CERN - European Organization for Nuclear Research

LCD-Note-2012-010

Measurement of $\sigma(e^+e^- \to H\nu\overline{\nu}) \times BR(H \to \tau\tau)$ at CLIC @ 1.4 TeV

A. Münnich[*]

[*] *CERN, CH-1211 Geneva 23, Switzerland*

March 26, 2013

Abstract

This detector benchmark study evaluates the statistical precision with which the $H \to \tau\tau$ branching ratio times cross section can be measured at CLIC running at $\sqrt{s} = 1.4$ TeV. Only the hadronic decays of τs are considered. Results for $M_H = 120$ GeV and 1.5 ab^{-1} of integrated luminosity are obtained using full detector simulation and including beam-induced backgrounds resulting in a statistical accuracy of cross section times branching ratio of 3.7%.

1 Introduction

The study reported in this note is carried out for a centre-of-mass energy of 1.4 TeV. The dominant production process for a Standard Model Higgs at this energy is $e^+e^- \rightarrow H\nu\bar{\nu}$ as shown in Figure 1. The goal is to measure the cross section for the Higgs decay into τs. This requires the reconstruction of τ leptons at high energies and in the presence of machine-induced backgrounds.

The investigated final state has two τ jets and missing energy. The corresponding branching ratios and cross section are listed in Table 1. Only hadronic decays of the τ are considered.
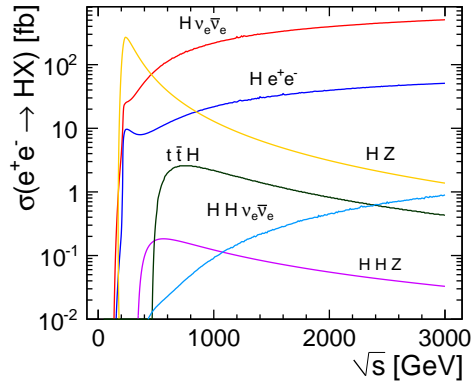

Figure 1: Cross sections for different Higgs production channels as a function of $\sqrt{s}$ for $M_{\mathrm{H}} = 120$ GeV.

Table 1: The Higgs mass, final state, branching ratios and cross section used in this study.

Process:	$e^+e^- \rightarrow H\nu\nu$
Particle masse:	$m_{\mathrm{H}} = 120$ GeV
Final state:	$\tau\tau\nu\nu$
Branching ratios:	$H \rightarrow \tau\tau$ (7%)
	$\tau \rightarrow$ hadrons (64.8%), $\tau \rightarrow e\nu\bar{\nu}$ (17.8%), $\tau \rightarrow \mu\nu\bar{\nu}$ (17.4%)
Cross section:	$\sigma = 248$ fb $\times$ 0.07 $= 17.5$ fb

2 Monte Carlo production

The physics events used for the study presented here were produced with the same procedures as used for the CLIC CDR [1]. Events were generated using the WHIZARD 1.95 [2] program. Initial and final state radiation (ISR and FSR) were enabled during the event generation. The luminosity spectrum expected at a 1.4 TeV CLIC machine was used during the event generation [3]. The hadronization of final state partons was simulated using Pythia [4]. The generated events were subsequently passed through the detector simulation program Mokka [5] which is based on the Geant4 [6] package. The CLIC_ILD [7] detector geometry model was used.

Events were overlayed with pileup from $\gamma\gamma \rightarrow$ hadrons interactions corresponding to 60 bunch crossings [8]. The reconstruction chain included an improved version [9] of the PandoraPFA [10] algorithm to reconstruct particle flow objects.

As the final state is equivalent to the one in the study of stau pair production [11] the data sets produced for the background channels can be resused. Additional backgrounds have been generated as the energy of the τs in this channel is lower than in the case of the stau analysis. Thus backgrounds arising via γe production are considered as well. Due to the high cross section of some of the background channels it is necessary to apply some cuts already before entering the full simulation. As some of the variables used in these cuts are sensitive to FSR the cuts can not be applied in WIZARD but only after the fragmentation in Pythia. This means the cuts are applied on the stdhep files before entering the Geant4 simulation. The following cuts were determined for the stau study and were applied for both τ candidates on almost all background channels:

- $10 < \theta_\tau < 170$ deg, where θ_τ is the polar angle of the τ candidate

- $p_\mathsf{T} > 20$ GeV

- $\Delta\phi < 178$ deg

- angle of τ system > 0.4 rad (23 deg)

- $40 <$ invariant mass of τ system < 650 GeV

For one background channel (ee $\rightarrow \tau\tau$) the full simulation was carried out on a data set with and without these cuts applied. A comparison of the number of events remaining in the data set after applying the pre-selection cuts confirmed that no bias is introduced by applying these cuts at this level. The cuts themselves are based on a study of signal and background channels at generator level.

An overview of all produced Monte Carlo (MC) samples is given in Table 2. The listed cross sections for the backgrounds are effective cross sections after applying the stdhep cuts.

3 Event reconstruction

The steps to reconstruct events with two τs from particle flow objects (PFOs) are described in this section. The presence of pileup from the process $\gamma\gamma \rightarrow$ hadrons increases the number of reconstructed PFOs in typical signal events by a factor 10 and the total visible momentum by

Table 2: Cross sections (after stdhep cuts) and integrated luminosities of the available Monte Carlo samples for the H → ττ study and the relevant backgrounds.The listed cross sections for the backgrounds are effective cross sections after applying the stdhep cuts. γe represents both γe$^+$ and γe$^-$ production.

Process	Cross section [fb]	Luminosity [ab^{-1}]
ee → Hνν → ττνν	17.5	4.4
ee → ττ	5.3	5.7
ee → ττνν	38.5	4.6
ee → eeττ	67.6	1.3
ee → μμττ	2.0	10.9
ee → qqνν	648.3	0.5
ee → qqee	225.9	0.4
γγ → ττ	404.3	0.7
γγ → ττνν	84.3	2.4
γγ → eeττ	2.6	5.9
γγ → μμττ	10.6	2.4
γγ → qqνν	0.63	15.0
γγ → qqee	12.2	1.6
γe → ττe	1795	0.5
γe → qqe	3283	0.4
γe → qqννe	20.97	1.7

a factor four. On the other hand, the background particles are emitted mostly in the forward direction.

A large fraction of the background can be rejected using combined timing and transverse momentum cuts [1]. The shape of the reconstructed τ energy distribution remains robust against the background. The τ reconstruction efficiency benefits from choosing the "tight selected PFOs".

The reconstruction of τ leptons is done with the `TauFinder` [12] which is essentially a seeded cone based jet clustering algorithm. Via steering parameters like the opening angle of the search and isolation cones the reconstruction algorithm can be optimized for a given τ signature. For this study a scan of the algorithm's parameters was carried out on events with quarks (ee → qqνν) to evaluate the fake rate, meaning mistaking a quark jet for a τ jet.

The following parameter set was chosen for the reconstruction of τ leptons:

- Minimum p_T to enter reconstruction: 1 GeV

- Minimum p_T for τ seed: 5 GeV

- Maximum for invariant mass of τ candidate: 2.5 GeV

- Opening angle of search cone: 0.07 rad

4

- Opening angle of isolation cone (relative to search cone): 0.3 rad

- Maximum energy allowed in isolation cone: 5.0 GeV

With these settings the efficiency to reconstruct a τ from the signal is 70%. The fake rate of 7% to mistake a quark for a τ is rather high but acceptable in this analysis as the background from quarks can be reduced significantly in the final event selection.

4 Event selection

The selection of H $\rightarrow \tau\tau$ signal events is performed in two steps. First, a cut-based pre-selection is applied. The remaining backgrounds are suppressed further using boosted decision trees in a second step. These two steps are described in the following two subsections.

4.1 Pre-Selection cuts

The pre-selection cuts used in the analysis have to be stronger than the cuts already applied at generator level to avoid a bias. As the generator level cuts were driven by the stau analysis the cut on p_T is slightly higher than one would choose if optimising for the lower energy τs in this study. Other cuts like the upper value of the invariant mass could be lowered considerably compared to the stau study. In addition a cut on the leptonic energy content in the τ candidate is applied to select only hadronic τ decays. Further cuts are applied to eliminate background dominated areas of the phase space.

- no leptons in τ

- $15 < \theta_\tau < 165$ deg

- p_T of $\tau > 25$ GeV

- $\Delta\phi < 177$ deg

- angle between the two τs > 0.5

- $45 <$ invariant mass of τ system < 130 GeV

- Thrust < 0.99

- $20 <$ transverse mass of τ system < 400 GeV

- Number of tracks in τ candidate either 1 or 3

The signal efficiency for the different steps is summarized in Table 3. After the pre-selection 8% of the produced signal statistics is left for the analysis. Already 60% of the statistics is lost by requiring both τs to decay hadronically.

Table 3: Efficiency flow for the different steps of the analysis.

Step	Efficiency or Branching Ratio
τ reconstruction	$\text{Eff}(\tau) = 0.7*0.7$
hadronic τ decay	$\text{BR}(\tau \rightarrow \text{hadrons}) = 0.64*0.64$
pre-selection cut	$\text{Eff}(\text{PreSel}) = 0.4$

4.2 Event selection using boosted decision trees

To distinguish between signal and background events further, the Toolkit for Multivariate Analysis (TMVA) [13] is used. Boosted Decision Trees (BDT) proved to be the most efficient classifiers for this analysis. For training purposes, 30% of the available events for each process are used. These events are not considered in the analysis to measure masses or cross sections.

Event classification

The boosted decision trees are trained using 14 variables describing the event topology and describing kinematic quantities of the reconstructed τ candidates:

- Missing transverse momentum $p_{\text{T,miss}}$

- Thrust and oblatenes of the τ system

- Sum of τ energies

- Sum of the transverse momenta of both τ candidates

- $\cos\theta_{\tau,1}$ and $\cos\theta_{\tau,2}$

- Invariant mass of the τ system

- Transverse mass of the τ system

- Angle between the two τ candidates

- θ^{miss}, where θ^{miss} is the polar angle of the missing momentum

- Accoplanarity ($\Delta\phi$) between the two τs

- Visible energy in the event

- Energy of the rest group (particles not belonging to a τ candidate)

As examples for the input variables, the accoplanarity of the τ system (left) and the angle between the two τ candidates in the event for the signal and all backgrounds are shown in Figure 2.
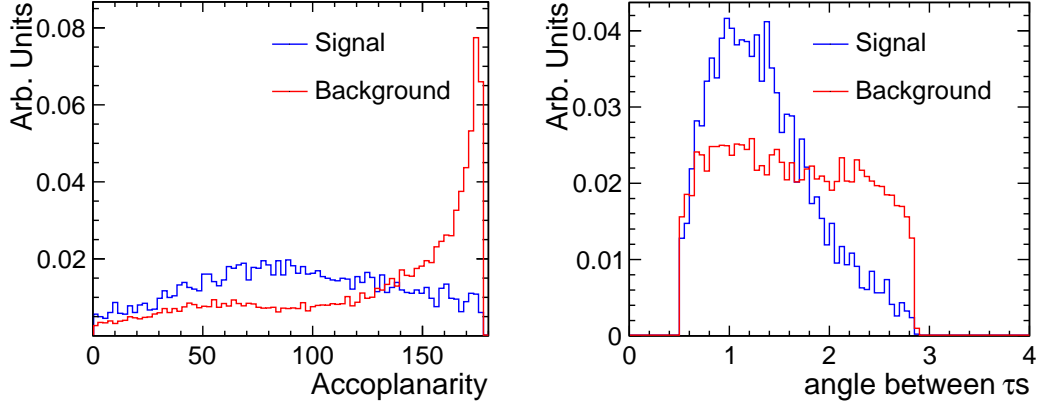
Figure 2: Accoplanarity of the τ system (left) and the angle between the two τ candidates (right) for the signal and all backgrounds. The normalization of all distributions is arbitrary to illustrate the different shapes.

Using the input variables described above, the classifier response for each event is computed which is referred to as BDT in the following. Figure 3 shows the distribution of the BDT for the signal and the different backgrounds. Signal events tend to be at higher BDT values than the backgrounds. Additionally, the signal efficiency, purity and significance for events passing the pre-selection as a function of the BDT cut value are shown. The highest significance is reached for a BDT of 0.05 which is then chosen for the analysis giving an efficiency of 55% and a purity of 43% for this selection. Combined with the efficiency from the reconstruction and the pre-selection this amounts to a signal efficiency of around 4% of the produced data.

The purely statistical uncertainty on the cross section can be calculated as $\sqrt{S+B}/S$. Figure 4 shows the dependence of this statistical uncertainty of the cross section on the selection efficiency and the BDT value.

Since this analysis is merely a counting experiment it is not necessary to limit the calculation of the statistical error by actually choosing a BDT cut. Each BDT bin with signal can be used to determine a value for $\Delta\sigma/\sigma$. These individual errors can be summed quadratically:

$$x_i = \frac{\sqrt{S_i + B_i}}{S_i}$$

$$\frac{\Delta\sigma}{\sigma} = \sqrt{\frac{1}{\sum_i \frac{1}{x_i^2}}}$$

This methods improves the statistical uncertainty for the cross section slightly from 4.3% to 3.7%. It is very stable and does not depend on the number of bins choosen. Reducing the number of bins from 200 to 25 the obtained result changes from 3.66% to 3.69%. The result is summarized in Table 4.
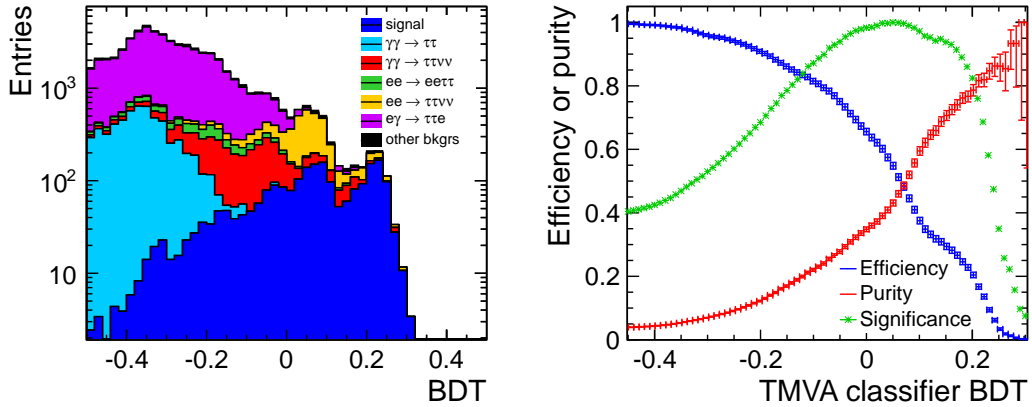
Figure 3: Distribution of the BDT values for the signal and the backgrounds for 1.5 ab^{-1} (left) and the selection efficiency, purity and significance in dependence on the chosen BDT cut value (right).
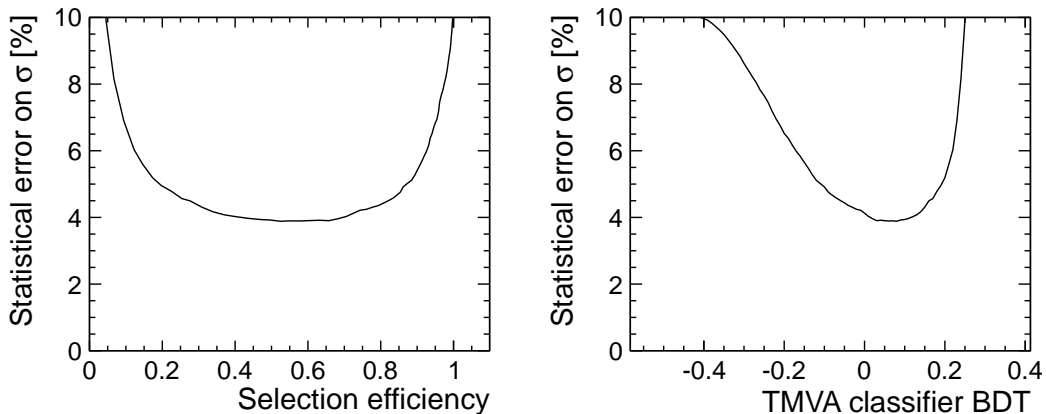

Figure 4: Statistical uncertainty of the cross section in dependence on the selection efficiency (left) and the chosen BDT cut value (right).

For a Higgs mass of 126 GeV instead of 120 GeV the branching ratio into $\tau\tau$ is reduced from 7.04% to 6.15%. The cross section drops by about 1%. The effective signal cross section for this analysis would therefor drop from 17.5 fb to 15.1 fb which is a 13% effect. Assuming one could translate that directly into a loss of 13% of the selected signal the statistical accuracy on the cross section times branching ratio would increase from 3.7% to 4.1%.

Table 4: Event selection performance and measured statistical uncertainty of the cross section for H → ττ decays.

Total signal events	2238
Signal events (above BDT cut)	1227
Background events (above BDT cut)	1620
Signal efficiency	55%
Signal purity	43%
Cross section statistical uncertainty	3.7%

5 Conclusion and Summary

The cross section for H → ττ decays was measured for the producion via ee → Hνν reconstructing hadronic decays of the two τs. The study was performed using full simulation and considering pileup from γγ → hadrons. A center-of-mass energy of 1.4 TeV and an integrated luminosity of 1.5 ab^{-1} is used. For $M_{\mathrm{H}} = 120$GeV the statistical uncertainty of the cross section times branching ratio is 3.7%. Given that this analysis reuses all background samples generated for the study of staus which produce τs of higher energies than in this analysis it is expected that the uncertainty on the cross section could be improved further if for example the cut on the transverse momentum of the reconstructed τ was lowered. As all the background samples already have these cuts applied at production level it would require a complete re-generation and simulation. Another way to improve the signal statistic is to include the leptonic τ decays. This would also require additional backgrounds to be produced.

References

[1] L. Linssen et al., *Physics and Detectors at CLIC: CLIC Conceptual Design Report*, CERN-2012-03

[2] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, arXiv:0708.4233 (2007); M. Moretti et al., *O'Mega: An Optimizing Matrix Element Generator*, LC-TOOL-2001-040, also arXiv:hep-ph/0102195 (2001).

[3] B. Dalena, J. Esberg and D. Schulte, *Beam-induced backgrounds in the CLIC 3 TeV CM energy interaction region*, Proceedings of LCWS11, arXiv:1202.0563 (2012).

[4] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual* , JHEP 05 (2006) 026.

[5] Simulator for the Linear Collider (MOKKA), http://www.lcsim.org/software/slic/.

[6] S. Agostinelli et al., *GEANT4 - a simulation toolkit*, Nucl. Instr. Meth. A 506 (2003) 250; J. Allison et al., *Geant4 Developments and Applications*, IEEE Trans. Nucl. Sci. 53 (2006) 270.

[7] A. Münnich and A. Sailer, *The CLIC ILD CDR Geometry for the CDR Monte Carlo Mass Production*, LCD-Note-2011-002 (2011).

[8] P. Schade, A. Lucaci-Timoce, *Description of the signal and background event mixing as implemented in the Marlin processor OverlayTiming*, LCD-Note-2011-006 (2011).

[9] J. Marshall and M.A. Thomson, *Redesign of the Pandora Particle Flow algorithm*, Report at IWLC2010 (2010).

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

[11] A. Münnich, *Measurement of $\tilde{\tau}_1$ pair production at CLIC*, LCD-Note-2012-004 (2012).

[12] A. Münnich, `TauFinder: A Reconstruction Algorithm for $\tau$ Leptons at Linear Colliders`, , LCD-Note-2010-009 (2010).

[13] A. Hoecker et al., *TMVA 4 - Toolkit for Multivariate Data Analysis with ROOT*, arXiv:physics/0703039, also CERN-OPEN-2007-007 (2009).