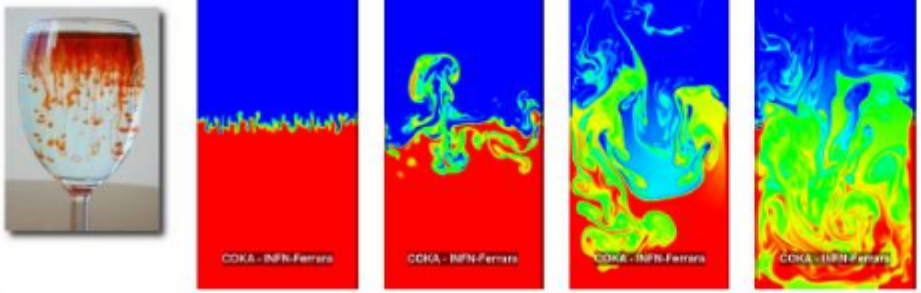
Lattice Boltzmann on the Tegra K1

GPU only

Lattice Boltzmann Methods: D2Q37



(*) Schifano et al.
A portable OpenCL Lattice Boltzmann code for multi- And many-core processor architectures,
Proc. Comp. Sci. 29, 2014



Performance comparison with K20

Propag. (MLUPS)		Collide (MLUPS)	
Tegra	K20	Tegra	K20
17.8	256.0	1.6	89.4

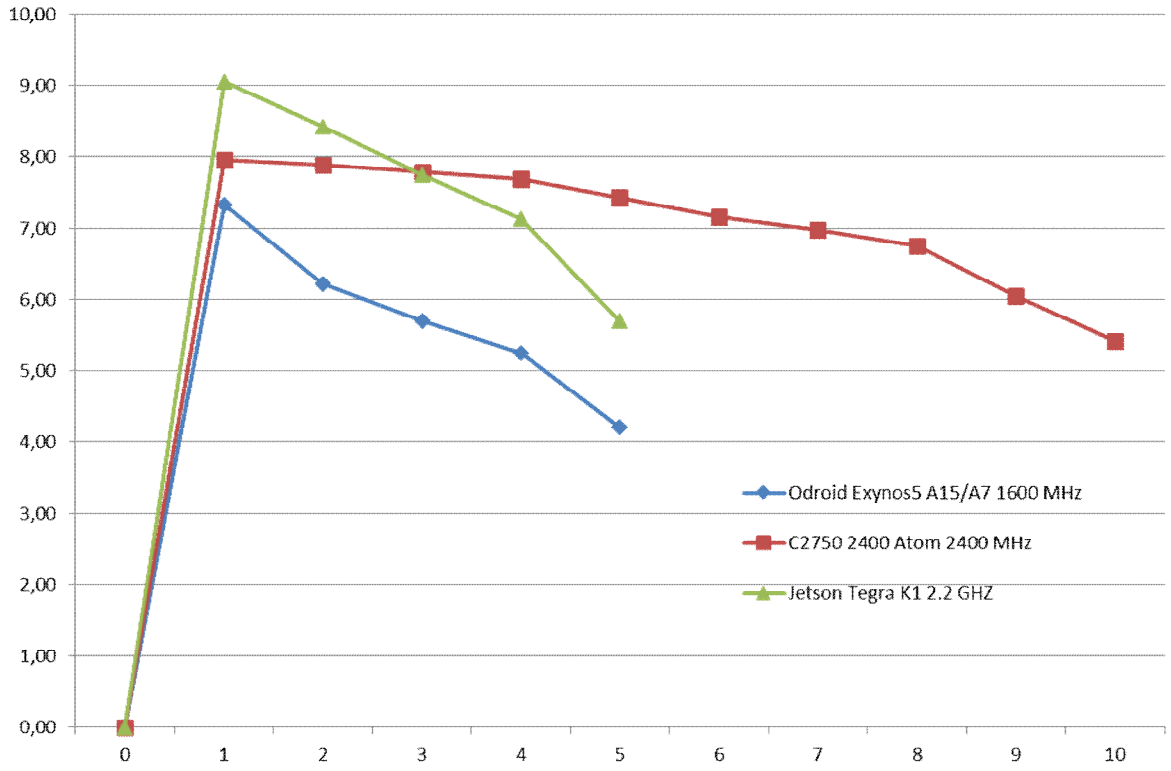
- Porting easier than expected
- Performance under investigation

See Enrico Calore talk

Lucia Morganti – INFN-CNAF

HS06

HS06 on Exynos5, TegraK1 and Atom C2750 – Per core loaded



(data from M.Michelotto@HEPIX 2014 https://indico.cern.ch/event/320819/session/3/contribution/30/material/slides/0.pptx)

Tests from CMS

PYTHIA

- Standard test MC05 (500 eventi, pp 14 TeV, Jets e loro caratteristiche)
- Suite completa di test (> 20)

Macchina	Real Time 05	Tutti i tests	Ratio 05 (X/Intel)
Odroid u2	30m22.565s	45m0.935s	9
Nvidia K1	13m41.977s	26m10.149s	2.8
IntelCore i7-2600 CPU @ 3.40GHz	4m32.135s	9m23.435s	1
Marvell ARMADA XP (Dell Copper)	51m1.198s	79m52.159s	8.8

ROOT

- Calcolo dei ROOTMarks

Macchina	ROOT Marks	Ratio ROOTMarks (X/Intel)
Odroid u2	309	0.14
Nvidia K1	583	0.26
IntelCore i7-2600 CPU @ 3.40GHz	2214	1
Marvell	157	0.07

CMS Analysis test

- Macro per calcolare fit alla massa dell'Higgs dati veri dell'analisi H to bb)
  - Macro root compilata
  - Comprende I/O sequenziale (~ 1 GB di root files)

Macchina	ZinvH	Ratio ZinvH(X/Intel)
Odroid u2	238 sec	10.8
Nvidia K1	110 sec	5.0
IntelCore i7-2600 CPU @ 3.40GHz	22 sec	1
Marvell ARMADA XP (Dell Copper)	443 sec	20

More on Pantaleo talk…

Low latency connection through FPGAs

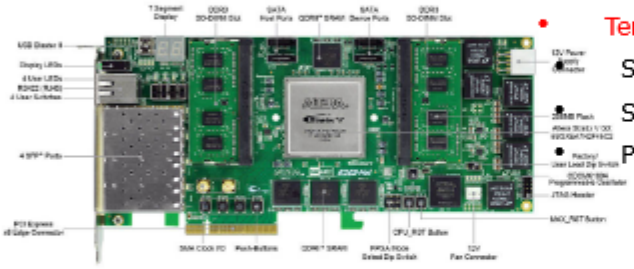
- FPGA SoCs are hybrid components that integrate a programmable hw component and a multicore low power ARM CPU
- Two main reasons to test them in COSA
  - Acceleration of computational task executed by microP embedded in the FPGA taking advantage of the RDMA capabilities of the APEnet architecture
    - APEnet v5, NaNet, PICOLO systems
  - Computational tasks executed on the ARM CPU taking advantage of the integrated low latency connection capabilities in the
    - Study of new architectures, i.e ExaNeSt

More on Lonardo talk…

FPGA+Custom network: sistemi di sviluppo

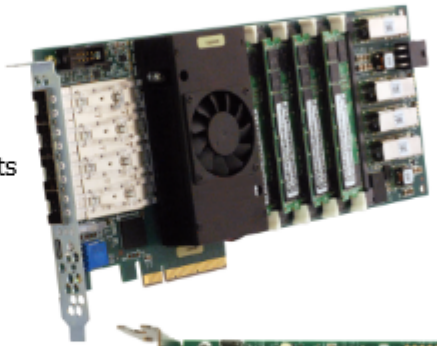
High-end o SoC con multicore CPU integrati

- APEnet+, Terasic, Nallatech, Bittware,…



- Terasic DE5-Net
  - Stratix V
  - SFP+ network ports
  - PCIe X8 Gen1/2/3

- Bittware S5-PCI-DS
  - Dual Stratix V
  - OpenCl enabled, PCI x16, huge memory banks…, SFP+ conn.



- Nallatech 395-AB
  - Stratix V AB
  - Highest density memory: 4x 32GB of DDR3
  - (4) SFP+ network ports supporting a range of network protocols and speeds

- Bittware A10PL4 (Arria10 based)
  - Altera Arria 10 GT/GX FPGA
  - PCIe x8 Gen1, Gen2, or Gen3
  - QSFP for 2x 100GigE, 2x 40GigE, or 8x 10GigE
  - Memory: up to 32 GBytes of DDR4

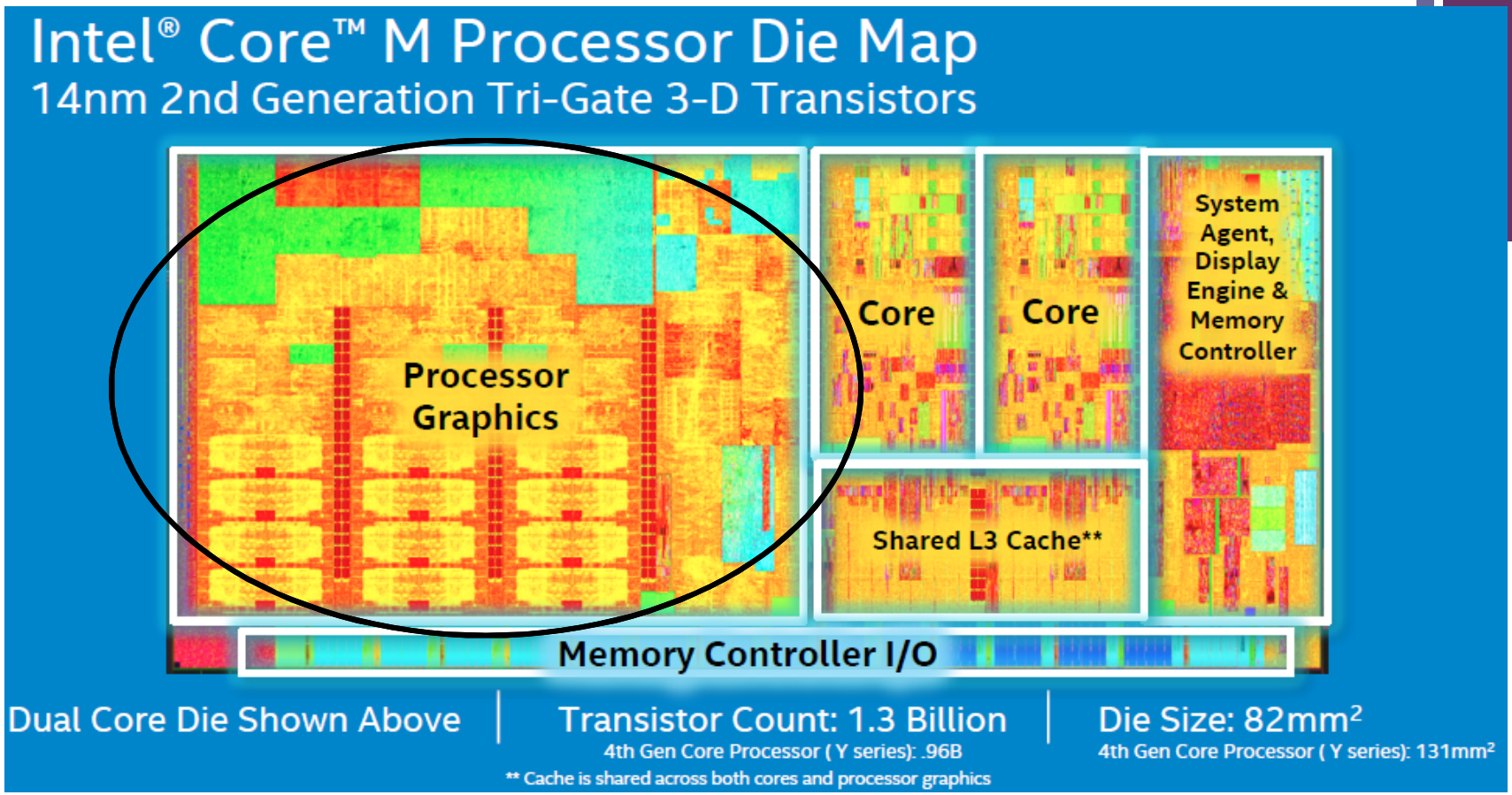
Only ARM based SoCs? And Intel?

- INTEL produce SoCs
- Some of them are low power
- Already 64bit
- Integrated GPU
  - CILK++ programmable
  - OpenCL programmable

Some low power from Intel

Product Name	Intel® Atom™ Processor E3845 (2M Cache, 1.91 GHz)	Intel® Core™ i7-5650U Processor (4M Cache, up to 3.20 GHz)	Intel® Xeon® Processor D-1540 (12M Cache, 2.00 GHz)	Intel® Core™ M-5Y71 Processor (4M Cache, up to 2.90 GHz)	Intel® Atom™ Processor C2750 (4M Cache, 2.40 GHz)
Code Name	Bay Trail	Broadwell	Broadwell	Broadwell	Avoton
Essentials					
Status	Launched	Launched	Launched	Launched	Launched
Launch Date	Q4'13	Q1'15	Q1'15	Q4'14	Q3'13
Processor Number	E3845	i7-5650U	D-1540	5Y71	C2750
Cache	2 MB L2 Cache	4 MB	12 MB	4 MB	4 MB
Instruction Set	64-bit	64-bit	64-bit	64-bit	64-bit
Embedded Options Available	Yes	Yes	No	No	No
Lithography	22 nm	14 nm	14 nm	14 nm	22 nm
Recommended Customer Price	TRAY: $52.00	TRAY: $426.00	TRAY: $581.00	TRAY: $281.00	TRAY: $171.00
Performance					
# of Cores	4	2	8	2	8
# of Threads	4	4	16	4	8
Processor Base Frequency	1.91 GHz	2.2 GHz	2 GHz	1.2 GHz	2.4 GHz
TDP	10 W	15 W	45 W	4.5 W	20 W
Graphics Specifications					
Processor Graphics ‡	Intel® HD Graphics	Intel® HD Graphics 6000	None	Intel® HD Graphics 5300	
Graphics Base Frequency	542 MHz	300 MHz		300 MHz	
Graphics Burst Frequency	792 MHz				
Intel® Quick Sync Video	Yes	Yes		Yes	
# of Displays Supported ‡	2	3		3	
Graphics Max Dynamic Frequency		1 GHz		900 MHz	

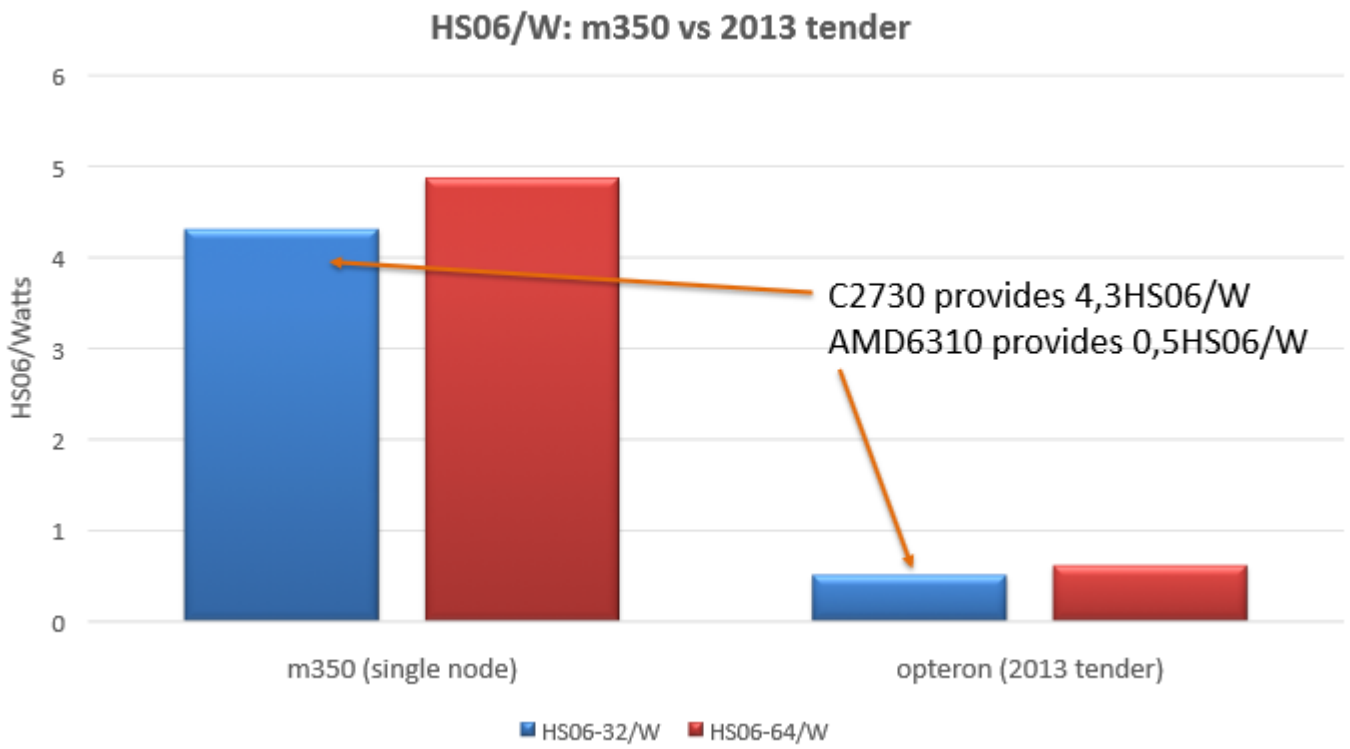
ATOM - BayTrail | i7 Mobile | XEON D | CORE M | AVOTON



- 2 cores + GPU
  - Intel HD Graphics
  - OpenCL 2.0 Support
- 4.5 Watt (TDP)

(*) http://www.notebookcheck.net/Intel-Core-M-5Y70-Broadwell-Review.130930.0.html
(**) http://www.intel.com/content/www/us/en/processors/core/core-m-processor-family-spec-update.html

AVOTON on HP Moonshot - HS06



(data from A.Chierici@HEPIX https://indico.cern.ch/event/305362/session/2/contribution/22/material/slides/0.pdf)

Conclusion

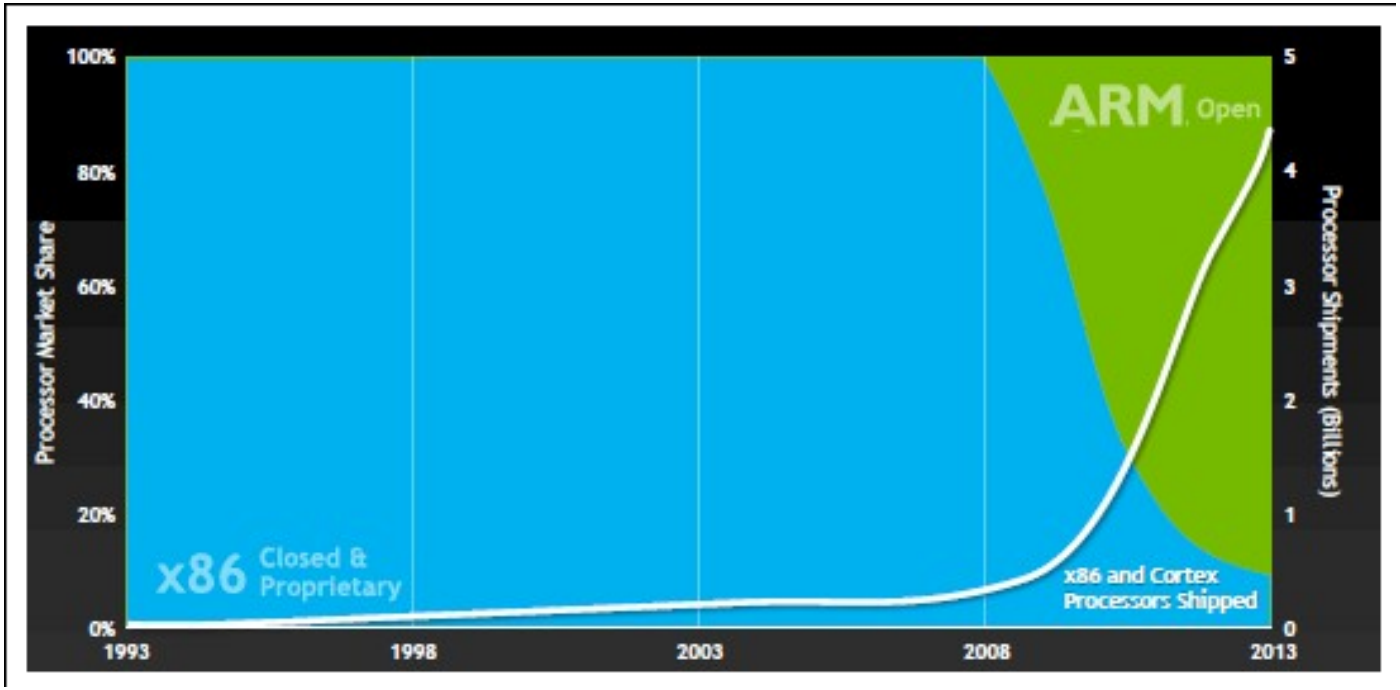
- COSA is showing that mobile and embedded low power System-on-Chip are becoming attractive for scientific computing
  - In particular if you manage to extract power from the GPU
- But hey still have many limitations for a production environment
- We are continuing to acquire know-how on hybrid architectures (both low power and traditional systems)
  - Operations
  - Performance Measurement
  - Application porting
  - Programming models
- It was decided to wait for the next generation 64bit SoCs before expanding the low power clusters
- COSA participated in the creation of the H2020 proposal for PICOLO
  - LEIT-ICT4 2014 call

Links and contacts

- http://www.cosa-project.it
- COSA@HEPIX Talk (https://indico.cern.ch/event/346931/session/5/contribution/57/material/slides/0.pdf)
- COSA@GDB talk (https://indico.cern.ch/event/319744/contribution/1/material/slides/0.pdf)
- COSA Indico (https://agenda.infn.it/categoryDisplay.py?categId=779)
- COSA mailing list: cosa-project__at__lists.infn.it

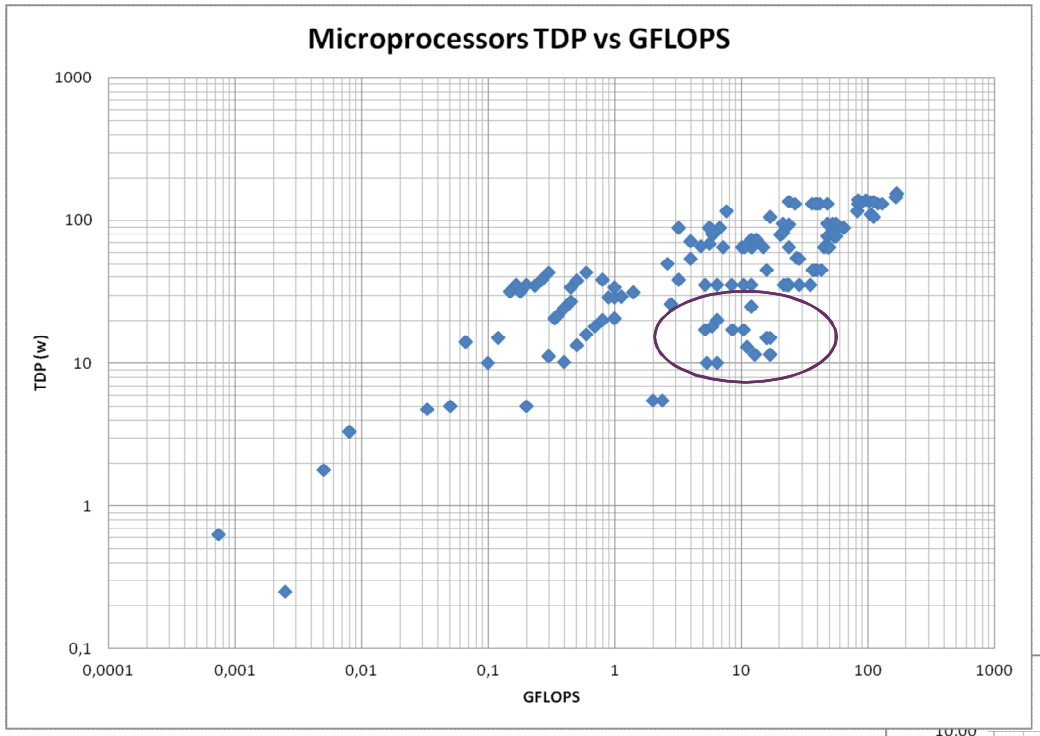
Backup

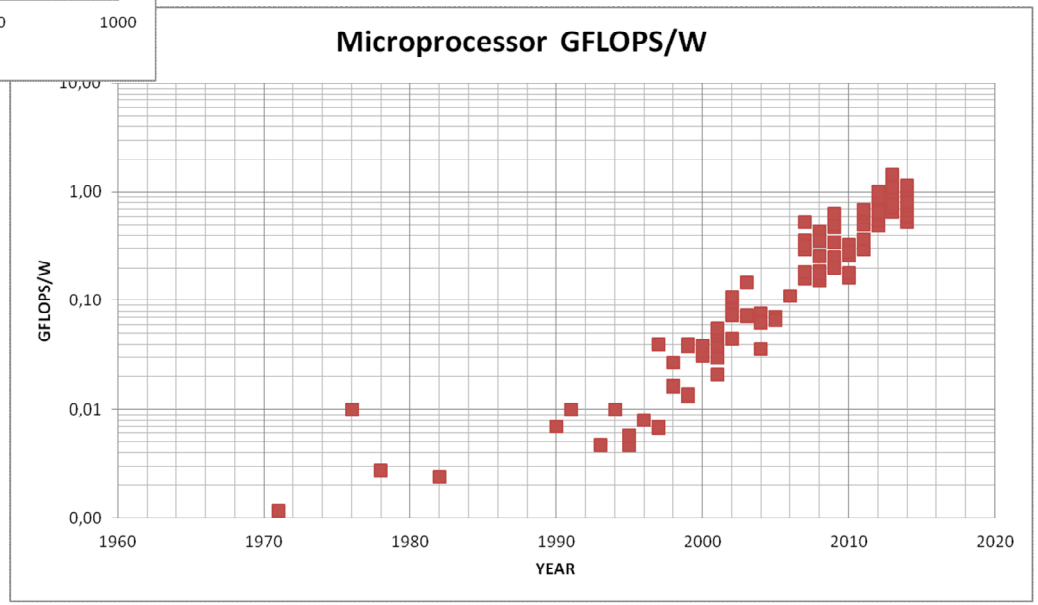
ARM based processor shipment



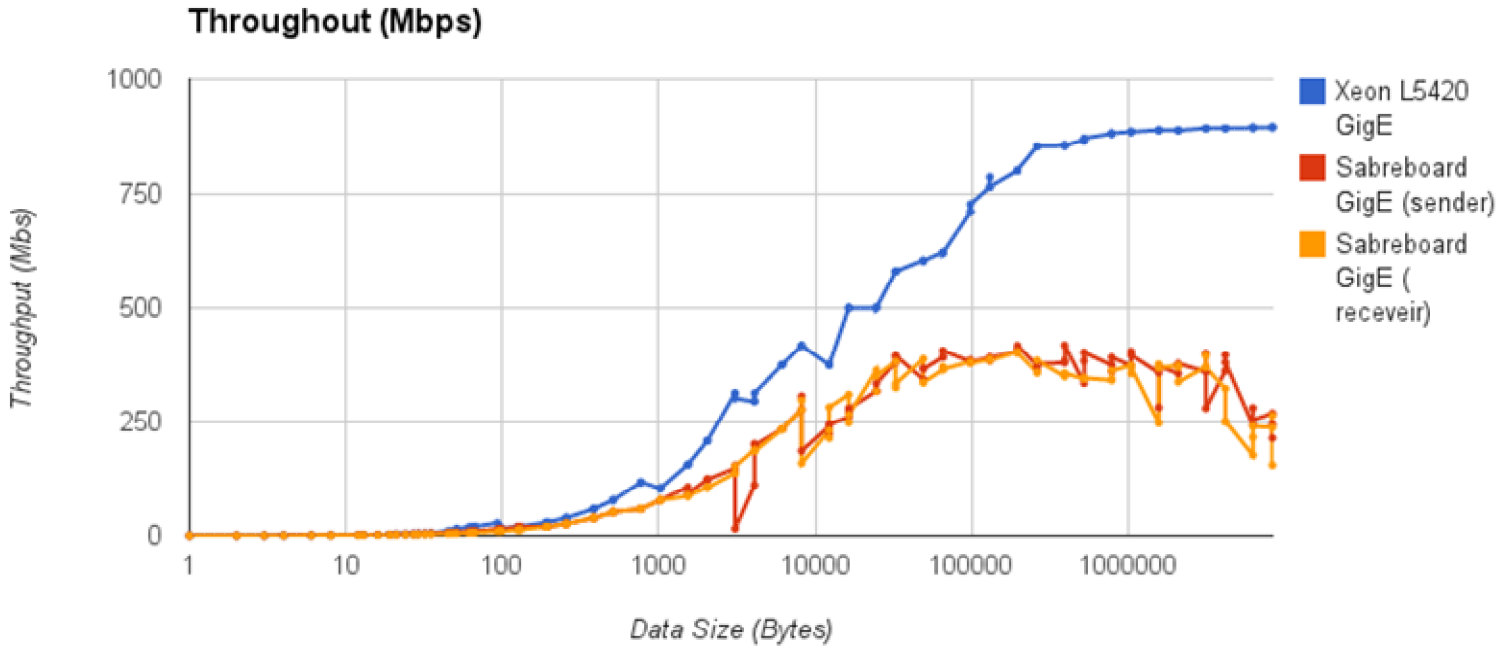
- ARM based processors are shipped in billions of units
  - ARM licences the Intellectual Properties to manufactures
    - …many manufactures ….
      - Samsung (Korea), MediaTek (China), Allwinner (China), Qualcomm (USA), NVIDIA (USA), RockChip (China), Freescale (USA), Texas Instruments (USA), HiSilicon(China), Xilinx (USA), Broadcom(USA), Apple(USA), Altera(USA), ST(EU) , WanderMedia(Taiwan), Marvel(USA), AMD(USA)etc..

CPU GFLOPS/Watt





Gbit Ethernet

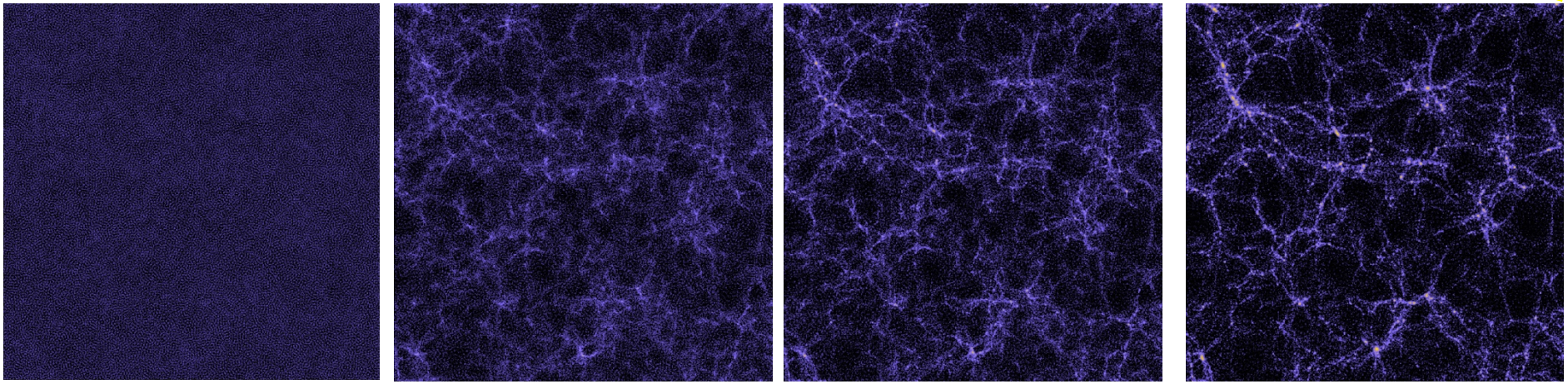


While latency was comparable to a server class 1Gb ethernet card (50/75 us)

Real use-case: evolving the Universe

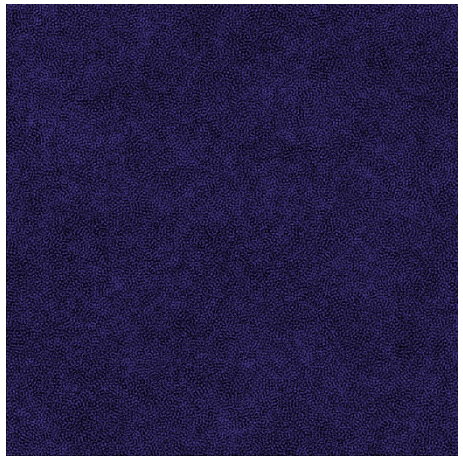
MULTI BOARD CPU ONLY

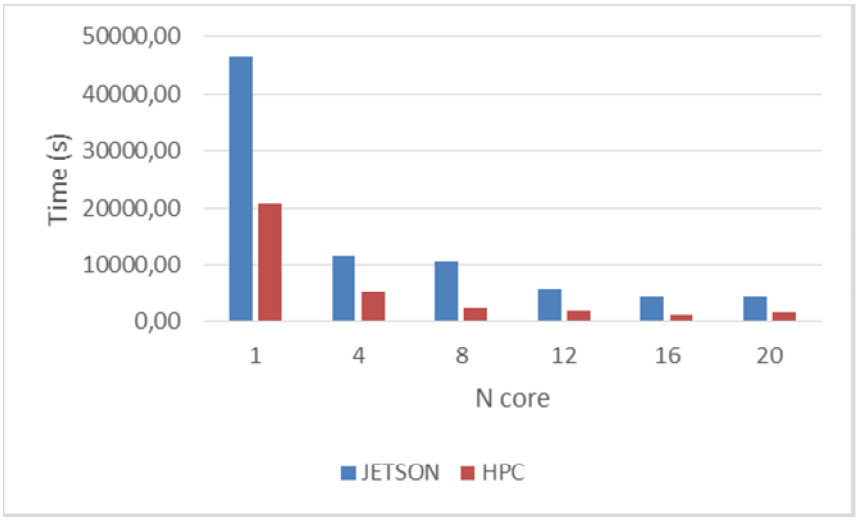
128^3 particles in a cube of 250 Mpc side, 13.6 Gyr

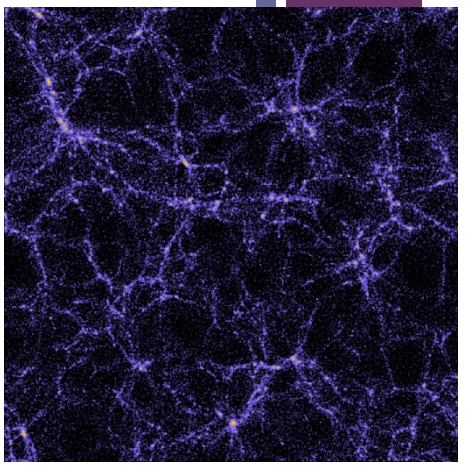


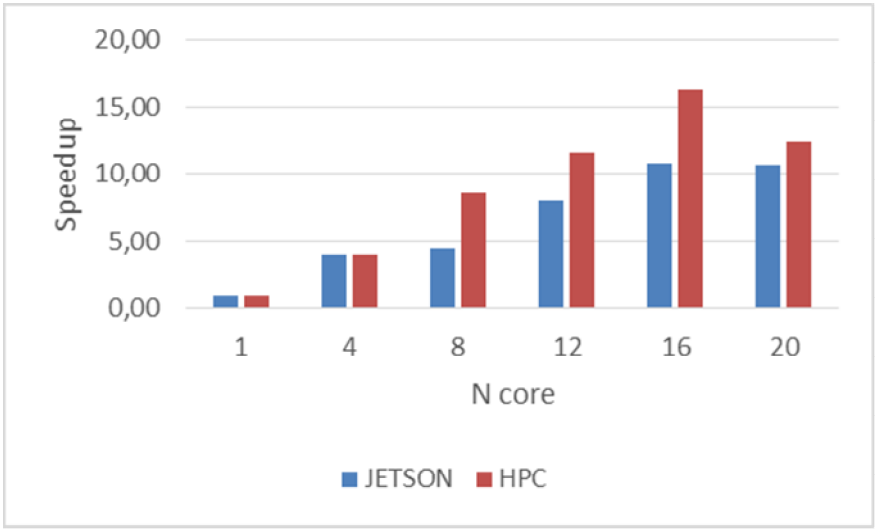
- Jetson-K1 about 2.6X slower than Xeon using 4 CPU cores, and about 4.4X slower using 8 CPU cores
  - Jetson-K1 **12.1 W**
  - Xeon **~220 W**

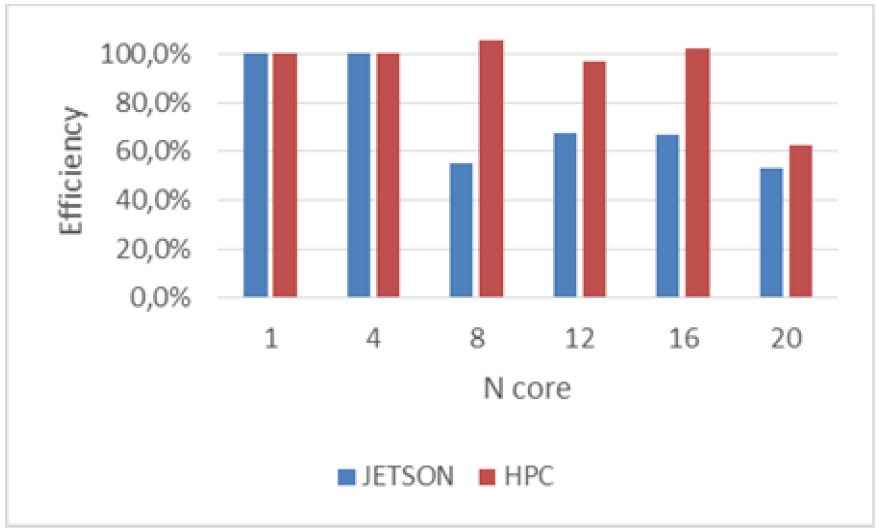
Real use-case: evolving the Universe



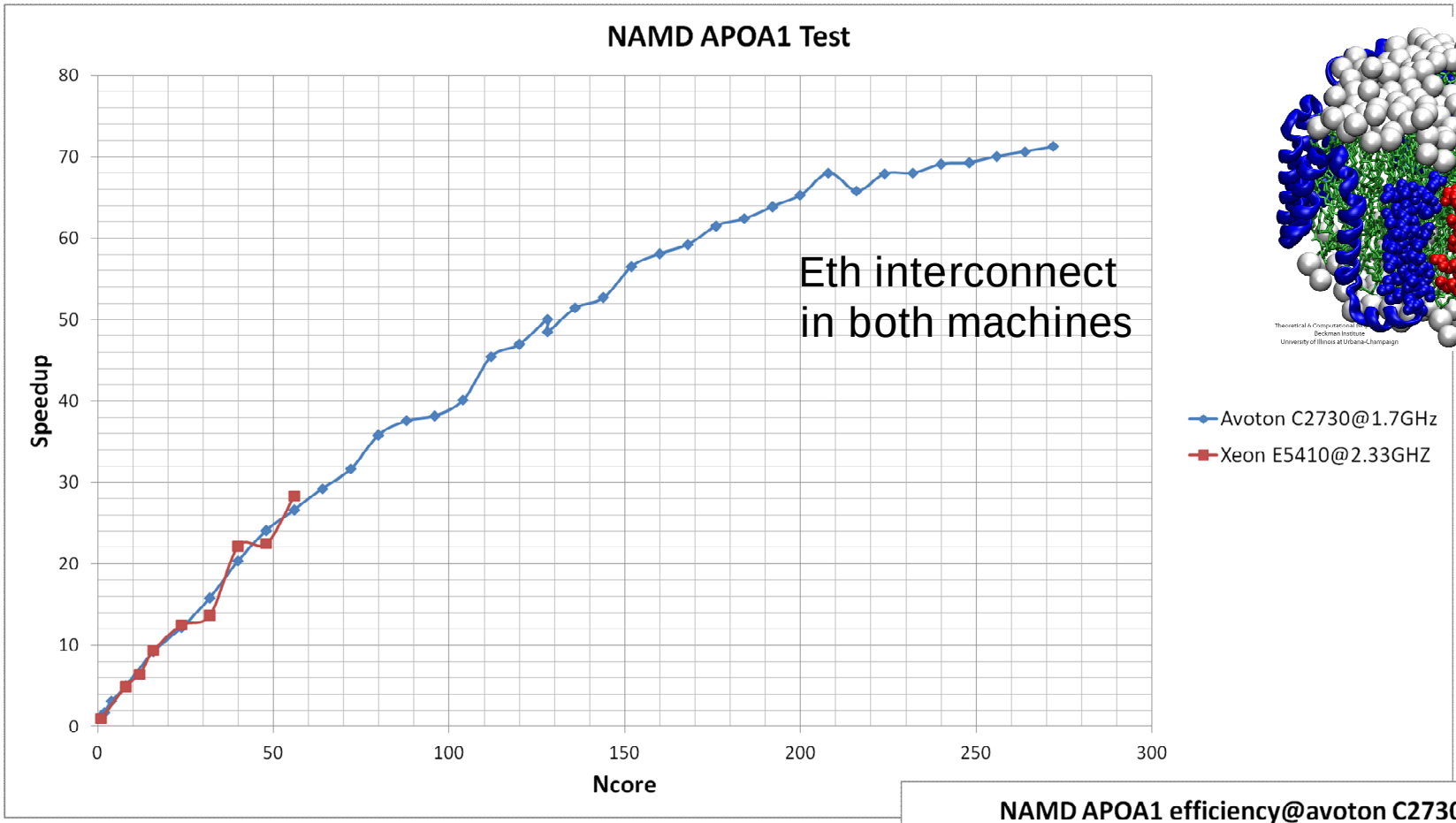


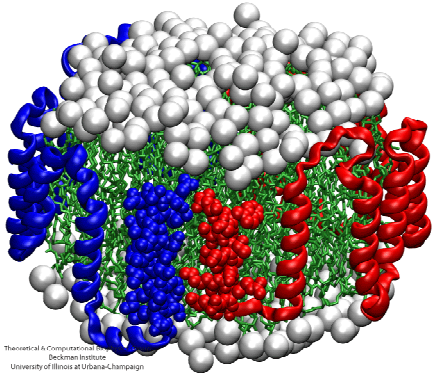


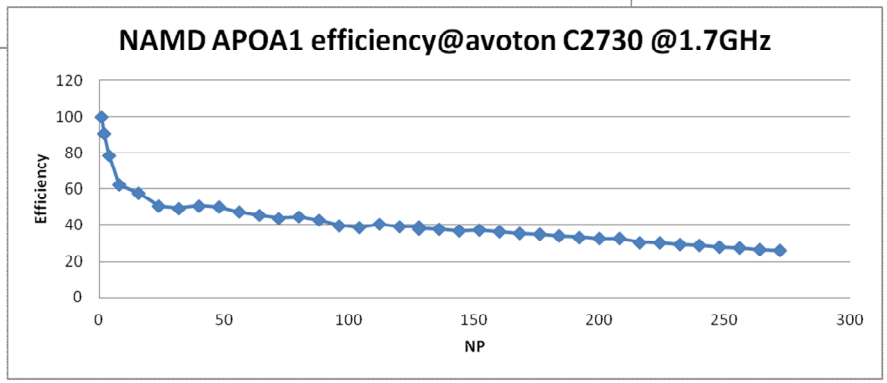




Test on Intel AVOTON

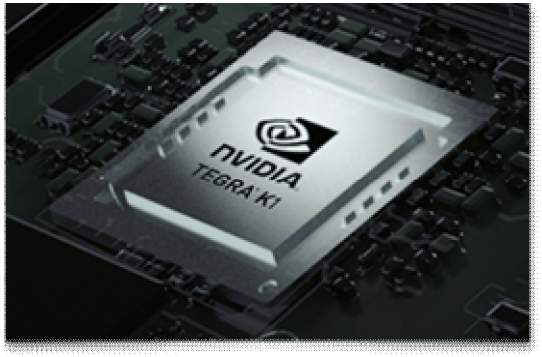






N.B. – Comparison with an old Penryn Xeon CPU

GPU acceleration in K1



4 core ARM A15 ~ 18 GFLOPS
Kepler SMX1 192 core ~ 300 GFLOPS (sp)
~ 15 W
~ 21 GFLOPS/W (sp)

2 x (E5-2673v2 (IvyBridge) 8 core)
~ 200 GFLOPS (dp)
220 W

NVIDIA TESLA K40 2880 core
~ 1400 GFLOPS (dp) ~ 4300 GFLOPS (sp)
235 W

CPU+GPU ~ 3 GFLOPS/W (dp)
~ 9 GFLOPS/W (sp)