

PAPER • OPEN ACCESS

Upgrades for the CMS simulation

To cite this article: M Hildreth *et al* 2017 *J. Phys.: Conf. Ser.* **898** 042040

View the [article online](#) for updates and enhancements.

Related content

- [Readiness of CMS simulation towards LHC startup](#)
S Banerjee
- [CMS reconstruction improvements for the tracking in large pileup events](#)
M Rovere
- [Fast emulation of track reconstruction in the CMS simulation](#)
Matthias Komm and CMS Collaboration

Upgrades for the CMS simulation

M Hildreth^{1,2}, V N Ivanchenko^{3,4,5} and D J Lange⁶ for the CMS Collaboration

¹ Université de Notre Dame du Lac, Notre Dame, IN, USA

² FNAL, Batavia, IL, USA

³ CERN, CH1211 Geneva 23, Switzerland

⁴ Geant4 Associates International Ltd, United Kingdom

⁵ Tomsk State University, Tomsk, Russia

⁶ Princeton University, Princeton, NJ, USA

E-mail: Vladimir.Ivantchenko@cern.ch

Abstract. We report the current status of the full simulation application developed by the CMS experiment. For LHC run-II, CMS is using Geant4 10.0p02 built in sequential mode. About 16 billion events were produced for run-II analysis during 2015-2016. A new method to handle pileup events has been developed and is now used for regular processing of simulation events. We plan to use Geant4 10.2p02 for 2017 production. In this work, we present CPU and memory performance of the CMS full simulation for different configurations and Geant4 versions considered during our testing. We also discuss technical aspects of the migration to Geant4 10.2 and a new premixing scheme for the simulation of pileup.

1. CMS simulation

The CMS simulation software [1-3] includes several components (figure 1). The generation of primary physics processes is performed using PYTHIA and other physics generators, and provides an input for the Geant4 [4], [5] simulation of particle transport in the detector and hit creation for digitization. Hits and Monte Carlo (MC) truth information are stored in the ROOT [6] persistent format. This first step of simulation is CPU demanding, so it is important to do it only once for a given generator, detector and beam-parameter configuration. All calibration, alignment corrections, inefficiency corrections, and pileup are applied at the next step, which is the simulation of the digitization process (“digi”). The output of this digitization step has the same format as real data, except it includes the additional MC truth information. Further, MC and data are processed by the same software; usually digitization and reconstruction are done together in the CMS processes model. The same initial geometry description is used to produce Geant4 geometry and reconstruction geometry formats. In this work, we will describe only the current status of hit and digi simulation for run-II and modifications prepared for continuation of run-II in 2017-2018. We will focus mainly on software aspects of these developments.

2. CMS simulation production for run-II

For run-II simulation hits production Geant4 10.0p02 was prepared [2], [3] at the time of the long LHC shutdown. During the preparation phase a number of technical fixes from the intermediate Geant4 version 10.1 were included in the CMS version of Geant4 in order to improve performance for large scale production. Multithreaded (MT) mode of simulation was already available with this version of Geant4 and implemented in CMS software (CMSSW) but not yet used for massive production. Overall,



the simulation speed in 2015-2016 is approximately 2 times faster than that of run-I. About 16 billion simulated events were produced so far during CMS run-II.

3. CMS hits simulation for 2017

For 2017 operations, the simulation application has been prepared to handle several modifications in the CMS hardware. First of all, a new pixel tracking detector, which provides more sensitive layers within the same solid angle, has been included. Additional muon chambers have been added and electronics of the hadronic end-cap calorimeter is updated. Because of these significant detector changes, previously produced simulated events cannot be used for analysis of the data obtained with the updated detector. Because the simulation should be redone for the new detector configuration, it was a good opportunity to add necessary modifications in all aspects of CMSSW including hit production and digitization.

CMS Monte Carlo Simulation approach

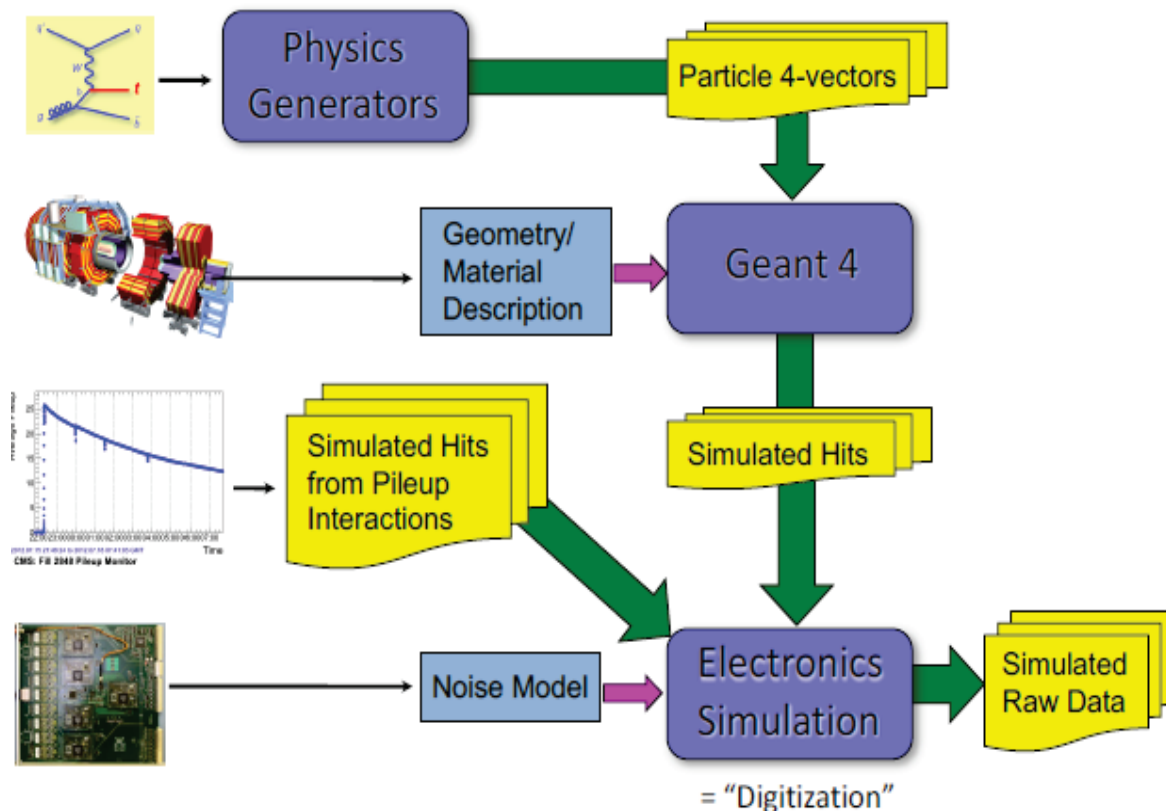


Figure 1. CMS simulation workflows.

3.1 Adaptation of Geant4 10.2

There was substantial evolution of Geant4 code after the release 10.0 where the MT mode was introduced for the first time. The CMS choice for the 2017-2018 production is the Geant4 version 10.2.p02 [5]. This choice was made based on the following main arguments:

- Geant4 10.2 is a consolidation release which includes many fixes and improvements compared to the initial version 10.0 and the following 10.1;

- It is fully c++11 compatible and compiles with recent gcc compilers adopted for the rest of CMSSW (specifically gcc5.3 and gcc6.2);
- There are many fixes in the geometry, transportation in magnetic fields, and physics sub-libraries of Geant4 which are relevant for CMS;
- It is expected to be more performant.

A slightly modified version of Geant4 10.2p02 was successfully integrated into CMSSW. The main modification was to use parameters of the FTFP model [5] from the Geant4 10.1, which provides similar calorimeter response to a high-energy hadronic shower as in previous run-II simulation production. It has now passed CMS validation procedure and physics validation versus the CMS test beam 2006 [7] and versus run-II data. As a result of these efforts [8], a new Geant4 Physics List was chosen as a CMS default: FTFP_BERT_EMM. The previous CMS default was QGSP_FTFP_BERT_EML. The main two differences between these two are:

- The Geant4 string model QGSP is replaced by the FTFP model recommended by the Geant4 collaboration [5];
- The CMS custom configuration for electromagnetic physics (EML) is changed to another one (EMM), in which the default Geant4 electromagnetic configuration is applied instead of EML within hadronic calorimeter, resulting in an improved simulation [8] of the calorimeter response.

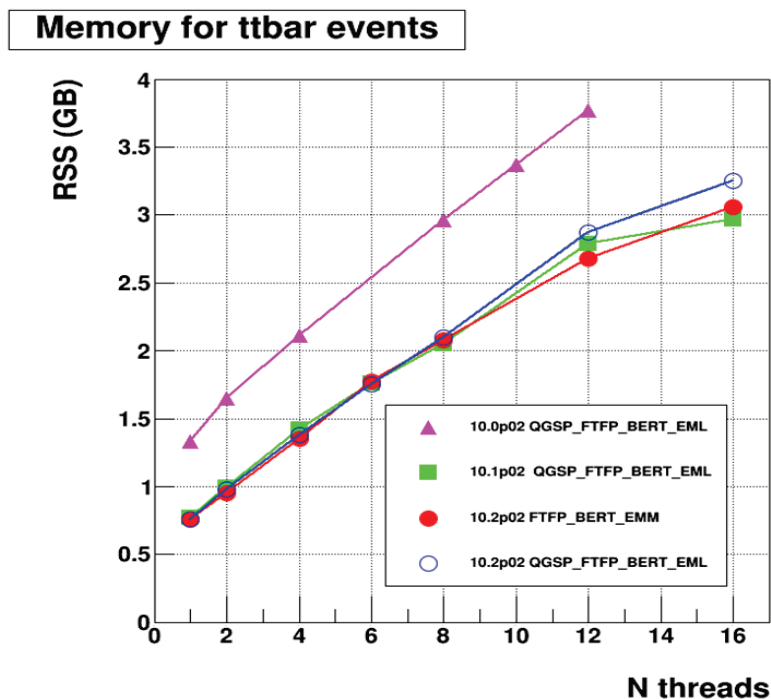


Figure 2. RSS memory for simulation of hard scattered events as a function of number of threads in the MT mode after 1000 events.

3.2 CPU and memory performance

For study of CPU and memory performance for the CMS hits simulation a typical processing node was used with 12 Intel processors. Geant4 and CMSSW were compiled with the current CMS default compiler gcc.5.3.0. Both single threading and MT modes were exercised. For the RSS memory check (figure 2) *ttbar* events were used and total RSS values were checked after 1000 events. It is clear from the plot that the main memory reduction was achieved with Geant4 10.1 and the RSS memory does not

depend on the selected Geant4 physics list. Memory usage depends linearly on the number of threads and substantial savings are observed (table 1). The main contributions to the memory are from CMS geometry and Geant4 voxelisation of this geometry, from CMS hits data structure, and from CMS framework software. The current memory reduction for one thread is mainly due to CMSSW improvements, while delta per thread reduction is mainly due to the usage of the new Geant4 version.

Table 1. RSS memory used by hard scattered events per thread.

<i>Release</i>	<i>1st thread</i>	<i>Delta per thread</i>
10.0p02	1.33 GB	0.23 GB
10.2p02	0.76 GB	0.19 GB

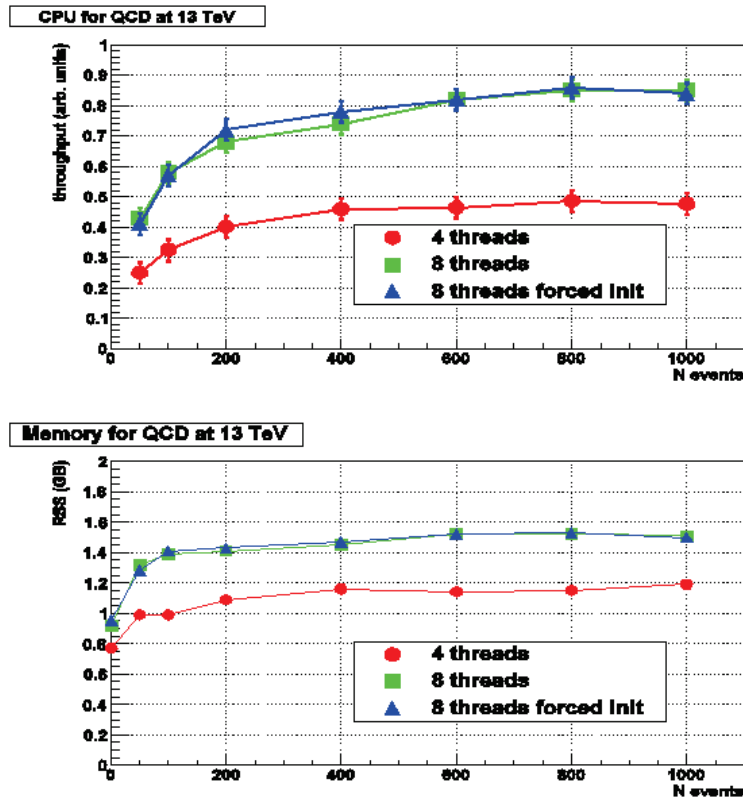


Figure 3. RSS memory (bottom) and event throughput (top) for simulation of QCD events with Geant4 10.2p02 as a function of the event number for 4 and 8 threaded runs.

For control on CPU usage, we studied event throughput as a function of number of produced events (figure 3). This allowed us to separate out initialization effects, as both Geant4 and CMSSW initialization require non-negligible CPU time. Geant4 hadronic models use the lazy initialization method to download only the data that is used in the simulation processing. Due to this, both memory and CPU per event grow with the event number and reach a plateau after ~ 500 events. Even if reading in all Geant4 hadronic physics data was forced for the Geant4 initialization, the results are practically the same. The same plots were obtained for different primary generators. To exclude effects of

initialization, we compare CPU performance for different Geant4 versions using runs with 600 events (table 2). To make comparisons in the same conditions as an 8-thread run in the MT mode, for a sequential run 8 jobs were submitted simultaneously and the result was averaged. Correspondingly, for the 4-threads MT run two jobs were submitted for comparison and averaged. Thus, CPU measurements were performed with the same load of the node. 8-thread runs were performed because such regime of production at the grid is planned for CMS, measurements were repeated several times to confirm stability of the results.

There is practically no significant CPU improvement with Geant4 10.2 compared to 10.0. A small speedup is observed for hard scattered events and some slowdown for QCD events. These results show that there is an additional CPU overhead due to the Geant4 and CMSSW multithread frameworks [9]. This overhead is independent of the event type, so the relative slowdown is larger for QCD events. Note that part of the Geant4 performance improvements were backported from Geant4 10.1 to the CMSSW production version as patches on top of the Geant4 version 10.0p02. CPU profiling performed using the *igprof* tool confirms the result reported for previous version [5]: 50-60 % of CPU for CMS hits simulation is spent for Geant4 code performing geometry navigation and tracking in the magnetic field. This fact should be taken into account for future optimisation of CMS simulation.

Table 2. CPU efficiency (event/thread/time) for different Geant4 and CMSSW configurations normalised to *tbar* value of 2015-2016 run-II production.

<i>Geant4 configuration</i>	<i>Number of threads</i>	<i>throughput for tbar events</i>	<i>throughput for QCD events</i>
10.0p02 seq.	1	1	6.0
10.2p02 MT	1	1.07	5.7
10.2p02 MT	4	1.06	5.1
10.2p02 MT	8	1.05	5.0

4. CMS digi simulation for 2017

For more effective digitization simulation a “premixing” approach for CMS was developed [2] and is now used for re-processing the existing run-II simulation. This approach means that when the pileup distribution for a certain running period is known (or else estimated), a digitized sample of pure pileup events (“premixed sample”) is prepared using QCD simulated events, including the effects of in-time and out-of-time interactions in the CMS detector. This prepared sample of pileup events is then reused for digitization of all types of hard scattered events. To make this method fully functional it was necessary to extend the raw format of CMS hits to ensure sufficient precision for making sums of small pulse heights in the digi step. This format saves approximately 90 % of input/output operations relative to the full pileup simulation used previously.

The premixing method brought a substantial operational improvements including the large reduction in I/O mentioned above. In addition, the CPU time required to digitize and reconstruct typical run-II events has been reduced by approximately a factor of two with this premixing approach. This is expected, because most of the digitization of the pileup interactions is performed as the premixed sample is created, and each event in the premixed sample is reused several times in the reconstruction of physics events. This event reuse does introduce statistical considerations. To ensure good statistical properties of the premixing library, it should be large enough and should be bigger if mean pileup will be increased. In 2016 the mean pileup was about 25, a corresponding sample of 200 million premixed events was produced and used for simulation production with various type of hard scattered events. This provides an acceptable level of QCD events re-use.

5. Conclusions

The CMS simulation for run-II has performed well in 2015-2016. We have produced samples of about 16 billion simulated events so far. For the 2017-2018 continuation of run-II, a number of improvements are introduced for the CMS simulation, including Geant4 10.2 in the MT mode and the premixing method. We expect stable and effective simulation production for the entire run-II data processing.

References

- [1] Banerjee S 2012 *J. Phys: Conf. Ser.* **396** 022003
- [2] Lange D J et al. 2015 *J. Phys: Conf. Ser.* **608** 012056
- [3] Hildreth et al. 2015 *J. Phys: Conf. Ser.* **664** 072022
- [4] The Geant4 Collaboration (Agostinelli S et al.) 2003 *Nucl. Instr. Meth. A* **506** 250-303
- [5] Allison J et al. 2016 *Nucl. Instr. Meth. A* **835** 186-225
- [6] Antcheva I et al. 2009 *Comp. Phys. Comm.* **180** 2499-2512
- [7] Abdullin S et al 2010 Calorimetry Task Force Report CMS-NOTE-2010-007; CERN-CMS-NOTE-2010-007 Geneva CERN 25 pp.
- [8] Banerjee S 2017 Validation of physics models of Geant4 using data from CMS experiment, this proceedings.
- [9] Jones C 2015 *J. Phys: Conf. Ser.* **664** 072026