HPC Geant4 Brainstorming

2 October 2019

Date

Motivation

- There exists the folklore that just investing effort in ‘re-writing’ the Geant4 code into Cuda, we could profit from HPCs in which the bulk of computing resources are realised with GPUs
  - e.g. HPC cross-experiment discussion (May 10th)
  - e.g. ExaScale project un US
- High-level people asking in various fora what Geant4, at least the CERN part, is doing for running simulation on HPCs using efficiently the accelerators (e.g. GPU's) that are there
  - e.g. LHCC, Scientific Computing Forum, etc.
- ORNL offering expertise to CERN to help to solve the computing problem of HL-LHC

CERN EP SFT

GeantV R&D

- We are now concluding GeantV R&D work with the following findings (my personal lessons conclusions)
  - Re-writing and modernising large parts of Geant4 cloud bring us a factor of ~2
    - compact code, better data formats, less virtual functions, etc.
  - Vectorisation (organising the work in baskets) does not bring any improvement
    - overhead in continuously reshuffling is too large, dealing with tails
  - The initial work (and very partial) with GPUs demonstrated to be impractical
    - same as vectorisation but larger baskets to make it efficient, copying data in and out from device, code duplication in and out

CERN EP SFT

Ongoing Work and Resources

- The current strategy is to develop along two main axes:
  - Improve, optimise and modernise the existing Geant4 code to gain in performance for the detailed simulation. We have had some recent successes but we need to do more
    - Guilherme started yesterday... + Andrei and Witek
  - Trade precision for performance using fast simulation techniques both with parameterisations or with ML methods and integrate them seemly in Geant4
    - Anna, Ioana, and new EP R&D fellow...
- Shall we add a third axis to explore accelerators in a different manner than what was done in GeantV?

CERN EP SFT

Objectives of today's meeting

- Explore ideas for transforming the current way particles are transported in Geant4 to make the problem more homogenous (with much less branches) and therefore better fit to GPU architectures
  - We must think out of the box
- What concrete 'demonstrators' should we invest in developing?
  - Produce a list of of things to test and check
  - Who will do it?
  - With what resources?
  - How much time do we have?

CERN EP SFT