High Performance Computing Based on Mobile Embedded Technology

Filippo Mantovani
Barcelona Supercomputing Center
Technical Coordinator
To develop an **European Exascale approach** leveraging **commodity** and **embedded** cost-effective technology.
Mont-Blanc objectives

- HPC prototype based on current mobile embedded technology
- Learn from the experience, plan for future architecture
- Port real scientific applications

Mont-Blanc 2

Extend
- Set of scientific applications
- OmpSs programming model
- Productivity tools
- Next generation Mont-Blanc architecture

Explore
- ARM 64-bit
- Fault tolerance and resiliency
- Market of ARM-based platforms for mini-clusters

Disseminate: End-User Group
Why are we doing this?

1 teraFLOPS supercomputer
ASCI Red
(Sandia – 1997)
Pentium Pro

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

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

~23M cores (Nov 2014)

<table>
<thead>
<tr>
<th>Year</th>
<th>Servers</th>
<th>PC</th>
<th>Smartphones</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>9.0M</td>
<td>316M</td>
<td>1000M</td>
</tr>
<tr>
<td>2014</td>
<td>9.3M</td>
<td>314M</td>
<td>1300M</td>
</tr>
</tbody>
</table>

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

---

Manchester, 01/07/2015

EMiT 2015
Outline

1. Hardware
   • History
   • Mont-Blanc prototype
   • Limitations

2. Software
   • System software
   • OmpSs programming model
   • Power monitor

3. Applications
   • Scalability
   • Energy to solution
   • End-User Group

4. Conclusions
   • Project status
   • Mont-Blanc impact
   • Remarks
1. **Hardware**
   - History
   - Mont-Blanc prototype
   - Limitations

2. **Software**
   - System software
   - OmpSs programming model
   - Power monitor

3. **Applications**
   - Scalability
   - Energy to solution
   - End-User Group

4. **Conclusions**
   - Project status
   - Mont-Blanc impact
   - Remarks
Prototypes are critical to accelerate software development
System software stack + applications
Mont-Blanc Server-on-Module (SoM)

CPU + GPU + Memory + Local Storage + Network
Form factor: 8.5 x 5.6 cm

Exynos5 Dual SoC
2x cores ARM Cortex-A15
1x GPU ARM Mali-T604

Memory
4 GB LPDDR3-1600

Network
USB3.0 to 1GbE bridge

Local Storage
microSD 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**
- 8 BullX chassis
- 72 Compute blades
- 1080 Compute cards
- 2160 CPUs
- 1080 GPUs
- 4.3 TB of DRAM
- 17.2 TB of Flash
- 35 TFLOPS
- 24 kWatt

<table>
<thead>
<tr>
<th></th>
<th>Mont-Blanc [GFLOPS/W]</th>
<th>Green500 [GFLOPS/W]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov 2011</td>
<td>0.15</td>
<td>2.0</td>
</tr>
<tr>
<td>Nov 2014</td>
<td>1.5</td>
<td>5.2</td>
</tr>
</tbody>
</table>
Limitations of commodity mobile technology

• 32-bit memory controller
  • Even if ARM Cortex-A15 offers 40-bit address space
• No ECC protection in memory
  • But surprisingly enough this is not affecting badly scalability (so far)
• No standard server I/O interfaces
  • No 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

Implementation decisions, not unsolvable problems

The only need is a business case to justify the cost of including the new features (e.g. the HPC and server markets)

Wait for next SoC producer | Design a new HPC SoC
Outline

1. Hardware
   • History
   • Mont-Blanc prototype
   • Limitations

2. Software
   • System software
   • OmpSs programming model
   • Power monitor

3. Applications
   • Scalability
   • Energy to solution
   • End-User Group

4. Conclusions
   • Project status
   • Mont-Blanc impact
   • Remarks
System software stack

- Open source system software
  - Ubuntu/Debian Linux OS
  - GNU compilers
  - JDK
  - Scientific libraries
  - Cluster management

- Runtime libraries
  - MPICH2, CUDA, OpenCL, OmpSs

- Productivity tools
  - Perf
  - Paraver, Scalasca
  - Allinea DDT debugger

+ Server software stack (LAMP, Hadoop, OpenStack)
OmpSs programming model

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

- Task graph provides look ahead
  - Exploit knowledge about the future
  - Allows exploration of scheduling policies

- It helps handling limitations of the hardware
  - Heterogeneity
  - Multiple address spaces
  - Low interconnect bandwidth
  - Synchronization
Power monitor – HW infrastructure
Power monitor – HW / SW interface

• Field Programmable Gate Array (FPGA)
  • Collects power consumption data from all 15 power measurement / sample interval: 70ms

• Board Management Controller (BMC)
  • Collects 1s averaged data from FPGA
  • Stores measurement samples in FIFO

• Mont-Blanc Pusher
  • Collects measurement data from multiple BMCs using custom IPMI commands
  • Forwards data using MQTT protocol through Collect Agent into key-value store
Power monitor – Block diagram

- Custom hardware monitoring
  - Mont-Blanc Pusher
- Standard hardware monitoring (out of band)
  - IPMI Pusher
- Standard hardware monitoring (in-band)
  - SysFS Pusher
- Standard building infrastructure monitoring
  - BACNet Pusher

MQTT Protocol

Collect Agent

Distributed Key-Value Store

Accessible like a single big data base!
Outline

1. Hardware
   • History
   • Mont-Blanc prototype
   • Limitations

2. Software
   • System software
   • OmpSs programming model
   • Power monitor

3. Applications
   • Scalability
   • Energy to solution
   • End-User Group

4. Conclusions
   • Project status
   • Mont-Blanc impact
   • Remarks
One application running: Alya RED

- Electro mechanics of the heart (of a rabbit)
- 10 steps measuring the parallel region of the code (MPI only)

Source: Constantino Gomez, Alejandro Rico
Understanding... at node level
Understanding... at blade level

Temperature (°C)
- temp (green)
- board_59 (red)
- pqii_59 (green)

Power (W)
- emb_59 (blue)

RPM x100
- fan_1a_59 (red)
- fan_1b_59 (green)
- fan_2a_59 (red)
- fan_2b_59 (green)
- fan_3a_59 (violet)
- fan_3b_59 (cyan)
- fan_4a_59 (gray)
- fan_4b_59 (brown)

Time (seconds)
Energy to solution

Himeno benchmark - L grid on mb-1105
Power Consumption (mW) vs. Exec Time (s)

QuantumESPRESSO/PWSCF - AUSURF 128 SDBs
Power Consumption (mW) vs. Exec Time (s) per 1 SDB
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
1. Hardware
   • History
   • Mont-Blanc prototype
   • Limitations

2. Software
   • System software
   • OmpSs programming model
   • Power monitor

3. Applications
   • Scalability
   • Energy to solution
   • End-User Group

4. Conclusions
   • Project status
   • Mont-Blanc impact
   • Remarks
Mont-Blanc project status

DONE!

- Prototype: design, development, deployment, monitor
- Deployment of HPC software stack on ARM
- Porting of HPC kernels and applications
- Test of non-HPC workload (Hadoop, OpenStack)

ON GOING…

- Next-generation architecture modelling
- ARM 64-bit exploration (mobile and server market)
- Porting of new applications
- Programming model enhancement
- Monitoring prototype for fault tolerant techniques
Mont-Blanc impact

Lower cost

Cheap computing
- Industrial applications
- Small labs
- Small datacenters

Mobile technology
- Cheap due to volumes
- Fast evolution
- Constrained by mobile requirements
- More aggressive

Automotive
- Sensor fusion
- Autonomous driving
- Hybrid fuel
- ...

Higher cost

New HPC node architecture
- Modular (IP)
- Requires industry involvement
- European architecture

Server technology
- Expensive
- Slower evolution
- More oriented to HPC (?)
- More conservative
Student Cluster Competition – ISC’15

- First team participating with ARM-based cluster
- First Spanish team
- 3 kW power budget
- 5+ applications
  - HPCC
  - LAMMPS
  - PyFR
  - Octopus
  - Some “secret” applications
- 3 awards
  - Highest HPL
  - Fan favorite
  - 1st, 2nd, 3rd overall places
Conclusions

• Mont-Blanc project explores the use of mobile embedded technology for scientific computing

• Hardware contributions
  • Several ARM-based prototypes have been developed
  • Companies and research institutions (EUG) can test MB prototypes

• Software contributions
  • System software + Power monitor
  • Programming model
  • Applications

• Even with “cheap” hardware it is possible to perform “decent” scientific computing

“The secret is to win going as slowly as possible.”

Niki Lauda
Mont-Blanc project

“The secret is to win going as slowly as possible.”

Niki Lauda