# The Fast Simulation of the CMS Experiment

To cite this article: Andrea Giammanco 2014 *J. Phys.: Conf. Ser.* **513** 022012

View the article online for updates and enhancements.

# The Fast Simulation of the CMS Experiment

**Andrea Giammanco**

Université Catholique de Louvain, Louvain-la-Neuve, Belgium
Fonds National de la Recherche Scientifique (FNRS), Brussels, Belgium
National Institute of Chemical Physics and Biophysics, Tallinn, Estonia

E-mail: andrea.giammanco@cern.ch

**Abstract.**   The CMS collaboration has developed a fast Monte Carlo simulation of the CMS detector with event production rates $\sim$100 times faster than the GEANT4-based simulation ("full simulation"), with nonetheless comparable accuracy for most of the physics objects typically considered in the analyses. We discuss basic technical principles of the CMS Fast Simulation and their implementation in the different components of the detector, and we illustrate the most recent developments towards a tighter integration with the full simulation and a better flexibility.

## 1. Introduction

The fast simulation of the CMS detector (henceforth FASTSIM) [1] is an object-oriented system, written in C++ and steered by python configuration files, in the CMS software framework (CMSSW) [2].

It is alternative and complementary to the GEANT4-based approach [3] (for which we employ the term "full simulation", including the electronic effects simulated by CMSSW-specific modules), with respect to which it is regularly validated and tuned. It differs from both parametric simulations like DELPHES [4], which translate generator-level information to the analysis level after smearing and applying efficiency and mistagging factors, and GEANT4 which is in many senses an *ab-initio* simulation; in FASTSIM, material effects are taken into account and parameterized at the hit level. These hits can then immediately be used as inputs of the same "higher-level algorithms" (track fitting, calorimeter clustering, $b$ tagging, electron identification, jet reconstruction and calibration, trigger algorithms, etc.) as in the full reconstruction and data analysis chain. This allows comparisons between fast and full simulations, as well as data, and subsequent tuning in a straightforward manner, with the use of identical analysis programs. Comparisons between FASTSIM and early CMS data have been shown in Ref. [5].

Reference [1] contains a more detailed description of the simulation modules, while this paper emphasizes recent developments and plans.

## 2. The Physics Processes

The central feature of the CMS detector is a superconducting solenoid providing a field of 3.8 T. Located within the solenoid are the silicon pixel and strip tracker, the crystal electromagnetic calorimeter (ECAL) and the brass/scintillator hadron calorimeter (HCAL). Muons are measured in gas-ionisation detectors embedded in the steel return yoke. In addition to the barrel and endcap detectors, a quartz-fiber Cherenkov detector extends the jet acceptance up to $|\eta| = 5$,

where the pseudorapidity $\eta$ is defined as $\eta = -\ln\tan\frac{\theta}{2}$, where $\theta$ is the polar angle of the particle or jet trajectory with respect to the counterclockwise beam direction. A much more detailed description of CMS can be found elsewhere [6].

The input of the fast simulation is a list of particles (originating from an event generator or a simple particle gun) characterized by their momentum and origin vertex, with mother and daughter relationships to follow the various decay chains in the event. Each of the (quasi)-stable particles in this list is then propagated in the CMS magnetic field to the different layers of the various CMS subdetectors, which it may interact with. While propagating, these quasi-stable particles are also allowed to decay according to their known branching fractions and decay kinematics, using a PYTHIA6 [7] routine adapted in order to take into account magnetic bending. The particles resulting from the interactions with the detector layers or from the decays in flight are added to the original list, and propagated/decayed in the same way.

The interactions simulated at present in the fast simulation are  1) electron Bremsstrahlung; 2) photon conversion;  3) charged particle energy loss by ionization;  4) charged particle multiple scattering;  5) nuclear interactions; and  6) electron, photon, and hadron showering. The first five processes are simulated for particles traversing the thin layers of the tracker (Section 3), while the latter is parameterized in the electromagnetic (Section 4.1) and hadron (Section 4.2) calorimeters. Muons propagate through the tracker, the calorimeters and the muon chambers with multiple scattering and energy loss by ionization taken into account during the propagation (Section 5).

## 3. Inner Tracker

A simplified version of the tracker geometry is used, made of several thin nested cylinders representing the sensitive layers of the pixel and strip detectors, interleaved with non-instrumented cylinders with dead material (cables, support, etc.) [6]. The material is also assumed to be uniformly distributed over each cylinder barrel or endcap. A "radiography" in the transverse plane of this simplified geometry is shown in Figure 1a, obtained from the vertices of converted photons in a large number of simulated events. For comparison, the higher level of details present in the full simulation is visible in Figure 1b.
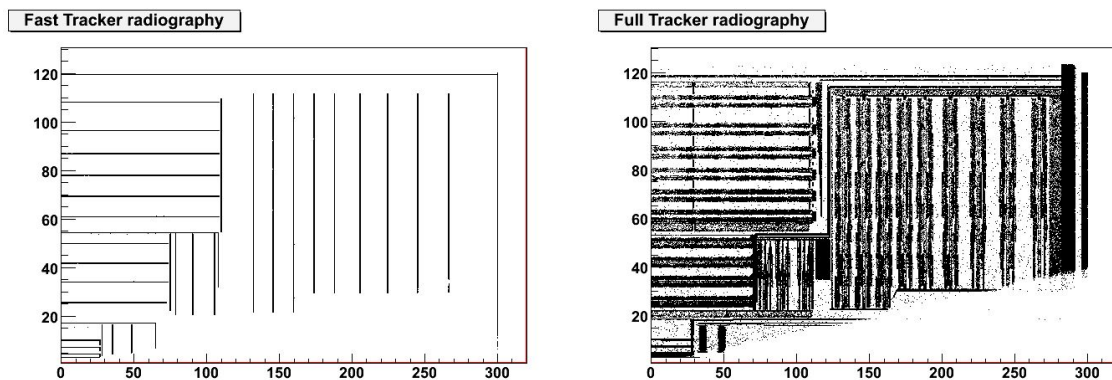


**Figure 1.** A radiography of a quarter of the simulated tracker geometry in the (a) fast and (b) full simulation.

The complete magnetic field map is used for the track propagation between two surfaces. While being propagated in the magnetic field through the tracker layers, charged particles experience multiple scattering and energy loss by ionization. The intersections between the modified trajectories and each tracker layer define the position of "simulated hits". Each simulated hit is turned with a certain efficiency to a "reconstructed hit", the position of which is

obtained from a smearing of the simulated hit position. In the silicon strip tracker, the Gaussian resolution in each of the two directions (longitudinal and transverse to the beam direction), obtained from a fit of the residuals with respect to the reconstructed charged particle tracks in the full simulation, is treated as a constant for each layer. In the pixel detector, the Gaussian resolution in each of the two directions is parameterized according to the pixel cluster size (itself generated according to its fully simulated $\eta$-dependent distribution) and on the incident angle of the particles with respect to the layer.

To save execution time, a "fast tracking" is performed in FASTSIM, that differs from the standard tracking [8] mostly in the fact that seeding is emulated by using the Monte Carlo truth, i.e. by checking that at least one combination of hits associated with the charged particle fulfils the seed selection criteria. As a result of the fact that no real pattern recognition is performed, there are no fake tracks. Despite these limitations, a good agreement is observed with the Run-I data [5].

Examples of high-level reconstruction modules that critically depends on the quality of the tracks are the typical b-tagging algorithms. Reference [9] compares the performance of the main algorithms in CMS between full and fast simulation, finding a good agreement overall but with discrepancies in efficiency (generally higher in FASTSIM) and fake rate (generally lower in FASTSIM) that grow with $p_T$ of the jets. These discrepancies are attributed mostly to two of the main approximations of the fast tracking:

(i) the lack of fake tracks (explaining the optimistic fake rate)

(ii) the fact that two hits simulated at the same position are never merged into a single reconstructed hit; this approximation is worse for high-$p_T$ jets, which tend to have a very large density of tracks.

Future runs with larger instantaneous intensity and larger amount of average pile-up per event will further challenge the accuracy of these approximations. Abandoning the paradigm of using Monte Carlo truth to speed-up tracking seems unfeasible, because running the real pattern recognition would make the overall execution time worse by an order of magnitude (see Sec. 8), and therefore the FASTSIM would loose most of its interest. However, a possibility for the future is to parameterize the fake rate and the efficiency degradations at track level by studying full-simulated samples with different track densities.

## 4. Calorimeters
### 4.1. Electromagnetic Calorimetry
The showers of electrons and photons which impinge on the ECAL are simulated in two steps.

In a first step, the shower is developed following the well-tuned Grindhammer parameterization [10, 11], as if the ECAL were an homogeneous medium. This approximation is realistic because the CMS calorimeter is made of contiguous crystals. In this parameterization, an electron shower consists of several thousands energy spots, longitudinally distributed according to a $\Gamma$ function, the parameters of which fluctuate from one shower to the other. The deposited energy is integrated over $2X_0$-thick longitudinal slices, including uncertainties due to the photo-statistical fluctuations and the longitudinal non-uniformity in the crystals.

In each slice, and as a second step, the energy spots are distributed in space according to the radial profile of the Grindhammer parameterization and placed into the actual crystal geometry, under the realistic assumption that no energy is lost in the small inter-crystal gaps. The time used in this step is kept to a reasonable value by limiting the two-dimensional spot-crystal assignment to a small $7 \times 7$ crystal grid (and even smaller for low energy electrons) in a plane perpendicular to the shower longitudinal development. The energy collection simulation is then refined by simulating a number of effects such as rear and front leakage, energy losses in the gaps between ECAL modules, and shower enlargement due to the magnetic field.

In front of the ECAL endcaps, electrons may first cross the preshower. In this case, the corresponding showers are developed from the preshower entrance, and the energy deposit in each layer is converted into a number of MIPs (with related statistical uncertainties), assigned to the relevant strips according to the shower radial profile. Very energetic electrons (above several hundred GeV) can extend their shower substantially beyond the ECAL. A $2X_0$-thick gap between the rear side of the crystals and the entrance of the HCAL is assumed, in which all the energy integrated from the $\Gamma$-distribution tail is lost. The rest is assigned to the HCAL towers according to the shower radial profile at this depth, and the energy of each spot is corrected for non-compensation.

The Grindhammer parameterization only applies to electrons. Photons are first converted in the ECAL (or preshower) material at a varying depth (according to the number of radiation lengths traversed). Each of the resulting $e^+e^-$ pairs gives rise to 2 separate showers generated as explained above along the same longitudinal direction and, therefore, with the same transverse crystal grids.

Finally, at rapidity values not covered by the electromagnetic calorimeter ($|\eta| > 3$), electrons and photons are propagated directly to the forward hadron calorimeter. Here, the detector response is tuned to the full simulation of electrons, in a way similar to that explained in Section 4.2.

### 4.2. Calorimeter Response to Hadrons

Charged and neutral hadrons are also propagated to the ECAL, HCAL and HF entrance after their interactions with the tracker layers. Their energy response is derived from the full simulation of single charged pion, and applied to all hadrons. In the CMSSW releases until 2012 [1] the energy smearing in HCAL was modeled through a Gaussian, while recently a new tuning has been deployed, based on a double-sided Crystal Ball function, i.e. a combination of a Gaussian and a power-law function for the tail.

This smeared energy is then distributed in the calorimeters using parameterized longitudinal and shower profiles, with shower-to-shower variations, following an approach similar to that of GFLASH [12]. The actual implementation in FASTSIM is, however, slightly different from that of the original GFLASH, so as to speed up the code and to better adjust the shower shape and the ECAL energy fraction to the full simulation.

### 4.3. Digitization and Local Reconstruction in the Calorimeters

Until 2012, simulated hits from the calorimeters were converted into reconstructed hit by a dedicated module that applies energy smearing, electronic noise (tuned to full simulation), zero suppression, hit inefficiency. The energy smearing was tuned to full simulation at the level of reconstructed hits, and therefore its parameters took already into account all electronic effects.

A recent development has been the modification of the FASTSIM sequences to allow the simulated hits from the calorimeters to be processed by the same digitizer modules as in the full simulation, following the example of what was already done for hits in the muon chambers (see Sec. 5). This option, which is now the new default, is advantageous for several reasons:

- it removes the need for frequent retuning of the models of the various electronic effects simulated in the smearing module, as these are completely delegated to the digitizers;
- it gives a simulation of these effects which is as accurate as in full simulation (see Fig. 2 for an example of an observable which is sensitive to the noise model);
- by exploiting the time-response simulations which are part of all digitizers, it makes the FASTSIM able to handle out-of-time signals; this is of great importance for a more realistic pile-up simulation (see Sec. 6).

The cost in terms of execution time was considered affordable, as the overall CPU time spent for digitization in FastSim is still a small fraction of the total, see Sec. 8.

The old smearing modules are still available as an option, as they can be of interest for example of studies of the performance of future detector designs, for which dedicated digitizer modules are not available and an extremely realistic simulation of electronic effects is not crucial.
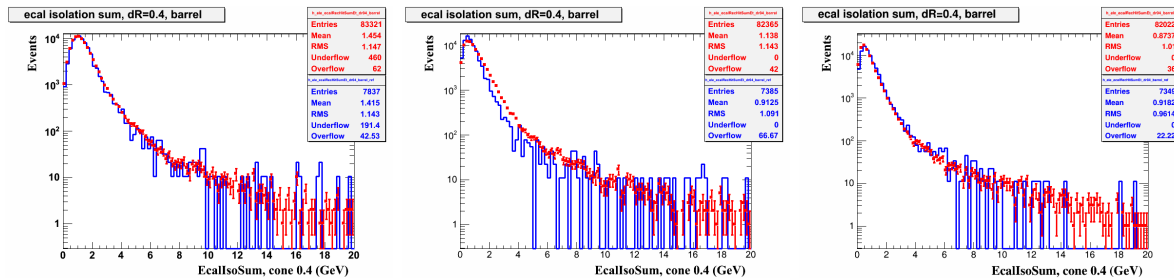


**Figure 2.** An example of high-level observable sensitive to the electronic response of the ECAL barrel, including the noise model and zero suppression: sum of the transverse energies of the reconstructed hits in a cone of $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ around the direction of the electron candidates in $Z \to e^+e^-$ events, for fast (red) and full (blue) simulation. Left: CMSSW 4.4, legacy release for 7 TeV analysis. Middle: CMSSW 5.3, which implements a different selective readout and zero suppression from 4.4; the FastSim noise model has been retuned to reproduce the bulk behaviour but it is inadequate in the tails. Right: CMSSW 6.2, released in 2013; the FastSim hits are processed through the digitizer module.

## 5. Muons

A muon, either coming from the main interaction vertex or produced from the decay of a hadron within the Tracker volume [1], is propagated in the CMS magnetic field through the calorimeters, the solenoid and the muon chambers. Because of this logic, no fake muons from punch through are present in FastSim, but this approximation is deemed acceptable because these are a negligible fraction of the muon candidates that pass typical analysis cuts [13].

At the time of writing, the implementation of the calorimeter response to muons is done independently from the propagation of the muon track, and in a way very much similar to that for pions (Section 4.2).

At the moment, only the multiple scattering and the energy loss by ionization are taken into account as physics processes for the muon in the fast simulation: therefore, no bremsstrahlung or delta-ray production are simulated. However, the effect of delta rays on hit efficiencies is parameterized as an analytical function of muon momentum.

The actual geometry of the CMS muon chambers (DT, CSC and RPC) is taken from a database, and simulation hits are positioned in the detector whenever the track trajectory crosses an active layer of those chambers. Following the same rationale as in Sec. 4.3, the simulated hits are processed through the full simulation digitizer modules and the resulting digis (raw data equivalent) are fed to the normal local and global muon reconstruction packages, to end up with the final muon objects to be used in the physics analyses.

## 6. Pile-up Simulation

Event pile-up, i.e. additional collisions superimposed to the primary (hard) event, is a crucial ingredient of any LHC simulation, and even more so for studies of future performance with

---

[1] At present, decays outside the Tracker outer radius are not considered in FastSim.

even larger integrated luminosities. At the time of writing, pile-up is simulated before the simulation of the detector response by overlaying minimum-bias pp collisions, randomly chosen from a pregenerated sample. In order to avoid correlations between event subsets overlapped to the same pile-up event sequence, minimum-bias samples are never reused in the same order. The distribution of the number of pile-up interactions is diced from a user-defined function or histogram, and the number of actual interactions is stored per each event, allowing to reweight a posteriori to match the pile-up profile observed in data. Only in-time pile-up, i.e. in the same bunch crossing as the primary event, is simulated in FASTSIM.

With an average of $\approx 2$ pp interactions per bunch crossing during the 2010 run, $\approx 10$ in 2011, and $\approx 21$ in 2012, and a bunch spacing of 50 ns, this approach has served us well so far. In preparation for the 25 ns bunch spacing planned for Run-II, and with the much larger pile-up averages expected in future runs, we are now considering a very different pile-up mixing scheme, designed to save execution time and memory usage and to naturally accommodate for out-of-time pile-up, i.e. signals originating from different bunch crossings.

In the proposed scheme, simulated hits in the calorimeters and the muon chambers would be treated as in the CMS full simulation [2], i.e. hits from the primary event and from the minimum-bias events would be overlayed when being fed to the digitizer modules. This possibility was not available before the recent modifications detailed in Sec. 4.3. In the inner tracker, we will mix already reconstructed tracks from primary and minimum-bias events, therefore skipping the most time-consuming module of FASTSIM for the charged particles from pile-up ($\approx 20$ per minimum-bias event.)

An appealing possibility with this new scheme is to mix primary events simulated with FASTSIM and minimum-bias events simulated with the full simulation, or the other way around.

## 7. Detector Upgrades

Recently, the preparation for simulated studies for the future detector upgrades [14] has reached a high priority in all the LHC experiments. From the FASTSIM point of view there are two major streams of development ongoing towards the upgrade studies (in addition to the new pile-up treatment outlined in Sec. 6): the generalization of the tracker and calorimeter geometries.

It is now possible to generate arbitrary geometries in the tracker by exploiting the tool presented in Ref. [15]. This tool produces a detailed reconstruction geometry file, from which the FASTSIM can derive the corresponding interaction geometry.

The effects of radiation damage (amplitude degradation) with very large integrated luminosities have been studied with FASTSIM. Calorimeter geometry, segmentation and material properties have been made fully configurable.

## 8. Timing

Table 1 shows the breakdown of cpu time for $t\bar{t}$ production at 13 TeV with FASTSIM, with CMSSW 7.0 on a 64 bits machine [2], with no pile-up and with a pile-up profile similar to the one observed during the 2012 data taking.

The digitization time does not scale in FASTSIM as in the full simulation, where also the tracker hits are digitized, and that the emulated seeding in the tracking step allows a major reduction of the reconstruction time, especially at large pile-up.

## 9. Conclusions

The CMS Fast Simulation, designed to deliver a percent-level accuracy with an overall execution time of the order of seconds (including reconstruction), has proven to be a very useful tool for analysis in CMS during Run I. Although the main use case in CMS is for small private

---

[2]  Intel(R) Xeon(R) CPU L5640 @ 2.27GHz

| Step | Full, no PU | Fast, no PU | Full, PU 2012 | Fast, PU 2012 |
|---|---|---|---|---|
| Generator (PYTHIA) | 0.02 | same | same | same |
| Detector simulation | 88 | 0.20 | same as no PU | 0.88 |
| Digitization | 0.7 | 0.24 | 3.2 | 0.30 |
| Reconstruction | 1.9 | 1.2 | 10.6 | 2.8 |

**Table 1.** Timing performance of fast and full simulation for two pile-up scenarios (no pile-up and 2012-like) in seconds per event, for $t\bar{t}$ production at 13 TeV. The same generator-level events have been used for this comparison. Time spent on detector simulation for the full simulation does not depend on the amount of pile-up because pile-up events are taken from a sample of minimum-bias events already processed through detector simulation, differently from FASTSIM where pile-up mixing happens before detector simulation.

productions of processes of interest of few groups, high-statistics Monte Carlo samples have been produced centrally with FASTSIM, providing the "simplified model scan" [16] signals for most CMS publications related to SuperSymmetry searches, and samples of $t\bar{t}$, single top, W+jets and Z+jets events with variations in the generator parameters for the purposes of evaluation of systematic uncertainties.

The recent trends in FASTSIM development have been towards a tighter integration with the full simulation, in particular by the usage of the same digitizers in the calorimeters and the muon chambers and by a new pile-up mixing scheme, and towards a generalization of the geometry to allow the flexibility needed by upgrade studies.

## References

[1] A. Giammanco, R. Kroeger, and R. Rahmat. The fast simulation of the CMS experiment. *J.Phys.Conf.Ser.*, 396:062016, 2012.
[2] CMS Collaboration. CMS technical design report, volume I: Detector performance and software. Technical Report CERN-LHCC-2006-001, CERN, 2006.
[3] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl.Instrum.Meth.*, A506:250–303, 2003.
[4] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, et al. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. 2013.
[5] CMS Collaboration. Comparison of the Fast Simulation of CMS with the first LHC data. (CMS-DP-2010-039), Oct 2010.
[6] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 03:S08004, 2008.
[7] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.
[8] CMS Collaboration. CMS Tracking Performance Results from early LHC Operation. *Eur. Phys. J.*, C70:1165–1192, 2010.
[9] CMS Collaboration. Status of b-tagging and vertexing tools for 2011 data analysis. Technical Report CMS-PAS-BTV-11-002, CERN, Geneva, 2011.
[10] G. Grindhammer, M. Rudowicz, and S. Peters. The fast simulation of electromagnetic and hadronic showers. *Nucl. Instrum. Meth.*, A290:469, 1990.
[11] G. Grindhammer and S. Peters. The parameterized simulation of electromagnetic showers in homogeneous and sampling calorimeters. 1993.
[12] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P.A. Dubois, et al. Geant4 developments and applications. *IEEE Trans.Nucl.Sci.*, 53:270, 2006.
[13] CMS Collaboration. Performance of muon identification in pp collisions at $\sqrt{s}$=7 TeV. *CMS Physics Analysis Summary*, CMS-PAS-MUO-10-002, 2010.
[14] CMS Collaboration. Technical Proposal for the Upgrade of the CMS detector through 2020. Technical Report CERN-LHCC-2011-006, LHCC-P-004, CERN, Geneva, Jun 2011.
[15] S. Mersi, D. Abbaneo, N. De Maio, and G. Hall. CMS Tracker Layout Studies for HL-LHC. *Physics Procedia*, 37(0):1070 – 1078, 2012. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).

[16] J. Alwall, P. C. Schuster, and N. Toro. Simplified models for a first characterization of new physics at the LHC. *Phys. Rev. D*, 79:075020, Apr 2009.