PhD Proposal:
HPC optimization for simulation code in astrophysics

Summary

The CTA (Cherenkov Telescope Array) [1] project aims to build and operate a network of about 100 Cherenkov telescopes located in the La Palma (Canary Islands, Spain) and Paranal (ESO [2], Chile) sites. CTA opens a new era in very-high energy gamma astronomy. CTA consortium gathers about 1200 members coming from about a hundred of research institutes in 32 countries.

Today the different processor generations show architectural improvements allowing performance gains in execution time (x2 to x16) and low energy consumption. Software programs should take into account the hardware specificities to fully exploit their performances.

The first objective of this PhD is to optimize the performances of the air shower simulation software used by CTA, by benefiting of the capabilities of new generation processors. Reducing the CPU time devoted to simulations would allow to reduce e-infrastructure costs as well as to improve scientific performances thanks to a better instrument characterization. The second objective of this thesis is to contribute to the development of a new general framework for the air shower simulations, common to several international projects in the cosmic-ray domain.

The PhD will take place in the context of a collaboration between computer scientists of the DALI team at UPVD-LIRMM/CNRS laboratory and astrophysicists of the LUPM/CNRS laboratory (PEPS Astro-Informatique [3]).

1. Scientific context

The CTA (Cherenkov Telescope Array) [1] project aims to build and operate a network of about 100 Cherenkov telescopes located in the La Palma (Canary Islands, Spain) and Paranal (ESO [2], Chile) sites. CTA opens a new era in very-high energy gamma astronomy. A first large telescope has being installed in La Palma and data acquisition started at the end of 2018. CTA consortium gathers about 1200 members coming from about a hundred of research institutes in 32 countries.

In order to characterize the instrument response to the Cherenkov light emitted when cosmic ray showers develop in the atmosphere, detailed Monte Carlo simulations will be regularly performed in parallel to CTA operation.

For the detailed simulation of air showers CTA uses the Corsika software [4], considered as the reference software in the whole community studying cosmic rays. The computation time associated to these simulations is and will remain very high, i.e. about 200 million normalized CPU hours per year. Typical Monte Carlo productions run on about 8000 cores in parallel, distributed at about 20 grid sites of the european grid EGI [5]. Reducing the computation time required for simulations is therefore a major challenge for the project.

A project of full re-writing of Corsika (Next Generation Corsika) [6] has started in 2018 with the contribution of different users communities: CTA, Auger [7], IceCube [8], etc. The current version is indeed the result of developments started in early 1990s. The capability to efficiently exploit new generation processors has been identified as a requirement of the new software.

2. Objectives

The objective of this thesis is twofold: firstly, to obtain the “maximum” performances of the current version of Corsika, which will remain in production for about 4-5 years and secondly, to contribute to the development of the Next Generation Corsika (NGC).

A preliminary study carried out in the framework of the Astro-Informatique PEPS, allowed us to show that automatic optimization tools have limits when confronted with complex codes like Corsika. The main idea is to explore the entire space of optimization opportunities by avoiding any constraints. We want to overcome

the limitations of static analysis and compiler heuristics, missing dataset information, and missing information on the target microarchitecture.

To achieve this end, the candidate will first have to perform "manual" program transformations that will take into account information on the datasets and the target micro-architecture.

In this transformation process, vectorization techniques will be considered, as well as the utilization or the development of optimized mathematical functions. It will then be necessary to measure and quantify performance improvements related to different constraints and to highlight the most promising ways to remove the locks of automatic optimization. This study will allow to identify the potential improvements of the automatic optimization process, in which is incorporated the numerical precision tuning.

In the second part of the thesis, the candidate will contribute to the development of the NGC. The goal is to deliver a general framework to maximise the benefit of using modern vector CPUs in a transparent way. The main idea is to explore a new paradigm to manage the simulation of 'particle transport', exploiting different levels of parallelism, bundling particles with similar properties. This approach is expected to yield significant speed-ups, which will be measured in realistic applications.

Finally, the optimized version of Corsika as well as the first version of NGC will be deployed on the EGI grid for their full validation with high statistics.

This work will be carried out in an international context in close collaboration with astrophysicists, members of the CTA collaboration, which will ensure the physics validity of the results obtained with the optimized Corsika and NGC.

3. Prerequisites

The retained candidate will have intermediate knowledge of processor architectures and compilation process. He/she should have solid experience in C/C++ projects. The knowledge of Fortran languages is a plus.

4. Duration, location, supervision and logistics

- Location: University of Perpignan Via Domitia, DALI/LIRMM team

- PhD duration: start in October 2019 for 3 years duration

- PhD directors: P. Langlois, Université de Perpignan, DALI/LIRMM/UPVD
  G. Vasileiadis, LUPM/CNRS

- Supervisors: Luisa Arrabito (LUPM)
  David Parello (DALI/LIRMM/UPVD)

- Contacts: arrabito@in2p3.fr, david.parello@univ-perp.fr

5. References

[1] The CTA Consortium 2017, Science with the Cherenkov Telescope Array, https://arxiv.org/abs/1709.07997
[2] https://www.eso.org/
[3] https://gite.lirmm.fr/cta-optimization-group/cta-optimization-project/wikis/home
[4] https://www.ikp.kit.edu/corsika/
[5] https://www.egi.eu/
[6] https://arxiv.org/abs/1808.08226
[7] https://www.auger.org/
[8] https://icecube.wisc.edu/