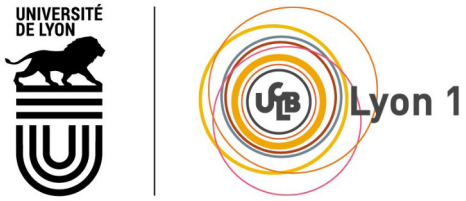
Etude du calorilmètre hadronique semi-digital et étude du canal physique e+ e- -H nu nu (H -> WW->qq qq) au collisionneur circulaire electron positon (CEPC)

Bing Liu



THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École doctorale ED52
Physique et d'Astrophysique de Lyon (PHAST)

Spécialité de doctorat : Physique des particules

Soutenue publiquement le 20/11/2020, par :
Bing LIU

Study of semi-digital hadronic calorimeter and study of physics channel $e^+e^- \to \nu_e\bar{\nu}_e\, H(H \to W^+W^- \to q\bar{q}'q\bar{q}')$ in the Circular Electron Positrion Collider of $\sqrt{s} = 240GeV$

Devant le jury composé de :

M. BRIENT Jean-Claude, Directeur de recherche CNRS, LLR **Rapporteur**
Mme. WANG Xiaolian, Professeure de Université de Sciences et Technologie de Chine, USTC **Rapporteur**
M. RUAN Manqi, Chercheur de Institute of High Energy Physics , IHEP **Examinateur**
Mme GASCON Susan, Professeur des universités, Université Lyon 1 **Examinatrice**

M. LAKTINEH Imad, Professeur des universités, Université Lyon 1 **Directeur de thèse**

M. YANG Haijun, Professeur de Shanghai Jiao Tong Université, SJTU **Directeur de thèse**

Abstract

The discovery of the Higgs boson by the ATLAS and CMS experiments in 2012 opens a new era for particle physics. Thus the high precision measurements of the properties of the Higgs boson is one of the most important goals for future leptonic collider projects such as ILC, CEPC, CLIC, FCC in the following decades. For those future experiments, obtaining excellent Jet Energy Resolution (JER) is very important for their physics study performance. In order to achieve this purpose, Particle Flow Algorithms (PFA) provides one of the most satisfying approaches. For the application of PFA, the high-granularity calorimeter is needed. The SDHCAL prototype which is the first technological one among the CALICE family of high granularity calorimeters, has been built and exposed to different particle beams in tests at PS and SPS of CERN. It is also one of the options of hadornic calorimeter equipping in ILD and CEPC baseline detectors.

Based on the test beam data of SDHCAL, this thesis presents the detail study about the SDHCAL technological prototype including the measurements of efficiency and multiplicity of glass-based RPC detectors, the homogenization process of detector response, the particle identification and energy reconstruction.

Using the collected beam muon events, the efficiency and multiplicity of GRPC is found to be around 96% ± 0.04 and 1.81 ± 0.19. Based on the test beam data of 50 GeV and 70 GeV pions taken at 2018, the relative deviation of response (nHit) of detectors improves from 4.9% and 3.9% to 1.8% and 1.4% for 50 and 70 GeV pions respectively. Based on the test beam data collected at 2015, this thesis adopts the BDT method to study the particle identification of SDHCAL technological prototype. It significantly improves the signal (pion) efficiency (>99%) and background rejection rates (>99%) comparing with the cut-based method.

Based on the simulation of SDHCAL technological prototype, this thesis adopts the MVA methods including BDT and MLP to improve the reconstructed energy linearity and resolution. With the input of extra energy-correlated information except the nHit information, the energy linearity is found to improve from the 3-4% to 1-2%. The energy resolution was relatively improved from 15% at 10 GeV and 3-5% at higher energies (20-80 GeV).

Finally, based on the simulation samples of CEPC operating at $\sqrt{s}$ = 240 GeV with baseline detector ((S)DHCAL option), this thesis studies the Higgs coupling constants g_{HWW} via the WW fusion process: $e^+e^- \to \nu_e\bar{\nu}_e\, H(H \to W^+W^- \to q\bar{q}'q\bar{q}')$. The precision of measurement of g_{HWW} is expected to be 2.24% (stats) ± 4.51% (sys) including both statistical and systematical uncertainties.

Résumé

La découverte du boson de Higgs par les expériences ATLAS et CMS en 2012 ouvre une nouvelle ère pour la physique des particules. Ainsi, les mesures de haute précision des propriétés du boson de Higgs sont l'un des objectifs les plus importants pour les futurs projets de collisionneurs leptoniques tels que ILC, CEPC, CLIC, FCC dans les décennies suivantes. Pour ces futures expériences, l'obtention d'une excellente résolution d'énergie de jet (JER) est très importante pour leurs performances. Afin d'atteindre cet objectif, les algorithmes de flux de particules (PFA) fournissent l'une des approches les plus satisfaisantes. Pour l'application de PFA, des calorimètres à grande granularité sont nécessaires. Le prototype SDHCAL qui est le premier prototype technologique de la famille de calorimètres granulaires développés par la collaboration internationale CALICE, a été construit et exposé à différents faisceaux de particules lors des tests au PS et au SPS du CERN. C'est également l'une des options du calorimètre hadornique proposées pour équiper les détecteurs de base ILD et CEPC.

Basée sur les données de faisceaux de test de SDHCAL, cette thèse présente l'étude détaillée du prototype technologique SDHCAL comprenant les mesures d'efficacité et de multiplicité des détecteurs RPC à base de verre, le processus d'homogénéisation de la réponse du détecteur, l'identification des particules et la reconstruction d'énergie.

En utilisant les événements muons du faisceau collectés, l'efficacité et la multiplicité du GRPC se situent autour de 96% ± 0.04 et 1.81 ± 0.19. Sur la base des données du faisceau d'essai de pions 50 GeV et 70 GeV prises en 2018, l'écart relatif de réponse (nHit) des détecteurs passe de 4.9% et 3.9% à 1.8% et 1.4% pour les pions 50 et 70 GeV respectivement. En exploitant les données de test sur faisceaux collectées en 2015, cette thèse adopte la méthode BDT pour étudier l'identification des particules du prototype technologique SDHCAL. Il améliore considérablement l'efficacité de sélection des pions (> 99 %) qui constituent notre signal et également les taux de rejet de (> 99 %) des autres particules qui constituent un fond pour notre signal, par rapport à une méthode classique utilisée auparavant et basée sur une sélection séquentielle.

Basée sur la simulation du prototype technologique SDHCAL, cette thèse adopte les méthodes MVA incluant BDT et MLP pour améliorer la linéarité et la résolution de l'énergie reconstruite. Avec l'utilisation des variables corrélées à l'énergie, à l'exception des informations nHit (déjà utilisé dans une méthode standard), la linéarité de l'énergie est améliorée de 3-4 % à 1-2 %. La résolution en énergie a été aussi améliorée de 15 % à 10 GeV et de 3-5 % à des énergies plus élevées (20-80 GeV).

Enfin, à partir des échantillons de simulation de CEPC fonctionnant à $\sqrt{s}$ = 240 GeV et en utilisant le calorimètre hadronique DHCAL, cette thèse étudie les constantes de couplage de Higgs g_{HWW} via le processus de fusion WW : $e^+e^- \rightarrow \nu_e\bar{\nu}_e\, H(H \rightarrow W^+W^- \rightarrow$

$q\bar{q}'q\bar{q}'$). La précision de la mesure de g_{HWW} devrait être de 2.24% (stats) ± 4.51 % (sys), y compris les incertitudes statistiques et systématiques.

Acknowledgment

Firstly, I want to deeply appreciate my supervisor Haijun Yang and co-supervisor Imad Laktineh. Thanks to the guide of Haijun, he opens the new world of high energy physics for me, thus I can have the opportunity to explore the secrets of the Universe. He also gives me many valuable suggestions not only for my research but also for my daily life. His strict attitude for research also importantly influences me. With the help of him, I can have the opportunity to study and work on the SDHCAL group. I also want to give my most sincere appreciation to Imad Laktineh. He guides me to the new area of calorimetry for high energy experiments. His strong curiosity and passion for the physics gives me deep impression and I benefited from it a lot. The discussions with him about research problem always inspire me. He also teaches me not to obey the authority and encourages me to say no and to question.

I also want to express my deep gratitude to the colleagues from SDHCAL group of CALICE Collaboration and SHanghai Jiao Tong University (SJTU). For SDHCAL group, Dr. Guillaume, he helps me a lot when I am fresh to join the SDHCAL group. His kind helps and humor make me more easily to adapt the life in Lyon. The discussion with him about SDHCAL gives me a lot inspiration. I would also to appreciate other SDHCAL colleagues: Gerald, Malik, Djamel, Mary-Cruz, post-doc Bo, Dr. Konstantin and Dr. Yvan. For SJTU, I'm also grateful to Jun Guo, Liang Li, Shu Li, Kun Liu and Weihao Wu, Qiuping Shen and Francois for their help.

I would like to express my appreciation to my Jury members: Jean-Claude Brient, Xiaolian Wang, Susan Gascon and Manqi Ruan.

I also appreciate all people who have shared the six years journey with me including Zirui Wang, Yushan Sun, Jing Li, Qiuping Shen and many others.

Finally, I want to appreciate my parents and my wife. They always support me and encourage me when I meet challengies.

Contents

Introduction

The standard Model represents a triumph of modern particle physics. All the particle predicted by Standard Model are completed by the discovery of Higgs Boson in 2012 with ATLAS and CMS detectors of LHC. It opens a new era for particle physics. Thus the high precision measurements of the Higgs boson properties is one of the most important goals for future high-energy physics experiments in the following decades. Thus, several leptonic colliders are proposed such as ILC, CEPC, CLIC and FCC. For those projects, obtaining excellent Jet Energy Resolution (JER) is very important for their physics study and Particle Flow Algorithms (PFA) provides one of the most satisfying approaches to achieve this purpose. For the application of PFA, the high-granularity calorimeter is needed. The SDHCAL prototype is the first technological prototype among the family of high granularity calorimeters developed by the CALICE collaboration. It is also one of the options of hadornic calorimeter equipping in ILD and CEPC baseline detectors. It has been built and exposed to different particle beams in tests at PS and SPS of CERN, thus this thesis mainly presents the work about the study of SDHCAL technological prototype with the use of test beam data collected.

In the chapter 1, a brief introduction to the Standard Model is given. The standard Model provides an overview of the fundamental particles and the interactions between these particles. In this chapter, an introduction to the the Higgs boson and Higgs mechanism is also given. Finally, some unanswered questions beyond the Standard Model are briefly discussed at the end of this chapter.

In the chapter 2, an introduction about Circular Electron Positron Collider (CEPC) is presented. It includes the physics potential, accelerator complex and the detectors. At the end of this chapter, a very brifly introduction about other future collider projects such as ILC, CLIC, FCC is given.

In the chapter 3, the basics of calorimetry are introduced which includes the interaction between the matter and particle, the electromagnetic and hadronic shower development.

In the chapter 4, a detail study about technological prototype is presented. It includes the study about the efficiency and multiplicity of glass-based RPC, the physics event reconstruction and the standard method of pion event selection and pion energy reconstruction.

In the chapter 5, we present a new method based on the boosted decision tree (BDT)

for particle identification. The comparison between BDT method and standard method is also discussed.

In the chapter 6, we present the new methods for pion energy reconstruction which are based on MultiVariate data Analysis (MVA) techniques: the Multi-Layer Perceptron (MLP) and the Boosted Decision Trees with Gradient Boost (BDTG).

In the chapter 7, we present a study about the measurement of Higgs coupling constants g_{HWW} in the CEPC via the WW fusion process: $e^+e^- \to \nu_e\bar{\nu}_e\ H(H \to W^+W^- \to q\bar{q}'q\bar{q}')$.

CHAPTER 1

The theory of Standard Model

The Standard Model is one of the most successful model of modern physics. The purpose of this chapter is to give a brief introduction to this model. It provides an overview of the fundamental particles and the interactions between these particles. It also gives an introduction to the properties of the Higgs boson which is the last particle predicated by the Standard Model to be observed. Finally, although the Standard Model is a remarkable achievement to provide a description of our world, there are some unanswered questions in this model, so a brief introduction to some open issues will be discussed at the end of this chapter.

Content

1.1 The introduction of the Standard Model

Particle physics is the core for understanding the Universe. Since 1930s, the work of thousands of physicists have led to a deep insight into the fundamental structure of matter that the Universe is made up of. It is found that the Universe is built of a small number of basic constituents called elementary particles and is governed by four fundamental interactions. Our current best understanding of this is encapsulated in the Standard Model of particle physics which also provides a unified picture about the interactions among elementary particles. Remarkably, the Standard Model provided good explanation of the results of a large number of experiments and it also accurately predicted a wide variety of phenomena. Therefore, Standard Model is the victory symbol of modern physics.

1.1.1 The fundamental particles

The matter of the Universe around us appears to be made of several different particles as the building blocks of matter. The known matter is made of atoms. Inside the atoms, there are the negatively charged electrons (e^-) spinning around a central nucleus comprised by the positively charged protons (p) and electrically neutral neutrons (n). In particular, the proton and neutron are also found to be composite objects at high energy scales as indicated in Fig. 1.1. The protons and neutrons are the bound states of genuinely elementary particles called **quarks** [1], with the proton composed of two up-quarks and a down-quark and the neutron composed of two down-quarks and a up-quark [1]. Then the basic building blocks of the low-energy Universe are completed by the nearly massless and electrically neutral electron neutrino (ν_e) which exists in some cases like nuclear $\beta-$decays of certain radioactive isotopes and the nuclear fusion. It is worth noting that electron was discovered by Joseph John Thomson, John Townsend and Harold A.Wilson [3] in 1897, while the electron neutrino was first discovered by Frederick Reines and Clyde Cowan in 1956 [4].

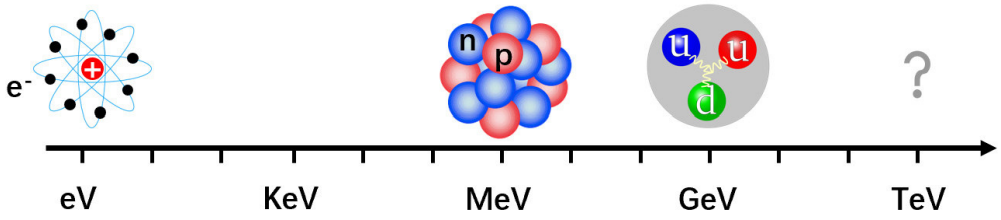


Figure 1.1: *The structure of the matter depending the energy scale.*

The electron, the electron neutrino, the up-quark and the down-quark are known collectively as the first generation. Indeed, with the development of modern particles especially the studies on the experiments based on high-energy colliders, more complexity

[1]The quantum number of hadrons (such as protons, neutrons) are determined by the so called valence quarks. Except for these, actually, any hadron can contain an indefinite number of virtual "sea" quarks, anti-quarks and gluons. [2].

was revealed. For each of the four particles in the first-generation, there are exactly two copies which have the same spin and electrically charge but with different masses. These extra eight particles are referred to as the second and the third generations. The second generation includes muon (μ^-), muon neutrino (ν_μ), strange-quark (s) and charm-quark (c). The third generation is filled by tau-lepton (τ^-), tau neutrino (ν_τ), bottom-quark (b) and top-quark (t). The electron, muon, tau-lepton and their corresponding neutrinos are collectively referred to as the **leptons** to distinguish them from quarks. For each of the quarks of the same generation, the electrical charge difference is $1e$ and the same for the leptons of the same generation. These twelve fundamental spin-half particles which are also called as **fermions** summarized in the Tab. 1.1 and Fig. 1.2.

	Leptons			Quarks		
Generation	Particle	Q Charge	Mass/c^2	Particle	Q Charge	Mass/c^2
First	e^-	$-1e$	511 keV	d	$-1/3e$	4.8 MeV
	ν_e	0	< 2.2 eV	u	$2/3e$	2.3 MeV
Second	μ^-	$-1e$	105.7 MeV	s	$-1/3e$	95 MeV
	ν_μ	0	< 0.17 MeV	c	$2/3e$	1.275 GeV
Third	τ^-	$-1e$	1.777 GeV	b	$-1/3e$	4.18 GeV
	ν_τ	0	< 15.5 MeV	t	$2/3e$	173.2 GeV

Table 1.1: *The properties of twelve elementary fermions.*

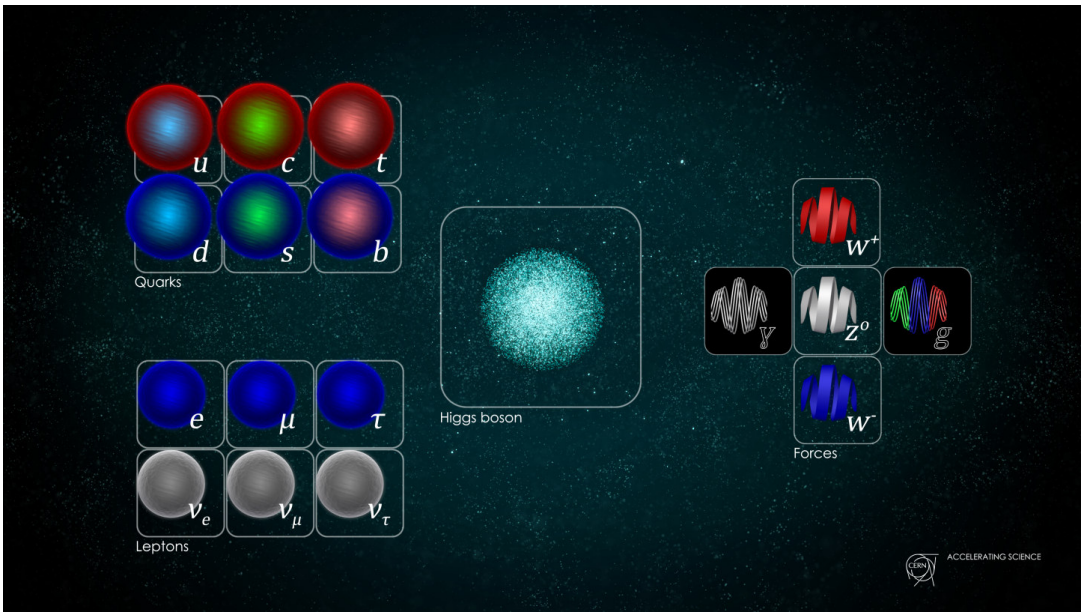


Figure 1.2: *The overview of the Standard Model.*

When seeing this pattern, it is natural to ask if there are further generations. For the time being, there is strong experimental evidence that there are only three generations [5]. One evidence is due to the resonance width of Z boson which gives the limit to the 4th-generation neutrino mass larger than approximately 45 GeV/c^2 but it indeed contradicts

the mass (not higher than 20 MeV/c^2) of all other three generation neutrinos. Hence the Universe is known to be built of these twelve fundamental particles belonging to three generations. It is also worthy to be mentioned here that the neutrinos (ν_e, ν_μ, ν_τ) of the three generations are in fact mixtures of three elementary neutrino states (ν_1, ν_2, ν_3). This distinction is only important for neutrinos travelling for long distances. The neutrino oscillation has been discovered for the first time in 1998 by the Super-Kamiokande experiment in Japan [6]. In addition, the twelve fundamental fermions are described by the Dirac equation of relativistic quantum mechanics, thus for each of them there are exactly twelve **antiparticles** with the same mass but opposite electrical charge.

1.1.2 The fundamental interactions

The section 1.1.1 introduced the elementary particles which compose the matter of the Universe. This section gives a brief description of the interactions among them. In our current knowledge, there are four fundamental interactions in the Universe:

- **Gravity**. Regardless of the distance, gravity exists between any two objects which have mass. Gravity is extremely weak and always attractive thus it is responsible for large-scale structure of the Universe like the case between earth and the Sun. Thus in the field of sub-atomic (about a few femtometres), the gravity can be neglected and it isn't included in the Standard Model.

- **Electromagnetism**. In general, any particle owning an electric charge can interact electromagnetically. For example, electrostatic attraction bounds the electrons and the nucleus together to form an atom which forms the physical and chemical texture of matter. This represents indeed the low-energy manifestation of electromagnetism.

- **Strong interaction**. The strong interaction is only effective over very short range, typically only working in the sub-atomic world. The particles which carry the "color" charge, described by Quantum ChromoDynamics (QCD), equivalent to the electric charge, can feel the strong interaction and experience it. Thus only quarks undergo strong interaction. For example, in the nucleus, the protons and neutrons are bound together by the strong interaction.

- **Weak interaction**. The weak interaction is also only effective over very short range like strong interaction. However, as its name shown, it is "effectively" the weakest interaction among the four interactions after the gravity. The weak interaction rules the physical phenomena like the β-decay of free-neutron. all fundamental particles can feel the weak interaction.

The twelve fundamental fermions are classified by their interactions, as listed in tab. 1.2.

In modern particle physics, each interaction in the Standard Model is described by the Quantum Filed Theory (QFT). According to the QFT, the interaction between particles is mediated by the exchange of force-carrier particles with spin-1, called **gauge boson**.

In the case of electromagnetism, it is described by the Quantum ElectroDynamics (QED), where the interaction between two charged particles is mediated by the massless force-carrier particle, called photon (γ) and represented by the A_μ field. In group theory,

	1st	2nd	3rd	strong	electromagnetic	weak
Quarks	d	s	b	√	√	√
	u	c	t	√	√	√
Leptons	e^-	μ^-	τ^-		√	√
	ν_e	ν_μ	ν_τ			√

Table 1.2: *The interactions experienced by twelve fundamental particles.*

the electromagnetism is represented by U(1) local gauge symmetry and its associated field B_μ. In the case of the weak interaction, all twelve fundamental fermions carry the weak-type charge (weak isospin), thus they all participate to the weak interaction. The weak interaction can be categorised following the force-carrier electrical charge; the weak charged-current interaction and the weak neutral-current interaction. The weak charged-current interaction is mediated by the exchange of charged W^+ and W^- bosons. The $W^\pm$ only couples with different flavour fermions. Because the $W^\pm$ have charge $\pm e$, in order to conserve the electric charge, the weak charged-current interaction only couples pairs of fundamental fermions which have a difference of one unit electric charge. The weak neutral-current interaction is mediated by the exchange of electrically neutral Z boson. In group theory, the weak interaction is associated with SU(2) local gauge symmetry. The physical $W^\pm$ bosons can be determined by the linear combinations:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp W_\mu^2\right) \tag{1.1}$$

where W_μ^k ($k = 1, 2, 3$) are the three gauge fields associated to SU(2) symmetry. The electromagnetic interaction and weak interaction then are unified by the electroweak model of Glashow, Salam and Weinberg (GSW) [7, 8], which can be described by the production of $SU(2) \times U(1)$. The photon and Z boson are both neutral corresponding to the gauge fields A_μ and Z_μ. In the GSW model, the U(1) gauge symmetry of electromagnetism are replaced by the new $U(1)_Y$ gauge symmetry which introduces a new gauge field B_μ coupling with a new type of charge , called hypercharge Y.

$$Y = 2(Q - I_W^{(3)}) \tag{1.2}$$

where the Q is electromagnetic charge and $I_W^{(3)}$ is the third component of the weak isospin. In the unified electroweak model, the photon and Z boson can be written as the linear combinations of the B_μ and neutral $W_\mu^{(3)}$ of the weak interaction,

$$\begin{aligned} A_\mu &= +B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \\ Z_\mu &= -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \end{aligned} \tag{1.3}$$

where the θ_W denotes the weak mixing angle which can be measured in experiments.

In the case of strong interaction, it is described by the QCD, where the force-carrier particle is eight massless gluon corresponding to the eight generators of the SU(3) local gauge symmetry. Like the electric charge of QED, the three conserved color charge r, b and g are used to describe the orthogonal basis states of the SU(3) color space. Then the

color assignment of eight physical gluons can be written as

$$r\bar{g},\ g\bar{r},\ r\bar{b},\ b\bar{r},\ g\bar{b},\ b\bar{g},\ \frac{1}{\sqrt{2}}(r\bar{r}-g\bar{g})\ and\ \frac{1}{\sqrt{6}}(r\bar{r}+g\bar{g}-2b\bar{b}).$$

The relative strength of the interactions, except the gravity, associated with their corresponding gauge bosons are summarized in Tab. 1.3.

Interaction	strength	Boson	Mass/GeV	Charge	Spin
strong	1	Gluon (g)	0	0	1
electromagnetic	10^{-3}	photon (γ)	0	0	1
weak	10^{-8}	$W^{\pm}$	80,379	$\pm 1e$	1
		Z	91,188	0	1

Table 1.3: *The properties of the three interactions associated with their corresponding force-carrier bosons. The W boson and Z boson mass are cited from Ref. [9]. The relative strength are approximately measured between two fundamental particles at distance of $1fm$.*

1.1.3 The Higgs boson

The Higgs boson, unlike all other fundamental fermions and the gauge bosons which are spin-half and spin-1 respectively,is a spin-0 (scalar) particle. The Higgs boson plays a important role in the Standard Model. All the particles in the Standard Model obtain their mass through the Higgs mechanism which will be discussed in detail in the next section 1.2.

The Higgs boson is the final particle observed of the Standard Model, which was discovered by the ATLAS [10] and CMS [11] experiments at Large Hadron Collider (LHC) in 2012. Thus the Nobel Prize in Physics was jointly awarded to François Englert and Peter Higgs "for the theoretical discovery of the mechanism contributing to our understanding of the origin of the mass of subatomic particles and recently confirmed by the discovery, by ATLAS and CMS experiments at CERN LHC, of the fundamental particle predicted by this theory", on October 8, 2013.

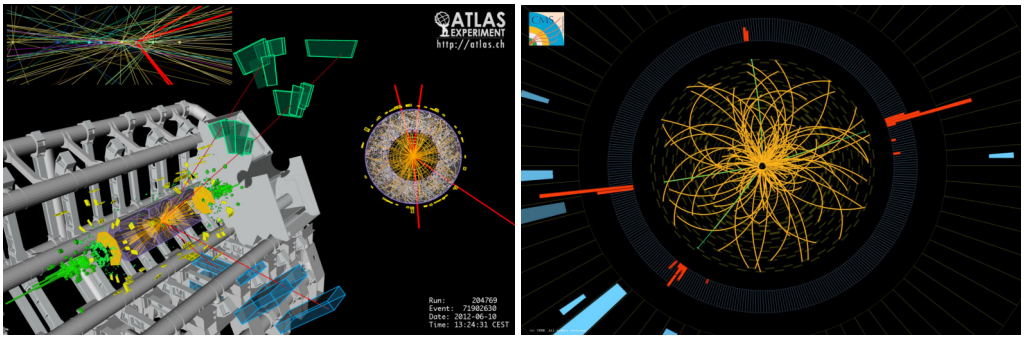


Figure 1.3: *Event display for a Higgs candidate of $H \to ZZ^* \to \mu^+\mu^-\ \mu^+\mu^-$ in the ATLAS (left) [12] and CMS detectors (right) [13].*

Both experiments ATLAS and CMS searched for Higgs boson in several states, like $\gamma\gamma$, $b\bar{b}$, ZZ^*, WW^* and $\tau^-\tau^+$. One of the most sensitive channel is $H \rightarrow ZZ^* \rightarrow 4l$. The event display of Higgs candidate $H \rightarrow ZZ^* \rightarrow 4l$ of ATLAS and CMS detector shown in Fig. 1.3. The invariant mass distribution of $4l$ final state is shown in Fig. 1.4. As shown in Fig. 1.4 (left), the ATLAS experiments observed the excess of events around 125 GeV can be attributed to the Higgs boson and the peak around 91 GeV is apparently from the Z boson production. The CMS also observed a comparable evidence for the Higgs boson at the same mass shown in Fig. 1.4 (right).

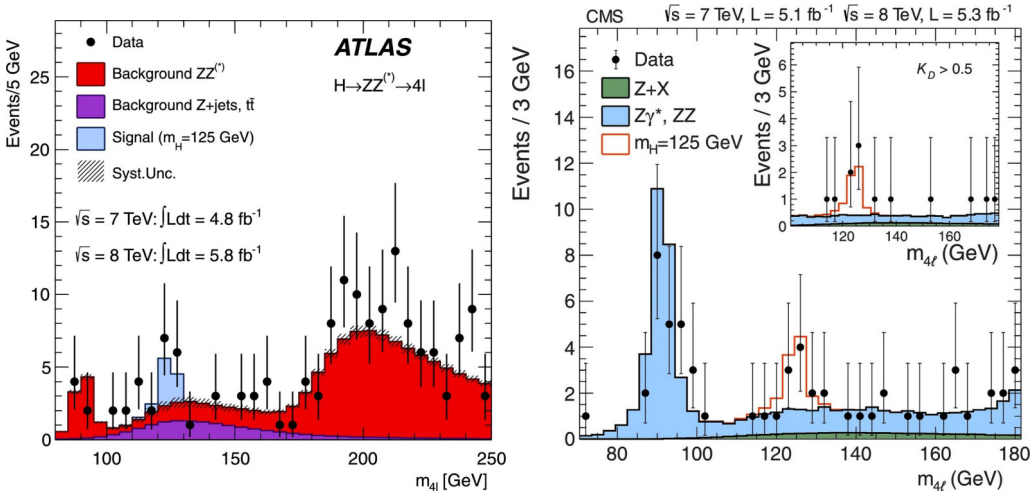


Figure 1.4: *The distribution of the four-lepton invariant mass. The points represent the data and the filled histograms is the background. The signal expectation of a Higgs boson in the Standard Model is also shown. The left plot is from the results of ATLAS detector [10] and the right plot is obtained from CMS detector [11].*

The combined measurements of the ATLAS and CMS experiments give the statistically compelling evidence for the discovery of the Higgs boson. The all measurements of the Higgs boson mass including the individual and combined Run 1 and the Ruan 2 data by ATLAS and CMS for the diphoton and four letpon final states, are summarized in the Fig. 1.5.

The latest experimentally measured Higgs boson mass , giving

$$m_H = 125.10 \pm 0.14 GeV \text{ [9].}$$

The discovery of the Higgs boson is a sucessful milestone of the modern particle physics but is not the end of the story, since further open questions are still open. Is there only one Higgs boson? Is Higgs boson a fundamental scalar particle or is it a composite? Furthermore, because of the huge and cumbersome backgrounds of hadronic colliders like LHC, the precision measurement of all Higgs properties is not satisfying and needs a future Higgs factory, the leptonic collider to explore, that should define the new era of precision Higgs boson measurements.

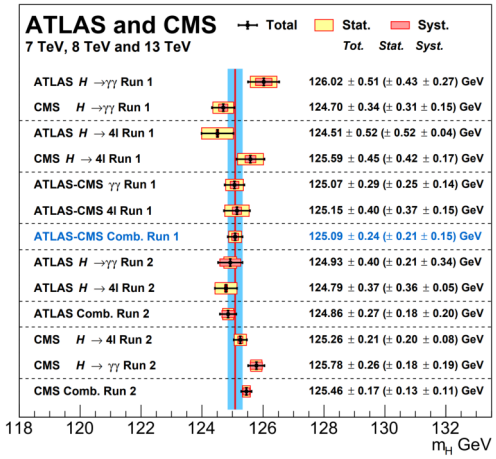


Figure 1.5: *The summary of the Higgs mass measurements by ATLAS and CMS in the* $H \to \gamma\gamma$ *and* $H \to ZZ^*$ *channels using Run 1and Run 2 data [14].*

1.2 The Brout-Englert-Higgs mechanism

The Brout-Englert-Higgs mechanism abbreviated Higgs mechanism is a fundamental part of the Standard Model. It provides the mechanism which generates the masses of the electroweak gauge bosons in an appropriate way that protects the local gauge invariance of the Standard Model. This mechanism also gives the mass to all fundamental fermions. The Higgs mechanism will be described in terms of the Standard Model Lagrangian.

1.2.1 The need for the Higgs mechanism

The interactions in the Standard Model are well described under the local gauge invariant principle. Especially the series of success of precision electroweak experimental measurements place this principle on a solid experimental basis. However, this principle is broken by the terms of particle masses in the Lagrangian. For instance, if we assume the photon is massive, the Lagrangian of QED would contain an extra term $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$,

$$\mathcal{L}_{QED} \to \overline{\psi}(i\gamma^\mu\partial_\mu - m_e)\psi + e\overline{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m_\gamma^2 A_\mu A^\mu.$$

Under the U(1) local gauge transformation: $\psi(x) \to \psi'(x) = e^{iq\chi(x)}\psi(x)$, the photon field transforms as

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu\chi$$

so the new mass term transforms as

$$\frac{1}{2}m^2 A_\mu A^\mu \to \frac{1}{2}m^2\left(A_\mu - \partial_\mu\chi\right)(A^\mu - \partial^\mu\chi) \neq \frac{1}{2}m^2 A_\mu A^\mu,$$

from which it is clear that the photon mass term do not respect the local gauge invariant principle. Thus the required U(1) local gauge symmetry can only be permitted by the gauge bosons which are massless. This restriction also works on the SU(2) and SU(3) gauge symmetry associated with weak and strong interactions. This cannot cause any problems for the QED and QCD where the gauge bosons are massless but it is in

apparent contradiction with the observation of the large masses of W boson and Z boson responsible for the weak interaction.

The problem of particle mass is not limited to the case of gauge bosons. For example, the mass term of electron Lagrangian can be written in terms of chiral states as

$$\begin{aligned} -m_e\overline{\psi}_e\psi_e &= -m_e\overline{\psi}_e[\frac{1}{2}(1-\gamma^5)+\frac{1}{2}(1+\gamma^5)]\psi_e \\ &= -m_e(\overline{\psi}_{eR}\psi_{eL}+\overline{\psi}_{eL}\psi_{eR}), \end{aligned} \tag{1.4}$$

where the subscripts L and R denotes the left-handed and right-handed chiral stats respectively. In the SU(2) gauge transformation of the weak interaction, left-handed particles transform as weak isospin doublets and right-handed particles as singlets, and therefore the mass term of Eq. (1.4) violates the gauge invariant requirement [15].

1.2.2 The spontaneous symmetry breaking

The principle of spontaneous symmetry breaking can be introduced by considering the complex scalar field,

$$\phi = \frac{1}{\sqrt{2}}(\phi_1+\phi_2),$$

for which the corresponding Lagrangian is

$$\mathcal{L} = (\partial_\mu\phi)^*(\partial^\mu\phi) - V(\phi) \quad with \quad V(\phi) = \mu^2(\phi^*\phi)+\lambda(\phi^*\phi)^2. \tag{1.5}$$

When this Lagrangian expressed in terms of two real scalar fields ϕ_1 and ϕ_2,

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}(\partial_\mu\phi_1)(\partial^\mu\phi_1)+\frac{1}{2}(\partial_\mu\phi_2)(\partial^\mu\phi_2) \\ &- \frac{1}{2}\mu^2(\phi_1^2+\phi_2^2) \qquad (1.6) \\ &- \frac{1}{4}\lambda(\phi_1^2+\phi_2^2). \qquad (1.7) \end{aligned}$$

The vacuum state is the lowest energy state of the field ϕ, which corresponds to the minimum of the potential of Eq. (1.5). In order to have a finite minimum, λ should be positive. The Lagrangian of Eq. (1.5) is invariant under U(1) local gauge symmetry. The shape of the potential determined by the sign of μ^2, as shown in Fig. 1.6 . when $\mu^2 > 0$, the minimum of the potential is obtained when ϕ_1 and$\phi_2 = 0$, corresponding the vacuum states of both fields being zero. If $\mu^2 < 0$, the potential has an infinite set of minima defined by,

$$\phi_1^2+\phi_2^2 = \frac{-\mu^2}{\lambda} = \nu^2,$$

as indicated by the red dash circle in Fig. 1.6.

For $\mu^2 < 0$, the lowest energy state does not occur at both ϕ_1 and ϕ_2= 0 and the fields is said to have a non-zero vacuum expectation value ν. The physical vacuum state will locate at a specific point on the dash circle shown in Fig. 1.6. The choice of the vacuum state breaks the symmetry of the Lagrangian, a process known as the **spontaneous symmetry breaking**. Without loss of generality, the vacuum state can be chosen to be in the real

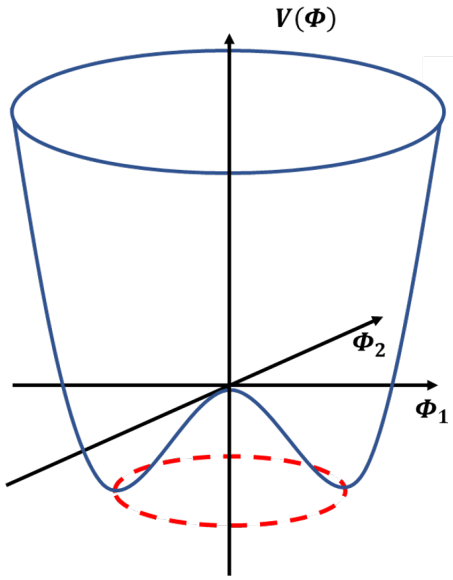


Figure 1.6: *The V(ϕ) potential for a complex scalar field for $\mu^2 > 0$.*

axis, $(\phi_1, \phi_2) = (v, 0)$, and then the complex scalar field ϕ can be expanded by writing $\phi_1(x) = \eta(x) + v$ and $\phi_2(x) = \xi x$,

$$\phi = \frac{1}{\sqrt{2}}(\eta + v + i\xi).$$

The Lagrangian associated with complex scalar field can be written in terms of η and ξ,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - V(\eta, \xi),$$

where the potential term is given by:

$$V(\eta, \xi) = \mu^2\phi^2 + \lambda\phi^4 \quad with \quad \phi^2 = \frac{1}{2}[(v + \eta)^2 + \xi^2].$$

then the $V(\eta, \xi)$ can be expanded in terms of η and ξ,

$$V(\eta, \xi) = -\frac{1}{4}\lambda v^4 + \lambda v^2\eta^2 + \lambda v\eta^3 + \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda\xi^4 + \lambda v\eta\xi^2 + \frac{1}{2}\lambda\eta^2\xi^2,$$

where the term proportional to η^2 can be treated as a mass, and the terms with other either three or four powers of the fields can be treated as interaction terms. Finally, the term $-\frac{1}{4}\lambda v^4$ is just a constant and has no physical consequence. Hence the Lagrangian can be written as

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \frac{1}{2}m_\eta^2\eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - V_{int}(\eta, \xi) \quad with \quad m_\eta = \sqrt{2\lambda v^2}, \tag{1.8}$$

and interactions given by

$$V_{int}(\eta, \xi) = \lambda v\eta^3 + \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda\xi^4 + \lambda v\eta\xi^2 + \frac{1}{2}\lambda\eta^2\xi^2.$$

The Lagrangian of Eq. (1.8) represents a scalar field η with mass $m_\eta = \sqrt{2\lambda v^2}$ and a massless scalar field ξ.

1.2.3 The Higgs mechanism and particle masses

In the Higgs mechanism , the spontaneous symmetry breaking of a two complex scalar field is embedded in the $U(1)_Y \times SU(2)_L$ local gauge symmetry of the electroweak part of the Standard Model. One of the scalar fields must be neutral, written as ϕ^0, and the other must be electrically charged such that ϕ^+ and $(\phi^+)^* = \phi^-$ correspond to the W^+ and W^-. Hence the minimal Higgs model can be written in a weak isospin doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}$$

The Lagrangian for this doublet of complex scalar field is

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \quad with \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi), \tag{1.9}$$

where the Higgs potential depends on two parameters μ and λ. After symmetry breaking of Higgs potential field, for $\mu^2 < 0$, the Higgs potential has an infinite set of finite minima satisfying

$$\phi^\dagger \phi = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda}.$$

then the fields can be expanded about this minimum by writting under the unitary gauge,

$$\phi_0(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.10}$$

Because of the existence of derivatives in Eq. (1.9) , the Lagrangian does not obey the $SU(2)_L \times U(1)_Y$ local gauge symmetry. In order to achieve this symmetry, the derivatives can be replaced by the appropriate covariant derivatives,

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + i g_W \vec{T} \cdot \vec{W}_\mu + i g' \frac{Y}{2} B_\mu,$$

where $\vec{T} = \frac{1}{2}\vec{\sigma}$ correspond to the three generators of the $SU(2)$ symmetry and $W(x)$ are the three gauge fields. Y is the weak hypercharge of GSW model with $Y = 2(Q - I_w^{(3)})$. The upper component of Higgs doublet is positively charged with $Q = 1$ and $I_w^{(3)} = \frac{1}{2}$, thus here the hypercharge of Higgs doublet is $Y = 1$. Hence the Lagrangian of Eq. (1.9) can be replaced by writing in the covariant derivatives,

$$\mathcal{L} = \left(D_\mu \phi\right)^\dagger \left(D^\mu \phi\right) - V(\phi).$$

The term $\left(D_\mu \phi\right)^\dagger \left(D^\mu \phi\right)$ of the Lagrangian of Eq. (1.2.3) is responsible for generating the mass of gauge bosons, which can be expanded by using Eq. (1.10),

$$\begin{aligned} (D_\mu \phi)^\dagger (D^\mu \phi) &= \frac{1}{2}(\partial_\mu h)(\partial^\mu h) \\ &+ \frac{1}{8} g_W^2 (W_\mu^{(1)} + i W_\mu^{(2)})(W^{(1)\mu} - i W^{(2)\mu})(v + h)^2 \\ &+ \frac{1}{8}(g_W W_\mu^{(3)} - g' B_\mu)(g_W W^{(3)\mu} - g' B^\mu)(v + h)^2. \end{aligned} \tag{1.11}$$

The masses of gauge bosons are determined by the terms of Eq. (1.11) which are quadratic in the gauge boson fields,

$$\frac{1}{8}g_W^2(W_\mu^{(1)}W^{(1)\mu} + W_\mu^{(2)}W^{(2)\mu}) + \frac{1}{8}v^2(g_W W_\mu^{(3)} - g'B_\mu)(g_W W^{(3)\mu} - g'B^\mu).$$

then the mass terms for the $W^{(1)}$ and $W^{(2)}$ can be written

$$\frac{1}{2}m_W^2 W_\mu^{(1)}W^{(1)\mu} \quad and \quad \frac{1}{2}m_W^2 W_\mu^{(2)}W^{(2)\mu}.$$

Thus the mass of W boson is

$$m_W = \frac{1}{2}g_W v. \tag{1.12}$$

Hence, the mass of the W boson is determined by the coulpling constant $SU(2)_L$ gauge interaction g_W and the vacuum expectation value of the Higgs field.

The terms is quadratic in the neutral $W^{(3)}$ and B fields in the Eq. (1.11) can be transformed to,

$$\frac{1}{8}v^2(g_W W_\mu^{(3)} - g'B_\mu)(g_W W^{(3)\mu} - g'B^\mu) = \frac{1}{8}v^2\begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} M \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix} \tag{1.13}$$

where the matrix M can be written as,

$$M = \begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix} \tag{1.14}$$

The elements of M couple with the $W^{(3)}$ and B fields, allowing them to mix. The mass of the physical boson are the eigenvalues of matrix M, which can be obtained from the euqation $(M - \lambda I) = 0$, giving

$$\lambda = 0 \quad or \quad \lambda = g_W^2 + g'^2.$$

Thus in the diagonal basis, the matrix of the right hand of Eq. (1.14) can be written,

$$\frac{1}{8}v^2\begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix}\begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix}\begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix} = \frac{1}{8}v^2\begin{pmatrix} A_\mu & Z_\mu \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & g_W^2 + g'^2 \end{pmatrix}\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} \tag{1.15}$$

$$= \frac{1}{2}\begin{pmatrix} A_\mu & Z_\mu \end{pmatrix}\begin{pmatrix} m_A^2 & 0 \\ 0 & m_Z^2 \end{pmatrix}\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} \tag{1.16}$$

where the A_μ and Z_μ correspond to the physical boson fields, giving

$$A_\mu = \frac{g'W_\mu^{(3)} + g_W B_\mu}{\sqrt{g_W^2 + g'^2}}$$

$$Z_\mu = \frac{g_W W_\mu^{(3)} - g'B_\mu}{\sqrt{g_W^2 + g'^2}}$$

By writing the ratio $\frac{g'}{g_W} = \tan\theta_W$, the relation of the Eq. (1.3) can be found. The m_A and m_Z are the corresponding masses of gauge bosons, giving

$$m_A = 0 \quad and \quad m_Z = \frac{1}{2}v\sqrt{g_w^2 + g'^2}.$$

The massless neutral gauge boson A is identified as photon and the massive neutral gauge boson can be identified as the Z boson. In the GSW model, there are four free parameters, the coupling constants g_W and g', and two parameters of Higgs potential μ and λ associated with the vacuum expectation value of Higgs field v and the mass of the Higgs boson m_H by

$$v^2 = \frac{-\mu^2}{\lambda} \quad and \quad m_H = \sqrt{2\lambda}v.$$

In the experiments, the g_W and m_W can be measured and the vacuum expectation value v is determined by the Eq. (1.12), giving

$$v = 246\ GeV.$$

The another parameter λ then can be obtained from the m_H measured at the LHC discussed in Section 1.1.3.

Finally, a simple discussion is given about the Higgs mechanism generating the fermions mass. In the Standard model, the fermions identified as left-handed chiral state are placed in SU(2) doublets written as L and the right-handed fermions are placed in SU(2) singlets wirtten as R. Then for all dirac fermions, the mass terms of Lagrangian can be constructed from either

$$\mathcal{L}_{mass} = -g_f[\bar{L}\phi R + (\bar{L}\phi R)^\dagger] \tag{1.17}$$

or

$$\mathcal{L}_{mass} = g_f[\bar{L}\phi_c R + (\bar{L}\phi_c R)^\dagger] \tag{1.18}$$

where the ϕ_c is the conjugate doublet of ϕ, $\phi_c = -i\sigma_2\phi^*$ and g_f is the Yukawa couplings of the fermions with the Higgs field, given by

$$g_f = \sqrt{2}\frac{m_f}{v}.$$

Without loss of generality, here we only consider the Eq. (1.17) and the Eq. (1.18) can be directly considered by taking the action of conjuagating based on the Eq. (1.17). After the spontaneous symmetry breaking, the Higgs doublet under the unitary gauge is the same form of Eq. (1.10), Thus the Eq. (1.17) becomes

$$\mathcal{L}_{mass} = -\frac{g_f}{\sqrt{2}}v(\bar{L}R + \bar{R}L) - \frac{g_f}{\sqrt{2}}h(\bar{L}R + \bar{R}L). \tag{1.19}$$

The first term in Eq. (1.19) is exactly the form required for the fermion masses. The second term represents the coupling between the fermion and the Higgs boson.

1.3 The physics beyond the Standard Model

All the particles predicted by Standard model are completed by the discovery of Higgs boson. Undoubtedly, the Standard Model provides the successful description of the

modern particles physics on solid experimental basis, thus it is one of the triumphs of modern physics. However, it does not mean the skyscraper of particle physics has been built up and there are some striking dark cloud in the sky of science. In this section, we will briefly discuss the physics beyond the Standard model (BSM). There are plenty of questions that can not be perfectly answered by the standard Model which require to investigate BSM physics.

- **(1) Neutrino mass**. In the standard Model, the right-handed chiral states ν_R does not couple with gluons or electrweak gauge bosons. Thus there is not directly experimental evidence to indicate the existence of them. However, neutrino oscillations experiments,shows that neutrinos (at least one of them) must have mass and thus the Lagrangian needs to include a corresponding mass term. Thus the right-handed chiral neutrinos must exist suggested by the Eq. (1.18). Thus the contradiction may imply that there are another mechnism to explain neutrino mass, for which one of interested mechanism is the seesaw mechanism.

- **(2) The relatively large number of free parameters of the Standard Model**. If neutrinos are Dirac fermions, the Standard Model has 25 (putting aside θ_{CP}) parameters which have to be input artificially. This may be a symptoms that a model in which the parameters are chosen to agree with observations, rather than coming from higher theoretical principles.

- **(3) The observed abundance of asymmetry between matter and anti-matter**. Based on the observation in the lab, the mass of the Universe is dominated by the matter. However, according to the theory of cosmology, the energy of Big Bang should have transformed into matter and anti-matter evenly. Thus it gives rise to the question about why anti-matter can not be observed like in the lab. The CP-violation mechanism of the standard model can not completely explain the existence of the huge asymmetry.

- **(4) The presence of dark matter**. Suggested by the experimental observation of astronomy, the majority of the Universe is composed of dark matter and dark energy, only the remaining about 4% is made of the particles of the Standard Model. The dark matter is neutral, colorless, non-baryonic and massive. The only possible particles in the Standard Model are neutrinos but they are too light. Thus a new type of weakly interacting massive particle (WIMP) is suggested and the direct detection of WIMPs is one of main goals in particle physics experiments at this time, such as the PandaX in China [16], the XENON in Italy [17] and LUX-ZEPLIN in the United States [18].

- **(5) The number of generation of fermions.** Why the fundamental fermions has three generations?

- **(6) The quantization of the electric charge.** The Standard Model contains the U(1) gauge symmetry, for which any charge is permitted. Why we find integer multiples of $\frac{1}{3}$ for electric charge, rather than $\sqrt{2}$ or others.

- **(7) The hierarchy of fermions' masses.** There is hierarchical structure in the three generations, but in the Standard Model, they are just free parameters.

- **(8) Supersymmetry (SUSY).** In SUSY model, each of the Standard Model particle has its corresponding super-partner "sparticle". The SUSY model provide natural solution to the Hierarchy problem [19] of Higgs boson in high energy scale and it also provide the possible solution to the dark matter. Thus the search for the production of SUSY particles is one of attractive focuses at experimental particles physics.

1.4 Summary

This chapter gives a brief introduction to the Standard Model of particle physics which includes the elementary particles and the basic interactions. The Standard Model particles is completed by the discovery of Higgs boson which is predicted by the Higgs mechanism. Thus the short discussion about the Higgs mechanism and the discovery of Higgs boson are also presented. Finally, although the Standard Model achieves remarkable success, there are also some open questions related to the physics beyond Standard Model.

CHAPTER 2

The future leptonic collider projects

The discovery of the Higgs boson by the ATLAS and CMS experiments in 2012 represents a new victory of the Standard Model. It also opens a new era for particle physics. Thus the high precision measurements of the properties of the Higgs boson is one of the main goals for physicists in the following decades. Because of the leptonic colliders providing cleaner environments than hadron-hadron colliders, they allow to precisely measure the Higgs properties and thus the high energy physics community is investigating several choices of leptonic collider, such as Circular Electron Positron Collider (CEPC), Future Circular Collider (FCC), International Linear Collider (ILC) and Compact Linear Collider (CLIC), which will be discussed in this chapter.

Content

2.1 The advantages of the leptonic collider

In the case of hadron-hadron colliders (pp or $p\bar{p}$), the beam particles are not the fundamental point-like particles. Thus it generates the interactions between the many constituents of the protons or anti-protons (gluons, quarks and antiquarks). Hence in the collision , the initial states are not defined and most of interactions occur at relatively low energy with the huge production of background and uninteresting events. This is the main disadvantage of hadronic colliders when compared with leptonic colliders. In contrast, lepton-antilepton collider can generate interactions between the fundamental point-like particles. The reactions generated are relatively simple to understand, the full energies of beams are available for physics production and there is very clean background of low energy events compared with hadronic colliders. Furthermore, if the center of mass energy is set equal to the mass of a suitable state of interest, then it can generate a large cross section through s-channel (the physical target-particle factory), in which a single state is generated by the interaction. In this case, the mass and quantum numbers of the state of our interest are constrained by the conditions of initial beams. If the energy spread of the beams is sufficiently narrow, then precision determination of masses and widths are possible. Thus the leptonic machines is very suitable for high precision measurements for Higgs properties.

2.2 The Circular Electron Positron Collider (CEPC) project

The discovery of the Higgs boson by the ATLAS and CMS experiments in 2012 constituted the beginning of a new era of modern particle physics. The Circular Electron Positron Collider(CEPC) [20, 21] is a large international scientific facility proposed by the Chinese particle physics community in 2012 to explore the physics program with abundance of content and then joined by international community.

The basic of the CEPC is that it is a double-ring collider, in a circular underground tunnel where the electron and positron beams circulating in opposite directions in isolated beam pipes. It has two interaction points (IPs) where two large detectors are to be located. The ring is approximately 100 km and it is designed to operate at around 91.2 GeV as a Z factory, at around 160 GeV of the WW production threshold, and at 240 GeV as a Higgs factory. The CEPC will produce close to one trillion Z bosons, 100 million W bosons and over one million Higgs bosons. The large amount of bottom quarks, charm quarks and τ-leptons subsequently produced in the decays of Z bosons also makes the CEPC, in addition of being a Higgs factory, to be an effective B-factory and τ-charm one. Therefore, the CEPC offers a excellent opportunity for high precision measurements and it also can serve for searching BSM physics.

The current tentative physics operation mode are summarized in the Tab. 2.1

2.2.1 The physics motivation for CEPC

The high precision measurement of Higgs boson properties will be a critical part of any ambitious collider project designed for high energy physics in the coming decades. The CEPC will precisely measure the Higgs boson properties in a model-independent way

Operation mode	$\sqrt{s}$ (GeV)	L per IP (10^{34}cm^{-2}s^{-1})	Running years	Total $\int L$ (ab^{-1}, 2 IPs)	Events yields
Higgs factory	240	3	7	5.6	1×10^6
Z factory	91.2	32 ($\star$)	2	16	7×10^{11}
W^+W^-	158 - 172	10	1	2.6	2×10^7($\diamond$)

Table 2.1: *The tentative operation plan of CEPC at different center-of-mass energy. ($\star$) The value presented here assumes a 2 Tesla solenoid used. ($\diamond$) The extra 9.4 $\times$ 10^7 W^+W^- events will be generated during the Higgs factory operation.*

compared with the (HL)-LHC and certainly reach a new better level of precision for the measurements of the W and Z bosons properties [20]. The cleaner collision environment of the CEPC will allow the search of potential and unknown decay channels that are extremely difficult to be found at the (HL-)LHC. Based on these precision measurements, the CEPC will uncover deviations from the SM predictions and reveal the existence of new particles that are beyond the capacity of current experiments. The precision measurements could also do unveil the hidden mechanism behind the electroweak phase transition. It also provides a key to explain the origin of the asymmetry between the matter and anti-matter inside the Universe.

Furthermore, the CEPC is also a sensitive and powerful machine serving for the search for new physics (e.g., dark matter, exotic Higgs decays), QCD precision measurement (e.g., precision α_s determination) and flavor physcis (e.g., Rare B decays) running at different modes.

2.2.1.1 The precision measurements for Higgs and electroweak

The CEPC will primarily run at a center-of-mass energy of 240 GeV as a Higgs factory. The main channel of Higgs boson production is through $e^+e^- \rightarrow ZH$. In this mode, the CEPC can achieve the integrated luminosity of 5.6 ab^{-1} corresponding to Higgs boson production of over one million. Based on the huge number of signal events and cleaner experimental environment than hadron-hadron colliders, the excellent precision measurements of Higgs properties can be performed by the CEPC. The reachable precision of measurements for Higgs couplings are shown in Fig. 2.1 in terms of the κ framework. The CEPC obtains the significant improvements of the precision for a set of κ parameters which limited by the huge backgrounds of LHC, such as κ_b, κ_c and κ_g [20]. In particular, the measurements of κ_Z improves with a factor of ten. In addition, considering this 7-parameter κ set, the only Higgs coupling that the HL-LHC can provide a comparable measurement is κ_γ, for which the CEPC sensitivity is restricted by statistics. Thus it is also the most important input provided by the HL-LHC for Higgs boson coupling measurements for CEPC, which underscores the significance of combining the results of these two colliders.

Operating in Z factory mode and near the W^+W^- threshold scan, the CEPC will allow precision measurements for electroweak model in complement to the Higgs precision measurements. The projected precision for the electroweak observables is shown in Fig. 2.2. It is clear that CEPC can improve the precision by one order of magnitude [20].

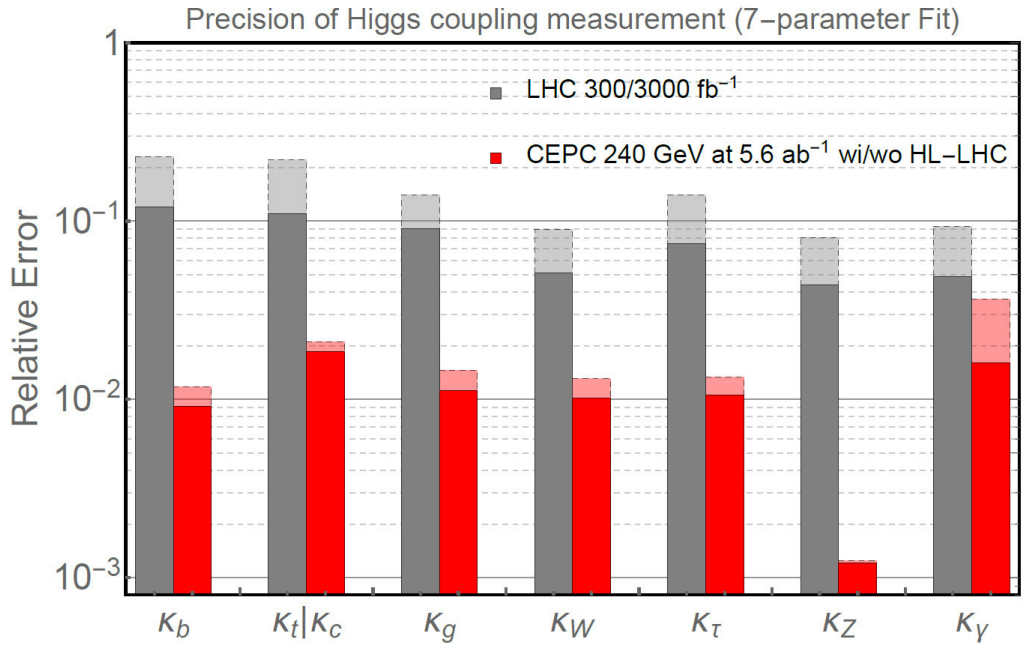


Figure 2.1: *Higgs coupling extraction in terms of the κ-framework. The CEPC results without and with combination with the HL-LHC input are shown as light red bars and red bars respectively. The LHC projections for an integrated luminosity of 3000 fb^{-1} are shown in gray bars and that of 300 fb^{-1} in light grey [20].*

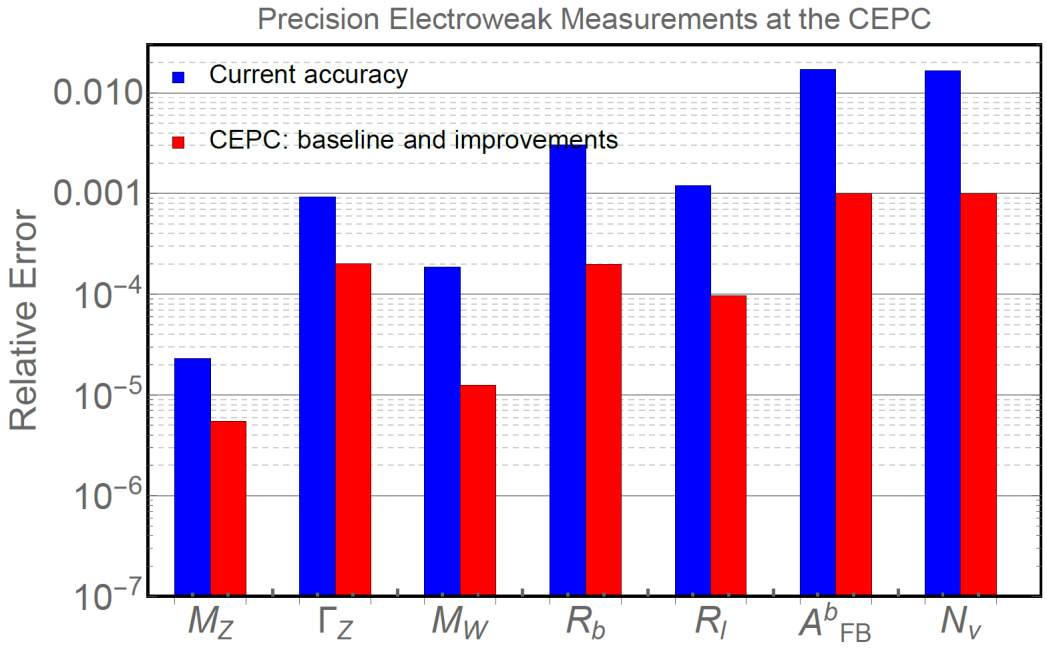


Figure 2.2: *The projection for the precision of the electroweak measurements of CEPC in Z-factory mode [20].*

More details about the precision measurements can be found in Ref. [20].

2.2.1.2 Exploring the new physics

The CEPC is capable of searching for different scenarios of new physics. For example, running as both a Higgs factory and a Z-factory, the exotic decays of the Higgs boson and the Z boson can be used to find new physics. CEPC can also contribute to the the detection of dark matter through both direct production and indirect mechanism. The different new physics cases will be discussed in this section.

- **Exotic Higgs boson decays** : Higgs boson exotic decays could be the manifestation of the new physics. The two-body Higgs boson decays into BSM particles, giving

$$H \rightarrow X_1 X_2$$

where the X_1 and X_2 are BSM particles and they can be allowed to subsequently decay to others. The channels of exotic Higgs boson decays are shown in Fig. 2.3. Compared with the HL-LHC, the improved coverage of exotic Higgs boson decay branching fractions is significant, varying from one to four orders of magnitude for the considered channels. The detailed analysis can be found in Ref. [22, 23, 24].

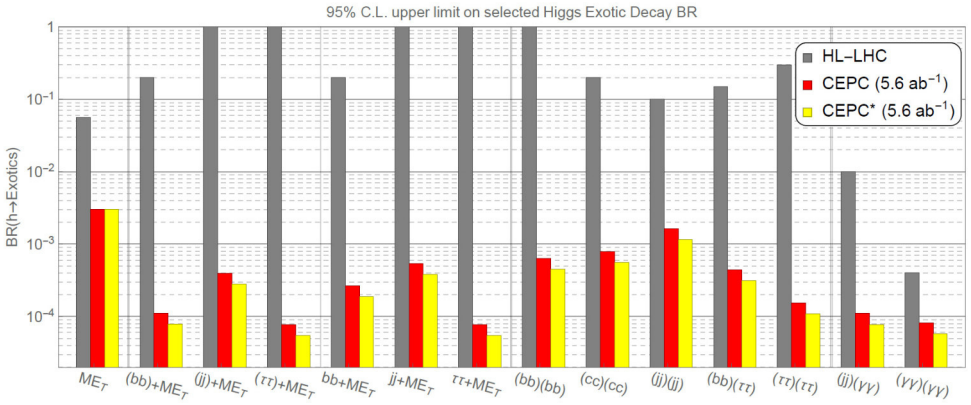


Figure 2.3: *The 95%CL upper limit on selected Higgs exotic decay branching fractions at HL-LHC and CEPC, based on Ref. [24]. The red bars is from the results using only leptonic decays of the spectator Z-boson. The yellow bars include extrapolation with the inclusion of the hadronic decays of the spectator Z-boson. This figure cited from Ref. [20].*

- **Exotic Z boson decays** : Operating at around 91.2 GeV as a Z factory, the CEPC owns the unique capability to search new physics existing in the exotic Z boson decays. The sensitivity of exotic Z boson decays achieved in CEPC is summarized in the Fig. 2.4.

- **Dark matter** :The one highlight of the CEPC is electroweak physics, through high precision measurements of properties of electroweak gauge bosons and Higgs bosons. Thus the most suitable place for the searches for dark matter particles at CEPC is in electroweak multiplets (e.g., doublets or triplets of $SU(2)_L$) or mixtures of electroweak multiplets (including admixtures of a singlet). The detail studies about this topic can be found in Ref. [26, 27, 28].

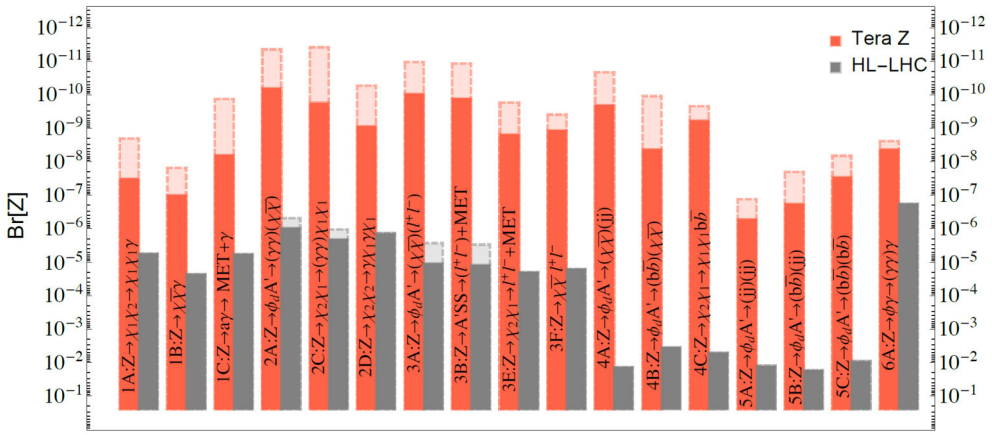


Figure 2.4: *The sensitivity of Z branching ratio for different exotic Z decay channels at CEPC (10^{12} Z), and the HL-LHC at 13 TeV with L = $3ab^{-1}$, based on Ref. [25]. This figure cited from Ref. [20].*

2.2.1.3 QCD precision measurements

The CEPC is an excellent machine to perform precise QCD measurements to improve our understanding of the strong interaction. The abundant topics include the measurement of the strong coupling constant α_s, jet rate physics and QCD event shapes and their utility in probing Yukawa-couplings of light quarks.

2.2.1.4 Flavor physics with CEPC as a Z factory

The CEPC is a high luminosity Z factory which will produce about 10^{12} Z bosons to provide unique opportunities for various flavor physics. Moreover, the decay of huge amount of Z bosons will generate approximately 10^{11} b-type hadrons, which is almost two orders of magnitude larger than the number of B mesons produced by the B factories like BaBar and Belle and it is comparable to the number of B mesons expected at Belle II. Whilst the Z factory of CEPC can be expected to produce 3×10^{10} $\tau^+\tau^-$ pairs from the Z decays. It is also comparable to the expected number of events for τ-lepton production at Belle II (roughly 5×10^{10} [29]). Hence, the CEPC can also be a powerful machine as B-factory and τ-charm factory. The possible interesting topics include: rare B decays, lepton-flavor violating τ decays and flavor violating Z decays.

2.2.2 The collider design of CEPC

2.2.2.1 The accelerator of the CEPC

CEPC is a double-ring collider with two interaction points (IPs) where the two advanced detectors are located. The detector concept will be discussed in detail in Section. 2.2.2.2. The CEPC accelerator complex consists of the Booster, the Collider, the linear accelerator (Linac), the damping ring (DR) and several transport lines as shown in Fig. 2.5. The Collider and the Booster are placed in an underground tunnel with a depth of approximately

100 m, whereas the Linac (the length is 1.2 km) and DR are built at ground level. As the Collider can operate simultaneously as a powerful synchrotron radiation (SR) light source, two gamma-ray beamlines are included in the design. In the tunnel, the additional space is reserved for a future upgrade to the pp collider, named as the Super Proton Proton Collider (SPPC) [30].

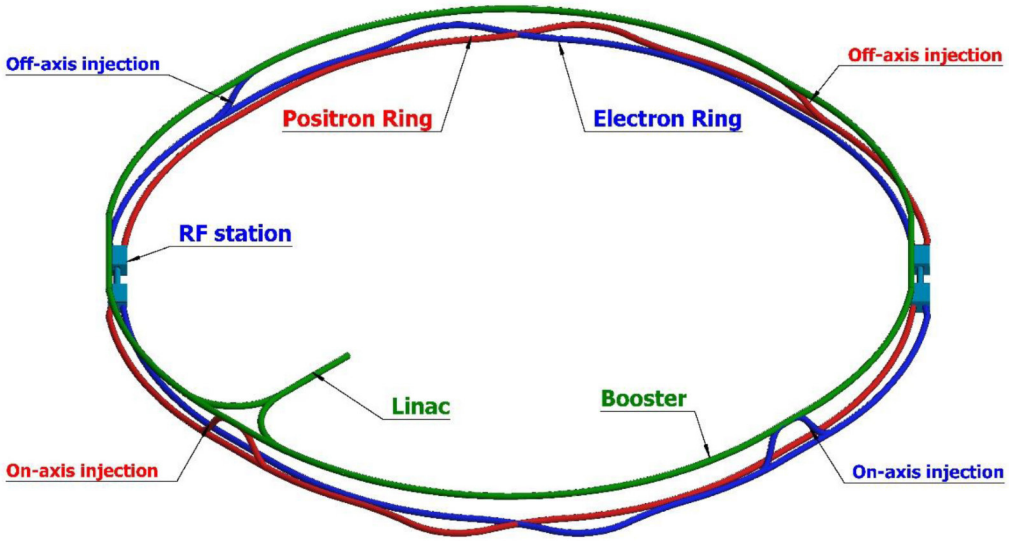


Figure 2.5: *The layout of CEPC [21].*

The electron and positron beams are accelerated in a 10 GeV Linac and then injected and accelerated in the booster to the specific energy of the three collider operating modes (Higgs, W and Z), as shown in Fig. 2.6.

There are 8 straight sections in the Collider: 2 interaction regions used for physics, 2 RF regions and 4 injection regions. Among them, the two on-axis injection regions are used only during operation in the Higgs mode, whereas two off-axis injection regions are to be used for operation in the Higgs, W and Z modes. The superconducting RF cavities are applied in the CEPC, which will be used to accelerate the electron and positron beams, to compensate for synchrotron radiation loss. It is also can provide sufficient RF voltage for energy acceptance and the required bunch length in the Booster and Collider. There are 240 cavities with 650 MHz RF system in the Collider. This cavity is made of bulk niobium and can achieve the gradient acceleration with 22 MV/m operating at 2 K. Each two cavities will share a klystron with 800 kW maximum output power and the beam power in two cavities are expected less than 70% of the maximum output power for any operation mode. The corresponding cavity module is shown in Fig. 2.7.

Some important parameters of CEPC accelerator are listed in the Tab. 2.2

2.2.2.2 The detector concepts of CEPC

The CEPC detector concepts are based on the strict performance requirements to perform a precision physics program that tests the Standard Model and searches for new physics of BSM over a wide range of center-of-mass energies at high beam luminosities. These requirements include large and precisely defined solid angle coverage, efficient vertex

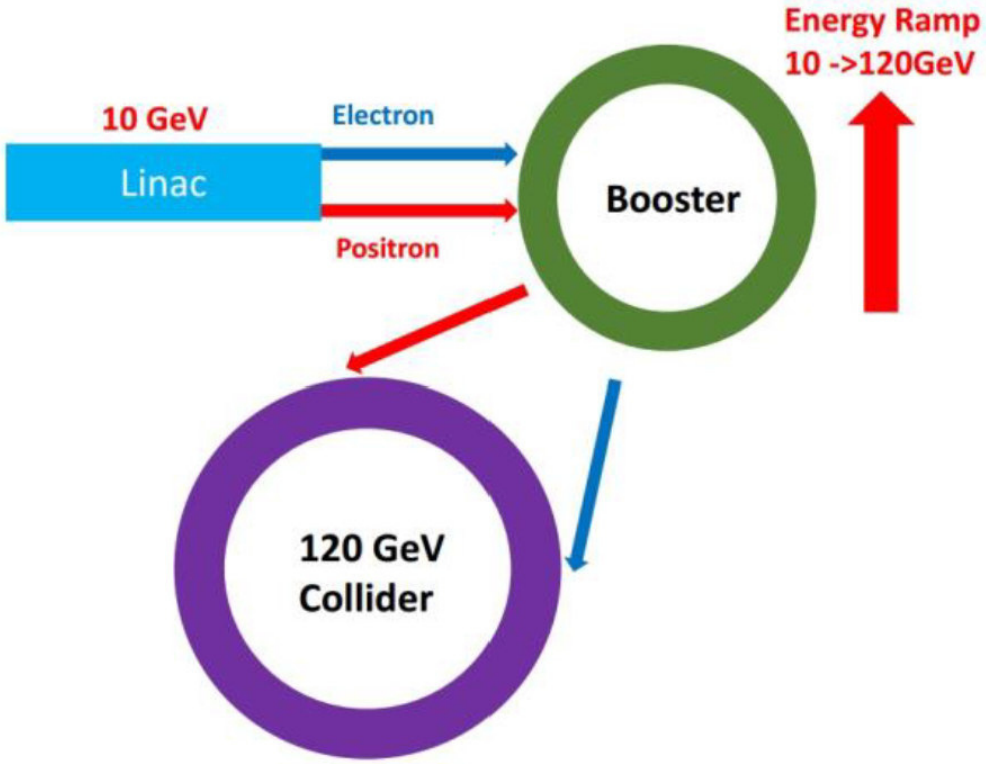


Figure 2.6: *The layout of the CEPC injection chain [21].*

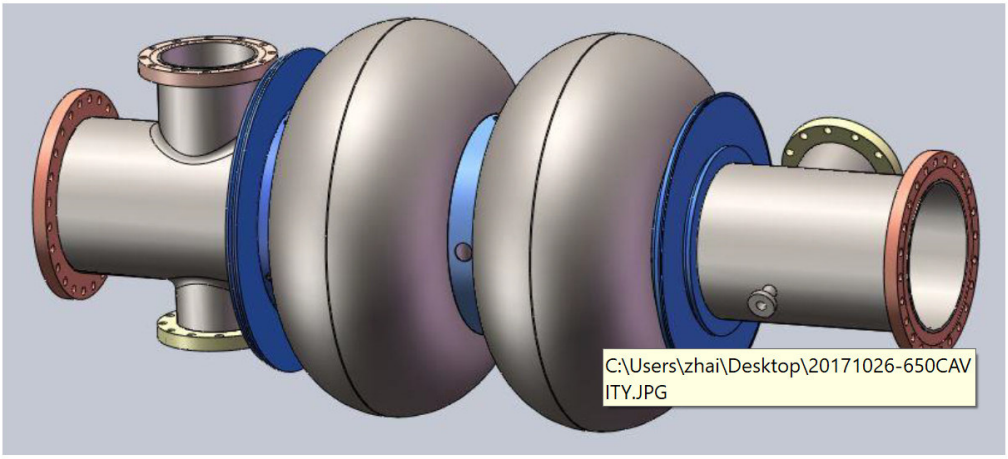


Figure 2.7: *The 3-D model of 650 MHz 2-cell cavity [21].*

reconstruction, excellent particle identification, precise particle energy/momentum measurements, excellent jet reconstruction and flavor tagging. Thus the demands of physics program drive the study of detector concepts.

Two primary detector concepts were studied.

- **Baseline detector concept** : The baseline detector concept is based on the Particle Flow Algorithm (PFA) [31], for which a precision vertex detector, a Time Projection Chamber (TPC), a silicon tracker, a 3 Tesla solenoid, a high granularity calorimeter and a muon detector are assembled together. The detail will be discussed in

Operation mode	Higgs factory	W^+W^-	Z factory (2T)
Number of IPs	2	2	2
Beam energy (GeV)	120	80	45.5
SR loss/turn (GeV)	1.73	0.34	0.036
Number of bunches	242	1524	12000
Bunch spacing (ns)	680	210	25
Beam size at IP σ_x/σ_y (μm)	20.9/0.06	13.9/0.049	6.0/0.04
RF voltage (GV)	2.17	0.47	0.10
Energy spread (%)	0.134	0.098	0.080

Table 2.2: *The CEPC parameters. The full parameters can be found in Ref. [21]*

Section 2.2.3.

- **IDEA concept** : Innovative Detector for Electron-positron Accelerator (IDEA) is an alternative detector concept, which is based on dual readout calorimetry with a precision vertex detector, a drift chamber tracker, a 2 Tesla solenoid, and a muon detector. The detail will be discussed in Section 2.18.

2.2.3 The baseline detector concept

Based on the experience of ILD concept [32], the baseline detector concept was developed and optimized for the CEPC collision environment. This baseline concept is guided by the PFA which can measure final state particles in the suitable sub-detector system. Thus it employs a low material tracking system to minimize the interaction of the final state particles in the tracking material, an ultra high granular calorimetry system to efficiently separate the final state particle showers, and a large volume 3 Tesla solenoid that encloses the entire calorimetry system. In addition, two options for its tracking system are being considered. The default option is a combination of a silicon tracker and a TPC. The other one is that the combination is replaced by a full silicon tracker. The overview of baseline is shown in Fig. 2.8.

From innermost to outermost, the baseline detector consists of a silicon pixel vertex detector, a silicon inner tracker, a TPC (Time Projection Chamber), a silicon external tracker, a silicon-tungsten based sampling Electromagnetic CALorimeter (ECAL), a steel-Glass Resistive Plate Chambers (GRPC) based sampling Hadronic CALorimeter (HCAL), a solenoid of 3 Tesla and a return yoke with an embedded muon detector.

2.2.3.1 The tracking system

In order to achieve the precision measurements for physics program of the CEPC, the excellent and robust tracking system is needed for CEPC baseline detector concept. The high performance tracking system consists of two major part, the vertex detector and tracker including the silicon tracker and TPC.

The vertex detector is composed of six concentric cylindrical pixel layers at radius between 1.6 and 6.0 cm, providing resolution of ≈ 5 μm. The material budget per layer is about 0.15% X_0. Outside the vertex detector is the silicon inner tracker consisting of

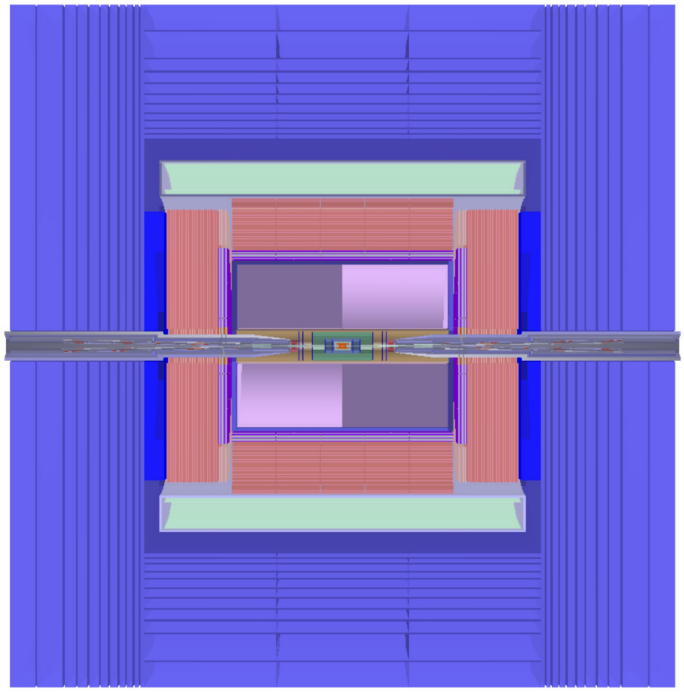


Figure 2.8: *The overview of the baseline detector of the CEPC [20].*

two micro-strip layers at radius of 15.3 cm and 30 cm before the TPC. The silicon external tracker consists of one micro-strip layer at radius of 181 cm after the TPC. The TPC are placed in between the silicon inner and external tracker, which has an inner radius of 0.3 m, an outer radius of 1.8 m, and a length of 4.7 m. The layout of tracking system are is shown in Fig. 2.9.

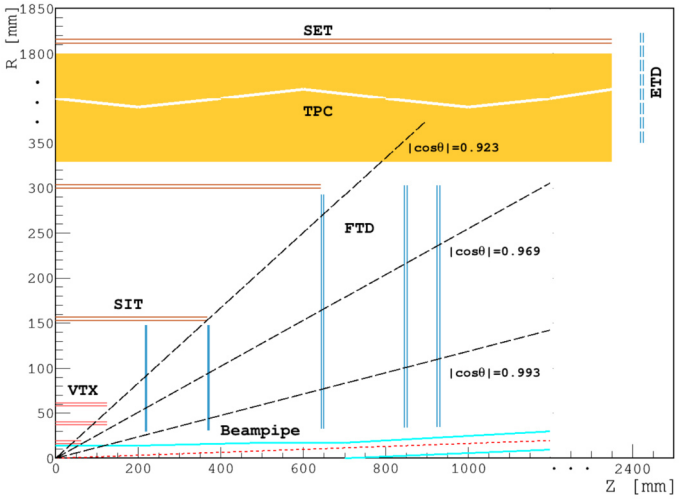


Figure 2.9: *The layout of tracking system of CEPC [20]. VTX is Vertex Detector. SIT and SET are Silicon Inner Tracker and Silicon External Tracker respectively. FTD and ETD are Forward Tracking Detector and Endcap Tracking Detector respectively.*

The TPC can measure charged particle single-hit with a spatial resolution of 100 μm in the $r-\phi$plane and 500 μm in the z direction. The CEPC TPC represents a material budget

of less than 1%X_0. The sketch of TPC is shown in Fig. 2.10. Combining the abilities of

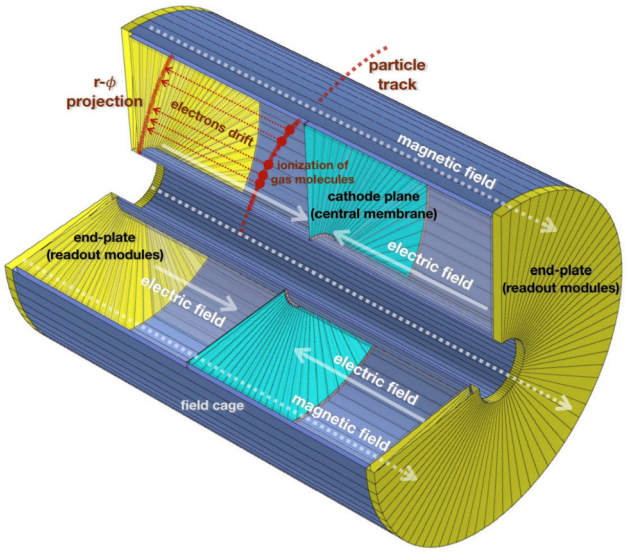


Figure 2.10: *The sketch of CEPC TPC [20].*

silicon tracking system and the TPC, the track momentum resolution can be achieved with a $\triangle(1/p_T) \approx 2{\times}10^{-5}$ GeV^{-1}. As a comparsion, the corresponding resolution of TPC alone is 10^{-4} GeV^{-1}. Here only the default tracking system are presented, the more detail can be found in Ref. [20].

2.2.3.2 Calorimetry

The CEPC baseline detector concept employs PFA-oriented electromagnetic calorimeter (ECAL) and hadronic calorimeter (HCAL) both with high granularity which provide 3-dimensional spatial information and high-resolution energy measurements for electrons, photons, taus and hadronic jets. For PFA-oriented calorimeters, extensive studies has been carried out by the CALICE collaboration and in world-wide detector R&D efforts for the ILC. A variety of detector technology options have been investigated to reach the strict performance requirements of physics case. A series of high granularity prototypes based on several technological options have been developed and exposed to particle beams providing deep understanding of the PFA-oriented calorimetry performance.

The PFA-oriented high granularity electromagnetic calorimeter (ECAL). The basic function of ECAL is to measure the energy deposited by the particles showering in the materials of ECAL. Thus the electron, photon and part of hadron's energy can be measured. In addition, the PFA-oriented ECAL provides imaging capabilities which can help to distinguish between electromagnetic and hadronic showers and separate by-close showers. In order to improve such performance especially for the close-by showers' separation, the high granularity is needed. Moreover, the small Moliere radius and the large ratio between interaction length (λ) and radiation length (X_0) are also preferred in order to obtain a more compact ECAL (saving the cost of detector system) and also for better close-by showers separation. To satisfy the restriction of spatial size of ECAL , the tungsten is one of the most suitable material as absorber. It has a radiation length of

3.50 mm, interaction length of 99 mm and a Moliere radius of 9 mm. Two options for the sensitive medium for ECAL are being considered. One is a silicon sensor that is considered as the baseline choice and the other is a scintillator-based which is an alternative option.

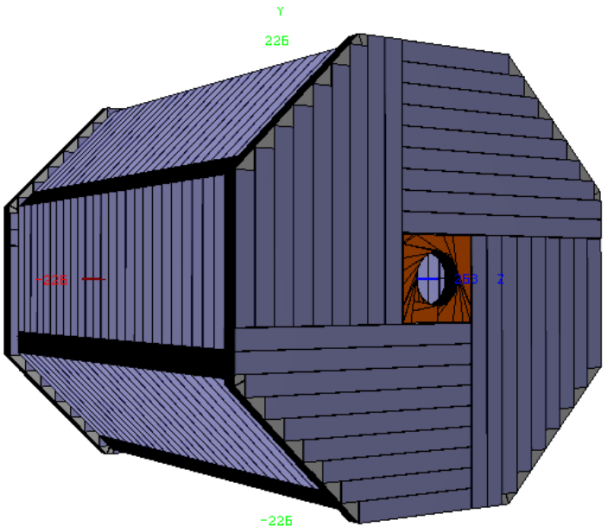


Figure 2.11: *The layout of CEPC baseline ECAL [20].*

- **Silicon-tungsten sandwich ECAL** : The silicon-tungsten ECAL is comprised by two parts. One is cylindrical barrel and another is two disk-like endcaps. For each of the two parts, it is organized in 30 layers with silicon-tungsten sandwich structure. The radius of the ECAL barrel is about 2028 mm in transverse direction, whilst the two ECAL endcaps are arranged at ±2635 mm of longitudinal direction respectively, the corresponding geometry of ECAL is shown in Fig. 2.11. There are several advantages for silicon as sensitive medium for ECAL, such as the stability, uniformity, flexibility, high signal-to-noise (S/N) ratio and timing capability. To balance the cost and physics performance, the size of the silicon sensor size of ECAL is chosen to be 10 mm × 10 mm with thickness 0.5 mm. Based on these parameters, the resolution of ECAL can achieve $\sigma_E/E = 17{,}1\%/\sqrt{E} \oplus 1{,}0\%$. The detailed study on the proposed silicon-tungsten ECAL can be found in Ref. [20].
- **Scintillator-tungsten sandwich ECAL** : Compared with silicon sensor, the scintillator-SiPM to be used to build the ECAL is less expensive and can also be cast in a compact way. This type of ECAL is similar to the silicon-tungsten ECAL except for the thickness and shape of the sensitive layers. The scintillator-tungsten ECAL is composed of sensitive layers with 2 mm thickness and 5mm × 45 scintillator strips with SiPM attached to one of the two strip's ends to read out the photon signals. The layout of the scintillator ECAL module is shown in Fig. 2.12. More information about this option can be found at Ref. [20]

The PFA-oriented high granularity hadronic calorimeter (HCAL). For CEPC project, the jet energy resolution will be a key factor for the success of precision measurements of Higgs properties. In general, the jet energy is the result of the energy deposition of charged and neutral hadronic showers. Even the jet energy of neutral component of

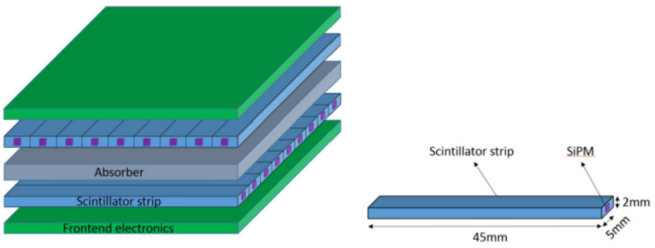


Figure 2.12: *The layout of CEPC baseline ECAL module with scintillator-SiPM [20].*

hadronic showers is only about 10% on average, it indeed has a wide range fluctuations from event to event. This effect limits the improvement of jet energy resolution when energy increasing to more than 100 GeV. This can be explained by that in the high energies (>100 GeV), the resolution described by PFA framework is limited by the confusion term which is influenced by the topological pattern recognition mismatch and the wrong segmentation for the hits belonging to charged and neutral hadronic showers. The high granularity provides a promising way to solve the confusion problem, thus the CEPC employs the PFA-oriented high granularity HCAL to obtain the satisfying jet resolution cooperating with other sub detector systems. The PFA-oriented HCAL are sampling calorimeters (classical sandwich structure) where the stainless-steel is appropriate materials (the interaction length is 17 cm, the radiation length is 1.8 cm) as the absorber. For the active layer, there two are main options being considered, one is gaseous detector (such as RPC, GEM) and another is scintillator tiles coupled with SiPM corresponding to the (Semi) Digital Hadronic Calorimeters ((S)DHCAL) and Analog Hadronic calorimeter (AHCAL). The schematic of CEPC HCAL layout is shown in Fig 2.13. It is composed of two parts, the barrel and two endcaps, for both of them which are made of 40 stainless steel absorbor plates (20 mm thickness) interleaved with sensitive layers. Here after is a brief summary of all the technological options of HCAL.

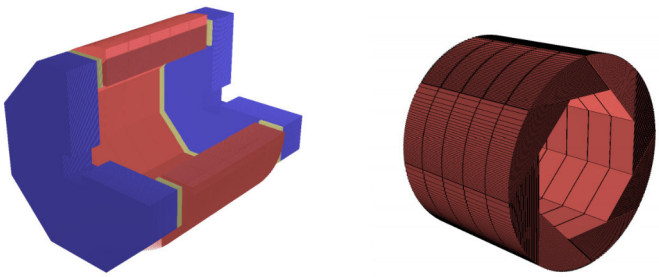


Figure 2.13: *The geometry layout of CEPC HCAL [20].*

- **The Semi-Digital Hadronic CALorimeter (SDHCAL)** : The SDHCAL has a fine lateral segmentation (1 cm × 1 cm) and negligible dead zone. There are two options; the baseline one is Glass Resistive Plate Chamber (GRPC) and the Thick Gas Electron Multiplier detectors (THGEM) as the alternative one. RPCs are really cost-effective and very fast-timing devices and they are widely used in particle experiments. The schematic of GRPC is shown in Fig. 2.14. Many investigations about the efficiency

and multiplicity of GRPC have been done, from which the average efficiency is 0.96 ± 0.03 and multiplicity is 1.76 ± 0.17. The first technological GRPC-based SDHCAL prototype was built in 2011 and has since been exposed many times to beam test at CERN (Fig 2.15). According to the beam test results, its energy resolution for hadron beams can achieve 7.5% at 80 GeV. A detailed discussion about GRPC-based SDHCAL is in Section. 4.1.

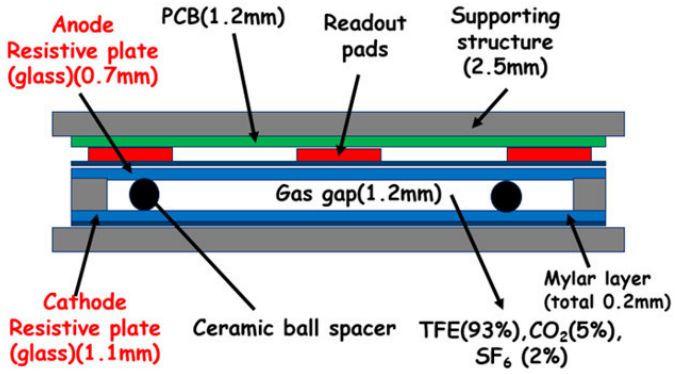


Figure 2.14: *The schematic of GRPC of CEPC SDHCAL.*

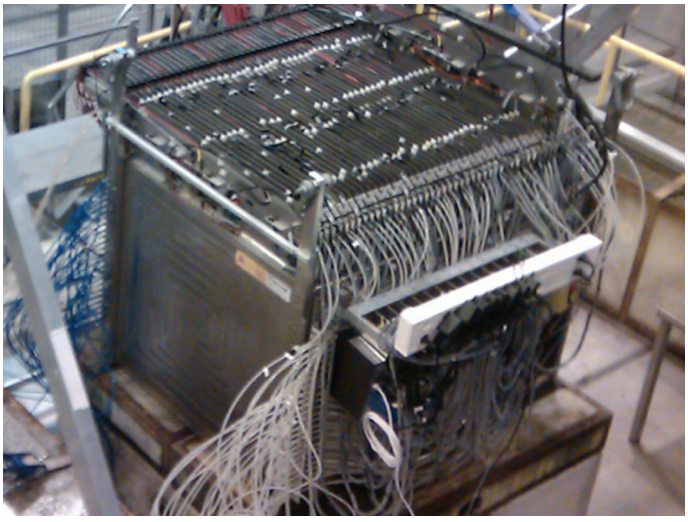


Figure 2.15: *The SDHCAL technological prototype exposed in the test beam.*

The THGEM is another appropriate detectors options for SDHCAL. Because it can be also produced in a cost effective way with large size and can provide flexible layout allowing high granularity. The THGEM also has good timing capability and very fast recovery time. The schematic of three different types of THGEM is shown in Fig. 2.16. The energy resolution achieves 20% when the prototype of THGEM tested by Fe^{55} radiation sources [20].

- **The Analog Hadronic Calorimeter (AHCAL)** : Under the collaboration with CALICE, the AHCAL is another technology options for CEPC HCAL. Without using gaseous detectors as active layer, the AHCAL adopts the scintillator tiles read out by SiPM as active layer, which is also found to fulfill the requirements of CEPC

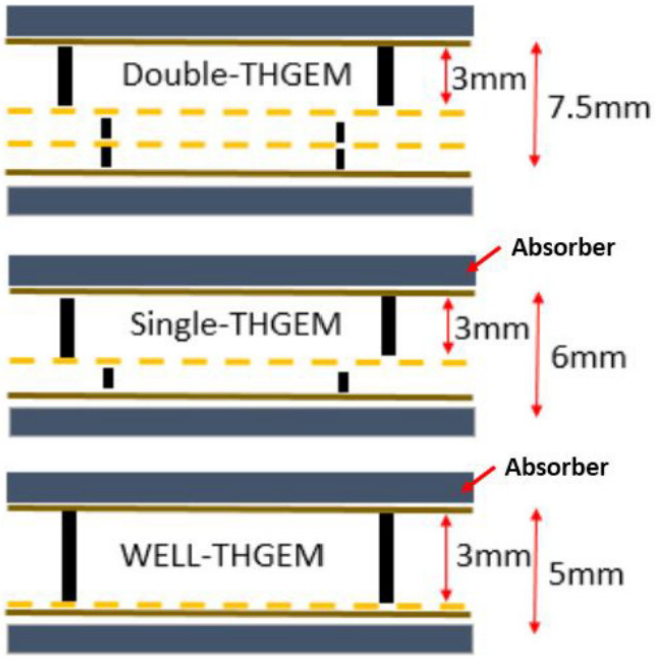


Figure 2.16: *The schematic of three different types of THGEMs [20].*

HCAL. The schematic of AHCAL single layer structure is shown in Fig. 2.17. The cell size of scintillator is chosen to be 30 mm × 30 mm following a cell size optimization study. A physical AHCAL prototype was built in 2006 [33]. The energy resolution of AHCAL achieves 11% at 20 GeV and 8% at 80 GeV. A new technological prototype with embedded electronics as for the SDHCAL was built in 2017. Energy resolution of the new technological AHCAL is to come soon.

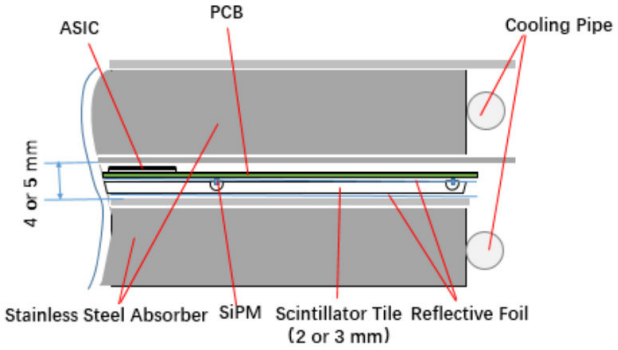


Figure 2.17: *The schematic of single AHCAL layer which is composed by a scintillator tile as the sensitive medium and stainless steel plate as the absorber [20].*

2.2.3.3 Muon detector system

The muon detector system is expected to enhance the capability of PFA-oriented calorimeter and tracking system for muon identification. It also can collect the energies leaked from

the hadronic calorimeters, thus it will help to precisely measure the jet energy resolution. There are several technology options for this system, the RPC-based, scintillator-based and the Micro Pattern Gas Detector (MPGD)-based detectors are all under current study to equip the muon system.

2.2.4 The IDEA detector concept

Apart from the baseline detector, there is another alternative detector concept: **Innovative Detector for Electron-positron Accelerator (IDEA)**. It is also the technology option for the detector concept of FCC-ee. The layout for the detector concept is that from the innermost to outermost in transverse direction, the vertex detector, the drift chamber , the preshower, the solenidal magnet system, the dual-readout calorimeter and the muon chamber are arranged in order. The corresponding schematic of IDEA detector concept is shown in Fig. 2.18.

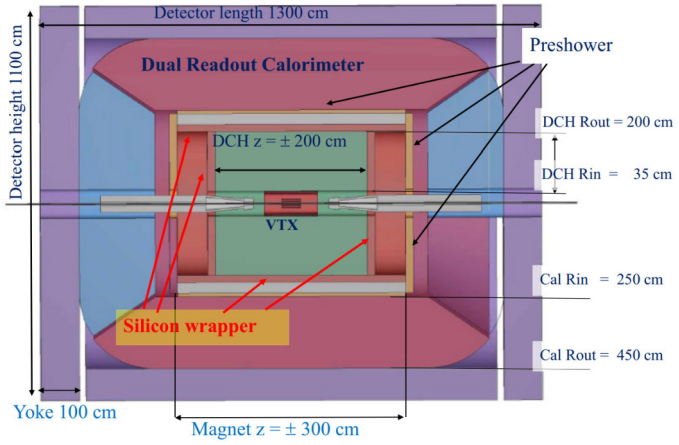


Figure 2.18: *The layout of IDEA concept.*

Comparing with the baseline detector concept, the IDEA concept has two main different features. one is the magnet system, for which the baseline detector concept has 3 Tesla magnetic field but IDEA is 2 Tesla magnetic field. The strength loss of magnetic field in the IDEA compensated by a large volume of tracking system. The another feature concerns the calorimeter system. To distinguish from baseline concept employing the PFA-oriented calorimeter, IDEA adopts the dual-readout calorimeter concept. One of the biggest challenges for PFA-oriented calorimeter is how to deal with huge amounts of readout channels which bring the problems such as cable management, cooling system and data processing. Thus the dual-readout calorimeters with relatively less channels are also pursued. With dual-readout concept, the ECAL and HCAL of IDEA are combined together into the same framework where both of them can measure the electromagnetic and hadronic shower energies. The principle of dual-readout method is that the showers can be independently sampled by two process, one is the photon emission from scintillation and another is the photon emission from Cerenkov. Thus the energy and EM faction of the shower can be simultaneously reconstructed by combining the two measurements. Based on the study of DREAM/RD52 collaboration, the resolution of dual-readout calorimeter energy is found to be: EM: $10\%/\sqrt{E}$, Hadronic: $60\text{-}70\%/\sqrt{E}$, within the energy range

from 10 GeV to 150GeV. This fulfills the requirements of CEPC. More information about dual-readout concept can be found in Ref. [34, 35, 36].

2.3 The other future leptonic collider projects

The charged particle (like electron , positron) will emit photons and lose energy when it changes direction (as it must in a circular collider). This effect, called synchrotron radiation, can be avoided by accelerating charged particles in a straight line. This is one of the advantages of linear collider. The two linear leptonic collider project, the International Collider and the Compact Linear Collider, will be introduced. Finally, another circular collider option Future Circular Collider will be briefly discussed.

2.3.1 The International Collider project (ILC)

The International Linear Collider ILC is a electron positron linear collider with high luminosity, based on 1.3 GHz superconducting radio-frequency (SCRF) accelerating technology [37]. Its centre-of-mass-energy $\sqrt{s}$ range is from 200 GeV to 500 GeV which is capable of upgrading to the 1 TeV with extending the main Linac to 50 km. For the baseline 500 GeV Machine, the luminosity of ILC can be expected to be 0.75 , 1.0 and 1.8 $\times 10^{34}\ cm^{-2}s^{-1}$ corresponding to $\sqrt{s}$ at 250, 350 and 500 GeV respectively. The layout of ILC is shown in Fig. 2.19.

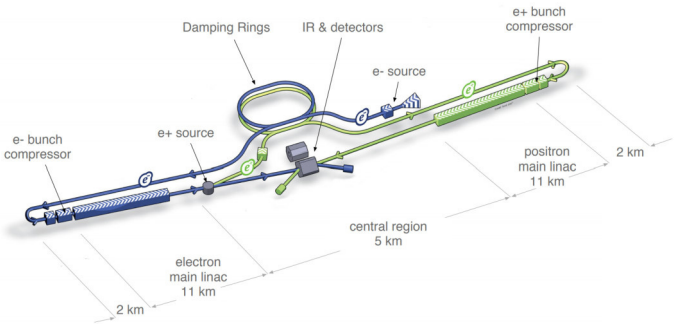


Figure 2.19: *The layout of ILC accelerator complex [37].*

The total length of ILC is approximately 31 km. The two main linacs is 11 km, using 1.3 GHz SCRF cavities for accelerating the beams at an average gradient 31.5 MV/m. One advantage of ILC is that its beam is polarized, which can help to obtain the data including more interesting physics events and simultaneously suppress a part of the backgrounds. Another advantage of ILC that it can operate at any energy within the a wide range allowed by its technology with small modification compared with circular collider. Thus many physics cases can be studied as follows:

- **91 GeV** : the study of precision measurements of electroweak with the physics process: $e^+e^- \rightarrow Z$.
- **160 GeV** : ultra-precision measurements of W boson mass with physics process: $e^+e^- \rightarrow W^+W^-$.

- **250 GeV** : precision measurements of Higgs couplings with $e^+e^- \rightarrow ZH$
- **350-400 GeV** : the top quark mass and its couplings; the precision W and Higgs couplings.
- **500 GeV** : Higgs self couplings, Higgs couplings to top quarks, the search for supersymmetry, extended Higgs states.
- **700-100 GeV** : Higgs self-coupling and new physics.

2.3.2 The Compact Linear Collider project (CLIC)

The CLIC is also a linear leptonic collider project developed by CERN whose main linac for acceleration is 42 km. The origin of "compact" due to the fact that it has very powerful accelerating gradient expected to be 100 MV/m. As a comparison, to achieve the same centre-of-mass-energy, the LHC need a very long distance about 840 km with its acceleration 5 MV/m. The CLIC is capable of extending the energy range from 380 GeV to 3 TeV using a novel technology called two beam acceleration (TBA) which replaces the klystrons (the classical power supplier of RF) by using intense particle beam, called the drive beam, to convert to the power of RF cavities.

The CLIC are expected to provide the answers for key open questions in particle physics such as the Higgs physics, hidden mechanism of electroweak symmetry, the top-quark physics, the dark matter, the other new phenomena of BSM and so on. The integrated luminosity is 2.3 and 5.9 $\times 10^{34} cm^{-2} s^{-1}$ corresponding to $\sqrt{s}$ at 500 GeV and 3 TeV respectively. The layout of CLIC accelerator complex and possible location are shown in Fig. 2.20.

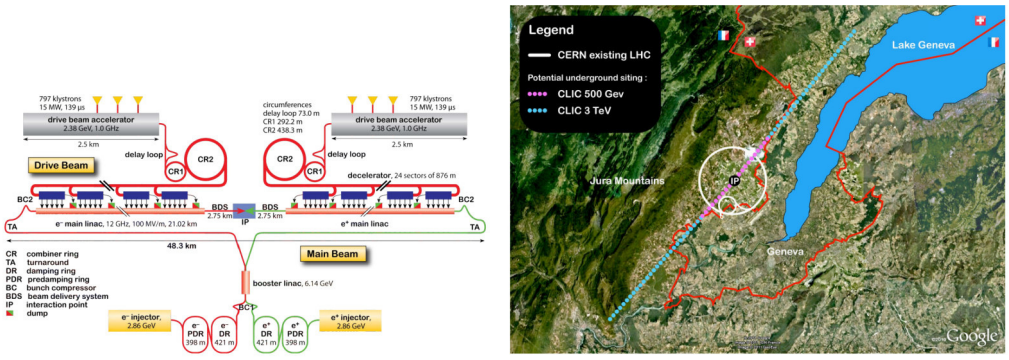


Figure 2.20: *Left: The layout of CLIC accelerator complex at 3 TeV. Right: the possible location of CLIC in the map. The two figures are cited from [38].*

2.3.3 Future Circular Collider project (FCC)

The FCC is a project similar to CEPC, to be hosted by the CERN. This projects consists of two steps, the first is high energy ($\sqrt{s}$ between 90 and 350 GeV) leptonic collider with highest-luminosity (FCC-ee) and as for the CEPC, it will have a second phase, a hadronic one at the energy-frontier of $\sqrt{s}$ =100 TeV) called FCC-hh. Both of them can share the same 100 km tunnel with minimal modifications. The rich physics case to be studied

by FCC-ee such as the electroweak precision measurements, the Higgs boson properties, the potential of new physics discovery and so on. The luminosity per IP of FCC-ee in different physics/energy modes is summarized in Tab. 2.3 .

Operation mode	Z	W^+W^-	ZH	$t\bar{t}$	
Beam energy (GeV)	45.6	80	120	175	182.5
Luminosity/IP ($\times 10^{34} cm^{-2} s^{-1}$)	230	28	8.5	1.8	1.55

Table 2.3: *The luminosity of FCC-ee for different operating modes. The full parameters of FCC-ee can be found in Ref. [39]*

The overview of FCC-ee accelerator complex and future location are shown in Fig. 2.21.

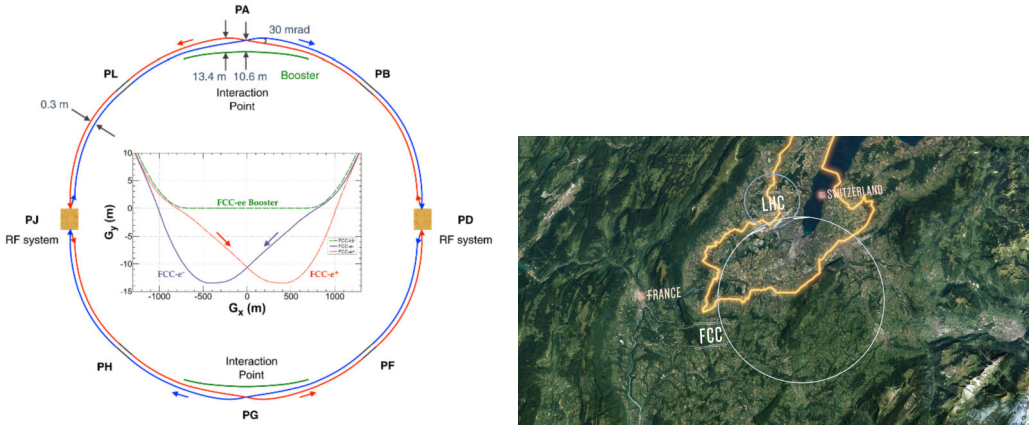


Figure 2.21: *Left: The layout of FCC-ee accelerator complex with two IPs. Right: the future location of FCC in the map. The two figures are cited from [39].*

2.4 Summary

The discovery of the Higgs boson opens a new era for particle physics. Particle physicists proposed several future collider projects such as ILC, CLIC, FCC and CEPC to precisely measure the Higgs properties and to search the new physics of BSM.

CEPC is a double-ring collider with approximately 100 km ring length and it is designed to operate at around 91.2 GeV as a Z factory, at around 160 GeV of the WW production threshold, and at 240 GeV as a Higgs factory. The abundant physics content can be studied at CEPC. In order to satisfy the demands of physics program of CEPC, the detector concepts of CEPC are also studied.

CHAPTER 3

The basics of calorimetry: the interactions of particles with matter.

In the high energy collider experiments, calorimeters are an essential part of the detection systems. The are used to measure the energy of electrons, photons and that of jets. They also contribute to particle identification. In order to understand the principle of calorimeter, it is necessary to know the interaction of particles with matter. According to the nature of the particles' interaction in matter, we can divide them into two categories: (a) the electromagnetic interactions ; (b) the strong interactions. In this chapter we will review the two kinds of interactions within calorimeters.

Content

3.1 Electromagnetic interactions of particles with matter

One of the important mechanisms related to charged particle energy loss in matter is through **Electromagnetic interactions**. These interactions are well understood and described by QED. There are several processes through which the electromagnetic interactions manifest themselves. The most important ones are: (a) the ionization; (b) the bremsstrahlung at high energies; (c) photon interactions.

3.1.1 The ionization energy loss

When a charged particle carrying certain energy passes through the matter, it electromagnetically interacts with the atomic electrons and thus loses part of its energy by ionizing the matter's atoms if its energy is enough to release the atomic electrons. Without a loss of generality, considering a charged particle with the charge ze and the velocity βc passing through the absorber with atomic number Z, the average ionization energy loss per unit length passed can be described by the Bethe-Bloch formula:

$$- < \frac{dE}{dx} >= Kz^2 \frac{Z}{A} \frac{1}{\beta^2} [\frac{1}{2} ln \frac{2m_e c^2 \beta^2 \gamma^2 T_{max} I^2}{-} \beta^2 - \frac{\delta}{2}], \tag{3.1}$$

where I is the mean excitation energy of the absorber, the T_{max} is the maximum energy transfer to electron in a single collision, the δ is the density effect correction term and K is constant term,

$$K = 4\pi N_A r_e^2 m_e c^2 = 0.3071 \frac{MeV}{g/cm^2}.$$

The term $< dE/dx >$ is often referred to as the stopping power. This formula can describe well (with accuracy of less than 10 %) the mean energy loss of particles in the range of approximately $0.1 \lesssim \beta\gamma \lesssim 1000$ for absorber with intermediate Z number[1]. Fig. 3.1 shows the relation between the $\beta\gamma$ and mean energy loss $< dE/dx >$ for μ^+ for different types of particles passing different materials (like liquid hydrogen, carbon, aluminum, iron and so on).

From the Fig. 3.1, we can find that at low energy region, the mean energy loss falls rapidly, approximately proportional to the factor $1/\beta^2$ and then reaches the relative wide area around the minimum at $\beta\gamma \approx 4$. The relativistic particles corresponding to this minimum are referred to as the minimum-ionization-particle (MIP). For the absorber materials whose ratio of $A/Z \approx 0.5$, the mean energy loss rate of MIP is about:

$$- < \frac{dE}{dx} >|_{min} \approx 1 - 2 \frac{MeV}{g/cm^2}.$$

It is worthy to be mentioned here, the muons even at high energy (few hundreds GeV) whose energy loss is dominated by the ionization loss and it is treated as the MIP in common, as shown in Fig 3.2. Thus the large amounts of absorber materials are needed to shield the penetration of muons. For this reason, many experiments which need to exclude the cosmic muon background events not interested are located in deep mines or in caves of high mountains. When charged particles traverse through relatively thin width

[1] In the case of $\beta\gamma > 1000$, the ionisation energy loss is negligible compared with bremsstrahlung.

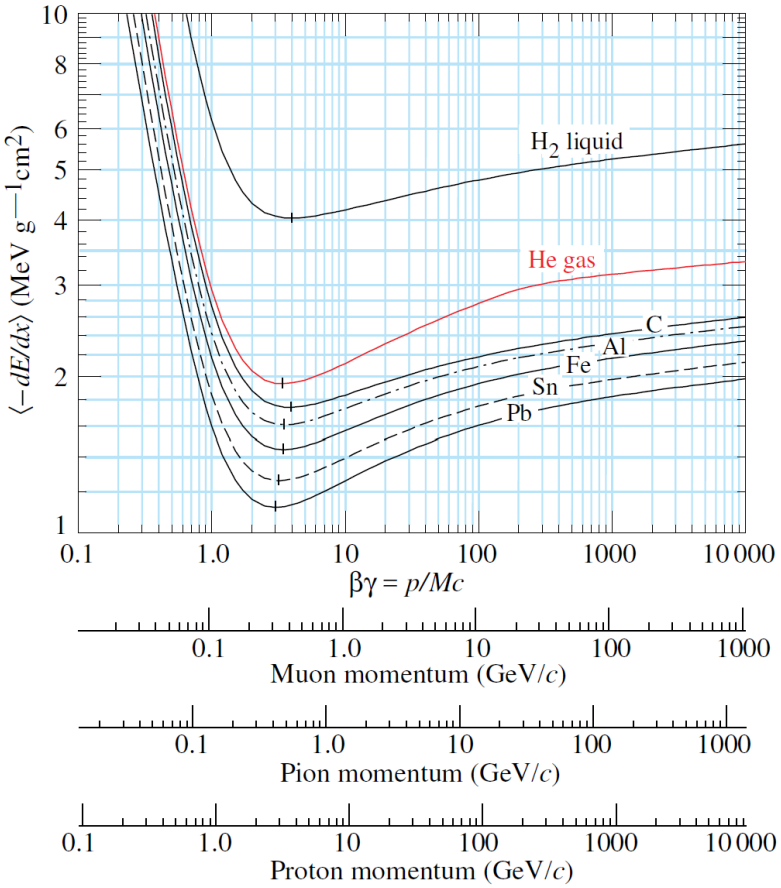


Figure 3.1: *The mean energy loss rate for different types of particles such as muons, pions and protons in different absorber materials. This figure is cited from Ref. [9].*

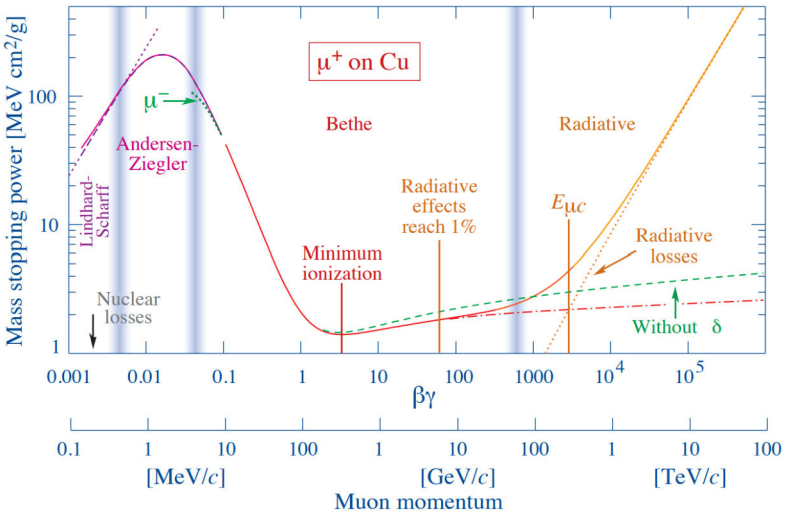


Figure 3.2: *The relation between the mean energy loss rate and $\beta\gamma$ for μ^+ passing through copper-made absorbers. This figure is cited from Ref. [9].)*

absorber like the absorber layer of typical sampling calorimeter, the total energy loss

$\triangle$E/$\triangle$x could largely differ the value integrated from the mean energy loss $< dE/dx >$. This can be explained by the relatively small number of collisions between particle and atomic electrons of absorber, and the large fluctuations of energy transfer that may occur in such collisions. Thus the most probable of energy loss distribution measured through thin materials is less than the mean energy loss and there is a large tail (referred to as Landau tail [40]) located in the large energy loss region.

3.1.2 Bremsstrahlung energy loss

When a relativistic charged particle passes through matters, it interacts with the electrostatic fields of the atom nucleus of matter and loss its kinetic energy by radiating photons. This process is named bremsstrahlung (German for braking radiation). At high energies, the energy loss of bremsstrahlung makes the primary contribution to the total energy loss. The equation of mean energy loss rate of bremsstrahlung can be given by:

$$- < \frac{dE}{dx} >= 4\alpha N_A \frac{Z^2}{A} z^2 (\frac{1}{4\pi\varepsilon_0} \frac{e^2}{mc^2}) E ln(\frac{183}{Z^{1/3}}), \tag{3.2}$$

where α is the fine-structure constant, z, m and E are the electric charge, mass and energy of incoming particle respectively. For the electron, its $-dE/dx$ can be transformed into:

$$- < \frac{dE}{dx} >= 4\alpha N_A \frac{Z^2}{A} z^2 r_e^2 E ln(\frac{183}{Z^{1/3}}), \tag{3.3}$$

where r_e is classical electron radius. It is clear that the energy loss of bremsstrahlung is proportional to incoming particle's energy E and inversely proportional to the square of its mass. The Eq. (3.3) can become:

$$- < \frac{dE}{dx} >= \frac{E}{X_0}, \tag{3.4}$$

where $\frac{1}{X_0} = 4\alpha N_A \frac{Z^2}{A} z^2 r_e^2 ln(\frac{183}{Z^{1/3}})$. The Eq. (3.4) defines a new physics quantity, **the radiation length** X_0, which represents the mean distance for a high energy electron losing all but $1/e$ fraction of its total energy through the bremsstrahlung. The radiation length is approximately proportional to Z^{-2} as can be deduced by estimating the Coulomb interaction cross-section between electrons and atom nucleus.

One fact is needed to be underlined here, that there is a competition between Ionization and bremsstrahlung processes to determine which kind of energy loss is the main component of the total energy loss. The critical energy, E_c is defined as the energy where the mean energy loss rate of ionization equals that of the mean energy loss rate by bremsstrahlung [41], giving:

$$for\ solids\ and\ liquids:\ Ec \approx \frac{610MeV}{Z + 1.24} \quad and \quad for\ gases:\ Ec \approx \frac{710MeV}{Z + 0.92},$$

as shown in Fig. 3.3. The electrons of interest in particle physics experiments are in the GeV and TeV range which are much larger than the critic energy and thus the energy loss is primarily from bremsstrahlung as shown in Fig 3.4. The energy loss of bremsstrahlung for other charged particles are suppressed by the factor $((m/m_e)^2)$. For example, the next-lightest charged particle is a muon whose mass is about 207 times heavier than the

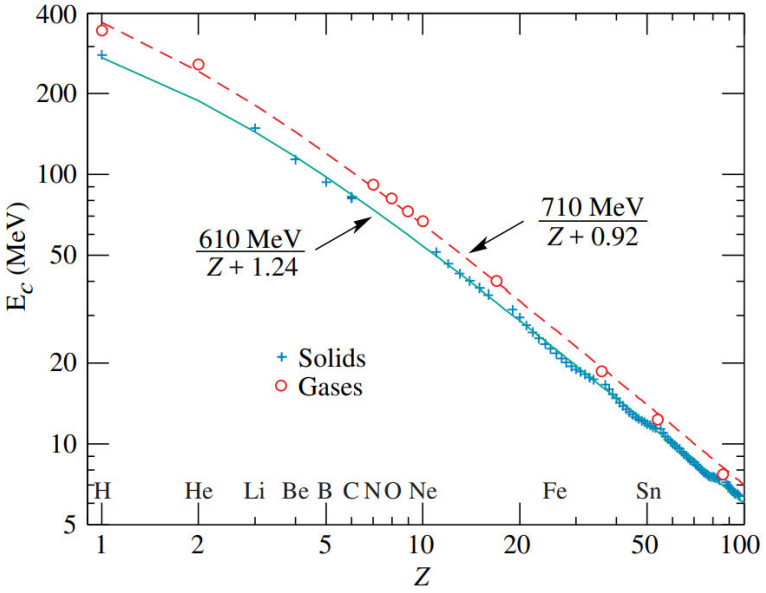


Figure 3.3: *The critical energy of electrons for different materials. This figure is cited from Ref. [9].)*

electron, thus the critic energy of a muon is 40000 times larger than that for an electron. This is the reason why the energy loss of muon is essentially due to ionization except for very high energies muons (>100 GeV).

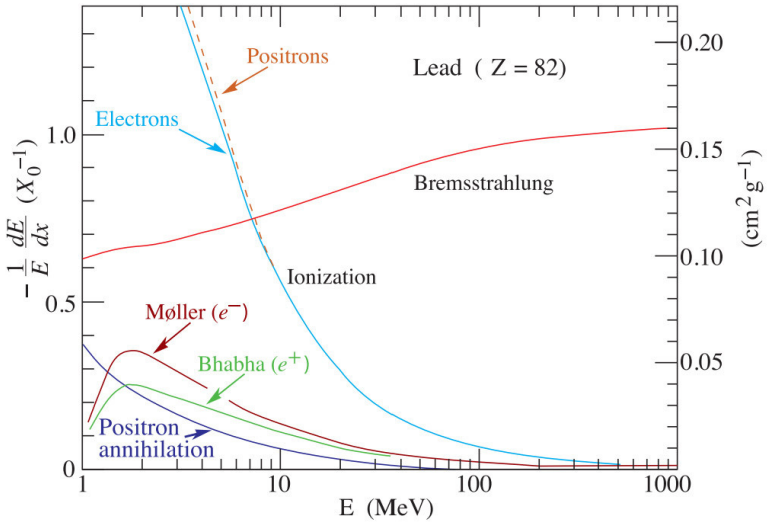


Figure 3.4: *the relation between injecting energy of electron or positron and fractional energy loss per radiation length in lead. This figure is cited from Ref. [9].*

3.1.3 The photon interactions

When the photon passing through matter, there are five different processes contributing to the interactions of photon with matter: (1) the photoelectric effect, (2) Rayleigh scattering,

(3) Compton scattering, (4) the pair production of electron-positron, (5) photonuclear absorption. Figure. 3.5 shows the corresponding cross-section for the different processes and the total photon cross-section is also shown.

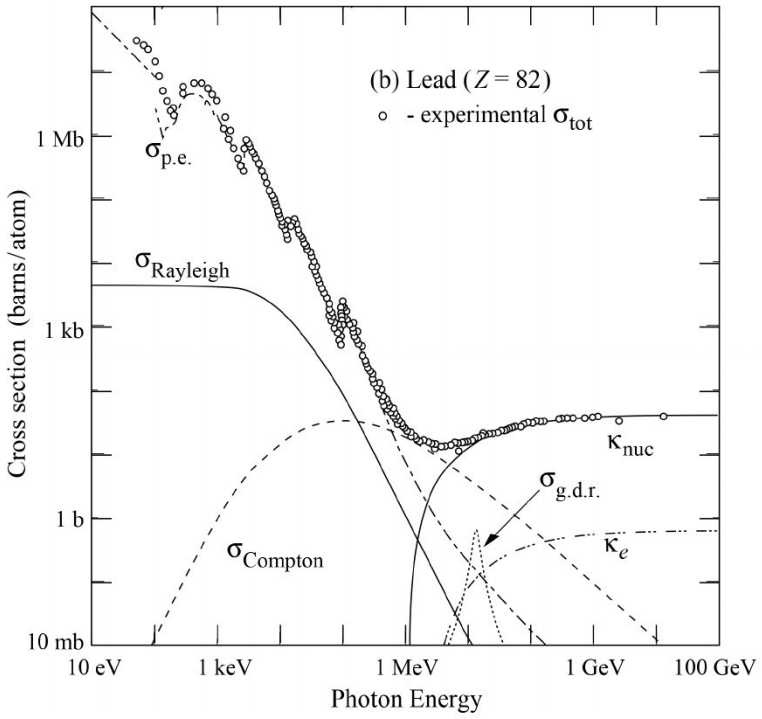


Figure 3.5: *the relation between photon energy and photon total cross-section. The branching cross-section as a function of photon energy for different processes of photon interaction with matter is also presented. This figure is cited from Ref. [9].*

- **The Photoelectric Effect**. For this process, an atomic electron can completely absorb a photon's energy and then escapes from the restriction of atomic nucleus. When the photoelectric effect occurs at the atomic inner shell (such as K-shell) , the vacancy of atomic inner shell can be filled by an electron of outer shell with high energy level. The energy difference $\triangle$E between these two shells results in the emission of Auger electrons or X-rays. The cross-section of photoelectric effect is proportional to the Z number of the materials with a factor varying between 4 and 5. It is the dominant process at low energies and rapidly loses its substantial contribution for energies exceeding 1 MeV.

- **Rayleigh Scattering process**. This is also important process at low energies where the photon is deflected by the atomic electrons of materials. One important fact it does not contribute to any energy loss of photon. Thus it only influence the spatial distribution of energy deposition.

- **Compton Scattering**. For Compton scattering, it describes that a photon is scattered by an atomic electron, with transfer of a fraction of momentum and energy to the electron. The cross section of Compton scattering is found to be proportional to the

Z value of the materials, because the atom nucleus has Z electrons. At high energies, its cross-section is proportional to $\frac{lnE}{E}$, where E is the photon energy. Thus in high energies (> 10 MeV), the energy loss of Compton scattering process is also limited as shown in Fig. 3.5.

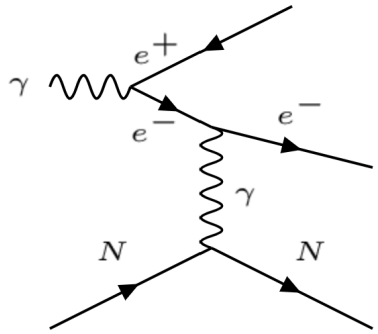


Figure 3.6: *The electron-positron pair production process. N represents an atom nucleus containing Z electrons.*

- **Electron-positron pair production**. For the pair production process as shown in Fig. 3.6, there is the lower limit energy threshold of photon's energy, it must be above electron-positron pair's invariant mass: 2 m_e≈1022 MeV considering $m_{nucleus}$ significantly larger than m_e. When photon energy is above 10 MeV, the pair production is dominant process comparing to others. The cross section of pair production approximately equals to:

$$\sigma_{pair} \approx \frac{7}{9}\frac{A}{N_A}\frac{1}{X_0}, \tag{3.5}$$

  where X_0 is the radiation length. The Equation. (3.5) indicates that the mean free path of photon equals to $\frac{9}{7}X_0$.

- **Photonuclear interactions** . This process occurs at energies in the range of 5 - 20 MeV. The cross-section of this process is found not to exceed 1% of total cross-section . It reaches the maximum when the photon energy is approximately equals to the marginal binding energy of atom nucleus.

3.1.4 The electromagnetic shower

When a high-energy electron or positron (e.g. multi-GeV) injects on the thick absorber, it can radiate a photon by bremsstrahlung process which only carries a fraction of the initial electron's (positron's) total energy. The photon then transforms into an electron-positron pair and lose its energy. The process of bremsstrahlung and electron-positron pair production subsequently continues to generate a cascade of photons, electron-positron pairs, which is referred to as the electromagnetic shower shown in Fig. 3.7. The case for high energy photons actually are similar to that of the electrons. They are first converted into electron-positron pair and then the electromagnetic shower following the

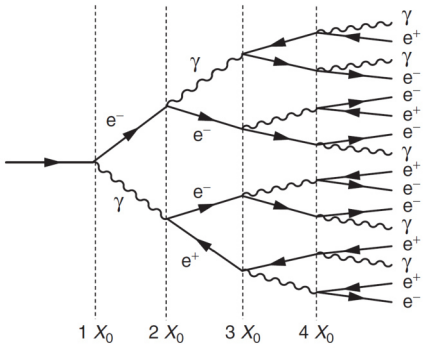


Figure 3.7: *The electromagnetic shower development .*

same mechanism as for the electrons. However, there are some differences between these two cases.

- **a**. The electromagnetic shower induced by an electron (or a positron) deposits its energy , on average, shallower inside the absorbers than the shower induced by an photon in the same energy. This can be explained by the nature of photons which crosses , on average, a mean free path of about $(9/7X_0)$ before their first interactions with materials. Thus the position of shower begin for electrons is shallower than for photons.
- **b**. The fluctuation of shower energy deposition for electrons is smaller than the photon-induced showers for passing the same width materials. This effect can be understood by the fact that beginning point of of shower induced by photons fluctuates from event by event but an electron (positron) immediately start its shower after entering materials.

With the development of shower, the number of particles rapidly increases and thus their average energy decreases. When the average energy of particles decreases to a value lower than a certain energy (approximately equals to the critical energy), the photons can still produce free electrons through the Compton and photoelectric effect processes but not through the pair production process. At the same time, the shower electrons (or positrons) are also more inclined to lose their energy by ionization process instead of through the bremsstrahlung process to generate photons. Hence, the shower development reaches its maximum and then gradually shrinks. By a rough estimation, one expects the number of electromagnetic shower particles to increase by a factor 2 after a radiation length equivalent of passed matter. Thus the number of particles of shower (with initial energy of E) after n radiation length can be estimated as 2^x and thus the average energy is approximately as $E/2^n$. When the shower particles' energy around critical energy, the shower development reaches its maximum, giving

$$Ec = \frac{E}{2^{n_{max}}} \Rightarrow n_{max} = \frac{ln(E/E_c)}{ln2}.$$

For instance, the lead is one of most often used materials in ECAL with critical energy $\approx$ 10 MeV, a 10 GeV electromagnetic shower reaches its maximum around 10 X_0 corresponding to about 6 cm lead thickness. Thus the electromagnetic showers can be restricted at

very compact space. The longitudinal profile of electromagnetic shower for a 30 GeV electron in the iron absorber, is shown in Fig. 3.8 .

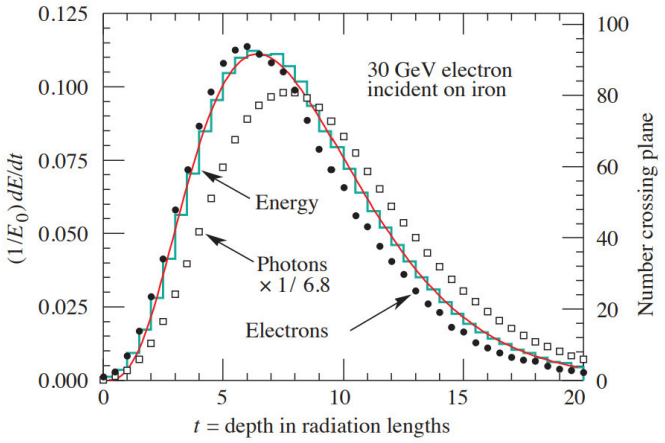


Figure 3.8: *The electromagnetic shower longitudinal profile of a 30 GeV electron in the iron absorber based on simulation. the unit of t is X_0. This figure is cited from Ref. [9].*

The electromagnetic shower development in the transverse plane can be described by the Molière radius. Its definition is,

$$R_M = X_0 \frac{E_s}{E_c},$$

where E_s is a scale energy defined as $\sqrt{4\pi/\alpha} m_e c^2$ ($= 21.2\ MeV$). On the average, the 90% of total energy of shower is inside of a cylinder with radius R_M around the shower axis. Finally, some important parameters of widely used materials are summarized in Tab. 3.1.

Material	Z	Density (g/cm^3)	X_0 (g/cm^2)	X_0 (cm)
Fe	26	7.87	13.84	1.757
Copper	29	8.96	12.86	1.436
Tungsten	74	19.30	6.76	0.350
lead	82	11.35	6.37	0.561
Uranium	92	18.95	6.00	0.317

Table 3.1: *The radiation length of different widely used materials for ECAL [42].*

3.1.5 The Electromagnetic calorimeter

In the high energy experiments, the energies of electrons (positrons) and photons can be measured by using Electromagnetic CALorimeter (ECAL), which also helps for particle identification. The ECAL can be broadly categorized into two classes, the homogeneous and sampling ones.

- **Homogeneous calorimeter**. Just like the manifestation of name itself, for homogeneous calorimeter, its whole volume are used as active media and thus make contributions to the signals. The main advantage of homogeneous calorimeter is that

it has excellent energy resolution. According to the nature of the active materials, it can be roughly divided into four categories: the semiconductor calorimeters; the Cherenkov calorimeters; the scintillator calorimeters and the noble liquid calorimeters. For example, in the CMS detector, it has a homogeneous ECAL which is built with approximately 75000 inorganic scintillator crystals of $PbWO_4$. This scintillator has a very small radiation length (X_0 = 0.89 cm) thus can contain the EM showers in a compact space. The important parameters of commonly used materials for homogeneous ECAL in high-energy collider experiments are summarized in Tab. 3.2.

	Scintillating crystals				Scintillating liquids	
	Na(Tl)	CsI(Tl)	BGO	$PbWO_4$	Liquid Kr	Liquid Xe
ρ (g/cm^3)	3.67	4.53	7.13	8.28	2.41	2.95
X_0 (cm)	2.59	1.85	1.12	0.89	4.70	2.77
R_M (cm)	4.5	3.8	2.4	2.2	4.70	4.2

Table 3.2: *The density, radiation length and Molière radius for commonly used materials in homogeneous ECAL.*

The resolution of homogeneous ECAL used by some important experiments are summarized in Tab. 3.3.

	Homogeneous ECAL		
Experiments	BaBar	LEP L3	CMS
Active medium	CsI(Tl)	BGO	$PbWO_4$
Resolution	$\frac{2.30\%}{\sqrt{E(GeV)}} \oplus 1.35\%$	$\frac{2.0\%}{\sqrt{E(GeV)}} \oplus 0.7\%$	$\frac{3.0\%}{\sqrt{E(GeV)}} \oplus 0.5\% \oplus 0.2\%/E(GeV)$

Table 3.3: *The resolution for different homogeneous ECAL with different active medium technologies.*

- **Sampling calorimeter**. Unlike the homogeneous calorimeter, the sampling calorimeter is built with two different components. One is the passive medium which is responsible for primary energy absorption of EM showers. The passive media are commonly made of high-density materials such as lead, iron, copper and uranium. The second is the active medium which is responsible for signal generation and a small fraction of (or negligible) energy absorption compared with the passive medium. According to the type of materials used for the active medium, the classification of sampling calorimeter also can be organized as: gas calorimeters, solid-state calorimeters, liquid calorimeters and scintillator calorimeters. Apart from the final case for which the signals are collected in the form of photons (fluorscent), in the other cases signals are collected in the form of electric charges. For example, the active medium of the ATLAS ECAL is Liquid Argon working at -183 ^{o}C. In sampling calorimeters, the active and passive media are made of layers interleaved with each other. This is commonly referred to as a "sandwich structure" as shown in Fig. 3.9.

  The resolution performance of sampling calorimeters generally are worse than ho-

mogeneous calorimeters as a result of the sampling fluctuation of its interleaved structure of active layers and passive layers. But on the other side, the sampling calorimeter is more flexible than the homogeneous one allowing to easily arrange the longitudinal and transverse segment, thus leading to better space resolution and particle identification performance.

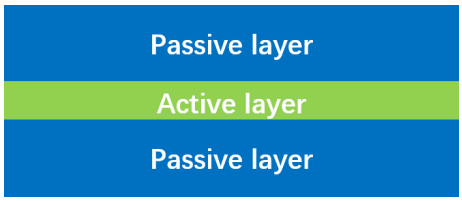


Figure 3.9: *The classical sandwich structure of sampling calorimeter.*

The resolution of sampling ECAL used by some famous experiments are summarized in Tab. 3.4.

	Sampling ECAL		
Experiments	ATLAS	ZEUS	DØ
Active medium	Liquid Argon	scintillator	Liquid Argon
Passive medium	lead(*)	Uranium	Uranium
Resolution	$\frac{10.0\%}{\sqrt{E(GeV)}} \oplus 0.4\% \oplus 0.3\%/E(GeV)$	$\frac{18\%}{\sqrt{E(GeV)}}$	$\frac{16.0\%}{\sqrt{E(GeV)}} \oplus 0.3\% \oplus 0.3\%/E(GeV)$

Table 3.4: *The resolution for different sampling ECAL. (*) Depending the postion of the ECAL segment, the passive medium (including lead, tungsten or copper) of ATLAS ECAL is different.*

In general, the energy resolution of electromagnetic calorimeter can be described by the equation,

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus c \oplus \frac{b}{E}, \tag{3.6}$$

where the mark "⊕" means a quadratic sum and E is expressed in the unit of GeV. The first tem of RHS of Eq. (3.6) is referred to as the "stochastic term". The second term is called ”constant term” and the third one is referred to as "noise term".

- **Stochastic term**. This term results from the statistical fluctuations related to shower developments, photo-electron statistics, sampling fluctuations and so on. For homogeneous calorimeters, because the energy of the incoming particles is fully deposited in the whole active volume and does not fluctuate from event to event, thus the intrinsic statistics-related fluctuations are small. Hence, this term for homogeneous calorimeter is small [2]. However, in sampling calorimeters, because of its interleaved structure of active and passive layers, its energy deposition presents a fluctuation from event to event. This fluctuation is referred to as "sampling fluctuation", which

[2] typically, it is around a few percents.

is found to be proportional to the $\sqrt{d/f_s}$. The parameter d represents the thickness of passive layer. The parameter f_s is the sampling fraction of MIPs [43]. The sampling fluctuation contributes significantly to the stochastic term of energy resolution. In general, the stochastic term for sampling calorimeters varies in the range from 10 to 20% with units in $1/\sqrt{E(GeV)}$.

- **Constant term**. Constant term includes the effect which does not rely on the energy of incoming particles. The primary source of the constant term b is from the non-uniformity of detector. For example, the abnormal shape of active and passive layer can contribute to this term. In addition, the radiation damage and mechanical defect of detector also affect it. In general, the constant term of ECAL should be protected to less than 1% especially in future high-energy colliders since in the high energy scale, this term is the dominant one for energy resolution.

- **noise term**. This term results from the electronic noise of the detector system.

3.2 Strong interactions of particles with matter

When charged hadrons (such as protons and charged mesons) pass through the thick absorbers, there are two kinds of interactions that contribute to their energy loss. One is through the ionization process of the electromagnetic interaction which is already discussed in the previous section. Another is through the strong interaction with the atomic nucleus of the absorbers. Because of its more complex characteristics, the manifestation of strong interaction is more complicated than the electromagnetic ones.

In hadronic showers, some π^0s generated through the strong interactions, they will primarily decay to two photons through the electromagnetic interaction, leading to pure electromagnetic component of hadronic shower. The fraction of the energy of this electromagnetic component is determined by the number of π^0s generated and strongly fluctuates from event to event. On average, the number of π^0s occupy 1/3 of the total number of mesons generated in the hadronic interactions [44].

When strong interaction occurs between the high energy hadron and the atomic nucleus of absorber, the most possible process for energy loss is from the nuclear spallation reactions. Through this way, the hadron could convert to tens of hadrons dramatically and at the same time, the atomic nucleus may experience the de-excitation from excited-state by γ-rays emission and the nucleus may also emit nucleons or nucleon aggregates if the energy is sufficient to cover the nuclear binding energy. The last process is very important because this part of energy is lost and thus does not make contributions to the calorimeter signal, referred as the invisible energy. There are large fluctuations from event to event for invisible energy that in some extreme case, the invisible energy is almost negligible and on the contrary in other cases it may rise up to 60% of shower energy. On average, the invisible energy represents 30-40% of the non-electromagnetic energies of the hadronic shower [43]. The large fluctuations certainly weaken the energy measurement performance of HCAL and thus the resolution of HCAL is generally worse than ECAL.

In nuclear spallation reactions, abundant "evaporation" neutrons are produced. Because of neutrons are electrically neutral, thus the only choice for neutron losing energy is through strong interaction (rarely from weak interaction) with the absorber. There

are two main ways for neutrons losing their energy including the scattering (elastic and inelastic) and the capture mechanisms.

- **Elastic scattering of neutron**. If the total kinetic energy of neutrons and atomic nucleus are kept the same before and after interaction, this process is referred to as the elastic scattering. The average kinetic energy loss for incoming neutron:

$$\Delta E = \frac{2TA}{(A+1)^2}$$

  where T is kinetic energy of neutron and A is the atomic number of the absorber material. It is clear from the previous formula that we can use low-Z materials to rapidly slow down neutrons. For example, to slow down a neutron from 2 MeV to 0.025 MeV (thermal neutron), it needs approximately 27 with hydrogen and 2175 collisions with uranium, respectively.

- **Inelastic scattering of neutron**. Unlike elastic scattering, the total kinetic energy of neutrons and atomic nucleus after the collisons are smaller than before collisions in inelastic scattering process. The kinetic energy loss will bring the nucleus to the excited state and then the excited neutron loss their energy through emitting γ-rays. Because of internal energy level complexity of different chemical elements, it is difficult to give global formula to describe its contributions for energy loss.

- **Neutron capture**. To distinct with scattering process, neutron also can be absorbed or captured by an atomic nucleus of materials which become an intermediated state nucleus. It may subsequently emit γ-rays, may emit charged particles such as protons, α particles. Even more, the intermediated nucleus may also emit neutron or neutrons.

3.2.1 The hadronic shower

The hadronic shower development is analogous to the electromagnetic shower development. The particles produced in the first strong interaction with atomic nucleus (such as protons, neutrons, mesons) will subsequently lose their energy through ionization process and/or nuclear reactions to continue the development of hadronic shower. However, the energy of shower particles decreases because of the number of particles rapidly increased with shower development. Thus at certain depth, the shower reaches its maximum and after the maximum, the number of shower particles gradually decrease (The absorption rate of particles is larger than their production rate.). Because of the complexity of strong interaction between hadron and atomic nucleus discussed by last section, the hadronic shower behaviour is more complicated than electromagnetic one. There are two components making the hadronic showers.

- **The electromagnetic (EM) component**. In hadronic shower, about 1/3 of mesons produced are π^0s, which subsequently decay to two photons arising the primary EM component of hadronic shower as shown in Fig. 3.10. The energy of this component largely fluctuates from event to event. The fraction of EM component energy included in the shower energy is referred to as the f_{EM}. The average EM

fraction can be parameterized by two factors following the power function law, giving

$$\langle f_{EM} = 1 - [(\frac{E}{E_0})^k],$$

where E_0 indicate the threshold energy (varying in the range of 0.7-1.3 GeV) for further hadronic shower multiplication and k is approximately -0.2.

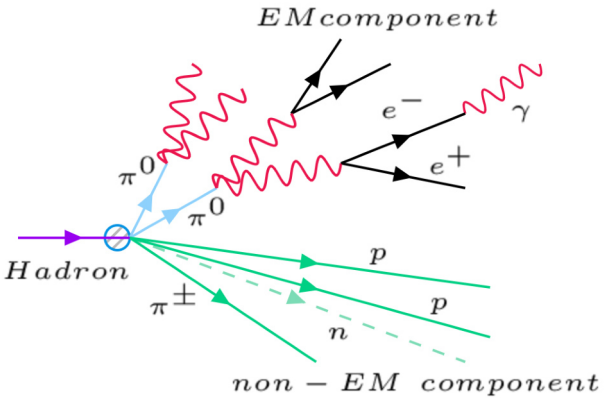


Figure 3.10: *The schematic of hadronic shower development.*

- **The non-EM component**. It includes all other processes through strong interaction to lose energy. As mentioned before calorimeters are blind to a fraction of energy which belongs to the non-EM component as shown in Fig. 3.10, the so called invisible energy. In addition, this invisible energy has very large fluctuations from event to event. The fraction of non-EM component energy included in the shower energy is referred to as f_h.

The large fluctuations of the two components of hadronic shower are behind the difficulties of improving the performance of practical hadronic calorimeter.

The hadronic shower development can be described by the parameter **Nuclear interaction length**, which is defined as the distance λ_I after which the hadron intensity is reduced to the $1/e$ of the original one as follows:

$$N = N_0 e^{-\frac{x}{\lambda_I}},$$

where x is the distance of an incoming particle passed. The λ_I is found to be inversely proportional to the total cross-section of strong interactions with atomic nucleus, giving

$$\lambda_I = \frac{A}{N_A \rho \sigma_t},$$

where ρ is the density of materials. The nuclear interaction length of some widely used materials are summarized in Tab. 3.5.

Material	Z	Density (g/cm^3)	λ_I (g/cm^2)	λ_I (cm)
Fe	26	7,87	132.1	16.77
Copper	29	8.96	137.3	15.32
Tungsten	74	19,30	191.9	9.946
lead	82	11.35	199.6	17.59
Uranium	92	18.95	209.0	11.03

Table 3.5: *The nuclear interaction length of different widely used materials for HCAL [42].*

3.2.2 The hadronic calorimeter (HCAL)

In high-energy collider experiments, the HCAL is an essential part of the detection systems. It is used to measure the hadronic shower energy and it also improve the particle identification. Unlike ECAL, almost all of HCALs used in high energy collider experiments are sampling calorimeters with a classic sandwich structure consisting of thick absorbers and thin active layers. Because of the large fluctuations of the EM component and invisible energy existing in hadronic showers, the performance of HCAL are worse than ECAL. Typically, we define the response of HCAL as the efficiency for hadronic shower energy deposited converting to calorimeter signals. Thus the "h" represents the response for purely hadronic energy converting to calorimeter signals and "e" is the response for EM energy converting to the HCAL signals. Now, considering a hadonic shower induced by a charged pion with energy E. The calorimeter signals E_π can be given,

$$\begin{aligned} < E_\pi > &= e < f_{EM} > E + h < f_h > E \\ &= e(< f_{EM} > +(1- < h_{EM} >)\frac{h}{e})E \end{aligned}$$

Thus the ratio of relative response between hadronic and electromagnetic showers , given

$$< \frac{E_\pi}{E_{EM}} >= \frac{\pi}{e} =< f_{EM} > +(1- < f_{EM} >)\frac{h}{e}$$

Thus, if $\frac{h}{e}$ =1, the HCAL is referred to as compensated calorimeter. If $\frac{h}{e} \neq 1$ (non-compensating), the fluctuations of f_{EM} will make contributions to HCAL energy resolution. The energy resolution of HCAL can be described

$$\frac{\sigma}{E} = \frac{a1}{\sqrt{E}} \oplus a2[(\frac{E}{E_0})^l]$$

where a_2 is the absolute value of (1 - $\frac{h}{e}$) and l is a constant [43]. Typically, the HCAL resolution will be given:

$$\frac{\sigma}{E} \gtrsim \frac{40\%}{\sqrt{E(GeV)}}.$$

The performance of sampling HCAL using different technologies are summarized in Tab. 3.6.

Calorimeter	Absorber	Active medium	Resolution	$< \frac{e}{h} >$
HELIOS	Uranium	Scintillator	$\frac{34\%}{\sqrt{E}}$	1.11
WA80	Uranium	Scintillator	$\frac{67\%}{\sqrt{E}}$	1.12
ZEUS FCAL	Uranium	Scintillator	$\frac{35\%}{\sqrt{E}}$	0.97
SPACAL	Pb	Scintillator	$\frac{30\%}{\sqrt{E}}$	1.15
D∅	Uranium	Liquid Arogon	$\frac{44\%}{\sqrt{E}}$	1.08

Table 3.6: *The resolution of different sampling HCAL with near-compensating [9].*

3.3 Summary

Calorimeter systems are important part of high-energy experiments. The interaction between particles and matter is the basics for understanding calorimeter. There two types of interactions: electromagnetic interaction and strong interaction which generate electromagnetic shower and hadroic shower respectively. To measure the two kinds of shower energy, the two categories calorimeter: ECAL and HCAL are studied. Due to the larger complexity of hadronic shower than electromagnetic one, the resolution of HCAL is commonly worse than ECAL.

CHAPTER 4

The high granularuty Semi-Digital Hadronic CALorimeter (SDHCAL)

For future high energy leptonic collider projects such ILC, CEPC and FCC-ee, obtaining excellent Jet Energy Resolution (JER) is very important for their physics study performance. In order to achieve this purpose, different calorimetry technology are pursued. Particle Flow Algorithms (PFA) provides one of the most satisfying approaches to reach this goal. On the other side, PFA encounters the "double counting" problem which is caused by the wrong assignment of shower components between charged particles and neutral ones. In order to alleviate such problem, high-granularity calorimeters are studied to improve the showers spatial separation capability. The high granularity SDHCAL concept is one of these investigated technologies. To demonstrate this concept, the SDHCAL prototype which is the first technological one among the CALICE family of high granularity calorimeters, has been built and exposed to different particle beams in tests at PS and SPS of CERN. This chapter will give a description of the SDHCAL prototype and its different sub-systems. Results of beam tests will also be presented.

Content

4.1 The SDHCAL technological prototype

For future high-energy collider projects, one of the most important physics cases is to precisely measure the Higgs boson properties. Hence, the precision measurement of jet energy is the one of the key factors of the success of these projects. The **Particle Flow Algorithms (PFA)** is one of the most attractive concepts. For successful and efficient application of PFA technology in reality, the calorimetry system should be equipped with high granularity to achieve excellent spatial resolution and tracking capability leading to the so-called imaging calorimetry. The CALICE collaboration developed a series of calorimeters with high granualrity and the SDHCAL prototype is first member of a new family of these PFA-oriented calorimeterer called technological. These technological prototypes try to fulfill the requirements of compactness, robustness and low power-consumption features that are needed in the future collider experiments in addition to the high granularity one to ensure their full success.

The SDHCAL prototype is a sampling HCAL with classic sandwich structure. It comprises 48 active layers, each of them equipped with a 1 m $\times$ 1 m Glass-based Resistive Plate Chamber (GRPC) as active medium and an Active Sensor Unit (ASU) of the same size hosting on one face (the one in contact with the GRPC), pickup pads of 1 cm $\times$ 1 cm and 144 HARDROC2 ASICs [45] on the other face. The GRPC and the ASU are assembled within a cassette made of two stainless steel plates, 2.5 mm thick each. The 48 cassettes are inserted in a self-supporting mechanical structure made of 49 plates, 15 mm thick each, of the same material as the cassettes, bringing the total absorber thickness to 20 mm per layer. The empty space between two consecutive plates is 13 mm to allow the insertion of one cassette of 11 mm thickness. The HARDROC2 ASIC has 64 channels to read out 64 pickup pads. Each channel has three parallel digital circuits whose parameters can be configured to provide 2-bit encoded information indicating if the charge seen by each pad has passed any of the three different thresholds associated to each digital circuit. This multi-threshold readout is proposed to improve on the energy reconstruction of hadronic showers at high energy (> 30 GeV) with respect to the simple binary readout mode which will be discussed after. The SDHCAL has more than 440k readout channels to read out the information provided by the pickup pads. The important amounts of data collected from such high-granularity prototype requires a sophisticated acquisition system.

A typical hadronic shower in SDHCAL prototype is shown in Fig. 4.1.

4.1.1 The Glass-based Resistive Plate Chamber (GRPC) system

The RPC is a type of gaseous detector which was developed by R.Cardarelli and R. Santonico in the 1980's [9]. The classic RPC structure is shown in Fig. 4.2. It consists of two parallel high-resistivity plates (typically is 10^9-$10^{13}\Omega$·cm.) as the anode and cathode respectively. The two resistive plates require to have an excellent flat surface and excellent planarity which help to keep the uniform electric field inside the chamber. Materials like glass have excellent smooth surface and flatness. Oiled Bakelite can also fulfill the requirements. The insulating spacers are used to form the stable gas gap between the two electrodes which can prevent against the gap deformation caused by the strong electrostatic force once high voltage is applied. Generally, the gap is about few mm with

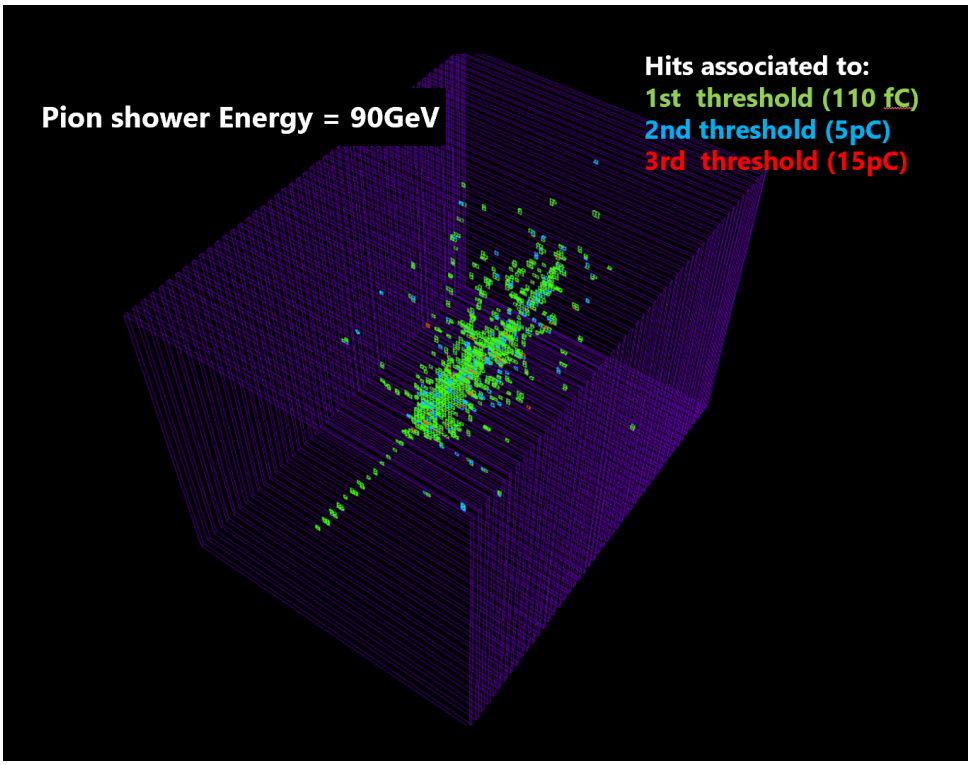


Figure 4.1: *The event display for typical hadronic shower (pion, 90 GeV) in the SDHCAL. The green, blue and red pixels represent hits associated with first, second and third threshold respectively.*

precision 5-10μm and sealed by an insulating frame. The gas gap is typically filled by a mixed gas which allows on the one hand the ionization and on the the other hand limits the avalanche transverse extension around the charged particle trajectory.
RPC can be operated with very high efficiency (>95%), excellent time resolution (20 ps - 1ns depending on the number of gas gaps) and spatial resolution which is determined principally by the size of pickup strips or pixels which are widely used to read out RPC signals. Therefore, it is commonly used in collider experiments (e.g., ATLAS and CMS), cosmic rays detectors and neutrino arrays.

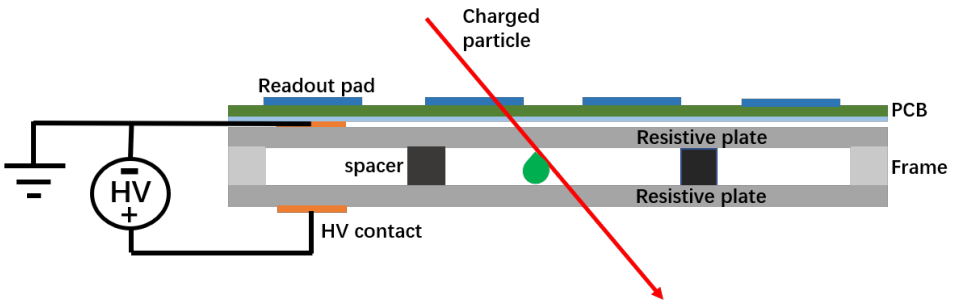


Figure 4.2: *The cross-section schematic of RPC (not scaled).*

GRPC detector is a part of the SDHCAL active medium as shown Fig. 4.3. It consists of two glass plates which has a thickness of 0.7 mm and 1.1 mm making the anode and the

cathode respectively. The resistive plates are painted by a uniform conductive graphite coating (0.5-2 M/□). The two resistive plates are separated with a gap of 1.2 mm which is maintained constant (to maintain the uniform electric field between the electrodes) by the use of insulating concentric cylindrical spacers. The spacer size and their number is optimized to achieve a negligible dead zone. The gap is sealed by an insulating frame and subsequently filled by a mixture of three different gases mixture: TetraFluoroEthane (TFE) (93%), CO2(5%) , SF6(2%). The first gas provides the avalanche electrons and ions, the second and third gas are responsible for quenching UV photons and electrons respectively to limit the size of avalanche development around the charged particle passage in the detector. The important steps for constructing GRPC are discussed below.

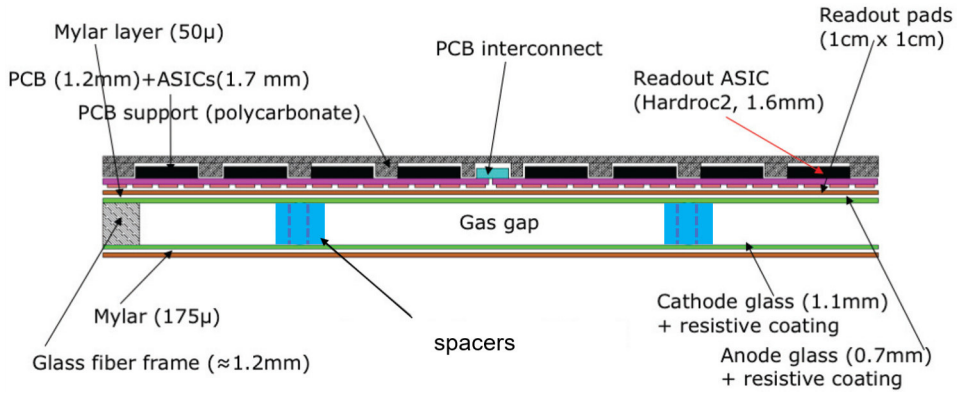


Figure 4.3: *The schematic of GRPC in the SDHCAL (not scaled).*

- **Resistive coating** In order to choose the suitable resistive coating, many different types of graphite-based products were tested and we finally chose a product made of two components. This product is found to be suitable for silk screen technique (easy for mass production), has a short period to reach a stable surface resistivity. In addition, it does not have, like many of the tested paintings, the migration problem of painting molecules located close to the the high voltage point contacts that results in losing the high voltage contact with the coating after a few weeks of operation. By changing the mixing ratio between the two coating components, we can adjust the surface resistivity and after optimization, the surface resistivity in the range 1-2MΩ/□ was reliably reproduced in the lab.

- **Spacer distribution** It is very important to keep constant the electric field between two electrodes to prevent large gain deviations. Thus, it is necessary to keep the the same gas gap in the whole chamber. But because of the gravity and electrostatic force, the two $1m^2$ plates have tendency to deform and thus need suitable spacers distribution to avoid the deformation. Therefore, the finite element analysis method is employed to optimize the spacer distribution and we finally achieved gas gap uniformity with less than 44 μm deformation when two closest spacers have a distance of 10 cm. The non-uniformity in this level is acceptable for gain fluctuations due to the gas gap deformation. The spacers distribution is shown in Fig. 4.4.

- **Gas distribution** To improve the gas distribution in RPC, we use a simple but efficient gas channelling-based system as the one shown in Fig. 4.4. Orange color

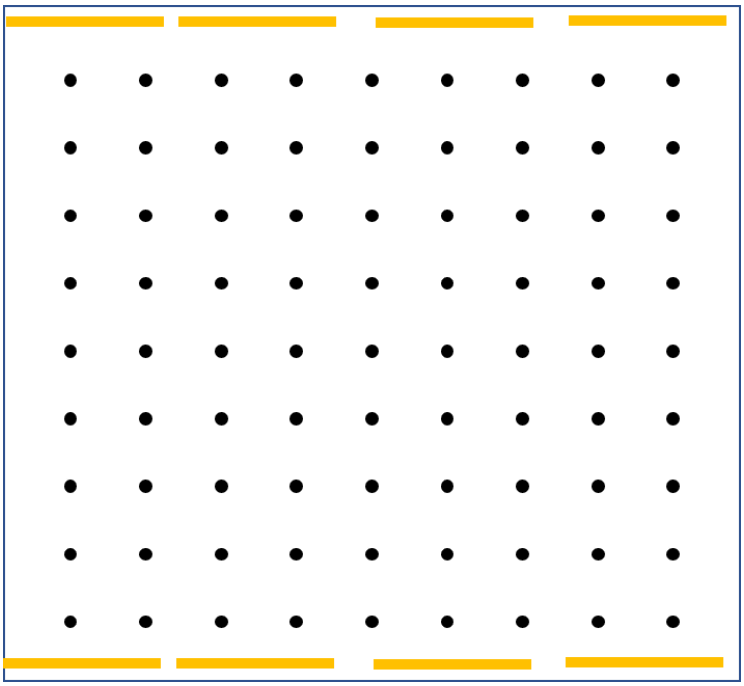


Figure 4.4: *The schematic of spacers distribution and gas channelling system in the GRPC (not scaled), the black points are spacers and short lines with orange color are gas pipes.*

lines in the figure represent the 1.2 mm diameter tubes with holes in between allowing the gas to be evenly distributed in the whole chamber from one side and then collected from the other side. Based on this design, simulation study confirmed the uniform gas distribution in our chambers.

In the gas mixture of GRPC, the SF_6 and $C_2H_2F_4$ are green house gases, thus they are not environment-friendly. The Global Warming Potential (GWP) are used as the measurement of of the green house gas absorbing heat in atmosphere compared with carbon dioxide (CO_2) . The GWP for CO_2 is 1. Therefore, the GWP of CO_2 is used as the unit of GWP. The GWP for SF_6 and $C_2H_2F_4$ are 22200 and 1430 respectively, as indicated in Fig. 4.5. Because of huge GWP of SF_6 and $C_2H_2F_4$, the study of other gases are started to solve this. For example, the new refrigerant gas Hydro-Fluoro-Olefin-1234yf (HFO-1234yf) is investigated to replace $C_2H_2F_4$, which is found to have very small GWP (≈ 4) as shown in Fig. 4.5. Moreover, in order to reduce the gas consumption and cost, we developed a new gas re-circulation system with the help of a gas group of CERN.

4.1.2 The electronic readout and active sensor unit

4.1.2.1 The electronic readout of GRPC

Once a charged particle passes the GRPC and generates an avalanche, the latter induces a signal in the pickup pads. As indicated before these pads have a size of 1 cm $\times$ 1 cm each as shown in Fig. 4.6 and embedded on the side facing GRPC as shown in Fig. 4.3. The induced signals are subsequently collected by ASICs (Application-specific Integrated

Figure 4.5: *The GWP of SF_6, $C_2H_2F_4$ and $C_3H_2F_4$ (HFO-1234yf) . The GWP for CO_2 is one.*

Circuit) which are designed (using SiGe 0.35μm process) for SDHCAL prototype and future ILC colldier projects. These ASICs are referred to as HARDROC (HAdronic Rpc Detector Readout Chip). One such ASIC is shown in Fig. 4.7. When the SDHCAL was built in 2012, we had two versions: HARDROC1 and HARDROC2. In the SDHCAL, we use the HARDROC2 because it offered multi-threshold readout while HARDORC1 offered only one.

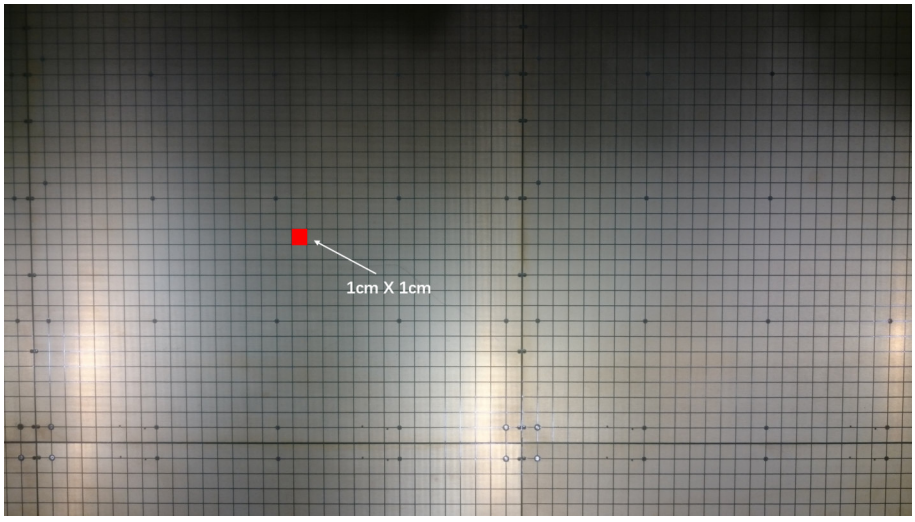


Figure 4.6: *The readout pads of GRPC with size 1cm×1cm.*

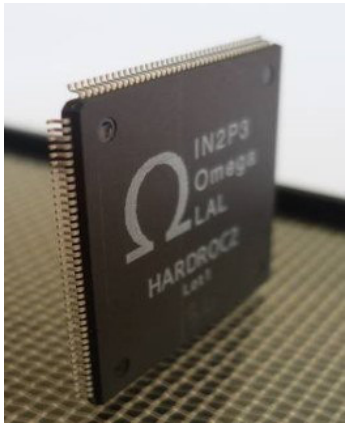


Figure 4.7: *The HARDROC2 designed for SDHCAL. It was produced by using 0.35μm process and 1.4 mm package with 160 pins.*

The one of the most important characteristics of HARDROC2 is that it has three threshold readout function which is behind the words "semi-digital" of the prototype name. This ASIC has 64 channels corresponding to the 64 readout pads of GRPC and each channel consists of a current pre-amplifier followed by three fast shapers (FSB) for each a low-offset discriminator is associated as shown in Fig. 4.8. The signal collected by the pickup pads subsequently are amplified by the pre-amplifier, then it goes through the shaping process in the FSB and finally enters in the discriminator. The three threshold values associated with three discriminators are set to low, medium and high levels corresponding to Vth0, Vth1 and Vth2 as shown in Fig. 4.8. The three reference levels are converted from the three thresholds by using the Digital Analog Converters (DAC). Furthermore, in order to precisely measure the conversion ratio between the injected charge and DAC unit value, two scans were performed for each channel. The fist scan consists of injecting a charge (Q_{inj}) and varying the threshold value and then determining the DAC value corresponding to the inflection point of the efficiency curve (called S-Curve). The second scan is a repetition of the previous with different charge values. These two scans provide the charge-DAQ conversion for the three FSBs. As an example, one of the charge-DAC response of FSB0 for one of the 64 channels is shown in Fig. 4.9.

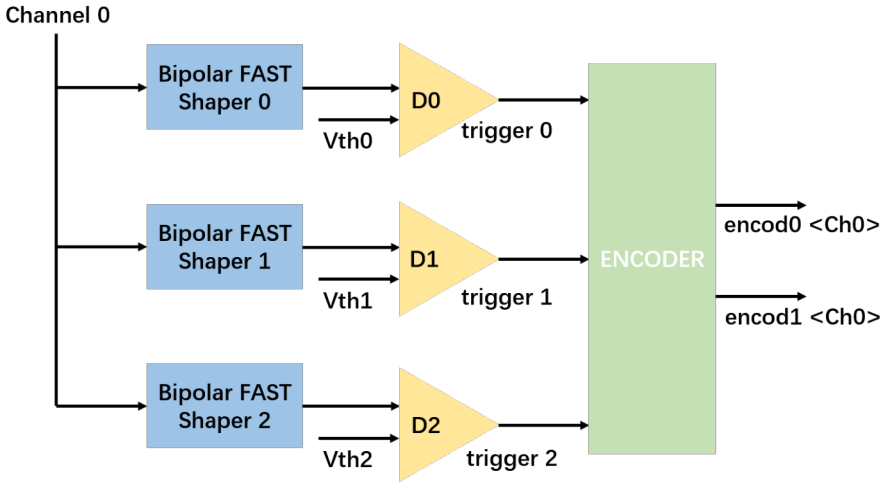


Figure 4.8: *The simple schematic of semi-digital logic. The D0, D1 and D2 are three discriminators which have three different thresholds Vth0, Vth1 and Vth2 associated with 110fC, 1pC and 15pC respectively.*

4.1.2.2 The Active Sensor Unit (ASU)

The ASU is a PCB hosting the electronic components which provides charge collection, ASICs hosting, commands transmission to the ASICs and communication between the ASICs and the data acquisition board. In general, an active layer of SDHCAL consists of six ASUs which fully cover the GRPC with 1 m×1 m. There is an important technological challenge in building a PCB of up to 1m length with good planarity to ensure a homogeneous contact of the pickup pads with the GRPC detector. In order to balance the complexity of combining the tens of small size ASUs to 1 m^2 and the difficulty of producing single ASU for the same area, we adopted a compromising solution, to divide

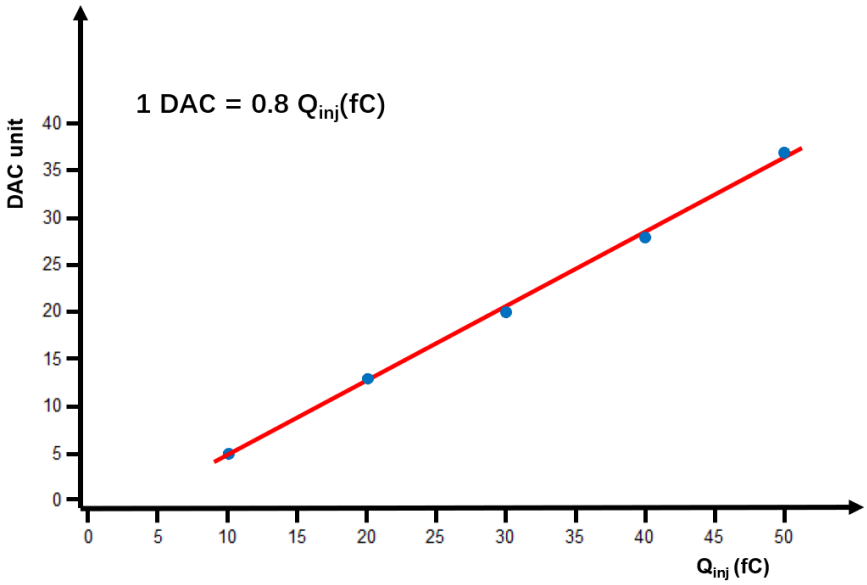


Figure 4.9: *The function between injected charge Q_{inj} and DAC unit value for FSB0. DAC unit value is determined by the 50% trigger efficiency of the S-curve.*

1 m^2 layer to six ASUs as shown in Fig .4.10. The ASU were assembled first two by two to form a slab of 33.3 cm $\times$ 100 cm. Each slab hosts 48 ASICs on the one side of PCB and thus host 48 $\times$ 64 pickup pads on the another side facing the GRPC. The three slabs were then soldered together to ensure the same grounding.

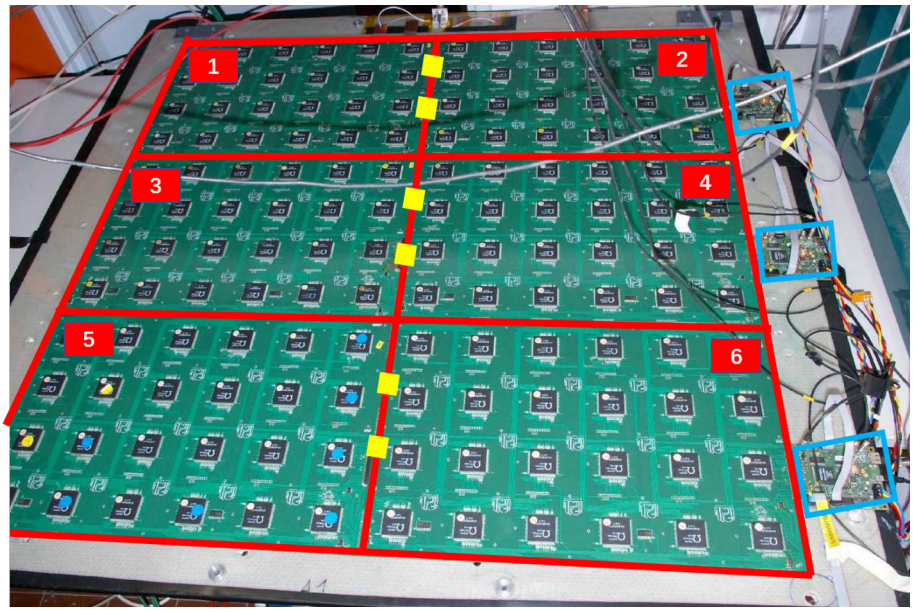


Figure 4.10: *The six ASUs are used for 1 m^2 GRPC detector. In the middle, there are six bridges (yellow) for connecting ASUs. In the right side, the three plugins (blue) are DIFs.*

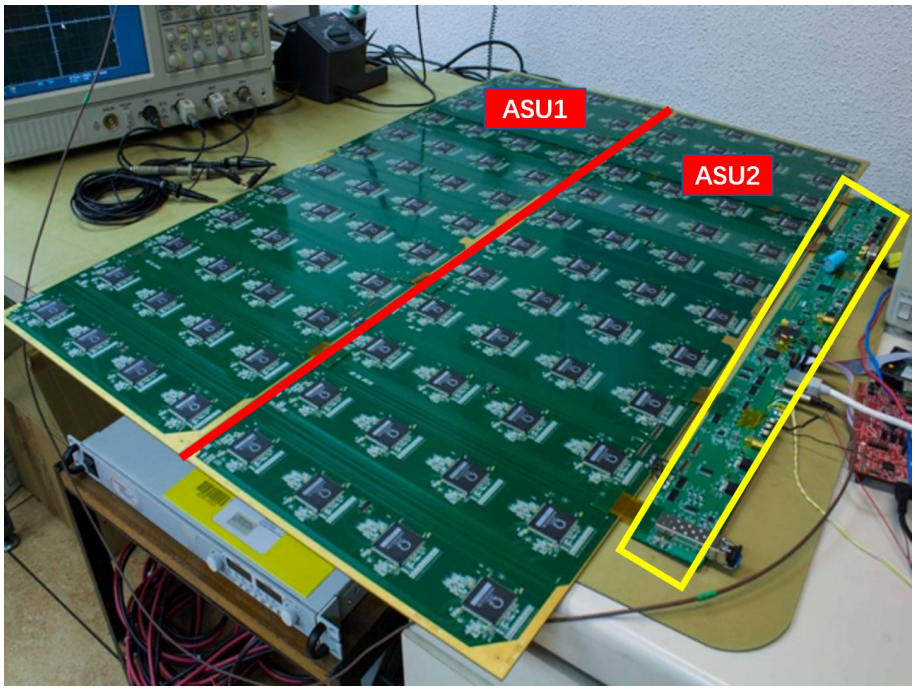


Figure 4.11: *The two new ASUs with size 1×0.33 m^2. A new version of DIF is also presented (yellow box).*

4.1.3 The SDHCAL data acquisition system

The data acquisition system plays an important role in the SDHCAL technological prototype which consists of both hardware and software sub-systems. The software sub-systems including three aspects: providing effective access based on C++ framework for low level hardware, the configuration database (SQL database hosted by Oracle server) for hardware and data collection as well as data quality online monitoring. For the hardware sub-system, the one of the most important component is the Detector InterFace (DIF) card which connects each SLAB (a third of one detector layer) to off-detector DAQ services. The primary features are listed below:

- **ASIC initialization and configurations.** After receiving the initialization and configuration commands from the off-detector DAQ system, the DIF will transfer these commands to the ASICs and at the same time check whether such process is successfully executed.

- **Digital readout.** Once the ASICs are successfully initialized and configured, the DAQ can be operated either in a trigger mode (that can be used for testing single chambers using a PMT-Scintillator system) or in triggerless mode for the SDHCAL prototype running as well as in future leptonic collider experiments.

- **Power-pulsing mode.** Because beams bunch crossing is expected to continue 1 ms per 200 ms in future ILC beam cycles, thus the analog readout part of the ASICs can be deactivated at this relatively long inactive time to decrease the power consumption and reduce thus the heat accumulation which eliminate the need of an active

cooling system that may result in worse PFA performance (due to a reduction of compactness and increased dead zones).

- **Data format.** In addition to the readout data from ASICs that include the ASIC number, the fired channels number and the associated threshold as well as the counter indicating the time slot in which the the channel was fired, the DIF will add extra information to the data stream for facilitating data analysis.

4.2 The performance of SDHCAL technologcial prototype

4.2.1 Physics event reconstrcution

To test the performance of SDHCAL technological prototype, it was exposed to beams many times. However, the data collected in the periods of beam test do not only include that related the physics events (such as those produced by pions, electrons and muons) but also the intrinsic noise produced by the detectors as well as the electronic system. Thus it is necessary to adopt suitable time algorithm to reconstruct the true physics event, based on the time information recorded by the time counter embedded in DIFs. This time counter will store the time difference of every fired ASIC channel (referred to as **hit**, here after and used as the fundamental constituent of SDCHAL events) with respect to the begin of the data acquisition and it increases with a time step of 200 ns. Thus, all hits recorded in the same time slot will have the same identity number called Bunch crossing ID (BCID). Actually, as mentioned before, apart from the BCID information, the hit information includes the ChannelID, AsicID and DifID which can be subsequently used to find the hit's position information and its corresponding threshold (1st, 2d or 3rd).

To reconstruct physics event from raw hits collection, a time clustering method is adopted. It consists of two steps: (1) Selecting all time slot candidates which have a number of hits larger than K_{hit}; (2) For every selected time slot as a candidate, we choose the time window of ± 200 ns (two neighbouring time slots) around the selected one and then we merge the hits belonging to these three time slots to reconstruct the physics event as shown Fig. 4.12. In addition, after time clustering algorithm applied, if hits of two events which share the common time slot, such events will be abandoned. Actually, because beam intensity is reduced to suitable level during test beam period given the limited rate capability of GRPC, this problem is not observed. In order to determine the value of K_{hit}, the hits-time histogram is built as shown in Fig. 4.13. Using the value $K_{hit} = 7$ was found to fulfill our requirements of rejecting almost all of the intrinsic noise and only losing a negligible part of physics events above a few GeV at the same time. For all the collected physics event candidates, there is a little fraction of events that are not related neither to beam particles nor cosmic rays crossing the detector. It consists of two types:

- **Coherent noises.** This type of noise is often observed with many hits appearing in the electronics of the same active layer or a very few of them not necessarily close by. The reason for this events could be explained by the imperfect electronic grounding

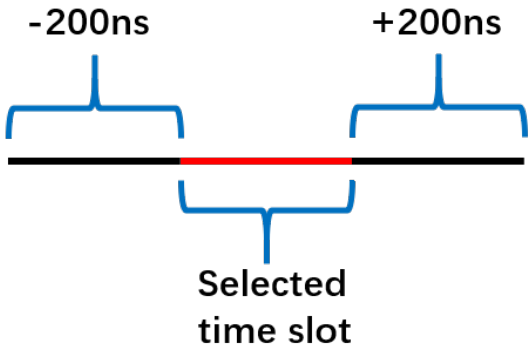


Figure 4.12: *The schematic for time clustering. The two neighbouring time slots are merged together with selected time slot to use for physics event reconstruction.*

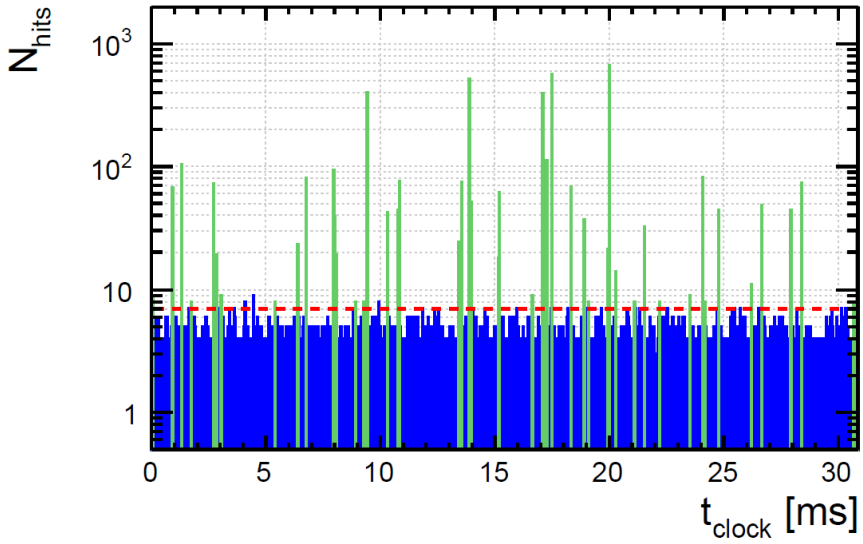


Figure 4.13: *The hits - time histograms with time slot 200 ns. The green lines represent the physics event candidates and the blue is from intrinsic noise. The K_{hit} is presented by the red dashed line. This figure is cited from Ref. [46].*

of prototype. This is confirmed by the reduced frequency of such events occurrence after the grounding quality of SDHCAL prototype was improved.

- **Intrinsic noises.** The intrinsic noises of prototype are caused by random avalanches introduced by electrons extracted from the cathode by the electric field (field effect) which increases when the temperature and high voltage values increase. The rate estimation of such a noise was found to be approximately 0.35 hit / 200 ns on average in the operation condition of the SDHCAL at the SPS and PS beamlines at CERN. Thus intrinsic noises hits of every physics events (including 3 time slots) is about 3 $\times$ 0.35 = 1.05 hits.

4.2.2 The efficiency and multiplicity of GRPC

To verify the performance of GRPC in SDHCAL prototype, we can measure two fundamental quantities, the efficiency (ε) and multiplicity (μ). The efficiency describes the possibility for GRPC successfully detecting a signal when a charged particle passing the detector. The multiplicity provides the description of the average number of hits produced when a single charged particle crosses the detector. Both quantities are measured by using the beam muon samples since high energy muons are able to cross the whole SDHCAL as shown in Fig. 4.14. For reconstructing muon tracks, we only select events whose total number of hits is found to be less than 200. This value is suggested by that a muon pass whole detector leaving at most 200 hits when it fires 4 pads in every layer corresponding to $4 \times 48 \approx 200$ hits, in extreme situation as shown Fig. 4.15.

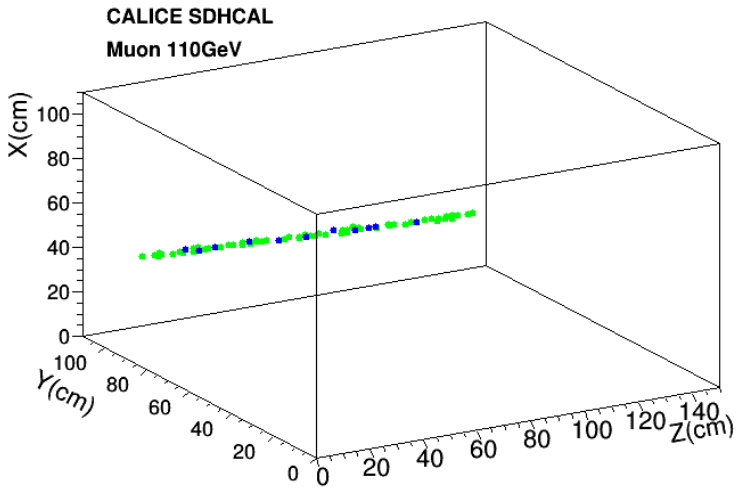


Figure 4.14: *The event display of a beam muon event and its track in the SDHCAL. The green and blue represent hits associated with first and second thresholds respectively. The hits associated with the third threshold do not exist in this event.*

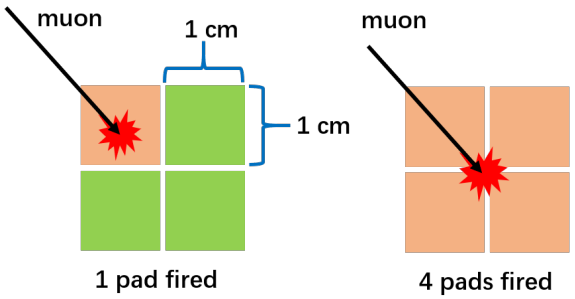


Figure 4.15: *Left: in normal case, a muon crosses one pad and only subsequently fires the crossed pad (light red), others are not fired (green). Right: in extreme case, a muon crosses the center space of 4 neighbouring pads and thus fires 4 pads (by induction). The red explosion mark represent the avalanche generated by the passage of a muon.*

4.2.2.1 Track reconstruction

For a given active layer, a so called nearest-neighbour clustering algorithm are adopted to combine all hits of this layer to clusters. This algorithm principle is that if two hits share a common edge, the two hits subsequently are regrouped to the same cluster as shown in Fig. 4.16. The cluster position (X_c, Y_c) is determined as the unweighted average of all hits position, given:

$$X_c = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$Y_c = \frac{\sum_{i=1}^{n} y_i}{n}$$

where n is the value of total number of hit in the cluster and x_i , y_i are position of hit i.

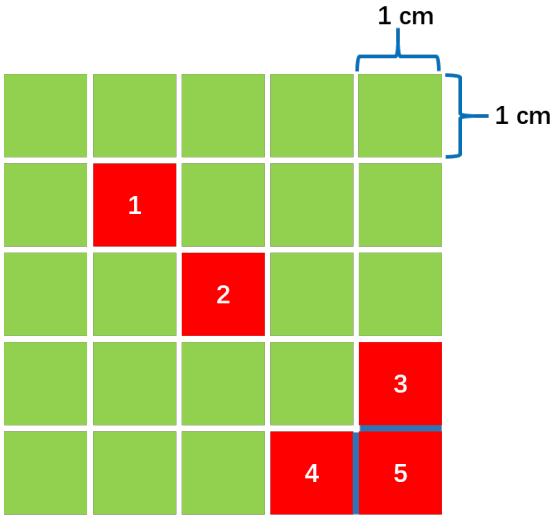


Figure 4.16: *The green block are not fired pads. The red block are fired pads (hits). The hit-3, hit4 and hit-5 share two edges (blue), thus they are combined to a same cluster. The hit-1 and hit-2 do not share any edge, thus they belong to two different clusters.*

For track reconstruction, a Hough Transform (HT) method is implemented in the cluster level. The basic idea for this method is that :

- Assume that we want to detect a straight line among many points (z_i, x_i) located in the Cartesian coordinates $z - x$ plane. Each of point in the (z, x) plane then can be transformed into a corresponding curve in the polar plane (ρ, θ) following the transformation formula:

$$\rho_i = z_i \cos\theta_i + x_i \sin\theta_i \quad for\ point\ i,$$

- After transformation process, the curve(ρ, θ) set corresponding to the points of the same straight line will cross a common node (ρ_0, θ_0). Therefore, we can find the points in the same straight line in the Cartesian coordinates $z - x$ plane by searching the node (ρ_0, θ_0) in the plane (ρ, θ).

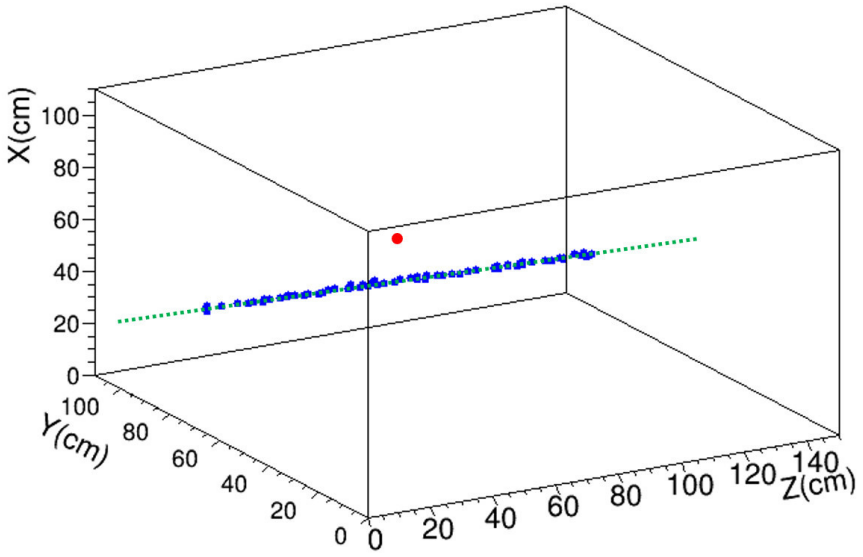


Figure 4.17: *The event display of a muon events. The read point is a noisy hit. The blue points are hits of muon track. The straight dashed line is obtained by fitting the hits selected by HT method. The red outlier is rejected by the HT process.*

Let's consider a muon event as the one shown in Fig. 4.17. The muon track is a 3D straight line which has a unique projection in the (z, x) plane and (z, y) plane giving two 2D-straight lines. Therefore, for any hit (x_i, y_i, z_i) of the muon event can transform two different polar planes (ρ^{zx}, θ^{zx}) and (ρ^{zy}, θ^{zy}) respectively, following

$$\rho_i^{zx} = z_i \cos\theta_i^{zx} + x_i \sin\theta_i^{zx} \quad for\ hit\ (x_i, y_i, z_i)$$
$$\rho_i^{zy} = z_i \cos\theta_i^{zy} + y_i \sin\theta_i^{zy} \quad for\ hit\ (x_i, y_i, z_i)$$

These two functions between ρ *and* θ can discretize to a 2D histogram respectively as shown in Fig. 4.18. Then the node (ρ_0, θ_0) can be found for each of the two corresponding histograms. It corresponds the bin with the maximum content. Finally, the common points existing in both 2D lines are selected to reconstruct the muon track which can be subsequently fitted by a linear function . Thanks to HT method which efficiently eliminates the outliers, the straight line can be more properly detected compared with the linear regression method.

For measuring the efficiency of a layer in SDHCAL prototype, the hits of each layer are first combined into clusters by a clustering algorithm. Then all clusters except those of the layer under study are used to reconstruct the track (using HT method) to avoid the bias introduced by using the layer itself. Then the interacting point of the studied layer is determined with the fitted track is determined. Finally, the efficiency of this layer is estimated by calculating the ratio between the number of tracks for which at least one hit is found around the intersection point within a distance less than 3 cm on and the total number of tracks. For hit multiplicity, it is estimated by counting how many number of hits in the cluster including a hit with a minimum distance with respect to the track.

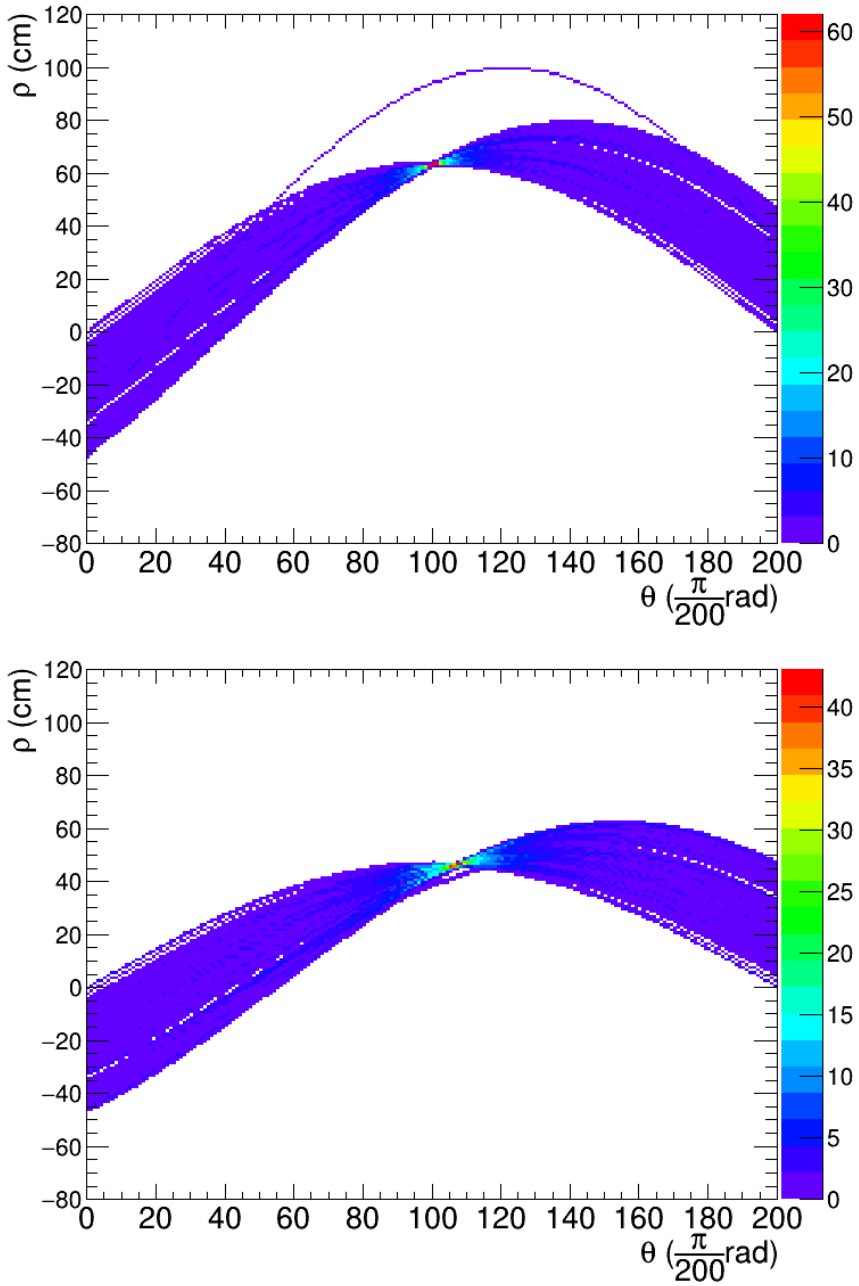


Figure 4.18: *Top: The hits projected to (z, x) plane are transformed to polar planes (ρ^{zx}, θ^{zx}). The one outlier (the red point) is detected. Bottom: The hits projected to (z, y) plane are transformed to polar planes (ρ^{zy}, θ^{zy}). The unit of x-axis for both figures are $\frac{\pi}{200}$ rad, thus 200 bins is exactly the one complete period for all curves . The curves of two histograms both through the bin with maximum content corresponds to the hits in the muon tracks.*

4.2.2.2 High voltage scan for GRPC efficiency study

It is important to measure GRPC efficiency under different values of applied high voltage to determine the optimal operational value that associate high efficiency and reduced

noise. Figure. 4.19 shows the relation between efficiency and high voltage in the scanning range [6200 V, 7200 V], for two GRPC layers of SDHCAL prototype. It is clear that the efficiency is almost proportional to high voltage from 6.2 kV to 6.6 kV and reaches a plateau around the 6.8 kV. Thus the best working voltage for GRPC of SDHCAL can be approximately found at range from 6.9 to 7.2 kV.

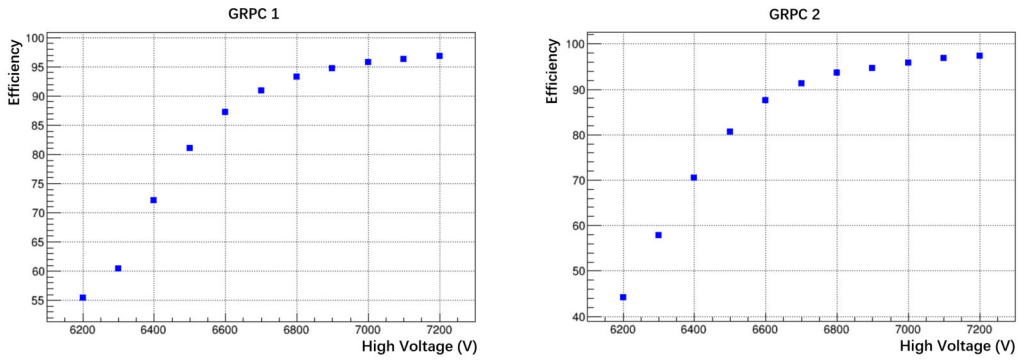


Figure 4.19: *The relation between efficiency and high voltage applied for two GRPC layers of SDHCAL in the 2017 test beam, at H2 of CERN. The plateau of efficiency appears at high voltage around 6.8 kV.*

4.2.2.3 Efficiency and multiplicity of GRPC in the beam test

For the beam test in 2012 at H6 of CERN, the efficiency and multiplicity for each GRPC layer (running at 6.9 kV) is shown in Fig. 4.20. The average efficiency and multiplicity of GRPC are found to be 0.96±0.04 and 1.75±0.16 respectively. The efficiency drop, to about 70% for layer 42, is observed and it is explained by the fact that one slab over the three slabs forming one electronics layer was faulty. In some layers, a relative large fluctuation of multiplicity is also observed which is possibly resulted from the gain fluctuation in some range of these GRPC layers (originating most probably from some non-uniformity of the painting or the gas gap or a tiny accidental distance between the pickup pads and the detectors of these layers).

For the beam test in 2015 at SPS of CERN, the efficiency and multiplicity are also measured for all 48 GRPC layers as shown in Fig .4.21. The layer 1 and layer 34 were not instrumented in this beam test period thus they are not included in Fig. 4.21. These two layers are not taken into account for the efficiency and multiplicity estimation. The efficiency of layer 2 is only about 70%, much lower than the other layers due to the fact that one slab of this layer was faulty, thus 1/3 of electronics controlled by it are absent. The average efficiency of 46 layers is found to be 0.95± 0.04 and the average multiplicity is 1.81±0.19.

Figure 4.22 shows efficiency map of layer 10 and 20. From this figure, we observe a few local areas with very low efficiency (the blue pixels) which result from masked channels to reduce their electronics noise. The regular shadow area (representing relative low efficiency) related to the spacer distribution which can be clearly see in Fig 4.23. The relative low efficiency around the border of the layer caused by the gas channeling pipes and the sealing frame.

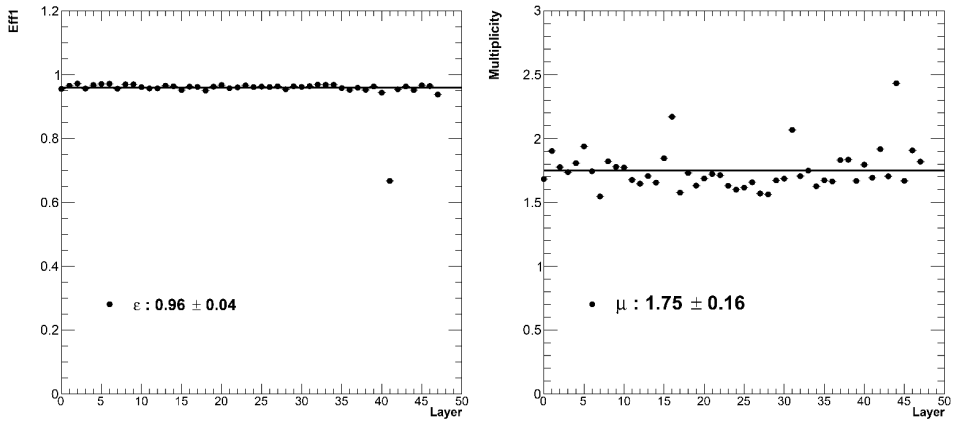


Figure 4.20: *Left: the efficiencies of 48 GRPC layers (LayerID from 0 to 47) of SDHCAL prototype. Right: the multiplicities of 48 GRPC layers of SDHCAL prototype. The results are obtained from the beam test in 2012 (H6 beamline of CERN). The all GRPC layers work at 6.9 kV.*

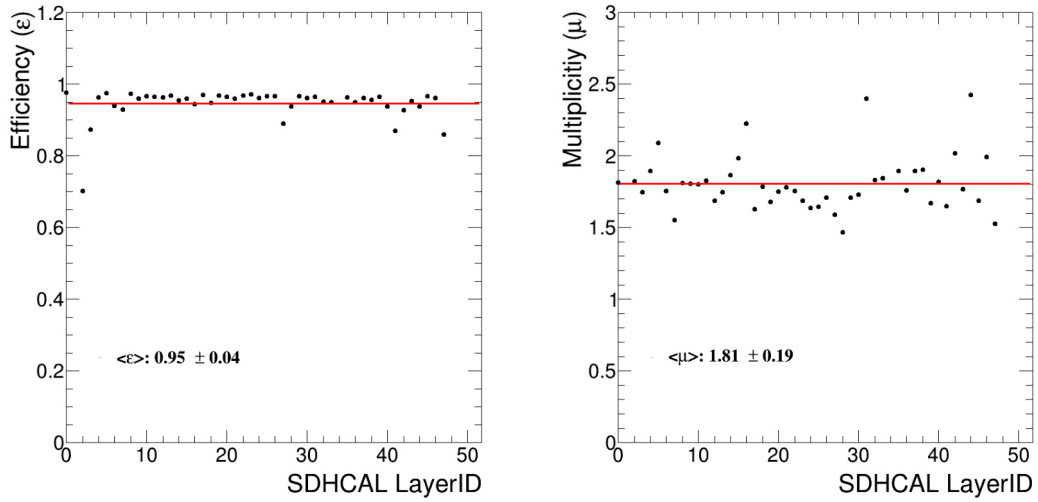


Figure 4.21: *Left: the efficiencies of 48 GRPC layers (LayerID from 0 to 47) of SDHCAL prototype. Right: the multiplicities of 48 GRPC layers of SDHCAL prototype. The results are obtained from the beam test in 2015 at SPS of CERN. The all GRPC layers work at 6.9 kV. The layer 1 and layer 34 are missing because they are out of service during beam test period.*

Figure 4.24 shows the multiplicity map for the two previous layers. The fluctuations of multiplicity are probably caused by non-homogeneity of the detectors resulting either from the resistive coating dispersion or the gas gap height or the distance between the pickup pads and the detector. The high multiplicity areas are strongly correlated with the borders between the six ASUs as shown in Fig. 4.25. The non-uniformity between PCB and GRPC layers for this area can result in induced charge with large fluctuations thus may produce the large multiplicity.

For beam test in 2017, the average efficiency and multiplicity of all available GRPC layers are shown in Fig. 4.26. The efficiency and multiplicity are 0.95±0.03 and 1.90±0.30

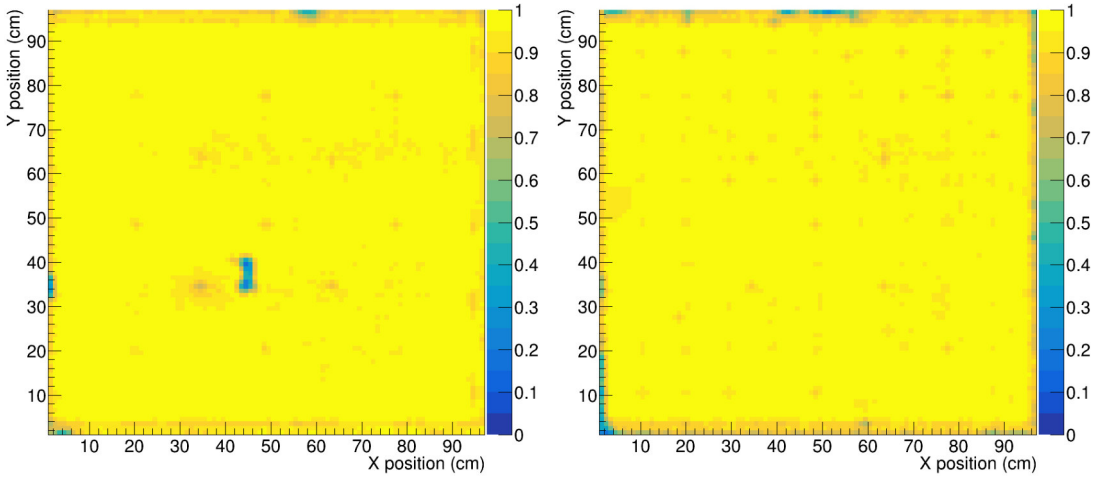


Figure 4.22: *The efficiency map $\epsilon(x, y)$ for GRPC layer 10 (left figure) and 20 (right figure) respectively.*

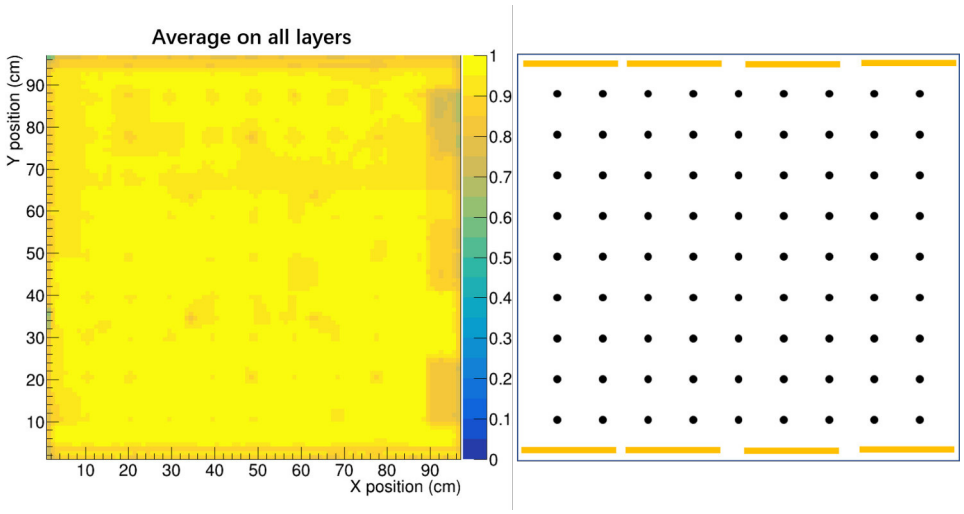


Figure 4.23: *The efficiency map $\epsilon(x, y)$ average on all layers shown in the left figure. The impact on efficiency due to spacers, gas channeling pipes is shown in the right figure. The results are obtained from 2015 beam test.*

respectively. The results are obtained by the measurements of 43 layers. The other 5 layers had some problems which could not be fixed for this measurement thus they are not used. The efficiency and multiplicity maps averaged on all the used layers are shown in Fig. 4.27. Similar phenomena as for the 2015 beam test are also observed.

For beam test in 2018, the average efficiency and multiplicity of all available GRPC layers are shown in Fig. 4.28. In this beam test, 11 GRPC layers of the SDHCAL prototype were absent. They were used in another experiment and among the remaining 37, two layers were faulty. Based on the 35 layers, the average efficiency and multiplicity are 0.89±0.04 and 1.61±0.14 respectively. The relative lower efficiency and multiplicity due to the use of a slightly different gas mixture (CO2 percentage was increased from 5% to

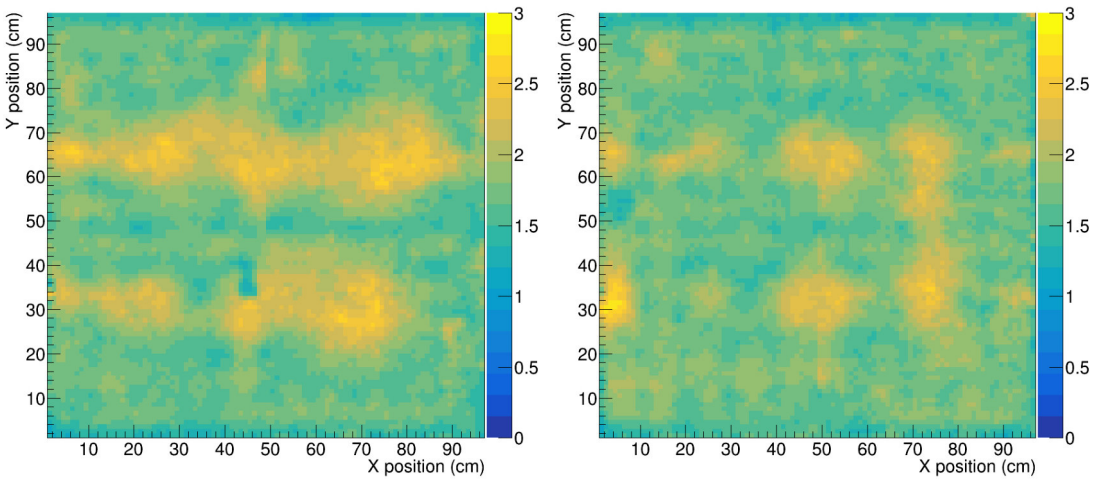


Figure 4.24: *The multiplicity map $\mu(x, y)$ for GRPC layer 10 (left figure) and 20 (right figure) respectively. The results are obtained from 2015 beam test.*

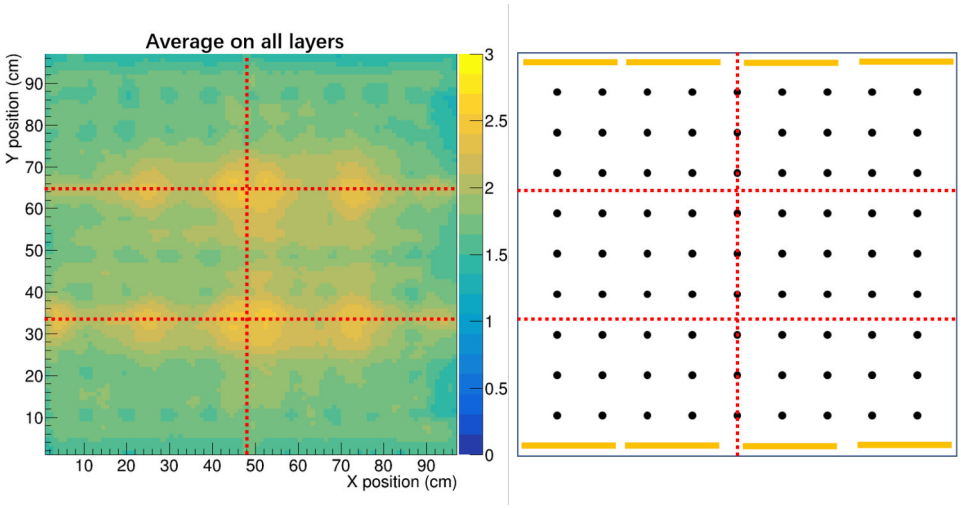


Figure 4.25: *The multiplicity map $\mu(x, y)$ average on all layers shown in the left figure. The non-homogeneity existing in the borders (red dashed lines) between six ASU PCBs has strong impact on multiplicity. The results are obtained from 2015 beam test.*

7.5%). The efficiency and multiplicity maps averaged on all the used layers are shown in Fig. 4.29 .

The results of efficiency and multiplicity measured in 2012, 2015, 2017 and 2018 beam tests are summarized in Tab. 4.1.

Based on the beam test results, the good global performance of GRPC in SDHCAL prototype is confirmed. However, some non-uniformities of the efficiency and multiplicity maps in GRPC plane are also found, which is resulted from the non-homogeneity of detectors. To fix such problems, one way is to look for new technologies capable of producing larger size PCBs which can reduce the number of PCBs used for one layer, avoiding the non-homogeneous connections in the border of ASU PCBs. Another way

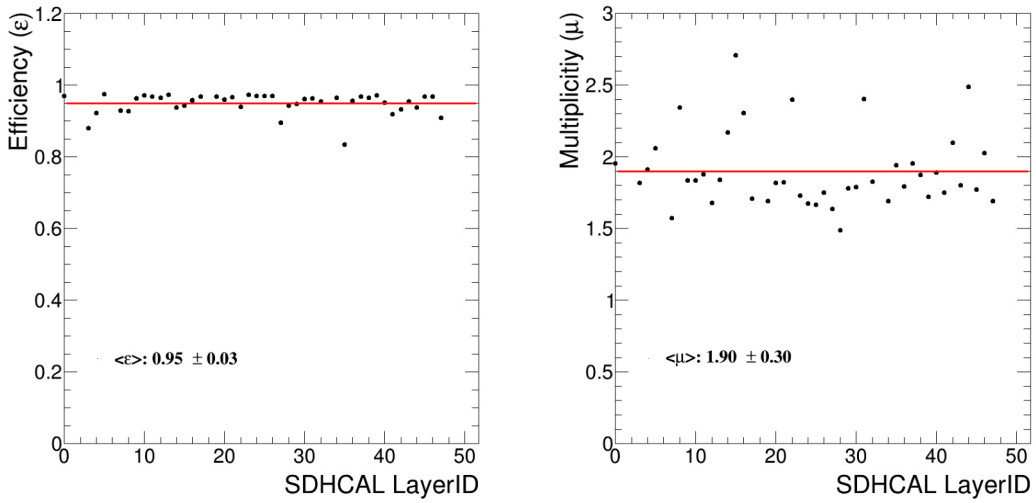


Figure 4.26: *Left: the efficiencies of 43 GRPC layers (five layers are not available in the test beam period in 2017) of SDHCAL prototype. Right: the multiplicities of 43 GRPC layers of SDHCAL prototype. The all GRPC layers work at 6.7 kV.*

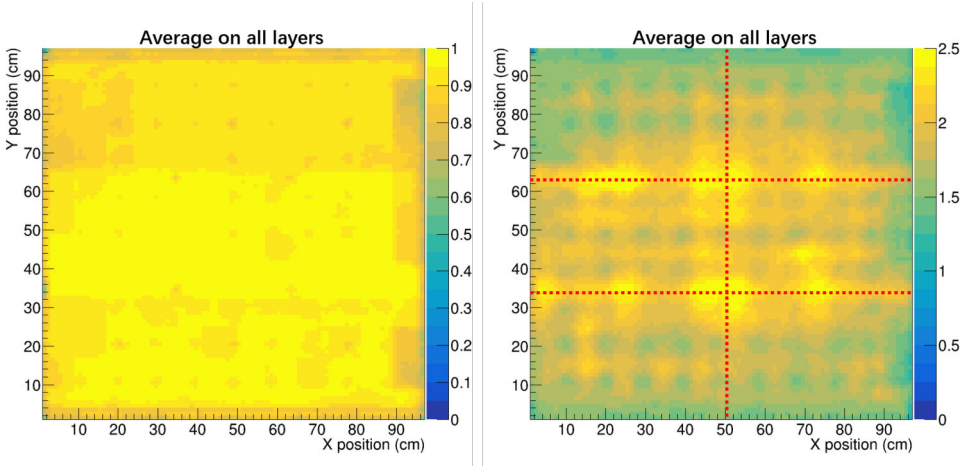


Figure 4.27: *Left: the efficiency map $\epsilon(x, y)$. Right: the multiplicity map $\mu(x, y)$. Both efficiency and multiplicity maps are averaged on 43 layers. The red dashed lines are related to the borders of six ASU PCBs. The results are obtained from 2017 beam test.*

Year	Voltage (kV)	efficiency (<ε>)	multiplicity (<μ>)
2012	6.9	0.96 ±0.04	1.75 ± 0.16
2015	6.9	0.95 ± 0.04	1.81 ± 0.19
2017	6.7	0.95 ± 0.03	1.90 ± 0.30
2018	6.7	0.89±0.04	1.61 ± 0.14

Table 4.1: *The average global efficiency and multiplicity of GRPCs in beam test at CERN. The results of 2018 is different with others due to the gas mixture changed.*

is to apply different configurations for the different ASICs, namely the thresholds and

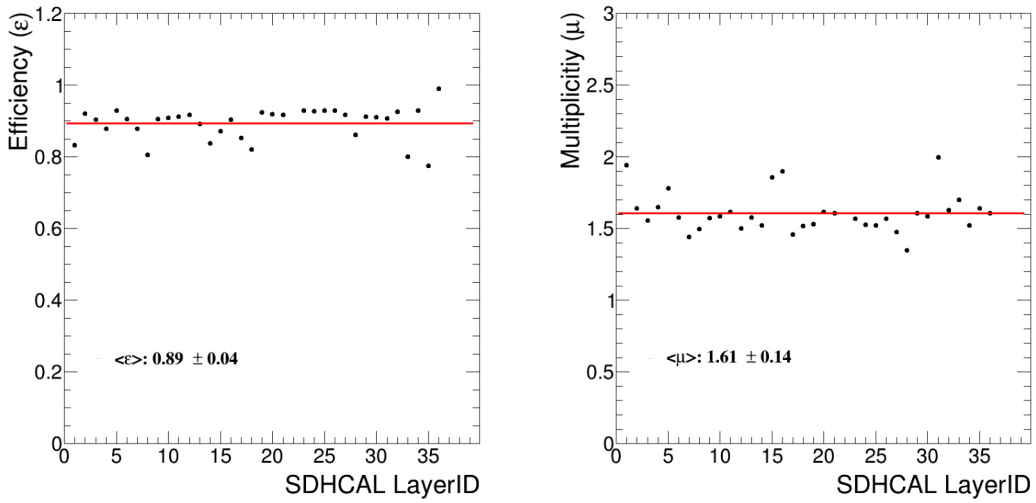


Figure 4.28: *Left: the efficiencies of 35 GRPC layers of SDHCAL prototype. Right: the multiplicities of 35 GRPC layers of SDHCAL prototype. The all GRPC layers work at 6.7 kV. The results are obtained from 2018 beam test.*

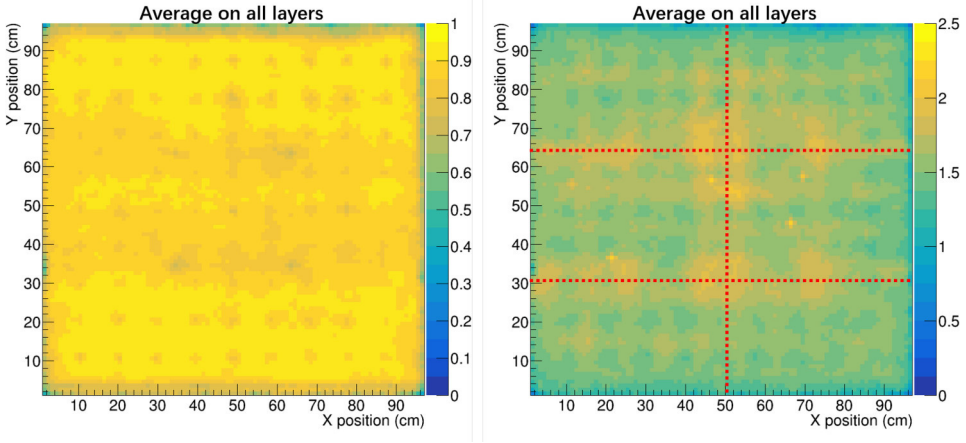


Figure 4.29: *Left: the efficiency map $\epsilon(x, y)$. Right: the multiplicity map $\mu(x, y)$. Both efficiency and multiplicity maps are averaged on 35 layers. The red dashed lines are related to the borders of six ASU PCBs. The results are obtained from 2018 beam test.*

gains, to help to uniform (or compensate) the detector response which will be discussed in Section. 4.2.3.

4.2.3 Homogenization process for SDHCAL prototype

As discussed in Section 4.2.2.3, the non-uniformity of multiplicity is observed in the GRPC layer. The multiplicity plays an important role in shower reconstruction and its energy estimation thus it is necessary to find a solution to reduce the non-uniformity of multiplicity. In the beam test of 2018 at H2 beamline of SPS, a homogenization process had been implemented on SDHCAL prototype to render uniform the distribution of multiplicity. The principle of this method is to apply different readout thresholds on the different ASICs in

order to achieve a uniform response of all the detectors. To initialize the homogenization process, the function between multiplicity as well as efficiency and thresholds should be determined in advance. Thanks to excellent muon tracking capability of SDHCAL, the prototype was exposed to muon beam to collect enough tracks in all the (x, y) plane of the GRPC detectors and this for different values of the different thresholds. For each of the thresholds set, the efficiency and multiplicity were estimated to obtain the relation between the efficiency and multiplicity on the one hand and the used thresholds of readout ASICs on the other hand as explained in Ref. [47]. The function between efficiency and readout threshold "t" is given by:

$$\varepsilon(t;\overline{q},\delta,\epsilon_0) = \varepsilon_0 \cdot \left(1 - \int_0^{thr} P(q;\overline{q},\delta)dq\right) \tag{4.1}$$

$$P(q;\overline{q},\delta) = \frac{1}{\Gamma\left(\frac{\overline{q}}{\delta}\right)\delta^{\frac{\overline{q}}{\delta}}}\, q^{\frac{\overline{q}}{\delta}-1}\, e^{-\frac{q}{\delta}} \tag{4.2}$$

where $\overline{q}$ is the average charge induced by a charged particle passing RPC, δ and ε_0 are two free parameters with positive values. For multiplicity, the function is found to be:

$$\mu(t; f, p, c) = f \cdot t^p + c \tag{4.3}$$

where f, p and c are three free parameters. These six free parameters ($\overline{q}$, δ and ε_0 for efficiency; f, p and c for multiplicity) of each ASIC can be experimentally determined through the threshold scan. Based on these two functions, we can adjust the readout thresholds of each ASIC to obtain the same average multiplicity every where in the GRPC detectors.

The average multiplicities of all the available layers before and after homogenization process are shown in Fig. 4.30. The average multiplicity before and after homogenization process is 1.61 ± 0.14 and 1.47 ± 0.03 respectively. It is clear that a smaller fluctuation of the multiplicity is observed after applying the homogenization process. The map of the averaged multiplicity of the used layers before and after homogenization is shown in Fig. 4.31. A more uniform multiplicity map is obtained after homogenization process and the large fluctuations located in the borders of ASUs is clearly reduced by applying the homogenization process.

To evaluate the impact of the homogenization process, the response of detectors to hadron beams is also estimated. In the 2018 beam test, different areas of the SDHCAL prototype were exposed to the hadron beam (center, top and top right) as shown in Fig. 4.32. For these different positions, we applied first the standard method and then the one based on the homogenization process for a few energy points.

The distribution of nHit (number of hits of a event) is one of the most important physics variables to gauge the SDHCAL prototype response of hadronic events. The comparison of nHit for 50 GeV hadron beam using the two different methods at two different positions is in Fig. 4.33 and that for the 70 GeV hadron beam in Fig. 4.34. The corresponding means of nHit obtained from fitting results of the two methods and the two energies are summarized in Tab. 4.2. In the standard method, the relative deviation between the two different position is calculated. For the 50 GeV and 70 GeV the deviation

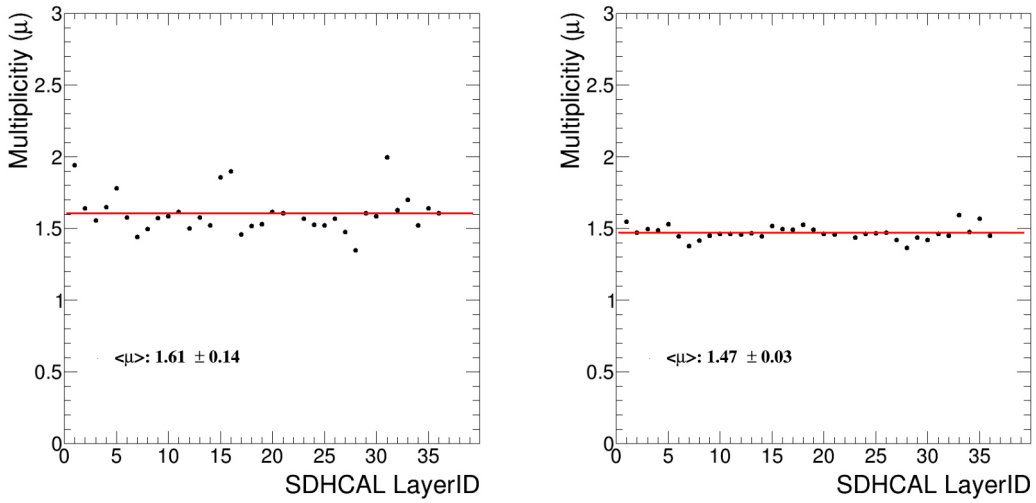


Figure 4.30: *Left: the average multiplicities of 35 GRPC layers of SDHCAL prototype before homogenization process. Right: the average multiplicities of 35 GRPC layers after homogenization process. The results are obtained from 2018 beam test.*

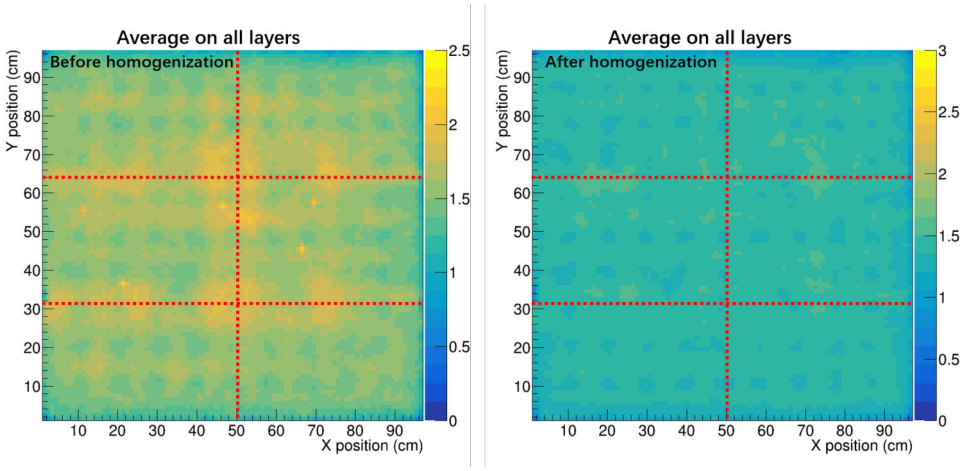


Figure 4.31: *Left: the multiplicity map $\mu(x, y)$ before homogenization process average on all layers. Right: the multiplicity map $\mu(x, y)$ after homogenization process average on all layers. The red dashed lines represent the borders of six ASU PCBs. The results are obtained from 2018 beam test.*

is 4.9% and 3.9% respectively. For the homogenisation-based case, the relative deviation for 50 GeV and 70 GeV are 1.8% and 1.4% respectively. The homogenisation process provides clearly a more uniform response than the standard method. The deviations described here before are summarized in Tab. 4.3.

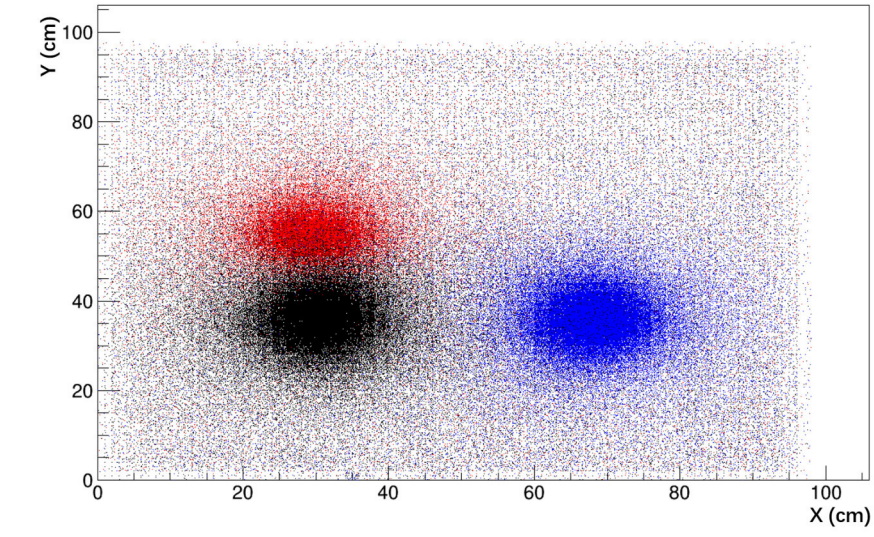


Figure 4.32: *The three hadron beam injecting positions such as center, top and top right are represented by red, black and blue areas respectively.*

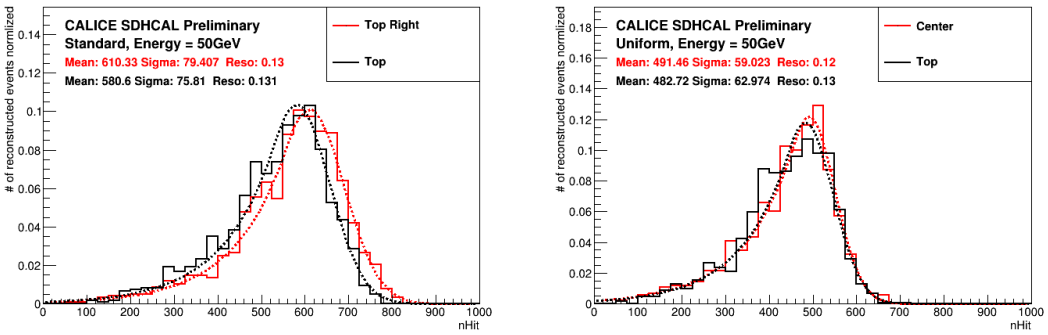


Figure 4.33: *Left: the Nhit (number of hits of a event) distribution of 50 GeV pion runs injected to top and top right areas without applying homogenization process. Right: the Nhit distribution of 50 GeV pion runs injected to center and top areas after applying homogenization process. The fit function of both figures are Crystal Ball function. The results are obtained from 2018 beam test.*

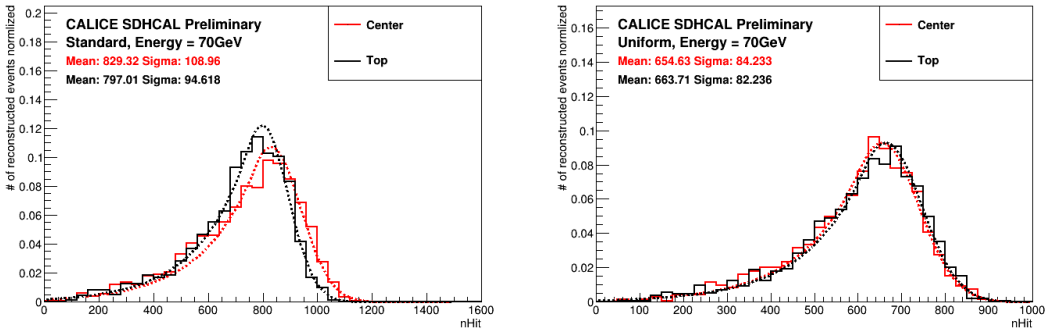


Figure 4.34: *Left: the Nhit (number of hits of a event) distribution of 70 GeV pion runs injected to center and top areas without applying homogenization process. Right: the Nhit distribution of 70 GeV pion runs injected center and top after applying homogenization process. The fit function of both figures are Crystal Ball function. The results are obtained from 2018 beam test.*

Energy	Center <nHit>	Top <nHit>	Top right <nHit>
	Standard case		
50 GeV	-	580.6 ± 5.7	610.3 ± 4.8
70 GeV	829.3 ± 1.3	797.0 ± 7.1	-
	Uniform case		
50 GeV	491.5 ± 1.3	482.7 ± 1.3	-
70 GeV	654.6 ± 1.5	663.7 ± 1.30	-

Table 4.2: *The means of nHit distribution for standard and uniform case, which are obtained from the fitting results by using Crystal Ball function based on 2018 beam test data.*

Energy (GeV)	Standard case	Uniform case
50	4.9%	1.8%
70	3.9%	1.4%

Table 4.3: *The relative deviation of means ($\frac{\triangle nHit}{nHit}$) for both standard and uniform cases, where nHit is the number of hits of a event.*

4.2.4 Simulation of SDHCAL prototype

In order to study the performance (e.g., particle identification, energy reconstruction) of SDHCAL prototype, it is necessary to develop a reliable simulation framework for simulating the response of GRPC detectors for different kinds of particles (e.g., pions, electrons, muons). Based on the GEANT4 toolkit [48] and the digitization process (called SimDigital algorithm), the simulation framework of SDHCAL prototype is found to reproduce the response of muons, electromagnetic showers and hadronic showers in the SDHCAL prototype [49]. Among the different Geant4 physics lists (physics lists apply a string model for the modeling of interactions of high energy hadrons), the FTF_BIC[1] [50] [51] was found to give the best agreement when comparing the simulation and the SDHCAL beam data [52].

4.2.5 The standard method for pion shower selection

For measuring the response of SDHCAL prototype to hadronic showers, the SDHCAL prototype was exposed several times to pion beams in SPS and PS at CERN. However, the pion beams, are contaminated by several different particles such as electrons as well as the muons produced by pions decaying before interacting with prototype . In addition, during the data taking the SDHCAL also detect the cosmic muons. Because of the absence of Cherenkov detectors in front of the SDHCAL during the beam tests, it is impossible to identify pions and eliminate the contamination by other particles. Therefore, it is necessary to develop other methods to select the pion events. In this section, the standard method (cut-based) for selection of pion showers is discussed. Another method, based on boosted decision tree method, will be discussed in Section. 5.

In order to simplify the description, the coordinate of SDHCAL prototype and beam direction are shown in Fig. 4.35. A common right-handed coordinates system is adopted for prototype whose entire 48 layers were arranged perpendicular to the incoming beams. The origin of the system is defined as the center of the first layer of the SDHCAL. The $x-y$ plane is parallel to the SDHCAL layers and referred to as the transverse plane while the z-axis runs parallel to the incoming beams. Based on this coordinates system, event displays for a pion, an electron and a muon of the 2015 collected data are shown in Fig. 4.36, Fig. 4.37 and Fig. 4.38 respectively. Pions in general penetrate several SDHCAL layers before to start showering due to the nature of their strong interaction. The compact part (red) of the hadronic shower is due to the EM component. Electrons, however, start showering almost immediately when they reach the first layer of the SDHCAL absorber since one SDHCAL absorber layer presents approximately 1.2 radiation length. Beam muons behave as mip particles in the SDHCAL leaving one straight track in the prototype. For a given event, the number of hits corresponding to induced charges passing the first , second and third threshold values are referred to as the nHit1, nHit2 and nHit3 respectively. The total number of hits is called nHit and it is equal to the sum of nHit1, nHit2 and nHit3. The nHit distribution of a 40 GeV pions is shown in Fig. 4.39.

[1]The FTF model is based on the Fritiof description of string excitation and fragmentation. BIC model means using Geant4 Binary cascade for primary protons and neutrons with energies below 10 GeV. BIC model better describes production of secondary particles produced in interactions of protons and neutrons with nuclei.

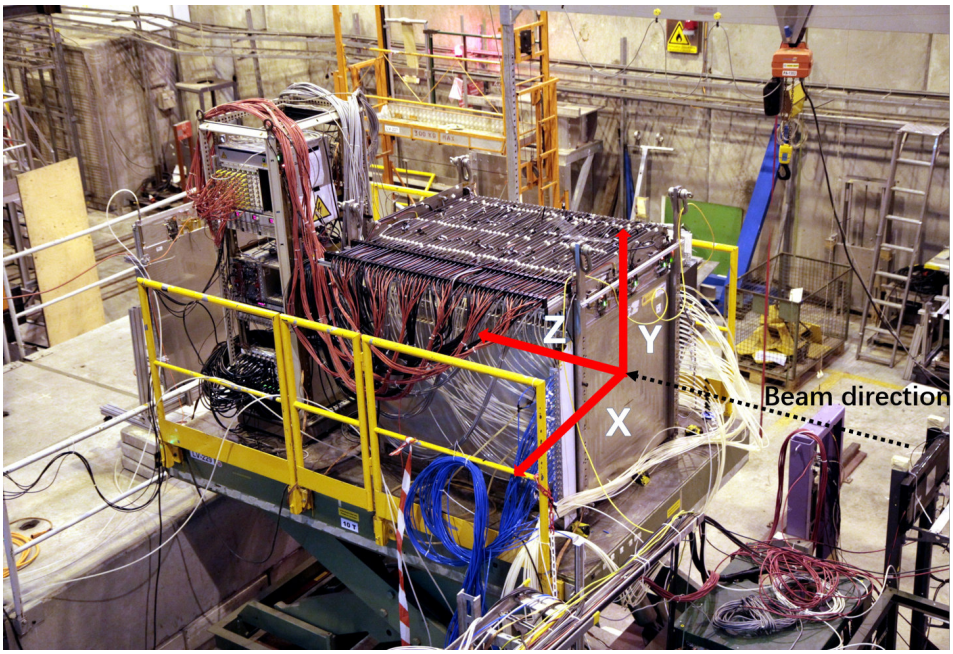


Figure 4.35: *A picture of the SDHCAL prototype on the SPS H2 beamline. The coordinates system used in the thesis are drawn as well as the beam axis and its direction.*

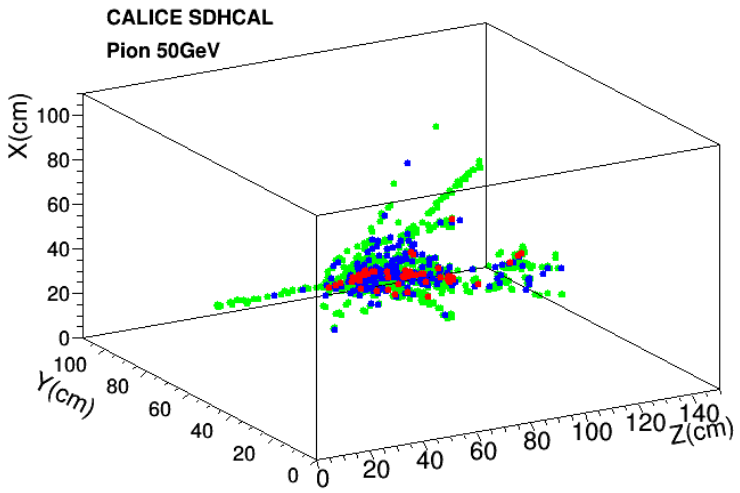


Figure 4.36: *The event displays of a 50 GeV pion. Hits pass the first threshold are depicted in green. Hits that pass the second in blue while the those that pass the third threshold are in red.*

The two peaks at top left show the contamination of muons (both beam ones and cosmic ones). The difference of nHit between these two peaks can be explained by the fact that most of cosmic muons only cross several layers with a large angle between its direction and the beam one. and beam muons is parallel to beam direction thus they almost can pass through whole 48 layers of prototype. The right wide peak is the contribution of hadronic showers. The electron contamination is hidden by the hadronic shower components thus not clearly seen in the figure. This is indicated in the Fig. 4.40

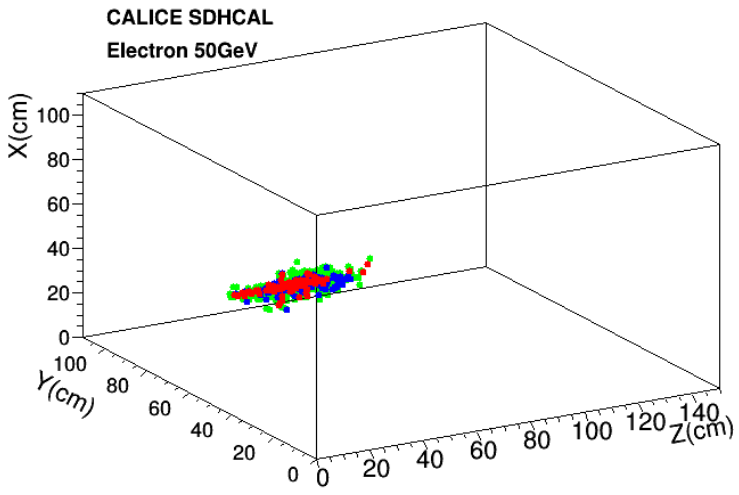


Figure 4.37: *The event displays of a 50 GeV electron. Hits pass the first threshold are depicted in green. Hits that pass the second in blue while the those that pass the third threshold are in red.*

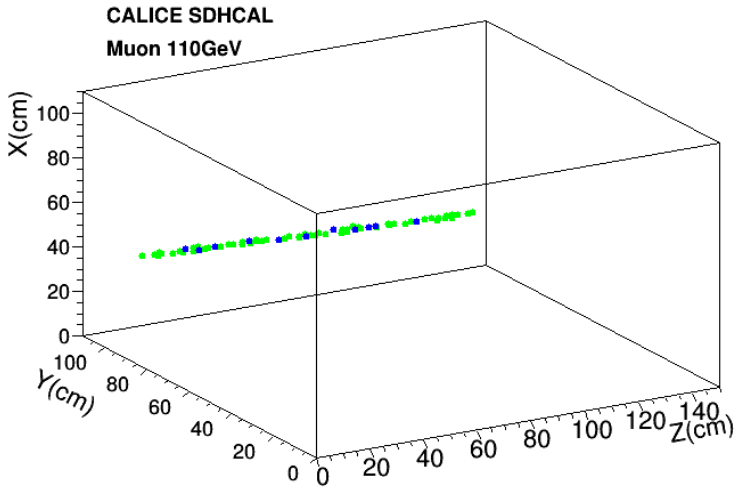


Figure 4.38: *The event displays of a 110 GeV muon . Hits pass the first threshold are depicted in green. Hits that pass the second in blue while the those that pass the third threshold are in red.*

that the nHit distribution of 40 GeV simulated electrons apparently overlaps with the nHit distribution of 40 GeV pions.

To apply standard method, several cut are studied.

- **Pre-selection**: The interaction area of the physics events should be around the impinging pion beam, if the events are outside this range they are removed. This can help to eliminate part of cosmic muons and noisy events. In addition , to eliminate neutral hadron events, events should have at least 4 hits in the first 5 layers.

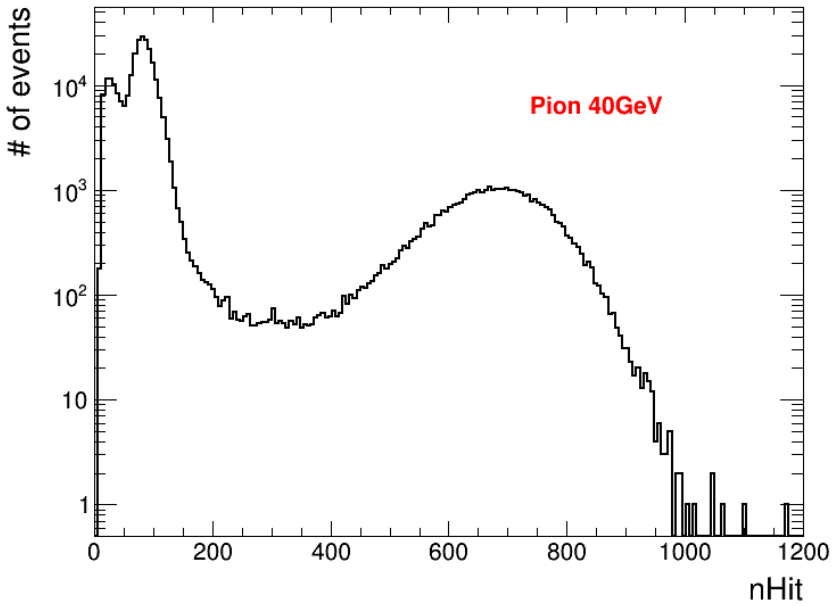


Figure 4.39: *The nHit distribution of 40 GeV pion beams taken at 2015. The two peaks at the top left are respectively contributed by the cosmic muons and muons produced by beam pion decaying. the relative wider peak at right is contributed by the hadronic showers.*

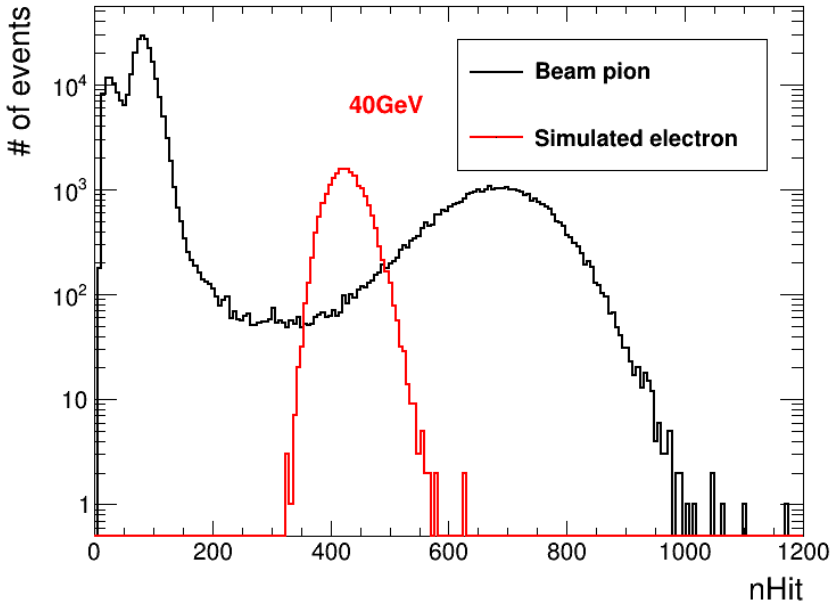


Figure 4.40: *The nHit distribution of 40 GeV pion beams (black) and simulated electron events (red). It is clear that electron contamination can be hidden by the huge number of pions due to the overlaps of the distributions of them*

- **Muon cut**: In order to eliminate the muons, the ratio of number of hits over number of fired layers (at least one hit in the layer) is required to be higher than 3. This cut

can reject the muon events with non-radiation. For the radiating muons as shown in Fig. 4.41, the another cut is considered.

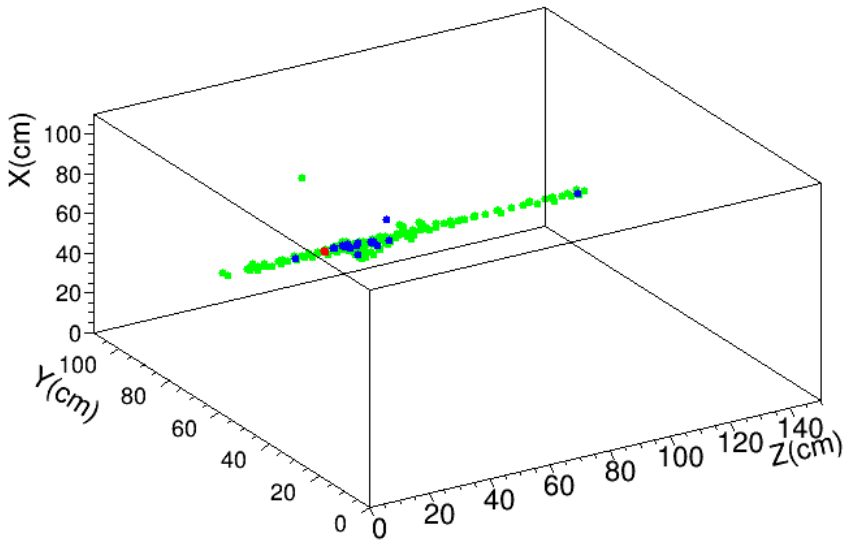


Figure 4.41: *The event display of a radiating muon in 2015 beam test. The red, blue, green color represents the hits associated with third, second and first thresholds.*

If a layer where the root mean square of the position (x, y) of hits is found to be greater than 5 cm , it is inferred to as a interacting layer. To eliminate the radiating muons, the ratio of number of interacting layers over number of fired layers is requested to be higher than 20%. These two muon cuts have limited impact on the selection of pions for which a large area of the detectors will be hit after the showering start. This is confirmed by applying the same cuts on a simulated pion events sample.

- **Electron cut**: Unlike most of the pions that generate hadronic showers after penetrating several layers, the electrons start showering almost in the first plate of the prototype. This result from the fact one SDHCAL layer corresponds to 1.14 radiation length given that the radiation length of steel is 17.6 mm. To eliminate almost of electron events, the events is requested to start its shower after the fifth SDHCAL layers corresponding to approximately $6X_0$. The layer of shower start showering is defined as the layer where we can find at least 4 hits followed by at least three layers with more than 4 hits. This definition avoids the fake shower starts due to noise or accidental high multiplicity. In addition, EM showers are found to be contained in less than 30 SDHCAL layers in case of electrons with energy lower than 80 GeV as shown in Fig. 4.42. Therefore, the electron cut is applied only for events with less than 30 fired layers.

The previous cuts are summarized in Tab. 4.4.

The performance of the standard method selection is shown in Fig. 4.43, Fig. 4.44 and Fig. 4.45 for pion data of three energy points 20, 50 and 80 GeV respectively.

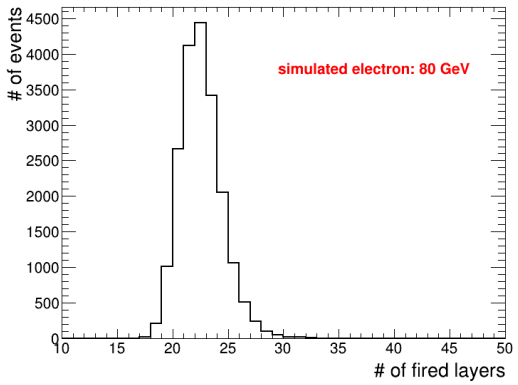


Figure 4.42: *The distribution of number of fired layers for simulated 80 GeV electrons.*

Cut type	detail
pre-selection	$\mathrm{nHit}_{first5layers} \geqslant 4$
Muon cut	$\mathrm{nHit} / \mathrm{nLayer}_{fired} > 3.0$
radiating muon cut	$\mathrm{nInetractingLayer} / \mathrm{nLayer}_{fired} > 20\%$
electron cut	$\mathrm{Layer}_{begin} \geqslant 5$ or $\mathrm{nLayer}_{fired} > 30$

Table 4.4: *The summary of standard method for pion events selection.*

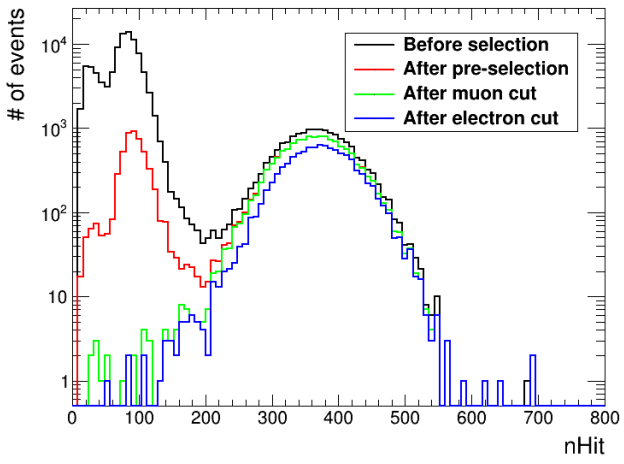


Figure 4.43: *The nHit distribution of 20 GeV pion beam runs before selection (black) and after cuts applied step by step.*

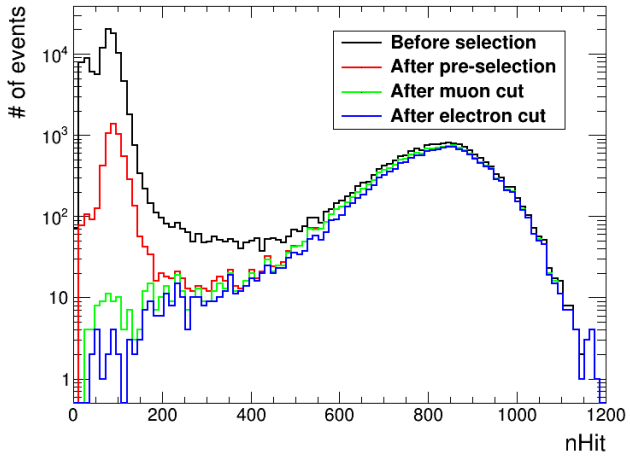


Figure 4.44: *The nHit distribution of 50 pion beam runs before selection (black) and after cuts applied step by step.*

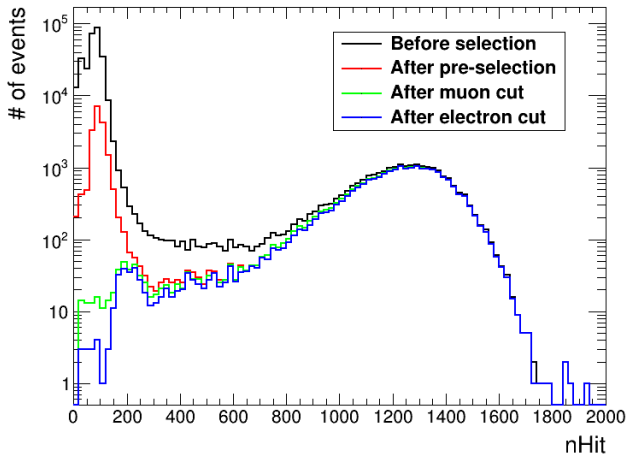


Figure 4.45: *The nHit distribution of 80 GeV pion beam runs before selection (black) and after cuts applied step by step.*

4.2.6 Time correction

Due to the fact that the efficiency of glass RPC decreases at high particle rate, the intensity of beam used for the SDHCAL prototype is optimized to balance efficiency and collecting as much events as possible. In beam test period, beam intensity was tuned to have less than 1000 particles per spill and thus limited particle flux to less than 100 particle/s·cm^2. However, the phenomenon that average number of hits of hadronic events decreases with the beam particle spill time was still observed. This can be explained by the fact that the rate of the produced particles within the shower may exceed the 100/s·cm^2 which is the limit for which the GRPC efficiency starts to decrease. The distributions of the average number of hits associated to the first (nHit1), the second (nHit2) and the third (nHit3) thresholds as a event spill time for 10, 40 and 80 GeV as a function of the event occurrence with respect tot the start of the spill are shown in Fig. 4.46, Fig. 4.47 and Fig. 4.48 respectively indicated by the black line. It is clear that this decrease is stronger in the higher energy pions when comparing results of 10, 40 and 80 GeV. In addition, this deviation is also more apparent in the number of hits related to the second and third thresholds than the first one for the the same energy. In order to correct this effect, a calibration method is proposed and a correction is applied. The functions describing the average number of hits related to the different thresholds with respect to the start of the spill time can be fitted by a second polynomial function:

$$N^{corr}_{thr,t} = N_{thr,t} - (a_0 + a_1 * t + a_2 * t^2) \tag{4.4}$$

where $N_{thr,t}$ represents the average numbers of hits associated with a given threshold, $a_{0,1,2}$[2] are the parameters of second polynomial fitting function and t is time difference with respect to the spill time start. The results after time correction described by Eq. (4.4) are shown in Fig. 4.46, Fig. 4.47 and Fig. 4.48 for 10, 40 and 80 GeV respectively.

[2] a_0 is negative in this correction.

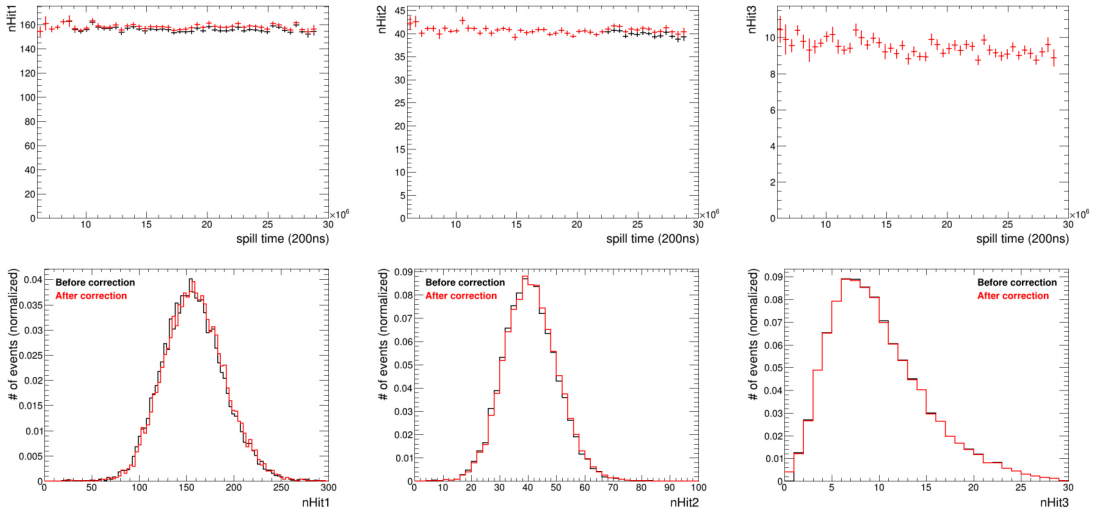


Figure 4.46: *Top: The functions describing average number of hits of hadronic showers for first (left) , second (middle) and third (right) thresholds with event spill time. The black symbols are the results before time correction and red symbols are the results after time correction using the second polynomial fit. Bottom: the number of hits of hadronic showers associated with first (left) , second (middle) and third (right) thresholds before (black line) and after (red line) time correction. The results are obtained from the 10 GeV beam data taken at 2015.*

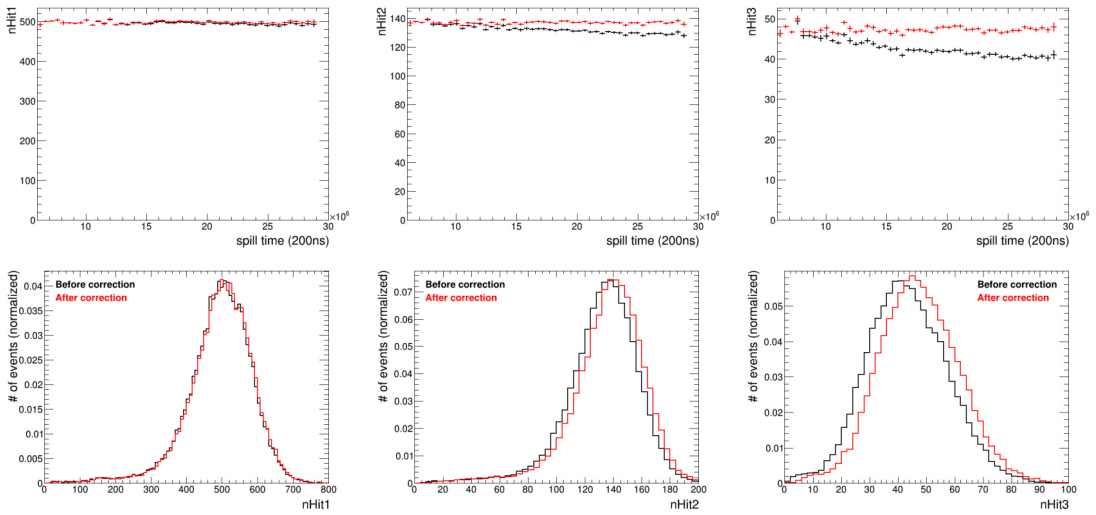


Figure 4.47: *Top: The functions describing average number of hits of hadronic showers for first (left) , second (middle) and third (right) thresholds with event spill time. The black symbols are the results before time correction and red symbols are the results after time correction using the second polynomial fit. Bottom: the number of hits of hadronic showers associated with first (left) , second (middle) and third (right) thresholds before (black line) and after (red line) time correction. The results are obtained from the 40 GeV beam data taken at 2015.*

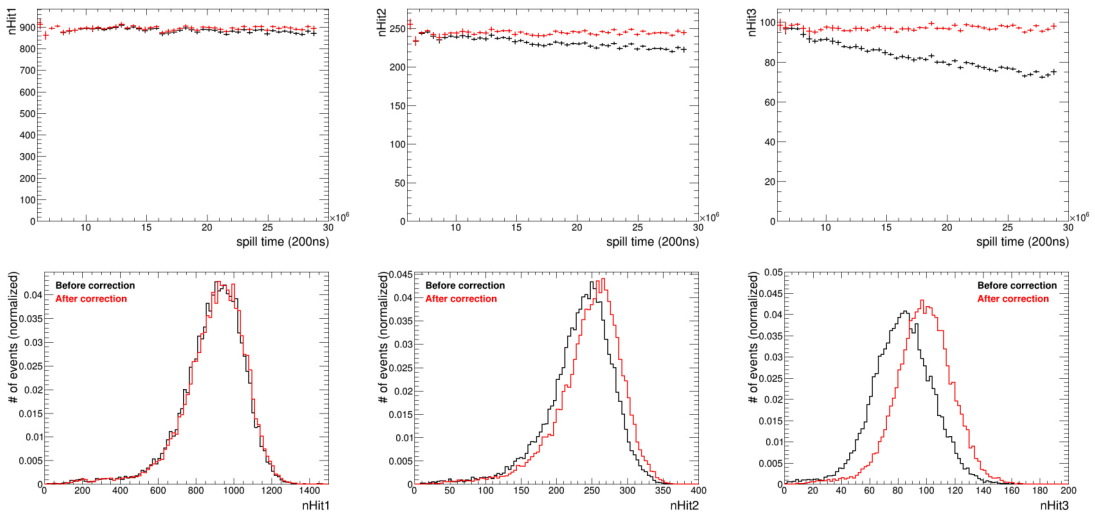


Figure 4.48: *Top: The functions describing average number of hits of hadronic showers for first (left) , second (middle) and third (right) thresholds with event spill time. The black symbols are the results before time correction and red symbols are the results after time correction using the second polynomial fit. Bottom: the number of hits of hadronic showers associated with first (left) , second (middle) and third (right) thresholds before (black line) and after (red line) time correction. The results are obtained from the 80 GeV beam data taken at 2015.*

4.2.7 Energy reconstruction of hadronic showers

After the pion events being selected by the standard method discussed in Section. 4.2.5, the one of the most important targets is to estimate the performance of the energy reconstruction of such selected events. In the SDHCAL prototype, there are two approaches for energy reconstruction such as the binary mode and the multi-threshold mode according to the number of thresholds one uses.

4.2.7.1 Binary mode

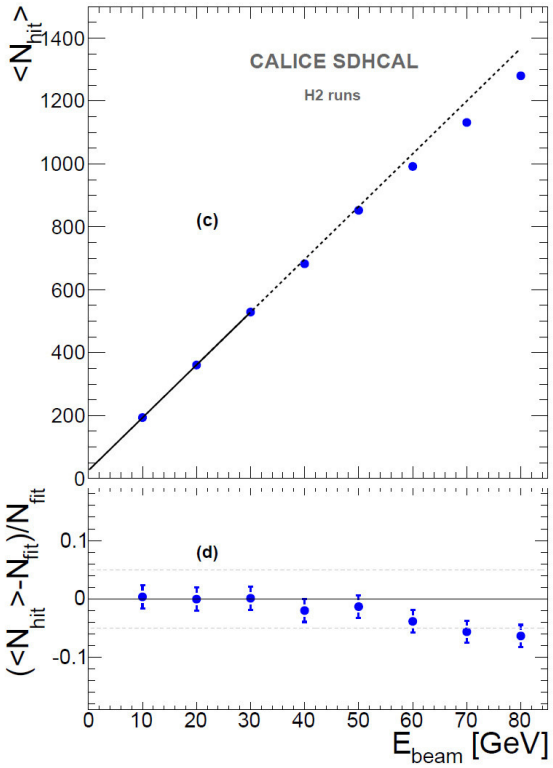


Figure 4.49: *The mean of total number of hits as a function of beam energy. The dashed line is a function <N_{hit}> = k * E_{beam}. The relative deviation $\frac{\Delta N_{hit}}{N_{hit}}$ as a function of beam energy is also shown. Right: the mean of number of hits associated with first (green square), second (blue triangle) and third red cross thresholds as a function of beam energy [46].*

In the binary model, only the information about whether a pad is fired or not is used (so sometimes it is called digital mode). The idea to reconstruct the energies of pion showers in this mode is implied by the fact that the mean total number of hits is correlated with the incoming pion's energy as shown in Fig. 4.49. It is clear that there is approximately linear dependency between average total number of the hits and the beam energy for energies less than 30 GeV. In the higher energy range (40-80 GeV), the deviation from linear relation is present and caused by the saturation effect of hadronic showers. This effect can be explained by that in the core of the hadronic shower pads

are crossed by many particles and thus the number of fired pads is not proportional to the number of produced particles which is, in principle, proportional to the average deposited energy. By taking into consideration of this saturation effect, several different energy reconstructed formulae are investigated. One formula was found to provide the best performance:

$$E_{reco} = \sum_{i=1}^{3} A_i * N_{hit}^i \tag{4.5}$$

where the $A_{i=1,2,3}$ are 3 parameters are determined by executing the χ^2-based optimizer:

$$\chi^2 = \sum_{i=1}^{N} \frac{(E_{reco}^i - E_{beam}^i)^2}{\sigma_i^2} \tag{4.6}$$

where N and E_{beam}^i are the total number of events used and beam energy of event i respectively and the E_{reco}^i is from Eq. 4.5 of event i. The σ_i equals to $\sqrt{E_{beam}^i}$ suggested by the fact that hadronic calorimeter resolution is expected to be approximately proportional to $\sqrt{E_{beam}}$.

After the optimization process based on Eq. 4.6 is completed, the three parameters $A_{i=1,2,3}$ are fixed and the energy of pion events are subsequently reconstructed by using Eq. 4.5. The results of energy reconstruction for 20 and 40 GeV pion data in binary mode are shown in Fig. 4.50. The distribution of reconstructed energy are fitted by a two-step

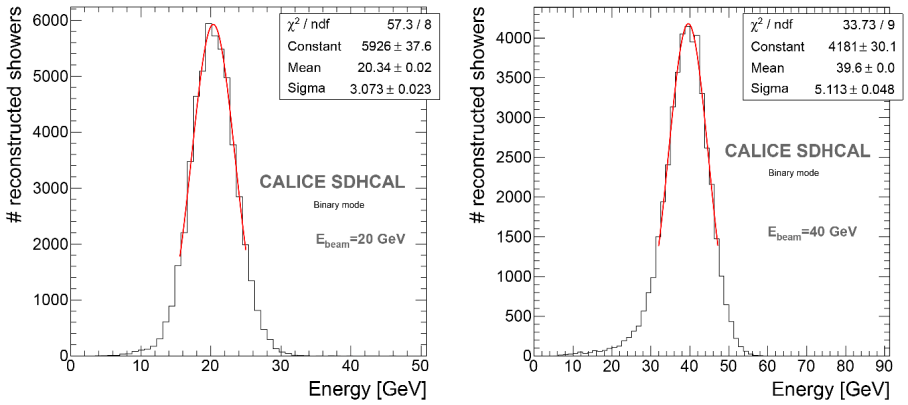


Figure 4.50: *The distribution of reconstructed energy in binary mode for 20 (left figure) and 40 GeV (right figure) pion data taken at 2012 [46]. A Gaussian function is used to fit both two figures in a $\pm 1.5\sigma$ range around the mean. The fitting results are also shown in the top right windows.*

Gaussian function process. First, the energy distribution is fitted by the Gaussian-1 over the full range. Then Gaussian-2 is used to fit the same energy distribution in a $\pm 1.5\sigma$ range around the mean value obtained from the fitting results of Gaussian-1. The mean μ and $\frac{\sigma}{\mu}$ obtained from Gaussian-2 are considered as the final reconstructed energy and the relative energy resolution. The results of reconstructed energy and its resolution in binary mode are shown in Fig. 4.51. The expected linear relation between reconstructed energy and beam energy was observed.

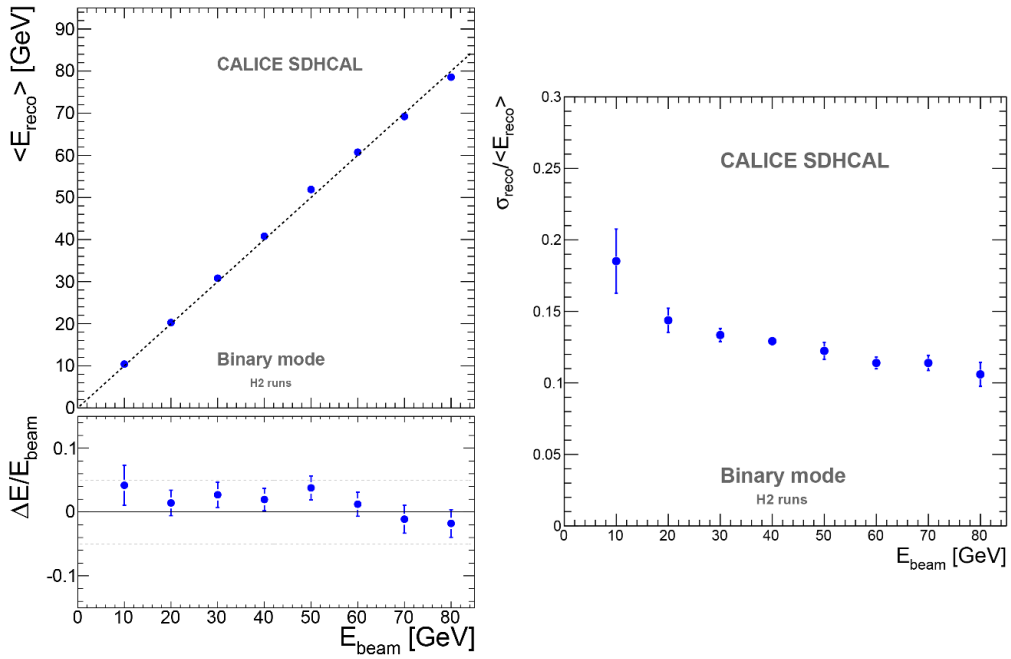


Figure 4.51: *Left: The mean of reconstructed energy and the relative deviation $\frac{\Delta E}{E}$ as a function of beam energy in binary mode. Right: the energy resolution as a function of beam energy in binary mode. The both figures are obtained from 2012 beam data [46].*

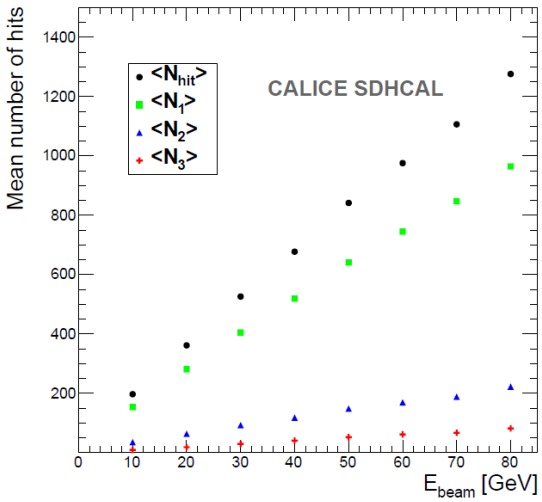


Figure 4.52: *The mean of number of hits associated with first (green square), second (blue triangle) and third red cross thresholds as a function of beam energy. The mean of total number of hits as a function of beam energy is also shown [46].*

4.2.7.2 Multi-threshold mode

Unlike the binary mode for which prototype can only provide an information whether a pad is fired or not independently of the number of particles that passed through it, in the multi-threshold mode it indicates whether a few, many or very many particles pass through the pad according to the induced charge. The evidence for such separation is

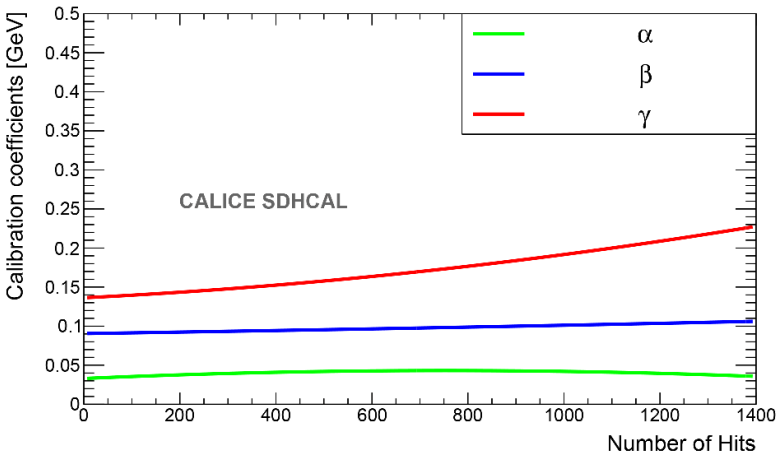


Figure 4.53: *The α, β and γ as a function of total number of hits [46].*

already shown in from Fig. 4.36 to Fig. 4.38. An electron produces a compact EM shower which presents the dense shower core indicated by the large fraction of hits associated with the third threshold. An important fraction of hits related to the third threshold is also observed in the core of hadronic showers due to the EM component. In multi-threshold mode, the average numbers of hits associated with first, second and third threshold as a function of beam energy are shown in Fig. 4.52. Inspired by the experience of energy reconstruction in binary mode, the energy reconstruction in multi-threshold mode can be expressed by the weighted formula with the energy expressed as a weighted sum of the average numbers of hits of the three different thresholds:

$$E_{reco} = \alpha N_1 + \beta N_2 + \gamma N_3 \tag{4.7}$$

where $N_{1,2,3}$ is the number of hits associated with its corresponding threshold. Due to the nature of hadronic shower, the three parameters α, β and γ should not be considered constant in a wide energy range. Therefore, these three parameters are subsequently parameterized as functions of the total number of hits. Based on the optimizer indicated by Eq. 4.6, several forms of function are investigated and the second polynomial function was found to achieve the best performance. Thus, these three parameters can be expressed by:

$$\begin{aligned} \alpha &= \sum_{i=0}^{2} \alpha_i * N_{hit}^i \\ \beta &= \sum_{i=0}^{2} \beta_i * N_{hit}^i \\ \gamma &= \sum_{i=0}^{2} \gamma_i * N_{hit}^i \end{aligned}$$

To determine the nine parameters ($\alpha_{i=0,1,2}$, $\beta_{i=0,1,2}$ and $\gamma_{i=0,1,2}$), optimization process based on optimizer discussed in Eq. 4.6 was applied on the collected data. The α, β and γ as a function of total number of hits are shown in Fig. 4.53.

The distributions of reconstructed energy for 20 and 40 GeV pion beam runs are shown in Fig. 4.54. The mean and resolution of reconstructed energy as a function of beam energy are shown in Fig. 4.55. The linearity of reconstructed energy is found in the range [-5%, +5%] and the resolution of 80 GeV is found to reach 7.7%. Comparing with Fig. 4.51, the apparent resolution improvement of multi-threshold mode is clearly observed when beam energy is higher than 30 GeV. This is explained by the fact that this mode takes into account the saturation effect in more suitable way than the binary mode.

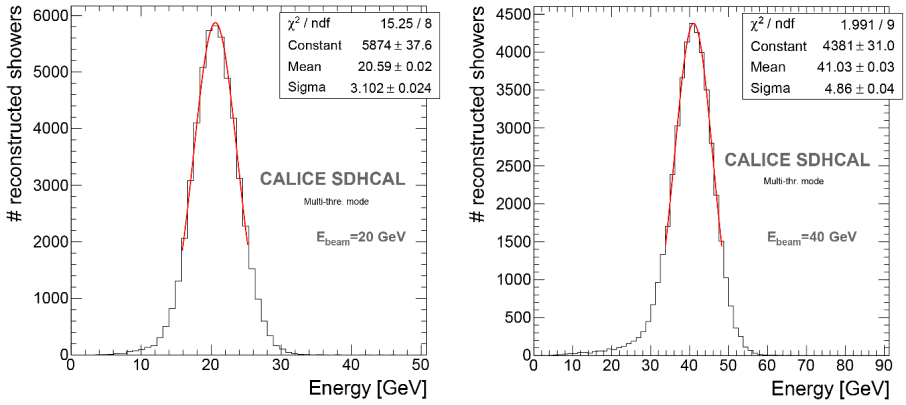


Figure 4.54: *The distribution of reconstructed energy in multi-threshold mode for 20 (left figure) and 40 GeV (right figure) pion data taken at 2012 [46]. A Gaussian function is used to fit both two figures in a $\pm 1.5\sigma$ range around the mean. The fitting results are also shown in the top right windows.*

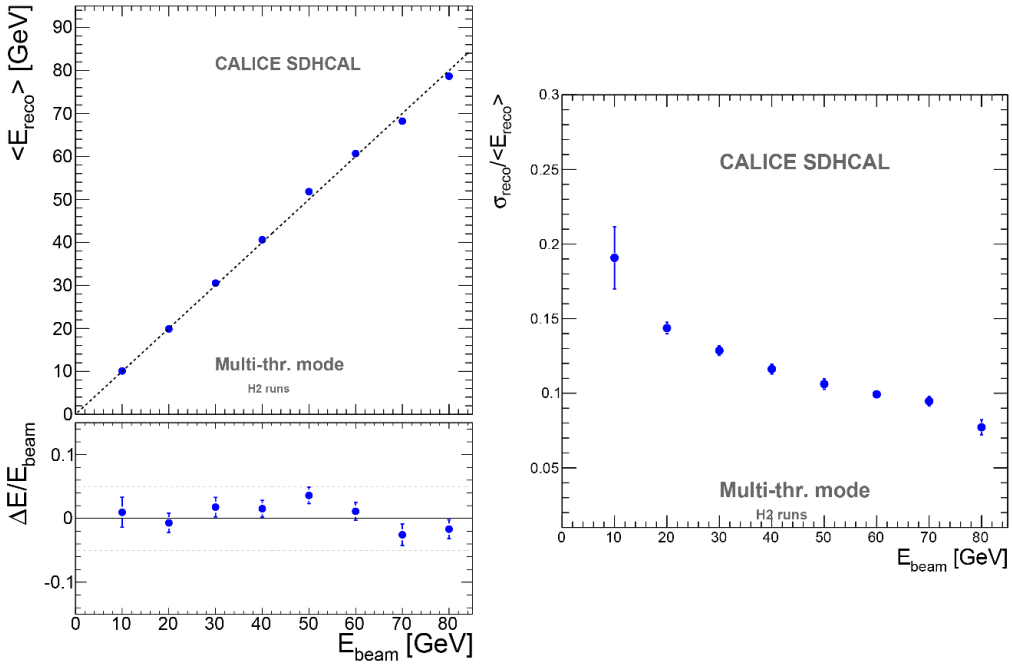


Figure 4.55: *Left: The mean of reconstructed energy and the relative deviation $\frac{\triangle E}{E}$ as a function of beam energy in multi-threshold mode. Right: the energy resolution as a function of beam energy in multi-threshold mode [46].*

Finally, apart from the energy reconstruction method discussed in this section, two other techniques are also investigated. One is based on Multi Variables Analysis method (that will be introduced in Section. 6) and another is based on hit density information that was covered in the Ref [47].

4.2.8 Optimization study for number of SDHCAL layers in CEPC

The SDHCAL prototype was built to confirm the performance of such a PFA-based hadronic calorimeter for the future ILC experiments for which the number of layers and their structure was optimized to reach the goal of jet energy resolution. The SDHCAL technology was afterwards also proposed to equip the detectors of the Circular Electron Positron Collider (CEPC). However in the case of the latter the detector volume needs to be reduced with respect to those of the ILC ones. This constraint is imposed by the reduced L^* value of the Machine-Detector-Interface (MDI) proposed in the case of the CEPC to reach the required luminosity.

To evaluate the impact of a reduced version of the SDHCAL calorimeter on its performance, the same procedure as the one detailed in the previous section on the energy reconstruction using the standard method is applied on the events of the SDHCAL using the hits information of only the first 36, 40, 44 layers respectively. The method is applied on both data and simulation events. For each scenario, we apply the same optimization procedure to the parameters α,β,γ for energy reconstruction. Fig. 4.57 shows the linearity and energy resolution obtained for these aforementioned SDHCAL configurations in addition to the 48-layer one using energy reconstruction standard method. As seen from Fig. 4.57, the choice of a SDHCAL with 40 layers could be a good compromise that reduces the hadronic calorimeter length by one over sixth while still providing very good performance. The same conclusion could be drawn by comparing the linearity and energy resolution when the procedure is applied to the simulated events as shown in Fig. 4.56.

However, when comparing Fig. 4.56 and Fig. 4.57, we found that the impact on the energy resolution is less important in the case of the simulation than in data when reducing number of layers. This could be due to the discrepancy we observed in the shower shape between data and simulation in SDHCAL. To confirm this difference, the number of hits within the shower as a function of the layer number are studied, which are shown in Fig. 4.58, Fig. 4.59 and Fig. 4.60 for 10 GeV, 40 GeV and 80 GeV respectively. From these plots, we observe a good agreement between data and simulation on the longitudinal profile of shower. However, a difference in the number of hits of different thresholds and in particular those of the second and the third one can be seen. Since our energy reconstruction algorithm based on the number of hits of the three thresholds, the difference may explain the origin of the degradation which is more important in data than in simulation.

4.3 Summary

The SDHCAL technological prototype is the first technological prototype of the family of High-granularity calorimeter. It fulfills the requirements of compactness, robustness and low power-consumption that are needed in the future collider experiment. The glass-

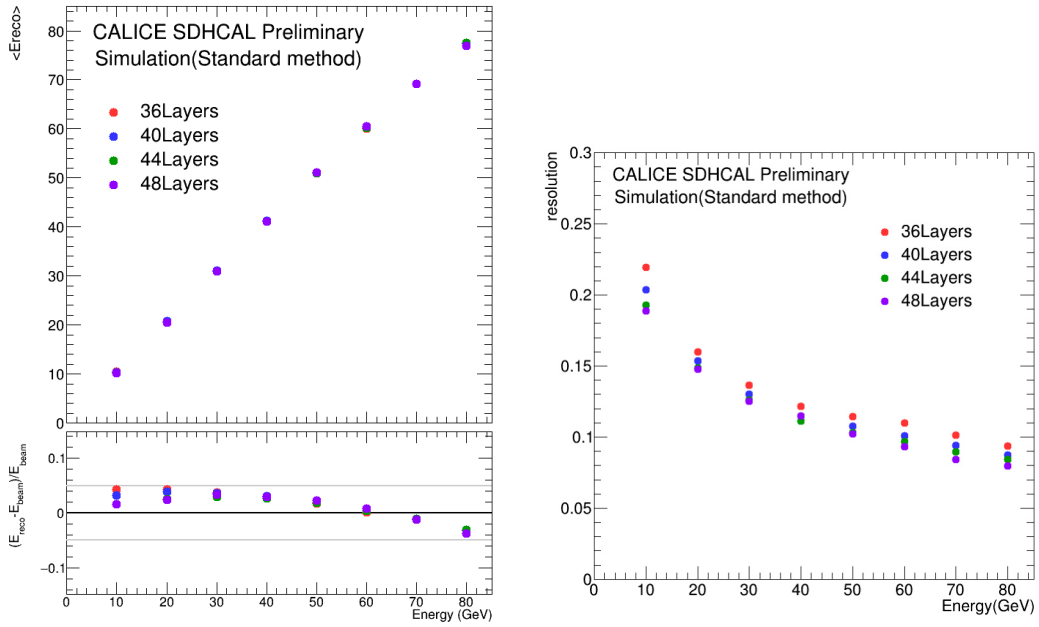


Figure 4.56: *Mean reconstructed energy for pion showers as a function of the simulation energy as well as relative deviation of the pion mean reconstructed energy with respect to the simulation energy (left) and resolution of the reconstructed hadron energy as a function of the simulation energy(right) for different scenarios: 36, 40, 44 active layers as well as for the standard SDHCAL scenario of 48 layers.*

based RPC detector is the essential part of SDHCAL technological prototype and it is found to have high efficiency, to allow fine lateral segmentation (1 cm $\times$ 1 cm) and to be cost-effective. It is exposed to different particle beams in tests at PS and SPS of CERN. To study its performance, the algorithms including reconstructing physics event, hadron event selection and energy reconstruction are studied. To improve the performance, the study of homogenization process of SDHCAL are also presented.

]

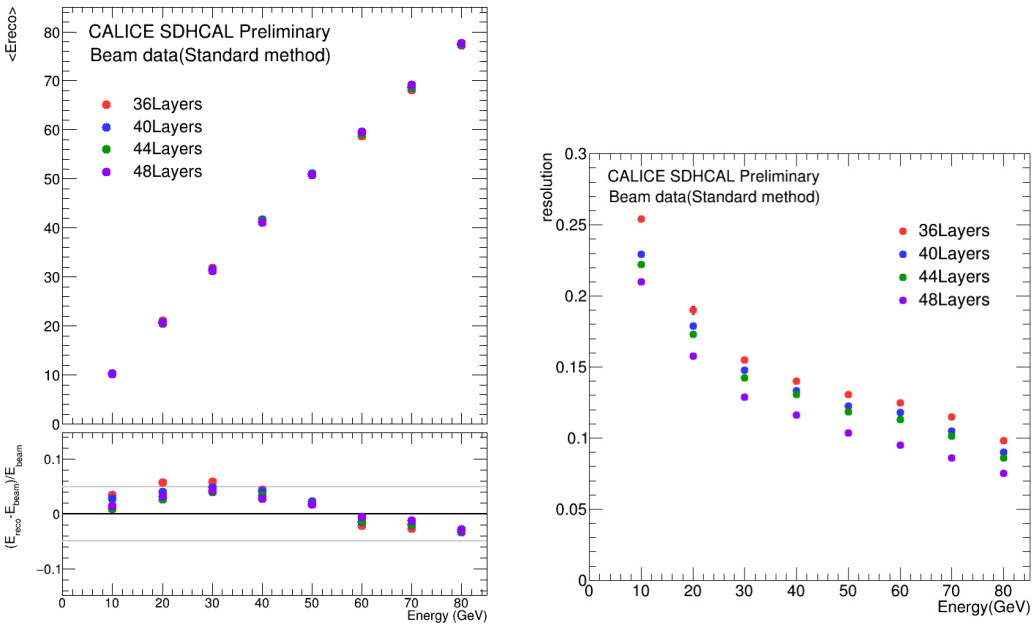


Figure 4.57: *Mean reconstructed energy for pion showers as a function of the beam energy as well as relative deviation of the pion mean reconstructed energy with respect to the beam energy (left) and resolution of the reconstructed hadron energy as a function of the beam energy(right) for different scenarios: 36, 40, 44 active layers as well as for the standard SDHCAL scenario of 48 layers.*

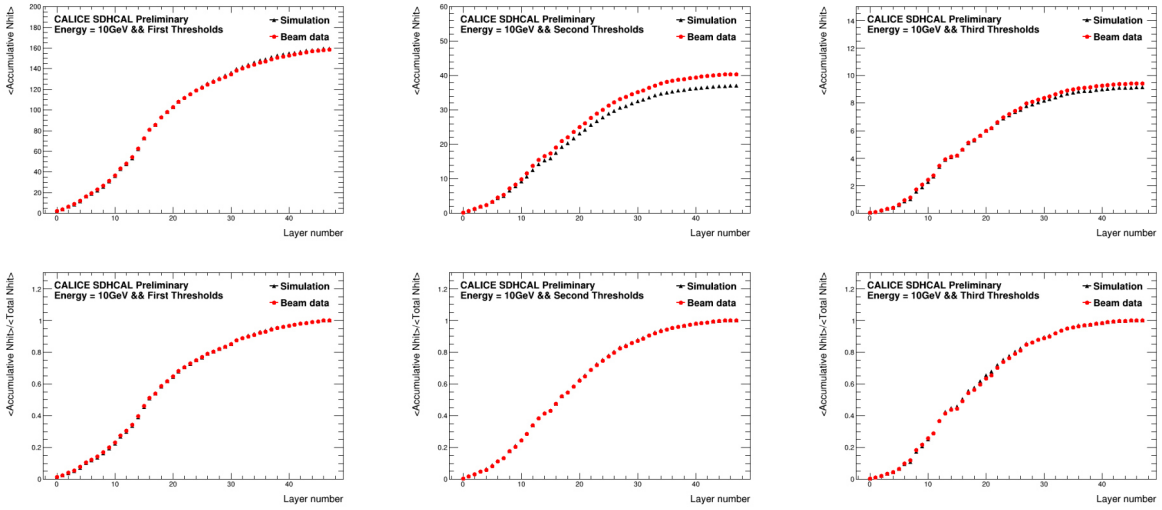


Figure 4.58: *The upper three plots show the comparison of accumulated number of hits of first(left),second(middle),third(right) thresholds as a function of the layer number when intergating over all layers between simulation and data for the 10GeV. The three plots below show the same plots as the upper but using the total number of hits normlized to 1.*

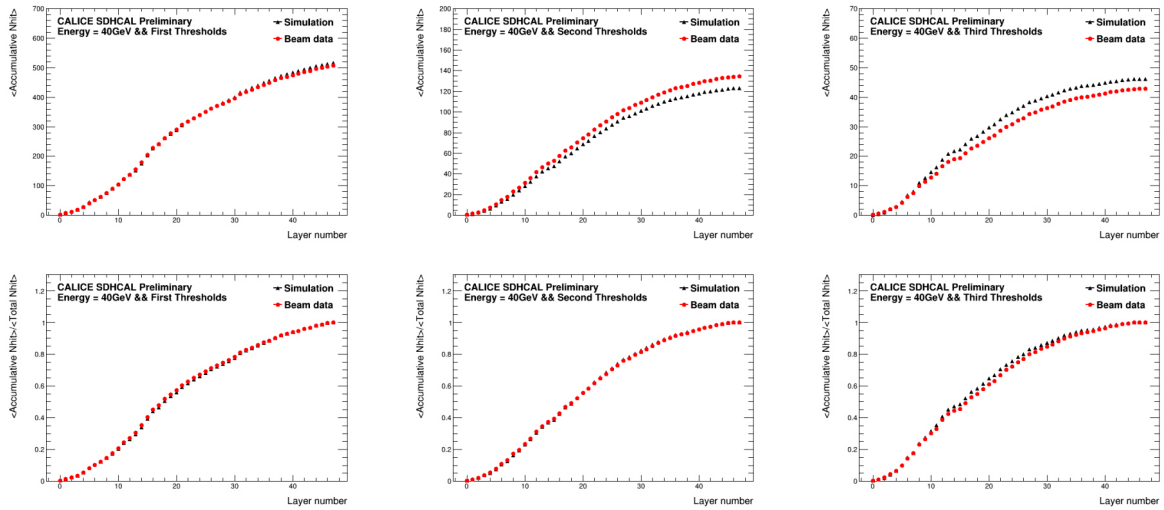


Figure 4.59: *The upper three plots show the comparison of number of hits belonging to First(left),Second(middle),Third(right) thresholds as a function of the layer number when intergating over all layers between simulation and data for the 40GeV. The three plots below show the same plots as the upper but using the total number of hits normlized to 1.*

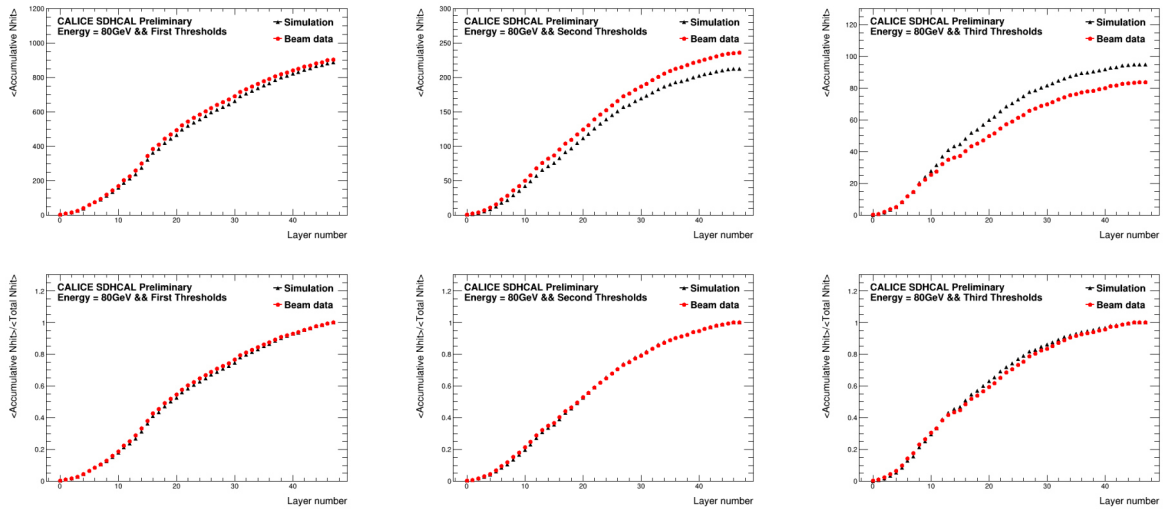


Figure 4.60: *The upper three plots show the comparison of number of hits belonging to First(left),Second(middle),Third(right) thresholds as a function of the layer number when intergating over all layers between simulation and data for the 80GeV. The three plots below show the same plots as the upper but using the total number of hits normalized.*

Chapter 5

The particle identification using boosted decision tree method in SDHCAL and the study of low beam energy data

In order to study the prototype's response for hadronic showers, new methods of selecting pion events and rejecting contamination by other particles are explored. The methods based on applying simple cut that was discussed in Chapter 4 will be referred to as the standard selection method.

A new method based on the boosted decision tree method will be discussed in this chapter. The use of such a method is possible thanks to the SDHCAL high granularity which provides detailed information of the different kinds of showers. In the 2015, the high-energy data in the range 10 to 80 GeV are taken at SPS and the low-energy data in the range 3 to 11 GeV are collected at PS. The new selection method is applied to these data. The performance (in terms of energy linearity and resolution) of the prototype using the hadronic selected data of both PS and SPS is then presented.

Content

5.1 Particle identification using boosted decision tree method in SDHCAL

5.1.1 The introduction of boosted decision tree (BDT) method

The decision tree is one of the most powerful methods for dealing with both regression and classification tasks. Therefore, it is widely used in high energy experiments. For example, it contributed to the discovery of Higgs boson at LHC. The decision tree is a special binary structure and to understand such a structure, an example of a decision tree for classification is created as shown in Fig. 5.1. Beginning from the root node of the tree (at depth 0), we have 150 samples corresponding to three class; 50 pion, 50 electron and 50 muon samples. The class of the node is determined by a majority law. Indeed the node is attributed to the class with the highest number of evens. If two or more classes have the same highest number then the class of the node is attributed to the first class according to their introduction order. Therefore, in the present root node, the class of this node is attributed to pion class (same number as the others but the first in order). In the root node, there is a condition (labelled by red text in the figure) which questions whether the radius of the event's shower is wider than 6 cm. If the answer is yes, then we go ahead to the left child node (depth 1, left) of the root node. If the child node has no further question to ask (no sub-child nodes), it will be named as a leaf node (represented by the green block in the figure) like the nodes at the depth 2. Now, if we take one event which has radius = 4.0 cm and density = 4, it will be attributed to node at left of depth 2 following the logic discussed before.

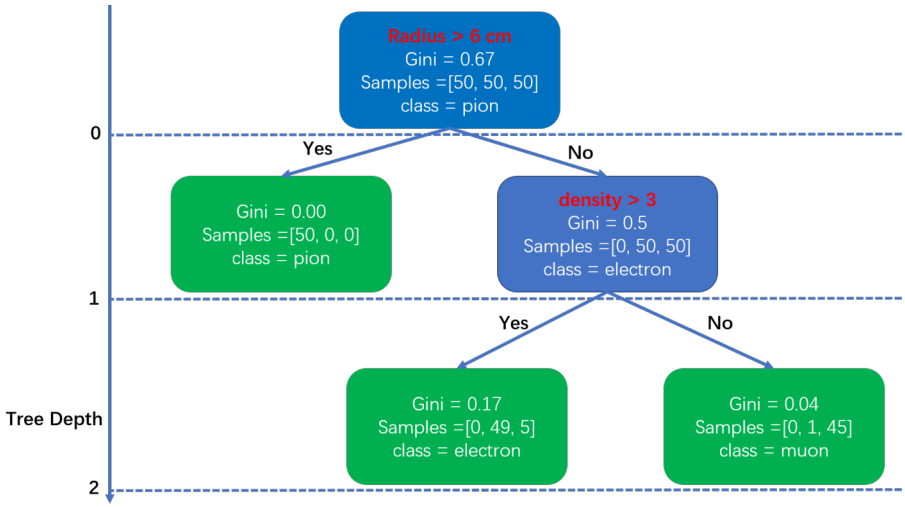


Figure 5.1: *The example tree for understanding decision tree for dealing with a classification task (to separate the pions, electrons and muons).*

For each of the nodes, A Gini value represents the impurity of the node and it can be written as:

$$G_k = 1 - \sum_{i=1}^{n} p_{k,\,i}^2\,, \tag{5.1}$$

where G_k is the Gini value of k-th node and $p_{k,i}$ is the ratio of the number of events of

class i over the total number of events in the k-th node. For example, in the node at left of depth 1, the Gini value equals to 1 - $(50/50)^2$ - $(0/50)^2$ - $(0/50)^2$ = 0. Another estimation of the impurity is the entropy which can be written as:

$$E_k = -\sum_{i=1}^{n} p_{k,\,i}\, log(p_{k,\,i}), \qquad p_{k,\,i} \neq 0 \tag{5.2}$$

Both estimations for impurity do not present a large difference in most of cases [53]. The Gini impurity spends relative less times for computation than entropy. Therefore it is suitable as the default choice. on the other side, the entropy tends to grow the trees in a more balanced way [54].

The basic algorithm for growing trees is the Classification And Regression Tree (CART) algorithm. The principle is very simple: Starting from the root node, the algorithms will split the training set as two subsets according to a feature k and its values x_k (e.g., the radius > 6 cm). For selecting the feature k and x_k, it tries to find the pair (k, x_k) which can minimize the cost function of the algorithm which can be written as:

$$L(k,\ x_k) = \frac{n_{left}}{n} G_{left} + \frac{n_{right}}{n} G_{right}$$

where $G_{left,right}$ is the Gini impurity of the left and right subsets respectively and $n_{left,right}$ is the number of events in the left and right subsets respectively. Once the best pair found, the node will split to two subsets and these two subsets are subsequently split to two subsets following the same logic, then the process can recursively continue until the maximum depth (should be defined by the users) is reached. The CART algorithm is a greedy algorithm which often finds the good solution but it cannot guarantee to find the global optimal tree which is considered as an NP-Complete problem [55].

Based on the idea of the decision tree, the boosted decision tree combines a series of weak trees to form a forest which achieves a much better performance than the simple single decision tree method. The two most widely used algorithms for boosting are **Adaptive Boost** [56] and **Gradient Boost** [57]. For our study, the adaptive boost is employed.

5.1.2 the motivation for event selection using BDT method

The SDHCAL was exposed several times to different kinds of particle beams in the CERN PS and SPS beamlines between 2012 and 2018. The energy reconstruction of hadronic showers within the SDHCAL using the associated number of fired pads with multi-threshold readout information is presented in Section. 4.2.7. The contamination of the SPS hadron beams such as electrons and muons and the absence of Cherenkov counters during data taking require the use of the event's topology to select hadronic events before reconstructing their energy. Although the rejection of muons based on the average number of hits per crossed layer is efficient, the rejection of electrons is more difficult because some hadronic showers (in particular at low energy) behave in similar way as the electromagnetic ones. To reject the electron events, the analysis presented in Section. 4.2.5 requires the shower to start after the fifth layer. Almost all of the electrons are expected to start showering before crossing the equivalent of 6 radiation lengths (X_0) [1]. Although

[1]The longitudinal depth of the SDHCAL prototype layer is about 1.2 X_0.

this selection is found to have no impact on the hadronic energy reconstruction, it keeps only hadrons that shower after about 0.6 interaction length (λ_I) of pions and thus reduces the amount of the hadronic showers available for analysis by about 54%.

To avoid the disadvantage (the loss of events) in the standard method, we explore another method to reject electron and muon contaminations that is not based on the shower start requirement and thus provide a larger sample for the energy reconstruction study. The new method is based on Boosted Decision Trees (BDT) [58, 59], a part of so-called MutiVariate Analysis (MVA) technique [60]. In the BDT, different variables associated to the topology of the event are exploited in order to distinguish between the hadronic and the electromagnetic showers, and also to identify muons including radiative ones that may exhibit a shower-like shape.

5.1.3 Monte Carlo samples and beam data samples

The SDHCAL prototype was exposed to pions, muons and electrons in the SPS of CERN in October 2015. Pion events at several energy points (10, 20, 30, 40, 50, 60, 70, 80 GeV) and muon events of 110 GeV were collected as well as electron events of 10, 20, 30, 40, 50 GeV. While the electron and muon beams are rather pure, the pion beams are contaminated by two sources. One is the electron contamination, despite the use of a lead filter to reduce their contribution. The other is the muon contamination resulting from pions decaying before reaching the prototype. All the active layers of the SDHCAL prototype were operational during the time for taking data except the layer number 34. Although this layer was physically present, its readout system was switched off due to an electronics problem when the SDHCAL was exposed to the pion beam. To apply the BDT method, six variables are selected and used in the Toolkit for MultiVariate data Analysis (TMVA) package [61] to build the decision tree.

To study the performance of the BDT method, the simulation samples are produced. For the training of the BDT, 10k events for each energy point from 10 GeV to 80 GeV with a step of 10 GeV for pions, muons and electrons were produced. In total, 160k events of pions, 160k events of electrons as well as 120k events of muons are used for this study.

The same amount of events of each species is produced and used to test the BDT method at the same time. Finally, the pure (> 99.5%) electron and muon data samples [2] are used as validation sets. The summary of data and simulation samples used for training and test is shown in Tab. 5.1

In order to render the particle identification independent of the energy of the different species and thus to extend the method applied here to a larger scope than the beam purification, the pion samples of different energies are mixed before using the BDT technique. The same procedure is applied for muon and electron samples.

5.1.4 BDT input variables

Thanks to the high granularity of the SDHCAL, we can use the MVA methods to mine the information of the shape of electromagnetic and hadronic shower to classify muons, electrons and pions. The six variables we use to distinguish hadronic showers from

[2] The purity of these samples is provided by the SPS electron and muon beams.

Particle type	Number of events
simulated pions	≈ 160k
simulated electrons	≈ 160k
simulated muons	≈ 120k
Electron data	≈ 30k
Muon data	≈ 11k

Table 5.1: *The summary of data and simulation samples (pion, electron and muon) used for training and test in BDT method.*

electromagnetic showers and from muons are described below. A common right-handed coordinate system is used throughout the SDHCAL whose 48 layers were placed perpendicular to the incoming beams. The origin of the system is defined as the center of the first of the 48 SDHCAL's layers (The x-y plane is parallel to the SDHCAL layers and referred to as the transverse plane while the z-axis runs parallel to the incoming beam).

- **First layer of the shower (Begin)** : The probability of a particle to interact in the calorimeter depends on the particle nature and the calorimeter material properties. The distribution of the coordinate z of the layer in which the first inelastic interaction takes place, follows an exponential law. It is proportional to $\exp(-\frac{z}{X_0})$ for electrons and to $\exp(-\frac{z}{\lambda_I})$ for pions, where X_0 and λ_I are the effective radiation length and nuclear interaction length for the SDHCAL material composition, respectively. To define the first layer in which the shower starts we look for the first layer along the incoming particle direction which contains at least 4 fired pads. To eliminate fake shower starts due to accidental noise or a locally high multiplicity, the following 3 layers after the first one are also required to have more than 4 fired pads in each of them. Particles crossing the calorimeter without interaction are assigned the value of 48, which is the case for most of the muons in the studied beam except the radiative ones. Figure 5.2 shows the distribution of the first layer of the shower in the SDHCAL prototype for pions, electrons and muons as obtained from the simulation and data.

- **Number of tracks segments in the shower (TrackMultiplicity)**: Applying the Hough Transform (HT) technique to single out the tracks in each event as described in Ref. [62], we estimate the number of tracks segments in the pion, electron and muon events. A HT-based segment candidate is considered as a track segment if there are more than 6 aligned hits with not more than one layer separating two consecutive hits. Electron showers feature almost no track segment while most of the hadronic showers have at least one. For muons, except for some radiative muons, only one track is expected as can be seen in Fig. 5.3.

- **Ratio of shower layers over total hit layers (NinteractingLayers/NLayers)**: This is the ratio between the number of layers in which the Root Mean Square (RMS) of the hits' position in the x-y plane exceeds 5 cm in both x and y directions and the total number of layers with at least one fired pad. It allows, as can be seen in Fig. 5.4, an

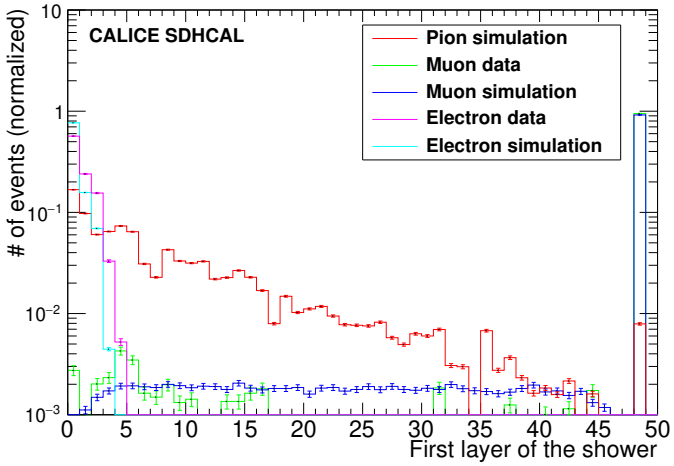


Figure 5.2: *Distribution of the first layer of the shower (Begin). Layer 0 refers to the first layer of the prototype. Layer 48 is the virtual layer after the last layer and used to tag events not fulfilling first layer criteria. In the standard method described in Section. 4.2.5, events that start showering before the fifth layer are eliminated.*

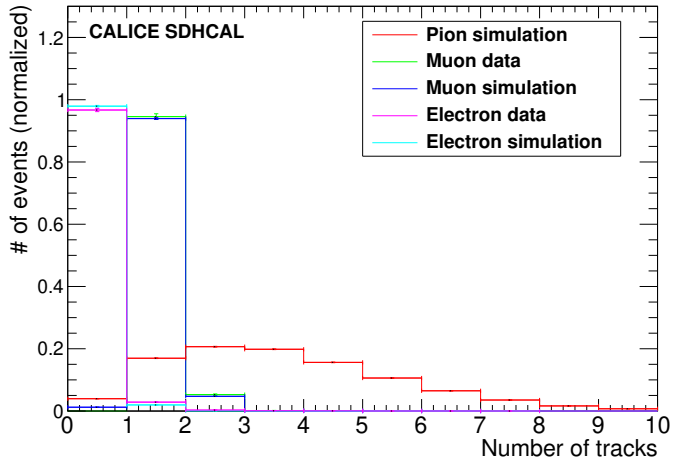


Figure 5.3: *Distribution of number of the tracks in the shower (TrackMultiplicity).*

easy discrimination of muons (even the radiative ones) from pions and electrons. It allows also a slight separation between pions and electrons.

- **Shower density (Density):** For each hit i, we count how many hits located in the 3×3 pads around it (including itself) to obtain N_i. The density is then defined as the average number of N_i following the formula: Density = $\sum_{i=1}^{N_{hit}} N_i/N_{hit}$, where N_{hit} is the total number of hits in the event. Figure 5.5 shows clearly that electromagnetic showers are more compact than the hadronic showers as expected.

- **Shower radius (Radius):** The RMS of each hit's distance from the event axis. To estimate the event axis, the average positions of the hits in each of the ten first fired

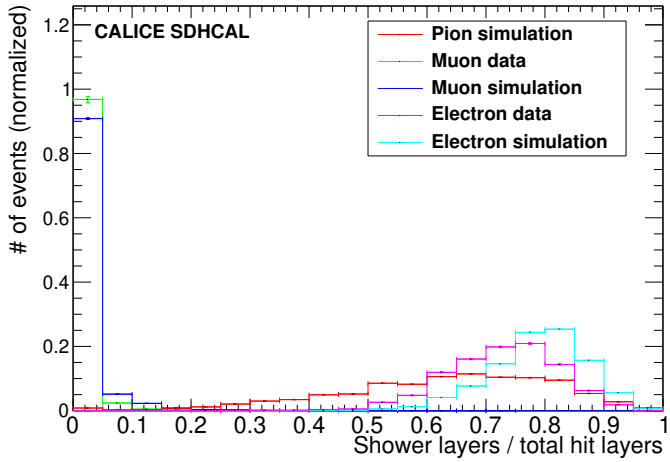


Figure 5.4: *Distribution of ratio of the number of layers in which RMS of the hits' position in the x-y plane exceeds 5 cm over the total number of fired layers (NinteractingLayers/NLayers).*

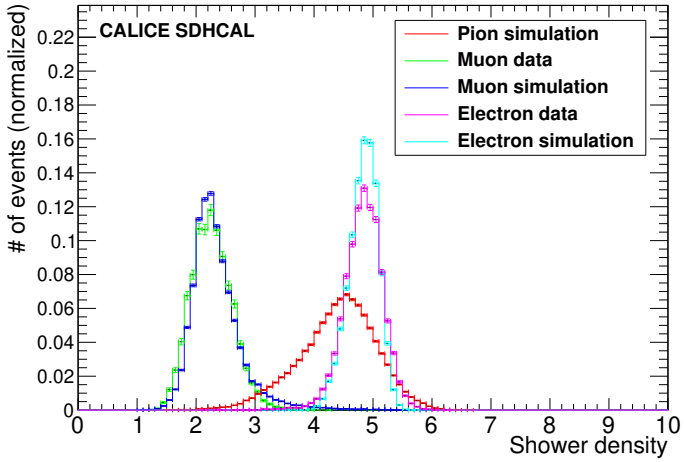


Figure 5.5: *Distribution of the average number of neighbouring hits surrounding one hit (Density).*

layers of an event are used to fit a straight line. The straight line is then used as the event axis. Figure 5.6 shows the average radius of the three particle species in the SDHCAL. Discrepancy of the muon radius distribution between data and simulation is due to the difference of hit multiplicity which is slightly larger in data with respect to simulation.

- **Shower maximum position (Length):** This is the distance between the shower start and the layer where the maximum RMS of hit transverse coordinates with respect to the shower axis is detected. The distribution of this variable for different particle species is shown in Fig. 5.7.

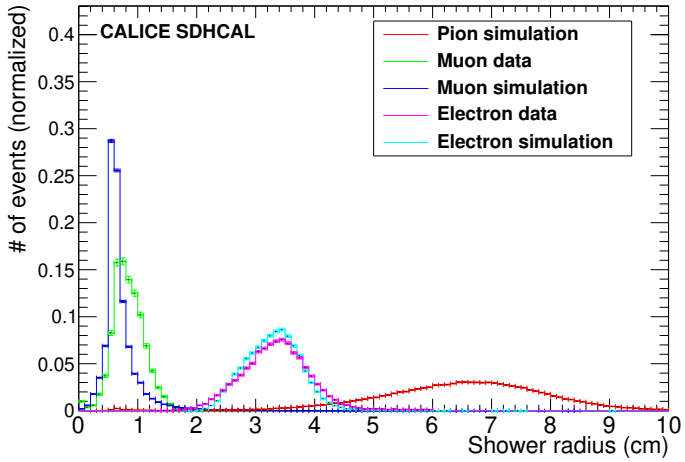


Figure 5.6: *Distribution of the average radius of the shower (Radius).*

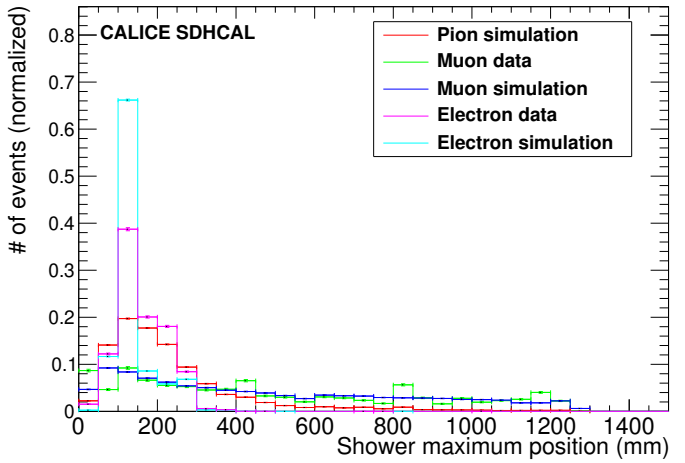


Figure 5.7: *Distribution of the position of the layer with the maximum radius (Length).*

Before using the variables listed above as input to the BDT method, we check that the variables distributions in the simulation are in agreement with data for the muon and electron beams which are quite pure. Figures 5.2 - 5.7 show that there is globally a good agreement for the six variables of the two species even though the agreement is not perfect in particular for electrons. The observed discrepancy is related to the difficulty to simulate precisely the saturation effect of electromagnetic showers in RPC detectors as explained in Ref.[49].

5.1.5 The two approaches to build the BDT-based classifier

In order to take into account the difference observed in some variable distributions between data and simulation, and to cross-check the particle identification using the BDT method, we adopt two different training strategies for the BDT-based classifier. The first

approach, referred to as MC Training, uses simulation samples of pions, electrons and muons as training sets. The second, referred to as Data Training, uses simulation samples of pions but electron and muon samples taken from data as training sets.

5.1.5.1 MC Training Approach

The six variables of the simulated pion, muon and electron events are used for the training and testing of the classifier. Events are chosen in alternating turns for the training and test samples as they occur in the source trees until the desired numbers of training and test events are selected. The training and test samples contain the same number of events for each event class. Independent samples of signal events (pions) and of the different background contributions (electron and muons) are used. The ratio between signal and each background (electron or muon) events is 1 for training and test samples. After the training, the BDT provides the relative weight of each variable as a measure of distinguishing signal from background. Two BDT-based classifiers are proposed here. The first ($BDT_{\pi\mu}$) is used to discriminate pions against muons and the second ($BDT_{\pi e}$) to discriminate against electrons. Table 5.2 shows the variable ranking according to their separation power in the $BDT_{\pi\mu}$ while Tab. 5.3 gives their separation power in the case of $BDT_{\pi e}$. The BDT algorithm using the variables and their respective weights is then applied to the test samples. The output of the BDT applied to each of the test sample events is a variable belonging to the interval [-1,1] with the positive value representing more signal-like events and the negative more background-like events.

Rank : Variable	Variable relative weight
1 : Length	0.233
2 : Density	0.225
3 : NInteractinglayer/Nlayer	0.163
4 : Radius	0.160
5 : Begin	0.139
6 : TrackMultiplicity	0.080

Table 5.2: *Variable ranking of separation power in the case of $BDT_{\pi\mu}$.*

Rank : Variable	Variable relative weight
1 : Radius	0.204
2 : NInteractinglayer/Nlayer	0.203
3 : Density	0.194
4 : Length	0.151
5 : Begin	0.145
6 : TrackMultiplicity	0.101

Table 5.3: *Variable ranking of separation power in the case of $BDT_{\pi e}$.*

Figure 5.8 (left) shows the output of the BDT for a test sample made of pions and muons

while Fig. 5.8 (right) shows the output for a test sample made of pions and electrons. The values differ significantly for signal and background suggesting thus a large separation power of the BDT approach. This is confirmed by Fig. 5.9. The pion selection efficiency versus the muon (electron) rejection of the test sample is shown in Fig. 5.10 (left) and Fig. 5.10 (right), respectively. A pion selection efficiency exceeding 99.0% with a muon and electron rejection of the same level (> 99.0%) can be achieved.

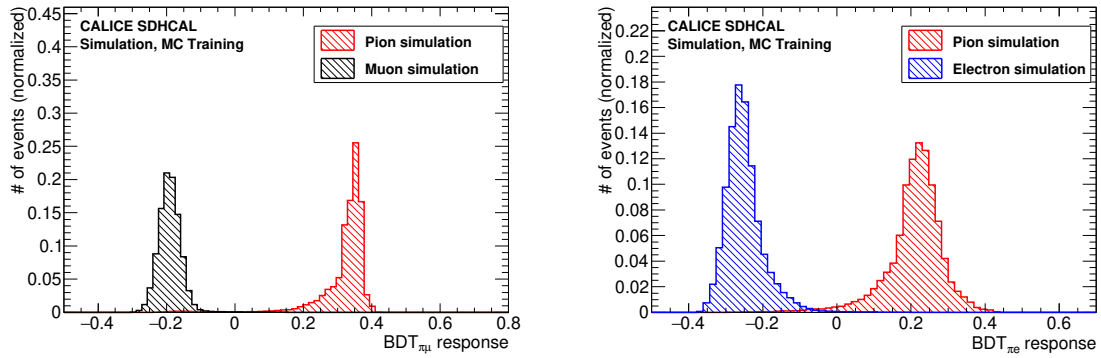


Figure 5.8: *The BDT output of the BDT$_{\pi\mu}$ (left) and BDT$_{\pi e}$ (right) built with simulation samples.*

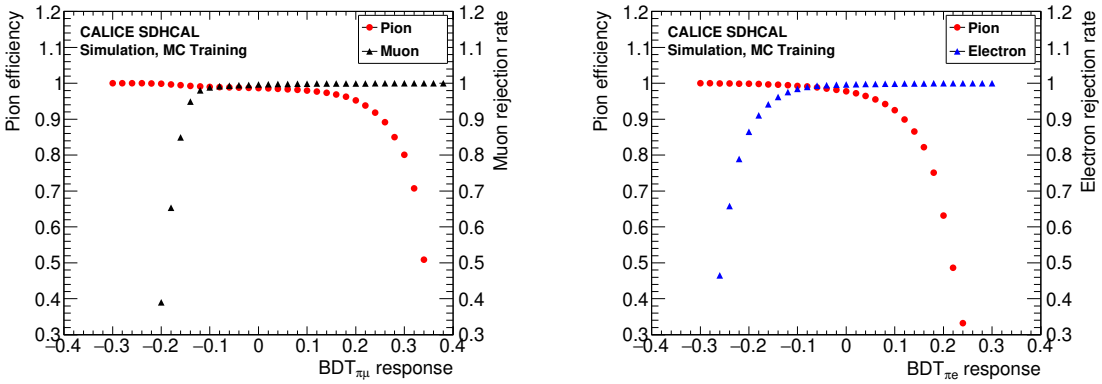


Figure 5.9: *Pion efficiency and muon rejection rate (left) and pion efficiency and electron rejection rate (right) as a function of the BDT output.*

In order to check the validity of these two classifiers, we use the beam samples of pure muons and electrons. Figure 5.11 (left) shows the BDT output of BDT$_{\pi\mu}$ and Fig. 5.11 (right) shows the case of BDT$_{\pi e}$. Beam muon results show a good agreement with respect to the simulated events. A shift of the beam electron shape is observed with respect to the one obtained from the simulated events. This difference is most probably due to the fact that the distribution of some variables in data and in the simulation are not identical. Next, as a first step in purifying the collected hadronic data events we apply the pion-muon classifier. Figure 5.11 (left) shows the BDT$_{\pi\mu}$ response applied to the collected hadron events in the SDHCAL. We can clearly see that there are two maxima.

One maximum in the muon range corresponds to the muon contamination of pion data and another one in the pion range. Hence, to ensure the rejection of the muons in the sample, the BDT variable is required to be > 0.1. The second step is to apply the $BDT_{\pi e}$ to the remaining of the pion sample. Figure 5.11 (right) shows the $BDT_{\pi e}$ output. In order to eliminate the maximum of the electrons contamination and get almost a pure (> 99.5%) pion sample with limited loss of pion events, we apply to the pion samples a $BDT_{\pi e}$ cut of 0.05.

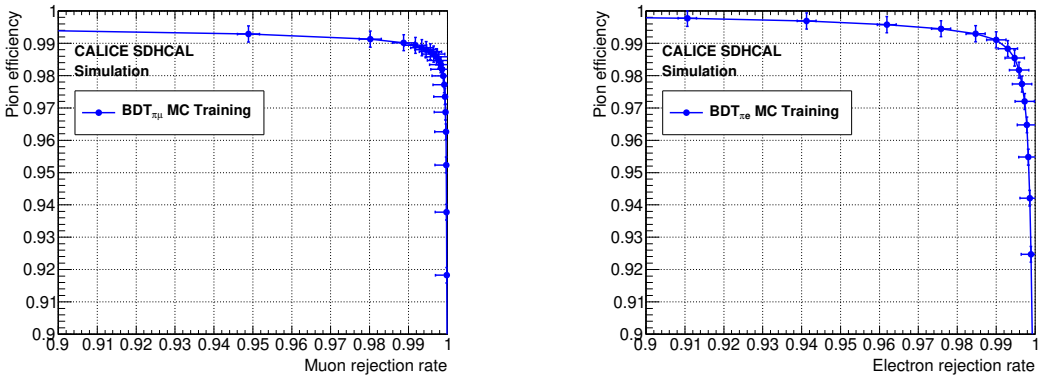


Figure 5.10: *Pion efficiency versus muon rejection rate(left) and pion efficiency versus electron rejection rate (right).*

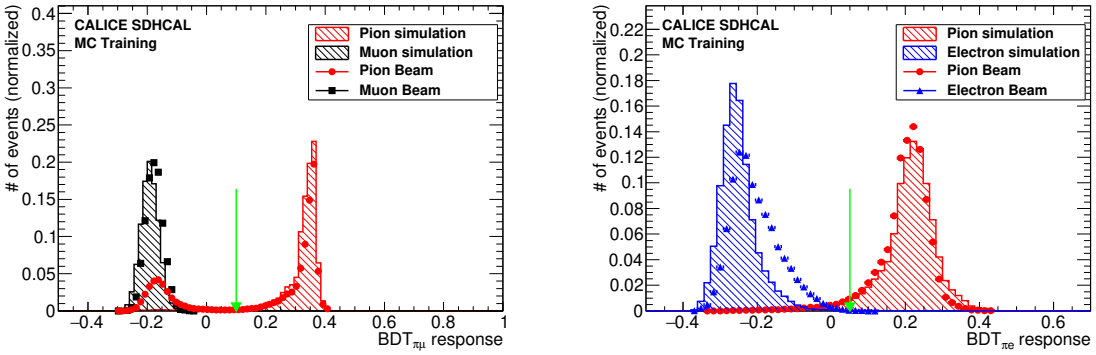


Figure 5.11: *The BDT output after using the $BDT_{\pi\mu}$ on the data pion sample (left) and the BDT output after using the $BDT_{\pi e}$ on the same data pion sample after classified by $BDT_{\pi\mu}$ (right). A green arrow is shown on both to indicate the BDT cut applied to clean the pion samples.*

5.1.5.2 Data Training Approach

We use the same variables of the MC Training approach on the data samples of muons (11k events) and electrons (30k events) but still on the simulated pion samples to build two classifiers. Then we apply the same procedure as the MC Training approach. Table 5.4

and 5.5 show the corresponding variables ranking for $BDT_{\pi\mu}$ and $BDT_{\pi e}$ according to their separation power importance.

Rank : Variable	Variable relative weight
1 : Length	0.300
2 : Radius	0.230
3 : Density	0.227
4 : Begin	0.103
5 : NInteractinglayer/Nlayer	0.080
6 : TrackMultiplicity	0.060

Table 5.4: *Variable ranking of separation importance in the case of $BDT_{\pi\mu}$.*

Rank : Variable	Variable relative weight
1 : Radius	0.195
2 : NInteractinglayer/Nlayer	0.191
3 : Density	0.189
4 : Length	0.151
5 : Begin	0.141
6 : TrackMultiplicity	0.131

Table 5.5: *Variable ranking of separation importance in the case of $BDT_{\pi e}$.*

The difference of variables weights of these two tables with respect to those obtained with MC training approach is explained by the slight difference of some variables distributions between data and simulation. Indeed, when dropping, in the BDT method, the variables for which the relative large discrepancy between data and simulation is present, namely the "radius" variable, similar weights are obtained for the remaining variables in the two approaches as shown in Tab. 5.6 and Tab. 5.7 for data training and MC training approach respectively.

Rank : Variable	Variable relative weight
1 : Density	0.263
2 : Length	0.230
3 : NInteractinglayer/Nlayer	0.186
4 : Begin	0.184
5 : TrackMultiplicity	0.137

Table 5.6: *Variable ranking of separation importance in the case of $BDT_{\pi\mu}$ for data training approach when drop the variable "radius".*

Figure 5.12 left (right) gives the results of pion efficiency and muon (electron) rejection rate. The left (right) plot of Fig. 5.13 shows the BDT output of the $BDT_{\pi\mu}$ ($BDT_{\pi e}$). Clearly

Rank : Variable	Variable relative weight
1 : Density	0.250
2 : Length	0.235
3 : NInteractinglayer/Nlayer	0.220
4 : Begin	0.184
5 : TrackMultiplicity	0.111

Table 5.7: *Variable ranking of separation importance in the case of $BDT_{\pi\mu}$ for MC training approach when drop the variable "radius".*

these two classifiers have very good separation power. We apply these classifiers to the raw pion beam samples. The results can be seen in Fig. 5.14. We apply a BDT cut value of 0.2 in the pion-muon separation stage and then a BDT cut value of 0.05 in the pion-electron separation stage.

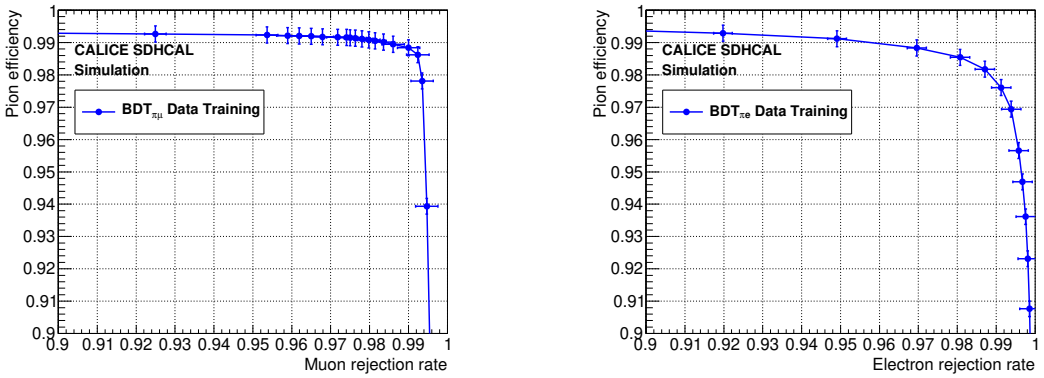


Figure 5.12: *Pion efficiency versus muon rejection rate (left) and pion efficiency versus electron rejection rate (right).*

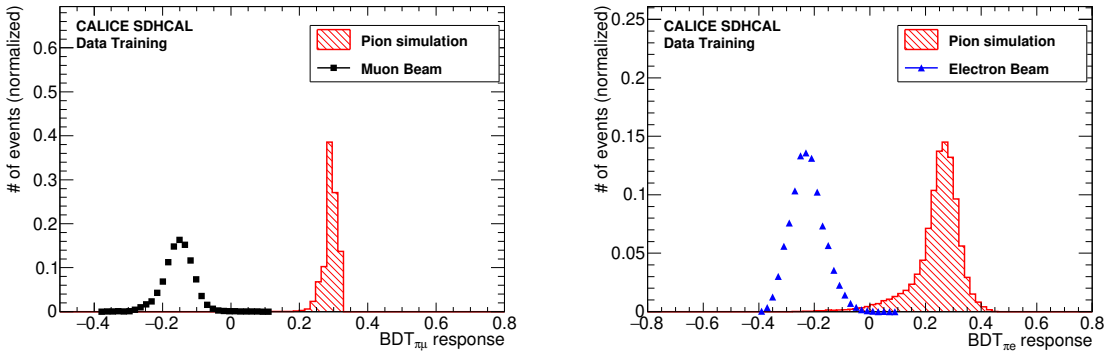


Figure 5.13: *BDT output of the BDT$_{\pi\mu}$ built with pure beam muons and simulated pion samples (left) and of the BDT$_{\pi e}$ built with pure beam electrons and simulated pion samples (right)*

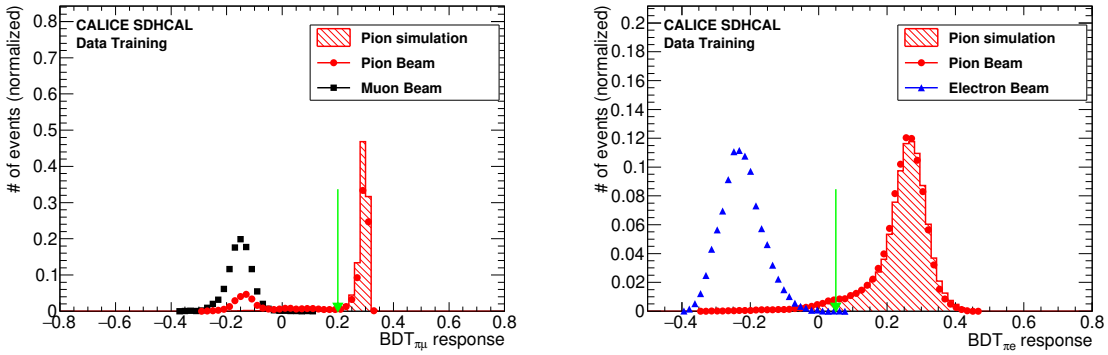


Figure 5.14: *The BDT output after using the BDT$_{\pi\mu}$ on the data pion sample (left) and the BDT output after using the BDT$_{\pi e}$ on the same pion sample after classified by BDT$_{\pi\mu}$ (right). A green arrow is shown on both to indicate the BDT cut applied to clean the pion samples.*

5.1.6 evaluation of two training approaches

The distributions of input variables for the data and simulation events of pion, muon and electron are shown in Fig. 5.15. Only the pion data sample distributions are obtained after applying the data-based BDT classifiers. A good agreement between the data and simulation events for pions is observed. This confirms the power of the BDT method. To compare the events selected by standard method and BDT method, the distributions of input variables for both two method are shown in Fig. 5.16.

The muon rejection rate obtained with the MC (Data) Training approach is found to be 99.9% (99.0%) and that of the electron is 99.8% (99.4%) respectively. To quantify the level of systematic uncertainty of pion efficiecny and background rejection rate, the difference in pion efficiency and background rejection rate between simulation and data for two training approaches are summarized in Tab. 5.8 (MC Training) and Tab. 5.9 (data Training). The difference between data and simulation using the two BDT training

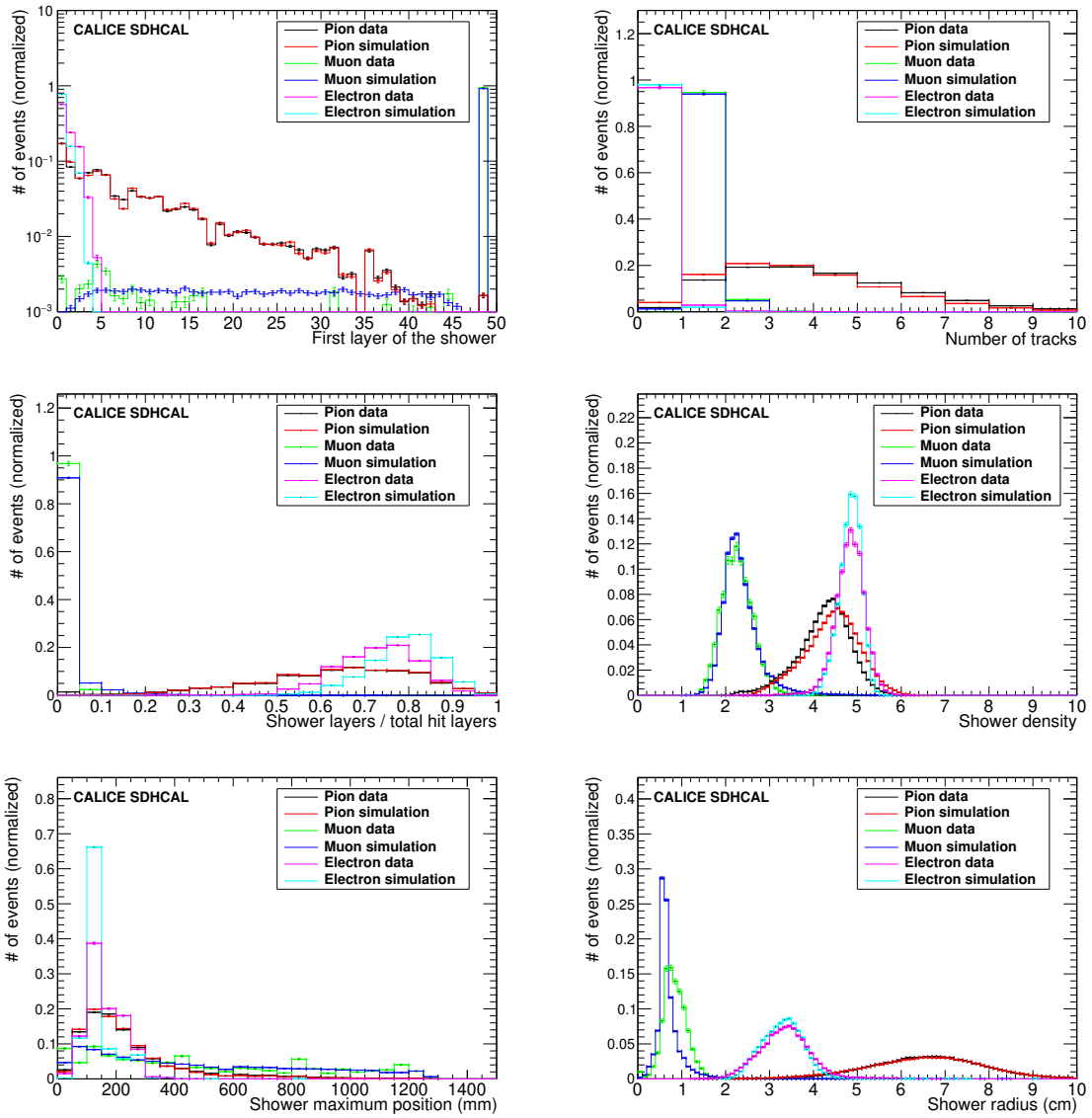


Figure 5.15: *Distributions of six input variables of electron, muon and pion samples. Pion distributions are obtained from pion data samples after applying the data-based training BDT selections.*

approches for different BDT cuts in the range of [0.05,0.25] in the case of the pion-muon separation shows a value smaller than 0.2% in both the pion efficiency and the muon rejection rate. In the case of the pion-electron separation, different BDT cuts in the range of [0.05, 0.15] result in a difference of less than 1.2% in the pion efficiency and less than 0.2 % in the electron rejection rate.

The rejection of muons and electrons presented in the pion data sample using the BDT allows us to have a rather pure pion sample as explained in the previous section. Figure 5.17 shows the results of comparison in event selection between the standard method and the BDT-based method using the simulation samples. For both simulation and beam data, the BDT method leads to a larger pure sample of hadrons comparing

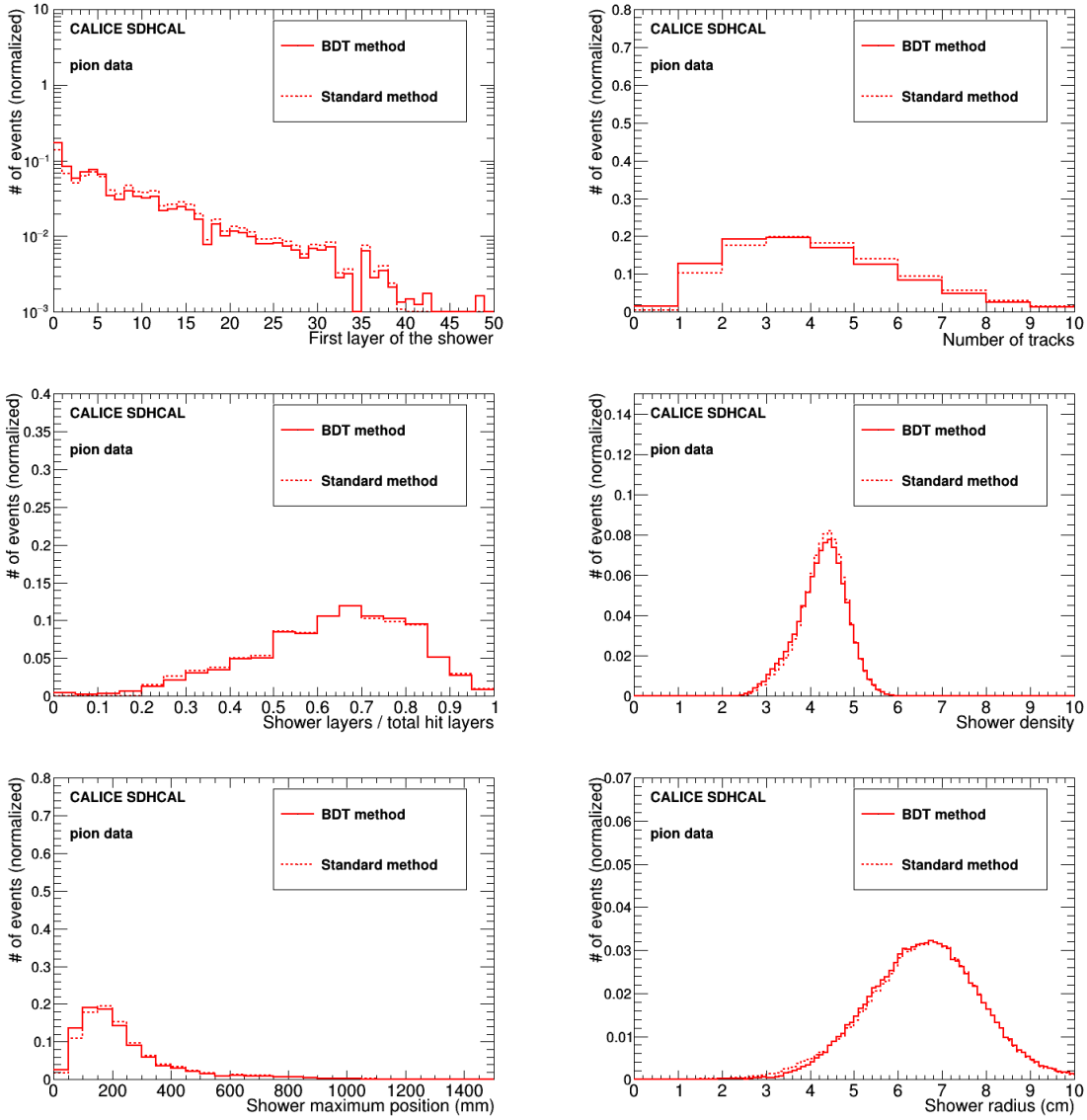


Figure 5.16: *Distributions of six input variables of electron, muon and pion samples which are selected by standard method (red dashed lines) and BDT method (red solid lines).*

to the standard method [46] in particular at low energy as shown in Fig. 5.18 for the comparison of the selected events as a function of the total number of hits for the 10 GeV pion beam data.

MC Training pion-muon separation		
Cut value	Systematic uncertainty of pion efficiency (%)	Systematic uncertainty of muon rejection rate (%)
0.05	0.2	0.2
$0.10^{(*)}$	0.2	0.1
0.15	0.1	0.1
MC Training pion-electron separation		
Cut value	Systematic uncertainty of pion efficiency (%)	Systematic uncertainty of electron rejection rate (%)
0.00	0.1	0.2
$0.05^{(\dagger)}$	0.4	0.1
0.10	1.2	0.1

Table 5.8: *The systematic uncertainty of pion efficiency, muon and electron rejection rate for MC training approach. (∗): the work point for pion-muon separation. (†): the work point for pion-electron separation.*

Data Training pion-muon separation		
Cut value	Systematic uncertainty of pion efficiency (%)	Systematic uncertainty of muon rejection rate (%)
0.15	0.1	0.2
$0.20^{(\star)}$	0.1	0.1
0.25	0.1	0.1
Data Training pion-electron separation		
Cut value	Systematic uncertainty of pion efficiency (%)	Systematic uncertainty of electron rejection rate (%)
0.00	0.2	0.1
$0.05^{(\ddagger)}$	0.4	0.1
0.10	0.6	0.0

Table 5.9: *The systematic uncertainty of pion efficiency, muon and electron rejection rate for Data training approach. (⋆): the work point for pion-muon separation. (‡): the work point for pion-electron separation.*

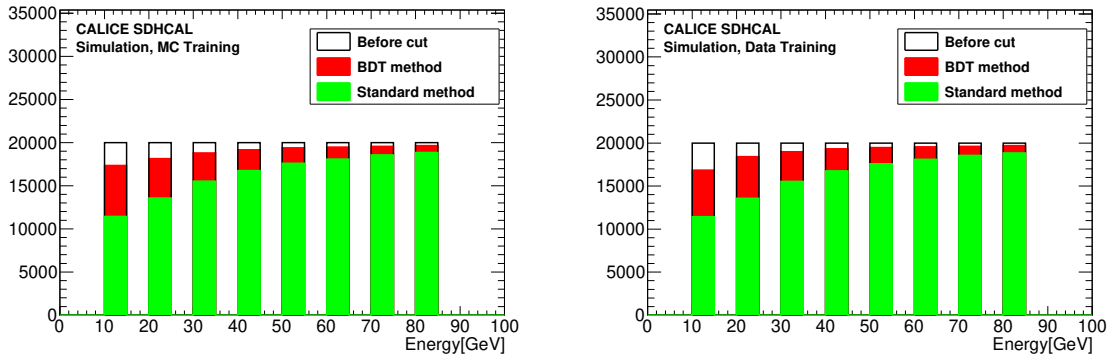


Figure 5.17: *The number of simulated events of different energy points from 10 GeV to 80 GeV before (white) and after applying the standard method (green) or BDT method (red). The left plot shows the results from BDT method with MC Training approach while the right one shows the results with Data Training approach.*

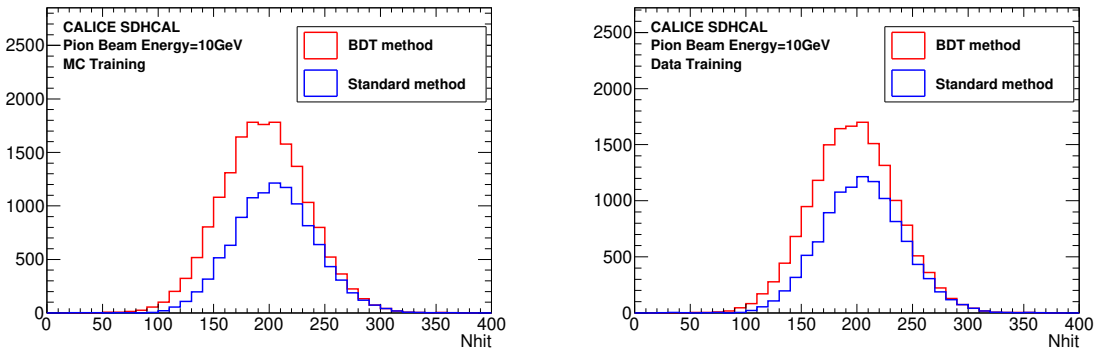


Figure 5.18: *Distribution of the total number of hits for the 10 GeV pion beam data selected by the standard method (blue) and the BDT method (red). The left plot shows the results from BDT method with MC Training approach while the right one shows the results with Data Training approach.*

5.1.7 Energy reconstruction

In order to check the validity of the BDT method, the same energy reconstruction technique presented in Section. 4.2.7 is applied to the pion samples selected with the BDT method as well as to the one selected by standard method. The same parameterization is used to estimate the pion energy of the samples selected by both methods. The reconstructed energy and associated energy resolution are obtained by fitting the energy distribution using Crystal Ball function that takes into account the tail due to shower leakage. The reconstructed energy distributions of 10, 40 and 80 GeV for simulated pion samples selected by two methods are shown in Fig. 5.19 - Fig. 5.21. Compared with simulation samples, the reconstructed energy distributions of 10, 40 and 80 GeV for pion beam samples are shown in Fig. 5.22 - Fig. 5.24.

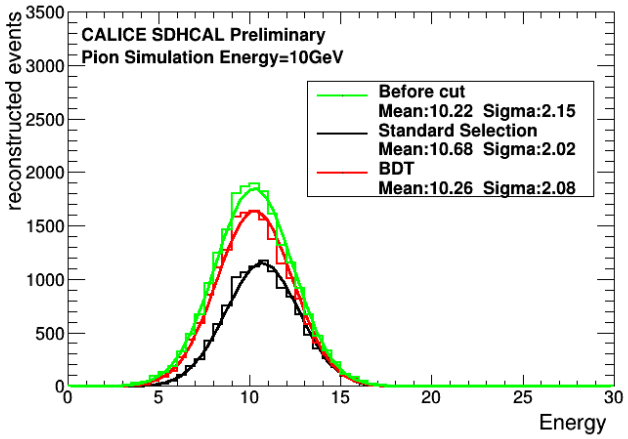


Figure 5.19: *The comparison of reconstructed energy between BDT and standard methods using simulation samples for 10 GeV. The distributions are fitted by Crystal Ball functions.*

Figure.5.25 (left) shows the energy reconstructed as well as the deviation with respect to the beam energy using the BDT method as well as the standard selection. In the Fig.5.25 (right), it is the comparison of energy resolution between standard selection and BDT method. We do not observe any significant deviation of energy resolution when applying the standard energy reconstruction on the pion events selected by the BDT method. Similar results are obtained with the both methods but using BDT we can get results with smaller statistical uncertainties than standard selection of Ref. [46].

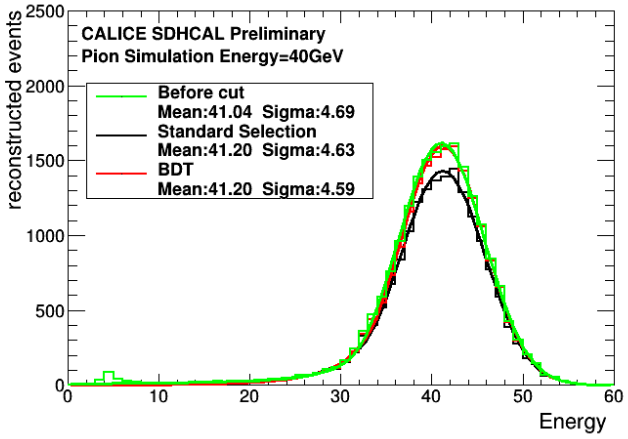


Figure 5.20: *The comparison of reconstructed energy between BDT and standard methods using simulation samples for 40 GeV. The distributions are fitted by Crystal Ball functions.*

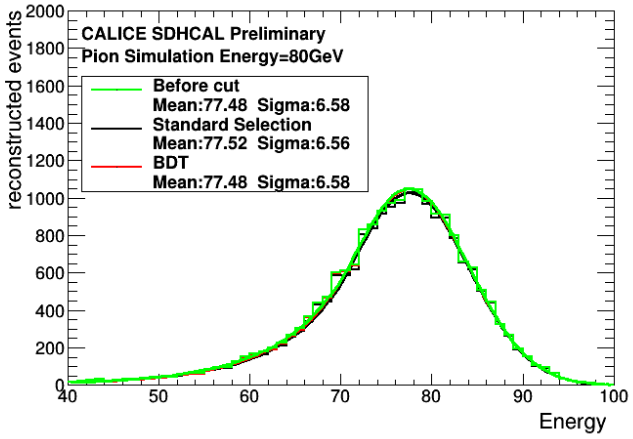


Figure 5.21: *The comparison of reconstructed energy between BDT and standard methods using simulation samples for 80 GeV. The distributions are fitted by Crystal Ball functions.*

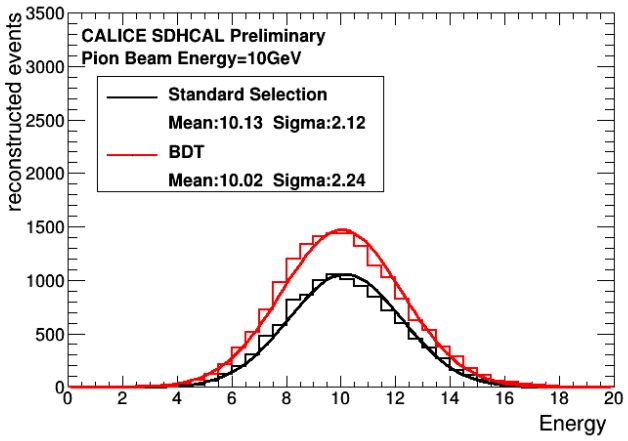


Figure 5.22: *The comparison of reconstructed energy between BDT and standard methods using pion beam samples for 10 GeV. The distributions are fitted by Crystal Ball functions.*

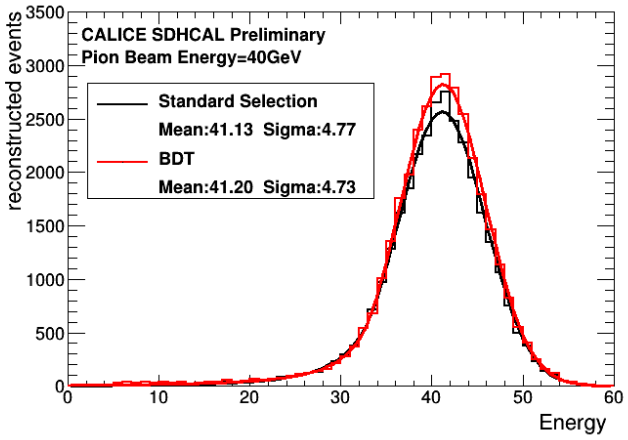


Figure 5.23: *The comparison of reconstructed energy between BDT and standard methods using pion beam samples for 40 GeV. The distributions are fitted by Crystal Ball functions.*

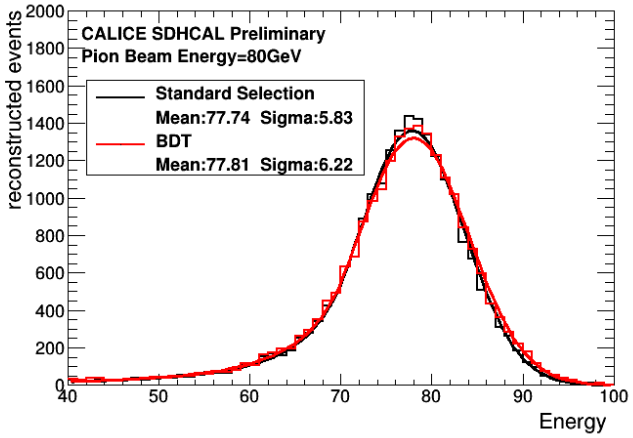


Figure 5.24: *The comparison of reconstructed energy between BDT and standard methods using pion beam samples for 80 GeV. The distributions are fitted by Crystal Ball functions.*

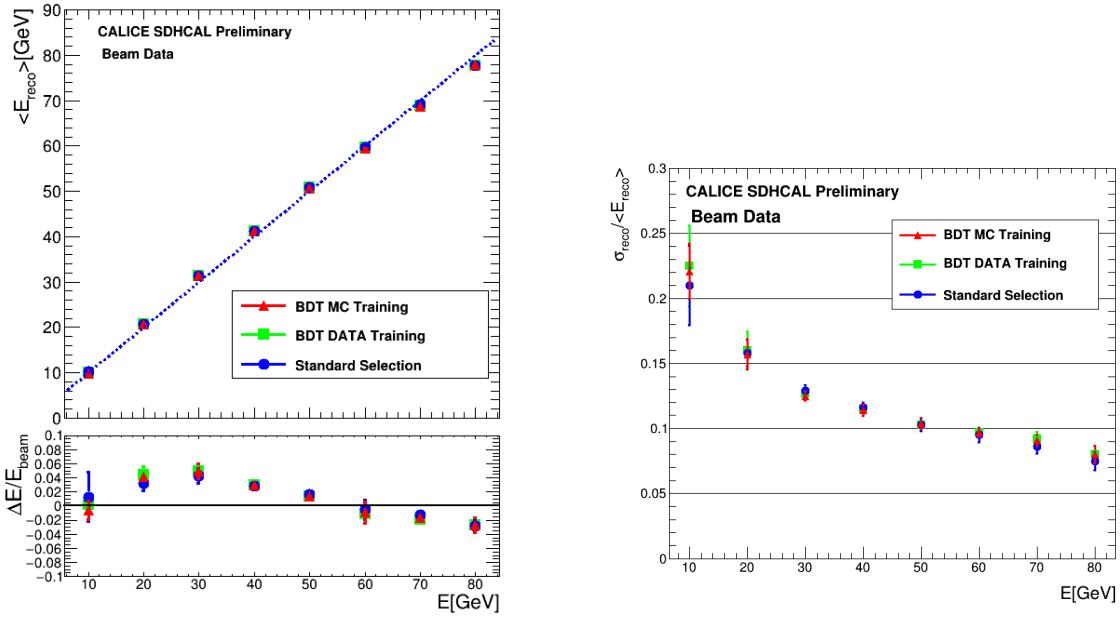


Figure 5.25: *Mean reconstructed energy for pion showers as a function of the beam energy as well as relative deviation of the pion mean reconstructed energy with respect to the beam energy (left) and resolution of the reconstructed hadron energy as a function of the beam energy (right). Both statistical and systematic uncertainties are included in the error bars.*

5.1.8 Uncertainties estimation for energy linearity and resolution

The linearity and energy resolution results presented previously include statistical and systematic uncertainties. We present here after the main contributions to the systematic uncertainties:

- The difference of the estimated energy before and after applying the selection criteria (BDT or standard selection) is evaluated using simulation samples of pions from 10 GeV to 80 GeV with 10 GeV energy step. The difference is used as one source of the systematic uncertainties.

- To account for the difference in shape of the hadronic showers that are found to be sparser in the data than in the simulation [52], the difference of reconstructed energy estimated using data samples on the one hand and the simulation samples on the other hand is considered as another source of systematic uncertainties. It is worth mentioning here that this uncertainty contribution is the main contribution to the large uncertainty observed at 10 GeV.

- For the standard selection, using all energy points data simples, each of the different selection criteria is varied by an arbitrary 5% in both directions with respect to the nominal values when this is possible. The maximum deviation with respect to the nominal value is used as the third source of systematic uncertainties in the case of the standard selection. For the BDT using MC training method, the BDT cut value is changed from 0.10 to 0.0 in pion-muon separation step and from 0.05 to 0.0 in pion-electron separation. The difference in energy of these two steps is added quadratically and taken as the third source of systematic uncertainties. For the BDT using data training, the same procedure is applied.

  By applying the BDT cut one may eliminate some of the pions that have an electron-like shape. To estimate such a possible bias, the energy of the simulated pion events is reconstructed without any selection and then by applying several values of the BDT cut. Figure 5.26 shows the reconstructed energies and the relative resolutions are not impacted except at 10 GeV where the energy resolution is slightly improved by applying the BDT cut. Since the electron-like pions are not limited to low energy, one may conclude that the BDT selection does not disfavor, in principle, the electron-like pions even though there is a slight difference at 10 GeV. The latter is duly included in the systematic uncertainties.

Although the statistical uncertainties are found to be negligible for almost all the runs with respect to systematic uncertainties, their contributions as well as the systematic uncertainties previously discussed are added quadratically to obtain the final uncertainties. The results are summarized in Table 5.10 and in Table 5.11.

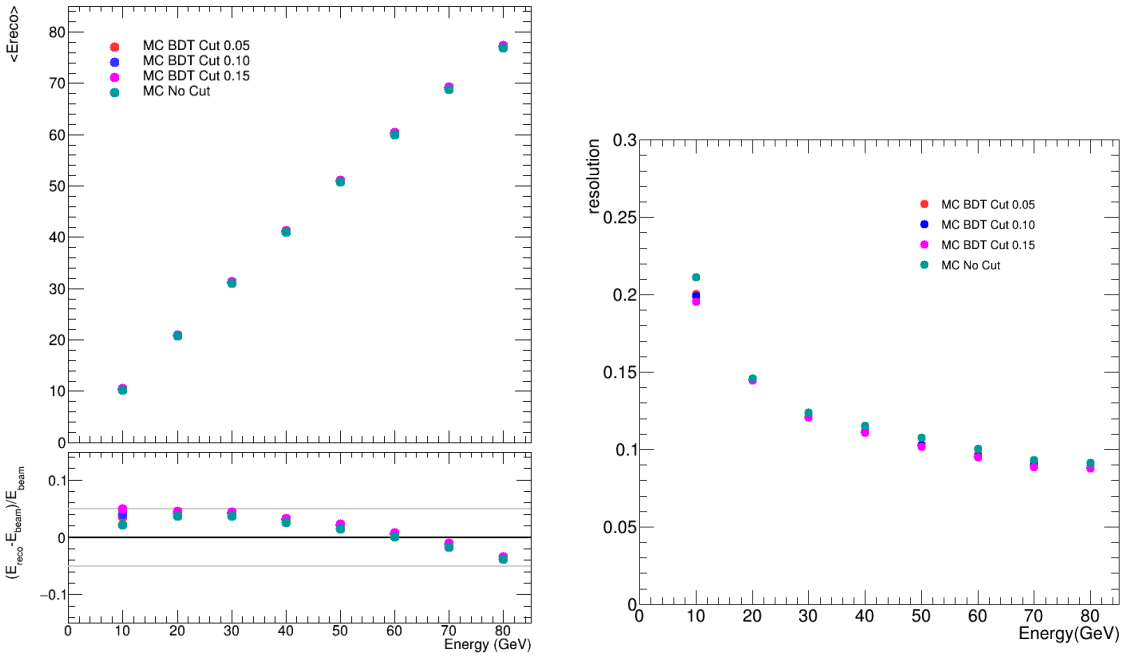


Figure 5.26: *Mean reconstructed energy for simulated pion showers as a function of the simulated energy as well as relative deviation of the pion mean reconstructed energy with respect to the simulated energy (left) and resolution of the reconstructed hadron energy as a function of the simualted energy (right) in different BDT cut value.*

Energy(GeV)	MC training	Data training	Standard selection
10	-0.006 ± 0.013	0.002 ± 0.016	0.013 ± 0.047
20	0.041 ± 0.011	0.044 ± 0.018	0.032 ± 0.013
30	0.049 ± 0.013	0.049 ± 0.015	0.043 ± 0.011
40	0.029 ± 0.004	0.030 ± 0.006	0.028 ± 0.006
50	0.014 ± 0.008	0.014 ± 0.008	0.017 ± 0.010
60	-0.009 ± 0.015	-0.011 ± 0.016	-0.005 ± 0.015
70	-0.017 ± 0.006	-0.019 ± 0.008	-0.012 ± 0.007
80	-0.027 ± 0.011	-0.027 ± 0.011	-0.028 ± 0.010

Table 5.10: *List of $\frac{\Delta E}{E}$ observed and associated uncertainties.*

Energy(GeV)	MC training	Data training	Standard selection
10	0.221 ± 0.021	0.225 ± 0.031	0.210 ± 0.031
20	0.157 ± 0.011	0.160 ± 0.015	0.158 ± 0.013
30	0.125 ± 0.013	0.127 ± 0.006	0.129 ± 0.011
40	0.114 ± 0.004	0.115 ± 0.003	0.116 ± 0.006
50	0.104 ± 0.004	0.103 ± 0.004	0.103 ± 0.010
60	0.097 ± 0.003	0.097 ± 0.003	0.095 ± 0.015
70	0.091 ± 0.004	0.092 ± 0.005	0.086 ± 0.007
80	0.080 ± 0.007	0.080 ± 0.007	0.075 ± 0.010

Table 5.11: *List of energy resolution observed and associated uncertainties.*

5.2 The study of beam data in the low energy range

5.2.1 Simulation and Beam data samples

In the year of 2015, The SDHCAL prototype was exposed to beams of hadrons, electrons and muons on the beamlines of PS in the CERN. In this period, the negatively charged pion beams of 3, 4, 5, 6, 7, 8, 9, 10 and 11 GeV were taken. We also collect the positively hadrons including the 10, 20, 30, 40, 50, 60, 70 and 80 GeV in the beamlines of SPS. Due to the analysis of 10 to 80 GeV beam data already already presented in Section. 5.1, this section we only focus on the beam pion data in the low energy range.

The simulation samples of pion, electron, muon for energy range from 3 GeV to 11 GeV are generated with 200k events for each energy point.

5.2.2 Pion events selection

In the low-energy range, the pion beam samples are also contaminated by two kinds of particles: electrons and muons. The muon contamination exists in all samples from 3 to 11 GeV. It include two different type of muons: cosmic muons and beam muons whcih are generated by pions decaying before arriving at prototype. Concerning the electron contamination, it is negligible in pion samples from 6 to 11 GeV since the successful use of electron eliminators but for the pion samples from 3 to 5 GeV, the electron contamination may be present. Inspired by the Section. 5.1, we propose to use the BDT method to reject the electron background.

5.2.2.1 Muon contamination rejection

One of the main contamination in our beam data is muons including beam muons and cosmic muons. To eliminate these two kinds of muons, we are inspired by the information based on different behavior in the prototype for muons and pions. Basically, muons cross the prototype and only leave a straight track in the prototype like Figure 5.27, so the mean of hits distance (desribed by the variable: meanRadius) of muons with respect to the shower axis is less than 1.5 cm ($\approx$ 1.5 pads) shown in Figure 5.28.

To eliminate most of the muon contamination, we required that the meanRadius is greater than 2 cm. To check the rejection power of muon cut, it was applied to the muon

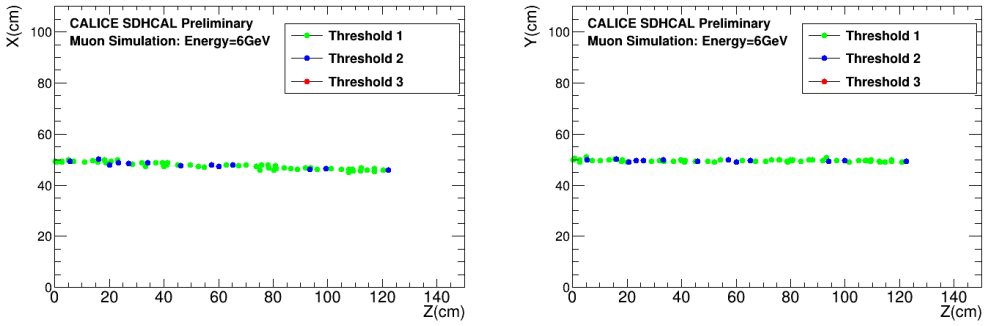


Figure 5.27: *Event display of the 6 GeV simulated muons projected in the x-z (left figure) and y-z (right figure) planes. The green, blue and red color indicate the hits associated with first, second and third threshold respectively.*

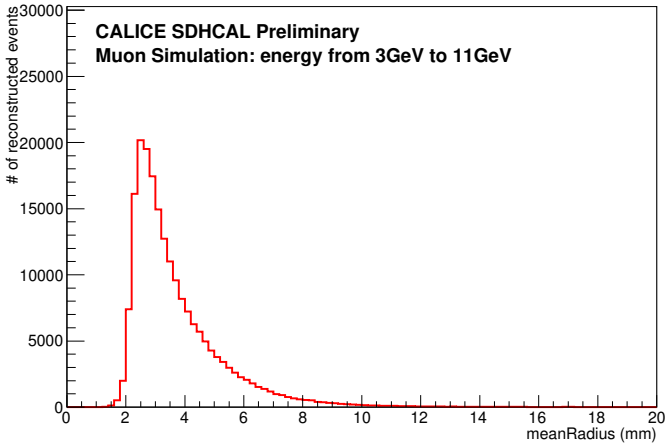


Figure 5.28: *Distribution of the meanRadius of shower by 3 GeV to 11 GeV muons as given by the simulation.*

runs. The Figure 5.29 shows the distribution of number of hits before and after muon rejection for 20 GeV muon runs. It shows clearly the rejection power of this selection which is higher than 99.0%.

5.2.2.2 Electron contamination rejection using Boosted Decision Tree (BDT)

Like the previous section discussed, for the 6 to 11 GeV pion runs, the electron contamination is negligible thanks to the use of electron eliminator. However for the 3 to 5 GeV pion runs, the electrons may exist. Therefore, it is necessary to check the electron contamination and to eliminate it if it is not negligible. Based on advantage of the high granularity of our prototype, we can use BDT method to mine the information of the shape of electromagnetic shower and hadrnoic shower to classify the electron and pion.

We finally choose eight features based on shower topology as the input of BDT model which can help to distinguish electrons and pions. The five variables of eight variables are

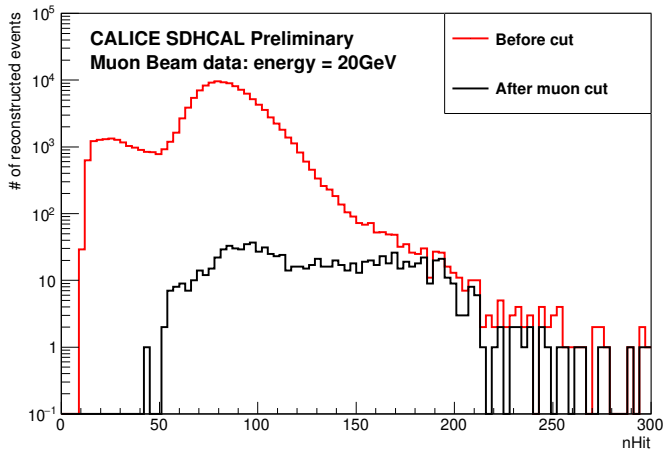


Figure 5.29: *Distribution of the number of hits for 20 GeV muon run before (red) and after (black) muon cut.*

already discussed in the Section 5.1, thus only extra three new variables will be described below.

- **Number of clusters of the shower (nCluster)**: All hits in a given layer are clustered using a nearest-neighboring algorithm described in Section. 4.2.2.1. It consists in merging in each GRPC plate the hits sharing a common edge. This variable defines the number of clusters of the shower shown in Fig. 5.30.

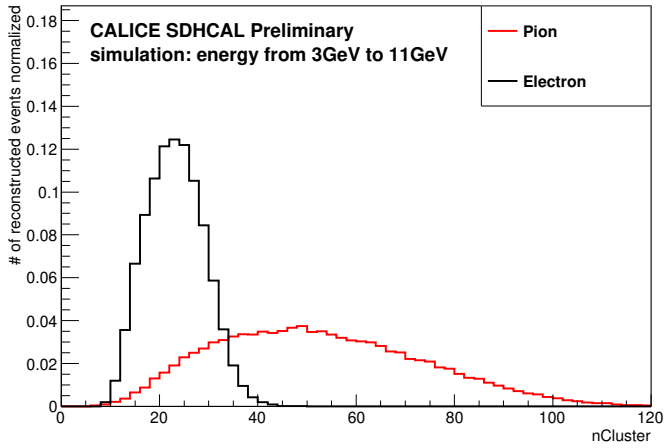


Figure 5.30: *Distribution of the number of clusters in the shower by 3 GeV to 11 GeV pions and electrons as given by the simulation. The red line is from pions and black one is from electrons.*

- **The average number of hits per fired layers (nHit/nLayer)**: This is the ratio between the total number of fired pads over the number of layers with at least one fired pad

in the prototype shown in Fig. 5.31.

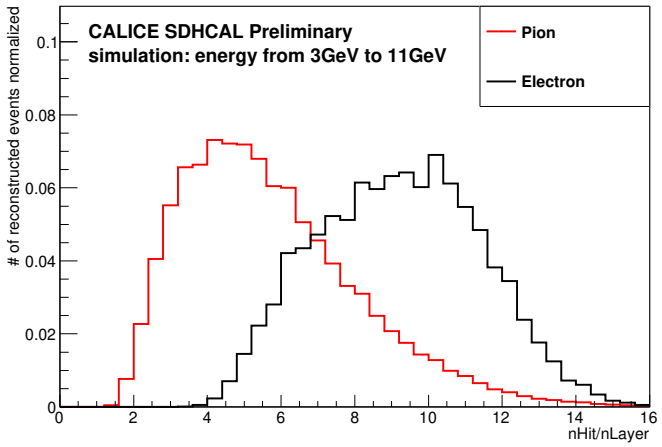


Figure 5.31: *Distribution of the average number of hits per fired layers of shower by 3 GeV to 11 GeV pions and electrons as given by the simulation. The red line is from pions and black one is from electrons.*

- **Ratio of the number of third threshold hits over the total number of hits (nHit3/nHit):** The nHit3 is the number of third threshold hits. The electron events have larger fraction of hits belonging to the third-threshold hit in the total number of hits than those of pion events. The distribution is shown in Fig. 5.32.

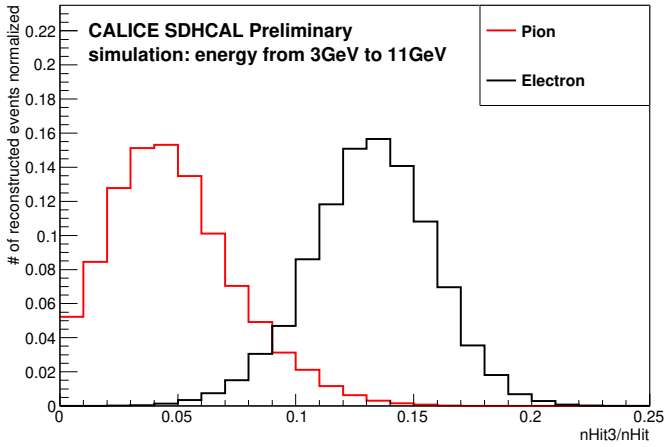


Figure 5.32: *Distribution of the ratio of number of third threshold hits over the total number of hits in the shower by 3 GeV to 11 GeV pions and electrons as given by the simulation. The red line is from pions and black one is from electrons.*

For the training and testing process, the 120000 pion events and 120000 electron events are used as training set (66.7%) and testing set (33.3%). The another independent 240000

events including pions and electrons are used as validation set. After hyperparameter optimization, the chosen set of hyperparameters such as maxDepth are described in Tab. 5.12. After feeding the eight topological variables to the BDT model using the

Option	Setting
Ntrees (Number of trees in the forest)	1000
nCuts (Number of steps during node cut optimisation)	20
MaxDepth (Max depth of the decision tree allowed)	4

Table 5.12: *The selected BDT hyperparameters .*

training and testing set, the performance of our model is shown in Figure. 5.33. It shows clearly the strong separation power between pion and electron. At the same time, The BDT response of validation set has very good agreement with training set. It confirms our model is generalized very well and far away from overfitting. Next, we apply our model in pion beams which are selected by using muon cut explained in section 5.2.2.1. The Figure. 5.34 shows the BDT output of 6 GeV (left plot) and 11 GeV (right plot) pion runs which are pure pions. The performance of pion beams matches the simulation pion very well in this figure, therefore it confirms our model is reliable. The Figure. 5.35 shows the result of 3 GeV and 5 GeV pion beam runs which are possible exist electron contamination. From this figure, the almost events locate in the region of pion class, so the electron contamination of pion runs from 3 GeV to 5 GeV is negligible. If requiring the BDT response larger than 0.0, we can obtain very pure pion events.

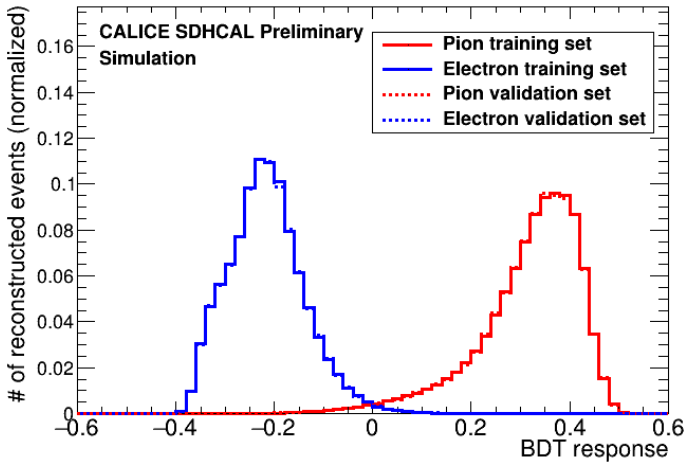


Figure 5.33: *Distribution of the BDT output of training and validation set using the simulated electron (blue) and pion (red) events from 1 GeV to 80 GeV. The solid line is from training set while the dashed one is from validation set.*

The result of the selection including the muon cut (meanRadius > 15 (mm)) and electron cut (BDT response > 0.0) is shown in Figure. 5.36 for beam data runs of 3 energies: 3, 7 and 11 GeV where the total number of hits is shown before and after selection.

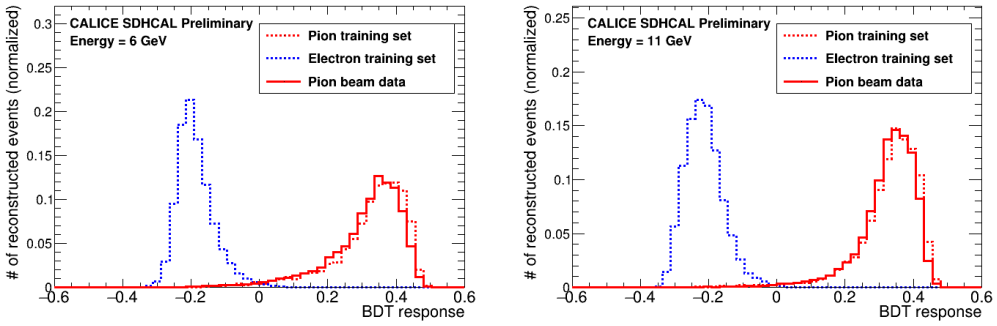


Figure 5.34: *The BDT ouput of 6 GeV (left) and 11 GeV (right) pion beam runs. The solid line is from pion beams and dashed one is from training set.*

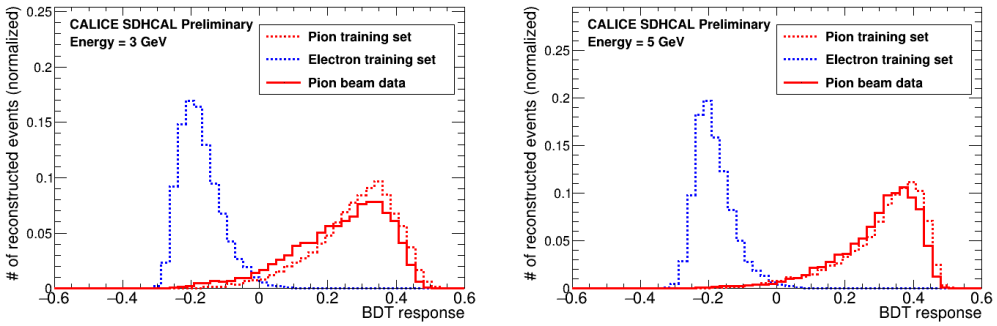


Figure 5.35: *The BDT ouput of 3 GeV (left) and 5 GeV (right) pion beam runs. The solid line is from pion beams and dashed one is from training set.*

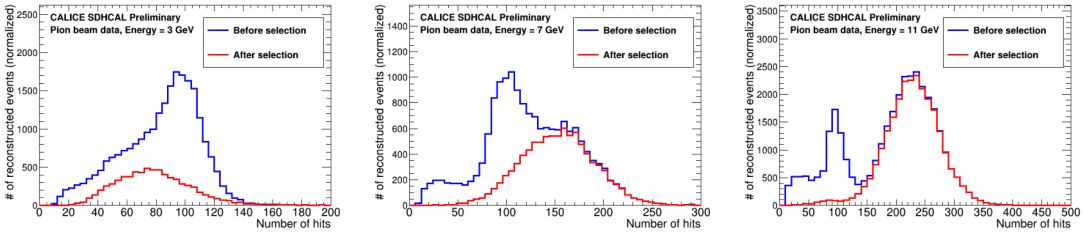


Figure 5.36: *The number of hits for 3, 7 and 11 GeV pion beam runs before (blue) and after (red) selection.*

5.2.3 Energy reconstruction

The rejection of muons and electrons present in the pion data sample using the BDT allows us to have pure pion sample as explained in the previous section. The selected pion events of beam energy from 3 GeV to 11 GeV then can be used to reconstruct energy, combined with the high energy data samples (from 20 GeV to 80 GeV) taken at SPS in 2015. The hadronic shower energy can be reconstructed as described in Section. 4.2.7.

The reconstructed energy result of 3, 7, 11 GeV pion data samples is shown in Fig. 5.37.

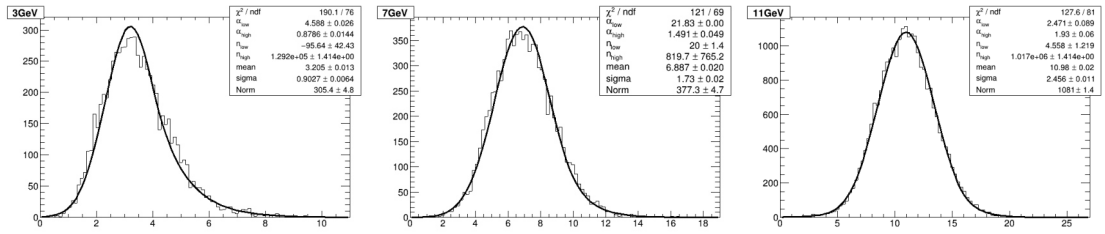


Figure 5.37: *Reconstructed energy distributions of 3 (left), 7 (middle) and 11 GeV (right) pion data samples. The distributions are fitted with a double sided Crystal Ball function.*

5.2.4 Energy resolution and linearity

After the pion events selection finished, the combined beam data including the data collected at SPS (20-80 GeV) and at PS (3-11 GeV) are used to reconstruct their energies by the standard method described in the Section 4.2.7.

After fitting the reconstructed energy distribution from 3 to 80 GeV, the mean value and standard deviation of the double sided Crystal Ball function are treated as the reconstructed energy and its resolution. Figure. 5.38 shows the corresponding energy linearity and resolution results of these combined beam data.

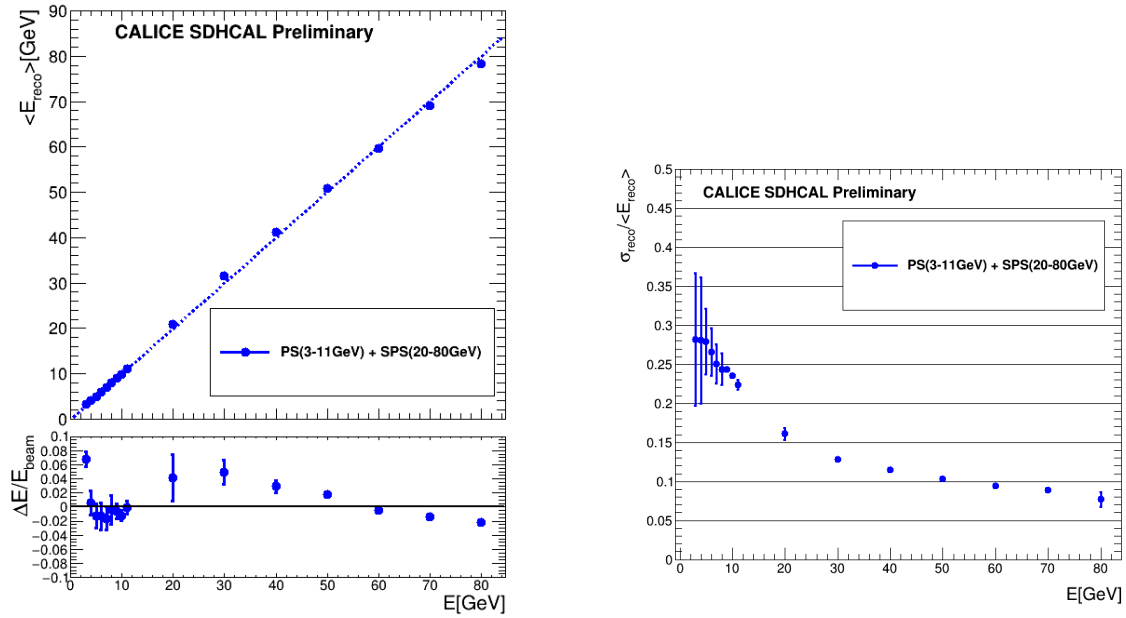


Figure 5.38: *Mean reconstructed energy of pion showers as a function of the beam energy as well as relative deviation of the pion mean reconstructed energy with respect to the beam energy (left) and resolution of the reconstructed hadron energy as a function of the beam energy (right). Both statistical and systematic uncertainties are included in the error bars.*

5.2.5 Uncertainties estimation

The linearity and energy resolution results presented in the Section. 5.2.4 include statistical and systematic uncertainties. We present here after the main contributions to the systematic uncertainties:

- For the reconstructed energy of all energy points, a double sided Crystal Ball fit function and a Gaussian fit function are used. The difference of fitting results obtained from these two fit functions are considered as the one source of systematic uncertainties.
- For the muon rejection, using all energy points data simples of PS, the meanRadius varied by an arbitrary 5% in both directions with respect to the nominal values. The maximum deviation with respect to the nominal value is used as the second source of systematic uncertainties.
- For the electron rejection using BDT method, the BDT cut value is changed from -0.05 to 0.05 with respect to the nominal values 0.0. The maximum deviation are taken as the third source of systematic uncertainties.

Although the statistical uncertainties are found to be negligible for almost all the runs with respect to systematic uncertainties, their contributions as well as the systematic uncertainties previously discussed are added quadratically to obtain the final uncertainties. The results are summarized in Tab. 5.13 and in Tab. 5.14. The uncertainty coming from the different fit function applied is the main component of total systematic uncertainties.

Energy(GeV)	Linearity
3	0.068 ± 0.011
4	0.006 ± 0.017
5	-0.013 ± 0.017
6	-0.013 ± 0.019
7	-0.016 ± 0.016
8	-0.004 ± 0.021
9	-0.006 ± 0.010
10	-0.012 ± 0.007
11	-0.001 ± 0.009
20	0.042 ± 0.033
30	0.049 ± 0.017
40	0.029 ± 0.008
50	0.018 ± 0.002
60	-0.005 ± 0.005
70	-0.014 ± 0.004
80	-0.022 ± 0.005

Table 5.13: *List of $\frac{\Delta E}{E}$ observed and associated uncertainties for beam data in the energy range from 3 to 80 GeV.*

Energy(GeV)	Resolution
3	0.282 ± 0.085
4	0.281 ± 0.081
5	0.279 ± 0.042
6	0.266 ± 0.030
7	0.251 ± 0.025
8	0.244 ± 0.020
9	0.244 ± 0.001
10	0.236 ± 0.001
11	0.224 ± 0.006
20	0.161 ± 0.008
30	0.128 ± 0.001
40	0.115 ± 0.002
50	0.103 ± 0.002
60	0.094 ± 0.003
70	0.089 ± 0.001
80	0.077 ± 0.009

Table 5.14: *List of energy resolution observed and associated uncertainties for beam data in the energy range from 3 to 80 GeV.*

5.2.6 Summary

A new particle identification method using BDT-based MVA technique is applied to purify the pion events collected at the SPS H2 beamline in 2015 by the CALICE SDHCAL prototype. The new method uses the topological shape of events associated to muons, electrons and pions in the CALICE SDHCAL to reject the two first species. A rejection rate of muons (electrons) exceeding 99.0% (99.4%) respectively with a pion selection efficiency of about 95.0% is obtained. A significant statistical gain is obtained with respect to the standard method used in the work presented in Section 4.2.5. This statistical gain is particularly significant at energies up to 40 GeV and can be explained by the fact that the showers that start in the first layers are not all rejected. This gain shows the better efficiency and separation power of the multivariate approach over the cut-based approach of the standard method. The BDT-based particle identification in CALICE SDHCAL is a robust and a reliable method as confirmed by the results of two different training approaches. Finally, a study of the linearity and the resolution of the reconstructed energy of the hadronic showers selected by the BDT-based method in the SDHCAL shows that no apparent deviation observed compared with the standard method.

In addition, the data collected in the wide energy range from 3 GeV to 80 GeV during the V2015 beam test are analyzed. The energy linearity (within 7%) and nice energy resolution (7.7% at 80 GeV) are achieved in the wide energy range. It confirms the robust performance of SDHCAL technological prototype.

CHAPTER 6

Energy reconstruction for a hadronic calorimeter using multivariate data analysis methods

The CALICE highly granular Semi-Digital Hadronic CALorimeter (SDHCAL) technological prototype provides rich information on the shape and structure of the hadronic showers. To exploit this information and to improve on the standard energy reconstruction method where only the total number of hits is used, we propose to use two methods based on MultiVariate data Analysis (MVA) techniques: the Multi-Layer Perceptron (MLP) and the Boosted Decision Trees with Gradient Boost (BDTG). The two new methods achieve better energy linearity ($\Delta E/E_{beam} \leq 2\%$) with respect to the classic method ($\Delta E/E_{beam} \leq 5\%$) and improve on the relative energy resolution. For instance, the MLP method achieves 6-7% relative improvement on the whole energy range when applied on samples of simulated π^- events in the SDHCAL.

Content

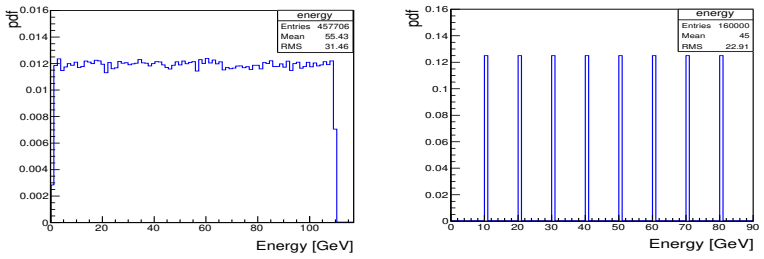
6.1 Introduction

The standard method for reconstructing energy is already discussed in the Section. 4.2.7. Although the performance of standard method to were found to be very good in the energy range between 3 and 80 GeV, the high granularity of the SDHCAL prototype was not optimally exploited. The structure and shape of the hadronic showers provide important information on its energy which was not used in the standard method for energy reconstruction.

Therefore, to improve energy reconstruction of hadrons, we firstly propose to use more information about hadronic showers in addition to the number of hits nHit1, nHit2 and nHit3. Secondly, advanced MVA methods are employed to combine the different variables and to reconstruct efficiently the hadron energy.

6.2 Monte Carlo Samples

The three sets of simulated pion events were generated by using FTF_BIC physics list. The first two sets with a total number of 457706 events were used as training (50%) and testing (50%) sets for classical method and MVA methods. The two sets have uniform distribution in the whole energy range of 1-110 GeV shown in Fig. 6.1(a). The third set of simulated events is used as validation set with discrete energy form 5 GeV to 80 GeV by a step of 5 GeV, with 20000 events for each energy point as shown in Fig. 6.1(b). The performance of the three energy reconstruction methods in terms of energy deviation from linearity and energy resolution is then given.



(a) The energy of training and testing samples.

(b) The energy of validation samples.

Figure 6.1: *MC samples used in this study. Figure (a) shows the energy distribution of training and testing sets. Figure (b) shows the energy distribution of validation set with 8 energy points. The energy range of the training and testing samples is a little bit wider than validation samples.*

6.3 The results of standard method

As a good reference, after applying the standard reconstruction method, the reconstructed energy results of discrete energy points from 10 GeV to 80 GeV with 10 GeV energy step are shown in Fig. 6.2. There are small peaks around 5 GeV for high energy beams in

Fig. 6.2. Its origin comes from the muon-like hadron events which do not generate hadronic shower and only leave a track in our prototype as in the case of a muon. The reconstructed energy distribution is fitted by a Gaussian function whose mean value and standard deviation provide the reconstructed energy (E_{reco}) and the associated resolution (σ) for each energy point. The results of these 8 energy points are shown in Tab. 6.7. The energy deviation from linearity $\Delta E/E_{beam}$, in which $\Delta E = E_{reco} - E_{beam}$, is about 3%-4% depending on the energy point. The relative energy resolution in Fig. 6.16 improves from 20% to about 9% as the energy increases from 10 GeV to 80 GeV.

6.4 Multivariate Data Analysis Methods

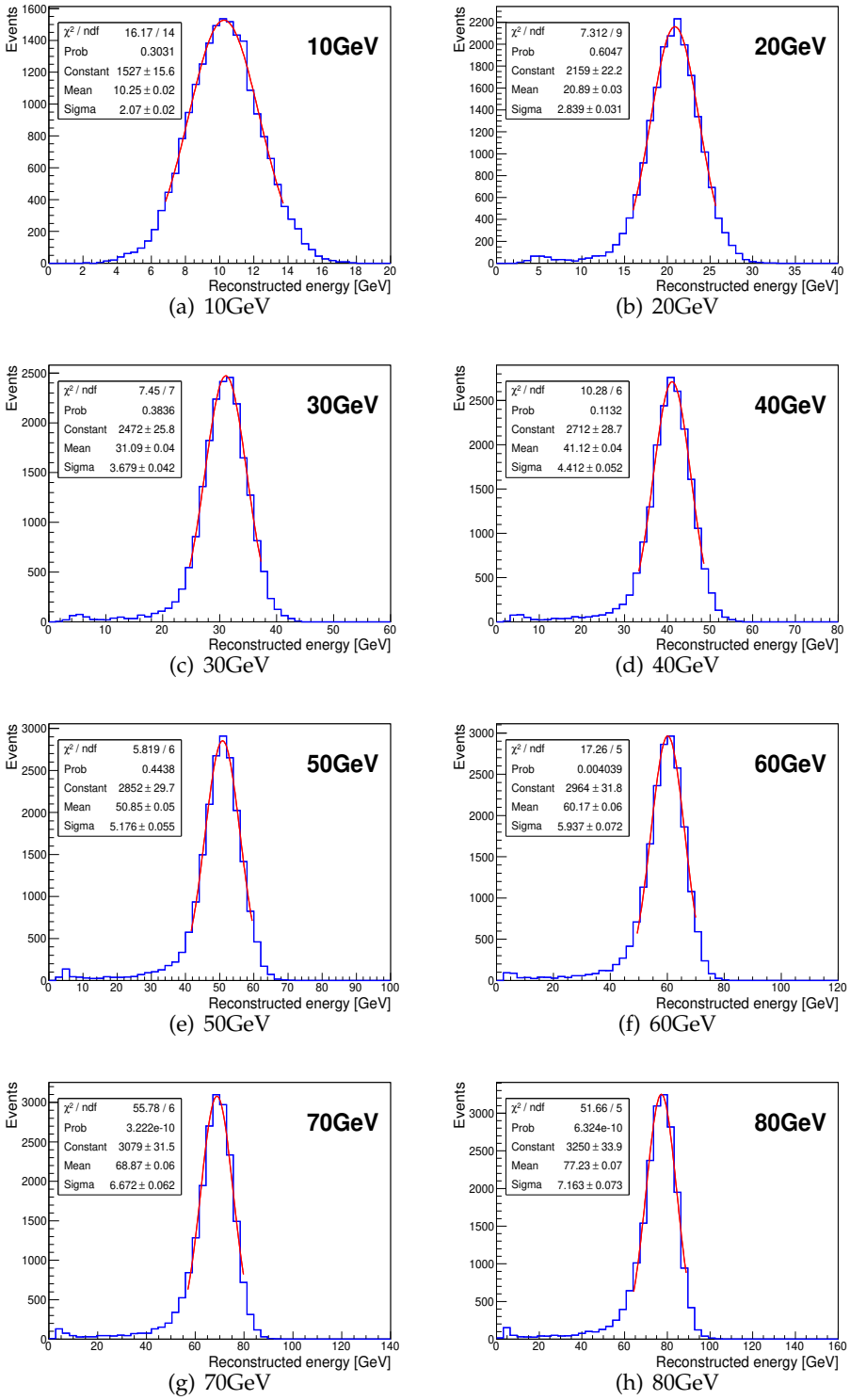
6.4.1 Input variables for MVA methods

In order to improve on the energy reconstruction performance, more information on the hadronic showers can be used. For instance, in addition to the ones used in the standard analysis: nHit1, nHit2 and nHit3, the spatial distribution of the hits and their density can help since it singles out the dense part of the shower which is in general electromagnetic-like for which the saturation effect is more important. All variables are listed in Tab. 6.1.

The presence of track segments within the hadronic shower and their number could also be helpful in estimating the energy as described in Ref. [62]. The distributions of all reconstructed variables are shown in Fig. 6.3 and Fig. 6.4. The two-dimensional distributions of all reconstructed variables versus energy are shown in Fig. 6.5 and Fig. 6.6. The correlation among input variables and beam energy is shown in Fig. 6.7 on the training samples. The variables such as nCluster, nLayer, Density, InterLayer listed in Tab. 6.1 have quite strong correlation with beam energy. These variables are used as input variables of MLP and BDTG methods.

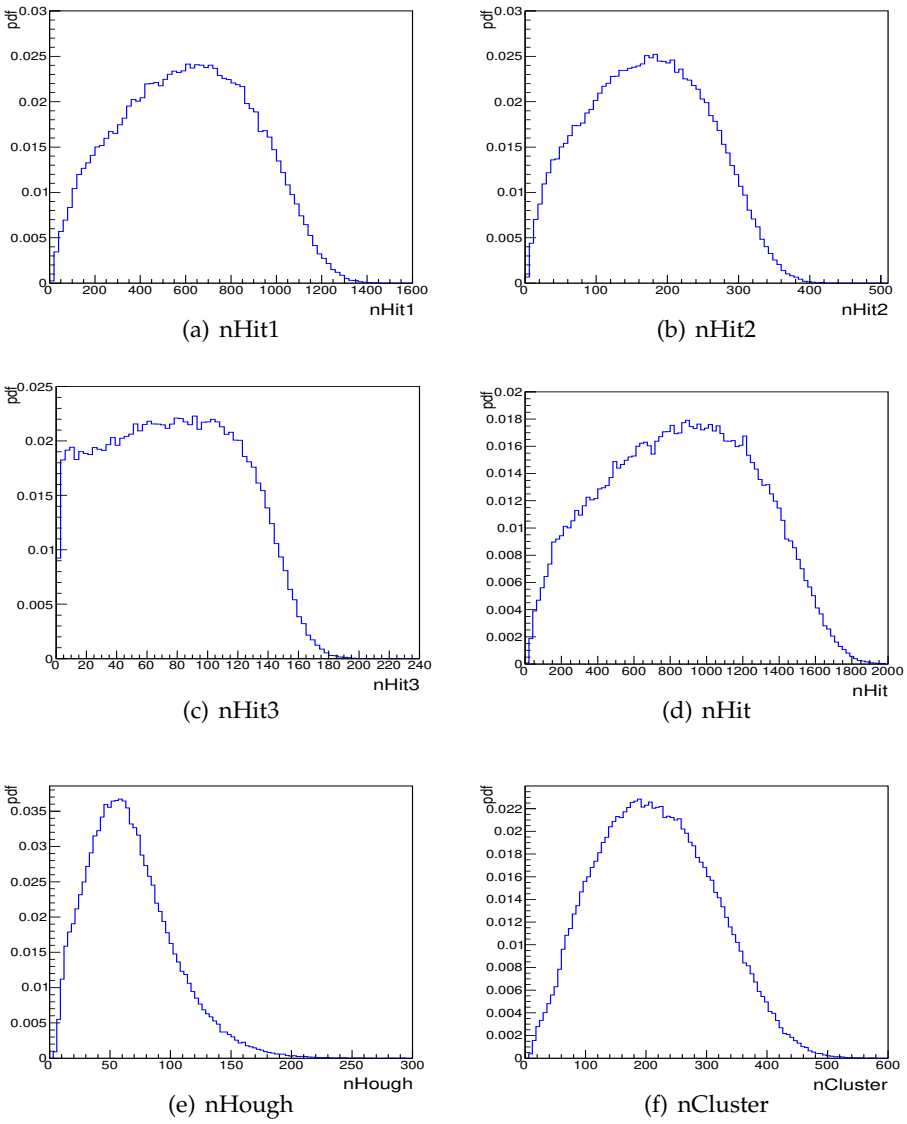
Input Variables	Description
$nHit1$	The number of hits only exceeding the threshold 1
$nHit2$	The number of hits exceeding the threshold 2 but not threshold 3
$nHit3$	The number of hits exceeding the threshold 3
$nHit$	$nHit = nHit1 + nHit2 + nHit3$
$nHough$	Number of hits used to do Hough Transformation
$nCluster$	Number of clusters
$nTrack$	Number of tracks
$nLayer$	Number of layers fired
$Density$	The density of hits
$meanRadius$	Mean of distance between tracks and hits
$InterLayer$	Number of layers when $meanRadius > 5cm$
$begin$	The number of the layer where the shower starts

Table 6.1: *Input Variables which include information of spatial distribution of hits and layers. The names defined here are the same as shown in Fig. 6.3 and Fig. 6.4.*



(a) 10GeV (b) 20GeV (c) 30GeV (d) 40GeV (e) 50GeV (f) 60GeV (g) 70GeV (h) 80GeV

Figure 6.2: *Pion energy reconstruction using classical method with energy points 10, 20, 30, 40, 50, 60, 70, 80* GeV *samples. The distributions are fitted with a Gaussian function in a* $\pm 1.65\sigma$ *(RMS>90%, it means the smallest region include the 90% events) range around the mean value. The mean value and standard deviation of the Gaussian function are treated as the reconstructed energy and its resolution.*

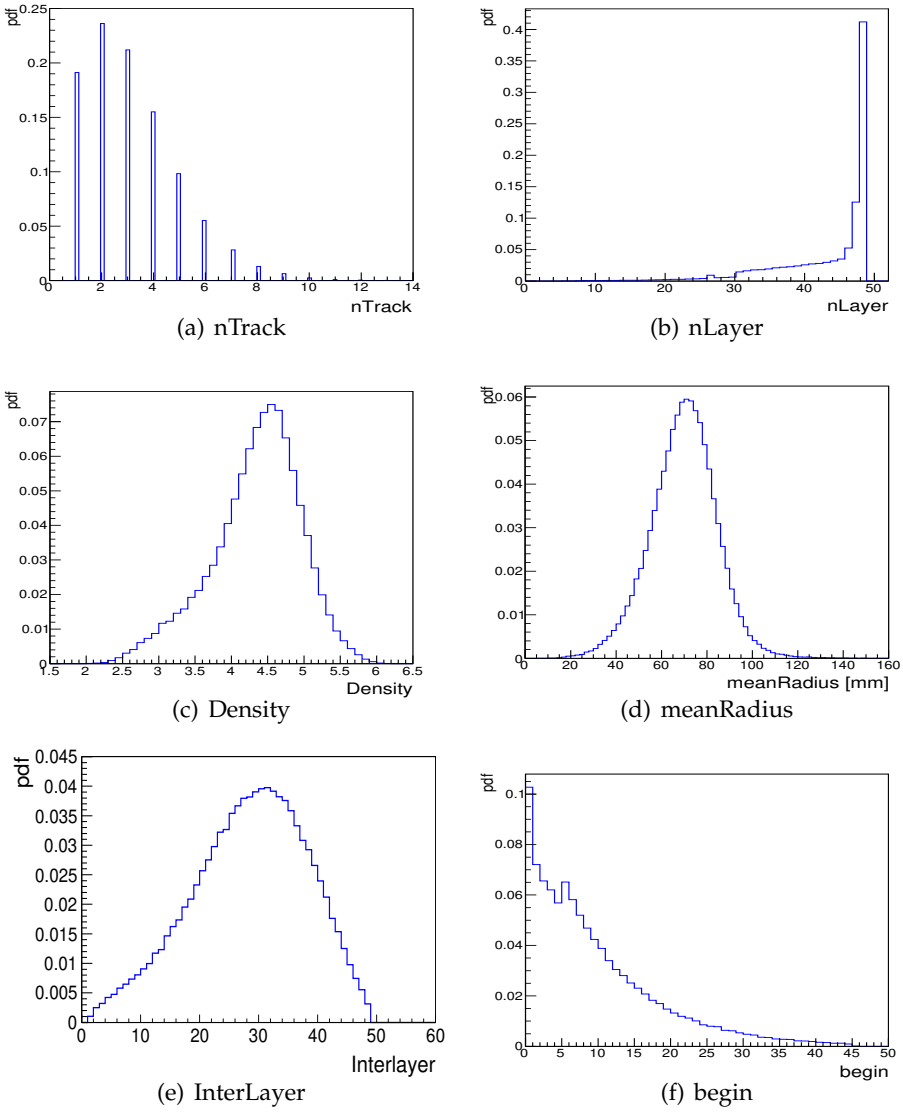


(a) nHit1 (b) nHit2

(c) nHit3 (d) nHit

(e) nHough (f) nCluster

Figure 6.3: *Probability density function/distribution of the input variables of events with uniform distributed energy from 1 GeV to 110 GeV for MVA methods.*

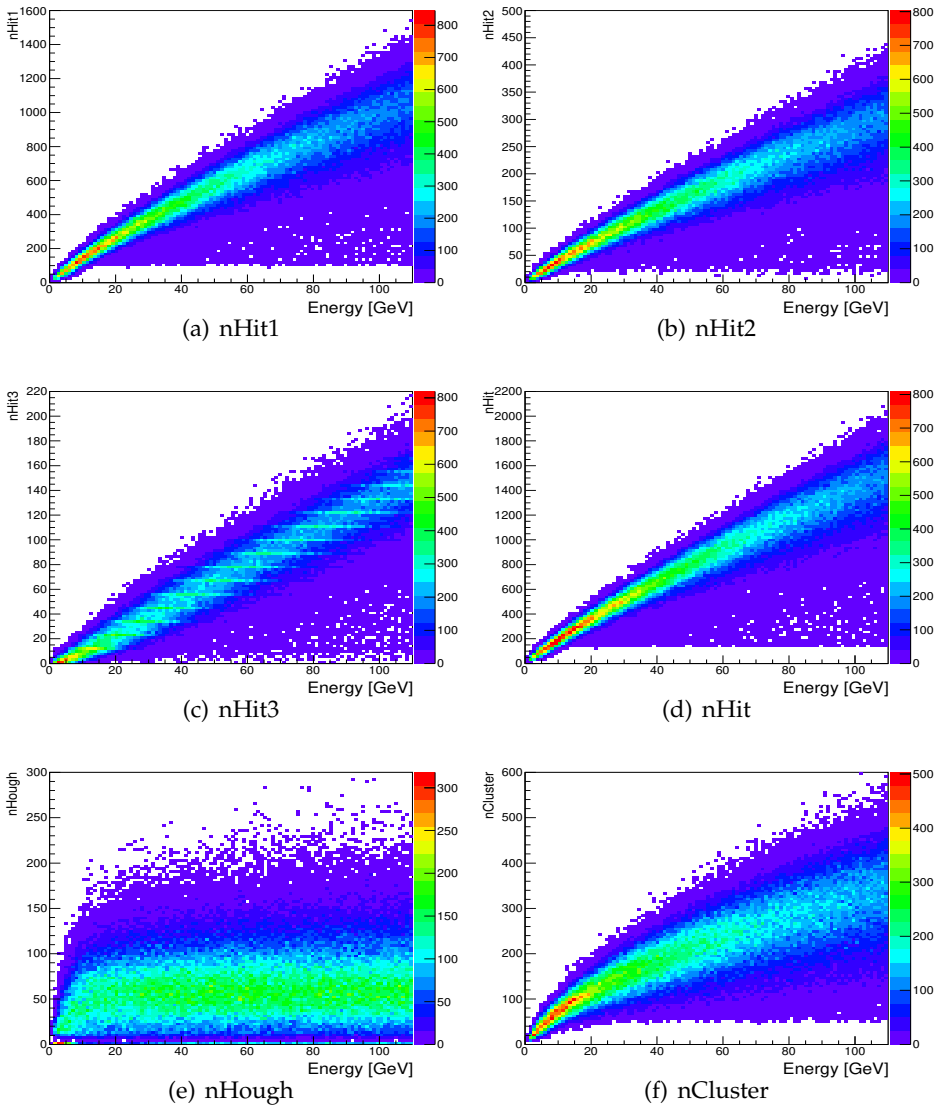


(a) nTrack (b) nLayer

(c) Density (d) meanRadius

(e) InterLayer (f) begin

Figure 6.4: *Probability density function/distribution of the input variables of events with uniform distributed energy from 1 GeV to 110 GeV for MVA methods.*

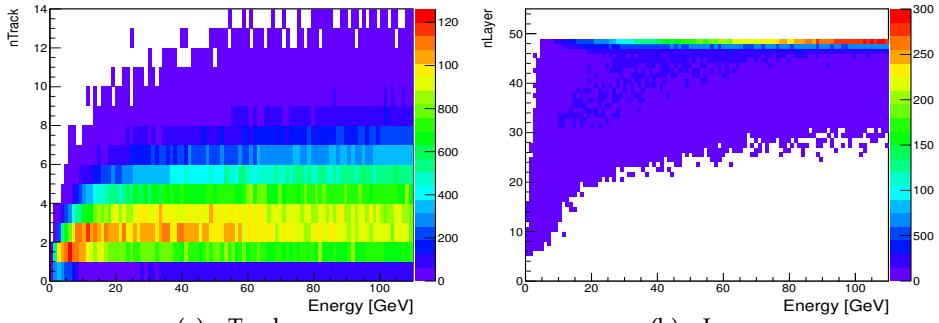


(a) nHit1 (b) nHit2

(c) nHit3 (d) nHit

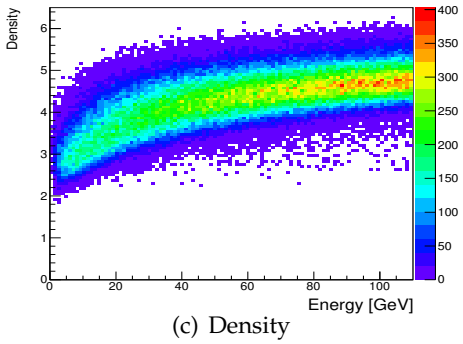
(e) nHough (f) nCluster

Figure 6.5: *The distributions of input variables versus beam energy using training and testing pion samples.*

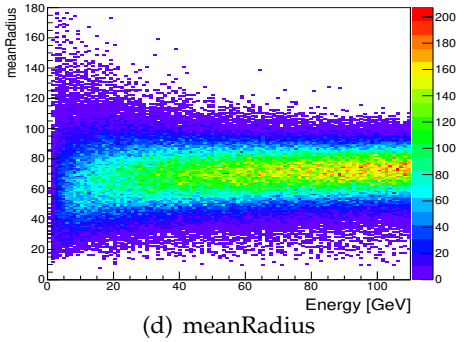


(a) nTrack

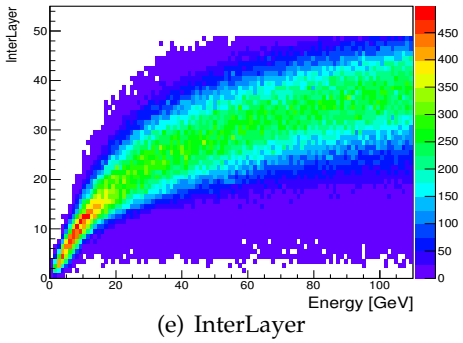
(b) nLayer



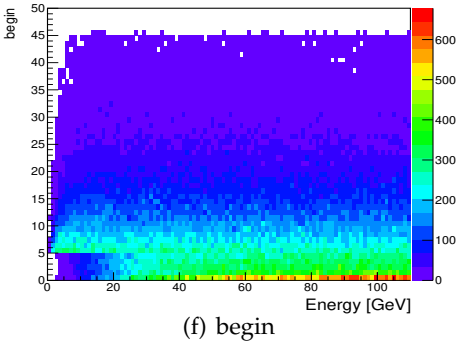
(c) Density



(d) meanRadius



(e) InterLayer



(f) begin

Figure 6.6: *The distributions of input variables versus beam energy using training and testing pion samples.*

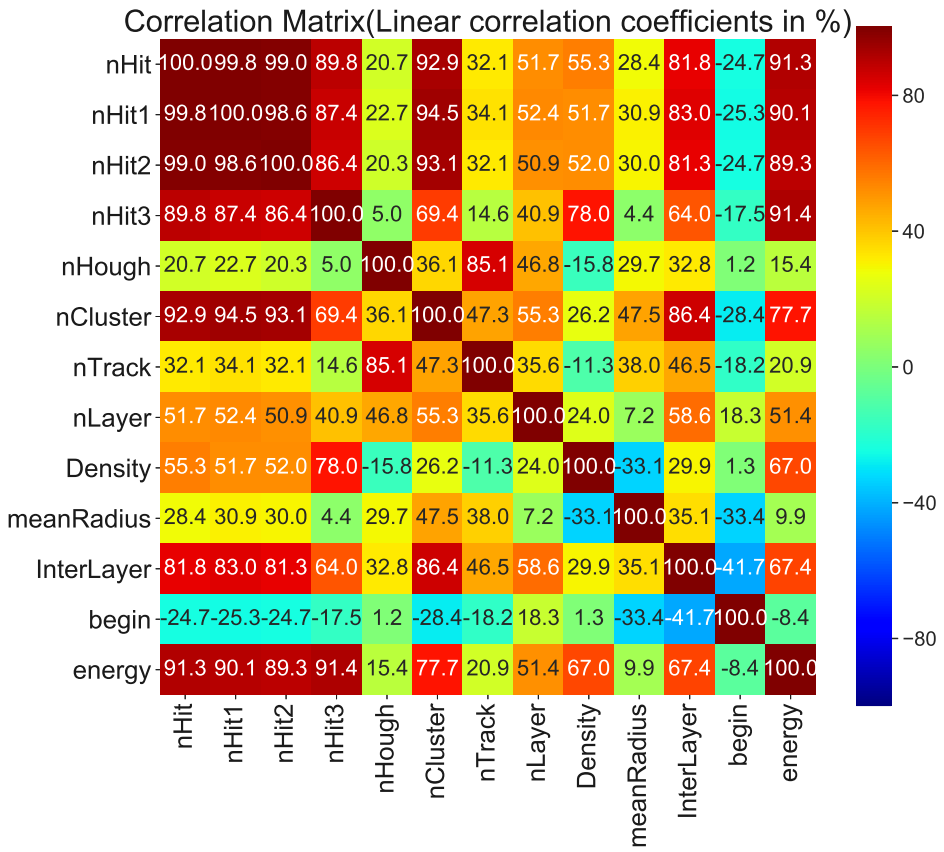


Figure 6.7: *The correlation matrix between input variables and beam energy. The colors and the numbers show the correlation between two of the input variables.*

6.4.2 Multi-Layer Perceptron (MLP)

A Multi-Layer Perceptron (MLP) is a class of feedforward artificial neural network which has been widely used in academic and industrial applications. The traditional MLP contains three layers: input layer, hidden layer and output layer. In some configurations, the hidden layer could be extended to several ones [63]. We have optimized the MLP model in this work by varying the number of layers from 3 to 6 and the number of neurons in each hidden layer from 2 to 12. We have also chosen the tanh function as the activation function after trying several other functions like the sigmoid and the Rectified Linear Unit (ReLU) function ($f(x) = x^+ = max(0, x)$). We have built this model with TMVA tool.

The training samples are fed into the network. Besides nHit1, nHit2, nHit3 and nHit variables, we have added the new variables described in Tab. 6.1 to the input layer of the neural network one by one. The order of adding the input variables is determined by the strength of their correlation with the beam energy as shown in Fig. 6.7. The performance of energy deviation from linearity and energy resolution of 9 (from 4 to 12 variables) sets of variables with 8 validation samples (8 energy points) are shown in Fig. 6.9, Tab. 6.2 and Tab. 6.3. The variables InterLayer, nLayer and nCluster help to improve energy reconstruction in low energy region. The resolution performance varies slightly as the number of input variables increases. We recommend the set of 11 variables (all except the

begin variable) to do the energy reconstruction. The finally chosen architecture is shown in Fig. 6.8. The energy reconstruction results of 11 variables are shown in Fig. 6.10 and Tab. 6.7, the energy deviation from linearity $\Delta E/E_{beam}$ is at most 1.6%, which is significantly better than that obtained with the classical method ($\leq 5\%$).

In order to avoid overtraining, training sets are used to fit the MLP model, while the testing sets are used for the assessment of the generalization goodness of the studied model. Here the two-dimensional distribution, which gives the deviation of the reconstructed energy with respect to the beam energy versus the beam energy for the chosen model is illustrated in Fig. 6.11(a) for training sets and Fig. 6.11(b) for testing sets respectively. The results of the two sets have shown similar distributions and most of the events are found in the neighborhood of $E_{reco} - E_{beam} = 0$. The model performs a good generalization and has not been obviously overtrained.

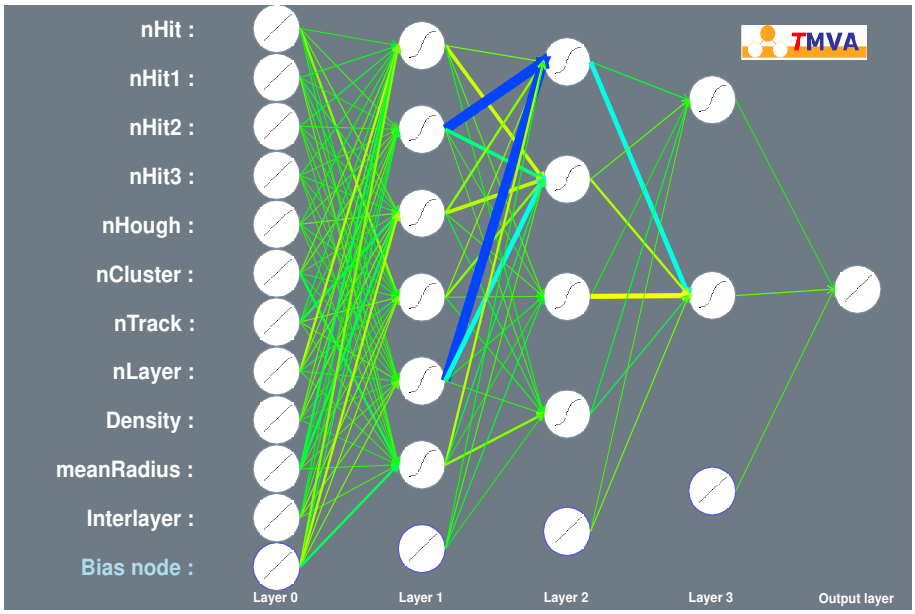
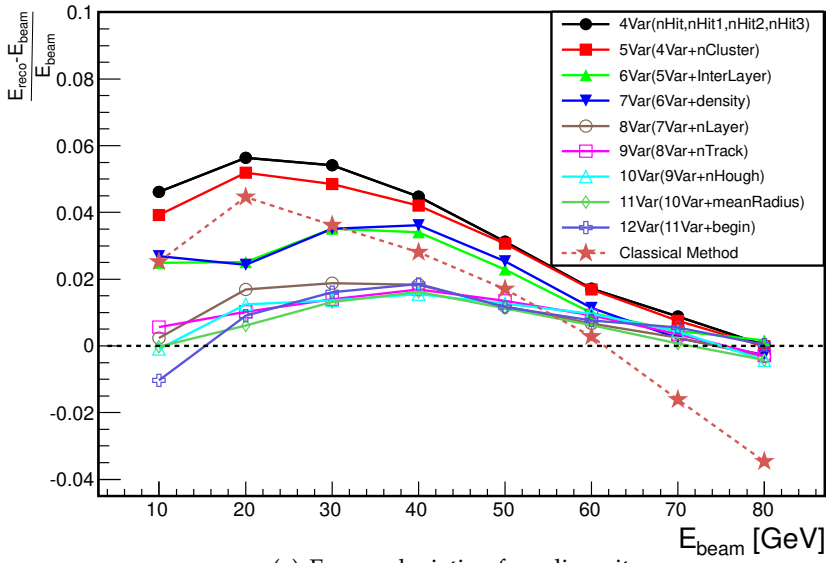
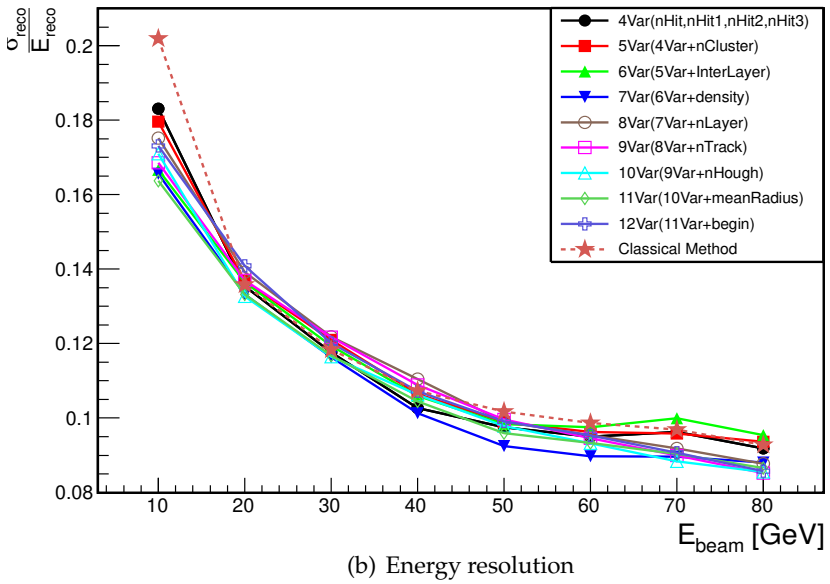


Figure 6.8: *The network architecture corresponds to the chosen set including 11 variables. The arrows represent the weighted connections of the neurons in neighboring layers. The first layer is the input layer and the final layer is the output one with one node for the reconstructed energy. The three middle layers are hidden layers.*



(a) Energy deviation from linearity



(b) Energy resolution

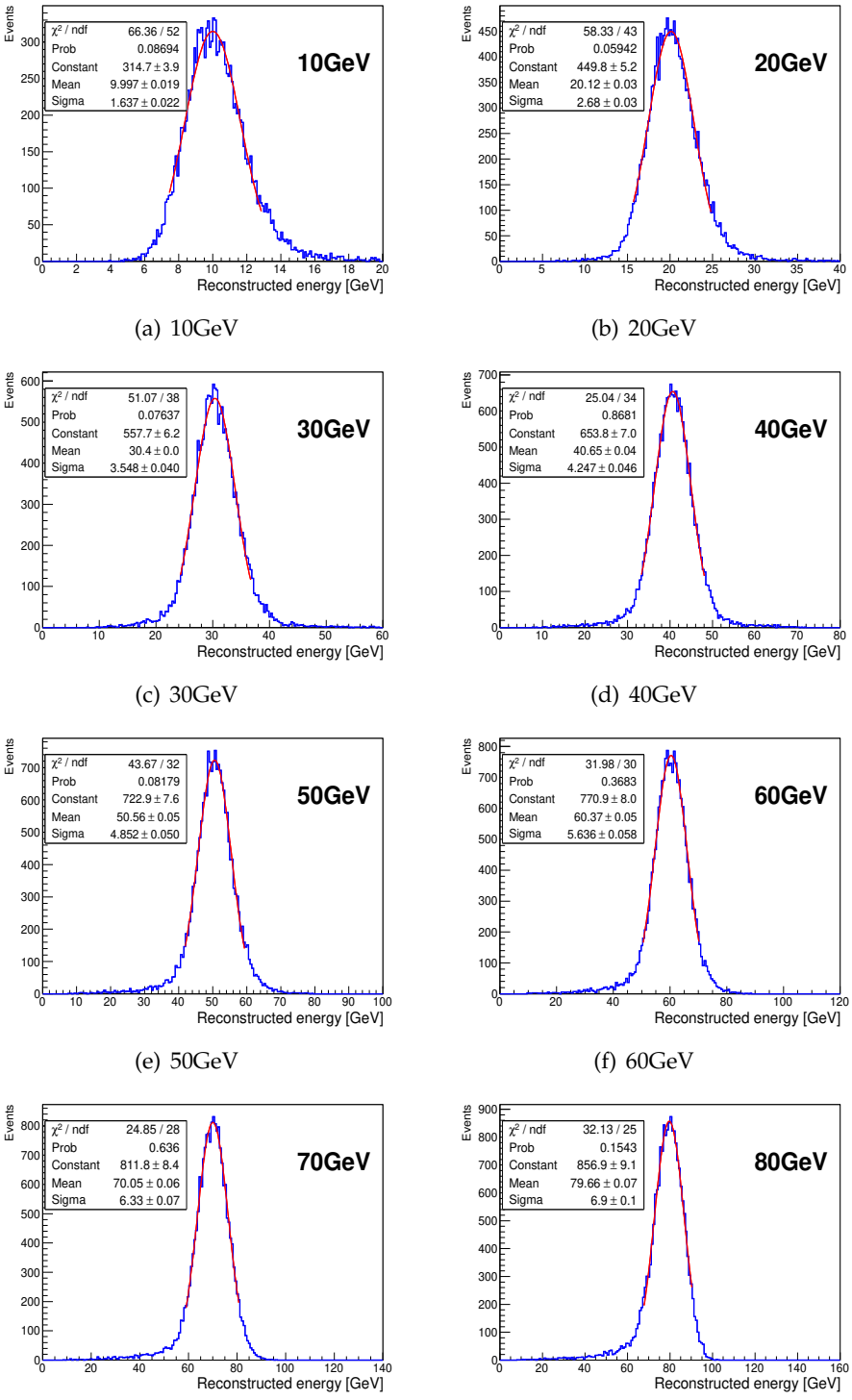
Figure 6.9: *The top plot shows the energy deviation from linearity ($\Delta E/E_{beam}$) and bottom one shows the energy resolution (σ/E_{reco}) results (from 4 to 12 variables) with MLP method. Different colors and markers denote for different sets of input variables. The variables are added one by one following the order of the strength of their correlation with the beam energy.*

$\Delta E/E_{beam}$ (%) E_{beam} / Variables	Energy of validation samples (GeV)							
	10	20	30	40	50	60	70	80
4 Var	6.4±0.3	8.7±0.1	6.4±0.1	4.3±0.1	2.9±0.1	1.8±0.1	1.0±0.1	0.1±0.1
5 Var	3.9±0.2	5.2±0.1	4.9±0.1	4.2±0.1	3.0±0.1	1.7±0.1	0.8±0.1	0.0±0.1
6 Var	2.5±0.2	2.5±0.1	3.5±0.1	3.4±0.1	2.3±0.1	1.0±0.1	0.5±0.1	0.2±0.1
7 Var	2.7±0.2	2.4±0.1	3.5±0.1	3.6±0.1	2.5±0.1	1.1±0.1	0.2±0.1	-0.3±0.1
8 Var	0.2±0.2	1.7±0.1	1.9±0.1	1.8±0.1	1.2±0.1	0.7±0.1	0.2±0.1	-0.3±0.1
9 Var	0.6±0.2	1.0±0.1	1.4±0.1	1.7±0.1	1.4±0.1	0.9±0.1	0.4±0.1	-0.3±0.1
10 Var	-0.1±0.2	1.2±0.1	1.4±0.1	1.6±0.1	1.3±0.1	1.0±0.1	0.4±0.1	-0.4±0.1
11 Var	0.0±0.2	0.6±0.1	1.3±0.1	1.6±0.1	1.1±0.1	0.6±0.1	0.1±0.1	-0.4±0.1
12 Var	-1.0±0.2	0.9±0.1	1.6±0.1	1.9±0.1	1.2±0.1	0.8±0.1	0.6±0.1	0.0±0.1

Table 6.2: *Summary table of the fitting results of $\Delta E/E_{beam}$ (%) with different input variables for MLP method. The data showed in the table correspond to the points in Fig. 6.9(a).*

σ/E_{reco} (%) E_{beam} / Variables	Energy of validation samples (GeV)							
	10	20	30	40	50	60	70	80
4 Var	20.3±0.3	13.4±0.2	11.2±0.1	10.0±0.1	9.8±0.1	9.7±0.1	9.5±0.1	9.2±0.1
5 Var	18.0±0.3	13.7±0.2	12.1±0.1	10.6±0.1	9.9±0.1	9.6±0.1	9.6±0.1	9.4±0.1
6 Var	16.7±0.2	13.7±0.2	11.9±0.1	10.6±0.1	9.8±0.1	9.8±0.1	10.0±0.1	9.5±0.1
7 Var	16.6±0.2	13.3±0.2	11.6±0.1	10.1±0.1	9.3±0.1	9.0±0.1	9.0±0.1	8.8±0.1
8 Var	17.5±0.2	13.9±0.2	12.2±0.1	11.1±0.1	9.9±0.1	9.6±0.1	9.2±0.1	8.8±0.1
9 Var	16.9±0.2	13.7±0.2	12.2±0.1	10.9±0.1	10.0±0.1	9.5±0.1	9.0±0.1	8.5±0.1
10 Var	17.1±0.2	13.3±0.2	11.7±0.1	10.6±0.1	9.8±0.1	9.3±0.1	8.8±0.1	8.6±0.1
11 Var	16.4±0.2	13.3±0.2	11.7±0.1	10.5±0.1	9.6±0.1	9.3±0.1	9.0±0.1	8.7±0.1
12 Var	17.3±0.2	14.1±0.2	12.0±0.1	10.7±0.1	9.9±0.1	9.5±0.1	9.1±0.1	8.6±0.1

Table 6.3: *Summary table of the fitting results of σ/E_{reco} (%) with different sets of input variables for MLP method. The data showed in the table correspond to the points in Fig 6.9(b).*



(a) 10GeV (b) 20GeV

(c) 30GeV (d) 40GeV

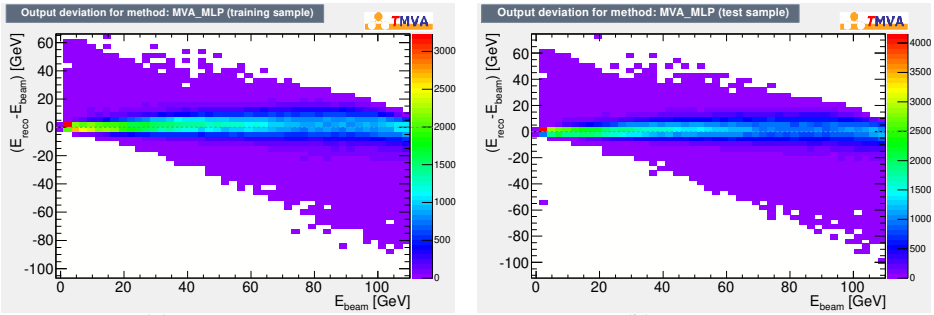
(e) 50GeV (f) 60GeV

(g) 70GeV (h) 80GeV

Figure 6.10: *Pion Energy reconstruction using MLP method with 11 variables for energy points 10*GeV*, 20*GeV*, 30*GeV*, 40*GeV*, 50*GeV*, 60*GeV*, 70*GeV*, 80*GeV *samples. The distributions are fitted with a Gaussian function in a* $\pm 1.65\sigma$ *(RMS>90%) range around the mean value. The mean value and standard deviation of the Gaussian function are treated as the reconstructed energy and its resolution.*

Option	Setting
Ntrees (Number of trees in the forest)	2000
BoostType (Boosting type for trees in the forest)	Grad
Shrinkage (Learning rate for GradBoost algorithm)	0.1
nCuts (Number of steps during node cut optimisation)	20
MaxDepth (Max depth of the decision tree allowed)	11

Table 6.4: *BDTG tuning parameters, Grad means Gradient. MaxDepth limits the scale of a single tree which could avoid over fitting.*



(a) Training sets (b) Testing sets

Figure 6.11: *Overtraining validation of the MLP method. The Y axis denotes deviation between reconstructed energy and the beam energy. The X axis is the beam energy. Figure (a) and (b) show the distribution of the training sets and testing sets respectively.*

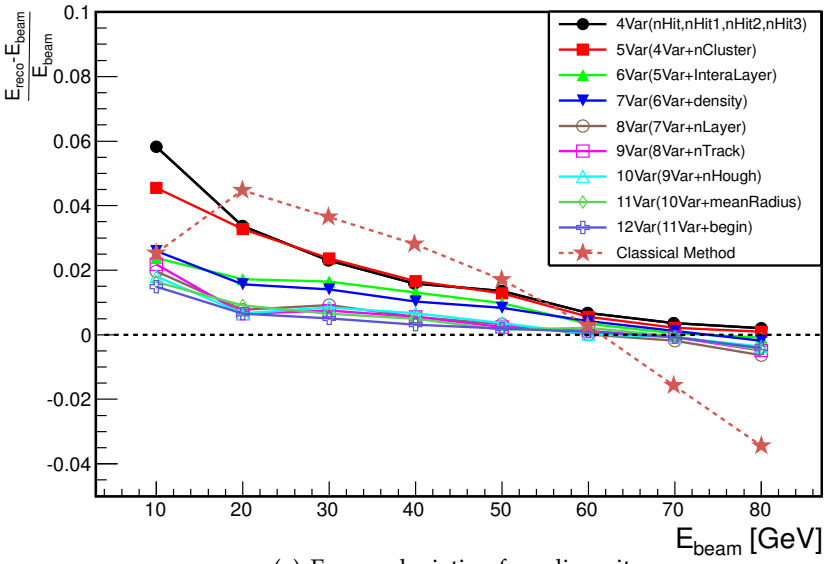
6.4.3 Boosted Decision Trees with Gradient Boost (BDTG)

With the help of BDTG method provided by TMVA tool, we build suitable model to solve this regression task. This model is optimized by varying the number of trees (Ntrees) from 1000 to 10000, and maxDepth from 5 to 13. Some tuning parameters for the 12 input variables of the BDTG tools are shown in Tab. 6.4, in which the name and the explanation of the parameters are listed.

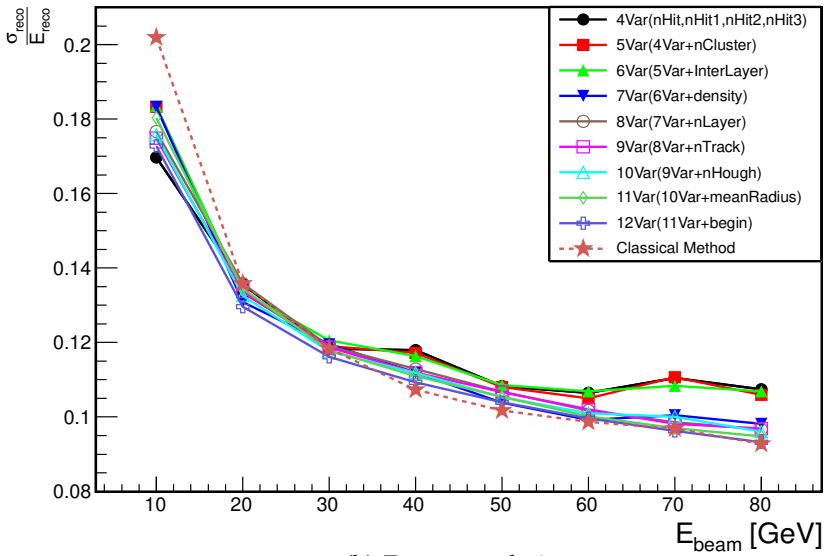
In order to determine the optimal combination of the input variables for the BDTG and to compare the two MVA methods on the same footing, we use the same samples for the training and that for the testing used in the MLP method. The results of validation set (8 energy points) with different sets of input variables are shown in Tab. 6.5, Tab. 6.6 and Fig. 6.12. Even though the variables InterLayer, nLayer and nCluster make an important contribution for the linearity performance, the other input variables still have nevertheless a significant contribution. Based on this study, we choose all 12 variables to reconstruct energy. The results are shown in Fig. 6.13 and Tab. 6.7. The energy deviation from linearity $\Delta E/E_{beam}$ is about 1.5% for E_{beam} = 10 GeV, and well within 1% for all the other energy points from 20 to 80 GeV. It is significantly better than that of the classical method and has marginal improvement with respect to that of the MLP method.

The overtraining validation is performed following the same procedure as in Section 6.4.2, which has been shown in Fig. 6.14(a) for the training set and in Fig. 6.14(b) for the testing set respectively. The BDTG model also has a good generalization performance

and has not been obviously overtrained and has similar results with the MLP method.



(a) Energy deviation from linearity



(b) Energy resolution

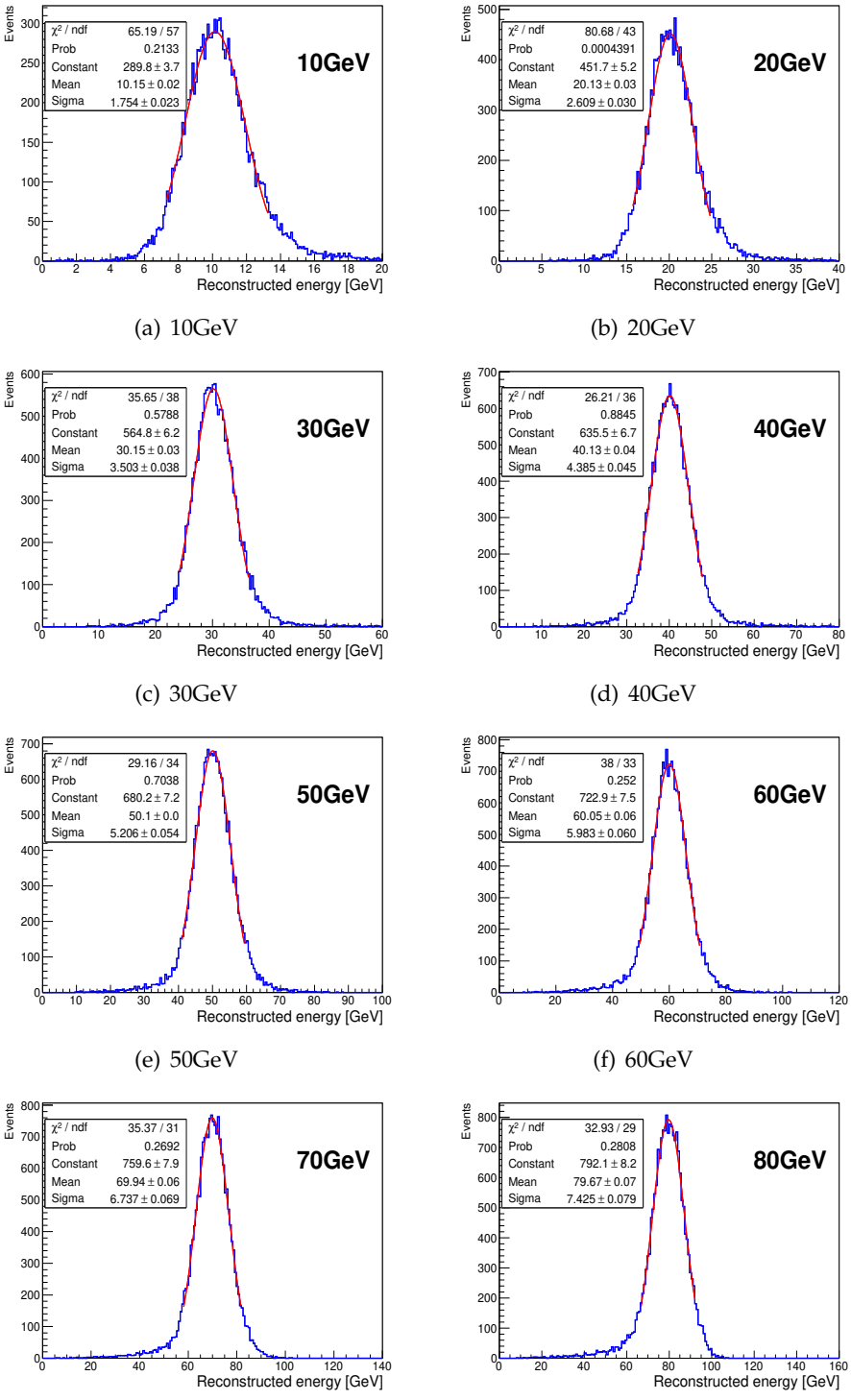
Figure 6.12: *The energy deviation from linearity ($\Delta E/E_{beam}$) and resolution (σ/E_{reco}) of the BDTG with different sets of input variables (from 4 to 12 variables). The colors and markers of the two line charts are the same as the MLP's. The results of classical method are also included in the plot for comparison.*

$\Delta E/E_{beam}$ (%) E_{beam} / Variables	Energy of validation samples (GeV)							
	10	20	30	40	50	60	70	80
4 Var	5.8 ± 0.2	3.4 ± 0.1	2.3 ± 0.1	1.6 ± 0.1	1.3 ± 0.1	0.7 ± 0.1	0.3 ± 0.1	0.2 ± 0.1
5 Var	4.6 ± 0.2	3.3 ± 0.1	2.4 ± 0.1	1.7 ± 0.1	1.3 ± 0.1	0.6 ± 0.1	0.2 ± 0.1	0.1 ± 0.1
6 Var	2.4 ± 0.2	1.7 ± 0.1	1.7 ± 0.1	1.3 ± 0.1	1.0 ± 0.1	0.3 ± 0.1	0.0 ± 0.1	-0.1 ± 0.1
7 Var	2.6 ± 0.2	1.6 ± 0.1	1.4 ± 0.1	1.0 ± 0.1	0.8 ± 0.1	0.4 ± 0.1	0.1 ± 0.1	-0.2 ± 0.1
8 Var	2.0 ± 0.2	0.8 ± 0.1	0.9 ± 0.1	0.6 ± 0.1	0.3 ± 0.1	0.0 ± 0.1	-0.2 ± 0.1	-0.6 ± 0.1
9 Var	2.2 ± 0.2	0.7 ± 0.1	0.8 ± 0.1	0.6 ± 0.1	0.2 ± 0.1	0.1 ± 0.1	-0.1 ± 0.1	-0.5 ± 0.1
10 Var	1.8 ± 0.2	0.7 ± 0.1	0.9 ± 0.1	0.7 ± 0.1	0.4 ± 0.1	0.0 ± 0.1	-0.1 ± 0.1	-0.4 ± 0.1
11 Var	1.6 ± 0.2	0.9 ± 0.1	0.7 ± 0.1	0.5 ± 0.1	0.2 ± 0.1	0.2 ± 0.1	-0.1 ± 0.1	-0.5 ± 0.1
12 Var	1.5 ± 0.2	0.7 ± 0.1	0.5 ± 0.1	0.3 ± 0.1	0.2 ± 0.1	0.1 ± 0.1	-0.1 ± 0.1	-0.4 ± 0.1

Table 6.5: *Summary table of the fitting results of $\Delta E/E_{beam}$ (%) with different sets of input variables for BDTG method. The data showed in table correspond to the points in Fig 6.12(a).*

σ/E_{reco} (%) E_{beam} / Variables	Energy of validation samples (GeV)							
	10	20	30	40	50	60	70	80
4 Var	17.0 ± 0.2	13.6 ± 0.2	11.8 ± 0.1	11.8 ± 0.1	10.8 ± 0.1	10.6 ± 0.1	11.0 ± 0.1	10.7 ± 0.1
5 Var	18.3 ± 0.3	13.3 ± 0.2	11.9 ± 0.1	11.7 ± 0.1	10.8 ± 0.1	10.5 ± 0.1	11.0 ± 0.1	10.6 ± 0.1
6 Var	18.4 ± 0.3	13.4 ± 0.2	12.1 ± 0.1	11.6 ± 0.1	10.9 ± 0.1	10.7 ± 0.1	10.8 ± 0.1	10.7 ± 0.1
7 Var	18.3 ± 0.3	13.1 ± 0.2	12.0 ± 0.1	11.2 ± 0.1	10.4 ± 0.1	9.9 ± 0.1	10.0 ± 0.1	9.8 ± 0.1
8 Var	17.7 ± 0.2	13.6 ± 0.2	11.9 ± 0.1	11.3 ± 0.1	10.7 ± 0.1	10.2 ± 0.1	9.9 ± 0.1	9.7 ± 0.1
9 Var	17.5 ± 0.2	13.4 ± 0.2	11.9 ± 0.1	11.2 ± 0.1	10.7 ± 0.1	10.2 ± 0.1	9.8 ± 0.1	9.7 ± 0.1
10 Var	17.6 ± 0.2	13.3 ± 0.2	11.8 ± 0.1	11.2 ± 0.1	10.5 ± 0.1	10.1 ± 0.1	10.0 ± 0.1	9.6 ± 0.1
11 Var	18.0 ± 0.2	13.5 ± 0.2	11.8 ± 0.1	11.1 ± 0.1	10.5 ± 0.1	10.1 ± 0.1	9.7 ± 0.1	9.5 ± 0.1
12 Var	17.3 ± 0.2	13.0 ± 0.2	11.6 ± 0.1	10.9 ± 0.1	10.4 ± 0.1	10.0 ± 0.1	9.6 ± 0.1	9.3 ± 0.1

Table 6.6: *Summary table of the fitting results of σ/E_{reco} (%) with different sets of input variables for BDTG method. The data showed in the table correspond to the points in Fig 6.12(b).*



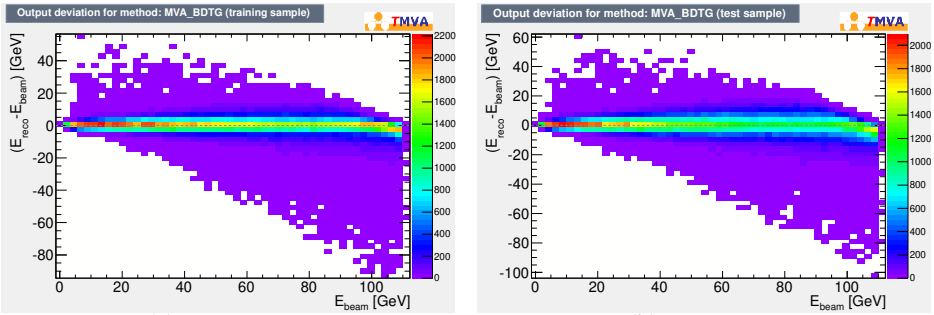
(a) 10GeV (b) 20GeV

(c) 30GeV (d) 40GeV

(e) 50GeV (f) 60GeV

(g) 70GeV (h) 80GeV

Figure 6.13: *12 input variables pion energy reconstruction using BDTG method with energy points 10*GeV*, 20*GeV*, 30*GeV*, 40*GeV*, 50*GeV*, 60*GeV*, 70*GeV*, 80*GeV *samples. The distributions are fitted with a Gaussian function in a* $\pm 1.65\sigma$ *(RMS>90%) range around the mean value. The mean value and standard deviation of the Gaussian function are treated as the reconstructed energy and its resolution.*



(a) Training set (b) Testing set

Figure 6.14: *Overtraining validation for BDTG method. The results are similar to the MLP method.*

6.5 Results

6.5.1 Comparison of the three methods

We have presented three energy reconstruction approaches based on π^- samples simulated in the frame of the SDHCAL prototype. The energy deviation from linearity and resolution of the three methods are shown in Tab. 6.7, Fig. 6.15 and Fig. 6.16. Both BDTG and MLP methods show improvements in linearity performance in the energy range between 10 and 80 GeV in comparison with the classical method. Using BDTG and MLP methods, the energy deviation from linearity is reduced from $\Delta E/E_{beam} \leq 5\%$ to $\Delta E/E_{beam} \leq 2\%$ for the broad energy range. For the energy resolution, these two methods also provide a good relative improvement. The BDTG method improves in the energy range between 10 GeV and 30 GeV with respect to the classic method while the MLP method achieves about 6-7% relative improvement on the whole energy range. It is worth mentioning here that in the two MVA-based methods the observed peak around 5 GeV due to the muon-like hadrons is absent. This can be explained by the fact that the muon-like hadrons are present in the hadron samples of all energies and are used in the training process. The MVA methods recognize probably these events but fail to assign to them the right energy. This was checked by analyzing the muon-like hadrons of a given energy through the MVA reconstruction. The reconstructed energy was found to be almost flat independently of the chosen energy. Although the muon-like hadrons represent a very small fraction of the hadron sample and do not affect the results of the MVA methods described above, it should be noticed that it is possible, as for muons, to separate them easily from those producing hadronic showers either by using standard methods exploiting their small average number of hits per layer or by using MVA techniques like the BDT-based one described in Section. 5.1. In addition, the momentum of these muon-like hadrons and therefore their energy can, as for muons, be well estimated in the presence of a magnetic field in the energy range studied in this paper.

The energy resolution could be written as:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus b \tag{6.1}$$

Results	Energy of the validation samples (GeV)							
	10	20	30	40	50	60	70	80
Classical Method								
Mean (GeV)	10.25 ± 0.02	20.89 ± 0.03	31.09 ± 0.04	41.12 ± 0.04	50.85 ± 0.05	60.17 ± 0.06	68.87 ± 0.06	77.23 ± 0.07
σ (GeV)	2.07 ± 0.02	2.84 ± 0.03	3.68 ± 0.04	4.41 ± 0.05	5.18 ± 0.05	5.94 ± 0.07	6.67 ± 0.06	7.16 ± 0.07
$\Delta E/E_{beam}$ (%)	2.5 ± 0.2	4.5 ± 0.1	3.6 ± 0.1	2.8 ± 0.1	1.7 ± 0.1	0.3 ± 0.1	-1.6 ± 0.1	-3.5 ± 0.1
σ/E_{reco} (%)	20.2 ± 0.2	13.6 ± 0.2	11.8 ± 0.1	10.7 ± 0.1	10.2 ± 0.1	9.9 ± 0.1	9.7 ± 0.1	9.3 ± 0.1
MLP Method								
Mean (GeV)	10.00 ± 0.02	20.12 ± 0.03	30.40 ± 0.04	40.65 ± 0.04	50.56 ± 0.05	60.37 ± 0.05	70.05 ± 0.06	79.66 ± 0.07
σ (GeV)	1.64 ± 0.02	2.68 ± 0.03	3.55 ± 0.04	4.25 ± 0.05	4.85 ± 0.05	5.64 ± 0.06	6.33 ± 0.07	6.90 ± 0.10
$\Delta E/E_{beam}$ (%)	0.0 ± 0.2	0.6 ± 0.1	1.3 ± 0.1	1.6 ± 0.1	1.1 ± 0.1	0.6 ± 0.1	0.1 ±0.1	-0.4 ± 0.1
σ/E_{reco} (%)	16.4 ± 0.2	13.3 ± 0.2	11.7 ± 0.1	10.5 ± 0.1	9.6 ± 0.1	9.3 ± 0.1	9.0 ± 0.1	8.7 ± 0.1
BDTG Method								
Mean (GeV)	10.15 ± 0.02	20.13 ± 0.03	30.15 ± 0.03	40.13 ± 0.04	50.10 ± 0.05	60.05 ± 0.06	69.94 ± 0.06	79.67 ± 0.07
σ (GeV)	1.75 ± 0.02	2.61 ± 0.03	3.50 ± 0.04	4.39 ± 0.05	5.21 ± 0.05	5.98 ± 0.06	6.74 ± 0.07	7.43 ± 0.08
$\Delta E/E_{beam}$ (%)	1.5 ± 0.2	0.7 ± 0.1	0.5 ± 0.1	0.3 ± 0.1	0.2 ± 0.1	0.1 ± 0.1	-0.1 ± 0.1	-0.4 ± 0.1
σ/E_{reco} (%)	17.3 ± 0.2	13.0 ± 0.2	11.6 ± 0.1	10.9 ± 0.1	10.4 ± 0.1	10.0 ± 0.1	9.6 ± 0.1	9.3 ± 0.1

Table 6.7: *Summary table of the fitting results of the validation samples with three methods. The mean value and sigma (σ) respectively stand for the mean value and standard deviation of the Gaussian function.*

Method	a (stochastic term)	b (constant term)
Classical	(55.2 ± 0.7) %	(6.7 ± 0.1) %
MLP	(49.1 ± 0.7) %	(6.9 ± 0.1) %
BDTG	(48.0 ± 0.7) %	(7.7 ± 0.1) %

Table 6.8: *The energy resolution formed by $\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus b$. The two terms a and b are obtained by fitting the subfigure of Fig. 6.16.*

where the $\oplus$ denotes the quadratic sum. We use this formula to fit the energy resolution of the three methods. The fitting results are shown in the bottom subfigure of Fig. 6.16 and Tab. 6.8.

6.5.2 Simulation model dependency

Different physics lists have different lateral and longitudinal developments leading to different responses in the SDHCAL. Compact hadronic shower presents more saturation behavior in the shower core and therefore we expect less linear relation between the reconstructed hadronic shower energy using the standard method and the beam energy. The MVA techniques seem to correct better for the saturation-related problem than the standard one and the difference of energy reconstruction at least among three of the physics lists is negligible when we use the same structure of MVA model as the one described in previous section. The results are shown in Fig. 6.17 and Fig.6.18.

6.6 Summary

The energy reconstruction for simulation samples generated by FTF_BIC physics list has been performed with the standard method and two MVA methods. Exploiting the rich information of hadronic shower structure provided by the highly granular SDHCAL

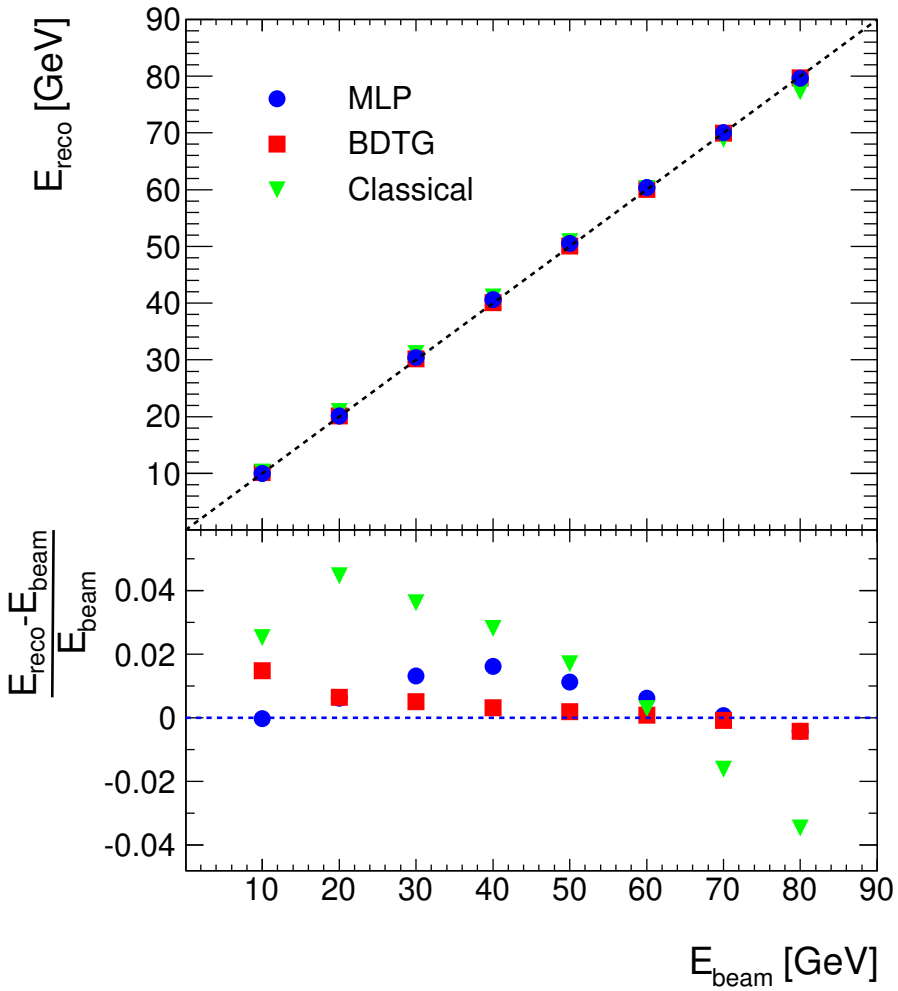


Figure 6.15: *Energy linearity comparison and relative deviation between the reconstructed energy and the beam energy of pion samples. The triangle markers represent the results of the classical method, the circles that of the MLP one and the squares the BDTG's ones.*

prototype, the MVA methods use efficiently the nonlinear relations between energy and the shower structure features providing better performance than the classical method. Both BDTG and MLP methods could be, therefore, considered as promising methods for hadron energy reconstruction. The improvement of these two methods on energy reconstruction with respect to the classic method is confirmed by applying them to different physics lists which have slightly different hadronic shower topologies.

The description of the data measured with the SDHCAL by the simulation is not yet satisfactory [49]. Therefore, it is not possible to apply those approaches to the data directly. In the future, once the simulation is able to reproduce the data in a satisfactory way, we can apply these methods to study the energy reconstruction. Otherwise, one can

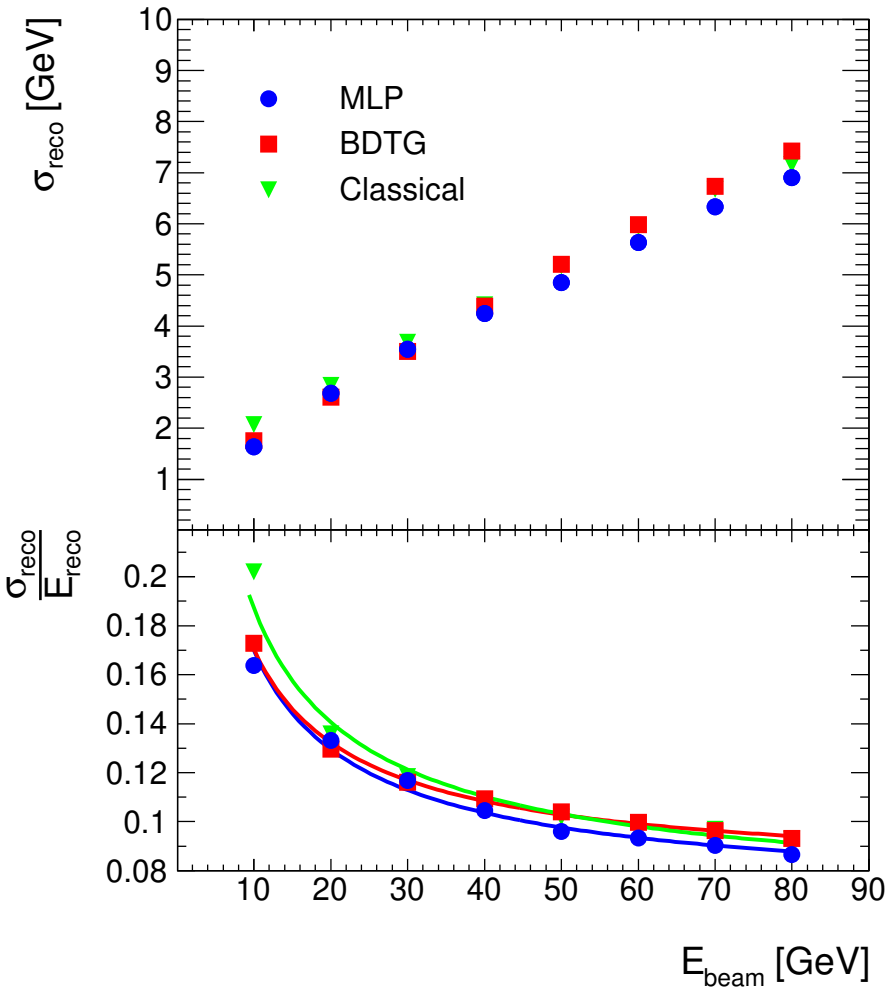
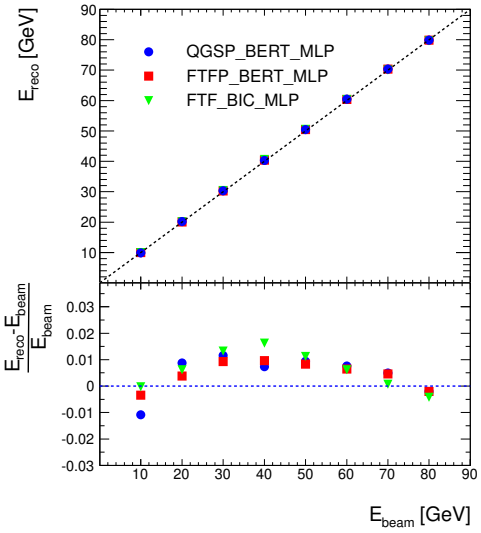
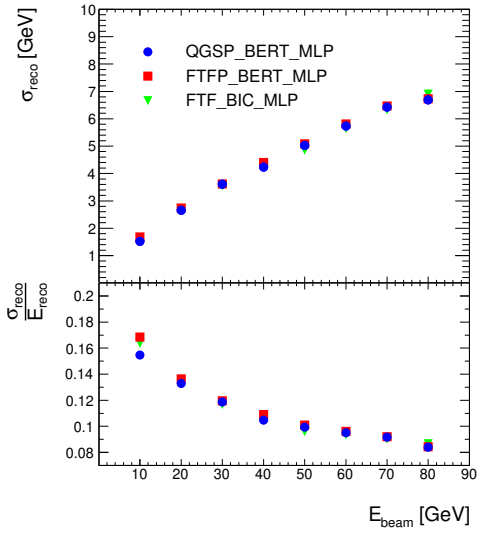


Figure 6.16: *Energy resolution comparison. The resolution is obtained by the fitted reconstructed energy of pion samples using Gaussian function. The triangle markers represent the results of the classical method, the circles that of the MLP one and the squares the BDTG's ones.*

expose the SDHCAL prototype to hadron beams with many energy points and use the data samples to train and test in a similar way to the one we used in this study.

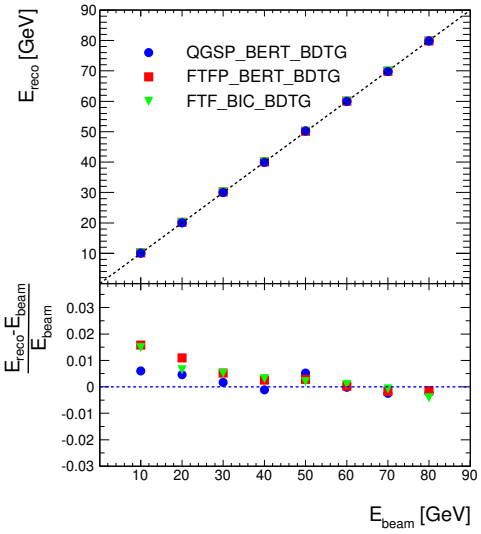


(a) Energy linearity comparison of MLP

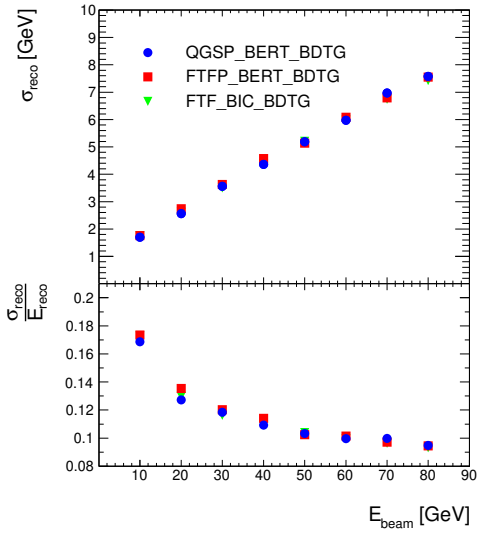


(b) Energy resolution comparison of MLP

Figure 6.17: *Energy linearity and resolution comparison between different physics lists using the MLP method.*



(a) Energy linearity comparison of BDTG



(b) Energy resolution comparison of BDTG

Figure 6.18: *Energy linearity and resolution comparison between different physics lists using the BDTG method.*

CHAPTER 7

Study of physics channel $e^+e^- \to \nu_e\bar{\nu}_e\, H(H \to W^+W^- \to q\bar{q}'q\bar{q}')$ in the CEPC with $\sqrt{s} = 240\ GeV$

The CEPC is a future leptonic collider project. One of the most important tasks for CEPC is to precisely measure the couplings of Higgs boson with other Standard Model particles. It gives the answer for the question if the discovered 125 GeV boson is the predicted Higgs of the Standard Model or not. In the CEPC operated at $\sqrt{s}$ = 240 GeV, there are three processes contributing to the Higgs production such as the ZH process, WW fusion process and ZZ fusion one. In order to measure the coupling constant of the Higgs boson with W boson referred to as g_{HWW}, the WW fusion process is employed. In this chapter we will describe how the CEPC will allow such a measurement and the expected statistical uncertainty.

Content

7.1 Introduction

The Higgs boson was discovered by both ATLAS and CMS experiments of LHC. The CEPC is one of the proposed leptonic collider projects as a Higgs factory. In order to check whether the Higgs boson is the particle predicted by the SM, the precision measurement of the couplings of Higgs boson with SM particles in CEPC will be its central task. Therefore, the simulation of physics event and reconstruction events after considering the detector response will play a key role in the study of measurement of Higgs couplings. This section will introduce the basic information about the simulation tools and its core algorithm referred to as Arbor PFA [31] for reconstructing the event. The present simulation flow of CEPC is shwon in Fig. 7.1.

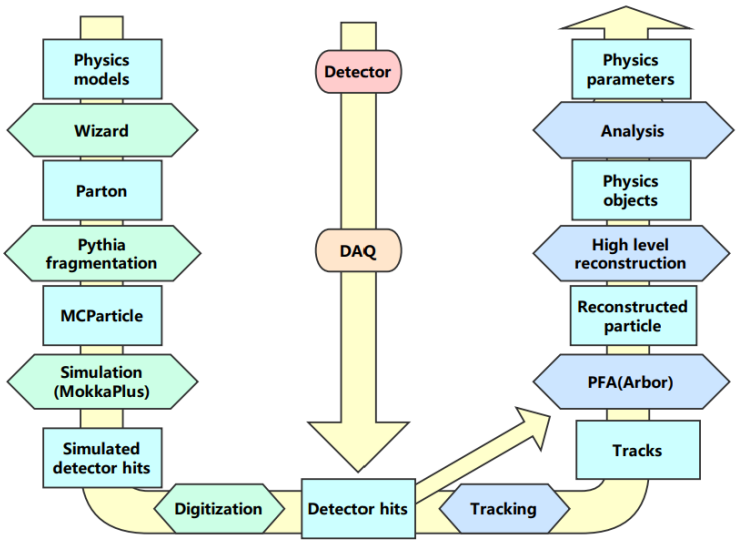


Figure 7.1: *The present simulation flow of CEPC [20]. The event generator is Whizard. The simulation of detector is based on MokkaPlus. The Arbor PFA is employed as the core algorithm to reconstruct event.*

7.1.1 Event and detector simulation

For the generation of physics events in the CEPC, the Whizard event generator [64] is adopted. This generator is also the one chosen by experiments such as ILC, CLIC and FCC. It provides efficient calculation of the cross section of many particle interactions. This includes not only the Standard Model interactions but also those of many BSM models. In the CEPC, in additon to the the Whizard event generator whic is employed for the simulation of Higgs signal and associated SM backgrounds, two other generators; Madgraph [65] and Pythia [66] are used to produce BSM events.

For the detector simulation of CEPC, the framework named MokkaPlus which is based on Geant4 package is used. This is an advanced version of the Mokka package which is used by the leptonic colliders projects in early times.

7.1.2 Event reconstruction

For the event reconstruction within the CEPC baseline detector, the Arbor PFA algorithm is employed. The basic idea of the Arbor PFA is inspired by the fact that the development of shower in space is similar to the topology of a tree. With the help of the PFA-oriented detector concept, especially the use of high granularity calorimeter system, the hits of a shower can be precisely recorded which can be subsequently used to reconstruct the tree-like structure as shown in Fig. 7.2 for a shower produced by a 20 GeV K_L^0. In the figure, on can clearly see the long tracks that are produced by the charged particles created in the shower. The Arbor PFA consists of two phase including the (a) clustering phase and

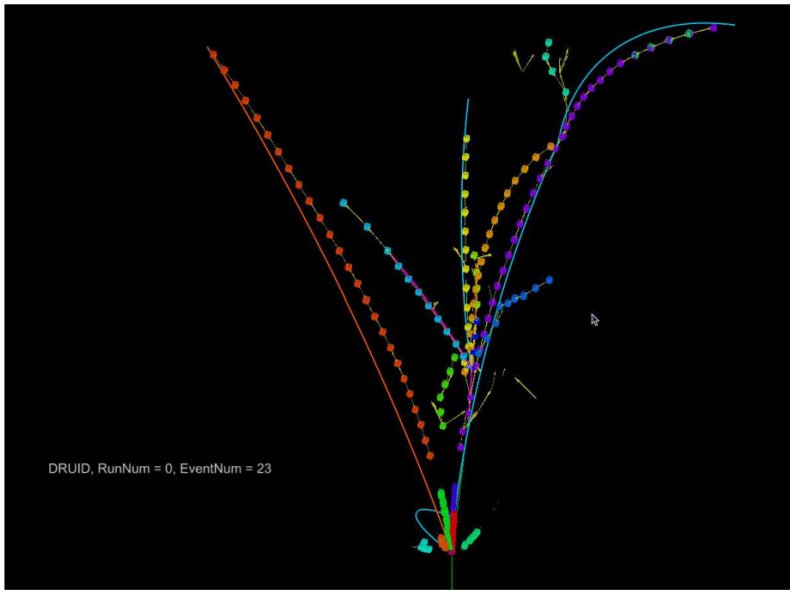


Figure 7.2: *The tree-like structure reconstructed by Arbor PFA for a 20 GeV K_L^0 shower [31].*

(b) matching phase. The clustering phase can be divided to two steps.

- **Building and cleaning connectors:** The fundamental unit of the information recorded by the PFA-oriented detector is a hit. For any pair of hits whose distance is shorter than a fixed vale (threshold), a local connector between these two hits will be built. This connector is represented by the oriented arrow linking this pair of hits. The arrow ends at the hit of the pair, which is located in the deeper position (downstream position with respect to the detector IP). For instance, in the event shown in Fig. 7.3 for which the hits are represented by points, the distance between the hit with black color and the other hits with red color is longer than the threshold value. therefore, there is no connector between the black hit and the other hits. However, for the red hit in the layer 2, its distance to the other hits is shorter than the threshold, thus five connectors linking this hit to the others are built, represented by the orange arrows. After building up the connectors, only one connector of all the connectors ending at a given hit will be kept, which is referred to as the connector cleaning. Indeed, for each hit, a reference direction is associated.

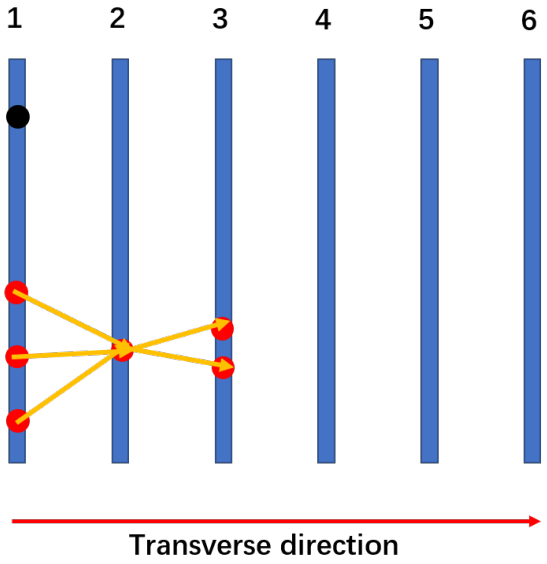


Figure 7.3: *The illustration of how to build the connector of a pair of hits in the Arbor PFA.*

This reference direction is calculated based on the information of the direction and the length of all connectors ending at this hit as explained in Ref.[31]. Only the connector which exhibits the minimal angle with the reference direction will be kept as shown in Fig. 7.4. In this figure, associated to the hit of the second layer there are three connector, the reference direction (represented by the blue dashed arrows) is calculated using these three connectors. Then among these three connectors (1, 2, 3), connector-2 has the minimal angle with the reference direction, thus it will be kept and the two other connectors are removed.

Figure 7.4: *The illustration of how to clean the connectors of a hit in the Arbor PFA. The blue dashed arrows are reference direction and the orange dotted lines represent the connectors removed.*

- **Building the branches (Clusters):** After the building and cleaning the connectors, we can obtain the connectors map which exhibits the tree-like topology. Figure. 7.5 shows an example of such a map. The yellow hits (in the 5 and 6th layers) represent the leaves where the connectors end and no connectors start. The oranges hit (in the 1st layer) is the seed where no connectors end but only start. Paths can be found by tracing back from each of the leaves to the seed. Then we can sort these paths according to their length. The longest path is defined as the main branch. Next,

we can then start again recursively to find further branches by following paths that end either at the seed or in the main branch. This process is illustrated on Fig. 7.5 where on can see two paths from leaves (yellow hits) to the seed (orange hit), a straight path and a bending one. The straight path presents longer length than the

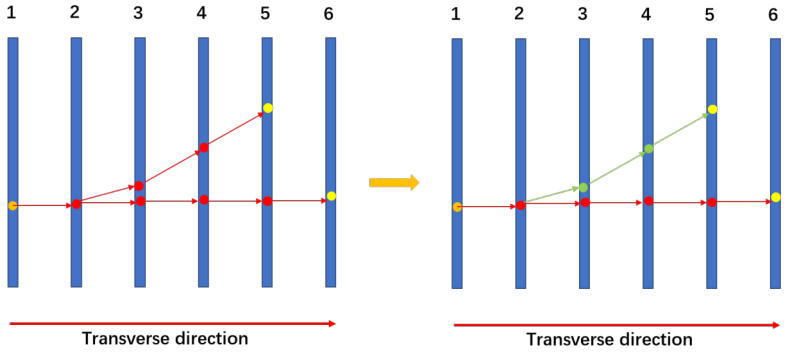


Figure 7.5: *The illustration of how to build the branches from connectors' map in the Arbor PFA. The red curve (right figure) represents the longest branch and green one is another branch. The orange and yellow hits represent the seed and leafs respectively.*

bending one, thus it is taken as the main branch (represented by the red line in the right figure of Fig. 7.5.) and then the second branch is found, represented by the green curve. In the Arbor PFA, the branches represent the tracks left by charged particles. For each of the branches, there two reference directions that are associated with its seed and its leaf respectively. These two direction can be subsequently used to link branches together to form the so-called PFA clusters[31]. Figure. 7.6 shows an example of application of the Arbor PFA algorithm to one event simulated using the CEPC baseline detector.

After the reconstruction by the Arbor PFA of the calorimeter clusters, another module of the Arbor PFA algorithm (called matching algorithm) is used to associate the reconstructed PFA clusters with tracks recorded in the tracker systems and subsequently to build the charged particles. The remaining clusters (with no matched track) are then considered as electrically neutral particles.

Based on these two phases, the Arbor PFA provides excellent interpretation allowing to use the information recorded in the sub-detectors system to build physics objects. These objects will be used in the physics analyses.

7.2 Higgs boson production

The cross section of Higgs boson production foreseen by the Standard Model as a function of the center of mass energy ($\sqrt{s}$) is shown in Fig. 7.7.

At $\sqrt{s}$ = 240 GeV, there are three processes contributing to the production of Higgs boson in the CEPC: (a) the ZH process (known as Higgsstrahlung) $e^+e^- \rightarrow ZH$; (b) the WW fusion process, $e^+e^- \rightarrow \nu\bar{\nu}_e H$; (c) the ZZ fusion process, $e^+e^- \rightarrow e^+e^- H$. The latter

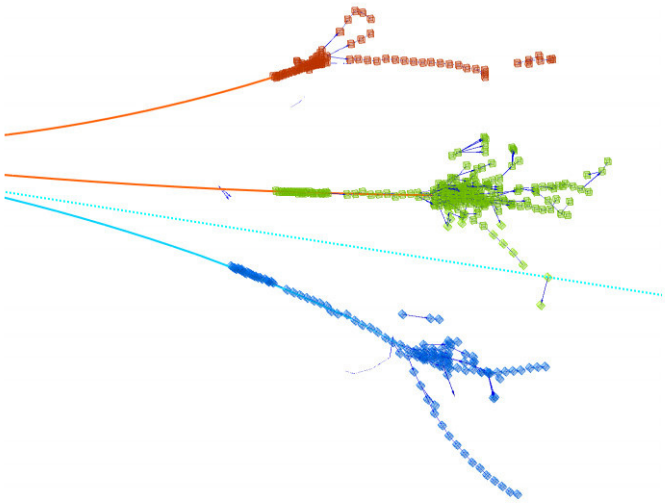


Figure 7.6: *The illustration of clustering reconstruction of Arbor PFA in the CEPC detector. The 3-prong decay of τ-lepton from Z→$\tau^+\tau^-$. Three reconstructed branches of calorimeter clusters represent a 5.7 GeV π^+ (red), a 27.4 GeV π^+ (green) and a 10.3 GeV π^- (blue). The Monte Carlo truth information of two$\pi^{+'}$s (red line), a π^-(cyan line) and a $\bar{\nu_\tau}$ (cyan dashed) [20].*

two process are collectively named as the vector boson fusion (VBF) process. The three corresponding Feynman diagrams (tree level) of these processes are shown in Fig. 7.8.

From Fig. 7.7, the cross section of ZH process increases with $\sqrt{s}$ and reaches its maximum around $\sqrt{s}$ = 250 GeV. When the $\sqrt{s}$ continues to increase, the cross section of ZH process decreases asymptotically as $1/\sqrt{s}$. For the Higgs production through VBF processes, its cross section increases following the logarithmic function os $ln^2(s/M_V^2)$, where M_V is the mass of the associated vector boson. Due to the fact that couplings between Z and the electron is small, the WW fusion process is dominant in the VBF process. For the CEPC operating as a Higgs factory at $\sqrt{s}$ = 240 GeV (Higgs mode), the summary of the cross sections for the three Higgs production processes are shown in Tab. 7.1.

Process	cross section (fb)
$e^+e^- \to ZH$	196.2
$e^+e^- \to \nu\bar{\nu}_e H$	6.19
$e^+e^- \to e^+e^- H$	0.28

Table 7.1: *The cross section of three Higgs production process at $\sqrt{s} = 240GeV$ in the CEPC.*

In addition to the difference of two bosons couplings to the electrons, The difference of the differenc of their cross sections can be understood by examining the Higgs boson decay branching ratios under the SM prediction as summarized in Tab. 7.2.

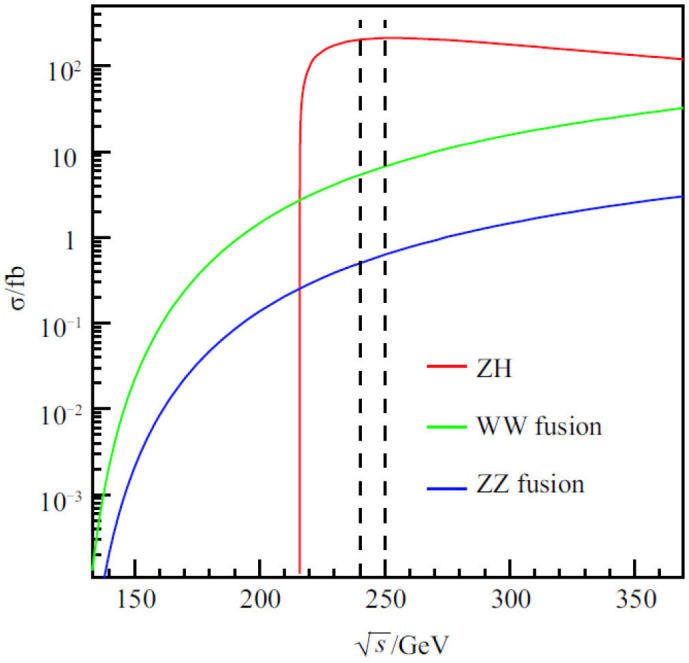


Figure 7.7: *The cross section for three process of Higgs production as a function of center mass of energy ($\sqrt{s}$) in the CEPC, under the assumption of Higgs mass of 125 GeV.*

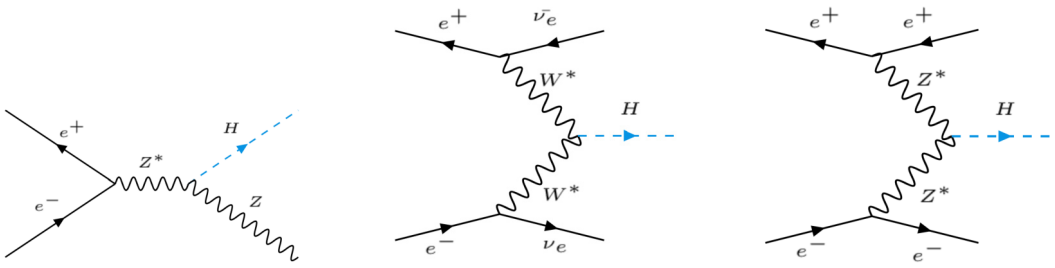


Figure 7.8: *The Feynman diagram of three Higgs production process. Left: ZH process. Middle: WW fusion process. Right: ZZ fusion process.*

7.2.1 Monte Carlo samples of Higgs signal

In the CEPC, the Higgs signal samples are generated as mentioned before using the Whizard generator, namely the version 1.95, with no beam polarization. The number of events are generated based on the expected integrated luminosity of 5ab^{-1} of $\sqrt{s}$ = 240 GeV. Because the ZH is the dominant process in the CEPC, the Higgs signal samples are categorised according to the final states with respect to the Z decay. For instance the the ZH with the Z ending into neutrinos will be in the same category as the WW fusion process.

Decay mode	Branching ratio (%)	relative uncertainty (%)
$H \to b\bar{b}$	57.7	+3.2, -3.3
$H \to c\bar{c}$	2.91	+12, -12
$H \to \tau^-\tau^+$	6.32	+5.7, -5.7
$H \to \mu^-\mu+$	2.19×10^{-2}	+6.0, -5.9
$H \to WW^*$	21.5	+4.3, -4.2
$H \to ZZ^*$	2.64	+4.3, -4.2
$H \to \gamma\gamma$	0.228	+5.0, -4.9
$H \to Z\gamma$	0.153	+9.0, -8.8
$H \to gg$	8.57	+10, -10

Table 7.2: *The Higgs boson decay modes and their corresponding branching ratios under the SM prediction of Higgs boson mass of 125 GeV. The uncertainty contributed by both theoretical and parametric sources [20].*

Tab. 7.3 summarizes the cross section of the different Higgs samples categories based on the final state content.

Final states	Cross section (fb)
$e^-e^+H^{(*)}$	7.04
$\mu^-\mu^+H$	6.77
$\tau^-\tau^+H$	6.75
$\nu\bar{\nu}H^{(\dagger)}$	46.29
$q\bar{q}H$	136.81

Table 7.3: *The summary of Higgs signal samples of SM generated by Whizard 1.95. (∗) This sample include the contributions from both ZH process and WW fusion process. (†) This sample include contributions from both ZH process and ZZ fusion process.*

7.3 background process

7.3.1 The two-fermion background process

The presence of the invisible neutrino and anti-neutrino in our studied signal $\nu\bar{\nu}$ H, explain the possibility to be contaminated by the SM two-fermion physics. There are indeed two processes responsible for this background. Their corresponding Feynman diagrams are shown in Fig. 7.9. One is s-channel process and the other is t-channel one.

The corresponding cross sections used for the generation of the two-fermion background samples are summarized in Tab. 7.4.

7.3.2 The four-fermion background process

According to the CEPC framwork of simulation samples production, the all four-fermion background can be categorised to six classes without overlap or omission, giving: (1) ZZ;

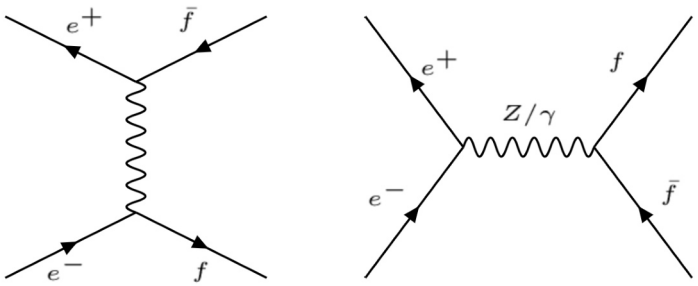


Figure 7.9: *Left: the Feynman diagram for two-fermions background through-channel process. the intermediate boson depends on the final states produced. Right: the Feynman diagram for two-fermions background through-channel process. The $f/\bar{f}$ represents the fermions/anti-fermions allowed (quarks, leptons).*

Final states	Cross section (fb)
$e^-e^+(*)$	24770.90
$\mu^-\mu^+$	5332.71
$\tau^-\tau^+$	4752.89
$\nu\bar{\nu}$	54099.51
$q\bar{q}$	54106.86

Table 7.4: *Summary of the SM two-fermion background samples generated by Whizard 1.95. (*).*

(2) WW; (3) ZZ or WW; (4) single Z; (5) single W; (6) single Z or single W. The more detail information about such classification can be found at Ref. [67].

The main Feynman diagrams of four-fermions backgrounds process are shown in Fig. 7.10. The internal curly lines can be any gauge bosons permitted not only W or Z bosons. For WW process, there is an extra Feynman diagram as shown in Fig. 7.11. For "single Z ", "single W" and "single Z or single W" process, there are two additional Feynman diagrams as shown in Fig. 7.12. Such classification is only important for easily allowing simulation production. For easier to connect the final stats with the class name, the each of six classes can be further grouped according to the content of its final states. The final states comprised by four quarks (qqqq), four leptons (l(ν)l(ν)l(ν)l(ν)) and the one including two quarks and two leptons(qql(ν)l(ν)) are denoted as "hadronic", "leptonic " and "semi-leptonic" respectively.

Due to the fact that there are two neutrinos and four jets in our signal channel, thus the four-fermion background with semi-leptonic final stats are the main backgrounds.

The cross sections used for generation four-fermions background are summarized in Tab. 7.5

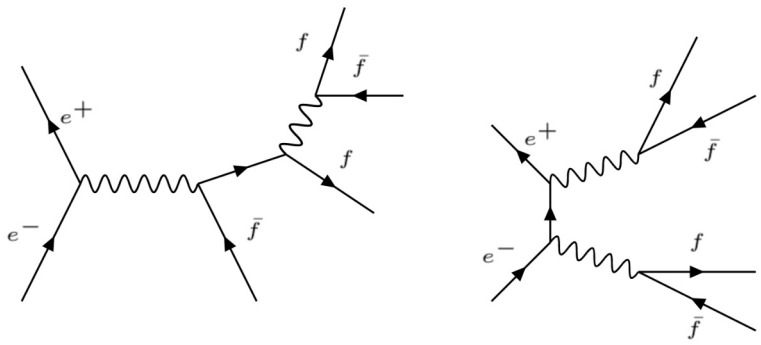


Figure 7.10: *The Feynman digrams for main process for four-fermions background.*

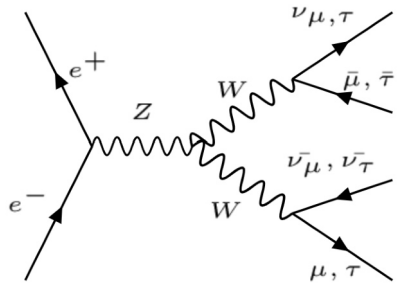


Figure 7.11: *The additional Feynman diagram for WW process, namely triple gauge coupling porcess. The lepton and its accompanying neutrino of the diagram can be replace by up quarks and down quarks respectively.*

process	Cross section (fb)
ZZ	1140.97
WW	9076.11
ZZ or WW	3650.15
single Z	1620.17
single W	3485.25
single Z or single W	249.48

Table 7.5: *The summary of cross sections for four-fermions background samples categorised as six classes.*

7.4 Analysis strategy

To measure the Higgs coupling constant the W boson (g_{HWW}) we propose to study the channel shown in Fig. 7.13. Here, the Higgs bosons are generated through the WW fusion process but then they also decay to two W bosons with these last ones decaying into jets.

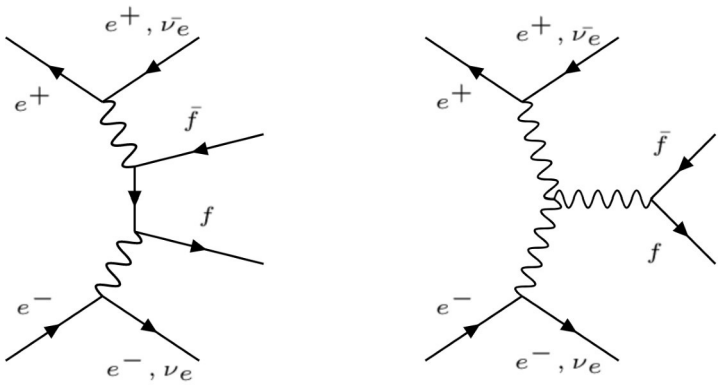


Figure 7.12: *The additional Feynman diagrams for "single Z", "single W" and "single Z or single W" process. The $f/\bar{f}$ is a (an) fermion/anti-fermion.*

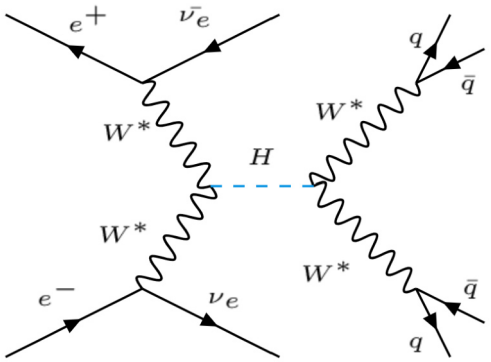


Figure 7.13: *The Feynman diagram of the studied channel. The q and $\bar{q}$ do not indicate they have the same flavor.*

7.4.1 Event selection

The event selection of our signal is made in two steps: the pre-selection step and the final selection step where the BDT method is used.

7.4.1.1 pre-selection for event selection

The cuts employed by pre-selection are the following:

- The events are reconstructed by using Arbor PFA to have four jets. As can be seen in Fig. 7.13, the invariant mass of the four jets (m_4jets) should be in the mass range of the Higgs boson. Therefore, the m_4jets are required to be in the range from 105 GeV to 145 GeV, 105 GeV < m_4jets < 145 GeV. The distribution of m_4jets for

backgrounds and signal are shown in the Fig. 7.14.

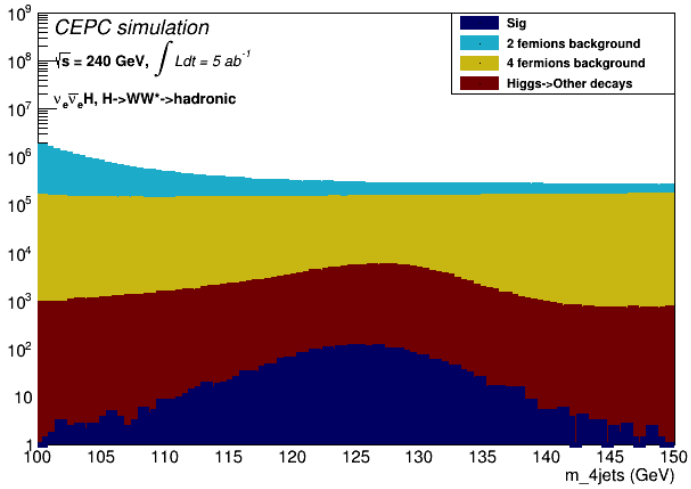


Figure 7.14: *The distribution of invariant mass of four jets for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and intergrated luminorsity 5 ab^{-1}.*

- For the higgs boson decaying to two W bosons, one is on-shell W boson (referred to as W_1) and another is the off-shell one (W_2). Therefore, one jet pair which has minimum invariant mass difference with respect to the invariant mass of W boson (quoted from PDG table) are selected among four jets. The invariant mass of this jet pair is referred to m_{W1} and the remaining jet pair are assigned to the off-shell W boson , referred to as m_{W2}. The corresponding distribution of m_{W1} and m_{W2} are shown in Fig. 7.15. 60 GeV < m_{W1} <100 GeV, 10 GeV < m_{W2} <60 GeV.

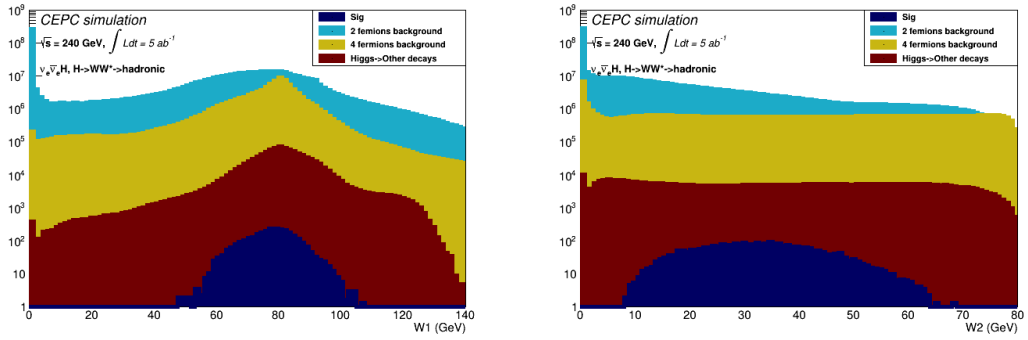


Figure 7.15: *The distribution of invariant mass of the jet pair decaying from on-shell W boson (left figure) and the jet pair decaying from off-shell W boson (right figure) for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and intergrated luminorsity 5 ab^{-1}.*

- the transverse momentum of the reconstructed on-shell W boson $P_t(W_1)$ shown in

Fig. 7.16. $P_t(W_1) < 80$ GeV.

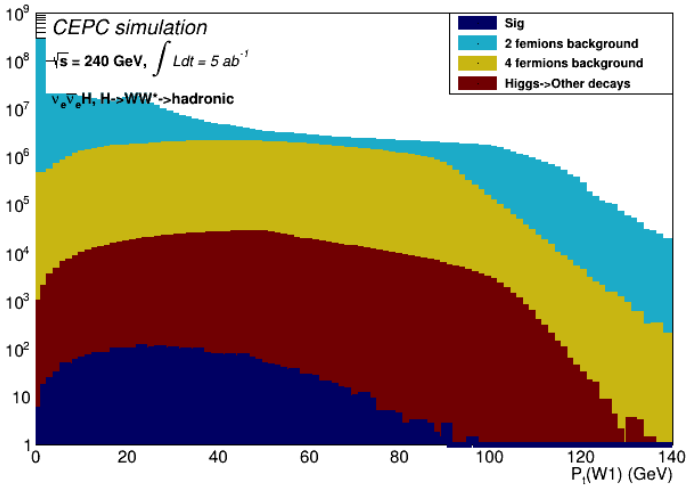


Figure 7.16: *The distribution of transverse momentum of reconstructed on-shell W boson for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and intergrated luminorsity 5 ab^{-1}.*

- the cos(JetAngle) where the "JetAngle" is the angle between two jets decaying from the one-shell W boson. Its distribution is shown in Fig. 7.17 and it is required to be less than 0.3.

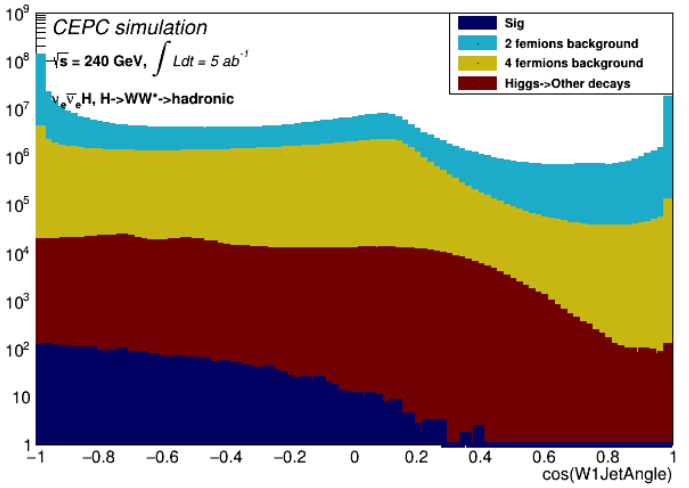


Figure 7.17: *The distribution of $cos(W1JetAngle)$ for signal and backgrounds, where the "W1JetAngle" is the angle between two jets decaying from on-shell W boson. The number of events for $\sqrt{s}$ = 240 GeV and intergrated luminorsity 5 ab^{-1}.*

- The jet reconstruction parameters such as $-log(y_{12})$, $-log(y_{23})$, $-log(y_{34})$, $-log(y_{45})$,

$-log(y_{56})$, $-log(y_{67})$. The corresponding cut value ranges are [0, 2.6], [0, 5.6], [0, 7.6], [0, 9.1] , [0, 9.5] and [0, 9.8] respectively. They are shown in Fig. 7.18.

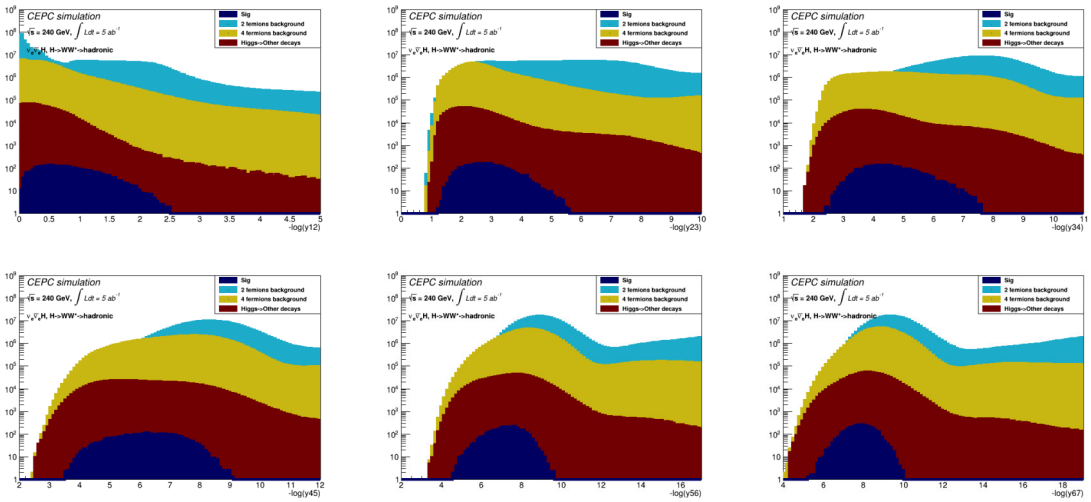


Figure 7.18: *The distribution of parameters of jet reconstruction for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and integrated luminosity 5 ab^{-1}.*

- The number of reconstructed particle flow objects (nPFOs). Because of the nature of the jets associated to the W decays in our siganl, the nPFOs is expected to be high which allows discriminating our signal from backgrounds with leptonic final states as can be seen in Fig. 7.19. Therefore, the nPFOS is required to be in the range [30, 70].

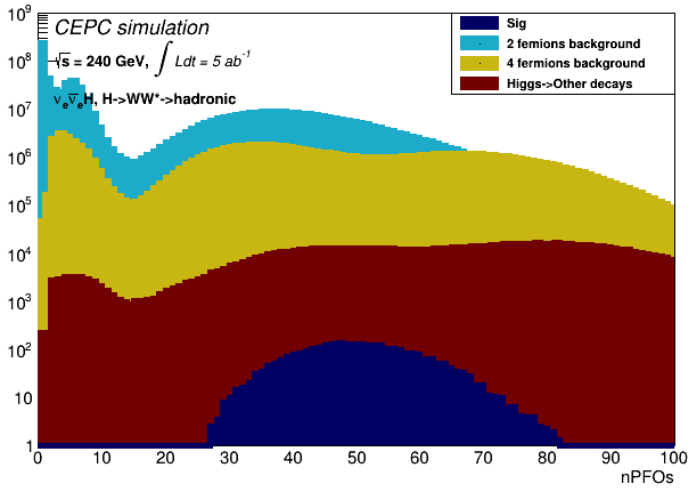


Figure 7.19: *The distribution of the nPFOs for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and integrated luminosity 5 ab^{-1}.*

- In the WW fusion process, two neutrinos are produced accompanying the production of Higgs boson. Therefore, the missing energy (E_{mis}) is not negligible as shown in Fig. 7.20. We require E_{mis} to be in the range [70 GeV, 130 GeV].

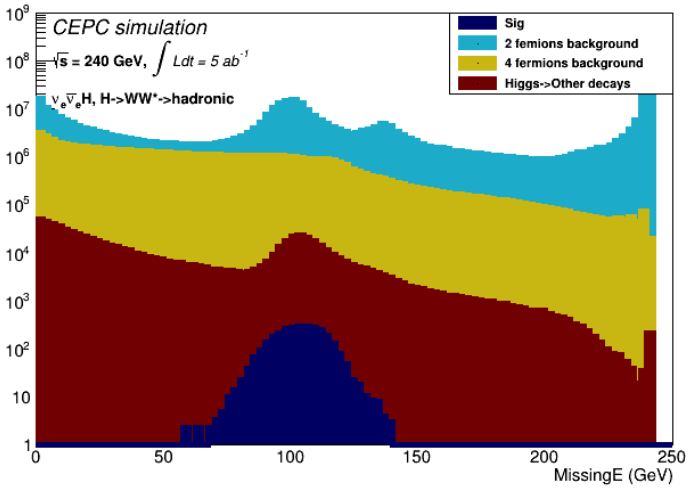


Figure 7.20: *The distribution of the missing energy for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and integrated luminosity 5 ab^{-1}.*

- The visible transverse momentum P_t and it is required to be less than 90 GeV. It is shown in Fig. 7.21.

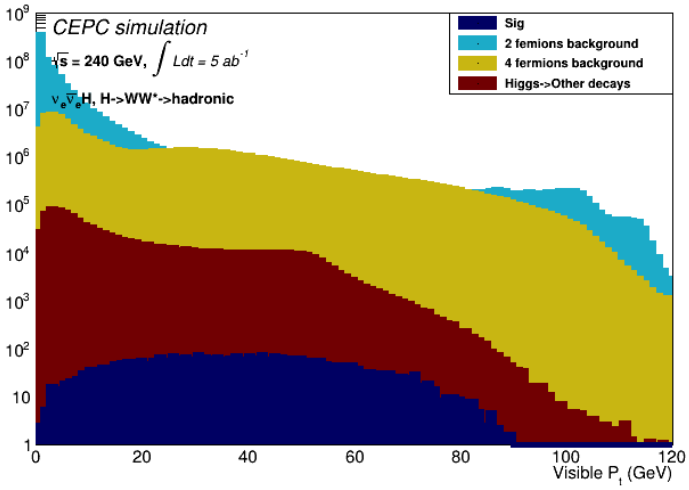


Figure 7.21: *The distribution of the visible transverse mentum for signal and backgrounds. The number of events for $\sqrt{s}$ = 240 GeV and intergrated luminorsity 5 ab^{-1}.*

After the pre-selection discussed above, the invariant mass of four jets is shown in Fig. 7.22. The efficiencies of pre-selection for signal and backgrounds are summarized in

the Tab. 7.6.

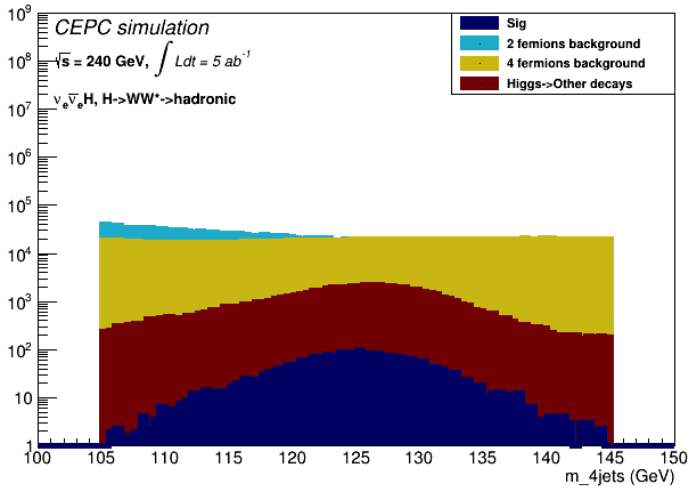


Figure 7.22: *The distribution of invariant mass of four jets for signal and backgrounds after the pre-selection applied. The number of events are normalized by using the integrated luminosity 5 ab^{-1} at $\sqrt{s}$ = 240 GeV.*

process	Cross section (fb)	ε_{pre}(%)	ε_{final}(%)
Signal	0.60	78.10	42.97
Higgs→ other decays	203.06	7.33	1.34
two-fermion backgrounds: $q\bar{q}$			
$e^-e^+ \to q\bar{q}$	54106.86	0.6	4.60×10^{-03}
four-fermion backgrounds: ZZ			
ZZ_h	516.67	0.117	5.93×10^{-03}
ZZ_sl	556.49	2.52	0.13
four-fermion backgrounds: WW			
WW_h	3825.46	0.045	2.76×10^{-03}
WW_sl	4846.99	0.403	0.11
four-fermion backgrounds: ZZ or WW			
ZZ or WW_h	3217.87	0.049	2.51×10^{-03}
four-fermion backgrounds: single Z			
SZE_sl	316.04	0.574	3.00×10^{-03}
SZNU_sl	145.62	3.14	0.18
four-fermion backgrounds: single W			
SW_sl	2612.62	1.25	0.02

Table 7.6: *The list of signal efficiency and background reduction efficiencies for applying pre-selection (ε_{pre}) and final selection (ε_{final}). Only the processes whose final efficiency exceed 10^{-7} are kept in this list. The "h", and "sl" of the end of the process name are referred to as "hadronic" and "semi-leptonic" final states respectively. The detail information for backgrounds classification can be found in Ref. [67].*

7.4.1.2 Final selection

After the pre-selection, the final selection based on the BDT method are performed. All samples after pre-selection applied are used to training and test phase of BDT. The kinematic variables of events are fed into the BDT including:

- The number of particle flow objects, nPFOs.
- The jet reconstruction parameters, $-log(y_{12})$, $-log(y_{23})$, $-log(y_{34})$, $-log(y_{45})$, $-log(y_{56})$, $-log(y_{67})$.
- The total missing energy, E_{mis}.
- The invariant masses of reconstructed four jets, on-shell and off-shell W boson referred to as m_4jets, m_{W1} and m_{W2}.
- The transverse momentum of missing object, on-shell and off-shell W bosons, Missing P_t, $P_t(W1)$, $P_t(W2)$.
- The angles between two jets of on-shell W boson and remaining two jets of off-shell W boson, cos(W1JetAngle), cos(W2JetAngle).
- The angle between on-shell and off-shell W boson, cos(W1W2).

Concerning the optimal cut value of the BDT output to separate the background and signal, it can be transformed to find the cut-off value which maximizes the significance S:

$$S = \frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}}$$

where N_{sig} and N_{bkg} are number of events of signal and backgrounds respectively. The S as a function of cut value of BDT output is shown in Fig. 7.23 .

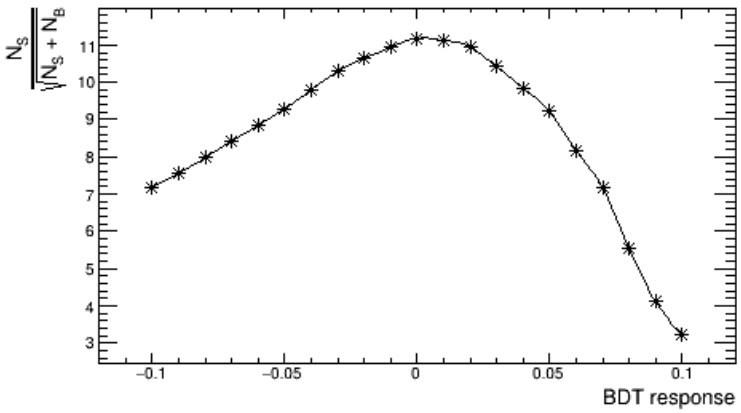


Figure 7.23: *The significance S as a function of cut value of BDT response. The maximum of significance (S = 11.17) can be obtained when cut value equals to 0.0.*

we can easily find the cut-off value and its corresponding optimal S = 11.17. The final selection efficiencies are summarized in the Tab. 7.6. The invariant mass of four jets after final selection is shown in Fig. 7.24.

7.4.2 The measurement of g_{HWW}

According to Fig. 7.13, the number of events associated to the studied channel can be calculated by the equation:

$$N_{sig} = L_{integrated} \times \sigma_{\nu_e\bar{\nu}_eH} \times \frac{Br(H \to WW \to 4jets)}{Br(W \to 2jets)^2} \times eff_{sig},$$

where the $L_{integrated}$ is the integrated luminosity of CEPC , $\sigma_{\nu_e\bar{\nu}_eH}$ and eff_{sig} are the cross-section of WW fusion process and efficiency of signal event selection respectively. The N_{sig} is proportional to the Higgs coupling constant g_{HWW}, given by:

$$N_{sig} \propto g^4_{HWW}.$$

Thus, the relative statistical uncertainty for g_{HWW}:

$$\frac{\triangle(\sigma \times Br)}{\sigma \times Br} = S^{-1} = 4\frac{\triangle g_{HWW}}{g_{HWW}},$$

which gives value of $\frac{\triangle g_{HWW}}{g_{HWW}} \sim 2.24\%$.

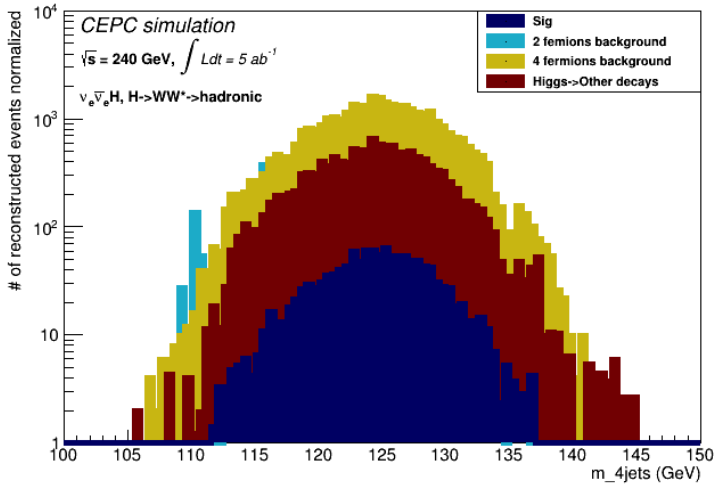


Figure 7.24: *The distribution of invariant mass of four jets for signal and backgrounds after the pre-selection and final-selection applied. The number of events are normalized by using the integrated luminosity 5 ab^{-1} at $\sqrt{s}$ = 240 GeV.*

7.4.2.1 Systematic uncertainty

The some possible sources of the systematic uncertainties of measurement of g_{HWW} are listed below.

- **The theory uncertainties**. The uncertainty of calculation of WW fusion process cross-section is negligible [20]. The relative uncertainty of $Br(H \rightarrow WW)$ is found to be 4.2%.
- **The experimental uncertainties**. (a) The relative uncertainty of luminosity is 0.1% [68]. (b) The relative uncertainty of event selection efficiency is found to be 1.63% which is estimated by varying BDT cut value from -0.001 to 0.001.

Thus the total systematic uncertainties equal to the quadratic sum of factors which are discussed in Section. 7.4.2.1 . It equals to the 4.51% shown in Tab. 7.7. The theoretical uncertainty is the dominant part of the total systematic uncertainties.

measurement (g_{HWW})	Uncertainties
$\triangle g_{HWW}/g_{HWW}$	2.24% ($stats$) ± 4.51% (sys)

Table 7.7: *The statistical uncertainty and systematic uncertainties of measurement of g_{HWW} at CEPC.*

7.5 Summary

The precision measurement of Higgs properties is a central task of CEPC project. With the PFA-oriented baseline detector, the precision measurement of Higgs coupling g_{HWW}

is studied through the process: $e^+e^- \to \nu_e\bar{\nu}_e\, H(H \to W^+W^- \to q\bar{q}'q\bar{q}')$. The statistical uncertainty of measurements of g_{HWW} is found to be 2.24%.

Conclusion and outlook

The standard Model is a remarkable achievements of modern particle physics. The Standard Model particle list are completed by the discovery of Higgs Boson in 2012.

In order to precisely measure the Higgs boson properties and search new physics of BSM, particle physicist community proposed several future collider projects such as ILC, CLIC, CEPC and FCC. CEPC is a circular leptonic collider projects proposed to locate in China. The abundant physics potential and technology options (accelerator complex and detector concept) are well studied.

The SDHCAL concept is one of most attractive technological HCAL options and the SDHCAL technological prototype is the first technological prototype of PFA-oriented high-granularity HCAL family. It exposed in many times of particle beam at PS and SPS of CERN. Based on the data collected in test beam periods, the glass-based RPC is found to have very good performance: 96% ± 0.04 (efficiency) and 1.81 ± 0.19 (multiplicity). Due to the fact of larger performance fluctuation existing in some areas of SDHCAL technological prototype, the homogenization process are also studied and it was found to obtain more better response of detector that the relative deviation of nHit improves from 4.9% and 3.9% to 1.8% and 1.4% for 50 and 70 GeV pions respectively after homogenization process applied.

Due to the fact that the standard method for hadronic event selection loses relative large fraction of events, the new event selection method based on BDT method are developed. Thanks to the high-granularity of SDHCAL technological prototype, the BDT-based method achieves much better performance than standard event selection method. The BDT method significantly improves the signal (pion) efficiency (>99%) and background rejection rates (>99%) comparing with the cut-based standard method. In the standard energy reconstruction method, only the information of number of hits is used. However, the SDHCAL technological providing rich topological information of events, thus a series of topological variables are built. Then, the MVA methods: MLP and BDT are used to exploit the correlation between such variables and energy by building corresponding regression models. These two methods are found to achieve better energy linearity and resolution. The energy linearity is found to improve from the 3-4% to 1-2%. The energy resolution was relatively improved from 15% at 10 GeV and 3-5% at higher energies (20-80 GeV).

Finally, using simulation samples produced with 5 ab^{-1} integrated luminosity, the physics channel ($e^+e^- \to \nu_e\bar{\nu}_e\, H(H \to W^+W^- \to q\bar{q}'q\bar{q}')$) are studied for measuring Higgs coupling constant g_{HWW}, with CEPC baseline detector in $\sqrt{s}$ = 240 GeV. The precison of measurement of Higgs coupling constant g_{HWW} is found to be 2.24% (stats) ± 4.51% (sys).

In the future, the SDHCAL will upgrade to have timing capability with the use of

multi-gap RPC, it will improve the separation between two hadronic showers located closely in space.

In addition, to apply the MVA methods to beam data for energy reconstruction, on the one side, we can try to improve the simulation performance to reproduce the data in better way, on the other side, we can expose the SDHCAL prototype to hadron beams with many energy points with fine energy steps and use the data samples to train and test.

Bibliography

[1] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[2] Povh, B and Rith, K and Scholz, C and Zetsche, F. *Particles and Nuclei*. Springer-Verlag Berlin Heidelberg, 2008.

[3] J. J. Thomson M.A. F.R.S. Xl. cathode rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(269):293–316, 1897.

[4] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the free neutrino: a confirmation. *Science*, 124(3212):103–104, 1956.

[5] D. DeCamp; et al. ALEPH Collaboration. Determination of the number of light neutrino species. *Physics Letters B*, 231(4):519 – 529, 1989.

[6] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[7] Abdus Salam and John Clive Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13:168–171, 1964.

[8] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[9] P.A. Zyla et al. (Particle Data Group). Review of Particle Physics. *Prog. Theor. Exp. Phys.*, 2020.

[10] Georges Aad, Tatevik Abajyan, B Abbott, J Abdallah, S Abdel Khalek, Ahmed Ali Abdelalim, R Aben, B Abi, M Abolins, OS AbouZeid, et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.

[11] Serguei Chatrchyan, Vardan Khachatryan, Albert M Sirunyan, Armen Tumasyan, Wolfgang Adam, Ernest Aguilo, Thomas Bergauer, M Dragicevic, J Erö, C Fabjan, et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.

[12] Observation of an excess of events in the search for the Standard Model Higgs boson in the $H \to ZZ^* \to 4l$ channel with the ATLAS detector. Technical Report ATLAS-CONF-2012-169, CERN, Geneva, Dec 2012.

[13] CMS Collaboration and Thomas McCauley, Lucas Taylor. Candidate events in the CMS Standard Model Higgs Search using 2010 and 2011 data. CMS Collection., Dec 2011.

[14] M Tanabashi, Particle Data Group, et al. Status of higgs boson physics. *Phys. Rev. D*, 98:030001, 2018.

[15] Kien Nguyen. The higgs mechanism. 2009.

[16] Qiuhong Wang, Abdusalam Abdukerim, Wei Chen, Xun Chen, Yunhua Chen, Chen Cheng, Xiangyi Cui, Yingjie Fan, Deqing Fang, Changbo Fu, Mengting Fu, Lisheng Geng, Karl Giboni, Linhui Gu, Xuyuan Guo, Ke Han, Changda He, Shengming He, Di Huang, Yan Huang, Yanlin Huang, Zhou Huang, Xiangdong Ji, Yonglin Ju, Shuaijie Li, Huaxuan Liu, Jianglai Liu, Wenbo Ma, Yugang Ma, Yajun Mao, Yue Meng, Kaixiang Ni, Jinhua Ning, Xuyang Ning, Xiangxiang Ren, Changsong Shang, Lin Si, Guofang Shen, Andi Tan, Anqing Wang, Hongwei Wang, Meng Wang, Siguang Wang, Wei Wang, Xiuli Wang, Zhou Wang, Mengmeng Wu, Shiyong Wu, Weihao Wu, Jingkai Xia, Mengjiao Xiao, Pengwei Xie, Binbin Yan, Jijun Yang, Yong Yang, Chunxu Yu, Jumin Yuan, Ying Yuan, Jianfeng Yue, Xinning Zeng, Dan Zhang, Tao Zhang, Li Zhao, Qibin Zheng, Jifang Zhou, Ning Zhou, and Xiaopeng Zhou. Results of dark matter search using the full pandax-ii exposure, 2020.

[17] E. Aprile, J. Aalbers, F. Agostini, M. Alfonsi, L. Althueser, F. D. Amaro, V. C. Antochi, E. Angelino, J. R. Angevaare, F. Arneodo, D. Barge, L. Baudis, B. Bauermeister, L. Bellagamba, M. L. Benabderrahmane, T. Berger, A. Brown, E. Brown, S. Bruenner, G. Bruno, R. Budnik, C. Capelli, J. M. R. Cardoso, D. Cichon, B. Cimmino, M. Clark, D. Coderre, A. P. Colijn, J. Conrad, J. P. Cussonneau, M. P. Decowski, A. Depoian, P. Di Gangi, A. Di Giovanni, R. Di Stefano, S. Diglio, A. Elykov, G. Eurin, A. D. Ferella, W. Fulgione, P. Gaemers, R. Gaior, M. Galloway, F. Gao, L. Grandi, C. Hasterok, C. Hils, K. Hiraide, L. Hoetzsch, J. Howlett, M. Iacovacci, Y. Itow, F. Joerg, N. Kato, S. Kazama, M. Kobayashi, G. Koltman, A. Kopec, H. Landsman, R. F. Lang, L. Levinson, Q. Lin, S. Lindemann, M. Lindner, F. Lombardi, J. Long, J. A. M. Lopes, E. López Fune, C. Macolino, J. Mahlstedt, A. Mancuso, L. Manenti, A. Manfredini, F. Marignetti, T. Marrodán Undagoitia, K. Martens, J. Masbou, D. Masson, S. Mastroianni, M. Messina, K. Miuchi, K. Mizukoshi, A. Molinario, K. Morå, S. Moriyama, Y. Mosbacher, M. Murra, J. Naganoma, K. Ni, U. Oberlack, K. Odgers, J. Palacio, B. Pelssers, R. Peres, J. Pienaar, V. Pizzella, G. Plante, J. Qin, H. Qiu, D. Ramírez García, S. Reichard, A. Rocchetti, N. Rupp, J. M. F. dos Santos, G. Sartorelli, N. Šarčević, M. Scheibelhut, J. Schreiner, D. Schulte, M. Schumann, L. Scotto Lavina, M. Selvi, F. Semeria, P. Shagin, E. Shockley, M. Silva, H. Simgen, A. Takeda, C. Therreau, D. Thers, F. Toschi, G. Trinchero, C. Tunnell, M. Vargas, G. Volta, H. Wang, Y. Wei, C. Weinheimer, M. Weiss, D. Wenz, C. Wittweg, Z. Xu, M. Yamashita, J. Ye, G. Zavattini, Y. Zhang, T. Zhu, J. P. Zopounidis, and X. Mougeot. Observation of excess electronic recoil events in xenon1t, 2020.

[18] MC Carmona-Benitez, DS Akerib, HM Araújo, X Bai, AJ Bailey, J Balajthy, P Beltrame, E Bernard, A Bernstein, A Bradley, et al. First results of the lux dark matter experiment. *Nuclear and particle physics proceedings*, 273:309–313, 2016.

[19] Csaba Csáki and Philip Tanedo. Beyond the standard model. *arXiv preprint arXiv:1602.04228*, 2016.

[20] The CEPC Study Group. Cepc conceptual design report: Volume 2 - physics & detector, 2018.

[21] The CEPC Study Group. Cepc conceptual design report: Volume 1 - accelerator, 2018.

[22] Guido Montagna, Oreste Nicrosini, and Fulvio Piccinini. On the qed radiator at order α^3. *Physics Letters B*, 406(3):243–248, Aug 1997.

[23] Guido Montagna, Fulvio Piccinini, Oreste Nicrosini, Giampiero Passarino, and Roberto Pittau. On a semi-analytical and realistic approach to e+ e- annihilation into fermion pairs and to bhabha scattering within the minimal standard model at lep energies. *Nuclear Physics B*, 401(1-2):3–66, 1993.

[24] Guido Montagna, Fulvio Piccinini, Oreste Nicrosini, Giampiero Passarino, and Roberto Pittau. Topaz0-a program for computing observables and for fitting cross sections and forward-backward asymmetries around the z0 peak. *Computer Physics Communications*, 76(3):328–360, 1993.

[25] Guido Montagna, Oreste Nicrosini, Fulvio Piccinini, and Giampiero Passerino. Topaz0 4.0 — a new version of a computer program for evaluation of deconvoluted and realistic observables at lep 1 and lep 2. *Computer Physics Communications*, 117(3):278–289, Mar 1999.

[26] Koya Abe, Kenji Abe, T. Abe, I. Adam, H. Akimoto, D. Aston, K. G. Baird, C. Baltay, H. R. Band, T. L. Barklow, and et al. Improved direct measurement of leptonic coupling asymmetries with polarizedzbosons. *Physical Review Letters*, 86(7):1162–1166, Feb 2001.

[27] G Abbiendi, C Ainsley, P.F Åkesson, G Alexander, J Allison, P Amaral, G Anagnostou, K.J Anderson, S Arcelli, S Asai, and et al. Measurement of heavy quark forward–backward asymmetries and average b mixing using leptons in hadronic z decays. *Physics Letters B*, 577(1-2):18–36, Dec 2003.

[28] Measurement of the forward-backward asymmetries of $e^+e^- \to z \to b\bar{b}$ and $e^+e^- \to z \to c\bar{c}$ using prompt leptons. *The European Physical Journal C*, 34(2):109–125, May 2004.

[29] Belle II Collaboration. T. Abe et al. Belle ii technical design report, 2010.

[30] Jingyu Tang, J. Scott Berg, Weiping Chai, Fusan Chen, Nian Chen, Weiren Chou, Haiyi Dong, Jie Gao, Tao Han, Yongbin Leng, Guangrui Li, Ramesh Gupta, Peng Li, Zhihui Li, Baiqi Liu, Yudong Liu, Xinchou Lou, Qing Luo, Ernie Malamud, Lijun Mao, Robert B. Palmer, Quanling Peng, Yuemei Peng, Manqi Ruan, GianLuca Sabbi, Feng Su, Shufang Su, Diktys Stratakis, Baogeng Sun, Meifen Wang, Jie Wang, Liantao Wang, Xiangqi Wang, Yifang Wang, Yong Wang, Ming Xiao, Qingzhi Xing, Qingjin Xu, Hongliang Xu, Wei Xu, Holger Witte, Yingbing Yan, Yongliang Yang, Jiancheng

Yang, Youjin Yuan, Bo Zhang, Yuhong Zhang, Shuxin Zheng, Kun Zhu, Zian Zhu, and Ye Zou. Concept for a future super proton-proton collider, 2015.

[31] Manqi Ruan. Arbor, a new approach of the particle flow algorithm. *arXiv preprint arXiv:1403.4784*, 2014.

[32] The ILD Collaboration and contact Ties Behnke. The ild detector at the ilc, 2019.

[33] CALICE Collaboration, C. Adloff, Y. Karyotakis, J. Repond, A. Brandt, H. Brown, K. De, C. Medina, J. Smith, J. Li, et al. Construction and commissioning of the CALICE analog hadron calorimeter prototype. *Journal of Instrumentation*, 5:5004, May 2010.

[34] A Cardini, M Cascella, Suyong Choi, D De Pedis, R Ferrari, S Franchino, G Gaudio, S Ha, J Hauptman, L La Rotonda, et al. The small-angle performance of a dual-readout fiber calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 808:41–53, 2016.

[35] N Akchurin, F Bedeschi, A Cardini, M Cascella, Fabrizio Cei, D De Pedis, R Ferrari, S Fracchia, S Franchino, M Fraternali, et al. The electromagnetic performance of the rd52 fiber calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 735:130–144, 2014.

[36] Richard Wigmans. New results from the RD52 project. *Nucl. Instrum. Meth. A*, 824:721–725, 2016.

[37] T. Behnke, J. E. Brau, B. Foster, J. Fuster, M. Harrison, J. McEwan Paterson, M. Peskin, M. Stanitzki, N. Walker, and H. Yamamoto. The International Linear Collider Technical Design Report - Volume 1: Executive Summary. *ArXiv e-prints*, June 2013.

[38] Lucie Linssen, Akiya Miyamoto, Marcel Stanitzki, and Harry Weerts. Physics and detectors at clic: Clic conceptual design report, 2012.

[39] A Abada, M Abbrescia, SS AbdusSalam, I Abdyukhanov, J Abelleira Fernandez, A Abramov, M Aburaia, AO Acar, PR Adzic, P Agrawal, et al. Fcc-ee: the lepton collider. *The European Physical Journal Special Topics*, 228(2):261–623, 2019.

[40] R. Kopp et al. A Measurement of Energy Loss Distributions of Energetic Muons in Iron. *Z. Phys. C*, 28:171, 1985.

[41] M Berger and S Seltzer. 10. tables of energy-losses and ranges of electrons and positrons. *Studies in penetration of charged particles in matter*, (39):205, 1964.

[42] Atomic and nuclear properties of materials for more than 350 materials. `https://pdg.lbl.gov/2020/AtomicNuclearProperties/index.html`. Last revised 04 June 2020.

[43] Michele Livan and Richard Wigmans. *Calorimetry for Collider Physics, an Introduction*. Springer, 2019.

[44] Michele Livan and Richard Wigmans. *Calorimetry for Collider Physics, an Introduction*. Springer, 2019.

[45] S. Callier, F. Dulucq, C. de La Taille, N. Martin-Chassard, and N. Seguin-Moreau. HARDROC, Readout chip of the Digital Hadronic Calorimeter of ILC. 2009.

[46] The CALICE Collaboration. First results of the CALICE SDHCAL technological prototype. *Journal of Instrumentation*, 11:P04001, April 2016.

[47] Guillaume Garillot. *Study of hadronic shower in a highly granular calorimeter and study of the e+e- → HZ (Z → qq) channel in future leptonic colliders*. Theses, Université de Lyon, February 2019.

[48] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[49] Z. Deng et al. Resistive Plate Chamber Digitization in a Hadronic Shower Environment. *JINST*, 11(06):P06014, 2016.

[50] V.V. Uzhinsky. The Fritiof (FTF) Model in Geant4. In *International Conference on Calorimetry for the High Energy Frontier*, pages 260–264, 2013.

[51] G Folger, VN Ivanchenko, and JP Wellisch. The binary cascade. *The European Physical Journal A-Hadrons and Nuclei*, 21(3):407–417, 2004.

[52] Arnaud Steen. *Hadronic shower study with the semi-digital hadronic calorimeter and comparison with theoretical models used in GEANT4*. Theses, Université Claude Bernard - Lyon I, November 2015.

[53] Laura Elena Raileanu and Kilian Stoffel. Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, 41(1):77–93, 2004.

[54] what are the advantages of the different impurity metrics. `https://sebastianraschka.com/faq/docs/decision-tree-binary.html`.

[55] Hyafil Laurent and Ronald L Rivest. Constructing optimal binary decision trees is np-complete. *Information processing letters*, 5(1):15–17, 1976.

[56] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[57] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

[58] Byron P Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 543(2-3):577–584, 2005.

[59] Hai-Jun Yang, Byron P Roe, and Ji Zhu. Studies of boosted decision trees for miniboone particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1-2):370–385, 2005.

[60] Joseph F. Hair. *Multivariate Data Analysis: An Overview*, pages 904–907. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[61] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. Tmva - toolkit for multivariate data analysis, 2007.

[62] Z. Deng, Y. Li, Y. Wang, Q. Yue, Z. Yang, D. Boumediene, C. Carloganu, V. Français, G. Cho, D-W. Kim, S. C. Lee, Z. Liu, W. Park, S. Vallecorsa, S. Cauwenbergh, M. Tytgat, A. Pingault, N. Zaganidis, O. Bach, E. Brianne, A. Ebrahimi, K. Gadow, P. Göttlicher, O. Hartbrich, A. Irles, K. Kotera, F. Krivan, K. Krüger, S. Lu, C. Neubüser, A. Provenza, M. Reinecke, F. Sefkow, S. Schuwalow, Y. Sudo, H.L. Tran, H. Hirai, K. Kawagoe, T. Suehara, H. Sumida, T. Yoshioka, E. Cortina Gil, S. Mannai, V. Buridon, C. Combaret, L. Caponetto, R. Eté, G. Garillot, G. Grenier, R. Han, J.C. Ianigro, R. Kieffer, T. Kurca, I. Laktineh, B. Li, N. Lumb, H. Mathez, L. Mirabito, A. Petrukhin, A. Steen, J. Berenguer Antequera, E. Calvo Alamillo, M.-C. Fouz, J. Marin, J. Navarrete, J. Puerta-Pelayo, A. Verdugo, F. Corriveau, M. Chadeeva, M. Gabriel, P. Goecke, C. Graf, Y. Israeli, N. van der Kolk, F. Simon, M. Szalay, H. Windel, S. Bilokin, J. Bonis, R. Pöschl, A. Thiebault, F. Richard, D. Zerwas, M. Anduze, V. Balagura, E. Becheva, V. Boudry, J-C. Brient, R. Cornat, F. Gastaldi, Y. Haddad, F. Magniette, J. Nanni, M. Ruan, M. Rubio-Roy, K. Shpak, T.H. Tran, H. Videau, D. Yu, S. Callier, F. Dulucq, Ch. de la Taille, G. Martin-Chassard, L. Raux, N. Seguin-Moreau, J. Cvach, M. Janata, M. Kovalcuk, J. Kvasnicka, I. Polak, J. Smolik, V. Vrba, J. Zalesak, and J. Zuklin. Tracking within hadronic showers in the calice sdhcal prototype using a hough transform technique. *Journal of Instrumentation*, 12(05):P05009, 2017.

[63] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[64] Wolfgang Kilian, Thorsten Ohl, and Jürgen Reuter. Whizard—simulating multiparticle processes at lhc and ilc. *The European Physical Journal C*, 71(9), Sep 2011.

[65] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. Madgraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), Jun 2011.

[66] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[67] Xin Mo and Gang Li. Generated Sample Status for CEPC Simulation Studies. `http://cepcdoc.ihep.ac.cn/DocDB/0000/000034/003/CEPCNoteCover.pdf`.

[68] The ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration, the OPAL Collaboration, and the LEP Electroweak Working Group. Electroweak measurements in electron-positron collisions at w-boson-pair energies at lep. *Physics Reports*, (4):119–244, Nov 2013.