Update on R&D activities

Witek Pokorski

12.12.2024

Geant4 Technical Forum

Content

- exploration of new hardware (GPUs)
  - general transport code prototypes– AdePT, Celeritas
  - domain specific application – Opticks/CaTS

Note: ML fast simulation not covered here – proposal to have a dedicated talk at the next Technical Forum

AdePT Project Targets

- Understand usability of GPUs for general particle transport simulation
  - Seeking potential speed up and/or usage of available GPU resources for HEP simulation
- Provide GPU-friendly simulation components
  - Physics, geometry, field, but also data model and workflow
- Integrate in a hybrid CPU-GPU Geant4 workflow

AdePT Recent developments

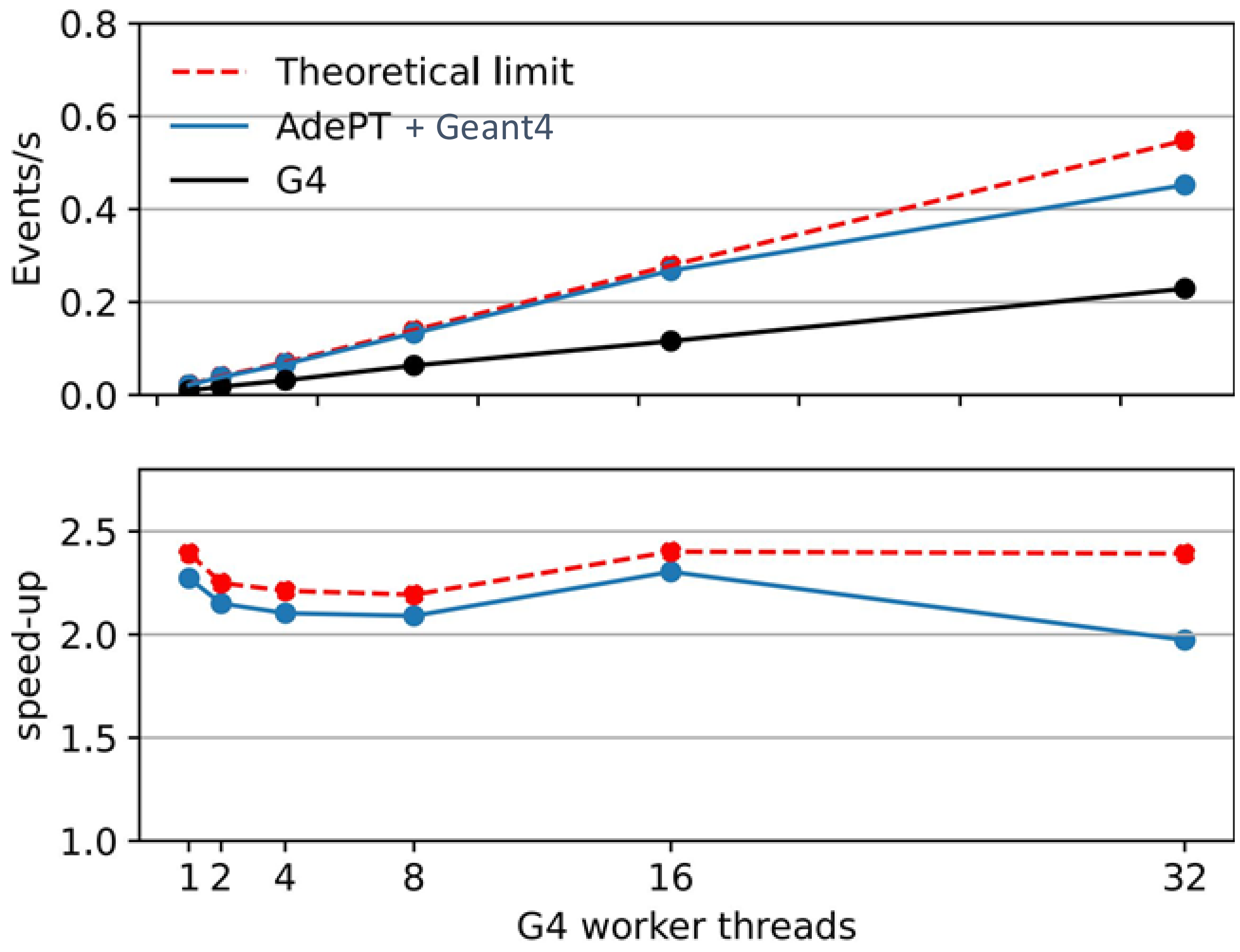
- New method for Geant4 integration
- New method for scoring
- Refactoring of AdePT into a library
  - There has been a major refactoring to reorganise the project into a library, simplifying integration into external applications
  - Example integration with the GeantVal HGCAL Test Beam app
- Gaussino Integration
- Progress towards an asynchronous AdePT backend
- Major development in VecGeom's Surface geometry model

AdePT - Geant4 Integration

- Integration using specialized G4VTrackingManager
  - Attaches AdePT to all EM particles as a process
  - Fully customizable
    - Possible filters based on region, energy, etc
- User only needs to register the AdePTPhysicsConstructor in their physics list (1 Line!)
- AdePT can be further configured through an API or macro commands
- Scoring: sending back hit information and calling the user-defined sensitive detector code on CPU
  - Sensitive volume information taken from the geometry
  - GPU hit information is used to reconstruct G4 steps
  - **No changes** to the user SD code are needed

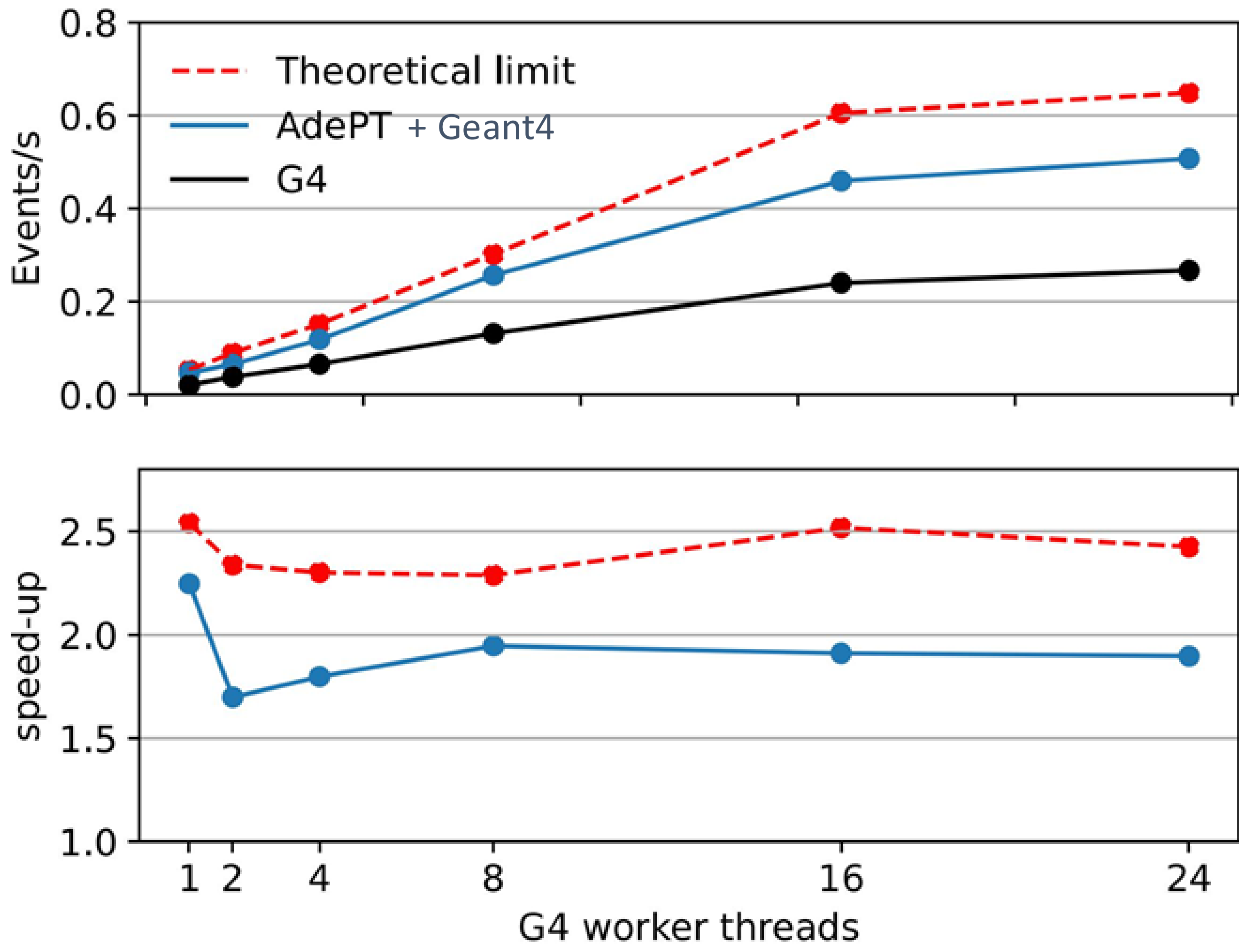
AdePT: current performance results

GPU: **Nvidia A100**
CPU: AMD EPYC 7752 96 cores
Input: 4 TTBar per thread
Geometry: CMS2018
No magnetic field
Simple scoring of energy deposition



AdePT: current performance results

GPU: **Nvidia RTX4090**
CPU: AMD Ryzen 9 16 cores
Input: 4 TTBar per thread
Geometry: CMS2018
No magnetic field
Simple scoring of energy deposition



Potential usage of consumer-grade GPUs?

LHCb’s Gaussino framework integration

- Allows to configure and to steer the different phases of detector simulation
- Provides wrappers for the Geant4 physics constructors and allows to build the Geant4 modular physics list using a simple Python configuration
- Gaussino has now been extended with such a wrapper for the AdePTPhysicsConstructor

```
GaussinoSimulation(
    PhysicsConstructors=[
        "GiGaMT_AdePTPhysics",
        "GiGaMT_G4EmStandardPhysics",
        "GiGaMT_G4EmExtraPhysics",
        "GiGaMT_G4DecayPhysics",
        "GiGaMT_G4HadronElasticPhysics",
        "GiGaMT_G4HadronPhysicsFTFP_BERT",
        "GiGaMT_G4StoppingPhysics",
        "GiGaMT_G4IonPhysics",
        "GiGaMT_G4NeutronTrackingCut"])
```

It can now be added to the simulation in a **single line!**

Gaussino Integration

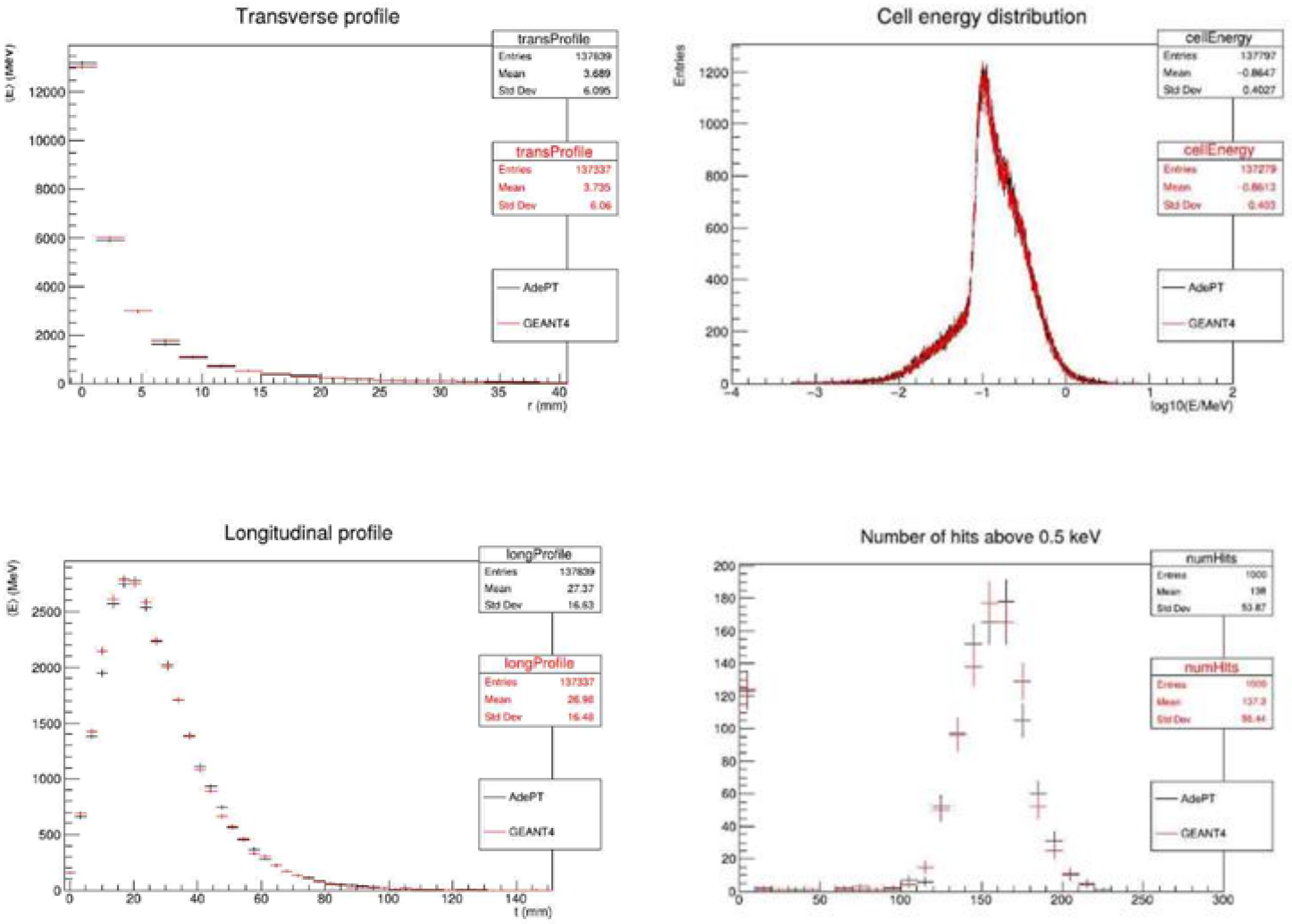
- Additional AdePT configuration can be passed through the Gaussino wrapper for Geant4 configuration macros

```
GiGaMTRunManagerFAC("GiGaMT.GiGaMTRunManagerFAC").InitCommands = [
    "/adept/setVecGeomGDML calochallenge.gdml",
    "/adept/addGPURegion CaloRegion" #"/adept/setTrackInAllRegions true"]
```

- Using the scoring mechanism discussed before
  - AdePT calls the appropriate Gaussino sensitive detectors to create hits as in a normal Geant4 simulation

Gaussino integration - Calo Challenge setup

- Gaussino - AdePT integration has been successfully used for the Calo Challenge setup
- Physics results show a good agreement with Geant4
- If there are enough particles sent to the GPU, the gains can be significant
    - Achieved 5x speedup with 4 CPU threads in initial tests with gamma-only events
    - this number would be lower for real events and complex detectors



Accelerating detector simulations with Celeritas

Performance improvements and new capabilities

Seth R Johnson

Celeritas Code Lead
Senior R&D Staff
Scalable Engineering Applications

Celeritas core team:

Elliott Biondo *(ORNL)*, Julien Esseiva *(LBNL)*, Hayden Hollenbeck *(UVA)*, Seth R Johnson *(ORNL)*, Soon Yung Jun *(FNAL)*, Guilherme Lima *(FNAL)*, Amanda Lund *(ANL)*, Ben Morgan *(U Warwick)*, Stefano Tognini *(ORNL)*

Celeritas core advisors:

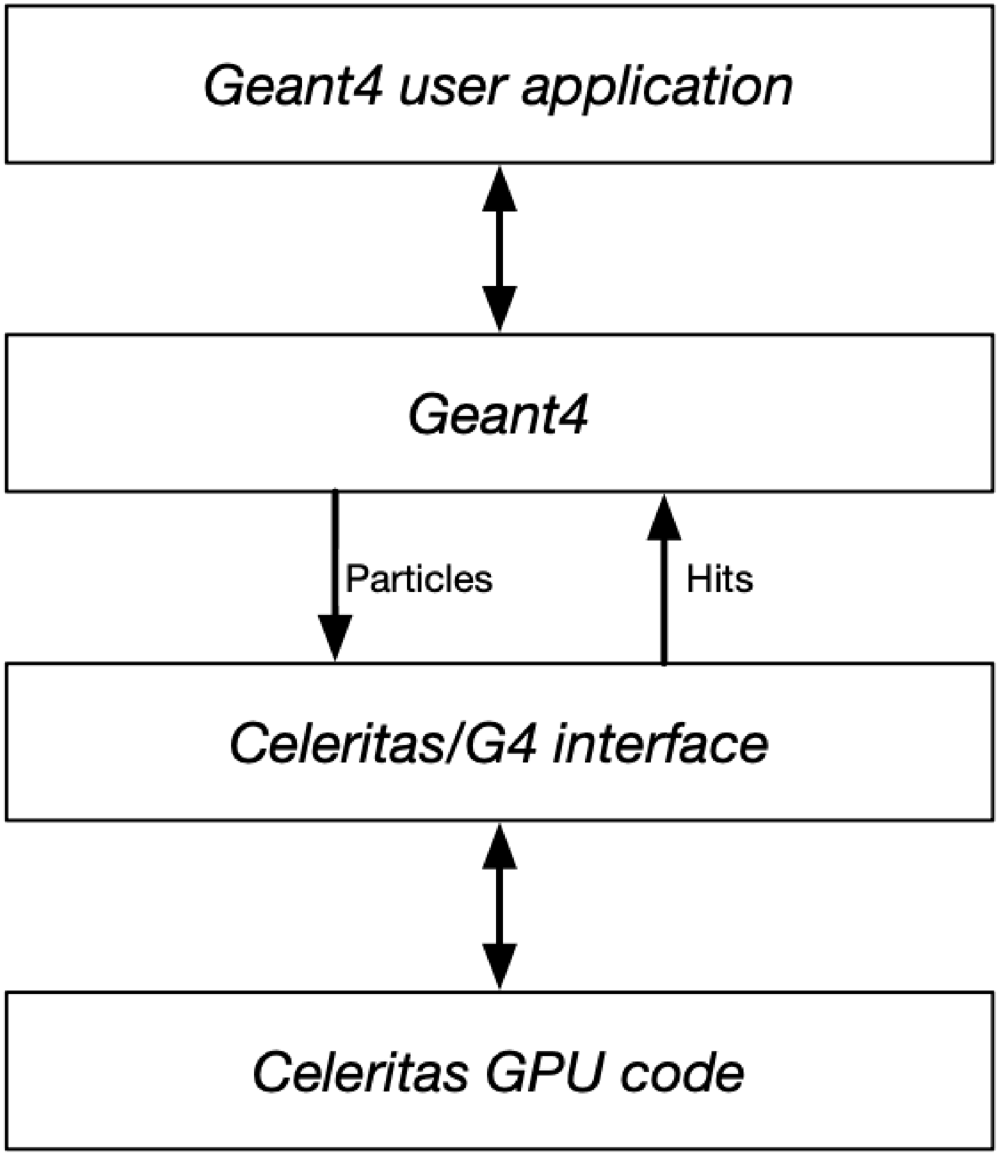
Tom Evans *(ORNL)*,
Philippe Canal *(FNAL)*,
Marcel Demarteau *(ORNL)*,
Paul Romano *(ANL)*

CELERITAS

CHEP
21 October, 2024

Geant4 integration

- Celeritas directly imports G4 data
  - Physics options and cross sections
  - Geometry translation to VecGeom/ORANGE
- e^-, e^+, γ sent to Celeritas (GPU)
  - Tracking manager or “user” offload
  - Tracks queued till buffer capacity/end of event
- Reconstructed “hits” *(energy deposition plus metadata)* sent back to user-defined detectors



Correctness and performance

1. Verify with basic EM tests

- ✓ Infinite medium
- ✓ Simplified calorimeter
- ✓ Thin transmission tests

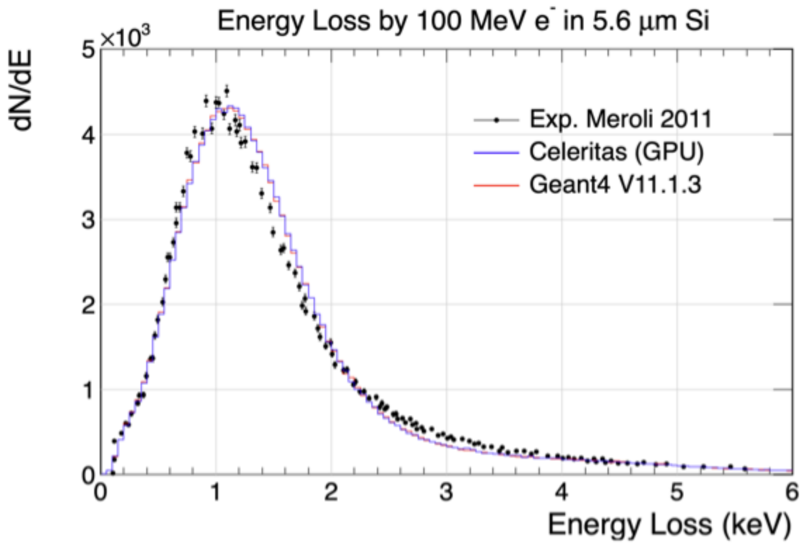
2. Assess standalone EM performance

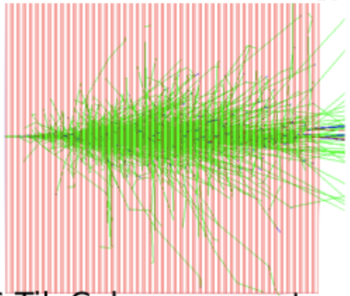
- ✓ Simplified calorimeter
- ✓ Subdetector components
- ✓ Full CMS Run 2

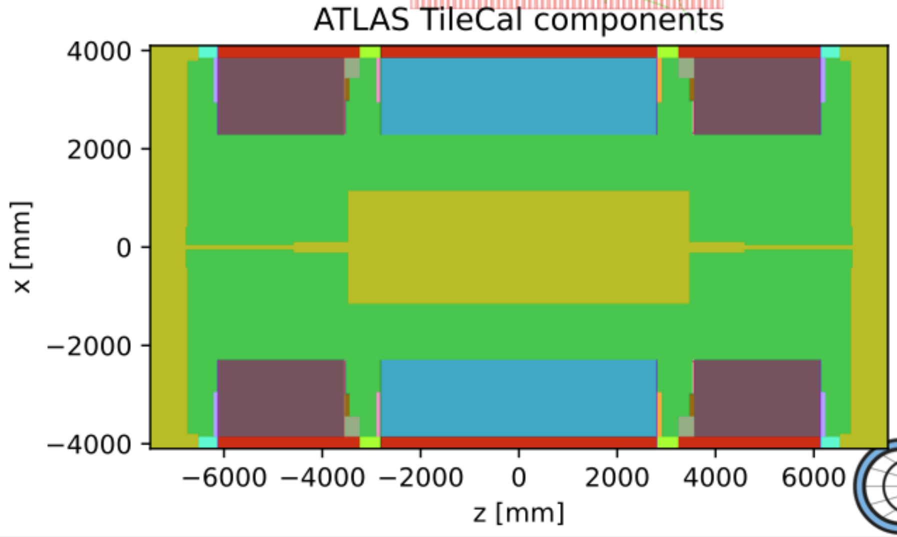
3. Validate with Geant4 hadronics and SDs

- ✓ Test beams: ATLAS TileCal and CMS HGCal
- ✓ Prototype CMSSW integration
- • ATLAS subdetectors and reconstruction*

* ATLAS/Celeritas/AdePT working meeting last October

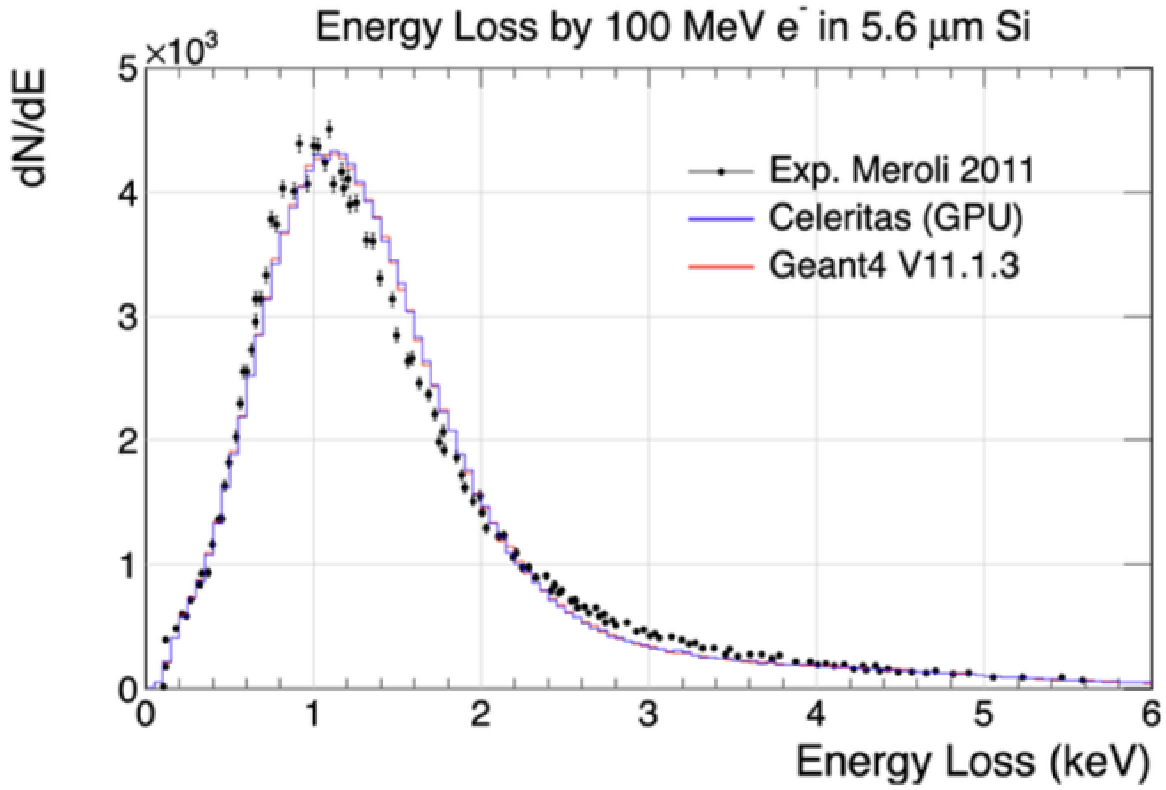


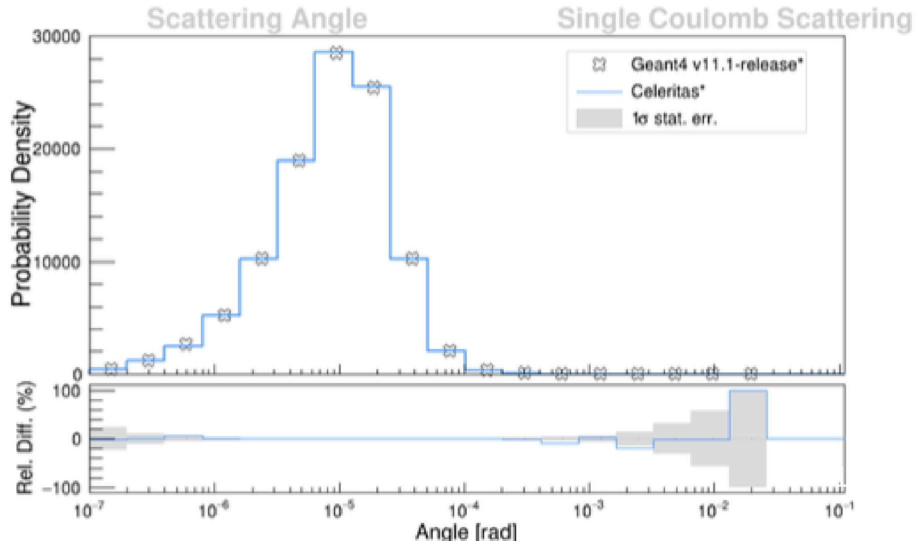




EM physics validation

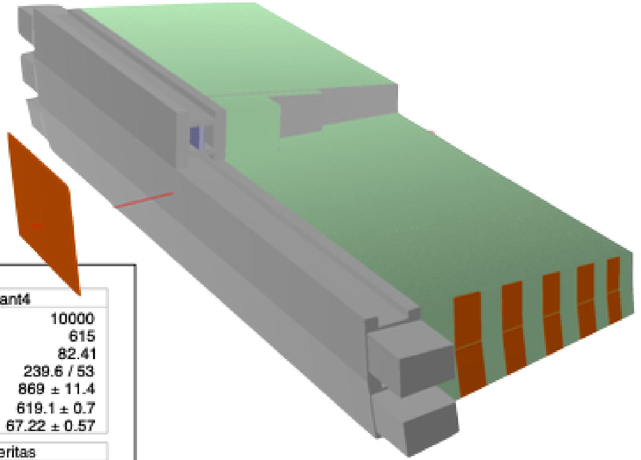
- Established good agreement with Geant4 for:
  - Energy loss fluctuations
  - Multiple scattering (azimuthal and longitudinal angle distribution)
  - Single Coulomb scattering
- **Experiment-specific validation required for acceptance**
  - GPU and CPU produce equivalent results
  - Recent work improves single-track performance for testing

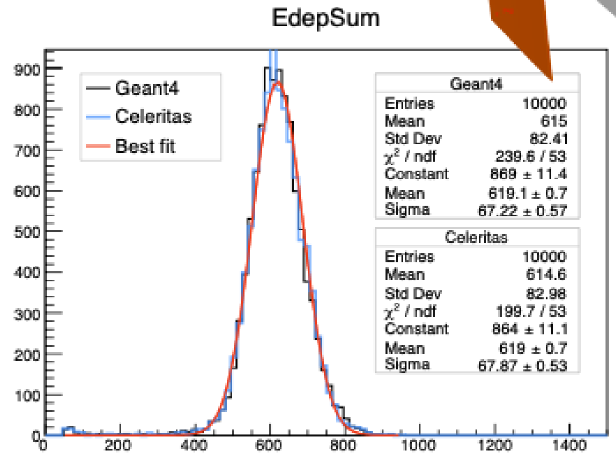


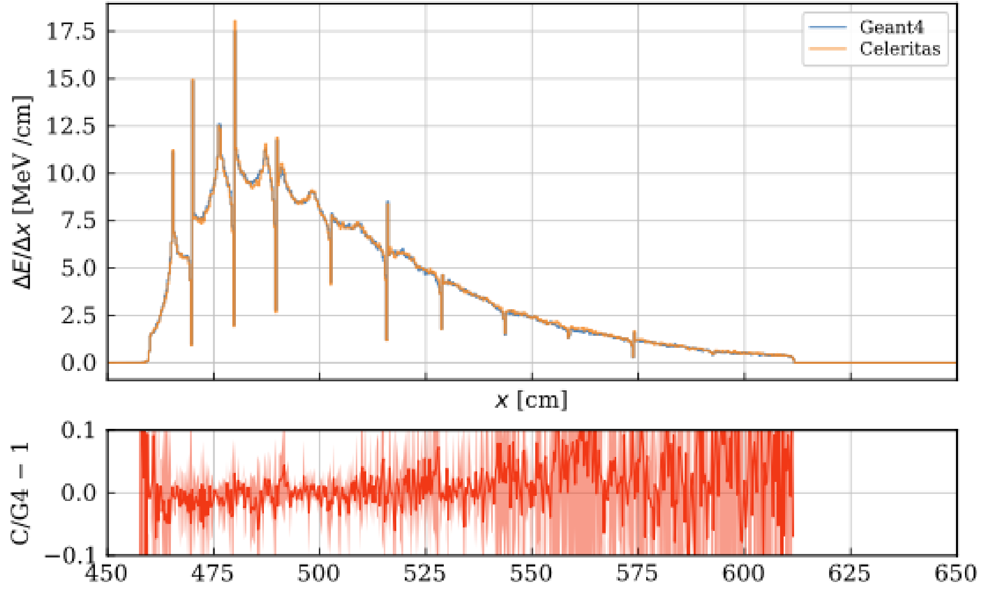


EM offloading with FullSimLight

- ATLAS FullSimLight: hadronic tile calorimeter module segment
  - 64 segments in full ATLAS, 2 in this test beam
  - 18 GeV π^+ beam, 64 per event, 32 events, no field
  - FTFP_BERT (default) physics list
    (includes standard EM)
- **~100 lines of code to integrate**
  - Offload e^-, e^+, γ to Celeritas
  - Celeritas reconstructs hits and sends to user-defined `G4VSensitiveDetector`
- **Good agreement** in energy deposition





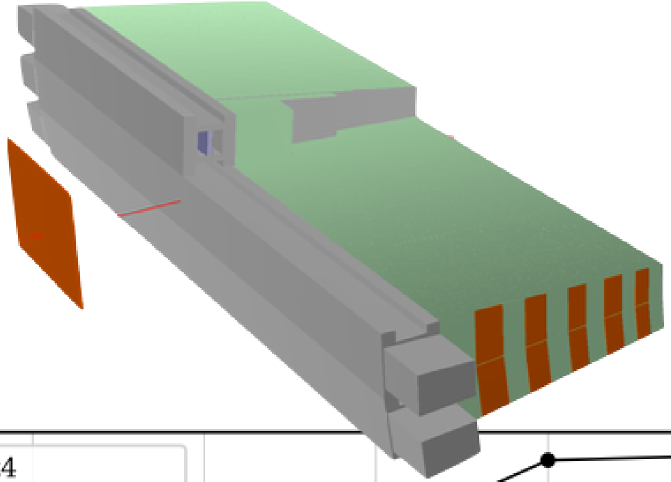


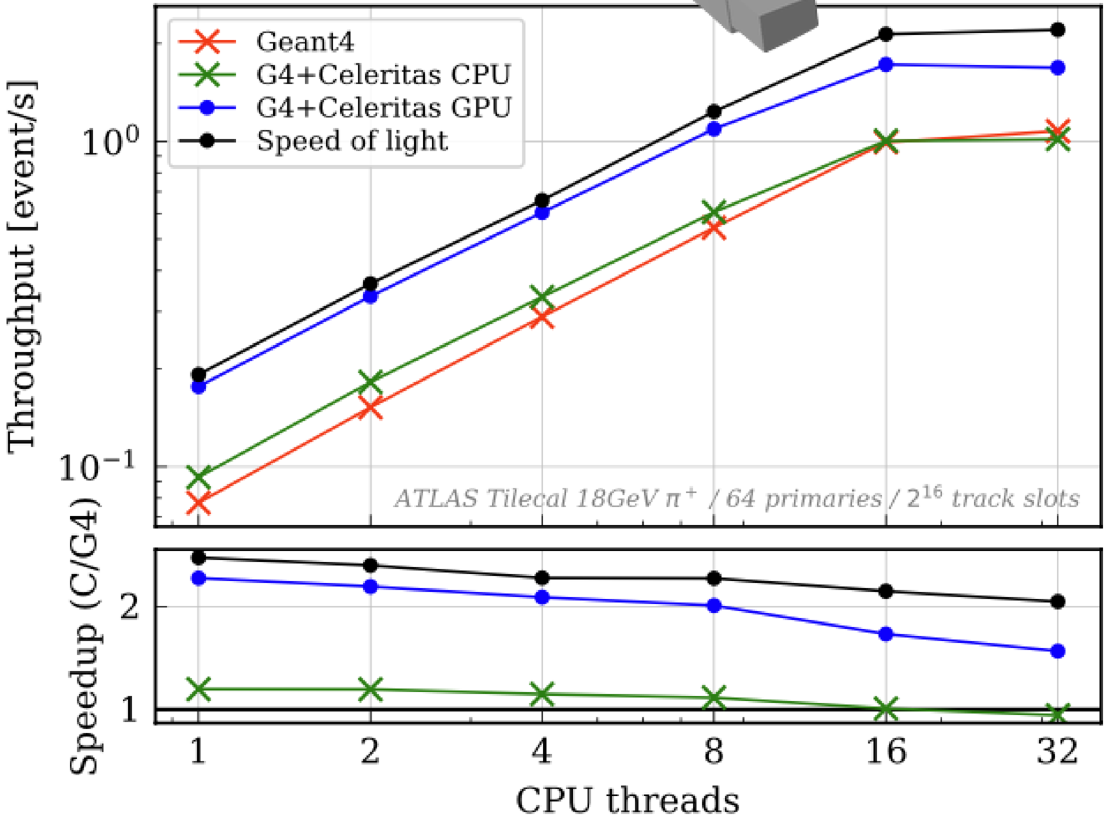
yz-integrated energy deposition

Source: Johnson, S.R. Geant4 Meeting 2023

Offload performance results

- 1/4 of a Perlmutter (NERSC) GPU node
  16 cores of AMD EPYC, 1 Nvidia A100
- Time includes startup overhead, Geant4 hadronic physics, track reconstruction, and SD callback
- GPU speedup: **1.5–2.3×** at full occupancy
  Using all CPU cores with a single GPU
- CPU-only speedup: **up to 1.2×**
- Theoretical maximum speedup: **2.0–2.5×**
  Instantly killing e-, e+, γ when born
- **One GPU is effective with many-CPU Geant4**





Note: higher workload *(primaries per CPU thread)* **increases** Celeritas performance

Opticks/CaTS and Integration to Geant4 and LArSoft

Hans Wenzel

FERMILAB-SLIDES-24-02

Opticks/G4CXOpticks/CaTS: and Integration with LArSoft

Opticks is an open source project that accelerates optical photon simulation by integrating NVIDIA GPU ray tracing, accessed via NVIDIA OptiX. Developed by Simon Blyth: https://bitbucket.org/simoncblyth/opticks/

It was developed for the Juno experiment→ Bound to NVIDIA hardware and software.

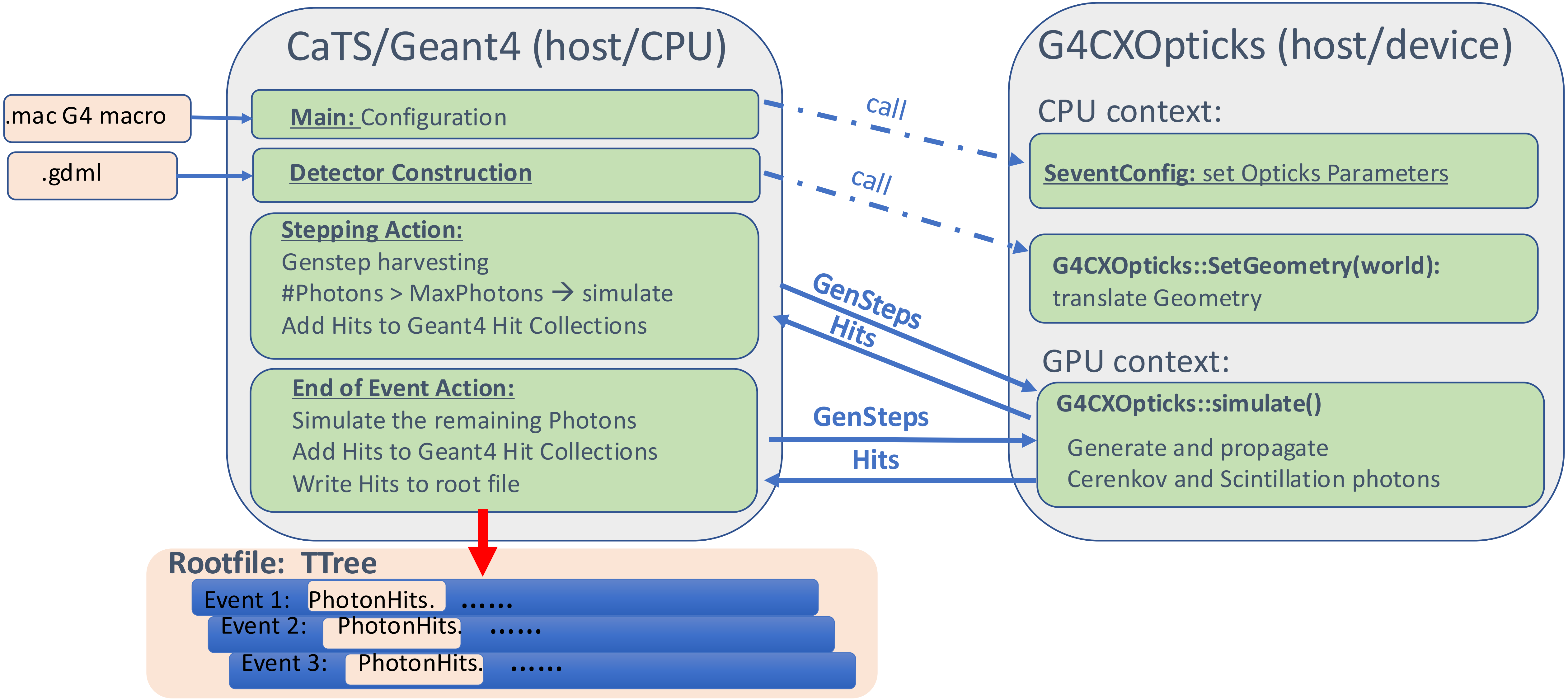
CaTS (Calorimeter and Tracker Simulation) is a flexible and extend-able framework. With respect to Opticks it interfaces Geant4 user code with Opticks and defines a hybrid workflow where generation and tracing of optical photons is optionally offloaded to Opticks (GPU) using the G4CXOpticks interface, while Geant4(CPU) handles all other particles.

CaTS an advanced Geant4 example based on legacy version of Opticks (based on Optix 6) was included in Geant4 11.0: https://geant4.kek.jp/lxr/source/examples/advanced/CaTS/ → not maintained, since Opticks based on Optix 6 is not maintained.

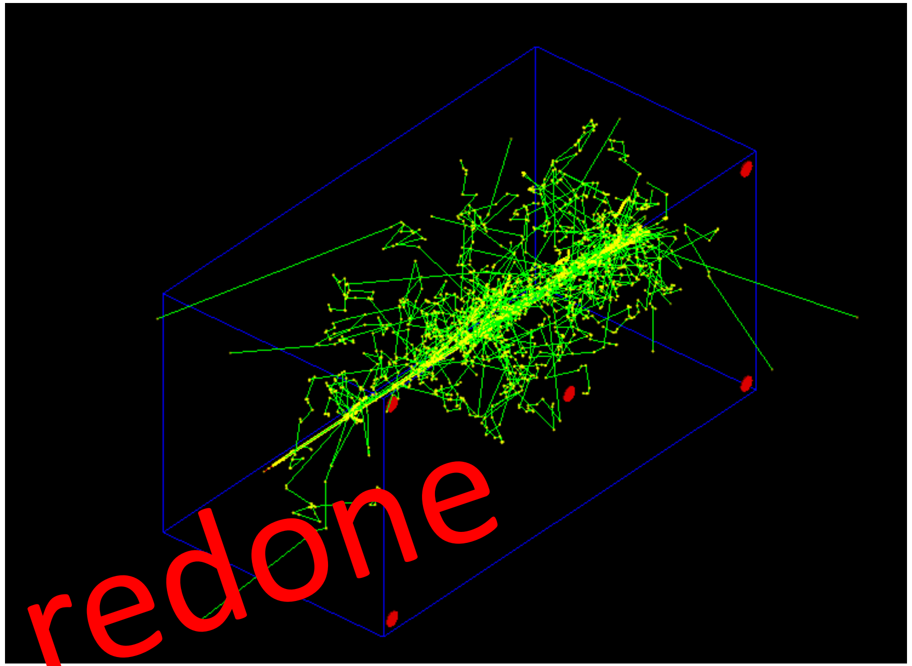
Status:

- **Opticks has been completely reengineered by Simon Blyth migrating to OptiX7.**
- **The new Opticks (NVIDIA OptiX7) has been fully functional (off and on) since January 2024.**
- **New Opticks APIs have been tested and successfully integrated with a modified workflow of CaTS!**
- **https://github.com/hanswenzel/CaTS. Optimization, physics validation and benchmarking of the new Opticks are ongoing!**

CaTS workflow using the new version of Opticks based on OptiX® 7:



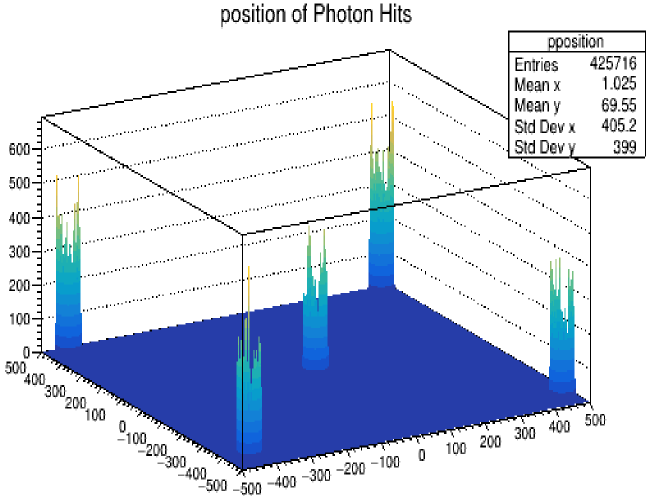
Performance: (Legacy Opticks, Needs to be redone)



Hardware:	
CPU	Intel® Core i9-10900k@ 3.7 GHz, 10 CPU cores
GPU	NVIDIA GeForce RTX 3090 @ 1.7 GHz, 10496 cores
Software:	
Geant4: 11.0, Opticks based on OptiX® 6	

To be redone

Number of CPU threads	Single threaded. Geant4 [sec/evt]	Opticks [sec/evt]	Gain/speed up
1	330	1.8	189x



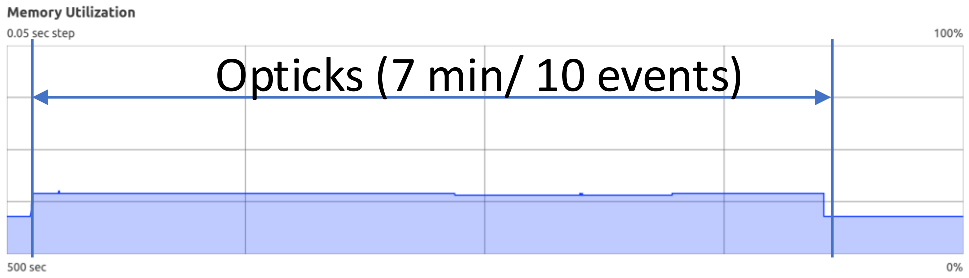
→It becomes feasible to run full optical simulation event by event! But comparison is to single threaded Geant4→ somehow unfair! Single geant4 threat can saturate the GPU and doesn't allow the use of multiple CPU cores.

Status

The new Opticks (with NVIDIA OptiX7) has been fully functional (on/off) since January 2024.
New Opticks APIs are tested and successfully integrated with a modified workflow of CaTS (v2xx).
. https://github.com/hanswenzel/CaTS . Physics validation and benchmarking of the new Opticks are ongoing!

With the legacy version of Opticks (based on NVIDIA Optix 6) we achieved speed up in the order of a few times 10^2, this depends strongly on detector geometry, hardware and settings.
So far with the new version of Opticks (same computing hardware and detector geometry, 2 GeV e-shower) we found:
Compared to the previous version, the speed up by GPU over CPU is only a factor 10 times compared to several 10^2 times previously. We found the GPU (computing and memory) resources are underutilized. The graphs below show that only 10% of the CPU and memory resources are utilized → the optimization of the GPU kernel launch promises to improve the performance significantly.
The work integrating Opticks with the liquid argon TPC software framework (LArSoft) is ongoing.

To be redone



Plans:

- Update the Geant4 advanced example CaTS to use the new Opticks. This will be part of the next release.
- Complete LArSoft integration.
- Provide detailed full-scale example of a liquid Argon Time projection Chamber (LArTPC) for optical simulation. (simple example exists).
- Provide detailed documentation of Geant4 optical physics processes and material properties (from literature) relevant to LArTPCs .

  “Liquid Argon optical properties for Geant4 Simulations”:

  DUNE-doc-31579 → make the information available to the Geant4 code base.
- Enhance Opticks functionality (implement Wavelength shifting process).
- CaTS: Improve RootIO using Root TBufferMerger when running in multi-threaded mode

Conclusion

- GPU protypes in the phase of integration tests with the experiments
  - first results of full integration with LHCb and ATLAS
  - delta assessment of the GPU prototypes planned for March 2025

- CaTS moved to new version of Opticks based on NVidia OptiX7
  - validation and benchmarking ongoing