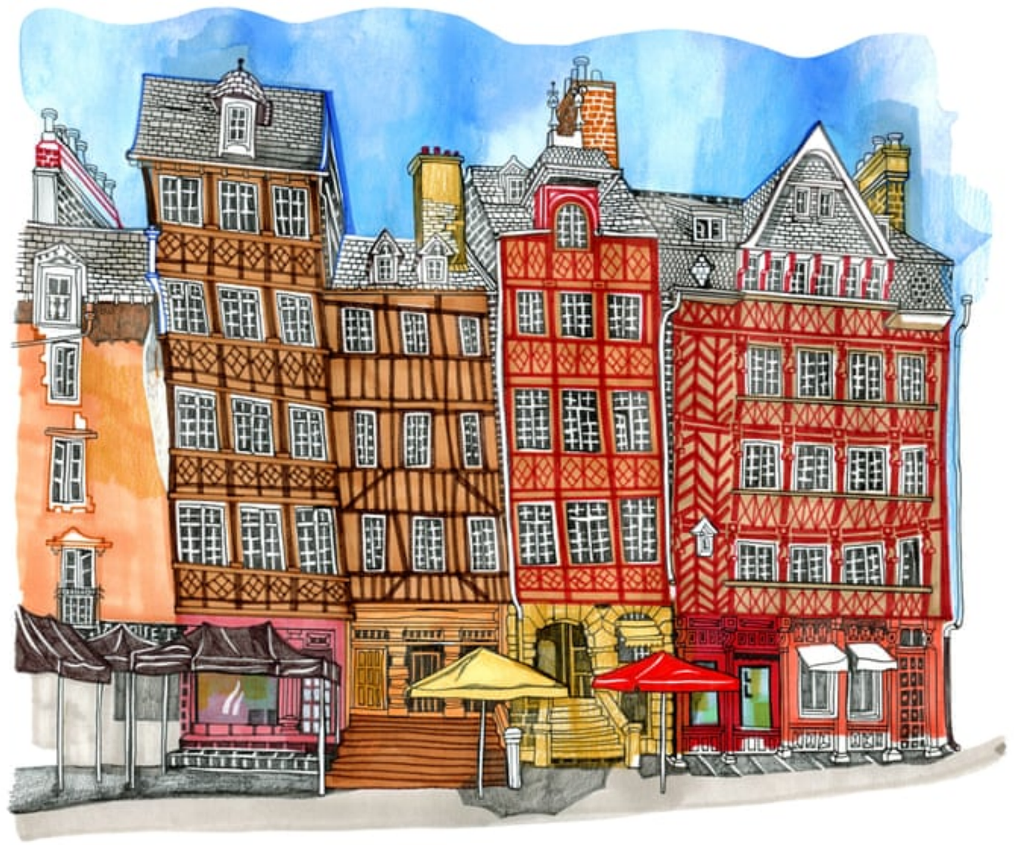
Computing Performance Results and Issues: ATLAS

27th Geant4 Collaboration Meeting

Rennes

26-30 September 2022



Marilena Bandieramonte (University of Pittsburgh)

Geant4 Simulation in ATLAS

- For **RUN3 Monte Carlo production ATLAS is using**:
  - **Geant4 10.6.patch03.atlas03**
    - It includes:
      - G4AtlasRK4 stepper
      - G4MagInt_Driver patch to reduce differences w.r.t. G4IntegrationDriver
      - G4GammaGeneralProcess patch from Geant4-10.7.2
  - **Geant4 10.7** might be used for the next year MC campaign (to be updated at the end of the year)

Geant4 Simulation Optimizations for Run3

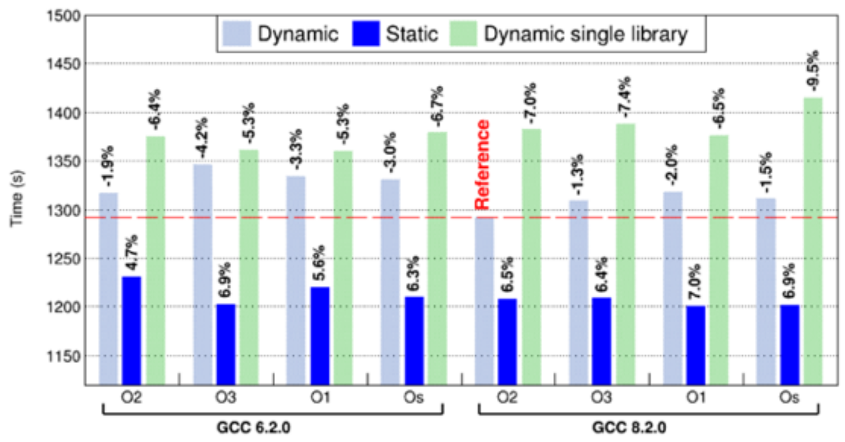
Improvements that made it to the start of the Run3 MC production

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***

- Potential area for further investigations such as LTO/PGO?



-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

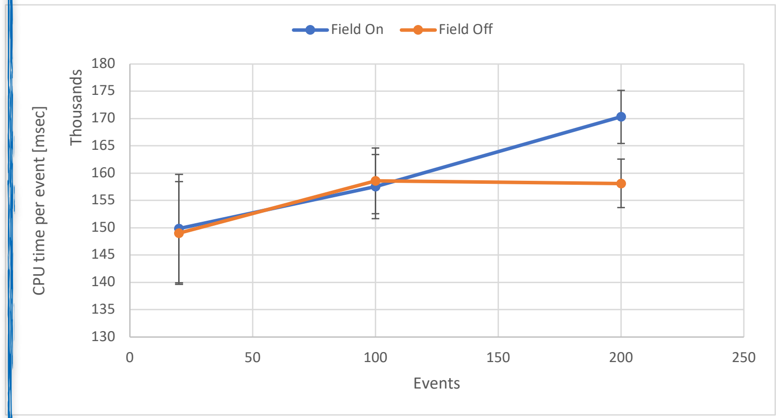
M. Bandieramonte [ATLASSIM-4791] [ATLPHYSVAL-818]

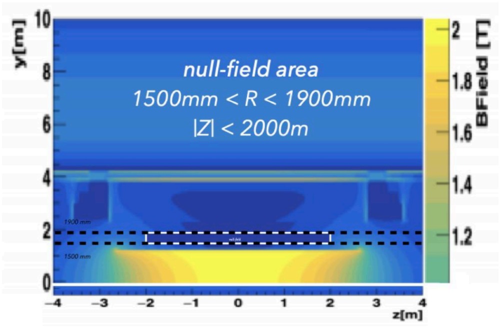
Reduce the number of steps/operations

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





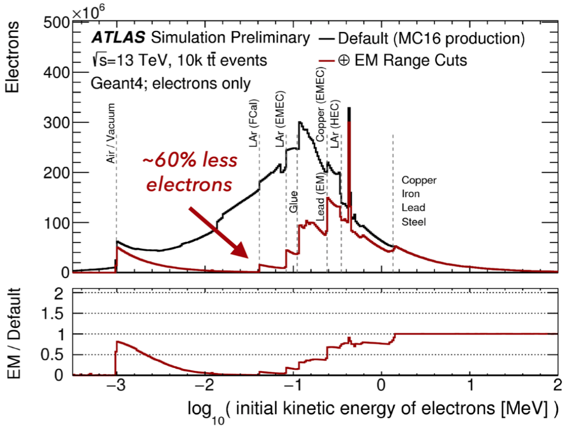
V. Kourlitis [ATLASSIM-4749] [ATLPHYSVAL-773]

EM range cuts

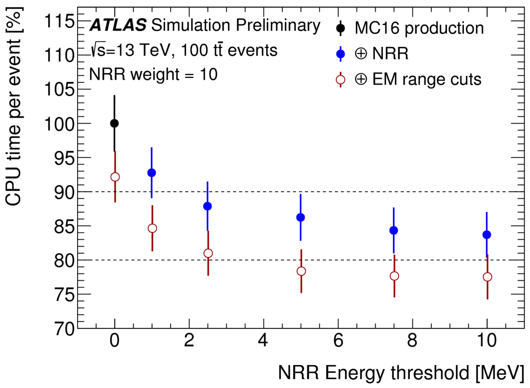
Explicitly activated for *Compton, photoelectric, pair-production*

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

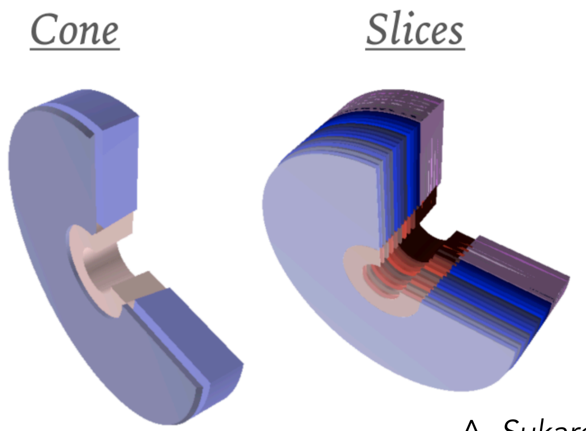
10% speedup with 2 MeV threshold

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**



A. *Sukarev* [ATLPHYSVAL-785]

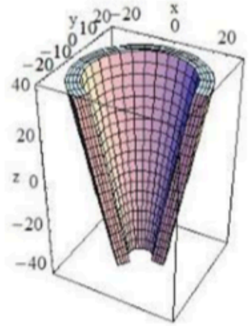
VecGeom*

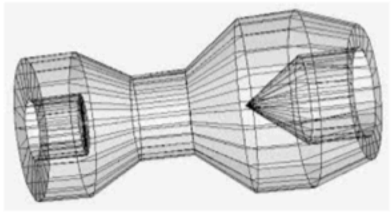
New and optimised implementation of geometrical shapes, designed to take advantage of explicit and implicit vectorisation:

- we replace only *polycons, cons and tubes*

-1.5-7% **

measured on 500 ttbar events in Athena





** Issue in production (see next slide)*

***more in benchmark slides*

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

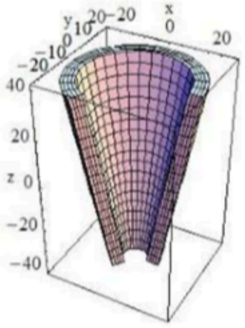
VecGeom integration

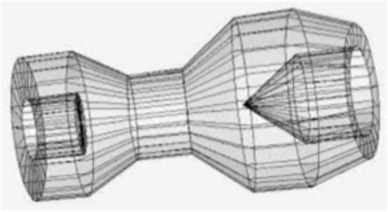
- *Issue discovered in production:*
  - *~10% of Minbias events crashing*
  - *Reproducible in standalone with FullSimLight*
    - *Minimal geometry example*
    - https://sft.its.cern.ch/jira/browse/VECGEOM-601
  - *Successfully debugged and solved (thanks Ben and Gabriele!)*
    - [VecGeom/VecGeom!874 (merged)]()

- *These days VecGeom is being restored in Athena*
  - Add local patch for consistent tolerance in cones on top of VecGeom1.1.20

-1.5-7% *

measured on 500 ttbar events in Athena





***see benchmark slides*

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

Ongoing effort

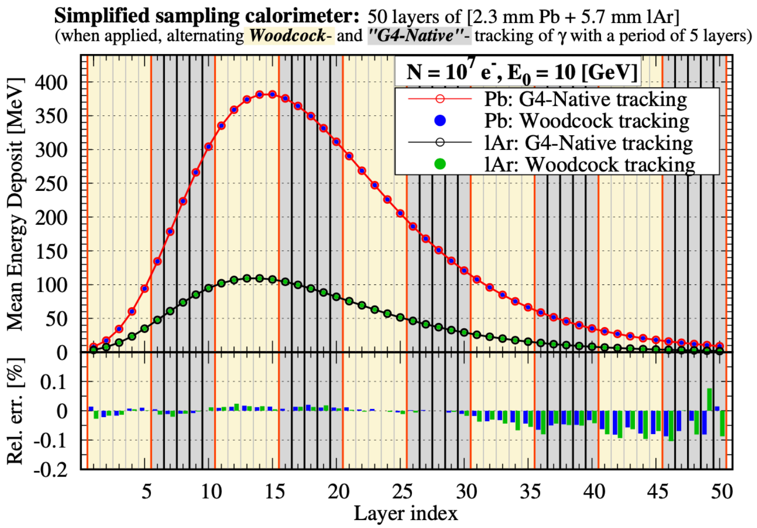
Targeting Run3 MC production of next year

Woodcock tracking

- Reduce the **number of steps** 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 **8-9 %** 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 (measured on 50 ttbar events)
    - **Full physics validation will start in the coming weeks!**

		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, W. Hopkins, J. Apostolakis: [ATLASSIM-5079]

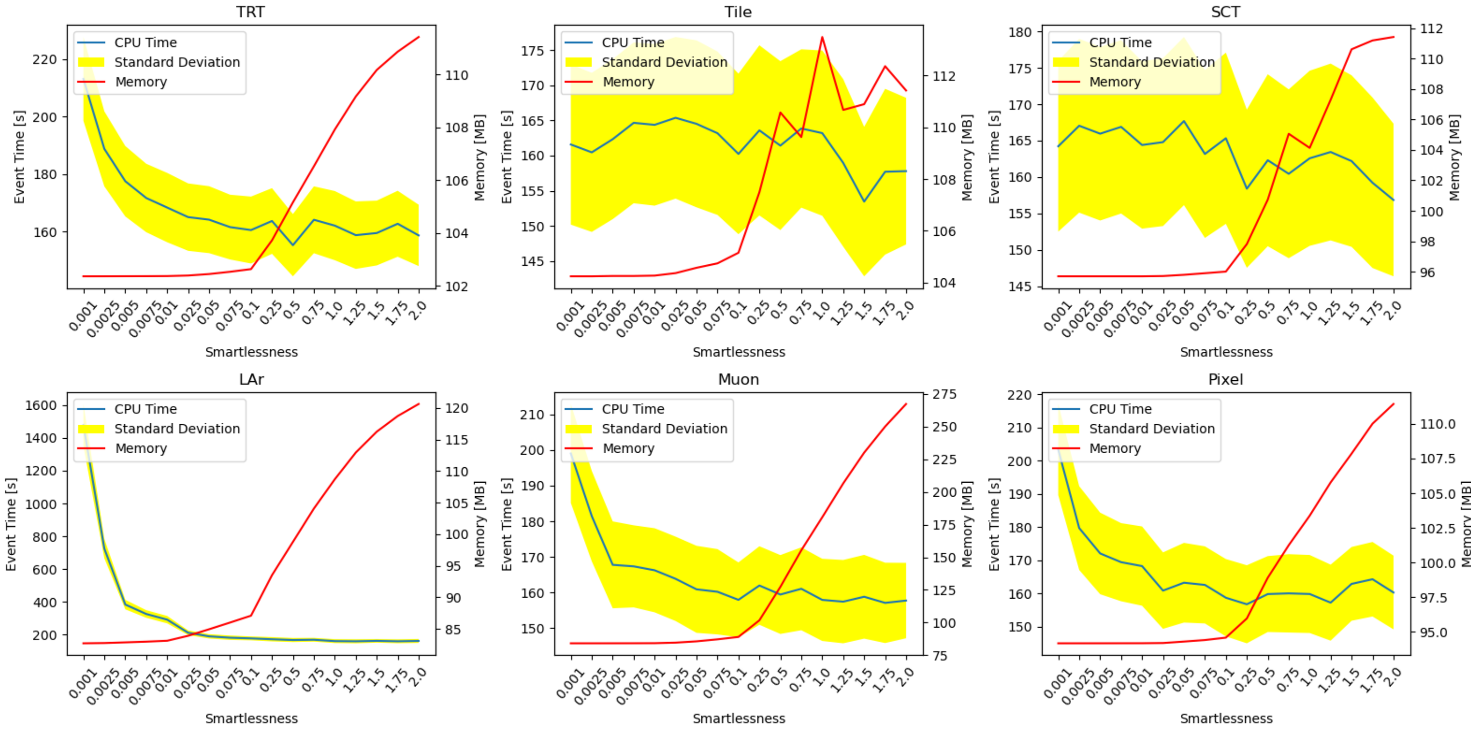
Voxel Density Tuning

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

- Optimal values tested give a preliminary:
  - **~5-6 %** speedup
  - **~40%** less memory footprint

- **Full physics validation will start in the coming weeks!**



smaller/finer voxels ⟷ *larger/coarser voxels*

Mustafa Andre Schmidt **[ATLASSIM-5876]**

Benchmarks

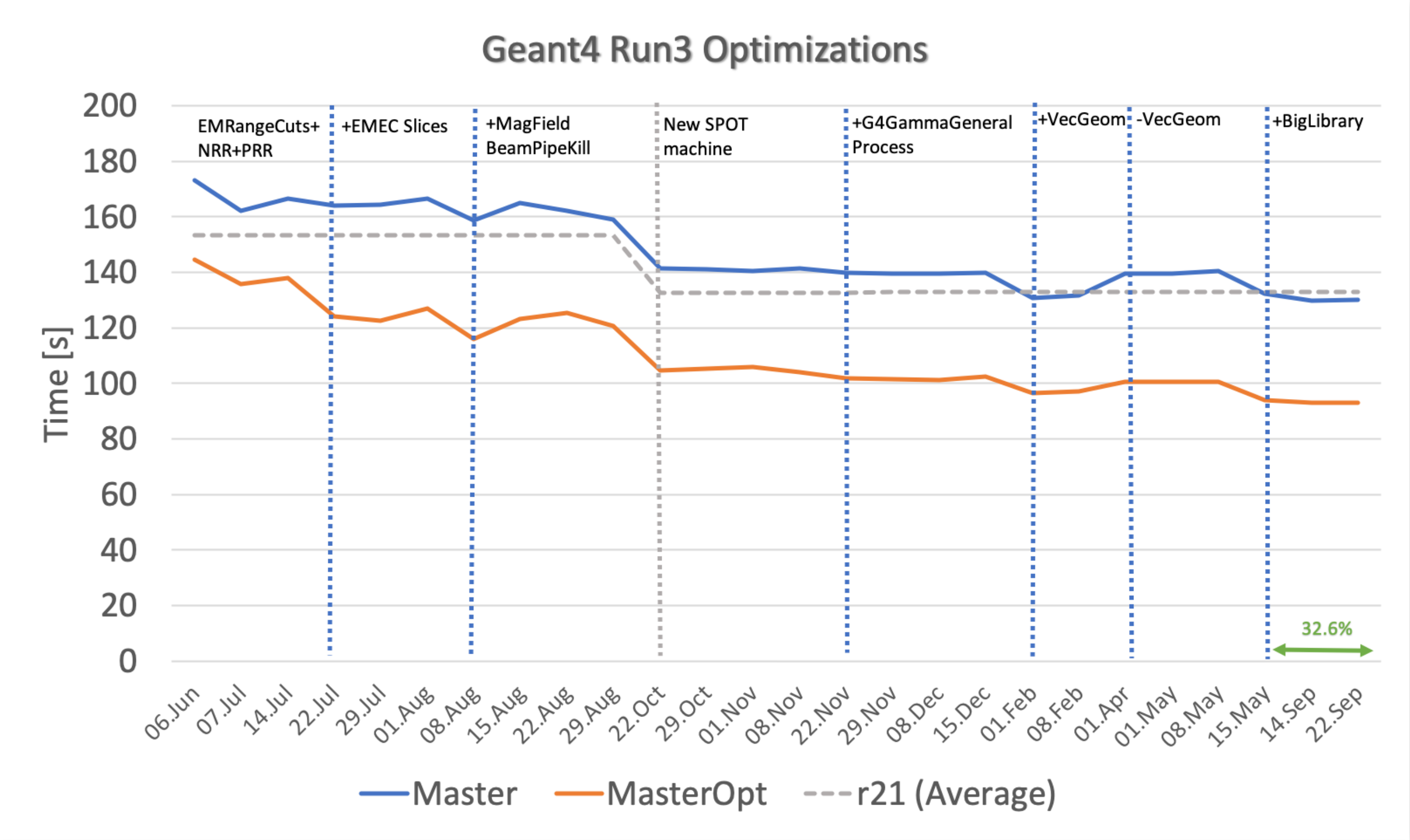
Run3 Geant4 Optimization Benchmarks

Benchmarks run on a quiet server:

2 x AMD [EPYC 7002 series]

- disable CPU boosting
- disable SMT (Simultaneous Multithreading)
- set the governor to performance mode w/ the upper limit set to base clock frequency

- Additional tricks (not yet implemented)
  - flushing the memory caches
  - disallowing swapping etc.

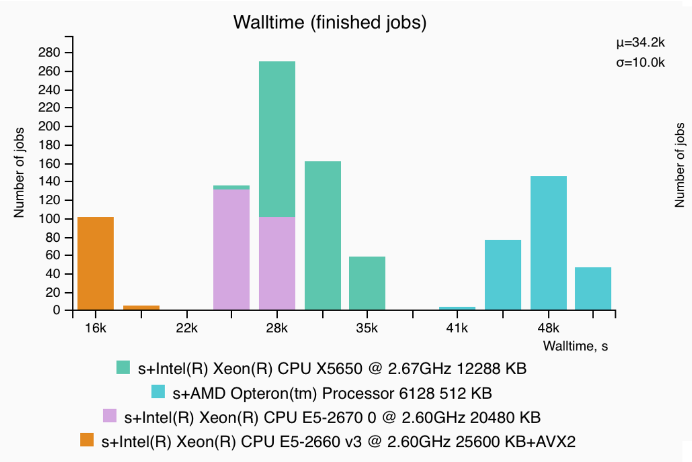


Data elaborated from the SPOT benchmarks

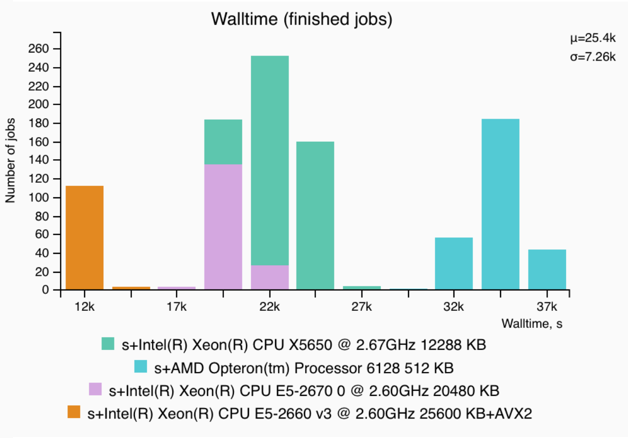
Run3 Geant4 Optimization GRID Benchmarks

Improvement is also observed in "realistic" production conditions – grid sites

Athena 22.0.47 No Opt
JobID: 27857061



Athena 22.0.47 +Run3 Opt
JobID: 27857078



Athena 22.0.47 + VecGeom No Opt
JobID: 27827044

Athena 22.0.47 + VecGeom+Run3Opt
JobID: 27827045

Australia-ATLAS Cluster, 1000 Jobs 100 tt-bar events/job	**Walltime,s**	**sigma**	**Uncertanty**	**Speedup Throughput**	**Speedup CPU time**
Athena 22.0.47 (baseline)	34.2k	10k	~0.92%	**-**	
Athena 22.0.47+Run3Opt	25.4k	7.26k	~0.9%	**34,64%**	**-25,73%**
Athena 22.0.47+VecGeom	33.7k	8.81k	~0.8%	**1,48%**	**-1,46%**
Athena 22.0.47+VecGeom+Run3Opt	22.7k	6.52k	~0.91%	**50,66%**	**-33,62%**

Assuming that **t1** is the initial CPU time/event and **t2** is the final one (after applying Run3Opt) the 2 speedups are:

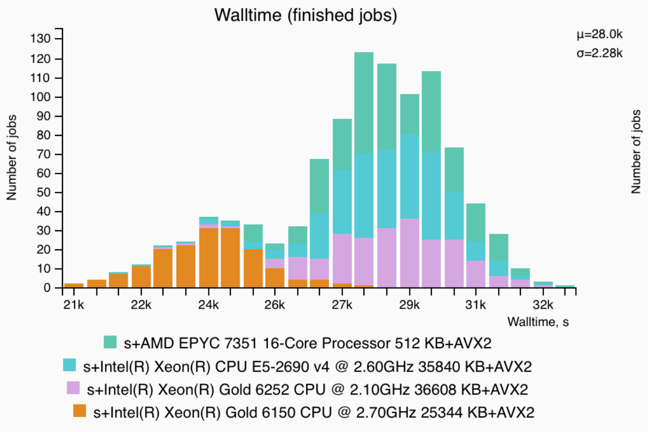
CPU time speedup=[(t2-t1)/t1]*100
Throughput speedup=[(t1-t2)/t2]*100

Still missing in the pipeline:

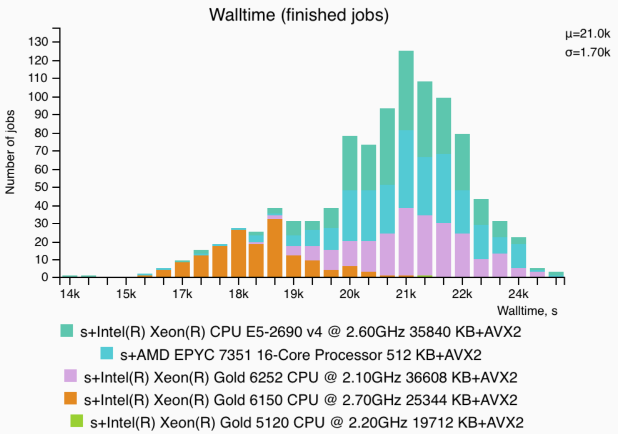
- *BigLibrary (~7%)*

Run3 Geant4 Optimization GRID Benchmarks

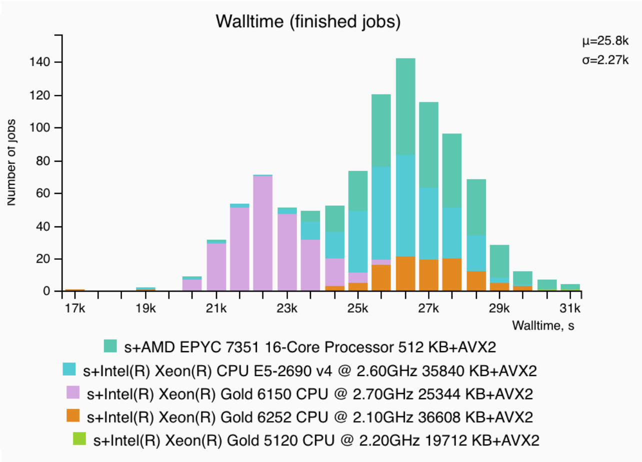
Athena 22.0.47 No Opt
JobID: 27857092



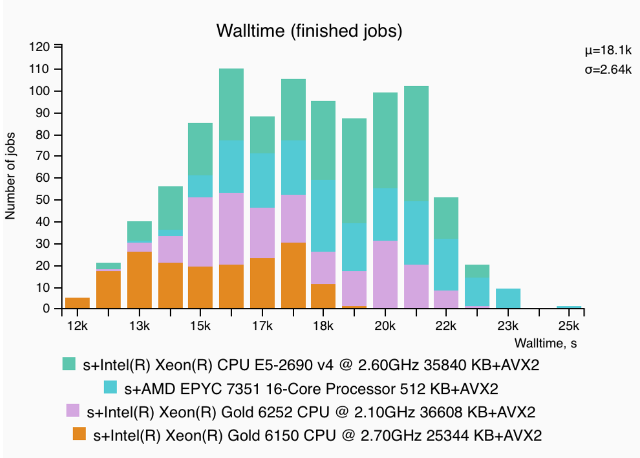
Athena 22.0.47 +Run3 Opt
JobID: 27857110



Athena 22.0.47 + VecGeom No Opt
JobID:27827017



Athena 22.0.47 + VecGeom+Run3Opt
JobID: 27827024



BNL Cluster, 1000 Jobs 100 tt-bar events/job	**Walltime,s**	**sigma**	**Uncertanty**	**Speedup Throughput**	**Speedup CPU time**
Athena 22.0.47 (baseline)	28.0k	2.28k	~0.25%	**-**	
Athena 22.0.47+Run3Opt	21.0k	1.7k	~0.26%	**33%**	**-25%**
Athena 22.0.47+VecGeom	25.8k	2.27k	~0.28%	**8,52%**	**-7,85%**
Athena 22.0.47+VecGeom+Run3Opt	18.1k	2.64k	~0.46%	**54,69%**	**-35,35%**

Assuming that **t1** is the initial CPU time/event and **t2** is the final one (after applying Run3Opt) the 2 speedups are:

CPU time speedup=[(t2-t1)/t1]*100
Throughput speedup=[(t1-t2)/t2]*100

Still missing in the pipeline:
- *BigLibrary (~7%)*

Issue: TLS impact reduction

- Detailed Athena profiling showed a slowdown due to the usage of Thread Local Storage (TLS):
  - Going from single-threaded to MT Geant4 has a 5-10% impact on performance
  - Noticeable when comparing r21 vs master performance
- Test on a dedicated machine with turboboost and hyper-threading turned off
  - 100 tt-bar events in Athena r21

G4 version	Event time	Slowdown
Geant4 10.1	193.6 +- 7.28	-
Geant4 10.1 MT	200.1 +- 7.525	3,35%
Geant4 10.6	198.6 +- 7.898	2,5%
Geant4 10.6 MT	195.3 +- 7.61	0,87%

- TLS is used both in Athena and Geant4:

- Athena -> Magnetic Field
- Geant4 -> Geometry

J. Apostolakis

Conclusions and Outlook

- Achieved a **33-35% CPU** speedup for Run3 MC campaign
  - Can produce **50-55%** more event with the same computing resources
- Other **optimisations ongoing** (coming soon!)
  - Woodcock Tracking (~7-8%)
  - Voxel density Optimisation (6-7%)
  - VecGeom (2-7%)
- Many interesting **longer-term developments** (please see backup slides):
  - G4HepEM library adoption **(already integrated in Athena! ~7-8% speedup)**
  - GPU-friendly EMEC implementation
  - New ISF particle killer
  - Quantized State Stepper (QSS) integration and testing
  - ML Correction for Aggressive Range Cuts
  - EM physics tuning

Thanks for your attention!

Marilena Bandieramonte
marilena.bandieramonte@cern.ch

Backup slides

Longer terms effort

G4HepEM library integration

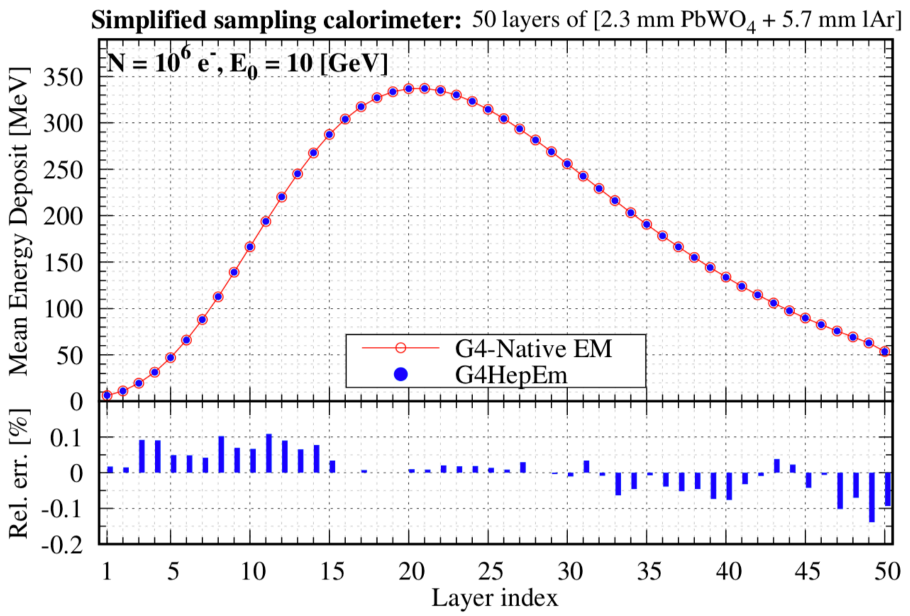
G4HepEM library is a new compact Geant4 EM library, from *Mihaly Novak, Jonas Hanfeld, Benjamin Morgan*

- 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 then 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



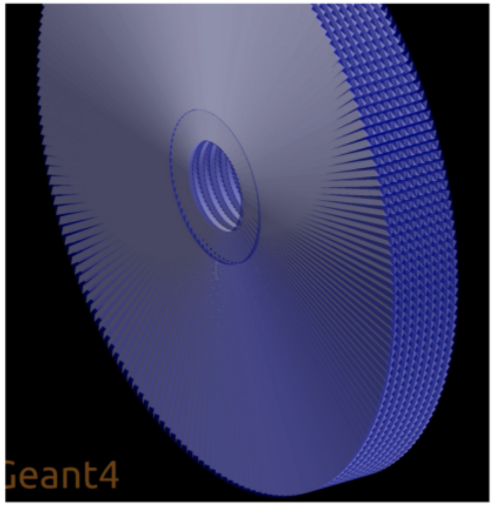
1 ‰ agreement between the native Geant4 and G4HepEM Simulation

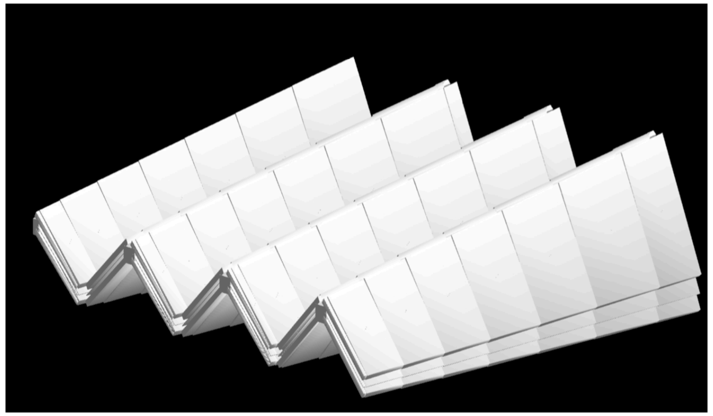
B. Wynne, M. Novak, J. Apostolakis [ATLASSIM-5410]

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 (see AdePT project)
- Repeated accordion volume implementations using:
  - `G4GenericTrap`(converted from `G4TwistedTrap`)
  - `Arb8`& `G4Trap`
- **Repository**: https://gitlab.cern.ch/avishwak/atlas_emec_g4
- STATUS: **resumed**





G4TwistedTrap

Benchmark run in FullsimLight for G4Trap and G4GenericTrap using 1000 events with 10 GeV electrons

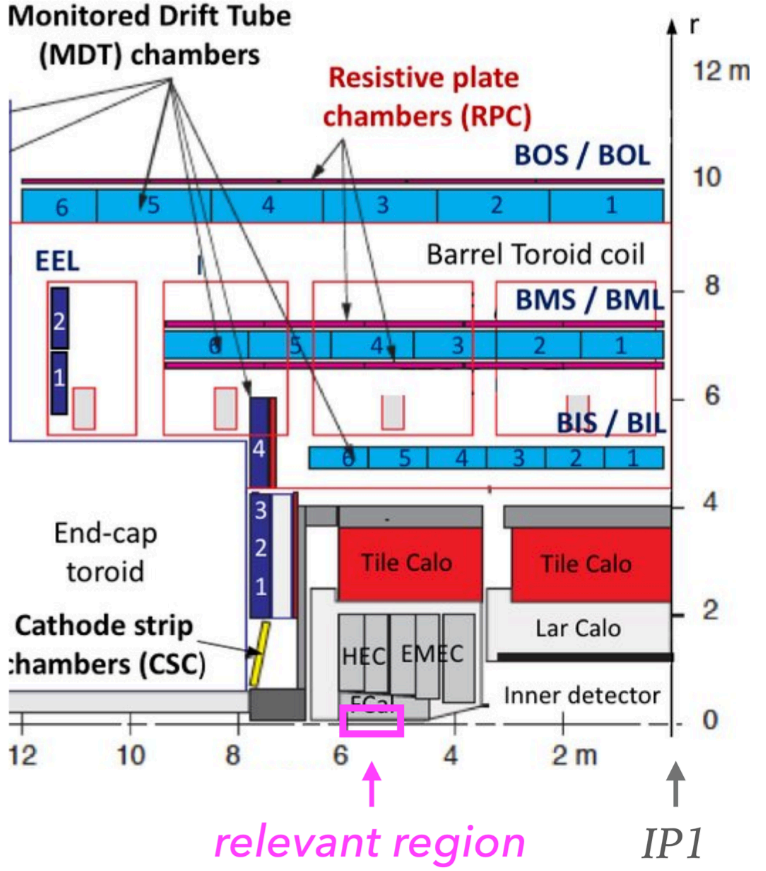
Geometry	Time (s)
G4Trap	111.67
G4GenericTrap	72.07

A. Vishwakarma, E. Tcherniaev [ATLASSIM-4752]

New ISF particle killer

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

- There is a huge amount of secondaries being created 5-6m away from (0,0,0) close to the beam-pipe
- Many of these secondaries will never cause any energy in the calorimeters or a muon hit
  - the primary particles that caused these interactions could just be dropped directly
- Approach:
  - generate a large sample of single particles with 4,5 < |η| < 6 and different energies
  - map out which eta/energy combinations can produce a relevant signal
  - drop the rest directly with a new ISF particle filter.



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

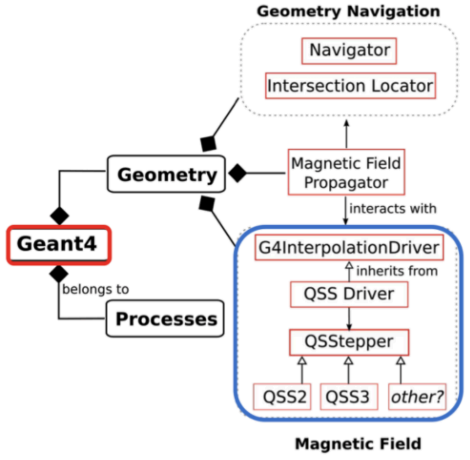
- QSS methods **discretize the system state** variables as opposed to the traditional methods that discretize the time

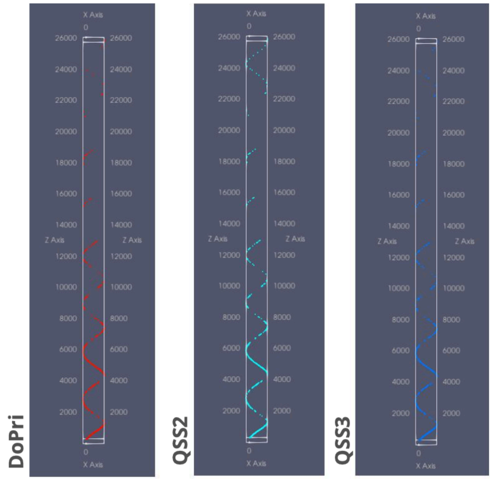
- This method handles very efficiently **discontinuities** in the simulation of continuous systems

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)

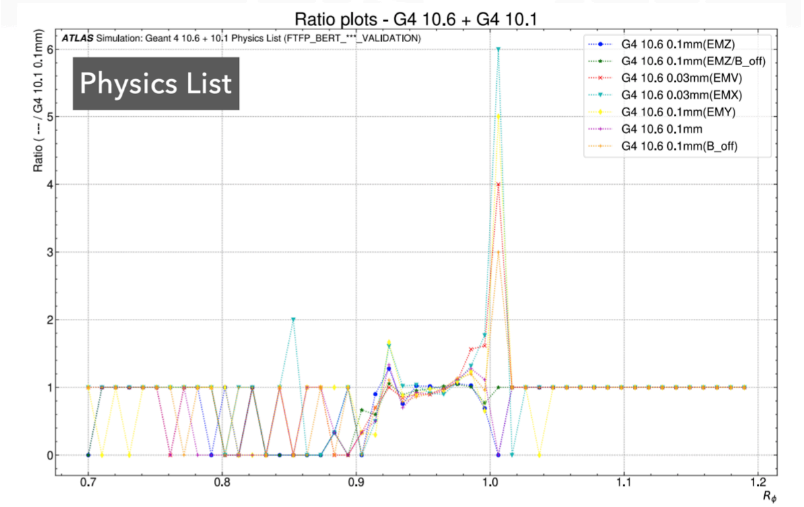
Goal: find the best compromise between simulation accuracy and speed in order to improve data-mc agreement

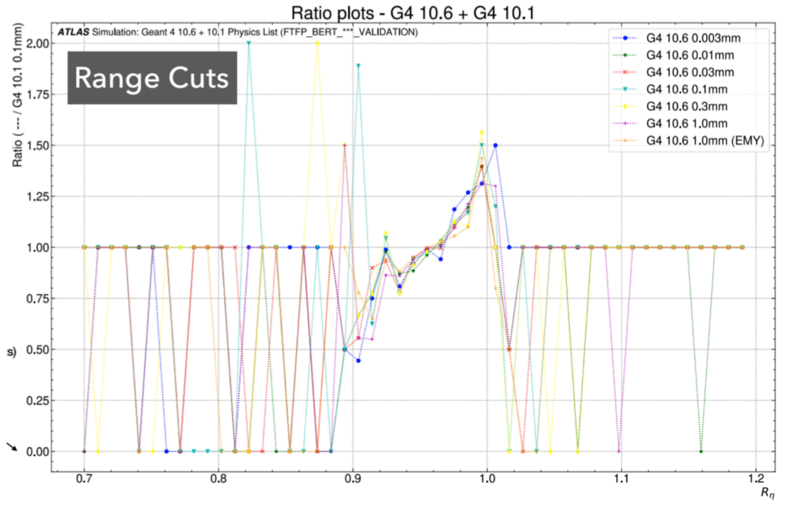
Parameters to look at:

- 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



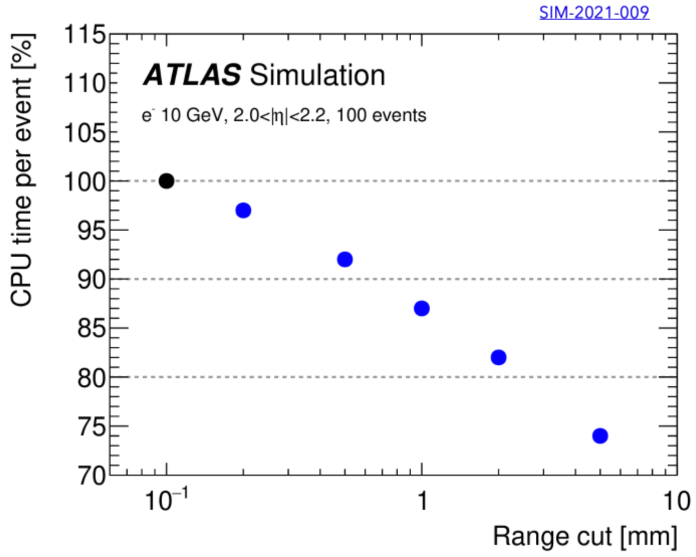


K. Dongwon [ATLASSCQT-34]

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

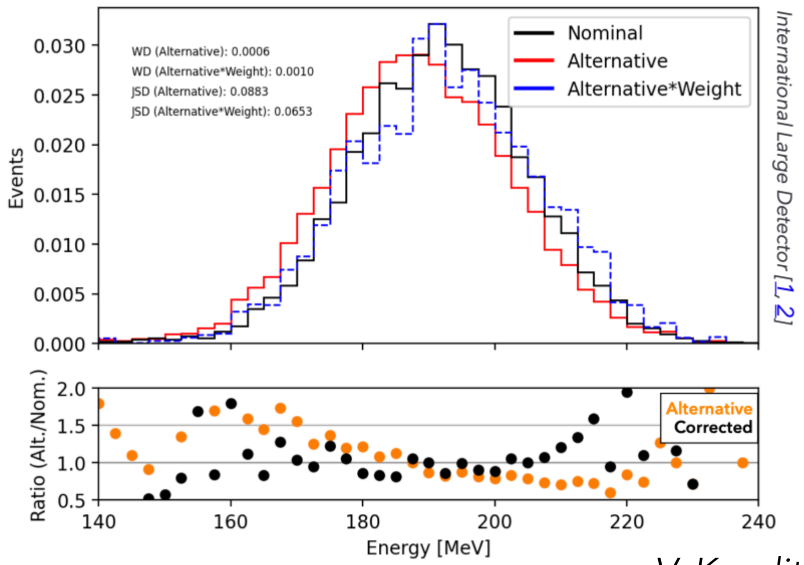
ML-based correction

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*

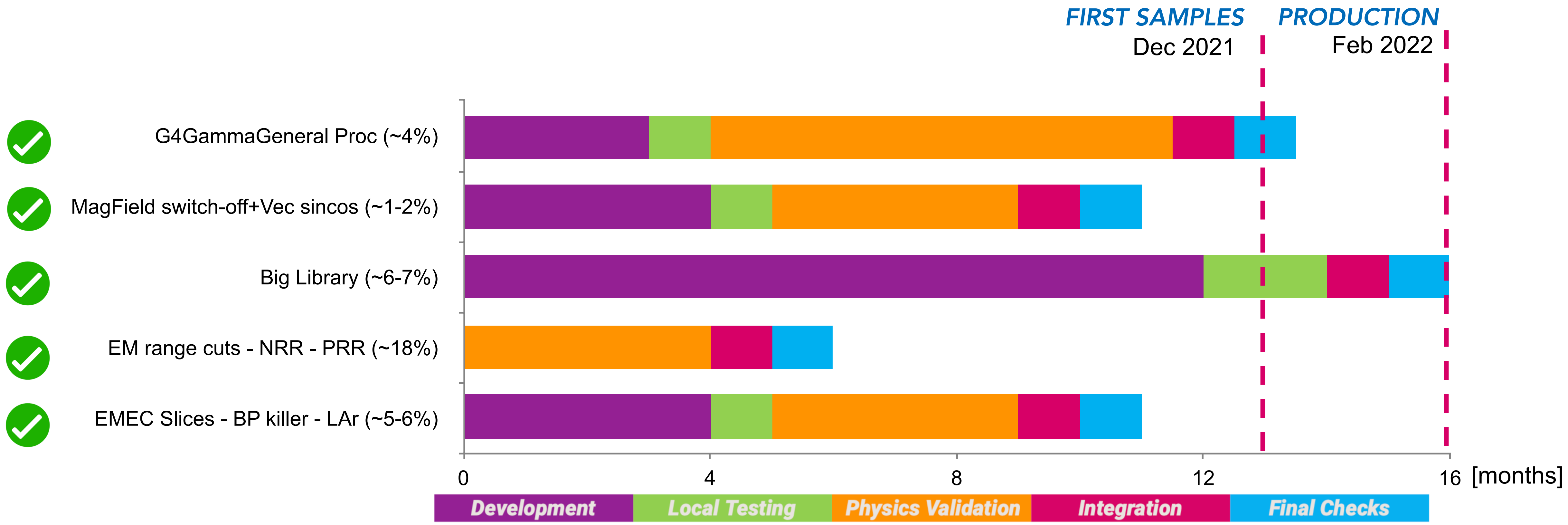
$$r(\vec{x}) = \frac{p(\vec{x}\,|\,\theta_p)}{q(\vec{x}\,|\,\theta_q)}$$

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



V. Kourlitis link

Geant4 Optimization task force



Improvements achieved for the start of Run3 MC production campaign