Plan of work for R&D activities

Witek Pokorski

11.04.2024

Geant4 Technical Forum

Content

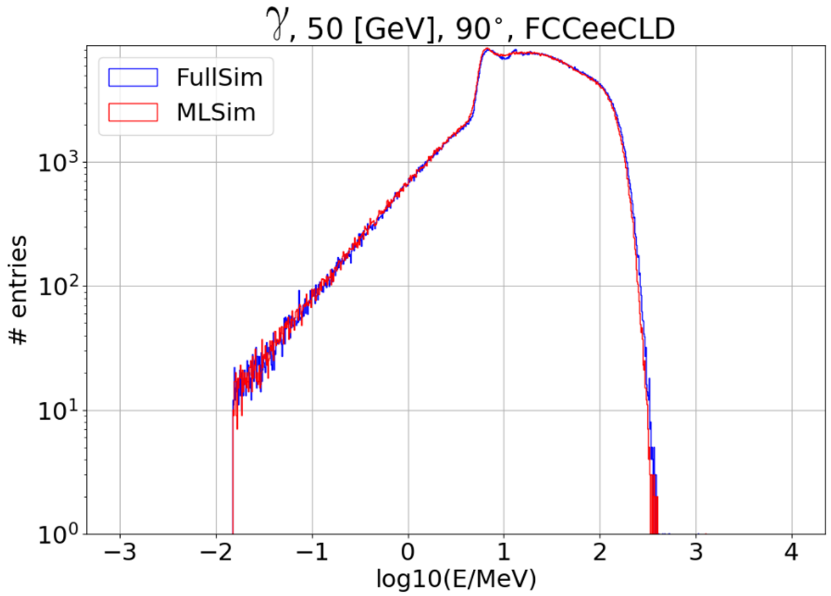
- fast simulation techniques
  - Machine-Learning based models
- exploration of new hardware (GPUs)
  - general transport code prototypes
  - domain specific application

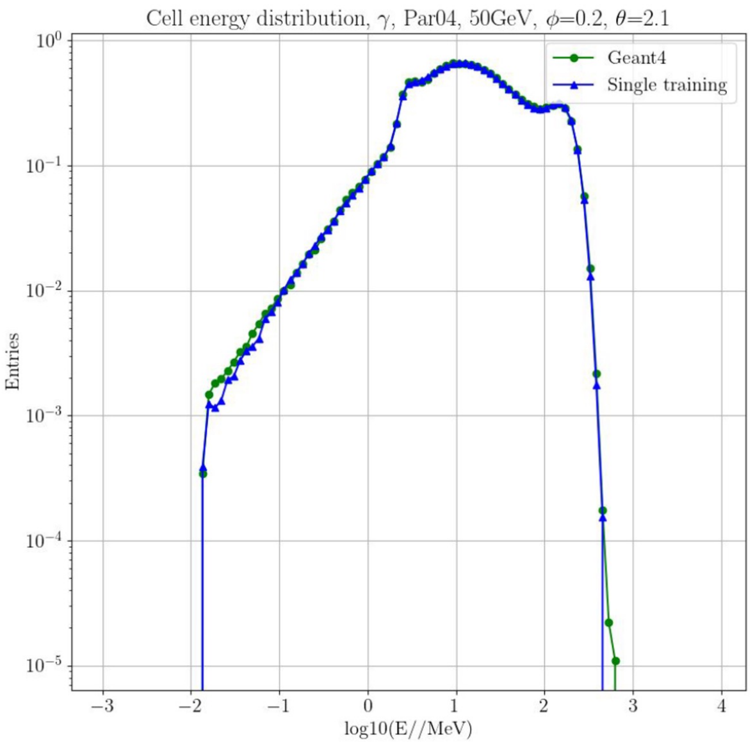
Anna Zaborowska
Piyush Raikwar
Peter McKeown

Fast shower simulation: ML models

Transformer-based ML models:

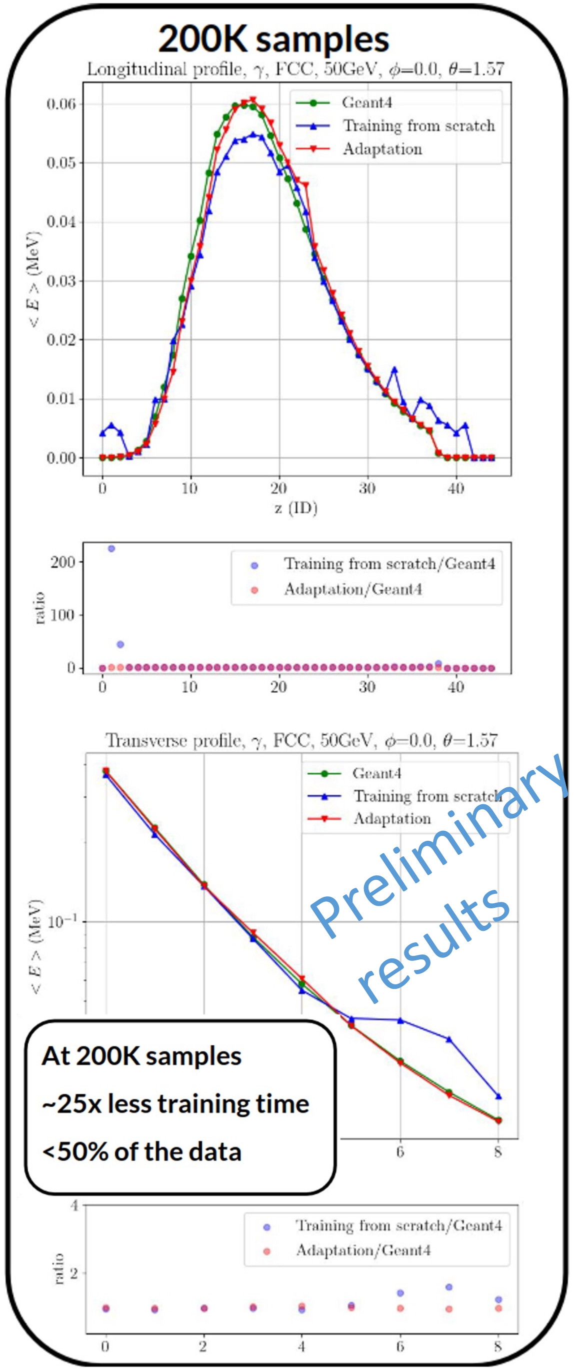
- Improve CaloDiT diffusion model (inference optimisation - making it faster)
- Investigate all potential architectures (VAE-based, GAN-based, diffusion-based) within IBM collaboration
- Test the benefits of starting from the pretrained model





ACAT 2024 presentation

FCC Physics Week



Fast shower simulation: Application of ML models

LHCb - infrastructure ready, joint work with LHCb on:

- Testing the CaloDiT on EM showers
- Optimisation of granularity for hadronic showers based on full sim studies
- Adaptation of transformer-based architecture to fit resulting hadronic data

ATLAS

- Support the work of making the “classical” shower parameterisation (FastCaloSimV2-based) able to work on any detector (with custom tuning), dependency on Geant4 only
- Contribute to the work on finding the best shower granularity and validation studies
- Prepare the infrastructure for testing the developed transformer-based models

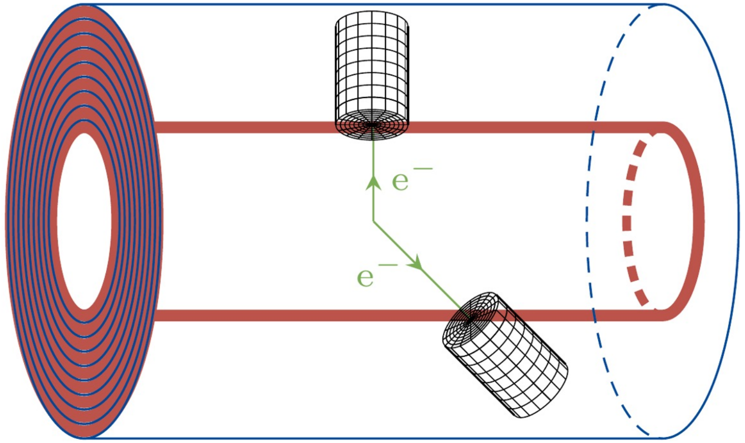
CMS

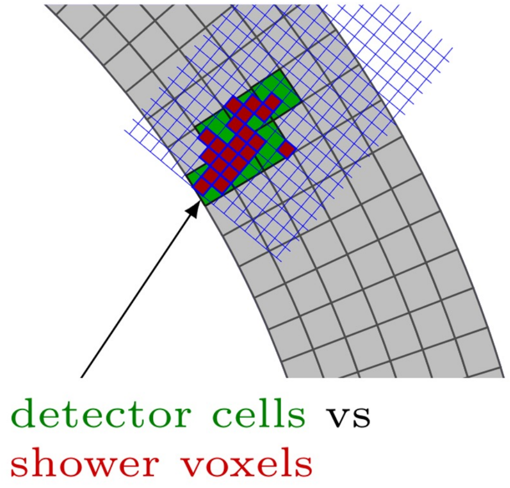
- Follow-up on started activities to produce a training sample for HGCal, implementing our proposed cell-independent infrastructure to test our (transformer-based) models

Oriented crystals detector

- Support work of Alexei Sytov et al. to speed-up oriented crystals simulation using fast shower simulation with ML models

Infrastructure: scoring mesh





Fast shower simulation: Community work

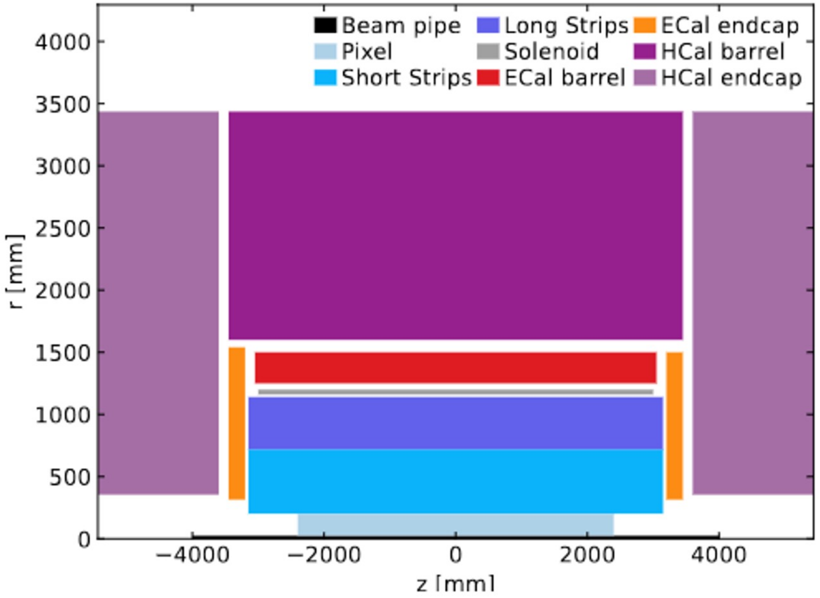
Open Data Detector

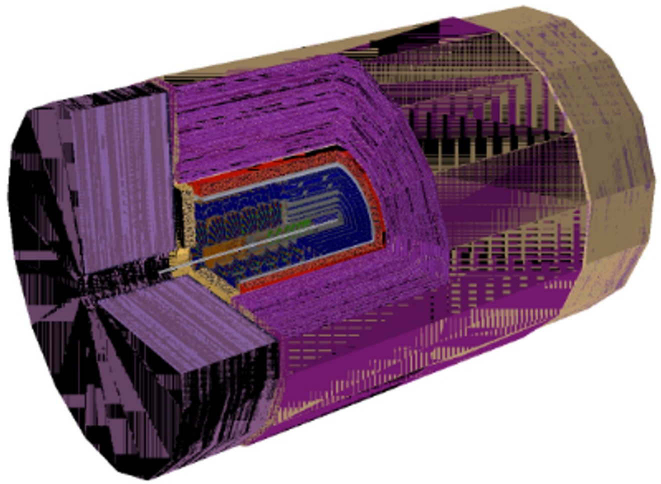
- Prepare input cell-level data for the demonstrator of ATLAS-derived FastCaloSimV2-based classical shower simulation, thus operating in the key4HEP software and using purely Geant4
- Generation of combined tracker-calorimeter dataset

Other community efforts

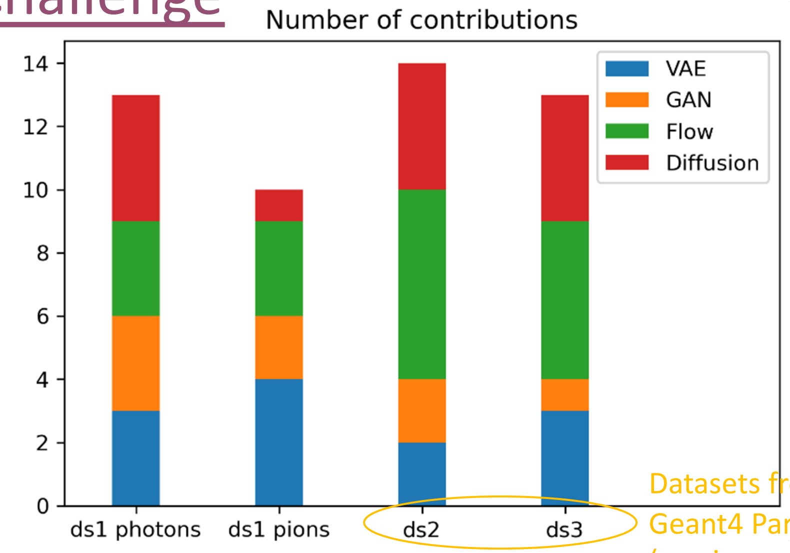
- Finalisation of CaloChallenge write-up
- Potential launch of the next edition, featuring more data, new detectors (e.g. ODD), and different data representation of the same showers, including the reconstruction validation

ML4jets 2023: Calo Challenge





ML4jets 2023: ODD



GPU-based transport codes

- Geant4 assessment of the R&D prototypes organized 13-14 December 2023
- AdePT and Celertitas projects presented in details and discussed
  - current status
  - main challenges
  - plans
  - possible synergy (common interface)

- Report from the assessment panel available on request

AdePT 2023 status

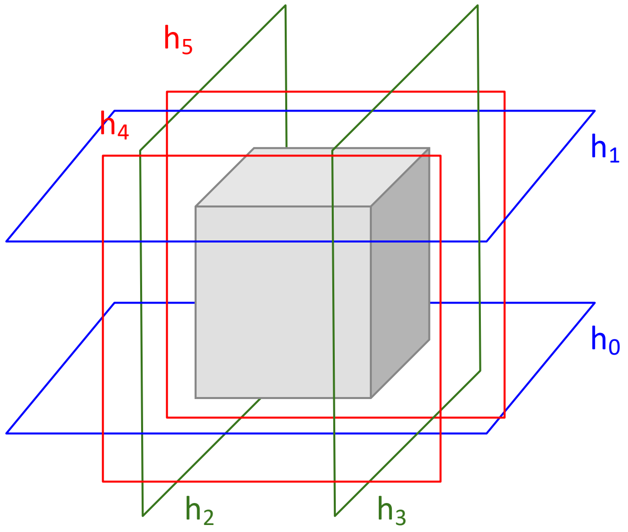
- Achieved the initial goals of the R&D
  - Understand usability of GPUs for general particle transport simulation, seeking for potential speed up and/or usage of available GPU resource for the HEP simulation
    - Prototype e+, e− and γ EM shower simulation on GPU, evolve to realistic use-cases
      - Geometry: VecGeom library, Physics: G4HepEm library
- Transport for EM particles working on GPUs for LHC-complexity geometries
  - Excellent physics agreement within statistical fluctuation
  - Reproducibility of the simulation achieved
- Full integration with Geant4 applications
  - Possible to plug AdePT into existing Geant4 applications with minimal extra code
    - reusing existing sensitive detector implementations
- Main bottleneck: geometry – being addressed now (see following slides)

AdePT Plans

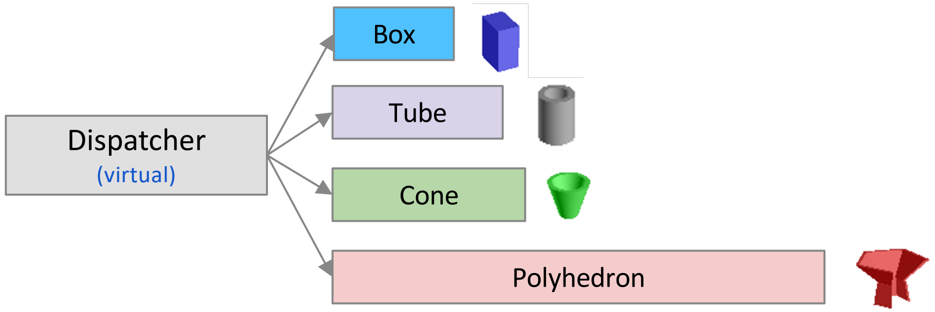
- Test the integration with LHC experiments
  - work ongoing for LHCb integration
  - ATLAS testbeam application implemented
  - CMS integration being studied
- Further testing and optimisation of magnetic field propagation
- Implementation of magnetic field map
- Benchmarking, profiling, etc
  - major testing benchmarking with the new geometry model (see next slides)
- Ongoing research into asynchronous scheduling strategies
  - The current strategy for scheduling work on the GPU is not optimal
    - In certain situations, this may even hinder CPU performance
      - GPU can block CPU threads when saturated
  - Initial results using a new scheduling strategy are promising

Bounded surface modeling for improving GPU performance

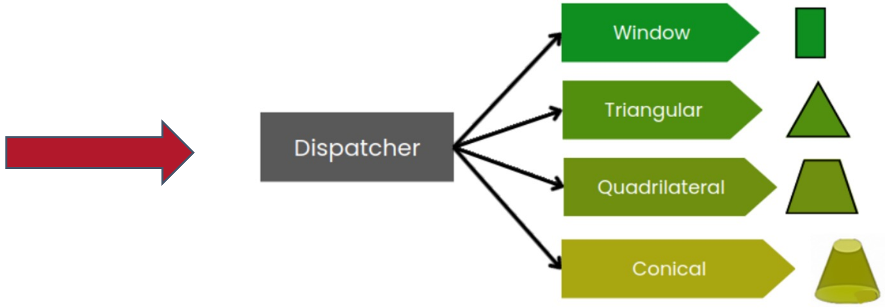
- Portable GPU-friendly header library
  - Algorithmic part independent on the backend, compilable with any native/portability compiler
  - Headers templated on the precision type to allow for a single-precision mode
- Code simplification and GPU performance
  - No virtual calls, no recursions, more work-balanced
  - Better device occupancy and kernel coherence
  - Reducing divergence and register usage on the GPU



Significant divergence in the solid model



Reduced divergence using surfaces



Status and plans

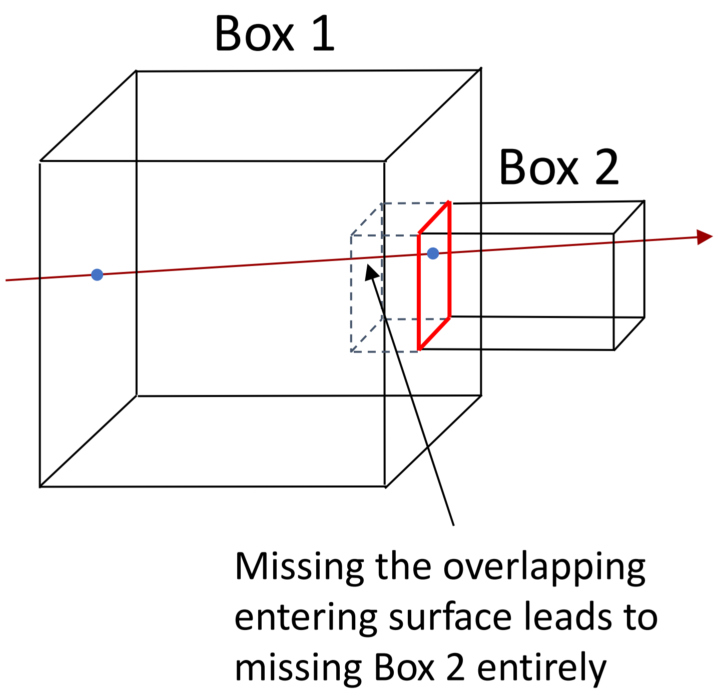
Surface model supports **all solids** required by the experiments

Current priorities:

- Debugging complex geometries (CMS, LHCb):
  Overlaps require relocations
- Implementing **Bounding Volume Hierarchy** accelerating structure

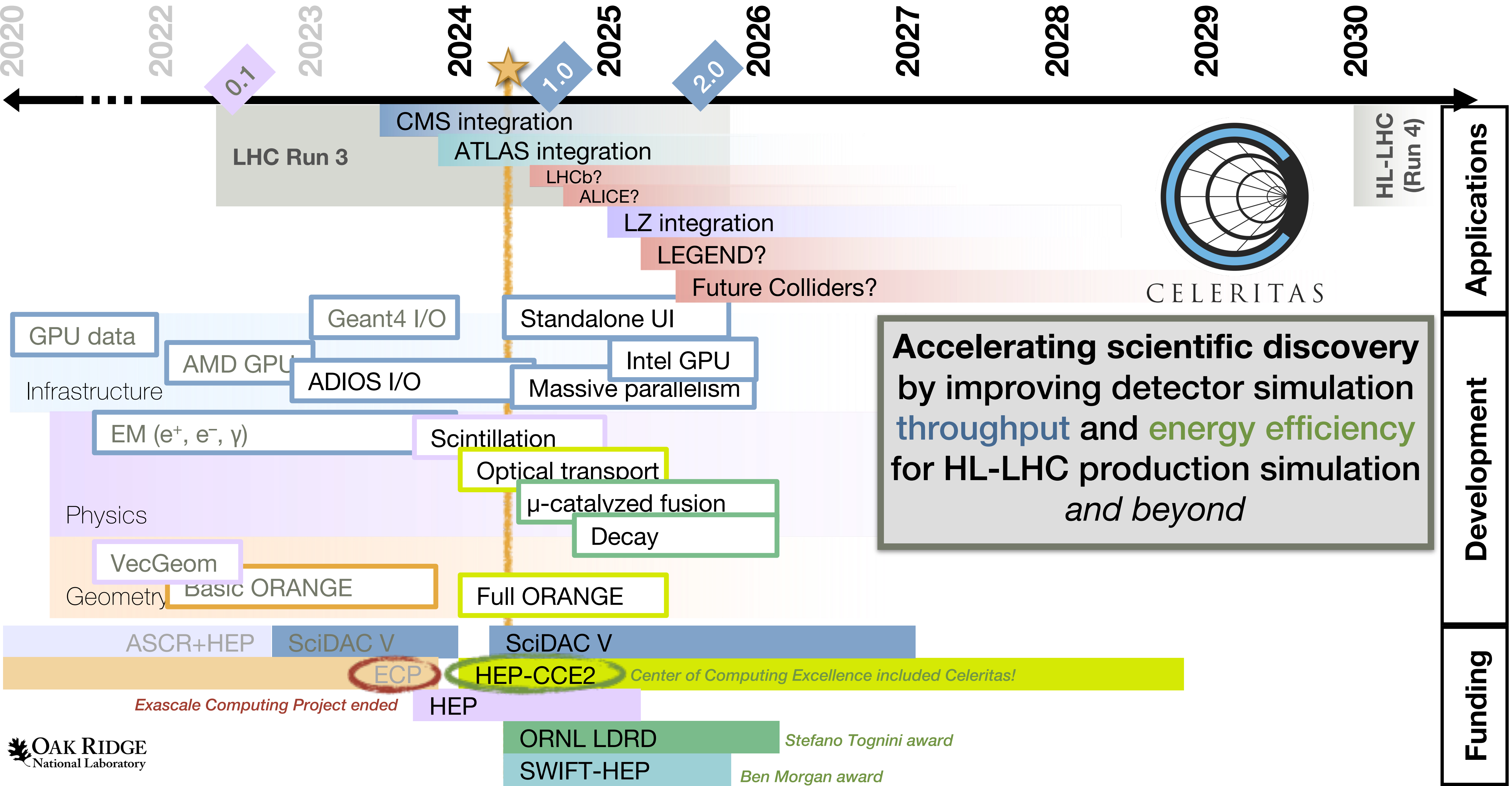
Next:

- Performance measurements, testing calorimeters (ATLAS EMEC, CMS HGCAL, LHCb ECAL)
- Lightweight portability layer (instead of pure CUDA)



10

Celeritas long-term timeline



Accelerating scientific discovery by improving detector simulation throughput and energy efficiency for HL-LHC production simulation ***and beyond***

Current tasks and next-year goals

- Integrate into **LHC experiment workflows**
- **Release Celeritas 1.0** through Spack
- Demonstrate **massive parallel execution** of Celeritas on DOE Leadership Computing Facility hardware

- Implement platform portable **optical physics** models (Nvidia/AMD) for standalone GPU
- Develop streamlined **user interface** for detector R&D

- Integrate GPU **optical photon event loop** into Celeritas with verification and baseline performance on (simplified) LZ geometry models
- Develop **GPU-enabled surface-based shape models** needed for ATLAS, CMS, and other experimental detector models

- Implement physics models for muon-catalyzed fusion

- Improve **Geant4 integration** using tracking manager
- Build **generic offload interface** with AdePT

US DOE	**SciDAC V**
US DOE	**CalVision**
US DOE	**HEP-CCE2**
	ORNL LDRD
UK	**SWIFT-HEP**

Funding Sources

Opticks/G4CXOpticks/CaTS: Full Integration of Opticks with Geant4. Status and Plans

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/

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 based on legacy version of Opticks (based on OptiX 6) was included in Geant4 11.0 as an advanced example.

https://geant4.kek.jp/lxr/source/examples/advanced/CaTS/

Status:

- Opticks has been completely reengineered by Simon Blyth migrating to OptiX7.
- The new Opticks (NVIDIA OptiX7) has been fully functional since January 2024.
- New Opticks APIs are 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!

Plans:

- Update the Geant4 advanced example CaTS to use the new Opticks. This will be part of the fall release.
- Provide easy build mechanism based on spack.
- Integrate with experimental frameworks (LArSoft).
- Enhance functionality (Wavelength shifting process).
- Provide detailed full-scale example of a liquid Argon Time projection Chamber (LArTPC) for optical simulation.
- Provide detailed documentation of Geant4 optical physics processes relevant to LArTPCs.
- Document the LArTPC material properties obtained from literature used in simulation.
- Contribute to Geant4 optical physics and documentation.

Conclusion

- generic ML fast simulation models start to be applied in the experiments with further developement planned
- significant progress in GPU protypes
  - planned testing of full integration with experimental frameworks
  - completion of the new surface model the coming months
- CaTS moving to new version of Opticks based on NVidia OptiX7