Evaluating the Scalability of HEP Software and Multi-core Hardware

S. Jarp, Alfio Lazzaro, J. Leduc, A. Nowak
CERN openlab

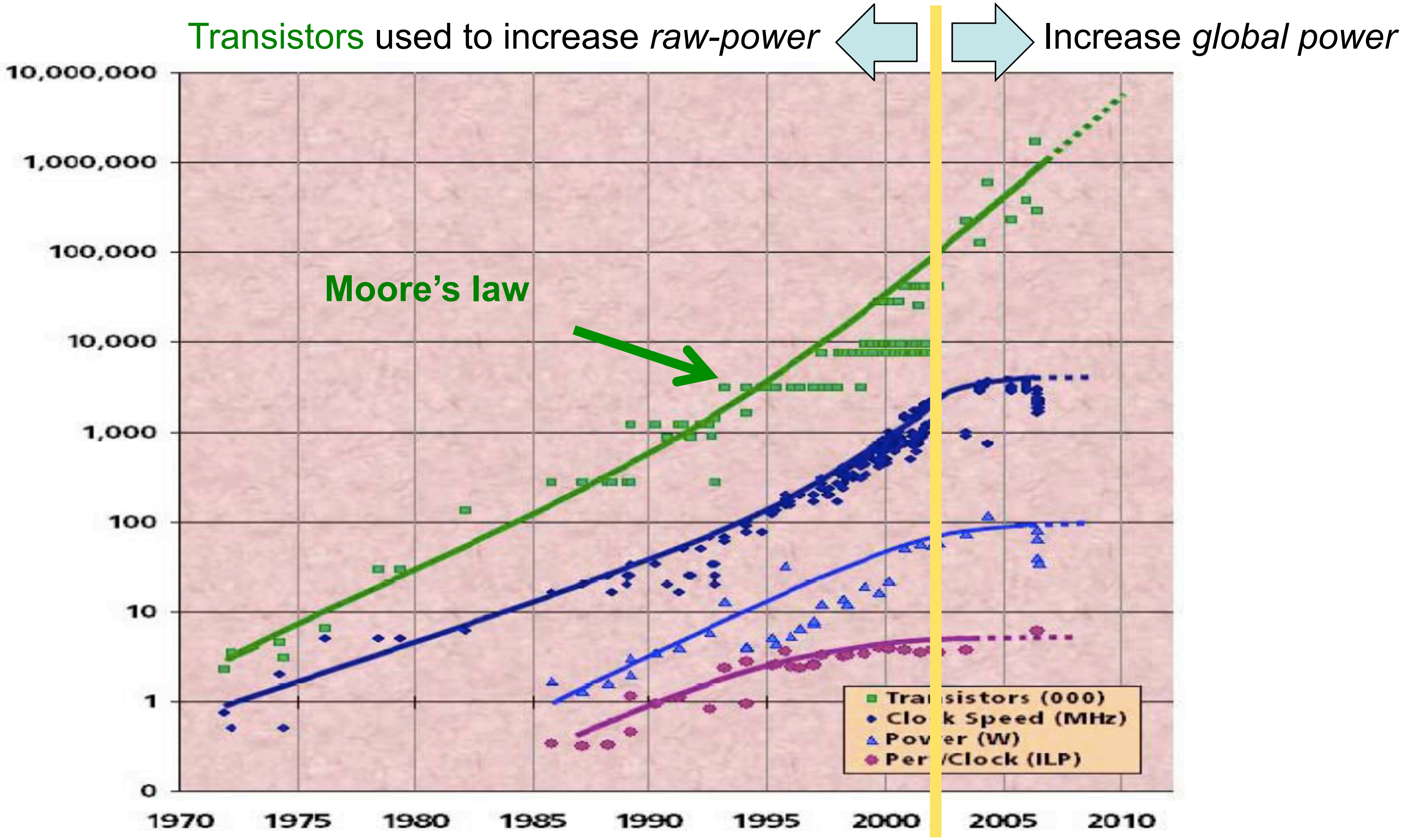
International Conference on Computing in High Energy and Nuclear Physics 2010 (CHEP2010)

October 18th, 2010

Academia Sinica, Taipei

Presentation on behalf of A. Nowak

Moore’s law



Transistors used to increase *raw-power*

Increase *global power*

Moore’s law

Consequence of the Moore's Law

- Hardware continues to follow Moore's law
  - More and more transistors available for computation
    - More (and more complex) execution units: hundreds of new instructions
    - Longer SIMD (Single Instruction Multiple Data) vectors
    - More hardware threading
    - More and more cores

Current Status in HEP

- Currently available nodes with up to 8 cores (4-cores dual-socket)
  - Soon this number will increase up to 48 cores
- Poor usage of multi-threading software
  - A machine with *N* cores is considered as *N* independent slots for *N* independent applications
  - No shared memory among the applications on the node
    - Memory usage increases linearly with *N*!
- Poor usage of hardware multi-threading (SMT), usually switched off by default
  - Current CPU can handle 2 hw-threads per core
  - For sequential applications the benefit of the SMT (10% - 30%) is small if compared to memory requirement (100% more memory required), but it is compute power for free in case of parallel applications!

In this presentation

- It is vital for HEP programmers to understand the scalability of their software on modern hardware and the opportunities for potential improvements
  - Move to multi-threaded version of the code
  - Reduce memory footprint using shared memory concepts

- This work aims to quantify the benefit of new mainstream architectures to the HEP community through practical benchmarking on recent hardware solutions, including the usage of parallelized HEP applications

Hardware (1)

- Westmere-EP
  - New “workhorse” of most of our computing centers
  - 2 sockets
    - 12 cores / 24 threads
  - Shrinking of the 45 nm Nehalem core
    - 32 nm process technology
    - Added 2 cores per CPU, with same L3 cache memory per each core (2 MB)
    - Same power consumption
  - X5670 specimen tested (2.93 GHz, 95W)

  - Reference: Nehalem-EP X5570 (2.93 GHz, 95W, 4 cores / 8 hw-threads)

Hardware (2)

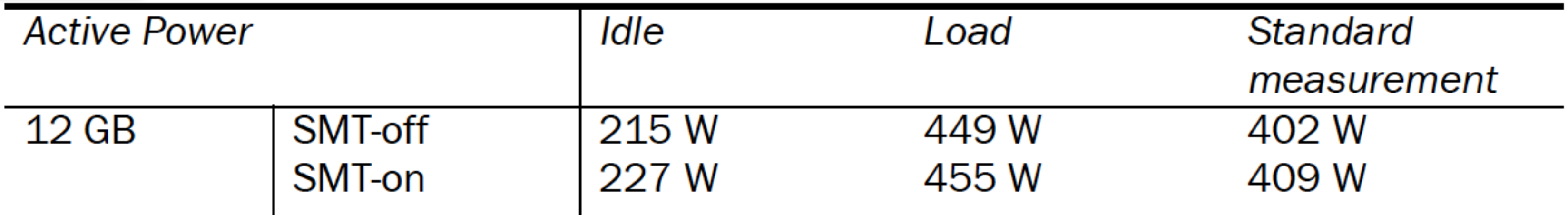
- Nehalem-EX
  - Designed for specialized multi-socket applications -- for a price of 1 Nehalem-EX chip you can get ~4 Westmere-EP chips
  - 4 sockets * 8 cores * 2 hw-thread = 32 cores / 64 hw-threads
  - Representative of the previous Nehalem generation
    - Older 45nm process technology
  - X7560 specimen tested (2.26 GHz, 130W)
  - Reference: Dunnington X7460 (2.66 GHz, 130W, 6 cores / no hw-threads)

Test setup

1. HEPSPEC06 performance
   - a standard HEP benchmark
2. Multi-threaded Geant4 prototype scalability (J. Apostolakis et al, *Multithreaded Geant4: Semi-automatic transformation into scalable thread-parallel software*, Europar 2010)
   - parallel implementation of the test40 example from Geant4
     - 200 random events per thread
   - ParFullCMSmt, a full CMS simulation ported to a parallel model
     - 100 pi- events per thread @ 300 GeV
3. MPI Parallel Maximum Likelihood (ML) fit with ROOT/RooFit (A. Lazzaro and L. Moneta, *MINUIT package parallelization and applications using the RooFit package*, *J. Phys.: Conf. Ser.* **219** 042044)
4. Power consumption vs performance
5. NUMA aspects (Nehalem-EX)

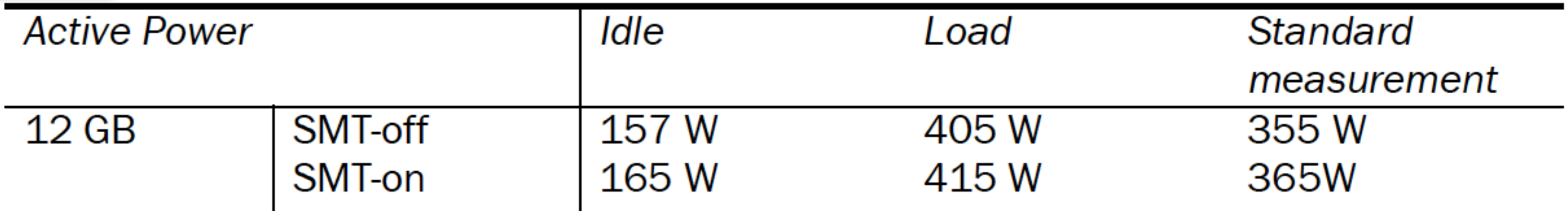
Westmere-EP – standard energy measurements

Two PSUs



Active Power		*Idle*	*Load*	*Standard measurement*
12 GB	SMT-off	215 W	449 W	402 W
	SMT-on	227 W	455 W	409 W

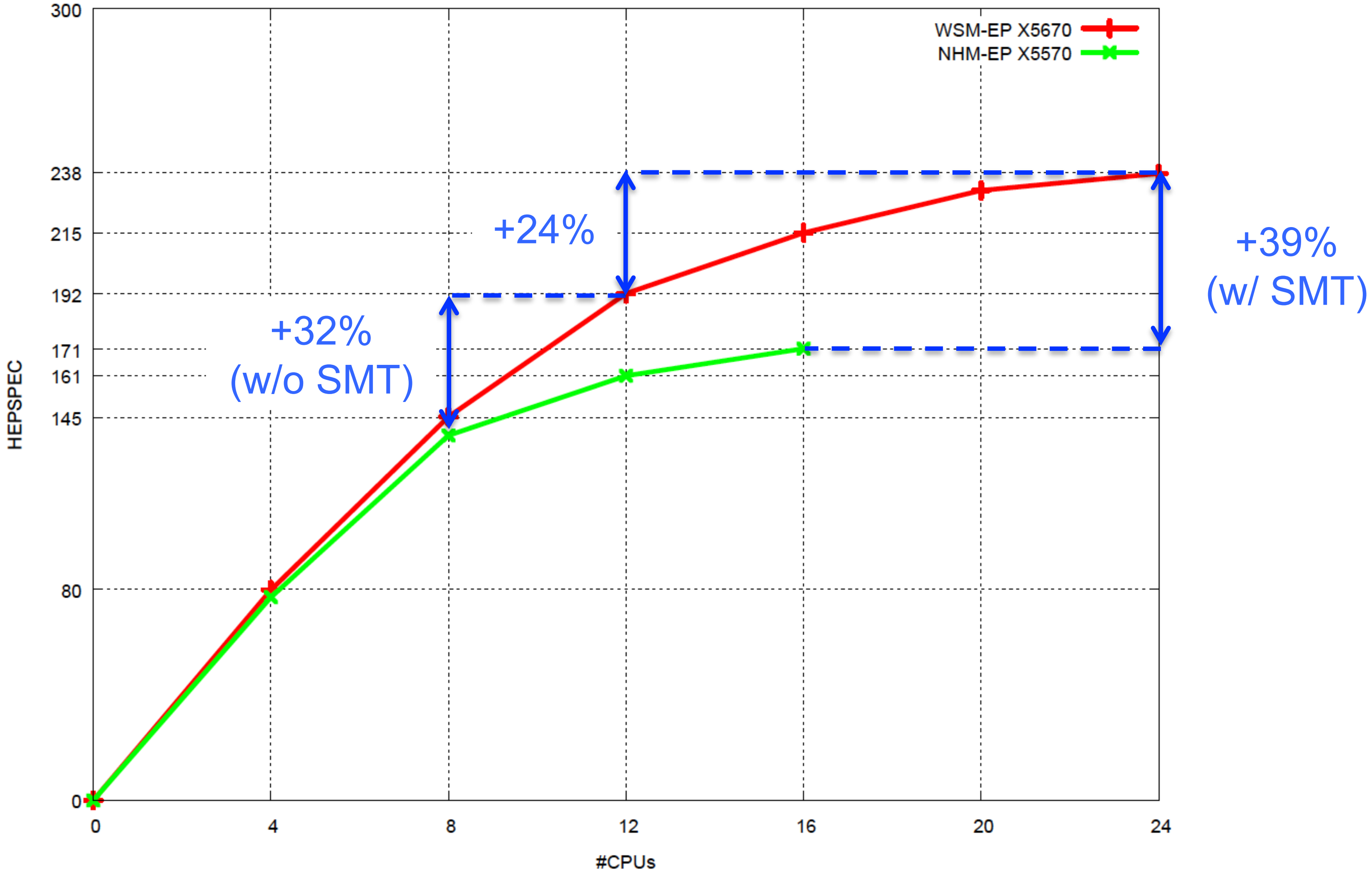
One PSU



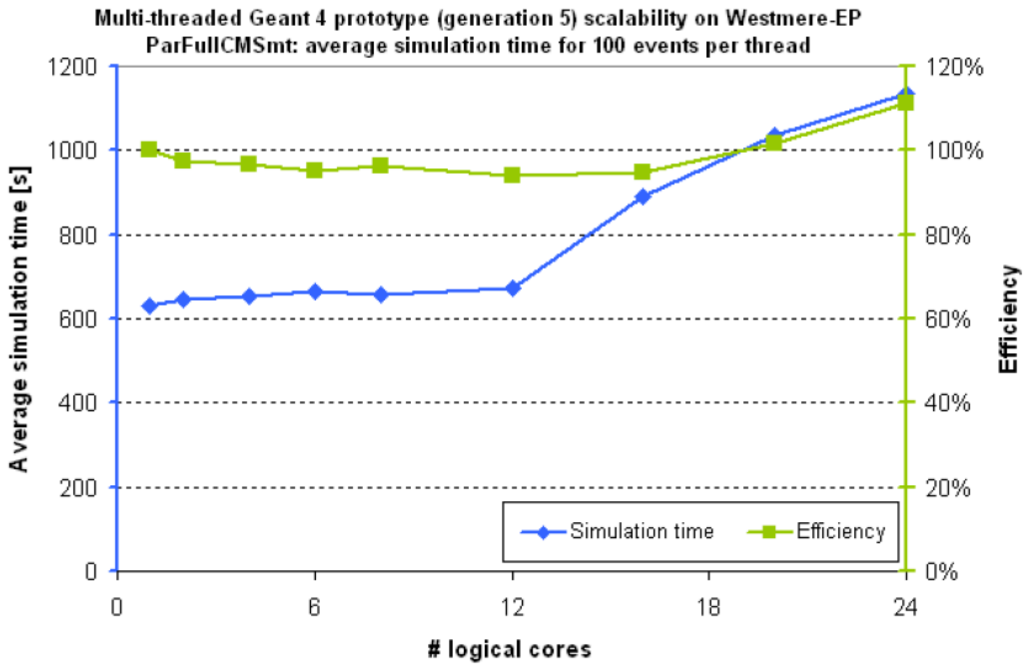
Active Power		*Idle*	*Load*	*Standard measurement*
12 GB	SMT-off	157 W	405 W	355 W
	SMT-on	165 W	415 W	365W

- Remarks:
  - 1 power supply vs. 2 makes a difference in power consumption
  - Turning SMT on introduces a minor penalty in power consumption: <5%

Westmere-EP – HEPSPEC06



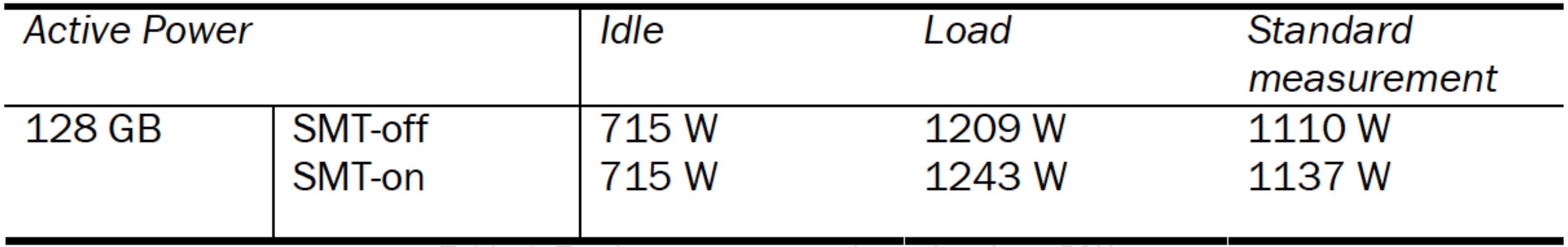
Westmere EP – ParFullCMSmt



- Test looking at throughput (TP), i.e. weak scaling
- Efficiency (% of max theoretical TP)
  - 97% @ 4 cores
  - 96% @ 8 cores
  - 94% @ 12 cores
- SMT benefit @ 24 threads:18% more real TP than 12 threads

Nehalem-EX – standard energy measurements

- Respectable power consumption:

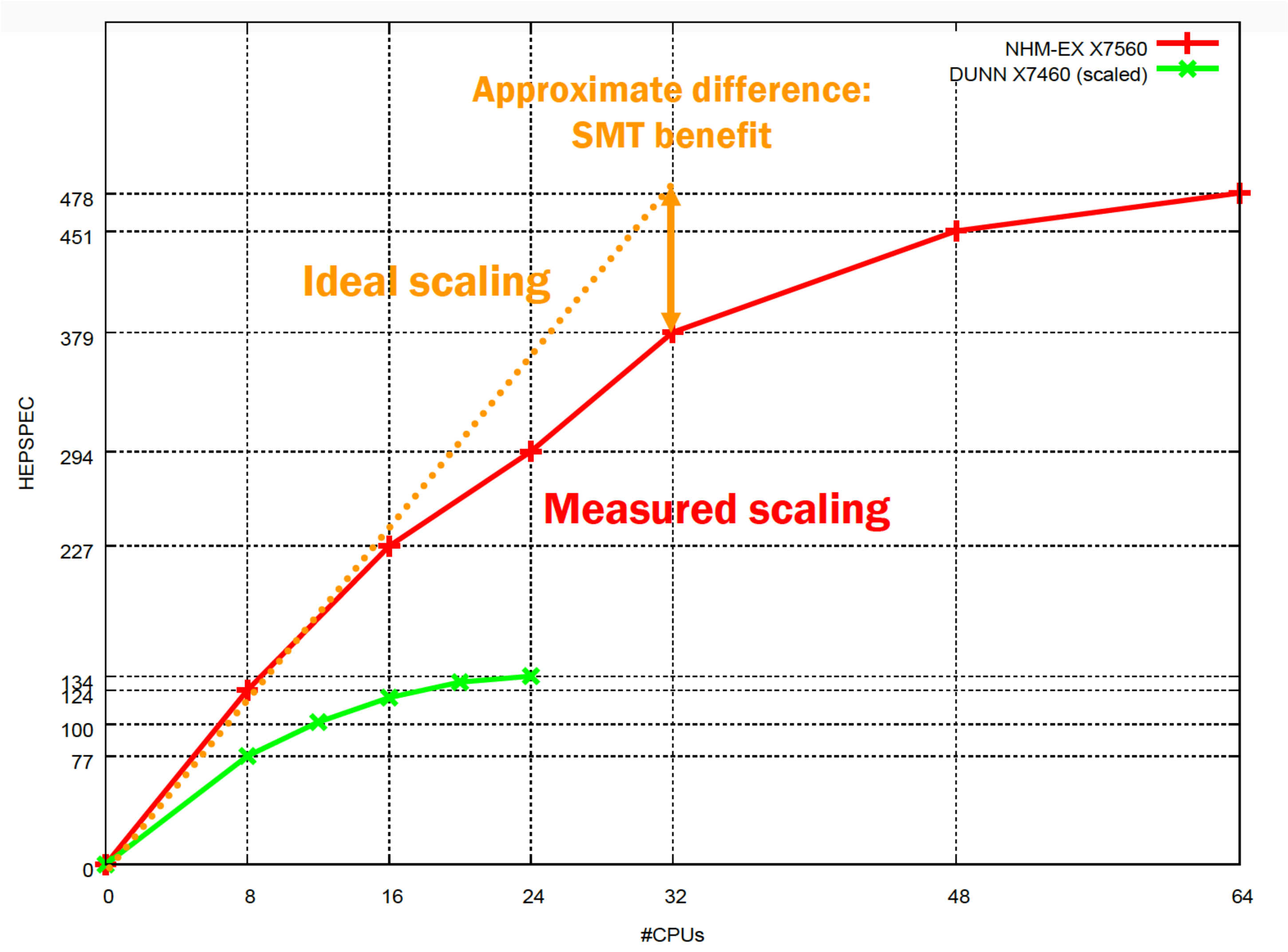


Active Power		*Idle*	*Load*	*Standard measurement*
128 GB	SMT-off	715 W	1209 W	1110 W
	SMT-on	715 W	1243 W	1137 W

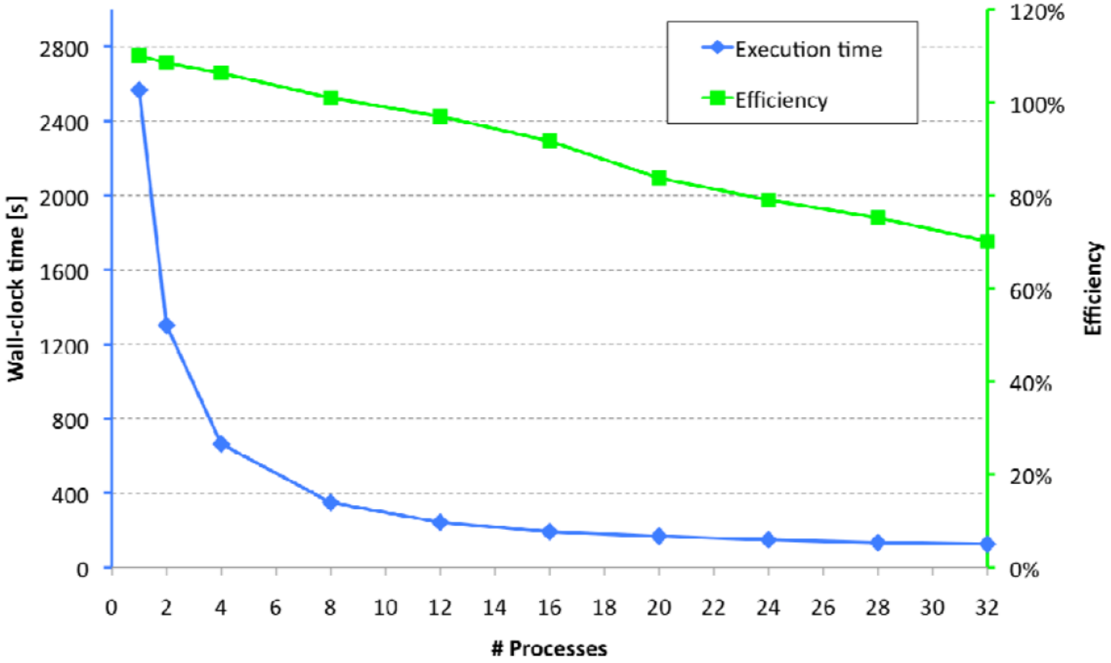
Table 1: Total power consumption using three PSUs

  - 450W (40%) is spent just on memory…
  - No comparison to Dunnington in this case

Nehalem-EX – HEPSPEC06



ML Fit



- Strong scaling:
  - fraction of execution time spend in code we can parallelize is 98.7%
  - Scaling as predicted by Amdahl’s law
- Test done with Turbo Mode on
  - Efficiency calculated wit respect to 1 process with Turbo Mode off

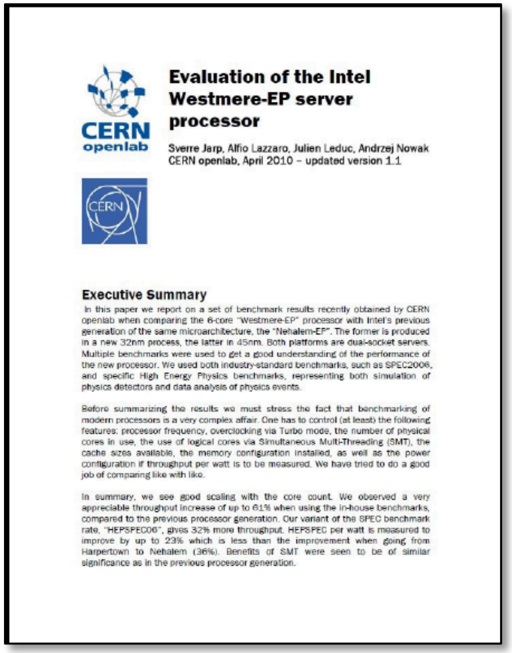
Conclusion

- Westemere-EP VS Nehalem-EP
  - 50% core increase, but HEPSPEC06 numbers only 32% better
  - Overall improvements between 39% and 61% (mostly due to core increase)
  - SMT benefit: 15% - 24% (unchanged)
  - 10% - 23% performance per Watt improvement
    - The previous transition (Core 2 -> Nehalem) was ~35%

- Nehalem-EX VS Dunnington (frequency scaled)
  - 33% core increase reflected in performance
  - Total TP increase: 3.5x on HEPSPEC06!
    - Credited to weak Dunnington performance
  - 47% - 87% more TP on in-house applications
  - SMT benefit: 19% - 28% (no SMT on Dunnington)
  - Significant power consumption

References

- Thanks to Intel collaborators
- All tests with more details are reported at openlab website (technical documents section 2010)
  - http://www.cern.ch/openlab



Evaluation of the Intel Westmere-EP server processor

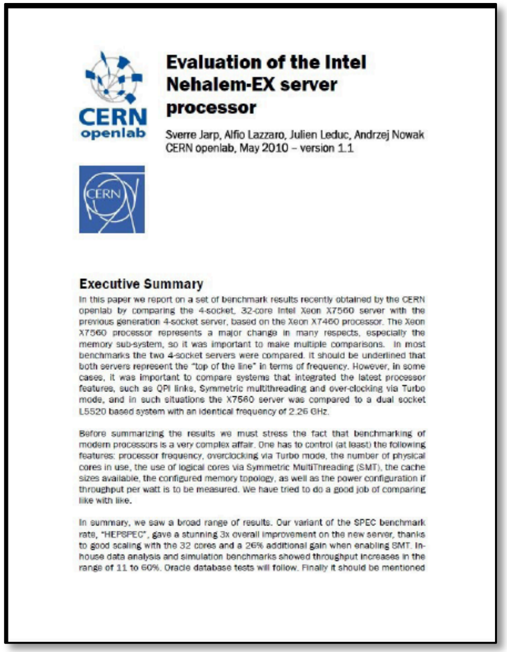
Sverre Jarp, Alfio Lazzaro, Julien Leduc, Andrzej Nowak
CERN openlab, April 2010 – updated version 1.1

Executive Summary

In this paper we report on a set of benchmark results recently obtained by CERN openlab when comparing the 6-core "Westmere-EP" processor with Intel's previous generation of the same microarchitecture, the "Nehalem-EP". The former is produced in a new 32nm process, the latter in 45nm. Both platforms are dual-socket servers. Multiple benchmarks were used to get a good understanding of the performance of the new processor. We used both industry-standard benchmarks, such as SPEC2006, and specific High Energy Physics benchmarks, representing both simulation of physics detectors and data analysis of physics events.

Before summarizing the results we must stress the fact that benchmarking of modern processors is a very complex affair. One has to control (at least) the following features: processor frequency, overclocking via Turbo mode, the number of physical cores in use, the use of logical cores via Simultaneous Multi-Threading (SMT), the cache sizes available, the memory configuration installed, as well as the power configuration if throughput per watt is to be measured. We have tried to do a good job of comparing like with like.

In summary, we see good scaling with the core count. We observed a very appreciable throughput increase of up to 61% when using the in-house benchmarks, compared to the previous processor generation. Our variant of the SPEC benchmark rate, "HEPSPEC06", gives 32% more throughput. HEPSPEC per watt is measured to improve by up to 23% which is less than the improvement when going from Harpertown to Nehalem (36%). Benefits of SMT were seen to be of similar significance as in the previous processor generation.



Evaluation of the Intel Nehalem-EX server processor

Sverre Jarp, Alfio Lazzaro, Julien Leduc, Andrzej Nowak
CERN openlab, May 2010 – version 1.1

Executive Summary

In this paper we report on a set of benchmark results recently obtained by the CERN openlab by comparing the 4-socket, 32-core Intel Xeon X7560 server with the previous generation 4-socket server, based on the Xeon X7460 processor. The Xeon X7560 processor represents a major change in many respects, especially the memory sub-system, so it was important to make multiple comparisons. In most benchmarks the two 4-socket servers were compared. It should be underlined that both servers represent the "top of the line" in terms of frequency. However, in some cases, it was important to compare systems that integrated the latest processor features, such as QPI links, Symmetric multithreading and over-clocking via Turbo mode, and in such situations the X7560 server was compared to a dual socket L5520 based system with an identical frequency of 2.26 GHz.

Before summarizing the results we must stress the fact that benchmarking of modern processors is a very complex affair. One has to control (at least) the following features: processor frequency, overclocking via Turbo mode, the number of physical cores in use, the use of logical cores via Symmetric MultiThreading (SMT), the cache sizes available, the configured memory topology, as well as the power configuration if throughput per watt is to be measured. We have tried to do a good job of comparing like with like.

In summary, we saw a broad range of results. Our variant of the SPEC benchmark rate, "HEPSPEC", gave a stunning 3x overall improvement on the new server, thanks to good scaling with the 32 cores and a 26% additional gain when enabling SMT. In-house data analysis and simulation benchmarks showed throughput increases in the range of 11 to 60%. Oracle database tests will follow. Finally it should be mentioned

Q&A