# The search and development of a new benchmarking tool for the grid

Jan Stypka

September 10, 2014

## 1 Introduction

The goal of the project was to find a benchmarking tool that scales linearly to HepSpec06, but is free and widely available at the same time. In order to test potential candidates for this task, an automatic system had to be rewritten and extended to run the benchmarks, manage data and visualise the results. The system is incorporated in *MonALISA* tool and results can be accessed through the webinterface.

## 2 Benchmarking system

The first step consists of submitting a job using *AliEN*, which is then automatically assigned to one of the machines on the grid. The job extracts information about the CPU and runs selected benchmarks, the results are written to separate files so it is easy to collect the data afterwards. When the benchmarks finish, script parses the files and sends all relevant information to the central *MonALISA* repository service through the *ApMon* library (UDP protocol).

On the server, the data is filtered and loaded into a *PostgreSQL* database. A user can then access the data either through a console and perform analytical tasks or the webinterface, where up-to-date results are dynamically calculated and visualised on the screen using *D3.js* library.

Different views on the data are available:

- Overview on all the gathered data
- Results filtered for one particular CPU configuration
- Results filtered for a particular site (e.g. CERN)
- Results filtered for one particular CPU configuration on one particular site

The tool shows results for several benchmarks as well as standard deviations of the results.

# 3 Candidates for replacing HepSpec06

Two potential candidates have been chosen to see if they behave similarly to the HepSpec06 benchmark.

## 3.1 stressHepix

A benchmark suite shipped with *ROOT* software consisting of several basic benchmarks. Each of them focuses on an other computer component, but in general they measure mostly CPU and I/O of a machine.

## 3.2 AliROOT Monte Carlo simulation

A basic Monte Carlo simulation run using the AliROOT&ROOT package. It is very CPU intensive and focuses almost only on the processor.

# 4 Tests

We have run over 200 000 jobs and gathered results for 115 different CPU configurations. In the meantime we downloaded HepSpec06 values for certain CPUs, which were freely available on the web on the official HEPIX website [1].

These values have been added to the database to calculate its ratio to the other benchmarks.

# 5 Results

We have aggregated the data, calculated averages and standard deviations and plotted correlations between different benchmarks.

In figure 1 the correlation between HepSpec and stressHepix is shown. Deviations in the results are significantly high, especially for the CPUs with hyper-threading turned on. There is a clear tendency in the results, however it is not strong and has several outliers with small standard deviations.

The situation is similar in figure 2. Big error bars indicate high dispersion of data, however a tendency can still be observed. The correlation coefficient is slightly higher than in figure 1, but this is mostly due to larger deviations in many data points.

Figure 3 shows the relationship between two candidates for replacing the HepSpec. Monte Carlo simulation time is inversely related to stressHepix, therefore all its values have been transformed $x \mapsto 1/x$. The correlation is very high with a coefficient of $R^2 = 0.895$, however the error bars are still very large and are not shown on the plot because of readability problems.

[1] `https://hepix.caspur.it/benchmarks/doku.php?id=bench:results`

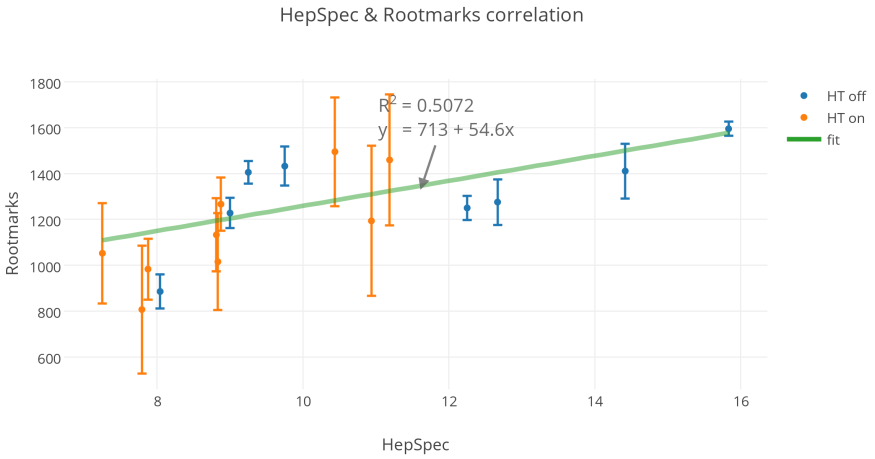


Figure 1: Correlation between the official HepSpec06 and stressHepix benchmark.

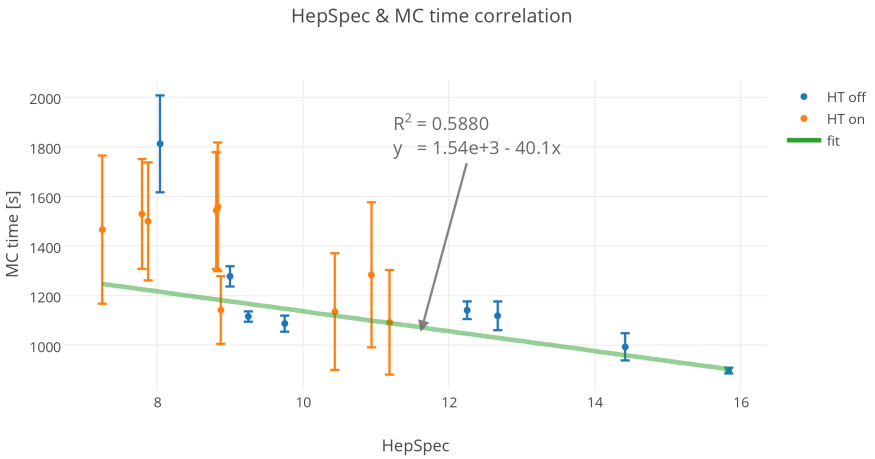


Figure 2: Correlation between the official HepSpec06 and Monte Carlo simulation benchmark.

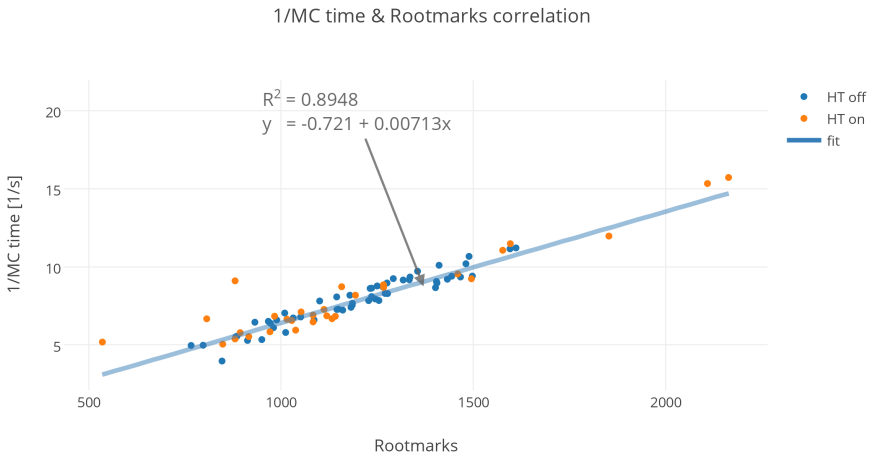


Figure 3: Correlation between the stressHepix and the inverse of MC simulation benchmark.

# 6 Conclusions

The purpose of this study was to investigate the correlation between stressHepix, Monte Carlo simulation and the official HepSpec06 benchmark. Results have shown that the first two candidates are very highly correlated and can be used interchangeably. However their similarity to HepSpec is questionable, there is a very slight correlation but not strong enough to draw significant conclusions.

The problem in the results are very high deviations for many data points. This is probably due to the difference in hardware specification (e.g. different motherboard, cooling system etc.) and particular circumstances in which the job was run i.e. what was happening on the other cores in the same time. Furthermore we do not know how the data from the Hepix website was collected and whether it does not contain any mistakes. The tests might have been carried out in completely different circumstances, which does not permit us to draw significant conclusions.

In order to enable to carry out this study, a system for running benchmarks has been developed. It consist of an *AliEn* job which is responsible for running the benchmarks, extracting the CPU information, parsing the results and sending them to a central repository. The server filters them and loads into the database where they can be queried by various clients. The web client visualises the data in various configurations using the *D3.js* library.