TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Mathematik und Naturwissenschaften Fachrichtung Physik, Institut für Kern- und Teilchenphysik

Lehrstuhl für Experimentelle Teilchenphysik

Master Thesis

Study of charged Higgs bosons search at the ILC for a collision energy of 1 TeV

Christian Drews

Born on: December 18, 1989 in Wismar

Matriculation number: 3781477

Matriculation year: 2011

to achieve the academic degree

Master of Science (M.Sc.)

Supervisor

Prof. Hitoshi Yamamot (Tohoku university)

First referee

Prof. Dr. Arno Straessner

Second referee

Prof. Hitoshi Yamamot (Tohoku university)

Submitted on: October 12, 2017

Todo list

Summary

Abstract (English):

The Two Higgs Doublet Model is a promising extension of the Standard Model where charged Higgs bosons appear. In this study, we assume the mass of the charged Higgs boson to be 350 GeV and perform a simulation study for the production and decay of charged Higgs pair at a linear collider. A charged Higgs boson is assumed to decay to a top quark and a bottom quark followed by the top quark decay to a b quark and a W boson. The final state is reconstructed in two modes: First both W bosons decay hadronically resulting in total of 8 jets and second one W boson decays leptonically and the other W boson decays hadronically resulting in 6 jets plus one lepton. The study is based on a full ILD simulation for collision energy of 1 TeV.
The event selection was conducted with static cuts as well as boosted decision trees both were optimized on signal significance or correctly paired signal significance. The mass measurement undertaken in a template fit and shape fitting methods. It is shown that the charged Higgs boson masses can be measured with 0.5 GeV precision assuming the production cross section to be 9 fb and $BR(\mathrm{H}^{\pm} \rightarrow \mathrm{bW}) = 90$ %, when using boosted decision trees based event selection optimized on correctly paired signal significance with a parameter reduced shape fitting method for the mass measurement.

Kurzfassung (German):

write Kurzfassung

Contents

1 Introduction

1.1 Motivation

As long as we can look back in the history, the human kind was wondering how did the world as we know began and where we came from. Since scientific thinking has evolved the scientific world postulates theories and tries to prove them in experiments. Today we still cannot be sure how the universe began. To satisfy this thirst for knowledge, we use particle collider to investigate our models for higher energies. The higher the energies are, the earlier universe we can learn about, because in the earliest universe the matter was very packed which is equivalent to high energy density. Nowadays the most advanced particle collider is the Large Hadron Collider (LHC). With the findings in July 2012 of a Higgs boson in ATLAS and CMS detector at the LHC, the Standard Model of particle physics (SM) was completed and the long awaited puzzle piece of electroweak symmetry breaking was provided. [16][1]

The SM is very promising and describes a wide ranges of particle physics' nature. But still there are many open questions in today's particle physics, such as baryon asymmetry, the hierarchy problem and the unknown nature of dark matter and dark energy, which cannot be answered by the SM. This makes one believe that there must be an extension of the SM (physics beyond the standard model (BSM)) and the SM must be the limiting model of this more general model. Aside from many other possible models, there are various Two Higgs Doublet Models (2HDM), which have the opportunity to answer some of these questions.

This study focuses on analysis of charged Higgs boson ($\mathrm{H}^{\pm}$) pair production and in particular the measurement of the charged Higgs boson mass $m_{\mathrm{H}^{\pm}}$. The direct search at an electron positron collider through on-shell Higgs bosons by s-channel production is fairly model independent and gives a solid limit on BSM in the contrary to a proton collider where most measurements are highly model dependent. However, the reach is limited by the collision energy which is lower than at the LHC. $\mathrm{H}^{+}\mathrm{H}^{-}$ pair production is especially interesting because the coupling to photons is fixed, thus the production cross section has a lower limit.

Recent combined results from BaBar, Belle and LHCb experiments showed a deviation to the SM of about four standard deviations [19]. On the base of the study B-mesons favoring decays involving muon and tau lepton. Since an increase of the decay into heavier particles was observed, a possible explanation could be the existence of charged Higgs bosons.

1.2 Two Higgs Doublet Model

Two Higgs Doublet Models (2HDM) are possible extensions of the Standard Model with an additional Higgs doublet with the same quantum numbers. Two Higgs doublets would introduce 8 degrees of freedom, where three get absorbed into the longitudinal polarization of the bosons of the weak force ($\mathrm{W}^{\pm}$ and Z), which leaves five Higgs bosons, two CP-even (H and h), where h is defined as the lighter one, one CP-odd Higgs boson (A) and two charged bosons ($\mathrm{H}^{\pm}$).

explain how it is in SM and motivation for second doublet

There are different types of 2HDMs. Models where the first doublet is fermiphobic and only the second doublet couples to fermions are referred to as type I. If the first doublet couples to up-type quarks and the second to down type quarks and charged leptons, it is usually called type II. The Higgs sector in the Minimal Supersymmetric extension of the Standard Model (MSSM) is a type II 2HDM. Type I models cannot be supersymmetric because of the hermitian conjugated doublet in the Yukawa terms for down-type quarks.

This study focuses on the direct search of charged Higgs bosons with a mass $m_{\mathrm{H}^{\pm}} = 350$ GeV. In general, MSSM was assumed. However, since the extended Higgs sector in most Supersymmetric models couples only to SM particles and a model is not explicitly chosen for Monte Carlo simulation, the results of this study can be applied to majority of models with extended Higgs sector. In MSSM at the decoupling limit[1] the coupling between charged Higgs bosons and gauge bosons are small and the coupling to fermions is dominant. Because the Yukawa couplings (Higgs couplings to fermions) are proportional to the mass of the fermions, the branching ratio of a charged Higgs boson with $m_{\mathrm{H}^{\pm}} > m_{\mathrm{t}} + m_{\mathrm{b}}$ to top and bottom quarks becomes dominant. In this study the branching ratio $BR(\mathrm{H}^{+} \to \mathrm{t}\bar{\mathrm{b}}) = BR(\mathrm{H}^{-} \to \bar{\mathrm{t}}\mathrm{b}) = 90\%$ was assumed. This leaves some space for decays to tau leptons or for smaller $\tan(\beta)$ decays to hW as well. $\tan\beta$ donates the ratio of vacuum exception values of the Higgs doublets. This has been chosen in consistency with [20]. The production cross section $\sigma(\mathrm{e}^{-}\mathrm{e}^{+} \to \mathrm{H}^{+}\mathrm{H}^{-})$ is assumed to be 9 fb. This is based on figure 1.2 which was taken from [22] and was interpreted for the considered mass.

Add diagram BR vs. mH+- with high and small tan beta

In the following signal refers to $\mathrm{e}^{-}\mathrm{e}^{+} \to \mathrm{H}^{+}\mathrm{H}^{-}$ where $\mathrm{H}^{\pm}$ decays into $\mathrm{b}\bar{\mathrm{t}}$ and $\bar{\mathrm{b}}\mathrm{t}$ respectively. Both t decay to W b. If both resulting W bosons decay into quarks, it will be referred to as hadronic signal (see Figure 1.1). If one $\mathrm{W} \to \ell\nu_{\ell}$ ($\ell = \mathrm{e},\mu$) and the other $\mathrm{W} \to q_{\mathrm{u}} q_{\mathrm{d}}$ ($q_{\mathrm{u}} = \mathrm{u,c}$ and $q_{\mathrm{d}} = \mathrm{d,s,b}$), it will be denoted by semi-leptonic signal.

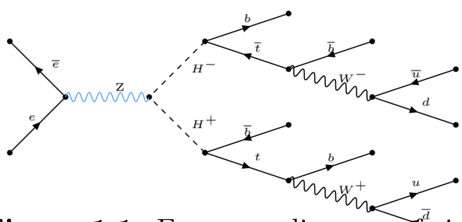

Figure 1.1: Feynman diagram of signal (hadronic channel)

[1]The decoupling limit denotes the situation with large mass of the CP-odd Higgs boson ($m_{\mathrm{A}} \to \infty$ or in a different way $m_{\mathrm{A}} \gg m_{\mathrm{Z}}$)

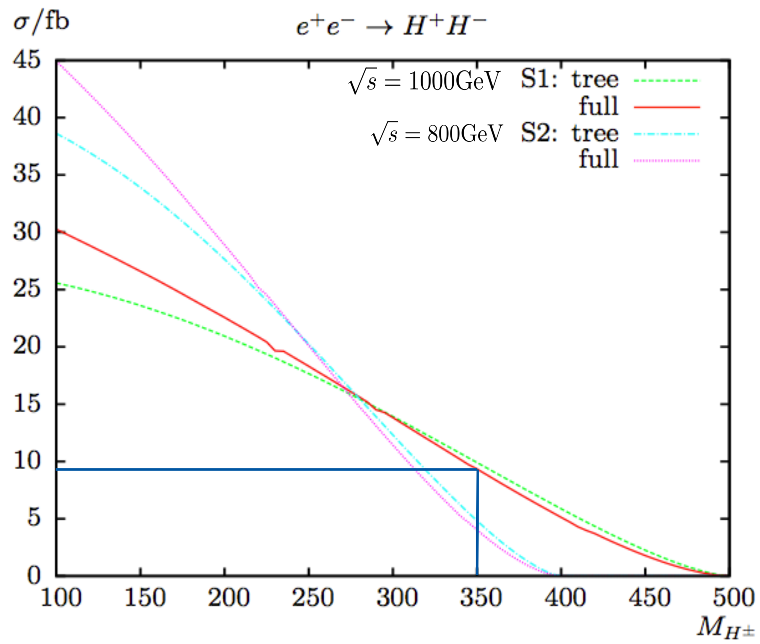

Figure 1.2: Tree-level and full one-loop corrected cross sections are shown for $\sqrt{s} = 1$ TeV and $\sqrt{s} = 800$ GeV with varied $m_{\mathrm{H}^\pm}$ (source: [22])

axis add "/GeV"

Current limits for charged Higgs bosons through direct search are from data of the Large Electron-Positron Collider (LEP). With CL 95% the limits $m_{\mathrm{H}^\pm} > 80$ GeV for type II 2HDMs and $m_{\mathrm{H}^\pm} > 72.5$ GeV for type I (from $\tau\nu$ and cs final states) were found [2]. The collision energy of LEP was $\sqrt{s} = 209$ GeV. The direct search for charged Higgs is limited by the accesible centre-of-mass energy, so translating this result naively to a linear collider with a $\sqrt{s} = 1$ TeV, a limit up to 400 GeV should be easily reachable.

Latest combined constrains from various experiments on the charged Higgs mass in different models can be found in [7]. For a wide range of models and $\tan\beta$ regions the tightest limit comes from the LEP search; in others models from flavor changing processes (typically for type II 2HDMs) the limit is around $m_{\mathrm{H}^\pm} \gtrsim 600$ GeV. This is because a light charged Higgs would have impact on flavor physics and various branching ratios of B mesons would be deviated. In some models and $\tan\beta$ regions the limit is from direct searches at the LHC over 1 TeV. This leaves a wide range of models and parameter regions to exclude at a future electron positron collider. However, the MSSM with type II 2HDM is already excluded with a charged Higgs boson mass of 350 GeV (see [7]).

Nevertheless the study here only chose the cross section from MSSM and the results are applicable to other models. Moreover, the developed methods can be transfered to higher $m_{\mathrm{H}^\pm}$ at electron positron collider with higher collision energy.

Since production cross-section [22] and the branching ratio [7] are compared to HA-channel relatively independent from $\tan\beta$, the $\mathrm{H}^-\mathrm{H}^+$-channel was chosen to be analyzed in this study. This is only true for the decoupling limit where $BR(\mathrm{H}^\pm \to \mathrm{hW}^\pm)$ becomes small. In addition

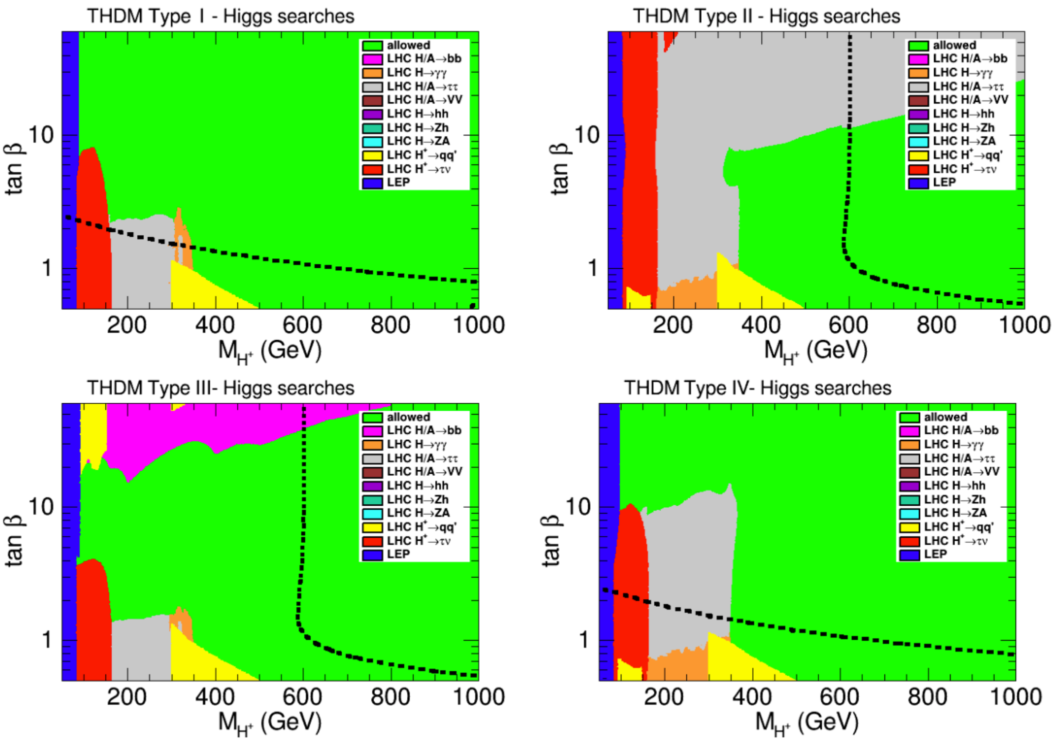

Figure 1.3: Constrains of $(m_{\mathrm{H}^{\pm}}, \tan\beta)$ parameter space of MSSM-like scenarios. The color coding corresponds to exclusion of 95 % C.L. by charged and neutral Higgs searches for the four different 2HDM types with different constraints, as given by the legend. The green region is allowed by all collider constraints. The dotted line frames the excluded area from flavor changing current observables, where the lower $\tan\beta$ side is excluded (source: [7])

the $\mathrm{H}^-\mathrm{H}^+$ production is interesting because $\tan\beta$ can be determined by the decay width of $\mathrm{H}^{\pm}$ [11]. Furthermore, pair production in general is a "clean" event where only the particles itself are produced and there are no byproduct. This simplifies the analysis and enhances precision. In addition, this channel has the opportunity to observe the CP-violation of the Higgs sector in branching ratio asymmetry, which is a possible explanation of baryon abundance. The CP-violation phase is defined as

$$\delta^{CP}_{f\overline{f}'} = \frac{BR(\mathrm{H}^+ \to f\overline{f}') - BR(H^- \to \overline{f}f')}{BR(\mathrm{H}^+ \to f\overline{f}') + BR(H^- \to \overline{f}f')}$$

check dependence on tan beta

$\delta^{CP}_{f\overline{f}'}$ depends on $\tan\beta$ [18][17].It is accessible in the semi-leptonic mode with the lepton charge and in the leptonic mode where both W bosons decay to lepton and neutrino pair. In the hadronic decay it may be reconstructible through the charge of the bottom jets.

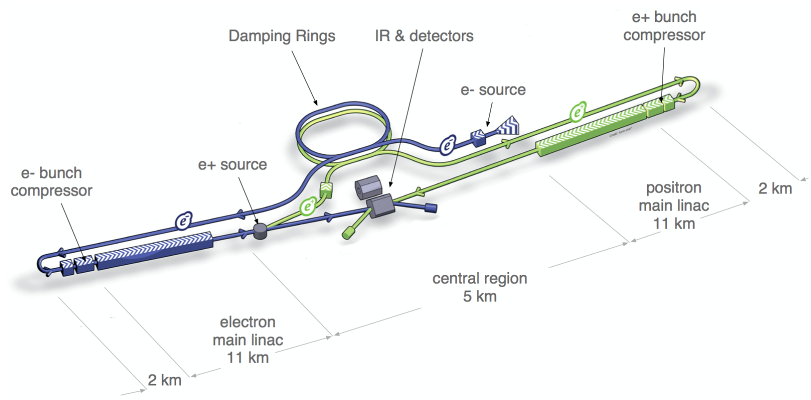

Figure 1.4: Schematic representation of the ILC [12]

1.3 International Linear Collider and International Large Detector

The International Linear Collider (ILC) (see figure 1.4) is a proposed electron positron collider with a tunable in center of mass energy in the range of 250 GeV to 500 GeV and can be upgraded to reach up to 1 TeV. The ILC evolved out of three projects, the Japanese GLC, European TESLA-collider and American NLC, and is now supported by the worldwide particle physics community. The ILC is planed to be constructed in Iwate prefecture in northern Japan. In 2013 the technical design report was published which reports detailed about the accelerator, detector and physics outcome of the project ([12][8][4][5][3]). At the ILC in comparison to a proton collider such as the LHC one needs fewer model assumptions, there is fewer background and the initial state is well known. It is even possible to polarize 80 % of the electron beam and 30 % of the positron beam. At a collision energy of 1 TeV the positron polarization is expected to lower to 20 %.
To ensure a cross check of the measurement the ILC will have two Detectors, the International Linear Detector (ILD) and the Silicon Detector (SiD), which will share the same interaction region by push-pull technique. In this analysis only the ILD is considered. It consists of a high-precision vertex detector surrounded by a hybrid tracking system with a silicon tracker and time-projection chamber. For optimal particle-flow performance a highly granular electromagnetic and hadron calorimeter system was developed. The whole detector barrel is contained by a 3.5 T solenoid [12].
The exact operation plan of the ILC will be decided from funding and discoveries in particle physics. The collision energy is relatively easy to adjust, so that depending on discoveries of the Large Hadron Collider (LHC) at CERN or other experiments the energy can be adjusted. A possible running scenario could be

- 91 GeV: Z boson peak for calibration and precise measurements of Z properties
- 160 GeV: $W^{\pm}$ production for precise measurements of W properties

- 250 GeV: Higgs factory through Higgs-Strahlung
- 350 GeV: Top quark factory through pair production
- 500 GeV: Top Yukawa coupling, BSM search, fermion pair production and Higgs through W-fusion
- 1 TeV: BSM search

This should not be an exclusive list but rather a quick overview on interesting physics accessible at a linear electron positron collider. 1 TeV as center mass energy is rather arbitrary but would give a new view on otherwise not accessible energy regions and gives a first mark on where to look at.
The accelerator of the ILC will be based on 1.3 GHz superconducting radio-frequency accelerating technology. The initial ILC will have a length of 31 km which can be extended to 50 km. With this length the ILC can reach 1 TeV or more. In the TDR a scenario A was proposed for 1 TeV [12] the luminosity is expected to be $L = 3.6 \cdot 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In this analysis an integrated luminosity is assumed to be $\mathcal{L} = 1\ \mathrm{ab}^{-1}$. This accounts for 324 days of running. Which calls for about three years of running at $\sqrt{s} = 1$ TeV considering service time.

1.4 Simulation and Reconstruction

In this study Monte Carlo data samples generated by Physsim and Whizard are analyzed. The signal of charged Higgs pair production is generated by Physsim [27] which is based on HELAS [28] for matrix element calculation. The SM background was generated by Wizard 95 . Parton shower and hadronization was performed by Pythia 6.4 . The beam spectrum is simulated by GuineaPig [30] and is incorporated in both the signal and background generators. In addition to the main event all data samples are overlaid with in average 4.1 events of $\gamma\gamma$ to hadron events with low transversal moment. This type of beam-induced background will be addressed further in chapter 2.3.1. The detector was simulated with Mokka on a full ILD model based on the Detailed Baseline Design (DBD) [3]. For reconstruction the Pandora Particle Flow Algorithm (PandoraPFA) was used. Pandora Particle Flow Algorithm uses the tracker detector to determine momentum of charged particles and only uses the calorimeter for energy determination of neutral particles. This improves the jet resolution and allows better separation of W, Z, H bosons and top quark by their invariant mass [32].
In the frame of this study the FastJetFinder [13] was used for beam-induced background reduction, LCFIplus [31] package for vertex reconstruction and flavor tagging accessed through Marlin [34]. For this analysis a dedicated Marlin processor was written for the data analysis. The output of the analyzer was stored to ROOT-files. After the event by event analysis with

add Wizard reference
add Pythia reference
add Wizard reference
add PandoraPFA reference
check name

add pyroot reference

Marlin ROOT [6] accessed with pyroot is then used for final analysis.
For computing the KEK Central Computer System [24] was used.
As Background only SM processes including various SM-like Higgs events in all final states are considered. Beam photon interactions, which include $\gamma\gamma$-annihilation and interactions with beam electrons or positrons were considered as well. A detailed list of data samples can be found in Table A.1.

2 Data Analysis

2.1 Analysis Strategy

All data samples used in this study are scaled to an integrated luminosity of $\mathcal{L} = 1\ \mathrm{ab}^{-1}$. The polarization of both beams is included as $P(\mathrm{e}^-, \mathrm{e}^+) = (-80\%, 20\%)$ [12]. The used samples in this analysis had two polarizations $P_L = (-100\%, 100\%)$ and $P_R = (100\%, -100\%)$. In order to obtain samples of correct polarization the weights are assigned to optain samples of correct polarization. The weights were assigned to the corresponding sample

$$w_{L,i} = \mathcal{L} \cdot 0.9 \cdot 0.6 \cdot \sigma_i \text{ and } \ w_{R,i} = \mathcal{L} \cdot 0.1 \cdot 0.4 \cdot \sigma_i$$

where $w_{L,i}$ stands for the weight of the data sample of process i where the electron is left handed and positron is right handed. $w_{R,i}$ is the weight for samples with opposite polarization, and σ_i are the corresponding cross sections. Processes with other polarization are weighted in an analogous manner. A full list of all samples with responding weights, expected number of events and generated number of events can be found in table A.1.

A flow diagram of the event by event based analysis used Marlin processors is shown in figure 2.1. For the hadronic mode the kt-algorithm with requesting eight jets (FastJet_kt_8) is used to reduce beam background, while for the semi-leptonic mode first a lepton is removed into a separate collection before kt-algorithm with requesting six jets (FastJet_kt_6) is used. The clustered event gets restored into tracks in a intermediate step (JetPFOs). Then vertex reconstruction (VertexFinder) and final jet clustering and b-tagging (JetClustering And FlavorTag) is done. Finally all relevant collections are analyzed and relevant observables are saved into a ROOT file (h2dmAnalysis).

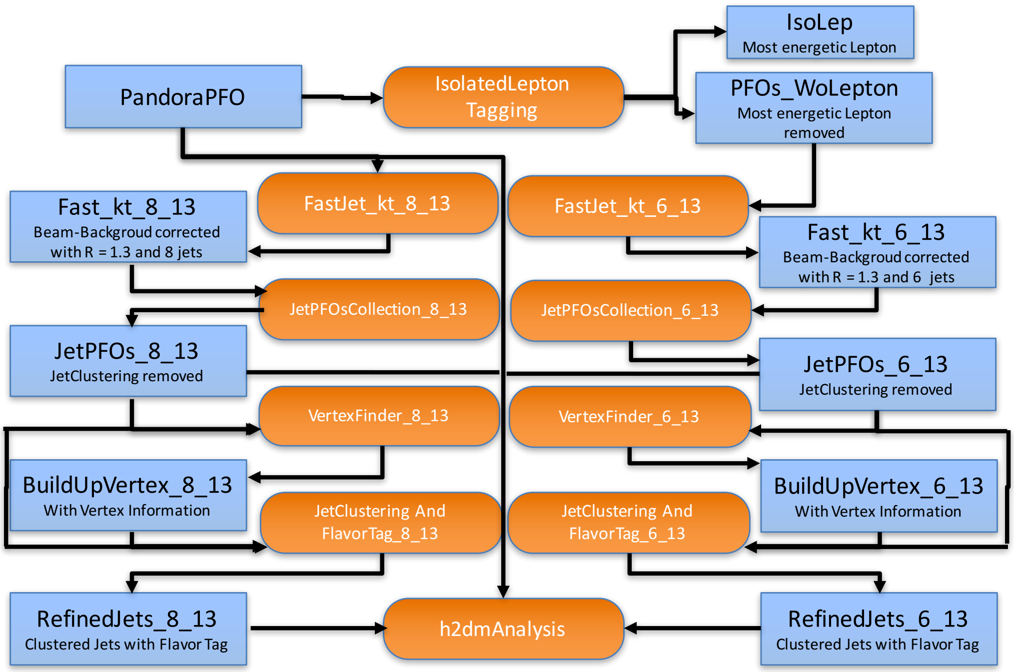

Figure 2.1: Schematic diagram of processor structure

make simpler and add if lepton found -> sl / no lep found -> h

2.2 Lepton selection

For the lepton selection the `IsolatedLeptonTaggingProcessor` [25] which is included in the MarlinReco package [34] since version `v01-12`. This processor uses the TMVA package (Toolkit for Multivariate Data Analysis [23] integrated in ROOT) to select one isolated lepton. Here weights[1] trained on four fermion processes at $\sqrt{s} = 500$ GeV because there are no weights trained on $\sqrt{s} = 1$ TeV available. Nevertheless the tagging efficiency is around 90 % high level (for details see table 2.1).

It is important to select the isolated lepton before the beam background reduction (chapter 2.3.1) because the used kt-algorithm is requiring six jets and removes particles which are far from those jets. So isolated leptons will be removed in some events. On the other hand it is very unlikely to select a particle of the beam background as isolated lepton. The lower efficiencies for the isolated lepton selection in table 2.2 is proving this.

In 2 % of the hadronic signal an isolated lepton is mistakenly selected. The reason for this unexpected large ratio maybe the weights trained on $\sqrt{s} = 500$ GeV because with larger energy in the event leptons in the jets have larger energy and might be selected mistakenly.

[1] `weights_isolated_electron_llh_gg_bbbb_500` and `weights_isolated_muon_llh_gg_bbbb_500` located at `/home/ilc/tianjp/analysis/PostDBD/IsolatedLeptonTagging/weights/`

	correct lepton selected	other particle selected
e	89.9 %	0.5 %
μ	90.6 %	0.4 %
τ	9.1 %	1.9 %

Table 2.1: Table of isolated lepton selection efficiencies; e stands for the semi-leptonic signal where $W \rightarrow e\nu_e$; τ and μ have analogous mining (isolated lepton selection is done before beam background removal)

	correct lepton selected	other particle selected
e	86.9	0.6
μ	88	0.45
τ	8.2	1.8

Table 2.2: Table of isolated lepton selection efficiencies; all numbers are given in percent, e stands for the semi-leptonic signal where $W \rightarrow e\nu_e$; τ and μ have analogous mining (beam background removal is done before isolated lepton selection)

2.3 Jet Reconstruction

2.3.1 Hadronic Beam-Induced-Background

The beam particles are bend under the electro-magnetic field of the oncoming beam and thus radiate photons. This is referred to as beamstahlung. In general these photons can react to produce e^+e^- pairs, most of which are very close to the beam line and get not detected by the main detector but are problematic in terms of radiation damage for materials and apparatuses in forward region.

In order to increase the luminosity at linear colliders an great effort has to be maid to focus the beams into a very small transverse size to collide. Thus the approaching beams are exposed to very large electro-magnetic field of the opposite bunch. The bunches are attracted to the center of the oncoming bunches of opposite charge and get focused even stronger which increases the luminosity. This is called pinch effect. Though relativistic effects the pinch effect becomes stronger with higher energy, which boosts on the one hand luminosity even more but on the other hand beamstrahlung as well.

The photons from beamstrahlung produce as well to quark pairs which effects this analysis and has a large impact on the resolution because of their high energy. In average 4.1 of these events were expected per bunch crossing for ILC at 1 TeV but a new not yet published study suggests a lower rate of 2.7[10]. Nevertheless a an average of 4.1 events where overlaid to the here used data samples.

These quark pairs, hadronising to various mesons, are in this study reduced with the kt-algorithm of the FastJetFinder ([13], [14]). This method was adapted from similar studies (e.g. [29]).

Generally speaking the kt-algorithm clusters all tracks to a requested number of jets. If a

track is closer to the beam line than to the closest jet, the track gets removed. To calculate the distance to the jet a generalized Radius R is used. This R value is used to optimize how many particles get removed.
In detail the kt-algorithm follows this steps:

1. Calculate the distance between all tracks

$$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2)\frac{\Delta R_{ij}}{R}$$

where $\Delta R_{ij} = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$, η is the pseudo rapidity and ϕ the azimuth (angle perpendicular to beam pipe) and p_{Ti}^2 is the transverse momentum of track i.

2. Find smallest d_{ij}

   a) If $d_{ij} < d_{iB} = p_{Ti}^2$, merge tracks

   b) If not, remove Track i
   (d_{iB} is the distance between track i and beam line)

3. Continue with the first step until there are only N tracks, where N is the number of requested jets [13]

The kt-algorithm applied in the range of 0.1 to 1.5. Then the now clustered event gets restored to tracks and reclustered by the Durham algorithm accessed through SatoruJetFinder from the MarlinReco package. The SatoruJetFinder rather than the LCFIplus was used because the computing time of LCFIplus is much longer but the clustering result is similar to the Durham algorithm.
To estimate which of the R values is appropriate for this study the mass of both charged Higgs bosons is calculated. To do so one of the four color singlets is connected to each jet as following: the color singlet which gave the largest contribution to the jet in terms of energy is assigned. This is done with generator information and can not be known in the real experiment.
Now the events are classified to three categories:

a) If all four color singlets have each two jets assigned, the assignment is final (good clustering)[2]

b) If color singlet k has only one jet assigned and color singlet j has three jets assigned, the jet with highest $\frac{E_{k,i}-E_{j,i}}{E_{k,i}+E_{j,i}}$ is reassigned to color singlet k (moderate clustering) where $E_{l,i}$ denotes the energy of jet i resulting from color singlet l

c) In other cases the event gets discarded for this calculation (failed clustering)

[2]occurrence is shown in table 2.3

Since from generator information it is known which color singlet originated form which charged Higgs, m_{H^+} is defined as the invariant mass of the jets assigned to the two color singlets from H^+. The invariant mass of the other four jets is m_{H^-}. For the events with failed clustering the relation between color singlet and jet stays unknown and the masses can not be reconstructed. Therefore, these events are discarded for this chapter. As one can see in

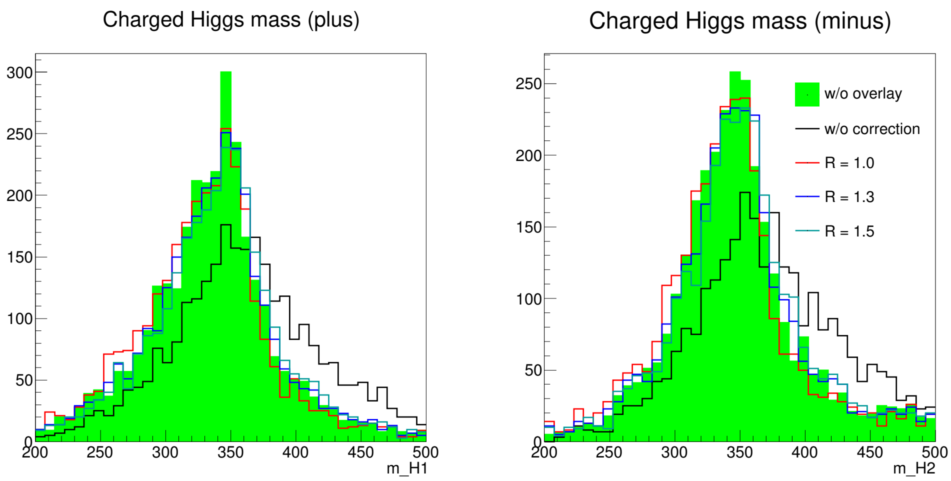

Figure 2.2: Charged Higgs mass (right: m_{H^+}, left: m_{H^-}) in green $\gamma\gamma$-background removed by generator in formation, black without any correction and other colors with corrected with kt-algorithm with varied R as noted in the legend (see figure A.1 for more values for R)

y-axis add "events /bin" and x-axis "/GeV" change to m_H+/-

figure 2.2 the contribution of the $\gamma\gamma$-background on the Higgs mass can be reduced with the used kt-algorithm. If the generalized radius R is chosen too small tracks from the real event tent to get removed. Thus energy in the event is missing and the reconstructed Higgs mass becomes smaller. Values between 1 and 1.3 for R were found to be appropriated. To avoid removing tracks of the real event a relatively high value of $R = 1.3$ was be chosen. This is consistent with an earlier study of the top-Yukawa-coupling at 1 TeV where the same final state was analyzed where $R = 1.2$ was chosen [29]. The influence of the background removal on jet pairing, b-tagging and clustering is shown in table 2.3. Jet pairing, b-tagging and clustering will be treated in the next chapters (2.3.3 and 2.3.2)

add plo of miss- clustere clustere $\gamma\gamma$- backgro and cor- rect clustere energy maybe not do i

The background removal with kt-algorithm was only studied on hadronic signal. Nevertheless the $\gamma\gamma$-background is corrected as well for semi-leptonic background in the same manner. After the lepton selection the kt-algorithm is run on the rest of the event while requesting six jets and setting $R = 1.3$.

2.3.2 Jet Clustering

The LCFIplus package [31] is used for the final jet clustering. LCFIplus uses the LCFIVertex package [9] and improves the clustering utilizing vertex information. At the same time LCFIplus provides a bottom quark likeliness called b-tag for every requested jet. The b-tagging is done with TMVA package and is essential in this study for the jet pairing and event selection (chapter 2.3.3 and 2.5). LCFIplus is using pretrained weights to calculate the b-tag values. Here the `6q1000_v02_p01` was used, which has been trained on events with six jet at $\sqrt{s} = 1$ TeV, however they are used for both hadronic and semi-leptonic mode because of the lack of weights trained on eight jet events.

2.3.3 Jet pairing

The jet pairing is performed with a chi square minimization. The here used χ^2 is defined as

$$\chi^2 = \left| \frac{(m_{j_1 j_2 j_3 j_4})^2 - (m_{j_5 j_6 j_7 j_8})^2}{2\sigma^2_{\mathrm{H}^\pm}} \right| + \left(\frac{m_{j_2 j_3 j_4} - m_{\mathrm{t}}}{\sigma_{\mathrm{t}}} \right)^2 + \left(\frac{m_{j_6 j_7 j_8} - m_{\mathrm{t}}}{\sigma_{\mathrm{t}}} \right)^2 + \left(\frac{m_{j_3 j_4} - m_{\mathrm{W}}}{\sigma_{\mathrm{W}}} \right)^2 + \left(\frac{m_{j_7 j_8} - m_{\mathrm{W}}}{\sigma_{\mathrm{W}}} \right)^2 \tag{2.1}$$

where j_1, j_2, j_5 and j_6 are b-jets and j_3, j_4, j_7 and j_8 are light jets from W decays. $\sigma_{\mathrm{H}^\pm}$ and σ_{t} have been chosen to 80 GeV and σ_{W} to 48 GeV. These values are taken from the width of the relevant mass distributions with the described jet pairing method in chapter 2.3.1 using generator information. In the first term of χ^2 for the Higgs mass, the difference of the two masses were introduced, rather than the deviation to the expected mass in order to not be biased towards the expected mass.

The total combinations of the eight jets is $N = 8! = 40320$. In order to obtain better quality of the jet pairing and reduce the number of possible jet pairing combinations, the following conditions are applied:

- The four jets with highest b-tag are required to be the jets from bottom quarks. This reduces the combination to $N = 4!^2 = 576$.
- Without exchanging the jets from a given W boson and without exchanging the two Higgs bosons with each other the combinations reduce to $N = \frac{4!^2}{2^4} = 36$.

With this reduced number of options the computing time is unproblematic and furthermore, the risk of getting a small χ^2 for a wrong combination is low.

From the method explained in chapter 2.3.1 the underlaying color singlet of the jets is known and can be compared to the χ^2 pairing. If the pairing agrees, it will be called correctly paired. About one quarter of the hadronic signal is correctly paired (see table 2.3).

In the case of semi-leptonic signal the same χ^2 pairing is used but the jets j_7 and j_8 are required to be the lepton four momentum and neutrino four momentum. The reconstruction of the neutrino will be treated in chapter 2.4.

	Uncorrected[†]	$R = 1.3$[††]	no $\gamma\gamma$-BG[†††]	Description
b-tag	38.0 %	42.5 %	44.6 %	The four b-jets have highest b-tag in the event
good clustering	40.2 %	49.5 %	50.7 %	As defined in chapter 2.3.1
working clustering	92.5 %	95.6 %	95.8 %	good and moderate clustering from chapter 2.3.1
correctly paired	17.2 %	24.5 %	27.8 %	Jet pairing agrees with major color singlet fraction in jet

† Overlay removed with generator information
†† Beam background corrected with kt-algorithm where $R = 1.3$
††† Without any correction

Table 2.3: Table of clustering, b-tagging and pairing efficiencies; all numbers are given in percent

add semi-leptonic

Even so the b-tagging efficiency is very high there are a number of events with low b-tags. For most of these events the clustering rather than pairing or b-tagging is problematic. Before two categories (b-jets and light jets) of jets were defined. However, for these events with bad clustering a more realistic pairing can be reached, if the following three categories are defined:

- 1. b-jets (with highest b-tag)
- 2. light jets (lowest b-tag)
- 3. unknown flavor jets (with medium b-tag)

Here the combinations are with two jets in third category $N = \frac{8*7*3!^2}{2^4} = 126$ or with four jets in third category $N = \frac{8*7*6*5*2!^2}{2^4} = \frac{8!*2^2}{4!*2^4} = 420$.
This method becomes effective for events with low b-tag but those events will be rejected by the background suppression (chapter 2.5) later on and has therefore no effect on the final result.

Another method to improve the jet pairing is the optimization of χ^2 or adding other terms. Therefore the following χ^2_{optim} was tested

$$\chi^2_{\text{optim}} = w_{\text{H}} \left| \frac{(m_{j_1j_2j_3j_4})^2 - (m_{j_5j_6j_7j_8})^2}{2\sigma^2_{\text{H}^\pm}} \right| + w_{\text{t}} \left(\frac{m_{j_2j_3j_4} - m_{\text{t}}}{\sigma_{\text{t}}} \right)^2 + w_{\text{t}} \left(\frac{m_{j_6j_7j_8} - m_{\text{t}}}{\sigma_{\text{t}}} \right)^2$$
$$+ w_{\text{W}} \left(\frac{m_{j_3j_4} - m_{\text{W}}}{\sigma_{\text{W}}} \right)^2 + w_{\text{W}} \left(\frac{m_{j_7j_8} - m_{\text{W}}}{\sigma_{\text{W}}} \right)^2 + w_\theta \left(\frac{\theta_{\text{H}^+\text{H}^-} - \pi}{\sigma_\theta} \right)^2$$
$$+ w_{\cos} \left(\frac{1 - \cos\theta_{\text{H}^+\text{H}^-}}{\sigma_{\cos}} \right)^2 + w_E \left(\frac{E_{\text{H}^-} - E_{\text{H}^+}}{\sigma_E} \right)^2$$

with

$$E_{\mathrm{H}^-} = \sum_{i=1}^{4} E_{ji} \quad \text{and} \quad E_{\mathrm{H}^+} = \sum_{i=5}^{8}$$

where E_{ji} is the energy of jet i. $\theta_{\mathrm{H}^+\mathrm{H}^-}$ is the production angle between the charged Higgs bosons formed by the reconstructed jets. The different widths were chosen to

$$\sigma_\theta = 0.3,\ \sigma_{\cos} = 0.18,\ \sigma_E = 117\ \mathrm{GeV}$$

with the same method as mentioned before.
By optimizing two of the weights at the same time the following optimal choice was found:

w_H	w_W	w_t	w_θ	$w_{\cos}$	w_E
1	2	3	0.6	0	0

This improves the pairing efficiency by about 1.6 % from 24.5 % to 26.1 %. The effect of this on the final result was not checked because of lack of time but the effect is expected to be small because the improvement is small, too. This was again only studied for hadronic signal.

In order to check jet clustering and pairing a 3D-display was developed for visual inspection on a event by event bases. In figure 2.3 an event where two jets got clustered to one and another jet got split in two. This kind of events are common but in most events with bad clustering it is difficult to figure out what is going on because of the large number of jets.

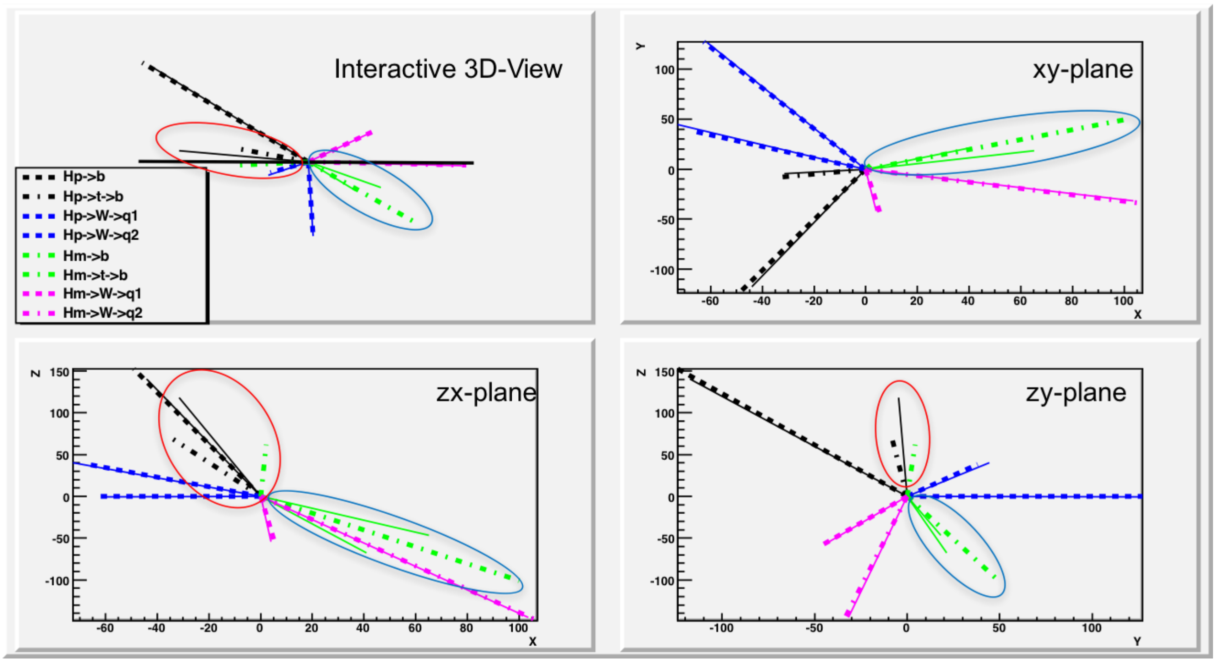

Figure 2.3: 3D-display to check jet clustering and pairing event by event. In the running program the upper left quarter can be turned with the mouse. The other quarters show the projection on one plane as noted. The dashed lines are displayed from simulator information and the solid lines show the reconstructed jets. Here the display shows an event where two jets got clustered to one and another jet got split in two. The ovals are added to indicate the issue.

2.4 Neutrino Reconstruction

For the neutrino reconstruction four methods where tested. Since the neutrino can not be detected it's four momentum has to be calculated from the missing momentum and energy in the event. The largest uncertainties for this is the beam spectrum, missing momentum from other neutrinos in the jets, beam background and beam background reduction.

2.4.1 Missing Energy Method (MEM)

The idea in MEM is simply using total four momentum of the event p_{vis} and subtract it from the momentum of center of mass system (CMS) p_{CMS}. Because the crossing angle will be 14 mrad and the collision energy 1 TeV[12], it is given as

$$p_{\text{CMS}} = (1\ \text{TeV}, 0, 0, 1\ \text{TeV} \cdot \sin(0.014/2))$$

p_{vis} is simply the sum over all Particle Flow Objects (PFO) which are the tracks

$$p_{\text{vis}} = \sum_{i=1}^{N_{\text{PFO}}} p_i$$

Thus the neutrino four momentum can be written as

$$p_{\nu,\mathrm{MEM}} = p_{\mathrm{CMS}} - p_{\mathrm{vis}} \quad (2.2)$$

This method is typically used for ILC analysis.

2.4.2 Missing Momentum Method (MMM)

This is a slight modification of MEM. Because the momentum resolution is better than the energy resolution and the neutrino is massless, the relation $E = p$ is adopted. Therefore we can write the neutrino four momentum as

$$p_{\nu,\mathrm{MMM}} = (|\vec{p}_{\nu,\mathrm{MEM}}|, \vec{p}_{\nu,\mathrm{MEM}})$$

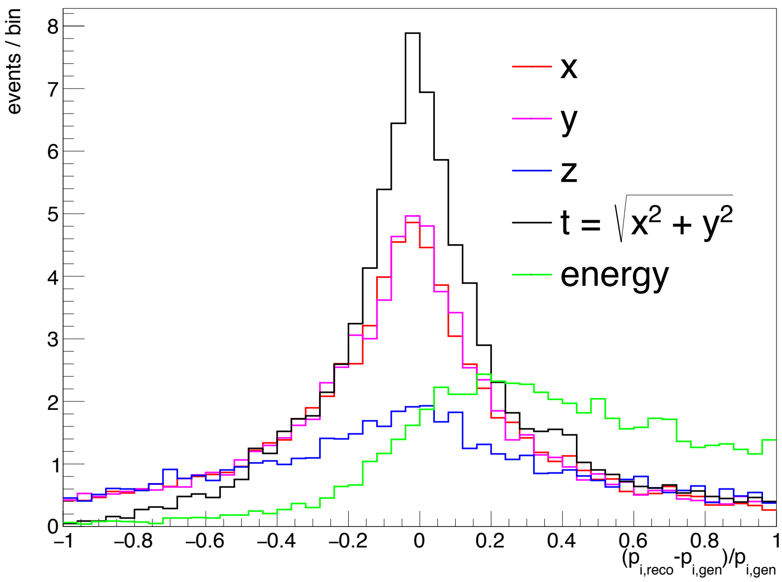

Figure 2.4: Comparison of the deviation between generated and reconstructed by MEM for the momentum components and the energy, where x and y stands for the momentum in x-direction and y-direction respectively and t is the transverse momentum

2.4.3 Missing Direction Method (MDM)

The invariant mass of the neutrino lepton system is the W boson mass. This information can be used to improve the resolution. Because the resolution of the direction of missing momentum is better than the energy resolution, the reasoning of the MDM is to calculate the neutrino Energy E_ν from the mass restraint. (compare figure 2.4)

We can write the W mass as

$$m_{\mathrm{W}}^2 = (p_\nu + p_\ell)^2 = p_\nu^2 + 2p_\nu p_\ell + p_\ell^2 = 2p_\nu p_\ell$$

where p_ν and p_ℓ denote the four momenta of neutrino and lepton respectively. The assumption $p_\nu^2 = p_\ell^2 = 0$ was used, which is obvious for neutrinos and reasonable for leptons, since muon and electron momentum are much larger than the mass.
Simplifying farther, we get

$$m_{\mathrm{W}}^2 = 2(E_\nu E_\ell - \vec{p_\nu}\vec{p_\ell}) = 2E_\nu E_\ell(1 - \cos\theta) \tag{2.3}$$

the assumption of negotiable mass was applied again in from of $E_i = |\vec{p}|$ and $\cos\theta$ is the decay angle of neutrino and lepton

$$\cos\theta = \frac{\vec{p}_{\nu,\mathrm{MEM}} \cdot \vec{p_l}}{|\vec{p}_{\nu,\mathrm{MEM}}||\vec{p_l}|}$$

now we can solve for E_ν and get the estimate of this method of the neutrino energy as

$$E_{\nu,\mathrm{MDM}} = \frac{m_{\mathrm{W}}^2}{2E_l(1 - \cos\theta)}$$

and the four momentum as

$$p_{\nu,\mathrm{MDM}} = (E_{\nu,\mathrm{MDM}}, E_{\nu,\mathrm{MDM}}\frac{\vec{p}_{\nu,\mathrm{MEM}}}{|\vec{p}_{\nu,\mathrm{MEM}}|})$$

An additional uncertainty of this method comes from the W width but is small in comparison to the uncertainty on the direction of missing momentum.

2.4.4 Missing Transversal Momentum Method (MTMM)

In this method the idea is to use only the missing momentum of the event in transversal direction orthogonal to the beam pipe. Looking at figure 2.4 it is easy to see that the resolution of the transversal direction is better than in z-direction for a number of reasons.

- **Beam background**: As discussed in chapter 2.3.1, beam background is mainly in forward direction as well as the beam background reduction, discussed in the same chapter. Remaining beam background or removed tracks from the main event contribute largely to the resolution in z-direction.

- **Beam spectrum**: The variance in the z-component of the beam electron and positron are much larger than in transverse components.

- **Undetected particles**: Particles of the main event can in general get lost in the beam pipe. Furthermore, the detectors in the barrel have a better accuracy than in the caps.

Equation 2.3 is reused as follows

$$\frac{m_{\mathrm{W}}^2}{2} = E_\nu E_\ell - \vec{p_\nu}\vec{p_\ell} = E_l\sqrt{p_{\nu x}^2 + p_{\nu y}^2 + p_{\nu z}^2} - p_{\nu x}p_{\ell x} - p_{\nu y}p_{\ell y} - p_{\nu z}p_{\ell z}$$

where $p_{\mathrm{p}i}^2$ donates the component i of p's momentum.
This is a fairly complicated polynomial second grade. Nevertheless it can be solved with the quadratic formula for the neutrino momentum in z-direction $p_{\nu z}$. The solution was found to be

$$p_{\nu z} = \frac{\pm K + p_{\ell z}[2(p_{\ell y}p_{\nu y} + p_{\ell x}p_{\nu x}) + m_{\mathrm{W}}^2]}{2(p_{\ell x}^2 + p_{\ell y}^2)}$$

with

$$K = E_l\sqrt{4[(2p_{\ell x}p_{\nu x} + m_{\mathrm{W}}^2)p_{\ell y}p_{\nu y} - p_{\ell x}^2 p_{\nu y}^2 - p_{\ell y}^2 p_{\nu x}^2 + m_{\mathrm{W}}^2 p_{\ell x}p_{\nu x}] + m_{\mathrm{W}}^4}$$

It has two solutions. In this study the solutions closer to the z-component of MEM $p_{\nu,\mathrm{MEM}z}$ is selected. Theoretically the square root in K can not become imaginary but from uncertainties there are cases where it would become imaginary. To prevent that the absolute value is used.

In figure 2.5 the energy deviation and deviation in z-component of the momentum to the generated value is shown for the methods explained. When comparing the methods MTMM is the best in the momentum but in the energy deviation MMM is a little better. Very badly reconstructed events can have a very large deviation from the real value for MTMM and MDM because of error evolution. However, MEM and MMM are stable for even those events. Since MEM and MMM are the exact same in the momentum but MMM is much better in the energy reconstruction, MMM was chosen for the further analysis. Furthermore, MMM does not fix the W mass and leaves the opportunity to use this value for the further analysis. Nevertheless MTMM could be a good alternative for most events.

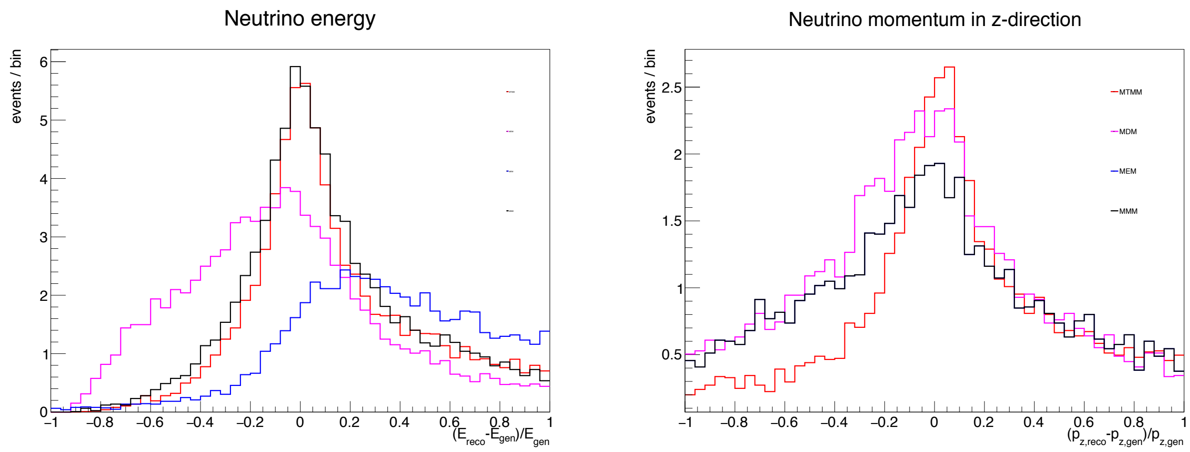

Figure 2.5: Comparison of the four methods for neutrino reconstruction; right figure shows deviation between reconstructed and generated energy; left figure shows the deviation for the z-component of the momentum

2.5 Event Selection

The event selection is optimized for maximal significance which is defined as

$$S = \frac{N_S}{\sqrt{N_S + N_B}}$$

where N_S is the number of signal events and N_B the total number of background events. In a simple counting experiment the statistical uncertainty would be the inverse of the significance.

$$\frac{\Delta N}{N} = \frac{1}{S}$$

One can optimize the event selection on the signal significance or on the correctly paired signal significance using the definition of correct pairing from chapter 2.3.3. In this study both has been tried out. In the case of optimization for correct pairing other signal was not added to N_B. Beside that hadronic signal was not considered as background when optimizing semi-leptonic signal and vice versa.

2.5.1 Static Cuts

The cuts in this chapter have been inspired by a similar study on charged Higgs bosons at the proposed Compact Linear Collider (CLIC) [26]. The cuts are shown in table 2.4 (2.5) for optimization for hadronic (semi-leptonic) signal significance and in table 2.6 (2.7) for optimization for correctly paired hadronic (semi-leptonic) signal significance.
In the following the cuts will be briefly explained.

- **(no) IsoLep** donates to isolated lepton selected as described in chapter 2.2

- **4 highest b-tag** is the sum of the highest four b-tags in the event. (see chapter 2.3.2)
- $\boldsymbol{E_{\text{vis}}}$ is defined as $E_{\text{vis}} = \sum_{i=1}^{N_{\text{PFO}}} E_i$ where N_{PFO} is the number of tracks in the event after beam background reduction and E_i is the reconstructed energy of track i. In case of the semi-leptonic mode the energy of the lepton is added as well to E_{vis}.
- $\boldsymbol{\chi_{\text{H}^\pm}}$ is the first term of the χ^2 used for jet pairing in equitation 2.1

$$\chi_{\text{H}^\pm} = \left| \frac{(m_{j_1 j_2 j_3 j_4})^2 - (m_{j_5 j_6 j_7 j_8})^2}{2\sigma_{\text{H}^\pm}^2} \right|$$

- $\boldsymbol{\chi_{\text{t}}}$ is the top quark related term of the χ^2 used for jet pairing in equitation 2.1

$$\chi_{\text{t}} = \left(\frac{m_{j_2 j_3 j_4} - m_{\text{t}}}{\sigma_{\text{t}}} \right)^2 + \left(\frac{m_{j_6 j_7 j_8} - m_{\text{t}}}{\sigma_{\text{t}}} \right)^2$$

- $\boldsymbol{y_{n(n+1)}}$ is provided by the LCFIplus package and obtained by the Durham algorithm which is briefly explained in chapter A.1. $y_{n(n+1)}$ is y_{cut} by the transition of $n+1$ to n requested jets.
- **Thrust cuts:** MinorThrust, PrincibleThrust and cosThrustAxis are provided by the ThrustReconstruction processor of MarlinReco. They are variables of the event shape or in other words the distribution of momentum in the space.
- $\boldsymbol{m_{\text{miss}}}$ is the missing mass in the event.

$$m_{\text{miss}} = \sqrt{(p_{\nu,\text{MEM}})^2}$$

$p_{\nu,\text{MEM}}$ was defined in equation 2.2

2.5.2 Boosted Decision Trees

The TMVA from ROOT was used as an alternative event selection. The boosted decision trees (BDT) and Boosted Decision Trees with gradient boosting (BDTG) were found to be the best methods for this purpose. To replace the event selection with static cuts very similar input values as the cut values as in the previous chapter were used. Only 4 highest b-tags was divided in two highest b-tags and next tow highest b-tags as well as χ_{t} was divided into its summands. As a preselection the "No IsoLep" criteria was used in the hadronic mode and "IsoLep" was used for semi-leptonic mode. BDT was found to be the best method. The results are shown in figure 2.6.

As a secondary background suppression especially to suppress background with same final

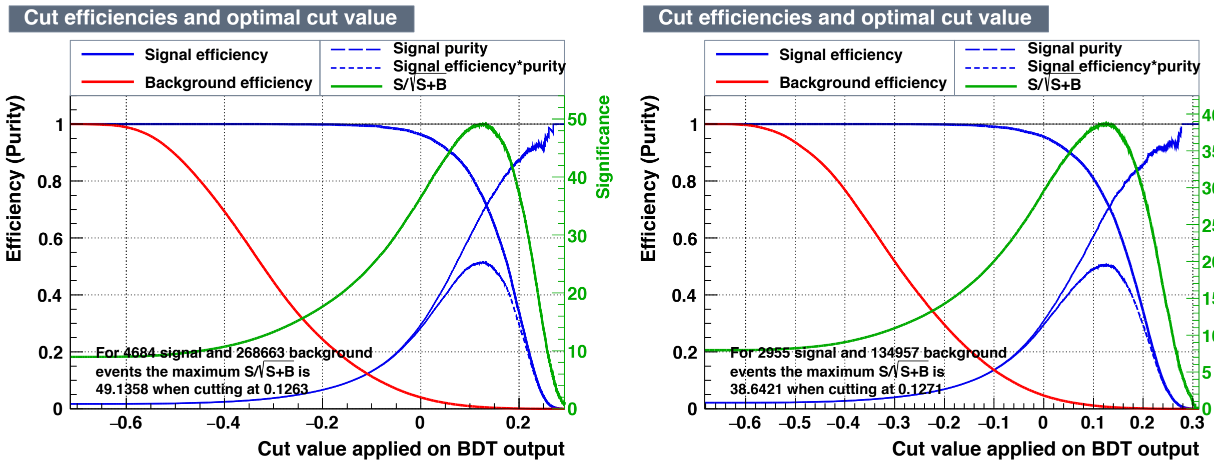

Figure 2.6: Results of primary BDT event selection for hadronic (left) and semi-leptonic signal (right)

state a second selection was trained after applying the static cuts from chapter 2.5.1. The input values were chosen to separate same final state signal.

- Invariant mass and decay angle of
  - Bottom quarks system
  - Top quarks system
  - Higgs bosons system
- Thrust information namely:
  - PrincipleThrust
  - MajorThrust
  - MinorThrust
  - CosThrustAxis
- y_{34}
- E_{vis}
- Energy of the top quarks
- Number of charged tracks in the event
- χ^2 (as defined in equation 2.1)
- Third and fourth highest b-tag
- $m_{\text{miss,t}}$
- Difference of momenta of bottom quarks

- Difference of momenta of Higgs bosons

For this secondary event selection BDTG showed an advantage over BDT. However, training and applying to improve the signal significance does not show an relevant effect over the primary selection with BDT on the other hand training and applying it on correctly paired signal significance shows an effect. The results are shown in figure 2.7. The main reason for this behavior is probably the large fraction of miss-clustered and miss-paired signal and the indistinguishably of this signal and background with same final state.

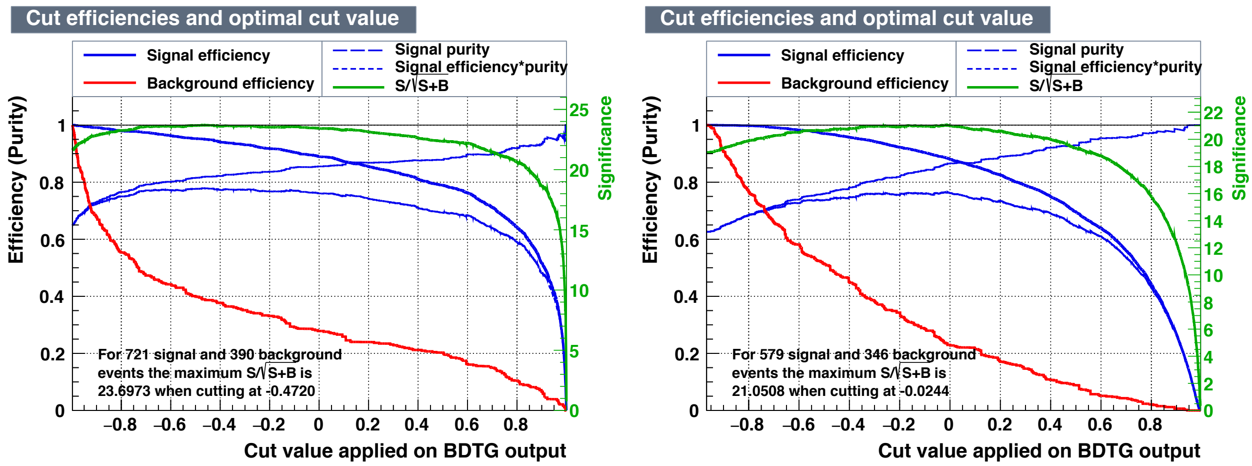

Figure 2.7: Results of secondary BDTG event selection for hadronic (left) and semi-leptonic signal (right) trained for correctly paired signal

The output of primary BDT and secondary BDTG event selection are combined with the previous ROOT-file from the Marlin analysis. After that the best cut values are selected. The corresponding cuts are shown in table 2.8 for hadronic and in table 2.9 for semi-leptonic mode. When optimizing for signal significance the optimal cut values can be taken from figure 2.6.

	had. signal	semi-l. signal	BG	Signif.	Effi.	Purity
Expected	4771	4597	$3.04 \cdot 10^8$	0.27	1.00	0.00
No IsoLep	4684	1642	$2.11 \cdot 10^8$	0.32	0.98	0.00
4 highest b-tag > 2.7	3606	1326	57006	14.65	0.76	0.06
E_{vis} 1200	3605	1326	56872	14.66	0.76	0.06
E_{vis} 760	3543	948	25223	20.89	0.74	0.12
$\chi_{\mathrm{H}^\pm} < 6$	3543	947	23814	21.42	0.74	0.13
$y_{45} > 0.002$	3487	896	8214	32.23	0.73	0.30
$\chi_{\mathrm{t}} < 9$	3487	896	8213	32.24	0.73	0.30
$y_{67} > 5 \cdot 10^5$	3477	875	7438	33.28	0.73	0.32
principleThrust < 0.81	3213	759	2361	43.03	0.67	0.58
minorThrust > 0.11	3209	756	2183	43.70	0.67	0.60
$\lvert$cosThrustAxis$\rvert < 0.91$	3127	736	1885	44.17	0.66	0.62
$m_{\mathrm{miss}} > 140$	3107	722	1803	44.34	0.65	0.63
$m_{\mathrm{miss,t}}$ 125	3094	587	1727	44.56	0.65	0.64
$m_{\mathrm{miss,z}}$ 210	3090	586	1708	44.61	0.65	0.64

Table 2.4: Cut table for hadronic signal / hadronic signal significance

	semi-l. signal	had. signal	BG	Signif.	Effi.	Purity
Expected	4597	4771	$3.04 \cdot 10^8$	0.26	1.00	0.00
IsoLep	2955	87	$9.27 \cdot 10^7$	0.31	0.64	0.00
4 highest b-tag > 2.5	2386	53	20712	15.70	0.52	0.10
$E_{\mathrm{vis}} < 330$	2298	53	12680	18.77	0.50	0.15
$E_{\mathrm{vis}} > -100$	2297	52	11993	19.22	0.50	0.16
$\chi_{\mathrm{H}^\pm} < 3$	2286	50	9891	20.72	0.50	0.19
$y_{45} > 0.001$	2237	50	3325	30.00	0.49	0.40
$\chi_{\mathrm{t}} < 41$	2237	50	3325	30.00	0.49	0.40
principleThrust < 0.815	2041	47	1190	35.90	0.44	0.63
minorThrust > 0.11	2033	47	1145	36.06	0.44	0.64
$\lvert$cosThrustAxis$\rvert < 0.94$	2001	46	1035	36.32	0.44	0.66
$m_{\mathrm{miss}} > -160$	1985	46	981	36.45	0.43	0.67
$m_{\mathrm{miss,t}} < 290$	1985	46	978	36.46	0.43	0.67
$m_{\mathrm{miss,z}} < 240$	1982	46	965	36.51	0.43	0.67

Table 2.5: Cut table for semi-leptonic signal optimized on signal significance

	cor. h. Sig.	other Sig.	BG	Signif.	Effi.	Purity
Expected	1166	8202	$3.04 \cdot 10^8$	0.07	1.00	0.00
No IsoLep	1165	5161	$2.11 \cdot 10^8$	0.08	1.00	0.00
4 highest b-tag $>$ 2.8	998	3705	42661	4.77	0.86	0.02
$E_{\text{vis}} < 1100$	997	3698	42207	4.80	0.86	0.02
$E_{\text{vis}} > 820$	951	2815	17225	7.05	0.82	0.05
$\chi_{\text{H}^\pm} < 0.4$	898	2346	10417	8.44	0.77	0.08
$y_{45} > 0.003$	862	2178	3041	13.83	0.74	0.22
$\chi_{\text{t}} < 0.4$	813	1513	1955	15.49	0.70	0.30
$y_{67} > 5 \cdot 10^{-5}$	810	1502	1815	15.86	0.69	0.31
principleThrust $<$ 0.8	749	1304	521	21.18	0.64	0.60
$\lvert$cosThrustAxis$\rvert$ $<$ 0.91	733	1271	458	21.43	0.63	0.63
$m_{\text{miss}} > -100$	726	1236	421	21.63	0.62	0.64
$m_{\text{miss,t}} < 95$	723	1155	394	21.84	0.62	0.66
$m_{\text{miss,z}} < 170$	721	1154	390	21.86	0.62	0.66

Table 2.6: Cut table for hadronic signal optimized on correctly paired signal significance

	cor. sl. Sig.	other Sig.	BG	Signif.	Effi.	Purity
Expected	1053	8315	$3.04 \cdot 10^8$	0.06	1.00	0.00
IsoLep	943	2099	$9.27 \cdot 10^7$	0.10	0.90	0.00
4 highest b-tag $>$ 2.85	741	1294	7334	8.25	0.70	0.09
$E_{\text{vis}} < 300$	703	1198	4266	9.97	0.67	0.14
$E_{\text{vis}} > -20$	701	1193	3798	10.45	0.67	0.16
$\chi_{\text{H}^\pm} < 1$	689	1107	2757	11.74	0.65	0.20
$y_{45} > 0.001$	676	1086	1300	15.21	0.64	0.34
$\chi_{\text{t}} < 1$	649	795	1003	15.97	0.62	0.39
principleThrust $<$ 0.815	594	717	395	18.90	0.56	0.60
minorThrust $>$ 0.11	591	715	388	18.90	0.56	0.60
$\lvert$cosThrustAxis$\rvert$ $<$ 0.935	582	703	358	18.98	0.55	0.62
$m_{\text{miss}} > -180$	580	701	350	19.02	0.55	0.62
$m_{\text{miss,t}} < 310$	580	701	350	19.02	0.55	0.62
$m_{\text{miss,z}} < 210$	579	699	346	19.04	0.55	0.63

Table 2.7: Cut table for semi-leptonic signal optimized on correctly paired signal significance

	cor. h. Sig.	other Sig.	BG	Signif.	Effi.	Purity
Expected	1166	8202	$3.04 \cdot 10^8$	0.07	1.00	0.00
No IsoLep	1165	5161	$2.11 \cdot 10^8$	0.08	1.00	0.00
pre BDT $>$ 0.13	1010	2914	1531	20.04	0.87	0.40
sec BDTG $>$ -0.025	865	936	190	26.63	0.74	0.82

Table 2.8: Cut table for hadronic signal optimized on correctly paired signal significance with TMVA outputs

	cor. sl. Sig.	other Sig.	BG	Signif.	Effi.	Purity
Expected	1053	8315	$3.04 \cdot 10^8$	0.06	1.00	0.00
IsoLep	943	2099	$9.27 \cdot 10^7$	0.10	0.90	0.00
pre BDT > 0.105	823	1606	1331	17.73	0.78	0.38
sec BDTG > 0.025	671	483	181	22.99	0.64	0.79

Table 2.9: Cut table for semi-leptonic signal optimized on correctly paired signal significance with TMVA outputs

2.6 Mass measurement

In order to develop a procedure for a possible mass measurement of the charged Higgs bosons data samples with variated mass were generated. To know in which margin the samples should be generated the mass distribution of correctly paired signal was fitted with a Breight-Wigner distribution and the failed pairing together with the background was fitted with a Gaussian distribution. In an added fit the two shapes were fitted together. Here the uncertainty on the mean of the Breight-Wigner distribution given by the used RooFit package [33] was about 1 GeV. With this very preliminary result it was decided to produce five data set in 2 GeV steps. In the further study the correlation between fitted means and generated mass was difficult to evaluate. That is the reason why two additional data sets at ± 10 GeV where generated (compare first rows of table A.1). The distribution of the invariant mass of the two reconstructed Higgs bosons of these seven samples is referred to as templates. For $m_{\mathrm{H}^\pm} = 350$ GeV twice as many events were generated to provide a statistically independent test data set in addition to the template at this mass. The test data set contains just the number of events from each sample, so that some of the events in the large sample stay unused.

In the following the charged Higgs mass distribution will be the invariant mass of the the first four jets j_1 to j_4 and last four jets j_5 to j_8 from the best selection in chapter 2.3.3. Both invariant masses are added to the same histogram. As well as hadronic and semi-leptonic signal are both added to this histogram.

maybe add a plot of different modes different invariant masses to show that there is now significant difference

The SM background which can not be reduced from the event selection (chapter 2.5) is flattened by a bifurcated Gaussian distribution (see figure 2.8). A bifurcated Gaussian is defined as

$$f(x) = \frac{1}{(\sigma_L + \sigma_R)\sqrt{\frac{\pi}{2}}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \text{with} \quad \sigma = \begin{cases} \sigma_L & x - \mu < 0 \\ \sigma_R & \text{otherwise} \end{cases}$$

where μ is the maximum[3] of the distribution and σ_L and σ_R are the widths of the left and right Gaussian respectively.

This flattening is necessary because the different data samples have different statistics and most samples have no additional statistic to provide an independent test data set. For example the data samples of Z boson to two quarks have very low statistic. Therefore they have to be weighted with 69.4. Only two events are selected by the static cuts which results in four large error bars in figure 2.8. Nevertheless it is assumed that the number of events (even with large uncertainty) is an appropriate approximation. However, for the purpose of getting a realistic distribution in the reconstructed Higgs mass a bifurcated Gaussian was chosen because it seams to fit well even if there is a slight change in the event selection.

[3] μ is sometimes referred to as mean but this is only true if $\sigma_L = \sigma_R$

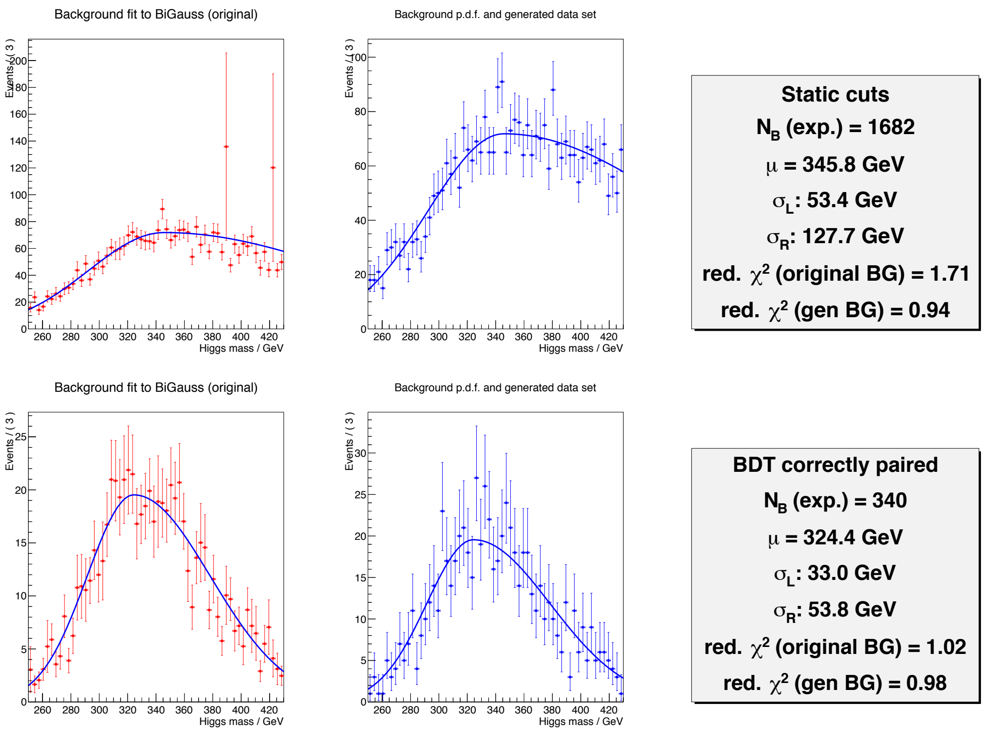

Figure 2.8: Charged Higgs mass distribution of SM background; upper plots show the background with static cut selection optimized on signal significance (chapter 2.5.1) and lower shows TMVA based selection optimized on correctly paired signal significance (chapter 2.5.2). Left hand plots reveal the original distribution and the fitted bifurcated Gaussian; in the middle the generated distribution which is used for the further mass determination is shown and on the right information to the fit is written.

2.6.1 Template method

For the template method a test set is compared to templates (see figure 2.9). To compare the distributions the minimum chi squared method is used. Here χ^2_{temp} is used as an observable for the difference of the distributions with the following definition

$$\chi^2_{\text{temp}} = \sum_{i=0}^{N} \frac{(T_i - S_i)^2}{S_i}$$

for histograms with N bins where T_i accounts for the expected number of events in bin i originate from the template; S_i is the corresponding number of events in bin i of the test set. The templates as well as the test data set includes the SM background as discussed before. All templates contain the same generated data set but the background set of the test sample is generated independently. The number of background events for the templates is the expected number from chapter 2.5. However, the corresponding number for the test set is a random number from Poisson distribution with mean of the expected number.

From the seven templates χ^2_{temp} values are plotted on the corresponding generated Higgs mass (see figure 2.10). In the case where the templates have the same statistics as the test data set

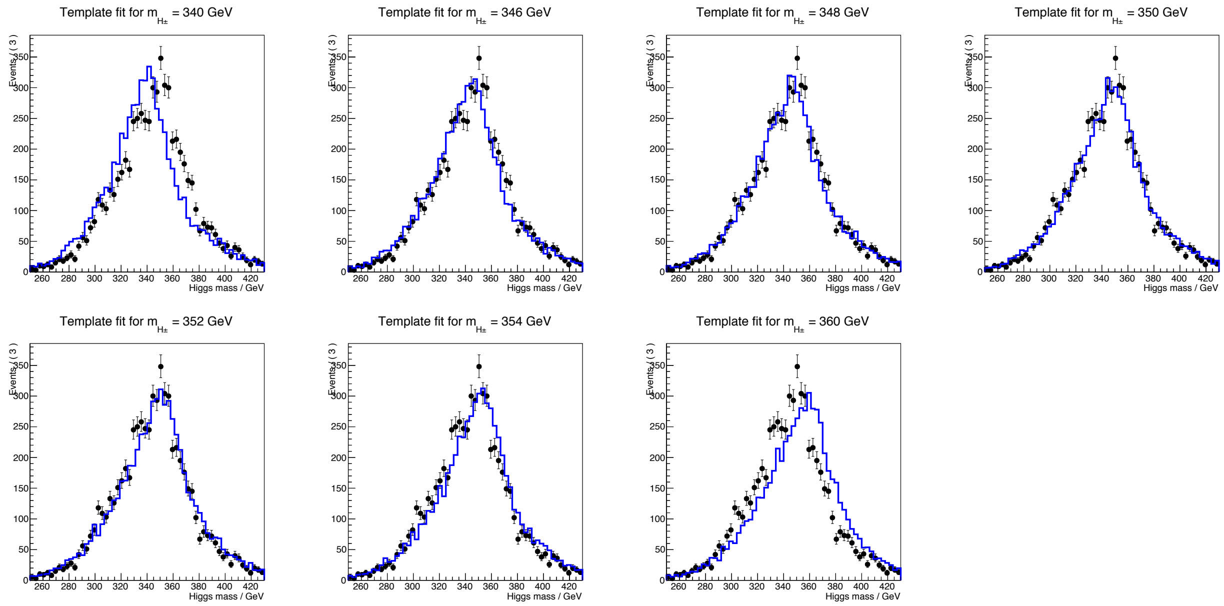

Figure 2.9: Template fit of variated $m_{\mathrm{H}^{\pm}}$; in blue histograms of templates and black data points from a test data set with $m_{\mathrm{H}^{\pm}} = 350\,\mathrm{GeV}$; the event selection is done by BDT optimized on correct pairing

the uncertainty on χ^2_{temp} would be $\Delta\chi^2_{\mathrm{temp}} = \sqrt{2N}$, where N is the number of bins. Here the expected number of events is different for all the modes but the statistic of the templates is same but at least about double of the test data set but the real shape of χ^2_{temp} is unknown, for that reason $\Delta\chi^2_{\mathrm{temp}} = \sqrt{2N}$ is used as rough estimate which is only a visual orientation and has no influence on the final result.

The χ^2_{temp} points in figure 2.10 are fitted with a parabola where the minimum is the estimate of the real experiment for the final result of the mass measurement. Therefore

$$m_{\mathrm{H}^{\pm}} = \chi^{-2}(\chi^2_{\mathrm{min}}) \text{ with } \left[\frac{d\chi^2_{\mathrm{temp}}(x)}{dx} = 0\right]_{x=\chi^2_{\mathrm{min}}}$$

where $\chi^2(x)$ is the fitted function and $\chi^{-2}(x)$ the inverse. The statistical uncertainty is given as

$$\Delta m_{\mathrm{H}^{\pm}} = \chi^{-2}(\chi^2_{\mathrm{min}} + 1 > m_{\mathrm{H}^{\pm}}) - \chi^{-2}(\chi^2_{\mathrm{min}} + 1 < m_{\mathrm{H}^{\pm}})$$

2.6.2 Shape method

In this method the aim is to identify the signal shape and fit its position to a test data set. Then a linear regression is applied to the position of the signal shape and the generated mass of the templates. From this information and the signal shape position of the test data set the underlying mass is reconstructed.

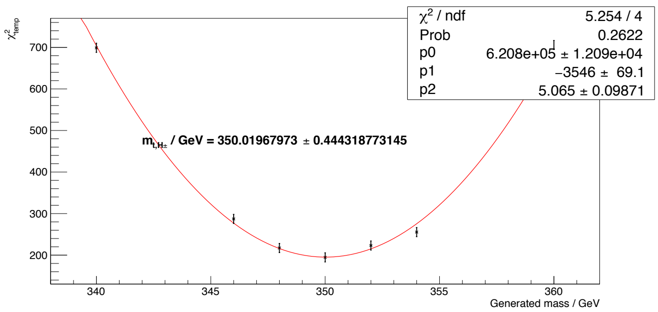

Figure 2.10: χ^2_{temp} obtained from the comparison of the templates and a test data set fitted by a parabola; the event selection is done by BDT optimized on correct pairing

make large

The signal shape is approximated with two bifurcated Gaussian distributions; a narrow one for correctly paired signal and a wide one for wrong paired signal. Correctly paired signal is selected as defined in chapter 2.3.3 and fitted with a bifurcated Gaussian and a normal Gaussian (left column of figure 2.11). Signal where the clustering has failed (definition in chapter 2.3.1) is fitted to the bifurcated Gaussian for wrong pairing which is displayed in the second column of figure 2.11. These tow preliminary fits fulfill only the purpose of gaining reliable starting values for fitting the total signal shape. There both bifurcated Gaussian distributions are fitted to the signal shape. This is shown in the right half of figure 2.11.
For the next step a generated background data set is added to templates and test data set in the same manner as explained before for the template method. Then all seven templates are fitted with three bifurcated Gaussian distributions for background, correctly and wrong paired signal. All parameters are fixed to the expected value except the maximum of correctly and wrong paired distributions which will be called μ_c and μ_w respectively in the following. The linear regression of the results for μ_c and μ_w and the generated mass are shown in figure 2.12.

The test data set is fitted in the same manner. This fit is shown for the four different event selections in figure 2.13. From this the estimate for shape method of the real experiment for

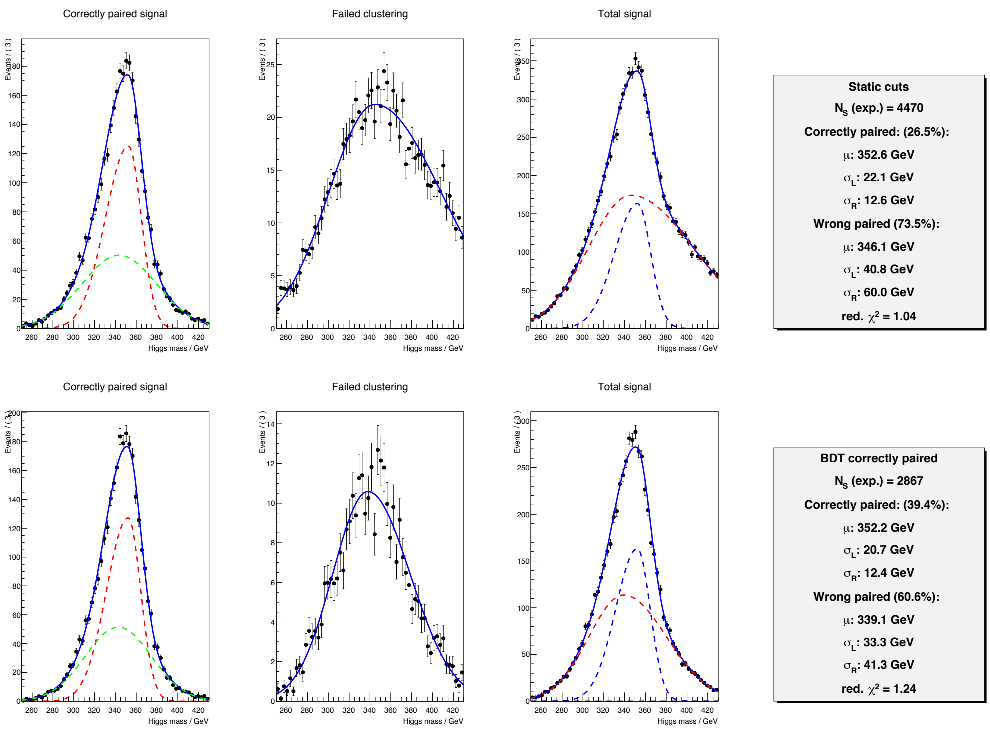

Figure 2.11: Charged Higgs mass distribution for signal and shape fitting; upper plots show the background with static cut selection optimized on signal significance (chapter 2.5.1) and lower shows TMVA based selection optimized on correctly paired signal significance (chapter 2.5.2). The first two columns show the preliminary fits to obtain start values for the final fit. The third column reveals the final fit where correctly paired (dashed blue) and wrong paired signal (dashed red) is fitted with bifurcated Gaussian distributions. In the left column information to the fit is shown.

the final result is given by

$$m_{\mathrm{H}^\pm} = b\mu + a$$

where b is the slope of the linear regression and a is the y-axis intercept. Therefore the uncertainty is given by

$$\begin{aligned}\Delta m_{\mathrm{H}^\pm} &= \sqrt{\left(\frac{\Delta\mu}{b}\right)^2 + \left(\frac{\Delta a}{b}\right)^2 + (\mu\Delta b)^2} \\ &= \sqrt{\Delta_{\text{fit}}^2 + \Delta_{\text{const}}^2 + \Delta_{\text{linear}}^2}\end{aligned}$$

In figure 2.12 the results from the test data set is shown in blue color.
The two results from wrong and correctly paired signal can be combined to one by weighted average.

$$m_{a,\mathrm{H}^\pm} = \frac{w_c m_{c,\mathrm{H}^\pm} + w_w m_{w,\mathrm{H}^\pm}}{w_c + w_w} \quad \text{with} \quad w_i = \frac{1}{(\Delta m_{i,\mathrm{H}^\pm})^2}$$

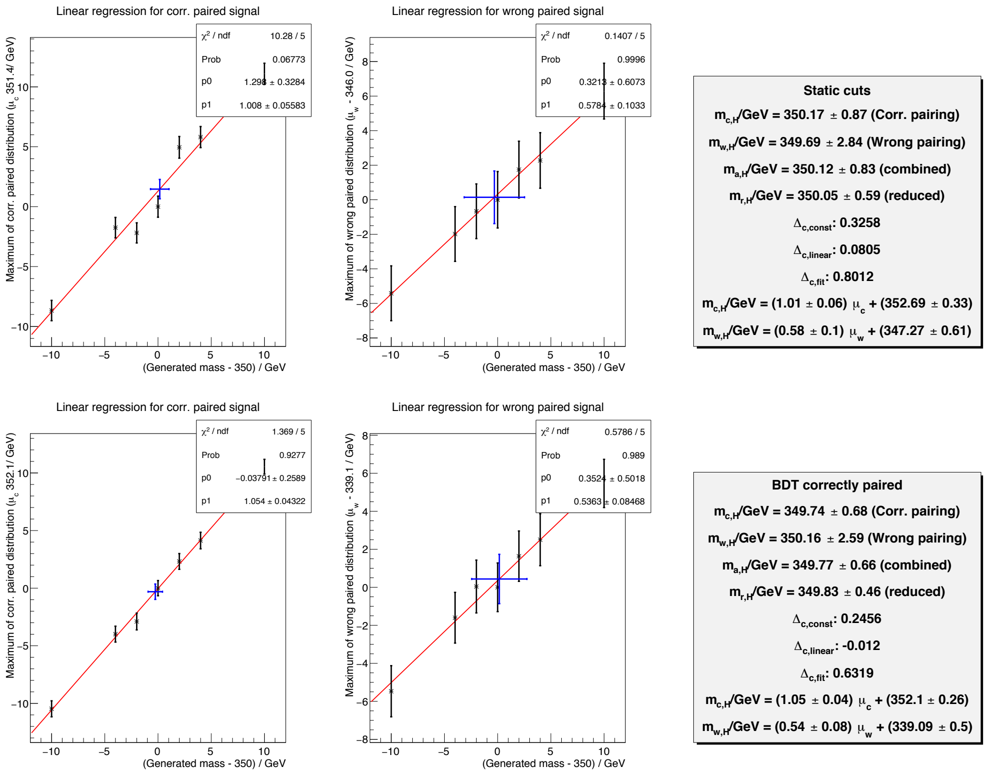

Figure 2.12: Linear regression of the generated mass to the maximum μ of the correctly paired (wrong paired) bifurcated Gaussian in the left (middle) column; upper plots show the background with static cut selection optimized on signal significance (chapter 2.5.1) and lower shows TMVA based selection optimized on correctly paired signal significance (chapter 2.5.2). In the left column information to the fit and the results for $m_{\mathrm{H}^\pm}$ are shown.

Hence, the uncertainty is

$$\Delta m_{a,\mathrm{H}^\pm} = \frac{1}{\sqrt{w_c + w_w}}$$

However, since $\Delta m_{c,\mathrm{H}^\pm} << \Delta m_{w,\mathrm{H}^\pm}$ the advantage of the weighted average over the value from estimated form the position of the correctly paired distribution is minimal (see 2.14).

2.6.3 Reduced shape method

This method is a variation of the shape method. Rather then combining two results as for $m_{a,\mathrm{H}^\pm}$, the fit can be reduced to only one variable, since the relations of $m_{\mathrm{H}^\pm}(\mu_c)$ and $m_{\mathrm{H}^\pm}(\mu_w)$ is known. RooFit provides a `RooFormulaVar` object to enable to connect a fitting parameter as μ with an formula to another parameter. Connecting tow of these objects with the formula gained from the linear regressions reduces the fit parameter to one which is directly the result of the estimate of the charged Higgs mass. The result is show as $m_{r,\mathrm{H}^\pm}$ for the examples in figure 2.12.

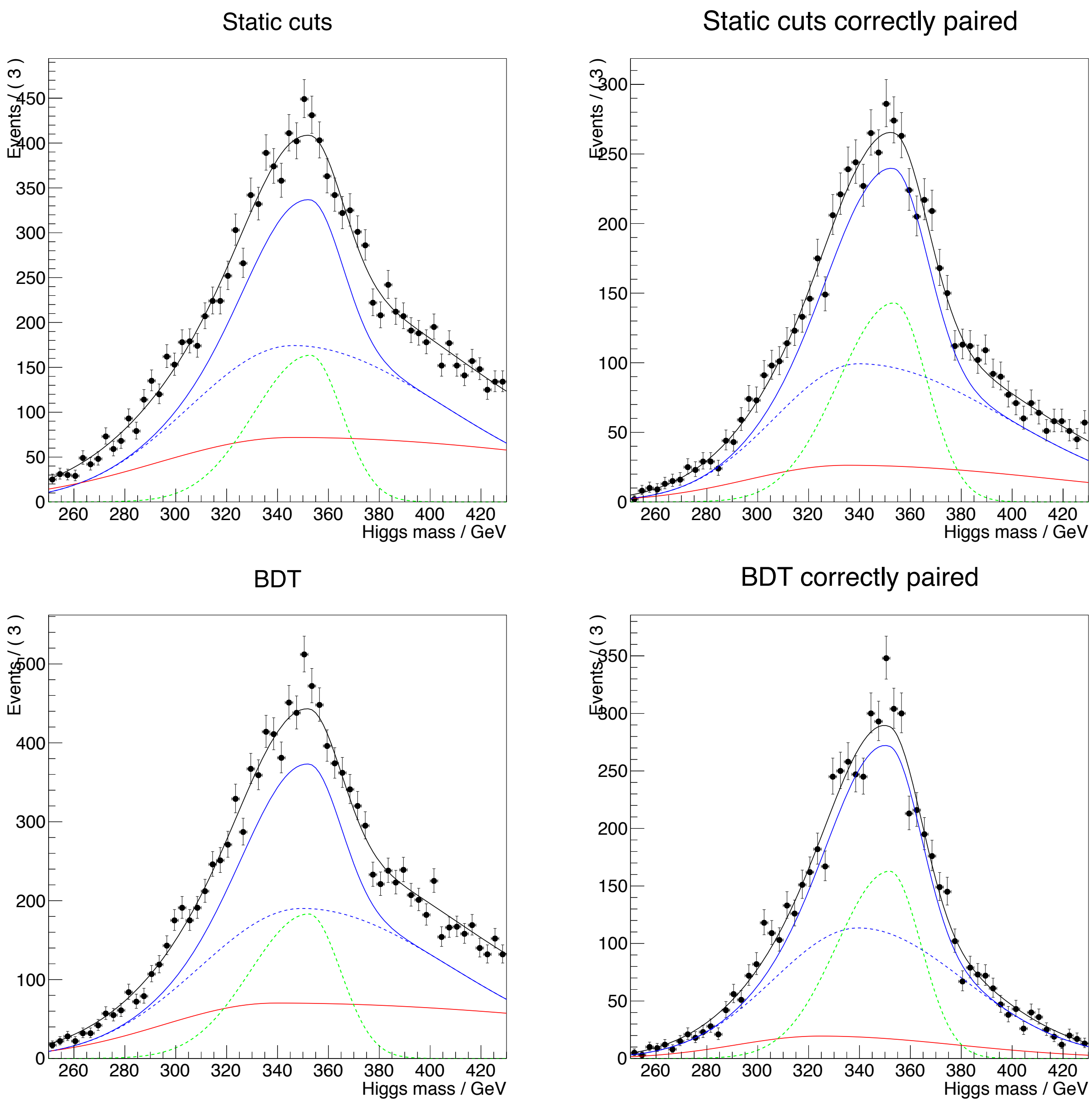

Figure 2.13: Fit of test data set by the shape method for different event selections as denote each figure; the data points is the test data set; the black function the total function, solid blue - total signal, dashed blue - wrong pairing, green - correct pairing and red is the background

In order to test the three methods described and the different event selections described in chapter 2.5 a Monte Carlo toy study was performed. Therefore a second template independent from the former template was taken from the second half of the data samples with $m_{\mathrm{H}^{\pm}} = 350$ GeV. On the bases of this template 10,000 test data sets are generated and the same procedures for the mass measurements are ran through. The number of signal events in the toy test data set can be varied and related to a cross section while the number of background events is kept constant.

The uncertainty and deviation from expected value of all mass measurement methods with the TMVA based selection optimized on correctly paired signal significance (chapter 2.5.2) is shown in figure 2.14. In figure 2.15 (figure 2.17 / figure 2.16) uncertainty and deviation of reduced shape method (template method / shape method) is shown with different event selections.

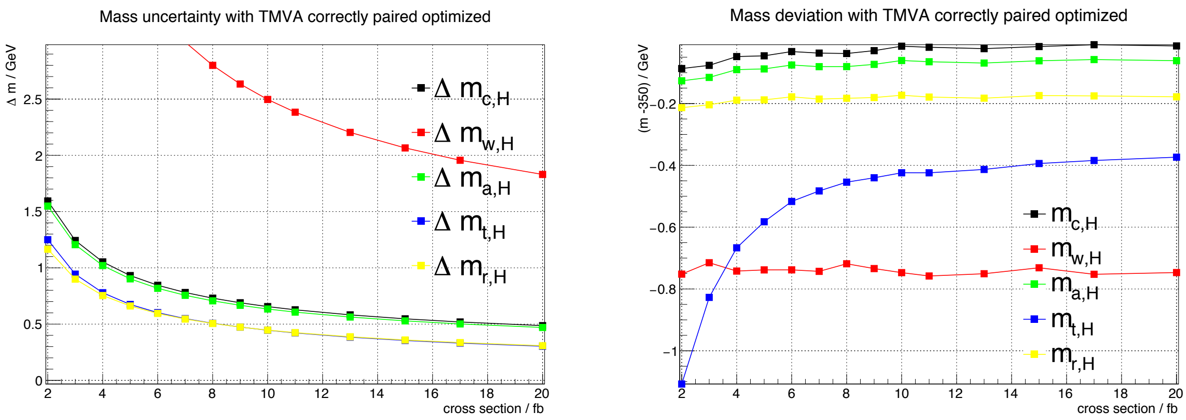

Figure 2.14: Mass uncertainty (right) and mass deviation of the expected value (left) for different mass measurement methods as noted in the legend; using the BDT based event selection optimized for correctly paired signal

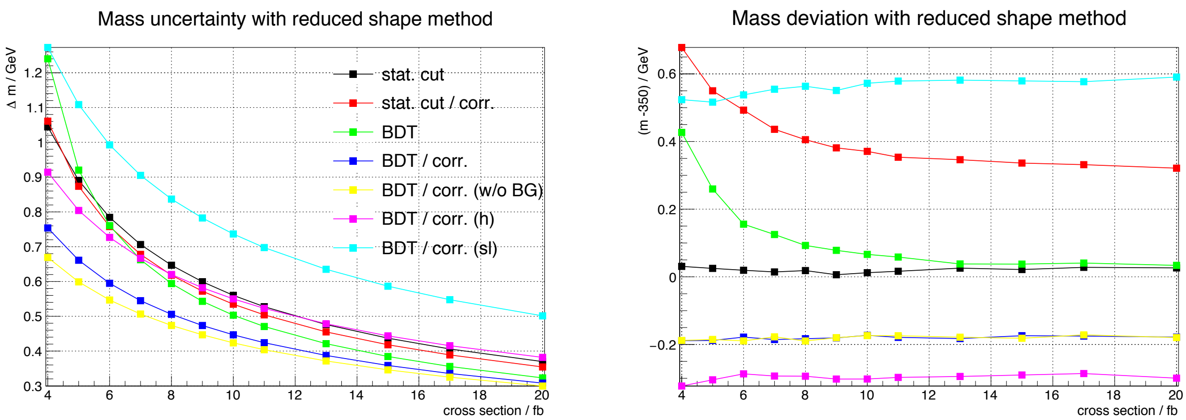

Figure 2.15: Mass uncertainty (right) and mass deviation of the expected value (left) from the reduced shape method for different event selections as noted in the legend

With the result from figure 2.14 it is shown that the uncertainty from the template method and reduced shape method is lowest. But the deviation of the template method is depended on the cross section. This is unexpected but the reason could be that the true function of χ^2_{temp} is not parabolic. If the true shape is unsymmetrical and has a steeper slope on one side, it could let the average result for the toy study deviate to the side of genital slope. A different explanation could be that the fraction of signal and background is different between test data set and templates. However, it has been made sure that apart from the Poisson fluctuation in the test data set there is no difference in the composition.

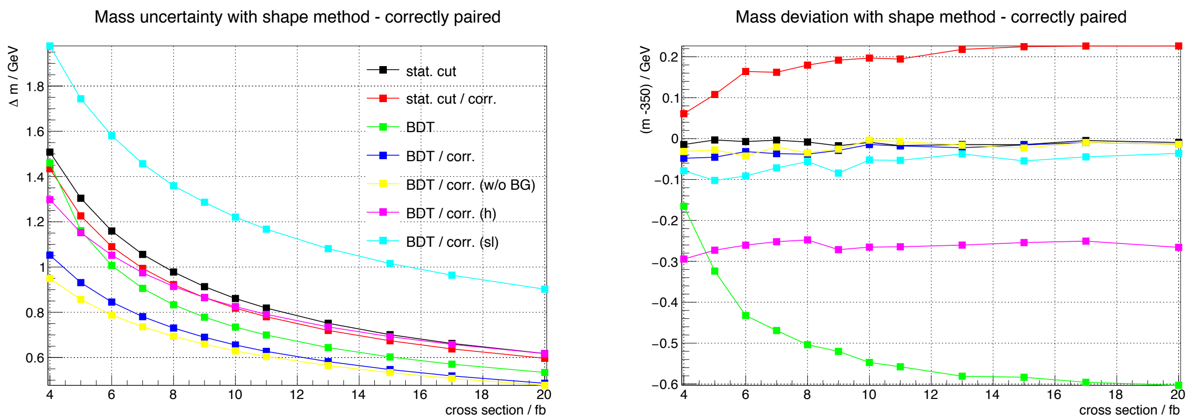

Figure 2.16: Mass uncertainty (right) and mass deviation of the expected value (left) from the shape method estimated form the position of the correctly paired distribution for different event selections as noted in the legend

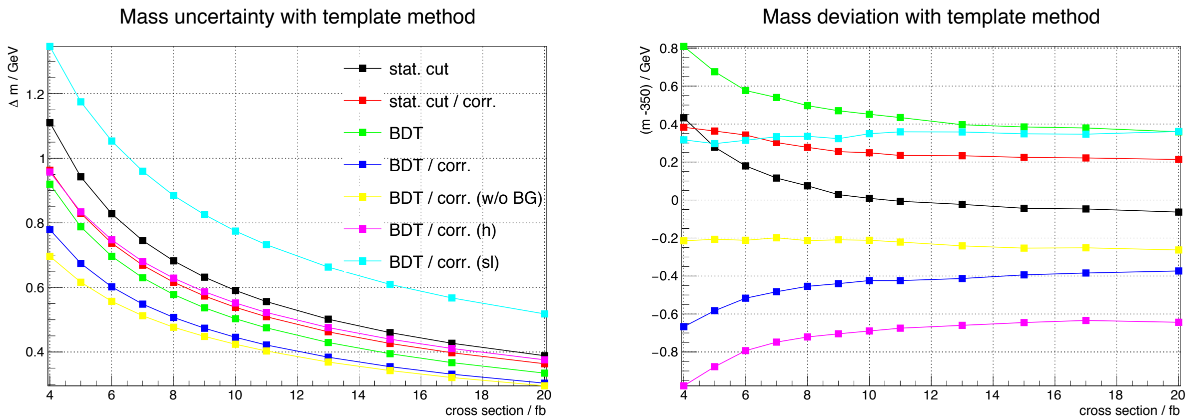

Figure 2.17: Mass uncertainty (right) and mass deviation of the expected value (left) from the template method for different event selections as noted in the legend

In the figures 2.15, 2.17 and 2.16 the uncertainty is lowest for event selection optimized for correctly paired signal based on BDT. Only the same selection fitted without background is better. In the case of static cuts, it is an advantage as well to optimize the cuts for correctly paired signal. For some mass measurement methods and event selections the deviation to the expected is depended on cross section. Eye-catching is that BDT selection optimized for signal significance is depended for all shown methods. The reason maybe that the wrong paired signal shape is very similar to the background shape (see 2.13. This can make the true function of χ^2_{temp} unsymmetrical. The other shape fitting methods are as well essentially χ^2 minimizations and the same effect can occur.

However, the deviation is in general smaller then the statistical uncertainty, so the results are reliable but for those methods which are dependent on cross section the deviation needs to be corrected or/and taken into account as systematic uncertainty.

3 Discussion

3.1 Result

For the neutrino reconstruction the Missing Momentum Method showed the best performance. However, the Missing Transversal Momentum Method has shown a good potential and might be useful for studies where very high precision is necessary.
The event selection has been conducted with static cuts as well as with the multi variable analysis toolkit from ROOT using boosted decision trees (BDT). The selection was optimized on signal significance or on correctly paired signal significance. An overview of significances, efficiencies, purities and mass precision of the selection can be found in table 3.1.

Cut type	Optim. type	Mode	Significance	Efficiency	Purity	mass precision
Static cuts		hadronic	44.61	65 %	64 %	0.60 GeV
Static cuts		semi-lep.	36.51	43 %	67 %	
Static cuts	corr. paired	hadronic	21.86	62 %	66 %	0.57 GeV
Static cuts	corr. paired	semi-lep.	19.04	55 %	63 %	
BDT		hadronic	49.14	73 %	67 %	0.54 GeV
BDT		semi-lep.	38.64	46 %	71 %	
BDT	corr. paired	hadronic	26.63	74 %	82 %	0.47 GeV
BDT	corr. paired	semi-lep.	22.99	64 %	79 %	

Table 3.1: Summary of significances, efficiencies, purities and mass precision with reduced shape method for 9 fb for the event selection

For the measurement of the charged Higgs mass three methods have been conducted. The best method was shown to be the reduced shape method with the BDT based event selection optimized on correctly paired significance. This configuration was found to have a statistical uncertainty of 0.5 GeV for the in MSSM expected 9 fb which relates to a relative uncertainty 0.14 %.

3.2 Outlook

With the parameter set used in this analysis the minimal supersymmetric standard model (MSSM) is excluded. However, for event generation it is not necessary to assume MSSM, as long as the mass, width and charge of the charged Higgs boson is fixed, all underlying probability density function for Monte Carlo simulation are fixed in this process. This makes this study transferable to a wide range of two Higgs doublet models. As well as the mass has been fixed in this study but if there is a discovery of a charged Higgs like particle the techniques and even the developed analysis program can be adjusted to the discovery.

There are open questions in this analysis. It has to be investigated why the deviation of the mean of result from toy Monte Carlo study is dependent on the cross section. Therefore it necessary to produce more mass samples in a smaller margin, to study the underlying χ^2 minimization.

In this analysis it was assumed that the simulation is conform with the real events. This assumption is needed to be able to compare the test data set, which will be real data in the ILC experiment, to the templates, which will be simulated. However, the simulation at this point will not be conform with what will be seen at the ILC. Nevertheless, when this analysis is conducted, the ILC project will have been running for several years and simulation will evolve with the project. Furthermore, the deviation of nature and simulation will be known from other measurements. For example at this point it is unknown, whether the real events and simulation behaves under the used beam background removal with kt-algorithm described in chapter 2.3.1. However, at the point this analysis will be conducted with real data similar background removal will have been used for other analyses such as top pair production or top Yukawa studies. With this experience, the influence can be corrected or/and the resulting systematic uncertainty will be better understood. The same can be said about the mass measurement, including the background estimation in it, and the neutrino reconstruction.

In addition it would be interesting to see the influence of using the Missing Transversal Momentum Method instead of the Missing Momentum Method for neutrino reconstruction on the final result. As well as the influence of the discussed jet pairing optimization. Nevertheless the influence is expected to be minor.

In this study only Standard Model background has been considered. In a similar study for the Compact Linear Collider at 3 TeV [11] it has been found that the SUSY background from heavy neutral Higgs bosons AH $\rightarrow$ bbbb is peaking in the same region. Since their mass may (depending on the model) change correlated to the charge Higgs mass and that way have major effect on the final result. In case of MSSM the mass of the charged Higgs boson is very similar to the mass of A and H. This would let the cross section of AH $\rightarrow$ tttt peak at the used parameter set and would become the major background.

4 Bibliography

[1] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys.Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[2] G. Abbiendi et al. "Search for Charged Higgs bosons: Combined Results Using LEP Data." In: *Eur. Phys. J.* C73 (2013), p. 2463. DOI: 10.1140/epjc/s10052-013-2463-1. arXiv: 1301.6065 [hep-ex].

[3] Halina Abramowicz et al. "The International Linear Collider Technical Design Report - Volume 4: Detectors." In: (2013). Ed. by Ties Behnke et al. arXiv: 1306.6329 [physics.ins-det].

[4] Chris Adolphsen et al. "The International Linear Collider Technical Design Report - Volume 3.I: Accelerator & in the Technical Design Phase." In: (2013). arXiv: 1306.6353 [physics.acc-ph].

[5] Chris Adolphsen et al. "The International Linear Collider Technical Design Report - Volume 3.II: Accelerator Baseline Design." In: (2013). arXiv: 1306.6328 [physics.acc-ph].

[6] I. Antcheva et al. "ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization." In: *Comput.Phys.Commun.* 182 (2011), pp. 1384–1385. DOI: 10.1016/j.cpc.2011.02.008.

[7] A. Arbey et al. "Status of the Charged Higgs Boson in Two Higgs Doublet Models." In: (2017). arXiv: 1706.07414 [hep-ph].

[8] Howard Baer et al. "The International Linear Collider Technical Design Report - Volume 2: Physics." In: (2013). arXiv: 1306.6352 [hep-ph].

[9] D. Bailey et al. "The LCFIVertex package: vertexing, flavour tagging and vertex charge reconstruction with an ILC vertex detector." In: *Nucl. Instrum. Meth.* A610 (2009), pp. 573–589. DOI: 10.1016/j.nima.2009.08.059. arXiv: 0908.3019 [physics.ins-det].

[10] Timothy Barklow. personal communication. Sept. 19, 2017.

[11] Marco Battaglia. "Charged Higgs Boson Physics at Future Linear Colliders." In: *PoS* CHARGED2010 (2010), p. 019. arXiv: 1102.1892 [hep-ex].

[12] Ties Behnke et al. "The International Linear Collider Technical Design Report - Volume 1: Executive Summary." In: (2013). arXiv: 1306.6327 [physics.acc-ph].

[13] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual." In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[14] S. Catani et al. "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions." In: *Nucl. Phys.* B406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

[15] S. Catani et al. "New clustering algorithm for multi - jet cross-sections in e+ e- annihilation." In: *Phys. Lett.* B269 (1991), pp. 432–438. DOI: 10.1016/0370-2693(91)90196-W.

[16] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Phys.Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[17] Ekaterina Christova et al. "CP violation in charged Higgs decays in the MSSM with complex parameters." In: *Nucl. Phys.* B639 (2002), pp. 263–280. DOI: 10.1016/S0550-3213(02)00542-4. arXiv: hep-ph/0205227 [hep-ph].

[18] E. Christova et al. "CP violation in charged Higgs decays in the MSSM." In: *JHEP* 02 (2007), p. 075. DOI: 10.1088/1126-6708/2007/02/075. arXiv: hep-ph/0612088 [hep-ph].

[19] Gregory Ciezarek et al. "A Challenge to Lepton Universality in B Meson Decays." In: *Nature* 546 (2017), pp. 227–233. DOI: 10.1038/nature22346. arXiv: 1703.01766 [hep-ex].

[20] Abdelhak Djouadi. "The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model." In: *Phys. Rept.* 459 (2008), pp. 1–241. DOI: 10.1016/j.physrep.2007.10.005. arXiv: hep-ph/0503173 [hep-ph].

[21] Yuri L. Dokshitzer et al. "Better jet clustering algorithms." In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[22] S. Heinemeyer and C. Schappacher. "Charged Higgs Boson production at e^+e^- colliders in the complex MSSM: a full one-loop analysis." In: *Eur. Phys. J.* C76.10 (2016), p. 535. DOI: 10.1140/epjc/s10052-016-4383-3. arXiv: 1606.06981 [hep-ph].

[23] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis." In: *PoS* ACAT (2007), p. 040. arXiv: physics/0703039 [PHYSICS].

[24] Go Iwai et al. "KEK Central Computer System (KEKCC)." In: *PoS* ISGC2016 (2016), p. 037.

[25] Claude Duerig Junping Tian. *Isolated lepton finder*. July 2015. URL: https://agenda.linearcollider.org/event/6787/contributions/33415/ (visited on 09/16/2017).

[26] Lucie Linssen et al. "Physics and Detectors at CLIC: CLIC Conceptual Design Report." In: (2012). DOI: 10.5170/CERN-2012-003. arXiv: 1202.5940 [physics.ins-det].

[27] Akiya Miyamoto. "Software tools for JLC studies." In: *AIP Conference Proceedings*. Vol. 578. 1. AIP. 2001, pp. 646–649.

[28] H. Murayama, I. Watanabe, and Kaoru Hagiwara. "HELAS: HELicity amplitude subroutines for Feynman diagram evaluations." In: (1992).

[29] Tony Price et al. "Full simulation study of the top Yukawa coupling at the ILC at $\sqrt{s} = 1$ TeV." In: *Eur. Phys. J.* C75.7 (2015), p. 309. DOI: 10.1140/epjc/s10052-015-3532-4. arXiv: 1409.7157 [hep-ex].

[30] Daniel Schulte. "Study of Electromagnetic and Hadronic Background in the Interaction Region of the TESLA Collider." PhD thesis. DESY, 1997. URL: http://inspirehep.net/record/888433/files/shulte.pdf.

[31] Taikan Suehara and Tomohiko Tanabe. "LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies." In: *Nucl. Instrum. Meth.* A808 (2016), pp. 109–116. DOI: 10.1016/j.nima.2015.11.054. arXiv: 1506.08371 [physics.ins-det].

[32] M.A. Thomson. "Particle Flow Calorimetry and the PandoraPFA Algorithm." In: *Nucl.Instrum.Me* A611 (2009), pp. 25–40. DOI: 10.1016/j.nima.2009.09.009. arXiv: 0907.3577 [physics.ins-det].

[33] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling." In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[34] O. Wendt, F. Gaede, and T. Kramer. "Event Reconstruction with MarlinReco at the ILC." In: *Pramana* 69 (2007), pp. 1109–1114. DOI: 10.1007/s12043-007-0237-8. arXiv: physics/0702171 [PHYSICS].

A Appendix

A.1 Durham algorithm

The Durham [15] algorithm works in the following manner.[21]

1. Calculate the distance between to all tracks $v_{ij} = 2(1 - \cos\theta_{ij})$
2. Find smallest v_{ij}
3. Calculate $y_{ij} = \min(E_i, E_j)v_{ij}$
   a) If $y_{ij} < y_{\text{cut}}$ merche the two tracks - udate table and start over with step 2
   b) If $y_{ij} > y_{\text{cut}}$ return to step 2 and look for next larger v_{ij}
4. If there are no tracks left to merge and there are more tracks than requested jets raise y_{cut} and go to step 2

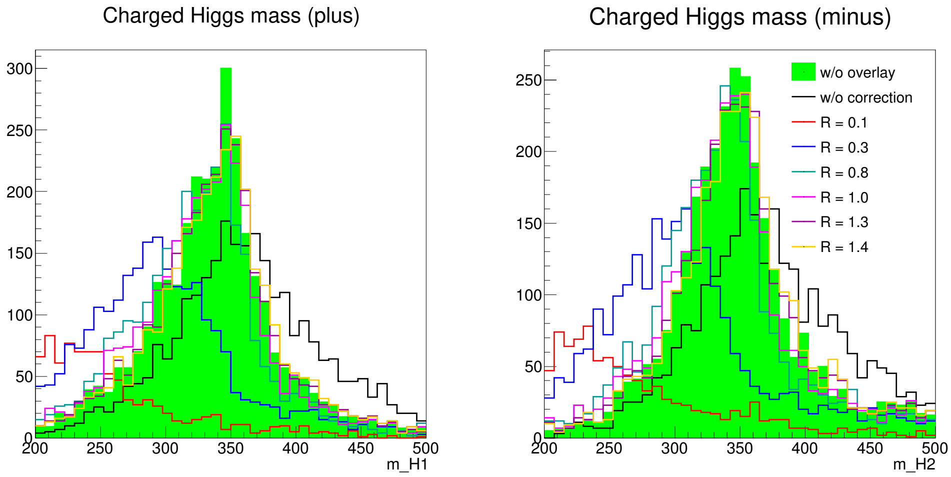

Figure A.1: Charged Higgs mass (right: m_{H^+}, left: m_{H^-}) in green $\gamma\gamma$-background removed by generator in formation, black without any correction and other colors with corrected with kt-algorithm with varied R (see legend)

keyword	weight	generated events	expected events	description
h2dm340_h_r	$3.39 \cdot 10^{-2}$	9702	$3.29 \cdot 10^{2}$	
h2dm340_h_l	$4.48 \cdot 10^{-1}$	9900	$4.44 \cdot 10^{3}$	

keyword	weight	generated events	expected events	description
h2dm346_h_r	$3.32 \cdot 10^{-2}$	9900	$3.29 \cdot 10^{2}$	
h2dm346_h_l	$4.48 \cdot 10^{-1}$	9900	$4.44 \cdot 10^{3}$	
h2dm348_h_r	$3.32 \cdot 10^{-2}$	9900	$3.29 \cdot 10^{2}$	
h2dm348_h_l	$4.87 \cdot 10^{-1}$	9108	$4.44 \cdot 10^{3}$	
h2dm350_h_r	$1.67 \cdot 10^{-2}$	19602	$3.29 \cdot 10^{2}$	
h2dm350_h_l	$2.26 \cdot 10^{-1}$	19602	$4.44 \cdot 10^{3}$	
h2dm352_h_r	$3.39 \cdot 10^{-2}$	9702	$3.29 \cdot 10^{2}$	
h2dm352_h_l	$4.48 \cdot 10^{-1}$	9900	$4.44 \cdot 10^{3}$	
h2dm354_h_r	$3.41 \cdot 10^{-2}$	9648	$3.29 \cdot 10^{2}$	
h2dm354_h_l	$4.48 \cdot 10^{-1}$	9900	$4.44 \cdot 10^{3}$	
h2dm360_h_r	$3.32 \cdot 10^{-2}$	9900	$3.29 \cdot 10^{2}$	
h2dm360_h_l	$4.67 \cdot 10^{-1}$	9504	$4.44 \cdot 10^{3}$	
h2dm340_slwm_r	$1.63 \cdot 10^{-2}$	9702	$1.58 \cdot 10^{2}$	
h2dm340_slwm_l	$2.20 \cdot 10^{-1}$	9702	$2.14 \cdot 10^{3}$	
h2dm340_slwp_r	$1.63 \cdot 10^{-2}$	9702	$1.58 \cdot 10^{2}$	
h2dm340_slwp_l	$2.20 \cdot 10^{-1}$	9702	$2.14 \cdot 10^{3}$	
h2dm346_slwm_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm346_slwm_l	$2.16 \cdot 10^{-1}$	9900	$2.14 \cdot 10^{3}$	
h2dm346_slwp_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm346_slwp_l	$2.29 \cdot 10^{-1}$	9306	$2.14 \cdot 10^{3}$	
h2dm348_slwm_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm348_slwm_l	$2.20 \cdot 10^{-1}$	9702	$2.14 \cdot 10^{3}$	
h2dm348_slwp_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm348_slwp_l	$2.20 \cdot 10^{-1}$	9702	$2.14 \cdot 10^{3}$	
h2dm350_slwm_r	$8.00 \cdot 10^{-3}$	19800	$1.58 \cdot 10^{2}$	
h2dm350_slwm_l	$1.08 \cdot 10^{-1}$	19800	$2.14 \cdot 10^{3}$	
h2dm350_slwp_r	$8.16 \cdot 10^{-3}$	19404	$1.58 \cdot 10^{2}$	
h2dm350_slwp_l	$1.11 \cdot 10^{-1}$	19206	$2.14 \cdot 10^{3}$	
h2dm352_slwm_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm352_slwm_l	$2.25 \cdot 10^{-1}$	9504	$2.14 \cdot 10^{3}$	
h2dm352_slwp_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm352_slwp_l	$2.20 \cdot 10^{-1}$	9702	$2.14 \cdot 10^{3}$	
h2dm354_slwm_r	$1.66 \cdot 10^{-2}$	9504	$1.58 \cdot 10^{2}$	
h2dm354_slwm_l	$2.16 \cdot 10^{-1}$	9900	$2.14 \cdot 10^{3}$	
h2dm354_slwp_r	$1.63 \cdot 10^{-2}$	9702	$1.58 \cdot 10^{2}$	
h2dm354_slwp_l	$2.16 \cdot 10^{-1}$	9900	$2.14 \cdot 10^{3}$	
h2dm360_slwm_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	

keyword	weight	generated events	expected events	description
h2dm360_slwm_l	$2.16 \cdot 10^{-1}$	9900	$2.14 \cdot 10^{3}$	
h2dm360_slwp_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm360_slwp_l	$2.25 \cdot 10^{-1}$	9504	$2.14 \cdot 10^{3}$	
h2dm_h_r	$3.39 \cdot 10^{-2}$	9702	$3.29 \cdot 10^{2}$	
h2dm_h_l	$4.67 \cdot 10^{-1}$	9504	$4.44 \cdot 10^{3}$	
h2dm_slwp_r	$1.60 \cdot 10^{-2}$	9900	$1.58 \cdot 10^{2}$	
h2dm_slwp_l	$2.16 \cdot 10^{-1}$	9900	$2.14 \cdot 10^{3}$	
h2dm_slwm_r	$1.66 \cdot 10^{-2}$	9504	$1.58 \cdot 10^{2}$	
h2dm_slwm_l	$2.16 \cdot 10^{-1}$	9900	$2.14 \cdot 10^{3}$	
2f_h_r	6.49	32032	$2.08 \cdot 10^{5}$	
2f_h_l	$6.94 \cdot 10$	72859	$5.06 \cdot 10^{6}$	
ttz_r	$2.40 \cdot 10^{-2}$	7253	$1.74 \cdot 10^{2}$	
ttz_l	2.08	3627	$7.57 \cdot 10^{3}$	
ttbb_r	$1.70 \cdot 10^{-2}$	3569	$6.06 \cdot 10$	
ttbb_l	$9.45 \cdot 10^{-1}$	1959	$1.85 \cdot 10^{3}$	
6f_ttbar_sl_r0	$4.33 \cdot 10^{-2}$	17191	$7.45 \cdot 10^{2}$	
6f_ttbar_sl_l0	$2.83 \cdot 10^{-1}$	128593	$3.64 \cdot 10^{4}$	
6f_ttbar_sl_r1	$6.39 \cdot 10^{-2}$	23345	$1.49 \cdot 10^{3}$	
6f_ttbar_sl_l1	$3.13 \cdot 10^{-1}$	200031	$6.26 \cdot 10^{4}$	
6f_ttbar_sl_r2	$4.35 \cdot 10^{-2}$	17141	$7.46 \cdot 10^{2}$	
6f_ttbar_sl_l2	$2.89 \cdot 10^{-1}$	127841	$3.69 \cdot 10^{4}$	
6f_ttbar_sl_r3	$6.46 \cdot 10^{-2}$	23094	$1.49 \cdot 10^{3}$	
6f_ttbar_sl_l3	$3.17 \cdot 10^{-1}$	198319	$6.28 \cdot 10^{4}$	
6f_ttbar_h_r0	$8.38 \cdot 10^{-2}$	13121	$1.10 \cdot 10^{3}$	
6f_ttbar_h_l0	$3.77 \cdot 10^{-1}$	121032	$4.56 \cdot 10^{4}$	
6f_ttbar_h_r1	$9.17 \cdot 10^{-2}$	11989	$1.10 \cdot 10^{3}$	
6f_ttbar_h_l1	$4.60 \cdot 10^{-1}$	99284	$4.56 \cdot 10^{4}$	
6f_ttbar_h_r2	$9.04 \cdot 10^{-2}$	12156	$1.09 \cdot 10^{3}$	
6f_ttbar_h_l2	$4.09 \cdot 10^{-1}$	111215	$4.55 \cdot 10^{4}$	
6f_ttbar_h_r3	$9.89 \cdot 10^{-2}$	11155	$1.10 \cdot 10^{3}$	
6f_ttbar_h_l3	$4.35 \cdot 10^{-1}$	105362	$4.58 \cdot 10^{4}$	
6f_other0	$1.38 \cdot 10^{-1}$	1000	$1.38 \cdot 10^{2}$	$xxW^+W^- \rightarrow xxveev$
6f_other1	1.65	1000	$1.65 \cdot 10^{3}$	$xxW^+W^- \rightarrow xxveev$
6f_other2	$2.20 \cdot 10^{-3}$	1000	2.20	$xxW^+W^- \rightarrow xxveev$
6f_other3	$2.22 \cdot 10^{-2}$	1000	$2.22 \cdot 10$	$xxW^+W^- \rightarrow xxveev$
6f_other4	$2.30 \cdot 10^{-1}$	1000	$2.30 \cdot 10^{2}$	$xxW^+W^- \rightarrow xxvelv$
6f_other5	1.71	1000	$1.71 \cdot 10^{3}$	$xxW^+W^- \rightarrow xxvelv$

keyword	weight	generated events	expected events	description
6f_other6	$4.41 \cdot 10^{-4}$	999	$4.41 \cdot 10^{-1}$	$xxW^+W^- \to$xxvelv
6f_other7	$6.62 \cdot 10^{-1}$	1000	$6.62 \cdot 10^{2}$	$xxW^+W^- \to$xxveyx
6f_other8	2.57	1930	$4.96 \cdot 10^{3}$	$xxW^+W^- \to$xxveyx
6f_other9	$1.17 \cdot 10^{-3}$	1000	1.17	$xxW^+W^- \to$xxveyx
6f_other10	1.70	1000	$1.70 \cdot 10^{3}$	$xxW^+W^- \to$xxvlev
6f_other11	$4.41 \cdot 10^{-4}$	1000	$4.41 \cdot 10^{-1}$	$xxW^+W^- \to$xxvlev
6f_other12	$3.81 \cdot 10^{-2}$	999	$3.81 \cdot 10$	$xxW^+W^- \to$xxvlev
6f_other13	2.36	1000	$2.36 \cdot 10^{3}$	$xxW^+W^- \to$xxvllv
6f_other14	$2.82 \cdot 10^{-3}$	1000	2.82	$xxW^+W^- \to$xxvllv
6f_other15	2.57	1914	$4.92 \cdot 10^{3}$	$xxW^+W^- \to$xxvlyx
6f_other16	$2.35 \cdot 10^{-3}$	1000	2.35	$xxW^+W^- \to$xxvlyx
6f_other17	2.57	1927	$4.95 \cdot 10^{3}$	$xxW^+W^- \to$xxxyev
6f_other18	$1.18 \cdot 10^{-3}$	1000	1.18	$xxW^+W^- \to$xxxyev
6f_other19	$1.10 \cdot 10^{-1}$	1000	$1.10 \cdot 10^{2}$	$xxW^+W^- \to$xxxyev
6f_other20	2.57	1914	$4.92 \cdot 10^{3}$	$xxW^+W^- \to$xxxylv
6f_other21	$2.35 \cdot 10^{-3}$	1000	2.35	$xxW^+W^- \to$xxxylv
6f_other22	$4.14 \cdot 10^{-2}$	1000	$4.14 \cdot 10$	xxZ $\to$xxxxxx
6f_other23	$8.39 \cdot 10^{-4}$	1000	$8.39 \cdot 10^{-1}$	xxZ $\to$xxxxxx
6f_other24	$8.82 \cdot 10^{-1}$	1000	$8.82 \cdot 10^{2}$	xxZ $\to$xxxxvv
6f_other25	$1.79 \cdot 10^{-3}$	1000	1.79	xxZ $\to$xxxxvv
6f_other26	$5.70 \cdot 10^{-2}$	1000	$5.70 \cdot 10$	xxZ $\to$xxxxll
6f_other27	$1.45 \cdot 10^{-3}$	1000	1.45	xxZ $\to$xxxxll
6f_other28	$3.11 \cdot 10^{-2}$	1000	$3.11 \cdot 10$	xxZ $\to$xxxxee
6f_other29	$8.22 \cdot 10^{-2}$	1000	$8.22 \cdot 10$	xxZ $\to$xxxxee
6f_other30	$3.53 \cdot 10^{-3}$	1000	3.53	xxZ $\to$xxxxee
6f_other31	$5.24 \cdot 10^{-3}$	1000	5.24	xxZ $\to$xxxxee
6f_other32	1.38	999	$1.38 \cdot 10^{3}$	xxZ $\to$vvvvxx
6f_other33	$1.29 \cdot 10^{-3}$	996	1.28	xxZ $\to$vvvvxx
6f_other34	2.52	999	$2.52 \cdot 10^{3}$	xxZ $\to$vvvvyy
6f_other35	$2.00 \cdot 10^{-3}$	998	1.99	xxZ $\to$vvvvyy
6f_other36	$1.16 \cdot 10^{-1}$	984	$1.14 \cdot 10^{2}$	$\ell^+\ell^-W^+W^- \to$llvelv
6f_other37	$7.42 \cdot 10^{-1}$	987	$7.33 \cdot 10^{2}$	$\ell^+\ell^-W^+W^- \to$llvelv
6f_other38	$2.14 \cdot 10^{-4}$	991	$2.12 \cdot 10^{-1}$	$\ell^+\ell^-W^+W^- \to$llvelv
6f_other39	$3.31 \cdot 10^{-1}$	1000	$3.31 \cdot 10^{2}$	$\ell^+\ell^-W^+W^- \to$llveyx
6f_other40	2.18	1000	$2.18 \cdot 10^{3}$	$\ell^+\ell^-W^+W^- \to$llveyx
6f_other41	$5.69 \cdot 10^{-4}$	1000	$5.69 \cdot 10^{-1}$	$\ell^+\ell^-W^+W^- \to$llveyx
6f_other42	$7.38 \cdot 10^{-1}$	991	$7.32 \cdot 10^{2}$	$\ell^+\ell^-W^+W^- \to$llvlev

keyword	weight	generated events	expected events	description
6f_other43	$2.13 \cdot 10^{-4}$	987	$2.11 \cdot 10^{-1}$	$\ell^+\ell^-W^+W^- \to$llvlev
6f_other44	$1.96 \cdot 10^{-2}$	979	$1.92 \cdot 10$	$\ell^+\ell^-W^+W^- \to$llvlev
6f_other45	$7.96 \cdot 10^{-1}$	978	$7.79 \cdot 10^{2}$	$\ell^+\ell^-W^+W^- \to$llvllv
6f_other46	$9.42 \cdot 10^{-4}$	975	$9.19 \cdot 10^{-1}$	$\ell^+\ell^-W^+W^- \to$llvllv
6f_other47	1.92	1000	$1.92 \cdot 10^{3}$	$\ell^+\ell^-W^+W^- \to$llvlyx
6f_other48	$1.13 \cdot 10^{-3}$	1000	1.13	$\ell^+\ell^-W^+W^- \to$llvlyx
6f_other49	2.17	1000	$2.17 \cdot 10^{3}$	$\ell^+\ell^-W^+W^- \to$llxyev
6f_other50	$5.69 \cdot 10^{-4}$	1000	$5.69 \cdot 10^{-1}$	$\ell^+\ell^-W^+W^- \to$llxyev
6f_other51	$5.59 \cdot 10^{-2}$	1000	$5.59 \cdot 10$	$\ell^+\ell^-W^+W^- \to$llxyev
6f_other52	1.92	1000	$1.92 \cdot 10^{3}$	$\ell^+\ell^-W^+W^- \to$llxylv
6f_other53	$1.13 \cdot 10^{-3}$	1000	1.13	$\ell^+\ell^-W^+W^- \to$llxylv
6f_other54	2.57	2258	$5.80 \cdot 10^{3}$	$\ell^+\ell^-W^+W^- \to$llxyyx
6f_other55	$7.28 \cdot 10^{-3}$	1000	7.28	$\ell^+\ell^-W^+W^- \to$llxyyx
4f_h0	$2.32 \cdot 10$	6994	$1.62 \cdot 10^{5}$	
4f_h1	1.72	2677	$4.61 \cdot 10^{3}$	
4f_h2	$2.32 \cdot 10$	77835	$1.81 \cdot 10^{6}$	
4f_h3	1.72	204	$3.51 \cdot 10^{2}$	
4f_h4	$2.32 \cdot 10$	64851	$1.50 \cdot 10^{6}$	
4f_h5	1.72	673	$1.16 \cdot 10^{3}$	
4f_sl0	$1.72 \cdot 10^{-1}$	3246	$5.59 \cdot 10^{2}$	
4f_sl1	2.66	835127	$2.22 \cdot 10^{6}$	
4f_sl2	$7.66 \cdot 10$	1796	$1.37 \cdot 10^{5}$	
4f_sl3	1.72	2702	$4.67 \cdot 10^{3}$	
4f_sl4	$1.62 \cdot 10$	75941	$1.23 \cdot 10^{6}$	
4f_sl5	1.72	580	$1.00 \cdot 10^{3}$	
4f_sl6	$3.44 \cdot 10$	24426	$8.41 \cdot 10^{5}$	
4f_sl7	$4.93 \cdot 10$	28218	$1.39 \cdot 10^{6}$	
4f_sl8	8.18	17105	$1.40 \cdot 10^{5}$	
4f_sl9	$1.60 \cdot 10^{2}$	565	$9.06 \cdot 10^{4}$	
4f_sl10	2.76	205344	$5.68 \cdot 10^{5}$	
4f_sl11	2.42	1999505	$4.84 \cdot 10^{6}$	
4f_sl12	$3.57 \cdot 10^{-1}$	265096	$9.46 \cdot 10^{4}$	
4f_sl13	$1.72 \cdot 10^{-1}$	1623	$2.79 \cdot 10^{2}$	
4f_WW_l_r	$1.16 \cdot 10^{-1}$	496	$5.75 \cdot 10$	
4f_WW_l_l	$6.64 \cdot 10$	2777	$1.84 \cdot 10^{5}$	
1f_3f0	9.62	572927	$5.51 \cdot 10^{6}$	
1f_3f1	9.51	1443407	$1.37 \cdot 10^{7}$	

keyword	weight	generated events	expected events	description
1f_3f2	$2.33 \cdot 10^2$	44683	$1.04 \cdot 10^7$	
1f_3f3	$4.69 \cdot 10^2$	51925	$2.43 \cdot 10^7$	
1f_3f4	$1.54 \cdot 10$	72726	$1.12 \cdot 10^6$	
1f_3f5	$3.18 \cdot 10$	82236	$2.62 \cdot 10^6$	
1f_3f6	$2.51 \cdot 10$	70294	$1.77 \cdot 10^6$	
1f_3f7	$1.71 \cdot 10^2$	25600	$4.39 \cdot 10^6$	
1f_3f8	$4.00 \cdot 10$	4057	$1.62 \cdot 10^5$	
1f_3f9	$1.87 \cdot 10$	21023	$3.94 \cdot 10^5$	
1f_3f10	$1.14 \cdot 10^4$	8	$9.18 \cdot 10^4$	
1f_3f11	$1.69 \cdot 10^4$	15	$2.53 \cdot 10^5$	
1f_3f12	$3.77 \cdot 10^3$	222	$8.38 \cdot 10^5$	
1f_3f13	$3.64 \cdot 10^4$	59	$2.15 \cdot 10^6$	
1f_3f14	$4.76 \cdot 10^3$	261	$1.24 \cdot 10^6$	
1f_3f15	$3.19 \cdot 10^4$	97	$3.10 \cdot 10^6$	
1f_3f16	6.68	549482	$3.67 \cdot 10^6$	
1f_3f17	6.07	1513034	$9.18 \cdot 10^6$	
1f_3f18	$1.39 \cdot 10^3$	3220	$4.50 \cdot 10^6$	
1f_3f19	$1.61 \cdot 10^2$	65037	$1.04 \cdot 10^7$	
1f_3f20	$7.69 \cdot 10$	90483	$6.96 \cdot 10^6$	
1f_3f21	$1.95 \cdot 10^2$	83564	$1.63 \cdot 10^7$	
1f_3f22	$1.31 \cdot 10^4$	448	$5.87 \cdot 10^6$	
1f_3f23	$3.90 \cdot 10^4$	343	$1.34 \cdot 10^7$	
1f_3f24	$3.51 \cdot 10^4$	252	$8.85 \cdot 10^6$	
1f_3f25	$2.10 \cdot 10^5$	96	$2.01 \cdot 10^7$	
1f_3f26	$4.12 \cdot 10^5$	38	$1.56 \cdot 10^7$	
1f_3f27	$1.75 \cdot 10^5$	149	$2.61 \cdot 10^7$	
1f_3f28	$7.37 \cdot 10^5$	32	$2.35 \cdot 10^7$	
1f_3f29	$5.63 \cdot 10^6$	7	$3.94 \cdot 10^7$	
1f_3f30	$7.75 \cdot 10$	8420	$6.52 \cdot 10^5$	
1f_3f31	$6.33 \cdot 10$	24951	$1.57 \cdot 10^6$	
1f_3f32	$2.04 \cdot 10^2$	5779	$1.18 \cdot 10^6$	
1f_3f33	$3.59 \cdot 10$	81398	$2.92 \cdot 10^6$	
tth_sl_r	$7.49 \cdot 10^{-3}$	3590	$2.68 \cdot 10$	
tth_sl_l	$3.59 \cdot 10^{-1}$	2245	$8.07 \cdot 10^2$	
tth_slnobb_r	$5.84 \cdot 10^{-3}$	3358	$1.96 \cdot 10$	
tth_slnobb_l	$1.09 \cdot 10^{-1}$	5394	$5.89 \cdot 10^2$	
tth_h_r	$8.83 \cdot 10^{-3}$	3161	$2.79 \cdot 10$	

keyword	weight	generated events	expected events	description
tth_h_l	$4.78 \cdot 10^{-1}$	1752	$8.38 \cdot 10^{2}$	
tth_hnobb_r	$5.38 \cdot 10^{-3}$	3787	$2.03 \cdot 10$	
tth_hnobb_l	$1.25 \cdot 10^{-1}$	4894	$6.12 \cdot 10^{2}$	
tth_l0	$2.43 \cdot 10^{-1}$	800	$1.94 \cdot 10^{2}$	
tth_l1	$1.61 \cdot 10^{-2}$	400	6.47	
tth_l2	$2.36 \cdot 10^{-1}$	600	$1.41 \cdot 10^{2}$	
tth_l3	$1.18 \cdot 10^{-2}$	400	4.72	

Table A.1: List of all used data samples; x stands for up-type quarks; y for down-type quarks; l for muon and tau leptons

add descriptions