AdePT/Celeritas Update

WARWICK
THE UNIVERSITY OF WARWICK

Ben Morgan

Detector Simulation Work and R&D

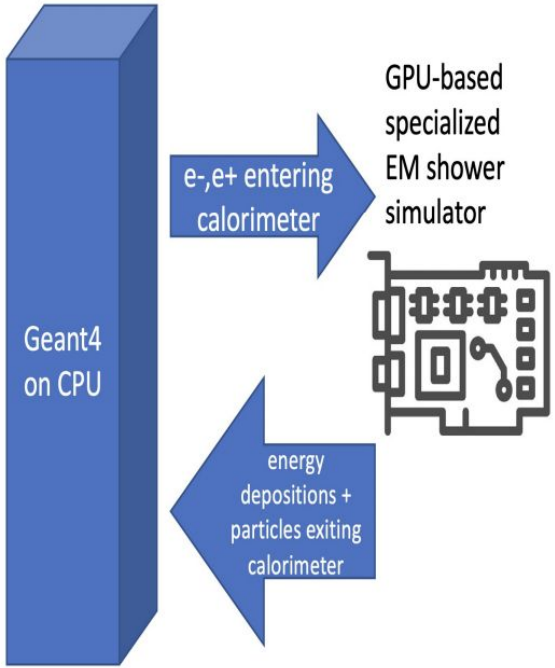
- Short overview of status today - follow links to recent presentations at Geant4 Collaboration Workshop, GPU R&D Review for in depth details
- **Geant4 Project (Worldwide)**
  - *https://geant4.web.cern.ch*
- **AdePT R&D Project (CERN-SFT)**
  - *https://github.com/apt-sim*
- **Celeritas R&D Project (ECP: ORNL, FNAL, Argonne, LBL)**
  - *https://github.com/celeritas-project*
- **Vecgeom/ORANGE Surface Based Geometry (CERN, Celeritas/ORNL)**
  - *https://gitlab.cern.ch/VecGeom/VecGeom (See surface_model branch)*
  - *https://github.com/celeritas-project/celeritas/tree/develop/src/orange*

Slight Aside: Geant4 Work in 2023/24

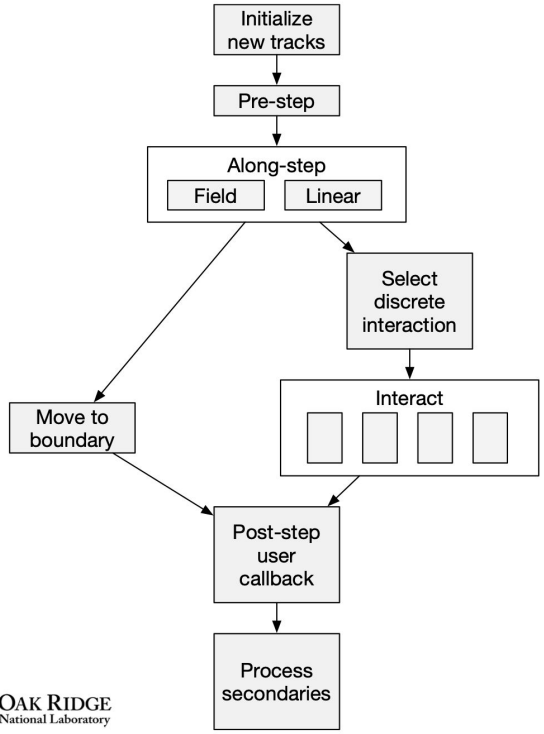
- Release 11.2 on 8th December 2023: Release Notes
  - *Minor release, with general improvements and fixes*
  - ***Significant UK input** on rollout of Qt6 and VTK visualization support (Cockroft, Warwick) despite minimal UK funding!*
- In 2024, UK (Warwick) expect to take leading role in Run/Event/Track **workflow management** topics:
  - *Implement any requirements coming from AdePT/Celeritas*
  - *Modernization/Sustainability of Multithreading/Tasking infrastructure (with CERN/JLAB)*
  - *Revival/Modernization of MPI support (with CERN)*

Geant4 R&D: e-/e+/γ Particle Transport on GPUs

- AdePT/Celeritas developing data structures, workflows to transport **EM particles($e^-/e^+/\gamma$)** using GPUs
  - *Goal: improve event/s/W throughput for LHC detector simulations*
  - ***Focus on EM showers as most expensive fraction of detector portion of "typical" production runs***
- Integrate with existing CPU Geant4 simulations by "offloading" EM particles to GPU, e.g. via "Fast Simulation" hooks in Geant4, with **main challenges**:
  - *Minimizing number/size of on/offload actions*
  - *Allow user-defined actions on GPU, such as scoring/hits*
  - *Synchronization between CPU/GPU (event boundaries)*
  - *Handing back particles (e.g. exiting particles, hadrons from photonuclear processes) from GPU to CPU*



Track-parallel Stepping Workflow on GPUs




```
extend_from_primaries          ▷ Copy primaries to device, create track initializers
while Tracks are alive do
  initialize_tracks            ▷ Create new tracks in empty slots
  pre_step                     ▷ Sample mean free path, calculate step limits
  along_step                   ▷ Propagation, slowing down
  boundary                     ▷ Cross a geometry boundary
  discrete_select              ▷ Discrete model selection
  launch_models                ▷ Launch interaction kernels for applicable models
  extend_from_secondaries      ▷ Create track initializers from secondaries
end while
```

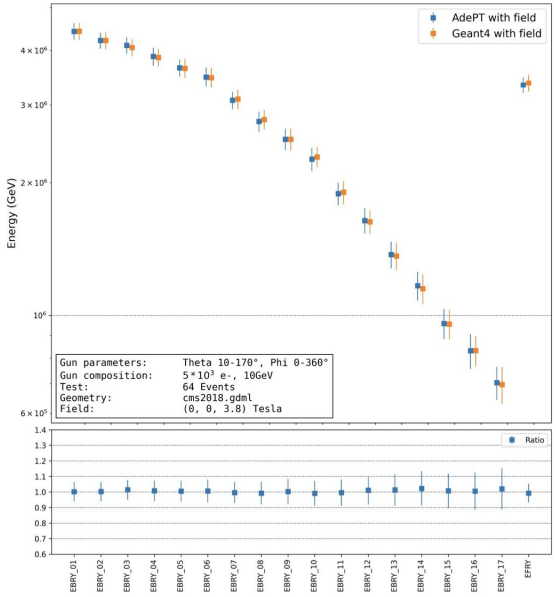
- CPU: parallel Events, ***sequential* Tracks**
- CPU+GPU: parallel Events, ***parallel* Tracks**
  - ***Action*** *based control flow*
  - *Kernels determine next* ***Action****, or perform an* ***Interaction***
  - *Example from Celeritas, AdePT's is similar though with larger, per-particle, kernels*

OAK RIDGE National Laboratory

Credit: Seth Johnson (ORNL)

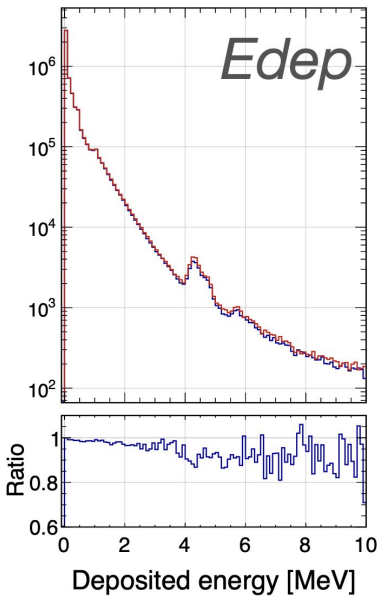
Physics Validation

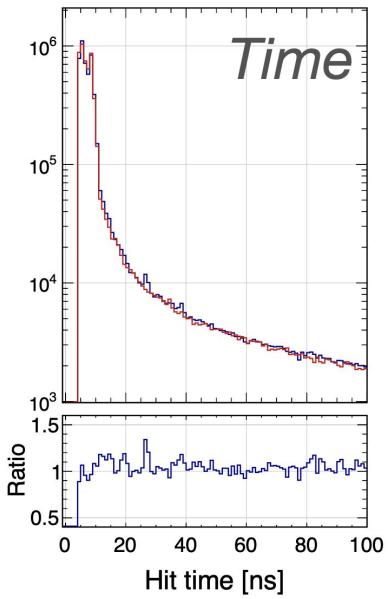
- G4HepEM in AdePT, CPU/GPU implementation of Geant4 models/data in Celeritas.
- ***Excellent agreement with Geant4, but studies ongoing across problem space***



Example: CMS ECAL
← *AdePT: 10GeV e⁻*

Celeritas: 14TeV $t\bar{t}$ →

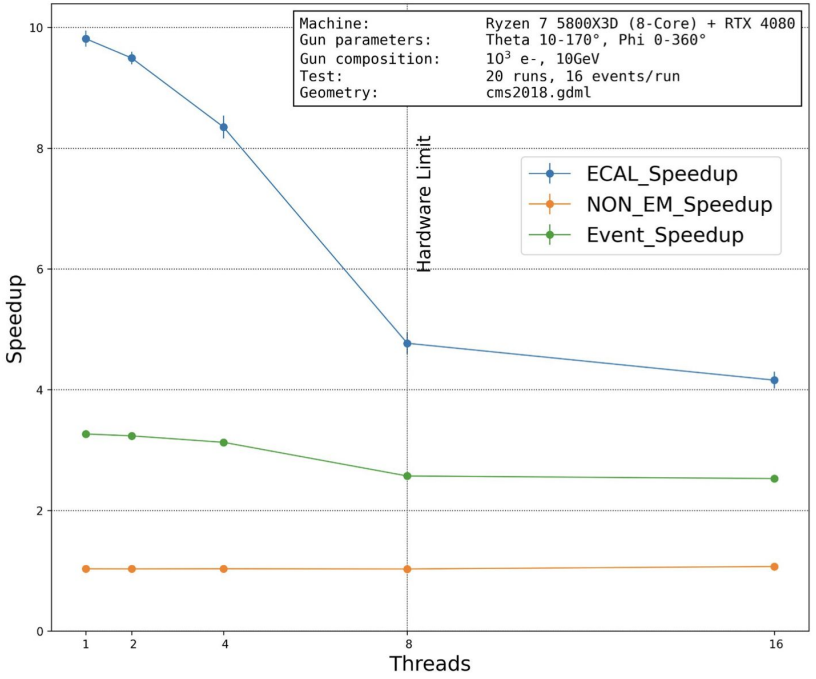




Credits: Jonas Hahnfeld (CERN), Seth Johnson (ORNL), Amanda Lund (Argonne)

Some Benchmarking Results: AdePT w/CMS2018

- Speedup of the ECAL simulation and overall event speedup (10GeV e-)
  - *AdePT does not affect the rest of the simulation, 1:1 ratio in the time spent outside ECAL*
- Vary number of Geant4 worker threads
  - *Decreasing AdePT speedup as the GPU becomes more saturated*

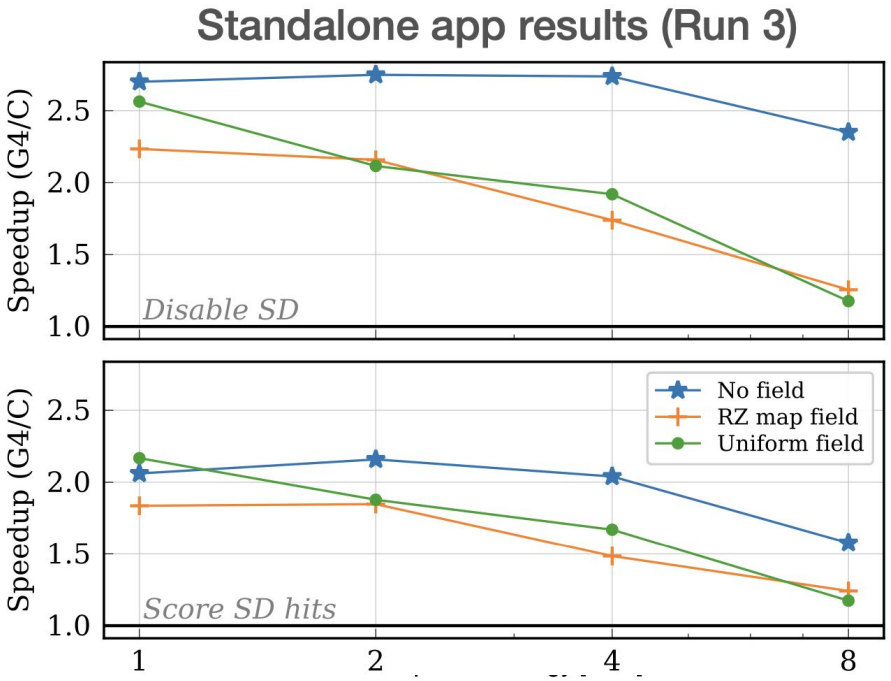


Speedup of the ECAL simulation and overall per-event speedup

Credit: [Jonas Hahnfeld (CERN)]

Some Benchmarking Results: Celeritas w/CMS Run3

- Initial performance comparison in standalone Geant4+Celeritas application
  - *CMS GDML geometry/Sensitive Detectors*
- 8CPU+1GPU standalone simulation with 14TeV tt **17-87% faster**
  - *Theoretical maximum speedup (all $e^-/e^+/g$ tracks take zero time) in full CMSSW ~230%*

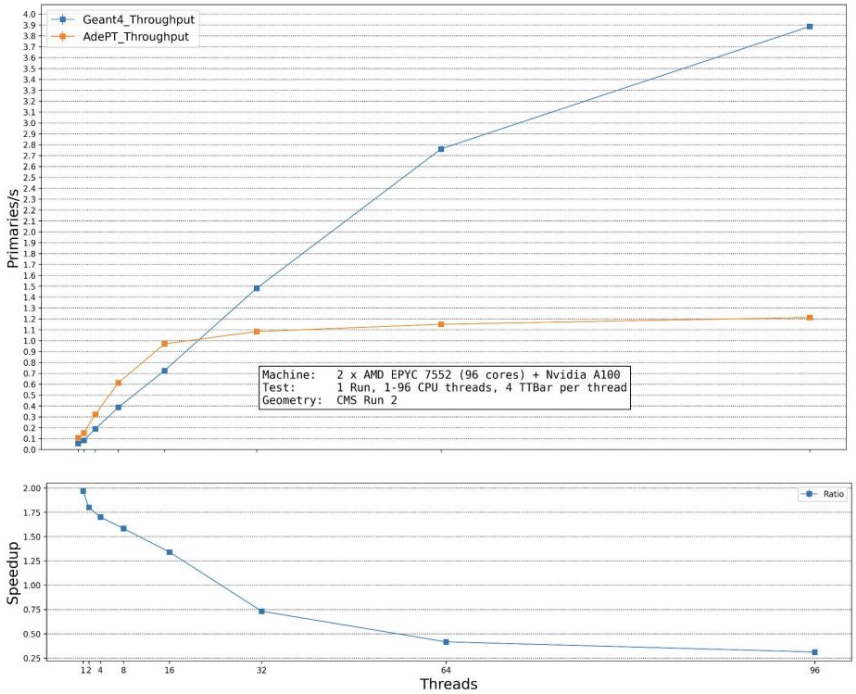


Hardware: Intel Xeon Gold 6152 CPU 22c 2.10GHz + NVIDIA Tesla V100 SXM2
Geometry: CMS detector (**Run 3 configuration**)
Input: 8 $t\bar{t}$ events @ 14 TeV from LHC pp collision

Credits:Seth Johnson (ORNL), Amanda Lund (Argonne)

Potential Optimization Strategies for Workflow

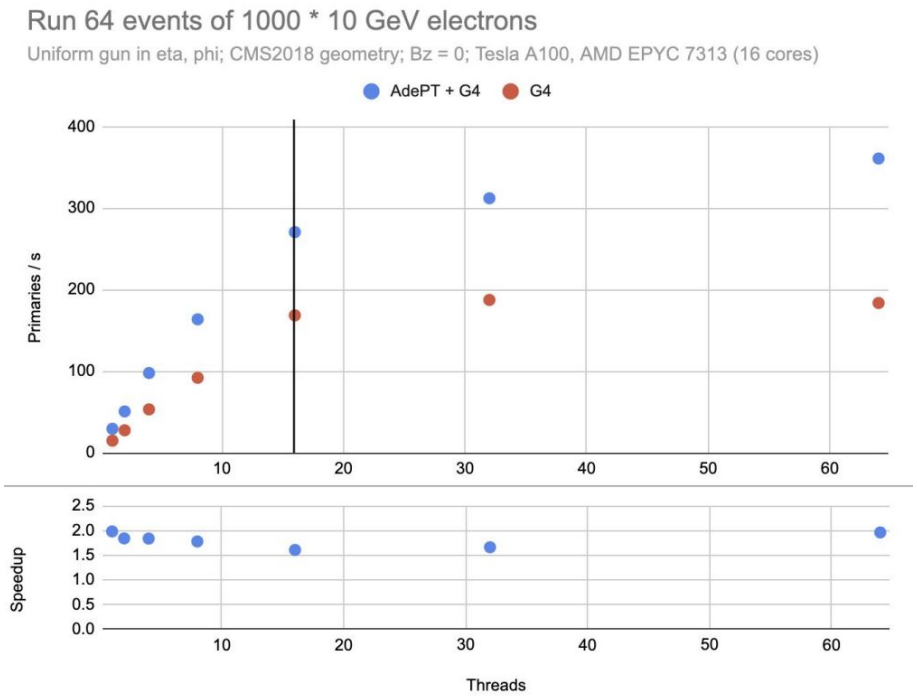
- Sorting tracks on type/energy/etc at specific points
  - *Reduce kernel grid sizes, maybe divergence*
- Use of single/mixed precision
  - *Mostly of benefit to consumer grade GPUs*
- Shared offload “service” to improve CPU/GPU concurrency
  - *Less blocking of CPU when GPU processes (e.g. Hadrons on CPU whilst GPU processes EM)*
  - *=> Event mixing on device*



Credit: [Juan Gonzalez (CERN)]

Potential Optimization Strategies for Workflow

- Sorting tracks on type/energy/etc at specific points
  - *Reduce kernel grid sizes, maybe divergence*
- Use of single/mixed precision
  - *Mostly of benefit to consumer grade GPUs*
- Shared offload “service” to improve CPU/GPU concurrency
  - *Less blocking of CPU when GPU processes (e.g. Hadrons on CPU whilst GPU processes EM)*
  - *=> Event mixing on device*



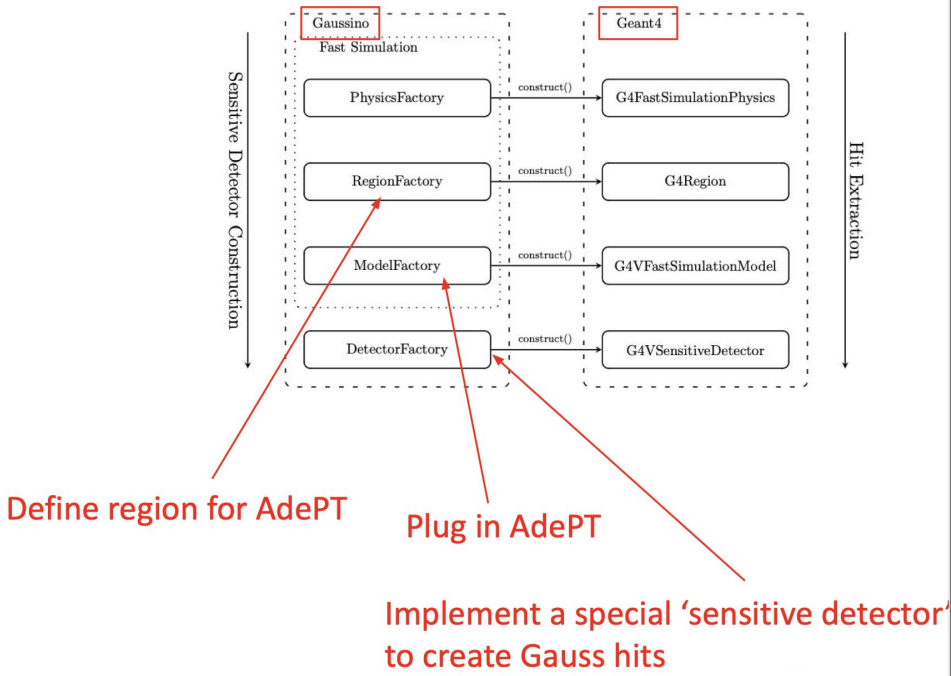
Credit: [Juan Gonzalez (CERN)]

Integration/Testing in Experiments: ATLAS

- TileCal test beam standalone Geant4 application as testbed
  - *Code: https://github.com/lopezzot/ATLTileCalTB (see presentation)*
- **AdePT**: Initial integration by Davide Costanzo:
  - *See https://indico.cern.ch/event/1215829/contributions/5306569/*
- **Celeritas**: integration both standalone and as FullSimLight plugin by Warwick
- Combined offload interface being developed by both teams
  - *Choose AdePT/Celeritas at runtime to compare, help consolidate/test*
  - *Enables easier integration, testing in Athena*
- **Topics being worked on as part of Warwick's contribution to ATLAS Full Simulation WG and both R&D teams in SWIFT-HEP**

Integration/Testing in Experiments: LHCb

- **AdePT**: Combine standalone application example with Gauss-on-Gaussino machinery
  - *Fill AdePT pipeline with particles entering LHCb Calo region*
  - *Generate Gauss hits from AdePT (to give equivalence with plain Geant4)*
  - *Working with Juan Bernardo Benavides (LHCb Doctoral Student)*
- **Celeritas**: Warwick will assist in trialling integration of last week's new combined interface this year.

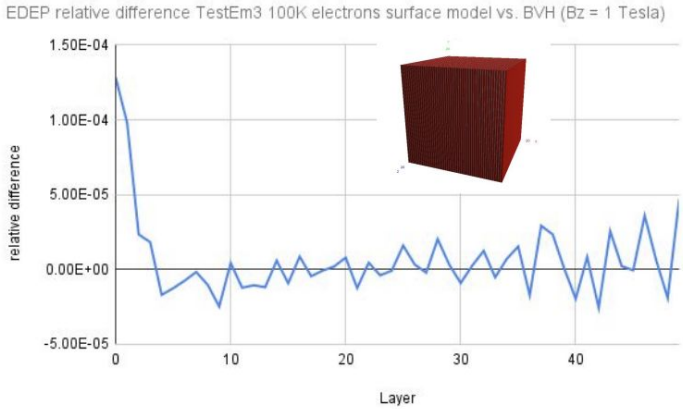


Surface-based Geometry: VecGeom, ORANGE

- Current CSG model of VecGeom known GPU bottleneck in AdePT and Celeritas
  - *Divergence from different algorithmic complexity in different solids, etc*
- Effort to develop and use surface- based geometry models, navigation
  - *Reduce divergence from smaller number of surfaces, simpler algorithms*
  - ***VecGeom**: bounded surfaces (explore potential for work reduction in LHC-complexity geometries by reducing checks on "virtual" crossings)*
  - ***ORANGE**: unbounded surfaces (approach from nuclear engineering codes for reactor geometries)*
- Defer to the following presentations at the Geant4 Collaboration meeting for details of developments and results:
  - *Surface-based GPU model in VecGeom, Andrei Gheata et al*
  - *ORANGE surface geometry progress, Seth Johnson et al*
- **RSEs at Warwick and Sheffield to profile and optimize under the ExaTEPP grant, also expect to develop common Geometry interface for VecGeom/ORANGE through SWIFT-HEP co-working.**
  - ***VecGeom/ORANGE surface models now in more stable, developed state than previously.***
  - ***Still some CSG->Surface conversion implementations to complete***

VecGeom Surface Model Status

- Optional usage of the surface model in AdePT example
  - *No relevant changes needed other than triggering the model conversion and the navigator type*
- "TestEm3" Sampling calorimeter simulation
  - *Pb + LAr box layers (w/ constant Bz field)*
  - *10 GeV electrons shot through layers*
- Numerical divergence small and understood
  - *Boundary crossing relocation*

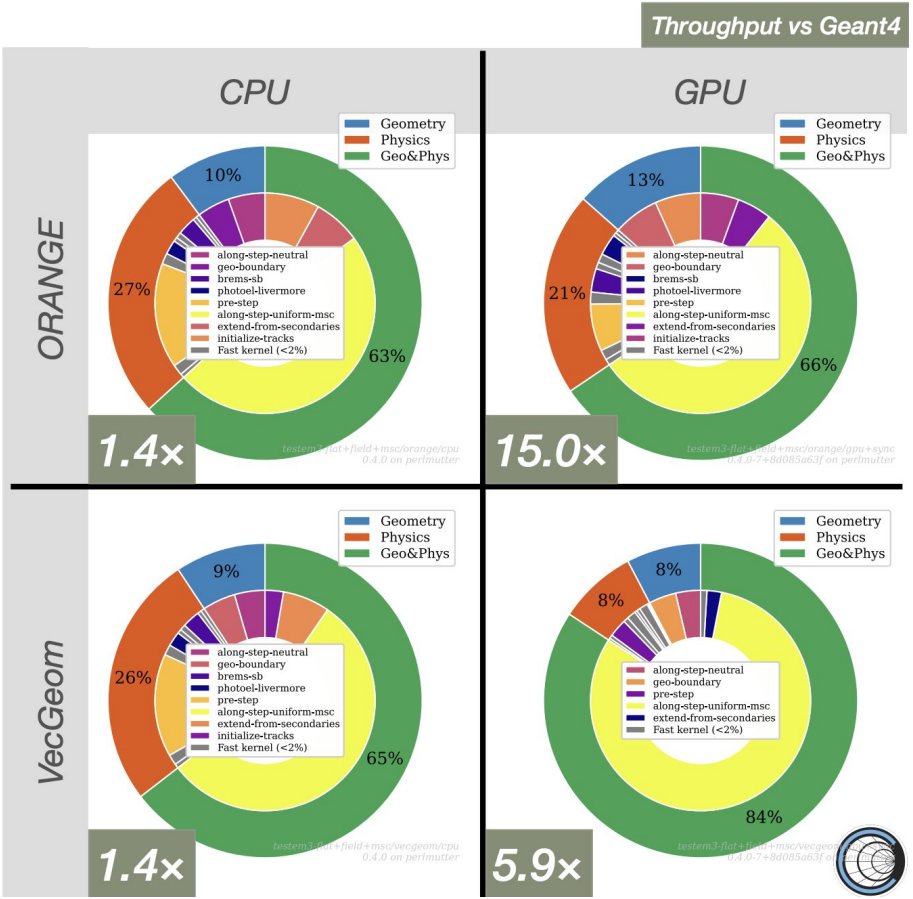


	BVH	no BVH surf
no field	152s	156s
Bz=1T	194s	184s

Credit: [Andrei Gheata (CERN)]

ORANGE Status

- Performance parity with VecGeom solids on CPU
- Physics time parity on GPU
- Step counts equivalent
- ORANGE speed up on GPU:
  - *1.4x: Neutral propagation*
  - *3.6x: Field propagation*
  - *1.5x: Boundary crossing*
- But… A very simple problem with “no” divergence as all boxes



Credit: [Seth Johnson (ORNL)]

Surface-based Geometry TODOs

- Development of common Geometry API to allow runtime choice between VecGeom/ORANGE backends
- Completion of solid-surface converters/surface builders to allow LHC-scale detectors to be modelled/profiled
  - *CMS Run 2 is immediate target*
  - *Full Geant4 conversion later*
- **Profiling, optimization of models and navigation**
  - ***Both codes, and AdePT/Celeritas, should now be in a more stable situation and with clearer targets for ExaTEPP to contribute to here***
  - ***Workplans in CERN SFT still being finalized, but expect to set up chats/meetings in next week or two.***

Other News/Topics of Interest

- Celeritas starting to develop optical photon physics/transport
  - *Interesting comparison with raytracing/rendering (NVidia OptiX, Mitsuba)*
  - *Broader scope of applications than just LHC (e.g. Dark Matter, Neutrinos, Detector Development)*
- Also some funding through SciDAC for neutron physics
- Potential impact of US P5 recommendations on funding for compute R&D/porting?: https://science.osti.gov/hep/hepap/Meetings/202312
- Power efficiency in simulation https://indico.cern.ch/event/1307331/contributions/5593673/
- Differentiable programming for simulation: https://indico.cern.ch/event/1307331/contributions/5594500/
  - *Related to broader topic of general use of AI in simulation*

Summary

- AdePT/Celeritas have developed significantly over the last six months
  - *Review meeting in December 2023*
  - *Integration testing in LHC frameworks*
- Efforts now underway on common APIs/codes
  - *Currently on Geant4<->GPU offload*
  - *Next on geometry calls*
- **VecGeom/ORANGE surface models reaching core functionality threshold**
  - ***Better opportunities for ExaTEPP contributions to profiling/optimization***