Feasibility of a minimum bias analysis of $e^+e^- \rightarrow ZH \rightarrow q\bar{q} + X$ at a 250 GeV ILC

Yacine Haddad
Laboratoire Leprince-Ringuet (LLR), École polytechnique, CNRS-IN2P3.
(Dated: April 14, 2014)

The precision measurements of the Higgs properties is crucial for a better understanding of electroweak symmetry breaking. It can be first achieved at the ILC [1] at $\sqrt{s} = 250$ via the Higgs-strahlung production process $e^+e^- \rightarrow ZH$. The hadronic decay mode $Z \rightarrow q\bar{q}$ constitutes more than 65% of the total, a factor 10 more than $Z \rightarrow \mu^+\mu^-$. An analysis based solely on the $Z$ jet pair information could thus lead to a high statistics and provide a minimum biased Higgs sample. A study of the feasibility of such analysis is shown here, based on $e^+e^-$ simulated collisions at 250 GeV in center of mass, for the equivalent integrated luminosity of 500 fb$^{-1}$ and using a fast simulation of the ILD detector.

I. Introduction

The measurement of the Higgs boson properties at the future $e^+e^-$ collider ILC [1] can be achieved via the known Higgs-strahlung process (Fig. 2). The $e^+e^- \rightarrow ZH$ represents the largest production cross section for a center-of-mass (c.o.m.) energy of $\sqrt{s} = 250$ GeV assuming a Higgs mass of 125 GeV (Fig. 1).

![Fig. 1. Higgs boson production cross section as function of the $e^+e^-$ center of mass energy. The purple dashed line represents the ILC operating at a center-of-mass energy of $\sqrt{s} = 250$ GeV](image)

Previous studies [3] have addressed the case where the $Z$ boson decays to $e^+e^-$ or $\mu^+\mu^-$ pairs. For these leptonic channels, the energy and the momentum can be precisely measured with a high tracking detector performance such as the ones proposed for the ILC detectors (SiD and ILD [1]) aiming at momentum resolution of about $\sigma_{(p_t)} \sim 2 \times 10^{-5}$ GeV$^{-1}$. A precision of about 2.6% on the $ZH$ cross section can be achieved at ILC [3].

However, analyses considering only the leptonic channels, are limited by the statistical precision due to the small branching ratio of $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ process ($\sim 3.3\%$). Thus the idea is to extend this analysis to hadronic decay mode of the $Z$ boson [1] $Z \rightarrow q\bar{q}$, which represent the large branching ratio of about $\sim 70\%$. The challenge is greater since it depends on the jet clustering algorithm which may introduce confusion in the combination process for hadronic decays of the Higgs boson (and $H \rightarrow \tau\tau$). Thus different reconstruction efficiency are expected for each decay mode of the Higgs boson.

The following study is based on a fast simulation. A smearing of the four momentum of the stable particles takes care of the expected performance of the ILD detector. The jets are then reconstructed using standard $e^+e^-$ jet clustering. The $Z$ boson jet pair is selected by having mass consistent with $Z$ boson. In order to reduce the background contamination, an event selection based on $s$ decay.
Boosted Decision Tree (BDT) [4] is exploited. The first preliminary results on this $ZH$ analysis are here shown and commented upon.

II. Simulation tools

A. Event generation

All events were generated using WHIZARD 1.95 Monte Carlo (MC) event generator [5] configured with ILC beam parameters, taking into account beamstrahlung and initial state radiation (ISR) photons. PYTHIA 6.4 [6] insured the final state QED and QCD parton showering, fragmentation and decay providing final-state observable particles.

Both signal and background events were produced at a c.o.m energy of $\sqrt{s} = 250$ GeV for a total integrated luminosity of 500 fb$^{-1}$ and a Higgs mass of 125 GeV. The $ZH$ process accounts for 95% of the total fermionic production cross section for the chosen Higgs mass (Fig. 1). The SM processes $e^+e^- \rightarrow WW$ and $e^+e^- \rightarrow ZZ$ decaying hadronically, leptonically or semi-leptonically are the main backgrounds. The Table. I summarizes the statistics of the generated samples, signal and background, as well as the corresponding cross section and event weight.

| Process | $N_{\text{events}}$ | $\sigma$ (fb) | weight $e_R^+e_R^-$ ($\mathcal{L} = 500$ fb$^{-1}$) |
|---------|----------------------|---------------|---------------------------------|
| $ZH \rightarrow q\bar{q} + X$ | 120409 | 346.01 | 1.41 |
| $WW \rightarrow q\bar{q}q\bar{q}$ | 321376 | 18781.00 | 60.48 |
| $WW \rightarrow q\bar{q}\nu\bar{\nu}$ | 181533 | 14874.30 | 52.30 |
| $ZZ \rightarrow q\bar{q}q\bar{q}$ | 120088 | 1422.14 | 4.45 |
| $ZZ \rightarrow q\bar{q}l\bar{l}$ | 178900 | 1402.06 | 6.46 |

Table. I. Statistic of the generated samples as well as the corresponding event weight.

Here only the $e_R^+e_R^-$ beam polarization was considered. It yields the largest production cross section and is the most challenging for the background suppression.

B. Generic fast simulation

The response of the ILD detector is modeled by a fast Monte-Carlo simulation (fast-MC), smearing each particle four-momentum according to the expected precision. A Particle Flow Algorithm (PFA) analysis is then used, in which charged particles are measured in the tracker (ignoring their calorimeter deposits) while only neutral particles (photon, neutral hadron) are measured by the calorimeters. Neutrinos are ignored.

The charged particle transverse momentum is smeared by a Gaussian of standard deviation:

$$\sigma_{1/p_T} = 2 \times 10^{-5} \text{ GeV}^{-1}$$

The Fig. 3 represent the relative dispersion of charged particle momentum. For photons ($\gamma$) and neutral hadrons ($h^0$) the energy is smeared by the expected calorimetric performance which correspond roughly to

$$\gamma: \sigma_E/E \sim 0.01 \oplus 0.1/\sqrt{E \text{ [GeV]}}$$

$$h^0: \sigma_E/E \sim 0.1 \oplus 0.5/\sqrt{E \text{ [GeV]}}$$

Each reconstructed particle is considered as Particle Flow Object (PFO). Only PFO with $|\eta| < 2.66$ and $p_t > 0.5$ GeV are kept to mimic the acceptance of the ILD detector.

![Fig. 3](image)

Fig. 3. Difference between the momentum of charged particle before and after smearing. The charged particles are well measured with respect to the performance of the tracker resolution.

III. Analysis Strategy

A. Jet Reconstruction

The fragmentation products of the hadronic system were clustered using the Durham-$k_t$ algorithm [7] [8] implemented in FastJet [9]. The Durham-$k_t$ algorithm has only one parameter $y_{\text{cut}}$ fixed here at $y_{\text{cut}} = 0.01$ without restricting the number of reconstructed jets. The particles are combined using the energy combination scheme where the four-vector of particle are summed. The Fig. [4] shows the distribution of number of jets of the signal. A peak around $N_{\text{jets}} \sim 4$ is observed for the signal as expected from the predominantly hadronic decays of the $Z$ and Higgs bosons.

All the possible jet pair combination is calculated. The pair matching at the best a $Z$ boson is selected using the $\chi^2 = (m_{jj} - m_Z)^2/\sigma_Z^2$ criterion.

B. Boosted Decision Tree (BDT) based selection

The separation of the signal and background event is performed by a Boosted Decision Tree (BDT) as imple-
Events surviving the BDT selection are used for the signal strength estimation.

Several options of BDT algorithm are available in the TMVA package, we choose the BDT-Gradient (BDTG) which shows a good signal-background separation. The BDT output variable ranges in $[-1, 1]$ (Fig. 5).

By varying a cut on the BDT output variable, the signal efficiency and the background rejection efficiency were determined. The signal efficiency is defined as the fraction of the event passing the cut, the background rejection efficiency is the fraction of the background event removed by the applied cut.

An appropriate cut can then applied on BDT output variables as shown in the Table. 1 to insure maximum background rejection as well as a good signal efficiency. The results shown that the four backgrounds are significantly reduced with an efficiency above 94%, for a signal efficiency above 44% (for the category $ZZ \rightarrow 2j$) and up
to 98% (for the category $WW \rightarrow 2j + 2l$)

| Category          | BDT cut $\varepsilon_{\text{sig}}$ | $1 - \varepsilon_{\text{bkg}}$ |
|-------------------|------------------------------------|---------------------------------|
| $WW \rightarrow 4j$ | $\text{BDT}_{\text{output}} > 0.8$ | 58%                             |
| $WW \rightarrow 2j + 2l$ | $\text{BDT}_{\text{output}} > -0.5$ | 96%                             |
| $ZZ \rightarrow 4j$ | $\text{BDT}_{\text{output}} > 0.6$ | 44%                             |
| $ZZ \rightarrow 2j + 2l$ | $\text{BDT}_{\text{output}} > 0.6$ | 88%                             |

Table II. Summary of the applied cut on the BDT output for the different categories. The efficiency $\varepsilon_{\text{sig}|bkg}$ is defined as the fraction of the event passing the cuts.

IV. Results

The recoiling Higgs mass distribution after the selection of the $ZH \rightarrow q \bar{q} + X$ is shown in the Fig. 7. A clear signal peak can be easily identified.

An overall background rejection of over 98% is notified (Table. III), but with a poor signal efficiency of about $\sim 14\%$. This low efficiency is due to the difficulties to remove background event (mainly $WW/ZZ \rightarrow q \bar{q}q \bar{q}$) having the same topology as the signal.

| Samples                      | Before cuts | After cuts | efficiency |
|------------------------------|-------------|------------|------------|
| $ZH$                         | 120409      | 16613      | 13.8%      |
| $WW \rightarrow 4j$         | 321376      | 2225       | 0.6%       |
| $WW \rightarrow 2j + 2l$    | 120088      | 246        | 0.1%       |
| $ZZ \rightarrow 4j$         | 120088      | 3085       | 2.5%       |
| $ZZ \rightarrow 2j + 2l$    | 178900      | 967        | 0.5%       |

Table III. Signal and background selection efficiency.

V. Conclusion and prospective

A method based on the multivariate technique for the analysis of the recoiling system of the $ZH$ where $Z$ decays hadronically is proposed. It uses a generic detector fast simulation, tuned on the ILD detector benchmarks. A first study based on four background categories shows a rejection of 98% of the considered backgrounds. However, the present selection suffers from the large signal suppression, only 14% of the signal surviving the cuts.

Next, the $q \bar{q}$ (continuous Drell-Yan background) has to be added. The $Z$ jet pair selection could be improved by introducing other criteria, such as the jets charge and Optimizing the MVA selection.

The current analysis cannot not yet conclude on the efficiency selection on each decay mode of the Higgs. Thus complementary studies are ongoing.

[1] T. Behnke, J. E. Brau, P. N. Burrows, J. Fuster, M. Peiskin, et al., The International Linear Collider Technical Design Report - Volume 4: Detectors.
[2] Y. Haddad, Presentation at LCWS 2013 (2013). URL https://agenda.linearcollider.org/contributionDisplay.py?sessionId=35&contribId=71&confId=6000
[3] H. Li, et al., HZ Recoil Mass and Cross Section Analysis in ILD.
[4] J. Therhaag, TMVA: Toolkit for multivariate data analysis, AIP Conf.Proc. 1504 (2009) 1013–1016.
[5] W. Kilian, T. Ohl, J. Reuter, WHIZARD: Simulating Multi-Particle Processes at LHC and ILC, Eur.Phys.J. C71 (2011) 1742.
[6] T. Sjostrand, S. Mrenna, P. Z. Skands, PYTHIA 6.4 Physics and Manual, JHEP 0605 (2006) 026.
[7] S. Catani, Y. Dokshitzer, B. Webber, The $k_T$-clustering algorithm for jets in deep inelastic scattering and hadron collisions, Physics Letters B 285 (3) (1992) 291 – 299.
[8] N. Brown, W. Stirling, Finding jets and summing soft gluons: a new algorithm, Zeitschrift fur Physik C Particles and Fields 53 (4) (1992) 629–636.
[9] M. Cacciari, G. P. Salam, G. Soyez, FastJet User Manual, Eur.Phys.J. C72 (2012) 1896.
[10] S. Brandt, C. Peyrou, R. Sosnowski, A. Wroblewski, The Principal axis of jets. An Attempt to analyze high-energy collisions as two-body processes, Phys.Lett. 12 (1964) 57 – 61.
[11] S. J. Brodsky, M. Burkardt, I. Schmidt, Perturbative QCD constraints on the shape of polarized quark and gluon distributions, Nucl.Phys. B441 (1995) 197–214.