CLICdp-Note-2014-001
30 March 2014

CERN - European Organization for Nuclear Research

Physics potential of the top Yukawa coupling measurement at a 1.4 TeV Compact Linear Collider using the CLIC_SiD detector

S. Redford*, P. Roloff*, M. Vogel†

* *CERN, Switzerland,* † *Pontificia Universidad Catolica de Chile, Chile*

Abstract

The physics potential of a direct measurement of the top Yukawa coupling is investigated, using the process $e^+e^- \to t\bar{t}H$ produced at a $\sqrt{s} = 1.4$ TeV Compact Linear Collider (CLIC). Final states with six and eight jets are reconstructed. This study addresses various aspects of the detector performance: jet clustering in complex hadronic final states, missing energy reconstruction, flavour-tagging and the identification of high-energy leptons. The analysis is based on full CLIC_SiD detector simulation using GEANT4. Beam-induced backgrounds from $\gamma\gamma \to$ hadrons interactions are overlaid on top of the physics events. The expected precision on the top Yukawa coupling is determined to be 4.27%, without beam polarisation.

Contents

1. Introduction

The Standard Model (SM) of particle physics predicts a linear dependence between the Higgs-fermion coupling strengths and the resulting fermion masses. The top quark is the heaviest known fundamental particle, therefore its coupling to the Higgs boson (the top Yukawa coupling g_{ttH}) represents the strongest of the Higgs couplings. Precise measurement of the top Yukawa coupling constitutes a test of the SM, and may provide sensitivity to physics beyond the SM.

The Compact Linear Collider (CLIC) [1] is a proposed future e^+e^- collider with possible staged centre-of-mass energies ($\sqrt{s}$) of 350 GeV, 1.4 TeV and 3 TeV. One of the main focuses of CLIC will be precision measurements of the properties of the Higgs boson. The top Yukawa coupling can be accessed through the process $e^+e^- \to t\bar{t}H$, which is sensitive to the strength of the coupling at the ttH vertex (see Figure 1).

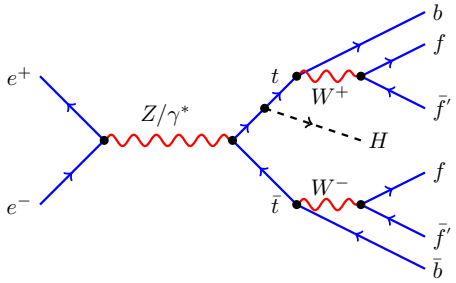


Figure 1: Feynman diagram of the process of interest, $e^+e^- \to t\bar{t}H$, in which a top quark radiates a Higgs boson.

In this note the statistical uncertainty of the top Yukawa coupling is measured at $\sqrt{s} = 1.4$ TeV assuming 1.5 ab^{-1} of integrated luminosity. Similar studies have been performed at $\sqrt{s} = 1$ TeV using the SiD [2] and ILD [3] detectors at the International Linear Collider (ILC). The signal and background cross-sections relevant for this analysis vary favourably as the centre-of-mass energy increases from 1 TeV to 1.4 TeV, however the levels of beam-induced background at CLIC are higher than at the ILC and must be controlled to perform precision measurements.

The $t\bar{t}H$ final state was not observed during Run I of the LHC [4, 5]. The expected precision on the top Yukawa coupling as measured at the $\sqrt{s} = 14$ TeV LHC is $14\% - 15\%$ with 300 fb^{-1}, and $7\% - 10\%$ with 3000 fb^{-1} of data [6].

2. Analysis framework

PHYSSIM [7] was used to generate all event samples, with the exception of the $t\bar{t}$ sample which was produced in WHIZARD 1.95 [8, 9]. The expected luminosity spectrum at CLIC was taken into account during the event generation. PYTHIA 6.422 [10] was used for hadronisation. Backgrounds from the beam-beam process $\gamma\gamma \to$ hadrons were simulated and overlaid on top of the physics processes. According to the expected bunch train structure and read-out electronics, backgrounds from 60 bunch-crossings were overlaid on top of each physics event. The full response of the CLIC_SiD detector [11, 12] was simulated for each event in GEANT4 [13, 14].

Events were reconstructed using the particle flow technique, implemented in Pandora PFA [15, 16]. The k_T algorithm including beam jets [17, 18], implemented in FastJet [19], was used in exclusive mode to cluster the jets of each event and to reject particles originating from beam-beam backgrounds. The LCFIPlus [20] package was used to re-cluster the signal jets using the Durham algorithm [21], and to flavour-tag these jets for charm and beauty probabilities. Simulation, reconstruction and analysis were done with ILCDIRAC [22].

3. The CLIC_SiD detector model

The CLIC_SiD detector is based on the SiD detector concept [23] for the ILC, and adapted for the beam conditions and higher centre-of-mass energy at CLIC. The detector design includes fine-grained calorimetry, to enable particle flow analysis techniques. Tungsten is used as the absorbing material throughout the ECAL and in the HCAL barrel, the HCAL endcap uses steel. Both calorimeters are within a 5 T solenoidal magnetic field. The tracking system is fully based on silicon pixel and strip technology. The vertex detector comprises ultra-thin hybrid pixel sensors, and uses power-pulsing and air-flow cooling to minimise the material budget. The outer tracking detector has five layers of thin silicon strip detectors. An instrumented iron yoke provides the magnetic flux return, and detection for penetrating particles.

4. Analysis strategy

The analysis strategy is presented in detail in the following sections, however a brief overview is given here. The $\mathrm{b\bar{b}}$ decay of the 126 GeV/c^2 Higgs boson is considered, corresponding to a branching ratio of 56%. Top quarks decay into a b jet and a W boson, where the W can subsequently decay either hadronically (68%) or leptonically (32%). There are therefore three possible final states of the top pair containing 0, 1 or 2 leptons depending on the decay of the two W bosons.

In this analysis two final states of the top pair are studied, the 'all jets' final state with a branching ratio of 46% and the 'lepton+jets' final state with a branching ratio of 45%. The 'di-lepton' $\mathrm{t\bar{t}}$ final state is not considered, as it has a branching fraction of only 9%, and much larger backgrounds. This results in two channels for the $\mathrm{t\bar{t}H}$ decay,

- the fully-hadronic channel, giving a $\mathrm{t\bar{t}H}$ final state of eight jets,
- the semi-leptonic channel, giving a $\mathrm{t\bar{t}H}$ final state of six jets, one lepton and one neutrino,

where the two channels are distinguished by the number of leptons present.

Leptons originating from W decays are identified in each event (described in Section 6). If zero leptons are found, the event is classified as fully-hadronic. If one lepton is found, the event is classified as semi-leptonic. Events in which more than one lepton are found are not analysed further.

All particles in an event which were not identified as leptons are clustered by the k_T algorithm into a specific number of jets, plus two beam jets (see Section 7). Events classified as fully-hadronic are clustered into eight jets, and those classified as semi-leptonic into six jets.

The particles clustered into the beam jets are assumed to have originated from beam-beam backgrounds, and are removed from the event.

The particles included in the six or eight jets are then re-clustered using the Durham algorithm, and flavour-tagging of each jet is performed (see Section 8). Jets are combined together using kinematic information to give $W^{\pm}$, top and Higgs candidates. A selection is used to separate signal from background, allowing the uncertainty on the $t\bar{t}H$ cross-section to be measured (see Section 9).

To translate the measured uncertainty on the cross-section into a measurement of the uncertainty on the top Yukawa coupling, a correction is applied to take into account the Higgsstrahlung diagram. This diagram contributes to $e^+e^- \to t\bar{t}H$ but is not sensitive to the top Yukawa coupling (see Figure 2). The correction factor is calculated by considering the cross-section for the process $e^+e^- \to t\bar{t}H$ as a function of the value of the top Yukawa coupling at $\sqrt{s} = 1.4$ TeV (see Figure 3). Beamstrahlung is included in the simulation, which reduces the effective centre-of-mass energy of the collisions. The correction factor is equal to the slope of the cross-section at the SM value of the top Yukawa coupling.

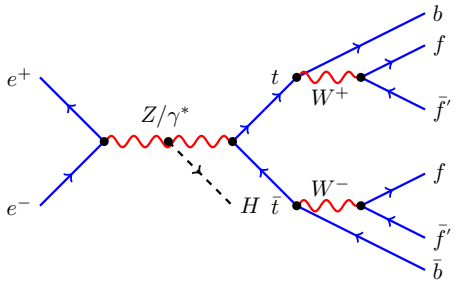


Figure 2: Feynman diagram of the Higgsstrahlung process, in which the $t\bar{t}H$ final state is achieved by the Z/γ^* radiating a Higgs boson.

5. Simulation samples

Table 1 gives details of the simulated samples used in this analysis. The fully-hadronic and semi-leptonic channels of interest are simulated, in addition to $t\bar{t}H$ decays to other final states and a number of background channels. Other eight-fermion final states included as background are $t\bar{t}Z$ and $t\bar{t}b\bar{b}$ decays, in which a top quark radiates a Z or $g \to b\bar{b}$. These samples are simulated separately for each final state of the top pair. The six-fermion final state channel of $t\bar{t}$ is simulated for all possible decay modes. This process is a significant background due to its comparatively large cross-section. All samples are simulated with unpolarised beams. The weight of each simulated sample is much smaller than 1, with the exception of the weight for the high cross-section $t\bar{t}$ sample which is of order 1.

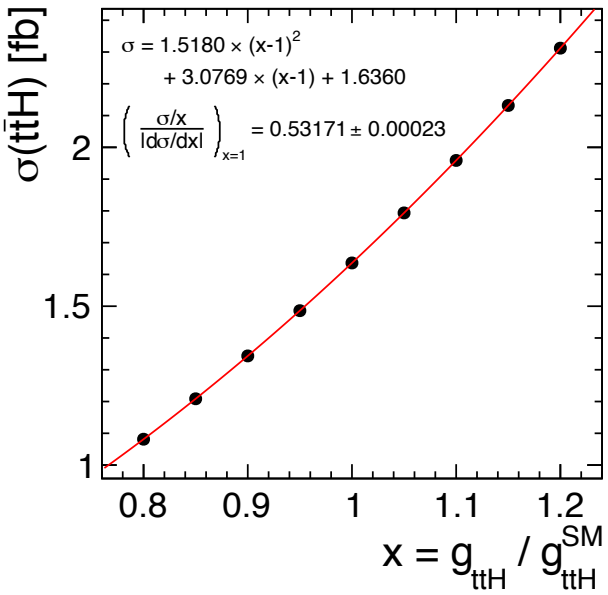


Figure 3: Cross-section for the process $e^+e^- \to t\bar{t}H$ as a function of the value of the top Yukawa coupling (g_{ttH}), calculated at $\sqrt{s} = 1.4$ TeV and including the effect of beamstrahlung.

Table 1: Simulated samples used in the analysis. The first two rows are the signal $t\bar{t}H$ channels, followed by the $t\bar{t}H$ decays to other final states and the background channels. Column 1 shows the simulated process. Column 2 shows the assumed cross-section. Column 3 shows the expected number of events in 1.5 ab^{-1}. Column 4 shows the sample weight. The number of jets refers to the $t\bar{t}$ decay only.

Process	Cross-section (fb)	Evt in 1.5 ab^{-1}	Sample weight
$t\bar{t}H$, 6 jet, $H \to b\bar{b}$	0.431	647	0.03
$t\bar{t}H$, 4 jet, $H \to b\bar{b}$	0.415	623	0.03
$t\bar{t}H$, 2 jet, $H \to b\bar{b}$	0.100	150	0.006
$t\bar{t}H$, 6 jet, $H \not\to b\bar{b}$	0.315	473	0.02
$t\bar{t}H$, 4 jet, $H \not\to b\bar{b}$	0.303	455	0.02
$t\bar{t}H$, 2 jet, $H \not\to b\bar{b}$	0.073	110	0.004
$t\bar{t}b\bar{b}$, 6 jet	0.549	824	0.03
$t\bar{t}b\bar{b}$, 4 jet	0.529	794	0.03
$t\bar{t}b\bar{b}$, 2 jet	0.127	191	0.008
$t\bar{t}Z$, 6 jet	1.895	2,843	0.1
$t\bar{t}Z$, 4 jet	1.825	2,738	0.1
$t\bar{t}Z$, 2 jet	0.439	659	0.03
$t\bar{t}$	135.8	203,700	1.5

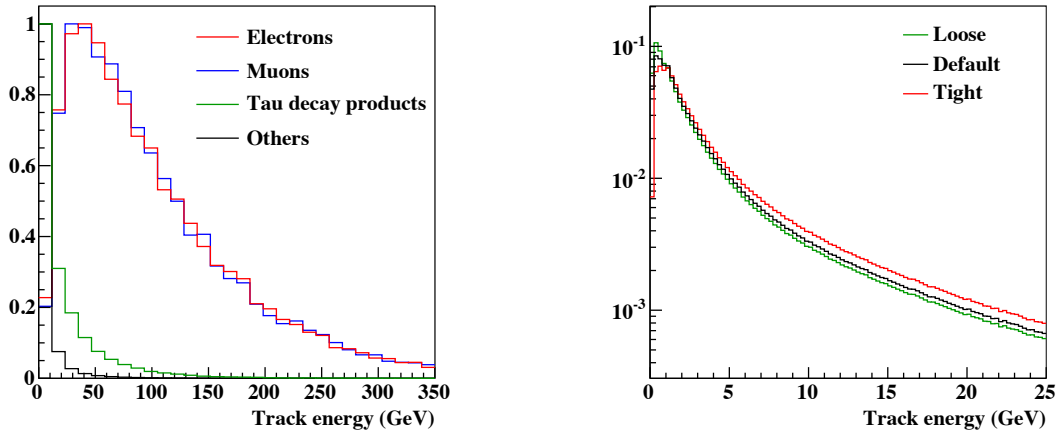
6. Lepton identification

The first step of the analysis is to identify and reconstruct leptons from $W^{\pm}$ decays. This section details how leptons are distinguished from all other reconstructed particles in the event. As expected due to their rapid decay, tau leptons demonstrate different behaviour to electrons and muons. Two searches are therefore implemented. Isolated leptons (electrons and muons) are searched for first, using a combination of track energy, calorimeter energy, impact parameter and isolation information. Then tau leptons are searched for, by considering highly energetic, low-multiplicity jets which must be isolated from other activity in the detector.

Simulator-level information can be used to match generated particles to reconstructed particles and vice-versa, a process known as 'truth-matching'. The properties of reconstructed leptons from $W^{\pm}$ decays are determined in this way (see Section 6.1). This allows the two search algorithms to be tuned, enabling them to distinguish between reconstructed leptons and other reconstructed particles (see Sections 6.2 and 6.3).

6.1. Properties of leptons from $W^{\pm}$ decays

The electrons and muons originating from $W^{\pm}$ decays have high reconstructed track energies, in the range of 50 GeV $-$ 350 GeV (see Figure 4). This is much higher than the energy of typical particles in jets. The decay products of tau leptons have much less energy than electrons and muons, however are still significantly more energetic than particles originating from showers within jets.



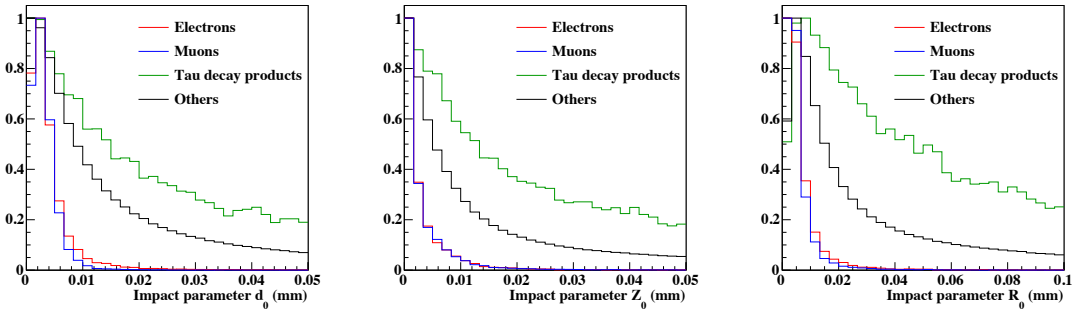
(a) (b)

Figure 4: Events from the simulated semi-leptonic $t\bar{t}H$ sample. (a): The reconstructed track energy of all reconstructed truth-matched electrons (red), muons (blue) and tau decay products (green) from $W^{\pm}$ decays. 'Others' denotes reconstructed tracks not truth-matched to a simulated signal lepton. The maximum of all histograms is set to one. (b): The track energy of all reconstructed particles not truth-matched to a lepton. 'Loose', 'default' and 'tight' refers to the level of beam-induced background suppression. Histograms are scaled to have unit area.

Due to the short lifetime of top quarks and $W^{\pm}$ bosons, the electrons and muons from $W^{\pm}$ decays typically originate from the primary vertex (PV) of the event. Conversely, b quarks and tau leptons have a longer lifetime and their decay products may originate from a displaced vertex. The impact parameter (IP) of a track describes the perpendicular distance between the track and PV, at the track's point of closest approach to the PV. It can be decomposed into longitudinal (Z_0) and radial (d_0) components, which combine to give the 3D IP:

$$R_0 = \sqrt{Z_0^2 + d_0^2}\,. \tag{1}$$

Figure 5 shows the three IP variables for reconstructed truth-matched leptons from $W^{\pm}$ decays, and for all other reconstructed particles. As expected, reconstructed electrons and muons have significantly smaller IP than other reconstructed particles in the event.



(a) (b) (c)

Figure 5: Impact parameter of truth-matched reconstructed electrons (red), muons (blue) and tau decay products (green) from $W^{\pm}$ decays, and all other reconstructed particles in the event not truth-matched to a lepton (black). (a): The radial IP component d_0. (b): The longitudinal IP component Z_0. (c): The combined 3D IP R_0 (see Equation 1).

Electrons, muons and hadrons have different interaction cross-sections in the materials which make up the detector. Therefore, information from the calorimeter system can be used to infer the identity of particles. The ratio of the energy deposited in the ECAL with the total energy deposited in both ECAL and HCAL

$$R_{\mathrm{CAL}} = \frac{E_{\mathrm{ECAL}}}{E_{\mathrm{ECAL}} + E_{\mathrm{HCAL}}}\,, \tag{2}$$

is shown in Figure 6 for reconstructed, truth-matched leptons from $W^{\pm}$ decays and all other reconstructed particles not truth-matched to a lepton. As electrons are contained within the ECAL, they form a peak at $R_{\mathrm{CAL}} = 1$. Muons deposit a minimum amount of ionisation energy throughout the calorimeters, and form a peak at $R_{\mathrm{CAL}} \approx 0.2$. Tau decay products can be leptonic or hadronic, and therefore can not be identified in this way.

A final method for identifying leptons from $W^{\pm}$ decays is to consider their isolation within the event. As most particles originate from showers within jets, they are reconstructed in 'busy'

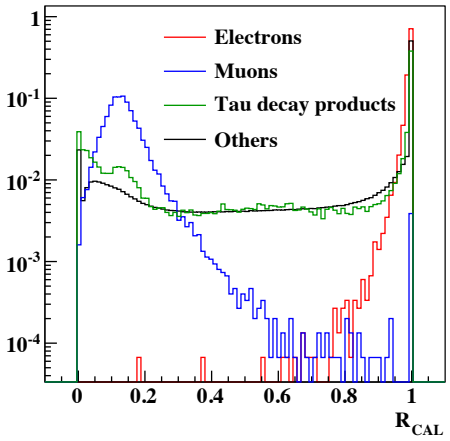
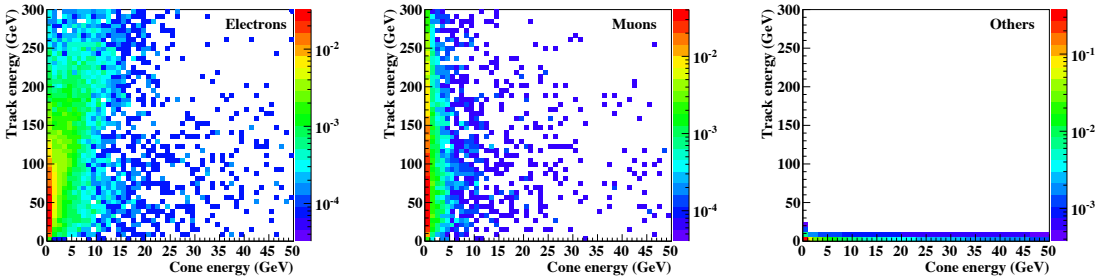


Figure 6: Calorimeter energy ratio (see Equation 2) of reconstructed, truth-matched electrons (red), muons (blue), tau decay products (green) and all other reconstructed particles not truth-matched to a lepton (black).

regions of the detector with high occupancy. The electrons and muons from $W^{\pm}$ decays are expected to be more isolated. Tau decay products, while not isolated from each other, should also exist within a lower occupancy region of the detector.

Figure 7 shows the energy within a cone of size $\cos(\theta) = 0.995$ around the particle's track as a function of track energy. Electrons are seen to radiate more than muons, as expected, although both flavours of leptons are typically highly energetic and highly isolated (low cone energy), in contrast to other particles. Considering cone energy as a function of track energy (rather than cone energy independently) allows the identification of highly energetic leptons which radiate photons at small angles.



(a) (b) (c)

Figure 7: Cone energy as a function of track energy. (a): Reconstructed, truth-matched electrons. (b): Reconstructed, truth-matched muons. (c): Other reconstructed particles not truth-matched to a lepton.

6.2. Selection of isolated leptons

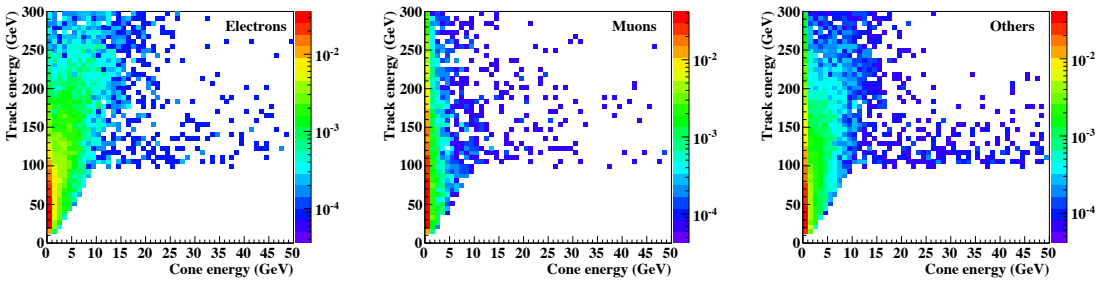
Using information from the above section, a selection based on track energy, calorimeter energy, impact parameter and isolation was devised to distinguish electrons and muons from $W^{\pm}$ decays from other particles in the event. The IsolatedLeptonFinder Marlin processor [24] was used.

To be selected as an electron or muon from a $W^{\pm}$ decay, a particle's track energy must be greater than 15 GeV. This selection retains 96.5% of reconstructed, truth-matched electrons and muons from $W^{\pm}$ decays and 9.6% of all other reconstructed particles not truth-matched to a lepton. A selection on IP is applied, requiring that d_0, Z_0 and R_0 are all less than 0.05 mm. The calorimeter energy ratio R_{CAL} is required to be greater than 0.9 or within the range $0.05 - 0.3$, to remove particles which do not behave as electrons or muons in the calorimeters. A polynomial isolation selection, dependent on the reconstructed particle's energy, is applied to the particle's cone energy (see Figure 8). The requirement is:

$$\text{Particle energy (GeV)} > \begin{cases} 100, & \text{if cone energy} > 10 \text{ GeV}. \\ 10 \times \text{cone energy (GeV)}, & \text{if cone energy} \leq 10 \text{ GeV}. \end{cases} \quad (3)$$

This removes non-leptons whilst retaining high-energy leptons which radiate a photon. Figure 8c appears significantly different from Figure 7c because the selection has removed the majority of particles. The remaining set of reconstructed particles contains a high fraction of leptons (within jets for instance), but these particles are not truth-matched to a lepton from a $W^{\pm}$ decay.

In combination these requirements retain 87.3% of truth-matched electrons and muons, and 0.4% of other reconstructed particles not truth-matched to a lepton from a $W^{\pm}$ decay.



(a) (b) (c)

Figure 8: Cone energy as a function of track energy, after application of the polynomial selection used to identify isolated leptons. (a): Reconstructed, truth-matched electrons. (b): Reconstructed, truth-matched muons. (c): Other reconstructed particles not truth-matched to a lepton.

6.3. Selection of tau leptons

The TauFinder Marlin processor [25] was adapted in the following ways to identify the products from tau decays, and to combine them to form reconstructed tau leptons. A 'seed' track is required to have p_{T} greater than 10 GeV/c. All particles within a cone of angle 0.04 rad are added

to the seed to form the tau candidate, providing each particle has p_T greater than 2 GeV/c and R_0 within the range 0.01 mm – 0.5 mm. These numbers were determined from truth-matched simulation to best describe tau lepton decays in this analysis. The reconstructed tau candidate must be composed of an odd number of charged tracks, and must have invariant mass less than 1.5 GeV/c^2. This is slightly lower than the tau invariant mass, however the undetected neutrino means that taus are rarely reconstructed with an invariant mass higher than this. An isolation ring is defined from 0.04 rad – 0.25 rad around the tau seed. There must be less than five particles in the isolation ring and their energy must sum to less than 5 GeV.

6.4. Performance of the lepton finder

The isolated lepton finder and tau finder process every event, using the criteria above to search for leptons from $W^{\pm}$ decays. In combination, they may find zero, one, or more leptons per event. The rate of lepton finding was studied for all event samples (see Table 2). It was found that 92% of hadronic $t\bar{t}H$ decays were correctly classified as having zero leptons, and 68% of semi-leptonic $t\bar{t}H$ events were found to have one lepton. This figure includes losses due to non-isotropic detector coverage and detection efficiency. In the semi-leptonic channel, when an isolated lepton is found it is the lepton from the $W^{\pm}$ decay 87% of the time. When a tau lepton is found, it is the tau from the $W^{\pm}$ decay 85% of the time. A figure showing the number of leptons found per event for each sample is presented in Figure 17.

Table 2: Results of the lepton finding per simulated event sample. Column 1 shows the simulated process. Column 2 shows the expected number of events in 1.5 ab^{-1}. Column 3 shows the number of events found to have 0 leptons, and this number as a percentage of the total number of expected events. Column 4 shows the number of events found to have 1 lepton, and this number as a percentage of the total number of expected events. The sum of columns 3 and 4 is not necessarily equal to the total number of events expected in 1.5 ab^{-1} as some events are found to have more than 1 lepton. The number of jets refers to the $t\bar{t}$ decay only.

Process	Evt in 1.5 ab^{-1}	Evt with 0 leptons		Evt with 1 lepton	
$t\bar{t}H$, 6 jet, $H \to b\bar{b}$	647	593	(91.8%)	49	(7.57%)
$t\bar{t}H$, 4 jet, $H \to b\bar{b}$	623	178	(28.5%)	420	(67.5%)
$t\bar{t}H$, 2 jet, $H \to b\bar{b}$	150	13	(8.42%)	61	(40.5%)
$t\bar{t}H$, 6 jet, $H \not\to b\bar{b}$	473	306	(64.8%)	127	(26.8%)
$t\bar{t}H$, 4 jet, $H \not\to b\bar{b}$	455	89	(19.5%)	246	(54.1%)
$t\bar{t}H$, 2 jet, $H \not\to b\bar{b}$	110	6	(5.92%)	33	(30.1%)
$t\bar{t}b\bar{b}$, 6 jet	824	737	(89.5%)	80	(9.68%)
$t\bar{t}b\bar{b}$, 4 jet	794	222	(28.0%)	533	(67.1%)
$t\bar{t}b\bar{b}$, 2 jet	191	16	(8.37%)	78	(41.1%)
$t\bar{t}Z$, 6 jet	2,843	2,335	(82.1%)	322	(11.3%)
$t\bar{t}Z$, 4 jet	2,738	711	(26.0%)	1,678	(61.3%)
$t\bar{t}Z$, 2 jet	659	54	(8.27%)	248	(37.7%)
$t\bar{t}$	203,700	111,020	(54.5%)	77,110	(37.9%)

7. Reconstruction of $\mathrm{W}^{\pm}$, top and Higgs candidates

The particles identified as leptons are removed from the event, and the remaining particles are clustered into jets. This is achieved using the FastJet implementation of the k_{T} algorithm, where the level of background suppression and the jet clustering radius are to be determined (see Section 7.1). Depending on the number of leptons found, the events are forced into six or eight jets and two additional beam-jets, which take into account the forward particles from beam-induced background.

Jets are combined to form $\mathrm{W}^{\pm}$, top and Higgs candidates by choosing the grouping which minimises the χ^2 function specific to the number of jets:

$$\begin{aligned}
\chi^2_6 &= \frac{(M_{12}-M_{\mathrm{W}^{\pm}})^2}{\sigma^2_{\mathrm{W}^{\pm}}} + \frac{(M_{123}-M_{\mathrm{t}})^2}{\sigma^2_{\mathrm{t}}} + \frac{(M_{45}-M_{\mathrm{H}})^2}{\sigma^2_{\mathrm{H}}}, \\
\chi^2_8 &= \frac{(M_{12}-M_{\mathrm{W}^{\pm}})^2}{\sigma^2_{\mathrm{W}^{\pm}}} + \frac{(M_{123}-M_{\mathrm{t}})^2}{\sigma^2_{\mathrm{t}}} + \frac{(M_{45}-M_{\mathrm{W}^{\pm}})^2}{\sigma^2_{\mathrm{W}^{\pm}}} + \frac{(M_{456}-M_{\mathrm{t}})^2}{\sigma^2_{\mathrm{t}}} + \frac{(M_{78}-M_{\mathrm{H}})^2}{\sigma^2_{\mathrm{H}}}
\end{aligned} \tag{4}$$

where the indices of M represent each of the jets.

The reconstructed masses (M) and resolutions (σ) of the $\mathrm{W}^{\pm}$, top and Higgs in the signal channels are determined by truth-matching reconstructed particles in the jets to the generated fermions. Example invariant mass spectra are shown in Figure 9, for a jet radius of 1.0 and default timing and momentum cuts. It can be seen that the invariant mass distributions are not symmetric. Therefore, RooFit [26] was used to fit each distribution with a modified Gaussian function:

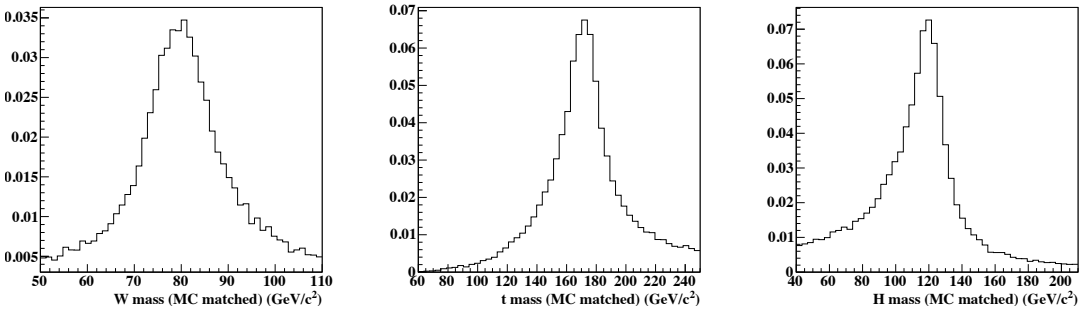
$$f = \exp\left(\frac{-(x-\mu)^2}{g}\right) \begin{cases} g = 2\sigma_{\mathrm{L}}^2 + \alpha_{\mathrm{L}}(x-\mu)^2, & x < \mu. \\ g = 2\sigma_{\mathrm{R}}^2 + \alpha_{\mathrm{R}}(x-\mu)^2, & x > \mu. \end{cases} \tag{5}$$

with mean μ, two widths σ_{L} and σ_{R}, and two parameters which modify the tails α_{L} and α_{R}. This allows the tails to be adjusted in a non-symmetric manner. The fitted distributions for this particular example can be seen in Figure 10.

7.1. Choice of background suppression level and jet clustering radius

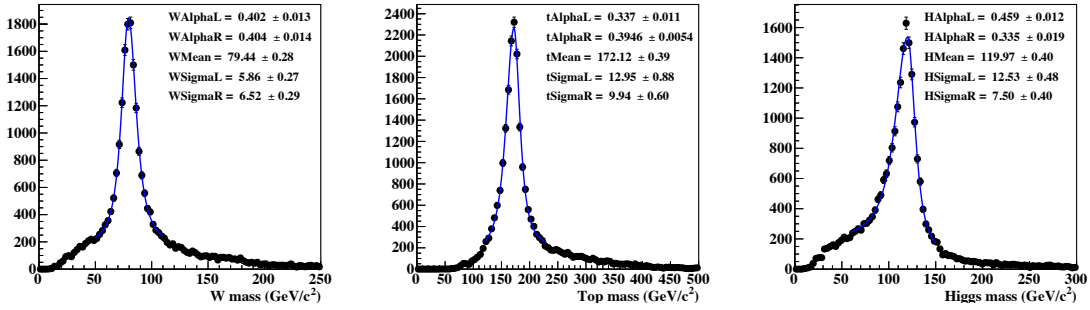
The level of background suppression and the jet radius used in the clustering algorithm were both optimised to give the best resolution possible for this analysis.

The presence of the $\gamma\gamma \to$ hadrons background is of significant consequence. At $\sqrt{s} = 1.4$ TeV, 1.3 $\gamma\gamma \to$ hadrons interactions occur per bunch-crossing, producing particles with a combined energy of over 200 GeV [27]. The majority of these particles have very small angle to the beam-axis, however approximately 20 GeV is deposited in the calorimeters per bunch-crossing due to this process. To reduce the effect of this background timing and momentum cuts are imposed on all reconstructed particles. The importance of these cuts for this analysis is demonstrated in Figure 11. Without the background suppression, the reconstructed $\mathrm{W}^{\pm}$ invariant mass distribution is smeared and shifted to higher mass. Implementing the background suppression allows the original $\mathrm{W}^{\pm}$ invariant mass distribution to be recovered almost exactly. Three levels of background suppression are available: 'loose', 'default' and 'tight'.



(a) (b) (c)

Figure 9: Normalised invariant mass distributions for the reconstructed, truth-matched candidates, using jet clustering radius 1.0 and default timing and momentum cuts. (a): $W^{\pm}$. (b): Top. (c): Higgs.



(a) (b) (c)

Figure 10: As Figure 9. Each distribution is fitted with a modified Gaussian function (see Equation 5).

The jet clustering radius determines which particles get combined to form a reconstructed jet. Too small a radius means that particles originating from the jet are lost, and the reconstructed mass is too low. Conversely, too large a radius means clustering particles which do not belong to the jet, causing the reconstructed mass to be too high. In both cases the resolution degrades. Jet radii from $0.5 - 1.5$ are considered.

The jet clustering was performed on the semi-leptonic signal sample, in combination with the application of background suppression cuts. There are thirty three unique combinations of background suppression (3) and jet radius (11). For each combination, the six jets were grouped into $W^{\pm}$, top and Higgs candidates by truth-matching the particles in each jet. The masses (M) and resolutions (σ) were determined by fitting each invariant mass distribution with a modified Gaussian (as in Figure 10). The fitted parameters for all combinations of background suppression and jet radius are presented in Figure 12. From this, it was decided to use jet radius 1.0 and the default level of background suppression, in order to minimise the low-mass width of the reconstructed Higgs (see Figure 12f). The corresponding mass and resolution parameters nec-

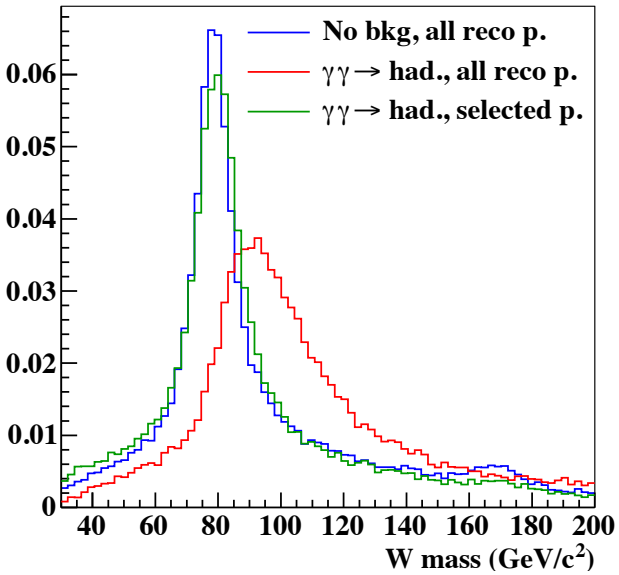


Figure 11: $W^{\pm}$ invariant mass distribution for the semi-leptonic $t\bar{t}H$ signal sample with no $\gamma\gamma \rightarrow$ hadrons background (blue), background and no timing and momentum cuts (red) and background and 'default' timing and momentum cuts (green). The small bump at invariant mass of 170 GeV/c^2 occurs when all three jets from a top decay are clustered into two jets.

essary for Equation 4 are shown in Table 3. The fitted mass parameters are below the generator values as some energy is lost via undetected neutrinos in the jets. For each possible permutation of jets within an event, the χ^2 function is evaluated using either σ_L, if the corresponding mass measurement is lower than μ, or σ_R if the mass measurement is higher than μ. Given a difference between the measured mass and the expected mass μ, this places a different penalty according to whether the measured mass is too high or too low.

Table 3: Invariant mass distribution parameters fitted from the optimal combination of default background suppression and a jet clustering radius of 1.0.

	Mass (GeV/c^2)	σ_L (GeV/c^2)	σ_R (GeV/c^2)
$W^{\pm}$	79.4	6.0	6.5
Top	172.2	13.0	10.0
Higgs	119.8	12.5	7.5

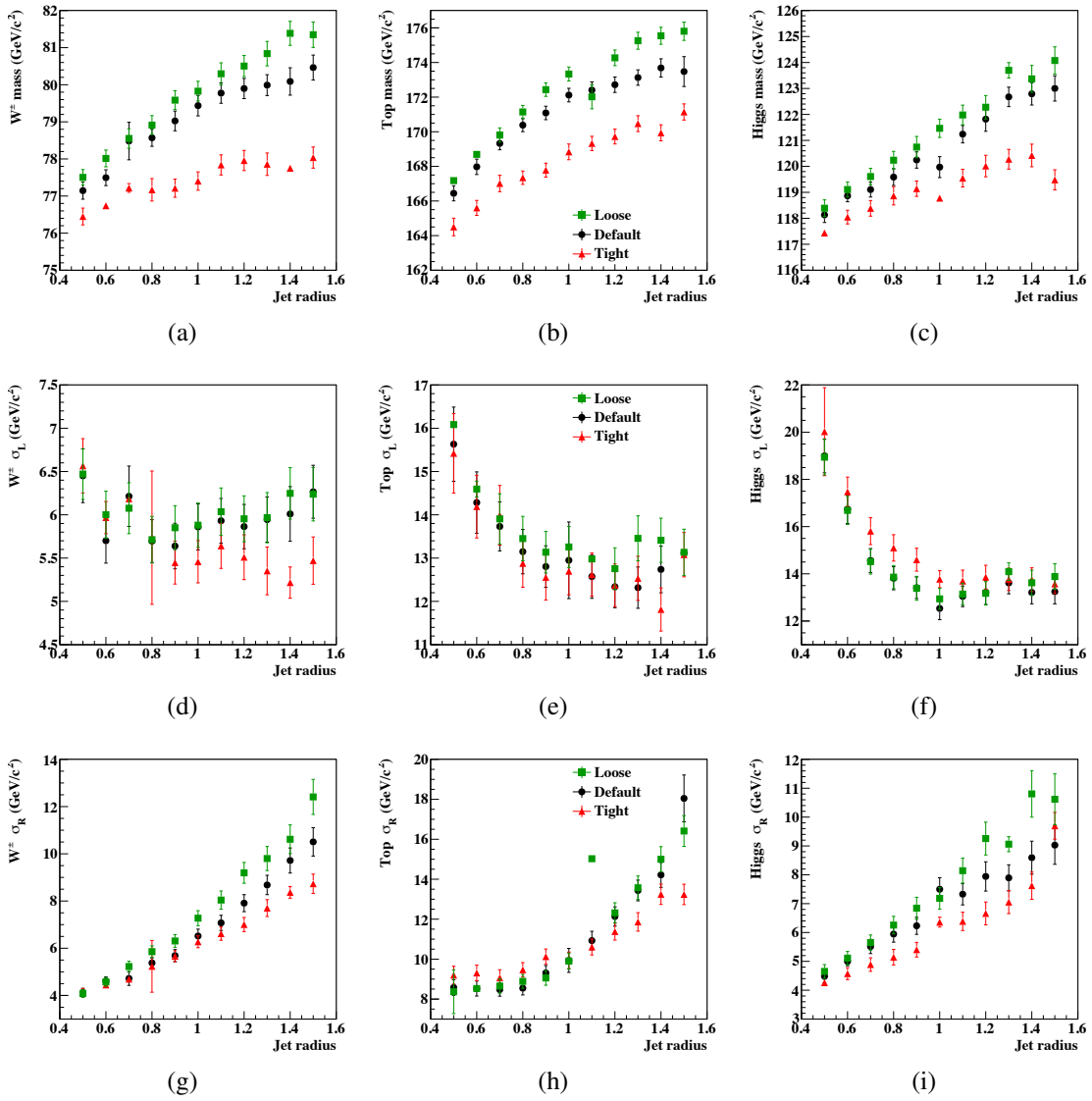
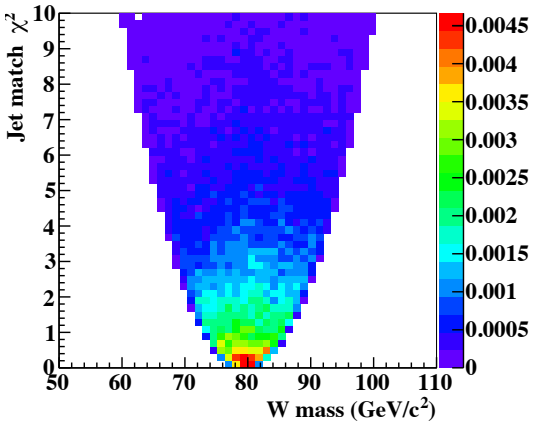


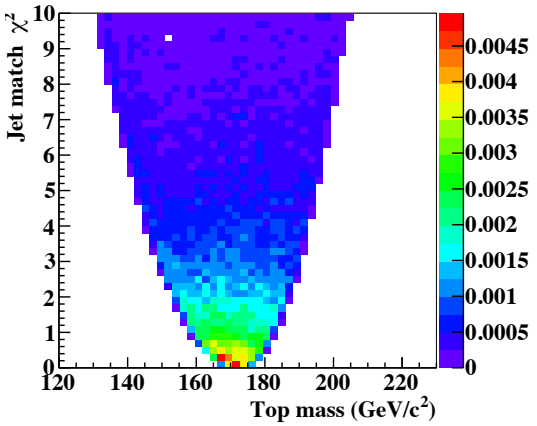
Figure 12: Fitted masses and widths of the reconstructed, truth-matched $W^{\pm}$, top and Higgs candidate invariant mass distributions as a function of jet clustering radius and background suppression level. The errors bars show the uncertainties on the fitted values. The semi-leptonic signal $t\bar{t}H$ sample is used. (a) – (c): The fitted masses (μ). (d) – (f): The fitted low-mass widths (σ_L). (g) – (i): The fitted high-mass widths (σ_R).

7.2. Result of χ^2 matching

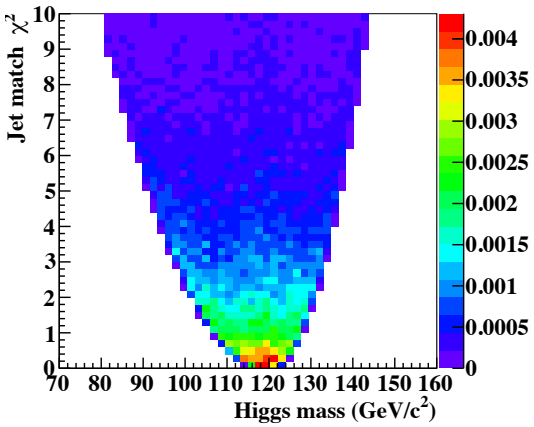
The result of performing the χ^2 matching of Equation 4 is shown in Figure 13, for the semi-leptonic signal sample. The majority of events form $W^{\pm}$, top and Higgs candidates close to the expected masses. Events in the peaks of the invariant mass distributions typically have the smallest χ^2 values.



(a)



(b)



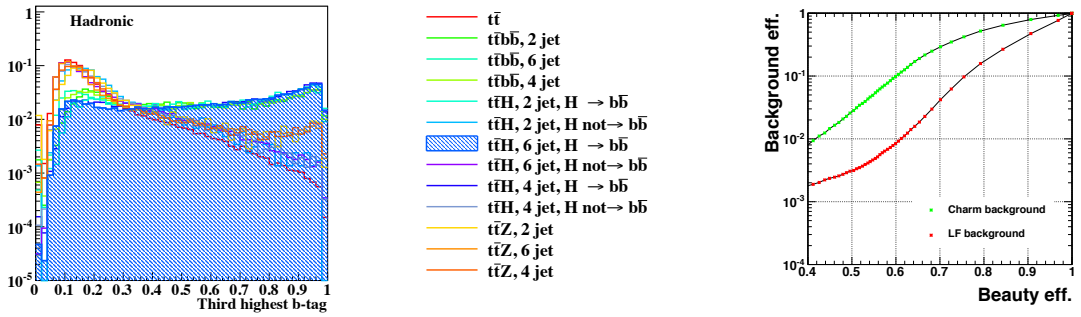
(c)

Figure 13: Reconstructed invariant masses using the χ^2 method to group the jets (no truth-matching). (a): $W^{\pm}$. (b): Top. (c): Higgs.

8. Beauty identification

As most of the backgrounds in this analysis do not have four b-jets in the final state, flavour-tagging the reconstructed jets provides discriminating power between signal and background.

The LCFIPlus package is used to re-cluster the particles which were not identified as beam-jets by FastJet. Secondary vertices are identified, and decay products from these vertices are constrained to be within the same jet as their parent. Flavour-tagging is performed, assigning each jet with a b- and c-tag probability (see for example Figure 14a). This is achieved using simulated samples of $e^+e^- \rightarrow qqqqqq$, where all six quarks have the same flavour. A high multiplicity of jets in the final states of these samples ensures the jets have similar kinematic qualities to those in the analysis samples. Eight boosted decision trees (BDTs) are created; for each of the two flavours, four BDTs are trained depending on the characteristics of the jets, for example the number of secondary vertices or the existence of high impact parameter tracks. Figure 14b shows the fake rates as a function of the efficiency for the b-jet flavour-tagging. For a 60% beauty efficiency, the retention rate for the charm background is 10%. Flavour-tag variables for both signal channels are shown in the Appendices (Figures 20, 21, 27, and 28).



(a) (b)

Figure 14: (a): Normalised beauty probability for the third highest b-tagged jet in the hadronic channel. (b): Beauty-tagging efficiency as a function of light flavour background efficiency (red) and charm background efficiency (green).

9. Event selection

Events are selected using a BDT as implemented in TMVA [28]. Gradient boosting is used. A pre-selection is made, requiring that all events have exactly zero or one isolated lepton, for the hadronic and semi-leptonic channels respectively. Two BDTs are trained independently for the analysis, one for each channel. A collection of variables is input to each BDT. There is large overlap in the variables used in the two channels, with a few variables specific to one or the other channel.

The following variables are used in both channels:

- the reconstructed Higgs mass. This discriminates background from $\mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{t\bar{t}Z}$, where no Higgs is present (see Figures 19e, 26e);
- the number of reconstructed particles in the event. This is strongly correlated to the number of jets in the event (see Figures 22e, 29c);
- the visible energy contained in jets (see Figures 22c, 29a);
- the amount of missing p_{T}, calculated by summing the momentum in the transverse plane for each jet and the possible one lepton. In the six-jet analysis, the neutrino from the leptonic $\mathrm{W}^{\pm}$ decay will cause some missing p_{T} (see Figures 22f, 29d);
- the χ^2 of the jet grouping, as described in Equation 4 (see Figures 22d, 29b);
- the event shape variables thrust, sphericity and aplanarity (see Figures 18, 25);
- the four highest b-tag values of the event, and the corresponding c-tag values (see Figures 20, 21, 27, 28);
- the cosine of the decay angle of the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay, and the cosine of the angles between the Higgs and each top quark (see Figures 23, 30);
- the distance value between the two closest jets, defined by the Durham jet clustering algorithm for jets i and $j = i+1$ as

$$y_{ij} = \frac{\min(E_i^2, E_j^2)(1-\cos\theta_{i,j})}{s}, \tag{6}$$

  where $\sqrt{s}$ is the centre-of-mass energy, E the jet energies and i, j are chosen to minimise the distance of the two jets which are merged. The jet transition values y_{45}, y_{56} and y_{67} are used in the BDT to discriminate signal from lower jet multiplicity backgrounds, in particular $\mathrm{t\bar{t}}$ (see Figures 24, 31).

Variables used only for the semi-leptonic analysis are:

- the cone energy of the isolated lepton (see Figure 22b);
- the ratio of energy deposits in the calorimeter of the isolated lepton, as described in Equation 2 (see Figure 22a).

Variables used only for the hadronic analysis are:

- the energy of the four lowest-energy jets (see Figure 32);
- the cosine of the angle of the two jets closest to the beam-axis (see Figure 30).

The event samples are split randomly in two. The first random half of each sample is input to TMVA, and is further split in two to train and test the BDT. These samples are weighted to represent $1.5\,\mathrm{ab}^{-1}$. Figure 15 shows the trained BDT response to the training and testing datasets. No sign of overtraining is observed.

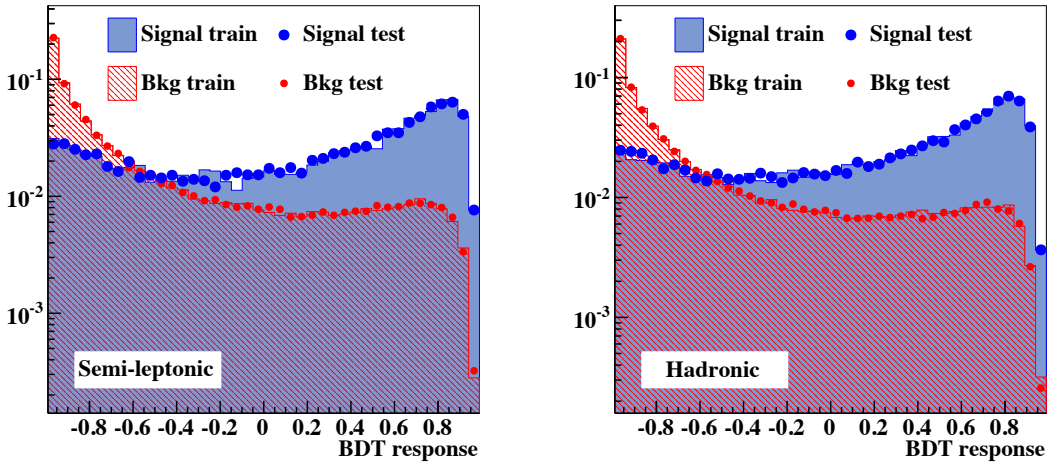


Figure 15: BDT overtraining check showing the normalised BDT response for training and testing samples. (a): Semi-leptonic channel. (b): Hadronic channel.

The selection on the BDT response value is chosen using the second random half of the event samples. The BDT response is calculated for each event (see Figure 16). The optimal selection is calculated by finding the maximum significance, defined as:

$$\frac{S}{\sqrt{S+B}}, \tag{7}$$

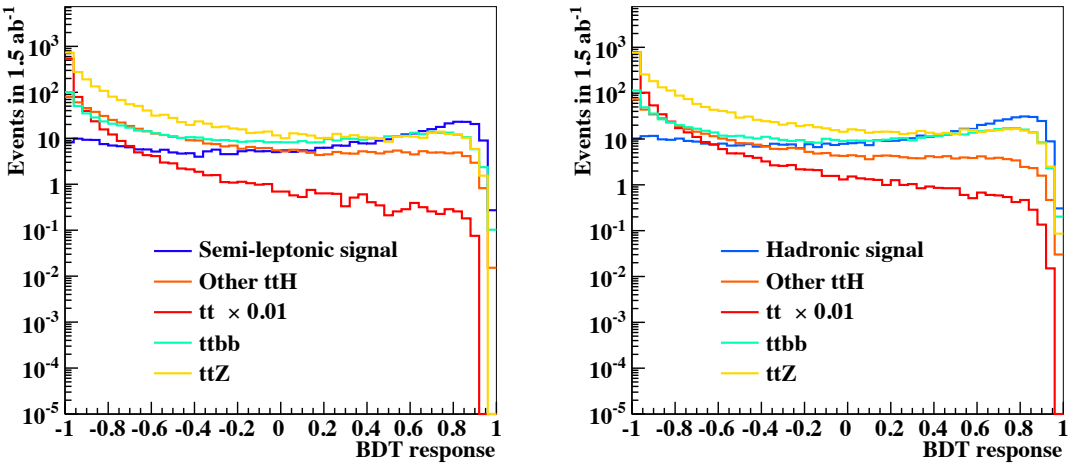
where S denotes the number of selected signal events, and B the number of selected background events. This results in a significance of 8.36 in the semi-leptonic channel, and 9.17 in the hadronic channel.

9.1. Selection efficiency

Applying the selection on BDT response to the event samples results in the efficiencies shown in Table 4.

Table 4: Selection efficiency for each event sample. Column 1 shows the simulated process. Column 2 shows the expected number of events in $1.5\,\mathrm{ab}^{-1}$. Column 3 shows the number of events in which 0 leptons were found. Column 4 shows the number (and percent) of these '0 lepton' events which pass the BDT trained for the hadronic channel. Column 5 shows the number of events in which 1 lepton was found. Column 6 shows the number (and percent) of these '1 lepton' events which pass the BDT trained for the semi-leptonic channel. The number of jets refers to the $\mathrm{t\bar{t}}$ decay only.

Process	Evt in $1.5\,\mathrm{ab}^{-1}$	Evt with 0 leptons	Evt pass Had BDT		Evt with 1 lepton	Evt pass SL BDT	
$\mathrm{t\bar{t}H}$, 6 jet, $\mathrm{H} \to \mathrm{b\bar{b}}$	647	593	357	(60.2%)	49	9	(18.8%)
$\mathrm{t\bar{t}H}$, 4 jet, $\mathrm{H} \to \mathrm{b\bar{b}}$	623	178	62	(35.1%)	420	233	(55.3%)
$\mathrm{t\bar{t}H}$, 2 jet, $\mathrm{H} \to \mathrm{b\bar{b}}$	150	13	1	(10.7%)	61	20	(32.5%)
$\mathrm{t\bar{t}H}$, 6 jet, $\mathrm{H} \not\to \mathrm{b\bar{b}}$	473	306	38	(12.3%)	127	8	(6.52%)
$\mathrm{t\bar{t}H}$, 4 jet, $\mathrm{H} \not\to \mathrm{b\bar{b}}$	455	89	5	(5.81%)	246	19	(7.82%)
$\mathrm{t\bar{t}H}$, 2 jet, $\mathrm{H} \not\to \mathrm{b\bar{b}}$	110	6	0	(1.52%)	33	1	(3.66%)
$\mathrm{t\bar{t}b\bar{b}}$, 6 jet	824	737	287	(38.9%)	80	8	(9.75%)
$\mathrm{t\bar{t}b\bar{b}}$, 4 jet	794	222	44	(19.6%)	533	175	(32.9%)
$\mathrm{t\bar{t}b\bar{b}}$, 2 jet	191	16	1	(8.71%)	78	14	(18.1%)
$\mathrm{t\bar{t}Z}$, 6 jet	2,843	2,335	316	(13.5%)	322	12	(3.68%)
$\mathrm{t\bar{t}Z}$, 4 jet	2,738	711	49	(6.86%)	1,678	170	(10.2%)
$\mathrm{t\bar{t}Z}$, 2 jet	659	54	1	(2.03%)	248	13	(5.23%)
$\mathrm{t\bar{t}}$	203,700	111,020	1,399	(1.26%)	77,110	523	(0.68%)



(a) (b)

Figure 16: BDT response for signal and background samples, scaled to the number of events expected in 1.5 ab^{-1}. (a): Semi-leptonic channel. (b): Hadronic channel.

10. Results

Sensitivity to the $t\bar{t}H$ cross-section can be calculated directly from the inverse of the signal significance. Assuming an integrated luminosity of 1.5 ab^{-1}, the $t\bar{t}H$ cross-section can be measured with an accuracy of 12.0% in the semi-leptonic channel and 10.9% in the hadronic channel. The combined precision of the two channels is 8.1%.

To extract the top Yukawa coupling (g_{ttH}) from the measured cross-section ($\sigma(t\bar{t}H)$), signal samples were simulated using different input values of the coupling. It was calculated (see Figure 3) that:

$$\frac{\Delta g_{ttH}}{g_{ttH}} = 0.53 \times \frac{\Delta\sigma(t\bar{t}H)}{\sigma(t\bar{t}H)}. \tag{8}$$

This factor is slightly bigger than 0.5 because of the additional signal-like events produced by Higgsstrahlung (see Figure 2).

Therefore, the uncertainty on the measured cross-section translates into a precision on the top Yukawa coupling of 4.27%. If beam polarisation were included in the analysis, it is expected that the precision would be less than 4%. This result is comparable with the $\sqrt{s} = 1$ TeV ILC study, which found a precision of 4.5% [2]. Both results from linear colliders are significantly more precise than the LHC experiments are expected to achieve with 3000 fb^{-1} at $\sqrt{s} = 14$ TeV.

Systematic uncertainties in the analysis are expected to be significantly smaller than the statistical uncertainty presented on the cross-section (8.1%). Systematic uncertainties from sources such as flavour-tagging, lepton reconstruction and the jet energy scale are all expected to be well below 10%. The systematic uncertainty due to the uncertainty of the luminosity spectrum reconstruction was discussed in another analysis at $\sqrt{s} = 1.4$ TeV [29] and was found to be small.

A. Number of leptons

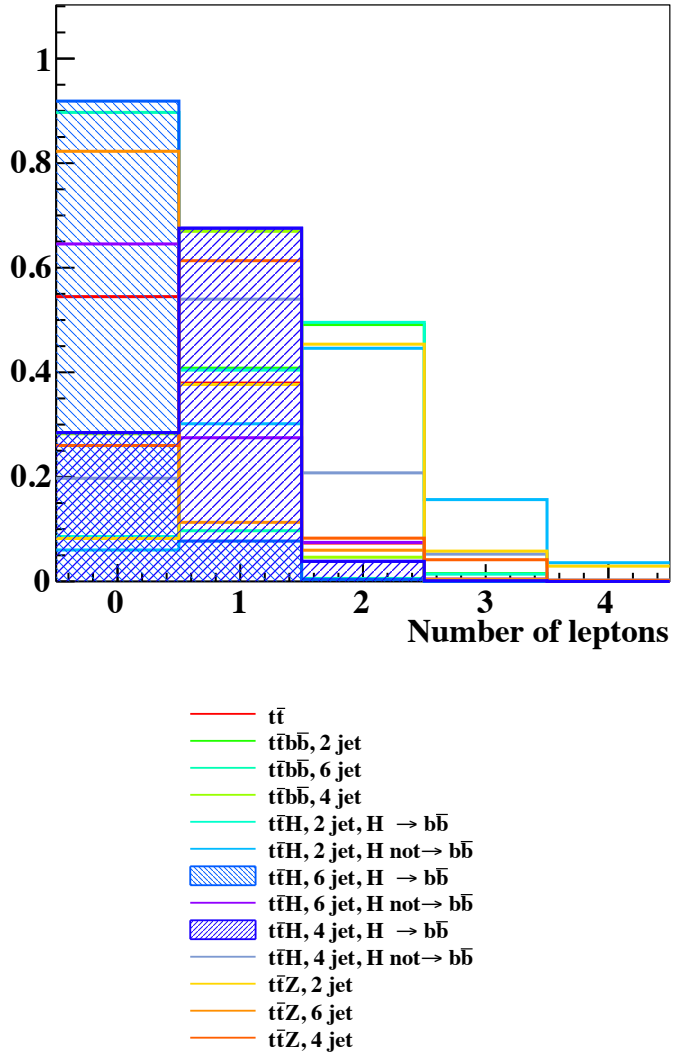


Figure 17: The number of leptons found per event in each sample, normalised to unit area.

B. Semi-leptonic analysis variables

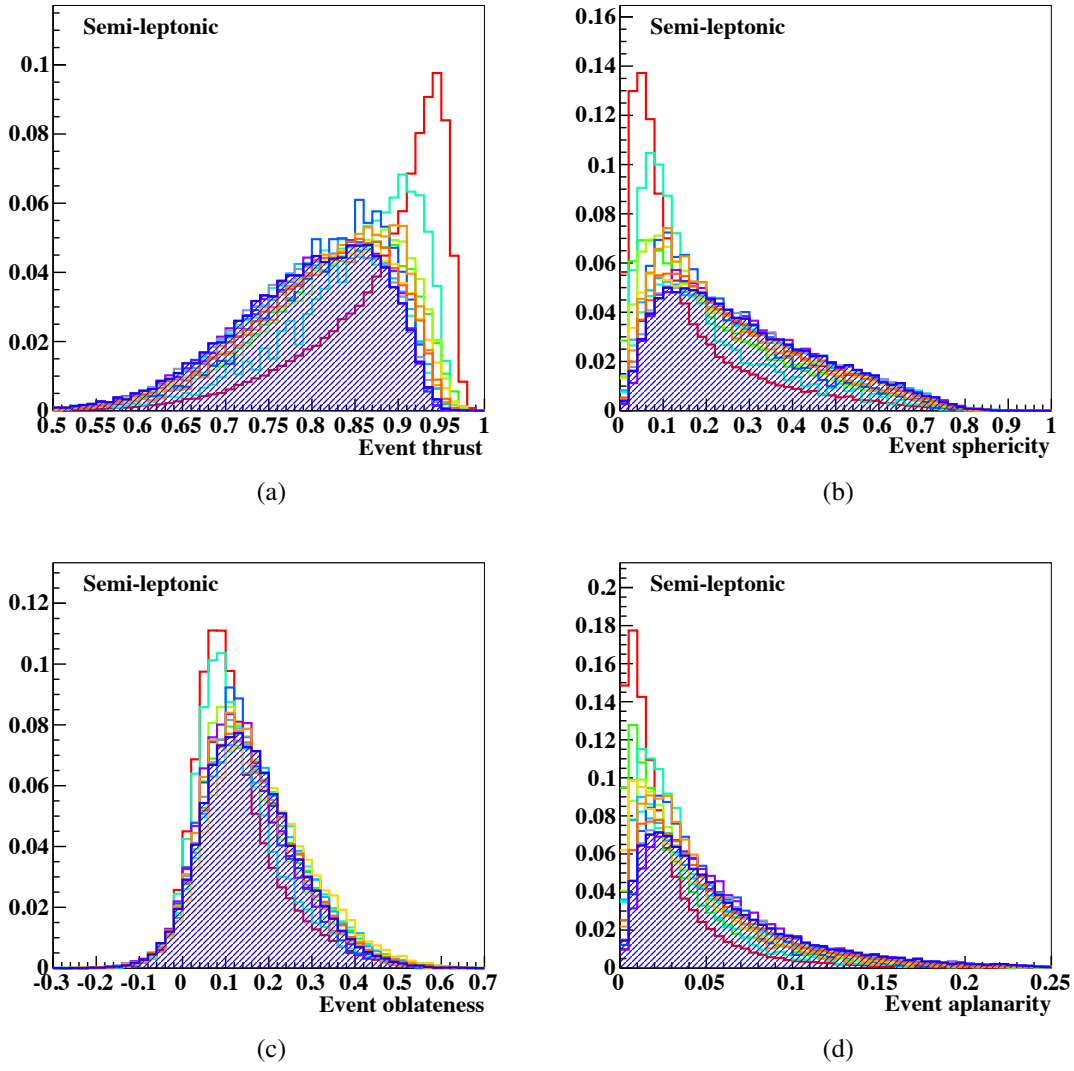


Figure 18: Event shape variable plots, calculated from all reconstructed particles in the event and normalised to unit area.

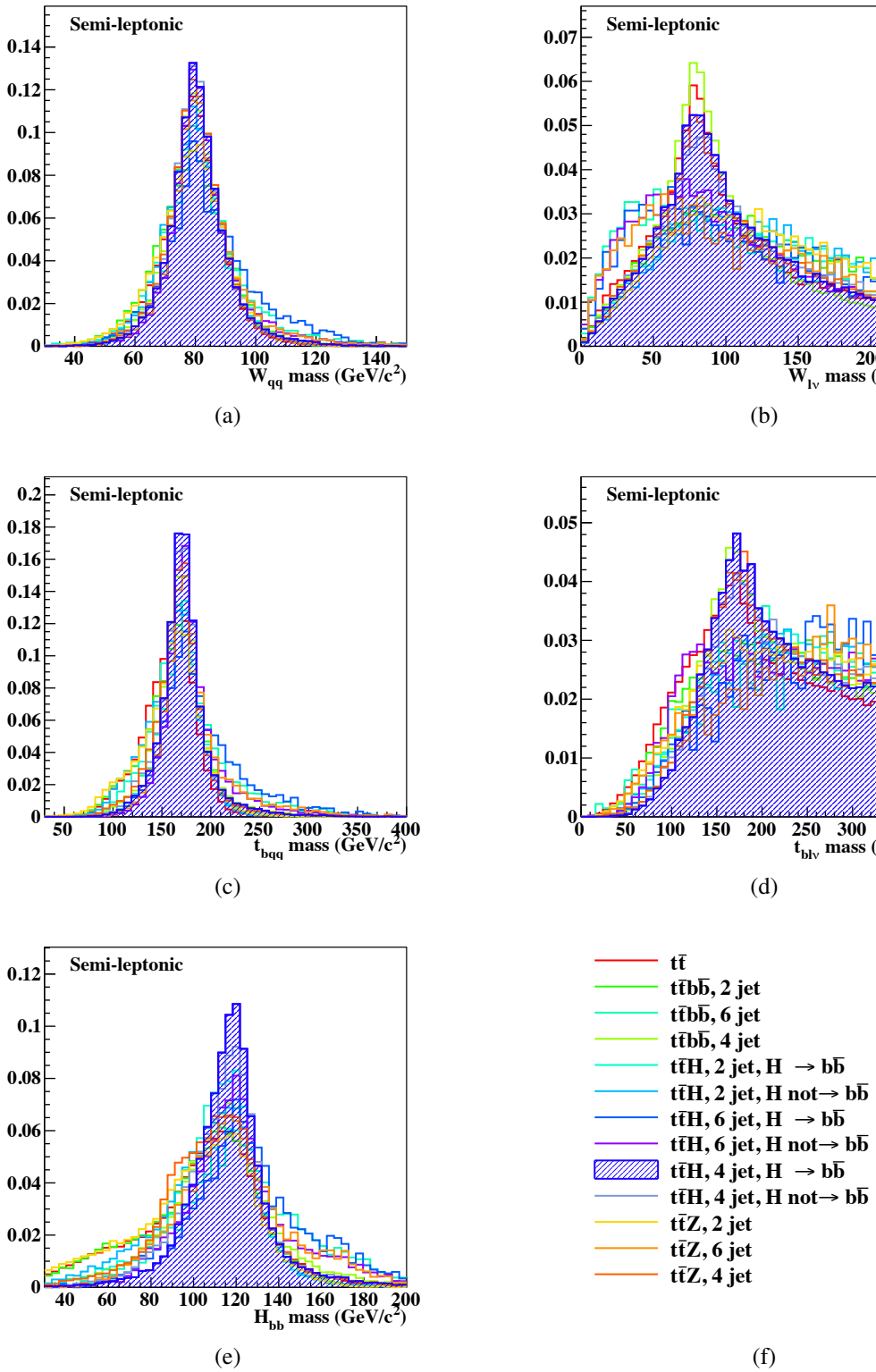
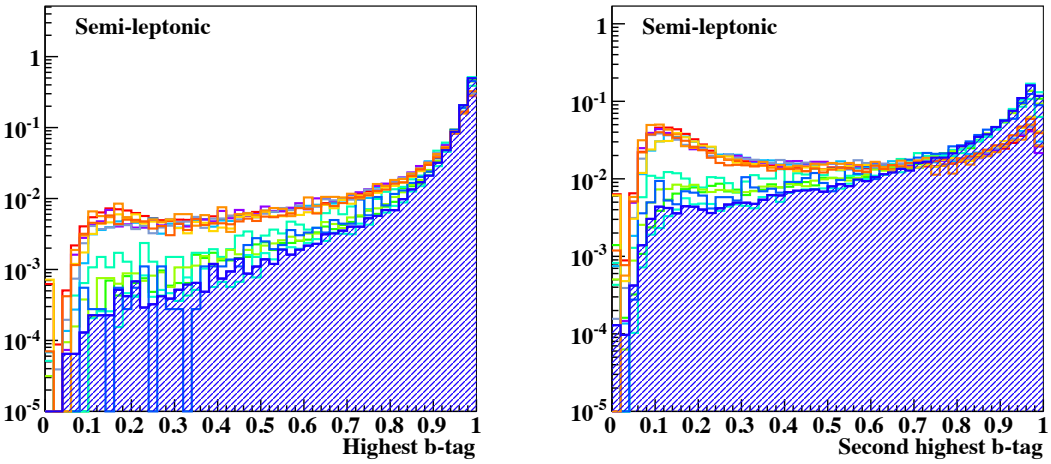
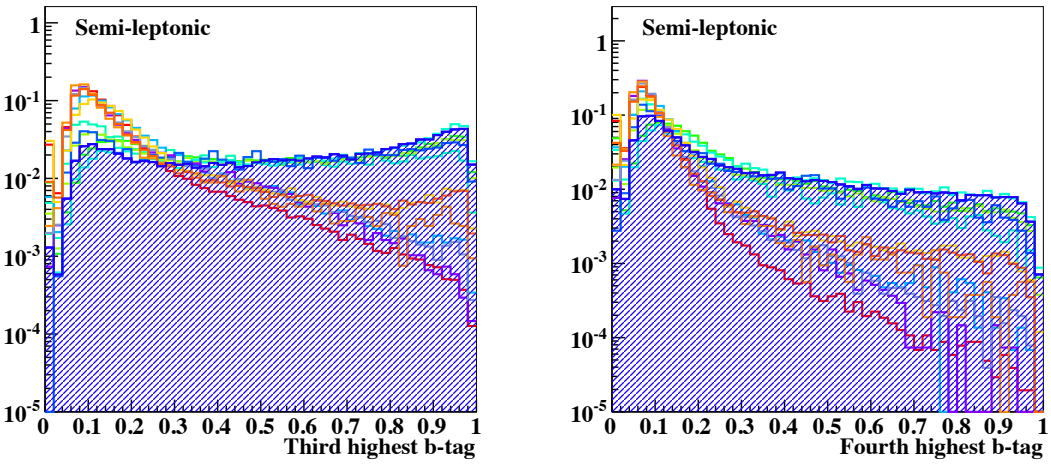


Figure 19: Reconstructed mass variable plots, normalised to unit area.

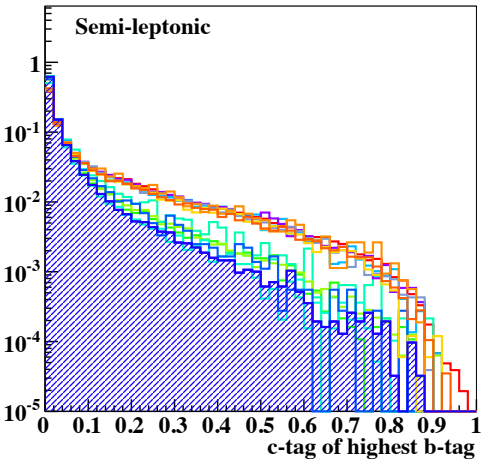


(a) (b)

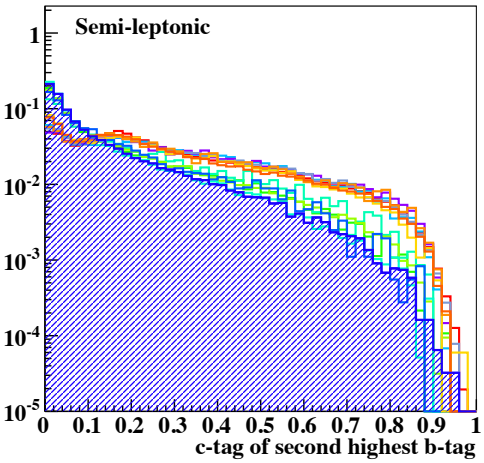


(c) (d)

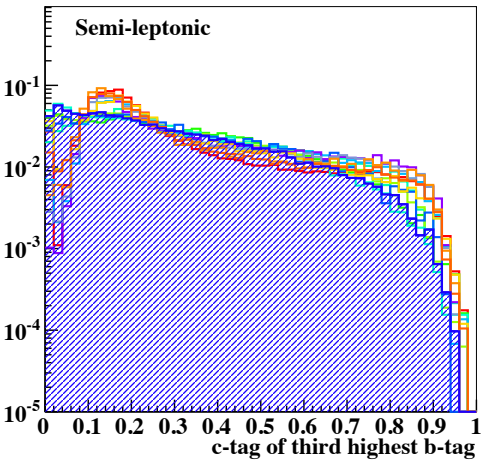
Figure 20: Four highest b-tag value plots, normalised to unit area.



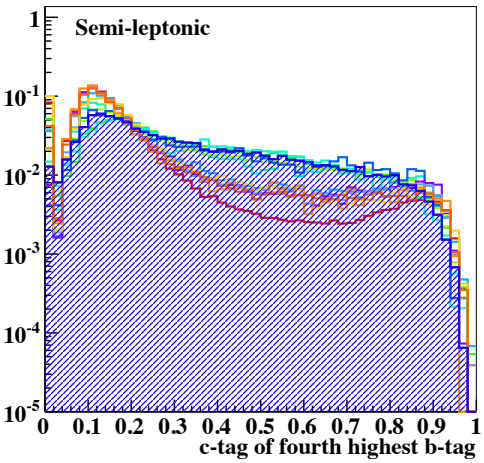
(a)



(b)

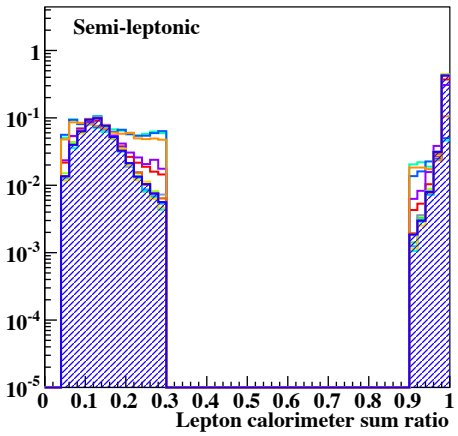


(c)

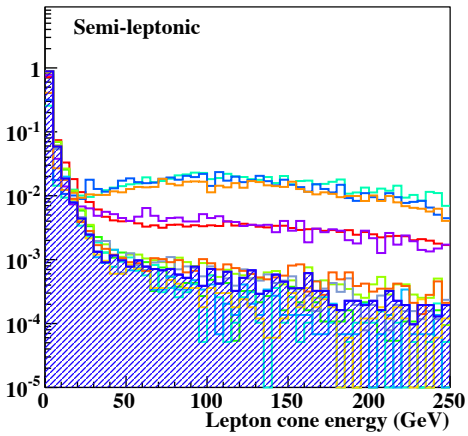


(d)

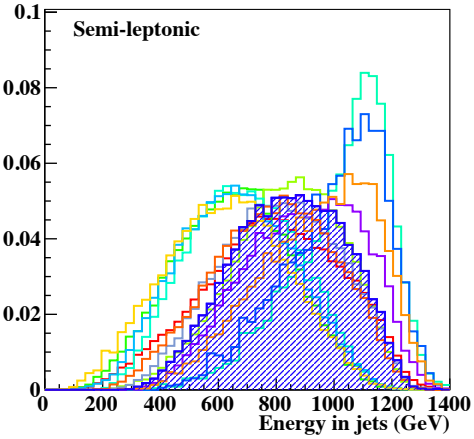
Figure 21: c-tag value plots of the four highest b-tagged jets, normalised to unit area.



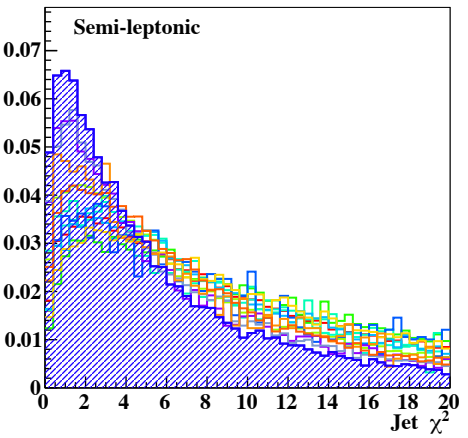
(a)



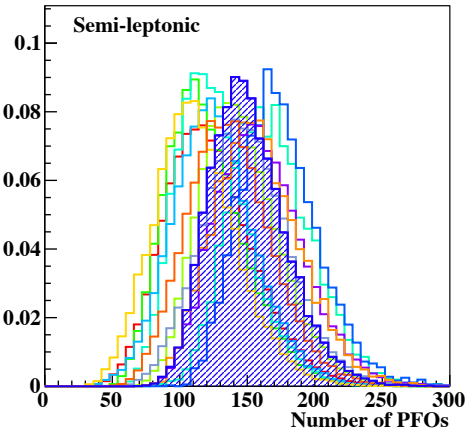
(b)



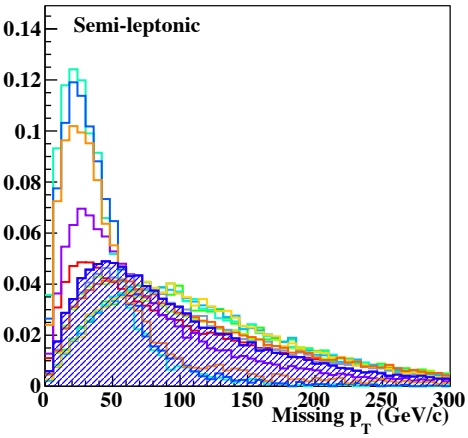
(c)



(d)

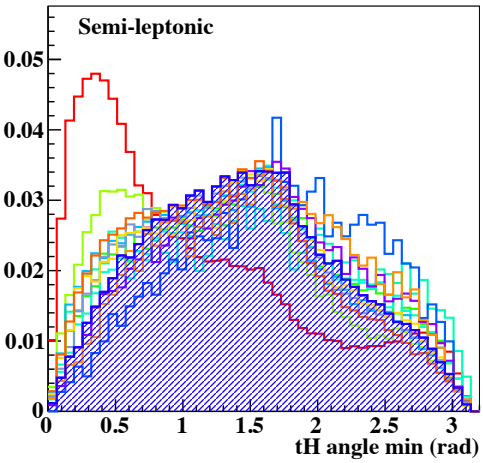


(e)

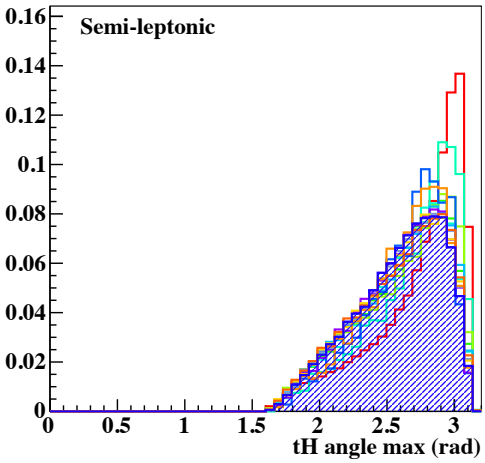


(f)

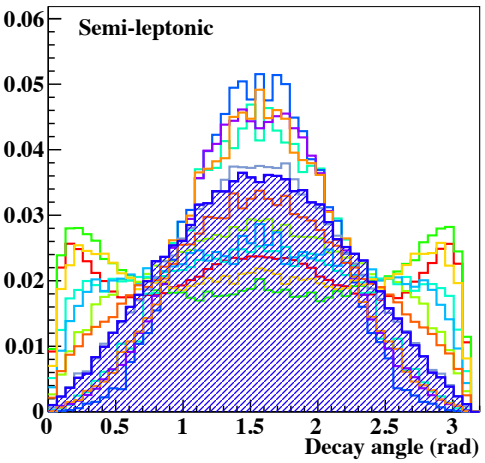
Figure 22: Lepton and event variable plots, normalised to unit area.



(a)

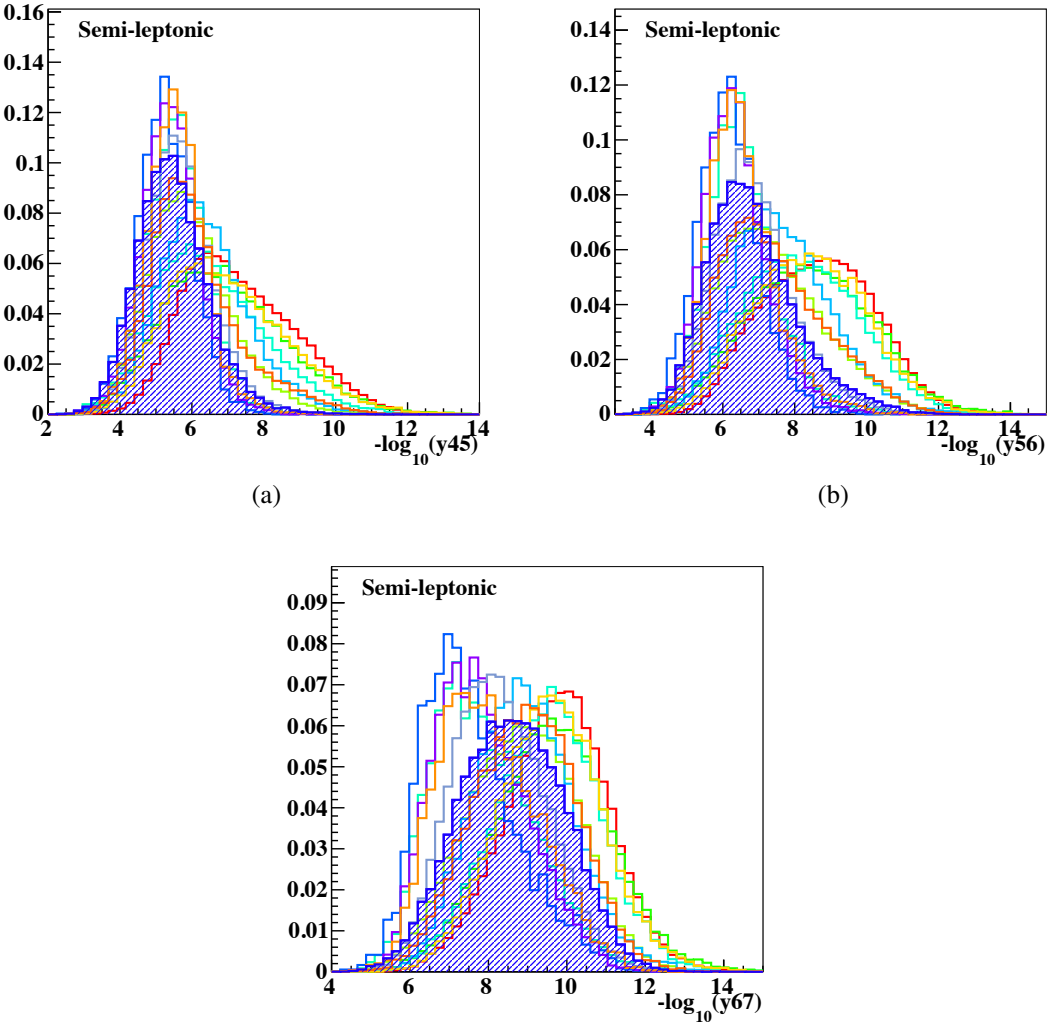


(b)



(c)

Figure 23: Angular variable plots, normalised to unit area.



(a) (b)

(c)

Figure 24: Jet transition variable plots, normalised to unit area.

C. Hadronic analysis variables

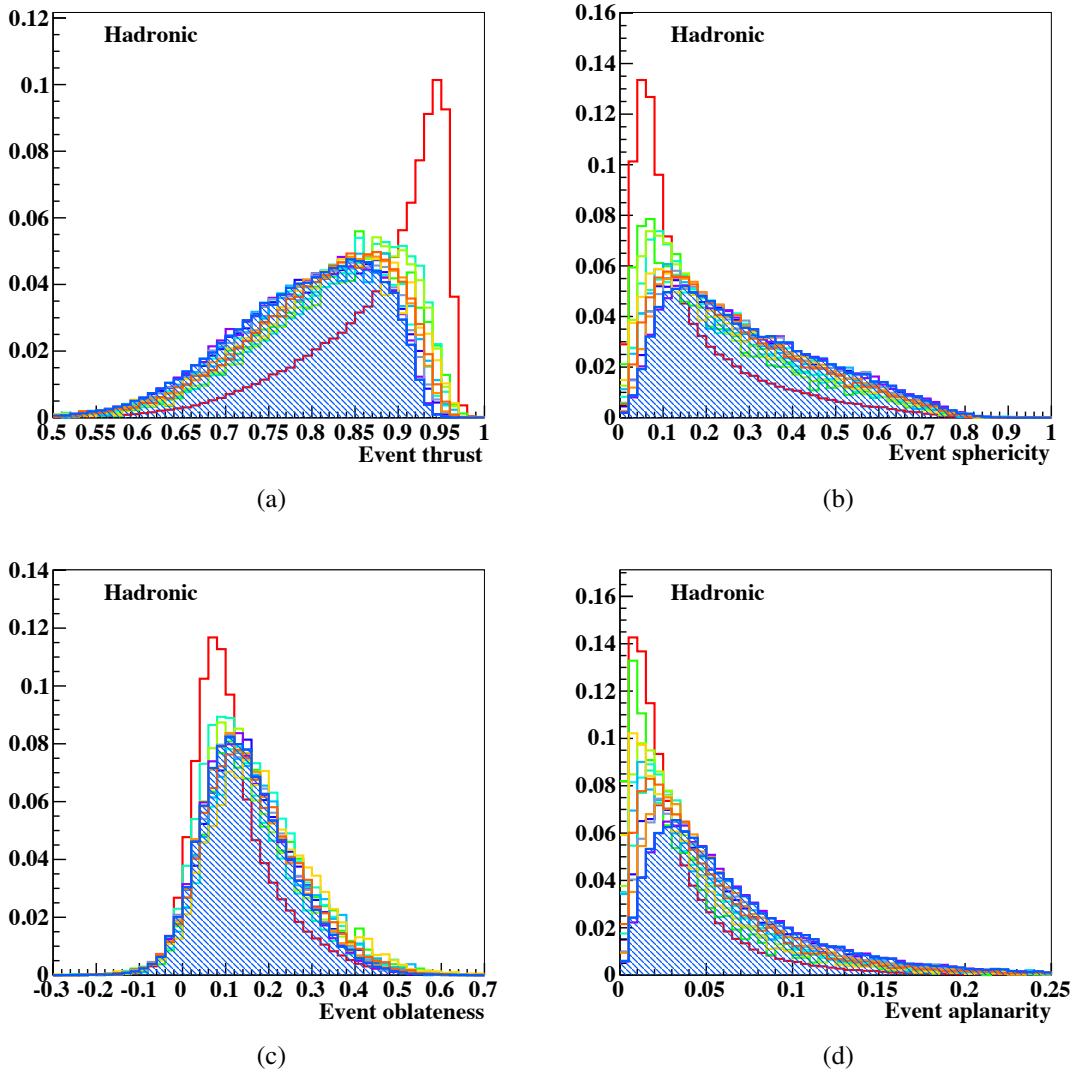
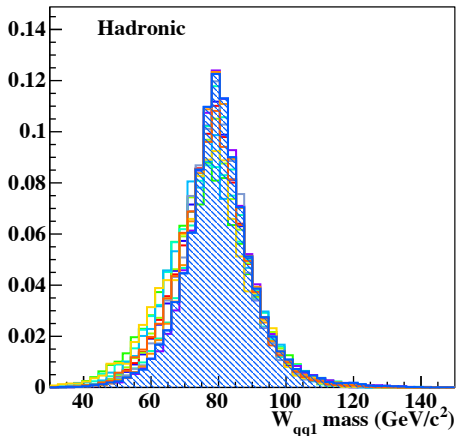
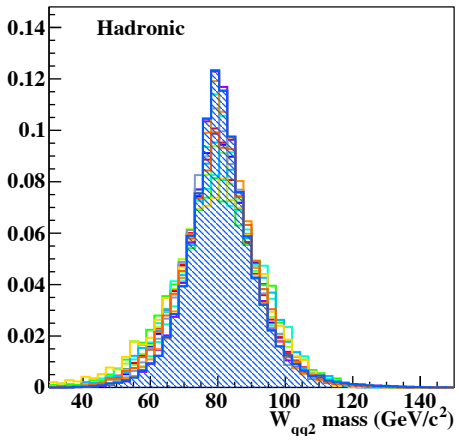


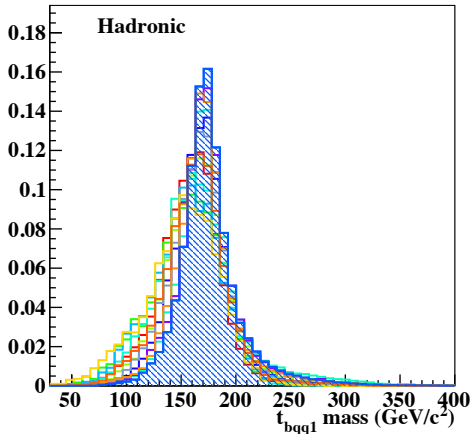
Figure 25: Event shape variable plots, normalised to unit area.



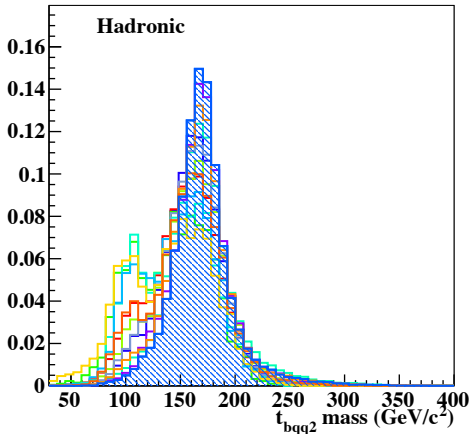
(a)



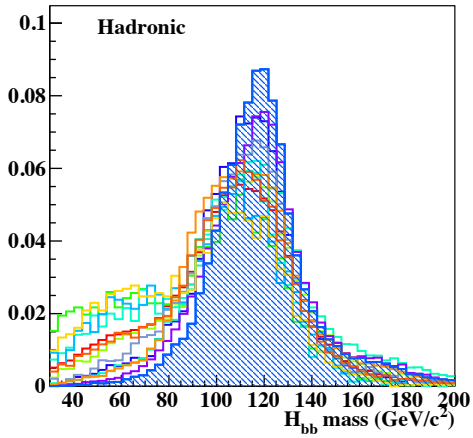
(b)



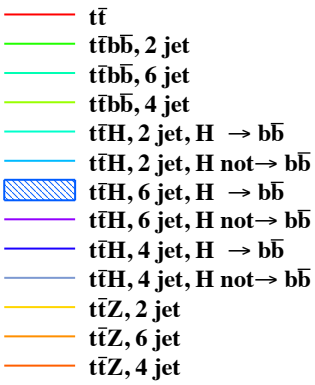
(c)



(d)

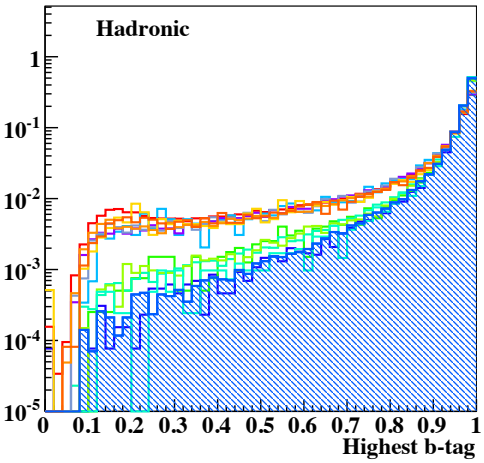


(e)

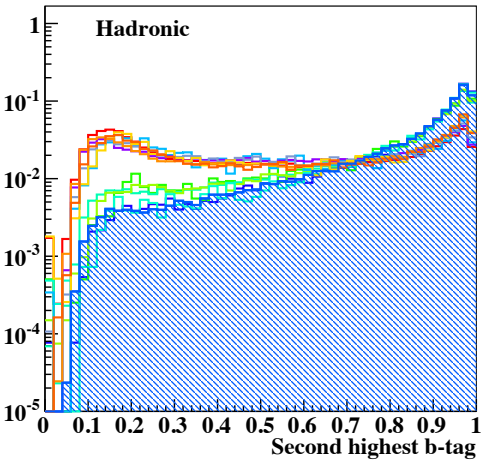


(f)

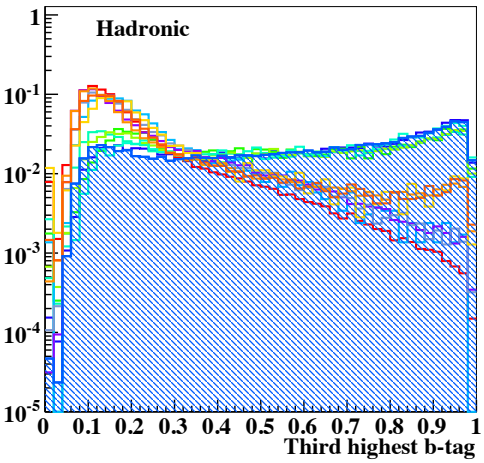
Figure 26: Reconstructed mass variable plots, normalised to unit area.



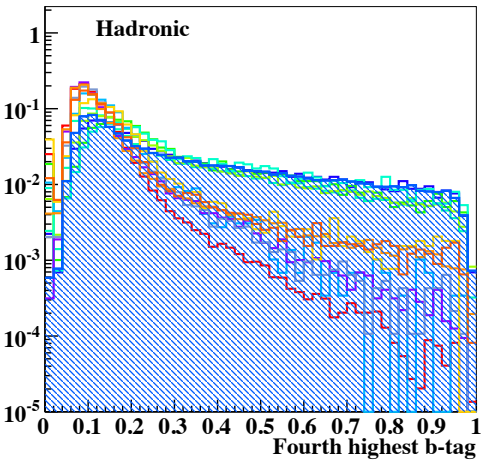
(a)



(b)

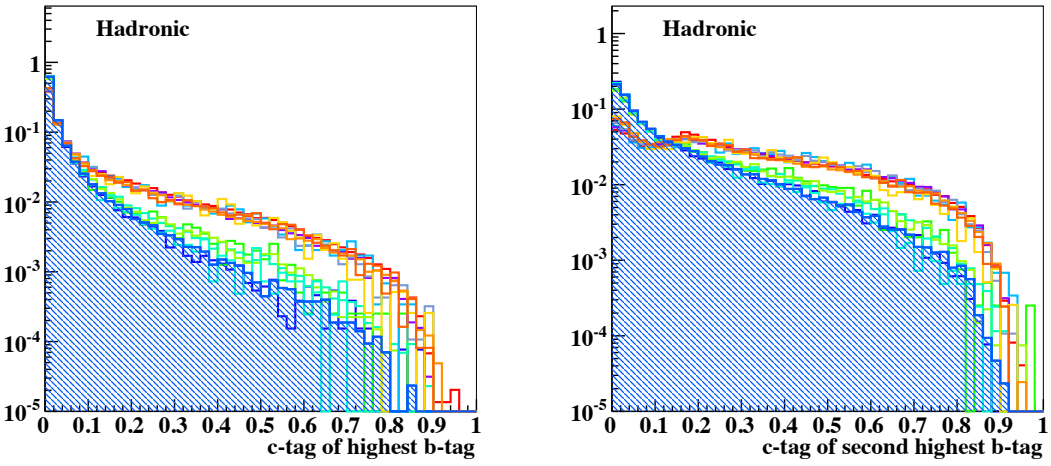


(c)

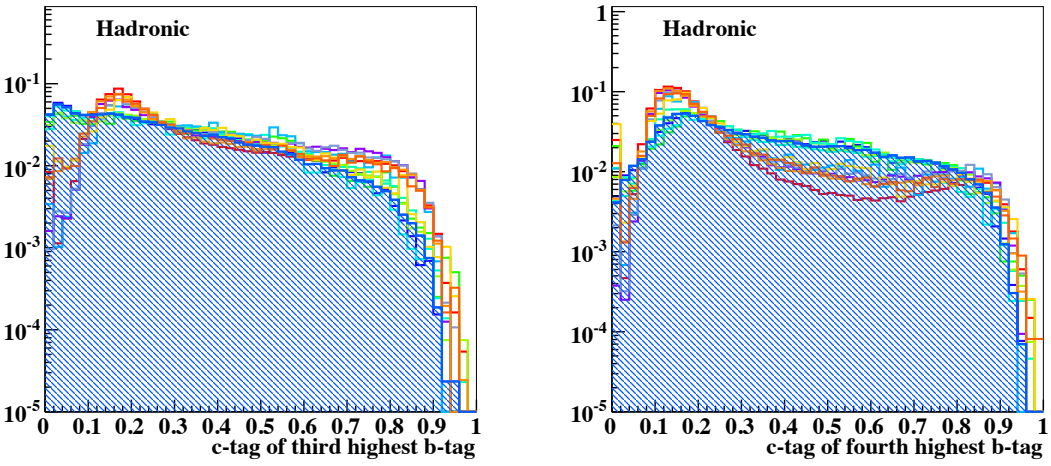


(d)

Figure 27: Four highest b-tag value plots, normalised to unit area.

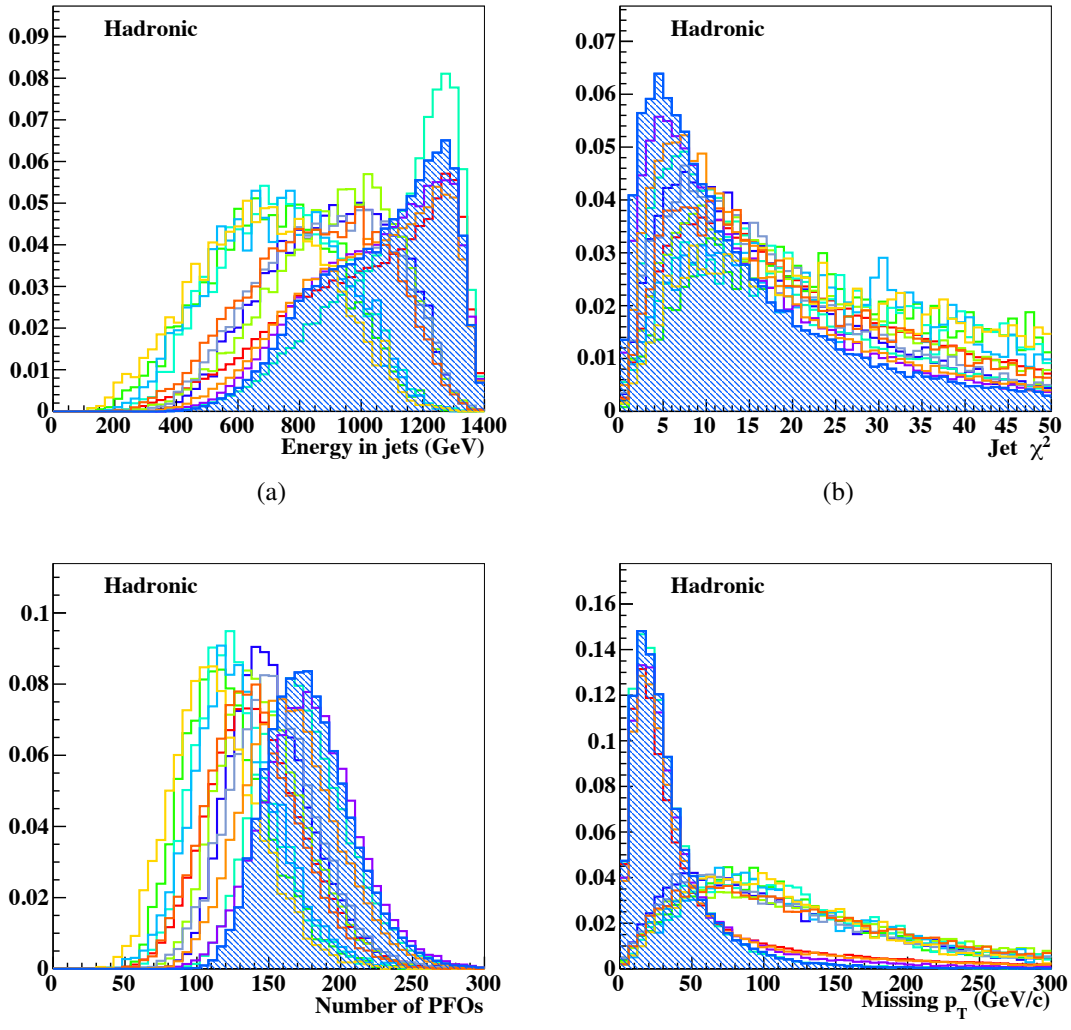


(a) (b)



(c) (d)

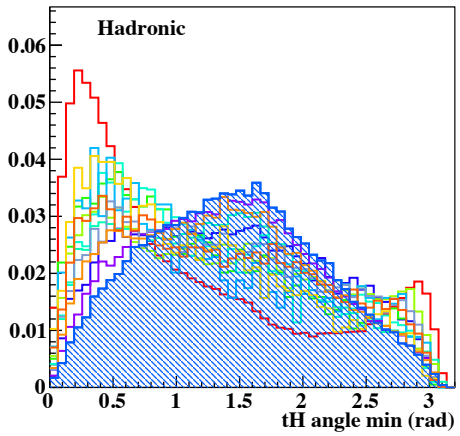
Figure 28: c-tag value plots of the four highest b-tagged jets, normalised to unit area.



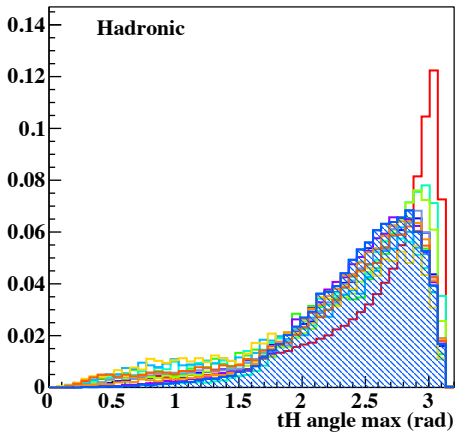
(a) (b)

(c) (d)

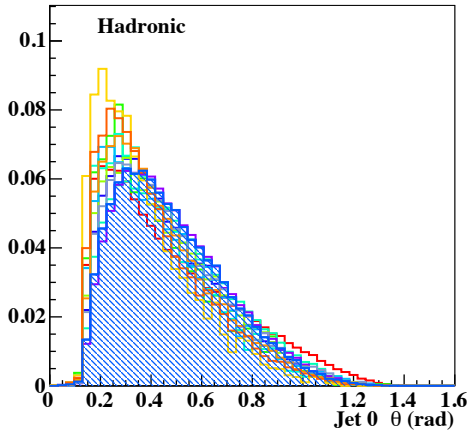
Figure 29: Event variable plots, normalised to unit area.



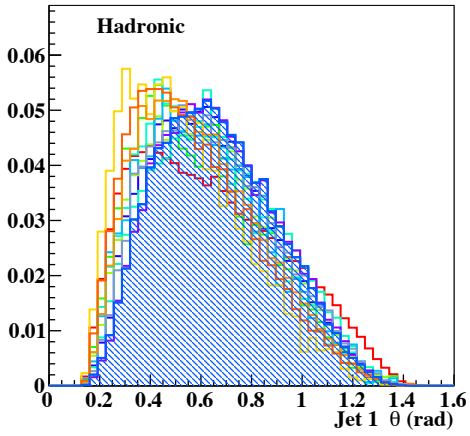
(a)



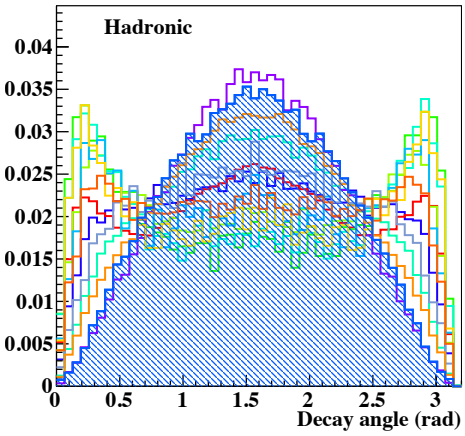
(b)



(c)

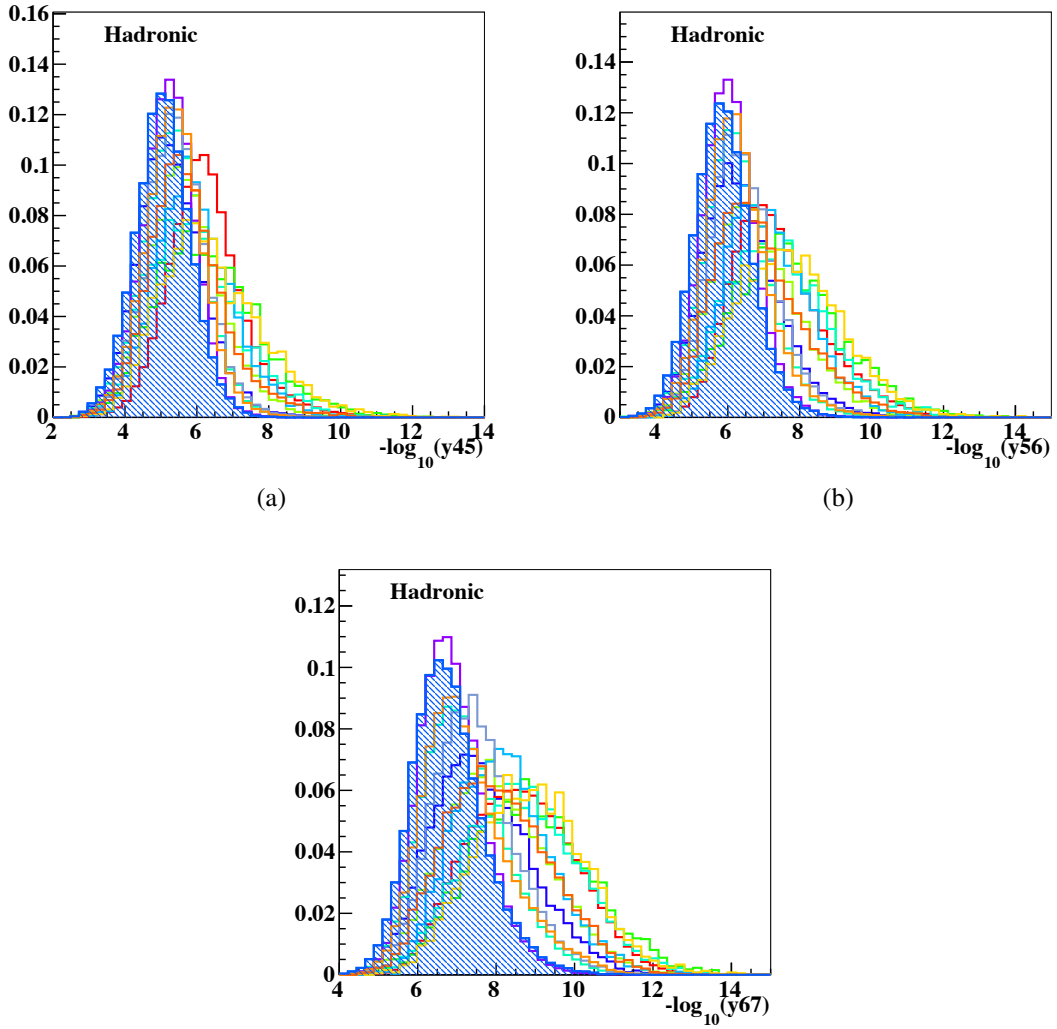


(d)



(e)

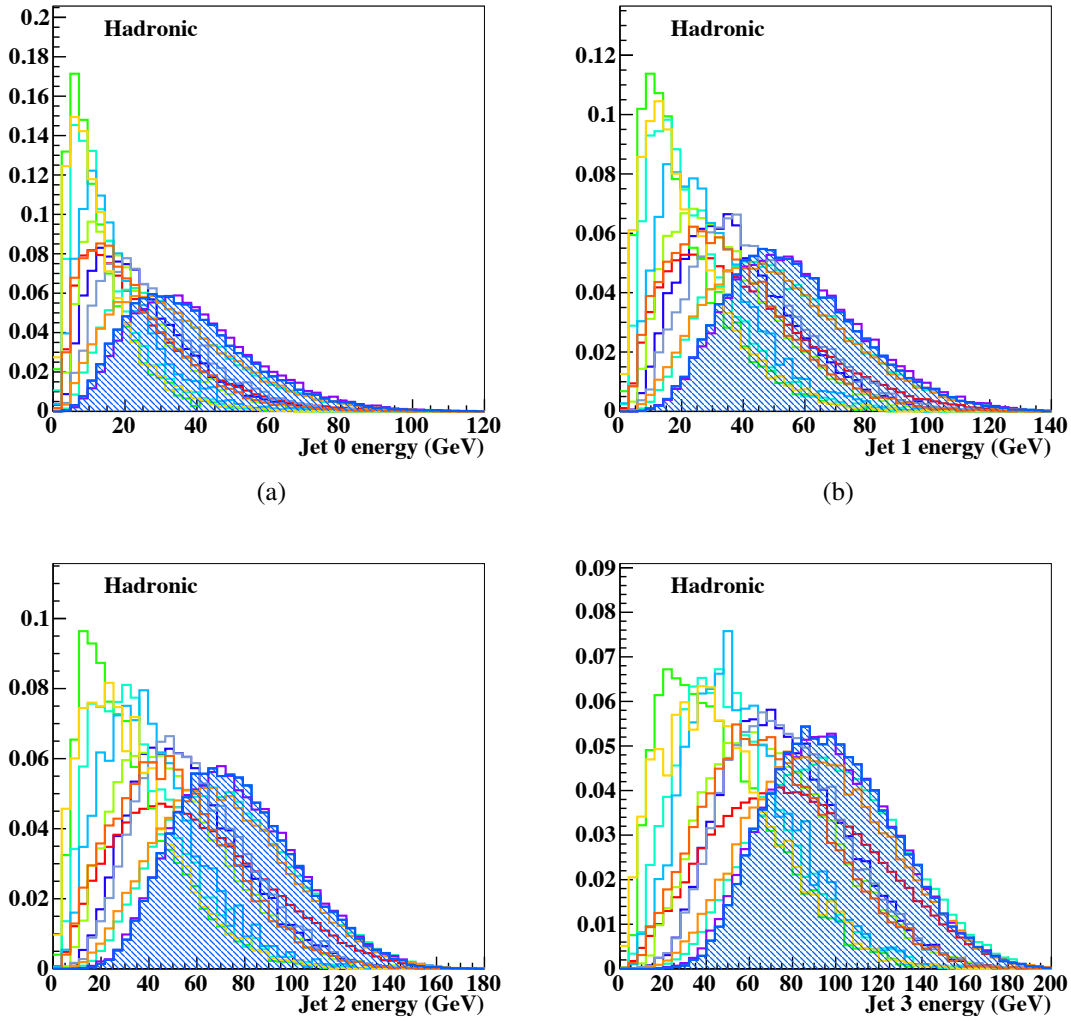
Figure 30: Angular variable plots, normalised to unit area.



(a) (b)

(c)

Figure 31: Jet transition variable plots, normalised to unit area.



(a) (b)

(c) (d)

Figure 32: Jet energy variable plots, normalised to unit area.

References

[1] M. Aicheler *et al.*, A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report, Tech. Rep. CERN-2012-007. SLAC-R-985. KEK-Report-2012-1. PSI-12-01. JAI-2012-001, Geneva, 2012

[2] P. Roloff and J. Strube, Measurement of the top Yukawa Coupling at a 1 TeV International Linear Collider using the SiD detector, LCD-NOTE 2013-001, 2013

[3] T. Price *et al.*, Measurement of the top Yukawa coupling at $\sqrt{s} = 1$ TeV using the ILD detector, LC-REP 2013-004, 2013

[4] Search for ttH production in the $H \to \gamma\gamma$ channel at $\sqrt{s} = 8$ TeV with the ATLAS detector, ATLAS-CONF 2013-080, CERN, Geneva, Jul 2013

[5] S. Chatrchyan *et al.*, Search for the standard model Higgs boson produced in association with a top-quark pair in pp collisions at the LHC, *JHEP* **1305** (2013) 145

[6] K. Agashe *et al.*, Snowmass 2013 Top quark working group report (2013)

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

[8] W. Kilian, T. Ohl and J. Reuter, WHIZARD: Simulating Multi-Particle Processes at LHC and ILC, *Eur. Phys. J.* **C 71** (2011) 1742

[9] M. Moretti, T. Ohl and J. Reuter, O'Mega: An Optimizing matrix element generator, LC-TOOL 2001-040, 2001

[10] T. Sjostrand, S. Mrenna and P. Z. Skands, PYTHIA 6.4 Physics and Manual, *JHEP* **0605** (2006) 026

[11] A. Miyamoto *et al.*, Physics and Detectors at CLIC: CLIC Conceptual Design Report, Tech. Rep. arXiv:1202.5940. CERN-2012-003. ANL-HEP-TR-12-01. DESY-12-008. KEK-Report-2011-7, Geneva, 2012, comments: 257 p, published as CERN Yellow Report CERN-2012-003

[12] C. Grefe and A. Muennich, The CLIC SiD CDR Detector Model for the CLIC CDR Monte Carlo Mass Production, LCD-Note 2011-009, Jun 2011

[13] S. Agostinelli *et al.*, Geant4 – a simulation toolkit, *Nucl. Instrum. Methods Phys. Res., Sect.* **A 506** (2003) 250

[14] J. Allison *et al.*, Geant4 developments and applications, *IEEE T. Nucl. Sci.* **53** (2006) 270

[15] M. Thomson, Particle Flow Calorimetry and the PandoraPFA Algorithm, *Nucl. Instrum. Meth.* **A 611** (2009) 25

[16] J. Marshall, A. Muennich and M. Thomson, Particle Flow Performance at CLIC, LCD-Note 2011-028, Jan 2012

[17] S. Catani *et al.*, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions, *Nucl. Phys.* **B 406** (1993) 187

[18] S. Ellis and D. Soper, Successive combination jet algorithm for hadron collisions, *Phys. Rev.* **D 48** (1993) 3160

[19] M. Cacciari, G. Salam and G. Soyez, FastJet User Manual, *Eur. Phys. J.* **C 72** (2012) 1896

[20] LCFIPlus, `https://confluence.slac.stanford.edu/display/ilc/LCFIPlus`

[21] S. Catani *et al.*, New clustering algorithm for multijet cross sections in e^+e^- annihilation, *Phys. Lett.* **B 269** (1991) 432

[22] G. Grefe *et al.*, ILCDIRAC, a DIRAC extension for the Linear Collider community, CLICdp-Conf 2013-003, CERN, Geneva, Nov 2013

[23] T. Behnke *et al.*, The International Linear Collider Technical Design Report - Volume 4: Detectors, arXiv 1306.6329, Jun. 2013

[24] F. Gaede and J. Engels, Marlin et al - A Software Framework for ILC detector R&D, EUDET-Report 2007-11, 2007

[25] A. Münnich, TauFinder: A reconstruction algorithm for τ leptons at linear colliders, LCD-Note 2010-009, 2010

[26] Verkerke, W. and Kirkby, D., The RooFit toolkit for data modeling, *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, 2005, proceedings of PHYSTAT 05

[27] D. Dannheim, Update on $\gamma\gamma \to$ hadrons samples at $\sqrt{s} = 1.4$ TeV, 2012, Slides

[28] A. Hoecker *et al.*, TMVA - Toolkit for Multivariate Data Analysis, *POSACAT* **040** (2007)

[29] P. Roloff, Physics performance for measurements of chargino and neutralino pair production at a 1.4 TeV CLIC collider, LCD-Note 2012-006, Feb 2013