Readiness of CMS Simulation towards LHC Startup

Sunanda Banerjee

Fermilab, P.O. Box 500, Batavia, Illinois 60510, USA

E-mail: `sunanda@fnal.gov`

Abstract. CMS experiment has used detector simulation software in its conceptual as well as technical design. With the detector construction near its completion, the role of simulation has changed toward understanding collision data to be collected by CMS in near future. CMS simulation software is becoming a data driven, realistic and accurate Monte Carlo programme. The software architecture is described with some detail of the framework as well as detector specific components. Performance issues are discussed as well.

1. Introduction

CMS experiment [1] has been using simulation software from its inception since 1992. Simulation software has helped to make a conceptual design of the experiment. The technical design reports of all the major detector components, trigger and data acquisition system have made heavy usage of the simulation software. Now that the detector components are all being built and the experiment is getting ready to take data shortly, the role of the software is changing its nature.

Monte Carlo samples in CMS will now be used to

(i) generate large amounts of signal and background events for use in physics analysis,

(ii) understand and demonstrate analysis procedures and methods based on data to derive calibrations, efficiencies, resolutions for high level physics objects (the closure tests of data based methods),

(iii) derive directly calibrations, efficiencies, resolutions for high level objects in cases where data are biased or not available.

A data driven realistic and accurate Monte Carlo programme is an essential tool.

Simulation effort started as a FORTRAN software package CMSIM [2] using the toolkit GEANT3[3]. It has evolved to the current design through four generation of simulation programmes. The two subsequent versions, OSCAR[4] as well as the current version embedded in the CMSSW framework [5] is written in C++ and is based on the tool kit GEANT4 [6].

Though in operation for a number of years, it is a live system. Goals, requirements and tools are evolving and will evolve throughout the life time of the experiment. The current version is based on the version GEANT 4.8.2 and in transition for the version 4.8.3 and then for 4.9.0.

In this report, the current status of the CMS simulation software is discussed. Section 2 will discuss the framework and its interface with GEANT4. The detector specific components are described in Section 3. Integration and performance issues are discussed in Section 4 and Section 5 provides the summary and outlook.

2. Framework and Interface with Geant4

The toolkit GEANT4 provides the following functionalities:

- physics processes for electromagnetic as well as hadronic interactions,
- tools for detector geometry and sensitive detector response,
- interfaces for tuning and monitoring particle tracking.

The new CMS offline framework and event data model [5], on the other hand,

- manages application control at run time,
- relies in the concept of event processing module (EDProducer),
- interface to common tools (Generators, Magnetic Field, Monte Carlo truth handling, infrastructure for Hits, Event Mixing, Digitization, · · ·),
- ensures provenance tracking and event immutability.

The communication among the independent modules of simulation (hit generation), digitization, reconstruction, etc. is done through the event bus which starts with event generation information and then adds to itself simulation hits, digitization data, reconstructed objects, · · ·. The various modules are guided by asynchronous data in the form of event setup information consisting of geometry, calibration and alignment data.

The core application is a framework based Event Data Producer and a customized `RunManager` as interface between GEANT4 and CMS event data model. The geometry record is created by a common source and is made available to both simulation as well as reconstruction. It uses XML-based Detector Description (DD) machinery, configurable at run time via a hierarchy of XML files. It converts solids and materials defined in the DD framework to GEANT4 counter parts. Sensitive detectors are associated with geometrical volumes through XML configuration files at run time.

Magnetic field description is based on dedicated geometry of magnetic volumes. This is provided by an independent sub-system again via the Event Setup. Selection of the magnetic field as well as the tuning parameters for propagating the particles are configurable at run time.

GEANT4 provides a large number of physics lists (LHEP, QGSP, QGSP_EMV, QGSC, FTFP, · · ·) for modeling physics processes. CMS software is flexible to select one of these physics processes at run time. Different production cuts on secondary particles are applied in different regions and they are again configurable at run time.

A large number of event generators (e.g. particle guns, PYTHIA, HERWIG, ALPGEN, · · ·) are interfaced to the simulation programme. The generator can be chosen at run time or can be run in a stand alone job and the output stored in HEPMC [7] format as a POOL [8] file.

Several notification classes are defined specific for begin or end of run, begin or end of event, begin or end of track, step, etc. Using a watcher mechanism, user can access information of GEANT4 objects at any stage with the help of these notification classes. This is extremely useful for monitoring, diagnostics, tuning and custom book keeping purposes.

The framework also allows to keep Monte Carlo truth record with history of decay/interaction of the generator particles and some selected tracks from GEANT4 simulation.

The framework also supports the possibility of event mixing and simulation of the detector electronics. During LHC running, there will be on average 3 inelastic collisions during the low luminosity runs and 25 inelastic collisions during the high luminosity runs. Detector electronics usually integrate the signal over a larger time window than the time interval between bunch crossings. So there are effects due to in-time pile-up between signal and minimum bias events as well as out-of-time pile-up coming from bunch crossings before and after the trigger event.

Currently the simulation programme is executed for minimum bias events first and a library of simulated hits is created. A bunch crossing scenario is created by adding the simulated hits

from the signal event and hits from a number of minimum bias events from the pre-generated library. To have the effect of out-of-time pile-up, the timing of the hits are suitably modified during this merging process.

Simulation of read out electronics, noise etc. is carried out by independent modules for each subsystem. Each of these modules takes simulated hits from a mixed event as input, computes detector dependent effects and adds digitized data to the event for the specific component.

## 3. Detector Specific Components
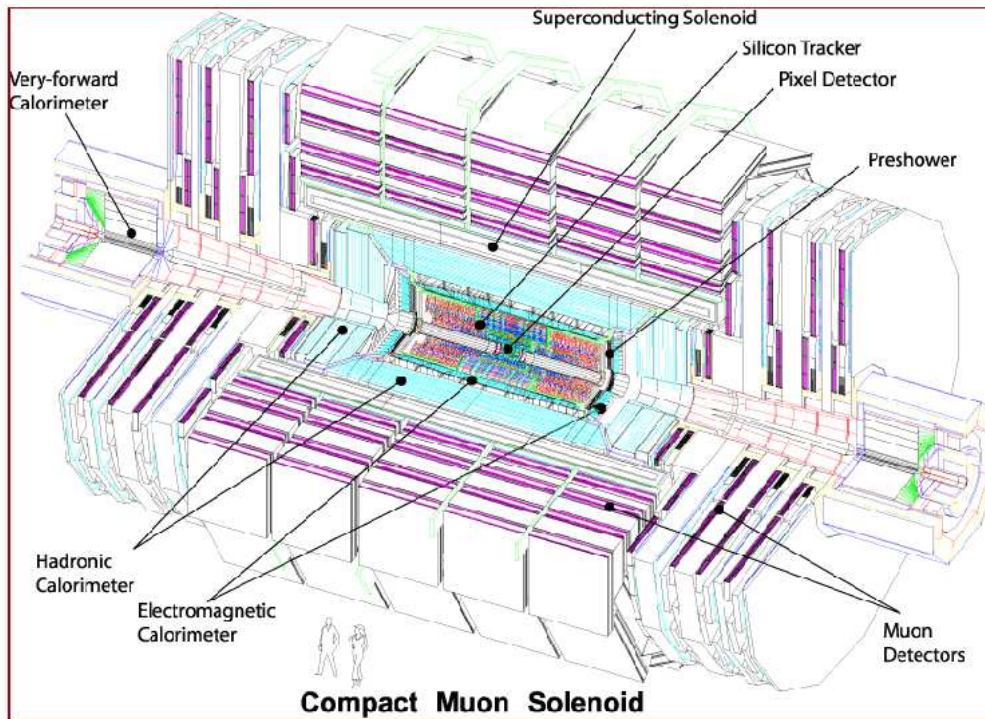
Figure 1. Perspective view of the CMS detector showing different sub systems

Compact Muon Solenoid (CMS) is a large complex detector system consisting of several subsystems each having its own different simulation requirement. This demands a region based optimization procedure which is described here. Figure 1 shows a view of the CMS detector. It is 22 m long, 15 m in diameter and weighs approximately 12.5 kilo tons. It is described within the simulation programme through 1.3 million of geometrical volumes using many complex shapes and 400 materials provided by GEANT4.

3.1. Tracker

The tracker subsystem demands a high degree of accuracy in the description of its active passive and active components. Recently each component of the tracker has been reviewed with full information from the integration centres. An exercise of weighing individual components and comparing the corresponding weight of the simulated component is performed. The agreement in weight between the real and simulated tracker is within 10% and the difference is being examined in detail. Figure 2 shows average number of radiation length as a function of pseudo-rapidity (η) within the CMS tracker. The individual contributions of support material, cables, cooling, electronics, sensitive and other components are shown by different shaded regions.
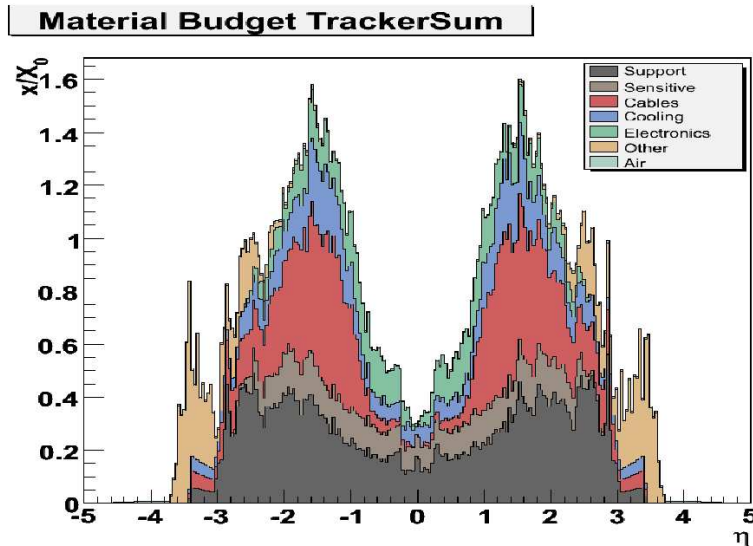

Figure 2. Material budget of the tracker plotted as a function of pseudo-rapidity, η. Component-wise decomposition is shown.

Tracker simulation also demands a correct and easily navigable Monte Carlo truth. For tracking electrons, it also needs a careful treatment of hard bremsstrahlung which has been requested to GEANT4 technical forum and has been provided by GEANT4 since version 4.7.1. The simulation has been extensively validated with Cosmic Ray data collected at the Tracker Integration Facility at CERN. Particular attention has been given to noise simulation, capacitive coupling, cluster charge, dE/dx, electron to ADC conversion. After correcting with the best knowledge of sensor thickness, the agreement between data and Monte Carlo is reasonable for TIB (Tracker Inner Barrel) modules as shown in Figure 3. The peak position is well reproduced, the width is slightly larger in the data and the discrepancy is under investigation.

3.2. Electromagnetic Calorimeter

Electromagnetic calorimeter (ECAL) simulation demands a very accurate description of geometry and material budget. Recently the geometry code has been reviewed and the old geometry description is now replaced with an algorithmic description of all the components, the barrel and the endcap parts as well as the preshower detector. This description defines super-modules (barrels) and super-crystals (endcap) as independent objects and there is a possibility
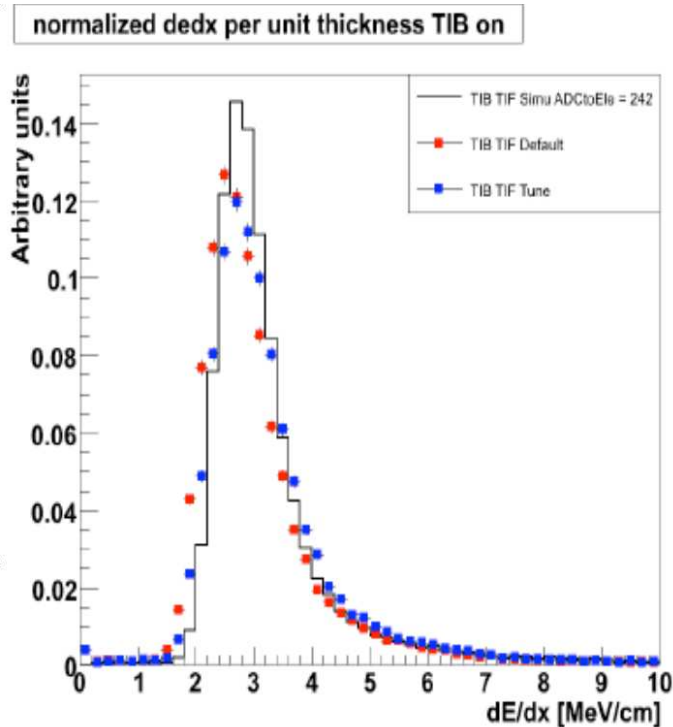
Figure 3. dE/dx distribution measured for TIB modules in the Tracker Integration Facility is compared with Monte Carlo simulation (shown by the solid histogram).

of independent alignment of these components. There is also an updated and more accurate description of support, colling, readout components, etc.

There is a demand for good and complete implementation of the physics processes, in particular the electromagnetic shower code. This is needed to understand the containment of showers both laterally and longitudinally. This knowledge is critical for understanding calibration issues of the calorimeter.

There has been an extensive test beam activity where the super modules are exposed to electron beams. The transverse shower profiles are measured and compared with simulation results. Figure 4 shows a plot of the ratio of energy deposit in the hottest crystal to the sum of energy measured in a 5×5 matrix surrounding it as a function of the crystal index measuring pseudo-rapidity (η). This measures the transverse shower profile and the data are compared with predictions from two versions of GEANT4. To obtain a realistic prediction from the simulation the beam line is also described to a high degree of accuracy. The more recent version of GEANT4 (4.8.2) gives a closer agreement to the data and the difference is less than 0.5%.

3.3. Hadron Calorimeter

Hadron calorimeter (HCAL) system, designed to make an accurate measurement of the hadronic jets, requires an accurate description of hadron as well as electromagnetic showers of the combined (electromagnetic + hadron) calorimeter system. Hadron calorimeter group undertook an extensive test beam studies between 2002 and 2007 with electron, muon and hadron beams covering a wide energy range from 2 GeV/c to 300 GeV/c. The beam line is equipped with detectors to identify the particle type to a high degree of accuracy. During 2006, the barrel hadron calorimeter prototype is preceded by one barrel ECAL supermodule and during 2007,
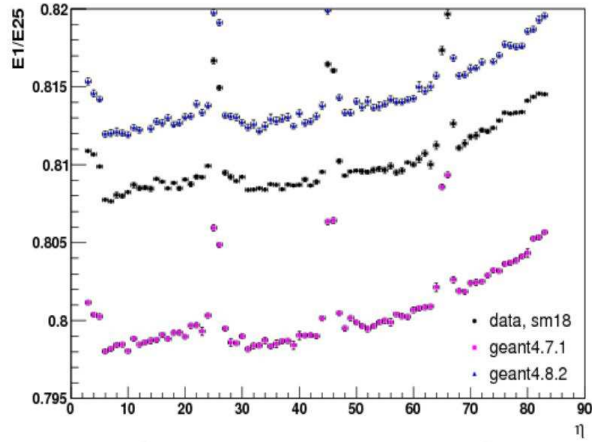
Figure 4. Ratio of energy deposit in the most energetic crystal to the sum of energy measured in a 5×5 matrix surrounding it as a function of the crystal index measuring pseudo-rapidity (η).

the endcap HCAL prototype has endcap ECAL super-crystals and layers of preshower detector preceding it.
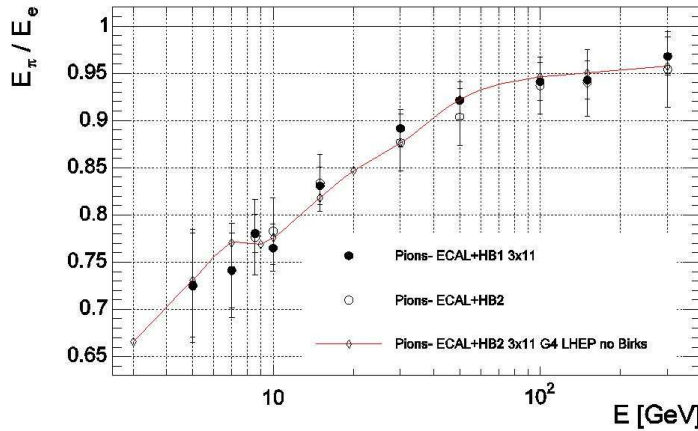
Figure 5. Measured π to electron response measured in the two HCAL barrel prototypes as a function of incident particle energy. The results are compared with prediction from GEANT4 LHEP physics model as in version 4.6.2.

Figure 5 shows a comparison of π to electron response ratio as a function of beam momentum measured in the 2004 test beam setup to LHEP physics list of GEANT4.6.2. The energy dependence is described reasonably well with the GEANT4 version. Measurements exist for energy response, energy resolution, transverse and longitudinal shower profiles. The low

energy hadron data show discrepancy in the energy deposit in the electromagnetic calorimeter. Figure 6 shows the "banana plot" - two dimensional distribution of energy deposits in hadron and electromagnetic calorimeter. GEANT4 model predicts too high energy deposits in the electromagnetic calorimeter. This is true for parametric model like LHEP and also for microscopic models (QGSP, QGSP_BERT, ···).
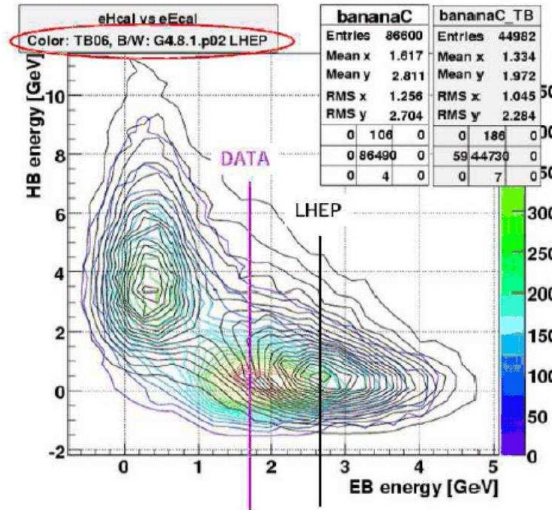
Figure 6. Energy measured in hadron calorimeter as a function of energy measured in electromagnetic calorimeter for 5 GeV/c pion beam run. The coloured contours are from the data and solid lines are from LHEP physics model as in version 4.8.1.

Hadron calorimeter simulation also needs faithful description of timings, noise, etc. Analysis of the Cosmic ray data has revealed additional noise and cross talks in the readout system. Simulation of hadron calorimeter now uses shower library, noise library, etc which are the first steps toward a data driven Monte Carlo system.
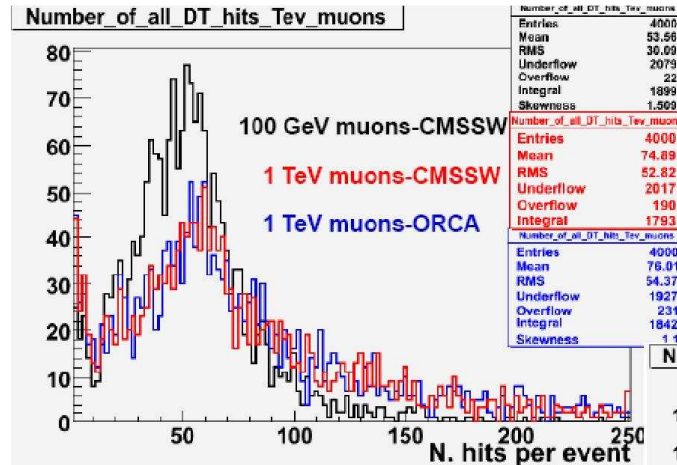
3.4. Muon System

Geometry of the muon detector system has been extensively tested using the Magnet Test and Cosmic Challenge (MTCC) data collected with a slice of CMS when the CMS magnet was operated at it full capacity (4 Tesla) for the first time.

Muon physics as described in GEANT4 has been extensively tested with data and validated over a broad energy range (up to 10 TeV). GEANT4 provides an improved description of μ bremsstrahlung, μ-nuclear effects as compared to its predecessor GEANT3. Treatment of multiple scattering has also been improved and the new results are in agreement with data.

Figure 7 shows number of hits per event in the drift tube chambers for 100 GeV/c and 1 TeV/c muons. The predictions are compatible with earlier simulation results and they are now being compared with the data.

3.5. Forward Detectors

CMS has a large number of detectors in the forward region. These are essential for studying diffractive physics and heavy ion interactions. The forward detectors comprise of (1) forward hadron calorimeter (HF) located at $\pm$ 11.5 m from the interaction point covering the angular region ($3 \leq |\eta| \leq 5.2$), (2) telescopes T1, T2 located at $\pm$ 7.5 m and $\pm$ 14.5 m, (3) CASTOR

Figure 7. Number of hits in the drift tube chamber of the muon system for 100 GeV/c and 1 TeV/c muons in the recent version of simulation software. Prediction for 1 TeV/c muons from earlier version of CMS simulation is also shown.

calorimeter at $z = \pm$ 14 m ($5.2 \leq |\eta| \leq 6.6$), (4) Zero Degree Calorimeter (ZDC) at $z = \pm$ 140 m ($8.3 \leq \eta$), (5) Roman pots at $z = \pm$ 147 m, $\pm$ 220 m, (6) a calorimeter system at 420 m. They use different technologies and serve for a tracking device or a calorimeter.

Simulations of stand alone systems exist for some time. Results from these simulations are compared with test beam studies for energy resolution, shower leakage, etc and are validated. Integrating these simulation softwares with that of the central CMS detector needs a slightly modified framework. The following procedure is foreseen:

- use a filter to separate particles from event generators to be processed through the central and the forward detectors
- use a separate transport code HECTOR [9] to transport particles filtered for forward detectors to a region close to the forward detectors
- add particles from beam halo from a library obtained from a separate simulation system MARS [10]
- trace the particles in the central as well as the forward detectors using GEANT4
- combine all the simulation hits to get the overall event

All the components of this scheme have been tested and the integrated approach will be brought to practice soon.

4. Integration and Performance Issues

There are several issues in integrating the software components, specifically in view of a data driven Monte Carlo system: (1) standardizing the event generator formats including particle identifiers, (2) mechanism for overlaying real events over Monte Carlo events (3) parameterizing showers using real data.

CMS is currently investigating how to substitute the library of hits from pre-simulated minimum bias events with one from real data to have a more realistic effect. Here the merging will be done not at the level of simulated hits but at the level of digitized information. However, the same framework can be reused for this purpose.

GFLASH model based on three probability density functions is used for parametrizing showers of electrons, positrons and photons. The parametrization can be switched on in some specific detector regions. In its current implementation this can be done in the barrel and endcap electromagnetic calorimeter. Figure 8 shows a comparison of simulation results for 50 GeV photons in the barrel electromagnetic calorimeter between parametrized shower models and full simulation. As can be seen from the figure, total energy deposits as well as shower profiles are well reproduced in the parametrized model.
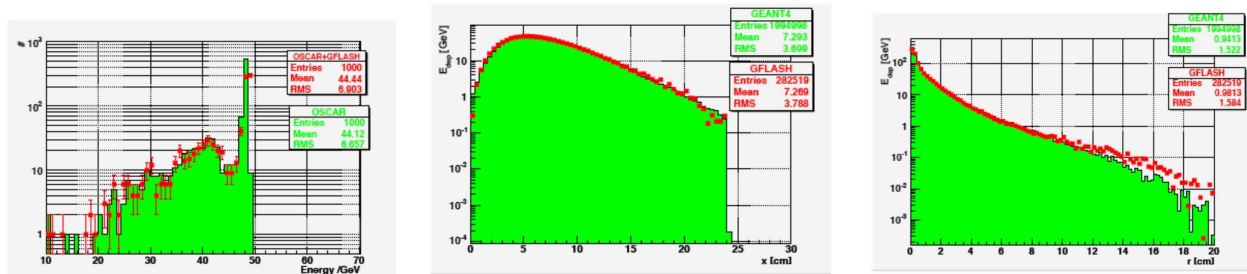
Figure 8. Comparisons of parametrized and complete simulation of electromagnetic shower from 50 GeV photons in the barrel electromagnetic calorimeter for (a) total energy deposit, (b) longitudinal shower profile, (c) transverse shower profile.

The framework is now extended to incorporate parametrization of hadronic showers. Using a wrapper class parametrized physics model can be implemented at the first inelastic hadron collision. The model is in the process of implementation and it will go through an extensive tuning programme.

Simulation software needs to be validated for the software components and also its physics contents. Physics content is validated by comparing the results of simulation with dedicated test beam or Cosmic Ray data. A dedicated framework of validation code has been written to ensure simulation (or other) software reliability. The exercise of validation is executed at every software release and/or at change of a GEANT4 version. Tools exist for validating geometry, checking overlaps, simulated hits, digitization and reconstructed hits/objects of individual detectors and also of the over all system.

With the new upgraded software, nearly 250 million events have been simulated by the production team since July 2006. 100 million of these events used GEANT4.7.1 during the computing challenge CSA06 and validation efforts, The remaining sample are generated using GEANT4.8.2 during the computing challenge CSA07. Extensive tests and validation are carried out for the transition of the GEANT4 version.

Failure rate in the production jobs is dominated by Grid related problems. The application software is very robust and failure rate here is better than 1 in 10^4. These are due to some arithmetic problems which are trapped and the event is skipped. There is hardly any crash due to segmentation violation. Memory leakage has been monitored and the simulation programme is found to have constant memory for thousands of events.

Using an Intel(R) Xeon(R) CPUE5335 processor at 2.00 GHz with 2 GB memory, the simulation programme takes 23 second on average for a minimum bias event and 170 second for a $t\bar{t}$ event. This speed refers to GEANT4 version of 4.8.2.p01 and QGSP_EMV physics list.

CMS has the strategy of equal number of simulated and real events and one expects to have approximately 1.5×10^9 events in a year. The current plan is to achieve this goal with a proper mixture of full and fast simulation.

5. Summary

A GEANT4-based Object Oriented simulation has been successfully implemented for the CMS experiment. The codes are now ported to a new framework. The results are widely validated and the success as well as shortcomings are reported back to the GEANT4 team. The software package has been used for detector and physics studies.

The simulation software is proven to be robust, powerful and maintainable. It is capable to fulfill the emerging requirements of the experiment. However, the project is still not complete and several developments are foreseen in near future. Geometry of the detector is yet to be finalized. One needs to simulate simultaneously the forward and central detectors. Shower parametrization is being extended for hadrons and a real data overlap tool is under construction. Effort is underway to improve the performance in terms of speed, robustness, memory utilization, etc. CMS is expected to have a robust, reliable simulation programme to meet the challenge of data coming from physics collisions.

References

[1] CMS: The Compact Muon Solenoid, CERN/LHC-94-38.
   CMS Physics Technical Design Report, CERN/LHCC 2006-001.
[2] CMSIM: *http://cmsdoc.cern.ch/cmsim/cmsim.html*
[3] R. Brun *etal.*, GEANT3 Users Guide, CERN Program Library W5013
[4] OSCAR: *http://cmsdoc.cern.ch/OSCAR/*
[5] CMSSW: *http://twiki.cern.ch/twiki/bin/view/CMS/WorkBook*
[6] GEANT4: *http://geant4.cern.ch*
[7] HEPMC: *https://lcgapp.cern.ch/project/simu/HepMC/*
[8] POOL: *https://pool.cern.ch/*
[9] HECTOR: *http://www.fynu.ud.ac.be/themes/he/ggamma/hector/index.html*
[10] MARS: *http://www-ap.fnal.gov/MARS/*