Evaluation of Avoton CPU

Andrea Chierici

WLCG Collaboration Workshop 2014

Outline

- Current situation
- Intel Avoton
- Performance measures
- Evaluation carried out by (some) experiments
- Conclusions

Disclaimer

- Some of the hardware tested is still under NDA

Current situation

Current state of computing room

- CNAF PUE 1,60 ☹
- Quite high KWh cost (at least in Italy!)
- Average monthly power required ~1MW
- Many old nodes have a poor HS06/W ratio

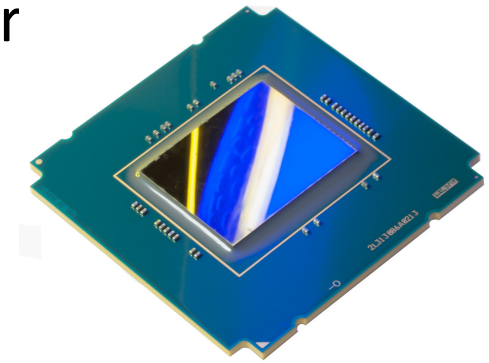
- What about these new low power CPU solutions?

Intel Avoton

Enter Intel Avoton CPU

- Intel® Atom™ Processor C2000
- System-On-Chip (SoC) built on Intel's 22-nanometer process technology.
- Up to 8 64-bit Intel® Atom™ processor cores (no hyperthreading supported) based on next generation Silvermont micro architecture, I/O controllers such as PCI Express, SATA, USB and the Intel QuickAssist Technology for crypto hardware acceleration.
- Created to address the needs of the Microserver, Entry Communication Infrastructure and Cloud Stor
- 12W Max TDP for C2730

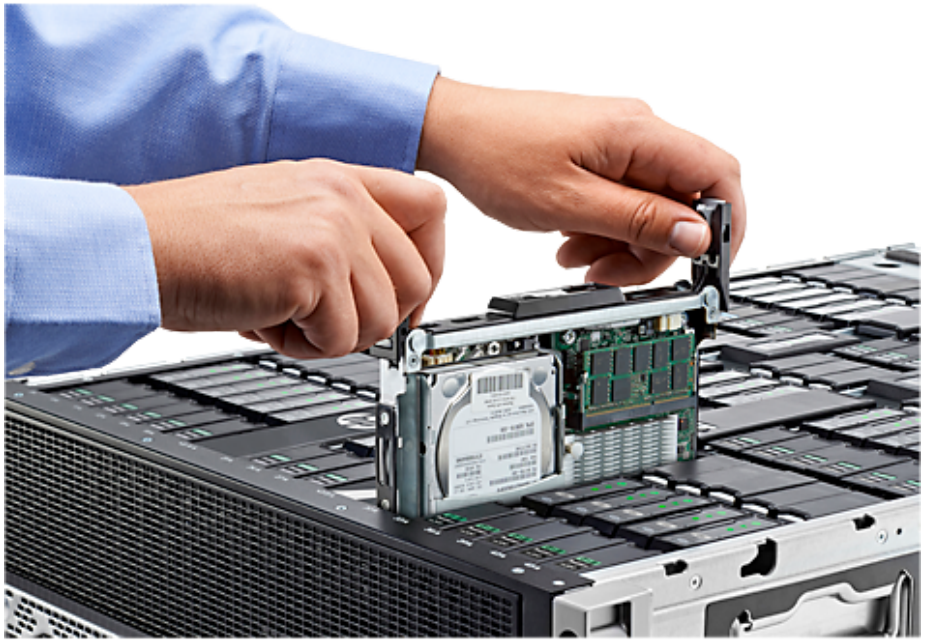
(source: ark.intel.com)



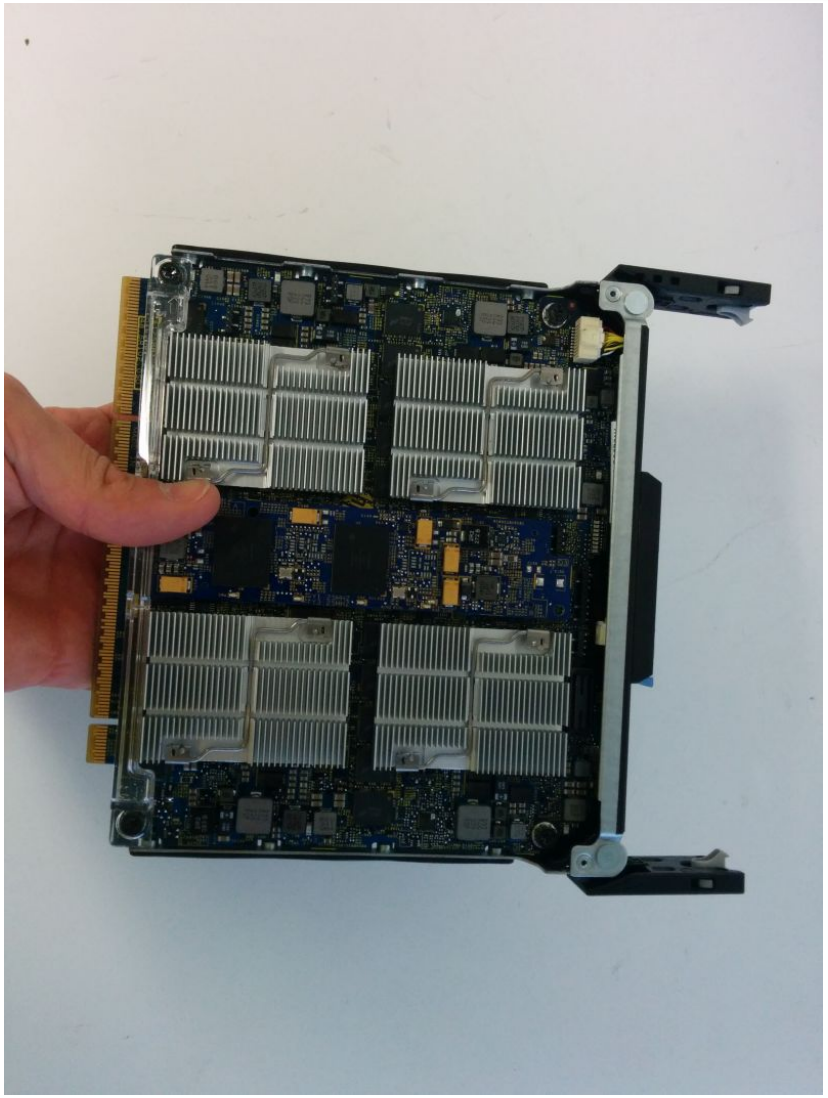
HP Moonshot

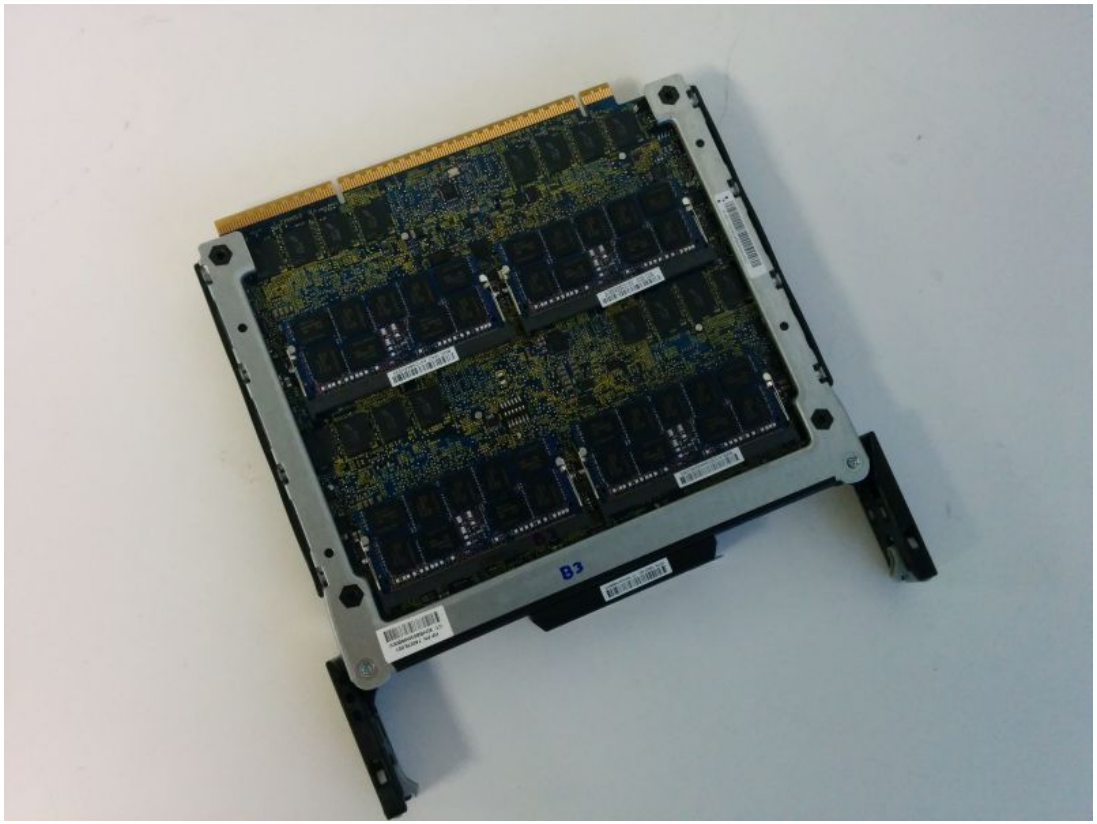
- HP Moonshot 1500 Chassis
  - 4,3U, 45 server cartridges, gigabit connection to two internal switches (4x40Gbps uplinks each)
- m350 server cartridges
  - 4 Avoton C2730 @ 1.7GHz
  - Memory: 16 GB DDR3 PC3-12800 SDRAM (1600 MHz)
  - 64GB integrated SSD
  - Networking: (Internal) dual port 1GbE per CPU

HP Moonshot



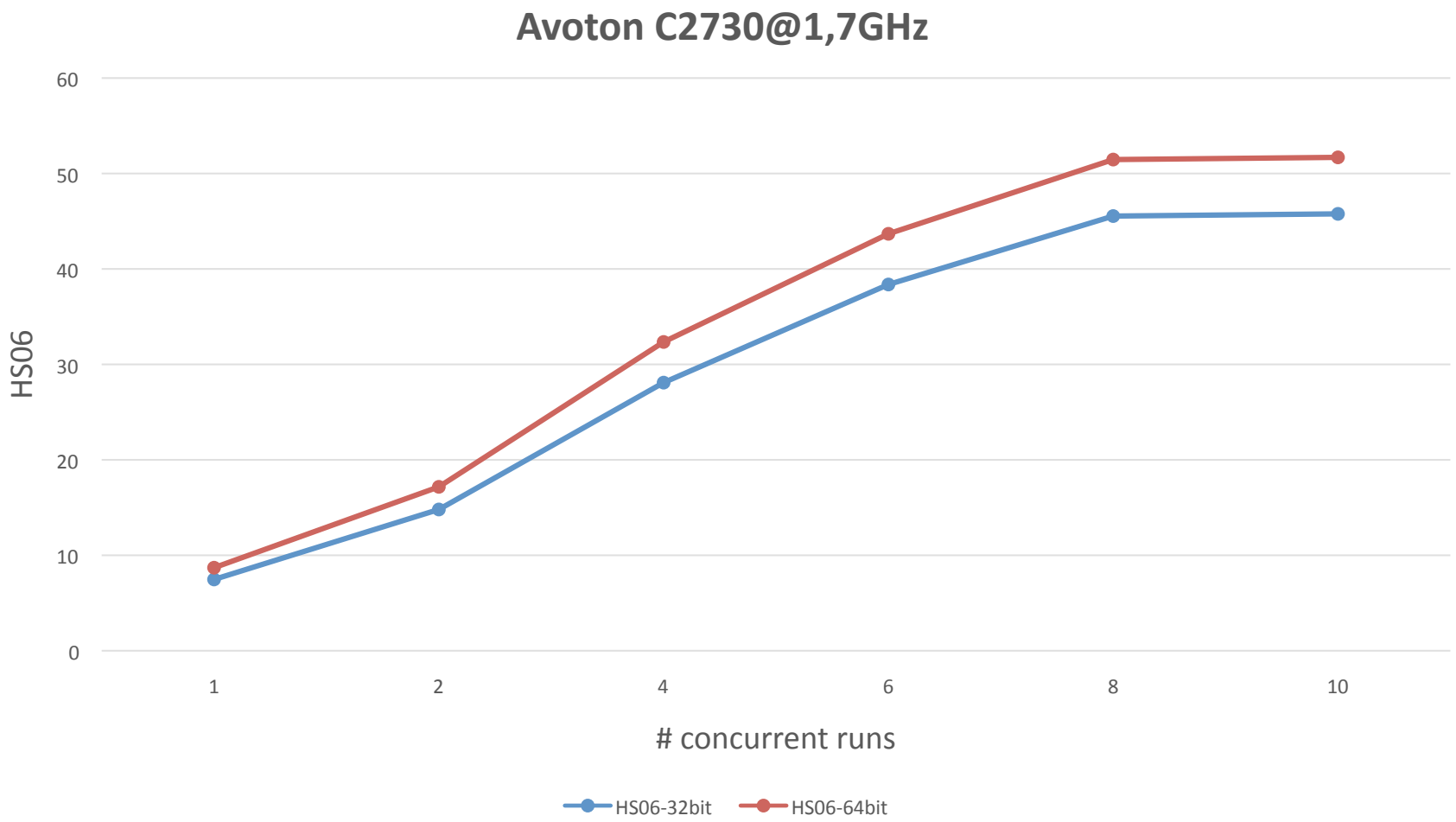
M350 Cartridge



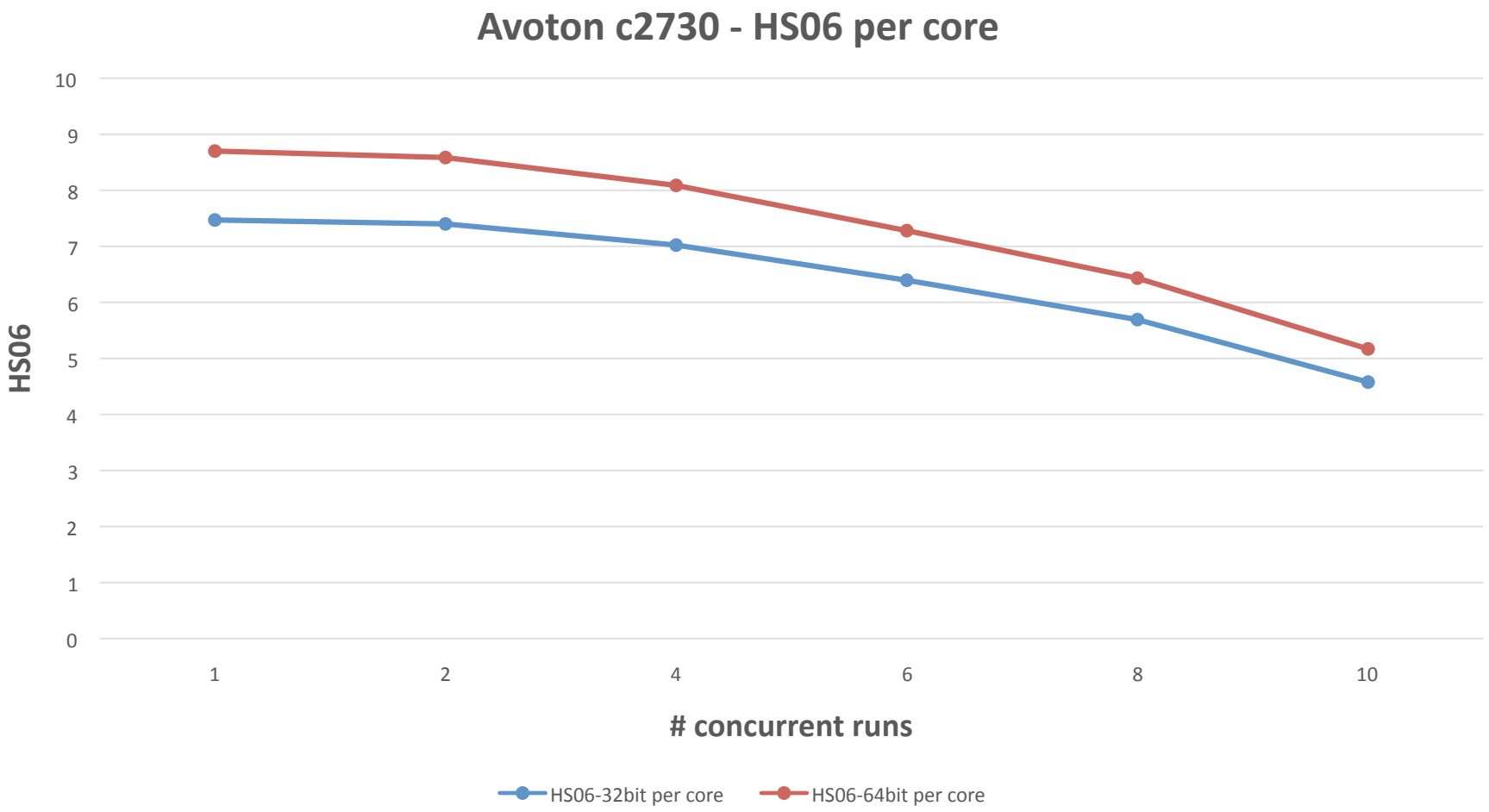


Performance measures

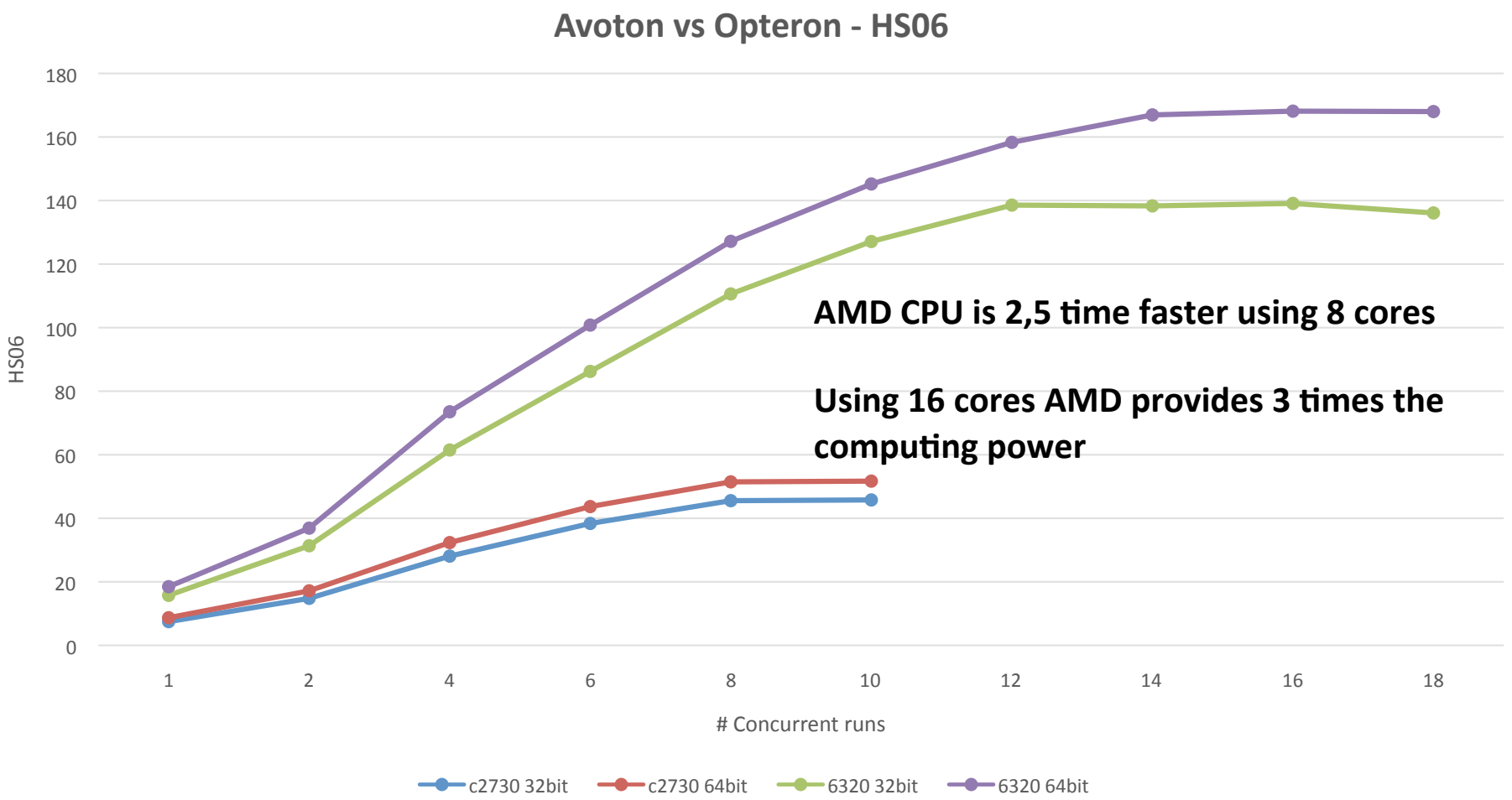
HS06 Performance



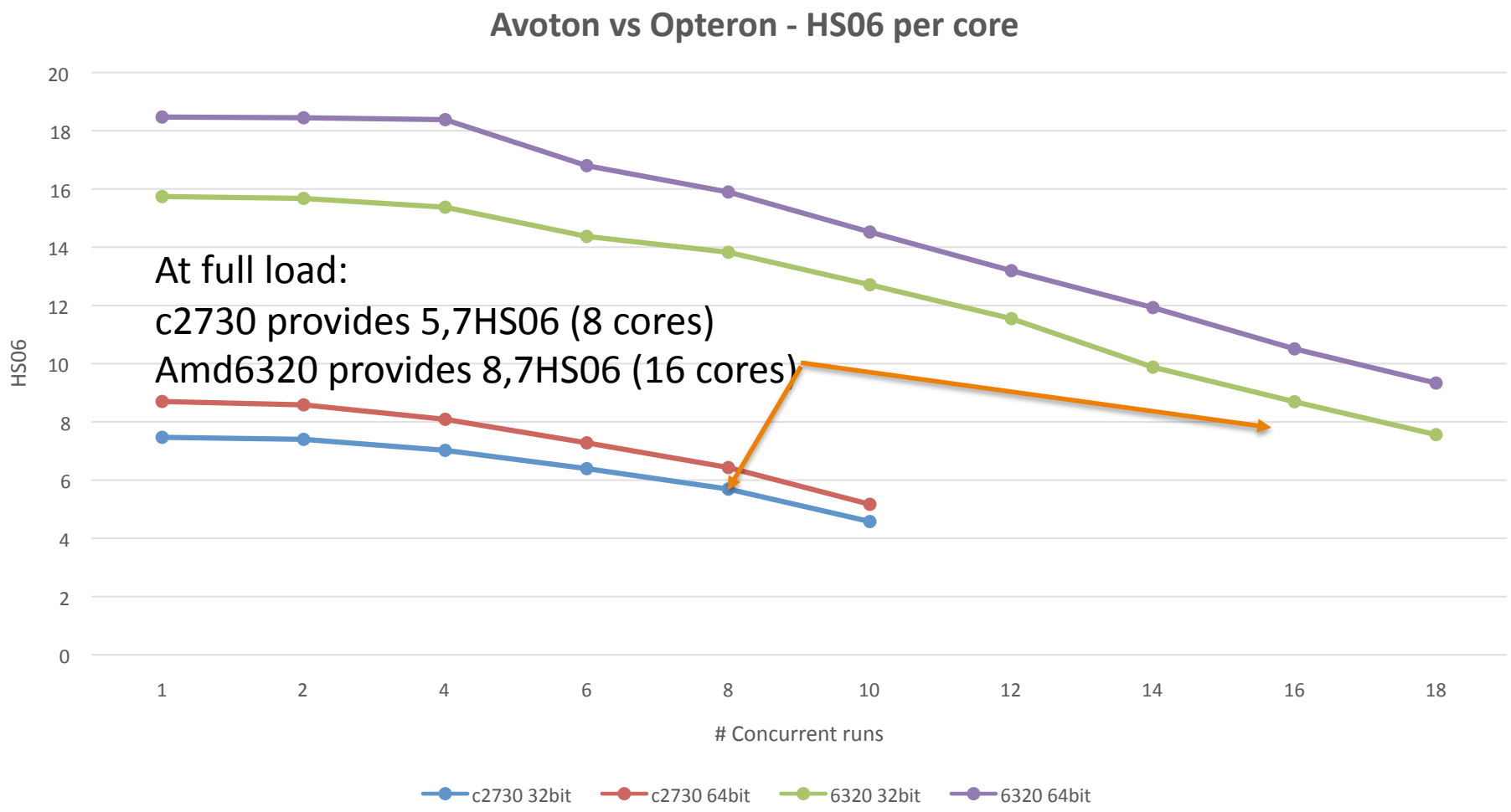
HS06 Performance



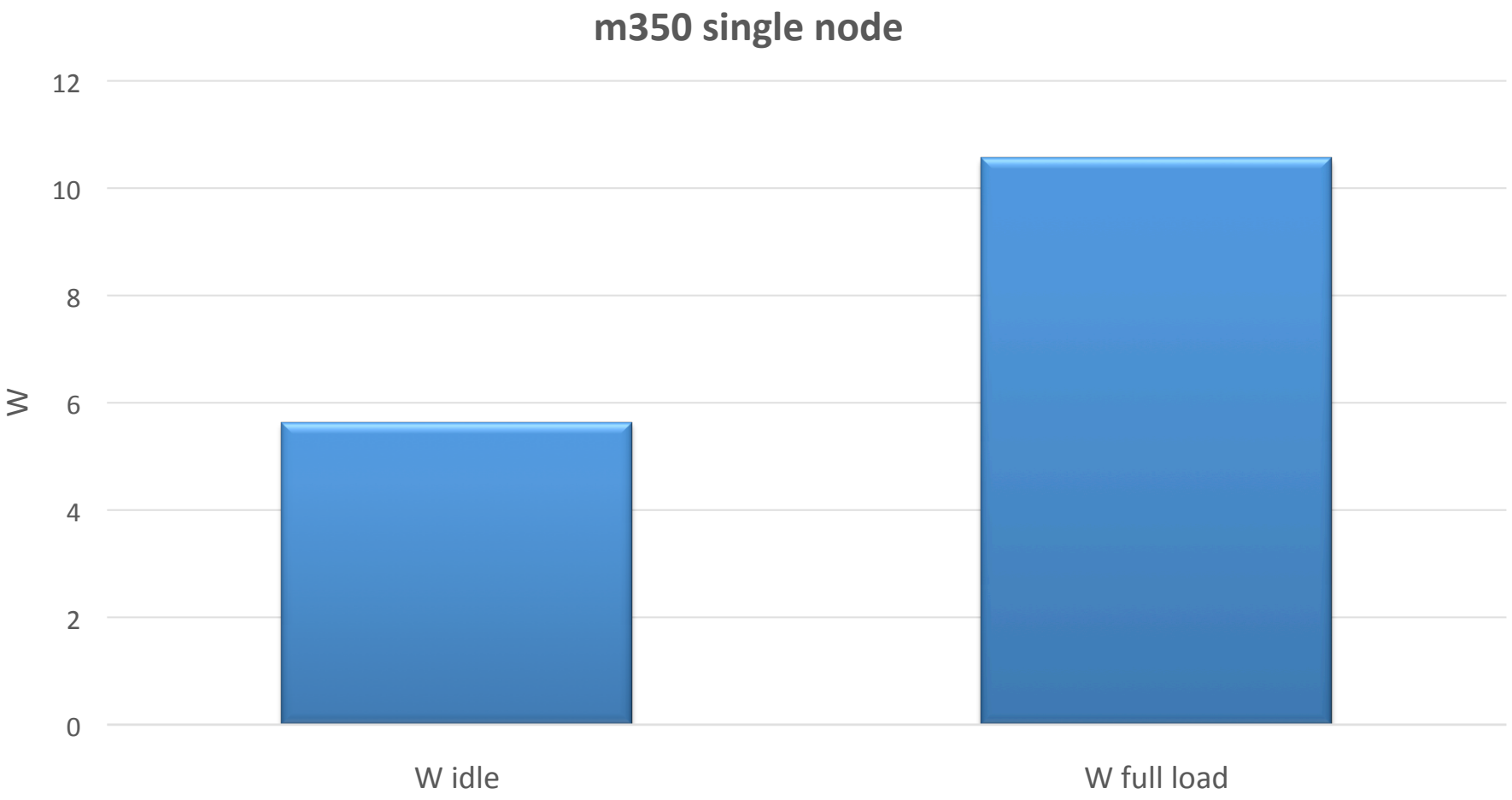
Avoton vs 2013 tender node



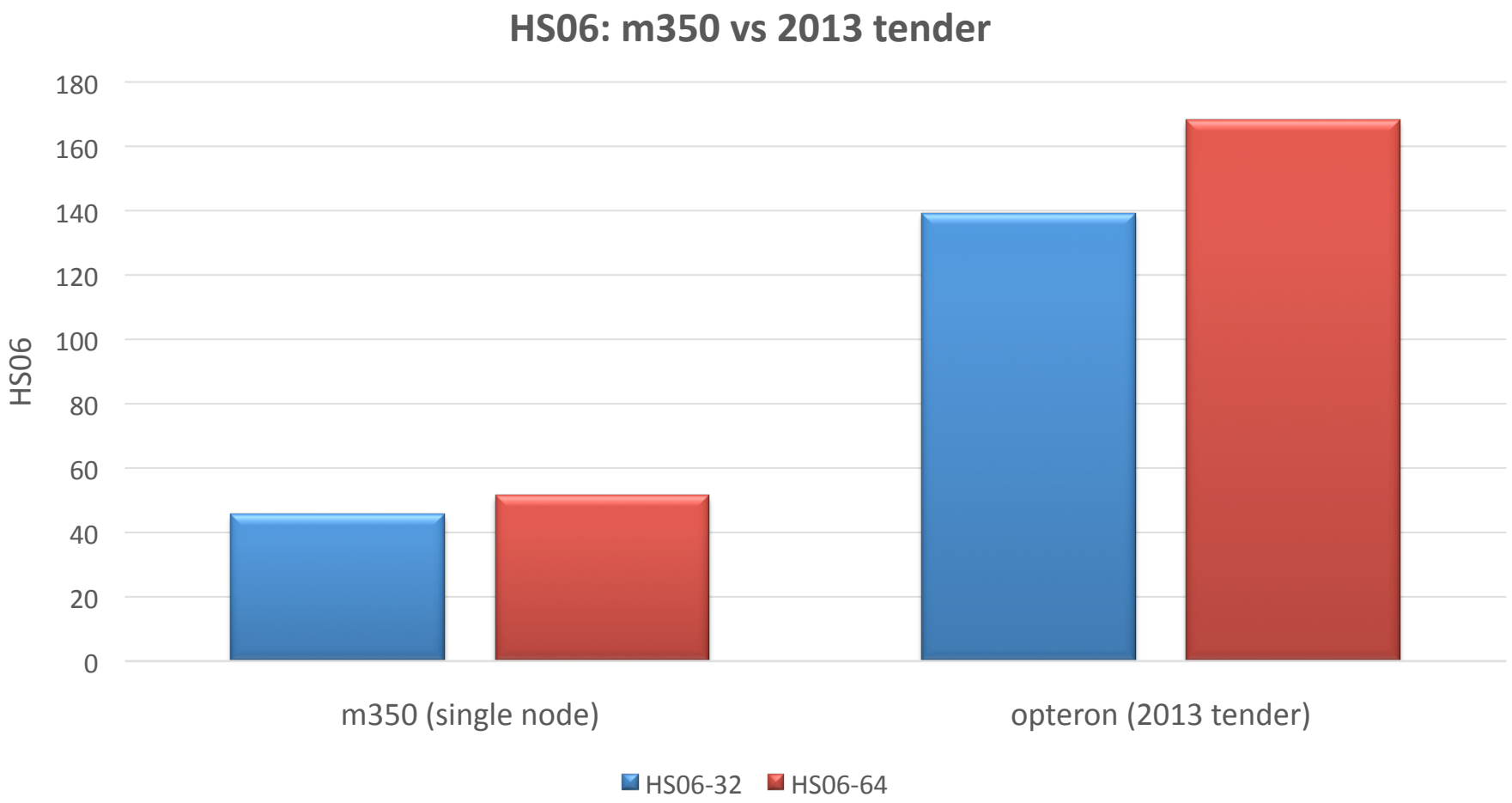
Avoton vs 2013 tender node



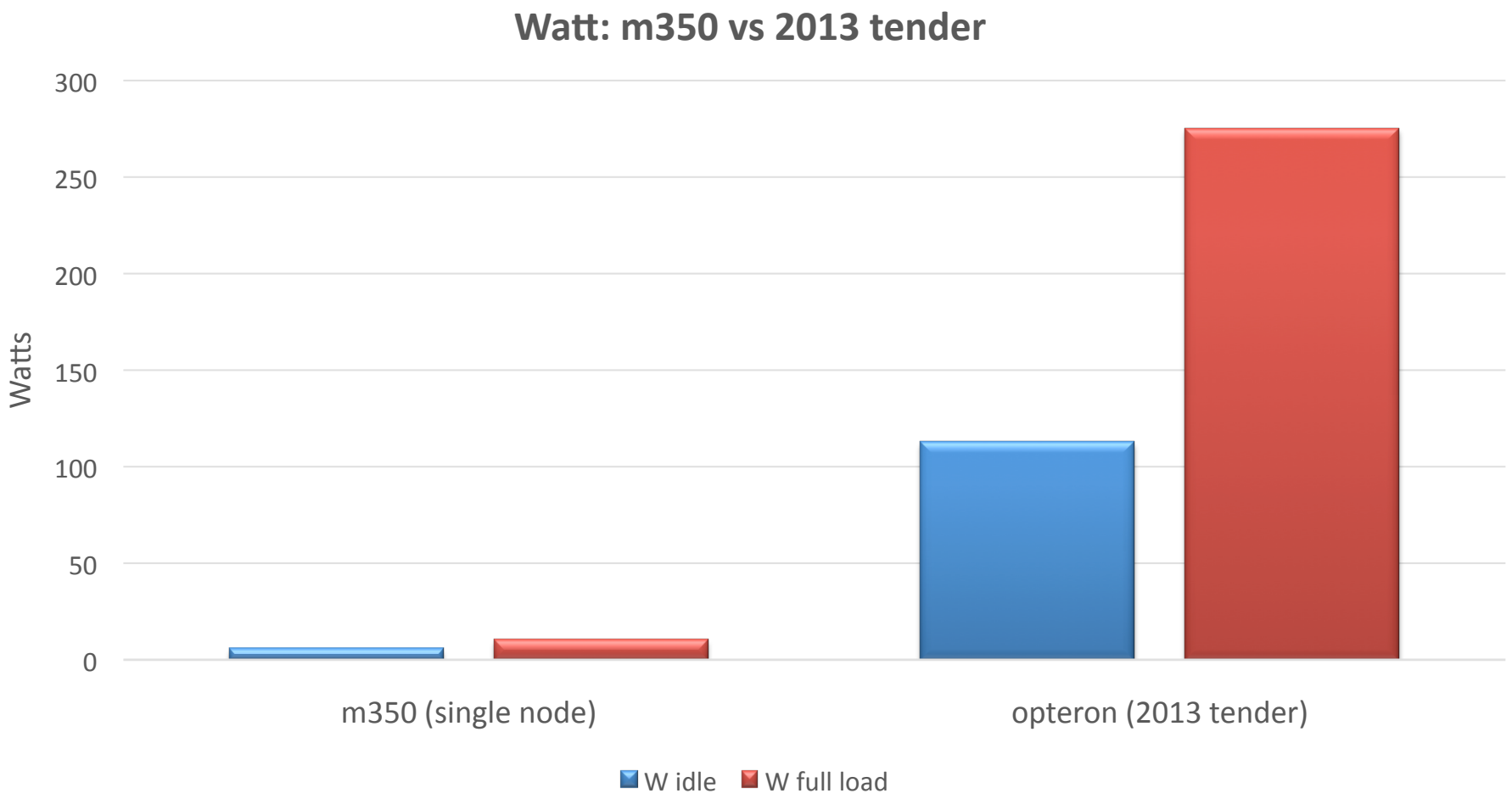
Power Consumption



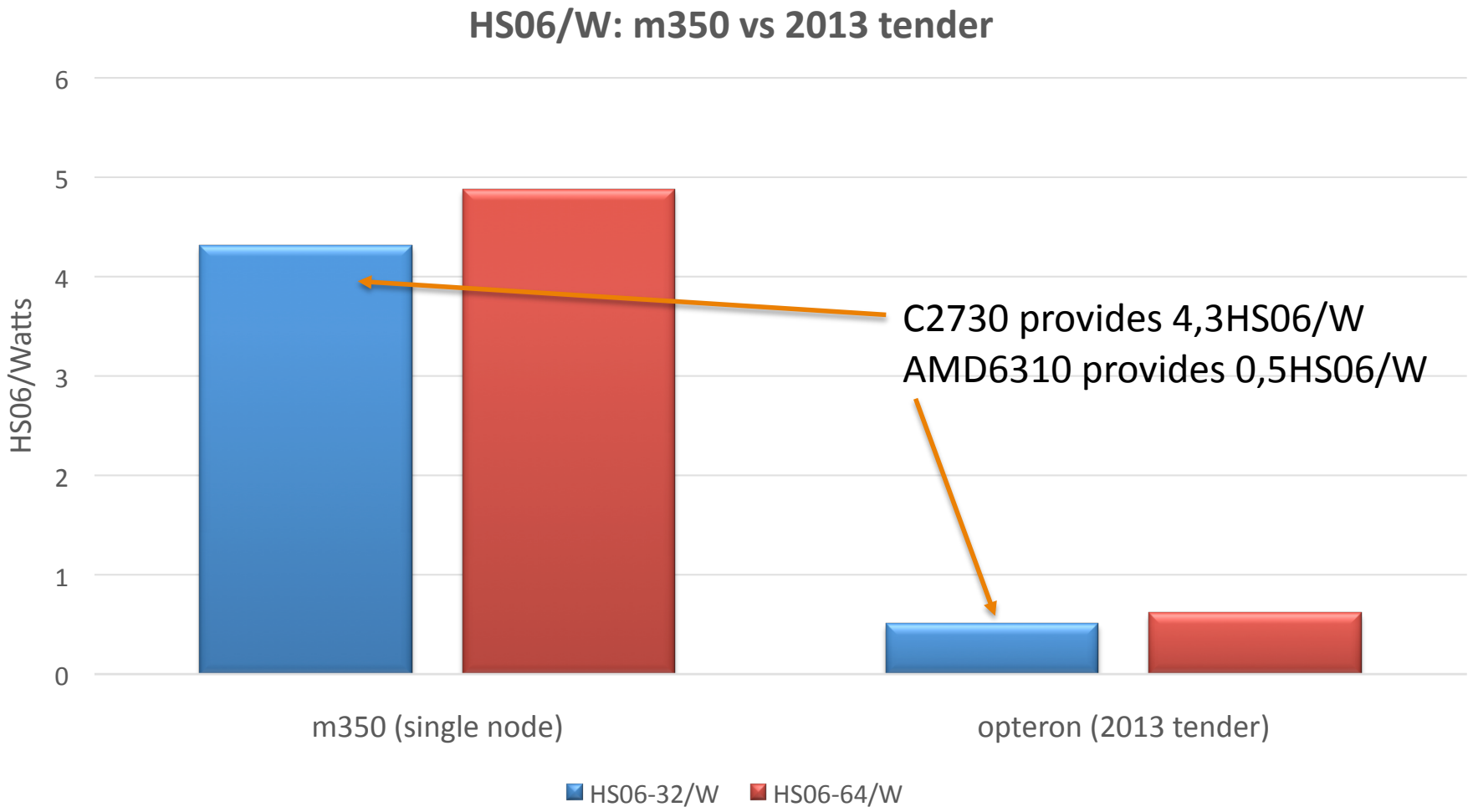
Avoton vs 2013 tender node



Avoton vs 2013 tender node

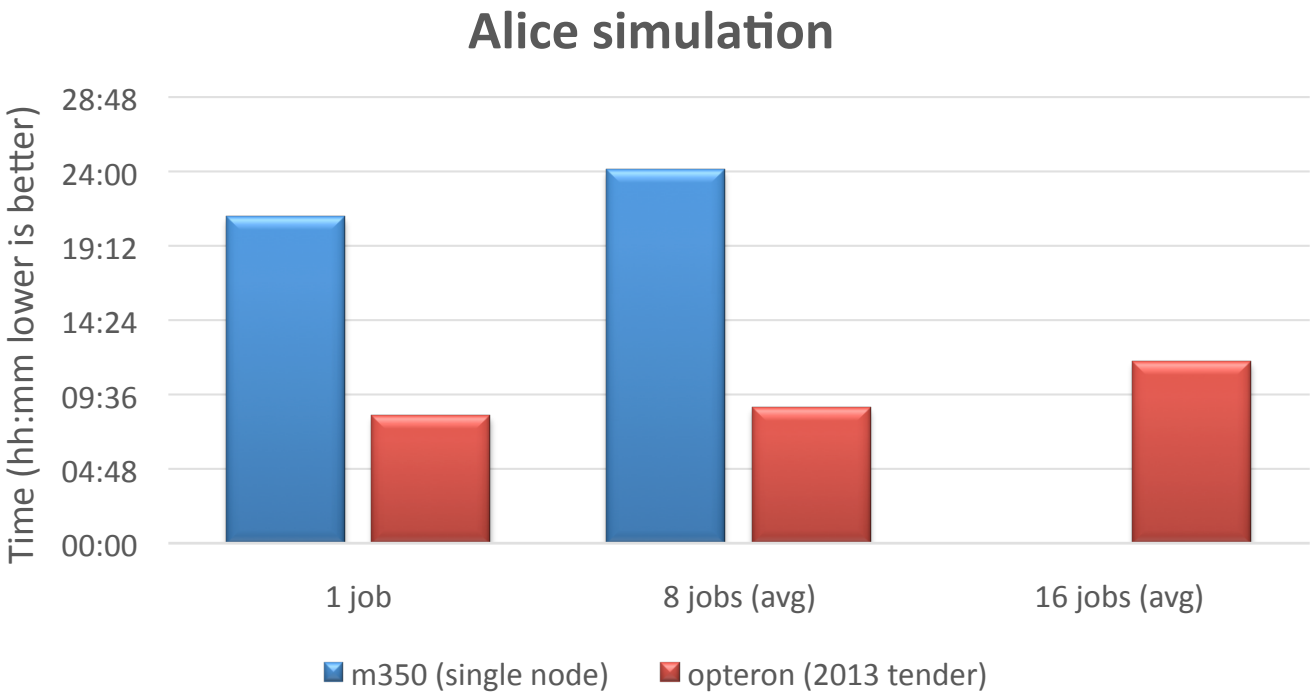


Avoton vs 2013 tender node



Evaluation carried out by Experiments

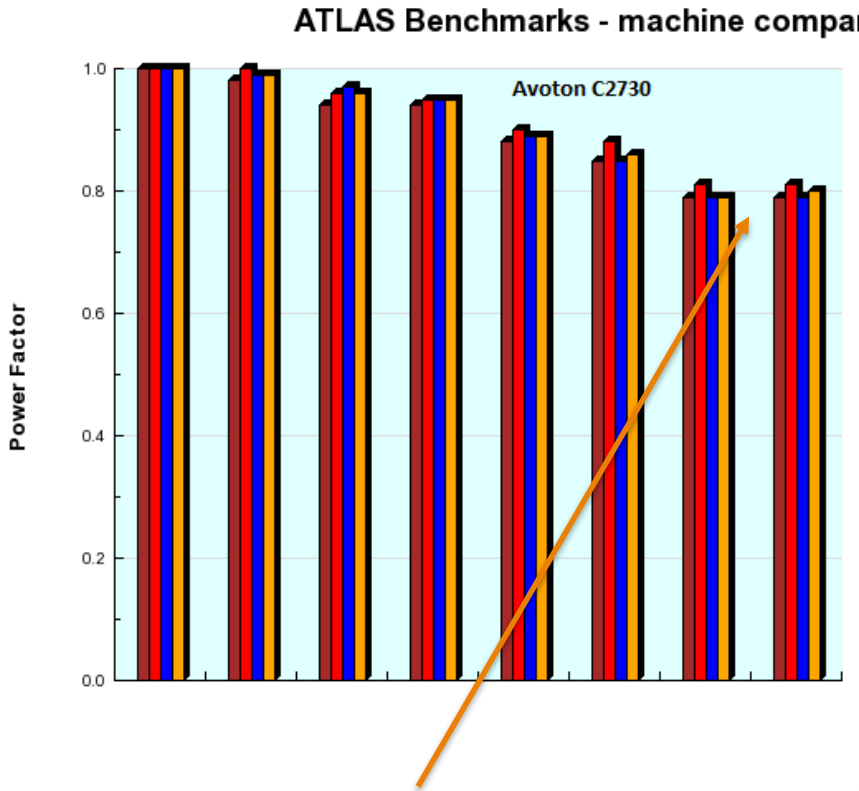
Alice (F. Noferini)

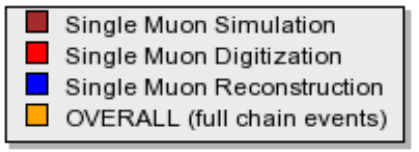


- Avoton 61% slower
- scales well

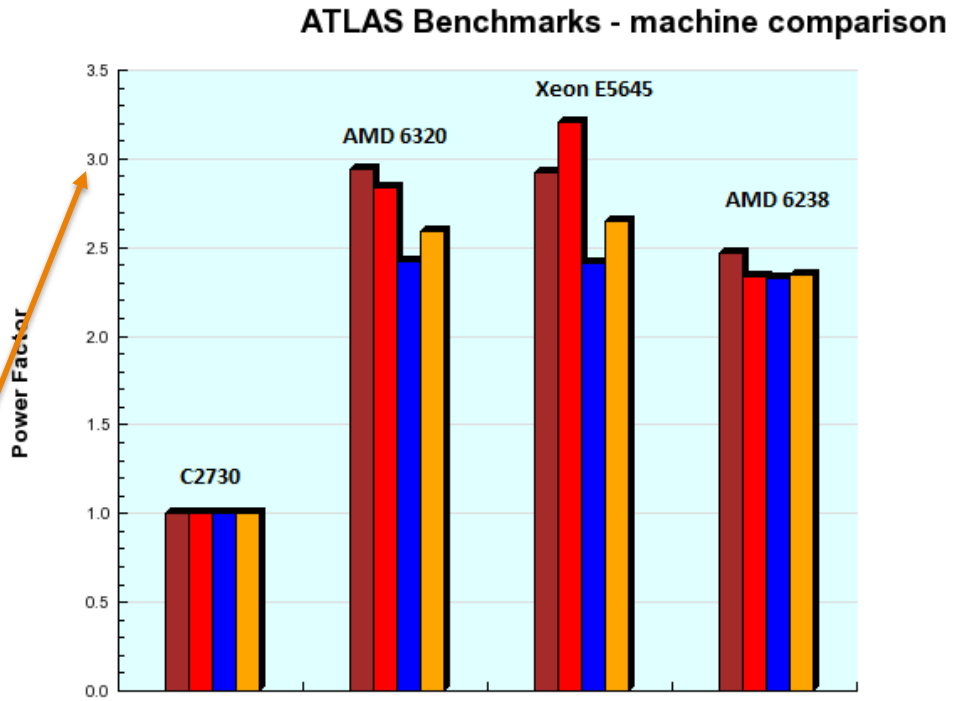
- 1 job is a simulation of 100 collisions proton-proton 7TeV (full detector simulation)
- Same seed for all the jobs

Atlas (A. De Salvo)



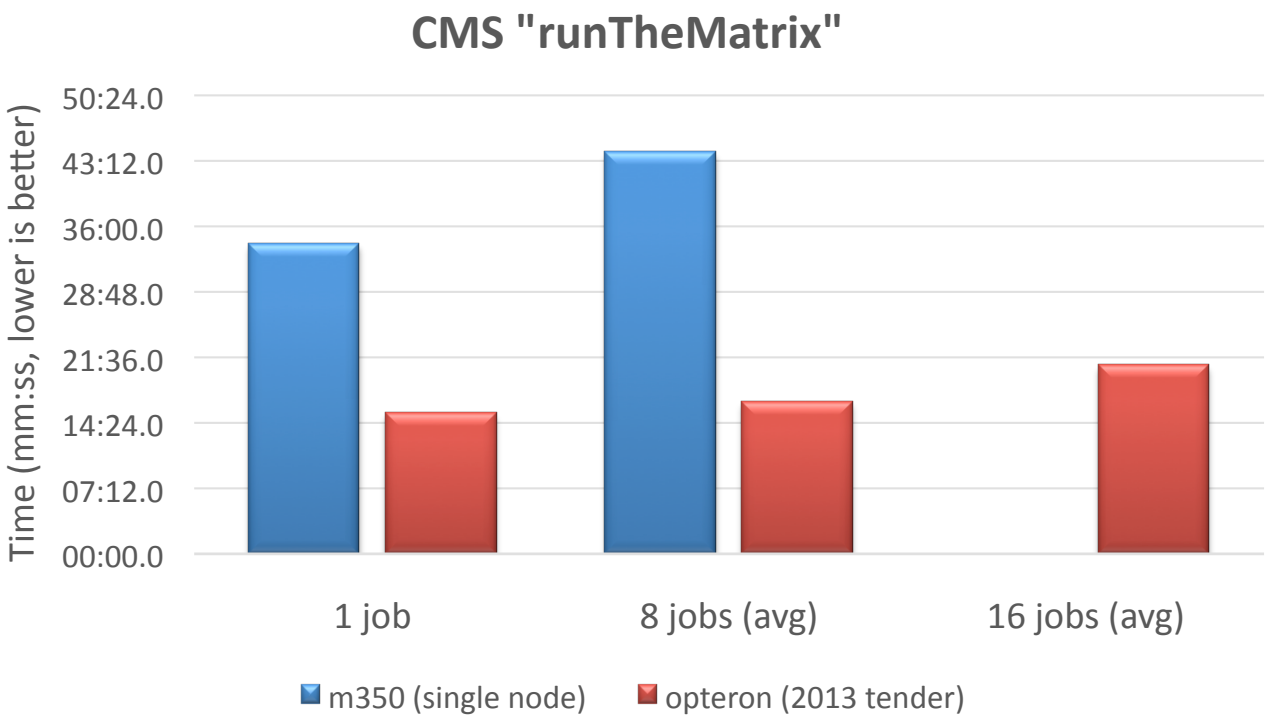


- Single muon “full chain” test



- Good scaling from 1 to 8 instances
- Avoton 3 times slower compared to atlas benchmark nodes

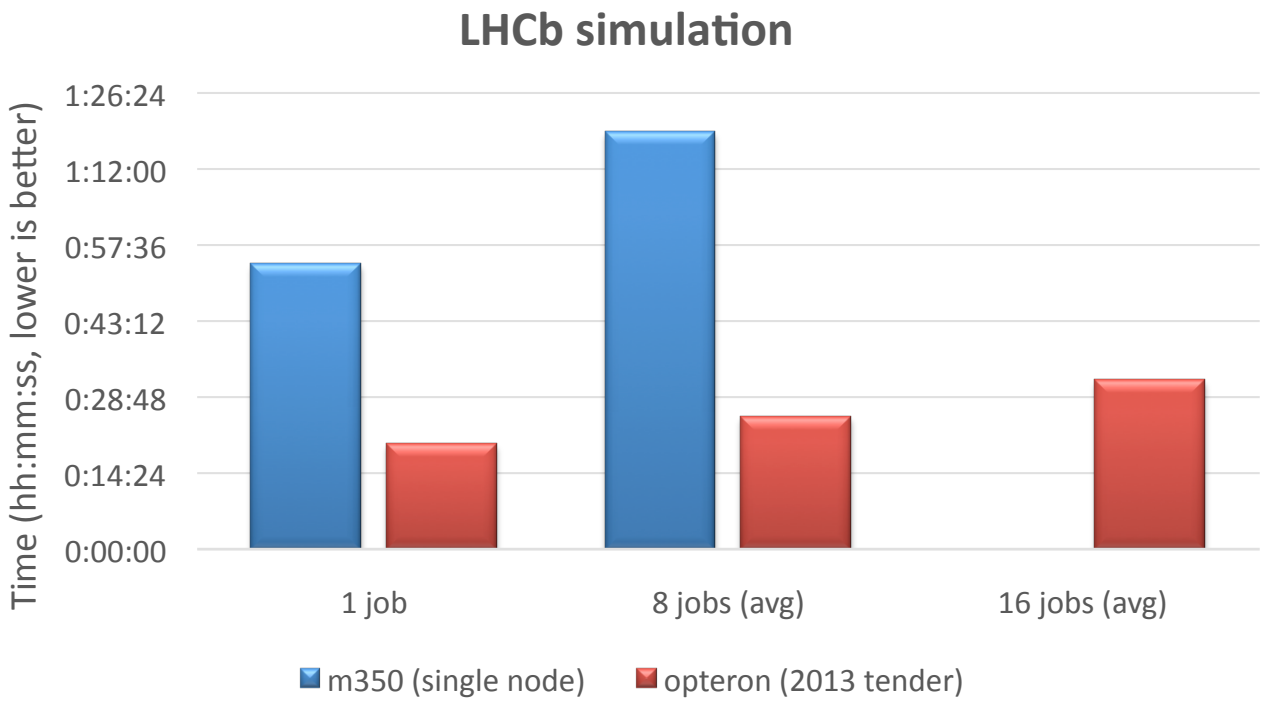
CMS (S. Taneja)



- 1 run is 55% slower
- 8 runs are 63% slower

- Completely CPU bounded
- Typical running time on an high-end x86_64 server is of the order of 15 min

LHCb (A. Falabella)



- 1 run is 63% slower
- 8 runs are 39% slower
- Still scaling well

- No significant I/O, mostly CPU bounded
- Cvmfs provides revelator status

Conclusions

PROs

- Switching to Avoton CPU is easy
  - No recompilation required, native x86_64 code
- CPU TDP 12W requires less cooling power
- 1U standard rack can host 73.710HS06 (little cabling both for power and network)
- Attractive if we think of “whole-node” and “multi-core jobs” paradigms
  - 8 cores per motherboard

CONs

- Not so powerful compared to “standard” server class CPUs
- Requires more motherboards to achieve the same computing power (approx. 3 times)
  - Implies greater human effort to keep all the nodes up & running
  - More hardware failures
  - Easier to reach the limit of batch system license
- M350 card requires an external storage to fulfill our requirements, this implies
  - Extra cost and rack space used for storage
  - Performance may be impacted by “slow” I/O

Final thought

- A simulation on 4 years scale shows the possibility to save 25% in TCO, compared to 2013 tender machines
  - Including higher hw costs
- We would like to have some feedback from WLCG community to better understand if a similar solution can be considered for a future tender

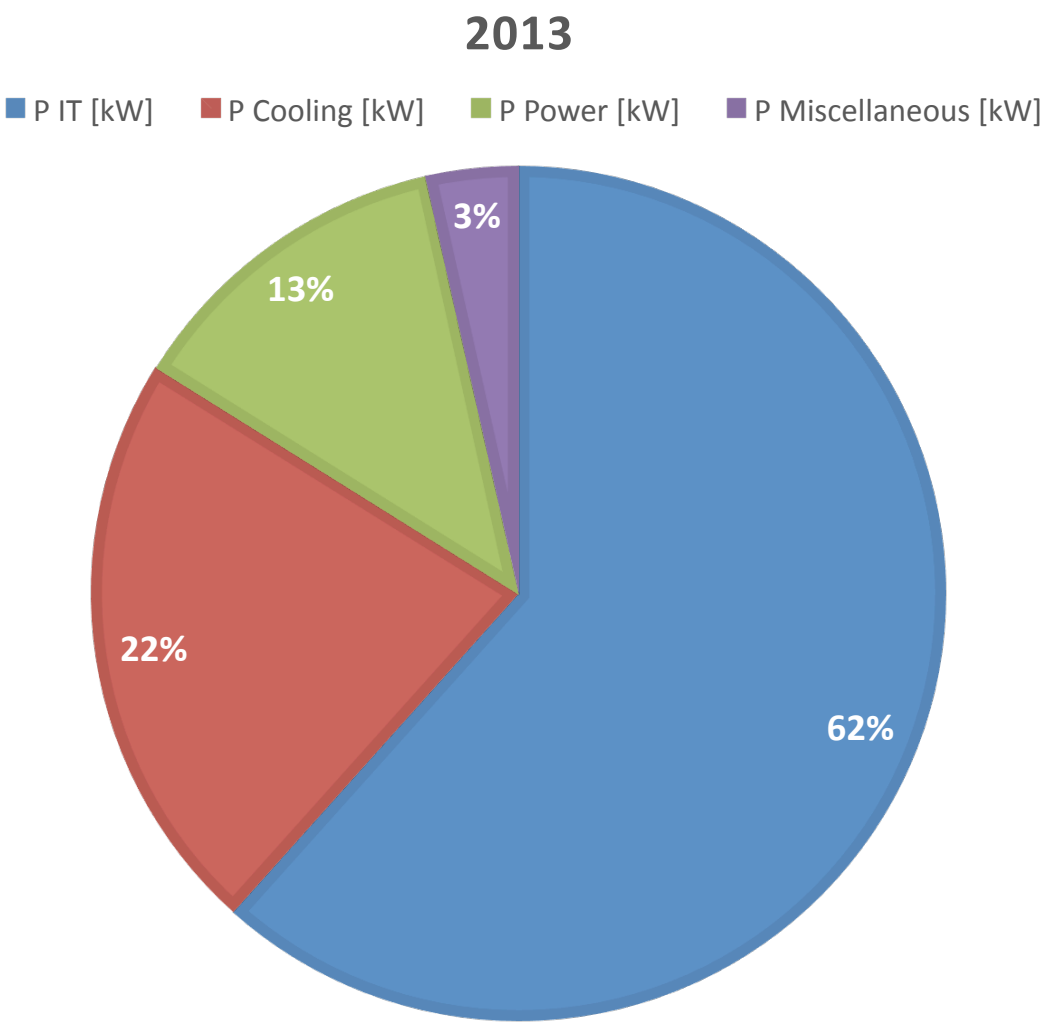
Backup slides

Power measurement

Voltech PM300

	Range	**Precision**
Voltage	From 2Vrms to 1000Vpk	±0.1%
Current	From 20mArms to 20Arms	±0.1%
Active power	From 0 to 1,999 GW	±0.2%
Cos(φ)	From 0.000 to ±1.000	±0.004
Frequence	From 5 Hz to 50kHz	0.2%

Partition of electric bill



Electricity vs HS06 per tender