



# Simulation – Geant4

Alberto Ribon
CERN EP-SFT
*On behalf of the Simulation team*

# Outline

- Three parts:
  - **1. Introduction** (1 slide)
    - Accuracy and Speed
  - **2. Towards Geant4 version 11.4** (3 slides)
    - Expected for December 5th
  - **3. Today's main focus** (9 slides)
    - G4HepEm, VecGeom, AdePT

# Introduction : Accuracy and Speed

- Simulations contribute to the **systematic uncertainties** of physics analyses – therefore simulations must be **accurate and fast**
- Effort for improving the accuracy of simulations, regardless of the speed, is always worth, even if not used in production
  - To investigate disagreements between simulation and experimental data
  - To tune or train fast simulation approaches
- Ways for improving the computing performances of simulations
  - By reviewing implementations, algorithms and approaches
    - Without compromising the physics accuracy : *e.g.* **G4HepEm** – more later
  - By using fast simulation solutions, *i.e.* simplifications with reduced accuracy
    - Many approximations are possible, matching different requirements and use-cases
    - Traditionally based on shower parameterisations; now based also on Machine Learning (ML)
  - By exploiting new technologies
    - *E.g.* GPU accelerators : *e.g.* **AdePT** & Celeritas – more later

# Status : Geant4 release 11.4 (expected for December 5th )

# Geometry and Transportation

- Improved robustness and performance of selected Geant4 solids
  - Reviewed implementation of selected solids in Geant4 for LHC production use
- Extensions and improvements of parallel geometry initialisation
  - Requested by CMS
- Optimisation of QSS (Quantized State Simulation) field driver
  - Improved robustness and speed of this alternative magnetic field driver

# Electromagnetic physics

- Getting close to the parallelisation of physics initialisation
  - With the parallelisation of geometry initialisation already achieved in G4 11.3, and improved in 11.4, this final milestone will complete the parallelisation of the initialisation
  - Requested by CMS
- Option to enable/disable fluctuations in the energy depositions per G4Region
  - Requested by ATLAS
- New approach to Cerenkov and Scintillation generation of optical photons
  - Two goals: flexibility in the usage of alternative models; easier offloading to GPUs
- Stable physics results
  - In particular for electromagnetic showers

# Hadronic physics

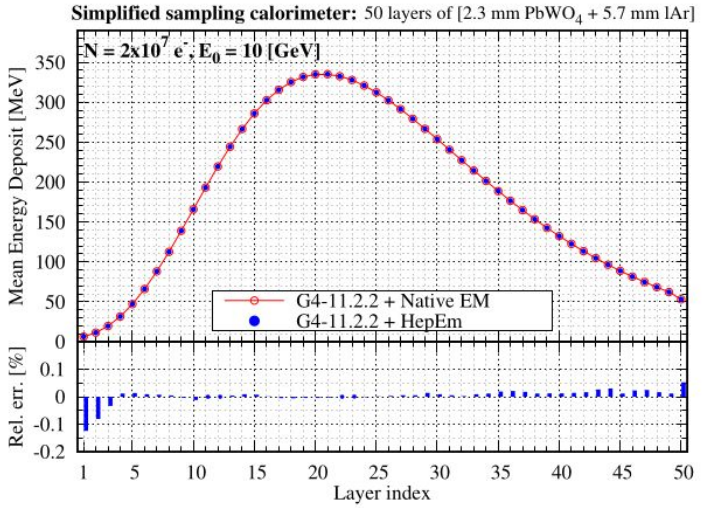
- Introduced option to roll back Bertini (BERT) model as in version G4 11.2
  - For the stability of hadronic showers, to ease migration of experiments to newer versions
- Tagging tracks produced by Fritiof (FTF) quasi-elastic or diffraction reactions
  - Requested by several experiments and users. It does not affect any physics results !
- Improved simulation of the charged exchange reaction
  - Requested by the NA64 experiment
- Improved simulation of low-energy nuclear physics
  - Revision of the Statistical Multi-Fragmentation (SMF) model
    - And its use before PreCompound model, to handle excited fragments produced by either hadronic string models or intra-nuclear cascade models
  - Alternative, new implementation of the Fermi-Break-Up (FBU) model
  - Alternative, new implementation of the Generalized Evaporation Model (GEM)

# Today’s main focus

# G4HepEm : from Conception to Production

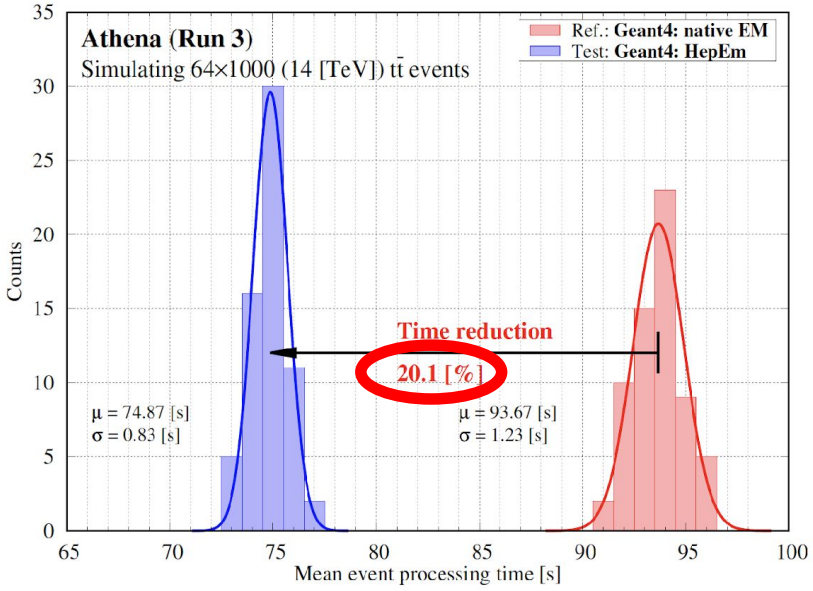
- Compact electromagnetic physics run-time library specialised for electron, positron and gamma, for HEP applications, running on both CPU and GPUs
  - Meant to reproduce the **same physics results as Geant4** default electromagnetic physics
    - G4 EM Opt0, used in all HEP experiments, *e.g.* in FTFP_BERT physics list
  - Seek speed-up with respect to Geant4 (on CPU) in two ways:
    1. Giving up the general and flexible capabilities of Geant4 electromagnetic physics
       - G4HepEm replaces the generic tracking loop and step computation of Geant4 with specialised ones for e±/γ, enabling the efficient implementation of further particle-specific algorithmic optimisations
    2. Efficient and streamlined implementation of the minimum physics sufficient for e±/γ in HEP
- Significant speed-up for EM showers in simplified calorimeters achieved in 2023
- Extended features for production-level readiness in 2024
  - Configuration per detector region; gamma/electron/positron-nuclear interactions; ”General Process”-like handling of cross sections; Woodcock tracking of gammas
- Integration in LHC experiments’ software framework reached in 2025
  - On-going “official” benchmarks on the grid; preliminary results by authors – see next slide
  - G4HepEm is the CMS default for the Run 3 production of 2026. For ATLAS, G4HepEm will be used in Run 4.

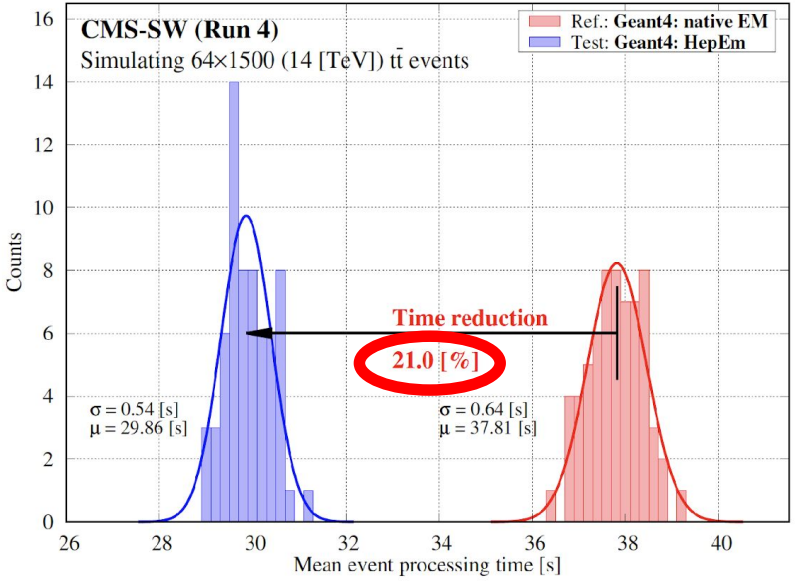
Courtesy of Mihaly Novak




| | | Geant4 EM | HepEm |
|---|---|---|---|
| $E_{dep}$ [MeV] | $PbWO_4$ | 6703.60 | 6708.67 |
| | lAr | 2535.42 | 2531.60 |
| #secondary | $\gamma$ | 4344.51 | 4347.29 |
| | $e^-$ | 1735.31 | 1734.87 |
| | $e^+$ | 425.97 | 426.05 |
| #steps | charged | 26932.3 | 12141.1 |
| | neutral | 40207.9 | 8586.5 |
| Time (60 threads) | | 2272.87 [s] | 1199.25 [s] |
| **Time reduction (speed-up)** | | | 47.27% (×1.9) |

# G4HepEm



# VecGeom : Surface geometry model

- The **surface geometry model** of VecGeom was developed with the hope that it could address GPU unfriendliness of the **solid geometry model**
  - No virtual function calls. No use of recursions. Less divergence due to fewer types. Less complex algorithms per surface.
- Unfortunately, this is not the case!
  - There are more objects. Still complex algorithms for handling traversals. Random memory access even worse due to more surfaces than solids. Expensive treatment of geometry overlaps.
  - No performance-breakthrough approach in sight, while feature completeness needs effort
- GPU thread divergence happens in the **BVH** (Bounded Volume Hierarchy) navigation, before the solid/surface is checked
  - A different navigation optimisation strategy must be investigated

*=> Giving up the initial plan of using VecGeom's surface geometry model on GPU, while keeping the solid geometry model only for CPU*

# VecGeom : Solid geometry model

- The solid geometry model is working
- The surface geometry model implemented important new features
  - These features can be ported in the solid geometry model, and some of them did already
- Targeting release **v2** in the coming weeks
  - With the surface geometry model as it is now (without the complete coverage on GPU as for the solid model), and with a partial clean-up and improvement of the solid model

=> The plan is to further clean-up and improve the **solid geometry model**

- The process can be incremental (with a few large chunks, *e.g.* removing virtual functions)
- Back-port the good ideas from the surface geometry model, *e.g.* handling of Boolean operations without recursion (*i.e.* linearized Boolean treatment)
- Improve performance where it matters most, *i.e.* the accelerating structure : implement the voxels-like navigation optimisation as in Geant4
- Remove explicit CUDA dependencies and implement a portability layer (*e.g.* Alpaka)

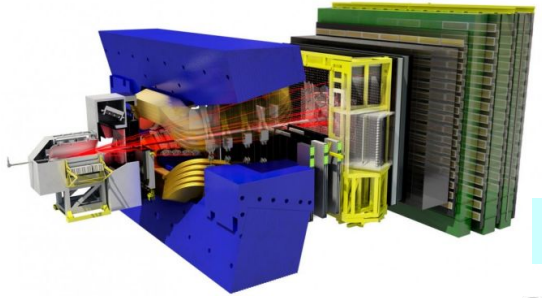
# AdePT : Simulation on GPU

- Goal: offloading to GPU the simulation of electrons, gammas and positrons
  - To gain in throughput, and possibly also in energy efficiency
  - The simulation of EM showers take a large fraction of the computing time of the LHC experiments
    - ~50-80%, with a potential (theoretical maximum) speed-up of 2-5 times
- AdePT is a lightweight Geant4 plugin
  - Components: 1.Core : kernel scheduling, particle tracking
    2. Physics : **G4HepEm**
    3. Geometry + Navigation : **VecGeom**
    4. Magnetic field integrator : Dormand-Prince
       - Using the **covfie** library for 3D field maps
- Differences in tracking on CPU and GPU
  - Geant4 tracks **sequentially** : one track is transported from beginning to end, in steps; then another track is transported, *etc.*
  - AdePT tracks **in parallel** : several tracks are transported in parallel, in steps
  - Parallel tracking requires different MC truth handling in experiments' software framework

# AdePT : Results in Gauss (LHCb)

Geometry: full LHCb detector
Field: 3D field map (v6r1 down)
Input: 20000 min bias events
GPU: **Nvidia RTX4090**
CPU: AMD Ryzen 9 **16 cores**

Full production settings except MC truth



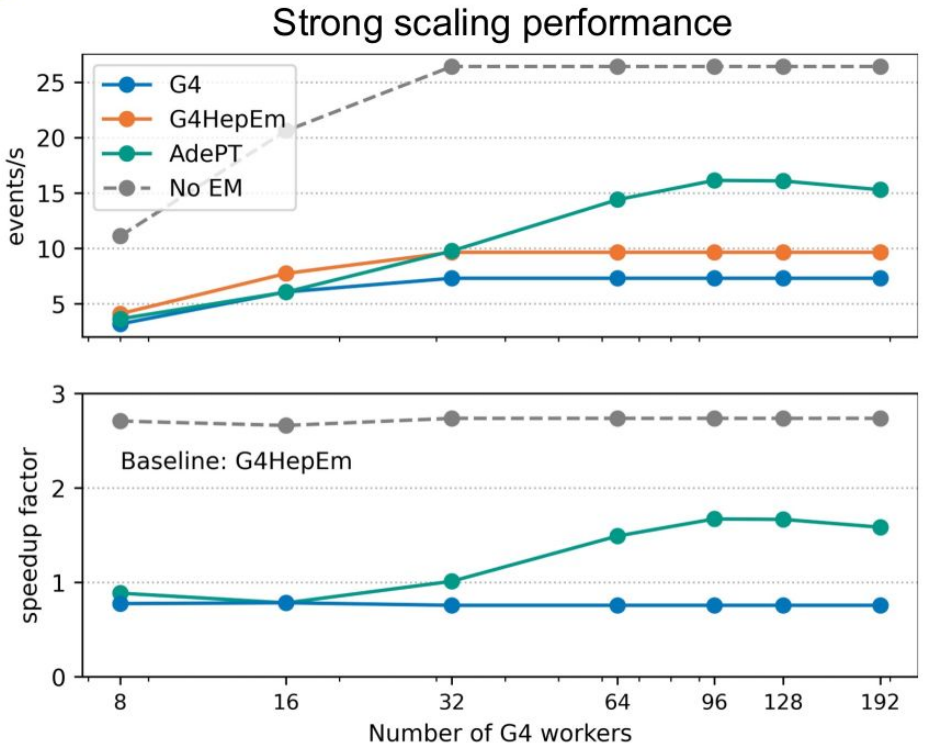
Courtesy of Severin Diederichs

**Full production settings except MC truth**

The GPU is **underutilized** in minimum bias events in Gauss

By oversubscribing the CPU, the GPU can be filled!

**1.7x** speedup w.r.t. G4HepEm, **2.1x** w.r.t. Geant4

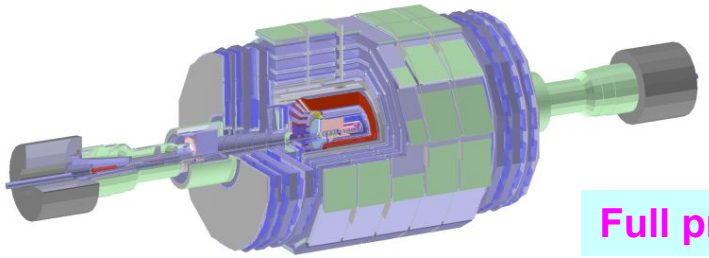


# AdePT : Results in CMSSW (CMS)

Application: CaloSimHitStudy
Geometry: full CMS detector,
Run 4 geometry (D110)
Field: 3D field map interpolated
Input: 50k ttbar events

Full production settings except MC truth

Offloading to GPU:
Everything except HCAL



Courtesy of Severin Diederichs

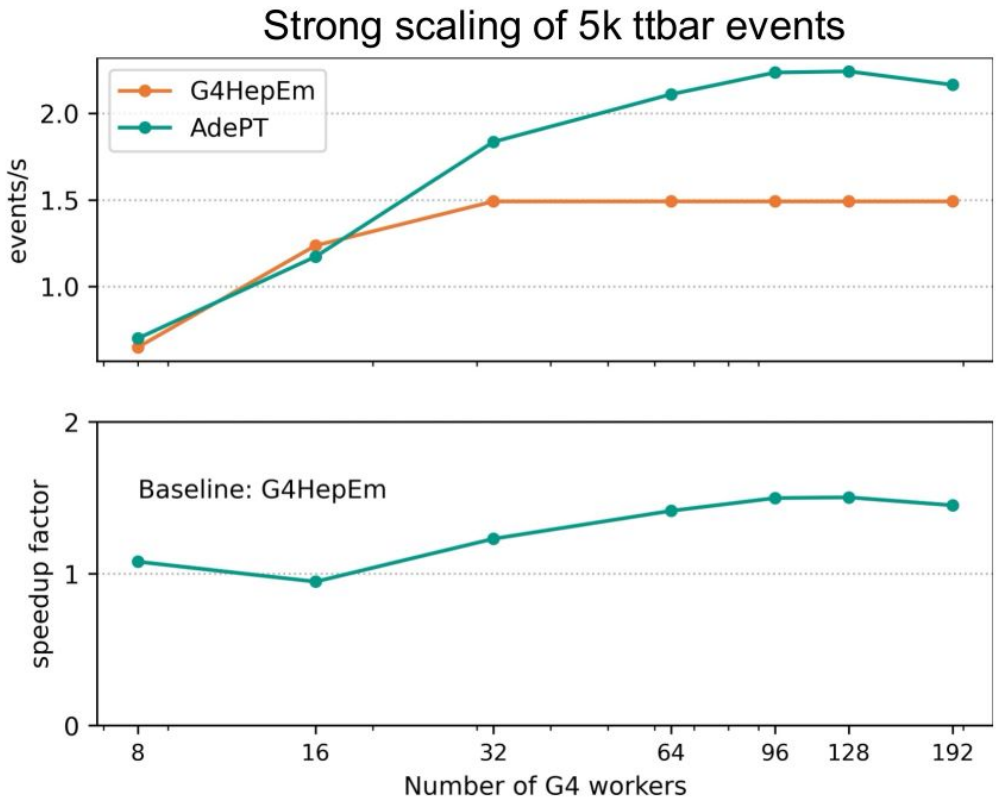
**Full production settings except MC truth**

**Hardware:**
GPU: **Nvidia RTX4090**
CPU: AMD Ryzen 9 **16 cores**

Current speedup:
**1.5x at 96 threads**

Large fraction of gammas (~2/3)
Woodcock tracking will be added

Note: changing cuts in HGCAL will influence run time – unclear in which direction

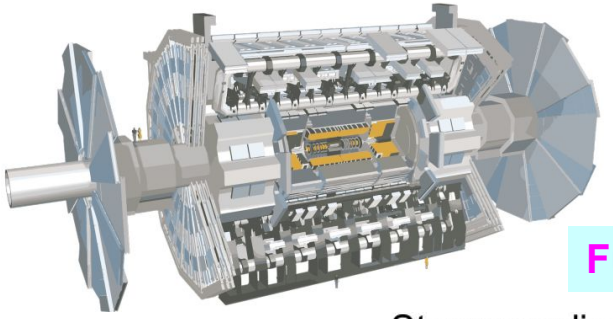


# AdePT : Results in Athena (ATLAS)

Geometry: full ATLAS detector,
Run 4 geometry
Field: 3D field map interpolated
Input: 10000 ttbar events

Full production settings except MC truth

Offloading to GPU:
EMEC, EMB, HEC, PreSampLAr



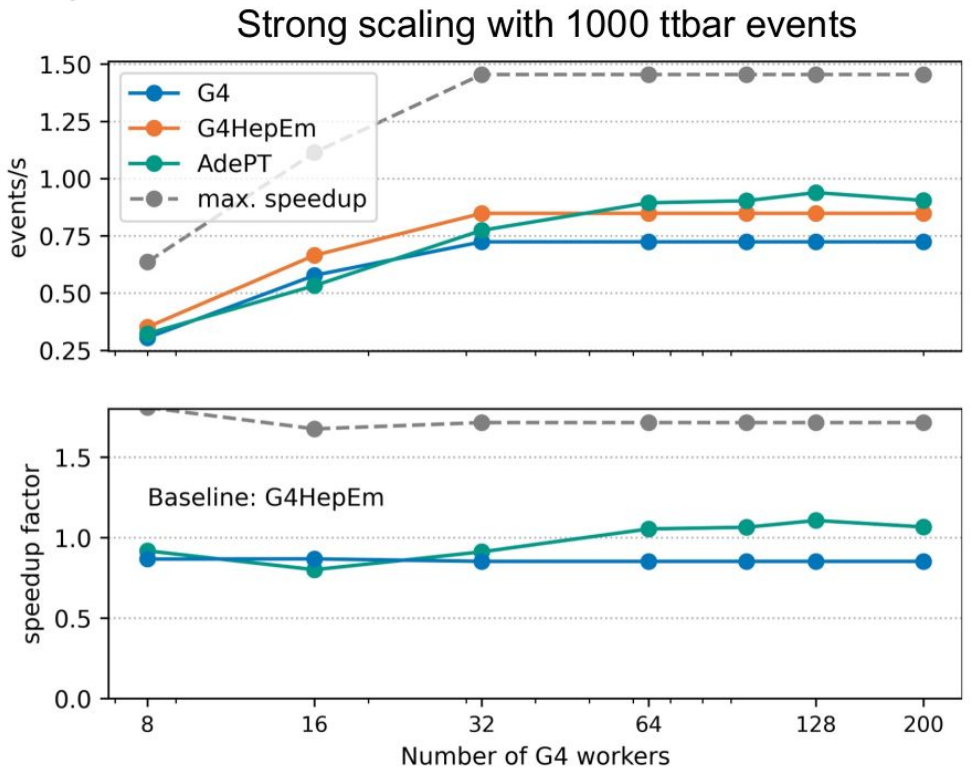
Courtesy of Severin Diederichs

**Full production settings except MC truth**

**Hardware:**
GPU: **Nvidia RTX4090**
CPU: AMD Ryzen 9 **16 cores**

Total Run 4 EM fraction: ~77%
Max. speedup factor: ~4.4x

Offloading to GPU:
EMEC, EMB, HEC, PreSampLAr
Max. speedup factor: **1.7x**
**Achieved: 1.1x**



# AdePT : Summary & Outlook

- Significant progress this year : integration into experiments' software framework is on-going, now it is possible to run **simulations close to production**
  - Excellent physics agreement between Geant4 on CPU and AdePT has been achieved
  - Still missing the handling of MC truth information
- **Performance results encouraging** and can be further improved
  - **Woodcook tracking**
    - Specialised tracking for gammas in granular calorimeters (*e.g.* ATLAS EMEC, CMS HGCal)
    - Already used in ATLAS EMEC on CPU in G4HepEm; to be implemented on GPU
  - **Geometry optimisations**
    - See previous slides on VecGeom
- Further steps
  - AdePT is currently CUDA-only : portability could be achieved via Alpaka (following CMS)
  - Energy efficiency needs to be assessed