http://www.montblanc-project.eu

The Mont-Blanc approach towards Exascale

Alex Ramirez

Barcelona Supercomputing Center

Disclaimer: Not only I speak for myself ... All references to unavailable products are speculative, taken from web sources. There is no commitment from ARM, Samsung, Intel, Nvidia, or others, implied.

This project and the research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 288777.

Outline

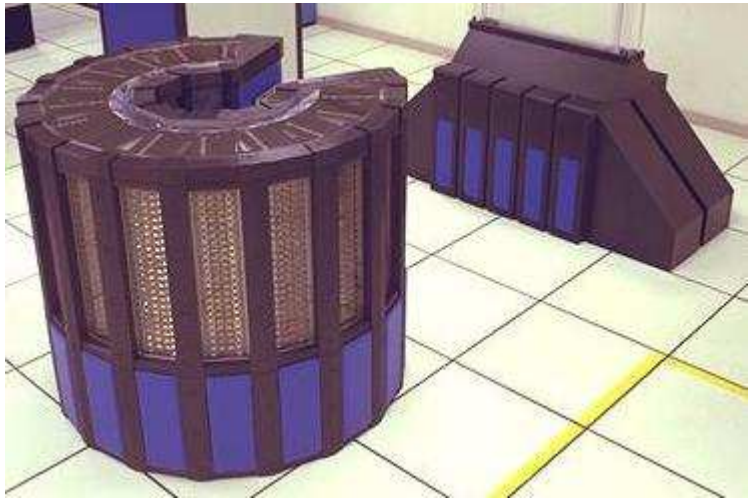
- A bit of history
  - Vector supercomputers
  - Commodity supercomputers
  - The next step in the commodity chain
- Supercomputers from mobile components
  - Homogeneous architecture
  - Compute accelerators
  - Rely on OmpSs to handle the challenges
- BSC prototype roadmap
- Mont-Blanc project goals

In the beginning ... there were only supercomputers

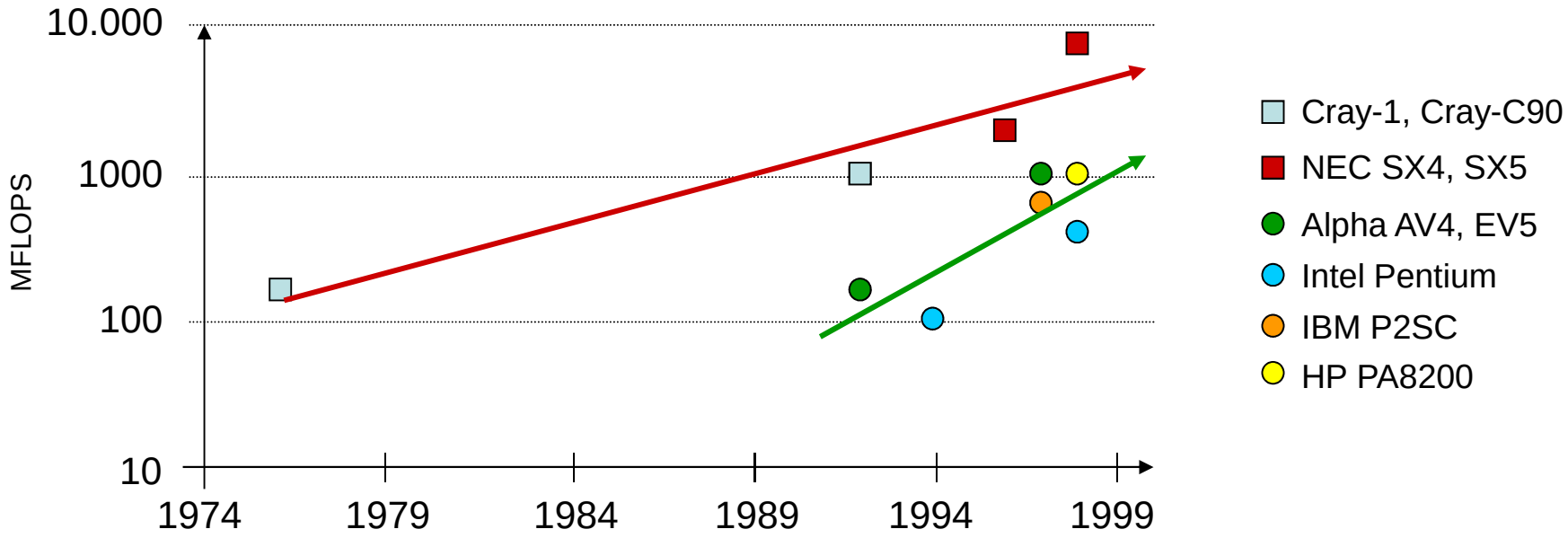
- Built to order
  - Very few of them
- Special purpose hardware
  - Very expensive

- Control Data, Convex, ...
- Cray-1
  - 1975, 160 MFLOPS
    - 80 units, 5-8 M$
- Cray X-MP
  - 1982, 800 MFLOPS
- Cray-2
  - 1985, 1.9 GFLOPS
- Cray Y-MP
  - 1988, 2.6 GFLOPS

- Fortran+vectorizing compilers



The Killer Microprocessors



- Microprocessors killed the Vector supercomputers
  - They were not faster ...
  - ... but they were significantly cheaper and greener
- Need 10 microprocessors to achieve the performance of 1 Vector CPU
  - SIMD vs. MIMD programming paradigms

Then, commodity took over special purpose

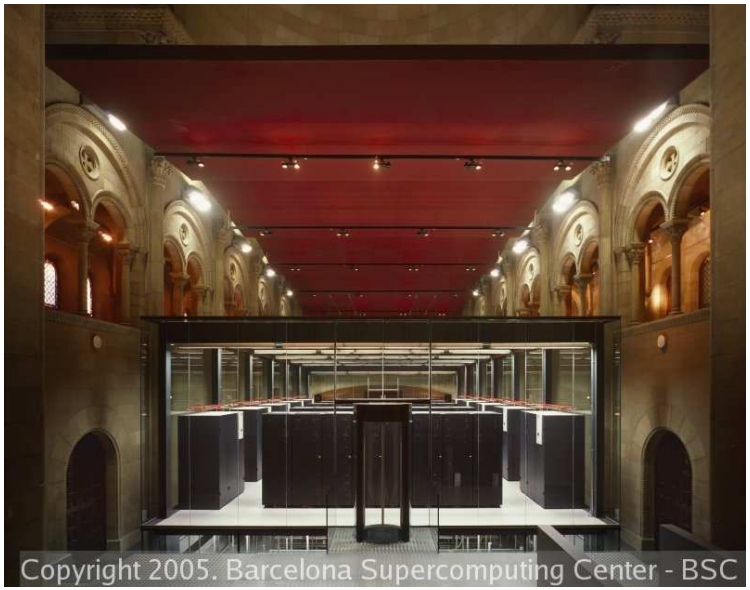
- ASCI Red, Sandia
  - 1997, 1 Tflops (Linpack),
  - 9298 cores @ 200 Mhz
  - 1.2 Tbytes
  - Intel Pentium Pro
    - Upgraded to Pentium II Xeon, 1999, 3.1 Tflops

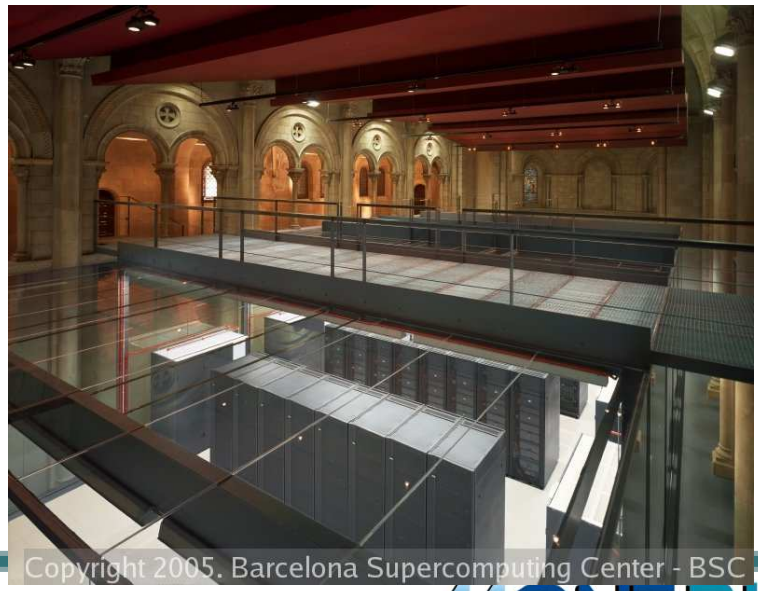
- ASCI White, LLNL
  - 2001, 7.3 TFLOPS
  - 8192 proc. @ 375 Mhz,
  - 6 Tbytes
  - (3+3) Mwats
  - IBM Power 3

Message-Passing Programming Models

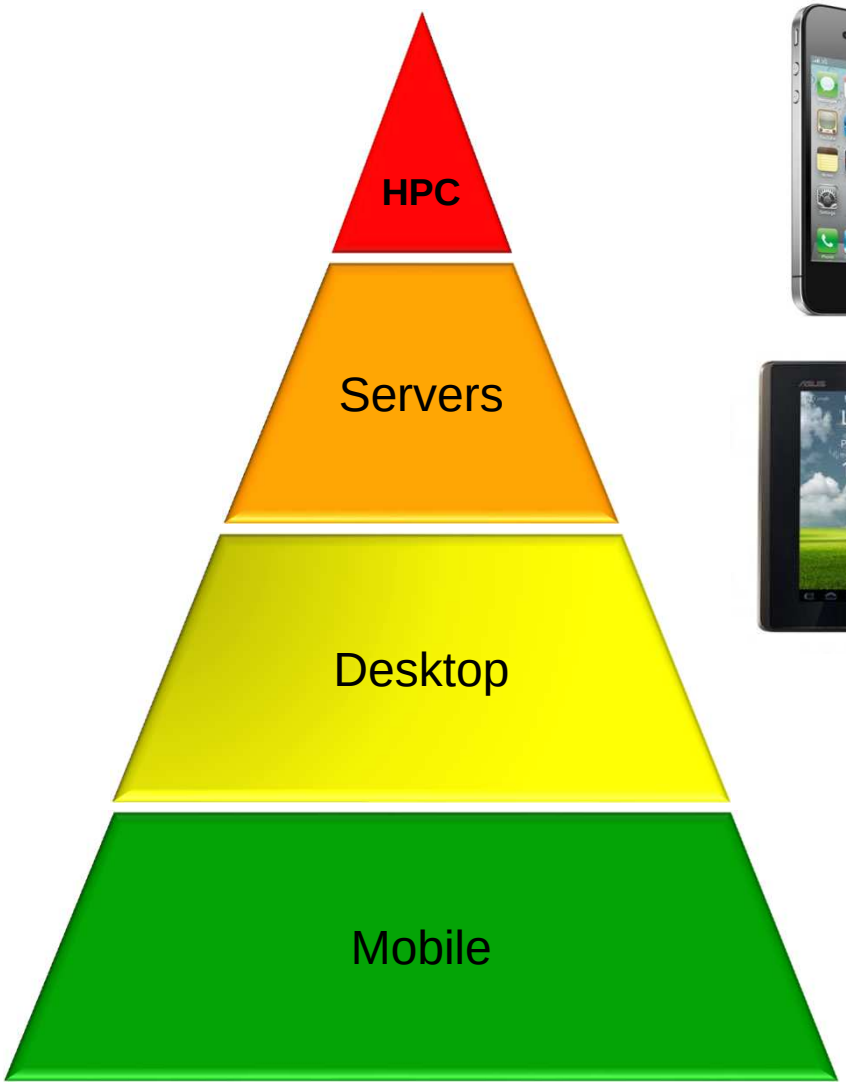
Finally, commodity hardware + commodity software

- MareNostrum
  - Nov 2004, #4 Top500
    - 20 Tflops, Linpack
  - IBM PowerPC 970 FX
    - Blade enclosure
  - Myrinet + 1 GbE network
  - SuSe Linux



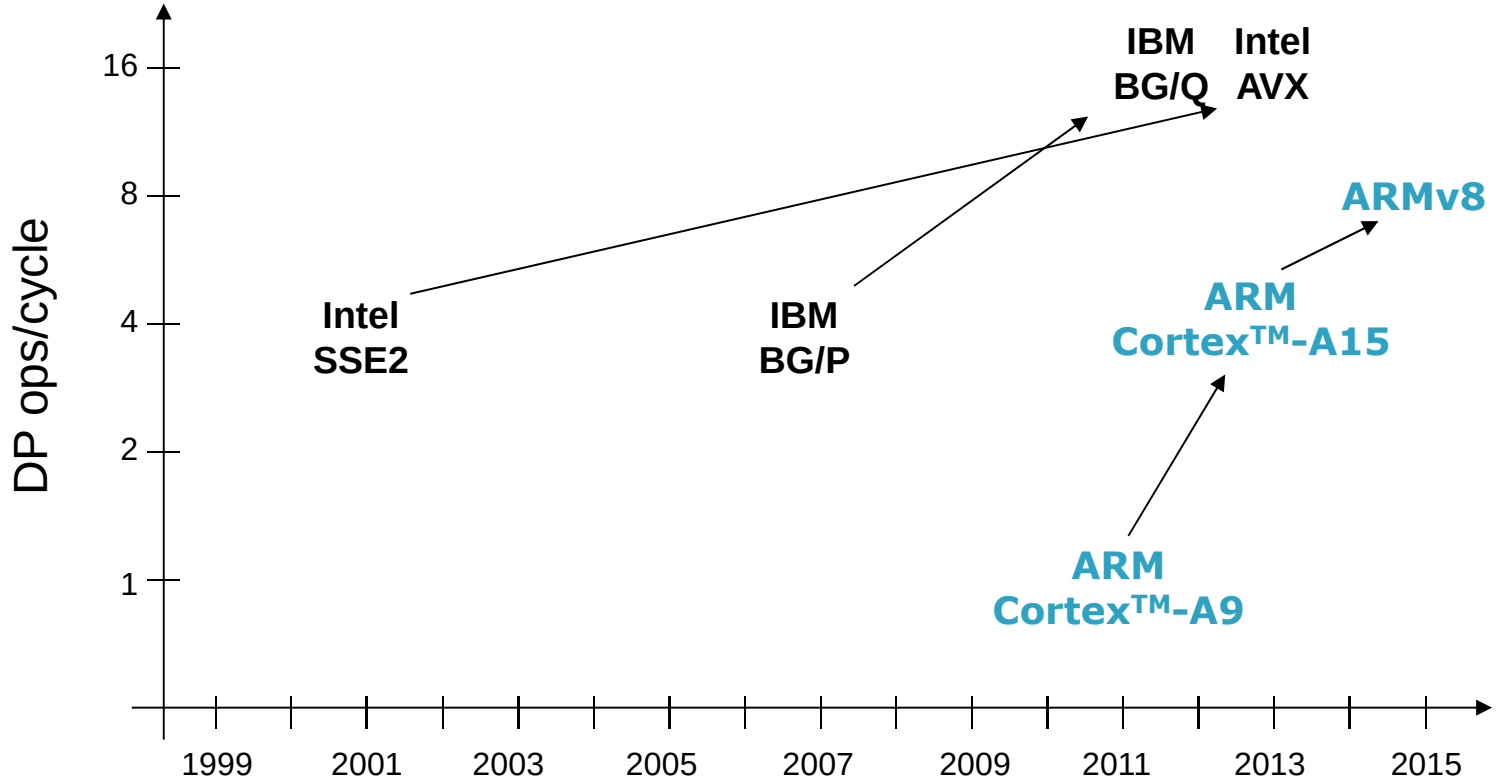


The next step in the commodity chain



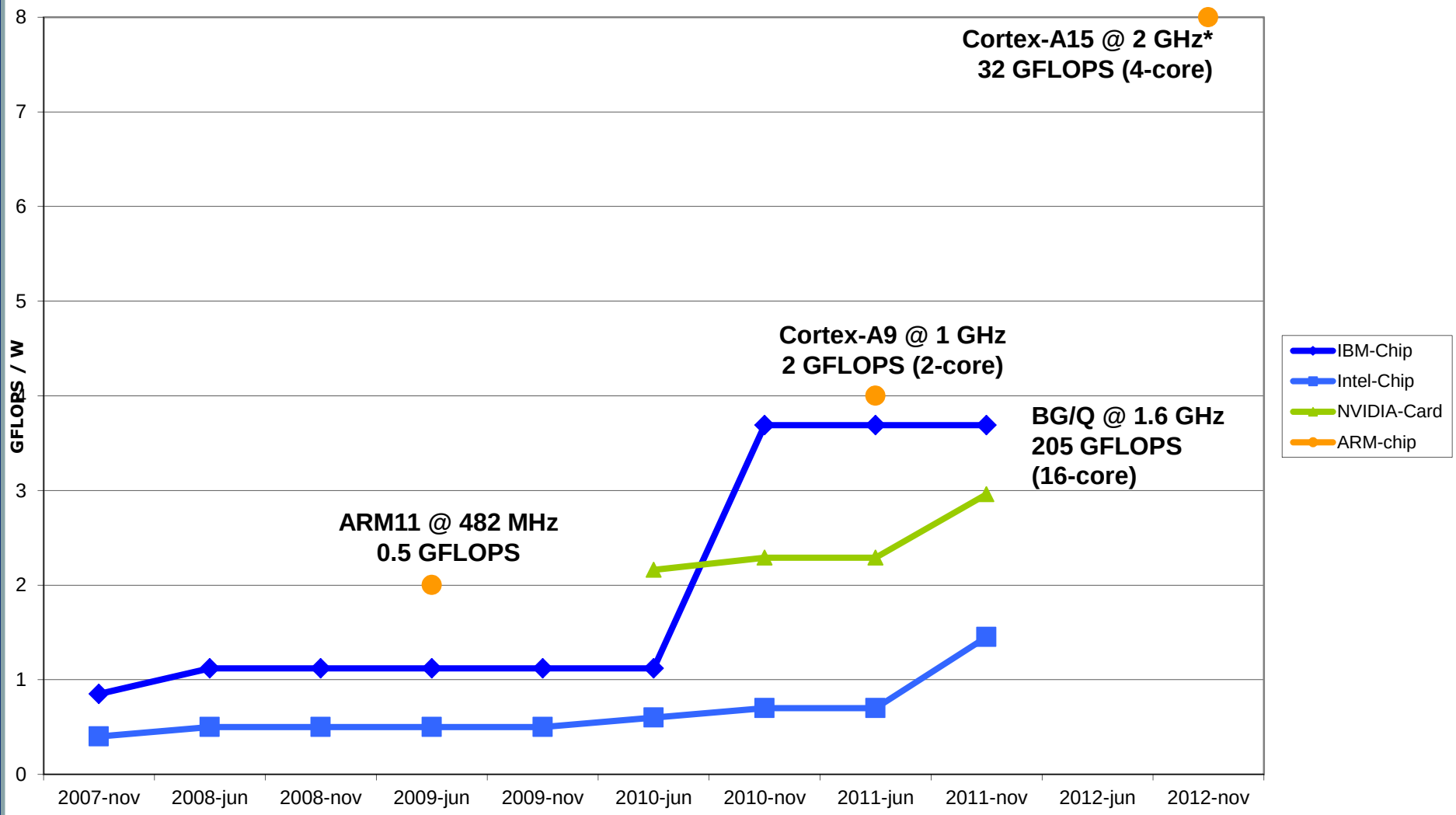
- Total cores in Jun'12 Top500
  - 13.5 Mcores
- Tablets sold in Q4 2011
  - 27 Mtablets
- Smartphones sold Q4 2011
  - > 100 Mphones

ARM Processor improvements in DP FLOPS



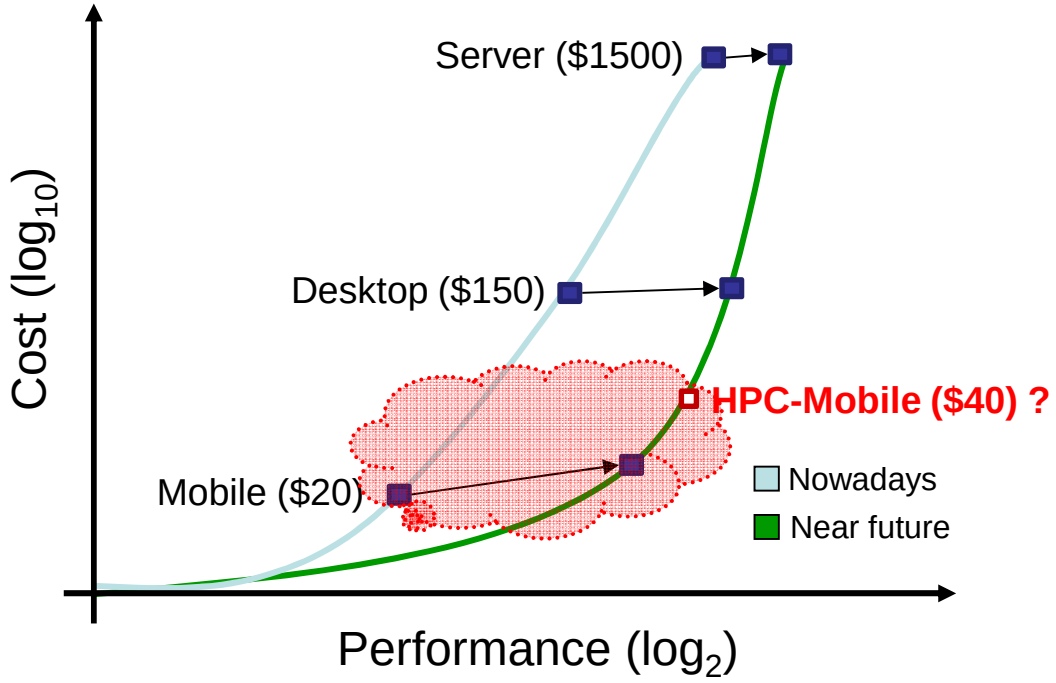
- IBM BG/Q and Intel AVX implement DP in 256-bit SIMD
  - 8 DP ops / cycle
- ARM quickly moved from optional floating-point to state-of-the-art
  - ARMv8 ISA introduces DP in the NEON instruction set (128-bit SIMD)

ARM processor efficiency vs. IBM / Intel / Nvidia



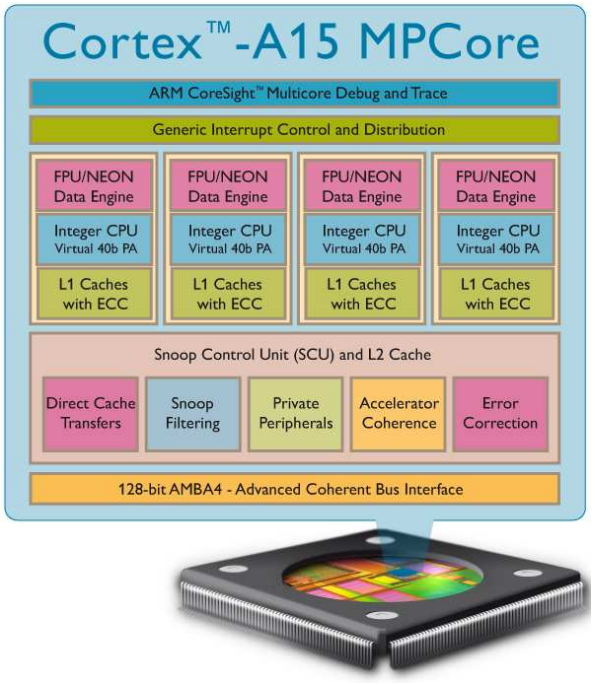
* Based on ARM Cortex-A9 @ 2GHz power consumption on 45nm, not an ARM comitment

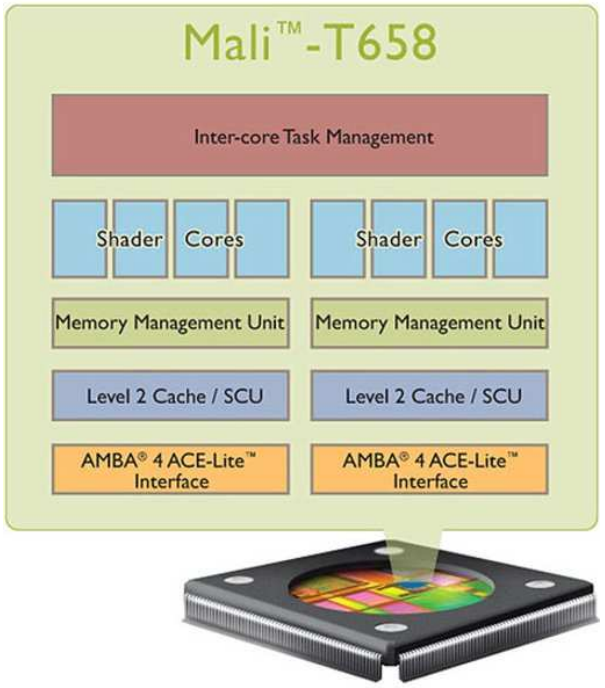
Are the “Killer Mobiles™" coming?



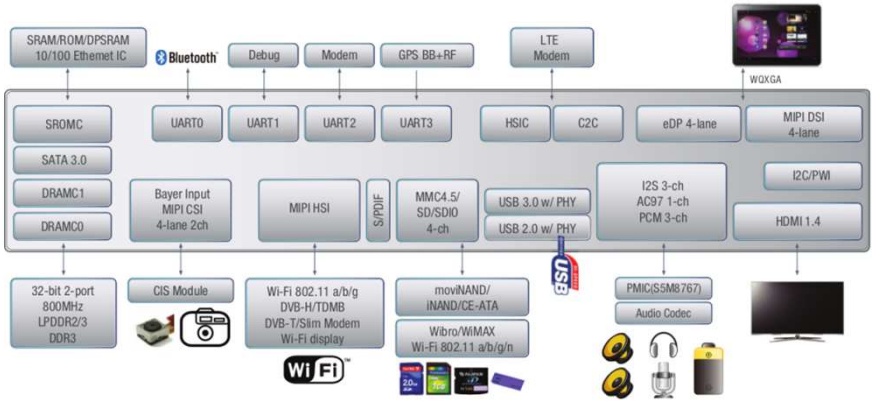
- Where is the sweet spot? Maybe in the low-end ...
  - Today ~ 1:8 ratio in performance, 1:100 ratio in cost
  - Tomorrow ~ 1:2 ratio in performance, still 1:100 in cost ?
- The same reason why microprocessors killed supercomputers
  - Not so much performance ... but much lower cost, and power

Killer mobile™ example: Samsung Exynos 5450 *





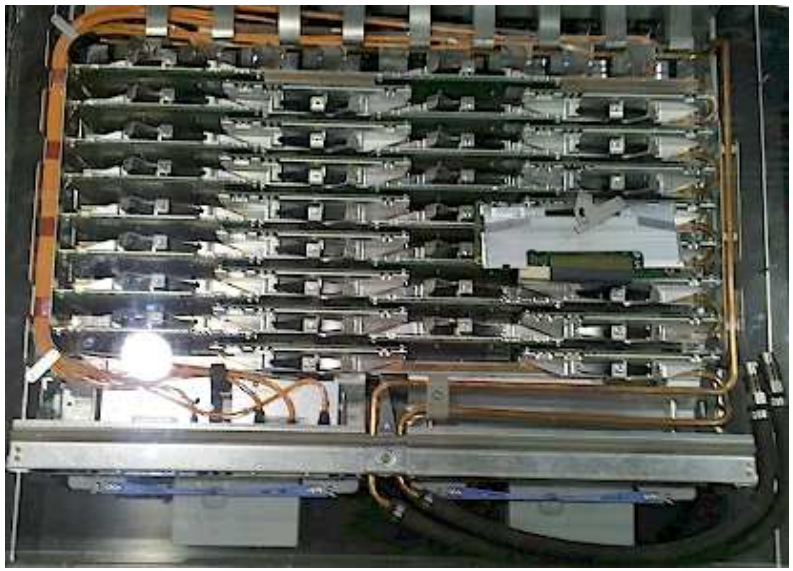
- 4-core ARM Cortex-A15 @ 2 GHz
  - 32 GFLOPS
- 8-core ARM Mali T685
  - 168 GFLOPS*
- Dual channel DDR3 memory controller
- **All in a low-power mobile socket**



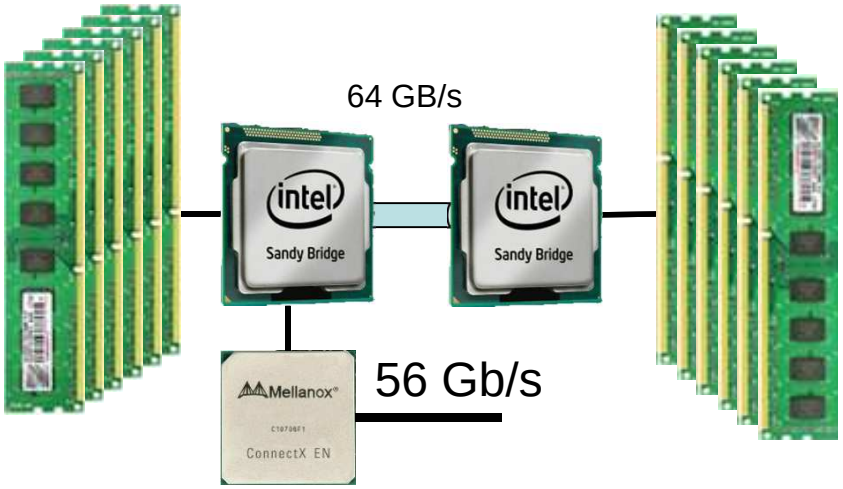
* Data from web sources, not an ARM or Samsung commitment

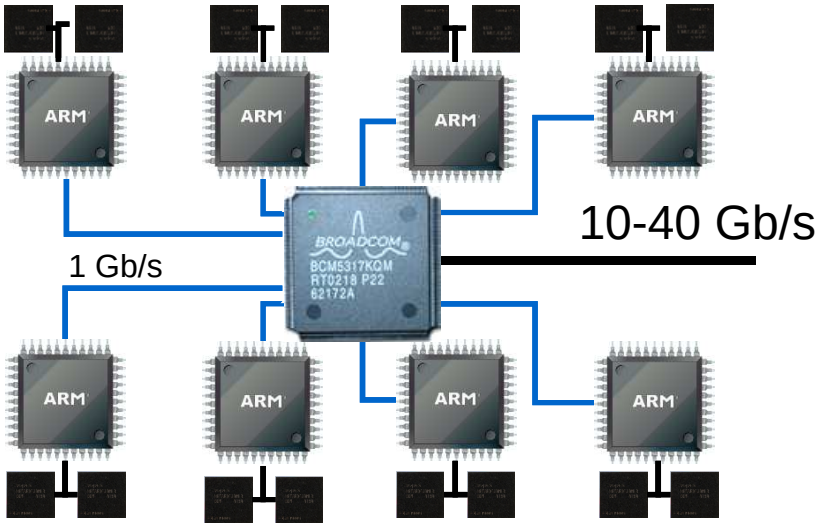
Are we building BlueGene again?

- Yes ...
  - Exploit Pollack's Rule in presence of abundant parallelism
    - Many small cores vs. Single fast core
- ... and No
  - Heterogeneous computing
    - On-chip GPU
  - Commodity vs. Special purpose
    - Higher volume
    - Many vendors
    - Lower cost
  - Lots of room for improvement
    - No SIMD / vectors yet ...
  - Build on Europe's embedded strengths



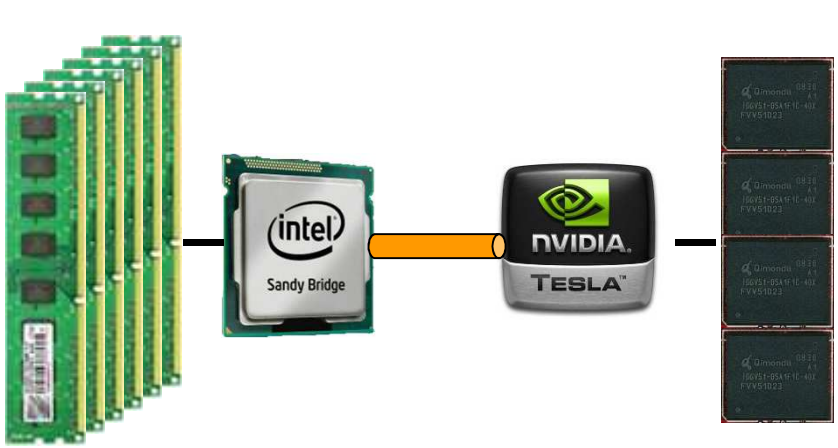
Can we achieve competitive performance?

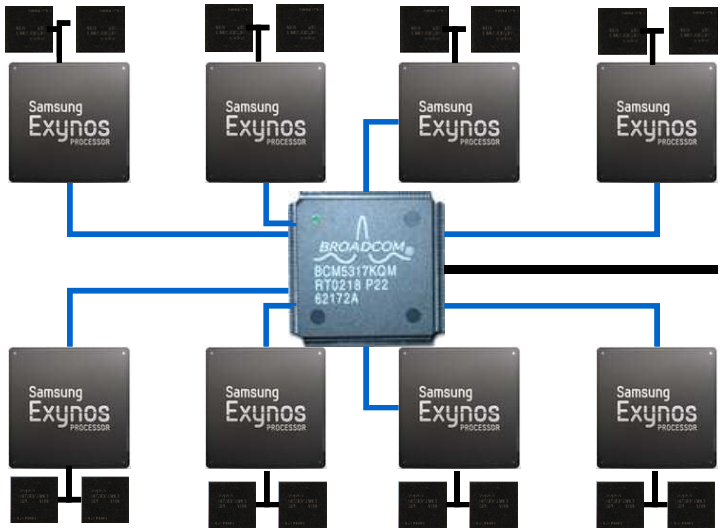




- 2-socket Intel Sandy Bridge
  - 370 GFLOPS
  - 1 address space
  - 44 MB on-chip memory
  - 136 GB/s
  - 64 GB/s intra-node (2 x QPI)
- 8-socket ARM Cortex A-15
  - 256 GFLOPS
  - 8 address spaces
  - 16 MB on-chip memory
  - 102 GB/s
  - 1 Gb/s intra-node (1 GbE)

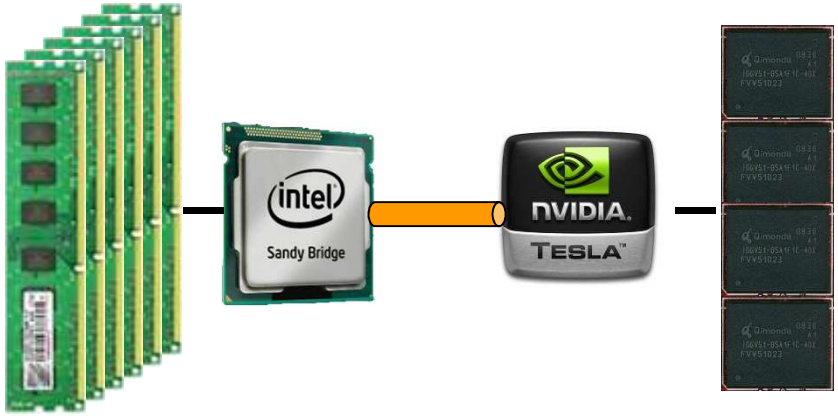
Can we achieve competitive performance?

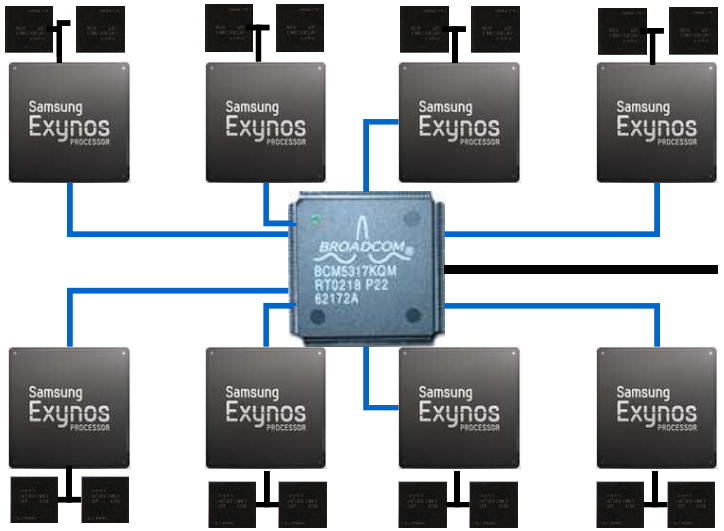




- Sandy Bridge + Nvidia K20
  - 1685 GFLOPS
  - 2 address spaces
  - 32 GB/s between CPU-GPU
    - 16x PCIe 3.0
  - 68 + 192 GB/s
- 8-socket Exynos 5450
  - 1600 GFLOPS
  - 16 address spaces
  - 12.8 GB/s between CPU-GPU
    - Shared memory
  - 102 GB/s

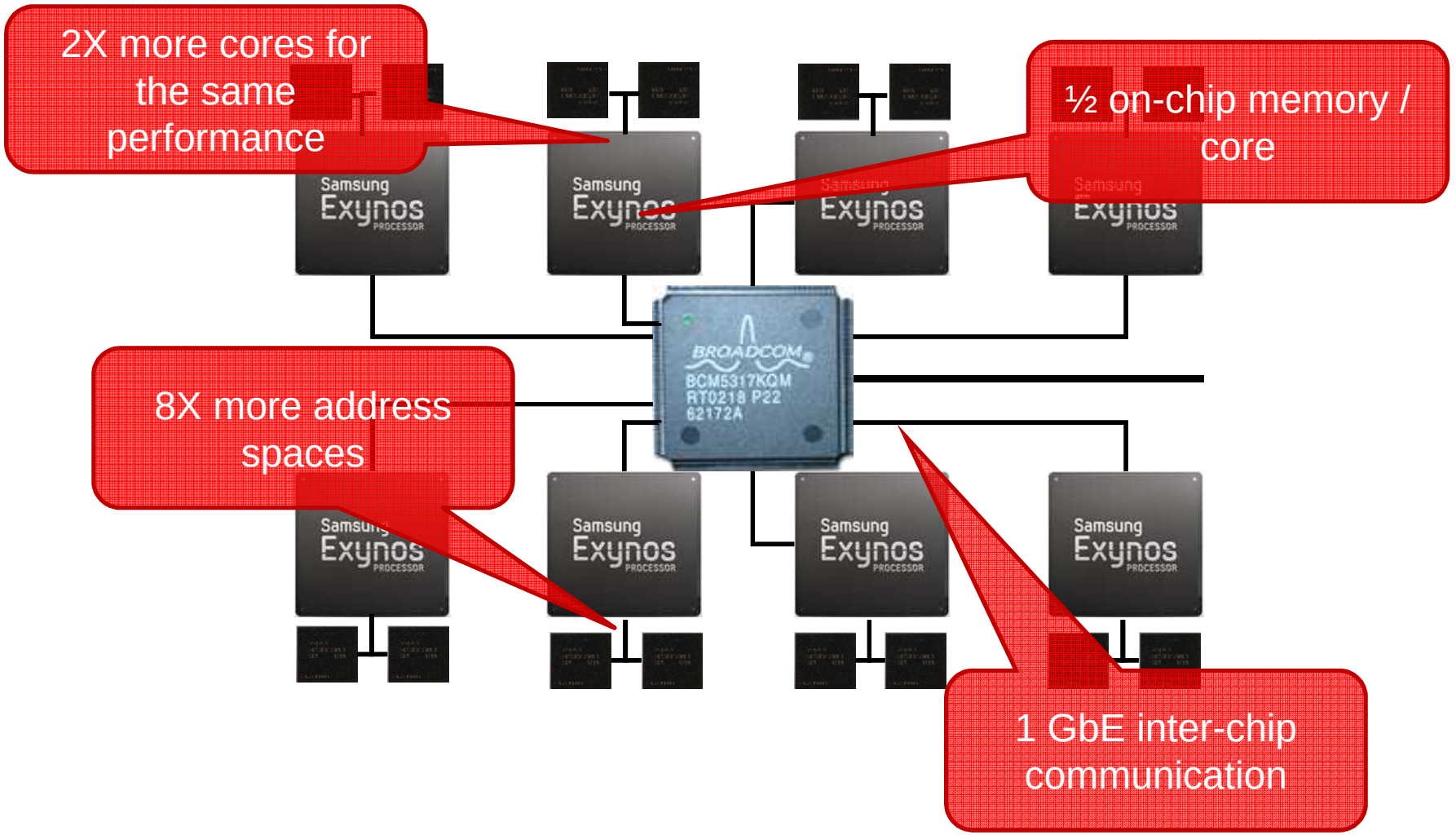
Then, what is so good about it?



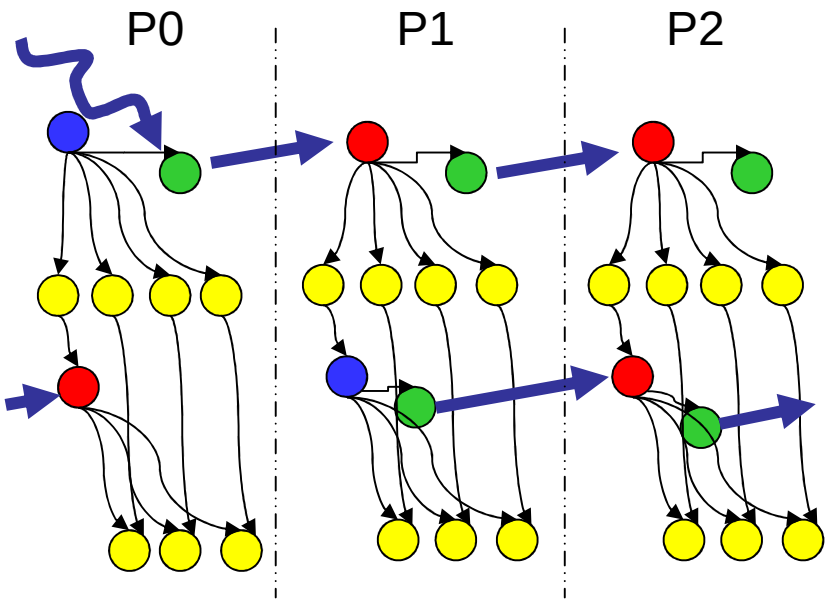


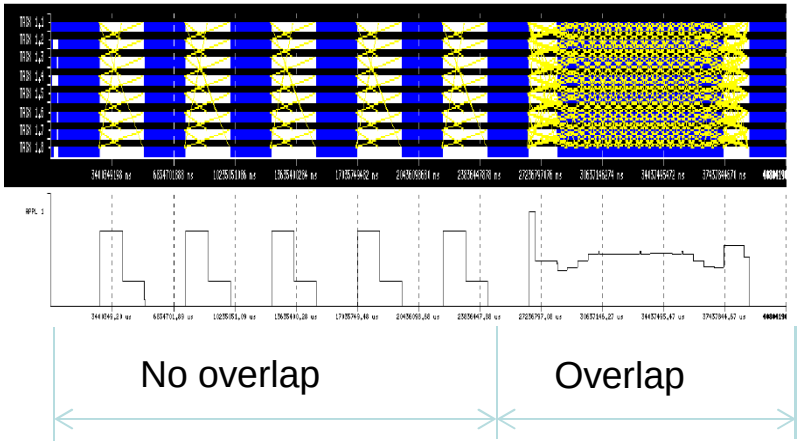
- Sandy Bridge + Nvidia K20
  - > $3000
  - > 400 Watt
- 8-socket Exynos 5450
  - < $200
  - < 100 Watt

There is no free lunch



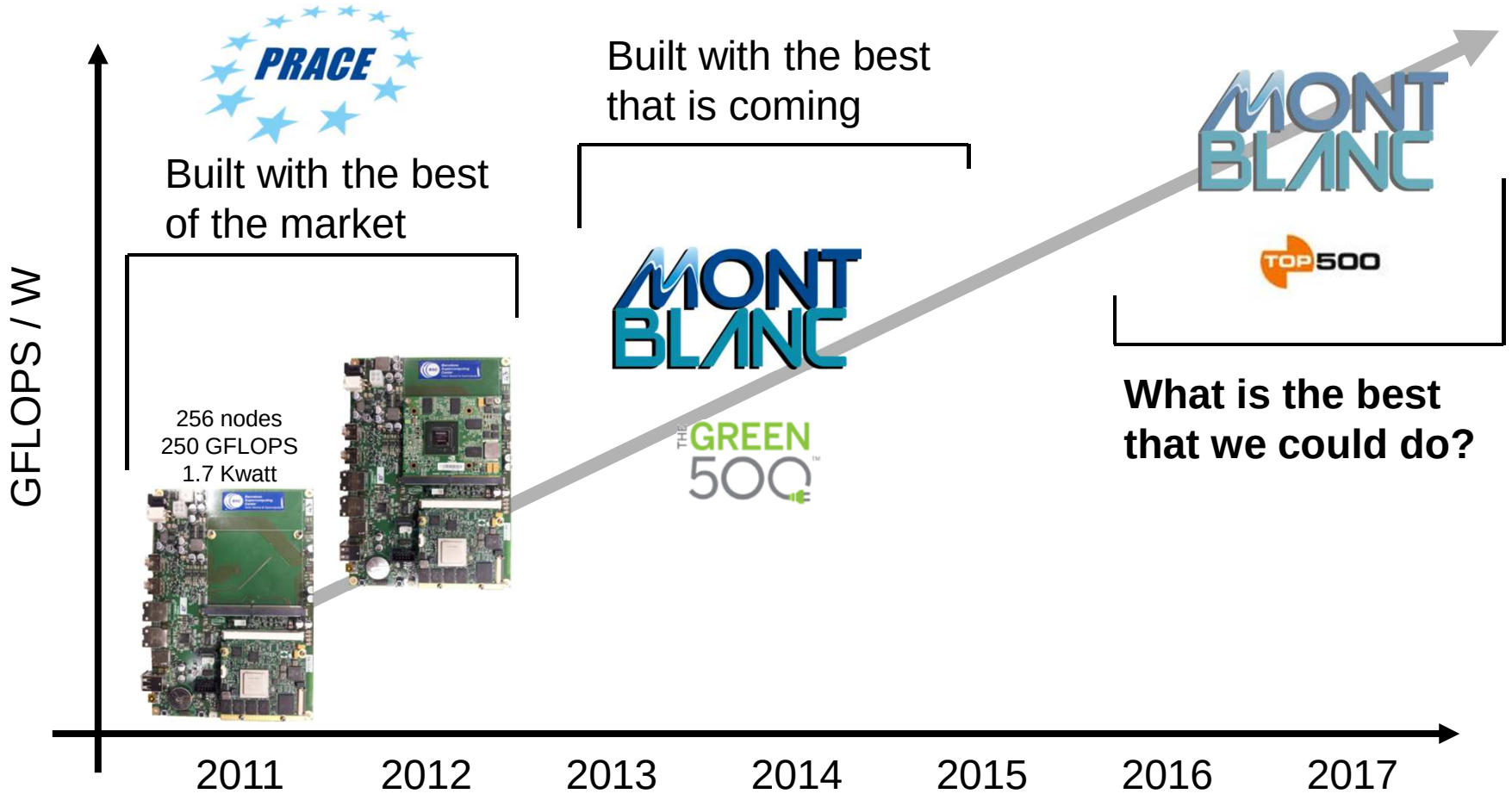
OmpSs runtime layer manages architecture complexity





- Programmer exposed a simple architecture
- Task graph provides lookahead
  - Exploit knowledge about the future
- Automatically handle all of the architecture challenges
  - Strong scalability
  - Multiple address spaces
  - Low cache size
  - Low interconnect bandwidth
- Enjoy the positive aspects
  - Energy efficiency
  - Low cost

A big challenge, and a huge opportunity for Europe

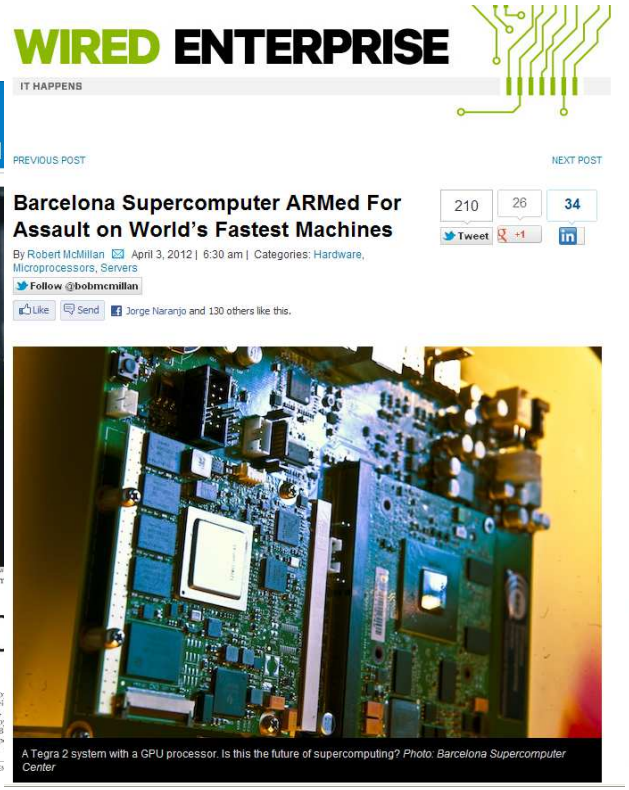


- Prototypes are critical to accelerate software development
  - System software stack + applications

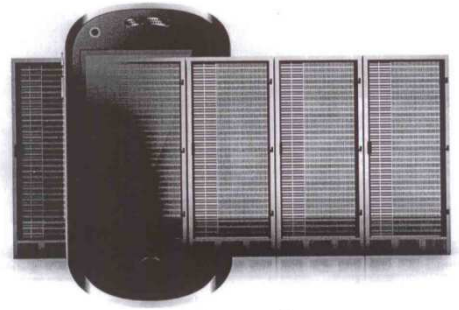
Very high expectations ...

- High media impact of ARM-based HPC
- Scientific, HPC, general press quote Mont-Blanc objectives
  - Highlighted by Eric Schmidt, Google Executive Chairman, at the EC's Innovation Convention

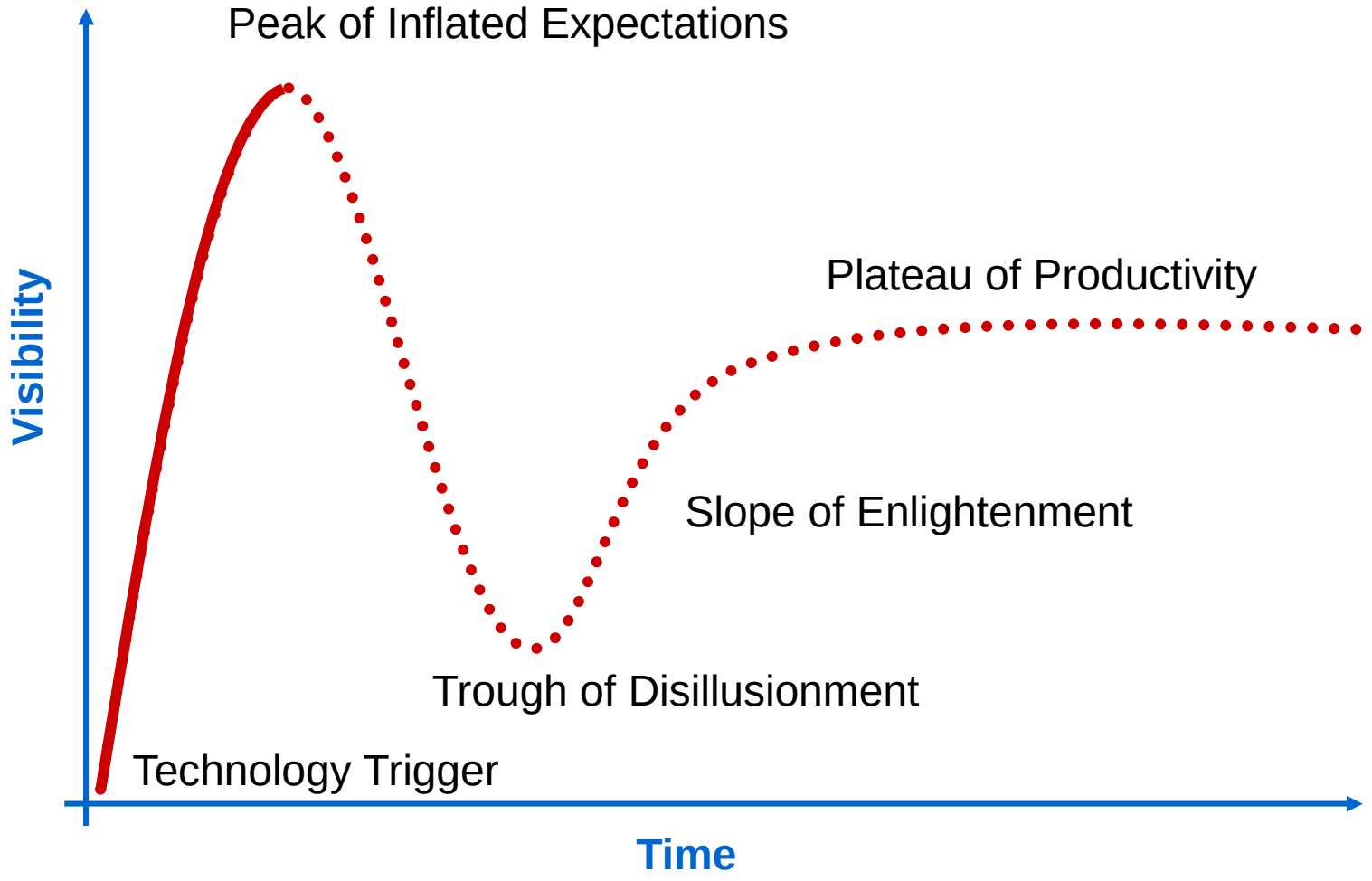
Científicos líderes. Àlex Ramírez, jefe de equipo del Barcelona Supercomputing Center, en la UPC. Detrás de él, la joya de la entidad, el Mare Nostrum



A Tegra 2 system with a GPU processor. Is this the future of supercomputing? *Photo: Barcelona Supercomputer Center*



The hype curve



- We'll see how deep it gets on the way down ...

Project goals

- To develop an **European** Exascale approach
- Based on embedded **power-efficient technology**

- Objetives
  - Develop a first prototype system, limited by available technology
  - Design a Next Generation system, to overcome the limitations
  - Develop a set of Exascale applications targeting the new system