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

“*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

Chapter 2

Detector

“ILC will be built next year”
— Mysterious person

overall

ILC

CLIC

calorimeter

ECal

HCal

Muon chamber

Forward detector

Tracker

CLCI-background

Chapter 3

Reconstruction

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

As previously stated, this document focus on the ILD and CLIC_ILD detectors. Due to the similarity, we often only discuss one detector to avoid the repetition. The difference in detectors will be stated if applicable.

3.1 Reconstruction overall

Reconstruction software runs in Marlin software framework [?], part of iLCSoft.

The reconstruction of a event into physical objects containing a few main steps: digitisation of calorimeter hits (simulated hits in this case), reconstruction of tracks in the tracking system using pattern recognition algorithms, and particle flow objects reconstruction with PandoraPFA. For the CLIC detectors, there is an extra step of $\gamma\gamma \rightarrow hadrons$ background suppression. We will discuss PandoraPFA citeThomson:2009rp,Marshall:2012ry,Marshall:2015rfa in details, where a lot work goes into, and the background suppression due to its relevance in later analysis.

3.2 Pandora

Inputs of PandoraPFA are digitised calorimeter hits and reconstructed tracks.

3.2.1 Track selection

Tracks are selected based on their topological properties, how likely they are from physical processes, and whether they are consistent with tracker resolution. Only selected tracks will be used for the subsequent reconstruction.

Using a helical fit of last 50 reconstructed hits, tracks are projected to the front of the ECal.

3.2.2 Calorimeter selection

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.

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 particle flow object (PFO) creation step.

3.2.3 Clustering

The main clustering scheme used in PandoraPFA is cone clustering, for grouping calorimeter hits. 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.

Typically a high energy calorimeter hit will be chosen 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. PandoraPFA will start to use projection of tracks at the ECal as seeds. When all tracks are used, the remaining hits will be used as seeds.

These cone clustering algorithms are widely used in the calorimeter in PandoraPFA, and they produce basic working objects, Clusters.

There are two standalone particle identification algorithms in PandoraPFA, muon identification and photon identification. Identified muons and photons will not participate in the clustering and re-clustering stages. Both algorithms aim to improve the clustering and the re-clustering. The photon identification and related algorithm will be discussed in details in chapter ??.

3.2.4 Clusters Merging

Initial clustering scheme is aggressive at splitting clusters. The next step is to merge clusters base on clear topological signatures. For clusters associated to tracks (charged clusters) and clusters not associated to tracks(neutral clusters), track like segments in the calorimeter are identified.

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.

3.2.5 Re-clustering

The clustering and cluster merging 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. herefore, it is important to re-cluster base on the compatibility of the cluster energy and the associated track momentum. A cluster may be split into two, or 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.

3.2.6 Photon identification

The neutral clusters are tested against an expected photon electromagnetic shower profile. The longitudinal shower profile for a photon cluster is required to be similar to a expected electromagnetic shower profile, with the discrepancy being smaller than a threshold.

3.2.7 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. This will be described in details in chapter ??.

3.2.8 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 ??.

3.3 Suppression of $\gamma\gamma \rightarrow \mathrm{hadrons}$ backround

For the CLIC, as discussed in section ??, significant $\gamma\gamma \rightarrow \mathrm{hadrons}$ 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 [?].

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 $\gamma\gamma \rightarrow \mathrm{hadrons}$ 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 $\gamma\gamma \rightarrow \mathrm{hadrons}$ 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 4

Analysis technique

"In preparing for battle I have always found that plans are useless, but planning is indispensable."
— Dwight D. Eisenhower

Automated analysis is the only way to deal with the vast amount of data generated in the high energy physics. In the last chapter we described the automated reconstruction tools in details. This chapter is dedicated to the common automated analysis tools and techniques, which will be used in the analysis described in subsequent chapters.

For the linear collider, thanks to the high granular calorimeter, the starting point for analysis would be individual Particle Flow Objects, as well as individual tracks. Each of the PFOs encodes four-momentum and position information. For tracks, they would have momentum and position information.

However, sometimes it is interesting to group PFOs and tracks into jets, where a jet is the result of hadronisation process from high energy particles like quarks or gulons.

4.1 Jet algorithm

A jet is typically a visually obvious structure in a event display. The momentum and the direction of a jet tend to resemble the originated particle. Despite the relative easiness of identifying jets visually, it presents a challenge for a pattern recognition program to identify jets effectively and efficiently.

Early work on jet finding started in 1977 [1], where later development can be found in reviews [2–4].

There are two large families of jet finding algorithm, cone based algorithm, and sequential combination algorithm. Cone based algorithm is briefly discussed in section 3.2.3.

Sequential combination algorithm typically calculate a pair-wise distance metric. Pairs with the smallest metric will be combined. The metric will be calculated and updated, and a pair with smallest metric will be combined. This procedure will be repeated until some stopping criterion are satisfied.

The chosen jet algorithm implementation is FastJet C++ software package [5, 6], providing a wide range of jet finding algorithms. The implementation in Marlin software package is called MarlinFastJet. The symbols in the subsequent discussion about specific jet algorithms will follow [5]

4.1.1 k_t algorithm

One of the common sequential combination algorithms for p $\overline{\text{p}}$ collider experiment, is longitudinally-invariant k_t algorithm [7, 8]. In the inclusive variant, The symmetrical pair-wise distance metric between particle i and j, and the beam distance, are defined as

$$d_{ij} = d_{ji} = \min\left({p_T}_i^2, {p_T}_j^2\right) \frac{\Delta R_{ij}^2}{R^2}, \tag{4.1}$$

$$d_{iB} = {p_T}_i^2, \tag{4.2}$$

where ${p_T}_i$ is the transverse momentum of particle i with respect to the beam (z) direction, and ΔR_{ij}^2 is the measurement of angular separation of particle i and j. Formally $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where $y_i = \frac{1}{2} \ln \frac{E_i + {p_z}_i}{E_i - {p_z}_i}$ and ϕ_i are particle i's rapidity and azimuthal angle. R is a free parameter controlling the jet radius.

If $d_{ij} < d_{iB}$, particle i and j are merged, with the four-momentum of particle i updated as the sum. Otherwise, particle i is set to be a final jet, and delete from the particle list. The above procedure is repeated until no particle left.

The exclusive variant is similar. First difference is that when $d_{iB} < d_{ij}$, the particle i is discarded and part of the beam jet. The second difference is that when both d_{ij} and d_{iB} are above some threshold, d_{cut}, the clustering will stop. In practise, exclusive mode allows a specified number of jets to be found, which will automatically choose the d_{cut}. The inclusive mode would fine as many jets as the algorithm allows.

4.1.2 Durham algorithm

Durham algorithm [9], also known as e⁻ e⁺ k_t algorithm, is commonly used e⁻ e⁺ collider experiment. It has a single distance metric:

$$d_{ij} = 2\min\left(E_i^2, E_j^2\right)(1 - \cos(\theta_{ij})), \tag{4.3}$$

where E_i is the energy of particle i. θ_{ij} is the polar angle difference between particle i and j. Durham algorithm can only be run at exclusive mode, which means that the clustering will stop when d_{ij} is above some threshold, d_{cut}.

Comparing to k_t algorithm, it uses energy instead of p_T in the distance metric, and it did not have a beam jet. This is because that for the e⁻ e⁺ collider in the past, the beam induced background was not severe and collisions energy is known, $\sqrt{s}$.

4.1.3 Jet algorithm for CLIC

Although CLIC is a e⁻ e⁺ collider, the significant beam-induced background adds a large amount of energy from $\gamma\gamma \to hadrons$ process. Therefore, traditional e⁻ e⁺ jet algorithms, like Durham algorithm, is not suitable for CLIC environment. Studies has shown that jet algorithms for p $\overline{p}$ collider have better performance [10, 11].

A more recent attempt at marrying merits from both Durham and k_t algorithms has resulted in Valencia jet algorithm [12]. It had shown promising improvement comparing to k_t algorithm.

4.1.4 y parameter

y parameter is a commonly used quantity to describe the transition of exclusive jet algorithm going from N clustered jets to $N+1$ clustered jets. For example, y_{23} would be the d_{cut} value for a exclusive jet algorithm, above which the jet algorithm returns 2 jets, below which the jet algorithm returns 3 jets.

Numerically y parameter is often much smaller than one. A typically way to convert the small number to a human acceptable range is to take the minus logarithm of the number.

4.2 Flavour tagging

The latest software package for jet flavour tagging is LCFIPlus [13]. It is based on the LCFIVertex package, which was used in the simulation studies for ILC Letter of Intent [14,15] and CLIC Concept Design Report [10]. Current software is built in mind of a future e^- e^+ collider. Although the software is modular, it will be described in order that it will be used in a physics analysis,

The vertex finding algorithms perform vertex fitting and identify primary and secondary vertex. There is a “V0” particle rejection, which is when neutral particles decay or convert into a pair of charged tracks. The topology is similar to the decay of b or c hadrons. Hence it is important to remove the V0 particles to improve the heavy quark falvour tagging.

Jet clustering ensures that the secondary vertices and the muons identified from semileptonic decay are combined. Therefore, it is consistent with the hadronic decay. Jet algorithms used are Durham and Durham modified algorithms.

Vertices are refined to improve the b jet identification from c jet. Two vertices is strongly correlated to a b jet. Hence the vertices refining will reconstruct as many secondary vertices correctly as possible.

The final flavour tagging of the jet is done using multivariate analysis, which will be discussed in section 4.3. Using TMVA software package [16], Boosted Decision Tree classier is used. A series of flavour sensitive variables are calculated, and the classification is divided four sub-set: jet with zero, one, or two properly reconstructed vertices, or a single-track pseudovertex. For each sub-set, a jet can either be a b jet, a c jet, or a light flavour quark jet (u, d or s). The multiclass classifier’s response is normalised across different sub-set, and they will be referred in the subsequent physics analysis as the tag value.

4.3 Multivariate analysis

Multivariate analysis (MVA) has become increasingly common in high energy physics. MVA can be viewed as an advanced tool for regression or classification. Comparing to the traditional cut based method, modern machine learning technique offers much improvement in data analysis.

Software package for MVA used throughout this document is TMVA [16].

A typical machine learning MVA classification involves two classes, also known as signal and background. A machine learning model, called classifier in TMVA, needs to be trained with training data. The model requires a set of discriminative variables, which should separate signal from background. The trained model will be applied onto the testing data, for signal extraction. Response of the model could be signal/background, or be a number in a continuous spectrum, where the user decides the value to separate signal from background.

Strictly, there should be three statistically independent samples for the MVA. One sample is for the training. Another sample for the validation, including optimisation and checking for overfitting. The last sample is for testing. However, due to technical reason, sometimes the same sample is used for the validation and the testing.

This classification scheme can be easily extended to multiple classes, implemented in TMVA with multiclass class.

4.3.1 Optimisation and overfitting

The optimisation of the model is to select the optimal free parameters of the model. One could build a complex model which fits the training samples very well, but it would not be optimal for another testing sample. A simple model is less prone to statistical fluctuation of samples, however, it might be too simple to achieve the optimal modeling. The former case is known as overfitting, or overtraining. The latter case is called underfitting, or undertraining.

The compromise is clear. The optimal model is one between overfitting and underfitting. In practice, this involves building the model with increasing complexity, and finding the point where overfitting occurs.

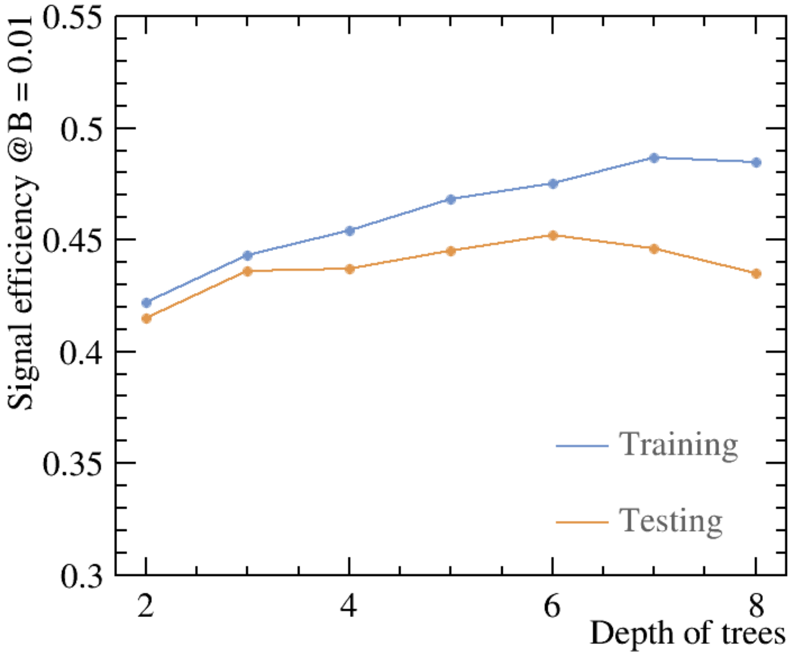


Figure 4.1: MVA overtraining

figure 4.1 shows a typical overfitting plot. Overfitting is defined when the efficiency of signal selection in the training samples increases, but the efficiency in the testing sample decrease. Here the example is chosen from double Higgs analysis, using Boosted Decision Tree model, for $\sqrt{s} = 3\,\text{TeV}$ samples. The efficiency of signal selection is defined as the signal efficiency when background efficiency is 1%, report by the TMVA training process. In the plot, the depth of the tree, or the number of layers in the tree, reflects the complexity of the model. From tree depth 2 to 5, the efficiency for both testing and training samples increases. From tree depth 6 onwards, the overfitting occurs. In this particular example, one should choose a tree depth fewer than 7 to avoid overfitting.

There are methods to assign the error on the selection efficiency. Thus one can make a better choice of parameters to avoid overfitting. These methods were not implemented due to the technical capacity provided by the TMVA.

4.3.2 Choice of models

The model, known as the classifier in TMVA, can be as simple as cut based, likelihood or linear regression. It can be complicated as non linear tree, non linear neutral network or support vector machine. Regardless of model complexity, the choice of most optimal classifier is often data driven. Also, given the free parameters in each model, the comparison between different models without individual tuning is not rigorous. Nevertheless, as researchers in the machine learning suggested, the boosted decision tree is probably the best out-of-the-box machine learning method. Neutral network could potentially be better than the boosted decision, but it requires more tuning, and it is less intuitive to interpret the model. For these reasons, boost decision tree (BDT) is often the choice of machine learning model in the high energy physics. And it is used in various physics analysis in this document.

Before describing BDT in detail, we will first visit the traditional rectangular cut model, and the Projective Likelihood method, which is used in the photon ID in the PandoraPFA.

4.3.3 Rectangular Cut

Probably the most intuitive model, the rectangular cut method optimise cuts to maximise some specific metric. The metric could be the signal efficiency for a particular background efficiency. Alternatively, the metric can be the significance, $\frac{\mathsf{S}}{\sqrt{\mathsf{S+B}}}$, where S and B are signal and background numbers, respectively.

Discriminative variables gives better separation power when they are gaussian-like and statistically independent. Therefore it is common to decoorelate the variables and gaussian transform them before using the rectangular cut MVA.

Because its simplicity, the cut method is often performed manual, much more often in the time pre-date the wide spread of machine learning methods. It is still commonly used for the pre-selection step before the MVA, and other simple usages. Unless specified, the optimal cuts proposed in this document for various physics analysis are found using the rectangular cut method manually.

4.3.4 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 in the TMVA 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.

TMVA implementation also fits an underlying function to the probability density. The PandoraPFA implementation simply uses binned likelihood ratio, R, as the output, due to the simplicity. The sub-categories for the PandoraPFA implementation are determined by the cluster energy.

Similarly to 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.

4.3.5 Boosted decision tree

Boost decision tree (BDT) is a non linear tree based model. Its rather complex nature requires a careful explanation of many concepts within the BDT.

Decision tree is a binary tree, where each node, the splitting point, uses a single discriminative variable to decide whether a event is signal-like ("goes down by a layer to the left"), or background-like ("goes down by a layer to the right"). At each node, samples are divided into signal-like and background-like sub-samples. The tree growing starts at the root node, and stops at certain criterion, which could be the minimum number of events in a node, the number of layers of the tree, or a minimum/maximum signal purity.

The training of the decision tree is to determine the optimal cut at the node. The the probability of the cut produces the signal is p. Three commonly used metrics for two-class classification are

1. Misclassification error: $1 - \max{(p,1-p)}$,
2. Gini index: $2p(1-p)$,
3. Cross-Entropy or deviance: $-p\log p - (1-p)\log{(1-p)}$.

The using of a trained decision tree is to transverse along the tree. The event is classified as signal or background depending on whether it falls in the signal-like or background-like end node.

Decision tree has a low bias, but high variance. This means it is very easy to construct a tree that fits the training data very well, but the tree would not be optimal for the testing sample. To overcome the instability of the decision tree, many methods have been developed. The most successful one is boosting.

Boosting: it is a technique where the misclassified events receives a higher weight than the correctly classified events. Therefore, when the training is iterated, the misclassified events would receive higher and higher wights and more likely to classify correctly. The boosting is done at every iteration, which can be few hundred or few thousand time. This will create a "forest" of many trees. The final output could be a majority vote, by transversing the event to the end node for each tree in the forest.

Bagging: also known as boot-strap, it is a method that select a simple random sub-set of the training sample, and apply the model. In this case, every boosting iteration takes a bagged sample, rather than the whole sample.

TMVA implementation of the BDT for the output is using a likelihood estimator, depending on how often a event is classified as signal in the forest. The likelihood number is later used to select signal from background.

4.3.6 Optimisation of Boosted decision tree

Many parameters can be tuned. Hence we dedicate a small section to describe the tuning parameters.

The most important parameter is the depth of a tree, which determines how many end nodes a tree has, or the degrees of freedom of a tree. The related parameter is the number of trees. Experience shows that using many small trees yields the best result.

The minimum number of events in a node, which is a stopping criteria for tree growing, affects the size of the tree. But it is less influential than the depth of the tree.

The learning rate, which controls how fast the weight changes for events in each boosting iteration. Experience shows small learning rate with many trees work better than large learning rate with few trees.

The usual choice of the metric for the optimal cuts is either Gini index or cross-entropy. (See section 4.3.5) We chose Gini index for out BDT usages, as it makes little difference to performances, comparing to the cross-entropy metric.

Number of bins per variables for the cut is necessary to make tree growing efficient. Discrete binned variables are faster to computer than continuous variables. The parameter does not impact the performance much. However, variables should be pre-processed before going into the model. For example, the variable should be limited to a sensible range to avoid the extremes. The variable should also be transformed to obtain a more uniform distribution, if the original distribution is highly skewed.

The boosting has two variant, adaptive boost and gradient boost. For all the BDT used in this document, adaptive boost is used.

For the end node, it is determined as either signal-like or background-like, based on the majority for the training event in the end node. Numerically, it corresponds to 1/0. However, the end node could also use signal purity as the output, resulting in a continues spectrum of [0,1].

4.3.7 Multiple classes

The above discussion is done assuming two classes - signal and background. The argument can be easily extended to multiple classes. 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. Multiclass is used in falvour tagging of jets, section 4.2, and in the tau lepton final state separation study, section ??.

4.4 Event shape varaibles

Event shape variables are some useful global variables to describe the shape of the event, for example whether it is back-to-back, or homogenous in the solid angle.

The classical event shape thrust [?], is defined as

$$T = \max_{\hat{t}} \frac{\sum_i \left|\hat{t} \cdot \vec{p_i}\right|}{\sum_i |\vec{p_i}|} \tag{4.4}$$

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 thrust value is useful in picking out back-to-back two-jet event. Thrust axis is useful to separate each jet in a back-to-back two-jet event.

Sphericity tensor [?], is defined as

$$\boldsymbol{S}^{\alpha\beta} = \frac{\sum_i p_i^{\alpha} p_i^{\beta}}{\sum_i |\vec{p_i}|^2}, \tag{4.5}$$

where $\vec{p_i}$ is the momentum vector of the particle i. Summation is over all particles in the event. α and β refer to the x, y, z coordinate axis. Eigenvalues of tensor $\boldsymbol{S}$ can be found, or in this case diagonalisation of the matrix $\mathbf{S}$, denoted with λ_1, λ_2, λ_3. The normalisation condition requires $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Sphericity, S, is defined in terms of λ,

$$S = \frac{3}{2}(\lambda_1 + \lambda_2). \tag{4.6}$$

S, is 0 for a perfect pencillike back-to-back two-jet event, and 1 for a perfect spherically symmetric event.

Aplanarity is another event shape varaible that distinguishes spherical symmetrical events from planar and linear events. The definition is

$$S = \frac{3}{2}(\lambda_1), \tag{4.7}$$

where λ_1 is the largest eigenvalue in the diagonalised sphericity tensor.

4.5 Miscellaneous

An event in a collider experiment refers to one collision and the subsequent energy deposition in the detector. An event corresponds to a certain type of physics process.

Event Number	True Signal	True Background
Selected Signal	N_S	N_1
Selected Background	N_2	N_B

Table 4.1: A toy example to demonstrate definitions of efficiency and purity.

Often we are dealing with extracting a type of events, from a large number of other events. The signal, or signal events refer to events of interests. Other events are referred to as the background, or background events.

Typical metrics of signal selection is efficiency and purity. This toy example illustrates definitions of efficiency and purity.

Signal selection efficiency is defined as $\frac{N_S}{N_S+N_2}$. Signal selection purity is defined as $\frac{N_S}{N_S+N_1}$. Significance is a quantity that is similar to purity, $\frac{N_S}{\sqrt{N_S+N_1}}$

When we are describing particles, light lepton, l_l, refer to electrons, e^-, and muons, μ^-. Light quarks, q_l, refer to up quark, u, down quark, d, and strange quark, s.

Computational intensive jobs are processed either on the Cambridge High Energy Physics grid, or the CLIC computing grid.

Chapter 5

Photon Reconstruction

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

Chapter 6

Tau Lepton Final State Separation

"MVA: Turn numbers into gold."
— TMVA

Chapter 7

Double Higgs Bosons Production Analysis

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

7.1 Motivation

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.

7.2 Theory

general higgs field

Lagrangian

current constraunt

single higgs coupling measurement done in higgs

Double higgs measurement

The main mechanism for double Higgs production

7.3 Analysis Straggly Overview

Proof-of-principle study was perform at CLIC using CLIC_ILD detector model for $\sqrt{s} = 1.4\,\mathrm{TeV}$ and $3\,\mathrm{TeV}$. Simulated samples, including those containing double higgs production were used. Signal events, events with double higgs production, were selected via a set of carefully designed and complicated methods. g_{HHH} and g_{WWHH} are extracted simultaneously with template fitting with modified couplings samples.

7.4 Monte Carlo Sample Generation

Single channel is defined as $\mathrm{e^-e^+} \rightarrow \mathrm{HH}\nu\bar{\nu}$. It is divided into sub-channel $\mathrm{HH} \rightarrow \mathrm{b\bar{b}W^+W^-}$ and $\mathrm{HH} \rightarrow \mathrm{b\bar{b}b\bar{b}}$ to allow closer examination and an improvement of signal selection when combined. In particular, I studied $\mathrm{HH} \rightarrow \mathrm{b\bar{b}W^+W^-}$ sub-channel.

Selected background samples, including processes initiated by photons, are considered in the analysis and listed in Table **??**. These background were expected share similar topologies with the signal process. When describing a multi-quark final state, it is referring to all final states of the same number of quarks, including final states with possible additional neutrinos and or leptons. A multi-quark final state does not include higgs production, unless explicitly stated.

The usual two-quark and four-quark final states were considered. 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.

Photon-electron/photon-photon interactions with four-quark final states were considered. Photon-electron interaction with two-quark final state, one Higgs, and one neutrino is considered. Photon-electron interaction with two-quark final state, one Higgs, and one lepton is not considered due to its negligible cross section.

Single higgs productions are not considered because topologies are very different to the single process. Six-quark final states were not considered due to computational limitation.

For processes involving Higgs production explicitly, simulated Higgs mass is 126 GeV. As multi-quark final state background samples could, in principle, contain double higgs production, they are generated with a Higgs mass of 14 TeV. This will produce negligible double higgs production cross section.

All samples are generated with WHIZARD 1.95 [], taking into account the expected CLIC luminosity spectrum. PYTHIA 6.4 [] tuned on LEP data [] is used to describe fragmentation, hadronisation processes, and Higgs decays. TAUOLA [] is used for τ lepton decays.

Simulation

For most background processes, events are simulated when invariant mass of quarks are above 50 GeV. For electron-photon interaction with four quarks and a neutrino 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 $\gamma\gamma \rightarrow hadrons$ is simulated and overlayed [] to all samples according to the integration time of each subdetector.

Channel	$\sigma(\sqrt{s} = 1.4\,\mathrm{TeV})$ / fb	$\sigma(\sqrt{s} = 3\,\mathrm{TeV})$ / fb
$e^-e^+ \rightarrow HH\nu\bar{\nu}$	0.588	0.149
$e^-e^+ \rightarrow q_l q_l H\nu\bar{\nu}$	0.86	1.78
$e^-e^+ \rightarrow c\bar{c}H\nu\bar{\nu}$	0.36	1.12
$e^-e^+ \rightarrow b\bar{b}H\nu\bar{\nu}$	0.31	1.91
$e^-e^+ \rightarrow qqqq$	1245.1	546.5
$e^-e^+ \rightarrow qqqq\ell\ell$	62.1	169.3
$e^-e^+ \rightarrow qqqq\ell\nu$	110.4	106.6
$e^-e^+ \rightarrow qqqq\nu\bar{\nu}$	23.2	71.5
$e^-e^+ \rightarrow qq$	4009.5	2948.9
$e^-e^+ \rightarrow qq\ell\nu$	4309.7	5561.1
$e^-e^+ \rightarrow qq\ell\ell$	2725.8	3319.6
$e^-e^+ \rightarrow qq\nu\nu$	787.7	1317.5
$e^-\gamma(BS) \rightarrow e^-qqqq$	1160.7	1268.7
$e^+\gamma(BS) \rightarrow e^+qqqq$	1156.3	1267.6
$e^-\gamma(EPA) \rightarrow e^-qqqq$	287.1	287.9
$e^+\gamma(EPA) \rightarrow e^+qqqq$	286.9	287.8
$e^-\gamma(BS) \rightarrow \nu qqqq$	136.9	262.5
$e^+\gamma(BS) \rightarrow \bar{\nu}qqqq$	136.4	262.3
$e^-\gamma(EPA) \rightarrow \nu qqqq$	32.6	54.2
$e^+\gamma(EPA) \rightarrow \bar{\nu}qqqq$	32.6	54.2
$e^-\gamma(BS) \rightarrow qqH\nu\nu$	15.8	58.6
$e^+\gamma(BS) \rightarrow qqH\nu\nu$	15.7	58.5
$e^-\gamma(EPA) \rightarrow qqH\nu\nu$	3.39	11.7
$e^+\gamma(EPA) \rightarrow qqH\nu\nu$	3.39	11.7
$\gamma(BS)\gamma(BS) \rightarrow qqqq$	21406.2	13050.3
$\gamma(BS)\gamma(EPA) \rightarrow qqqq$	4018.7	2420.6
$\gamma(EPA)\gamma(BS) \rightarrow qqqq$	4034.8	2423.1
$\gamma(EPA)\gamma(EPA) \rightarrow qqqq$	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. Simulated W has invariant mass of 80.385 GeV.

7.5 Physics object and event reconstruction

Simulation is performed by MOKKA, interfacing GEANT 4. 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 a compromise of signal selection and background rejection. The latest function flavour tagging processor exist in iLCSoft v01-16, which prevented the usage of more recent iLCSoft.

For my signal channel, $\mathrm{HH} \rightarrow \mathrm{b\bar{b}W^{+}W^{-}}$, there is no lepton in the final state. Hence a effective lepton identifier would improve the signal identification. 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 reject more background at the cost of losing a bit more signals, to increase the signal significance. It was found that the same set of parameters work well under $\sqrt{s}$ = 1.4 TeV and 3 TeV.

7.5.1 Electron and muon identification

IsolatedLeptonFinderProcessor

In Marlin package, IsolatedLeptonFinderProcessor has been used. The optimal parameters ware chosen in collaboration and tested. The particle is identified as an isolated light lepton if it passes a chain of cuts.

A charge track is considered if it has more than 15 GeV energy. An electron is identified if the energy in the ECal is over 90% of the total calorimetric energy. A muon is identified if the energy in the ECal is between 5% and 25% of the total calorimetric energy. Furthermore, only primary track is selected, which requires the Euclidean distance in the x-y plane, the in z direction, and in the x-y-z three dimensional space of the track starting point to the impact point to be less than 0.02 mm, 0.03mm, and 0.04 mm, respectively. The isolation criteria states that

$$E_{cone}^{2} \leqslant 5.7 \times E_{l} - 50 \tag{7.1}$$

where, E_{cone} is the total energy of PFOs within an opening angle of $\cos^{-1}(0.995)$ of the light lepton, and E_{l} is the energy of the light lepton.

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,

$$E_l \geqslant 23 \times \sqrt{E_{cone}} + 5 \tag{7.2}$$

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,

$$E_l \geqslant 28 \times \sqrt{E_{cone}} + 30 \tag{7.3}$$

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 7.2

7.5.2 Tau identification

TauFinderProcessor

With a decay length of $87\mu m$, 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 [17], 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,
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 BonoTauFinderProcessor 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 7.2

7.5.3 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.

Selection / Efficiency (1.4 TeV)	Signal	$qqq\ell\nu$
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%
Processor / Efficiency (3 TeV)	Signal	$qqq\ell\nu$
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 7.2: isolated lepton finder processors performance on the signal and selected background samples.

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 $\sqrt{s} = (\text{ TeV } 3)$, the BeamCal detection efficiency is provided by a software package []. For $\sqrt{s} = (\text{ TeV } 1.4)$, 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

$$\varepsilon = \begin{cases} 0, & \text{if } E < 50\,GeV \\ 0.99 \times \frac{(\mathrm{erf}(E-100)+1)}{2}, & \text{otherwise} \end{cases} \tag{7.4}$$

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.

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

Table 7.3: Very forward electron and photon finder performance on the signal and selected background samples.

7.5.4 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.

7.6 Jet reconstruction

The signal channel, $HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}qqqq$, 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.

7.6.1 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 $\sqrt{s} = 1.4\,\text{TeV}$ and $\sqrt{s} = 3\,\text{TeV}$.

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, $\mathrm{HH} \rightarrow \mathrm{b\bar{b}W^+W^-} \rightarrow \mathrm{b\bar{b}}qqqq$, 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, $\mathrm{HH} \rightarrow \mathrm{b\bar{b}W^+W^-} \rightarrow \mathrm{b\bar{b}}qqqq$, 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, $m_{\mathrm{H_{bb}}}$, $m_{\mathrm{H_{WW^*}}}$, and two W masses m_{W}, $m_{\mathrm{W^*}}$. 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, $m_{\mathrm{H_{bb}}}$, $m_{\mathrm{H_{WW^*}}}$, 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:

$$f(m) = Ae^{-\frac{(m-\mu)^2}{g}} \begin{cases} g = 2\sigma_L + \alpha_L(m-\mu), & \text{if } m < \mu \\ g = 2\sigma_R + \alpha_R(m-\mu), & \text{if } m \geqslant \mu \end{cases} \tag{7.5}$$

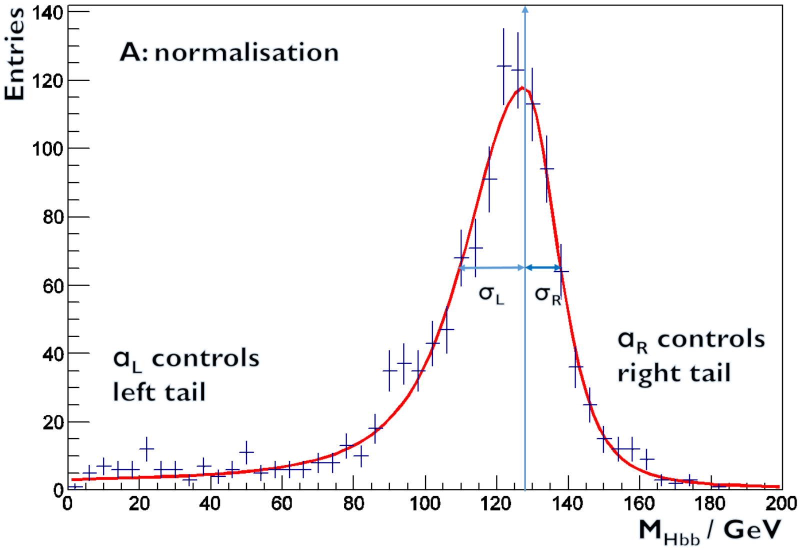


Figure 7.1: A typical example of MC mass fit of $m_{H_{bb}}$ for double higgs analysis. Red line indicates the best fit. Vertical arrow indicates the fitted peak position.

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 $\sqrt{s} = 1.4\,\text{TeV}$, 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 7.3 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 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 7.4, 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 7.5, 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

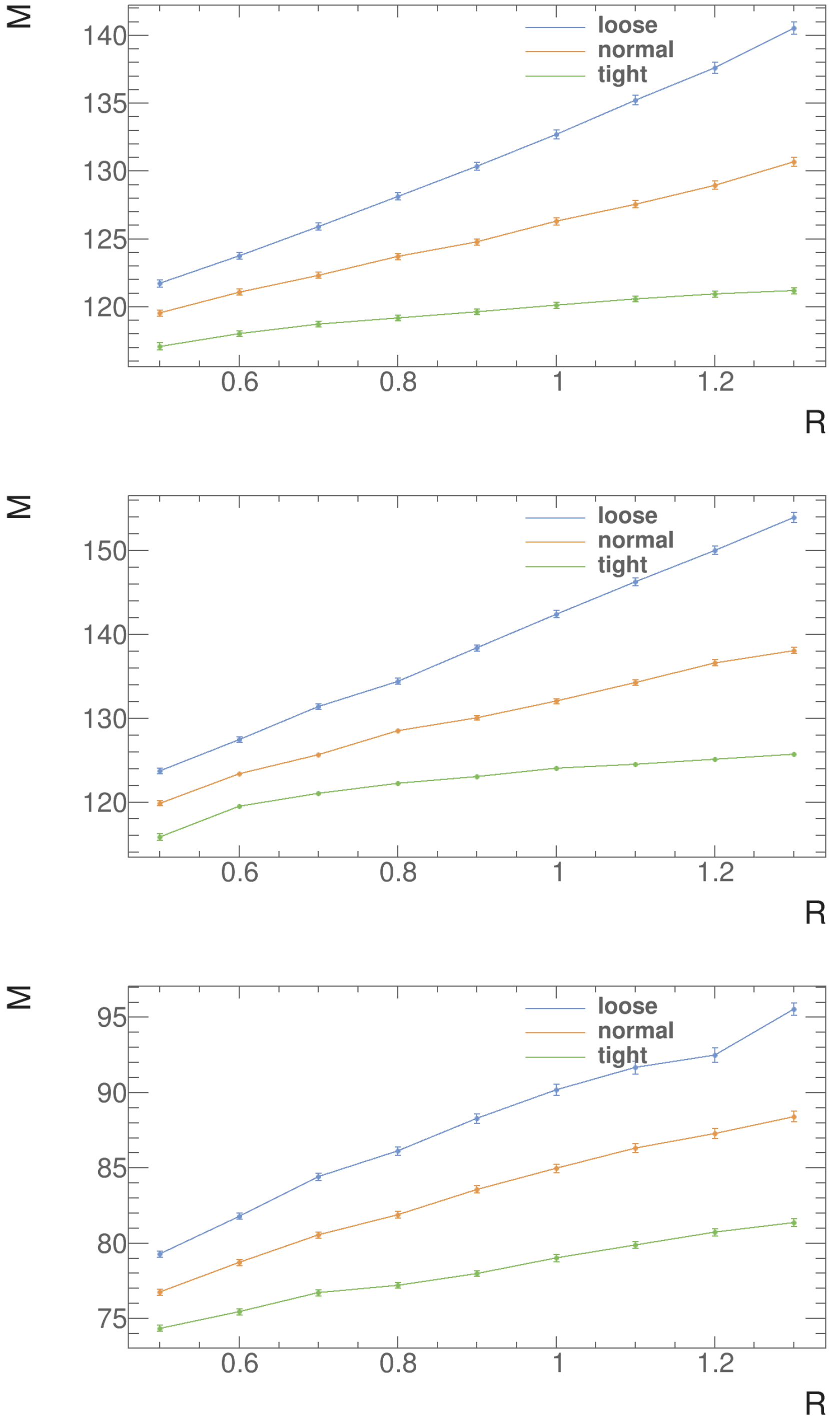


Figure 7.2: Fitted mass and statistical error of H_{bb}, H_{WW^*} and W for $\sqrt{s} = 1.4\,\text{TeV}$, for loose, normal and tight selected PFO against R parameter.

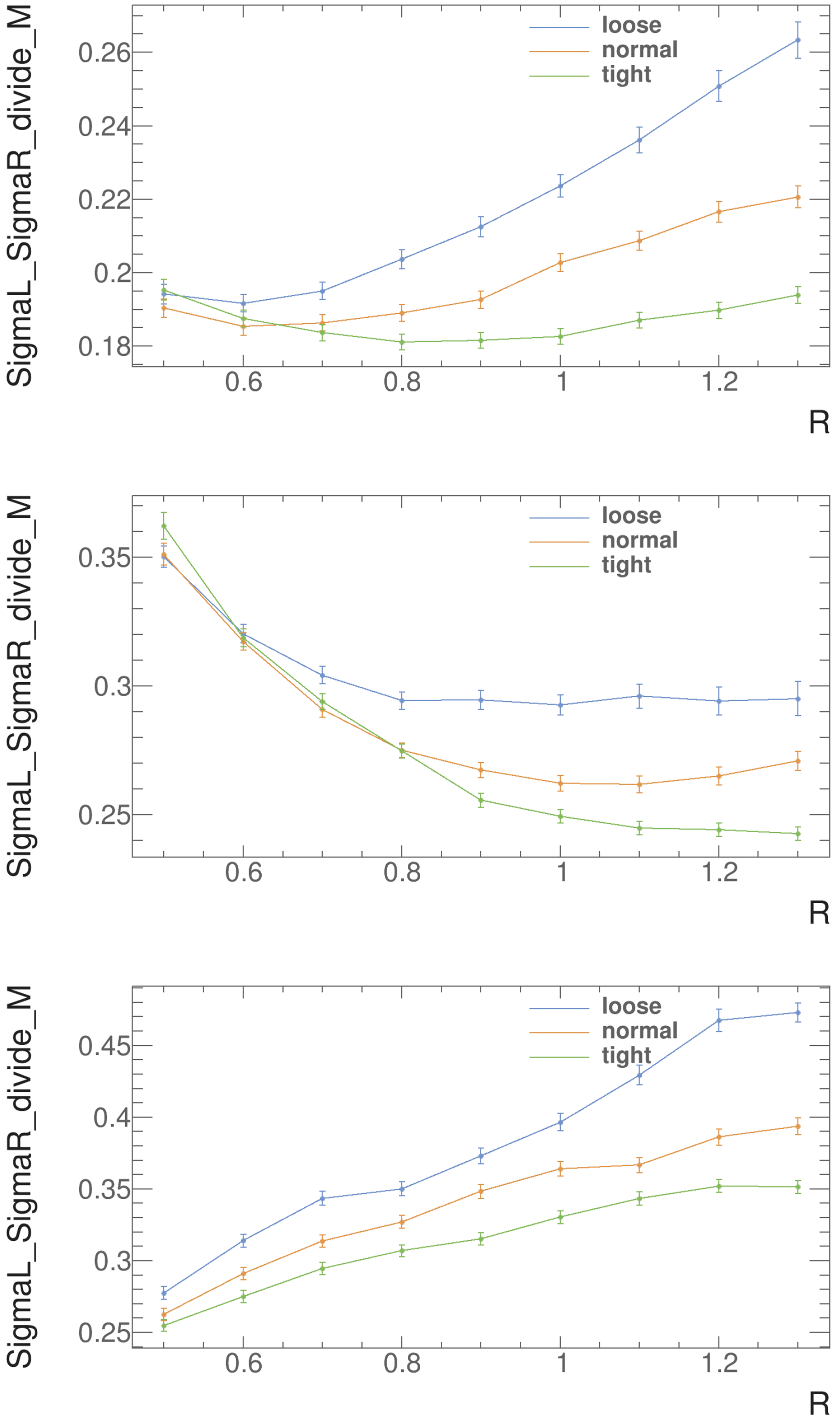


Figure 7.3: Fitted combined width and statistical error of $\mathrm{H_{bb}}$, $\mathrm{H_{WW^*}}$ and W for $\sqrt{s} = 1.4\,\mathrm{TeV}$, for loose, normal and tight selected PFO against R parameter.

Jet Parameters	$\sqrt{s}$ = 1.4 TeV	$\sqrt{s}$ = S 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 7.4: The extracted fitted parameters of optimal jet reconstructions, normal selected PFO with $R = 0.7$ for $\sqrt{s}$ = 1.4 TeV and tight selected PFO with $R = 0.7$ for $\sqrt{s}$ = 3 TeV.

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 TeV is tight selected PFO with $R = 0.7$.

The extracted fitted parameters of optimal jet reconstructions are summarised in table 7.4.

7.6.2 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 $e^-e^+ \to Z\nu\bar{\nu}$, where Z decays to $q_l\bar{q}_l$, $b\bar{b}$, or $c\bar{c}$ at $\sqrt{s}$ = 1.4 TeV and $\sqrt{s}$ = 3 TeV, because they have similar event topology as the signal, and they have only two jets in the final state.

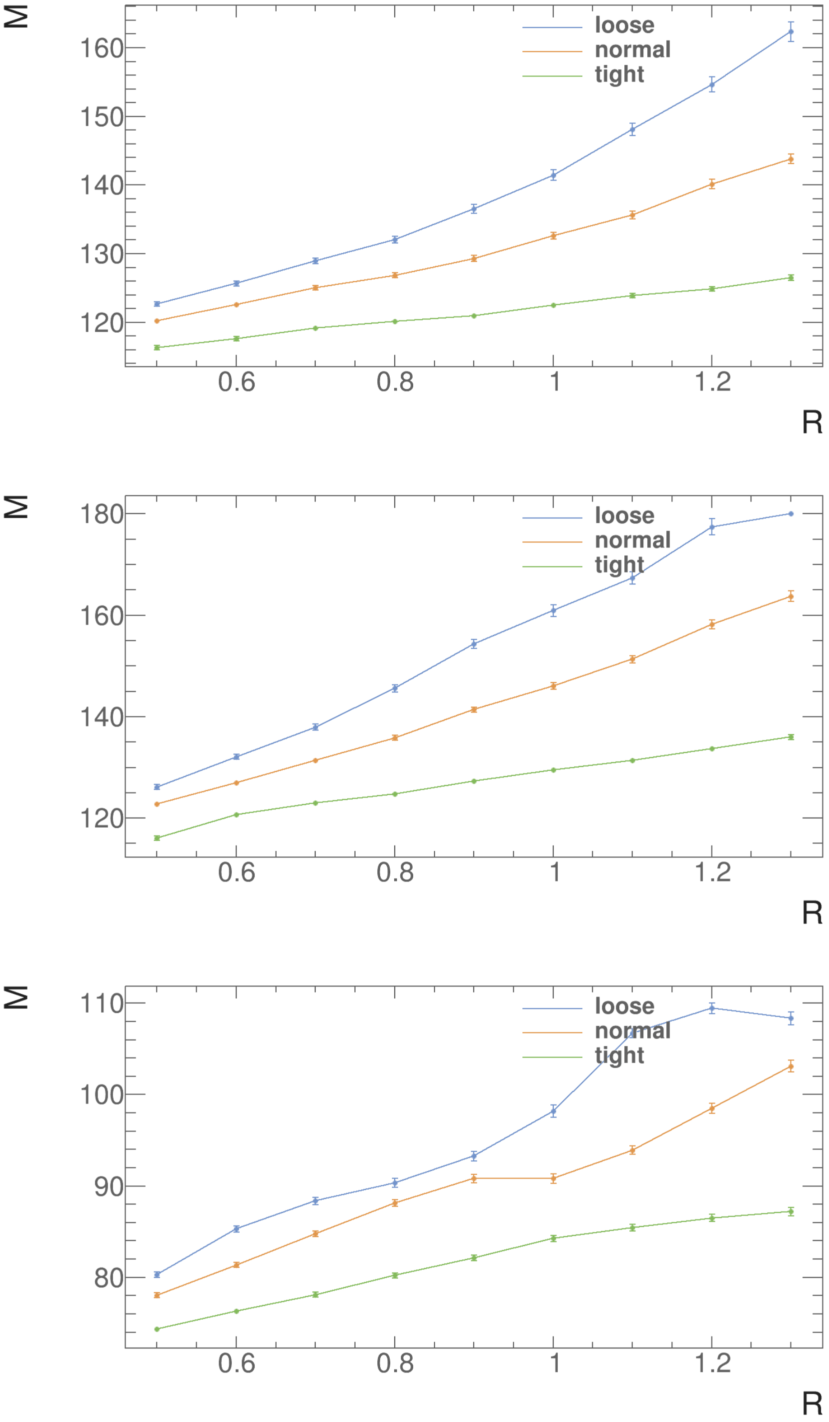


Figure 7.4: Fitted mass and statistical error of H_{bb}, H_{WW^*} and W for $\sqrt{s} = 3\,\text{TeV}$, for loose, normal and tight selected PFO against R parameter.

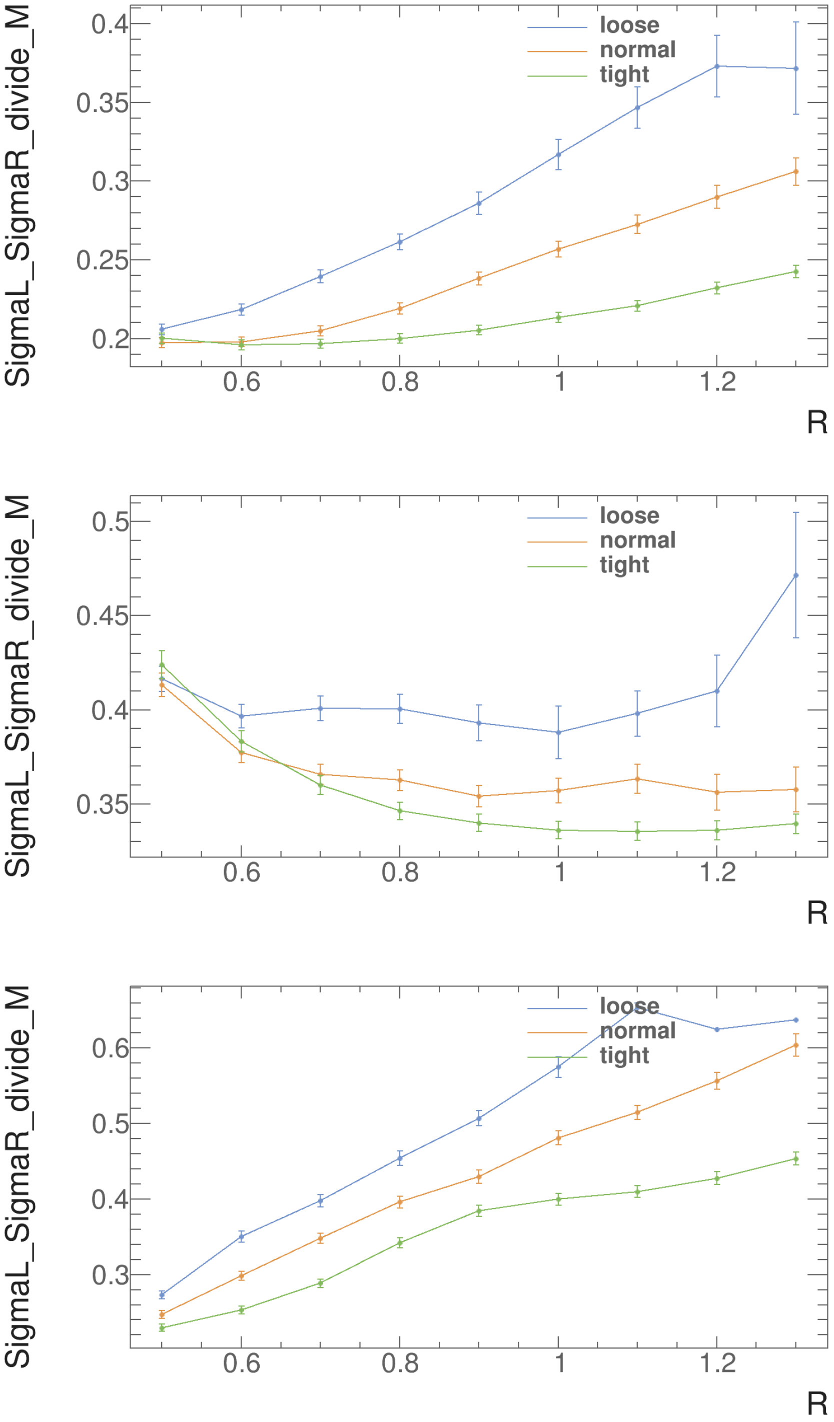
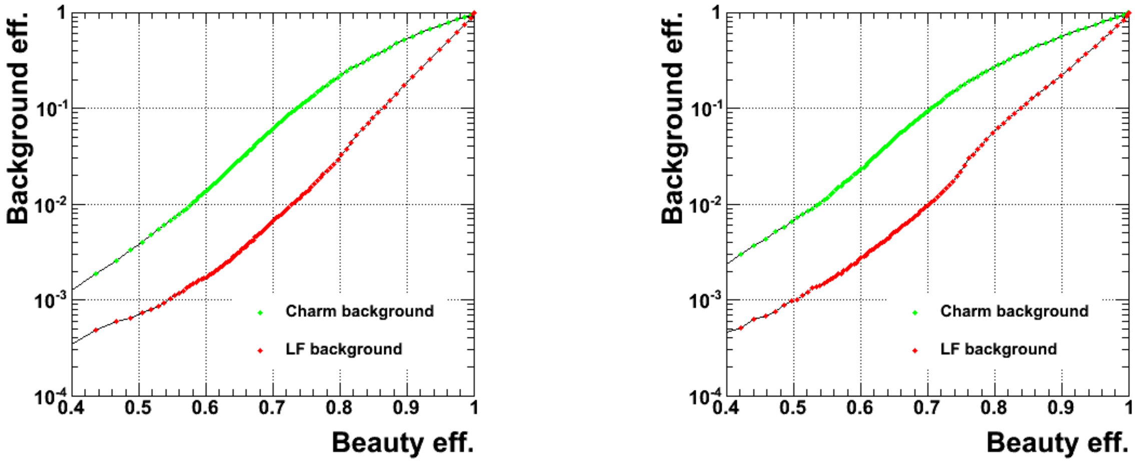


Figure 7.5: Fitted combined width and statistical error of H_{bb}, H_{WW^*} and W for $\sqrt{s} = 3\,\text{TeV}$, for loose, normal and tight selected PFO against R parameter.



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

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

Figure 7.6: Performance of b-jet tagging with trainning samples

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.

7.6.3 Jet pairing

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

The actual pairing is done via a minimisation

$$\chi^2 = \left(\frac{m_{ij} - \mu_{H_{bb}}}{\sigma'_{H_{bb}}}\right)^2 + \left(\frac{m_{klmn} - \mu_{H_{WW*}}}{\sigma'_{H_{WW*}}}\right)^2 + \left(\frac{m_{kl} - \mu_W}{\sigma'_W}\right)^2, \qquad (7.6)$$

where, $\mu_{H_{bb}}$ and $\sigma'_{H_{bb}}$ are the fitted invariant mass, and the fitted width, respectively. Both are obtained in section 7.6.1. $\sigma'_{H_{bb}}$ 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 7.4. 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.

7.7 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.

7.7.1 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,

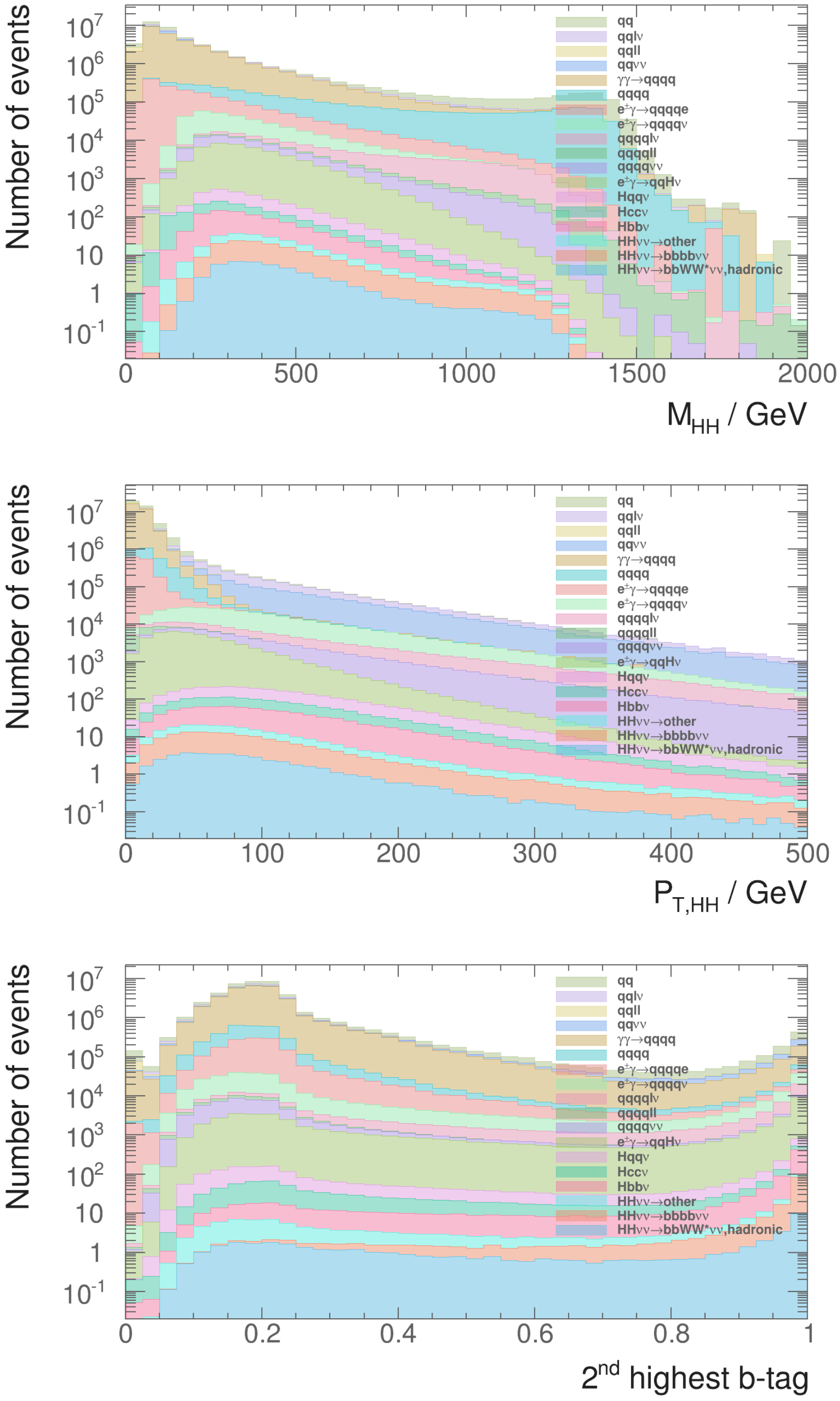


Figure 7.7: Discriminative pre-selection variables for $\sqrt{s} = 1.4\,\text{TeV}$, after rejecting events with identified leptons, and jet pairing

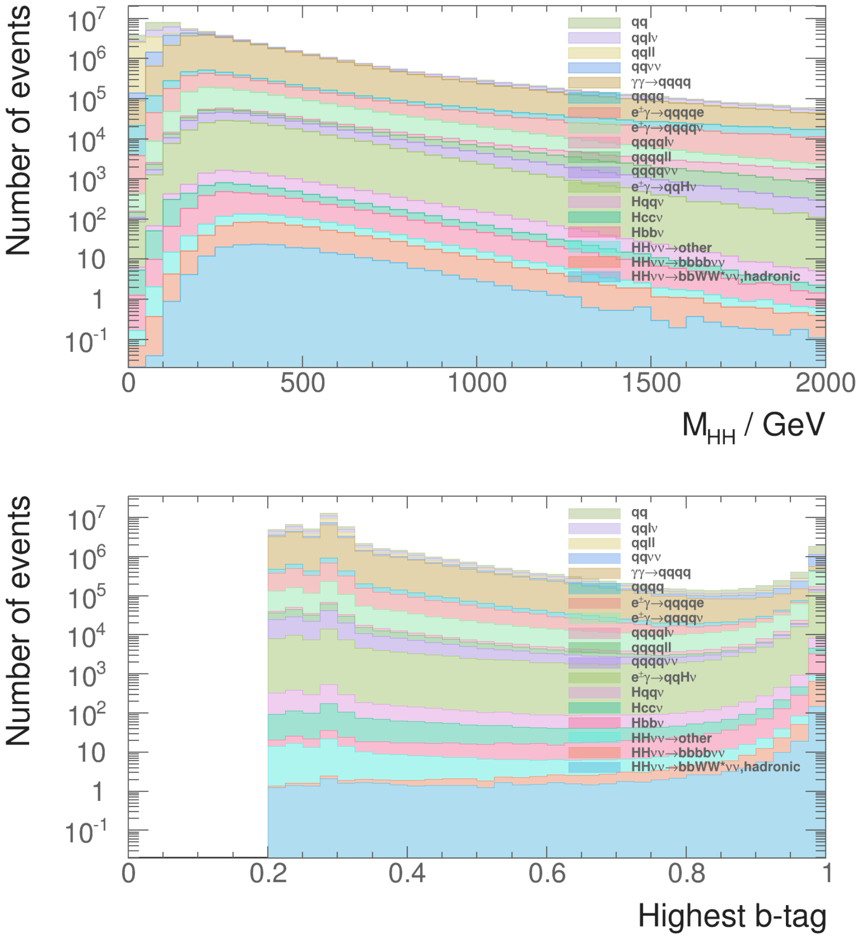


Figure 7.8: Discriminative pre-selection variables for $\sqrt{s} = 3\,\text{TeV}$, after rejecting events with identified leptons, and jet pairing

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.

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

For $\sqrt{s} = 3\,\mathrm{TeV}$, invariant masses for $\mathrm{H_{bb}}$, $\mathrm{H_{WW^*}}$, 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 **??**.

7.7.3 Mutually exclusive cuts for $\mathbf{HH \rightarrow b\bar{b}W^+W^-}$ and $\mathbf{HH \rightarrow b\bar{b}b\bar{b}}$

Since the analysis for $\mathrm{e^-e^+ \rightarrow HH\nu\bar{\nu}}$ channel is divided into two subchannels, $\mathrm{HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}}\mathbf{qqqq}$ and $\mathrm{HH \rightarrow b\bar{b}b\bar{b}}$, 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 7.9, two subchannels can be clearly separated in the two dimensional parameter space. The optimal rectangular cuts were selected by scanning the two

Channel / Efficiency $\sqrt{s} = 1.4\,\text{TeV}$	Expected number of events	Lepton ID and jet pairing	$m_{HH} > 150\text{GeV}$	$B_2 > 0.2$	$p_T > 30\text{GeV}$
$e^-e^+ \to HH\nu\bar{\nu} \to$ $b\bar{b}W^+W^-\nu\bar{\nu}$, 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%
$e^-\gamma(BS) \to qqH\nu\nu$	1241.0	61.2%	60.0%	45.4%	34.5%
$e^+\gamma(BS) \to qqH\nu\nu$	1232.8	61.2%	60.0%	45.4%	34.2%
$e^-\gamma(EPA) \to qqH\nu\nu$	354.7	31.9%	31.3%	23.8%	18.2%
$e^+\gamma(EPA) \to qqH\nu\nu$	355.1	32.2%	31.7%	23.9%	18.7%
$\gamma(BS)\gamma(BS) \to qqqq$	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 7.5: 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.

Channel / Efficiency $\sqrt{s} = 3\,\text{TeV}$	Expected number of events	Lepton ID and jet pairing	$m_{HH} > 150\text{GeV}$	$B_1 > 0.7$
$e^-e^+ \to HH\nu\bar{\nu} \to b\bar{b}W^+W^-\nu\bar{\nu}$, 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%
$e^-e^+ \to q_lq_lH\nu\bar{\nu}$	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%
$e^-\gamma(BS) \to qqH\nu\nu$	92588.0	58.3%	56.2%	37.3%
$e^+\gamma(BS) \to qqH\nu\nu$	92430.0	58.1%	56.0%	37.1%
$e^-\gamma(EPA) \to qqH\nu\nu$	23400.0	30.1%	29.2%	19.4%
$e^+\gamma(EPA) \to qqH\nu\nu$	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 7.6: 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.

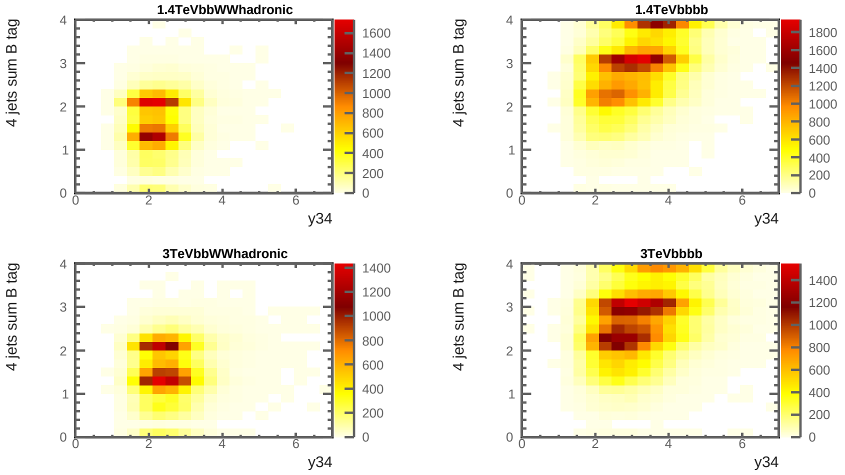


Figure 7.9: Sum of b tag against y_{34}, shown for signal samples

$\sqrt{s}$	$\mathit{selection}$	$HH \to b\bar{b}qqqq$ Selection Efficiency	$HH \to b\bar{b}b\bar{b}$ 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 7.7: Mutually exclusive cuts, for full signal samples

parameters, and maximising

$$\varepsilon = P(\mathit{subchannel}_1|\mathit{selection}) \times P(\mathit{subchannel}_2|\neg\mathit{selection}) \tag{7.7}$$

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

Variables tested includes ΣB_{4jets}, $\sum_1^3 B_{4jets}$, y_{34}, y_{45}, y_{56}, y_{67} and other related variables. The best separation was summarised in table 7.7.

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

7.8 Discriminative Variables

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

Channel / Efficiency	Sanity $\sqrt{s}$ = 1.4 TeV	Mutually exclusive $\sqrt{s}$ = 1.4 TeV	Sanity $\sqrt{s}$ = 3 TeV	Mutually exclusive $\sqrt{s}$ = 3 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%
$e^-\gamma(BS) \to qqH\nu\nu$	34.5%	30.2%	37.2%	30.2%
$e^+\gamma(BS) \to qqH\nu\nu$	34.2%	30.3%	37.1%	30.2%
$e^-\gamma(EPA) \to qqH\nu\nu$	18.2%	16.0%	19.0%	15.7%
$e^+\gamma(EPA) \to qqH\nu\nu$	18.7%	16.4%	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 7.8: List of signal and background samples with the corresponding expected number at $\sqrt{s} = 1.4$ TeV and $\sqrt{s} = 3$ 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.

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

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 **??**.

7.9 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.

7.10 Signal selection results

7.11 Couplings extration

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

Colophon

This thesis was made in LaTeX 2$_{\varepsilon}$ using the “hepthesis” class [18].

Bibliography

[1] G. Sterman and S. Weinberg, Phys. Rev. Lett. **39**, 1436 (1977).

[2] S. Moretti, L. Lonnblad, and T. Sjostrand, JHEP **08**, 001 (1998), hep-ph/9804296.

[3] G. P. Salam, Eur. Phys. J. **C67**, 637 (2010), 0906.1833.

[4] A. Ali and G. Kramer, Eur. Phys. J. **H36**, 245 (2011), 1012.2288.

[5] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. **C72**, 1896 (2012), 1111.6097.

[6] M. Cacciari and G. P. Salam, Phys. Lett. **B641**, 57 (2006), hep-ph/0512210.

[7] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, Nucl. Phys. **B406**, 187 (1993).

[8] S. D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993), hep-ph/9305266.

[9] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, Phys. Lett. **B269**, 432 (1991).

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

[11] M. Battaglia and F. P., CERN Report No. LCD-Note-2010-006, 2010 (unpublished).

[12] M. Boronat, J. Fuster, I. Garcia, E. Ros, and M. Vos, Phys. Lett. **B750**, 95 (2015), 1404.4294.

[13] T. Suehara and T. Tanabe, Nucl. Instrum. Meth. **A808**, 109 (2016), 1506.08371.

[14] Linear Collider ILD Concept Group -, T. Abe *et al.*, (2010), 1006.3396.

[15] H. Aihara *et al.*, (2009), 0911.0006.

[16] A. Hocker *et al.*, PoS **ACAT**, 040 (2007), physics/0703039.

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

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

List of figures

List of tables