

# HEPScore

## the HEP-specific benchmark

D. Giordano (CERN)

on behalf of

HEPiX Benchmarking WG & WLCG HEPScore Deployment TF

CERN Openlab Technical Workshop

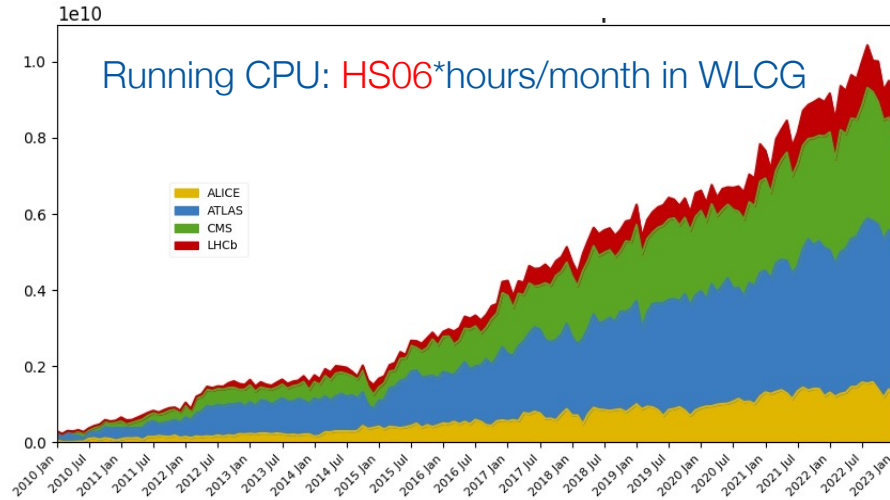
16/03/2023



# Compute benchmarking in HEP

☞ CPU benchmarking plays a central role in the resource-sharing model of WLCG

- Needed for
  - Site pledges & accounting reports
  - Procurement procedures
  - Performance studies
- WLCG distributed infrastructure counts
  - ~1.4 million CPU cores spread over 170 data centres



☞ Since 2009 the HEP community is using **HEP-SPEC2006 (HS06)**

- *de-facto* recognised *currency* to value resources

# Need to evolve

- 🔗 HS06 is a suite of 7 C++ applications CPU-intensive
  - Subset of SPEC CPU® 2006 benchmark
  - *None of the applications comes from HEP*
  - HS06 is not supported anymore by the SPEC org. since 2018



- 🔗 HEP community needed to identify a new benchmark

- 🔗 Which benchmark for HEP?



- *“The first step in performance evaluation is to **select the right measures of performance**”*
- *“The types of applications of computers are so numerous that **it is not possible to have a standard measure of performance [...] for all cases.**”*

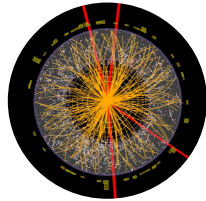
From "Art of Computer Systems Performance Analysis Techniques For Experimental Design Measurements Simulation And Modeling"  
(Raj Jain , Wiley Computer Publishing, John Wiley & Sons, Inc)



# HEP applications

HEP applications consist of

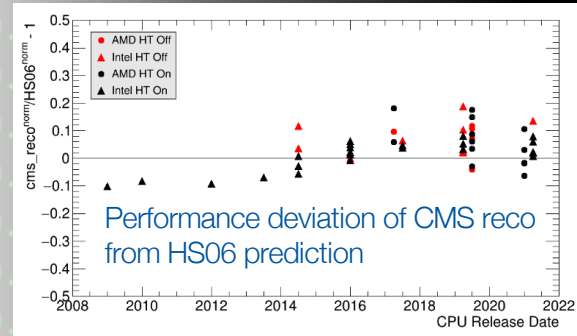
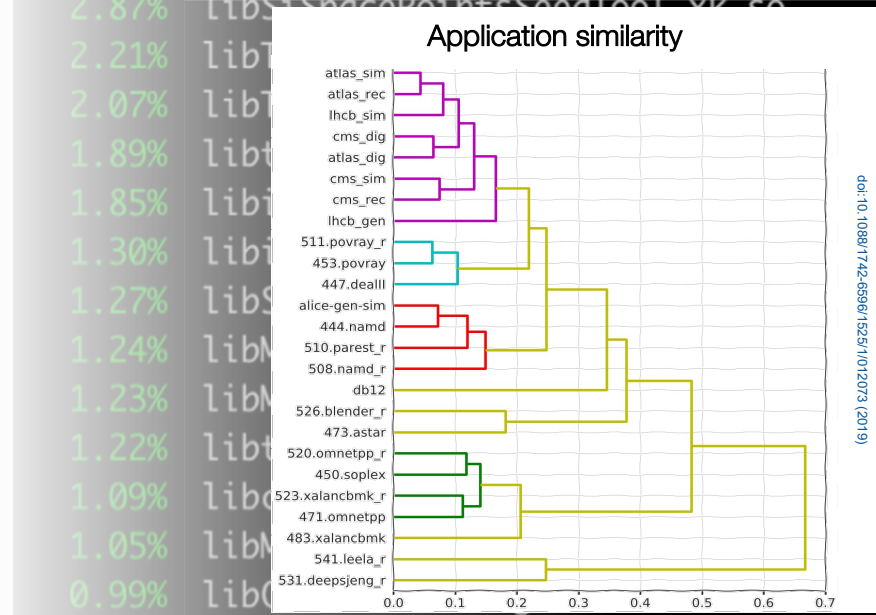
- Several hundred algorithms and no hotspots
- Complex frameworks
- Event based



Experiment software is evolved since 2009

- Adoption of new programming approaches (multi-threading and vectorization) and heterogeneous resources (CPUs, GPUs)

An effective benchmark must scale with the average performance of the applications running in WLCG



doi:10.1088/1742-6596/1525/1/012073 (2019)

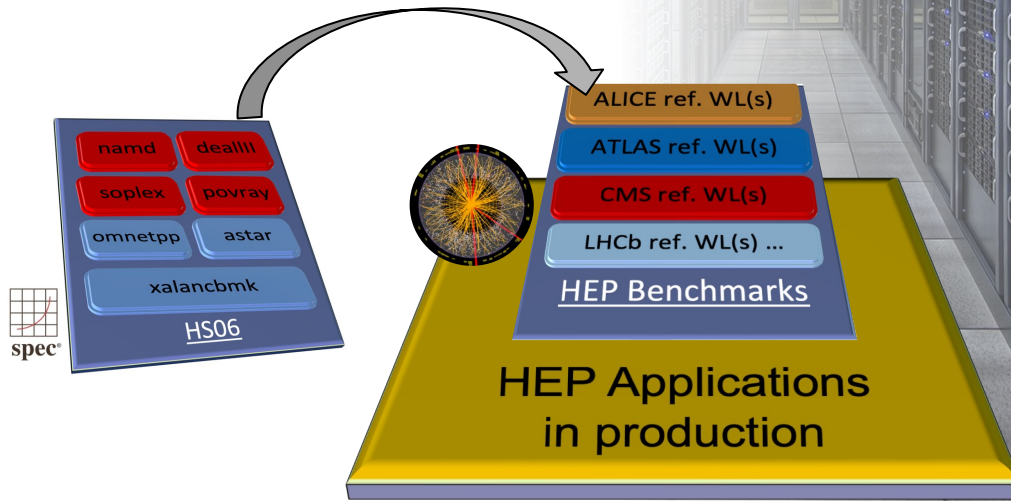
so  
Lib.so  
.8  
0



# A change of paradigm

## Adopt field-specific benchmark applications

- More accurate representation of the real-world performance of a system in HEP



# HEP Benchmarks project

🔗 **HEPScore** has been proposed as alternative to HS06

- Uses the workloads of the HEP experiments
- Combine them in a single benchmark score

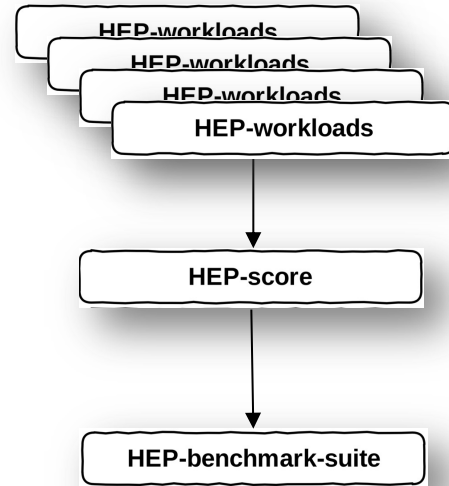
🔗 HEPscore relies on **HEP Workloads**

- Reference HEP applications from multiple experiments

🔗 In addition, **HEP Benchmark Suite**

- Orchestrator of multiple benchmark (HEPScore, HS06, SPEC CPU'17)
- Central collection of benchmark results

🔗 All released under GPLv3 license





# HEP Workloads in containers

## Challenge:

Collect, maintain, execute workloads from several HEP experiments

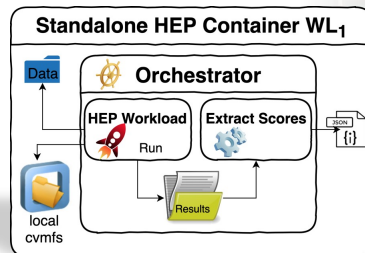
- More than 30 workload versions from 7 experiments prepared

## Requirements

- Zero burden from accessing remote data, databases, etc
- Provide consistent CLI, reporting structure, error logging
- Portable, not too large distribution
- Reproducible results

## Strategy: build standalone containers encapsulating **all and only** the dependencies needed to run the benchmarks

- Support Docker and Apptainer



<https://pixabay.com/photos/belgium-antwerp-shipping-container-1601920/>

# A long process to adopt a new benchmark

🕒 2017

- First proposal of HEP Benchmark with containerized HEP applications (HEPiX Benchmarking WG)

🕒 2018/20

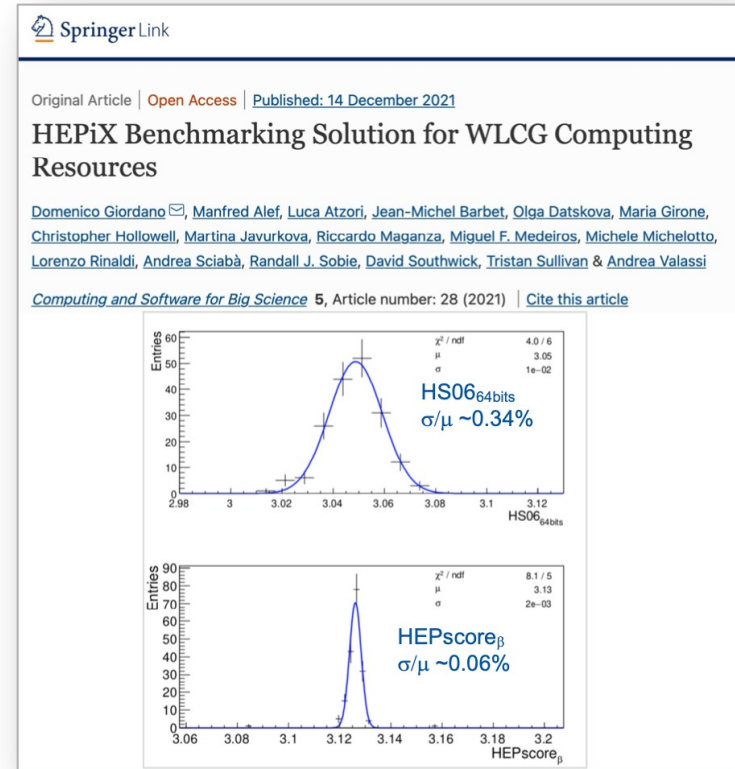
- Design, prototype, validate, deliver

🕒 2020/21

- Proven the technical feasibility of an HEP-Benchmark: HEPscore $\beta$  using experiments' applications from LHC Run2 & Belle2
- Include new applications: from LHC Run 3, Juno, IGWN

🕒 2022/23

- Define the final HEPscore composition validated in a large measurement campaign
- Transition from HS06 to HEPscore

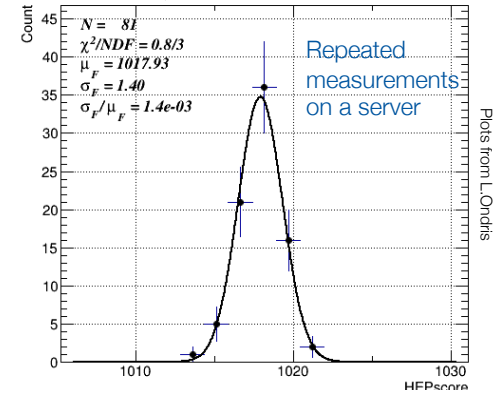




# HEPScore23

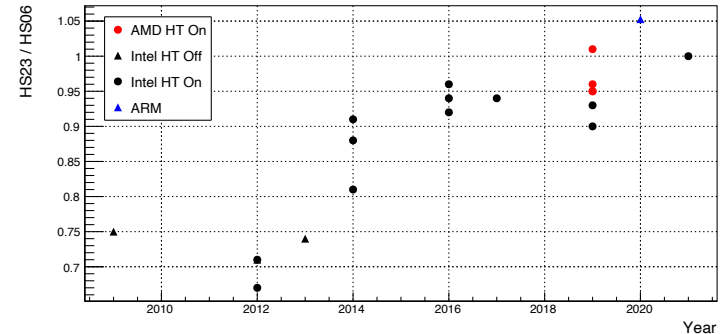
HEPScore23 (HS23) is the configuration of HEPScore to replace HS06

- 7 workloads from 5 experiments included
- All workloads have the most recent version of the experiments' software
- Support **x86** and **aarc64**, extendable to other archs.
  - Enables studies of energy consumption vs performance



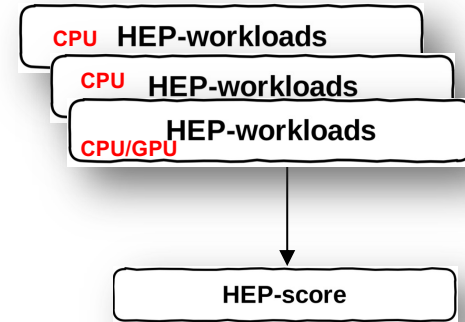
HS23 better represents than HS06 the modernization occurred in HEP applications

- Resolution ( $\sigma/\mu$  typically  $\ll 1\%$ )
- Runtime ( $\sim 4$  hours)



# Extend HEPScore to heterogeneous resources

- 🗨 In the future WLCG resources will include GPUs
  - This is already true for the online farms
  - HEP experiments have/are re-writing their offline applications to use also GPUs
- 🗨 HEP Benchmarks project: growing support for heterogeneous workloads
  - Madgraph4gpu (see talk by A. Valassi)
  - CMS HLT (see talk by F. Pantaleo)
  - ML/AI train AI model (see talk by D. Southwick)
  - Prototyping other workloads, such as analysis
- 🗨 Inclusion of GPU workloads into HEPScore is straightforward
  - If production HEP workloads are available!



# Conclusions

- 🕒 The replacement of HS06 with HEPSScore for CPUs will happen in 2023
  - Identified the “golden” HEPSScore composition of 7 workloads
  - Run on x86 and ARM
  - Being validated on an increasing number of CPU models available in WLCG sites
  
- 🕒 The design of HEPSScore includes its extension as a benchmark for heterogeneous resources

