AdePT
Accelerated Particle Transport

AdePT Perspectives

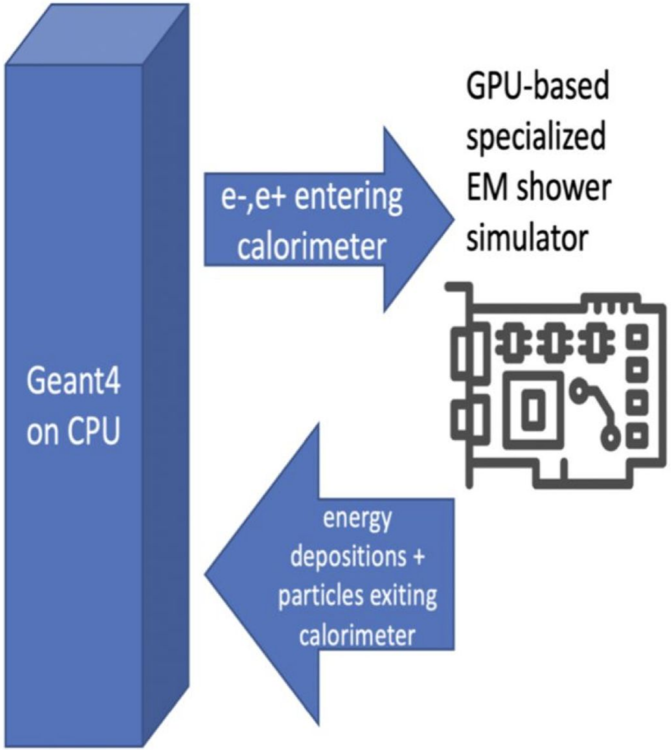
AdePT Developers
6 May 2022

Overview of AdePT today

- AdePT has implemented a complete simulation of EM showers on GPU, validated against equivalent Geant4 CPU problems
- Some remaining features in development
  - Map-based magnetic fields
  - User defined scoring: hooks for this present, requires demonstration of feasibility for use cases more complex than energy deposit summation
- No blockers currently identified that would prevent addressing the full complexity of the LHC and other HEP experiments
- Nevertheless, much work to be done to fully optimize performance for production scenarios and ensure easy integration by experiment frameworks

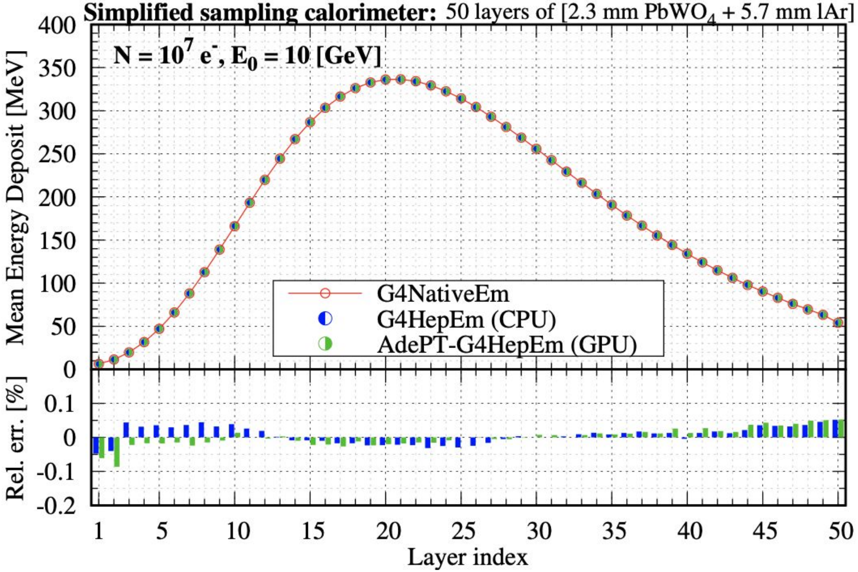
AdePT as an accelerator

- Proof-of-concept use in hybrid CPU-GPU applications driven by Geant4 on the host
- Use of standard Geant4 fast simulation hooks to offload $e^-/e^+/\gamma$ in specific regions to GPU
- Other particles/processes left to CPU, enabling full simulation of HEP events with tracks on appropriate hardware
- Should provide a route for experiments to integrate in their workflows



EM Physics with G4HepEM

- Geant4 R&D project to optimize EM physics models for the HEP use case
- Targeting CPU, but design with GPU in mind enabled use in AdePT
- Validated against standard Geant4 models on both CPU and GPU
- Candidate for inclusion in Geant4 for CPU use, and experiments already testing CPU version



	GEANT4	G4HepEm			
	G4Em (CPU)	HepEm (CPU)	Rel. err. [%]	AdePT (GPU)	Rel. err. [%]
Energy deposit per material [MeV]					
$PbWO_4$	6730.00	6729.46	-0.008	6729.46	-0.008
lAr	2566.18	2566.55	0.014	2566.52	0.013
Number of secondaries					
γ	4456.06	4454.57	-0.033	4451.46	-0.103
e^-	8066.48	7953.90	-1.40*	7953.22	-1.40*
e^+	429.103	429.146	0.010	429.147	0.010
Number of steps					
charged	36696.7	36283.2	-1.13*	36292.0	-1.10*
neutral	40377.9	40426.6	0.121	40597.9	0.545

Magnetic Field

- Interfaces and workflow for track propagation in fields in place
  - Already tested and partially optimized in the AdePT advanced examples
  - Current implementation revealed to be an important source of GPU work imbalance in complex setups, introducing larger performance degradation than on the CPU
- Challenge to reduce divergence due to different number of segments/trials
  - Important for performance - at least at 10-20% level in current tested geometries , more complex geometries require testing
  - Tuning and revised scheme underway
- Development of Runge-Kutta field integration in preparation
  - After validation, will proceed to method / parameter tuning for performance, and enhancement
  - Keen to incorporate evaluation of non-uniform fields via ‘texture map’ (e.g. from ACTS)

Overview of Performance

- AdePT and Celeritas demonstrate that we can run HEP simulation workflows on GPU systems
  - Even if full potential not yet realized, it means utilization of these resources is now possible
- AdePT's prototypes shows equivalent peak performance between an oversubscribed dual core machine with 2x16 cores and a medium class consumer GPU card for a standalone sampling calorimeter example.
- Key development item will be a set of HEP "problems" for benchmarking realistic production workflows on standard/expected hardware.
  - Discussed in more detail later

Opportunities for Optimization

- Profiling has revealed several areas for improvement:
  - Splitting kernels to reduce divergent paths
  - Track batch sizes or concurrent tracks from multiple events to maximize tracks-in-flight
  - Data structures and management to improve coalesced access
  - Suitability of energy fluctuation algorithm for GPU
- Detailed analysis indicates greatest potential for improvement lies in restructuring geometry code for GPU
  - Reducing complexity
  - Use of surface-based models to allow more balanced workloads
  - Support non-CUDA toolchains/accelerators
- Working with Celeritas on investigations of geometry (ORANGE), with input from other projects considering the geometry topic such as ACTS

Optimization Challenges

- Scheduling of off/onload of tracks to GPU in hybrid CPU/GPU applications is still an open area for R&D
  - Some limitations known, but many scenarios to explore both for performance and usability/compatibility with experiment's frameworks
  - Welcome input from experiments here, and can work with you to explore options.
- Larger performance degradation on GPU for strong magnetic fields in complex geometry compared to the CPU
  - Mainly due to the large work imbalance specific to field propagation, penalizing the GPU more than the CPU
  - An important work site for evolving the prototype
  - Benchmarking with realistic geometry/fields critical to give accurate picture of bottlenecks and necessary code and kernel structure

Future Directions

- AdePT (and Celeritas) have clearly demonstrated that HEP workflows for EM calorimetry can be run on GPU, but much work still to do across many areas
- **Performance:**
  - Optimization of identified bottlenecks, metrics for realistic HEP production scenarios
- **Common Components:**
  - Identify components projects could share and reuse (e.g. physics, geometry)
- **Use of AdePT for GPU off/onload within Geant4-based applications**
  - Give experiments flexibility in use whilst yielding sufficient performance improvement

Common Benchmarks for Performance

- Critical that a set of realistic benchmark "problems" are established for HEP use cases, to allow comparison of pure Geant4 CPU against hybrid Geant4+GPU, with common input parameters:
  - Geometry
  - Physics
  - Input event samples
- Whilst unit tests of individual components are beneficial for developers, only benchmarking a full workflow provides an accurate measure of the performance achievable in a production setting
- We would value discussion with experiments on what you would like to see these problems output and measure, e.g.
  - Physics quantities per-event or ensemble to be scored for validation/regression testing
  - Host/Device performance metrics to record or otherwise measure
  - Range of Host/Device hardware configurations the above should be produced for
  - Number of jobs to run for a given configuration
- Initial discussions with Celeritas on these topics so we can use the same setups, but need input from the experiments to ensure we are optimizing for what you require!

Identifying Common Components

- AdePT and Celeritas have developed independently but meet regularly to discuss progress and exchange knowledge
  - Different approaches to the problem extremely beneficial in exploring designs and expect this to continue through R&D phase
- Codes are however mature enough to start identifying areas for closer collaboration, or convergence towards common code for specific components
  - Has already begun with the VecGeom/ORANGE geometry codes and evolution
- Several other areas we could explore:
  - Data management/organization on device
  - Physics data
  - Methods for integration in a hybrid Geant4 CPU/GPU application
- Equally, are there components we can/should reuse from the broader HEP/HPC communities?
  - Contribution/engagement with effort and ideas very welcome!

Integration in Hybrid Geant4 Workflows

- Only off/onload to AdePT via Geant4's fast simulation hooks has been explored so far.
- Want to investigate and benchmark other options, including the customizable per-particle type G4VTrackingManager in Geant4 v11.0.
- Another area of commonality to explore with Celeritas (the Acceleritas sub-project) especially given the different implementations of workflow
- Important to track GPU workflow evolutions that affect CPU/GPU integration, for example, how to merge hits from concurrent tracks across multiple events on the GPU back into hits-per-event data on the CPU.
- As with benchmarking, we would value input from experiments here to ensure we are providing the interfaces you require and can deliver the necessary performance improvement.

Conclusions

- We can run HEP EM Calorimetry workflows on GPU with physics validated against Geant4!
- Key performance bottlenecks identified, with ongoing work to investigate these in both GPU-only and CPU+GPU workflows
- Work now to develop set of HEP benchmark problems to establish and optimize performance in realistic production scenarios
  - Input from the experiments vital here
- Work with Celeritas colleagues valuable in exploring different designs
  - Starting to explore areas for common code
- Topic of Geometry on GPU a major one going forward, and VecGeom/ORANGE collaboration very valuable here

AdePT
Accelerated Particle Transport

G4HepEm R&D