Reconstruction of leptonic physic objects at future e+e- Higgs factory

Dan Yu

université
PARIS-SACLAY

NNT : 2018SACLX018

Thèse de doctorat

Reconstruction of leptonic physics objects at future e+e- Higgs factory

Thèse de doctorat de l'Université Paris-Saclay
préparée à Ecole Polytechnique

Ecole doctorale n°576 particules hadrons énergie et noyau : instrumentation, image, cosmos et simulation (Pheniics)
Spécialité de doctorat : Physique des particules

Thèse présentée et soutenue à Palaiseau, le 08/02/2018, par

DAN YU

Composition du Jury :

Zhiqing Zhang LAL-IN2P3	Président
Mark Thomson University of Cambridge	Rapporteur
Luca Malgeri CERN	Rapporteur
Florian Beaudette LLR-IN2P3	Examinateur
Vincent Boudry LLR-IN2P3	Co-directeur de thèse
Manqi RUAN IHEP-CAS	Co-directeur de thèse

Abstract

The Standard Model of elementary particle interactions is the outstanding achievement of the past forty years of experimental and theoretical activity in particle physics. Since the discovery of the Higgs boson in 2012 by the experiments at the Large Hadron Collider (LHC), precise measurement of Higgs boson has become the challenge in high energy physics experiments. Many electron-positron Higgs factories with improved accuracy on the Higgs total width measurements have been proposed, including the International Linear Collider (ILC), the Circular Electron Positron Collider (CEPC), the Future Circular Collider e^+e^- (FCCee). The Higgs physics program to be carried out in the future e^+e^- colliders has been evaluated and the reachable precision on many of couplings is estimated to percent or sub-percent levels. In order to achieve this precision, the use of Particle Flow Algorithm (PFA) has become the paradigm of detector design for the high energy frontier. The key idea is to reconstruct every final state particle in the most suited sub-detectors, and reconstruct all the physics objects on top of the final state particles. The PFA oriented detectors have high efficiency in reconstructing physics objects such as leptons, jets, and missing energy.

The lepton identification is essential for this physics programs, especially for the precise measurement of the Higgs boson.

In this thesis, a PFA based lepton identification (Lepton Identification for Calorimeter with High granularity (LICH) has been developed for detectors with high granularity calorimeters. Using the conceptual detector geometry for the CEPC, featuring typical calorimeter granularity of 1000 and 400 cells / cm^3 respectively for the electromagnetic and hadronic parts, and samples of single charged particles with energy larger than 2 GeV, LICH identifies electrons or muons with efficiencies higher than 99.5% and controls the mis-identification rate of hadron to muons or electrons to better than 1% or 0.5% respectively. Reducing the calorimeter granularity by 1 or 2 orders of magnitude, the lepton identification performance is stable for particles with E > 2 GeV. Applied to fully simulated eeH or $\mu\mu$ H events at $\sqrt{s} = 250$GeV, the lepton identification performance is consistent with the single particle case: the efficiency of identifying all the high energy leptons in an event ranges between 95.5% and 98.5%.

Oppositely to muons and electrons, τ's are extremely intriguing physics objects as their Yukawa coupling to the Higgs boson is relatively large. Due to their rich decay products, properties such as the Higgs CP and EW parameters at a Z-factory can be measured. The τ-decay products have low multiplicity and in high energy colliders are tightly collimated and have low multiplicity, providing excellent signatures to probe. In this thesis, the H→ $\tau\tau$ channel is analyzed in different Z decay modes with SM background taken into account. The combined final accuracy of $\sigma \times Br(H \to \tau\tau)$ is expected to be 0.89%.

Résumé

Le Modèle Standard des interactions des particules élémentaires est la réalisation en cours des quarante dernières années d'activité expérimentale et théorique en physique des particules. Depuis la découverte du boson de Higgs en 2012 par les expériences du Grand collisionneur de hadrons (LHC), une mesure précise de Higgs boson est devenu le défi dans les expériences de physique des hautes énergies. De nombreux électrons-positons usines de Higgs avec une meilleure précision sur les mesures de largeur totale de Higgs ont été proposées, y compris le collisionneur linéaire international (ILC), la circulaire de sollicitation collisionneur électron-positon (CEPC), future collisionneur circulaire e^+e^- (FCCee). Le programme de la physique du Higgs à réaliser dans l'avenir e^+e^- collisionneurs a été évaluée et la précision accessible à un grand nombre d'accouplements est estimé à cent ou niveaux au dessous de pour-cent. Pour atteindre cette précision, l'utilisation de l'algorithme de flux de particules (PFA) est devenu le paradigme de la conception du détecteur pour la frontière de haute énergie. L'idée principale est de reconstruire chaque particule d'état final dans les sous-détecteurs les plus adaptés, et de reconstruire tous les les objets de la physique au-dessus des particules d'état final. les détecteurs orientés PFA ont une efficacité élevée dans la reconstruction des objets physiques tels que leptons, jets, et de l'énergie manquante.

L'identification des leptons est essentielle pour ce programme de physique, en particulier pour la mesure précise du boson de Higgs. L'identification du lepton est fondamentale pour les mesures de Higgs. Environ 7% des bosons de Higgs au CEPC ou au ILC sont générés avec une paire d'électrons ou de muons. Ces événements sont les signaux d'or pour l'analyse de recul de Higgs, qui est l'ancre pour les mesures absolues de Higgs. Une fraction indéfinissable du boson de Higgs se désintègre, directement ou par cascade, en états finaux avec des leptons. C'est-à-dire que 0.02 % des SM Higgs se désintègrent en muons; les leptons sont les bougies essentielles de l'identification des états finaux $H \rightarrow WW/ZZ \rightarrow$ leptoniques / semi-leptoniques. En outre, une fraction significative des événements Higgs $\rightarrow bb/cc$ génère des leptons dans leur cascade de désintégration. Une identification du lepton à haute efficacité est également très appréciée pour les mesures EW. Le système de suivi et le système calorimétrique hautement granulaire fournissent des variables discriminantes pour l'identification des particules, et la boîte à outils TMVA offre une utilisation optimale de ces variables.

Dans cette thèse, un PFA basé identification des leptons (leptons identification pour calorimètre avec une granularité élevée (LICH) a été mis au point pour les détecteurs avec calorimètre haute granularité. En utilisant la géométrie du détecteur conceptuel du CEPC, avec une granularité de calorimètre typique de 1000 et 400 cellules / cm ^{3} respectivement pour les parties électromagnétiques et hadroniques, et des échantillons de particules individuelles chargées avec une énergie supérieure à 2 GeV, LICH identifie des électrons ou muons avec des rendements supérieurs à 99,5 % et contrôle la vitesse identification erronée de hadrons à muons ou des électrons à mieux que 1 % ou 0,5

% respectivement. la réduction de la granularité du calorimètre par 1 ou 2 ordres de grandeur, la performance d'identification de lepton est stable pour des particules avec E> 2 GeV. appliquée à eeH entièrement simulé ou $\mu\mu événementsH$ à $\sqrt{s} = 250$ GeV, les performances d'identification de lepton est compatible avec le cas de particules unique: l'efficacité de l'identification de tous les leptons de haute énergie dans un événement se situe entre 95,5 % et 98,5 %.

À l'opposé de muons et électrons, les τ sont des objets de physique extrêmement intrigante que leur couplage Yukawa au boson de Higgs est relativement importante. En raison de leurs produits riches en désintégration, propriétés telles que les paramètres CP Higgs et EW à Z-usine peut être mesurée. Le $g(H\tau\tau)$ devrait être mesuré avec une précision relative supérieure à 1% au CEPC. La mesure de la polarisation τ au Z-pole conduit à une détermination précise de l'asymétrie $A_{FB}(\tau)$. La reconstruction des fonctions spectrales tau a également un potentiel convaincant au CEPC. Dans cette thèse, la reconstruction de tau couvre le canal de Higgs se désintégrant en $tau\, tau$ accompagné de leptons ou de jets. L'idée de base est de profiter de la haute granularité et de la propriété de la multiplicité. Les $tauproduits - decayde$ ont une faible multiplicité et à colliders haute énergie sont étroitement collimaté et ont une faible multiplicité, offrant d'excellentes signatures de sonde. dans ce mémoire, le H $rightarrow$ $tau\, tau$ canal est analysé en différents modes de désintégration de Z avec le fond de SM pris en compte. La précision finale combinée de $\sigma \times Br(H \to \tau\tau)$ devrait être 0,89 %.

Contents

1 Introduction **1**

2 Theory **4**
- 2.1 Standard Model . . . 5
  - 2.1.1 The Electroweak symmetry breaking . . . 6
  - 2.1.2 Higgs mechanism . . . 7
- 2.2 Beyond Standard Model . . . 13

3 e^+e^- Collider as Higgs factory **15**
- 3.1 Production processes . . . 17
- 3.2 The Circular Electron-Positron Collider (CEPC) . . . 19
  - 3.2.1 Accelerator design . . . 19
  - 3.2.2 Machine Detector Interface (MDI)[1] . . . 22
- 3.3 The International Linear Collider (ILC) . . . 23
  - 3.3.1 ILC Subsystems[2] . . . 24
- 3.4 The Future Circular Collider (FCC) and High Luminosity LHC (HL-LHC) 28

4 Detector **29**
- 4.1 Particle Flow Algorithm (PFA) oriented detector . . . 30
- 4.2 Detector design . . . 31

5 Softwares and Particle Flow Algorithm (PFA) **34**
- 5.1 Particle Flow Algorithm . . . 34
  - 5.1.1 Jet Energy Resolution . . . 34
  - 5.1.2 PandoraPFA . . . 35
  - 5.1.3 Arbor . . . 37
- 5.2 Tools . . . 42
  - 5.2.1 LCIO[3] . . . 42
  - 5.2.2 Simulation . . . 42
  - 5.2.3 Marlin Framework[4] . . . 42
- 5.3 Detector optimization . . . 43
  - 5.3.1 ECAL optimization . . . 43

5.3.2 HCAL and B field optimization 45

6 Particle identification **48**
6.1 Detector geometry and sample 48
6.2 Discriminant variables and the output likelihoods 49
6.3 Performance on single particle events 56
6.4 Lepton identification performance on single particle events for different geometries 62
6.5 Performance on physics events 66
6.6 Conclusion 68

7 Measurement of $H \to \tau\tau$ Branching Ratio **73**
7.1 Introduction 73
7.1.1 τ physics 73
7.1.2 τ decay modes 74
7.1.3 Measurements and precisions 77
7.2 Samples 77
7.3 Leptonic channels 78
7.3.1 $Z \to \mu\mu$ 79
7.3.2 $Z \to \nu\nu$ 85
7.4 Hadronic channel, $Z \to qq$ 87
7.5 Combined Results 97
7.6 Extrapolating in ILC 99
7.7 Discussion 99

8 Conclusion **101**

References **103**

Chapter 1

Introduction

The basic rules of the Universe are always attractive to the physicists.They focus on the elementary particles, the fundamental interactions, the beginning and the future of the Universe, etc. Ever since the discovery of the way to combine the electromagnetic and weak interactions by Sheldon Glashow in 1961[5], and the Higgs mechanism incorporated by Steven Weinberg and Abdus Salam[6], the Standard Model has been developed to describe the fundamental structure of matter and its interactions[7, 8, 9, 10]. With this model, all matter can be built from twelve particles of spin 1/2 and their anti-particles. The interactions between these particles can be explained by the existence of four fundamental forces mediated by spin 1 or 2 quanta. The last unverified part of this model, the Higgs boson, was successfully discovered at the Large Hadron Collider (LHC) by ATLAS and CMS experiments in 2012[11, 12], after decades of hunting, from LEP to Tevatron. Up to the most recent measurements, this Higgs boson behaves as the SM predicts. However, more precise measurements are still needed to fully validate the Higgs mechanism.

The Standard Model agrees with the experimental observations. Nevertheless, there are questions not answered by SM: why are there three generations of elementary fermions, why is the mass hierarchy so enormous, what is the nature of gravitational forces, what is the nature of dark matter and dark energy, why is there such an matter-antimatter asymmetry of the Universe.... These questions are expected to be solved in the new physics beyond the Standard Model. Even though the models proposed vary from each other, most of them predict deviations of Higgs couplings of $O \sim 1\%$[13, 14, 15].

While the LHC has huge discovery power, its final accuracy is always limited by the usage of protons as colliding particles, as the huge QCD backgrounds leads to a low signal to background ratio. On the contrary, the electrons and positrons - in the current state of knowledge - are point-like objects which interact through electroweak interactions (much weaker than the strong interactions), yielding events that are relatively free of background debris. This makes possible to treat the events as a whole and to constrain the new particle properties with the knowledge of the initial state. Two advanced proposals of e^+e^- Higgs factories are the International Linear Collider (ILC)[16, 17], the Circular Electron-Positron Collider (CEPC)[18], and the Future Circular Collider e^+e^- (FCCee)[19]. The ILC provides polarized beams and leaves the possibility to be upgraded to higher energy, while the CEPC provides higher luminosity and can be upgraded to a proton-proton collider. The FCC is a design study of CERN to extend the research after LHC reaches the end of its lifespan, and FCCee is part of it.

The high precision to be reached at e^+e^- Higgs factories imposes stringent requirements on the detector. A typical event at ILC or CEPC will feature a multi-jet final state topology. Many physics channels have to be reconstructed with unconstrained kinematics, e.g. each time neutrinos are involved. A calorimetric system is then required with resolution far beyond what has been achieved so far. An approach named Particle Flow (PF), which exploits the synergy of hardware and software developments to the level of individual particle reconstruction and identification, is believed to address these requirements[20]. It consists in reconstructing every visible particle in an event, using at best each of the sub-detectors. In turn, detectors with high efficiency and reliability, maximum hermeticity, and a highly segmented calorimeter allowing particle shower separation, are mandatory. Thus the baseline of the detectors at e^+e^- Higgs factories contains a tracking system with excellent resolution and a highly granular calorimeter system.

The lepton identification is fundamental to the Higgs measurements. About 7% of Higgs bosons at the CEPC or ILC are generated together with a pair of electrons or muons. Those events are the golden signals for the Higgs recoil analysis, which is the anchor for the absolute Higgs measurements. A unneglagable fraction of the Higgs boson decays, directly or via cascade, into final states with leptons[21]: i.e., 0.02% of SM Higgs decays into muons; the leptons are the essential candles of the identification of $H \to WW/ZZ \to$ leptonic /semi-leptonic final states. In addition, a significant fraction of Higgs $\to bb/cc$ events generate leptons in their decay cascade. A highly efficiency lepton identification is also highly appreciated for the EW measurements. The tracking system and highly granular calorimetic system provide discriminant variables for the particle identification, and the TMVA toolkit[22] offers optimal utilization of these variables.

The τ lepton[21] is an extremely intriguing physics object. As the heaviest lepton in the SM, τ has a large Yukawa coupling $g(H\tau\tau)$ to the Higgs boson, leading to a significant branching ratio $Br(H \to \tau\tau)$. The $g(H\tau\tau)$ is expected to be measured with a better than 1% relative accuracy at the CEPC. Measuring the τ polarization at the Z pole leads to a precise determination of the backward-forward asymetry $A_{FB}(\tau)$[23]. The reconstruction of the tau spectral functions also have compelling potential at the CEPC. In this thesis, the τ reconstruction is covering the channel of Higgs decaying to $\tau\tau$ accompanied with leptons or jets. The basic idea is to take advantage of the high granularity and the property of multiplicity.

The content of this paper is organized as follows. A brief overview of the Standard Model is described in Chapter 2. In Chapter 3, we describe the CEPC and ILC as e^+e^- colliders. The PFA oriented detectors will be introduced in Chapter 4, followed by the presentation of two PFAs and their application in detector optimization in Chapter 5. The description of the particle identification package and its performance on single particles as well as in fully simulated events are discussed in Chapter 6. In Chapter 7, we will discuss the signal strength of the Higgs boson decaying into tau lepton pairs at the CEPC, taking into account all the SM backgrounds.

Chapter 2

Theory

The Standard Model of elementary particle interactions is the outstanding achievement of the past forty years of experimental and theoretical activity in particle physics. In one word, the Standard Model is a field-theory description of strong and electroweak interactions at the energy of several hundred GeV. So far it is a theoretical structure which has worked splendidly. In the Standard Model, the fundamental fermionic constituents of matter are quarks and leptons[24]. Both of them have spin $\frac{1}{2}$ and are point-like at the smallest distances currently probed by the highest-energy accelerators. There are three generations of these particles, namely: (a)(u, d) and (ν_e, e), (b)(c, s) and (ν_μ, μ), (c)(t, b) and (ν_τ, τ). We have a relatively simple picture of quarks and leptons with their interactions (gravitation excepted). These interactions are mediated by spin 1 particles following the Bose-Einstein statistics[**?**]. They are referred as "bosons". Gluons correspond to strong interactions, W and Z to the weak interactions and gamma to electromagnetic. The weak interactions involve pairs of quarks and leptons, these are sources for the $W^\pm$ and Z^0 fields. Charged particles are sources for the photon field, which is the medium of electromagnetic interaction. The theory is to describe the forces between fermions by the exchange of these bosons[25]. The elementary particles and there interactions are shown in Figure 2.1.

In modern physics, symmetry almost is one of the highest principles of the new laws of physics for a physicist to explore.

When physicists want to invent a new mechanism (for example, construct a Lagrangian quantity) to explain some new phenomenon, this mechanism has to meet certain symmetry and to adjust within this framework to try to find the necessary mechanisms. According to Noether's theorem, any differentiable symmetry of the action of a physical system has a corresponding conservation law. We know that the action of a classical physical system is the integral over time of a Lagrangian function, and it is invariant due to conservation laws. In this chapter, we will develop this subject for relativistic field

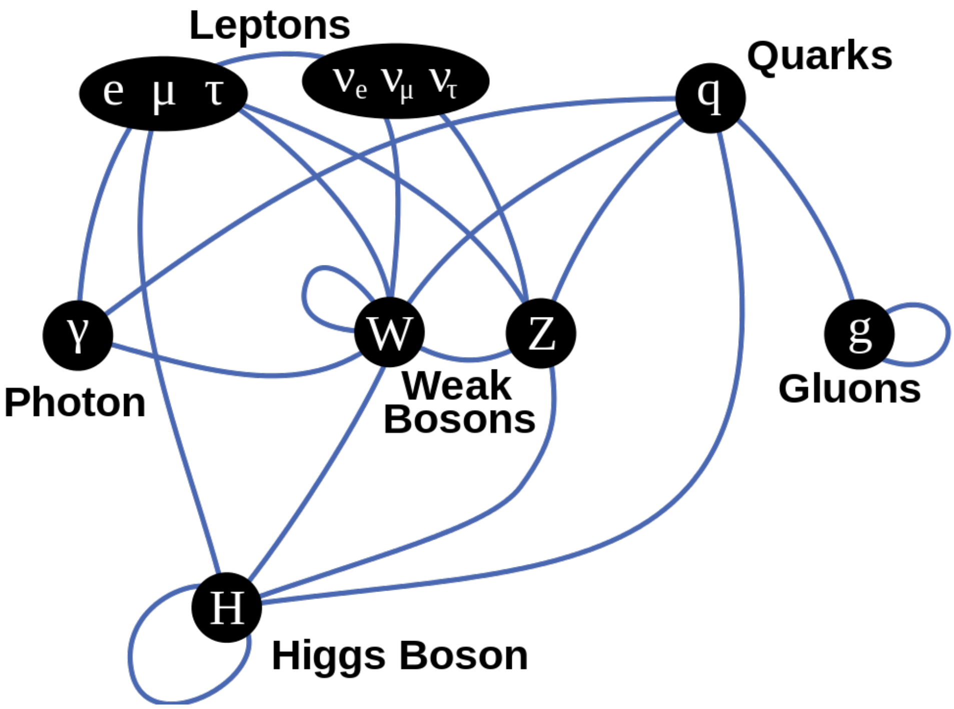

Figure 2.1: The Standard Model particles and the interaction in between them

theories. If the transformation is identically performed at every point in space-time, we call it a global symmetry. In gauge theory, it is required that the system is invariant under a local symmetry, which means that the transformation is labeled by a spacetime-dependent phase so that the transformation can be applied in a local area without influencing other areas. These transformations are called gauge transformations. For each set of interaction mediating boson, the Lagrange function in gauge transformations, therefore these bosons are called gauge bosons. In fact, the gauge transformation is an element of a unitary group called gauge group[26]. For strong interaction, the gauge group is SU(3), and it is SU(2) $\times$ U(1) for electroweak interactions. In group theoretical language, the Standard Model is encoded in the symmetry group SU(3) $\times$ SU(2) $\times$ U(1).

2.1 Standard Model

The Standard Model Lagrangian is written in three parts, the kinematic terms, the coupling terms, and mass terms, it can be written in a simplified formula as:

$$\begin{aligned}\mathcal{L} &= -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &+i\bar{\psi}\not{D}\psi + h.c. \\ &+\psi_i y_{ij}\psi_j\phi + h.c. \\ &+\left|D_\mu\phi\right|^2 - V(\phi)\end{aligned} \tag{2.1.1}$$

the terms in this formula are: the scalar product of the field strength tensor $F_{\mu\nu}$ containing the mathematical encoding of all interaction particles except the Higgs boson, the term describing how interaction particles interact with matter particles, the term describing how matter particles couple to the Brout–Englert–Higgs field ϕ and obtaining mass, how the interaction particles couple to the BEH field, and the potential of the BEH field.

2.1.1 The Electroweak symmetry breaking

The Lagrangian of a classical theory subjected to a non-zero vacuum expectation value describes a system with n real scalar fields $\phi_i(x)$ (vectors) by[25]:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\phi^i\right)^2 + \frac{1}{2}\mu^2(\phi^i)^2 - \frac{\lambda}{4}(\phi^i)^4 \tag{2.1.2}$$

where a sum over all $i = 1, ..., n$ is conducted in each term. μ^2 corresponds to the ordinary mass term m^2 with a changed sign. We identify the first term as the kinetic energy of the system and deduce that the rest is the potential $V(\phi^i)$. The altered sign of the mass $m^2 \to \mu^2$ term will allow for a potential with negative minima, which will be crucial in our depict of symmetry breaking. This is an example of a self-interacting theory where λ is a dimensionless coupling constant describing the strength of the interaction (a more basic example than the QED Lagrangian which also encodes a self-interacting theory). By setting an even power of the fields, we will be able to obtain positive definite energies (and scalar field theories with fields of an even higher order than 4 will not be renormalizable). The lowest energy value of $\mathcal{L}$ is obtained when we are dealing with a uniform constant field $\phi(x) = \phi_0^i$. It is chosen as the field which minimizes the potential term in $\mathcal{L}$, i.e.:

$$V(\phi^i) = -\frac{1}{2}\mu^2(\phi^i)^2 + \frac{\lambda}{4}(\phi^i)^4 \tag{2.1.3}$$

This optimization problem is straightforward to solve and we find

$$(\phi^i)^2 = \frac{\mu^2}{\lambda} \quad (2.1.4)$$

However, this equation only defines the length of the vector ϕ_0^i leaving its direction arbitrary. In two dimensions, this can be investigated visually as the two fields are then constrained by

$$\phi_1^2 + \phi_2^2 = \frac{\mu^2}{\lambda} \quad (2.1.5)$$

which corresponds to a circle. Drawing the potential as in Figure 2.2, we discern that the minima will be found on this circle and not where $\phi_1^2+\phi_2^2 = 0$. The system, therefore, has an infinite number of possible solutions that obey this minima condition as any point on the circle will do. Moreover, the system may choose one of these spontaneously and in doing so, its $O(2)$-symmetry is hidden from our experimental surveys since we cannot perceive the other solutions not chosen. The symmetry is broken spontaneously by the choice of one of the solutions.

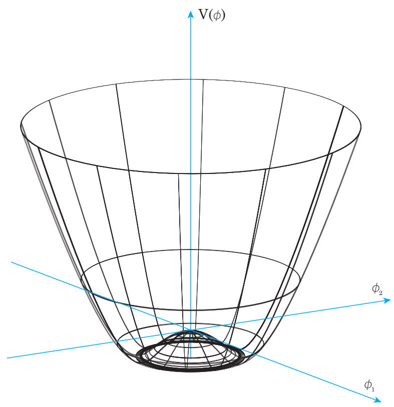

Figure 2.2: A visualization of the potential V in the case where n = 2. Notice that the minima where V has negative values are found on the circle defined by 2.1.5 which physically correspond to a set of degenerate vacua.

2.1.2 Higgs mechanism

Introducing a complex scalar field ϕ will satisfy Lorentz invariance as well as rotational invariance, due to its scalar nature. This field might yield a non-zero expectation value of the vacuum as we have seen in the calculations above and let us construct a gauge invariant Lagrangian which gauge bosons will acquire mass. A usual choice is to call this complex scalar field ϕ and to write it as:

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{2.1.6}$$

where ϕ_1, ϕ_2 are real fields. This field is known as the Higgs field. Moreover, by combining two of these in a doublet, we transform them in a SU(2) spinor, i.e. we are in a model with a spinorial representation of SU(2). Let us use the rotational freedom of SU(2) to compute the vacuum expectation value of this field as

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix} \text{with } v = \sqrt{\frac{\mu^2}{\lambda}} \tag{2.1.7}$$

Once again we are on the circle and our symmetry is broken.

Boson mass

To see how this affects the Lagrangian of the system, we have to investigate how its kinetic term, involving the covariant derivative arising from the gauge symmetry, couples to this new field. We will find some terms which we are able to recognize as "mass terms" as we did previously for the linear sigma model. This can be done if we insert the covariant derivative of SU(2)

$$D_\mu\phi = \left(\partial_\mu + igA^a_\mu\tau^a\right)\phi \tag{2.1.8}$$

where the index $a = 1, 2, 3$ runs over all of the generators $i\tau^a = i\sigma^a/2$ of SU(2) in its two dimensional representation (τ^a Hermitian matrices). In the part of the Lagrangian corresponding to a kinetic term and let it couple to the field, and the charge of the Lagrangian $\Delta\mathcal{L}$ be written as

$$\Delta\mathcal{L} = \frac{g^2v^2}{8}A_\mu A^\mu \tag{2.1.9}$$

with the mass coefficient for the three gauge bosons as

$$m_A = \frac{gv}{2} \tag{2.1.10}$$

That particle can obtain mass by interacting with a field of this kind is known as the Higgs mechanism or with more names occasionally the Englert–Brout–Higgs mechanism[27, 28].

Given that the weak interactions are to be mediated by our gauge vector bosons, we thus required three vector mesons $W_\mu^a (a = 1,2,3)$, at this stage all massless. The simplest group that contains the required three generators is SU (2). However, it is clear that this is not enough if we wish to include electromagnetic interaction as well. Given that the W_μ^acouple in a parity-violating fashion only to the left-handed parts of the leptons, as required for the weak interactions, whereas the electromagnetic interaction conserves parity and involves both left and right parts of the leptons. Thus we need one further gauge vector meson, B_μ, and correspondingly a group with one generator, U (1). The overall gauge group is then U (1)×SU (2)$_L$ with a total of four generators. The subscript L on SU (2)$_L$ indicates that among fermions, only left-handed states transform nontrivially under weak isospin. For the electroweak force, fermions live in representations of the hypercharge U(1) and weak isospin SU(2) which are tensored together. Its mediating particles, the $W_\pm$-bosons, Z_0-boson and the photon span the complexified adjoint representation.

Since we desire to end up with three heavy vector bosons associated with the weak interactions and a massless vector boson, the photon, we require 4 independent scalar fields. The simplest choice is a doublet of complex scalar fields, one charged, one neutral:

$$\psi = \begin{pmatrix} \psi^+ \\ \psi^0 \end{pmatrix} \tag{2.1.11}$$

The 2×2 matrices representing the generators of U (1) and SU (2) are just the unit matrix I and the Pauli matrices divided by two, and the Lagrangian should be:

$$\mathcal{L} = (D^\mu \psi)^* (D_\mu \psi) - V(\psi) \tag{2.1.12}$$

where the potential V is to produce spontaneous symmetry breaking, and D_μ should have the form:

$$D_\mu = \partial_\mu + \frac{i}{2} g_1 W_\mu^a \tau^a + \frac{i}{2} g_2 I B_\mu \tag{2.1.13}$$

Generally, we put

$$\begin{cases} B_\mu &= \cos\omega_W A_\mu + \sin\omega_W Z_\mu \\ W_\mu^3 &= \sin\omega_W Z_\mu - \cos\omega_W A_\mu \end{cases} \tag{2.1.14}$$

where ω_W is called Weinberg angle and we shall adjust its value so that A_μ turns out to be the photon field and Z_μ will be then the massive neutral boson. The term concerning W_μ^3 and B_μ in 2.1.12 will become:

$$\begin{aligned} & \tfrac{i}{2}\left(g_1 W_\mu^3 \tau^3 + g_2 I B_\mu\right)\psi \\ = \quad & \tfrac{i}{2}[A_\mu\left(g_1\tau^3 \sin\omega_W + g_2 I \cos\omega_W\right) \\ & Z_\mu\left(g_2 I \sin\omega_W - g_1 \tau^3 \cos\omega_W\right)] \end{aligned} \tag{2.1.15}$$

The photon field A_μ couples through the unbroken generator with the charge e, thus:

$$e = g_1 \sin\omega_W = g_2 \cos\omega_W \tag{2.1.16}$$

Introducing the charged field $W_\mu^\pm$ as

$$W_\mu^\pm = \sqrt{\frac{1}{2}}\left(W_\mu^1 \mp i W_\mu^2\right) \tag{2.1.17}$$

corresponding to the gauge bosons $W^\pm$, and by using the vacuum Higgs configuration in 2.1.7, the Lagrangian in 2.1.12 becomes

$$\mathcal{L} = \frac{1}{8} g_1^2 v^2 \left[2 W_\mu^+ W^{-\mu} + \frac{Z^2}{\cos_W}\right] + \frac{1}{2}\partial_\mu v \cdot \partial^\mu v \tag{2.1.18}$$

A_μ does not appear in this equation, which means that a massless electromagnetic field exists as required. The charged W-boson masses can be read off directly as

$$M_{W^\pm} = \frac{1}{2} g_2 v \tag{2.1.19}$$

because the term proportional to the bosons corresponds to charged intermediate boson masses. And we can define the mass of neutral gauge boson Z by using the relations in 2.1.16, we get

$$M_Z = \frac{M_W}{\cos\omega_W} = \frac{v}{2}\sqrt{g_1^2 + g_1^2} \tag{2.1.20}$$

Thus the masses for the W bosons and Z^0-boson have been found, and of course a similar method can be used for a larger and more complicated group SU(3)× SU(2) × U(1), which leads to the construction of the complete Standard Model.

Fermion mass

The fermion term of the Lagrangian is:

$$\mathcal{L} = -m\bar{\psi}\psi = -m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \tag{2.1.21}$$

However this lagrangian is not gauge invariant since the left handed fermions form an isospin doublet and the right handed fermions form isospin singlets. In order to construct an SU(2)L × U(1)Y invariant term for fermions, we used the complex doublet introduced in the previous section, which gives:

$$\mathcal{L} = -\lambda_f(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi\psi_L) \tag{2.1.22}$$

where λ_f is the so-called Yukawa coupling between the fermions and the scalar field.

For all generations of quarks and leptons, the complete Lagrangian for the Yukawa interaction with the Higgs field can be expressed as:

$$\begin{aligned} L &= Y^d_{ij}\bar{Q}^i_L\phi d^j_R + Y^u_{ij}\bar{Q}^i_L\tilde{\phi}u^j_R + Y^l_{ij}\bar{L}^i_L\phi l^j_R + h.c. \\ &= \frac{1}{\sqrt{2}}Y^d_{ij}\left(\bar{u}^i_L,\bar{d}^i_L\right)\begin{pmatrix}0\\v\end{pmatrix}d^j_R + \frac{1}{\sqrt{2}}Y^u_{ij}\left(\bar{u}^i_L,\bar{d}^i_L\right)\begin{pmatrix}v\\0\end{pmatrix}u^j_R + \frac{1}{\sqrt{2}}Y^l_{ij}\left(\bar{\nu}^i,\bar{e}^i\right)\begin{pmatrix}0\\v\end{pmatrix}e^j_R + h.c. \\ &= \frac{Y^d_{ij}\cdot v}{\sqrt{2}}\bar{d}^i_L d^j_R + \frac{Y^u_{ij}\cdot v}{\sqrt{2}}\bar{u}^i_L u^j_R + \frac{y_{ii}\cdot v}{\sqrt{2}}\bar{e}^i e^i \end{aligned} \tag{2.1.23}$$

where i and j run over all generations. Thus the mass of quarks matrix is introduced as: $\frac{Y^{u,d}_{ij}\cdot v}{\sqrt{2}}$ and lepton mass of each generation as: $\frac{y_l\cdot v}{\sqrt{2}}\bar{l}l$ (The Yukawa matrix for lepton is diagonal and the neutrino are massless in this model).

Higgs coupling

It is interesting to study details of the Higgs boson properties like its coupling to fermions and gauge bosons as that determines if and how the Higgs boson is produced in experiments and what the event topology will be.

If we parameterize the scalar field ϕ in 2.1.6 to be:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+h \end{pmatrix} \tag{2.1.24}$$

where v is the vacuum expectation of ϕ and h is a fluctuating real valued field with $\langle h \rangle = 0$.

Rewriting the Lagrangian in the unitary gauge, the protential energy term takes the form:

$$\mathcal{L}_V = -\mu^2 h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 \tag{2.1.25}$$

The field h is thus a scalar particle with mass $m_h = \sqrt{2}\mu^2 = \sqrt{\frac{\lambda}{2}}v$. This particle is known as Higgs boson.

Rewriting the Lagrangian in 2.1.12, the kinematic energy term yields the gauge boson mass term plus additional terms involving the Higgs boson field:

$$\mathcal{L}_{boson} = \frac{1}{2}\partial_\mu h \cdot \partial^\mu h + \left[M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}M_Z^2 Z_\mu Z^\mu\right] \cdot \left(1 + \frac{h}{v}\right)^2 \tag{2.1.26}$$

where the M_W and M_Z are given in the previous section.

Rewriting the Lagrangian that couples the Higgs doublet to the fermion fields, these terms in unitarity gauge can be evaluated:

$$\mathcal{L}_{fermion} = -m_f \bar{f} f \left(1 + \frac{h}{v}\right) \tag{2.1.27}$$

From 2.1.25, 2.1.26 and 2.1.27, the coupling of the Higgs boson to other particles of the weak interaction theory are proportional to the masses of those particles.

2.2 Beyond Standard Model

Even though the Standard Model explains some of the matters in particle physics, it is not a truly fundamental theory. There are quite some open problems left: the reason for three generations of elementary fermions, the scheme of grand unification, the hierarchy problem[29], the nature of gravitational forces, dark matter and dark energy, the matter over anti-matter dominance in the Universe, etc.

The only naturally defined mass of the SM is the Planck Mass $M_{Pl} = 2.4 \times 10^{18} GeV/c^2$, sitting 16 orders of magnitude above the ElectroWeak mass scales. The radiative corrections to the Higgs being quadratic in energy and masses, the tuning of the SM parameters requires an unrealistic precision over such a large scale gap. This is the hierarchy problem. It is partly solved by the Grand Unification which sets a unification of forces at $\sim 10^{15} GeV/c^2$, but for which a scheme compatible with observations has to be defined, or by SuperSymmetry which cancels out the corrections above a scale which could be not too far from the EW one. On cosmological grounds, the evolution of the Universe metric suggests a content of the universe made of 68% of Dark Energy and 27% of Dark Matter (for 5% of standard matter) of unknown nature, no corresponding particle having been observed (hence the "dark" quality). Some ideas such as SUSY, extra dimensions, or Minimal Dark Matter are proposed to describe this issue. What is observed is the complete predominance of matter over anti-matter, whereas initial conditions of the Big-Bang predicts symmetry. No symmetry breaking mechanism has proven strong enough in the SM to explain this fact. That is why the theorists proposed the mechanism in SUSY, extended Higgs sector, etc. The Einstein theory of gravity, which has proven correct in all tests (the latest being the existence of gravitational waves) is not yet compatible with quantum theory. New theories of gravity exists but are still far beyond experimental scope.

In order to explain these problems, plenty of models are proposed by theoretical physicists and to be tested at the future e+e- colliders.

Grand Unification The basic hypothesis of grand unification states that SU(3)$\times$ SU(2) $\times$ U(1) is the remnant of a larger, simple or semi-simple group G, whose symmetry is lost at currently reachable energies. Several groups have been used for grand unification, including SU(5), SO(10), E_6 or E_8[30, 31, 32].

Supersymmetry[33] Supersymmetry (SUSY) is a symmetry relating particles of integer spin, i.e. spin–0 and spin-1 bosons, and particles of spin $\frac{1}{2}$, i.e. fermions. The basic idea of SUSY is that the generators transform fermions into bosons and vice-versa. When the symmetry is exact, the bosonic fields, i.e. the scalar and gauge fields of spin 0 and spin 1, respectively, and the fermionic fields of spin $\frac{1}{2}$ have the same masses and quantum numbers, except for the spin. The particles are combined into super fields and the simplest case is the chiral or scalar super field which contains a complex scalar field with two degrees of freedom and a Weyl fermionic field with two components.

In the breaking of Supersymmetry, we obviously need to preserve the gauge invariance and the renormalizability of the theory and, also, the fact that there are still no quadratic divergences in the Higgs boson mass squared. Since up to now there is no completely satisfactory dynamical way to break SUSY, a possibility is to introduce by hand terms that break SUSY explicitly and parametrize our ignorance of the fundamental SUSY–breaking mechanism. This gives a low energy effective SUSY theory, the most economic version being the Minimal Supersymmetric Standard Model (MSSM), providing candidate dark matter particles. In a supersymmetric theory, Planck-scale quantum corrections cancel between partners and superpartners (owing to a minus sign associated with fermionic loops). Thus the hierarchy between the electroweak scale and the Planck scale is achieved in a natural manner. Besides, the running of the gauge couplings are modified, and precise high-energy unification of the gauge couplings is achieved.

The precise measurement of the Higgs boson is a key to verify these proposed models. If there is new physics beyond the Standard Model, the coupling deviates from the Standard Model prediction. The deviation depends on the new physics beyond the Standard Model but is estimated to be $O(\sim 1\%)$ in many models[34, 35, 36]. Therefore, a precision of a few percent or less is required to shed light on a signal of new physics concealed in the coupling constants, which can be achieved with the next generation of colliders.

Chapter 3

e^+e^- Collider as Higgs factory

After the discovery of the Higgs boson, the precise measurement of its properties has become the challenge in high energy physics experiments. Several projects as the next generation of LHC are proposed for this purpose. The ATLAS and CMS experiments at the LHC will continue to improve the measurement of the Higgs boson properties including couplings to gauge bosons and Yukawa couplings. It will integrate into a High Luminosity LHC with an integrated luminosity to 3000 fb^{-1}[37], however, the accuracy of HL-LHC will be at the levels of a few percent achievable for some of the couplings, which does not meet the requirement needed to explore new physics regime.

In the Large Hadron Collider (LHC), the proton-proton collisions result in many fragmented pieces of what was originally a proton, each fragment producing its own shower of particles or jets. On the contrary, the e^+e^- are point-like particles which interact through forces much weaker than the strong interactions at LHC, so that the annihilations produce events that are relatively free of background debris. This makes it possible to analyze the events as a whole and to use all of the details to constrain the particle properties. In LHC, the huge QCD backgrounds leads to a low signal to background ratio. The total signal produced is estimated to 10^8 events in HL-LHC, the efficiency for the signal is to the order of 10^{-3}, while this efficiency for e^+e^- collider is of order 1. Another strong advantage of the e^+e^- collider is that the Higgs can be detected through the recoil mass method by reconstructing the Z boson decay only, without examining the Higgs decays. This method establishes the denominator for an absolute measurement of branching fractions, and will consequently allow the incorporation of the LHC results to obtain the best world averages. The recoil mass method also provides the best probe into the Higgs invisible decays and search for dark matter and exotic particles produced in the Higgs decays. The experimental conditions will be much cleaner, allowing the reconstruction of detectors with unprecedented precision in energy and momentum measurement. For example, as compared to the detectors designed for LHC events, the ILC detectors will have only one-tenth of the amount of material in front of the calorimeters that measure photon energies.

In conclusion, the e^+e^- collider is an appreciated collider for precision measurements with high sensitivity to effects of new physics.

Various proposals are claimed to be the e^+e^- Higgs factory, including linear and circular. The International Linear Collider (ILC) is a flagship program of linear ones, based on Superconducting RF technology. While the Circular Electron Positron Collider(CEPC) and Future Circular Collider of e^+e^- (FCCee) are two of the proposals for the circular ones. These two kinds of collider have the examples in the previous century, the two e^+e^- Z-factories, the circular LEP and the linear SLC. Both of them were successfully designed, constructed and operated, and both achieved important physics results.

The main difficulty for the linear collider comes from the high cost of the project. Recently the Japan HEP community proposed to build a 250 GeV center of mass linear collider in Japan as the first stage of the ILC serving as a Higgs factory[38]. The advantage of ILC is that the beams are polarised, and there is potential for an energy upgrade.

For circular collider, the technology is kind of mature, since all circular e^+e^- colliders are similar except for the sizes, and there are several which have been successfully constructed sharing a number of common features. The challenge for CEPC is that, due to high beam intensity and small beam size, the beamstrahlung (synchrotron radiation of individual particles in the opposing beam's field) will limit the beam lifetime. High synchrotron radiation power is another major challenge. The main advantage of a circular e^+e^- collider of sufficiently large size is to offer a higher luminosity than a linear one at 240 GeV and below. Also, a circular collider can accommodate more than one interaction point. Even though the energy is limited by synchrotron radiation and thus has no potential for an energy upgrade, a circular e^+e^- collider could be converted to a pp collider in the future as the next energy frontier, which is a plan for CEPC to SPPC. Another disadvantage is that there is no polarization in CEPC.

Plenty of issues have been studied for ILC, CEPC and FCCee. For ILC, the Technical Design Report[39] was published in 2013 and recently the project for 250GeV[38] has been reported and waits for an action from the Japanese government. According to the timeline of ILC, once there is a positive decision, there will be 4 to 6 years of preparation and about 9 years of construction and 20 years of operation. For CEPC, the Preliminary Conceptual Design Report (PreCDR)[40] was published by the end of 2014 and the CDR is under preparation and supposed to come out in the beginning of 2018. The R&D, as well as the Engineering Design, is ongoing until 2022, and the construction is estimated to be finished by the end of 2030, that means CEPC data-taking will start before the LHC program ends around 2035. After the operation of ten years, the CEPC will be upgraded to SPPC, if needed. For FCCee, or TLEP, the studies are set up since 2014, and is part and parcel of the FCC design study.

In this chapter, the ILC and the CEPC will be introduced in detail, including the physics of these colliders and their technologies.

3.1 Production processes

As shown in Figure 3.1, the leading production processes for the SM Higgs boson at e^+e^- collider operating at 250 GeV are: a) $e^+e^- \rightarrow ZH$ (Higgsstrahlung or ZH), b) $e^+e^- \rightarrow \nu\nu H$ (WW fusion), c) $e^+e^- \rightarrow e^+e^-H$ (ZZ fusion), as shown in Figure 3.2, and the estimated statistics for CEPC ($5ab^{-1}$) and ILC ($1ab^{-1}$) are shown in Table 3.1 and Table 3.2, the polarization for ILC 250GeV is either $P(e^+, e^-) = (+30\%, -80\%)$ or $P(e^+, e^-) = (-30\%, +80\%)$.

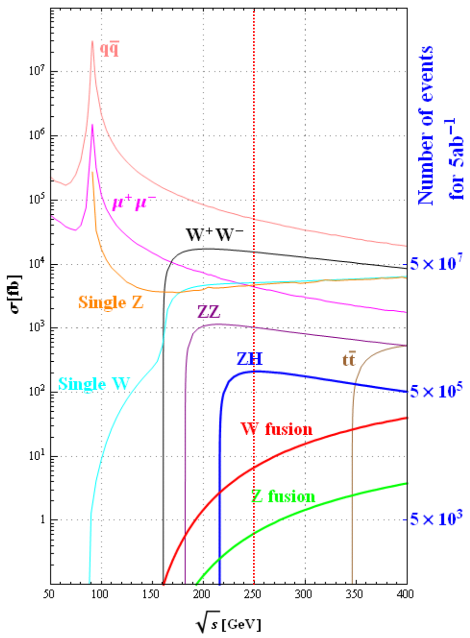

Figure 3.1: Unpolarized cross sections of main standard model processes of $e + e-$ collisions as functions of center- of-mass energy (from 50GeV to 400GeV), the dotted line indicates 250GeV

At the energy of 250 GeV, near the peak of the cross section for $e^+e^- \rightarrow ZH$, the Z boson recoil can tags the Higgs boson events. At higher energy, the WW fusion process of Higgs production, $e^+e^- \rightarrow \nu\nu$H, turns on. Measurement of this process at the full ILC energy of 500 GeV gives a model-independent precision measurement of the total Higgs boson width. Experiments at 350 GeV and 500 GeV also allow first measurements of the Higgs boson coupling to the top quark and of the Higgs boson self-coupling with

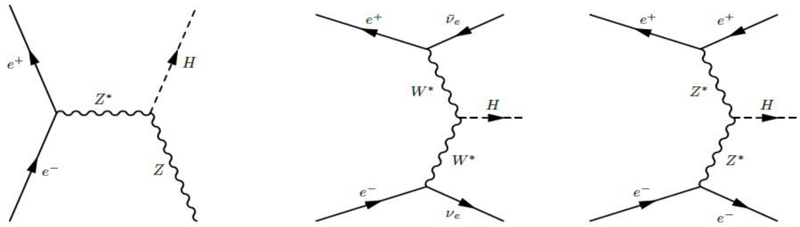

Figure 3.2: Feynman diagrams of the Higgs production processes in e^+e^- collider

Table 3.1: The cross section (fb^{-1}) of various SM processes for CEPC and ILC. eL.pR represents electron left polarized and positron right polarized, eR.pL represents right polarized and positron left polarized

Process	CEPC	ILC (eL.pR)	ILC (eR.pL)
qq	50216	129148	71272
ll	4404	21226	16470
Single Z	4733	2192	1506
Single W	5144	13335	114
Bhabha	25060	25286	24228
WW	15483	35219	323
ZZ	1033	2982	1418
ffH	219	515	319

Table 3.2: The cross section (fb^{-1}) of Higgs signal for CEPC and ILC. eL.pR represents electron left polarized and positron right polarized, eR.pL represents right polarized and positron left polarized

Process	CEPC	ILC (eL.pR)	ILC (eR.pL)
eeH	7.60	17.60	11.16
$\mu\mu H$	7.10	17.14	10.98
$\nu\nu H$	48.96	128.64	65.10
qqH	143.39	173.01	110.98

the $t\bar{t}$ events. The measurement of the forward-backward asymmetry is a probe to new physics.

3.2 The Circular Electron-Positron Collider (CEPC)

The CEPC is a circular electron-positron collider in a tunnel with a circumference of 100 km and is envisioned to operate with a center-of-mass energy of 250 GeV where the Higgs events are produced primarily through the interaction e^+e^-. With a nominal luminosity of $2 \times 10^{34} cm^{-2}s^{-1}$ about 1 million clean Higgs events will be produced by CEPC over a period of 10 years. The large statistics of this Higgs sample will enable CEPC to measure the Higgs boson production cross sections and most of its properties with precisions far beyond what is achievable at the LHC. The CEPC can also serve as a high luminosity ($10^{35\sim36} cm^{-2}s^{-1}$) Z factory at a centre of mass energy of 91 GeV, i.e. $10^{10\sim11}$ Z boson in one year.

The beam current at CEPC, determined by the synchrotron radiation budget, is 100 MW for two beams. The preliminary layout of 50km tunnel CEPC (2014) is shown in Figure 3.3, the CEPC collider is designed with four interaction points, where IP1 and IP3 are for e+e- collisions, while the other two IP's are reserved for the future pp collider, SPPC. The progressed collider[41] circumference is 100 km, including 8 arcs of 5852.8 m, 4 arc straight sections of 849.6 m each and 4 interaction region straights of 1132.8 m each.

3.2.1 Accelerator design

The CECP design aims to be a Higgs factory producing 10^6 Higgs operating at 250 GeV center of mass energy and a $W\&Z$ factory producing 10^{10} Z^0 operating at 90 GeV or 160 GeV center of mass energy. It should also leave the opportunity to be upgraded to a 100TeV proton-proton collider.

The CEPC contains several subsystems[40]:

- **Injector** In this part, 10GeV electrons /positrons will be produced and sent to the Booster. A strong focusing lattice consisting of several tens of quadrupoles maintains the transverse beam size. A pair of x-y correction dipoles and a stripline beam position monitor are associated with each quadrupole for trajectory correction. High resolution profile monitors are located along the Linac. Monitors for the energy, energy spectrum, and emittance growth are placed near the end of the Linac to allow either automatic or operator controlled correction during operations.

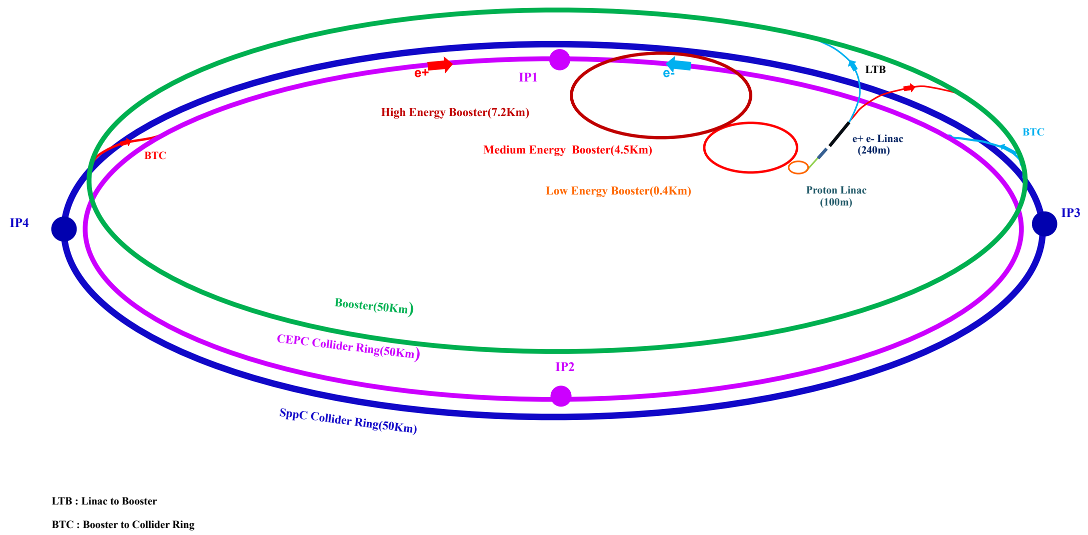

Figure 3.3: CEPC preCDR Layout

- **Electron Source** The CEPC electron source is a thermionic gridded cathode driven by high voltage pulser for the baseline design. After leaving one of these guns, the bunches pass through a Y bend and into two sub-harmonic bunching cavities. Two operation modes are required: one is to provide a 3.2 nC bunch charge for electron injection, and the other is to provide an 11 nC bunch charge as the primary electron beam for positron production. The electron beams are accelerated to 200MeV before going into the same accelerating section as positrons.

- **Positron Source** In CEPC, positrons are generated using a 4 GeV electron beam impinging on a high-Z, high density tungsten target. The positron yield per incident electron is approximately proportional to the electron energy so that the positron current is proportional to the incident power of electron beam. The large transverse emittance of the positron beam emerging from the target is transformed to match the capture section aperture with a pseudo-adiabatically changing solenoidal field. Three constant-gradient accelerator sections will boost the captured positrons to 200 MeV. The positrons are then transported back to the beginning of Linac through a quadrupole lattice and reinjected into the Linac where they are accelerated to 10 GeV.

- **Damping Rings** The primary purpose of the damping ring (DR) is to reduce the transverse phase space of the positron beam to a suitably small value at the beginning of the linac and also to adjust the time structure of the positron beam for reinjection into the Linac. a bunch compressor system is added after the damping ring to reduce the bunch length in the ring, thus to minimize wake field effects in the Linac.

- **Accelerating section** In CEPC, the klystrons and their associated modulators are the keys to acceleration. A first acceleration section containing 11 klystrons of 18 MeV/m is providing 1.1GeV electrons and positrons before the positrons are sent to the Damping Rings. Then the second acceleration section containing 20 klystrons of 27 MeV/m accelerate the beams to 4GeV, where the electron beam is used to produce the positron beams. Finally, the beams are accelerated to 10GeV through a third section containing 42 klystrons of 27 MeV/m. The procedure for acceleration in the Linac is shown in Figure 3.4

• **Booster** After being accelerated to 10GeV, electron and positron beams are injected from the Linac through the LTB transfer line (Linac to Booster) into the Booster. In CEPC, the Booster is in the same tunnel as the collider, placed 2m above the collider ring and has about same circumference (10km). Bypasses are arranged to avoid the detectors at IPs. Because of the very low synchrotron radiation damping rate, a scheme of single bunch injection from Linac to Booster is adopted. The two

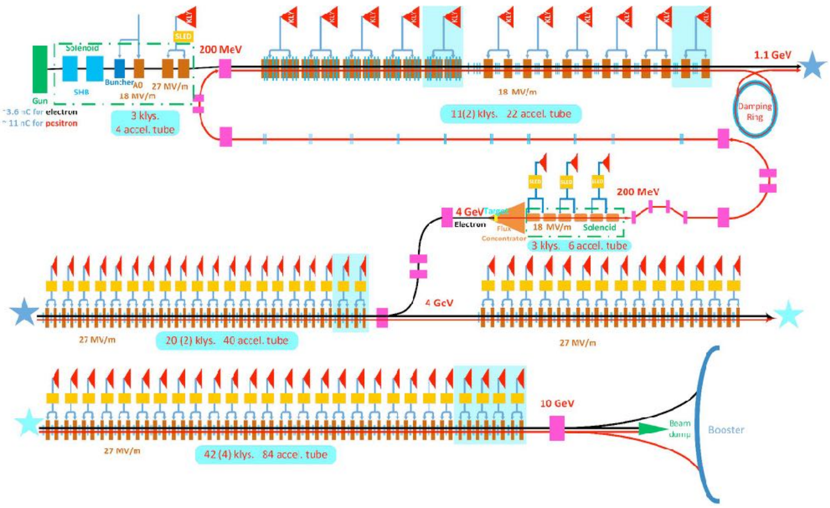

Figure 3.4: Schematic layout of the CEPC Linac, the stars represent the continuous of figure[42]

radiofrequency cavities (RFs) regions of 84 cavities each, with the cavity frequency of 1.3GHz, is ramping the energy of electron and positron beams to 45GeV(Z factory) or 120GeV (Higgs factory). Then the beams are extracted from the Booster through BTC transfer line (Booster to Collider Ring) into the Main Ring.

- **Main Ring** The Main Ring is a double ring system and is in the same channel with the Booster[43]. Two stations of radiofrequency cavities (RFs) are shared by these two rings for Higgs production, with a cavity frequency of 650MHz. Twin-aperture dipoles and quadrupoles are adopted in the arc region to reduce the power. The distance between two beams is 0.35m. For W/Z production, only half the number of cavities will be used and bunches can be filled in full ring, to lower the impedance. The layout of the double ring accompanying with the Booster is shown in Figure 3.5.

3.2.2 Machine Detector Interface (MDI)[1]

MDI plays a very important role on the way to achieve the physics goals at the electron positron collider. The MDI for CEPC is about $\pm$7m long from the Interaction Points. The interaction region of the CEPC partial double ring consists of two beam pipes, and the positron and electron beams collide with a 33 mrad crossing angle and the final focusing length is 2.2m. The accelerator components inside the detector without shielding

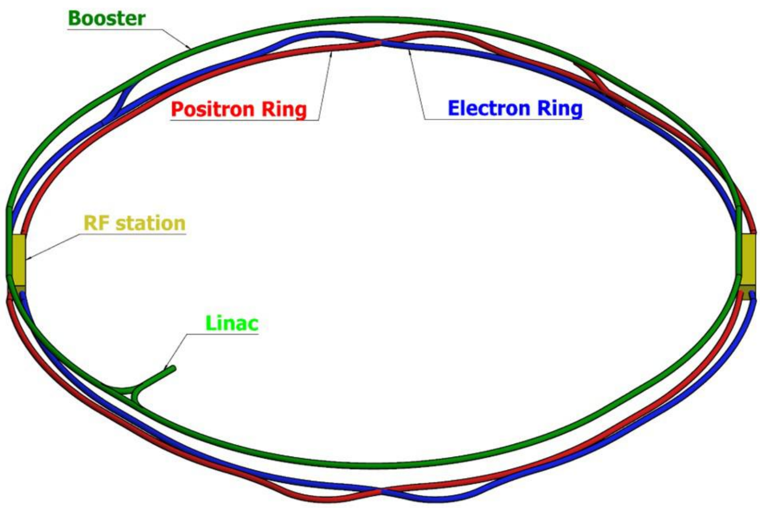

Figure 3.5: The CEPC Booster and Double Ring Layout

are within a conical space with an opening angle of $\cos\theta = 0.993$. There are two high gradient quadrupole magnets (QD0 for horizontal and QF1 for vertical) in the interaction region, inside the detector solenoid magnet which has a field of about 3.0 T. The distance from IP to the last quadrupole (QD0) is 2.2m, which is much smaller than for the ILC. To minimize the effect of the longitudinal detector solenoid field on the accelerator beam, anti-solenoid coils are used. Their magnetic field direction is opposite to the detector solenoid field, and the strength is 7.0 T to make the combined total integral longitudinal field generated by the detector solenoid and anti-solenoid coils are nearly zero. A Luminosity Calorimeter (Lumical) will be installed on the outgoing beam at a distance of 0.95 $\sim$1.11 m, with an inner radius 28.5 mm and outer radius 100 mm.

3.3 The International Linear Collider (ILC)

The ILC is one of the most mature among all the proposed particle accelerators. Both beams at ILC will have the capability to be polarized which is important for many measurements. The left- and right-handed electrons couple differently to the SU(2) and U(1) components of the Standard Model gauge group, so the different polarized reactions access different slices of the electroweak interaction. This increases the power of the ILC in several different respects.

The overall layout of the baseline in the TDR is shown in Figure 3.7. The latest ILC staging report 2017 proposes that ILC will collide electrons and positrons with initial

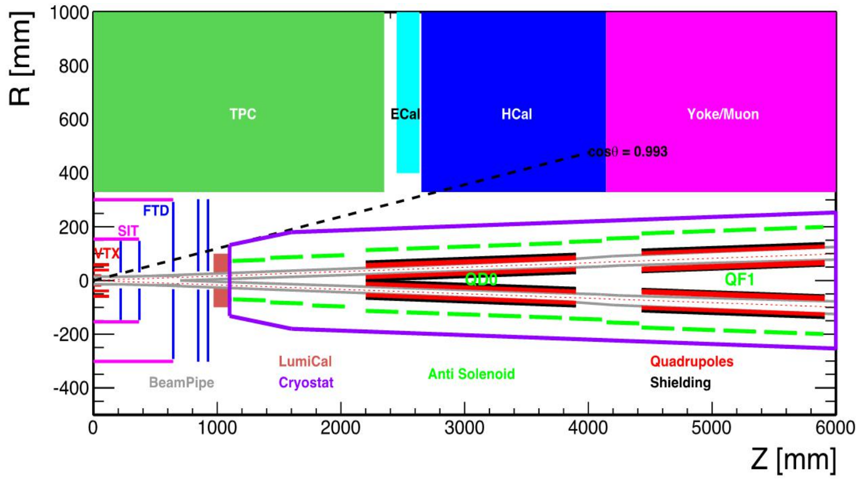

Figure 3.6: The CEPC MDI Layout

center of mass energy 250GeV, as shown in Figure 3.8. The beam power of ILC250 is 5.26MW, with the total luminosity to be 1.35 $10^{-34} cm^{-2} s^{-1}$. Following several years of successful operation of the initial ILC250, a luminosity upgrade is possible. The basic change in the luminosity upgrade is the increase in the number of bunches from 1312 to 2625.

The ILC will leave the opportunity to operate at higher center of mass energy: 350GeV, 500 GeV or 1TeV.

3.3.1 ILC Subsystems[2]

The accelerating system of ILC contains several subsystems:

- **Electron Source** The required trains of polarized electron bunches are produced with a laser hitting a photocathode in a DC gun, then bunched and pre-accelerated in normal-conducting structures. The beam is then accelerated in a superconducting linac. The spin vector is rotated into the vertical plane by superconducting solenoids, and a separate superconducting RF structure is used for energy compression before the beam is transported to the Damping Ring.

- **Positron Source** After accelerated to suitable energy, the electron beam is then extracted to a parallel beam line to create positrons and return the positrons to the electron main linac.

  In ILC the electrons pass through a helical undulator and a dogleg, generating a

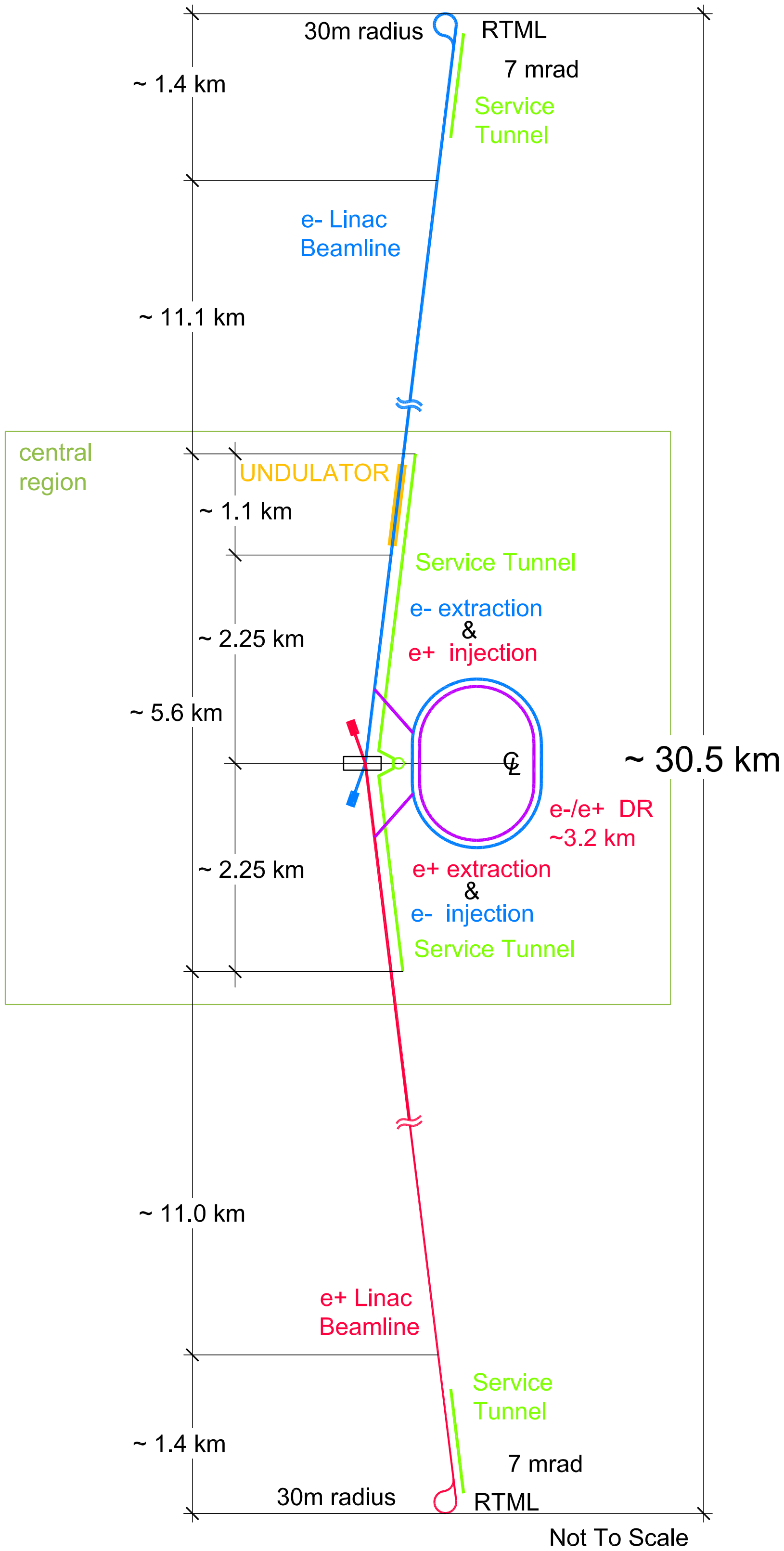

Figure 3.7: Schematic layout of the ILC complex for 500 GeV CM[2]

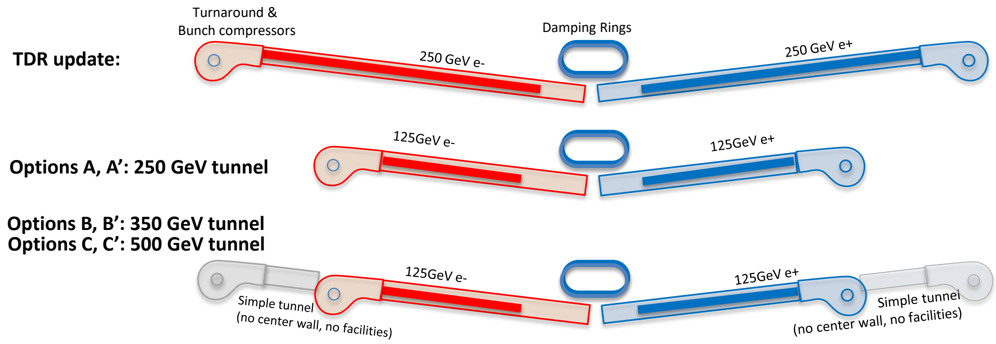

Figure 3.8: Schematic layout of the ILC250GeV staging options

monochromatic and polarized photon beam of about 10MeV. Part of this polarization is conserved when the photons hit a rotating Ti-alloy target to produce electron and positron pairs. An alternative approach uses Compton scattering of a laser beam on an electron beam from a storage ring or a linac. The laser beam is stored in optical cavities that provide several interaction points. The scattered photons are polarized. This polarization is kept with a high purity during their conversion on a fixed target. The resulting positrons are stacked in the damping ring. The independence of the system avoids the disturbance of the main electron beam due to the pass through the undulator. The cavities and the laser system are still in the focus of R&D work.

The beam is then captured, focused and pre-accelerated. After separation and dumping of the electrons and photons, the positrons enter another phase of acceleration (to 400MeV) with focusing, then transported further downstream in a superconducting linac that accelerates them to 5GeV. Before injection into the damping rings, the spin vector is rotated to the vertical direction and energy compression is performed. The polarization of the beam is about 30% and is foreseen to be upgraded to 60% later.

- **Damping Rings** In ILC, in order to achieve the design luminosity, the beam emittance has to be lowered by five orders of magnitude. In the central region, two separate damping rings, one for positron and the other one for electrons with a circumference of $\sim$ 6.7 km are housed in a single tunnel. A low operation energy of 5 GeV has been chosen. The frequency of the integrated superconducting RF system is half the frequency used in the main linac to be able to easily handle different bunch patterns.

- **Main Linac** The compressed bunch is ready to enter the Main Linac. At a distance of about 11 km, the beam particles will be accelerated to 250 GeV in ILC[1]. The un-

[1]This is the design for 500 GeV ILC. The Main Linac has been reduced to 125 GeV for 250GeV ILC,

derlying technology is based on supra-conducting 1.3 GHz RF units. The average accelerating gradient is 31.5MV/m. Three cryomodules, containing 26 nine-cell cavities make up the so-called RF units. About 280 of those are needed for each of the main linacs. This makes some 17.000 cells in total. High resolution beam pair monitors will allow having precise orbit control in order to preserve the small beam emittances over the acceleration.

- **Beam Delivery System** After exiting the main linacs the beam enters the Beam Delivery System. One of the first things needed is a measurement of the beam (energy, polarization, and emittance). Corrections are then applied on the way to the Interaction Point (IP), including the removal of the beam halo to avoid large backgrounds in the detector. A fast extraction system can be used to protect the detector and the beam line in case of failure or miss-steered beams.

- **Machine Detector Interface MDI** In ILC part of the beam delivery system will be integrated into the detector. The beam passes through a conical beam-pipe of minimal radius, as low as 15 mm at the IP. In the very forward region, sub-detector systems will record remnants of the interaction and monitor beam properties. These detectors will suffer big radiation doses.

  The beam crossing angle of ILC is 14 mrad. This angle reduces the cross section for the interaction. To provide effective head-on collisions, Crab cavities will be used to turn the beams in the horizontal plane. After the interaction and a second measurement of their properties to cross-check their stability, the beams are extracted and dumped.

  In ILC, the interaction region is shared by two detectors in a so-called "push-pull" configuration. The quadrupoles for final focus closest to the interaction point are integrated into the detector to facilitate the push-pull operation.

and 5km of each side have been reduced

3.4 The Future Circular Collider (FCC) and High Luminosity LHC (HL-LHC)

The FCC is a post-LHC particle accelerator project proposed by CERN [19], with different particle collider scenarios explored with the aim of significantly expanding the current energy and luminosity frontiers. The FCC-ee project is part of it, it is a high-luminosity, high-precision e^+e^- circular collider with a center-of-mass energy from 90 to 400 GeV, envisioned in a new 80∼100 km tunnel in the Geneva area.

The HL-LHC is an update of LHC with luminosity increased by a factor of 10 beyond the LHC's design value. The up-to-date(Oct. 2017) instantaneous luminosity have already achieved $2.0 \times 10^{34} cm^{-2} s^{-1}$. A instantaneous ultimate luminosity of $7.5 \times 10^{34} cm^{-2} s^{-1}$ and integrated luminosity to 3000 fb^{-1} is expected[37]. The preliminary studies which have been done in CMS and Atlas show that HL-LHC can extend the precision of measurements on Higgs boson couplings, Higgs width, Higgs self-couplings, etc.

Chapter 4

Detector

Detectors at the electron positron collider face a very different set of challenges compared to the previous state-of-the-art employed for LEP and hadron colliders. While the detectors at ILC and CEPC will enjoy lower rates, less background and lower radiation doses than those at the LHC, the electron positron collider will be pursuing physics that places challenging demands on precision measurements and particle tracking and identification. The reasons for this can be illustrated by several important physics processes, namely measuring the properties of a Higgs boson, identifying strong electroweak symmetry breaking, identifying supersymmetric (SUSY) particles and their properties. Taking W and Z for example, in order to distinguish them in their hadronic decay mode, the di-jet mass resolution should be comparable to their natural width, say a few GeV or less. Besides, the detector at an e^+e^- collider should be able to distinguish the Higgs signal from the SM background and to classify the Higgs events according to the generation/decay modes of the Higgs boson.

Except for the basic demands of Higgs measurements, there are slight differences between detectors at CEPC and ILC. For CEPC, the EW measurements are mostly limited by the systematics, which makes alignments, calibration, and stability crucial for the detector. For example, the CEPC detector is required to determine the luminosity to a relative accuracy of 10^{-3} for the Higgs measurements, and an accuracy of 10^{-4} for the Z pole operation. For higher energy ILC, the measurement requirements for new physics should be satisfied. For example, the low mass difference between SUSY states requires an adequate detector in the very forward direction, including an electron veto capability in the extreme forward region.

In order to meet the need for precise measurement, the Particle Flow, a full concept of detectors involving trackers and calorimeters to reconstruct individual particles is proposed as a solution.

4.1 Particle Flow Algorithm (PFA) oriented detector

PFA[44, 45] is an algorithm reconstructing all the final state particles instead of measuring jet energies globally without identifying particles. With all the final state particles correctly reconstructed, the final physics objects can be recognized with a high efficiency and purity. For example, in the flavor physics, the charged kaons/pions separation is very important.

The requirement of detector for PFA is that it should contain different sub-detectors suitable for different kind of particles. By combining the information in these sub-detectors, the PFA oriented detector design could significantly enhance the reconstruction efficiency of the key physics objects and largely improve the accuracy of jet energy resolution, since the majority of jet energy is stored in the charged hadrons, whose momentum is usually measured with a much better accuracy than its cluster energy measured at the calorimeter system.

A PFA oriented detector requires a precise tracking system with limited material budget and limited dead space between different sub-detectors. Low-material tracker is required to limit the probability of interactions before the particle reaches the calorimeter, i.e., via multi-scattering, bremsstrahlung, and hadron-nuclear interactions. To fully reconstruct individual particles from the interaction, an efficient separation of showers from charged particles, photons, and neutral hadrons in the calorimeter is required. That implies a high granularity calorimeter system. Besides, the short readout time is needed because of the high granularity.

The PFA is widely used in data analyses, both for the existing experiments and for the projects under developments, for highly granular calorimetry and for experiments without highly granular calorimetry. At the LHC, the high granularity calorimetry has already been proposed into CMS (CMS-HGC)[46] and ATLAS (ATLAS-HPTD)[47] as part of their HL-LHC upgrade program. The PFA have already been used in CMS[46], the overall JER takes a value between 6% (at P_t < 20 GeV) to 3% (at P_t > 100 GeV). The two detector designs for ILC, ILD and SiD are PFA oriented[39]. In CEPC, the baseline of detector (CEPC_v1) takes the ILD as a reference. In order to accommodate the CEPC collision environment, some necessary changes have been made to the sub-detector design. Recently another version of detector (APODIS) has been reported with optimized parameters.

4.2 Detector design

The proposed concept is designed as a multi-purpose detector, which meets the requirements in spatial and energy measurement over a large solid angle. The prototype and components of ILD and CEPC_v1 are similar, as shown in Figure 4.1, namely the multi-layer pixel-vertex detector (VTX) for reconstruction of vertices; the central silicon components SIT, SET, and ETD, providing extra precise space points to track; the large volume time projection chamber (TPC), measuring tracks with a large number of three-dimensional space points (providing a point resolution of better than 100 μm for the complete drift and a double hit resolution of less than 2 mm); the calorimetry system containing the ECAL to identify photons and measure their energy complemented by a HCAL to measure neutral hadrons; LCAL in the very forward region to measure the luminosity and in ILC the BCAL is to monitor beam parameters; the iron yoke instrumented to measure showers escaping the hadron calorimeter, and the confining magnetic field. Here the CEPC_v1 detector is introduced in detail.

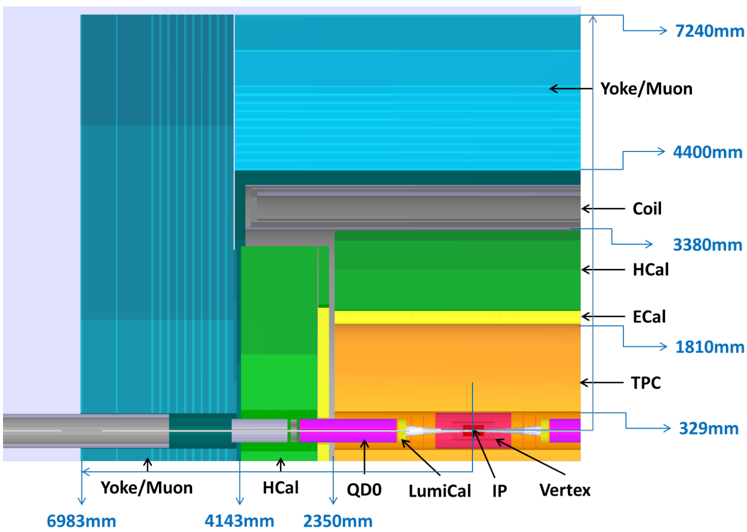

Figure 4.1: Overview of the CEPC detector in the baseline of preCDR.

- **VTX - Vertex Detector** The VTX consists of six layers of silicon pixels grouped in pairs. Optimal point resolution ($< 3\mu m$) while keeping a low material budget ($< 0.15\% X_0$/layer) is the primary design goals. This needs to be combined with a first measurement point very close to the interaction point (i.e. 16mm), which is imposed by the extreme radiation conditions as well as the strong pair background at this distance. The vertices reconstructed in VTX are important in many physics events, such as the b/c quark tagging and tau tagging.

- **FTD - Forward Tracking Discs** A set of disks equipped with silicon-pixels or silicon-strips extends the tracking down to essentially the radius of the beam tube.

- **SIT - Silicon Internal Tracker** The strong background imposes another constraint: even with a strong magnetic field, the core component of the tracking, the Time Projection Chamber, has to be kept at a distance of approximately 30 cm from the IP. To provide linking points between the VTX and the TPC, two layers of Si strips are installed in the barrel region. This will not only improve pattern recognition and momentum resolution but give also time stamps for each bunch crossing.

- **TPC - Time Projection Chamber** TPC is a cylinder with a radius of 1.8m and half-length of 2.35m. The advantage of a TPC over a silicon-based tracking system (e.g. as used in LHC experiments) is the high number of space points provided per track. The position resolution provided by TPC can be 100μm in $r - \phi$. This will play a major role in achieving the goal of a visual tracking. It will not only be possible to identify backscattering from the calorimeters, to see kinks in a track, V_0 reconstruction, as well as to recover pair production or hadronic interactions in the tracker region. Another advantage over silicon tracking is the lower material budget, a must for the best calorimeter performance. Additionally, particle ID can be performed by measuring dE/dx. This holds for K separation to isolate Kaon modes as well as for electron separation that is especially important at low energies where ID based on the calorimeter is not so good.

- **SET - Silicon External Tracker** Another set of two layers of silicon strip detectors in the barrel region are providing additional high precision spacepoints. These will not only improve the precision of the momentum measurement but can also be used to align the TPC in interplay with the SIT. Furthermore, a measurement point so close to the ECAL entry can be used as starting point for clustering algorithms.

- **ECAL** The particle flow approach requires excellent pattern recognition in the calorimeters to reconstruct individual particles. This is only possible with a short Moliere radius and with a very high granularity, cell sizes inferior to the Moliére Radius. The design of the calorimeters is driven by this goal and not by the optimization of single particle energy resolutions, although these needs still to be taken into consideration in order to achieve the desired jet energy resolutions. Both the electromagnetic as well as the hadronic calorimeters are planned as sampling calorimeters with highly segmented active layers. The materials proposed for the ECAL are tungsten as absorber and silicon as active material. It has a high longitudinal (30 layers, 24 X_0) as well as transversal segmentation ($5 \times 5 mm^2$ cell size), as shown in Fig. 4.2. Alternative designs include signal collection in scintillators, implemented as strips with alternating orientation to match effectively the separation capabilities of smaller area square cells, as well as a concept for a digital ECAL, realized with Monolithic Active Pixel Sensors (MAPS). Pixel-sizes in the order of 50 μm can ensure linearity up to high energies, leading to a total number of pixels of the order of 10^{12} for the complete ECAL.

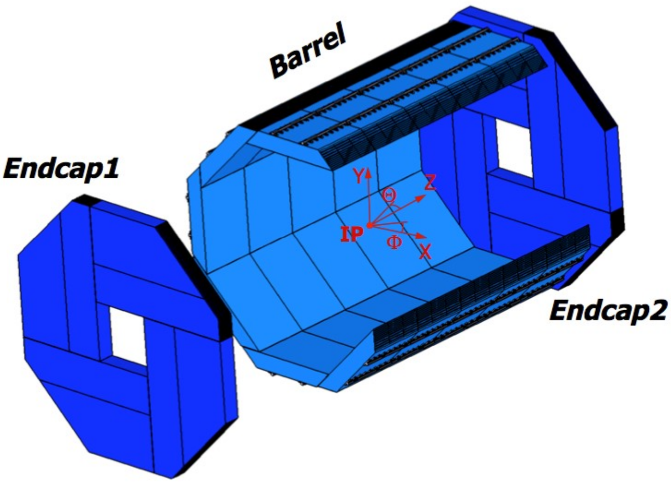

Figure 4.2: The electromagnetic calorimeter within the CEPC Detector.

- **HCAL** The HCAL is as highly segmented (48 layers for CEPC_v1 and 40 layers for APODIS, 1 cm^2 cellsize). It is a sampling calorimeter with steel as the absorber and scintillator tiles or gaseous devices with embedded electronics. The proposed structure of active layer is Glass Resistive Plate Chamber(GRPC) at CEPC. To handle the readout of such a high granularity, cells would not read out unless in a digital or semi-digital mode.

- **Coil** A superconducting coil providing a nominal field of 3.5 Tesla and representing 2.2 interaction lengths surrounds the two calorimeters. A field of this strength will contain the core of the pair background in the beampipe. Also the curvature of the track of a charged particle scales proportional to B. This means improvement in the momentum resolution with higher field strength as well as a better separation of charged tracks from neutrals at a given inner radius of the calorimeter.

- **Yoke** A magnetic field of this strength has to be closed to minimize stray fields. An iron yoke is used for this purpose. This yoke is then instrumented with RPC's. The system serves like this as tagger for high energy muons.

Chapter 5

Softwares and Particle Flow Algorithm (PFA)

To accomplish the goal of future electron positron collider, the hadronic decays of W and Z bosons should be separated via the reconstruction of the di-jet invariant masses. This implies that a di-jet mass resolution of about 3.5% for jets has to be achieved. A broadly accepted approach to reach these resolutions is the Particle Flow concept. In this chapter, it will be shown that this method will impose constraints on the detector that demand a very special design that has never been attempted before. The tools used are also introduced.

5.1 Particle Flow Algorithm

Several Particle Flow Algorithms (PFA) have been developed, such as GARLIC (GAmma Reconstruction at a LInear Collider)[48], specified to identify photons in the high granularity calorimeter, or global to identify and measure particles reaching the semi-digital hadron calorimeter, with good separation between nearby showers, such as PandoraPFA[45] and Arbor[49].

5.1.1 Jet Energy Resolution

A jet is defined as a narrow cone of particles produced by the hadronization of a quark or gluon, it is an important object to be observed in particle physics experiments because of its high production cross section. In the traditional calorimetry, jet energy is obtained from the sum of energies deposited in ECAL and HCAL, pointing to a jet energy resolution with a stochastic term greater than 60%[50], which does not allow to separate the hadronic decays of W and Z and does not meet the requirements of ILC and CEPC. In PFA, a jet is the sum of the individual particles divided into three part: charged particles whose momenta are measured in the tracking detectors (providing a momentum resolution as good as $\sigma_{tracker} \sim 5 \cdot 10^{-6} p_T^2$), photons whose energies are best measured in ECAL (with energy resolution typically of $\sigma(E)/E \sim 0.16/\sqrt{E}$) and neutral hadrons whose energy obtained from the HCAL (with energy resolution of $\sigma(E)/E \sim 0.5/\sqrt{E}$). Since the average jet energy content is of 65% from the charged track(s), 26% from the photon(s) and 9% from neutral hadron(s), the HCAL which has the worst resolution used to measure only less than 10% of the energy in the jet. Thus the energy resolution of a jet can be as good as needed. Since $\sigma(E)/E = a/sqrtE \bigoplus b/E \bigoplus c$ where $a/sqrtE$, b/E and c are the stochastic response, electronic noise term and constant term caused by dead material, the assumption that the constant term for ECAL and HCAL to be 1% and 2% can be made (more dead zones in HCAL), while the noise term for ECAL and HCAL assumed to be $0.3/E$ and $0.1/E$ (more electronics in ECAL). Taking the above resolutions as hypothesis, one can see in Figure 5.1 and Figure 5.2 that the tracker measurement would only be beaten by calorimeters for particles above 500 and 700 GeV for electrons and hadrons, which is not the case in 250 GeV e^+e^- colliders, see Figure 5.3.

5.1.2 PandoraPFA

PandoraPFA has been created by Mark Thomson[45] after the 2005 Snowmass workwhop on the Linear Collider. There are eight main steps to reconstruct particle flow in PandoraPFA:

1) **Track topology** Tracking is done separately in PandoraPFA, track topologies of neutrals in the detector volume are identified and classified according to their ways of decays, and they are projected onto the front face of the ECAL.

2) **Calorimeter Hit Selection and Ordering** Isolated hits defined by proximity to others in the calorimeter are removed at this stage, and the selected hits are stored with four-vector information after calibration, geometry, isolation, MIP identification and ordering.

3) **Clustering** Hits are either added to existing clusters (if a hit lies within the cone defined by existing cluster, and is suitably close) or they are used to seed new clusters (if the hit is unmatched) in this stage. This process starts at innermost layers and works outward, considering each calorimeter hit in turn. In order to follow tracks in the calorimeters, the algorithm clusters are assigned a direction (or potentially directions) in which they are propagating.

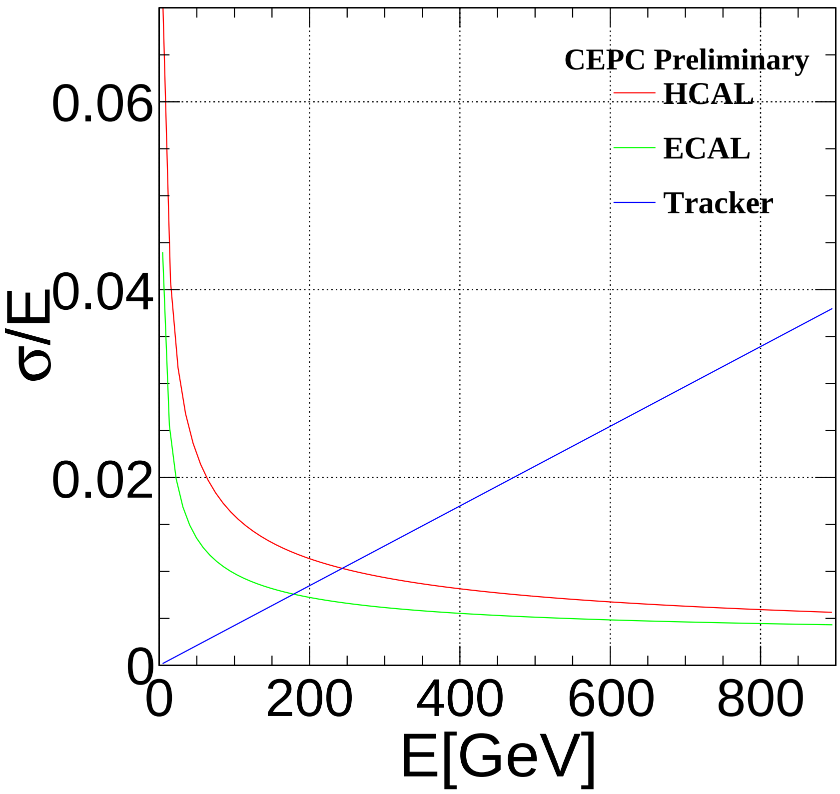

Figure 5.1: The energy resolution of TPC, ECAL and HCAL at different energy (for a direction perpendicular to the magnetic field).

4) **Topological Cluster Merging** Clusters which have not been identified as photons are associated together making use of high granularity for tight cluster association, or clear topologies.

5) **Statistical Re-clustering** For jets with energy higher than 50GeV, the performance degrades due to the increasing overlap between hadronic showers from different particles. If a significant discrepancy between the energy of a cluster and momentum of its associated track is identified, this stage is applied by altering clustering parameters, or changing clustering algorithm entirely, until cluster splits in such a way that sensible track-cluster associations are obtained.

6) **Photon Identification and Recovery** The tagging of photons is improved by applying photon identification algorithm to the clusters and the cases where a primary photon is merged with a hadronic shower from a charged particle are recovered.

7) **Fragment Removal** Relevant clusters are merged together in this stage by removing neutral clusters (no track-associations) that are really fragments of charged (track-associated) clusters and merging them with the appropriate parent charged cluster.

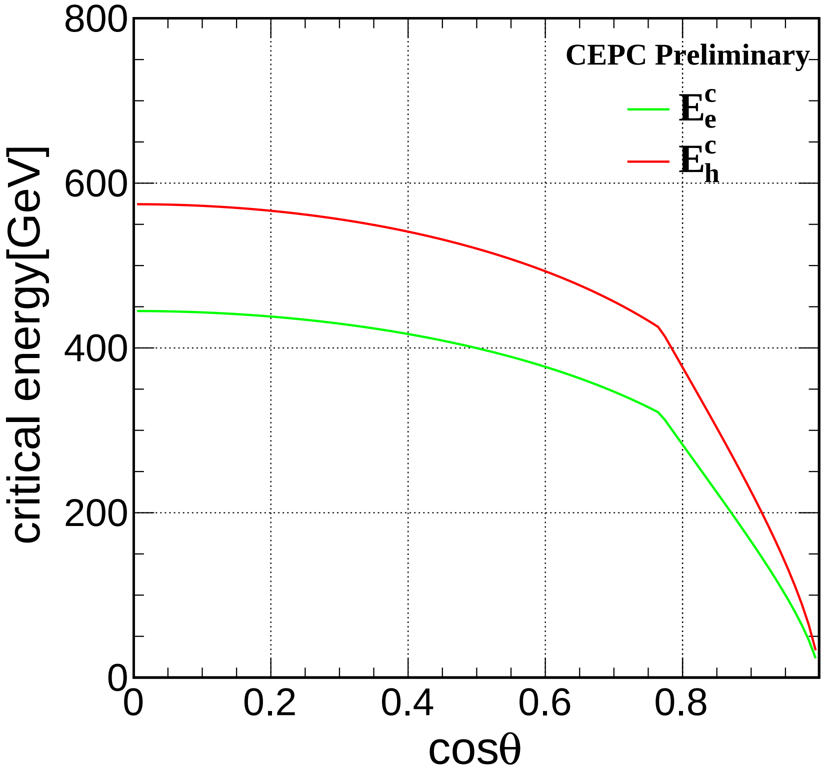

Figure 5.2: The critical energy where the energy resolution of ECAL or HCAL is the same as TPC for different direction.

8) **Formation of Particle Flow Objects** The final stage of PandoraPFA is to build Particle Flow Objects (PFOs) from the results of the associated clustering combined with tracks. Relatively primitive particle identification is applied and the reconstructed PFOs, including four-momenta, are written out in LCIO(Linear Collider I/O)) format, which will be introduced in next section.

For R&D study in ILD, the JER got from Pandora can reach 3% for high energy jets, as shown in Figure 5.4.

5.1.3 Arbor

Arbor algorithm is inspired by the fact that the shower spatial development follows the topology of a tree.[49] With a granularity calorimeter, Arbor could efficiently separate nearby particle showers and reconstruct the inner structure of a shower. Arbor also maintains a high efficiency in collecting the shower hits or energy, which is appreciated for the shower energy estimation.

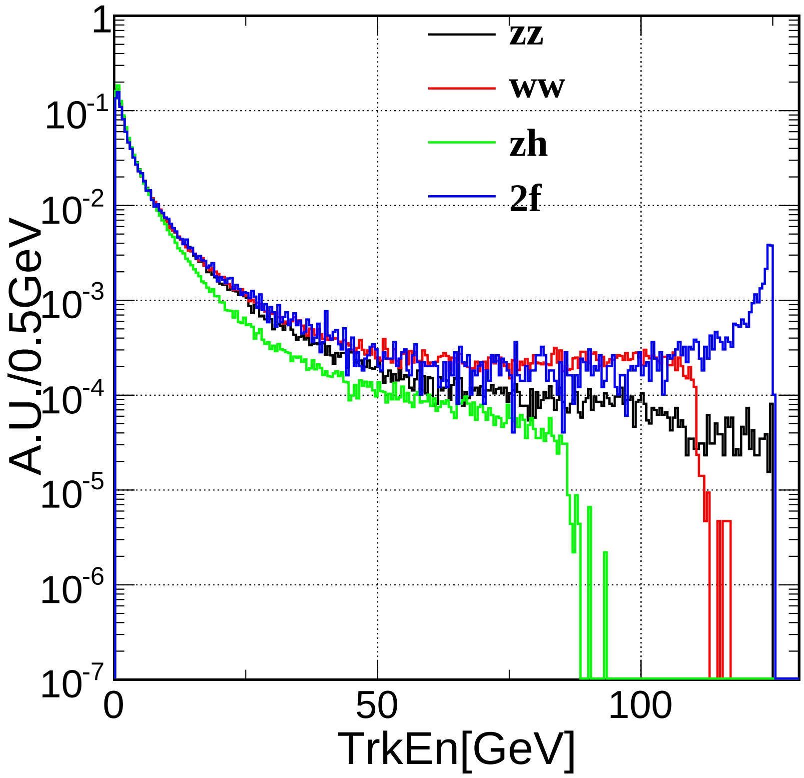

Figure 5.3: Charged particle energy spectra for different physics processes with different final states: ZH, WW, ZZ, or 2 fermions events at center of mass 250 GeV

The steps to reconstruct particle flow in Arbor is:

1) **Hits Connecting** After necessary hit cleaning, if the distance between any pair of hits is smaller than a given threshold, a local connector is build. The connector is an orientated arrow which links a pair of hits and ends at the hit with larger transverse distance to the origin.

2) **Clean Connectors** After the first step, there can be multiple connectors end or begin at a given hit. Using the directions and length of these connectors as well as the spatial position of the hit, a reference direction can be calculated. From all the connectors ending at this hit, Arbor keeps at most one connector that has the minimal angle to the reference direction. Therefore, no loop structure will be kept after the cleaning and a tree structure based on the connectors emerges.

3) **Iteration** New connectors can be added according to the relative positions between hits as well as their reference directions, and the set of connectors can always be cleaned with similar criteria. The purpose of the iteration is simply to find the best connector configurations, in the sense that every branch should be as smooth as possible and al-

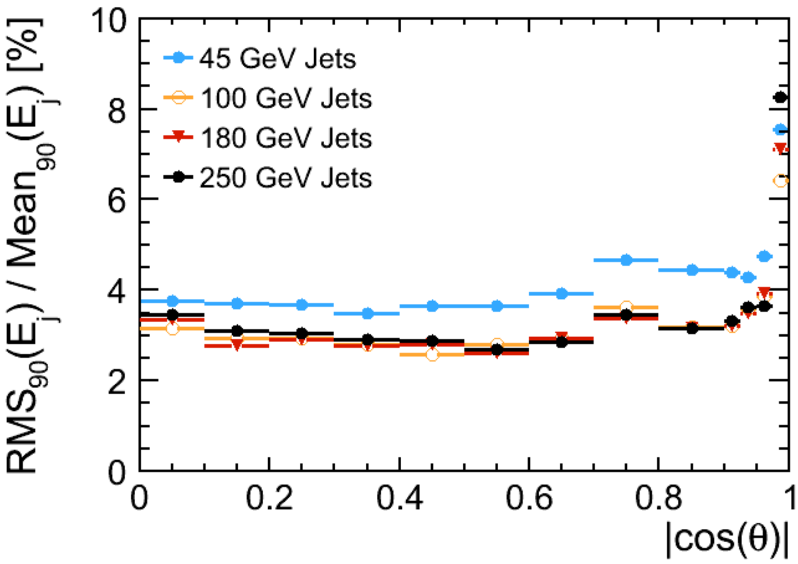

Figure 5.4: Jet Energy Resolution (JER) of PandoraPFA for various angles and energies for ILC using $Z \to uds$ samples. For $\cos(\theta) < 0.95$ and energy > 45 GeV, it meets the requirement of separating W and Z bosons. The JER is expressed in RMS_{90}, the RMS in the smallest range of reconstructed energy which contains 90% of the events.

lowance for long connectors.

4) **Clustering** After the last step the tree structure is built and decoupled into sets of branches. The topology of each cluster is used in a pre-identification.

5) **Building Particle Flow Objects** The final stage is to build Particle Flow Objects (PFOs) from the results of the associated clustering combined with tracks, similar as for PandoraPFA.

The performance of Arbor PFA can be revealed in two aspects:

- The separation performance, i.e., to successfully reconstruct nearby incident particle.
- The jet reconstruction performance.

As shown in Figure. 5.5 and Figure 5.6, Arbor could efficiently separate nearby particle showers and reconstruct the inner structure of the shower. For physics events with only two jets, the boson mass could be measured to a relative accuracy better than 4% at CEPC reference detectors.

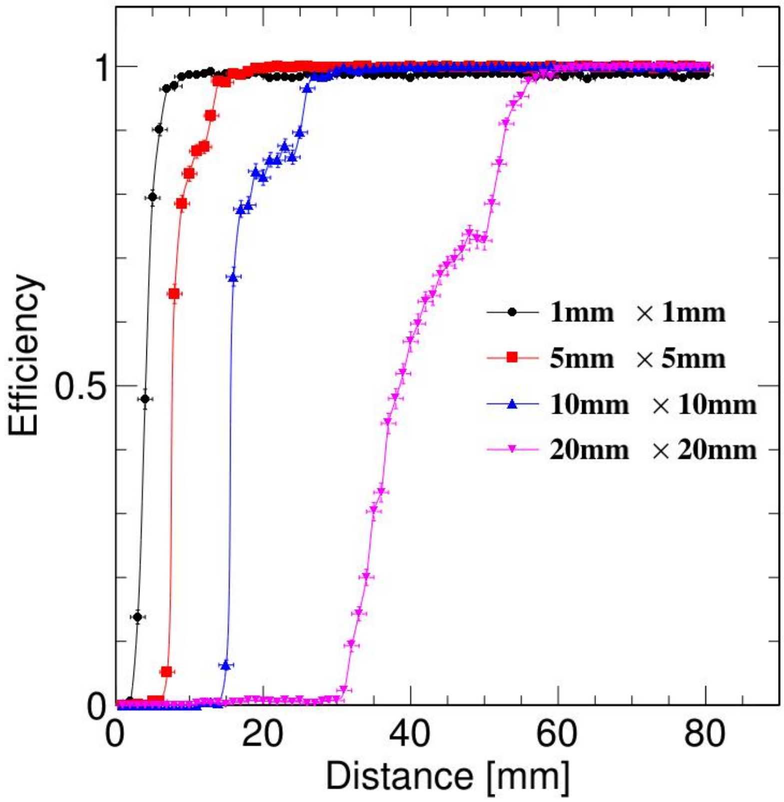

Figure 5.5: Reconstruction efficiency depending on distance of the di-photon system. The different lines corresponds to different ECAL cell sizes. The efficiency is defined as the probability of successfully reconstructing two photons with anticipated energy and incident positions.

BMR(Boson Mass Resolution), the resolution of the mass of Higgs boson in $\nu\nu H$ with $H \rightarrow qq$ events is used as a standard expression of performance in CEPC. In order to focus on the performance of the detectors or reconstructions, the events with ISR photons, with neutrinos from Higgs, or with jets shooting to the endcaps are not taken into account. As shown in Figure 5.7.

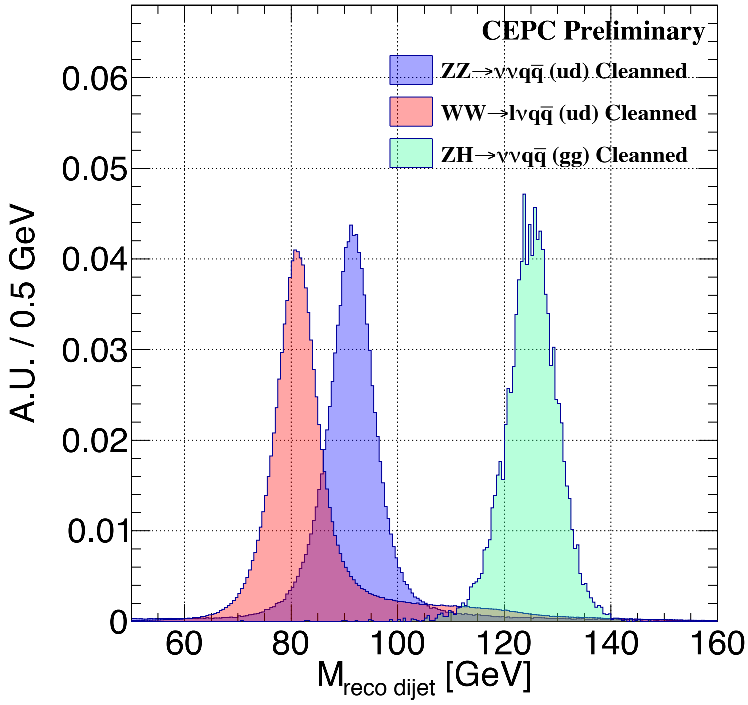

Figure 5.6: Reconstructed boson masses from cleanned $\nu\nu$ events , $l\nu qq$ events and $\nu\nu H$ with $H \to qq$ events. Here only events with final state jets to be fragmented from either light flavor quarks or gluons are taken into account. The events with ISR are also excluded.

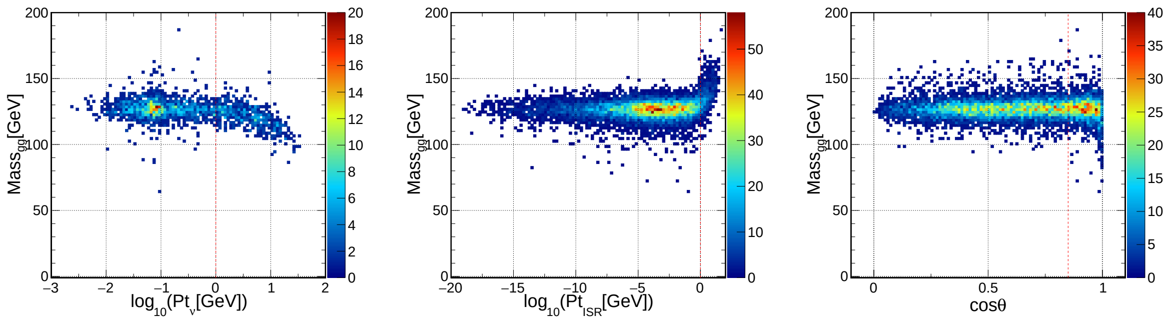

Figure 5.7: Reconstructed boson masses from $\nu\nu H$ with $H \to qq$ events depending on P_t of ISR, P_t of Higgs decayed neutrino, and $\cos\theta$ where θ is the maximum angle of the jet direction. The dashed line shows the event selection for BMR.

5.2 Tools

5.2.1 LCIO[3]

LCIO (Linear Collider Input/Output) is a persistence framework and event data model for linear collider detector studies. It is intended to be used in both simulation studies and analysis frameworks. Its lightweight and portability make it also suitable for use in detector R&D testbeam applications. It provides a C++ and a Java implementation with a common interface (API): a Fortran interface to the C++ the implementation also exists.

Using a common persistence format and event data model allows to easily share results and compare reconstruction algorithms. LCIO is used by almost all groups involved in linear collider detector studies and thus has become a de facto standard.

5.2.2 Simulation

The tool applied in this report for simulation is MOKKA[51], based on GEANT4[52]. In order to run Mokka, the first step is to set up the environment parameters, defining the global environment variables such as the working directory, where GEANT4 is installed, the implementation of the Mokka database, the installation of LICO and GEAR, as well as the shared libraries path to be scanned when running Mokka. After Mokka is built, a steering file containing the information of the simulation should be prepared. This file defines the database and user to obtain the geometry information, the output files, the Macro file to give commands, detector mode (one can change the geometry of detector by removing subdetectors) and so on. The physics list (see GEANT4) is also chosen in this file, which is used to describe the modeling of the interaction of high energy hadrons, here QGSP. In the Macro file, the information of the particles can be generated from particle gun (where the particle type, position, direction, smearing and others are set) or by events generated from elsewhere (from HEPevt input file), the energy and events number of simulation are also defined in this file.

5.2.3 Marlin Framework[4]

The software tool used for full simulation is Mokka, based on Geant4, which can write an LCIO file defining the parameters for subdetectors. After the generation of the events, and the simulation of the detector response using MOKKA, reconstruction software is used to reconstruct and analyze the events. In order to identify individual particles, new tools for reconstruction are required.

Marlin(**M**odular **A**nalysis and **R**econstruction for the **LIN**ear collider) is a modular `C++` application framework for ILC detector reconstruction and analysis LCIO data. Marlin is first configured by an XML steering file containing parameters defined for individual processors or globally, the order in which the processors are called and the conditions applied to *Processors* (plug-in modules that can be loaded at runtime to implement some core functionality) evaluating with the runtime. The LCIO files, which contain data such as hits, tracks, and clusters, will be used by processors according to the need for reconstruction.

5.3 Detector optimization

The optimization of detectors for CEPC and ILC is a balance between the budget and the performance. In this section, two examples of optimization using the tools above will be shown.

5.3.1 ECAL optimization

The cost of detectors for CEPC and ILC is always a matter to consider. Therefore optimization is ongoing to reduce the price and mantain good performance at the same time. The ECAL is the major cost of ILD, because of the high price of silicon wafers. This provides options to optimize, such as the inner radius of ECAL, the number of Si layers in the ECAL, etc. In this section, the performances of modified detector with a reduced radius and number of Si layers in ECAL is studied. The detector model used here is an ILD detector with the TPC radius reduced from 1800 to 1400mm (the length is modified accordingly), and the ECAL layer number reduced from 30 layers to 26/20 layers. The total absorber thickness, the ratio of W thickness between inner and outer absorber layers, carbon fiber, cooling layers, Si thickness, etc., remain the same for the three models. The $Z \to qq$ events with the centre of mass energy range from 91GeV to 500GeV are generated and reconstructed with PandoraPFA, after calibration to set the digitization constant depending on different sampling fraction in the ECAL of each model. The resolution is expressed with RMS90, defined as the RMS in the smallest range of reconstructed energy which contains 90% of the events, in order to handle properly the non-Gaussian energy distribution with a tail corresponding to the population of events where the confusion is significant. As shown in Figure 5.8, the JER increases 10% to 91 GeV di-jets and less than 5% for 100 GeV di-jets by decreasing the number of Si layers from 30 to 20. At the 250GeV e^+e^- colliers the typical jet energy is less than 70GeV, as shown in Figure 5.9, corresponding to the 91 GeV di-jets.

A comparison with Arbor using the invariant mass resolution of 250GeV $\nu\nu gg$ events is

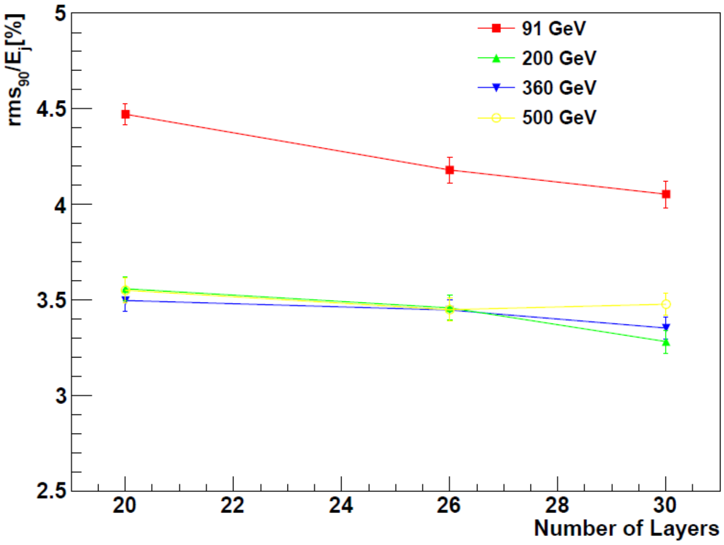

Figure 5.8: JER comparison for different jets energy in function of layer numbers, a cut $|\cos(\theta_{jet})| < 0.7$ is applied to avoid the endcap area.

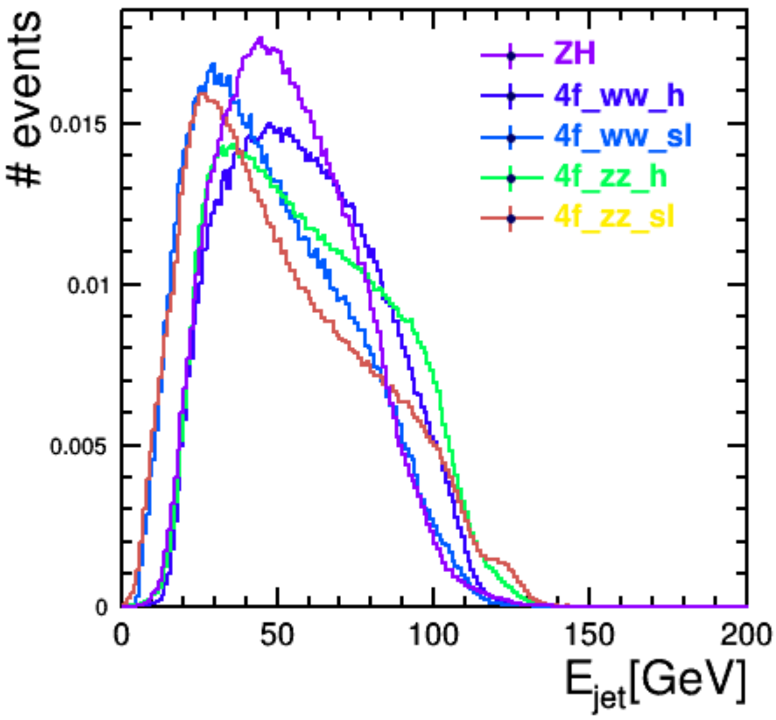

Figure 5.9: jet energy spectra for different physics processes with different final states: ZH or 4 fermions from W bosons or Z bosons decays, at center of mass 250 GeV

shown in Figure 5.10, with the resolution expressed in BMR and not only the number of layers but also the ECAL cell size is taken into account. Since the total number of readout will decrease with the cell size, the cooling system might be inactive if the cell size enlarged. The events with ISR and events with jet direction to the endcaps are excluded. It is shown that the degradations of performance using the two frameworks are similar to each other.

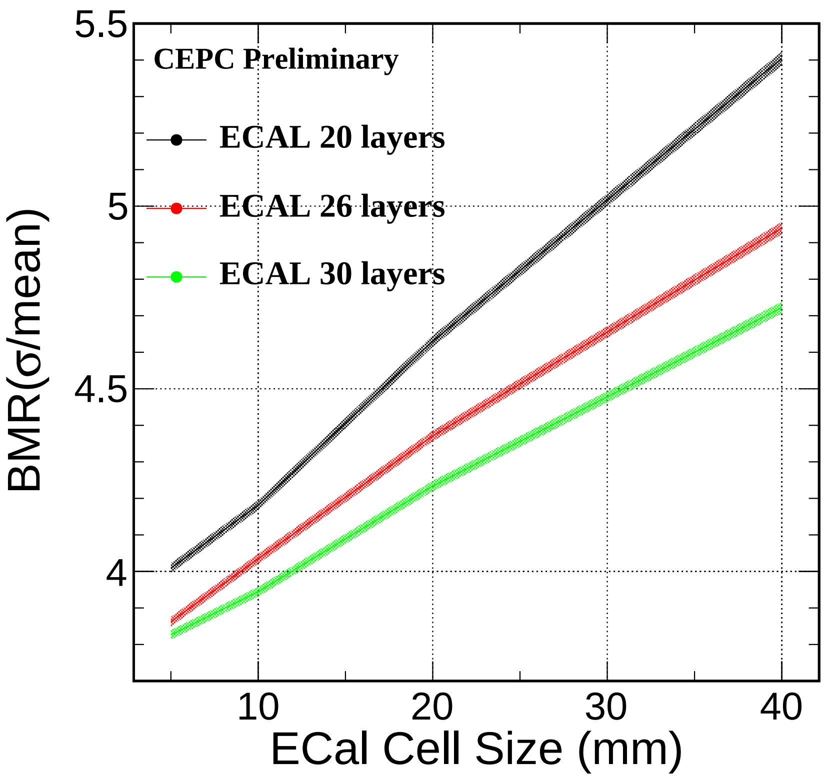

Figure 5.10: The invariant mass resolution of 250GeV $\nu\nu gg$ events in CEPC for different number of ECAL layers and different ECAL cell sizes.

5.3.2 HCAL and B field optimization

For HCAL, the optimization is done for a reduced number of layers while the thickness of each layer remains the same. The B field is allowed to be reduced because of the high granularity. The $\nu\nu gg$ events are generated in CEPC detector with HCAL layers range from 20 to 48 and B field to be (2.5T, 3.0T, 3.5T) and reconstructed with Arbor (v3.3). The resolution is expressed as the resolution of the reconstructed invariant mass, with final state jets from either light flavor quarks or gluons and the events with ISR excluded. As shown in Figure 5.11 and Figure 5.12, the resolution degrades by 0.1 while the number of layers reduces from 48 to 20. This result also leaves an opportunity to degrade the B field in CEPC to 3 Tesla, which is appreciated by the MDI and will be applied for the baseline of CDR. In the new version of CEPC detector, the baseline of HCAL layer number is chosen to be 40.

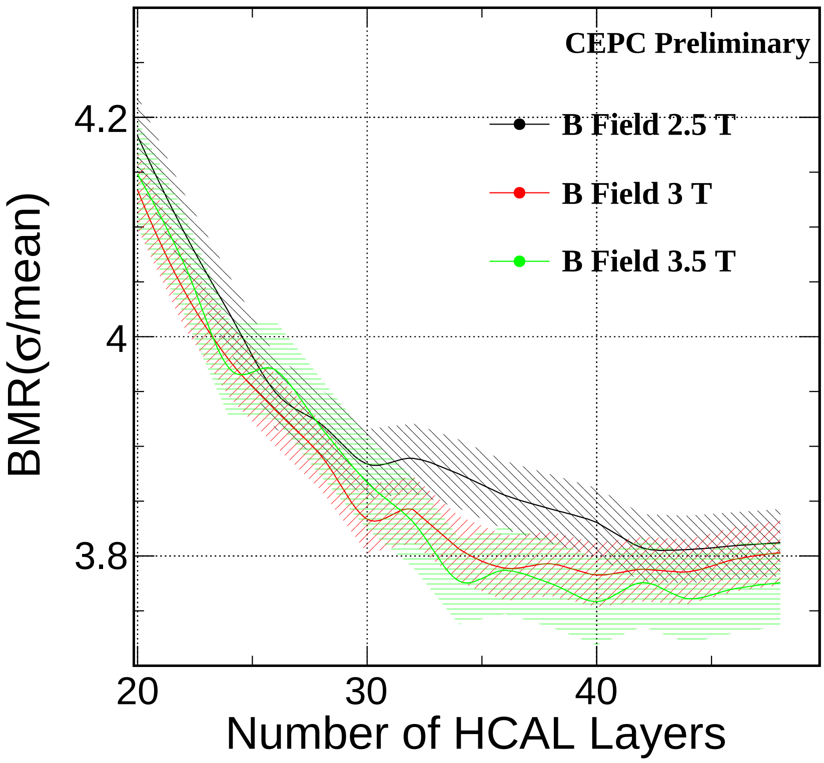

Figure 5.11: The invariant mass resolution of 250GeV $\nu\nu gg$ events in CEPC for different number of HCAL layers.

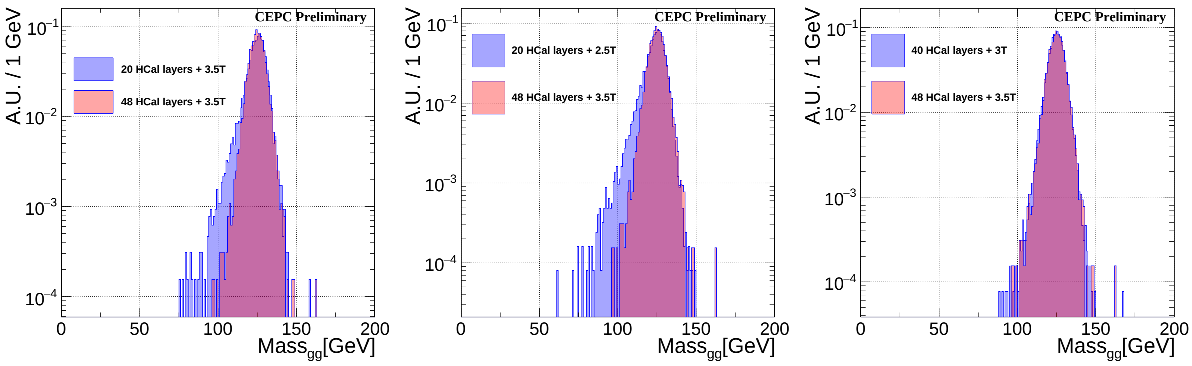

Figure 5.12: The Higgs boson invariant mass for 250 GeV $\nu\nu qq$ events, with different B fields and different HCAL layer numbers, comparing with the baseline geometry in preCDR. The last plot is the baseline for CDR.

Chapter 6

Particle identification

The lepton identification is essential to the precise Higgs boson measurements. The Standard Model Higgs boson has roughly 10% chance to decay into final states with leptons, for example, $H \to WW^* \to ll\nu\nu/l\nu qq$, $H \to ZZ^* \to llqq$, $H \to \tau\tau$, $H \to \mu\mu$, etc. The SM Higgs also has a branching ratio Br(H→bb) = 58%, where the lepton identification provides an important input for the jet flavor tagging and the jet charge measurement. On top of that, the Higgs boson has a significant chance to be generated together with leptons. For example, in the ZH events, the leading Higgs generation process at 240-250 GeV electron-positron collisions, about 7% of the Higgs bosons are generated together with a pair of leptons (Br(Z→ee) and Br(Z→ $\mu\mu$) = 3.36%). At the electron-positron collider, ZH events with Z decaying into a pair of leptons is regarded as the golden channel for the HZZ coupling and Higgs mass measurement[53]. Furthermore, leptons are intensively used as a trigger signal for the proton colliders to pick up the physics events from the huge QCD backgrounds.

6.1 Detector geometry and sample

In this section, the reference geometry is the CEPC conceptual detector [18], which is developed from the ILD geometry.

To study the lepton identification performance, we simulated single particle samples (pion+, muon-, and electron-) over an energy range of 1-120 GeV (1, 2, 3, 5, 7, 10, 20, 30, 40, 50, 70, 120 GeV). At each energy point,100k events are simulated for each particle type. These samples follow a flat distribution in theta and phi over the 4π solid angle.

These samples are reconstructed with Arbor (version 3.3). To disentangle the lepton identification performance from the effect of PFA reconstruction and geometry defects, we select those events where only one charged particle is reconstructed. The total number of these events is recorded as $N_{1Particle}$, and the number of these events identified with correct particle types is recorded as $N_{1Particle,T}$. The performance of lepton identification is then expressed as a migration matrix in Table 6.1, its diagonal elements ϵ_i^i refer to the identification efficiencies (defined as $N_{1Particle,T}/N_{1Particle}$), and the off diagonal element P_j^i represent the probability of a type i particle to be mis-identified as type j.

Table 6.1: Migration Matrix

	$e^- like$	$\mu^- like$	$\pi^+ like$	undefined
e^-	ϵ_e^e	P_μ^e	P_π^e	P_{und}^e
μ^-	P_e^μ	ϵ_μ^μ	P_π^μ	P_{und}^μ
π^+	P_e^π	P_μ^π	ϵ_π^π	P_{und}^π

6.2 Discriminant variables and the output likelihoods

LICH takes individual reconstructed charged particles as input, extracts 24 discriminant variables for the lepton identification, and calculates the corresponding likelihood to be an electron or a muon. These discriminant variables can be characterized into five different classes:

- **dE/dx**

  For a track in the TPC, the distribution of energy loss per unit distance follows a Landau distribution. The dE/dx estimator used here is the average of this value but after cutting tails at the two edges of the Landau distribution (first 7% and last 30%). The dE/dx has a strong discriminant power to distinguish electron tracks from others at low energy (under 10 GeV) (Figure 6.1).

- **Fractal Dimension**

  The fractal dimension (FD) of a shower is used to describe the self-similar behavior of shower spatial configurations, following the original definition in [54], the fractal dimension is directly linked to the compactness of the particle shower. The FD of a shower is expressed as $FD_\beta = \langle log()R_{\alpha,\beta}/log\alpha\rangle + 1$ where $R_{\alpha,\beta} = N_\beta/N_\alpha$ represents the ratio of the number of hits at different scales. Here β range from $10mm$ to $150mm$ and α is $10mm$.

  At a fixed energy, the EM showers are much more compact than the muon or hadron shower, leading to a large FD. The muon shower usually takes the config-

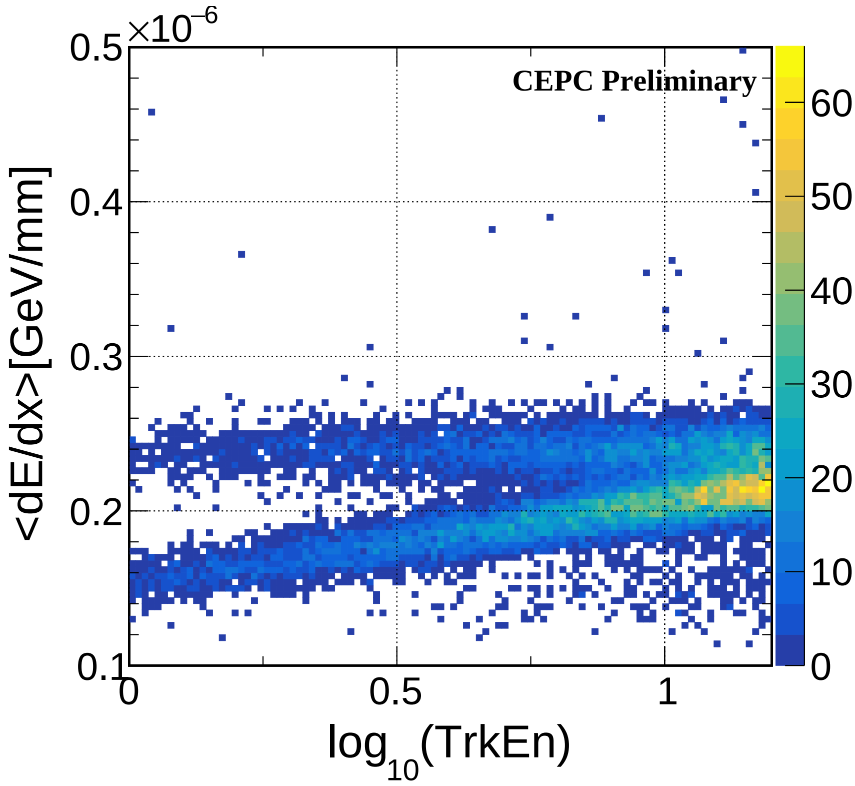

Figure 6.1: dE/dx for e^-, μ^- and π^+, for electrons it is stable around 2.4×10^{-7}, for muon and pion it is smaller at energy lower than 10 GeV and after that they start mixing with electron

uration of a 1-dimensional MIP(Minimum Ionizing Particle) track, therefore has an FD close to zero. The FD of the hadronic shower usually lays between the EM and MIP tracks, since it contains both EM and MIP components. A typical distribution of F_all (the fractal dimension using both ECAL and HCAL) for 40 GeV showers is presented in Figure 6.2,

For any calorimeter cluster, LICH calculates 5 different FD values: from its ECAL hits, HCAL hits, hits in 10 or 20 first layers of ECAL, and all the calorimeter hits.

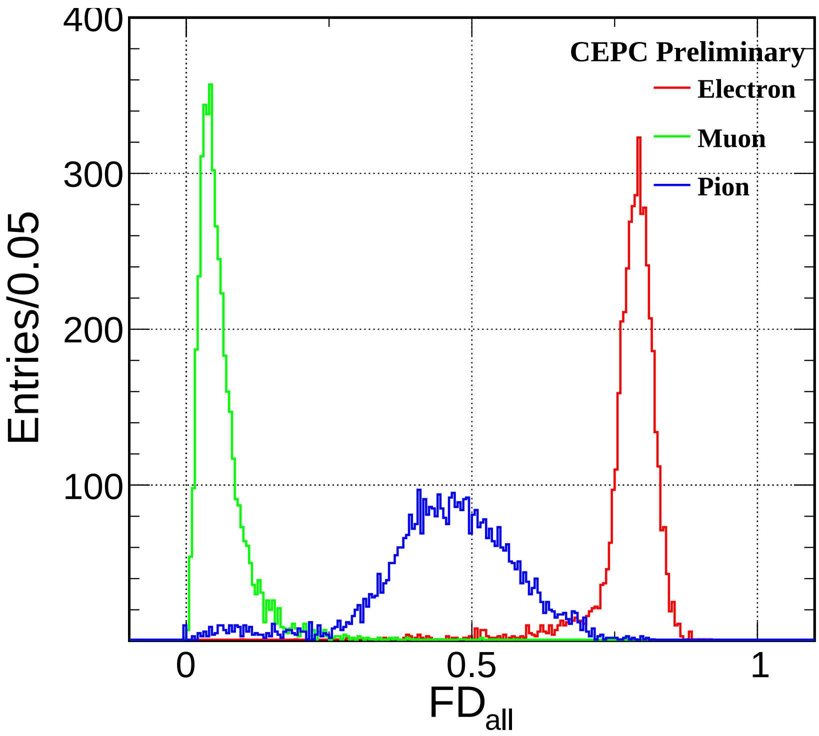

Figure 6.2: Fractal dimension using both ECAL and HCAL for e^-, μ^- and π^+ at 40 GeV

- **Energy Distribution**

  LICH builds variables out of the shower energy information, including the proportion of energy deposited in the first 10 layers in ECAL to the entire ECAL, or the energy deposited in a cylinder around the incident direction with a radius of 1 and 1.5 Moliere radius.

- **Hit Information**

  Hits information refers to the number of hits in ECAL and HCAL and some other information obtained from hits, such as the number of ECAL (HCAL) layers hit by the shower, number of hits in the first 10 layers of ECAL.

- **Shower Shape, Spatial Information**

  The spatial variables include the maximum distance between a hit and the extrapolated track, the maximum distance and average distance between shower hits and the axis of the shower (defined by the innermost point and the center of gravity of the shower), the depth (perpendicular to the detector layers) of the center of gravity, and the depth of the shower defined as the depth between the innermost hit and the outermost hit.

The correlations of those variables at energy 40 GeV are summarized in Figure 6.4, the definitions of all the variables are:

- NH_ECALF10: Number of hits in the first 10 layers of ECAL
- FD_ECALL20: FD calculated using hits in the last 20 layers of ECAL
- FD_ECALF10: FD calculated using hits in the first 10 layers of ECAL
- AL_ECAL: Number of ECAL layer groups (every five layers forms a group) with hits
- av_NHH: Average number of hits in each HCAL layer groups (every five layers forms a group)
- rms_Hcal: The RMS of hits in each HCAL layer groups (every five layers forms a group)
- EEClu_r: Energy deposited in a cylinder around the incident direction with a radius of 1 Moliere radius
- EEClu_R: Energy deposited in a cylinder around the incident direction with a radius of 1.5 Moliere radius
- EEClu_L10: Energy deposited in the first 10 layers of ECAL
- MaxDisHel: Maximum distance between a hit and the helix
- minDepth: Depth of the innermost hit
- cluDepth: Depth of the cluster position
- graDepth: Depth of the cluster gravity center
- EcalEn: Energy deposited in ECAL
- avDisHtoL: Average distance between a hit to the axis from the innermost hit and

the gravity center

- maxDisHtoL: Maximum distance between a hit to the axis from the innermost hit and the gravity center
- NLHcal: Number of HCAL layers with hits
- NLEcal: Number of ECAL layers with hits
- HcalNHit: Number of HCAL hits
- EcalNHit: Number of ECAL hits

The distribution of all the variables used in TMVA are shown in Figure 6.3

It is clear that the dE/dx, measured from tracks, does not correlate with any other variables which are measured from calorimeters. Some of the variables are highly correlated, such as FD_ECAL (FD calculated from ECAL hits) and EcalNHit (number of ECAL hits). However, all these variables are kept because their correlations change with energy and polar angle.

LICH uses TMVA[22] methods to combine these input variables into two likelihoods, corresponding to electrons and muons. Multiple TMVA methods have been tested and the Boosted Decision Trees with Gradient boosting (BDTG) method is chosen for its better performance. The e-likeness (L_e) and μ-likeness (L_μ) for different particles in a 40 GeV sample are shown in Figure 6.5.

The overtraining check of Muon BDT response at 40GeV is shown in 6.6 as an example.

The weight of the 24 variables varies with different energies, at 2GeV the 5 most important variables are: dE/dx, cluDepth, EcalNHit, E_r, and maxDisHtoL, while at 40GeV the 5 most important variables are: E_10, FD_all, NLEcal, EcalNHit, and avDisHtoL. Taking the 5 GeV energy point as an example, the charged particle identification efficiency for 15, 10, 5 variables are shown in Table 6.2.

Table 6.2: The efficiency of charged particle identification at 40 GeV (%), training with different number of variables

Number of variables	5	10	15	24
e^-	96.3	98.3	98.7	99.7
μ^-	97.1	99.2	99.2	99.9
π^+	94.7	97.7	98.2	99.3

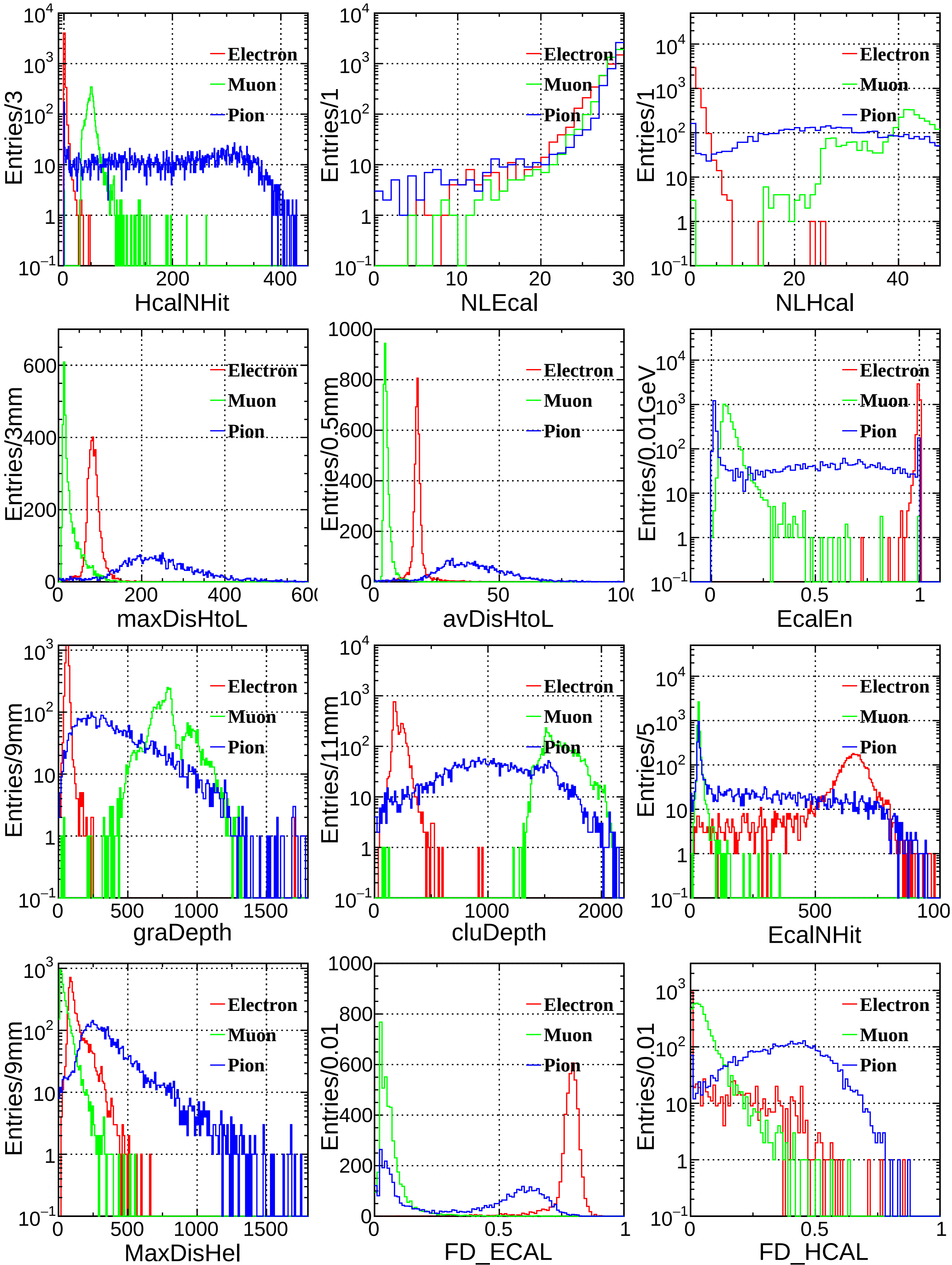

Figure 6.3: Calorimeter based variables used in TMVA (40GeV) (to be continued)

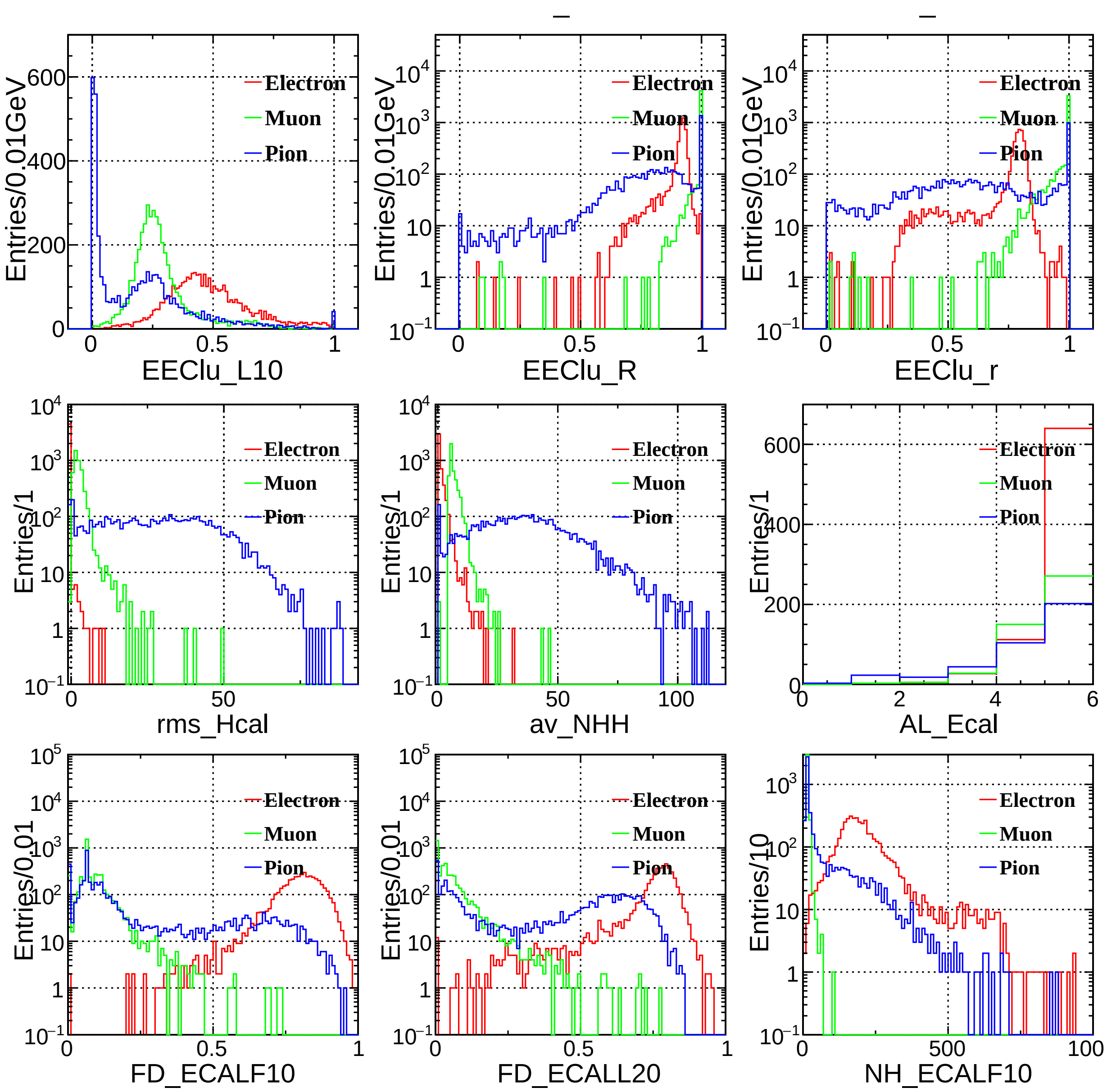

Figure 6.3: Calorimeter based variables used in TMVA (40GeV)

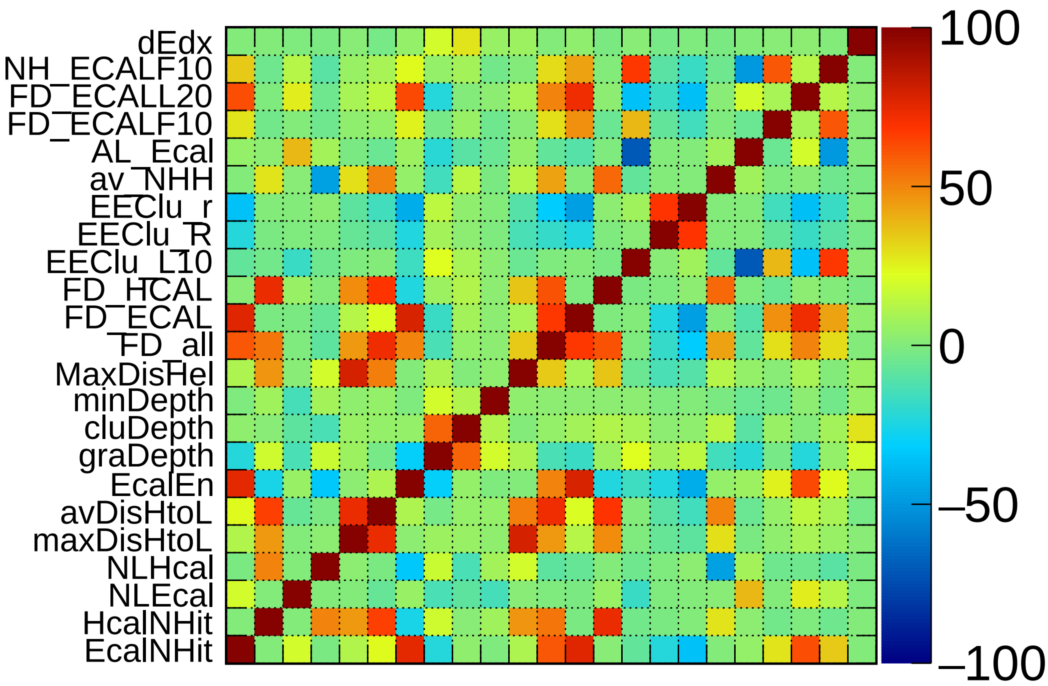

Figure 6.4: The correlation matrix of all the variables

6.3 Performance on single particle events

The phase space spanned by the lepton-likelihoods (L_e and L_μ) can be separated into different domains, corresponding to different catalogs of particles. The domains for particles of different types can be adjusted according to physics requirements. In this paper, we demonstrate the lepton identification performance on single particle samples using the following catalogs:

- Muon: $L_\mu > 0.5$
- Electron: $L_e > 0.5$
- Pion: 1-(L_μ+L_e)> 0.5
- Undefined: $L_\mu < 0.5$ & $L_e < 0.5$ & 1-(L_μ+L_e) < 0.5

The probabilities of undefined particles are very low (<10^{-3}) at single particle samples with the above catalog.

Since the distribution of these variables depends on the polar angle of the initial particle

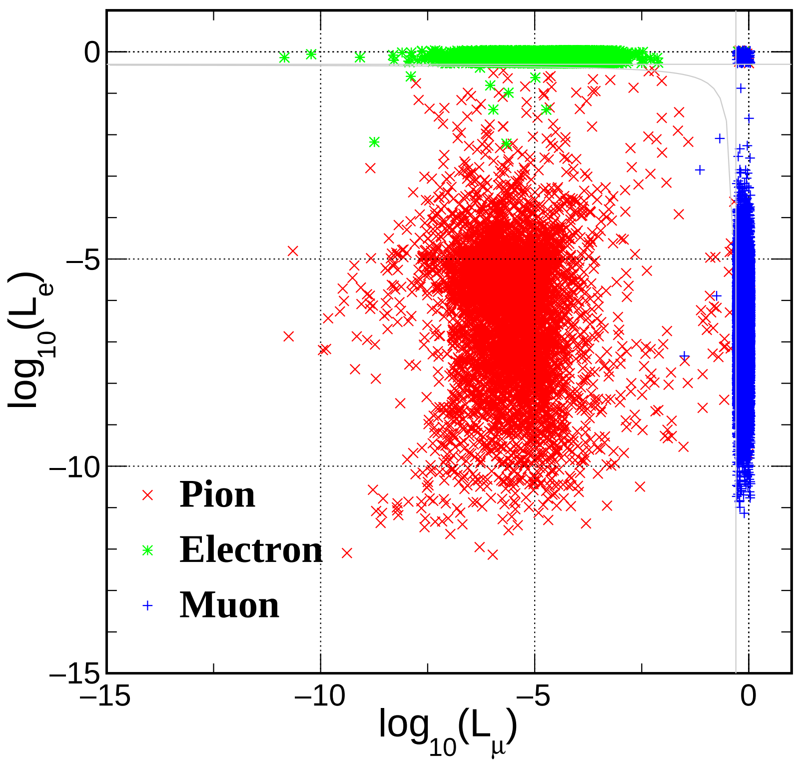

Figure 6.5: The e-likeliness and μ-likeness of e^-, μ^- and π^+ at 40 GeV, grey lines are the cuts for different catalogs in next section

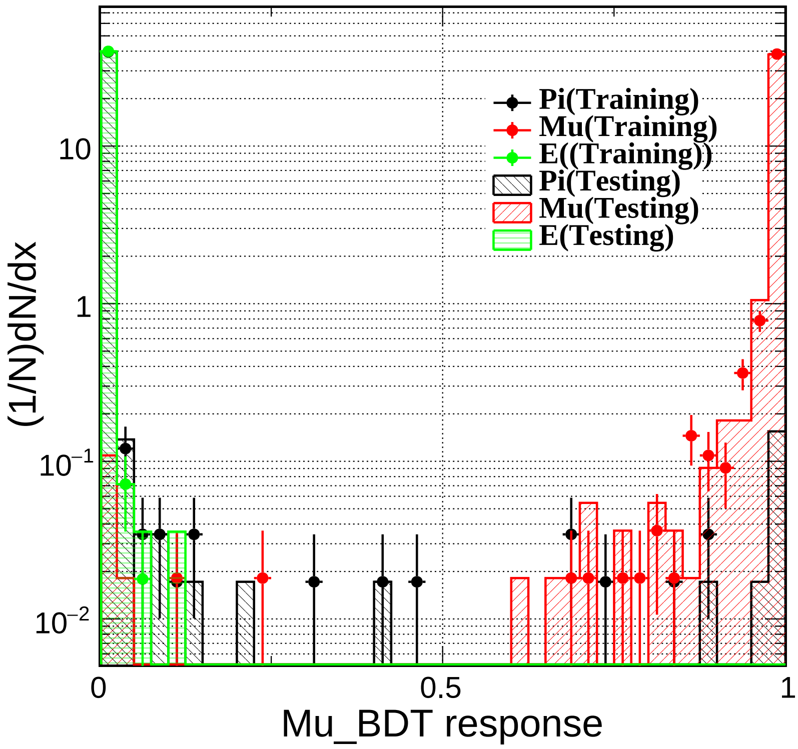

Figure 6.6: Muon BDT response of e^-, μ^- and π^+ at 40GeV (training and test samples)

(θ), the TMVA is trained independently on four subsets:

- **barrel 1**: middle of barrel ($|\cos\theta| < 0.3$),
- **barrel 2**: edge of barrel ($0.3 < |\cos\theta| < 0.7$),
- **overlap**: overlap region of barrel and endcap ($0.7 < |\cos\theta| < 0.8$),
- **endcap**: ($0.8 < |\cos\theta|$ < 0.98).

Take the sample of 40 GeV charged particle as an example, the migration matrix is shown in Table 6.3. Comparing this table to the result of ALEPH for energetic taus[55], the efficiencies are improved, and the mis-identification rates from hadrons to leptons are significantly reduced.

Table 6.3: Migration Matrix at 40 GeV (%)

Type	$e^- like$	$\mu^- like$	$\pi^+ like$
e^-	99.71 ± 0.08	< 0.07	0.21 ± 0.07
μ^-	< 0.07	99.87 ± 0.08	0.05 ± 0.05
π^+	0.14 ± 0.05	0.35 ± 0.08	99.26 ± 0.12

The lepton identification efficiencies (diagonal terms of the migration matrix) at different energies are presented in Figure 6.7 for the different regions. The identification efficiencies saturate at 99.9% for particles with energy higher than 2 GeV. For those with energy lower than 2 GeV, the performance drops significantly, especially in **barrel2** and **overlap** regions. For the overlap region, the complex geometry limits the performance; while for the **barrel2** region, charged particles with Pt < 0.97 GeV cannot reach the barrel, they will eventually hit the endcaps at large incident angle, hence their signal is more difficult to catalog.

Concerning the off-diagonal terms of the migration matrix, the chances of electrons to be mis-identified as muons and pions are negligible ($P^e_\mu, P^e_\pi < 10^{-3}$), the crosstalk rate P^μ_e is observed at even lower level. However, the chances of pions to be mis-identified as leptons (P^π_e, P^π_μ) are of the order of 1% and are energy dependent. In fact, these mis-identifications are mainly induced by the irreducible physics effects: pion decay and π^0 generation via π-nucleon collision. Meanwhile, the muons also have a small chance to be mis-identified as pions at energy smaller than 2 GeV. Figure 6.8 shows the significant crosstalk items (P^π_e, $P^\pi_\mu$and P^μ_π) as a function of the particle energy in the endcap region. The green shaded band indicates the probability of pion decay before reaching the calorimeter, which is roughly comparable with P^π_μ.

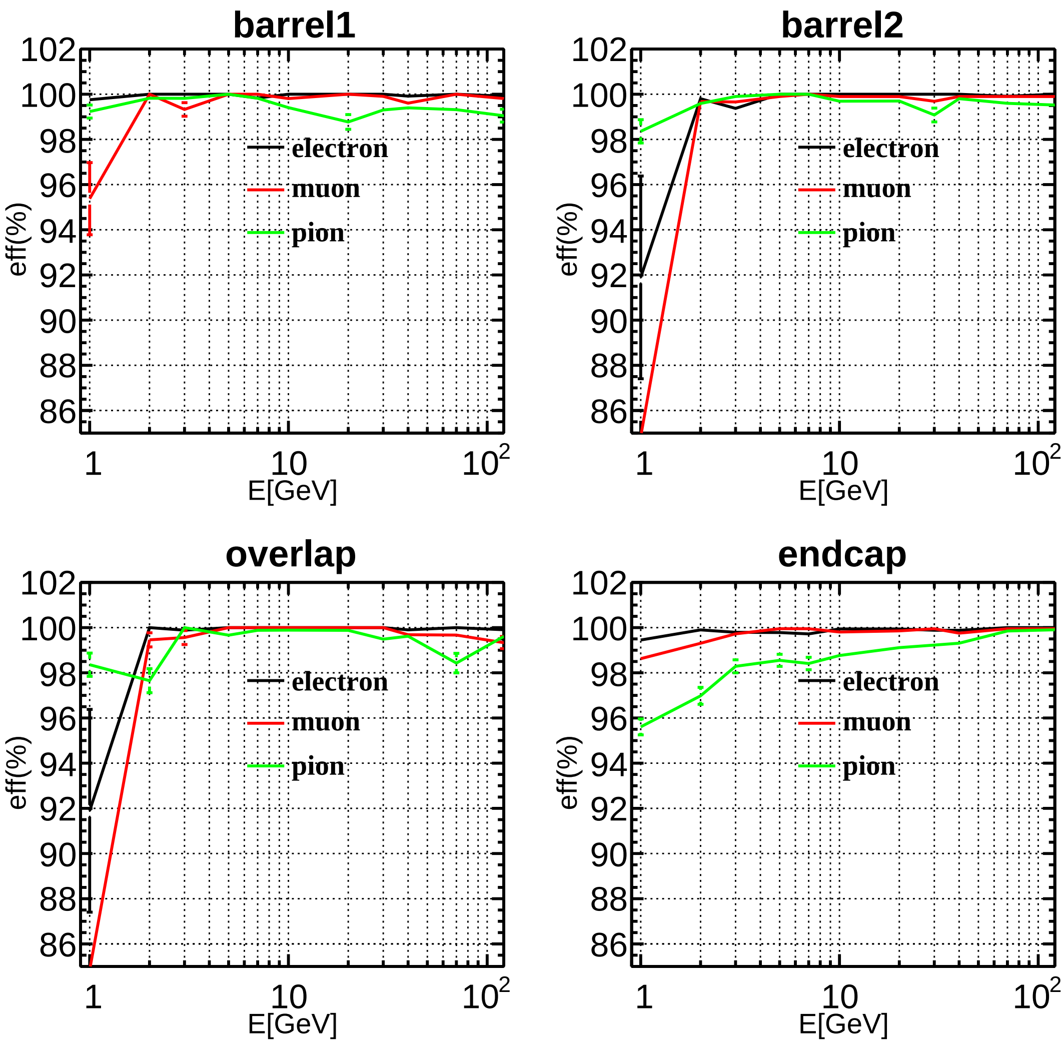

Figure 6.7: The efficiency of lepton identification for e^-, μ^- and π^+ as function of particle energy in the four regions

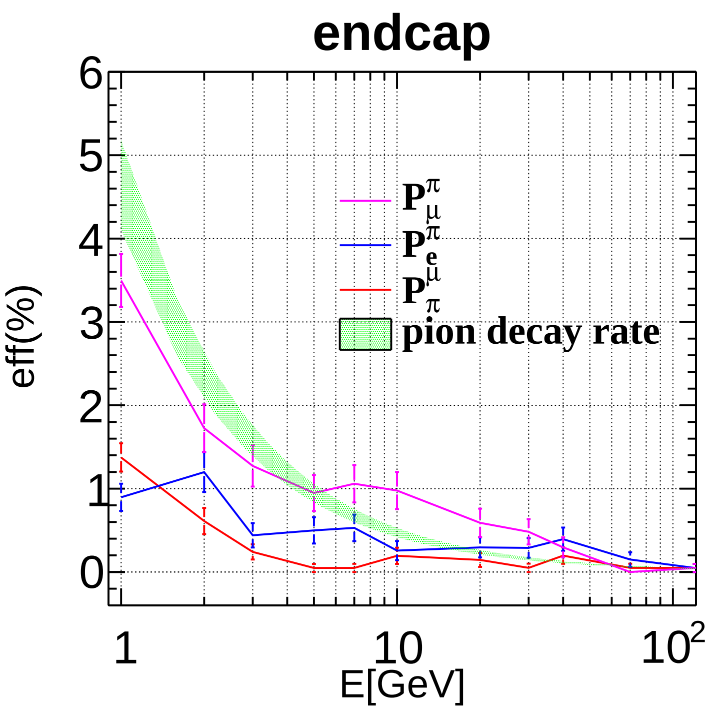

Figure 6.8: The mis-identification rates of lepton identification for μ and π in $\sim$ 5000 events for the endcap region; Pion decay rate band (to account for the polar angle spread) is indicated for comparison

6.4 Lepton identification performance on single particle events for different geometries

The power consumption and electronic cost of the calorimeter system scale with the number of readout channels. It's important to evaluate the physics performance of different calorimeter granularities, at which the LICH performance is analyzed.

The performance is scanned over certain ranges of the following parameters:

- the number of layers in ECAL, taking the value of 20, 26, 30 (total absorber thickness unchanged);
- the number of layers in HCAL: 20, 30, 40, 48 (absorber thickness of each layer unchanged);
- the ECAL cell size = 5×5 mm^2, 10×10 mm^2, 20×20 mm^2, 40×40 mm^2
- HCAL cell size = 10×10 mm^2, 20×20 mm^2, 40×40 mm^2, 60×60 mm^2, 80×80 mm^2

In general, the lepton identification performance is extremely stable over the scanned parameter space. Only for HCAL cell size larger than 60×60 mm^2 or HCAL layer number less than 20, marginal performance degradation is observed: the efficiency of identifying muons degrades by 1-2% for low energy particles ($E \leq 2$ GeV), and the identification efficiency of pion degrades slightly over the full energy range, see Figure 6.9 to Figure 6.12.

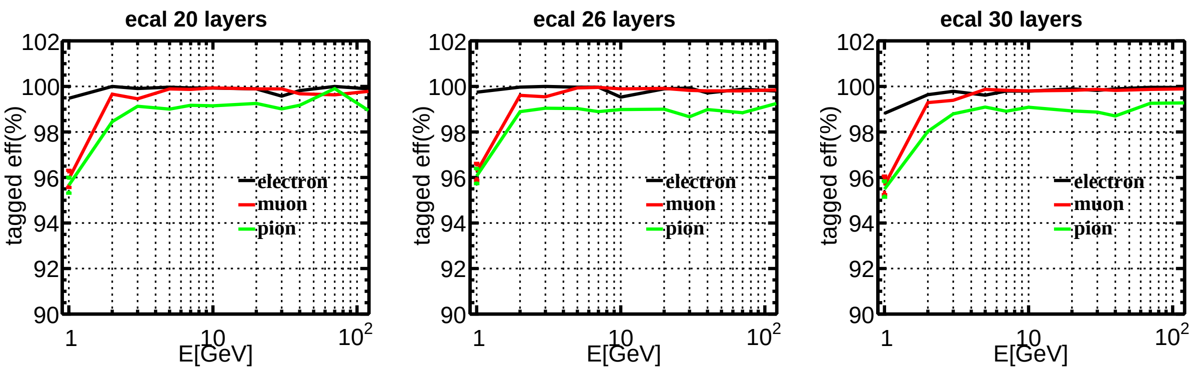

Figure 6.9: The efficiency of lepton identification for different ECAL layer number

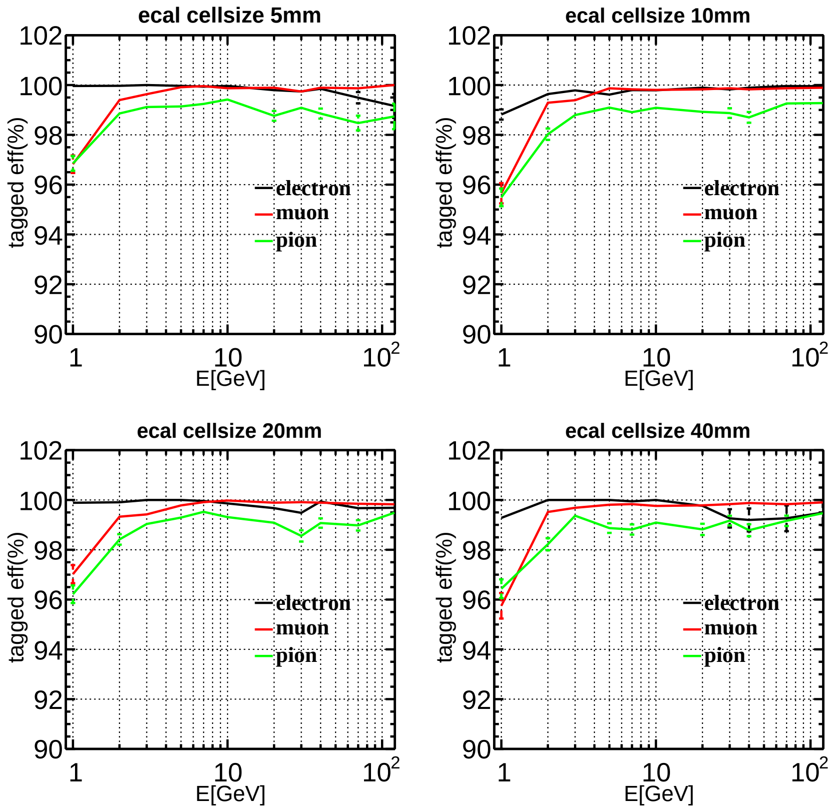

Figure 6.10: The efficiency of lepton identification for different ECAL cell size

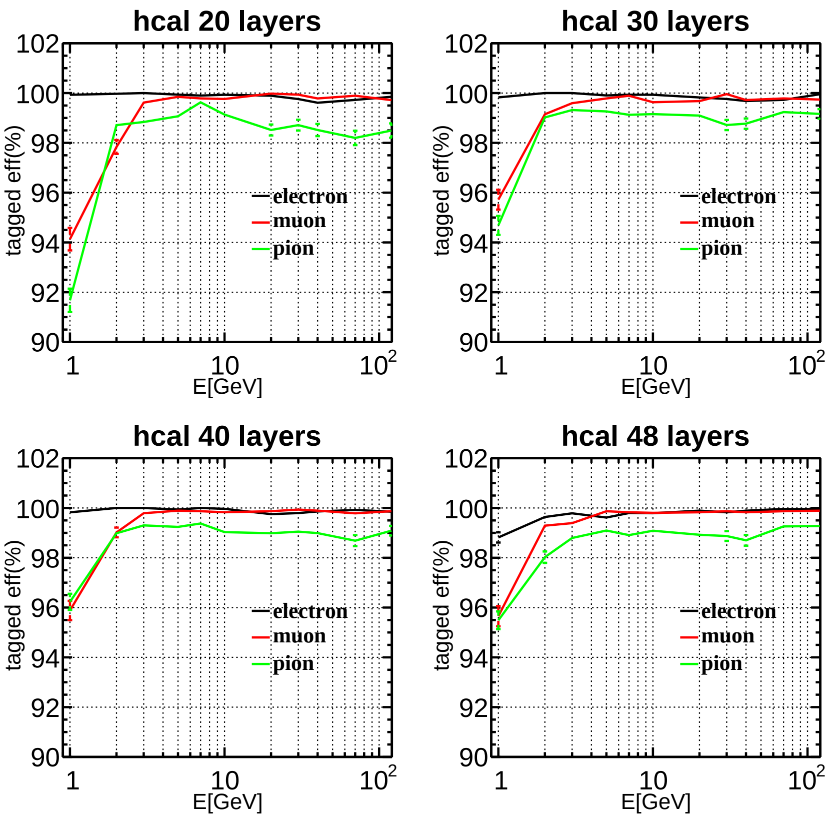

Figure 6.11: The efficiency of lepton identification for different HCAL layer number

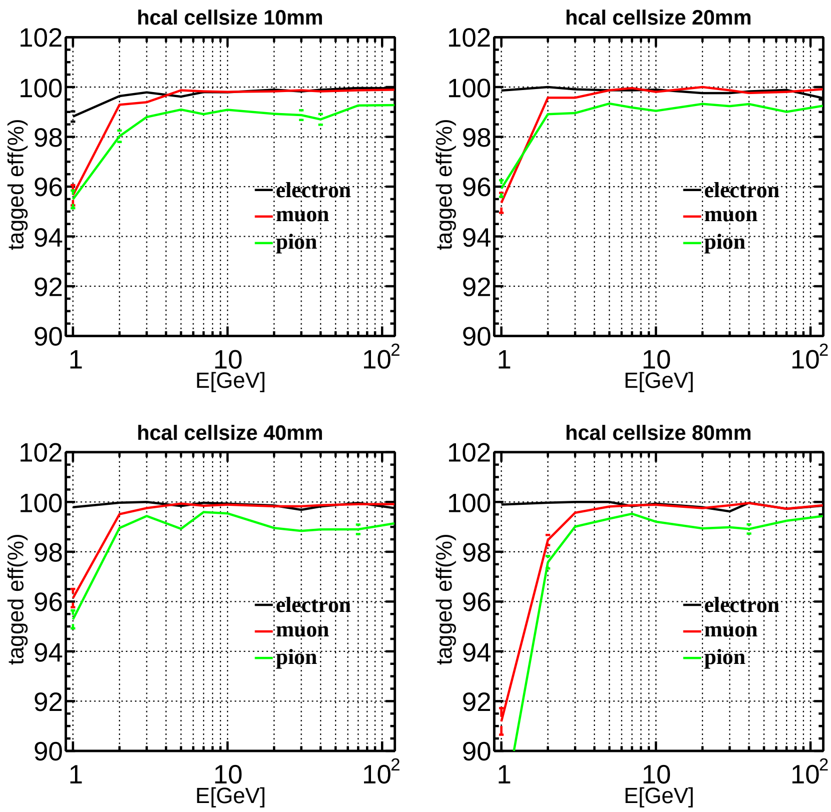

Figure 6.12: The efficiency of lepton identification for different HCAL cell size

6.5 Performance on physics events

The Higgs boson is mainly generated through the Higgsstrahlung process (ZH) and more marginally through vector boson fusion processes at electron-positron Higgs factories. A significant part of the Higgs bosons will be generated together with a pair of leptons (electrons and muons). These leptons are generated from the Z boson decay of the ZH process. For the electrons, they can also be generated together with the Higgs boson in the Z boson fusions events, see Figure 6.13. At the CEPC, 3.6×10^4 $\mu\mu$H events and 3.9×10^4 eeH events are expected at an integrated luminosity of 5 ab^{-1}. In these events, the particles are rather isolated.

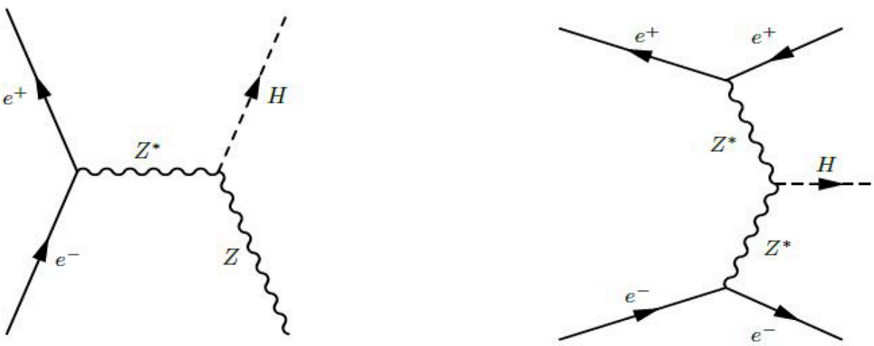

Figure 6.13: Feynman diagrams of major Higgs production with leptons at CEPC: the Higgsstrahlung and ZZ fusion processes.

The eeH and $\mu\mu$H events provide an excellent access to the model-independent measurement to the Higgs boson using the recoil mass method [53]. The recoil mass spectrum of eeH and $\mu\mu$H events is shown in Figure 6.14, which exhibits a high energy tail induced by the radiation effects (ISR, FSR, bremsstrahlung), while in CEPC the beamstrahlung effect is negligible. The bremsstrahlung effects for the muons are significantly smaller than that for the electrons, therefore, it has a higher maximum and a smaller tail.

Figure 6.15 shows the energy spectrum for all the reconstructed charged particles in 10k eeH/$\mu\mu$H events. The leptons could be classified into 2 classes, the initial leptons (those generated together with the Higgs boson) and those generated from the Higgs boson decay cascade. For the eeH events, the energy spectrum of the initial electron exhibits a small peak at low energy, corresponding to the Z fusion events. The precise identification of these initial leptons is the key physics objective for the lepton identification performance of the detector.

Since the lepton identification performance depends on the particle energy, and most of the initial leptons have an energy higher than 20 GeV, we focused on the performance

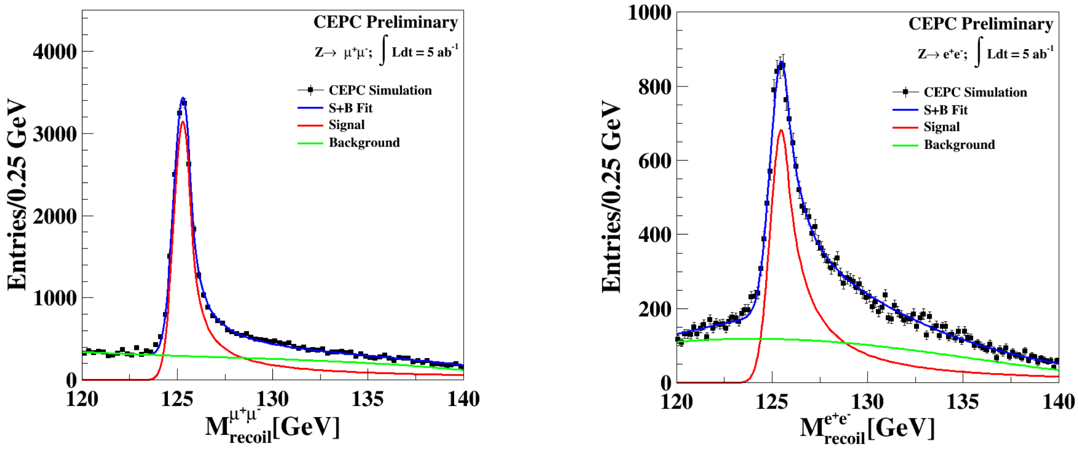

Figure 6.14: The recoil mass spectrum of ee/$\mu\mu$

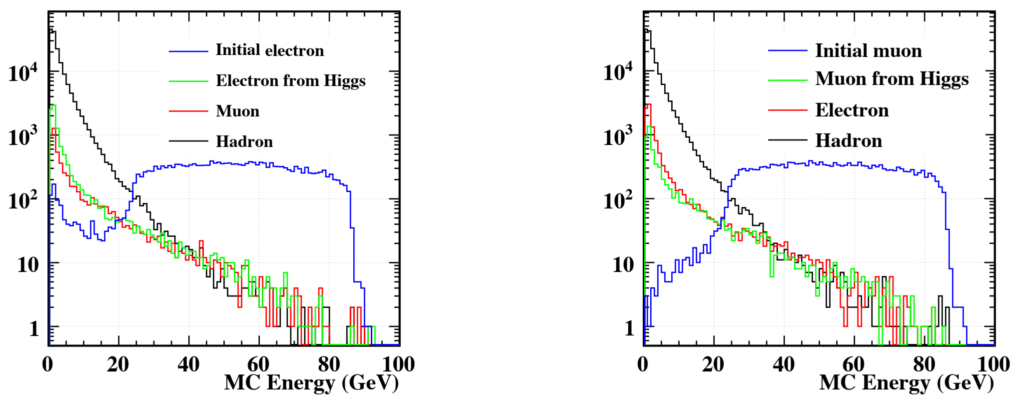

Figure 6.15: Energy Spectrum of charged particles in eeH and $\mu\mu H$ event at 250 GeV center of mass energy, low energy peak in eeH corresponds to the Z fusion events

Table 6.4: $\mu\mu$H/eeH events lepton identification efficiency (ε) and purity (η) (for leptons with erergy > 20GeV)

	Geom 1 (ECAL and HCAL Cell Size 10×10 mm^2)		Geom 2 (ECAL and HCAL Cell Size 20×20 mm^2)	
	$\mu\mu$H	eeH	$\mu\mu$H	eeH
μ definition	L_μ>0.1	L_μ>0.1	L_μ>0.1	L_μ>0.1
e definition	L_e>0.01 L_μ<0.1	L_e>0.001 L_μ<0.1	L_e>0.01 L_μ<0.1	L_e>0.001 L_μ<0.1
ε_e	93.41 ± 0.92	$\mathbf{98.64 \pm 0.08}$	91.60 ± 1.02	$\mathbf{97.89 \pm 0.11}$
η_e	92.02 ± 1.00	$\mathbf{99.74 \pm 0.04}$	89.89 ± 1.10	$\mathbf{99.67 \pm 0.04}$
ε_μ	$\mathbf{99.54 \pm 0.05}$	95.53 ± 0.76	$\mathbf{99.19 \pm 0.06}$	86.48 ± 1.26
η_μ	$\mathbf{99.60 \pm 0.04}$	96.31 ± 0.70	$\mathbf{99.83 \pm 0.03}$	95.38 ± 0.81
ε_{event}	98.53 ± 0.13	97.06 ± 0.19	97.24 ± 0.18	95.40 ± 0.24

study of lepton identification on these high energy particles at detectors with two different sets of calorimeter cell sizes.

The μ-likeliness and e-likeliness of electrons, muons, and pions, for eeH events and $\mu\mu$H events are shown in Figure 6.16 and Figure 6.17. Table 6.4 summarizes the definition of leptons and the corresponding performance under different conditions. The identification efficiencies for the initial leptons are degraded by 1-2% with respect to the single particle case. This degradation is mainly caused by the shower overlap, and is much more significant for electrons as electron showers are much wider than that of muon, leading to a larger chance of overlapping. The electrons in $\mu\mu$H events and vice versa are generated in the Higgs decay. Their identification efficiency and purity still remain at a reasonable level. For charged leptons with energy lower than 20 GeV, the performance degrades by about 10% because of the high statistics of background and the cluster overlap, as shown in Table 6.5. The event identification efficiency, which is defined as the chance of successfully identifying both initial leptons, is presented in the last row of Table 6.4. The event identification efficiencies are roughly the square of the identification efficiency of the initial leptons. Comparing the performance of both geometries, it is shown that when the number of readout channels is reduced by 3/4, the event reconstruction efficiency is degraded by 1.3% and 1.7%, for $\mu\mu$H and eeH events respectively.

6.6 Conclusion

The high granularity calorimeter is a promising technology for detectors in collider facilities of the High Energy Frontiers. It provides good separation between different final state particles, which is essential for the PFA reconstructions. It also records the shower

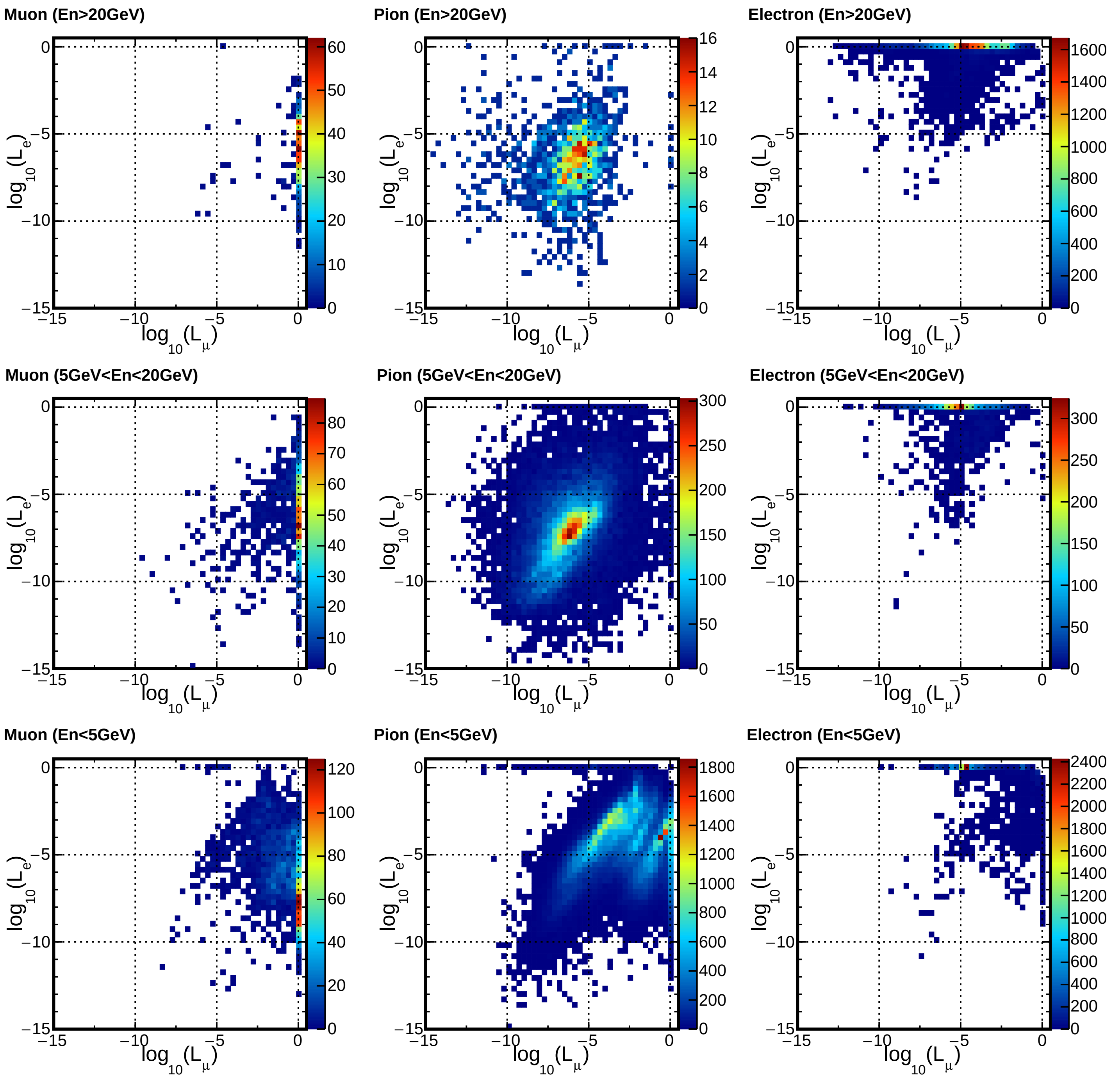

Figure 6.16: e-likelihood and μ-likelihood of charged particles with different energy bins in eeH event

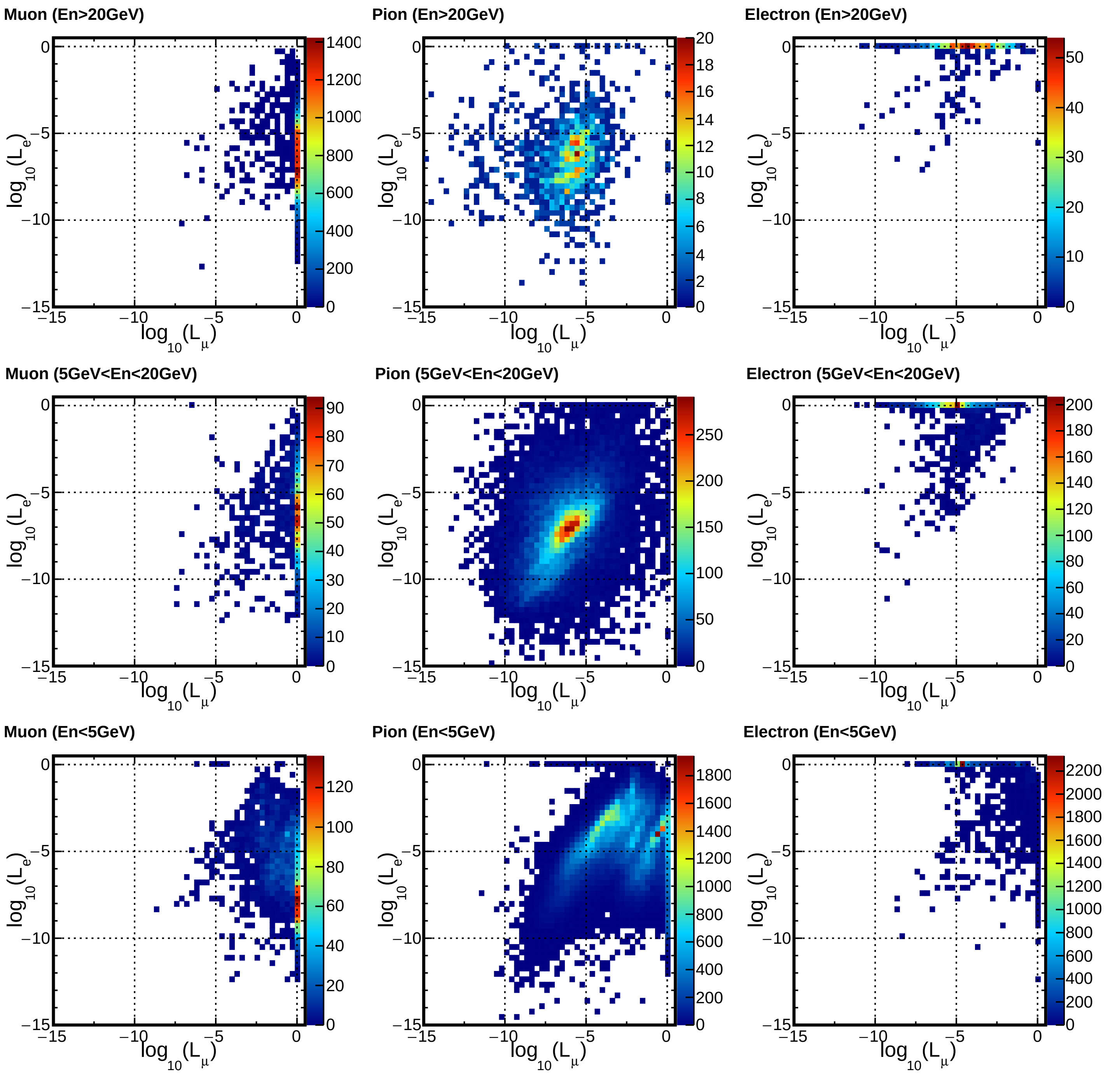

Figure 6.17: e-likelihood and μ-likelihood of charged particles with different energy bins in $\mu\mu$H event

Table 6.5: $\mu\mu$H/eeH events lepton identification efficiency (ε) and purity (η) (for leptons with erergy < 20GeV)

	5GeV < En < 20GeV		En < 5GeV	
	$\mu\mu$H	eeH	$\mu\mu$H	eeH
μ definition	L_μ>0.1	L_μ>0.1	L_μ>0.1	L_μ>0.1
e definition	L_e>0.001 L_μ<0.1	L_e>0.001 L_μ<0.1	L_e>0.001 L_μ<0.1	L_e>0.001 L_μ<0.1
ε_e	91.30 ± 0.71	92.52 ± 0.52	94.86 ± 0.29	95.31 ± 0.27
η_e	70.24 ± 0.92	80.22 ± 0.65	81.90 ± 0.47	79.27 ± 0.47
ε_μ	79.92 ± 0.99	79.89 ± 1.02	60.78 ± 0.95	61.11 ± 0.98
η_μ	82.25 ± 0.96	81.69 ± 0.99	22.73 ± 0.49	22.42 ± 0.50

spatial development and energy profile to an unprecedented level of details, which can be used for the energy measurement and particle identifications.

To exploit the capability of lepton identification with high granularity calorimeters and also to provide a viable toolkit for the future Higgs factories, LICH, a TMVA based lepton identification package dedicated to high granular calorimeter, has been developed. Using mostly the shower description variables extracted from the high granularity calorimeter and also the dE/dx information measured from tracker, LICH calculates the e-likeness and μ-likeness for each individually reconstructed charged particle. Based on these output likelihoods, the leptons can be identified according to different physics requirement.

Applied to single particle samples simulated with the CEPC_v1 detector geometry, the typical identification efficiency for electron and muon is higher than 99.5% for energies higher than 2 GeV. For pions, the efficiency is reaching 98%. These efficiencies are comparable to the performance reached by ALEPH, while the mis-identification rates are significantly improved. Ultimately, the performances are limited by the irreducible confusions, in the sense that the chance for muon to be mis-identified as electron and vice versa is negligible, the mis-identification of pion to muon is dominated by the pion decay.

The tested geometry uses an ultra-high granularity calorimeter: the cell size is 1 by 1 cm^2 and the layer number of ECAL/HCAL is 30/48. In order to reduce the total channel number, LICH is applied to a much more modest granularity, it is found that the lepton identification performance degrades only at particle energies lower than 2 GeV for an HCAL cell size bigger than 60×60 mm^2 or with an HCAL layer number less than 20.

The lepton identification performance of LICH is also tested on the most important physics events at CEPC. In these events, multiple final state particles could be produced in a single collision, the particle identification performance will potentially be degraded by the overlap between nearby particles. The lepton identification on eeH/$\mu\mu$H event at 250 GeV collision energy has been checked. The efficiency for a single lepton identification is consistent with the single particle results. The efficiency of finding two leptons decreases by 1∼2 % when the cell size doubles, which means that the detector needs 2∼4% more statistics in the running. In eeH events, the performance degrades because the clustering algorithm still needs to be optimized.

To conclude, ultra-high granularity calorimeter designed for ILC provides excellent lepton identification ability, for operation close to ZH threshold. It may be a slight overkill for CEPC and a slightly reduced granularity can reach a better compromise. And LICH, the dedicated lepton identification for future e+e- Higgs factory, is prepared.

Chapter 7

Measurement of $H \to \tau\tau$ Branching Ratio

7.1 Introduction

In this chapter, the Higgs boson decaying into tau lepton pairs will be discussed. After τ lepton was discovered in the 1970s at SLAC, its properties have been studied in several experiments and projects. The world average for the τ mass is $1776.86 \pm 0.12 MeV$, and the average for the τ lifetime is $290.3 \pm 0.5 fs$[21]. As the heaviest SM lepton, τ has a larger coupling to Higgs than μ or e, i.e., a larger cross section, which makes $H \to \tau\tau$ channel a tool to test the Higgs properties and search for new physics at higher scales.

7.1.1 τ physics

QCD The mass of τ is heavy enough to decay to hadrons, this turns out to be useful for studying strong interaction effects at low energies. This makes the τ useful as a probe for QCD and many electroweak phenomena. Decays including strangeness enable measurements of the mass of the strange quark and the CKM matrix element V_{us}[56].

The polarity and spin are measured in hadronic decay with a better precision than in the case of leptonic decays. In leptonic decays, one cannot reconstruct the direction of the polarimeter vector, the polarization measurement cannot be performed with the full sensitivity of the polarimeter. The polarization vector can be reconstructed for the hadronic decays in one or two pions and so the angle between the polarization vector and the τ direction can be measured. A measurement of the distribution will then allow conclusions on the τ polarization.

Leptonic decay The leptonic decays of the τ lepton probe the structure of the weak currents and the universality of their couplings to gauge boson. One of the basic ideas in the SM is that all lepton doublets have identical couplings to the Z and W bosons. Comparing the measured decay widths of leptonic or semi-leptonic decays which only differ in the τ decay, one can test experimentally that the interaction is indeed the same, i.e., that $g_e = g_\mu = g_\tau \equiv g$[57, 58].

New physics The τ is also an important probe to the new physics, by observing the coupling constants deviation from the Standard Model prediction or exploring lepton flavor violating τ decay. A few samples are heavy scalar resonances decaying to a τ lepton pair and charged Higgs bosons decays predicted in the MSSM[59]. In the HH searches, the H$\to \tau\tau$ decay channel is one of the most sensitive to both SM and many BSM production modes[60]. Besides, differences in the τ^+ and τ^- lifetimes would indicate the violation of CPT[61].

B physics The τ lepton could also be used also a probe of some particular process where heavy meson decays into final states containing τ leptons[62]. Decays such as $B^- \to \tau^-\bar{\nu}_\tau$, $B \to D^*\tau^-\bar{\nu}_\tau$, $B_c^- \to \tau^-\bar{\nu}_\tau$ or $D_s^- \to \tau^-\bar{\nu}_\tau$ involve the heaviest elementary fermions that can be directly produced at flavor factories, providing important information about the underlying dynamics mediating these processes.

Higgs measurement[63, 64] The τ is the heaviest SM lepton, which leads to a large coupling to Higgs, i.e. a significant fraction of the SM Higgs boson decays into $\tau\tau$ final states. This makes it possible to measure $g(H \to \tau\tau)$ with a better accuracy.

As one of the most important channel in the future e^+e^- Higgs factory, $H \to \tau\tau$ channel performance also provides evidence for detector optimization and the PFA developments. The requirement to separate photons and hadrons decayed from τ should be satisfied by a relatively high granularity and an efficient PFA. On the other hand, to distinguish different τ decay modes, the PFA should provide reasonable particle identification.

7.1.2 τ decay modes

The leptonic decay of τ lepton follows $\tau^- \to \nu_\tau l^- \bar{\nu}_l$, with $l = e, \mu$. These two neutrinos make it difficult to reconstruct the τ mass. In the hadronic decays, only one neutrino is involved, its direction can thus be reconstructed by measuring all other decay products. This is not used in this thesis, but can be a continuetion to the studies. The hadronic decay of τ lepton can be classified in:

- final state without photon: $\tau^- \to \nu_\tau h^-$, with $h = \pi, K$

- final state with two photons dominated by ρ production: $\tau^- \rightarrow \nu_\tau \rho^- \rightarrow \nu_\tau \pi^- \pi^0$ and $\pi^0 \rightarrow \gamma\gamma$
- final state with four photons dominated by a_1^- production: $\tau^- \rightarrow \nu_\tau a_1^- \rightarrow \nu_\tau \pi^- 2\pi^0$ and $\pi^0 \rightarrow \gamma\gamma$.

The branching ratio of these dominant τ decay modes [21] is shown in table 7.1.

Table 7.1: τ^- decay modes and branching fraction (%). The first five decay modes with only one track in final state are called "1-prong", and the decay modes with three track in final state are "3-prong" decay

$e^- \bar{\nu}_e \nu_\tau$	17.82 ± 0.04
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.39 ± 0.04
$\pi^- \nu_\tau$	10.82 ± 0.05
$\pi^- \pi^0 \nu_\tau$	25.49 ± 0.09
$\pi^- 2\pi^0 \nu_\tau$	9.26 ± 0.10
$\pi^- \pi^+ \pi^- \nu_\tau$	9.31 ± 0.05
others	< 10

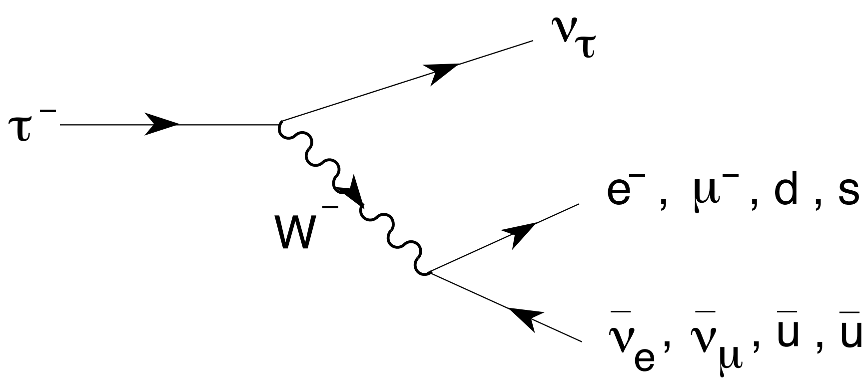

Figure 7.1: Feynman diagram for $\tau \rightarrow \nu_\tau X$ decay modes

The topology of τ in the high granular detector is shown in the event display in Figure 7.2.

As shown in the event display, the τ decay in high energy colliders is tightly collimated and low multiplicity, which provide excellent signatures to probe.

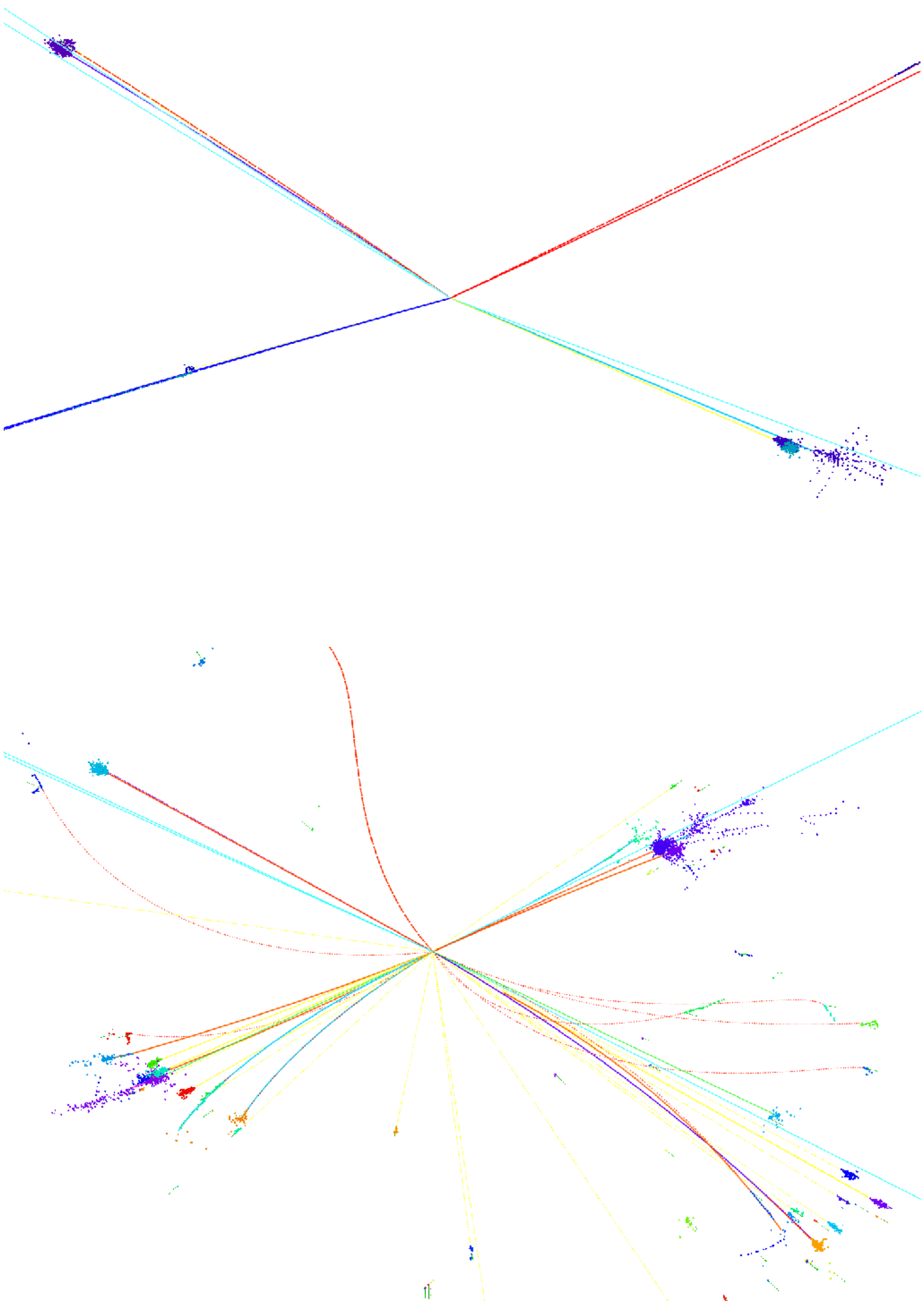

Figure 7.2: Event display of a $\mu\mu\tau\tau$ event with one $\tau \to e^-\bar{\nu}_e\nu_\tau$ and the other $\tau \to \pi^-\nu_\tau$ (up) and a $qq\tau\tau$ event with one $\tau \to e^-\bar{\nu}_e\nu_\tau$ and the other $\tau \to \pi^-\pi^+\pi^-\nu_\tau$(down) at CEPC (reconstructed with Arbor)

7.1.3 Measurements and precisions

The deviation of coupling constants from the Standard Model prediction to new physics beyond the Standard Model depends on the new physics model, and this deviation is estimated to be 1% level by many models proposed. At the LHC, the process of the Higgs boson decaying into tau pairs will be measured using proton–proton collision. This decay has been studied by the ATLAS and the CMS experiments, who reported a combined signal yield consistent with the Standard Model expectation, with a combined observed significance at the level of 6σ. With an uncertanty of 9% at HL-LHC (300 fb^{-1}), the LHC experiment may not have sufficient sensitivity for new physics described in the previous section.

On the other hand, previous studies of the Higgs boson decaying into tau pairs at the ILC show that the measurement can be of the order of a few percent[65] and that the measurement at the ILC plays a crucial role after the LHC experiments. However, these studies did not take into account some of the relevant background processes (such as $\nu\nu$H), nor based on the jet clustering algorithm. Therefore in this thesis, this channel is studied independently from the jet clustering while taking into account the whole SM background.

7.2 Samples

The CEPC luminosity is supposed to be 5000 $\mathbf{fb^{-1}}$. For the ZH signal, the cross section for different Z decay modes is summarized in table Chapter 2, as well as the branching ratio of Higgs decaying to $\tau\tau$. All the samples in this chapter are generated by the MC generator Whizard, version 1.95[66]. The detector used in the simulation is the CEPC detector.

The cross section shown here gives the first view to the efficiency and purity that need to be achieved. Taking qqH channel, for example, the statistics for signal qq$\tau\tau$ and backgrounds are 44872 and 488 million respectively. Using the simple expression of accuracy as $\sqrt{S+B}/S$, if the efficiency to identify qq$\tau\tau$ event is 80%, the background should be suppressed by 99.98% in order to achieve the 1% accuracy.

The studies on Higgs decaying into the τ channel are treated individually for each Z decaying channel, in order to distinguish the signal with the different type of backgrounds. The selection of events is done in two steps:

- **Pre-selection** Due to the limited computing resource, the inclusive ZH events, and SM categories background events are filtered by some preselection using MC truth information to simplify the samples. The excellent performance of PFA ensures that this preselection would not lose information. The information used in the preselection is different for each Z decaying channel, including the number of muons ($N_{\mu^{+/-}}$), the recoil mass of the muon pair (M_{recoil}), the invariant mass of the muon pair ($M_{invariant}$), the missing mass ($M_{missing}$), the total visible mass (M_{tot}), the transverse momentum (p_T), the visible energy (E_{vis}), the number of charged particles (N_{charge}).

- τ **tagging** The τ tagging process is applied using the topology of events. The impact parameters are used in order to deduce the statistics of signal and backgrounds.

A successful reconstruction of the τ lepton is not a trivial task, for the τ lepton could be generated with various different event topology, and it has diverse decay final states. In the e^+e^- collision environment, we summarize the τ events into two categories according to the event topology, in which the reconstruction algorithm and performances have been studied separately.

7.3 Leptonic channels

The first category is the leptonic one, where no physics objects, or only lepton / photon / missing energy is generated together with the τ candidates.[1] These events include, for example:

- ZH, Z→ l^+l^- / $\nu\nu$, $H \rightarrow \tau\tau$ events; golden channel for $g(H\tau\tau)$ measurements
- ZZ, l^+l^- / $\nu\nu$ / $\tau\tau$ events
- WW events with $l\nu\tau\nu$ final states.
- $Z \rightarrow \tau\tau$ events at Z pole operation.

In these events, the global multiplicity is limited while the additional physics objects, if they exists, are easy to identify. A successful identification of these events relies highly on the reconstruction of photons and charged hadrons. In the following section, the physics performances of τ reconstruction at $\mu\mu H$ and $\nu\nu H$ channel are shown as well as their Br($H \rightarrow \tau\tau$) measurement.

[1]The charge is ignored for event classifications.

7.3.1 $Z \to \mu\mu$

The easiest channel to study is the $\mu\mu H$ channel since the two muons are easy to be vetoed by calculating their invariant mass. According to the different behavior of $\mu\mu H$ and backgrounds shown in Figure 7.3, the preselection applied to select $\mu\mu H$ are:

- $N_{\mu^+} > 1$, $N_{\mu^-} > 1$
- $110GeV < M_{recoil} < 180GeV$
- $40GeV < M_{invariant} < 180GeV$

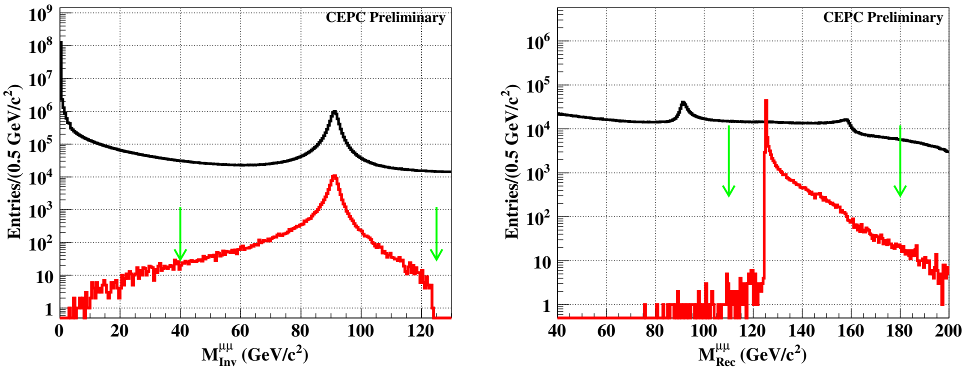

Figure 7.3: Distribution of invariant mass and recoil mass (MC information) for $\mu\mu H$ and backgrounds at $\sqrt{s} = 250$GeV, the red/black line is for signal($\mu\mu H$)/background(inclusive), the green arrows indicates the cuts applied in the preselection. The invariant mass of $\mu\mu H$ peaks at the mass of Z while only the ZZ background have this peak. The recoil mass of $\mu\mu H$ peaks at the mass of H while ZZ background peaks at the mass of Z

Thank the excellent efficiency and purity of the muon identification, the efficiency of this preselection can achieve 97.68%, while most of the SM backgrounds are vetoed except for $\mu\mu$ (3.51% remaining).

Most of the decaying modes of τ are with one or three tracks and an even number of photons, as can be seen in Table 7.1, this is the main idea in the τ tagging. From the decay modes, the topology of τs is simpler than jets, which provides the way to distinguish τ events from the others. The steps for di-τ events tagging are:

- Veto the μs decayed from Z by choosing the μ pair with invariant mass closest to Z mass
- Find the leading track among the remaining particles and collect the tracks and photons close to this track (< 1 rad, to be grouped in region A), and their numbers are noted as NTrkA and NPhA.

- Collect the rest tracks and photons and group them in region B with their numbers noted as NTrkB and NPhB.

- Get the angle between the leading tracks in region A or B and the furthest track in this region, noted as Cone$_{T-T}$(A/B).

- Cone$_{T-P}$(A/B) is the angle between the leading tracks in region A or B and the furthest photon in this region.

- Cone$_{P-P}$(A/B), the angle between the leading photon in a region and the furthest photon in this region.

The distributions of these numbers in τ events and other decay channels of Higgs is shown in Figure 7.4 and the cuts of NTrk and NPh are chosen to be less than 6 and less than 7.

Table 7.2: Cut Flow of MC sample for $\mu\mu H \rightarrow \tau\tau$ selection on signal and inclusive SM backgrounds

	$\mu\mu H\tau\tau$	$\mu\mu H$ inclusive bkg	ZZ	WW	single$W$	single$Z$	$2f$
total generated	2292	33557	5711445	44180832	15361538	7809747	418595861
after preselection	2246	32894	122674	223691	0	86568	1075886
$N_{Trk}(A/B) < 6$ & $N_{Ph}(A/B) < 7$	2219	1039	2559	352	0	9397	25583
BDT>0.78	2135	885	484	24	0	157	161
efficiency	93.15%	2.63%	<0.01%	<0.01%	<0.01%	<0.01%	<0.01%

After the cut of the number of tracks and photons, these parameters are trained in TMVA and optimized to the signal significance giving the BDT cut to 0.78, the cut flow is summarized in Table 7.2, the efficiency of the signal after training is 93%. The correlation matrix and overtraining check are shown in Figure 7.5 and Figure 7.6.

However, the channels such as Higgs decaying into W and W leptonic decay are the main backgrounds after the selections. This is due to the topologies of these events are similar to our signal.

By looking at the starting points for the tracks, those stemming from τ decays are further away from the vertex than the others. From the sum of transverse and longitudinal impact parameters (D0 / Z0[2]) of the two leading tracks in regions A and B normalized by their uncertainty $\sigma_{D0/Z0}$, a "pull" can be defined as: $D0^2/\sigma_{D0}^2 + Z0^2/\sigma_{Z0}^2$, since $D0$ and $Z0$ are comparable in CEPC detector, the pull are simplified as $D0^2 + Z0^2$. The pull

[2] The impact parameter D0 is the signed distance from the origin to the point of closest approach in the $r - \phi(x - y)$ plane. The impact parameter Z0 is the Z position of the perigee.

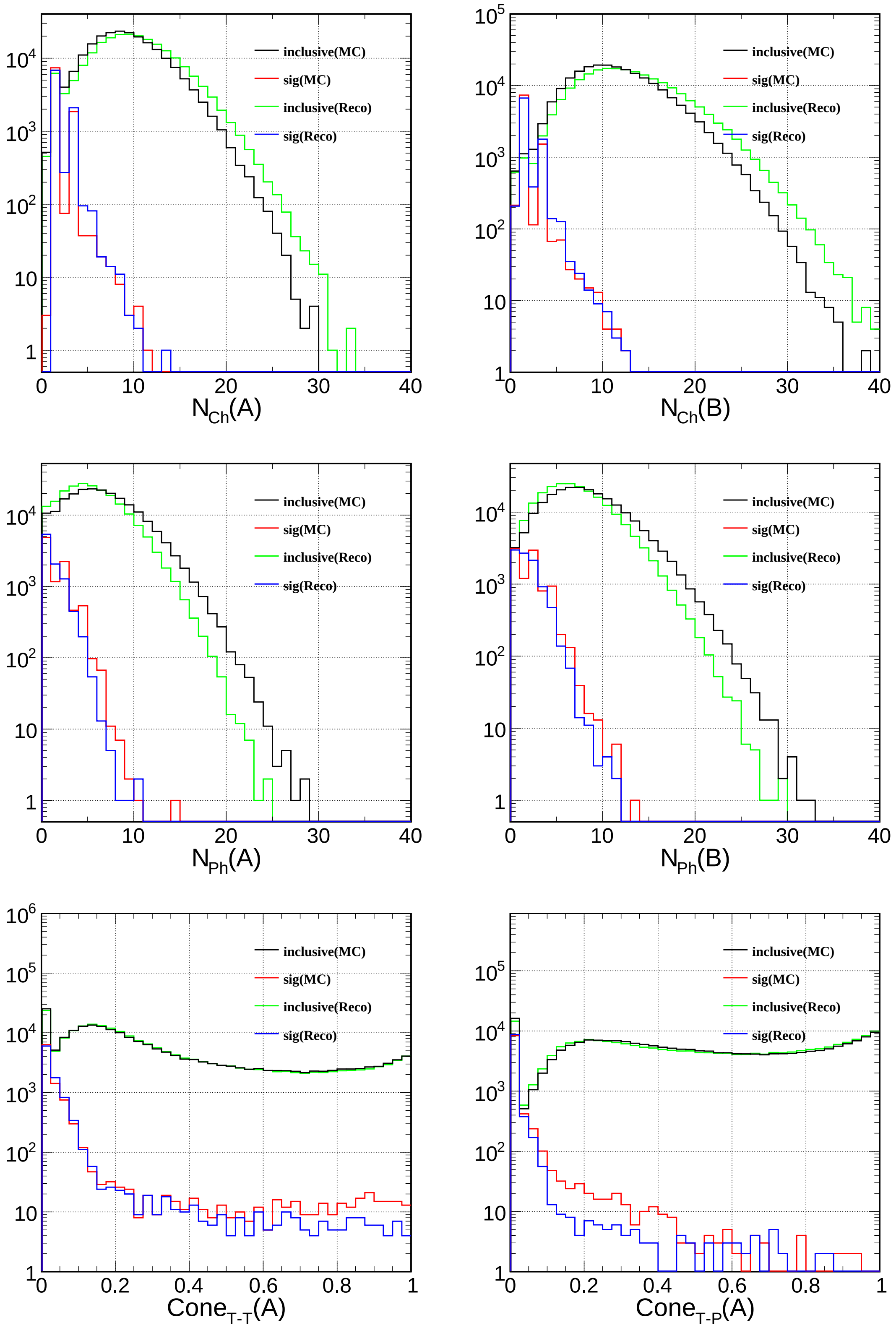

Figure 7.4: Distribution of number of tracks and photons, the angle between track to track, track to photon, or photon to photon in the two opposite regions A and B. The black/red line represents the MC information of the inclusive $\mu\mu H$ backgrounds / signal ($\mu\mu H \to \tau\tau$), the green/blue line is for the reconstructed information.

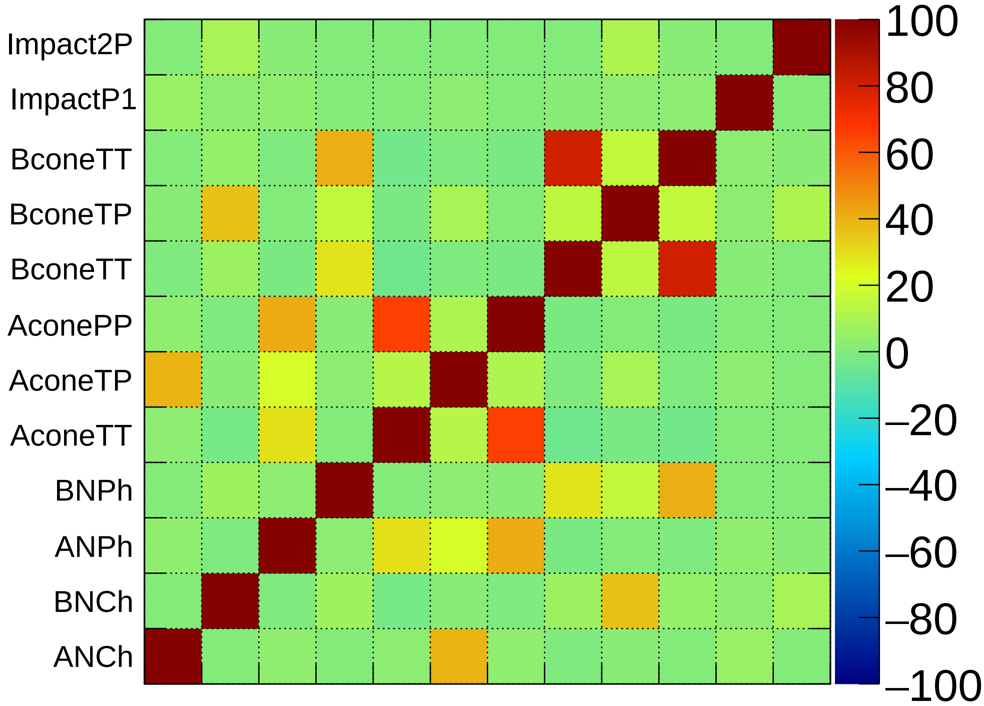

Figure 7.5: The correlation matrix of all the variables

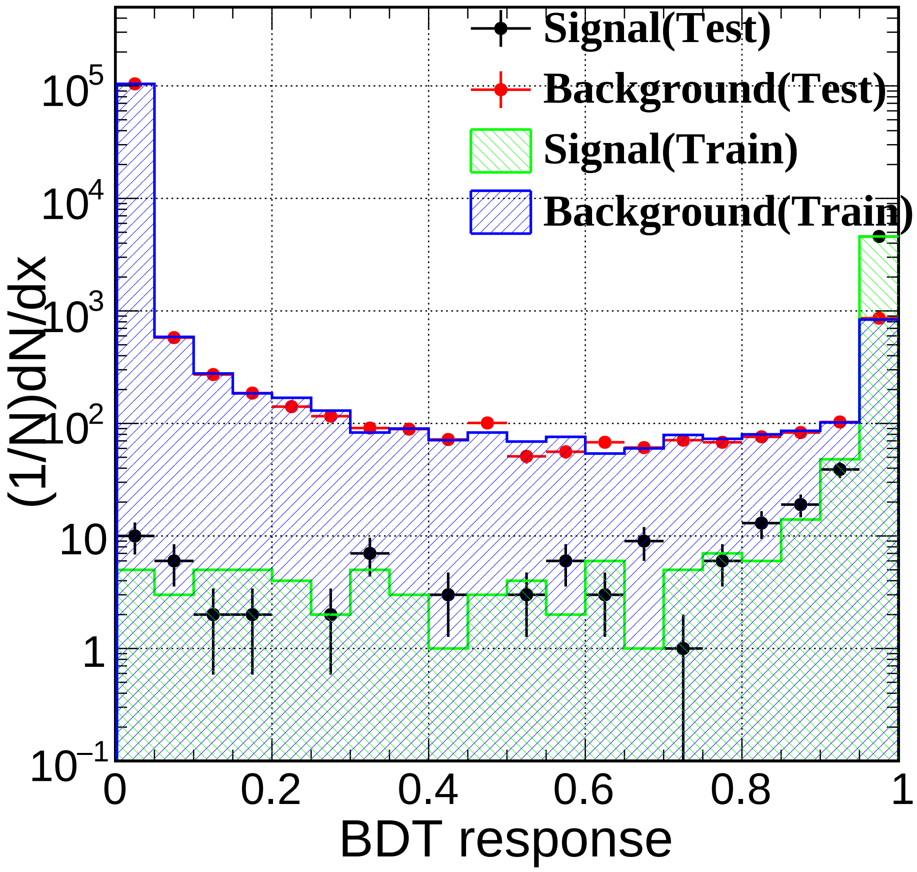

Figure 7.6: BDT response of di-τ finding

distribution is shown in Figure 7.7 for signal and SM inclusive background with a fit.

The branching ratio $Br(H \to \tau\tau)$ can be calculated from the fitted signal event number S, the total event number T and previous selection efficiency ε, as $Br = S/(\varepsilon \cdot T)$, to be 6.40 ± 0.18. The expected accuracy $\sigma \times BR = \delta(S)/S$ to be 2.68%, where the $\delta(S)$ is the fitted signal event number error.

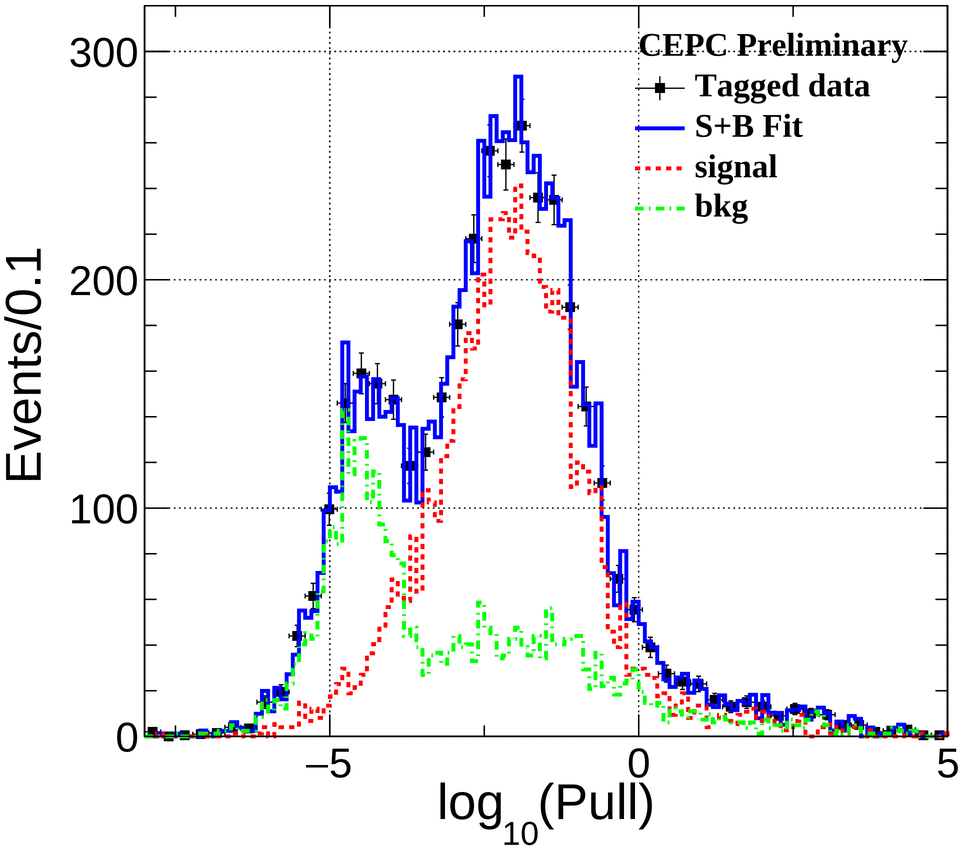

Figure 7.7: Fit of the sum of D0^2 and Z0^2 of the leading tracks of two cones with SM background included

Assuming that the efficiency of $\tau\tau$ event tagging is the same for $\mu\mu H$ and eeH events, the accuracy for the eeH event can be extrapolated. The difference between this two channel is that the efficiency for preselection is not the same, as shown in Table 7.3. The extrapolated accuracy or eeH event is deduced to be 2.72%.

Table 7.3: Preselection efficiency for eeH selection on signal and inclusive SM backgrounds

	eeH	ZZ	WW	single$W$	single Z	$2f$
total generated	38357	5711445	44180832	15361538	7809747	418595861
after preselection	37901	4075	4072	256892	561237	5278241

7.3.2 $Z \to \nu\nu$

According to the different behavior of $\nu\nu$H and backgrounds, the cut flow of the preselection for $\nu\nu$H events is:

- $65GeV < M_{missing} < 225GeV$
- $M_{total} > 50GeV$
- $10GeV < p_T < 100GeV$

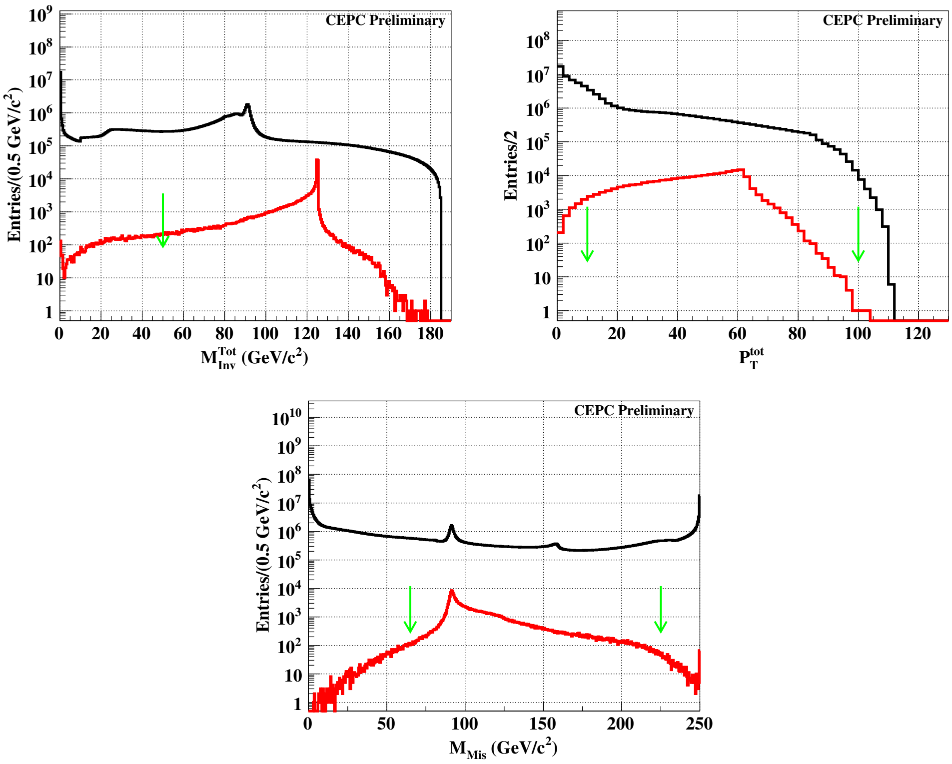

Figure 7.8: Distribution of total invariant mass M_{Inv}^{tot}, transverse momentum P_T^{tot}, and missing mass M_{Mis} for $\nu\nu$H and backgrounds at $\sqrt{s}$ = 250GeV, the red/black line is for signal($\nu\nu$ H)/background(inclusive), the green arrows indicates the cuts applied in the preselection.

However, a bias exists on the different signal channel in this cut flow, which leads to a

1.7% degradation of $BR(H \to \tau\tau)$ and the final result needs to be corrected according to this number.

The procedure of τ tagging in $Z \to \nu\nu$ event is similar to the one in $Z \to \mu\mu$, but without the step to veto the μ pair. However, there exists a huge irreducible background coming from WW and W$\to \nu\tau$, whose impact parameters are not distinguishable, as shown in Figure 7.9. Therefore the only statistic result is deduced in this channel by ignoring the error of the fraction of signal and background.

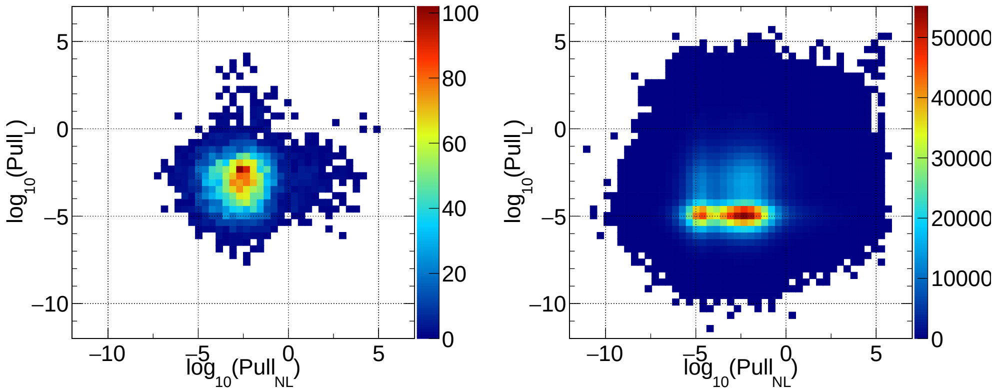

Figure 7.9: Pull of the leading track and the next to leading track for $\nu\nu$H and backgrounds.

The efficiency of tagging after TMVA training is 95%, and the τ event number is calculated from the statistics, as shown in Table 7.4. The branching ratio $Br(H \to \tau\tau)$ can be calculated from the statistics result and previous selection efficiency to be 6.19 ± 0.27, as well as the expected accuracy to be 4.29%.

Table 7.4: Cut Flow of MC sample for $\nu\nu$H$\to \tau\tau$ selection on signal and inclusive SM backgrounds

	$\nu\nu$H$\tau\tau$	$\nu\nu$H inclusive bkg	ZZ	WW	singleW	single Z	$2f$
total generated	15497	231670	5711445	44180832	17361538	7809747	418595861
after preselection	9434	214830	1239457	7463105	3327803	956694	12826280
$N_{Trk}(A/B) < 6$ & $N_{Ph}(A/B) < 7$	9260	8858	24760	1354852	17389	676185	1535029
BDT > 0.78	8836	6587	15450	89729	1355	10739	11243
efficiency	57.02%	2.84%	0.27%	0.20%	<0.01%	0.14%	<0.01%

7.4 Hadronic channel, $Z \to qq$

The second catalog is the hadronic one, where the τ lepton(s) are always observed with jets. For instance, we have:

- ZH, $Z \to qq$, $H \to \tau\tau$
- $ZZ \to qq\tau\tau$
- $WW \to qql\tau$
- ZH, $Z \to qq$, $H \to WW \to l\nu\tau\nu$

The most difficult channel is Z decaying to quarks since these quarks cannot be vetoed from the invariant mass without jet clustering.

The preselection applied to choose the qqH events is:

- $E_{visible} > 100GeV$
- $N_{charge} > 8$
- $P_t < 93GeV$
- $M_{Mis} < 120GeV$

Since the background is still too large, a second preselection is applied to choose the qqH$\to$ $\tau\tau$ events is:

- $115GeV < E_{visible} < 245GeV$
- $M_{Mis} > 2GeV$

The distribution of these variables for preselection is shown in Figure 7.11.

Since the qqH process is more complex than $\mu\mu$H and $\nu\nu$H, the preselection is not that powerful as the previous ones. Keeping the preselection efficiency high leads to nearly half of ZZ and WW semi-leptonic decay remaining. That's a huge number of events to study, therefore the backgrounds are not analyzed in the whole sample but on smaller statistics (10k per sub channel) and scaled to 5 ab^{-1}.

After the preselection, the tagging method is no longer for di-τ but to tag the τ jets in the whole space in an event. The steps are:

- Find tracks with energy higher than a defined E_{min} as the seed

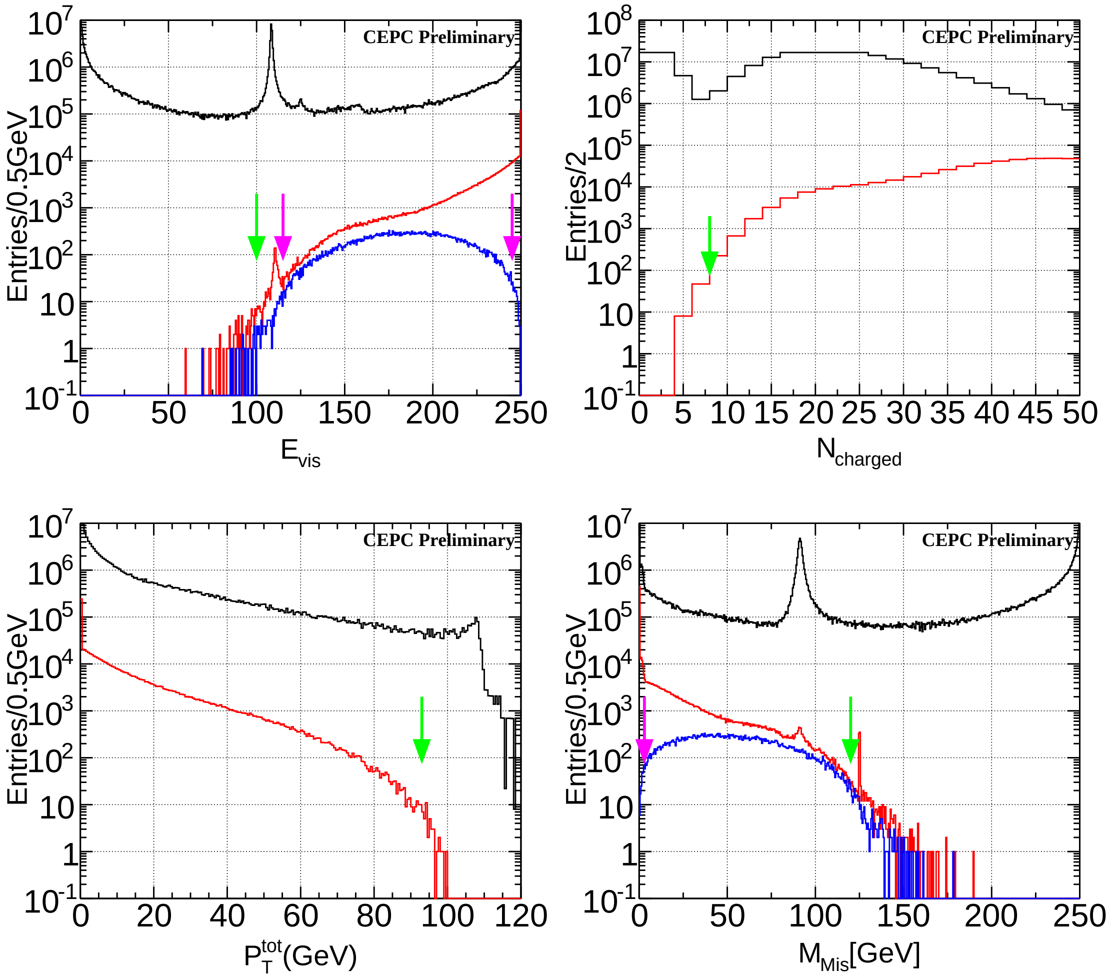

Figure 7.10: Distribution of total visible energy $E_{visible}$, number of charged particles N_{charge}, total visible energy E_{vis}, transverse momentum P_T^{tot}, and missing mass M_{Mis} for qqH$\tau\tau$, qqH and backgrounds at $\sqrt{s}$ = 250GeV, the blue/red/black line is for signal (qqH)/background(inclusive). The green arrows indicates the cuts applied in the first preselection and the pink arrows indicate the second preselection.

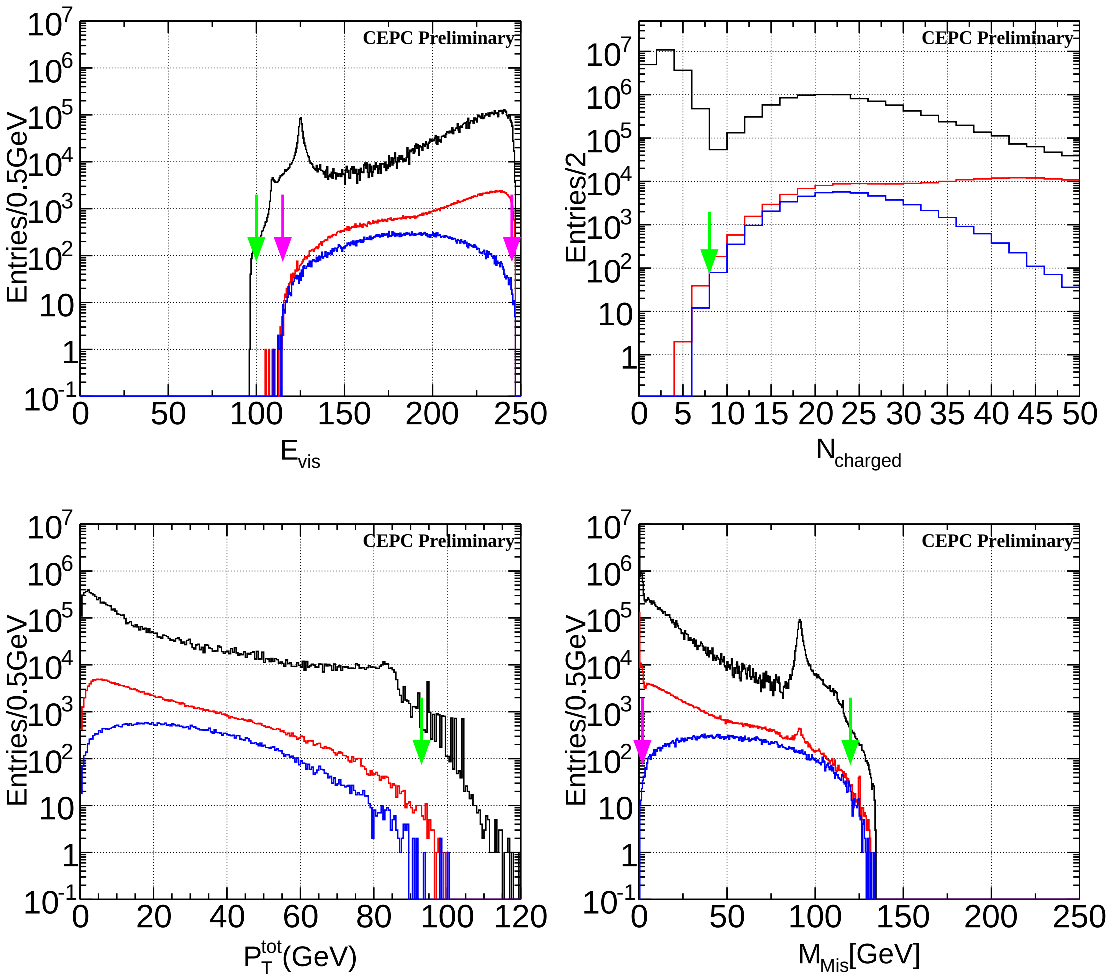

Figure 7.11: Distribution of total visible energy $E_{visible}$, number of charged particles N_{charge}, total visible energy E_{vis}, transverse momentum P_T^{tot}, and missing mass M_{Mis} for qqH$\tau\tau$, qqH and backgrounds at $\sqrt{s} = 250$GeV, with the other cuts applied.

- Collect tracks and photons within an angle ConeA
- Calculate invariant mass with these particles
- Calculate the D0 and Z0 of the leading track
- Calculate the energy in a larger cone ConeB around the seed.

The cut of τ tagging is:

- Number of tracks/photons smaller than 6/8
- Energy proportion in the smaller cone larger than R_{En}
- Invariant mass of the $\tau\tau$ system larger than M_{min} GeV and smaller than M_{max} GeV
- Invariant mass of the qq system (the particles except for τs) smaller than M_{qq} GeV.

Here the parameters E_min, ConeA, ConeB, R_{En}, M_{min} and M_{max} are optimized to the value $\epsilon \cdot p$, where ϵ is the efficiency of finding an opposite charged τ pair in qq$\tau\tau$ events and p is the probability of tagging a opposite charged τ pair in the backgrounds. The value of these parameters are: E_{min} = 1.5 GeV, ConeA = 0.15 rad, ConeB = 0.45 rad, M_{min} = 0.2 GeV, M_{max} = 2.0 GeV, R_{En} = 0.92, the optimized $\epsilon \cdot p$ is 56%. However, this is a rough optimization without background normalization taken into account.

After these cuts, the remaining τs in an event is collected and the two leading energetic ones with opposite charge are chosen to calculate the invariant mass of the di-τ, as shown in Figure 7.12. The distribution of each type of background in Figure 7.13 shows that the 2f background is reduced in this step, as well as the events with "fake" taus reconstructed.

The events with at least a pair of τs and the invariant mass in a range of (20, 120 GeV) are chosen as a Higgs decaying to the $\tau\tau$ event. The particles except for these have been chosen to form the two leading energetic ones with opposite charge are used to get the invariant mass of the qq system and the cut of $70 < M_{qq} < 105 GeV$ is chosen as the selection of signal, as shown in Figure 7.14. In Figure 7.15, it is shown that the ZH background and WW background can be reduced, where the invariant mass of qq leads to a Higgs mass or it is a flat distribution. The ZZ background is still an important one since the invariant mass of qq is also peaking at Z mass.

The recoil mass of the qq system is used to reduce the ZZ backgrounds, as shown in Figure 7.16 and Figure 7.17, the background $ZZ \rightarrow qq\tau\tau$ are reduced because the recoil mass of the qq leads to the mass of Z.

The cut chain is summarized in Table 7.5 and the efficiency for the τ events tagging is

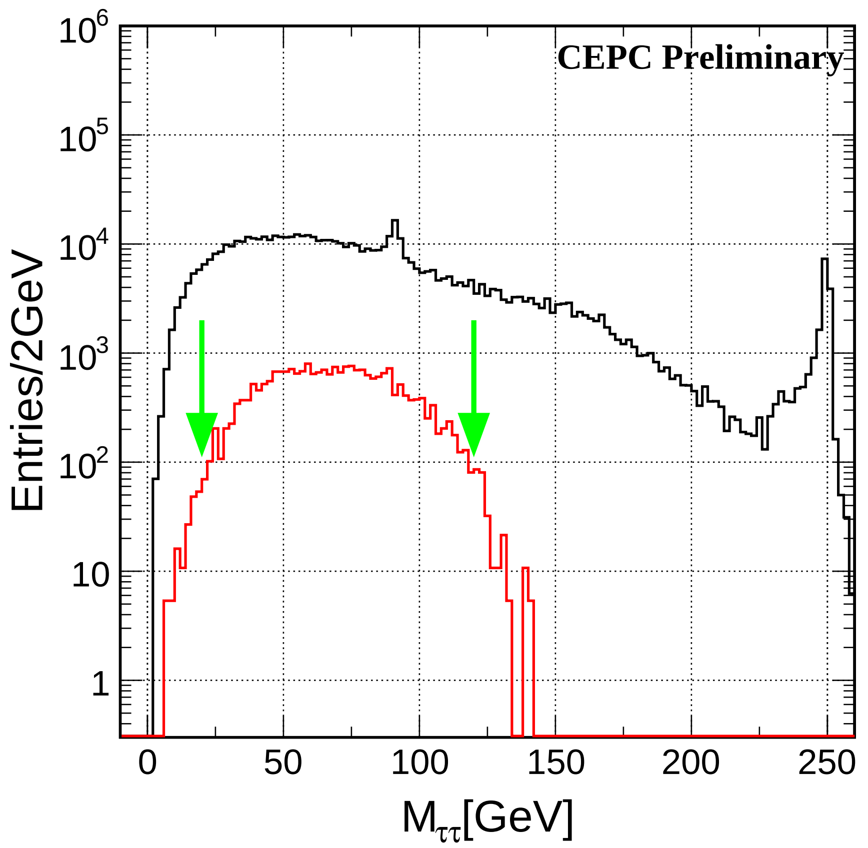

Figure 7.12: Distribution of the invariant mass of the di-τ, $M_{\tau^+\tau^-}$ for qqH$\tau\tau$, and backgrounds at $\sqrt{s} = 250$GeV, the red/black line is for signal ($qqH\tau\tau$)/background(inclusive). The arrows indicates the cuts applied.

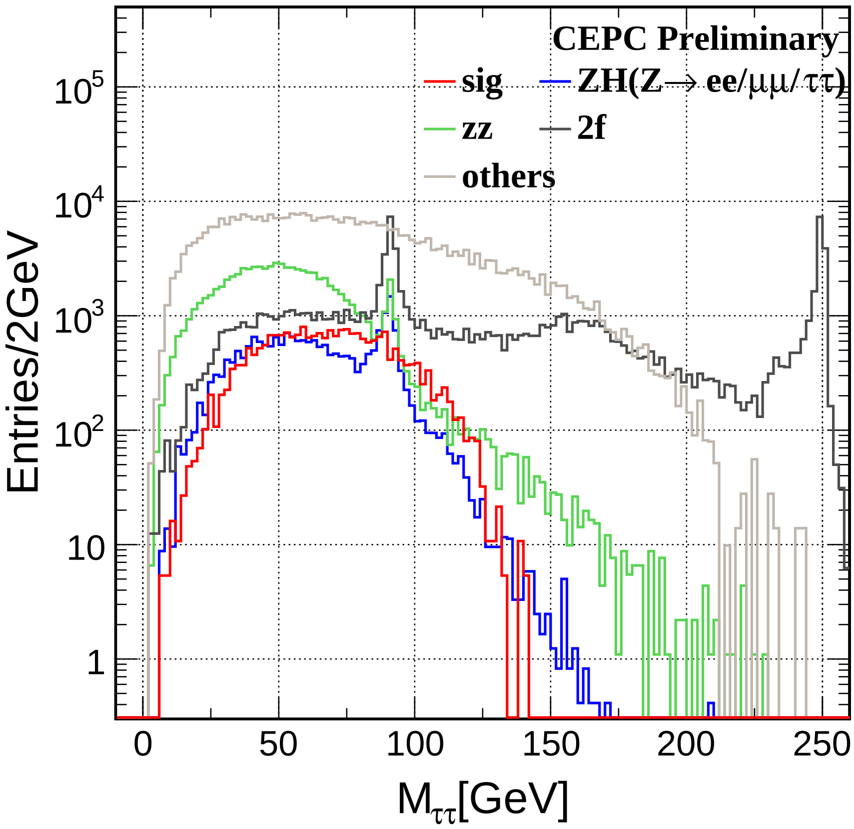

Figure 7.13: Distribution of the invariant mass of the di-τ, $M_{\tau^+\tau^-}$ for qqH$\tau\tau$, and each backgrounds at $\sqrt{s} = 250$GeV .

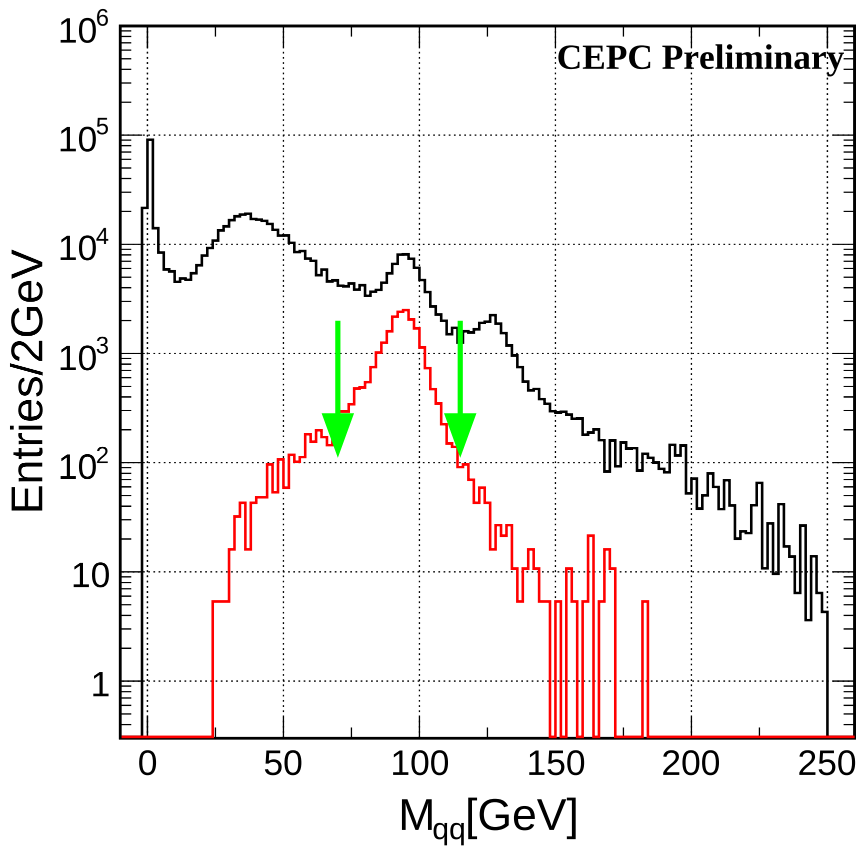

Figure 7.14: Distribution of the invariant mass of the qq, M_{qq} for qqH$\tau\tau$ and backgrounds at $\sqrt{s} = 250$GeV, the red/black line is for signal ($qqH\tau\tau$)/background(inclusive). The arrows indicates the cuts applied.

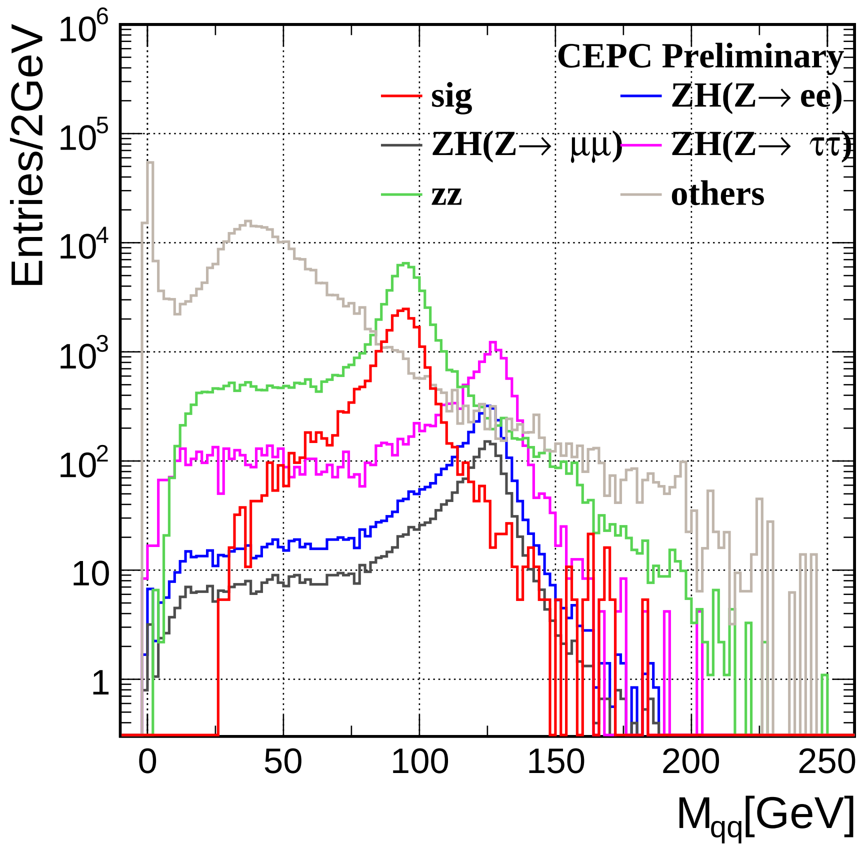

Figure 7.15: Distribution of the invariant mass of the qq, M_{qq} for qqH$\tau\tau$ and each backgrounds at $\sqrt{s} = 250$GeV after the previous cuts.

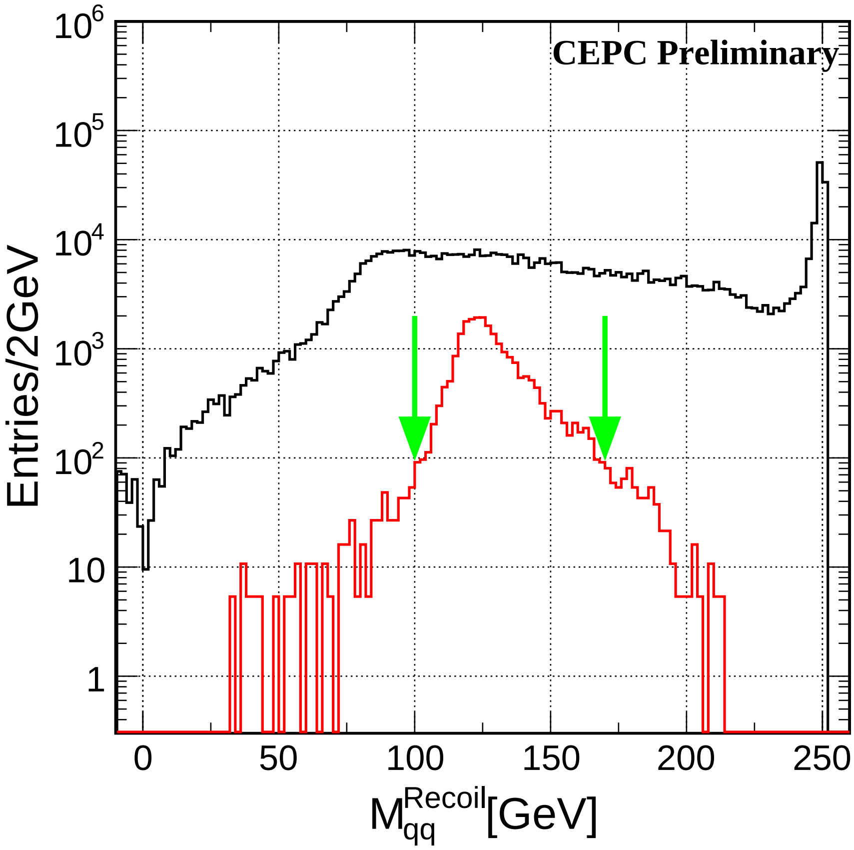

Figure 7.16: Distribution of the recoil mass of the qq, M_{qq}^{recoil} for qqH$\tau\tau$ and backgrounds at $\sqrt{s} = 250$GeV, the red/black line is for signal ($qqH\tau\tau$)/background(inclusive). The arrows indicates the cuts applied.

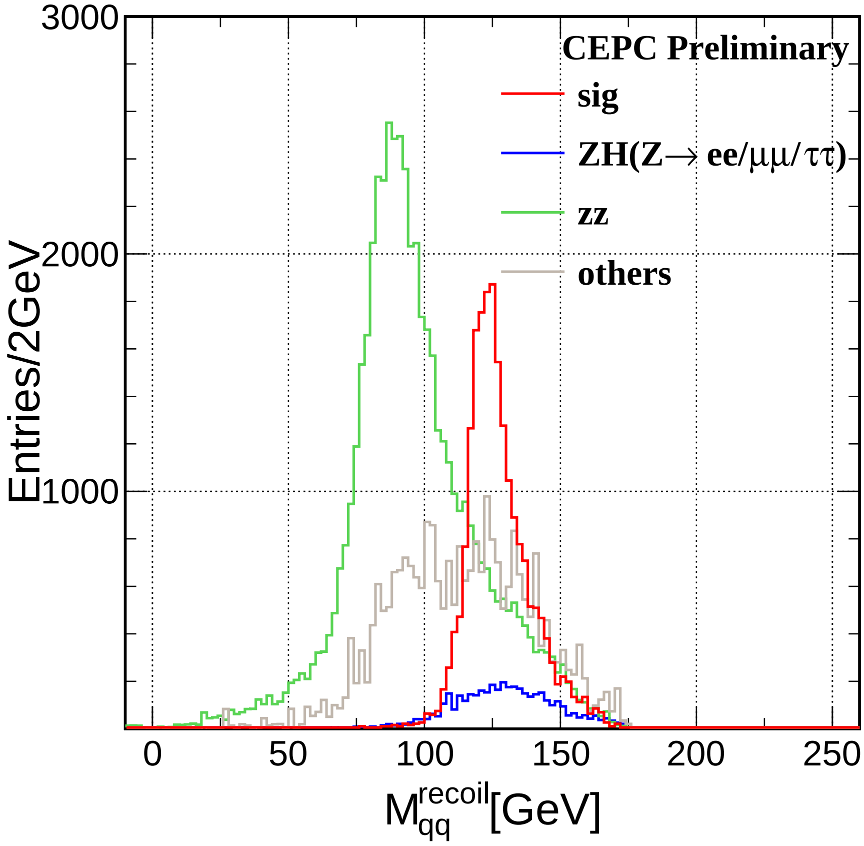

Figure 7.17: Distribution of the recoil mass of the qq, M_{qq}^{recoil} for qqH$\tau\tau$ and each backgrounds at $\sqrt{s} = 250$GeV after the previous cuts

49.97%. The finding efficiency of each τ can be expressed as: N_{found}/N_{truth}, where N_{found} is the number of tagged τ's and the leading track is close to a τ decayed track from the MC information, while N_{truth} is the number of MC τ's. Here the efficiency in the $qqH\tau\tau$ channel is 70.7%. In the similar way the purity defined as N_{found}/N_{total} where N_{total} is the total number of tagged τ's, in the $qqH\tau\tau$ channel, is 70.1%.

Table 7.5: Cut Flow of MC sample for qqH→ $\tau\tau$ selection on signal and inclusive SM backgrounds

	$qqH\tau\tau$	qqH inclusive bkg	ZH inclusive bkg	ZZ	WW	single$W$	single$Z$	$2f$
total generated (scaled to 5 ab^{-1})	45597	678158	357249	5711445	44180832	17361538	7809747	418595861
1st preselection	45465	677854	310245	5039286	42425195	1267564	1398362	148401031
2nd preselection	45145	174650	226059	293306	12452091	125735	117306	547402
$N_{\tau^+} > 0, N_{\tau^-} > 0$	24674	7342	33721	93955	723989	33887	54386	103642
$20GeV < M_{\tau^+\tau^-}$ < 120GeV	24284	6290	32344	88245	597480	24927	36039	56615
$70GeV < M_{qq}$ <110GeV	22937	2103	4887	65625	21718	738	1893	556
$100GeV < M_{qq}^{Rec}$ <170GeV	22703	2045	4524	23789	13154	315	306	193
efficiency	49.97%	0.31%	1.26%	0.41%	0.04%	<0.01%	<0.01%	$<$ 0.01%

From the table, the background of WW and ZZ are more important than the others, this is because of the sub channel of their semi-leptonic decay with q jets and leptons or even τs, which is irreducible. The statistics of signal and the main backgrounds are shown in Figure 7.18. The branching ratio $Br(H \to \tau\tau)$ can be calculated from the fit result and previous selection efficiency to be 6.25 ± 0.04, and the expected accuracy to be 1.30%.

7.5 Combined Results

To conclude, the τ reconstruction at the CEPC is currently catagorized into leptonic and hadronic events and reconstructed using different strategies and τ finding algorithms. In the leptonic events, where the τ lepton is generated only in association with leptons, photons or missing energy, the τ events identification relies strongly on a successful reconstruction of the photons and charged hadrons.

In the hadronic events, it is more difficult to suppress the background, for further study, the correlation with other channels might be applied.

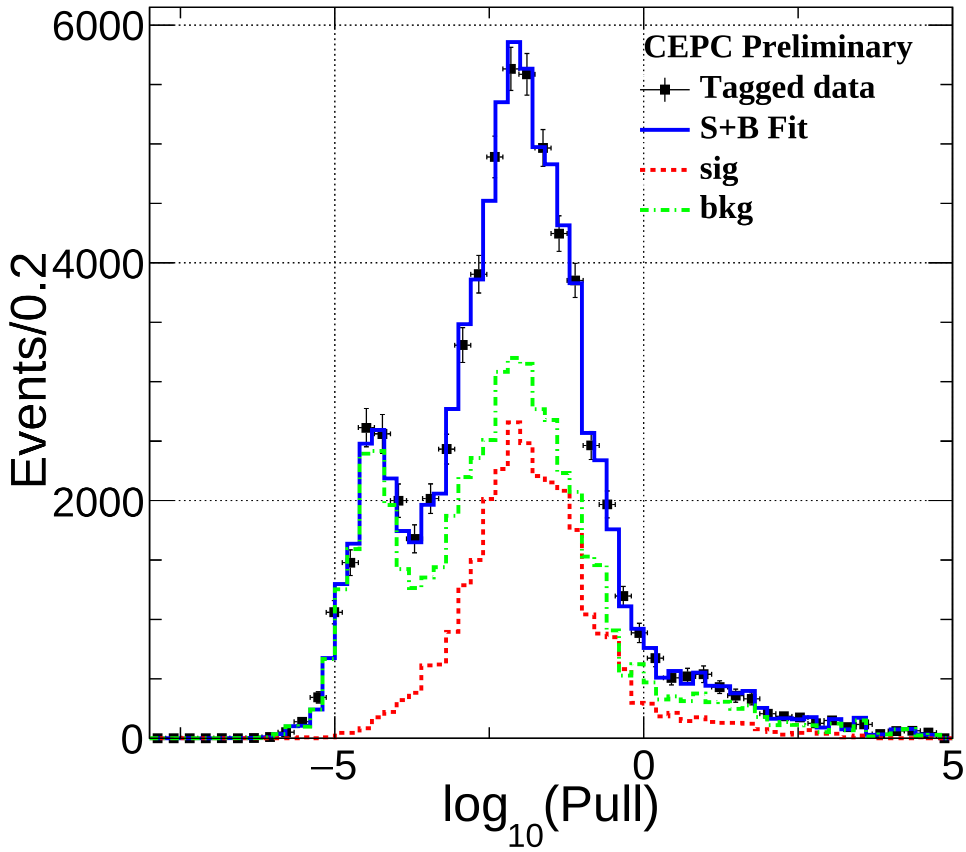

Figure 7.18: Fit of the sum of $D0^2$ and $Z0^2$ of the leading tracks of two cones with SM background included

With these channels analyzed and the cross section of Higgs decaying to $\tau\tau$ can be summarized as in Table 7.6

Table 7.6: Combined cross section

	BR (H→ $\tau\tau$)	δ ($\sigma\times$ BR)/($\sigma\times$ BR)
$\mu\mu$H	6.40	2.68%
eeH(extrapolated)	6.37	2.72%
$\nu\nu$H	6.26	4.38%
qqH	6.23	0.93%
combined	6.28	0.81%

In both cases, a precise reconstruction of the impact parameter is essential for the τ events identification, as shown in the figures, the statistics can be fitted only if the position resolution is good enough to distinguish the two peaks for τs and backgrounds.

7.6 Extrapolating in ILC

The cross section for three polarization scenarios in ILC at 250GeV is shown in Chapter 2.

Comparing these cross sections with the cross section at CEPC as shown in previous section, a simple extrapolation can be done as in Table 7.7. The assumption here is that the efficiency for each signal and background stays the same for ILC and CEPC.

Table 7.7: Extrapolated accuracy δ ($\sigma\times$ BR)/($\sigma\times$ BR) in ILC 250GeV (2000 fb^{-1})

	CEPC	ILC(L)	ILC(R)
Luminosity(ab^{-1})	5	2	2
Polarization(e^-, e^+)	-	(0.8, -0.3)	(-0.8, 0.3)
Total Higgs	1.06M	0.60M	0.40M
Accuracy(%)	0.81	1.13	1.22

7.7 Discussion

In this chapter, different channels with Higgs decaying into $\tau\tau$ at CEPC have been studied and the combined accuracy is reaching 1% level. This result is also extrapolated to ILC and also gives the reasonable accuracy.

This accuracy has still space to be improved. One choice is to use the collinear approximation to recover the momentum of neutrino(s) from τ. This method needs to assume that τ decay products almost flight back-to-back. The collinear approximation will help to reconstruct the invariant mass of tau pair system and its comparison to the Higgs mass could be a powerful variable to suppress ZZ/WW backgrounds with τ final states.

Another method is to fully reconstruct hadronically decaying τ momenta by making use of the interaction point position, the impact parameters of the τ decay products, and the transverse momentum of the Z boson recoiling against the $\tau\tau$ system [67]. Since more than 60% of τs decays into hadrons, this method will help to improve the performance of these channels.

Besides, a jet clustering algorithm can be applied in the qqH channel in order to suppress the 2f backgrounds with jets.

However, this study here is based on a perfect vertex detector, the resolution is not taken into account. Since the result was obtained from the impact parameter, the influence of the vertex detector design to the performance should be studied in the future.

Chapter 8

Conclusion

This thesis covers the aspects of detector optimization, the particle identification, and tau analysis, in the concept of CEPC, but not limited to CEPC. The requirement on accuracy to 1% order by the new physics appreciates the future e^+e^- colliders with a cleaner environment. In these colliders, the Particle Flow concept becomes a trend for the detector design. The Particle Flow aims at reconstructing all the final state particles, leads to a higher efficiency and purity on the final physics objects. In order to reconstruct the particles correctly with the most suited sub-detector system, the detector design requires a precise tracking system and high granularity calorimeter system. While the subdetector prototypes are designed and adjustable, full simulation studies are performed to define the characteristics and physics capabilities of the final detector.

Taking the Higgs mass resolution of 250 GeV ZH ($Z \to \nu\nu$, H$\to gg$) events as the reference to compare the performance, several models with different ECAL layer number, HCAL layer number, and magnetic field have been studied. The result shows that by degrading the transverse granularity of ECAL by 1/3 ($\sim$1/2 budgets for ecal), we lose 6% of resolution. The influence of thickness and cell size of the Si sensors also gives hints for the engineering. The previous detector design of CEPC takes most of the ILD detector as the framework, however, the HCAL was designed for higher energy. The result on the HCAL layer numbers and magnetic field provide proves to safely reduce the number of HCAL to 40 layers and to reduce the magnetic field from 3.5T to 3T, which is appreciated by the MDI. With this optimization, the CEPC will release a new version of CEPC detector in the CDR on preparing.

The particle identification is essential to the precise Higgs measurements. In the PFA oriented detectors, the segmentation between clusters, detailed energy and spatial information, and track information are provided. Taking full advantage of this information, a dedicated lepton identification algorithm for Higgs factories, LICH, has been developed. For the single particles with energy higher than 2 GeV, LICH reaches an efficiency better than 99.5% in identifying the muons and the electrons, and 98% for pions. The algorithm is also tested in full simulated events, showing that LICH is powerful in these events to select high energy leptons, In the jet environment, the performance is limited by the isolation performance and the unbalanced statistics for leptons and hadrons. Since the particle identification requires high granularity for the segmentation, the performance of different granular calorimeters has been studied, showing that the efficiency of finding two leptons decreases by 1∼2 % when the cell size doubles, which means that the detector needs 2∼4% more statistics in the running. Another advantage of LICH is that the identification condition is adjustable according to the analysis. In the preparation of CEPC CDR, most of the physics are analyzed with LICH.

The reconstruction of all final state particles in PFA also allows it to reconstruct τ events with higher efficiency. Since the multiplicity of τ is much smaller than that of jets, the $H \rightarrow \tau\tau$ events can easily be recognized in the leptonic channels, where the leptons decayed from Z can be vetoed by their recoil mass. In hadronic events, the method of defining well-isolated cones with smaller multiplicity is used to choose the τ candidates. The reconstructed τ candidates are selected to deduce the information of the di-τ system and the qq system. The irreducible backgrounds such as ZZ and ZH with Z decaying to $\tau\tau$, are reduced to 1% level.

Thanks to the efficient vertex detector, the starting point of particles can be measured with excellent resolution. Therefore, the impact parameter is used in the tagging of τ, as a method to get the statistic of signals and backgrounds. At the center-of-mass energy of 250 GeV and 5000 fb^{-1}, the obtained precisions for the production cross section times the branching ratio, $\Delta(\sigma \times BR)/(\sigma \times BR)$, is 2.68% for $\mu\mu H$, 4.29% for $\nu\nu H$, and 1.05% for qqH. After extrapolating the result for $\mu\mu H$ to eeH, the combined accuracy of the $H \rightarrow \tau\tau$ is 0.89%, the simple extrapolation to ILC gives an accuracy of 0.84%.

In conclusion, the CEPC detector design is still under optimization, by using the particle flow algorithm. The lepton identification not only provides tools for analysis but also helps to optimize the detector. The τ analysis shows that at current detector design, the accuracy can achieve the 1% level or even better, satisfying the requirement of new physics.

Bibliography

[1] Sha Bai, Jie Gao, Yiwei Wang, Qinglei Xiu, Weichao Yao, and Teng Yue. MDI Design in CEPC partial double ring. In *7th International Particle Accelerator Conference (IPAC'16), Busan, Korea, May 8-13, 2016*, pages 3802–3804. JACOW, Geneva, Switzerland, 2016.

[2] Chris Adolphsen. The International Linear Collider Technical Design Report-Volume 3. II: Accelerator Baseline Design. Technical report, Argonne National Laboratory (ANL), Argonne, IL (United States); Thomas Jefferson National Accelerator Facility (TJNAF), Newport News, VA (United States); Brookhaven National Laboratory (BNL), Upton, NY (United States); SLAC National Accelerator Laboratory (SLAC), Menlo Park, CA (United States); Fermi National Accelerator Laboratory (FNAL), Batavia, IL (United States), 2013.

[3] Frank Gaede, Ties Behnke, Norman Graf, and Tony Johnson. LCIO-A persistency framework for linear collider simulation studies. *arXiv preprint physics/0306114*, 2003.

[4] O Wendt, F Gaed, and T Krämer. Marlin: Modular Analysis and Reconstruction for the LINear collider. Technical report, LC-DET-2007-001, 2007.

[5] Sheldon L Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22 (4):579–588, 1961.

[6] Steven Weinberg. A model of leptons. *Physical review letters*, 19(21):1264, 1967.

[7] M Veltman et al. Regularization and renormalization of gauge fields. *Nuclear Physics B*, 44(1):189–213, 1972.

[8] Harald Fritzsch, Murray Gell-Mann, and Heinrich Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365–368, 1973.

[9] FJ Hasert, S Kabe, W Krenz, J Von Krogh, D Lanske, J Morfin, K Schultze, H Weerts, G Bertrand-Coremans, Jean Sacton, et al. Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment. *Nuclear Physics B*, 73(1):1–22, 1974.

[10] Ugo Amaldi, Albrecht Böhm, LS Durkin, Paul Langacker, Alfred K Mann, William J Marciano, Alberto Sirlin, and HH Williams. Comprehensive analysis of data pertaining to the weak neutral current and the intermediate-vector-boson masses. *Physical Review D*, 36(5):1385, 1987.

[11] Serguei Chatrchyan, Vardan Khachatryan, Albert M Sirunyan, Armen Tumasyan, Wolfgang Adam, Ernest Aguilo, T 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] Georges Aad, T Abajyan, B Abbott, J Abdallah, S Abdel Khalek, AA Abdelalim, O Abdinov, R Aben, B Abi, M Abolins, 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.

[13] Hans Peter Nilles. Supersymmetry, supergravity and particle physics. *Physics Reports*, 110(1-2):1–162, 1984.

[14] Howard E Haber and Gordon L Kane. The search for supersymmetry: probing physics beyond the standard model. *Physics Reports*, 117(2-4):75–263, 1985.

[15] John Ellis, Giovanni Ridolfi, and Fabio Zwirner. Radiative corrections to the masses of supersymmetric Higgs bosons. *Physics Letters B*, 257(1-2):83–91, 1991.

[16] Ties Behnke, James E Brau, Brian Foster, Juan Fuster, Mike Harrison, James McEwan Paterson, Michael Peskin, Marcel Stanitzki, Nicholas Walker, and Hitoshi Yamamoto. The International Linear Collider Technical Design Report-Volume 1: Executive Summary. *arXiv preprint arXiv:1306.6327*, 2013.

[17] Abdelhak Djouadi, Joseph Lykken, Klaus Mönig, Yasuhiro Okada, Mark Oreglia, and Satoru Yamashita. International Linear Collider reference design report volume 2: physics at the ILC. *arXiv preprint arXiv:0709.1893*, 2007.

[18] CEPC-SPPC study group et al. CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector. Technical report, IHEP-CEPC-DR-2015-01, 2015.

[19] Johannes Gutleber. Future Circular Collider, overview. Technical report, FCC-DRAFT-MGMT-2016-005, 2014.

[20] Henri Videau. Energy flow or particle flow-the technique of" energy flow" for pedestrians. In *International Conference on Linear Colliders-LCWS04*, pages 105–120. Ecole Polytechnique Palaiseau, 2004.

[21] C Patrignani, Particle Data Group, et al. Review of particle physics. *Chinese physics C*, 40(10):100001, 2016.

[22] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, Helge Voss, M Backes, T Carli, O Cohen, A Christov, et al. TMVA-Toolkit for multivariate data analysis. *arXiv preprint physics/0703039*, 2007.

[23] G. Alexander et al. A Precise measurement of the tau polarization and its forward - backward asymmetry at LEP. *Z. Phys.*, C72:365–375, 1996. doi: 10.1007/s002880050257.

[24] Gordon L Kane. *Modern elementary particle physics: the fundamental particles and forces*. Addison-Wesley, 1993.

[25] Michael Edward Peskin. *An introduction to quantum field theory*. Westview press, 1995.

[26] Howard Georgi. *Lie algebras in particle physics: from isospin to unified theories*, volume 54. Westview press, 1999.

[27] François Englert and Robert Brout. Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13(9):321, 1964.

[28] Peter W Higgs. Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13(16):508, 1964.

[29] Lisa Randall and Raman Sundrum. Large mass hierarchy from a small extra dimension. *Physical Review Letters*, 83(17):3370, 1999.

[30] Ali H Chamseddine, Ro Arnowitt, and Pran Nath. Locally supersymmetric grand unification. *Physical Review Letters*, 49(14):970, 1982.

[31] Luis Ibáñez. Locally supersymmetric SU (5) grand unification. *Physics Letters B*, 118(1-3):73–78, 1982.

[32] John Baez and John Huerta. The algebra of grand unified theories. *Bulletin of the American Mathematical Society*, 47(3):483–552, 2010.

[33] Julius Wess and Jonathan Bagger. *Supersymmetry and supergravity*. Princeton university press, 1992.

[34] C Arzt, MB Einhorn, and J Wudka. Patterns of deviation from the standard model. *Nuclear Physics B*, 433(1):41–66, 1995.

[35] Damien M Pierce, Jonathan A Bagger, Konstantin T Matchev, and Ren-jie Zhang. Precision corrections in the minimal supersymmetric standard model. *Nuclear Physics B*, 491(1-2):3–67, 1997.

[36] Nima Arkani-Hamed, Savas Dimopoulos, and Gia Dvali. The hierarchy problem and new dimensions at a millimeter. *Physics Letters B*, 429(3-4):263–272, 1998.

[37] G Apollinari, I Béjar Alonso, Oliver Brüning, M Lamont, and Lucio Rossi. High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report. Technical report, Fermi National Accelerator Laboratory (FNAL), Batavia, IL (United States), 2015.

[38] Lyn Evans and Shinichiro Michizono. The International Linear Collider Machine Staging Report 2017. 2017.

[39] Ties Behnke et al. The international linear collider technical design report-volume 4: detectors. Technical report, Argonne National Laboratory (ANL), Argonne, IL (United States); Pacific Northwest National Laboratory (PNNL), Richland, WA (United States); SLAC National Accelerator Laboratory (SLAC), Menlo Park, CA (United States); Fermi National Accelerator Laboratory (FNAL), Batavia, IL (United States), 2013.

[40] CEPC-SPPC Study Group et al. CEPC-SPPC Preliminary Conceptual Design Report, Volume II-Accelerator. *IHEP, Beijing, China, Rep. IHEP-AC-2015-01*, 2015.

[41] CEPC-SPPC Study Group et al. Cepc-sppc progress report, accelerator (2015–2016). Technical report, IHEP-CEPC-DR-2017-01, IHEP-CEPC-AC-2017-01, 2017.

[42] Cai Meng, Jingru Zhang, Dou Wang, Xiaoping Li, Guoxi Pei, Shilun Pei, and Yunlong Chi. CEPC Linac Design and Beam Dynamics. 2017.

[43] Dou Wang, Jie Gao, Feng Su, Yuan Zhang, Jiyuan Zhai, Yiwei Wang, Sha Bai, Huiping Geng, Tianjian Bian, Xiaohao Cui, et al. CEPC partial double ring scheme and crab-waist parameters. In *The Future of High Energy Physics: Some Aspects*, pages 179–188. World Scientific, 2017.

[44] Jean-Claude Brient and Henri Videau. The calorimetry at the future e+ e-linear collider. *arXiv preprint hep-ex/0202004*, 2002.

[45] MA Thomson. Particle flow calorimetry and the PandoraPFA algorithm. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 611(1):25–40, 2009.

[46] Florian Beaudette. The CMS Particle Flow Algorithm. *arXiv preprint arXiv:1401.8155*, 2014.

[47] Collaboration ATLAS. Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment. Technical report, 2012.

[48] Daniel Jeans, Marcel Reinhard, and Jean-Claude Brient. GAmma Reconstruction at a LInear Collider. 2013.

[49] Manqi Ruan. Arbor, a new approach of the Particle Flow Algorithm. *arXiv preprint arXiv:1403.4784*, 2014.

[50] Damir Buskulic, D Casper, I De Bonis, D Decamp, P Ghez, C Goy, J-P Lees, M-N Minard, P Odier, B Pietrzyk, et al. Performance of the ALEPH detector at LEP. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 360(3):481–506, 1995.

[51] P Moras de Freitas et al. MOKKA: A detailed Geant4 simulation for the International Linear Collider detectors, 2003.

[52] Sea Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G Barrand, et al. GEANT4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[53] Zhenxing Chen, Ying Yang, Manqi Ruan, Dayong Wang, Gang Li, Shan Jin, and Yong Ban. Study of Higgsstrahlung Cross Section and Higgs Mass Measurement Precisions with ZH (Z$\rightarrow$ $\mu_{+}\mu^{-}$) events at CEPC. *arXiv preprint arXiv:1601.05352*, 2016.

[54] Manqi Ruan, Daniel Jeans, Vincent Boudry, Jean-Claude Brient, and Henri Videau. Fractal Dimension of Particle Showers Measured in a Highly Granular Calorimeter. *Physical review letters*, 112(1):012001, 2014.

[55] Aleph Collaboration et al. Measurement of the Tau Polarisation at LEP. *arXiv preprint hep-ex/0104038*, 2001.

[56] Eric Braaten, Stephan Narison, and A Pich. QCD analysis of the tau hadronic width. *Nuclear Physics B*, 373(3):581–612, 1992.

[57] Antonio Pich. Precision tau physics. *Progress in Particle and Nuclear Physics*.

[58] CLEO collaboration et al. Experimental tests of lepton universality in {tau} decay. *Physical Review, D*, 55(5), 1997.

[59] DP Roy. The hadronic tau decay signature of a heavy charged Higgs boson at LHC. *Physics Letters B*, 459(4):607–614, 1999.

[60] Matthew J Dolan, Christoph Englert, and Michael Spannowsky. Higgs self-coupling measurements at the LHC. *Journal of High Energy Physics*, 2012(10):112, 2012.

[61] Seong-Youl Choi and Manuel Drees. Signals for CP violation in scalar tau pair production at muon colliders. *Physical review letters*, 81(25):5509, 1998.

[62] Nathan Isgur and Mark B Wise. Weak decays of heavy mesons in the static quark approximation. *Physics Letters B*, 232(1):113–117, 1989.

[63] Serguei Chatrchyan, V Khachatryan, AM Sirunyan, A Tumasyan, W Adam, T Bergauer, M Dragicevic, J Eroe, C Fabjan, M Friedl, et al. Search for neutral Higgs bosons decaying to tau pairs in pp collisions at. *Physics Letters B*, 713(2): 68–90, 2012.

[64] Georges Aad, Brad Abbott, Jalal Abdallah, S Abdel Khalek, O Abdinov, Rosemarie Aben, Babak Abi, Maris Abolins, OS AbouZeid, Halina Abramowicz, et al. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *Journal of high energy physics*, 2015(4):117, 2015.

[65] Trong Hieu Tran, Vladislav Balagura, Vincent Boudry, J-C Brient, and Henri Videau. Reconstruction and classification of tau lepton decays with ILD. *The European Physical Journal C*, 76(8):468, 2016.

[66] Wolfgang Kilian, Thorsten Ohl, and Jürgen Reuter. Whizard—simulating multi-particle processes at lhc and ilc. *The European Physical Journal C*, 71(9):1742, 2011.

[67] Daniel Jeans. Tau lepton reconstruction at collider experiments using impact parameters. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 810:51–58, 2016.

Acknowledgement

Firstly, I would like to thank my two supervisors, Vincent and Manqi, for your supports during my PhD study and research. Your guidance helped me in the research and also while writing this thesis. Your patience and advices encouraged me to finish my thesis.

I would also like to thank Mark Thomson and Luca for accepting the roles of the reporter as well as Florian, especially Zhiqing Zhang, for presiding over my thesis jury.

And thank you, Jean-Claude, for accepting me to do this thesis at the LLR, and for your ideas and guidances. I would like to express my sincere gratitude to Henri Videau, you are really helpful to me, you always answer my stupid question, correct my error, I would not have my thesis without you.

My thanks also goes to my fellow labmates in LLR, Vladik, Hieu, Yacine, Kostia, and so on, all of you have helped me during this three years, or since even earlier.

I would also like to take this opportunity to express my gratitude to my family: my parents and my sister. It is your supports that guaranteed my study in France and kept me away from the worries.

Thanks to my neighbors in Building 76, especially K & S, for feeding me with delicious soups, as well as the others, thank you for all the fun we have had and for providing me miraculous life in the campus. To my friends at IHEP, thank you for welcoming me and helped me adapting the life in Beijing.

université PARIS-SACLAY

ÉCOLE DOCTORALE
PHENIICS

Titre : Reconstruction des objets leptoniques dans future e+e- usine de Higgs

Mots clés : Higgs, canaux tau, calorimétrie

Résumé : Depuis la découverte du boson de Higgs en 2012 par les expériences du Large Hadron Collider (LHC), la mesure précise est devenue le défi dans les expériences de physique des hautes énergies. De nombreuses usines de Higgs électron-positon avec une précision améliorée sur les mesures de largeur totale de Higgs ont été proposées, y compris le collisionneur linéaire international (ILC) et le collisionneur à électrons positrons circulaires (CEPC). Afin d'atteindre la précision à des niveaux de pourcentage ou de sous-pourcentage, l'utilisation de l'algorithme de flux de particules (PFA) est devenue le paradigme de la conception de détecteurs pour la frontière à haute énergie. L'idée clé est de reconstruire chaque particule d'état finale dans les sous-détecteurs les plus adaptés, et de reconstruire tous les objets physiques au-dessus des particules d'état finales. Les détecteurs orientés à PFA ont une grande efficacité dans la reconstruction d'objets physiques tels que les leptons, les jets et l'énergie manquante.
Dans cette thèse, une identification par lepton basée sur PFA (Lepton Identification pour calorimètre à haute granularité) a été développée pour des détecteurs utilisant des calorimètres à haute granularité. Utilisation de la géométrie du détecteur conceptuel pour le CEPC, et les échantillons de particules chargées uniques d'énergie supérieure à 2 GeV, LICH identifie les électrons ou les muons avec des rendements supérieurs à 99,5% et contrôle le taux de désinscription du hadron aux muons ou aux électrons 1% ou 0.5 %. Réduisant la granularité du calorimètre de 1 ou 2 ordres de grandeur, la performance d'identification du lepton est stable pour les particules avec E $>$ 2 GeV Appliquée à des événements eeH ou $\mu\mu H$ simulés à $\sqrt{s} = 250$ GeV, la performance d'identification du lepton est cohérente avec le cas d'une seule particule: l'efficacité d'identifier tous les leptons de haute énergie dans un événement est de 95,5 $\sim$ 98,5 %.
Les produits τ-decay dans les collisionneurs de haute énergie sont étroitement collimatés et ont une faible multiplicité, fournissant d'excellentes signatures à sonder. Dans cette thèse, les canaux H $\rightarrow \tau\tau$ sont analysés dans différents modes de désintégration Z avec le contexte SM pris en compte. La précision finale combinée de $\sigma \times Br(H \rightarrow \tau\tau)$ devrait être de 0.89 %.

Title : Reconstruction of leptonic physics objects at future e+e- Higgs factory

Keywords : Higgs, tau channel, calorimetry

Abstract : Since the discovery of the Higgs boson in 2012 by the experiments at the Large Hadron Collider (LHC), precise measurement of Higgs boson has become the challenge in high energy physics experiments. Many electron-positron Higgs factories with improved accuracy on the Higgs total width measurements have been proposed, including the International Linear Collider (ILC), the Circular Electron Positron Collider (CEPC), the Future Circular Collider e^+e^- (FCCee). In order to achieve the precision estimated to percent or sub-percent levels, the use of Particle Flow Algorithm (PFA) has become the paradigm of detector design for the high energy frontier. The key idea is to reconstruct every final state particle in the most suited sub-detectors, and reconstruct all the physics objects on top of the final state particles. The PFA oriented detectors have high efficiency in reconstructing physics objects such as leptons, jets, and missing energy.
In this thesis, a PFA based lepton identification (Lepton Identification for Calorimeter with High granularity (LICH) has been developed for detectors with high granularity calorimeters. Using the conceptual detector geometry for the CEPC, and samples of single charged particles with energy larger than 2 GeV, LICH identifies electrons or muons with efficiencies higher than 99.5% and controls the mis-identification rate of hadron to muons or electrons to better than 1% or 0.5% respectively. Reducing the calorimeter granularity by 1 or 2 orders of magnitude, the lepton identification performance is stable for particles with E $>$ 2 GeV. Applied to fully simulated eeH or $\mu\mu$ H events at $\sqrt{s} = 250$GeV, the lepton identification performance is consistent with the single particle case: the efficiency of identifying all the high energy leptons in an event ranges between 95.5% and 98.5%.
The τ-decay products have low multiplicity and in high energy colliders are tightly collimated and have low multiplicity, providing excellent signatures to probe. In this thesis, the H$\rightarrow \tau\tau$ channel is analyzed in different Z decay modes with SM background taken into account. The combined final accuracy of $\sigma \times Br(H \rightarrow \tau\tau)$ is expected to be 0.89%.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France