+

The INFN COSA project and possible applications

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

http://www.cosa-project.it

Istituto Nazionale di Fisica Nucleare

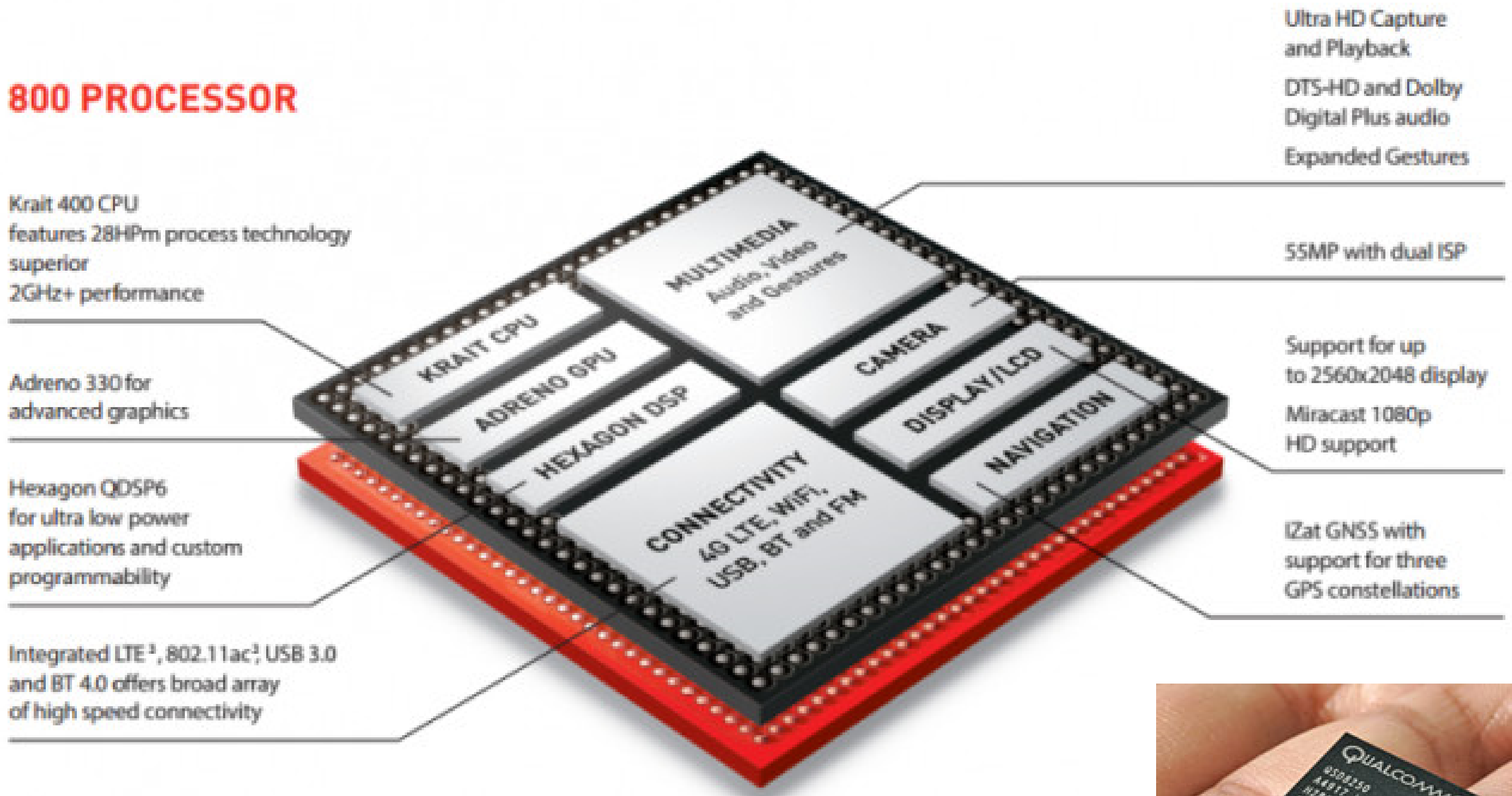
INFN COSA project

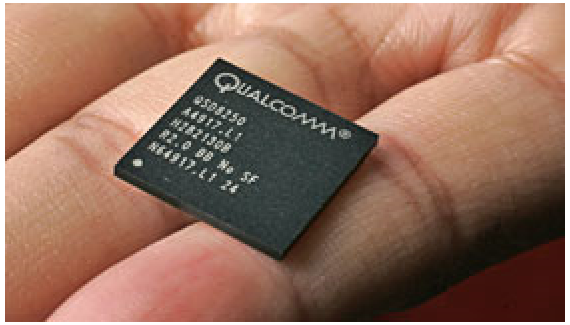
- COSA: Computing On SOC Architecture
- Duration: 3 years from January 2015
- Departments: 7 INFN
  - CNAF, PI, PD, ROMA1, FE, PR, LNL
- BUDGET :51.5 kEuro Year1, 42kEuro Year2
  - Funded by INFN CSN5

Objectives

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

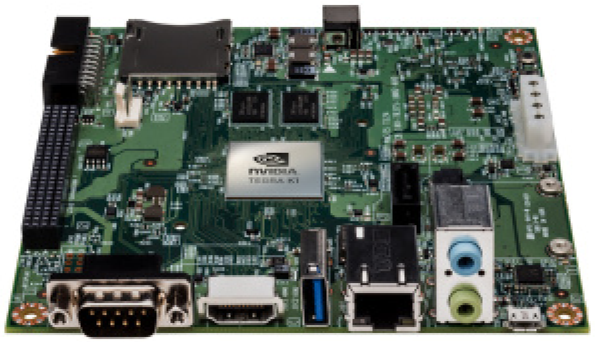
Low-Power System on Chip (SoCs)





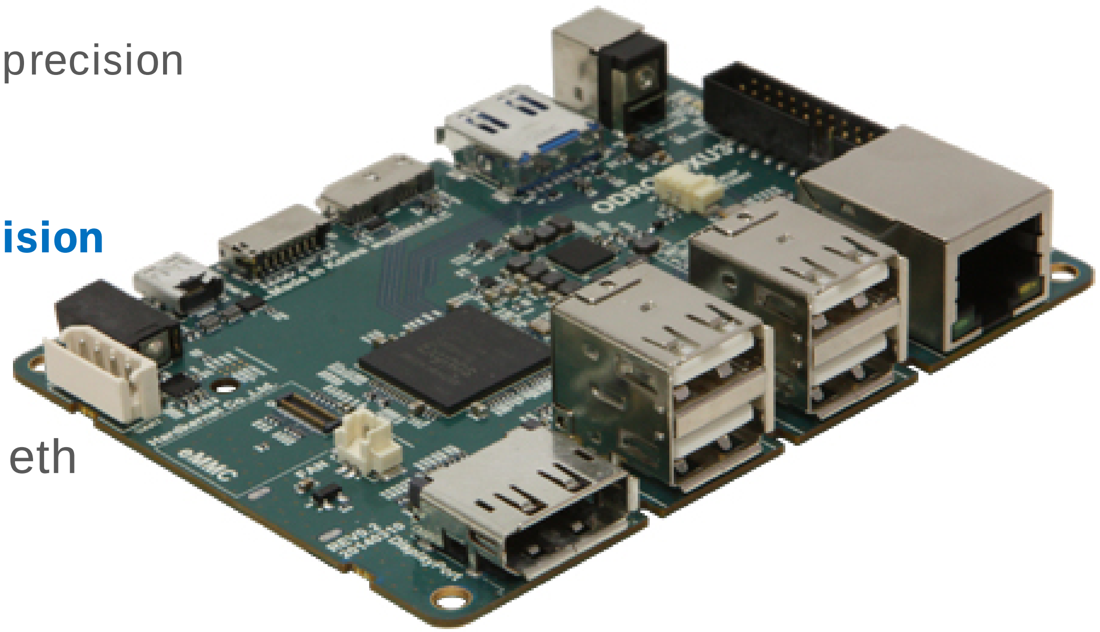
Ok, but then....an iPhone cluster?

- 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, 1x1000Gbs eth
- Ubuntu 14.4
- HDMI 1.4 port
- 64 GB flash storage

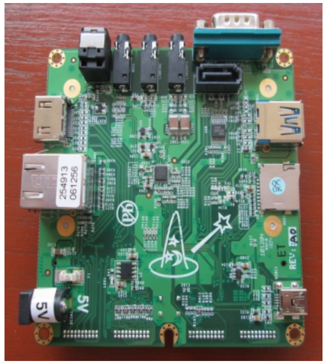


Power consumption max ~ 15 W

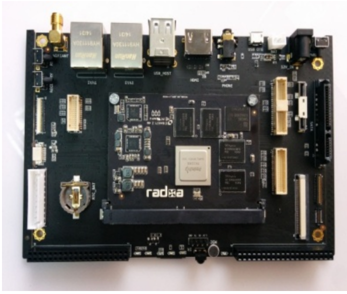
Costs 150 euro!

Other nice boards…

…during the old good times of ARM 32bit



WandBoard



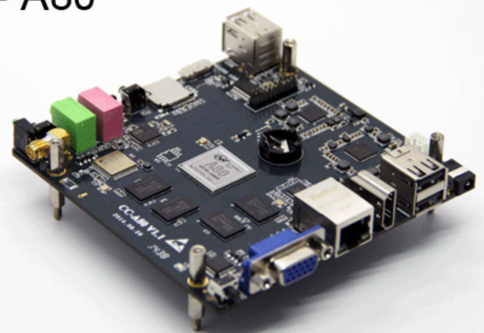
Rock2Board

PandaBoard

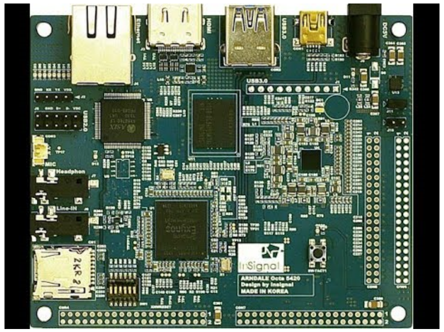
DragonBoard

SabreBoard

- A80



CubieBoard



Arndale OCTA Board

Texas Instruments EVMK2H

http://elinux.org/Development_Platforms

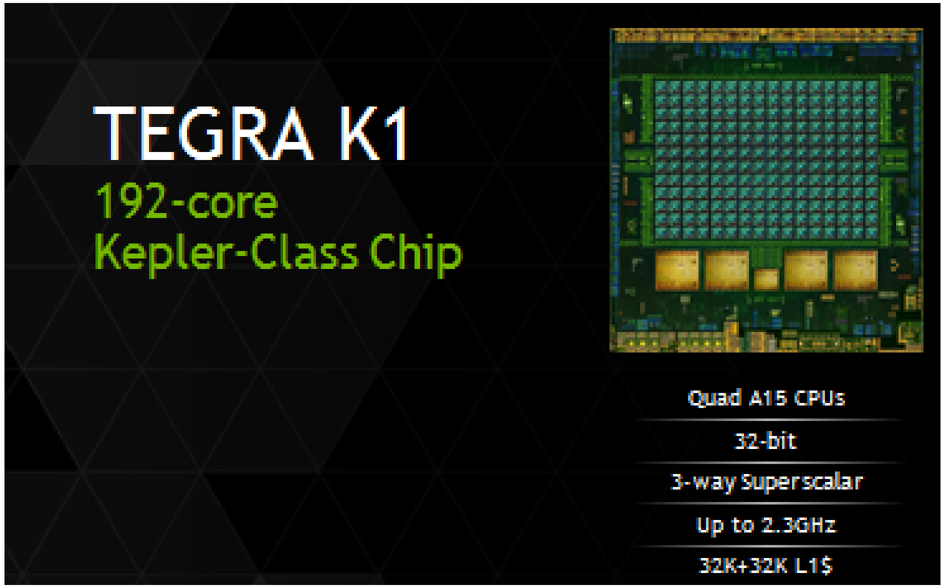
- …and counting…

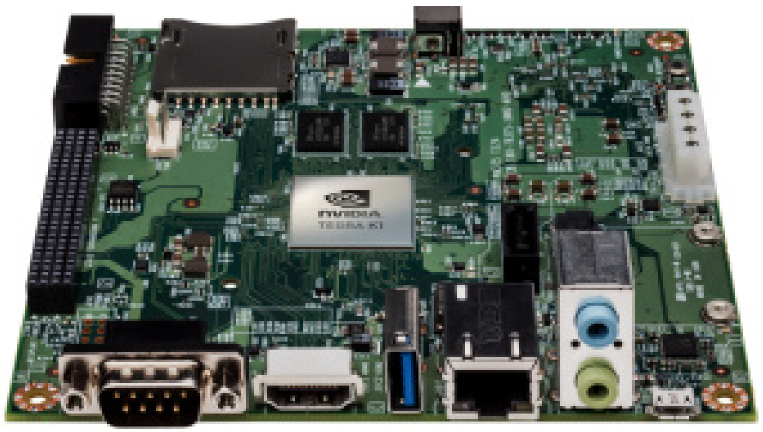
Some specs

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)**

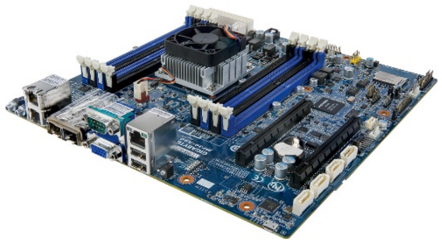
- ... for less than 200 Euros
- 32bit
  - 64bit version announced

ARMv8 64bit boards…

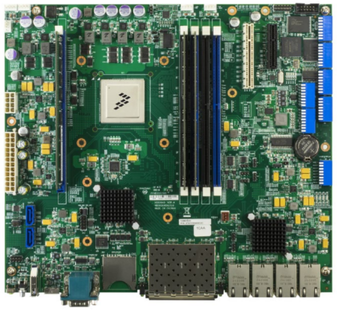
…harder times

Server Grade platform

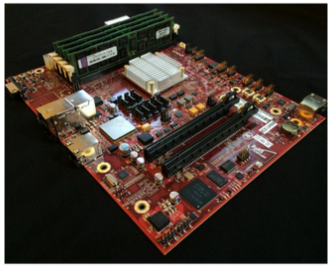
ARM Juno Board
r1: 2xA57 + 4xA53
r2: 2xA72 + 4xA53
DRAM: 8 Gbytes
4 PCI-E (Gen.2, 4x)
r1: 5000$
r2:7000$



Gigabyte MP30-AR0
AppliedMicro X-Gene1 8core
DRAM:max128GB
2 x 10GbE SFP+
2 x 1GbE LAN ports
2 x PCI-Express slots (Gen.3, 8x)
700eu

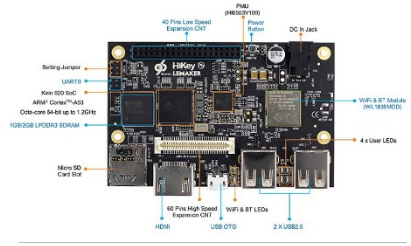


FreescaleQorIQ LS2085A
8 x Cortex-A57 cores
DRAM:max 16GB
PCI Gen3 (x8)
4 x 10 GbE SFP
4 x 10 GbE RJ45
About 3000$

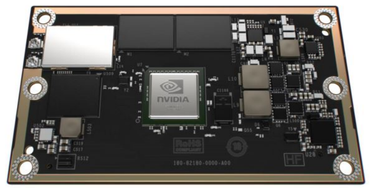


AMD Opteron A1100
16GB RAM
2x10Gbs
Cost 2000$

Embedded platform

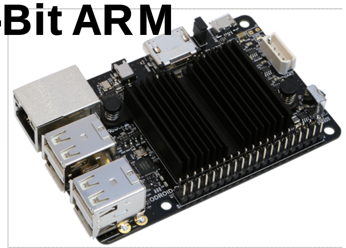


HiKey 96boards
1/2GB LPDDR3 SDRAM
8 x Cortex-A53 cores
Cost: $100 (2GB)



NVIDIA Jetson TX1
4x A57 2 MB di L2; 4x A53 512 KB di L2
256 core di GPU NVIDIA Maxwell
600$

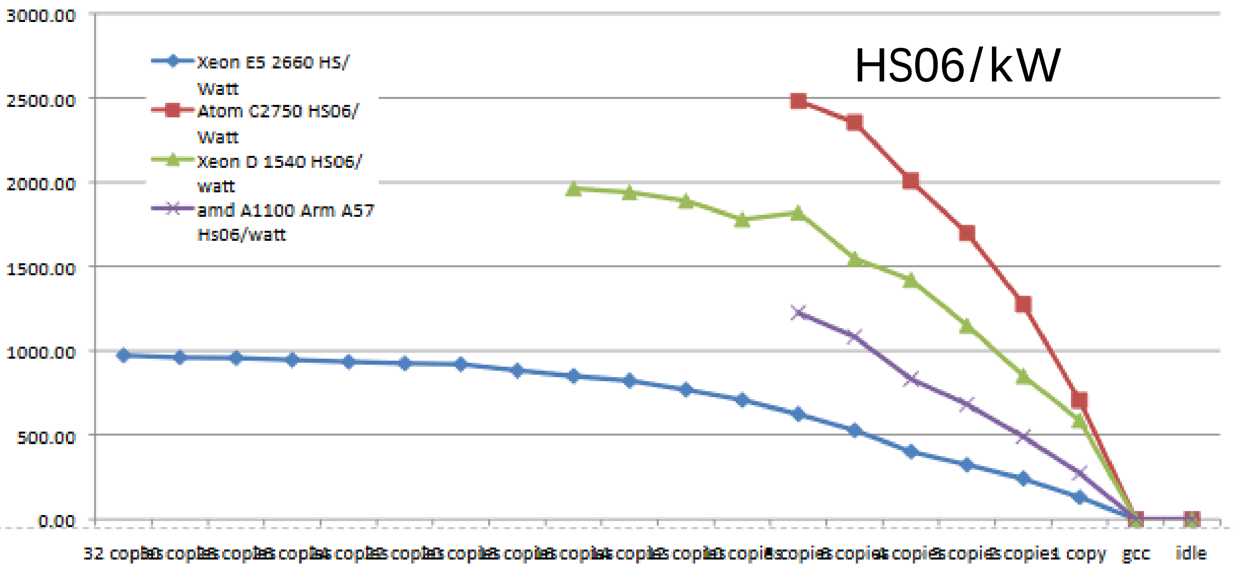
ODROID-C2 64-Bit ARM
4xA53@2GHz
Mali™-450 GPU
2GB RAM
1Gbs ETH



HS06 for AMD A1100 – ARM A57 8 cores

HS06/Watt on A57 and Xeon D

- Xeon D perform as expected
- Amd A1100 – A57 has good performances but disappointing power consumption. Not yes production grade configuration. Power supply?



39

@Michele Michelotto – COLA workshop – https://agenda.infn.it/getFile.py/access?contribId=8&resId=0&materialId=slides&confId=10434

Low Power from intel

CPU	Brand	Microarchitecture	Family	#	CORES	RAM (GB)	POWER (W)	HS06	HS/W
E5-2683v3	XEON	Haswell		2	56 (HT)	128	370	573	1.55
D-1540	XEON	Broadwell		1	16 (HT)	16	80	151	1.89
C2750	ATOM	Silvermont	Avoton	1	8	16	20	55	2.50
N3700	PENTIUM	Airmont	Braswell	1	4	16	7	28	4.00

- Tested in COSA during last 6 months
  - E5-2683v3 not low power – used as reference

- 29th Aril 2016: Intel announced that will exit the smartphone and tablet mobile SoC business by ending its struggling Atom chip product line

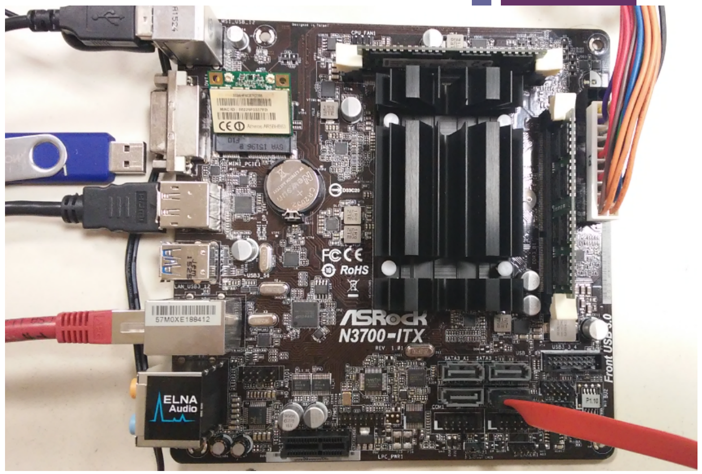
Comparison of Intel's Atom SoC Platforms

	Node	Release Year	Smartphone	Tablet	Netbook Notebook
Saltwell	32 nm	2011	Medfield Clover Trail+	Clover Trail	Cedar Trail
Silvermont	22 nm	2013	Merrifield Moorefield	Bay Trail-T	Bay Trail-M/D
Airmont	14 nm	2015	'Riverton'	Cherry Trail-T	Braswell
Goldmont	14 nm	2016	**Broxton (cancelled)**	**Willow Trail (cancelled)** Apollo Lake	Apollo Lake

- Apollo Lake not cancelled
  - Goldmont microarch
  - Integrated graphics
  - 14nm
  - 4 cores

Low power from Intel

Product Name	Intel® Pentium® Processor N3700 (2M Cache, up to 2.40 GHz)	Intel® Pentium® Processor J3710 (2M Cache, up to 2.64 GHz)	Intel® Pentium® Processor N3710 (2M Cache, up to 2.56 GHz)
Code Name	Braswell	Braswell	Braswell
Processor Number	N3700	J3710	N3710
Cache	2 MB L2 Cache	2 MB L2 Cache	2 MB L2 Cache
Instruction Set	64-bit	64-bit	64-bit
Embedded Options Available	No	No	Yes
Lithography	14 nm	14 nm	14 nm
Recommended Customer Price	TRAY: $161.00	N/A	N/A
Datasheet	Link		Link
Conflict Free	Yes	Yes	Yes
Additional Information URL	Link		Link
Performance			
# of Cores	4	4	4
# of Threads	4	4	4
Processor Base Frequency	1.6 GHz	1.6 GHz	1.6 GHz
Burst Frequency	2.4 GHz	2.64 GHz	2.56 GHz
TDP	6 W	6.5 W	6 W
Scenario Design Power (SDP)	4 W		4 W
Memory Specifications			
Max Memory Size (dependent on memory type)	8 GB	8 GB	8 GB
Memory Types	DDR3L-1600	DDR3L-1600	DDR3L-1600
Max # of Memory Channels	2	2	2
ECC Memory Supported ‡	No	No	No
Graphics Specifications			
Processor Graphics ‡	Intel® HD Graphics	Intel® HD Graphics 405	Intel® HD Graphics 405
Graphics Base Frequency	400 MHz	400 MHz	400 MHz
Graphics Burst Frequency	700 MHz		700 MHz



INTEL N3700

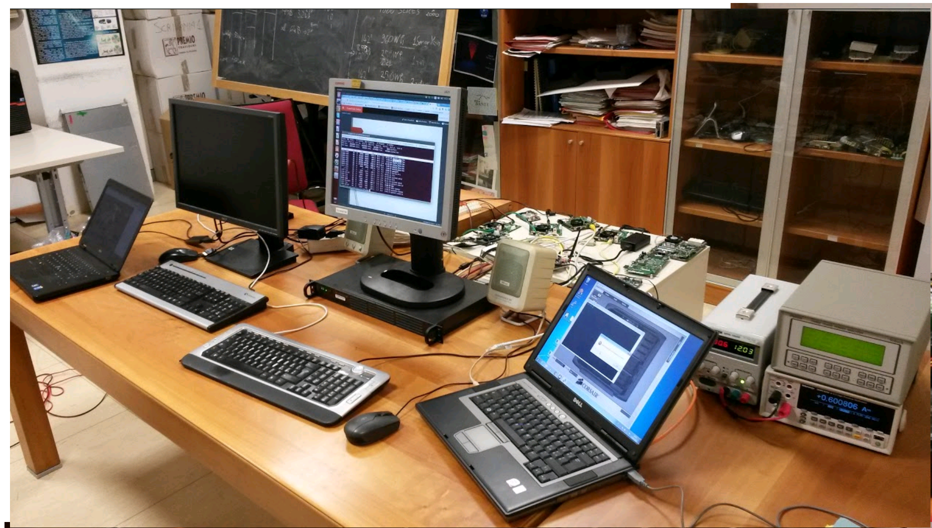
- 4 cores / Intel HD Graphics
- 6W
- Airmont microarchitecture
  (64 bit, No AVX/AVX2)
- 16GB
- SATA ports / PCIe 2.0 1x
- Fanless
- 100 euro !!!

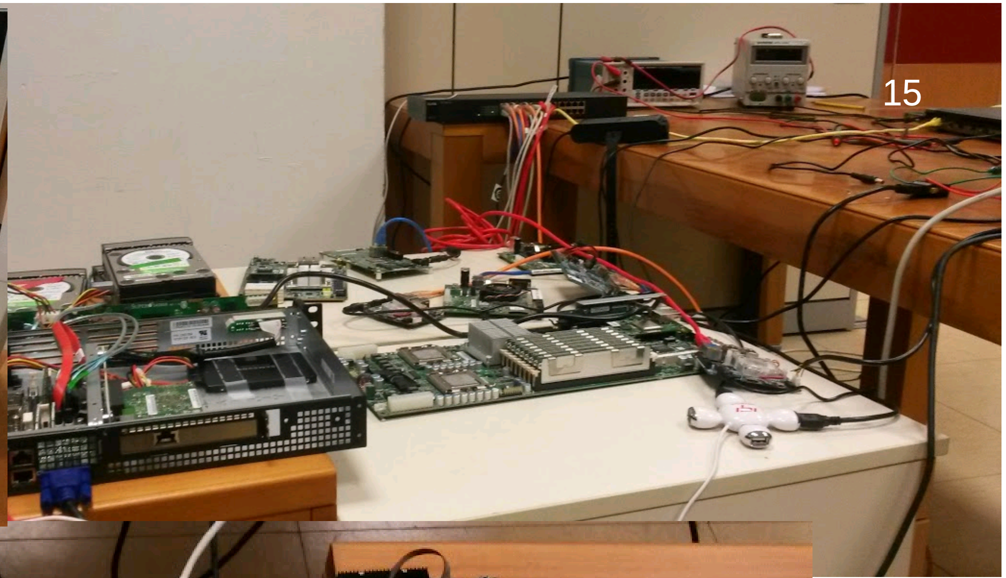
Low power from Intel - 2

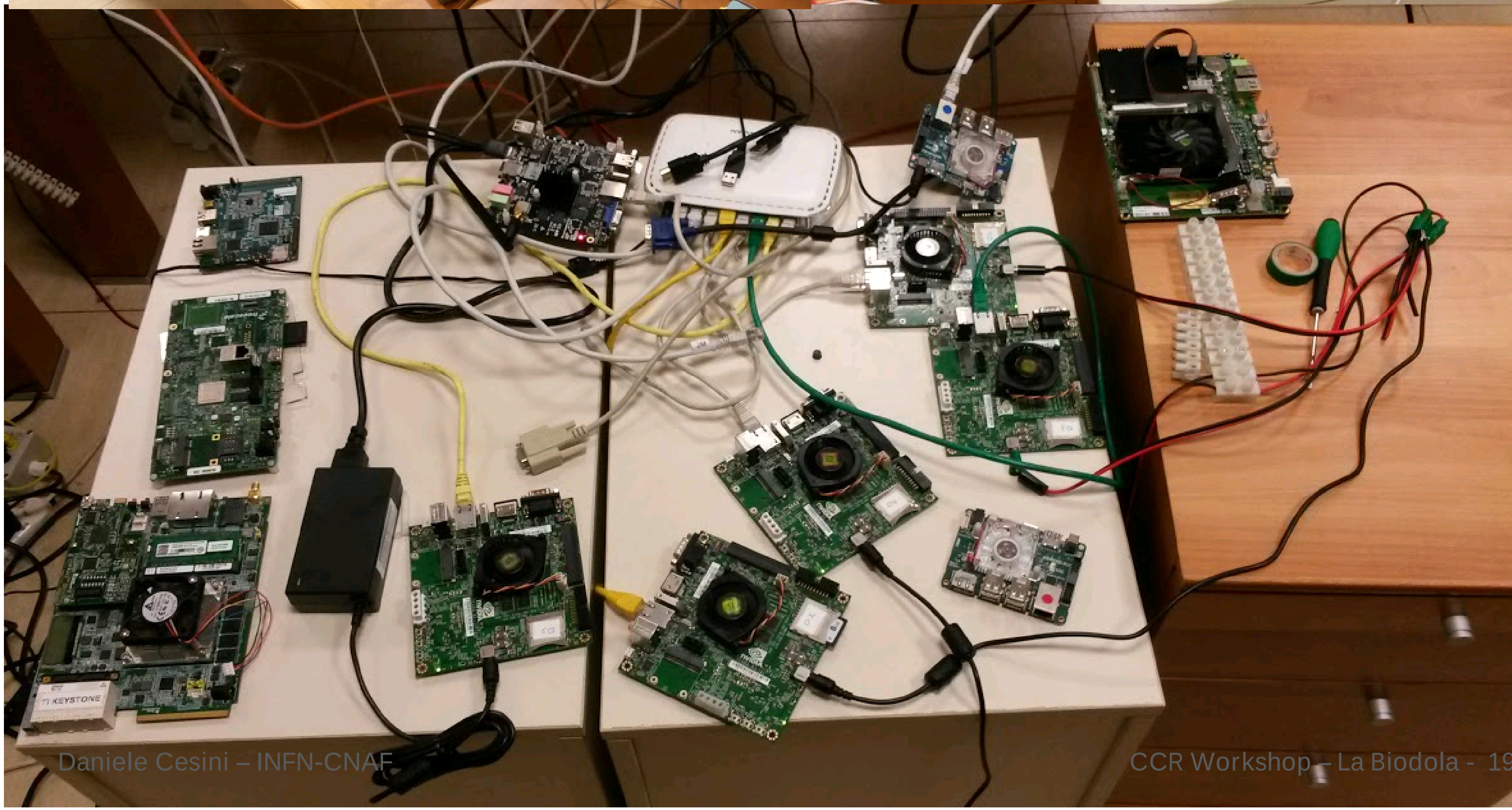
Product Name	Intel® Core™ m5-6Y54 Processor (4M Cache, up to 2.70 GHz)	Intel® Core™ i7-6500U Processor (4M Cache, up to 3.10 GHz)	Intel® Xeon® Processor D-1540 (12M Cache, 2.00 GHz)	Intel® Atom™ Processor C2750 (4M Cache, 2.40 GHz)
Code Name	Skylake	Skylake	Broadwell	Avoton
Essentials				
Status	Launched	Launched	Launched	Launched
Launch Date	Q3'15	Q3'15	Q1'15	Q3'13
Processor Number	M5-6Y54	i7-6500U	D-1540	C2750
Cache	4 MB Intel® Smart Cache	4 MB Intel® Smart Cache	12 MB	4 MB
Instruction Set	64-bit	64-bit	64-bit	64-bit
Instruction Set Extensions	SSE4.1/4.2, AVX 2.0	SSE4.1/4.2, AVX 2.0	AVX 2.0	
Embedded Options Available	No	No	No	No
Lithography	14 nm	14 nm	14 nm	22 nm
Recommended Customer Price	TRAY: $281.00	TRAY: $393.00	TRAY: $581.00	TRAY: $171.00
Datasheet	Link	Link	Link	Link
Product Brief	Link	Link	Link	
Scalability			1S Only	
Performance				
# of Cores	2	2	8	8
# of Threads	4	4	16	8
Processor Base Frequency	1.1 GHz	2.5 GHz	2 GHz	2.4 GHz
Max Turbo Frequency	2.7 GHz	3.1 GHz	2.6 GHz	2.6 GHz
TDP	4.5 W	15 W	45 W	20 W

CORE M — i7 Mobile — XEON D — AVOTON

Intel® HD Graphics 515/520

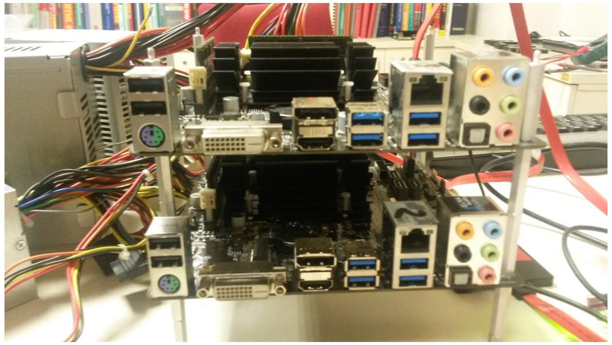






Low Power COSA Clusters@CNAF

5 mini racks



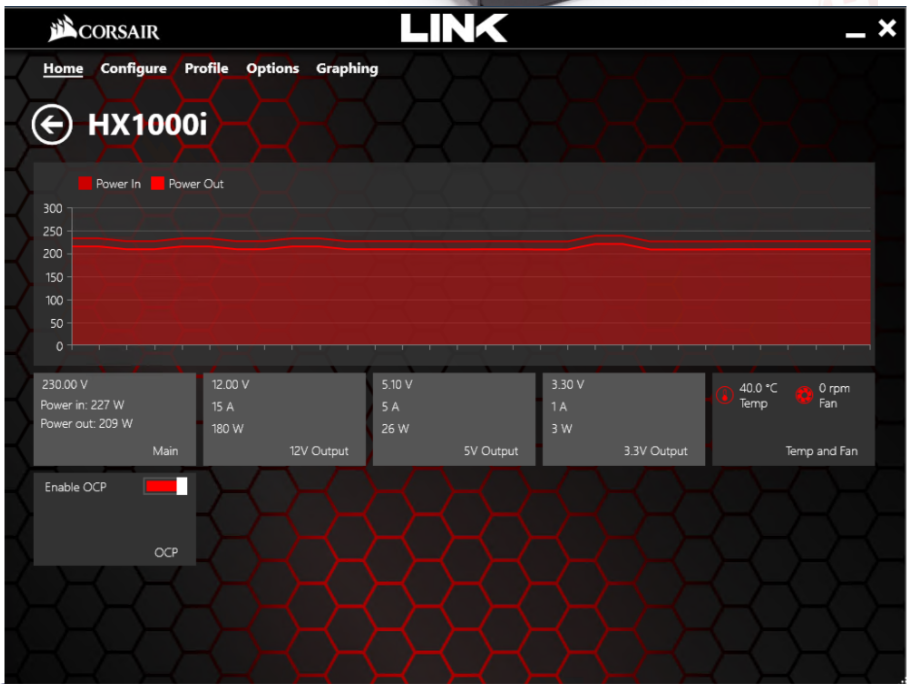
4xINTEL N3700

16xARMv7
2xARMv8

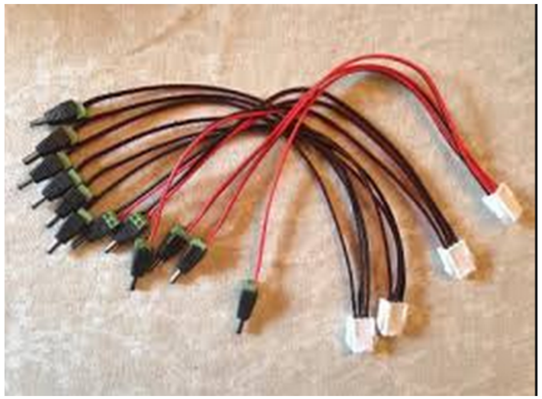
4xINTEL AVOTON C-2750
4xINTEL XEOND-1540

PSU&Cables

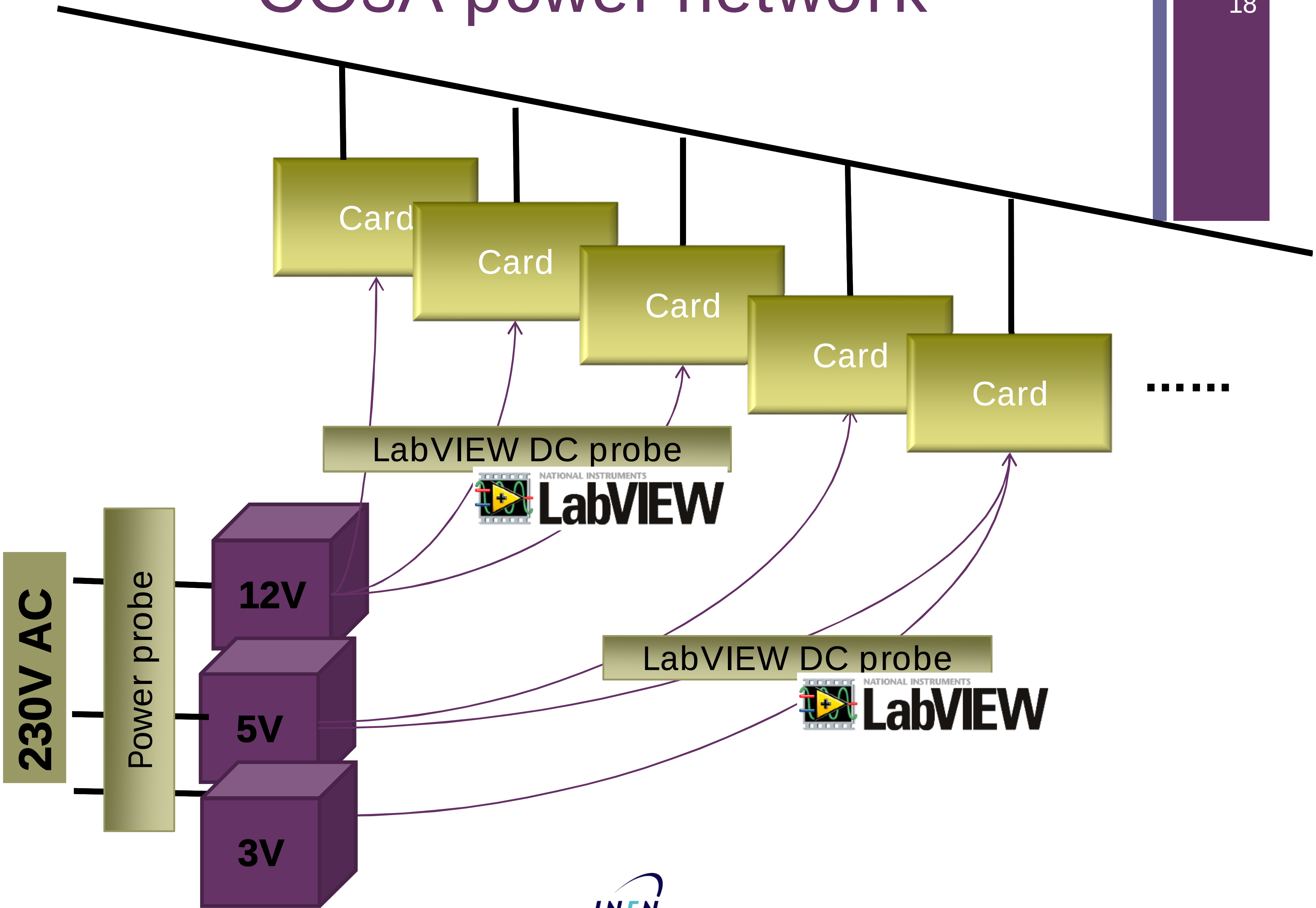
- PSU HX1000i
  - 12 lines@12V (Jetson+Intel)
  - 6 lines@5V (other boards)
  - 2lines@3V (n3700)



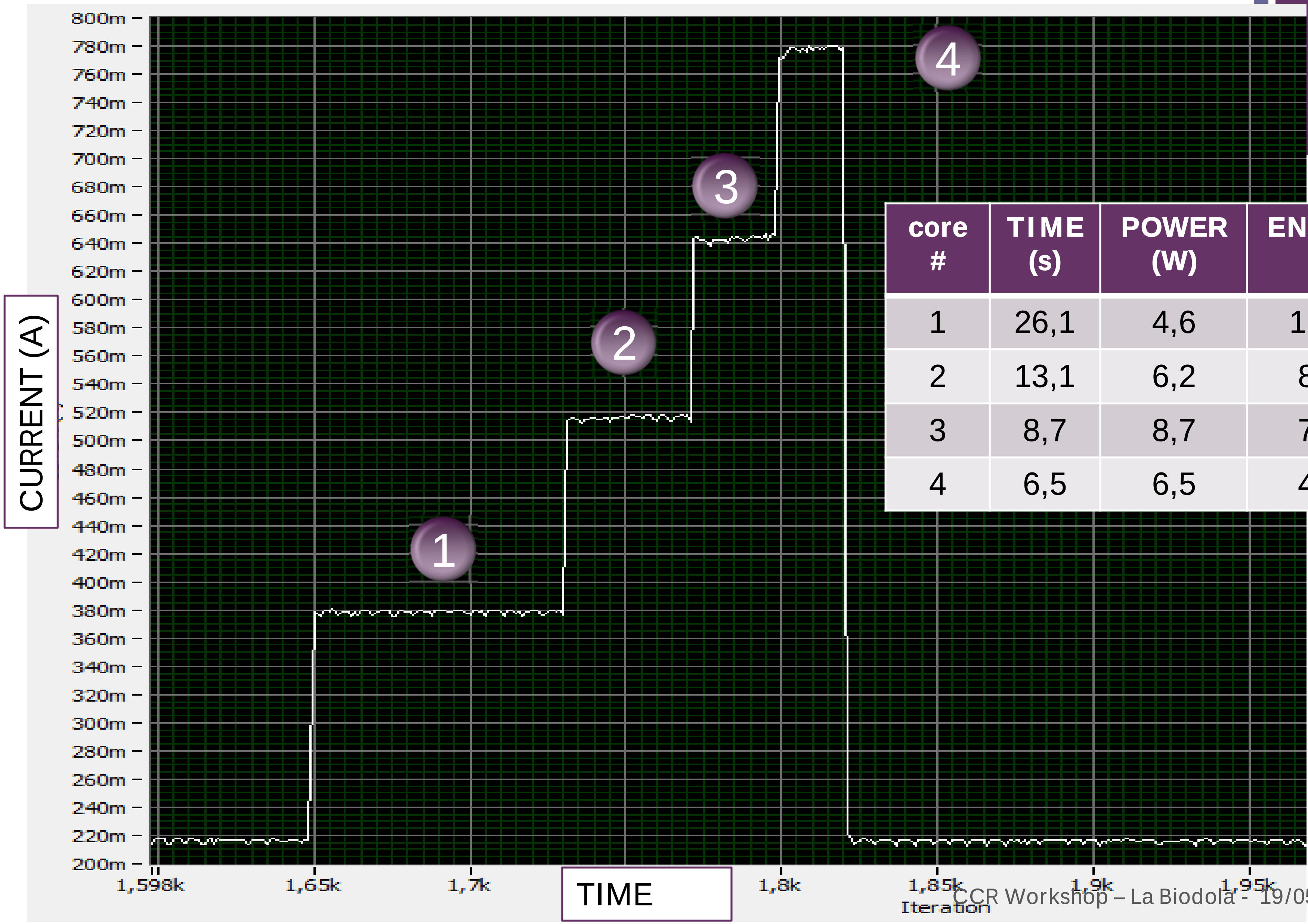
- GRIDSEED Cable
  - 1 MOLEX ➔ 3 BARREL



COSA power network



OPENMP π computation on Jetson



core #	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

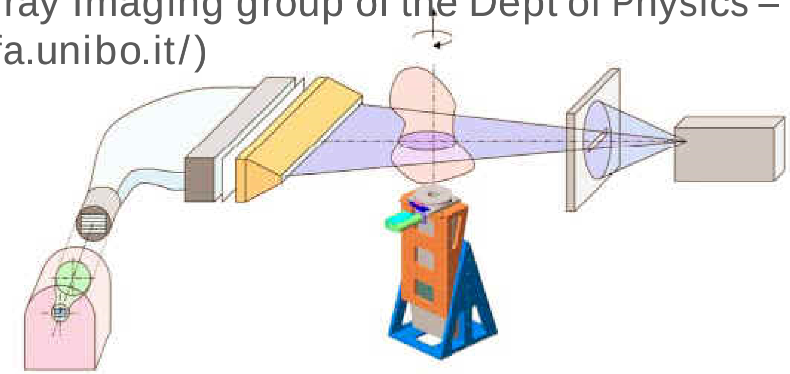
Applications

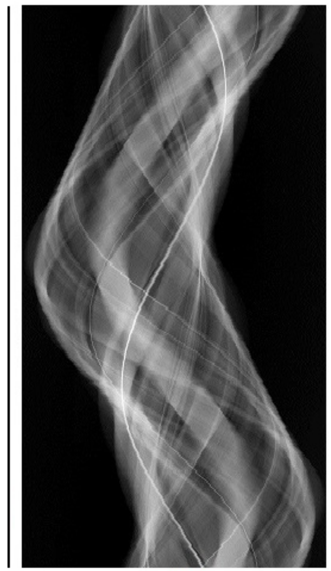
- Experimental Physics
  - Montecarlo and analysis of LHC experiments
  - HEP experiments High Level Trigger and Data Acquisition applications
  - Applications needing portable systems
    - Computer tomography
- Theoretical Physics
  - Parallel applications usually run in HPC environments
    - Relativistic astrophysic
    - Lattice Quantum ChromoDynamics simulations
    - Lattice Boltzmann fluid dynamics
    - Monte Carlo simulations of Spin-Glass systems
- Neural Networks
  - DPSNN-STDP code
- Synthetic Tests
  - HEPSPEC06
  - HPL

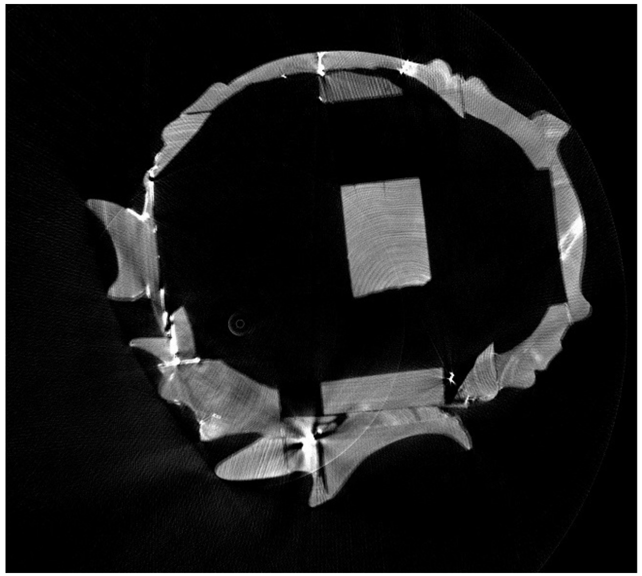
Computer tomography

Filtered Backprojection Algorithm

In collaboration with the X-ray Imaging group of the Dept of Physics – Bologna University (http://xraytomography.difa.unibo.it/)



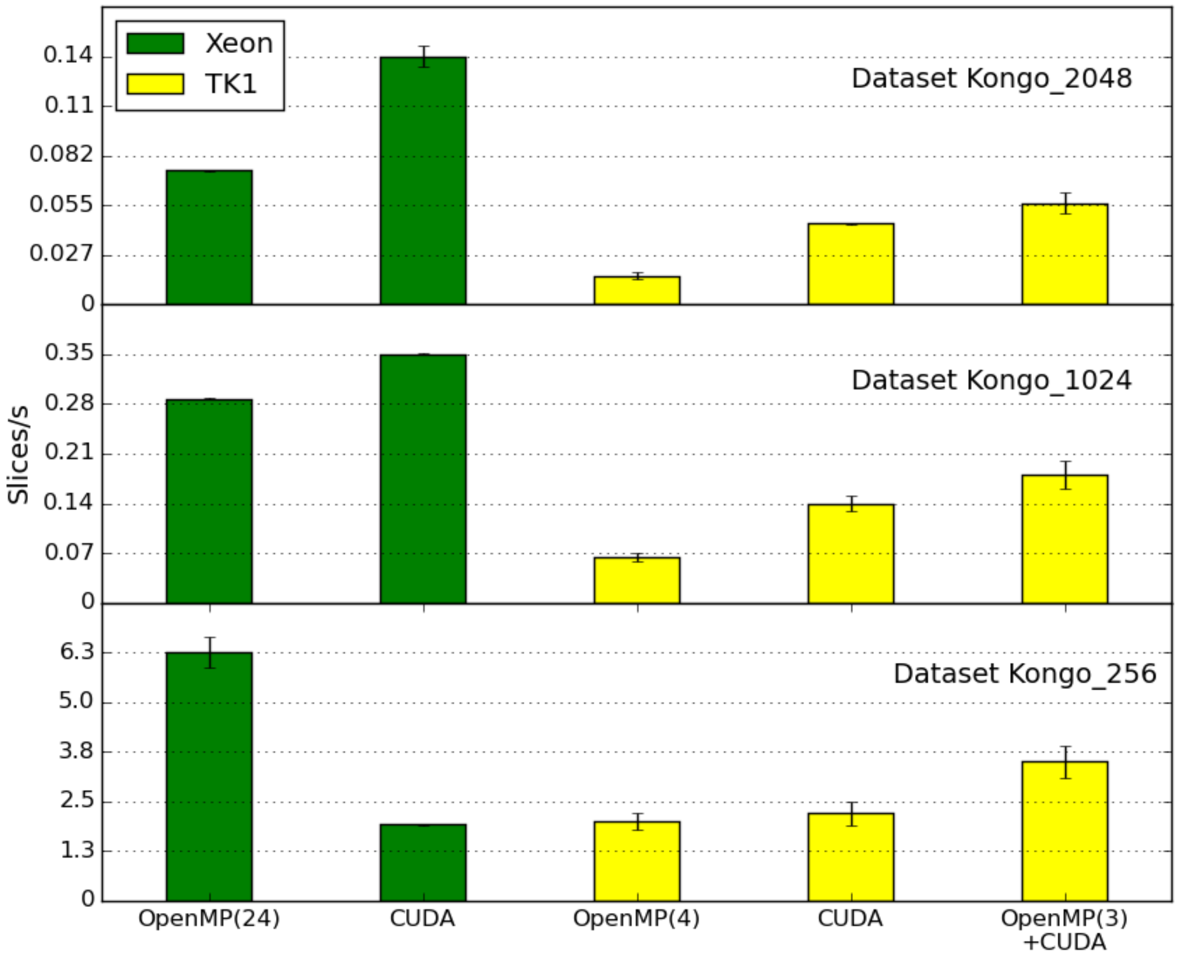




Real-Time Reconstruction for 3-D CT Applied to Large Objects of Cultural Heritage, R. Brancaccio, M. Bettuzzi, F. Casali, M. P. Morigi, G. Levi, A. Gallo, G. Marchetti, and D. Schneberk, IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 58, NO. 4, AUGUST 2011

FBP Algorithm - Productivity

Number of reconstructed slices for time unit

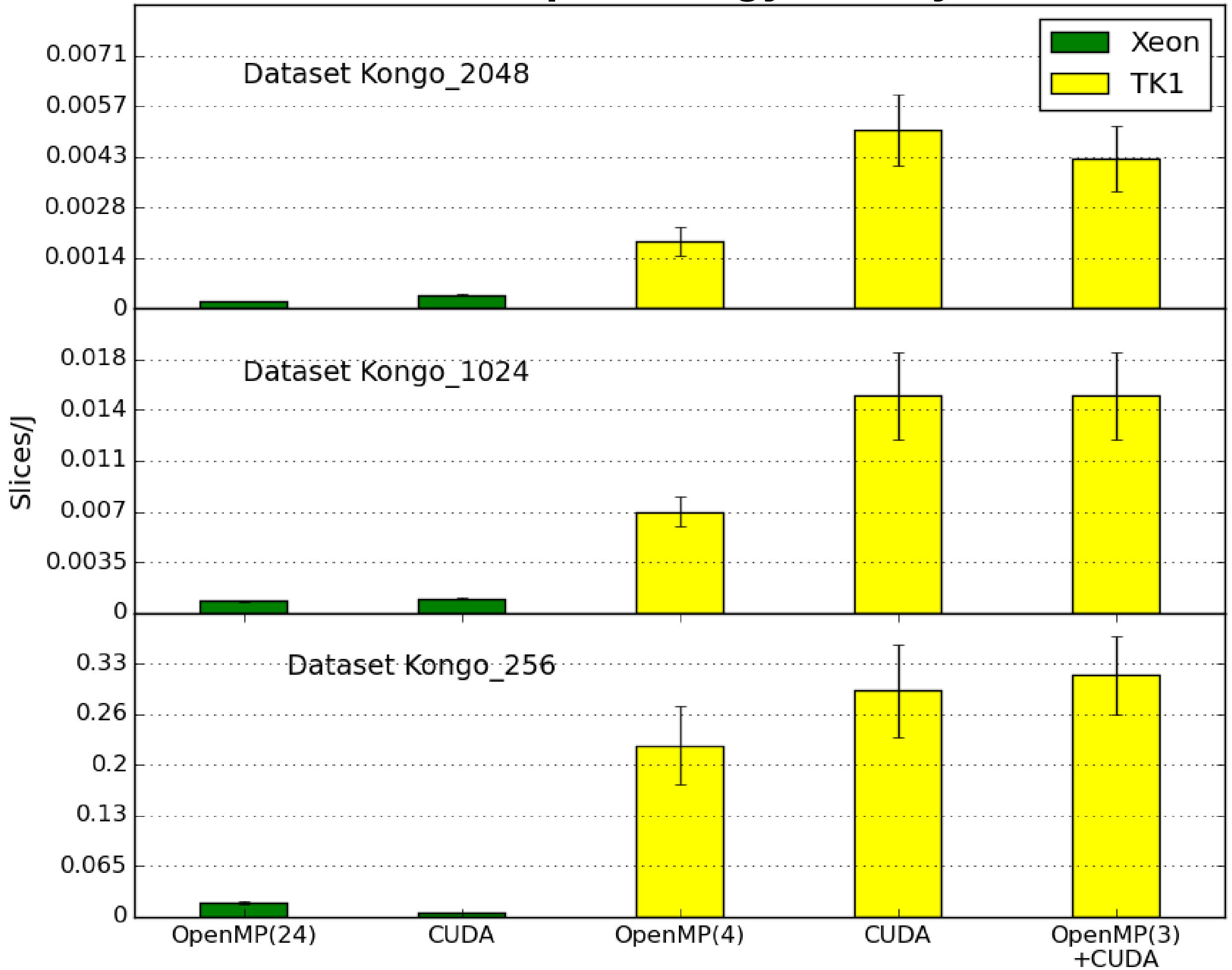


Xeon is a dual E5-2620 + NVIDIA K20
TK1 is the NVIDIA Jetson TK1

- Not surprisingly, the Xeon guarantees a higher speed than the SoC architecture
- The multi-threaded version of the algorithm is faster than the GPU version for small sizes of the slice when the application performances are broken by data transfer to and from device

FBP Algorithm - Energy efficiency

Reconstructed slices per energy unit by different runs



Xeon is a dual E5-2620 + NVIDIA K20
TK1 is the NVIDIA Jetson TK1

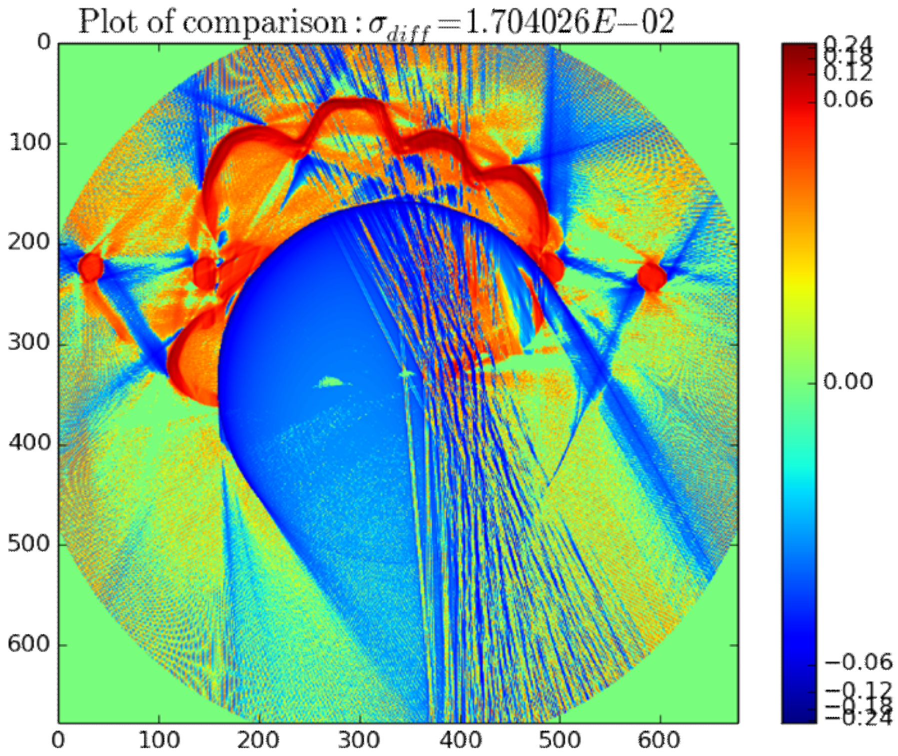
In one hour, considering the 1024x1024 slice and combining the CUDA and OMP runs on both architectures:

- **5** TK1:
  2340 slices consuming **41W**
- **2x**E5-2620+**1**K20
  2268 slices consuming **350W**

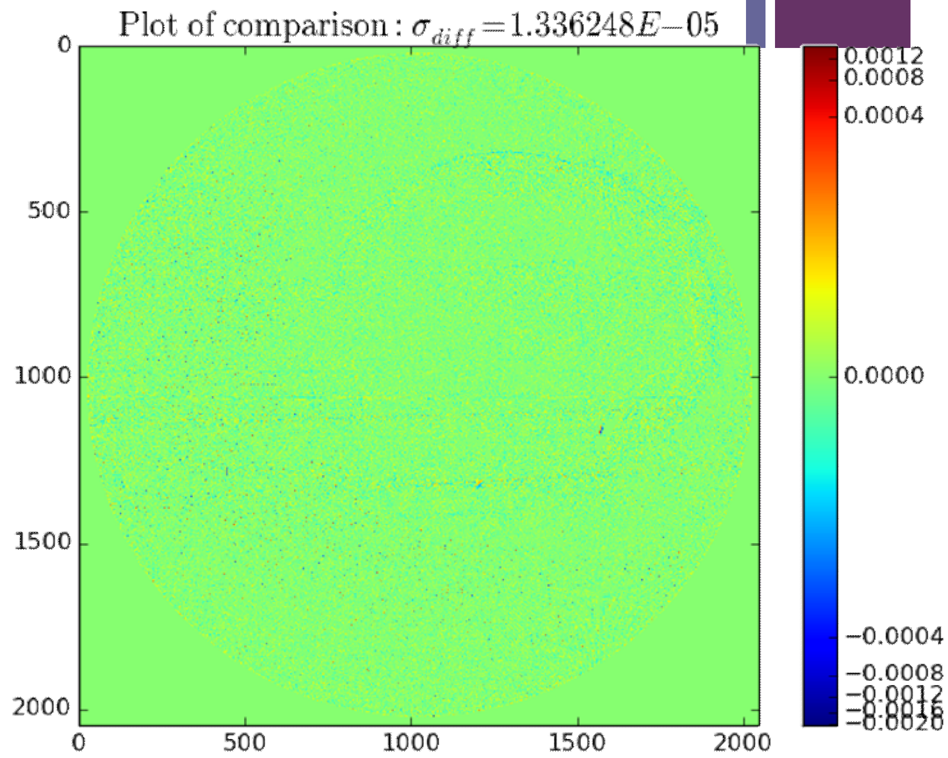
Master thesis of Elena Corni:
Implementazione dell'algoritmo Filtered Back-Projection (FBP) per architetture Low-Power di tipo Systems-On-Chip

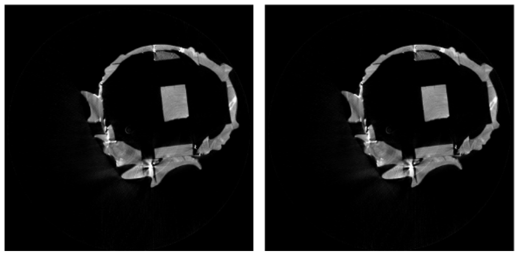
+ Numerical correctness

Wrongly reconstructed slice



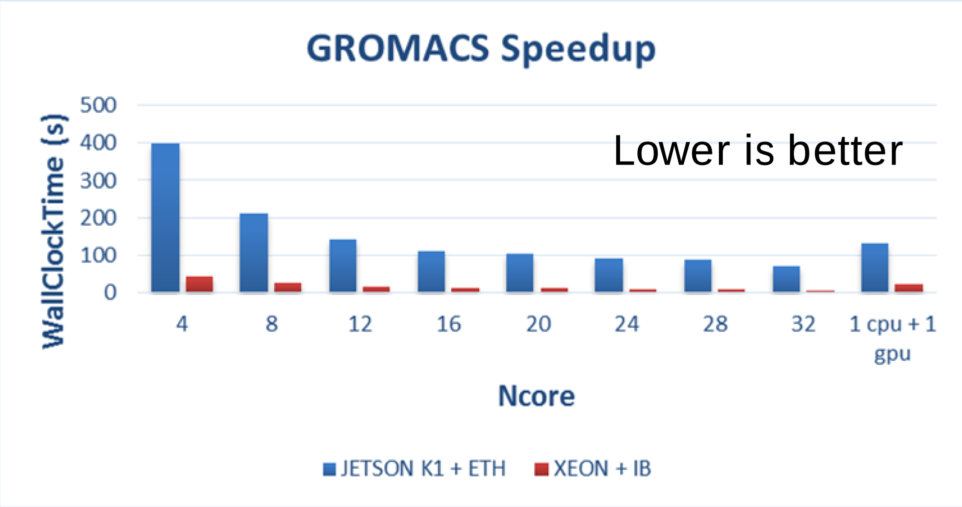
Accurately reconstructed slice

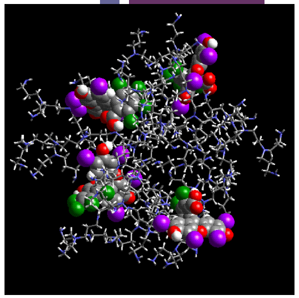


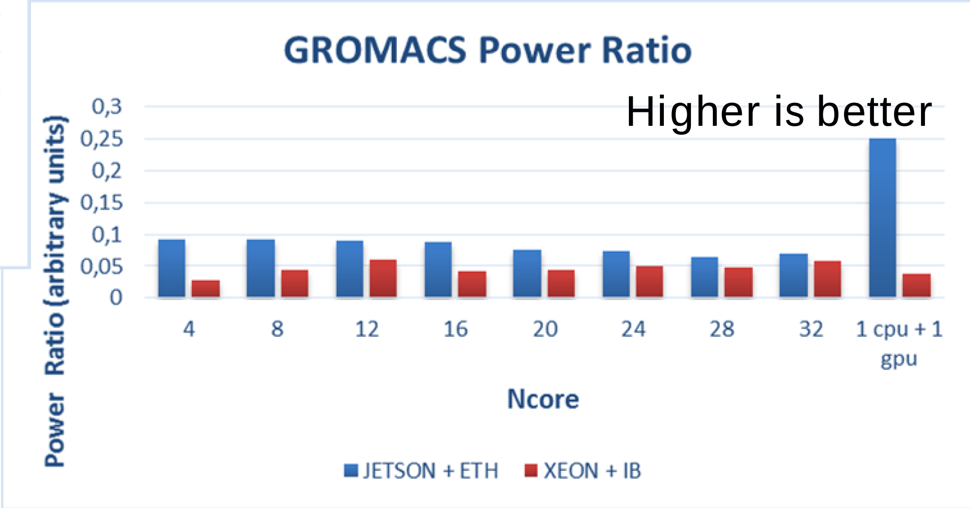


Molecular Dynamics on Jetson-TK1

Parallel application for CPU and GPU





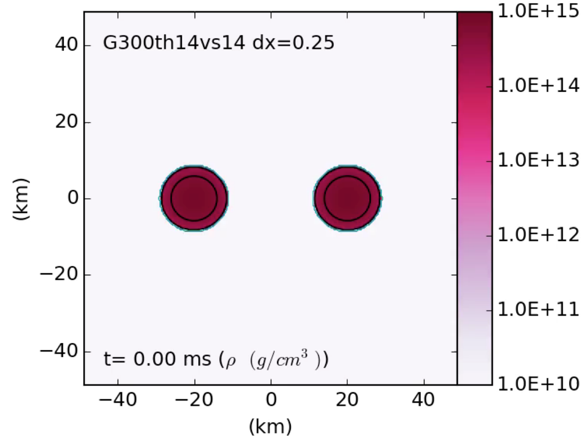


- Jetson-TK1 about 10X slower using the same number of cores
- Jetson-TK1 about 10X slower using the GPU (vs. an NVIDIA Tesla K20)
  - Jetson-TK1 13.5Watt
  - Xeon+K20 ~320Watt

+ # Einstein Toolkit

Roberto De Pietri , Roberto Alfieri - INFN Parma and Parma University

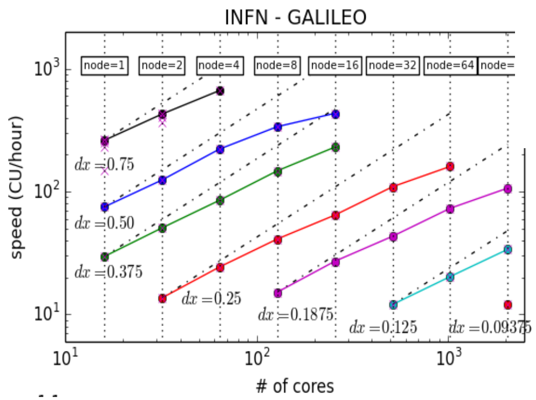
- **The scientific case: high resolution simulation of inspiral and merger phase of binary neutron stars system**
  - **one of source of the gravitational waves that are the observational target of the LIGO/VIRGO experiment**



- **Computation performed using The Einstein ToolKit**

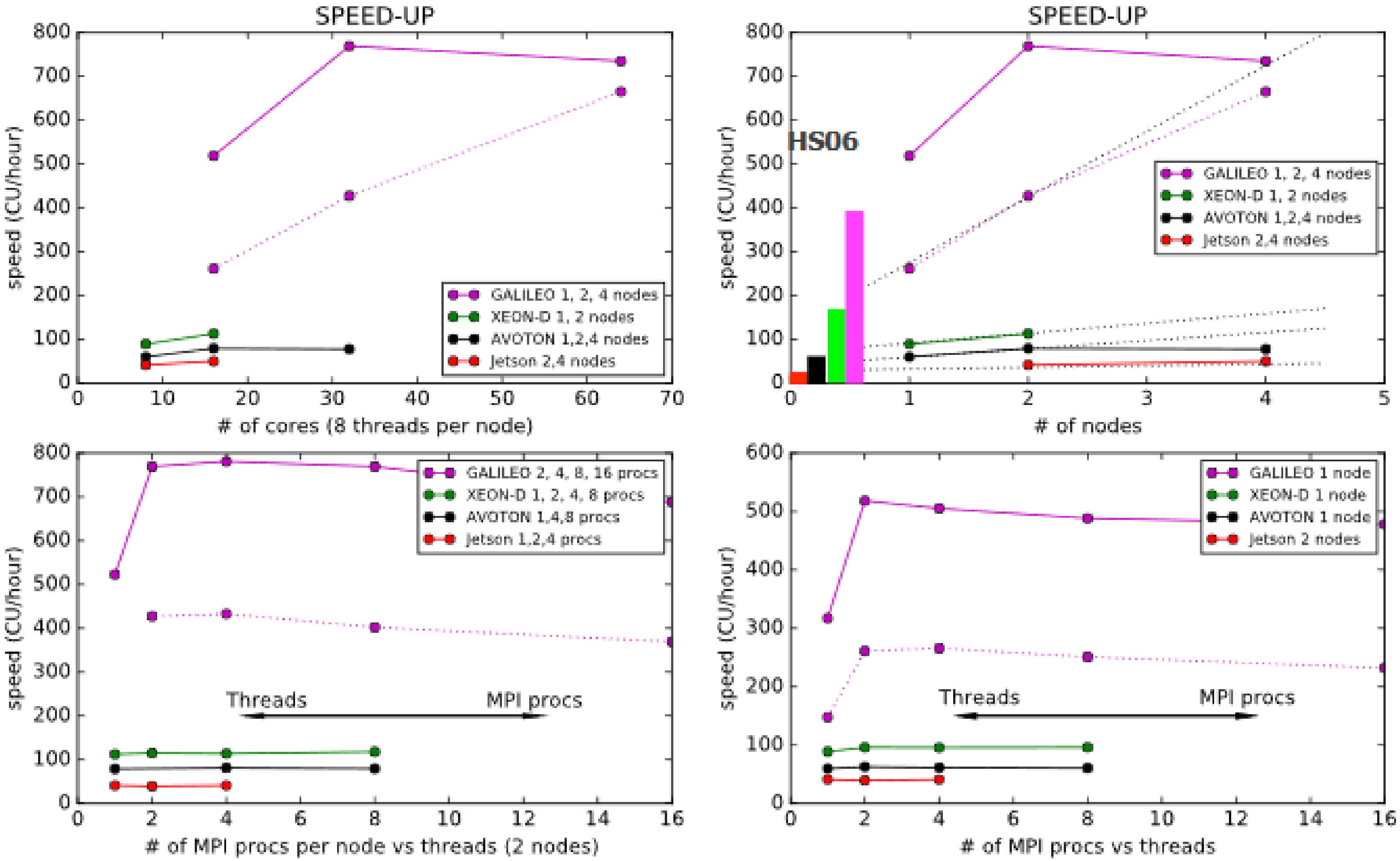
- **Result obtained on Galileo at CINECA**

- **COSA low power systems**
  - **Basic performance analysis**
  - **Porting of the application**
  - **Comparative results analysis**



Einstein Toolkit on COSA low power systems

Roberto De Pietri , Roberto Alfieri - INFN Parma and Parma University



https://agenda.infn.it/getFile.py/access?contribId=19&resId=0&materialId=slides&confId=10434

Einstein Toolkit on COSA low power systems

Roberto De Pietri , Roberto Alfieri - INFN Parma and Parma University

- GOOD NEWS: the framework **works** on LOW-POWER architectures

- BAD NEWS: performance not up to the par of traditional High-End Processor. Memory limitation would require an even higher number of nodes interconnected with **a high speed network**

- In order to run our application on Low Power architectures at production level **we need to exploit the accelerator present on the system (GPU)** in order to speed up the computation

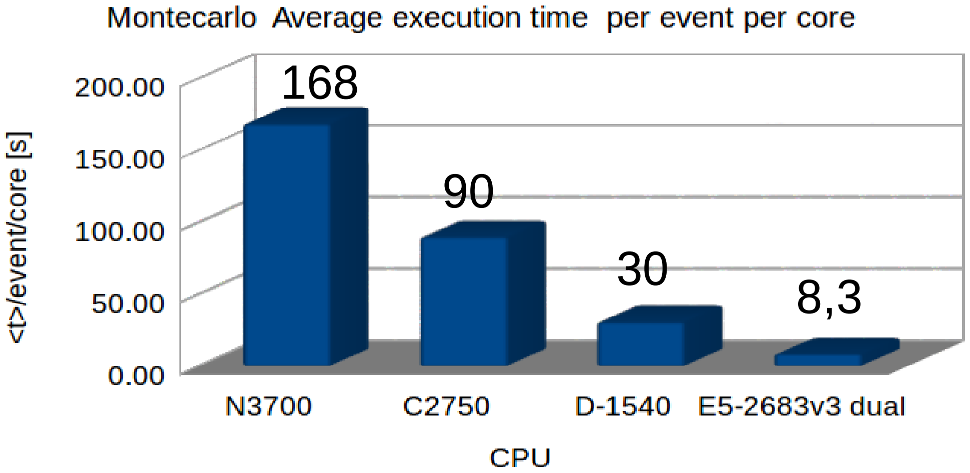
….but…..

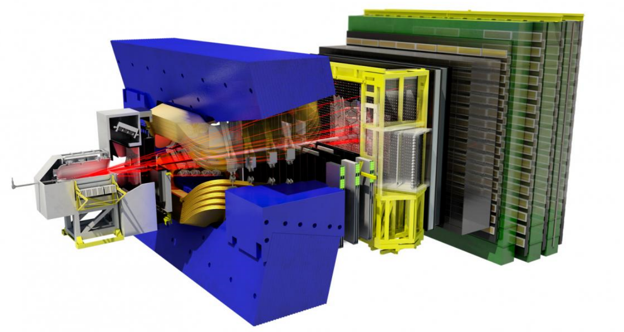
- We cannot change the code for every new hardware device

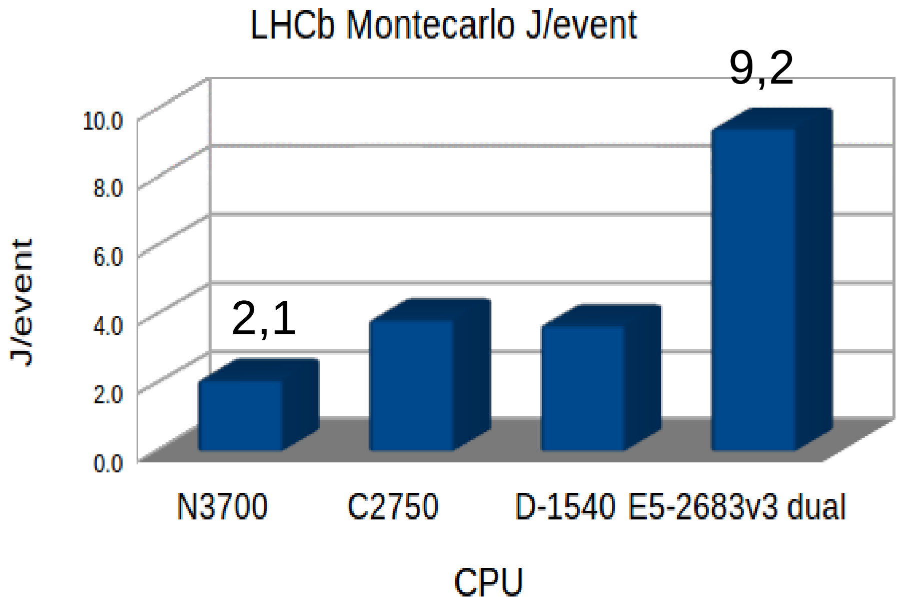
LHCb Montecarlo software test

A.Falabella@INFN-CNAF

Gauss: event simulation + detector description based on Geant



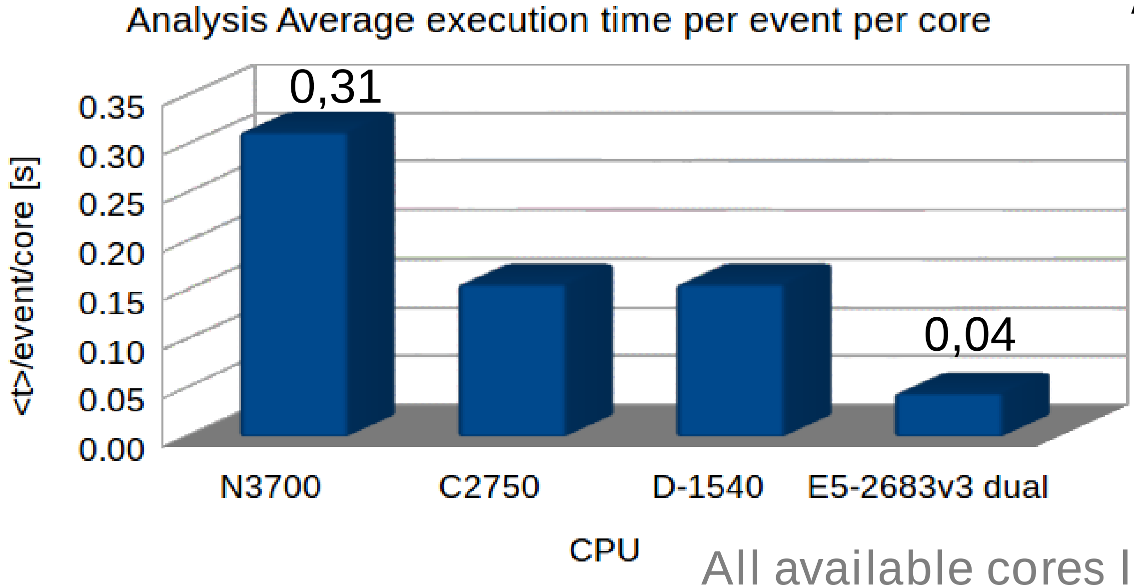




- Porting was not difficult
  - Just recompilation
- All the platform can provide enough RAM per core for the LHCb sw

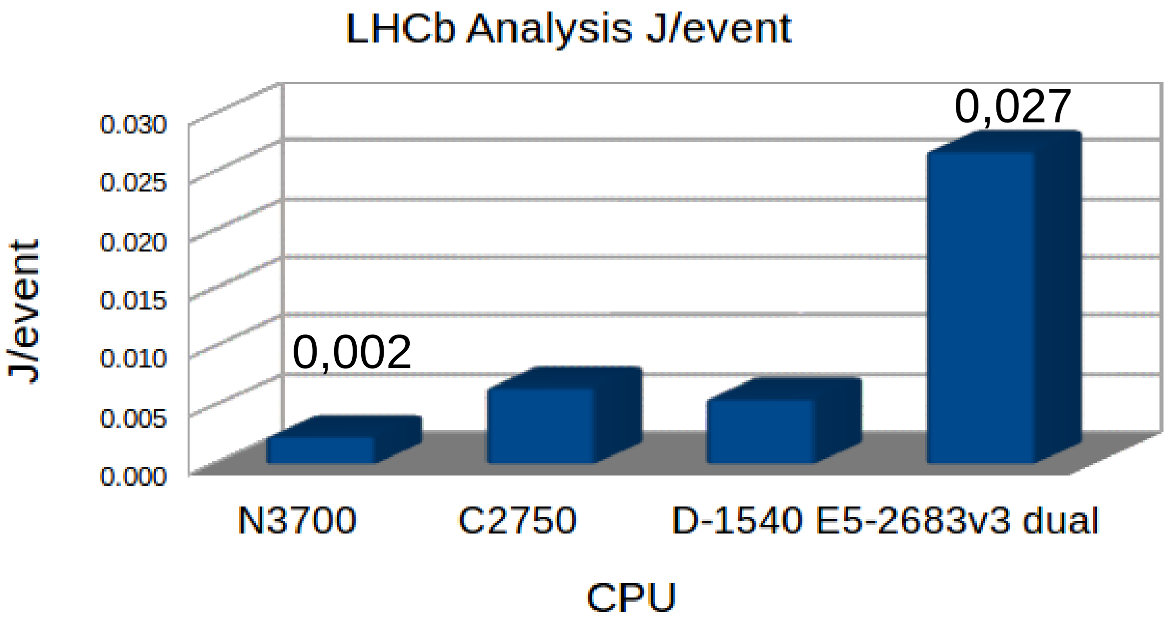
LHCb Analysis software test

A.Falabella@INFN-CNAF



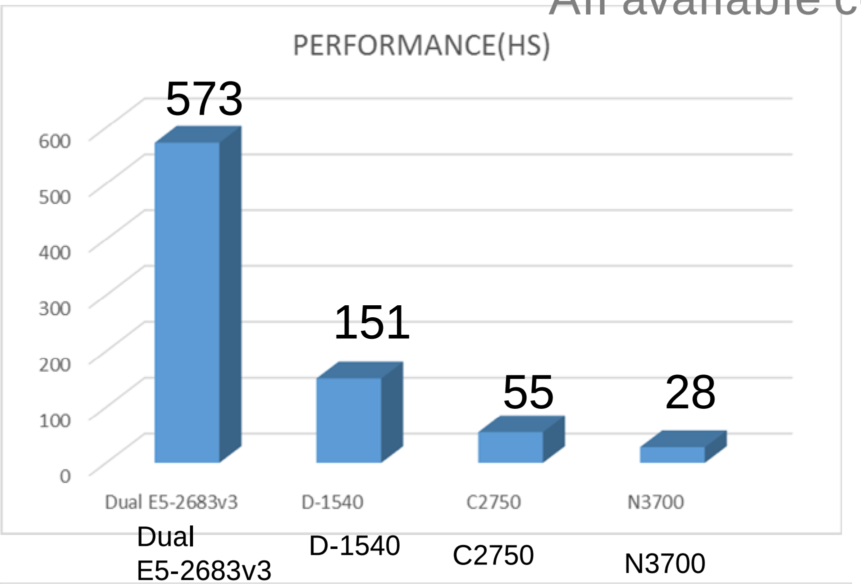
Brunel: offline reconstruction algorithms

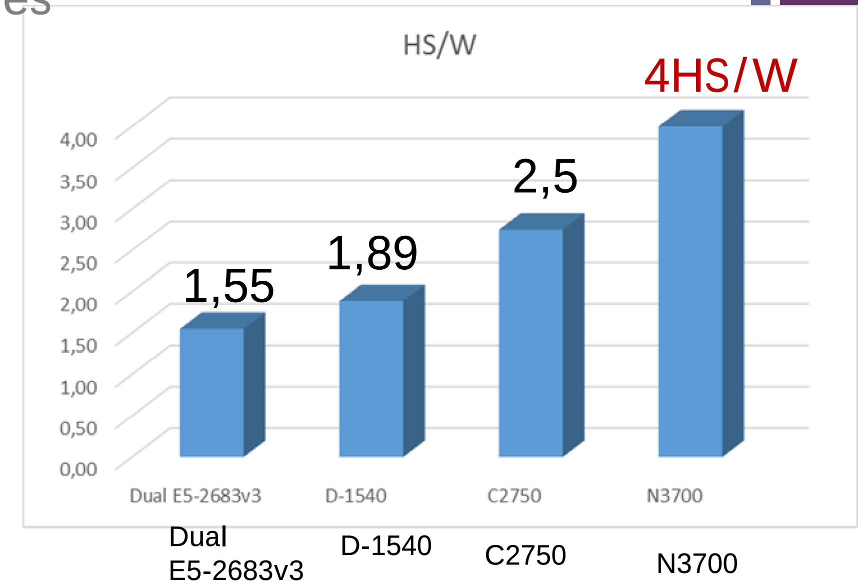
All available cores loaded

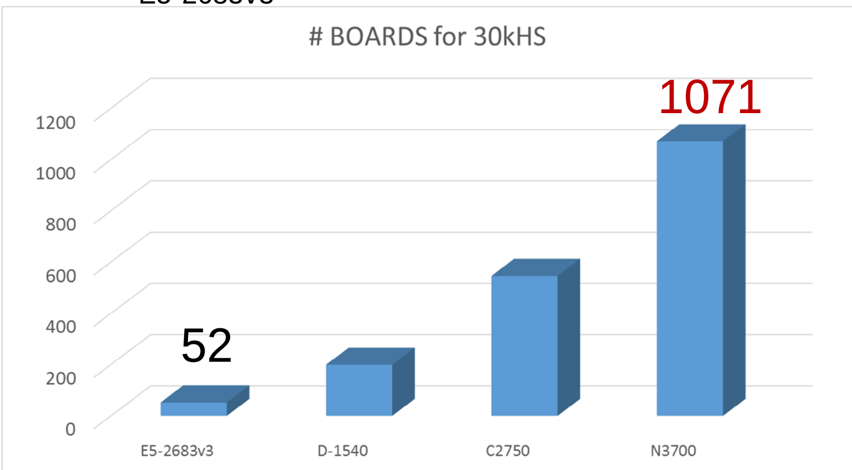


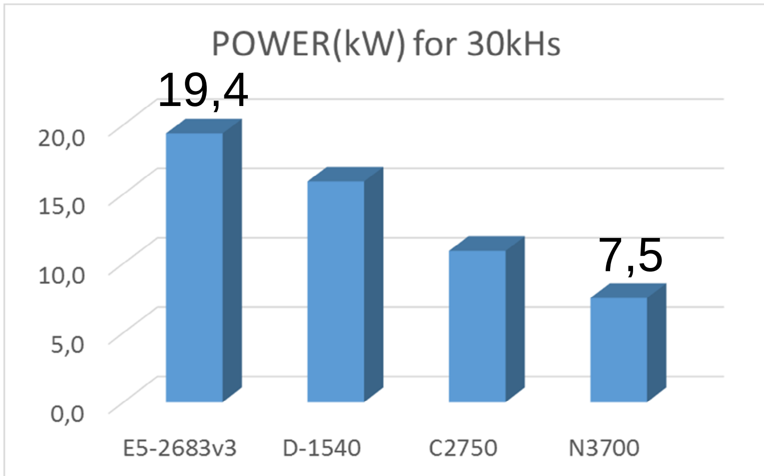
HS06 on Intel platforms

All available cores









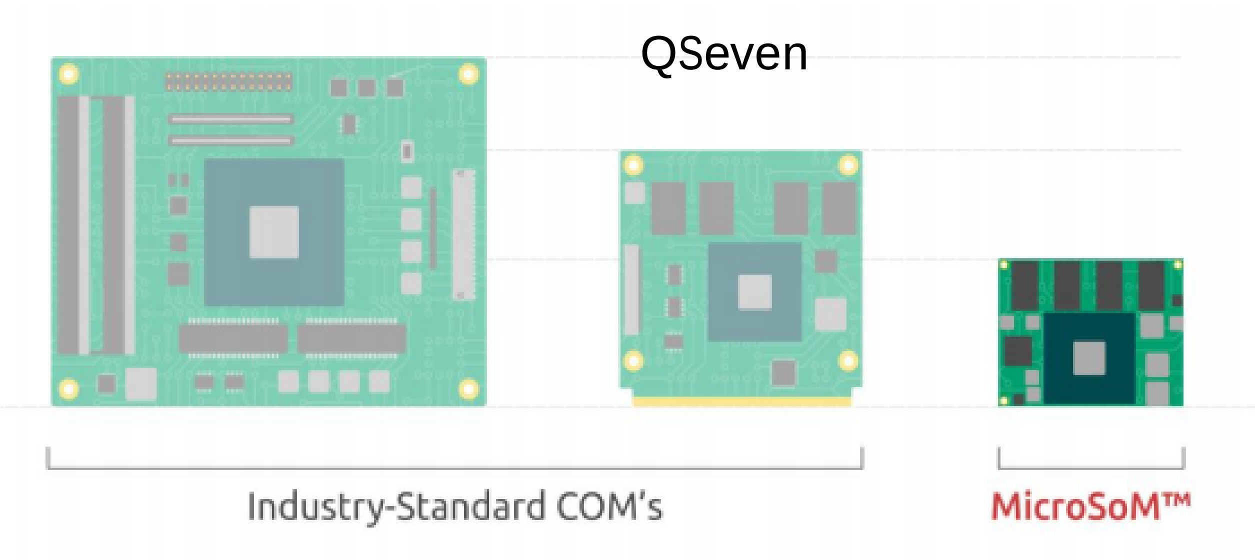
EXTREME INTEGRATION NEEDED!!

Integration with SoMs

- Can we integrate something using existing SoMs ?
  - Similar to HP moonshot with off-the-shelf components
  - Standard communication links
    - Easy to change the SoMs

PCI COM Express 10

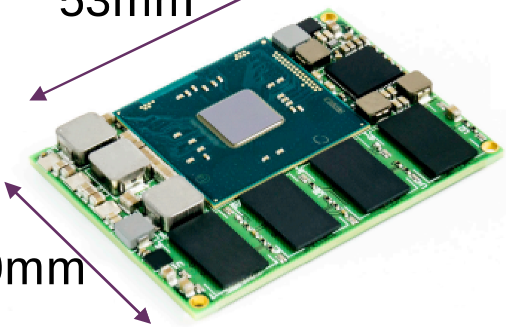
QSeven



SOMs and carriers with PENTIUM N3700

53mm

40mm



Braswell MicroSoM

Pentium N3710, 8GB, 1Gbe, SATA, PCI2

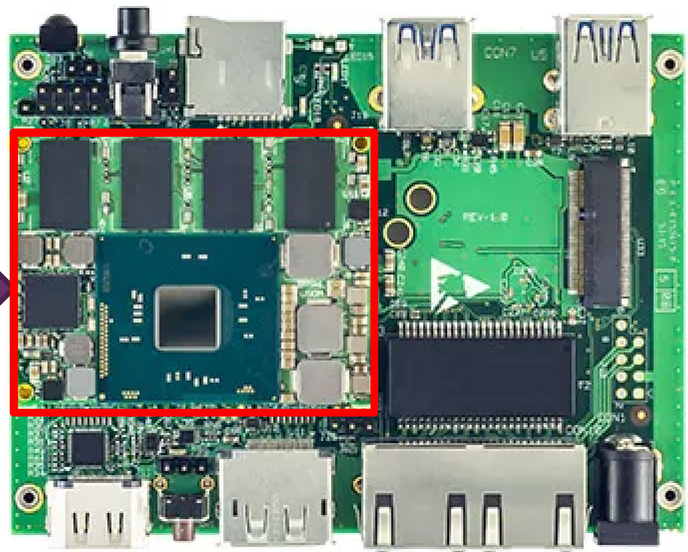
SolidRun

Do Your Thing

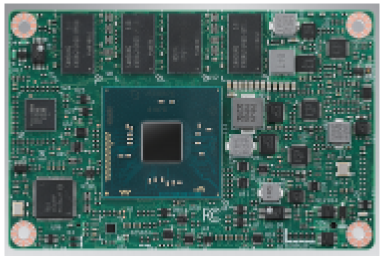
Proprietary bus

100mm

80mm



Carrier Board

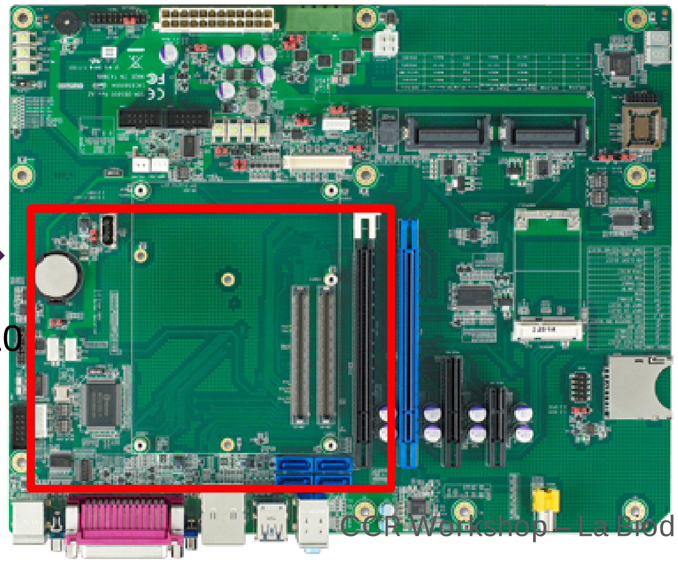


SOM-7568

Pentium N3710, 8GB, 1Gbe, SATA, PCI2

ADVANTECH

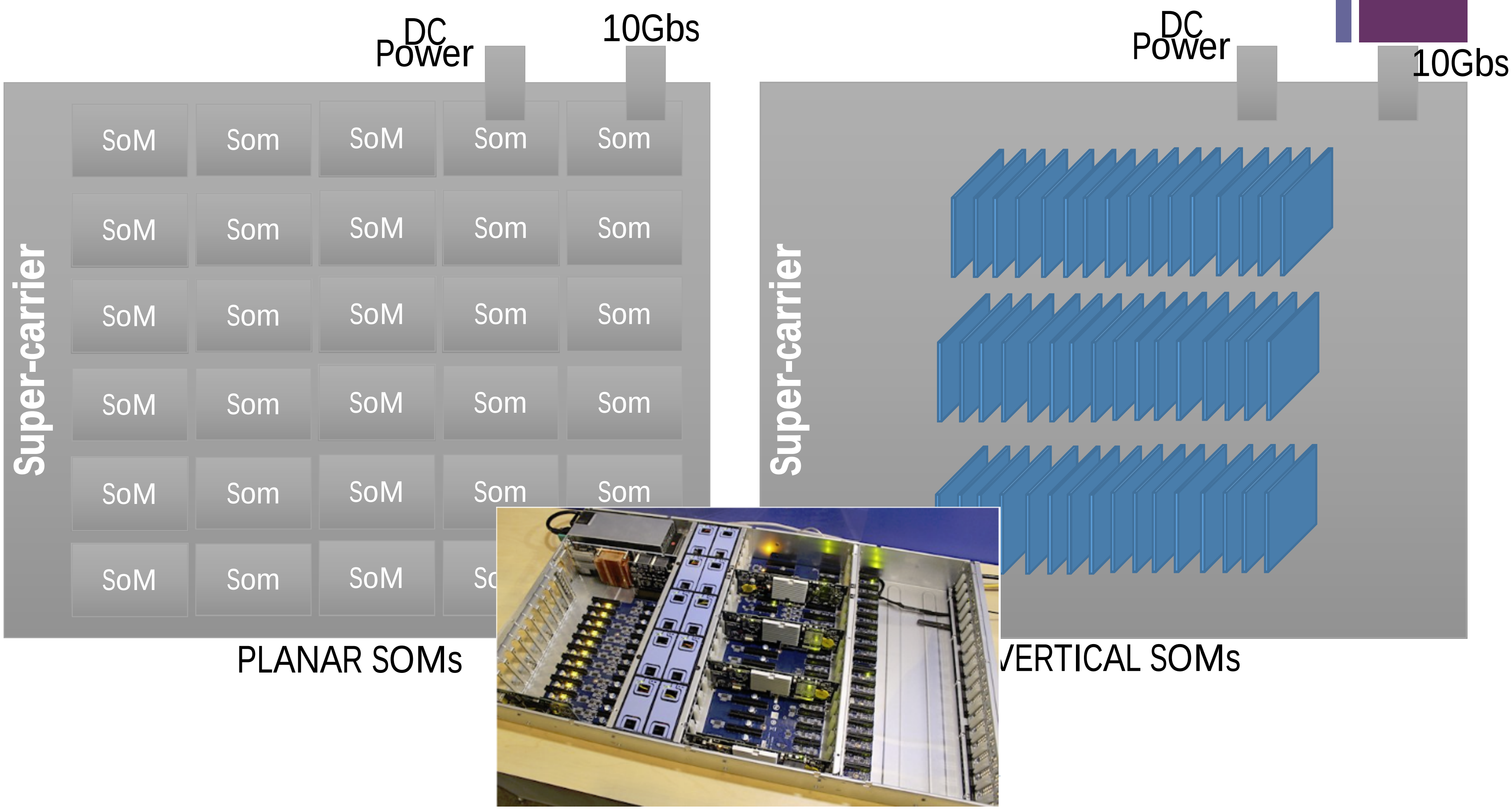
COM Express® Type 10



ATX form factor

SUPER-CARRIER (350W, 1344HEPSPEC)

- 48 SOMs (every SOM: 7W, 28HEPSPEC)
- 48 2.5'' HDDs (every HDD is face-to-face with every SOM)
- DC Power connector
- 10 Gbit connector

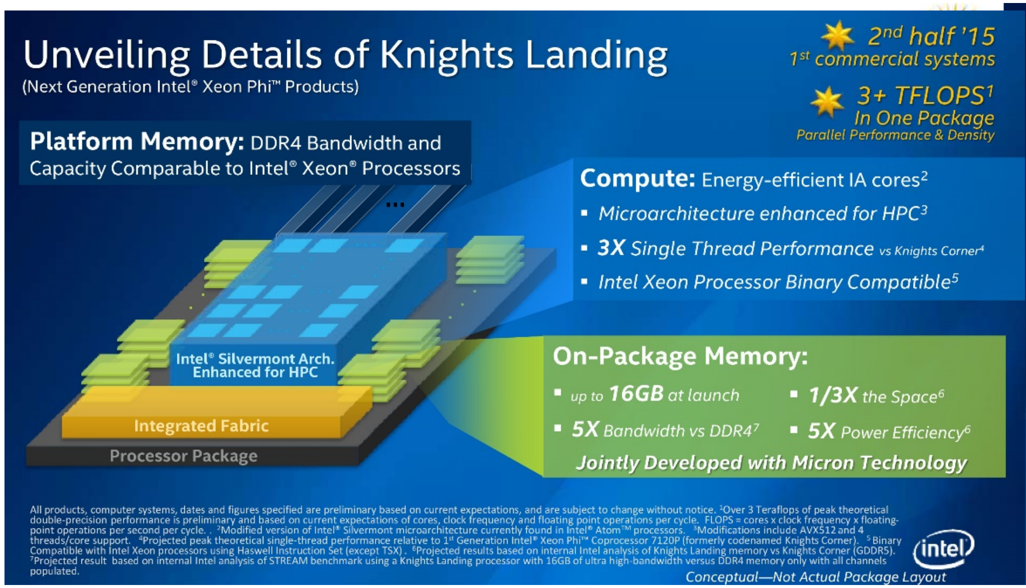


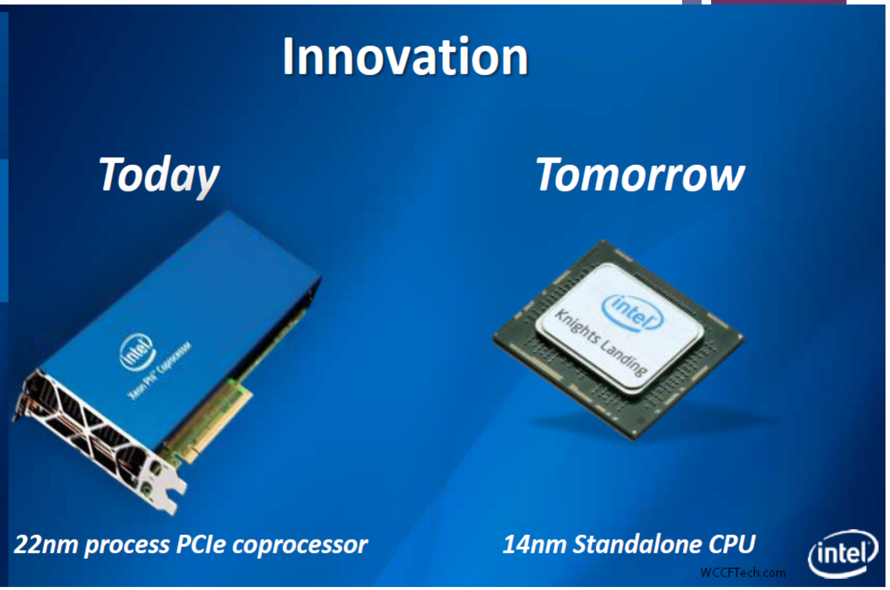
PLANAR SOMs

VERTICAL SOMs

NEC 368 avoton core in 2U

Is the same of few KNL XEON-PHIs?

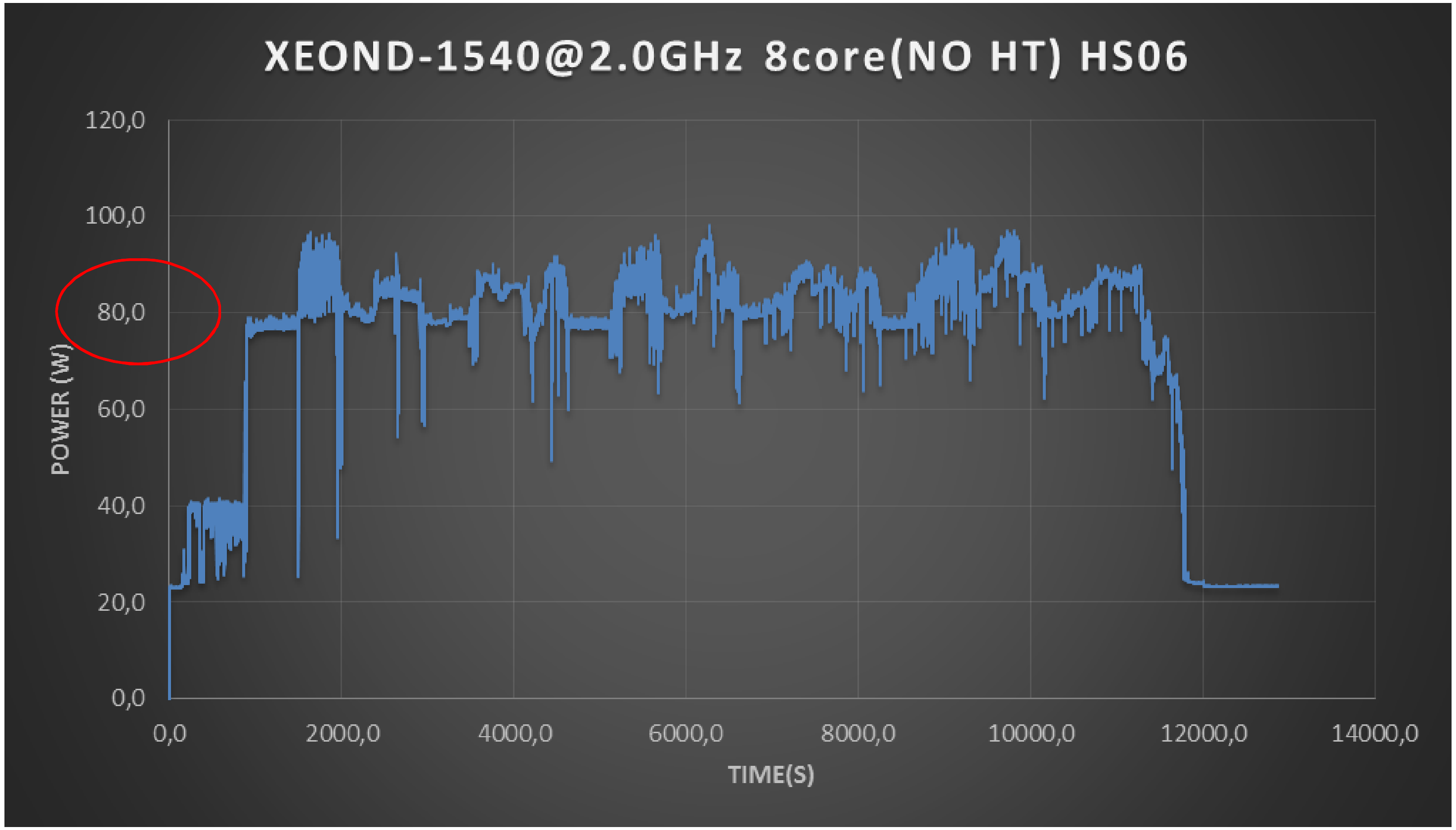




- up to 72 Airmont (Atom) cores with four threads per core
- up to 384 GB of "far" DDR4 RAM and 8–16 GB of stacked "near" 3D MCDRAM
- Each core will have two 512-bit vector units and will support AVX-512 SIMD instructions
- Standalone version
- Onmipath fabric integrated (2x100gbs)

XEON D-1540 HS06 benchmark

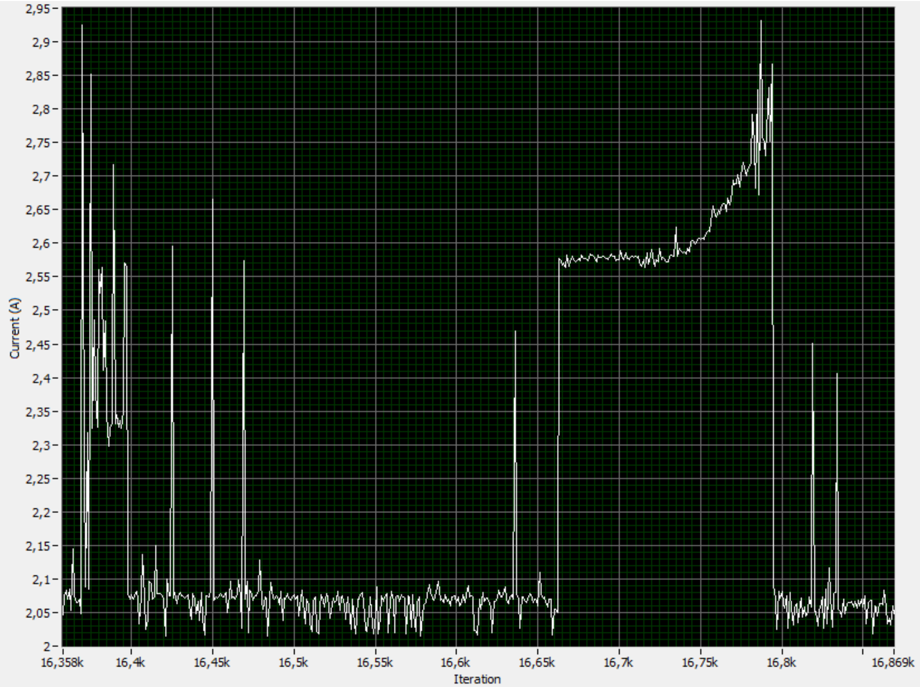
- Based on the all_cpp benchmark subset of the SPEC® CPU2006 benchmark suite
- Widely used in INFN tenders



Netpipe test Power Consumption @10Gbs

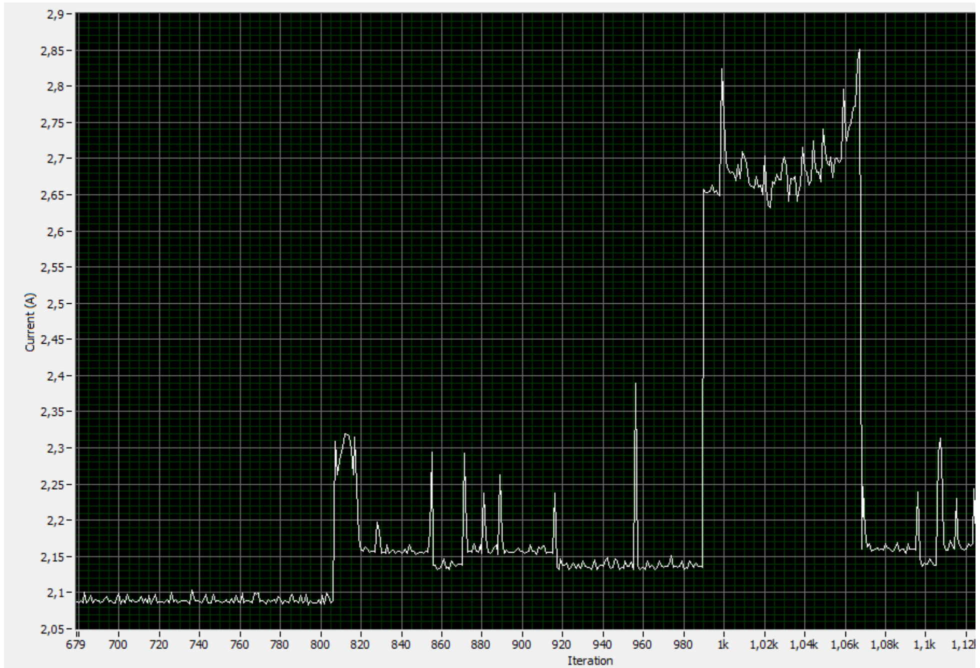
- Ondemand governor
- C-STATE enabled
- NO HyperThread

- 10Gb PCI HBA



Bandwidth 9.0Gb/s
Latency 26.0 us
Max 33W

- 10Gb XEOND integrated

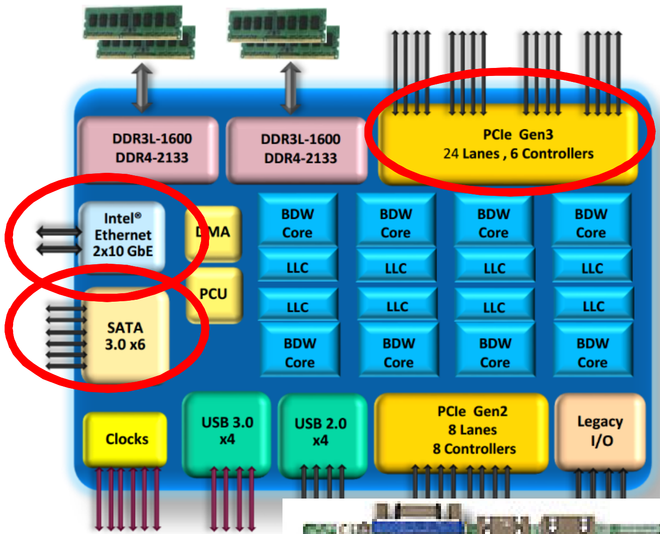


Bandwidth 8.8Gb/s
Latency 24.0 us
Max 34W

XEON D-1540 for storage bricks

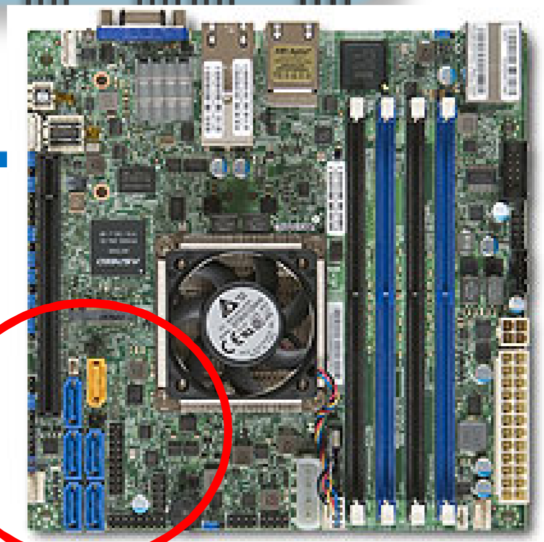
Intel® Xeon® Processor D - SoC Architecture

CPU	2-8 Core Intel® Xeon™ (14nm) CPUs
L1 cache	32K data, 32k instruction per core
L2 cache	256K per core
LLC cache	1.5MB per core
Addressing	46 bits physical / 48 bits virtual
Memory	DDR4 up to 2133 MT/s DDR3L up to 1600 MT/s Two Channels (2 DIMMs/Channel)
Memory Capacity	RDIMM: 128 GB (32 GB/DIMM) UDIMM/SODIMM: 64 GB (16 GB/DIMM)
DIMM Types	SODIMM, UDIMM, RDIMM with ECC and non-ECC
Memory RAS	Enhanced ECC Single bit Error Correction – Dual bit Error Detection (SEC-DED) covers address and data paths, DDR scrambler to reduce error rate.
PCI-E*	x24 PCIe Gen3 with up to 6 controllers x8 PCIe Gen 2 with up to 8 controllers
Integrated IO	Intel® Ethernet 2x10 GbE , x4 USB 3.0, x4 USB 2.0, and x6 SATA 3
Technologies	Intel® VT, Core RAPL, PECI over SMBUS, PSE
Power Management	FIVR, PCPS, EET, UFS Hardware PM
Legacy I/O	SPI for boot flash, SMBus, UART LPC, GPIO, 8259, I/O APIC, 8254 Timer, RTC



Supermicro X10SDV-TLN4F

Performance	
# of Cores	8
# of Threads	16
Processor Base Frequency	2 GHz
Max Turbo Frequency	2.6 GHz
TDP	45 W



Broadwell-DE, Xeon D 8-Core

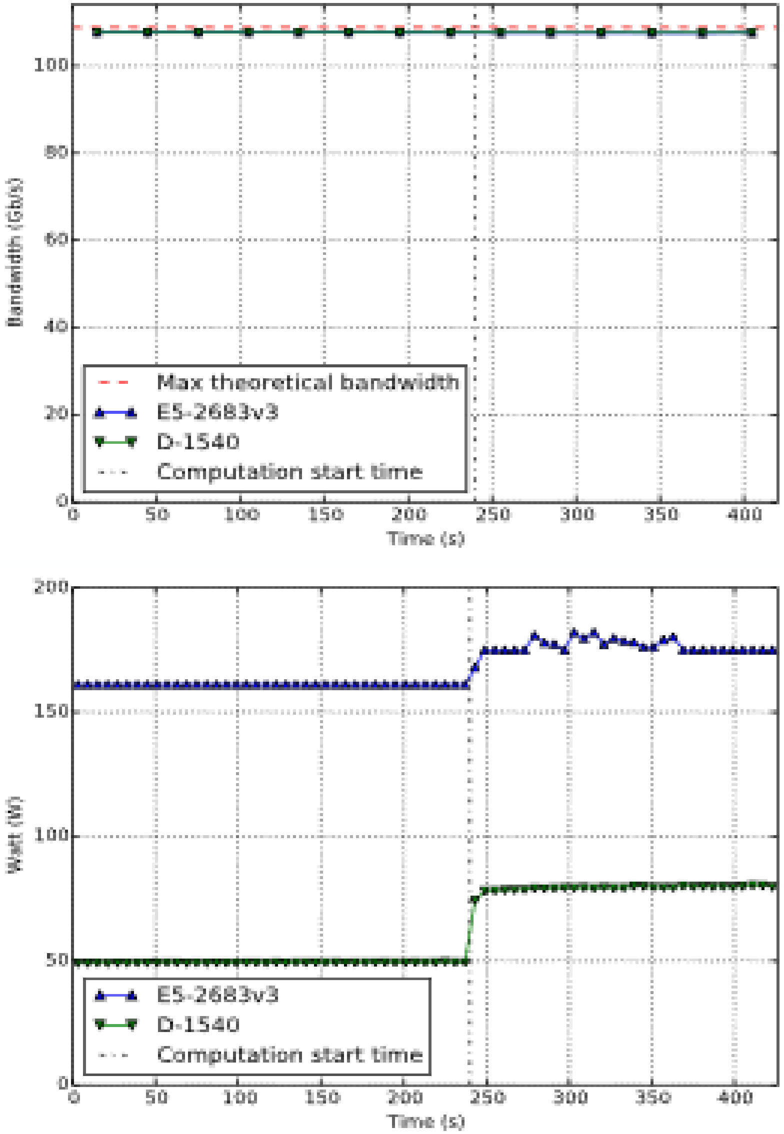
HDD/SSDs

INFN Istituto Nazionale di Fisica Nucleare

LHCb Event Building on the XEOND

Matteo Manzali – INFN-CNAF & UniFe

- Sw designed to simulate the event building on InfiniBand based network
  - It relies on the verbs library to perform RDMA operations
- In the second part of the test a pure computation process is started on four cores
  - in order to simulate a software trigger
- The performances of the Event Builder are comparable, but the D-1540 requires a third of the power consumption of the E5-2683v3



Conclusion

- COSA is testing two types of SoCs
  - Low-Power SoCs from the mobile/embedded world
    - still have many limitations for a production environment
  - Low Power SoCs from the server world
    - very expensive and in some cases not really low power
    - 10Gbs/Infiniband networking easier to obtain
- SoCs are becoming attractive for real life scientific applications
  - In particular if you manage to extract power from the integrated GPU
  - CPU porting was easier than expected
- Low-power/low cost dominated by ARM until last year, now INTEL is becoming competitive in this segment
  - No porting required for the CPU
- Advanced HW integration needed to maintain a reasonable size of the system