Higgs to $\tau \tau$ analysis in the future $e^+e^-$ Higgs factories

Dan Yu$^1$, Manqi Ruan$^{a,1}$, Vincent Boudry$^2$, Henri Videau$^2$, Jean-Claude Brient$^2$

$^1$IHEP, China
$^2$LLR, Ecole Polytechnique, France

Received: date / Accepted: date

Abstract The Circular Electron Positron Collider and International Linear Collider are two electron positron Higgs factories. They are designed to operate at center-of-mass energy of 240 and 250 GeV and accumulate 5.6 and 2 $ab^{-1}$ of integrated luminosity. Using CEPC official samples, the signal strength for Higgs to $\tau \tau$ events are analyzed. The combined accuracy of the signal strength for $H \rightarrow \tau \tau$ at CEPC achieves 0.8%. Extrapolating this analysis to the ILC setup, we conclude the ILC can reach a relative accuracy of 1.1% or 1.2%, corresponding to two benchmark settings of the beam polarization.

1 Introduction

Since the discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC), the precise measurement of the Higgs boson becomes the focus of the high energy physics experiments. Comparing to the LHC, the $e^+e^-$ Higgs factories have clean environment, well known and adjustable initial state, and can determine the absolute value of Higgs boson couplings and decay width. Because of these advantages, many electron-positron Higgs factories are proposed, including the International Linear Collider (ILC)[1], the Circular Electron Positron Collider (CEPC)[2], the Future Circular Collider $e^+e^-$ (FCCee)[3], and the CLIC[4].

The CEPC has a main ring circumference of 100 km, it can be operated as a Z factory ($\sqrt{s} = 91.2$ GeV) and a Higgs factory ($\sqrt{s} = 240$ GeV). It could also perform a W threshold scan at $\sqrt{s} = 160$ GeV and determines precisely the mass and width of the W boson. After the electron-positron collision phase, a proton collider (SPPC) with a center-of-mass energy around 100 TeV can be installed in the same tunnel. The CEPC has a nominal integrated luminosity of 5.6 $ab^{-1}$ and is expected to produce one million of Higgs bosons [5].

The CEPC has very clean collision environment and its detector system can record almost all the Higgs events. This clean, large-statistic Higgs sample provides crucial information on top of the Higgs program at the HL-LHC, and can boost the precision of Higgs boson property measurements by one order of magnitude [5].

Another $e^+e^-$ Higgs factory, the ILC, has been intensively studied in the past 20 years. Comparing to the circular colliders, the center-of-mass of the ILC is much easier to upgrade. In its staging scenario, the ILC will start the collision at a center-of-mass energy of 250 GeV, serving as a Higgs factory with nominal luminosity of 2 $ab^{-1}$. The ILC can upgrade its center-of-mass energy to 380 GeV, 500 GeV, and eventually, to 1 TeV. These high-energy collisions give access to the $t\bar{t}$, the $t\bar{t}H$, the $v\bar{v}HH$ and the $ZH$ events, and also improves significantly the Higgs width measurements. Another significant advantage of the ILC is the capability of beam polarization. Since the left and right handed fermions have different quantum numbers in the Electroweak interaction, the beam polarization provides precision degree of freedoms to control the initial state of the collision. The beam polarization could significantly enhance the physics performance, for instance the $\sin^2(\theta_W)$ measurements. According to the ILC TDR, there are two official settings of the ILC beam polarization, $P(-0.8, 0.3)$ and $P(0.8, -0.3)$, the first/second number represents the electron/positron polarization status, and the minus sign refers to the left-hand polarization. In terms of the Higgs property measurement, the polarization could also enhance the signal yields, and/or suppress the SM background. Table 1 shows the inclusive cross section, the nominal luminosity, and the expected total Higgs events at the CEPC and the ILC.

As the heaviest lepton in the SM, a significant fraction of the SM Higgs boson decays into di-$\tau$ final states, making the $H \rightarrow \tau \tau$ a sensitive probe to the new physics. In this paper, the expected accuracy of the $H \rightarrow \tau \tau$ signal strength mea-

---

$^a$e-mail: ruanmq@ihep.ac.cn
measurement is analyzed using the official CEPC software and samples. Two different analysis methods are developed, corresponding to the signal with or without jets in the final state. For the events without jets (llH and vvH), the signals are identified using the multiplicity information. For the events with jets (qqH), a dedicated tau-finding algorithm TAU RUS is developed. The results for each channel are combined, showing that the signal strength can be measured to a relative accuracy of 0.8% at the CEPC. This result is also extrapolated to ILC and an accuracy of 1.1%/1.2% is achieved for left or right-hand polarization.

This paper is organized as follows. Section 2 describes the detector model, softwares, Monte Carlo (MC) simulation and samples used in the studies. Section 3 presents two different ττ event finding methods for different channels. In Section 4, the combination and extrapolation of the results for different channels are discussed. A conclusion and general discussion is summarized in Section 5.

2 Samples and software chain

The SM Higgs bosons are mainly generated via the ZH process and the vector boson fusion at the e⁺e⁻ colliders, as shown in Figure 1. The cross section of each channel is shown in Table 2.

At the Higgs runs of both the CEPC and the ILC, the inclusive SM background within the detector fiducial volume, is roughly 2-3 orders of magnitudes higher than that of the Higgs signal. In our analysis, the backgrounds are characterized according to the number of final state fermions at parton level. At 240 or 250 GeV center-of-mass energy, the leading SM backgrounds are the 2-fermion and 4-fermion backgrounds. The 2-fermion backgrounds are the q\(\bar{q}\), b\(\bar{b}\) and \(\tau\)τ background; and the 4-fermion backgrounds includes the ZZ, the WW, the single W, the single Z, and the interfering processes. There are some combinations of particles could come from both Z and W boson, the corresponding processes form the interfering processes include z\(\bar{z}\)or w\(\bar{w}\) process and zorw process. Their cross sections are shown in Table 2.

The detector model used in the simulation is the CEPC baseline detector[2], a Particle Flow Oriented detector. It comprises of a low-material tracking system, a high granularity calorimeter system, and a 3-Tesla large radius solenoid that host both ECAL and HCAL inside. A baseline CEPC simulation-reconstruction software has been established. It uses the Whizard as the generator [6], the Mokka [7] for the full detector simulation, the Clupatra [8] for tracking, and the Arbor [9] for the PFA reconstruction.

Using the CEPC baseline geometry, an official massive Monte-Carlo production is performed, corresponding to the nominal setting of the CEPC Higgs runs. Weighting method are applied to this massive production. For the Higgs processes with small cross section, typically under 20 \(fb\), the sample is simulated to a minimal statistic of 100k. For leading 2-fermion standard model background, the massive production only simulate a fraction of the expected statistics, to save the computing resource. The scaling factor for q\(\bar{q}\), n\(\bar{n}\) and b\(\bar{b}\)backgrounds is 2.7, 5.7, and 3.1 accordingly. For all 4-fermion backgrounds, the samples are generated with full statistics.

3 Signal strength analysis

In this paper, the τ events are classified into two categories according to their final states: without jets (\(\mu\mu H\), \(eeH\) and \(\nu\nuH\)) and with jets (qqH). The \(\tau\tau H\) channel are not discussed in this paper. The statistics of the \(H\rightarrow \tau\tau\) signal at different channels are listed in Table 3.

For the qqH channel, TAU RUS, a dedicated τ finding algorithm has been developed and optimized for this analysis. TAU RUS identifies all the τ candidate in an event, from which the Higgs decay final states are identified. The remaining particles are recognized as the di-jet system, whose
invariant and recoil masses are used to distinguish the signal from the background.

No specific $\tau$ finding algorithm is used for the events without jets. Instead, these signals are identified mainly using the multiplicity of the charged particles and the photons. For the $\mu\mu H$ and $eeH$, the prompt leptons are identified using their invariant and recoil mass information, which significantly reduces the SM background.

Roughly 40% of SM $\tau$s decay into a single charged particle and neutrino(s). To ensure a high signal efficiency, the isolated charged particles are intentionally identified as $\tau$ candidate and neutrino(s). To ensure a high signal efficiency, the baseline detector is equipped with high precision vertex system. Because the $\tau$ leptons has a $ct\tau$ of 89 $\mu m$, the leading tracks decayed from the $\tau$ candidates has a significant impact parameter. Therefore, the track impact parameter is used to distinguish the 1-prong decayed $\tau$ lepton from the prompt isolated tracks.

The analyses of different sub-channels are discussed in detail in the following sections.

3.1 $\mu\mu H$ channel

The analysis of the $\mu\mu H$ channel is presented in this section. Each signal event consists a pair of $\mu$ decayed from the Z boson and a pair of $\tau$ lepton decay from the Higgs boson. The cut chain is shown in Table 4.

The prompt $\mu$ is a critical signature of the signal. The baseline design of CEPC provides a high-efficiency and high purity identification of the leptons. Requesting a pair of $\mu$ preserves 97% of the signal and reduces the entire SM background by 40 times. After the first selections, the prompt di-$\mu$ system is identified as the combination with the closest mass to the 91.2 GeV. The backgrounds are further reduced by applying a constrain on the invariant and recoil mass of the di-$\mu$ system. These requirements suppress the SM background by another 50 times, and the remaining backgrounds are dominated by 2f events. Using the selection condition defined in Table 4 (3 first lines), the signal efficiency is 88.5% and the background rejection rate is 99.95%.

The remaining particles are identified as the di-$\tau$ system. The leading charged particle is identified, and all the particles within 1 radius to the leading particle are identified as one $\tau$ candidate, while the remaining particles are identified as the other $\tau$ candidate. Since the $\tau$ lepton decays into small number of charged particles and photons, the charged particle and photon multiplicity of each $\tau$ candidate is required to be smaller than 6 and 7, respectively. At this step, the 2 fermion backgrounds including $\mu\mu$ and $\tau\tau$ (mainly 1 prong after the $\mu$ selection) are suppressed by requiring the existence of the $\tau$ candidates after the $\mu$ pair suppressed.

Several variables are extracted from the di-$\tau$ system, and combined using the TMVA method, to further suppress the background. These variables includes:

- the angle between the leading tracks and the furthest track in each region.
- the angle between the leading tracks and the furthest photon in each region.
- the angle between the leading photon and the furthest photon in each region.

The impact parameter of the leading track in the $\tau$ candidates is also used in the TMVA. By looking at the impact parameters of the tracks, those stemming from $\tau$ decays are further away from the vertex than the others. The impact parameters along the transverse and longitudinal directions (D0 and Z0\(^1\)) of the leading track from each $\tau$ candidate can be extracted.

The BDT distribution for the signal and the ZH backgrounds events are shown in Figure 2. After an optimized cut at BTD value of 0.78, the background reduced further by 30%, at the cost of lost 5% of the signal. See the 7th line of Table 4.

![Fig. 2 The BDT values for the signal and the ZH backgrounds events](image)

After the TMVA event selection in Table 4, the remaining backgrounds are Single Z and ZZ events.

\(^1\)The impact parameter D0 is the signed distance from the origin to the point of closest approach in the $r-\phi(x-y)$ plane. The impact parameter Z0 is the Z position of the perigee.
Table 4 \( \mu\mu H \) cut flow

| \( \mu\mu H \) | 2f | sv | sz | WW | ZZ | mixed | ZH | total Bkg | \( \sqrt{S + B}/S(\%) \) |
|----------------|----|----|----|----|----|-------|----|-----------|------------------|
| total generated | 2388 | 8011 | 15207 | 8972946 | 50826211 | 6389424 | 21839941 | 1102582 | 909900581 | 1263.17 |
| \( N_{\tau -} > 1, N_{\tau +} > 1 \) | 2251 | 658199 | 0 | 17760 | 111340 | 56516 | 99822 | 957 | 944594 | 48.02 |
| 50 GeV < \( M_{\text{recoil}} \) < 100 GeV | 2111 | 864842 | 0 | 31145 | 111376 | 56642 | 99874 | 987 | 962066 | 48.08 |
| 60 GeV < \( M_{\text{invariant}} \) < 100 GeV | 2042 | 662042 | 0 | 31145 | 111376 | 56642 | 99874 | 987 | 962066 | 48.08 |
| \( \mu^+ + \mu^- > 1, \mu^+ - > 1 \) | 1900 | 78 | 0 | 996 | 2576 | 8019 | 29 | 105 | 11803 | 6.16 |
| \( N_{\tau -} (A/B) < 6 \) & \( N_{\tau +} (A/B) < 7 \) | 1823 | 0 | 0 | 264 | 231 | 3682 | 9 | 39 | 4225 | 4.26 |
| \( BDT > 0.78 \) | 2026 | 658199 | 0 | 17760 | 111340 | 56516 | 99822 | 957 | 944594 | 48.08 |

The invariant mass of the \( \tau \) pair is calculated using the collinear approximation (assuming the momentum of neutrino or neutrinos is proportional to the \( \tau \)'s). The distribution is shown in Figure 3 for signal and SM inclusive background with a fit.

From the fit result on the \( \tau \tau \) collinear mass, the expected accuracy \( \Delta (\sigma \times BR)/(\sigma \times BR) = \delta(S)/S \) to be 2.8\%, where the \( \delta(S) \) is the fitted signal event number error.

3.2 eeH channel

The analysis of the eeH channel is similar to that of the \( \mu\mu H \) channel. Because of the Z fusion, the signal statistic of the eeH channel increases by 4\% comparing to the \( \mu\mu H \) channel. The electron identification efficiency performance is slightly worse than the muon identification, because the electrons has much significant bremsstrahlung effect. The eeH channel analysis also have much severe single W backgrounds. The cut chain for eeH channel is shown in Table 5. After optimizing the parameters in the cut chain, 57\% of the signal (1422) events survives and the entire SM background are reduced to 12k statistics, leading to a relative uncertainty of 8.3\%, about 50\% worse than that of the \( \mu\mu H \) channel analysis.

Similar to the \( \mu\mu H \) channel analysis, a fit is performed on the \( \tau \tau \) collinear mass. The signal strength of \( H \to \tau\tau \) measured from the eeH channel reaches 5.3\%.

3.3 vνH channel

Comparing to the llH channels, the vνH signal also doesn’t has jet in its final states and has 6 times larger the statistics. However, the vνH has no prompt lepton pairs with clean invariant and recoil mass signature, and the background for the vνH channel is much larger. This makes the signal strength accuracy of vνH channel is worse than in \( \mu\mu H \) channel.

For the vνH channel, the criteria for event selection is that the missing mass is larger, the transverse momentum is also a method for the selection. The steps for di-\( \tau \) tagging is similar as in \( \mu\mu H \) channel except for that there is no need to veto the lepton pair. The cut chain of vνH channel is shown in Table 7. After the \( \tau \) candidates found, the angles between the \( \tau \)s are applied to reduce the 2f backgrounds.

In this channel, the collinear approximation can not be used, so only statistic result of 7.9\% from the cut chain is used as the accuracy.

3.4 qqH channel

The qqH channel is critical for the \( H \to \tau\tau \) signal strength measurement, since 70\% of the Higgs events at the CEPC are generated via this channel. The cut chain for this analysis is summarized in Table 7. It includes 4 steps, relying on the information of the general event description, the di-\( \tau \) system, the di-jet system, and the vertex system correspondingly.

The first step uses the information of the charged particle multiplicity, the total visible energy, and the leading lepton energy. The multiplicity of the charged particles is required to be larger than 10. This requirement eliminates...
Table 5 eeH cut flow

| cut          | 2f | sw | sz | WW | ZZ | mixed | ZH | total Bkg | $\sqrt{S+B}/S(\%)$ |
|--------------|----|----|----|----|----|-------|----|-----------|-------------------|
| total generated | 2483 | 801152078 | 19517399 | 9072946 | 50826211 | 6389424 | 21839941 | 81486 | 909630945 | 184.68 |
| $N_\tau > 1, N_{sc} > 1$ | 2073 | 252768538 | 19020426 | 2069390 | 4793593 | 226473 | 2566203 | 519907 | 273877330 | 79.83 |
| $110GeV < M_{uuss} < 180GeV$ | 1803 | 8931425 | 39252624 | 683298 | 19359 | 10732 | 59181 | 2322 | 13811718 | 206.13 |
| $40GeV < M_{uuss} < 180GeV$ | 1705 | 3046092 | 643288 | 337928 | 51155 | 4422 | 195532 | 859 | 4279266 | 121.35 |
| $N_{coll} / A(B) < 6$ & $N_{col} / A(B) > 7$ | 1598 | 4729 | 22737 | 22453 | 4410 | 474 | 651 | 107 | 136561 | 23.26 |
| BDT > 0.78 | 1422 | 2533 | 3150 | 6315 | 225 | 175 | 271 | 39 | 12708 | 8.35 |

Table 6 nnH cut flow

| cut          | 2f | sw | sz | WW | ZZ | mixed | ZH | total Bkg | $\sqrt{S+B}/S(\%)$ |
|--------------|----|----|----|----|----|-------|----|-----------|-------------------|
| total generated | 16331 | 801152078 | 19517399 | 9072946 | 50826211 | 6389424 | 21839941 | 81486 | 909630945 | 184.68 |
| $110GeV < M_{nuuu} < 225GeV$ | 15709 | 48775523 | 2140070 | 160084 | 2357138 | 651545 | 1795752 | 9307 | 57329491 | 48.21 |
| $M_{nuuu} > 20GeV$ | 14874 | 12307462 | 1879196 | 1155787 | 1080687 | 565093 | 1525347 | 6811 | 18520383 | 28.94 |
| $10GeV < p_T < 80GeV$ | 13010 | 8728105 | 1393911 | 867874 | 612205 | 357230 | 1250288 | 5964 | 13215577 | 27.95 |
| $E_{vis} < 45GeV, E_{vis} < 65GeV$ | 11898 | 7603289 | 691168 | 740355 | 220275 | 343815 | 709917 | 5681 | 9726280 | 26.22 |
| $N_{coll} / A(B) < 6$ & $N_{col} / A(B) > 7$ | 10363 | 288566 | 51018 | 145838 | 130567 | 69682 | 608076 | 1398 | 4382734 | 20.10 |
| BDT > 0.78 | 9960 | 862244 | 270754 | 59187 | 51522 | 32776 | 35473 | 405 | 1613631 | 12.86 |
| $2 < \theta_e < 3$ | 9551 | 430975 | 107021 | 35780 | 40216 | 17950 | 141658 | 381 | 782081 | 9.3 |
| $2 < \delta_{\phi_e} < 3$ | 9007 | 206717 | 94070 | 29353 | 39593 | 14229 | 114486 | 357 | 498805 | 7.9 |

Table 7 Cut Flow of MC sample for qqH → ττ selection on signal and inclusive SM backgrounds. $E_{vis}/E_{Lab}$ represents the energy of the leading lepton or muon, $M_{coll}^\tau$ is the $\tau$ mass calculated with collinear approximation, Pull1 and Pull2 are the pulls of the leading $\tau$ pairs.

| cut          | 2f | sw | sz | WW | ZZ | mixed | ZH | total Bkg | $\sqrt{S+B}/S(\%)$ |
|--------------|----|----|----|----|----|-------|----|-----------|-------------------|
| $N_{total}$ | 48266 | 801152078 | 19517399 | 9072946 | 50826211 | 6389424 | 21839941 | 81486 | 909630945 | 184.68 |
| $N_{Ch} > 10$ | 47347 | 27292986 | 1376307 | 1699792 | 47052263 | 5756249 | 18020636 | 331843 | 359894260 | 40.07 |
| $110GeV < E_{vis} < 235GeV$ | 46183 | 17358096 | 13150996 | 942644 | 31297172 | 3239464 | 5134115 | 264535 | 22764887 | 32.67 |
| $E_{vis} < 45GeV, E_{vis} < 65GeV$ | 44093 | 16593968 | 3413790 | 707027 | 22428227 | 2918136 | 4955026 | 237240 | 20423014 | 32.41 |
| $N_{\tau} > 1, N_{sc} > 0$ | 22414 | 401147 | 212183 | 13999 | 1129502 | 171380 | 193055 | 16821 | 2138087 | 6.55 |
| $90GeV < M_{coll}^\tau < 160GeV$ | 17176 | 9717 | 21483 | 1689 | 135538 | 62721 | 7722 | 5305 | 244175 | 2.97 |
| $70GeV < M_{coll}^\tau < 110GeV$ | 16257 | 1596 | 4119 | 1012 | 26823 | 52307 | 1818 | 717 | 88392 | 1.98 |
| $M_{coll}^\tau (GeV) > 160GeV$ | 16211 | 0 | 1463 | 637 | 11071 | 13814 | 1265 | 647 | 28897 | 0.93 |

Efficiently the full leptonic SM background. According to the visible energy distribution at the signal, the total visible energy is limited to (110, 235) GeV. The up limit of 235 GeV efficiently reduces the fully visible hadronic events, such as the $ee \rightarrow j j (j = uds)$ and the $WW$/$ZZ \rightarrow 4q$. The single W and single Z boson backgrounds have energetic final state leptons, and can be reduced by an up limit on the leading lepton energy.

A $\tau$ finding algorithm, TAURUS, is used to identify the $\tau$ candidates. Its parameters are optimized for the signal strength analysis at qqH channel. An overall $\tau$ finding efficiency/purity of 80%/90% is achieved at the $qqH, H \rightarrow \tau \tau$ signal, see Figure 4. More details can be found in the appendix A, ref. The leading $\tau$ candidate of each charge, if exist, is identified as the decay product of the Higgs boson. The invariant mass of this pair is calculated using the collinear approximation (assuming the momentum of neutrino or neutrinos is proportional to the tau’s), as shown in Figure 5. The second step requires a pair of $\tau$ candidates, and its collinear mass is limited to (90, 160) GeV.

![Fig. 4 $\tau$ finding purity and efficiency at $qqH \rightarrow \tau \tau$ channel.](image)

The first step is a gentle selection that preserves 90% of the signal and reduces the SM background by almost 4 times. The second step reduces the remaining background...
by 3 orders of magnitudes at the cost of losing 60% of the signal events. The second step, especially the $\tau$ finding performance, is critical for this measurement.

After the first two steps, the main backgrounds include the backgrounds with the same final states at the parton level: $ZZ \rightarrow qq \tau\tau$ and $ZH(Z \rightarrow \tau\tau, H \rightarrow qq)$. In addition, a few $WW$, single $W$ backgrounds survive after the previous steps, where one of the $\tau$ candidates might be generated from the mis-identification of TAUROS. These backgrounds can be significantly reduced by using the information of the remaining final state particles, which is defined as the di-jet system. The $ZH$ and $WW$ background can be reduced using the invariant mass of di-jet system, since the signal peaks at the $Z$ boson mass, while the $ZH$ background peaks at the Higgs mass and the $WW$ background has a flat distribution, see Figure 6. After the restriction on di-jet invariant mass (line 7 of Table 7), the $ZZ$ background became a dominant one since its di-jet invariant mass also peaks at $Z$ boson mass. However, the recoil-mass of the di-jet system can clearly separate the signal from the $ZZ$ background, see Figure 7 (line 8 of Table 7).

The PFA oriented detector design and reconstruction provide an accurate reconstruction of the di-jet system. Using the invariant and recoil mass of the di-jet system, the backgrounds are suppressed by one order of magnitude, and the cost of lost 5% of the remaining signal, and the accuracy improves more than a factor of 2, see line 7 and line 8 of Table 7. After these event selections, the dominant backgrounds are $WW$ and $ZZ$, especially their semi-leptonic decay.

For each track, the pull is defined as: $D_h^2 = \sigma_h^2 + Z_h^2$, where $\sigma_h$ is the uncertainty of D0 and Z0. The pull parameter of the event is then extracted as the sum of the two $\tau$ candidates decayed from Higgs, which distinguishes the $\tau$ candidates from the prompt tracks, see Figure 8. The signal strength accuracy of the $H \rightarrow \tau\tau$ is extracted to be 0.93%.

4 Combination of results and extrapolation

To the first order, the measurements of different channels ($eeH$, $\mu\muH$, $\nu\nuH$ and $qqH$) are independent. The cross section of Higgs decaying to $\tau\tau$ can be summarized as in Table 8. A total accuracy of 0.8% is achieved for $5.6 \text{ ab}^{-1}$. In the $eeH$ and $\nu\nuH$ channel, it is shown that the accuracy is much worse than $\mu\muH$ channel, even though the signal statistics in these two channels are larger. In $eeH$ channel, the signal efficiency is about 10% smaller than the $\mu\muH$ channel, caused by the electron identification efficiency performance. On the same time, the single $W$ and single $Z$ backgrounds
with $e^+e^-$ final states are not easy to suppress, leading to a three times larger SM backgrounds comparing with $\mu\mu H$ channel. Therefore, the accuracy in $eeH$ channel is about two times worse than $\mu\mu H$ channel. In $\nu\nu H$ channel, the $Z$ decay final states can not be used, leading to a large amount of backgrounds. On the other hand, the collinear approximation can not be applied to get the fit result.

Table 8 Combined cross section

|          | $\delta (\sigma \times \text{BR})/\langle \sigma \times \text{BR} \rangle$ |
|----------|----------------------------------------------------------------------------------|
| $\mu\mu H$ | 2.8%                                                                              |
| $eeH$     | 5.3%                                                                              |
| $\nu\nu H$ | 7.9%                                                                              |
| $qqH$     | 0.9%                                                                              |
| combined  | 0.8%                                                                              |

The result is extrapolated to ILC. The ILC will be operated at 250 GeV center-of-mass energy with polarized beams, therefore, the signal and background cross sections are different from that at the CEPC. We calculated the cross sections and the expected number of events at the ILC. Assuming that the efficiency for each signal and background stays the same for ILC and CEPC, an extrapolation can be done as in Table 9. Comparing with the result in ILC[11], the independent analysis in this paper leads to an accuracy 10% better, this improvement is mainly from the di-jet system information.

5 conclusion

In this paper, the expected signal strength of different channels with Higgs decaying into $\tau\tau$ at CEPC have been studied and the combined accuracy is reaching 0.8% level. This result is also extrapolated to ILC and gives an accuracy of 1.2% and 1.1% for two polarization set. The analysis is done with two kinds of signal events including events with or without jets.

The PFA oriented detector and the reconstruction at the CEPC is critical for this analysis. At channels without jets, including $\mu\mu H$, $eeH$, and $\nu\nu H$, the $\tau$ events identification relies strongly on a successful reconstruction of the photons and charged particles. The PFA reconstructs proper number of particles, providing the critical multiplicity information. In the $qqH$ events, a dedicated $\tau$ finding algorithm is developed thanks to the precise reconstruction of final state particles. Even more, the invariant and recoil mass of the qq system provide the excellent selection of signal and backgrounds, 90% backgrounds are reduced while 94% signals remained. From further study on the PFA performance, it shows that the qqH signal strength accuracy degrades to 1.3% if the boson mass resolution degrades from 3.7% to 6%.

In both cases, a precise reconstruction of the impact parameter is essential for the $\tau$ events identification. In $llH$ channel, the impact parameters are used in the TMVA training and it can improve about 1/3 of result than the TMVA without these parameters. In $qqH$ channel, the fit on impact parameter improves almost 50% with the final result of cut chain.

In the channels with Z boson decaying to visible final states, the collinear mass of the $\tau$ pair is an important parameter for the signal selection. In $\mu\mu H$ and $eeH$ channel, the fit on the collinear mass improves the accuracy by 1 time, in $qqH$ channel the cut on this mass reduces 90% backgrounds.

The precise measurement might be influenced by the systematic uncertainties coming from luminosity, the fit procedure, and other experimental effects. The integrated luminosity could be measured using small angle radiative Bhabha scattering and the expected precision is better than $10^{-3}$. There is a number of other experimental effects such as acceptance, uncertainties of the $\tau$ finding or the influence of passive detector material. Further quantitative analysis of these effects is still needed. The uncertainty of fitting procedure could be estimated by changing the background shape and fitting range, and the difference in the measurement is taken as the systematic uncertainty.

Table 9 Extrapolated accuracy $\delta (\sigma \times \text{BR})/\langle \sigma \times \text{BR} \rangle$ in ILC 250GeV (2000 fb$^{-1}$)

|          | CEPC | ILC(L) | ILC(R) |
|----------|------|--------|--------|
| Luminosity($ab^{-1}$) | 5.6  | 2      | 2      |
| Polarization($e^-,\gamma^+$) | -   | (0.8, -0.3) | (-0.8, 0.3) |
| Total Higgs | 1.18M | 0.60M  | 0.40M  |
| Accuracy(%) | 0.8  | 1.09   | 1.21   |
6 Appendix 1: TAURUS(TAU ReconstrUction toolS)

The package for τ finding in CEPC is a double cone based algorithm, the steps are:

- Find tracks with energy higher than a defined $E_{\text{min}}$ as the seed
- Collect tracks and photons within an angle ConeA
- Calculate invariant mass with these particles
- Calculate the D0 and Z0 of the leading track
- Calculate the energy in a larger cone ConeB around the seed.

The variables of τ tagging is:

- Number of tracks/photons
- Energy proportion in the smaller cone
- Invariant mass of the $\tau\tau$ system (the particles except for $\tau$s) $M_{\text{qq}}$
- Recoil mass of the qq system (the particles except for $\tau$s) $M_{\text{recoil}}$

Here the parameters $E_{\text{min}}$, ConeA, ConeB, are optimized to the value $\epsilon \cdot p$, where $\epsilon$ is the efficiency of finding $\tau$ in $qq\tau\tau$ events (defined as the number of truth $\tau$ and found divided by the number of truth $\tau$), and $p$ is the purity of the tagged $\tau$s (defined as the number of truth $\tau$ and found divided by the number of tagged $\tau$). The value of these parameters are: $E_{\text{min}} = 1.5$ GeV, ConeA = 0.15 rad, ConeB = 0.45 rad, $M_{\text{min}} = 0.2$ GeV, $M_{\text{max}} = 2.0$ GeV, $R_{\text{En}} = 0.92$.

References

1. T. Behnke, J.E. Brau, P.N. Burrows, et al, The International Linear Collider Technical Design Report-Volume 4: Detectors[J]. arXiv preprint arXiv:1306.6329, 2013.
2. M. Ahmad et al (The CEPC-SPPC Study Group), CEPC Conceptual Design Report: Volume 2 - Physics & Detector, arXiv:1811.10545 [hep-ex].
3. M. Bicer et al., TLEP Design Study Working Group collaboration, First Look at the Physics Case of TLEP, JHEP 01 (2014) 164.
4. CLICdp, CLIC collaboration, M. J. Boland et al., Updated baseline for a staged Compact Linear Collider.
5. M. Ahmad et al (The CEPC-SPPC Study Group), CEPC Conceptual Design Report: Volume 1 - Accelerator , arXiv:1809.00285 [physics.acc-ph].
6. Wolfgang Kilian, Thorsten Ohl, and Juergen Reuter. Whizardâ ˘AˇTsimulating multi- particle processes at lhc and ilc. The European Physical Journal C, 71(9):1742, 2011.
7. P Moras de Freitas et al. MOKKA: A detailed Geant4 simulation for the International Linear Collider detectors, 2003
8. Frank Gaede. Clupatra, Topological TPC pattern recognition, LCWS 2011
9. M. Ruan, Arbor, a new approach of the Particle Flow Algorithm. arXiv:1403.4784 (2014).
10. A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, ... & D. Dannheim (2007), TMVA-Toolkit for multivariate data analysis. arXiv preprint physics: 0703039.
11. Shin-ichi Kawada. Study of Higgs boson decays to tau pairs at the International Linear Collider.