The Mont-Blanc Project: European Summit of Exascale System Architectures

In October 2011, the EU funded project Mont-Blanc started as one of three Exascale projects within the research theme "Information and Communication Technologies" of the 7th Framework Programme of the European Community.

The objectives of Mont-Blanc are the development of a fully functional energy-efficient HPC prototype based on the low-power commercially available embedded ARM technology, the design of a next-generation HPC system in order to overcome the limitations identified in the prototype system, and the development of a portfolio of Exascale applications to be run on this new generation of HPC systems.

The project is coordinated by the Barcelona Supercomputing Centre (BSC) and brings together a purely European consortium that joins industrial technology providers and research supercomputing centres: Bull, as a major HPC system vendor, ARM, as the world leader in embedded high-performance processors, and Gnodal, as interconnect partner that focuses its new product on scalability and power efficiency. Besides the technology providers, Mont-Blanc unites the supercomputing centres from the four Tier-0 hosting partners in PRACE who have leading roles in system software and Exascale application development: Germany (Forschungszentrum Jülich GmbH (FZJ), Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften (BADW-LRZ)), France (GENCI, CNRS), Italy (CINECA), and Spain

(BSC). The budget of the project is over 14 Million Euro, including over 8 Million Euro funded by the European Commission. With energy efficiency being a key issue, supercomputers are expected to achieve 200 Petaflop/s in 2017 with a power budget of 10 MW, and 1,000 Petaflop/s (1 Exaflop/s) in 2020 with a power budget of 20MW. That means an increase in energy-efficiency of more than 20 times compared to the most efficient supercomputers today is necessary. It must be taken into account that not all energy is used for computing within the cores. In current systems the processors consume the lion's share of the energy, often 40% or more. The remaining energy is used to power up the memory, interconnection network, and storage system. Furthermore, a significant fraction is wasted in power supply overheads, and in thermal dissipation (cooling), which do not contribute to performance at all.

During the Mont-Blanc project, a new prototype will be developed. By using ARM processors it is expected to achieve from 4 to 10 times increase in energy-efficiency compared with current technologies. Integrating the low-power processors by providing very high-capacity interconnects is the role of Gnodal, the British company currently offering the lowest latency chip in the world.

Figure 1-a shows a first board (in the front) providing the Nvidia Tegra2 processor (dual-core ARM Cortex A9 @ 1 GHz), 1 GB of DDR2 memory, 16 GB of flash

storage, and PCIe connectivity. This card will soon be upgraded to Tegra3 (quad-core ARM Cortex-A9 @ 1.5 GHz).

Figure 1-b shows a MXM mobile GPU accelerator, a CUDA capable GPGPU with double-precision floating-point support. A wide variety of GPU options are available. The carrier board connects the Tegra2 to a native 100 Mbit Ethernet port and a 1 Gbit Ethernet port through PCIe-x1, and the GPU card from Figure 1-b through PCIe-x4. These cards and the carrier board originate from a PRACE project that has started right now. For the Mont-Blanc project, this prototype serves as a starting platform to port scientific applications to the ARM architecture and to utilize GPGPU accelerators. In addition to being of larger scale than the existing prototype system, the final Mont-Blanc prototype will feature professional and tighter system integration, which is necessary to achieve the energy efficiency goals of the project.

The supercomputing centres run thousands of real applications on a daily basis on their Tier-0 and Tier-1 systems, coming from a vast number of scientific domains and serving a large community of academic and industrial users. In order to assess the different hardware and software components made available during the project, an incremental approach will be used, working on both the porting and the optimisation of small kernels, and then on real end-users' scientific applications.

The interests of the end users must be taken into account in devising the architectures of Exascale systems. FZJ contributes by analyzing applications from different areas of research: the parallel Coulomb solver PEPC, the massively

parallel multi-particle collision software MP2C, the software for molecular mechanics of proteins SMMP, and the protein folding and aggregation simulator ProFASi. Furthermore, JSC is involved in the field of performance analysis. Scalasca, the world leader in scalable portable analysis tools will be provided for the Mont-Blanc platform.
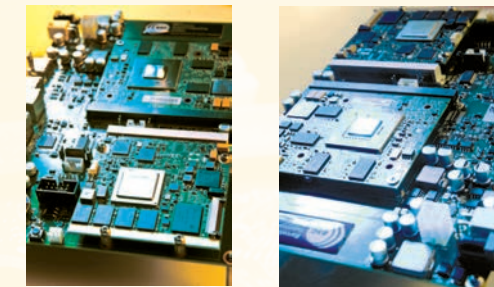
Figure 1-a (left): Nvidia Tegra2 processor (dual-core ARM Cortex A9 @ 1 GHz)

Figure 1-b (right): MXM mobile GPU accelerator

BADW-LRZ contributes BQCD, a highly scalable simulation from the area of quantum chromodynamics. Furthermore, with its experience in the area of monitoring and optimization of energy consumption, BADW-LRZ provides a significant contribution to accomplish the objectives of the prototype system concerning energy efficiency. For this, BADW-LRZ is developing energy-aware system management software and is working closely with Bull to ensure an efficient system integration of the prototype.

At Exascale, it might be necessary to broaden the range of programming models. OmpSs is an effort made by BSC to extend the shared memory parallel programming model OpenMP with support for asynchronous parallelism and for heterogeneous devices, which will be exploited in the Mont-Blanc project. Up to now, for some of the selected applications an MPI version has already been ported to the ARM based prototype and current work focuses on the identification of tasks to develop an OmpSs version in order to use the high level of parallelism in the system most efficiently.

• Thomas Fieseler[1]
• Willi Homberg[1]
• Axel Auweter[2]

[1] Forschungszentrum Jülich

[2] BADW-LRZ