R&D Task Force status report

Witek Pokorski
Jonathan Madsen

21.09.2020

[Virtual] 25th Geant4 Collaboration Meeting

Geant4 Task Force for R&Ds

- promote and survey research activities on
  - potential software architectural revisions of Geant4
  - exploitation of emerging technologies or computing architecture that would be beneficial for Geant4
- ensure the visibility that the Collaboration is open to such exploitations and act as the focal point for such activities inside and outside Geant4
- assess performance of different improvements and effort for integration
  - Make recommendations to the Geant4 SB
- intended as catalyzer for short-cycle integration of performance developments
  - Coming from within and outside the collaboration

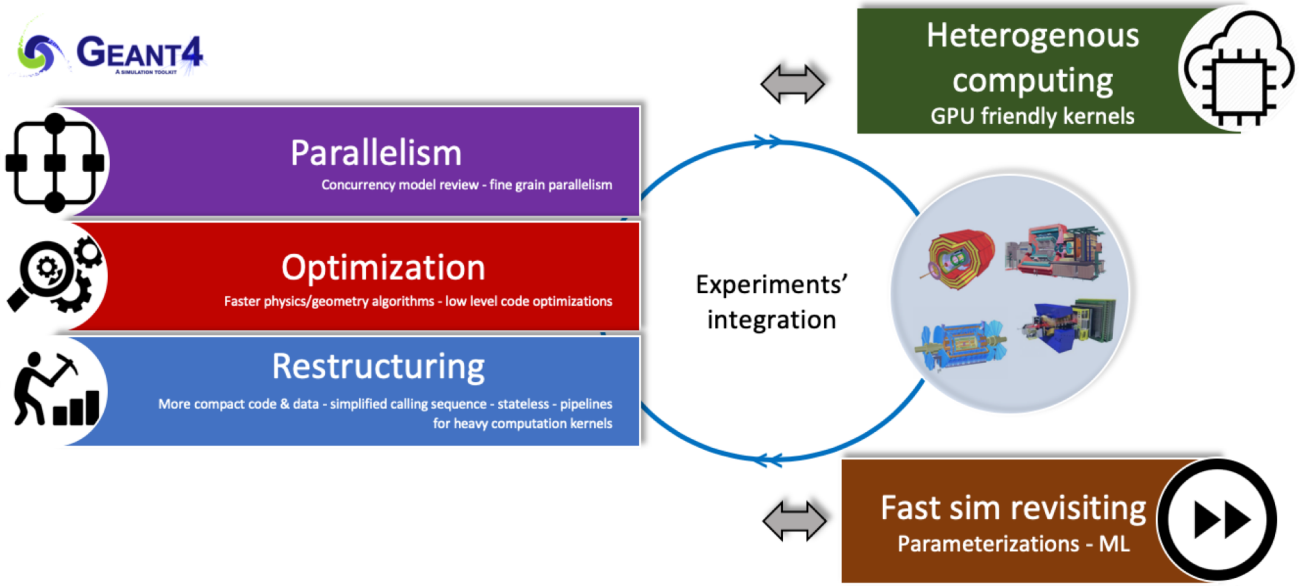
Activity

- started with brainstorming meetings, gathering ideas, investigating ongoing R&Ds and looking at the most promising directions for the future

- 10 meetings so far (since the creation of TF in February 2019) with presentations on refactoring Geant4, Fast Simulation and use of compute accelerators
  - inviting both members of Geant4 collaboration as well as users to discuss R&Ds with potential interest for simulation

Three Main Axes of Development

- **Improve, optimise and modernise** the existing Geant4 code to gain in performance for the detailed simulation
  - Re-structure the code to make possible major changes (task-oriented concurrency, specialisation of the physics, better data formats, etc.)



- Trade precision for performance using **fast simulation techniques** both with parameterisations and with ML methods, and integrate them seamlessly in Geant4
  - Use detailed simulation to 'train networks' or to 'fit parameters' that later can deliver approximative detector responses well integrated within Geant4
- Investigate the **use of accelerators** such as GPUs
  - With novel approaches for organising the computational work

Web page

GEANT4
A SIMULATION TOOLKIT

Download | User Forum
Contact Us | Bug Reports

Geant4

Home » Collaboration

Task Force for R&Ds

The role of Task Force for R&Ds is to

- Promote longer-term R&D efforts on the exploitation of emerging technologies, computing architectures or software architectural revisions, and new or better physics ideas that would be beneficial to Geant4.
- Make timely assessments on these R&Ds for their feasibility, benefits and required efforts.

The Task Force has identified three development axis that should be followed in order to evolve Geant4 to meet the requirements of the forthcoming experiments:

- Improvement, optimization, modernization and refactoring of the existing Geant4 code
- Development and integration of fast simulation techniques
- Investigation of the potential use of accelerators

Several R&D activities have started exploring the potential evolution in those three areas and currently they include:

- Geant4 task-based prototype
- Stateless Geant4 prototype
- Single precision usage in simulation components
- Instruction and data cache optimizations
- Alternative e-/e+ and gamma transport simulation highly specialised for HEP detector simulations

- Electromagnetic shower parametrisation
- Machine Learning-based fast simulation tools
- Validation tools for fast simulation of electromagnetic showers

- GPU-based simplified simulation prototypes
- GPU vendor libraries in particle transport
- Portability frameworks for accelerator-based particle transport
- Celeritas - Exascale particle transport prototype
- Full integration of Opticks with Geant4
- AdePT - Accelerated demonstrator of electromagnetic Particle Transport

The meetings of the Task Force are open to the public and contributions are welcome.

Task Force coordinator: Witek Pokorski (CERN)
Deputy Task Force coordinator: Jonathan R. Madsen (NERSC/LBL)

- several R&D projects proposed
  - covering 3-axes mentioned
  - ranging from ‘preliminary investigations’ to ‘entering production’ (tasking, Opticks)
- some will require consolidation after the first phase of prototyping

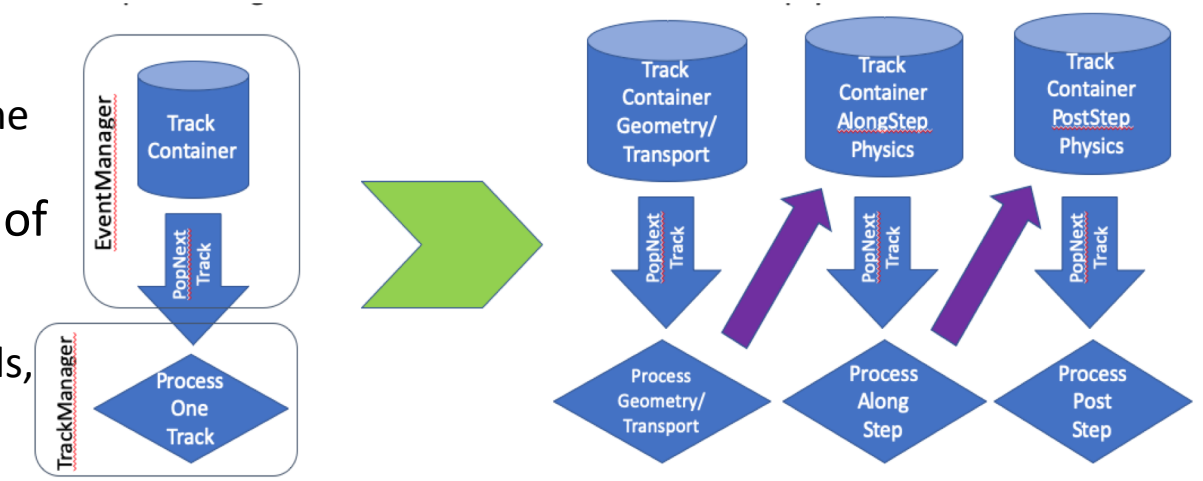
Improve, optimise and modernise Geant4 architecture

- some selected topics
  - Task-based framework – entering production (see dedicated session later today by Jonathan)
  - Instruction and data cache optimisations – several fixes and improvements in the code and developed new performance regression testing tools (see presentation tomorrow by Guilherme)
  - ‘stateless Geant4’ prototype – demonstrated the feasibility of architectural changes to explore fine-grained parallelism (see next slide and presentation last Tuesday)

Stateless Geant4 prototype

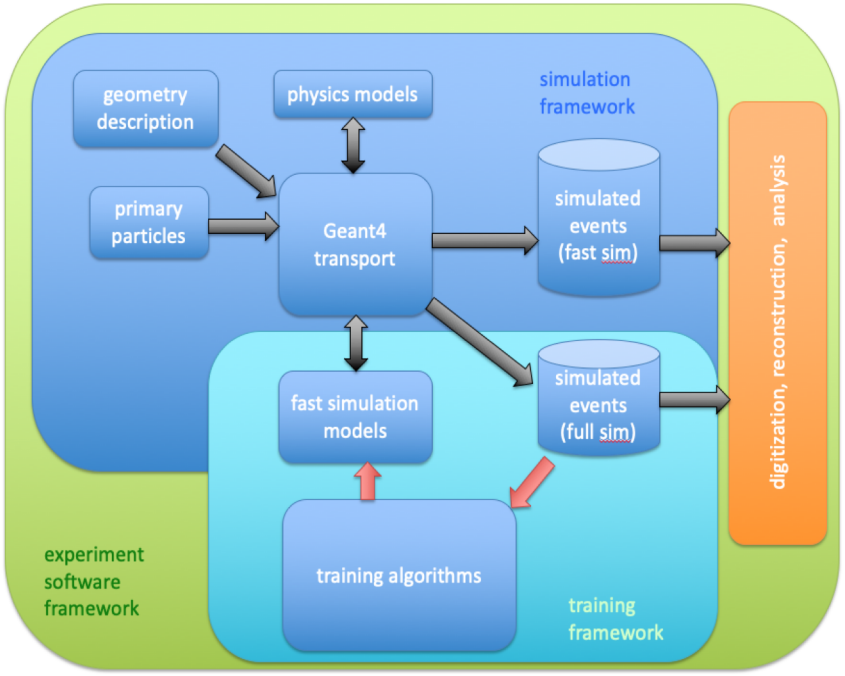
- to explore benefits of instruction-level locality and vectorization, sub-step-level parallelism needs to be possible
  - to group tracks ‘doing the same thing’
  - requires Geant4 engine to be ‘stateless’ (not keeping state of the current track)
- R&D work demonstrated the necessary changes in the code to achieve this
  - removing ‘state’ information from managers, navigators and some processes and attaching it to the G4Track
  - splitting stepping into several stages
  - introducing containers for each stage
    - detailed report presented at R&D parallel session last Tuesday

- changes are substantial, but feasible
  - first step could be making G4Navigator stateless (in line with the plans for Geometry WG)
- could allow potential performance improvements of certain computational phase by locality and vectorization
  - magnetic field propagation, MSC, some physics models, etc



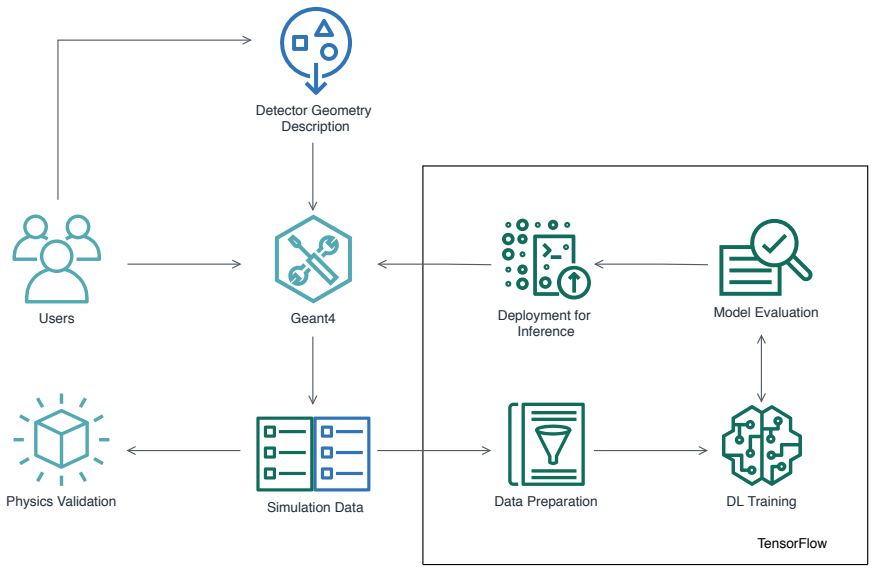
Fast Simulation

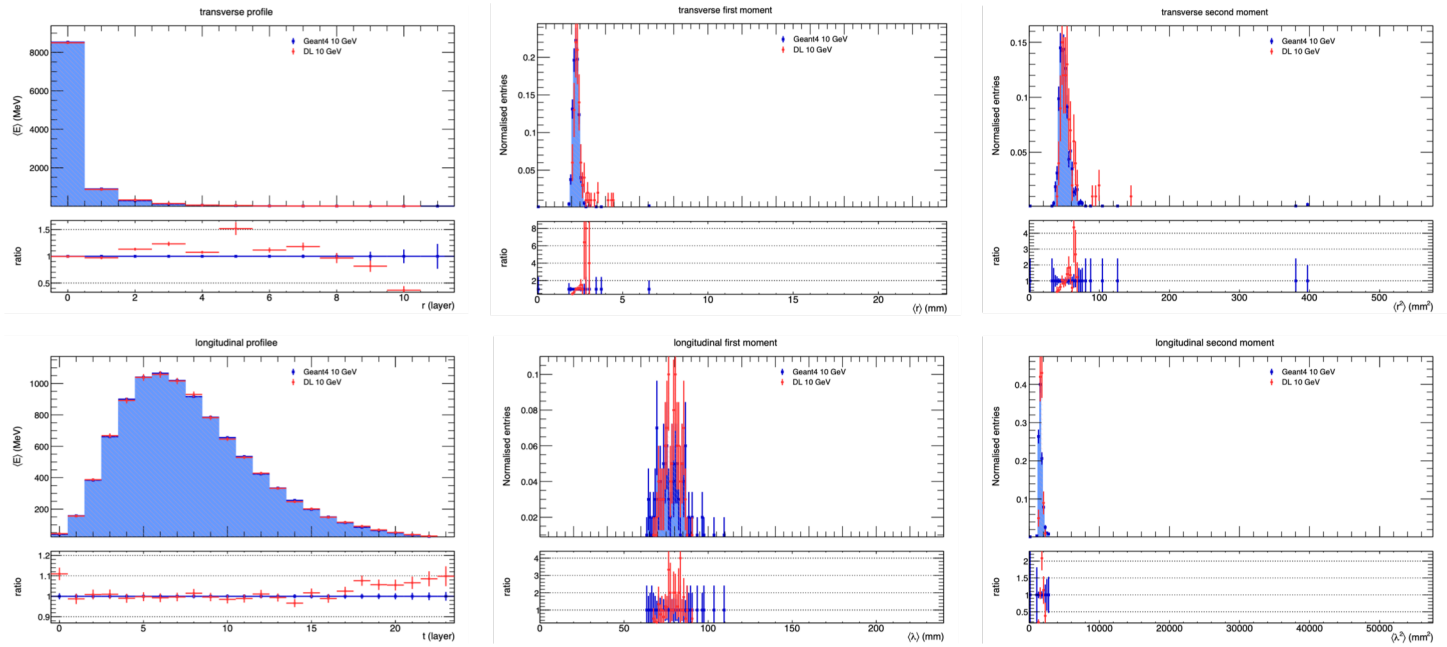
- ‘classical’
  - revisited GFlash
  - worked on automation of EM shower parametrisation - tuning of parameters for users’ geometry
    - see talk by Anna on Thursday
- Machine Learning-based fast-simulation tools
  - new interesting avenue
  - using trained generative models to obtain detector response
  - encouraging first results but a general, re-usable tools not yet there
    - detailed physics validation required



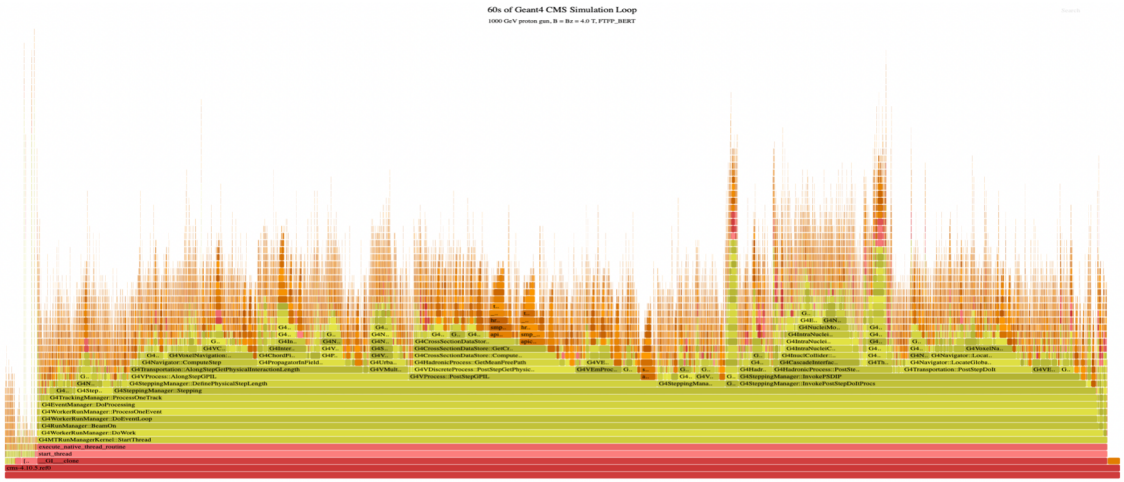
Machine Learning-based Fast Simulation

- implemented a full cycle system for Geant4 integration of Fast Simulation Deep Learning utilities, from data production through inference integration and results validation
- developed auto-regressive network for calorimeter shower simulation
  - very promising first results
  - further validation and generalization ongoing
- see Ioana's presentation last Tuesday





Investigate Use of Accelerators



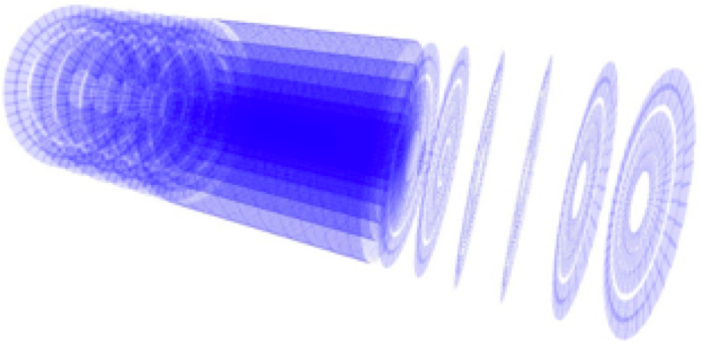
- need to efficiently use the evolving hardware
- general HEP simulation code is **not a natural candidate to run on GPUs**
  - Large complex codes, computation spread in many areas, many branches and special cases
  - Work needed to be done not known a priori (stochasticity)
  - Amdahl's law not helping here (any sequential code in CPU will limit maximum speedup)
- some successes on reduced and simpler problems
  - Low energy electromagnetic, medical app. with simple geometries, neutrons transport, optical photons, etc.
- there may be other alternatives (non-GPU based HPC), but GPUs are certainly widely available
  - Pressure from funding agencies to make efficient use of large HPC installations
  - We won’t get the necessary speedup by running on CPUs
- big issue on the sustainability of the code
  - No standard GPU programming language

GPU R&Ds

- Goal: transform the very heterogenous Geant4 HEP particle transportation into a more homogenous computational problem
  - impractical to ‘port’ full Geant4 in its current form to GPU
- Development of representative ‘demonstrators’
  - Before embarking in a massive conversion and re-engineering of large parts of the code we need to demonstrate its feasibility
- Study sustainability aspects of the code and the use of portability libraries
  - Gain experience in using libraries such as Kokkos, Alpaka, etc.

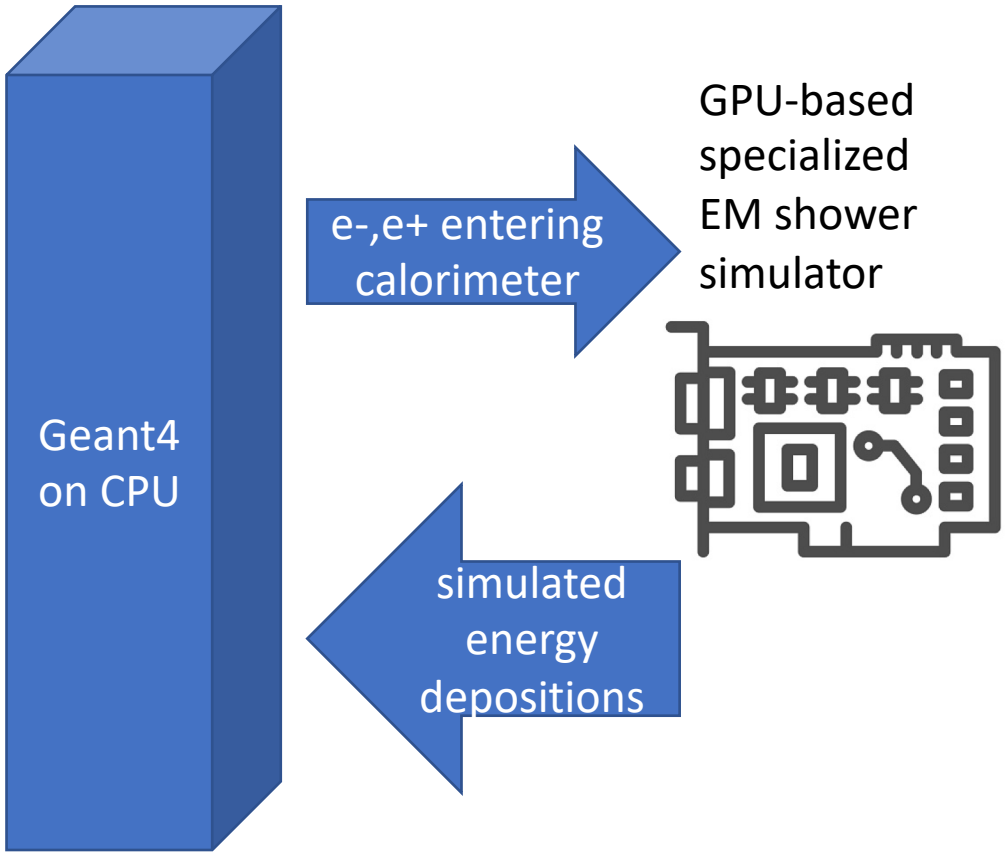
Specific investigations

- Opticks – very successful application of Nvidia Optix library to optical photon simulation (see talk at R&D parallel session last Tuesday by Hans)
  - can we use similar approach (vendor libraries) for more general simulation (e+, e-, gamma)?
- VecGeom – candidate modeler available both on CPU and GPU
  - A demonstrator ray-tracing utility using arbitrary geometry was developed (see talk at R&D parallel session last Tuesday by Andrei)
    - Work started to make geometry efficient for simulation on GPU
      - smaller navigation state caching transformation matrices
      - single precision navigation
    - Need for navigator class optimized for GPU
- AdePT
  - Accelerated demonstrator of electromagnetic Particle Transport
    - talk by Andrei last Tuesday



AdePT - Accelerated demonstrator of electromagnetic Particle Transport

- significant amount of work need to be performed in one go on the device due to high cost of gathering and copying data from CPU to GPU
- decided to focus on prototyping specialized GPU code to perform electromagnetic shower simulation in a calorimeter
  - specialized set of physics models and geometry
  - pre-defined scoring capabilities
- Geant4 would off-load simulation of EM showers coming from specific particles to the GPU library
  - similar concept to 'fast-simulation' processes, but doing full simulation on GPU
- dedicated GitHub [repository]
  - starting with CUDA utilities for track data handling



Summary

- R&D Task Force playing an important role in establishing communication between different R&D initiatives
  - providing forum for discussion and setting the direction
- several projects ongoing in the three axes
  - improvements to the current Geant4 code base
  - Fast Simulation
  - use of accelerators
- aiming at consolidating the efforts and developing demonstrators of new technologies to be integrated into Geant4

Final thoughts on R&D

- recognizing R&D activities is of crucial importance
  - both for the Collaboration as well as for the potential contributors and funding agencies
- by endorsing specific R&D projects, the Collaboration could really help in achieving progress and would profit from it
  - some activities, without a more formal endorsement, cannot even start due to lack of effort
- it would be mutually beneficial if Collaborating Institutes could have effort explicitly dedicated to supported R&D activities