Commodity embedded technology and mobile GPUs for future HPC systems

Filippo Mantovani
Barcelona Supercomputing Center
Technical Coordinator
Mont-Blanc project goals

- To develop an European Exascale approach
- Leverage commodity and embedded power-efficient technology

Supported by EU FP7 with 16M€ under two projects:

- Mont-Blanc: October 2011 – September 2014 + 6 months
  14.5 M€ budget (8.1 M€ EC contribution), 1095 Person-Month

  11.3 M€ budget (8.0 M€ EC contribution), 892 Person-Month
Mont-Blanc: Project objectives

• To deploy a prototype HPC system based on currently available energy-efficient embedded technology
  • Scalable to 50 PFLOPS on 7MWatt
  • Competitive with Green500 leaders in 2014
  • Deploy a full HPC system software stack

• To design a next-generation HPC system and new embedded technologies targeting HPC systems that would overcome most of the limitations encountered in the prototype system
  • Scalable to 200 PFLOPS on 10MWatt
  • Competitive with Top500 leaders in 2017

• To port and optimize a small number of representative Exascale applications capable of exploiting this new generation of HPC systems
  • Up to 11 full-scale applications
Mont-Blanc 2: Project objectives

• Continue support for the Mont-Blanc consortium
  • Mont-Blanc prototype(s) operation
  • OmpSs developer support
  • Increased dissemination effort (End-User Group)

• Complement the effort on the Mont-Blanc system software stack
  • Development tools: debugger, performance analysis
  • Resiliency
  • ARMv8 ISA

• Continue tracking and evaluation of ARM-based products
  • Deployment and evaluation of small developer kit clusters
  • Evaluation of their suitability for HPC

• Initial definition of future Mont-Blanc Exascale architectures
  • Performance & power models for design space exploration
Why are you doing this?

- RISC processors replaced vectors
- x86 (commodity) processors replaced RISC
  - Vector processors survive as (widening) SIMD extensions
Why are you doing this?

**Commodity**

- **Intel**
- **AMD**
- **IBM**

**HPC**

- **First teraFLOPS supercomputer**
  - ASCI Red (Sandia – 1997)
  - Pentium Pro

- **First petaFLOPS supercomputer**
  - Roadrunner (IBM / Los Alamos NL - 2008)
  - AMD Opteron + PowerXCell 8i

- **First >10 petaFLOPS supercomputer**
  - Titan (Cray / Oak Ridge NL - 2012)
  - AMD Opteron + Nvidia K20
What's commodity nowadays?

~22M cores (June '14)

<table>
<thead>
<tr>
<th>Year</th>
<th>Servers</th>
<th>PC</th>
<th>Smartphones</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>8.7M</td>
<td>350M</td>
<td>725M</td>
</tr>
<tr>
<td>2013</td>
<td>9.0M</td>
<td>+3%</td>
<td>315M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-9.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000M</td>
<td>+38%</td>
</tr>
</tbody>
</table>

...and we are still ignoring tablets: >200M

Source: International Data Corporation
In case of “invasion”, we want to be ready...
Prototypes are critical to accelerate software development
System software stack + applications
Mont-Blanc Server-on-Module (SoM)

CPU + GPU + DRAM + storage + network all in a compute card just 8.5 x 5.6 cm

Exynos5 Dual: 2x ARM Cortex-A15
ARM Mali-T604

USB 3.0 to 1GbE bridge

4 GB DDR3-1600
μSD slot up to 64 GB
The Mont-Blanc prototype

**Exynos 5 compute card**
- 2 x Cortex-A15 @ 1.7GHz
- 1 x Mali T604 GPU
- **6.8 + 25.5 GFLOPS**
- 15 Watts
- **2.1 GFLOPS/W**

**Carrier blade**
- 15 x Compute cards
- 485 GFLOPS
- **1 GbE to 10 GbE**
- 300 Watts
- 1.6 GFLOPS/W

**Blade chassis 7U**
- 9 x Carrier blade
- 135 x Compute cards
- 4.3 TFLOPS
- 2.7 kWatts
- 1.6 GFLOPS/W

**Rack**
- 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

<table>
<thead>
<tr>
<th>Mont-Blanc</th>
<th>Green500</th>
</tr>
</thead>
<tbody>
<tr>
<td>[GFLOPS/W]</td>
<td>[GFLOPS/W]</td>
</tr>
<tr>
<td>Nov 2011</td>
<td>0.15</td>
</tr>
<tr>
<td>Jun 2014</td>
<td>1.5</td>
</tr>
</tbody>
</table>

GPU ~ 2/3 peak
CPU ~ 1/3 peak
Everything perfect, then?

- 2x more cores for the same performance
- Lower I/O bandwidth
- Hybrid computation CPU+GPU
- 3/2 on chip memory/core
- 8x more cluster nodes
OmpSs runtime manages architecture complexity

- Programmer exposed to a simple architecture
  - Tasks
  - Data dependencies
  - Target (heterogeneity)

- Task graph provides lookahead
  - Exploit knowledge about the future

- Automatically handle many of the architecture challenges
  - Strong scalability
  - Multiple address spaces
  - Low cache size
  - Low interconnect bandwidth
Porting applications to Mont-Blanc

BQCD
Particle physics

BigDFT
Elect. Structure

COSMO
Weather forecast

EUTERPE
Fusion

MP2C
Multi-particle collisions

PEPC
Coulomb + Grav. Forces

ProFASI
Protein folding

Quantum ESPRESSO
Elect. Structure

SMMP
Protein folding

SPECFEM3D
Wave propagation

YALES2
Combustion

GPU capable
(CUDA or OpenCL)
OmpSs capable

+ 4 applications of Mont-Blanc2
End-User Group

- Develops a synergy among industry, research centers and partners of the project
- Validates the novel HPC technologies produced by the project
- Provides feedback to the project

Mont-Blanc provides EUG members with:
- Remote access to Mont-Blanc prototype platforms
- Support in platform evaluation and performance analysis
- Invitation to the Mont-Blanc training program
Micro-benchmarks results ( multicore + GPU):

Architecture summary:

2011
- 2 x ARM Cortex-A9 @ 1GHz
- 1 x 32-bit DDR2-333 channel
- 32KB L1 + 1MB L2

2012
- 4 x ARM Cortex-A9 @ 1.3GHz
- 1 x 32-bit DDR3-750 channel
- 32KB L1 + 1MB L2

2012
- 2 x ARM Cortex-A15 @ 1.7GHz
- 2 x 32-bit DDR3-800 channels
- 32KB L1 + 1MB L2
- GPU Mali T604 (4 shader)

2012
- 4 x Intel SandyBridge @ 2.4GHz
- 2 x 64-bit DDR3-800 channels
- 32KB L1 + 1MB L2 + 6MB L3

Under evaluation:
- Samsung Exynos 5 Octa big.LITTLE
  - IKS (in-kernel switcher)
  - MP (multi-processing)
  - NVIDIA Tegra K1

What do we learn from this?
What we learned so far:

- 32-bit memory controller
  - Even if ARM Cortex-A15 offers 40-bit address space
- No ECC protection in memory
  - Limited scalability, errors will appear beyond a certain number of nodes
- No standard server I/O interfaces
  - Do NOT provide native Ethernet or PCI Express
  - Provide USB 3.0 and SATA (required for tablets)
- No network protocol off-load engine
  - TCP/IP, OpenMX, USB protocol stacks run on the CPU
- Thermal package not designed for sustained full-power operation

All these are implementation decisions, not unsolvable problems. Only need a business case to justify the cost of including the new features (e.g. the HPC and server markets)
e.g. in the meantime things improve...
Conclusions:

- Need sustainable EFLOPS technology
  - $\min(\text{power + space + cost + ...})$

- Europe has a strong position in embedded computing
  - Energy efficiency
  - Commodity market

- BSC has a strong position in parallel programming models
  - OmpSs tasking model extends OpenMP 4.1

- Mont-Blanc offers a nice bag of Lego bricks to play with
  - Mont-Blanc prototype(s)
  - HPC software stack for embedded devices

- Leverage on all this to build a new class of sustainable computer faster, cheaper, more efficient

montblanc-project.eu
MontBlancEU
@MontBlanc_EU