



HAL
open science

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

Dan Yu

► **To cite this version:**

Dan Yu. Reconstruction of leptonic physics objects at future e^+e^- Higgs factory. High Energy Physics - Experiment [hep-ex]. Université Paris Saclay (COMUE); Institute of high energy physics (Chine), 2018. English. NNT: 2018SACLX018 . tel-01852267

HAL Id: tel-01852267

<https://pastel.hal.science/tel-01852267>

Submitted on 1 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



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
Mark Thomson
University of Cambridge
Luca Malgeri
CERN
Florian Beaudette
LLR-IN2P3
Vincent Boudry
LLR-IN2P3
Manqi RUAN
IHEP-CAS

Président
Rapporteur
Rapporteur
Examineur
Co-directeur de thèse
Co-directeur de thèse

1 Abstract

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

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

19 In this thesis, a PFA based lepton identification (Lepton Identification for Calorimeter
20 with High granularity (LICH) has been developed for detectors with high granularity
21 calorimeters. Using the conceptual detector geometry for the CEPC, featuring typical
22 calorimeter granularity of 1000 and 400 cells / cm^3 respectively for the electromagnetic
23 and hadronic parts, and samples of single charged particles with energy larger than 2
24 GeV, LICH identifies electrons or muons with efficiencies higher than 99.5% and con-
25 trols the mis-identification rate of hadron to muons or electrons to better than 1% or
26 0.5% respectively. Reducing the calorimeter granularity by 1 or 2 orders of magnitude,
27 the lepton identification performance is stable for particles with $E > 2$ GeV. Applied to
28 fully simulated eeH or $\mu\mu H$ events at $\sqrt{s} = 250\text{GeV}$, the lepton identification perfor-
29 mance is consistent with the single particle case: the efficiency of identifying all the high
30 energy leptons in an event ranges between 95.5% and 98.5%.

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

39 **Résumé**

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

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

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

79 % respectivement. la réduction de la granularité du calorimètre par 1 ou 2 ordres de
80 grandeur, la performance d'identification de lepton est stable pour des particules avec
81 $E > 2$ GeV. appliquée à eeH entièrement simulé ou $\mu\mu$ événements H à $\sqrt{s} = 250$ GeV, les
82 performances d'identification de lepton est compatible avec le cas de particules unique:
83 l'efficacité de l'identification de tous les leptons de haute énergie dans un événement se
84 situe entre 95,5 % et 98,5 %.

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

Contents

100

101	1 Introduction	1
102	2 Theory	4
103	2.1 Standard Model	5
104	2.1.1 The Electroweak symmetry breaking	6
105	2.1.2 Higgs mechanism	7
106	2.2 Beyond Standard Model	13
107	3 e^+e^- Collider as Higgs factory	15
108	3.1 Production processes	17
109	3.2 The Circular Electron-Positron Collider (CEPC)	19
110	3.2.1 Accelerator design	19
111	3.2.2 Machine Detector Interface (MDI)[1]	22
112	3.3 The International Linear Collider (ILC)	23
113	3.3.1 ILC Subsystems[2]	24
114	3.4 The Future Circular Collider (FCC) and High Luminosity LHC (HL-LHC)	28
115	4 Detector	29
116	4.1 Particle Flow Algorithm (PFA) oriented detector	30
117	4.2 Detector design	31
118	5 Softwares and Particle Flow Algorithm (PFA)	34
119	5.1 Particle Flow Algorithm	34
120	5.1.1 Jet Energy Resolution	34
121	5.1.2 PandoraPFA	35
122	5.1.3 Arbor	37
123	5.2 Tools	42
124	5.2.1 LCIO[3]	42
125	5.2.2 Simulation	42
126	5.2.3 Marlin Framework[4]	42
127	5.3 Detector optimization	43
128	5.3.1 ECAL optimization	43

129	5.3.2	HCAL and B field optimization	45
130	6	Particle identification	48
131	6.1	Detector geometry and sample	48
132	6.2	Discriminant variables and the output likelihoods	49
133	6.3	Performance on single particle events	56
134	6.4	Lepton identification performance on single particle events for different	
135		geometries	62
136	6.5	Performance on physics events	66
137	6.6	Conclusion	68
138	7	Measurement of $H \rightarrow \tau\tau$ Branching Ratio	73
139	7.1	Introduction	73
140	7.1.1	τ physics	73
141	7.1.2	τ decay modes	74
142	7.1.3	Measurements and precisions	77
143	7.2	Samples	77
144	7.3	Leptonic channels	78
145	7.3.1	$Z \rightarrow \mu\mu$	79
146	7.3.2	$Z \rightarrow \nu\nu$	85
147	7.4	Hadronic channel, $Z \rightarrow qq$	87
148	7.5	Combined Results	97
149	7.6	Extrapolating in ILC	99
150	7.7	Discussion	99
151	8	Conclusion	101
152		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 a 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

181 of background debris. This makes possible to treat the events as a whole and to con-
 182 strain the new particle properties with the knowledge of the initial state. Two advanced
 183 proposals of e^+e^- Higgs factories are the International Linear Collider (ILC)[16, 17], the
 184 Circular Electron-Positron Collider (CEPC)[18], and the Future Circular Collider e^+e^-
 185 (FCCee)[19]. The ILC provides polarized beams and leaves the possibility to be up-
 186 graded to higher energy, while the CEPC provides higher luminosity and can be up-
 187 graded to a proton-proton collider. The FCC is a design study of CERN to extend the
 188 research after LHC reaches the end of its lifespan, and FCCee is part of it.

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

202 The lepton identification is fundamental to the Higgs measurements. About 7% of
 203 Higgs bosons at the CEPC or ILC are generated together with a pair of electrons or
 204 muons. Those events are the golden signals for the Higgs recoil analysis, which is the
 205 anchor for the absolute Higgs measurements. A unnegligable fraction of the Higgs bo-
 206 son decays, directly or via cascade, into final states with leptons[21]: i.e., 0.02% of SM
 207 Higgs decays into muons; the leptons are the essential candles of the identification of
 208 $H \rightarrow WW/ZZ \rightarrow$ leptonic /semi-leptonic final states. In addition, a significant fraction
 209 of Higgs $\rightarrow bb/cc$ events generate leptons in their decay cascade. A highly efficiency
 210 lepton identification is also highly appreciated for the EW measurements. The track-
 211 ing system and highly granular calorimetric system provide discriminant variables for
 212 the particle identification, and the TMVA toolkit[22] offers optimal utilization of these
 213 variables.

214 The τ lepton[21] is an extremely intriguing physics object. As the heaviest lepton in the
 215 SM, τ has a large Yukawa coupling $g(H\tau\tau)$ to the Higgs boson, leading to a significant
 216 branching ratio $Br(H \rightarrow \tau\tau)$. The $g(H\tau\tau)$ is expected to be measured with a better than
 217 1% relative accuracy at the CEPC. Measuring the τ polarization at the Z pole leads to a
 218 precise determination of the backward-forward asymmetry $A_{FB}(\tau)$ [23]. The reconstruc-
 219 tion of the tau spectral functions also have compelling potential at the CEPC. In this
 220 thesis, the τ reconstruction is covering the channel of Higgs decaying to $\tau\tau$ accompa-

221 nied with leptons or jets. The basic idea is to take advantage of the high granularity and
222 the property of multiplicity.

223 The content of this paper is organized as follows. A brief overview of the Standard
224 Model is described in Chapter 2. In Chapter 3, we describe the CEPC and ILC as e^+e^-
225 colliders. The PFA oriented detectors will be introduced in Chapter 4, followed by the
226 presentation of two PFAs and their application in detector optimization in Chapter 5.
227 The description of the particle identification package and its performance on single par-
228 ticles as well as in fully simulated events are discussed in Chapter 6. In Chapter 7, we
229 will discuss the signal strength of the Higgs boson decaying into tau lepton pairs at the
230 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

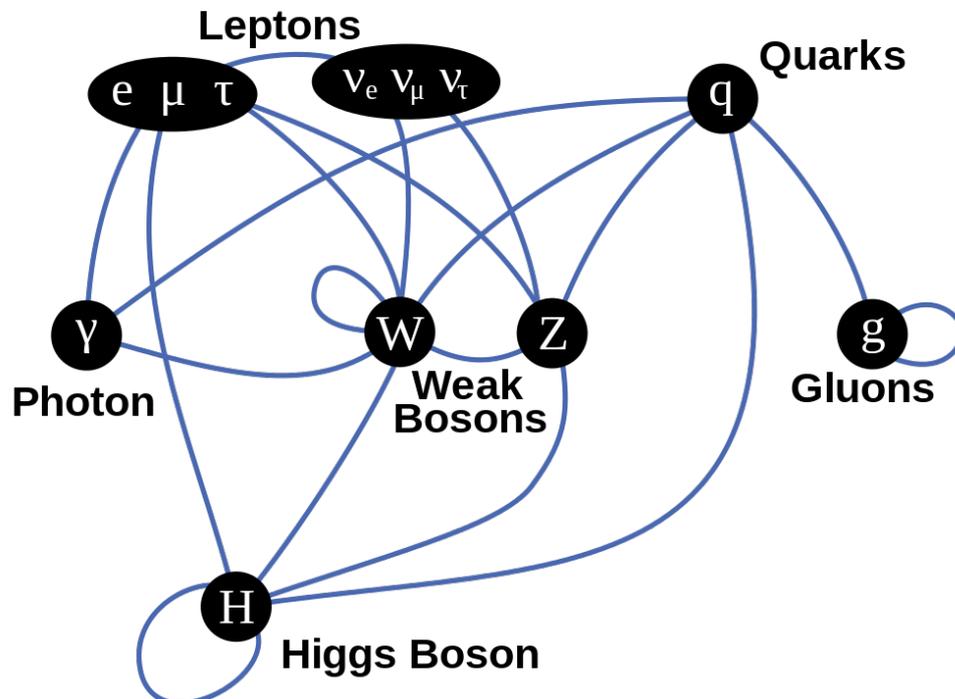


Figure 2.1: The Standard Model particles and the interaction in between them

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

269 2.1 Standard Model

270 The Standard Model Lagrangian is written in three parts, the kinematic terms, the cou-
 271 pling terms, and mass terms, it can be written in a simplified formula as:

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

272 the terms in this formula are: the scalar product of the field strength tensor $F_{\mu\nu}$ con-
273 taining the mathematical encoding of all interaction particles except the Higgs boson,
274 the term describing how interaction particles interact with matter particles, the term de-
275 scribing how matter particles couple to the Brout–Englert–Higgs field ϕ and obtaining
276 mass, how the interaction particles couple to the BEH field, and the potential of the BEH
277 field.

278 2.1.1 The Electroweak symmetry breaking

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

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

281 where a sum over all $i = 1, \dots, n$ is conducted in each term. μ^2 corresponds to the ordi-
282 nary mass term m^2 with a changed sign. We identify the first term as the kinetic energy
283 of the system and deduce that the rest is the potential $V(\phi^i)$. The altered sign of the mass
284 $m^2 \rightarrow \mu^2$ term will allow for a potential with negative minima, which will be crucial in
285 our depict of symmetry breaking. This is an example of a self-interacting theory where
286 λ is a dimensionless coupling constant describing the strength of the interaction (a more
287 basic example than the QED Lagrangian which also encodes a self-interacting theory).
288 By setting an even power of the fields, we will be able to obtain positive definite en-
289 ergies (and scalar field theories with fields of an even higher order than 4 will not be
290 renormalizable). The lowest energy value of \mathcal{L} is obtained when we are dealing with a
291 uniform constant field $\phi(x) = \phi_0^i$. It is chosen as the field which minimizes the potential
292 term in \mathcal{L} , i.e.:

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

293 This optimization problem is straightforward to solve and we find

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

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

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

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

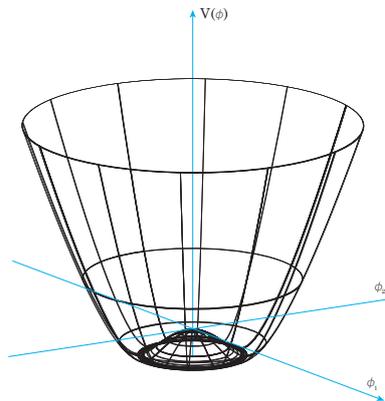


Figure 2.2: A visualization of the potential V in the case where $n = 2$. Notice that the minima where V has negative values are found on the circle defined by 2.1.5 which physically correspond to a set of degenerate vacua.

304 2.1.2 Higgs mechanism

305 Introducing a complex scalar field ϕ will satisfy Lorentz invariance as well as rotational
 306 invariance, due to its scalar nature. This field might yield a non-zero expectation value
 307 of the vacuum as we have seen in the calculations above and let us construct a gauge

308 invariant Lagrangian which gauge bosons will acquire mass. A usual choice is to call
309 this complex scalar field ϕ and to write it as:

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

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

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

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

315 **Boson mass**

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

$$D_\mu \phi = (\partial_\mu + igA_\mu^a \tau^a) \phi \quad (2.1.8)$$

321 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
322 two dimensional representation (τ^a Hermitian matrices). In the part of the Lagrangian
323 corresponding to a kinetic term and let it couple to the field, and the charge of the
324 Lagrangian $\Delta\mathcal{L}$ be written as

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

325 with the mass coefficient for the three gauge bosons as

$$m_A = \frac{gv}{2} \quad (2.1.10)$$

326 That particle can obtain mass by interacting with a field of this kind is known as the
 327 Higgs mechanism or with more names occasionally the Englert–Brout–Higgs mechanism[27,
 328 28].

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

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

$$\psi = \begin{pmatrix} \psi^+ \\ \psi^0 \end{pmatrix} \quad (2.1.11)$$

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

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

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

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

351 Generally, we put

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

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

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

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

$$e = g_1 \sin \omega_W = g_2 \cos \omega_W \quad (2.1.16)$$

356 Introducing the charged field W_μ^\pm as

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

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

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

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

$$M_{W^\pm} = \frac{1}{2} g_2 v \quad (2.1.19)$$

361 because the term proportional to the bosons corresponds to charged intermediate boson
 362 masses. And we can define the mass of neutral gauge boson Z by using the relations in
 363 2.1.16, we get

$$M_Z = \frac{M_W}{\cos \omega_W} = \frac{v}{2} \sqrt{g_1^2 + g_2^2} \quad (2.1.20)$$

364 Thus the masses for the W bosons and Z^0 -boson have been found, and of course a
 365 similar method can be used for a larger and more complicated group $SU(3) \times SU(2) \times$
 366 $U(1)$, which leads to the construction of the complete Standard Model.

367 Fermion mass

368 The fermion term of the Lagrangian is:

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

369 However this lagrangian is not gauge invariant since the left handed fermions form
 370 an isospin doublet and the right handed fermions form isospin singlets. In order to
 371 construct an $SU(2)_L \times U(1)_Y$ invariant term for fermions, we used the complex doublet
 372 introduced in the previous section, which gives:

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

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

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

$$\begin{aligned} L &= Y_{ij}^d \bar{Q}_L^i \phi d_R^j + Y_{ij}^u \bar{Q}_L^i \tilde{\phi} u_R^j + Y_{ij}^l \bar{L}_L^i \phi l_R^j + h.c. \\ &= \frac{1}{\sqrt{2}} Y_{ij}^d (\bar{u}_L^i, \bar{d}_L^i) \begin{pmatrix} 0 \\ v \end{pmatrix} d_R^j + \frac{1}{\sqrt{2}} Y_{ij}^u (\bar{u}_L^i, \bar{d}_L^i) \begin{pmatrix} v \\ 0 \end{pmatrix} u_R^j + \frac{1}{\sqrt{2}} Y_{ij}^l (\bar{\nu}^i, \bar{e}^i) \begin{pmatrix} 0 \\ v \end{pmatrix} e_R^j + h.c. \\ &= \frac{Y_{ij}^d \cdot v}{\sqrt{2}} \bar{d}_L^i d_R^j + \frac{Y_{ij}^u \cdot v}{\sqrt{2}} \bar{u}_L^i u_R^j + \frac{y_{ii} \cdot v}{\sqrt{2}} \bar{e}^i e^i \end{aligned} \quad (2.1.23)$$

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

379 **Higgs coupling**

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

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

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

384 where v is the vacuum expectation of ϕ and h is a fluctuating real valued field with
 385 $\langle h \rangle = 0$.

386 Rewriting the Lagrangian in the unitary gauge, the potential energy term takes the
 387 form:

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

388 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
 389 as Higgs boson.

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

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

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

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

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

395 From 2.1.25, 2.1.26 and 2.1.27, the coupling of the Higgs boson to other particles of the
 396 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

435 two degrees of freedom and a Weyl fermionic field with two components.

436 In the breaking of Supersymmetry, we obviously need to preserve the gauge invari-
437 ance and the renormalizability of the theory and, also, the fact that there are still no
438 quadratic divergences in the Higgs boson mass squared. Since up to now there is no
439 completely satisfactory dynamical way to break SUSY, a possibility is to introduce by
440 hand terms that break SUSY explicitly and parametrize our ignorance of the fundamen-
441 tal SUSY-breaking mechanism. This gives a low energy effective SUSY theory, the most
442 economic version being the Minimal Supersymmetric Standard Model (MSSM), provid-
443 ing candidate dark matter particles. In a supersymmetric theory, Planck-scale quantum
444 corrections cancel between partners and superpartners (owing to a minus sign asso-
445 ciated with fermionic loops). Thus the hierarchy between the electroweak scale and
446 the Planck scale is achieved in a natural manner. Besides, the running of the gauge
447 couplings are modified, and precise high-energy unification of the gauge couplings is
448 achieved.

449 The precise measurement of the Higgs boson is a key to verify these proposed models.
450 If there is new physics beyond the Standard Model, the coupling deviates from the
451 Standard Model prediction. The deviation depends on the new physics beyond the
452 Standard Model but is estimated to be $O(\sim 1\%)$ in many models[34, 35, 36]. Therefore,
453 a precision of a few percent or less is required to shed light on a signal of new physics
454 concealed in the coupling constants, which can be achieved with the next generation of
455 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

484 events, the ILC detectors will have only one-tenth of the amount of material in front of
485 the calorimeters that measure photon energies.

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

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

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

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

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

522 and parcel of the FCC design study.

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

525 3.1 Production processes

526 As shown in Figure 3.1, the leading production processes for the SM Higgs boson at
527 e^+e^- collider operating at 250 GeV are: a) $e^+e^- \rightarrow ZH$ (Higgsstrahlung or ZH), b)
528 $e^+e^- \rightarrow \nu\nu H$ (WW fusion), c) $e^+e^- \rightarrow e^+e^-H$ (ZZ fusion), as shown in Figure 3.2,
529 and the estimated statistics for CEPC ($5ab^{-1}$) and ILC ($1ab^{-1}$) are shown in Table 3.1
530 and Table 3.2, the polarization for ILC 250GeV is either $P(e^+, e^-) = (+30\%, -80\%)$ or
531 $P(e^+, e^-) = (-30\%, +80\%)$.

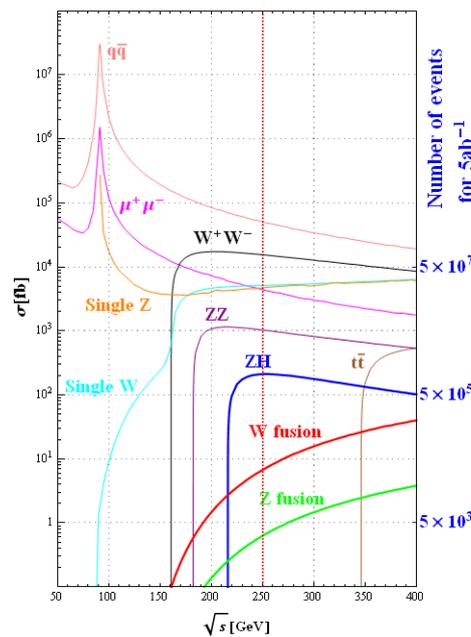


Figure 3.1: Unpolarized cross sections of main standard model processes of e^+e^- collisions as functions of center- of-mass energy (from 50GeV to 400GeV), the dotted line indicates 250GeV

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

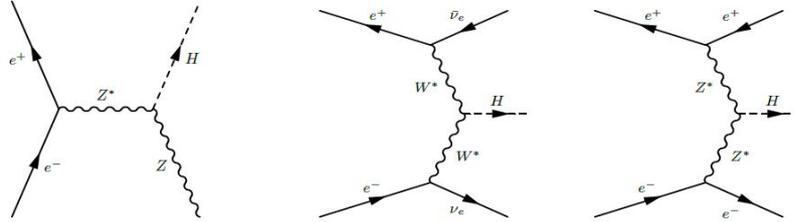


Figure 3.2: Feynman diagrams of the Higgs production processes in e^+e^- collider

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

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

540 3.2 The Circular Electron-Positron Collider (CEPC)

541 The CEPC is a circular electron-positron collider in a tunnel with a circumference of
542 100 km and is envisioned to operate with a center-of-mass energy of 250 GeV where
543 the Higgs events are produced primarily through the interaction e^+e^- . With a nominal
544 luminosity of $2 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$ about 1 million clean Higgs events will be produced by
545 CEPC over a period of 10 years. The large statistics of this Higgs sample will enable
546 CEPC to measure the Higgs boson production cross sections and most of its properties
547 with precisions far beyond what is achievable at the LHC. The CEPC can also serve as
548 a high luminosity ($10^{35\sim 36} \text{cm}^{-2}\text{s}^{-1}$) Z factory at a centre of mass energy of 91 GeV, i.e.
549 $10^{10\sim 11}$ Z boson in one year.

550 The beam current at CEPC, determined by the synchrotron radiation budget, is 100 MW
551 for two beams. The preliminary layout of 50km tunnel CEPC (2014) is shown in Figure
552 3.3, the CEPC collider is designed with four interaction points, where IP1 and IP3 are
553 for e^+e^- collisions, while the other two IP's are reserved for the future pp collider, SPPC.
554 The progressed collider[41] circumference is 100 km, including 8 arcs of 5852.8 m, 4 arc
555 straight sections of 849.6 m each and 4 interaction region straights of 1132.8 m each.

556 3.2.1 Accelerator design

557 The CEPC design aims to be a Higgs factory producing 10^6 Higgs operating at 250 GeV
558 center of mass energy and a $W\&Z$ factory producing 10^{10} Z^0 operating at 90 GeV or 160
559 GeV center of mass energy. It should also leave the opportunity to be upgraded to a
560 100TeV proton-proton collider.

561 The CEPC contains several subsystems[40]:

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

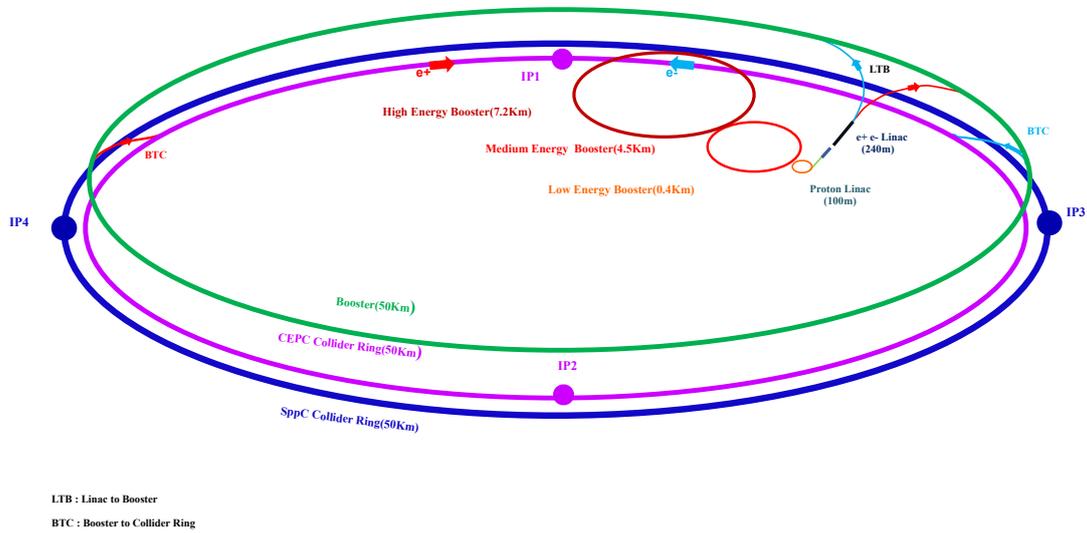


Figure 3.3: CEPC preCDR Layout

- 570 – **Electron Source** The CEPC electron source is a thermionic gridded cathode
571 driven by high voltage pulser for the baseline design. After leaving one of
572 these guns, the bunches pass through a Y bend and into two sub-harmonic
573 bunching cavities. Two operation modes are required: one is to provide a
574 3.2 nC bunch charge for electron injection, and the other is to provide an
575 11 nC bunch charge as the primary electron beam for positron production.
576 The electron beams are accelerated to 200MeV before going into the same
577 accelerating section as positrons.
- 578 – **Positron Source** In CEPC, positrons are generated using a 4 GeV electron
579 beam impinging on a high-Z, high density tungsten target. The positron yield
580 per incident electron is approximately proportional to the electron energy so
581 that the positron current is proportional to the incident power of electron
582 beam. The large transverse emittance of the positron beam emerging from
583 the target is transformed to match the capture section aperture with a pseudo-
584 adiabatically changing solenoidal field. Three constant-gradient accelerator
585 sections will boost the captured positrons to 200 MeV. The positrons are then
586 transported back to the beginning of Linac through a quadrupole lattice and
587 reinjected into the Linac where they are accelerated to 10 GeV.
- 588 – **Damping Rings** The primary purpose of the damping ring (DR) is to reduce
589 the transverse phase space of the positron beam to a suitably small value at
590 the beginning of the linac and also to adjust the time structure of the positron
591 beam for reinjection into the Linac. a bunch compressor system is added after
592 the damping ring to reduce the bunch length in the ring, thus to minimize
593 wake field effects in the Linac.
- 594 – **Accelerating section** In CEPC, the klystrons and their associated modula-
595 tors are the keys to acceleration. A first acceleration section containing 11
596 klystrons of 18 MeV/m is providing 1.1GeV electrons and positrons before
597 the positrons are sent to the Damping Rings. Then the second acceleration
598 section containing 20 klystrons of 27 MeV/m accelerate the beams to 4GeV,
599 where the electron beam is used to produce the positron beams. Finally,
600 the beams are accelerated to 10GeV through a third section containing 42
601 klystrons of 27 MeV/m. The procedure for acceleration in the Linac is shown
602 in Figure 3.4
- 603 • **Booster** After being accelerated to 10GeV, electron and positron beams are injected
604 from the Linac through the LTB transfer line (Linac to Booster) into the Booster.
605 In CEPC, the Booster is in the same tunnel as the collider, placed 2m above the
606 collider ring and has about same circumference (10km). Bypasses are arranged to
607 avoid the detectors at IPs. Because of the very low synchrotron radiation damping
608 rate, a scheme of single bunch injection from Linac to Booster is adopted. The two

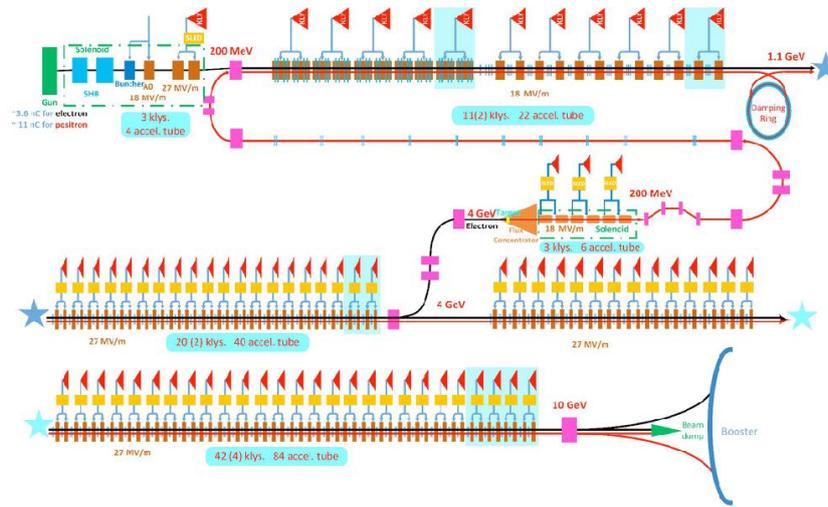


Figure 3.4: Schematic layout of the CEPC Linac, the stars represent the continuous of figure[42]

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 $\pm 7\text{m}$ 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

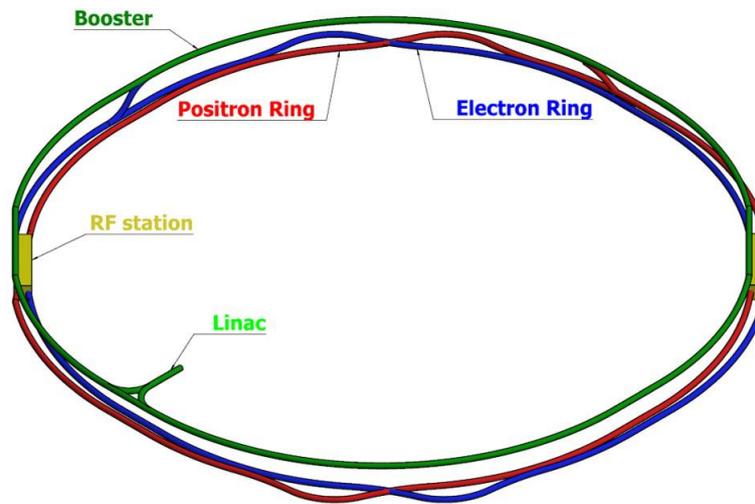


Figure 3.5: The CEPC Booster and Double Ring Layout

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

637 3.3 The International Linear Collider (ILC)

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

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

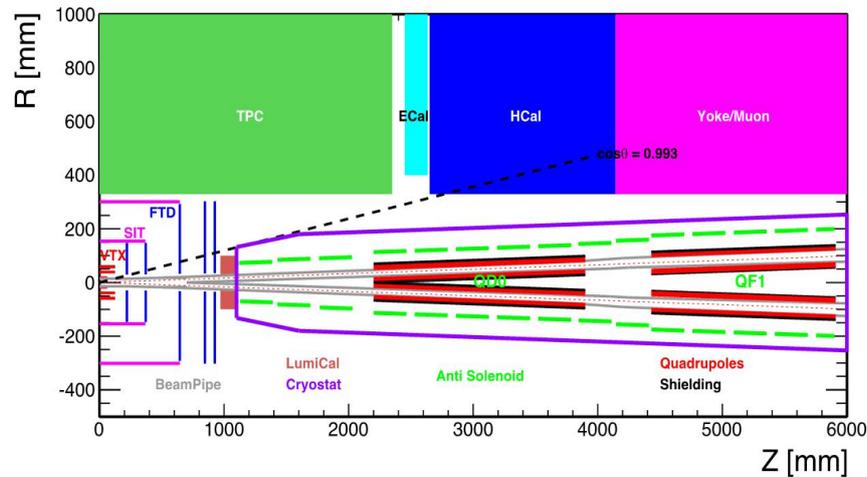


Figure 3.6: The CEPC MDI Layout

646 center of mass energy 250GeV, as shown in Figure 3.8. The beam power of ILC250 is
 647 5.26MW, with the total luminosity to be $1.35 \cdot 10^{-34} \text{cm}^{-2} \text{s}^{-1}$. Following several years of
 648 successful operation of the initial ILC250, a luminosity upgrade is possible. The basic
 649 change in the luminosity upgrade is the increase in the number of bunches from 1312 to
 650 2625.

651 The ILC will leave the opportunity to operate at higher center of mass energy: 350GeV,
 652 500 GeV or 1TeV.

653 3.3.1 ILC Subsystems[2]

654 The accelerating system of ILC contains several subsystems:

- 655 • **Electron Source** The required trains of polarized electron bunches are produced
 656 with a laser hitting a photocathode in a DC gun, then bunched and pre-accelerated
 657 in normal-conducting structures. The beam is then accelerated in a superconduct-
 658 ing linac. The spin vector is rotated into the vertical plane by superconducting
 659 solenoids, and a separate superconducting RF structure is used for energy com-
 660 pression before the beam is transported to the Damping Ring.
- 661 • **Positron Source** After accelerated to suitable energy, the electron beam is then
 662 extracted to a parallel beam line to create positrons and return the positrons to the
 663 electron main linac.

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

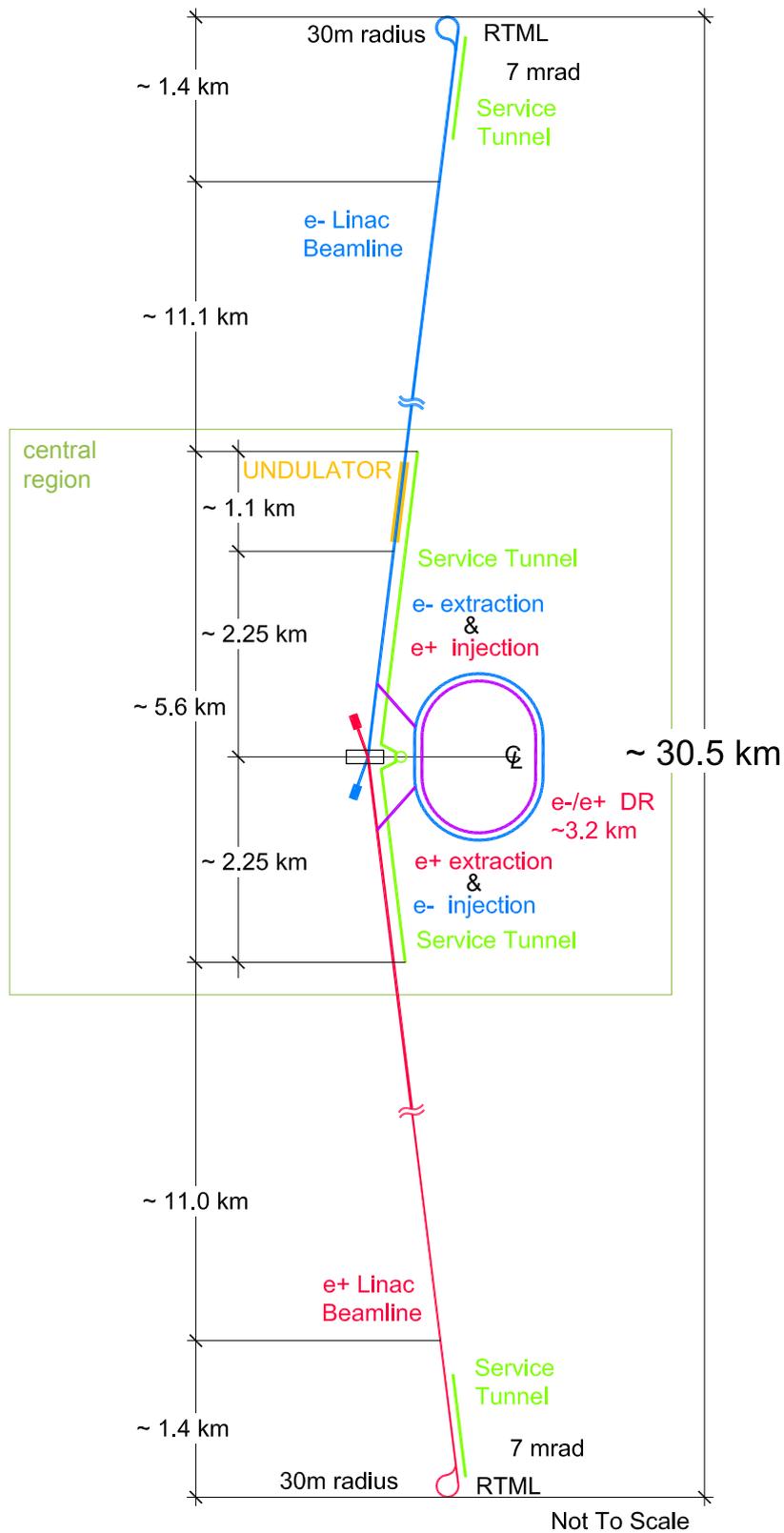


Figure 3.7: Schematic layout of the ILC complex for 500 GeV CM[2]

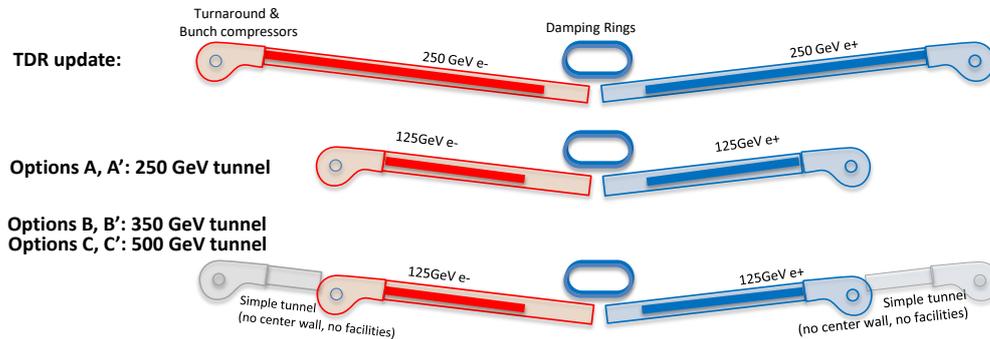


Figure 3.8: Schematic layout of the ILC250GeV staging options

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

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

- 682 • **Damping Rings** In ILC, in order to achieve the design luminosity, the beam emit-
 683 tance has to be lowered by five orders of magnitude. In the central region, two
 684 separate damping rings, one for positron and the other one for electrons with a
 685 circumference of ~ 6.7 km are housed in a single tunnel. A low operation energy
 686 of 5 GeV has been chosen. The frequency of the integrated superconducting RF
 687 system is half the frequency used in the main linac to be able to easily handle
 688 different bunch patterns.
- 689 • **Main Linac** The compressed bunch is ready to enter the Main Linac. At a distance
 690 of about 11 km, the beam particles will be accelerated to 250 GeV in ILC¹. The un-

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

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

- 697 • **Beam Delivery System** After exiting the main linacs the beam enters the Beam
698 Delivery System. One of the first things needed is a measurement of the beam
699 (energy, polarization, and emittance). Corrections are then applied on the way to
700 the Interaction Point (IP), including the removal of the beam halo to avoid large
701 backgrounds in the detector. A fast extraction system can be used to protect the
702 detector and the beam line in case of failure or miss-steered beams.
- 703 • **Machine Detector Interface MDI** In ILC part of the beam delivery system will be
704 integrated into the detector. The beam passes through a conical beam-pipe of min-
705 imal radius, as low as 15 mm at the IP. In the very forward region, sub-detector sys-
706 tems will record remnants of the interaction and monitor beam properties. These
707 detectors will suffer big radiation doses.

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

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

and 5km of each side have been reduced

716 **3.4 The Future Circular Collider (FCC) and High Lumi-** 717 **osity LHC (HL-LHC)**

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

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

756 4.1 Particle Flow Algorithm (PFA) oriented detector

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

762 The requirement of detector for PFA is that it should contain different sub-detectors suit-
763 able for different kind of particles. By combining the information in these sub-detectors,
764 the PFA oriented detector design could significantly enhance the reconstruction effi-
765 ciency of the key physics objects and largely improve the accuracy of jet energy resolu-
766 tion, since the majority of jet energy is stored in the charged hadrons, whose momentum
767 is usually measured with a much better accuracy than its cluster energy measured at the
768 calorimeter system.

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

777 The PFA is widely used in data analyses, both for the existing experiments and for
778 the projects under developments, for highly granular calorimetry and for experiments
779 without highly granular calorimetry. At the LHC, the high granularity calorimetry has
780 already been proposed into CMS (CMS-HGC)[46] and ATLAS (ATLAS-HPTD)[47] as
781 part of their HL-LHC upgrade program. The PFA have already been used in CMS[46],
782 the overall JER takes a value between 6% (at $P_t < 20$ GeV) to 3% (at $P_t > 100$ GeV).
783 The two detector designs for ILC, ILD and SiD are PFA oriented[39]. In CEPC, the
784 baseline of detector (CEPC_v1) takes the ILD as a reference. In order to accommodate
785 the CEPC collision environment, some necessary changes have been made to the sub-
786 detector design. Recently another version of detector (APODIS) has been reported with
787 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 $100 \mu m$ 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.

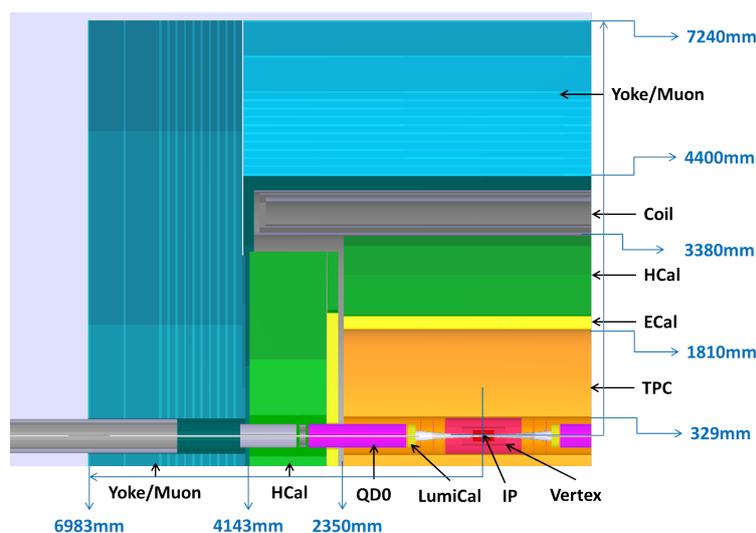


Figure 4.1: Overview of the CEPC detector in the baseline of preCDR.

- **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 ($< 0.15\% X_0/\text{layer}$) 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.

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

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

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

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

- 835 • **ECAL** The particle flow approach requires excellent pattern recognition in the
836 calorimeters to reconstruct individual particles. This is only possible with a short
837 Moliere radius and with a very high granularity, cell sizes inferior to the Moliere
838 Radius. The design of the calorimeters is driven by this goal and not by the op-
839 timization of single particle energy resolutions, although these needs still to be
840 taken into consideration in order to achieve the desired jet energy resolutions.
841 Both the electromagnetic as well as the hadronic calorimeters are planned as sam-
842 pling calorimeters with highly segmented active layers. The materials proposed
843 for the ECAL are tungsten as absorber and silicon as active material. It has a high
844 longitudinal (30 layers, $24 X_0$) as well as transversal segmentation ($5 \times 5mm^2$ cell
845 size), as shown in Fig. 4.2. Alternative designs include signal collection in scin-
846 tillators, implemented as strips with alternating orientation to match effectively
847 the separation capabilities of smaller area square cells, as well as a concept for a

848
849

digital ECAL, realized with Monolithic Active Pixel Sensors (MAPS). Pixel-sizes in the order of $50 \mu\text{m}$ can ensure linearity up to high energies, leading to a total number of pixels of the order of 10^{12} for the complete ECAL.

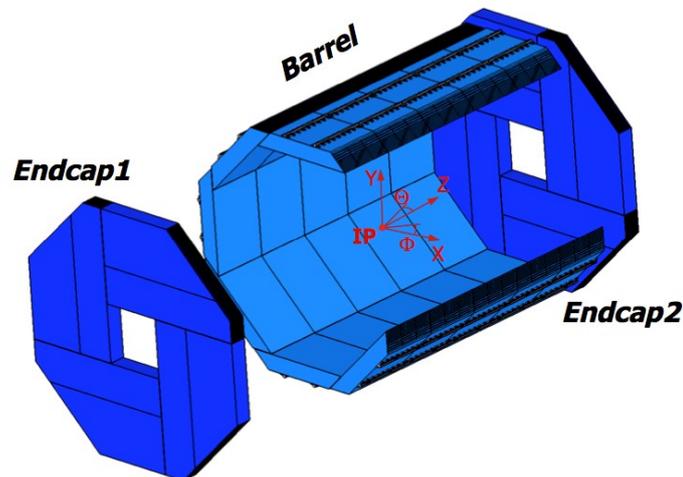


Figure 4.2: The electromagnetic calorimeter within the CEPC Detector.

850

851
852
853
854
855
856

- **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.

857
858
859
860
861
862
863

- **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.

864
865
866

- **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 trigger for high energy muons.

867 Chapter 5

868 Softwares and Particle Flow Algorithm 869 (PFA)

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

877 5.1 Particle Flow Algorithm

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

883 5.1.1 Jet Energy Resolution

884 A jet is defined as a narrow cone of particles produced by the hadronization of a quark
885 or gluon, it is an important object to be observed in particle physics experiments because
886 of its high production cross section. In the traditional calorimetry, jet energy is obtained

887 from the sum of energies deposited in ECAL and HCAL, pointing to a jet energy res-
 888 olution with a stochastic term greater than 60%[50], which does not allow to separate
 889 the hadronic decays of W and Z and does not meet the requirements of ILC and CEPC.
 890 In PFA, a jet is the sum of the individual particles divided into three part: charged par-
 891 ticles whose momenta are measured in the tracking detectors (providing a momentum
 892 resolution as good as $\sigma_{tracker} \sim 5 \cdot 10^{-6} p_T^2$), photons whose energies are best measured
 893 in ECAL (with energy resolution typically of $\sigma(E)/E \sim 0.16/\sqrt{E}$) and neutral hadrons
 894 whose energy obtained from the HCAL (with energy resolution of $\sigma(E)/E \sim 0.5/\sqrt{E}$).
 895 Since the average jet energy content is of 65% from the charged track(s), 26% from the
 896 photon(s) and 9% from neutral hadron(s), the HCAL which has the worst resolution
 897 used to measure only less than 10% of the energy in the jet. Thus the energy resolution
 898 of a jet can be as good as needed. Since $\sigma(E)/E = a/\sqrt{E} \oplus b/E \oplus c$ where a/\sqrt{E} ,
 899 b/E and c are the stochastic response, electronic noise term and constant term caused
 900 by dead material, the assumption that the constant term for ECAL and HCAL to be 1%
 901 and 2% can be made (more dead zones in HCAL), while the noise term for ECAL and
 902 HCAL assumed to be $0.3/E$ and $0.1/E$ (more electronics in ECAL). Taking the above
 903 resolutions as hypothesis, one can see in Figure 5.1 and Figure 5.2 that the tracker mea-
 904 surement would only be beaten by calorimeters for particles above 500 and 700 GeV for
 905 electrons and hadrons, which is not the case in 250 GeV e^+e^- colliders, see Figure 5.3.

906 5.1.2 PandoraPFA

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

910 1) **Track topology** Tracking is done separately in PandoraPFA, track topologies of neu-
 911 trals in the detector volume are identified and classified according to their ways of de-
 912 cays, and they are projected onto the front face of the ECAL.

913 2) **Calorimeter Hit Selection and Ordering** Isolated hits defined by proximity to oth-
 914 ers in the calorimeter are removed at this stage, and the selected hits are stored with
 915 four-vector information after calibration, geometry, isolation, MIP identification and or-
 916 dering.

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

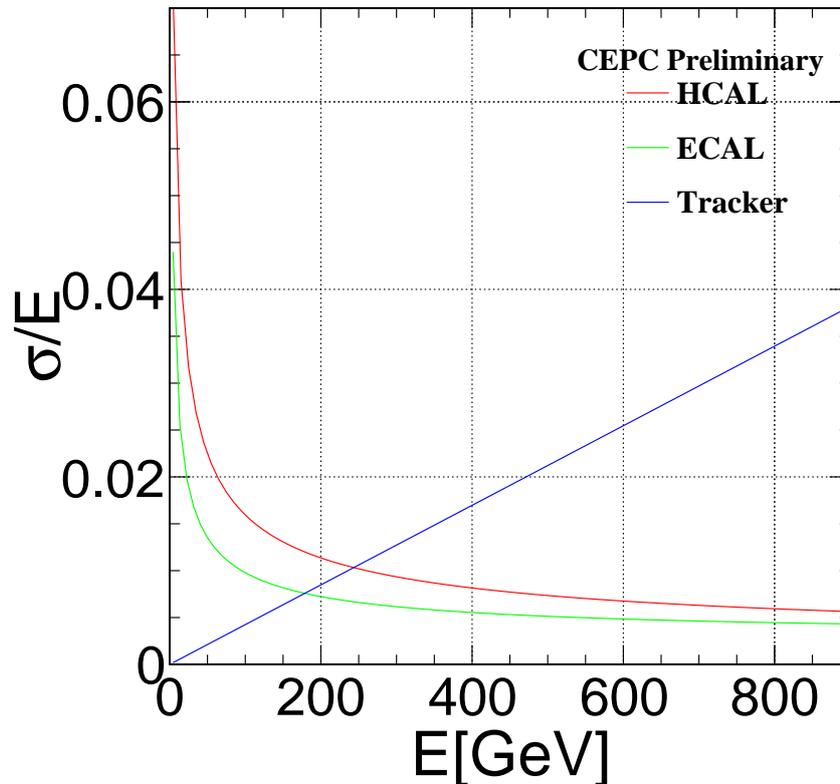


Figure 5.1: The energy resolution of TPC, ECAL and HCAL at different energy (for a direction perpendicular to the magnetic field).

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

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

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

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

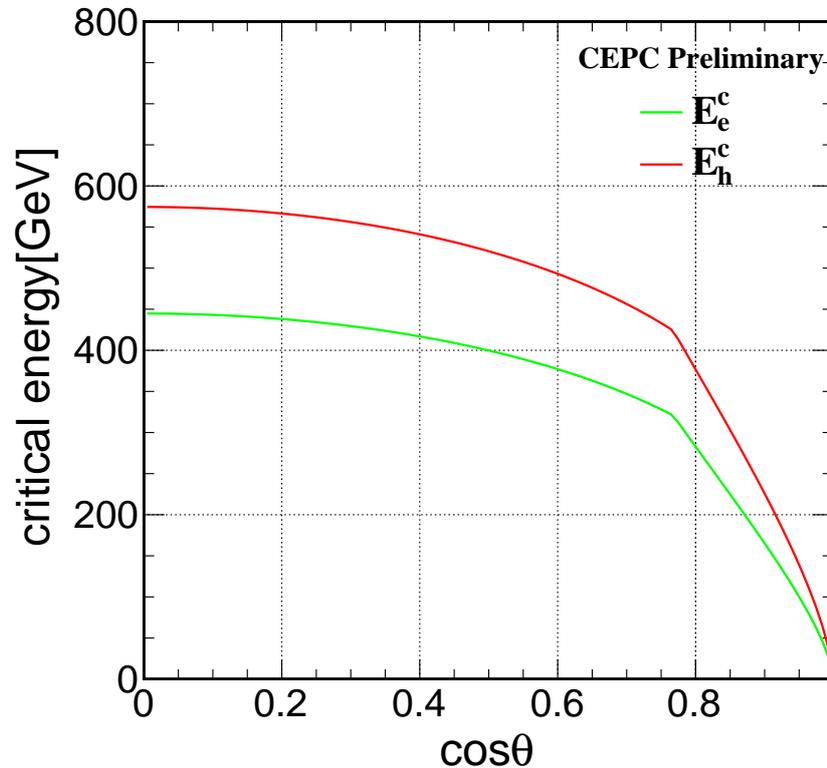


Figure 5.2: The critical energy where the energy resolution of ECAL or HCAL is the same as TPC for different direction.

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

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

945 5.1.3 Arbor

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

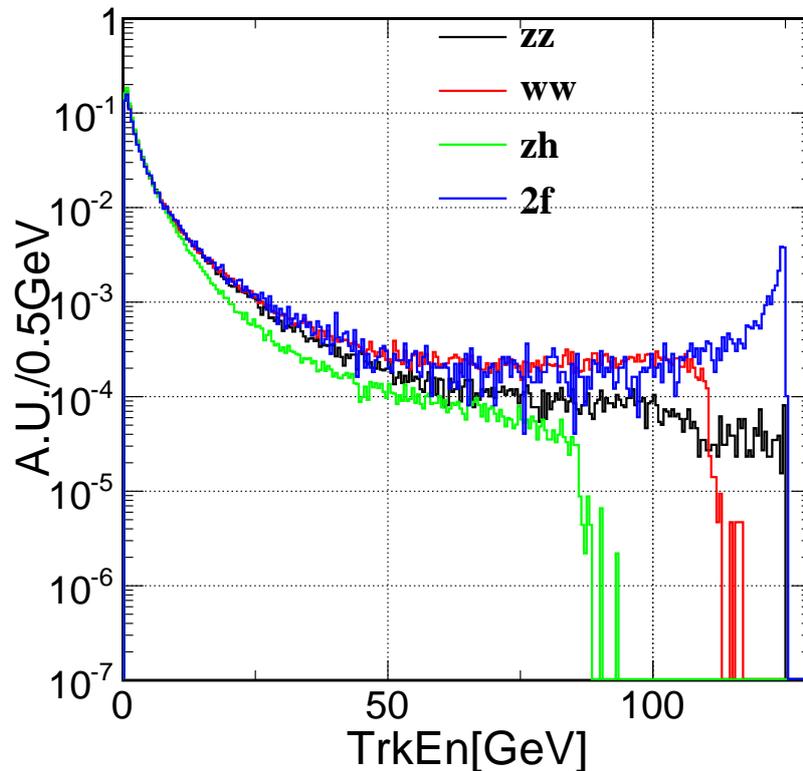


Figure 5.3: Charged particle energy spectra for different physics processes with different final states: ZH , WW , ZZ , or 2 fermions events at center of mass 250 GeV

951 The steps to reconstruct particle flow in Arbor is:

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

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

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

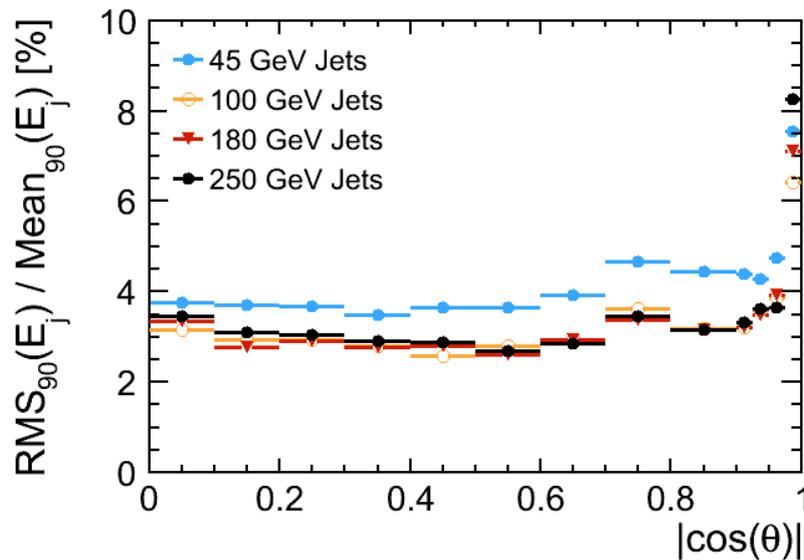


Figure 5.4: Jet Energy Resolution (JER) of PandoraPFA for various angles and energies for ILC using $Z \rightarrow uds$ samples. For $\cos(\theta) < 0.95$ and energy > 45 GeV, it meets the requirement of separating W and Z bosons. The JER is expressed in RMS_{90} , the RMS in the smallest range of reconstructed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

966 lowance for long connectors.

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

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

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

- 973 • The separation performance, i.e., to successfully reconstruct nearby incident par-
 974 ticle.
- 975 • The jet reconstruction performance.

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

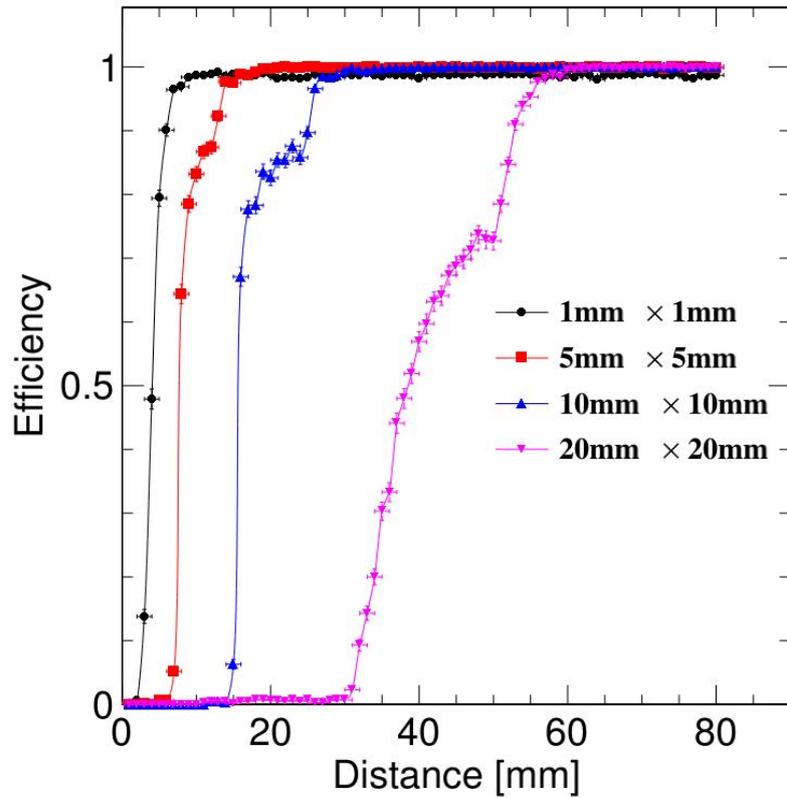


Figure 5.5: Reconstruction efficiency depending on distance of the di-photon system. The different lines corresponds to different ECAL cell sizes. The efficiency is defined as the probability of successfully reconstructing two photons with anticipated energy and incident positions.

980 BMR(Boson Mass Resolution), the resolution of the mass of Higgs boson in $\nu\nu H$ with
 981 $H \rightarrow qq$ events is used as a standard expression of performance in CEPC. In order
 982 to focus on the performance of the detectors or reconstructions, the events with ISR
 983 photons, with neutrinos from Higgs, or with jets shooting to the endcaps are not taken
 984 into account. As shown in Figure 5.7.

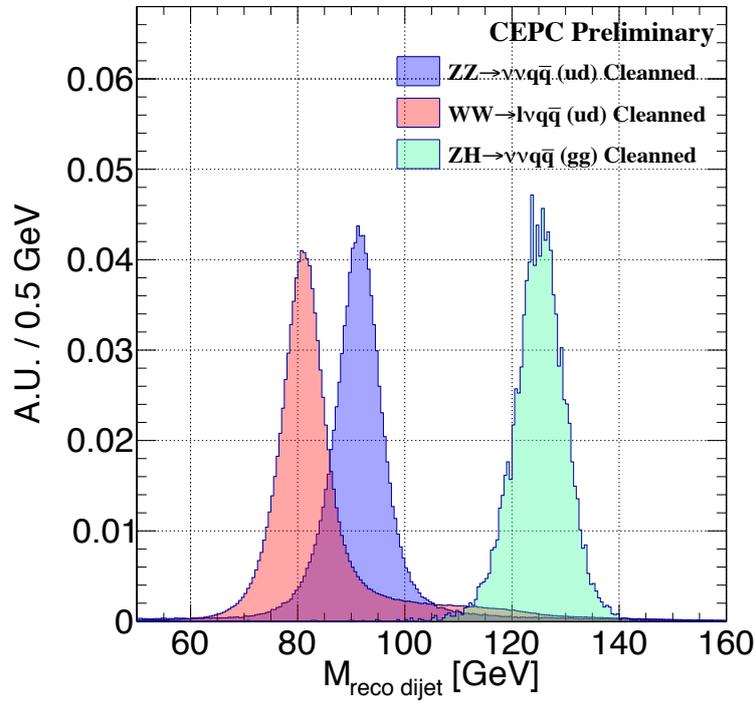


Figure 5.6: Reconstructed boson masses from cleaned $\nu\nu$ events, $lvq\bar{q}$ events and $\nu\nu H$ with $H \rightarrow qq$ events. Here only events with final state jets to be fragmented from either light flavor quarks or gluons are taken into account. The events with ISR are also excluded.

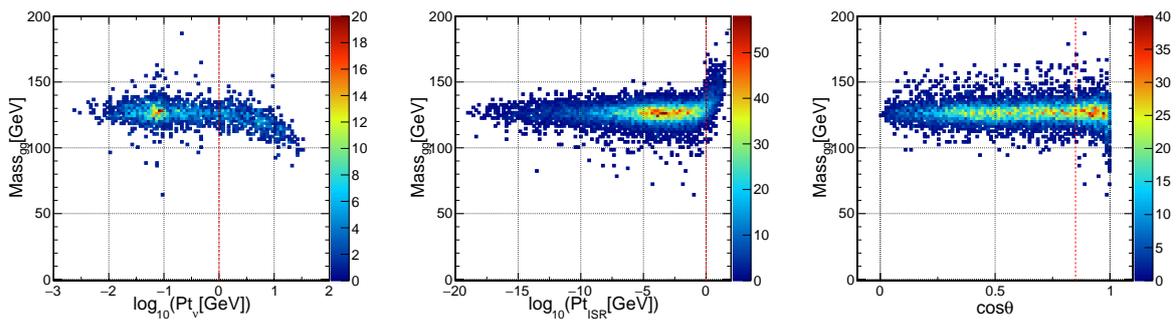


Figure 5.7: Reconstructed boson masses from $\nu\nu H$ with $H \rightarrow qq$ events depending on P_t of ISR, P_t of Higgs decayed neutrino, and $\cos\theta$ where θ is the maximum angle of the jet direction. The dashed line shows the event selection for BMR.

985 5.2 Tools

986 5.2.1 LCIO[3]

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

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

995 5.2.2 Simulation

996 The tool applied in this report for simulation is MOKKA[51], based on GEANT4[52].
997 In order to run Mokka, the first step is to set up the environment parameters, defining
998 the global environment variables such as the working directory, where GEANT4 is in-
999 stalled, the implementation of the Mokka database, the installation of LICO and GEAR,
1000 as well as the shared libraries path to be scanned when running Mokka. After Mokka
1001 is built, a steering file containing the information of the simulation should be prepared.
1002 This file defines the database and user to obtain the geometry information, the output
1003 files, the Macro file to give commands, detector mode (one can change the geometry of
1004 detector by removing subdetectors) and so on. The physics list (see GEANT4) is also
1005 chosen in this file, which is used to describe the modeling of the interaction of high
1006 energy hadrons, here QGSP. In the Macro file, the information of the particles can be
1007 generated from particle gun (where the particle type, position, direction, smearing and
1008 others are set) or by events generated from elsewhere (from HEPevt input file), the en-
1009 ergy and events number of simulation are also defined in this file.

1010 5.2.3 Marlin Framework[4]

1011 The software tool used for full simulation is Mokka, based on Geant4, which can write
1012 an LCIO file defining the parameters for subdetectors. After the generation of the
1013 events, and the simulation of the detector response using MOKKA, reconstruction soft-
1014 ware is used to reconstruct and analyze the events. In order to identify individual par-
1015 ticles, new tools for reconstruction are required.

1016 Marlin(Modular Analysis and Reconstruction for the LINear collider) is a modular C++
1017 application framework for ILC detector reconstruction and analysis LCIO data. Marlin
1018 is first configured by an XML steering file containing parameters defined for individual
1019 processors or globally, the order in which the processors are called and the conditions
1020 applied to *Processors* (plug-in modules that can be loaded at runtime to implement
1021 some core functionality) evaluating with the runtime. The LCIO files, which contain
1022 data such as hits, tracks, and clusters, will be used by processors according to the need
1023 for reconstruction.

1024 5.3 Detector optimization

1025 The optimization of detectors for CEPC and ILC is a balance between the budget and
1026 the performance. In this section, two examples of optimization using the tools above
1027 will be shown.

1028 5.3.1 ECAL optimization

1029 The cost of detectors for CEPC and ILC is always a matter to consider. Therefore op-
1030 timization is ongoing to reduce the price and maintain good performance at the same
1031 time. The ECAL is the major cost of ILD, because of the high price of silicon wafers.
1032 This provides options to optimize, such as the inner radius of ECAL, the number of Si
1033 layers in the ECAL, etc. In this section, the performances of modified detector with a
1034 reduced radius and number of Si layers in ECAL is studied. The detector model used
1035 here is an ILD detector with the TPC radius reduced from 1800 to 1400mm (the length
1036 is modified accordingly), and the ECAL layer number reduced from 30 layers to 26/20
1037 layers. The total absorber thickness, the ratio of W thickness between inner and outer
1038 absorber layers, carbon fiber, cooling layers, Si thickness, etc., remain the same for the
1039 three models. The $Z \rightarrow qq$ events with the centre of mass energy range from 91GeV
1040 to 500GeV are generated and reconstructed with PandoraPFA, after calibration to set
1041 the digitization constant depending on different sampling fraction in the ECAL of each
1042 model. The resolution is expressed with RMS90, defined as the RMS in the smallest
1043 range of reconstructed energy which contains 90% of the events, in order to handle
1044 properly the non-Gaussian energy distribution with a tail corresponding to the pop-
1045 ulation of events where the confusion is significant. As shown in Figure 5.8, the JER
1046 increases 10% to 91 GeV di-jets and less than 5% for 100 GeV di-jets by decreasing the
1047 number of Si layers from 30 to 20. At the 250GeV e^+e^- colliders the typical jet energy is
1048 less than 70GeV, as shown in Figure 5.9, corresponding to the 91 GeV di-jets.

1049 A comparison with Arbor using the invariant mass resolution of 250GeV $\nu\nu gg$ events is

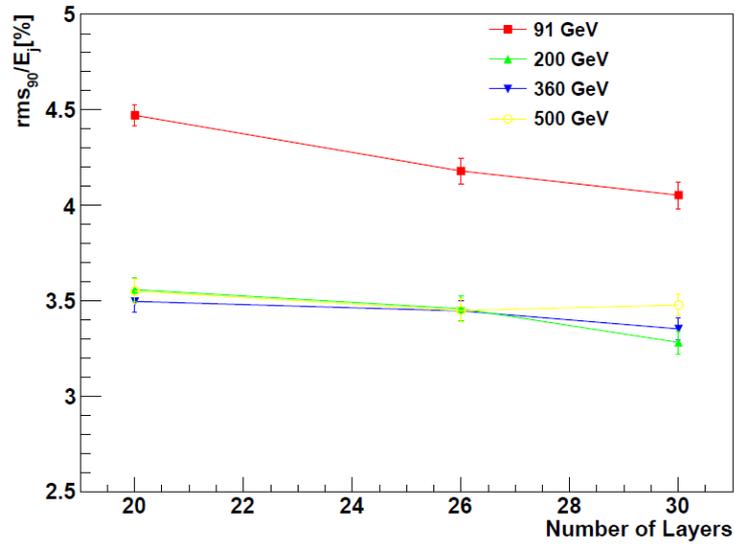


Figure 5.8: JER comparison for different jets energy in function of layer numbers, a cut $|\cos(\theta_{jet})| < 0.7$ is applied to avoid the endcap area.

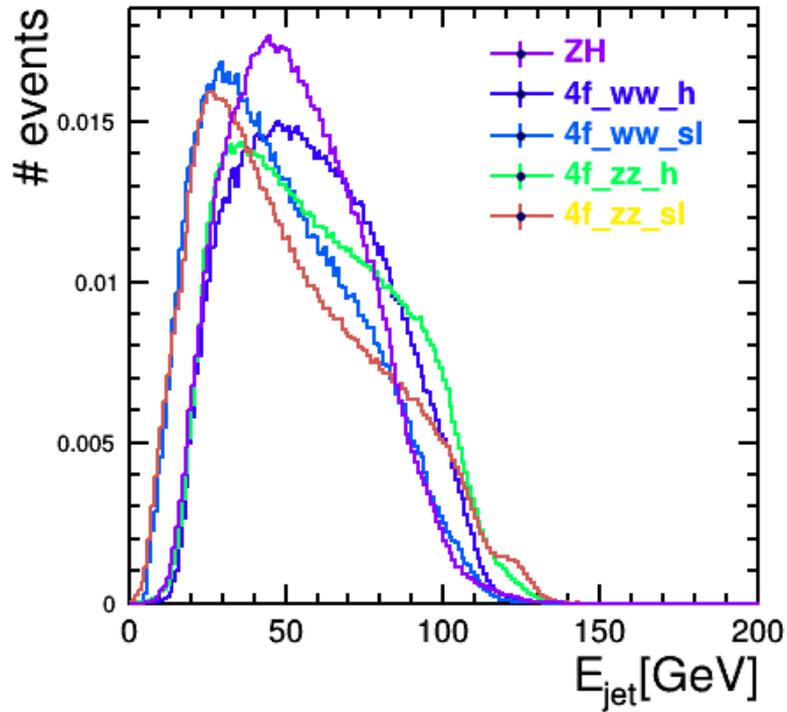


Figure 5.9: jet energy spectra for different physics processes with different final states: ZH or 4 fermions from W bosons or Z bosons decays, at center of mass 250 GeV

1050 shown in Figure 5.10, with the resolution expressed in BMR and not only the number
 1051 of layers but also the ECAL cell size is taken into account. Since the total number of
 1052 readout will decrease with the cell size, the cooling system might be inactive if the cell
 1053 size enlarged. The events with ISR and events with jet direction to the endcaps are
 1054 excluded. It is shown that the degradations of performance using the two frameworks
 are similar to each other.

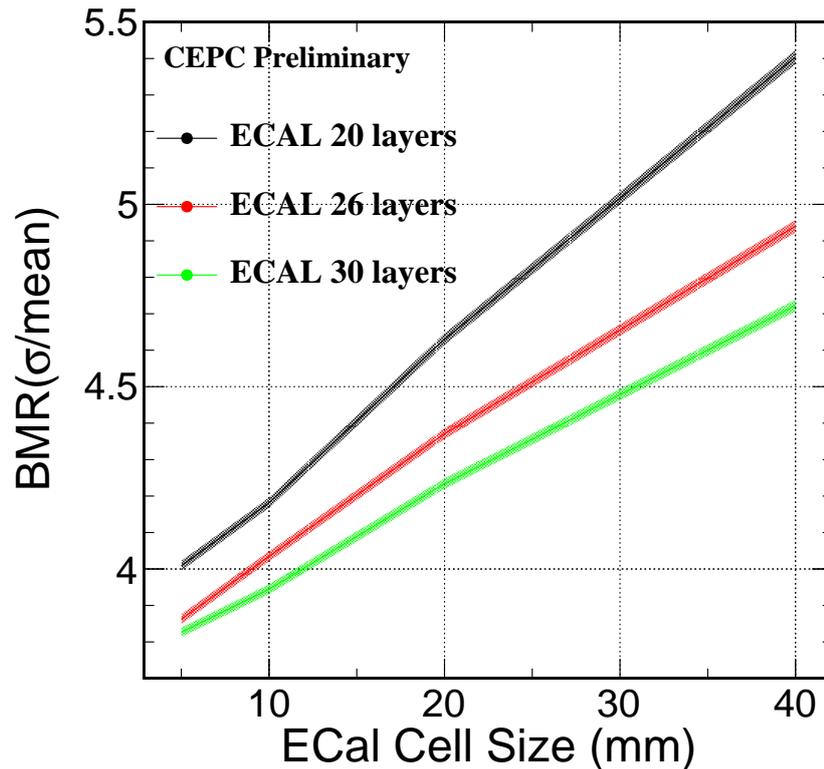


Figure 5.10: The invariant mass resolution of 250GeV $\nu\nu gg$ events in CEPC for different number of ECAL layers and different ECAL cell sizes.

1055

1056 5.3.2 HCAL and B field optimization

1057 For HCAL, the optimization is done for a reduced number of layers while the thickness
 1058 of each layer remains the same. The B field is allowed to be reduced because of the high
 1059 granularity. The $\nu\nu gg$ events are generated in CEPC detector with HCAL layers range
 1060 from 20 to 48 and B field to be (2.5T, 3.0T, 3.5T) and reconstructed with Arbor (v3.3). The
 1061 resolution is expressed as the resolution of the reconstructed invariant mass, with final
 1062 state jets from either light flavor quarks or gluons and the events with ISR excluded. As

1063 shown in Figure 5.11 and Figure 5.12, the resolution degrades by 0.1 while the number
 1064 of layers reduces from 48 to 20. This result also leaves an opportunity to degrade the
 1065 B field in CEPC to 3 Tesla, which is appreciated by the MDI and will be applied for
 1066 the baseline of CDR. In the new version of CEPC detector, the baseline of HCAL layer
 1067 number is chosen to be 40.

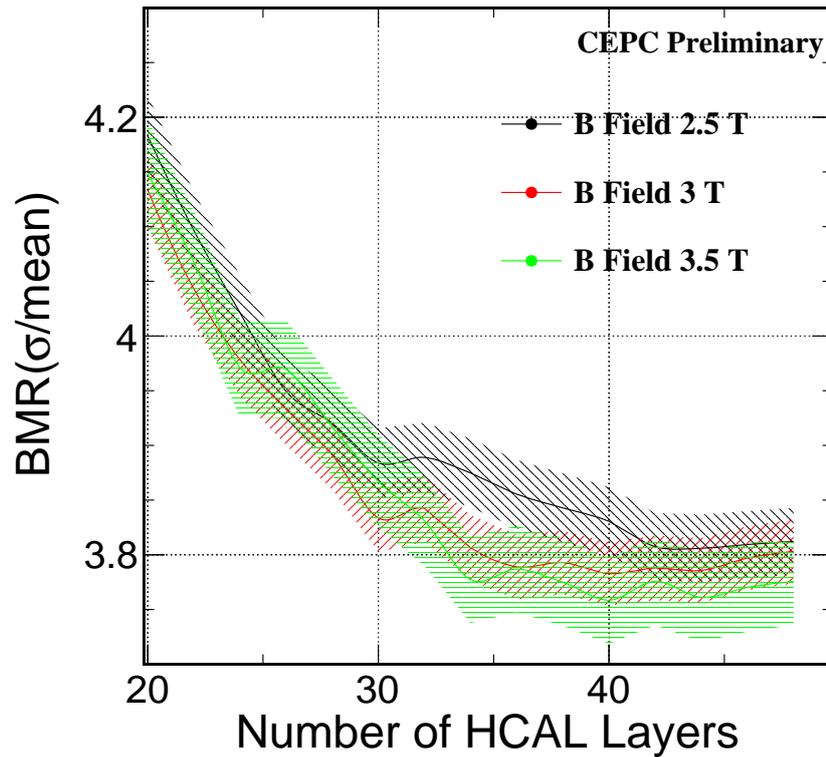


Figure 5.11: The invariant mass resolution of 250GeV $\nu\nu gg$ events in CEPC for different number of HCAL layers.

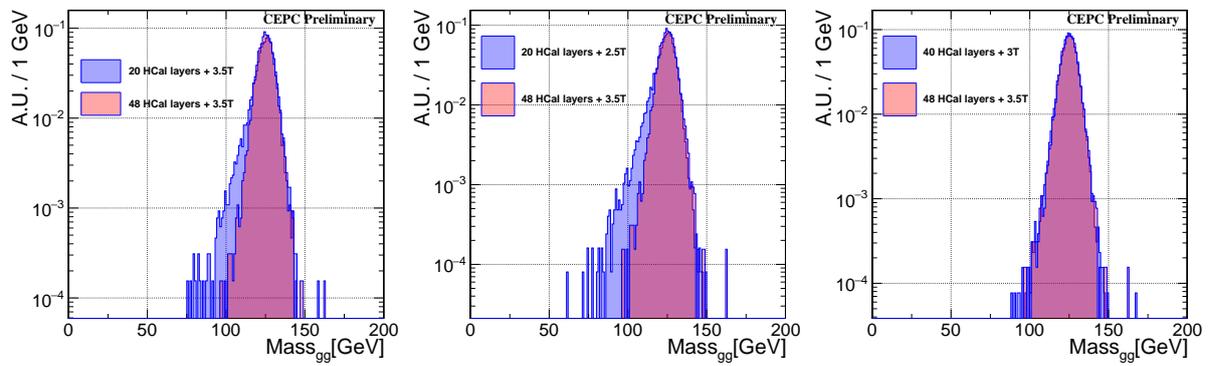


Figure 5.12: The Higgs boson invariant mass for 250 GeV $\nu\nu qq$ events, with different B fields and different HCAL layer numbers, comparing with the baseline geometry in preCDR. The last plot is the baseline for CDR.

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, $H \rightarrow WW^* \rightarrow ll\nu\nu/l\nu qq$, $H \rightarrow ZZ^* \rightarrow llqq$, $H \rightarrow \tau\tau$, $H \rightarrow \mu\mu$, etc. The SM Higgs also has a branching ratio $\text{Br}(H \rightarrow bb) = 58\%$, 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 ($\text{Br}(Z \rightarrow ee)$ and $\text{Br}(Z \rightarrow \mu\mu) = 3.36\%$). 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

1091 identification performance from the effect of PFA reconstruction and geometry defects,
 1092 we select those events where only one charged particle is reconstructed. The total num-
 1093 ber of these events is recorded as $N_{1Particle}$, and the number of these events identified
 1094 with correct particle types is recorded as $N_{1Particle,T}$. The performance of lepton identifi-
 1095 cation is then expressed as a migration matrix in Table 6.1, its diagonal elements ϵ_i^i refer
 1096 to the identification efficiencies (defined as $N_{1Particle,T}/N_{1Particle}$), and the off diagonal
 1097 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	μ^-like	π^+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

1098 6.2 Discriminant variables and the output likelihoods

1099 LICH takes individual reconstructed charged particles as input, extracts 24 discriminant
 1100 variables for the lepton identification, and calculates the corresponding likelihood to
 1101 be an electron or a muon. These discriminant variables can be characterized into five
 1102 different classes:

1103 • dE/dx

1104 For a track in the TPC, the distribution of energy loss per unit distance follows a
 1105 Landau distribution. The dE/dx estimator used here is the average of this value
 1106 but after cutting tails at the two edges of the Landau distribution (first 7% and last
 1107 30%). The dE/dx has a strong discriminant power to distinguish electron tracks
 1108 from others at low energy (under 10 GeV) (Figure 6.1).

1109 • Fractal Dimension

1110 The fractal dimension (FD) of a shower is used to describe the self-similar behav-
 1111 ior of shower spatial configurations, following the original definition in [54], the
 1112 fractal dimension is directly linked to the compactness of the particle shower. The
 1113 FD of a shower is expressed as $FD_\beta = \langle \log()R_{\alpha,\beta}/\log\alpha \rangle + 1$ where $R_{\alpha,\beta} = N_\beta/N_\alpha$
 1114 represents the ratio of the number of hits at different scales. Here β range from
 1115 10mm to 150mm and α is 10mm.

1116 At a fixed energy, the EM showers are much more compact than the muon or
 1117 hadron shower, leading to a large FD. The muon shower usually takes the config-

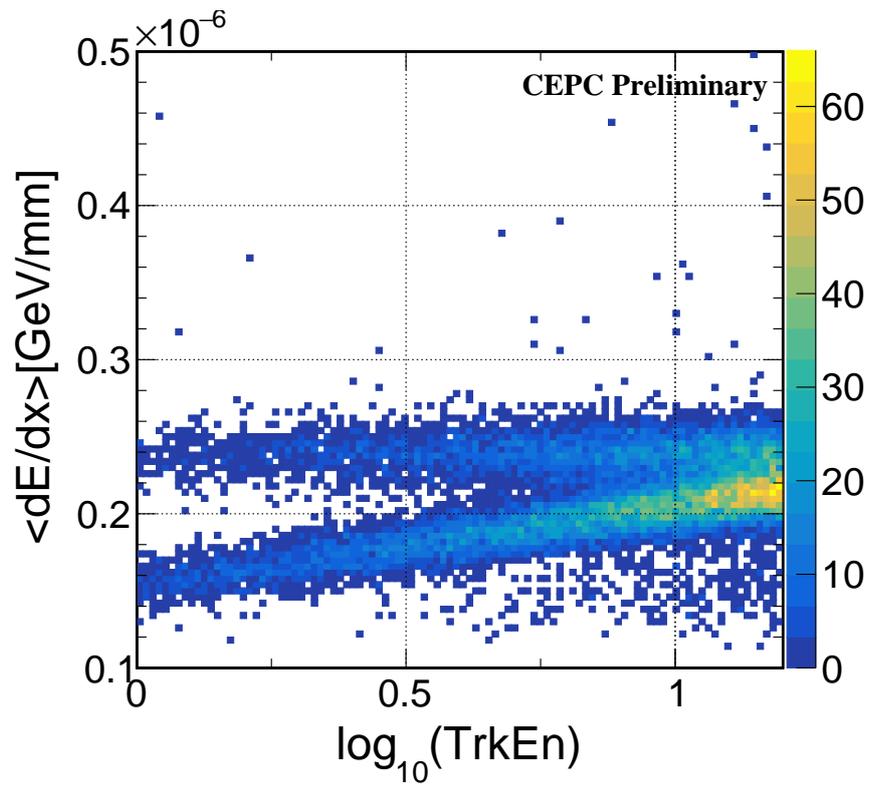


Figure 6.1: dE/dx for e^- , μ^- and π^+ , for electrons it is stable around 2.4×10^{-7} , for muon and pion it is smaller at energy lower than 10 GeV and after that they start mixing with electron

1118 uration of a 1-dimensional MIP(Minimum Ionizing Particle) track, therefore has
 1119 an FD close to zero. The FD of the hadronic shower usually lays between the EM
 1120 and MIP tracks, since it contains both EM and MIP components. A typical dis-
 1121 tribution of F_{all} (the fractal dimension using both ECAL and HCAL) for 40 GeV
 1122 showers is presented in Figure 6.2,

1123 For any calorimeter cluster, LICH calculates 5 different FD values: from its ECAL
 1124 hits, HCAL hits, hits in 10 or 20 first layers of ECAL, and all the calorimeter hits.

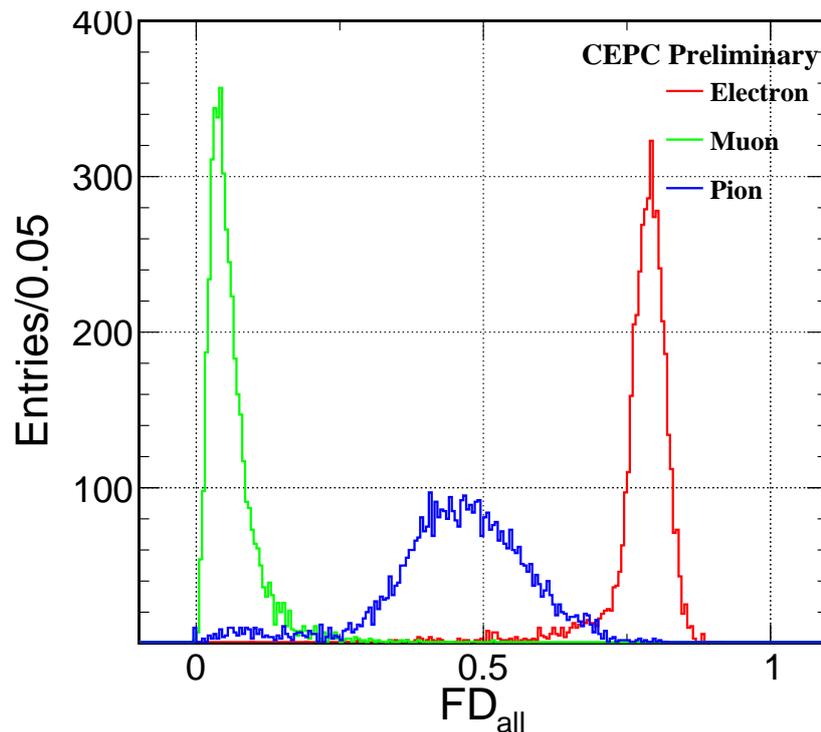


Figure 6.2: Fractal dimension using both ECAL and HCAL for e^- , μ^- and π^+ at 40 GeV

1125 • Energy Distribution

1126 LICH builds variables out of the shower energy information, including the pro-
 1127 portion of energy deposited in the first 10 layers in ECAL to the entire ECAL, or
 1128 the energy deposited in a cylinder around the incident direction with a radius of
 1129 1 and 1.5 Moliere radius.

1130 • Hit Information

1131 Hits information refers to the number of hits in ECAL and HCAL and some other
 1132 information obtained from hits, such as the number of ECAL (HCAL) layers hit
 1133 by the shower, number of hits in the first 10 layers of ECAL.

1134 • **Shower Shape, Spatial Information**

1135 The spatial variables include the maximum distance between a hit and the extrap-
 1136 olated track, the maximum distance and average distance between shower hits
 1137 and the axis of the shower (defined by the innermost point and the center of grav-
 1138 ity of the shower), the depth (perpendicular to the detector layers) of the center of
 1139 gravity, and the depth of the shower defined as the depth between the innermost
 1140 hit and the outermost hit.

1141 The correlations of those variables at energy 40 GeV are summarized in Figure 6.4, the
 1142 definitions of all the variables are:

- 1143 • NH_ECALF10: Number of hits in the first 10 layers of ECAL
- 1144 • FD_ECALL20: FD calculated using hits in the last 20 layers of ECAL
- 1145 • FD_ECALF10: FD calculated using hits in the first 10 layers of ECAL
- 1146 • AL_ECAL: Number of ECAL layer groups (every five layers forms a group) with
 1147 hits
- 1148 • av_NHH: Average number of hits in each HCAL layer groups (every five layers
 1149 forms a group)
- 1150 • rms_Hcal: The RMS of hits in each HCAL layer groups (every five layers forms a
 1151 group)
- 1152 • EEclu_r: Energy deposited in a cylinder around the incident direction with a ra-
 1153 dius of 1 Moliere radius
- 1154 • EEclu_R: Energy deposited in a cylinder around the incident direction with a ra-
 1155 dius of 1.5 Moliere radius
- 1156 • EEclu_L10: Energy deposited in the first 10 layers of ECAL
- 1157 • MaxDisHel: Maximum distance between a hit and the helix
- 1158 • minDepth: Depth of the innermost hit
- 1159 • cluDepth: Depth of the cluster position
- 1160 • graDepth: Depth of the cluster gravity center
- 1161 • EcalEn: Energy deposited in ECAL
- 1162 • avDisHtoL: Average distance between a hit to the axis from the innermost hit and

- 1163 the gravity center
- 1164 • maxDisHtoL: Maximum distance between a hit to the axis from the innermost hit
 - 1165 and the gravity center
 - 1166 • NLHcal: Number of HCAL layers with hits
 - 1167 • NLEcal: Number of ECAL layers with hits
 - 1168 • HcalNHit: Number of HCAL hits
 - 1169 • EcalNHit: Number of ECAL hits

1170 The distribution of all the variables used in TMVA are shown in Figure 6.3

1171 It is clear that the dE/dx , measured from tracks, does not correlate with any other vari-
 1172 ables which are measured from calorimeters. Some of the variables are highly corre-
 1173 lated, such as FD_ECAL (FD calculated from ECAL hits) and EcalNHit (number of
 1174 ECAL hits). However, all these variables are kept because their correlations change
 1175 with energy and polar angle.

1176 LICH uses TMVA[22] methods to combine these input variables into two likelihoods,
 1177 corresponding to electrons and muons. Multiple TMVA methods have been tested and
 1178 the Boosted Decision Trees with Gradient boosting (BDTG) method is chosen for its
 1179 better performance. The e-likeness (L_e) and μ -likeness (L_μ) for different particles in a 40
 1180 GeV sample are shown in Figure 6.5.

1181 The overtraining check of Muon BDT response at 40GeV is shown in 6.6 as an example.

1182 The weight of the 24 variables varies with different energies, at 2GeV the 5 most impor-
 1183 tant variables are: dE/dx , cluDepth, EcalNHit, E_r , and maxDisHtoL, while at 40GeV
 1184 the 5 most important variables are: E_{10} , FD_all, NLEcal, EcalNHit, and avDisHtoL.
 1185 Taking the 5 GeV energy point as an example, the charged particle identification effi-
 1186 ciency 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

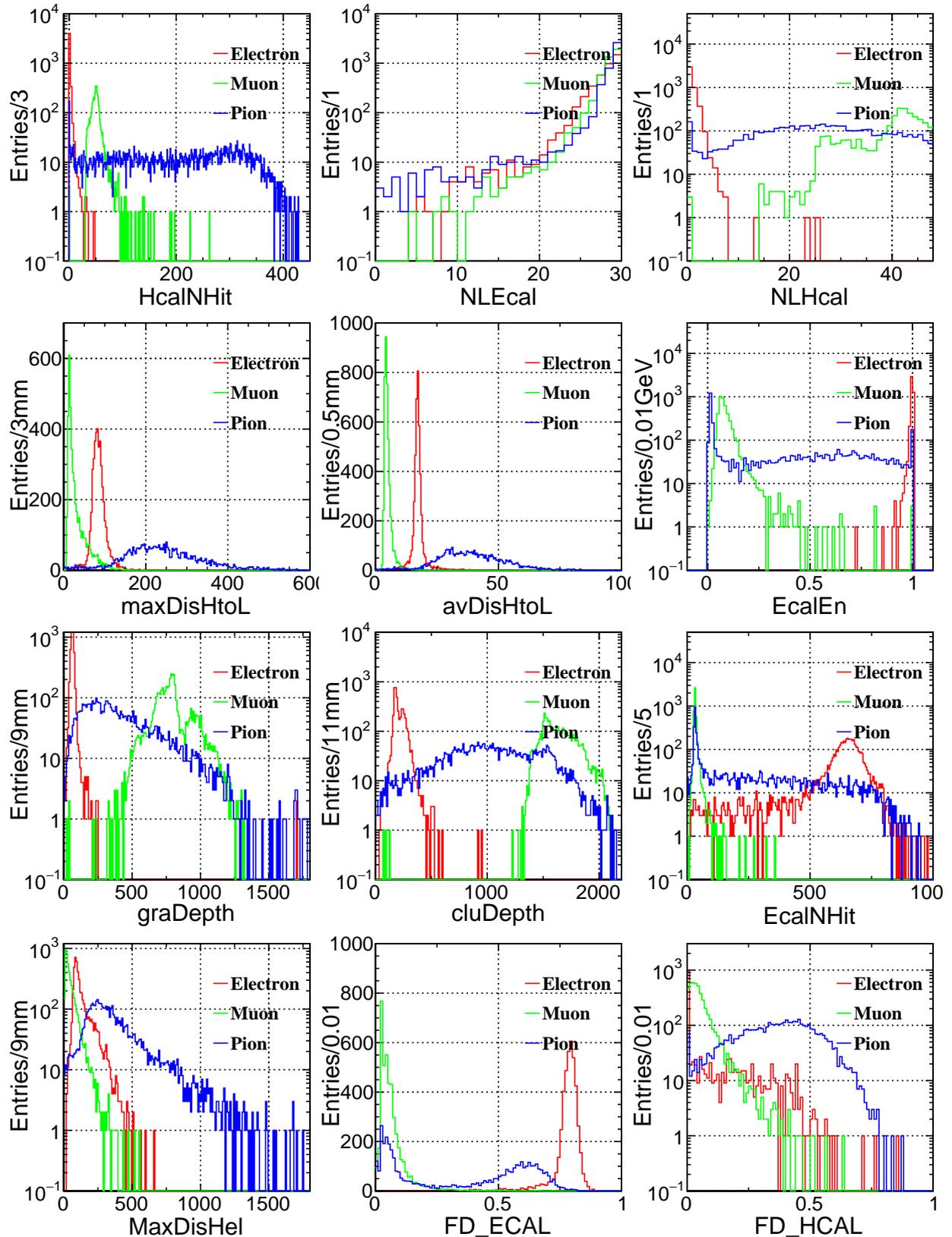


Figure 6.3: Calorimeter based variables used in TMVA (40GeV) (to be continued)

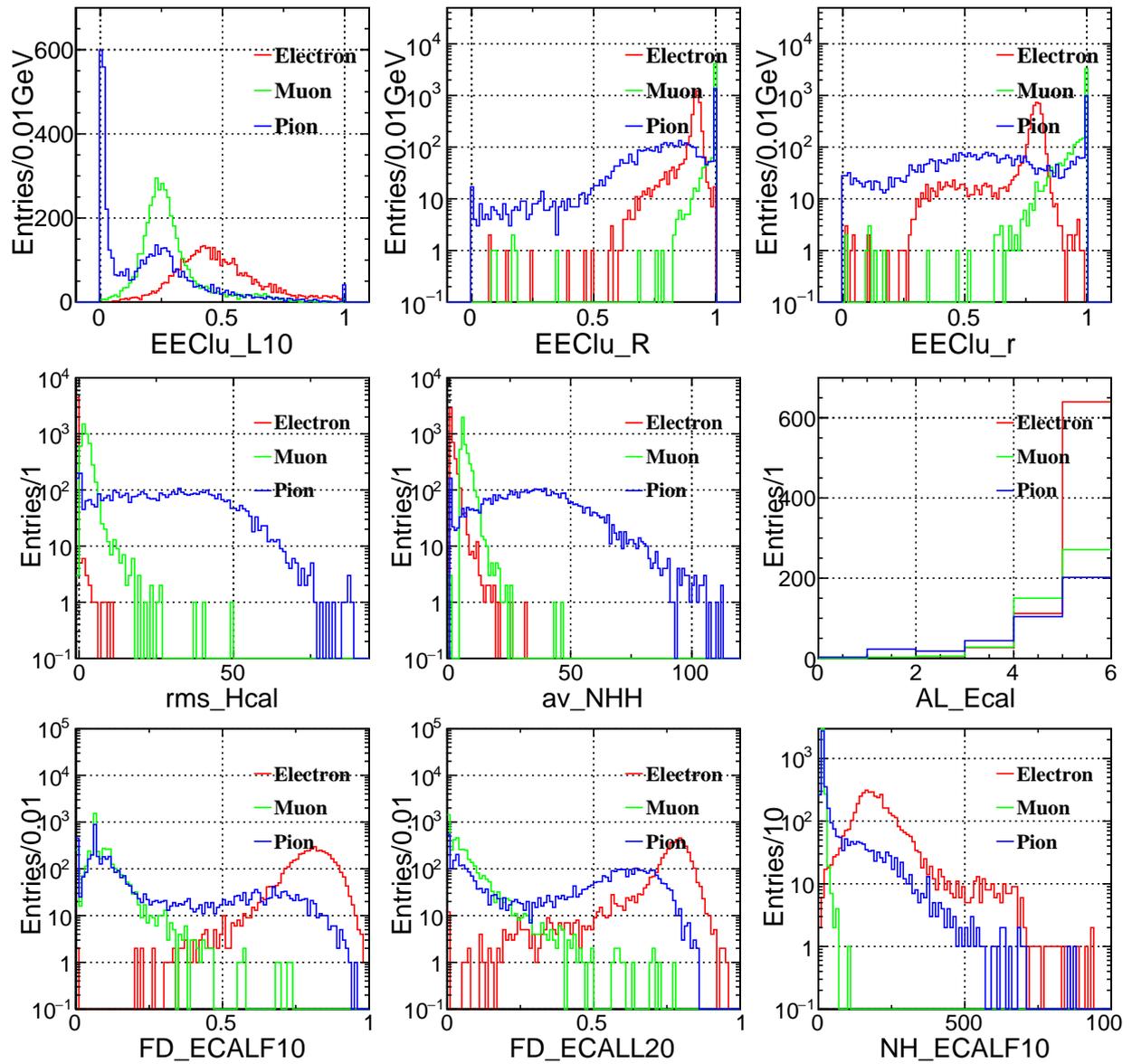


Figure 6.3: Calorimeter based variables used in TMVA (40GeV)

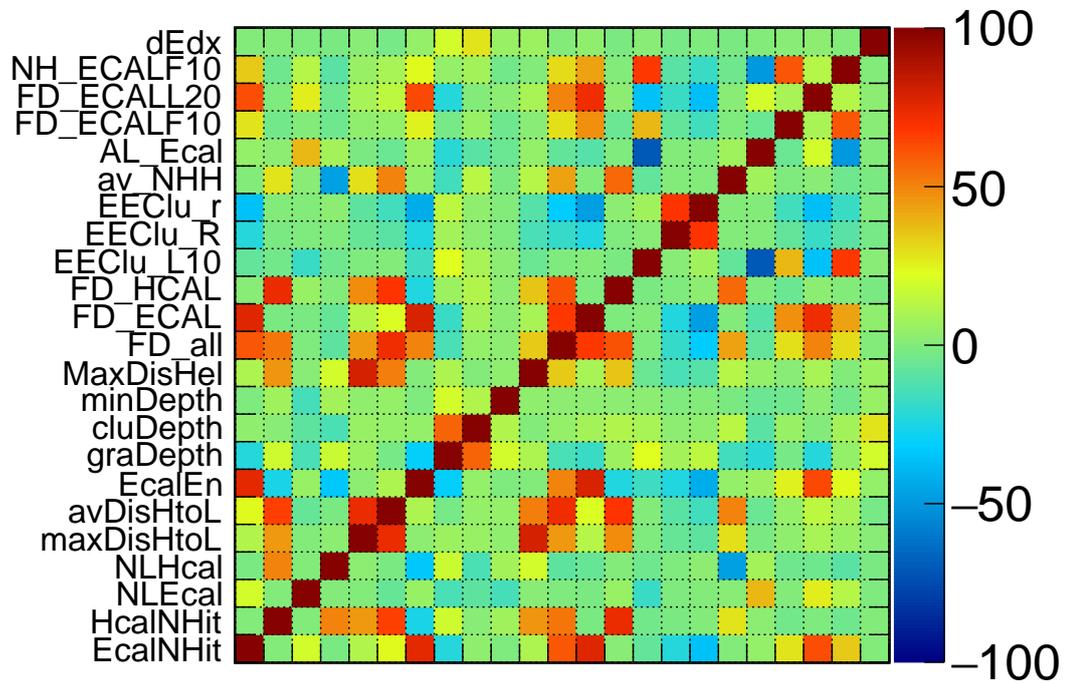


Figure 6.4: The correlation matrix of all the variables

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

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

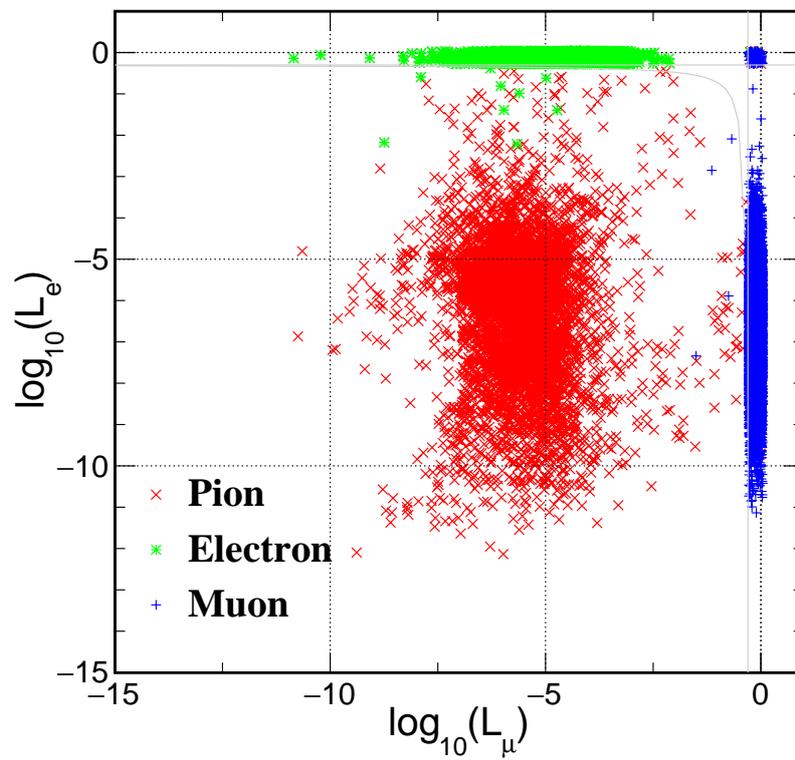


Figure 6.5: The e-likeliness and μ -likeness of e^- , μ^- and π^+ at 40 GeV, grey lines are the cuts for different catalogs in next section

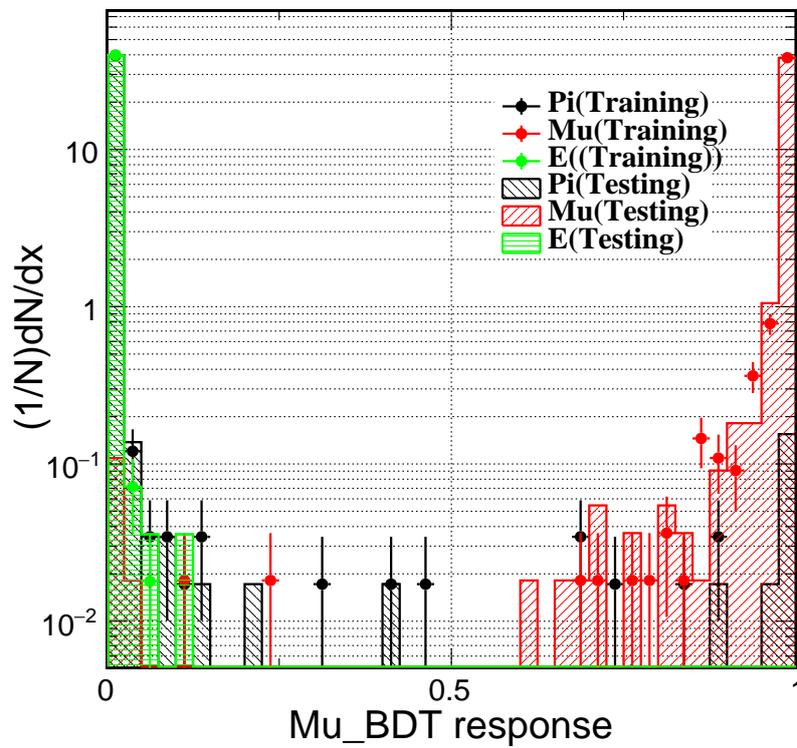


Figure 6.6: Muon BDT response of e^- , μ^- and π^+ at 40GeV (training and test samples)

1200 (θ), the TMVA is trained independently on four subsets:

- 1201 • **barrel 1**: middle of barrel ($|\cos \theta| < 0.3$),
- 1202 • **barrel 2**: edge of barrel ($0.3 < |\cos \theta| < 0.7$),
- 1203 • **overlap**: overlap region of barrel and endcap ($0.7 < |\cos \theta| < 0.8$),
- 1204 • **endcap**: ($0.8 < |\cos \theta| < 0.98$).

1205 Take the sample of 40 GeV charged particle as an example, the migration matrix is
 1206 shown in Table 6.3. Comparing this table to the result of ALEPH for energetic taus[55],
 1207 the efficiencies are improved, and the mis-identification rates from hadrons to leptons
 1208 are significantly reduced.

Table 6.3: Migration Matrix at 40 GeV (%)

Type	e^-like	μ^-like	π^+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

1209 The lepton identification efficiencies (diagonal terms of the migration matrix) at dif-
 1210 ferent energies are presented in Figure 6.7 for the different regions. The identification
 1211 efficiencies saturate at 99.9% for particles with energy higher than 2 GeV. For those with
 1212 energy lower than 2 GeV, the performance drops significantly, especially in **barrel2** and
 1213 **overlap** regions. For the overlap region, the complex geometry limits the performance;
 1214 while for the **barrel2** region, charged particles with $Pt < 0.97$ GeV cannot reach the bar-
 1215 rel, they will eventually hit the endcaps at large incident angle, hence their signal is
 1216 more difficult to catalog.

1217 Concerning the off-diagonal terms of the migration matrix, the chances of electrons to
 1218 be mis-identified as muons and pions are negligible ($P_{\mu}^e, P_{\pi}^e < 10^{-3}$), the crosstalk rate
 1219 P_e^{μ} is observed at even lower level. However, the chances of pions to be mis-identified
 1220 as leptons (P_e^{π}, P_{μ}^{π}) are of the order of 1% and are energy dependent. In fact, these
 1221 mis-identifications are mainly induced by the irreducible physics effects: pion decay
 1222 and π^0 generation via π -nucleon collision. Meanwhile, the muons also have a small
 1223 chance to be mis-identified as pions at energy smaller than 2 GeV. Figure 6.8 shows the
 1224 significant crosstalk items (P_e^{π}, P_{μ}^{π} and P_{π}^{μ}) as a function of the particle energy in the
 1225 endcap region. The green shaded band indicates the probability of pion decay before
 1226 reaching the calorimeter, which is roughly comparable with P_{μ}^{π} .

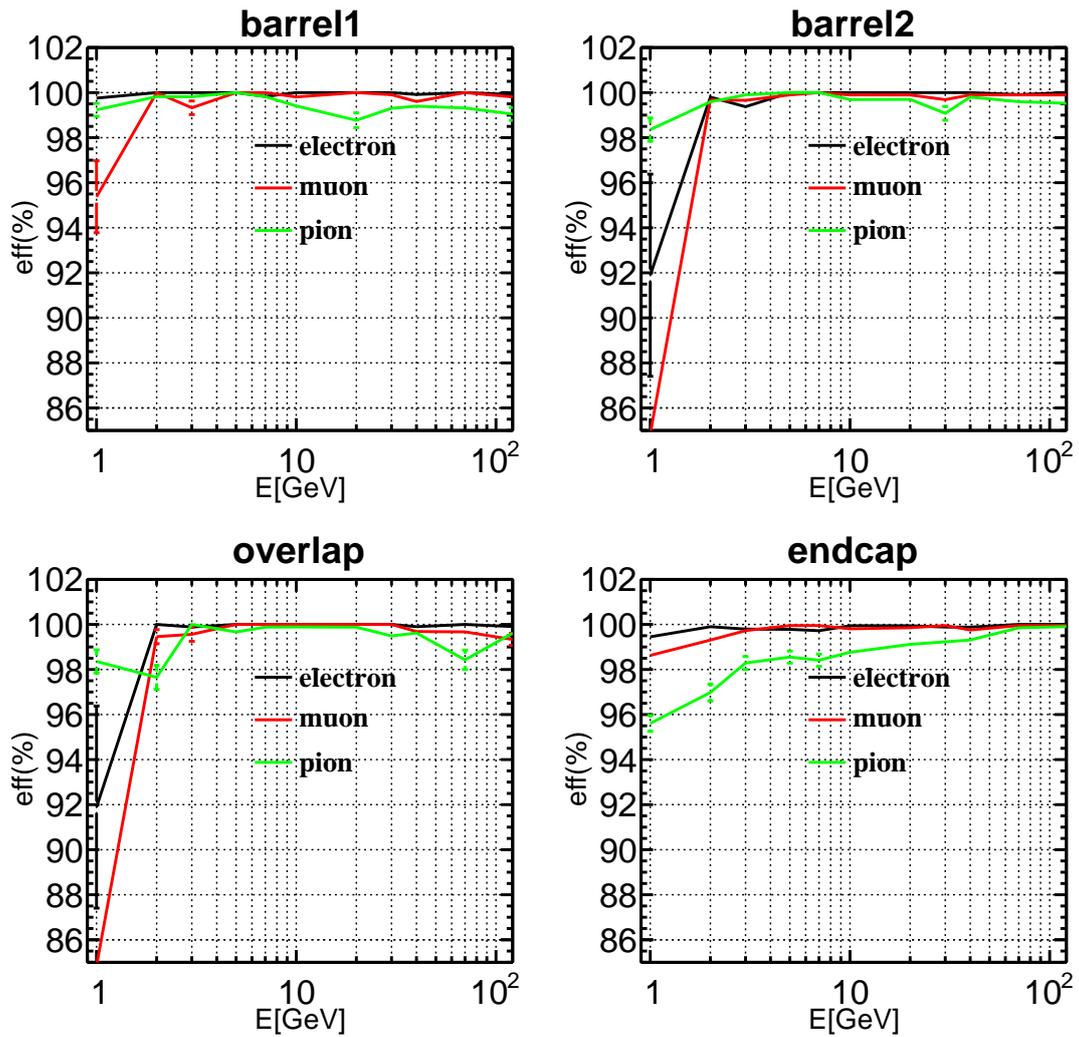


Figure 6.7: The efficiency of lepton identification for e^- , μ^- and π^+ as function of particle energy in the four regions

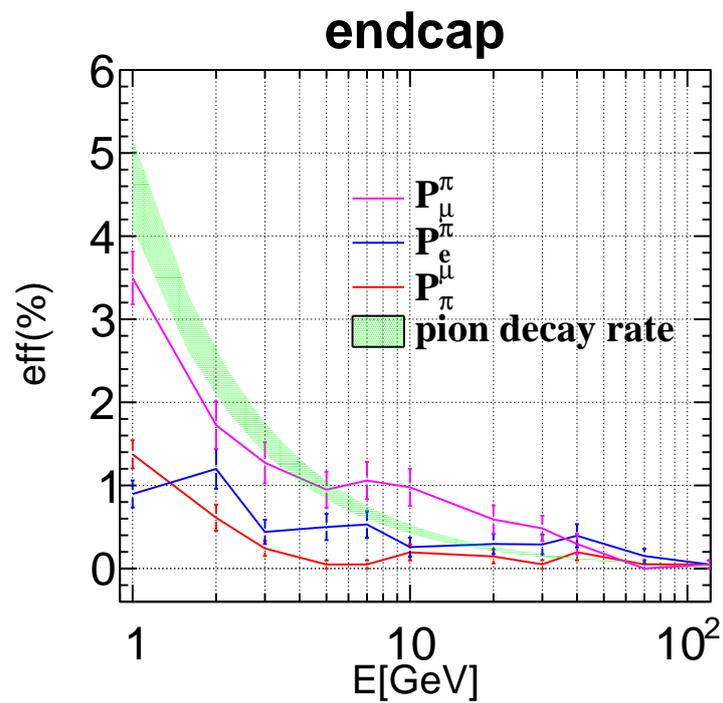


Figure 6.8: The mis-identification rates of lepton identification for μ and π in ~ 5000 events for the endcap region; Pion decay rate band (to account for the polar angle spread) is indicated for comparison

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 = $5 \times 5 \text{ mm}^2$, $10 \times 10 \text{ mm}^2$, $20 \times 20 \text{ mm}^2$, $40 \times 40 \text{ mm}^2$
- HCAL cell size = $10 \times 10 \text{ mm}^2$, $20 \times 20 \text{ mm}^2$, $40 \times 40 \text{ mm}^2$, $60 \times 60 \text{ mm}^2$, $80 \times 80 \text{ mm}^2$

In general, the lepton identification performance is extremely stable over the scanned parameter space. Only for HCAL cell size larger than $60 \times 60 \text{ mm}^2$ 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 ($E \leq 2 \text{ GeV}$), and the identification efficiency of pion degrades slightly over the full energy range, see Figure 6.9 to Figure 6.12.

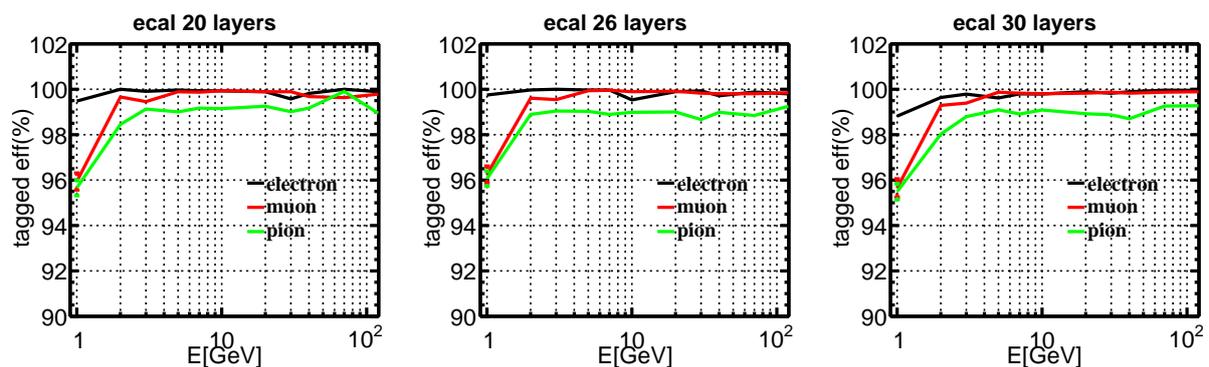


Figure 6.9: The efficiency of lepton identification for different ECAL layer number

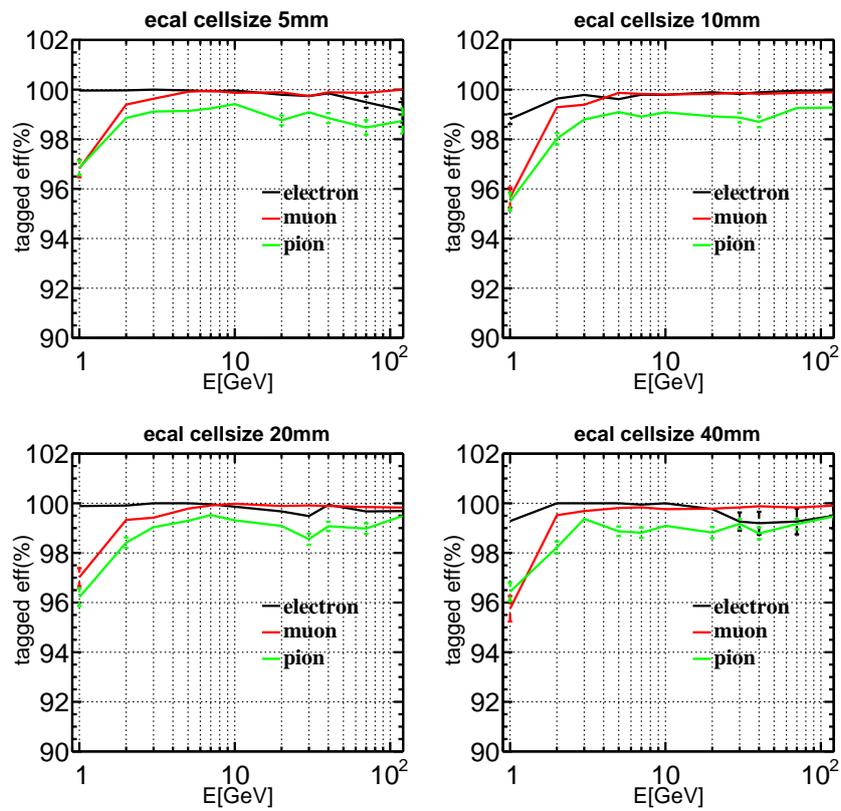


Figure 6.10: The efficiency of lepton identification for different ECAL cell size

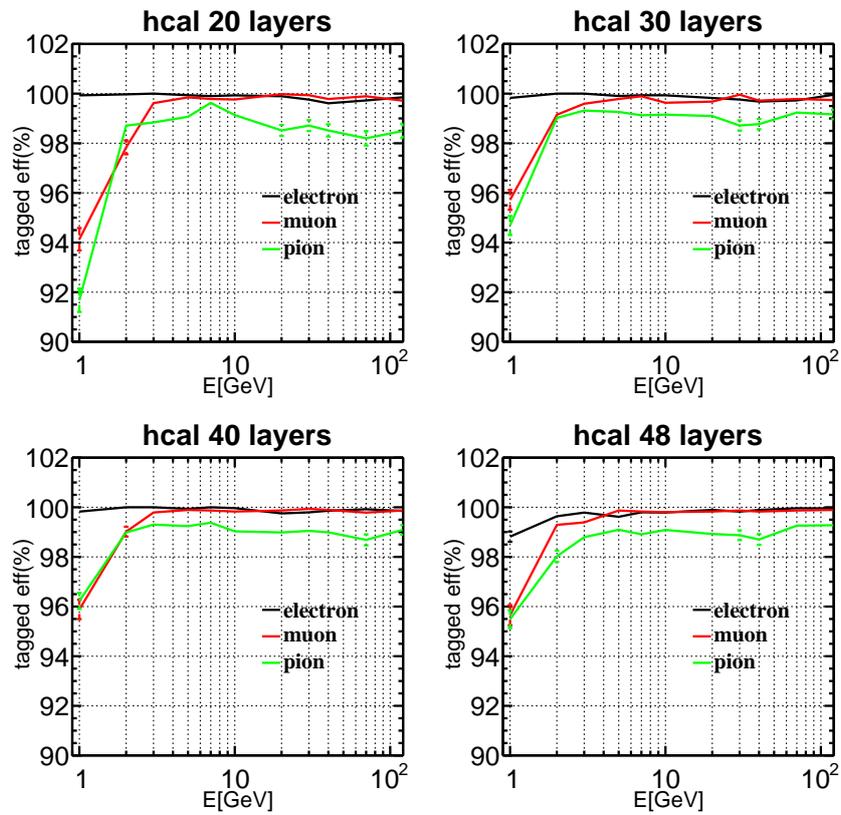


Figure 6.11: The efficiency of lepton identification for different HCAL layer number

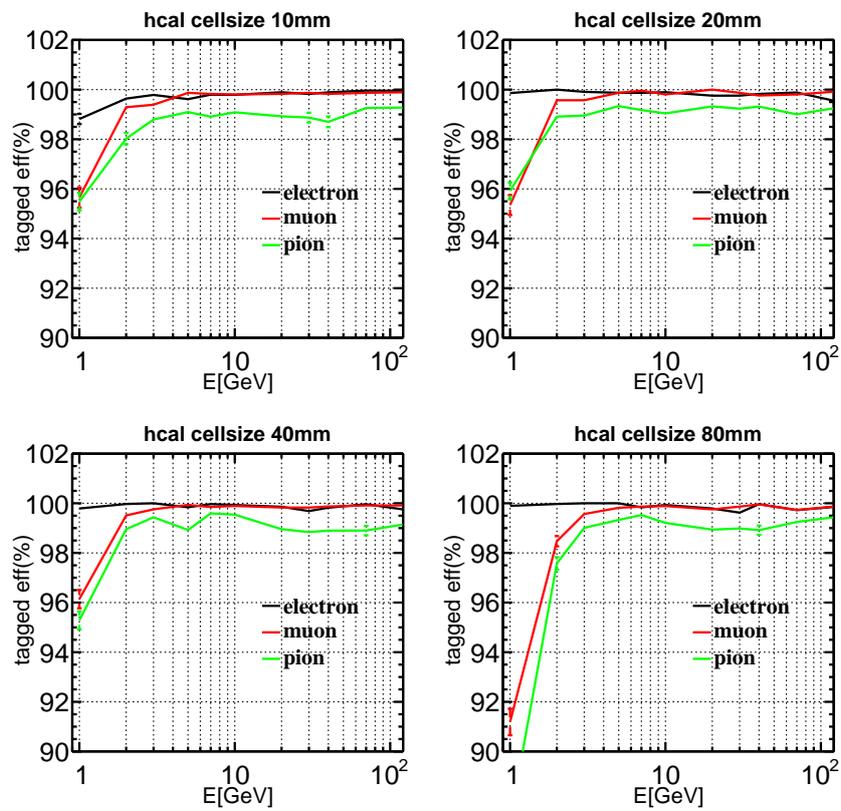


Figure 6.12: The efficiency of lepton identification for different HCAL cell size

1245 6.5 Performance on physics events

1246 The Higgs boson is mainly generated through the Higgsstrahlung process (ZH) and
 1247 more marginally through vector boson fusion processes at electron-positron Higgs fac-
 1248 tories. A significant part of the Higgs bosons will be generated together with a pair of
 1249 leptons (electrons and muons). These leptons are generated from the Z boson decay of
 1250 the ZH process. For the electrons, they can also be generated together with the Higgs
 1251 boson in the Z boson fusions events, see Figure 6.13. At the CEPC, $3.6 \times 10^4 \mu\mu H$ events
 1252 and $3.9 \times 10^4 eeH$ events are expected at an integrated luminosity of 5 ab^{-1} . In these
 1253 events, the particles are rather isolated.

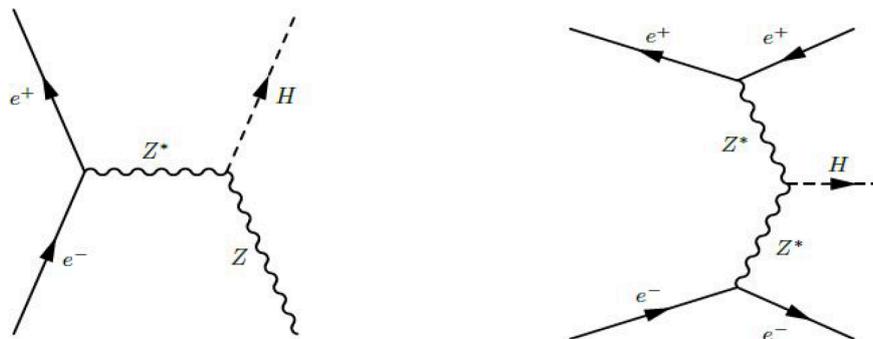


Figure 6.13: Feynman diagrams of major Higgs production with leptons at CEPC: the Higgsstrahlung and ZZ fusion processes.

1254 The eeH and $\mu\mu H$ events provide an excellent access to the model-independent mea-
 1255 surement to the Higgs boson using the recoil mass method [53]. The recoil mass spec-
 1256 trum of eeH and $\mu\mu H$ events is shown in Figure 6.14, which exhibits a high energy tail
 1257 induced by the radiation effects (ISR, FSR, bremsstrahlung), while in CEPC the beam-
 1258 strahlung effect is negligible. The bremsstrahlung effects for the muons are significantly
 1259 smaller than that for the electrons, therefore, it has a higher maximum and a smaller tail.

1260 Figure 6.15 shows the energy spectrum for all the reconstructed charged particles in 10k
 1261 $eeH/\mu\mu H$ events. The leptons could be classified into 2 classes, the initial leptons (those
 1262 generated together with the Higgs boson) and those generated from the Higgs boson
 1263 decay cascade. For the eeH events, the energy spectrum of the initial electron exhibits
 1264 a small peak at low energy, corresponding to the Z fusion events. The precise identi-
 1265 fication of these initial leptons is the key physics objective for the lepton identification
 1266 performance of the detector.

1267 Since the lepton identification performance depends on the particle energy, and most of
 1268 the initial leptons have an energy higher than 20 GeV, we focused on the performance

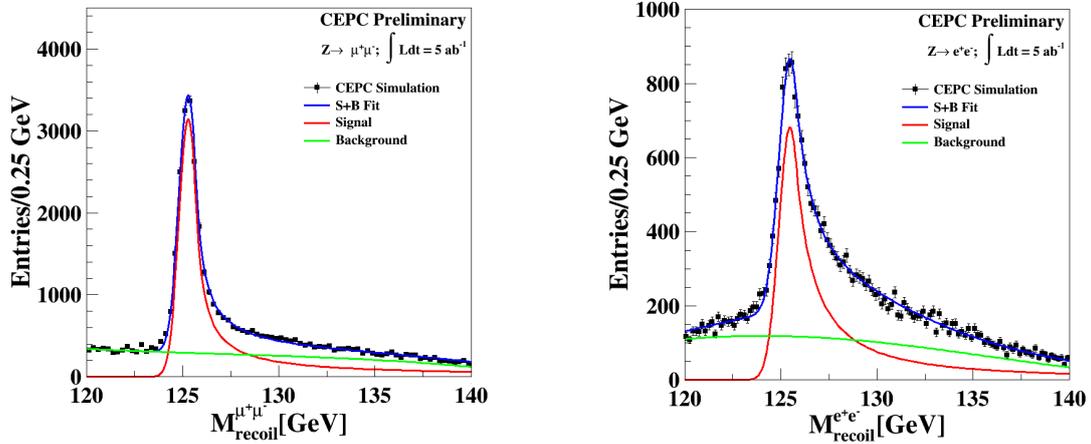


Figure 6.14: The recoil mass spectrum of ee/μμ

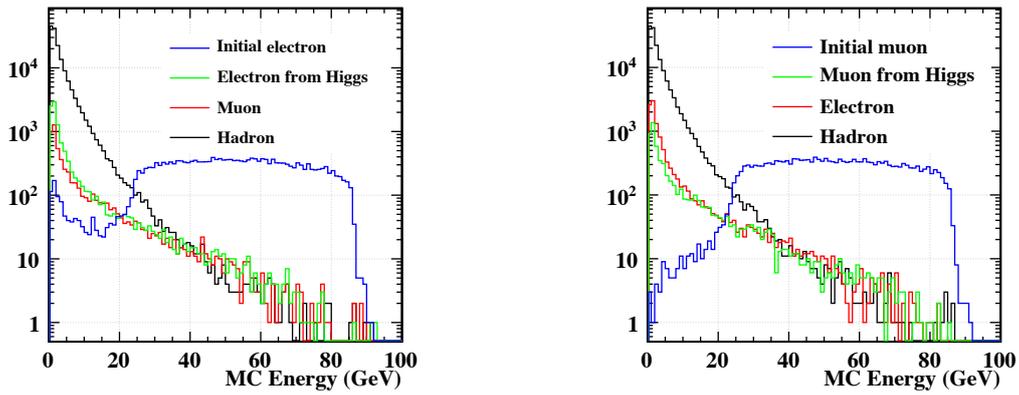


Figure 6.15: Energy Spectrum of charged particles in eeH and μμH event at 250 GeV center of mass energy, low energy peak in eeH corresponds to the Z fusion events

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

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

1269 study of lepton identification on these high energy particles at detectors with two dif-
1270 ferent sets of calorimeter cell sizes.

1271 The μ -likeliness and e-likeliness of electrons, muons, and pions, for eeH events and
1272 $\mu\mu\text{H}$ events are shown in Figure 6.16 and Figure 6.17. Table 6.4 summarizes the defi-
1273 nition of leptons and the corresponding performance under different conditions. The
1274 identification efficiencies for the initial leptons are degraded by 1-2% with respect to
1275 the single particle case. This degradation is mainly caused by the shower overlap, and
1276 is much more significant for electrons as electron showers are much wider than that of
1277 muon, leading to a larger chance of overlapping. The electrons in $\mu\mu\text{H}$ events and vice
1278 versa are generated in the Higgs decay. Their identification efficiency and purity still
1279 remain at a reasonable level. For charged leptons with energy lower than 20 GeV, the
1280 performance degrades by about 10% because of the high statistics of background and
1281 the cluster overlap, as shown in Table 6.5. The event identification efficiency, which is
1282 defined as the chance of successfully identifying both initial leptons, is presented in the
1283 last row of Table 6.4. The event identification efficiencies are roughly the square of the
1284 identification efficiency of the initial leptons. Comparing the performance of both ge-
1285 ometries, it is shown that when the number of readout channels is reduced by 3/4, the
1286 event reconstruction efficiency is degraded by 1.3% and 1.7%, for $\mu\mu\text{H}$ and eeH events
1287 respectively.

1288 6.6 Conclusion

1289 The high granularity calorimeter is a promising technology for detectors in collider fa-
1290 cilities of the High Energy Frontiers. It provides good separation between different final
1291 state particles, which is essential for the PFA reconstructions. It also records the shower

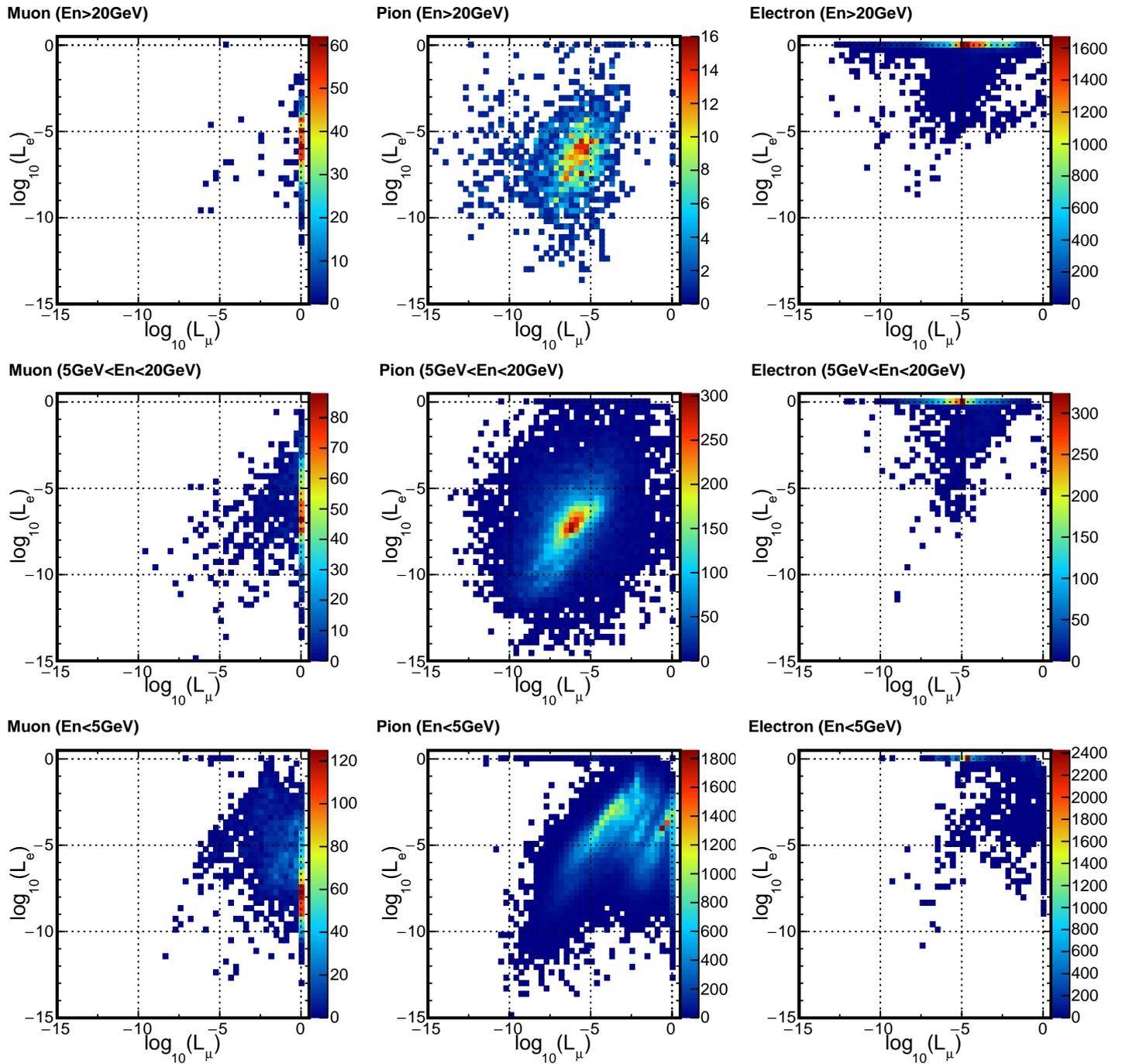


Figure 6.16: e-likelihood and μ -likelihood of charged particles with different energy bins in eeH event

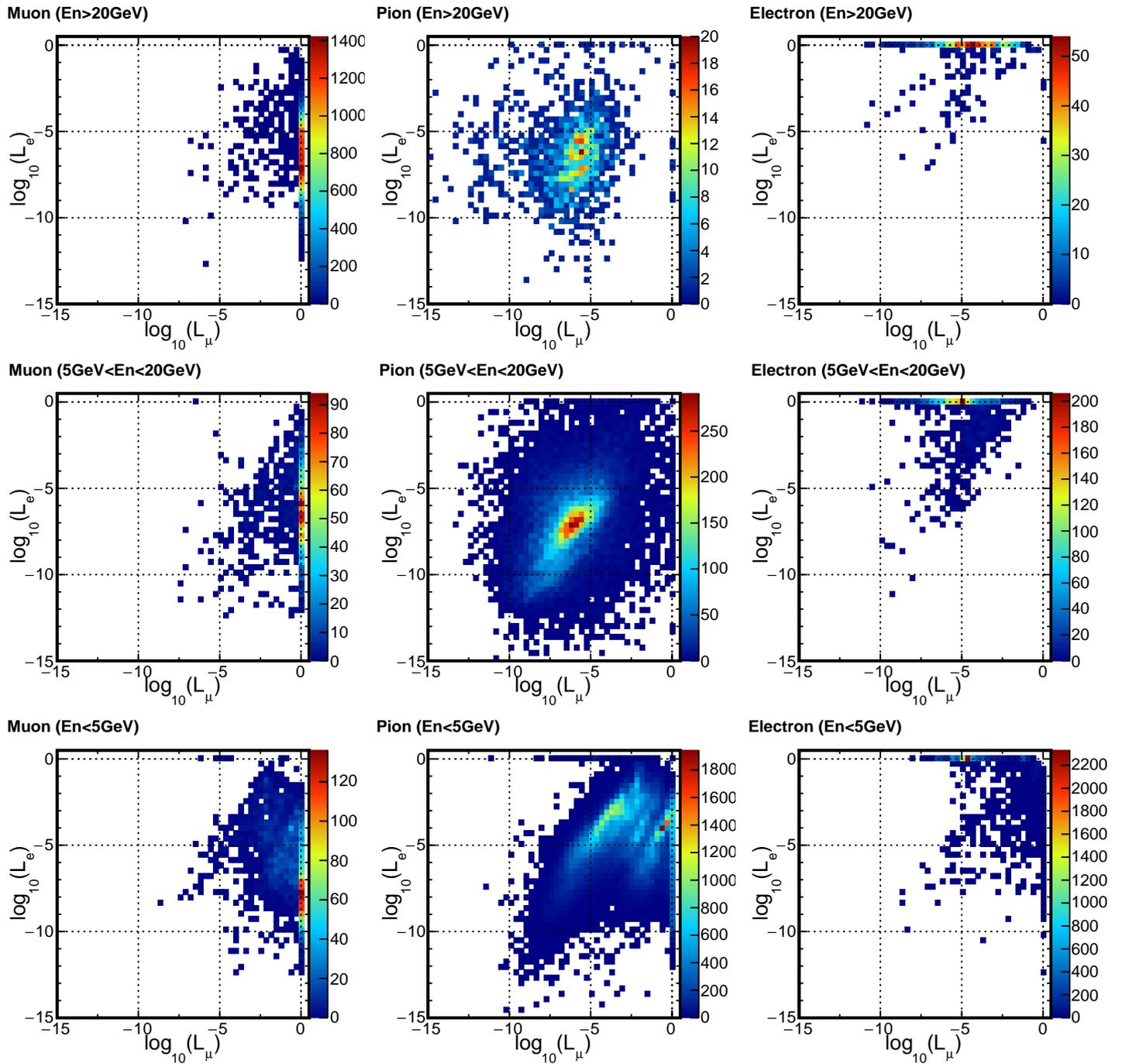


Figure 6.17: e-likelihood and μ -likelihood of charged particles with different energy bins in $\mu\mu H$ event

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

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

1292 spatial development and energy profile to an unprecedented level of details, which can
 1293 be used for the energy measurement and particle identifications.

1294 To exploit the capability of lepton identification with high granularity calorimeters and
 1295 also to provide a viable toolkit for the future Higgs factories, LICH, a TMVA based
 1296 lepton identification package dedicated to high granular calorimeter, has been devel-
 1297 oped. Using mostly the shower description variables extracted from the high granu-
 1298 larity calorimeter and also the dE/dx information measured from tracker, LICH calcu-
 1299 lates the e-likeness and μ -likeness for each individually reconstructed charged particle.
 1300 Based on these output likelihoods, the leptons can be identified according to different
 1301 physics requirement.

1302 Applied to single particle samples simulated with the CEPC_v1 detector geometry, the
 1303 typical identification efficiency for electron and muon is higher than 99.5% for ener-
 1304 gies higher than 2 GeV. For pions, the efficiency is reaching 98%. These efficiencies are
 1305 comparable to the performance reached by ALEPH, while the mis-identification rates
 1306 are significantly improved. Ultimately, the performances are limited by the irreducible
 1307 confusions, in the sense that the chance for muon to be mis-identified as electron and
 1308 vice versa is negligible, the mis-identification of pion to muon is dominated by the pion
 1309 decay.

1310 The tested geometry uses an ultra-high granularity calorimeter: the cell size is 1 by 1
 1311 cm^2 and the layer number of ECAL/HCAL is 30/48. In order to reduce the total channel
 1312 number, LICH is applied to a much more modest granularity, it is found that the lepton
 1313 identification performance degrades only at particle energies lower than 2 GeV for an
 1314 HCAL cell size bigger than $60 \times 60 \text{mm}^2$ or with an HCAL layer number less than 20.

1315 The lepton identification performance of LICH is also tested on the most important
 1316 physics events at CEPC. In these events, multiple final state particles could be produced
 1317 in a single collision, the particle identification performance will potentially be degraded
 1318 by the overlap between nearby particles. The lepton identification on eeH/ $\mu\mu\text{H}$ event

1319 at 250 GeV collision energy has been checked. The efficiency for a single lepton identifi-
1320 cation is consistent with the single particle results. The efficiency of finding two leptons
1321 decreases by 1~2 % when the cell size doubles, which means that the detector needs
1322 2~4% more statistics in the running. In eeH events, the performance degrades because
1323 the clustering algorithm still needs to be optimized.

1324 To conclude, ultra-high granularity calorimeter designed for ILC provides excellent lep-
1325 ton identification ability, for operation close to ZH threshold. It may be a slight overkill
1326 for CEPC and a slightly reduced granularity can reach a better compromise. And LICH,
1327 the dedicated lepton identification for future e+e- Higgs factory, is prepared.

Chapter 7

Measurement of $H \rightarrow \tau\tau$ 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 $H \rightarrow \tau\tau$ 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.

1350 **Leptonic decay** The leptonic decays of the τ lepton probe the structure of the weak
 1351 currents and the universality of their couplings to gauge boson. One of the basic ideas
 1352 in the SM is that all lepton doublets have identical couplings to the Z and W bosons.
 1353 Comparing the measured decay widths of leptonic or semi-leptonic decays which only
 1354 differ in the τ decay, one can test experimentally that the interaction is indeed the same,
 1355 i.e., that $g_e = g_\mu = g_\tau \equiv g$ [57, 58].

1356 **New physics** The τ is also an important probe to the new physics, by observing the
 1357 coupling constants deviation from the Standard Model prediction or exploring lepton
 1358 flavor violating τ decay. A few samples are heavy scalar resonances decaying to a τ
 1359 lepton pair and charged Higgs bosons decays predicted in the MSSM[59]. In the HH
 1360 searches, the $H \rightarrow \tau\tau$ decay channel is one of the most sensitive to both SM and many
 1361 BSM production modes[60]. Besides, differences in the τ^+ and τ^- lifetimes would indi-
 1362 cate the violation of CPT[61].

1363 **B physics** The τ lepton could also be used also a probe of some particular process where
 1364 heavy meson decays into final states containing τ leptons[62]. Decays such as $B^- \rightarrow$
 1365 $\tau^- \bar{\nu}_\tau$, $B \rightarrow D^* \tau^- \bar{\nu}_\tau$, $B_c^- \rightarrow \tau^- \bar{\nu}_\tau$ or $D_s^- \rightarrow \tau^- \bar{\nu}_\tau$ involve the heaviest elementary fermions
 1366 that can be directly produced at flavor factories, providing important information about
 1367 the underlying dynamics mediating these processes.

1368 **Higgs measurement**[63, 64] The τ is the heaviest SM lepton, which leads to a large
 1369 coupling to Higgs, i.e. a significant fraction of the SM Higgs boson decays into $\tau\tau$ final
 1370 states. This makes it possible to measure $g(H \rightarrow \tau\tau)$ with a better accuracy.

1371 As one of the most important channel in the future e^+e^- Higgs factory, $H \rightarrow \tau\tau$ channel
 1372 performance also provides evidence for detector optimization and the PFA develop-
 1373 ments. The requirement to separate photons and hadrons decayed from τ should be
 1374 satisfied by a relatively high granularity and an efficient PFA. On the other hand, to
 1375 distinguish different τ decay modes, the PFA should provide reasonable particle iden-
 1376 tification.

1377 7.1.2 τ decay modes

1378 The leptonic decay of τ lepton follows $\tau^- \rightarrow \nu_\tau l^- \bar{\nu}_l$, with $l = e, \mu$. These two neutrinos
 1379 make it difficult to reconstruct the τ mass. In the hadronic decays, only one neutrino is
 1380 involved, its direction can thus be reconstructed by measuring all other decay products.
 1381 This is not used in this thesis, but can be a continuation to the studies. The hadronic
 1382 decay of τ lepton can be classified in:

- 1383 • final state without photon: $\tau^- \rightarrow \nu_\tau h^-$, with $h = \pi, K$

- 1384 • final state with two photons dominated by ρ production: $\tau^- \rightarrow \nu_\tau \rho^- \rightarrow \nu_\tau \pi^- \pi^0$
 1385 and $\pi^0 \rightarrow \gamma\gamma$
- 1386 • final state with four photons dominated by a_1^- production: $\tau^- \rightarrow \nu_\tau a_1^- \rightarrow \nu_\tau \pi^- 2\pi^0$
 1387 and $\pi^0 \rightarrow \gamma\gamma$.

1388 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

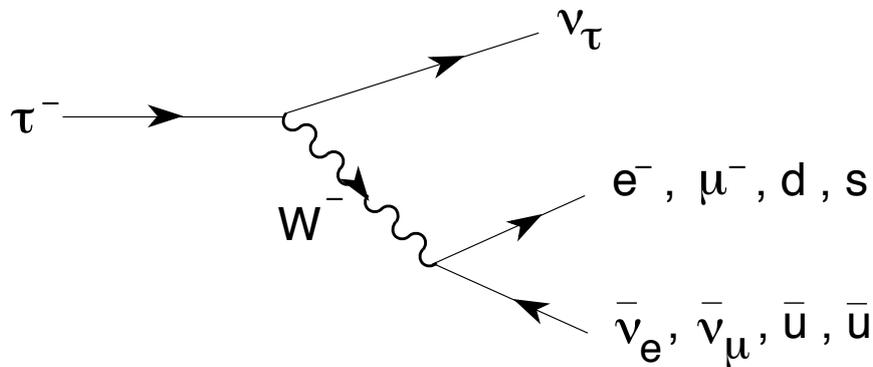


Figure 7.1: Feynman diagram for $\tau \rightarrow \nu_\tau X$ decay modes

1389 The topology of τ in the high granular detector is shown in the event display in Figure
 1390 7.2.

1391 As shown in the event display, the τ decay in high energy colliders is tightly collimated
 1392 and low multiplicity, which provide excellent signatures to probe.

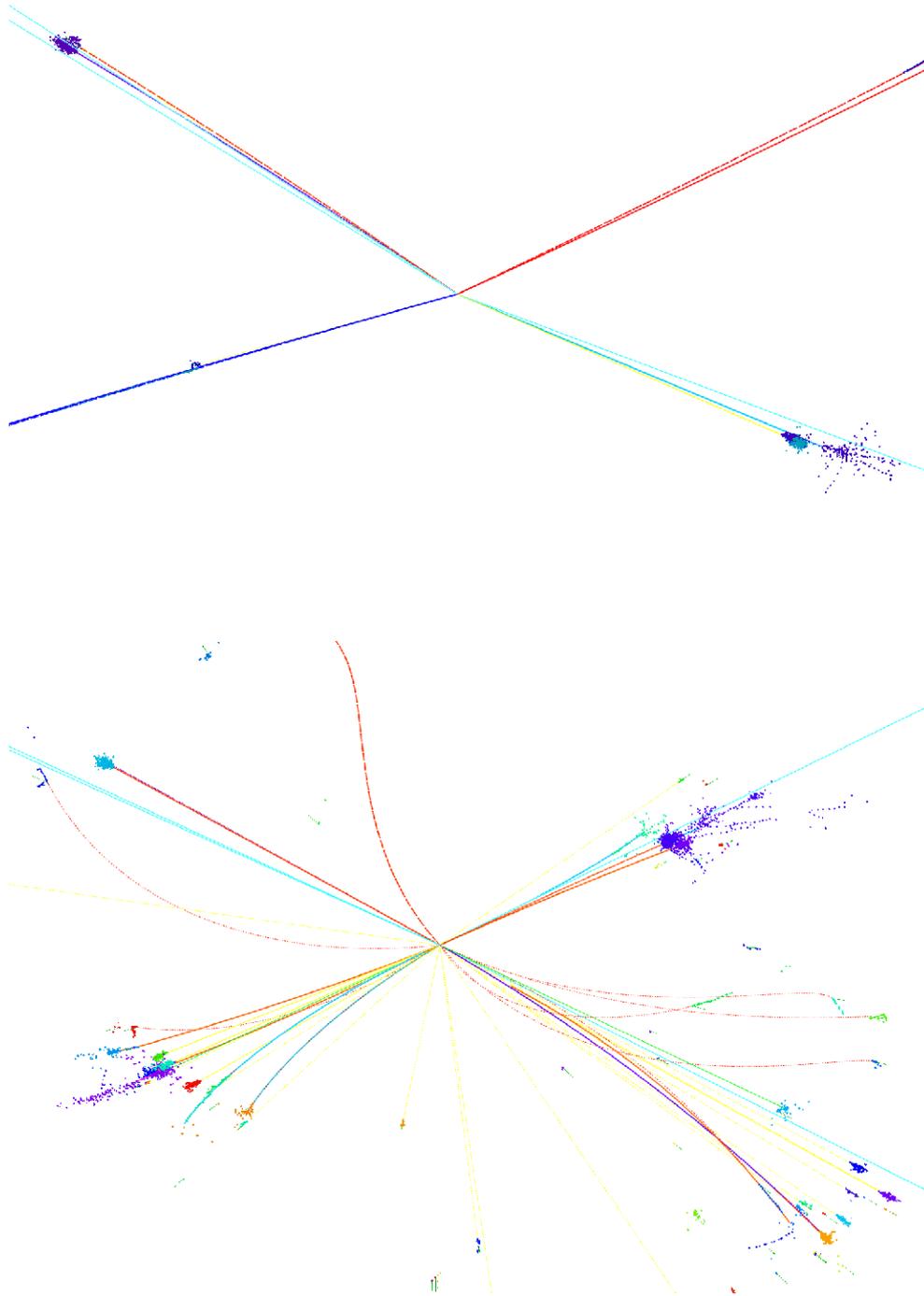


Figure 7.2: Event display of a $\mu\mu\tau\tau$ event with one $\tau \rightarrow e^- \bar{\nu}_e \nu_\tau$ and the other $\tau \rightarrow \pi^- \nu_\tau$ (up) and a $qq\tau\tau$ event with one $\tau \rightarrow e^- \bar{\nu}_e \nu_\tau$ and the other $\tau \rightarrow \pi^- \pi^+ \pi^- \nu_\tau$ (down)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

1393 7.1.3 Measurements and precisions

1394 The deviation of coupling constants from the Standard Model prediction to new physics
 1395 beyond the Standard Model depends on the new physics model, and this deviation is
 1396 estimated to be 1% level by many models proposed. At the LHC, the process of the
 1397 Higgs boson decaying into tau pairs will be measured using proton–proton collision.
 1398 This decay has been studied by the ATLAS and the CMS experiments, who reported a
 1399 combined signal yield consistent with the Standard Model expectation, with a combined
 1400 observed significance at the level of 6σ . With an uncertainty of 9% at HL-LHC (300 fb^{-1}),
 1401 the LHC experiment may not have sufficient sensitivity for new physics described in the
 1402 previous section.

1403 On the other hand, previous studies of the Higgs boson decaying into tau pairs at the
 1404 ILC show that the measurement can be of the order of a few percent[65] and that the
 1405 measurement at the ILC plays a crucial role after the LHC experiments. However, these
 1406 studies did not take into account some of the relevant background processes (such as
 1407 $\nu\nu H$), nor based on the jet clustering algorithm. Therefore in this thesis, this channel is
 1408 studied independently from the jet clustering while taking into account the whole SM
 1409 background.

1410 7.2 Samples

1411 The CEPC luminosity is supposed to be 5000 fb^{-1} . For the ZH signal, the cross section
 1412 for different Z decay modes is summarized in table Chapter 2, as well as the branching
 1413 ratio of Higgs decaying to $\tau\tau$. All the samples in this chapter are generated by the MC
 1414 generator Whizard, version 1.95[66]. The detector used in the simulation is the CEPC
 1415 detector.

1416 The cross section shown here gives the first view to the efficiency and purity that need
 1417 to be achieved. Taking qqH channel, for example, the statistics for signal $qq\tau\tau$ and
 1418 backgrounds are 44872 and 488 million respectively. Using the simple expression of
 1419 accuracy as $\sqrt{S+B}/S$, if the efficiency to identify $qq\tau\tau$ event is 80%, the background
 1420 should be suppressed by 99.98% in order to achieve the 1% accuracy.

1421 The studies on Higgs decaying into the τ channel are treated individually for each Z de-
 1422 caying channel, in order to distinguish the signal with the different type of backgrounds.
 1423 The selection of events is done in two steps:

- 1424 • **Pre-selection** Due to the limited computing resource, the inclusive ZH events, and
 1425 SM categories background events are filtered by some preselection using MC truth

1426 information to simplify the samples. The excellent performance of PFA ensures
 1427 that this preselection would not lose information. The information used in the
 1428 preselection is different for each Z decaying channel, including the number of
 1429 muons ($N_{\mu^{+/-}}$), the recoil mass of the muon pair (M_{recoil}), the invariant mass of the
 1430 muon pair ($M_{invariant}$), the missing mass ($M_{missing}$), the total visible mass (M_{tot}),
 1431 the transverse momentum (p_T), the visible energy (E_{vis}), the number of charged
 1432 particles (N_{charge}).

- 1433 • **τ tagging** The τ tagging process is applied using the topology of events. The
 1434 impact parameters are used in order to deduce the statistics of signal and back-
 1435 grounds.

1436 A successful reconstruction of the τ lepton is not a trivial task, for the τ lepton could be
 1437 generated with various different event topology, and it has diverse decay final states.
 1438 In the e^+e^- collision environment, we summarize the τ events into two categories ac-
 1439 cording to the event topology, in which the reconstruction algorithm and performances
 1440 have been studied separately.

1441 7.3 Leptonic channels

1442 The first category is the leptonic one, where no physics objects, or only lepton / photon
 1443 / missing energy is generated together with the τ candidates.¹ These events include,
 1444 for example:

- 1445 • $ZH, Z \rightarrow l^+l^- / \nu\nu, H \rightarrow \tau\tau$ events; golden channel for $g(H\tau\tau)$ measurements
- 1446 • $ZZ, l^+l^- / \nu\nu / \tau\tau$ events
- 1447 • WW events with $l\nu\tau\nu$ final states.
- 1448 • $Z \rightarrow \tau\tau$ events at Z pole operation.

1449 In these events, the global multiplicity is limited while the additional physics objects, if
 1450 they exists, are easy to identify. A successful identification of these events relies highly
 1451 on the reconstruction of photons and charged hadrons. In the following section, the
 1452 physics performances of τ reconstruction at $\mu\mu H$ and $\nu\nu H$ channel are shown as well
 1453 as their $\text{Br}(H \rightarrow \tau\tau)$ measurement.

¹The charge is ignored for event classifications.

1454 **7.3.1** $Z \rightarrow \mu\mu$

1455 The easiest channel to study is the $\mu\mu H$ channel since the two muons are easy to be
 1456 vetoed by calculating their invariant mass. According to the different behavior of $\mu\mu H$
 1457 and backgrounds shown in Figure 7.3, the preselection applied to select $\mu\mu H$ are:

- 1458 • $N_{\mu^+} > 1, N_{\mu^-} > 1$
- 1459 • $110\text{GeV} < M_{recoil} < 180\text{GeV}$
- 1460 • $40\text{GeV} < M_{invariant} < 180\text{GeV}$

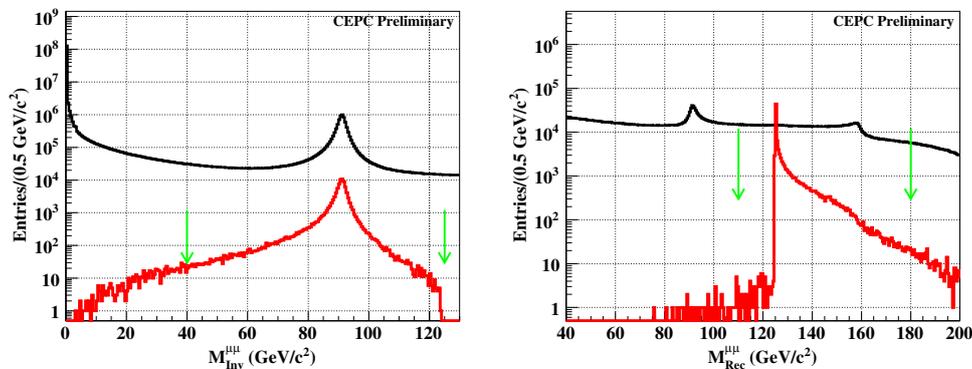


Figure 7.3: Distribution of invariant mass and recoil mass (MC information) for $\mu\mu H$ and back-
 grounds at $\sqrt{s} = 250\text{GeV}$, the red/black line is for signal($\mu\mu H$)/background(inclusive), the
 green arrows indicates the cuts applied in the preselection. The invariant mass of $\mu\mu H$ 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

1461 Thank the excellent efficiency and purity of the muon identification, the efficiency of this
 1462 preselection can achieve 97.68%, while most of the SM backgrounds are vetoed except
 1463 for $\mu\mu$ (3.51% remaining).

1464 Most of the decaying modes of τ are with one or three tracks and an even number of
 1465 photons, as can be seen in Table 7.1, this is the main idea in the τ tagging. From the decay
 1466 modes, the topology of τ s is simpler than jets, which provides the way to distinguish τ
 1467 events from the others. The steps for di- τ events tagging are:

- 1468 • Veto the μs decayed from Z by choosing the μ pair with invariant mass closest to
 1469 Z mass
- 1470 • Find the leading track among the remaining particles and collect the tracks and
 1471 photons close to this track (< 1 rad, to be grouped in region A), and their numbers
 1472 are noted as NTrkA and NPhA.

- 1473 • Collect the rest tracks and photons and group them in region B with their numbers
1474 noted as NTrkB and NPhB.
- 1475 • Get the angle between the leading tracks in region A or B and the furthest track in
1476 this region, noted as $\text{Cone}_{T-T}(A/B)$.
- 1477 • $\text{Cone}_{T-P}(A/B)$ is the angle between the leading tracks in region A or B and the
1478 furthest photon in this region.
- 1479 • $\text{Cone}_{P-P}(A/B)$, the angle between the leading photon in a region and the furthest
1480 photon in this region.

1481 The distributions of these numbers in τ events and other decay channels of Higgs is
1482 shown in Figure 7.4 and the cuts of NTrk and NPh are chosen to be less than 6 and less
1483 than 7.

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

	$\mu\mu H_{\tau\tau}$	$\mu\mu H$ inclusive bkg	ZZ	WW	singleW	singleZ	$2f$
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%

1484 After the cut of the number of tracks and photons, these parameters are trained in
1485 TMVA and optimized to the signal significance giving the BDT cut to 0.78, the cut flow
1486 is summarized in Table 7.2, the efficiency of the signal after training is 93%. The corre-
1487 lation matrix and overtraining check are shown in Figure 7.5 and Figure 7.6.

1488 However, the channels such as Higgs decaying into W and W leptonic decay are the
1489 main backgrounds after the selections. This is due to the topologies of these events are
1490 similar to our signal.

1491 By looking at the starting points for the tracks, those stemming from τ decays are further
1492 away from the vertex than the others. From the sum of transverse and longitudinal
1493 impact parameters ($D0 / Z0^2$) of the two leading tracks in regions A and B normalized
1494 by their uncertainty $\sigma_{D0/Z0}$, a "pull" can be defined as: $D0^2/\sigma_{D0}^2 + Z0^2/\sigma_{Z0}^2$, since $D0$
1495 and $Z0$ are comparable in CEPC detector, the pull are simplified as $D0^2 + Z0^2$. The pull

²The 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.

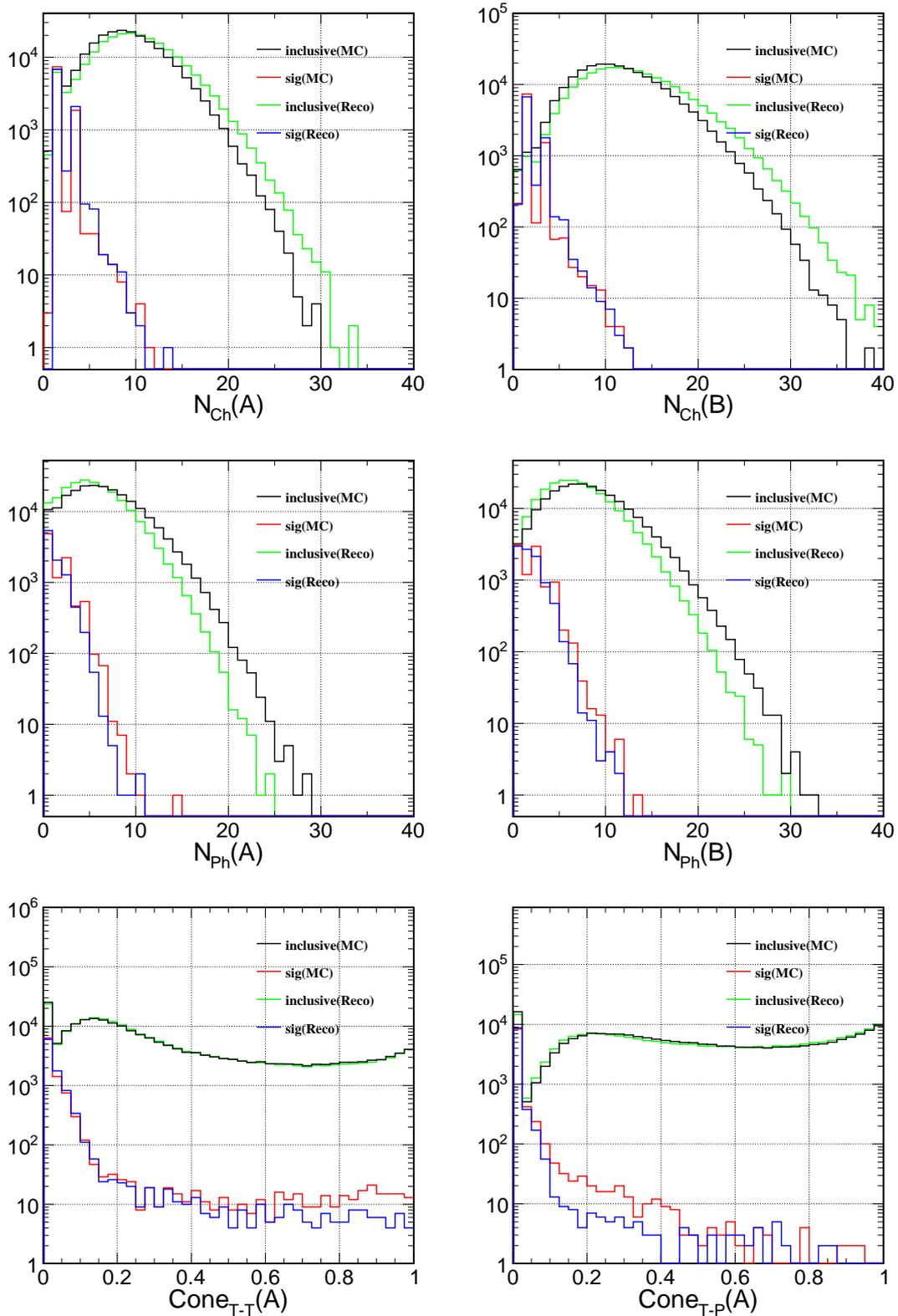


Figure 7.4: Distribution of number of tracks and photons, the angle between track to track, track to photon, or photon to photon in the two opposite regions A and B. The black/red line represents the MC information of the inclusive $\mu\mu H$ backgrounds / signal ($\mu\mu H \rightarrow \tau\tau$), the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

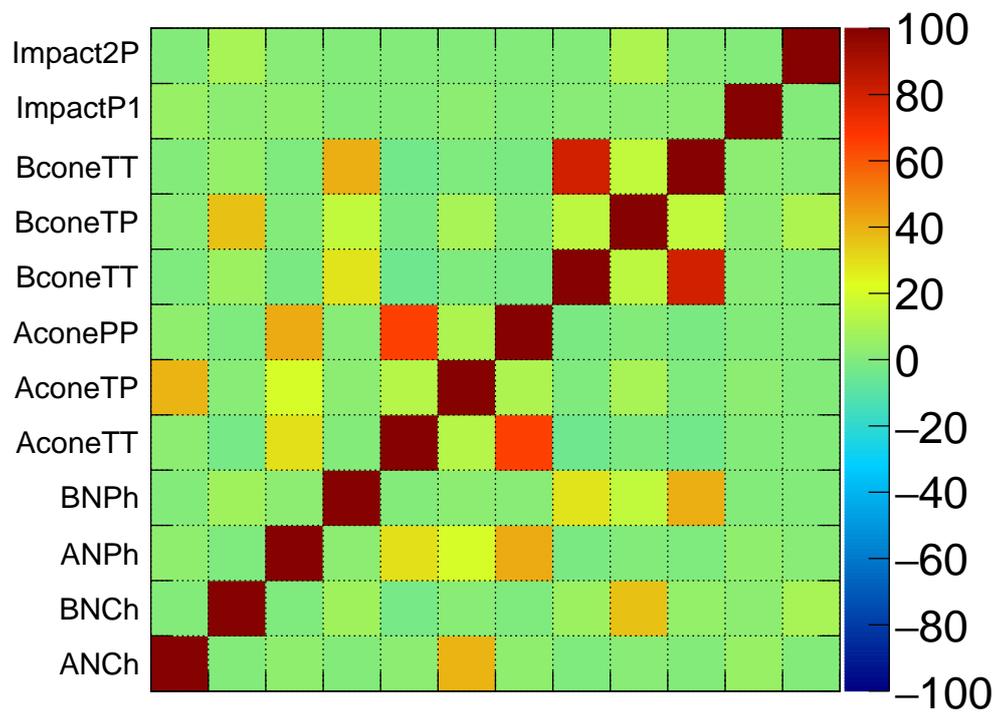


Figure 7.5: The correlation matrix of all the variables

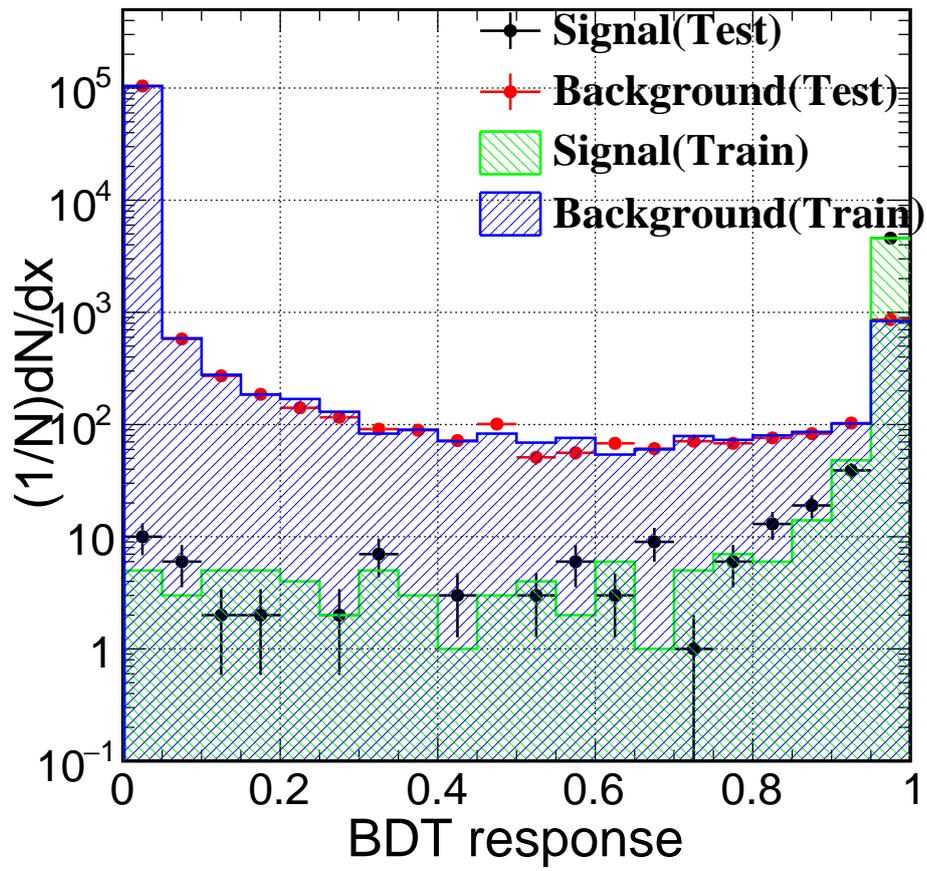


Figure 7.6: BDT response of di- τ finding

1496 distribution is shown in Figure 7.7 for signal and SM inclusive background with a fit.

1497 The branching ratio $Br(H \rightarrow \tau\tau)$ can be calculated from the fitted signal event number
 1498 S , the total event number T and previous selection efficiency ε , as $Br = S/(\varepsilon \cdot T)$, to be
 1499 6.40 ± 0.18 . The expected accuracy $\sigma \times BR = \delta(S)/S$ to be 2.68%, where the $\delta(S)$ is the
 1500 fitted signal event number error.

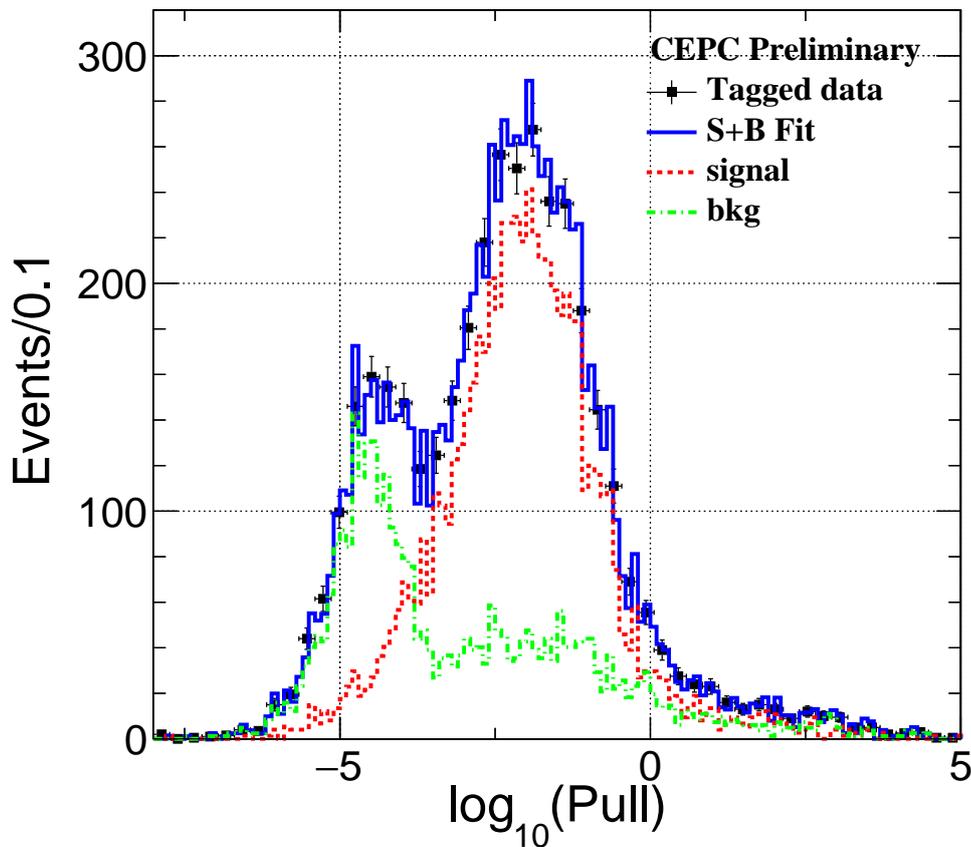


Figure 7.7: Fit of the sum of $D0^2$ and $Z0^2$ of the leading tracks of two cones with SM background included

1501 Assuming that the efficiency of $\tau\tau$ event tagging is the same for $\mu\mu H$ and eeH events,
 1502 the accuracy for the eeH event can be extrapolated. The difference between this two
 1503 channel is that the efficiency for preselection is not the same, as shown in Table 7.3. The
 1504 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

1505 **7.3.2** $Z \rightarrow \nu\nu$

1506 According to the different behavior of $\nu\nu H$ and backgrounds, the cut flow of the pre-
1507 lection for $\nu\nu H$ events is:

- 1508 • $65\text{GeV} < M_{\text{missing}} < 225\text{GeV}$
- 1509 • $M_{\text{total}} > 50\text{GeV}$
- 1510 • $10\text{GeV} < p_T < 100\text{GeV}$

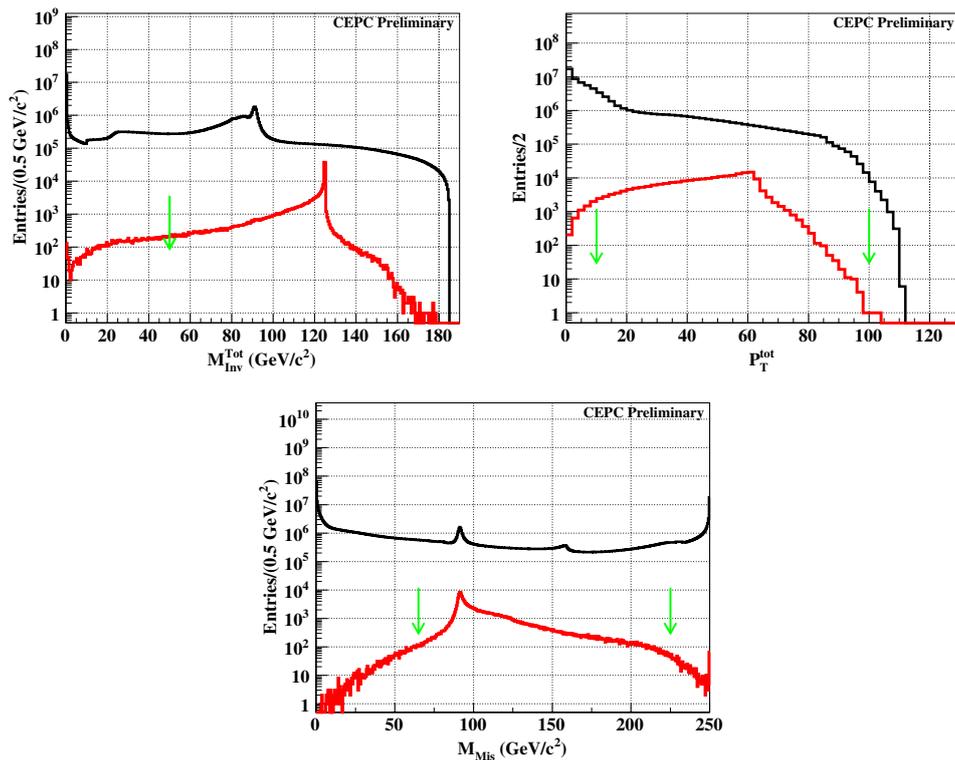


Figure 7.8: Distribution of total invariant mass $M_{\text{Inv}}^{\text{tot}}$, transverse momentum P_T^{tot} , and missing mass M_{Mis} for $\nu\nu H$ and backgrounds at $\sqrt{s} = 250\text{GeV}$, the red/black line is for signal($\nu\nu H$)/background(inclusive), the green arrows indicates the cuts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

1511 However, a bias exists on the different signal channel in this cut flow, which leads to a

1512 1.7% degradation of $BR(H \rightarrow \tau\tau)$ and the final result needs to be corrected according
 1513 to this number.

1514 The procedure of τ tagging in $Z \rightarrow \nu\nu$ event is similar to the one in $Z \rightarrow \mu\mu$, but without
 1515 the step to veto the μ pair. However, there exists a huge irreducible background coming
 1516 from WW and $W \rightarrow \nu\tau$, whose impact parameters are not distinguishable, as shown in
 1517 Figure 7.9. Therefore the only statistic result is deduced in this channel by ignoring the
 1518 error of the fraction of signal and background.

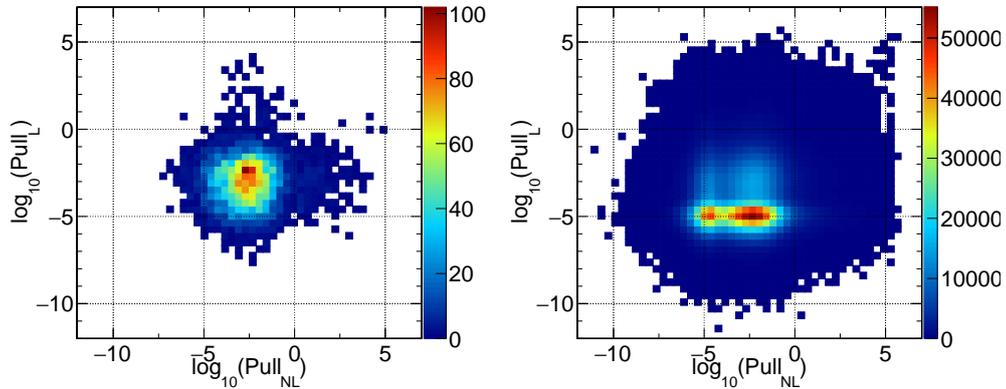


Figure 7.9: Pull of the leading track and the next to leading track for $\nu\nu H$ and backgrounds.

1519 The efficiency of tagging after TMVA training is 95%, and the τ event number is calcu-
 1520 lated from the statistics, as shown in Table 7.4. The branching ratio $Br(H \rightarrow \tau\tau)$ can be
 1521 calculated from the statistics result and previous selection efficiency to be 6.19 ± 0.27 , as
 1522 well as the expected accuracy to be 4.29%.

Table 7.4: Cut Flow of MC sample for $\nu\nu H \rightarrow \tau\tau$ selection on signal and inclusive SM back-grounds

	$\nu\nu H_{\tau\tau}$	$\nu\nu H$ 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%

1523 7.4 Hadronic channel, $Z \rightarrow qq$

1524 The second catalog is the hadronic one, where the τ lepton(s) are always observed with
1525 jets. For instance, we have:

- 1526 • $ZH, Z \rightarrow qq, H \rightarrow \tau\tau$
- 1527 • $ZZ \rightarrow qq\tau\tau$
- 1528 • $WW \rightarrow qql\tau$
- 1529 • $ZH, Z \rightarrow qq, H \rightarrow WW \rightarrow l\nu\tau\nu$

1530 The most difficult channel is Z decaying to quarks since these quarks cannot be vetoed
1531 from the invariant mass without jet clustering.

1532 The preselection applied to choose the qqH events is:

- 1533 • $E_{visible} > 100GeV$
- 1534 • $N_{charge} > 8$
- 1535 • $P_t < 93GeV$
- 1536 • $M_{Mis} < 120GeV$

1537 Since the background is still too large, a second preselection is applied to choose the
1538 $qqH \rightarrow \tau\tau$ events is:

- 1539 • $115GeV < E_{visible} < 245GeV$
- 1540 • $M_{Mis} > 2GeV$

1541 The distribution of these variables for preselection is shown in Figure 7.11.

1542 Since the qqH process is more complex than $\mu\mu H$ and $\nu\nu H$, the preselection is not that
1543 powerful as the previous ones. Keeping the preselection efficiency high leads to nearly
1544 half of ZZ and WW semi-leptonic decay remaining. That's a huge number of events to
1545 study, therefore the backgrounds are not analyzed in the whole sample but on smaller
1546 statistics (10k per sub channel) and scaled to 5 ab^{-1} .

1547 After the preselection, the tagging method is no longer for di- τ but to tag the τ jets in
1548 the whole space in an event. The steps are:

- 1549 • Find tracks with energy higher than a defined E_{min} as the seed

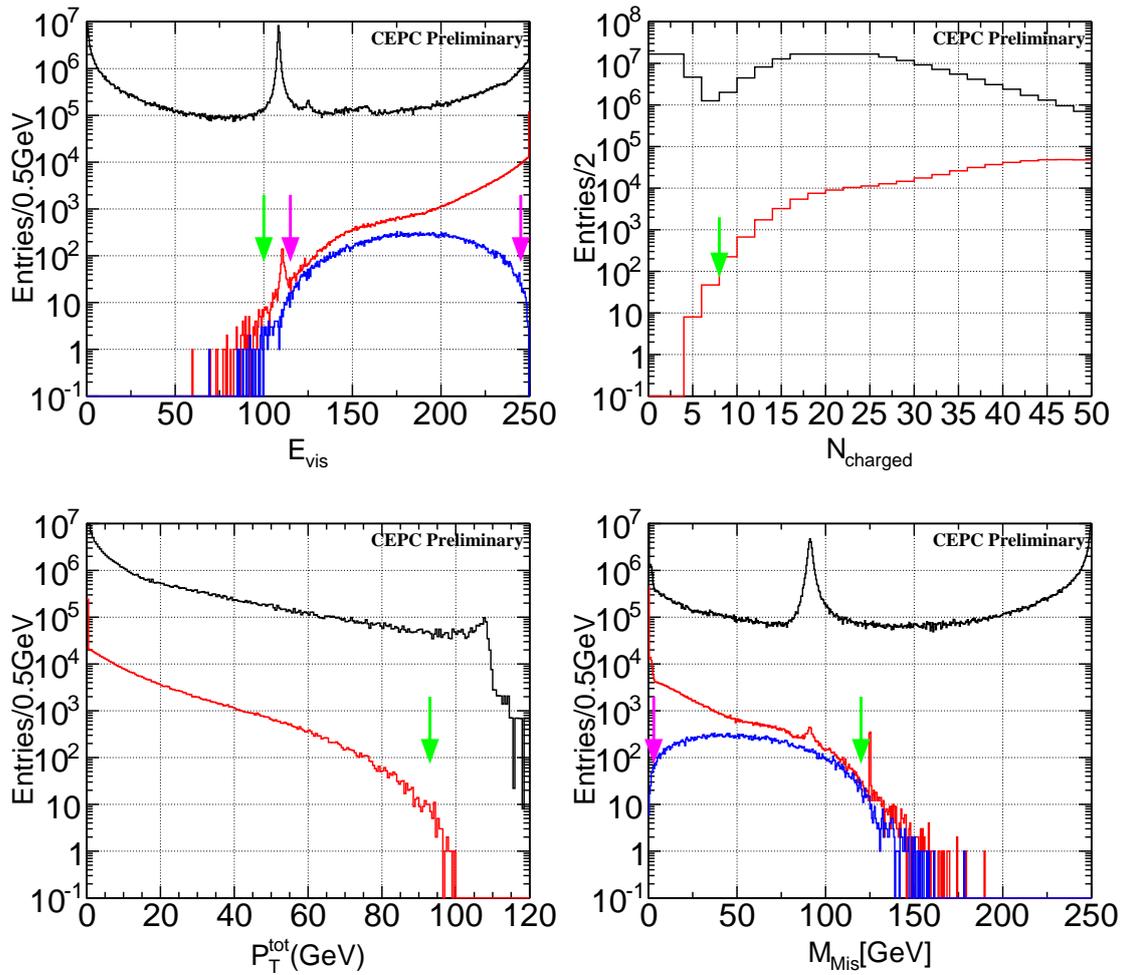


Figure 7.10: Distribution of total visible energy $E_{visible}$, number of charged particles N_{charge} , total visible energy E_{vis} , transverse momentum P_T^{tot} , and missing mass M_{Mis} for $qqH\tau\tau$, qqH and backgrounds at $\sqrt{s} = 250\text{GeV}$, the blue/red/black line is for signal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

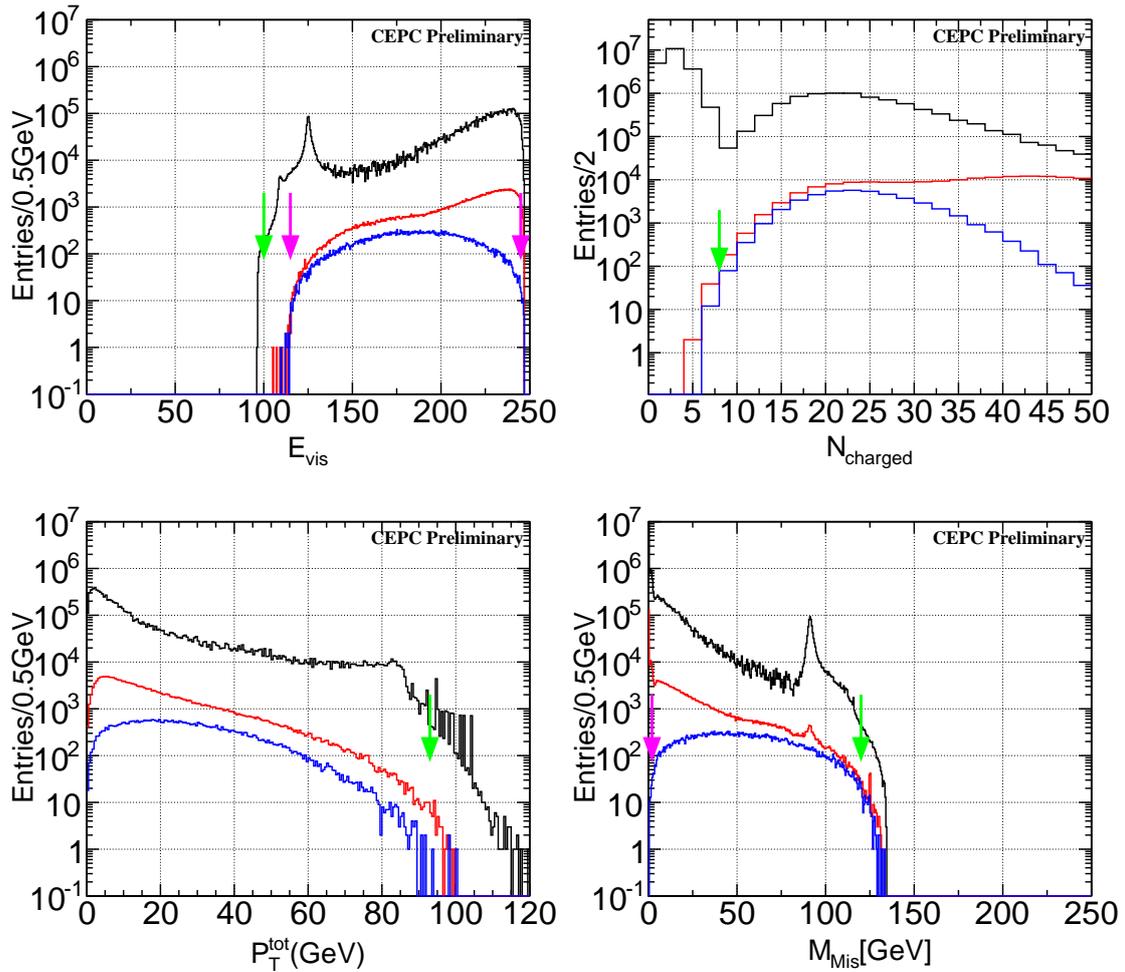


Figure 7.11: Distribution of total visible energy $E_{visible}$, number of charged particles N_{charge} , total visible energy E_{vis} , transverse momentum P_T^{tot} , and missing mass M_{Mis} for $qqH\tau\tau$, qqH and backgrounds at $\sqrt{s} = 250\text{GeV}$, with the other cuts applied.

- 1550 • Collect tracks and photons within an angle ConeA
- 1551 • Calculate invariant mass with these particles
- 1552 • Calculate the D0 and Z0 of the leading track
- 1553 • Calculate the energy in a larger cone ConeB around the seed.

1554 The cut of τ tagging is:

- 1555 • Number of tracks/photons smaller than 6/8
- 1556 • Energy proportion in the smaller cone larger than R_{En}
- 1557 • Invariant mass of the $\tau\tau$ system larger than M_{min} GeV and smaller than M_{max} GeV
- 1558 • Invariant mass of the qq system (the particles except for τ s) smaller than M_{qq} GeV.

1559 Here the parameters E_{min} , ConeA, ConeB, R_{En} , M_{min} and M_{max} are optimized to the
 1560 value $\epsilon \cdot p$, where ϵ is the efficiency of finding an opposite charged τ pair in qq $\tau\tau$ events
 1561 and p is the probability of tagging a opposite charged τ pair in the backgrounds. The
 1562 value of these parameters are: $E_{min} = 1.5$ GeV, ConeA = 0.15 rad, ConeB = 0.45 rad,
 1563 $M_{min} = 0.2$ GeV, $M_{max} = 2.0$ GeV, $R_{En} = 0.92$, the optimized $\epsilon \cdot p$ is 56%. However, this is
 1564 a rough optimization without background normalization taken into account.

1565 After these cuts, the remaining τ s in an event is collected and the two leading energetic
 1566 ones with opposite charge are chosen to calculate the invariant mass of the di- τ , as
 1567 shown in Figure 7.12. The distribution of each type of background in Figure 7.13 shows
 1568 that the 2f background is reduced in this step, as well as the events with "fake" taus
 1569 reconstructed.

1570 The events with at least a pair of τ s and the invariant mass in a range of (20, 120 GeV)
 1571 are chosen as a Higgs decaying to the $\tau\tau$ event. The particles except for these have
 1572 been chosen to form the two leading energetic ones with opposite charge are used to
 1573 get the invariant mass of the qq system and the cut of $70 < M_{qq} < 105$ GeV is chosen as
 1574 the selection of signal, as shown in Figure 7.14. In Figure 7.15, it is shown that the ZH
 1575 background and WW background can be reduced, where the invariant mass of qq leads
 1576 to a Higgs mass or it is a flat distribution. The ZZ background is still an important one
 1577 since the invariant mass of qq is also peaking at Z mass.

1578 The recoil mass of the qq system is used to reduce the ZZ backgrounds, as shown in
 1579 Figure 7.16 and Figure 7.17, the background $ZZ \rightarrow qq\tau\tau$ are reduced because the recoil
 1580 mass of the qq leads to the mass of Z.

1581 The cut chain is summarized in Table 7.5 and the efficiency for the τ events tagging is

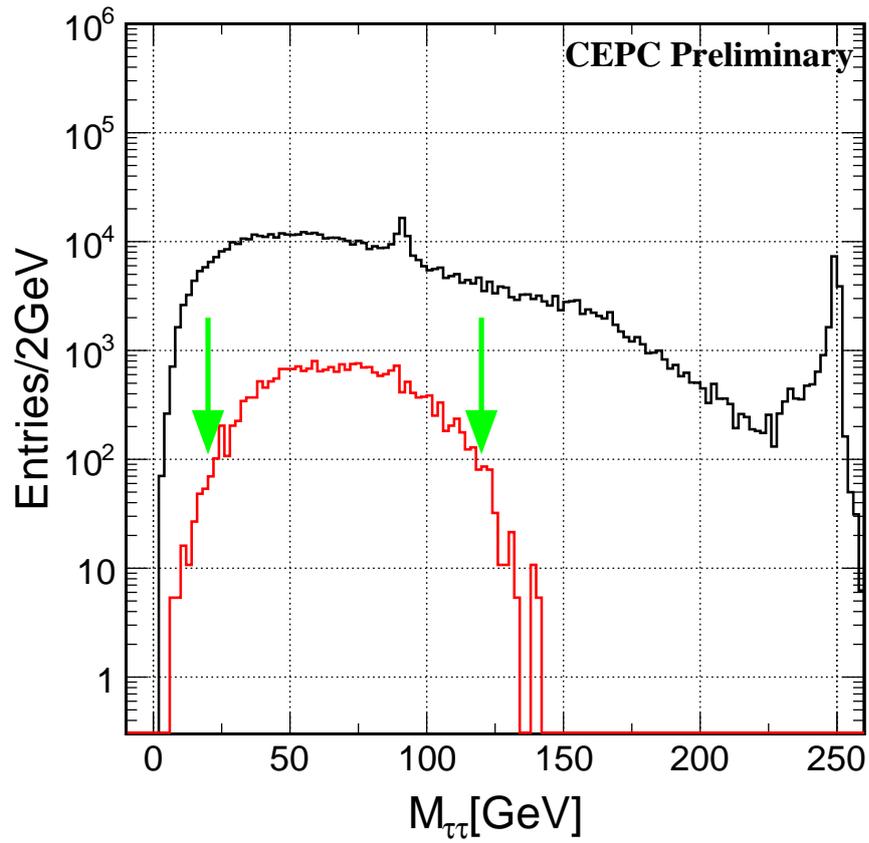


Figure 7.12: Distribution of the invariant mass of the di- τ , $M_{\tau+\tau-}$ for $qqH\tau\tau$, and backgrounds at $\sqrt{s} = 250\text{GeV}$, the red/black line is for signal ($qqH\tau\tau$)/background(inclusive). The arrows indicates the cuts applied.

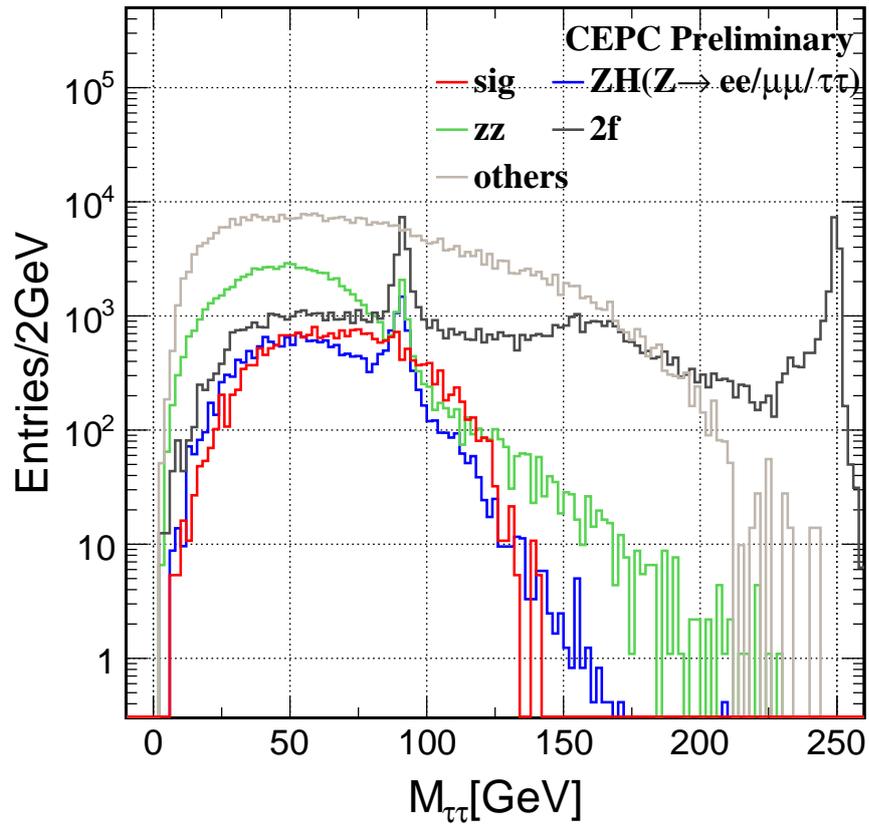


Figure 7.13: Distribution of the invariant mass of the di- τ , $M_{\tau+\tau^-}$ for $qqH\tau\tau$, and each background at $\sqrt{s} = 250$ GeV .

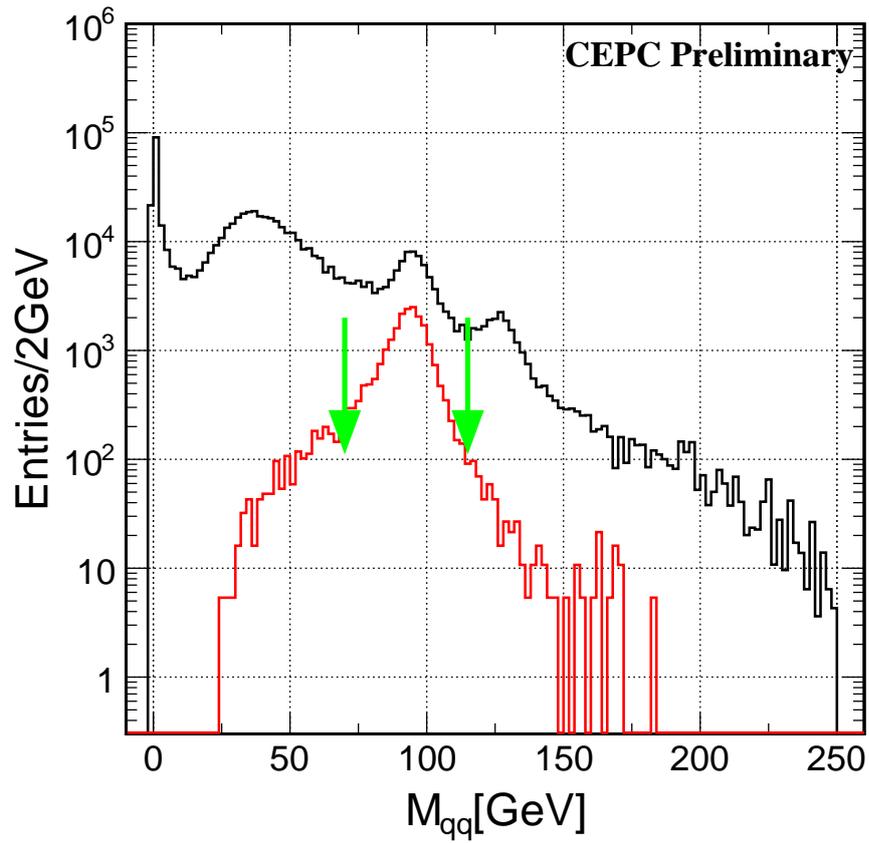


Figure 7.14: Distribution of the invariant mass of the qq , M_{qq} for $qqH\tau\tau$ and backgrounds at $\sqrt{s} = 250\text{GeV}$, the red/black line is for signal ($qqH\tau\tau$)/background(inclusive). The arrows indicates the cuts applied.

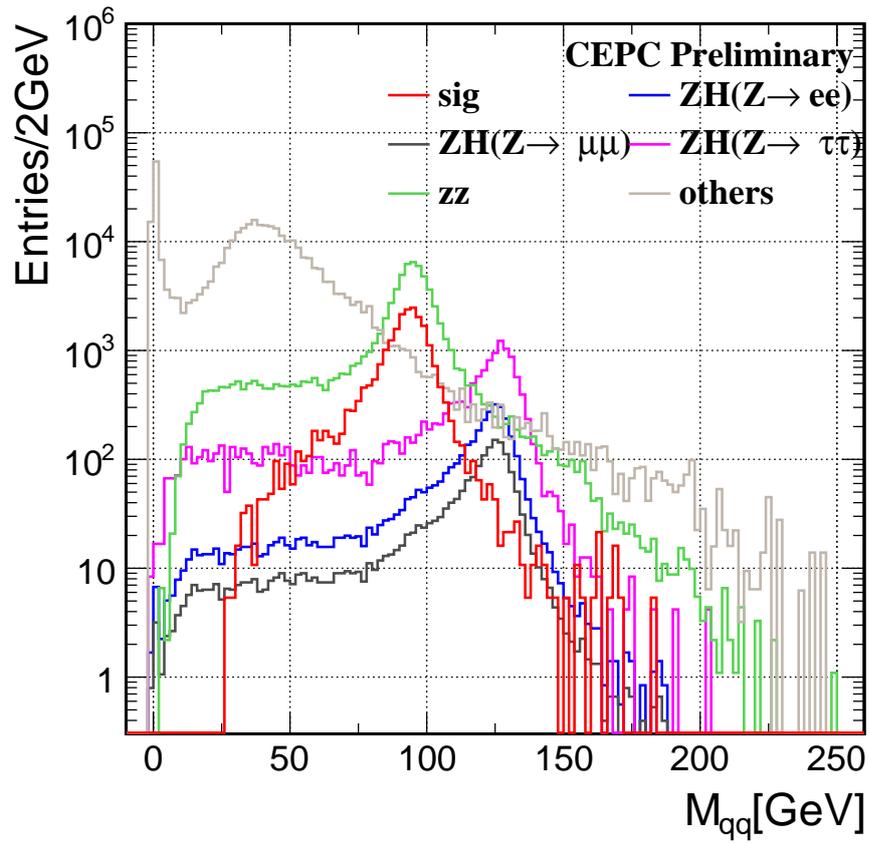


Figure 7.15: Distribution of the invariant mass of the qq , M_{qq} for $qqH\tau\tau$ and each backgrounds at $\sqrt{s} = 250\text{GeV}$ after the previous cuts.

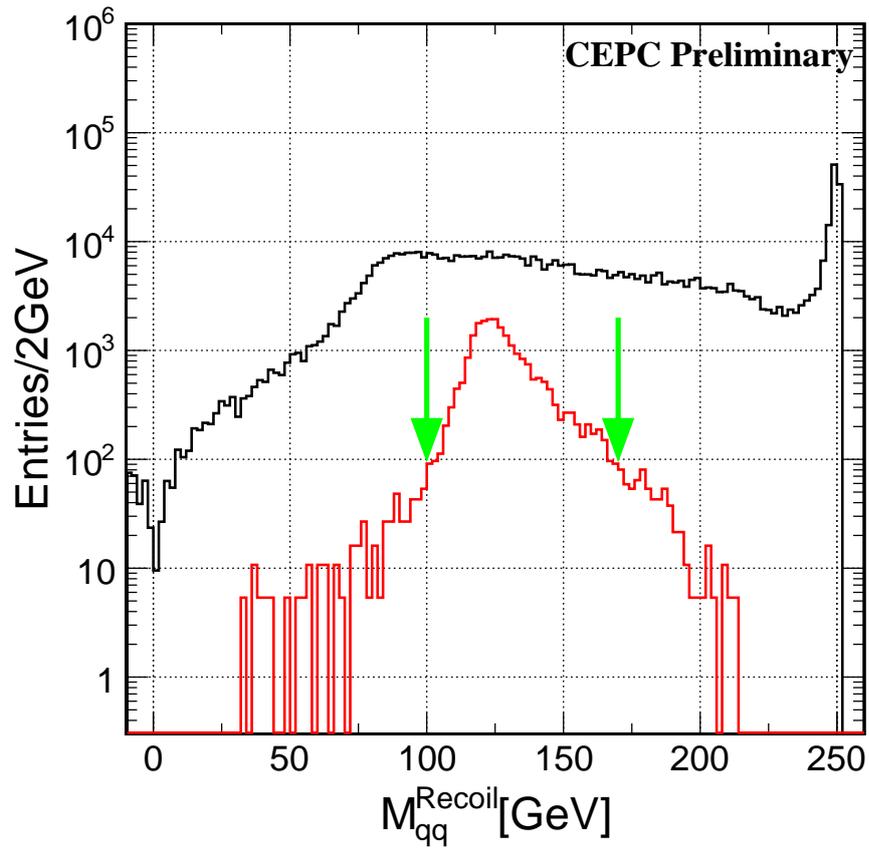


Figure 7.16: Distribution of the recoil mass of the qq , M_{qq}^{recoil} for $qqH\tau\tau$ and backgrounds at $\sqrt{s} = 250\text{GeV}$, the red/black line is for signal ($qqH\tau\tau$)/background(inclusive). The arrows indicates the cuts applied.

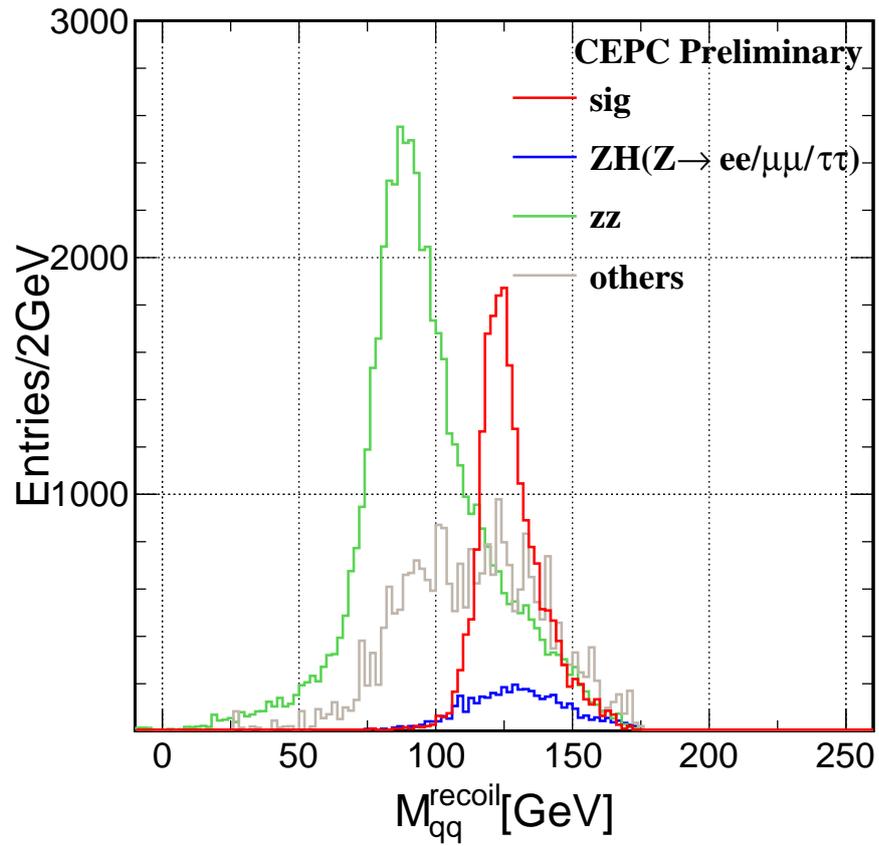


Figure 7.17: Distribution of the recoil mass of the qq , M_{qq}^{recoil} for $qqH\tau\tau$ and each backgrounds at $\sqrt{s} = 250\text{GeV}$ after the previous cuts

1582 49.97%. The finding efficiency of each τ can be expressed as: N_{found}/N_{truth} , where N_{found}
 1583 is the number of tagged τ 's and the leading track is close to a τ decayed track from the
 1584 MC information, while N_{truth} is the number of MC τ 's. Here the efficiency in the $qqH\tau\tau$
 1585 channel is 70.7%. In the similar way the purity defined as N_{found}/N_{total} where N_{total} is
 1586 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 $qqH \rightarrow \tau\tau$ selection on signal and inclusive SM back-
 grounds

	$qqH\tau\tau$	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
$20 \text{ GeV} < M_{\tau^+\tau^-}$ $< 120 \text{ GeV}$	24284	6290	32344	88245	597480	24927	36039	56615
$70 \text{ GeV} < M_{qq}$ $< 110 \text{ GeV}$	22937	2103	4887	65625	21718	738	1893	556
$100 \text{ GeV} < M_{qq}^{Rec}$ $< 170 \text{ GeV}$	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%

1587 From the table, the background of WW and ZZ are more important than the others, this
 1588 is because of the sub channel of their semi-leptonic decay with q jets and leptons or even
 1589 τ s, which is irreducible. The statistics of signal and the main backgrounds are shown in
 1590 Figure 7.18. The branching ratio $Br(H \rightarrow \tau\tau)$ can be calculated from the fit result and
 1591 previous selection efficiency to be 6.25 ± 0.04 , and the expected accuracy to be 1.30%.

1592 7.5 Combined Results

1593 To conclude, the τ reconstruction at the CEPC is currently catagorized into leptonic and
 1594 hadronic events and reconstructed using different strategies and τ finding algorithms.
 1595 In the leptonic events, where the τ lepton is generated only in association with leptons,
 1596 photons or missing energy, the τ events identification relies strongly on a successful
 1597 reconstruction of the photons and charged hadrons.

1598 In the hadronic events, it is more difficult to suppress the background, for further study,
 1599 the correlation with other channels might be applied.

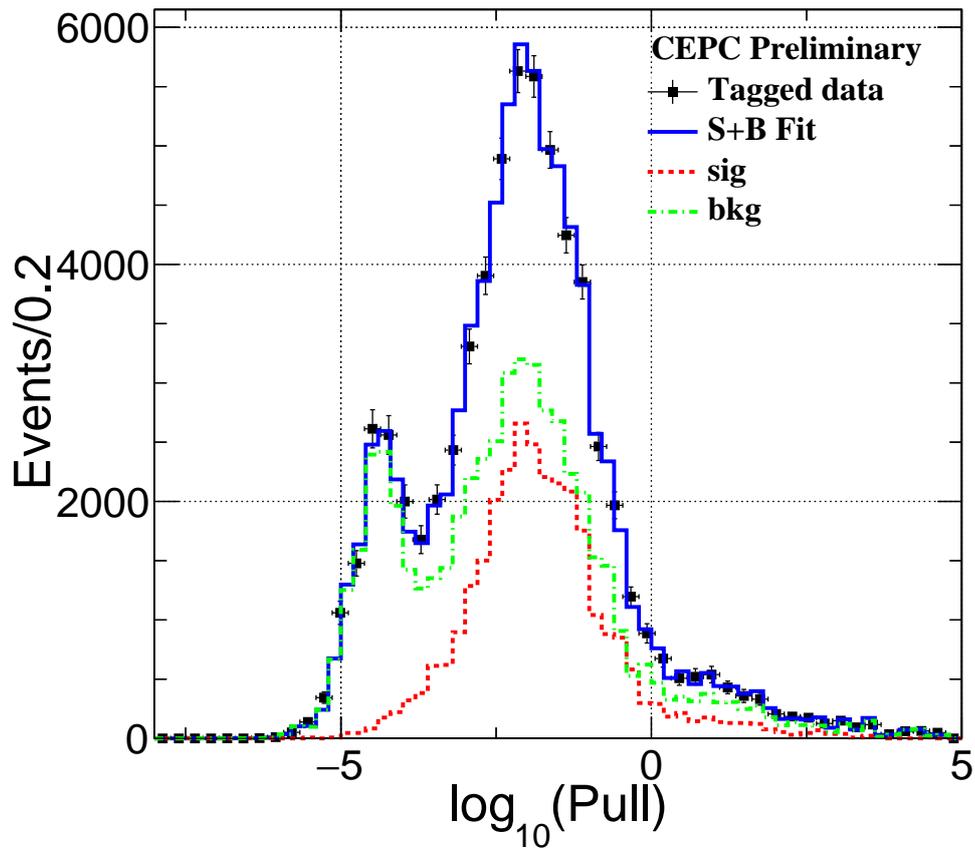


Figure 7.18: Fit of the sum of $D0^2$ and $Z0^2$ of the leading tracks of two cones with SM background included

1600 With these channels analyzed and the cross section of Higgs decaying to $\tau\tau$ can be sum-
 1601 marized as in Table 7.6

Table 7.6: Combined cross section

	BR ($H \rightarrow \tau\tau$)	$\delta (\sigma \times \text{BR}) / (\sigma \times \text{BR})$
$\mu\mu H$	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%

1602 In both cases, a precise reconstruction of the impact parameter is essential for the τ
 1603 events identification, as shown in the figures, the statistics can be fitted only if the posi-
 1604 tion resolution is good enough to distinguish the two peaks for τ s and backgrounds.

1605 7.6 Extrapolating in ILC

1606 The cross section for three polarization scenarios in ILC at 250GeV is shown in Chapter
 1607 2.

1608 Comparing these cross sections with the cross section at CEPC as shown in previous
 1609 section, a simple extrapolation can be done as in Table 7.7. The assumption here is that
 1610 the efficiency for each signal and background stays the same for ILC and CEPC.

Table 7.7: Extrapolated accuracy $\delta (\sigma \times \text{BR}) / (\sigma \times \text{BR})$ 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

1611 7.7 Discussion

1612 In this chapter, different channels with Higgs decaying into $\tau\tau$ at CEPC have been stud-
 1613 ied and the combined accuracy is reaching 1% level. This result is also extrapolated to
 1614 ILC and also gives the reasonable accuracy.

1615 This accuracy has still space to be improved. One choice is to use the collinear ap-
1616 proximation to recover the momentum of neutrino(s) from τ . This method needs to
1617 assume that τ decay products almost flight back-to-back. The collinear approximation
1618 will help to reconstruct the invariant mass of tau pair system and its comparison to the
1619 Higgs mass could be a powerful variable to suppress ZZ/WW backgrounds with τ final
1620 states.

1621 Another method is to fully reconstruct hadronically decaying τ momenta by making use
1622 of the interaction point position, the impact parameters of the τ decay products, and the
1623 transverse momentum of the Z boson recoiling against the $\tau\tau$ system [67]. Since more
1624 than 60% of τ s decays into hadrons, this method will help to improve the performance
1625 of these channels.

1626 Besides, a jet clustering algorithm can be applied in the qqH channel in order to sup-
1627 press the 2f backgrounds with jets.

1628 However, this study here is based on a perfect vertex detector, the resolution is not taken
1629 into account. Since the result was obtained from the impact parameter, the influence of
1630 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 ($Z \rightarrow \nu\nu$, $H \rightarrow gg$) 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 effi-

1659 ciency better than 99.5% in identifying the muons and the electrons, and 98% for pions.
1660 The algorithm is also tested in full simulated events, showing that LICH is powerful
1661 in these events to select high energy leptons, In the jet environment, the performance
1662 is limited by the isolation performance and the unbalanced statistics for leptons and
1663 hadrons. Since the particle identification requires high granularity for the segmentation,
1664 the performance of different granular calorimeters has been studied, showing that the
1665 efficiency of finding two leptons decreases by 1~2 % when the cell size doubles, which
1666 means that the detector needs 2~4% more statistics in the running. Another advantage
1667 of LICH is that the identification condition is adjustable according to the analysis. In
1668 the preparation of CEPC CDR, most of the physics are analyzed with LICH.

1669 The reconstruction of all final state particles in PFA also allows it to reconstruct τ events
1670 with higher efficiency. Since the multiplicity of τ is much smaller than that of jets, the
1671 $H \rightarrow \tau\tau$ events can easily be recognized in the leptonic channels, where the leptons
1672 decayed from Z can be vetoed by their recoil mass. In hadronic events, the method of
1673 defining well-isolated cones with smaller multiplicity is used to choose the τ candidates.
1674 The reconstructed τ candidates are selected to deduce the information of the di- τ system
1675 and the qq system. The irreducible backgrounds such as ZZ and ZH with Z decaying
1676 to $\tau\tau$, are reduced to 1% level.

1677 Thanks to the efficient vertex detector, the starting point of particles can be measured
1678 with excellent resolution. Therefore, the impact parameter is used in the tagging of τ , as
1679 a method to get the statistic of signals and backgrounds. At the center-of-mass energy
1680 of 250 GeV and 5000 fb⁻¹, the obtained precisions for the production cross section times
1681 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%
1682 for qqH . After extrapolating the result for $\mu\mu H$ to eeH , the combined accuracy of the
1683 $H \rightarrow \tau\tau$ is 0.89%, the simple extrapolation to ILC gives an accuracy of 0.84%.

1684 In conclusion, the CEPC detector design is still under optimization, by using the particle
1685 flow algorithm. The lepton identification not only provides tools for analysis but also
1686 helps to optimize the detector. The τ analysis shows that at current detector design,
1687 the accuracy can achieve the 1% level or even better, satisfying the requirement of new
1688 physics.

Bibliography

1689

- 1690 [1] Sha Bai, Jie Gao, Yiwei Wang, Qinglei Xiu, Weichao Yao, and Teng Yue. MDI De-
1691 sign in CEPC partial double ring. In *7th International Particle Accelerator Confer-*
1692 *ence (IPAC'16), Busan, Korea, May 8-13, 2016*, pages 3802–3804. JACOW, Geneva,
1693 Switzerland, 2016.
- 1694 [2] Chris Adolphsen. The International Linear Collider Technical Design Report-
1695 Volume 3. II: Accelerator Baseline Design. Technical report, Argonne National Lab-
1696 oratory (ANL), Argonne, IL (United States); Thomas Jefferson National Accelerator
1697 Facility (TJNAF), Newport News, VA (United States); Brookhaven National Labo-
1698 ratory (BNL), Upton, NY (United States); SLAC National Accelerator Laboratory
1699 (SLAC), Menlo Park, CA (United States); Fermi National Accelerator Laboratory
1700 (FNAL), Batavia, IL (United States), 2013.
- 1701 [3] Frank Gaede, Ties Behnke, Norman Graf, and Tony Johnson. LCIO-A persistency
1702 framework for linear collider simulation studies. *arXiv preprint physics/0306114*,
1703 2003.
- 1704 [4] O Wendt, F Gaed, and T Krämer. Marlin: Modular Analysis and Reconstruction
1705 for the LINear collider. Technical report, LC-DET-2007-001, 2007.
- 1706 [5] Sheldon L Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22
1707 (4):579–588, 1961.
- 1708 [6] Steven Weinberg. A model of leptons. *Physical review letters*, 19(21):1264, 1967.
- 1709 [7] M Veltman et al. Regularization and renormalization of gauge fields. *Nuclear*
1710 *Physics B*, 44(1):189–213, 1972.
- 1711 [8] Harald Fritzsch, Murray Gell-Mann, and Heinrich Leutwyler. Advantages of the
1712 color octet gluon picture. *Physics Letters B*, 47(4):365–368, 1973.
- 1713 [9] FJ Hasert, S Kabe, W Krenz, J Von Krogh, D Lanske, J Morfin, K Schultze, H Weerts,
1714 G Bertrand-Coremans, Jean Sacton, et al. Observation of neutrino-like interactions
1715 without muon or electron in the Gargamelle neutrino experiment. *Nuclear Physics*
1716 *B*, 73(1):1–22, 1974.

- 1717 [10] Ugo Amaldi, Albrecht Böhm, LS Durkin, Paul Langacker, Alfred K Mann, William J
1718 Marciano, Alberto Sirlin, and HH Williams. Comprehensive analysis of data per-
1719 taining to the weak neutral current and the intermediate-vector-boson masses.
1720 *Physical Review D*, 36(5):1385, 1987.
- 1721 [11] Serguei Chatrchyan, Vardan Khachatryan, Albert M Sirunyan, Armen Tumasyan,
1722 Wolfgang Adam, Ernest Aguilo, T Bergauer, M Dragicevic, J Erö, C Fabjan, et al.
1723 Observation of a new boson at a mass of 125 GeV with the CMS experiment at the
1724 LHC. *Physics Letters B*, 716(1):30–61, 2012.
- 1725 [12] Georges Aad, T Abajyan, B Abbott, J Abdallah, S Abdel Khalek, AA Abdelalim,
1726 O Abidinov, R Aben, B Abi, M Abolins, et al. Observation of a new particle in the
1727 search for the Standard Model Higgs boson with the ATLAS detector at the LHC.
1728 *Physics Letters B*, 716(1):1–29, 2012.
- 1729 [13] Hans Peter Nilles. Supersymmetry, supergravity and particle physics. *Physics Re-*
1730 *ports*, 110(1-2):1–162, 1984.
- 1731 [14] Howard E Haber and Gordon L Kane. The search for supersymmetry: probing
1732 physics beyond the standard model. *Physics Reports*, 117(2-4):75–263, 1985.
- 1733 [15] John Ellis, Giovanni Ridolfi, and Fabio Zwirner. Radiative corrections to the masses
1734 of supersymmetric Higgs bosons. *Physics Letters B*, 257(1-2):83–91, 1991.
- 1735 [16] Ties Behnke, James E Brau, Brian Foster, Juan Fuster, Mike Harrison, James McE-
1736 wan Paterson, Michael Peskin, Marcel Stanitzki, Nicholas Walker, and Hitoshi Ya-
1737 mamoto. The International Linear Collider Technical Design Report-Volume 1: Ex-
1738 ecutive Summary. *arXiv preprint arXiv:1306.6327*, 2013.
- 1739 [17] Abdelhak Djouadi, Joseph Lykken, Klaus Mönig, Yasuhiro Okada, Mark Oreglia,
1740 and Satoru Yamashita. International Linear Collider reference design report vol-
1741 ume 2: physics at the ILC. *arXiv preprint arXiv:0709.1893*, 2007.
- 1742 [18] CEPC-SPPC study group et al. CEPC-SPPC Preliminary Conceptual Design Re-
1743 port. 1. Physics and Detector. Technical report, IHEP-CEPC-DR-2015-01, 2015.
- 1744 [19] Johannes Gutleber. Future Circular Collider, overview. Technical report, FCC-
1745 DRAFT-MGMT-2016-005, 2014.
- 1746 [20] Henri Videau. Energy flow or particle flow-the technique of "energy flow" for
1747 pedestrians. In *International Conference on Linear Colliders-LCWS04*, pages 105–120.
1748 Ecole Polytechnique Palaiseau, 2004.
- 1749 [21] C Patrignani, Particle Data Group, et al. Review of particle physics. *Chinese physics*
1750 *C*, 40(10):100001, 2016.

- 1751 [22] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von To-
1752 erne, Helge Voss, M Backes, T Carli, O Cohen, A Christov, et al. TMVA-Toolkit for
1753 multivariate data analysis. *arXiv preprint physics/0703039*, 2007.
- 1754 [23] G. Alexander et al. A Precise measurement of the tau polarization and its for-
1755 ward - backward asymmetry at LEP. *Z. Phys.*, C72:365–375, 1996. doi: 10.1007/
1756 s002880050257.
- 1757 [24] Gordon L Kane. *Modern elementary particle physics: the fundamental particles and*
1758 *forces*. Addison-Wesley, 1993.
- 1759 [25] Michael Edward Peskin. *An introduction to quantum field theory*. Westview press,
1760 1995.
- 1761 [26] Howard Georgi. *Lie algebras in particle physics: from isospin to unified theories*, vol-
1762 ume 54. Westview press, 1999.
- 1763 [27] François Englert and Robert Brout. Broken symmetry and the mass of gauge vector
1764 mesons. *Physical Review Letters*, 13(9):321, 1964.
- 1765 [28] Peter W Higgs. Broken symmetries and the masses of gauge bosons. *Physical*
1766 *Review Letters*, 13(16):508, 1964.
- 1767 [29] Lisa Randall and Raman Sundrum. Large mass hierarchy from a small extra di-
1768 mension. *Physical Review Letters*, 83(17):3370, 1999.
- 1769 [30] Ali H Chamseddine, Ro Arnowitt, and Pran Nath. Locally supersymmetric grand
1770 unification. *Physical Review Letters*, 49(14):970, 1982.
- 1771 [31] Luis Ibáñez. Locally supersymmetric SU (5) grand unification. *Physics Letters B*,
1772 118(1-3):73–78, 1982.
- 1773 [32] John Baez and John Huerta. The algebra of grand unified theories. *Bulletin of the*
1774 *American Mathematical Society*, 47(3):483–552, 2010.
- 1775 [33] Julius Wess and Jonathan Bagger. *Supersymmetry and supergravity*. Princeton uni-
1776 versity press, 1992.
- 1777 [34] C Arzt, MB Einhorn, and J Wudka. Patterns of deviation from the standard model.
1778 *Nuclear Physics B*, 433(1):41–66, 1995.
- 1779 [35] Damien M Pierce, Jonathan A Bagger, Konstantin T Matchev, and Ren-jie Zhang.
1780 Precision corrections in the minimal supersymmetric standard model. *Nuclear*
1781 *Physics B*, 491(1-2):3–67, 1997.
- 1782 [36] Nima Arkani-Hamed, Savas Dimopoulos, and Gia Dvali. The hierarchy problem
1783 and new dimensions at a millimeter. *Physics Letters B*, 429(3-4):263–272, 1998.

- 1784 [37] G Apollinari, I Béjar Alonso, Oliver Brüning, M Lamont, and Lucio Rossi. High-
1785 Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report. Tech-
1786 nical report, Fermi National Accelerator Laboratory (FNAL), Batavia, IL (United
1787 States), 2015.
- 1788 [38] Lyn Evans and Shinichiro Michizono. The International Linear Collider Machine
1789 Staging Report 2017. 2017.
- 1790 [39] Ties Behnke et al. The international linear collider technical design report-volume
1791 4: detectors. Technical report, Argonne National Laboratory (ANL), Argonne,
1792 IL (United States); Pacific Northwest National Laboratory (PNNL), Richland,
1793 WA (United States); SLAC National Accelerator Laboratory (SLAC), Menlo Park,
1794 CA (United States); Fermi National Accelerator Laboratory (FNAL), Batavia, IL
1795 (United States), 2013.
- 1796 [40] CEPC-SPPC Study Group et al. CEPC-SPPC Preliminary Conceptual Design Re-
1797 port, Volume II-Accelerator. *IHEP, Beijing, China, Rep. IHEP-AC-2015-01*, 2015.
- 1798 [41] CEPC-SPPC Study Group et al. Cepec-sppc progress report, accelerator (2015–2016).
1799 Technical report, IHEP-CEPC-DR-2017-01, IHEP-CEPC-AC-2017-01, 2017.
- 1800 [42] Cai Meng, Jingru Zhang, Dou Wang, Xiaoping Li, Guoxi Pei, Shilun Pei, and Yun-
1801 long Chi. CEPC Linac Design and Beam Dynamics. 2017.
- 1802 [43] Dou Wang, Jie Gao, Feng Su, Yuan Zhang, Jiyuan Zhai, Yiwei Wang, Sha Bai, Huip-
1803 ing Geng, Tianjian Bian, Xiaohao Cui, et al. CEPC partial double ring scheme and
1804 crab-waist parameters. In *The Future of High Energy Physics: Some Aspects*, pages
1805 179–188. World Scientific, 2017.
- 1806 [44] Jean-Claude Brient and Henri Videau. The calorimetry at the future e+ e-linear
1807 collider. *arXiv preprint hep-ex/0202004*, 2002.
- 1808 [45] MA Thomson. Particle flow calorimetry and the PandoraPFA algorithm. *Nuclear
1809 Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, De-
1810 tectors and Associated Equipment*, 611(1):25–40, 2009.
- 1811 [46] Florian Beaudette. The CMS Particle Flow Algorithm. *arXiv preprint
1812 arXiv:1401.8155*, 2014.
- 1813 [47] Collaboration ATLAS. Letter of Intent for the Phase-II Upgrade of the ATLAS Ex-
1814 periment. Technical report, 2012.
- 1815 [48] Daniel Jeans, Marcel Reinhard, and Jean-Claude Brient. GAMMA Reconstruction
1816 at a LInear Collider. 2013.
- 1817 [49] Manqi Ruan. Arbor, a new approach of the Particle Flow Algorithm. *arXiv preprint
1818 arXiv:1403.4784*, 2014.

- 1819 [50] Damir Buskulic, D Casper, I De Bonis, D Decamp, P Ghez, C Goy, J-P Lees, M-N
1820 Minard, P Odier, B Pietrzyk, et al. Performance of the ALEPH detector at LEP.
1821 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrom-*
1822 *eters, Detectors and Associated Equipment*, 360(3):481–506, 1995.
- 1823 [51] P Moras de Freitas et al. MOKKA: A detailed Geant4 simulation for the Interna-
1824 tional Linear Collider detectors, 2003.
- 1825 [52] Sea Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai,
1826 D Axen, S Banerjee, G Barrand, et al. GEANT4—a simulation toolkit. *Nuclear in-*
1827 *struments and methods in physics research section A: Accelerators, Spectrometers, Detec-*
1828 *tors and Associated Equipment*, 506(3):250–303, 2003.
- 1829 [53] Zhenxing Chen, Ying Yang, Manqi Ruan, Dayong Wang, Gang Li, Shan Jin, and
1830 Yong Ban. Study of Higgsstrahlung Cross Section and Higgs Mass Measurement
1831 Precisions with $ZH (Z \rightarrow \mu_+ \mu_-)$ events at CEPC. *arXiv preprint arXiv:1601.05352*,
1832 2016.
- 1833 [54] Manqi Ruan, Daniel Jeans, Vincent Boudry, Jean-Claude Brient, and Henri Videau.
1834 Fractal Dimension of Particle Showers Measured in a Highly Granular Calorimeter.
1835 *Physical review letters*, 112(1):012001, 2014.
- 1836 [55] Aleph Collaboration et al. Measurement of the Tau Polarisation at LEP. *arXiv*
1837 *preprint hep-ex/0104038*, 2001.
- 1838 [56] Eric Braaten, Stephan Narison, and A Pich. QCD analysis of the tau hadronic
1839 width. *Nuclear Physics B*, 373(3):581–612, 1992.
- 1840 [57] Antonio Pich. Precision tau physics. *Progress in Particle and Nuclear Physics*.
- 1841 [58] CLEO collaboration et al. Experimental tests of lepton universality in {tau} decay.
1842 *Physical Review, D*, 55(5), 1997.
- 1843 [59] DP Roy. The hadronic tau decay signature of a heavy charged Higgs boson at LHC.
1844 *Physics Letters B*, 459(4):607–614, 1999.
- 1845 [60] Matthew J Dolan, Christoph Englert, and Michael Spannowsky. Higgs self-
1846 coupling measurements at the LHC. *Journal of High Energy Physics*, 2012(10):112,
1847 2012.
- 1848 [61] Seong-Youl Choi and Manuel Drees. Signals for CP violation in scalar tau pair
1849 production at muon colliders. *Physical review letters*, 81(25):5509, 1998.
- 1850 [62] Nathan Isgur and Mark B Wise. Weak decays of heavy mesons in the static quark
1851 approximation. *Physics Letters B*, 232(1):113–117, 1989.

- 1852 [63] Serguei Chatrchyan, V Khachatryan, AM Sirunyan, A Tumasyan, W Adam,
1853 T Bergauer, M Dragicevic, J Eroe, C Fabjan, M Friedl, et al. Search for neutral
1854 Higgs bosons decaying to tau pairs in pp collisions at. *Physics Letters B*, 713(2):
1855 68–90, 2012.
- 1856 [64] Georges Aad, Brad Abbott, Jalal Abdallah, S Abdel Khalek, O Abidinov, Rosemarie
1857 Aben, Babak Abi, Maris Abolins, OS AbouZeid, Halina Abramowicz, et al. Evi-
1858 dence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detec-
1859 tor. *Journal of high energy physics*, 2015(4):117, 2015.
- 1860 [65] Trong Hieu Tran, Vladislav Balagura, Vincent Boudry, J-C Brient, and Henri
1861 Videau. Reconstruction and classification of tau lepton decays with ILD. *The Euro-
1862 pean Physical Journal C*, 76(8):468, 2016.
- 1863 [66] Wolfgang Kilian, Thorsten Ohl, and Jürgen Reuter. Whizard—simulating multi-
1864 particle processes at lhc and ilc. *The European Physical Journal C*, 71(9):1742, 2011.
- 1865 [67] Daniel Jeans. Tau lepton reconstruction at collider experiments using impact pa-
1866 rameters. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,
1867 Spectrometers, Detectors and Associated Equipment*, 810:51–58, 2016.

1868 Acknowledgement

1869 Firstly, I would like to thank my two supervisors, Vincent and Manqi, for your supports
1870 during my PhD study and research. Your guidance helped me in the research and also
1871 while writing this thesis. Your patience and advices encouraged me to finish my thesis.

1872 I would also like to thank Mark Thomson and Luca for accepting the roles of the reporter
1873 as well as Florian, especially Zhiqing Zhang, for presiding over my thesis jury.

1874 And thank you, Jean-Claude, for accepting me to do this thesis at the LLR, and for your
1875 ideas and guidances. I would like to express my sincere gratitude to Henri Videau, you
1876 are really helpful to me, you always answer my stupid question, correct my error, I
1877 would not have my thesis without you.

1878 My thanks also goes to my fellow labmates in LLR, Vladik, Hieu, Yacine, Kostia, and so
1879 on, all of you have helped me during this three years, or since even earlier.

1880 I would also like to take this opportunity to express my gratitude to my family: my
1881 parents and my sister. It is your supports that guaranteed my study in France and kept
1882 me away from the worries.

1883 Thanks to my neighbors in Building 76, especially K & S, for feeding me with delicious
1884 soups, as well as the others, thank you for all the fun we have had and for providing
1885 me miraculous life in the campus. To my friends at IHEP, thank you for welcoming me
1886 and helped me adapting the life in Beijing.

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

Mots clés : Higgs, canaux tau, calorimétrie

1887

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 granula-

rité. 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%.

