Detector optimisation for future linear collider

Boruo Xu
of King's College

A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

Abstract

This is my abstract. To be or not to be.

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the respective Degree Committee.

Boruo Xu

Acknowledgements

Of the many people who deserve thanks, some are particularly prominent, such as my supervisor…

Preface

This will be my preface. Where is Wolly?

Contents

1 Let's make introduction great again **1**

1.1 Future Linear Colliders 1

1.2 Motivation 1

2 Theoretical overview **3**

2.1 Overview of the Standard Model 3

2.2 Notations and conventions 5

2.3 Quantum electrodynamics 5

2.4 Quantum chromodynamics 6

2.5 The electroweak interaction 6

2.6 Higgs Mechanism 8

2.7 Yukawa couplings 8

2.8 Standard Model Higgs boson 9

2.9 Higgs beyond the Standard Model 10

3 Detector and Physics at Future Linear Colliders **15**

3.1 ILC 15

3.1.1 ILD vs SID 16

3.2 CLIC 16

3.3 Physics at future linear colliders 18

3.4 Physics requirements on the detector design 18

3.5 CLIC vs ILC 19

3.6 International Large Detector 20

3.7 Overview of ILD sub-detectors 22

3.7.1 Vertex Detector 22

3.7.2 Tracking Detectors 22

3.7.3 Electromagnetic Calorimeter 23

3.7.4 Hadronic Calorimeter 24

3.7.5 Solenoid 25

3.7.6 Yoke and Muon system . . . 26
3.7.7 Very Forward Calorimeters . . . 26

4 Simulation and Reconstruction **29**
4.1 Monte Carlo event generation . . . 29
4.2 Event Simulation . . . 30
4.3 Event Reconstruction . . . 30
4.4 PandoraPFA . . . 30
4.4.1 Track selection . . . 31
4.4.2 Calorimeter selection . . . 31
4.4.3 Cone Clusters Algorithm . . . 32
4.4.4 Particle Identification . . . 33
4.4.5 Clustering . . . 33
4.4.6 Topological cluster association . . . 33
4.4.7 Track-cluster association . . . 33
4.4.8 Re-clustering . . . 34
4.4.9 Fragment removal . . . 34
4.4.10 Particle Flow Object Creation . . . 34
4.5 Suppression of $\gamma\gamma \to hadrons$ backgrounds . . . 34

5 Photon Reconstruction in PandoraPFA **37**
5.1 Overview of photon reconstruction in PandoraPFA . . . 37
5.2 Photon reconstruction algorithm . . . 38
5.2.1 Form photon clusters . . . 38
5.2.2 Reconstruct photon candidates . . . 39
5.2.3 Photon ID test . . . 40
5.2.4 Photon Fragment removal . . . 40
5.3 Two dimensional peak finding algorithm for photon candidate . . . 40
5.3.1 Candidate close to track projection . . . 41
5.3.2 Peak filtering . . . 42
5.3.3 Inclusive mode . . . 42
5.4 Likelihood classifier for photon ID . . . 42
5.4.1 Overview of Projective Likelihood . . . 43
5.4.2 Projective Likelihood in PandoraPFA . . . 43
5.5 Photon fragment removal algorithm in the ECAL . . . 44
5.6 High energy photon fragment recovery algorithm . . . 47
5.7 Photon splitting algorithm . . . 50

5.8 Photon reconstruction performance improvement 51
5.9 Breakdown of photon reconstruction improvement 54
5.10 Photon reconstruction performance 55

6 Tau Lepton Final State Separation 57
6.1 Introduction 57
6.2 Simulation and reconstruction 58
6.3 Generator level cut 58
6.4 Decay modes 59
6.5 Discriminative variables 60
6.6 Multivariate Analysis 64
6.7 Result 65
6.8 Electromagnetic calorimeter optimisaiton 66

7 Double Higgs Bosons Production Analysis 71
7.1 Analysis Straggly Overview 72
7.2 Monte Carlo Sample Generation 72
7.3 Physics object and event reconstruction 73
7.3.1 Electron and muon identification 75
7.3.2 Tau identification 77
7.3.3 Very forward electron identification 79
7.3.4 Other lepton identification processors 79
7.4 Jet reconstruction 80
7.4.1 Jet reconstruction optimisation 80
7.4.2 Jet flavour tagging 85
7.4.3 Jet pairing 85
7.5 Pre-selection 86
7.5.1 Discriminative pre-selection cuts 87
7.5.2 Sanity cuts 92
7.5.3 Mutually exclusive cuts for $\mathrm{HH} \rightarrow \mathrm{b\bar{b}W^+W^-}$ and $\mathrm{HH} \rightarrow \mathrm{b\bar{b}b\bar{b}}$ . . . 92
7.6 Discriminative Variables 93
7.7 Multivariate analysis 96
7.8 Signal selection results 96
7.9 Couplings extration 96

Bibliography 103

List of figures 107

List of tables 109

"Two bags of pork scratchings are worth a bag of gold."

— Joris the Dutch

Chapter 1

Let's make introduction great again

"Introduction means introdcution"
— Theresa Trump

Introduction

1.1 Future Linear Colliders

Basic intro. LHC.

Next challenge

Future Options. FCC vs LC

LC options

1.2 Motivation

Photon - passage through matter. Photon electromagnetic shower

Since Higgs discovery in the LHC in 2012, Higgs

Ha there is a higgs.

We found higgs. Higgs is cool. It explains mass.

Why double higgs. Double higgs coulpling is unique to linear collider. It can revel much about the BSM models.

Generator level study has performed. ILC has done this this and that. gHHH in CLIC before

Here we do things differently. First subchannels, then extract both couplings simultaneously.

Chapter 2

Theoretical overview

"ILC will be built next year"
— Mysterious person

This chapter provides a theoretical overview which would be used in the subsequent chapters. A short review of the Standard Model of Particle Physics, the current best particle theory, is provided, with an emphasis on the Higgs mechanism and the Higgs boson. A general parametrisation of the Higgs theory, beyond the Standard Model, is discussed, and supplies the theoretical background for the physics analysis in the chapter 7.

2.1 Overview of the Standard Model

The Standard Model (SM) is a quantum field theory concerning three fundamental interactions of nature: the electromagnetic, weak and strong integrations. The SM also describes the interactions between the sub-atomic particles. The deployment and the experimental verification of the SM throughout the second half of the 20th century is one of the greatest triumph of the particle physics. The most recent discovery of the Higgs boson in 2012 [1] further verified the theory. This chapter summarise the SM based on the summaries of the SM [2–5].

The fundamental particles in the SM consist of three categories: force exchange bosons, leptons and neutrinos, and quarks. In the SM, the force exchange bosons carries the fundamental forces between particles. For example, photon is the force carrier of the electromagnetic force. W^+, W^-, and Z are the force carriers of the weak force. Gluon, g, is the force carrier of the strong force. These bosons will be discussed further in section 2.5. There is also the Higgs boson from the spontaneous symmetry breaking of the Higgs filed, which is discussed in section 2.8

Another category of fundamental particles contains leptons and neutrinos. These particles are fermions. For each fermion in the SM, there is an anti-fermion with same mass and spin, but opposite charge. Leptons and neutrinos have three generations. Each generation has same interaction properties, but different masses. Although neutrinos could not be directly detected, measurements of the Z decay width strong suggested the three generations of neutrinos [6]. Leptons and neutrinos experience weak forces as well as electromagnetic forces, which will be further discussed in section 2.5.

The last category of fundamental particles are quarks, which are also fermions and have three generations. Each generation has a positively charged up type quark and a negatively charged down type quark. Quarks experiences all three fundamental forces described by the SM.

The SM has enjoyed great success with theoretical predictions being experimentally varied. Some highlights included the discovery of the top quark in 1995 cite, the tau netrino in 2000 cite and the Higgs bosons in 2012 citeAad:2012tfa. However, there are observations which are not explained by the SM. One issue is that the SM does not incorporate the gravitational force. There have been attempts to modify the SM but no conclusive theory yet. Another issue is that the SM does not allow neutrino masses and mixings. There have been many theories beyond the Standard Model (BSM). One such example is the generalisation of the Higgs theory to allow non SM coupling strengths. This will be discussed in section 2.9.

Notations and conventions will be introduced. The overview of the Standard Model starts with the quantum electrodynamics, and its generalisation to quantum chromodynamics. The unification of electromagnetism and weak interaction, electroweak gauge theory will be discussed. Afterwards, Higgs mechanism and Yukawa couplings will be introduced to explain bosons and fermions masses whilst preserving the Lagrangian symmetry. This will be followed by a detailed discussion on the Standard Model Higgs boson, its mass and interactions with other particles. The chapter finishes with explanation for possible Higgs theories beyond the Standard Model, the Lagrangian of the Higgs interaction, and observables, which forms the theoretical background for the analysis on the double Higgs production in the chapter 7.

2.2 Notations and conventions

The natural unit is used in this document, $\hbar = c = 1$. The metric is mostly-minus, $\eta^{\mu\nu} = \text{diag}(1, -1, -1, -1)$. The Dirac gamma matrices are represented with γ^μ, with μ goes from 0 to 3. $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. $\overline{\psi} = \psi^\dagger\gamma^0$. Einstein summation convention is used as well this document.

This set of notations allows a contracted pair to be a Lorentz invariant. For a Weyl spinor, ψ_α, the mass term in the lagrangian is of the form $\psi^\alpha\psi_\alpha$, which is the Majorana mass term. The contracted pair between two different Weyl spinors would form a Dirac mass term.

2.3 Quantum electrodynamics

The natural starting point to introduce the SM is the quantum electrodynamics (QED). The QED is a quantum field theory explaining electromagnetic interactions. The theory involves a spin-half Dirac (electron) field ψ and a vector (photon) field A_μ. When the local (gauge) symmetry is imposed, which is equivalent to the Lagrangian invariance under transformations,

$$\psi \rightarrow e^{ie\phi(x)}, \; A_\mu \rightarrow A_\mu - \partial^\mu\phi(x), \tag{2.1}$$

the Lagrangian is fixed to be

$$\mathcal{L}_{\text{QED}} = \overline{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.2}$$

if up to cubic terms are allowed in the fields. There are two free parameters in the QED, m the electron mass, and e the electron charge. The mass term for the photon, $\nu^2 A_{nu} A^\nu$, is forbidden by gauge invariance.

The QED has been verified experimentally. One of the greatest prediction is the spin magnetic dipole moment of the electron, defined as $\vec{\mu} = g_s \frac{Qq}{2m}\vec{s}$. The g_s is predicted to be 2 by the Dirac equations. The small corrections to the value comes from the electron's interaction with virtual photons, so called higher "loop" corrections in Feynman diagrams. The precise agreement of the theoretical prediction and the experimental value is a success of the QED.

2.4 Quantum chromodynamics

Like the QED, the quantum chromodynamics (QCD) a quantum field theory explaining strong interactions. There are eight gauge bosons, gluons, coupling to nine fermions, quarks. Unlike the QED, the theory is invariant under local non-Abelian SU(3) transformations. Gulons can interact with other gluons, and carry colour charge (red, green, and blue). Nine quarks transform as colour triplets. The QCD Lagrangian is

$$\mathcal{L}_{\text{QCD}} = \sum_{f \in u,d,s,c,b,t} \overline{\psi} \left(i\gamma^{\mu}\partial_{\mu} - g_s \gamma^{\mu} G^{a}_{\mu} \frac{\lambda^{a}}{2} - m_f \right) \psi - \frac{1}{4} G^{a}_{\mu\nu} G^{a\mu\nu}, \tag{2.3}$$

where g_s is the strong coupling constant. a is the colour charge. λ is the Gell-Mann matrices. $G^{a}_{\mu\nu}$ is the gluon field strength, given by

$$G^{a}_{\mu\nu} = \partial_{\mu}\gamma^{a}_{\nu} - \partial_{\nu}\gamma^{a}_{\mu} - g_s f_{abc} G^{b}_{\mu} G^{c}_{\nu}. \tag{2.4}$$

The last extra term comparing to QED indicates the non-Abelian nature of the QCD.

2.5 The electroweak interaction

The electroweak interaction can be thought of a extension to the QED to incorporate the weak force, and to explain different coupling strength to left-handed and right-handed fermions. However, the Lagrangian does not allow the massive electroweak force exchange bosons and fermions, which are explained by the Higgs mechanism and the Yukawa interactions.

There are four vector boson fields in the theory, 3 W and 1 B field. The Lagrangian can be divided into two parts: the bosonic self interaction and the couplings to the fermions.

$$\mathcal{L}_{\text{Electroweak}} = \mathcal{L}_{\text{Boson}} + \mathcal{L}_{\text{Fermion}} \tag{2.5}$$

The bosonic self interaction Lagrangian, $\mathcal{L}_{\text{Boson}}$, is given by

$$\mathcal{L}_{\text{Boson}} = -\frac{1}{4} W^{i}_{\mu\nu} W^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.6}$$

where

$$W^i_{\mu\nu} = \partial_\nu W^i_\mu - \partial_\mu W^i_\nu - g\epsilon^{ijk} W^j_\mu W^k_\nu \tag{2.7}$$

$$B_{\mu\nu} = \partial_\nu B_\mu - \partial_\mu B_\nu \tag{2.8}$$

B field is invariant under U(1). W field is invariant under non-Abelian SU(2) transformations. g is the coupling strength of the W field. The indices, i, j, and k indicates 3 W fields, going from 1 to 3.

The fermionic part of the Lagrangian, $\mathcal{L}_{\text{Fermion}}$, has different components for the left-handed and right-handed fermions, given by

$$\mathcal{L}_{\text{Fermion}} = \sum_{\psi \in \text{fermions}} \overline{\psi}_L \gamma^\mu D^L_\mu \psi_L + \overline{\psi}_R \gamma^\mu D^R_\mu \psi_R \tag{2.9}$$

D^L_μ and D^R_μ are defined as

$$D^L_\mu = \partial_\mu + ig\frac{\tau_i}{2} W^i_\mu + ig' Y_\psi B_\mu \tag{2.10}$$

$$D^R_\mu = \partial_\mu + ig' Y_\psi B_\mu \tag{2.11}$$

This Lagrangian allows W and B field to couple with left-handed fermions, but only B field couples to right-handed fermions. The τ_i matrices are the generators of the SU(2). Pauli spin matrices are one of the representations. Y_ψ is the hypercharge associating with the fermion field ψ. g' is the B field strength.

Physical bosons W^+, W^- only couples to left-handed fermions. Z and γ couples to both left-handed and right-handed fermions. Hence W^1 and W^2 are associated to W^+ and W^-. Mass eigenstates for Z and γ, Z_μ and A_μ are mixture of W^3_μ and B_μ.

$$Z_\mu = \cos(\theta_W) W^3_\mu - \sin(\theta_W) B_\mu \tag{2.12}$$

$$A_\mu = \sin(\theta_W) W^3_\mu + \cos(\theta_W) B_\mu \tag{2.13}$$

θ_W is the Weinberg mixing angle [7], which is determined experimentally. So far the SU(2)⊗U(1) gauge theory explains the parity violating nature of the weak interaction.

The explicit fermion mass are not allowed in the gauge symmetry. The higgs mechanism via spontaneous symmertry breaking would introduce mass terms for fermions.

2.6 Higgs Mechanism

A complex scalar Higgs field, Φ_H, is added to the electroweak Lagrangian. Φ_H transforms as a doublet of SU(2) with hypercharge $Y = \frac{1}{2}$

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi_H)^\dagger (D^\mu \Phi_H) - \mu^2 \Phi_H^\dagger \Phi_H - \lambda \left(\Phi_H^\dagger \Phi_H\right)^2 \tag{2.14}$$

with

$$D_\mu \Phi_H = \left(\partial_\mu + ig\frac{\tau_i}{2} W_\mu^i + ig'\frac{1}{2} B_\mu\right) \Phi_H \tag{2.15}$$

For negative μ^2, the Higgs filed potential

$$\mu^2 \Phi_H^\dagger \Phi_H + \lambda \left(\Phi_H^\dagger \Phi_H\right)^2 \tag{2.16}$$

is minimised with a Higgs vacuum potential $\frac{v}{\sqrt{2}} = \sqrt{\frac{v^2}{2\lambda}}$. After the symmetry breaking, the $\mathcal{L}_{\text{Higgs}}$ provides the mass terms for W^+, W^-, Z and γ via terms in the Lagrangian:

$$\frac{(gv)^2}{4} W_\mu^+ W^{-\mu} + \frac{\left(g^2 + g'^2\right) \mu^2}{8} Z_\mu Z^\mu \tag{2.17}$$

This provides equal mass for W^+ and W^- with massless photon.

2.7 Yukawa couplings

Section 2.6 explains the Higgs mechanism for gauge bosons gaining masses. The fermions gain masses in a similar fashion. Consider a Higgs field transforming as a doublet of SU(2) with hypercharge $Y = \frac{1}{2}$, the Yukawa couplings is given

$$\mathcal{L}_{\text{Yukawa}} = -\lambda^u \overline{q_L} \Phi_H^c u_R - -\lambda^d \overline{q_L} \Phi_H d_R - \lambda^e \overline{l_L} \Phi_H e_R + \text{h.c.} \tag{2.18}$$

$\Phi_H^c \equiv i\sigma^2 H^*$ is an SU(2) double field with hypercharge $Y = -\frac{1}{2}$ and σ is the Pauli spin matrix. u, d, and e are fields for up-type quark, down-type quark and leptons. The Lagrangian is summed over all possible quarks and leptons. The interactions terms in the $\mathcal{L}_{Yukawa}$ become mass terms when the Higgs vacuum expectation value is substituted. The fermion masses are given by

$$m_u = \frac{\lambda^u v}{\sqrt{2}}, \ m_d = \frac{\lambda^d v}{\sqrt{2}}, \ m_e = \frac{\lambda^e v}{\sqrt{2}} \tag{2.19}$$

2.8 Standard Model Higgs boson

So far, interactions between different fields in the Standard Model, as well as the mass obtaining mechanism have been discussed. Only left for discussion is the Higgs bosons, and its interactions with other fields.

For the Higgs doublet complex field in the SM, there are four real scalar degrees of freedom. By choosing the unitary gauge, three degree of freedoms are manifestly eaten. The Higgs field becomes

$$H(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.20}$$

$h(x)$ is real scalar field for the Higgs boson. It is not charged under electromagnetism as it is real. The Higgs boson interaction terms in the Lagrangian with other particles can be shown by replacing v with $v + h(x)$ in previous expressions. For fermions field, ψ_i, the Higgs boson interaction term is given by

$$\mathcal{L} \supset -\frac{m_i}{v} h \overline{\psi}_i \psi_i \tag{2.21}$$

From the equation 2.17, the Higgs boson interaction terms for bosons can be shown as

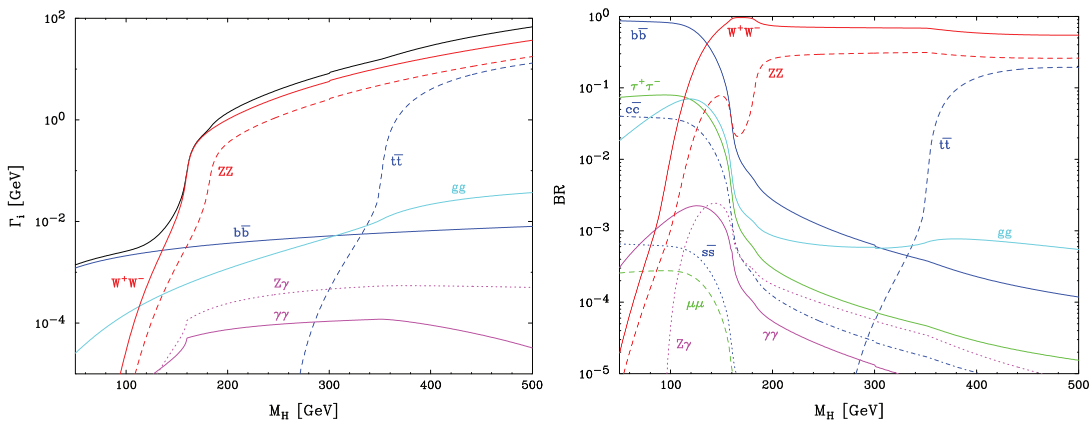
$$\mathcal{L} \supset m_W^2 \left(\frac{2h}{v} + \frac{h^2}{v^2} \right) W_\mu^+ W^{-\mu} + \frac{m_Z^2}{2} \left(\frac{2h}{v} + \frac{h^2}{v^2} \right) Z_\mu Z^\mu \tag{2.22}$$

The Higgs boson self interactions are obtained from the Higgs field potential

$$\mathcal{L} \supset \frac{\mu^2}{2}(\nu+h)^2 - \frac{\lambda}{4}(\nu+h)^4 \supset -\lambda\nu^2h^2 - \lambda\nu h^3 - \frac{\lambda}{4}h^4 = -\frac{m_h^2}{2}h^2 - \frac{m_h^2}{2\nu}h^3 - \frac{m_h^2}{8\nu^2}h^4. \tag{2.23}$$

The Higgs boson mass, m_h is $2\lambda\nu^2$. The triple and quartic self interaction strengths are $-\frac{m_h^2}{2\nu}$ and $\frac{m_h^2}{8\nu^2}$. Once the m_h is determined, λ can be worked out. The Higgs boson decay width and branching fraction can be roughly worked out. For example, figure 2.1a and figure 2.1b show partial decay width and the branching ratios as a function of m_h.

The Higgs boson decaying to a pair of heavier particles, such as W^+W^- or $Z\ Z$ is forbidden kinematically. However, figure 2.1 shows that the Higgs decaying to W^+W^- dominates before the mass threshold $m_H = 2m_W \sim 160$ GeV. This is because one of the $W^{\pm}$ gauge bosons is virtual and not on the mass shell, which is allowed by the quantum field theory. The virtual gauge boson subsequently decays to real on-shell particles.



(a) **(b)**

Figure 2.1: figure 2.1a shows Standard Model Higgs boson partial widths as a function of its mass, M_H. The total width is the black curve figure 2.1b shows selected Standard Model Higgs boson branching ratios as a function of its mass, M_H. Both plots are taken from [8]

2.9 Higgs beyond the Standard Model

Since the SM like Higgs boson discovery in 2012, it becomes important to understand of role of the Higgs boson in the electroweak spontaneous symmetry breaking. In the absence of the Higgs boson, the coupling strength of the longitudinally polarised vector bosons grows with energy and becomes strong at TeV scale. The SM Higgs bosons moderates the interacting strength, allowing the extraction of the weak coupling at short distances. In this scenario, the SM Higgs couplings are constrained and predicted in one parameter only, the Higgs mass. But other alternative scenarios could allow the behaviour of the SM Higgs at low energy. One such example, motivated by the hierarchy problem and the electroweak data, is that the light and narrow Higgs-scalar is a composite bound state of some strongly interacting sector at the TeV scale. The couplings of the Higgs to fermions and bosons would be different to those in the SM. If the composite Higgs is the pseudo Nambu-Goldstone boson from a spontaneous global symmetry breaking, the Higgs can be naturally light [9]. Another scenario is that a composite dilaton, the pseudo Nambu-Goldstone boson arose from a spontaneous scale invariance breaking, partially behaves like a light Higgs [10]. In both scenarios, the interaction of Higgs becomes strong at high energy. The coupling of the Higgs would deviate to those in the SM.

An important physics channel for testing the Higgs theory is the double Higgs production via vector boson fusion at high energy [11–13]. For the composite Higgs scenario, the scattering amplitude increases with the energy. For the dilaton scenario, no energy defence on the scattering amplitude is expected. It is difficult for the Large Hardon Collider to measure the cross section due to the large SM background rate [12]. However, a multi-TeV linear electron position collider, such as Compact Linear Collider, would be able to precisely measure the cross section [14].

Following the assumption made in the [12, 13], the self interaction of the light scalar Higgs, h, and its coupling to other SM bosons can be described by the following Lagrangian. The notation in the [13] is followed. After the electroweak symmetry breaking, the bosonic part of the Lagrangian reads:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu h\right) - V(h) + \left(m_W^2 W_\mu^+ W^{-\mu} + \frac{m_Z^2}{2} Z_\mu Z^\mu\right)\left[1 + 2a\frac{h}{v} + b\frac{h^2}{v^2} + \dots\right], \tag{2.24}$$

where $V(h)$ is the h field potential,

$$V(h) = \frac{1}{2}m_h^2 h^2 + d_3\left(\frac{m_h^2}{2v}\right)h^3 + d_4\left(\frac{m_h^2}{8v^2}\right)h^4 + \dots. \tag{2.25}$$

a, b, d_3 and d_4 are arbitrary dimensionless parameters. Higher order terms in h are omitted. a and b are proportional to the coupling strength of the VVh and $VVhh$ vertices, where $V = W^\pm, Z$. d_3 and d_4 are proportional to the triple and quartic h self coupling strength. Comparing with equation 2.22 and equation 2.23, the SM Higgs suggests $a = b = d_3 = d_4 = 1$ and all higher order terms vanish. The dilaton scenario imposes the relation, $a = b^2$.

The scattering amplitude for $V_L V_L \to hh$ can be written as

$$A = a^2(A_{SM} + A_1\delta_b + A_2\delta_{d_3}), \tag{2.26}$$

where A_{SM} is the SM amplitude and

$$\delta_b \equiv 1 - \frac{b}{a^2}, \tag{2.27}$$

$$\delta_{d_3} \equiv 1 - \frac{d_3}{a}. \tag{2.28}$$

A_1 grows like energy squared at large center-of-mass energy, $E \gg m_V$. A_{SM} and A_2 has no energy depended. Therefore, δ_b controls the magnitude of the scattering amplitude increasing as a function of energy. δ_{d_3}, however, determines the magnitude at threshold. In an electron-positron collider, this scattering process and be studied via $e^+e^- \to \nu\bar{\nu}hh$ channel. The cross section of the channel can be written as

$$\sigma = a^4\sigma_{SM}\left(1 + A\delta_b + B\delta_{d_3} + C\delta_b\delta_{d_3} + D\delta_b^2 + E\delta_{d_3}^2\right), \tag{2.29}$$

where σ_{SM} is the cross section predicted by the SM. With suitable kinematic cuts, high-energy behaviour can be disentailed from the physics at threshold, allowing the extraction of δ_b, δ_{d_3} and hence the coupling strength g_{VVHH} and g_{HHH}. Suitable observables are variables that increases with increasing centre-of-mass energy. Two examples of such variables are the invariant mass of the two Higgses system, m_{hh}, and the sum of their transverse momenta, H_T. figure **??** shows that the m_{hh} and H_T distributions are sensitive to the values of δ_b and δ_{d_3}. The figure shows the result of a general level study performed in [13].

In the equation 2.29, a, which is proportional to g_{VVH}, only entres as an overall factor. However, a also appears in the definition of δ_b and δ_{d_3}. To extract δ_b and δ_{d_3}, a should be a constant ideally. For a multi-TeV electron-positron collider, the cross section for single Higgs production is far greater than that of the double Higgs. Figure 2.3 shows the comparison of the cross section as a function of the centre-of-mass energy. Therefore, g_{VVH} and a will be measured precisely before investigating the double Higgs production.

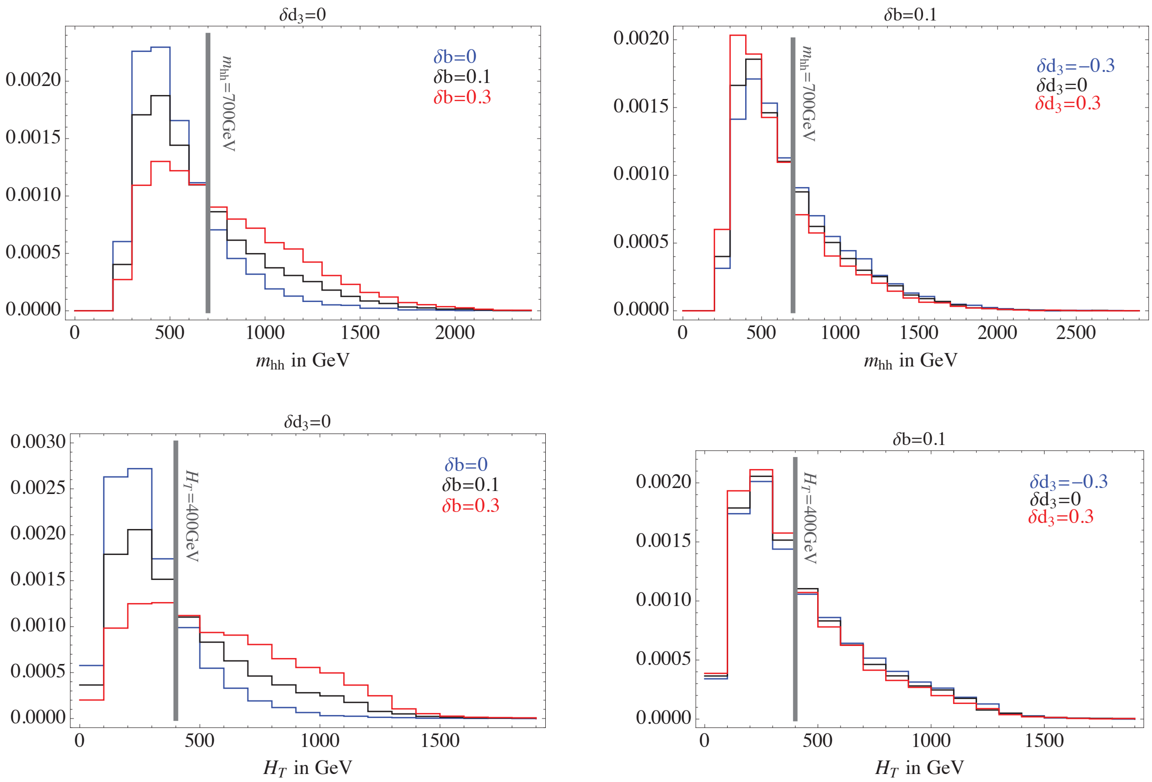


Figure 2.2: Normalized differential cross sections $d\sigma/dm_{hh}$ and $d\sigma/dH_T$ for $e^+e^- \to \nu\bar{\nu}hh$ at the Compact Linear Collider, with $\sqrt{s} = 3\,\text{TeV}$ TeV after the identification cuts, for several values of δ_b and δ_{d_3}. Plot is taken from [13].

For the purpose of measuring g_{VVHH} and g_{HHH} via double Higgs production, a in the equation 2.29 can be treated as a constant.

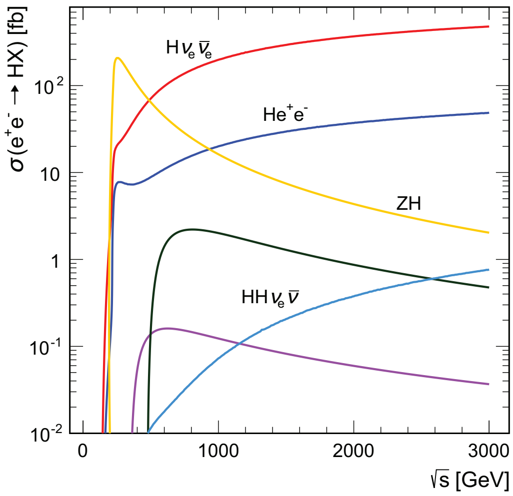


Figure 2.3: Cross section as a function of centre-of-mass energy for the Higgs production processes at an electron-positron collider for a Higgs mass of 126GeV. The values shown correspond to unpolarised beams and do not include the effect of beamstrahlung. Plot is taken from [15].

Chapter 3

Detector and Physics at Future Linear Colliders

"ILC will be built next year"
— Mysterious person

Since the discovery of a particle consistent with being the SM Higgs boson in LHC at 2012 [1, 16], our understanding of Standard Model has improved greatly. Yet limited by the underlying QCD interaction from proton-anti-proton collision, one has great difficulty to measure the properties of the Higgs precisely. Next generation electron-positron linear collider could hopefully make precision measurements of the Higgs sector and the Top quark sector [?].

3.1 ILC

Two leading candidates for next generation electron-positron linear collider are the International Linear Collider (ILC) [17], and the Compact Linear Collider (CLIC) [18]. The ILC is a high-luminosity electron-positron linear collider with centre-of-mass energy from 200 GeV up to 1 TeV. Thirty years of development leads to the the technical design report in 2013 [19]. A layout of the collider complex is shown in figure 3.1.

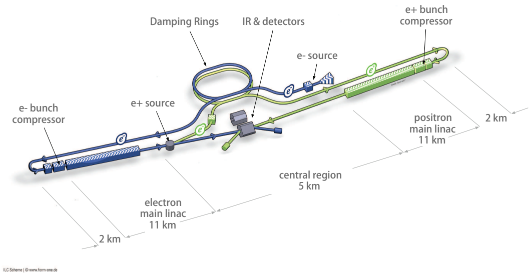


Figure 3.1: A layout of the International Linear Collider complex, taken from [19].

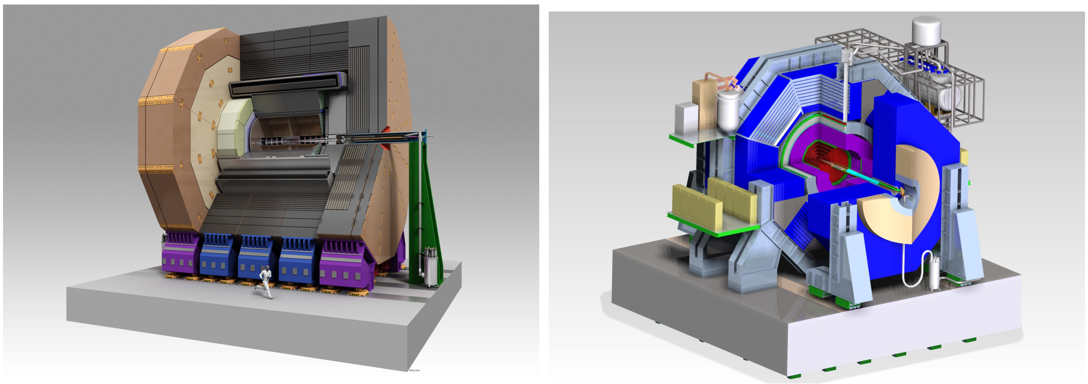
3.1.1 ILD vs SID

Two detectors concepts have been designed for the ILC to deliver the physics program. Precision tests for Standard Model requires an excellent jet energy resolution and di-jet mass reconstruction. Particle Flow Algorithms based event reconstruction meets the requirement and motivates the detector designs. For the best performance of the PFA, high granular calorimeter systems and highly efficient tracking systems are designed. The requirement to separate W and Z bosons in di-jet final states requires a jet energy resolution below 3.5%. The momentum resolution of 5×10^{-5} GeV is motivated by the Higgs boson recoil reconstruction in the Higgs-strahlung.

Motivation for two detectors is to have multiple independent measurements within one collider for cross-checking, complementary measurements and competition between collaborations. Two detectors are both general purpose detectors. Silicon Detector, SiD, is a compact detector with a large magnetic filed of 5 T. It uses silicon tracking modules. The International Large Detector, ILD, is a larger detector with a time projection chamber as the main tracking unit. Both detectors have high granular calorimeters optimised for the particle flow. A view of both detector concepts can be seen in figure 3.2

3.2 CLIC

The other potential next generation electron-positron linear collider, the Compact Linear Collider (CLIC), has a higher reach of the centre-of-mass energy up to 3 TeV. Although the accelerating systems would be different, the calorimeter designs are similar, both highly granular for the optimal particle flow. A layout of the CLIC complex is shown in figure 3.3



(a) (b)

Figure 3.2: Figure 3.2a and figure 3.2b show the view of the International Large Detector and the Silicon Detector for the International Linear Collider, taken from [19].

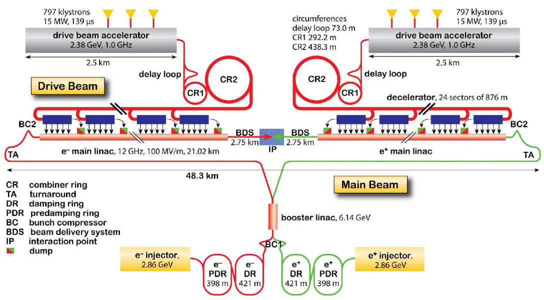


Figure 3.3: A layout of the Compact Linear Collider at 3 TeV, taken from [20].

3.3 Physics at future linear colliders

The physics program for the CLIC and the ILC, which is a driving force for the detector design, have some common goals. ILC has a reach of centre-of-mass from 200 GeV to 1 TeV, whilst CLIC can reach from 350 GeV to 3 TeV. Both machines are capable of precision higgs coupling measurements, top mass and coupling measurements, search for new physics such as supersymmertry particles. The ILC can also operate at low energy to be a Z and a H factory for ultra precision Z mass and H mass measurement. CLIC, however, has the advantage of higher energy reach, which allows measurements of rare events, such as higgs triple self-couplings and quartic couplings. The discovery potential for the CLIC is also greater.

3.4 Physics requirements on the detector design

The physics goal of jet energy resolution at the ILC and the CLIC is to separate W and Z hadronic decays via reconstruction their di-jet masses [18, 19]. This translate to a requirement of 3.5-5% of the energy resolution. This level of precision is unlikely to be achieved with a traditional calorimetry design. On the contrary, particle flow approach has demonstrated it ability to reach the goal [21, 22].

In a typical jet, using measurements on the particle composition from the LEP [23, 24], about 62% of the jet energy is from charged particles, 27% from photons, 10% from long-lived neutral hadrons, and 1.5% from neutrinos. In a traditional approach to calorimetry, about 72% jet energy is measured in the electromagnetic (ECAL) and the hadronic (HCAL) calorimeters combines. The jet energy resolution is thus limited by the energy resolution of the hadronic calorimeters, which typically is $\gtrsim 55\%/\sqrt{\mathsf{E(GeV)}}$.

The particle flow approach to calorimetry improves the jet energy resolution by fully reconstructing the four momenta of all visible particles in the detector. The jet energy is the sum of the individual particles' energies, where the energy of the charge particles are measured in the tracking detectors, and the energy of neutral particles are measured in calorimeters. In this manner, the hadronic calorimeter only measures about 10% of the energy, which would greatly improves the overall energy measurement.

The particle flow calorimetry requires to fully reconstruct particles and associate calorimeter hits to tracks. This is demanding for the software design and the detector design. The software details of the PandoraPFA, which is a successful particle flow implementation, are described in section 4.4. The detector needs to be highly granular for the excellent spatial resolution to be able to correctly associate calorimeter hits to the inner detector tracks. This forms the main motivation in the detector design and it is illustrated with a brief description of the particle flow principle.

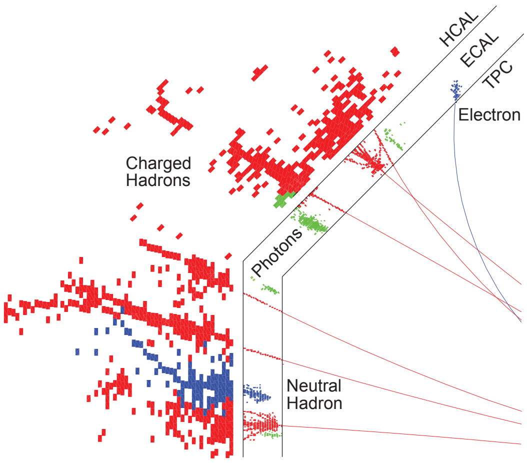


Figure 3.4: A typical topology of a 250 GeV jet, simulated with CLIC_ILD detector concept, taken from [22].

Figure 3.4 is a typical topology of a 250 GeV jet, simulated with CLIC_ILD detector concept. The particles with the calorimeter hits and tracks are labelled with different colours. Clusters of calorimeter hits in the highly granular ECAL and HCAL are associated with tracks from the inner tracking detector, TPC. Photons are identified using the characteristics longitudinal and transverse electromagnetic shower profiles. Hadronic showers are separated from electromagnetic showers due to the small transverse spread of the electromagnetic shower. Therefore, the inner tracking detector should be highly efficient and has very little material. For the calorimeter, the ECAL and HCAL, both should be highly granular. The material of the calorimeter should be dense and has a large ratio of interaction length to radiation length.

3.5 CLIC vs ILC

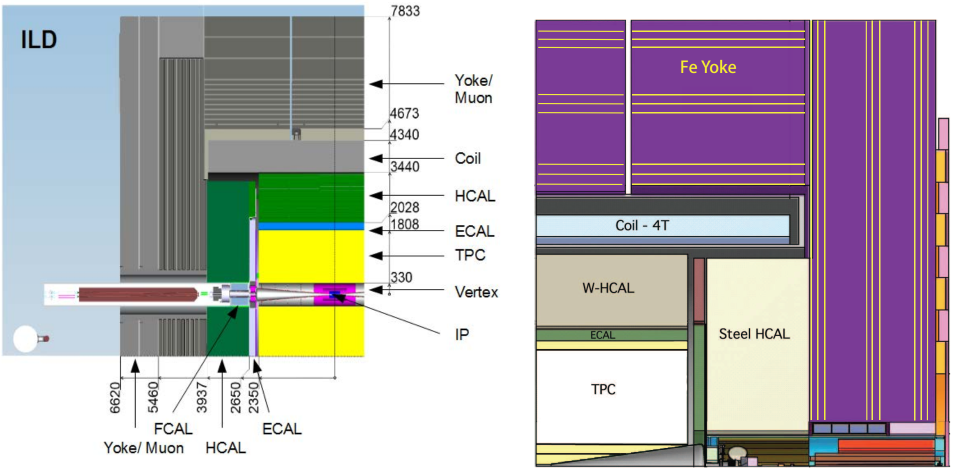
Due to the similarities of the two two linear collider programs, the development with CLIC detector concepts start with ILC detector concepts. CLIC_ILD and CLIC_SiD are developed based on ILD and SiD. The two main differences are the high centre-of-mass energy and the 0.5 ns between bunch crossings at CLIC. More incoherent pairs and more hadronic two-photon events are produced at high energy. Particles produced in the forward region via t-channel are more important due to a stronger boost. These differences leads to a modification in the detector design and the reconstruction software for the CLIC. A comparison of CLIC_ILD and ILD longitudinal cross sections can be seen in figure 3.5. A comparison of key parameters of the ILD and CLIC_ILD detector concepts is shown in table 3.1.

Concept	ILD	CLIC_ILD
Tracker	TPC/Silicon	TPC/Silicon
Solenoid Field (T)	3.5	4
Solenoid Field Bore (m)	3.3	3.4
Solenoid Length (m)	8.0	8.3
VTX Inner Radius (mm)	16	31
ECAL r_{min} (m)	1.8	1.8
ECAL Δr (mm)	172	172
HCAL Absorber B / E	Fe	Fe / W
HCAL Interaction Length	5.5	7.5
Overall Height (m)	14.0	14.0
Overall Length (m)	13.2	12.8

Table 3.1: A comparison of key parameters of the ILD and CLIC_ILD detector concepts. ECAL r_{min} is the smallest distance from the calorimeter to the main detector axis. HCAL Absorber B / E indicates the absorber material for the barrel (B) and the endcap (E). The table is adapted from [18].

3.6 International Large Detector

The International Large Detector, ILD, is a detector concept at the International Linear Collider, ILC. The ILD detector concept has been optimised in the view of the particle flow techniques. Particle flow approach to event reconstruction has shown to deliver the best possible jet reconstruction with proof-of-principle implementation such as PandoraPFAchapter 5. Each individual particles are reconstructed with the particle flow approach. For charged particles, calorimeter hits are associated with the tracks. The measurement of charged particle relies on the excellent tracking system resolution. Neutral particle reconstruction require fine spatial resolution of the calorimeters. These form the requirements for the detector designs and optimisations.



(a) (b)

Figure 3.5: Figure 3.5a and Figure 3.5a shows longitudinal cross section of top quadrant of the ILD and the CLIC_ILD detector concepts, taken from [19] and [18] respectively. From interaction point (IP) outwards, there is a tracking system compromising a large time projection chamber (TPC) augmented with silicon tungsten layer, highly granular electromagnetic calorimeters (ECAL) and hadronic calorimeters (HCAL), muon chambers, forward calorimeters (FCAL), magnetic coils and iron yokes. Numbers are in units of mm.

The particle flow paradigm requires topological information for individual particle reconstruction. The sub-detector systems need to have the spatial resolution to separate charged particles from neutral particles. The result is a highly granular calorimeters with a central tracking system with excellent momentum resolution. Longitudinal cross section of top quadrant of the ILD detector concept, taken from [?], is shown in figure 3.5a From interaction point (IP) outwards, there is a tracking system compromising a large time projection chamber (TPC) augmented with silicon tungsten layer, highly granular electromagnetic calorimeters (ECAL) and hadronic calorimeters (HCAL), muon chambers, forward calorimeters (FCAL), magnetic coils and iron yokes. Numbers are in units of mm.

This section will describe the sub-systems of the ILD detector concept in the ILD technical design report [], the ILD_o1_v05 option in Mokka simulation. This detector concept has been used in studies described in subsequent chapters. The ILD detector concept has been optimised and documented in previous documents, such as the letter of intent []. The CLIC_ILD detector concept for the CLIC in the conceptual design report [18] is a modified version of the ILD, adapted to the CLIC colliding environment.

The differences between ILD and CLIC_ILD can be seen in figure 3.5 and table 3.1, and are addressed in the discussion below.

3.7 Overview of ILD sub-detectors

The ILD detector concept is designed as a general purpose detector. Closest to the interaction points are the precision a vertex detector and a tracking system. The tracking system consists of silicon tracking with a time projection chamber. Surrounding the tracking system is a high granular calorimeter system. The outer solenoid provides a magnetic field of 3.5 T. The most outer iron return yoke acts as a muon calorimeter.

3.7.1 Vertex Detector

The pixel-vertex detector(VTX) needs to be close to the interaction point to reconstruct secondary vertex. As the TPC is the main tracking detector, the VTX mainly measures the impact parameter of tracks. The structure is three double layers with a barrel geometry. Double layer lowers the material budget and improves the impact parameter measurements. The first double layer is half length of the other two to avoid the high occupancy region of direct low omentulum hits from the incoherent pair background.

For the CLIC, the same structure is used. The first layer is moved outwards due to a larger high occupancy region with higher centre-of-mass energy. The detector is also required to provided time stamping at nanoseconds level.

3.7.2 Tracking Detectors

The hybrid tracking system is consists of a large volume time projection chamber (TPC), a Silicon Inner Tracker (SIT), a Silicon External Tracker (SET) in the barrel region, a end cap tracking component (ETD) behind the endplate of the TPC, and a silicon forward tracker (ETD) in the forward region. The SIT, SET, and ETD are made up two single-sided strip layers tilted by a small angle. The ETD is a system of two silicon-pixel disks and five silicon-strip disks. The silicon envelope tracking system and the TPC are shown in figure 3.6.

The main part of the tracking system, the TPC, can measure a large number of three dimensional spatial points. Continuous tracking allows precise reconstruction of non-pointing tracks. The TPC is optimised for point resolution and minimum material, as required for the best calorimeter and particle flow performance.

The barrel silicon trackers improve the the overall momentum resolution. They provide additional high precision space points and additional redundancy between the TPC, the VTX, and the calorimeters. The ETD provide the low angle coverage which is not covered by the TPC.

For the CLIC_ILD, the hybrid structure is used. The outer silicon tracking system is more important at the CLIC to achieve a high momentum resolution at high centre-of-mass energy, as it is challenging using a TPC to sperate two tracks in high energy jets and to identify events in the collection of 312 bunch crossings in 156 ns.

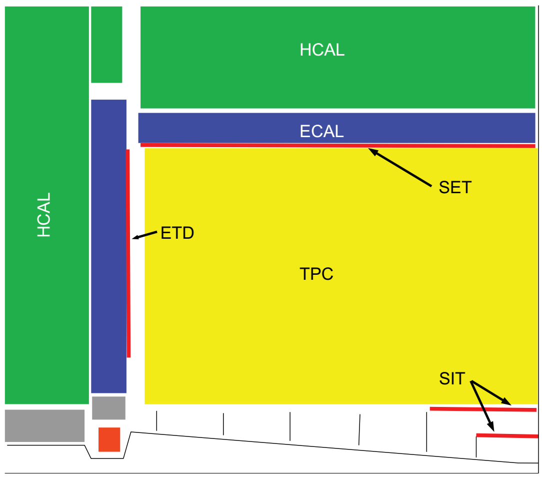


Figure 3.6: A top quadrant view of the ILD silicon envelope system, SIT, SET, ETD, and ETD as included in MOKKA full simulation, adapted from the figure in [19].

3.7.3 Electromagnetic Calorimeter

The Silicon-Tungsten sampling electromagnetic calorimeters in the ILD consist of a nearly cylindrical barrel and two end cap systems, optimised for particle flow. The ECAL measures photon energies and separates photons from other particles. The fine granular ECAL also sits inside the HCAL, which hosts the first part of the hadronic showers and greatly helps to separate hadronic showers.

The particle flow paradigm has a large impact on the ECAL design with many requirements. In addition for the ECAL to measure and separate photons, it also needs to reconstruct detail shower profiles to separate electromagnetic showers from hadronic showers, as approximately 50% of hadronic showers starts in the ECAL. These requirements can be fulfilled with an excellent three dimensional granular ECAL.

From test beam data and simulation studies, a sampling calorimeter with longitudinal and transverse segmentation below one Molière radius and below one radiation length at the front the calorimeter is needed. The most compact design is realised with Tungsten as absorber material and silicon pad diodes as active material. A cross section of the ECAL is shown in figure 3.7. Tungsten is a dense material with a large ratio of interaction length to radiation length. This helps to separate electromagnetic showers from hadronic showers by making electromagnetic showers transversely narrow. Silicon pad size of 5.1 by 5.1 mm cover large areas. They are simple and reliable to operate. The choice of thin silicon layers offers a great spatial resolution at a cost of the energy resolution in favour of the particle flow.

The longitudinal segregation is a compromise between the cost and the performance. The total 30 layers, which is about 20 cm, provides about 24 radiation lengths. The first 20 layers use 2.1 mm thick absorber plates, which is twice finer sampling than the last 10 layers with 4.2mmthick absorber plates. The test beam data with electron shows the energy resolution of the ECAL concept to be $16.6/\sqrt{E(\text{ GeV})} \oplus 1.1\%$, which is compatible with the values assumed for the full ILD detector simulation.

For the CLIC_ILD, the same ECAL from the ILD is assumed, as the requirements of a CLIC detector are satisfied. The increased centre-of-mass energy results in extra energy leakage. But only a small fraction of particles are affected and the leakage is controlled by the HCAL.

3.7.4 Hadronic Calorimeter

The requirements of sampling hadronic calorimeter is, again, driven by the need of the particle flow. The need of three dimensional granularity in transverse and logistically direction is satisfied by a sampling calorimeter.

The principle role of the HCAL is to separate neutral hadron showers from other particles, and to measure neutral hadron energies. The neutral hadron contribution of the jet energy is around 10% on average. A moderate fine granular HCAL is a good

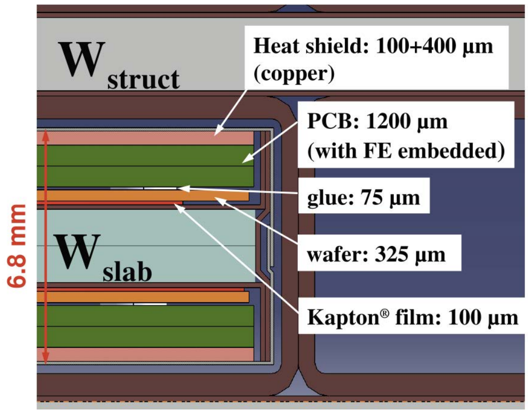


Figure 3.7: A cross section through electromagnetic calorimeter layers, taken from [19].

balance between cost and performance. The chosen layout is 48 longitudinal layers with 3 by 3cmscintillator tiles, using an analogue read out system. The layout of a technological prototype, the "EUDET prototype" is shown in figure 3.8a [25].

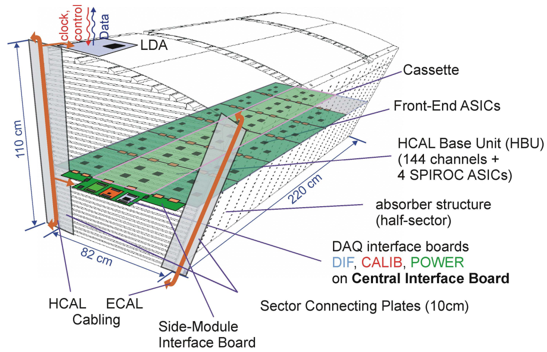
The longitudinal system provide about 6 radiation lengths including the ECAL, which is sufficient to contain the hadronic showers. The transverse cell sizes has been optimised for the best jet energy resolution. It is found that no substantial gain below 3 cm and performance degradation above 3 cm. Hence 3 cm cell size is chosen for the HCAL. The jet energy resolution as a function of HCAL scintillator cell size with different jet energies is shown in figure 3.8b.

The absorber material, stainless steel is chosen for mechanical and calorimetric reasons. Steel allows a self-supporting structure without auxiliary supports. Also steel has a moderate ratio of interaction length to radiation length.

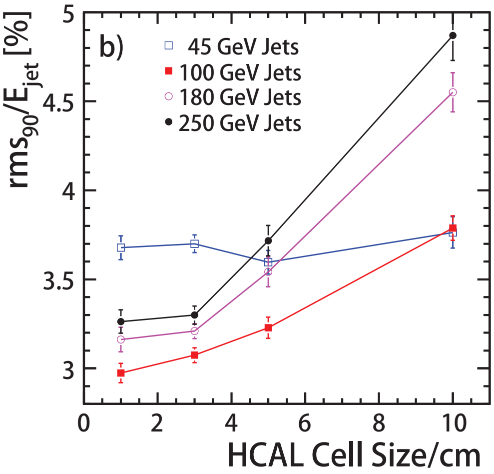
For the CLIC_ILD, extra layers are added to contain the hadronic shower at high energy. The increased thickeners is justified by the simulation studies, where the jet energy resolution degrades quickly for a thinner HCAL.

3.7.5 Solenoid

A large superconducting solenoid outside the calorimeters produces a nominal 3.5 T magnetic field. For the CLIC_ILD, the magnetic filed is increased to 4 T for a better performance at a high centre-of-mass energy.



(a)



(b)

Figure 3.8: Figure 3.8a shows the schematic view of a CALICE AHCAL technological prototype module. Figure 3.8b shows the jet energy resolution as a function of the hadronic calorimeter scintillator cell sizes, with different energies. Both figures are taken from [19].

3.7.6 Yoke and Muon system

An iron yoke instrumented with scintillator strips active layers returns the magnetic flux, and acts as a muon detector and tail catcher calorimeter at the same time. The layout is shown in figure 3.9. The agreed maximum magnetic field at 15 m radial distance from the detector is 50 Gauss to ensure safety [26]. A highly efficient muon detector is provided by the 3 by 3 cm scintillator strips. As a tail catcher calorimetry, the first layer of the muon detector, catches the energy leakage from the HCAL and the ECAL. It has been shown a 10% improvement of single particle energy resolution is possible with the tail catcher [27].

For the CLIC_ILD, due to the different magnetic field strength, the iron yoke thickness differs to that of the ILD.

3.7.7 Very Forward Calorimeters

The forward region detectors provide luminosity measurements and forward coverage of calorimeters. A system of precision and radiation resistent calorimeters are required. The luminosity calorimeter counts Bhabha scattering to measure the luminosity to precision

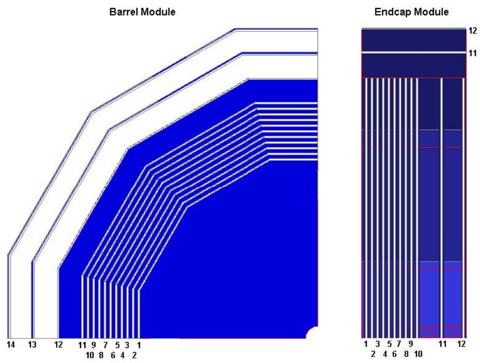


Figure 3.9: Sensitive Layers of the ILD muon system, taken from [19].

of 10^{-3} at 500 GeV centre-of-mass energy. The beam calorimeter (BeamCAL) extend the forward coverage, which are hit by many beamstrahlung pairs after each bunch corssing. BeamCAL estimates a bunch-by-bunch luminosity. An additional hadron calorimeter, LHCAL, at the forward region extends the angular coverage of the HCAL to that of the LumiCAL. Electron tagging is possible with the very forward calorimeters [28], which aids event reconstruction at high centre-of-mass energy.

The CLIC_ILD adopted a similar very forward calorimetry system as that of the ILD. Modifications to the design due to CLIC 3 TeV centre-mass-of energy can be found in [18].

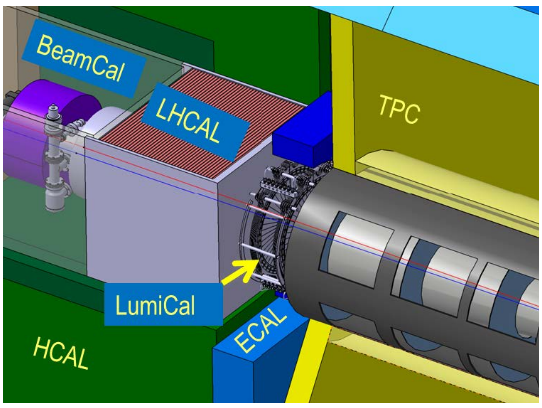


Figure 3.10: The forward calorimeters of the ILD, taken from [19]. LumiCal, BeamCal, and LHCAL are the luminosity calorimeter, beam calorimeter, and forward hadronic calorimeter.

Chapter 4

Simulation and Reconstruction

"How to open a pandora box?"
— A wise Chinese

In previous chapters, overviews of the theory and the future linear collider experiments have been described. Since the work presented in this document is for future collider, simulation and monte carlo method is used throughout the document. In this chapter, the simulation and event reconstruction chain are discussed, with emphasis on the PandoraPFA event reconstrcution, which provides the background for the photon reconstruction algorithms in chapter 5.

Simulation and reconstruction of events for the future Linear Colliders, ILC and CLIC, share common software framework. The noticeable difference will be discussed.

4.1 Monte Carlo event generation

For the simulated study, Monte Carlo (MC) event generation is the first step. Most events, electron-positron interaction, are generated with WHIZARD software, [29, 30], with no polarisation of the electron and positrons. Some simple events are generated with HEPEVT. PYTHIA [31] is used to describe parton showering, hadronisation and fragmentation. The parameters for PYTHIA are tuned to OPAL data from the LEP [32]. TAUOLA [33] debrides the tau lepton decay with correct spin correlations of the day products. The Initial State Radiation (ISR) is simulated in WHIZARD with the ISR photons being collinear with the beam direction. The Final State Radiation (FSR) is simulated with default parameters in PYTHIA.

For the CLIC simulated samples, the luminosity spectrum, which is generated with GUINEAPIG [34], is simulated in WHIZARD. The $\gamma\gamma \rightarrow \mathtt{hadrons}$ background events are hadronised with PYTHIA, and superimposed on the physics process simulations to save computational resources.

4.2 Event Simulation

The event simulation software is GEANT4 [35], and the detector geometry description is provided by MOKKA [36]. QGSP_BERT physics list is used to describe the hadronic showers decay in the detector.

4.3 Event Reconstruction

Reconstruction software runs in Marlin framework [37], as a part of the iLCSoft. Event reconstruction contains following steps: digitisation of simulated calorimeter hits, reconstruction of tracks in the tracking system using pattern recognition algorithms, and particle flow objects (PFOs)reconstruction with PandoraPFA [21, 22]. Reconstruction does not include calorimeters hits in the forward calorimeters, due to computational reasons.

For the CLIC detector simulations, suppression of $\gamma\gamma \rightarrow \mathtt{hadrons}$ background is performed.

Details of the reconstruction can be found in [17, 18]. Particle flow reconstruction via PandoraPFA will be discussed in details, which provides the background for the photon reconstructions in PandoraPFA in chapter 5.

4.4 PandoraPFA

Tradition calorimetric approach is unable to meet the mass and energy requirements for the future linear collider. The particle flow approach with PandoraPFA has a proof-of-principle demonstration of its capability to reach required resolution. The particle flow approach also put stringent requirements on the detector design, which is described in chapter 3. By associating calorimeter hits to the tracks, around 60% of the jet energy from charged particles is measured by the tracker, which has a much better resolution than the calorimeter. Small cell sizes of the calorimeters are required to identify hits from different particles. The traditional sum of calorimeter cell energies is replaced by particle flow reconstruction algorithms, a complex pattern recognition problem. The PandoraPFA algorithm has been developed and used in the ILC and CLIC simulation studies.

Started with the ILD detector concept, PandoraPFA has been adapted to the CLIC condition and shows its ability to deliver required energy resolutions [18]. Recent the code base of the PandoraPFA has been restructured. The core base codes for basic object and memory managements are factorised in the Pandora C++ Software Development Kit [38]. There are over 60 linear collider specific reconstruction algorithms, each aims to address a particular topological in the reconstruction.

In the subsequent paragraphs, the main steps in the PandoraPFA reconstructions are described. The details of the reconstruction can be found in citeThomson:2009rp,Marshall:2012ry,Marshall

Inputs of PandoraPFA are digitised calorimeter hits and reconstructed tracks. The output are reconstructed particles with four-momenta, Particle Flow Objects (PFOs).

4.4.1 Track selection

Tracks from the tracking system are selected based on their topological properties, how likely they are from physical processes, and whether they are consistent with the tracker resolution. Tracks passed the selection are used for the subsequent reconstruction.

Using a helical fit of last 50 reconstructed hits, tracks are projected to the front of the ECAL. Tracks are also classified based on their likely origin. The information are stored and passed on to the subsequent reconstruction.

4.4.2 Calorimeter selection

The input information of a calorimeter hit is the position, the layer in the calorimeter and the energy response from the calorimeter.

Calorimeter hits are selected based on a series of criterion. The selected hits need to have energies above the threshold, using the conversion of a minimum ionising particle (MIP) equivalent, and using directly the converted energy. Similar to tracks, only calorimeter passed the selection are used in later steps.

Geometry information and likelihood of the hit originated from a minimum ionising particle (MIP) are calculated.

Isolated hits, often originated from low energy neutrons in a hadronic shower, are difficult to associate to the correct hadronic shower. They are identified and not used in the clustering. But their energy is added in the very last particle flow object (PFO) creation step.

4.4.3 Cone Clusters Algorithm

Before discussing the rest of the PandoraPFA reconstruction, it is necessary to introduce the cone based clustering algorithm, which is widely used in the calorimeter in PandoraPFA. The clustering algorithm produces basic working objects, Clusters.

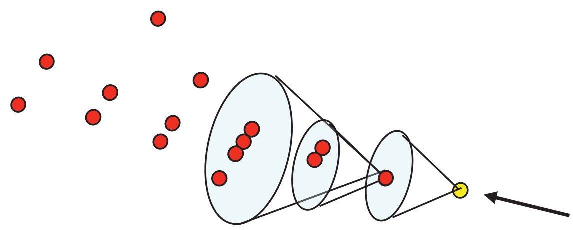


Figure 4.1: Illustration of the cone clustering algorithm, taken from [39]

There are two main types of clustering algorithms: cone based and sequential combination (see section **??**). The main clustering scheme PandoraPFA is cone clustering, for grouping calorimeter hits. Illustrated in figure 4.1, cone clustering has a specified opening angle of the seed hit. Because the direction of particle flows is largely unchanged from the originated particle, whether it is a electromagnetic shower, QCD radiation or hadronisation, these cone clusters have similar direction and energy to the originated particle. Therefore it is applicable to use cone based clustering algorithms for building clusters.

The seed for the cone clustering is typically the projection of a energetic track to the front of the ECAL. A high energy calorimeter hit can also be used as a seed. A cone with a specified opening angle and depth will be formed around the seed. The four-momentum of calorimeter hits sum to the cone's four-momentum.

4.4.4 Particle Identification

Dedicated particle identification algorithms aim to identify muons and photons before associating calorimeter hits to tracks. The details of the photon reconstruction algorithms and photon related algorithms are described in chapter 5. By removing the hits from muons and photons, the reconstruction of charged particles is improved as it reduce the pattern recognition problem. Identified muons and photons do not participate in the clustering and re-clustering stages, but re-entre the construction at the fragment removal stage (see section 4.4.9).

4.4.5 Clustering

The cone clustering algorithm described in section 4.4.3 is used to group calorimeter hits from innermost to outmost pseduo-layer. The output Clusters are further processed, merged or split based on their topological properties.

4.4.6 Topological cluster association

Initial clustering scheme is aggressive at splitting clusters. Small clusters are merged based on clear topological signatures. These merging signatures include combining track segments, connecting tack segments with gaps, connecting track segment to a hadronic shower, and merging clusters when they are within close proximity.

4.4.7 Track-cluster association

Clusters are associated to tracks, according to the proximity of the first layer of the cluster and the track projection to the front of the ECAL.

4.4.8 Re-clustering

The cluster association scheme work well for low energy (less than 50 GeV) jet. For a high energy jet, particles and the subsequent hadronic showers are more boosted and more likely to overlap each other. Therefore, it is important to re-cluster based on the compatibility of the cluster energy and the associated track momentum. A cluster may be split into two. Two clusters maybe be re-clustered based on the track-cluster association. The re-clustering algorithm is applied iteratively to find a more correct clustering of calorimeter hits.

4.4.9 Fragment removal

The late stage of the reconstruction will focus on merging low energy clusters, especially non-photon neutral clusters. These neutral clusters are likely to be fragments of charged clusters, instead of being a physical particle. The merging criterion are mostly based on the proximity and the energy comparison.

One algorithm will attempt to split up photon clusters, where each is originated from two close by photons. Photon related algorithms are described in details in chapter 5.

4.4.10 Particle Flow Object Creation

Particle Flow Objects (PFOs) are created at the last step. Tracks are associated to the clusters based on the proximity. Simple but effective particle identification for electrons, muons are applied. Photon identifications have been applied at various stages of the reconstruction.

PFOs are the output of the PandoraPFA reconstruction. The four-momentum of these PFOs are used heavily for the downstream analysis. The electron, muon and photon identification are also used in physics analysis, such as one described in chapter 6.

4.5 Suppression of $\gamma\gamma \rightarrow \mathtt{hadrons}$ backgrounds

For the CLIC, significant $\gamma\gamma \rightarrow \mathtt{hadrons}$ background is present. It is crucial to remove the beam induced background as they don't represent the underlying physics process.

Two Marlin process has been developed to suppress these background, a track selector and a PFO selector [22].

The track selector aims to remove poor quality and fake tracks. It places simple quality cut and a simple time of arrival cut. If the arrival time of the track at the front of the ECAL, using the helical fit, differs more than 50 ns from using a straight line fit, the track will be rejected.

The PFO selector utilise the high spatial resolution from the high granular calorimeter. PFOs from $\gamma\gamma \to \text{hadrons}$ often have low p_T and have a range of time. PFOs from physics processes have a range of p_T, and have time close to the brunch crossing time. These two distinctive features allow $\gamma\gamma \to \text{hadrons}$ background to be separated. The optimal suppression uses different p_T and time cuts for the central part of the detector, and for the forward part of the detector, and uses different cuts for photons, neutral PFOs and charged PFOs. Three configurations of these cuts are developed, namely "loose", "normal", and "tight" selections. As the name suggested, "loose" selection corresponds to a looser cut of p_T and time. The optimal configuration depends on the $\sqrt{s}$ of the collision, and the physics process to study.

The background suppression is used in analysis described in chapter 7

Chapter 5

Photon Reconstruction in PandoraPFA

"Photons have mass? I didnąŕt even know they were Catholic."
— Woody Allen

Photon reconstruction is an important part of particle reconstruction. A good photon reconstruction provides a good single photon completeness and purity, as well as a good photon separation resolution. Such a good photon reconstruction is crucial for reconstructing heavy particles, for many physics processes involving these particles decaying into photons, such as τ lepton and π^0.

5.1 Overview of photon reconstruction in PandoraPFA

PandoraPFA provides a framework for particle reconstruction [], as described in chapter 5. In the linear collider content, it has a vast library of algorithms developed through years by many people. Each algorithm addresses one topological issue in the particle reconstruction []. The essential part of the PandoraPFA is track-cluster association and reclustering to find the best track-cluster pair. Algorithms that removes trackless clusters, such as removing muon clusters or photon clusters, would provide a clean environment for the track-cluster association, hence improving the jet energy resolution.

Photon identification in the PandoraPFA has two main mechanisms. The basic mechanism (see section 4.4.10) performs photon identification after track-cluster association and the reclustering processes. The second more sophisticated photon identification, the photon reconstruction algorithm, is performed before the track-cluster association and reclustering process. This algorithm identifies photon electromagnetic shower cores carefully in the dense jet environment.

The photon reconstruction algorithm in PandoraPFA version 1 improves jet energy resolution by correctly identifying photon electromagnetic shower cores and leaving a cleaner environment for the track-cluster association. However, the peripheral calorimeter hits to the shower cores may be left as fragments, and reconstructed as separate particles. This lowers the reconstructed photon completeness and makes the number of reconstructed photons a less useful physical quantity. Also, the algorithm in PandoraPFA version 1 leaves rooms for improvement of photon separation resolution.

This section presents a solution to the photon fragments issue. The newly introduced PandoraPFA algorithms also improves the photon separation resolution. Algorithms related to photon reconstruction, fragmental removal and photon splitting, which are written or introduced by authors, will be discussed below.

5.2 Photon reconstruction algorithm

The photon reconstruction algorithm refers to the more sophisticated photon identification of the two main identification mechanisms, before the track-cluster association and reclustering process (see section 4.4.4). The algorithm has the following steps: coarsely forming photon clusters, reconstructing photon candidate, photon ID test, and optional fragment removals. Reconstructing photon candidate requires further explanation of the two dimensional peak finding algorithms in section 5.3. The photon ID test involves a multi dimensional likelihood classifier, which is described in section 5.4.

5.2.1 Form photon clusters

This step finds large potential photon clusters. All calorimeter hits in the ECAL, which are not used in previous algorithms, are grouped into clusters using a cone based clustering algorithm.To find neutral photon clusters, which do not deposit energies in the tracking system, the cone clustering algorithm is seeded with energetic hits. The parameters for the cone clustering are generous, allowing potentially two or three photons in one cluster.

5.2.2 Reconstruct photon candidates

The large photon clusters are split into smaller photon candidates, using two-dimensional shower profiles. The candidates close to a track projection are deemed as non-photons. Identifying photon candidates within a large photon cluster relies on the characteristic electromagnetic showers, in particular the transverse distribution. A energetic photon or electron hits the absorber layers of the ECAL, it initiates an electromagnetic shower, where electron pair production and bremsstrahlung produce more low-energy photons and electrons. The transverse distribution is characterised by a narrow cone, widening while the shower develops.

To view the transverse shower distribution, a two-dimensional energy deposition projection is constructed in the plane perpendicular to the direction of the cluster. figure 5.1 shows the energy deposition projection of two photons candidates. U and V axis are two arbitrary orthogonal axis in the transverse plane perpendicular to the direction of photons. Z axis shows the sum of the calorimeter hit energy in GeV. The bin size corresponds to the square ECAL cell size.

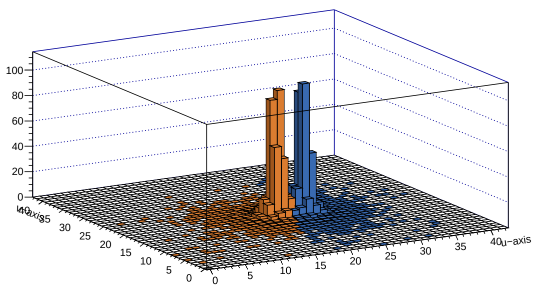


Figure 5.1: Two 500 GeV photons (yellow and blue), just resolved in the transverse plane perpendicular to the direction of the flight, of their energy deposition in electromagnetic calorimeter. U and V axis are two arbitrary axis perpendicular to each other in the plane. Z axis is the sum of the calorimeter hit energy in each particular bin in 2D plane in GeV.

By using the two-dimensional energy deposition projection, separating photons translates to separating peaks in the projection. Therefore a high performance two dimensional peak finding algorithm is the key to identify multiple photons. The peak finding algorithm will be discussed in section 5.3

The output of this step is a collection of photon candidates from a photon cluster, which will be fed to the photon ID test.

5.2.3 Photon ID test

Photon ID test decides if a a candidate is a photon. If a candidate is not a photon, the calorimeter hits of the candidate will be passed on to the next stage of the reconstruction. The photon ID test is a multidimensional likelihood classifier. The classifier is trained with discriminating variables, which exploit features electromagnetic showers. The classifier will be discussed in section **??**

5.2.4 Photon Fragment removal

The optional photon fragment removal aims to merge small photon fragment to main photons. Since this step shares the same logic as the algorithm in section **??**, only differing in the cut-off values for merging metrics, this step be discussed in section **??**.

This step marks the end of the photon reconstruction algorithm. The output are a collection of reconstructed photons, separated from non-photon calorimeter hits.

5.3 Two dimensional peak finding algorithm for photon candidate

As discussed in section 5.2.2, separating photon candidates from a cluster is same as identifying peaks in a two dimensional histogram. An example of two photons is shown in the figure 5.1. The basic algorithm treats all clusters as potential photon clusters. Since charged hadrons would deposit tracks in the tracking system, extra care is taken when a cluster is close to the projection of the track in the front of the ECAL. The basic peak finding algorithm has two main functions: identifying peaks, and assigning bins to peaks.

A two dimensional histogram is constructed using a plane orthogonal to the direction of the flight of the photon cluster. The U and V axis in figure 5.1 are two orthogonal axes in the plane. The width of the histogram is determined by the size of the ECAL square cell. The calorimeter hits of the photon cluster are then projected onto the two dimensional histogram. The height of the bin is the sum of the calorimeter hit energies in the bin.

A local peak is defined as a bin where its height is above all eight neighbouring bins. After all peak bins are found, non-peak bins are associated to one peak bin, by choosing the peak bin that minimise the metric

$$\frac{d}{\sqrt{E_{peak}}} \tag{5.1}$$

where d is the Euclidean distance between a non-peak bin and a peak bin on the histrogram, and E_{peak} is the height of the peak bin, which is the energy. Alternative metrics provided in the algorithm include d, $\frac{d}{E_{peak}}$, and $\frac{d}{E^2_{peak}}$. The default metric is chosen due to a good balance between distance and energy of the peak.

5.3.1 Candidate close to track projection

If a cluster or a photon candidate is close to the projection of the track in the front of the ECAL, it is likely that the cluster or the candidate is a charged hadron. Misidentifying a charged hadron as a photon leads to significant degradation in reconstruction performance. However, if a photon next to a charged hadron is carefully reconstructed, the overall reconstruction is improved. Hence this step aims to carefully identifies photon candidate next to charged hadrons, by using track information and features of the electromagnetic shower. Photon induced electromagnetic shower in the ECAL typically start in the first few layers. As the shower develops, the direction of the shower core does not change much.

If a peak bin is within the eight neighbouring bins of the track projection onto the two dimensional plane, the peak and its associated bins are flagged as non-photons. Furthermore, the ECAL is sliced longitudinally to help identify photon candidates. For example, the default three slices will result in three ECAL fiducial spaces, each contains space from the front of the ECAL to a third, two thirds and the back of the ECAL, respectively. The peaking finding algorithm is repeated for the same cluster divided in each ECAL fiducial space. The peak is only preserved as a photon candidate if the peak exists in every fiducial space, and if its position is shifted by no more than one neighbouring bin between fiducial spaces.

5.3.2 Peak filtering

The performance of the two dimensional peaking finding algorithm is improved by clever programming and physics arguments. For a given two dimensional histogram, such as the one in figure 5.1, major peaks most likely correspond to physical photons, while the minor peaks more likely come from fluctuations in energy deposition. To select major peaks, every time after non-peak bins are associated with peak bins, minor peaks with fewer than three bins associated (including the peak bin) are discarded. These bins are then associated with non-discarded peaks. The algorithm also allows bins with height below a critical value to not participate in the peak finding. The default value is set such that only empty bins are not used.

5.3.3 Inclusive mode

The two dimensional histogram is iterated a few times during the algorithm. The time complexity is $O(n^2)$ for a $n \times n$ histogram (Default $n = 41$). Therefore, for the purpose of speed, it is undesirable to have a very large histogram. However, since the histogram has a finite size, only energy deposition projected on the histogram would be considered for peak finding. This behaviour is suitable for photon reconstruction (section 5.2.2) and test for photon fragment removal (section 5.5). However, for photon splitting (section 5.7), there should be no calorimeter hits loss from splitting a photon. Hence inclusive mode of the peak finding algorithm is developed, and it allows energy deposition projected outside the histogram to be associated with identified peaks.

5.4 Likelihood classifier for photon ID

Section 5.2.3 outlines the photon ID test in the photon reconstruction algorithm. This section describes the multidimensional likelihood classifier in details, including discriminating variables. For each photon candidate, a set of kinematic variables are calculated. The classifier training typically uses simulated jet events.

5.4.1 Overview of Projective Likelihood

Projective likelihood model (PDE) is used in PandoraPFA for the photon ID due to its simplicity and low requirement on computing resources.

PDE implemented calculates the probability density for each discriminative variable, for signal and background. The overall signal and background likelihood are defined as products of the individual probability density. The likelihood ratio, R, is then defined as the signal likelihood over signal plus background likelihood.

To use the likelihood ratio, one way is to fit an underlying function to the probability density, which is implemented the TMVA software package. The other way is to use binned likelihood ratio, R, as the output, due to the simplicity. This is implemented in the PandoraPFA. Similarly to classifier like the rectangular cut method, PDE works better with decorrelated, gaussian like variables. The PandoraPFA implementation did not decorrelate nor transform the variables, to keep implementation fast.

5.4.2 Projective Likelihood in PandoraPFA

Kinematic variables to obtain the probability distribution exploit the differences between a characteristic electromagnetic shower and a hadronic shower, and the fact that a photon is more likely to be isolated from other showers and charged tracks. Two variables use the longitudinal shower distribution: the first ECAL layer of the shower, and the difference between a expected longitudinal distribution and the observed. Two variable uses the transverse shower distribution: in the transverse plane with two orthogonal axes, the r.m.s. distance of associated bins to the peak bin, and the smallest ratio of the two r.m.s. distances in each axis direction. One variable is the ratio between the photon candidate energy to the photon cluster energy. The last kinematic variable is the distance between the photon candidate and the closest track projection. The distributions of kinematic variables are normalised to probability distribution, stored in binned histograms.

TODO explain variables explain longitudinal and transverse

Furthermore, the classifier is improved by realising the kinematic variable distributions depend on the photon energy. Thus these distributions are divided by bins of photon candidate energy. The numbers of photon and non-photon candidates in each energy bin are also different, which helps the ID test. The default energy bins edges are 0.2, 0.5, 1, 1.5, 2.5, 5, 10, 20 GeV, which covers a good range of photon energies. Candidate with energy below 0.2 GeV would not be examined in this step, as it is very unlikely to be a photon.

For a given candidate, which falls in a energy bin, the likelihood classifier output is given by

$$\mathtt{pid} = \frac{N\prod P_i}{N\prod P_i + N'\prod P'_i} \tag{5.2}$$

where P_i and P'_i are the probability of i^{th} kinematic variable of photon and non-photon candidates. N and N' are the number photon and non-photon candidates. These are obtained during classifier training.

During classification, a candidate passes the photon id test if

$$\begin{cases} \mathtt{pid} > 0.6, & \text{if } 0.2 < E < 0.5\,\mathrm{GeV} \\ \mathtt{pid} > 0.4, & \text{if } E \geqslant 0.5\,\mathrm{GeV} \end{cases} \tag{5.3}$$

where E is the candidate energy. Two values of the $\mathtt{pid}$ cuts reflect the confidence of the id test with different candidate energy. The test is more cautious with low energy candidate.

5.5 Photon fragment removal algorithm in the ECAL

During the reconstruction, it is possible that a core of the photon electromagnetic shower is identified as a photon (the main photon). The outer part of the shower is reconstructed as a separate particle, and wrongly identified as a photon or a neural hadron (the photon/neutral fragment). figure 5.2 shows a typical creation of such a photon fragment. The fragment does not have the electromagnetic shower structure, and typically it is has much lower energy than the main photon. If a photon-fragment pair is merged, the pair should be consistent with a one-particle profile. These characteristics are used to merge fragments to main photons.

Photon fragment removal algorithms can exist in multiple step in the reconstruction, at the end of the photon reconstruction (see Section 5.2.4), or at the end of the reconstruction.

(a) (b) (c)

Figure 5.2: An event display of a typical 10 GeV photon (figure 5.2a), reconstructed into a main photon (figure 5.2b) and a photon fragment (figure 5.2c).

Since these algorithms share the same base class, the latter one will be discussed. The former differs mostly in the default cut-off values for merging metrics.

A photon and a potential fragment form a pair of particles (photon-fragment pair), if they are spatially close. Kinematic and topological properties of the photon-fragment pair are examined. The pair is merged when the properties pass a set of cuts, developed by comparing true photon-fragment pairs and non photon-fragment pair. This merging test is iterated over all possible photon-fragment pairs. If multiple photon-fragment pairs pass the merging test, the pair with closest distance metric, d, will be merged.

The photon-fragment pairs is classified into photon-photon-fragment pairs and photon-neutral-hadron-fragment pairs, because they have different kinematic and topological distributions. The pairs are further classified into low energy and high energy pair, depending on whether the fragment energy (E_p) above 1 GeV. The cuts for merging pairs, are classified which will be explained later, are listed in table 5.1.

Table 5.1 lists cuts for merging photon-photon-fragment pairs and photon-neutral-hadron-fragment pairs for both low energy and high energy fragments. d, d_c and d_h are the mean energy weighted intra-layer distance between each PFO in the pair, the distance between centroids, the minimum distance between calorimeter hits of each PFO in the pair, respectively. E_m and E_f are the main photon energy and the fragment energy. E_{p1} and E_{p2} are the two largest peaks and associated calorimeter hits, found by the two dimensional peak finding algorithm (section 5.3), ordered by descending energy, using the pair as input. N_{calo} is the number of the ECAL hits in the fragment. $|\cos(\theta_Z)|$ is the absolute cosine of the polar angle of the main photon, where beam direction is the z-axis.

Low E_f	Photon-photon	Photon-neutral-hadron
transverse shower comparison	$d < 30$, $\frac{E_{p1}}{E_m+E_f} > 0.9$, $\frac{E_{p2}}{E_f} < 0.5$, $E_{p1} > E_m$	-
close proximity	-	$d < 20$, $d_c < 40$
low energy fragment	$d < 20$, $E_p < 0.4$	-
small fragment 1	$d < 30$, $N_{calo} < 40$, $d_c < 50$	$d < 50$, $N_{calo} < 10$, $d_h < 50$
small fragment 2	$d < 50$, $N_{calo} < 20$	-
small fragment forward region	$N_{calo} < 40$, $d_c < 60$, $E_f < 0.6$, $\lvert\cos(\theta_Z)\rvert > 0.7$	-
relative low energy fragment	$d < 40$, $d_h < 20$, $\frac{E_f}{E_m} < 0.01$	$d < 40$, $d_h < 15$, $\frac{E_f}{E_m} < 0.01$
High E_f	**Photon-photon**	**Photon-neutral-hadron**
transverse shower comparison	$\frac{E_{p1}}{E_m+E_f} > 0.9$, $E_{p2} = 0$ or $(\frac{E_{p2}}{E_f} < 0.5$, $E_{p1} > E_m)$	$\frac{E_{p1}}{E_m+E_f} > 0.9$, $E_{p2} = 0$ or $(\frac{E_{p2}}{E_f} < 0.5$, $E_{p1} > E_m)$
relative low energy fragment 1	$d < 40$, $d_h < 20$, $\frac{E_f}{E_m} < 0.02$	$d < 40$, $d_h < 20$, $\frac{E_f}{E_m} < 0.02$
relative low energy fragment 2	-	$d < 40$, $d_h < 20$, $\frac{E_f}{E_m} < 0.1$, $E_f > 10$
relative low energy fragment 3	-	$d < 20$, $d_h < 20$, $\frac{E_f}{E_m} < 0.2$, $E_f > 10$

Table 5.1: The cuts for merging photon-photon-fragment pairs and photon-neutral-hadron-fragment pairs for both low energy and high energy fragments. d, d_c and d_h are the mean energy weighted intra-layer distance of the pair, the distance between centroids, the minimum distance between calorimeter hits of the pair. E_m and E_f are the main photon energy and the fragment energy. E_{p1} and E_{p2} are the two largest peaks, found by peak finding algorithm, ordered by descending energy. N_{calo} is the number of the calorimeter hits in the fragment. $\lvert\cos(\theta_Z)\rvert$ is the absolute cosine of the polar angle, where beam direction is the z-axis.

Three distance measurements have subtle difference. d_c gives the distance between centroids of each PFO in the pair, which is a quick but crude measurement. d_h is the minimum distance between calorimeter hits of each PFO in the pair. For a true photon-fragment, d_h should be close to zero as the pair should be spatially close. d is the mean energy weighted intra-layer distance between each PFO in the pair:

$$d = \frac{\sum_i^{layers} d_{l,i}\, E_{f,i}}{\sum_i^{layers} E_{f,i}} \tag{5.4}$$

where i indicates i^{th} pseudo-layer of the ECAL. $d_{l,i}$ is the minimum distance between calorimeter hits of the pair in the i^{th} pseudo-layer. $E_{f,i}$ is the energy of the fragment in the the i^{th} pseudo-layer. d is a better measurement of the closeness of the pair. Similar to d_h, d will be very small for a true photon-fragment pair.

One logic for merging is when the fragment is small with low energy and is close to the main photon. The other logic is when the pair looks like one photon in two-dimensional energy deposition projection (see section 5.2.2 and figure 5.1). Comparing low E_f and high E_f cut, the cuts are similar. High E_f cuts are more relaxed on the energy comparison for small fragment test. Comparing photon-photon-fragment pair and photon-neutral-hadron-pair, cuts for photon-neutral-hadron-pair are more conservative for low E_f, but more relaxed for high E_f. This reflects that the neutral hadron fragments originated from charged particles are more likely to be low energy.

Since all possible photon-fragment pairs are compared, this is a costly cooperation with $O(n^2)$ time complexity for n particles. The speed is improved by considering only the pairs with $d < 80$mm. The algorithm occurs at the end of the reconstruction.

5.6 High energy photon fragment recovery algorithm

Section 5.5 descried effective algorithms to removal photon fragments that are peripheral to the main photon, or the electromagnetic shower core. An example of such fragment is shown in figure 5.2. There is another type of fragment which is the leakage effect of the ECAL. When the high energy photon shower is not fully contained in the ECAL, shower deposits energy in the HCAL, which often forms a neutral hadron in the HCAL. Photon reconstruction, as described in section 5.2, considers only calorimeter hits in the ECAL.

An example of a 500 GeV photon reconstructed into a main photon in the ECAL (yellow) and a neutral hadron fragment in the HCAL (blue) is shown in figure 5.3. For the ILD detector, this ECAL leakage effect appears when the photon energy is above 50 GeV.

Figure 5.3: An event display of a typical 500 GeV photon, reconstructed into a main photon in the ECAL (yellow) and a neutral hadron fragment in the HCAL (blue).

With figure 5.3 as an example, high energy fragments in the HCAL is spatially close to the main photon. A fitted cone from the main photon covers most of the fragment if extended to the HCAL. These features allow a set of cuts developed to merge high energy fragments, listed in table 5.2

This algorithm would collect the photons and neutral hadrons in the HCAL as inputs. It occurs after the first pass of topological association in the reconstruction, which connects tracks to clusters in the ECAL and the HCAL. The algorithm would iterate over all pairs of reconstructed photons and neutral hadrons in the HCAL. For each pair, a set of variables are calculated and compared to a set of cuts (table 5.2). Photon-fragment pairs passing the cuts will be merged.

Fragment in the HCAL should be spatially close to the main photon, measured by three metrics. d_c^l is the distance between centroids of the last outer layer of the main photon and the first inner layer of the fragment. d_{cone}^l is the distance between fitted cones using the last outer layer of the main photon and the first inner layer of the fragment. d_{cone} is the distance between fitted cones using the main photon and the fragment.

The direction of the fragment should be similar to that of the main photon. r_f, the r.m.s. mean energy weighted distance of a calorimeter hit in the fragment to the direction of the main photon, has to be small for merging.

High energy fragment recovery	Cuts
distance comparison	$d_c^l \leqslant 173$ mm, $d_{cone}^l \leqslant 100$ mm, $d_{cone} \leqslant 100$ mm
shower width comparison	$0.3 \leqslant \frac{w_f^l}{w_m^l} \leqslant 5$
projection comparison	$r_f \leqslant 45$ mm
energy comparison	$\frac{E_f}{E_m} \leqslant 0.1$
cone comparison	$\%N_{calo,cone} \geqslant 0.5$

Table 5.2: The cuts for merging high energy photon fragment in the HCAL to the main photon in the ECAL. d_c^l is the distance between centroids of the last outer layer of the main photon and the first inner layer of the fragment. d_{cone}^l is the distance between fitted cones using the last outer layer of the main photon and the first inner layer of the fragment. d_{cone} is the distance between fitted cones using the main photon and the fragment. w_m^l and w_f^l are the r.m.s. width of the last outer layer of the main photon and the first inner layer of the fragment. r_f is the r.m.s. mean energy weighted distance of a calorimeter hit in the fragment to the direction of the main photon. E_m and E_f are the main photon energy and the fragment energy. $\%N_{calo,cone}$ is the fraction of the calorimeter hits in the fragment in the extended fitted cone of the main photon.

Another feature of the fragment and the main photon is that the shower width should be similar. w_m^l and w_f^l are the r.m.s. width of the last outer layer of the main photon and the first inner layer of the fragment. The ratio $\frac{w_f^l}{w_m^l}$ needs to be in the range of 0.3 to 5. The generous upper bound is due to the HCAL is coarser than the ECAL.

When a fitted cone from the main photon is extended to the HCAL, the cone should contain a significant amount of the fragment. $\%N_{calo,cone}$, the fraction of the calorimeter hits in the fragment in the extended fitted cone of the main photon, has to be no less than 0.5 for the merging.

The last criteria is the fragment should has low energy relative to the main photon. E_m and E_f are the main photon energy and the fragment energy. The ratio, $\frac{E_f}{E_m}$, has to be less than 0.1 for the merging.

If multiple photon-fragment pairs pass the cuts with the same fragment, the pair with highest $\%N_{calo,cone}$ will be merged.

5.7 Photon splitting algorithm

Algorithms described above deal with forming photons from calorimeter hits in the ECAL, merging photon fragments in the ECAL and the HCAL. Another aspect in photon reconstruction is splitting accidentally merged photons. During the particle reconstruction, it is possible that photons are accidentally merged if they are spatially close. Hence another algorithm at the end of the particle reconstruction addresses this issue and tries to split merged photons.

Merged photon is typically energetic. The merged photon should be consistent with topologies of a spatially closed photon pair. Extra care should be taken if the photon is close to a charged PFO. Many PandoraPFA algorithms deal with track clusters association and there is a greater confidence in clusters associated with tracks. These features form logics behind the algorithm.

Photon splitting	Cuts
Cuts	$E > E_{c1}$, $E_{p2} > E_{c2}$, $N_p < 5$
E_{c1} and E_{c2} values	
0 nearby charged PFO	$E_{c1} = 10$, $E_{c2} = 1$
1 nearby charged PFO	$E_{c1} = 10$, $E_{c2} = 5$
> 1 nearby charged PFO	$E_{c1} = 20$, $E_{c2} = 10$

Table 5.3: The cuts for splitting photons, and the values for energy cut-off points. E is the photon energy. E_{p2} is energy if the second largest peak from the two dimensional peak finding. N_p is the number of peaks identified by the peak finding. E_{c1} and E_{c2} are the energy cut-off values, determined by the number of nearby charged PFOs.

The table 5.3 shows values for the splitting a photon. E is the photon energy. E_{p2} is energy if the second largest peak from the two dimensional peak finding. N_p is the number of peaks identified by the peak finding. E_{c1} and E_{c2} are the energy cut-off values, determined by the number of nearby charged PFOs. When a energetic photon is identified, and a energetic second peak can be found by the peak finding, the photon is likely from a photon pair. N_p cut is because a reconstructed photon is unlikely from more than four photons. The values of E_{c1} and E_{c2} allow more conservative approach when a photon is close to charged PFOs.

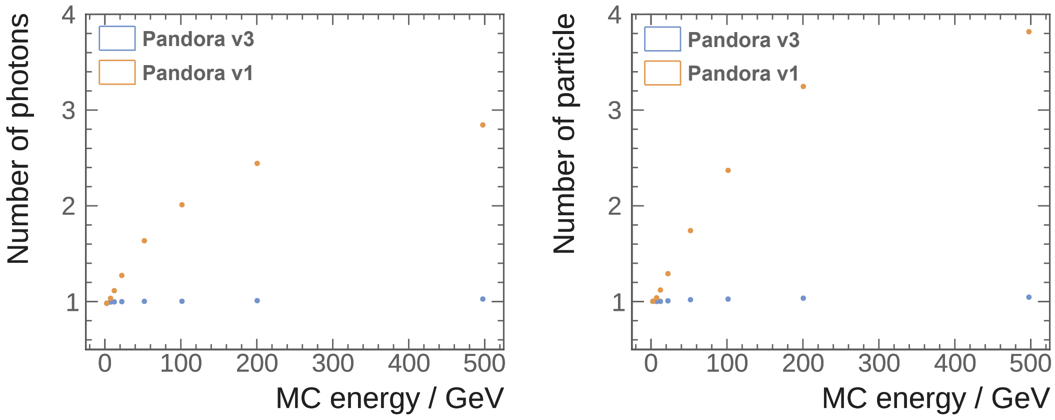
5.8 Photon reconstruction performance improvement

Motivations and implementations of four different algorithms for photon reconstruction, fragment removal and photon splitting have been described in the above. The main photon reconstruction algorithm in section 5.2 improves the photon completeness and the photon pair resolution, due to the improved two dimensional peak finding algorithm in section 5.3. The fragment removal algorithms in section 5.5 and section 5.6 further reduce the photon fragments in the ECAL and the HCAL. The photon splitting algorithm in section 5.7 exploits the peak finding algorithms and improves the photon separation resolution. Because of the high photon reconstruction completeness, the jet energy resolution receives a small improvement.

This section reviews the performance improvement with the introduced algorithms, using single photon, photon pair and jet samples. The performance was compared using PandoraPFA version 1 and version 3, where the photon algorithms were introduced in PandoraPFA version 2. The ILD detector model is used. The photon pair simulated events were generated with a uniform distribution in the solid angle for a range of the opening angles between the pair. The events selected such that there is no early photon conversion and the monte carlo photon deposits energies in the calorimeter. The events are further restricted to photon decaying in barrel and end cap region only, to minimise the detector effect.

figure 5.4a shows the reduction in fragments identified as photons, using a single photon per event sample. For the blue dots, the average number of photon stays below 1.05, where the true value is 1, even at high energy. A similar trends shows in figure **??**, where the extra fragments identified as neutral hadrons have taken into account. For a 100 GeV photon, the average numbers of photon and particle are reduced to 1 from 2 and 2.4. For a 500 GeV photon, the average numbers of photon and particle are reduced to 1.05 from 2.8 and 3.8.

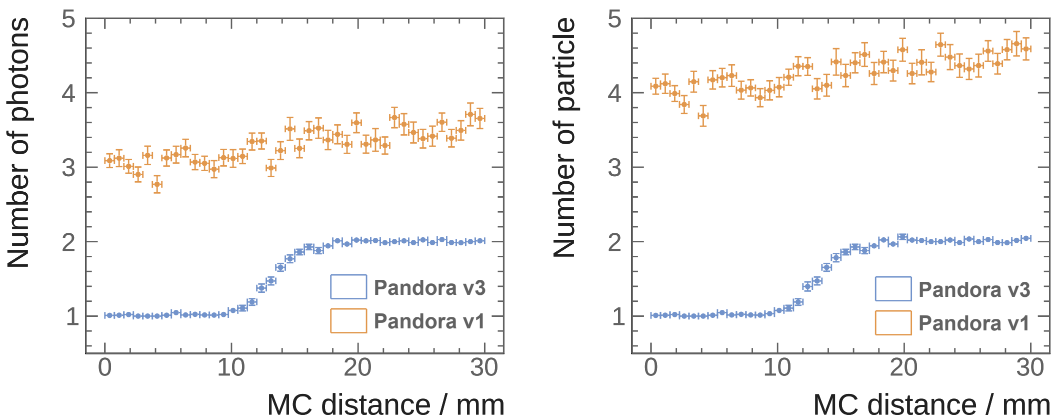
figure 5.5 illustrates a similar reduction in the photon fragments and the neutral hadron fragments using two photons of 500 and 50 GeV per event sample. The high energy photon are more likely to create fragments. And the imbalance in the two photon energies makes it more difficult to separate correctly. The figure shows the MC distance separation from 0 to 30 mm, which corresponds to approximately 6 ECAL square cell. In both figure 5.5a and figure 5.5b, the average numbers of photon and particle are below



(a) (b)

Figure 5.4: figure 5.4a and figure 5.4b shows the average number of reconstructed photons and reconstructed particles, as a function of their true energy using a single photon per event sample. The top orange and bottom blue dots are reconstructed with PandoraPFA version 1 and version 3. The photon reconstruction is changed in PandoraPFA version 2.

2.05 at 30 mm apart, which is significantly better than reconstruction in PandoraPFA version 1. Two photons start to be resolved at 10 mm apart, and fully resolved at 20 mm apart. The resolution is better than reconstruction in PandoraPFA version 1, which is difficult to extract due to excess fragments.



(a) (b)

Figure 5.5: figure 5.5a and figure 5.5b shows the average number of reconstructed photons and reconstructed particles, as a function of the MC distance separation in the calorimeter, using two photons of 500 and 50 GeV per event sample. The top orange and bottom blue dots are reconstructed with PandoraPFA version 1 and version 3. The photon reconstruction is changed in PandoraPFA version 2.

Another metric to reflect the improvement in photon reconstruction is the fragment energy fraction of the total energy as function of the distance separation. Shown in figure 5.6, using two photons of 500 and 50 GeV per event sample, a reduction in fragment energy can be seen clearly. With improved reconstruction, the average fragment energy fraction is below 0.1% up to 30 mm apart, whilst around 5% energy would be in fragments with reconstruction in PandoraPFA version 1.

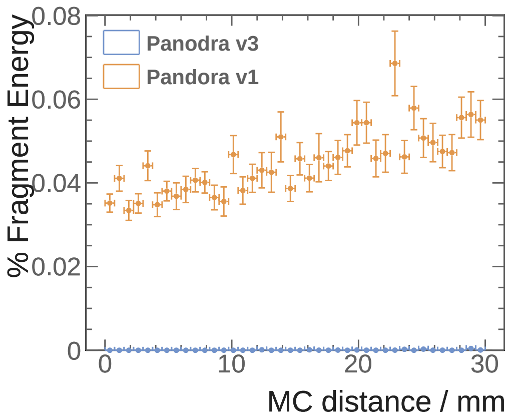


Figure 5.6: figure 5.6 shows the average fraction fragments energies of the total energy, as a function of the MC distance separation in the calorimeter, using two photons of 500 and 50 GeV per event sample. The top orange and bottom blue dots are reconstructed with PandoraPFA version 1 and version 3. The photon reconstruction is changed in PandoraPFA version 2.

The improvement in completeness and resolution in photon reconstruction, as shown in single photon and double photon reconstruction, leads to a small improvement in the jet energy resolution at high energy. Jet energy resolution is defined as the root mean squared divided by the mean for the smallest width of distribution that contains 90% of entries, using $Z' \rightarrow u/d/s$ sample. The di-jet energy is sampled at 91, 200, 360 and 500 GeV. Shown in figure 5.7, the jet energy resolutions are better at 360 and 500 GeV with improved photon reconstruction.

The improvement of the photon is also demonstrated in chapter 6, where tau lepton decay modes are classified. Excellent photon reconstruction leads to a high classification rate.

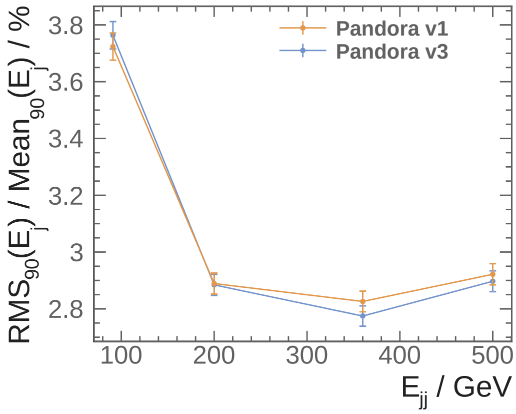


Figure 5.7: figure 5.7 shows jet energy resolution as a function of the di-jet energy using $Z' \rightarrow u/d/s$ sample. The top orange and bottom blue dots are reconstructed with PandoraPFA version 1 and version 3. The photon reconstruction is changed in PandoraPFA version 2.

5.9 Breakdown of photon reconstruction improvement

As stated before, photon reconstruction algorithm in section 5.2 and photon splitting algorithm in section 5.7 improves the photon completeness and the photon pair resolution. The fragment removal algorithm in section 5.5 removes fragments in the ECAL. High energy fragment removal algorithm in section 5.6 removes fragments in the HCAL. To show the incremental improvement, the average number of particle for a high energy photon pair, 500 - 500 GeV is shown in figure 5.8. With fragment removal algorithm in the ECAL, the number of fragment is reduced significantly comparing to figure 5.5b, shown as the orange dots. The high energy fragment removal algorithm further reduces the number of fragments, shown as the green dots. At 40 mm apart, with both fragment removal algorithms, there is less than 0.05 fragment per photon pair, which is similar to the best performance. The introduction of the revised photon reconstruction and photon splitting improves the photon separation resolution. Photons pair starts to be resolved at 5 mm instead of 10 mm for 500 - 500 GeV pair. Also two photons are fully resolved at 15 mm instead of 40 mm apart.

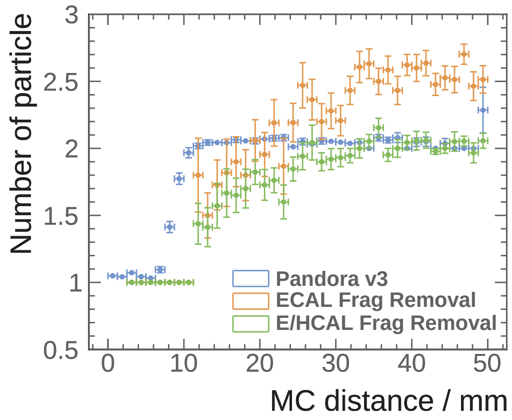


Figure 5.8: Figure shows the average number of photons, as a function of the MC distance separation in the calorimeter, using two photons of 500 and 500 GeV per event sample. The blue, orange, and green dots are reconstructed with PandoraPFA version 3, PandoraPFA version 1 with fragment removal in the ECAL (section **??**), and PandoraPFA version 1 with fragment removal in the ECAL and the HCAL. The photon reconstruction is changed in PandoraPFA version 2.

5.10 Photon reconstruction performance

In section 5.8, the improved performance of the photon reconstruction is demonstrated with different metrics, using single photon, double photons and jet samples. In this section, the features of the photon reconstruction will be described.

For simple samples such as two photons per event, there are very few fragments. Shown in figure 5.9a for 500 and 50 GeV photons pair sample, the average number of photons beyond 20 mm apart is 2 within errors. The average number of particle is less than 0.05 larger than the average number of photons.

The resolving power of a photon pair depends on energies of two photons. figure 5.9b is an example of average number of photon reconstructed for differen photon pairs. When the energies of two photons are similar, the resolving distance is shorter. This is because that the two photon showers have similar sizes, and the peak finding algorithm can exploit the symmetry. For example, 500 - 500 GeV photon pair and 10 - 10 GeV photon pair start to be resolved at 6 mm apart, which is about 1 ECAL cell. The asymmetrical photon pair, 500 - 50 GeV and 100 - 10 GeV pair, starts to be resolved at 10 mm apart, which is about 2 ECAL cell.

For the energetic photon, it is more difficult to remove fragments, but it is easier to identify the photon. The electromagnetic shower core is more dominant than the

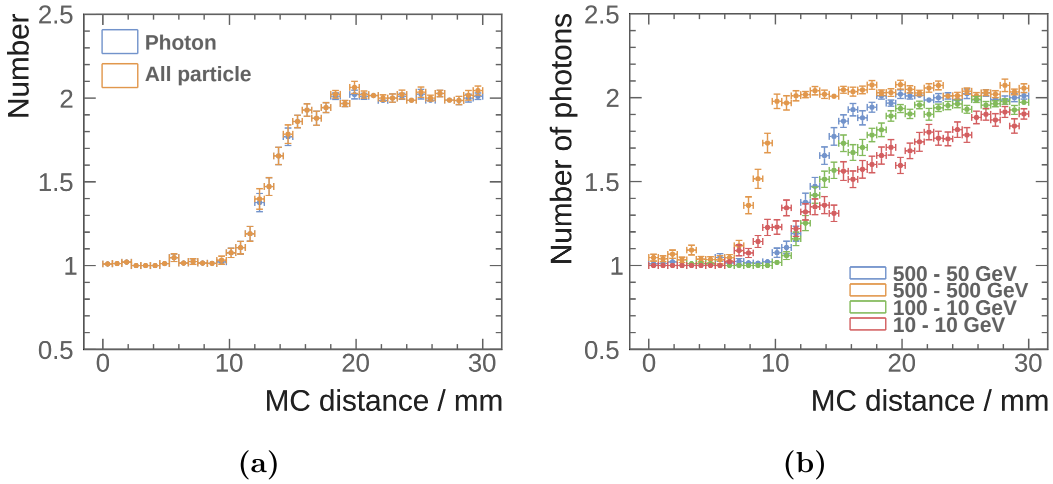


Figure 5.9: figure 5.9a shows the average numbers of photon and particle using two photons of 500 and 50 GeV per event sample. figure 5.9b shows the average numbers of photon for four different photon pairs: 500 - 50, 500 - 500, 100 - 10 and 10 - 10 GeV.

peripherals. Therefore separating two energetic photons is easier than separating two low energy photons. This can be seen in figure 5.9b. At 20 mm apart, two photons in 500 - 500 GeV pair are fully resolved, where approximately 60% of two photons in 10 - 10 GeV pair are resolved.

This set of photon related algorithms have been incorporated into the default reconstruction chain in PandoraPFA. The CLIC simulation studies have benefited from the improved photon reconstructions in various physics process, such as $H \rightarrow \gamma\gamma$.

Chapter 6

Tau Lepton Final State Separation

"MVA: Turn numbers into gold."
— TMVA

6.1 Introduction

Why study tau

Tau lepton has been examined closely in the past. The decay and the spin of the decay product were direct tests to the standard model. The spin of the decay product, using a Higgs decaying to tau tau channel, allows one to determine the spin of the higgs. Also, as tau is short-lived, only its decay products can be detected and reconstructed in the detector. Therefore, the ability to reconstruct and separate different tau decay modes is benchmark of detector performances.

This chapter will describe a tau final decay separation study. The processor developed for the study is used to test different detector models, as a proof-of-principle of detector optimisation using tau decay separation. Lastly, the spin of the Z was studied using one Z decaying to two tau tau channel.

6.2 Simulation and reconstruction

$e^-e^+ \rightarrow \tau^-\tau^+$ channel is used for the tau decay mode separation study. Generator software WHIZARD 1.95 [29] is used to generated simulated Monte Carlo (MC) samples. Hadronisation is described with PYTHIA 6.4 [31], which is tuned to the LEP results []. The spin effect of tau lepton decay is described by TAUOLA [33].

Final state radiation (FSR) was simulated. The initial state radiation (ISR) and the beam induced background were not simulated.

Events were simulated with the CLIC_ILD detector concept, using software with MOKKA [36], based on the GEANT 4 package [35]. Events were reconstructed with ilcsoft version v01-17-07 [40] and PandoraPFA version v02-02-00 [38], where the photon reconstruction is described in [41].

6.3 Generator level cut

To study the difference between different tau decay modes, clear topological difference is required. Therefore, events was considered if the event passes a set of cuts at generator level, listed here

- the final state photons not converting to electron pair in the tracker,
- the tau leptons decaying in the barrel and the end cap regions, which are defined as polar angle between 0.3 to 0.6 rad and 0.8 to 1.57 rad, and
- the visible energy of the tau lepton decay products more than 5 GeV, where the visible energy of the tau lepton decay is defined as the energy of the tau minus the energy of the tau neutrino.

The angular requirement is due to the gap region between the barrel and and the end cap of calorimeters, which degrades the PFO resolution significantly.

Around two million events were simulated for this study.

Table 6.1: Branching ratios of the seven major τ^- decays, taken from [2]. τ^+ decays similarly to τ^-.

Decay final state	Branching ratio / %
$e^- \ \overline{\nu}_e \ \nu_\tau$	17.83 ± 0.04
$\mu^- \ \overline{\nu}_\mu \ \nu_\tau$	17.41 ± 0.04
$\pi^- \ \nu_\tau$	10.83 ± 0.06
$\rho(\pi^-\pi^0)_{770}\nu_\tau$	25.52 ± 0.09
$a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$	9.30 ± 0.11
$a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$	8.99 ± 0.06
$\pi^- \ \pi^- \ \pi^+ \ \pi^0 \ \nu_\tau$	2.70 ± 0.08

6.4 Decay modes

Seven major decay final states of the tau lepton shown in table 6.1 were studied, covering 92.58 % of all tau decays [2]. Decay modes not listed in the table have branching fractions lower than 1% each. These final states can be classified into three categories: leptonic decays ($e^- \ \overline{\nu}_e \ \nu_\tau$ and $\mu^- \ \overline{\nu}_\mu \ \nu_\tau$), one-prong with photons ($\pi^- \ \nu_\tau$, $\rho(\pi^-\pi^0)_{770}\nu_\tau$ and $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$), and three-prong with photons ($a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$ and $\pi^- \ \pi^- \ \pi^+ \ \pi^0 \ \nu_\tau$).

The studied channel, $e^-e^+ \to \tau^-\tau^+$, contains two τ decaying in opposite directions. To select decay products of one τ, the fiducial detector space was divided into two halves. Event shape variable thrust is used to separate two halves. The classical event shape thrust [42], is defined as

$$T = \max_{\hat{t}} \frac{\sum_i \left|\hat{t} \cdot \vec{p_i}\right|}{\sum_i |\vec{p_i}|} \tag{6.1}$$

where $\vec{p_i}$ is the momentum vector of the particle i. Summation is over all particles in the event. Thrust axis, $\hat{t}$, is a unit vector. (Principle) Thrust value, T, is 1 for a perfect pencillike back-to-back two-jet event, and 0.5 for a perfect spherical event. The sign of dot product between thrust axis and PFO momentum determines which half the PFO falls into.

6.5 Discriminative variables

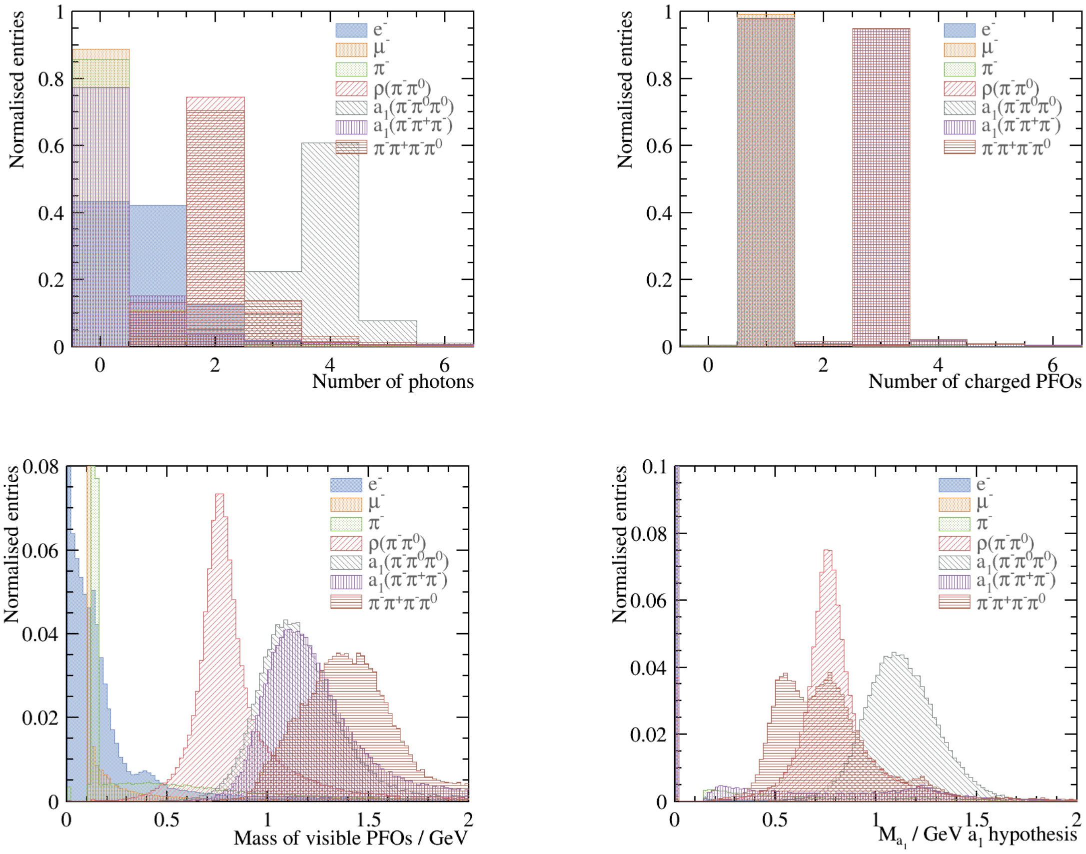


Figure 6.1: Normalised distribution for selected discriminative variables for seven final states, $e^- \ \overline{\nu}_e \ \nu_\tau$, $\mu^- \ \overline{\nu}_\mu \ \nu_\tau$, $\pi^- \ \nu_\tau$, $\rho(\pi^-\pi^0)_{770}\nu_\tau$, $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$, $a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$ and $\pi^- \ \pi^- \ \pi^+ \ \pi^0 \ \nu_\tau$, separated using truth information, for $\sqrt{s} = 100\,\text{GeV}$ for nominal CLIC_ILD detector model. The top left, top right, bottom left and bottom right plots are the normalised entries against the number of photons, number of charged PFOs, invariant mass of visible PFOs, and the invariant mass of $a_1(\pi^-\pi^0\pi^0)_{1260}$ for hypothesis test, respectively. There is a clear distinction between different final states in each plot.

Some variables with most discriminative power are shown in figure ??. In total 29 variables used in the multivariate analysis. The reason for the large number of variables is due to training seven decay modes at once, which will be discussed later.

Here is a full list of all variables used in the multivariate analysis. Energy of the τ is assume to be the same as the energy of $e^\pm$ colliding beam, which is half of the $\sqrt{s}$ $=$ e TeV nergy. Recoil momenta were calculated assuming the $e^- \ e^+$ collision happened at the centre of mass energy. Both assumptions are largely valid when there is no ISR contribution.

- $\frac{E_{ECal,HCal}}{E_{tot}}, charged$: Sum of energy deposited in ECal and HCal, divided by the energy of charged particles
- $\frac{E_{ECal,HCal}}{E_{tot}}, all$: Sum of energy deposited in ECal and HCal, divided by the energy of all particles
- m_{vis}: Invariant mass of visible particles in GeV
- $\frac{E_{vis}}{E_{\tau^-}}$: Sum of energy of all particles, divided by the energy of τ^-
- $\frac{E_{charged}}{E_{\tau^-}}$: Sum of energy of charged particles, divided by the energy of τ^-
- $\frac{E_{\mu^-}}{E_{\tau^-}}$: Sum of energy of muons, divided by the energy of τ^-
- $\frac{E_{e^-}}{E_{\tau^-}}$: Sum of energy of electrons, divided by the energy of τ^-
- $\frac{E_{\gamma}}{E_{\tau^-}}$: Sum of energy of photons, divided by the energy of τ^-
- $\frac{E_{\pi^-}}{E_{\tau^-}}$: Sum of energy of charged pions, divided by the energy of τ^-
- $N_{charged}$: Number of charged particles
- N_{μ^-}: Number of muons
- N_{e^-}: Number of electrons
- N_{γ}: Number of photons
- N_{π^-}: Number of charged pions
- m_{γ}: Invariant mass of photons in GeV
- $m_{charged}$: Invariant mass of charged particles in GeV
- $m_{neutral}$: Invariant mass of neutral particles in GeV
- m_{π^-}: Invariant mass of charged pions in GeV
- $m_{\pi^0}, \rho(\pi^-\pi^0)_{770} hypothesis$: Fitted invariant mass of π^0 for $\rho(\pi^-\pi^0)_{770}$ hypothesis test

- $m_{\rho(\pi^-\pi^0)_{770}}, \rho(\pi^-\pi^0)_{770}\mathit{hypothesis}$: Fitted invariant mass of $\rho(\pi^-\pi^0)_{770}$ for $\rho(\pi^-\pi^0)_{770}$ hypothesis test

- $m_{\pi^0 1}, a_1(\pi^-\pi^0\pi^0)_{1260}\mathit{hypothesis}$: First fitted invariant mass of π^0, for $a_1(\pi^-\pi^0\pi^0)_{1260}$ hypothesis test, ordered by closeness to the true π^0 mass

- $m_{\pi^0 2}, a_1(\pi^-\pi^0\pi^0)_{1260}\mathit{hypothesis}$: Second fitted invariant mass of π^0, for $a_1(\pi^-\pi^0\pi^0)_{1260}$ hypothesis test, ordered by closeness to the true π^0 mass

- $m_{a_1(\pi^-\pi^0\pi^0)_{1260}}, a_1(\pi^-\pi^0\pi^0)_{1260}\mathit{hypothesis}$: Second fitted invariant mass of $a_1(\pi^-\pi^0\pi^0)_{1260}$, for $a_1(\pi^-\pi^0\pi^0)_{1260}$ hypothesis test

- $\bar{E}_{cell}$: Average energy deposited in a calorimeter cell in GeV

- $d_{trans,shower}$: Transverse shower width for electromagnetic shower profile, averaged for all clusters in the ECal

- $l_{long,shower}$: Longitudinal start layer for electromagnetic shower profile, averaged for all clusters in the ECal

- $\Delta l_{long,shower}$: Longitudinal discrepancy for electromagnetic shower profile, averaged for all clusters in the ECal

- $\%MIP$: Fraction of calorimeter hits registered as minimum ionised particles, averaged for all clusters in the ECal

- $\frac{E}{P}$: Energy divided by momentum, averaged for all clusters in the ECal

Number of photons is an important variable for separating decay modes. This information is only available due to the excellent photon reconstruction. Shown in figure **??**, the majority of $\mu^- \ \bar{\nu}_\mu \ \nu_\tau$, $\pi^- \ \nu_\tau$ and $a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$ final states have zero photon reconstructed. The $e^- \ \bar{\nu}_e \ \nu_\tau$ final state event have one photon reconstructed instead of zero, due to the FSR effect. $\rho(\pi^-\pi^0)_{770}\nu_\tau$ and $\pi^- \ \pi^- \ \pi^+ \ \pi^0 \ \nu_\tau$ have nearly 80% events with two reconstructed photons, whilst $a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$ have over 60% events with four reconstructed photons. The loss is efficiency is due to the increasing difficulty to separate nearby photons.

The number of charged PFOs can clearly separate the leptonic and 1-prong final states, from the 3-prong final states, shown in figure **??**. The efficiency of leptonic final states are over 98%.

The invariant mass of the visible PFOs shows clear differences between different final states. $\rho(\pi^-\pi^0)_{770}\nu_\tau$, $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$ and $a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$ distribution show clear resonance at ρ and $a_1(1260)$. $e^-\ \overline{\nu}_e\ \nu_\tau$, $\mu^-\ \overline{\nu}_\mu\ \nu_\tau$ and $\pi^-\ \nu_\tau$ distribution show much smaller invariant mass and $\pi^-\ \pi^-\ \pi^+\ \pi^0\ \nu_\tau$ shows a large invariant mass than $a_1(1260)$. The $e^-\ \overline{\nu}_e\ \nu_\tau$ final state has a long tail of invariant mass due to the extra photons from the FSR.

$\frac{E_{ECal,HCal}}{E_{tot}}, charged$ and $\frac{E_{ECal,HCal}}{E_{tot}}, all$ are both very effective at picking out leptonic decay modes.

For final states containing ρ and $a_1(1260)$ resonance, it is useful to use minimisation test for right pairing of the resonance. We will use $a_1(\pi^-\pi^0\pi^0)_{1260}$ as an example. $\rho(\pi^-\pi^0)_{770}$ is very similar.

The minimisation of $a_1(\pi^-\pi^0\pi^0)_{1260}$ hypothesis states

$$\chi^2_{a_1(1260)} = \left(\frac{m_{a_1(1260),fit} - m_{a_1(1260)}}{\sigma_{a_1(1260)}}\right)^2 + \left(\frac{m_{\pi^0,fit} - m_{\pi^0}}{\sigma_{\pi^0}}\right)^2 + \left(\frac{m_{\pi^{0*},fit} - m_{\pi^0}}{\sigma_{\pi^0}}\right)^2, \quad (6.2)$$

where $m_{\pi^0,fit}$ and $m_{\pi^{0*},fit}$ are the invariant masses of all possible two photons combinations, $\sigma_{a_1(1260)}$ and σ_{π^0} are the half width of the invariant mass distribution of reconstructed $a_1(1260)$ and π^0 using the truth information, and $m_{a_1(1260)}$ and m_π are the masses of $a_1(1260)$ and π^0, taken from [2]. If there are only two or three photons, the $\chi^2_{a_1(1260)}$ expression will be reduced and not including $m_{\pi^{0*},fit}$ term, assuming two photons are merged in the reconstruction. If there are fewer than two photons, the $\chi^2_{a_1(1260)}$ expression would only contain $m_{a_1(1260),fit}$ term.

For the $\rho(\pi^-\pi^0)_{770}\nu_\tau$ final state, a similar $\chi^2_{\rho(770)}$ test for $\rho(770)$ hypothesis is used to extract $m_{\rho(770),fit}$ and $m_{\pi^0,fit}$ variables. $\chi^2_{\rho(770)}$ is similar to $\chi^2_{a_1(1260)}$ with $\rho(770)$ replacing $a_1(1260)$ and only one $m_{\pi^0,fit}$ term.

figure **??** shows the $m_{a_1(1260),fit}$ where $\rho(\pi^-\pi^0)_{770}\nu_\tau$, $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$ and $\pi^-\ \pi^-\ \pi^+\ \pi^0\ \nu_\tau$ final states contribute to the $a_1(1260)$ resonance, although only $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$ final has a real $a_1(1260)$ resonance. This is due to the structure of the $\chi^2_{a_1(1260)}$ minimisation function allowing final states with more than two photons and one $\pi^\pm$ to contribute.

Last six variables in the list help to differentiate an electron final state to that of a charged pion. A charged pion that starts showering early in the calorimeter could have a similar topology to an electromagnetic shower. Nevertheless, a good separation between the two can be achieved with the help of these variables.

6.6 Multivariate Analysis

For the multivariate analysis, the multiclass class of the TMVA package [43] was used to perform a multiclass classification, which trains the seven final states simultaneously. The multiclass class is an extension of the standard two-class signal-background classifier.

There are two ways for the training. "One v.s. one" is each class is trained against each other class. And the overall likelihood is normalised. The second way to train is called "one v.s. all", which is when each class is trained against all other classes.

Using a three-class example, A, B and C, "one v.s. one" scheme trains A against B, B against C, and C against A. Then the likelihood is normalised. "One v.s. all" would train A against B plus C, B against A plus C, and C against A plus B.

TMVA multiclass implementation uses "one v.s. all" scheme. For each final state, the multiclass classifier will train the final state as the signal against all other final states as the background. This process is repeated for each final state. The classifier output for a single event is a normalised response for each final state, where the sum is one. The response of each final state of a event can be treated as the likelihood. The event is classified into a particular final state if the final state has the highest classifier output response. The advantage of using the multiclass is that the correlation between different final states are accounted for and the classifier output are correctly adjusted for multiple final states, hence one event can only be classified into one final state. The issue with the multiclass is that discriminative variables for each final state need enter the training stage, resulting in a large number of variables.

Half of the randomly selected samples were used in the training process and the other half were used for testing.

The TMVA multiclass classifier used is boosted decision tree with gradient boosting (BDTG), as it was found to give for the best performance. The MVA classifier is trained and optimised to give the best overall separation across all final states. MVA will be discuss further in section **??**

6.7 Result

Reco ↓ True →	$e^-\ \overline{\nu}_e$	$\mu^-\ \overline{\nu}_\mu$	π^-	$\rho(\pi^-\pi^0)$	$a_1(\pi^-\pi^0\pi^0)$	$a_1(\pi^-\pi^-\pi^+)$	$\pi^-\ \pi^-\ \pi^+\ \pi^0$
$e^-\ \overline{\nu}_e$	**99.8**	-	0.9	1.1	0.8	-	-
$\mu^-\ \overline{\nu}_\mu$	-	**99.5**	0.5	-	-	-	-
π^-	-	0.3	**93.2**	0.9	-	0.4	-
$\rho(\pi^-\pi^0)$	-	-	4.1	**93.0**	10.5	0.6	2.8
$a_1(\pi^-\pi^0\pi^0)$	-	-	-	4.3	**88.2**	-	1.0
$a_1(\pi^-\pi^-\pi^+)$	-	-	1.0	0.3	-	**96.6**	6.9
$\pi^-\ \pi^-\ \pi^+\ \pi^0$	-	-	-	0.4	0.4	2.4	**89.3**

Table 6.2: The percentage of reconstructed decay modes corresponds to underlying true decay modes, with $\sqrt{s} = 100\,\text{GeV}$ for nominal CLIC_ILD detector model. Bold numbers show the correctly reconstructed percentages. Numbers less than 0.25% are not shown. Statistical uncertainties are less than 0.25%. Final states include ν_τ, which is not shown.

The reconstruction efficiencies for the seven final state of the tau decaying with c.o.m. energy of 100 GeV for the nominal CLIC_ILD detector are shown in table 6.2. The perfect reconstruction would result in only terms in the diagonal.

The unprecedented high classification rate has been achieved. The improvement of photon reconstruction described in section **??** improved the ability to separate 1-prong final state. Most notably, figure **??** shows number of photons have a high correct reconstruction efficiency.

For leptonic decay, the selection efficiency is above 99.5% as the tracking system have much better resolution than the calorimeter.

The $\mu^-\ \overline{\nu}_\mu$ final state has very clear topology, as muon deposits energy in the muon chamber. Therefore, there is little confusion with other final states.

$e^-\ \overline{\nu}_e$ final state is well separated, due to the specialised variables aimed to differentiate early hadronic shower to electromagnetic shower. However, there is still about 1% confusion in one prong final state.

For one prong final states, π^- , $\rho(\pi^-\pi^0)$, and $a_1(\pi^-\pi^0\pi^0)$, the confusion is mainly due to the imperfect separation of nearby photons, originated form π^0.

Similarly the confusion between 3-prong final state, $a_1(\pi^-\pi^-\pi^+)$, and $\pi^- \pi^- \pi^+ \pi^0$ is caused by the inability to resolve photon pairs.

6.8 Electromagnetic calorimeter optimisaiton

As discussed above, the tau decay mode separation is an benchmark test of detector performance. The ability to resolve photon pairs is crucial to separate different 1-prong states, and different 3-prong state. One of the main feature of calorimeter design affecting the photon resolution is the size of electromagnetic calorimeter (ECal) cell for the high granular calorimeter. The finer ECal cell size is, the better resolution of reconstructing individual photons.

The classification is being tested with the impact of different $\sqrt{s}$ and different ECal square cell sizes. Around two million events were simulated at each $\sqrt{s}$ = 100, 200, 500 and 1000 GeV, with each different ECal square cell sizes of 3, 5, 7, 10, 15 and 20 mm. Events were simulated and reconstructed in the same way as described above, with same selection applied. MVA classifier was trained individually for each $\sqrt{s}$ and each ECal square cell size, with same set of discriminative variables.

To access the impact different ECal square size on detector performance, in particular ECal performance, the correct reconstruction efficiency for 1-prong and 3-prong final states is used as metric. The higher $\sqrt{s}$ of the collision would degrade the performance, as photons are more boosted and more difficult to resolve.

The leptonic decay correct reconstruction efficiency is not used as a metric as they are similar across different ECal cell sizes. This is because the $e^\pm$ and $\mu^\pm$ identifications mostly rely on the tracking system, which was not varied in this study. The energy deposited in the calorimeter are used for the association to the tracks but it has a small impact on the lepton identification.

figure 6.2 shows that as the ECal cell sizes increase, the reconstruction efficiencies generally decrease. Larger cell sizes have lower spatial resolutions, making the separating of nearby photons more difficult.

For the $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$ final state, the selection efficiency for 500 GeV rises from ECal cell sizes 15 mm to 20 mm and the one for 1000 GeV rises from 7 , to 20 mm actually goes up as cell size increases. This is because when the algorithm can not reconstruct

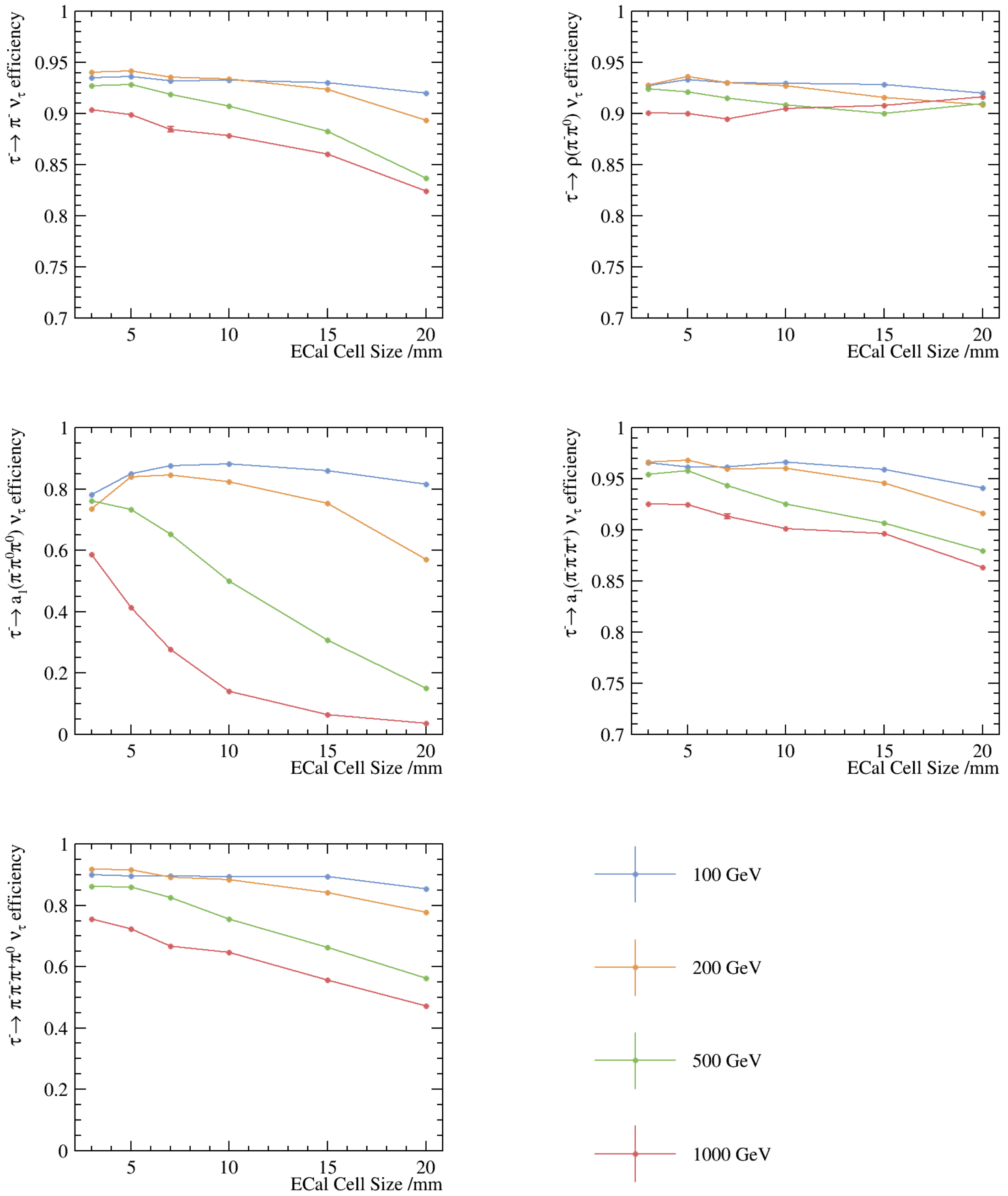


Figure 6.2: The selection efficiencies for various final states against the ECal cell size for different c.o.m. energies with the nominal CLIC_ILD detector model are shown. The top left, top right, middle left, middle right and bottom left plots are for the $\pi^- \nu_\tau$, $\rho(\pi^-\pi^0)_{770}\nu_\tau$, $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$, $a_1(\pi^-\pi^-\pi^+)_{1260}\nu_\tau$ and $\pi^- \pi^- \pi^+ \pi^0 \nu_\tau$ final states respectively. From the top to the bottom, blue, orange, green and red lines are representing the $\sqrt{s} = 100$, 200, 500 and 1000 GeV respectively.

four photons in the $a_1(\pi^-\pi^0\pi^0)_{1260}\nu_\tau$ final state, and the event topology would be very similar to the $\rho(\pi^-\pi^0)_{770}\nu_\tau$ final states.

For the $\sqrt{s} = =$TeV 100 and 200 GeV, the selection efficiency of the 5 mm ECal cell size is better than that of the 3 mm. One possible explanation is that the and the PandoraPFA have been optimised for the nominal ILD detector with the 5 mm ECal cell size, which shares the same ECal structure with the nominal CLIC_ILD detector.

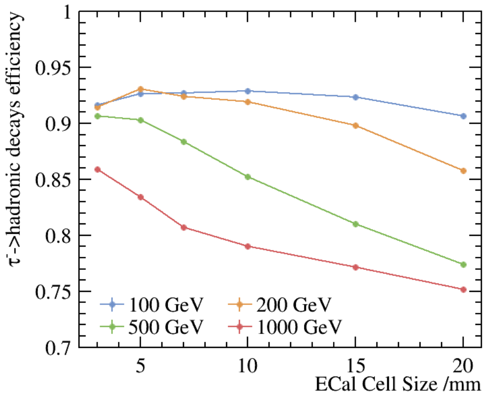


Figure 6.3: The τ hadronic decay efficiency against the ECal cell size for different $\sqrt{s} =$ e TeV nergies with the nominal CLIC_ILD detector model are shown. The blue, orange, green and red lines are representing the $\sqrt{s} = 100$, 200, 500 and 1000 GeV respectively.

To effectively compare the overall separation power of all hadronic final states across $\sqrt{s}$ and ECal square cell sizes, we constructed a single parameter function, the τ hadronic decay final state efficiency function,

$$\varepsilon_{had} = \frac{\Sigma_i Br_i \varepsilon_i}{\Sigma_i Br_i}, \tag{6.3}$$

where Br_i is the branching fraction of a hadronic final state after the generator level cut. ε_i is the correct reconstruction efficiency of the final state, and the i is summing over five hadronic decay final state of τ. Leptonic decays, $e^-\ \overline{\nu}_e\ \nu_\tau$ and $\mu^-\ \overline{\nu}_\mu\ \nu_\tau$, were not included, because the variation of the leptonic decay selection efficiency is small.

In the figure 6.3, τ hadronic decay final state efficiency, ε_{had}, against the ECal cell size with different $\sqrt{s}$ is shown. ε_{had} decreases when cell sizes increases and when $\sqrt{s}$ increases. ε_{had} of the 5 mm ECal cell size is better than that of the 3 mm for $\sqrt{s} = 100$

and 200 GeV lines possibly due the optimisation of the software for the nominal ILD 5 mm ECal square cell size.

The ε_{had} is above 90% for the ECal cell size from 3 to 20 mm for the $\sqrt{s} = 100$ GeV. For $\sqrt{s} = 200$ GeV, the ε_{had} decreases from over 90% to 86% for the ECal square cell size from 3 to 20 mm. The degradation of the ε_{had} is more significant for the $\sqrt{s} = 500$ and 1000 GeV, where the ε_{had} drops from over 90% to 77% and from 86% to 75% respectively, over the same range of ECal square cell size.

For $\sqrt{s} = 100$ and 200 GeV, up to 15 mm cell sizes of ECal will give a good performance for τ hadronic decay modes separation, and the ε_{had} is above 90%. For $\sqrt{s} = 500$ and 1000 GeV, it is preferential to have a small ECal cell size for a good τ hadronic decay modes separation. There is about 15% degradation of ε_{had} for ECal square cell size from 3 to 20 mm.

6.9

Chapter 7

Double Higgs Bosons Production Analysis

"Two is better than one"
— Sir Steve Orange, 1785–1854

Since the discovery of Higgs boson in the LHC in 2012 [1,16], it is crucial to understand the properties of the Higgs boson and test if it is a Standard Model Higgs. The Higgs mechanism and the Higgs boson in the Standard Model have been explained in the chapter 2. A few theories of Higgs beyond the Standard Model could be tested via the double higgs production in an electro-positron collider (section 2.9). A generator level studies has shown that the precision reached by a multi-TeV linear collider, such as the Compact Linear Collider (CLIC) is much better than it by the Large Hadron Collider [13]. In this chapter, a full CLIC_ILD detector simulation studies has been performed for the double higgs production, $\mathrm{e^-e^+} \to \mathrm{HH}\nu\bar{\nu}$. First event generation and simulation will be briefly discussed. An overview of the analysis, including lepton finding and jet reconstruction, is presented. This is followed by a multivariate analysis with results on the selection efficiency. Lastly, a uncertainty on extracting triple Higgs self couplings, $\mathrm{g_{HHH}}$, and quartic coupling, $\mathrm{g_{WWHH}}$, is presented.

7.1 Analysis Straggly Overview

Proof-of-principle generator study was perform at CLIC using CLIC_ILD detector model for $\sqrt{s} = 1.4\,\mathrm{TeV}$ and $3\,\mathrm{TeV}$ [18]. For the full CLIC_ILD detector simulation, firstly suitable signal and background channels are identified. To help selecting the signal, light lepton and tau lepton are found and b-jet tagging is used. An event is group into a number of jets depending on the final state, followed by pre-selections cuts and multivariate analysis.

7.2 Monte Carlo Sample Generation

The double higgs production process is referring to $e^-e^+ \rightarrow HH\nu\bar{\nu}$. It is divided into sub-channel $HH \rightarrow b\bar{b}W^+W^-$ and $HH \rightarrow b\bar{b}b\bar{b}$ to allow a more detailed examination, providing a cross checks between two sub-channels and an improvement of signal selection when combined. In this chapter, the W hadronic decay mode of the $HH \rightarrow b\bar{b}W^+W^-$ channel is studied, and called the signal channel. It is chosen because the hadronic decay has the largest cross section and does not produce neutrinos, which means that each W can be reconstructed using a di-jet.

Selected background samples are considered in the analysis and listed in Table **??**. These background were expected share similar topologies with the signal process. The ignored background samples either have different different event topologies, or they have very small cross sections. For example $e^{\pm}\gamma() \rightarrow qqH\ell$ is ignored as the cross section is vert small even at $\sqrt{s} = (\,\mathrm{TeV}\ 3)$ (0.07 fb with photon from EPA, 0.6 fb with photon from BS).

Most two-quark and four-quark final states were considered. Six-quark final states were not considered due to computational limitation. Since the significant presence of beamstrahlung, where photon produced due to the high electric field generated by the colliding beams, processes initiated by photons are also included. Processes invloving real photons from beamstrahlung (BS) and "quasi-real" photons are generated separately. For the "quasi-real" photon initiated processes, the Equivalent Photon Appproximation (EPA) has been used.

Single higgs productions are considered and it is shown later that it is different to completely remove single higgs events when studying double higgs events.

For processes involving Higgs production explicitly, simulated Higgs mass is 126 GeV. Multi-quark final state background samples could, in principle, contain higgs production. Therefore, they are generated with a Higgs mass of 14 TeV. This will produce negligible double higgs production cross section. The cross section for sub-channels of the signal, such as $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$, are corrected manually according to [44], as the internal value of the event generation software (PYTHIA) is inaccurate.

The simulation and reconstruction chain are described in chapter 5. For most background processes, events are simulated when invariant mass of quarks are above 50 GeV. For electron-photon interaction with $qqqq\nu$ final state, events are simulated when invariant mass of quarks are above 120 GeV. These limits are necessary to generate a large amount of background samples in a feasible time, without losing much signal samples.

Finally, the main beam induced background $\gamma\gamma \to hadrons$ is simulated and overlayed to all samples according to the integration time of each subdetector. Details can be found in section 4.5.

7.3 Physics object and event reconstruction

The reconstruction is done via Marlin in iLCSoft v01-16. Separate software package (processor) exists for identification of electrons, muons, taus, and jet reconstruction. New processors have been developed and existing processors have been optimised for signal selection and background rejection. The latest functioning flavour tagging processor exist in iLCSoft v01-16. Thus newer versions of iLCSoft can not be used in this analysis.

For the signal channel, $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-} \to \mathrm{b\bar{b}}qqqq$, there is no lepton in the final state. Hence effective lepton identifiers would improve the signal identification. The lepton finding processors are wither developed or optimised with samples at $\sqrt{s} = 1.4$ TeV, and checked against samples at $\sqrt{s} = 3$ TeV. Because the expected signal significance would be low, the processors are optimised to aggressively reject more background, to increase the final signal significance. It was found that the same set of parameters for lepton identifiers work well under $\sqrt{s} = 1.4$ TeV and 3 TeV.

Channel	$\sigma(\sqrt{s}=1.4\,\mathrm{TeV})$ / fb	$\sigma(\sqrt{s}=3\,\mathrm{TeV})$ / fb
$e^-e^+ \to HH\nu\bar{\nu}$	0.149	0.588
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-$,hadronic	0.018	0.07
$e^-e^+ \to HH \to b\bar{b}b\bar{b}$	0.047	0.19
$e^-e^+ \to HH \to$ others	0.085	0.34
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	0.86	1.78
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	0.36	1.12
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	0.31	1.91
$e^-e^+ \to qqqq$	1245.1	546.5*
$e^-e^+ \to qqqq\ell\ell$	62.1*	169.3*
$e^-e^+ \to qqqq\ell\nu$	110.4*	106.6*
$e^-e^+ \to qqqq\nu\bar{\nu}$	23.2*	71.5*
$e^-e^+ \to qq$	4009.5	2948.9
$e^-e^+ \to qq\ell\nu$	4309.7	5561.1
$e^-e^+ \to qq\ell\ell$	2725.8	3319.6
$e^-e^+ \to qq\nu\nu$	787.7	1317.5
$e^-\gamma(BS) \to e^-qqqq$	1160.7	1268.7*
$e^+\gamma(BS) \to e^+qqqq$	1156.3	1267.6*
$e^-\gamma(EPA) \to e^-qqqq$	287.1	287.9*
$e^+\gamma(EPA) \to e^+qqqq$	286.9	287.8*
$e^-\gamma(BS) \to \nu qqqq$	79.8†	262.5*
$e^+\gamma(BS) \to \bar{\nu}qqqq$	79.3†	262.3*
$e^-\gamma(EPA) \to \nu qqqq$	17.4†	54.2*
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	17.3†	54.2*
$e^-\gamma(BS) \to qqH\nu$	15.8*	58.6*
$e^+\gamma(BS) \to qqH\nu$	15.7*	58.5*
$e^-\gamma(EPA) \to qqH\nu$	3.39*	11.7*
$e^+\gamma(EPA) \to qqH\nu$	3.39*	11.7*
$\gamma(BS)\gamma(BS) \to qqqq$	21406.2*	13050.3*
$\gamma(BS)\gamma(EPA) \to qqqq$	4018.7*	2420.6*
$\gamma(EPA)\gamma(BS) \to qqqq$	4034.8*	2423.1*
$\gamma(EPA)\gamma(EPA) \to qqqq$	753.0*	402.7*

Table 7.1: List of signal and background samples with the corresponding cross sections at $\sqrt{s} = 3\,\mathrm{TeV}$ and $\sqrt{s} = 1.4\,\mathrm{TeV}$. q can u, d, s, b or t. Unless specified, q, ℓ and ν represent particles and its corresponding anti-particles. γ (BS) represents a real photon from beamstrahlung (BS). γ (EPA) represents a "quasi-real" photon, simulated with the Equivalent Photon Approximation. For processes involving Higgs production explicitly, simulated Higgs mass is 126 GeV. Otherwise, Higgs mass is set to 14 TeV. For processes labeled with * and †, the generator level cut requires invariant mass of quarks greater than 50 and 120 GeV, respectively.

7.3.1 Electron and muon identification

IsolatedLeptonFinderProcessor

IsolatedLeptonFinderProcessor in Marlin package is used, modified, and optimised. This processor identifies high energy electrons and muons that are away from other particles, "isolated". The optimal parameters below are chosen in collaboration and tested using the signal channel and the $e^-e^+ \rightarrow qqqq\ell\nu$ channel.

Electrons induced electromagnetic showers are mostly contained in the ECAL, while muons would deposit some energies in the ECAL. The inner detector tracks of electrons and muons from the electron-positron interaction are primary tracks, which are very close to the interaction point. The isolation criteria requires the lepton to be far away from other high energy particles. These form the logics for the IsolatedLeptonFinderProcessor. The value for the selection cut is listed in table 7.2. E_{ECAL} is the energy deposited in the ECAL. d_0, z_0, and r_0 are the Euclidean distance of the track starting point to the interaction point in the x-y plane, the in z direction, and in the x-y-z three dimensional space. The performance of the processor is shown in table 7.3.

IsolatedLeptonFinderProcessor	Selection
High Energy (GeV)	$E > 15$
Energy deposition $e^\pm$	$\frac{E_{ECAL}}{E} > 0.9$
Energy deposition $\mu^\pm$	$0.25 > \frac{E_{ECAL}}{E} > 0.05$
Primary Track (mm)	$d_0 < 0.02, z_0 < 0.03, r_0 < 0.04$
Isolation (GeV)	$E^2_{cone} \leqslant 5.7 \times E - 50$

Table 7.2: Optimised selection criterion for IsolatedLeptonFinderProcessor

BonoLeptonFinderProcessor

The IsolatedLeptonFinderProcessor is rather conservative. I developed a new more aggressive light lepton selection processor, BonoLeptonFinderProcessor, that utilises calorimetric information provided by PandoraPFA.

The processor uses two chains of cuts.

First chain uses the particle ID information from PandoraPFA. A electron is identified if it is a “PandoraPFA” electron and the energy in the ECal is over 95% of the total calorimetric energy. A muon is identified if it is a “PandoraPFA” muon. Primary track selection states the Euclidean distance in the x-y-z three dimensional space of the track starting point to the impact point to be less than 0.015 mm, and the PFO energy is more than 10 GeV. The light lepton either satisfy the high p_T requirement of at least 40 GeV, or the isolation criteria,

$$E_l \geqslant 23 \times \sqrt{E_{cone}} + 5 \tag{7.1}$$

where E_{cone} and E_l have the same definition as in the IsolatedLeptonFinderProcessor.

Second chain of cuts is similar to the IsolatedLeptonFinderProcessor. An electron is identified if the energy in the ECal is over 95% of the total calorimetric energy. A muon is identified if the energy in the ECal is between 5% and 20% of the total calorimetric energy. Primary track selection states the Euclidean distance in the x-y-z three dimensional space of the track starting point to the impact point to be less than 0.5 mm, and the PFO energy is more than 10 GeV. The light lepton either satisfy the high p_T requirement of at least 40 GeV, or the isolation criteria,

$$E_l \geqslant 28 \times \sqrt{E_{cone}} + 30 \tag{7.2}$$

where, E_{cone} is the total energy of PFOs within an opening angle of $\cos^{-1}(0.99)$ of the light lepton, and E_l is the energy of the light lepton.

Comparison: IsolatedLeptonFinderProcessor v.s. BonoLeptonFinderProcessor

Two processors share similar criterion for light lepton identification. The main difference is that the BonoLeptonFinderProcessor allows high p_T light lepton to be identified in a potential non-isolated environment, which leads to the more aggressiveness of the BonoLeptonFinderProcessor. The performance of two processors on the signal and selected background samples is shown in table 7.3.

7.3.2 Tau identification

TauFinderProcessor

With a decay length of 87μm, tau leptons decay before reaching the detector and can only be identified through the reconstruction of their decay products. The leptonic decay of tau can be identified using the two isolated lepton finder processor. Therefore tau identification will focus on the hadronic decay.

TauFinderProcessor [45], an existing processor Marlin package, has been tuned in collaboration and tested. The a collection of tau decay productions are identified they pass a chain of cuts.

Particles are not considered if p_T is less than 1 GeV or $|\cos(\theta_Z)|$ is more than 1.1 rad, as they are more likely from beam induced background. A seed is considered if a charged particle has p_T more than 10 GeV. A search cone of opening angle 0.03 rad is then formed. The search cone is rejected if it has more than 3 charged particles, more than 10 particles or its invariant mass more than 2 GeV. An isolation cone is formed with opening angle between 0.03 and 0.33 rad of the seed. The seed is rejected if there are more than 3 GeV in the isolation cone.

BonoTauFinderProcessor

The TauFinderProcessor's performance is decent, but there is room for improvement. I developed a new more aggressive tau lepton selection processor, BonoTauFinderProcessor, that utilises calorimetric information provided by PandoraPFA.

Similar to the previous processor, PFOs with p_T less than 1 GeV are rejected. A tau seed is defined as a charged particle with p_T at least 5 GeV. The search cone has an opening angle of $\cos^{-1}(0.999)$. Particles are iteratively added to the search cone according to the size of the opening angle to the seed. A temporary search cone is then considered if it has one or three charged particles, and the invariant mass is less than 3 GeV. The search cone needs to satisfy one of isolation criterion.

1. No particle in the large isolation cone, and p_T of search cone at least 10 GeV,
2. One charged particle in the search cone, one particle in the large isolation cone, and r_0 larger than 0.01 mm,

Efficiency (1.4 TeV)	Signal	$e^+e^- \to qqqq\ell\nu$
IsolatedLeptonFinderProcessor	99.3%	50.3%
BonoLeptonFinderProcessor	99.1%	39.9%
TauFinderProcessor	97.5%	52.3%
BonoTauFinderProcessor	89.7%	38.5%
ForwardFinderProcessor	98.9%	95.1%
Combined	86.6%	16.8%
Efficiency (3 TeV)	Signal	$e^+e^- \to qqqq\ell\nu$
IsolatedLeptonFinderProcessor	99.5%	66.8%
BonoLeptonFinderProcessor	99.0%	52.5%
TauFinderProcessor	97.7%	79.5%
BonoTauFinderProcessor	86.3%	60.3%
ForwardFinderProcessor	95.9%	80.7%
Combined	81.0%	23.3%

Table 7.3: isolated lepton finder processors performance on the signal and selected background samples.

3. Three charged particle in the search cone, one particle in the large isolation cone, p_T of search cone at least 10 GeV, and search cone opening angle less than $\cos^{-1}(0.9995)$,
4. One charged particle in the search cone, no particle in the small isolation cone, r_0 larger than 0.01 mm, and p_T of search cone at least 10 GeV,
5. Three charged particle in the search cone, no particle in the small isolation cone, p_T of search cone at least 10 GeV, and search cone opening angle less than $\cos^{-1}(0.9995)$,

where large and small isolation cone are defined as opening angle of $\cos^{-1}(0.95)$, and $\cos^{-1}(0.99)$ respectively. If there are multiple temporary search cone of a same seed passing the isolation criteria, the cone with smallest opening angle is chosen for output.

Comparison: TauFinderProcessor v.s. BonoTauFinderProcessor

Two processors share similar size of search cone and isolation cone. The BonoTauFinder-Processor has looser cut on minimum p_T and invariant, but stricter isolation criterion. This leads to a more aggressive tau finder. The performance of two processors on the signal and selected background samples is shown in table 7.3

7.3.3 Very forward electron identification

Certain background channels, for example photon-electron interactions, contain electrons in the very forward part of the detector, namely LCal and BCal. These forward calorimeters were not simulated due to computational limitation. Most particle in these detector would be very forward particles from beam induced background. However, previous study has shown [] that high energy electrons can be identified with high efficiency. Due to the lack of tracking in these region, electrons and photons would have the same electromagnetic shower profile, with the given calorimeter resolution. MC photons and electrons are checked if they fall in the LCal or the BCal, and checked against the known detection efficiency.

Beam Calorimeter acceptance is defined as $|\cos(\theta_Z)|$ is between 0.01 and 0.04 rad and length in z direction is between 3181 and 3441 mm. Luminosity Calorimeter acceptance is defined as $|\cos(\theta_Z)|$ is between 0.038 and 0.11 rad and length in z direction is between 2539 and 2714 mm. For $\sqrt{s} =$ (TeV 3), the BeamCal detection efficiency is provided by a software package []. For $\sqrt{s} =$ (TeV 1.4), the same software for the BeamCal is used, by scaling the energy of the MC particle by a factor of $\frac{3}{1.4}$. For the LumiCal, the identification efficiency is defined as

$$\varepsilon = \begin{cases} 0, & \text{if } E < 50\,\text{GeV} \\ 0.99 \times \frac{(\mathrm{erf}(E-100)+1)}{2}, & \text{otherwise} \end{cases} \tag{7.3}$$

where E is the energy of the electron or the photon.

The background rejection is significant, shown in table ?? for the signal and selected background.

7.3.4 Other lepton identification processors

Other isolated lepton selection processors available in Marlin package, including IsolatedLeptonTagging and TauJetClustering, have been tested. The results, after some tuning of parameters, were unsatisfactory. They either performed poorly comparing to the processors above, or became redundant after the processors above. Therefore, these processors were not used in this analysis.

Selection / Efficiency (1.4 TeV)	Signal	$e^-\gamma(BS) \to e^- qqqq$
Combined light lepton finder	87.6%	67.5%
ForwardFinderProcessor	98.9%	53.6%
Combined	86.6%	30.8%
Processor / Efficiency (3 TeV)	Signal	$e^-\gamma(BS) \to e^- qqqq$
Combined light lepton finder	84.4%	72.7%
ForwardFinderProcessor	95.9%	55.4%
Combined	81.0%	33.4%

Table 7.4: Very forward electron and photon finder performance on the signal and selected background samples.

7.4 Jet reconstruction

The signal channel, $HH \to b\bar{b}W^+W^- \to b\bar{b}qqqq$, is a four-jet final state. A useful technique for the analysis is to reconstruct the four-jet final state using jet algorithms. This allows discriminative variables to be calculated.

7.4.1 Jet reconstruction optimisation

Longitudinal invariant, k_t, jet algorithm was chosen for the jet clustering. Due to the presence high level of beam induced background at the CLIC, it has been shown that a jet algorithm designed for hadron colliders are more effective than those traditional designed for the electron-positron collider, such as Durham algorithm. []

The free parameters for k_t algorithm is the R parameter, which controls the fatness of the jet. There is also the choice of the PFO collection, which incorporate different level of time and p_T cuts, to reduce beam induce background. Both parameters are optimised for $\sqrt{s} = 1.4$ TeV and $\sqrt{s} = 3$ TeV.

The details of jet algorithm can be found in section **??**.

The R parameter of the k_t jet algorithm, and the collection of the PFOs are chosen to give the best invariant mass resolution. When there are a few suitable candidate, analysis were performed in parallel. Decision were made to give the highest signal significance.

k_t jet algorithm was used as part of the FastJet algorithms available in the Marlin package.

The samples containing the signal, $HH \to b\bar{b}W^+W^- \to b\bar{b}qqqq$, was used for the optimisation of the jet reconstruction. The signal events were chosen using MC truth information.

Jet algorithm was run in exclusive mode, where number of jets is chosen to be six.

For the signal, $HH \to b\bar{b}W^+W^- \to b\bar{b}qqqq$, one Higgs decays to two b quarks, resulting in two jets from hadronisation. Similarly the other Higgs decays to two W bosons, where each W boson decays into two quarks. Therefore, the expected number of jets is six.

Jets produced by the k_t jet algorithm are pairred up using MC truth information, to the corresponding Higgs and W boson. Four invariant mass distributions are obtained: two Higgs masses, $m_{H_{bb}}$, $m_{H_{WW^*}}$, and two W masses m_W, m_{W^*}. W^* indicates the off-mass-shell W boson, because when a Higgs decays into two W bosons, one W is off the mass shell, as the Higgs mass is less than the sum two W masses.

Three mass distributions are worth comparing for different jet reconstruction, namely, $m_{H_{bb}}$, $m_{H_{WW^*}}$, and m_W. The ideal jet reconstruction should produce the a sharp mass peak around the particle's true mass.

To quantitatively access the mass distribution, a gaussian like fit is performed to extract the position of the peak, and the width of the distribution. The fit has the form:

$$f(m) = Ae^{-\frac{(m-\mu)^2}{g}} \begin{cases} g = 2\sigma_L + \alpha_L(m-\mu), & \text{if } m < \mu \\ g = 2\sigma_R + \alpha_R(m-\mu), & \text{if } m \geqslant \mu \end{cases} \tag{7.4}$$

The fit represents an asymmetrical gaussian function, where m is binned mass distribution, with 50 bins in range [0, 200] GeV. The fitted mass peak is denoted by μ. σ_L and σ_R allow asymmetrical width of the distribution. α parameter controls the fit of tails. Inspired by the $t\bar{t}$ analysis [], the use of the α parameter allows the fit in the whole mass range, otherwise only the peak of the distribution should be fitted with a gaussian like function. A is the normalistion factor. An example of the fit of $m_{H_{bb}}$ is shown in figure 7.1.

For $\sqrt{s} = 1.4$ TeV, shown in figure 7.2, normal selected PFO with $R = 0.7$ give a good fitted mass for H_{WW^*} and W. The mass is slightly too low for the H_{bb}. figure **??** shows the combined relative fitted width for the H_{bb}, H_{WW^*} and W. Normal selected PFO with $R = 0.7$ gives an almost optimal relative width for H_{bb}, while achieving a good balance

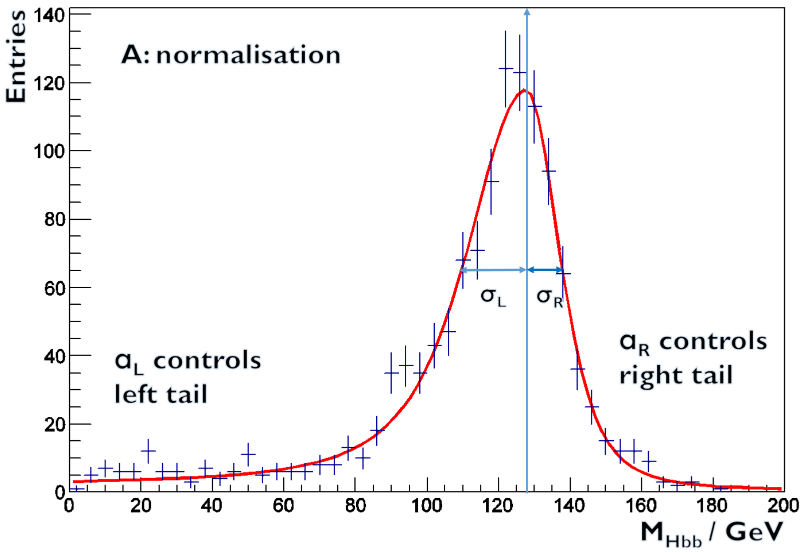


Figure 7.1: A typical example of MC mass fit of $m_{H_{bb}}$ for double higgs analysis. Red line indicates the best fit. Vertical arrow indicates the fitted peak position.

for H_{WW^*} and W. Therefore, normal selected PFO with $R = 0.7$ is chosen to be the optimal jet reconstruction parameters.

For $\sqrt{s} = 3\,\text{TeV}$, the choice is a bit more complicated. Shown in figure 7.3, fitted mass for H_{bb} favours normal selected PFO with $R = 0.8$. Fitted mass for H_{WW^*} favours tight selected PFO with $R = 0.9$. Fitted mass for W favours tight selected PFO with $R = 0.8$. Looking at the combined relative fitted width for the H_{bb}, H_{WW^*} and W, shown in figure ??, normal selected PFO gives a larger width than tight selected PFO. Within tight selected PFO, small R values provide a shaper width for H_{WW^*} and H_{bb}, but a broader width for W. Therefore, tight selected PFO with $R = 0.7$ and $R = 1$ are both chosen for parallel analysis.

Later it was shown that tight selected PFO with $R = 0.7$ gives a better signal significance. Therefore the optimal choice of jet reconstruction for $\sqrt{s} = 3\,\text{TeV}$ is tight selected PFO with $R = 0.7$.

The extracted fitted parameters of optimal jet reconstructions are summarised in table 7.5.

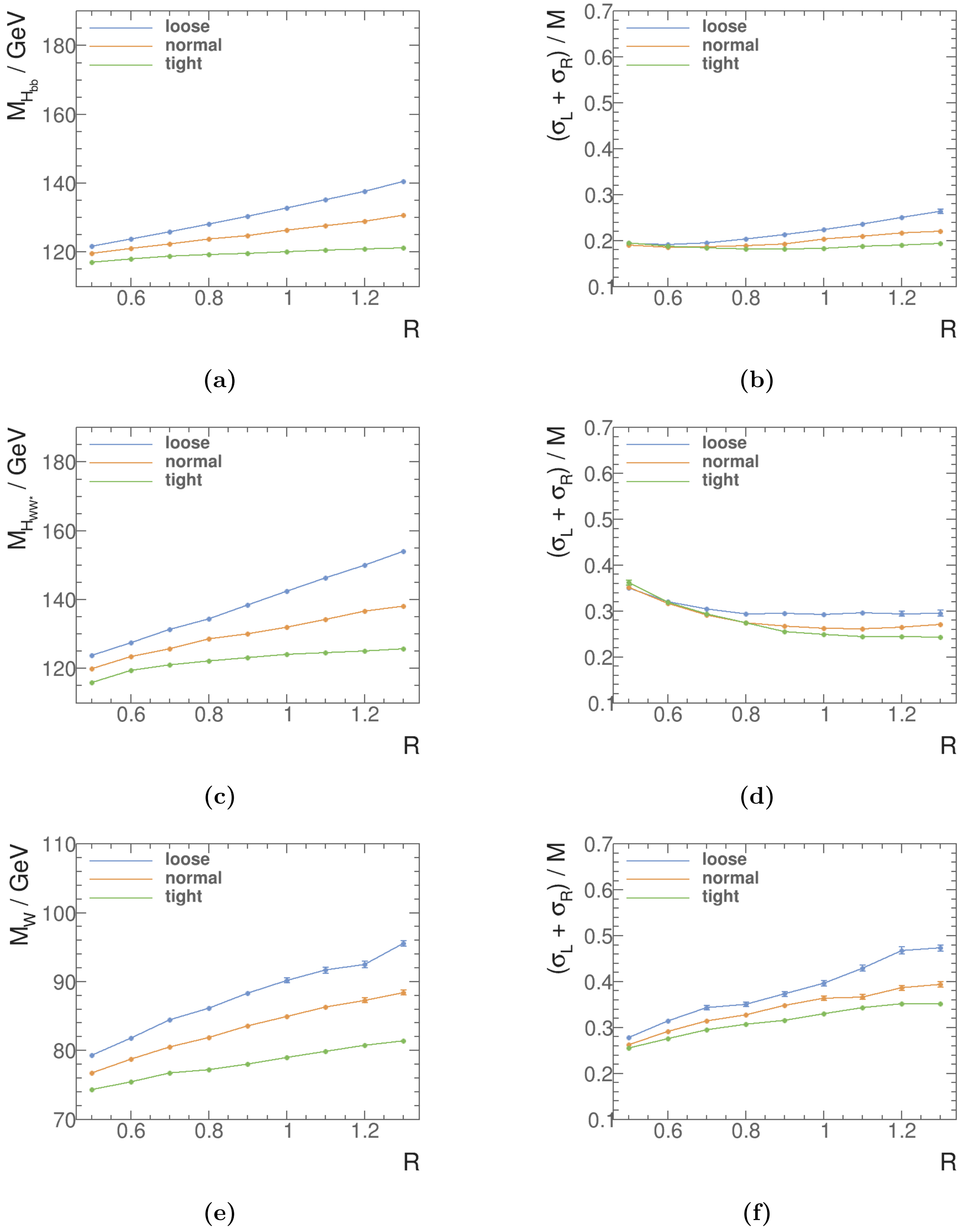


Figure 7.2: figure 7.2a, 7.2c, and 7.2e show fitted mass of H_{bb}, H_{WW^*}, and W, respectively, for loose, normal and tight selected PFO against R parameter, with $\sqrt{s} = 1.4\,\text{TeV}$. figure 7.2b, 7.2d, and 7.2f show fitted combined widths divided by the fitted masses of H_{bb}, H_{WW^*}, and W, respectively, for loose, normal and tight selected PFO against R parameter, with $\sqrt{s} = 1.4\,\text{TeV}$.

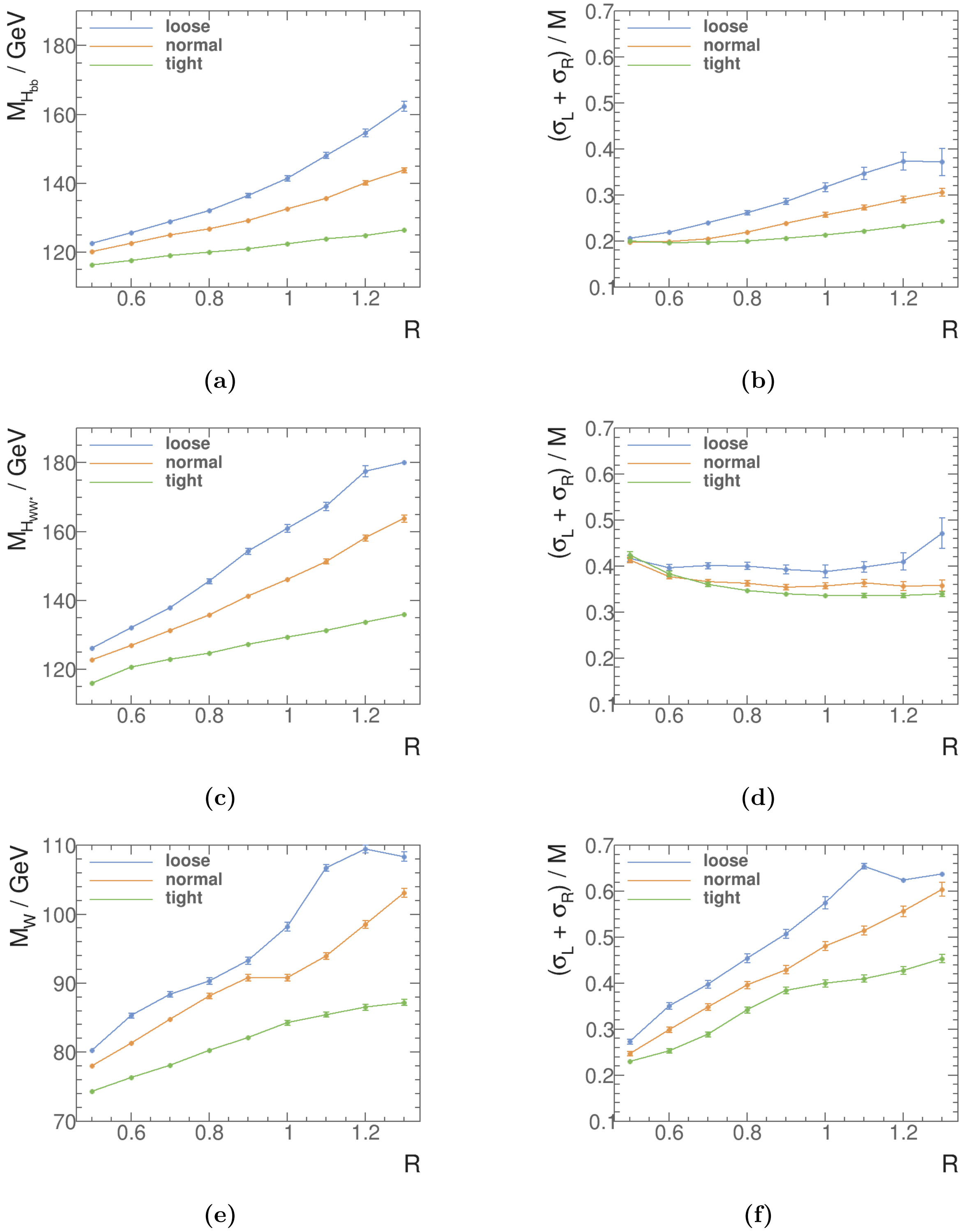


Figure 7.3: figure 7.3a, 7.3c, and 7.3e show fitted mass of H_{bb}, H_{WW^*}, and W, respectively, for loose, normal and tight selected PFO against R parameter, with $\sqrt{s} = 3\,\text{TeV}$. figure 7.3b, 7.3d, and 7.3f show fitted combined widths divided by the fitted masses of H_{bb}, H_{WW^*}, and W, respectively, for loose, normal and tight selected PFO against R parameter, with $\sqrt{s} = 3\,\text{TeV}$.

Jet Parameters	$\sqrt{s} = 1.4\,\text{TeV}$	$\sqrt{s} = 3\,\text{TeV}$
$\mu_{H_{bb}}$	$122.3_{\pm 0.2}$	$119.1_{\pm 0.3}$
$\sigma_{L,H_{bb}}$	$15.2_{\pm 0.2}$	$15.0_{\pm 0.3}$
$\sigma_{R,H_{bb}}$	$7.55_{\pm 0.16}$	$8.4_{\pm 0.2}$
$\mu_{H_{WW^*}}$	$125.7_{\pm 0.2}$	$123.0_{\pm 0.3}$
$\sigma_{L,H_{WW^*}}$	$29.4_{\pm 0.3}$	$36.6_{\pm 0.6}$
$\sigma_{R,H_{WW^*}}$	$7.18_{\pm 0.17}$	$7.4_{\pm 0.2}$
μ_{W}	$80.5_{\pm 0.2}$	$78.1_{\pm 0.3}$
$\sigma_{L,W}$	$16.2_{\pm 0.3}$	$13.1_{\pm 0.4}$
$\sigma_{R,W}$	$9.03_{\pm 0.16}$	$9.5_{\pm 0.2}$

Table 7.5: The extracted fitted parameters of optimal jet reconstructions, normal selected PFO with $R = 0.7$ for $\sqrt{s} = 1.4\,\text{TeV}$ and tight selected PFO with $R = 0.7$ for $\sqrt{s}$ $= 3\,\text{TeV}$.

7.4.2 Jet flavour tagging

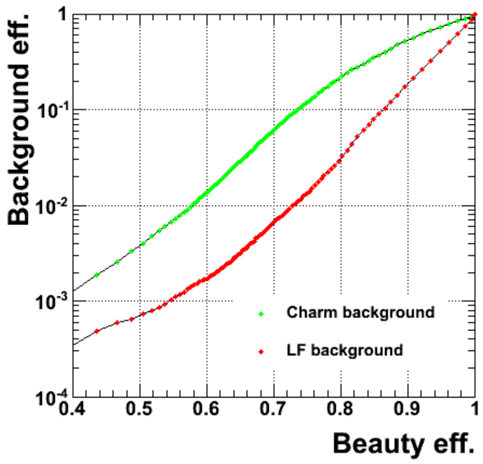
Two b-jets out of six jets in final states are identified with flavour tagging processors. The processor calculates a set of discriminatively variables for a jet. After training the MVA, the MVA is applied to jets to produce a likelihood for b-jet and c-jet. For details see section **??**.

The existing LCFIPlus processor in Marlin package is used. The training sample of the flavour tagging processor is $e^-e^+ \to Z\nu\overline{\nu}$, where Z decays to $q_l\overline{q}_l$, $b\overline{b}$, or $c\overline{c}$ at $\sqrt{s} =$ $1.4\,\text{TeV}$ and $\sqrt{s} = 3\,\text{TeV}$, because they have similar event topology as the signal, and they have only two jets in the final state.

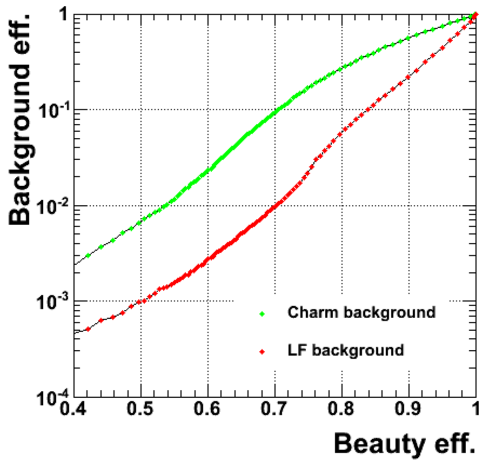
The selection efficiency of b-jets and c-jets with trainning samples are shown in figure **??**. Flavour tagging performs better at low energy. Because at high energy, particles at more collimated and more difficult to separate.

7.4.3 Jet pairing

The jet pairing was performed by checking combination of jets that are compatible with signal $HH \to b\overline{b}W^+W^-$.



(a) $\sqrt{s} = 1.4\,\mathrm{TeV}$



(b) $\sqrt{s} = 3\,\mathrm{TeV}$

Figure 7.4: Performance of b-jet tagging with trainning samples

The actual pairing is done via a minimisation

$$\chi^2 = \left(\frac{m_{ij} - \mu_{H_{bb}}}{\sigma'_{H_{bb}}}\right)^2 + \left(\frac{m_{klmn} - \mu_{H_{WW^*}}}{\sigma'_{H_{WW^*}}}\right)^2 + \left(\frac{m_{kl} - \mu_W}{\sigma'_W}\right)^2, \tag{7.5}$$

where, $\mu_{H_{bb}}$ and $\sigma'_{H_{bb}}$ are the fitted invariant mass, and the fitted width, respectively. Both are obtained in section 7.4.1. $\sigma'_{H_{bb}}$ is $\sigma_{L,H_{bb}}$ when $m_{ij} < m_{H_{bb}}$, and $\sigma_{R,H_{bb}}$ otherwise. Similarly $\mu_{H_{WW^*}}$ and μ_W are fitted mass, and $\sigma'_{H_{WW^*}}$ and σ'_W are fitted invariant mass, and the fitted width, respectively. Out of the six jets from the jet clustering, indicated by subscript i, j, k, l, m, n, two are used for H_{bb}, two for W and four for H_{WW^*}. The fitted parameters used are listed in table 7.5. Additional requirement is that at least one of two jets forming H_{bb} needs to have a b-jet tag of 0.2 or greater.

With the χ^2, all possible combinations are tested, and the one with smallest χ^2 is chosen.

7.5 Pre-selection

Discriminative variables were calculated. Some are used as to discard background events, whilst hurting the signal events a bit. This allows MVA to concentrate on events where it is difficult to separate in a single parameter space.

7.5.1 Discriminative pre-selection cuts

As discussed before, events with identified leptons are rejected. Jet pairing implies that events with the largest b-jet tag less than 0.2 are rejected.

For $\sqrt{s} = 1.4\,\text{TeV}$, a range of variables were tested and three where chosen as pre-selection cuts.

Event with invariant mass of two higgs less than 150 GeV is rejected. The cut above 120 GeV is needed as some background samples were generated only for invariant mass greater than 120 GeV. Shown in table **??** and figure **??**, this cut is effective against samples with two quark final states.

Event with second highest b-jet tag less than 0.2 is rejected. This stricter cut than the jet pairing helps to reduce samples with no b-jets.

Event with p_T of two higgs less than 30 GeV is rejected. This is extremely effective against samples with no neutrinos in the final state.

For $\sqrt{s} = 3\,\text{TeV}$, event with invariant mass of two higgs less than 150 GeV is rejected, for the reason similar to $\sqrt{s} = 1.4\,\text{TeV}$.

In addition, event with highest b-jet tag less than 0.7 is rejected. It is found that b-jet tag is less efficient at a higher $\sqrt{s}$. Therefore, a stricter cut at b-jet tag is useful to compensate for the tagging efficiency loss.

These set of cuts are stricter than usual analysis. The cross sections of signal channel for both $\sqrt{s}$ are extremely small, comparing to the background. Hence only the signal events with very clear characteristic topologies would be able to pass the final selection, in order to achieve a decent signal-to-background ratio. Therefore, a strict pre-selection cut would not hurt the final signal selection. On the contrary, final signal selection would benefit from MVA being able to focus the difficult background events, where their topologies are too similar to the signal events to separate in any single parameter space.

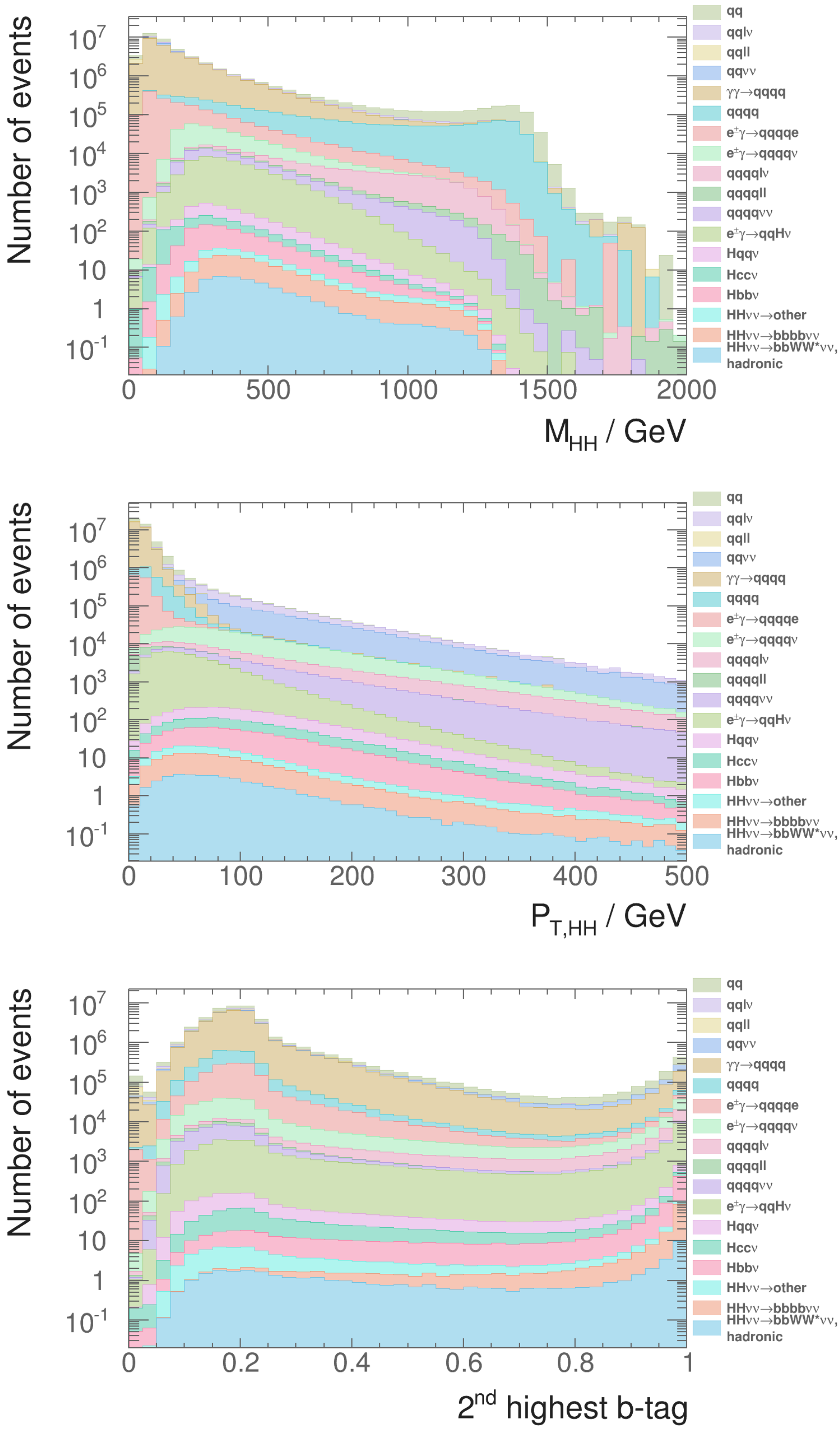


Figure 7.5: Discriminative pre-selection variables for $\sqrt{s} = 1.4\,\mathrm{TeV}$, after rejecting events with identified leptons, and jet pairing

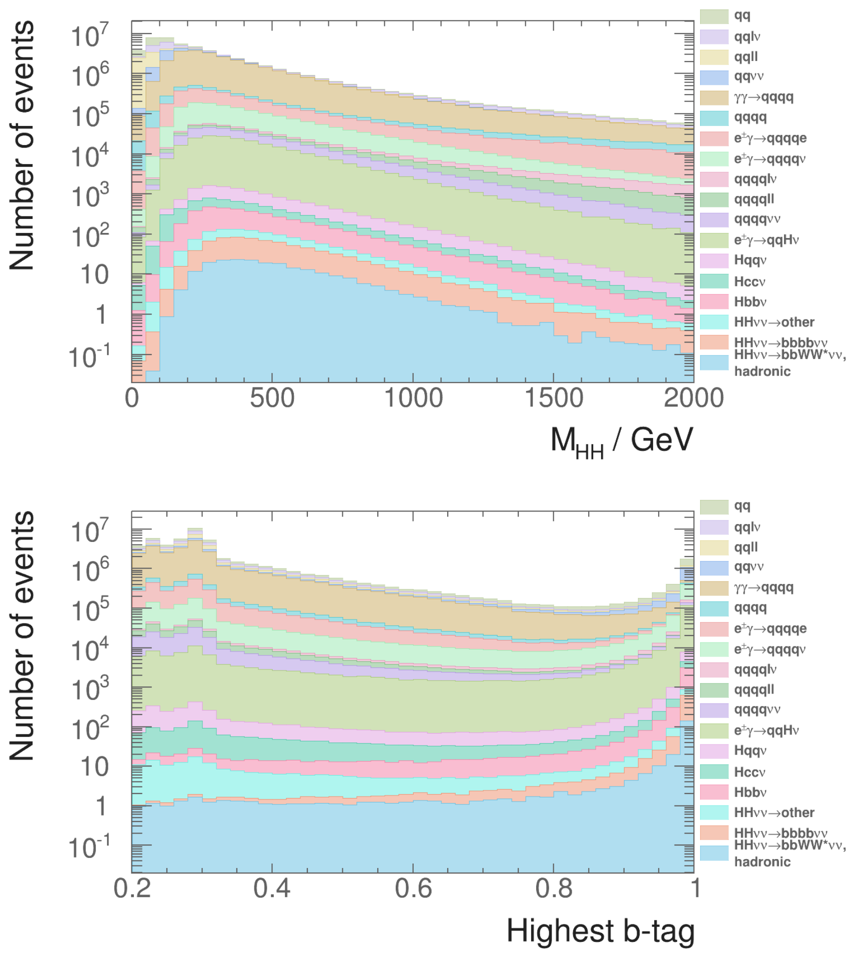


Figure 7.6: Discriminative pre-selection variables for $\sqrt{s} = 3\,\text{TeV}$, after rejecting events with identified leptons, and jet pairing

Channel / Efficiency $\sqrt{s} = 1.4$ TeV	Expected number of events	Lepton ID and jet pairing	$m_{HH} >$ 150GeV	$B_2 > 0.2$	$p_T >$ 30GeV
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, hadronic	27.9	85.8%	85.6%	73.7%	66.4%
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	67.6	90.8%	90.5%	90.1%	80.6%
$e^-e^+ \to HH\nu\bar{\nu} \to$ other	128.0	36.2%	35.3%	27.7%	24.7%
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	1304.0	60.7%	59.8%	44.9%	42.0%
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	546.1	67.4%	57.7%	46.5%	43.4%
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	463.0	73.9%	72.6%	68.7%	64.2%
$e^-e^+ \to qqqq$	1867650.0	48.8%	46.1%	17.3%	4.7%
$e^-e^+ \to qqqq\ell\ell$	93150.0	5.0%	4.9%	1.5%	0.3%
$e^-e^+ \to qqqq\ell\nu$	165600.0	15.1%	15.1%	12.4%	11.4%
$e^-e^+ \to qqqq\nu\bar{\nu}$	34800.0	50.7%	50.0%	20.1%	18.8%
$e^-e^+ \to qq$	6014250.0	54.5%	17.5%	8.4%	2.2%
$e^-e^+ \to qq\ell\nu$	6464550.0	14.1%	5.3%	2.0%	1.6%
$e^-e^+ \to qq\ell\ell$	4088700.0	13.0%	1.1%	0.6%	0.1%
$e^-e^+ \to qq\nu\nu$	1181550.0	60.1%	12.3%	6.2%	5.8%
$e^-\gamma(BS) \to e^-qqqq$	1305787.5	23.3%	10.6%	4.4%	0.4%
$e^+\gamma(BS) \to e^+qqqq$	1300837.5	23.4%	10.5%	4.3%	0.4%
$e^-\gamma(EPA) \to e^-qqqq$	430650.0	11.1%	5.4%	2.2%	0.3%
$e^+\gamma(EPA) \to e^+qqqq$	430350.0	11.1%	5.3%	2.1%	0.3%
$e^-\gamma(BS) \to \nu qqqq$	89775.0	58.3%	56.8%	31.0%	27.7%
$e^+\gamma(BS) \to \bar{\nu}qqqq$	89212.5	57.6%	56.1%	30.3%	27.3%
$e^-\gamma(EPA) \to \nu qqqq$	26100.0	29.6%	28.9%	15.4%	13.9%
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	25950.0	29.2%	28.5%	15.0%	13.7%
$e^-\gamma(BS) \to qqH\nu$	17775	61.0%	59.8%	45.5%	34.6%
$e^+\gamma(BS) \to qqH\nu$	17662.5	61.1%	60.0%	45.6%	34.6%
$e^-\gamma(EPA) \to qqH\nu$	5085	31.8%	31.2%	23.7%	18.2%
$e^+\gamma(EPA) \to qqH\nu$	5085	31.9%	31.3%	23.8%	18.4%
$\gamma(BS)\gamma(BS) \to qqqq$	2054951.5	56.3%	23.9%	9.6%	0.3%
$\gamma(BS)\gamma(EPA) \to qqqq$	4521037.5	33.6%	14.2%	5.7%	0.4%
$\gamma(EPA)\gamma(BS) \to qqqq$	4539150.0	33.7%	14.2%	5.7%	0.4%
$\gamma(EPA)\gamma(EPA) \to qqqq$	1129500.0	21.1%	9.1%	3.7%	0.4%

Table 7.6: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4$ TeV, assuming a luminosity of $1500\,\text{fb}^{-1}$. The selection efficiencies are presented in a "flow" fashion, as the every selection cut contains all the cuts to the left of it.

Channel / Efficiency $\sqrt{s} = 3\,\text{TeV}$	Expected number of events	Lepton ID and jet pairing	$m_{HH} > 150\mathrm{GeV}$	$B_1 > 0.7$
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, hadronic	146.0	80.2%	79.9%	69.7%
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	355.0	83.4%	82.9%	81.2%
$e^-e^+ \to HH\nu\bar{\nu} \to$ other	675.0	36.7%	35.8%	25.2%
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	6115.4	59.5%	58.5%	40.4%
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	2249.9	64.8%	58.4%	39.3%
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	2197.7	69.7%	68.4%	64.2%
$e^-e^+ \to qqqq$	1093000.0	48.5%	39.7%	3.0%
$e^-e^+ \to qqqq\ell\ell$	338600.0	14.7%	14.2%	0.7%
$e^-e^+ \to qqqq\ell\nu$	213200.0	19.7%	19.4%	10.0%
$e^-e^+ \to qqqq\nu\bar{\nu}$	143000.0	58.4%	57.3%	11.9%
$e^-e^+ \to qq$	5897800.0	62.8%	13.2%	2.7%
$e^-e^+ \to qq\ell\nu$	11121800	28.3%	11.9%	0.3%
$e^-e^+ \to qq\ell\ell$	6639200.0	38.3%	2.9%	0.7%
$e^-e^+ \to qq\nu\nu$	2635000.0	71.4%	24.1%	5.3%
$e^-\gamma(BS) \to e^-qqqq$	2004388.1	23.3%	21.5%	0.8%
$e^+\gamma(BS) \to e^+qqqq$	2002334.1	23.4%	21.6%	0.8%
$e^-\gamma(EPA) \to e^-qqqq$	575600.0	12.0%	11.0%	0.5%
$e^+\gamma(EPA) \to e^+qqqq$	575600.0	12.0%	10.9%	0.4%
$e^-\gamma(BS) \to \nu qqqq$	414750.0	61.7%	59.5%	20.4%
$e^+\gamma(BS) \to \bar{\nu}qqqq$	414434.0	61.2%	59.1%	19.4%
$e^-\gamma(EPA) \to \nu qqqq$	108400.0	30.9%	29.9%	9.6%
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	108400.0	30.7%	29.7%	9.1%
$e^-\gamma(BS) \to qqH\nu$	92588.0	58.3%	56.2%	37.3%
$e^+\gamma(BS) \to qqH\nu$	92430.0	58.1%	56.0%	37.1%
$e^-\gamma(EPA) \to qqH\nu$	23400.0	30.1%	29.2%	19.4%
$e^+\gamma(EPA) \to qqH\nu$	23400.0	29.7%	28.6%	18.8%
$\gamma(BS)\gamma(BS) \to qqqq$	18009413.9	54.2%	49.2%	1.9%
$\gamma(BS)\gamma(EPA) \to qqqq$	3824548.1	33.5%	30.2%	1.2%
$\gamma(EPA)\gamma(BS) \to qqqq$	3828498.1	33.7%	30.3%	1.2%
$\gamma(EPA)\gamma(EPA) \to qqqq$	805400.0	22.0%	19.8%	0.8%

Table 7.7: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 3\,\text{TeV}$, assuming a luminosity of $2000\mathrm{fb}^{-1}$. The selection efficiencies are presented in a "flow" fashion, as the every selection cut contains all the cuts to the left of it.

7.5.2 Sanity cuts

A set of very loose cuts, aiming to reduce the range of some discriminative variables to increase the effectiveness of MVA. (See section **??** on MVA) These cuts are very loose and physics motivated.

For $\sqrt{s} = 1.4\,\text{TeV}$, invariant masses for H_{bb}, H_{WW^*}, W, and HH are smaller than 500, 800, 200, and 1400 GeV, respectively.

For $\sqrt{s} = 3\,\text{TeV}$, invariant masses for H_{bb}, H_{WW^*}, W, and HH are smaller than 500, 800, 200, and 3000 GeV, respectively.

The selection efficiencies after sanity cuts and other pre-selection cuts stated above, are listed in table **??**.

7.5.3 Mutually exclusive cuts for $HH \to b\bar{b}W^+W^-$ and $HH \to b\bar{b}b\bar{b}$

Since the analysis for $e^-e^+ \to HH\nu\bar{\nu}$ channel is divided into two subchannels, $HH \to b\bar{b}W^+W^- \to b\bar{b}qqqq$ and $HH \to b\bar{b}b\bar{b}$, it is convenient to divided samples, both signal and background, into two mutually exclusive sets. This will make combining subchaneels much easier, as correlations between subchannels do not need to be considered.

The most distinctive difference between two subchannels, is that they have different number of jets, and different number of b-jets in the final state. So variables related to number of b-jets or a number of jets are suitable for separating two subchannels.

Shown in figure 7.7, two subchannels can be clearly separated in the two dimensional parameter space. The optimal rectangular cuts were selected by scanning the two parameters, and maximising

$$\varepsilon = P(\mathsf{subchannel}_1|\mathsf{selection}) \times P(\mathsf{subchannel}_2|\neg\mathsf{selection}) \tag{7.6}$$

where $\mathsf{selection}$ represents the mutually exclusive cuts, $\neg\mathsf{selection}$ indicates the phase space not covered by the $\mathsf{selection}$.

Variables tested includes $\Sigma B_{4\mathrm{jets}}$, $\sum_1^3 B_{4\mathrm{jets}}$, y_{34}, y_{45}, y_{56}, y_{67} and other related variables. The best separation was summarised in table 7.8.

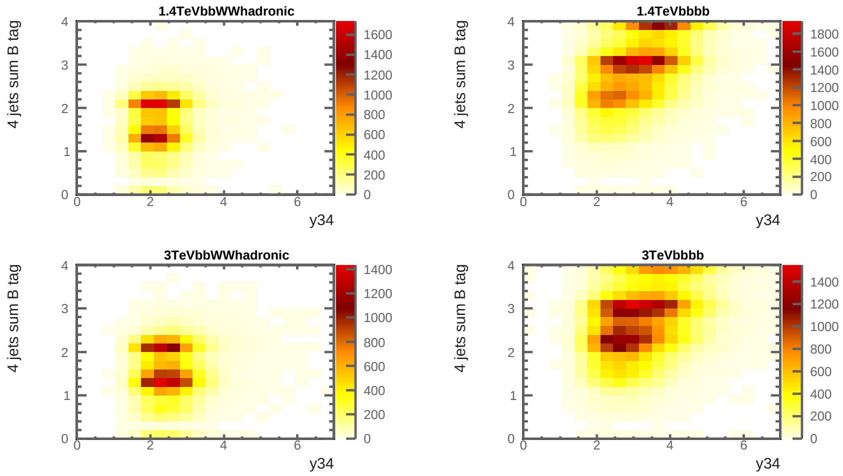


Figure 7.7: Sum of b tag against y_{34}, shown for signal samples

$\sqrt{s}$	selection	$HH \rightarrow b\bar{b}qqqq$ Selection Efficiency	$HH \rightarrow b\bar{b}b\bar{b}$ Selection Efficiency
1.4 TeV	$\Sigma B_{4jets} < 2.3$ and $y_{34} < 3.7$	86%	78%
3 TeV	$\Sigma B_{4jets} < 2.3$ and $y_{34} < 3.6$	89%	82%

Table 7.8: Mutually exclusive cuts, for full signal samples

The selection efficiencies after mutually exclusive cuts and other pre-selection cuts stated above, are listed in table ??.

7.6 Discriminative Variables

A series of discriminative variables were calculated, and fed into MVA for signal selection.

The full list of variables can be found in table ??. Same set of variables are used for $\sqrt{s} = 1.4$ TeV and $\sqrt{s} = 3$ TeV.

figure ?? hows the the variable XX which gives a good discrimination of signal against background.

The optimal set were chosen to give the best MVA performance, whilst no strong pair-wise correlation between any two variables, shown in figure ??.

Channel / Efficiency	Sanity $\sqrt{s} = 1.4\,\text{TeV}$	Mutually exclusive $\sqrt{s} = 1.4\,\text{TeV}$	Sanity $\sqrt{s} = 3\,\text{TeV}$	Mutually exclusive $\sqrt{s} = 3\,\text{TeV}$
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, hadronic	66.4%	59.7%	69.5%	61.7%
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	80.6%	15.4%	81.1%	18.8%
$e^-e^+ \to HH\nu\bar{\nu} \to$ other	24.7%	20.5%	25.1%	20.0%
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	42.0%	39.5%	40.3%	35.9%
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	43.4%	31.7%	39.2%	26.2%
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	64.2%	25.2%	64.2%	25.9%
$e^-e^+ \to qqqq$	4.6%	3.4%	2.5%	1.4%
$e^-e^+ \to qqqq\ell\ell$	3.3%	3.1%	0.7%	0.6%
$e^-e^+ \to qqqq\ell\nu$	11.4%.	9.8%	9.2%	7.2%
$e^-e^+ \to qqqq\nu\bar{\nu}$	18.8%	16.6%	11.8%	9.0%
$e^-e^+ \to qq$	2.0%	0.8%	2.5%	1.4%
$e^-e^+ \to qq\ell\nu$	1.6%	0.9%	0.3%	0.1%
$e^-e^+ \to qq\ell\ell$	0.1%	0.1%	0.7%	0.4%
$e^-e^+ \to qq\nu\nu$	5.8%	4.0%	5.3%	3.1%
$e^-\gamma(BS) \to e^-qqqq$	0.4%	0.3%	0.8%	0.7%
$e^+\gamma(BS) \to e^+qqqq$	0.4%	0.4%	0.8%	0.7%
$e^-\gamma(EPA) \to e^-qqqq$	0.3%	0.2%	0.4%	0.4%
$e^+\gamma(EPA) \to e^+qqqq$	0.3%	0.3%	0.4%	0.3%
$e^-\gamma(BS) \to \nu qqqq$	27.7%	25.3%	20.3%	16.8%
$e^+\gamma(BS) \to \bar{\nu}qqqq$	27.3%	24.9%	19.3%	15.9%
$e^-\gamma(EPA) \to \nu qqqq$	13.9%	12.6%	9.4%	7.8%
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	13.7%	12.3%	8.9%	7.3%
$e^-\gamma(BS) \to qqH\nu$	34.6%	30.6%	37.2%	30.2%
$e^+\gamma(BS) \to qqH\nu$	34.6%	30.6%	37.1%	30.2%
$e^-\gamma(EPA) \to qqH\nu$	18.2%	16.0%	19.0%	15.7%
$e^+\gamma(EPA) \to qqH\nu$	18.4%	16.1%	18.4%	15.2%
$\gamma(BS)\gamma(BS) \to qqqq$	0.3%	0.3%	1.9%	1.7%
$\gamma(BS)\gamma(EPA) \to qqqq$	0.4%	0.3%	1.1%	1.0%
$\gamma(EPA)\gamma(BS) \to qqqq$	0.4%	0.3%	1.1%	1.0%
$\gamma(EPA)\gamma(EPA) \to qqqq$	0.4%	0.3%	0.7%	0.6%

Table 7.9: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4\,\text{TeV}$ and $\sqrt{s} = 3\,\text{TeV}$, assuming a luminosity of 1500 and $2000\,\text{fb}^{-1}$, respectively. The selection efficiencies are presented in a “flow” fashion, as the every selection cut contains all the cuts to the left of it.

Variable	Description
$m_{H_{bb}}$	Invariant mass of H_{bb}
$m_{H_{WW^*}}$	Invariant mass of H_{WW^*}
m_W	Invariant mass of W
m_{HH}	Invariant mass of HH
E_{W^*}	Energy of W^*
E_{mis}	Missing energy, assuming collision at $\sqrt{s}$
$p_{TH_{bb}}$	Transverse momentum of H_{bb}
$p_{TH_{WW^*}}$	Transverse momentum of H_{WW^*}
p_{THH}	Transverse momentum of HH
η_{mis}	Pseudorapidity of missing momentum, assuming collision at $\sqrt{s}$
p_{THH}	Transverse momentum of HH
$-\ln(y_{23})$	minus ln of y_{23}. See section ?? for y parameter. See section ?? for the ln transformation
$-\ln(y_{34})$	minus ln of y_{34}.
$-\ln(y_{45})$	minus ln of y_{45}.
$-\ln(y_{56})$	minus ln of y_{56}.
$B_{1,H_{bb}}$	Highest b-jet tag value of two jets forming H_{bb}.
$B_{2,H_{bb}}$	Lowest b-jet tag value of two jets forming H_{bb}.
$B_{1,W}$	Highest b-jet tag value of two jets forming W.
B_{1,W^*}	Highest b-jet tag value of two jets forming W^* .
$C_{1,H_{bb}}$	Highest c-jet tag value of two jets forming H_{bb}.
$C_{1,W}$	Highest c-jet tag value of two jets forming W.
$\lvert\mathbf{S}\rvert$	Modulus of sphericity, **S**. See section ??.
$acol_{H_{bb}}$	Acolinearity of two jets forming H_{bb}.
$acol_W$	Acolinearity of two jets forming W.
$acol_{HH}$	Acolinearity of H_{bb} and H_{WW^*}.
$N_{H_{bb}}$	Number of PFOs forming H_{bb}.
$N_{H_{WW^*}}$	Number of PFOs forming H_{WW^*}.
N_W	Number of PFOs forming W.
N_{W^*}	Number of PFOs forming W^*.
$\cos\left(\theta^*_{H_{bb}}\right)$	Cosine of opening angles of two jets forming H_{bb}, in their rest frame.
$\cos\left(\theta^*_{H_{WW^*}}\right)$	Cosine of opening angles of W and W^*, forming H_{WW^*}, in their rest frame.
$\cos\left(\theta^*_{W}\right)$	Cosine of opening angles of two jets forming W, in their rest frame.
	Cosine of opening angles of two jets forming W^*, in their

7.7 Multivariate analysis

Multivariate analysis was performed with TMVA package. The classifier that performs the best was found to be the boosted decision tree. See section **??** for details on boosted decision tree.

The parameters for boosted decision tree were optimised and checked for overtraining. The most important variables are the depth of the tree and the number of trees. Other parameters includes the minimum number of nodes in a leaf, the number of cuts of a variable, the learning rate, the sampling fraction, the yes/no or purity leaf, adaBoost or gradient boost.

The optimisation and overtraining test was done with $\sqrt{s} = 3\,\text{TeV}$ samples. $\sqrt{s} = 1.4\,\text{TeV}$ samples produce similar results.

Half of the samples were used for training, and the other half used for testing.

7.8 Signal selection results

7.9 Couplings extration

Channel / Efficiency $\sqrt{s} = 1.4\,\mathrm{TeV}$	Expected number of events	Pre-selection efficiency	MVA efficiency	Number of events after MVA
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, hadronic	27.9	59.8%	8.2%	1.37
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	67.6	15.4%	0.5%	0.05
$e^-e^+ \to HH\nu\bar{\nu} \to$ other	128.0	20.4%	1.7%	0.45
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	1304.0	39.5%	0.05%	0.29
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	546.1	31.6%	0.1%	0.16
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	463.0	24.7%	0.3%	0.37
$e^-e^+ \to qqqq$	1867650.0	3.3%	-	-
$e^-e^+ \to qqqq\ell\ell$	93150.0	0.3%	-	-
$e^-e^+ \to qqqq\ell\nu$	165600.0	9.8%	0.01%	2.06
$e^-e^+ \to qqqq\nu\bar{\nu}$	34800.0	16.5%	0.002%	0.10
$e^-e^+ \to qq$	6014250.0	0.8%	-	-
$e^-e^+ \to qq\ell\nu$	6464550.0	0.9%	-	-
$e^-e^+ \to qq\ell\ell$	4088700.0	0.08%	-	-
$e^-e^+ \to qq\nu\nu$	1181550.0	4.0%	-	-
$e^-\gamma(BS) \to e^-qqqq$	1305787.5	0.3%	-	-
$e^+\gamma(BS) \to e^+qqqq$	1300837.5	0.4%	-	-
$e^-\gamma(EPA) \to e^-qqqq$	430650.0	0.3%	-	-
$e^+\gamma(EPA) \to e^+qqqq$	430350.0	0.3%	-	-
$e^-\gamma(BS) \to \nu qqqq$	89775.0	25.4%	0.005%	1.09
$e^+\gamma(BS) \to \bar{\nu}qqqq$	89212.5	24.9%	0.004%	0.96
$e^-\gamma(EPA) \to \nu qqqq$	26100.0	12.6%	-	-
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	25950.0	12.4%	0.008%	0.27
$e^-\gamma(BS) \to qqH\nu$	17775	30.8%	0.02%	1.00
$e^+\gamma(BS) \to qqH\nu$	17662.5	30.6%	0.02%	1.16
$e^-\gamma(EPA) \to qqH\nu$	5085	16.0%	0.04%	0.33
$e^+\gamma(EPA) \to qqH\nu$	5085	16.2%	0.08%	0.62
$\gamma(BS)\gamma(BS) \to qqqq$	2054951.5	0.2%	-	-
$\gamma(BS)\gamma(EPA) \to qqqq$	4521037.5	0.4%	-	-
$\gamma(EPA)\gamma(BS) \to qqqq$	4539150.0	0.3%	-	-
$\gamma(EPA)\gamma(EPA) \to qqqq$	1129500.0	0.3%	-	-

Table 7.11: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4\,\mathrm{TeV}$, for a luminosity of $1500\,\mathrm{fb}^{-1}$. The number of events, selection efficiency of pre-selection, selection efficiency of MVA after pre-selection, number of events after MVA are shown. - represents no events passing the MVA.

Channel / Efficiency $\sqrt{s} = 3\,\text{TeV}$	Expected number of events	Pre-selection efficiency	MVA efficiency	Number of events after MVA
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, hadronic	146.0	61.7%	11.6%	10.43
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	355.0	18.8%	1.5%	1.01
$e^-e^+ \to HH\nu\bar{\nu} \to$ other	675.0	20.0%	3.6%	4.93
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	6115.4	36.0%	0.4%	9.42
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	2249.9	26.3%	0.5%	3.13
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	2197.7	25.8%	1.2%	6.82
$e^-e^+ \to qqqq$	1093000.0	1.4%	0.01%	1.43
$e^-e^+ \to qqqq\ell\ell$	338600.0	0.6%	-	-
$e^-e^+ \to qqqq\ell\nu$	213200.0	7.3%	0.05%	8.35
$e^-e^+ \to qqqq\nu\bar{\nu}$	143000.0	9.0%	0.05%	6.35
$e^-e^+ \to qq$	5897800.0	1.4%	-	-
$e^-e^+ \to qq\ell\nu$	11121800	0.1%	-	-
$e^-e^+ \to qq\ell\ell$	6639200.0	0.4%	-	-
$e^-e^+ \to qq\nu\nu$	2635000.0	3.1%	-	-
$e^-\gamma(BS) \to e^-qqqq$	2004388.1	0.7%	-	-
$e^+\gamma(BS) \to e^+qqqq$	2002334.1	0.7%	-	-
$e^-\gamma(EPA) \to e^-qqqq$	575600.0	0.4%	-	-
$e^+\gamma(EPA) \to e^+qqqq$	575600.0	0.3%	-	-
$e^-\gamma(BS) \to \nu qqqq$	414750.0	16.8%	0.04%	30.7
$e^+\gamma(BS) \to \bar{\nu}qqqq$	414434.0	15.9%	0.05%	30.3
$e^-\gamma(EPA) \to \nu qqqq$	108400.0	7.8%	0.04%	3.37
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	108400.0	7.3%	0.03%	2.63
$e^-\gamma(BS) \to qqH\nu$	92588.0	30.2%	0.2%	67.5
$e^+\gamma(BS) \to qqH\nu$	92430.0	30.3%	0.2%	54.2
$e^-\gamma(EPA) \to qqH\nu$	23400.0	15.4%	0.2%	7.88
$e^+\gamma(EPA) \to qqH\nu$	23400.0	15.2%	0.3%	10.2
$\gamma(BS)\gamma(BS) \to qqqq$	18009413.9	1.6%	-	-
$\gamma(BS)\gamma(EPA) \to qqqq$	3824548.1	1.0%	-	-
$\gamma(EPA)\gamma(BS) \to qqqq$	3828498.1	1.0%	-	-
$\gamma(EPA)\gamma(EPA) \to qqqq$	805400.0	0.6%	-	-

Table 7.12: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 3\,\text{TeV}$, for a luminosity of $2000\,\text{fb}^{-1}$. The number of events, selection efficiency of pre-selection, selection efficiency of MVA after pre-selection, number of events after MVA are shown. - represents no events passing the MVA.

Channel / Efficiency $\sqrt{s} = 3\,\text{TeV}$	Expected number of events	Pre-selection efficiency	MVA efficiency	Number of events after MVA
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, semi-leptonic	96.8	44.6%	21.9%	9.48
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	355.0	13.3%	10.9%	5.16
$e^-e^+ \to HH\nu\bar{\nu} \to$ other	724.2	13.1%	13.6%	12.89
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	6115.4	7.4%	13.7%	62.63
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	2249.9	6.3%	12.1%	17.10
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	2197.7	15.9%	5.1%	18.03
$e^-e^+ \to qqqq$	1093000.0	0.6%	0.2%	15.04
$e^-e^+ \to qqqq\ell\ell$	338600.0	1.0%	0.06%	1.85
$e^-e^+ \to qqqq\ell\nu$	213200.0	27.6%	0.5%	270.33
$e^-e^+ \to qqqq\nu\bar{\nu}$	143000.0	1.9%	1.6%	43.78
$e^-e^+ \to qq$	5897800.0	0.4%	0.3%	60.82
$e^-e^+ \to qq\ell\nu$	11121800	0.3%	0.08%	21.24
$e^-e^+ \to qq\ell\ell$	6639200.0	0.6%	0.2%	84.14
$e^-e^+ \to qq\nu\nu$	2635000.0	0.4%	0.9%	92.55
$e^-\gamma(BS) \to e^-qqqq$	2004388.1	1.2%	-	-
$e^+\gamma(BS) \to e^+qqqq$	2002334.1	1.2%	-	-
$e^-\gamma(EPA) \to e^-qqqq$	575600.0	1.1%	-	-
$e^+\gamma(EPA) \to e^+qqqq$	575600.0	1.1%	-	-
$e^-\gamma(BS) \to \nu qqqq$	414750.0	3.7%	1.5%	226.77
$e^+\gamma(BS) \to \bar{\nu}qqqq$	414434.0	3.5%	1.6%	225.68
$e^-\gamma(EPA) \to \nu qqqq$	108400.0	11.2%	0.9%	107.90
$e^+\gamma(EPA) \to \bar{\nu}qqqq$	108400.0	10.7%	0.8%	92.75
$e^-\gamma(BS) \to qqH\nu$	92588.0	7.9%	10.7%	779.36
$e^+\gamma(BS) \to qqH\nu$	92430.0	7.9%	10.1%	741.57
$e^-\gamma(EPA) \to qqH\nu$	23400.0	22.9%	6.9%	369.52
$e^+\gamma(EPA) \to qqH\nu$	23400.0	22.7%	7.2%	381.33
$\gamma(BS)\gamma(BS) \to qqqq$	18009413.9	0.4%	-	-
$\gamma(BS)\gamma(EPA) \to qqqq$	3824548.1	1.0%	-	-
$\gamma(EPA)\gamma(BS) \to qqqq$	3828498.1	1.0%	0.08%	28.85
$\gamma(EPA)\gamma(EPA) \to qqqq$	805400.0	1.1%	-	-

Table 7.13: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 3\,\text{TeV}$, for a luminosity of $2000\,\text{fb}^{-1}$. The number of events, selection efficiency of pre-selection, selection efficiency of MVA after pre-selection, number of events after MVA are shown. - represents no events passing the MVA.

Colophon

This thesis was made in $\LaTeX\,2_\varepsilon$ using the “hepthesis” class [46].

Bibliography

[1] ATLAS Collaboration, G. Aad *et al.*, Phys.Lett. **B716**, 1 (2012), 1207.7214.

[2] Particle Data Group, K. A. Olive *et al.*, Chin. Phys. **C38**, 090001 (2014).

[3] M. Thomson, *Modern particle physics* (Cambridge University Press, New York, 2013).

[4] D. Tong, Lectures on quantum field theory, 2006.

[5] B. Gripaios, Lectures on gauge field theory, 2017.

[6] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3, S. Schael *et al.*, Phys. Rept. **427**, 257 (2006), hep-ex/0509008.

[7] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[8] D. Rainwater, Searching for the Higgs boson, in *Proceedings of Theoretical Advanced Study Institute in Elementary Particle Physics : Exploring New Frontiers Using Colliders and Neutrinos (TASI 2006): Boulder, Colorado, June 4-30, 2006*, pp. 435–536, 2007, hep-ph/0702124.

[9] D. B. Kaplan and H. Georgi, Phys. Lett. **B136**, 183 (1984).

[10] W. D. Goldberger, B. Grinstein, and W. Skiba, Phys. Rev. Lett. **100**, 111802 (2008), 0708.1463.

[11] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, JHEP **06**, 045 (2007), hep-ph/0703164.

[12] R. Contino, C. Grojean, M. Moretti, F. Piccinini, and R. Rattazzi, JHEP **05**, 089 (2010), 1002.1011.

[13] R. Contino, C. Grojean, D. Pappadopulo, R. Rattazzi, and A. Thamm, JHEP **02**,

006 (2014), 1309.7038.

[14] V. Barger, T. Han, P. Langacker, B. McElrath, and P. Zerwas, Phys. Rev. **D67**, 115001 (2003), hep-ph/0301097.

[15] H. Abramowicz *et al.*, (2016), 1608.07538.

[16] CMS Collaboration, S. Chatrchyan *et al.*, Phys.Lett. **B716**, 30 (2012), 1207.7235.

[17] J. Brau *et al.*, (2007).

[18] L. Linssen, A. Miyamoto, M. Stanitzki, and H. Weerts, (2012), 1202.5940.

[19] H. Baer *et al.*, (2013), 1306.6352.

[20] M. Aicheler *et al.*, (2012).

[21] M. Thomson, Nucl.Instrum.Meth. **A611**, 25 (2009), 0907.3577.

[22] J. S. Marshall, A. MÍźnnich, and M. A. Thomson, Nucl. Instrum. Meth. **A700**, 153 (2013), 1209.4039.

[23] I. G. Knowles and G. D. Lafferty, J. Phys. **G23**, 731 (1997), hep-ph/9705217.

[24] M. Green, *Electron-Positron Physics at the Z*Studies in high energy physics, cosmology, and gravitation (Taylor & Francis, 1998).

[25] CALICE, C. Adloff, (2011), 1105.0511.

[26] B. Parker *et al.*, (2009).

[27] CALICE, JINST **7**, P04015 (2012), 1201.1653.

[28] A. Sailer, *Radiation and Background Levels in a CLIC Detector Due to Beam-beam Effects: Optimisation of Detector Geometries and Technologies* (Humboldt Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, 2012).

[29] W. Kilian, T. Ohl, and J. Reuter, European Physical Journal C **71** (2011).

[30] M. Moretti, T. Ohl, and J. Reuter, p. 1981 (2001), hep-ph/0102195.

[31] T. Sjostrand, (1995), hep-ph/9508391.

[32] OPAL, G. Alexander *et al.*, Z. Phys. **C69**, 543 (1996).

[33] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, Comput. Phys. Commun. **76**, 361

(1993).

[34] D. Schulte, (1999).

[35] GEANT4, S. Agostinelli *et al.*, Nucl.Instrum.Meth. **A506**, 250 (2003).

[36] P. Mora de Freitas and H. Videau, p. 623 (2002).

[37] F. Gaede, Nucl. Instrum. Meth. **A559**, 177 (2006).

[38] J. S. Marshall and M. A. Thomson, Eur. Phys. J. **C75**, 439 (2015), 1506.05348.

[39] J. S. Marshall, Presentation on pandorapfa with lc reconstruction, https://github.com/PandoraPFA/Documentation/blob/master/Pandora_LC_Reconstruction.pdf, 2017.

[40] F. Gaede and J. Engels, EUDET Report (2007).

[41] B. Xu, Improvement of photon reconstruction in PandoraPFA, in *Proceedings, International Workshop on Future Linear Colliders (LCWS15): Whistler, B.C., Canada, November 02-06, 2015*, 2016, 1603.00013.

[42] E. Farhi, Phys. Rev. Lett. **39**, 1587 (1977).

[43] TMVA Core Developer Team, J. Therhaag, AIP Conf.Proc. **1504**, 1013 (2009).

[44] S. Dittmaier *et al.*, (2012), 1201.3084.

[45] A. MÍźnnich, CERN Report No. LCD-Note-2010-009, 2010 (unpublished).

[46] A. Buckley, The hepthesis LaTeX class.

List of figures

4.1 Illustration of the cone clustering algorithm, taken from [39] 32

5.1 Two 500 GeV photons (yellow and blue), just resolved in the transverse plane perpendicular to the direction of the flight, of their energy deposition in electromagnetic calorimeter. U and V axis are two arbitrary axis perpendicular to each other in the plane. Z axis is the sum of the calorimeter hit energy in each particular bin in 2D plane in GeV. 39

5.3 An event display of a typical 500 GeV photon, reconstructed into a main photon in the ECAL (yellow) and a neutral hadron fragment in the HCAL (blue). 48

7.1 Example MC mass fit for double higgs analysis 82

7.2 Fitted mass, and resolution of H_{bb}, H_{WW^*} and W for $\sqrt{s} = 1.4\,\mathrm{TeV}$ . . . 83

7.3 Fitted mass, and resolution of H_{bb}, H_{WW^*} and W for $\sqrt{s} = 3\,\mathrm{TeV}$ 84

7.4 Performance of b-jet tagging with trainning samples 86

7.5 Discriminative pre-selection variables for $\sqrt{s} = 1.4\,\mathrm{TeV}$ 88

7.6 Discriminative pre-selection variables for $\sqrt{s} = 3\,\mathrm{TeV}$ 89

7.7 Sum of b tag against y_{34} 93

List of tables

6.1 Branching ratios of the seven major τ^- decays, taken from [2]. τ^+ decays similarly to τ^-. 59

7.3 isolated lepton finder processors performance on the signal and selected background samples. 78

7.4 Very forward electron and photon finder performance on the signal and selected background samples. 80

7.5 The extracted fitted parameters of optimal jet reconstructions 85

7.6 List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4\,\text{TeV}$, assuming a luminosity of 1500fb^{-1}. The selection efficiencies are presented in a "flow" fashion, as the every selection cut contains all the cuts to the left of it. 90

7.7 List of signal and background samples with the corresponding expected number at $\sqrt{s} = 3\,\text{TeV}$, assuming a luminosity of 2000fb^{-1}. The selection efficiencies are presented in a "flow" fashion, as the every selection cut contains all the cuts to the left of it. 91

7.8 Mutually exclusive cuts 93

7.9 List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4\,\text{TeV}$ and $\sqrt{s} = 3\,\text{TeV}$, assuming a luminosity of 1500 and 2000fb^{-1}, respectively. The selection efficiencies are presented in a "flow" fashion, as the every selection cut contains all the cuts to the left of it. 94

7.10 List of variables used in MVA 95

7.11 List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4\,\text{TeV}$, for a luminosity of $1500\mathtt{fb}^{-1}$. The number of events, selection efficiency of pre-selection, selection efficiency of MVA after pre-selection, number of events after MVA are shown. - represents no events passing the MVA. 97

7.12 List of signal and background samples with the corresponding expected number at $\sqrt{s} = 3\,\text{TeV}$, for a luminosity of $2000\mathtt{fb}^{-1}$. The number of events, selection efficiency of pre-selection, selection efficiency of MVA after pre-selection, number of events after MVA are shown. - represents no events passing the MVA. 98

7.13 List of signal and background samples with the corresponding expected number at $\sqrt{s} = 3\,\text{TeV}$, for a luminosity of $2000\mathtt{fb}^{-1}$. The number of events, selection efficiency of pre-selection, selection efficiency of MVA after pre-selection, number of events after MVA are shown. - represents no events passing the MVA. 99