

Recent openlab benchmark results on the Dell R820 4-socket Sandy Bridge-EP system

Liviu Vâlsan, CERN openlab » 20th of November 2012



CERN openlab Minor Review » November 2012

What has been benchmarked and compared?

- » 32-core, quad socket Sandy Bridge-EP server
 - Dell PowerEdge R820
 - Intel Xeon E5-4650

- » 16-core, dual socket Sandy Bridge-EP server
 - Intel S2600CP
 - Intel Xeon E5-2680

- » 40-core, quad socket Westmere-EX server
 - Intel / QCI QSSC-S4R
 - Intel Xeon E7-4870

	Quad socket SNB-EP	Dual socket SNB-EP	Quad socket WSM-EX
Processor	Intel Xeon E5-4650	Intel Xeon E5-2680	Intel Xeon E7-4870
Nominal processor speed	2.70 GHz	2.70 GHz	2.40 GHz
Turbo processor speed	3.30 GHz	3.46 GHz	2.80 GHz
Number of cores / socket	8	8	10
Total number of cores	32	16	40
L3 cache	20 MB	20 MB	30 MB
Processor silicon process	32 nm	32 nm	32 nm
Motherboard	Intel C600 series	Intel S2600CP	QCI QSSC-S4R
Total memory installed	256 GB	64 GB	128 GB
Memory configuration	16 x 16 GB DDR3L 1333 MHz 1.35V	16 x 4 GB DDR3L 1333 MHz 1.35V	32 x 4 GB DDR3 1066 MHz 1.5V

- » **OS:** Scientific Linux CERN 6.3
- » **Kernel:** 2.6.32-279.5.2.el6.x86_64
- » **GCC:** 4.4.6 (standard SLC 6 version)
- » **ICC:** 13.0.1 (Composer XE update 1)

- » A ZES-Zimmer LMG450 power analyser has been used (accuracy of 0.1%)
- » Load stressing tools:
 - **The Intel Optimized LINPACK Benchmark** is a generalization of the LINPACK 1000 benchmark. It is used to load both the memory system and the CPU. The memory consumption depends on the size of the generated matrices and is easy to adapt to fit the needs.
 - **CPU Burn-in** was originally written as a tool for overclockers, so that they can stress the overclocked CPUs, and check if they are stable. It runs Floating Point Unit intensive operations to get the CPUs under full load, allowing the highest power consumption to be measured from the CPU.

» Two different running conditions:

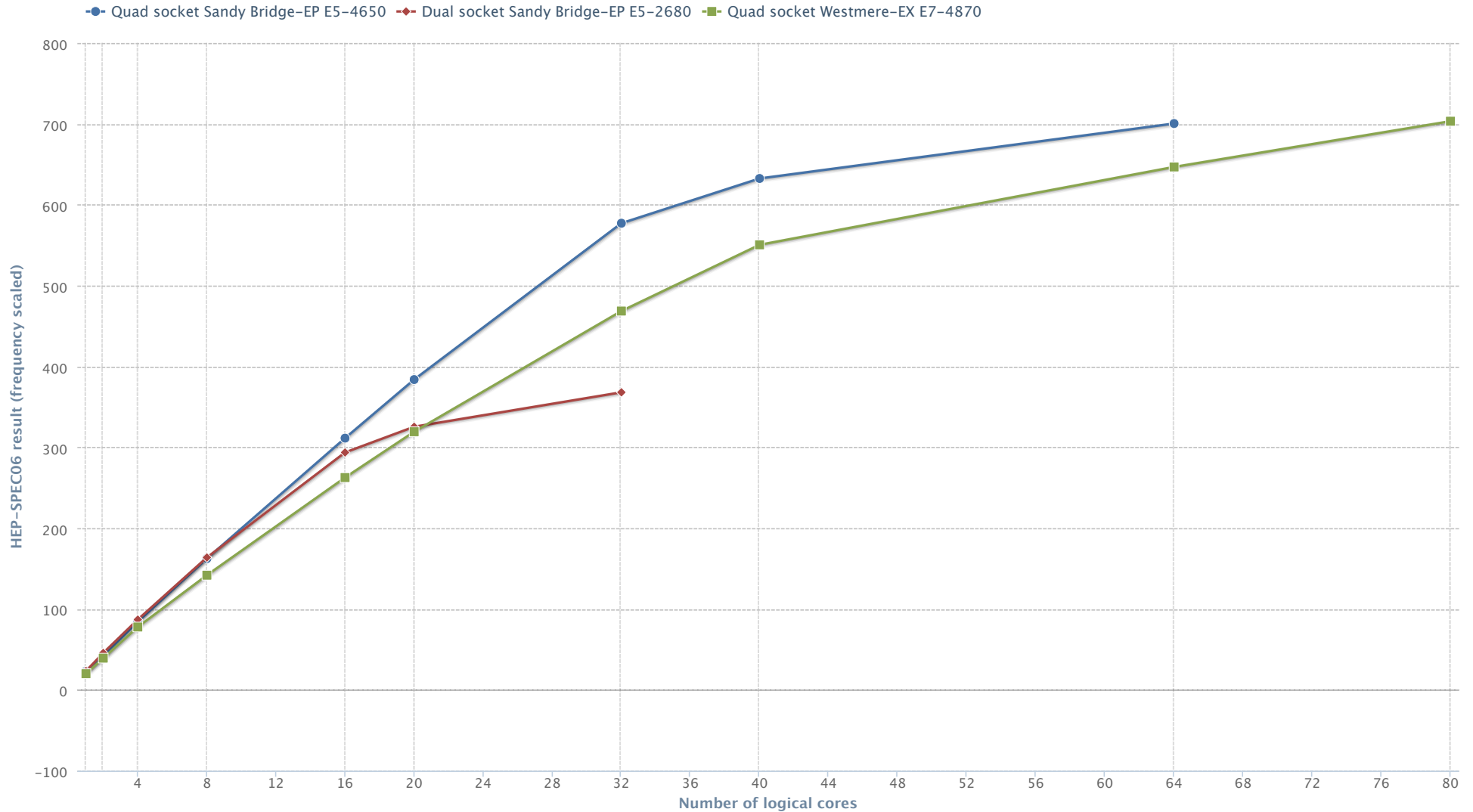
- **Idle:** the system is booted within the Operating System and is left idle.
- **Load:** the system is running CPU Burn-in on half of the cores, and LINPACK on all the other cores, using all the installed memory.
 - With SMT disabled the system appears as having 32 cores in total, so the load stress test consists of running 16 instances of CPU Burn-in along with 16 instances of LINPACK (using 16 GB of memory each).
 - With SMT enabled the system was considered as a 64 core server, meaning that the load stress test was conducted by running 32 instances of CPU Burn-in along with 32 instances of LINPACK (using 8 GB of memory each)

	Idle	Load	Standard measurement *
SMT-off	149 Watts	534 Watts	457 Watts
SMT-on	150 Watts	577 Watts	492 Watts

* The standard energy measurement is a mix of the active power under the Idle condition, accounting for 20%, and the active power under Load condition, accounting for 80%.

HEP-SPEC06 results and comparisons

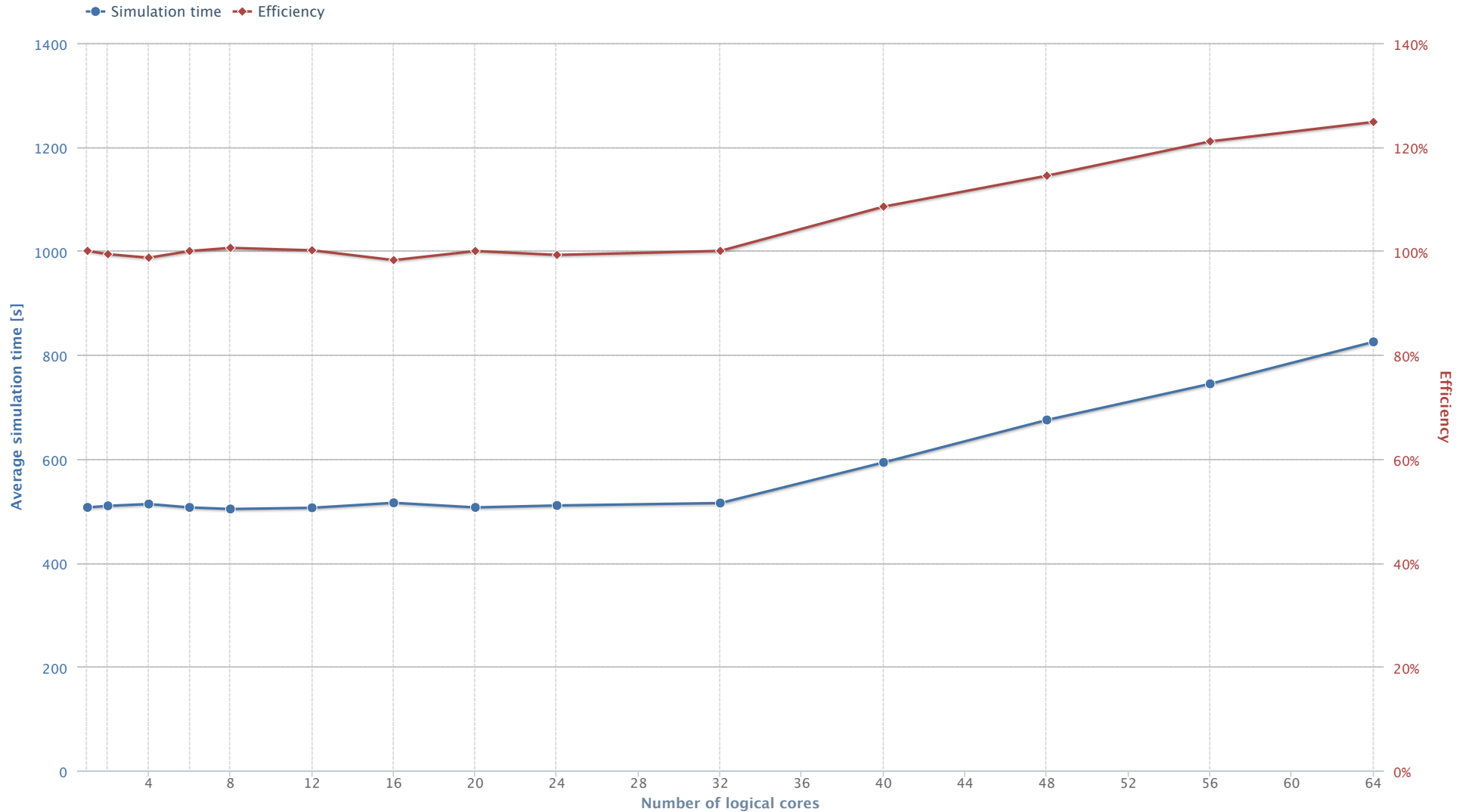
HEP-SPEC06 performance comparison Turbo Boost Technology disabled (higher is better)



- » gcc 4.4.6 has been used to compile the benchmark on all systems (in 64bit mode).
- » When considering the maximum number of physical cores shared by each pair, the performance per core offered by the Sandy Bridge-EP E5-4650 processor is:
 - 6% more than that offered by the Sandy Bridge-EP E5-2680
 - 23% more than that of a Westmere-EX E7-4870 core
- » When considering the performance per core at the maximum CPU occupancy:
 - equivalent to that of the dual socket Sandy Bridge-EP E5-2680
 - 31% more than that of a Westmere-EX E7-4870 core

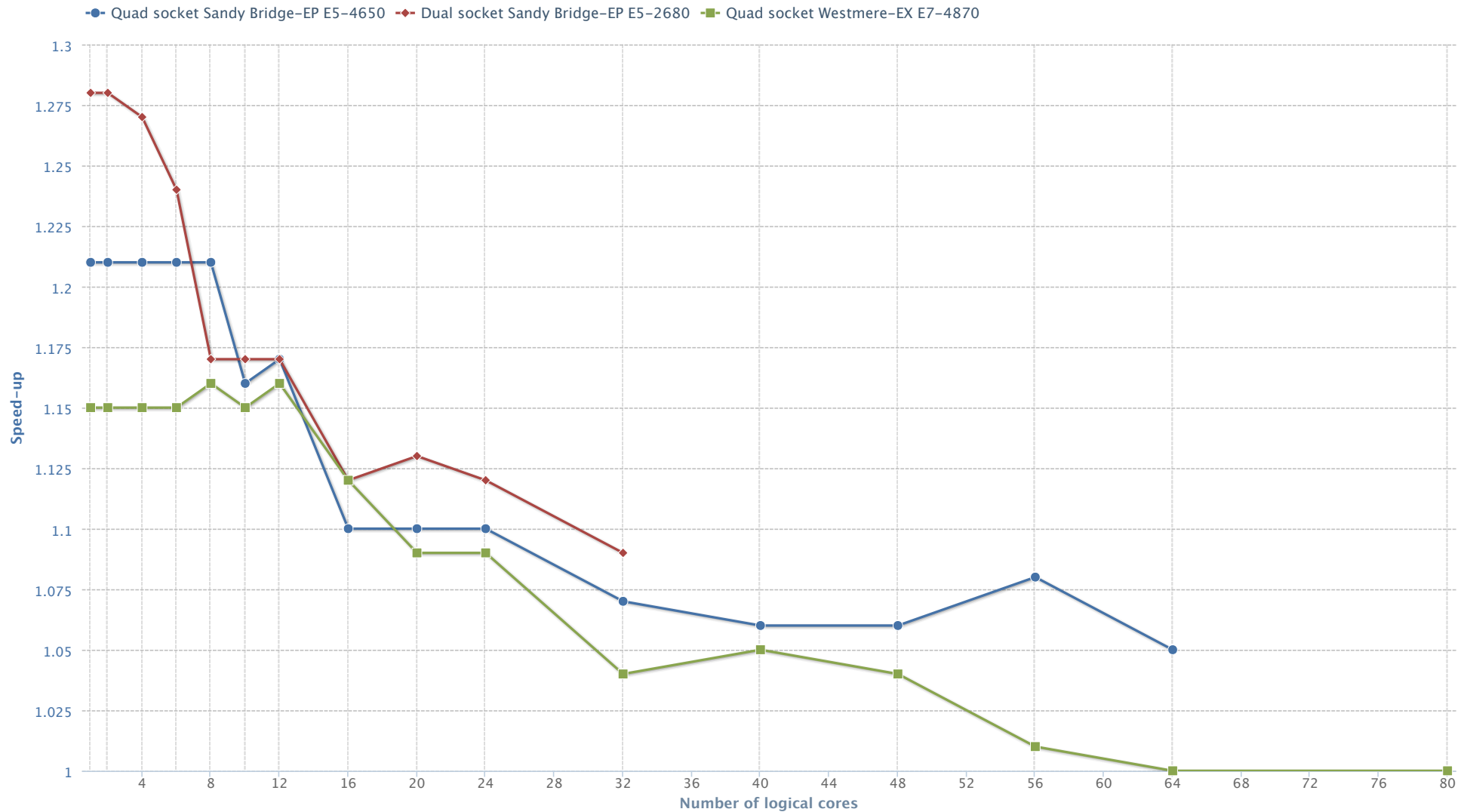
Multi-threaded Geant4 prototype results

Multi-threaded Geant 4 prototype (generation 6) scalability on quad socket Sandy Bridge-EP
 ParFullCMSmt: average simulation time for 100 pi- events per thread

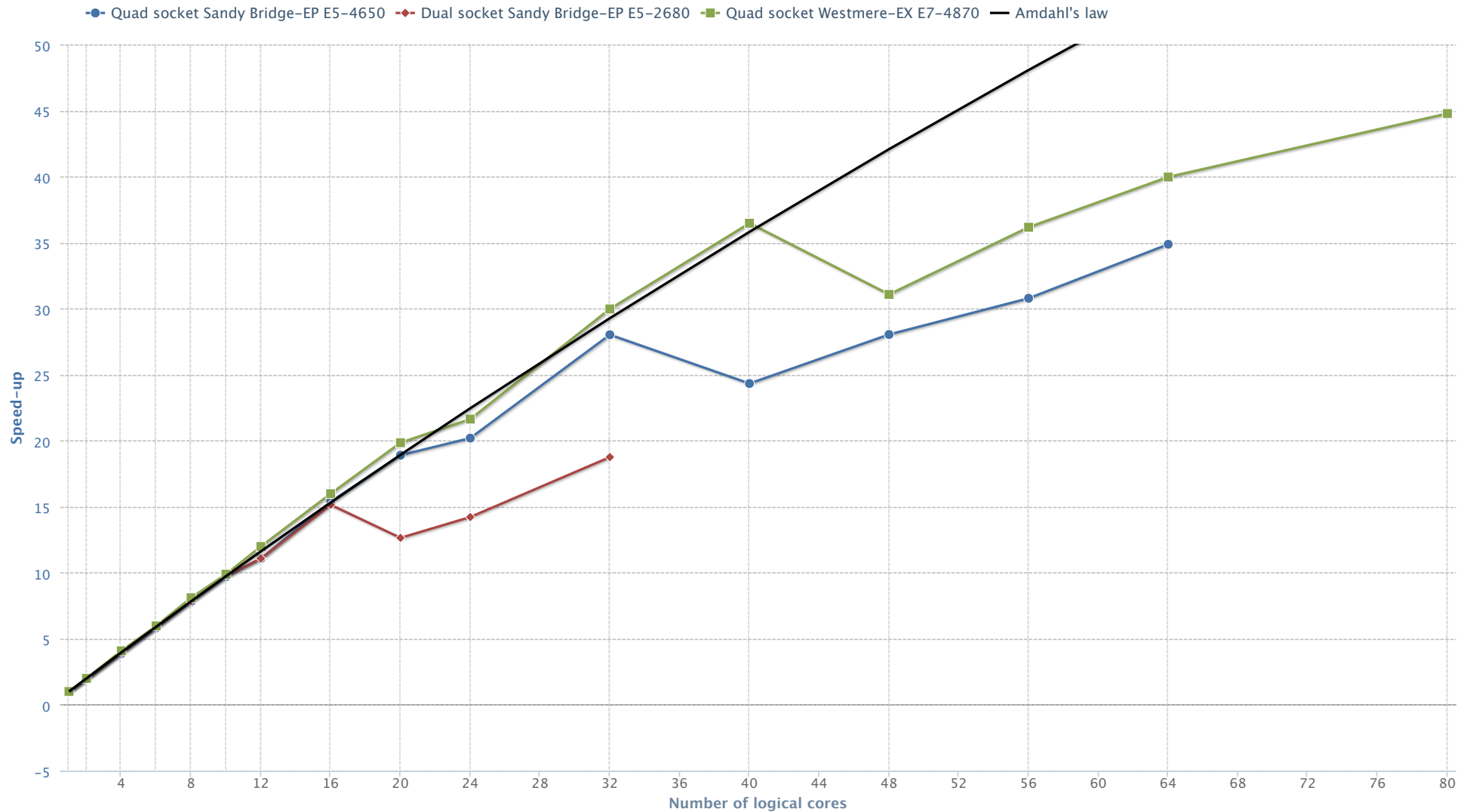


MLFit Turbo Boost Technology speed-up

Speed-up ratios obtained from the comparison of the execution times with Turbo Boost Technology disabled and Turbo Boost Technology enabled

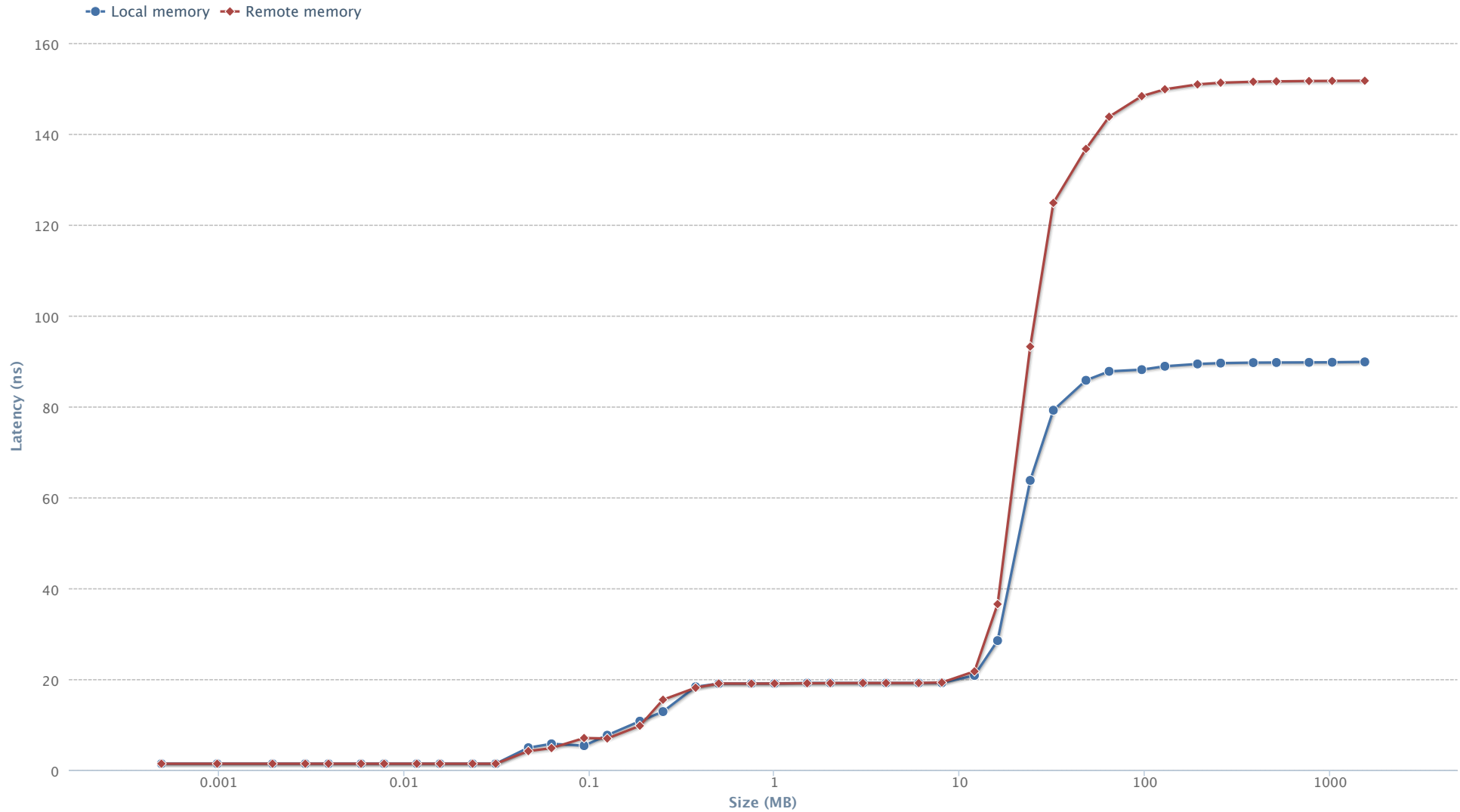


Scalability results.
 The solid black line is the theoretical speed-up obtained by Amdahl's law with a parallel fraction of 99.7%.



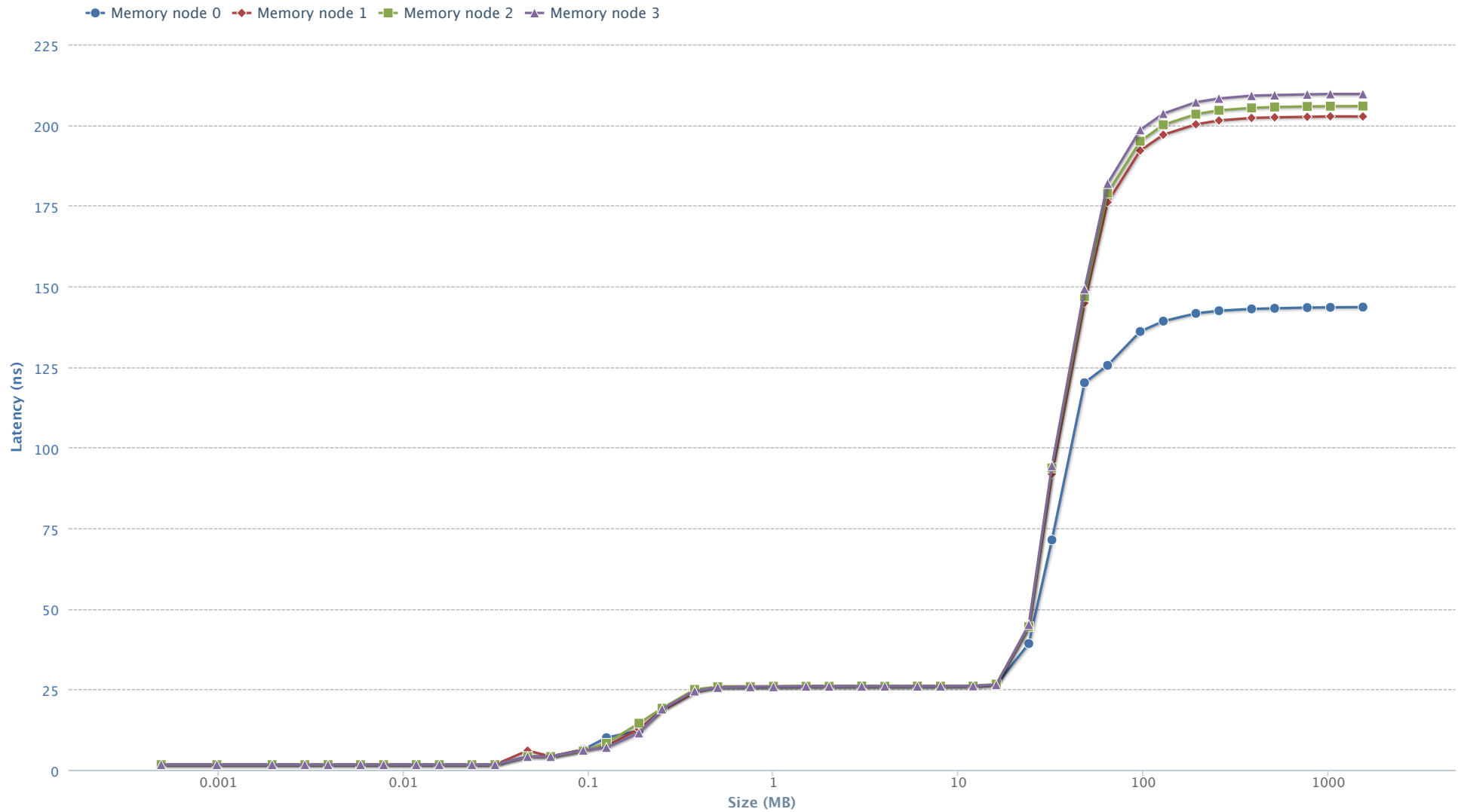
Memory Latency Measurements

Sandy Bridge-EP E5-2690 memory latency



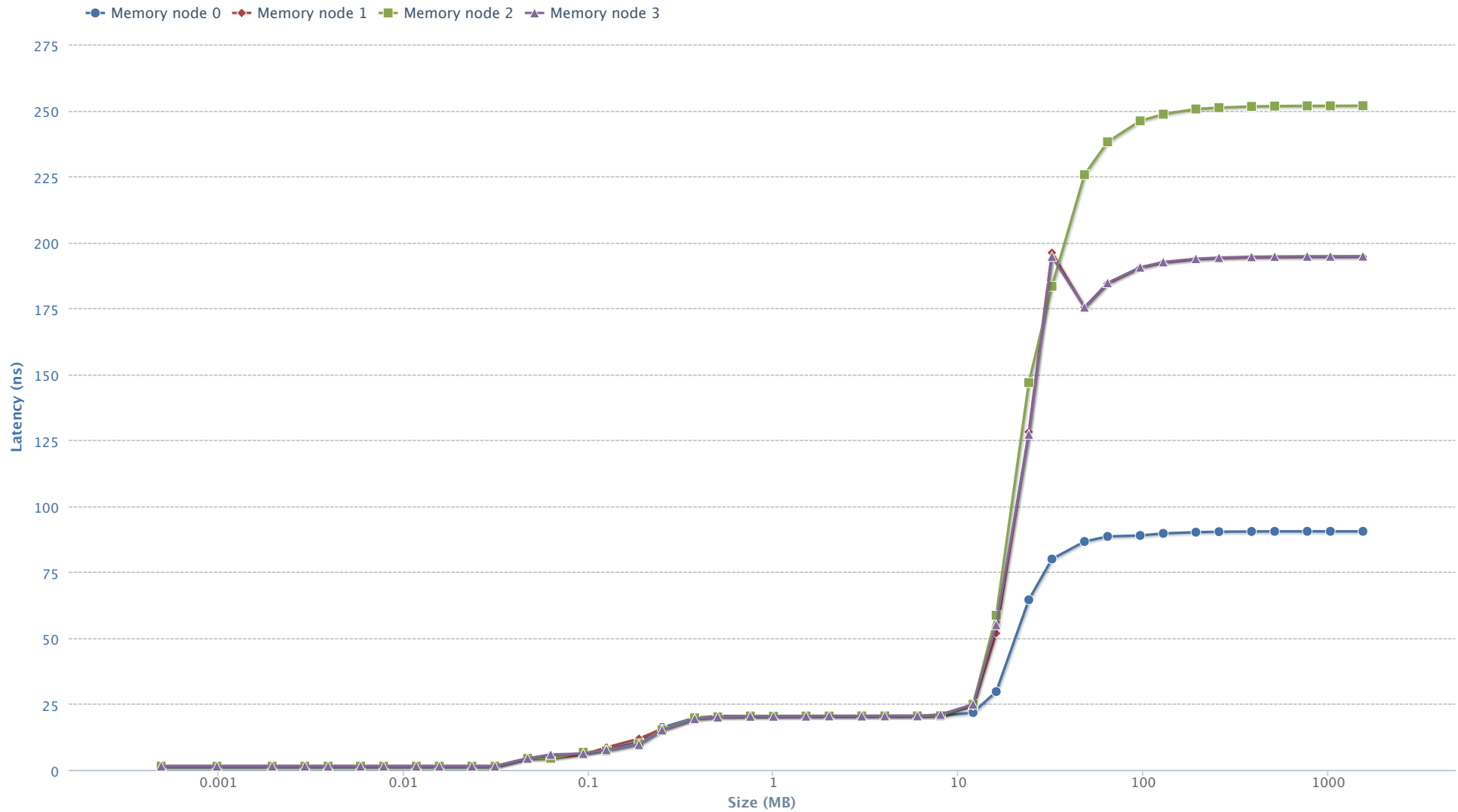
Memory Latency Measurements

Westmere-EX E7-4870 memory latency



Memory Latency Measurements

Sandy Bridge-EP E5-4650 memory latency



- » The overall performance of the quad socket Sandy Bridge-EP server matches the performance of the quad socket Westmere-EX server. This is particularly interesting given the differences between the two:
 - The Westmere-EX comes with 25% more cores
 - The Westmere-EX has 20% larger L3 cache
- » We have obtained between 23% and 48% better performance per core for the quad socket Sandy Bridge-EP when compared to the Westmere-EX
- » The quad socket Sandy Bridge-EP showed very good scalability figures, in line with those observed for the dual socket version
- » Energy efficiency is particularly impressive for the quad socket Dell R820
- » In case of non-uniform memory access it's important to pay attention to architectural system differences.

Q & A



CERN
openlab

Other questions?

liviu.valsan@cern.ch