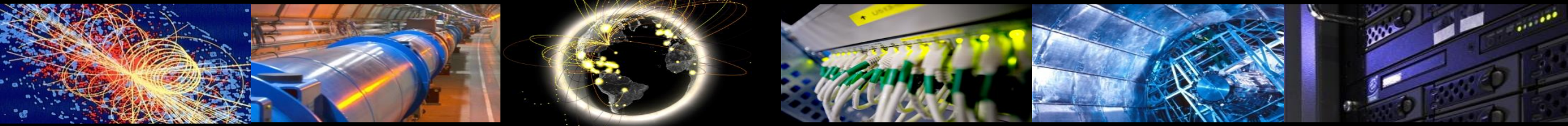
Update on CPU benchmarking with production jobs

Andrea Sciabà
Andrea Valassi

12 June 2018, pre-GDB



WLCG
Worldwide LHC Computing Grid

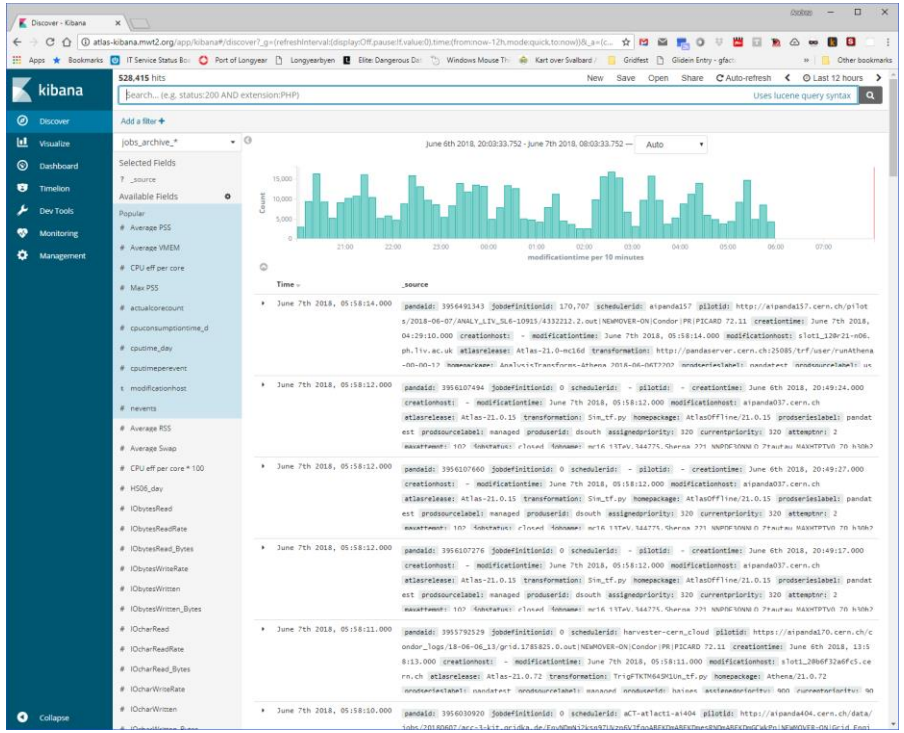
Motivation

- At the previous pre-GDB on benchmarking (February 2017) an attempt at using real jobs for “passive” benchmarking was presented
  - Test the idea of using different types of jobs to measure the approximate speed of CPUs (or sites)
  - See if the choice of the type of job matters or not
- What is the same?
  - The method
- What is new?
  - New CPU models
  - Better data in ElasticSearch (CPU model added for CMS jobs)
  - CMS Tier-0 analysis

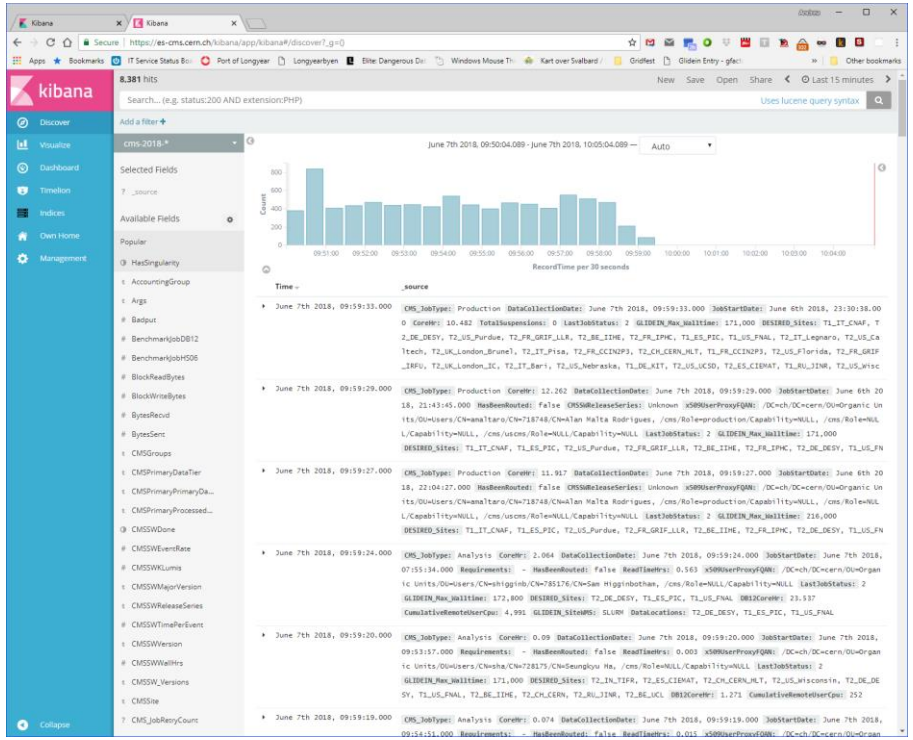
WLCG Worldwide LHC Computing Grid

Data sources

- ATLAS and CMS use ElasticSearch databases for job metadata
  - Kibana available for data exploration and fast prototyping of analyses
  - Very easy to make complex queries and data aggregations



atlas-kibana.mwt2.org



es-cms.cern.ch

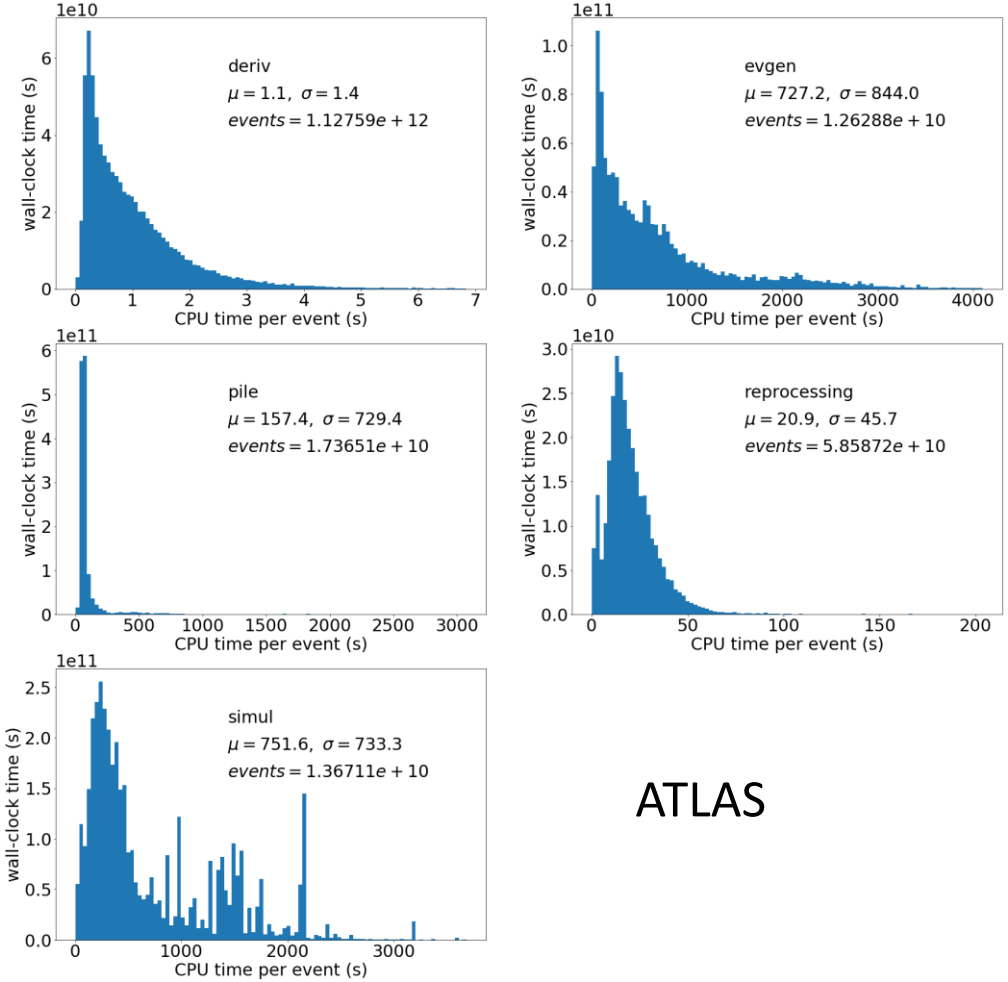
Fitting CPU speeds

- Basic assumptions
  - The CPU "speed" and the average CPU time/event of a given job are inversely proportional
  - All jobs in the same task are comparable, i.e. they run the very same executable but on different portions of a dataset
- Limitations
  - Limited information about the worker node: only the CPU model is known, but the SMT status is not, as the nature (virtual vs. physical) of the machine or any overprovisioning, etc.
  - There is no "absolute scale" for the CPU speed
    - Only speed ratios of different CPU models can be measured

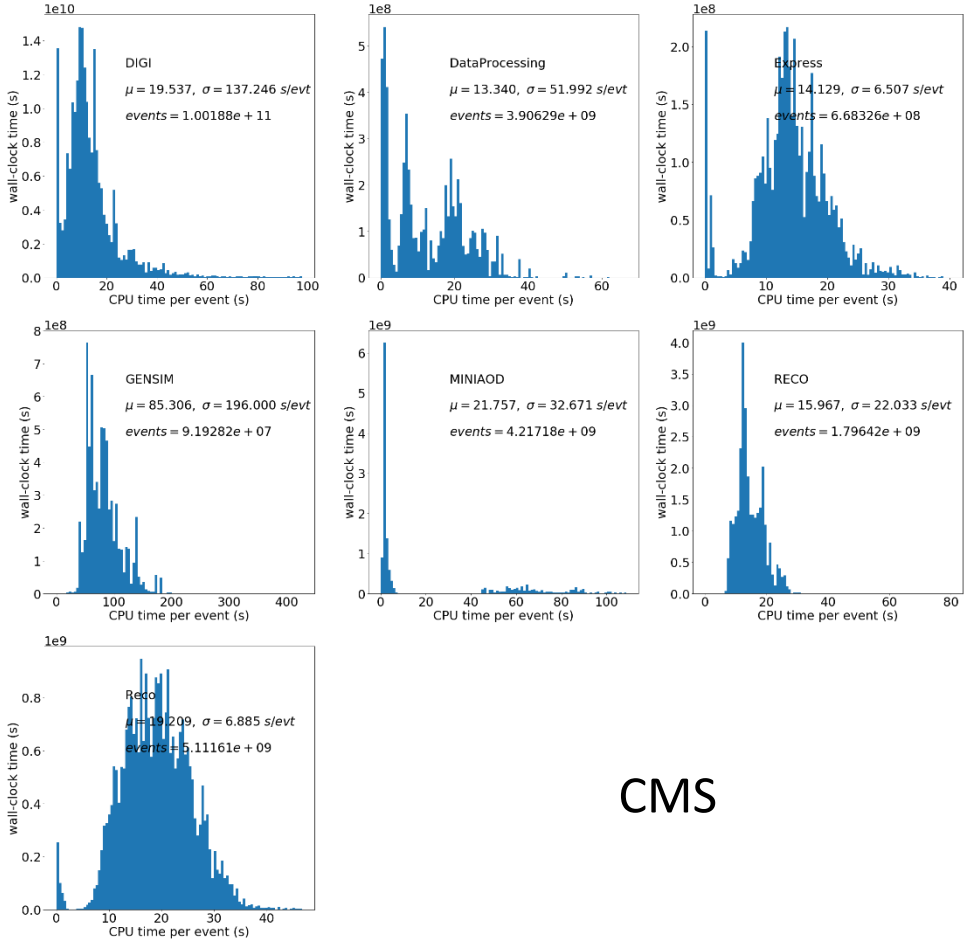
WLCG Worldwide LHC Computing Grid

CPU per event

- Different job types have very different CPU utilisations
  - CPU time is expected to be an equally good estimator in all cases



ATLAS



CMS

Analysis and fitting

- Each CPU model is associated to a “speed factor”
- All jobs in the same task should approximately have the same CPU time/event when this is normalized to the speed factor of the CPU they run on
- The fit of the speed factors is performed over a data sample of O(1 year)
  - Refer to the 2017 talk for more details on the procedure

WLCG Worldwide LHC Computing Grid

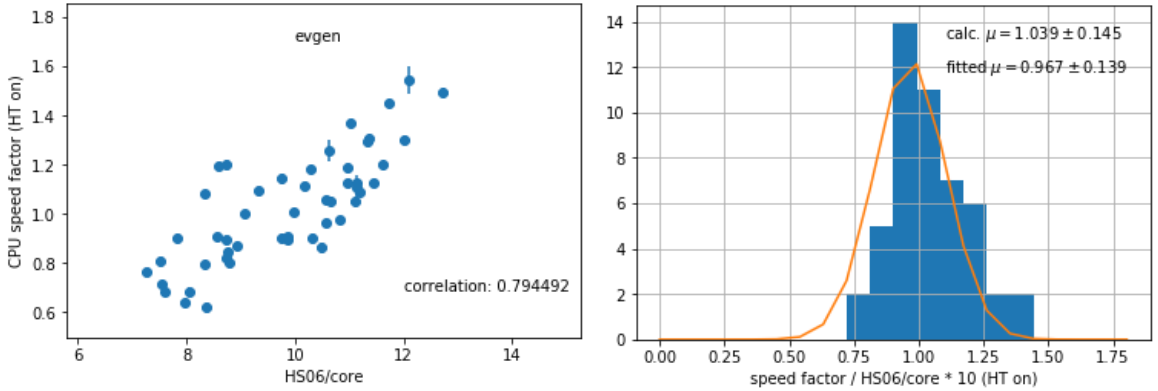
Comparing to HS06

- Used values published on HEPiX benchmarking WG page
  - http://w3.hepix.org/benchmarking/sl6-x86_64-gcc44.html
  - Only ~40 CPU models of the ~90 found in production jobs have a score
- Hyperthreading is sometimes enabled, sometimes not
  - HT on increases total score by ~25%
- HS06 per hardware thread heavily depends on HT status
  - HT off: HS06 per hardware thread ~2/1.25=1.6 larger than for HT on
  - Impossible to know for jobs if the WN had HT on or off
    - Large systematic uncertainty!

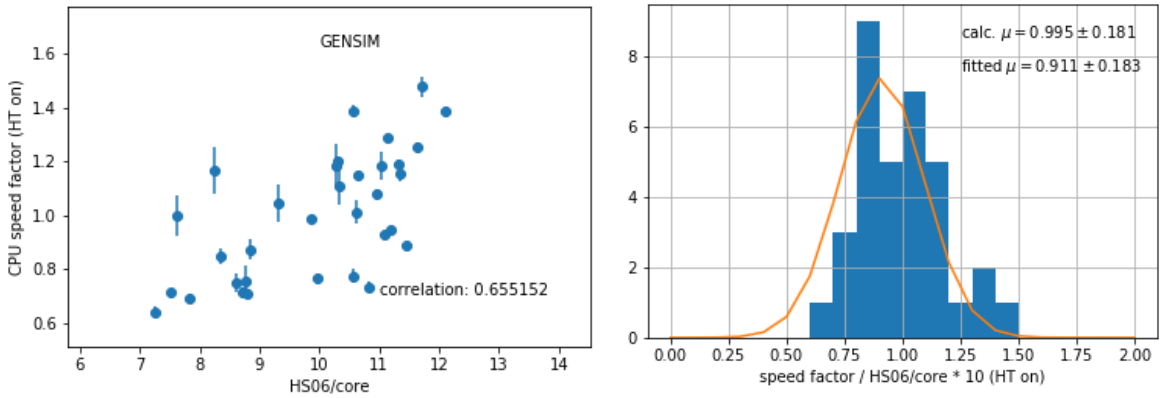
WLCG Worldwide LHC Computing Grid

Effect of HS06 rescaling

- Rescaling by HS06/core reduces the spread of the speed factor distribution
  - Scenarios assuming HT off or on
  - Several systematic uncertainties
- Correlation value slightly favours the scenario with HT on
- Precision of the passive benchmarking is around 15%
  - Mostly all systematic error



ATLAS



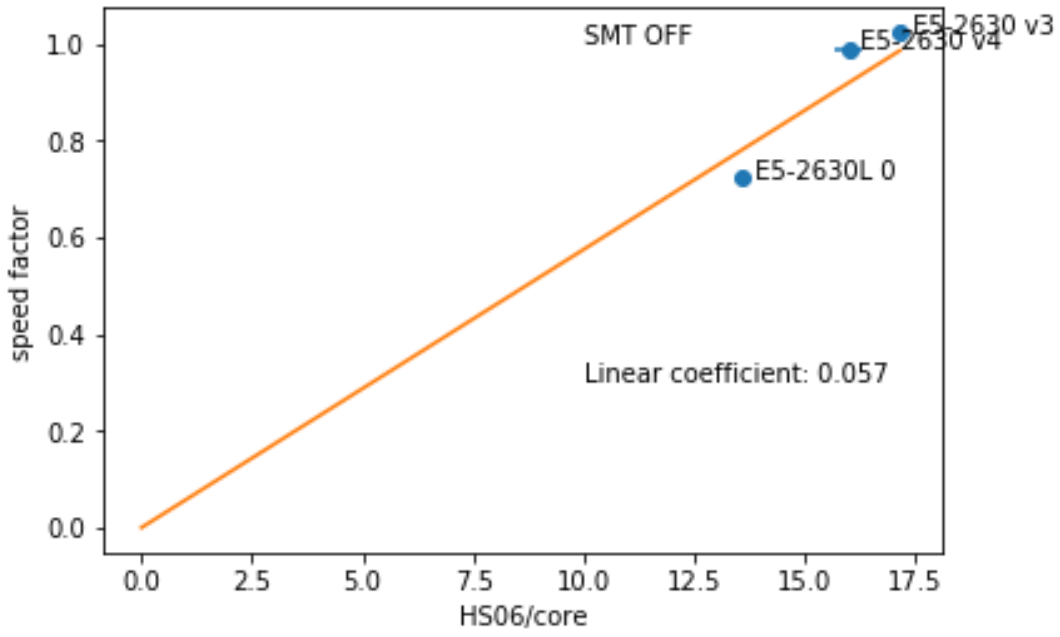
CMS

WLCG Worldwide LHC Computing Grid

ATLAS Tier-0 case

- The analysis was applied to jobs run at the ATLAS Tier-0
  - June 2017 – June 2018: SMT off
- Only three CPU models are found to be used in this period

Model name	Architecture	Physical cores
Xeon E5-2630 v3 2.40 GHz	Haswell	8
Xeon E5-2630 v4 2.20 GHz	Broadwell	10
Xeon E5-2630L 0 @ 2.00GHz	Sandy Bridge	6



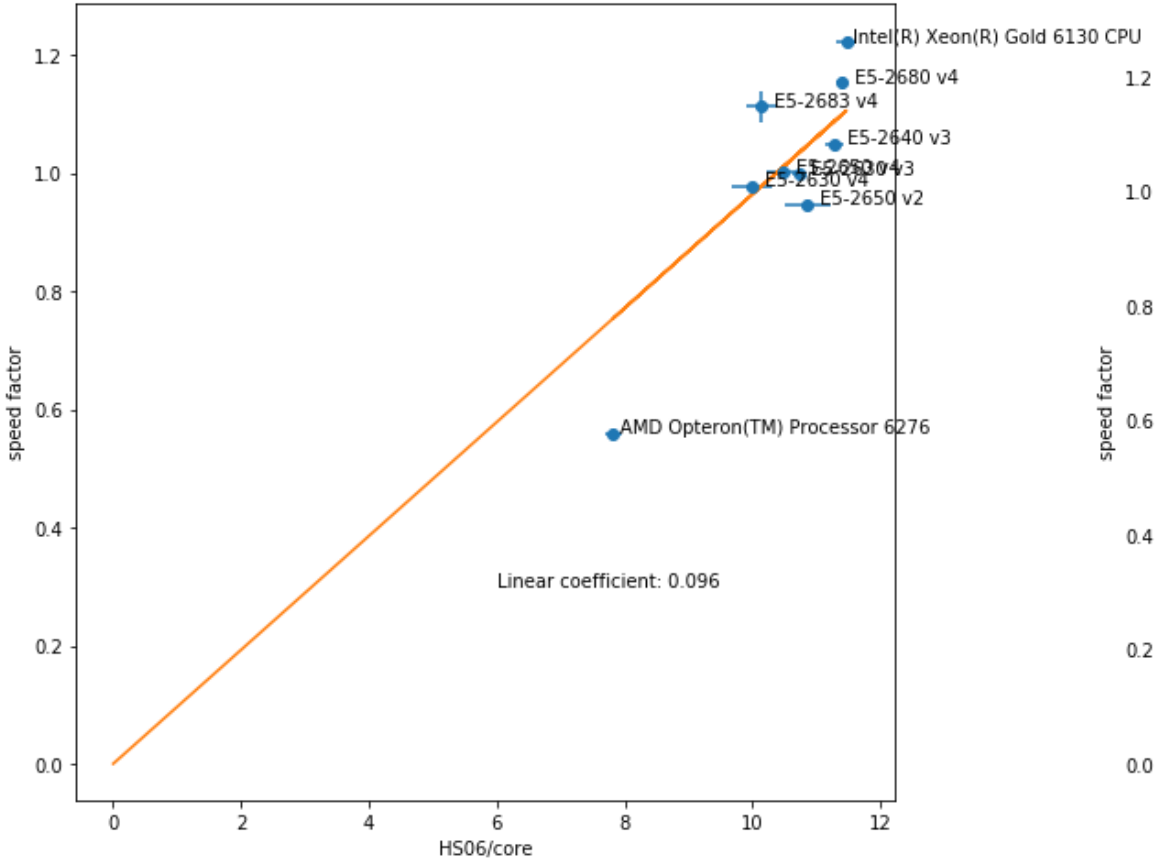
- HS06 scores taken from the CERN procurement database
- Sandy Bridge slower than HS06 would predict
  - Same conclusions as in March 2017

CMS Tier-0 case (new!)

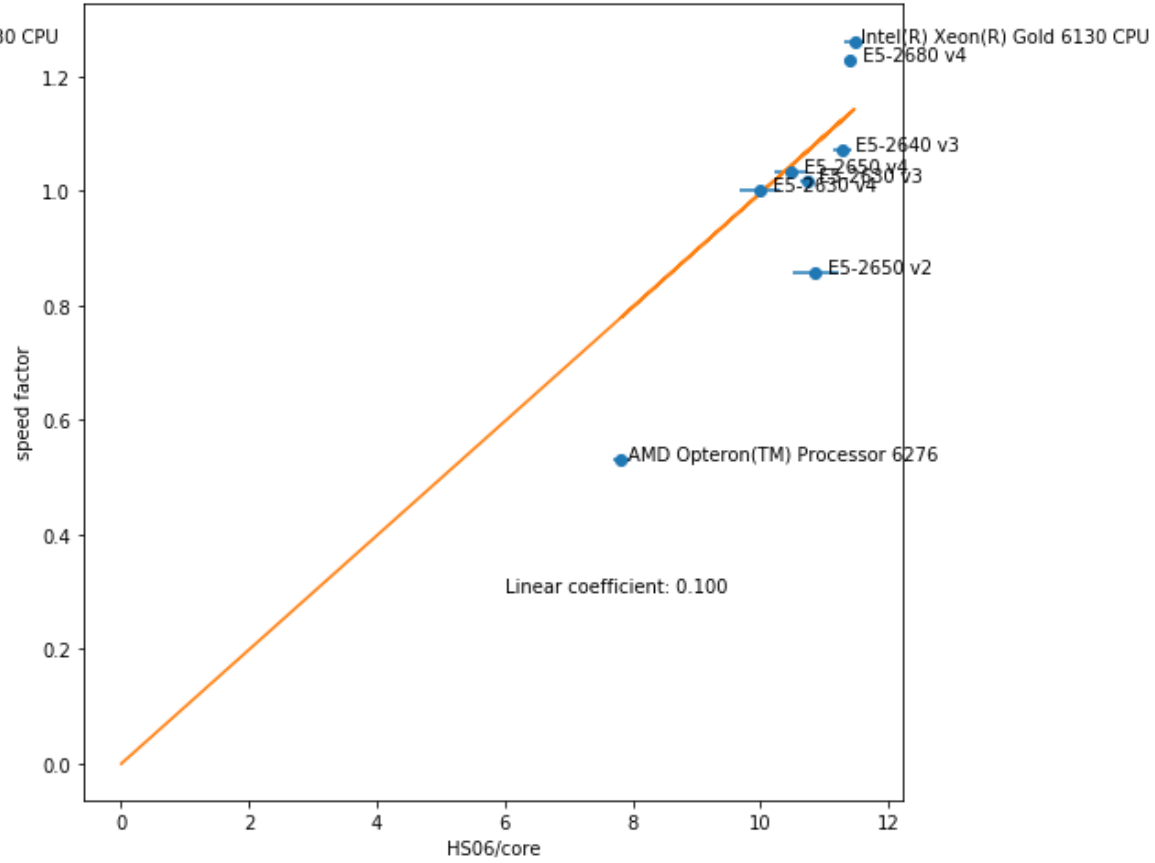
- The analysis was applied to jobs run at the CMS Tier-0
  - January 2018 – June 2018: SMT on
- Several CPU models used
- Results almost identical when using DIGI vs. RECO jobs (see next slide)
  - Broadwell and Skylake perform better than the rest for the same HS06 score

Model name	Architecture	Physical cores
Xeon E5-2630 v3 2.40 GHz	Haswell	8
Xeon E5-2630 v4 2.20 GHz	Broadwell	10
Xeon E5-2630L 0 @ 2.00GHz	Sandy Bridge	6
Intel(R) Xeon(R) CPU E5-2640 v3	Haswell	8
Intel(R) Xeon(R) CPU E5-2683 v4	Broadwell	16
Intel(R) Xeon(R) CPU E5-2680 v4	Broadwell	14
Intel(R) Xeon(R) CPU E5-2650 v4	Broadwell	12
Intel(R) Xeon(R) Gold 6130 CPU	Skylake	16
Intel(R) Xeon(R) CPU E5-2650 v2	Ivy Bridge	8
AMD Opteron(TM) Processor 6276	Bulldozer	16

CMS Tier-0 speed factors vs HS06



DIGI jobs



RECO jobs

Conclusions

- Passive benchmarking with Grid jobs can be used to measure (with a relatively poor precision) the speed of CPUs
  - The type of job does not matter much
  - Precision is around 15%
    - Because the same CPU model can lead to different performance depending on site configuration
- If applied to a single site (e.g. CERN) it can be much more precise
  - It can be correlated to HS06 to show for example differences in scaling for different architectures – much more useful

WLCG Worldwide LHC Computing Grid