Large scale low power architectures computing system: status of ExaNeSt and EuroExa projects

P. Vicini[1] for the INFN ExaNeSt team

[1] INFN, Sezione di Roma, Italy

E-mail: piero.vicini@roma1.infn.it

Abstract. The next generation HPC systems will be characterized by ExaFlops performances (i.e. 10^{18} Floating Point Operations per second), and will be used for computing broad class of large size, computing-demanding scientific and industrial applications, ranging from modeling and simulation of complex physical systems to biotechnology, cloud computing, big data and analytics.

At very high level of abstraction, an ExaFlops scale HPC system will be composed of millions low power consumption computing cores, tightly interconnected by a low latency high performance network, equipped with a distributed storage architecture, densely packaged but cooled by an appropriate technology. A simple system power estimation, based on current state-of-the-art technologies and 10^6 CPUs, returns a total power consumption of $\sim 100MW$. In the next future one can expect (at best) an improvement of a factor 3 for the ratio power/flops and a factor 5 for the ratio flops/volume.
So, while in the past the main characterizing parameters of supercomputing platforms were computational peak performances and procurement costs, today power consumption and cooling effective mechanisms have become key factors driving the process of HPC system design.

The ExaNeSt project, started on December 2015 and funded in EU H2020 research framework (call H2020-FETHPC-2014, n. 671553), is a European initiative combining industrial and academic research expertise and aiming to evaluate technologies, to design architecture and to deploy fully functional demonstrators of a novel system-level interconnect, distributed NVM (Non-Volatile Memory) storage and advanced cooling infrastructure for an ARM-based ExaFlops-class supercomputers.

The project leverages on last generation high end SoC FPGA (System on Chip FPGA) - the Xilinx Zynq UltraScale+ with 4 ARM Cortex-A53 embedded cores running up at $1.5GHz$ - to integrate 1000+ ARM cores system prototype interconnected by *ExaNet*, a novel, unified (for data and storage traffic), low latency, high throughput, RDMA-based interconnect architecture. ExaNet is designed and validated, using passive copper and/or active optical cable technology, on a network testbed built on the same SoC FPGA of final ExaNeSt prototype. ExaNet integration will allow us to explore different interconnection topologies, to evaluate advanced routing functions for congestion control and fault tolerance and to design specific hardware components for acceleration of collective operations. Tests on first ExaNet release are really encouraging showing a sub-μS latency for a single hop, node-to-node roundtrip packet transfer and bandwidth performances for small-medium size packets very close to the peak limit.

Starting for September 2017 a new H2020 EU initiative, named EuroExa and funded under FETHPC-2016 program, will build on ExaNeSt results to deliver a world-class, ARM-based HPC platform prototype. EuroExa will be co-designed and benchmarked through a large set of key scientific and engineering applications and will be used as medium-scale demonstrator of a novel European extreme scale computing system.

In this talk, after a brief introduction of the motivations, goals and plan of activities of ExaNeST and EuroExa projects, I will report on the status of system R&D in particular on network architecture design, test and preliminary bandwidth and latency measurements.