Upgrades for the CMS simulation

D J Lange[1], M Hildreth[2], V N Ivantchenko[3], and I Osborne[4] for the CMS Collaboration

[1]Lawrence Livermore National Laboratory, Livermore, Ca USA 94550.
[2]University of Notre Dame, Notre Dame, In, 64556, USA
[3]CERN, Geneva, Switzerland.
[4]Fermi National Laboratory, Batavia, Il USA 60510.

Abstract. Over the past several years, the CMS experiment has made significant changes to its detector simulation application. The geometry has been generalized to include modifications being made to the CMS detector for 2015 operations, as well as model improvements to the simulation geometry of the current CMS detector and the implementation of a number of approved and possible future detector configurations. These include both completely new tracker and calorimetry systems. We have completed the transition to Geant4 version 10, we have made significant progress in reducing the CPU resources required to run our Geant4 simulation. These have been achieved through both technical improvements and through numerical techniques. Substantial speed improvements have been achieved without changing the physics validation benchmarks that the experiment uses to validate our simulation application for use in production. In this presentation, we will discuss the methods that we implemented and the corresponding demonstrated performance improvements deployed for our 2015 simulation application.

1. Introduction

The increasing beam energy and pileup conditions of the LHC will bring new challenges to the simulation program of the CMS experiment for the next decade. Given both short and long term upgrades to the accelerator and experimental apparatus, there are significant challenges to be faced to ensure good physics performance of the experiment for both today and tomorrow. In particular, higher LHC beam energy means more complex interactions, a larger number of interactions per bunch crossing (higher pileup) which are more time consuming to simulate and reconstruct. At the same time, CMS plans a much higher output rate of trigger (~1 kHz, or 2.5x higher than in Run 1), which means a demand for correspondingly larger samples of simulated events. In the longer term, simulating a potential CMS detector for the high-luminosity LHC program (HL-LHC) has additional challenges including a still to be determined CMS detector configuration, a higher output rate of trigger (potentially 10 kHz), and further increases in luminosity and pileup conditions (i.e., more than 140 interactions per crossing).

These proceedings discuss a number of recent changes made to the CMS simulation application to allow CMS to continue to rely on a high-fidelity detector simulation as the basis for its analysis and detector optimization work. The detector response for the primary CMS simulation application is performed using Geant4 [1]. Our recent work has brought substantial technical performance improvements without changing the physics accuracy of our simulation.

2. CMS Detector upgrades

The CMS experiment is working on instrumentation changes on three different timescales. The most immediate detector upgrades are being implemented during the on-going LHC shutdown period and will be fully completed before the start of the Run 2 data collection. The most significant changes are to replace the central region of the beam pipe, to complete the muon detector coverage as it was originally designed, and to update the electronics in part of the hadronic calorimeter.

The "Phase-I" upgrades begin in 2017 and continue through the next long LHC downtime in 2018 [2,3,4]. The Phase-I upgrade will include a new pixel detector installed during the annual shutdown between 2016 and 2017, and more extensive changes during the LHC technical stop. These include substantial changes to the hadronic calorimeter electronics and Level-1 trigger.

Finally, the proposed "Phase-II" upgrades of CMS are major detector improvements for operation in the high luminosity LHC (HL-LHC) conditions. HL-LHC presents increased challenges for triggering, tracking and calorimetry, in particular for low to medium P_T objects. As summarized in Figure 1, CMS is presently working on a Technical Proposal including a tracker replacement and extensive forward calorimetry and muon improvements. These extensive changes require a flexible simulation program. While physics results can be derived using a parameterized simulation, a full detector simulation is invaluable for doing validation of any fast simulation, in particular for a beam intensity regime that is very different from that of the LHC Run 1.

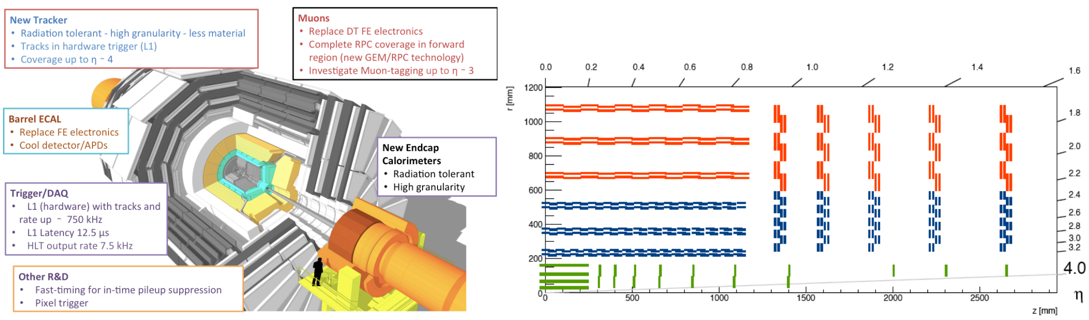

Figure 1: Summary of CMS detector upgrade concepts for Phase II (left), and cross section of upgraded tracker detector showing simulated hits (right).

3. Changes to CMS simulation for 2015

CMS has made a number of changes in our simulation application and infrastructure to facilitate both Run 2 and long-term upgrade work within the experiment. Users for both applications rely on a full Geant4-based detector simulation for reliable results.

An important aspect of this work is an underlying geometry model that is very flexible [5]. We are able to simultaneously support numerous geometries and sub-detector developers can easily update their XML-based geometry definitions as detector requirements are updated.

Goals of our recent work include achieving significantly better throughput (reduced time to simulate each event), easier configuration management in order to support additional detectors, and to enable very high pileup simulation within the CMS computing resource constraints. A number of projects have been undertaken over the course of several years in anticipation of the Run 2 requirements for larger simulation samples and for more complex events.

This Section describes our approach and the results of our work in these areas. We focus primarily on changes made to achieve large gains in throughput, and the significant changes we have implemented in our pileup simulation.

3.1. Monte Carlo sampling techniques: Russian Roulette

A significant portion of the CMS simulation time is due the tracking of low-energy particles, particularly in the electromagnetic and hadronic calorimeter systems. We investigated and have now implemented a sampling technique used widely in Monte Carlo applications such as neutron shielding calculations ("Russian Roulette") that significantly reduces the time spent tracking these particles [6]. Specifically, we changed our simulation to track only a small fraction of gamma and neutron particles below a threshold energy (5 and 10 MeV, respectively) where the fraction and threshold were tuned such that the final physics output of our simulation was not affected. We found that it was necessarily to have sampling factors and thresholds that depend on both detector region and particle type.

Figure 2 illustrates the technique for a particle interaction. In this example, only one of the six low-energy particles in the shower resulting from the original particle is tracked using Geant. The energy deposits from this particle and its daughters are given an additional weight, W, corresponding to the inverse of the fraction of particles kept. We found that W=0.1 for neutrons was an adequate compromise between loss of physics performance and improvement in CPU performance.

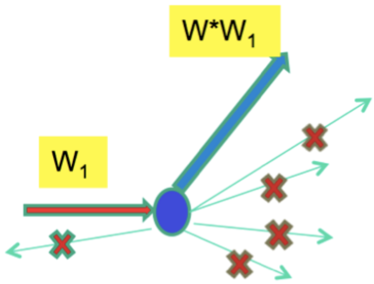

Figure 2: Russian Roulette sampling technique.

The performance improvement we observe from the application of this technique depends on the physics channel being simulated, but is typically between 25% and 40%. Our validation process showed that the energy and shower shape response in our high-resolution barrel electromagnetic calorimeter were the most sensitive to parameter tuning of the Russian Roulette technique.

3.2. Library packaging

One purely technical improvement to the CMS simulation is that we have repackaged all shared libraries within our software that depend on Geant4 into a single static library. We also packaged all of the Geant4 libraries that we use into a single static library. These changes allow us to hide Geant4 from the rest of the CMSSW code base, and subsequently to evaluate the benefit of more aggressive linker options than used for the rest of CMSSW. The result of these investigations was that adding "-*flto -Wl,--exclude-libs,ALL*" to our linker options gave the best performance improvement, providing approximately an 10% reduction in the CPU needed per event with no detrimental impact on memory or other constrained resources.

This technique does impose a number of constraints. While these are not a large burden for simulation developments, they do need to be considered before extending this technique to more of the CMSSW codebase.

First, we must control dependencies to use Geant4 only within this single library. This is relatively straightforward for CMSSW code that depends on Geant4, as only a small fraction of our libraries (<2%) depend on it. Second, there is some impact on simulation developers due to the creation of this single large library. We have minimized this impact by keeping the original shared libraries cached in release. Therefore, any developer needing to rebuild one of the affected libraries just incurs an extra step of rebuilding the static library rather than all of the packages that depend on Geant4.

In limited experience so far, we have had no problems controlling dependencies that would impact the effectiveness of this single static library. More experience will allow us to better understand any operational overhead that we have introduced.

3.3. Summary of improvements

The use of Monte Carlo sampling techniques and other purely technical changes in code optimization, together with the integration of Geant4.10, have provided a total reduction of about 50% in the CPU per event needed for the CMS simulation at 13 TeV within only the past one year of development. This improvement has been achieved without introducing a reduction in physics accuracy in our detector simulation. The largest components are the Russian Roulette technique and other CMS simulation code optimizations. We continue to work on other optimizations as we anticipate the need to further reduce the computing resources needed for our simulation application in the future.

4. Pileup Simulation

An import component of the CMS Monte Carlo simulation is the modeling of the minimum bias pileup events, both in the bunch crossing of the hard-scatter event simulated and in the nearby out-of-time crossings. Here we describe our original approach and recent developments undertaken to improve performance at high pileup.

4.1. Run 1 approach

CMS has always simulated pileup interactions by separately generating simulated samples of hard-scatter interactions and minimum bias events for pileup. We have employed this approach both to save processing time (by using events in the minimum bias samples more than one time) and to increase our flexibility to change the distribution of the number of pileup events much after the simulation samples themselves are generated. For example, this approach allows us to change the pileup distribution during a re-digitiziation and re-reconstruciton pass through the simulation samples.

Figure 3 shows the pileup simulation approach. The combination is done at the "SimHit" level as the input to the electronics simulation step. These SimHits corresponds to the full information saved from the Geant4 detector simulation.

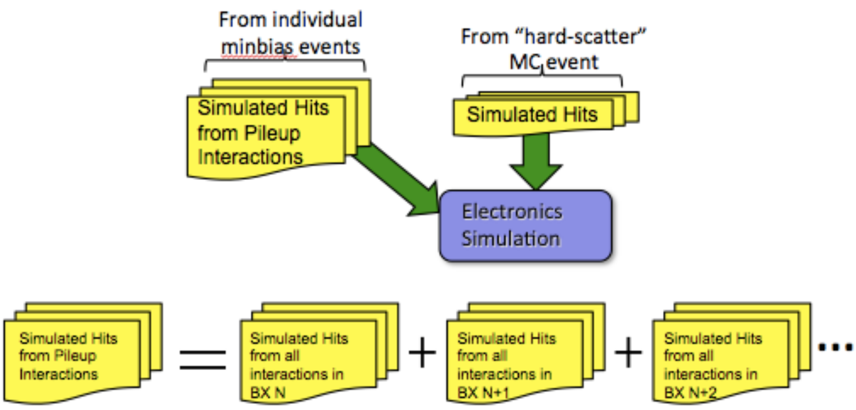

Figure 3 Original pileup simulation approach.

The figure shows the critical limitation of this approach. All SimHits from all pileup interactions (from all 16 bunch crossings that we model) are loaded into memory simultaneously. Given practical limits on the memory footprint of the simulation application, this limits our pileup simulation essentially the Run 1 operating conditions.

4.2. Addressing the memory footprint problem

Given expectations for higher pileup in both Run 2 and upgrade simulations, we refactored the pileup simulation to process each interaction sequentially, as shown in Figure 4. This work required a substantial rewrite of the digitization code and the re-organization of internal event processing framework.

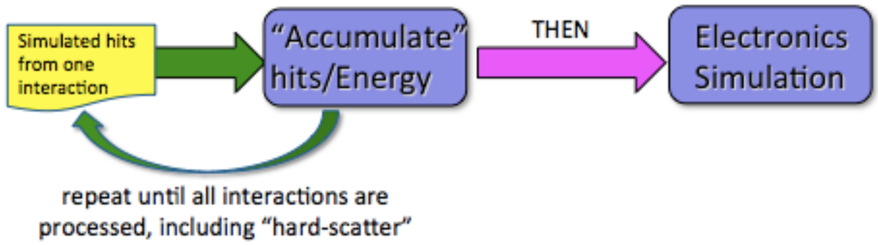

Figure 4: New pileup simulation approach.

The benefit of this new approach is that the content of individual interaction events is dropped once they have been processed so that only one event is in memory at any given time. This means that essentially an arbitrarily number of pileup events can be included in the digitization within the memory footprint limits of our computing resources. This work has been successfully deployed and used in a number of simulation production campaigns (primarily for the upgrade program).

4.3. Addressing the I/O problem

While our current pileup-mixing scheme has reduced the memory requirement for this simulation to fit within resource constraints for the foreseeable future, we continue to read a large number of minimum bias event one-by-one to simulate pileup. In particular, we still read more than 2000 minimum bias events to produce a single event with appropriate pileup at HL-LHC luminosities in order to simulate the interactions in all of the bunch crossings needed to accurately model signals processed by the CMS detector.

We have developed a potential solution to the I/O component of the pileup simulation, which we call "pre-mixing". The idea, illustrated in Figure 5, is to create a library of events containing only pileup contributions, following pre-determined luminosity profile to calculate how many interactions to include. The result is saved in a raw data format. The hard-scatter sample is created and processed through the digitization step with no pileup, convert to our raw data format. The final step is then that the two streams are merged

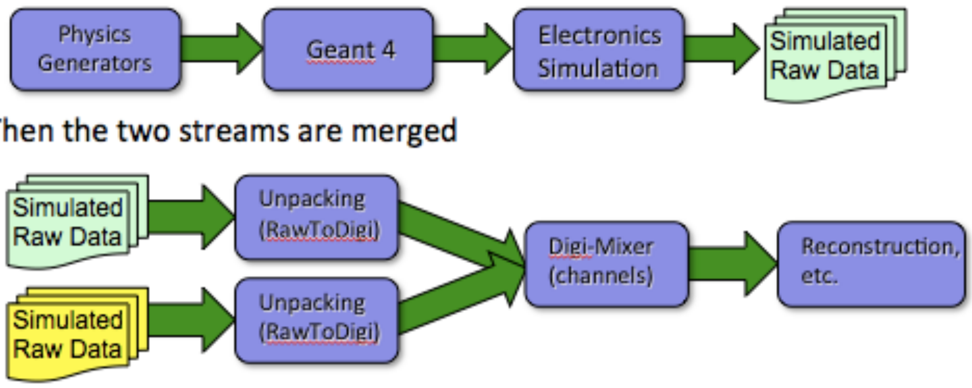

Figure 5: Pre-mixing pileup simulation approach.

The advantage of this approach is that only one pileup event is needed for the pileup simulation of each hard-scatter event. Thus it is much easier to process the pre-mixing approach through computing infrastructure once the initial premixed sample is created.

The initial version of pre-mixing has been implemented and deployed for part of our large 2015 preparation samples. Production tests went smoothly and demonstrated gains in I/O and CPU efficiency at high pileup. We have done an extensive validation against our normal mixing simulation approach. At this point, there are only a few known physics issues left to solve (saturation effects).

Another remaining issue is that we need to take additional care with the re-use of the individual minimum bias events. While our current mixing scheme reuses individual minimum bias events, the pre-mixing scheme can potentially reuse an entire set of minimum bias events, thus enhancing the effect of any important correlations between minimum bias events. We could avoid this entirely by generating one pre-mixed event for each hard-scatter event (e.g., 10 billion), but this would significantly reduce the potential benefit of this approach. We are evaluating the statistical issues with reuse of pileup events, but in any case, care is needed in the Monte Carlo production system to avoid excessive reuse of pre-mixed events.

5. Conclusion

We have summarized a number of areas of recent development work in CMS that have brought a big reduction in simulation resource needs for 2015 even in the face of higher event complexity and trigger rates. CMS detector upgrades push us to use today's software/computing infrastructure to simulate tomorrow's event complexity. In particular, the detector upgrade developments have proven to be an excellent platform for the quick deployment of new simulation features.

References

[1] Agostinelli, S, *et. al.*, "Geant4 – a simulation toolkit", *Nucl. Inst. Meth.* **A506** (2003) 250.

[2] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade", CERN-LHCC-2012-016, (2012).

[3] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", CERN-LHCC-2012-015, (2012).

[4] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", CERN-LHCC-2013-011, (2013).

[5] Osborne I, et al., "CMS geometry through 2020," CHEP 2013.

[6] Kalos, M H, "Importance sampling in Monte Carlo shielding calculations, *Nuci. Sci. Eng.*, **16**, 227, 1963.