Celeritas: Bringing GPU capabilities to HEP detector simulation

Oak Ridge National Laboratory
Argonne National Laboratory

Stefano Tognini*
Amanda Lund

*togninis@ornl.gov

ORNL	Elliott Biondo, Marcel Demarteau, Thomas Evans, Rashika Gupta, Seth Johnson, Stefano Tognini
ANL	Amanda Lund, Paul Romano
FNAL	Philippe Canal, Soon Yung Jun, Guilherme Lima
LBNL	Julien Esseiva
BNL	Sakib Rahman
U. of Virginia	Hayden Hollenbeck
U. of Warwick	Benjamin Morgan

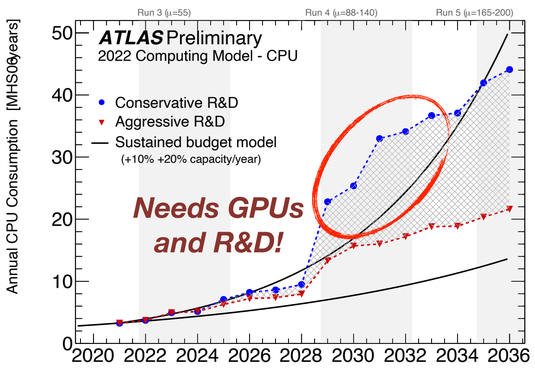
MISSION & STRATEGY

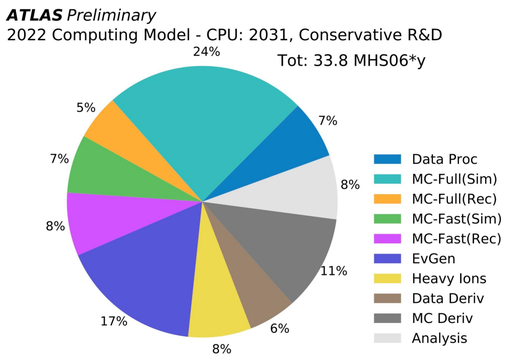
Enable scientific discovery in HEP by leveraging GPUs

- Detector simulation is a fundamental component of HEP frameworks
- HPC/GPU hardware has become an essential tool in the field

MOTIVATION

Computational demand



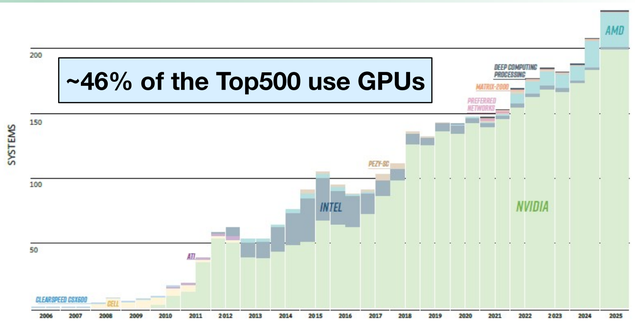


ATLAS Collaboration: CERN-LHCC-2022-005; LHCC-G-182

- Unprecedented computing demand by next-gen HEP experiments

Computational supply

- Heterogeneous architectures are increasingly common
  - AI/ML will accelerate the trend
  - HEP is also expanding AI/ML usage from event generation to data analysis

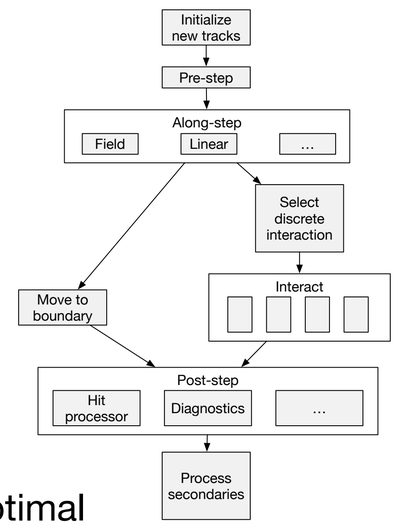


https://www.nextplatform.com/2025/06/10/top500-supers-even-accelerators-cant-bend-performance-up-to-the-moores-law-line/

CODE DESIGN

- **Overview**
  - Composition-based, data-oriented design
  - Platform-portable (CUDA, HIP, OpenMP)
  - Cross-section data imported from Geant4
  - **Fully reproducible on device**
- **Extensibility**
  - Physics models are written in pure C++
  - Launch wrappers for kernel executions
  - Small kernels
- **Geometry navigation**
  - **VecGeom**: HEP community, CUDA only
  - **ORANGE**: ORNL, platform portable, GPU-optimal

Production-ready software/library



Interface with existing Geant4 applications

- Specialized tasks are offloaded to GPU
- Return sensitive hit objects for digitization and I/O

CURRENT COLLABORATIONS

Logos are ™ and © their respective owners and are not formally affiliated with Celeritas

PHYSICS PROCESSES

Electromagnetic		
Particle	**Process**	**Model(s)**
γ	Photon conversion	Bethe-Heitler
	Compton scattering	Klein-Nishina
	Photoelectric effect	Livermore
	Rayleigh scattering	Livermore
e±	Ionization	Moller-Bhabha
	Bremsstrahlung	Seltzer-Berger, relativistic
	Multiple scattering	Urban (extended to high E)
	Coulomb scattering	Wentzel
e+	Pair annihilation	Positron to two gammas
μ±	Bremsstrahlung	Bremsstrahlung
	Pair production	Muon pair production
	Multiple scattering	Urban / WentzelVI
	Ionization	Bethe-Bloch
μ+	Ionization	Bragg
μ−	Ionization	ICRU73QO

Magnetic field	
Integrators	**Supported inputs**
Analytical helix	uniform
Runge-Kutta	mapped
Dormand-Prince	parameterized

Optical physics
Cherenkov
Scintillation
WLS
Rayleigh scattering
Absorption
Mie scattering
Surface physics

- **Ongoing**
  - Surface Physics
- **Planned**
  - Decay
  - Hadronic

Implemented | Not implemented

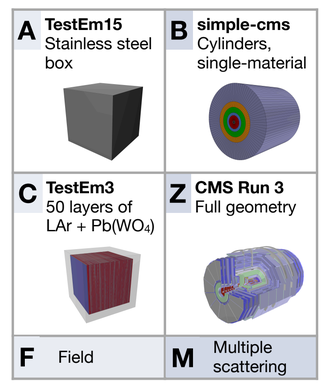
PERFORMANCE

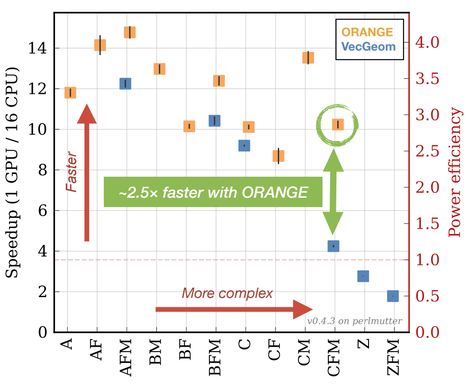
NERSC

Results use ¼ of a Perlmutter node:
1 Nvidia A100 vs 16 cores of AMD EPYC 7763 (32 hardware threads)

Standalone Celeritas EM-only app

- Efficiency: **4×**
  Performance per watt
- Capacity: **~2–15×**
  For 1 GPU vs. 16 CPU cores
- **~95%** of CMS Run 3 runtime spent in geometry kernels
  Using VecGeom

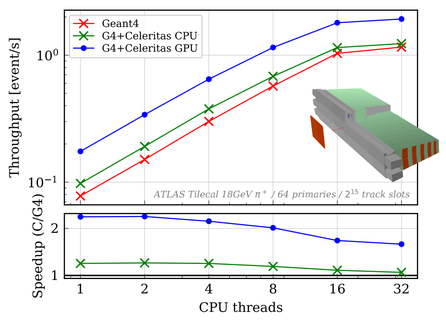




Geant4 full physics + Celeritas EM: ATLAS subdetector

- Speedup with Celeritas + GPU: **~1.8–2.2×**
  With up to 32 threads (16 cores x 2 hyperthreads/core)
- Celeritas CPU-only speedup: **~1.1–1.3×**
  Despite integration overhead

Theoretical maximum speedup **~2.2–2.5×**
Based on Amdahl's Law due to Geant4 full physics time



IMCC CHALLENGES & OPPORTUNITIES

Beam-induced background simulation challenges

- BIB is an unprecedented challenge on both simulation and hardware
- Relies on CPU-only codes such as FLUKA for R&D
- Multi-TeV EM showers require weeks on full-time multithreaded CPU executions to achieve a fraction of a bunch crossing
  - Leads to compromises (e.g. high threshold energy for e± and γ)

Opportunities

- Heterogeneous architectures: faster + higher-fidelity BIB simulations
- Integrate modern computing paradigms into the framework
- **Celeritas** is addressing one of the P5 recommendations
  Support emerging technologies and computing hardware to enable scientific discovery
  - Extensibility allows for the addition of tailored physics
  - Expand R&D avenues for detector and reconstruction R&D
  - Path for long-term integration between IMCC and DOE LCFs

ACKNOWLEDGEMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of High Energy Physics, **Scientific Discovery through Advanced Computing (SciDAC)** program. This research used resources of the **Oak Ridge Leadership Computing Facility** at the **Oak Ridge National Laboratory**, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. This research used resources of the **National Energy Research Scientific Computing Center (NERSC)**, a U.S. Department of Energy Office of Science User Facility located at **Lawrence Berkeley National Laboratory**, operated under Contract No. DE-AC02-05CH11231 using NERSC award HEP-ERCAP-0023868. This research was supported by the **Fermi National Accelerator Laboratory**, managed and operated by Fermi Forward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy.