HEPscore23 Workload Analysis

Ladislav Ondris

7 March 2023

Workload Evolution Study

Workload Evolution

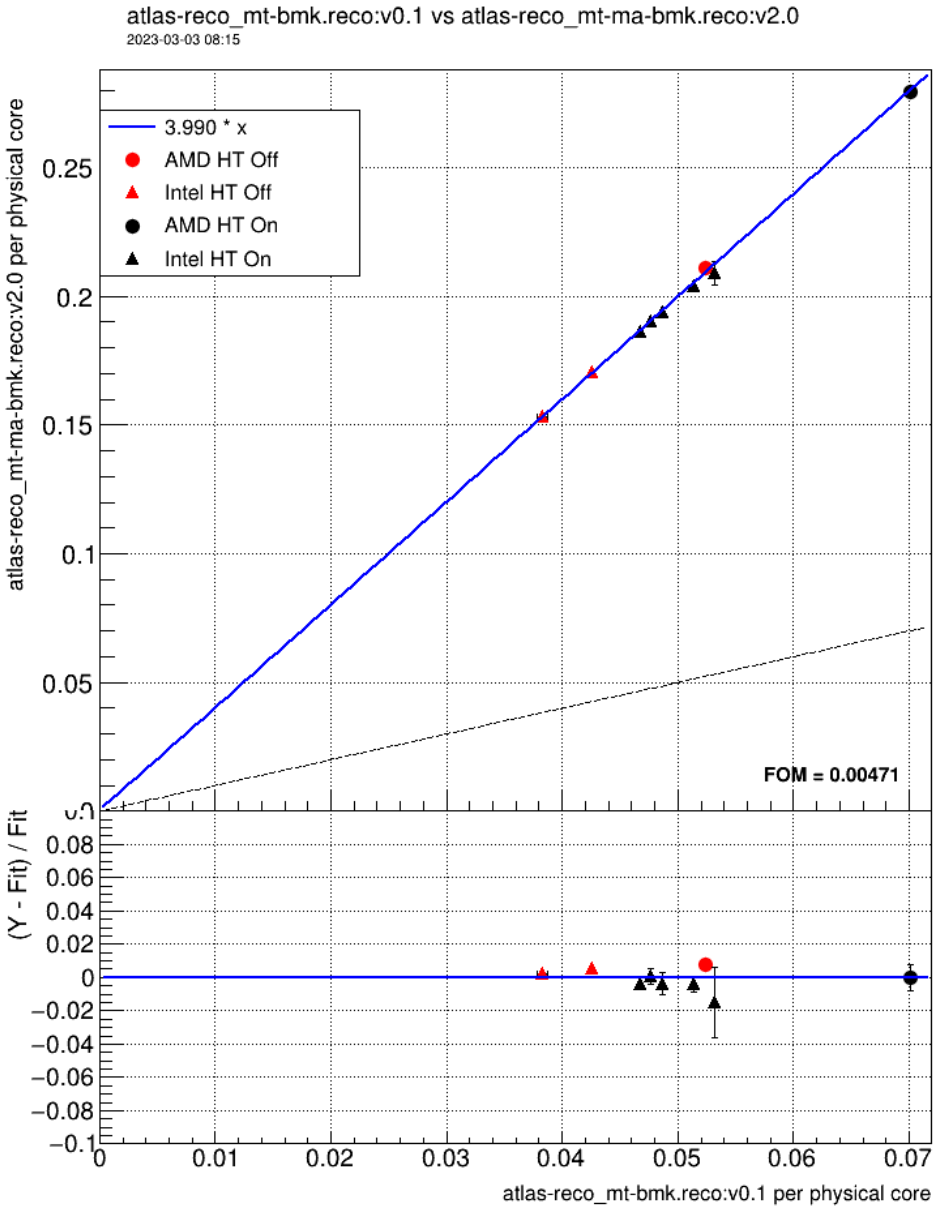
- Workloads have changed since the HEPscore workshop in Sep 2022
  - New software versions for all the applications inside the WLs
  - Support for x86 and ARM (multi-architecture)
- How do the scores (event throughput) scale for the testbed machines?
- Correlation plots are presented in the next slides
  - WL_m_i (v_x) vs. WL_m_j (v_y)
  - E.g., CMS gen-sim (v0.6) vs. CMS gen-sim-ma (v1.0)

Name	Version	Metric	Count
cms-digi-run3-bmk	v0.4	digi	533
cms-digi-run3-bmk	v0.6	digi	2254
cms-digi-run3-bmk	v0.7	digi	3
cms-digi-run3-ma-bmk	v1.0	digi	1882

ATLAS reco

- Comparison between old and new
  - Points represent CPUs
  - The goodness of fit measured by Figure of merit (FOM)
  - Slope of 1 represented by the gray dashed line
  - Error bars show the standard deviation
- Event throughput is 4x the old one
  - Moved from CPU time to Wall time
- Relative discrepancy $\lesssim$ 1.5%

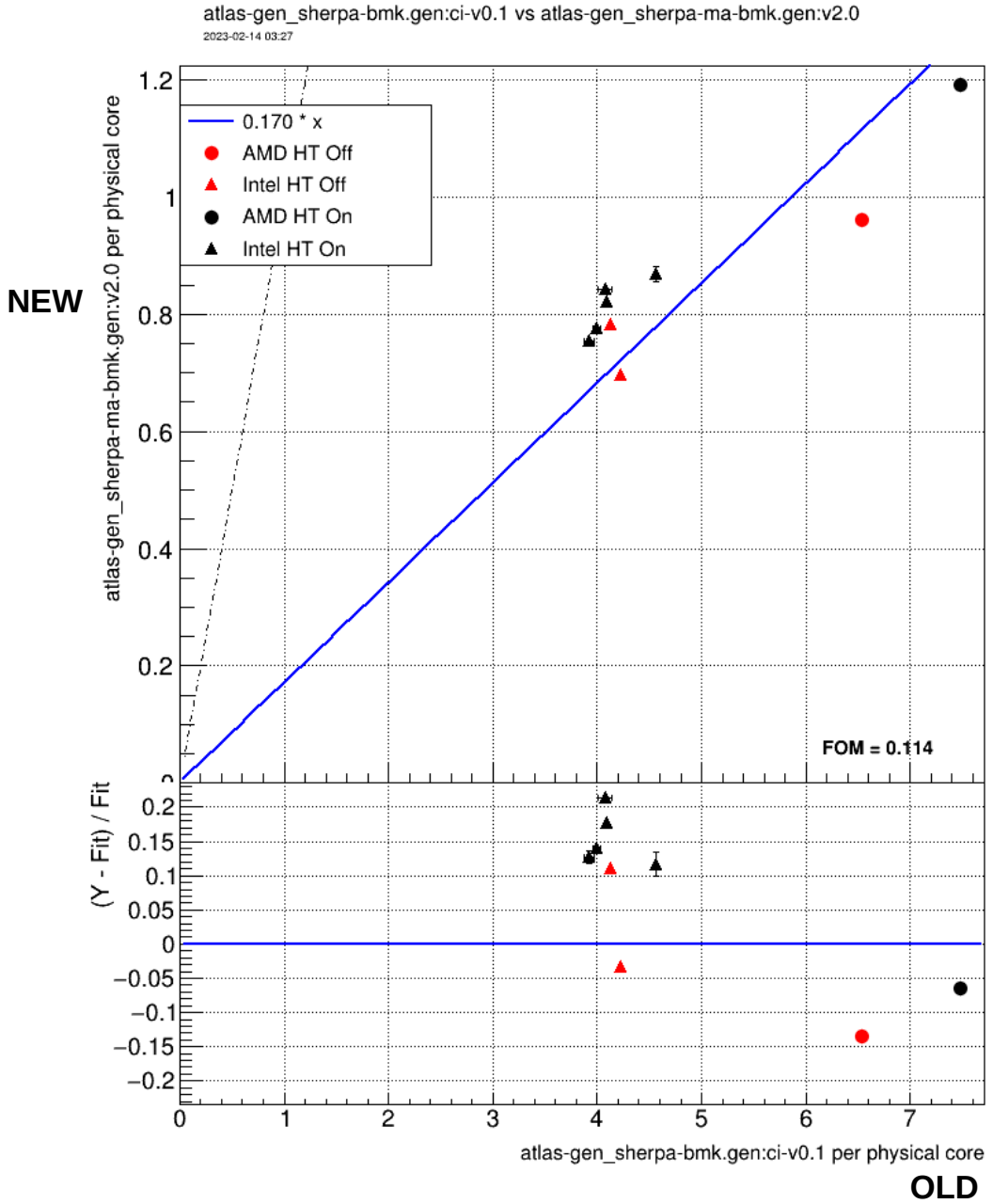
NEW



OLD

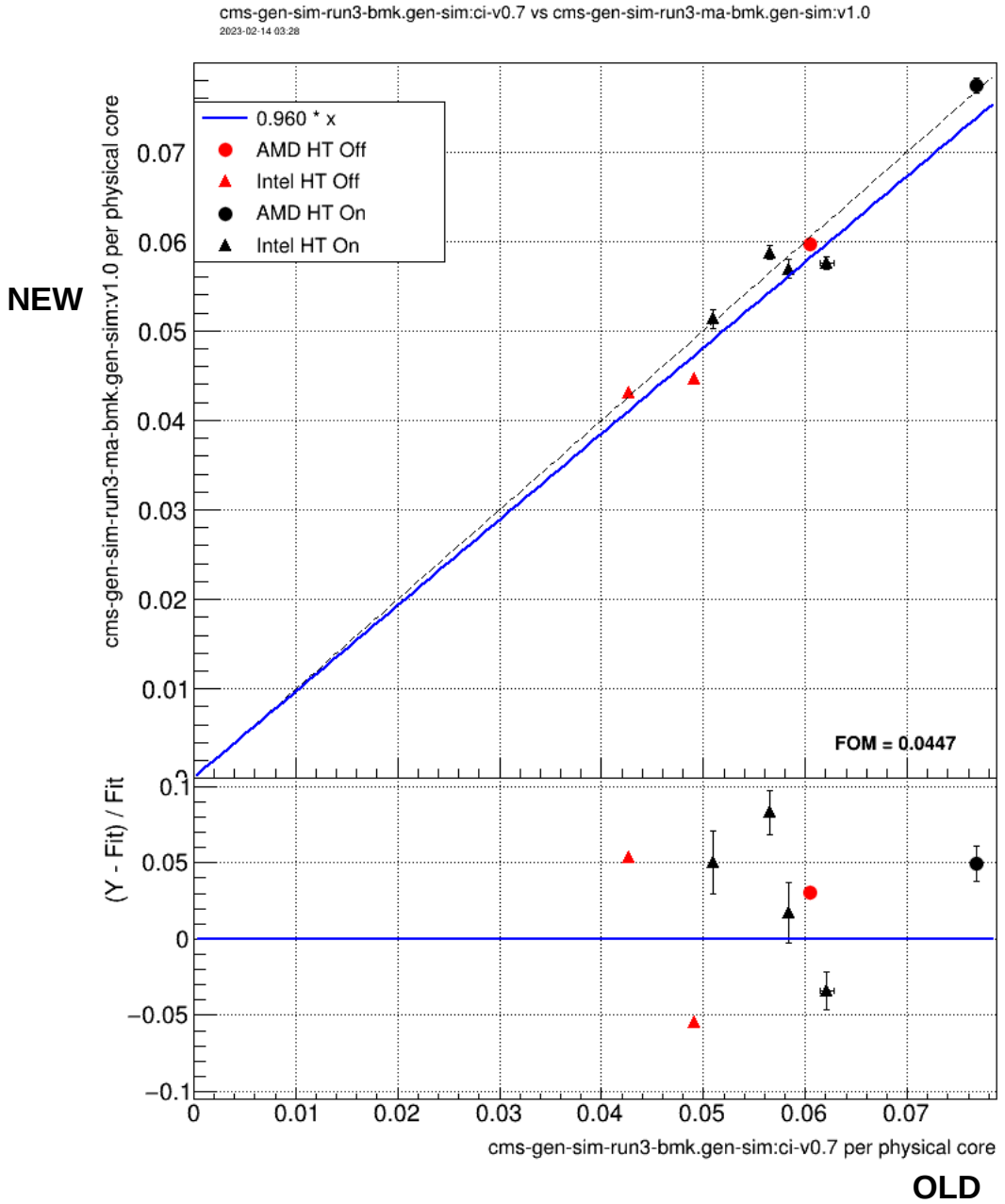
ATLAS gen_sherpa

- Reminder: a bug in the old workload (used the wrong duration for score computation)
- New event throughput ×0.17 the old one
- Relative discrepancy ≲ 20%



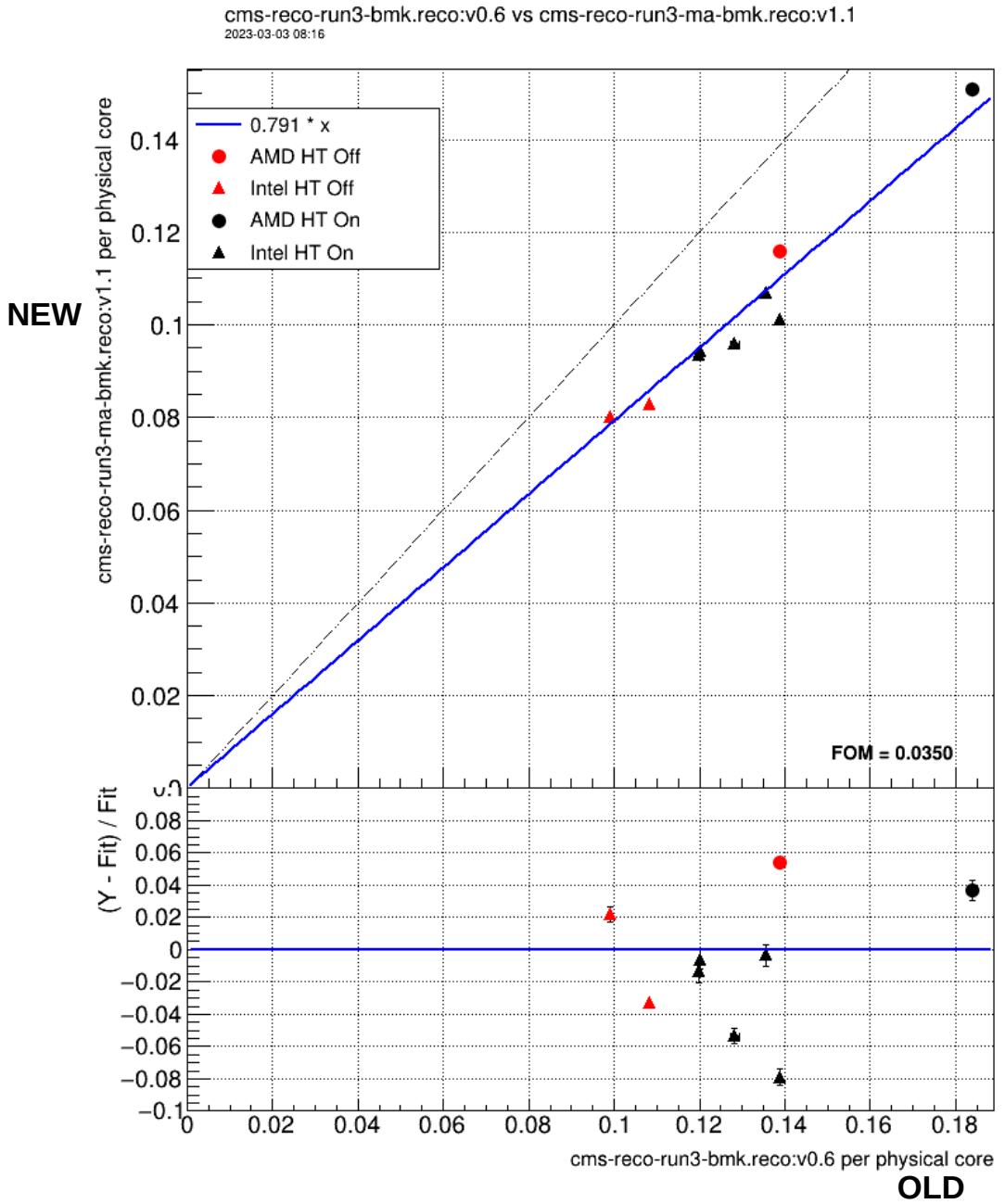
CMS gen-sim

- Same event throughput
- Relative discrepancy ≲ 8%



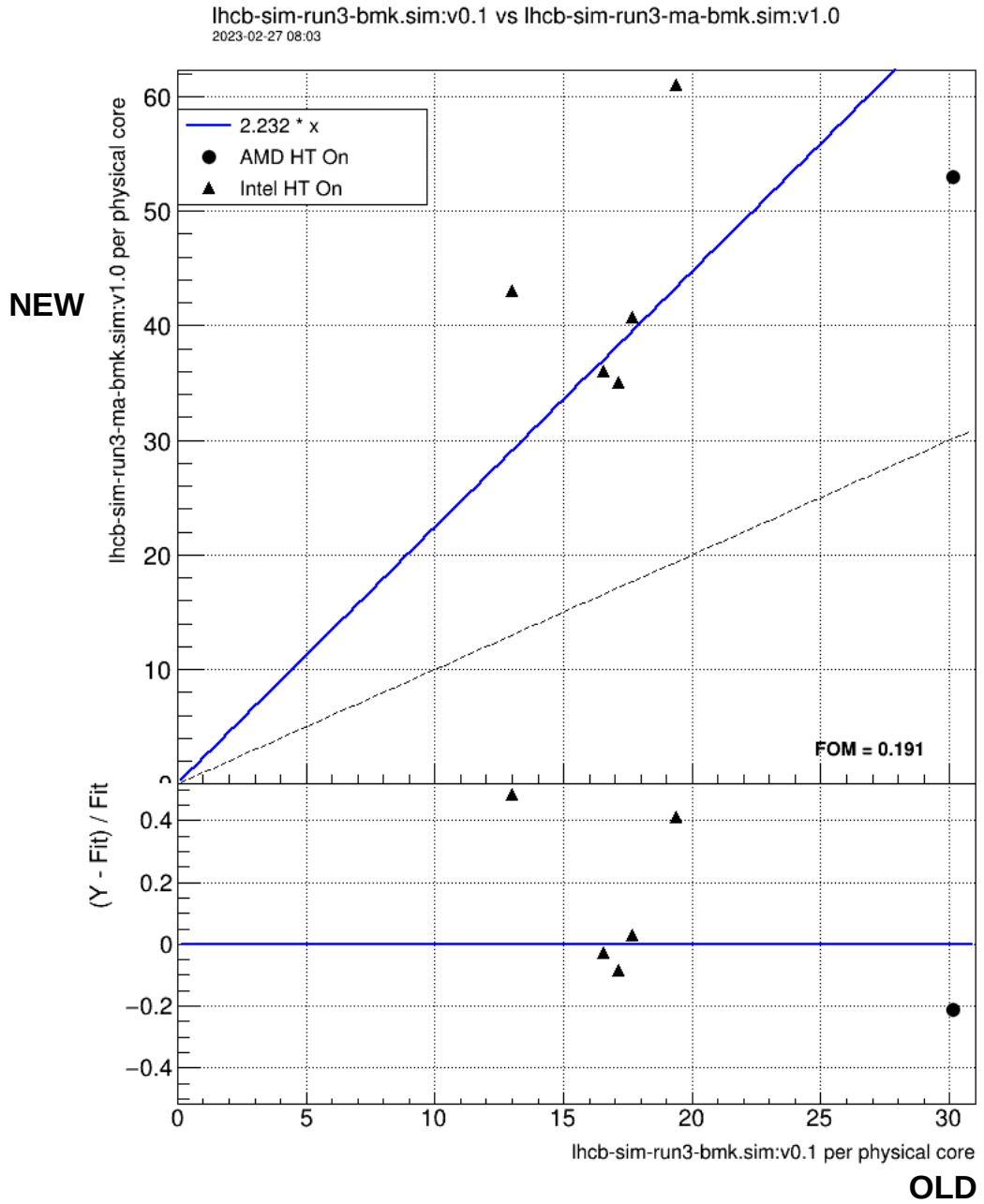
CMS reco

- Lower event throughput
  - Discussed with CMS experts
  - Software undergoes development
- Relative discrepancy ≲ 8%



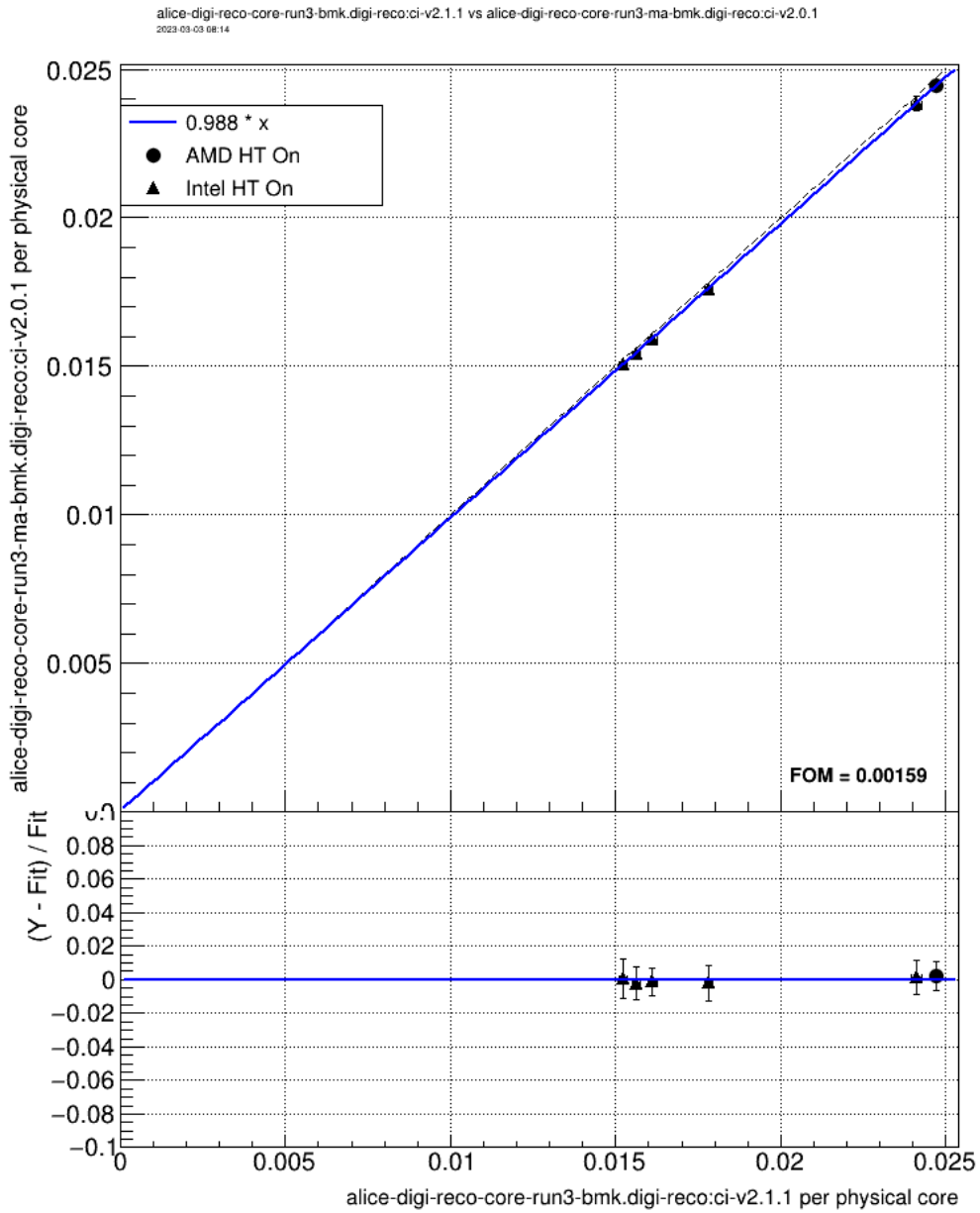
LHCb sim

- Event throughput 2× of the old one
  - Significant software improvements
- Relative discrepancy ≲ 44%



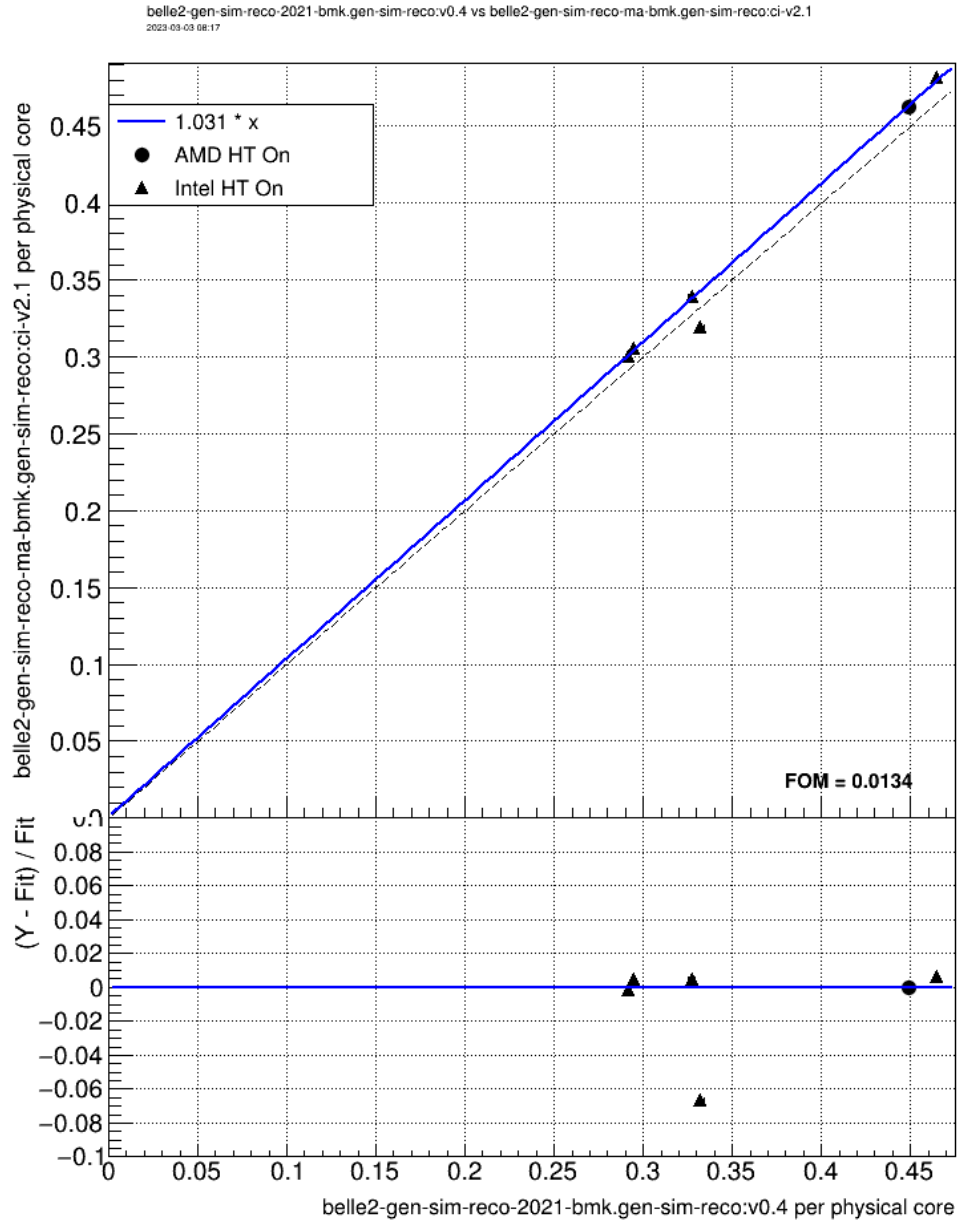
ALICE digi-reco

- Both workloads are recent
- Same event throughput
- Relative discrepancy $\lesssim 0.3\%$



Belle2 gen-sim-reco

- Same event throughput
- Relative discrepancy ≲ 7%



NEW

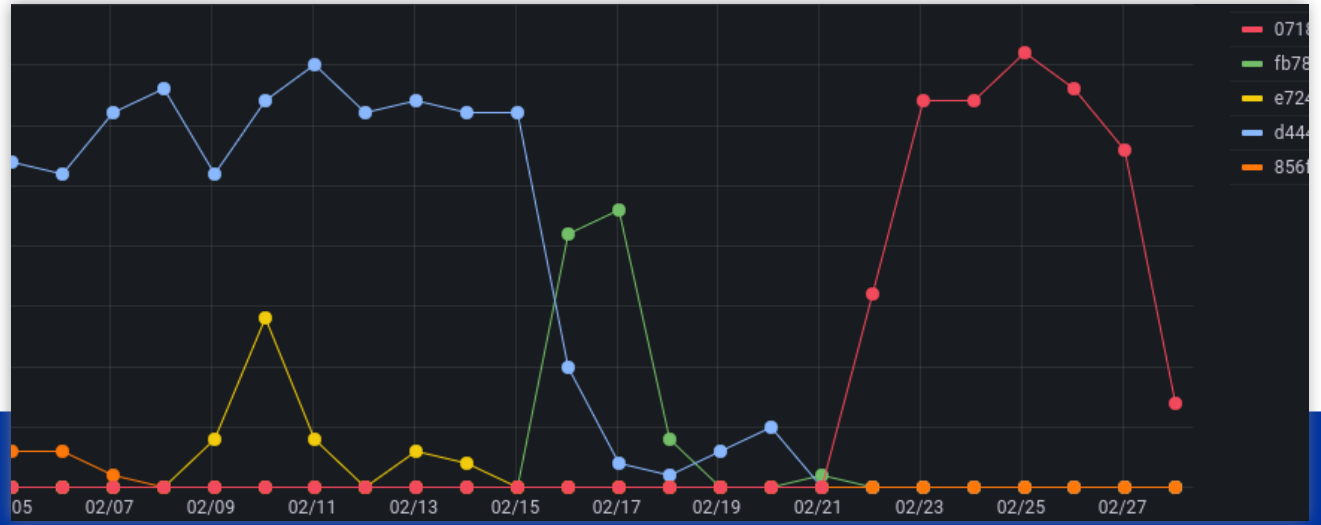
OLD

Impact of Changes in HEPscore Composition

HEPscore Configurations

- Multiple HS23 configurations built in the past months
  - All multi-architecture
  - Some only x86
  - Missing some WLs under development
- Tracking configurations by their hash

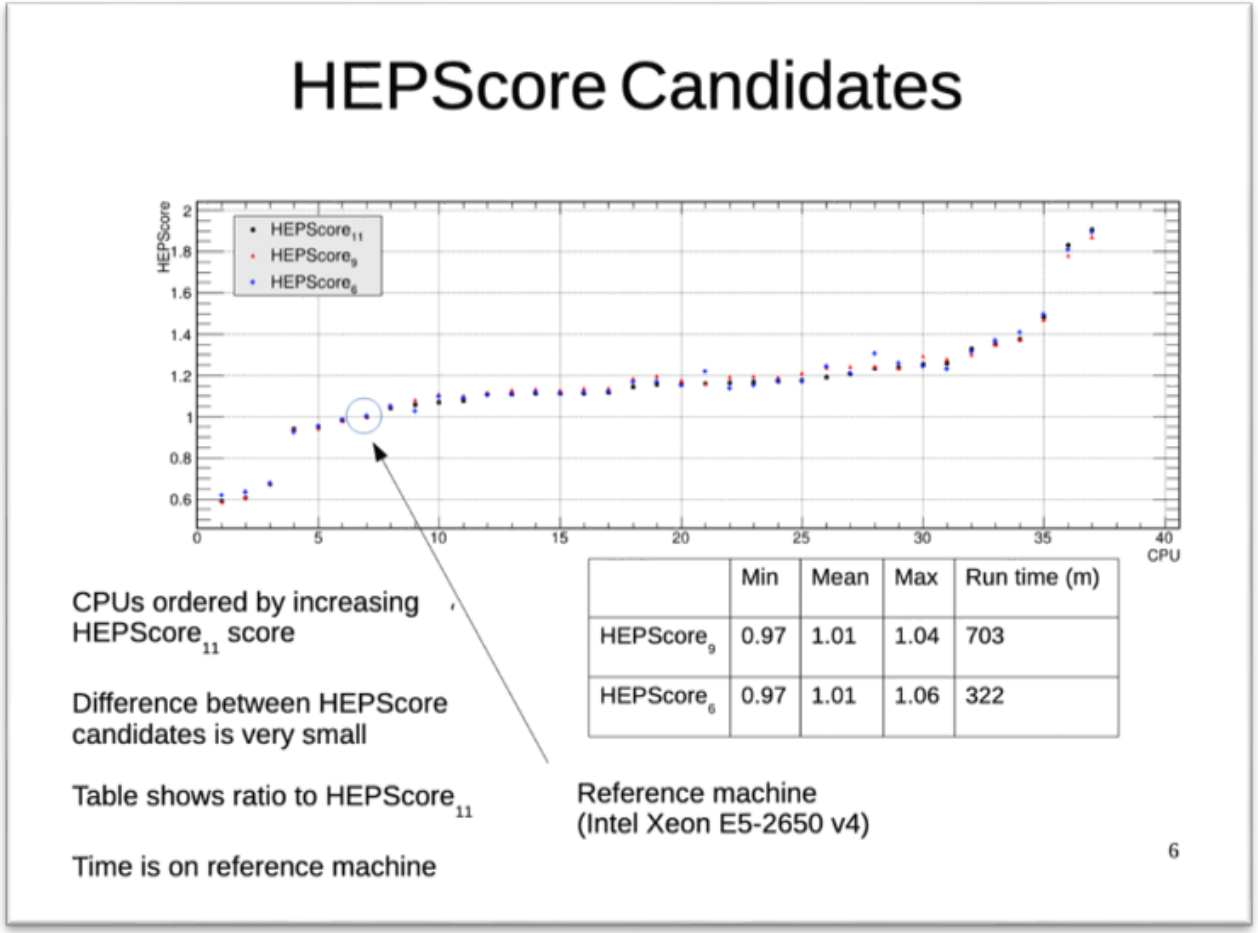
Hash	Description
071	Final HS23 (all multi-architecture)
fb7	Without LHCb (only six workloads)
d44	Without LHCb (only six workloads)
e72	ALICE, Belle2 and LHCb are not MA
856	Without ALICE; Belle2 and LHCb not MA



Effects of HEPscore Configurations

Presented by Tristan Sullivan

- Already proven that the differences among candidates can be small
  - Limited effect on CPUs ordering

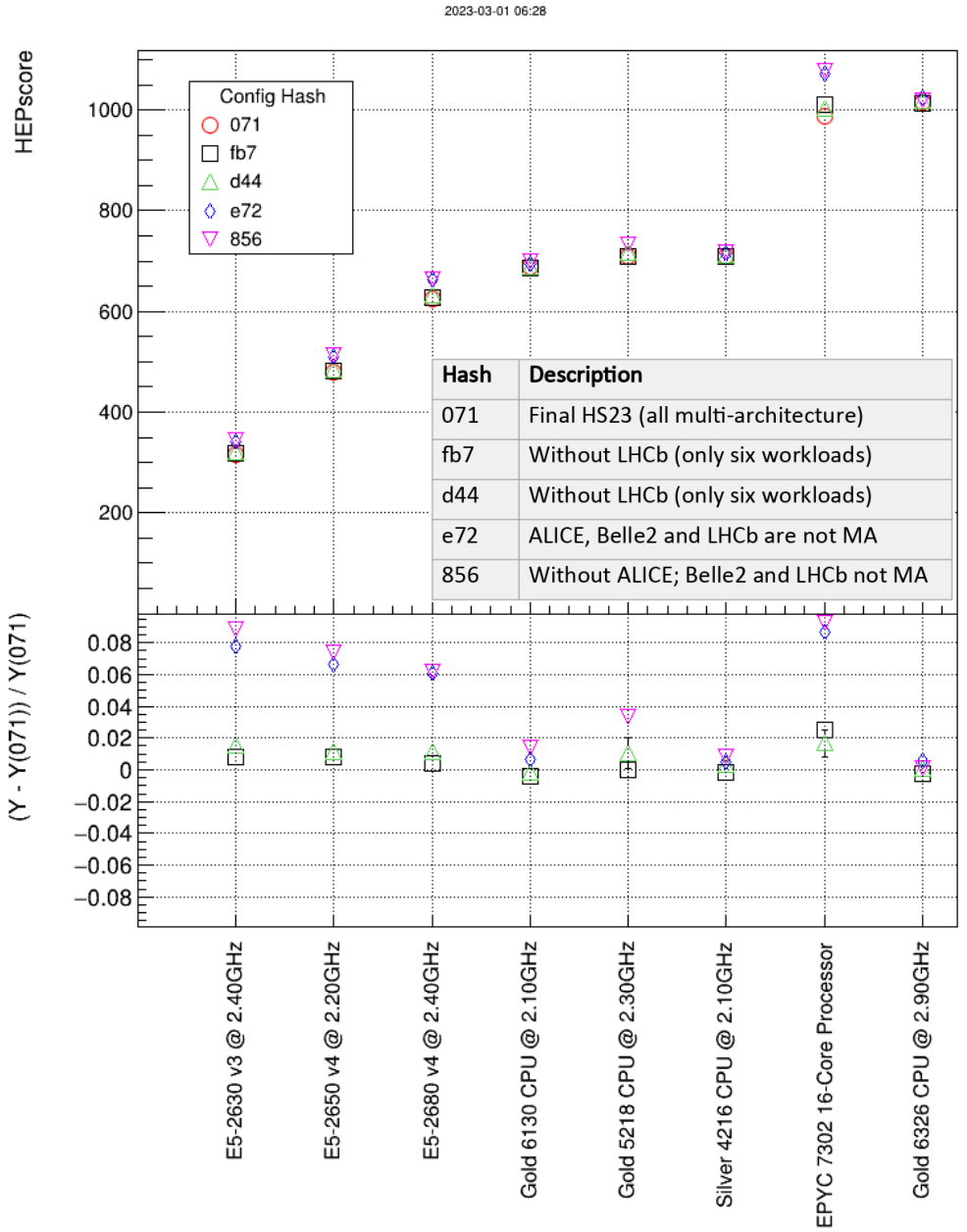


	Min	Mean	Max	Run time (m)
HEPScore$_9$	0.97	1.01	1.04	703
HEPScore$_6$	0.97	1.01	1.06	322

https://indico.cern.ch/event/1170924/contributions/4951092/attachments/2510486/4314832/HEPScoreCandidates_TristanSullivan.pdf

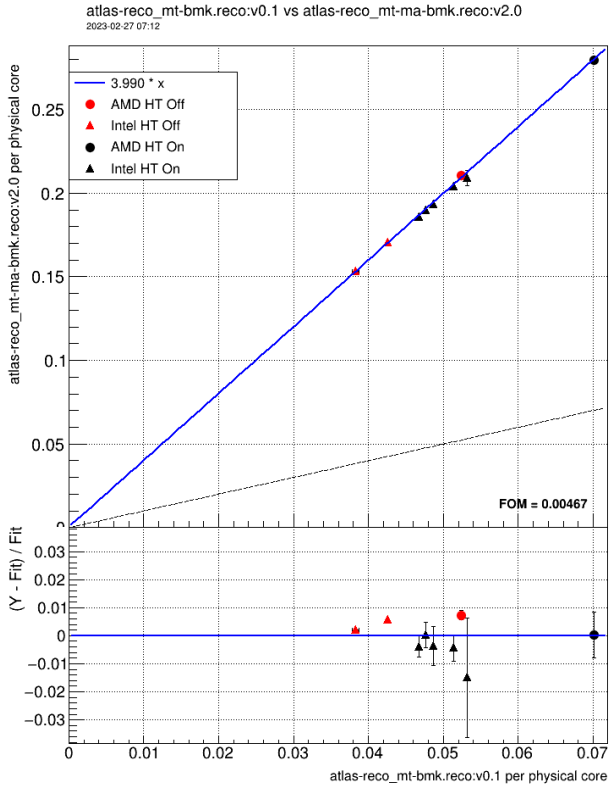
Comparing Configurations

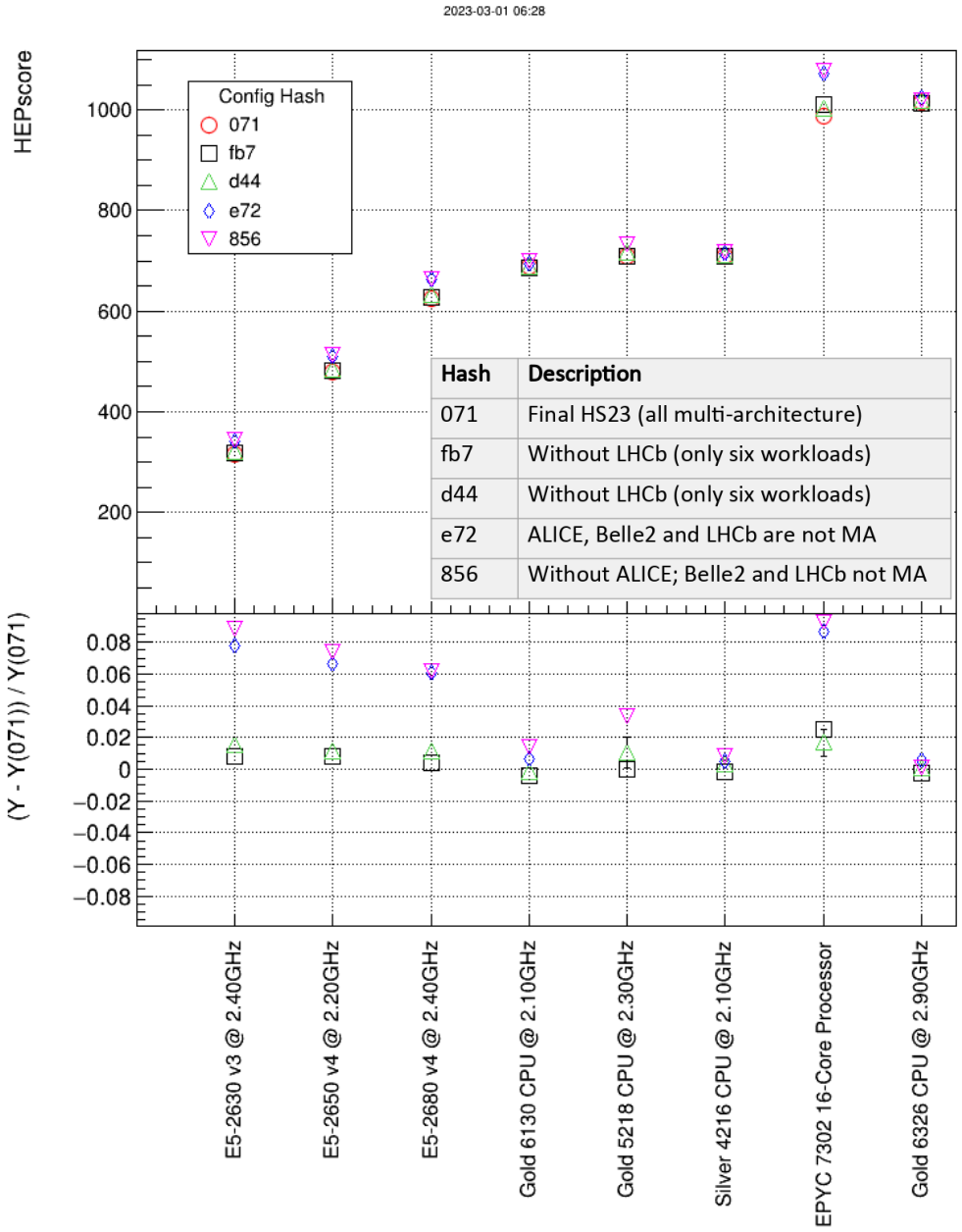
- Examining the impact of CPU choice on HEPscore for different configurations
- The configurations can be considered as different candidates
  - Relative discrepancy $\lesssim 8\%$
  - Low-effect of changes on the result
- 071 is the latest configuration
  - Serves as a reference for the rest



Effects of Workload Changes

- HEPscore increases for all CPUs
- Ordering is maintained

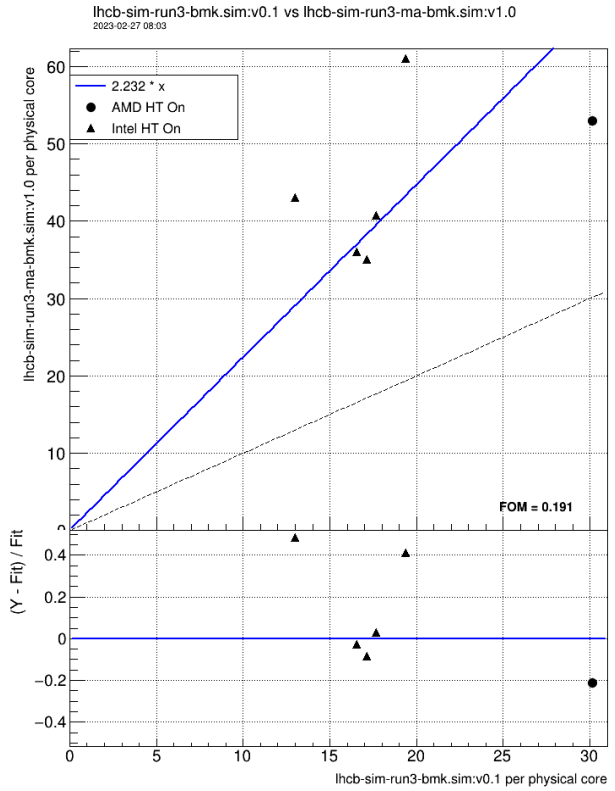


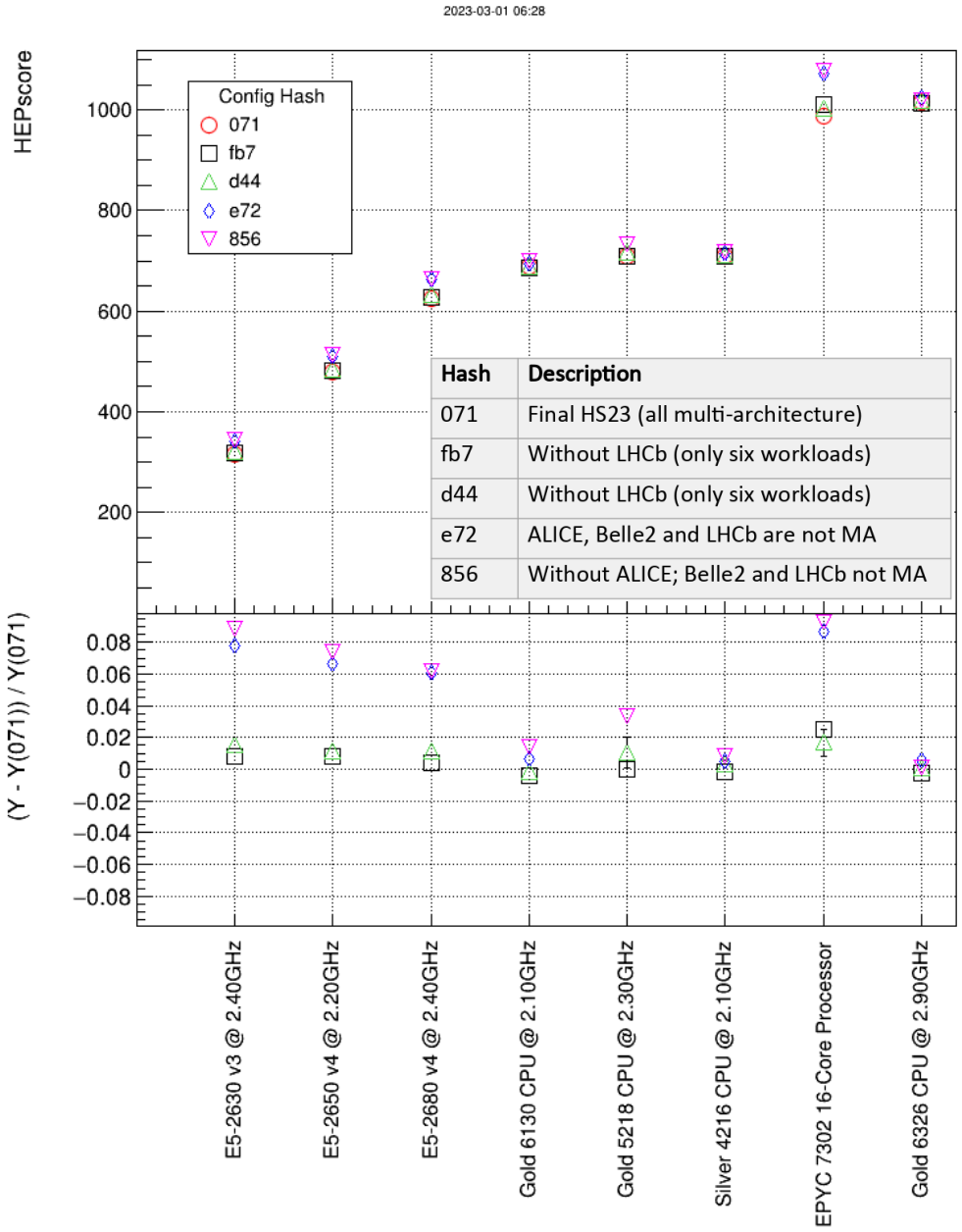


Hash	Description
071	Final HS23 (all multi-architecture)
fb7	Without LHCb (only six workloads)
d44	Without LHCb (only six workloads)
e72	ALICE, Belle2 and LHCb are not MA
856	Without ALICE; Belle2 and LHCb not MA

Effects of Workload Changes

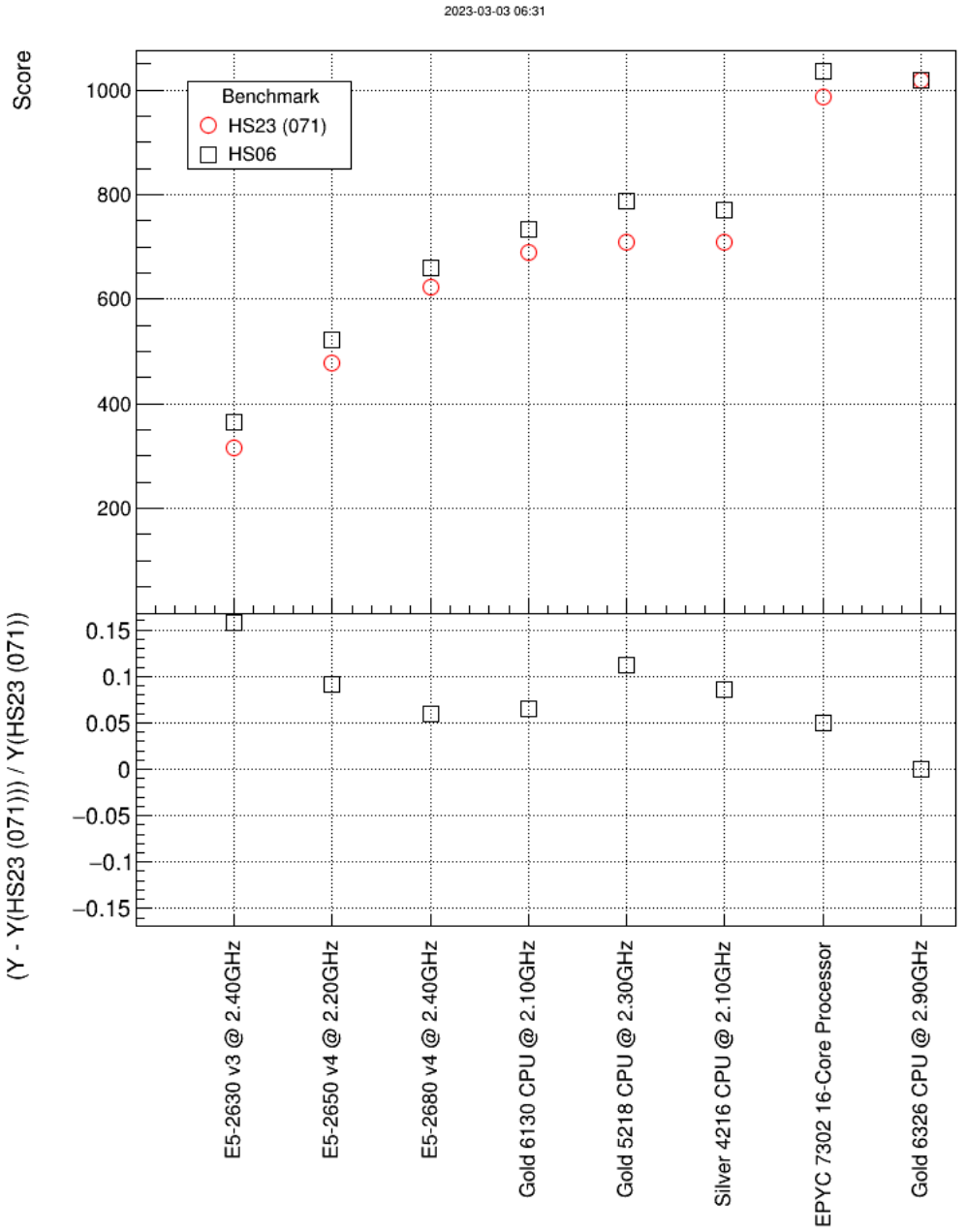
- HEPscore changes differently for different CPUs
- CPU ordering may change
- No drastic change if only a minority of the workloads changes





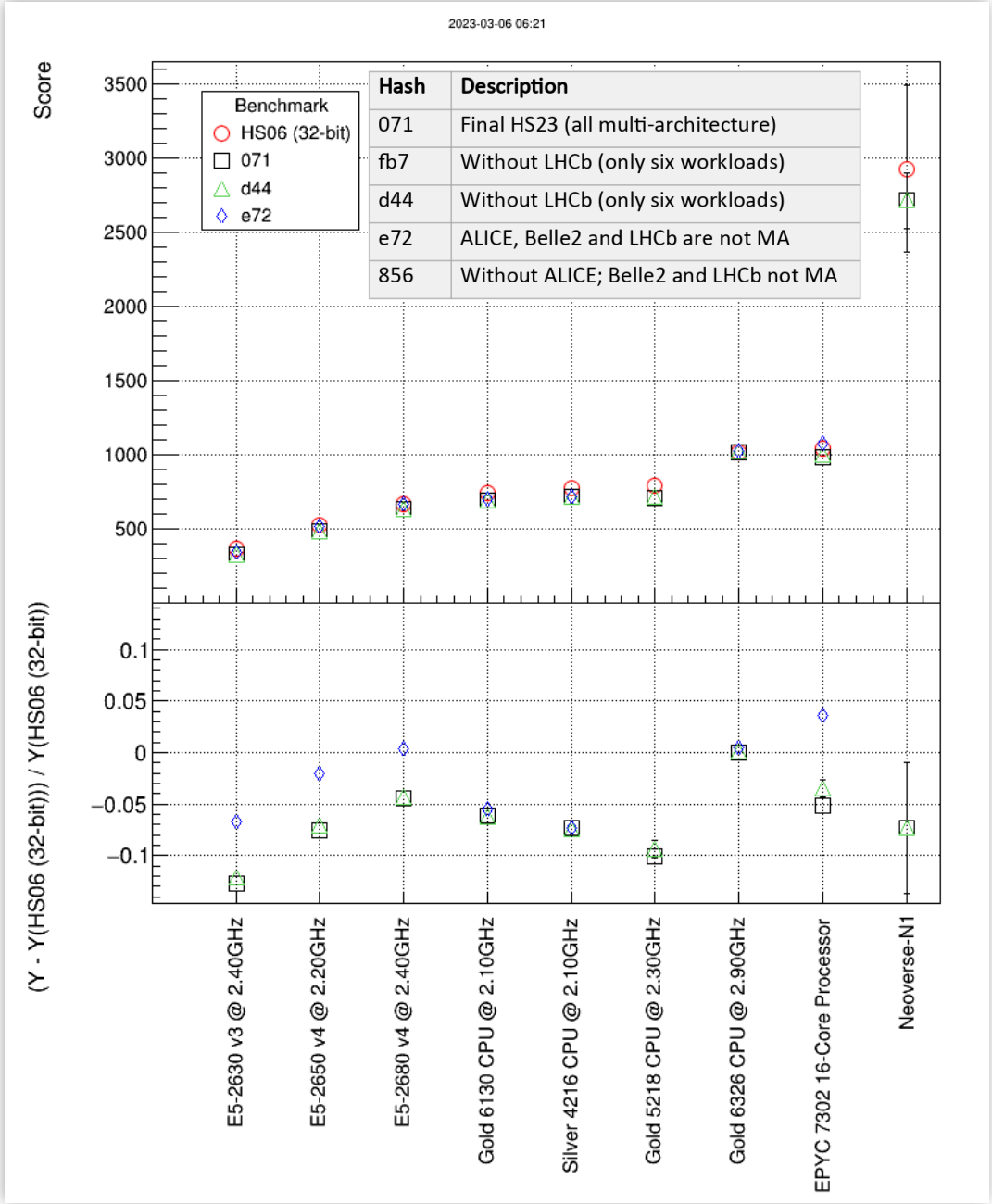
HS23 vs HS06

- Examining the discrepancy between the HS23 and HS06
- Equal scores for the reference machine
- Relative discrepancy ≲ 16%



HS06 vs. HS23 configurations

- HS06 as a reference
- Some configurations have lower discrepancy than others
  - Such as the e72 configuration



Conclusion

- Correlation studies of old vs. new workloads show expected differences for some workloads as a result of large improvements or bug fixes
- The different composition of HEPscore (w/o LHCb and/or ALICE, old LHCb, etc) would affect the servers' HEPscore by < 8%


CERN