Available on CMS information server

CMS CR -2009/090

The Compact Muon Solenoid Experiment

Conference Report

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

12 May 2009 (v2, 13 May 2009)

Data Driven Approach to Calorimeter Simulation in CMS

Sunanda Banerjee (on behalf of Calorimeter Simulation Task Force for CMS)

Abstract

CMS is looking forward to tune detector simulation using the forthcoming collision data from the LHC. CMS established a task force in February 2008 in order to understand and reconcile the discrepancies observed between the CMS calorimetry simulation and the test beam data recorded during 2004 and 2006. Within this framework, significant effort has been made to develop a strategy of tuning fast and flexible parameterizations describing showering in the calorimeter with available data from test beams. These parameterizations can be used within the context of Full as well as Fast Simulation. The study is extended to evaluate the use of first LHC collision data, when it becomes available, to rapidly tune the CMS calorimeter.

Presented at *Computing in High Energy and Nuclear Physics (CHEP09),March 23-27, 2009,Prague,Czeck Republic,15/05/2009*

Data Driven Approach to Calorimeter Simulation in CMS

Sunanda Banerjee
(on behalf of Calorimeter Simulation Task Force for CMS)
Fermilab, P. O. Box 500, Batavia, Illinois 60510, USA

E-mail: sunanda@fnal.gov

Abstract. CMS is looking forward to tune detector simulation using the forthcoming collision data from the LHC. CMS established a task force in February 2008 in order to understand and reconcile the discrepancies observed between the CMS calorimetry simulation and the test beam data recorded during 2004 and 2006. Within this framework, significant effort has been made to develop a strategy of tuning fast and flexible parameterizations describing showering in the calorimeter with available data from test beams. These parameterizations can be used within the context of Full as well as Fast Simulation. The study is extended to evaluate the use of first LHC collision data, when it becomes available, to rapidly tune the CMS calorimeter.

1. Introduction

The Compact Muon Solenoid (CMS) experiment [1] uses two independent packages for simulating response of the detector system. "FullSim" [2] refers to a detailed detector simulation using the GEANT4 toolkit [3]. The detector geometry is described in detail and the particle interactions with the detector material is calculated from first physics principles. FullSim can take several minutes of CPU time per event. "FastSim" [4] refers to a detector simulation using a simplified parameterization of the particle response and typically takes less than a second per event. In order to analyze the early collision data from the Large Hadron Collider (LHC) and to have reliable physics results, the detector simulation should agree with the data to within 5% for kinematic distributions obtained from Standard Model enriched control samples.

The CMS calorimeter response was measured in a number of test beam studies using prototypes and final detector modules in the CERN SPS beam lines. Response to electrons in the electromagnetic calorimeter was studied in the H4 beam line [5] while the response of the combined calorimeter system (mainly to hadron beams) was studied in the H2 beam line [6, 7]. The comparisons of the measured response to the full GEANT4 detector simulation showed some significant discrepancies.

A task force was established by the CMS management to investigate how the simulation can be improved to better describe the test beam results. The charge of the task force is

(i) to evaluate and "fix" or tune the shower models used by GEANT4 to improve the agreement between the results of FullSim and the test beam data for the linearity of the response, the resolution, and the shower shape;

(ii) to develop a GFLASH [8] based parameterization for the electromagnetic and hadron shower shapes within the framework of FullSim and to tune this to the test beam data;

(iii) to tune the parameterization of the electromagnetic and hadron showers in FastSim to match with the results of the Full Simulation and to the test beam data;

(iv) to provide a concise strategy to tune both the Full and Fast Simulation to collider data where the strategy will include the specification of a trigger path to record the necessary data as well as the tools for analysis and tuning of the simulation.

The work of the task force is summarized here where the program of tuning, validation, and cross-validation of the CMS calorimeter simulations is outlined. The retuning of the simulation, using the procedures and tools developed by the task force, will need to be performed with the first LHC collision data in order to respond to the needs of the physics research program of CMS. Many of these improvements have been included in the CMS standard software system CMSSW [9].

2. Geant4 based Full Simulation

The Full Simulation software for CMS can choose at run time the physics processes of Geant4 to be used during the simulation. CMS has been using the physics list QGSP. This physics list utilizes the Quark-Gluon String model (QGS) [10] at high energies with the PreCompound model [11] for nuclear de-excitation and the Low Energy Parameterized model (LEP) [12] for low energies. The test beam setups are also described within the same framework and simulation results are compared with results obtained from the data.

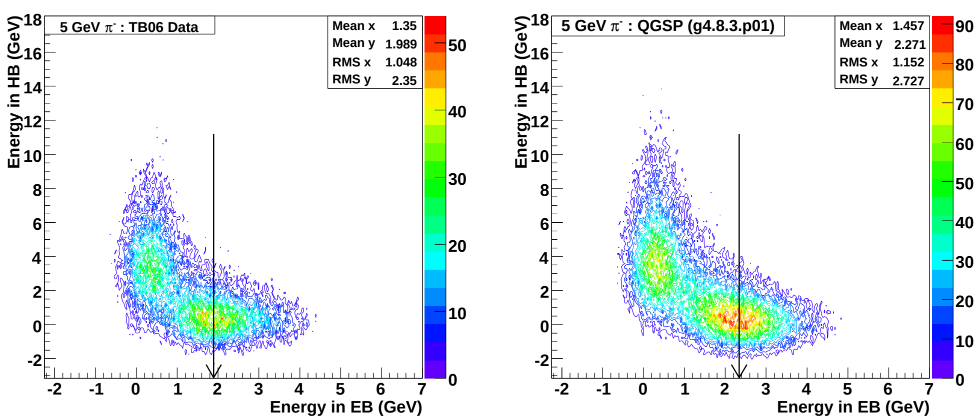

Figure 1. The measured energy in HCAL vs ECAL for 5 GeV π^- in the data (left), Geant4 prediction (right) using the QGSP physics list. The simulated energy in ECAL is higher than what is observed for the test beam.

Figure 1 shows a comparison in the energy measurement in the hadron (HCAL) and electromagnetic (ECAL) calorimeters between simulation and data for 5 GeV/c π^-. The data show energy peak in ECAL at a much smaller value than in the simulation. The lower response in the data may be explained by the effects of saturation in scintillation light emission as observed in organic scintillator by Birks [13]. Saturation effects have been observed in inorganic crystals like BGO and BaF_2, however, they have not been experimentally studied in $PbWO_4$ crystals which constitute the sensitive part of the ECAL for CMS. Enabling Birks' law in the simulation using the measured coefficient for BGO gives a better description of the measured energy in the ECAL. The description of energy deposits in the HCAL alone can be improved by the introduction of Birks' law in the plastic scintillator of HCAL.

Enabling Birks' law and using the QGSP (or QGSP_EMV) physics list results in a significant underestimation of the mean energy measurement. The agreement can be recovered by introducing a physics list that results in larger energy deposits at lower energies such as QGSP_BERT (or QGSP_BERT_EMV). This physics list uses Bertini cascade model [14, 11] for energies below 9.5 GeV. The combination of the new physics list and Birks' law leads to a better agreement between data and simulation as shown in Figure 2 for the mean energy response.

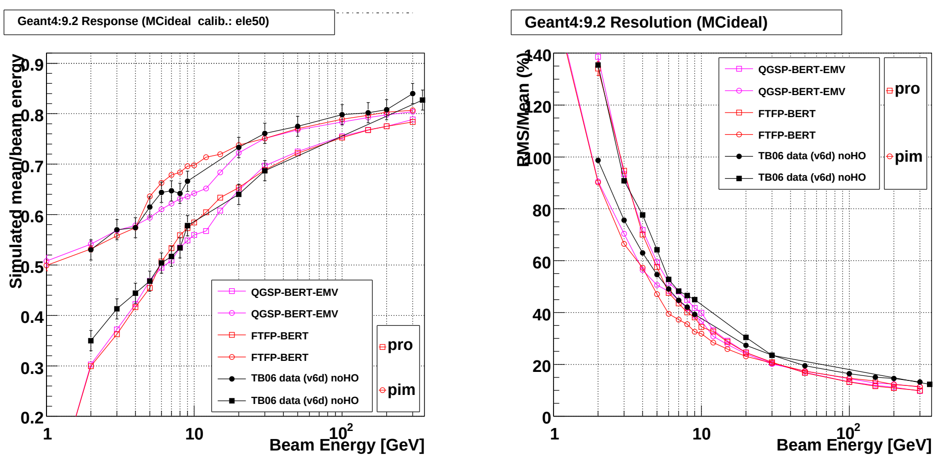

Figure 2. (Left) Mean energy responses of π^- (shown as circles) and protons (shown by squares) are compared with the predictions of QGSP_BERT_EMV physics list (shown by empty points) as a function of beam momentum. (Right) Energy resolutions for π^- and proton are compared with predictions of Full Simulation as a function of beam momentum.

An important consideration is to ensure that the use of Birks' law preserves the already good description of the electromagnetic showers. The effect on the absolute energy scale and on the transverse shape of the shower are studied on electrons samples. The crystal response depends strongly on the electron impact point on the crystal face, therefore only the electrons impacting close to the maximum containment point, defined as the impact point which yields the highest measured energy, of the crystal are used. There is almost no effect on the transverse shape nor on the absolute energy with the parameterization used for $PbWO_4$.

The shower profiles for the hadrons are also compared between data and Monte Carlo. The transverse shower profile, measured in the ECAL (Figure 3), and the longitudinal shower profile, measured with the specially designed HB module (Figure 3) show the new parameterization describes the data better than the previously used physics lists QGSP and QGSP_EMV.

A critical measurement for the calibration of the HCAL is the fraction of MIP (Minimum Ionizing Particle) events in the ECAL as a function of beam energy. The MIP events are selected by the amount of energy measured in the ECAL (below 0.8 GeV). The data show a decrease with energy till $\sim$ 10 GeV, then rather a flat energy dependence and again a smooth decline above 100 GeV. The early version of GEANT4 did not have the smooth decline above 100 GeV and a sharp discontinuity around 10 GeV. The first effect has been identified due to the lack of bremsstrahlung and pair production of high energy hadrons while the second is due to joining of two models (Bertini cascade and low energy parameterization) in the physics lists. In the recent (9.2) release of GEANT4 both these issues have been addressed and the recent comparisons show much better agreement with the data (Figure 4).

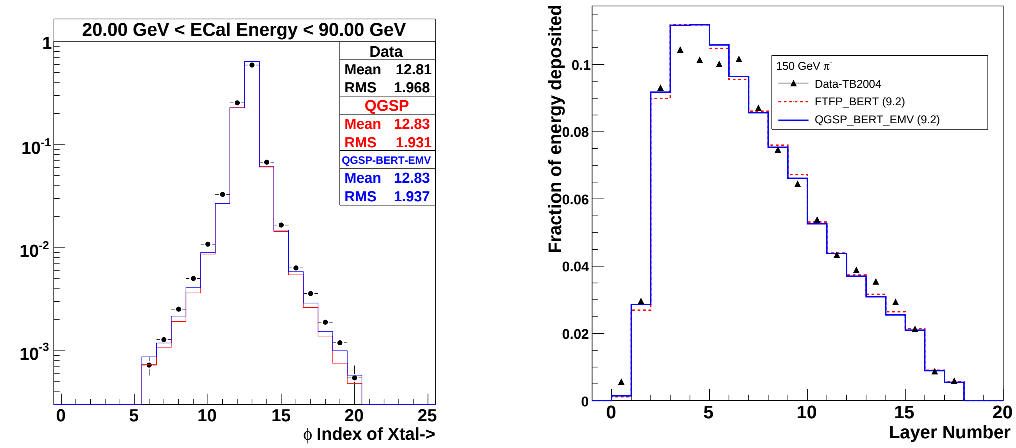

Figure 3. (left) Transverse shower profile of 100 GeV/c π^- and (right) longitudinal shower profile of 150 GeV/c π^- compared with the predictions of GEANT4 using different physics lists.

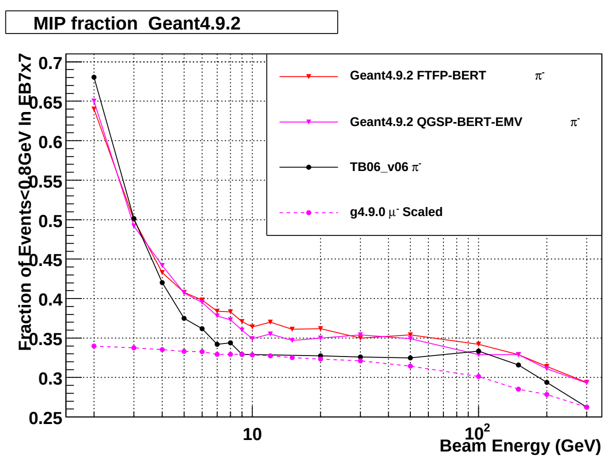

Figure 4. The fraction of π^-s with a MIP like signal in ECAL as a function of beam energy. Test beam results are compared with predictions of GEANT4 using the QGSP_BERT and FTFP_BERT physics lists.

3. GFLASH in the Full Simulation

GFLASH is a parameterized simulation of electromagnetic (EM) and hadronic showers. It uses parameterizations to describe the longitudinal and lateral profiles in homogeneous and sampling calorimeters taking into account individual shower fluctuations and correlations. GFLASH provides a significantly faster calorimeter simulation than detailed GEANT4 based simulations, with greater flexibility of tuning the response to data. Details of the original GFLASH program can be found in [8, 15].

The simulation of showers in GFLASH is divided into the spatial distribution of deposited energy (given by a parameterization) and the energy fraction of the deposited energy which is visible in the active medium. The energy fraction of the deposited energy which is visible in the active medium is determined by the sampling fraction (or scaling factors determined by calibration) and their fluctuations.

In general, the EM shower parameterization can be applied to any homogeneous detector provided one expresses the energy in units of critical energy (E_c) and the length in units of the radiation length (X_o). Since the CMS electromagnetic calorimeter (ECAL) consists of a homogeneous material (namely $PbWO_4$) and is relatively long ($\sim 25X_o$), the original GFLASH approach works for the ECAL shower model without major modifications.

The hadron shower parameterization is far more complicated due to π^0 contributions from both the primary and secondary inelastic hadronic interactions. The original GFLASH parameterization was developed for the H1 calorimeter assuming that the detector is described as a single effective medium within a repetitive sampling structure. The H1 parameterization, especially the longitudinal parameterization, is not suitable for the CMS calorimeter which has a hybrid structure consisting of crystal followed by the hadron calorimeter with some intermediate passive material. A new parameterization for the longitudinal hadronic shower is developed to take into account the specific structure of the CMS detector (see [16]).

The longitudinal and lateral profile of GFLASH are first tuned to those of GEANT4 to verify that the original parameterization is valid for the CMS detector and to study any CMS specific variations. Then, a minimal set of GFLASH parameters are changed to match the EM energy response in the $N \times N$ crystal tower between test beam data and simulated data of GFLASH. The results of the tuning are compared to the test beam data in the form of mean energy deposit in the most energetic crystal E1 or in matrices 3×3 (E9) or 5×5 (E25) in Figure 5. All the energies are normalized to the incident beam energy.

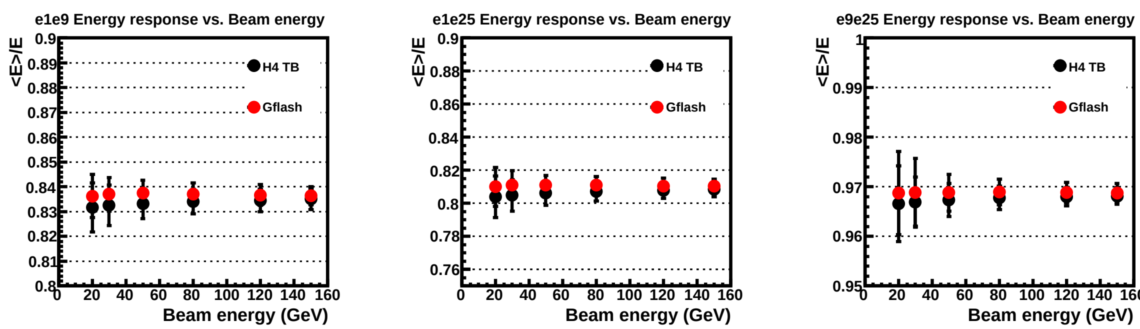

Figure 5. Mean fractional energy deposit in 1×1 (E1), 3×3 (E9) and 5×5 (E25) crystal matrices in the test beam data and tuned GFLASH parameterization.

The hadron showers are treated separately depending on the position of the first inelastic hadronic interaction (within ECAL crystals, inside HCAL or in-between). However, the intrinsic representation of a shower for a given energy in each detector region (ECAL, HCAL) is kept the same irrespective of the starting point. Also the default GFLASH parameterization of longitudinal shower profile is replaced by log-normal distributions.

The hadronic parameterization of GFLASH is tuned to the test beam data following the same procedure as used in tuning the GEANT4 simulation. For each beam energy, the following distributions are compared: the total energy, the energy measured in the electromagnetic and hadron calorimeters and also energy in HCAL when there is a MIP like signal in ECAL. Figure 6 shows a comparison of mean energy response and energy resolution of the finally tuned results with test beam data for π^- beam as a function of beam momentum. As can be seen from these plots, a very good agreement has been obtained. Similar results have been obtained for proton beams as well.

The results of the tuning are incorporated in the CMS simulation program. The remaining issues are the tuning of anti-proton response (which has clearly different response from protons) and treatment of shower leakage. GFLASH parameterization has been applied to physics channels

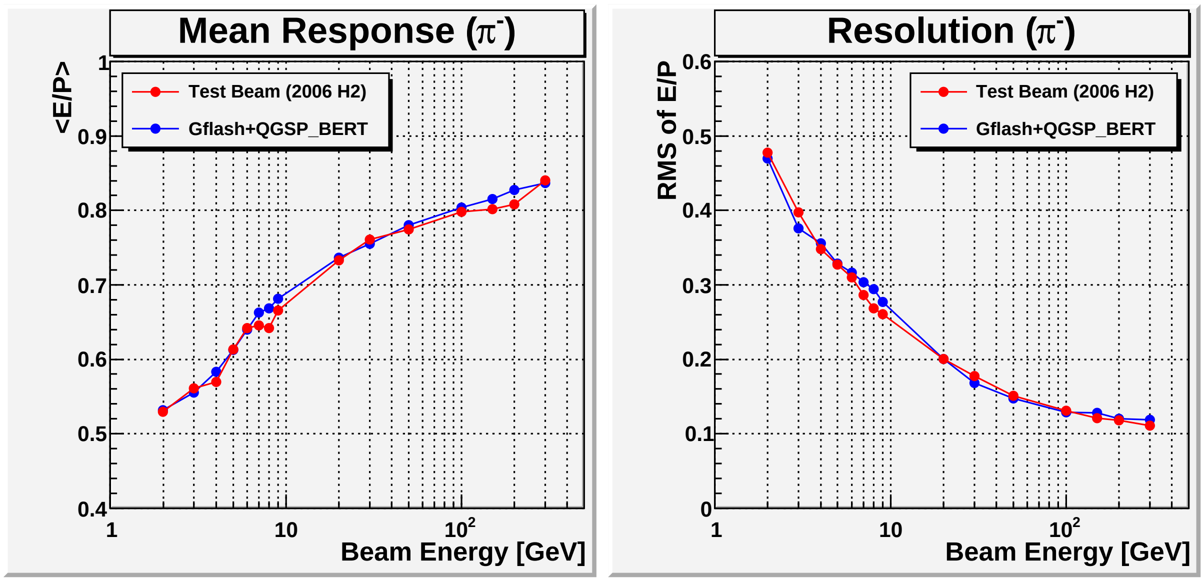

Figure 6. Mean energy response and energy resolution of π^- as a function of momentum from the test beam data and tuned GFLASH parameterization.

like $t\bar{t}$ production or inclusive Z production where Z decays to a pair of e^+e^-. The results from GFLASH compare well with those from GEANT4 as can be seen in Figure 7. One also sees a substantial gain in CPU time ($\sim 50\%$) with the parameterized approach.

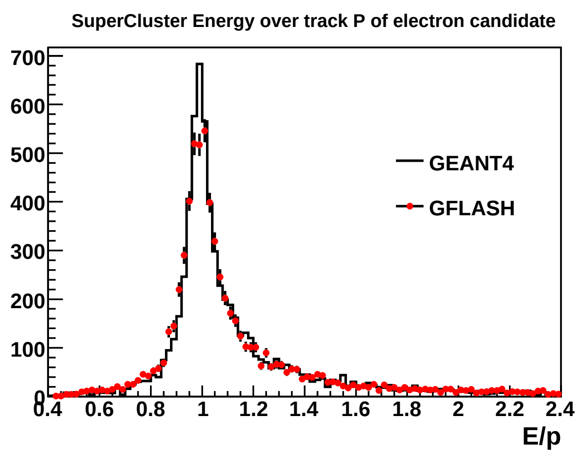

Figure 7. Effective mass distribution of e^+e^- system from simulation of inclusive Z production sample using GEANT4 and tuned GFLASH parameterization.

4. Fast Simulation

FastSim provides a fast and fairly accurate simulation of the CMS detector. The event format after passing through the detector simulation is the same for both FullSim and FastSim and the same offline higher level reconstruction code can be run on the data produced by the two different simulation packages. FastSim also uses the same geometry description as FullSim.

But to achieve very Fast Simulation (a few seconds per event), a number of assumptions and simplified parameterizations describing showering are used. The details of the calorimeter shower development, both electromagnetic and hadronic, is parameterized in FastSim and is being tuned to the detailed GEANT4 CMS simulation.

The electromagnetic shower parameterization in FastSim has been tuned to the test beam data by mimicking the test beam configuration in the FastSim framework. The lateral profile of the electron shower measured in the data are used to tune the simulation. The results of the tuning shows that the original Grindhammer parameterization used in FastSim provides a good description of the the lateral shower spread when its default parameters are used (see Figure 8). The parameterization predicts 83.4% of the energy will be contained in 3×3 crystal window while test beam measures the containment to be 83.5% for 120 GeV electrons. The tail on the low energy side is due to electron bremsstrahlung in the long beam line which is not simulated in the FastSim application.

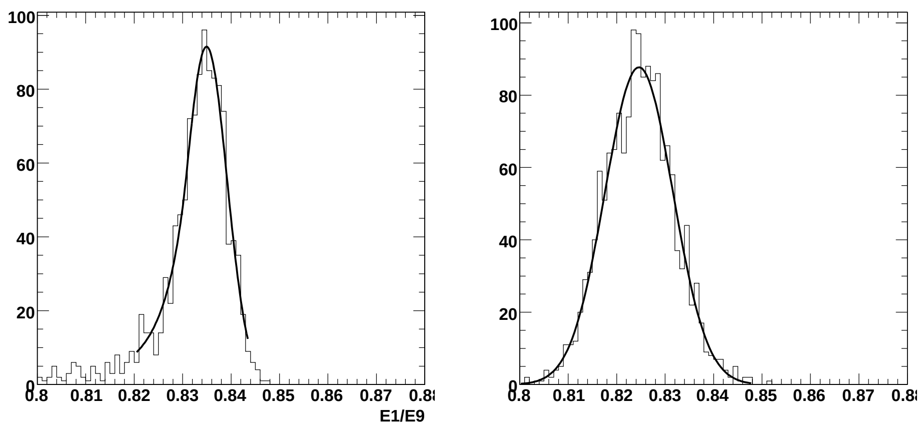

Figure 8. Distribution of the simulated E1/E9 ratio from the test beam data from 120 GeV electron (left) and with FastSim with the default setting of the Moliére radius (right).

Hadronic shower is parameterized with original GFLASH approach with the mean response obtained using FullSim with GEANT4 at several energy points. Also showers starting in ECAL are treated differently from those starting later. The recent tuning to FullSim clearly shows improvement in the agreement (see Figure 9).

Experience gained from tuning FastSim to FullSim (and partially to the test beam data) will be vital in quickly providing a more accurate detector simulation by tuning to the LHC collision data.

5. Tuning to Collision Data

CMS simulation software has been so far tuned with data obtained in the test beam setups. Though the data come from a controlled experimental environment, they will not have some of the effects which are foreseen in the final CMS detector. The main effect will be due to material and magnetic field. Test beam setups often use prototypes and material between ECAL and HCAL or in front of ECAL are either not present or not proper. The magnetic field will cause scintillation brightening and some increased path length inside sampling scintillator (due to bending). In view of this, certain strategy has been drawn to tune the simulation to collision data.

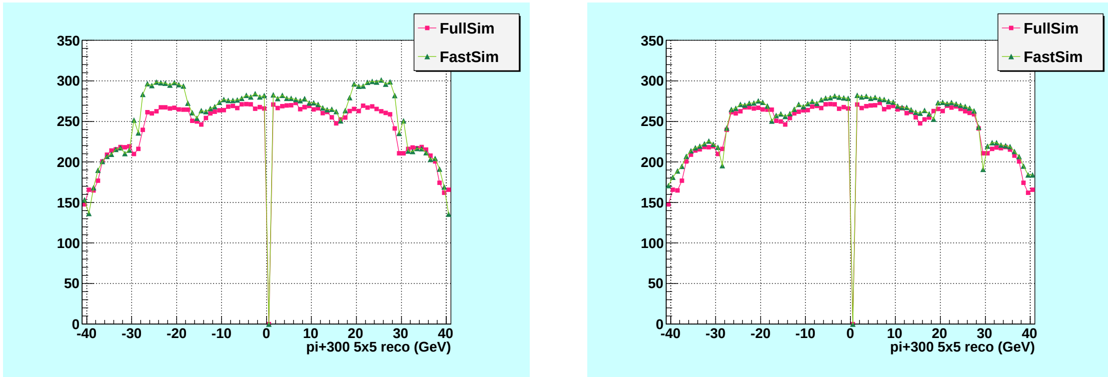

Figure 9. Mean energy response of 300 GeV/c π^- as a function of pseudo-rapidity (η) before (left) and after (right) the tuning.

A dedicated trigger path has been developed in order to record a high rate of single charged particles (mainly pions) which will be used for HCAL calibration. This data will be used to inter-calibrate the cells and well as to provide an estimate of the energy scale. The rich sample of charged pions collected by this trigger will be used to tune the simulation to the data.

Multi-jet QCD production is the dominant source of single charged particles and the rate of such events will be large. The trigger will have to be highly prescaled in order to keep the data volume within the bandwidth limitations. Data required to calibrate HCAL will preferentially have charged hadrons at intermediate and high energies. However, tuning of the shower response will require lower energy hadrons. Samples of minimum or unbiased trigger data are also looked into to cover this energy region.

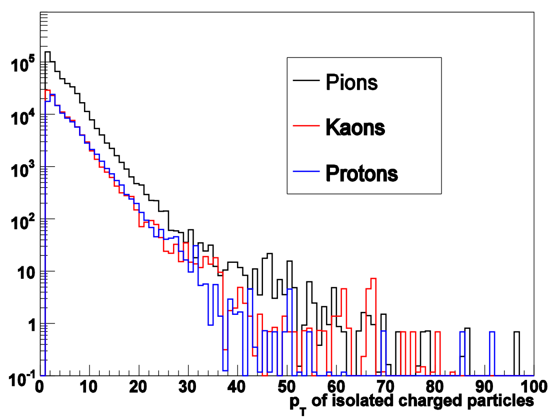

Figure 10. Momentum spectrum of isolated charged particles expected from 10 pb^{-1} of LHC data at a centre-of-mass energy of 10 TeV.

Figure 10 shows momentum spectrum of isolated charged particles expected from 10 pb^{-1} of LHC data at a centre-of-mass energy of 10 TeV. As can be seen from the figure, the momentum spectrum is peaking heavily at low energy. Offline selection rate is 2.2 Hz for particles above 10 GeV/c from single jet trigger of 30 GeV threshold at a luminosity of 10^{31} cm^{-2} s^{-1}. These particles are dominantly charged pions (66.7%) with contamination of charged kaons (16.0%)

and protons (14.2%). No particle identification will be present in these data and tuning of the shower simulation will require typically 100 thousand isolated charged particles in each energy region.

6. Summary

Simulation of showers in the electromagnetic and hadron calorimeter of CMS has been given a special attention by the collaboration. The quality of simulation is improved for both Geant4 based Full Simulation and parameterized version in Full and Fast Simulation. The results of the simulation are tested against the data from several test beam experiments and the current level of agreement between data and Monte Carlo is quite acceptable. There is a plan to continue the tuning with collision data from the LHC.

Acknowledgment

The author would like to thank the members of Calorimeter Simulation Task Force and simulation team of CMS.

References

[1] The CMS Collaboration 1994, *The Compact Muon Solenoid Technical Proposal* CERN/LHCC 94/38, LHCC/P1.
[2] F. Cossutti, *Contribution to the 2009 Conference for Computing in High-Energy and Nuclear Physics (CHEP 09), Prague, Czech Republic*; S. Banerjee, Journal of Physics Conference Series **119** (2008) 032006.
[3] Geant4: S. Agostinelli *et al* , Nuclear Instruments and Methods **A506** (2003) 250; J. Allison *et al* , IEEE Transactions on Nuclear Science **53** (2006) 78.
[4] D. Orbaker, *Contribution to the 2009 Conference for Computing in High-Energy and Nuclear Physics (CHEP 09), Prague, Czech Republic.*
[5] A. Zabi *et al* , CMS DN-2007/005.
[6] G. Baiatian *et al* , CMS Note 2006/143.
[7] S. Abdullin *et al* , EPJ **C60** (2009) 359.
[8] G. Grindhammer, M. Rudowicz, and S. Peters, *"The Fast Simulation of Electromagnetic and Hadronic Showers"*, Nucl. Instrum. and Methods **A290** (1990) 469.
[9] Cmssw: *http://twiki.cern.ch/twiki/bin/view/CMS/WorkBook*
[10] G. Folger and J. P. Wellisch, *In the Proceedings of 2003 Conference for Computing in High-Energy and Nuclear Physics (CHEP 03), La Jolla, California, 24-28 Mar 2003, pp MOMT007* [arXiv:nucl-th/0306007].
[11] A. Heikkinen, N. Stepanov and J. P. Wellisch, *In the Proceedings of 2003 Conference for Computing in High-Energy and Nuclear Physics (CHEP 03), La Jolla, California, 24-28 Mar 2003, pp MOMT008* [arXiv:nucl-th/0306008].
[12] D. H. Wright *et al* , AIP Conf. Proc. **867** (2006) 479; T. Koi *et al* , AIP Conf. Proc. **896** (2007) 21.
[13] J. Birks, Theory and Practice of Scintillation Counting, Pergamon Press, 1964; J. B. Birks, Proc. Phys. Soc. **A64** (1951) 874.
[14] D. H. Wright *et al* , AIP Conf. Proc. **896** (2007) 11.
[15] G. Grindhammer and S. Peters, "The parameterized simulation of electromagnetic showers in homogeneous and sampling calorimeters," *Int. Conf. of Monte-Carlo Simulation in High Energy and Nuclear Physics, Tallahassee, Florida, USA(1993),* arXiv:hep-ex/0001020.
[16] Calorimetry Task Force Report, CMS Note 2008/025.