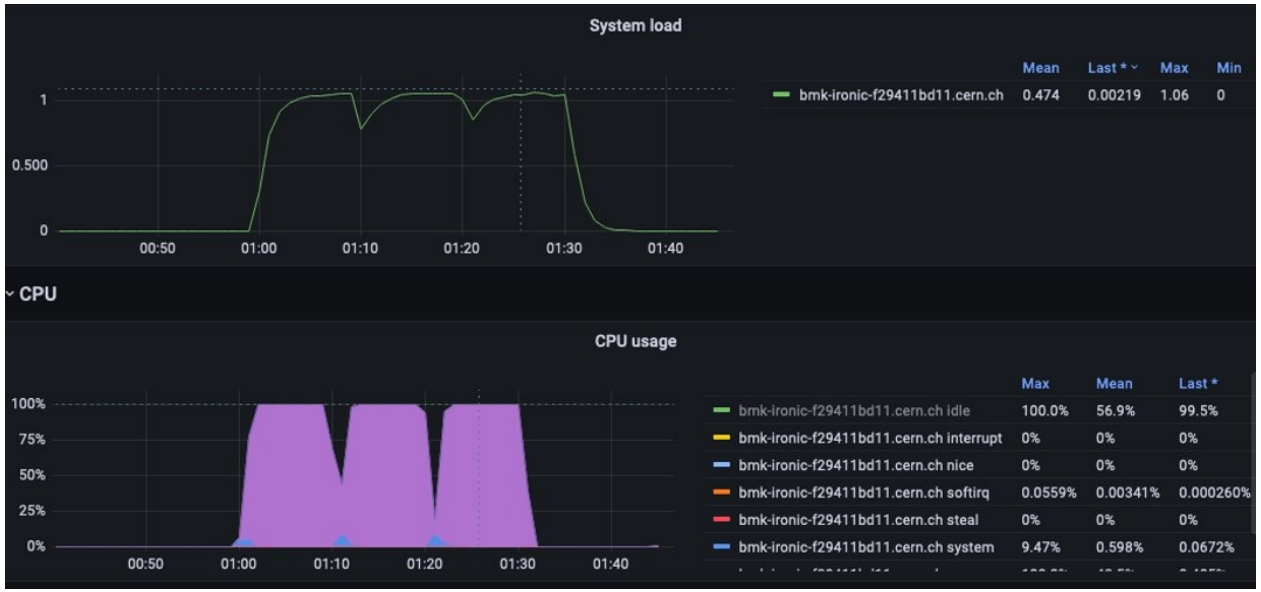
HEPscore23 Workload Analysis

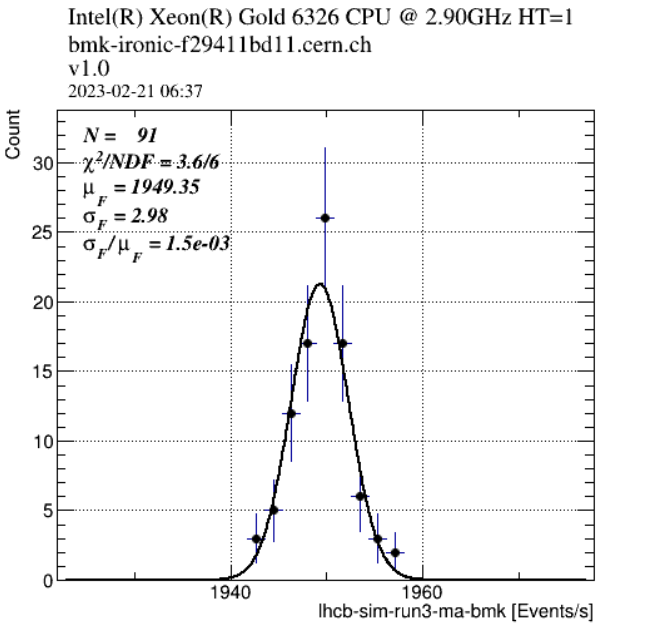
Ladislav Ondris

1 March 2023

Validation of LHCb MA v1.0

- Server utilization as expected
- Spread ≲ 1%
- Reference value: 1949





lhcb-sim-run3-ma

site	CPU	Online CPUs	# reps	# copies	Threads x copy	Events x thread	Count	50th percentile ↓	Spread
CERN	Neoverse-N1	0-159	3	160	1	10	121	5009	1.01%
CERN	Intel(R) Xeon(R) Gold 6326 CPU @ 2.90GHz	0-63	3	64	1	10	74	1949	0.533%
CERN	AMD EPYC 7302 16-Core Processor	0-63	3	64	1	10	77	1695	0.398%
CERN	Intel(R) Xeon(R) Silver 4216 CPU @ 2.10GHz	0-63	3	64	1	10	72	1390	0.441%
CERN	Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz	0-63	3	64	1	10	49	1377	0.475%
CERN	Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz	0-55	3	56	1	10	72	1138	0.620%
CERN	Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz	0-47	3	48	1	10	72	863	0.599%
CERN	Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz	0-31	3	32	1	10	72	560	0.975%

Validation of HEPscore

- Server utilization as expected
- Spread ≲ 2.3%
- Reference value: 1019

Version	site	CPU	Online CPUs	# reps	Count	Score ↓	spread	hash
v1.5rc9	CERN	Neoverse-N1	0-159	3	8	2712	0.707%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) Gold 6326 CPU @ 2.90GHz	0-63	3	18	1019	0.482%	071892
v1.5rc9	CERN	AMD EPYC 7302 16-Core Processor	0-63	3	18	984	0.905%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) Silver 4216 CPU @ 2.10GHz	0-63	3	12	712	2.30%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz	0-63	3	6	708	0.465%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz	0-63	3	2	690	0.124%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz	0-55	3	11	631	2.03%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz	0-47	3	11	481	1.52%	071892
v1.5rc9	CERN	Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz	0-31	3	11	319	2.32%	071892

HEPscore

- Computed as a geometric mean of scores of pre-selected workloads
- Configuration file lists seven workloads:
  - ATLAS gen_sherpa
  - ATLAS reco
  - CMS gen-sim
  - CMS reco
  - LHCb sim
  - ALICE digi-reco
  - Belle2 gen-sim-reco

```
hepscore_benchmark:
  benchmarks:
    atlas-gen_sherpa-ma-bmk:
      results_file: atlas-gen_sherpa-ma_summary.json
      ref_scores:
        gen: 38.58
      weight: 1.0
      version: v2.0
      args:
        threads: 1
        events: 200
    atlas-reco_mt-ma-bmk:
      results_file: atlas-reco_mt-ma_summary.json
      ref_scores:
        reco: 9.062
      weight: 1.0
      version: v2.0
      args:
        threads: 4
        events: 100
```

⋮

```
  settings:
    name: HEPscore23Beta
    reference_machine: "E423521X1B04810-B Gold 6326 CPU @ 2.90GH
    registry: oras://gitlab-registry.cern.ch/hep-benchmarks/hep-
    addarch: true
    method: geometric_mean
    repetitions: 3
    retries: 1
    scaling: 1018
    container_exec: singularity
```

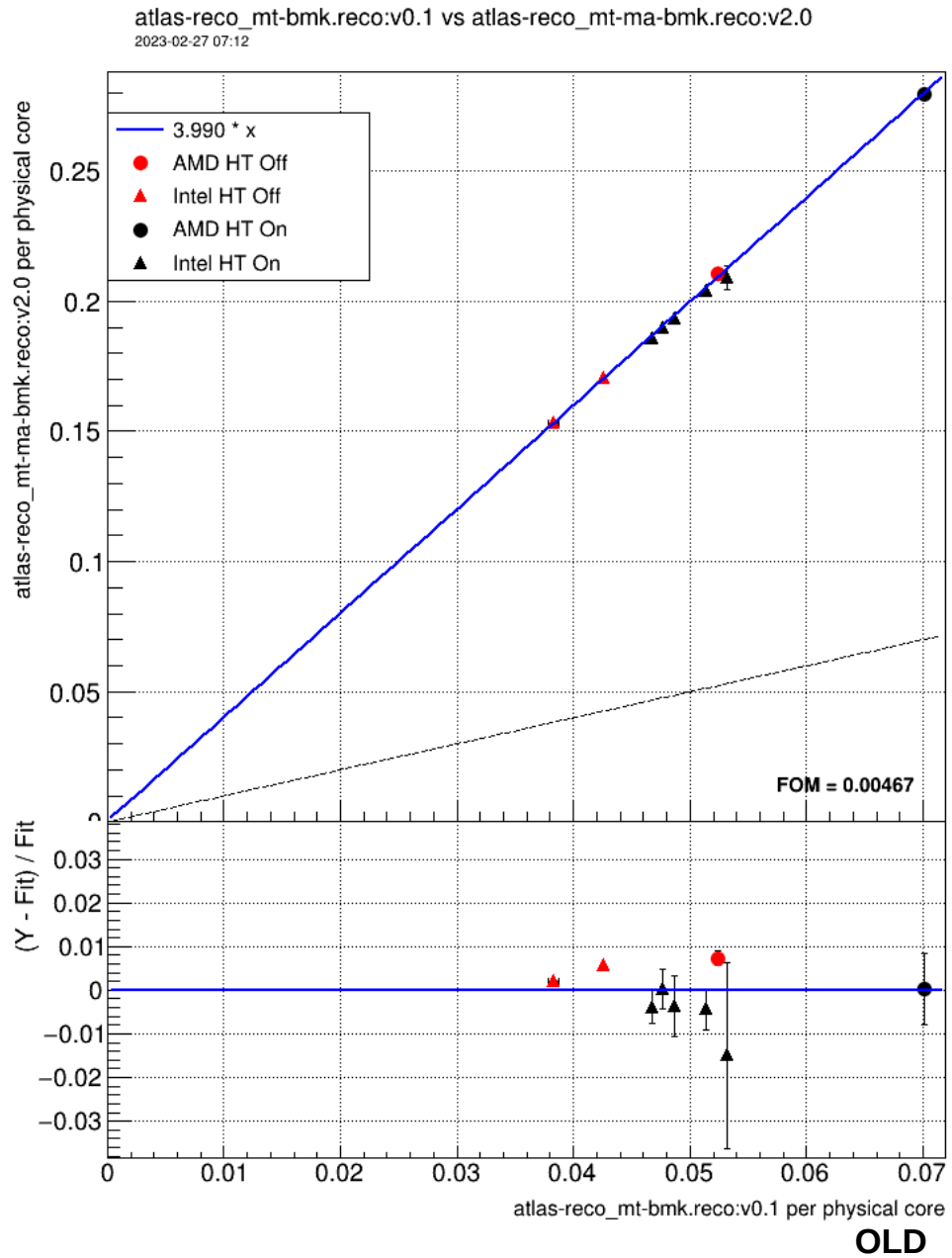
Workload Evolution

- 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 (vx) vs. WL_m (vy)
  - E.g., CMS gen-sim (v0.6) vs. CMS gen-sim-ma (v1.0)

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 4× the old one
- Relative discrepancy ≲ 1.5%

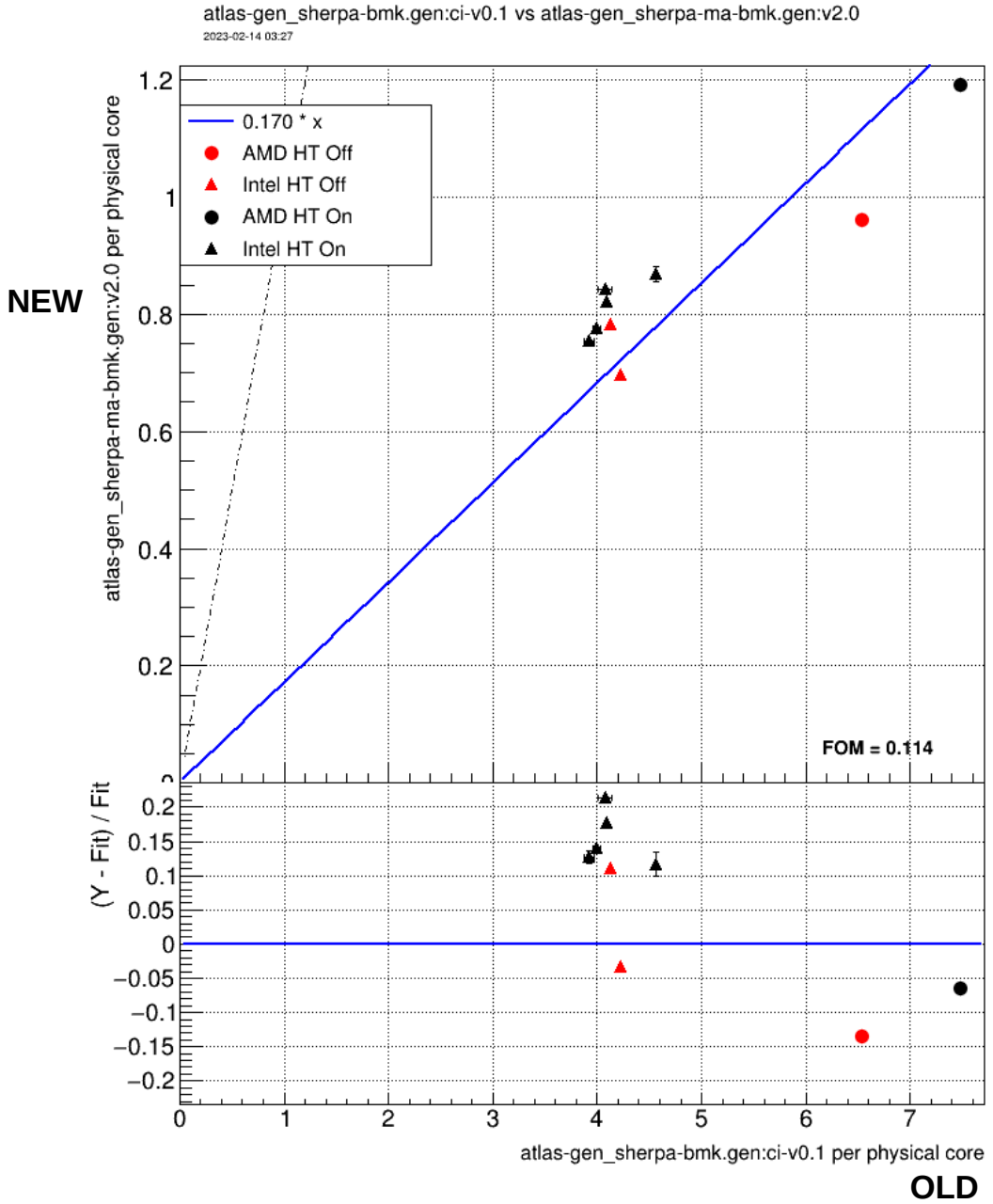
NEW



OLD

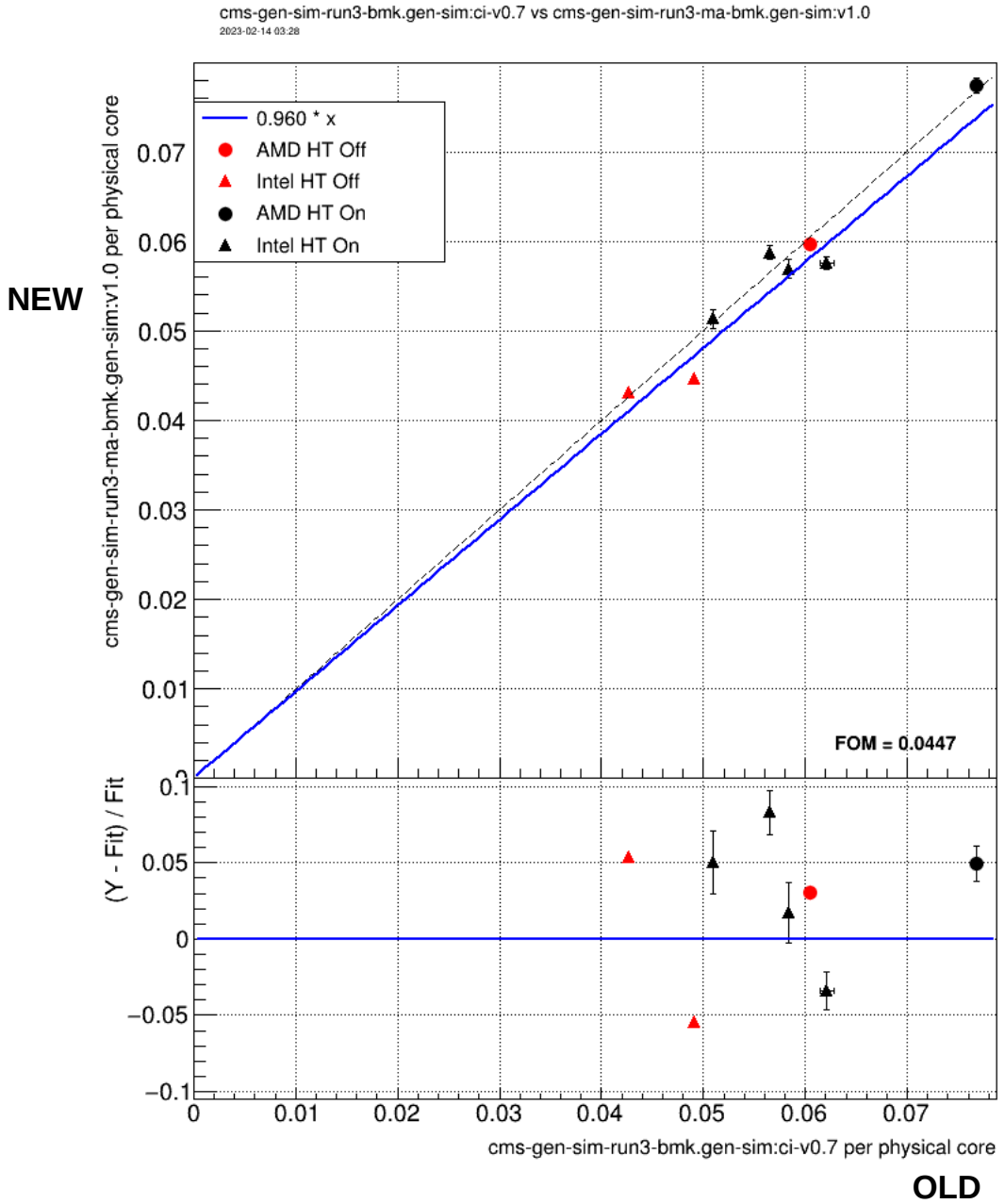
ATLAS gen_sherpa

- 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%



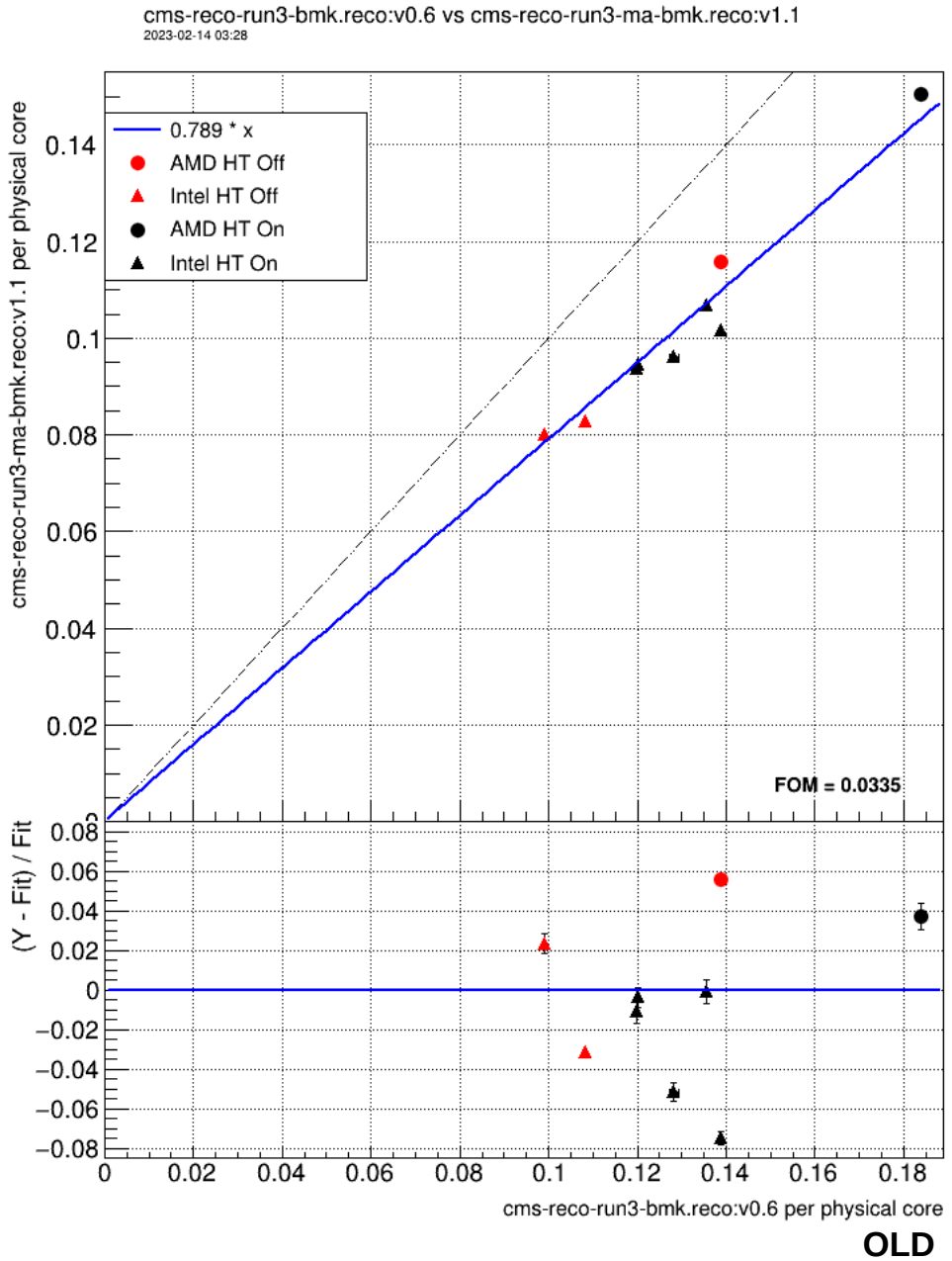
CMS gen-sim

- Same event throughput
- Relative discrepancy ≲ 8%



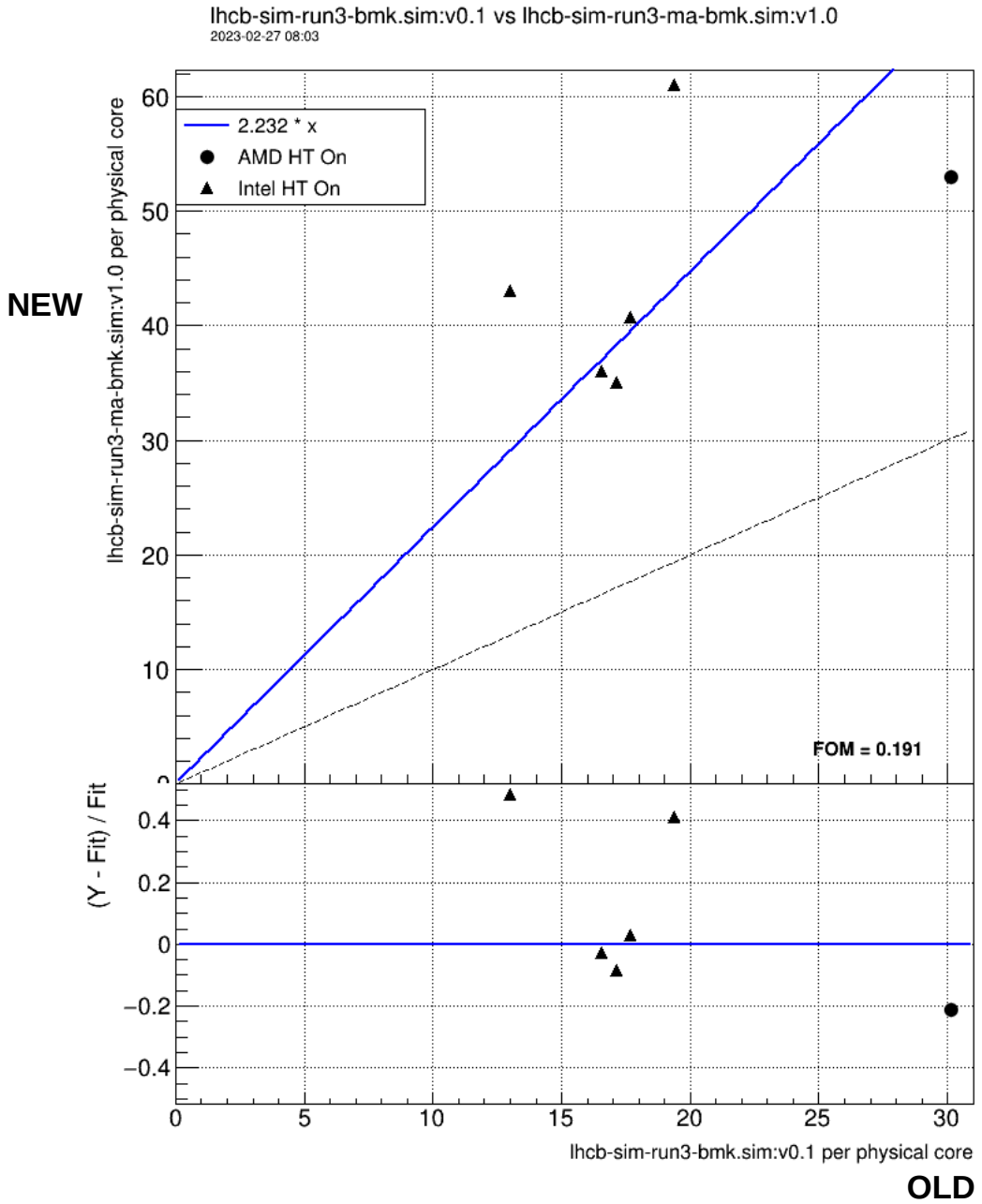
CMS reco

- Lower event throughput
  - Waiting for CMS experts to confirm that it is expected
- Relative discrepancy ≲ 8%



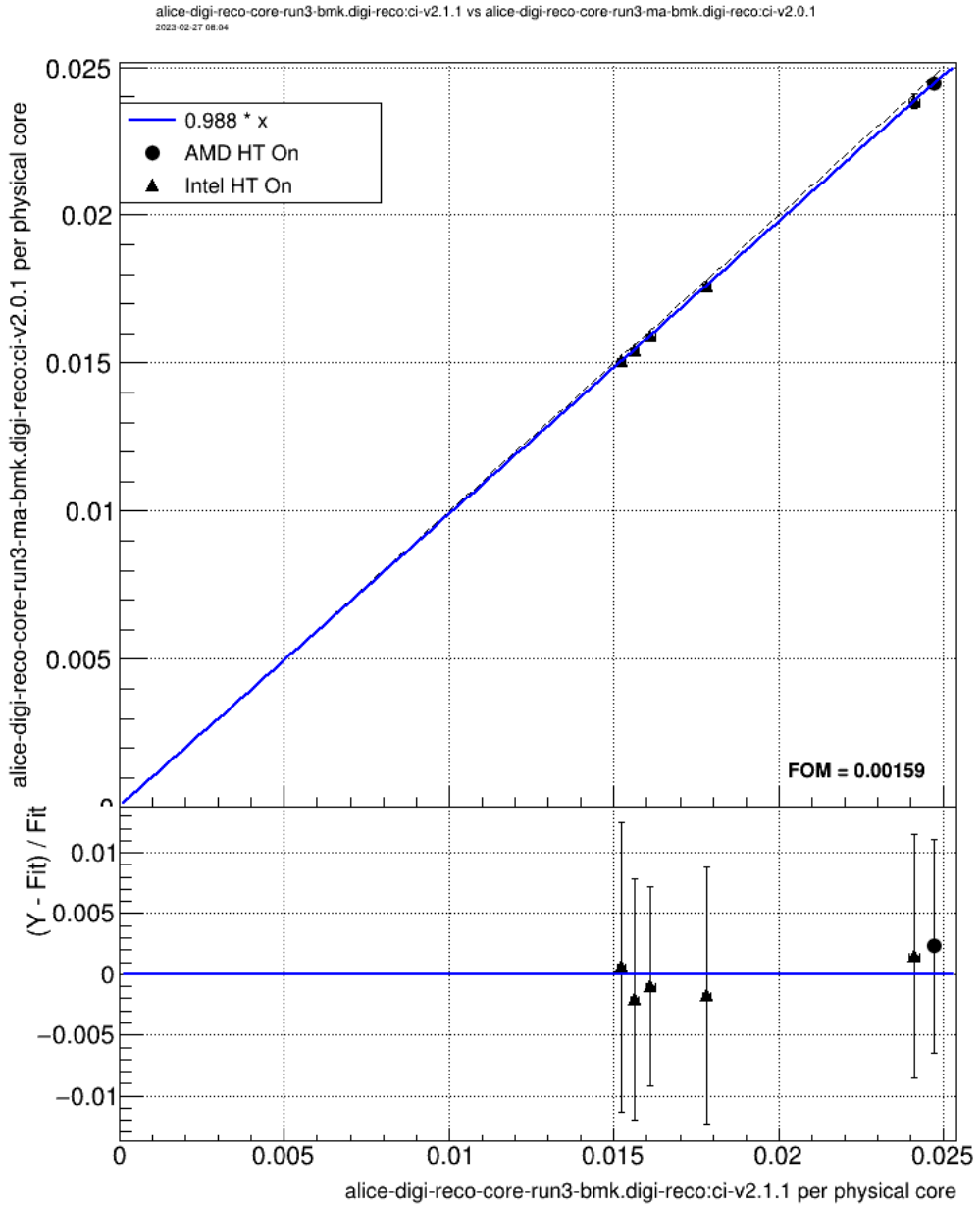
LHCb sim

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



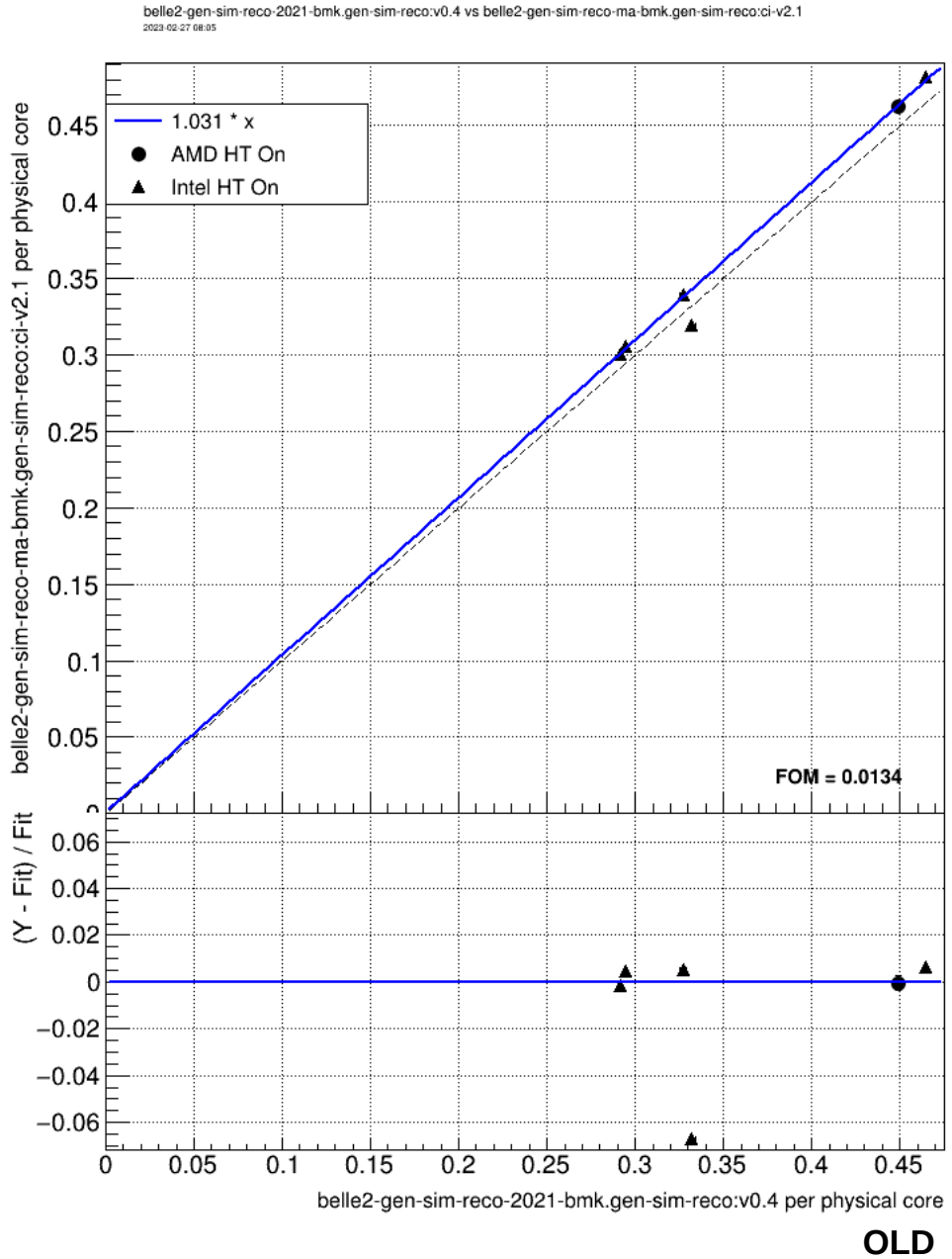
ALICE digi-reco

- Both workloads are recent
- Same event throughput
- Relative discrepancy ≲ 0.3%



Belle2 gen-sim-reco

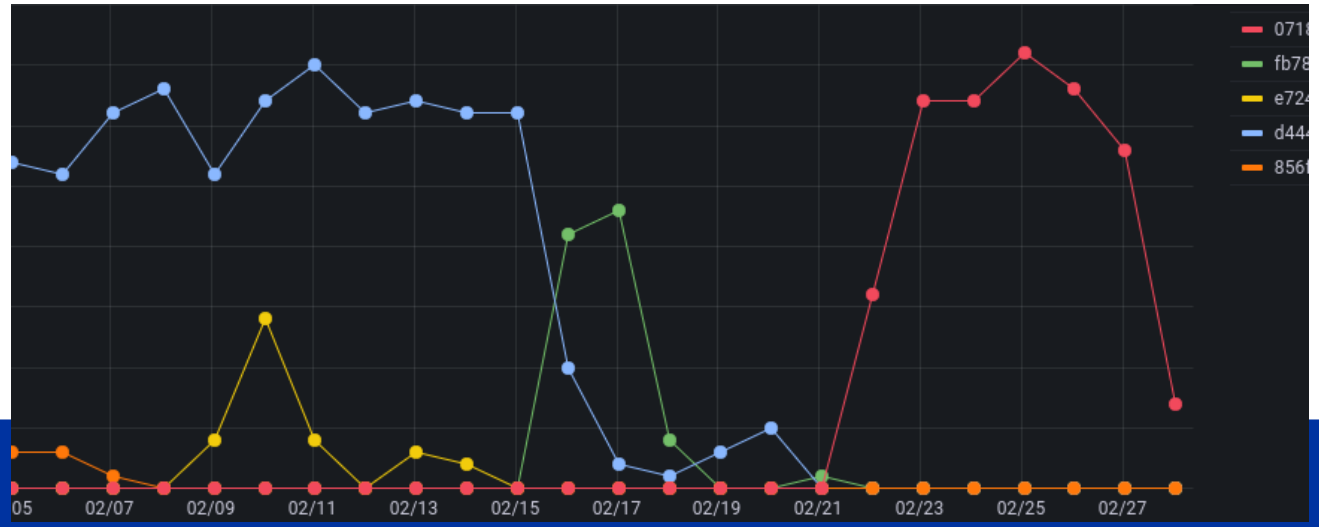
- Same event throughput
- Relative discrepancy ≲ 7%



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

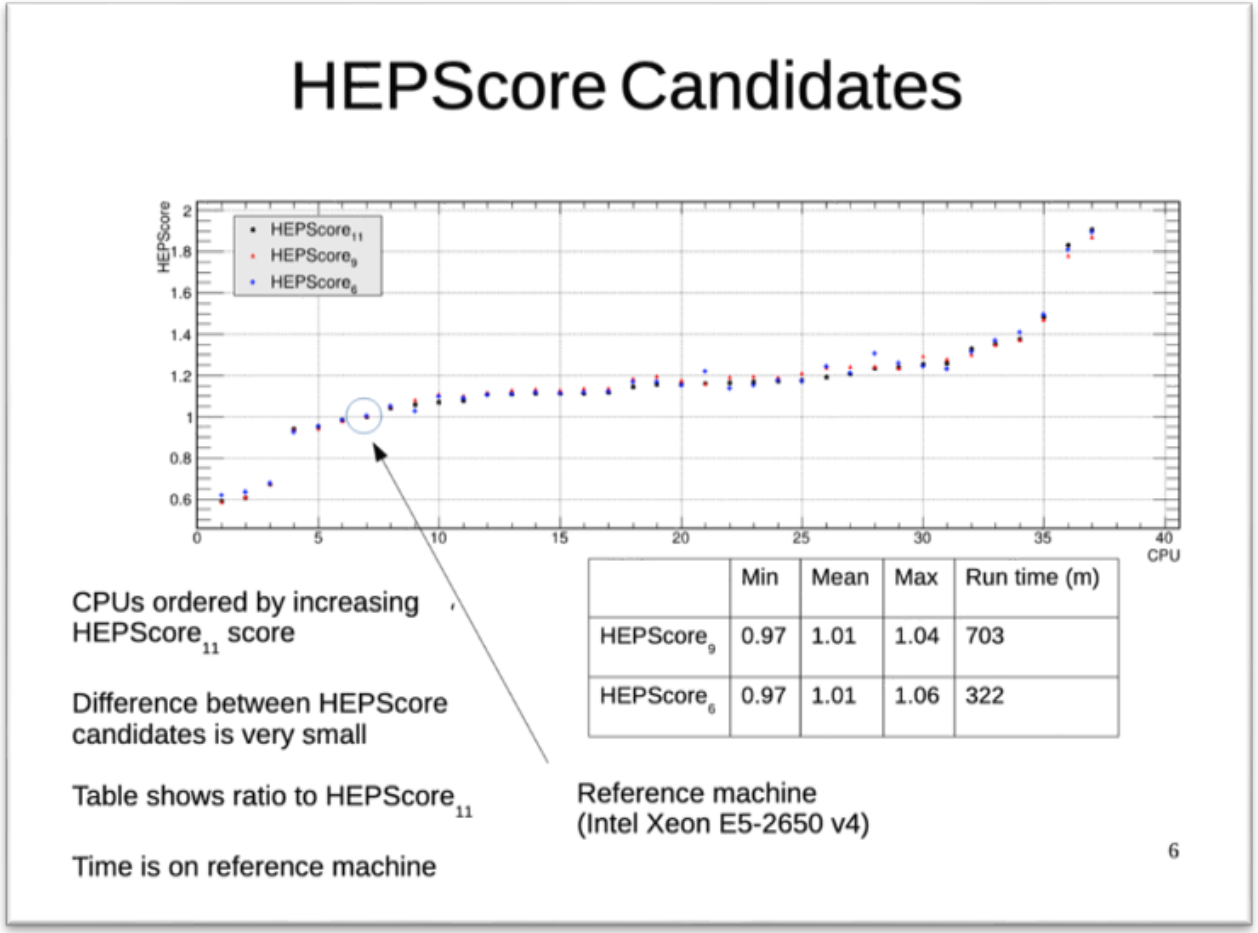
Hash	Description
071	Final HS23 (all multi-architecture)
fb7	Without LHCb (only six workloads)
d44	Without LHCb (only six workloads)
e72	ALICE, Belle2 and LHCb are not MA
856	Without ALICE; Belle2 and LHCb not MA



Effects of HEPscore Configurations

Presented by Tristan Sullivan

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

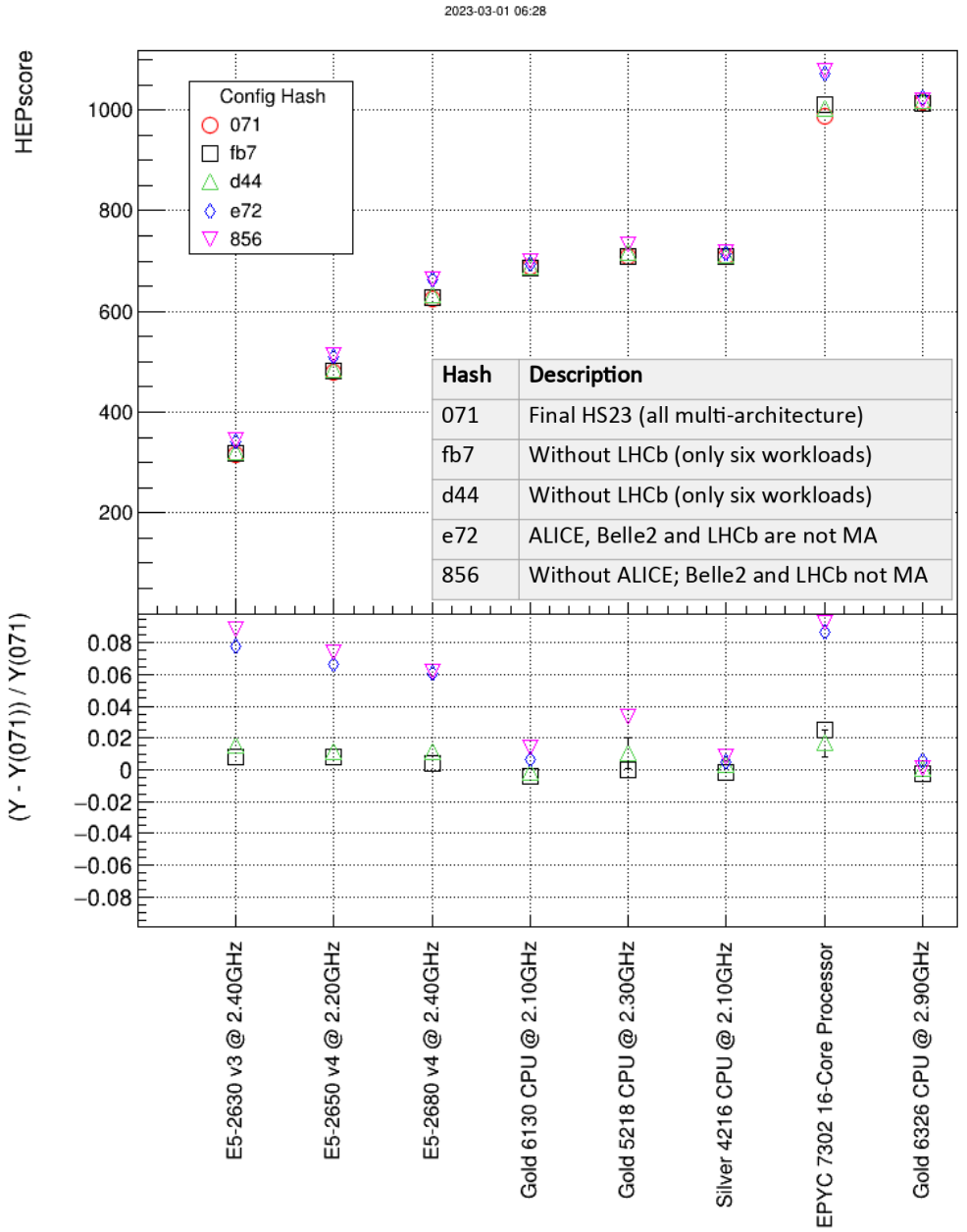


	Min	Mean	Max	Run time (m)
HEPScore$_9$	0.97	1.01	1.04	703
HEPScore$_6$	0.97	1.01	1.06	322

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

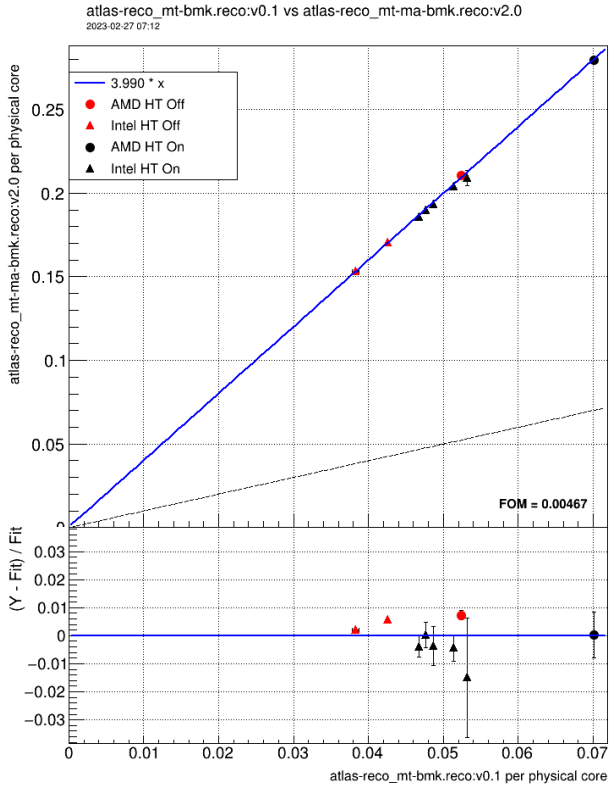
Comparing Configurations

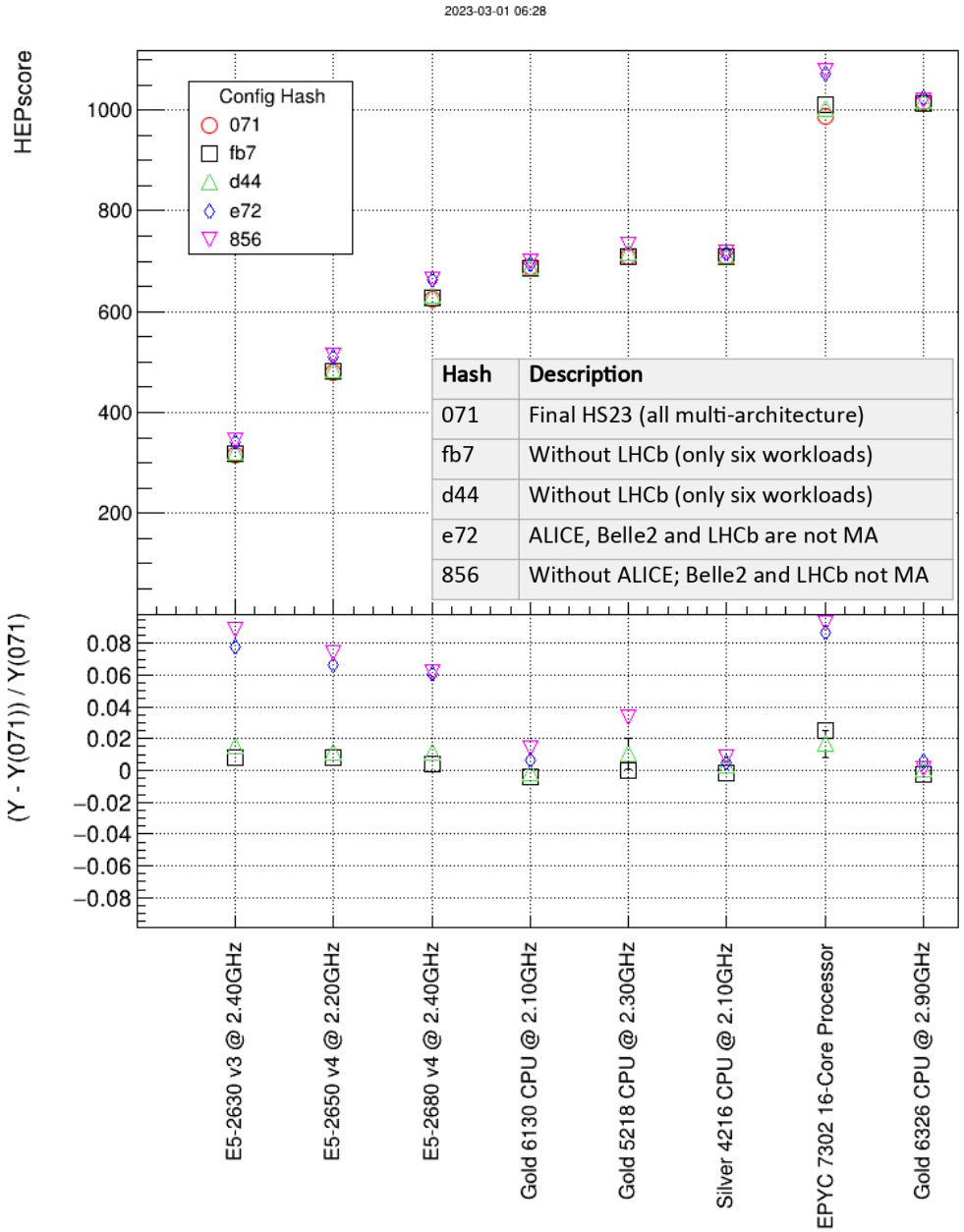
- Examining the impact of CPU choice on HEPscore for different configurations
- The configurations can be considered as different candidates
  - Relative discrepancy $\lesssim 8\%$
  - Low-effect of config hashes on the result
- fb7 is the latest configuration
  - Serves as a reference for the rest
  - Overlaps with d44 as expected



Effects of Workload Changes

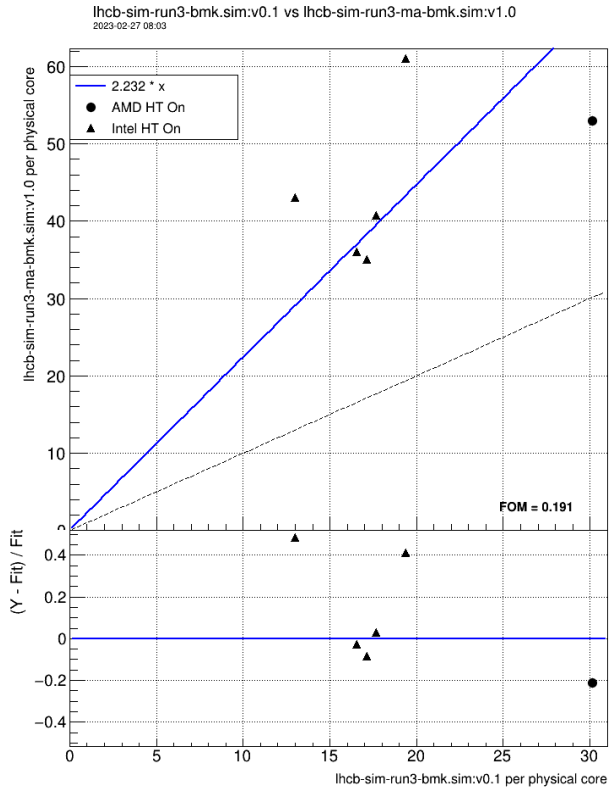
- HEPscore increases for all CPUs
- Ordering is maintained

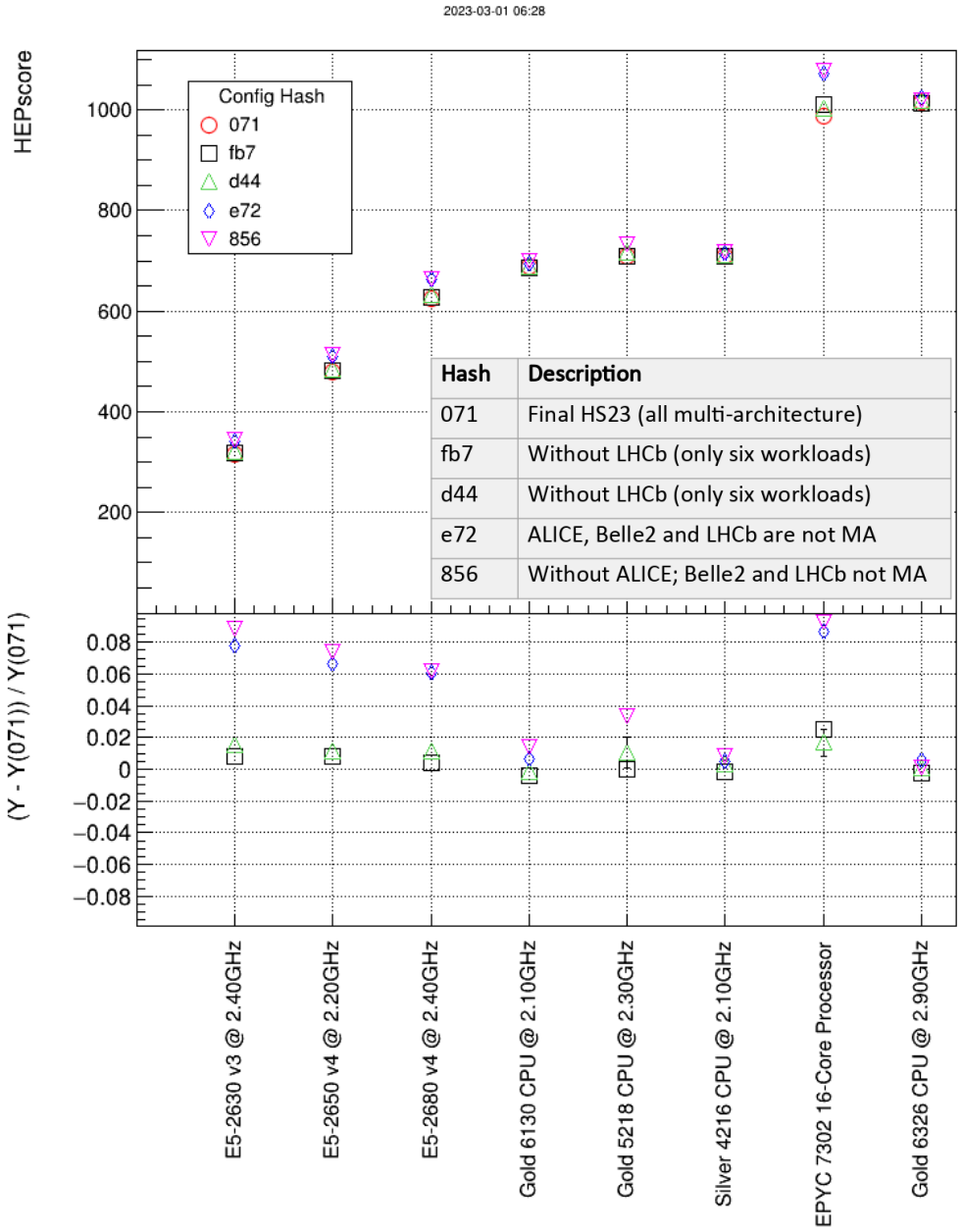




Effects of Workload Changes

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





Conclusion

- LHCb validation confirms stability and reproducibility of results
  - Calibration value for the ref machine is extracted
- HEPscore23 validation on test bed confirms stability and reproducibility
- Correlation studies of old vs. new workloads show expected differences for some workloads 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