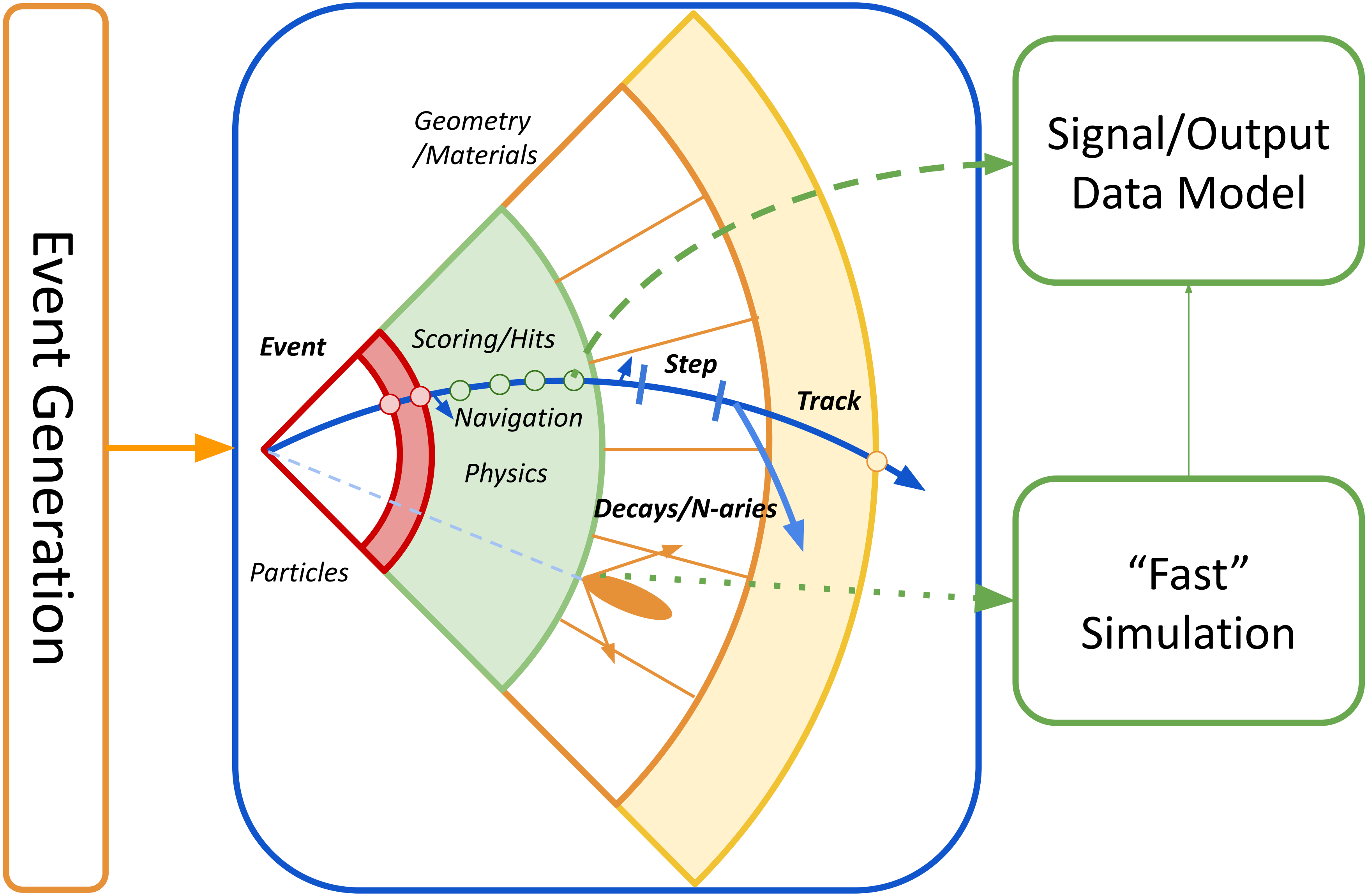
Detector Simulation in ExCALIBUR-HEP and towards SwiftHEP

Ben Morgan

SwiftHEP/ExCALIBUR-HEP Workshop
Online, 14th January 2021



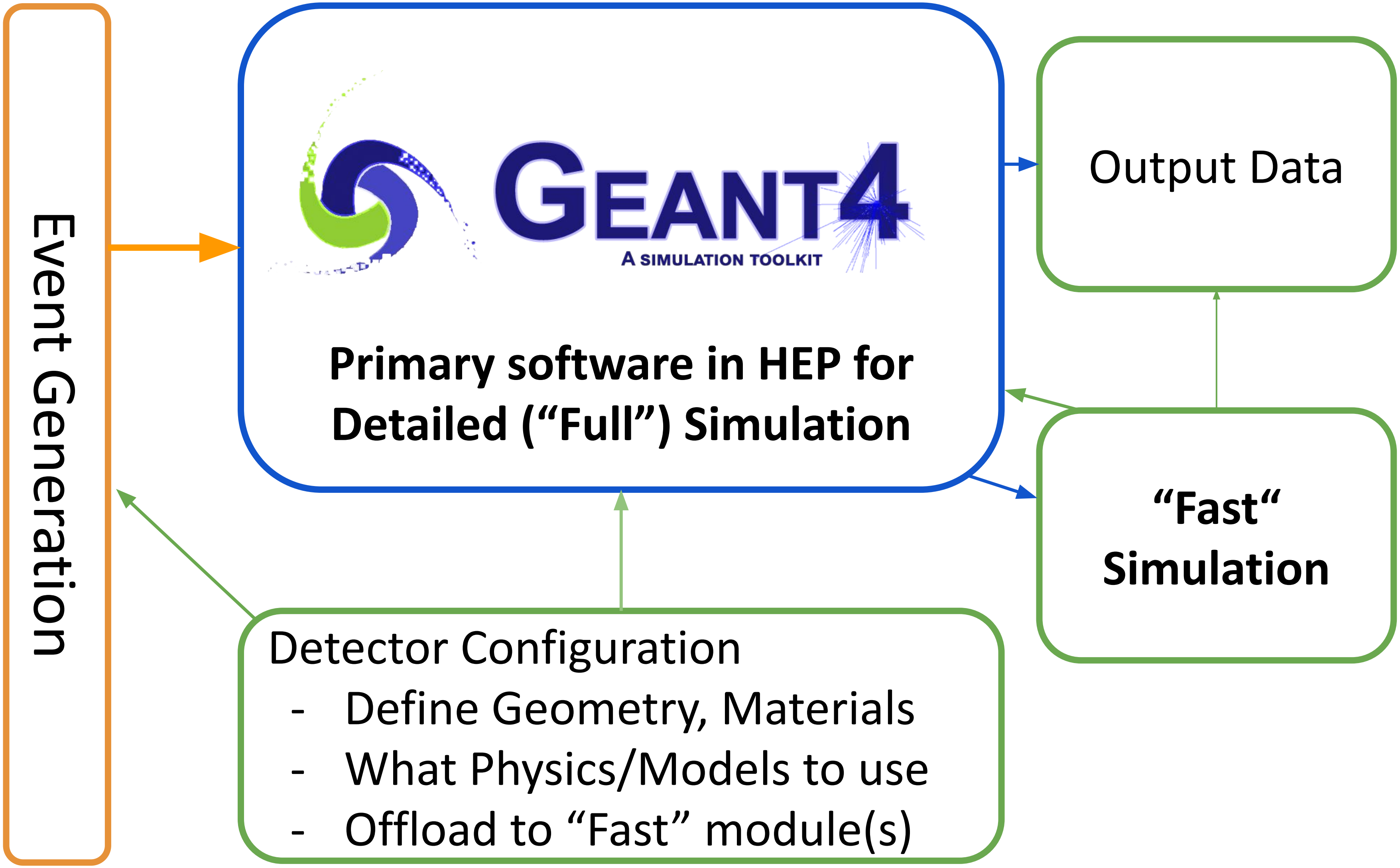
Event Generation

Signal/Output Data Model

“Fast” Simulation

Detector Simulation in Two Slides...

Event Generators covered earlier.



Primary software in HEP for Detailed ("Full") Simulation

Detector Configuration

- Define Geometry, Materials
- What Physics/Models to use
- Offload to "Fast" module(s)

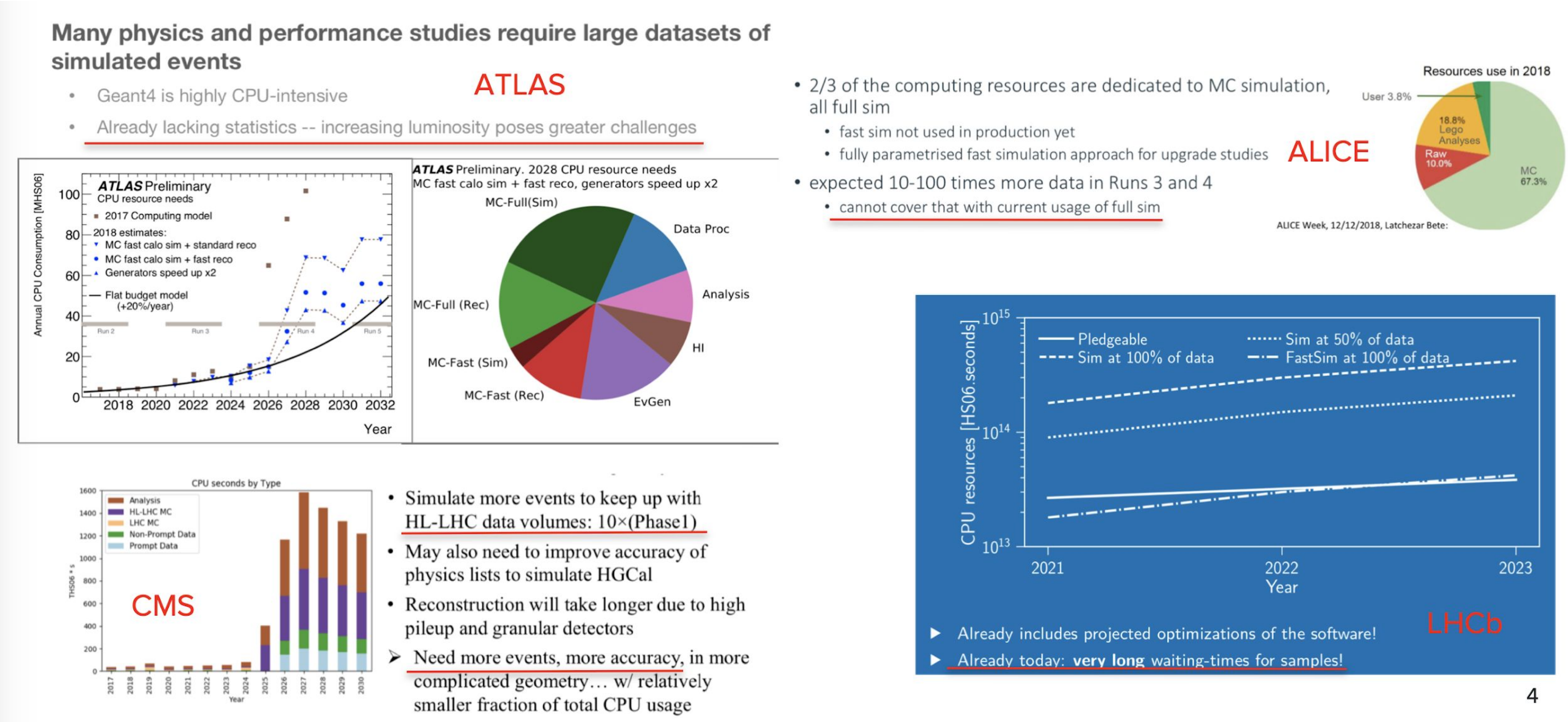
Detector Simulation in Two Slides...

http://geant4.cern.ch

Geant4 Development Objectives

1. **Long term service to experiment/stakeholder requirements for production use**
   ○ *Functionality, physics, and API needs*
   ○ *Support, maintenance, testing, and validation*
2. **Continuous development to improve:**
   ○ *Precision of, add new, physics models*
   ○ *Computational performance*
   ○ *Functionality, maintaining stability and sustainability for experiment/stakeholders*
3. **Monitor and plan for experiment/computing changes**
   ○ *Computation challenges: e.g. HL-LHC, Intensity Frontier*
   ○ *Evolution of architectures/APIs: GPU, FPGA, ARM etc*

Forecast Simulation Needs: LHC



Many physics and performance studies require large datasets of simulated events

- Geant4 is highly CPU-intensive
- Already lacking statistics -- increasing luminosity poses greater challenges

ATLAS

- 2/3 of the computing resources are dedicated to MC simulation, all full sim
  - fast sim not used in production yet
  - fully parametrised fast simulation approach for upgrade studies
- expected 10-100 times more data in Runs 3 and 4
  - cannot cover that with current usage of full sim

ALICE

ALICE Week, 12/12/2018, Latchezar Bete:

CMS

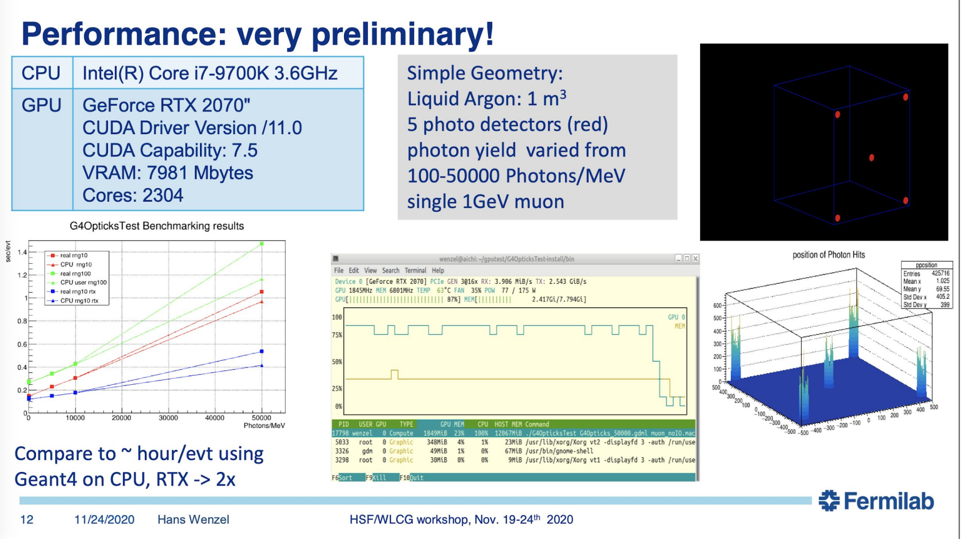
- Simulate more events to keep up with HL-LHC data volumes: 10×(Phase1)
- May also need to improve accuracy of physics lists to simulate HGCal
- Reconstruction will take longer due to high pileup and granular detectors

➢ Need more events, more accuracy, in more complicated geometry… w/ relatively smaller fraction of total CPU usage

- ▶ Already includes projected optimizations of the software!
- ▶ Already today: **very long** waiting-times for samples!

LHCb

- [R&D directions for performance improvement in Geant4 (Andrei Gheata)]
- [Detector simulation requirements from HEP Experiments (Marilena Bandieramonte)]



Performance: very preliminary!

CPU	Intel(R) Core i7-9700K 3.6GHz
GPU	GeForce RTX 2070" CUDA Driver Version /11.0 CUDA Capability: 7.5 VRAM: 7981 Mbytes Cores: 2304

Simple Geometry:
Liquid Argon: 1 m^3
5 photo detectors (red)
photon yield varied from
100-50000 Photons/MeV
single 1GeV muon

G4OpticksTest Benchmarking results

wenzel@aichi: ~/gputest/G4OpticksTest-install/bin

```
Device 0 [GeForce RTX 2070] PCIe GEN 3@16x RX: 3.906 MiB/s TX: 2.543 GiB/s
GPU 1845MHz MEM 6801MHz TEMP  63°C FAN  35% POW  77 / 175 W
GPU[||||||||||||||||||||||||||| 87%] MEM[||||||||||      2.417Gi/7.794Gi]
```

```
  PID  USER GPU    TYPE      GPU MEM    CPU  HOST MEM Command
17798 wenzel   0 Compute  1849MiB  23%  100%  12067MiB ./G4OpticksTest G4Opticks_50000.gdml muon_noIO.mac
 5033   root   0 Graphic   348MiB   4%    1%     23MiB /usr/lib/xorg/Xorg vt2 -displayfd 3 -auth /run/use
 3326    gdm   0 Graphic    49MiB   1%    0%     67MiB /usr/bin/gnome-shell
 3298   root   0 Graphic    30MiB   0%    0%      9MiB /usr/lib/xorg/Xorg vt1 -displayfd 3 -auth /run/use

F6Sort    F9Kill    F10Quit
```

position of Photon Hits

pposition	
Entries	425716
Mean x	1.025
Mean y	69.55
Std Dev x	405.2
Std Dev y	399

Compare to ~ hour/evt using
Geant4 on CPU, RTX -> 2x

Fermilab

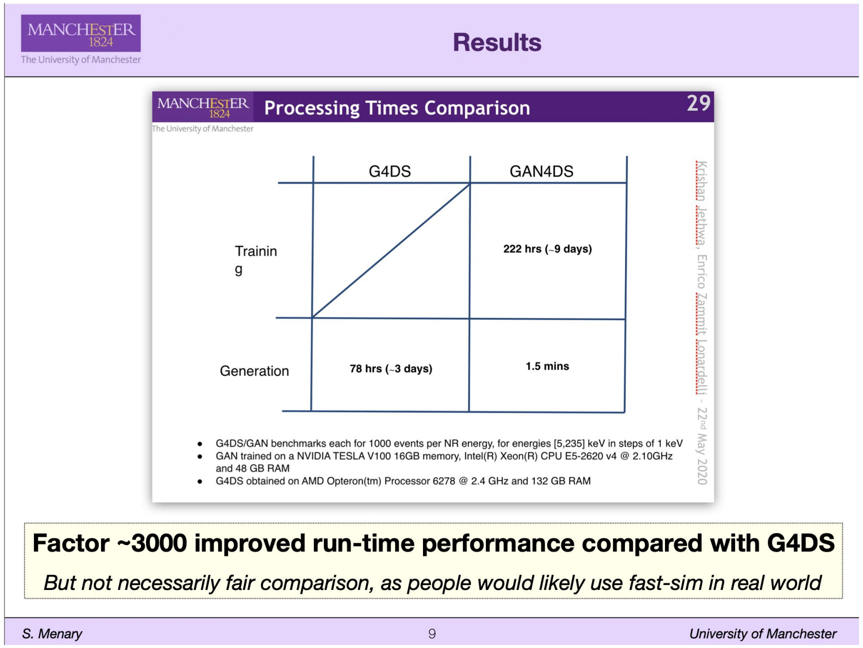
12 11/24/2020 Hans Wenzel HSF/WLCG workshop, Nov. 19-24th 2020

Using Geant4 and Opticks to simulate Liquid Argon TPCs, Hans Wenzel

Hybrid CPU+GPU "Full" Simulation

Towards a Fast Simulation of Light Detection in LAr TPCs, Stephen Menary

"Fast" Simulation of optical photons



MANCHESTER 1824
The University of Manchester

Results

MANCHESTER 1824 The University of Manchester

Processing Times Comparison 29

	G4DS	GAN4DS
Training		222 hrs (~9 days)
Generation	78 hrs (~3 days)	1.5 mins

- G4DS/GAN benchmarks each for 1000 events per NR energy, for energies [5,235] keV in steps of 1 keV
- GAN trained on a NVIDIA TESLA V100 16GB memory, Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz and 48 GB RAM
- G4DS obtained on AMD Opteron(tm) Processor 6278 @ 2.4 GHz and 132 GB RAM

Krishan Jethwa, Enrico Zammit Lonardelli - 22nd May 2020

Factor ~3000 improved run-time performance compared with G4DS

But not necessarily fair comparison, as people would likely use fast-sim in real world

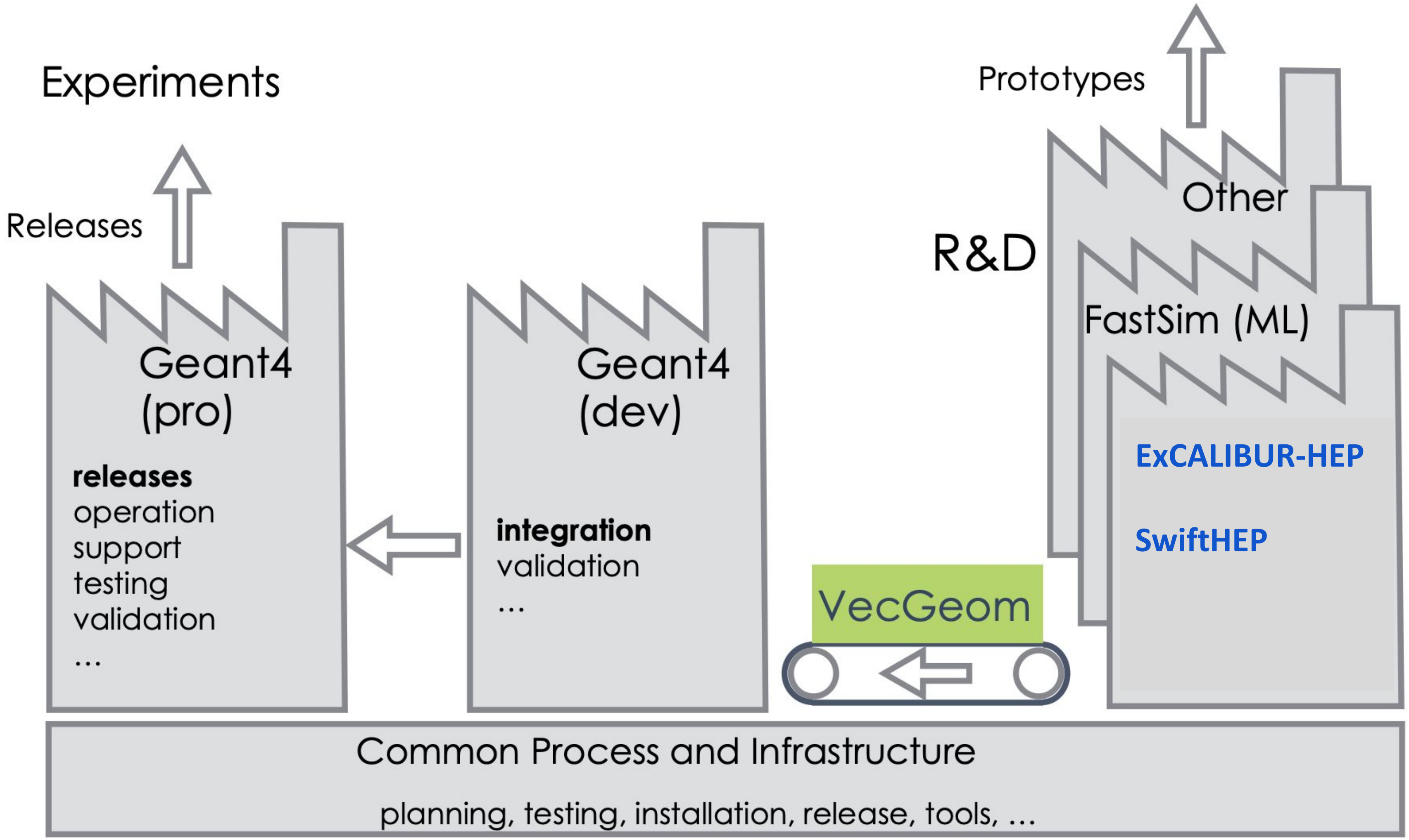
S. Menary 9 *University of Manchester*

Neutrino/Dark Matter Challenges

Overall CPU needs lower than LHC, but some specific challenges such as optical photons

Three axes to meet challenges

1. **Modernize and optimize "Full" Geant4 toolkit**
   - ○ *Inc. code architecture to help items 2 and 3*
   - ○ *E.g. Task-based parallelism released in v10.7.0*
2. **Trade precision for performance with "Fast" methods**
   - ○ *E.g. Parameterization, Machine Learning, Biasing*
   - ○ *"Full" used to fit/train/validate "Fast"*
   - ○ *Ensure seamless integration with "Full" Geant4*
3. **R&D on algorithms and data structures for radiation transport on new architectures**
   - ○ *Not a simple "port X to GPU" problem!*
   - ○ ***ExCALIBUR-HEP and SwiftHEP's prime area of work on simulation***



Evolution Strategy

Geant4 Task Force for R&D

HSF Detector Simulation WG

Simulation in ExCALIBUR-HEP

- ExCALIBUR-HEP: 15 month project since May 1st 2020 in UK funded under UKRI Excalibur High Priority Use Cases Phase 1
- **Work package on simulation, focussing on use of GPUs for EM Physics (typically highest CPU consumer)**
- Two lines of investigation:
  - *Secondary generation/management, understanding where/how to avoid divergences, Host/GPU transfer*
  - *Contribute to investigations on applicability of RT methods to charged particle transport*
- Strong overlap with Portability Work Package (oneAPI, Alpaka, etc)

Global Picture

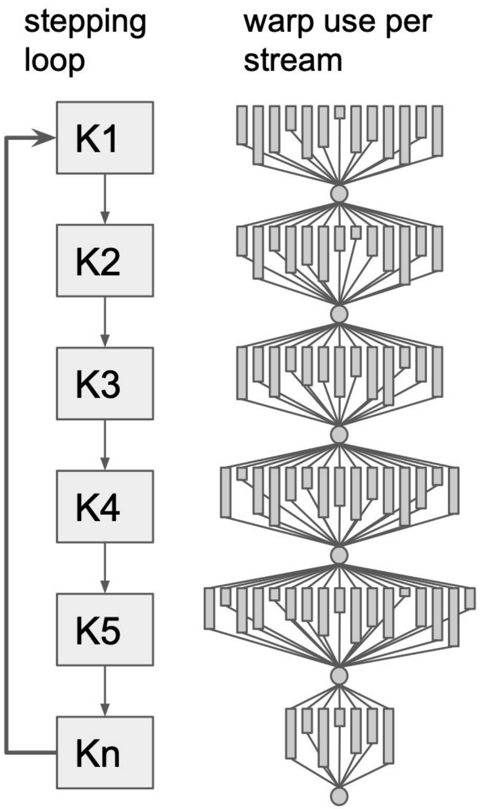
- Links and collaboration established with other R&D efforts to share ideas and avoid reinvention/deadends whilst allowing rapid prototyping and independent goals
- AdePT Project (*CERN-SFT*)
  - *https://github.com/apt-sim*
- Celeritas Project (*ECP: ORNL, FNAL, Argonne, LBL*)
  - *https://github.com/celeritas-project*
- Regular working and strategy meetings between projects
  - *Also HSF and Geant4 workshops and meetings*

AdePT

- **See [Andrei Gheata's HSF/WLCG presentation] for full overview**
- **Demonstrate realistic EM shower simulation workflow on GPU**
  - ○ *Start with a basic "Fisher-Price" like workflow demonstrator*
  - ○ *Single particle type carrying minimal state, two processes (energy loss and secondary generation), energy deposits per cell as output, dynamic track population*
  - ○ *Develop basic framework controlling a dynamic track workflow*
- **Evolve as e+/e-/γ simulator in simple calorimeter setup**
  - ○ *Magnetic field*
  - ○ *VecGeom-based transport manager as first implementation*
  - ○ *Gradually evolved physics processes to simulate EM showers*
  - ○ *Simple "hits" as output, transferred to host*
- **Maintain CPU compatibility for the entire simulation**
  - ○ *Correctness validation, performance analysis, reference baseline*

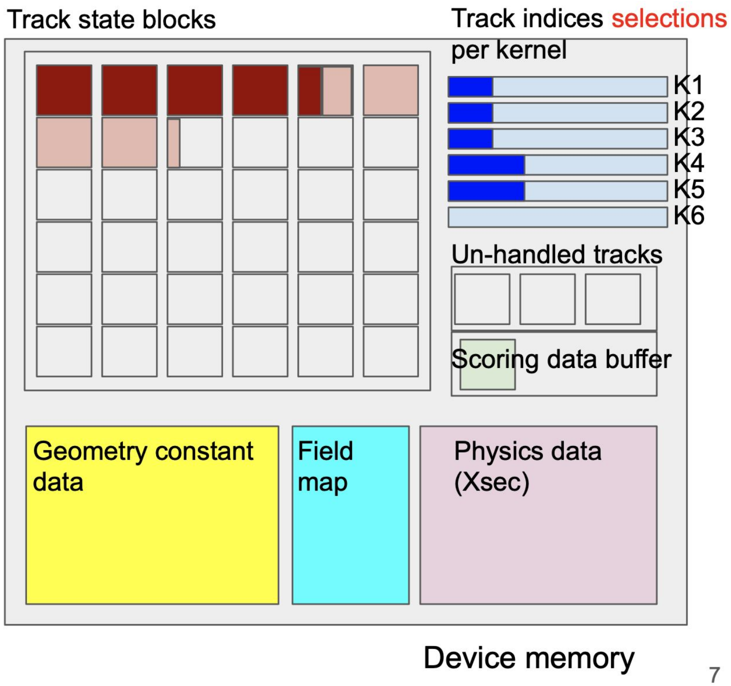
AdePT: Bootstrapping

- Basic Host/Device utilities
  - *Data structures, examples*
  - *Discussion/advice from LHCb Allen project on portability*
- Study refactor of stepping loop into sequence of kernels
  - *Move from: “list of actions to do for a track” to “list of tracks doing the same action”*
  - *Partially mitigate kernel synchronization overheads and thread divergence by overlapping stepping loop execution in multiple streams*
  - *Partially mitigate memory latency issues by coalescing accesses for reused state data*



AdePT: Compute Workflow

- Workflow organisation:
  - *Pre-allocate data on device: constant data, buffers for state (tracks) and output (hits) data*
  - *Fill tracks from host, start stepping loop*
    - *Manage dynamically selections for the "next" kernel*
    - *Manage dynamically holes produced by killed tracks*
  - *Fill output "hits"*
  - *Copy "hits" to host*



Celeritas

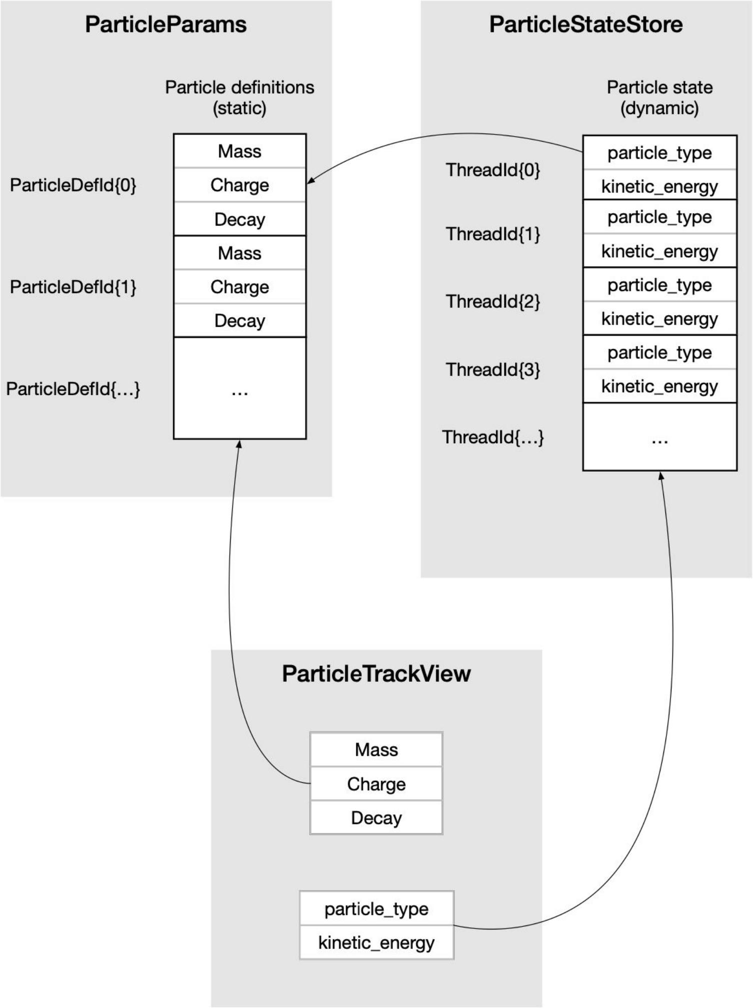
- **See Seth Johnson’s HSF/WLCG presentation for full overview**
- Deliver a GPU-accelerated particle transport application for HEP detector simulations
- Efficiently use leadership class hardware (GPUs) to increase particle tracking throughput with concurrent improvements in I/O and post- processing analysis
- Use Celeritas to offload EM physics in a standard Geant4-constructed application
- Use Celeritas as part of a broader LHC workflow for complete detector simulation
- Combinations of both approaches should be possible

Celeritas: Ongoing Work

- Infrastructure
  - *Cohesive framework (GitHub wiki)*
  - *Import GPU algorithms from Shift GPU code*
- Physics
  - *Leverage Geant4 physics implementations and documentation*
  - *Export preprocessed Geant4 data where possible*
- Geometry
  - *Interface with VecGeom CUDA library*
  - *Links established with VecGeom and NVidia*

Celeritas: CPU/GPU Data, Portability

- Kernel code is separate from memory management and memory layout
  - *Enables development, testing, and debugging on CPU*
  - *Allows isolated experimentation with data layouts to optimize GPU performance*
  - *Could enable (for example) Kokkos data management*
- Kernels are plain C++ with annotations
  - *Macros to enable CUDA device code, extensible to HIP and other performance abstraction layers*
  - *Runtime initialization code is host-only*



Celeritas: “Acceleritas” Mini-app

- Collaboration between Celeritas and ExCALIBUR-HEP
  - *Develop a Geant4 application using the new tasking system to demonstrate offload of tracks to a GPU for further processing*
  - *https://github.com/celeritas-project/acceleritas*
- Similar to Opticks workflow for optical photons
- Provides a testbed for performance and physics validation in a hybrid system
- AdePT will also use it for cross-project comparisons

ExCALIBUR-HEP engagement activities

- Apart from AdePT/Celeritas links, other engagements and opportunities have been enabled
  - *Davide Costanzo will take over from John Allison as STFC representative on Geant4 Oversight Board*
  - *Ben Morgan co-convener in HSF Detector Simulation WG for 2021*
  - *Building contacts with Lattice QCD community, including joint ExCALIBUR-HEP/ExaLat Data Parallel School (See tomorrow's talks)*
- ExCALIBUR cross-cutting themes activity
  - *Proposal on "computational geometrical modelling and simulation of particle and nuclear radiation"*
  - *Re-establish links with Medical, Bio-, Space, Nuclear, Industry fields using Geant4 and other codes in this area*
  - *Facilitate exchange of ideas, including training*
  - *Develop common applications to run on modern HPCs and investing in skilled RSEs to develop these applications*

Simulation in SwiftHEP

1. **Build on R&D from ExCALIBUR-HEP on algorithms for EM Physics on GPU**
   - *Continue and strengthen collaboration with AdePT and Celeritas projects*
   - *Extend "Acceleritas" example to fully demonstrate, profile and validate hybrid CPU/GPU Geant4 application*
2. **Contribute to development of example application using Geant4+Opticks for hybrid CPU/GPU optical photon simulation**
   - *In collaboration with FNAL and Geant4 R&D Task Force*
   - *Build UK links with, and expertise in, Opticks*

Summary

- Much activity on detector simulation since ECHEP Edinburgh meeting, despite the challenges of 2020!
- ExCALIBUR-HEP has built strong links with global R&D efforts on detector simulation on GPUs
  - *Working towards demonstrator applications for EM physics in 2021*
  - *Hope to build broader UK community: cross-cutting activity proposal*
- SwiftHEP will continue these efforts and extend them to cover hybrid CPU/GPU simulations of optical photons
- **Encourage all UK users/developers of detector simulation to contribute ideas, experience, developments to community:**
  - *HSF Simulation Working Group Meetings*
  - *Geant4 Task Force for R&D Meetings*