LC-REP-2013-004

Measurement of the top Yukawa coupling at $\sqrt{s} = 1$ TeV using the ILD detector

Tony Price[a], Tomohiko Tanabe[b], Keisuke Fujii[c], Victoria Martin[d], Nigel Watson[a]

[(a)]University of Birmingham, United Kingdom; [(b)]ICEPP, The University of Tokyo, Japan; [(c)]High Energy Accelerator Research Organization, Tsukuba, Japan; [(d)]University of Edinburgh, United Kingdom

Abstract

We evaluate the potential of the International Large Detector (ILD) to measure the top Yukawa coupling of a 125 GeV Higgs boson at the International Linear Collider (ILC) operating at $\sqrt{s}$=1 TeV. The hadronic and semileptonic decay modes of the process, $e^+e^- \to t\bar{t}H$, are considered with the Higgs boson decaying via the dominant $b\bar{b}$ channel. Two methods are used to perform the analysis, a cut based approach and a multivariate approach using the TMVA package. When combining the hadronic and semileptonic decay modes a statistical uncertainty on the top Yukawa coupling of 4.3% can be achieved using the multivariate approach.

1 Introduction

The International Linear Collider (ILC) is a proposed e^+e^- linear collider which will operate at centre-of-mass energies up to 1 TeV. Due to the nature of e^+e^- colliders the state of the initial interaction can be controlled and calculated with a high degree of certainty and thus high precision measurements can be achieved. This will allow a rich physics program to be completed including accurate measurements of the couplings of the Higgs boson to matter. According to the Standard Model (SM) of particle physics, the strength of the Yukawa coupling of the Higgs boson to matter fermions scales with the mass of these particles by

$$g_{ffH} = \frac{M_f}{v} \tag{1}$$

where M_f is the mass of fermion and $v = 246$ GeV is the vacuum expectation value.

Figure 1 demonstrates the expected experimental precision of the Yukawa couplings of the Higgs boson after running several years at each center-of-mass energies at 250 GeV, 500 GeV, and 1 TeV. As the top quark is the heaviest

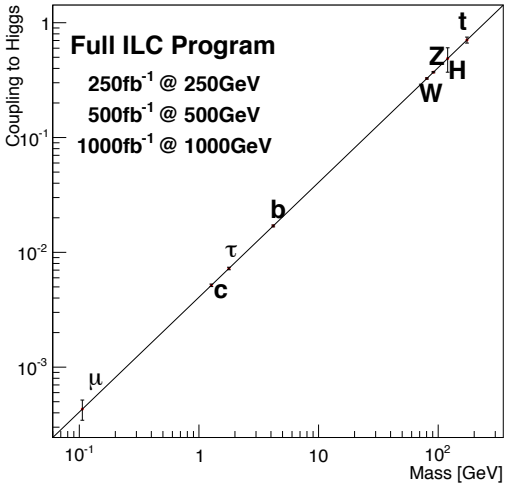

Figure 1: Expected experimental precision of the Yukawa couplings of the Higgs boson following a full ILC program running at 250 GeV, 500 GeV, and 1 TeV [1].

known particle in the SM, its Yukawa coupling is expected to be the largest. Any deviations away from the SM in the

couplings would be an indication of new physics. It is thus essential to evaluate the capability of the ILC to measure the top Yukawa coupling. It is also worthwhile to note that the measurement of the top quark polarization combined with the beam polarization information offers the possibility of determining structure of the $t\bar{t}H$ coupling, although this is beyond the scope of this analysis.

In principle, the top Yukawa coupling could be measured indirectly. This could be achieved either through top pair production near threshold provided that the theoretical uncertainties are reduced to $O(1)\%$ level, or via the $H\to gg$ decay channel where the top quark contribution dominates in the loop. However, whilst the top quark, due to it having the largest mass, is favoured in the loop there are other interfering processes which are reliant on Higgs couplings which lead to increased theoretical uncertainties. Direct measurements of the top Yukawa coupling is thus desired.

At the ILC, the direct measurement of the top Yukawa coupling becomes possible starting at around $\sqrt{s} = 500\,\text{GeV}$ via the $e^+e^- \to t\bar{t}H$ reaction. The optimal centre-of-mass energies to measure the top Yukawa coupling are $\sqrt{s} = 700$–$800\,\text{GeV}$, around which the $e^+e^- \to t\bar{t}H$ production cross section attains its maximum. However, one must also keep in mind the concurrency of various measurements at the ILC. By far the most demanding in terms of integrated luminosity is the Higgs self-coupling measurement. At around $\sqrt{s} = 500\,\text{GeV}$ the $e^+e^- \to ZHH$ reaction reaches its maximum cross section, making 500 GeV the first important energy for this measurement. The $e^+e^- \to \nu\bar{\nu}HH$ reaction via WW fusion rises at higher energies, which demands a centre-of-mass energy as high as possible. At the ILC, $\sqrt{s} = 1\,\text{TeV}$ is the highest centre-of-mass energy that is considered to be technologically feasible. It also remains to show that the ILC detectors are capable of performing physics measurements at $\sqrt{s} = 1\,\text{TeV}$. Several reactions were chosen as detector benchmarks including the top Yukawa coupling measurement.

The final state consists of eight fermions with at least four jets, potential leptons, and missing energy and this study will also demonstrate the ability of the ILD detector to reconstruct complicated topologies.

In this note, the precision of the top Yukawa coupling is estimated using the International Large Detector (ILD) concept.

2 Signal and Backgrounds

Figure 2 summarizes the cross sections for signal and typical background processes as a function of the center-of-mass energy. Figure 3 illustrates the lowest order Feynman diagrams for the process $e^+e^- \to t\bar{t}H$. The diagram $e^+e^- \to ZH$

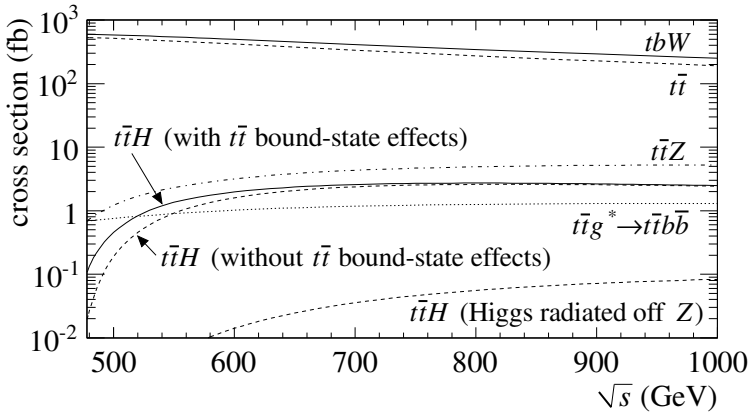

Figure 2: Summary of cross sections for signal and background processes for unpolarized initial beams [2].

(Higgs-strahlung) with $Z\to t\bar{t}$ which does not contain the top Yukawa coupling has small yet non-negligible contribution to the total cross section. The size of this effect is studied by looking at how the $e^+e^- \to t\bar{t}H$ cross section varies when modifying the top Yukawa coupling from the SM value using the cross section calculation of the event generator. The result is shown in Fig. 4 for $e^-_L e^+_R$ initial state (left) and $e^-_R e^+_L$ initial state (right). It is found that a quadratic curve models the behavior of the cross section dependence on the top Yukawa coupling quite well. The slope $d\sigma/dy_t$ is extracted at the SM value of the top Yukawa coupling to compute the factor needed to convert the precision in the cross

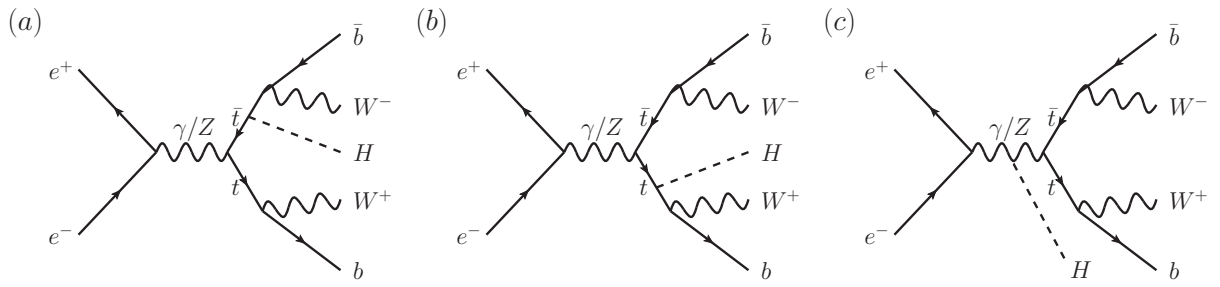

Figure 3: The lowest order Feynman diagrams for the process $e^+e^- \to t\bar{t}H$. In (a) and (b) the Higgs boson is radiated from a top quark and (c) is the background Higgs-strahlung process where the Higgs boson is radiated from the Z boson.

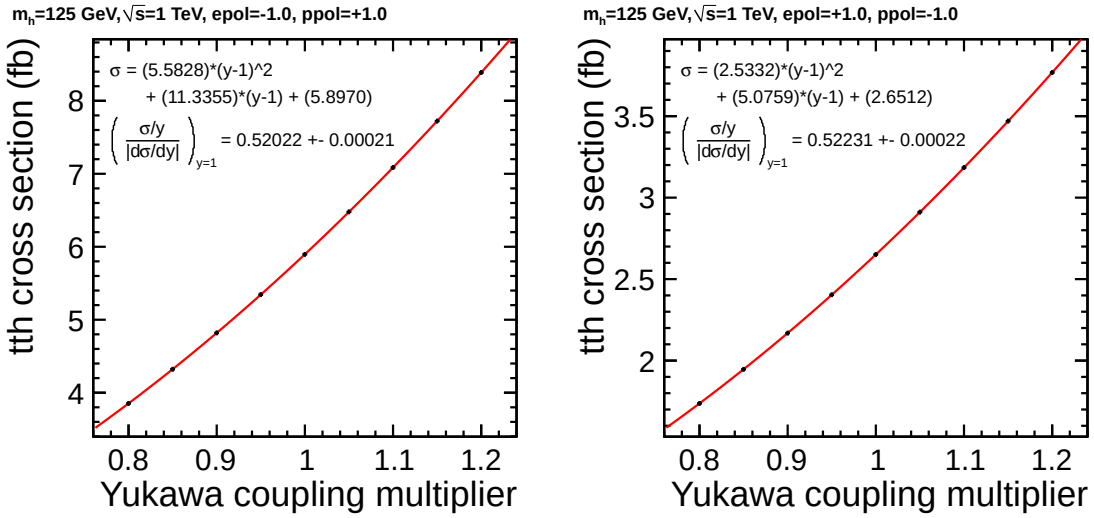

Figure 4: Sensitivity of the cross section to the top Yukawa coupling for $e_L^- e_R^+$ initial state (left) and $e_R^- e_L^+$ initial state (right).

section into the precision of the top Yukawa coupling, using the following equation:

$$\frac{\Delta y_t}{y_t} = \left(\frac{\sigma/y_t}{|d\sigma/dy_t|}\right)_{y_t=y_t(SM)} \frac{\Delta\sigma}{\sigma}. \quad (2)$$

The pre-factor in the r.h.s. is exactly equal to 1/2 if there is no contribution of the Higgs-strahlung diagram. In actuality, we find that $\left(\frac{\sigma/y_t}{|d\sigma/dy_t|}\right)_{y_t=y_t(SM)} = 0.52$. which indicates that the Higgs-strahlung contribution is about 4% of the total cross section. This factor will be used in the final extraction of the top Yukawa coupling precision.

For this study the semileptonic and hadronic decays of the $t\bar{t}$ system were studied with the Higgs decaying via the dominant decay mode into a $b\bar{b}$ pair. For the fully hadronic decay channel this leads to a signal of 8 jets, 4 of which should be tagged as b-jets. In the semileptonic mode the final signal in the detector consists of 6 jets, 4 of which should be tagged as b-jets, an isolated lepton, and missing energy and momentum from a neutrino. These two competing modes act not only as a signal process but also as a background to each other.

Irreducible backgrounds to these processes arise from the 8 fermion final states of $t\bar{t}Z$ where the Z decays into a $b\bar{b}$ pair and $t\bar{t}b\bar{b}$ where the $t\bar{t}$ system radiates a hard gluon which forms a $b\bar{b}$ pair. A large background contribution arises from $t\bar{t}$ due to the huge relative cross section compared to the signal. There is also a contribution from the other decay modes of the $t\bar{t}H$ system such as the Higgs not decaying to a $b\bar{b}$ pair and the fully leptonic decays of the top quarks.

The signal is selected using the variables described in Section 6 via the methods described in Sections 7 and 8.

3 Sample Generation

The samples for this study were generated using two event generators, WHIZARD v1.95 [3], and PHYSSIM [4]. The $t\bar{t}$ samples were generated using WHIZARD whereas the $t\bar{t}H$, $t\bar{t}b\bar{b}$, and $t\bar{t}Z$ samples used PHYSSIM due to its improved phase space integration time for the eight fermion final states compared to WHIZARD. We include as background all six fermion final states generated with WHIZARD which are compatible with the final states of the $t\bar{t}$ process; these processes will be called "$t\bar{t}$" processes despite their containing contributions which are non-resonant in $t\bar{t}$.

The detector simulations were conducted using the Mokka/GEANT4 software with the **ildconfig-v01-14-01-p00** configuration and the **ILD_o1_v5** detector geometry model. The "o1" model indicates the choice of silicon-tungsten electromagnetic calorimeter with 5×5 mm^2 cell sizes, and scintillator-steel hadronic calorimeter with 3×3 cm^2. The simulation detector model is validated by software experts for correctness of the material and geometry, such as gaps and/or overlaps. The $\gamma\gamma \to$ hadrons pile-up backgrounds are overlayed at the level of simulated hits, with an average of 4.1 interactions per bunch crossing. The reconstruction used software packages consistent with the versions in iLCSoft v01-16 [5, 6] (ildconfig-v01-16-p03) including; MarlinTrk, a new Kalman-filter based track finder; PandoraPFA [7] particle flow algorithm, and LCFIPlus [8] for jet clustering and flavour tagging. The analysis was carried out using the Data Summary Tables (DSTs) which were centrally produced at DESY [9] and reweighted to give an integrated luminosity of 1000 fb^{-1} split equally between two polarisation states, (e^+, e^-) = (±0.2, ∓0.8).

4 Event Reconstruction

The final state of the $t\bar{t}H$ decays contain eight fermions, eight jets for the hadronic and 6 jets, a lepton and a neutrino in the case of the semileptonic mode. This leads to a large number of possible combinations to reconstruct the final state of two W bosons, two top quarks and a Higgs boson. The optimal combination of jets was found by minimising

$$\chi^2 = \frac{(M_{t_1} - M_t)^2}{\sigma_{t_1}^2} + \frac{(M_{t_2} - M_t)^2}{\sigma_{t_2}^2} + \frac{(M_{b\bar{b}} - M_H)^2}{\sigma_H^2} \quad (3)$$

where the top quarks are formed by combining a b-jet with a W boson, and the Higgs boson from the remaining two b-jets. The W bosons are formed using the four least b-like jets in the hadronic mode, and the two least b-like jets form one W in the semileptonic mode with the other formed from the isolated lepton and neutrino. The resolutions σ_i are taken to be all roughly equal and are found to be insensitive to small variation for the final mass resolutions.

The neutrino was reconstructed assuming that all of the missing momenta is attributed to a neutrino with zero mass. Figure 5 shows that this assumption is valid for events with small amounts of ISR but as the amount of ISR increases

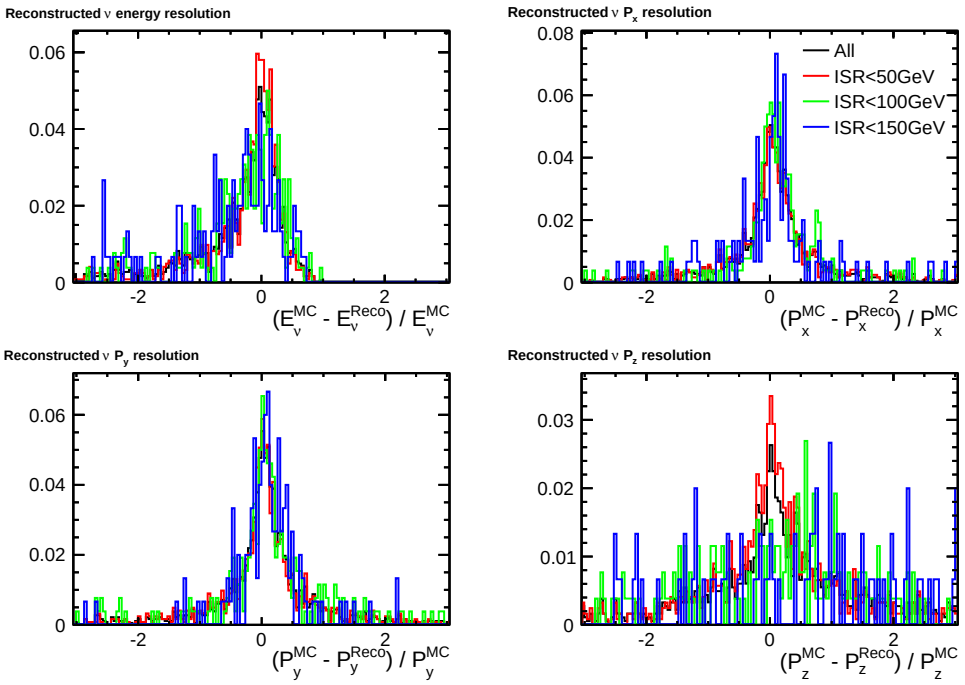

Figure 5: The neutrino reconstruction for varying amounts of ISR and its effect on the energy reconstruction (top left), p_x (top right), p_y (bottom left), and p_z (bottom right).

the neutrino reconstruction suffers. This is primarily observed in the z-direction as the ISR photons have small p_t but large p_z.

The effect of the neutrino reconstruction can be seen in Figure 6 where the leptonic W and top reconstruction has a much broader width than the equivalent hadronic particles. However, this behaviour is expected and as the masses peak in the correct places with the correct shapes this is not an issue. It can also be seen that the reconstructed Higgs mass peaks at 125 GeV as expected and the total mass of the final system has a maximum at the expected 470 GeV. These properties demonstrate that the semileptonic decay mode can be reconstructed well.

As with the semileptonic case, the hadronic events can be successfully reconstructed as demonstrated by Figure 7. The expected behaviour is observed for the W boson, top quarks, Higgs boson, and total masses. The W boson and top quark histograms have twice the statistics of the Higgs boson due to there being two of these particles in each event.

5 $\gamma\gamma \rightarrow$hadrons Overlay Removal

In addition to the underlying event, there is an extra component to the event from unrelated $\gamma\gamma \rightarrow$hadrons "pile-up" events. On average there are 4.1 pile-up events per bunch-crossing leading to an additional 50 GeV in the reconstructed event. Figure 9 demonstrates that without the removal of the pile-up the reconstructed energy in the event is overestimated. This is due to the Durham algorithm forcing all of the particles in an event into the jets. In order to use the Durham algorithm implemented within LCFIPlus these pile-up events must be removed.

The particles in the pile-up events are generally low p_T and happen at angles close to the beam axis as shown in Figure 8. As the pile-up events are, in general, separate from the underlying event they were removed using the kt algorithm with optimised values of R. The optimal values of R was found to be 1.2. The isolated leptons were removed first followed by the $\gamma\gamma \rightarrow$hadrons removal to ensure that none were discarded with the pile-up events.

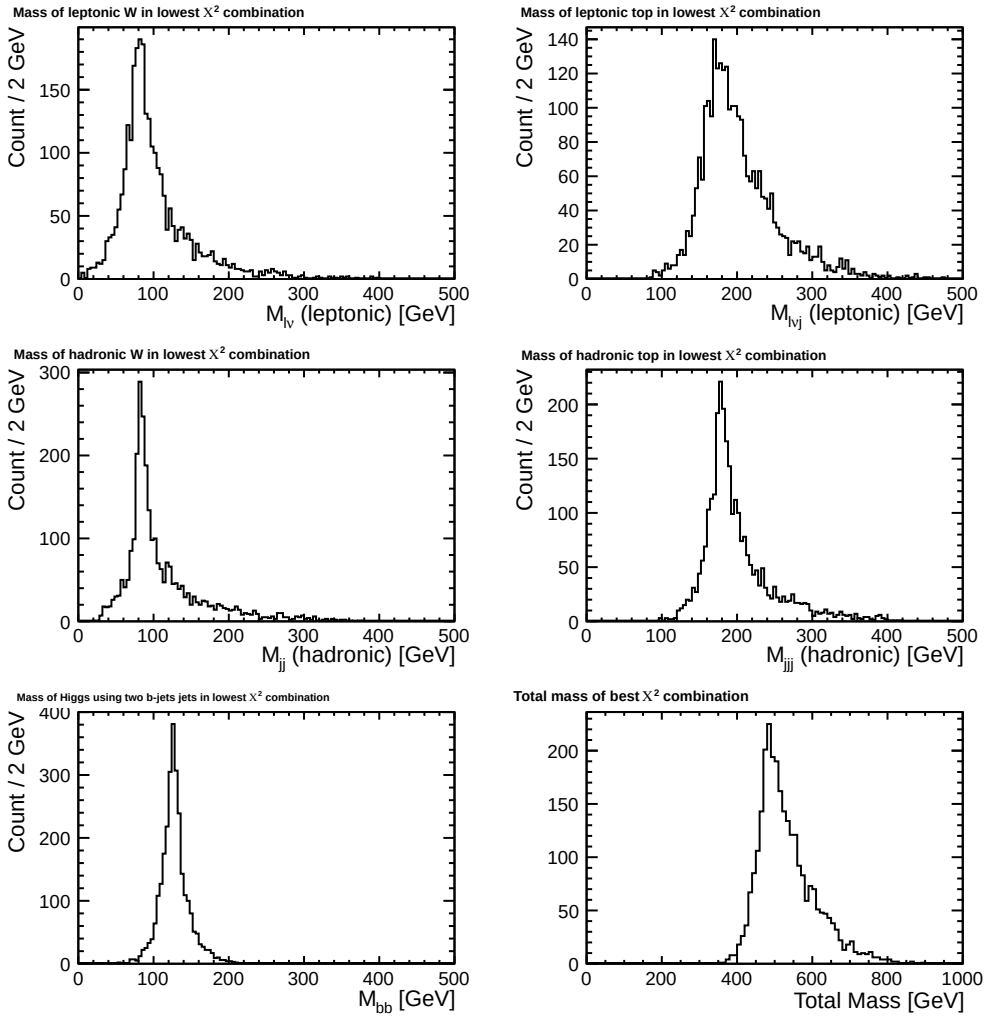

Figure 6: Reconstructed masses for the semileptonic decay mode for the leptonic W boson (top left), leptonic top quark (top right), hadronic W boson (middle left), hadronic top quark (middle right), Higgs boson mass (bottom left), and the total mass of the system (bottom right).

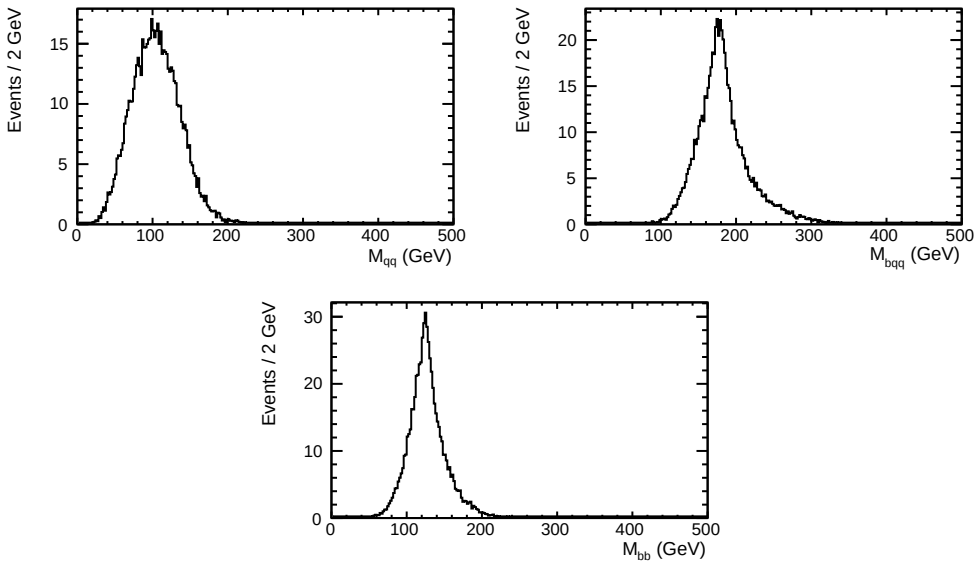

Figure 7: Reconstructed masses for the hadronic decay mode for the W boson (top left), top quark (top right), Higgs boson mass (bottom).

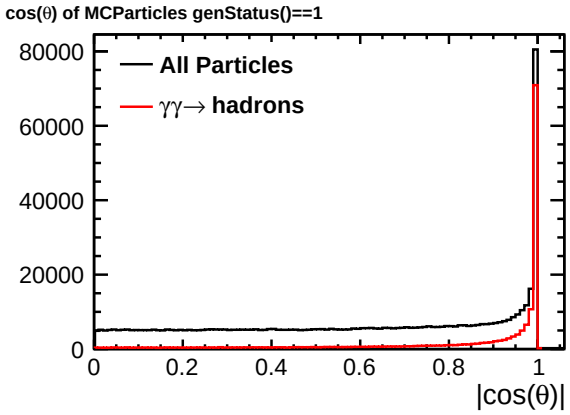

Figure 8: The angle relative to the beam axis of all MCParticles in the events (black), and the $\gamma\gamma \rightarrow$hadrons pile-up only (red).

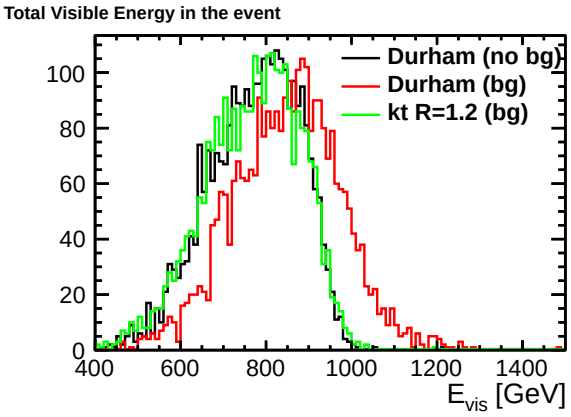

Figure 9: The total visible energy for events with and without $\gamma\gamma \rightarrow$hadrons clustered using the Durham algorithm, and the kt algorithm for the semileptonic mode.

6 Selection Variables

The signal is selected and the background reduced using a series of selection variables. These include the number of isolated leptons in the event, the total visible energy and number of Particle Flow Objects (PFOs) in the final state, the thrust of the event and jet clustering variables from the Durham algorithm (y_{ij}), the number of jets tagged as good b-jets in LCFIPlus, the consistency of the final reconstructed masses with the expected values, and the helicity of the $b\bar{b}$ pair associated with the Higgs boson. Figures 11–15 illustrate the normalised distributions for the cuts used for the semileptonic and hadronic modes and the $t\bar{t}Z$, $t\bar{t}b\bar{b}$, $t\bar{t}H \rightarrow$other, and the $t\bar{t}$ backgrounds.

6.1 Lepton Isolation

To ensure that the samples used for the semileptonic and hadronic modes were completely independent the samples were first split using the number of isolated leptons found utilising the LAL Lepton Finder isolation method as presented here [10]. The PFOs in the events were forced into 8 jets and the isolation of constituent particles within these jets checked using

$$z = \frac{E_{lep}}{E_{jet}} \tag{4}$$

where E_{lep} and E_{jet} are the energies of the lepton and the jet within which the lepton resides, and

$$x_T = \frac{p_T}{M_{jet}} \tag{5}$$

where p_T is the transverse momentum of the lepton and M_{jet} is the mass of the jet. The values of z and p_T were optimised to yield the best performance at values of $x_T > 0.25$, and $z > 0.6$ as shown in Figure 10. Table 1 shows the efficiency of this method for locating isolated leptons, coupled with the purity of the selection and the source of the isolated leptons. An efficiency of 82% (89%) and purity of 95% (97%) for electrons (muons) from W decays is observed.

Lepton	Efficiency	Composition			
		W$\rightarrow$e, μ	W$\rightarrow \tau \rightarrow$e, μ	Other e, μ	Fake e, μ
Electrons	84.0%	94.2%	2.9%	1.6%	2.3%
Muons	90.5%	96.3%	2.4%	1.2%	0.7%

Table 1: The performance of the isolated lepton finder including the efficiency of selecting a genuine lepton from a W decay, and the composition of the selected samples including fake leptons.

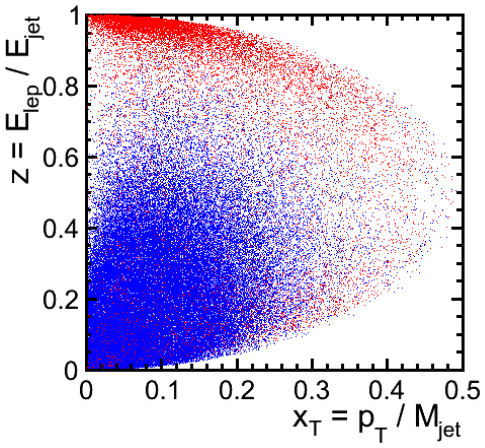

Figure 10: A scatter plot showing the variables used in the lepton isolation for leptons originating from a W boson (red) and all other leptons (blue).

6.2 Total Visible Energy

Events were selected within a window of the total visible energy, E_{vis}^{tot}, in an event in order to remove events with a large number of particles removed alongside the $\gamma\gamma \rightarrow$hadrons due to very forward jets, events with a large amount of ISR, and events which contained multiple neutrinos in the final state.

6.3 Number of Particle Flow Objects

Events with a small number of PFOs were rejected to further protect against backgrounds with multiple neutrinos especially those containing two leptons, two neutrinos, and just four jets. This is of particular relevance for the $t\bar{t}$ background which contains two lepton processes with high cross sections in the final state.

6.4 Thrust

The thrust of an event is given by

$$T = \max \frac{\sum_i |\hat{n} \cdot \vec{p}_i|}{\sum_i |\vec{p}_i|} \tag{6}$$

where p_i is the momentum of the jet. The thrust of an event with two back to back jets is one whereas for an event where the jets are spherically symmetric in the detector is 0.5. As the $t\bar{t}$ system is effectively a two fermion system the thrust of the event is larger than the $t\bar{t}H$, $t\bar{t}Z$, and $t\bar{t}b\bar{b}$.

6.5 Jet Clustering

The signal and backgrounds were clustered using the Durham algorithm into 6 and 8 jet final states. The distance parameter between n and (n+1) jets in the final state is given by

$$Y_{ij} = \frac{\min(E_i^2, E_j^2)(1 - \cos\theta_{ij})}{E_{CM}^2} \tag{7}$$

where i=n and j=(n+1) jets, and $\cos\theta_{ij}$ is the angle between the jets. If the event is forced into more jets than necessary, Y_{ij} becomes small. As the values of i and j are signal dependent the two signals used different values of Y_{ij}. For the semileptonic decay mode the values of Y_{45} and Y_{56} were interesting to study the change between 6 and 7 jets, and the hadronic mode used Y_{67} and Y_{78}.

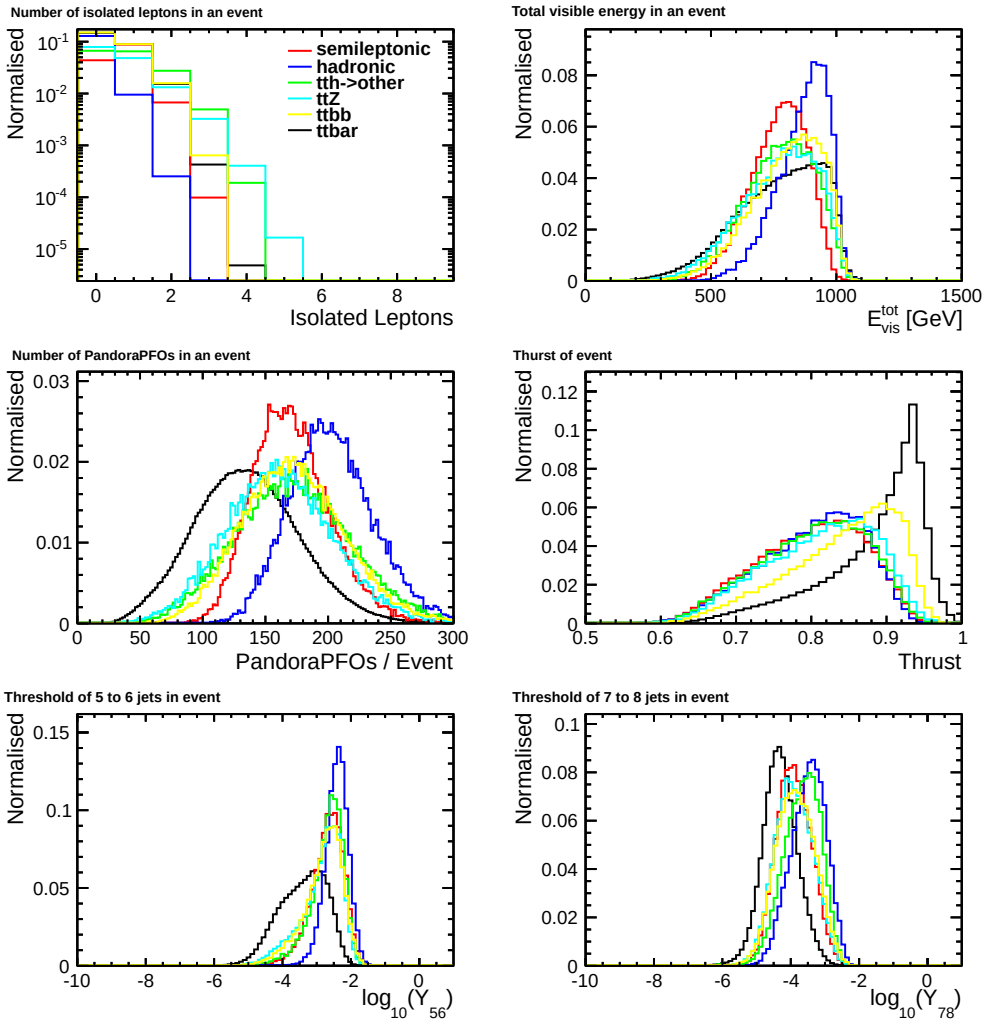

Figure 11: Normalised histograms of the number of identified isolated leptons (top left), total visible energy (top right), number of PandoraPFOs (middle left), thrust (middle right), and the jet parameters (bottom right and bottom left) within the events for the semileptonic (red) and hadronic (blue) modes alongside the backgrounds.

6.6 Flavour Tags

Both the semileptonic and the hadronic signals have four b-jets in the final state, two from the top decays and two from the Higgs decay. The $t\bar{t}$ background will only contain two b-jets from the top decays as will ∼80 % of $t\bar{t}Z$. The flavour tagging is thus a very good discriminant between the signal and backgrounds. The b-tag values from LCFIPlus were ordered from largest to smallest and cuts were made on the third and fourth values to remove the events containing just two b-jets.

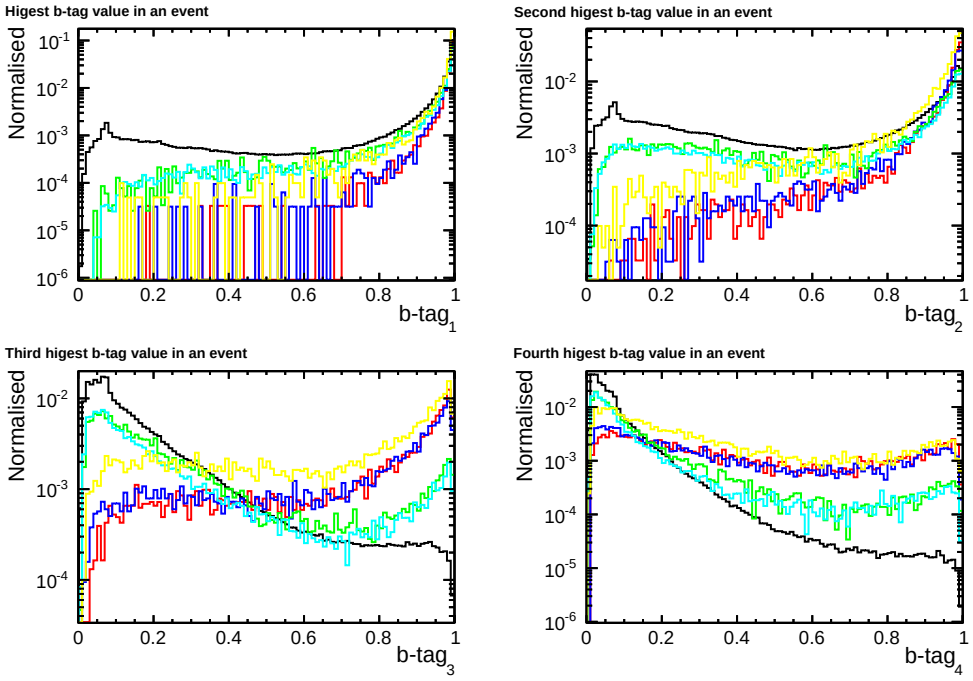

Figure 12: Normalised histograms for the response of the b-tagging from LCFIPlus for the highest ranked jet in an event (top left), second highest (top right), third highest (bottom left), and fourth highest (bottom right) for the semileptonic (red) and hadronic (blue) signal modes and all backgrounds.

Figure 12 shows the differences in the expected profiles of the third and fourth largest b-tags. For the background, there are small peaks at higher values of the b-tag due to the small contributions from genuine backgrounds with four b-jets, but the shape, especially $t\bar{t}$, is dominated by the peak at 0.1.

6.7 Event Reconstruction and Masses

After finding the optimal combination of jets from Equation 3 and reconstructing the event, cuts were made on the resulting masses of the top quarks, Higgs boson and total masses to ensure that the final event is consistent with $t\bar{t}H$.

Figure 13 demonstrates the final masses for all events which have exactly one isolated lepton. This requirement was essential for the reconstruction of the leptonically decaying top quark in the semileptonic mode. The statistics in the hadronic plots of Figure 13 are poor due to the excellent rejection of events which do not contain any isolated leptons. This is also reflected by the lack of a peak in the reconstructed mass of the leptonic W for the hadronic sample as there are no genuine isolated leptons for the reconstruction. A peak is observed in all of the backgrounds due to the finite fraction of expected leptonic top decays in the samples. There is a large fraction of events within the $t\bar{t}$ background which do not reconstruct a good W boson and in turn a good hadronic top due to the low multiplicity in the events containing multiple neutrinos.

Figure 14 shows the equivalent plots when there are no identified isolated leptons. In general, the reconstruction of the events for the hadronic mode is improved across all of the signal and background channels. This is a consequence of the events which have been selected by the isolation criteria. The events which pass this cut have hadronically decaying W bosons leading to higher particle multiplicities which removes the reconstruction issues of the semileptonic samples.

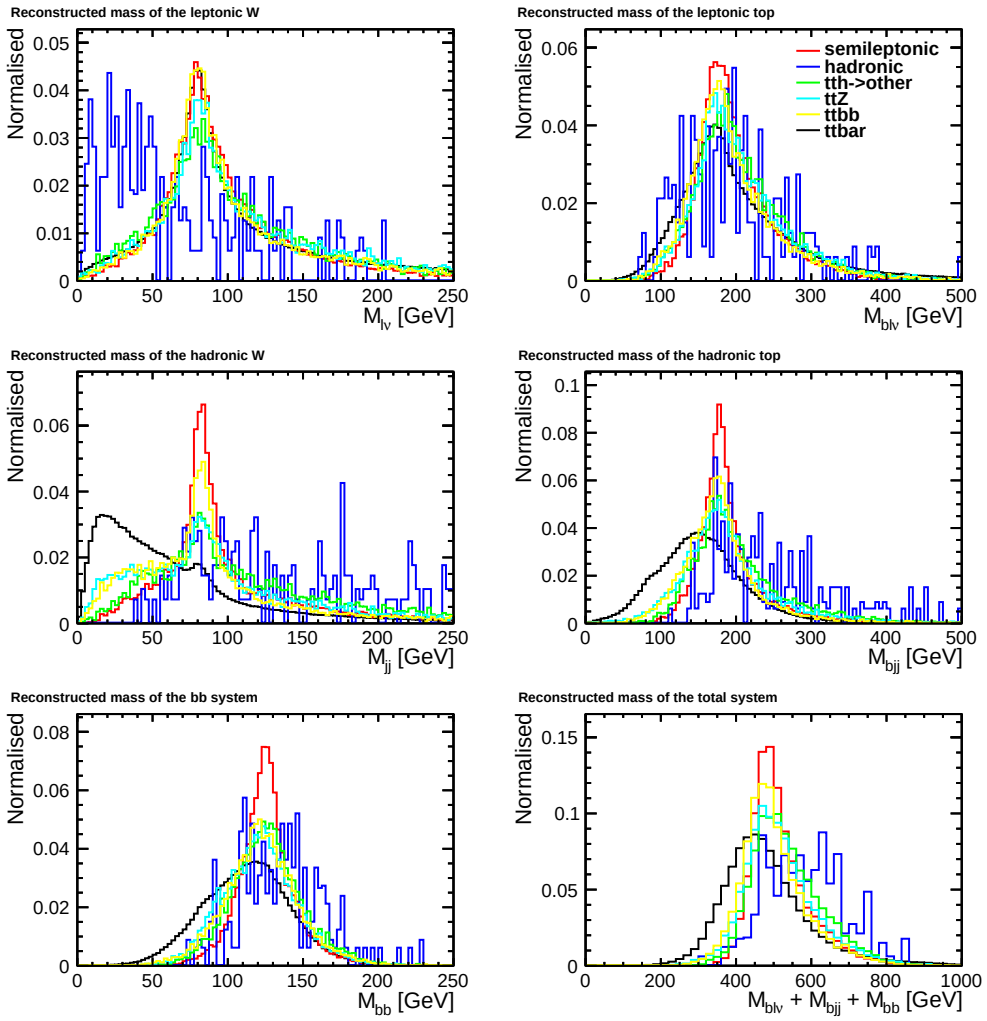

Figure 13: The reconstructed masses of the leptonic W boson (top left), leptonic top quark (top right), hadronic W boson (middle left), hadronic top quark (middle right), Higgs boson (bottom left) and total mass (bottom right) when there is exactly one identified isolated lepton in the event.

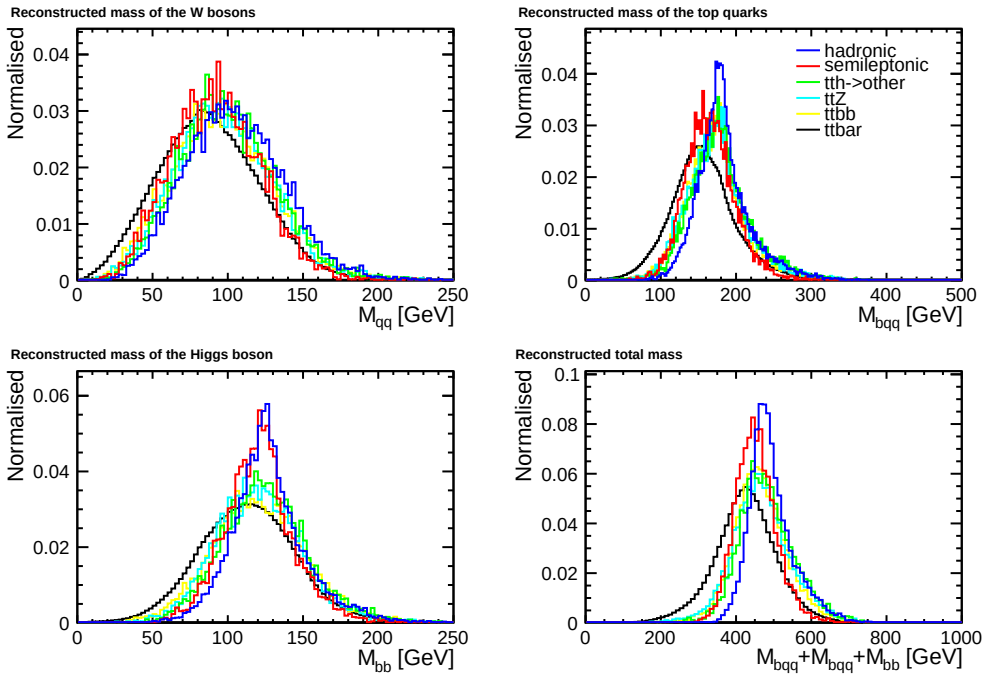

Figure 14: The reconstructed masses of the W boson (top left), top quark (top right), Higgs boson (bottom left) and total mass (bottom right) when there are no identified isolated leptons.

6.8 Helicity of Higgs Decay

The nature of the Higgs boson, especially its spin, is expected to be well-established by the time this analysis will commence. We make use of the fact that in the SM the Higgs boson is a scalar particle, whose decay products are distributed isotropically (and back-to-back) in the rest frame of the Higgs boson. We calculate the helicity angle of the Higgs decay as defined by the angle of the two b jets, in the rest frame of the Higgs boson, with respect to the direction of the Higgs momentum. This can be used to discriminate the $t\bar{t}Z$ process in which the Z boson is vector particle and results in non-uniform helicity angle distribution. From Figure 15 it is also found that this variable is also useful in discriminating $t\bar{t}b\bar{b}$ and $t\bar{t}$ backgrounds.

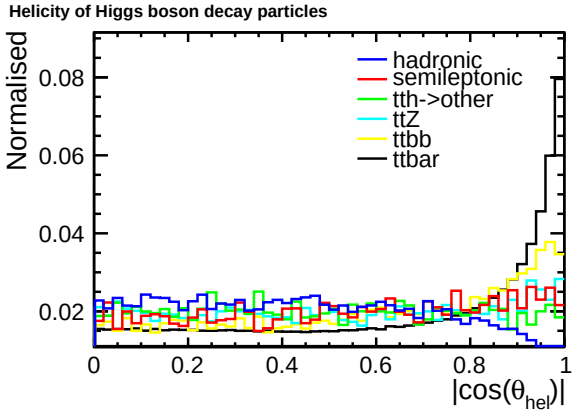

Figure 15: The helicity of the decay products associated with the Higgs boson for the semileptonic (red) and hadronic (blue) decay modes relative to the Higgs direction.

Cut	leptonic	semileptonic	hadronic	$t\bar{t}H \rightarrow$other	$t\bar{t}Z$	$t\bar{t}b\bar{b}$	$t\bar{t}$	$\frac{S}{\sqrt{S+B}}$
Total Events	151.4	628.7	652.7	1046.1	5332.4	1434.5	308800.9	1.11
N_{isolep}=1	74.6	363.5	5.0	371.8	1581.5	439.9	101295.2	1.13
610< E_{vis} <1000	49.6	338.5	4.7	312.7	1228.9	373.8	75507.1	1.21
nPFOs>154	15.0	235.0	4.1	195.0	589.0	194.5	12605.9	2.00
Thrust<0.88	12.5	205.6	3.7	168.9	492.6	140.0	6092.3	2.44
$\log_{10}(Y_{45})$ >-2.25	7.7	151.3	3.2	108.5	295.2	91.0	2067.2	2.90
$\log_{10}(Y_{56})$ >-3.35	6.9	145.1	3.2	106.2	277.6	86.0	1836.1	2.92
b-tag$_1$ >0.96	6.7	135.1	2.8	79.8	216.9	78.3	1367.8	3.11
b-tag$_2$ >0.91	6.1	118.2	2.3	41.2	135.1	66.9	715.2	3.59
b-tag$_3$ >0.67	5.5	102.1	1.6	5.7	59.4	56.0	137.2	5.33
b-tag$_4$ >0.04	5.5	100.5	1.5	5.4	58.3	54.8	128.7	5.34
χ^2 <450	5.3	100.0	1.4	5.2	56.8	53.7	126.0	5.36
364< M_{Total} <808	5.2	99.7	1.4	5.2	56.5	53.5	124.7	5.36
98< M_H <234	4.6	95.1	1.2	4.7	46.1	48.6	109.9	5.40

Table 2: The number of events passed each cut when the cut values are optimised to select the semileptonic signal with maximum significance. The $t\bar{t}H \rightarrow$other is the background where the Higgs boson does not decay to a $b\bar{b}$ pair.

7 Cut Based Analysis

The variables described in Section 6 were applied to the data sets to select the signal and reject the background events. The variables were optimised to maximise the significance $\mathscr{S} = S/\sqrt{S+B}$, of selecting the signal where S and B are the number of signal and background events passed the cut. Each cut was applied and optimised to the subset of events which passed all of the previous cuts.

Table 2 shows the optimised cut values for the semileptonic analysis and the number of events expected passed each cut for 1000 fb^{-1} split between the two polarisation states. The $t\bar{t}H$ events where the Higgs does not decay to a $b\bar{b}$ are incorporated into the tth$\rightarrow$other events. Whilst the cuts before the b-tags demonstrate excellent background reduction, the main discriminant in this analysis is the b-tag$_3$ cut as this yields the largest increases in the signal significance of the sample. The small increases observed when applying the mass cuts is a result of the excellent background suppression from the previous cuts leading to only events which are consistent with $t\bar{t}H$.

The expected number of events for the hadronic selection are shown in Table 3. As observed in the semileptonic channel the main discriminant is the b-tagging. The final hadronic sample has a much larger significance than the semileptonic channel due to the cut on the number of isolated leptons removing less than one percent of the events compared with 42% for the semileptonic study. The large number of semileptonic events in the final hadronic sample is believed to be a result of the lepton isolation criteria which rejects the events which contain taus.

Cut	leptonic	semileptonic	hadronic	tth$\rightarrow$other	$t\bar{t}Z$	$t\bar{t}b\bar{b}$	$t\bar{t}$	$\frac{S}{\sqrt{S+B}}$
Total Events	151.4	628.7	652.7	1046.1	5332.4	1434.5	308800.9	1.11
N_{isolep}=0	20.9	261.2	647.9	556.7	3226.1	932.5	188911.4	1.47
E_{vis}> 650	9.8	221.0	636.2	497.5	2743.5	849.3	157389.6	1.58
Thrust<0.87	8.1	187.8	577.6	440.1	2219.7	540.9	46916.1	2.56
$\log_{10}(Y_{78})$ >-4	3.7	143.6	549.5	415.5	1926.6	474.6	27472.1	3.12
b-tag$_4$ >0.38	1.9	81.0	275.0	17.6	230.0	209.6	680.6	7.11
$\cos(\theta_{hel})$ <0.9	1.6	73.8	263.7	16.5	215.9	189.2	584.9	7.19
M_t >120	1.5	68.9	255.4	15.6	207.8	178.5	530.93	7.20

Table 3: The number of events passed each cut when the cut values are optimised to select the hadronic signal with maximum significance. The $t\bar{t}H \rightarrow$other is the background where the Higgs boson does not decay to a $b\bar{b}$ pair.

Figure 16 demonstrates the power of the background suppression. Before any cuts have been applied the reconstructed Higgs mass peak in the signal is completely swamped by background events. Following the cuts the number of signal and background events remaining are comparable with each other and the Higgs mass peak is clearly visibly above the signal.

After the cuts have been applied the efficiency of signal selection, ε_{sig}, is 39.1% (15.1%) with a selected sample purity, $\rho^{\text{sel}}_{\text{sample}}$, of 20.3% (30.6%) for the hadronic (semileptonic) mode. This yields a signal significance of 7.20 (5.40) and a statistical uncertainty of 7.2% (9.6%) on the value of $g_{t\bar{t}H}$. The uncertainty is reduced in the hadronic channel

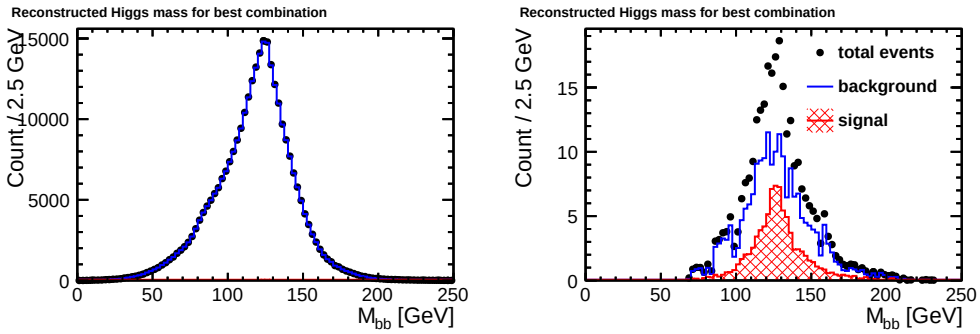

Figure 16: The reconstructed Higgs boson mass for the optimal combination of jets in the semileptonic decay mode for all events (left) and only the events which pass all of the cuts (right).

due to a higher relative cross section as the tau decays are rejected from the semileptonic channel. As the samples were split by the number of isolated leptons and minimal amounts of the signals remain as backgrounds in the final selections the final numbers can be combined to yield an overall significance of 9.01 and statistical uncertainty of 5.8% on the value of $g_{t\bar{t}H}$.

8 TMVA Analysis

A multivariate analysis using the TMVA toolkit [11] was implemented using boosted decision trees (BDT) to improve the signal and background separation. It was found that BDT with gradient boosting (BDTG) offers superior performance over BDT with adaptive boosting. The variables used as an input to the analysis are the same as those defined in Section 6 after the lepton isolation leading to input variables of the; visible energy; number of PandoraPFOs; thrust; jet clustering parameters; b-tags for the four highest jets; total mass of the final system; masses of the W bosons, top quarks, and Higgs boson, and the helicity of the decay products of the Higgs.

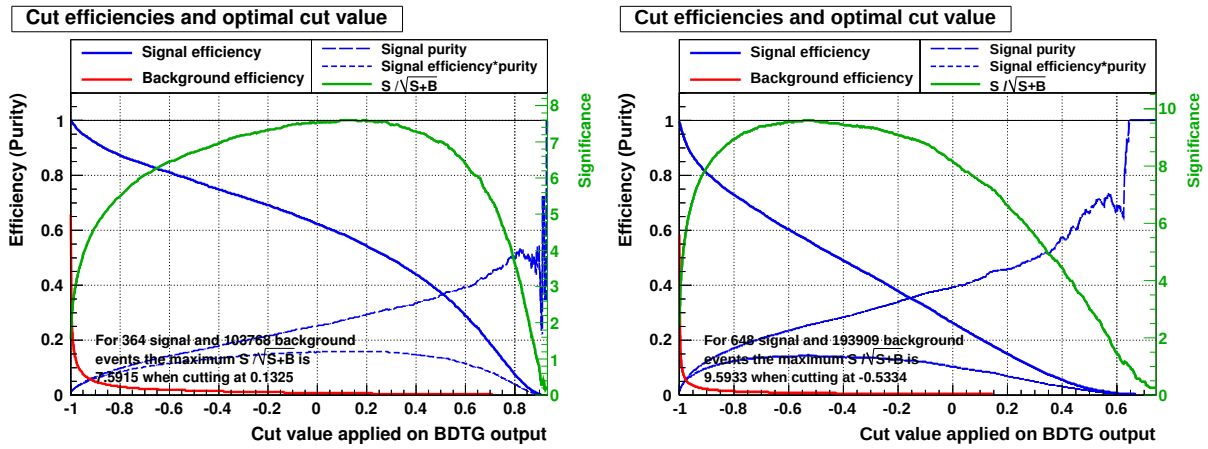

Figure 17: The response of the multivariate analysis for the semileptonic (left) and hadronic (right) decay modes.

Figure 17 shows the cut efficiencies, purities, and significance for the BDTG output value for both modes. The optimal cut values of -0.5334 (0.1325) yield a maximal significance of 9.59 (7.59) leading to a statistical uncertainty on $g_{t\bar{t}H}$ of 5.4% (6.9%) for the hadronic (semileptonic) modes. The performance for each mode is significantly improved with the use of a TMVA over the cut based method. As with the cut based method the performance in the hadronic channel is observed to be better than the semileptonic channel as expected. When the results are combined an uncertainty on $g_{t\bar{t}H}$ of 4.3% is obsevered. The efficiencies and purities of the final results are summarized in Tabs. 4-6.

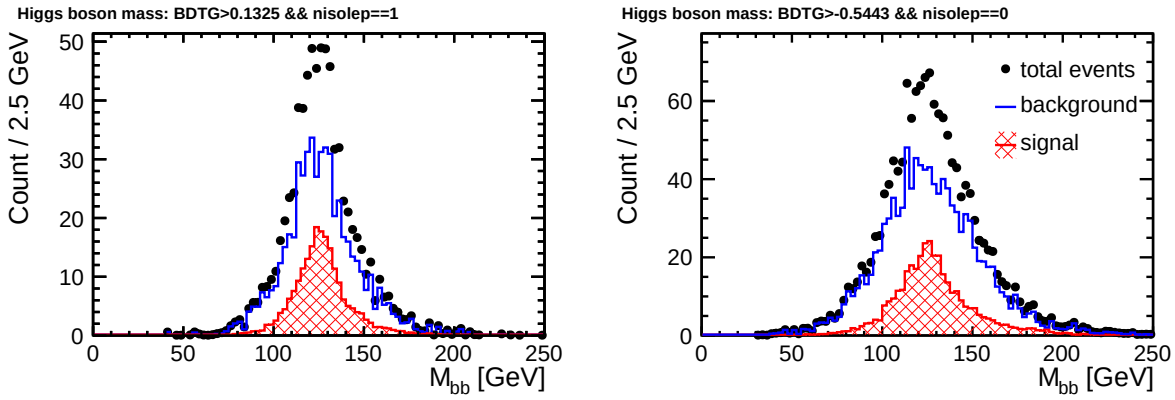

Figure 18: Reconstructed masses for the Higgs boson candidate after applying a cut on the multivariate classifier for the semileptonic (left) and hadronic (right) decay modes.

Cut	leptonic	semileptonic	hadronic	tth→other	$t\bar{t}Z$	$t\bar{t}b\bar{b}$	$t\bar{t}$	$\frac{S}{\sqrt{S+B}}$
Total Events	151.4	628.7	652.7	1046.1	5332.4	1434.5	308800.9	1.11
$\mathrm{BDTG}_{\mathrm{semil}} > 0.1325$	18.7	208.0	2.1	10.1	126.1	125.4	261.2	7.59

Table 4: The number of events passed each cut in the TMVA analysis for the semileptonic channel.

Cut	leptonic	semileptonic	hadronic	tth→other	$t\bar{t}Z$	$t\bar{t}b\bar{b}$	$t\bar{t}$	$\frac{S}{\sqrt{S+B}}$
Total Events	151.4	628.7	652.7	1046.1	5332.4	1434.5	308800.9	1.11
$\mathrm{BDTG}_{\mathrm{had}} > -0.5334$	0.3	65.5	365.6	25.0	260.5	222.6	513.6	9.59

Table 5: The number of events passed each cut in the TMVA analysis for the hadronic channel.

	Efficiency	Purity	Significance
Semileptonic (Cut)	15.1%	30.6%	5.40
Hadronic (Cut)	39.1%	20.3%	7.20
Semileptonic (TMVA)	33.3%	28.0%	7.59
Hadronic (TMVA)	56.0%	25.2%	9.59

Table 6: Summary of efficiencies, purities, and significances.

9 Consideration of systematic uncertainties

Given that the final sample used in the extraction of the top Yukawa coupling contains signal and background events in comparable proportion, it is expected that the estimation of the background will be a dominant source of systematic uncertainties. The total cross section is expected to be calculable from theory to very good precision for $t\bar{t}Z$ and $t\bar{t}$ processes. The $t\bar{t}b\bar{b}$ cross section may present an issue; in principle the measurement of the gluon splitting rate at relevant energies will provide a handle to estimate its size. The most crucial aspect is expected to be the estimation of the efficiencies. For this, one needs to know how well the event selection variables are modeled. Here we illustrate how one might arrive at control samples for different background sources in order to control the efficiency of each background component.

It is foreseen that the $t\bar{t}Z$ final state can be reconstructed in a similar fashion to the $t\bar{t}H$ final state. For hadronically decaying Z, the number of jets in the final state will be the same as in the $t\bar{t}H$ analysis. For our nominal integrated luminosities, 1400 events are expected for $t\bar{t}H(\rightarrow b\bar{b})$ whereas 800 events are expected for $t\bar{t}Z(\rightarrow b\bar{b})$, taking into account the $Z \rightarrow b\bar{b}$ branching ratio. The other hadronic decays of the Z will have large $t\bar{t}$ background due to the absence of the two b jets. The Z leptonic decays may help increase the sensitivity. Overall, one can naively expect that the statistical uncertainty for $t\bar{t}Z$ will be similar to that of $t\bar{t}H$, i.e. at the few percent level.

In the case of $t\bar{t}$, although its cross section is large, it should be checked on data whether the events passing the $t\bar{t}H$ event selection come from the core or the tail of the $t\bar{t}$ system. Once it is found that the core of $t\bar{t}$ is responsible for most of the $t\bar{t}$ background, the systematic uncertainty can be evaluated in the same way as in $t\bar{t}Z$.

For $t\bar{t}b\bar{b}$, the gluon splitting rate is expected to be measured for $t\bar{t}$ processes at various energies.

Other sources of systematic uncertainties such as the luminosity measurement, jet energy scale, and flavor tagging are typically at the 1% level or better for e^+e^- colliders.

10 Conclusions

The uncertainty on the measurement of the top Higgs Yukawa coupling has been studied at 1 TeV for the hadronic and semileptonic decay modes using a conventional cut-based approach and a multivariate analysis. The semileptonic analysis leads to an uncertainty of 6.9% (9.6%) for the multivariate (cut based) methods and the hadronic analysis yields improved values of 5.4% (7.2%). When the samples were combined an uncertainty on $g_{t\bar{t}H}$ of 4.3% is achieved via the multivariate method. This demonstrates the robustness of the physics reconstruction of high jet multiplicity final states at $\sqrt{s} = 1$ TeV under realistic simulation conditions.

11 Acknowledgments

The authors of the study would like to thank their colleagues in the Linear Collider community for various support and discussions at numerous occasions; in particular, T. Barklow, M. Berggren, and A. Miyamoto for the event generation; J. Engels for the mass production of samples; P. Roloff and J. Strube for the cross-checking of the results with the SiD analysis; R. Poeschl and A. Ishikawa for the review of the results in this study. This work was supported in part by the Grant-in-Aid for Specially Promoted Research No. 23000002 by the Japan Society for Promotion of Science.

References

[1] H. Baer *et al.*, "Physics at the International Linear Collider". `http://lcsim.org/papers/DBDPhysics.pdf`. to be published in the ILC Detailed Baseline Design Report.

[2] R. Yonamine, K. Ikematsu, T. Tanabe, K. Fujii, Y. Kiyo, *et al.*, "Measuring the top Yukawa coupling at the ILC at $\sqrt{s} = 500$ GeV" *Phys.Rev.* **D84** (2011) 014033, `arXiv:1104.5132 [hep-ph]`.

[3] W. Kilian, T. Ohl, and J. Reuter, "WHIZARD: Simulating Multi-Particle Processes at LHC and ILC" *Eur.Phys.J.* **C71** (2011) 1742, `arXiv:0708.4233 [hep-ph]`.

[4] "Physics Study Libraries". `http://www-jlc.kek.jp/subg/offl/physsim/`.

[5] F. Gaede, "Marlin and LCCD: Software tools for the ILC" *Nucl.Instrum.Meth.* **A559** (2006) 177–180.

[6] F. Gaede, T. Behnke, N. Graf, and T. Johnson, "LCIO: A Persistency framework for linear collider simulation studies" *eConf* **C0303241** (2003) TUKT001, `arXiv:physics/0306114 [physics]`. SLAC-PUB-9992, CHEP-2003-TUKT001.

[7] M. Thomson, "Particle Flow Calorimetry and the PandoraPFA Algorithm" *Nucl.Instrum.Meth.* **A611** (2009) 25–40, arXiv:0907.3577 [physics.ins-det]. CU-HEP-09-11.

[8] T. Tanabe *et al.*, "LCFIPlus". http://ilcagenda.linearcollider.org/getFile.py/access?contribId=16&sessionId=1&resId=0&materialId=slides&confId=5496. ILD Workshop, Kyushu, May, 2012.

[9] "DBD Samples". http://ilcsoft.desy.de/dbd/generated/.

[10] J. Rouene, "Measurement of $t\bar{t}$ asymmetries with the ILD at the ILC". http://ilcagenda.linearcollider.org/getFile.py/access?contribId=50&sessionId=10&resId=0&materialId=slides&confId=5468, 2012.

[11] A. Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, *et al.*, "TMVA - Toolkit for Multivariate Data Analysis" *PoS* **ACAT** (2007) 040, arXiv:physics/0703039 [PHYSICS].