CLICdp-Conf-2016-003
20 January 2016

Physics potential of the $BR(H \to WW^*)$ measurement at a $\sqrt{s}$=350 GeV and $\sqrt{s}$=1.4 TeV CLIC collider

Mila Pandurović*[1)], Mark Thomson†

On behalf of the CLICdp collaboration

* *Vinca Institute of Nuclear Sciences, University of Belgrade, Serbia,* † *University of Cambridge, United Kingdom*

Abstract

Precision measurements of the number of properties of the Higgs boson, like invariant mass and couplings to the Standard Model particles, represent one of the key measurements of the CLIC physic program. The CLIC energy staging scenario allows to perform these measurements using different Higgs production channels. The Higgs decay to a WW pair, which is analysed at two CLIC energy stages, plays an important role in this program, as it gives access to the relative Higgs couplings to the vector bosons and to the total Higgs decay width. The studies presented here are part of an ongoing effort to investigate the full physics potential of the CLIC collider.

Talk presented at the International Workshop on Future Linear Colliders (LCWS14), Belgrade, Serbia, 6–10 October 2014.

[1)] milap@vinca.rs

1 Introduction

The careful design of the energy staging scenario at CLIC is optimised to provide a number of measurements foreseen by the rich physics program, elaborated in detail in [1]. The first stage, $\sqrt{s}$ = 350GeV, is mainly devoted to the precision Standard Model (SM) physics, namely Higgs and top physics. The second energy stage, $\sqrt{s} = 1.4$ TeV, gives access to possible New Physics phenomena and expands the spectrum of the Higgs precision measurements to the top-Yukawa coupling, Higgs potential and rare Higgs decays. The highest energy stage, $\sqrt{s} = 3$ TeV, covers complete SM precision physics and provides optimal sensitivity to New Physics phenomena.

Measurements at CLIC devoted to the investigation of the properties of the SM-like Higgs boson are one of the cornerstones of the CLIC program. In order to explore the full physics potential of the CLIC machine, a set of full simulation studies, which concern the statistical precision of the Higgs boson couplings to the SM particles that can be achieved at different CLIC energy stages, are performed.

At the first CLIC energy stage, the leading Higgs production channel is the Higgsstrahlung process ($e^+e^- \to HZ$). This is a s-channel type of Higgs production, (Figure 1), with a total cross-section of 134 fb. At the next energy stage, the dominant production goes through the t-channel WW-fusion process ($e^+e^- \to H\nu_e\bar{\nu}_e$), (Figure 2), with a total cross-section of 278 fb.

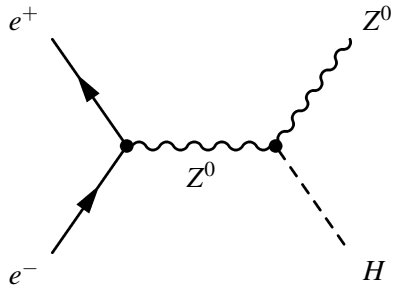

Figure 1: Higgsstrahlung production process.

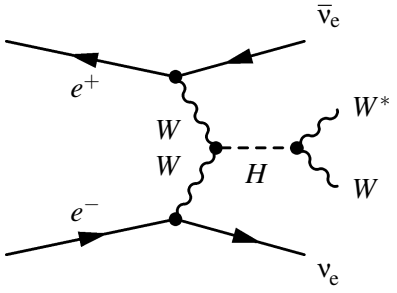

Figure 2: WW-fusion Higgs production process.

The measurement of the Higgsstrahlung process enables a model-independent determination of the absolute coupling of the Higgs to Z bosons, g_{HZZ}, utilising the Higgs recoil mass distribution obtained by using either leptonic or hadronic Z boson decays. The obtained absolute coupling can be used further as an input to a number of other Higgs coupling measurements.

The two analyses presented here are the measurements of the fully hadronic Higgs decay, ($H \to WW^* \to q\bar{q}q\bar{q}$), at the first and the second CLIC energy stage. When the Higgs boson is produced in Higgsstrahlung, this decay gives access to the relative coupling of the Higgs to the W bosons at CLIC, through the observable $(\frac{g^2_{HZZ} \cdot g^2_{HWW}}{\Gamma_H})$, where Γ_H is the total Higgs decay width. Using the input from the recoil mass measurements of the Higgs coupling to the Z boson, a relative Higgs to W boson coupling can be calculated. On the other hand, the same Higgs decay mode, but in a different Higgs production channel, WW-fusion, can be used to obtain Γ_H, by measuring the observable $(\frac{g^4_{HWW}}{\Gamma_H})$, using as an input the combined result from the independent measurements of the g^2_{HWW} coupling.

2 Simulation and analysis tools

Both Higgs production channels, Higgsstrahlung and WW fusion, as well as background processes, were simulated using the Whizard 1.95 [2] event generator, including initial state radiation and a realistic CLIC luminosity spectrum. The CLIC luminosity spectrum and beam-induced processes were simulated by GuineaPig 1.4.4 [3]. The hadronisation and fragmentation of the Higgs and vector bosons were simulated

using Pythia 6.4 [4]. Background coming from $\gamma\gamma$ to hadrons were overlaid over each generated event-sample before reconstruction. Particle reconstruction and identification was done using the particle flow technique, implemented in the Pandora particle-flow algorithm (PFA) [5, 6]. The ILCSoft package [7] is used for the lepton isolation, jet clustering and b tagging. The response of the detector was simulated with the CLIC_ILD [8] detector model. Signal and background separation was obtained using multivariate classification analysis, implemented in the TMVA package [9].

3 Fully hadronic Higgs decay to WW*

The two analyses presented are using the fully hadronic final state, $H \to WW^* \to q\bar{q}q\bar{q}$, at two CLIC energy stages, $\sqrt{s}$ = 350 GeV and 1.4 TeV, using the leading Higgs production channel at each stage. Higgs decay to a WW* pair is the sub-leading Higgs decay mode with a 21.5 % branching fraction. The successive fully hadronic WW* decay, with a 45.6 % branching fraction, leads to a 10% of all Higgs decays. This final state allows complete reconstruction of the invariant mass of the Higgs boson. Although the studied Higgs decay mode is the same for both analyses, the kinematics of the process is different and mainly dictated by the Higgs production channel.

3.1 H $\to$ WW* $\to$ qq̄qq̄ at $\sqrt{s}$ = 350 GeV

The signature of the studied Higgs decay, produced in the Higgsstrahlung process, reflects the s-channel nature of the production, leading to a central topology for all jets and leptons produced.
Depending on the Z-decay, two types of final states are produced: the fully hadronic one, where both W bosons and the Z boson decay hadronically, giving six jets in the final state, and the semi-leptonic channel with four jets coming from the Higgs decay and a pair of leptons from the Z decay. In this analysis only the Z decays to electrons and muons are considered.

3.1.1 Event samples

At 350 GeV centre-of-mass energy, the SM Higgs boson with invariant mass of 126 GeV is produced with a cross-section of 134 fb. Assuming four years of operation of CLIC at the nominal peak luminosity of 1.3 x 10^{-34} $cm^{-2}s^{-1}$, with 50% of data-taking efficiency, the total integrated luminosity would amount to 0.5 ab^{-1}. The list of considered signal and background processes is given in Table 1.

3.1.2 Event selection

The event selection is performed in several distinct steps. Firstly, the two types of final states are distinguished by the presence of the Z-decay leptons. The leptons are separated from other particle-flow objects (PFOs), using the condition E_{track}> 12 GeV, where E_{track} is the energy of a track of the lepton candidate. Figure 3 shows E_{track} as a function of the E_{cone} plane for signal events with Z decaying to an electron-positron pair. E_{cone} is the energy contained in a cone with an opening angle of $\Theta = 5.7^\circ$ around the lepton track.

In the second step, the PFOs assigned to the Z-decay leptons are removed from the event and the rest of the PFOs are clustered into jets. If the number of isolated leptons is equal to two, the event is classified as semi-leptonic, and these events are clustered into four jets. The events containing zero isolated leptons are declared as fully hadronic ones, and these events are clustered into six jets. Additionally, to suppress the background coming from the Higgs decay to a $b\bar{b}$ pair, the event is also clustered into two jets, and the b-tagging and c-tagging probability is assigned to each jet. The next step is the identification of the Higgs, W, W* and Z boson candidates, by grouping of the reconstructed jets and leptons in the event.

For the semi-leptonic final states, the obtained leptons are used to form a Z boson candidate. The four jets in the event, which are comprising the Higgs boson, are grouped to form real and virtual W

Table 1: List of considered signal and background processes, with the corresponding cross-sections at $\sqrt{s}$ = 350 GeV. The signal processes are listed separately, depending on the type of the Z decay. The table also lists the preselection efficiencies (ε_{pres}) and the total selection efficiency (ε_{total}), for two types of final states.

Process	$\sigma[fb]$	$N_{jets}=4$, $N_{lept}=2$		$N_{jets}=6$	
		ε_{pres}	ε_{total}	ε_{pres}	ε_{total}
$e^+e^- \to HZ, WW^* \to q\bar{q}q\bar{q}$					
$Z \to q\bar{q}$	9.16			72 %	50%
$Z \to e^+e^-$	0.453	83 %	30 %		
$Z \to \mu^+\mu^-$	0.454	91 %	35%		
$H \to$ other Higgs decays	92.2	62 %	3.0%	35 %	0.29%
$e^+e^- \to q\bar{q}q\bar{q}$	5847	0.2 %	$< 10^{-5}$%	18 %	0.15%
$e^+e^- \to q\bar{q}l^+l^-$	1704	5.8 %	0.0006%	0.2 %	< 0.0001%
$e^+e^- \to q\bar{q}l\nu$	5914	2.6 %	0.001%	0.1 %	$<10^{-4}$%
$e^+e^- \to q\bar{q}\nu\bar{\nu}$	324.6	2.6 %	$<10^{-5}$%	0.0001%	$<10^{-5}$%
$e^+e^- \to H\nu\bar{\nu}$	51.4	0.01 %	$<10^{-5}$%	0.0006%	$<10^{-5}$ %

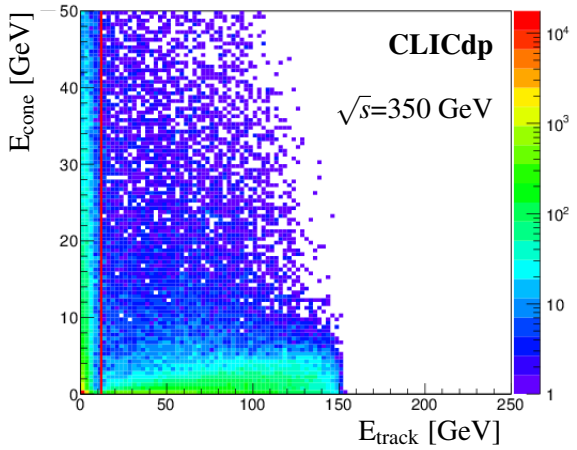

Figure 3: The energy distribution of a track versus the energy within a cone around the lepton track, for the semi-leptonic final state, where the Z boson decays to an electron-positron pair. The cut value of $E_{track} > 12$ GeV is marked with the red line.

boson candidates. Among the three possible combinations of jet pairs (ij), the one with the smallest mass distance, Δ_{min}, between the measured m_{ij} and invariant mass of W boson, m_W, is chosen.

$$\Delta_{min} = min|m_w - m_{ij}|. \qquad i,j = 1,4 \tag{1}$$

For the hadronic final state, the six jets are grouped into pairs to form the Higgs boson, and the real W and Z boson candidates. The combination that minimises the following χ^2 was chosen:

$$\chi^2 = \frac{(m_{ij}-m_w)^2}{\sigma_W^2} + \frac{(m_{kl}-m_Z)^2}{\sigma_Z^2} + \frac{(m_{ijmn}-m_H)^2}{\sigma_H^2}, \qquad i,j,k,l,m,n = 1,6 \tag{2}$$

where σ_W and σ_H are the widths of the W and H mass peaks obtained from the signal events in the semi-leptonic channel, and σ_Z is the width of the Z peak obtained from the signal sample with both W bosons decaying leptonically, and Z decaying hadronically.

Since the cross-sections of the considered background processes for both types of final states are several orders of magnitude higher for the signal events (Table1), the background to signal ratio is minimised by the set of preselection criteria prior to the final selection. The final selection is performed using the multivariate analysis, the Boosted Decision Tree (BDT) classifier. In both signal channels, the backgrounds with the relevant cross-section after the preselection are used as an input to the BDT classifier.
The preselection criteria are optimized separately for the leptonic and hadronic type of signal final state. The list of the preselection variables and the respective cut-off values, for semi-leptonic final states, are:

- the invariant mass of the Z boson candidate, $m_Z > 40$ GeV;
- the invariant mass of real W boson candidate, 45 GeV $< m_W <$ 95 GeV;
- visible energy in the event, 100 GeV $< E_{vis} <$ 300 GeV;
- transverse momentum of each jet in the event, $p_T < 40$ GeV;

and the T value at which the jet number is making the transition from the i-th to the j-th number of jets in the event -log(y_{ij}),

- -log(y_{23}) < 2.5;
- -log(y_{34}) < 4.0.

The preselection efficiencies for the signal and background processes are given in Table 1. For the semi-leptonic signal type, the preselection completely removes the background ($q\bar{q}q\bar{q}$, $q\bar{q}\nu\bar{\nu}$, $H\nu\bar{\nu}$). The other backgrounds, with leptons in the final state, $q\bar{q}l^+l^-$, $q\bar{q}l\nu$ and non-signal Higgs decays (referred to as 'other Higgs decays') are minimised by the preselection and used as an input to the multivariate analysis.

The kinematic variables that are used to discriminate signal and background are: the invariant masses of both W bosons, Z and Higgs boson, m_W, m_{W^*}, m_Z, m_H; number of particle-flow objects (NPFO) in the event; the visible energy, E_{vis}; transverse momentum of a jet, p_{T}; jets transitions -log(y_{23}), -log(y_{34}); the event shape variables (thrust, oblatness, sphericity and aplanarity); the polar angle of Z decay electrons (θ_{lept}); and the flavor tagging probabilities for the two jet hypothesis (btag, ctag).

For the fully hadronic signal final states, the following preselection criteria were used:

- the invariant mass of the Z boson candidate, $m_Z > 70$ GeV;
- visible energy, $E_{vis} > 250$ GeV;
- jet transitions, -log(y_{12}) < 2.0, -log(y_{23}) < 2.6, -log(y_{34}) < 3.0, -log(y_{45}) < 3.2;
- event shape variable, thrust < 0.9.

The effect of the preselection on the signal and background processes is given in Table 1. After the preselection, the semi-leptonic backgrounds are removed ($q\bar{q}l^+l^-$, $q\bar{q}l\nu$). The relevant backgrounds after the preselection, $q\bar{q}q\bar{q}$ and other Higgs decays, are further minimised by using a multivariate analysis. The list of discriminating input variables include: the invariant masses of both W bosons, Z and Higgs boson, m_W, m_{W^*}, m_Z, m_H; number of particle-flow objects (NPFO) in the event; the visible energy, E_{vis}; transverse momentum of jets that comprise the Higgs boson p_T^{Higgs}; jets transitions -log(y_{12}), -log(y_{23}), -log(y_{34}), -log(y_{45}), -log(y_{56}), -log(y_{67}); event shape variables (thrust, oblatness, sphericity and aplanarity); and flavor tagging probabilities for the two jet hypothesis (btag, ctag).

The final selection is based on the BDT classifier cut value which is chosen to minimise the statistical uncertainty of the cross-section $\sigma(HZ) \times BR(H \to WW^*)$, that is to minimise the ratio

Table 2: List of considered signal and background processes for the $\sqrt{s}$ = 1.4 TeV case, with the corresponding cross-sections. The table lists also the preselection efficiencies (ε_{pres}) and the total selection efficiency for a likelihood cut of $\mathscr{L} > 0.35$ and the expected number of events passing event-selection for an integrated luminosity of 1.5 ab^{-1}.

Process	$\sigma[fb]$	ε_{pres}	$\varepsilon_{\mathscr{L}>0.35}$	$N_{\mathscr{L}>0.35}$
$e^+e^- \to H\nu_e\bar{\nu}_e$	244.1	14.61%	3.0%	11101
$H \to WW^* \to q\bar{q}q\bar{q}$	23.9	32.4 %	18.1%	7518
$H \to WW^* \to q\bar{q}l\nu$	23.0	4.4 %	0.6%	253
$H \to b\bar{b}$	136.9	1.9 %	0.4%	774
$H \to c\bar{c}$	6.8	8.1 %	2.1%	209
$H \to gg$	20.7	19.1 %	7.1%	1736
$H \to ZZ$	7.1	12.0 %	5.0%	556
$H \to$ other Higgs decays	25.6	0.7 %	0.2%	55
$e^+e^- \to q\bar{q}\nu\bar{\nu}$	788.0	4.6%	0.2 %	2225
$e^+e^- \to q\bar{q}q\bar{q}l\nu$	115.3	0.1%	<0.1 %	43
$e^+e^- \to q\bar{q}q\bar{q}\nu\bar{\nu}$	24.7	0.8%	0.4 %	130
$\gamma e^+(\gamma e^-) \to q\bar{q}q\bar{q}\nu$	254.3	1.8%	0.4 %	1389

$$\frac{\Delta\sigma}{\sigma} = \frac{N_S}{\sqrt{(N_S + N_B)}}, \tag{3}$$

where N_S, N_B are the number of signal and background events after the final selection, respectively.

The analysis is still ongoing and this paper presents only the adopted strategy. A refinement of the lepton isolation is foreseen, as well as addition of backgrounds with six fermions in the final state.

3.2 $H \to WW^* \to q\bar{q}q\bar{q}$ at $\sqrt{s}$ = 1.4 TeV

The Higgs production at the second CLIC energy stage, $\sqrt{s}$ = 1.4 TeV, is dominated by the t-channel WW fusion process. Therefore the studied Higgs decay is characterized by soft, forward-peaked jets and missing energy from the production process. The total invariant mass in the event has to be consistent with the Higgs mass, and the mass of one of the jet-pairs has to be consistent with the invariant mass of the W boson. The considered background processes are listed in Table 2. An integrated luminosity of 1.5 ab^{-1} is assumed.

The main backgrounds are coming from two types of processes. One from other pure hadronic Higgs decays, like $H \to b\bar{b}$, $H \to c\bar{c}$ and $H \to gg$, which have an invariant mass consistent with the Higgs mass. The second group of the final states are large cross-section backgrounds with a signal-like jet/missing energy combination in the final state. In the first stage of the analysis, each event is clustered into four jets using the k_T jet-finder. Of three possible jet associations with the candidate W bosons, the one containing the pair with the invariant mass closest to m_W is selected. In addition, in order to suppress the main Higgs decay background, $H \to b\bar{b}$, with the BR($H \to b\bar{b}$) = 57.7 %, the events are forced into two jets, and b-tagging probability is assigned to each jet. In the next step, the following set of preselection criteria is applied:

- the invariant mass of the Higgs boson, 70 GeV $< m_H <$ 150 GeV;
- the invariant mass of the W boson candidate, 45 GeV $< m_W <$ 95 GeV;
- the invariant mass of the W* boson candidate, $m_{W^*} <$ 65 GeV;
- visible energy, 125 GeV $< E_{vis} <$ 600 GeV;

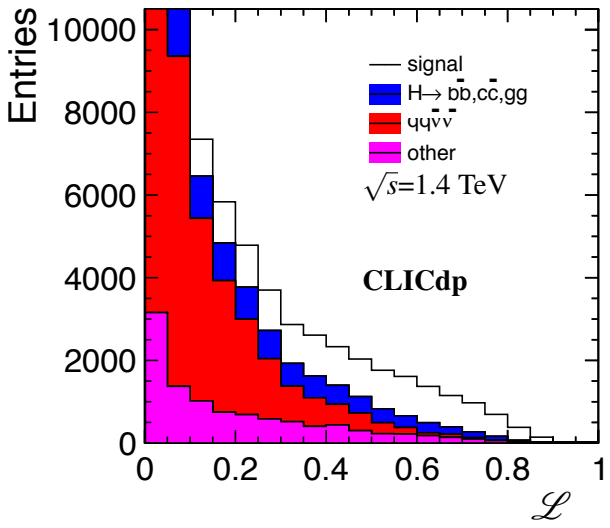

Figure 4: The relative likelihood distribution for the preselected events, for the signal and background samples of an integrated luminosity of 1.5 ab^{-1} at $\sqrt{s}$ = 1.4 TeV.

- transverse momentum of a jet, $p_T > 90$ GeV;
- jet transitions, -log(y_{23}) < 2.75, -log(y_{23}) < 3.5;
- flavor-tagging probabilities for the two-jet hypothesis, btag$_1$, btag$_2 < 0.95$.

After the preselection, the main backgrounds are $e^+e^- \to q\bar{q}\nu\bar{\nu}$ and $\gamma e^{\pm} \to q\bar{q}q\bar{q}\nu$ and other Higgs decays, predominantly $H \to b\bar{b}$ and $H \to gg$, as can be seen from Table 2. These backgrounds are inputs to the classification using a relative likelihood selection. For each of the five types of events, including signal, a relative likelihood distribution, $\mathscr{L}$, is constructed as following:

$$\mathscr{L}_i = \frac{L_i}{L_{WW^*} + L_{b\bar{b}} + L_{gg} + L_{q\bar{q}\nu\bar{\nu}} + L_{q\bar{q}q\bar{q}\nu}}, \tag{4}$$

where L_i are the absolute likelihood distributions for each of the five event-types. The absolute likelihood distributions for signal and four types of background, L$_i$, are formed from the normalised probability distributions $P_i(x_i)$ for the three two-dimensional distributions, invariant masses m_H vs. m_W , jet transitions -log(y$_{23}$) vs. -log(y$_{23}$), and b-tag probabilities btag$_1$ vs. btag$_2$:

$$L_i = P(m_w, m_H) \times P(-\log(y_{23}), -\log(y34)) \times P(\text{btag}_1, \text{btag}_2). \quad i = 1,5 \tag{5}$$

The use of the two-dimensional distribution is taking into account the most significant correlations between the likelihood variables. The constructed relative likelihood distributions are shown in Figure 4.

The selection efficiencies and expected number of events for the signal-dominated region, $\mathscr{L} > 0.35$, are listed in Table 2. The expected precision for BR($H \to WW^*$) is extracted from a fit to the likelihood distributions of Figure 4.

A χ^2 fit to the expected $\mathscr{L}$ distribution is performed by scaling independently five components: the signal, $H \to WW^*$, and four types of backgrounds, $H \to b\bar{b}$, $H \to c\bar{c}$, $H \to gg$ and all other backgrounds. The latter are dominated by $e^+e^- \to q\bar{q}\nu\bar{\nu}$ and $\gamma e^{\pm} \to q\bar{q}q\bar{q}\nu$.

The background components coming from the Higgs decays ($H \to b\bar{b}$, $H \to c\bar{c}$, $H \to gg$) were constrained using the branching ratios obtained by the independent analyses performed at CLIC, at the same

energy stage. The systematic uncertainty in the non-Higgs background, denoted by $\bar{h}$, is taken to be 1%. The constrains are implemented by modifying the χ^2 function to include these penalty terms:

$$\chi^2 \to \chi^2 + \frac{(S_{b\bar{b}}-1)^2}{\sigma^2_{b\bar{b}}} + \frac{(S_{c\bar{c}}-1)^2}{\sigma^2_{c\bar{c}}} + \frac{(S_{gg}-1)^2}{\sigma^2_{gg}} + \frac{(S_{ZZ^*}-1)^2}{\sigma^2_{ZZ^*}} + \frac{(\bar{h}-1)^2}{\sigma^2_{\bar{h}}}. \quad (6)$$

The resulting statistical uncertainty on the $H \to WW^*$ branching ratio is

$$\delta[\sigma(H\nu\bar{\nu}_e)xBR(H \to WW^*)] = 1.4\%. \quad (7)$$

4 Conclusions

Two measurements of relative statistical uncertainty of the branching fraction of the $H \to WW^*$ decay, in two Higgs production channels, accessible at different CLIC energy stages, 350 GeV and 1.4 TeV, have been presented.

The first analysis is performed at the first CLIC energy stage, $\sqrt{s}$= 350 GeV, where the dominant Higgs production channel is the Higgsstrahlung process, $e^+e^- \to HZ$. Two types of final states were analysed, the fully hadronic and the semi-leptonic, and the adopted analysis strategy is presented. The presented work is still ongoing and the future plans incorporate the refinement of the criteria for the lepton isolation, inclusion of additional backgrounds with six fermions in the final state and accommodation of the preselection criteria and final selection to these new backgrounds.

The second analysis of the studied Higgs decay has been performed at the second CLIC energy stage, $\sqrt{s}$= 1.4 TeV, using the leading Higgs production channel $e^+e^- \to H\nu_e\bar{\nu}_e$. It has been shown that the branching ratio $BR(H \to WW^*)$ can be measured with a statistical uncertainty of $\delta[\sigma(H\nu\bar{\nu}_e)x\ BR(H \to WW^*)] = 1.4\%$.

References

[1] L. Linssen et al., eds., *Physics and Detectors at CLIC: CLIC Conceptual Design Report*, ANL-HEP-TR-12-01, CERN-2012-003, DESY 12-008, KEK Report 2011-7.

[2] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, Eur. Phys. J. **C 71** (2011) 1742, DOI: 10.1140/epjc/s10052-011-1742-y.

[3] D. Schulte, *Beam-beam simulations with GUINEA-PIG*, CERN-PS-99-014-LP, 1999.

[4] T. Sjostrand, S. Mrenna and P.Z. Skands, *Pythia*, JHEP **0605** (2006) 026, DOI: 0.1088/1126-6708/2006/05/026.

[5] M. Thomson, *Particle flow calorimetry and the PandoraPFA algorithm*, Nucl. Instrum. Meth **A 611** (2009) 25, DOI: 10.1016/j.nima.2009.09.009.

[6] J. Marshall, A. Münnich. Thomson, *Performance of particle flow calorimetry at CLIC*, Journal of Physics: Conference Series **700** (2012) 153, DOI: 10.1016/j.nima.2012.10.038.

[7] F. Gaede, *Marlin and LCCD: Software tools for the ILC*, Nucl. Instrum. Methods **A 559** (2006) 177, DOI: 10.1016/j.nima.2005.11.138.

[8] A. Münnich, A. Sailer, *The CLIC ILD CDR Geometry for the CDR Monte Carlo Mass Production*, LCD-Note-2011-002 (2012).

[9] J. Therhaat, *TMVA - Toolkit for multivariate data analysis*, AIP Conf.Proc. **1504** (2013) 1013, DOI: 10.1063/1.4771869.