

CMS Full Simulation for Run 3

Vladimir Ivanchenko^{1,2,*}, *Sunanda Banerjee*³, *Gabrielle Hugo*², *Sergio Lo Meo*^{4,5}, *Ianna Osborne*⁶, *Kevin Pedro*³, *Danilo Piparo*², *Dmitry Sorokin*⁷, *Phat Srimanobhas*⁸, and *Carl Vuosalo*⁹ on behalf of the CMS Collaboration

¹Tomsk State University, 634050 Tomsk, Russia

²CERN, CH1211, Geneva 23, Switzerland

³Fermi National Accelerator Laboratory, Batavia, IL, USA

⁴ENEA, Bologna, Italy

⁵INFN Sezione di Bologna, Italy

⁶Princeton Institute for Computational Science & Engineering, Princeton University, New Jersey, USA

⁷Moscow Physics Technical University, Russia

⁸Dept. of Physics, Faculty of Science, Chulalongkorn University, Thailand

⁹Dept. of Physics, University of Wisconsin-Madison, USA

Abstract. We report the status of the CMS full simulation for Run 3. During the long shutdown of the LHC a significant update has been introduced to the CMS code for simulation. The CMS geometry description is reviewed. Several important modifications were needed. CMS detector description software is migrated to the DD4Hep community developed tool. We will report on our experience obtained during the process of this migration. Geant4 10.7 is the CMS choice for Run 3 simulation productions. We will discuss arguments for this choice, the strategy of adaptation of a new Geant4 version, and will report on the physics performance of the CMS simulation. A special Geant4 Physics List configuration FTFP_BERT_EMM will be described, which provides a compromise between simulation accuracy and CPU performance. A significant fraction of time for simulation of CMS events is spent on tracking of charged particles in a magnetic field. In the CMS simulation a dynamic choice of Geant4 parameters for tracking in field is implemented. A new method is introduced into simulation of electromagnetic components of hadronic showers in the electromagnetic calorimeter of CMS. For low-energy electrons and positrons a parametrization of GFlash type is applied. Results of tests of this method will be discussed. In summary, we expect about 25% speedup of the CMS simulation production for Run 3 compared to the Run 2 simulations.

1 Introduction

The CMS full simulation scheme for Run 2 was described in our previous reports [1-3]. It is based on the Geant4 toolkit [4-6] and multi-threaded software framework used for CMS

* Corresponding author: Vladimir.Ivanchenko@cern.ch

software (CMSSW). The Geant4 physics configuration was optimized for the CMS simulation production [7]. During long shutdown two (LS2) at LHC, significant modifications of simulation software were required to provide effective simulation for Run 3, which is expected to start at the beginning of 2022. Simulation for Run 3 should be stable during analysis and the Geant4 version for Run 3 should not be changed. Considering this, significant efforts were provided for validation of Geant4 10.7 [8]. In parallel, several methods to speed-up simulation production were implemented and will be described below. These developments are also important in view of the long-term high luminosity HL-LHC program.

2 CMS geometry update

Several modifications on top of the Run 2 geometry are implemented for Run 3 (Fig.1) including a description of the new beam-pipe, a more accurate description of support elements for the tracker, an improved description of the muon system, and few other improvements. These updates concern mainly elements of construction and no changes to sensitive detectors, except for a description of a few new muon GEM chambers [9]. An additional very forward sub-detector for scattered protons (PPS [10]) was added to the CMS general description. The subdetector consists of 4 Roman Pot stations at 214 m and 220 m from the CMS interaction point.

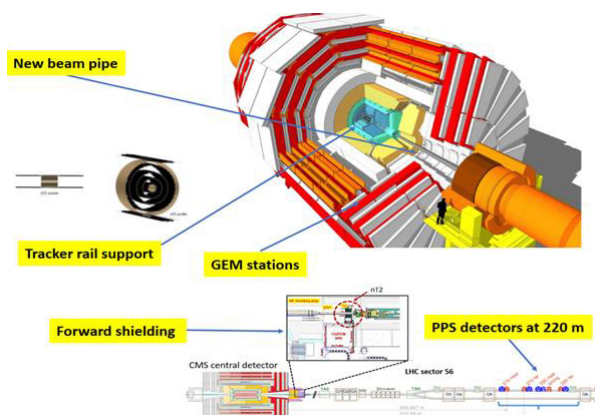


Fig. 1. Updated and new components of CMS for Run 3.

3 Migration to DD4Hep

During LS2 the CMS detector description has been migrated to the DD4Hep community developed tool [11, 12]. One of the advantages of the tool, besides its future sustainability, is a more flexible description of simulation and reconstruction geometries. Developments of the Phase-2 detectors of CMS require frequent modifications of a particular detector geometry before final choice of the configuration is made. One of the goals of the DD4Hep migration was to review and update the CMSSW sub-libraries for geometry description, making these classes simpler and more compact. Such an improvement facilitates long-term support of the CMS geometry description. During this migration, XML files describing the CMS geometry were verified and improved, for example, all geometry parameters have now explicit units. The strategy of the migration is the following: first providing the DD4Hep geometry for Run 3, after that the implementation of the DD4Hep geometry for Phase-2. Two geometry descriptions (old CMS native and the new DD4Hep based) will co-

exist for some time until there is evidence that DD4Hep provides full functionalities for all possible tasks within CMS. The process of adapting of DD4Hep for CMS allows also for improving interfaces and performance of the DD4Hep software. Finally, the Run 3 geometry construction requires the same CPU time for the old and new descriptions. This was achieved after optimisations introduced in both CMS code and the DD4Hep software. Checks on overlaps for simulation geometry show a significantly reduced number of overlaps in the DD4Hep geometry. In the Run 2 geometry, there were ~ 1000 tiny overlaps below 0.1 mm, all are between passive elements. With the current DD4Hep geometry, there are only 8 such overlaps.

4 Adaptation of Geant4 version 10.7

For Run 2 Monte Carlo production Geant4 version 10.4p03 [6] was used and about 100 billion events were produced. The Geant4 Collaboration provided the new version of the toolkit, 10.7 (December 2020), which is the last version of the series 10, which includes technical improvements developed during the seven years. It also includes the high-fidelity physics models provided by the Geant4 team and it is the fastest version of the series. The Geant4 Collaboration offers long term support for Geant4 10.7, preparation of patches and bug fixes. As the first step toward to the most advanced version, Geant4 10.6p02 was adopted as the default version for CMSSW in 2020.

To integrate Geant4 10.7 we continued the approach developed for Run 2 [7], the adaptation was started with the beta version (June 2020). Each new monthly reference version of Geant4 was integrated into CMSSW in a special git branch allowing validation of Geant4 physics. This approach allowed for necessary preparations in CMSSW sub-libraries in advance and for providing feedback to the Geant4 team. The validation demonstrates stable results, and it is possible to conclude that Geant4 10.7p01 and 10.6p02 predictions are very close to each other. For the same Run 3 workflows Geant4 10.6p02 is about 5% faster than 10.4p03, and 10.7p01 is about 3% faster than 10.6p02.

For more effective monitoring of CPU performance of the simulation, the initialisation of Geant4 physics per thread in CMSSW was moved from the event loop to the initialisation phase. This allows for CPU performance measurements using only 200 events (Fig.2).

5 Configuration of Geant4 physics

With Geant4 10.7, additional new features are available. In CMSSW extra physics configuration options are available for the Run 3 Monte Carlo:

- Gamma general process - a method to handle all gamma processes together [13], depending on setup it may bring few percent CPU speed-up.
- Energy limit for muon propagation used to stop the tracking of a muon; if energy transfer exceeds this limit, the track is killed, is added to the list of secondary particles with a new track number, and a new vertex is created.
- Optionally enable the simulation of rare processes such as muon pair production by gamma conversion or annihilation of an ultra-relativistic positron with an electron of a media producing a pair of muons or hadrons.
- Data for nuclear gamma levels are uploaded at initialisation stage, not in the event loop as it was for Run 2 simulation.
- Time consuming initialisation of the data structure for muon induced e^+e^- pair production may be performed using data files.

Part of these features may be used in the mainstream simulation production; other will be used for special simulations.

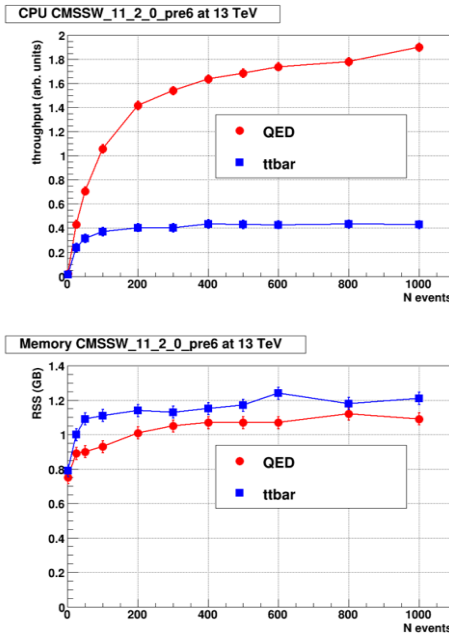


Fig. 2. Event throughput (top) and RSS memory (bottom) for MinBias (red) and ttbar (blue) as a function of number of simulated events.

Several Geant4 physics lists were developed specifically for CMS:

- FTFP_BERT_EMM – the CMS default.
- QGSP_FTFP_BERT_EML – alternative physics for cross-check of high energy hadron response.
- FTFP_BERT_EMN – physics list with the accurate simulation of electromagnetic processes for Phase 2 test-beam studies.
- FTFP_BERT_XS_EML and FTFP_BERT_HP_EML – physics lists for neutron background studies.

In all these physics lists, the configuration of pion-nuclear interaction is different from Geant4 recommendations: the Bertini cascade model is used from 3 GeV to 12 GeV. Electromagnetic physics options of FTFP_BERT_EMM different from the Geant4 default are the following: cuts are applied on secondaries from gamma processes and “simple” limitation of step of e^+ and e^- due to multiple scattering everywhere except in the hadronic calorimeter.

The integration of charged particle trajectories in a magnetic field was optimised. It was shown in [3] that geometry and tracking in magnetic field accounts for more than 50% of the CPU for the CMS simulation. The Geant4 default integrator is *G4DormandPrince745* since version 10.6. It is used in CMSSW as the default. There are diverse requirements for tracking in field depending on particle energy and detector region. In the volume occupied by the silicon tracker, it is necessary to obtain a precision of track intersections with the sensors better than 1 micron. Also, high accuracy is needed for muon chambers. In all other regions where the field itself is not measured as accurately; it is not necessary to satisfy such a stringent requirement. The solution identified consists of the dynamic change of Geant4 tracking parameters. Before each Geant4 step, a check of energy, detector region,

and particle type is performed. For all other particles but the magnetic monopole, three sets of parameters are applied (Table 1):

- Set 1 - high accuracy in central detector region $R < 8$ m, $|Z| < 11$ m, for energetic particles $E > 200$ MeV.
- Set 2 - low accuracy for low-energy particles $E < 15$ MeV.
- Set 3 - medium accuracy for the rest.

In the Run 2 simulation there are 2 sets of parameters. The Run 3 approach provides improved accuracy for relativistic particles without loss of CPU speed. The parameter set 2 is needed to avoid problems for tracking of low-energy particles in a non-uniform magnetic field.

Table 1. Three sets of parameters for Geant4 tracking in field.

Geant4 parameters	Parameter set 1	Parameter set 2	Parameter set 3
DeltaIntersection (mm)	10^{-6}	0.01	10^{-4}
DeltaOneStep (mm)	10^{-4}	0.1	10^{-3}
DeltaChord (mm)	10^{-3}	0.1	$2 \cdot 10^{-3}$
MaxStep (cm)	150	50	150

6 Fast parameterisation of electromagnetic showers

A new method (GFElowE) has been developed to speed up simulation of hadronic showers inside the CMS electromagnetic calorimeter (ECAL). The idea of GFElowE is to parameterize the sampling of hits from low energy e^+ and e^- . The ECAL is built from $PbWO_4$ crystals, and so particles are transported in a medium which is substantially uniform. This allows us to make a relatively simple parameterisation.

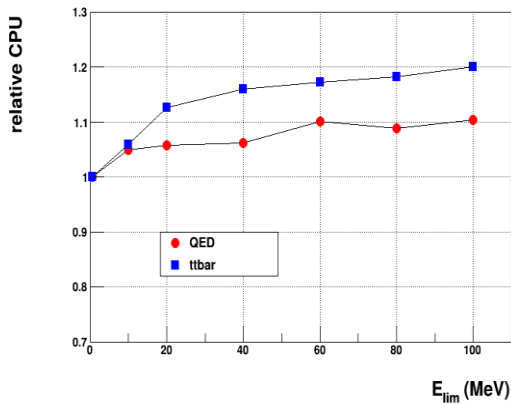


Fig. 3. Relative CPU improvement as a function of energy threshold for the GFElowE method for MinBias and ttbar for CMS simulation.

The GFElowE parameterisation is applied to e^+ or e^- , when their energy at a certain simulation step becomes lower than a threshold referred as to E_{lim} . The parameterisation includes three functions: $S(E)$ – fraction of energy delivered at point, $R(E)$ – radial distribution around the track direction, $Z(E)$ – longitudinal energy distribution. During the sampling, the energy $E \cdot S(E)$ delivered at the initial point, remaining $E \cdot (1 - S(E))$ is distributed in the space round the track according to $R(E)$ and $Z(E)$. The effect on the CPU usage as a function of E_{lim} is shown in Fig.3. From this plot it is possible to conclude that the limit 20 MeV provides a speed-up between 20% and 10%, depending on the process.

Validations of GFElowE were performed using simulated data with $Z \rightarrow e^+e^-$ and $t\bar{t}$ events. It was shown that parameters of high energy electromagnetic showers are affected. Similar effects are much smaller for hadronic showers. As the working variant, the GFElowE approach is applied only on hadronic showers for Run 3. In the simulation of showers produced by primary gamma, e^+ , or e^- from the interaction region this method is not used.

7 Summary

The high-fidelity simulation plan of CMS for Run 3 has been reviewed. The geometry description of the CMS experiment was migrated to the DD4Hep tool. The major innovation is the adoption of Geant4 10.7 as the baseline release of the Geant4 toolkit. On top of it the CMS specific configurations of physics models for the simulation are implemented. Due to migration to Geant4 10.7 the Run 3 simulation production will become 8% faster than that of Run 2 without any loss of physics performance. Considering the application of the GFElowE parameterisation of low-energy electrons and positrons inside ECAL, we expect the Run 3 simulation production become even faster.

References

1. D.J. Lange et al., J. Phys.: Conf. Ser. **608**, 012056 (2015)
2. M. Hildreth et al., J. Phys.: Conf. Ser. **664**, 072022 (2015)
3. M. Hildreth et al., J. Phys.: Conf. Series **898**, 042040 (2017)
4. The Geant4 Collaboration (S. Agostinelli et al.), Nucl. Instr. Meth. A **506**, 250-303 (2003)
5. J. Allison et al., IEEE Trans. Nucl. Sci. **53**, 270-278 (2006)
6. J. Allison et al., Nucl. Instr. Meth. A **835**, 186-225 (2016)
7. V. Ivanchenko and S. Banerjee, EPJ Web of Conferences **214**, 02012 (2019)
8. <https://geant4.web.cern.ch/support/download>
9. D. Abbaneo et al., Nucl. Instr. Meth. A **918**, 67-75 (2019)
10. M.M. Obertino, JINST **15**, C05049 (2020)
11. F. Gaede et al., EPJ Web of Conferences **245**, 02004 (2020)
12. C. Vuosalo et al., EPJ Web of Conferences **245**, 02032 (2020)
13. V. Ivanchenko et al., EPJ Web of Conferences **245**, 02009 (2020)