Eur. Phys. J. C (2021) 81:879
https://doi.org/10.1140/epjc/s10052-021-09653-0

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Physics potential for the H→ ZZ* decay at the CEPC

Ryuta Kiuchi[1], Yanxi Gu[2], Min Zhong[2], Lingteng Kong[3], Alex Schuy[4], Shih-Chieh Hsu[4,a], Xin Shi[1,b], Kaili Zhang[1]

[1] Institute of High Energy Physics, Chinese Academy of Science, Beijing 100049, China
[2] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China
[4] Department of Physics, University of Washington, Seattle 98195-1560, USA

Received: 18 March 2021 / Accepted: 13 September 2021 / Published online: 6 October 2021

Abstract The precision of the yield measurement of the Higgs boson decaying into a pair of Z bosons process at the Circular Electron Positron Collider is evaluated. Including the recoil Z boson associated with the Higgs production (Higgsstrahlung) a total of three Z bosons are involved for this channel, from which final states characterized by the presence of a pair of leptons, quarks, and neutrinos are chosen for the signal. Two analysis approaches are compared and the final statistical precision of $\sigma_{\mathrm{ZH}}\cdot\mathrm{BR}(H \to ZZ^*)$ is estimated to be 6.9% using a multivariate analysis technique, based on boosted decision trees. The relative precision of the Higgs boson width, using this $H \to ZZ^*$ decay topology, is estimated by combining the obtained result with the precision of the inclusive ZH cross section measurement.

1 Introduction

After the discovery of the Higgs boson [1,2], a great effort is going into measuring properties of the Higgs boson. One of the motivations for these studies is to obtain hints for physics beyond the Standard Model (SM), whose existence is suggested by several experiment facts, such as dark matter and cosmological baryon–antibaryon asymmetry. The Circular Electron Positron Collider (CEPC) [3,4] is a proposed future circular e^+e^- collider, with a main ring circumference of ∼100 km. As a Higgs factory, the CEPC is planned to operate at a center of mass energy of $\sqrt{s} = 240$ GeV with an integrated luminosity of 5.6 ab^{-1} corresponding to the production of more than 10^6 Higgs bosons. This will allow an order of magnitude improvement on measurements of several Higgs boson properties over the final LHC precision.

The Higgs production mechanisms in e^+e^- collision at $\sqrt{s} = 240$ GeV will be the Higgsstrahlung process $e^+e^- \to Z^* \to ZH$ (hereafter, denoted as ZH process) and the vector boson fusion processes, $e^+e^- \to WW^*\nu_e\bar{\nu}_e \to H\nu_e\bar{\nu}_e$ and $e^+e^- \to ZZ^*e^+e^- \to He^+e^-$. The Higgs production cross section via the ZH process reaches at its maximum at $\sqrt{s}$ of 240–250 GeV and it is dominating over all of the others at this center of mass energy [5]. The tagging of the ZH events using the recoil mass method against the Z boson is unique to lepton colliders. Therefore, the ZH process provides a clean environment for the Higgs boson measurements such as model independent measurement of the inclusive $e^+e^- \to ZH$ production cross section σ_{ZH}. By efficiently identifying the Z boson of the ZH process, individual decay channels of the Higgs boson will be explored subsequently.

The decay of the Higgs boson into a pair of Z bosons, will be of key importance at the CEPC. Like the other decay modes, the Branching ratio $\mathrm{BR}(H \to ZZ^*)$ can be obtained from the measurement of the signal yield, since the yield allows to extract the observable $\sigma_{ZH} \times \mathrm{BR}(H \to ZZ^*)$. In addition, the Higgs boson width Γ_H can be inferred as well. Under the assumption that the coupling structure follows the SM, the branching ratio is proportional to $\mathrm{BR}(H \to ZZ^*) = \Gamma(H \to ZZ^*)/\Gamma_H \propto g^2_{HZZ}/\Gamma_H$, therefore, Γ_H can be deduced with precision determined from the measurements of the coupling g^2_{HZZ} ($\sigma_{ZH} \propto g^2_{HZZ}$) and the signal yield. Note that the cross section of the WW fusion process $\nu\bar{\nu}H$ via the $e^+e^- \to \nu\bar{\nu}H \to \nu\bar{\nu}b\bar{b}$ measurement in combination with measurements of the $\mathrm{BR}(H \to b\bar{b})$ and $\mathrm{BR}(H \to WW^*)$ from the ZH process can also provide the Γ_H value independently, hence the final value will be determined from the combination of the two estimations [5].

The study of $H \to ZZ^*$ channel via the ZH process has an unique feature among the other decays that is origi-

[a] e-mail: schsu@uw.edu (corresponding author)
[b] e-mail: shixin@ihep.ac.cn (corresponding author)

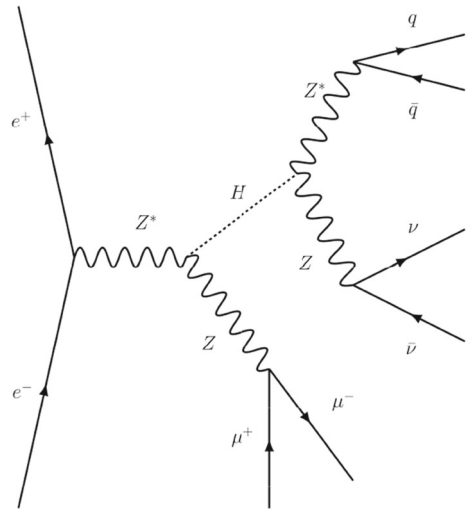

Fig. 1 Example Feynman diagram of the signal process which is characterized by the presence of a pair of muons, jets and neutrinos. In this example, the initial Z boson recoiling against the Higgs boson is decaying into muons. Final states with all of cyclic permutation of the decay products from three Z bosons are considered throughout this analysis

nated from its event topology where two on-shell Z bosons and one off-shell Z boson are involved. Considering that Z bosons can decay to any fermion anti-fermion pair except a top quark pair, the topology diverges into lots of final states. The $H \to ZZ^* \to 4l$ decay is the so-called "golden channel" of the Higgs boson study at the LHC, as it has the cleanest signature of all the possible Higgs boson decay modes [6,7]. However, the statistics of this leptonic channel at the CEPC may not allow to study the properties with required precision. Conversely, fully hadronic channel can provide enough statistics, but difficulties in identifying and matching jets with proper Z bosons, as well as efficient separation from the SM backgrounds have to be overcome. Between these two extremes, the decay channels having a pair of leptons, two jets and two neutrinos are most promising candidates for studying $H \to ZZ^*$ properties, owing to its clear signature and larger branching fraction than the leptonic channel. Therefore, this final state has been chosen as the signal for the evaluation of the $H \to ZZ^*$ properties. Among charged leptons, muons have advantage on discrimination of isolated candidates from those produced via semi-leptonic decays of heavy flavor jets. Therefore, the final states including a pair of muons are finally selected as the signal process: $Z \to \mu^+\mu^-$, $H \to ZZ^* \to \nu\bar{\nu}q\bar{q}$ (Fig. 1) and its cyclic permutations, $Z \to \nu\bar{\nu}$, $H \to ZZ^* \to q\bar{q}\mu^+\mu^-$ and $Z \to q\bar{q}$, $H \to ZZ^* \to \mu^+\mu^-\nu\bar{\nu}$, where the q represents all quark flavors except for the top quark.

In this article, we report on the estimation of relative precision of the yield measurement for the $H \to ZZ^*$ decay at the CEPC using the signal processes characterized by the presence of a pair of muons, jets and neutrinos. In Sect. 2, we briefly introduce the CEPC detector design and the Monte Carlo (MC) simulation scheme. The details of the event selection on generated samples is described in Sect. 3. The statistical procedure and results of the estimated precision of the signal yield is presented in Sect. 4 followed by a brief discussion in Sect. 5. Finally, conclusions are summarized in Sect. 6.

2 Detector design and simulation samples

The CEPC will hosts two interaction points (IP) on the main ring, where the detector at each IP records collision data under different center of mass energies varying from $\sqrt{s} = 91.2$ GeV as a Z factory to $\sqrt{s} = 240$ GeV as a Higgs factory. To fulfill the physics goals, a baseline concept of the detector is developed based on the International Large Detector (ILD) concept [8] with further optimizations for the CEPC environment. From the most inner sub-detector component, the detector concept is composed of a silicon vertex detector, a silicon inner tracker consisting of micro strip detectors, a Time Projection Chamber (TPC), a silicon external tracker, ultra-fine segmented calorimeters, an Electromagnetic CALorimeter (ECAL) and an Hadronic CALorimeter (HCAL), a 3T superconducting solenoid, and a muon detector [4].

The CEPC simulation software package implements the baseline concept detector geometry. Events for the SM processes are generated by the Whizard [9] including the Higgs boson signal, where the detector configuration and response is handled by the GEANT4-based simulation framework, MokkaPlus [10]. Modules for digitization of the signals at each sub detector creates the hit information. Particle reconstruction has been taken place with the Arbor algorithm, which builds the reconstructed particles using calorimeter and track information [11]. A set of MC samples at $\sqrt{s} = 240$ GeV has been generated with this scheme where the Higgs boson signal also contain the WW/ZZ fusion processes. All of the SM background samples, which can be classified according to number of fermions in their final states, two-fermion processes ($e^+e^- \to f\bar{f}$) and four-fermion processes ($e^+e^- \to f\bar{f}f\bar{f}$), are produced as well. More details about the samples and their classification can be found in Ref. [12].

3 Event selection

Event selection is performed in several stages. The preselection builds higher-level objects, such as isolated muons, jets, and missing four momentum from the Particle Flow (PF) objects which are reconstructed by the ArborPFA. The isolation requirements on muons, identified by the PFs, are

imposed. For muons with energy higher than 3 GeV, tracks inside of a cone with a half-angle θ around the candidate are examined and it is identified as an isolated muon, when a ratio of the energy of the muon candidate to summation of the energy from all of the tracks except for the candidate in a volume defined by the cone is greater than 10 with $\cos\theta = 0.98$. Jets are clustered from the PFs but except for isolated lepton candidates, using the k_t algorithm for the e^+e^- collision ($ee-kt$) with the FastJet package [13]. Exclusive requirement ($N_{jet} = 2$) on number of jets is imposed. Events are requested to have a pair of isolated muons of positive and negative charged, and two jets successfully clustered.

The events satisfying the pre-selection criteria are separated into six categories. Depending on which physics objects ($\mu\mu/q\bar{q}/\nu\bar{\nu}$) form the tagged Z boson (hereafter denoted it as initial Z boson), the signal samples can be classified into three categories. Furthermore, distinguishing the status between having a pair of objects suppose to be decaying from the on-shell Z boson and from the off-shell Z boson where $H \to ZZ^*$ decay is assumed, enhances the efficiency of the event selection by applying different selection criteria for each respectively. Following notation is adopted for denoting each category: $\mu\mu\mathrm{H}\nu\nu qq$ ($\mu\mu\mathrm{H}qq\nu\nu$) category is defined to be most sensitive to signal events having reconstructed invariant mass $M_{\mu\mu}$ of two muons in the range 80–100 GeV where two top characters in the notation represent a pair of muons decaying from the initial Z boson, with the reconstructed invariant mass of missing term M_{miss} due to escaping neutrinos is larger (smaller) than dijet invariant mass M_{jj}. The mass range of the initial Z boson for the other categories are chosen as 75–110 GeV for the $\nu\nu\mathrm{H}\mu\mu qq$ and $\nu\nu\mathrm{H}qq\mu\mu$ categories, 75–105 GeV for the $qq\mathrm{H}\nu\nu\mu\mu$ and $qq\mathrm{H}\mu\mu\nu\nu$ categories, taking into account the reconstructed mass resolution for this analysis. The recoil mass against the initial Z boson is required to be in the range of 110–140 GeV. To ensure that the events are separated into categories exclusively, further requirements on recoil mass distributions of a pair of objects are applied that is described later.

On total six categories, $\mu\mu\mathrm{H}\nu\nu qq$, $\mu\mu\mathrm{H}qq\nu\nu$, $\nu\nu\mathrm{H}\mu\mu qq$, $\nu\nu\mathrm{H}qq\mu\mu$, $qq\mathrm{H}\nu\nu\mu\mu$, $qq\mathrm{H}\mu\mu\nu\nu$, further event selection criteria are optimized separately. Two different analysis approaches are exploited for this stage, the one where requirements are imposed on a set of kinematic variables (referred to "cut-based" analysis) and the one which uses a multivariate analysis technique, based on the boosted decision tree (BDT) implemented within scikit-learn package [14], in order to achieve better separation between signal and background (referred to "BDT" analysis).

For the cut-based analysis, the signal to background ratio is maximized by set of criteria. The variable list used in the cut-based analysis is as follows:

- $M_{\mu\mu}$, M_{jj}, M_{miss}: invariant mass of the di-muons, di-jets and the missing mass
- $M^{\mathrm{recoil}}_{\mu\mu}$, M^{recoil}_{jj}, M_{vis}: recoil mass of the di-muons, di-jets and invariant mass of all visible particles
- N_{PFO}: number of particle flow objects
- $\cos\theta_{\mathrm{vis}}$: polar angle of the sum of all visible particles
- $\Delta\phi_{ZZ}$: angle between a Z boson reconstructed from the two muons and that reconstructed from the two jets
- d0, z0: transverse and z impact parameter of the muon
- btag_1, btag_2: b-likeness for each jet.

The reconstructed masses $M_{\mu\mu}$, M_{jj} and M_{miss} are required to fall into the mass window around the Z (Z^*) boson. Number of PFO N_{PFO} in the event is required to be larger than a threshold value, which is decided by the condition whether jets are originated from an on-shell Z boson or not, as well as suppression of background contributions where the jets are reconstructed from any objects other than quark seeds. Cut on the polar angle $\cos\theta_{\mathrm{vis}}$ is applied to further reject background components, such as two-fermion processes which tend to be back-to-back along the beam axis. The angle between the di-muons and di-jets systems $\Delta\phi_{ZZ}$ is used to reduce background components as well. In order to suppress muons originating from in-flight decays of τ-leptons, each muon track is additionally requested to have tighter impact parameters d0, z0 for the $\nu\nu\mathrm{H}qq\mu\mu$ and $qq\mathrm{H}\nu\nu\mu\mu$ categories. Jets containing b-hadrons are identified via the LCFIPlus package [15] with a 80% efficiency working point. In order to reduce the $Z \to \mu^+\mu^-$, $H \to b\bar{b}$ background components, the events having two b-tagged jets are rejected for the $\mu\mu\mathrm{H}qq\nu\nu$ and $qq\mathrm{H}\mu\mu\nu\nu$ categories.

Kinematic properties of two on-shell Z bosons has significant overlap at $\sqrt{s} = 240$ GeV. As a result, a signal process and its *conjugate* process that is the signal process by exchanging decay objects from on-shell Z bosons, e.g. $Z \to \mu^+\mu^-$, $H \to (Z{\to}q\bar{q}, Z^*{\to}\nu\bar{\nu})$ and $Z \to q\bar{q}$, $H \to (Z{\to}\mu^+\mu^-, Z^*{\to}\nu\bar{\nu})$, have considerable overlaps in the kinematical phase space. To ensure that the events are grouped into mutually exclusive categories which are optimized based on their kinematic properties, two exclusive regions in the two-dimensional phase space of recoil mass distributions of di-objects, are defined and are used to further restrict categories. For example, in the $M^{\mathrm{recoil}}_{\mu\mu}$-$M^{\mathrm{recoil}}_{jj}$ phase space, a region covering majority of $Z \to q\bar{q}$, $H \to (Z{\to}\mu^+\mu^-, Z^*{\to}\nu\bar{\nu})$ signal events is defined as

$$M^{\mathrm{recoil}}_{\mu\mu} - M_H > \left|M^{\mathrm{recoil}}_{jj} - M_H\right| \quad (M^{\mathrm{recoil}}_{\mu\mu} > M_H)$$
$$< -\left|M^{\mathrm{recoil}}_{jj} - M_H\right| \quad (M^{\mathrm{recoil}}_{\mu\mu} < M_H)$$

where M_H represents the Higgs boson mass of 125 GeV. A requirement, denoted by "*not-qqHZZ*", has been added to the cut sequence for the $\mu\mu\mathrm{H}qq\nu\nu$ category where events are rejected if a set of reconstructed recoil mass ($M^{\mathrm{recoil}}_{\mu\mu}$,

Table 1 Overview of the requirements applied when selecting events (cut-based)

Pre-selections						
$N(l) = 2$, where leptons(l) should pass the isolation criteria $N(\mu^+) = 1$, $N(\mu^-) = 1$ with $E(\mu^\pm) > 3$ GeV $N(jet) = 2$						
Selection (cut-based)	$\mu\mu$H$\nu\nu qq$	$\mu\mu$H$qq\nu\nu$	$\nu\nu$H$\mu\mu qq$	$\nu\nu$H$qq\mu\mu$	qqH$\nu\nu\mu\mu$	qqH$\mu\mu\nu\nu$
Mass order	$M_{\text{miss}} > M_{jj}$	$M_{\text{miss}} < M_{jj}$	$M_{\mu\mu} > M_{jj}$	$M_{\mu\mu} < M_{jj}$	$M_{\text{miss}} > M_{\mu\mu}$	$M_{\text{miss}} < M_{\mu\mu}$
$M_{\mu\mu}$ (GeV)	[80, 100]		[60, 100]	[10, 60]	[15, 55]	[75, 100]
M_{jj} (GeV)	[15, 60]	[60, 105]	[10, 55]	[60, 100]	[75, 105]	
M_{miss} (GeV)	[75, 105]	[10, 55]	[75, 110]		[70, 110]	[10, 50]
$M_{\mu\mu}^{\text{recoil}}$ (GeV)	[110, 140]		–	–	[175, 215]	[115, 155]
M_{vis} (GeV)	–	[175, 215]	[110, 140]		[115, 155]	[185, 215]
M_{jj}^{recoil} (GeV)	[185, 220]	–	–	–	[110, 140]	
N_{PFO}	[20, 90]	[30, 100]	[20, 60]	[30, 100]	[40, 95]	[40, 95]
$\lvert\cos\theta_{\text{vis}}\rvert$				< 0.95		
$\Delta\phi_{ZZ}$ (degree)	[60, 170]	[60, 170]	< 135	< 135	–	[120, 170]
|d0| and |z0| (mm)	–	–	–	< 0.015	< 0.015	–
btag_1 and btag_2	–	< 0.8	–	–	–	< 0.8
Region masking	*not-*$\nu\nu$*HZZ* and *not-qqHZZ*		*not-*$\mu\mu$*HZZ* and *not-qqHZZ*		*not-*$\nu\nu$*HZZ* and *not-*$\mu\mu$*HZZ*	

M_{jj}^{recoil}) satisfies above condition. Similarly, total two kinds of "*not*-xx*HZZ*" (xx: $\mu\mu$ or $\nu\nu$ or qq) cuts are added in the selection for each category. Table 1 summaries the selection criteria applied across all the categories considered.

The signal and background reduction efficiencies together with expected number of events running at $\sqrt{s} = 240$ GeV corresponding to a total integrated luminosity of 5.6 ab^{-1} after the event selection are listed in the Table 2. For the signal events, Table 2 reports the number of events for the dominant and sub-dominant signal process separately, where the sub-dominant signal process in the category is always the *conjugate* process. The rest of signal processes other than the two processes are not listed in the table since their contributions are found to be very small. In general, the analysis achieves a strong background rejection, while the signal selection efficiencies of approximately 25% and higher are kept.

The major background processes left in each category are following: In the $\mu\mu$H$\nu\nu qq$ category, a dominant source of background is $Z \to \mu^+\mu^-$, $H \to WW^*$, with a W boson decays hadronically and the other W boson decays semileptonically. About 80% of the semileptonic decay is $W \to \tau\nu$. The decay products of the τ-lepton (or a lepton from the semileptonic decay) are merged into jets, resulting in the contamination in the signal region. $Z \to q\bar{q}$, $H \to WW^*$ background is relevant for the $\nu\nu$H$qq\mu\mu$ and qqH$\nu\nu\mu\mu$ categories. The background contamination from the processes where the muons are supplied from semileptonic decays of τ pairs, such as $Z \to q\bar{q}$, $H \to \tau^+\tau^-$ and $e^+e^- \to ZZ \to \tau^+\tau^- q\bar{q}$, arises because of the low invariant mass of di-muons with wider kinematic acceptance allowed for neutrinos for these categories. They are suppressed by the impact parameter requirements on the muons while a fraction of $Z \to q\bar{q}$, $H \to WW^*(\to \mu\nu\mu\nu/\mu\nu\tau\nu)$ events passes the final selection and this process is dominant in the categories. In the qqH$\mu\mu\nu\nu$ and $\mu\mu$H$qq\nu\nu$ categories, $Z \to \mu^+\mu^-$, $H \to WW^*/b\bar{b}$ and $e^+e^- \to ZZ \to \mu^+\mu^- q\bar{q}$ populate the signal regions. Vetoing events with two b-tagged jets reduces the $Z \to \mu^+\mu^-$, $H \to b\bar{b}$ contribution by approximately a factor of two, though the process is still one of the major background sources. The decay mode of $Z \to \mu^+\mu^-$, $H \to WW^*$ background in these categories is the same as in the case of $\mu\mu$H$\nu\nu qq$ category. Despite the difference of kinematic quantities of di-muons and di-jets, the SM four-fermion process is irreducible background to the categories due the large production cross section of $e^+e^- \to ZZ$ together with the low missing four momentum requirement.

For the BDT analysis, simpler selection criteria are applied prior to the BDT discrimination. The invariant and recoil mass of the initial Z boson are required to be in the region of the signal mass window. The selection requirements on the number of particle flow objects and the polar angle of the sum of all visible particles are also applied as used in the cut-based analysis.

A boosted decision tree is then trained on remaining signal and background events for each category separately. The boosting algorithm utilized in this analysis is the AdaBoost scheme [16]. In addition to the selection variables used in the

Table 2 Summary of the selection efficiency ϵ and the number of expected events $N_{evt.}$ for each category after the final event selection in the cut-based analysis. The numbers in parentheses refer to the number of events for dominant background processes in the category

	(Cut-based)	$\mu\mu$H$\nu\nu qq$		$\nu\nu$H$\mu\mu qq$	
	Process	ϵ [%]	$N_{evt.}$	ϵ [%]	$N_{evt.}$
Signal	"dominant"	38	54	54	75
	"sub"	6	8	6	9
Background	Higgs decays (total)	2.4×10^{-3}	28	6.3×10^{-4}	7
	$Z \to \mu^+\mu^-, H \to WW^*$		(25)		(5)
	SM four-fermion	3.7×10^{-6}	4	8.4×10^{-6}	9
	SM two-fermion	0	0	0	0
		$\nu\nu$H$qq\mu\mu$		qqH$\nu\nu\mu\mu$	
	Process	ϵ [%]	$N_{evt.}$	ϵ [%]	$N_{evt.}$
Signal	"dominant"	33	47	23	32
	"sub"	8	11	6	8
Background	Higgs decays (total)	3.5×10^{-3}	40	4.4×10^{-3}	50
	$Z \to q\bar{q}, H \to WW^*$		(24)		(32)
	SM four-fermion	1.0×10^{-5}	11	1.8×10^{-5}	19
	SM two-fermion	0	0	0	0
		qqH$\mu\mu\nu\nu$		$\mu\mu$H$qq\nu\nu$	
	Process	ϵ [%]	$N_{evt.}$	ϵ [%]	$N_{evt.}$
Signal	"dominant"	20	28	31	44
	"sub"	4	5	7	10
Background	Higgs decays (total)	8.2×10^{-3}	94	4.8×10^{-2}	546
	$Z \to \mu^+\mu^-, H \to WW^*$		(40)		(313)
	$Z \to \mu^+\mu^-, H \to b\bar{b}$		(39)		(211)
	SM four-fermion (total)	1.2×10^{-4}	129	3.0×10^{-4}	326
	$e^+e^- \to ZZ \to \mu^+\mu^- q\bar{q}$		(127)		(317)
	SM two-fermion	0	0	0	0

cut-based analysis, several variables are added to the set of input variables for the training:

- $M_{\mu\mu}$, M_{jj}, M_{miss}, $M_{\mu\mu}^{\text{recoil}}$, M_{jj}^{recoil}, M_{vis}
- N_{PFO}, $\cos\theta_{\text{vis}}$, $\Delta\phi_{ZZ}$, $\text{d0}_{\mu 1}$, $\text{d0}_{\mu 2}$, $\text{z0}_{\mu 1}$, $\text{z0}_{\mu 2}$
- $P_{\mu\mu}$, P_{jj}, P_{vis}, $P_{t,\text{vis}}$: magnitude of the momentum and transverse momentum
- E_j^{lead}, E_j^{sub}, $P_{t,j}^{lead}$, $P_{t,j}^{sub}$: energy and transverse momentum of the leading/sub-leading jet
- btag_1, btag_2, ctag_1, ctag_2: b/c-likeness for each jet
- y_{12}, y_{23}, y_{34}, y_{45}: jet transition values.

The analysis exploits the increased sensitivity by combining total 28 input variables into the final BDT discriminant. Figure 2 shows the obtained BDT score distributions for signal and background samples. For the final separation of signal and background events, the cut value on the BDT score is chosen so as to maximize a significance measure $S/\sqrt{S+B}$, where for a chosen cut, S (B) is the number of signal (background) events above this cut. The signal selection efficiency of the analysis with this cut value ranges from 21 to 67% which are slightly higher than that from the cut-based analysis, except the qqH$\nu\nu\mu\mu$ category while a comparable ($\nu\nu$H$\mu\mu qq$ category) or better significance measure is obtained. The cut value on the BDT score with the final signal selection efficiency as well as the other selection criteria are summarized in Table 3.

4 Result

An unbinned maximum likelihood fit is performed to extract the signal yield for each of six categories. The obtained signal and background distributions of recoil mass spectrum M_Z^{recoil} against the initial Z boson in the range 110–140 GeV, are added to make up a pseudo-experimental result, while the likelihood template is constructed from sum of the Probability Density Function (PDF) describing the distributions of M_Z^{recoil} for the signal and the background individually. The normalized distribution of M_Z^{recoil} for signal events in a cat-

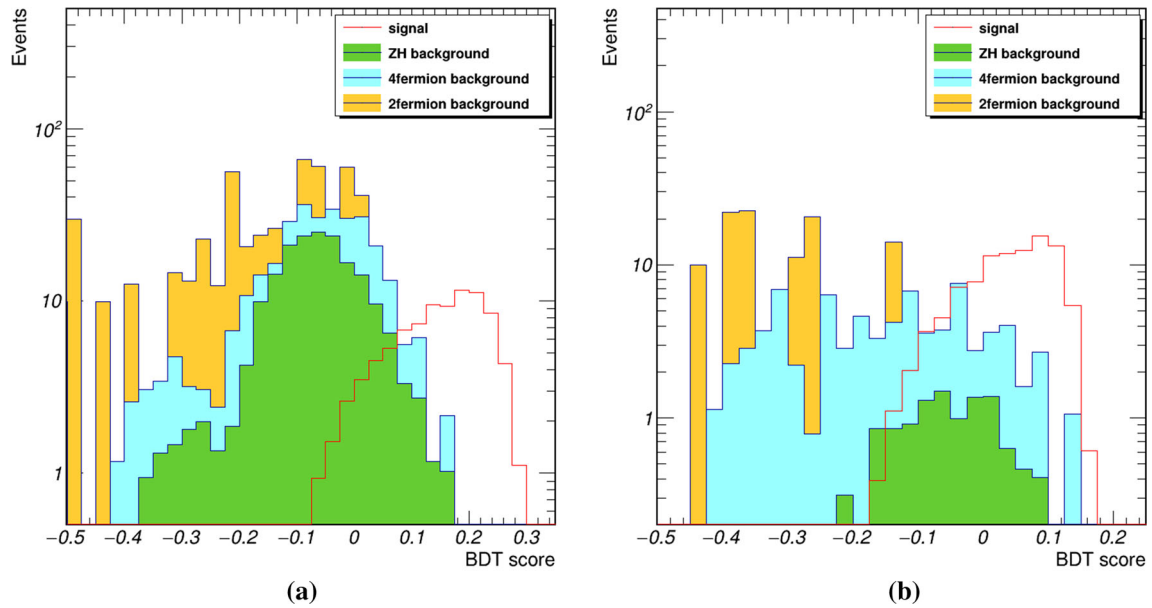

Fig. 2 BDT score distributions for two of most sensitive categories: **a** $\mu\mu\mathrm{H}\nu\nu qq^{\mathrm{mva}}$ category and **b** $\nu\nu\mathrm{H}\mu\mu qq^{\mathrm{mva}}$ category. The signal distribution is shown with a red histogram while background contributions, ZH (green), four-fermion (cyan) and two-fermion (yellow), are drawn

Table 3 Overview of the requirements applied when selecting events (BDT-based). The last row shows the final selection efficiency for the signal where the dominant and sub-dominant signal processes are summed

Pre-selections						
$N(l) = 2$, where leptons(l) should pass the isolation criteria $N(\mu^+) = 1$, $N(\mu^-) = 1$ with $E(\mu^\pm) > 3$ GeV $N(jet) = 2$						
Selection (BDT)	$\mu\mu\mathrm{H}\nu\nu qq$	$\mu\mu\mathrm{H}qq\nu\nu$	$\nu\nu\mathrm{H}\mu\mu qq$	$\nu\nu\mathrm{H}qq\mu\mu$	$qq\mathrm{H}\nu\nu\mu\mu$	$qq\mathrm{H}\mu\mu\nu\nu$
Mass order	$M_{\mathrm{miss}} > M_{jj}$	$M_{\mathrm{miss}} < M_{jj}$	$M_{\mu\mu} > M_{jj}$	$M_{\mu\mu} < M_{jj}$	$M_{\mathrm{miss}} > M_{\mu\mu}$	$M_{\mathrm{miss}} < M_{\mu\mu}$
$M_{\mu\mu}$ (GeV)	[80, 100]		–	–	–	–
M_{jj} (GeV)	–	< 105	–	–	[75, 105]	
M_{miss} (GeV)	–	–	[75, 110]		–	–
$M_{\mu\mu}^{\mathrm{recoil}}$ (GeV)	[110, 140]	[120, 140]	–	–	–	> 110
M_{vis} (GeV)	–	–	[110, 140]		–	–
M_{jj}^{recoil} (GeV)	–	–	–	–	[110, 140]	
N_{PFO}	[20, 90]	[30, 100]	[20, 60]	[30, 100]	[40, 95]	[40, 95]
$\lvert\cos\theta_{\mathrm{vis}}\rvert$	< 0.95					
Region masking	*not-*$\nu\nu$*HZZ* and *not-qqHZZ*		*not-*$\mu\mu$*HZZ* and *not-qqHZZ*		*not-*$\nu\nu$*HZZ* and *not-*$\mu\mu$*HZZ*	
BDT score	> 0.06	> -0.01	> -0.10	> -0.00	> 0.01	> -0.03
Signal Eff. ϵ (%)	52	42	67	42	21	26

egory is described by sum of a double sided Crystal Ball function and small Gaussian tails for the signal process with the initial Z boson decaying to di-muon and a Breit–Wigner function convolved with a Gaussian for the rest of signal processes. For the SM background components, a continuous PDF is constructed using the kernel density estimation technique [17] for each component. The background events from the other Higgs decay channels are modeled by the same PDF as the signal in terms of decay objects from the initial Z boson, except for the channels having small number of entries (< 20) where a PDF from the kernel density estimation is used to describe the shape.The background components mentioned above are combined according to their fraction and are normalized to the number of events left in the category. The template model used to the likelihood fit is then expressed as $\mu \cdot N_{sig} \cdot f_{sig} + N_{bkg} \cdot f_{bkg}$, where f_{sig} (f_{bkg}), N_{sig} (N_{bkg}) are the combined PDF and total number of events for signal (background) events, μ is a free parameter determined by

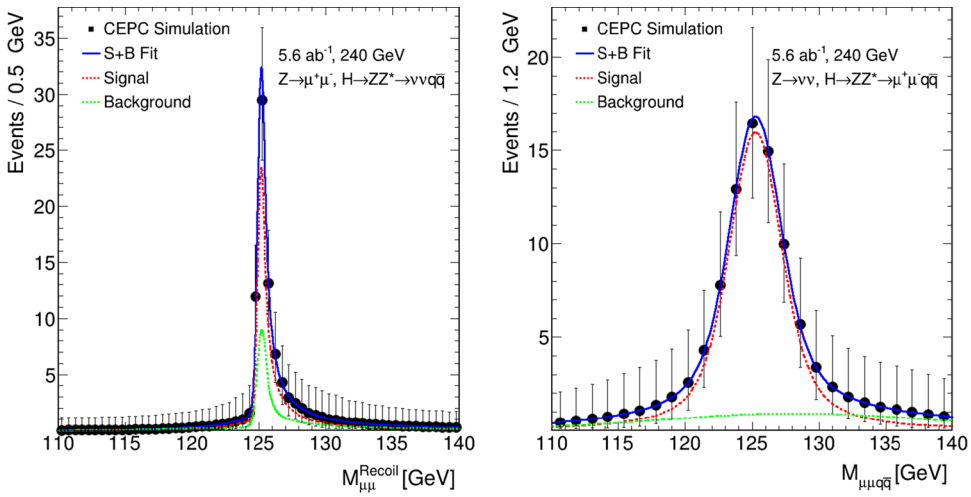

Fig. 3 Recoil mass distributions in the **a** $\mu\mu\mathrm{H}\nu\nu qq^{\mathrm{cut}}$ and **b** $\nu\nu\mathrm{H}\mu\mu qq^{\mathrm{cut}}$ categories from the cut-based analysis. The black dots represent the predicted results at the CEPC and the solid blue line shows the fitted model which is broken down into signal (dashed red line) and background (dashed green line) components

the fit. Note that nuisance parameters, such as uncertainty of the total luminosity, are fixed to the expected values, so that those systematic uncertainties are ignored at this stage. The recoil mass distribution together with the fitting results for two of the most sensitive categories from the cut-based analysis is shown in Fig. 3. The dominant background component observed in the Fig. 3a is the $Z \to \mu^+\mu^-$, $H \to WW^*$ process which shows the same position of the Higgs signal mass peak.

The number of expected signal events can be simply represented by $N_{sig} = \mathcal{L}\cdot\epsilon\cdot\sigma_{ZH}\cdot\mathrm{BR}(H \to ZZ^*) \cdot \prod_{X=\mu,\nu,q} \mathrm{BR}(Z \to X\bar{X})$, where $\mathcal{L}$ is the total luminosity and ϵ represents efficiencies including the detector acceptance and the analysis selection. The uncertainty of the fitting parameter μ is then regarded as the statistical uncertainty of $\sigma_{ZH}\cdot\mathrm{BR}(H \to ZZ^*)$ by neglecting other systematic uncertainties. Table 4 summarizes the derived relative precision on the product of the inclusive ZH cross section and the branching ratio $\Delta(\sigma\cdot\mathrm{BR})/(\sigma\cdot\mathrm{BR})$ from the cut-based analysis and the BDT analysis. The bottom row shows the combined precision that is calculated from the standard error of the weighted mean, $\sigma = 1/\sqrt{\sum_{i=1}^{n}\sigma_i^{-2}}$, where σ_i is the precision for each category. The final result for the relative statistical uncertainty of the $\sigma_{ZH} \times \mathrm{BR}(H \to ZZ^*)$ is estimated to be 7.7% for the cut-based analysis and 6.9% for the BDT analysis.

The systematic uncertainty related to the analysis is discussed in the following. Several sources of systematic uncertainties on Higgs measurements at the CEPC (e.g. the uncertainty on $\mathcal{L}$, σ_{ZH}, etc.) are described in Ref. [18]. Although the study in Ref. [18] has been performed with a slightly different detector configuration and operation scenario, the order of magnitude of these estimated systematic uncertainties $\mathcal{O}(0.1)\%$ can be also assumed for the current HZZ analysis, and thus their contributions are ignored.

The uncertainty related to the number of background events due to a potential contamination of two-fermion processes is evaluated. Those processes have relatively large scaling factor, which is the ratio between an expected and analyzed number of events, owing to the limited statistics of generated MC samples. Although the process are completely rejected during the event selection for both analyses, a fraction of events may survive the final selection with increased statistics of MC samples, especially for the two categories $\mu\mu\mathrm{H}\nu\nu qq$ and $\nu\nu\mathrm{H}\mu\mu qq$ where two-fermion background events remain until the later selection stage. Assuming one two-fermion process entry exists after the final event selection for the two categories with its scaling factor of 10 that is calculated from statistics of the data for $e^+e^- \to \mu^+\mu^-$ process, the precision of both categories is scaled down by the change of the measure $S/\sqrt{S+B}$ when the number of background count is increased by 10. After combination with the other four categories, the final calculated precision on the signal yield is estimated as 8.0% for the cut-based analysis and 7.1% for the BDT analysis.

During the signal fitting procedure, the signal strength μ for the HZZ signal yield is the only parameter to be determined while the other Higgs decay BRs are fixed to the expected values from the SM. In order to estimate the impact of uncertainties in the measurements of other Higgs decays on the final results, the simultaneous fit is performed with the other Higgs BRs being as floating as well. Each range over which the Higgs BR is allowed to change in the fit is restricted by the expected precision at the CEPC listed in Ref. [5] and the resulting precision of the HZZ signal yield is obtained as 8.6% for the cut-based analysis and 7.7% for the BDT analysis. Regarding the worsening of the precision from the statistical only uncertainty as a quadratic sum of the systematic effect, the HZZ signal yield precision after combining two systematic effects to the statistical uncertainty is

Table 4 Statistical uncertainties on the product of the ZH cross section and the branching ratio. The bottom row shows the result of combined value of the six categories

Category	$\frac{\Delta(\sigma \cdot BR)}{(\sigma \cdot BR)}$ [%]	
	Cut-based	BDT
$\mu\mu \mathrm{H} \nu\nu qq^{\mathrm{cut/mva}}$	15	13
$\mu\mu \mathrm{H} qq \nu\nu^{\mathrm{cut/mva}}$	46	37
$\nu\nu \mathrm{H} \mu\mu qq^{\mathrm{cut/mva}}$	12	12
$\nu\nu \mathrm{H} qq \mu\mu^{\mathrm{cut/mva}}$	18	15
$qq \mathrm{H} \nu\nu \mu\mu^{\mathrm{cut/mva}}$	24	24
$qq \mathrm{H} \mu\mu \nu\nu^{\mathrm{cut/mva}}$	48	32
Combined	7.7	6.9

calculated as 8.8% for the cut-based analysis and 7.9% for the BDT analysis.

The signal yield $\sigma_{ZH} \cdot \mathrm{BR}(H \rightarrow ZZ^*)$ combined with independently determined σ_{ZH} allows the Higgs width Γ_H to be extracted as described in Sect. 1. Hence the precision of the Higgs width can be evaluated from the $H \rightarrow ZZ^*$ decay channel. Using the following relationship

$$\sigma_{ZH} \cdot \mathrm{BR}(H \rightarrow ZZ^*) \propto g^2_{HZZ} \cdot \frac{\Gamma(H \rightarrow ZZ^*)}{\Gamma_H} \propto \frac{g^4_{HZZ}}{\Gamma_H}$$

the relative uncertainty of the extracted Higgs width is obtained where the relative uncertainty on square of the coupling g^2_{HZZ} of 0.5% taken from Ref. [5] is assumed. From the BDT analysis, it is 7.9% (6.9%) using the precision on the HZZ signal yield with (without) systematics. As mentioned in Sect. 1, the measurement of the $\mathrm{BR}(H \rightarrow b\bar{b})$, $\mathrm{BR}(H \rightarrow WW^*)$ and the WW fusion cross section via the $e^+e^- \rightarrow \nu\bar{\nu}H \rightarrow \nu\bar{\nu}b\bar{b}$ channel will give another estimation on the precision of the Higgs width in the similar manner discussed above. It is shown that the precision using the WW fusion process is expected to reach 3.5% at the CEPC [5], therefore, the final combined precision of the Higgs width is dominated by the HWW vertex measurements. It should be also mentioned that the effective field theory (EFT) is also widely accepted as an alternative approach to explore the Higgs couplings, where additional terms for the interaction between Higgs and Z boson in the Lagrangian collapse the simple picture above [5,19].

5 Discussion

Our estimation of the precision of the yield measurement $\sigma_{ZH} \times \mathrm{BR}(H \rightarrow ZZ^*)$ does not reach the level 5.1% reported in Ref. [5]. Although the definition of the signal regions and the signal processing is slightly different, the recoil mass distributions in the previous work (Fig.13 in Ref. [5]) can be compared with the obtained distributions as shown in Fig. 3. A noticeable difference between the two results resides in the lower event selection efficiencies, and consequently the lower signal rates as well as more background events in our analysis. A possible difference may exist on more sophisticated treatment of background estimation in current analysis. Although improvement on the precision could be achieved by performing further elaborated studies to suppress backgrounds more effectively, the improvement is also expected by considering more final states of the $H \rightarrow ZZ^*$ decay in the ZH process since only small fraction ($< 3\%$) of the entire decay events has been chosen as signals and analyzed.

Broadening analysis channel of the $H \rightarrow ZZ^*$ decay will provide crucial qualitative improvements in studying other HZZ related topics as well. For example, the application of EFT frameworks on the HZZ decay vertex for the study of Higgs CP properties and anomalous couplings to gauge bosons in the presence of beyond the SM physics, has been discussed so far on the production channel ($ee \rightarrow Z^* \rightarrow ZH$) for future lepton colliders [20,21]. Increasing the signal statics under sufficient background rejection will allow further study on the $H \rightarrow ZZ^*$ vertex directly.

6 Summary

The precision of the yield measurement $\sigma_{ZH} \times \mathrm{BR}(H \rightarrow ZZ^*)$ at the CEPC is evaluated using MC samples for the baseline concept running at $\sqrt{s} = 240$ GeV with an integrated luminosity of 5.6 ab^{-1}. Among the various decay modes of the $H \rightarrow ZZ^*$, the signal process having two muons, two jets and missing momentum in final states has been chosen. After the event selection, relative precision is evaluated with the likelihood fitting method on signal and background. The final statistical uncertainty combined from all of six categories is 7.7% from the cut-based analysis and 6.9% from the BDT analysis. The relative precision of the Higgs boson width from the $H \rightarrow ZZ^*$ measurement, is estimated to be 7.9% from the BDT analysis by combining the uncertainty on $\sigma_{ZH} \times \mathrm{BR}(H \rightarrow ZZ^*)$ including the systematic effects with the precision of the inclusive ZH cross section measurement.

Acknowledgements The authors would like to thank the CEPC computing team for providing the simulation tools and MC samples. We also thank Yaquan Fang, Manqi Ruan, Gang Li, and Yuhang Tan for helpful discussions.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The data associated with this paper are available from the corresponding author on reasonable request.]

Funded by SCOAP[3].

References

1. ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B **716**, 1–29 (2012). arXiv:1207.7214 [hep-ex]
2. CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B **716**, 30–61 (2012). arXiv:1207.7235 [hep-ex]
3. The CEPC Study Group, *CEPC Conceptual Design Report: Volume 1—Accelerator* (2018). arXiv:1809.00285 [physics.acc-ph]
4. The CEPC Study Group, *CEPC Conceptual Design Report: Volume 2—Physics & Detector* (2018). arXiv:1811.10545 [hep-ex]
5. F. An et al., Precision Higgs physics at the CEPC. Chin. Phys. C **43**, 043002 (2019). arXiv:1810.09037 [hep-ex]
6. ATLAS Collaboration, Measurement of the Higgs boson coupling properties in the $H \to ZZ^* \to 4l$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector. J. High Energy Phys. **03**, 095 (2018). arXiv:1712.02304 [hep-ex]
7. CMS Collaboration, Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV. J. High Energy Phys. **11**, 047 (2017). arXiv:1706.09936 [hep-ex]
8. T. Behnke et al., *The International Linear Collider Technical Design Report—Volume 4: Detectors* (2013). arXiv:1306.6329 [physics.ins-det]
9. W. Kilian, T. Ohl, J. Reuter, WHIZARD—simulating multi-particle processes at LHC and ILC. Eur. Phys. J. C **71**, 1742 (2011). arXiv:0708.4233 [hep-ph]
10. P. Mora de Freitas, H. Videau, Detector simulation with MOKKA/GEANT4: present and future. In *Presented at the International Workshop on Linear Colliders (LCWS 2002)* (2002), pp. 623–627. https://inspirehep.net/literature/609687
11. M. Ruan et al., Reconstruction of physics objects at the Circular Electron Positron Collider with Arbor. Eur. Phys. J. C **78**, 426 (2018). arXiv:1806.04879 [hep-ex]
12. X. Mo et al., Physics cross sections and event generation of e^+e^- annihiliations at the CEPC. Chin. Phys. C **40**, 033001 (2016). arXiv:1505.01008 [hep-ex]
13. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. C **72**, 1896 (2012). arXiv:1111.6097 [hep-ph]
14. F. Pedregosa et al., Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011). arXiv:1201.0490 [cs.LG]
15. T. Suehara, T. Tanabe, LCFIPlus: a framework for jet analysis in linear collider studies. Nucl. Instrum. Meth.ods **A808**, 109–116 (2016). arXiv:1506.08371 [physics.ins-det]
16. Y. Freund, R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**, 119–139 (1997)
17. K. Cranmer, Kernel estimation in high-energy physics. Comput. Phys. Commun. **136**, 198–207 (2001). arXiv:hep-ex/0011057
18. Z. Chen et al., Cross section and Higgs mass measurement with Higgsstrahlung at the CEPC. Chin. Phys. C **41**, 023003 (2017). arXiv:1601.05352 [hep-ex]
19. T. Barklow et al., Improved formalism for precision Higgs coupling fits. Phys. Rev. D **97**, 053003 (2018). arXiv:1708.08912 [hep-ph]
20. I. Anderson et al., Constraining anomalous HVV interactions at proton and lepton colliders. Phys. Rev. D **89**, 035007 (2014). arXiv:1309.4819 [hep-ph]
21. N. Craig et al., Beyond Higgs couplings: probing the Higgs with angular observables at future e^+e^- colliders. J. High Energy Phys. **03**, 050 (2016). arXiv:1512.06877 [hep-ph]