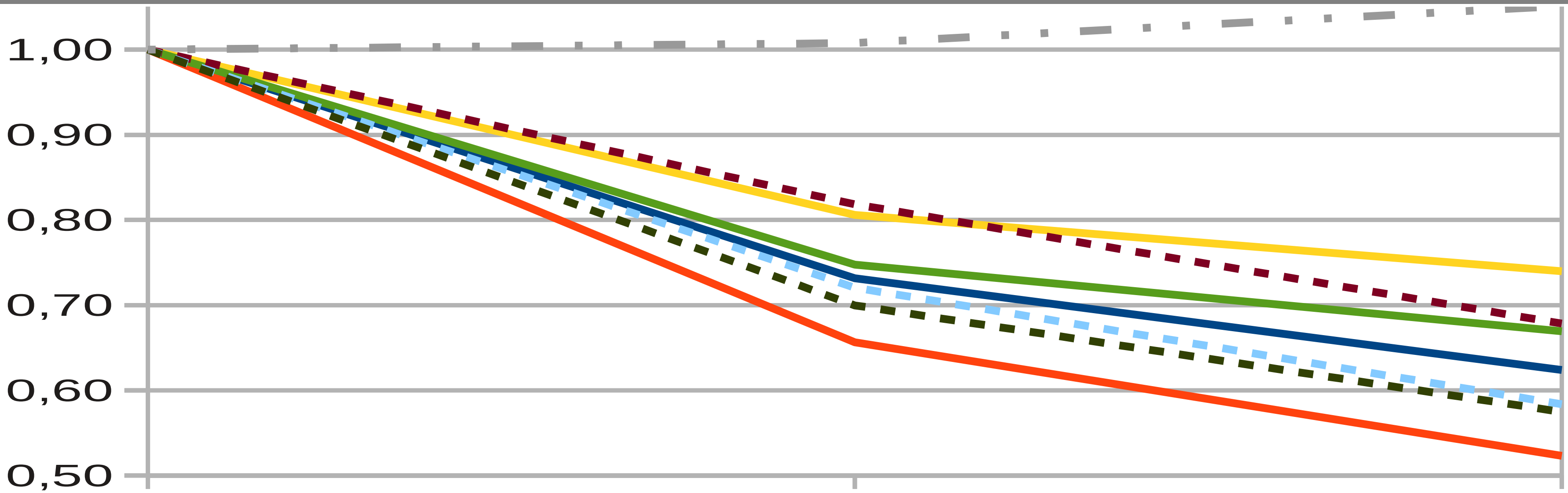
Correlation of Benchmark Results, Job Performance, and the Number of Job Slots per Core

Manfred Alef

STEINBUCH CENTRE FOR COMPUTING (SCC)



Preamble

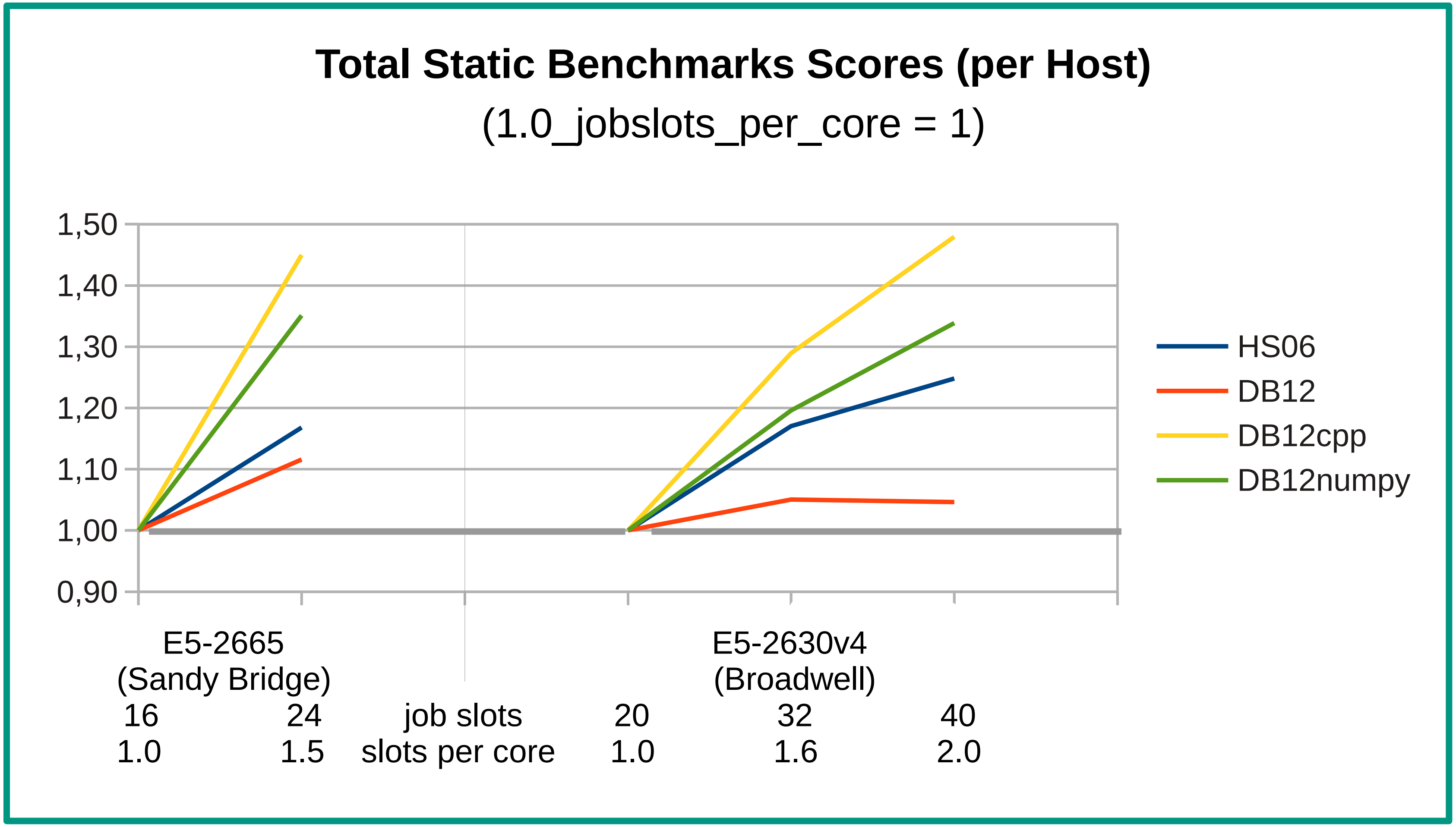
- Many presentations (e.g. GDB, HEPiX Benchmarking Working Group) comparing job performance (events/s) with benchmark results
  - → Focus on CPU model:
    "Magic Boost" of modern processor architectures
  - → No presentations are reporting possible dependencies on special WN configuration items like the number of job slots per core
    - There are however a lot of presentations comparing benchmarks (HS06, DB12*, KV, ttbar, ...)
    - Few talks showing dependencies implicitly
- Evidence of influence of configuration details on job performance
  - → Wide range of benchmark results when varying the number of benchmark copies

Still Open Question: Influence of Number of Job Slots

- Impact of the number of configured job slots (or benchmark copies) per physical core on the job performance?
  - ➔ No common strategy by sites about the number of job slots per WN
    - #job_slots == #physical_cores (HT on or off)
    - #cores < #job_slots < #logical_processors (HT enabled)
    - #job_slots == #logical_processors (HT enabled)
  - ➔ Undisclosed overcommitment of cloud IaaS hypervisors
    - (see the examples in previous talks by D. Giordano)

Still Open Question: Influence of Number of Job Slots

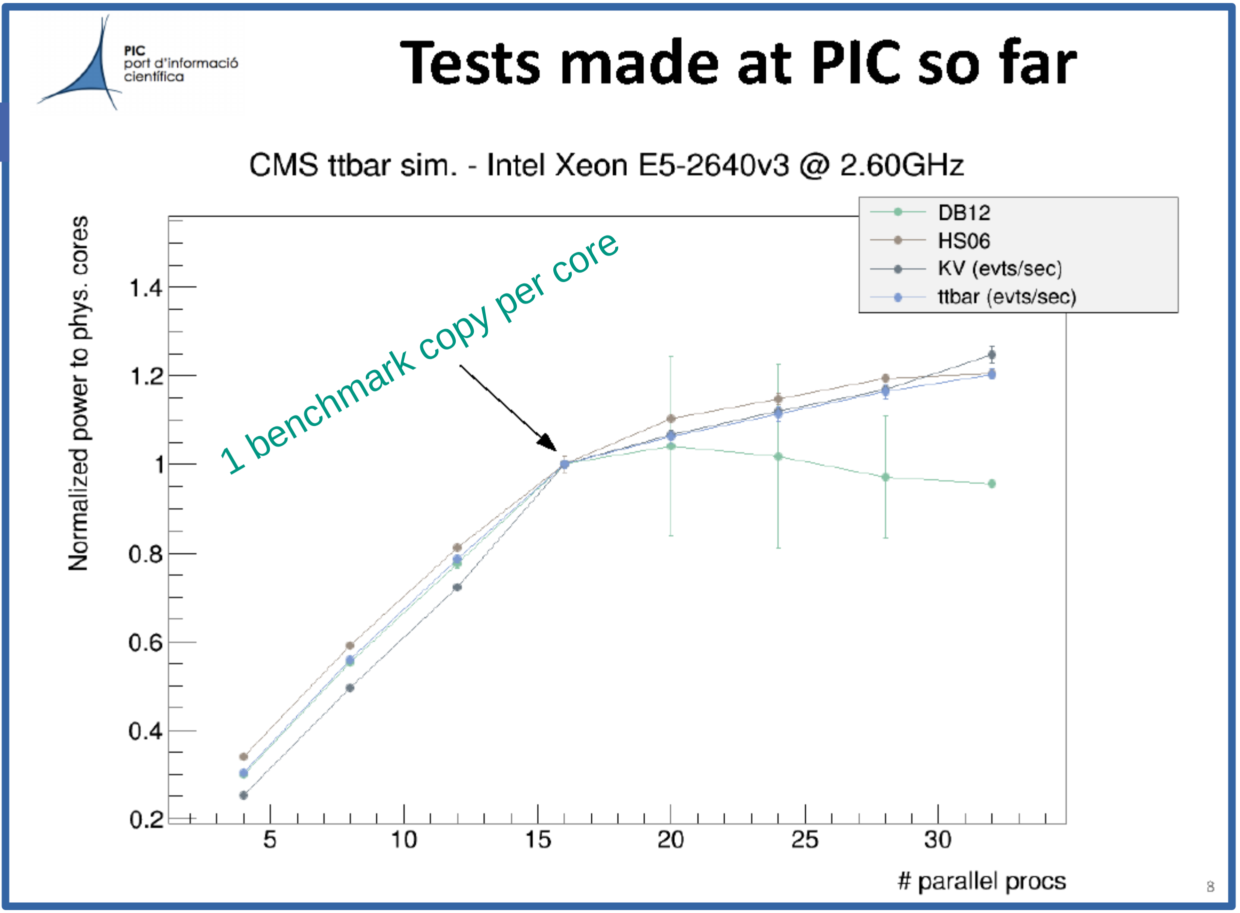
- Impact of the number of configured job slots ...



Still Open Question: Influence of Number of Job Slots

- Impact of the number of configured job slots (or benchmark copies) per physical core on the job performance?

J. Flix et al.



Still Open Question: Influence of Number of Job Slots

- Impact of the number of configured job slots per physical core on the job performance?
  - ➔ What about the corresponding application performance (in units of events/s)?

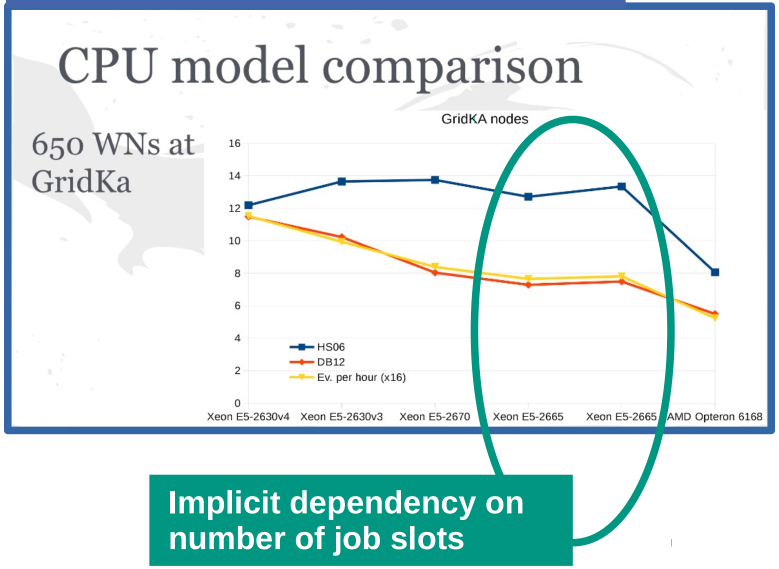
Still Open Question:
Influence of Number of Job Slots

- Background:
  - → The number of job slots per WN is not just a configuration item
    - Requests: RAM per slot (2GB), disk space per slot, ...
  - → Sites are providing the most economical solution (max. HS06/EUR)
    - Tenders at GridKa: 1.5 ... 1.6 job slots per core
  - → However, if the performance is not better when configuring the number of slots > number of cores (as indicated by initial DB12), then 1 slot per core would become the most economic solution
    - Impact on future procurements!

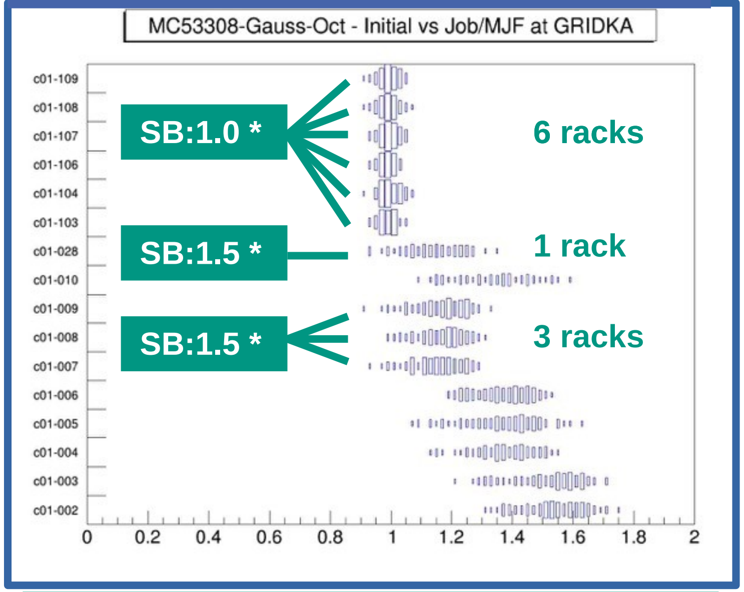
Still Open Question:
Influence of Number of Job Slots

- Investigations by experiments:
  - ➔ Most presentations comparing job performance vs. benchmarks per CPU model, neglecting the configured number of job slots
  - ➔ Few exceptions (and private communication), for instance:

Alice (C. Grigoras, pre-GDB Feb 2017)



LHCb (P. Charpentier, private communication)



*** Intel Xeon E5-2665 (6+1 racks) and -2670 (3 racks) configured with 1 or 1.5 job slots per physical core**

Application Performance Assessments by Site Monitoring

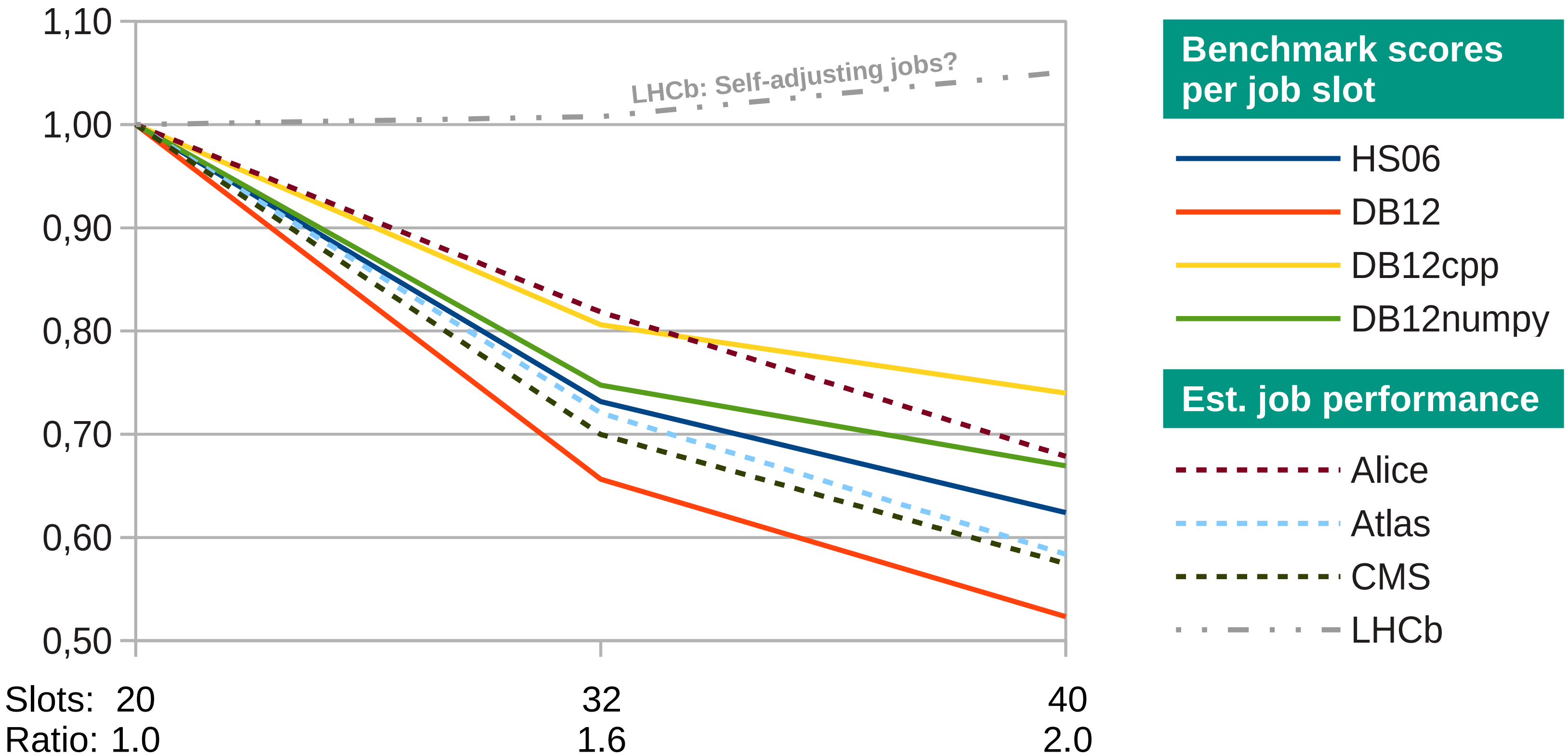
- (See talk 2017-05-05)
- Runtime of pilot payloads estimated by monitoring top commands (aliroot, root.exe, athena.py, cmsRun, python)
  - → Ignoring pre- and postprocessing
  - → Healthy nodes only
  - → Job slot utilization almost always at 100%
  - → Quick-and-dirty assessment, doesn't take into account the job type (simulation, reconstruction, ...), or whether the jobs are part of the same production or not
    - 1 measurement per hour,
      average of 2.5 weeks (2017-04-29 ... 2017-05-15)
  - → Analysis per VO

Results

- Hardware model: Intel E5-2630v4 (Broadwell)
  - ➔ Job slots:
    - 20 job slots (1.0 per core): 40 hosts
    - 32 job slots (1.6 per core): 160 hosts
    - 40 job slots (2.0 per core): 40 hosts
  - ➔ Memory per slot ≥ 2.4 GB
    - Typical mix of low and high memory jobs can run on each type of hosts

Estimated Job Performance vs. Benchmarks

Hardware Platform:
Intel Xeon E5-2630v4 (Broadwell) – 2.2 GHz, 10-core – 2 sockets
RAM per job slot ≥ 2.4 GB



Conclusions

- When comparing job performance vs. benchmark results, not only the CPU model but also WN configuration details (esp. the number of job slots per physical core) should be taken into account
  - → MJF item: $JOBFEATURES/jobslots

Conclusions

- Wider range of benchmark results than job performance?
  - ➔ Quick-and-dirty assessment of job performance by site monitoring over more than 2 weeks
  - ➔ More investigations by experiments desired to move to a more reliable approach
  - ➔ Deeper analysis in case of LHCb required
  - ➔ In case of GridKa:
    - HS06 as well as original DB12-at-boot scores available from MJF
    - Corresponding DB12cpp-at-boot and DB12numpy-at-boot scores are available on demand