Evaluating the power efficiency and performance of multi-core platforms using HEP workloads

April 14\textsuperscript{th}, CHEP 2015, Okinawa

Pawel Szostek, Vincenzo Innocente
Outline

› “Haswell-EP” microarchitecture refresher
› Description of the tested setup
› Experiments and results: scalability, vectorization, single core performance, power efficiency
› Conclusions
“Haswell-EP” microarchitecture refresher

- **dual-socket** platform, 22nm, AVX2 256b vectors, DDR4 support
- 10-core and higher core counts come with **two memory controllers** (Cluster-on-Die),
- beefed-up core (already in single socket HSW):
  - 8 instead of 6 execution units
  - bigger OoO execution buffer (192 vs. 168 entries)
“Haswell-EP” microarchitecture refresher (II)

› comes with a couple of interesting power-saving features:
  - frequency/voltage scaled on a per-core basis
  - uncore frequency/voltage scaling independent from cores
  - configurable TDP
  - energy efficient turbo mode

› As in the single socket, AVX slows down the clock. Switching to AVX is penalized
Setup for the tests

› **Haswell-EP (dual socket):**
  - E5-2699v3 – 18 cores, 2.3Ghz, 145W TDP
  - E5-2698v3 – 16 cores, 2.3GHz, 135W
  - E5-2683v3 – 14 cores, 2.0GHz, 120W

› **Ivy Bridge-EP:**
  - E5-2695v2 – 12 cores, 2.4Ghz, 115W

› **Sandy Bridge-EP:**
  - E5-2690 – 8 cores, 2.9GHz, 135W
Setup for the tests (II)

- 2U Intel Bobcat Pass chassis with 4 servers in an enclosure
- 8x8GB 2133MT/s DDR4 DIMMs
- 2xIntel SSD DC S3500 240GB, with LVM stripping
- Turbo Boost disabled, SMT enabled, P- and C- states disabled unless stated otherwise
- SLC 6.6 with the 2.6.32-504.12.2
The benchmarks

› **Standard:** HEPSPEC06
  - not optimized for next-gen hardware

› **Analysis:**
  - MLFit
    - Threaded (pthreads, MPI, OpenMP, TBB)
    - Vectorized (Cilk+)
  - VIFit (Vincenzo Innocente Fit)
    - lightweight version of MLfit
    - NUMA-aware memory management

› **Simulation:** up-to-date ParFullCMS with Geant 4 v.10.01 patch 01 [MT]
  - multi-threaded, not vectorized
HEP-SPEC06 scalability (freq. scaled)

- Sandy Bridge-EP E5-2690
- Ivy Bridge-EP E5-2695v2
- Haswell-EP E5-2683v3, COD enabled
- Haswell-EP E5-2683v3, COD disabled
- Haswell-EP E5-2698v3 COD enabled
- Haswell-EP E5-2698v3, COD disabled
- Haswell-EP E5-2699v3, COD enabled
- Haswell-EP E5-2699v3, COD disabled
HEP-SPEC06 per core (freq. scaled)

- Sandy Bridge-EP E5-2690
- Ivy Bridge-EP E5-2695v2
- Haswell-EP E5-2683v3, COD enabled
- Haswell-EP E5-2683v3, COD disabled
- Haswell-EP E5-2698v3 COD enabled
- Haswell-EP E5-2698v3, COD disabled
- Haswell-EP E5-2699v3 COD enabled
- Haswell-EP E5-2699v3, COD disabled

Number of threads vs. HEP-SPEC06 per core, frequency scaled.
COD gains in HS06

| System                  | Total HS06 Score | COD Disabled | COD Enabled |
|-------------------------|------------------|--------------|-------------|
| Haswell-EP E5-2683v3    | 600.00           | 500.00       | 400.00      |
| Haswell-EP E5-2698v3    | 700.00           | 600.00       | 500.00      |
| Haswell-EP E5-2699v3    | 800.00           | 700.00       | 600.00      |

COD disabled
COD enabled
Data throughput scalability

- Sandy Bridge-EP E5-2690
- Ivy Bridge-EP E5-2695v2
- Haswell-EP E5-2683v3, COD enabled
- Haswell-EP E5-2683v3, COD disabled
- Haswell-EP E5-2698v3 COD enabled
- Haswell-EP E5-2698v3, COD disabled
- Haswell-EP E5-2699v3 COD enabled
- Haswell-EP E5-2699v3, COD disabled
ParFullCMS (II)

Power efficiency scalability

- Sandy Bridge-EP E5-2690
- Ivy Bridge-EP E5-2695v2
- Haswell-EP E5-2683v3, COD enabled
- Haswell-EP E5-2683v3, COD disabled
- Haswell-EP E5-2698v3 COD enabled
- Haswell-EP E5-2699v3 COD enabled
- Haswell-EP E5-2699v3, COD disabled
- Haswell-EP E5-2699v3, COD disabled
VIFit speed-up
(freq. scaled, SNB is the baseline)
Vectorization speedup
-no-vec is the baseline

Sandy Bridge-EP
E5-2690

Ivy Bridge-EP
E5-2695v2

Haswell-EP
E5-2683v3,
COD enabled

Haswell-EP
E5-2683v3,
COD disabled

Haswell-EP
E5-2698v3

Haswell-EP
E5-2698v3,
COD enabled

Haswell-EP
E5-2698v3,
COD disabled

Haswell-EP
E5-2699v3,
COD enabled

Haswell-EP
E5-2699v3,
COD disabled

no vec

SSE4.2

AVX

AVX2
ParFullCMS – energy efficiency

Three power settings (on Haswell-EP E5-2699v3 – 18 cores/socket)

|                   | Turbo | P- and C- states | Uncore scaling | Energy Efficient Turbo |
|-------------------|-------|------------------|----------------|------------------------|
| “standard”        | NO    | NO               | YES            | NO                     |
| Low power         | YES   | YES              | YES            | YES                    |
| High power        | YES   | NO               | NO             | NO                     |
ParFullCMS – energy efficiency (II)

Power efficiency scalability

Data throughput scalability

- Normal power
- Low power
- High power

Number of threads vs. events/J

Number of threads vs. events/s
Conclusions

- Haswell-EP offers major improvements in the uncore part, which can be mainly exploited by multi-threaded (NUMA-aware) applications.
- Otherwise, we “just” get a massive core-count increase.
- Haswell-EP provides higher vectorization gains than previous dual socket platforms.
- New power saving features allow lowering down the absolute power consumption, but for a cost of lower power efficiency.
Questions?

Other questions?
pawel.szostek@cern.ch
Detailed COD diagram
Haswell EP Die Configurations

| Chop  | Columns | Home Agents | Cores   | Power (W) | Transistors (B) | Die Area (mm²) |
|-------|---------|-------------|---------|-----------|-----------------|----------------|
| HCC   | 4       | 2           | 14-18   | 110-145   | 5.69            | 662            |
| MCC   | 3       | 2           | 6-12    | 65-160    | 3.84            | 492            |
| LCC   | 2       | 1           | 4-8     | 55-140    | 2.60            | 354            |

Not representative of actual die-sizes, orientation and layouts – for informational use only.
Core-to-core bandwidth
Backup – VI$\text{F}$it with a core disabled

VI$\text{F}$it runtime with a core disabled

mean values over 100 runs

Pawel Szostek, Vincenzo Innocente – CERN openlab
Backup – power consumption

- Sandy Bridge-EP E5-2690
- Ivy Bridge-EP E5-2695v2
- Haswell-EP E5-2683v3, COD enabled
- Haswell-EP E5-2683v3, COD disabled
- Haswell-EP E5-2698v3, COD enabled
- Haswell-EP E5-2698v3, COD disabled
- Haswell-EP E5-2699v3, COD enabled
- Haswell-EP E5-2699v3, COD disabled

Legend:
- blue: idle
- red: loaded, turbo off
- green: loaded, turbo on

14/04/2015

Pawel Szostek, Vincenzo Innocente – CERN openlab