Available on CMS information server

**CMS CR -2012/146**

The Compact Muon Solenoid Experiment

**Conference Report**

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland

**15 June 2012 (v2, 20 June 2012)**

# CMS Simulation Software

Sunanda Banerjee for the CMS Collaboration

**Abstract**

The CMS simulation, based on the Geant4 toolkit, has been operational within the new CMS software framework for more than four years. The description of the detector including the forward regions has been completed and detailed investigation of detector positioning and material budget has been carried out using collision data. Detailed modeling of detector noise has been performed and validated with the collision data. In view of the high luminosity runs of the Large Hadron Collider, simulation of pile-up events has become a key issue. Challenges have raised from the point of view of providing a realistic luminosity profile and modeling of out-of-time pileup events, as well as computing issues regarding memory footprint and IO access. These will be especially severe in the simulation of collision events for the LHC upgrades; a new pileup simulation architecture has been introduced to cope with these issues.

The CMS detector has observed anomalous energy deposit in the calorimeters and there has been a substantial effort to understand these anomalous signal events present in the collision data. Emphasis has also been given to validation of the simulation code including the physics of the underlying models of Geant4. Test beam as well as collision data are used for this purpose. Measurements of mean response, resolution, energy sharing between the electromagnetic and hadron calorimeters, shower shapes for single hadrons are directly compared with predictions from Monte Carlo. A suite of performance analysis tools has been put in place and has been used to drive several optimizations to allow the code to fit the constraints posed by the CMS computing model.

Presented at *CHEP 2012: International Conference on Computing in High Energy and Nuclear Physics*

# CMS Simulation Software

**S. Banerjee**
**(On behalf of CMS Experiment)**
Saha Institute of Nuclear Physics, 1/AF Salt Lake, Kolkata 700064, India
E-mail: sunanda@fnal.gov

**Abstract.** The CMS simulation, based on the Geant4 toolkit, has been operational within the new CMS software framework for more than four years. The description of the detector including the forward regions has been completed and detailed investigation of detector positioning and material budget has been carried out using collision data. Detailed modeling of detector noise has been performed and validated with the collision data. In view of the high luminosity runs of the Large Hadron Collider, simulation of pile-up events has become a key issue. Challenges have raised from the point of view of providing a realistic luminosity profile and modeling of out-of-time pileup events, as well as computing issues regarding memory footprint and IO access. These will be especially severe in the simulation of collision events for the LHC upgrades; a new pileup simulation architecture has been introduced to cope with these issues.

The CMS detector has observed anomalous energy deposit in the calorimeters and there has been a substantial effort to understand these anomalous signal events present in the collision data. Emphasis has also been given to validation of the simulation code including the physics of the underlying models of Geant4. Test beam as well as collision data are used for this purpose. Measurements of mean response, resolution, energy sharing between the electromagnetic and hadron calorimeters, shower shapes for single hadrons are directly compared with predictions from Monte Carlo. A suite of performance analysis tools has been put in place and has been used to drive several optimizations to allow the code to fit the constraints posed by the CMS computing model.

## 1. Introduction

The CMS experiment chooses to use a data driven, realistic and accurate Monte Carlo program. The simulation code is based on the Geant4 [1, 2] toolkit and special effort has been made to validate the physics models of Geant4 using an elaborate set of measurements in dedicated test beam setups. The single particle response in the prototype CMS calorimeters has been measured in these experiments using electron and hadron beams over a large range of energy. Measurements of the single particle response from collision data at the Large Hadron Collider are also used in testing the predictions of Geant4.

The simulation code is built like any Cmssw [3] (CMS specific software package) application in the form of special shared object libraries "plugins". In practice this means that there is only one command to run all applications:

```
cmsRun <some-configuration-file>
```

The configuration files are written in python. Though the simulation program has been in operation for a number of years, it is a live system - goals, requirements, tools evolve throughout the lifetime of the experiment.

It is currently based on the 9.4.p03 version (patch 03 of 9.4 with some CMS specific corrections) of Geant4 which provides: (1) the physics processes for electromagnetic and hadronic interactions; (2) tools for building detector geometry and sensitive element response; (3) interfaces for tuning and monitoring particle tracking. The CMS offline framework and Event Data Model manage the control at run time. The framework relies on the concept of event processing module (EDProducer) and provides interfaces to common tools (generators, magnetic field, MC truth handling, infrastructure for hits, event mixing, digitization, etc.). It also ensures provenance tracking and event immutability.

## 2. Core Simulation Software

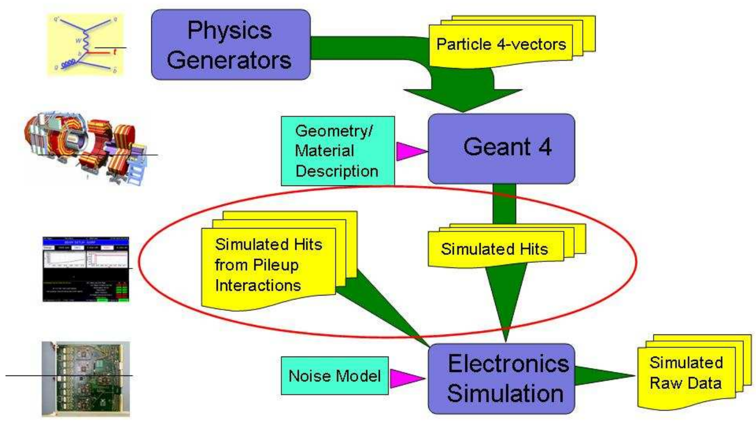


**Figure 1.** Block diagram showing the work flow of the simulation program.

Figure 1 shows a schematic view of the work flow of the simulation program. Physics generators provide four vectors of generated particles which are traced by Geant4 based simulation code. A limited number of built-in physics event generators are available. For standard production, event generator outputs are stored in HepMC [4] format and are read in and interfaced to the Event. The geometry of the detectors as well as detailed material content is provided via the EventSetup of the framework. This uses an XML-based Detector Description (DD) machinery which is configurable at run time through a hierarchy of XML files. The geometry modeler converts DD solids and materials to Geant4 counterparts. Sensitive detectors are associated with geometrical volumes through the XML configuration files at run time. The XML content is now also stored in the database together all other conditions for better tracking of the version used in production, and this is what is used for standard Monte Carlo production jobs.

The magnetic field is based on a dedicated geometry of magnetic volumes and is provided by independent subsystem via EventSetup. Selection of the field, field integrator as well as the tuning of propagation are configurable at run time. A large variety of physics lists is available for modeling physics processes. The simulation program provides tools to make run-time selection of the physics list and production cuts. This includes activation or tailoring of individual processes (the default physics list is QGSP_FTFP_BERT_EML).

There are specific user action classes which allow access to Geant4 objects at any stage (run, event, track, step). These are used for tuning, diagnostics and custom bookkeeping. A Monte Carlo truth record is kept with history of decays and interactions of the generator level particles and also of selected secondaries from Geant4 simulation.

Pure Monte Carlo inputs are used to simulate pileup interactions. The current default for pileup is minimum bias events generated through Pythia6 [5] Tune Z2 and Pythia8 [6] Tune 4C Monte Carlo. Distribution of the number of interactions per beam crossing is chosen in advance to simulate a desired luminosity profile. For each event, the instantaneous luminosity is chosen from the input distribution at random. The number of in- and out-of-time interactions to be overlaid are selected individually from a Poisson distribution based on the chosen luminosity and the total inelastic cross section (71.3 mb). Out-of-time interactions are simulated for each beam crossing that is "in scope" for a given production run. Any arbitrary bunch configuration can be generated in 25 ns steps. Times of Geant4 SimHits are shifted to match bunch assignment and the shifted times are considered in generating pulse shapes in the the digitization simulation. Typically, ±125 ns worth of bunch crossings are simulated. Collection of Geant4 SimHits from all of the minimum bias events and the hard-scatter "signal" event are merged, and then processed by digitization or electronics simulation. Simulation of "double-hard-scatter" have not yet been considered.

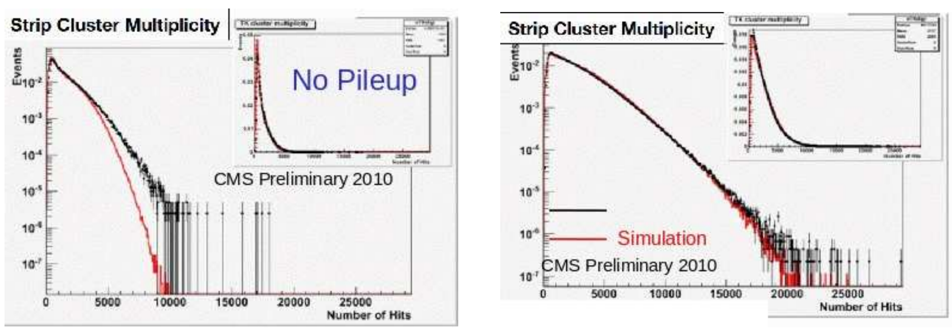


**Figure 2.** Comparison between data and Monte Carlo for occupancy in the silicon strip tracker with or without the effect of pileups.

Figure 2 shows comparison between data and Monte Carlo for tracker occupancy without or with the effect of pileup. The figure indicates that the current implementation of pile-up is very successful at "modest" luminosity . The low-level as well as the high-level objects are well-modeled.

## 3. Detector Specific Requirements and Performance

### *3.1. Tracker*

The requirements for simulating the CMS tracker are the following: (a) a high degree of accuracy in the description of active and passive components, (b) correct, navigable Monte Carlo truth information, (c) a very precise treatment of hard electron bremsstrahlung. The tracker geometry has been very carefully built describing each component in great detail. Each of these components is reviewed with full information from integration centres and then verified by weighing individual components prior to testing the simulation against collision data. Special care is taken in the choice of the physics model of Geant4 for $\delta$-ray production and electron bremsstrahlung. The simulation results have been extensively validated in terms of signal simulation, tracking, dE/dx, etc.

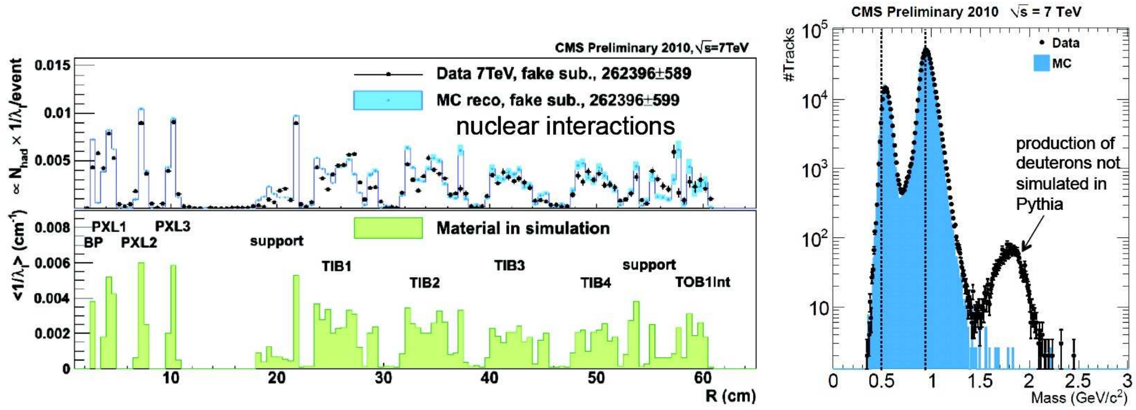


**Figure 3.** (a) Study of the tracker material using nuclear interactions in the data and in the Monte Carlo. (b) Estimation of mass of the charged particles from measurements of the momentum and dE/dx in the data and in the Monte Carlo.

Figure 3 shows some examples of these validation tests. Material content within the tracker is estimated from photon conversion and nuclear interaction. The measurements are found to be consistent. The studies are repeated in data and in Monte Carlo. The figure shows a comparison between data and Monte Carlo for studies of nuclear interaction within the tracker. As can be seen from the figure the frequency of interaction as a function of depth within the tracker is well reproduced in the Monte Carlo. The figure also shows a comparison for particle identification capability between data and Monte Carlo using dE/dx measurements. The data are well reproduced for kaon and protons while the data show a third peak at around 2 GeV. This is absent in Monte Carlo and the discrepancy has been traced back to the generator Pythia which does not produce deuteron in the minimum bias sample.

### *3.2. Electromagnetic Calorimeter*

Simulation of the electromagnetic calorimeter requires (a) an accurate description of geometry and material budget, (b) a good and complete implementation of the electromagnetic physics process. The geometry of the electromagnetic calorimeter is provided with possible independent alignment of modules, super-crystals, wafers etc. The updated distributions of support, cooling, readout are again tested by making specific measurements and finally with collision data. Precise electron and photon identification and the energy measurement require a good understanding of the transverse shower profile (containment, calibrations) as well as longitudinal shower profile (leakage). The simulation has been validated extensively with test beam and collision data for energy measurement and transverse shower profiles. Figure 4 shows a comparison of widths of electromagnetic shower (in the non bending plane) between data and Monte Carlo. The width is narrower in the barrel calorimeter which corresponds to lower noise level. The Monte Carlo distributions match very well with those observed in the data.

### *3.3. Hadron Calorimeter*

Simulation of the hadron Calorimeter requires faithful description of timings, noise as well as energy response as a function of energy. The simulation utilizes shower libraries, noise libraries to overcome some of the limitations (performance, description of different types of noises). The HCAL community made several comparisons for single particle measurements between test beam data (2002-2007) and MC, with different HCAL modules, preceded by real ECAL super-module

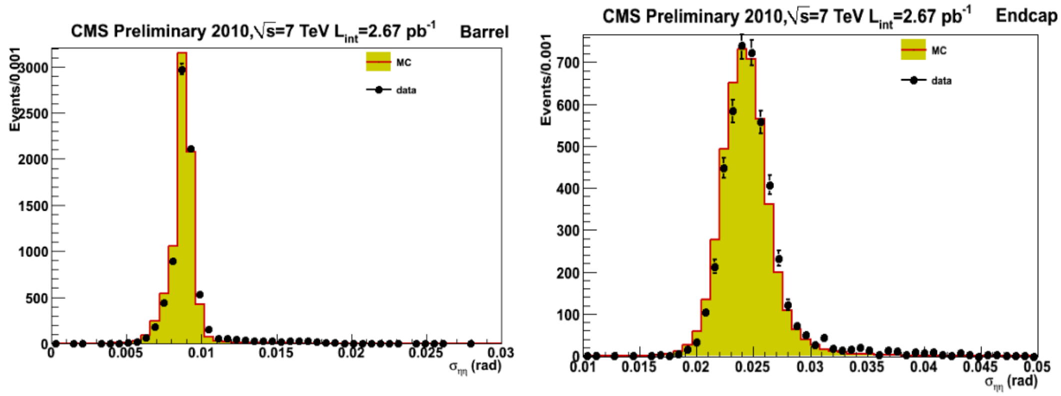


**Figure 4.** Comparison of widths of electromagnetic showers (along $\eta$ direction) between data and Monte Carlo in the barrel and in the endcap electromagnetic calorimeter.

or prototype, to beams of hadrons, electrons and muons over large energy range. The studies on energy resolution and linearity, e/h ratio, and shower profile are instrumental in validating GEANT4 hadronic physics models. Some of these measurements are repeated using isolated charged hadrons in the collision data. The momenta of these particles are well measured by the tracker. To ensure that there is no contamination due to other nearby particles, suitable isolation criteria are applied.

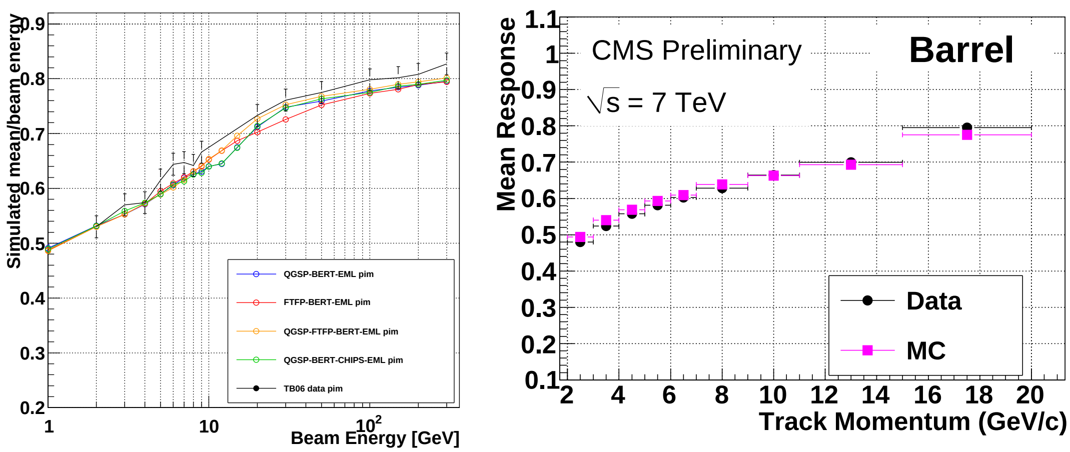


**Figure 5.** (a) Mean energy responses for $\pi^-$ as a function of incident particle momentum. Predictions of the four physics lists QGSP_BERT_EML, QGSP_FTFP_BERT_EML, QGSP_BERT_CHIPS_EML and FTFP_BERT_EML are also shown. (b) Mean response measurements as a function of track momentum in the barrel calorimeter. Monte Carlo prediction is compared with the data.

Figure 5(a) shows mean energy responses for $\pi^-$ as a function of incident particle momentum measured in a dedicated test beam experiment compared with predictions of a number of hadronic models of GEANT4. Figure 5(b) shows mean energy response of isolated charged hadrons measured from 7 TeV collision data in the barrel calorimeter being compared with Monte Carlo predictions with the CMS default physics list. As can be seen from these figures, Monte Carlo provides a good description of the data

*3.4. Muon Detector*
For the muon system, the geometry description in simulation is verified using the Cosmic data collected during the various cosmic runs (MTCC, CRAFT, etc.) and finally from the collision data. Muon physics in GEANT4 is extensively tested and validated in the energy range 10 GeV 10 TeV. An improved description of $\mu$ bremsstrahlung, $\mu$-nuclear effects and a better description of multiple scattering (in agreement of data) have been introduced in recent versions of GEANT4. These new descriptions are validated with earlier simulation and with collision data. Figure 6 shows comparison between data and Monte Carlo for reconstruction and isolation efficiency of muons. Reconstruction efficiency depends on good description of geometry and alignment while isolation depends on estimation of punch through. As can be seen from the figure, all aspects are well described.

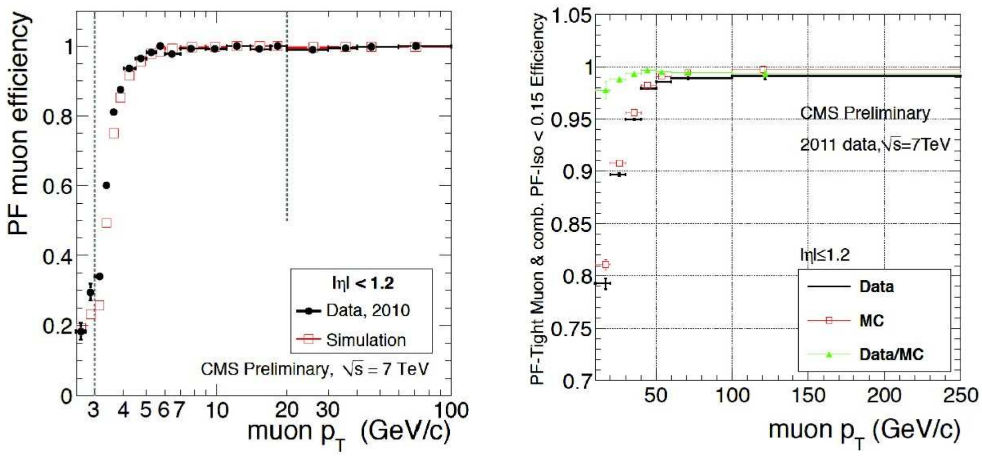


**Figure 6.** Reconstruction and isolation efficiency of muons as a function of momentum are compared between data and Monte Carlo.

*3.5. Forward Detectors*
There are a number of forward detectors in the CMS: Totem telescopes T1, T2, the CASTOR and ZDC calorimeters, Roman pots belonging to the Totem system. Simulation with central as well as forward detectors is done by using a filter to separate particles from event generators to be processed through central and forward detectors. A separate transport code "Hector" [7] is used to transport particles within acceptance of forward detectors close to the forward detectors. Beam interactions creating beam halo effect are taken from a library obtained using the shower code of MARS [8]. The particles in the central detector and also in the forward detector region are transported using GEANT4 and all the simulated hits are combined to get the description of an overall event.

## 4. Abnormal Calorimeter Hits
Rare but anomalous high energy deposits are observed in the calorimeters. Isolated single crystal hits are seen in the barrel ECAL calorimeter (EB) and long tails are observed in the distributions of energy deposit in the forward hadron calorimeter (HF). These tails are seen in either short or long fibres. Figure 7 shows evidence of two types of anomalous energy deposits. The plot on the left shows distributions of energy deposits in the barrel electromagnetic calorimeter where a clear excess is seen at the high energy tail. The plots on the right shows the correlation of

energy deposits between the long and the short fibres where large tails are seen with only one of the fibres fired.

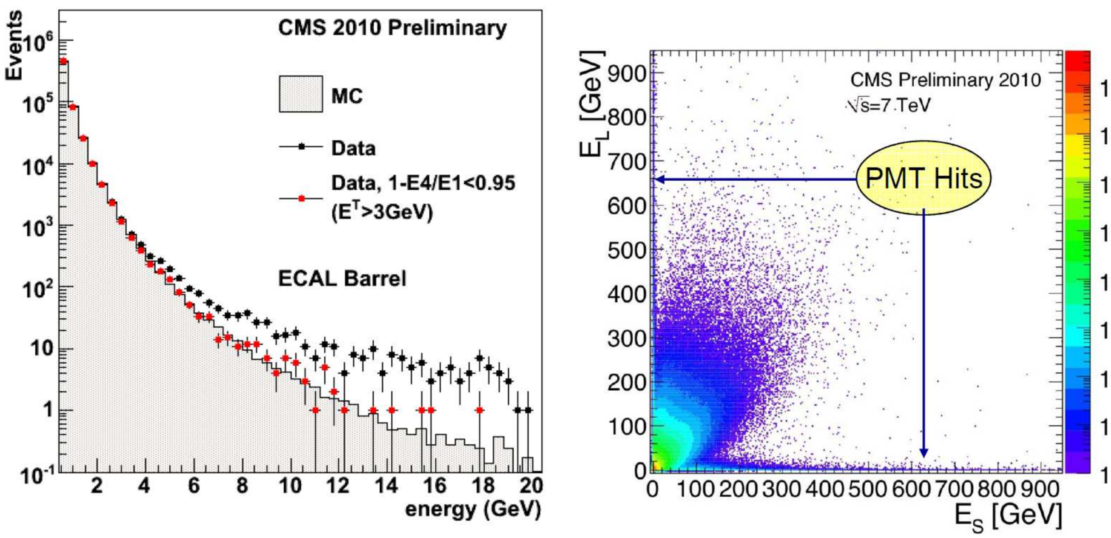


**Figure 7.** Distributions of energy deposit in the barrel ECAL (left) compared to the prediction of Monte Carlo. The red points refer to the data after the application of the noise filter. The plot on the right refers to a scatter plot of energy deposits in the long and in the short fibres.

Filters based on topological properties and timings are developed to remove these abnormal hit events. These tools are very effective in eliminating the abnormal hits from the data sample. For example, the left plot of Figure 7 shows the data distribution after the noise filter where one observes good agreement between data and Monte Carlo. However, one needs to understand the source of the hits and the effect of these in a crowded environments.

Anomalous crystal hits in the barrel electromagnetic calorimeter are identified to be due to energy deposits in the thin silicon layers of the APD's (induced by heavily ionizing particles produced in the epoxy layers by neutrons). Anomalous hits in the forward hadron calorimeter are due to Cerenkov light produced by the penetrating component of hadron showers in the fibres and in the windows of the Photo Multiplier tubes sitting behind the absorber.

The silicon layers in the APD and the PMT window material (together with the fibre bundle behind the HF absorber) are treated as sensitive detector in the simulation. An improved neutron transport code is used from Geant4. The resulting simulation code could explain the distribution sensitive to spikes in EB. This variable which compares the energy of 2×2 matrix versus the energy in the trigger crystal has a spike in the data which could not (could) be explained by Monte Carlo without (with) inclusion of energy deposit in the APD (see Figure 8).

The showers in the forward hadron calorimeter are produced by transporting all hadrons within HF using Geant4 hadronic models. The electromagnetic components of showers in HF are replaced using a parametrization. The resulting simulation code explains the measured energy spectrum as well as anomalous hit rates in HF rather well (see Figure 9).

## 5. Summary

CMS simulation, based on recent version of Geant4, is providing adequate service to understand the collision data. CMS has chosen QGSP_FTFP_BERT_EML as the default physics list based on performance criteria (both physics as well as computing).

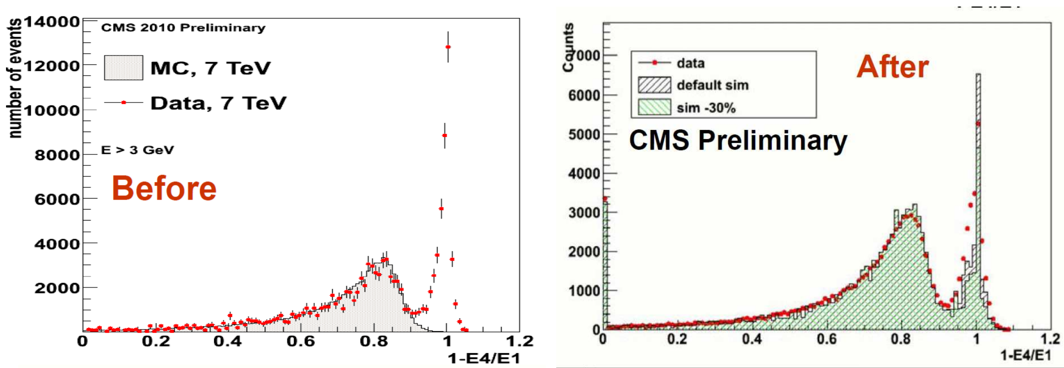


**Figure 8.** Comparison of data and Monte Carlo for the ratio of energy deposit in a 2×2 crystal matrix to the energy in the most energetic crystal. The Monte Carlo distributions are obtained without (left plot) and with (right plot) the inclusion of energy deposits in the thin silicon layer of the APD.

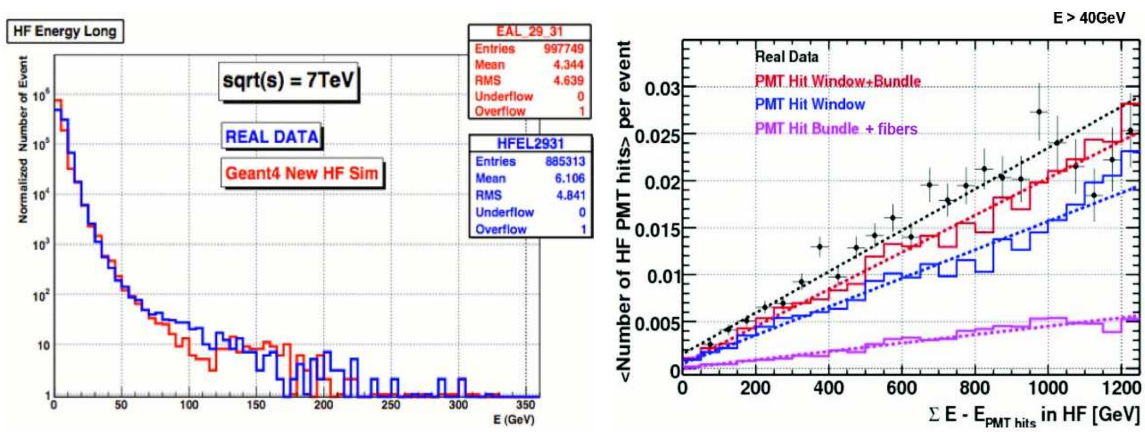


**Figure 9.** The left plot refers to energy measured by the long fibres in HF for collision data at 7 TeV compared to Monte Carlo predictions with hits produced in the absorber, PMT window and fibre bundles. Plot on the right shows the correlation of mean number of PMT hits with energy in the absorber.

CMS has been validating the physics models inside GEANT4 using its test beam as well as collision data. Several physics lists inside the most recent version of GEANT4 provide good agreement of the energy response, resolution of $\pi^{\pm}$ and protons. More work is needed to improve the physics for $K^{\pm}$, anti-protons and hyperons. Electromagnetic physics in GEANT4 gives a good description of shower shapes for electron and photon candidates in the collision data.

Detector geometry and description of in- and out of- time pile-up events are well modeled so that Monte Carlo can give a good description of the data. Full simulation helps in understanding some of the sources of anomalous hits in the detector.

## References


[1] S. Agostinelli *etal.*, Nuclear Instruments and Methods **A506** (2003) 250.
[2] J. Allison *etal.*, IEEE Transactions on Nuclear Science **53** (2006) 278.
[3] *http://twiki.cern.ch/twiki/bin/view/CMS/WorkBook*
[4] M. Dobbs and J. B. Hansen, Computer Physics Communications **134** (2001) 41.

[5] T. Sjöstrand, S. Mrenna, P. Skands, PYTHIA 6.4 Physics and Manual, JHEP **0605** (2006) 026.
[6] T. Sjöstrand, S. Mrenna and P. Skands, A Brief Introduction to PYTHIA 8.1, Comp. Phys. Comm., **178** (2008) 852 [arXiv:hep-ph/0170.3820].
[7] X. Rouby *etal.*, Journal of Instrumentation **2** (2007) P09005.
[8] N. V. Mokhov, S. I. Striganov, Fermilab-Conf-07-008-AD (2007).