Analyse des performances et de la consommation des processeurs embarqués

Jean-François Méhaut
UJF-CEA/DRT, LIG, Grenoble

4 Octobre 2012

Agenda

- Embedded Computing Technologies
  - ARM core processors
  - ARM Accelerators
- European Mont-Blanc Project
  - Objectives
  - Consortium
  - Platforms

ARM Ltd

- Founded in November 1990
  - Spun out of Accorn Computers
- Designs the ARM range of RISC processors
- Licenses ARM core designs to semiconductors partners who fabricate and sell to their customers
  - ARM does not fabricate silicon itself
- Also develop technologies to assist with the design-in of the ARM architecture
  - Software tools, Boards, debug hardware, application software, bus architecture, peripherals,...

ARM Partnership Model

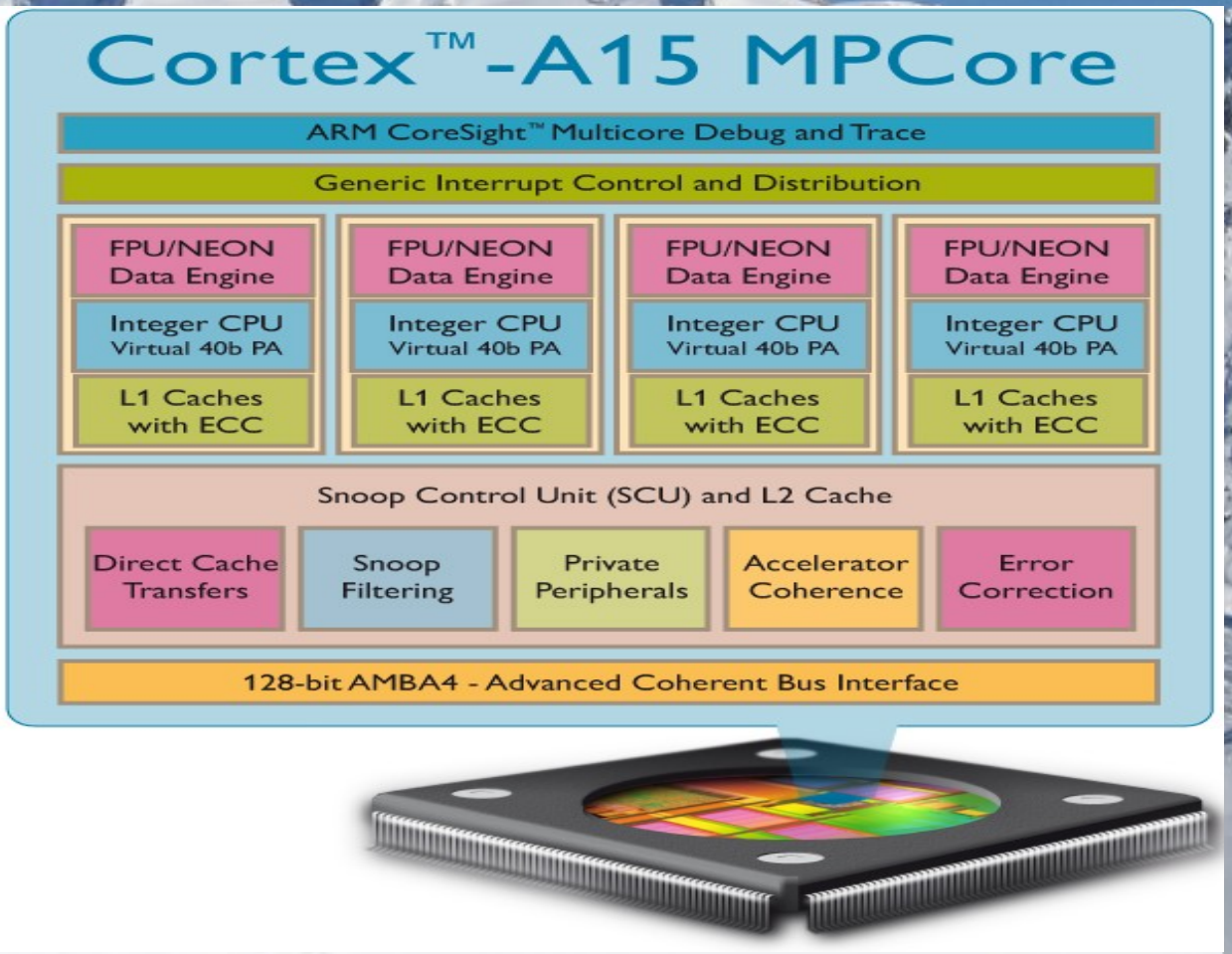
ARM Powered Products

Overview of ARM Processors

- Focusing on Cortex A9 & Cortex A15
- ARM ships no processors but only IP cores
  - For SoC integration
- Targeting markets
  - Netbooks, tablets, smart phones, game console
  - Digital Home Entertainement
  - Home and Web 2.0 Servers
  - Wireless Infrastructure
- Design Goals
  - Performance, power, Easy Synthesis
  -

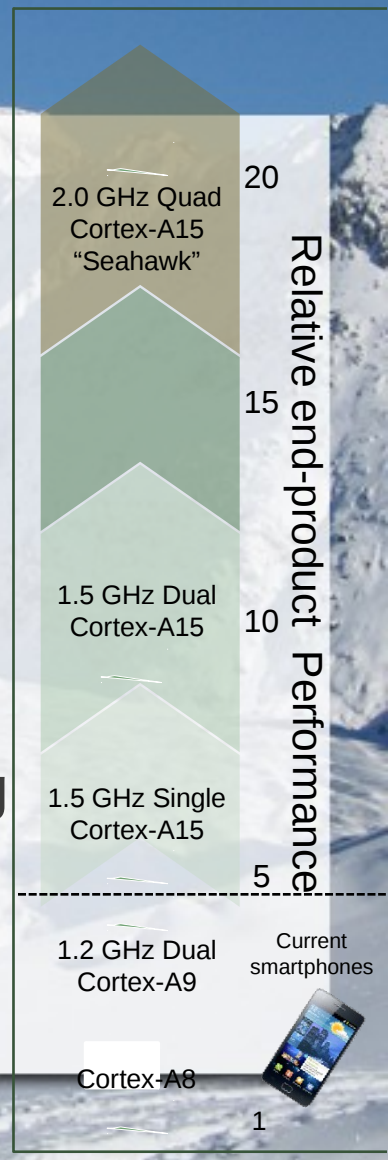
ARM Cortex A9/A15

- 1-4 Cores
- Out-of-Order Superscalar
- Branch Predictor
- 32KB L1 I/D caches
- ~4M L2 caches with coherency
- NEON (SIMD) & FPU
- 32/28nm (A15)
  45nm (A9)

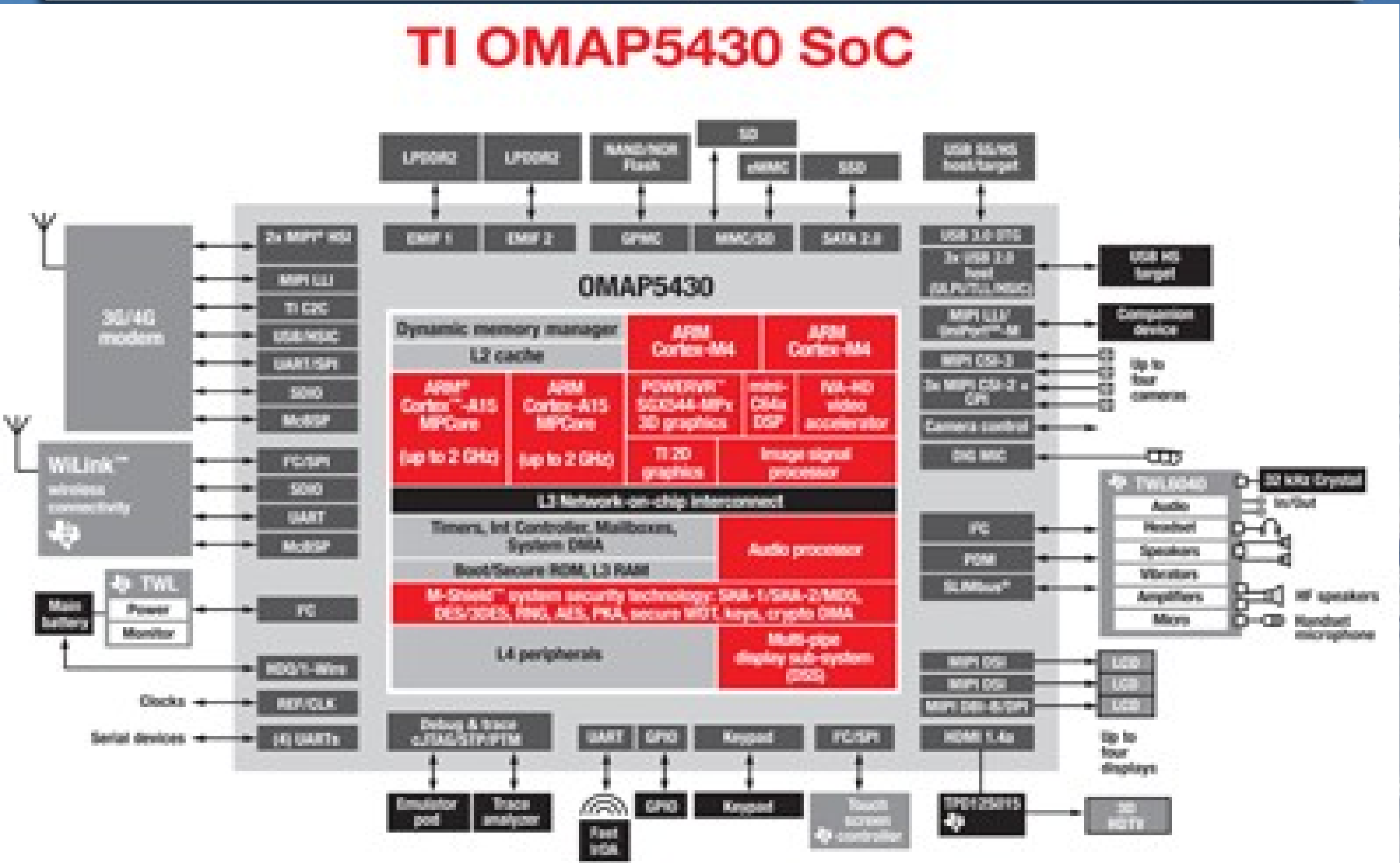


Cortex-A15 – Low Power Computing

- CORTEX-A15 is ARM's highest performance applications processor to date
  - Delivers 2x performance of current super smartphone in same low-power envelope
  - Can deliver up to 8x performance for mobile applications
- Wide range of end markets
  - Consummers electronics, digital TVs, mobile computing
  - Enterprise computing, servers, clouds, networking
- First SoC arriving in market shortly



Texas Instrument OMAP 5



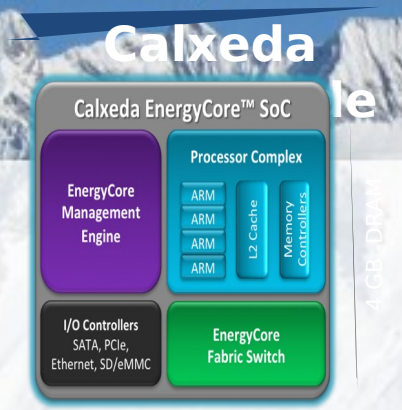
Comparison of ARM, Atom, i7

	Cortex A15 (no L2, 32 nm)	Cortex A9 (no L2, 40 nm)	Atom N270 (45 nm)	I7 960 (45 nm)
Number of cores	2 (4 maximum)	2 (4 maximum)	1 Core 2 HT Threads	4 Cores 8 HT Threads
Frequency	1 Ghz-2.5 Ghz	800 Mhz, 2 Ghz	1,6 Ghz	3,2 Ghz
Out of Order?	Yes	Yes	No	Yes
L1 cache size	32 KB I/D	32 KB I/D	32 KB I/D	32 KB I/D
L2 cache size	N/A	N/A	512KB	1 MB + 8MB L3
Issue width	4	4	2	4?
Pipeline Stages	?	8	16	14~24 (?)
Supply Voltage		1,05V	0,9 – 1,116V	0.8
Transistor count	?	26,000,000	47,000,000	731,000,000
Die Size	?	4,6 mm2 6,7 mm2	26 mm2	263mm2
Power Consumption	?	0,5W 1,9 W	2,5 (TDP)	130W (TDP)

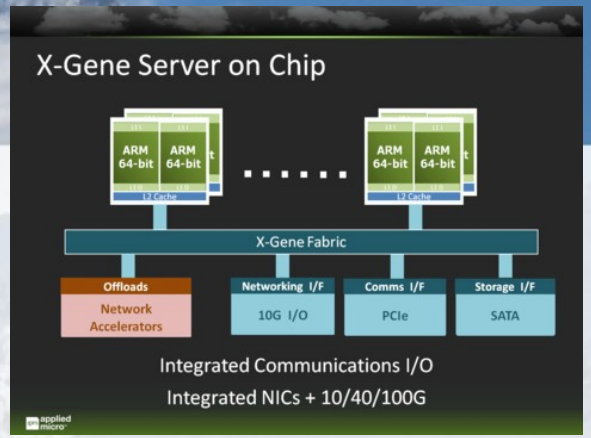
Comparison of ARM SoC, Atom, i7

	TI OMAP5 (28nm)	Nvida Tegra 2 (40 nm)	Atom N450 (45 nm)	I7 2600S (32 nm)
CPU cores	2 x A15 2 x M4	2 x A9	1 Core, 2 HT threads	4 Cores, 8 HT threads
CPU Frequency	2 Ghz (A15)	1 Ghz	1,66Ghz	2,6 Ghz
GPUs ASICs	Video, Audio, Encryption, Display 2D/3D	8 x GPUs, Audio, Video, ISP	1 GPU	1 GPU
L2	?	1 MB	512 KB	1 MB + 8 MB
Die Size	?	49 mm2	66m2	?
Transistors	?	260,000,000	123,000,000	?
Package Size	17x17 mm2	23x23mm2	22x22mm2	37.5x37.5 mm2
Power Consumption	?	150~500 W?	5,5 W (TDP)	65 W (TDP)

Extending Technology into the Data Centers



- Quadcore Cortex-A9
- 1.2GHz
- 5W power including 4GB DRAM
- Integrated server fabric
- PCIe
- GbE
- USB 2.0



- Up to 32, 64-bit cores per SoC. 128 coherent processors per system
- 3GHz
- Network processors
- PCIe
- SATA
- 10GbE

CALXEDA
Power your tomorrow™

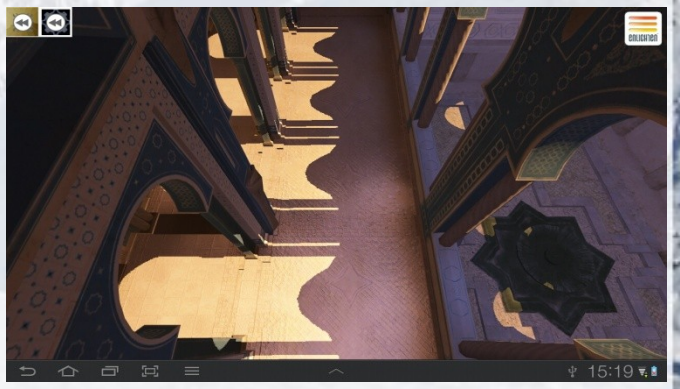
applied micro

The challenges for Accelerators

- Benefit of offload must also outweigh the cost of offload
  - Additional software complexity
  - Additional data transfers
  - Overcome Transfer Latencies and associated power management
- The General Purpose Processor's ISA also advances
  - Floating point moves to FPU within CPU
  - DSP, SIMD, Crypto,...
- Only the large API abstracted, or edge accelerators appear to have a viable future
  - Menwhile the GPP can start to optimize through specialization

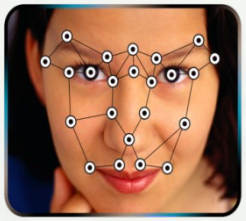
The Integration of Graphics

- Graphics is a large (very large) API abstracted, “edge” accelerator
- Graphics is already a priority for anything with a screen
  - Smartphones,DTVs, Tablets, hand-held game consoles, In-Car Entertainment
- Large and Growing Market for GPU
  - 4 billion Internet Connected Screens in 2016, most with embedded graphics
- Looking towards a GP-GPU future?
  - … but is it then still at the on the edge?



Mali for Graphics and GPU Computing

Performance



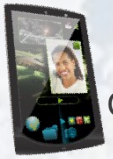
Skrymir

Mali-T658
High end solution
Maximum compute capability

Mali-T604
First Midgard architecture product
Scalable to 4 cores

Date of production chips

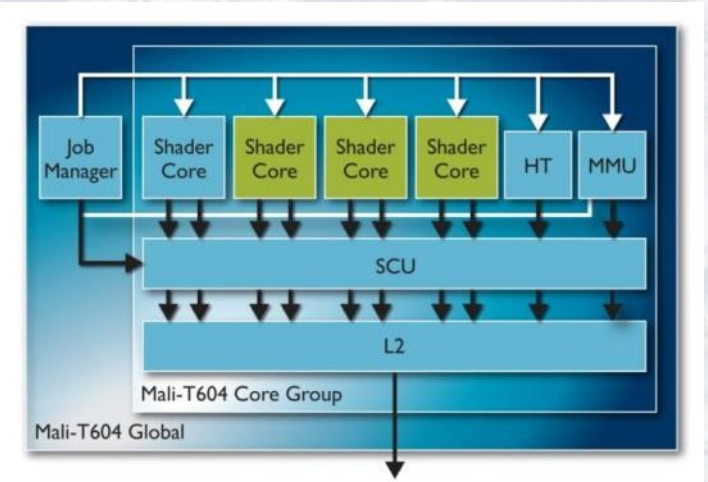
2012 2013 2014



- Designed for GPU Computing
  - Uncompromised support for OS/API choice
- Closer CPU-GPU links
  - Efficient Use of all device resources
  - Maximize performance and battery life
- Protecting partner investments
  - Common software platform reduces cost and TTM
  - Multicore delivers performance scalability over multiple form factors
- Roadmap for further share gains
  - Mali-T600 silicon shipping in consumer products in H2 2012
  - Skrymir driving design wins in next generation super-smartphines and mobile computeDesigned for GPU Computing

Mali-T604: ARM's First GP-GPU

- Innovative GPU architecture
  - Designed for performance and flexibility
  - Leading graphics performance
  - Architected for GPU computing
- State of the art bandwidth reduction
  - Optimized tile-based architecture
  - Transaction elimination
  - Hierarchical tiling
- A system approach
  - ARM CPU + GPU + Memory + Interconnect
  - Efficient, high-throughput design



Mont-Blanc - Objectives

- Develop prototypes of HPC clusters using low power commercially available embedded technology.
- Design the next generation in HPC systems based on embedded technologies and experiments on the prototypes
- Develop a portfolio of existing applications to test these systems and optimize their efficiency, using BSC's OmpSs programming model (11 existing applications were selected for this portfolio)

Mont-Blanc Consortium

- **Barcelona SuperComputing (BSC)**
  - **Prototype hosting, programming model,...**
- **Bull SA**
  - **System Integration, System software**
- **ARM**
  - **Architecture (multicore, node, system)**
- **Gnodal**
  - **Interconnect switch**
- **Genci**
  - **Applications (CEA INAC), Architecture (CEA Leti)**
- **CINECA**
  - **Apps.**
- **JSC, LRZ**
  - **Apps, power consumption**
- **CNRS**
  - **Architecture (LIRMM), Apps (CORIA, LMA, LIG) Auto tuning (LIG)**

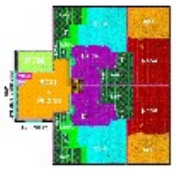
Mont-Blanc - Applications

Code	Scientific Domain	Contact	Institution
YALES2	Combustion	V. Mouveau	CNRS/CORIA
EUTERPE	Fusion	T. Akgun	BSC
SPECFEM3D	Wave propagation	D. Komatitsch	CNRS
MP2C	Multi-particle collision	G. Sutmann, A. Schiller	JSC
BigDFT	Elect. Structure	T. Deutsch	**CEA INAC**
Quantum Expresso	Elect. Structure	C. Cavazzoni	CINECA
PEPC	Coulomb + gravitational forces	P. Gibbon, L. Arnold	JSC
SMMP	Protein folding	J. Meinke	JSC
ProFASI	Protein folding	S. Mohanty	JSC
COSMO	Weather forecast	M. Culpo	CINECA
BQCD	Particle physics	M. Allalen	LRZ

Montblanc Platforms

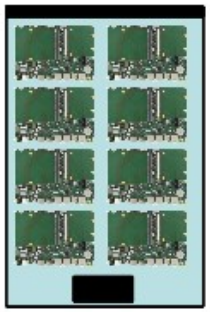
13 blades of 8 harmony Tegra 2 boards already deployed

- 35W/blade
- 16 ARM Cortex A9 1GHz cores/blade
- Will be pushed to 256 nodes (32 blades)



Tegra2 SoC:
2x ARM Corext-A9 Cores
2 GFLOPS
0.5 Watt

Tegra2 Q7 module:
1x Tegra2 SoC
2x ARM Corext-A9 Cores
1 GB DDR2 DRAM
2 GFLOPS
~4 Watt
1 GbE interconnect



1U Multi-board container:
1x Board container
8x Q7 carrier boards
8x Tegra2 SoC
16x ARM Corext-A9 Cores
8 GB DDR2 DRAM
16 GFLOPS
~35 Watt

Rack:
32x Board container
10x 48-port 1GbE switches
256x Q7 carrier boards
256x Tegra2 SoC
512x ARM Corext-A9 Cores
256 GB DDR2 DRAM
512 GFLOPS
~1.7 Kwatt

300 MFLOPS / W

Snowball Board (ST Ericsson)

- Full embedded system
  - Dual Core **ST-Ericsson** ARM 1GHz SoC
  - Neon Floating Point Unit
  - Integrated GPU
  - 1 Gbyte RAM
  - HDMI output
  - Ethernet
  - Runs Linux (Linaro Ubuntu) or Android
  - Igloo community for development and support
  - 2.5W maximum energy consumption

Conclusions - Mont-Blanc

- Porting Apps to tibidabo is a success
- Most limitation come from the prototype and should be lifted in the next version
- Second round of porting is to begin
- SPECFEM3D and BigDFT kernels have been selected specifically for optimization