Abstract

Higgs mass and cross-section measurements have been examined to assess the capability of a 500 GeV CLIC machine, operating at centre-of-mass energies of 350 GeV and 500 GeV. A Higgs mass of 120 GeV and a luminosity of 500 fb$^{-1}$ were assumed. Model-independent measurements were performed by examining the recoil of the $Z$ in the Higgsstrahlung process, with the $Z$ subsequently decaying to a pair of muons or electrons. At 350 GeV, the muon channel yielded a mass precision of 133 MeV and a $HZ$ cross-section precision of 4.9 %. At 500 GeV, the measurement accuracy decreased due to reduced lepton momentum resolution and increased background.

Model-dependent studies at 500 GeV investigated the four-jet final state and the two-jet neutrino final state, which includes contributions from both Higgsstrahlung and WW-fusion. The four-jet study yielded a mass precision of 104 MeV and a $HZ$ cross-section precision of 1.6 %. The two-jet neutrino study provided a mass precision of 97 MeV and an inclusive cross-section precision of 1.01 %. Fitting the $p_T$ distribution of selected events allowed the relative Higgsstrahlung and WW-fusion normalisations to be measured with a precision of 5.1 %, thereby measuring the ratio of the couplings $g_{HZZ}/g_{HWW}$ with a precision of 2.6 %.
Contents

1 Introduction 3

2 Event Generation, Simulation and Reconstruction 3

3 350 GeV HZ Recoil Analysis 4
   3.1 Analysis Samples ........................................... 4
   3.2 Lepton Identification ...................................... 5
   3.3 Background Rejection ...................................... 6
   3.4 Fit Procedure ............................................. 9
   3.5 Results .................................................. 11
      3.5.1 $\mu\mu X$ Channel .................................... 11
      3.5.2 $eeX$ Channel ........................................ 12

4 500 GeV HZ Recoil Analysis 15
   4.1 Analysis Samples ........................................... 15
   4.2 Results .................................................. 16
      4.2.1 $\mu\mu X$ Channel .................................... 16
      4.2.2 $eeX$ Channel ........................................ 18

5 500 GeV HZqq Analysis 19
   5.1 Analysis Samples ........................................... 19
   5.2 Jet Reconstruction and $b$-Tagging ......................... 19
   5.3 Kinematic Fit ............................................. 20
   5.4 Background Rejection ..................................... 22
   5.5 Fit Procedure ............................................. 26
   5.6 Results .................................................. 26

6 500 GeV $H\nu\nu$ Analysis 28
   6.1 Analysis Samples ........................................... 28
   6.2 Background Rejection ..................................... 29
   6.3 Fit Procedure ............................................. 33
   6.4 Results .................................................. 34
   6.5 Higgsstrahlung and WW-Fusion: Relative Normalisations 34

7 Summary 37
1 Introduction

CLIC is a proposed linear collider designed to perform electron-positron collisions at centre-of-mass energies up to 3 TeV, with a design luminosity of $5.9 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$. The construction of CLIC would most likely be performed in discrete stages and a strategic document, “The CLIC Programme: towards a staged $e^+e^-$ Linear Collider exploring the Terascale, CLIC Conceptual Design Report” [1], describes an example of such a staged approach. Beginning with a centre-of-mass energy of 500 GeV, the strategic document envisages an intermediate stage with a centre-of-mass energy of 1.4 TeV, before reaching the design energy of 3 TeV.

This LCD-Note investigates the precisions with which measurements of the Higgs mass and cross-sections can be made at a 500 GeV, first stage, CLIC machine. The Higgs candidate recently observed at the LHC has a mass of approximately 125 GeV and should provide a rich spectrum of production and decay modes that can be measured with high precision at CLIC. The studies presented in this document began before this recent discovery, so a Higgs mass of $M_H = 120\text{GeV}$ was assumed for all analyses.

Using the 500 GeV CLIC machine to run at centre-of-mass energies of 350 GeV and 500 GeV allows detailed examination of Higgs production via Higgsstrahlung and $WW$-fusion processes. In particular, measurement of the $Z$ recoil in $HZ$ production can provide a model-independent mechanism for measuring cross-sections and the Higgs mass. This model-independent analysis is unique to a lepton collider and requires precise knowledge of the initial state, before $HZ$ production. The $Z$ can be reconstructed most accurately when it decays to two leptons, so $Z \rightarrow \mu\mu$ and $Z \rightarrow ee$ channels are investigated. The ultimate precision of the measurements will depend on the lepton momentum resolution and the effects of Initial State Radiation, which alters the initial state.

The measurement precision offered by the model-independent analyses will decrease upon moving from 350 GeV to 500 GeV. This is due to reduced lepton momentum resolution at higher energies, reduction of the $HZ$ cross-section and overall less favourable background conditions. At 500 GeV, the mass and cross-section measurements can be studied in a model-dependent manner, assuming Standard Model Higgs decays. The Higgs can be reconstructed from its final-state two-quark decay, predominantly $H \rightarrow b\bar{b}$. Two final states are investigated: the four-jet final state, originating from the process $e^+e^- \rightarrow HZ \rightarrow q\bar{q}q\bar{q}$, and the two-jet neutrino final state, which receives contributions from both Higgsstrahlung $e^+e^- \rightarrow HZ \rightarrow q\bar{q}\nu\bar{\nu}$ and $WW$-fusion $e^+e^- \rightarrow Hv\bar{\nu} \rightarrow q\bar{q}v\bar{\nu}$. The luminosity assumed for all studies is $500 \text{fb}^{-1}$.

2 Event Generation, Simulation and Reconstruction

The physics events used for the studies presented in this document were produced in the same manner as those used for the CLIC Conceptual Design Report (CDR) [2]. The Monte Carlo event samples were generated using the WHIZARD [3, 4] program, assuming zero polarization of the electron and positron beams. Initial and Final State Radiation (ISR and FSR) were enabled during event generation and the expected CLIC luminosity spectra were used [2, 5]. Different luminosity spectra were required for the 500 GeV CLIC machine when running at centre-of-mass energies of 350 GeV and 500 GeV. Parton showering, hadronisation and fragmentation
were performed using PYTHIA \[6\].

Generated events were passed through the detector simulation program MOKKA \[7\], which implements the detector model CLIC_ILD_CDR500\[8\] in the GEANT4 \[9, 10\] framework. The QGSP_BERT physics list was used to simulate the detailed development of hadronic showers in the detector. The initial state is of great importance to the model-independent analyses examined in this document and, at CLIC, the beam-crossing angle is 20 mrad. This crossing angle was introduced in the simulation stage, following the generation of head-on collisions.

The MARLIN \[11\] framework was used for the digitisation, reconstruction and analysis of the simulated events. Following digitisation of the simulated hits and full reconstruction of tracks in the inner detector trackers, a particle flow reconstruction was performed by PANDORA-PFA \[12, 13\]. This program uses a large number of pattern-recognition algorithms to trace the paths of individual particles through the detector. The energy and momentum for each particle can then be extracted from the detector subsystem in which the measurements are likely to be most accurate. In this way, four-vectors were reconstructed for all visible particles in the event. Particle identification was also performed, and photons and charged leptons were flagged with high efficiency \[2\].

Background from $\gamma\gamma \rightarrow$ hadrons was added to all samples, with hits from simulated background events added to those from the underlying simulated physics event prior to digitisation \[14\]. Background was added to each of the 300 bunch crossings in a time window around the physics event. The number of events added to each bunch crossing was drawn from a Poisson distribution, with a mean of 0.3 events per bunch crossing at 500 GeV and 0.0464 events per bunch crossing at 350 GeV. This background could be efficiently removed by applying a combination of timing and momentum cuts to the reconstructed particle collection created by the particle flow reconstruction \[2\]. Three levels of timing cuts were available for use in physics analyses, labelled DEFAULT, LOOSE and TIGHT.

3 350 GeV HZ Recoil Analysis

3.1 Analysis Samples

The generated signal and background samples for the model-independent recoil analysis at 350 GeV are detailed in Table 1. Two separate analyses were performed, examining the cases where the $Z$ decays to muons and to electrons. The signal samples are characterised by a pair of oppositely-charged high $p_T$ leptons, which have an invariant mass equal to that of the $Z$. Possible background processes are those with a similar pair of leptons in the final state. These background final states can be divided into two-fermion samples, and four-fermion samples. The two-fermion samples ($ee, \mu\mu$ and $\tau\tau$) were simple to remove, with all events in test-samples failing the initial set of selection cuts outlined in Section 3.3. These samples were therefore neglected throughout the remainder of the model-independent analyses.

The irreducible backgrounds for the recoil analyses are the four-fermion backgrounds, which have final states consisting of a pair of oppositely-charged leptons (of the relevant flavour) and any other possible fermion pair. For both the muon and electron channels (henceforth referred to as simply $\mu\mu X$ and $eeX$), the background has a total cross-section approximately 1000 times
greater than the signal. This presented a challenge not only for background rejection at the analysis stage, but also for the initial event generation.

Given the small signal cross-sections, excess events were generated, simulated and reconstructed. A safety factor of approximately 50 was applied, and events were then weighted to the correct normalisations throughout the analysis. This procedure helped remove the effects of statistical fluctuations in the signal sample. For the background samples, however, cuts at the generator level were required. The aim was simply to avoid simulating and reconstructing background events that would only proceed to fail the most trivial analysis cuts. The following cuts were applied to the background samples:

- $p_T^{\ell+\ell-} > 10$ GeV. The Higgsstrahlung process will tend to produce a high $p_T$ $Z$, and hence a high $p_T$ lepton pair.

- $|\cos(\theta_{\ell+\ell-})| < 0.95$. The cross-sections for the Higgsstrahlung process are expected to decrease towards the forward/backward regions, whilst these directions are favoured for processes involving the production of $W$ or $Z$ pairs.

### 3.2 Lepton Identification

The reconstructed particles available at the analysis stage have already been processed by particle identification algorithms, during the particle flow reconstruction. For the model-independent recoil analysis, the signal selection procedure consisted of an examination of the particles available in the DEFAULT collection of PANDORAPFA particle flow objects (PFOs). For the lepton flavour of interest (e.g. muons for the $\mu\muX$ channel), separate lists of negatively and positively charged leptons were created. If both lists were populated, the event would be flagged as a signal candidate for further analysis. If either list contained multiple entries, all possible di-lepton combinations were considered and the lepton pair with invariant mass closest to the $Z$ mass chosen.

The selection of charged leptons was implemented in a MARLIN processor. Following selection of the lepton pair, quantities such as the $p_T$ and polar angle of the lepton pair were calculated, as were properties such as the acollinearity and acoplanarity of the leptons. This information was stored in a ROOT [15] tree, dubbed a physics analysis ntuple (PAN). The efficiency of the lepton-pair selection for the different samples is shown in Table 2.
3.3 Background Rejection

The process of background rejection for the model-independent recoil analyses can be divided into two parts. Firstly, hard selection cuts were applied to event properties that have significantly different distributions for the signal and background samples. In the second part of the selection procedure, the TMVA [16] package was used to perform a multivariate analysis, examining multiple properties in order to classify each event as signal-like or background-like. A cut on the output from the multivariate analysis completed the selection procedure, at which point the signal-rich distributions could be examined to extract Higgs mass and cross-section measurements.

Note that all figures displayed throughout this section refer to the $\mu\mu X$ channel. The process of background rejection, however, was very similar for both channels, with application of the same hard selection cuts and the same properties used as inputs for the multivariate analysis.

For the first stage of the selection procedure, the distributions examined are displayed in Figure 1. The properties investigated were the mass of the selected lepton pair, the $p_T$ of the lepton pair, and the “recoil mass”. The recoil mass was calculated by subtracting the four-vector of the lepton pair from the initial state four-vector. The result should yield the four-vector of the Higgs and this is the key to the model-independent analyses: knowledge of the initial state, followed by precision measurement of the $Z$ allows precision inferences to be made about the Higgs. The recoil mass is the invariant mass calculated using the inferred Higgs four-vector.

The distributions shown in Figure 1 motivated the following simple selection cuts:

- $40 \text{ GeV} < M_{\ell^+\ell^-} < 120 \text{ GeV}$
- $95 \text{ GeV} < M_{\text{recoil}} < 290 \text{ GeV}$
- $p_T,\ell^+\ell^- > 60 \text{ GeV}$

After the application of the selection cuts, the distributions of the $p_T$ and mass of the lepton pair were used as inputs to a multivariate analysis. Further inputs to the multivariate analysis are shown in Figure 2 and include the cosine of the polar angle of the lepton pair (calculated from the vector sum of their individual three-momenta) and the acollinearity and acoplanarity of the leptons. The acollinearity is the opening angle of the two leptons, whilst the acoplanarity is the projection of the opening angle into the $r - \phi$ plane. The acollinearity and acoplanarity are calculated as follows:

\begin{align*}
\text{acol} &= \cos^{-1}(\frac{p_1 \cdot p_2}{|p_1||p_2|}) \\
\text{acop} &= \cos^{-1}(\frac{p_{T1} \cdot p_{T2}}{|p_{T1}||p_{T2}|})
\end{align*}

where $p_i$ is the three-momentum of lepton $i$. 

### Table 2: Lepton selection efficiencies for generated signal and background samples.

<table>
<thead>
<tr>
<th>WHIZARD Process Id.</th>
<th>hzmumu</th>
<th>e2e2ff</th>
<th>hzee</th>
<th>e1e1ff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>98%</td>
<td>80%</td>
<td>97%</td>
<td>73%</td>
</tr>
</tbody>
</table>

Note that all figures displayed throughout this section refer to the $\mu\mu X$ channel. The process of background rejection, however, was very similar for both channels, with application of the same hard selection cuts and the same properties used as inputs for the multivariate analysis.
Figure 1: (a) Invariant mass of identified lepton pair. (b) Recoil mass, calculated using measurement of lepton pair and knowledge of the initial state. (c) $p_T$ of the lepton pair.

The final input to the multivariate analysis was calculated as the difference between the $p_T$ of the lepton pair and the $p_T$ of the highest energy photon in the event, presumed to be created by FSR. This quantity is expected to peak at zero for the background samples, where the di-lepton $p_T$ must be balanced by the $p_T$ of the emitted FSR photon. In the case of the signal, the $HZ$ recoil reaction will ensure that the di-lepton pair has significant $p_T$.

$$\Delta p_{T,\text{balance}} = p_{T,\ell^+\ell^-} - p_{T,\gamma}$$  \hspace{1cm} (3)

Following tests of a number of the multivariate techniques offered by TMVA, the Boosted Decision Tree (BDT) was found to offer the optimal signal purity for a given signal efficiency [17] and was chosen for use in the analyses. Figure 3(a) shows the distribution of the BDT output value for signal and background samples, whilst Figure 3(b) shows the signal purities and efficiencies obtained by cutting on the BDT output value. The efficiency is defined as the number
Figure 2: (a) Acollinearity of the lepton pair. (b) Acoplanarity of the lepton pair. (c) Cosine of the polar angle of the lepton pair. (d) Difference between the $p_T$ of the lepton pair and that measured for the highest energy photon in the event; presumed to be FSR.

Figure 3(c) displays the variation of the signal significance as a function of the cut on the BDT value. The significance is defined as the number of selected signal events, divided by the square root of the total number of selected events. Figure 3(d) shows the relationship between the selected signal significance and the signal efficiency. The BDT cut selected for use in the final analyses was that maximising the signal significance. Figure 4 shows the resulting di-lepton mass and recoil mass distributions for the final selected events.
3.4 Fit Procedure

Following the background rejection procedures, described in Section 3.3, the recoil mass distribution was examined in order to extract measurements of the Higgs mass and the $HZ$ production cross-section. The approach was to model the shapes of the signal and background components of the recoil mass distribution, then to use these shapes in a fit to the observed distribution. The fit parameters were the Higgs mass, the signal normalisation and the background normalisation. The shape of the signal distribution was modelled using the technique of SIMPLIFIED KERNEL.
Figure 4: (a) Invariant mass of the lepton pair for 350 GeV events passing the selection procedure. (b) The recoil mass, calculated for events passing the selection procedure.

**Estimation.** The signal distribution was approximated using the following functions:

\[ F_S(x) = \frac{1}{N} \sum_{j=1}^{m} n_j G(x; t_j; h_j) \]

\[ h_j = \left( \frac{4}{3} \right)^{1/5} N^{-1/5} \Delta x \sqrt{\frac{N}{n_j}} \]

where the summation is over the \( m \) bins in a high statistics reference sample of the signal recoil mass distribution. \( N \) is the total number of events in the signal reference sample, whilst \( n_j \) is the number of events assigned to bin \( j \) and \( \Delta x \) is the width of bins used (here assumed constant). The signal distribution was constructed via the summation of weighted Gaussian distributions, \( G \), each of which have mean \( \mu = t_j \) (where \( t_j \) is the centre of bin \( j \) in the reference sample) and width \( \sigma = h_j \).

The Simplified Kernel Estimation approach was used only to model the shape of the signal distribution. The normalisation of the signal distribution could be varied in the fit, providing sensitivity to the number of signal events. Use of the transformation \( x \rightarrow x' = x - M_H \) then allowed sensitivity to the value of the Higgs mass. Using Simplified Kernel Estimation removed the difficulty of finding parent functions to describe the signal shape, but was only possible because high statistics samples were available to provide reference distributions. Figure 5(a) shows the reference distribution for the signal recoil mass distribution, together with the resulting signal approximation.

For the background component of the recoil mass distribution, the available statistics were only sufficient to represent a single 500 fb\(^{-1} \) sample. Simplified Kernel Estimation could not be used, so the approach was to simply fit a fourth-order polynomial function. The single sample background and the fitted polynomial are displayed in Figure 5(b). The polynomial provides a reasonable description of the background, although future studies may benefit from higher statis-
tics background samples to improve understanding of background events that pass the selection procedure. Only the shape of the background distribution was considered in the fit, providing sensitivity to the number of background events.

The fit itself was performed via comparison of the total observed selected recoil mass distribution, with predicted distributions made for different combinations of the fit parameters: the Higgs mass, the number of signal events and the number of background events. The recoil mass distributions were divided into bins and a bin-by-bin comparison allowed calculation of a negative log likelihood, using the following likelihood function:

$$-\ln \mathcal{L} = \sum_{j=1}^{n_{\text{bins}}} n_{\text{pred},j} - n_{\text{obs},j} \ln(n_{\text{pred},j})$$

(6)

where $n_{\text{bins}}$ is the number of bins for the recoil mass distribution, $n_{\text{obs},j}$ is the number of events observed in bin $j$ and $n_{\text{pred},j}$ is the number of events predicted to be in bin $j$ for specific values of the fit parameters. Variation of the fit parameters, controlled by the minimisation program MINUIT [18], allowed the most likely values of the parameters to be identified; the values for which the negative log likelihood value is minimal. The accuracy with which each parameter can be measured was identified by examining the variation of the negative log likelihood value as each parameter was varied individually. The error assigned to a given parameter represents the change in the parameter value required for the negative log likelihood value to increase, from its best-fit value, by $\Delta(-\ln \mathcal{L}) = +0.5$ (i.e. $\Delta \chi^2 = +1.0$).

3.5 Results

3.5.1 $\mu\mu X$ Channel

In order to assess the accuracy with which the Higgs mass and the number of signal events (and hence $HZ$ production cross-section) can be measured, 1000 test “data” samples were pro-
duced. Each sample was created by adding the high statistics selected signal sample (scaled to the correct normalisation) to the smooth fourth-order polynomial background, then fluctuating the resulting smooth distribution to create a representative 500 fb$^{-1}$ data sample. Each of the 1000 test samples created in this way was used as the input to the fit procedure described in Section 3.4. The distribution of best-fit values could then be examined, as could distributions of the reported precisions for the fitted parameters.

Figure 6(a) displays the results of fitting a typical test sample for the $\mu\mu X$ channel. The points show the recoil mass distribution for the fluctuated test sample, whilst separate distributions are displayed for the fitted signal and background distributions. Repeating this procedure for 1000 test samples allowed the distributions in Figure 6(b) to be constructed. This figure shows the distribution of the reported best-fit Higgs mass for the 1000 samples, the distribution of the reported Higgs mass precisions and the distribution of the reported precisions for the number of signal events.

The results for the 1000 $\mu\mu X$ test samples are summarised in Table 3. It should be noted that the RMS of the fitted Higgs mass distributions is in close agreement with the mean of the precisions for the Higgs mass, as reported by MINUIT. The Higgs mass was determined with a precision of 133 MeV. The number of signal events, and so the $HZ$ production cross-section, was determined with a precision of 4.9%.

<table>
<thead>
<tr>
<th>350 GeV Recoil</th>
<th>Mean</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu\mu X$</td>
<td>$M_H$</td>
<td>119.950.4 MeV</td>
</tr>
<tr>
<td></td>
<td>$\Delta M_H$</td>
<td>133.3 MeV</td>
</tr>
<tr>
<td></td>
<td>$\Delta N_{sig}$</td>
<td>4.91 %</td>
</tr>
</tbody>
</table>

Table 3: Summary of results and measurement precisions obtained for $\mu\mu X$ channel at 350 GeV.

### 3.5.2 eeX Channel

The procedures for analysing the $eeX$ channel were very similar to those for the $\mu\mu X$ channel. However, the $eeX$ channel introduced a new complication: Bremsstrahlung of the final state electrons. Bremsstrahlung reduces the number of events in the peak of the recoil mass distribution and increases the population of the tail of the distribution. This harms measurements of both the Higgs mass and the $HZ$ production cross-section. The radiation from the electrons also impacts the distributions used as inputs to the multivariate selection procedures, increasing the difficulty of the background rejection.

In an effort to improve the precision of measurements from the $eeX$ channel, an attempt was made to recover Bremsstrahlung photons. Following the ideas implemented in [19], events were examined in order to find photons that were collinear with the reconstructed electrons. The four-momenta of the electrons were then combined with those of the candidate Bremsstrahlung photons. Figure 7 displays the impact of this Bremsstrahlung recovery procedure on the distribution of the invariant mass of the lepton pair and the distribution of the recoil mass. The Bremsstrahlung recovery significantly increased the number of events in the peak of the dis-
Figure 6: (a) Result of fitting a typical $\mu\mu X$ test sample at 350 GeV. (b) Distributions of best-fit Higgs masses, mass precisions and signal event precisions for 1000 test samples.

distributions, but also increased the width of the peak. This increase in width was expected, due to use of calorimeter information for the lepton momentum reconstruction, plus incorrect or incomplete Bremsstrahlung photon identification.

Figure 8 displays the results of fitting typical test samples for the $eeX$ channel, with and without Bremsstrahlung recovery. The measurement precisions from this channel are summarised in Table 4. The Higgs mass was determined with a precision of 300 MeV, with the measurement accuracy falling to 394 MeV with the use of Bremsstrahlung recovery. The $HZ$ production cross-section was determined with a precision of 8.1 %, with the use of Bremsstrahlung recovery improving the precision to 7.9 %.
Figure 7: (a) Invariant mass of the lepton pair before (left) and after (right) Bremsstrahlung recovery at 350 GeV. (b) The recoil mass, calculated before (left) and after (right) Bremsstrahlung recovery.

Table 4: Summary of results and measurement precisions obtained for $eeX$ channel at 350 GeV.
Figure 8: Result of fitting typical eeX test samples at 350 GeV, (a) before and (b) after Bremsstrahlung recovery.

4 500 GeV HZ Recoil Analysis

4.1 Analysis Samples

The analyses described in Section 3 were repeated for a 500 GeV CLIC machine running at a centre-of-mass energy of 500 GeV. The generated signal and background samples for these analyses are detailed in Table 5. For both $\mu\mu X$ and $eeX$ channels, the signal to background ratios are much lower at a centre-of-mass energy of 500 GeV, than at 350 GeV. When combined with the reduced lepton momentum resolution at higher energies, and the increased radiative effects at 500 GeV, the expectation is that the precision of Higgs mass and $HZ$ production cross-section measurements achieved via model-independent recoil analyses will be significantly worse at 500 GeV, compared to 350 GeV. This expectation was strengthened by a first comparison of the shape of the recoil mass distributions for $\mu\mu X$ signal samples at the two energies, displayed in Figure 9.

<table>
<thead>
<tr>
<th>WHIZARD Process Id.</th>
<th>Cross-section / fb</th>
<th>Cross-section (gen. cuts) / fb</th>
<th>Events / 500fb$^{-1}$</th>
<th>Available events</th>
</tr>
</thead>
<tbody>
<tr>
<td>hzmumu (signal)</td>
<td>2.451</td>
<td>2.451</td>
<td>1225</td>
<td>59381</td>
</tr>
<tr>
<td>e2e2 f/f</td>
<td>4.389</td>
<td>833.6</td>
<td>416 800</td>
<td>416 440</td>
</tr>
<tr>
<td>hzee (signal)</td>
<td>2.450</td>
<td>1225</td>
<td>993 000</td>
<td>927 000</td>
</tr>
<tr>
<td>e1e1 f/f</td>
<td>5.854</td>
<td>1854</td>
<td>60 379</td>
<td>51 000</td>
</tr>
</tbody>
</table>

Table 5: Generated event samples for the model-independent recoil analysis at 500 GeV.

The analyses at 500 GeV were performed in the same manner as those at 350 GeV, described in
Section 3. The same generator-level cuts were placed on the four-fermion background samples, and the process of background rejection was again split into hard selection cuts, followed by input of the same distributions to a BDT multivariate analysis. The hard selection cuts were adjusted to those shown below, following use of the TMVA cut optimisation package:

- $30 \text{ GeV} < M_{\ell^+\ell^-} < 120 \text{ GeV}$
- $0 \text{ GeV} < M_{\text{recoil}} < 380 \text{ GeV}$
- $p_T,\ell^+\ell^- > 65 \text{ GeV}$

4.2 Results

4.2.1 $\mu \mu X$ Channel

The fit procedure for the 500 GeV model-independent recoil analyses was identical to that used at 350 GeV, described in Section 3. Simplified Kernel Estimation was again used to model the shape of the signal distribution, whilst a fourth-order polynomial was used to model the shape of the background. By fluctuating the sum of a high statistics signal sample (scaled to the correct normalisation) and the smooth background fit, a set of 1000 representative $500 \text{ fb}^{-1}$ data samples was created.

Figure 10(a) displays the results of fitting a typical test sample for the $\mu \mu X$ channel at 500 GeV. Figure 10(b) shows the distributions of best-fit Higgs masses, Higgs mass precisions and signal event precisions reported for the 1000 samples. The measurement precisions are summarised in Table 6. The Higgs mass was determined with a precision of 646 MeV, whilst the $HZ$ production cross-section was determined with a precision of 11%. As expected, these measurement precisions are significantly worse than those obtained from the study of the $\mu \mu X$ channel at 350 GeV.
Figure 10: (a) Result of fitting a typical $\mu \mu X$ test sample at 500 GeV. (b) Distributions of best-fit Higgs masses, mass precisions and signal event precisions for 1000 test samples.

<table>
<thead>
<tr>
<th>500 GeV Recoil</th>
<th>Mean</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_H$</td>
<td>119.866.8 MeV</td>
<td>689.2 MeV</td>
</tr>
<tr>
<td>$\Delta M_H$</td>
<td>645.6 MeV</td>
<td>87.3 MeV</td>
</tr>
<tr>
<td>$\Delta N_{\text{sig}}$</td>
<td>11.03 %</td>
<td>1.08 %</td>
</tr>
</tbody>
</table>

Table 6: Summary of results and measurement precisions obtained for $\mu \mu X$ channel at 500 GeV.
4.2.2 eeX Channel

Figure 11 displays the results of fitting typical test samples for the eeX channel at 500 GeV, with and without Bremsstrahlung recovery. The measurement precisions from this channel are summarised in Table 7. The Higgs mass was determined with a precision of 1323 MeV, with the measurement accuracy falling to 2057 MeV with the use of Bremsstrahlung recovery. The HZ production cross-section was determined with a precision of 16.7 %, with the use of Bremsstrahlung recovery improving the precision to 15.9 %. As expected, these measurement precisions are significantly worse than those obtained from the study of the eeX channel at 350 GeV.

![Figure 11: Result of fitting typical eeX test samples at 500 GeV, (a) before and (b) after Bremsstrahlung recovery.](image)

<table>
<thead>
<tr>
<th>500 GeV Recoil</th>
<th>Mean</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>eeX</td>
<td>$M_H$</td>
<td>119793.7 MeV</td>
</tr>
<tr>
<td></td>
<td>$\Delta M_H$</td>
<td>1322.9 MeV</td>
</tr>
<tr>
<td></td>
<td>$\Delta N_{sig}$</td>
<td>16.68 %</td>
</tr>
<tr>
<td>eeX Bremsstrahlung recovery</td>
<td>$M_H$</td>
<td>120041.0 MeV</td>
</tr>
<tr>
<td></td>
<td>$\Delta M_H$</td>
<td>2057.1 MeV</td>
</tr>
<tr>
<td></td>
<td>$\Delta N_{sig}$</td>
<td>15.85 %</td>
</tr>
</tbody>
</table>

Table 7: Summary of results and measurement precisions obtained for eeX channel at 500 GeV.
5 500 GeV HZqq Analysis

5.1 Analysis Samples

In addition to the model-independent studies, a 500 GeV CLIC machine also offers excellent opportunities for measuring the Higgs mass and cross-sections in a model-dependent manner. Standard Model Higgs decays were assumed, and opportunities for reconstructing the Higgs from its final-state two-quark decay (predominantly $H \rightarrow b \bar{b}$) were considered. This section describes an analysis of the four-jet final state, originating from the process $e^+ e^- \rightarrow HZ \rightarrow q\bar{q}qq$. The generated signal and background samples for this study are displayed in Table 8.

The signal sample consists exclusively of $HZ$ production, followed by hadronic decay of the $Z$ and Higgs bosons. The final state should therefore consist of four jets, with the two jets from the Higgs most likely to be created by $b$-quarks. The analysis approach was to use a jet reconstruction algorithm to force each event into four jets, then assign pairs of jets to the $Z$ and Higgs. Possible backgrounds are processes providing final states that can plausibly be reconstructed as four jets. The $qqqq$ final-state is therefore significant, as are any final states that actually only contain only two quarks: $qq$ and $qq\nu\nu$.

An excess of events was generated for the signal sample, providing a sample free from statistical fluctuations. Significant numbers of background events were required, but there were no generator-level cuts that could be identified as completely “safe”, without potential impact on the study. For the background samples, as many events as possible were generated and weights greater than unity were applied to each event in order to scale the samples to the correct normalisation for a luminosity of 500 fb$^{-1}$.

<table>
<thead>
<tr>
<th>WHIZARD Process Id.</th>
<th>Cross-section / fb</th>
<th>Events / 500 fb$^{-1}$</th>
<th>Events available</th>
</tr>
</thead>
<tbody>
<tr>
<td>$hzqq$ (signal)</td>
<td>34.353</td>
<td>17,177</td>
<td>91,708</td>
</tr>
<tr>
<td>$qqqq$</td>
<td>4050.395</td>
<td>2025 198</td>
<td>548 449</td>
</tr>
<tr>
<td>$qq$</td>
<td>13,212.307</td>
<td>6606 154</td>
<td>598 647</td>
</tr>
<tr>
<td>$qq\nu\nu$</td>
<td>353.343</td>
<td>176 672</td>
<td>292 123</td>
</tr>
</tbody>
</table>

Table 8: Generated event samples for model-dependent four-jet final state analysis at 500 GeV.

5.2 Jet Reconstruction and $b$-Tagging

Jet reconstruction was performed by the FASTJET library [20]. Following the studies documented in [21], the $k_t$ algorithm was selected, with a $\Delta \eta - \Delta \phi$ metric. The jet reconstruction was performed in exclusive mode, with each event forced into four jets. Three different values for the jet cone size were considered: $R = 0.7, 1.0$ and $1.3$. The analysis was completed for each $R$-value and the value providing the best measurement precisions (particularly for the $HZ$ production cross-section) was identified. Figure 12(a) shows the distribution of the total reconstructed jet energy sum for the signal samples with each of the jet cone sizes.

The expectation for the Standard Model Higgs is that its dominant two-quark final state will be $bb$. For this reason, it is of great help to the analysis to tag the jets likely associated with
Figure 12: (a) Total jet energy sum for HZqq signal samples, using the $k_t$ algorithm with jet cone sizes $R = 0.7$, 1.0 and 1.3. (b) Purities and efficiencies obtained for $b$-quark selection.

$b$-quarks. The LCFI FLAVOUR TAGGING [22] package was used to provide $b$-tagging information. Following reconstruction of vertex information, and other properties of the jets, the algorithm uses a neural network to provide $b$ and $c$ jet probabilities for each jet in an event. For these studies, the neural network was trained using approximately 10,000 signal events, with hadronic decays of the $Z$ providing light quarks for identification of $b$ and $c$ backgrounds. Figure 12(b) shows the purities and efficiencies obtained for the $b$-quark selection, when testing a statistically independent signal sample.

5.3 Kinematic Fit

Following the jet reconstruction and $b$-tagging, the next analysis step was the assignment of jets to the $Z$ and Higgs bosons. This assignment was performed using a kinematic fit, implemented as part of the MARLINFIT [23] package. The fit uses specified kinematic constraints to improve the precision of parameters of interest in the event. The constraints used in the kinematic fit were:

- Energy conservation, $\sum E_i = 500\text{GeV}$.
- Momentum conservation, $\sum_i (p_{x,i}, p_{y,i}, p_{z,i}) = (5, 0, 0)\text{GeV}$, where the non-zero initial $p_x$ value is due to the beam crossing-angle of 20 mrad.
- Mass constraint, the mass of one pair of the jets is constrained to be equal to the $Z$ mass.

The input to the kinematic fit was the four jets in an event. The six possible unique assignments of the four jets to the $Z$ and Higgs were considered, and the jet energies and momenta were varied in order to fulfil the constraints. A fit probability was calculated and the combination with the greatest probability was used to finalise the assignment of jets to $Z$ and Higgs. The variation of the jet parameters, and calculation of the fit probability used the following jet energy and angular resolutions, as determined in [21]:

\[ \sigma_{E} = 4.5\% \cdot E_{\text{jet}} \]
\[ \sigma_{\theta} = 0.27 \text{ rad} \cdot \text{GeV}/\sqrt{E_{\text{jet}}} \]
\[ \sigma_{\varphi} = 0.25 \text{ rad} \cdot \text{GeV}/\sqrt{E_{\text{jet}}} \]

After the assignment of jets to bosons had been finalised, it actually proved beneficial to remove the Z mass constraint and re-run the kinematic fit for the chosen combination with only the energy and momentum constraints. This completed the reconstruction of the jet four-vectors and allowed calculation of the Higgs and Z masses. A number of jet-shape variables, such as thrust and oblateness, could also be calculated using the reconstructed four-vectors. These shape variables later prove to be of use in the background rejection procedures. Figure 13 shows the effect of the kinematic fit process described above, compared to a simple procedure of taking the initial jet four-momenta and identifying the jet pair with invariant mass closest to the Z. The Figure shows reconstructed Higgs mass distributions obtained with a jet cone size of \( R = 1.0 \).

![Figure 13: Impact of the kinematic fit on the Higgs mass reconstruction, compared to results obtained using a simple assignment of jets to the Z and Higgs bosons.](image)

It is possible for events to fail the kinematic fit and these events were excluded from the subsequent analysis. Table 9 shows the fraction of the signal and background events passing the kinematic fit, for jet reconstruction with a cone size of \( R = 1.0 \). About 89\% of signal events passed the fit, whilst very few of the \( qq \) and \( qq\nu\nu \) events passed, aiding the process of background rejection.

<table>
<thead>
<tr>
<th>WHIZARD Process Id.</th>
<th>( hzqg )</th>
<th>( qqqq )</th>
<th>( qq )</th>
<th>( qq\nu\nu )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events passing kinematic fit</td>
<td>89%</td>
<td>78%</td>
<td>21%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Table 9: Fractions of signal and background events passing kinematic fit.
5.4 Background Rejection

The task for the background rejection procedure is illustrated in Figure 14. This figure shows the reconstructed Higgs mass distributions, following the kinematic fit, for the signal and combined background samples. The shapes of the distributions are also shown, with the background sample broken down into its different components. The distributions shown correspond to a jet cone size of $R = 1.0$, as do all distributions displayed in this section.

![Figure 14: (a) Reconstructed Higgs mass distributions, after kinematic fit, for signal and background. (b) Shape of fitted Higgs mass distributions for signal and background.](image)

The background rejection uses a TMVA Boosted Decision Tree, which combines information from eight different reconstructed quantities to provide a single variable for signal/background discrimination. In order to achieve optimal selection in the mass-region of interest, the signal/background distributions passed to the BDT consider only events with fitted Higgs masses in the region $95 \text{GeV} < M_H < 150 \text{GeV}$. Having been trained using only events from within this mass window, the BDT was then used to classify all input events.

The first two quantities used by the BDT were the highest $b$-tag and second-highest $b$-tag probabilities assigned to jets in the event. The distributions of these quantities are shown in Figure 15. Use of these variables reflects the dominant Higgs decay into $b \bar{b}$ and highlights the intent to select a sample rich in $H \rightarrow b \bar{b}$ final states.

Figure 16 displays the distributions of the remaining quantities used as inputs to the BDT selection process. These quantities are described below:

- $d_{3,4}$: This is a variable provided by the jet reconstruction algorithm. It provides a measure of the probability that the event should be reconstructed into four jets (as specified), rather than three.

- $M_{Z, \text{fit}} - M_{Z, \text{start}}$: This is a measure of how much the kinematic fit alters the invariant mass of the two jets assigned to the $Z$. 

22
Figure 15: (a) Highest and (b) second-highest $b$-tag values assigned to a jet in signal and background events. Events contain four jets, each of which has a $b$-tag probability.

- $h_T$: This is the sum of the transverse momenta of the four jets:
  \[ h_T = \sum_i^j |p_{T,i}| \]

- $S$: This jet shape variable is calculated by constructing the sphericity tensor, then manipulating the eigenvalues of this tensor:
  \[ S_{\alpha\beta} = \sum_i^{n_jets} p_i^\alpha p_i^\beta \sum_i^{n_jets} |p_i|^2 \]
  Eigenvalues: $\lambda_1 \geq \lambda_2 \geq \lambda_3$, Sphericity: $S = \frac{3}{2}(\lambda_1 + \lambda_2)$

- $T$: This jet shape variable is calculated by varying the thrust axis until the direction yielding the maximum projection of the jet three momenta is found:
  \[
  \text{Thrust: } T = \max_{|n| = 1} \frac{\sum_i^{n_jets} |n \cdot p_i|}{\sum_i^{n_jets} |p_i|}
  \]

- $O$: This jet shape variable is obtained by evaluating the difference between the thrust, when evaluated with respect to the major and minor axes (these axes, together with the thrust axis, form an orthogonal basis):
  \[
  T_{\text{major}} = \max_{|n| = 1, n \cdot v_1 = 0} \frac{\sum_i^{n_jets} |n \cdot p_i|}{\sum_i^{n_jets} |p_i|}
  \]
  Oblateness: $O = T_{\text{major}} - T_{\text{minor}}$
Figure 16: Distributions used by the TMVA Boosted Decision Tree to provide signal/background discrimination. These variables are fully described in Section 5.4.
Figure 17(a) shows the distribution of the BDT output value for signal and background samples, whilst Figure 17(b) shows the signal purities and efficiencies obtained by cutting on the BDT output value. The efficiency is defined with respect to the number of signal events passing the kinematic fit. Figure 17(c) displays the variation of the signal significance as a function of the cut on the BDT value. The significance is defined as the number of selected signal events, divided by the square root of the total number of selected events. The BDT cut selected for use in the final analysis was that maximising the signal significance. Figure 17(d) shows the resulting fitted Higgs mass for the final selected events.

![Graphs showing distribution, purities, efficiencies, and significance.](image)

Figure 17: (a) BDT output value for signal and background samples. (b) Selected signal purities and efficiencies obtained using BDT. (c) Selected signal significance obtained using BDT. (d) Fitted Higgs mass distribution for selected events.
5.5 Fit Procedure

The procedure for extracting measurements of the Higgs mass and HZ production cross-section was as described in Section 3.4. Simplified Kernel Estimation was used to approximate the shape of the fitted Higgs mass distribution for the signal sample. The background shape was approximated using a fourth-order polynomial. The fit was performed for the mass range $90\text{GeV} < M_H < 200\text{GeV}$. This included the full mass-region of interest, but discarded the high and low-mass regions that offer lower sensitivity to the fit parameters.

Figure 18 displays the high statistics reference sample for the signal and the Simplified Kernel Estimation signal approximation. The same figure also displays the fitted Higgs mass distribution for background events passing the selection procedure. The number of available background events is rather low, with events carrying normalisation weights greater than unity, so the selected background distribution displays statistical fluctuations. The fourth-order polynomial fit to the background is shown.

![Figure 18](image)

Figure 18: (a) Reference distribution for signal fitted Higgs mass distribution, with resulting signal approximation. (b) Single sample background and fitted polynomial.

5.6 Results

By fluctuating the sum of a high statistics signal sample (scaled to the correct normalisation) and the smooth background fit, a set of 1000 representative 500 fb$^{-1}$ data samples was created. Figure 19(a) displays the results of fitting a typical test sample (jet cone size $R = 1.0$), whilst Figure 19(b) shows the distributions of best-fit Higgs masses, Higgs mass precisions and signal event precisions for the 1000 test samples. The measurement precisions are summarised in Table 10. For $R = 1.0$, the Higgs mass was determined with a precision of 104 MeV, whilst the HZ production cross-section was determined with a precision of 1.6%. The selected signal sample comprised 99.1% $H \rightarrow b\bar{b}$ decays and 0.9% $H \rightarrow c\bar{c}$ decays.

It may be noted that the peak of the fitted Higgs mass signal distribution in Figure 19(a) is
Figure 19: (a) Result of fitting a typical $HZqq$ test sample at 500 GeV. (b) Distributions of best-fit Higgs masses, mass precisions and signal event precisions for 1000 test samples.

a little offset from the input Higgs mass of 120 GeV. This offset varies with the cone size used in the jet reconstruction. Provided that systematics can be controlled relative to the treatment of Monte Carlo samples, the offset is not a problem for the analysis. The fit proceeds by creating template distributions for different values of the fit parameters. The template distribution for a given Higgs mass value will account for any offsets or smearing that occurs during the reconstruction and selection procedures.
Table 10: Summary of results and measurement precisions obtained for a model-dependent analysis of $HZ$ four-jet final state at 500 GeV.

6 500 GeV $H\nu\nu$ Analysis

6.1 Analysis Samples

This section describes a second model dependent analysis at a centre of mass energy of 500 GeV: the two-jet two-neutrino final state is considered. This final state receives contributions from both Higgsstrahlung, $e^+e^- \rightarrow HZ \rightarrow q\bar{q}\nu\bar{\nu}$, and $WW$-fusion, $e^+e^- \rightarrow H\nu\bar{\nu} \rightarrow q\bar{q}\nu\bar{\nu}$. The generated signal and background samples for this study are displayed in Table 11.

Table 11: Generated event samples for two-jet neutrino final state analysis at 500 GeV.

The signal sample was an inclusive sample comprising both Higgsstrahlung and $WW$-fusion processes. Unlike the signal sample for the $HZqg$ study, in Section 5, the sample contained all Standard Model Higgs decays, not just $H \rightarrow q\bar{q}$. The analysis procedure was to use a jet reconstruction algorithm to force each event into two jets, then search for $H \rightarrow b\bar{b}$ final states. The relevant background processes were therefore those providing $qqqg$, $qqg$ and $qq\nu\nu$ final states. The aim was to select events from the inclusive signal sample with high purity and efficiency, even though the two processes contributing to the signal produce events with rather different topologies. Following event selection, the Higgs mass and the total inclusive production cross-section could be extracted. The angular distributions of selected signal events could then be examined in order to extract a measurement of the relative normalisations for Higgsstrahlung and $WW$-fusion. This provides a measurement of the ratio of couplings $g_{HZZ}/g_{HWW}$. 

<table>
<thead>
<tr>
<th>WHIZARD Process Id.</th>
<th>Cross-section / fb</th>
<th>Events / 500fb$^{-1}$</th>
<th>Events available</th>
</tr>
</thead>
<tbody>
<tr>
<td>$hnunu$ (inclusive signal)</td>
<td>80.693</td>
<td>40 347</td>
<td>287 712</td>
</tr>
<tr>
<td>$qqqg$</td>
<td>4 050.395</td>
<td>2 025 198</td>
<td>537 658</td>
</tr>
<tr>
<td>$qq$</td>
<td>13 212.307</td>
<td>6 606 154</td>
<td>587 417</td>
</tr>
<tr>
<td>$qq\nu\nu$</td>
<td>353.343</td>
<td>176 672</td>
<td>286 920</td>
</tr>
</tbody>
</table>

28
It is not possible to perfectly distinguish between Higgsstrahlung and WW-fusion processes in the inclusive signal sample. However, examination of the Monte Carlo particle collections allows separation with acceptable accuracy. Monte Carlo particles representing the two final-state neutrinos were identified and the invariant mass of the neutrino-pair was calculated. If this invariant mass was within 5 GeV of the Z mass, the event was declared to be Higgsstrahlung. The remaining events were flagged as WW-fusion. The $k_t$ jet reconstruction algorithm was used in exclusive mode, with each event forced into two jets. Figure 20(a) displays the reconstructed Higgs masses (the invariant masses of the two jets) for the inclusive signal sample and for those signal events identified as Higgsstrahlung. Separate distributions are shown for jet cone sizes of $R = 0.7, 1.0$ and 1.3. Figure 20(b) shows the $p_T$ distribution of the jet pair, with $R = 1.0$, for the inclusive signal sample and its Higgsstrahlung and WW-fusion components.

![Figure 20](image)

Figure 20: (a) The reconstructed Higgs mass for inclusive Higgsstrahlung and WW-fusion signal samples, obtained using the $k_t$ algorithm with three jet cone sizes. Distributions are also shown for an exclusive Higgsstrahlung sample. (b) The distribution of di-jet $p_T$ for the inclusive signal sample and its Higgsstrahlung and WW-fusion components.

### 6.2 Background Rejection

The task for the background rejection procedure is illustrated in Figure 21. This figure shows reconstructed Higgs mass distributions for the signal and combined background samples. The shapes of the distributions are also highlighted, with the background sample broken down into its different components. The distributions represent a jet cone size of $R = 1.0$, as do all distributions displayed in this section. As with the $HZqq$ study, described in Section 5, the $Hvv$ analysis was performed for three different $R$-values, $R = 0.7, 1.0$ and 1.3, and the value providing the best measurement precisions (particularly for the production cross-section) was identified.

The background rejection used a TMVA Boosted Decision Tree, which combined information from nine different reconstructed quantities to provide a single variable for signal/background discrimination. In order to achieve optimal selection in the mass-region of interest, the sig-
Figure 21: (a) Reconstructed Higgs mass distributions for signal and background. (b) Shape of reconstructed Higgs mass distributions for signal and background.

Signal/background distributions passed to the BDT considered only events with reconstructed Higgs masses in the region $95\text{GeV} < M_H < 150\text{GeV}$. Having been trained using only events from within this mass window, the BDT was then used to classify all input events. Distributions of the quantities used as inputs to the BDT are shown in Figures 22 and 23. These quantities are described below:

- **Highest $b$-tag**: The intention was to select a signal sample rich in $H \rightarrow b\bar{b}$ decays. As with the $HZqq$ study, $b$-tagging information was provided by the LCFI Flavour Tagging package. The neural network was trained using 10,000 signal and background events.

- **Second-highest $b$-tag**: The favoured signal events should contain $H \rightarrow b\bar{b}$ final states, so the signal sample should contain events for which both jets receive high $b$-tag values.

- **$d_{1,2}$**: This variable is provided by the jet reconstruction algorithm and indicates the probability that the event should be reconstructed into two jets, rather than a single jet.

- **$d_{2,3}$**: This variable indicates the probability that the event should be reconstructed into three jets, rather than two jets.

- **$p_T,qq$**: This is the $p_T$ of the di-jet system, which is expected to extend to much higher values for the signal and $qq\nu\nu$ background, than for the $qq$ and $qqqq$ backgrounds.

- **$\cos(\theta_{qq})$**: This is the cosine of the polar angle of the di-jet system. The $qq$ and $qqqq$ background samples will tend to favour the forward/backward regions, whilst the signal and $qq\nu\nu$ background are expected to have much flatter distributions.

- **Acollinearity**: This is the opening angle between the two jets, calculated via Equation 1.
Figure 22: Distributions used by the TMVA Boosted Decision Tree to provide signal/background discrimination. These variables are fully described in Section 6.2.

- **Acoplanarity**: This is the projection of the angle between the two jets into the $r - \phi$ plane, calculated as shown in Equation 2.

- **Missing mass**: A “missing” four-vector can be obtained by subtracting the four-vector of the di-jet system from the initial state four-vector. This variable is the invariant mass calculated from the missing four-vector.

Figure 24(a) shows the distribution of the BDT output value for signal and background samples, whilst Figure 24(b) shows the signal purities and efficiencies obtained. Figure 24(c) displays the variation of the signal significance as a function of the cut on the BDT value. The significance is defined as the number of selected signal events, divided by the square root of the total number of selected events. The BDT cut selected for use in the final analysis was that maximising the signal significance. Figure 24(d) shows the selected Higgs mass distribution.
Figure 23: Distributions used by the TMVA Boosted Decision Tree to provide signal/background discrimination. These variables are fully described in Section 6.2.
Figure 24: (a) BDT output value for signal and background samples. (b) Selected signal purities and efficiencies obtained using BDT. (c) Selected signal significance obtained using BDT. (d) Reconstructed Higgs mass distribution for selected events.

6.3 Fit Procedure

The procedure for extracting measurements of the Higgs mass and inclusive production cross-section (including both Higgsstrahlung and WW-fusion) was as described in Section 3.4. Simplified Kernel Estimation was used to approximate the shape of the reconstructed Higgs mass distribution for the signal sample. The shape of the background, which decreases as the reconstructed Higgs mass increases, was approximated using a Landau distribution. The fit was performed for the mass range 90GeV < M_H < 150GeV.

Figure 25 displays the high statistics reference sample for the signal and the Simplified Kernel Estimation signal approximation for a jet cone size of R = 1.0. The same figure also displays the reconstructed Higgs mass distribution for background events passing the selection procedure.
The number of available background events was rather low, with events carrying normalisation weights greater than unity, so the selected background distribution displays statistical fluctuations. The Landau distribution used to fit the background is shown.

![Figure 25: (a) Reference distribution for signal Higgs mass distribution, with resulting signal approximation. (b) Single sample background and fitted Landau distribution.](image)

6.4 Results

By fluctuating the sum of a high statistics signal sample (scaled to the correct normalisation) and the smooth background fit, a set of 1000 representative 500 fb$^{-1}$ data samples was created. Figure 26(a) displays the results of fitting a typical test sample (jet cone size $R = 1.0$), whilst Figure 26(b) shows the distributions of best-fit Higgs masses, Higgs mass precisions and signal event precisions for the 1000 test samples. The measurement precisions are summarised in Table 12. For $R = 1.0$, the Higgs mass was determined with a precision of 97 MeV, whilst the inclusive production cross-section was determined with a precision of 1.01 %. The selected signal sample comprised 90.7 % $b\bar{b}$ final states, 2.2 % $c\bar{c}$, 2.5 % $\tau\bar{\tau}$, 2.4 % $WW$ and 1.5 % $gg$. The remaining 0.7 % contained other Standard Model Higgs decays, most notably $H \rightarrow ZZ$.

6.5 Higgsstrahlung and $WW$-Fusion: Relative Normalisations

The final step in the analysis was to examine the $p_T$ distribution for selected events and to extract a measurement of the relative Higgsstrahlung and $WW$-fusion normalisations. The $p_T$ distribution was chosen because it offered clear discrimination (see Figure 20(b)) between the two components of the signal sample. Events were selected from signal and background samples using the selection procedure described in Section 6.2. However, because this selection was optimised for a mass-region of 90 – 150 GeV, an additional cut on the reconstructed Higgs mass was applied in order to prevent the large increase in background events at low reconstructed Higgs masses. The best value for this additional cut was identified as $M_H > 95$ GeV.
The method described in Section 6.1 was used to flag events in the selected signal sample as “true” Higgsstrahlung or WW-fusion.

The technique of Simplified Kernel Estimation was used to model separate shapes for the $p_T$ distributions of selected Higgsstrahlung events and selected WW-fusion events. The $p_T$ distribution for selected background events was modelled using a fourth-order polynomial, as illustrated in Figure 27. This approach was a little unsatisfactory, as the $p_T$ distributions are rather different for the individual background samples. However, given the small event samples available for the backgrounds, an approach such as Simplified Kernel Estimation could not be used. The fourth-order polynomial provided a reasonable description of the background $p_T$.
Table 12: Summary of results and measurement precisions obtained for a model-dependent analysis of the Higgsstrahlung and $WW$-fusion two-jet neutrino final state at 500 GeV.

<table>
<thead>
<tr>
<th>$500 \text{ GeV } HZ\nu\nu$</th>
<th>Mean</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R=0.7$</td>
<td>$M_H$</td>
<td>$120\ 034.3 \text{ MeV}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta M_H$</td>
<td>$95.9 \text{ MeV}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta N_{\text{sig}}$</td>
<td>$1.061 %$</td>
</tr>
<tr>
<td>$R=1.0$</td>
<td>$M_H$</td>
<td>$120\ 015.0 \text{ MeV}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta M_H$</td>
<td>$97.2 \text{ MeV}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta N_{\text{sig}}$</td>
<td>$1.010 %$</td>
</tr>
<tr>
<td>$R=1.3$</td>
<td>$M_H$</td>
<td>$120\ 006.5 \text{ MeV}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta M_H$</td>
<td>$110.8 \text{ MeV}$</td>
</tr>
<tr>
<td></td>
<td>$\Delta N_{\text{sig}}$</td>
<td>$1.057 %$</td>
</tr>
</tbody>
</table>

Figure 27: $p_T$ distribution for selected events from a rather low statistics background sample, normalised to $500 \text{ fb}^{-1}$. The polynomial approximation used in the fit is displayed.

distribution, but did not model the dip at low values caused by the sudden decrease in the number of $qq$ events.

The nominal ratio of the number of Higgsstrahlung to $WW$-fusion events in the original inclusive signal sample was 0.209. The selection efficiencies then differed for the two processes, so the application of the BDT selection cut altered the ratio to 0.225. The application of the reconstructed Higgs mass cut then increased the ratio to 0.256. A measurement of this final ratio was extracted by using MINUIT to control a simple fit to the $p_T$ distribution, where the shapes of the two signal components and the background were fixed. The fit parameters were the number of Higgsstrahlung events, the number of $WW$-fusion events and the number of background events.

Using the high statistics inclusive signal sample, together with the smooth background polynomial, then introducing fluctuations allowed the creation of 1000 representative $500 \text{ fb}^{-1}$ data
samples. Figure 28 displays the results of fitting a typical test sample \( k_t \) algorithm, jet cone size \( R = 1.0 \). The mean of the 1000 \( HZ/WW \) values was 0.2563, whilst the RMS of the distribution of \( HZ/WW \) values was 0.0131. This implies that the relative Higgsstrahlung and \( WW \)-Fusion normalisations can be determined with a precision of 5.1 \%, so the ratio of the couplings, \( g_{HZ}/g_{WW} \), can be measured with a precision of 2.6 \%.

![Graph](image)

**Figure 28:** Result of fitting a typical 500 fb\(^{-1}\) test sample, extracting the relative Higgsstrahlung and \( WW \)-fusion normalisations from the \( p_T \) distribution of selected events.

### 7 Summary

Higgs mass and cross-section measurements have been examined using samples of events representing the results of operating a 500 GeV CLIC machine at centre-of-mass energies of 350 GeV and 500 GeV. A Higgs mass of 120 GeV and a luminosity of 500 fb\(^{-1}\) were assumed. Measurement of the Z recoil in \( HZ \) production provided a model-independent mechanism for measuring the Higgs mass and \( HZ \) production cross-section. Two channels were investigated, at 350 GeV and 500 GeV, with the Z decaying to a pair of muons or a pair of electrons. The analysis procedure was to measure the di-lepton four-vector and subtract it from the initial state four-vector, providing an inferred four-vector for the Higgs.

At 350 GeV, the muon channel provided a Higgs mass measurement precision of 133 MeV and a \( HZ \) production cross-section measurement accuracy of 4.9 \%. For the electron channel, these measurement accuracies were 300 MeV and 8.1 \% without Bremsstrahlung recovery and
394 MeV and 7.9 % with Bremsstrahlung recovery. At 500 GeV, the measurement accuracy decreased due to reduced lepton momentum resolution and overall less favourable background conditions. The muon channel provided a Higgs mass measurement precision of 646 MeV and a HZ production cross-section measurement accuracy of 11 %. For the electron channel at 500 GeV, the measurement accuracies were 1323 MeV and 16.7 % without Bremsstrahlung recovery and 2057 MeV and 15.9 % with Bremsstrahlung recovery.

At 500 GeV, model-dependent studies were also performed, assuming Standard Model Higgs decays. The Higgs was reconstructed from its final-state two-quark decay, primarily $H \rightarrow b\bar{b}$. Two final states were investigated: the four-jet final state, originating from the Higgsstrahlung process $e^+e^- \rightarrow HZ \rightarrow q\bar{q}q\bar{q}$, and the two-jet neutrino final state, which receives contributions from both Higgsstrahlung $e^+e^- \rightarrow HZ \rightarrow q\bar{q}\nu\bar{\nu}$ and WW-fusion $e^+e^- \rightarrow H\nu\bar{\nu} \rightarrow q\bar{q}\nu\bar{\nu}$. With the $k_t$ algorithm and a jet cone size of $R = 1.0$, the four-jet final state provided a Higgs mass measurement precision of 104 MeV and a HZ production cross-section measurement accuracy of 1.6 %.

With the $k_t$ algorithm and a jet cone size of $R = 1.0$, the two-jet neutrino final state provided a Higgs mass measurement precision of 97 MeV and an inclusive Higgsstrahlung and WW-fusion production cross-section measurement accuracy of 1.01 %. Fitting the $p_T$ distribution allowed measurement of the relative Higgsstrahlung and WW-fusion normalisations with a precision of 5.1 %. This provides a measurement of the ratio of the couplings $g_{HZZ}/g_{HWW}$ with a precision of 2.6 %.

References


