Tuning and optimization of the CMS simulation software

Fabio Cossutti, on behalf of the CMS Offline project

Abstract

The CMS simulation has been operational within the new CMS software framework for more than 3 years. While the description of the detector, in particular in the forward region, is being completed, during the last year the emphasis of the work has been put on fine tuning of the physics output. The existing test beam data for the different components of the calorimetric system have been exploited to adjust different parts of the Geant4 models for hadronic and electromagnetic showers, as well as the CMS custom code used, in close collaboration with Geant4 developers. Significant improvements have been achieved in describing the data, albeit at the cost of a notable increase in cpu time. A big effort has therefore been undertaken to put in place a suite of performance analysis tools and a number of optimisations have now been introduced that allow the code to fit the resource constraints posed by the CMS computing model.

Presented at CHEP
1 Introduction

Experience has been gained in simulating the CMS detector [1] in a series of large scale productions over several years with the Full Simulation application based on the CMSSW software framework. A description of the basic characteristics of the simulation can be found in [2].

The complete geometrical description of the apparatus and of its material budget is provided through a custom-developed package, DetectorDescription [3], which is an integral part of CMSSW. It is based on an admixture of xml based static descriptions and C++ code to build algorithmically parts of the model. The simulation of the electronics (digitization) and the reconstruction algorithms use a sub-set of this description for the sensitive detectors, the so called reconstruction geometry, which is implemented in a set of dedicated objects mostly built out of the complete model, and in any case required to be fully consistent with it. The most recent development has been to allow the storage of both the complete and the reconstruction geometries in the condition database: the detailed description is transformed back into a single static xml-based file, while the minimal amount of information needed to build the sensitive detectors model objects for each subsystem is stored separately.

The model for the interaction of particles with matter is provided by the Geant4 toolkit [4], and by the specialized description of each sensitive detector response. Currently the most recent version of the toolkit, 9.2.patch01, is being tested in view of next round of massive production. The input to this stage is provided by the event generators in HepMC format, and the output consists of collections of simulated tracks and vertexes in the tracker, and of hits recorded in all the sensitive elements of each subsystem.

The simulation of the electronic readout response is provided by custom specialized code for each subsystem to describe the conversion of the energy deposit in the sensitive elements into digital signal. It provides an output in the same format of the real readout, to be used by the L1 and High Level triggers, and to be converted in the RAW format which can be processed by the reconstruction code. In this step, there is optionally the possibility to simulate the pileup from other collision events and from other background sources (like cosmic and beam halo/beam gas induced events) by combining multiple simulated events in merged collections to be used as input to the digitization stage. The algorithm dedicated to such mixing operation correctly moves the time of the different signals to mimic the provenance from different bunch crossings according to the desired luminosity scenario. A module to perform a similar operation by mixing real data and simulated events is currently under development and test.

Recently, the emphasis of the work on the simulation has moved from establishing the basic setup for the first detector performance studies, to making it robust enough for routine comparison with real collision data so that it can be used in a reliable way for physics analysis preparation. This requires validation of the material budget description of the detector as well as the physics models used to simulate the passage of particles through the detector through the ability of the simulation to describe real data, such as laboratory measurements, cosmic and test beam events. At the same time, software performance must be constantly monitored during the code development cycle, and when possible improved in order to be compatible with available resources, as described in the CMS computing model [5].

2 Tuning and validation of the geometry and material budget

The complete description of each part of the detector has been reviewed during recent years, according to the final engineering drawings and using measurements done in the integration facilities where the different components have been assembled. Besides getting the correct position and size of each piece in the setup, attention has been mainly put in ensuring that the material budget described is as much as possible in agreement with the real one, as this is what finally matters to correctly describe the particle interaction rate. The development of the geometrical description has tried to balance the level of accuracy, not only in the sensitive components but also in all the passive structures (mechanical support, cables, electronic cards, cooling, etc) and the need for an affordable model from a point of view of the navigation during the simulation, which implies introducing some simplifications whose impact has to be evaluated.

As an example, the tracker subsystem has undergone a very long and detailed review of the description of each of its parts. In Figure 1 the evolution of the description of a sensitive tracker component is shown, and compared with the real detector part. It is pretty evident the improvements which have been introduced in the latest versions of the model.

The weight of the different parts measured in the laboratories have been used to constrain and adjust the material composition based on the engineers specifications, accounting for possible simplifications introduced in the modeling of the geometry itself. This composition can be, to some extent, tuned to better match the real data available.
Figure 1: View of the evolution in the description of a tracker sensitive detector and comparison with the image of the real object.

Figure 2: Radiation length distribution vs $\eta$ for the tracker geometrical model.

As an overall test, the weight of the tracker setup brought at the CMS cavern for the final installation has been measured and this value has been compared with the one derived from the simulated description, accounting for the extra components needed (missing cooling fluid, installation and packaging parts, etc.). Using an estimate for these latter additions, a 3% level agreement between the measured and simulated material budget has been found (corresponding to a final discrepancy of about 170 kg). Detailed measurements have been done for the innermost part of the system, the barrel pixel, showing (without cables and coolant) a preliminary level of agreement with the model used in the simulation of about 6%. The evolution of the description during the development has been constantly monitored in detail, see for instance figure 2.

Similar activities have been done also for the calorimeters. For instance ECAL crystals are well known, and constitute the bulk of the material budget. But in order to have a reliable simulation of the interactions before them (mostly relevant for electromagnetic showers) and in the material behind them just in front of HCAL (important for hadronic showers), the measured weight of all the passive components has been used to fine tune a detailed model of them up to a 3% level of agreement.

### 3 Tuning and validation of the model of particles interaction with matter

While cosmic events are useful to understand the response of the electronics in tracking systems, they are of limited use for the validation of the description of particles interaction with matter. For this part, the bulk of the useful data comes from the calorimeters test beam campaigns that have been done during the last 5 years to study the response of different parts of the calorimetric system to electrons and hadrons of known energy and direction. Most of the
results come from two sets of measurements done in 2006 at the H2 and H4 beam lines at CERN. The former was a combined calorimetry test (ECAL barrel super-module and HCAL barrel part behind it), mostly useful to understand hadronic showers modeling, while the latter was an ECAL only test, using an electron beam, therefore useful for detailed study of electromagnetic showers. A set of HCAL data collected in 2004 has been also used.

At the beginning of 2008, the comparison with the simulation setup based at that moment on Geant4.8.3.patch01, was showing a number of problems. Electromagnetic shower transverse profiles agreed reasonably well between data and simulation to within 0.5% level, but there were changes of about 1% from one version to another without a clear understanding of the cause. The situation was much worse for hadronic showers: they were narrower and shorter in shape in the simulation, with a mean measured energy and a rate of early interactions in ECAL overestimated in the simulation compared to data. For this reason, a Calorimetry Task Force was appointed to work on the fine tuning of the simulated model on the test beam data, so as to have an improved agreement available for Monte Carlo productions before the beginning of the data taking [7]. The work has been done in close contact with the Geant4 developers, and this has proved to be very fruitful, an activity from which both the basic Geant4 toolkit and the CMS simulation have greatly benefited.

The preliminary results of the work of the task force have been presented in reference [8] and updated results are discussed in [7], but the work is still continuing at the present moment, enlarging its scope with the analysis of the data from the 2007 campaigns (for the forward components of the detector). The main consequence of these activities has been a contribution to several improvements contained in the latest Geant4 versions, up to the adoption of the most recent Geant4.9.2.patch01. For the electromagnetic sector, the revision of the models for the basic processes, with a better interpolation of physics tables and separate tuning of the multiple scattering for each particle type, has been the key for a better understanding of the electromagnetic shower behavior. Bremsstrahlung and pair production from hadrons have proven to be necessary to explain the hadron interaction rate in ECAL. Scintillation saturation effects have been introduced in the description of the ECAL crystals and HCAL fibers to match the real data. The hadronic part of the physics lists has been deeply reviewed, moving to the adoption of the Bertini cascade model. The improved description of the quasi elastic scattering processes in both the high (QGSP model) and low (Bertini model) energy regimes has been of great importance, as well as the improvement in the Bertini cascade cross section. The physics list adopted by CMS as a baseline choice has moved to QGSP_BERT_EMV.

### 3.1 Electromagnetic showers

The transverse shape of electromagnetic showers in ECAL is experimentally measured as ratios of the energy deposited in the central crystal of the shower (the one with the highest energy deposit) and matrices of crystals centered on it ($E3 \times 3$, $E5 \times 5$, ...). The behavior of these ratios as a function of the incidence angle is in good agreement between data and simulation, and stable with respect to the change of Geant4 version, see Figure 3, but their absolute values at a given incidence differ from the real data by an offset which has fluctuated in recent Geant4 versions by up to 1%. The detailed revision of the multiple scattering description, and its separated tuning for each particle type has allowed to bring back the agreement in these transverse shape observables to a 0.5% level in the most recent version, with a deeper understanding of the reasons for its fluctuation (due to different stages of the revision of the multiple scattering model).

The adoption of Birks law to describe scintillation light saturation in crystals has been motivated by the observation that the energy deposit in ECAL was overestimated in the simulation. Unfortunately no measurement of the Birks parameters is available for PbWO$_4$, the material of ECAL crystals. The values measured for BGO have been used. The electromagnetic showers have been checked to be unaffected in their energy scale by this change, which is affecting the relative response of hadrons and $\pi^0$ induced cascades in hadronic showers.

### 3.2 Hadronic showers

The mean energy measured in the combined H2 calorimetric setup as a function of the hadron beam incident energy is shown in Figure 5 compared to the Geant4.8.3.patch01 results. The clear discrepancies look to have been significantly reduced when moving to the most recent Geant4.9.2.patch01 description combined with the description of scintillation light saturation effect in the sensitive detectors, as can be seen in Figure 6.

Another important observable to study in hadron interactions is the fraction of MIP-like hadrons in ECAL, e.g. depositing no more than 0.8 GeV in ECAL. The comparison between the results reached in [8] with Geant4.9.1.patch02 and the most recent ones with Geant4.9.2.patch01 are shown in Figure 7. The discrepancies at high energies have been strongly reduced thanks to the addition of bremsstrahlung for hadrons, while the improved treatment of quasi elastic scattering in the Bertini cascade has significantly diminished the discrepancies at low energy.
Figure 3: Ratio E1/E5×5 of the ECAL energy deposits in the central crystal of the electromagnetic shower and in a 5×5 matrix around it as a function of the electron incidence angle for different Geant4 versions.

Figure 4: Energy deposit by 5 GeV pions in ECAL. The adoption of the Birks law in the ECAL response improves the agreement with data.
Figure 5: Mean hadron energy measured in the combined calorimetry test beam for all selected particles (left) and for those MIP-like in ECAL (right) compared with the Geant4.8.3.patch01 prediction.

Figure 6: Mean hadron energy measured in the combined calorimetry test beam for all selected particles (left) and for those MIP-like in ECAL (right) compared with the Geant4.9.2.patch01 prediction.

Figure 7: Fraction of MIP-like hadrons in ECAL measured in the combined calorimetry test beam compared with the prediction of Geant4.9.1.patch02 (left) and Geant4.9.2.patch01 (right).
Hadronic shower shapes have been among the first observables to be checked, and to trigger concerns about data-Monte Carlo disagreements. Their current level of reproducibility is shown in Figure 8, and is pretty satisfactory.

### 3.3 GFlash

The significant discrepancies between real data and Geant4 based predictions originally found at the beginning of the work of the Calorimetry Task Force, and the fact that tuning Geant4 is a lengthy and complex process, have suggested the investigation of alternative approaches to the description of calorimetric showers which could be more directly tunable. The GFlash project interfaces with Geant4 the fast parameterization of showers originally developed by Grindhammer [9] and used in H1, replacing the electromagnetic shower or the hadronic one from the first inelastic interaction onwards. A tuning on the test beam data of the parameters of the chosen function forms, which have been optimized for the CMS calorimetric setup, is ongoing. Although the possibility of a more direct tuning on data was the first target of this project, its enhanced computing speed, compared to Geant4, is of course interesting. Studies are ongoing to extend the improvements observed at single particle level, see Figure 9, to real full generated physics events, by suitably defining the thresholds beyond which the GFlash approach should take over the full Geant4 description.

### 4 Tuning and validation of the digitization model

Cosmic events have been extensively used, mainly by the tracking systems, to validate and adjust the model of the electronic digital response to energy deposits. As an example, the data taken already at the Tracker Integration Facility have been exploited to study the characteristics of the signal development in the tracker strips [6], allowing the fine tuning of several relevant quantities, such as capacitive couplings. The effect of this work can be seen in Figure 10, where the size of the hits clusters, measured in number of strips, are compared for the different parts of the tracker as a function of the incidence angle: the agreement has definitely improved compared to the first versions, and is already pretty satisfactory.
Figure 10: Distribution of the reconstructed hit cluster size, measured in number of strips, for the different parts of the tracker as a function of the incidence angle of the impinging cosmic muon.

Table 1: CPU speed for PYTHIA minimum bias and $t\bar{t}$ events measured on a CERN build machine. Numbers are shown separately for the SIM (e.g. particle interaction through Geant4) and DIGI (digitization) steps.

|               | SIM (s/ev) | DIGI (s/ev) |
|---------------|------------|-------------|
| Single muon $p_t=10$ GeV | 0.53       | 0.95        |
| Single electron $E=1$ TeV   | 115.01     | 0.92        |
| Single pion $E=1$ TeV       | 70.80      | 1.03        |
| Minimum bias            | 12.00      | 1.02        |
| $t\bar{t}$              | 104.09     | 1.95        |

5 Computing performances

Extensive studies have been done to fully understand and optimize whenever possible all the different aspects of the computing performance of the CMS simulation (as well as of the whole CMS framework): CPU speed, number of memory allocations, memory footprint, event size. For a detailed description of the tools and techniques used see reference [10]. The basic CPU speed measured in controlled load conditions on a CERN build machine (Intel Xeon 5160, dual processor dual core running at 3 GHz, with a RAM of 8 GB) for the Geant4 and digitization stages separately on two typical benchmark processes, PYTHIA minimum bias and $t\bar{t}$ production events, and a few particle guns, are shown in Table 1.

Another point on which there is constant attention and work to improve is the memory footprint, which should be constrained in the 1 GB budget per job, according to the computing model. This is particularly critical for high multiplicity events, pileup simulation or heavy ions events.

The move from the QGSP_EMV to the QGSP_BERT_EMV physics list suggested by the studies of the Calorimetry Task Force on test beam data has resulted in an increase of about 50% of both the CPU time and of the simulated event size. This has triggered investigations to understand the cause for this deterioration and possibly recover it.

The main cause for the increased event size has been found in a huge number of very low energy hits produced in the Bertini cascade model, see Figure 11. As it can be seen, the net contribution to the total energy deposit from the lowermost part of the spectrum is negligible, and can be safely removed. This allows a gain in the event size, and memory occupancy, with no real impact on the physics output (as verified also on reconstructed quantities).

The memory footprint can be squeezed by moving the management of this hit suppression to a primary-by-primary track basis, since any physics event can be seen as an array of independent particle guns. This is due to the fact that the Geant4 particle stack is used in CMS simulation always in the "last in first out" mode. In this way, the hit objects rejected can be reused for following tracks reducing both the total number of allocations and the overall memory allocated at any moment. A similar strategy has been used to reject tracks for the persistent storage. This strategy has allowed a gain of about 50 MB on the average memory footprint of a typical $t\bar{t}$ event run.
Figure 11: Distribution of the log$_{10}$(Energy) for simulated ECAL hits (top) and fraction of the total energy deposit in a crystal determined by the corresponding energy bin (bottom).

6 Conclusion

An extensive program of validation and tuning of the different components of the full simulation has been carried out by CMS in the last year. Great attention has been paid to balance the improvements in the physics quality of the simulation with respect to the computing performances, so as to make the full simulation an affordable tool for routine mass productions in view of data taking and physics analysis. The current level of data and simulation agreement reached by using test beam and cosmic events will form the basis from which a new round of comparison on collision data will start as soon as possible.

Acknowledgment

The author would like to thank the members of the CMS Simulation team, of the CMS Calorimetry Task Force and the CMS members that have reviewed the material presented in this work.

References

[1] The CMS Collaboration 1994 The Compact Muon Solenoid Technical Proposal CERN/LHCC 94/38, LHCC/P1
[2] Banerjee S 2008 J.Phys.Conf.Ser. 119:032006
[3] Liendl M et al 2001 The Role of XML In The CMS Detector Description Database CMS CR 2001/008
Case M et al 2001 Detector Description Domain Architecture and Data Model CMS Note 2001/057
[4] Agostinelli S et al 2003 Nucl. Inst. Meth. A 506 250;
Allison J. et al 2006 IEEE Trans. Nucl. Sci. 53 N. 1 270.
[5] The CMS Collaboration 2005 The CMS Computing Project Technical Design Report CERN/LHCC-2005-023
[6] Adam W et al 2008 Silicon Strip Tracker Detector Performance with Cosmic Ray Data at the Tracker Integration Facility CMS Note 2008/032
[7] Banerjee S 2009 *Data Driven Approach to Calorimeter Simulation in CMS* in these proceedings

[8] Abdoullin S et al 2008 *Calorimetry Task Force Report* CMS Note 2008/025

[9] Grindhammer G et al 1990 *Nucl. Inst. Meth.* **A290** 469

[10] Elmer P 2009 *CMS Software Performance Strategies* in these proceedings