# Optimizing the ATLAS Geant4 Detector Simulation

Advanced Computing and Analysis Techniques in Physics Research ACAT 2024


Mustafa Schmidt on behalf of the ATLAS collaboration


March 13, 2024


University of Wuppertal

# Introduction

- ATLAS experiment at LHC and physics analysis of taken data heavily depend on simulated event samples produced with full Geant4 detector simulations
- Monte Carlo (MC) simulation based on Geant4 major consumer of computing resources during 2018 data-taking year
- Anticipated to remain dominating resource users in High-Lumi LHC era
- Performing simulations with Geant4 and Athena (ATLAS Offline software)
- Many proposed and already validated simulation optimizations available for improving CPU and memory consumption without sacrificing physics accuracy

**Topics covered within this talk:**

Implemented Optimizations

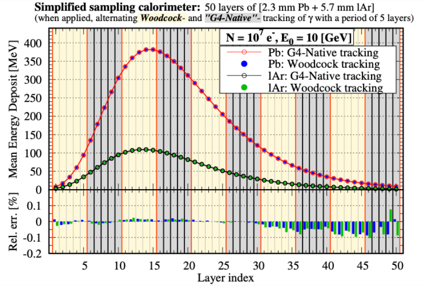
Performance Studies

Ongoing Tasks

# Implemented Optimizations

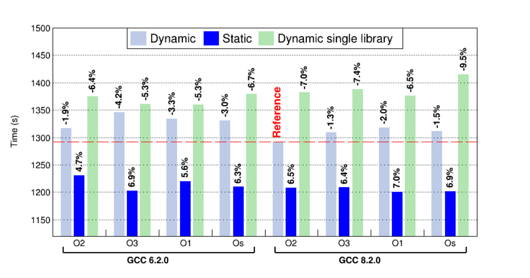
# Woodcock Tracking for Gammas

- **Goal:** Tracking optimization technique for neutral particles (especially photons/gammas) in highly segmented detectors → simulation steps in detector volumes limited by geometric boundaries (not physics)
- **Advancements:**
  - Performs tracking in a unified geometry made by material with highest macroscopic cross-section (Pb) related to $Z$ and number of atoms per unit volume (physics does not change)
  - Interaction then occurs with probability equal to ratio of CR of true material and Pb
  - Implemented as special wrapper process for gammas in Geant4




- **Result:**
  - Reducing ≈ 50% of steps in EMEC region
  - Overall measured speedup in Athena: 17.5%

# Static Linking



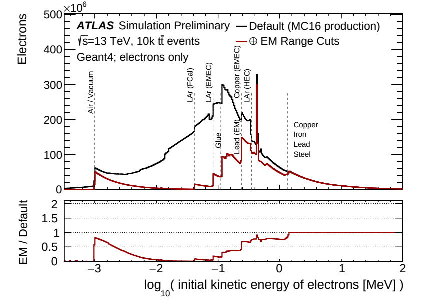

Reference: HepExMT Benchmark

**Results:** Speed-up of $\approx 5 - 7\%$ depending on compiler version

- **Goal:** Use Geant4 as static library to avoid lookup table delays
- **Approach:** Define BigSimulation shared library by grouping all libraries from Athena packages that use Geant4
- **Advantage:** States are identical $\rightarrow$ No physics validation required for this task

# EM Range Cuts

- **Motivation:** Different energy thresholds in Geant4 (only production threshold for secondary particles relevant)
- **Approach:**
  - Secondary production threshold for ionization ($e^-$) and bremsstrahlung ($\gamma$) at cross-section level
  - Setting secondary production cut in length units: minimum range of secondary $e^-$ and minimum absorption length of gammas
  - Below threshold: energy deposited at the end of production step

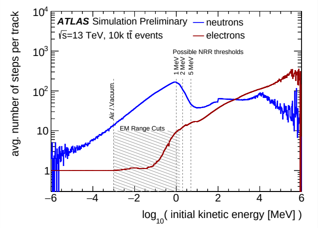


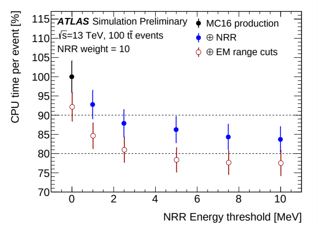

Reference: ATLAS Geant4 Performance Optimization Plots

- **Results:** CPU speed-up of ≈ 8% and significant reduction of simulated low-energy electrons (≈ 60%)

Reference: Geant4 performance optimization in the ATLAS experiment

# Russian Roulette



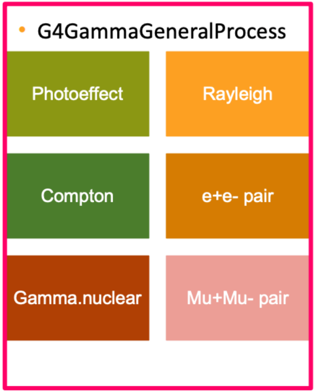

Reference: ATLAS Geant4 Performance Optimization Plots

- **Motivation:** Neutrons & photons take majority of CPU time → Barrel and endcap EM calorimeters most resource-intensive
- **Approach:** Photon/Neutron Russian Roulette (PRR/NRR): randomly discard particles below energy threshold and weight the energy deposits of remaining particles accordingly
- **Results:** NRR performance increased by 10% speed up with 2 MeV threshold for neutrons

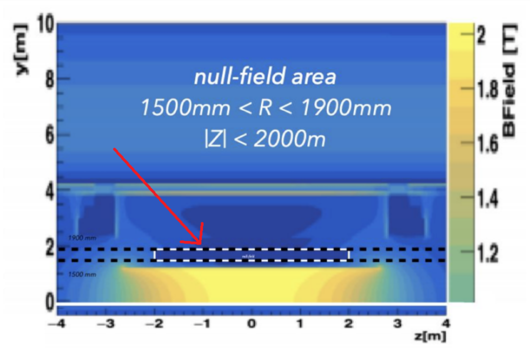
Reference: Geant4 performance optimization in the ATLAS experiment

# G4GammaGeneralProcesss

- **Motivation:** Many different gamma processes increase computation time
- **Approach:** Use only one collective physics process for photons $\rightarrow$ reducing number of instructions/calculations on geometry boundary crossings
- **Results:** Improvement about 3% CPU speed-up



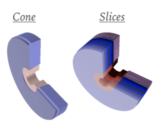
# Magnetic Field Optimizations



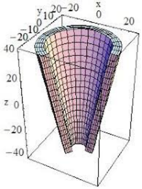

- **Motivation:** Tracking particles in magnetic field use many resources
- **Approach:** Switching off magnetic field in LAr region without affecting shower shape (not used for muons or high energetic $e^+/e^-$
- **Results:** Speed-up of around 3%
- **Further Progress:** Possibility to extend approach to other detector regions as well

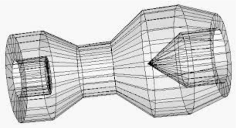
# Geometry Optimizations

- **EMEC Custom Solid**
- **Motivation:** Described with G4Polycone
- **Approach:**
  - Replacing G4Polycone with G4ShiftedCone – outer wheel divided into two conical-shaped sections
  - Slices new wheel — divided into many thick slices along Z axis
- **Result:** Speed-up of 5–6%




- **VecGeom Integration**
- **Approach:**
  - Optimised implementation of geometrical shapes → taking advantage of explicit and implicit vectorisation
  - Only replacement of polycons, cones, tubes relevant
- **Result:** Speed-up of 2–7%

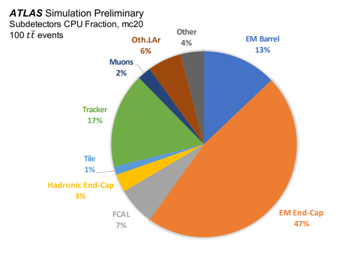


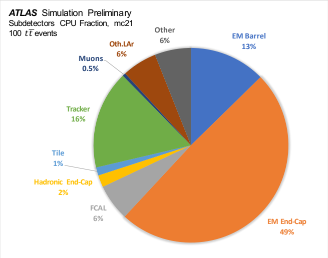


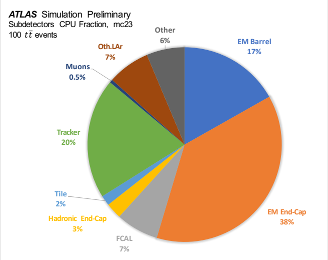
# Performance Studies

# Computing Fractions

- **CPU fraction:** CPU time distribution among different subdetectors through mc20, mc21 and mc23 campaigns
- Largest CPU fraction in EMEC followed by tracker and Barrel EMC
- Tile detector and muons showing smallest impact on performance

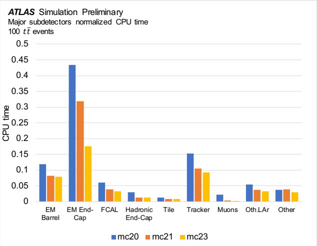


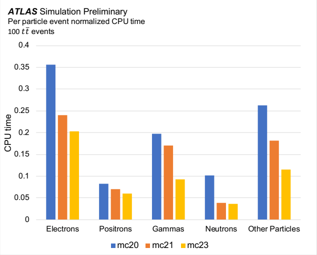



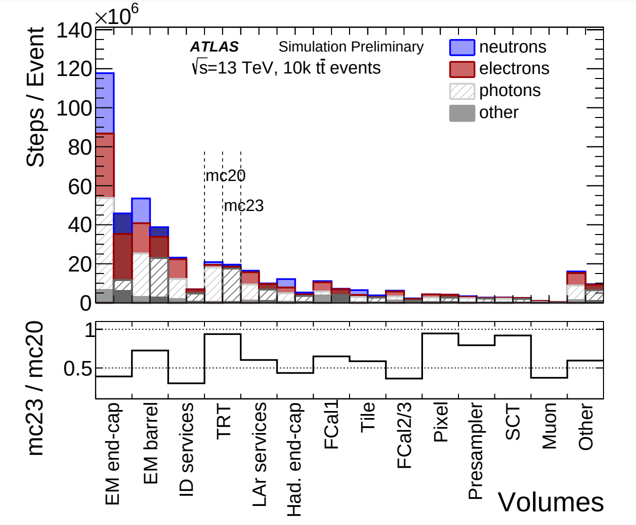
# CPU Time

- Time spent per event simulating 100 $t\bar{t}$ events (important benchmark channel for simulations):
  - for each of the major subdetectors (top)
  - for each of the each particle type (bottom)
- Color indicating different configurations:
  - Left bar: Run 2 configuration (mc20)
  - Middle bar: Previous Run 3 configuratin (mc21)
  - Rigth bar: run 3 configuration with latest optimizations (mc23)
- In total improvement of 50% CPU time reached compared to Run 2 configuration



# Stepping Plots



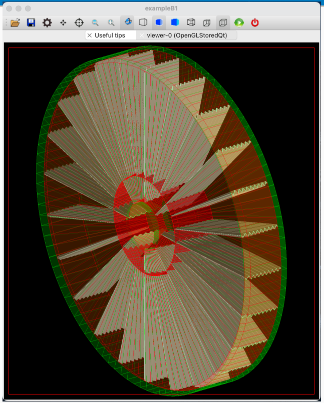

Latest improvements include:

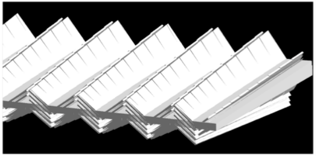
- EM range cuts
- Photon Russian Roulette
- Neutron Russian Roulette
- Geant4 built as a single library
- G4GammaGeneralProcess
- Woodcock tracking in EM end-cap
- Magnetic field optimization
- Optimized EM end-cap
- Using VecGeom

# Ongoing Tasks

# EMEC Geometry Optimization

- **Motivation:** Current implementation of Endcap EM Calorimeter (EMEC) with custom solids based on G4Polycone
- **Approach:**
  - Define geometry of EMEC with Standard G4 shapes (G4Trap) to speed up simulation as well as allow usage in other architectur (GPU)
  - Additional slices in $z$-direction result in further improvements
  - Ongoing comparison between different geometry options and configurations
- **Results:** Solid improvement in CPU time of $\approx 19\%$ (physics validation to be done)



# Celeritas/AdePT

- R&D projects aiming to accelerate simulation of electromagnetic showers on GPUs
- Celeritas: Collaboration between several US national labs (Poster https://indico.jlab.org/event/459/contributions/11818/)
- AdePT: mainly CERN-SFT initiative
- Using VecGeom (vectorized geometry) library to handle complex detector geometries
- Implementing data structures & workflows for track-parallel stepping on GPUs
- Using G4HepEM (CPU/GPU implementation of Geant4) for EM physics models
- Continuously validating results against Geant4
- Demonstrate significant speedup over Geant4 on CPU with simple geometries

# Advanced Compiler Optimization

- **Goal**: Speed up simulation with link time and profile-guided optimization
- Advanced compiler optimizations can lead to non-negligible speed-up factors
- CMS report $\approx$ 10% speed in their software
- Two approaches for reducing application runtime relying on the compiler: smarter usage of compiler $\rightarrow$ more throughput $\rightarrow$ efficient use of computing resources

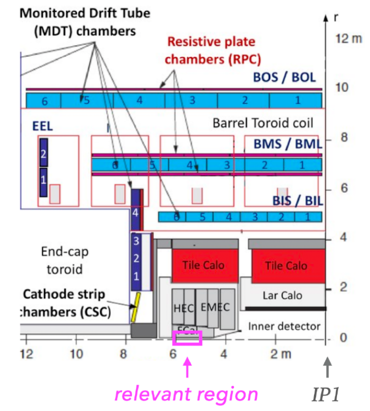
**LTO**: Compilation units with metadata

- Consults to optimize when building shared objects
- Expands scope of inter-procedural optimizations to encompass all objects visible at link time
- Preliminary benchmark: 3–4% speed-up (physics validation planned)

**PGO**: Optimize full executable

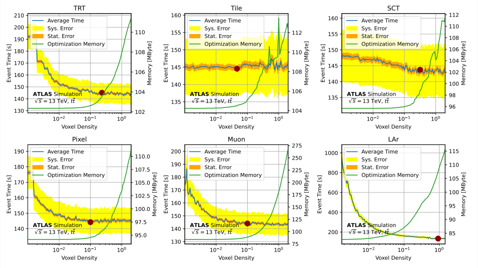
- Build instrumented binaries, producing profile for application, rebuild from sources and profile
- Inlining, block ordering, register allocation, conditional branch optimization, virtual call speculations, ...
- Preliminary benchmark: 3–5% speed-up (physics validation done but following up)

- **Goal**: Kill primary particles generating secondaries close to the beam pipe at 5 – 6 m
  - Huge number of secondaries being produced 5–6 m away from IP with small $r$ (close to beam pipe)
  - Many of these secondaries will not cause any energy deposits in the calorimeters or a muon hit
  - Primary particles causing interactions could be dropped directly
- **Approach**:
  1. Generate a large sample of single particles with $4.5 < |\eta| < 6$ with different energies
  2. Map out $\eta$ and $E$ combinations with relevant signals
  3. Drop others directly with new particle killer



# Voxel Density Optimization

- **Goal**: Find optimal values of voxel density for optimization of CPU time and memory consumption
- **Advancements:**
  - Size/Granularity of the voxels can be tuned
  - Voxel density - member variable in Geant4 logical volume class
  - Improvement in memory consumption for geometry optimizations




- **Results:**
  - Small improvement in CPU time and memory consumption seems possible
  - Physics validation completed
  - Follow-up study is ongoing (increase of CPU observed)

# Conclusion & Outlook

**Conclusion:**

- Significant optimizations implemented in ATLAS Geant4 simulation improved CPU time and memory efficiency without sacrificing physics accuracy
- Total reduction of $> 50\%$ possible compare to run 2 samples
- Key advancements include Woodcock tracking, EM range cuts, optimized geometry, and new gamma processes
- Excellent collaboration with Geant4 team for implementing optimizations

**Outlook:**

- Ongoing improvements focus on leveraging modern computing architectures
- Innovative approaches continue to refine simulation precision and resource usage
- Continuous validation against Geant4 ensures high standards of physics accuracy

# Thank you very much for your attention!

# Backup Slides

# In-Field Parameter Tuning

- **Motivation:** Previous tuning performed by CMS (full simulation for Run 3)
- **Goal:** Find the optimal values of the in-field tracking parameters for physics performance and CPU savings
- **Approach:**
  - Lists of tuning parameters + descriptions for tracking in a magnetic field
    - DeltaIntersection: accuracy of intersection with boundary volume
    - DeltaOneStep: accuracy for endpoint of 'ordinary' integration step
    - DeltaChord: approximation of curve with linear sections
    - MaxStep: maximum step length
  - Cross-correlation between different parameters → global optimization required
  - Tuning to be done for different detector regions and various particle energies
- **Current Progress:** Implementation in simplified framework FullSimLight completed → next step testing influence in full ATLAS simulations with Athena

# Energy Loss Fluctuation Studies

- **Goal:** Evaluate performance gain from disabling energy loss fluctuation in Geant4 simulations
- **Study Focus:**
  - Determine the computational benefits of switching off the energy loss fluctuation feature in Geant4
  - Assess potential impacts on physics accuracy
- **Approach:**
  - Use dedicated Geant4 command to disable energy loss fluctuations
  - Initial tests to be conducted in FullSimLight followed by full ATLAS simulations in Athena
- **Current Progress:** Performance gains are being investigated with physics validation to follow for assessing impact