Detector optimisation for future linear collider

Boruo Xu
of King's College

A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

Abstract

This is my abstract. To be or not to be.

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the respective Degree Committee.

Boruo Xu

Acknowledgements

Of the many people who deserve thanks, some are particularly prominent, such as my supervisor...

Preface

This will be my preface. Where is Wolly?

Contents

1 Double Higgs Bosons Analysis **1**

1.1 Motivation 1

1.2 Theory 1

1.3 Analysis Straggly Overview 2

1.4 Monte Carlo Sample Generation 2

1.5 Physics object and event reconstruction 3

1.5.1 Electron and Muon identification 5

Bibliography **13**

List of figures **15**

List of tables **17**

“*Two bags of pork scratchings are worth a bag of gold.*”

— Joris the Dutch

Chapter 1

Double Higgs Bosons Analysis

1.1 Motivation

Ha there is a higgs.

We found higgs. Higgs is cool. It explains mass.

Why double higgs. Double higgs coulpling is unique to linear collider. It can revel much about the BSM models.

Generator level study has performed. ILC has done this this and that. gHHH in CLIC before

Here we do things differently. First subchannels, then extract both couplings simultaneously.

1.2 Theory

general higgs field

Lagrangian

current constraunt

single higgs coupling measurement done in higgs

Double higgs measurement

The main mechanism for double Higgs production

1.3 Analysis Straggly Overview

Proof-of-principle study was perform at CLIC for $\sqrt{s} = 1.4\,\text{TeV}$ and $3\,\text{TeV}$. Simulated samples, including those containing double higgs production were used. Signal events, events with double higgs production, were selected via a set of carefully designed and complicated methods. g_{HHH} and g_{WWHH} are extracted simultaneously with template fitting with modified couplings samples.

1.4 Monte Carlo Sample Generation

Single channel is defined as $e^-e^+ \to \mathrm{HH}\nu\bar{\nu}$. It is divided into sub-channel $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$ and $\mathrm{HH} \to \mathrm{b\bar{b}b\bar{b}}$ to allow closer examination and an improvement of signal selection when combined. In particular, I studied $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$ sub-channel.

Selected background samples, including processes initiated by photons, are considered in the analysis and listed in Table **??**. These background were expected share similar topologies with the signal process. When describing a multi-quark final state, it is referring to all final states of the same number of quarks, including final states with possible additional neutrinos and or leptons. A multi-quark final state does not include higgs production, unless explicitly stated.

The usual two-quark and four-quark final states were considered. Since the significant presence of beamstrahlung, where photon produced due to the high electric field generated by the colliding beams, processes initiated by photons are also included.

Processes invloving real photons from beamstrahlung (BS) and “quasi-real” photons are generated separately. For the “quasi-real” photon initiated processes, the Equivalent Photon Appproximation (EPA) has been used.

Photon-electron/photon-photon interactions with four-quark final states were considered. Photon-electron interaction with two-quark final state, one Higgs, and one neutrino is considered. Photon-electron interaction with two-quark final state, one Higgs, and one lepton is not considered due to its negligible cross section.

Single higgs productions are not considered because topologies are very different to the single process. Six-quark final states were not considered due to computational limitation.

For processes involving Higgs production explicitly, simulated Higgs mass is 126 GeV. As multi-quark final state background samples could, in principle, contain double higgs production, they are generated with a Higgs mass of 14 TeV. This will produce negligible double higgs production cross section.

All samples are generated with WHIZARD 1.95 [], taking into account the expected CLIC luminosity spectrum. PYTHIA 6.4 [] tuned on LEP data [] is used to describe fragmentation, hadronisation processes, and Higgs decays. TAUOLA [] is used for τ lepton decays.

Simulation

For most background processes, events are simulated when invariant mass of quarks are above 50 GeV. For electron-photon interaction with four quarks and a neutrino final state, events are simulated when invariant mass of quarks are above 120 GeV. These limits are necessary to generate a large amount of background samples in a feasible time, without losing much signal samples.

Finally, the main beam induced background $\gamma\gamma \to hadrons$ is simulated and overlayed [] to all samples according to the integration time of each subdetector.

1.5 Physics object and event reconstruction

Simulation is performed by MOKKA, interfacing GEANT 4. The reconstruction is done via Marlin in iLCSoft. Separate software package (processor) exists for identification of electrons, muons, taus, and jet reconstruction. New processors have been developed and existing processors have been optimised for a compromise of signal selection and background rejection.

For my signal channel, $\mathrm{HH} \to \mathrm{b\bar{b}W^+W^-}$, there is no lepton in the final state. Hence a effective lepton identifier would improve the signal identification. Processors are wither developed or optimised with samples at $\sqrt{s}$ = 1.4 TeV, and checked against samples at $\sqrt{s}$ = 3 TeV. Because the expected signal significance would be low, the processors are optimised to reject more background at the cost of losing a bit more signals, to increase the signal significance. It was found that the same set of parameters work well under $\sqrt{s}$ = 1.4 TeV and 3 TeV.

Channel	$\sigma(\sqrt{s} = 3\,\mathrm{TeV})$ / fb	$\sigma(\sqrt{s} = 1.4\,\mathrm{TeV})$ / fb
$e^-e^+ \to HH\nu\bar{\nu}$	0.588	0.149
$e^-e^+ \to q_l q_l H\nu\bar{\nu}$	1.78	0.86
$e^-e^+ \to c\bar{c}H\nu\bar{\nu}$	1.12	0.36
$e^-e^+ \to b\bar{b}H\nu\bar{\nu}$	1.91	0.31
$e^-e^+ \to qqqq$	546.5	1245.1
$e^-e^+ \to qqqq\ell\ell$	169.3	62.1
$e^-e^+ \to qqqq\ell\nu$	106.6	110.4
$e^-e^+ \to qqqq\nu\bar{\nu}$	71.5	23.2
$e^-e^+ \to qq$	2948.9	4009.5
$e^-e^+ \to qq\ell\nu$	5561.1	4309.7
$e^-e^+ \to qq\ell\ell$	3319.6	2725.8
$e^-e^+ \to qq\nu\nu$	1317.5	787.7
$e^-\gamma(BS) \to e^-qqqq$	1268.7	1160.7
$e^+\gamma(BS) \to e^+qqqq$	1267.6	1156.3
$e^-\gamma(EPA) \to e^-qqqq$	287.9	287.1
$e^+\gamma(EPA) \to e^+qqqq$	287.8	286.9
$e^-\gamma(BS) \to \nu qqqq$	262.5	136.9
$e^+\gamma(BS) \to \bar{\nu} qqqq$	262.3	136.4
$e^-\gamma(EPA) \to \nu qqqq$	54.2	32.6
$e^+\gamma(EPA) \to \bar{\nu} qqqq$	54.2	32.6
$e^-\gamma(BS) \to qqH\nu\nu$	58.6	15.8
$e^+\gamma(BS) \to qqH\nu\nu$	58.5	15.7
$e^-\gamma(EPA) \to qqH\nu\nu$	11.7	3.39
$e^+\gamma(EPA) \to qqH\nu\nu$	11.7	3.39
$\gamma(BS)\gamma(BS) \to qqqq$	13050.3	21406.2
$\gamma(BS)\gamma(EPA) \to qqqq$	2420.6	4018.7
$\gamma(EPA)\gamma(BS) \to qqqq$	2423.1	4034.8
$\gamma(EPA)\gamma(EPA) \to qqqq$	402.7	753.0

Table 1.1: List of signal and background samples with the corresponding cross sections at $\sqrt{s} = 3\,\mathrm{TeV}$ and $\sqrt{s} = 1.4\,\mathrm{TeV}$. q can u, d, s, b or t. Unless specified, q, ℓ and ν represent particles and its corresponding anti-particles. γ (BS) represents a real photon from beamstrahlung (BS). γ (EPA) represents a "quasi-real" photon, simulated with the Equivalent Photon Approximation. For processes involving Higgs production explicitly, simulated Higgs mass is 126 GeV. Otherwise, Higgs mass is set to 14 TeV.

1.5.1 Electron and Muon identification

IsolatedLeptonFinderProcessor

In Marlin package, IsolatedLeptonFinderProcessor has been used. The optimal parameters ware chosen in collaboration and tested. The particle is identified as an isolated light lepton if it passes a chain of cuts.

A charge track is considered if it has more than 15 GeV energy. An electron is identified if the energy in the ECal is over 90% of the total calorimetric energy. A muon is identified if the energy in the ECal is between 5% and 25% of the total calorimetric energy. Furthermore, only primary track is selected, which requires the Euclidean distance in the x-y plane, the in z direction, and in the x-y-z three dimensional space of the track starting point to the impact point to be less than 0.02 mm, 0.03mm, and 0.04 mm, respectively. The isolation criteria states that

$$E_{cone}^2 \leqslant 5.7 \times E_l - 50 \tag{1.1}$$

where, E_{cone} is the total energy of PFOs within an opening angle of $\cos^{-1}(0.995)$ of the light lepton, and E_l is the energy of the light lepton.

BonoLeptonFinderProcessor

The IsolatedLeptonFinderProcessor is rather conservative. I developed a new more aggressive light lepton selection processor, BonoLeptonFinderProcessor, that utilises calorimetric information provided by PandoraPFA.

The processor uses two chains of cuts.

First chain uses the particle ID information from PandoraPFA. A electron is identified if it is a "PandoraPFA" electron and the energy in the ECal is over 95% of the total calorimetric energy. A muon is identified if it is a "PandoraPFA" muon. Primary track selection states the Euclidean distance in the x-y-z three dimensional space of the track starting point to the impact point to be less than 0.015 mm, and the PFO energy is more than 10 GeV. The light lepton either satisfy the high p_T requirement of at least 40 GeV, or the isolation criteria,

$$E_l \geqslant 23 \times \sqrt{E_{cone}} + 5 \tag{1.2}$$

where E_{cone} and E_l have the same definition as in the IsolatedLeptonFinderProcessor.

Second chain of cuts is similar to the IsolatedLeptonFinderProcessor. An electron is identified if the energy in the ECal is over 95% of the total calorimetric energy. A muon is identified if the energy in the ECal is between 5% and 20% of the total calorimetric energy. Primary track selection states the Euclidean distance in the x-y-z three dimensional space of the track starting point to the impact point to be less than 0.5 mm, and the PFO energy is more than 10 GeV. The light lepton either satisfy the high p_T requirement of at least 40 GeV, or the isolation criteria,

$$E_l \geqslant 28 \times \sqrt{E_{cone}} + 30 \tag{1.3}$$

where, E_{cone} is the total energy of PFOs within an opening angle of $\cos^{-1}(0.99)$ of the light lepton, and E_l is the energy of the light lepton.

Comparison: IsolatedLeptonFinderProcessor v.s. BonoLeptonFinderProcessor

Two processors share similar criterion for light lepton identification. The main difference is that the BonoLeptonFinderProcessor allows high p_T light lepton to be identified in a potential non-isolated environment, which leads to the more aggressiveness of the BonoLeptonFinderProcessor. The performance of two processors on the signal and selected background samples is shown in table 1.2

TauFinderProcessor

With a decay length of $87\mu m$, tau leptons decay before reaching the detector and can only be identified through the reconstruction of their decay products. The leptonic decay of tau can be identified using the two isolated lepton finder processor. Therefore tau identification will focus on the hadronic decay.

TauFinderProcessor, an existing processor Marlin package, has been tuned in collaboration and tested. The a collection of tau decay productions are identified they pass a chain of cuts.

Particles are not considered if p_T is less than 1 GeV or $|\cos(\theta_Z)|$ is more than 1.1 rad, as they are more likely from beam induced background. A seed is considered if a charged particle has p_T more than 10 GeV. A search cone of opening angle 0.03 rad is then formed. The search cone is rejected if it has more than 3 charged particles, more than 10 particles or its invariant mass more than 2 GeV. An isolation cone is formed with opening angle between 0.03 and 0.33 rad of the seed. The seed is rejected if there are more than 3 GeV in the isolation cone.

BonoTauFinderProcessor

The TauFinderProcessor's performance is decent, but there is room for improvement. I developed a new more aggressive tau lepton selection processor, BonoTauFinderProcessor, that utilises calorimetric information provided by PandoraPFA.

Similar to the previous processor, PFOs with p_T less than 1 GeV are rejected. A tau seed is defined as a charged particle with p_T at least 5 GeV. The search cone has an opening angle of $\cos^{-1}(0.999)$. Particles are iteratively added to the search cone according to the size of the opening angle to the seed. The temporary search is then considered if it has one or three charged particles, and the invariant mass is less than 3 GeV. The search cone needs to satisfy one of isolation criterion.

1. No particle in the large isolation cone, and p_T of search cone at least 10 GeV,
2. One charged particle in the search cone, one particle in the large isolation cone, and $r_0$larger than 0.01 mm,
3. Three charged particle in the search cone, one particle in the large isolation cone,p_T of search cone at least 10 GeV, and search cone opening angle less than $\cos^{-1}(0.9995)$,
4. One charged particle in the search cone, no particle in the small isolation cone, $r_0$larger than 0.01 mm, and p_T of search cone at least 10 GeV,
5. Three charged particle in the search cone, no particle in the small isolation cone,p_T of search cone at least 10 GeV, and search cone opening angle less than $\cos^{-1}(0.9995)$,

where large and small isolation cone are defined as opening angle of $\cos^{-1}(0.95)$, and $\cos^{-1}(0.99)$ respectively.

First chain uses the particle ID information from PandoraPFA. A electron is identified if it is a "PandoraPFA" electron and the energy in the ECal is over 95% of the total calorimetric energy. A muon is identified if it is a "PandoraPFA" muon. Primary track selection states the Euclidean distance in the x-y-z three dimensional space of the track

Processor / Efficiency (3 TeV)	Signal	$qqq\ell\nu$
IsolatedLeptonFinderProcessor	99.5%	66.8%
BonoLeptonFinderProcessor	99.0%	52.5%
TauFinderProcessor	97.7%	79.5%
BonoTauFinderProcessor	86.3%	60.3%
ForwardFinderProcessor	95.9%	80.7%
Combined	81.0%	23.3%
Selection / Efficiency (1.4 TeV)	Signal	$qqq\ell\nu$
IsolatedLeptonFinderProcessor	99.3%	50.3%
BonoLeptonFinderProcessor	99.1%	39.9%
TauFinderProcessor	97.5%	52.3%
BonoTauFinderProcessor	89.7%	38.5%
ForwardFinderProcessor	98.9%	95.1%
Combined	86.6%	16.8%

Table 1.2: isolated lepton finder processors performance on the signal and selected background samples.

starting point to the impact point to be less than 0.015 mm, and the PFO energy is more than 10 GeV. The light lepton either satisfy the high p_T requirement of at least 40 GeV, or the isolation criteria,

Other Processors

Other isolated lepton selection processors available in Marlin package, including IsolatedLeptonTagging and TauJetClustering, have been tested. The results, after some tuning of parameters, were unsatisfactory. They either performed poorly comparing to the processors above, or became redundant after the processors above. Therefore, these processors were not used in this analysis.

For the identification of electrons and muons the IsolatedLeptonFinderProcessor [] has been used. The processor uses a combination of track energy, calorimeter energy, impact parameter, and isolation information to distinguish between light leptons and other objects. In order to improve the identification efficiency, the method has been integrated with the particle flow information from Pandora []. Particle ID from Pandora uses shower shape information which is complementray to the IsolatedLeptonFinderProcessor.

This results in a significant rejection of background events with forward electrons and positrons, at the price of a moderate loss of signal events, O(%).

For the identification of tau lepton decays, the TauFinder [] has been used. The processor aims to identify high energetic, low multiplicity jets, including criteria on impact parameter and cone isolation. This method has been improved by searching for 1-prong and 3-prong decays specifically, which account for X% of the tau lepton decays.

Colophon

This thesis was made in $\LaTeX\,2_\varepsilon$ using the "hepthesis" class [1].

Bibliography

[1] A. Buckley, The hepthesis LaTeX class.

List of figures

List of tables

1.1 List of signal and background samples with the corresponding cross sections at $\sqrt{s} = 3\,\text{TeV}$ and $\sqrt{s} = 1.4\,\text{TeV}$. q can u, d, s, b or t. Unless specified, q, ℓ and ν represent particles and its corresponding anti-particles. γ (BS) represents a real photon from beamstrahlung (BS). γ (EPA) represents a "quasi-real" photon, simulated with the Equivalent Photon Approximation. For processes involving Higgs production explicitly, simulated Higgs mass is 126 GeV. Otherwise, Higgs mass is set to 14 TeV. 4

1.2 isolated lepton finder processors performance on the signal and selected background samples. 8