Response of CMS High Granularity Calorimeter Prototype to Charged Pions

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Niramay Vinayak Gogate

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2020

Supervisor: Dr. Seema Sharma

Certificate

This is to certify that this dissertation entitled *Response of CMS High Granularity Calorimeter Prototype to Charged Pions* towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Niramay Vinayak Gogate at Indian Institute of Science Education and Research under the supervision of Dr. Seema Sharma, Associate Professor, Department of Physics, during the academic year 2019-2020.

Dr. Seema Sharma

Niramay Vinayak Gogate

Committee:

Dr. Seema Sharma

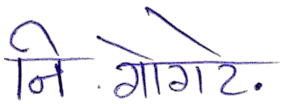
Dr. Rajdeep Chatterjee

This thesis is dedicated to my grandfather, Shankar Gogate

Declaration

I hereby declare that the matter embodied in the report entitled *Response of CMS High Granularity Calorimeter Prototype to Charged Pions*, is the result of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Seema Sharma and the same has not been submitted elsewhere for any other degree.

Dr. Seema Sharma



Niramay Vinayak Gogate

Acknowledgment

I am grateful to my project advisor Dr. Seema Sharma, associate professor, IISER for her guidance throughout the project, for availing me CERN resources which were necessary for this project as well as for introducing me to the HGCAL beam tests and other members of the HGCAL project. Special thanks to Mr. Shubham Pandey, IISER for patiently solving my doubts about HGCAL, providing me all the data analysis tools, root files and for explaining to me the test-beam setup and various terminologies associated with calorimetry. I would like to thank Dr. Rajdeep Chatterjee, the University of Minnesota for giving me advice on the thesis. I would also like to thank group members, Dr. Vinay Hegde, Dr. Aditee Rane, Bhumika Kansal and my colleagues at IISER, Shivani Lomte, Alpana Sirohi and Nukulsinh Parmar for their inputs during the weekly group meetings and the offline discussions.

I would like to thank the Department of Science and Technology, Government of India for providing me financial assistance through the KVPY fellowship during the course of this project. I would also like to thank IISER for providing me cluster access for computational and storage purposes and for providing GPU access for data analysis and machine learning tasks.

Abstract

This thesis presents the calibration and performance study of CMS High Granularity Calorimeter (HGCAL) based on the data collected during the beam tests conducted in 2018. The muon run data was used to calibrate the silicon sensors of HGCAL which record the energy deposits of the incoming high energy particles. The calibration values thus found were then used to study the signal-to-noise ratio of these sensors. The pion run data was later used to check the energy response and resolution of the detector. The simulation of the passage of charged pions through HGCAL test-beam setup was used to understand the leakage and absorbed energy profile. With multiple active layers available, the shower start algorithm was used to predict the approximate location of the first hadronic interaction of the pion with the detector material. In order to improve the pion energy resolution, a simple machine learning regression model was considered which gave better results than the conventional methods, suggesting the possibility for improvement with more complex machine learning models.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Standard Model (SM) of particle physics is the theory describing three of the four known fundamental forces, the electromagnetic, weak, and strong interactions as well as classifying all known elementary particles. The last particle to be found was the Higgs boson which was discovered in 2012 at the European Organization for Nuclear Research (CERN), the world's largest accelerator facility.

Large Hadron Collider (LHC) at CERN, the biggest particle collider in the world has been successful in verifying most of the results predicted by the standard model. LHC is capable of colliding two proton beams with the center of mass energy of 13TeV.

Despite experimental verification of SM, there are still many open questions like neutrino oscillations, CP violation that SM has failed to explain. To continue to probe the nature and look for signatures of newly proposed theories, the LHC will be upgraded to higher luminosity from $\sim 10^{34} cm^{-2} s^{-1}$ to $\sim 10^{35} cm^{-2} s^{-1}$ around the year 2026. Higher luminosity implies more intense beams and hence more chances of producing particles of interest - the processes involving heavy masses or smaller cross-sections. The High Luminosity LHC (HL-LHC) will integrate 10 times more luminosity than the LHC, posing significant challenges for radiation tolerance and event pileup on detectors, especially for forward calorimetry, and hallmarks the issue for future colliders.

1.1 HGCAL Detector: motivation

To cope with the high radiation environment and large pileup rates at the HL-LHC, the CMS collaboration plans to upgrade the endcap side of the detector extending in the range $1.5 < |\eta| < 3$. Currently, the CMS detector contains the following two types of calorimeters to measure the energy of electromagnetic objects (electrons, photons) and the hadronic objects (protons, neutrons, pions, kaons) respectively.

- **Electromagnetic Calorimeter (ECAL):** This homogenous calorimeter is made of $PbWO_4$ crystals. A particle passing through the crystals generates scintillating photons which are then detected by photodetectors to obtain the raw signal.

- **Hadronic Calorimeter (HCAL):** This sampling calorimeter has alternating layers of active and passive material. The active material consists of plastic scintillators and passive material consists of brass and steel.

These calorimeters will be upgraded with a silicon and scintillator based technology to provide high granularity for multi-dimensional shower reconstruction. This new detector is known as the High Granularity Calorimeter (HGCAL). Figure 1.1 depicts the proposed design of HGCAL.

1.2 HGCAL Beam tests

In order to validate the basic design of the HGCAL, a series of beam tests are currently being performed on the proposed detector configuration. Beam tests are carried out to test various components of the detector such as the electromagnetic and hadronic calorimeters, scintillators and the readout electronics. In the current Beam tests electrons, muons and pions of known energy were shot at the prototype of HGCAL. The electrons create electromagnetic showers, pions create hadronic showers and muons do not create any shower (Refer to chapter 2 for details). This way, choosing this set of particles enables the study of the detector's response to different types of particle-matter interactions.

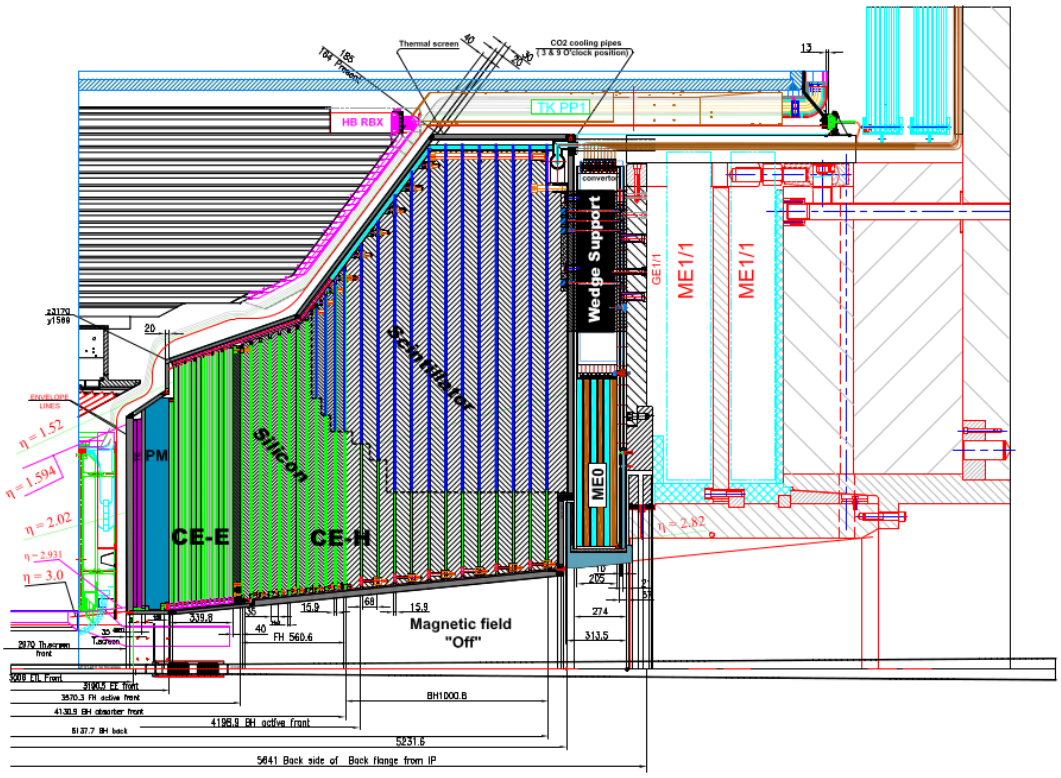


Figure 1.1: Schematic diagram of HGCAL in the CMS end-cap region [1]

Particle	Beam Energy [in GeV]
Muon	200, 300
Electron	20, 50, 80, 100, 120, 200, 250, 300
Pion	20, 50, 80, 100, 120, 200, 250, 300

Table 1.1: HGCAL Test-beam : beam energies

1.3 HGCAL test-beam setup

The analysis presented here made use of the experimental data acquired from the beam tests conducted at CERN in October 2018. The HGCAL test-beam setup consisted of the beam source, trigger system, DWC, and HGCAL detector [Refer to Figure 1.2].

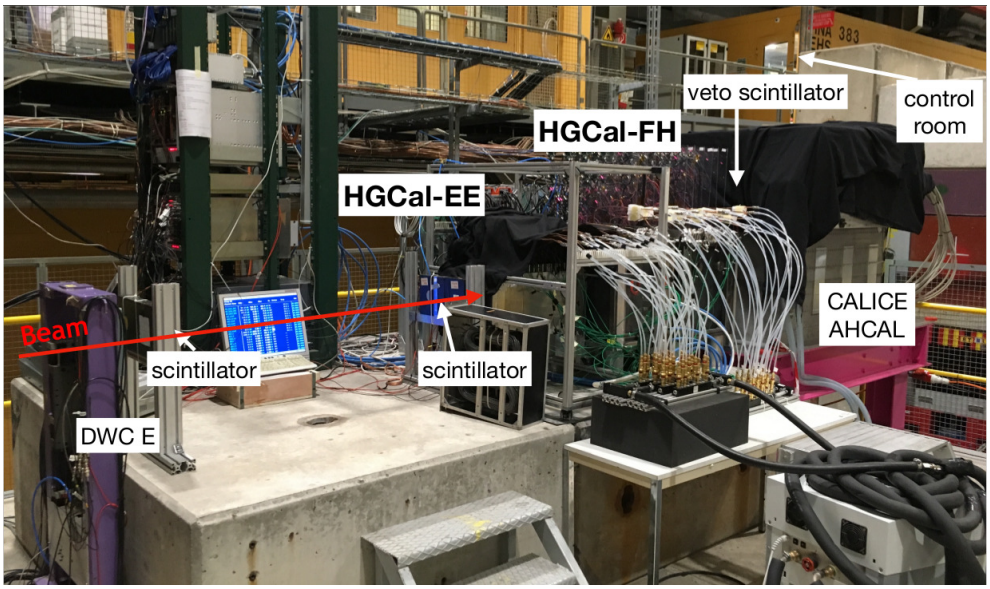


Figure 1.2: Actual test-beam setup photo [2]

1.3.1 Beam source

The Super Proton Synchrotron (SPS) at CERN produces proton beams of energies upto 400 GeV. These proton beams then collide with Beryllium plates of different thicknesses at site T2 to produce secondary beams of desired momenta and purity [3]. The secondary beam mainly contains hadrons like protons, pions and kaons [4]. The secondary beam is carried by the H2 beamline to the Experimental Hall North 1 (EHN1) area at CERN where the prototype HGCAL detector is placed. Usually, there is 1 μs time gap between the consecutive particle bunches. The test beam experiments were carried out with protons, muons, electrons, and pions. These particles were generated from the initial proton beam through the following physical processes.

- Pions: The proton beam collides with the fixed target to produced charged and neutral hadrons as the scattered products.

- Electrons: The neutral hadrons produced in the above process decay immediately (half-life $\sim 10^{-22}s$)to two photons. These two photons give two electron-positron pairs. [4]

- Muons: The charged pions decay to give muon and neutrinos ($\sim 10^{-8}s$)

1.3.2 Trigger system

The readout electronics stores the information received from the HGCAL sensors only when it receives an external trigger. In the case of HGCAL beam tests, two consecutive scintillators were kept before the starting of the HGCAL setup. The trigger was set when both the scintillators recorded a signal at the same time. The scintillators had dimensions of 2cm$\times$2cm and 4cm$\times$4cm while a single hexaboard sensor was 15cm wide. The EE part of HGCAL had only one hexaboard module per active layer passing through the beam axis. So the beam particles entering both the scintillators thus ensured that the readings were taken only for the events with the beam particles travelling close to the beam axis.

1.3.3 Delay Wire Chamber (DWC)

HGCAL will have a silicon tracker at CMS to predict the particle trajectories. In the case of beam tests, four delay wire chambers were kept one after the other to mimic the tracker and for trajectory reconstruction.
A delay wire chamber is made of wire mesh connected to a voltage source [Refer Figure 1.3]. A charged particle passing through it ionizes the gas around the wires. The electrons thus created then travel to the anode, creating an avalanche that can be detected by the system hardware. DWC is used for precision position measurements and comparatively lower cost. The DWCs used for HGCAL testbeam had a position resolution of $\sim$1mm.

1.4 HGCAL test beam - detector configuration

The HGCAL prototype had three main components namely EE, FH and AH which were analogous to electromagnetic, hadronic calorimeter and scintillator parts of the proposed HGCAL. The first two parts of the detector, ie, EE and FH sampling calorimeters, consisting of a number of passive layers which were primarily meant to absorb incoming particle's energy and a number of active layers made of small silicon channels which detect the charged particles passing through them. The relevant length scale for the hadronic shower is the interaction length (λ_{int}); defined for a material as the depth till which $\frac{1}{e}$ ($\sim 36\%$)fraction of incoming pions will survive the detector without starting a hadronic shower. The detector's

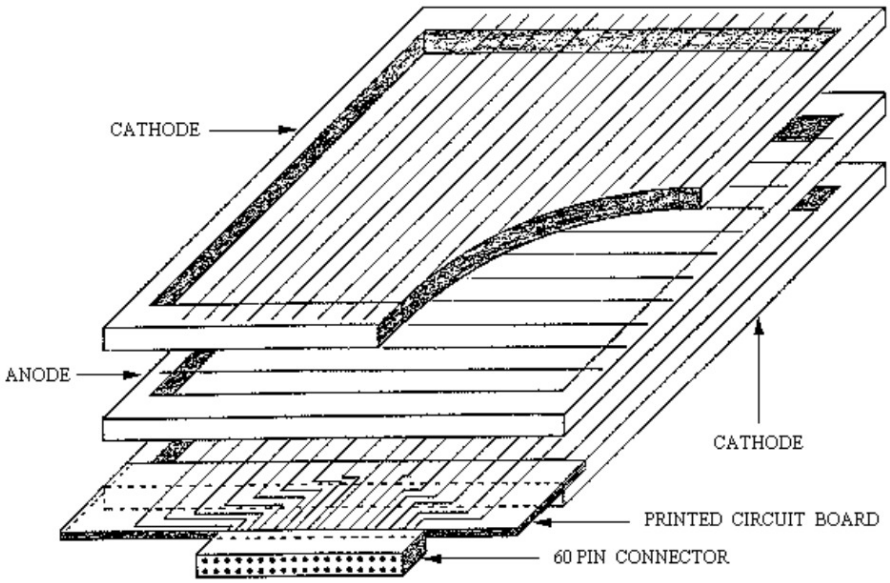


Figure 1.3: Schematic diagram of a typical DWC [5]

thickness can hence be defined in terms of interaction length. Table 1.2 gives information about the different components of the HGCAL detector.

Detector	**Absorber material**	**No. of active layers**	**Depth (in λ_{int})**
EE	Lead	28	1.6
FH	Steel	12	3.4
AH	Steel	39	4.6

Table 1.2: HGCAL test-beam detector configuration

The active layers of EE ad FH were made of hexaboard modules each comprising of 128 small silicon sensors (also called channels) each capable of collecting charge in the event of charged particle passing. All the 28 active layers of EE contained one hexaboard module per layer, whereas the first 9 active layers of FH had 7 hexaboard modules per layer (Figure 1.4). The last 3 active layers of FH had 1 hexaboard module per layer.

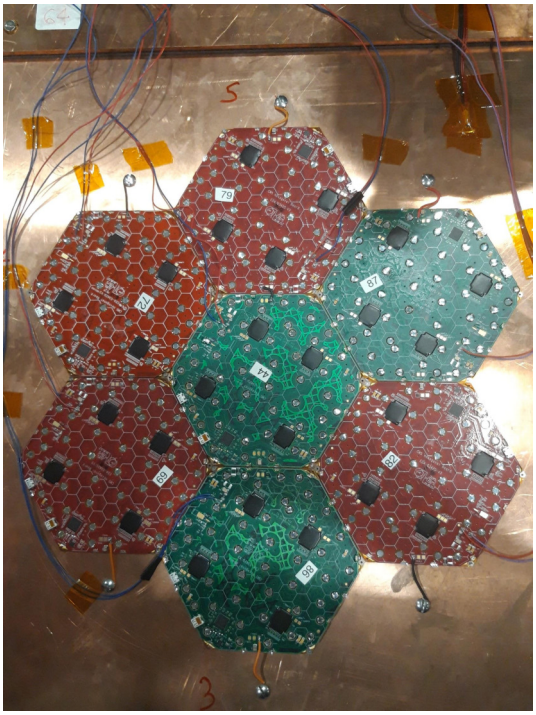


Figure 1.4: Hexaboard modules in daisy structure

Chapter 2

Calorimetry

Calorimetry involves measuring the energy of particles entering the detector. The high energy particles (energy ∼GeV) such as photons, electrons, protons, pions generate a large number of secondary particles when they pass through the detector. This collection of newly generated particles is called a particle shower. Therefore measuring the energy of incoming particles requires the study of particle showers. The particle showers are broadly categorized into the following two categories based on the physical processes involved in making the shower.

2.1 Electromagnetic Showers

Electromagnetic showers are created by electrons and photons. These particles deposit their energy in the detector via following different physical processes.

Photon

- **Pair production :** In the presence of detector material, the photon converts into electron-positron pair. This requires a minimum of 1MeV amount of energy to create

an e+e- pair. The cross-section for this process is proportional to z^2, where z is the atomic number of the material

- **Compton scattering:** The photon scatters electron of the material. The cross-section goes approximately as $\frac{lnE}{E}$.
- **Photo-electric effect:** The photon knocks out an electron from the valence shell of the material atoms. Some part of the energy is utilized to overcome the work function of the material and the rest of the energy is transferred to the electron.

Electron

- **Bremsstrahlung:** The electron interacts with the nuclei in the nucleus and emits a photon in the process. This process is dominant for high energies of the electron. This process is the most dominant process until the electron's energy reaches certain threshold energy, called critical energy (ϵ_c).
- **Ionization:** The incoming electron ionizes electrons of the material, producing secondary electrons. This process dominates for lower energies, ie, below the critical energy.

The secondary particles created through various processes as described above continue to lose energy in the same fashion, thereby creating a large number of shower particles.

2.2 Hadronic Showers

Charged hadrons interact with the detector material and create a hadronic shower. A hadronic shower has a purely hadronic component composed of secondary hadrons like $p^+, n, \pi^\pm$ as well as an electromagnetic component which comes from the neutral pions which decay to two photons. The following are the processes involved in hadronic shower development.

- **Hard interaction :** This is a fast process where the incoming hadron interacts with nucleons of the material to produce secondary hadrons ($\pi^\pm, \pi^0$). The secondary hadrons

are produced in the forward direction. The secondary hadrons production requires a minimum of 280 MeV energy ($\sim 10^{-22}s$).

- **Nuclear Evaporation:** The hard interaction changes the baryonic configuration of the material nuclei, making them unstable. These nuclei undergo radioactive decay, producing new hadrons. This stage is called nuclear evaporation and lasts for a longer period of time ($\sim 10^{-16}s$)

- γ **deexcitation:** Neutrons which were converted to protons during the hadronic interaction capture the electron from the closest orbit to get back the neutron; thereby releasing energy in the form of photons.

- **Fission:** The unstable nuclei can also decay into lighter nuclei which then further interact as they pass through the matter

2.3 Energy fluctuations

The charged particles created in the electromagnetic shower then generate a signal for the calorimeter either by creating scintillating photons, by ionizing the gas molecules or by creating electron-hole pair. The energy measured by the calorimeter depends on the number of charged particles created in the shower as well as the amount of signal created by these charged particles. Hence the intrinsic fluctuations in the signal arise from the following two methods:

1. **Signal quantum fluctuations:** The incoming charged particles ionize the detector atoms producing free electrons (or photons in case of scintillators) which are then recorded as signal of the detector. The fluctuations in the number of these electrons/photons produced in the event are called signal quantum fluctuations. The recording of a signal is essentially a particle counting problem and follows a Poisson distribution. Hence the relative energy resolution coming solely from the quantum

fluctuations can be given by:

$$p(n|\lambda) = \frac{\lambda^n . e^{-\lambda}}{n!}$$
$$\therefore \langle n \rangle = \lambda$$
$$\therefore \Delta n = \sqrt{\lambda}$$
$$\therefore \frac{\Delta n}{\langle n \rangle} = \frac{\sqrt{\lambda}}{\lambda} = \frac{1}{\sqrt{\lambda}}$$

In general the reconstructed energy is proportional to the number of electron/photons recorded. In such cases,

$$E \propto n \implies \langle E \rangle \propto \lambda$$
$$\therefore \frac{\Delta E}{\langle E \rangle} = \frac{\Delta n}{\langle n \rangle} = \frac{1}{\sqrt{\lambda}}$$
$$\therefore \frac{\Delta E}{\langle E \rangle} \propto \frac{1}{\sqrt{\langle E \rangle}}$$

where,
n = number of electrons/photons recorded
λ = mean number of electrons/photons recorded
E = reconstructed energy.

In general, the average number of such electrons/photons produced by high energy charged particles is of the order of 10^4. For example, the number of electron-hole pairs generated by a minimum ionizing particle in a $300\mu m$ thick silicon semiconductor is about 23,000 [6]. Therefore the relative energy fluctuation caused by quantum effects, which goes as $\frac{1}{\sqrt{n}}$ is about 0.6%, which is significantly lower than the other sources of energy fluctuations as discussed in the later sections.

2. **Visible energy fluctuations:** Visible energy corresponds to the energy of the particle which is carried by the ionized electrons/ scintillating photons which are then detected by the detector. In the case of hadronic showers, the incoming hadrons interact with the nucleons in the detector material. Some energy of the hadron is used in overcoming binding energy of the quarks in the nucleon whereas some energy is used in creating new

hadrons. This energy lost by the hadron cannot be detected by the detector in terms of ionized electrons/scintillating photons and hence this energy is called invisible energy. The invisible energy contributes to over 30-40% of the total non-electromagnetic component of the deposited energy.([7] Table 2.6) and the fluctuations in this energy contribute to about $15 - 20\%$ [7]

3. **Shower fluctuations:** The fluctuation in the number of secondary particles created in a particle shower is called shower fluctuations. In the case of charged pion shower, the secondary particles which carry most of the energy are charged and neutral pions. The secondary charged pions interact with the detector material and contribute to the hadronic shower whereas the neutral pion decays to two photons (half-life $\sim 10^{-16}s$) which in turn contribute to the electromagnetic component of the shower. The purely hadronic component of the shower loses some of its energy in terms of the binding energy as described in the previous point which further degrades the energy resolution. Hence fluctuations in the amount of energy that goes to electromagnetic and hadronic components are called shower fluctuations. These contribute significantly to net energy fluctuation. For example, in the case of QFCAL (See appendix C for more information) prototype scintillator, these contributed about 28% compared to the other sources which gave around 33% resolution [8].

2.4 Parametrizing energy resolution

For a given particle and a given detector, the energy resolution as a function of beam energy can be parametrized in the following way :

$$\frac{\Delta E}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c$$

with the coefficients a, b, c being parameters of the model. The three terms are added quadratically assuming that they are not mutually correlated. The three coefficients represent the dominant sources of the fluctuations in the following manner.

- a : This represents random energy fluctuations that come from the intrinsic noise in the system. This includes thermal noise from the different electronic components of the detector and the common mode noise in the system. Since these fluctuations are

independent of the incident energy of the particle, their relative energy fluctuations have inverse beam energy dependance.

- b : This represents stochastic fluctuations. In the case of both quantum fluctuations as well as the shower fluctuations, the detector signal is the addition of energy depositions from different individual quanta. For N particles each giving a signal x_i, the signal fluctuations can be given as:

$$E = k\sum_{i=1}^{N} x_i = kN\langle x\rangle \implies (\Delta E)^2 = k^2\sum_{i=1}^{N}(\Delta x_i)^2 = k^2(\sigma_x^2)N$$

$$\therefore \frac{\Delta E}{E} = \frac{k\sigma_x\sqrt{N}}{kN\langle x\rangle} \propto \frac{1}{\sqrt{N}} \propto \frac{1}{\sqrt{E}}$$

  In the case of quantum fluctuations, x_i is the energy deposited by each ionized electron while in case of shower fluctuations, x_i is the total energy deposited by the shower particle in the material.

- c : This represents the instrumental effects such as lateral, longitudinal leakage and albedo. For high energy hadrons in a sampling calorimeter, these effects produce an energy independant relative energy resolution [9]. These effects dominate the high beam energy region where the stochastic and noise terms become negligible due to their inverse dependance on energy.

Chapter 3

Signal to Noise Ratio Analysis

Electronic measuring instruments such as the silicon sensors used in HGCAL have inherent noise coming from various sources such as thermal fluctuations, common-mode noise etc. Hence it becomes very crucial to define the signal and noise of the measuring system and then distinguish them. In the case of HGCAL sensors, the signal is meant to be the electric signal given by the sensor when a mip-like particle passed through it whereas noise is meant to be the electric signal given by the sensor when no particle passed through it. When measuring the energy deposited by an incoming particle, it is important to reject the hits which get recorded due to mere detector noise. These noisy hits can be rejected by defining an energy threshold for the recorded hits. This threshold needs to be set so as to reject maximum noise events and pass maximum signal events. Knowing the signal to noise ratio (SNR) helps to set the scale for this energy threshold. For example, if the noise and SNR of the sensor are N, SNR respectively, then the energy threshold can be set as high as $N \times SNR$. For this reason, finding SNR is important for HGCAL sensors. On the other hand, the stability of SNR across the sensors is one indication that all the sensors are performing uniformly.

This section briefly describes the mechanism of silicon sensors, definitions of signal and noise and SNR characteristics of HGCAL detector.

3.1 Silicon pixel detectors

Silicon sensors are used in the HGCAL detector as the active material in EE and the FH parts. The signal and noise of a measuring instrument depend on the structure and mechanism of the instrument. Hence it is important to first study the overall structure and principle of a general semiconductor detector.

A general semiconductor detector consists of a p-n junction with the voltage applied across the junction in reverse biased mode (Fig. 3.1). The detector is kept at a fixed reverse-biased voltage and the output of the detector is measured using the readout electronics in terms of the charged collected.

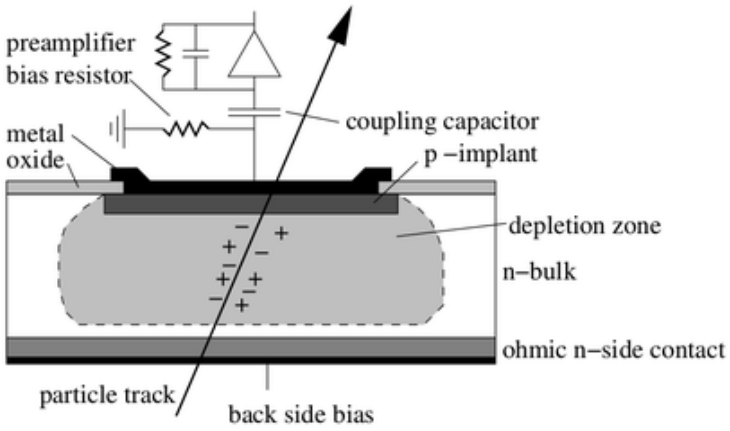


Figure 3.1: Silicon pad sensor- schematic diagram

A p-n junction, in general, has free electron-hole pairs that can move across the sensor. When reverse biased, electrons, holes in the p-n junction get attracted towards the oppositely charged electrodes, creating a depletion region near the junction site which is void of e-h pairs in equilibrium. The formation of the depletion region breaks the circuit and no further current passes through the junction. In the event a charged particle passes through this region, it ionizes atoms in the material, creating extra e-h pairs. These newly created e-h pairs then travel to the oppositely charged electrodes and thus current is set in the system. This way, a current is generated whenever a charged particle passes through the system.

To obtain maximum current in the event of charged particle passing, the depletion region is increased as much as possible by applying more and more potential difference. The voltage can be increased until the depletion region covers the entire pn junction. Increasing the voltage further creates an avalanche of electrons that overcome the potential barrier in the depletion region. This critical voltage is called the breakdown voltage.

In HGCAL silicon sensors, voltage is set below the breakdown voltage to avoid the breakdown while also making sure that the sensor is fully depleted. In the current HGCAL setup, 128 of such sensors collectively make one hexaboard module and one hexaboard module is used as the smallest mechanical unit of the active layer components.

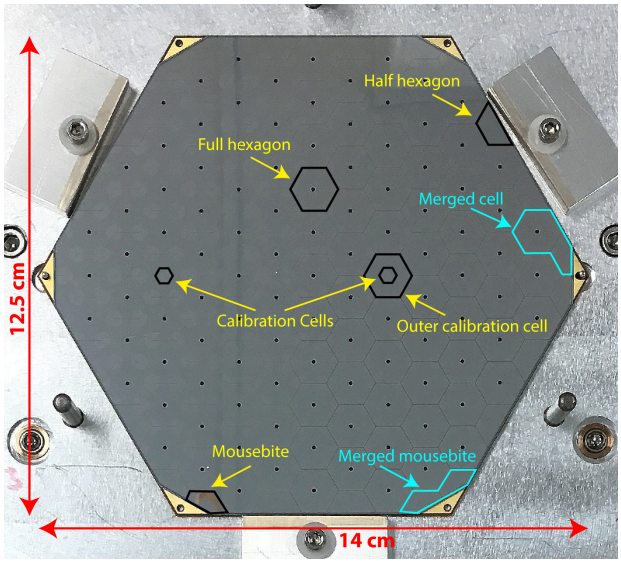


Figure 3.2: Hexaboard module

3.2 Readout electronics

The readout electronics attached to the sensor collects charge obtained from the sensor. This analog signal is converted to a digital signal and is measured in terms of analog-to-digital counts (ADC counts). The HGCAL readout system has the following main components:

1. **Pre-amplifier** : the actual current generated due to passage of charged particle is extremely low (in nano amperes) and normal electronic components cannot detect

such small quantity hence a preamplifier is used to amplify the raw current to a level which is detectable by the rest of the electronics

2. **trigger** : Trigger is an electronic signal given to the system which tells the system to store the reading. In the case of beam tests, there are two scintillators kept before the starting of HGCAL. The trigger is set when both the scintillators record signals simultaneously. Such a triggering mechanism makes sure that a charged particle passed through both of the scintillators which also implies that it would have entered the HGCAL.

3. **pulse shaper**: The reading is recorded with different time stamps. Hence in the event of particle passing, the detector receives the signal for different time stamps. The pulse shaper fits a predefined function to this set of data and reports the ADC counts for given event

4. **Data Acquisition**: The output of pulse shaper is then read by the data acquisition system (DAQ) and is then sent out as the final output of the sensor.

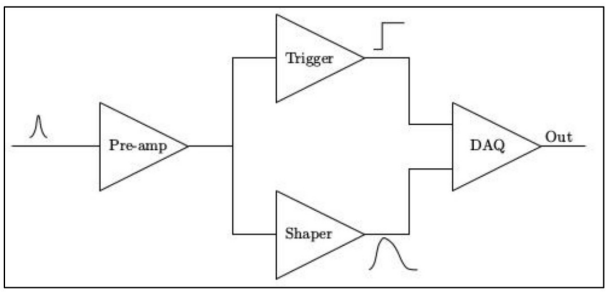


Figure 3.3: Generic Readout mechanism

3.3 Noise

In the case of HGCAL beam tests, the noise is thought of as the electric signal recorded by the sensor which does not correspond to the energy deposited by passage of charged particle. Depending on the nature of the noise, the noise can be divided into two categories.

1. **Electronic noise:** The silicon sensors used in the HGCAL are fully depleted reverse-biased p-n junctions. The electric signal is generated in the sensor due to the creation of

electron-hole pairs. Under the thermal equilibrium, electron-hole pairs are constantly generated and annihilated at the same rate in a p-n junction; but when reverse biased, the spontaneously created electron-holes pair are constantly removed from the junction due to the presence of the electric field. These newly generated charge carriers then travel through the oppositely charged electrodes and give rise to the inherent noise in the system. Every silicon sensor of HGCAL setup independantly develops electronic noise in this fashion.

2. **Common-mode noise:** Unlike the electronic noise which depends on the individual silicon sensors and has no correlation between different sensors, there exists another type of noise called common-mode noise which shows correlation in different sensors. This type of noise originates from sources such as proximity of many electric wires carrying current, contamination from visible light etc.

Apart from these two categories, it was seen that the mean ADC counts received in the absence of the beam source was off from zero. To account for the offset, pedestal was defined to be the mean ADC counts recorded in the absence of any beam source [17] and was subtracted from the raw ADC counts for further analysis in order to make the mean of the ADC distribution zero.

Hence in the event of charged particle passage, the total electric charge collected from a sensor has a contribution from both the electric noise and the energy deposited by the incoming charged particle. Therefore to measure the energy deposition by the charged particle, it is important to estimate the electric noise. In this analysis, the noise levels were estimated in terms of the ADC counts.

3.3.1 Noise extraction

The HGCAL sensors recorded electronic signals in 13-time samples after receiving the trigger. In order to record noise, bypassing the pre-selection criteria, the time sample with the highest overall ADC counts was chosen for all the trigger events and the raw ADC counts received in that particular time sample were considered for the analysis. In the analysis, the 12th time sample was considered for this purpose. The ADC distribution thus obtained is then

fitted with a gaussian and the standard deviation of the gaussian is reported as the noise of the given channel. Figure 3.4 shows the noise fitting for one of the channels. Pedestals were first subtracted from the ADC counts used for the noise analysis. For this reason, we can see negative entries in the following graph.

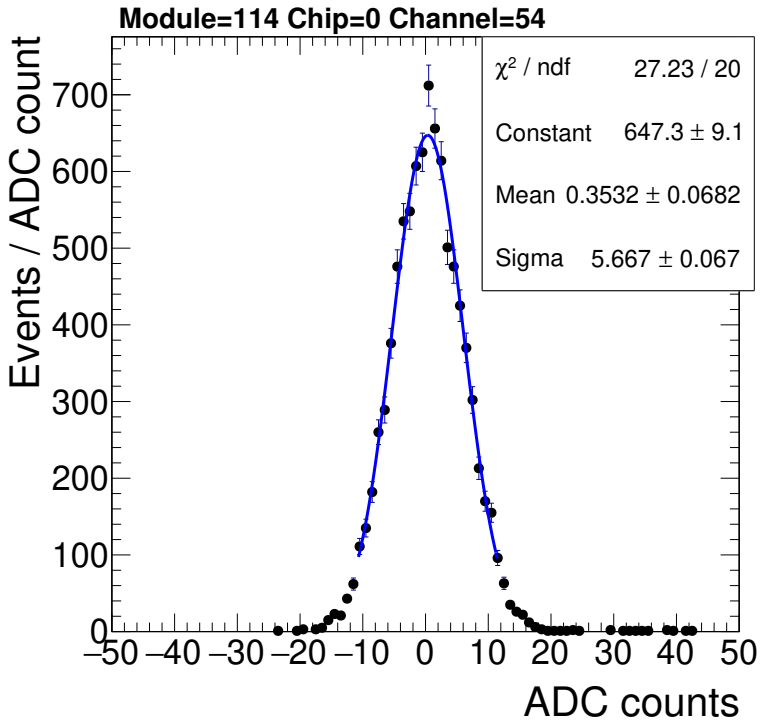


Figure 3.4: Example of noise fitting (FH layer 7)

3.3.2 Noise Analysis

The electronic noise arises from the spontaneous creation of electron-hole pairs. A reverse-biased p-n junction can be thought of as a charged capacitor with the distance between the plates to be the depletion width. For a capacitor, the electronic noise increases with the separation of the plates. In the case of FH modules, the voltage across the channels was set such that their depletion width was around 300 microns. Hence even for the channels with a different surface area, the noise levels were the same. Figure 3.5 Left shows the noise profile of one of the FH modules.

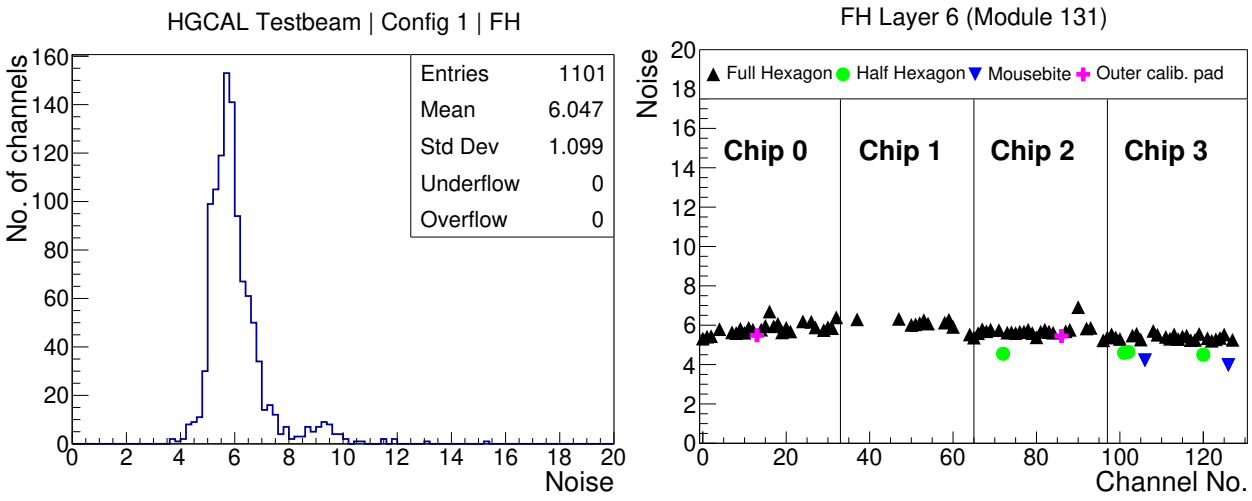


Figure 3.5: FH Noise distribution. **Left:** Noise distribution for all the available FH channels; **Right:** Variation of noise within a module

3.4 Signal

3.4.1 Signal extraction

To calculate the signal for silicon sensors, the experimental data from 200 GeV muon runs taken in October 2018 at the CERN test beam facility was used. For each muon passing event, the silicon sensors recorded electric signals in 13 time steps. These 13 data points were then fitted with a known pulse shape to get the final output of the sensor in terms of Analog to Digital Converter (ADC) counts. The ADC counts were further filtered by subtracting pedestal and correcting for common mode noise. For every muon-passing event, for every channel, the pulse fitting was followed by pre-selection criteria which made sure that the pulse fitting actually captured a signal the charged particle energy deposit. If $S(t)$ represents the signal received at time t, the pre-selection criteria were defined in the following way.

$$\begin{aligned} S(t_{max}) &> S(t_{max}+3) \\ \&\ S(t_{max}+1) &> S(t_{max}+3) \\ \&\ S(t_{max}) &> 20 \end{aligned}$$

where,
t_{max} = time at which the maximum signal was recorded.

The energy deposited by a charged particle through ionization in a thin layer of material follows a Landau distribution [10]. Hence the signal for the sensor was defined to be the most probable value (MPV) of the Landau fitted of over the ADC distribution. Considering the intrinsic electric noise in the detector system and a possibility of two muons passing simultaneously through the sensors, the following fit model was used to acquire MPV of the Landau. Figure 3.6 shows signal extraction for a typical silicon sensor.

$$\begin{aligned} f(x\ |\{mpv, \sigma_{noise}, \sigma_{mpv}, bkg\}) = & \ \mathbf{nnoise}.G(x\ |0, \sigma_{noise}) \\ & + \mathbf{n1mip}.L(x\ |mpv, \sigma_{mpv}) * G(x\ |0, \sigma_{noise}) \\ & + \mathbf{n2mip}.L(x\ |2mpv, \sigma_{mpv}) * G(x\ |0, \sigma_{noise}) \\ & + \mathbf{nbkg} \end{aligned}$$

where,
nnoise, n1mip, n2mip, nbkg are the weights given to each of the cases.
* denotes convolution.
L,G are Landau and Gaussian functions give by:

$$L(x\ |\mu, \sigma) = \int_0^\infty e^{-t}.cos\left[t(\tfrac{x-\mu}{\sigma}) + \tfrac{2t}{\pi}.ln(\tfrac{t}{c})\right] dt$$

$$G(x\ |\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}.exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

Since the x axis had a finite range, the individual probability distributions couldn't integrate out to unity in the given range. This caused fitting problems. The constant term **nbkg** was added to overcome this problem.

The fit model defined above gave good results with a few exceptions (Figure 3.7 Left). The signal fitting for these channels improved when the x-axis range chosen for fitting was changed from (0,200) to (20,200). The pre-selection criteria defined in section 3.4.1 ignored

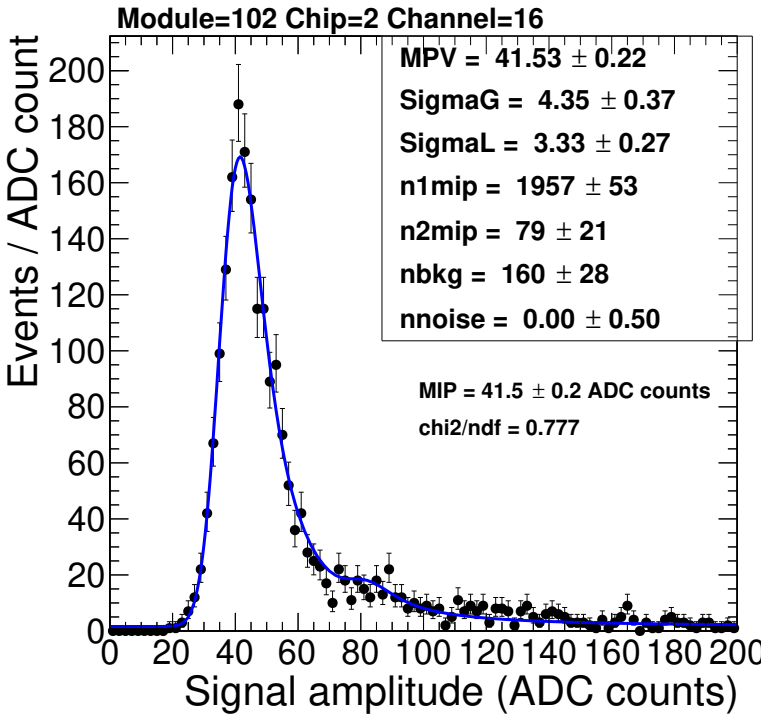


Figure 3.6: Example of signal fitting (FH Layer 4)

the events where the maximum ADC counts received in a time sample were less than 20. This motivated the change of fitting range.

3.4.2 Signal analysis

The MIP signal was calculated for all the central modules of HGCAL. But the analysis presented here focusses only on the central modules from 12 layers of FH. The stability of the signal was checked by looking into the overall distribution of the signal (Figure 3.8 Left).

There were two prominent peaks seen in the histograms making the signal distribution spread wider. Further investigation of the smaller peak revealed that the maximum contribution to the smaller peak came from the channels of layer 8,11 modules (Refer Fig.3.8 Right). The blue points represent the average signal recorded in the layer and the vertical error bars represent the standard deviation of signal in the layer.

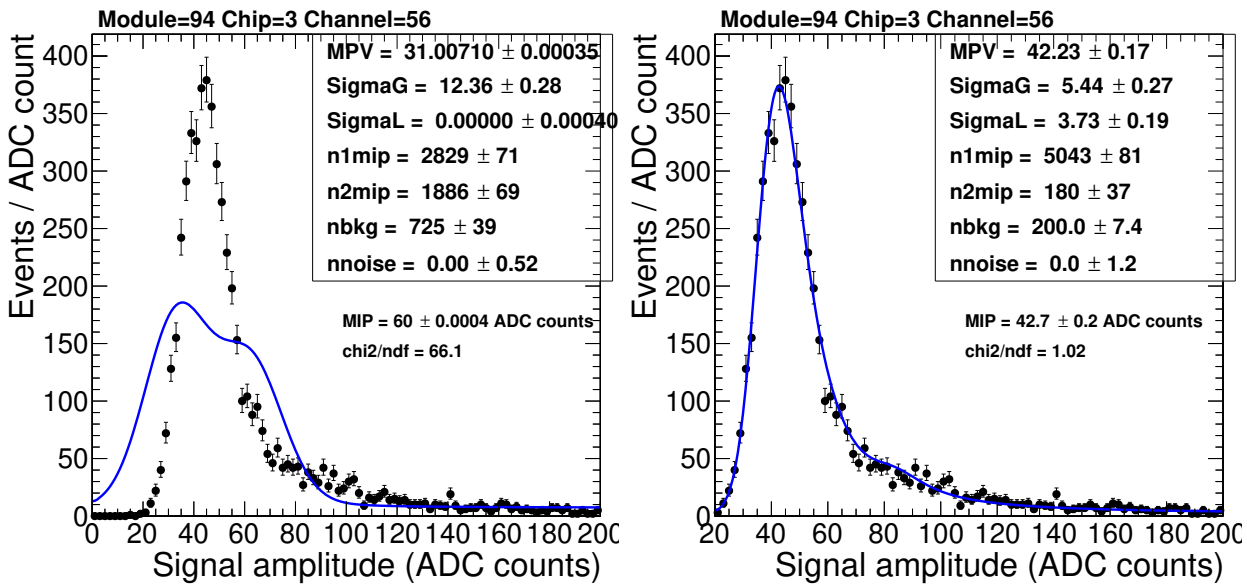


Figure 3.7: Example Bad fits (FH Layer 3). **Left:** Default settings ; **Right:** Modified settings. (More plots in the appendix A.1)

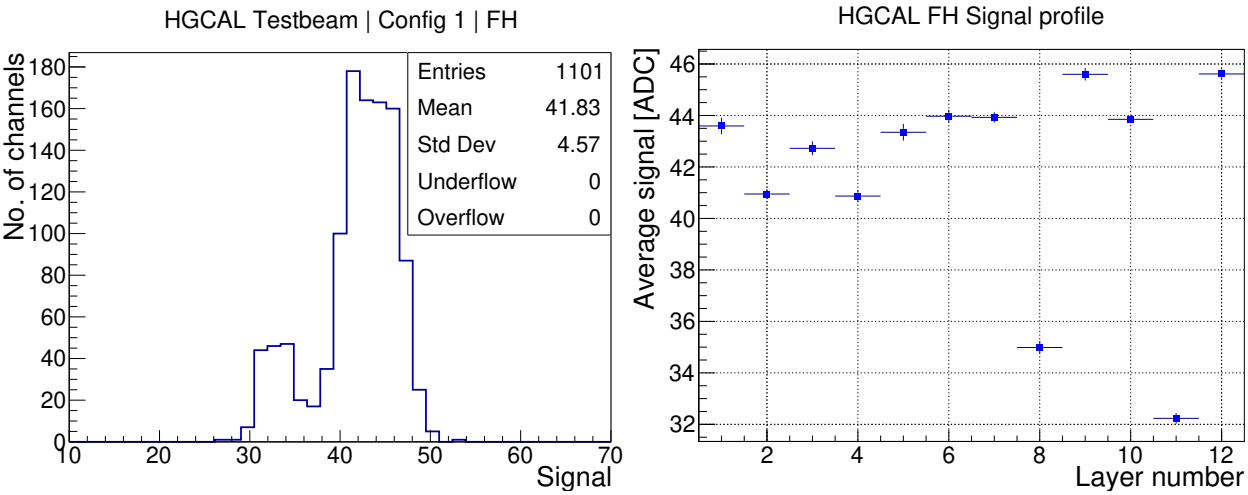


Figure 3.8: FH Signal distribution. **Left:** Overall Signal distribution; **Right:** Layer-wise signal variation (For 1-D signal distributions refer to A.3)

Within a single module, the electric charges from the silicon sensors are collected using 4 Skiroc2 CMS chips [11]. Each chip receives a signal from 32 silicon sensors. Any fault in the chip can affect the readings of all the 32 corresponding channels. So it becomes crucial to look at the signal variation across the chips for the given module. Figure 3.9 shows an example of a signal profile for one module. All the modules considered show overall stable

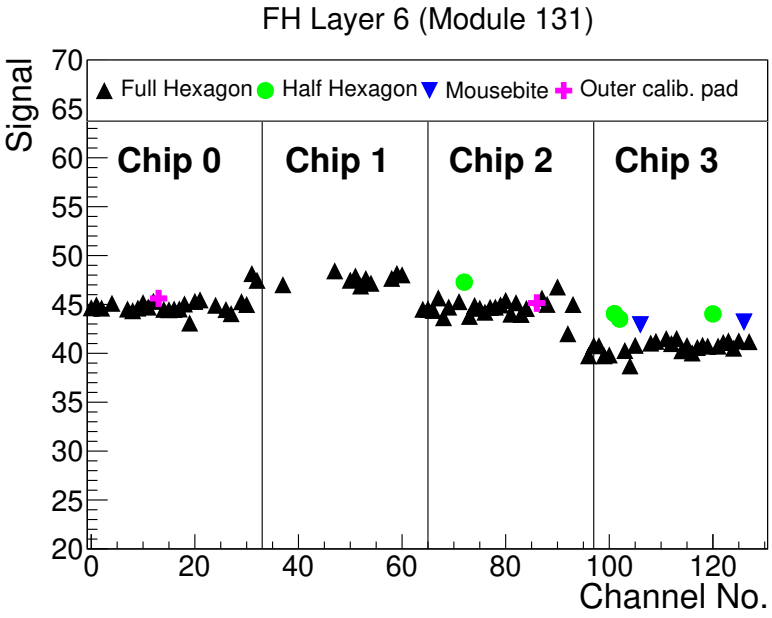


Figure 3.9: Variation of signal within a module

signals within a chip.

3.5 Signal to Noise Ratio

As discussed in section 3.4, the signal was defined to be the most probable value of the ADC distribution obtained from MIP-like events. On the other hand, as discussed in the section 3.3, the noise was defined to be the standard deviation of the gaussian fitted over the ADC distribution obtained for the events where no MIP-signal was expected. In this case, the signal to noise ratio (SNR) was defined to be

$$SNR = \frac{Signal}{Noise}$$

SNR was calculated for all the available channels separately. The SNR variation was checked per hexaboard module while differentiating between channels with different shapes. Figure 3.10(Right) shows the SNR profile for one of the hexaboard modules. Full hexagonal modules, which had the biggest surface area; showed minimum SNR (black points) while for example, the half hexagons, which had half the area of full hexagon, showed higher SNR

(green points).

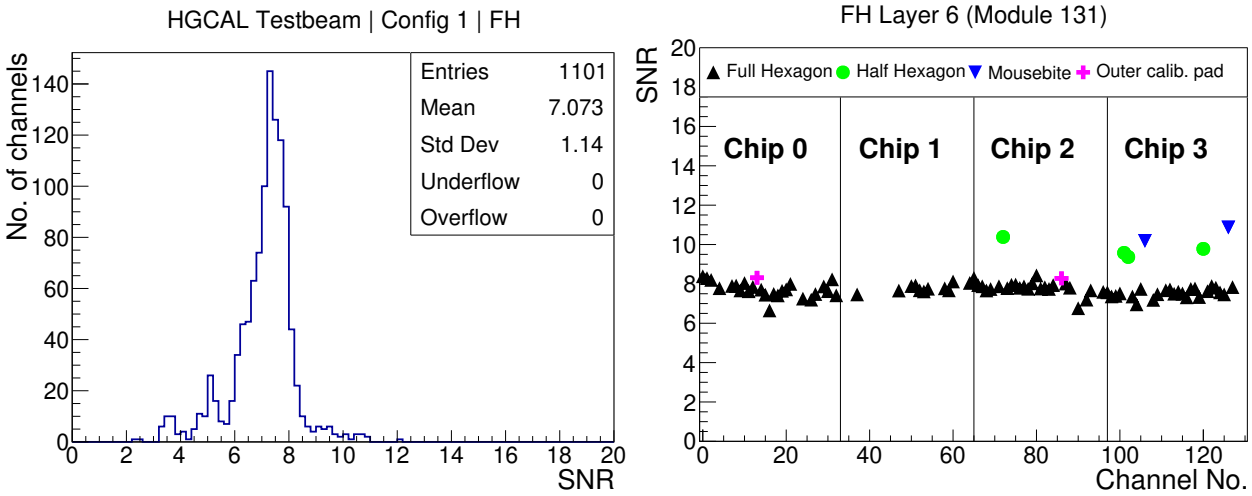


Figure 3.10: FH SNR distribution. **Left:** SNR distribution for all the available FH channels; **Right:** Variation of SNR within a module

Chapter 4

Pion Shower Analysis

4.1 Energy reconstruction

The energy recorded by a single sensor is called as Reconstructed Hit Energy (rechit energy). The rechit energy is measured in terms of MIPs. One MIP is the number of ADC counts that the sensor records if a Minimum Ionizing Particle (MIP) passes through it. In an experiment with a sampling calorimeter, the initial energy of the charged particle is reconstructed using rechit energies of all the available sensors. This is generally done by applying conversion factors to total rechit energy recorded in every active layer. The total rechit energy recorded by the detector depends on the following factors:

1. **Absorber material**: Pion's interaction with absorber depends on absorber's physical properties such as the atomic mass number and mass density. In a sampling calorimeter, most of the particle's energy is deposited in the absorbers and only a fraction ($\sim$ 5%) of energy is deposited in the active layers.

2. **Number of active layers**: More number of active layers means more frequent sampling of the shower which means more total rechit energy.

3. **Size of the detector:** A charged pion creates a hadronic shower while passing through the detector material. So maximum rechit energy is recorded only when the detector is able to fully contain the hadronic shower. Failing to contain the hadronic shower results

in secondary particles escaping the detector without depositing all of their energy. The energy carried by these escaped particles is called leakage.

In the case of test-beam setup, EE, FH, and AH have different absorber materials, different absorber thicknesses, different numbers of active layers and different geometrical dimensions. Therefore while reconstructing pion's initial energy, instead of applying a constant MIP-to-GeV factor to total rechit energy in the whole HGCAL detector, the rechit energies in EE, FH, AH are added by first applying constant weights to them and then applying a common MIP-to-GeV conversion factor. The weights are supposed to take care of the above-mentioned differences in the 3 detector components.

4.2 Weighted rechit energy sum

As described in the previous section, the pion energy was reconstructed by applying relative weights to the total rechit energies in EE, FH, and AH respectively. The final form of the reconstructed energy was as shown below:

$$E_{Reco} = \gamma_1[E_{EE} + \beta(E_{FH} + \alpha E_{AH})] \qquad \ldots \textit{for shower start in EE}$$
$$E_{Reco} = \gamma_2(E_{FH} + \alpha E_{AH}) \qquad \ldots \textit{for shower start in FH}$$

Where $\gamma_1, \gamma_2, \alpha, \beta$ were free parameters which were found using test-beam data. (Appendix B)
and E_{EE}, E_{FH}, E_{AH} were total rechit energies in EE, FH, AH respectively.

The relative weights applied in this fashion could have taken care of the material differences and the different sampling frequencies of the detector but it could not have accounted for variable shower start locations. The pion can start hadronic shower at any point inside the detector and is modeled using the interaction length as described in Section 4.4. If the pion starts shower in FH, it acts as a MIP in EE, thereby depositing a very less amount of energy in EE. As a result, in such cases, total rechit energy in EE could be ignored in energy reconstruction. As can be seen from figure 4.1 (blue curve), the mean longitudinal leakage increases for the late showering pions. (See sec. 4.5 for details of compartmentalization)

For all these reasons, the relative weights were calculated separately for events with

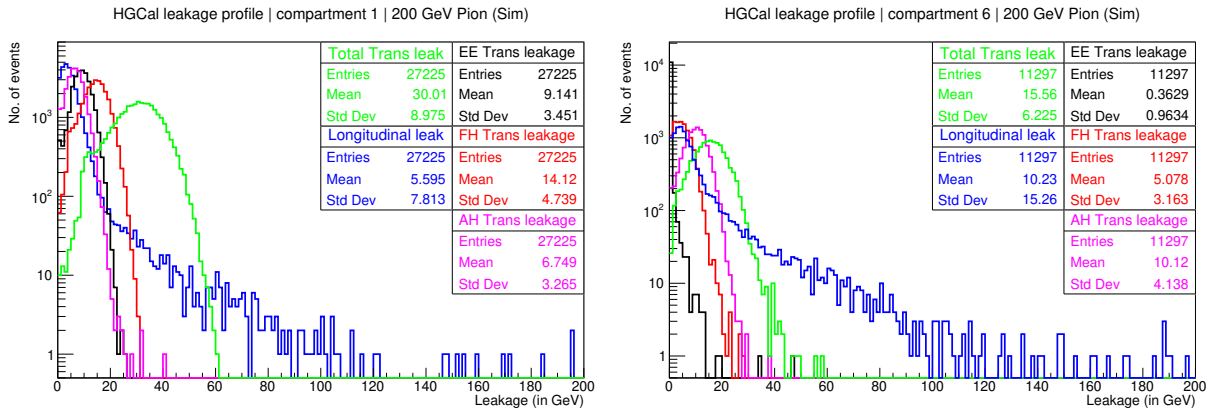


Figure 4.1: 200 GeV Pion leakage profile w.r.t. different shower start locations

hadronic shower starting in EE, FH respectively.

The procedure of finding these weights, as described in Appendix B was followed for different beam energies and therefore finally, the relative weights calculated were a function of both the pion beam energy (20 GeV to 300 GeV) and pion shower start location (EE or FH)

4.3 HGCAL test-beam simulation

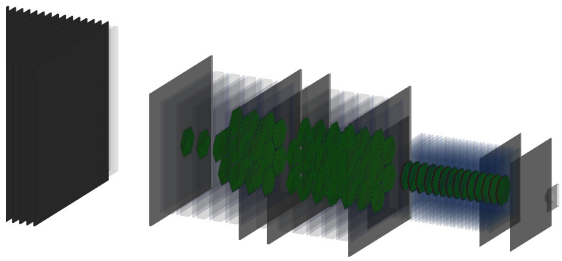


Figure 4.2: HGCAL testbeam setup (Simulation)

The beam tests on the pion beam can only provide information about energy deposits in the active layers. In order to understand the pion energy deposition in detail, it is important to know the amount of energy deposited in the passive layers and the energy leaked. An

idealized trend of these values can be obtained by simulating the HGCAL testbeam setup and passage of pions through them. HGCAL beam tests were simulated using the Geant4 simulation toolkit [12]. Figure 4.2 shows the simulation of HGCAL testbeam setup. In addition to the rechit energy, the simulation provided the following information:

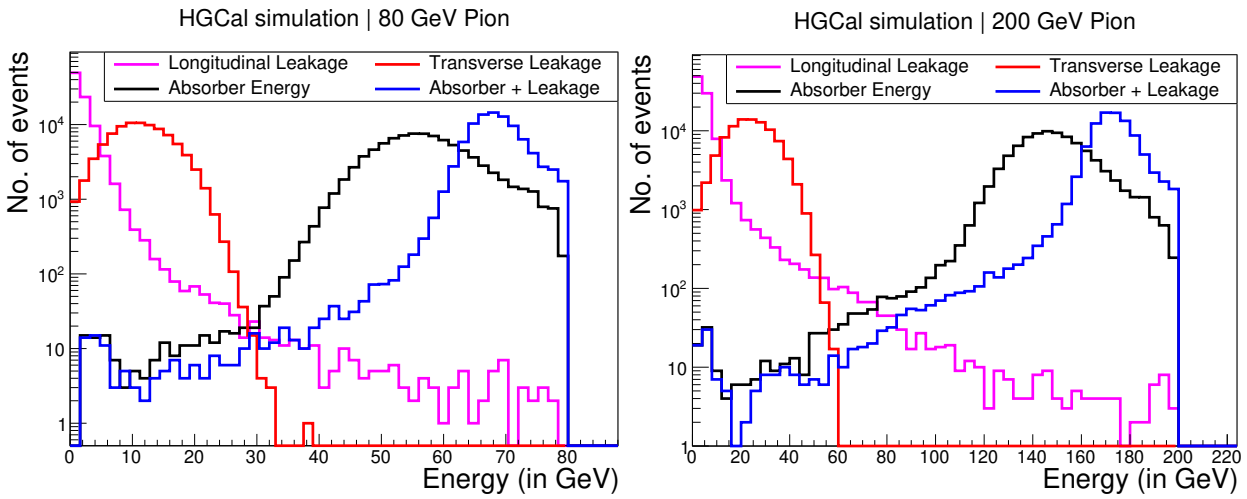


Figure 4.3: Pion absorber and leakage profile (Rest of the plots in the appendix A.4)

- **Energy in absorber**: total energy deposited in EE, FH and AH. (3 variables for the 3 detector components respectively)

- **Transverse leakage:** Total energy leakage in the transverse direction. This was calculated by inspecting the last recorded coordinates of the particles which escaped the detector. The last recorded z-coordinate would determine whether the particle escaped from EE or FH or AH.

- **Longitudinal leakage:** Leakage in the longitudinal direction. This was calculated by summing energies of the particles that escaped the detector and had the last z-coordinate more than that of the last active layer of AHCal.

Figure 4.3 shows these variables for 80 and 200 GeV pion samples. Note that summing up absorber energy and leakage doesn't exactly give back the initial energy of pion. This is because hadronic interactions always contain undetectable energies (eg. energy lost in breaking the nuclei).

4.4 Shower start location

The incoming pion may start hadronic shower at variable depths. For a given collection of high energy pions, the number of pions reaching a depth of z inside the detector without creating a hadronic shower varies as:

$$\frac{dN}{dz} = -\frac{1}{\lambda}N$$

where,
z=depth into the detector material
N=No. of hadrons which didn't start hadronic shower till depth z
λ= interaction length of the material.
If the depth into the detector is measured in terms of the interaction, length, (say $t = \frac{z}{\lambda}$) then the equation further simplifies to:

$$\frac{dN}{dt} = -N$$
$$\therefore N(t) = N_0 e^{-t}$$

where $N_0 = N(0)$

4.4.1 Shower start algorithm

The HGCAL detector in the Test-Beam geometry has a finite volume and hence its energy reconstruction performance highly depends on where exactly does the pion starts showering. Having multiple active layers in the detector, it is possible to estimate the location of the shower start based on the energy deposits in the active layers (Appendix D).

A shower start algorithm has been developed which returns the location of the active layer nearest to the shower start location. The algorithm checks for the total energy deposits in all the active layers in a small area around the beam axis and then returns the layer for which energy deposition increases twofold compared to its previous layer. The main logic behind this is that the shower is started when the hard interaction of the incoming pion produces two or more secondary hadrons which further give rise to new particles.

4.4.2 Shower start distribution

One way to check the validity of the shower start algorithm will be to look at the distribution of shower start locations. As described in the previous section, an exponentially decreasing behavior is expected for this distribution. The shower start algorithm only predicts the active layer nearest to the shower start location. This means the actual location of the shower start could be anywhere between two consecutive layers. Hence the original differential equation can be modified as:

$$\frac{\Delta N}{\Delta t} \sim N$$

where,
ΔN = number of showers starting at depth t
Δt = distance between consecutive active layers
Now substituting the closed solution form for N(t),

$$\Delta N = N_0 \Delta t e^{-t}$$

Hence a histogram of shower start locations will be nothing but a graph of ΔN vs. t. If the distance between consecutive active layers is (Δt) is constant for all layers, this histogram is expected to show an exponential behavior. In the case of HGCAL, the active layers had variable distances between them, so to retrieve the exponential behavior, the histogram was further normalized by dividing each bin by the corresponding

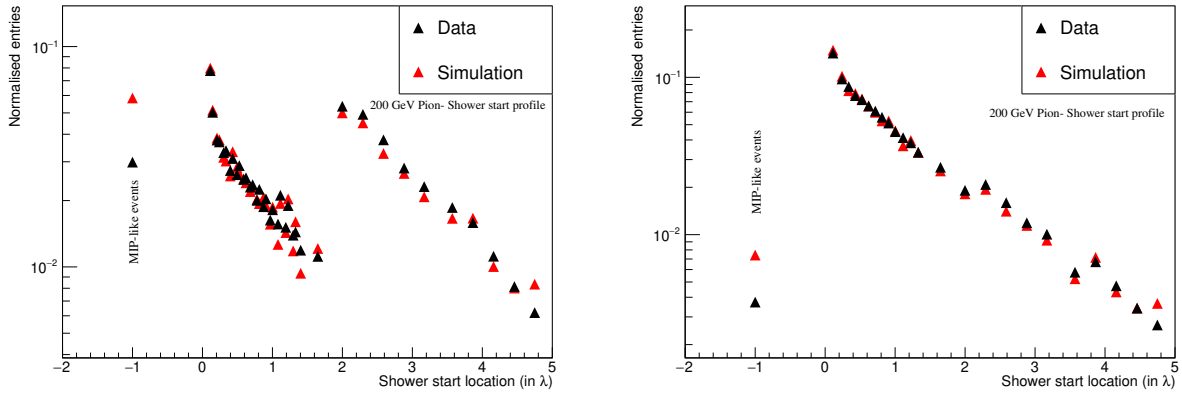


Figure 4.4: no. of shower starts per layer **Left**: ΔN vs shower start location, **Right**: $\frac{\Delta N}{\Delta t}$ vs shower start

Compartment	Depth (in λ)	Detector	Layers	Thickness (in λ)
1	0 to 0.35	EE	1 to 6	0.35
2	0.35 to 0.78	EE	7 to 14	0.43
3	0.78 to 1.12	EE	15 to 22	0.34
4	1.12 to 1.8	EE	23 to 28	0.68
5	1.8 to 2.7	FH	1 to 3	0.9
6	2.7 and beyond	FH	4 to 12	2.4

Table 4.1: HGCAL compartments

4.5 Compartmentalization of the detector

The active layers in the HGCAL Test beam setup were placed at varying depths in terms of interaction lengths. The thicknesses of the absorbers were also different in EE and FH. For these reasons, instead of analyzing each active layer separately, the active layers were divided into compartments (Table 4.1). The compartments were made with the goal that each compartment will have a comparable thickness in terms of interaction length while also having a comparable number of shower start events in each compartment. Keeping in mind the exponentially falling behavior of shower start events, the later compartments were made longer to gather enough statistics.

For the rest of the analysis, pion showering events were classified based on in which compartment do the pions start showering.

4.6 Energy Resolution

The primary function of HGCAL is to estimate the energy of the incident particle. The performance of HGCAL in energy reconstruction is measured based on its energy resolution. The energy resolution is a measure of the sensitivity of the detector in measuring the energy of the incoming particle. It can be calculated experimentally by shooting particles of known energy and measuring the energy recorded by the detector. Energy resolution can be calculated in two ways:

1. Take reconstructed energy distribution, fit a Gaussian to it and calculate $\frac{\sigma}{\mu}$

2. Take reconstructed energy distribution and calculate $\frac{RMS}{Mean}$

where, σ, μ are the mean and standard deviation of the Gaussian respectively. The two ways described will give the same numbers if the reconstructed energy distribution closely resembles a Gaussian. In other cases, these methods give different results and are reported separately.

In this study, energy resolution was calculated for three variables:
1. Total absorbed energy (EE, FH, AH combined)
2. Total absorbed energy + total transverse leakage (EE, FH, AH combined)
3. The weighted sum of rechit energy

The weighted sum of rechit energy was calculated for both data and simulation and became the basis to compare data-simulation. The other two variables were considered to check for the effect of fluctuations of absorber and leakage on reconstructed energy resolution. The energy resolution was calculated for showers starting in different compartments. So for one beam energy, 6 energy distributions were obtained corresponding to shower starting in six compartments. Figure 4.5 shows one such distribution where all the pions starting shower in compartment 3 from 200Gev sample were considered. Each of these energy distributions could have peaked at different energies but for visual comparison, the energies were scaled by an appropriate factor so that all of their means coincide with the beam energy (Fig. 4.5 - right).

The raw energy distributions were then fitted with gaussian around their cores to extract μ, σ. Figure 4.6 shows compartment-wise energy resolution for 200 GeV pions.

It can be seen that the energy resolution changes according to the shower start location. The EE part of HGCAL had only one silicon module per layer which covered approximately 7cm of distance in the radial direction (Figure 3.2). As the shower starts deeper in EE, the shower maximum shifts deeper into FH where more energy can be recorded due to more no. of modules per layer. This reduces statistical fluctuations on the rechits energy, so energy resolution decreases. When shower starts in the last compartment of EE, statistical fluctuation in EE rechits energy is considerable, so energy resolution increases. As the shower starts deeper in FH, the shower cannot be fully contained in the detector ($L_{95} \sim$

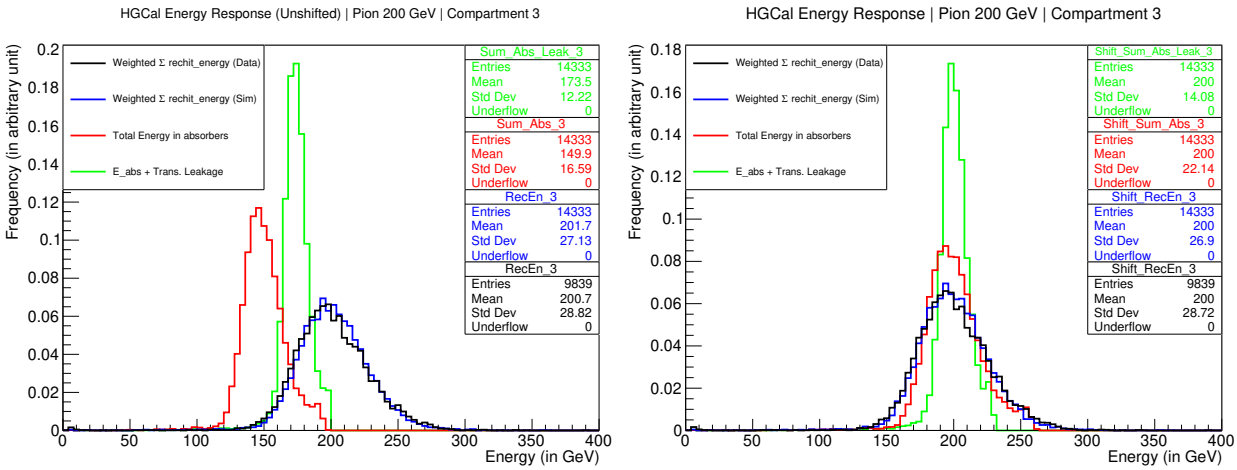


Figure 4.5: Energy distributions. **Left:** raw energy distribution, **Right:** energy distributions with means scaled appropriately to match beam energy (Rest of the plots in the appendix A.5 onwards)

$11\lambda_{int}$). So there is large longitudinal leakage. So resolution increases. The dependence of transverse leakage on the energy resolution can be seen clearly from the green points. After adding transverse leakage to the weighted rechit energy sum, the resultant energy resolution improves and remains constant for all the compartments except the sixth compartment. In the case of the shower starting in FH, the energy resolution degrades with a deeper shower start location mainly because of the longitudinal leakage. This is also reflected in the increasing energy resolution even after adding transverse leakage to the weighted rechit energy sum.

To understand the general trend of energy resolution for given beam energy, energy distributions inclusive of all the showering pions irrespective of their shower start location was also considered (Fig. 4.6). The process was repeated for all the available energy samples. Error on the energy resolution was propagated through the error on the fitting parameters σ, μ in the following manner:

$$\Delta(\frac{\Delta E}{E}) = \frac{\sigma}{\mu}.\sqrt{(\frac{\Delta\sigma}{\sigma})^2 + (\frac{\Delta\mu}{\mu})^2}$$

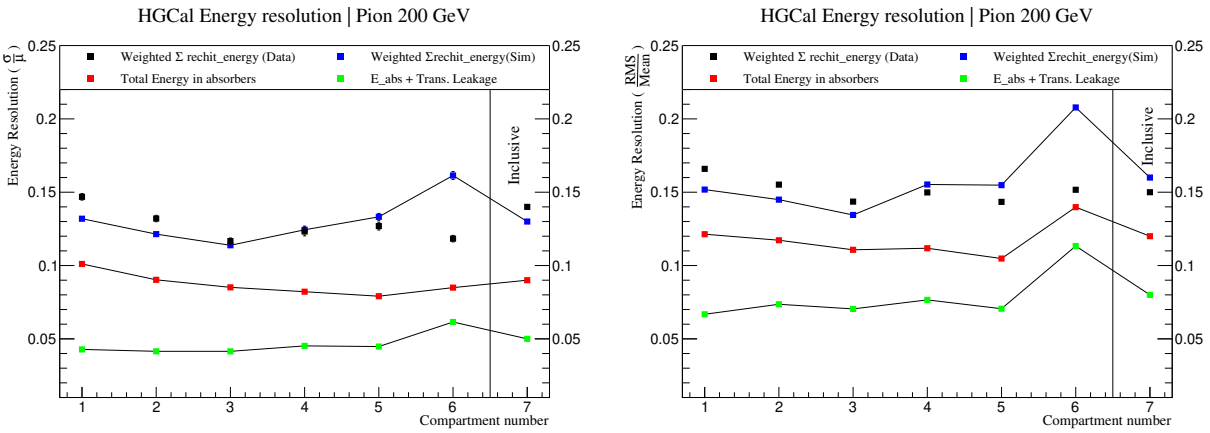


Figure 4.6: Energy resolution for different compartments

4.7 Parametrizing energy resolution

The energy resolution of calorimeters is usually parameterised as : $\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c$ where E is the energy of the incoming particle and the parameters a, b, c represent:

- **a :** Statistical fluctuations in measuring E.
- **b :** random noise in the system. (eg. thermal noise)
- **c :** constraints imposed by detector geometry. (eg. insufficient detector thickness, which can cause leakage)

The three terms are added quadratically assuming they are pairwise uncorrelated and the sum is represented by the symbol $\oplus$. Figure 4.7 shows the trend of inclusive energy resolution for different beam energies as well as the contribution of different fluctuations to the net energy resolution.

At lower beam energies (eg. 20, 50 GeV), the shower is mostly contained in the detector and hence the contribution coming from the leakage, which is included in the constant term (c) is lower compared to the statistical fluctuations (Figure 4.8). For higher energies,(eg. 250,300 GeV) the statistical fluctuations (red points) reduce due to increase in the average number of showering particles/ionized electrons produced and hence the statistical fluctuations reduce but since the detector is not large enough to contain the full shower, the amount

of longitudinal leakage increases and hence contributes the most to the energy fluctuations. It can also be noted that the noise term (black points) is a few orders of magnitude lower than the rest of the terms. This can be attributed to the pre-selection criteria and the pulse-shaping.

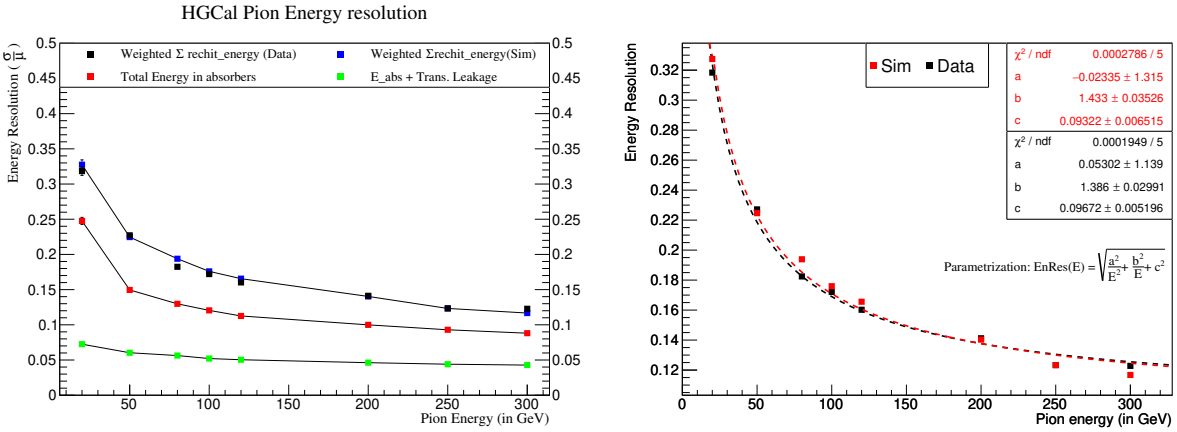


Figure 4.7: HGCAL pion energy resolution. **Left:** Beam energy wise comparison of resolution for different methods of energy reconstruction, **Right:** Parametrization of energy resolution for energy reconstructed using weighted rechit energy sum for data and simulation

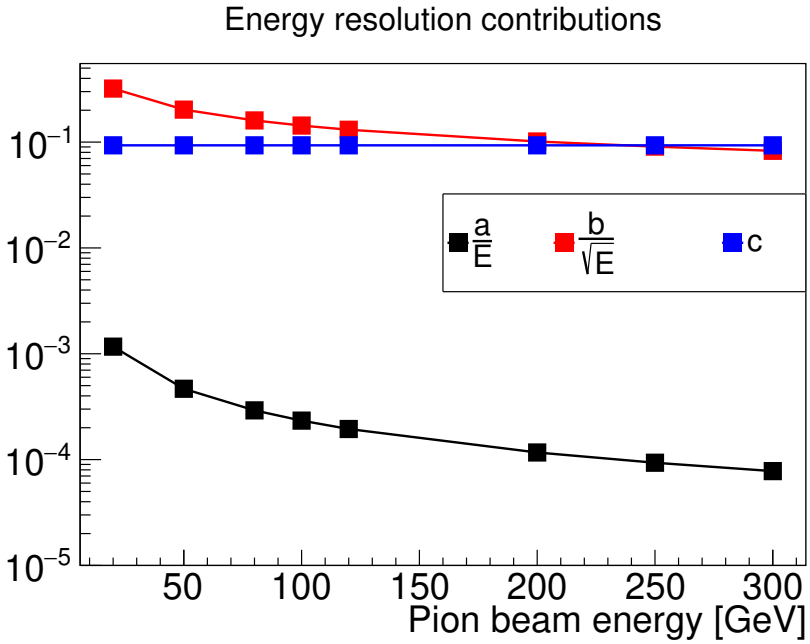


Figure 4.8: contribution of different terms in energy resolution (simulation)

With this parametrization, the pion energy resolution profile can be compared with other relevant calorimeters. HGCAL will be replacing the current endcap calorimeters so it will be important to compare its energy resolution with the current ones. Table 4.2 shows comparison of pion energy resolution for three calorimeter prototypes. One should note that the active layers in CMS ECal, HCal had scintillators in contrast to HGCAL EE and FH which had silicon sensors. Furthermore, the energy resolution parametrization used for CMS ECal, HCal had considered only the terms $\frac{\sigma}{E} = \frac{b}{\sqrt{E}} \oplus c$ while the term $\frac{a}{E}$ was calulated separately using pedestal estimation. With these differences, a common feature that can be seen in all of them is that the stochastic term dominates the resolution except for high beam energies around 300GeV.

Prototype	a [GeV]	b [GeV$^{1/2}$]	c
HGCAL	0.05	1.38	0.09
CMS HCal	0.38	1.06	0.04
CMS ECal+HCal	0.38	1.18	0.04

Table 4.2: Pion energy resolution comparison of different calorimeters. CMS ECal and HCal numbers were taken from the corresponding beam test results. [13]

In conclusion,

- the energy reconstructed using the weighted rechit energy sum shows a correlation with the shower starting compartment.
- The simulation studies set a lower limit on the minimum achievable energy resolution.
- Energy resolution parametrization reveals that instrumental factors and stochastic fluctuations dominate the energy resolution.

Chapter 5

Machine Learning for energy reconstruction

There are various methods of implementing machine learning. Artificial neural networks were used in the analysis as a method of machine learning. A neural network takes an array of numbers as input and applies mathematical operations to give an array as an output. The most commonly used mathematical operation is sequential logistic regression. In this method, logistic regression is applied multiple times to arrive at the final result. One instance of logistic regression is called a layer which takes an input array and gives an output array in the following manner.

$$X^{i+1}_{m\times 1} = \sigma(A^i_{m\times n} X^i_{n\times 1} + B^i_{m\times 1})$$

Here, input array X^i is converted to output array X^{i+1} using matrices A^i, B^i and by applying an activation function σ. The elements of A, B are free parameters of the model and are found out during training.

5.1 Introduction

5.1.1 Loss function

During training, a neural network tries to predict the right output by processing on the given input. The model parameters are updated in every epoch in order to better predict the output. The goodness of the prediction is quantified by a loss function. A loss function is a function of predicted output and the true output. Following are the commonly used loss function.

$$
\begin{aligned}
F(x_i, y_i) &= \sum_{i=1}^{N} \frac{(Y_i - y_i)^2}{\sigma_i^2} && \cdots \chi^2 \\
&= \frac{1}{N} \sum_{i=1}^{N} (Y_i - y_i)^2 && \cdots \text{Mean Squared Error (MSE)} \\
&= \sum_{i=1}^{N} -Y_i ln(y_i) + (1 - Y_i) ln(1 - y_i) && \cdots \text{Binary Cross Entropy (BCE)}
\end{aligned}
$$

Where,
x_i, y_i = the i^{th} input and output of the training dataset.
Y_i = the output predicted by the neural network for the i^{th} training dataset.
σ_i = uncertainity on the true output.
N = total number of training data points.

5.1.2 Gradient Descent

In general, the loss functions used in machine learning are of the form:

$$F(\theta, X) = \frac{1}{n} \sum_{i=1}^{n} f(\theta, X_i)$$

Where,
θ is a vector of model parameters
X is a vector of data points
X_i is the i^{th} datapoint

Gradient descent is a method to find out optimum model parameters iteratively. The algorithm updates values of the model parameters using the gradient of the loss function in the following manner:

$$\theta' = \theta - \alpha.\nabla F(\theta, X)$$

where,
θ' is the updated model parameter vector
α is the learning rate; a parameter decided at the beginning of the program.

5.1.3 Stochastic gradient descent

Computing gradient of the loss function can turn out to be computationally inefficient for large amount of data. The parameters can be estimated faster if the data is split into smaller batches. This becomes the basis for the stochastic gradient descent algorithm. In stochastic gradient descent, (SGD), the gradient of the whole loss function is approximated by a small batch of data. Hence for every epoch, the parameters get updated multiple times.

Following is the pseudocode for stochastic gradient descent:

```
for i in (1,n_epoch)
for j in (1,n_batch)
theta = theta - LR*grad()
end
end
```

5.1.4 Adam Optimizer

Adam optimizer is an improvement on the stochastic gradient descent algorithm using higher-order corrections [14]. This algorithm is used as the optimizer for all the machine learning algorithms presented in this analysis.

5.1.5 Parameter Initialization

While using Keras for training, the weights of each layer were initialized with a uniform probability distribution defined in the following way

$$\begin{aligned} p(x) &= \frac{1}{2l} && \cdots \text{for } x \in [-l, l] \\ &= 0 && \ldots \text{ elsewhere} \end{aligned}$$

where,
$l = \sqrt{\frac{6.s}{n_{in}+n_{out}}}$
$s =$ scaling. (free parameter.)
$n_{in}, n_{out} =$ no. of input and output nodes for the layer
Example:
If a hidden layer has 3 inputs and 1 output, then with $s = 1$

$$l = \sqrt{\frac{6.1}{3+1}} = \sqrt{\frac{3}{2}}$$

5.1.6 Activation function

In each of the hidden layer, The product of matrix multiplication and bias addition is further modified using an activation function. Activation function is introduced to bring in non-linearity to the network for better results. The main motivation for this came from the biology analog of activation potential of neurons. The activation functions are essentially used to choose a particular set of output nodes using a mathematical condition. For example while using relu activation function, only those nodes will contribute to the next layer which give a positive output. The negative outputs will be suppressed to zero and won't contribute to the further layers. Following are the commonly used activation functions:

$$\sigma(x) : \mathbb{R} \to \mathbb{R}$$

$$\begin{aligned} \sigma(x) &= \frac{1}{1+e^{-x}} && \cdots \text{Sigmoid} \\ &= max(0, x) && \cdots \text{Rectified Linear unit (ReLU)} \\ &= \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} && \cdots \tanh \end{aligned}$$

5.2 Linear regression using ML

The pion energy reconstruction was carried out using supervised learning. The simplest ML regression model is the linear model with one hidden layer. In its basic form, the machine learning framework accepted rechit energies recorded in the event in different parts of the detector and gave the reconstructed energy as the output. The goodness of the reconstruction was tested by comparing it with the true beam energy. The ML model was tuned with different loss functions, activation functions, and input modes to arrive at the best possible model.

The passage of high energy pions through the HGCAL test-beam setup was simulated using the GEANT4 simulation toolkit. The input to the neural network was in the form of total rechit energy recorded in EE, FH, AH. The pion beam energies considered were 20,50,80,100,120,200,250,300 GeV. Simulation samples were generated centrally by the HGCAL simulation group for each of the mentioned energies. At the time of analysis, the ML model was trained separately for different energies. A portion of the dataset was used for training and the remaining was used for testing. The additional details of the machine learning model used are listed below.

- Machine learning framework: Python(v 3.6.9), tensorflow(v 1.14.0), Keras(v 2.2.5)
- Machine learning model: $E_{Reco} = W_1 E_{EE} + W_2 E_{FH} + W_3 E_{AH}$
- Training done for 75% of available events
- Testing done for remaining events (25%)
- Loss functions: χ^2, $\frac{Std.\ dev}{mean}$, $Std.\ Dev$
- Activation function: Linear
- Number of iterations: 10,000
- Parameter estimation : using Adam optimizer

5.2.1 Choosing the loss function

The ML model was trained separately using following loss functions and the results were compared in terms of energy response and resolution (Figure 5.3).

1. Average Error : $\langle (y - y_{true})^2 \rangle$
2. Modified χ^2 : $\langle \frac{(y-y_{true})^2}{y_{true}} \rangle$
3. RMS/Mean : $\frac{\sqrt{\langle y^2 \rangle - \langle y \rangle^2}}{\langle y \rangle}$

The simplest possible loss function which will correspond to the deviation of reconstructed energy from the beam energy is the average error; hence it was the first choice of the loss function. To consider the deviation of reconstructed energy relative to the beam energy, modified χ^2 was considered. Since the end goal of the analysis was to reduce the energy resolution and RMS/Mean being one of the measures of the energy resolution, it was considered to be the third choice.
Energy resolution was then calculated with loss function χ^2 using following methods.

1. Gaussian is fitted around ± 1.5 *std. dev.* of mean of reconstructed energy distribution with parameters σ, μ. Then energy resolution $= \frac{\sigma}{\mu}$
2. Calculate std. dev, mean of the reconstructed energies. Then energy resolution $= \frac{Std.\ dev.}{Mean}$
3. Calculate std. dev for region around mean which contains 90% of the data. Then energy resolution $= \frac{Std.\ dev.}{Mean}$

The three methods were considered because the reconstructed energy distributions are not always Gaussian and there is a possibility of fluctuations far away from the mean.

In the case of RMS/Mean as cost function, the reconstructed energies had good resolution but the response was bad, as can be seen from the figure 5.3 upper left graph. Since there was no involvement of beam energy in this cost function, the ML model didn't care what the response was. Hence to reduce RMS/Mean, it increased the mean irrespective of the

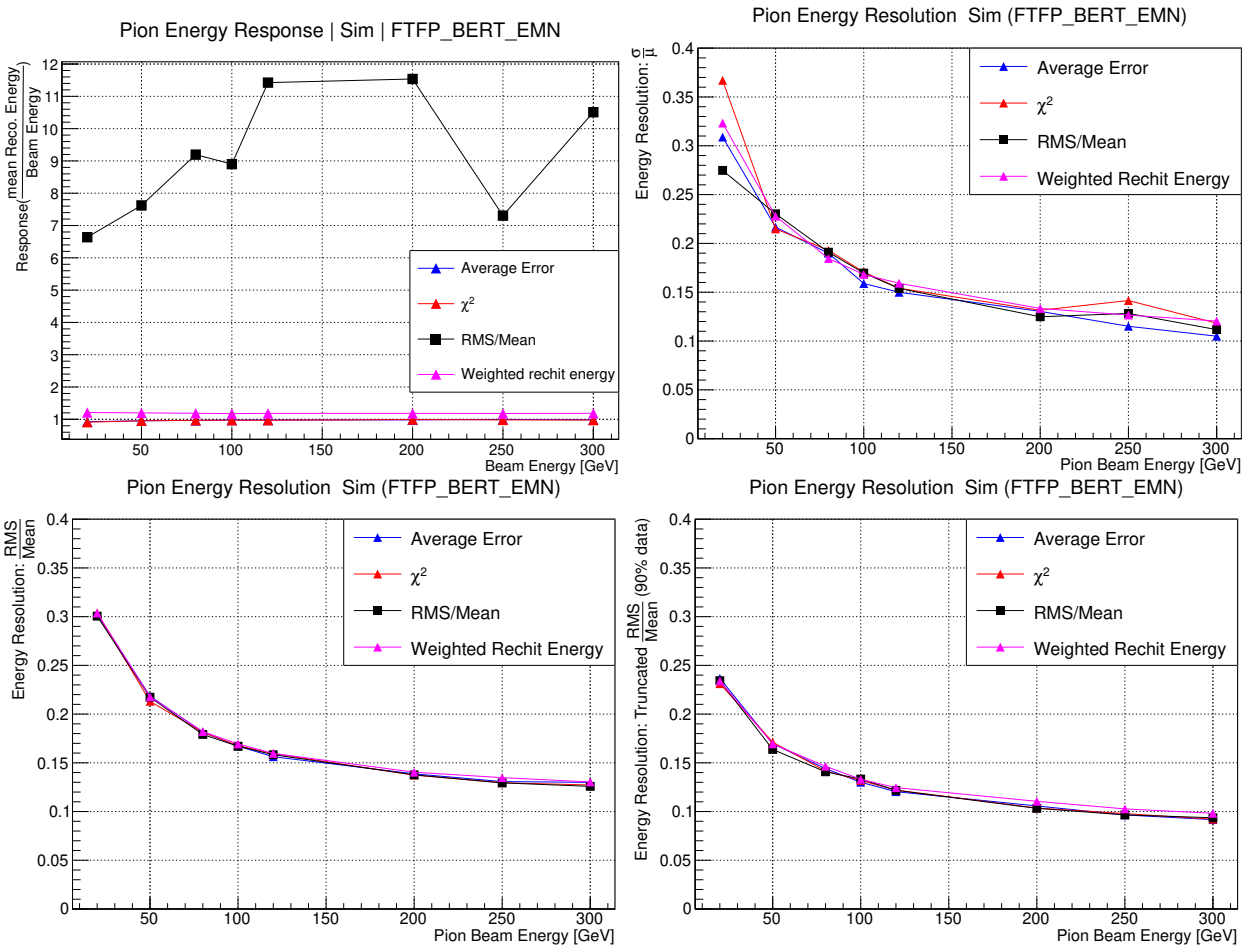


Figure 5.1: Pion Energy resolution

beam energy. On the other hand, since the ML model was trained separately for different energy samples, within a given energy sample, average error and the modified χ^2 differed only by a constant scale factor. As a result, there was no noticeable difference between the corresponding in terms of response and resolution. Therefore the modified χ^2 was considered as the cost function for the rest of the analysis. (See appendix A.8 and A.9 for reconstructed energy graphs)

5.2.2 Choosing Input nodes

The following two sets of inputs were considered for the machine learning analysis.

1. **Set 1** : 3 numbers representing total rechit energy recorded in EE, FH, AH.

2. **Set 2** : 79 numbers representing total rechit energy recorded in all 79 active layers of HGCAL respectively.

The entire machine learning analysis was done with the total rechit energy recorded by the sub-detectors EE, FH, AH. Since the information of total rechit energy recorded by each active layer in these sub-detectors was available, the second set of inputs, representing total rechit energy recorded by each active layer of HGCAL was considered for the comparison with the previous set. Since there were in total 79 active layers in HGCAL, the number of model parameters for the second set was also 79. In general, the second set of inputs was expected to give better energy response and resolution because of the finer sampling.

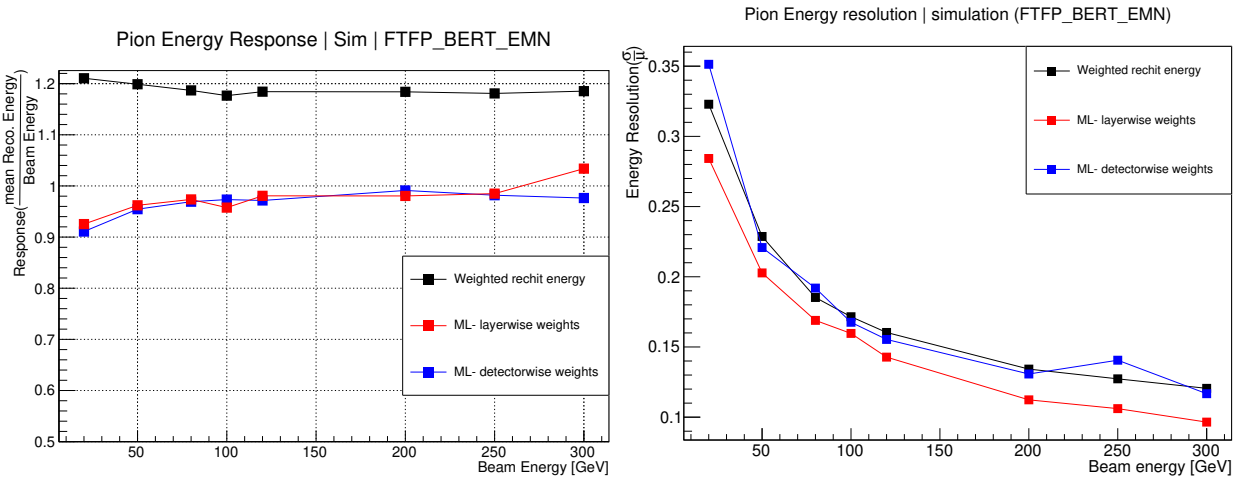


Figure 5.2: Left: Energy response ; Right: Energy resolution

Refer to figure 5.2. The energy response for both input sets deviates on an average by 0.8% while the energy resolution deviates on an average by 14%. Considering the simplicity of the model with input set 1 and no significant change in the behaviour for input set 2, the first input set consisting of total rechit energies recorded by EE, FH, AH was used for further analysis.

Shower start location	EE rechits	FH rechits
EE	> 100	no-cut
FH	< 100	> 60
mip-like	< 100	< 60

Table 5.1: Event classification

5.2.3 Event classification

The pion passing events were classified according to their shower start location. The shower start algorithm discussed in section 4.4.1 was found to be inefficient for lower beam energies. Hence a simplified criterion was used to estimate the shower start location.

The pion events were classified into three categories: shower starting in EE, shower starting in FH and mip-like events. The classification was done based on the total rechit energy recorded by EE and FH (Refer to Table 5.1). The rechit energy thresholds were set by comparing muon and pion rechit energy distributions. MIP-like events were ignored in the analysis and the neural network was trained separately for events starting shower in EE and FH. Figure 5.3 shows the energy resolution for ML models trained on different shower starting locations. It can be seen that the resolution is the best for pions starting a shower in FH. This was expected since there are seven modules per layer in FH and hence there is less amount of transverse leakage through the detector and hence lesser fluctuations in the deposited energy. The resolution improvement is the most significant for lower energies because the lower energy pions require a lesser amount of material to fully contain their energy. On the other hand, resolution for showers starting in EE is comparatively worse because EE has only one module per layer and hence there is a greater chance of transverse leakage.

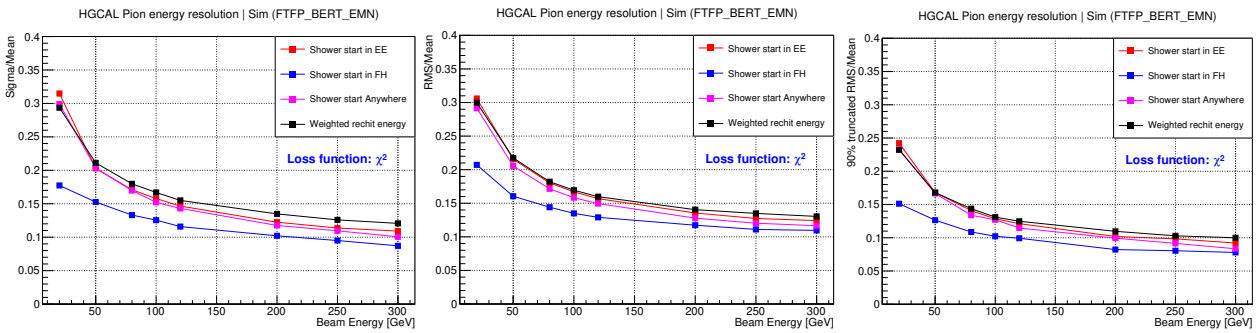


Figure 5.3: Pion Energy resolution

5.3 Conclusion

Hadronic energy reconstruction is a challenging task because of the event-by-event fluctuations in the electromagnetic fraction, uncertainty in neutral pion formation and the requirement of the large length of the detector to fully contain the shower. To address these problems, the analysis presented here included a very crude machine learning regression model. Even with this model, the energy resolution was slightly better than the previously formulated method of weighted rechit energy sum. The performance of machine learning can be improved further by considering more complex ML models.

Chapter 6

Results and Summary

The MIP calibration of the HGCAL silicon sensors was carried out using muon run data for all the central modules of FH. The MIP signal for a sensor was then defined to be the most probable value of the ADC counts. The fit model defined for this purpose considered the cases with zero, one and two muon passing events. To calculate noise for the sensors, ADC counts obtained in the 12th-time sample were used by bypassing the pre-selection criteria. The noise was defined to be the standard deviation of the gaussian fitted over the ADC distribution. The signal and noise defined in this manner were used to study signal to noise ratio (SNR) of each individual silicon sensors. The SNR calculated for about 1100 silicon sensors showed a stable behavior and is evident from a narrow peak in the histogram of SNR distribution (Figure 3.10).

The performance of pion inside HGCAL was evaluated using the energy response and resolution. To arrive at the best possible energy resolution, the rechit energy recorded in each of the subdetector of the HGCAL prototype was added by applying appropriate weights. These weights were made available from the offline discussion with Mr. Shubham Pandey. These weights were calculated separately for pions starting a shower in EE and FH. The approximate location of the pion shower starting point was predicted by the shower start algorithm (Appendix D). The accuracy of the algorithm was evident from the linearity of the distribution of shower starting points (Figure 4.4.2). The first 40 active layers of HGCAL which covered the subdetectors EE, FH were divided into 6 compartments such that each compartment has a comparable thickness in terms of interaction lengths as well

as comparable number of shower start events. Using this compartmentalization, pion energy was reconstructed separately for pions starting shower in different compartments and was compared to the transverse leakage and the energy deposited in the absorbers using GEANT4 simulation of HGCAL beam test (Figure 4.5). The energy resolution was calculated for different pion beam energies by fitting Gaussian over the reconstructed energies as well as by considering the mean and standard deviation of the whole energy distribution. The contribution of different factors to energy resolution was predicted by parametrizing the energy resolution (Figure 4.7). For the lower energy pions, statistical fluctuations dominated the energy resolution whereas for the high energy pions the detector effects dominated (Figure 4.8).

To further improve the energy resolution, a simple regression using machine learning was attempted. The analysis presented here concluded that with the total rechit energy recorded in EE, FH, AH respectively as the input to the regression model the best possible energy resolution is achieved for modified χ^2 as the loss function and for the showers starting in FH (Figure 5.3) as determined by the event selection criteria (Table 5.1).

Scope of improvement and future work

The shower start algorithm discussed in section 4.4.1 was a crude algorithm. One can use machine learning approach to classify events with different shower start locations. With a proper definition of shower start, the machine learning framework can be trained and tested on simulation data.

A very primitive neural network with only three inputs and a single hidden layer was used for energy reconstruction. The energy resolution thus obtained was comparable to the method of weighted rechit energy sum. A more complex neural network with multiple hidden layers and more number of input parameters can be used to further improve the energy resolution. The number of input parameters can be increase by including rechit energy per layer or channel-wise recorded rechit energy. Also the other known machine learning approaches such as Convolutional Neural Network, parametrized machine learning could give better results and that has to be tested.

Bibliography

[1] CMS Collaboration, "The Phase-2 Upgrade of the CMS Endcap Calorimeter," CERN-LHCC-2017-023 ; CMS-TDR-019.

[2] Thorben Quast, "Beam Tests for the CMS HGCal upgrade," The 14^{th} Gentner Day talk.

[3] N. Abgrall *et al.* [NA61 Collaboration], "NA61/SHINE facility at the CERN SPS: beams and detector system," JINST **9**, P06005 (2014) doi:10.1088/1748-0221/9/06/P06005 [arXiv:1401.4699 [physics.ins-det]].

[4] F. Tellander and N. Charitonidis, "Tertiary particle production and target optimization of the H2 beam line in the SPS North Area," CERN-ACC-NOTE-2016-0060 [arXiv:1610.09066 [physics.acc-ph]]

[5] Spanggaard, J., "Delay Wire Chambers - A Users Guide" SL-Note-98-023-BI (1998) http://cds.cern.ch/record/702443

[6] L. Rossi, P. Fischer, T. Rohe, N. Wermes. "Pixel Detectors From Fundamentals to Applications," ISBN-10 3-540-28332-3 Springer Berlin Heidelberg New York

[7] R. Wigmans, "Calorimetry: Energy measurement in particle physics," Int. Ser. Monogr. Phys. **107**, 1 (2000).

[8] N. Akchurin *et al.*, "Beam test results from a fine-sampling quartz fiber calorimeter for electron, photon and hadron detection," Nucl. Instrum. Meth. A **399**, 202 (1997). doi:10.1016/S0168-9002(97)00789-4

[9] Ugo Amaldi, "Fluctuations in Calorimetry Measurements" Phys. Scr. 23 409 (1981).

[10] L. D. Landau, "On the Energy Loss of Fast Particles by Ionization, Collected Papers of L.D.Landau" doi:10.1016/B978-0-08-010586-4.50061-4.

[11] J. Borg, S. Callier, D. Coko, F. Dulucq, C. de La Taille, L. Raux, T. Sculac and D. Thienpont, "SKIROC2_CMS an ASIC for testing CMS HGCAL," JINST **12**, no.02, C02019 (2017) doi:10.1088/1748-0221/12/02/C02019

[12] S. Agostinelli *et al.* [GEANT4 Collaboration], Nucl. Instrum. Meth. A **506**, 250 (2003). doi:10.1016/S0168-9002(03)01368-8

[13] G. Baiatian *et al.* [CMS HCAL], "Design, performance, and calibration of CMS hadron endcap calorimeters," CERN-CMS-NOTE-2008-010.

[14] D. P. Kingma and J. L. Ba, "ADAM : A Method for stochastic optimization," arXiv:1412.6980 [cs.LG], 2014.

[15] A. Penzo *et al.* [CMS Collaboration], "The CMS-HF quartz fiber calorimeters," J. Phys. Conf. Ser. **160**, 012014 (2009). doi:10.1088/1742-6596/160/1/012014

[16] Experimental techniques - Tejinder S. Virdee

[17] N. Akchurin *et al.*, "First beam tests of prototype silicon modules for the CMS High Granularity Endcap Calorimeter," JINST **13**, no. 10, P10023 (2018). doi:10.1088/1748-0221/13/10/P10023

[18] A. Lechner (CERN), "Particle interactions with matter", CAS, Erice, Italy, March 11th , 2017

[19] D. Contardo, M. Klute, J. Mans, L. Silvestris and J. Butler, "Technical Proposal for the Phase-II Upgrade of the CMS Detector," CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02.

[20] C. W. Fabjan and F. Gianotti, "Calorimetry for particle physics," Rev. Mod. Phys. **75**, 1243 (2003). doi:10.1103/RevModPhys.75.1243

[21] F. Beaudette [CMS Collaboration], arXiv:1401.8155 [hep-ex].

Appendix A

Histograms and graphs

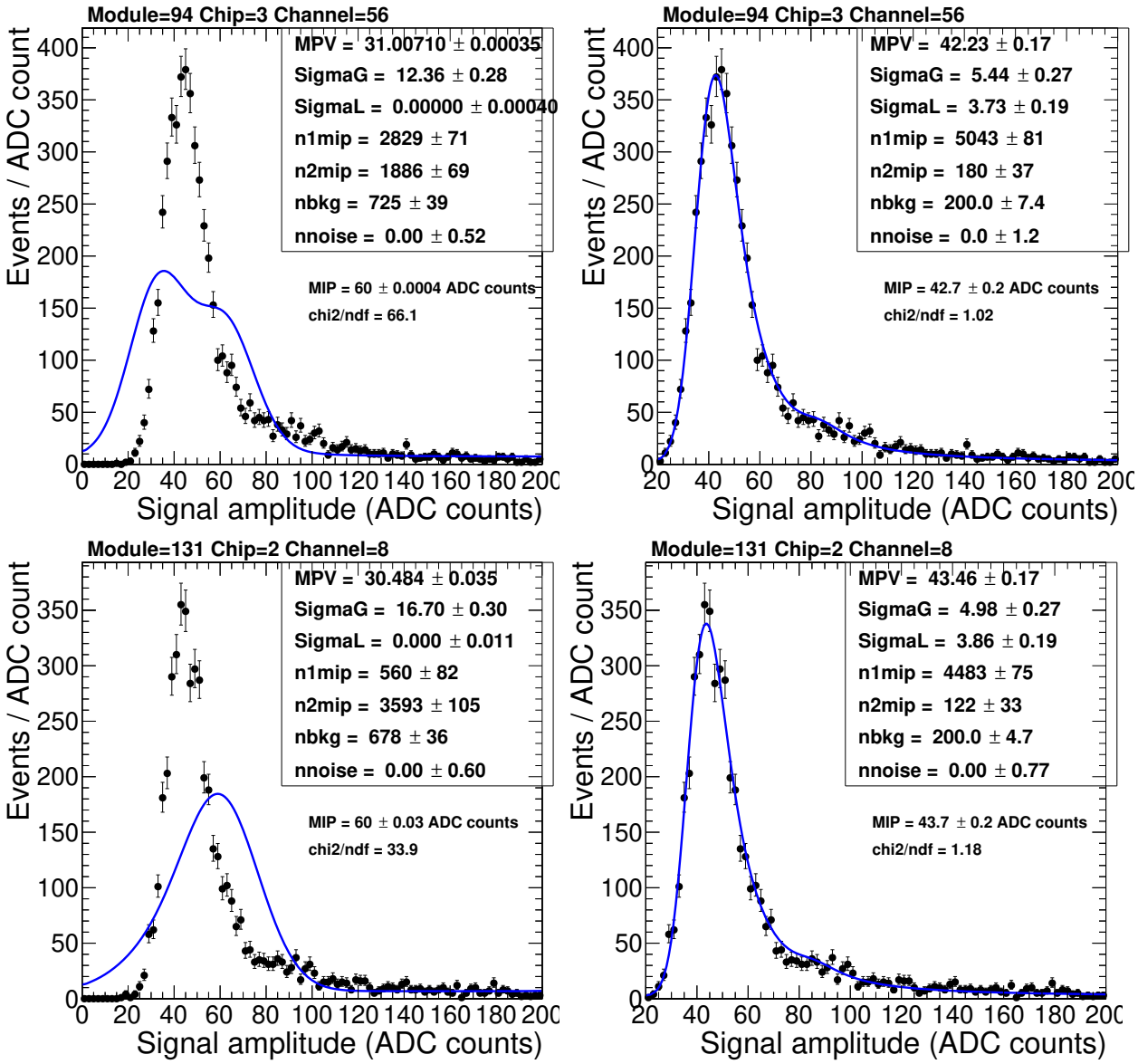


Figure A.1: HGCAL FH noise - bad fits **Left:** Default settings ; **Right:** Modified settings.

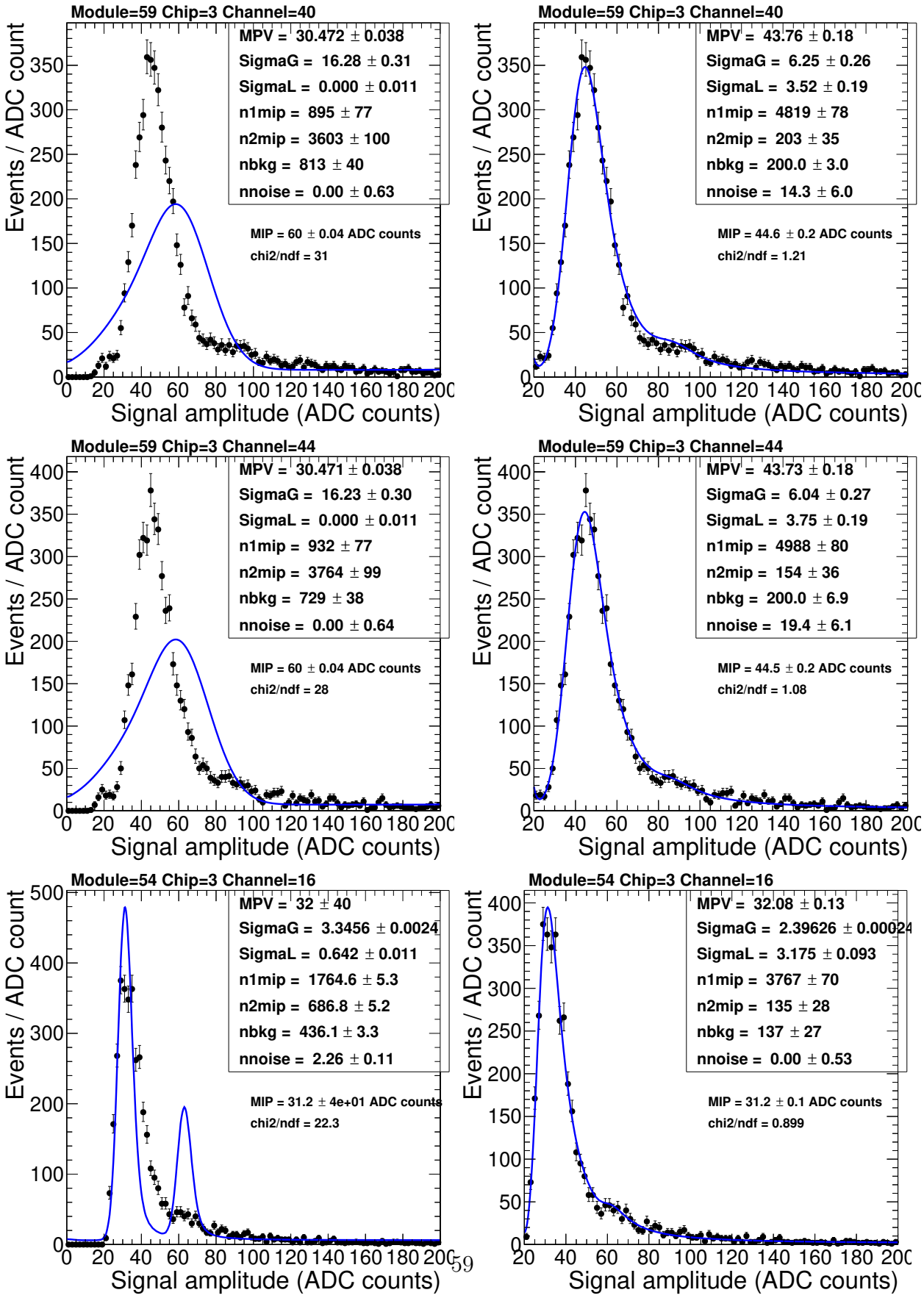


Figure A.2: HGCAL FH noise - bad fits (continuation)

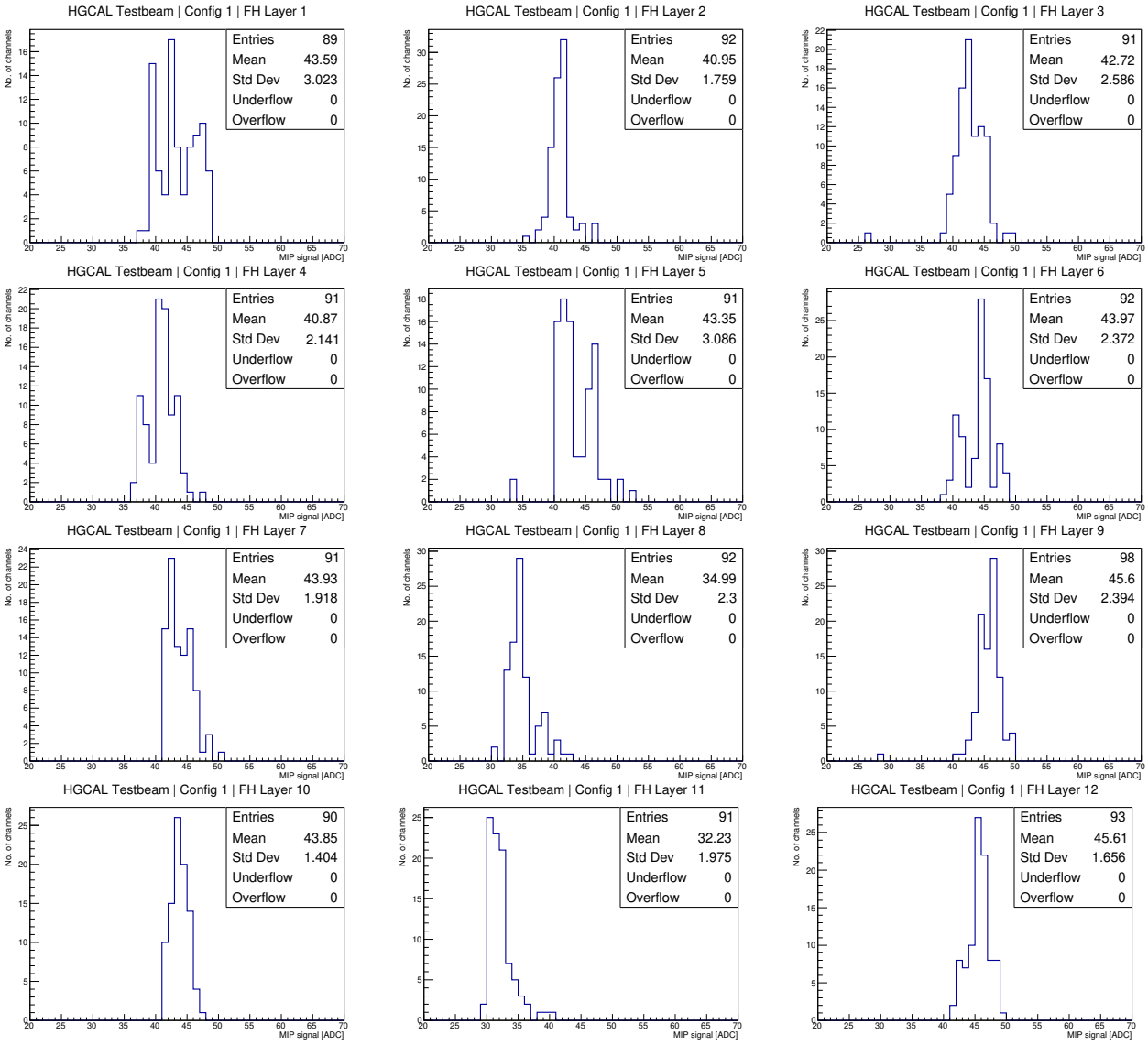


Figure A.3: Signal variation per layer

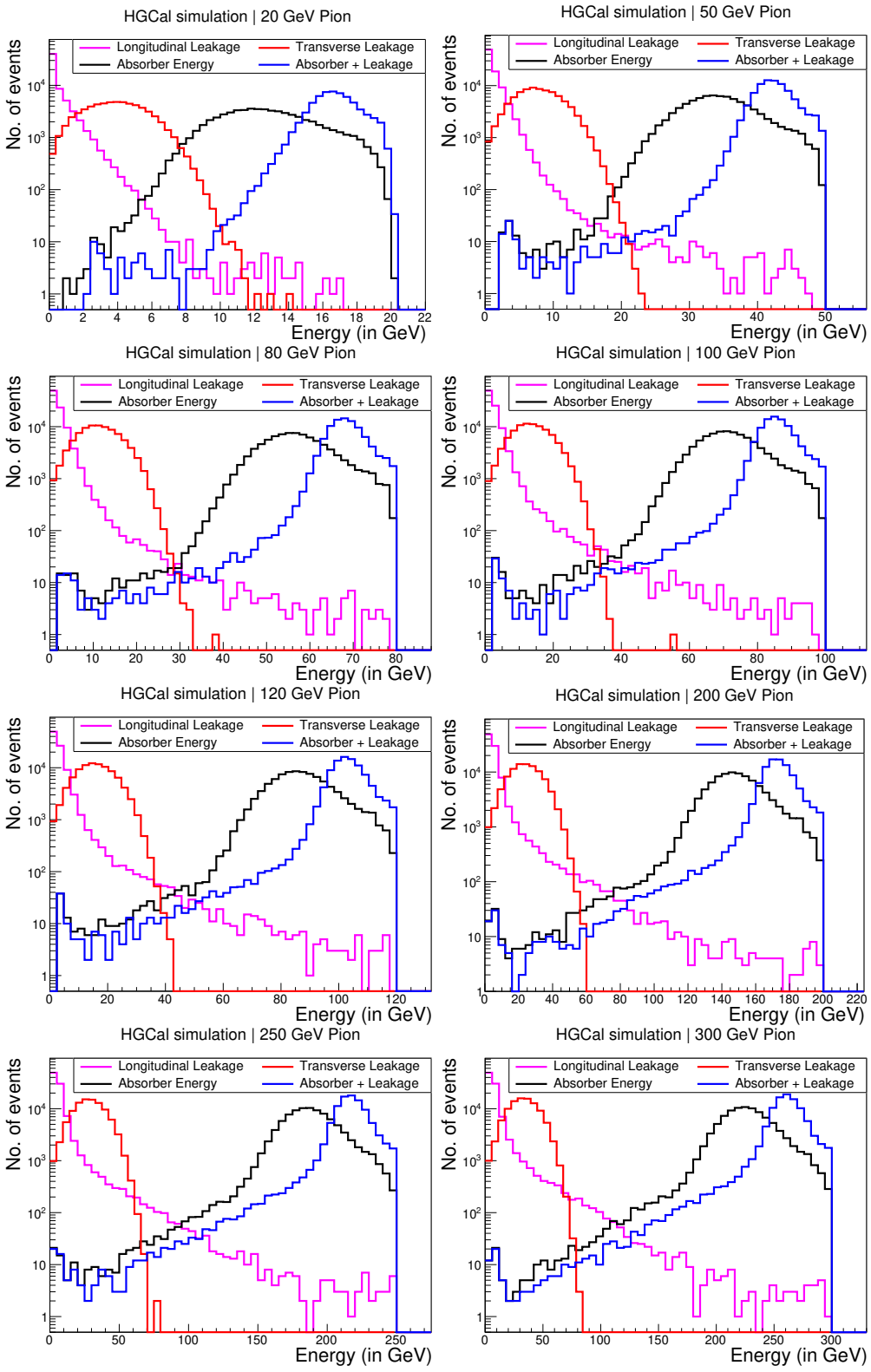


Figure A.4: Pion absorbed and leakage energy

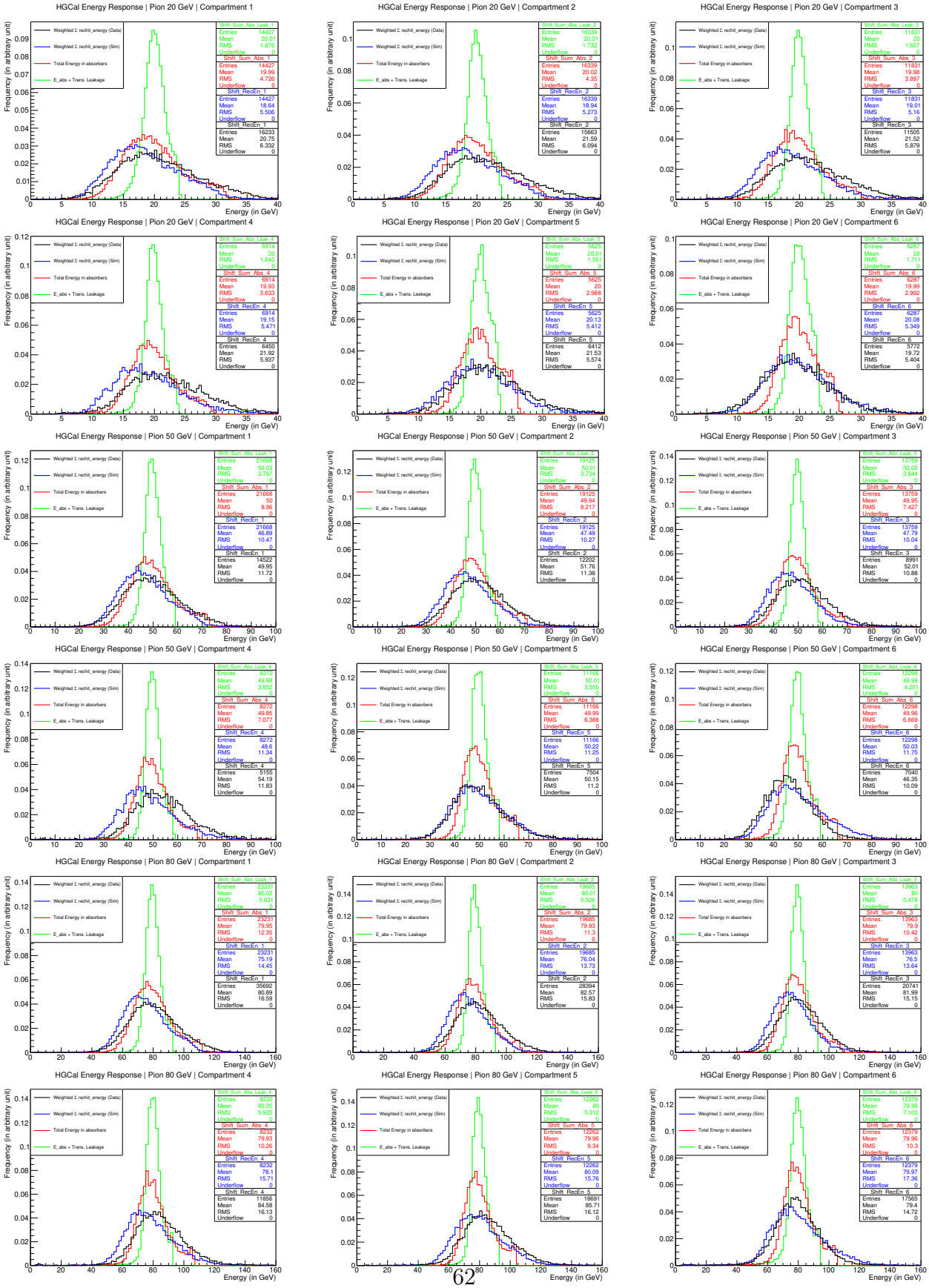


Figure A.5: Compartment-wise reconstructed energies (Pion Energies 20, 50, 80 GeV)

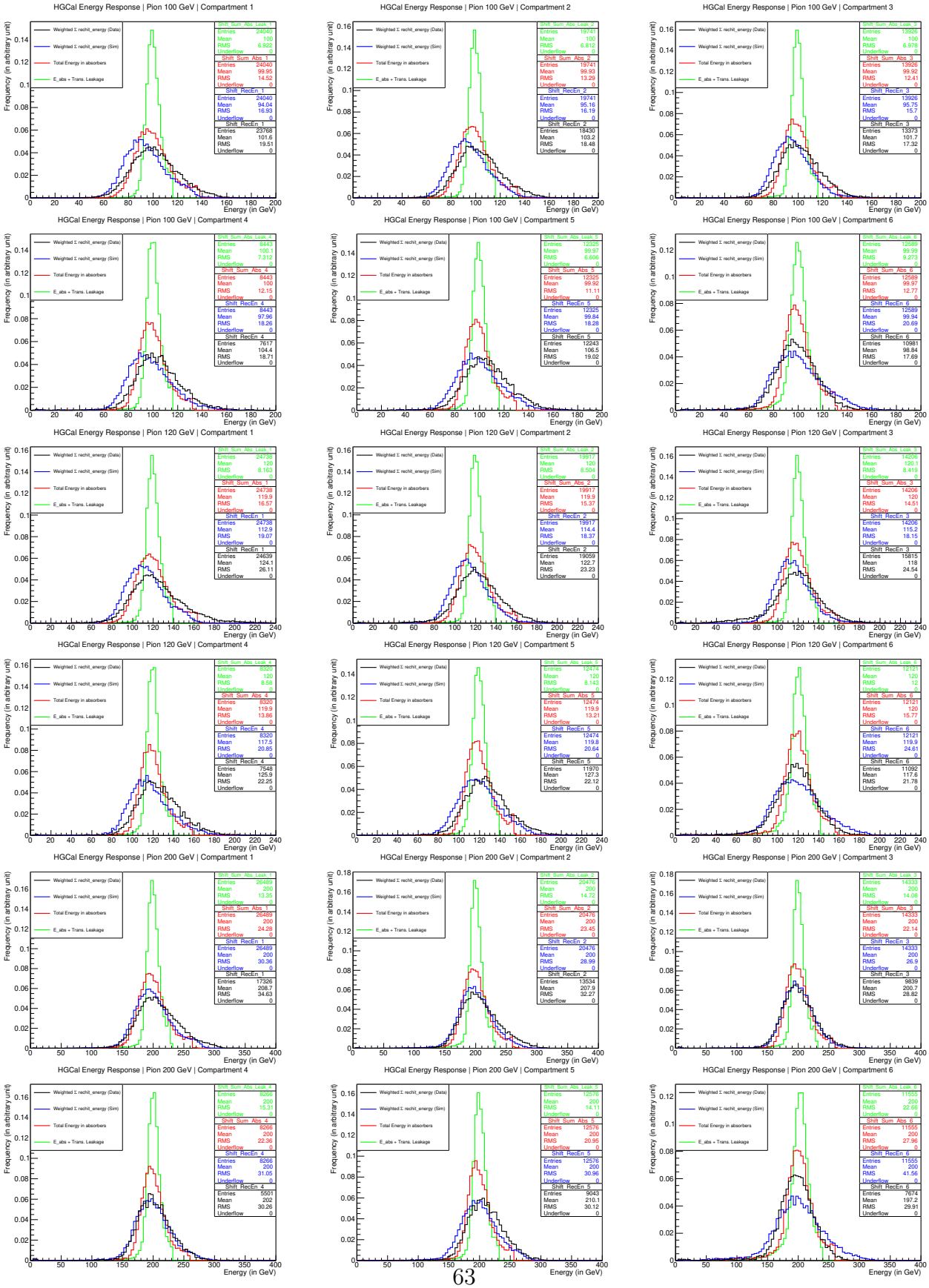


Figure A.6: Compartment-wise reconstructed energies (Pion Energies 100, 120, 200 GeV)

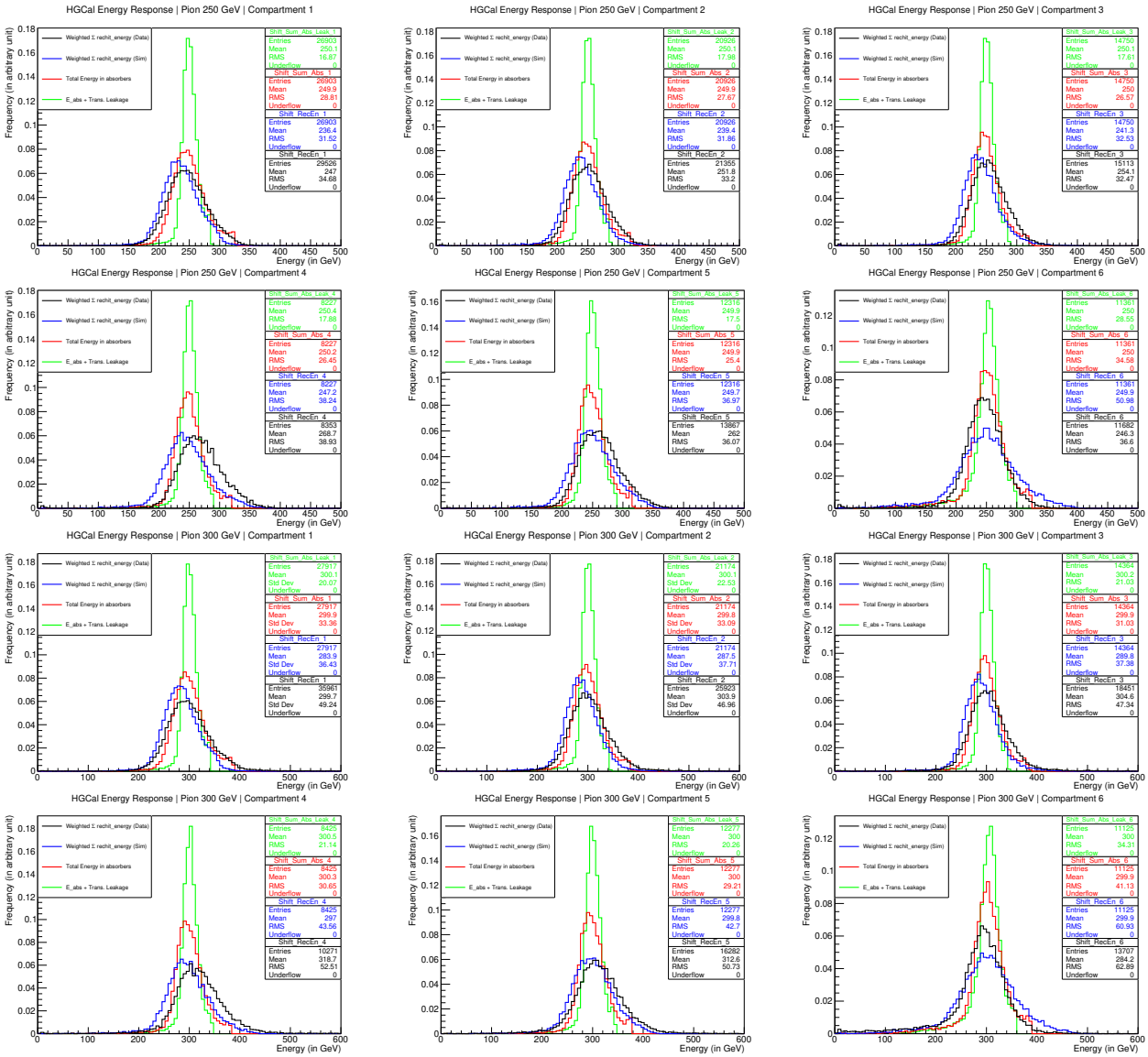


Figure A.7: Compartment-wise reconstructed energies (Pion Energies 250, 300 GeV)

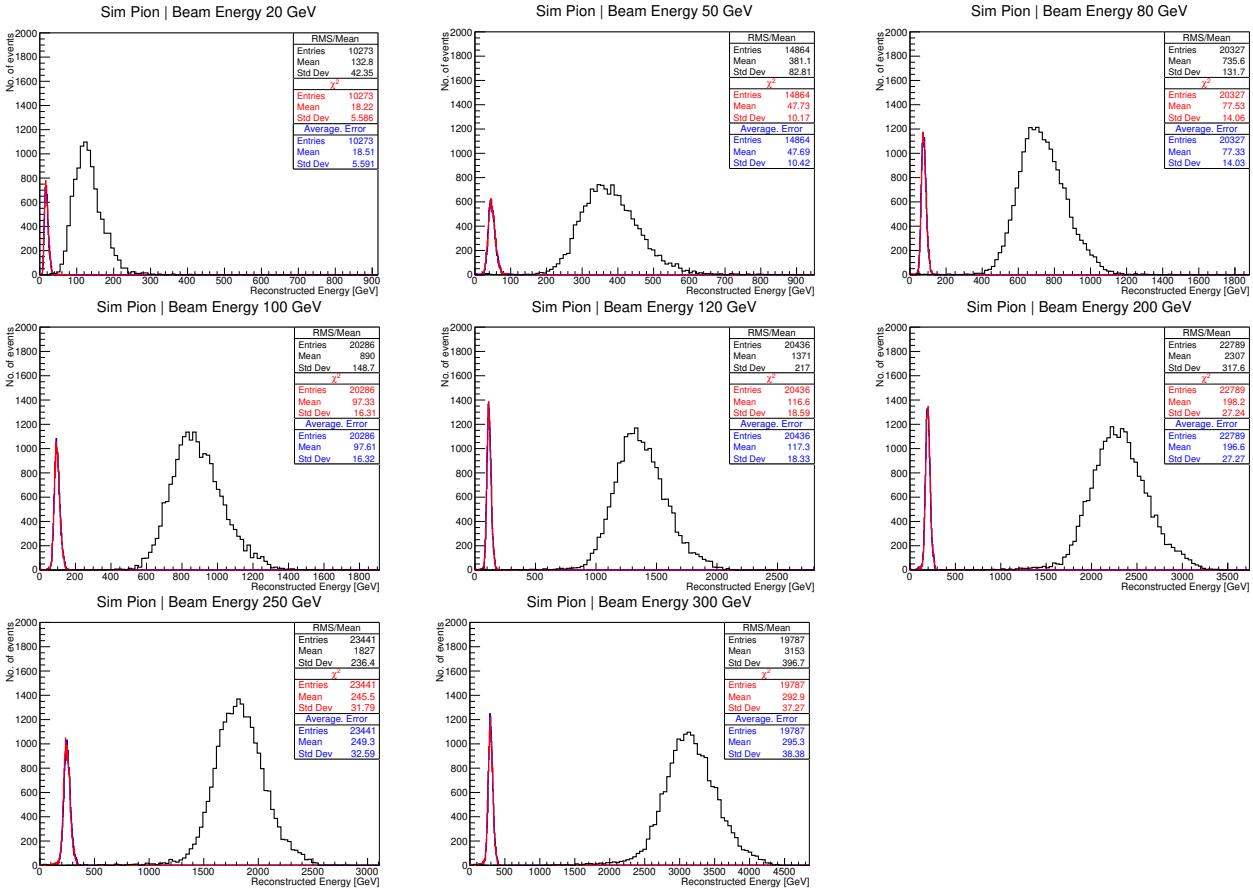


Figure A.8: Raw reco. energy distributions for different cost functions

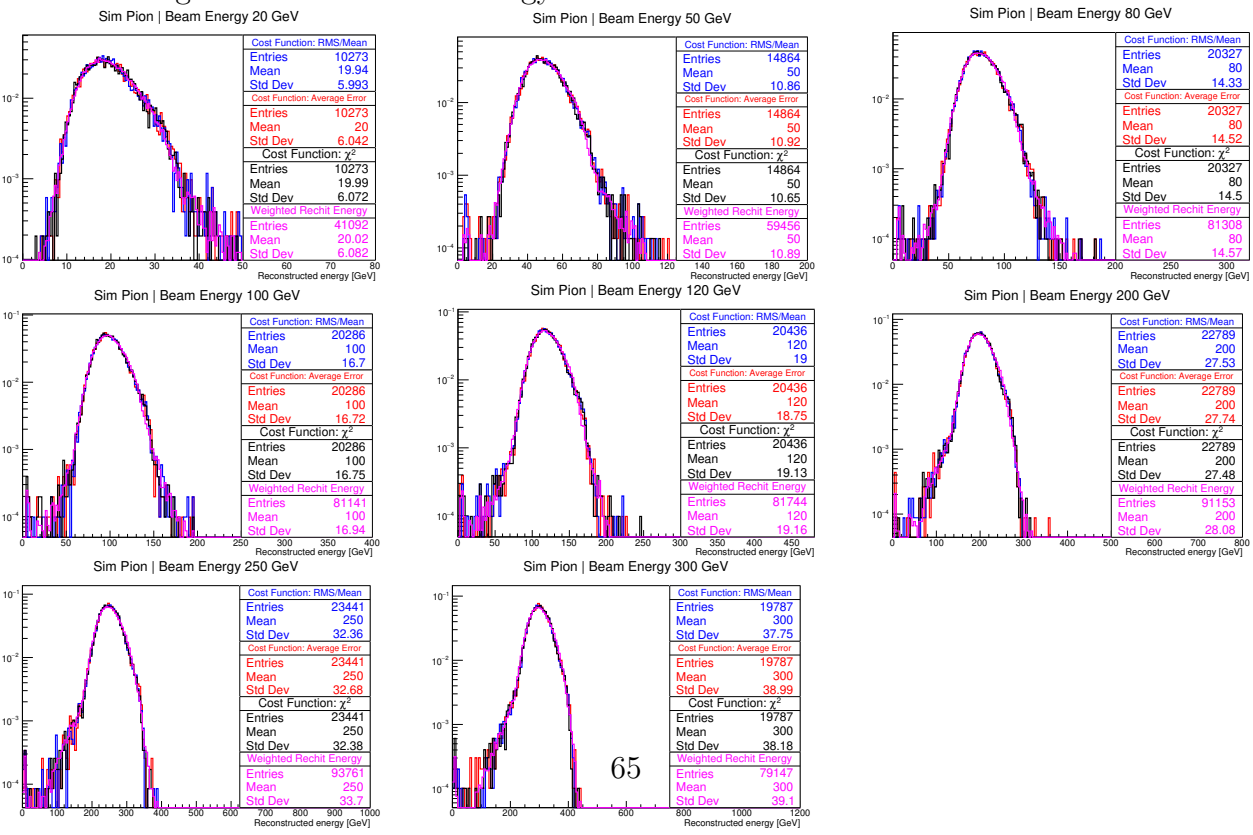


Figure A.9: rescaled reco. energy distributions for different cost functions

Appendix B

Weighted rechit energy

B.1 Finding α

1. Use the shower start algorithm and consider events where shower started in FH
2. Choose some value for α.
3. Plot $E_{FH} + \alpha\ E_{AH}$ for these events
4. Fit gaussian to the histogram and extract the std. Dev.
5. Plot a graph of std.dev vs α
6. Choose that α for which you get minimum in the graph.

B.2 Finding β

1. Use the shower start algorithm and consider events where shower started in EE
2. Choose some value for β.
3. Plot $E_{EE}+ \beta\ (E_{FH} + \alpha\ E_{AH})$ for these events. (take A as calculated previously)
4. Fit gaussian to the histogram and extract the std. Dev.

5. Plot a graph of std.dev vs β

6. Choose that β for which you get minimum in the graph.

B.3 Finding γ

γ is a conversion factor between MIP to GeV. Once α, β are evaluated, γ can be obtained by:

$$\gamma = \frac{beam\ energy\ (in\ GeV)}{E_{EE} + \beta(E_{FH} + \alpha E_{AH})}$$

Appendix C

QFCAL

QFCAL stands for Quartz Fibre Calorimeter. This is a scintillating detector currently used in the forward region of CMS ($3 < \eta < 5$) (also known as Hadron Forward calorimeter) designed especially for radiation hardness and good optical transparency. This detector is essentially sensitive only to the electromagnetic showers with e/h$\sim$5 [15]. Hence in the case of hadron showers, the energy is mainly measured from their electromagnetic component.

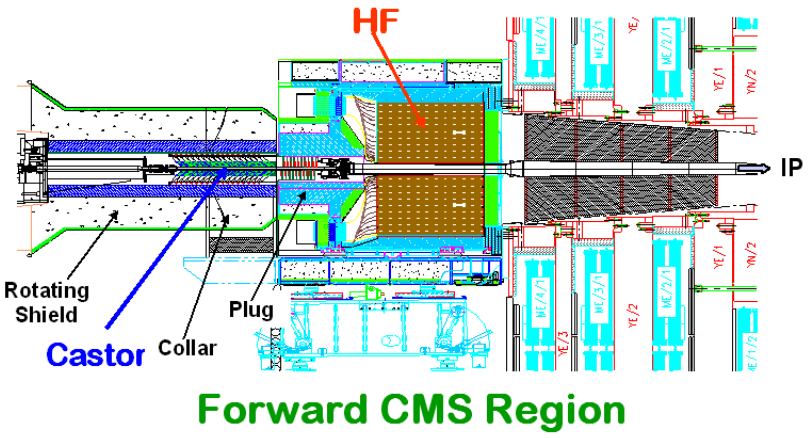


Figure C.1: Forward CMS region - schematic diagram [15]

Appendix D

Shower start algorithm

Aim: To predict index of layer closest to the shower start.
SSI = the layer closest to the shower start
E_2 = array that stores layerwise total energy deposit within radius of 2cm around the beam axis

1. For every rechit, calculate dR between rechit and the intersection of beam axis and the layer
2. If $E_2[1] > 20$, SSI= 1
3. Else if $E_2[2] > 20$ and also $E_2[2] > 2.E_2[1]$ then SSI=2
4. Else SSI = L where L (in the range (3,40)) satisfies :
    - $E_2[L] > 20$
    - $E_2[L] > 2.E_2[L-1]$
    - $E_2[L] > 2.E_2[L-2]$
5. Else pion will be considered MIP-like. (SSI=-1)

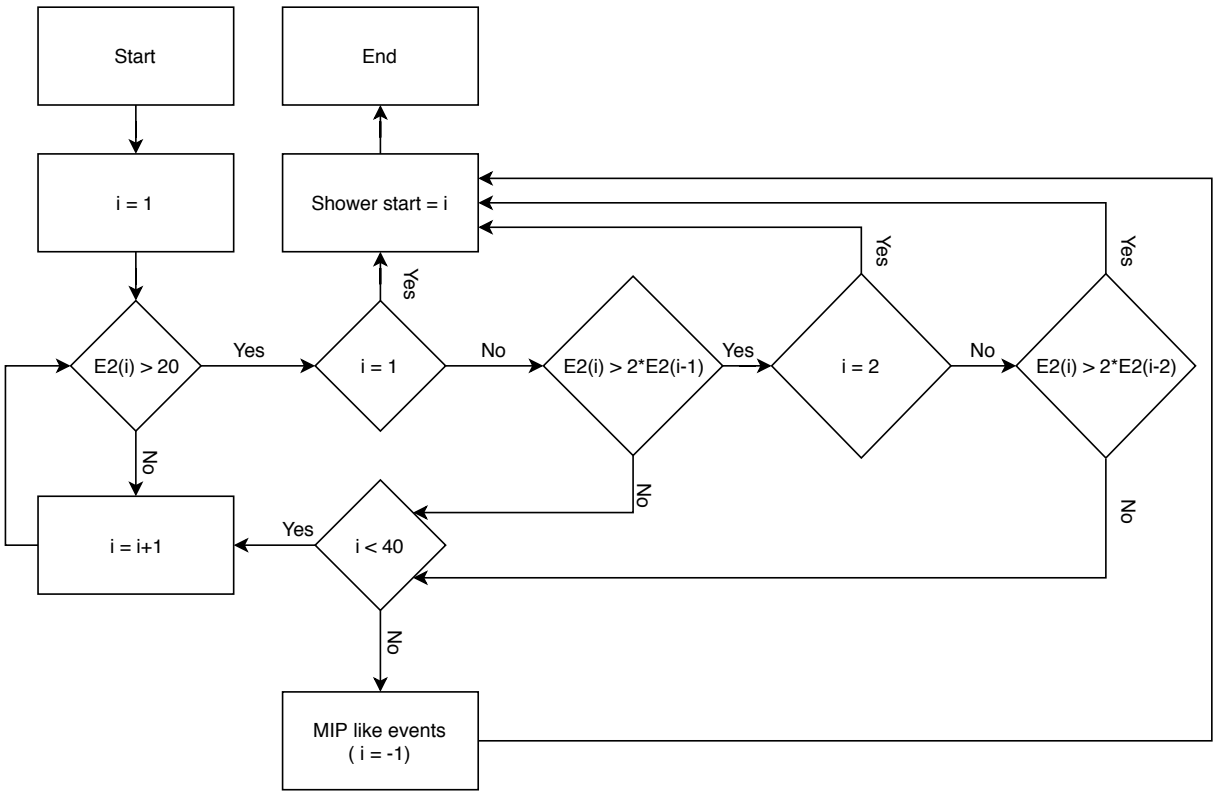


Figure D.1: Flowchart: Shower-start algorithm. E2(i) represents total rechit energy of all the rechits within 2cm radius of the beam axis in the i^{th} layer