ORNL/SPR-2024/3719

Celeritas midterm SciDAC report

Seth R. Johnson[1]
Amanda Lund[3]
Philippe Canal[2]
Stefano C. Tognini[1]
Julien Esseiva[4]
Elliott Biondo[1]
Soon Yung Jun[2]
Guilherme Lima[2]
Ben Morgan[5]
Marcel Demarteau[1]
Thomas Evans[1]
Paul Romano[3]

2 December 2024

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]Fermi National Accelerator Laboratory, Batavia, IL, USA
[3]Argonne National Laboratory, Lemont, IL, USA
[4]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[5]University of Warwick, Coventry, United Kingdom

OAK RIDGE National Laboratory

ORNL IS MANAGED BY UT–BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/SPR-2024/3719

Computational Sciences & Engineering Division

CELERITAS MIDTERM SCIDAC REPORT

Seth R. Johnson[1]
Amanda Lund[3]
Philippe Canal[2]
Stefano C. Tognini[1]
Julien Esseiva[4]
Elliott Biondo[1]
Soon Yung Jun[2]
Guilherme Lima[2]
Ben Morgan[5]
Marcel Demarteau[1]
Thomas Evans[1]
Paul Romano[3]

2 December 2024

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

ABSTRACT

Celeritas is a new Monte Carlo (MC) code that helps satisfy the increasing demand for high energy physics (HEP) detector simulation, using Graphics Processing Unit (GPU) hardware on high performance computing (HPC) systems to model Large Hadron Collider (LHC) experiments and beyond. This report details the project's progress midway through its SciDAC funding period, highlighting the first complete implementation of standard electromagnetic (EM) physics on GPUs, initial results for performance and scalability on Leadership Computing Facilities (LCFs), and preliminary integration into the CMS and ATLAS experiments. By integrating HEP domain knowledge with expertise in MC transport, Celeritas has catalyzed a shift in the HEP community's perception of GPU platforms as the future for HPC simulations.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADIOS	Adaptable Input Output System
ALICE	A Large Ion Collider Experiment
API	Application Programming Interface
ASCR	Advanced Scientific Computing Research
ATLAS	A Toroidal LHC Apparatus
BNL	Brookhaven National Laboratory
CERN	Conseil Européen pour la Recherche Nucléaire
CI	continuous integration
CMS	Compact Muon Solenoid
CMSSW	CMS software
CPU	Central Processing Unit
DOE	U.S. Department of Energy
DUNE	Deep Underground Neutrino Experiment
ECP	Exascale Computing Project
EIC	Electron-Ion Collider
EM	electromagnetic
ePIC	Electron-Proton/Ion Collider
ESPPU	European Strategy for Particle Physics Update
GPU	Graphics Processing Unit
HEP	high energy physics
HL-LHC	High Luminosity Large Hadron Collider
HPC	high performance computing
JSON	JavaScript Object Notation
LBNL	Lawrence Berkeley National Laboratory
LCF	Leadership Computing Facility
LEGEND	Large Enriched Germanium Experiment
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LZ	LUX–ZEPLIN
MC	Monte Carlo
MPI	Message Passing Interface
MSC	multiple scattering
ORANGE	Oak Ridge Adaptable Nested Geometry Engine
PI	Principal Investigator
SciDAC	Scientific Discovery through Advanced Computing
SD	sensitive detector
SIMD	single instruction, multiple data

1. INTRODUCTION

In early 2019 a pilot effort, cosponsored by the U.S. Department of Energy (DOE)'s HEP program and the Exascale Computing Project (ECP), explored GPU-accelerated particle transport for HEP detector simulation. Building on early successes with GPU prototype simulations, this nascent "Celeritas" project secured a SciDAC grant with a reduced scope and timeline, later extended from two years to five.

At the beginning of the project, the use of GPUs for detector simulation faced widespread skepticism, in part due to unsuccessful attempts to adapt Geant4 (Agostinelli et al. 2003; John Allison et al. 2006; J. Allison et al. 2016) to architectures such as single instruction, multiple data (SIMD). Thanks to the investment of SciDAC to date, Celeritas has driven a profound shift in this perception by demonstrating the benefits of a multidisciplinary team collaborating closely with the broader HEP community.

This report summarizes the progress of the Celeritas project midway through the SciDAC funding period. It highlights advancements in physics modeling, advanced computing, and the community engagement essential for enabling new scientific discoveries.

2. PHYSICS

Celeritas has achieved a major milestone by completing the first implementation of LHC EM physics on GPUs, enabling efficient simulation of EM showers, the most computationally intensive aspect of LHC detector simulations. Although some software components in Celeritas are directly analogous to parts of Geant4—processes define physical phenomena and models describe their theoretical underpinnings—the code is a fully independent implementation of these models, rewritten for platform portability and optimized for GPUs. Table 1 enumerates the standard EM physics processes and models used by LHC. The

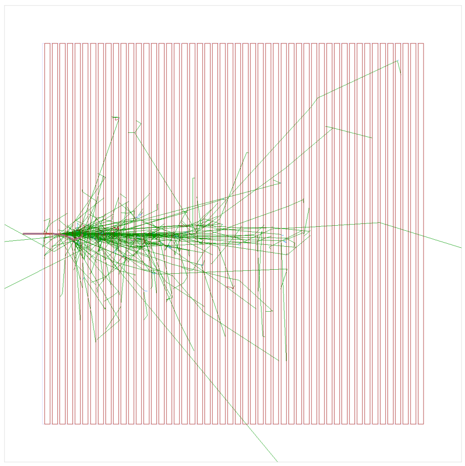
Table 1. Current status of Celeritas EM physics.

Particle	Process	Model(s)
γ	photon conversion	Bethe–Heitler
	Compton scattering	Klein–Nishina
	photoelectric effect	Livermore
	Rayleigh scattering	Livermore
$e^{\pm}$	ionization	Møller–Bhabha
	bremsstrahlung	Seltzer–Berger, relativistic
	pair annihilation	EPlusGG
	Coulomb scattering	eCoulombScattering
	multiple scattering	Urban, WentzelVI

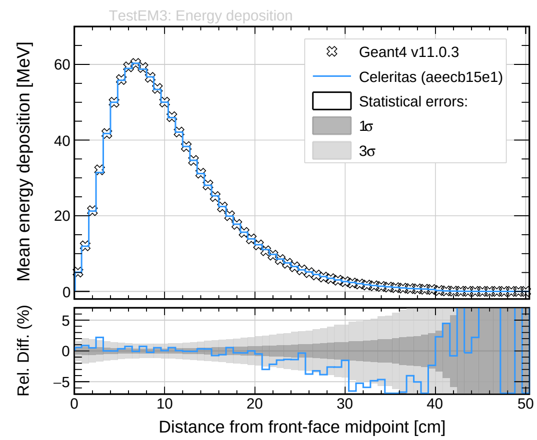
extensible physics design central to Celeritas allows additional processes and models to be straightforwardly added, allowing additional projects to develop and integrate muon physics processes (fusion, decay, and EM) and optical physics.

Aiming to improve standards for testing and verification within the HEP community, Celeritas applies best practices from scientific computing to the particle physics domain. Physics models are improved by component-based programming, rigorous unit testing, detailed model verification (Tognini et al. 2022), experimental validation, and high-level test problems. The idealized calorimeter problem "TestEM3" represents the simplest and most important integral test case. As shown in Figure 1, this problem is a 40 cm cube with 50 interleaved layers of liquid argon and lead, with an incident pencil beam of 1 GeV electrons. The simulated energy deposition, both radially and axially, verifies the equivalence of the Celeritas and Geant4 physics implementations. Additional code-to-code comparisons of internal diagnostics, including step distributions and process interaction frequencies, show further agreement.

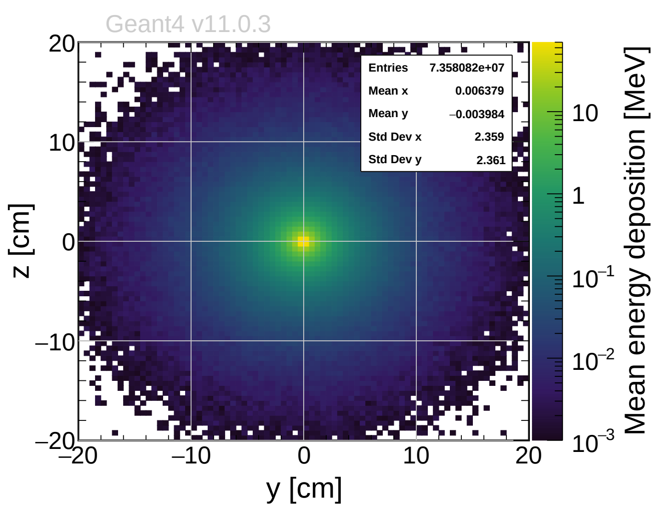
Figure 2 shows two EM validation problems comparing Geant4 and Celeritas results to experimental measurements. The first tests multiple scattering (MSC) phenomenology, which determines the exiting angular distribution of 15.7 MeV electrons scattered off of a 19 µm thin gold foil (Hanson et al. 1951). The second measures energy loss fluctuations during the slowing down of 100 MeV electrons in silicon (Meroli, Passeri, and Servoli 2011).



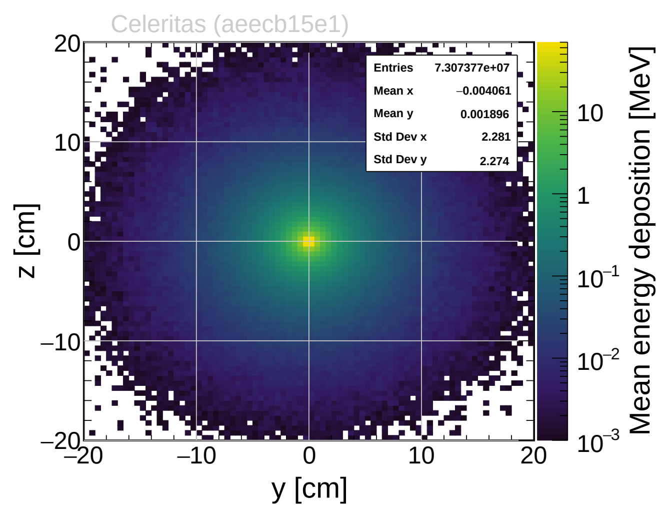
(a) Visualization of an event



(b) Axial energy deposition

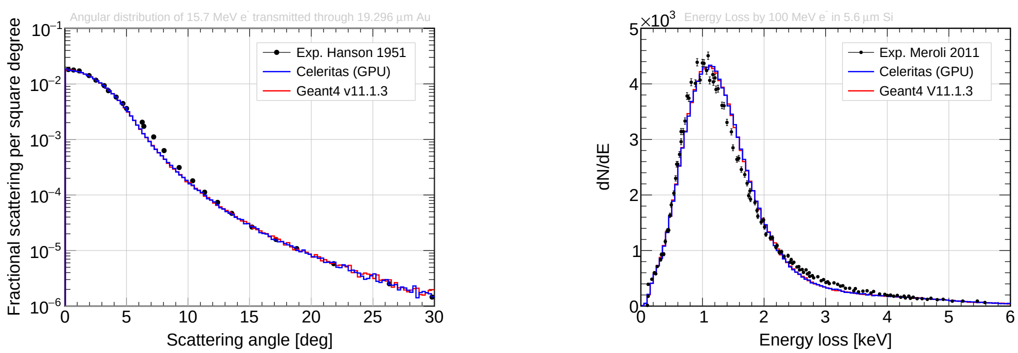


(c) Radial energy deposition (Geant4)



(d) Radial energy deposition (Celeritas)

Figure 1. Validation overview of the TestEM3 problem. Figure shows (a) a Celeritas simulated event, highlighting the path of the electron primary; (b) the energy deposition as a function of the distance from the midpoint of the front-face of the geometry, for both Celeritas and Geant4; and the perpendicular energy distribution for (c) Geant4 and (d) Celeritas.



(a) Thin foil scattering

(b) Silicon energy loss

Figure 2. Experimental results for two thin-target validation problems compared against simulated Geant4 and Celeritas results.

3. ADVANCED COMPUTING

Celeritas has made novel advancements in scientific computing by addressing the algorithmic challenges of particle transport simulation in HEP (Johnson et al. 2021) and by demonstrating performance on multiple HPC systems (Johnson, Esseiva, et al. 2024).

3.1 NOVEL ALGORITHMS FOR HEP

The core algorithm in Celeritas is to perform a *loop interchange* (Allen and Kennedy 1984) between particle tracks and steps. The classical, serial way of simulating an event (Fig. 3a) is to have an outer loop over tracks and an inner loop over steps, and inside each step are the various actions applied to a track such as evaluating cross sections, calculating the distance to the nearest geometry boundary, and undergoing an interaction to produce secondaries. There is effectively a data dependency between the track at step i and step $i + 1$ that prevents vectorization. The approach Celeritas takes to "vectorize" the stepping loop on GPU (Fig. 3b) is to have an outer loop over "step iterations" and an inner loop over "track slots", which are elements in a fixed-size vector of tracks that may be in flight.

3.2 SCALABLE SOFTWARE

The original proposal for Celeritas development emphasized a standalone simulation application to maximize performance gains on HPC systems. However, as the project matured, it became clear that integration into experiment frameworks was essential for widespread acceptance and eventual uptake for production use. To make this a priority, the initial development of Celeritas has focused on integration in HEP workflows through an external Geant4 framework, with a standalone app for performance testing and scaling in HPC environments.

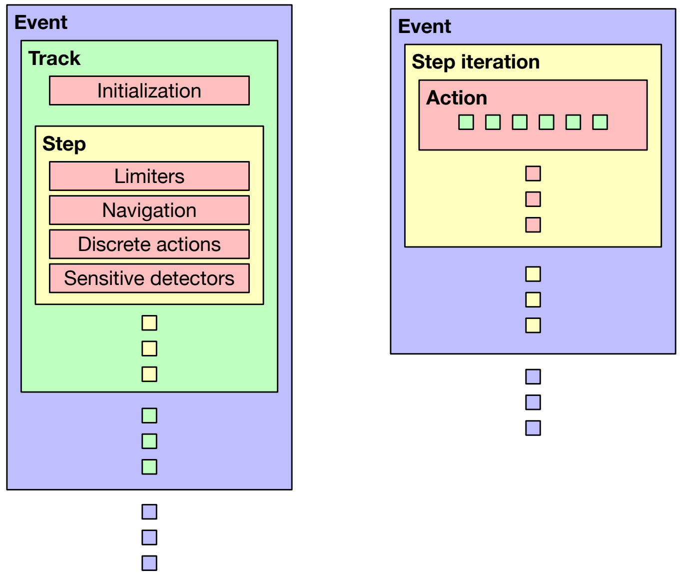
At the same time, in preparation for future scalability work, the RAPIDS2 team completed a preliminary integration of Adaptable Input Output System (ADIOS) for high-performance output of detailed simulation diagnostics, colloquially referred to as "MC truth." The first implementation revealed challenges specific to HEP workflows: rather than being large contiguous arrays of distributed-memory data written sequentially, the output from simulations are large irregular blobs of small record-like data. Although ADIOS excels in handling large data chunks, the high metadata overhead for small records resulted in exceedingly large output files. For now, the Celeritas team uses ROOT (Brun et al. 2020), which was designed specifically for such output formats, to write this data. Future work is to incorporate GPU ROOT output for improved bandwidth.

Although basic Message Passing Interface (MPI) support has been incorporated into Celeritas, the lack of support from experiment workflows means it has not been a priority. Instead, GNU parallel has been used experimentally to launch 800 independent multicore jobs on the Frontier supercomputer, demonstrating efficient scaling across nodes (Maheshwari et al. 2024).

3.3 STANDALONE PERFORMANCE RESULTS

The performance of Celeritas is regularly measured using a set of benchmark problems (Tognini et al. 2022) of increasing complexity that evaluate the cost of incremental features such as multiple materials, multiple geometric regions, magnetic fields, and additional physics. Most of these problems use 1300 10 GeV electron primaries per event, representing the amount of energy per LHC collision deposited into the material.

The gold standard for Celeritas benchmarks is the DOE capacity machine Perlmutter (Center 2022), which is of particular importance in HEP as an analysis tool for ATLAS. Of the available DOE machines, it



(a) CPU loop **(b) GPU loop**

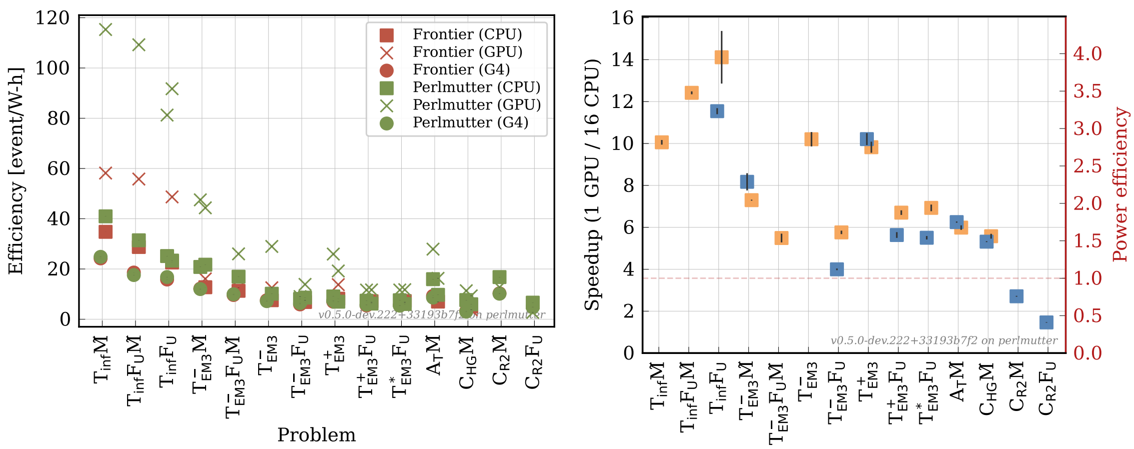
Figure 3. Monte Carlo loop structure (a) on traditional CPU architectures and (b) with a loop interchange for performance on GPU. Vertical dots imply sequential replication of the same-colored loop, and the horizontal green dots are tracks operating in parallel.

is the only one to use Nvidia GPUs, necessary for comparing full detector problems due to the CUDA-specific implementation of VecGeom (Apostolakis et al. 2019). Some runs using the Oak Ridge Adaptable Nested Geometry Engine (ORANGE) geometry implementation include results from Frontier (Atchley et al. 2023), which uses AMD GPUs that VecGeom does not support.

On each supercomputer, each benchmark problem is run on a single compute node with one process per discrete GPU. These independent processes run simultaneously with different starting seeds for the pseudo-random number generator in order to give a better estimate of the variance over a range of potential events. One set of benchmarks is run with each instance using one CPU and one discrete GPU, and a second set executes an OpenMP-multithreaded run of C/G CPUs per instance, where C is the number of CPUs per node and G is the number of discrete GPUs. Comparing the two gives an effective measure of the GPU-to-CPU core equivalence for each benchmark.

Figure 4 shows (a) power efficiency estimated by theoretical chipset power draws and (b) the relative speedup by using GPUs available in the machine versus neglecting them and using only CPUs. The GPU-to-CPU equivalence is the speedup multiplied by the number of CPUs used in each run: on Perlmutter, an A100 provides the same throughput as 10–100 cores of an AMD EPYC. Equivalently, given that hardware configuration, ignoring the GPUs leads to a 32–93% performance loss: for some problems with Celeritas, the CPU powers only 7% of the total throughput. These performance numbers, which highlight the importance of utilizing the GPUs on DOE systems, have fluctuated over the course of the project as new capabilities are added and performance optimizations implemented.

Further results, details, and discussions are presented in (Johnson, Castro Tognini, et al. 2024) and (Johnson, Esseiva, et al. 2024).



(a) Estimated power efficiency

(b) GPU performance on Perlmutter

Figure 4. Efficiency and capacity results for the benchmark problems. Blue markers (offset left of the grid line) are for VecGeom and orange (offset right of the grid line) are for ORANGE. Problem T_{inf} is an infinite medium, T_{EM3} is the idealized calorimeter with modifiers $-/+/*$ for increasingly complex representation and materials, A_T is the ATLAS tile calorimeter, C_{HG} is the CMS high granularity calorimeter, and C_{R2} is the full CMS Run 2 (2018) geometry. A $\tilde{M}$ suffix means MSC was disabled, and F_U indicates the presence of a 1T uniform axial magnetic field.

4. IMPACT

The advancements in computing developed by the Celeritas team have opened the door to using DOE computing facilities for next-generation detector simulation. Yet the *physics discovery* impact of Celeritas requires both scientific advancements and community building. Accelerating production-level physics simulations in LHC experiment frameworks is the ultimate goal of this SciDAC and requires careful collaboration with and integration into the HEP community. Celeritas has been a leader in GPU simulation: driving innovation in the GPU detector simulation space, working constantly toward production-quality interfaces to Geant4, and motivating further investment into GPU integration.

4.1 GEANT4 INTEGRATION

Because effectively all HEP applications use Geant4 for EM detector simulations, it is essential for Celeritas to integrate as a plug-and-play replacement to minimize the work required by experiments. With this interface, user applications set up the Geant4 toolkit, Geant4 sends EM tracks to the main Celeritas stepping loop, and Celeritas reconstructs Geant4 "step" objects that are sent to user-implemented sensitive detectors (SDs) (Fig. 5).

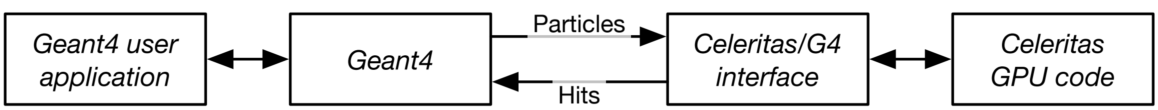


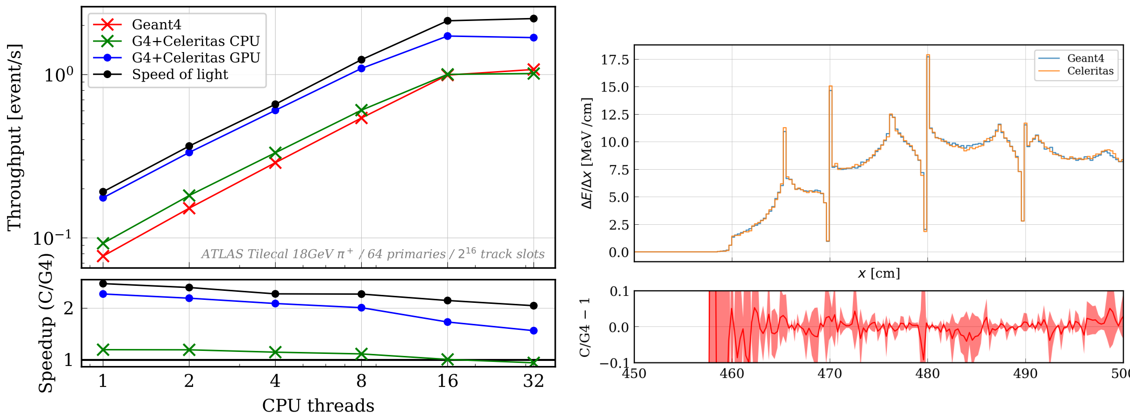
Figure 5. Geant4/Celeritas interface.

An early integration success with Celeritas was a standalone test beam of 18 GeV π^+ incident on an ATLAS tile calorimeter (TileCal) module (Lachnit, Pezzotti, and Konstantinov 2022). The team adapted the test problem to be part of the ATLAS FullSimLight (Bandieramonte, Bianchi, and Boudreau 2020) test application and ran independent test problems to validate computing and physics results (Fig. 6). With approximately 100 lines of mostly boilerplate interface code, Celeritas increases simulation throughput by about a factor of two, close to the maximum theoretical speedup (determined by killing all EM tracks at birth) with the same physics result. This positive result increased interest from the ATLAS team and established the feasibility of reconstructing hits on GPU and sharing a single GPU among multiple CPUs.

A standalone application, `celer-g4`, was developed to provide a programmatic interface to Geant4 with and without Celeritas offloading. It integrates full hadronic physics with EM offloading and serves as a testbed for performance comparisons for realistic detector simulations. The relative speedup of offloading EM tracks to Celeritas on an Nvidia A100 GPU with the CMS High Luminosity Large Hadron Collider (HL-LHC) detector configuration and the three different field setups is shown in Figure 7. The theoretical maximum speedup varies between 3–6×, depending on the field configuration. The overhead of reconstructing the SD hits is relatively small, accounting for about 15% of the runtime in the zero field case and 5–10% with a magnetic field. These test problems proved the potential for using Celeritas to accelerate full detector simulation problems rather than single subdetector components as originally targeted by other GPU detector simulation projects.

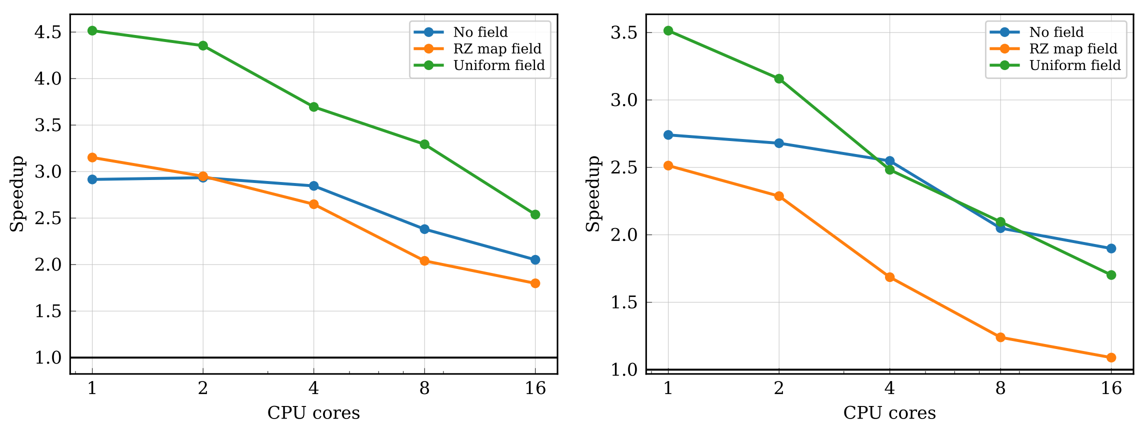
4.2 EXPERIMENT FRAMEWORK INTEGRATION

The Celeritas project aims first and foremost to increase the physics output of the DOE-sponsored experiments. To that end, it is essential to involve those experiments, and in particular their software framework



(a) Single-GPU strong scaling **(b) Energy deposition comparison**

Figure 6. Results for the ATLAS Tilecal. Performance results are on $\frac{1}{4}$ of a Perlmutter node.



(a) CMS Run 3 geometry **(b) CMS Run 4 geometry**

Figure 7. Relative speedup of offloading EM tracks to the GPU through Celeritas as a function of CPU threads. 32 $t\bar{t}$ events were simulated using the CMS Run 3 and Run 4 geometries with different magnetic field configurations. Performance was measured on an AMD EPYC 7532 32-core processor @ 2.4 GHz and a 40 GB Nvidia A100 GPU card.

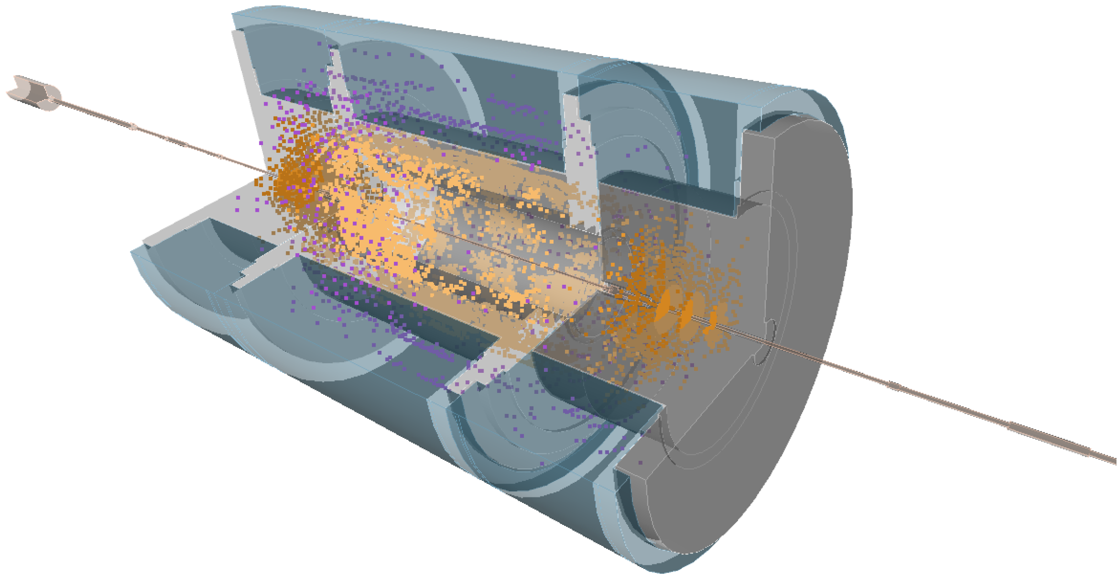


Figure 8. Event display of the ATLAS calorimeters. Celeritas-produced hits within the Athena framework.

developers as early as possible to ensure a smooth adoption and to align the development with their requirements.

Integration with the CMS experiment framework, led largely by Fermilab and thus closely connected to this SciDAC project, began as soon as a minimally functional EM physics simulation was implemented. Adding Celeritas as an "external" package to CMS software (CMSSW) (CMS Offline Software and Computing 2021) and then integrating into the Geant4 simulation required extensive collaboration with the CMS software team, meeting over Zoom, conversing over Slack, and working at two in-person hackathons. The resulting prototype, capable of offloading EM tracks to Celeritas and returning hits to CMSSW, uses full EM physics and a cylindrical magnetic field geometry.

The prototype identified key requirements, such as region-dependent physics options, for exact comparisons. It also underscored the importance of active involvement from framework maintainers: frequent updates to the CPU-based code and limited availability of CMS personnel to review and test the prototype caused the implementation to fall out of sync with mainline development. The Celeritas code library remains an optional CMS external, and efforts to resume work on CMS integration are planned for early 2025, supported by a new group of developers dedicated to enhancing CMS simulation capabilities.

Building on maturing engagement with the ATLAS experiment through Lawrence Berkeley National Laboratory (LBNL) and the SWIFT-HEP collaboration, ATLAS became the next target for Celeritas integration. This year in Geneva, a hands-on hackathon between the Celeritas team and the ATLAS Simulation group worked to integrate Athena framework (ATLAS Collaboration 2021) and Celeritas. As with CMS integration, several challenges were identified and mostly overcome; these are tracked on a Celeritas GitHub issue to maintain communication and status updates between the teams. Two outcomes of the hackathon were the first visualization of ATLAS hits performed on GPU (Fig. 8) and a successful preliminary validation of hit distributions. While the current integration remains at the prototype stage, it lays the groundwork for future refinement and eventual adoption within the Athena framework.

Throughout the SciDAC, the Celeritas project has been in contact with the simulation teams in ALICE, LHCb and LUX–ZEPLIN (LZ) in order to understand requirements and plan work to support those experiments. More recently the Celeritas team has engaged with the Electron-Proton/Ion Collider (ePIC) (Electron-Ion Collider (EIC)), LEGEND, and Deep Underground Neutrino Experiment (DUNE) experiments on integration specifically targeted at optical photons.

4.3 OPEN SOURCE SOFTWARE

One of the Celeritas project's tenets is to not only leverage existing open source software but to contribute advancements in such software back to the community.

A particularly difficult technical challenge faced by the project is the CUDA feature "Relocatable Device Code," which is required by the VecGeom library but was apparently not designed for external library compatibility. Proper support of this type of configuration required the development of custom CMake code to automate the arcane rules necessary for the system to work properly. This CMake code is now being shared by multiple projects, and as it matures it will be upstreamed to the official Kitware repository.

The HPC-oriented package manager Spack (Gamblin et al. 2015) has been extensively used by the Celeritas project in developer environments and the continuous integration (CI) system. The Celeritas team has actively engaged with and contributed to multiple Spack recipes and core capabilities. The close collaboration between the Celeritas and ROOT projects has also strengthened the stability and usability of the ROOT build and execution environment, particularly in Spack-based setups and on LCF hardware.

Members of the Celeritas team, as primary developers of ORANGE, have also been working in close collaboration with the CERN-led VecGeom project. One key outcome of this collaboration has been to inspire VecGeom to reimplement the core tracking methods using elements of ORANGE novel surface-based design, which has been shown to work more effectively on GPU.

Finally, as part of the Geant4 collaboration, five Celeritas core team members have worked with the Geant4 and VecGeom projects to improve their functionality. Numerous bugs and compatibility issues in Geant4 and VecGeom have been reported and fixed, benefiting the entire HEP software ecosystem.

4.4 BROADENING THE COMMUNITY

The Celeritas project is actively seeking community engagement for development. To this end, the project repository contains defined, documented, and maintained a set of policies for community standards, code contributions, code reviews, and code releases. The popularity of Celeritas within the scientific software community has been increasing steadily as measured by the number of GitHub "stars" (Fig. 9).

In the three years since the project's inception, four students of diverse backgrounds have worked with our team, all leveraging external funding sources. The first implemented a muon Bremsstrahlung model, another explored possible performance optimizations for the magnetic field driver, a third wrote the Coulomb single scattering kernel and is working on optical photon physics, and the most recent implemented a new geometry model for involute surfaces. The core Celeritas team works closely with the students to ensure code quality as measured by unit tests, documentation, and validation problems.

The Celeritas team is also leading by example, by demonstrating the cost effectiveness and increase in code quality that result from modern software management techniques and coding patterns. Frequent discussions with and presentations to members of the HEP community have emphasized the importance and utility of such practices.

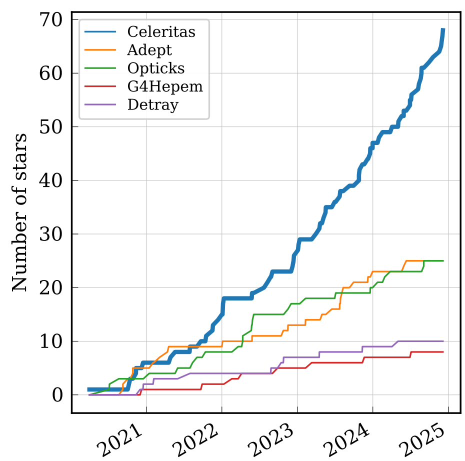


Figure 9. Github popularity of Celeritas and adjacent HEP GPU projects.

5. PROJECT PLAN

The priorities and work plan for the remainder of the project are based on the adjusted SciDAC scope (see Appendix A), which emphasizes scalability. The next phase of the project is to mature its capabilities to production readiness, and to provide novel workflows optimized for DOE LCFs.

We will work with the previously mentioned HEP experiments to stabilize the Celeritas/Geant4 integration interface and incorporate the minimum set of functionality to test full production runs. Once completed, Celeritas 1.0 will be distributed via Spack for HPC systems and experiment-specific external packages.

To ensure portability across a wide range of HPC architectures, the core transport engine will adopt experimental C++ parallelism functionality (e.g., "stdpar"), which is implemented by Nvidia and AMD. This "parallel algorithm" implementation will serve as a stepping stone to an implementation for Intel GPUs, if that platform continues to be viable.

To enhance scalability and provide new capabilities that can only exist at large scale, we will develop a standalone workflow targeted at computationally difficult HEP problems that need data reductions such as average detector responses. The first such application will be an app that constructs computationally expensive optical photon maps for dark matter experiments, including those used to simulate S2 signals in LZ (Akerib et al. 2021).

After the initial stable release of Celeritas, the application and library interfaces will be extended with Python-based drivers for facilitating user setup. These prototype interfaces will use JavaScript Object Notation (JSON) and REST Application Programming Interfaces (APIs) to allow both execution and data exploration in a multitude of environments, including remotely and through virtualization containers. Such interfaces will greatly lower the barrier to entry for students and other researches not enmeshed in HEP, and they can be more easily integrated with next-generation experiment frameworks such as Gaussino and Key4hep.

6. CONCLUSIONS

The evolution of GPU adoption within the HEP community reflects a profound transformation over the past few years. In the 2019 HEP computing roadmap (The HEP Software Foundation et al. 2019), the detector simulation section mentioned SIMD three times but referred to GPUs only once, reflecting the minimal interest in GPU-based solutions at the time. By contrast, in the recent European Strategy for Particle Physics Update (ESPPU) planning meeting, seven of nine presentations included substantial discussions of GPU simulation, and there was unanimous verbal agreement that GPU-enabled workflows are essential for the future of HEP software and computing. Celeritas has played a critical role in catalyzing this shift, demonstrating the feasibility of GPU-accelerated particle transport and leading the way for the integration of advanced computing technologies in HEP simulation.

The work presented in this report, along with many artifacts enumerated at https://celeritas.ornl.gov/ (43 presentations, 2 journal articles, 3 conference papers, and 5 major code releases), demonstrates remarkable progress in fulfilling the objectives of the Celeritas SciDAC. Substantial advancements have been made in physics modeling, advanced computing, and integration with HEP experiment frameworks. The successful implementation of standard EM physics on GPUs represents a major milestone, showcasing the potential for high-fidelity, platform-portable particle transport simulations on next-generation computing architectures.

Beyond technical achievements, Celeritas has had a notable impact on the HEP computing ecosystem. Its integration efforts with CMS and ATLAS have advanced the state of GPU-accelerated simulation and fostered closer collaboration between the physics and computer science domains, preparing the way for production use of Celeritas to transform advanced computing technologies into meaningful scientific discovery.

Looking ahead, the near-term priorities for Celeritas focus on expanding its physics capabilities, improving scalability on LCFs, and strengthening integration with experiment frameworks. These efforts will position Celeritas to meet the growing computational demands of the HEP community, ensuring compatibility with evolving HPC systems and delivering tools that empower researchers to answer computationally challenging hypotheses. With continued support and collaboration, Celeritas is poised to become an indispensable resource for next-generation detector simulation and a model for innovation at the intersection of physics and computing.

ACKNOWLEDGMENTS

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. Work for this paper was supported by Oak Ridge National Laboratory (ORNL), which is managed and operated by UT–Battelle, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC05-00OR22725 and by Fermi National Accelerator Laboratory, managed and operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy. 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.

APPENDIX A. PERSONNEL AND PROJECT CHANGES

APPENDIX A. PERSONNEL AND PROJECT CHANGES

The project has had several changes over its lifetime. First, between the proposal submission and start of funding, Vincent Pascuzzi (Brookhaven National Laboratory (BNL) PI) left the institution and it was removed from the SciDAC corresponding to the reduced funding level. After the two year mark, commencing with the extension of the project and reflecting the changing roles in the group, PIs Demarteau and Romano were replaced with Johnson and Lund, respectively.

Originally the project was scoped for five years with a cumulative $10M budget, but the initial grant reduced the level to two years for a cumulative $1.2M, three quarters of which was supported by the HEP office. The scope was similarly reduced to implement only EM physics. At the end of that period, a follow-on grant under the same B&R code effectively extended the project by three years with a total of $1.3M, over half of which is supported by Advanced Scientific Computing Research (ASCR). The extension is focused primarily on maturing the independent Celeritas front-end application into a HPC-optimized workflow.

APPENDIX B. REFERENCES

APPENDIX B. REFERENCES

Agostinelli, S., J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, et al. 2003. "Geant4—a simulation toolkit." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506, no. 3 (July): 250–303. ISSN: 01689002, accessed March 23, 2018.

Akerib, D.S., C.W. Akerlof, A. Alqahtani, S.K. Alsum, T.J. Anderson, N. Angelides, H.M. Araújo, et al. 2021. "Simulations of Events for the LUX-ZEPLIN (LZ) Dark Matter Experiment." *Astroparticle Physics* 125 (February): 102480. ISSN: 09276505, accessed April 28, 2023. https://doi.org/10.1016/j.astropartphys.2020.102480. https://linkinghub.elsevier.com/retrieve/pii/S0927650520300529.

Allen, John, and Ken Kennedy. 1984. "Automatic Loop Interchange." In *SIGPLAN Notlce8,* 19:233–246. June.

Allison, J., K. Amako, J. Apostolakis, P. Arce, M. Asai, T. Aso, E. Bagli, et al. 2016. "Recent developments in Geant4." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835:186–225. ISSN: 0168-9002. https://doi.org/https://doi.org/10.1016/j.nima.2016.06.125. https://www.sciencedirect.com/science/article/pii/S0168900216306957.

Allison, John, et al. 2006. "Geant4 developments and applications." *IEEE Trans. Nucl. Sci.* 53:270. https://doi.org/10.1109/TNS.2006.869826.

Apostolakis, John, Gabriele Cosmo, Andrei Gheata, Mihaela Gheata, Raman Sehgal, and Sandro Wenzel. 2019. "A Vectorization Approach for Multifaceted Solids in VecGeom." Edited by A. Forti, L. Betev, M. Litmaath, O. Smirnova, and P. Hristov. *EPJ Web of Conferences* 214:02025. ISSN: 2100-014X, accessed July 11, 2023. https://doi.org/10.1051/epjconf/201921402025. https://www.epj-conferences.org/10.1051/epjconf/201921402025.

Atchley, Scott, Christopher Zimmer, John Lange, David Bernholdt, Veronica Melesse Vergara, Thomas Beck, Michael Brim, et al. 2023. "Frontier: Exploring Exascale." In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis.* SC '23. Denver, CO, USA: Association for Computing Machinery. https://doi.org/10.1145/3581784.3607089. https://doi.org/10.1145/3581784.3607089.

ATLAS Collaboration. 2021. *Athena.* V. 21.0.127, May. https://doi.org/10.5281/zenodo.2641996. https://zenodo.org/records/4772550.

Bandieramonte, Marilena, Riccardo Maria Bianchi, and Joseph Boudreau. 2020. "FullSimLight: ATLAS Standalone Geant4 Simulation." Edited by C. Doglioni, D. Kim, G.A. Stewart, L. Silvestris, P. Jackson, and W. Kamleh. *EPJ Web of Conferences* 245:02029. ISSN: 2100-014X, accessed July 13, 2023. https://doi.org/10.1051/epjconf/202024502029. https://www.epj-conferences.org/10.1051/epjconf/202024502029.

Brun, Rene, Fons Rademakers, Philippe Canal, Axel Naumann, Olivier Couet, Lorenzo Moneta, Vassil Vassilev, et al. 2020. *root-project/root: v6.18/02.* V. v6-18-02, June. https://doi.org/10.5281/zenodo.3895860. https://doi.org/10.5281/zenodo.3895860.

Center, National Energy Research Scientific Computing. 2022. *Perlmutter.* https://www.nersc.gov/systems/perlmutter.

CMS Offline Software and Computing. 2021. *Evolution of the CMS Computing Model towards Phase-2.* CMS Note 2021/001. Geneva. Switzerland: CERN, January. https://cds.cern.ch/record/2751565/files/NOTE2021_001.pdf.

Gamblin, Todd, Matthew LeGendre, Michael R. Collette, Gregory L. Lee, Adam Moody, Bronis R. de Supinski, and Scott Futral. 2015. "The Spack Package Manager: Bringing Order to HPC Software Chaos." Supercomputing 2015 (SC'15). Austin, Texas, US, November. https://tgamblin.github.io/pubs/spack-sc15.pdf.

Hanson, A. O., L. H. Lanzl, E. M. Lyman, and M. B. Scott. 1951. "Measurement of Multiple Scattering of 15.7-Mev Electrons." *Phys. Rev.* 84 (4): 634–637. https://doi.org/10.1103/PhysRev.84.634. https://link.aps.org/doi/10.1103/PhysRev.84.634.

Johnson, Seth R., Stefano Castro Tognini, Elliott Biondo, Thomas Evans, Julien Esseiva, Philippe Canal, Amanda Lund, et al. 2024. *Celeritas R&D Report: Accelerating Geant4.* ORNL/TM-2023/3204. January 1, 2024. Accessed January 23, 2024. https://doi.org/10.2172/2281972. https://www.osti.gov/servlets/purl/2281972/.

Johnson, Seth R., Julien Esseiva, Elliott Biondo, Philippe Canal, Marcel Demarteau, Thomas Evans, Soon Yung Jun, et al. 2024. "Celeritas: Accelerating Geant4 with GPUs." *EPJ Web of Conferences* 295:11005. https://doi.org/10.1051/epjconf/202429511005.

Johnson, Seth R., Stefano C. Tognini, Philippe Canal, Thomas Evans, Soon Yung Jun, Guilherme Lima, Amanda Lund, and Vincent R. Pascuzzi. 2021. "Novel Features and GPU Performance Analysis for EM Particle Transport in the Celeritas Code." Edited by C. Biscarat, S. Campana, B. Hegner, S. Roiser, C.I. Rovelli, and G.A. Stewart. *EPJ Web of Conferences* 251:03030. ISSN: 2100-014X, accessed March 31, 2023. https://doi.org/10.1051/epjconf/202125103030. https://www.epj-conferences.org/10.1051/epjconf/202125103030.

Lachnit, Stephan, Lorenzo Pezzotti, and Dmitri Konstantinov. 2022. *Standalone Geant4 validation of the ATLAS Tile Calorimeter.* Technical report CERN-STUDENTS-Note-2022-069. CERN, August.

Maheshwari, Ketan, William Arndt, Ahmad Maroof Karimi, Junqi Yin, Frédéric Suter, Seth R Johnson, and Rafael Ferreira da Silva. 2024. "GNU Parallel: Enabling Low-Overhead HT-HPC Workflows at Extreme Scale." The 19th Workshop on Workflows in Support of Large-Scale Science. Atlanta, GA, November.

Meroli, S, D Passeri, and L Servoli. 2011. "Energy Loss Measurement for Charged Particles in Very Thin Silicon Layers." *Journal of Instrumentation* 6, no. 06 (June): P06013–P06013. ISSN: 1748-0221, accessed May 7, 2024. https://doi.org/10.1088/1748-0221/6/06/P06013.

The HEP Software Foundation, Johannes Albrecht, Antonio Augusto Alves, Guilherme Amadio, Giuseppe Andronico, Nguyen Anh-Ky, Laurent Aphecetche, et al. 2019. "A Roadmap for HEP Software and Computing R&D for the 2020s." *Computing and Software for Big Science* 3, no. 1 (December): 7. ISSN: 2510-2036, 2510-2044, accessed October 1, 2019. https://doi.org/10.1007/s41781-018-0018-8. http://link.springer.com/10.1007/s41781-018-0018-8.

Tognini, S. C., P. Canal, T. M. Evans, G. Lima, A. L. Lund, S. R. Johnson, S. Y. Jun, V. R. Pascuzzi, and P. K. Romano. 2022. *Celeritas: GPU-accelerated particle transport for detector simulation in High Energy Physics experiments.* ArXiv:2203.09467 [hep-ex, physics:physics], March. Accessed March 31, 2023. http://arxiv.org/abs/2203.09467.