Low Power Computing with modern Systems on Chip

Daniele Cesini – INFN-CNAF
Andrea Ferraro – INFN-CNAF
Lucia Morganti – INFN-CNAF
…and many others….
www.cosa-project.it

Outline

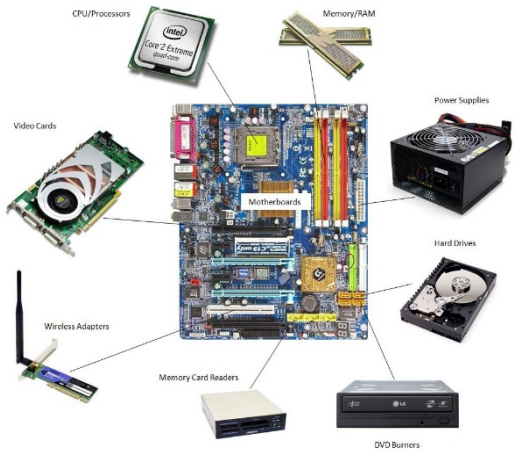
- Introduction
- A little bit of history
- Modern Systems on Chip
- Computing on System on Chip, the INFN experience
- European Projects on low power computing
- Conclusion

Outline

- Introduction
- A little bit of history
- Modern Systems on Chip
- Computing on System on Chip, the INFN experience
- European Projects on low power computing
- Conclusion

Glossary: Low, Power and Computing

- POWER: electrical power in Watts (W)
- COMPUTING: We will focus on computation (CPU+GPU)
  - But computing is much more

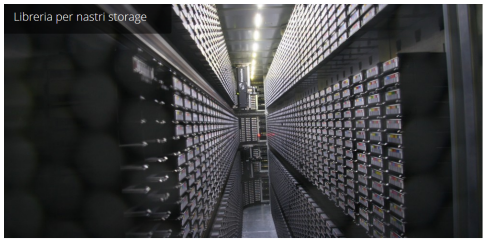


- LOW: Low Power, Low Cost
  - We need a context

A possible context

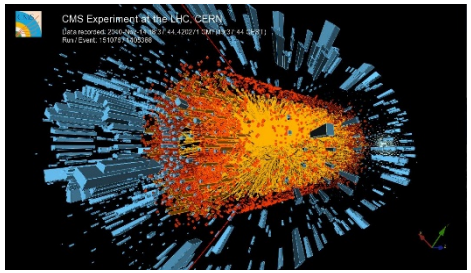
The INFN-CNAF datacenter

- 20.000 CPU core



- 20PB disk storage
- 35PB tape storage

- 1.2MW Electrical Power (10GWh/year)
  - 2Meur/year (in Italy)



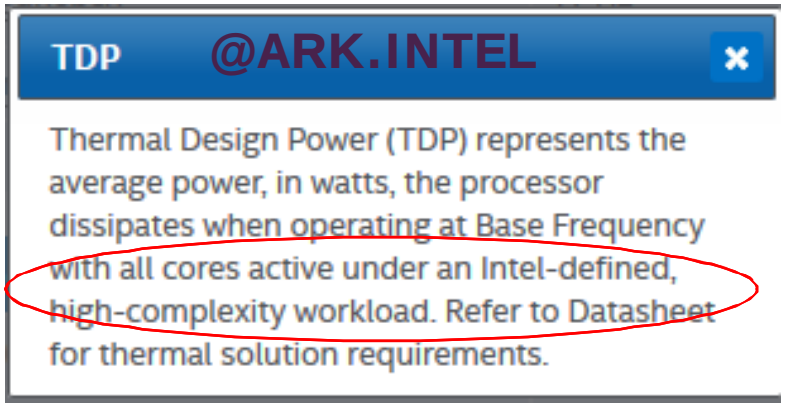
- Almost all embarrassingly parallel jobs for high energy physicists

Glossary: GFLOPS and TDP

- GFLOPS: Billions of Floating Point Operations per second
  - Max GFLOPS of a system can be calculated using:

$$\text{GFLOPS} = \text{sockets} \times \frac{\text{cores}}{\text{socket}} \times \text{clock} \times \frac{\text{FLOPs}}{\text{cycle}} \qquad \text{(Clock in GHz)}$$

- TDP: **Thermal Design Power** is the maximum amount of heat generated by the CPU that the cooling system in a computer is required to dissipate in **typical** operation (*)



(*) From wikipedia

Outline

- Introduction
- A little bit of history
- Modern Systems on Chip
- Computing on System on Chip, the INFN experience
- European Projects on low power computing
- Conclusion

Once upon a time....

The vector machines

- Serial number 001 Cray-1™
  - Los Alamos National Laboratory in 1976
  - $8.8 million
  - 80 MFLOPS scalar, 160/250 MFLOPS vector
  - 1 Mword (64 bit) main memory
  - 8 vector registers
    - 64 elements 64bit each
  - Freon refrigerated
  - 5.5 tons including the Freon refrigeration
  - 115 kW of power
    - 330 kW with refrigeration

(*) Source: Wikipedia

- Serial number 003 was installed at the National Center for Atmospheric Research (NCAR) in 1977 and decommissioned in 1989

Wireless technology inside

CRAY-XMP...Vector MultiProcessor

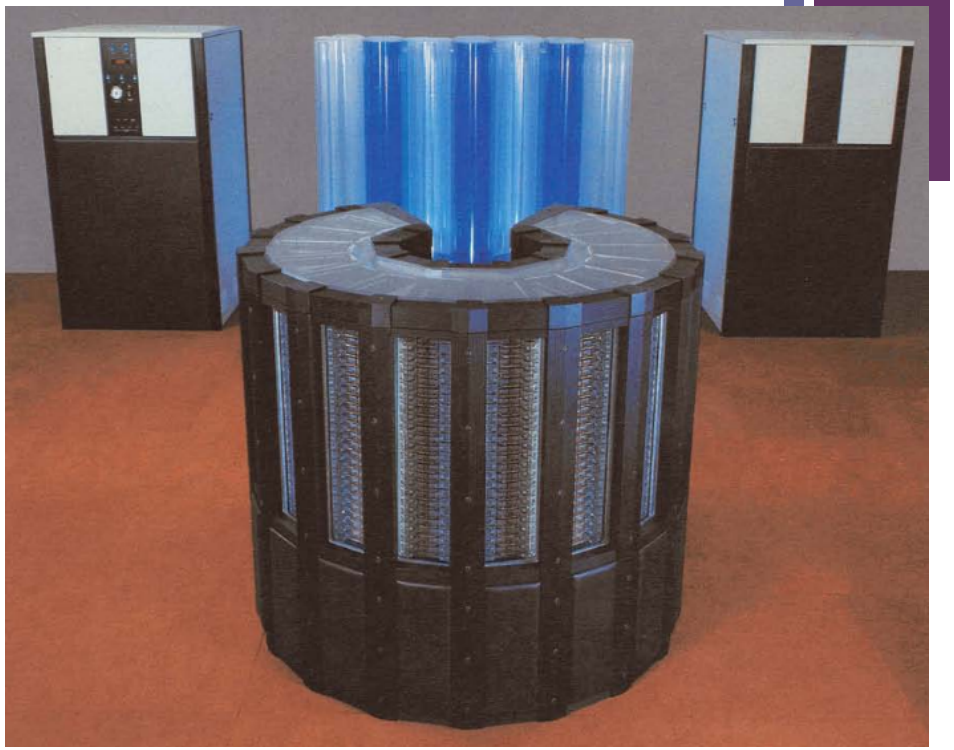
- 1982 CRAY-XMP 2 processors
  - 9.5 ns clock cycle (105 MHz)
  - 2x200MFLOPS
  - 2Mwords (64 bit) = 16MB

- 1984 CRAY-XMP four processors
  - 800 MFLOPS
  - 8Mwords
    - 64 MB main memory
  - about US$15 million
    - plus the cost of disks!!!

CRAY-XMP48 @ CERN in 1984
(*) source:
http://cerncourier.com/cws/article/cern/29150

The Cray-2

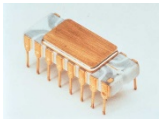
- The **Cray-2** released in 1985
  - 4 processors
  - 250MHz (4.1 ns)
  - 256 Mword (64bit) Main Memory
    - 2 GByte
  - 1.9 GFLOPS
  - 150 - 200 kW
  - Fluorinet cooling
  - 16 sq ft floor space
  - 5500 pounds
  - About $17 million



An inert fluorocarbon liquid circulates in the mainframe cabinet in direct contact with the integrated circuit packages. This liquid immersion cooling technology allows for the small size of the CRAY-2 mainframe and is thus largely responsible for the high computation rates.

(*) http://archive.computerhistory.org/resources/text/Cray/Cray.Cray2.1985.102646185.pdf

Microprocessors



The Intel 4004
1971

The Intel 8080
1974

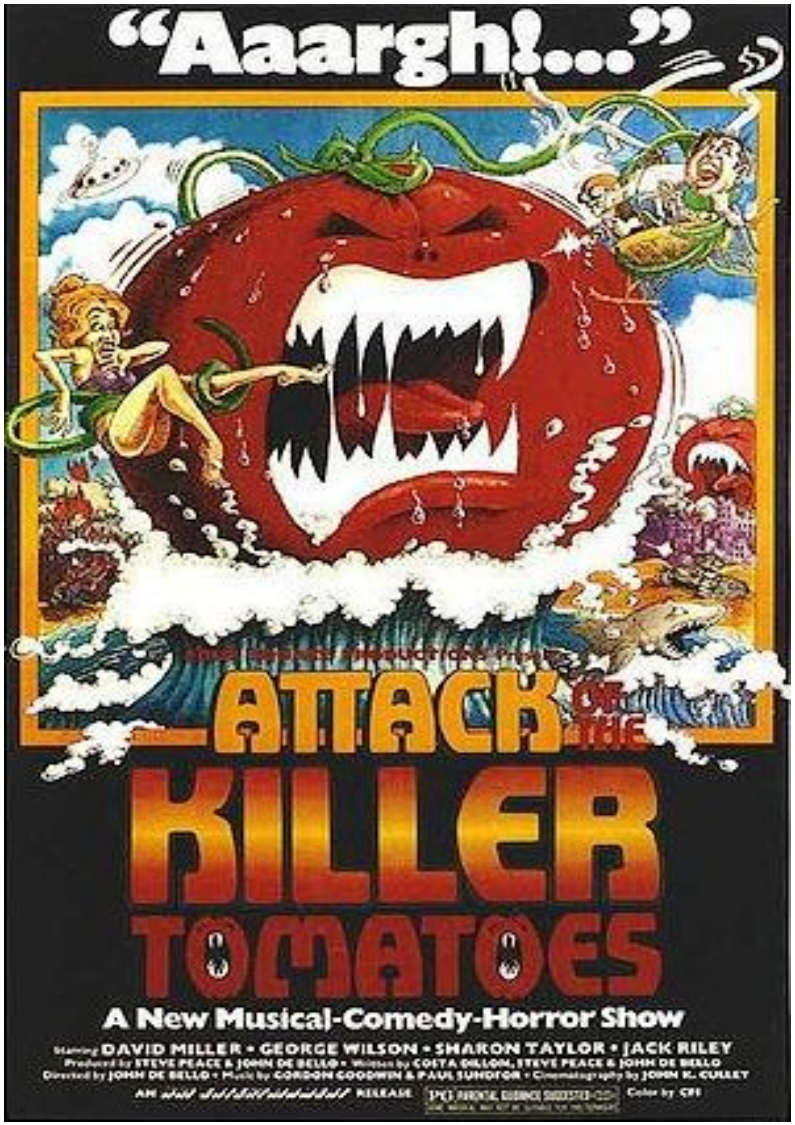
The Intel 8086
1978

The Intel 80286
1982

The Intel 80486
1989

In the 80s started the replacement of vector supercomputers with Massively Parallel Processors (MPP) and Clusters assembled from a larger number of lower performing microprocessors

The Killer Tomatoes



Attack of the Killer Tomatoes is a 1978 comedy horror film directed, produced, edited, scored and co-written by John DeBello

Had three sequels!!!

Massively Parallel Processor (MPP)

- A single computer with many networked processors
  - Specialized interconnect networks
  - Low latency interconnection
- Up to thousands of processors
- Some examples
  - Connection Machines (CM-1/2/200/5)
  - Intel Paragon
  - ASCI series
  - IBM SP
  - IBM BlueGene

ASCI Red MPP

- 1996 At Sandia Laboratories
- Based on the Paragon architecture
- Fastest supercomputer from 1997 to 2000
  - 1.4 TFLOPS (peak) in 1997
    - 9152 cores
  - 3,2 TFLOPS (peak) in 1999
    - 9632 cores
- 1st supercomputer above 1 TFLOPS

TERA SCALE

(*) Source: Wikipedia

IBM BlueGene/Q MPP

PETA SCALE

- **Trading the speed of processors for lower power consumption**
- System-on-a-chip design. All node components were embedded on one chip
- A large number of nodes
- 5D xTorus interconnect
- Compute chip is an 18 core chip
  - The 64-bit PowerPC A2
  - 4-way simultaneously multithreaded per core
  - 1.6 GHz
  - a 17th core for operating system functions
  - chip manufactured on IBM's copper SOI process at 45 nm.
  - 204.8 GFLOPS and 55 watts per processor
- Up to 20 PFLOPS (peak)
  - 16384 cores

(*) Source: Wikipedia

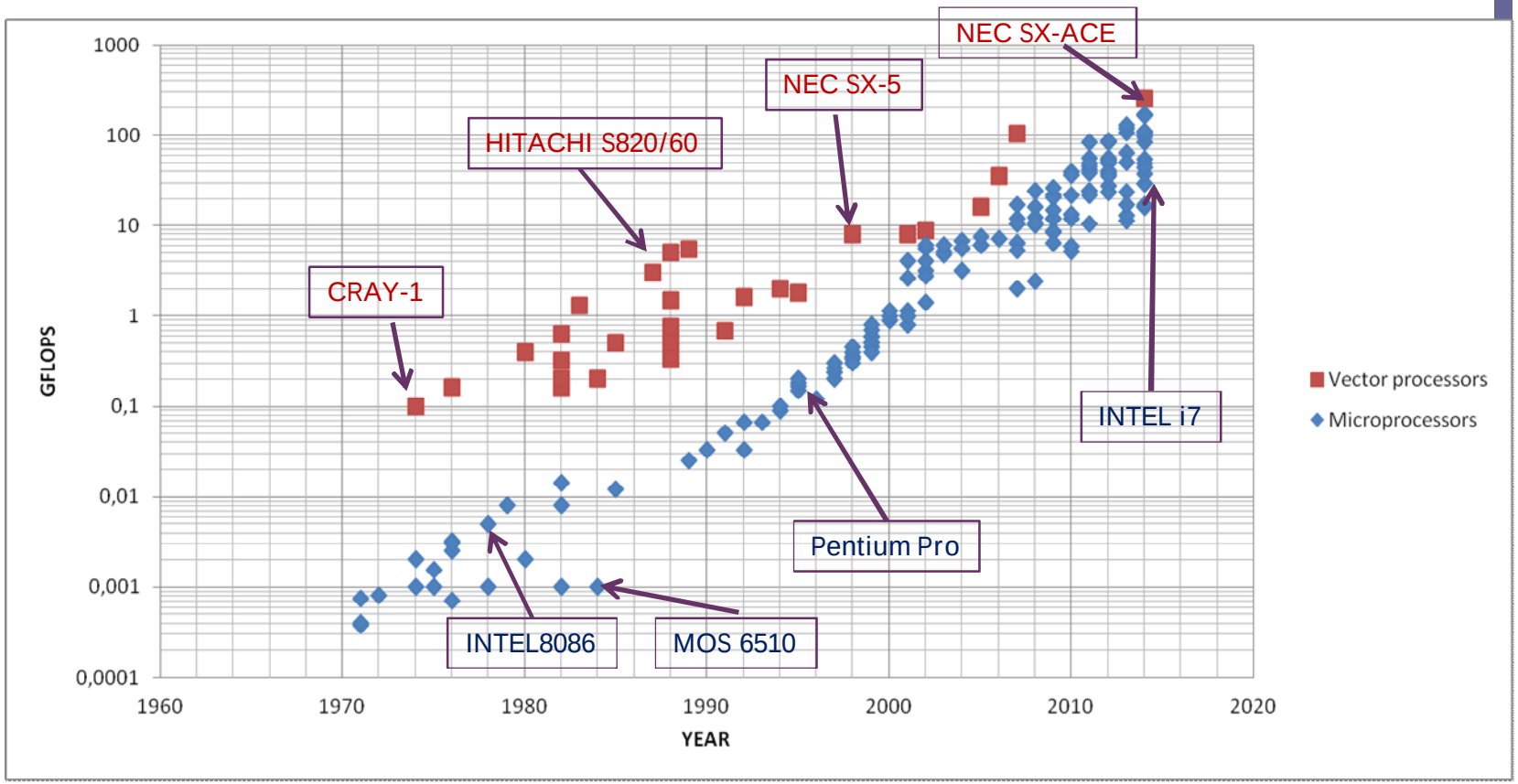
Clusters

[a cluster is a] parallel computer system comprising an integrated collection of independent nodes, each of which is a system in its own right, capable of independent operation and derived from products developed and marketed for other stand-alone purposes

Dongarra et al. : "High-performance computing: clusters, constellations, MPPs, and future directions", Computing in Science & Engineering (Volume:7 , Issue: 2)

(*) Picture from: http://en.wikipedia.org/wiki/Computer_cluster

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

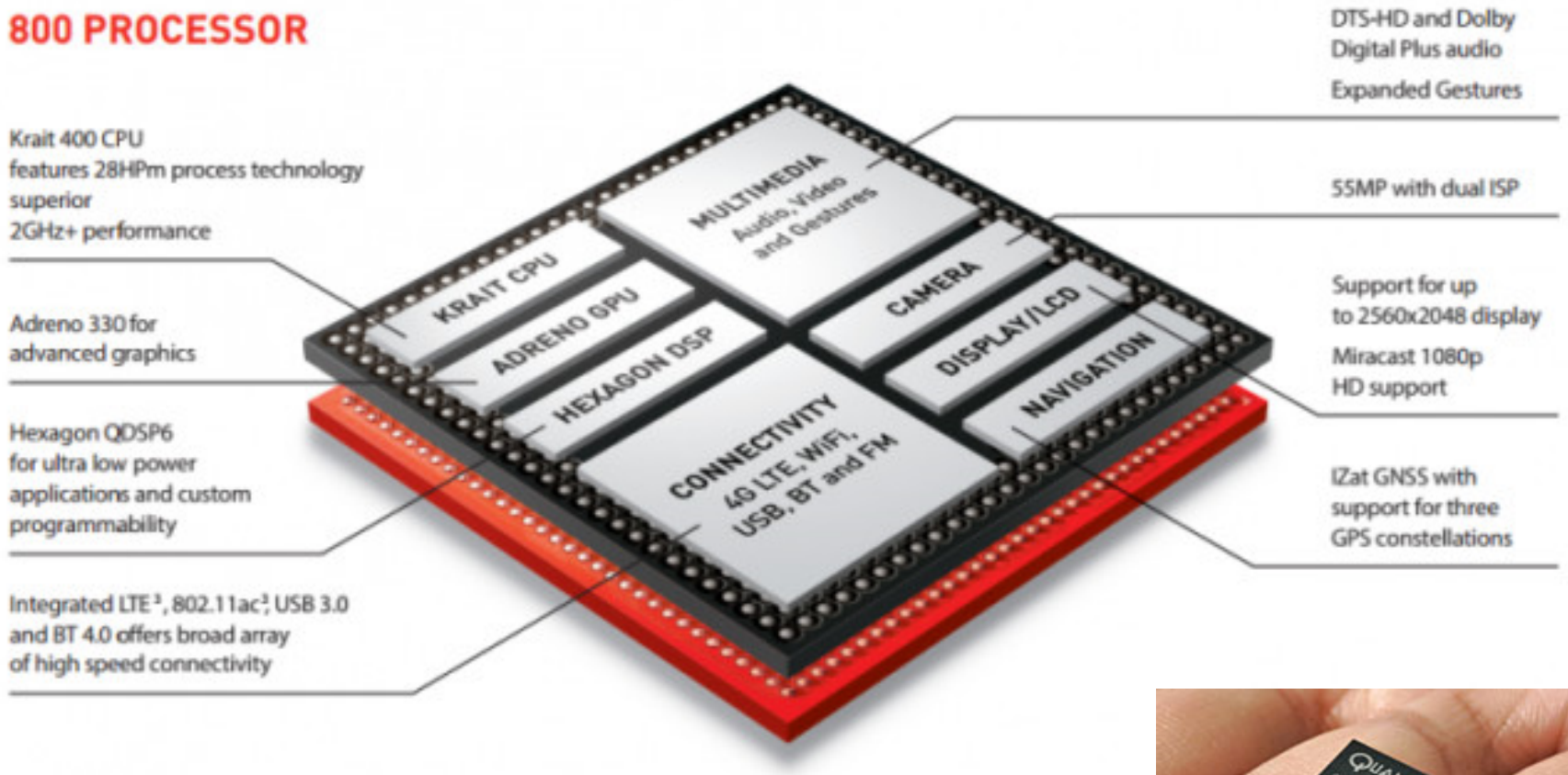
Cool people would say "greener"

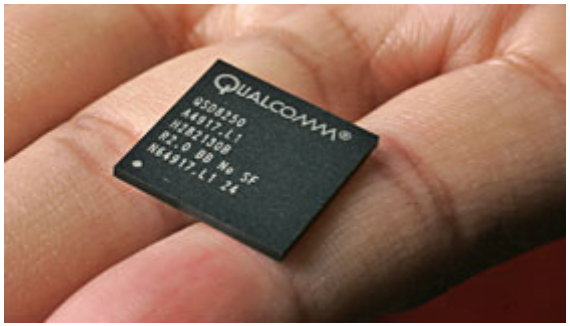
Commodity hardware

- Microprocessors started to be mass produced and used in everyday life
  - Personal computer at home
  - Office automation
  - Gaming

What's commodity nowadays?

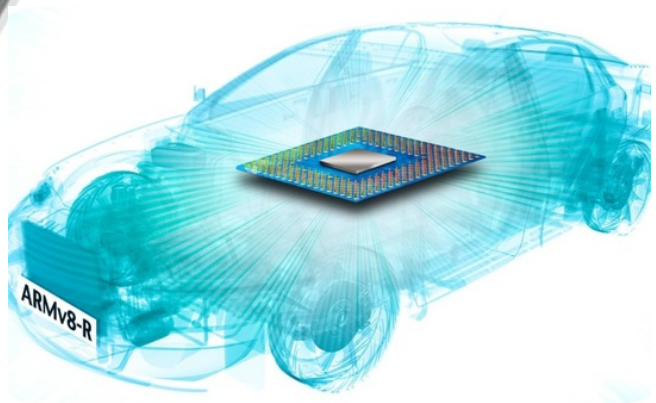
Low-Power System on Chip (SoCs)



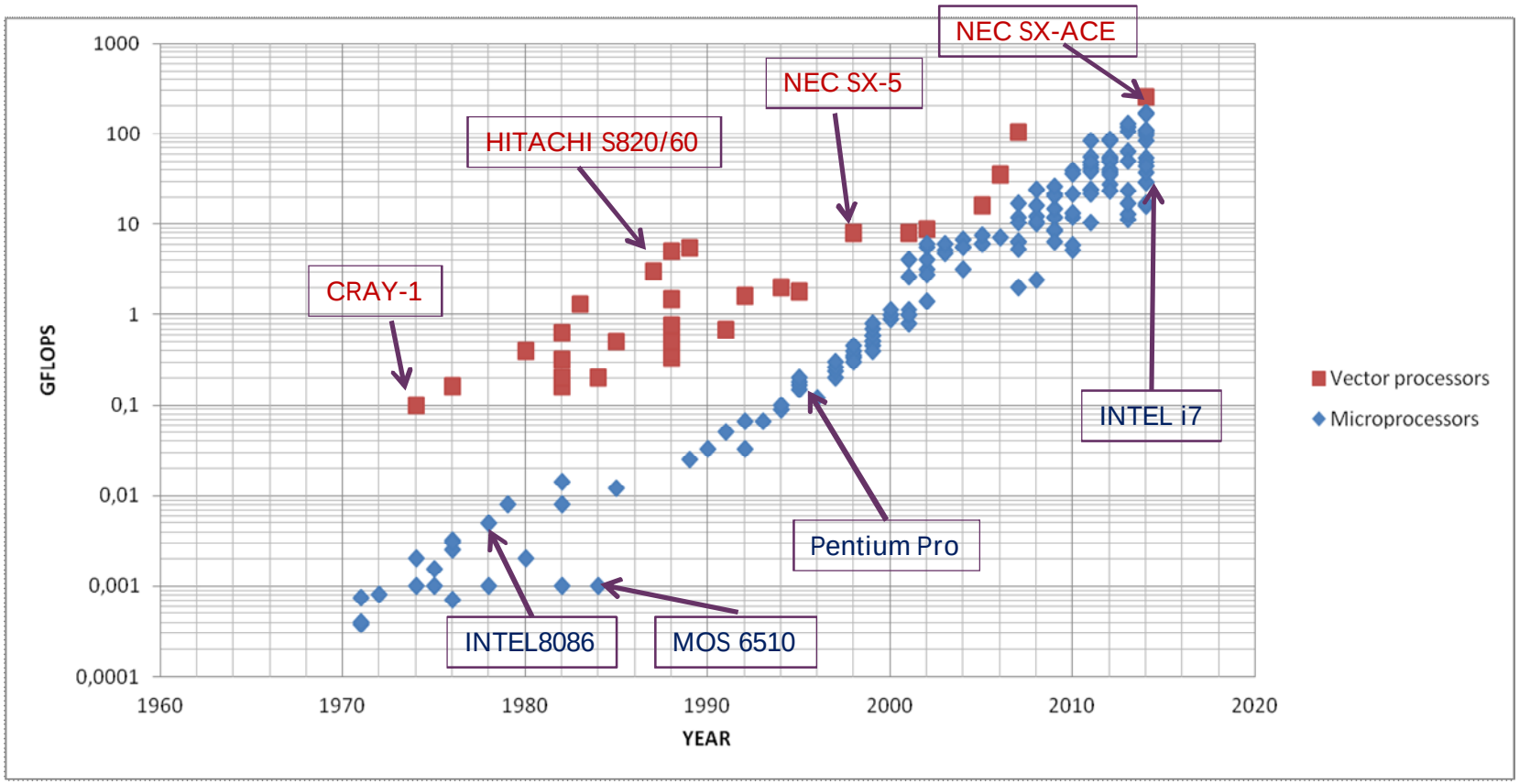


Where do I find a SoC?

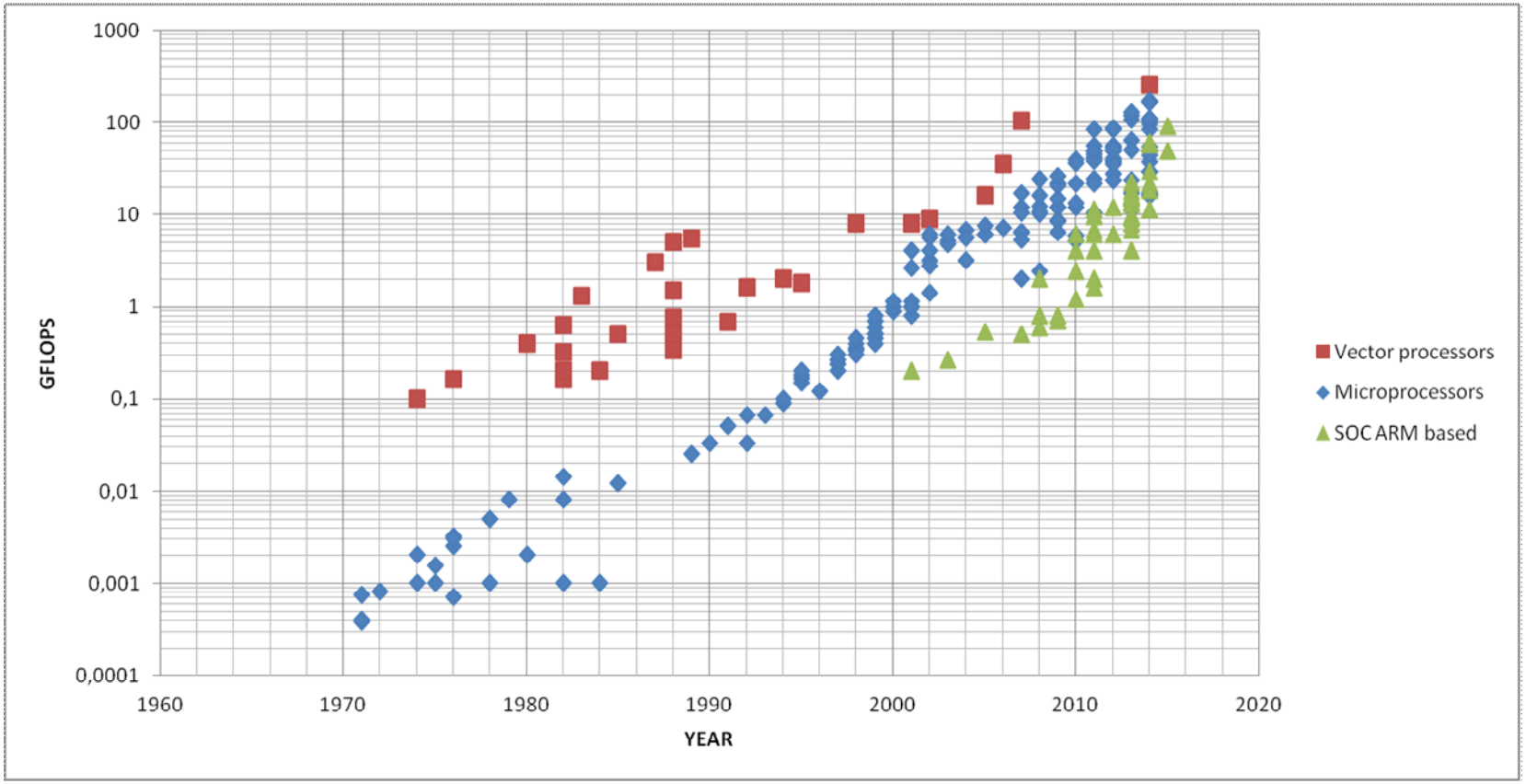
- Mobile
- Embedded



Vector vs Micro computing power



Vector vs Micro vs ARM based



- Is history repeating?

Shaping the future is difficult

- People in the foreseen 2015 look like barbarians to us
  - No internet
  - No google
  - No facebook
  - No smartphones

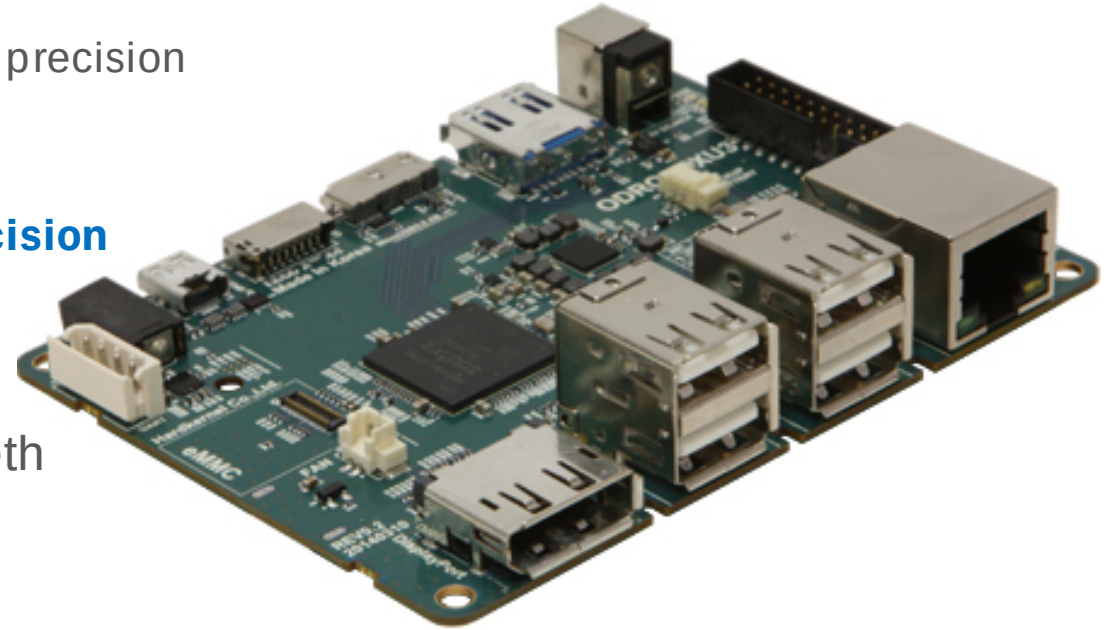
…but they had flying cars!!

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

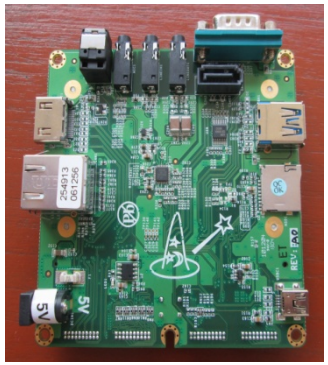
- Powered by ARM® big.LITTLE™ technology, with a **Heterogeneous Multi-Processing (HMP)** solution
  - 4 core ARM A15 + 4 cores ARM A7
- Exynos 5422 by Samsung
  - ~ 20 GFLOPS peak (32bit) single precision
- **Mali-T628 MP6 GPU**
  - **~ 110 GFLOPS peak single precision**
- 2 GB RAM
- 2xUSB3.0, 2xUSB2.0,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



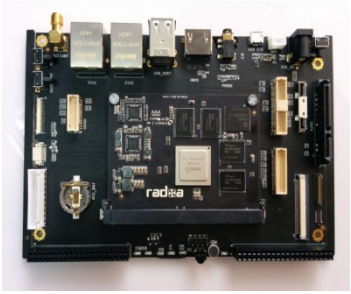
Power consumption max ~ 15 W

Costs 150 euro!

Other nice boards



WandBoard

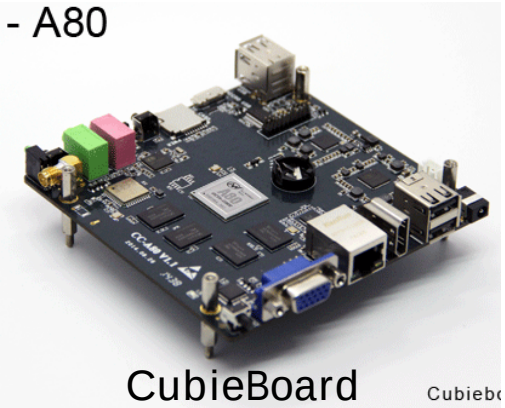


Rock2Board

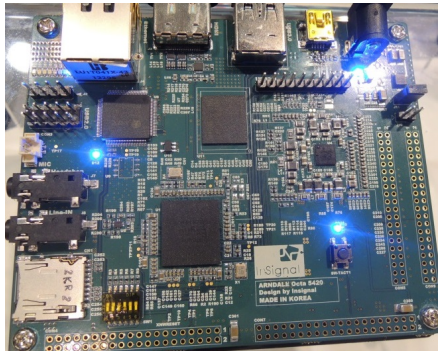
PandaBoard

DragonBoard

SabreBoard



CubieBoard



ArndaleBoard

Texas Instruments EVMK2H

http://elinux.org/Development_Platforms

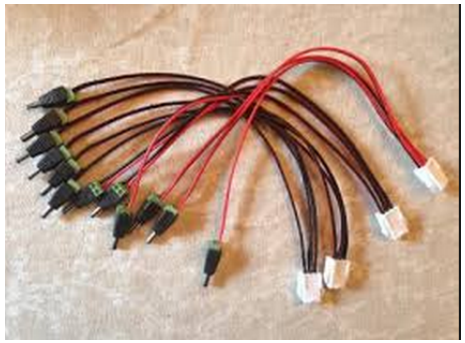
- ...and counting...

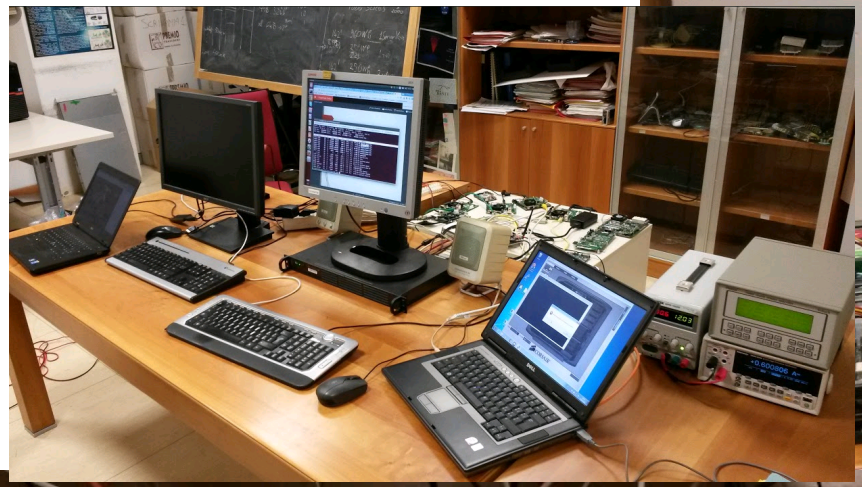
COSA Project Clusters

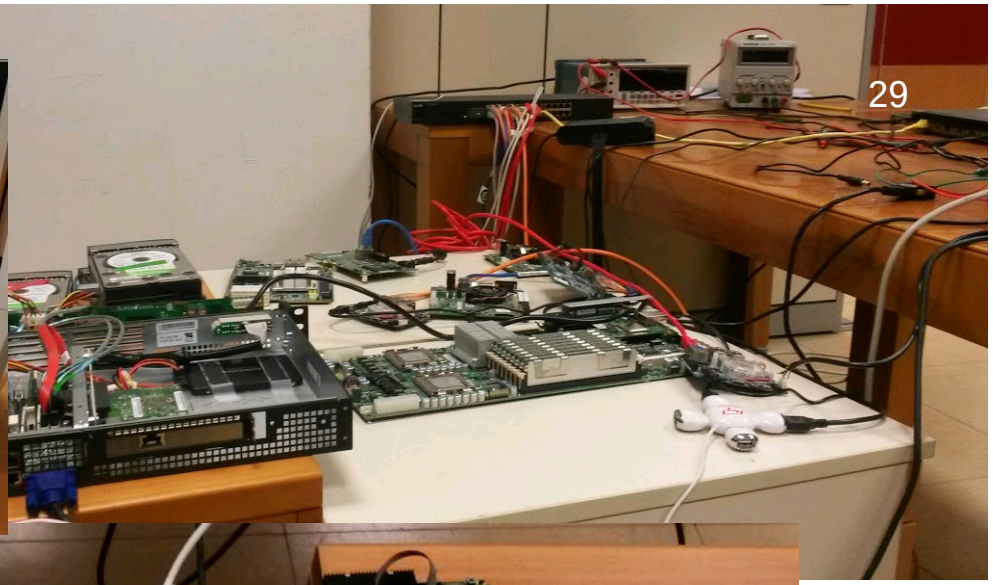
funded by INFN

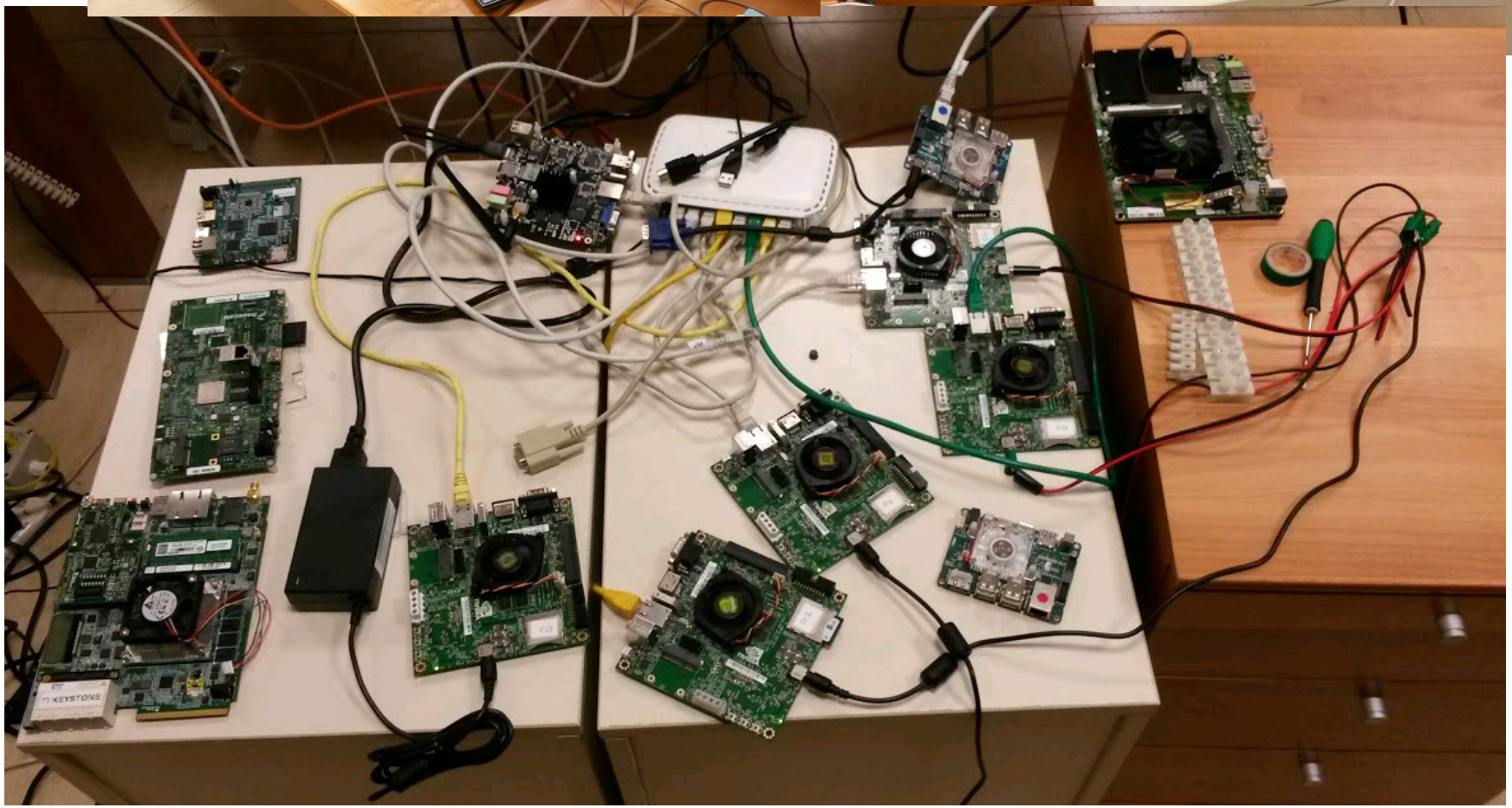
ARMv7 Cluster

X86_64 Cluster








KEYSTONE

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)	**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)

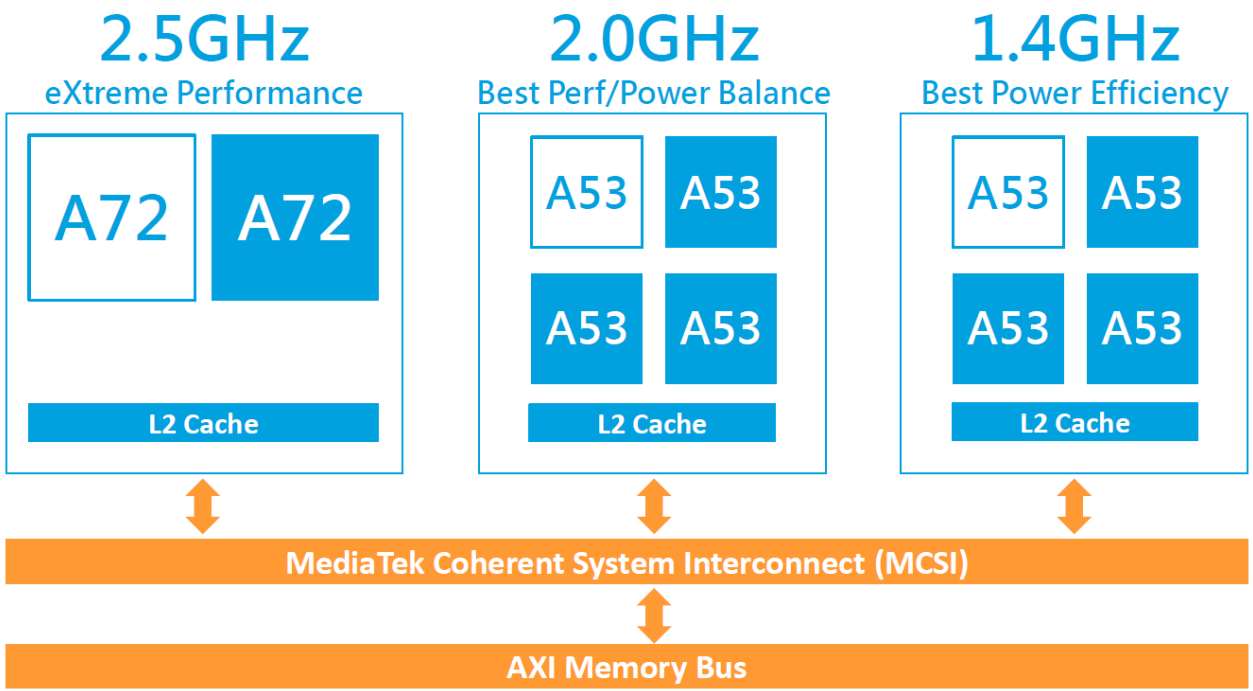
TDP beetween 5W and 15W

(EVMK2H > 15W)

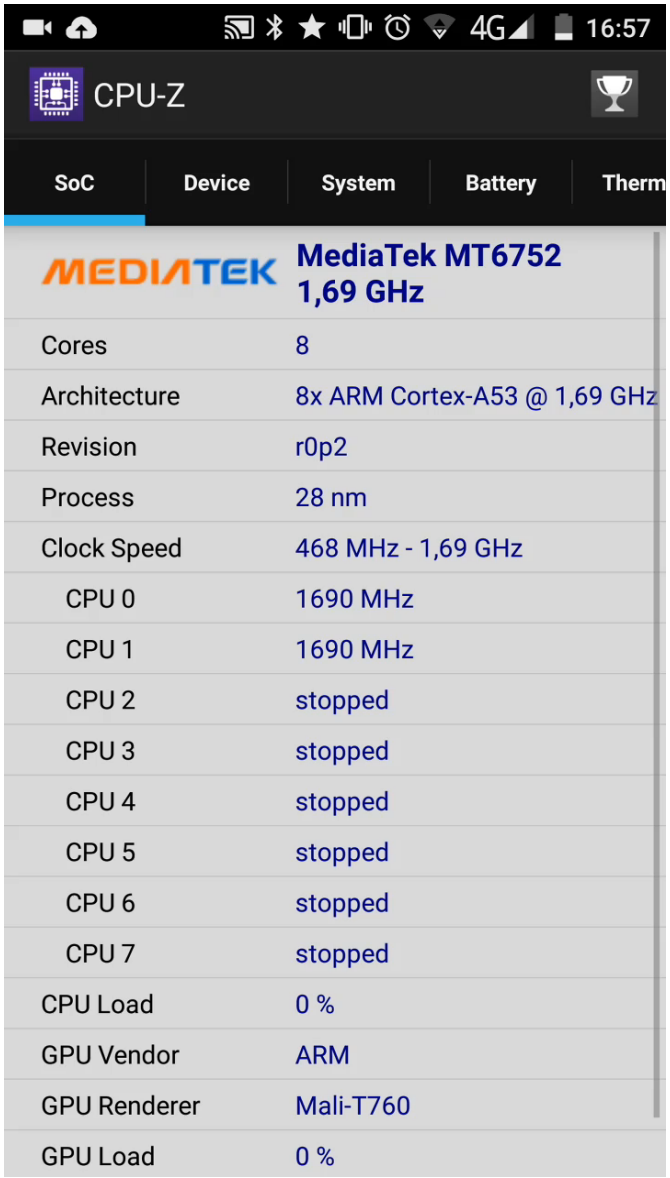
SoC Multicore Madness

big.Medium.LITTLE

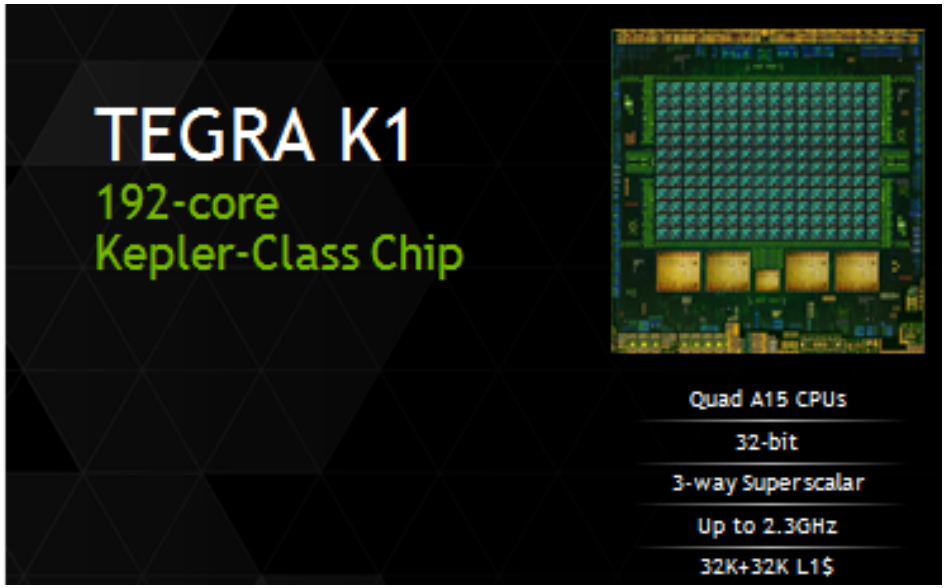
Deca/10-Core CPU Architecture

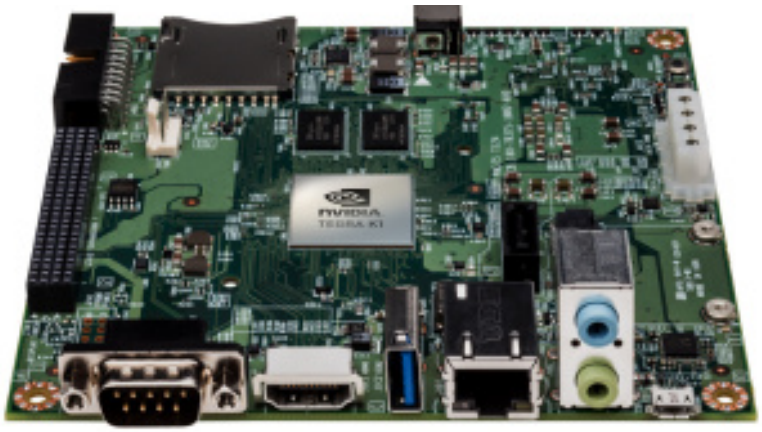


Modern SoCs


16:57
CPU-Z
SoC
Device
System
Battery
Therm
MEDIATEK
MediaTek MT6752
1,69 GHz
Cores 8
Architecture 8x ARM Cortex-A53 @ 1,69 GHz
Revision r0p2
Process 28 nm
Clock Speed 468 MHz - 1,69 GHz
CPU 0 1690 MHz
CPU 1 1690 MHz
CPU 2 stopped
CPU 3 stopped
CPU 4 stopped
CPU 5 stopped
CPU 6 stopped
CPU 7 stopped
CPU Load 0 %
GPU Vendor ARM
GPU Renderer Mali-T760
GPU Load 0 %

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 is available for ARM CPUs (for free)
- OpenCL for the GPU
  - If you are lucky enough to find working drivers
- CUDA on the JETSON TK1
- Cross compilation
  - If you dare!

Limitations of these boards

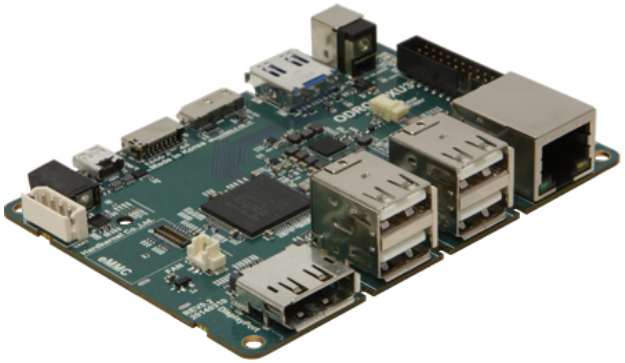
- Moreover commodity SoCs and development boards have a number of limitations:
  - 32 bit
    - 64bit boards are coming (maybe)
  - Small caches
  - Small RAM size
  - No ECC memory
  - Frequent failures and system crashes
  - Slow connections in some cases
  - HW bugs

If anything can go wrong, it will. (Murphy)

Outline

- Introduction
- A little bit of history
- Modern Systems on Chip
- Computing on System on Chip, the INFN experience
- European Projects on low power computing
- Conclusion

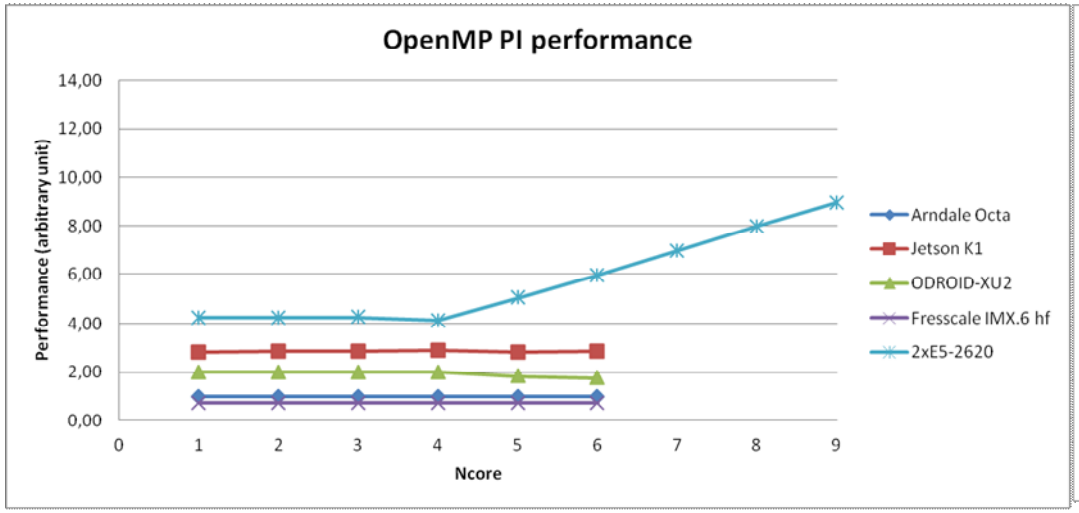
That's not fair…

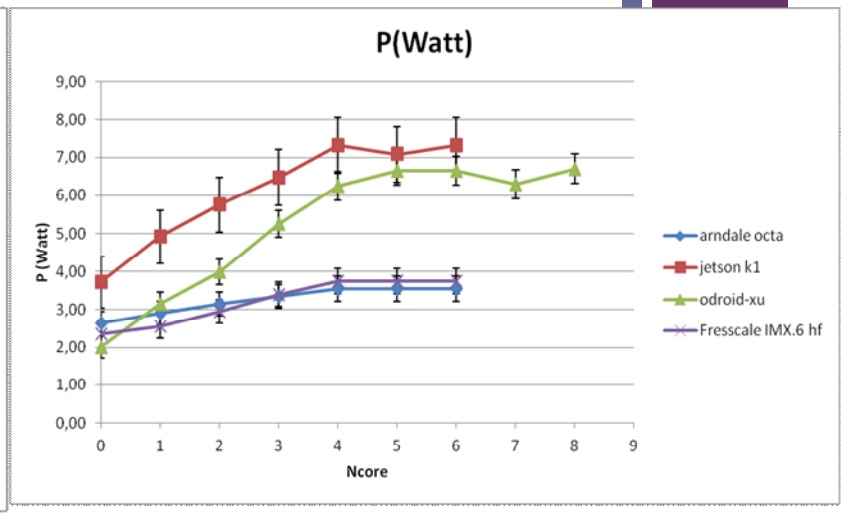


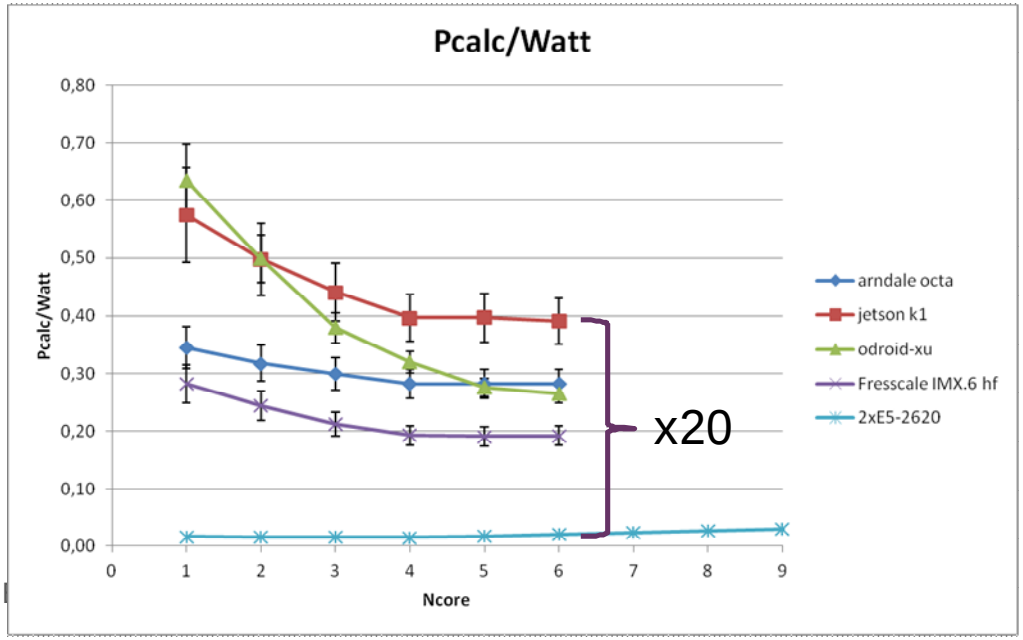
≠

OpenMP π computation

CPU ONLY

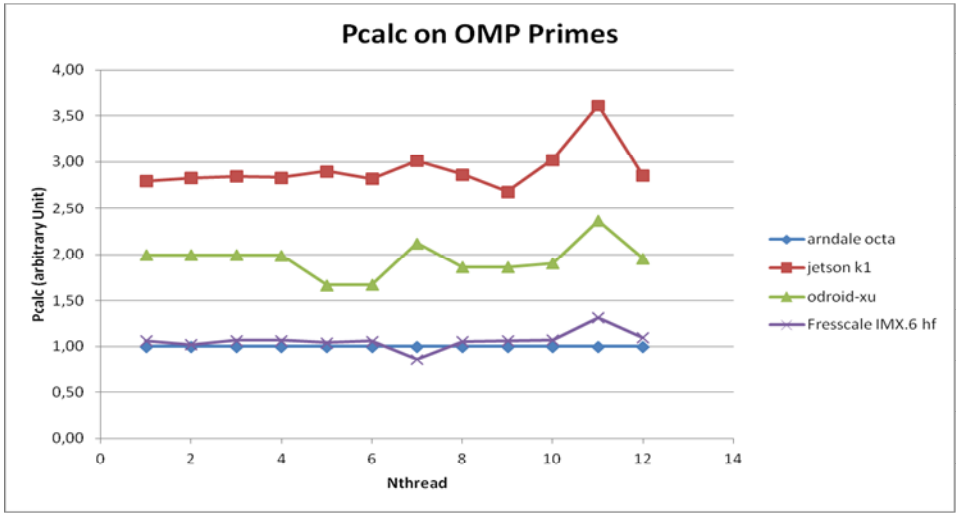


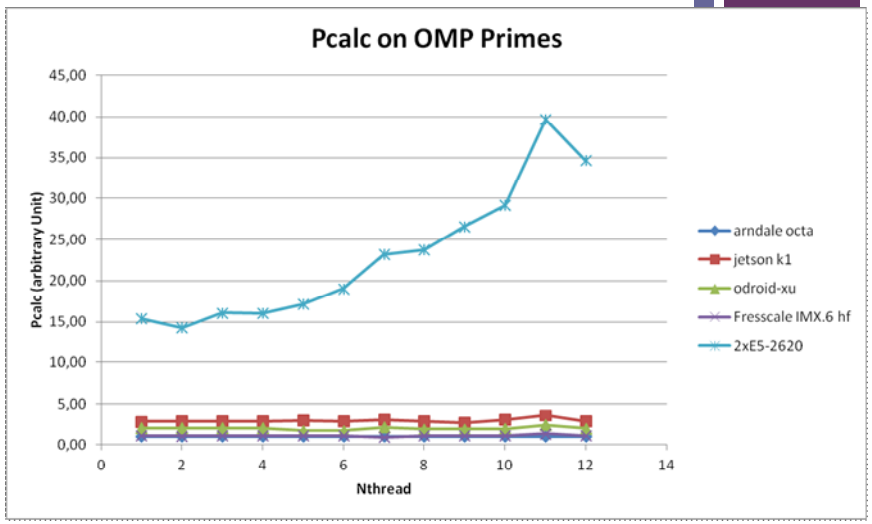


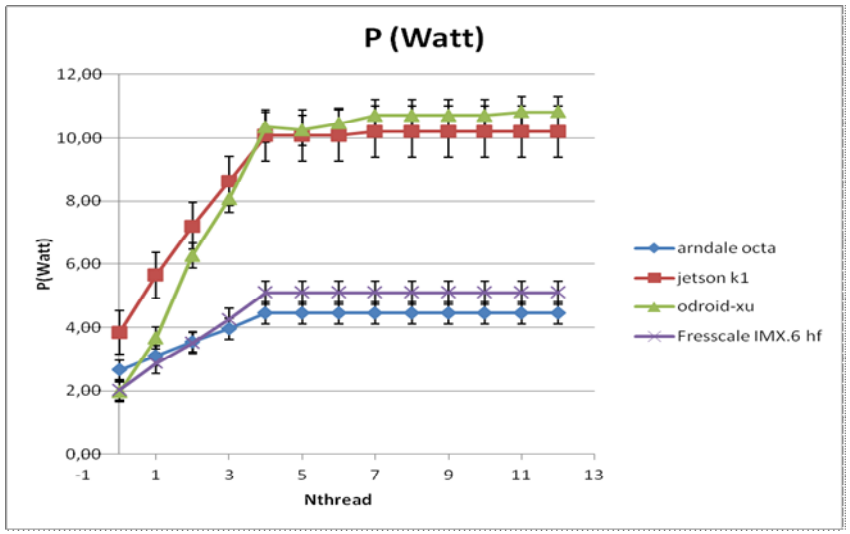


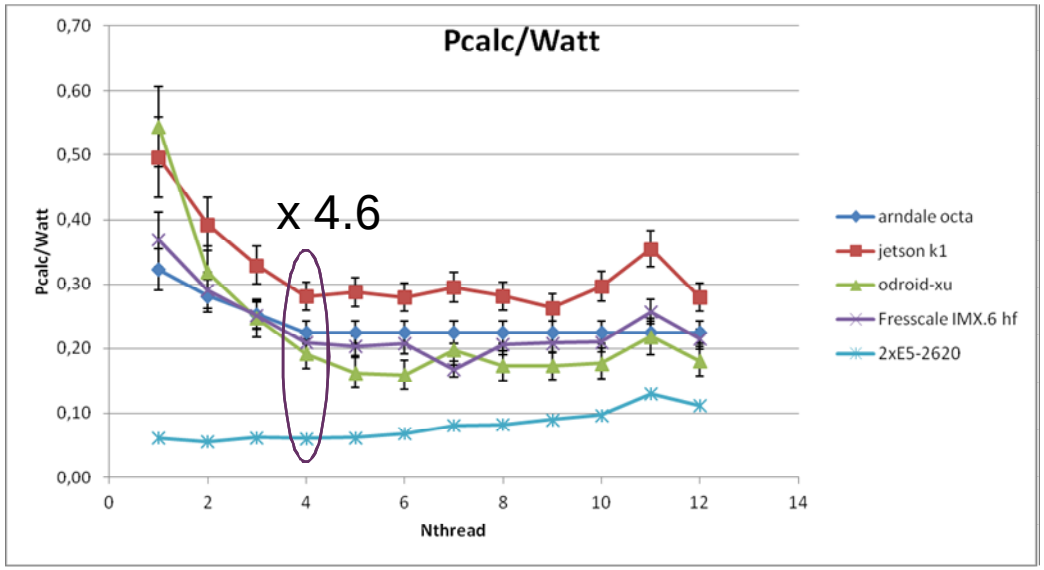
Prime numbers computation

CPU ONLY

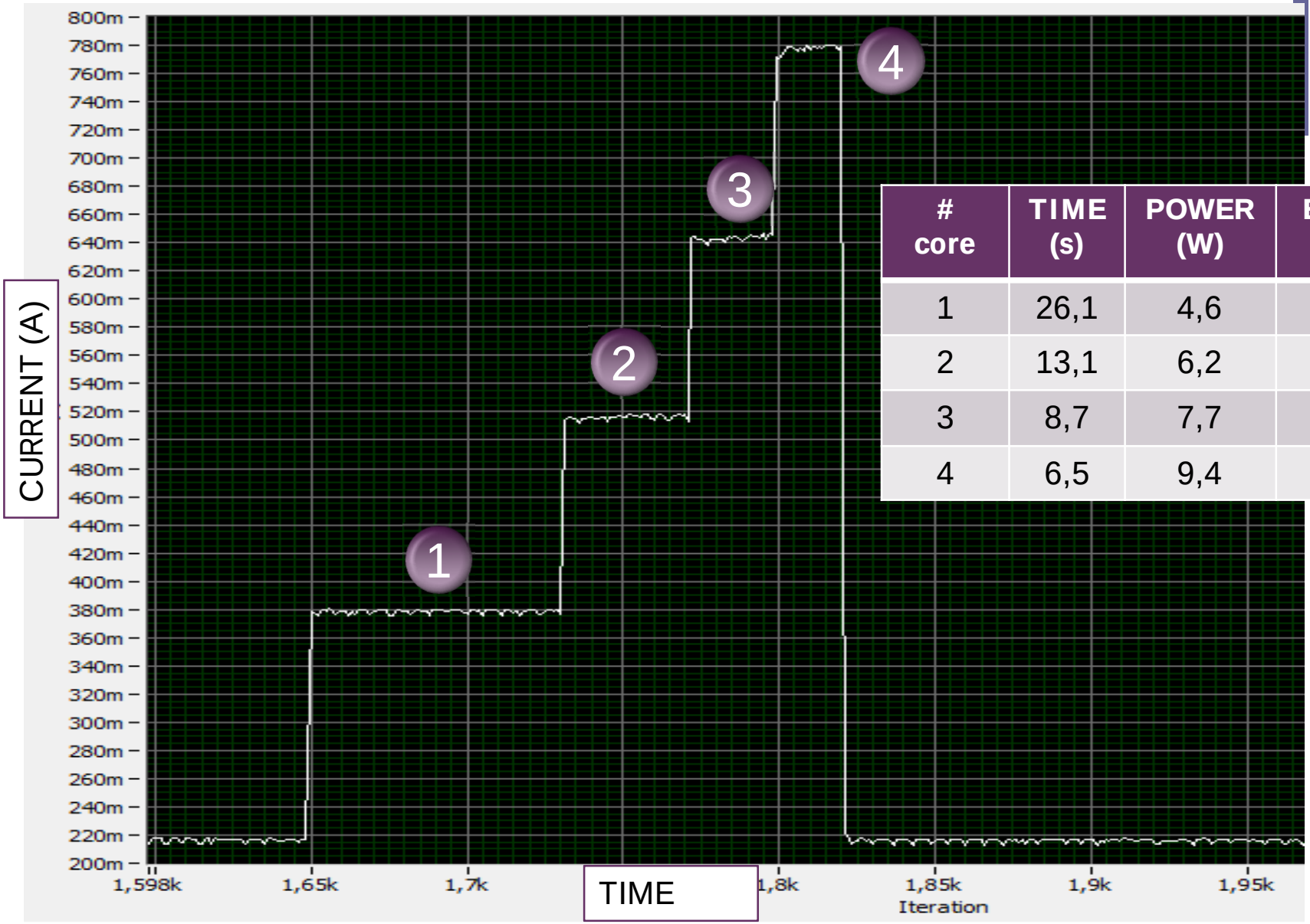








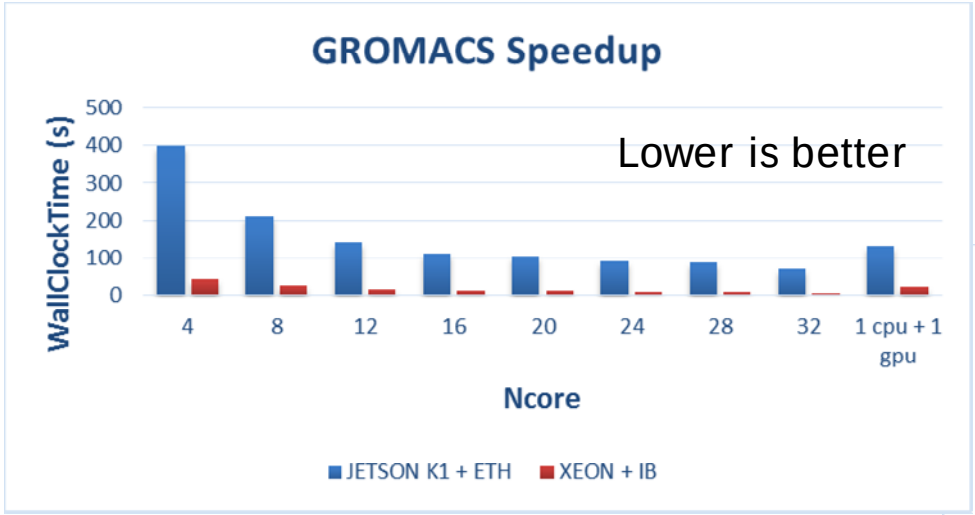
OPENMP π computation on Jetson



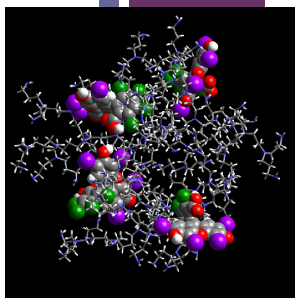
# core	TIME (s)	POWER (W)	ENERGY (J)
1	26,1	4,6	120,1
2	13,1	6,2	81,2
3	8,7	7,7	66,8
4	6,5	9,4	60,8

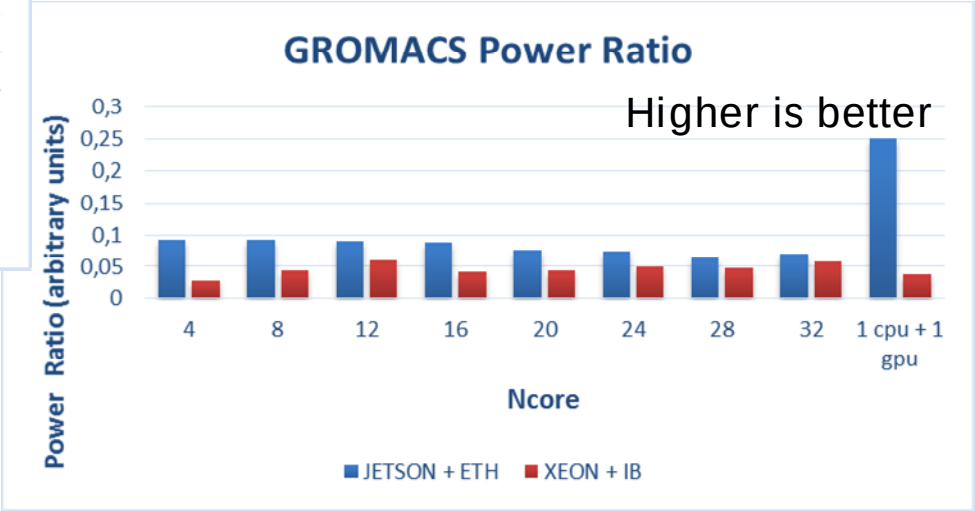
Molecular Dynamics on Jetson-K1

Parallel application for CPU and GPU



OpenMPI/CUDA runs
52809 Atoms
2ps simulation





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

CMS@CERN results

CPU ONLY

High Energy Physics MonteCarlo simulations

Table 1. Results of run time tests for single core CMSSW (GEM-SIM) and a multi-threaded version of the Geant4 benchmark "FullCMS" with 4 threads (G4MT).

Type	Cores	Power (TDP)	GEN-SIM Events /minute /core	GEN-SIM Events /minute /Watt	G4MT Events /minute (threads)	G4MT Events /minute /Watt
ODROID U2	4	4W	1.08	1.08	34.2 (4)	8.6
ODROID XU+E	4/4	5W	1.47	1.07	47 (4)	9.4
dual Xeon L5520 @2.27GHz	2×4	120W	3.37	0.22	307.2 (16)	2.6
dual Xeon E5-2630L @2.0GHz	2×6	120W	3.46	0.35	N/A	N/A

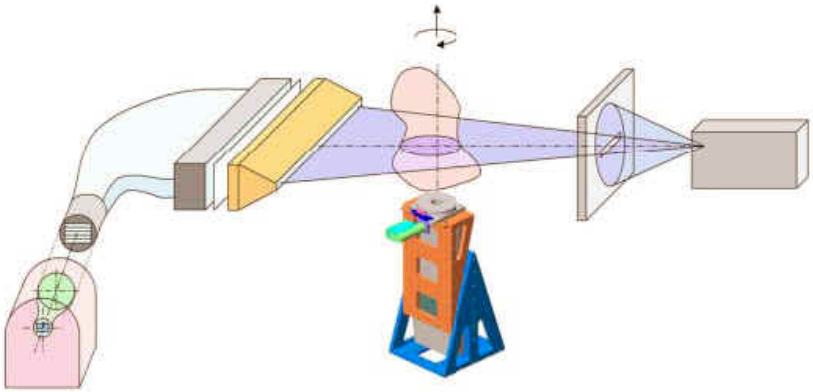
David Abdurachmanov *et al* 2014 *J. Phys.: Conf. Ser.* **513** 052008 doi:10.1088/1742-6596/513/5/052008

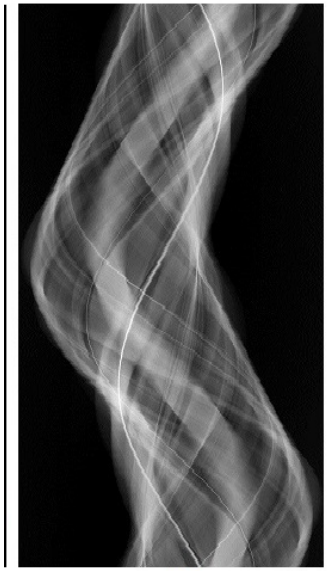
ARM slower by a factor 3 or 4 but…

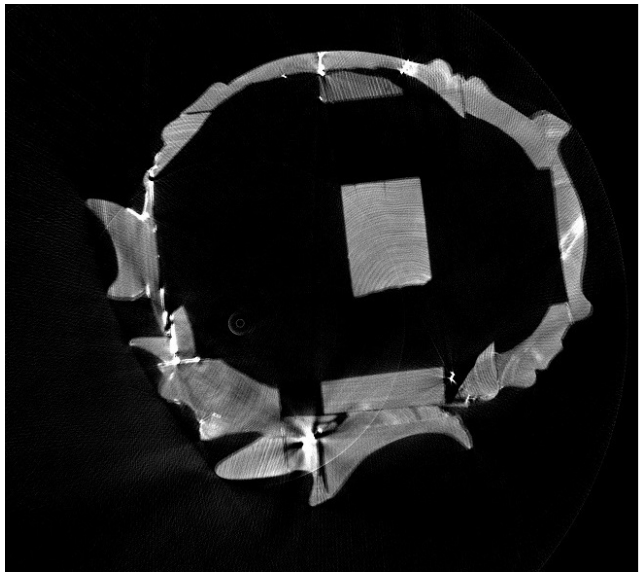
…ARM better by a factor 3 or 5 on the power ratio

Computer Tomography - 1

Filtered Backprojection Algorithm





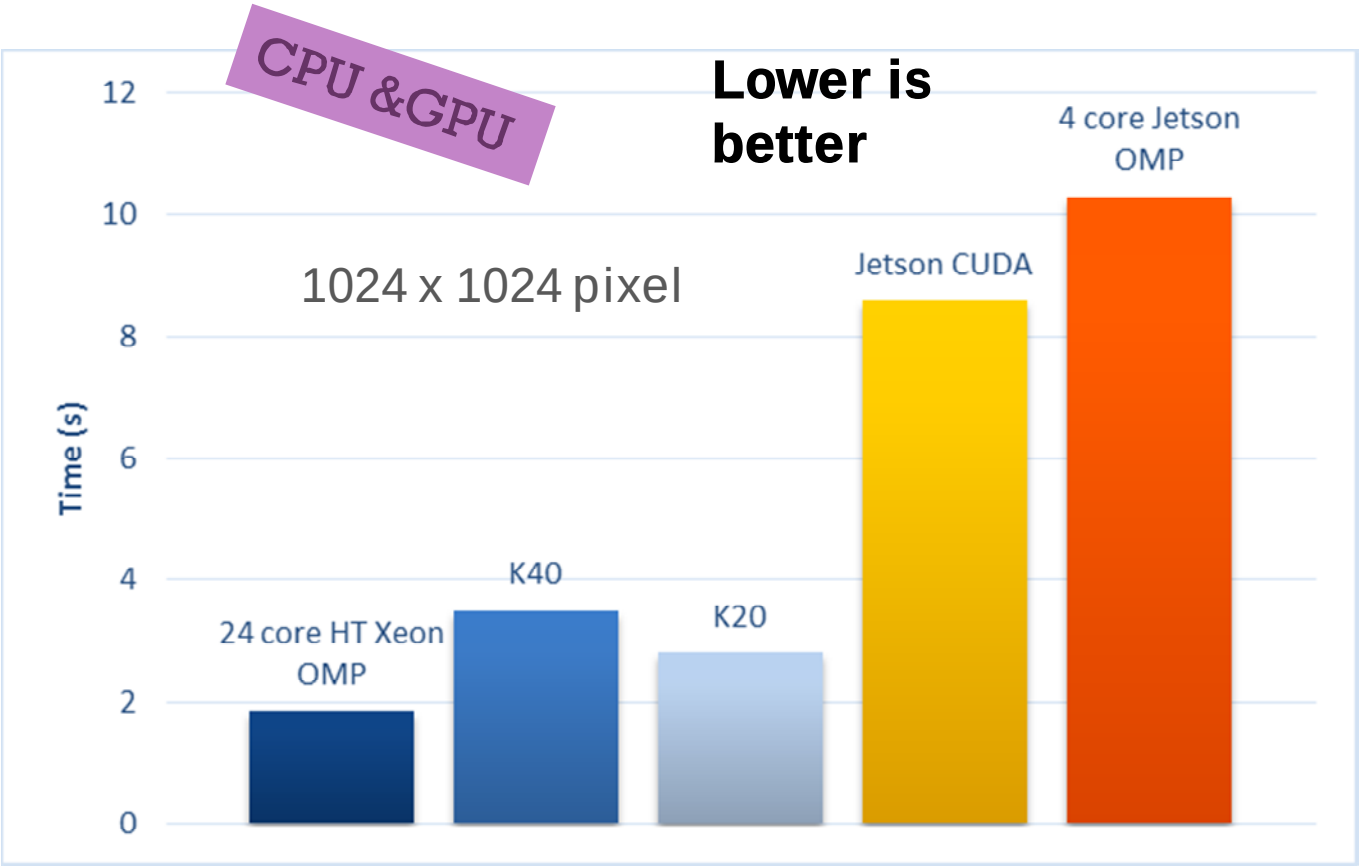


Real-Time Reconstruction for 3-D CT Applied to Large Objects of Cultural Heritage, R. Brancaccio, M. Bettuzzi, F. Casali, M. P. Morigi, G. Levi, A. Gallo, G. Marchetti, and D. Schneberk, IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 58, NO. 4, AUGUST 2011

Computer Tomography - 2

Filtered Backprojection Algorithm

(*) 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

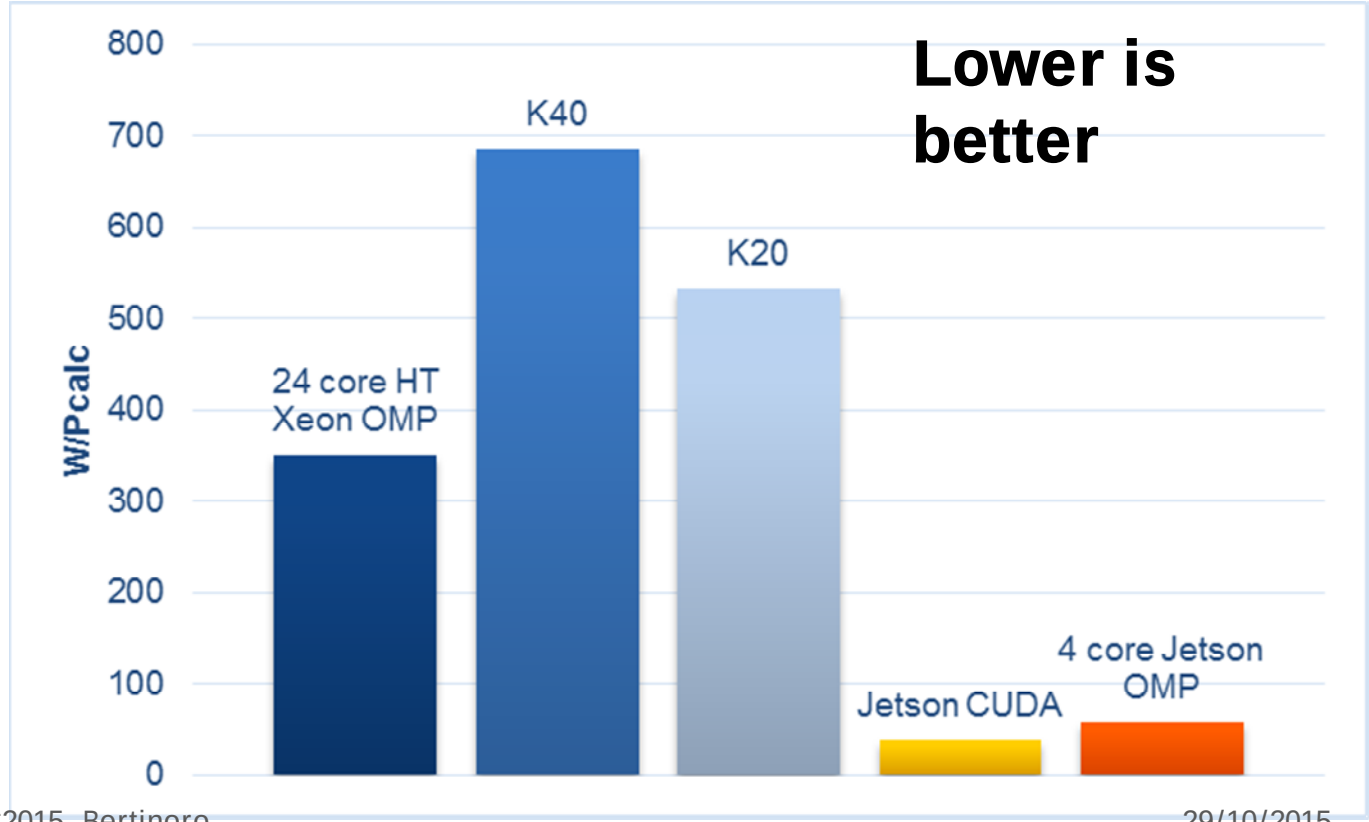
Computer Tomography - 3

Filtered Backprojection Algorithm

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

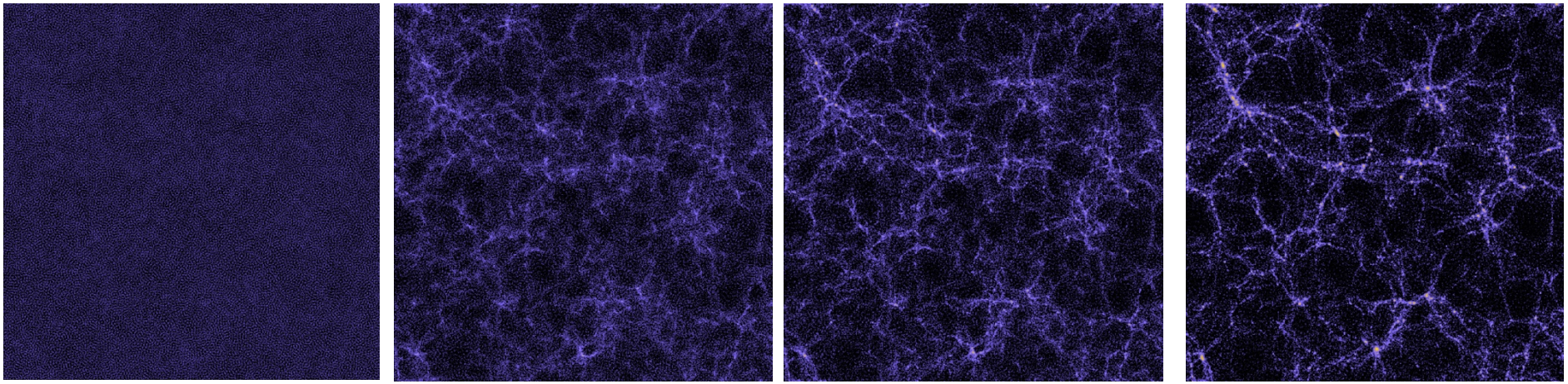
On **5** x Jetson-K1: 3840 slices in 1 h: **41** Wh

1024 x 1024
pixel per slice



Evolving the Universe with Gadget2

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**
  - Dual Xeon **~220 W**

Time (s)

N cores	Jetson K1	Xeon	MIC
4	11649,30	4445,58	
8	10563,76	2383,3	
16		1344,78	10608,96

Neural Networks: DPSNN-STDP

CORTICONIC

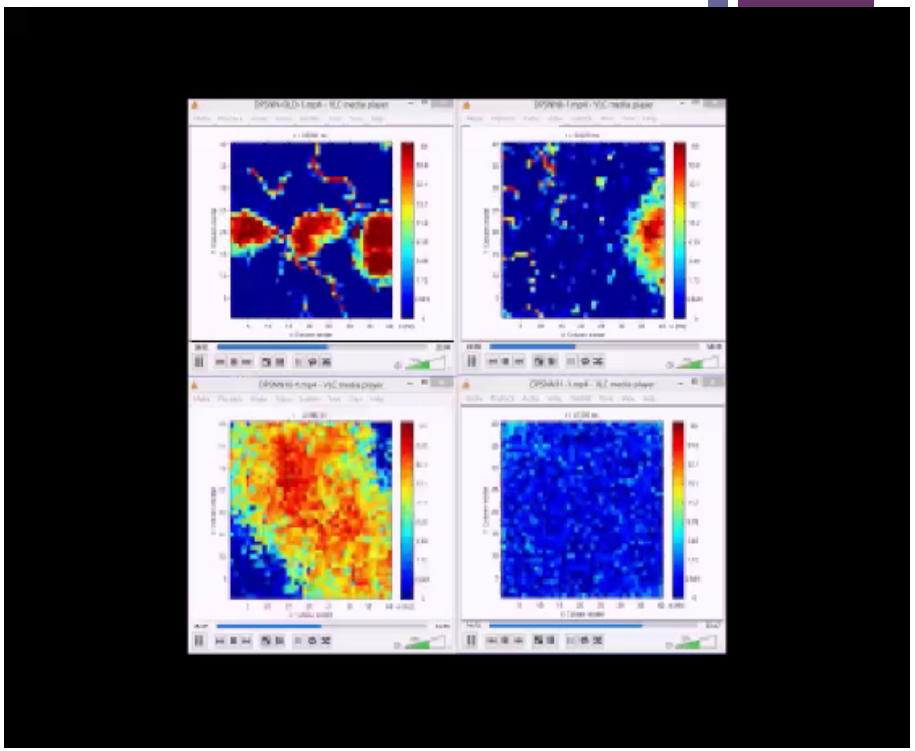
A Challenging Problem

- **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

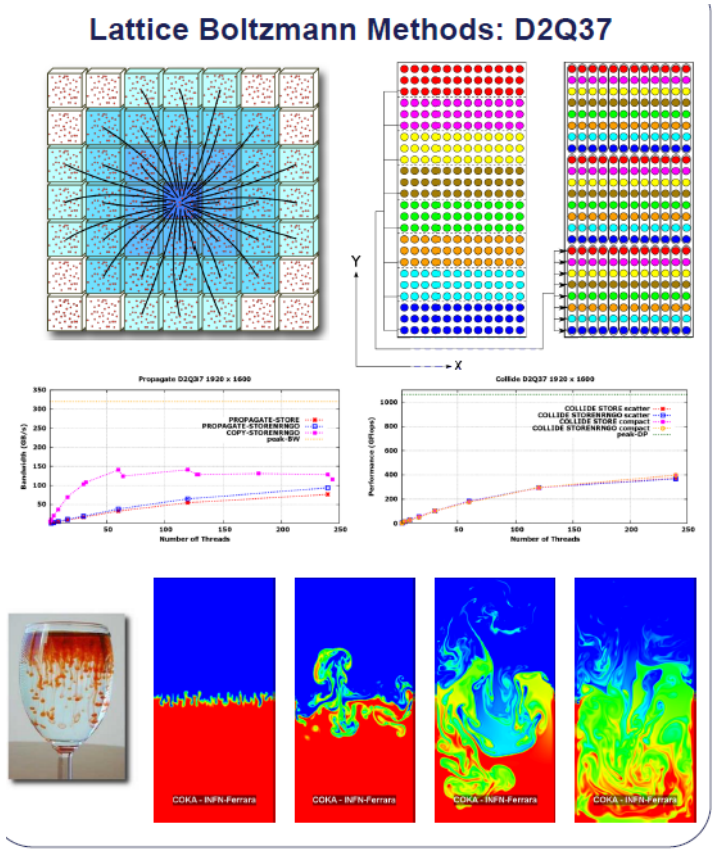
COSA Meeting - Roma - 2015 05 18



(*) **Power, Energy and Speed of Embedded and Server Multi-Cores applied to Distributed Simulation of Spiking Neural Networks: ARM in NVIDIA Tegra vs Intel Xeon quad-cores**
Authors: Pier Stanislao Paolucci, Piero Vicini et al

Lattice Boltzmann on the Tegra K1

GPU only



(*) Schifano et al. ; A portable OpenCL Lattice Boltzmann code for multi- And many-core processor architectures;
Procedia Computer Science Volume 29, 2014, Pages 40-49,
doi: 10.1016/j.procs.2014.05.004

LBM Performance Comparison (*)

	Xeon-Phi 7120		Tesla K20Xm			i7-4930K	
Code Version	OCL	C(*)	OCL	CUDA SM_20	CUDA SM_35	OCL	C(*)
propagate T/iter [msec]	30.46	37.67	14.89	15.40	15.38	186.42	162.00
GB/s	76.42	61.8	156.33	151.16	151.36	12.48	14.54
$\mathcal{E}_p$	22%	17%	62%	60%	60%	21%	24%
bc T/iter [msec]	3.20	4.61	7.08	5.68	5.70	4.30	4.87
collide T/iter [msec]	72.79	79.14	93.27	83.33	43.96	440.18	307.42
GFLOPS (DP)	410	377	320	358	680	68	97
MLUPS	54.02	49.69	42.16	47.19	89.44	8.93	12.94
$\mathcal{E}_c$	34%	31%	24%	27%	52%	42%	59%
μJ / site	5.55	6.03	5.57	4.98	2.63	14.55	10.04
T_{WC}/iter [msec]	106.45	121.42	115.24	104.42	65.03	630.90	489.98
MLUPS	36.94	32.38	34.12	37.65	60.46	6.23	8.12

On Tegra-K1

15 GFLOPS
12 GB/s
P_e~ 10 Watt

40x slower than a K20m

- Porting easier than expected
- Performance under investigation

Computing on ARM SoCs

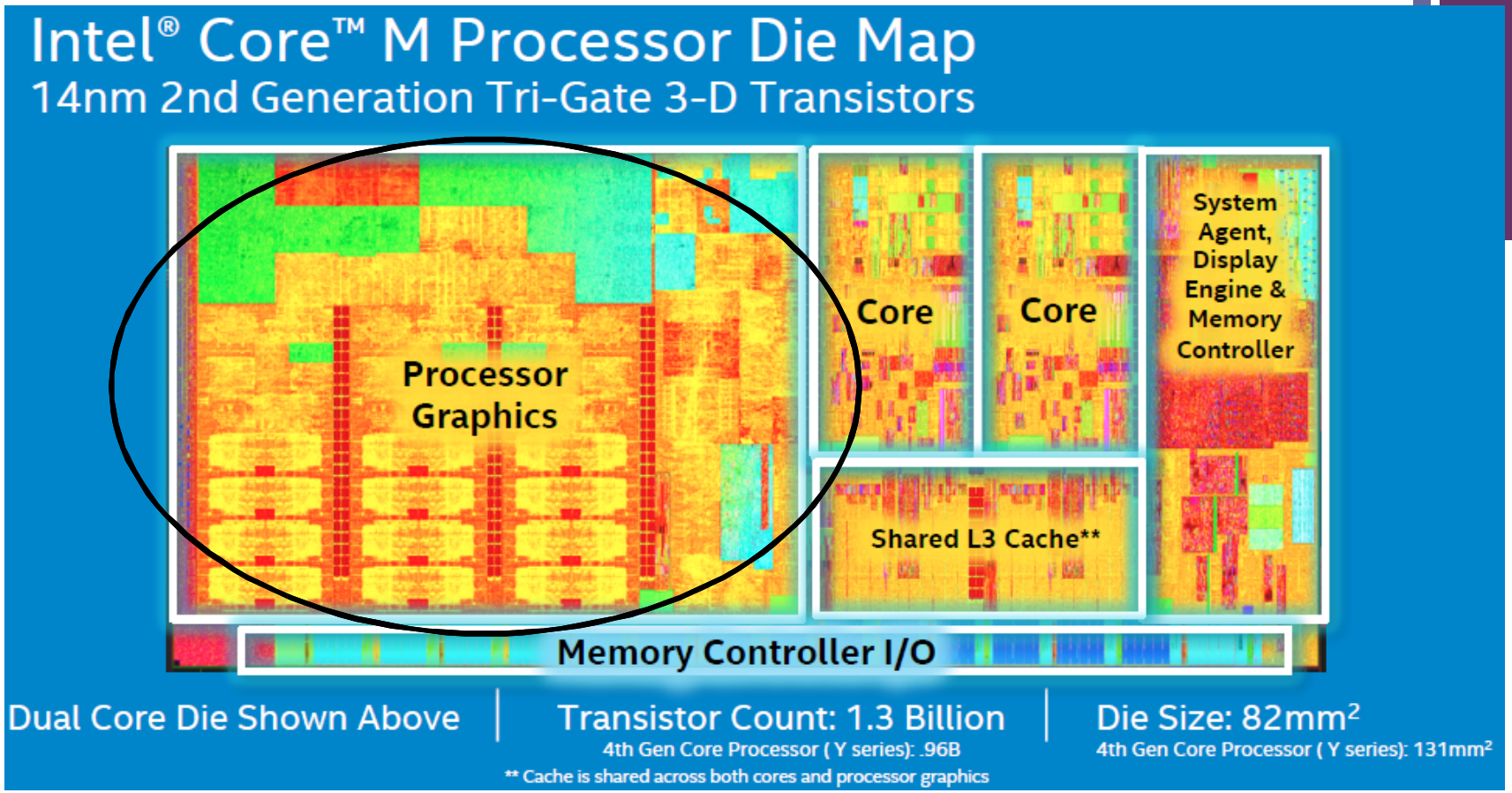
- ARM based SoCs are becoming attractive for scientific computing
- Most of the computing gains are due to the GPU
  - It can be very difficult to exploit it
- Porting x86 MPI/OPENMP/CUDA applications is quite easy
- They still have many limitations
  - Don't just look at specs and GFLOPS count!!!
- Performance strongly depends on the application as in every computing platform

Only ARM based SoCs? And Intel?

- INTEL produce SoCs
  - Probably you have one in your laptop or smartphone/tablet
- Already 64bit
- Integrated GPU
  - CILK++ programmable
  - OpenCL programmable

Some Low Power from Intel

Nome prodotto	Intel® Core™ m7-6Y75 Processor (4M Cache, up to 3.10 GHz)	Intel® Core™ M-5Y71 Processor (4M Cache, up to 2.90 GHz)	Intel® Core™ i7-6650U Processor (4M Cache, up to 3.40 GHz)	Intel® Pentium® Processor N3700 (2M Cache, up to 2.40 GHz)	Intel® Atom™ x7-Z8700 Processor (2M Cache, up to 2.40 GHz)	Intel® Xeon® Processor D-1540 (12M Cache, 2.00 GHz)	Intel® Atom™ Processor C2750 (4M Cache, 2.40 GHz)
Nome in codice	Skylake	Broadwell	Skylake	Braswell	Cherry Trail	Broadwell	Avoton
Informazioni di base							
Stato	Launched	Launched	Launched	Launched	Launched	Launched	Launched
Data di lancio	Q3'15	Q4'14	Q3'15	Q1'15	Q1'15	Q1'15	Q3'13
Numero del processore	M7-6Y75	5Y71	i7-6650U	N3700	x7-Z8700	D-1540	C2750
Cache	4 MB Intel® Smart Cache	4 MB	4 MB Intel® Smart Cache	2 MB L2 Cache	2 MB	12 MB	4 MB
Set di istruzioni	64-bit	64-bit	64-bit	64-bit	64-bit	64-bit	64-bit
Estensioni set di istruzioni	SSE4.1/4.2, AVX 2.0	SSE4.1/4.2, AVX 2.0	SSE4.1/4.2, AVX 2.0			AVX 2.0	
Opzioni integrate disponibili	No	No	No	Yes	No	No	No
Litografia	14 nm	14 nm	14 nm	14 nm	14 nm	14 nm	22 nm
Prezzo consigliato per il cliente	TRAY: $393.00	TRAY: $281.00	TRAY: $415.00	TRAY: $161.00	TRAY: $35.00	TRAY: $581.00	TRAY: $171.00
Numero di core	2	2	2	4	4	8	8
Numero di thread	4	4	4	4	4	16	8
Frequenza base del processore	1.2 GHz	1.2 GHz	2.2 GHz	1.6 GHz	1.6 GHz	2 GHz	2.4 GHz
Frequenza turbo massima	3.1 GHz	2.9 GHz	3.4 GHz			2.6 GHz	2.6 GHz
TDP	4,5 W	4,5 W	15 W	6 W		45 W	20 W
TDP-up configurabile	7 W	6 W					
TDP-down configurabile	3,5 W	3,5 W	9,5 W				
SDP (Scenario Design Power)		3,5 W		4 W	2 W		
Specifiche della grafica							
Grafica del processore ‡	Intel® HD Graphics 515	Intel® HD Graphics 5300	Intel® Iris™ Graphics 540	Intel® HD Graphics	Intel® HD Graphics	None	
Frequenza di base grafica	300 MHz	300 MHz	300 MHz	400 MHz	200 MHz		
Frequenza dinamica massima grafica	1 GHz	900 MHz	1.05 GHz				
Memoria massima video grafica	1.7 GB		1.7 GB				



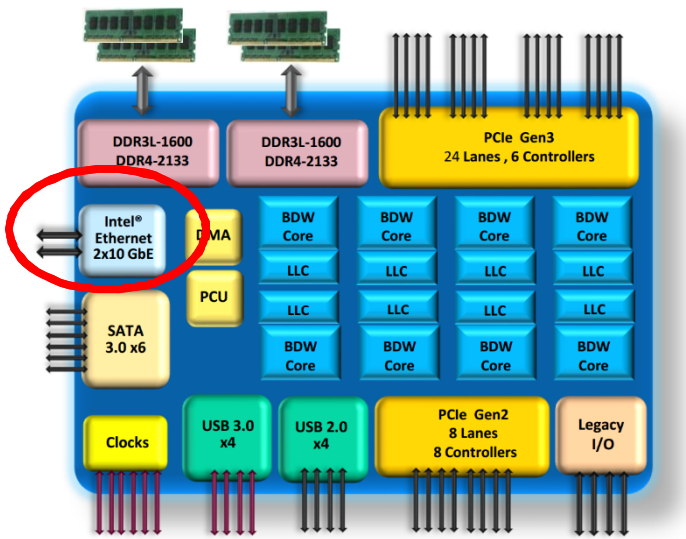
- 2 cores + GPU
  - Intel HD Graphics
  - OpenCL 2.0 Support
- 4.5 Watt (TDP)

(*) 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

XEON D-1540

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



Product Brief	Link

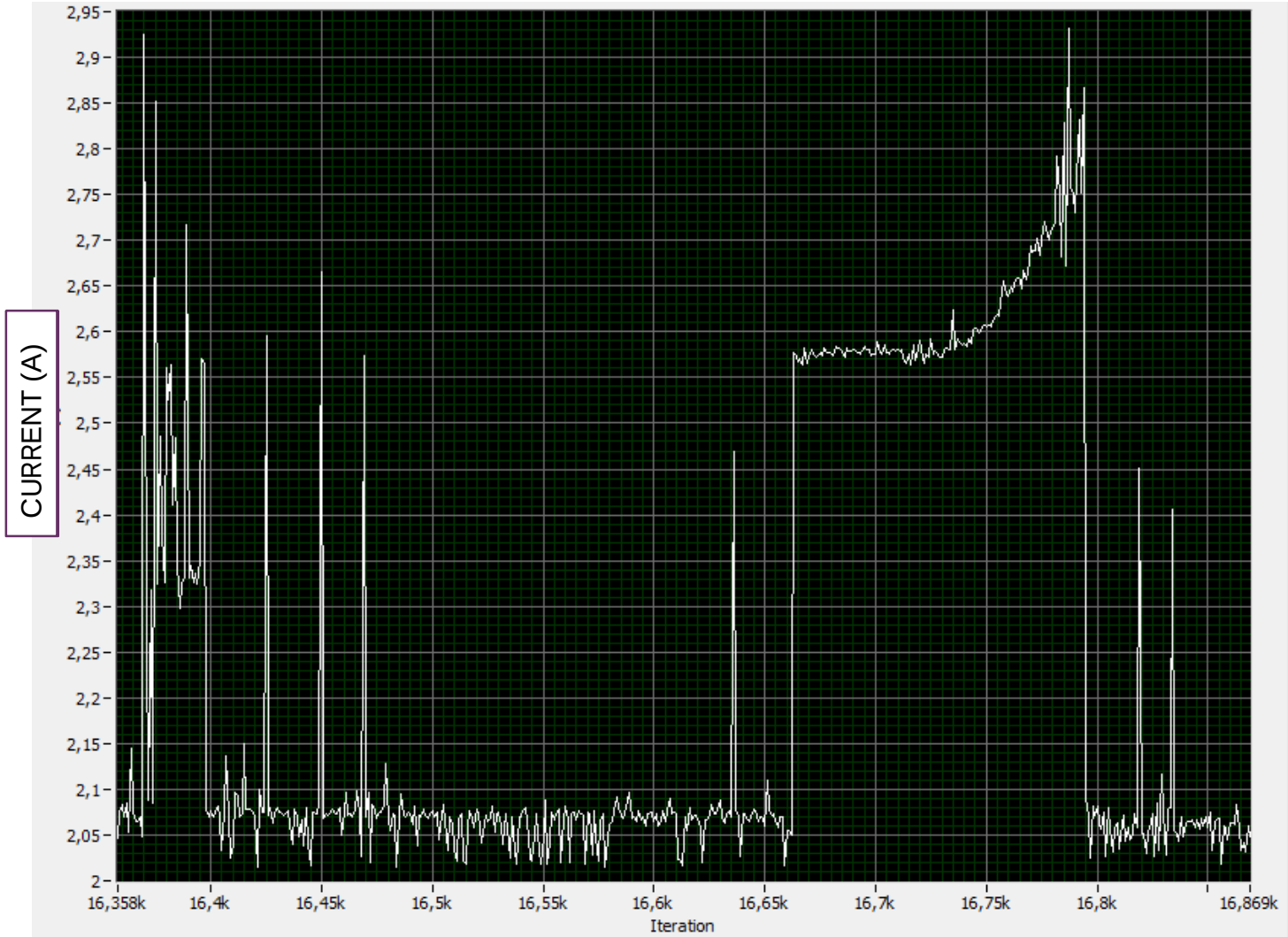
Performance	
# of Cores	8
# of Threads	16
Processor Base Frequency	2 GHz
Max Turbo Frequency	2.6 GHz
TDP	45 W

Broadwell-DE, Xeon D 8-Core

SSDs

XEON D-1540

Netpipe test Power Consumption @ 10Gbs
Cstate enabled, on demand governor

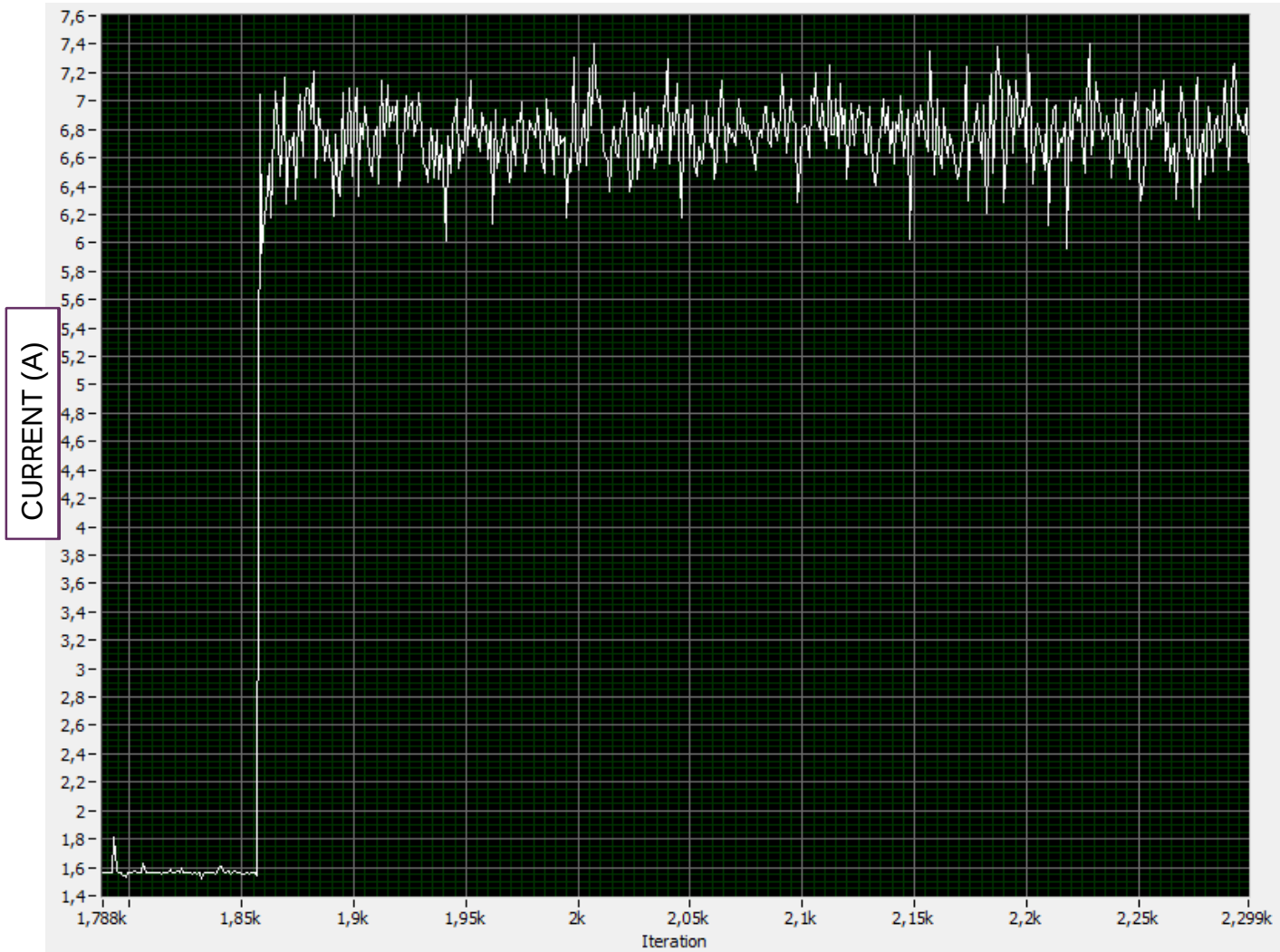


Power consumption is about 34W @10Gbs

TDP is 45W

XEON D-1540

Stess test, 8 core Power Consumption
Cstate enabled, on demand governor



Power consumption is about 90W under the stress test

TDP is 45W

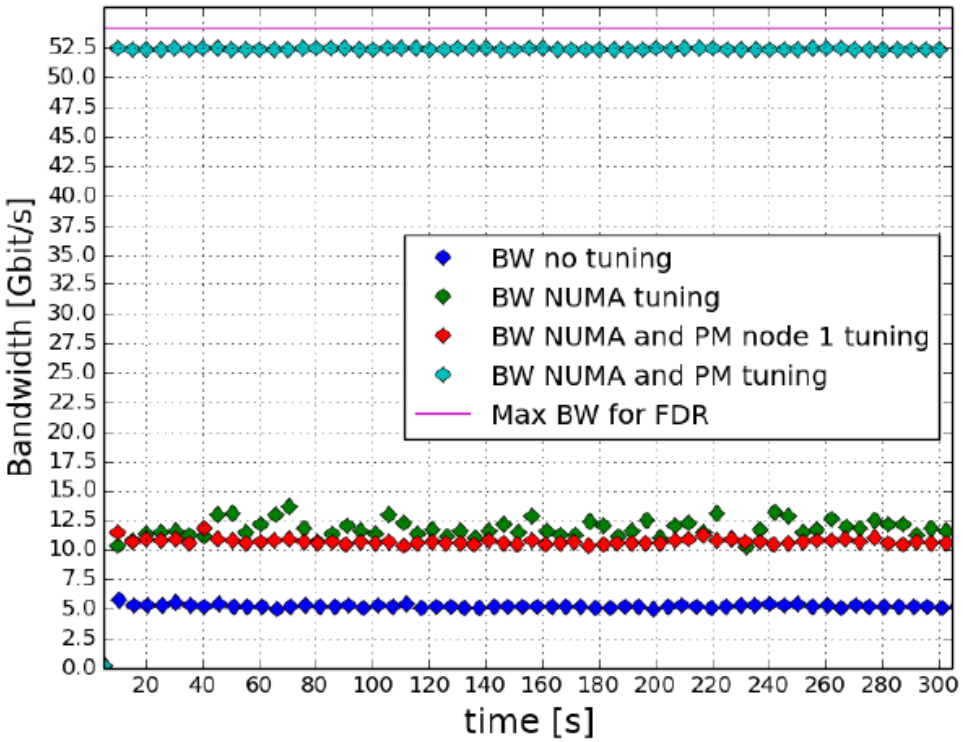
Low Power Techniques

- **A P-state is a voltage and frequency operating point**
- **C-states are idle states**
  - The basic idea of these modes is to cut the clock signal and power from idle units inside the CPU. The more units you stop, reduce the voltage or even completely shut down, more energy you save
- C-states and P-states are also orthogonal

Mode	Name	What it does	CPUs
C0	Operating State	CPU fully turned on	All CPUs
C1	Halt	Stops CPU main internal clocks via software; bus interface unit and APIC are kept running at full speed.	486DX4 and above
C1E	Enhanced Halt	Stops CPU main internal clocks via software and reduces CPU voltage; bus interface unit and APIC are kept running at full speed.	All socket LGA775 CPUs
C1E	—	Stops all CPU internal clocks.	Turion 64, 65-nm Athlon X2 and Phenom CPUs
C2	Stop Grant	Stops CPU main internal clocks via hardware; bus interface unit and APIC are kept running at full speed.	486DX4 and above
C2	Stop Clock	Stops CPU internal and external clocks via hardware	Only 486DX4, Pentium, Pentium MMX, K5, K6, K6-2, K6-III
C2E	Extended Stop Grant	Stops CPU main internal clocks via hardware and reduces CPU voltage; bus interface unit and APIC are kept running at full speed.	Core 2 Duo and above (Intel only)
C3	Sleep	Stops all CPU internal clocks	Pentium II, Athlon and above, but not on Core 2 Duo E4000 and E6000
C3	Deep Sleep	Stops all CPU internal and external clocks	Pentium II and above, but not on Core 2 Duo E4000 and E6000; Turion 64
C3	AltVID	Stops all CPU internal clocks and reduces CPU voltage	AMD Turion 64
C4	Deeper Sleep	Reduces CPU voltage	Pentium M and above, but not on Core 2 Duo E4000 and E6000 series; AMD Turion 64
C4E/C5	Enhanced Deeper Sleep	Reduces CPU voltage even more and turns off the memory cache	Core Solo, Core Duo and 45-nm mobile Core 2 Duo only
C6	Deep Power Down	Reduces the CPU internal voltage to any value, including 0 V	45-nm mobile Core 2 Duo only

C-States trap

…when you are doing I/O only…
….pay attention to the power management



(@ Falabella A., Manzali M. - Update on the LHCb Event Builder developments for the Upgrade)

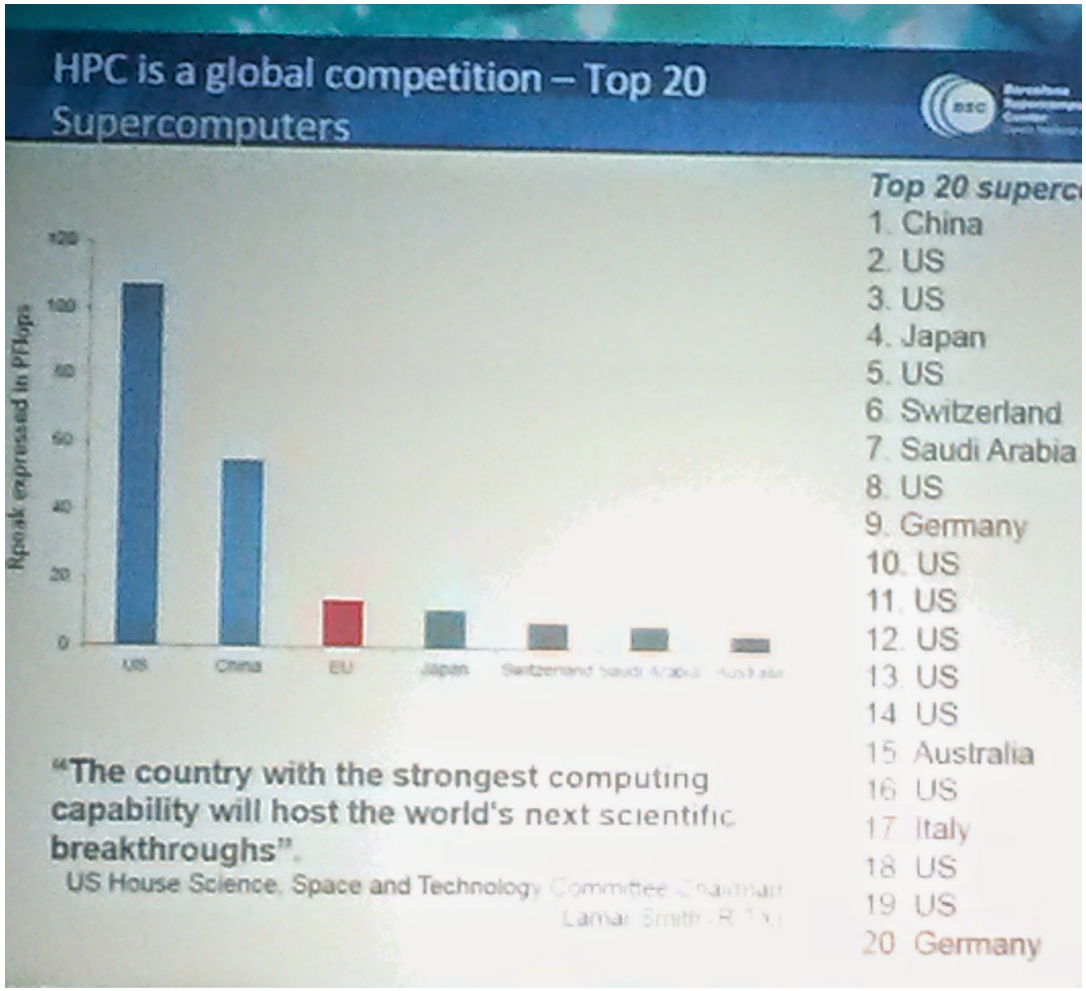
- Client-Server Infiniband tests (FDR) with C-state transitions allowed/not allowed on the nodes
- Close to saturation only if C-states transition are not allowed

Outline

- Introduction
- A little bit of history
- Modern Systems on Chip
- Computing on System on Chip, the INFN experience
- European Projects on low power computing
- Conclusion

HPC Global Competition

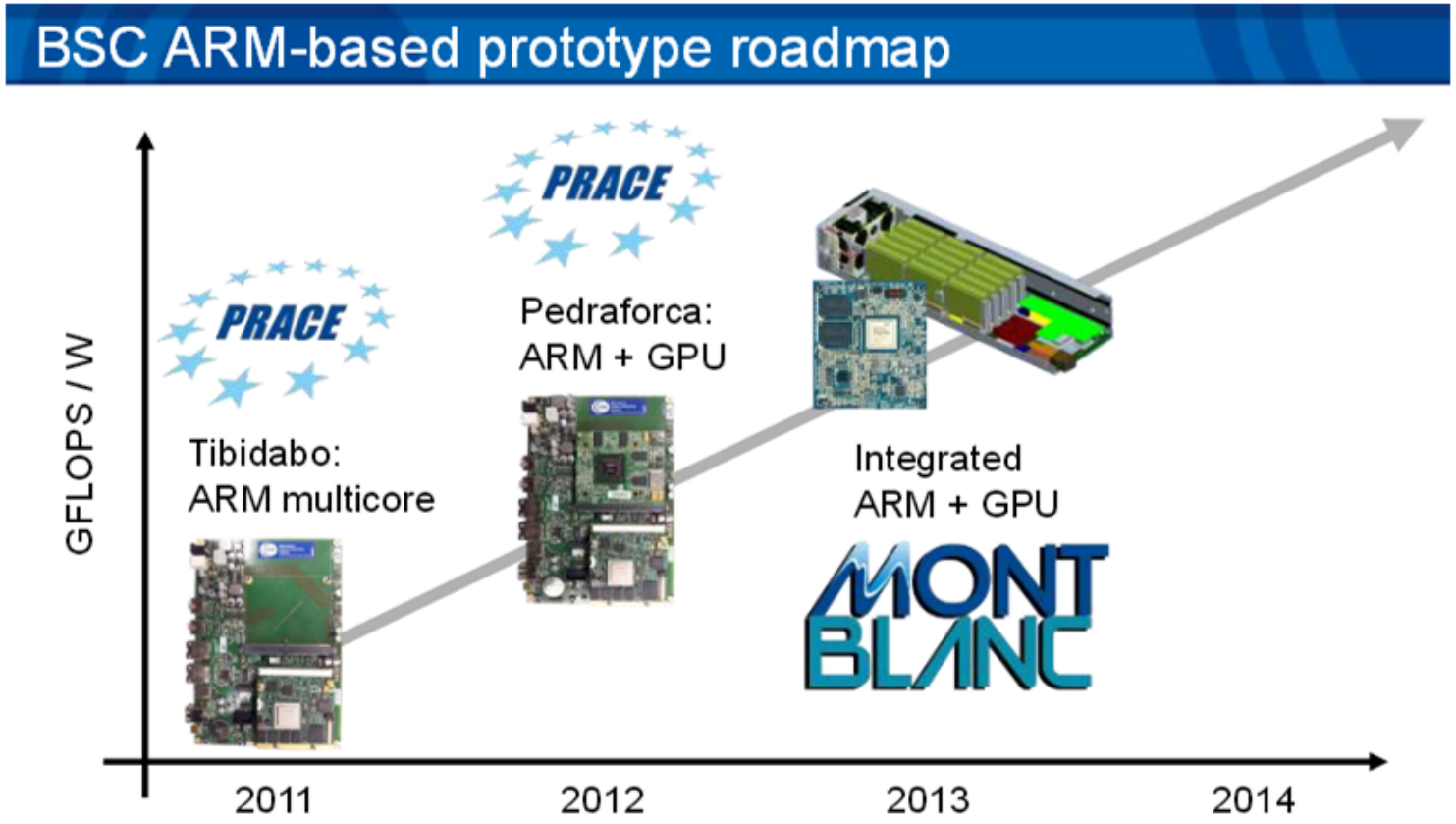
EC ICT2015 Event in Lisbon



European leadership in low power?

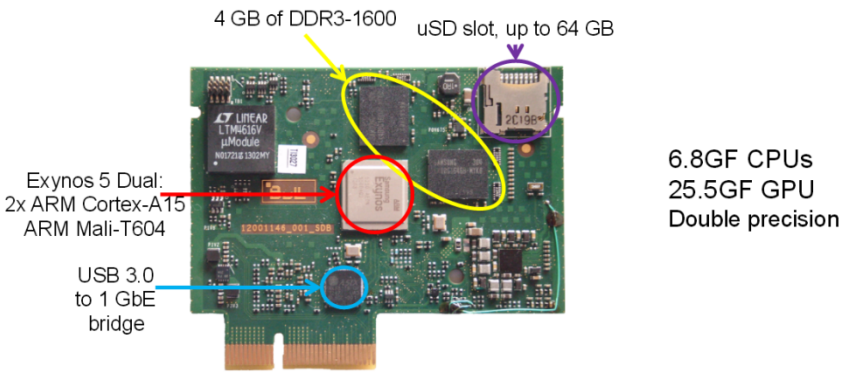
- Mobile CPU and GPU (IP licenses)
  - ARM
  - Imagination Technology (iPhone GPU, MIPS tech)
- Embedded/Automotive/Avionics
  - Siemens
  - Bosch
  - ST
  - Infineon

The MontBlanc Project - 1



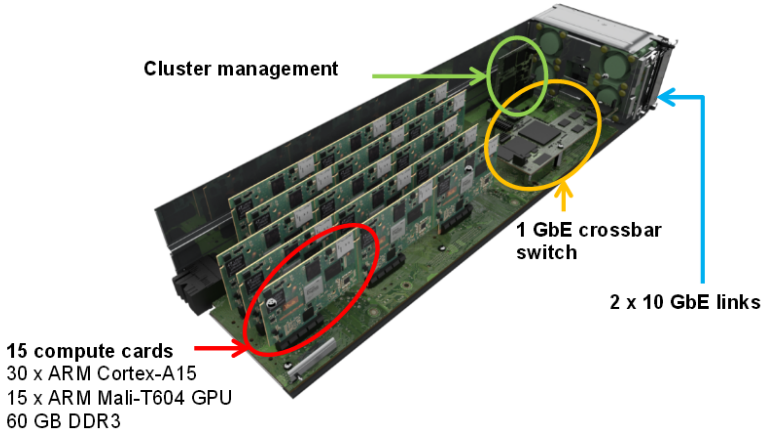
- Prototypes are critical to accelerate software development
  - System software stack + applications

The MontBlanc Project - 2

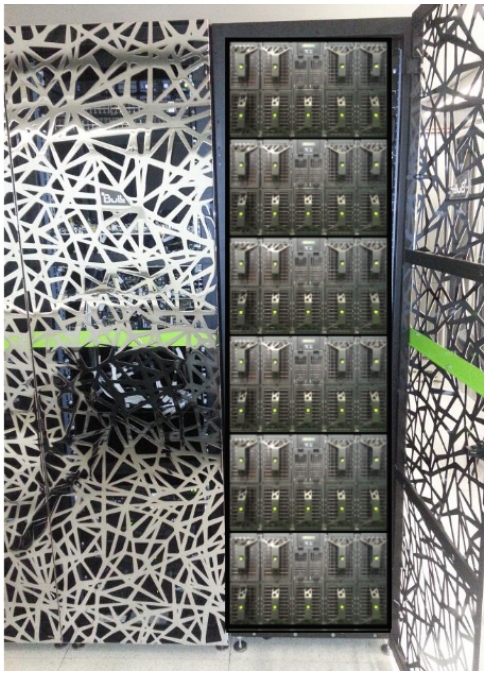


6.8GF CPUs
25.5GF GPU
Double precision

- 15 node-cluster in a standard Bull B505 enclosure



- 6 BullX chassis
- 54 Compute blades
- 810 Compute cards
  - 1620 CPU
  - 810 GPU
  - 3.2 TB of DRAM
  - 52 TB of Flash
- 26 TFLOPS
- 18 kWatt



ExaNest + ExaNoDe Projects

European Exascale System Interconnect and Storage

ExaNeSt

- Home
- Partners
- Logo
- News / Press
- Publications
- Contact
- Private Area

This Horizon2020 FET-HPC *ExaNeSt* project develops and prototypes solutions for some of the crucial problems on the way towards production of Exascale-level Supercomputers:

WHAT we do:

ExaNeSt develops and prototypes solutions for **Interconnection Networks**, **Storage**, and **Cooling**, as these have to evolve in order for the production of *exascale-level supercomputers* to become feasible. We tune real HPC **Applications**, and we use them to evaluate our solutions. *Click* for more information...

WHY we do it:

HPC is a precious tool for all of modern technology, science, and society. For the next generation of HPC systems, we need millions of low-power-consumption computing cores, tightly interconnected and packaged together and appropriately cooled, and with a new storage architecture. *Click* to see why...

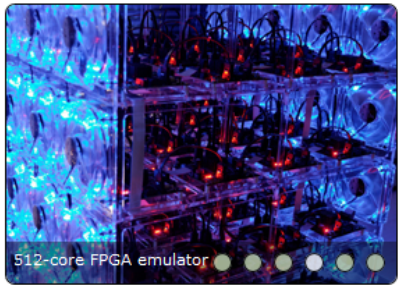
HOW we do it:

We use the *UNIMEM* Global Address Space and zero-copy send/receive operations, and we tune our applications for such architectures. We develop efficient networking with all its frequent-path functions implemented in hardware, including advanced congestion management for full utilization at low latency; we distribute the file storage on top of that. *Click* to see more...

WHO we are:

The ExaNeSt Consortium combines industrial and academic research expertise, especially in the areas of system cooling and packaging, storage, interconnects, and the HPC applications that drive all of the above. *Click* to see who we are

performance_low_power - Excel



Latest News

- **30 September 2015**: FET-HPC Projects Meeting
- **25 September 2015**: HPC User Forum

View all news

Project Info:

Call: H2020-FETHPC-1-2014 – (a) core technologies and architectures

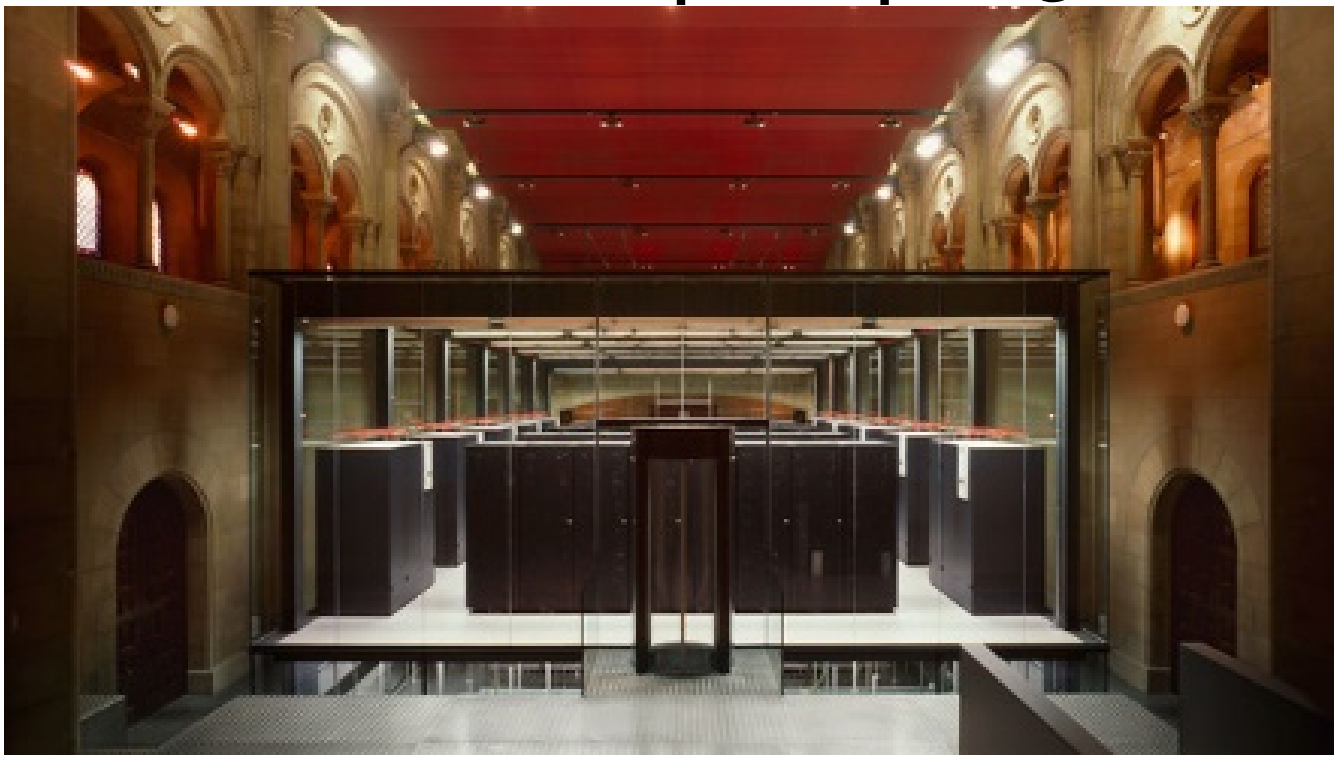
Project Number, CORDIS Info: 671553

Duration: 01 Dec. 2015 – 30 Nov. 2018 (36 months).

Budget: Cost: 8.44 M€, EU contribution: 8.44

The Road to ExaScale

Mare Nostrum supercomputer @BSC



Architecture	Intel Xeon E5 (Sandy Bridge)
Power	1,015.60 kW
Operating system	SUSE Linux
Memory	104.6 TB
Storage	2 PB of GPFS disk storage
Speed	1,017.04 TFlops
Ranking	TOP500: 77, June 2015
Web site	http://www.bsc.es

1MW for 1PFLOPS

1GW for 1EFLOPS

"And what comes after exascale? We can look forward to zettascale (10^{21}) and yottascale (10^{24})…..
….Then we run out of prefixes."
(from: http://www.kurzweilai.net/designing-the-exascale-computers-of-the-future)

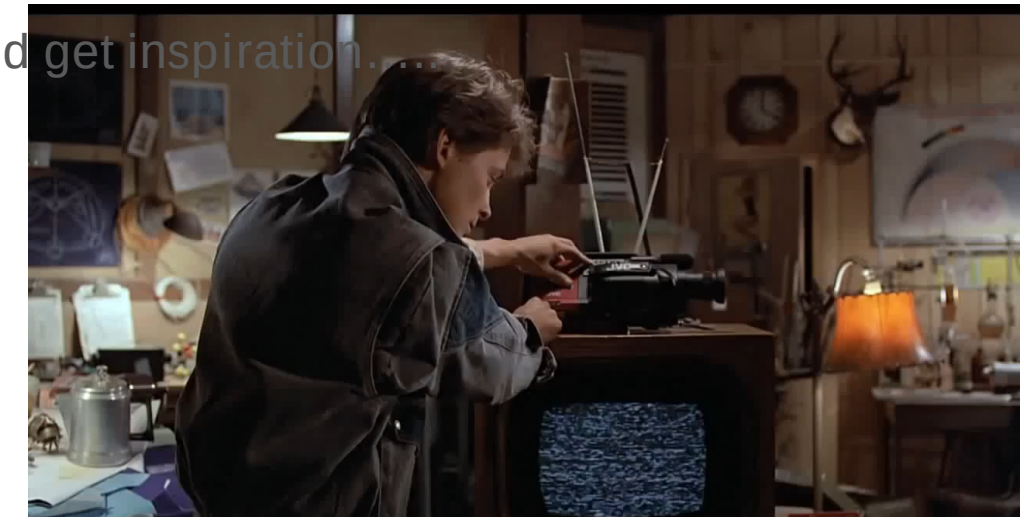
The Road to ExaScale

You have to choose….

- Use a dedicated nuclear reactor

Or…

- …use low power components

Or…

- …..again look back to the 80s and get inspiration….