Computing Performance Results and Issues: ATLAS

28th Geant4 Collaboration Meeting

Hokkaido University, *Sapporo*

25-29 September 2023

Marilena Bandieramonte (University of Pittsburgh)

Geant4 Simulation in ATLAS

- For RUN3 Monte Carlo production ATLAS uses:
  - Geant4 10.6.patch03.atlas07
    - It includes:
      - `G4AtlasRK4` stepper
      - `G4MagInt_Driver` patch to reduce differences w.r.t. `G4IntegrationDriver`
      - `G4GammaGeneralProcess` patch from Geant4-10.7.2
      - Woodcock Tracking
- We are working on a (2) new patch(es):
  - Geant4 10.6.patch03.atlas(08)09
    - c++20, gcc13 support
    - Make Woodcock tracking more robust against corner cases
    - Fix for some rare crashes in hadronic physics (`G4AntiNeutron` by default unstable)

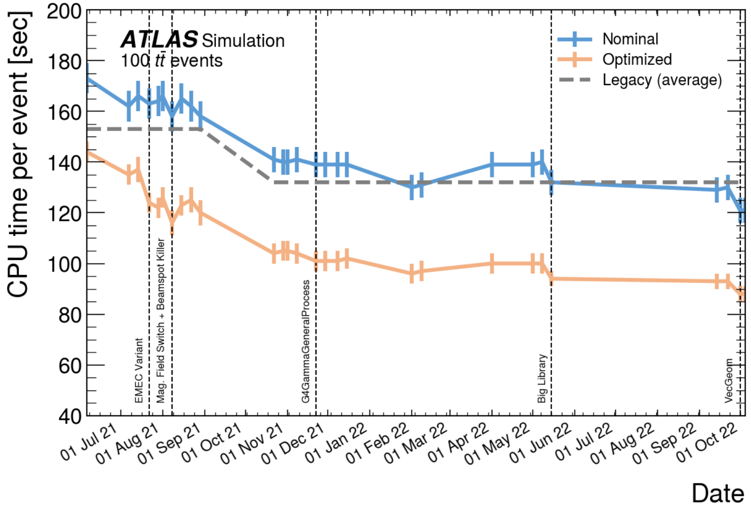
Geant4 Simulation Optimizations for Run3

New Improvements that are used in production in the Run3 MC campaigns*

*since last year presentation, details in backup slides

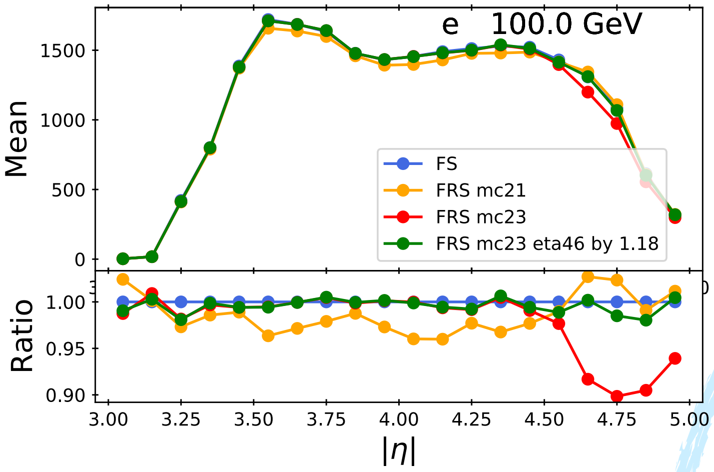
Full Geant4 Simulation for Run3

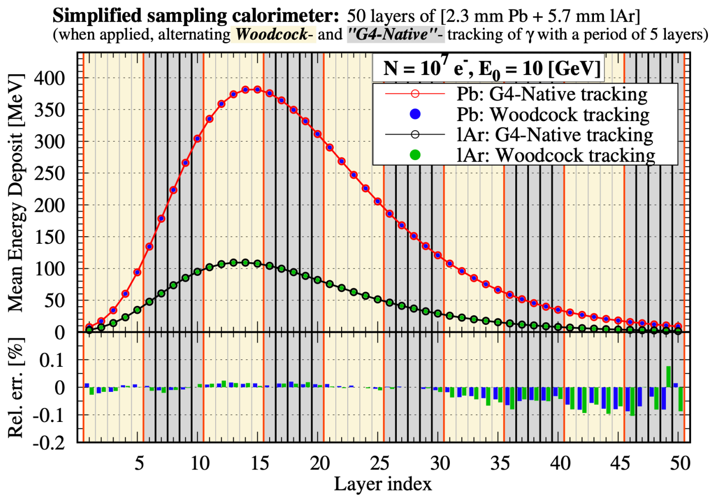
- MC campaign for start of Run3 (mc21) already **>30% faster** than mc16.



- **EM range cuts**
- **Neutron** and **Photon Russian Roulette**
- New **EMEC Slices** variant
- **BigLibrary**
- **G4GammaGeneralProcess**
- **Magnetic Field** tailored **switch-off**

- In addition **MC campaign for Run3 (mc23)** features:
  - **VecGeom**: new vectorized geometry library, we replace only *polycons, cones and tubes (2-3%)*
  - **Frozen Showers** update (tuned to the new Geant4 version)
  - **Woodcock tracking** [1]: smart tracking for highly segmented detectors (~20%)

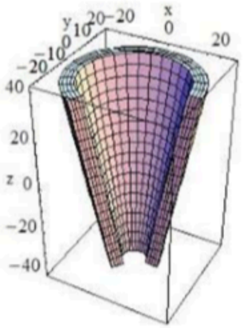




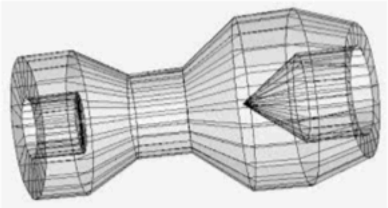
VecGeom: vectorized Geometry

New and optimised implementation of geometrical shapes, designed to take advantage of explicit and implicit vectorisation:

- we replace only *polycons, cones and tubes*



2-3%



measured on 500 ttbar events in the Atlas software framework (Athena)

VecGeom alone

BNL Cluster, 1000 Jobs 100 tt-bar events/job	**Walltime,s**	**sigma**	**Speedup CPU time**	**Job ID**
Athena 22.0.47 (baseline)	280	22.8	-	27857092
Athena 22.0.63 with VecGeom (no BigLibrary)	263	23	-6.1%	31379861
Athena 22.0.63 with VecGeom (no BigLibrary) +Run3Opt	195	15.9	-30.4%	31379871

VecGeom on top of the Big Library

BNL Cluster, 1000 Jobs 100 tt-bar events/job	**Walltime,s**	**sigma**	**Speedup CPU time**	**Job ID**
Athena 22.0.47 (baseline)	280	22.8	-	27857092
Athena 22.0.67 "out the box" (BigLibrary only)	270	21.3	-3.6%	31366397
Athena 22.0.67 "out the box" (BigLibrary only) +Run3Opt	189	15.7	-32.5%	31366406
Athena 23.0.7 "out the box" (BigLibrary + VecGeom)	252	21.6	-10%	31352919
Athena 23.0.7 "out the box" (BigLibrary + VecGeom)+Run3Opt	**185**	**15.8**	**-33.9%**	**31352932**

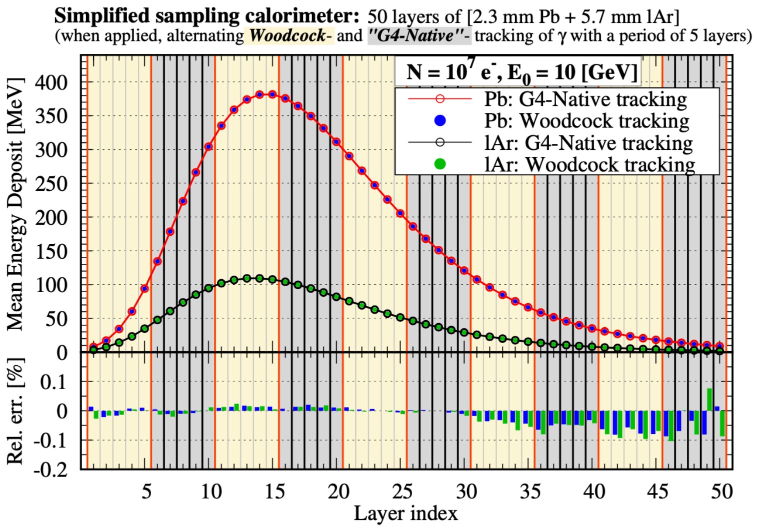
Ben Morgan

Woodcock tracking

- Reduce the **number of steps for gammas** in highly segmented detectors (e.g. EMEC), where the geometric boundaries limit the step, rather than the physics
- Perform the transport in a geometry with no-boundaries made of the densest material (i.e. Pb)
- The interaction probability is proportional to the cross-section ratio between the real material and the “fake" one
- Integrated in FullSimLight *and in **Athena***
  - Shows a ~**10-15%** performance gain when using *Woodcock-tracking* of gamma photons in the EMEC detector region on top of the gain already provided by the *Gamma-general* process
  - **Full physics validation successful and in production for Run3!**

		FTFP_BERT_ATL	_WDCK-*GammaGeneral*	_WDCK *Woodcock*(EMEC)
#secondary	γ	3.054e+05	3.06e+05	3.062e+05
	e^-	6.240e+05	6.204e+05	6.201e+05
	e^+	2.186e+04	2.19e+04	2.193e+04
#steps	charged	3.548e+06	3.548e+06	3.550e+06
	neutral	**8.501e+06**	**8.464e+06**	**4.215e+06**

FTFP_BERT_ATL	_WDCK-*GammaGeneral*	_WDCK *Woodcock*(EMEC)
base line	[2 - 4] %	[10 - 14] %



M. Novak, J. Apostolakis

Ongoing effort

Targeting Run3 MC reprocessing, or future Run3 MC campaigns

LTO/PGO and auto-FDO

- **PGO** (Profile Guided Optimization) and **FDO** (Feedback-Directed Optimization) are both based on the instrumentation of the binary

- **Auto-FDO** (Automatic Feedback-Directed Optimization) is a system to simplify the deployment of feedback-directed optimization (FDO).

    - The system works by sampling hardware performance monitors on production machines and using those profiles to guide optimization

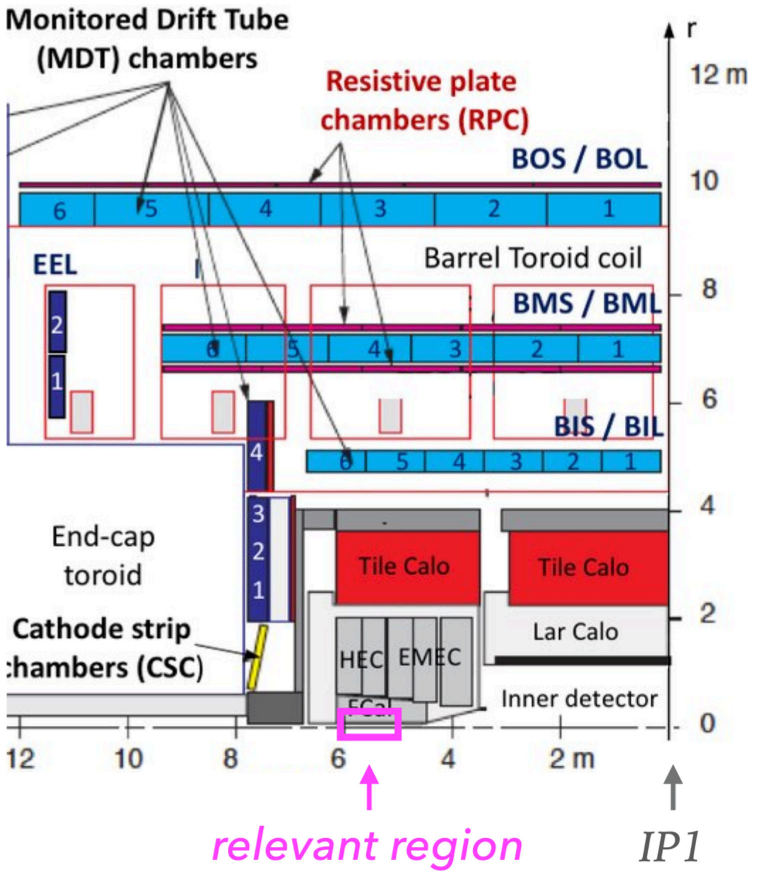
- First test running FullSimLight with LTO enabled:
    - full static linking plus LTO yields a significant speed up
    - dynamic loading plus LTO doesn’t achieve the same speedup, but a **~4-5%** decrease in runtime is observable

B. *Morgan* and C. *Marcon*

High-η particle rejection (ISF particle killer)

Goal: Kill primary particles generating secondaries close to the beam-pipe at 5-6 m

- Many particles in the collision are at high |η| (no ID hits) with little energy compared to the calorimeter noise.
- The goal is to kill these particles as we expect that their secondaries will never cause any energy in the calorimeters or a muon hit
- Approach:
  - generate a large sample of single particles with $4{,}5 < |\eta| < 6$ and different energies
  - map out which eta/energy combinations can produce a relevant signal
  - drop the rest directly with a new ISF (Integrated Simulation Framework) particle filter.
- Technical studies on how to classify particles as significant or not are on-going.



Dongwon Kim

GPU-friendly EMEC implementation

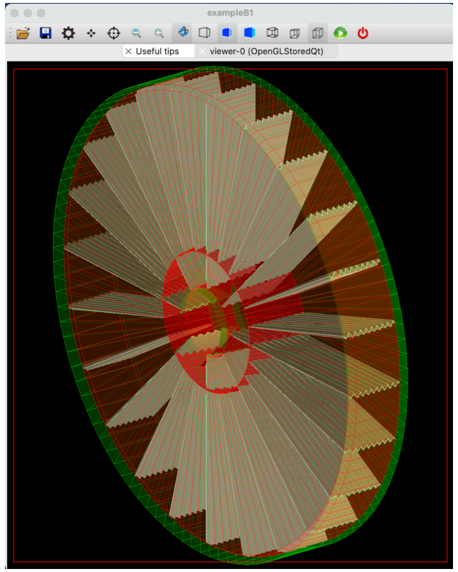
Description of the EMEC with Geant4/VecGeom standard shapes:

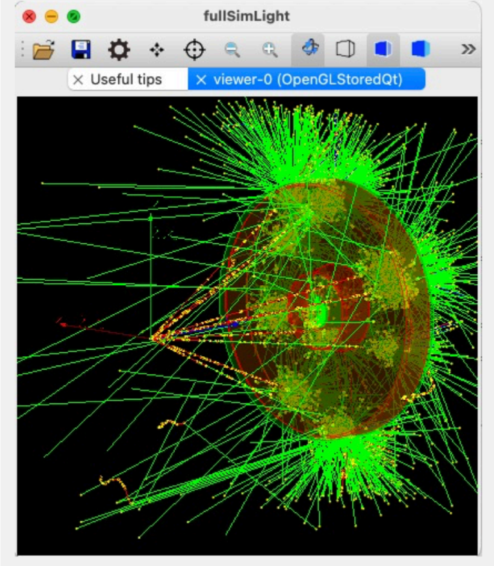
Advantages:

- Possible speedup in VecGeom on CPU making use of internal vectorisation
- Possibility for the ATLAS geometry to be standard and GPU-friendly

STATUS:

- Implemented by using `G4GenericTrap`
- The CPU consumption of various EMEC implementations was measured using FullSimLight
- Preliminary benchmarks showed that a speedup of ~2 is achievable!
  - This work lead to the discovery of a bug in the calculation of the Safety distance of the `G4GenericTrap` (work is ongoing to fix it)
- Integration in Athena to be started soon and physics validation will follow





A. Vishwakarma, E. Tcherniaev

Using Adept/Celeritas in FullSimLight/TileCal

- Initial work earlier this year by Julien/Seth to integrate **Celeritas** offload/transport in the Tile calorimeter test beam simulation validation application:
  - https://github.com/celeritas-project/atlas-tilecal-integration
  - *From:* https://github.com/lopezzot/ATLTileCalTB
  - Now also runnable through FullSimLight as a plugin, albeit currently built using a fork of GeoModel:
    - https://github.com/celeritas-project/GeoModel
- **AdePT** example of the Tile calorimeter test beam simulation implemented by Davide:
  - https://indico.cern.ch/event/1215829/contributions/5306569/
  - Now beginning work to implement this as a FullSimLight plugin as has been done for Celeritas

M. Bandieramonte, B. Morgan, Davide Costanzo, A. Dell'Acqua, J. Esseiva, S. Johnson

G4HepEM library integration

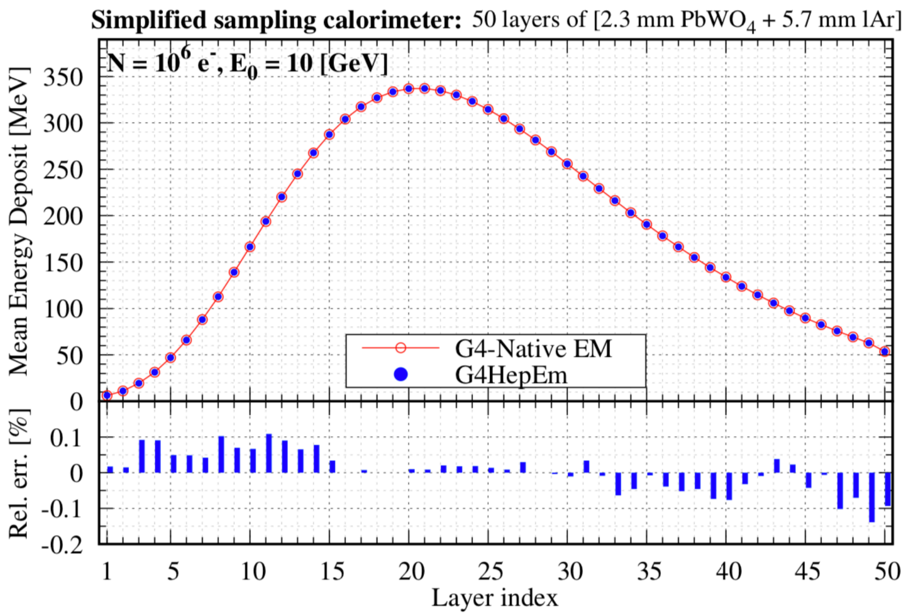
G4HepEM library is a new compact Geant4 EM library, from *Mihaly Novak, Jonas Hanfeld, Benjamin Morgan*

- Optimized to be used for HEP electromagnetic showers development and transport
    - more compact and GPU-friendly
    - It provides significant speedup w.r.t the general Geant4 EM library.

Integrated in *FullSimLight* first and first tests in *Athena*

	Physics List	Specialised Tracking	difference
G4NativeEm	2889 s	2747 s	-4.9 %
G4HepEm	2847 s	2660 s	-6.6 %
difference	-1.5 %	-3.2 %	-7.9 %

Reference

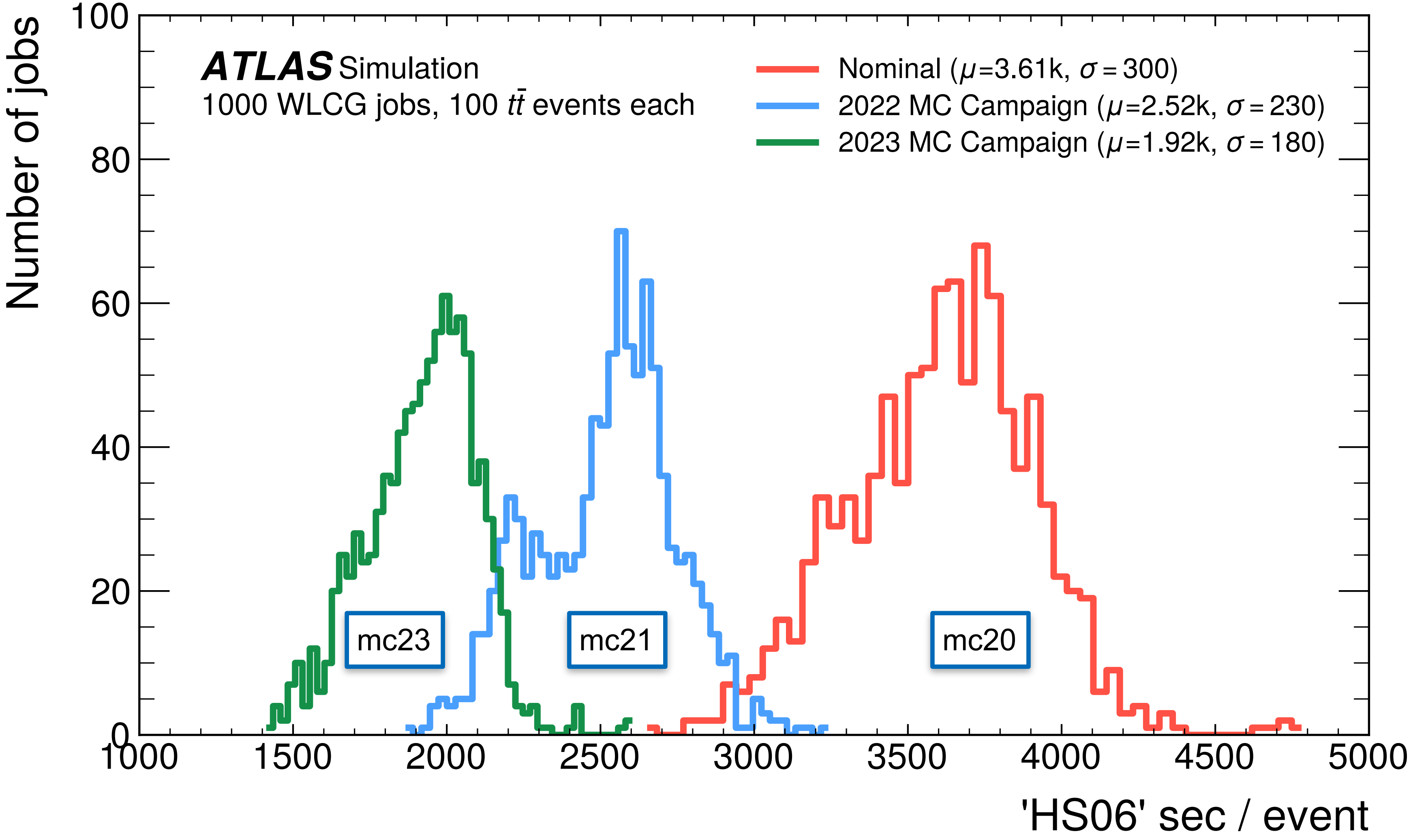


1 ‰ agreement between the native Geant4 and G4HepEM Simulation

B. Wynne, M. Novak, J. Apostolakis

Benchmarks

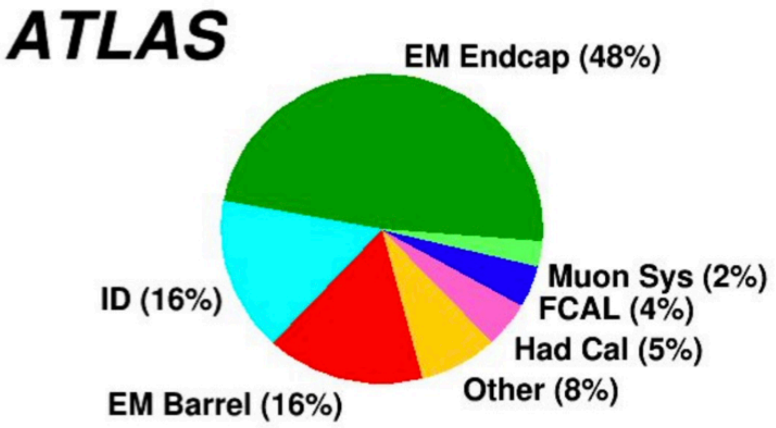
Performances of ATLAS Simulation



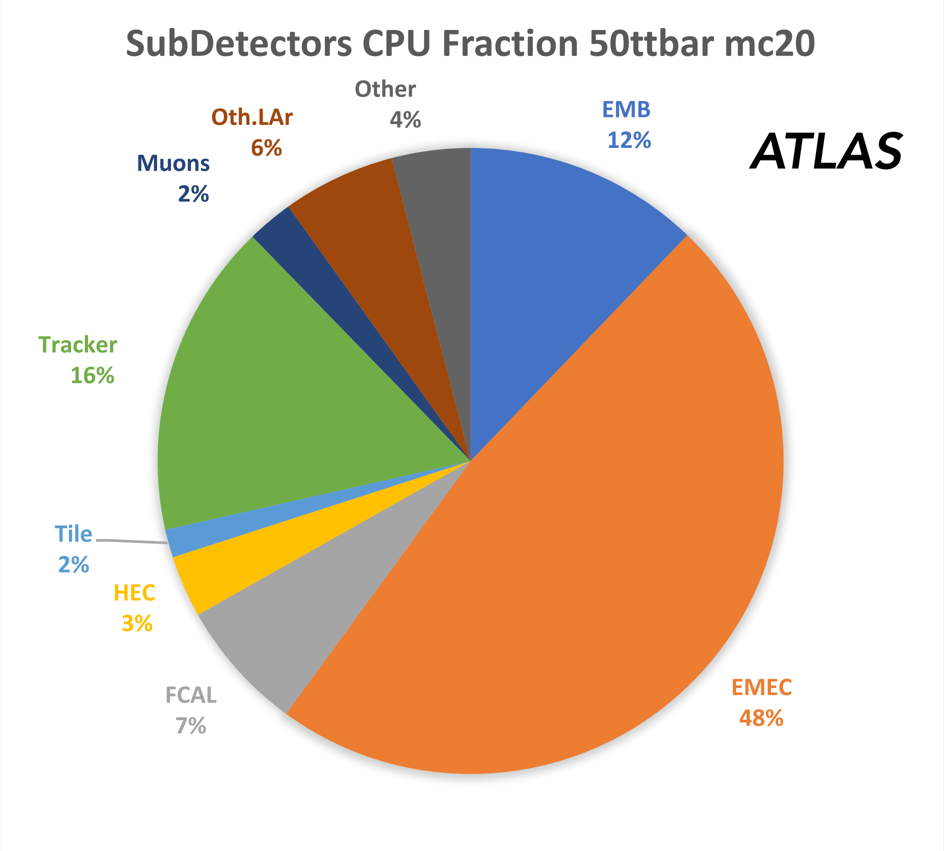
Run3 simulation ~2 times faster than the Run2 one!

Performances of ATLAS Simulation

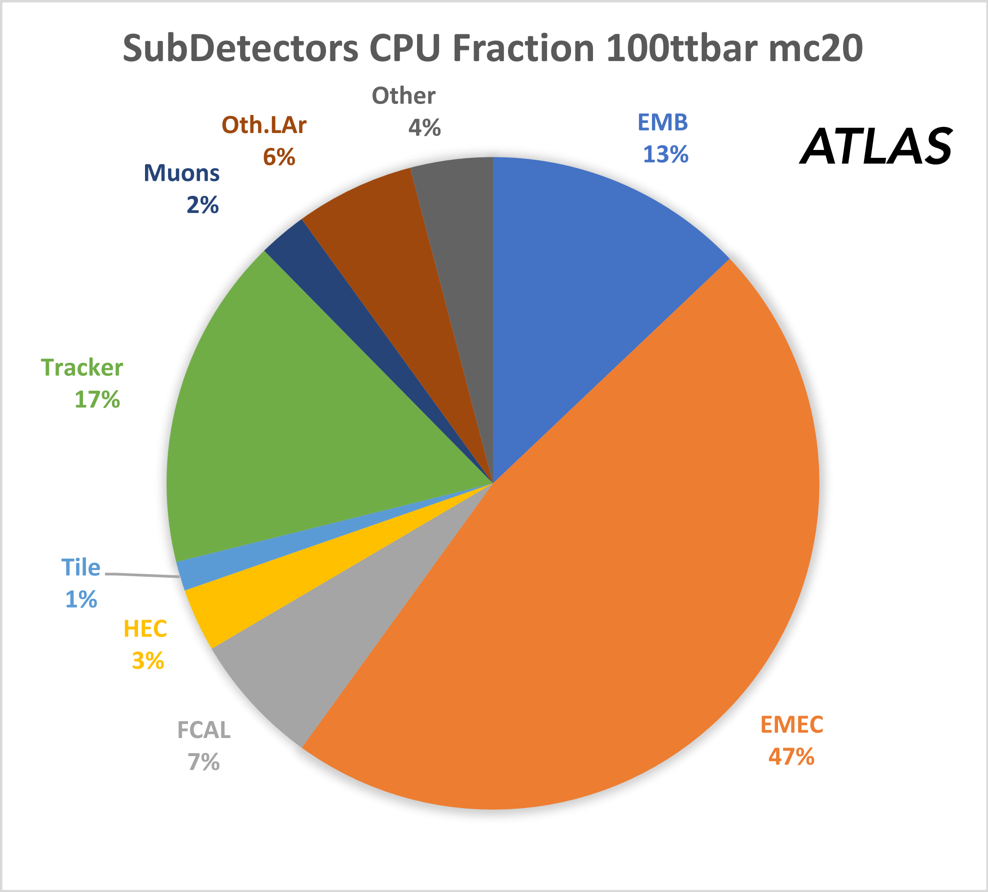
- The ATLAS **EMEC** (EM EndCap calorimeter), is the dominant contribution to ATLAS Full Simulation
- Following are the **Inner Detector** (Tracker) and the **EM Barrel**

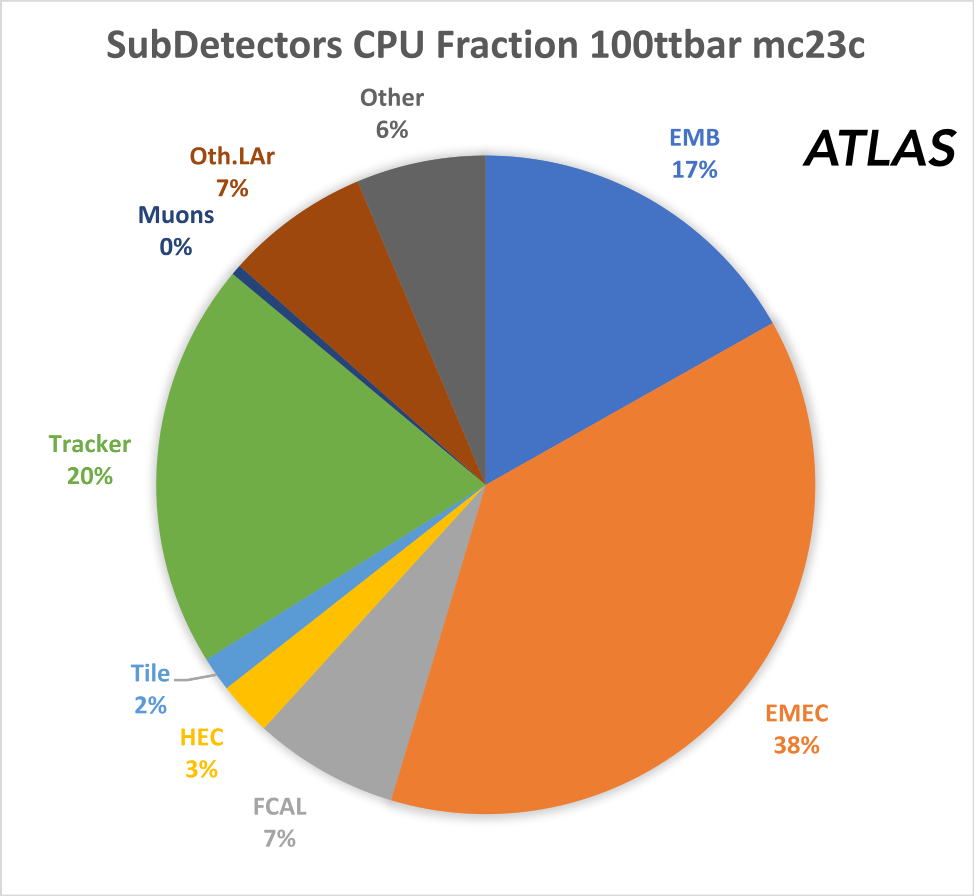


Subdetector CPU fraction for 50 ttbar events
MC16 Candidate Release



Performances of ATLAS Simulation

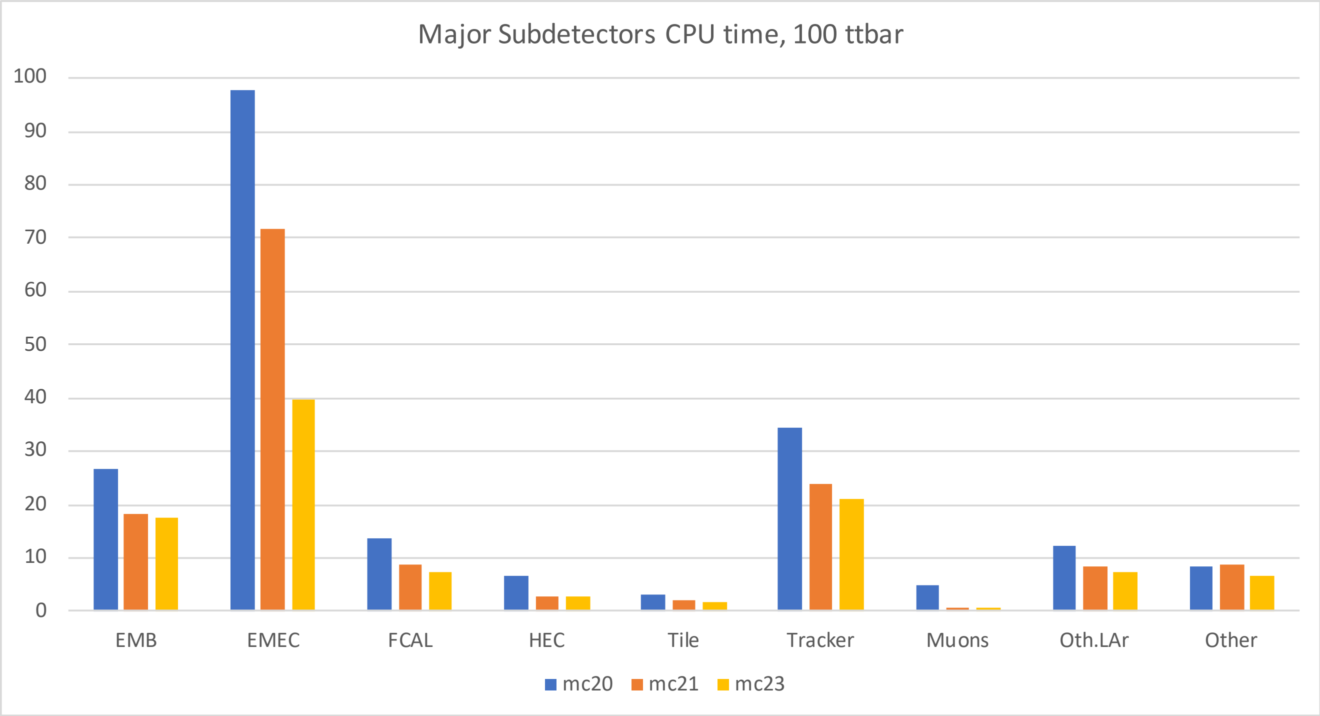




- In mc23c we have a clear reduction of the fraction of time spent in the EMEC (from 47% to 38%), but it is still the dominant contribution
- Following are still the Inner detector (Tracker) and the EM barrel

Performances of ATLAS Simulation

- Looking from a different perspective (time spent per sub-detector and per particle type) gives us more insights:

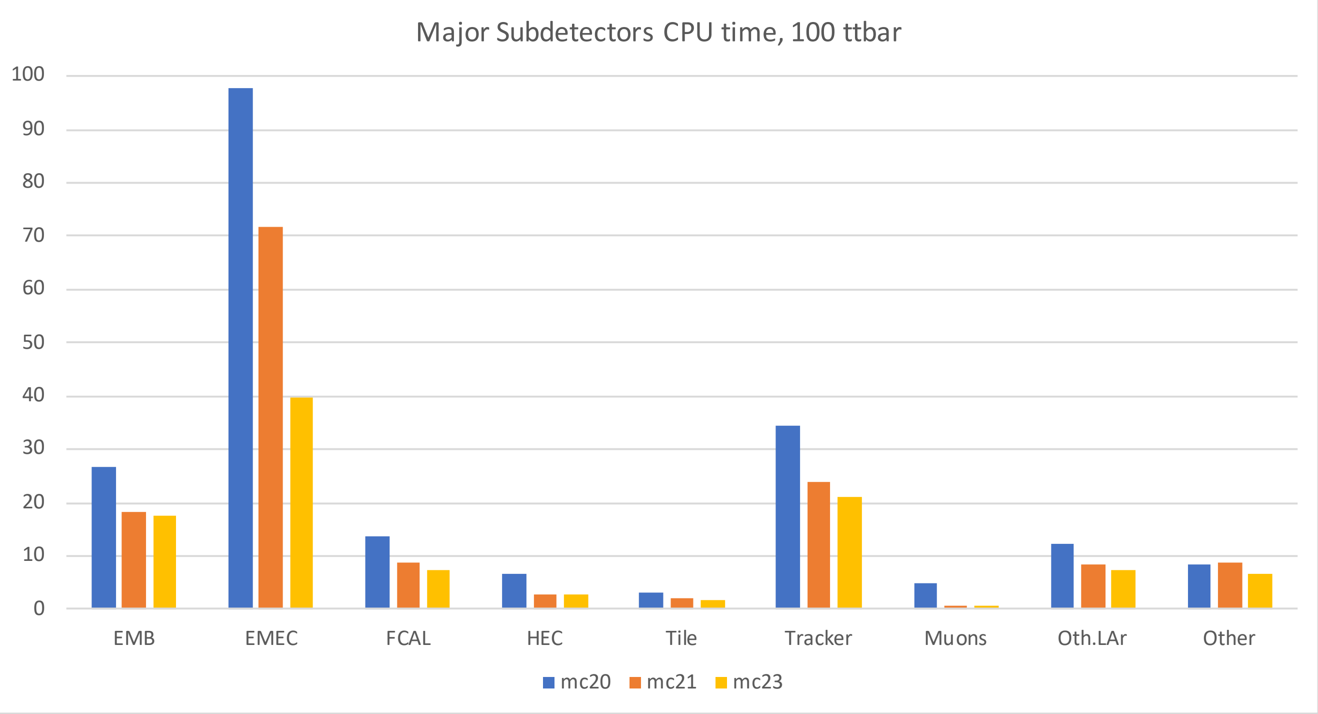


Improvements over different campaigns

- **mc20, MC reprocessing campaign for Run2:**
  - TRT range cut (sets the range cut for e-/e+ in the TRT volumes filled with Xenon)
  - Frozen Showers
- **mc21, MC campaign for start of Run3: (in addition to mc20)**
  - EM range cuts: expect to reduce the n. of low energy electrons
  - PRR: expect to reduce the n. of tracked photons
  - NRR: expect to reduce the n. of tracked neutrons
  - EMEC slices: improve the CPU time in the EMEC
  - Big Library: improve overall the Full Simulation Time
  - MagField Off: improve transport in magField
  - G4GammaGeneralProcess: improve time spent to transport gammas
  - BeamPipe killer: kills particles at eta>5.5, that do not make to the detector
- **mc23c (in addition to the mc21), MC campaign for Run3:**
  - Woodcock Tracking: improve timing of gammas in the EMEC
  - VecGeom: Improve CPU time for detectors that use Polycons, Tubes and Cones

Performances of ATLAS Simulation

- Looking from a different perspective (time spent per sub-detector and per particle type) gives us more insights:

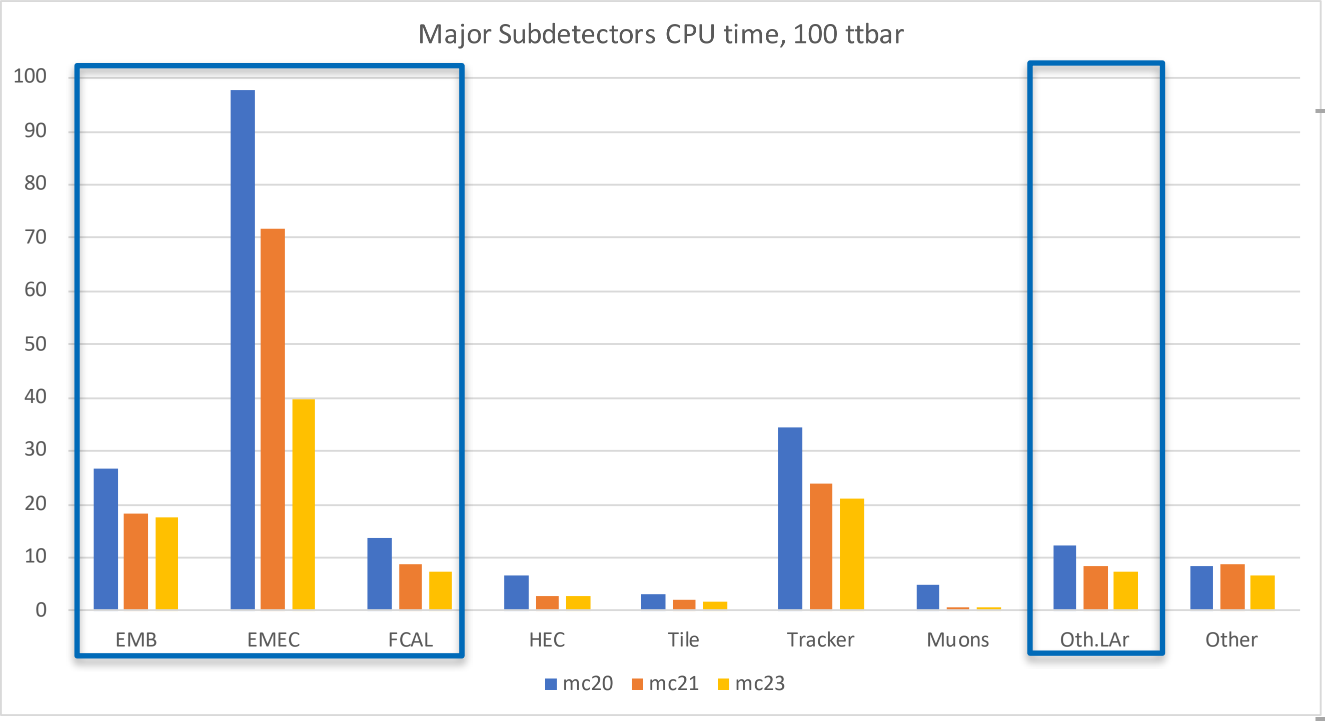


Changes affecting all subsystems

- **mc20, MC reprocessing campaign for Run2:**
  - TRT range cut (sets the range cut for e-/e+ in the TRT volumes filled with Xenon)
  - Frozen Showers
- **mc21a, MC campaign for start of Run3: (in addition to mc20)**
  - EM range cuts: expect to reduce the n. of low energy electrons
  - PRR: expect to reduce the n. of tracked photons
  - NRR: expect to reduce the n. of tracked neutrons
  - EMEC slices: improve the CPU time in the EMEC
  - Big Library: improve overall the Full Simulation Time
  - MagField Off: improve transport in magField
  - G4GammaGeneralProcess: improve time spent to transport gammas
  - BeamPipe killer: kills particles at eta>5.5, that do not make to the detector
- **mc23c (in addition to the mc21), MC campaign for Run3:**
  - Woodcock Tracking: improve timing of gammas in the EMEC
  - VecGeom: Improve CPU time for detectors that use Polycons, Tubes and Cones

Performances of ATLAS Simulation

- Looking from a different perspective (time spent per sub-detector and per particle type) gives us more insights:

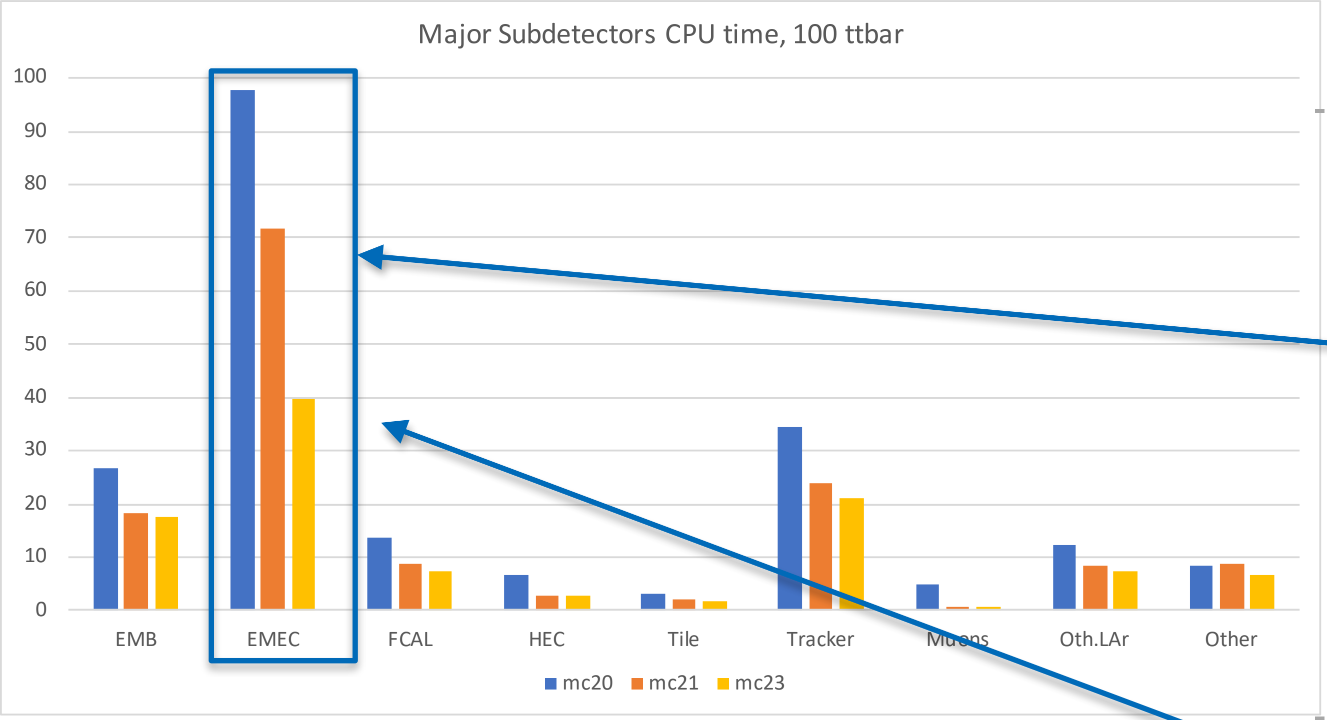


Changes impacting mainly EM calorimeters

- **mc20, MC reprocessing campaign for Run2:**
  - TRT range cut (sets the range cut for e-/e+ in the TRT volumes filled with Xenon)
  - Frozen Showers
- **mc21a, MC campaign for start of Run3: (in addition to mc20)**
  - EM range cuts: expect to reduce the n. of low energy electrons
  - PRR: expect to reduce the n. of tracked photons
  - NRR: expect to reduce the n. of tracked neutrons
  - EMEC slices: improve the CPU time in the EMEC
  - Big Library: improve overall the Full Simulation Time
  - MagField Off: improve transport in magField
  - G4GammaGeneralProcess: improve time spent to transport gammas
  - BeamPipe killer: kills particles at eta>5.5, that do not make to the detector
- **mc23c (in addition to the mc21), MC campaign for Run3:**
  - Woodcock Tracking: improve timing of gammas in the EMEC
  - VecGeom: Improve CPU time for detectors that use Polycons, Tubes and Cones

Performances of ATLAS Simulation

- Looking from a different perspective (time spent per sub-detector and per particle type) gives us more insights:



Changes targeting the EMEC

- **mc20, MC reprocessing campaign for Run2:**
  - TRT range cut (sets the range cut for e-/e+ in the TRT volumes filled with Xenon)
  - Frozen Showers
- **mc21a, MC campaign for start of Run3: (in addition to mc20)**
  - EM range cuts: expect to reduce the n. of low energy electrons
  - PRR: expect to reduce the n. of tracked photons
  - NRR: expect to reduce the n. of tracked neutrons
  - EMEC slices: improve the CPU time in the EMEC
  - Big Library: improve overall the Full Simulation Time
  - MagField Off: improve transport in magField
  - G4GammaGeneralProcess: improve time spent to transport gammas
  - BeamPipe killer: kills particles at eta>5.5, that do not make to the detector
- **mc23c (in addition to the mc21), MC campaign for Run3:**
  - Woodcock Tracking: improve timing of gammas in the EMEC
  - VecGeom: Improve CPU time for detectors that use Polycons, Tubes and Cones

Conclusions and Outlook

- Further **Geant4 optimizations in Run3:** speedup from 33% to 48%
  - full simulation runs **2 times faster!**

- Other **optimisations ongoing**
  - GPU-friendly EMEC implementation **(expected 8-20% speedup)**
  - LTO/PGO **(preliminary tests ~3-4% speedup)**
  - G4HepEM library adoption **(preliminary tests in Athena ~7-8% speedup)**
  - High-η particle rejection (ISF particle killer) **(few % expected)**

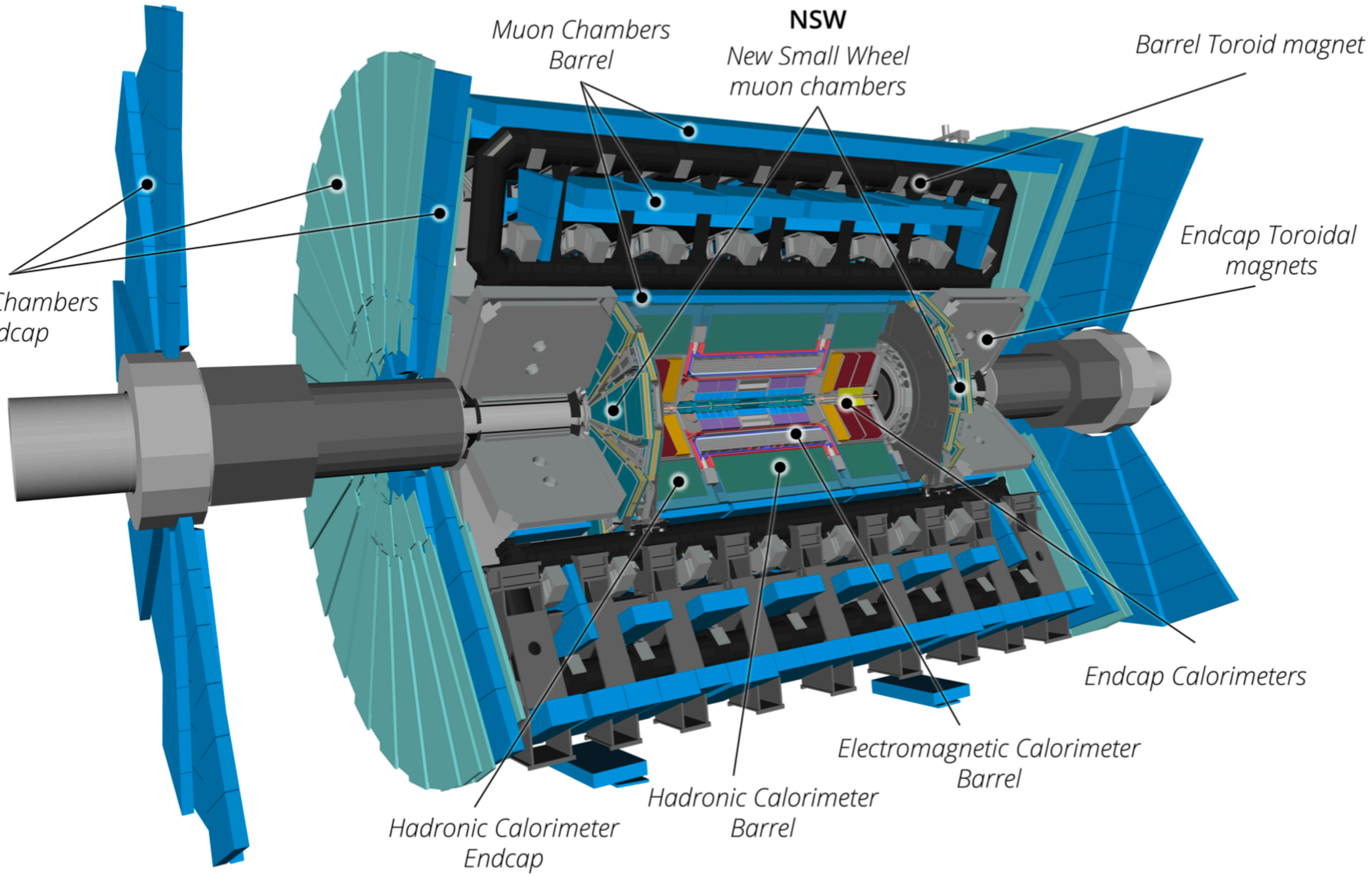
- Many interesting **longer-term developments** (please see backup slides):
  - Quantized State Stepper (QSS) integration and testing
  - ML Correction for Aggressive Range Cuts
  - EM physics tuning
  - G4 Field parameter tuning
  - Voxel Density Optimization

Thanks for your attention!

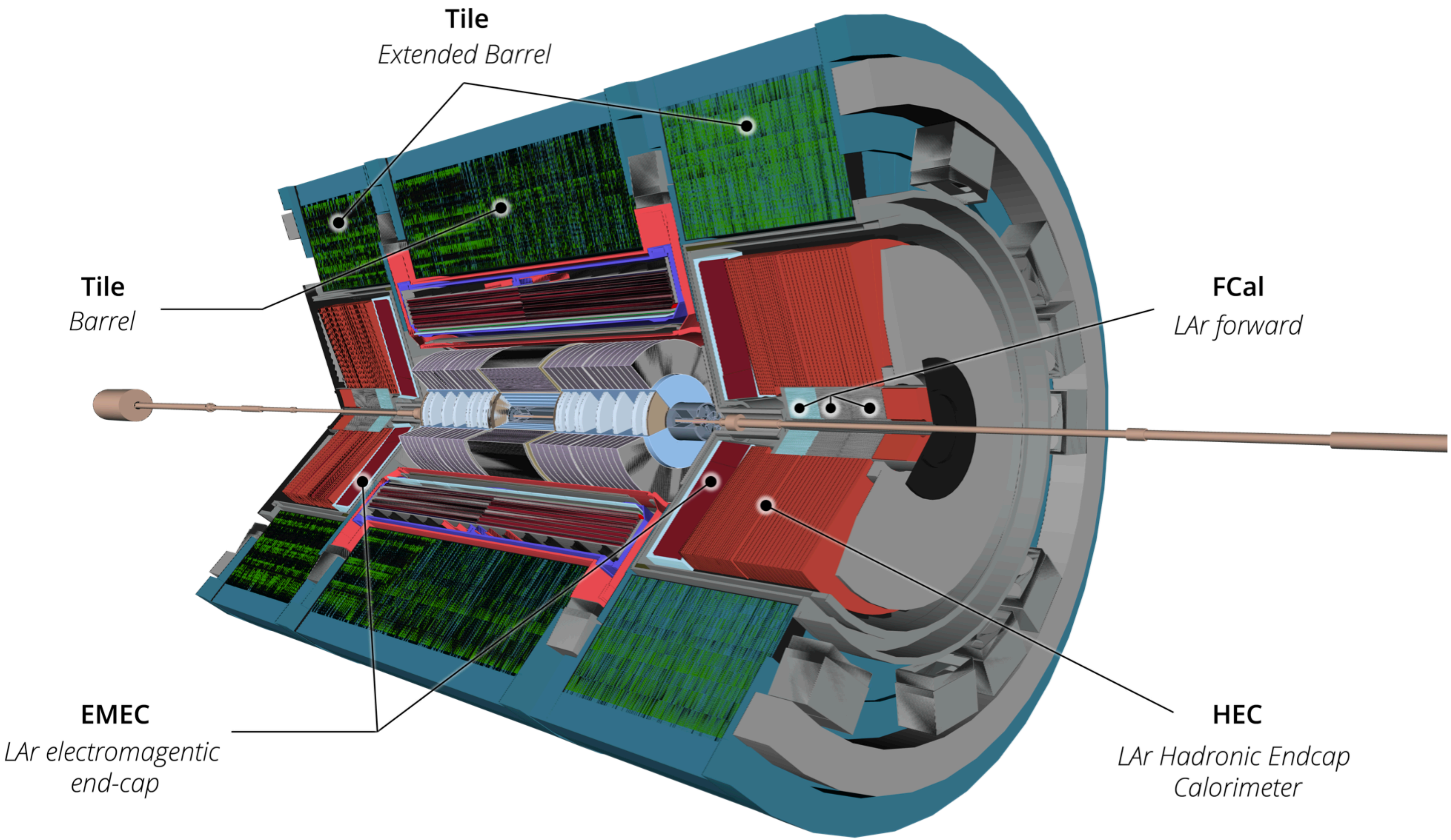
Marilena Bandieramonte
marilena.bandieramonte@cern.ch

Backup slides

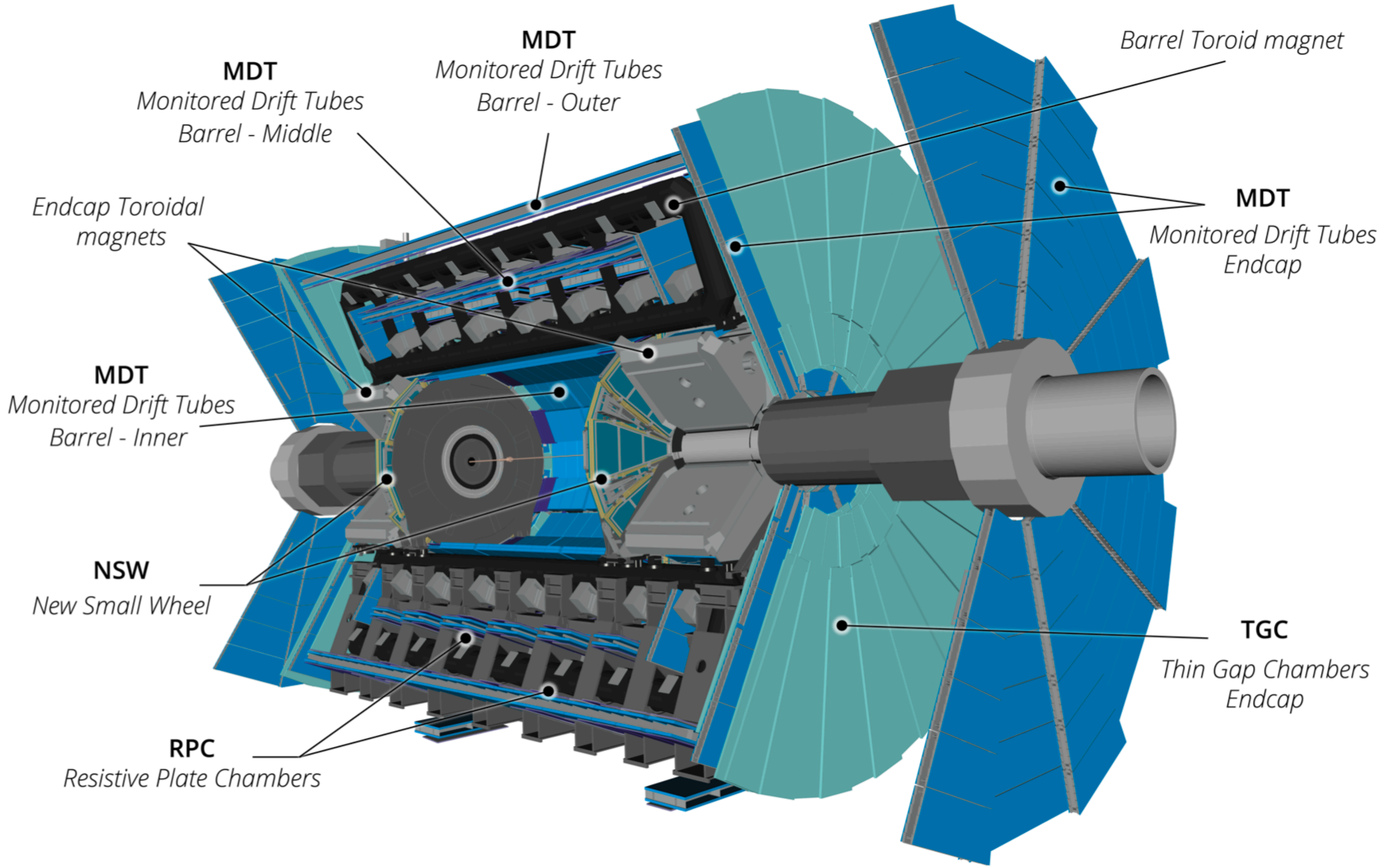
ATLAS detector in Run3



ATLAS detector in Run3 - Calorimeters



ATLAS detector in Run3: Muons System



Geant4 Simulation Optimizations for Run3

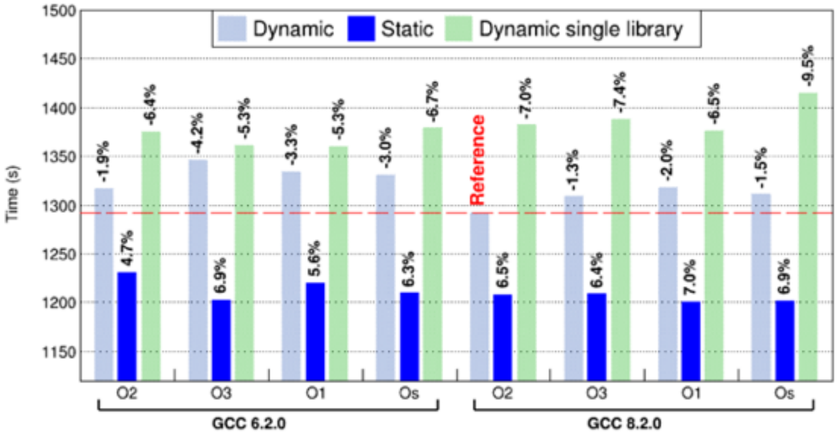
Improvements deployed for the start of Run3 MC production

**details in the presentation of last year*

Geant4 intrinsic improvements

Geant4 static linking (Big Library)

Combine all Athena libraries/components that use Geant4 into a ***single shared library*** linking to Geant4 ***static libraries***



-7%

measured in Athena

B. Morgan, C. Macron [ATLASSIM-3150]

G4GammaGeneralProcess

A single general process for gammas that interacts with the Stepping Manager:
- significantly reduce the number of operations needed

-4.3%

measured on 100 ttbar events in Athena

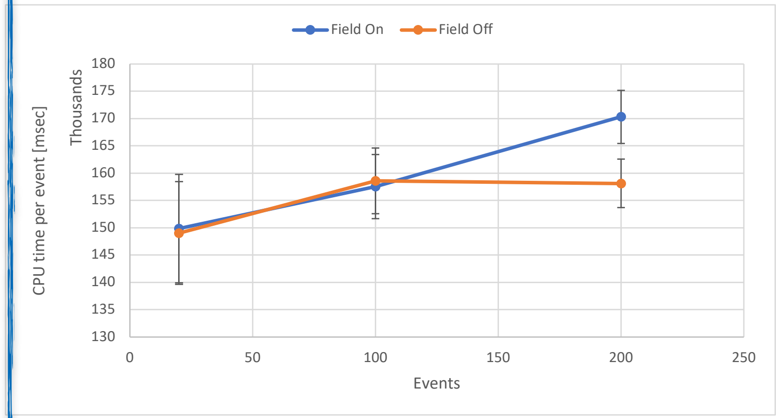
M. Bandieramonte [ATLASSIM-4791] [ATLPHYSVAL-818]

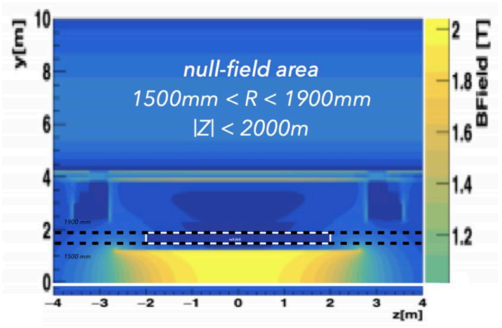
Reduce the number of steps/operations

Magnetic Field Tailored Switch-off

Speedup observed when switching off magnetic field in LAr calorimeter (except for muons) without affecting shower shapes

- **Detailed studies showed smaller null-field area needed**
  - ~1-2% speedup for full ttbar events





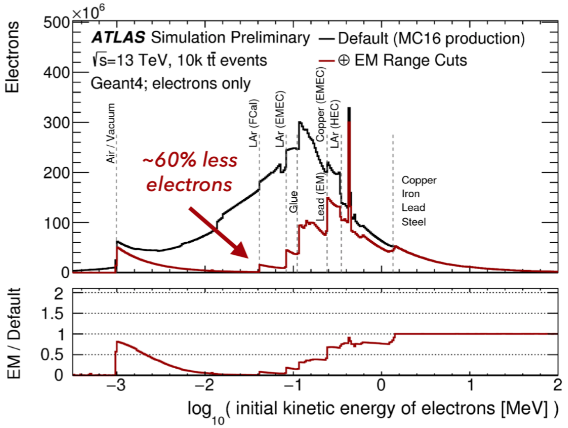
V. Kourlitis [ATLASSIM-4749] [ATLPHYSVAL-773]

EM range cuts

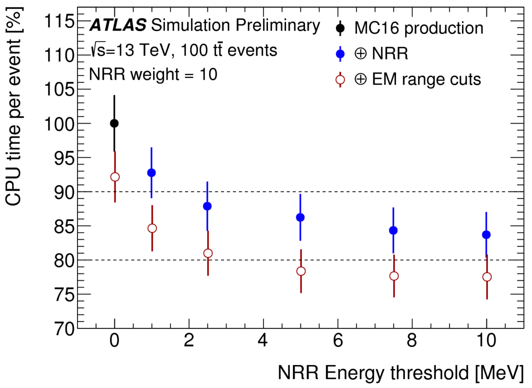
Explicitly activated for *Compton, photoelectric, pair-production*

Avoid creating secondaries and transporting a gamma if its energy is below a certain value

6-7% speedup



Photon and Neutron Russian roulette



Randomly discard particles below the energy threshold and weight accordingly the surviving ones

10% speedup with 2 MeV threshold

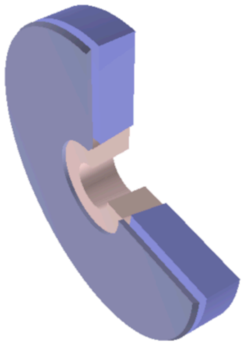
M. Muskinja [SIM-2019-001] [ATLASSIM-3924] [ATLPHYSVAL-632]

Geometry optimizations

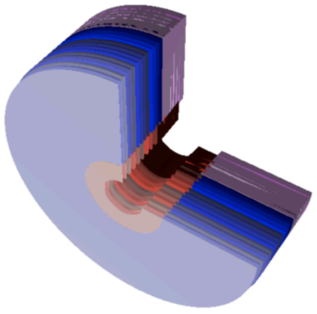
New EMEC custom solid variants

- Wheel the default LArWheelSolid with G4Polycone
- Cone improved LArWheelSolid with G4ShiftedCone - outer wheel divided into two conical-shaped sections
- **Slices** new LArWheelSliceSolid — each wheel is divided into many thick slices along Z axis:
  - **best candidate: 5-6% speedup**

Cone



Slices



A. *Sukarev* [ATLPHYSVAL-785]

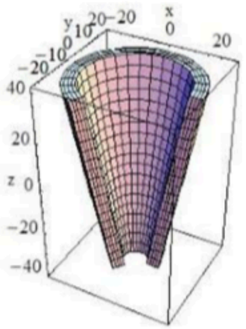
VecGeom

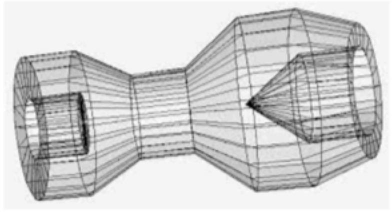
New and optimised implementation of geometrical shapes, designed to take advantage of explicit and implicit vectorisation:

- we replace only *polycons, cons and tubes*

-1.5-7% **

measured on 500 ttbar events in Athena





Ben Morgan [ATLASSIM-4750] [ATLPHYSVAL-831]

Longer terms effort

G4 Field parameter tuning

Goal: Find the optimal values of the in-field tracking parameters for physics performance and CPU savings

Inspired by CMS efforts to optimize the below parameters

- DeltaIntersection: accuracy of intersection with boundary volume
- Epsilong_max: relative accuracy for endpoint of 'ordinary' integration step (delta_one_step / step_size)
- DeltaChord: approximation of curve with linear sections
- MaxStep: maximum step length

Magnetic field parameters	**Strict**	**Intermediate**	**Loose**
DeltaIntersection (mm)	10^{-6}	10^{-4}	0.01
Epsilon_max	10^{-4} /step_size	10^{-3} /step_size	0.1/step_size
DeltaChord (mm)	10^{-3}	$2 \cdot 10^{-3}$	0.1
MaxStep (mm)	150	50	150

N. *Nitika* [ATLASSIM-6147]

Quantized State Stepper (QSS)

Quantized State System (QSS): numerical methods to solve the ordinary differential equations that describe the movement of particles in a field

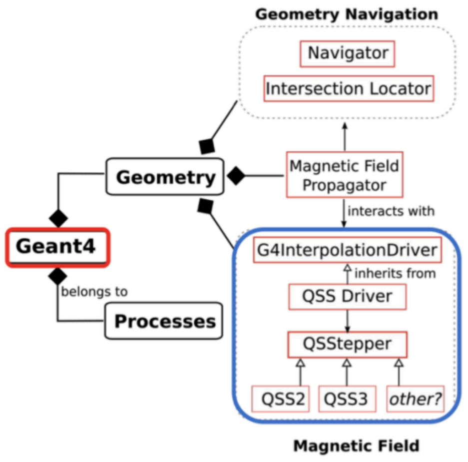
- QSS methods **discretize the system state** variables as opposed to the traditional methods that discretize the time

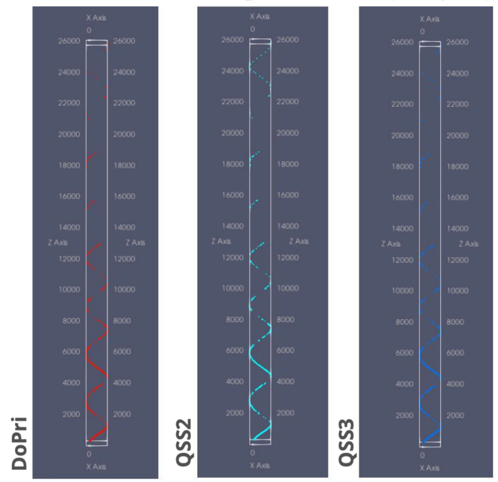
- This method handles very efficiently **discontinuities** in the simulation of continuous systems

Reference: Efficient discrete-event based particle tracking simulation for high energy physics

STATUS

- Successfully ported QSS stepper from Geant4 v10.5 to v10.7.2
  - Test within the ATLAS geometry:
    - Results using the N02 model qualitatively indistinguishable compared to those using the G4DormandPrince745
- Integrated in **FullSimLight**
  - ATLAS geometry & magnetic field map
  - Performance profiling ongoing





R. Castro, L. Santi,Leandro Gómez Vidal, Alejandro Mignanelli: [link]

EM Physics tuning

Geant4 simulation can be tuned to the needs of the specific simulation features (detector, physics, required accuracy, computational time)

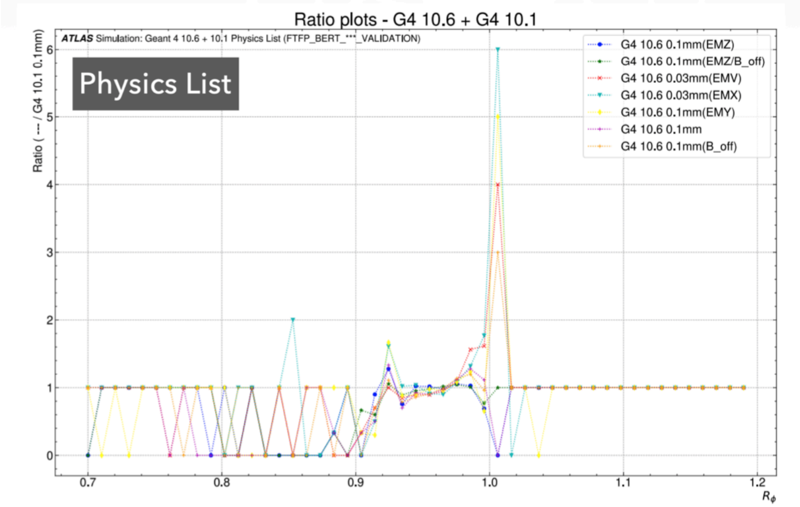
Goal: find the best compromise between simulation accuracy and speed in order to improve data-mc agreement

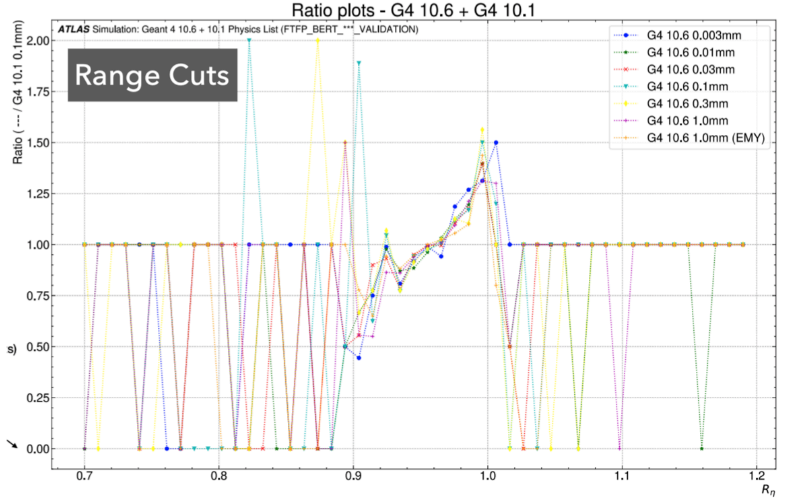
Parameters to look at:

- Physics Lists
- Range cuts
- Different Geant4 versions
- MSC range factor

G4 simulation is compared to measurements available from the ATLAS CP groups, especially gamma and jet/emiss

- Studies ongoing



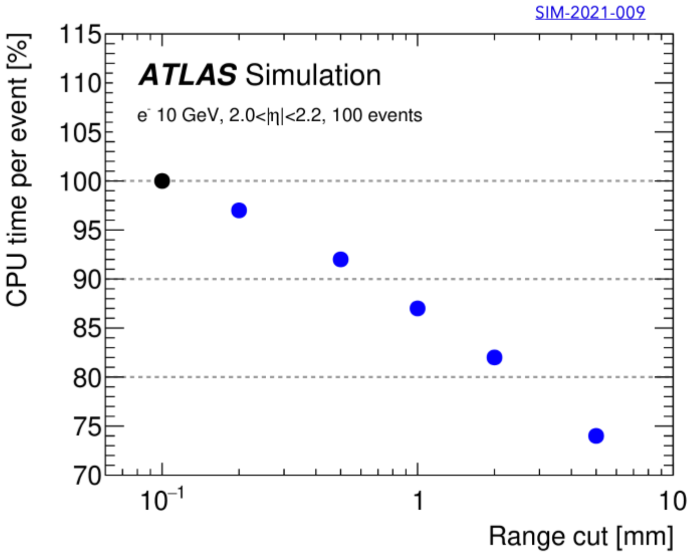


K. Dongwon [ATLASSCQT-34]

ML Correction for Aggressive Range Cuts

Increased range cuts can **reduce the number of photons**, thus reduce the transportation steps and **increase computational performance**

EM calorimeters dominate the simulation load due to low-energy photons from electron scattering, ~90% of these are transportation processes



Side-effect:
"High" range cuts can degrade the accuracy of the simulation

The ML correction applied as a post-processing step utilizing **batch processing** and **accelerator hardware** achieving **~15% speed up** in example geometries – ML inference time negligible compared to simulation time reduction.

Solution to be implemented/tuned for the ATLAS EMEC

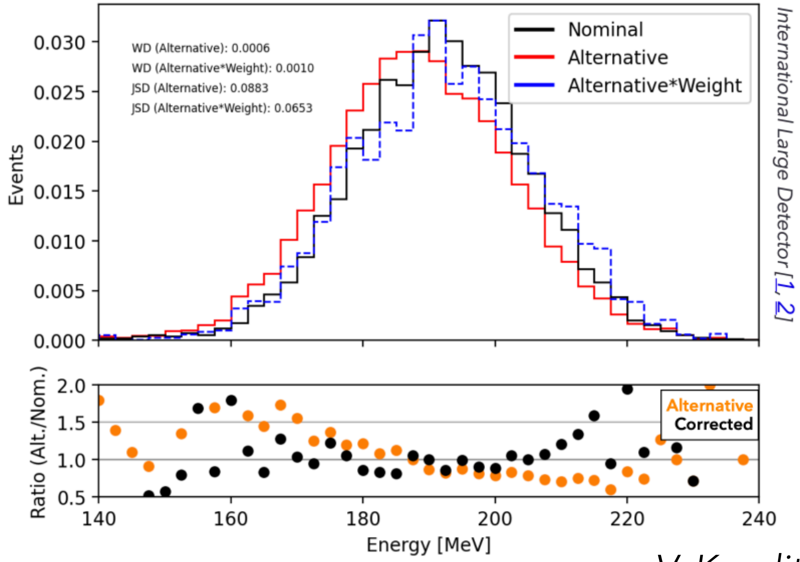
ML-based correction

Classification NN to learn correction weights [ref]

*Re-weight the **alternative** simulation to the **nominal** one by learning multi-dimensional weights considering all cell energy deposits*

$$r(\vec{x}) = \frac{p(\vec{x}\,|\,\theta_p)}{q(\vec{x}\,|\,\theta_q)}$$

θ be the range cut, $\boldsymbol{x}$ the energy deposits

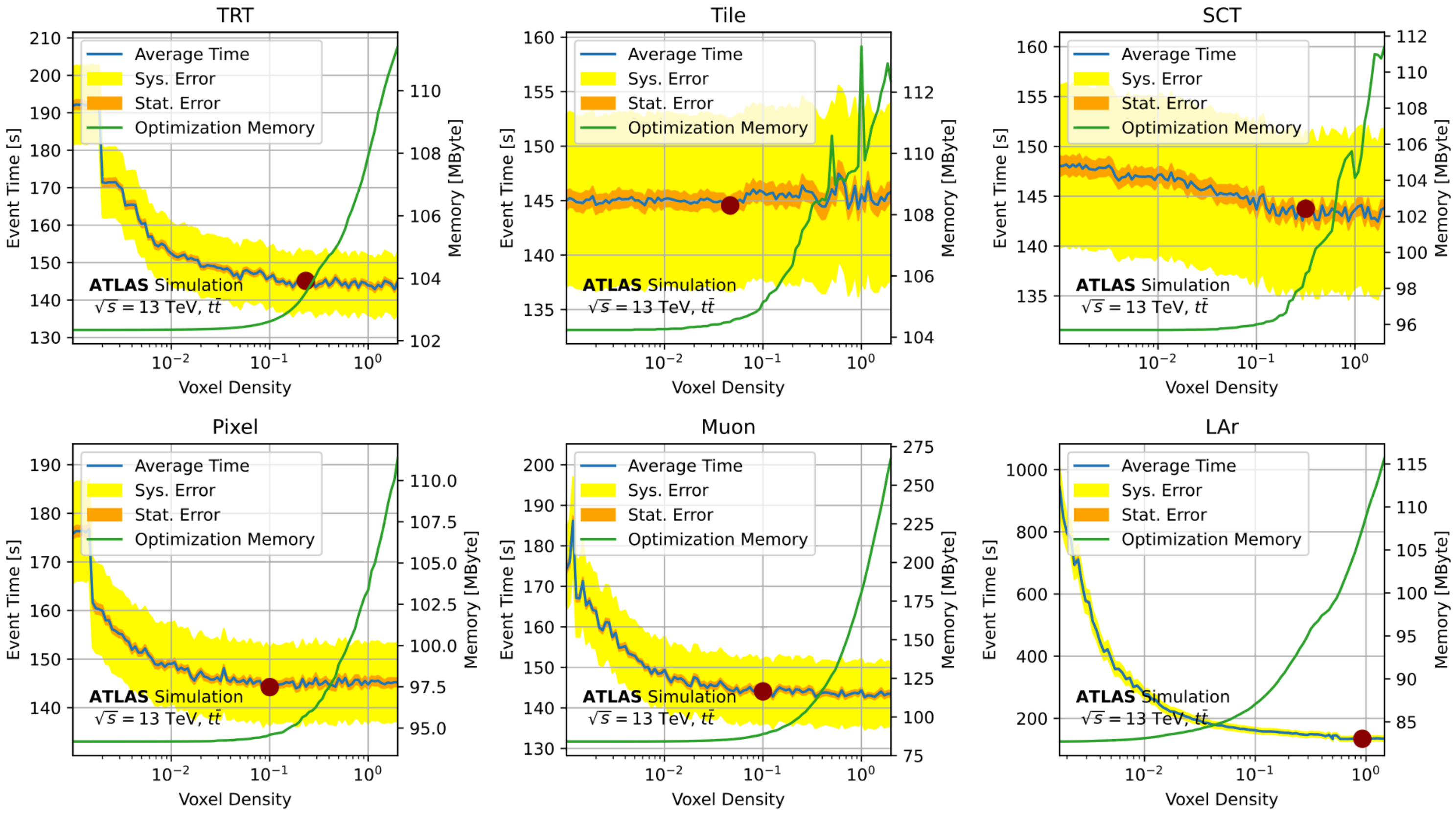


V. Kourlitis link

Voxel Density Optimization

Tracking can be optimized by voxelization, the **size/granularity of the voxels** can be tuned

- **Goal**: Determine optimal value for the voxel density
    - Balance between memory used for the detector description and CPU time for simulation, without compromising the accuracy
- Changes studied in FullSimLight & Athena
- ~40% memory footprint reduction for detector description [ATLPHYSVAL-887]
- **No expected speedup/slowdown: some recent tests on the GRID showed a slowdown when VecGeom is active. Currently investigating.**



Mustafa A. Schmidt [ATLASSIM-5876]