

The expected measured precision of the branching ratio of the Higgs decaying to the di-photon at the CEPC

Fangyi Guo^{a,b}, Yaquan Fang^{a,b}, Gang Li^{a,b} and Xinchou Lou^{a,b,c}

^a*Institute of High Energy Physics (IHEP), Chinese Academy of Science, Beijing, 100049, China*

^b*University of Chinese Academy of Science, Beijing, 100049, China*

^c*University of Texas at Dallas, Richards, 75080-3021, TX, USA*

ARTICLE INFO

Keywords:

CEPC

Higgs

di-photon

ABSTRACT

This paper presents the prospects of measuring $Br(H \rightarrow \gamma\gamma)$ of the Standard Model Higgs boson via the processes $e^+e^- \rightarrow ZH$ in which $H \rightarrow \gamma\gamma$, $Z \rightarrow q\bar{q}/\mu^+\mu^-/\nu\bar{\nu}$ using the baseline conceptual detector with $\sqrt{s} = 240\text{GeV}$ at the Circular Electron Positron Collider (CEPC). The simulated Monte Carlo events are generated and scaled to an integrated luminosity of 5.6ab^{-1} to mimic the data. Extrapolated results to 20ab^{-1} are also shown. The expected statistical precision of this measurement after combining 3 channels of Z boson decay is 7.7%. With some preliminary estimation on the systematical uncertainty, the total precision is 9.4%. The performance of CEPC electro-magnetic calorimeter (ECAL) is studied by smearing the photon energy resolution in simulated events in $e^+e^- \rightarrow ZH \rightarrow q\bar{q}\gamma\gamma$ channel. In present ECAL design, the stochastic term in resolution plays the dominant role in the precision of Higgs measurements in $H \rightarrow \gamma\gamma$ channel. The impact of the resolution on the measured precision of $Br(ZH \rightarrow q\bar{q}\gamma\gamma)$ as well as the optimization of ECAL constant term and stochastic term are studied for the further detector design.

1. Introduction


In 2012, the ATLAS and CMS collaboration announced the discovery of Higgs Boson at the Large Hadron Collider (LHC) [1, 2]. In the following years the precise measurements of Higgs properties become one of the main goals in particle physics, hoping to answer the remaining basic questions in nature and find the new physics. For this purpose, the hadron collider like LHC may not be the best choice due to large amount of background processes and corresponding lower ratio between the signals and backgrounds. Instead, a lepton collider machine can provide cleaner experiment environment and well-known initial states, which is crucial for high precision studies to find the hints of new physics. Thus several future lepton collider experiments are proposed, including the International Linear Collider (ILC) [3], the Circular Electron Positron Collider (CEPC) [4], the Future Circular Collider e^+e^- (FCC-ee) [5], and the Compact Linear Collider (CLIC) [6].

The CEPC is designed to be a circular lepton collider hosted in a tunnel with a circumference of 100 km and operate at a center of mass energy $\sqrt{s} = 240\text{GeV}$ as a Higgs factory. After 10 years running period, the CEPC will collect 5.6ab^{-1} data, corresponding to more than 1 million Higgs boson. With this clean and large Higgs sample, the precision for the measurements of the Higgs properties is expected to be enhanced with one order of magnitude comparing with the LHC [7].

The Higgs boson interacts with photon through the top quark loop and massive boson loop. This mechanism gives low $H \rightarrow \gamma\gamma$ branching ratio in the Standard Model (SM) but also makes it a good channel to test the new physics beyond

the SM. Besides, high energy photons from the Higgs boson decay can be identified and measured well in the detector, providing a high signal-over-background ratio. So this channel also serves as a good benchmark for the performance of the electromagnetic calorimeter (ECAL) study. Current measurement of the inclusive Higgs boson signal strength in the diphoton channel in LHC is $1.04^{+0.10}_{-0.09}$ in ATLAS [8] and $1.03^{+0.11}_{-0.09}$ in CMS [9] with the pp collision data collected by ATLAS and CMS from 2015 to 2018. The results are consistent with the SM prediction with the present precision. In the HL-LHC period the ATLAS expects to collect 3ab^{-1} data. The projected precision of the $H \rightarrow \gamma\gamma$ signal strength is 6.0% to 3.7% depending on different considerations of the systematic uncertainties S1 or S2 in [10]. After the combination with CMS, a total precision of 2.6% can be reached in the optimistic systematic scenario S2.

A previous analysis studied the expected Higgs precision in various Higgs decay channels [7] including $H \rightarrow \gamma\gamma$. A result of 6.8% precision in $\sigma(ZH) \times Br(H \rightarrow \gamma\gamma)$ is provided for CEPC-v4 concept. However, this result is based on the fast simulation of Monte Carlo samples and cut-based analysis method. In recent study [11] the CEPC accelerator study group proposed an update on the radiation power, resulting in an increase of the instantaneous luminosity of 66%. Based on this a new nominal data-taking scenario is developed. It aims at ten years of data taking at $\sqrt{s} = 240\text{GeV}$ with two interaction points (IPs), accumulating an integrated luminosity of 20ab^{-1} Higgs data [12]. And one new conceptual detector design is also on-going. A homogeneous ECAL is considered to replace the previous silicon-tungsten sampling calorimeter [12, 13, 14]. So it is worth to re-study the $H \rightarrow \gamma\gamma$ process with the latest benchmark, update the analysis method from the cut-based

 guofangyi@ihep.ac.cn (F. Guo)

ORCID(s): 0000-0002-3864-9257 (F. Guo)

to multi-variable analysis (MVA) and investigate the impact from the detector.

This paper is organized as following. Sec. 2 briefly introduces the CEPC detector and the simulated Monte-Carlo samples used in this analysis. Sec. 3 presents the object reconstructions and event selections. Sec. 4 describes the developed MVA method for this work. Sec. 5 studies the signal and background models. The results are summarized in Sec. 7. In Sec. 8 we investigate how these results can be influenced by the CEPC ECAL resolution, that can be a guide for detector optimization. The conclusions are drawn in Sec. 9.

2. CEPC detector and Monte-Carlo simulation

The CEPC detector is designed to match the physics goals that all possible final states can be separately identified and reconstructed with high resolution. The baseline detector concept utilizes the particle flow approach (PFA) idea [15], with a precise vertex detector, a Time Projection Chamber (TPC), a silicon tracker, a high granularity Silicon-Tungsten sampling ECAL and a GRPC-based high granularity hadronic calorimeter (HCAL). All the system is imbedded in 3 Tesla magnetic field. The outermost of the detector is a muon chamber. The details can be found in Ref. [4].

The Higgs production mechanisms at the CEPC are Higgs-strahlung $e^+e^- \rightarrow ZH$, W/Z fusion $e^+e^- \rightarrow \nu\bar{\nu}H$ and $e^+e^- \rightarrow e^+e^-H$ as illustrated in Figure 1. In this analysis Higgs production via ZH process decaying to diphoton final state $e^+e^- \rightarrow ZH \rightarrow f\bar{f}\gamma\gamma$ at $\sqrt{s} = 240$ GeV is considered as the dominant signal. It is further divided into 3 sub-channels, depending on Z decaying to $q\bar{q}$, $\mu^+\mu^-$ and $\nu\bar{\nu}$. $Z \rightarrow e^+e^-$ channel is abandoned due to the known extremely large Bhabha background, and $Z \rightarrow \tau^+\tau^-$ channel is dropped as well because of the complexity of τ jet identification. W/Z fusion process is counted in ZH , $Z \rightarrow \nu\bar{\nu}$ sub-channel. The background process only counts the 2-fermion background $e^+e^- \rightarrow f\bar{f}$ in CEPC with at least 2 radiation photons. The Higgs resonant background, 4-fermion processes and possible reducible background in the experiment are expected to be negligible. These SM physics processes are generated with Whizard [16] at leading order (LO) interfaced with Pythia 6 [17] for parton showering and hadronization with parameters based on the Large Electron Positron Collider (LEP) [18] data. Initial state radiation (ISR) and final state radiation (FSR) effects are taken into account. The total energy spread caused by beamstrahlung and synchrotron radiation is studied by Monte-Carlo simulation and determined to be 0.1629% at CEPC [19]. Table 1 lists the cross sections of physics processes and MC sample statistics used in the analysis. Event yields are normalized to 5.6 ab^{-1} . Detailed configurations can be found in Ref. [20].

The simulation of detector configuration and response is handled by MokkaPlus [21], a GEANT4 [22] based framework. The full detector simulation is performed for signal

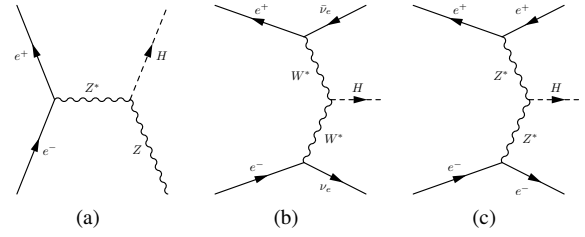


Figure 1: Feynman diagrams of the Higgs boson production processes at the CEPC: (a) $e^+e^- \rightarrow ZH$, (b) $e^+e^- \rightarrow \nu\bar{\nu}H$ and (c) $e^+e^- \rightarrow e^+e^-H$.

Process	σ	statistics
<i>q$\bar{q}$$\gamma\gamma$ sub-channel</i>		
$e^+e^- \rightarrow ZH \rightarrow q\bar{q}\gamma\gamma$	0.31 fb	100 k
$e^+e^- \rightarrow q\bar{q}$	54.1 pb	20 M
<i>$\mu^+\mu^-\gamma\gamma$ sub-channel</i>		
$e^+e^- \rightarrow ZH \rightarrow \mu^+\mu^-\gamma\gamma$	0.15 fb	100 k
$e^+e^- \rightarrow \mu^+\mu^-$	5.3 pb	20 M
<i>$\nu\bar{\nu}\gamma\gamma$ sub-channel</i>		
$e^+e^- \rightarrow ZH \rightarrow \nu\bar{\nu}\gamma\gamma$	0.11 fb	100 k
$e^+e^- \rightarrow \nu\bar{\nu}H \rightarrow \nu\bar{\nu}\gamma\gamma$		
$e^+e^- \rightarrow \nu\bar{\nu}$	54.1 pb	20 M

Table 1

Cross section and the simulated MC sample statistics. In $q\bar{q}\gamma\gamma$ and $\mu^+\mu^-\gamma\gamma$ channels ZH is the only considered process, and in $\nu\bar{\nu}\gamma\gamma$ channel both ZH $Z \rightarrow \text{inv.}$ and W/Z fusion processes are considered.

process only. The background process uses a fast simulation which smears the truth particles with the parameterized detector resolution and efficiency to save the computing resource.

3. Object reconstruction and event selection

The CEPC follows the PFA scheme in the event reconstruction, with a dedicated toolkit ARBOR [23, 24]. The tracks are firstly reconstructed with the hits in the tracking detector by Clupatra module [25]. Then ARBOR collects the tracks from Clupatra and hits in calorimeter, and composes the Particle Flow Objects (PFOs) by its clustering and matching modules. These PFOs are identified as charged particles, photons, neutral hadrons and unassociated fragments. With this approach the photon is identified with an EM-like cluster in calorimeter without any matched track. Converted photons are not considered yet, which counts 5-10% in central region and 25% in forward region [4]. The lepton (e^\pm , μ^\pm) is defined by a track-matched particle. A likelihood-based algorithm, LICH [26], is implemented in ARBOR to separate electrons, muons and hadrons. Jets are formed from the particles reconstructed by ARBOR with the Durham clustering algorithm [27] after excluding the interested particles. The jet energy is calibrated by the MC simulation currently, while is foreseen to be re-calibrated with physics events like $W \rightarrow q\bar{q}$ and/or $Z \rightarrow q\bar{q}$ in

Selections	Higgs signal	$q\bar{q}\gamma\gamma$ background
Exclusive 2 jets and 2 photons	85.56%	69.57%
$E_{\gamma 1} > 25$ GeV	100.00%	2.35 %
$E_{\gamma 2} \in [35, 95]$ GeV	98.37%	35.33%
$\cos \theta_{\gamma\gamma} > -0.95$	95.20%	68.01%
$\cos \theta_{jj} > -0.95$	90.86%	85.54%
$pT_{\gamma 1} > 20$ GeV	93.42%	56.94%
$pT_{\gamma 2} > 30$ GeV	93.25%	54.54%
$m_{\gamma\gamma} \in [110, 140]$ GeV	97.50%	21.14%
$E_{\gamma\gamma} > 120$ GeV	99.47%	98.41%
$\min \cos \theta_{\gamma j} < 0.9$	71.67%	48.05%
Total eff	44.08%	0.01%
Yields in 5.6 ab^{-1}	766.64	26849.38

Table 2

Selection criteria and the corresponding efficiencies in $q\bar{q}\gamma\gamma$ channel. $\gamma 1(\gamma 2)$ is defined as the photon with lower (higher) energy. $\cos \theta_{\gamma\gamma}(\cos \theta_{jj})$ is the polar angle of di-photon (di-jet) system. $\min|\cos \theta_{\gamma j}|$ is the minimum $\cos \theta$ of the photon-jet pairs.

Selections	Higgs signal	$\mu^+\mu^-\gamma\gamma$ background
Exclusive 2 muons and 2 photons	70.18%	5.18%
$E_{\gamma} > 35$ GeV	99.21%	8.39%
$ \cos \theta_{\gamma} < 0.9$	83.79%	38.14%
$pT_{\gamma 1} \in [10, 70]$ GeV	99.84%	86.30%
$pT_{\gamma 2} \in [30, 100]$ GeV	99.96%	95.59%
$m_{\gamma\gamma} \in [110, 140]$ GeV	98.08%	37.62%
$M_{\gamma\gamma}^{\text{recoil}} \in [85, 105]$ GeV	80.12%	21.29%
$E_{\gamma\gamma} \in [125, 145]$ GeV	99.88%	95.86%
Total eff	45.69%	0.01%
Yields in 5.6 ab^{-1}	39.32	2662.77

Table 3

Selection criteria and the corresponding efficiencies in $\mu^+\mu^-\gamma\gamma$ channel. $\gamma 1(\gamma 2)$ is defined as the photon with lower (higher) energy. $M_{\gamma\gamma}^{\text{recoil}}$ is the recoil mass of di-photon system in CEPC $\sqrt{s} = 240$ GeV: $(M_{\gamma\gamma}^{\text{recoil}})^2 = (\sqrt{s} - E_{\gamma\gamma})^2 - p_{\gamma\gamma}^2 = s - 2E_{\gamma\gamma}\sqrt{s} + m_{\gamma\gamma}^2$.

CEPC. No flavor tagging approach is used in this analysis for simplification.

The event selections are applied to improve the signal significance and background modeling. In three sub-channels individual strategies are considered depending on the topology of the physics process. In $ZH \rightarrow \nu\bar{\nu}\gamma\gamma$ channel 2 photons are required inclusively in the final state. In $ZH \rightarrow \mu^+\mu^-\gamma\gamma$ channel the 2 leading photons and 2 muons are selected exclusively, requiring a veto of other particles, the missing energy E_{missing} and missing mass M_{missing} less than 10 GeV and the invariant mass of the muon pair close to Z boson mass.

In $ZH \rightarrow q\bar{q}\gamma\gamma$ channel, 2 leading photons are firstly selected, and other particles are reconstructed into 2 jets with Durham algorithm. Some dedicated cuts are applied on the kinematic variables of these final state objects as listed in Table 2, 3, 4, along with the final efficiency and expected event yields.

4. MVA-based analysis

The Multi-Variate Analysis (MVA) method is employed to further suppress the background. It uses the machine learning (ML) packages to combine the separation power from several variables into a unique variable. In this analysis we choose the Gradient Boosted Decision Tree (BDTG)

Selections	Higgs signal	$\nu\bar{\nu}\gamma\gamma$ background
Inclusive 2 photons	85.51%	0.34%
$E_{\gamma} > 30$ GeV	99.81%	20.13%
$ \cos \theta_{\gamma} < 0.8$	70.48%	11.56%
$pT_{\gamma} > 20$ GeV	99.97%	99.26%
$M_{\text{missing}} > 60$ GeV	98.17%	99.71%
$m_{\gamma\gamma} \in [110, 140]$ GeV	97.51%	22.86%
$E_{\gamma\gamma} \in [120, 150]$ GeV	99.16%	99.58%
Total eff	57.08%	0.002%
Yields in 5.6 ab^{-1}	335.89	3640.20

Table 4

Selection criteria and the corresponding efficiencies in $\nu\bar{\nu}\gamma\gamma$ channel. M_{missing} is the missing mass calculated from the total visible objects.

method with TMVA toolkit [28]. In each sub-channel the ZH and 2 fermion processes are considered as the signal and background for the BDTG. All events from MC are separated into 2 sets for the 2-fold validation [29] to avoid the risk of overtraining. Following principles are considered while constructing the input variables for BDTG:

- The basic information is the Lorentz vector of the final state particles. These include the momentum (P), transverse momentum (p_T), energy (E), polar angle ($\cos \theta$), recoil mass for photons, fermions, systems, and the ΔP , ΔE , $\Delta \Phi$, $\Delta \cos \theta$, ΔR for 2 objects or systems, and the missing mass M_{missing} .
 - Use the separation $\langle S^2 \rangle$ defined in Eq. 1 to quantify the discrimination power between signal and background of a given variable, where y represents the discriminating variable, and $\hat{y}_s(y)$ and $\hat{y}_b(y)$ are the corresponding distributions of the variable for signal and background samples.
- $$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_s(y) - \hat{y}_b(y))^2}{\hat{y}_s(y) + \hat{y}_b(y)} dy. \quad (1)$$
- To ensure the application of 2D model described in Sec. 5, which requires an assumption of independence between the BDTG response and $m_{\gamma\gamma}$, the constructed variable should have low linear correlation with $m_{\gamma\gamma}$: $|\text{Corr}_{v-m_{\gamma\gamma}}| < 30\%$.
 - To reduce the redundancy for the training, the linear correlation between any two variables should be small: $|\text{Corr}_{v1-v2}| < 40\%$. The one with lower separation power is removed.

Table 5-7 lists the selected variables along with their definition and $\langle S^2 \rangle$ for BDTG.

5. Signal and background models

The Higgs signal is extracted by fitting the $m_{\gamma\gamma}$ and the shape of the BDTG responses. The resonant peak above a

Variable	Definition	Separation
$pT_{\gamma 1}$	Transverse momentum of the sub-leading photon	0.209
$\cos\theta_{\gamma 2}$	Polar angle of the leading photon	0.197
$\Delta\Phi_{\gamma\gamma}$	Azimuthal angle between two photons	0.147
$\min\Delta R_{\gamma,j}$	Minimum ΔR between one of the two photons and one of the jets	0.054
E_{j1}	Energy of the sub-leading jet	0.041
$\Delta\Phi_{\gamma\gamma,jj}$	Azimuthal angle between the diphoton and dijet system	0.033
pT_{j2}	Transverse momentum of the leading jet	0.032
$\cos\theta_{j1}$	Polar angle of the sub-leading jet	0.032
$\cos\theta_{\gamma\gamma,jj}$	Polar angle difference between diphoton and dijet system $\cos(\theta_{\gamma\gamma} - \theta_{jj})$	0.024
$\cos\theta_{\gamma 1,j1}$	Polar angle difference between sub-leading photon and sub-leading jet $\cos(\theta_{\gamma 1} - \theta_{j1})$	0.023

Table 5

 Input variables for BDTG in $q\bar{q}\gamma\gamma$ channel.

Variable	Definition	Separation
$\min\Delta R_{\gamma,\mu}$	Minimum ΔR between one of the two photons and one of the muons	0.335
$E_{\mu\mu}$	Energy of the di-muon system	0.259
$\cos\theta_{\gamma 1,\mu 1}$	Polar angle difference between sub-leading photon and sub-leading muon	0.189
$E_{\gamma 2}$	Leading photon energy	0.160
$\Delta\Phi_{\gamma\gamma}$	Azimuthal angle between two photons	0.090
$\cos\theta_{\gamma 2}$	Polar angle of the leading photon	0.072
$\Delta\Phi_{\gamma\gamma,\mu\mu}$	Azimuthal angle between diphoton and dimuon system	0.034
$\cos\theta_{\mu 1}$	Polar angle of the sub-leading muon	0.014

Table 6

 Input variables for BDTG in $\mu^+\mu^-\gamma\gamma$ channel.

Variable	Definition	Separation
$pT_{\gamma 1}$	Transverse momentum of the sub-leading photon	0.089
$\cos\theta_{\gamma 2}$	Polar angle of the leading photon	0.079
$\Delta\Phi_{\gamma\gamma}$	Azimuthal angle between two photons	0.054
$pT_{\gamma\gamma}^{\perp}$	Diphoton p_T projected perpendicular to the diphoton thrust axis	0.042
$pT_{\gamma 2}$	Transverse momentum of the leading photon	0.037

Table 7

 Input variables for BDTG in $\nu\bar{\nu}\gamma\gamma$ channel.

smooth $m_{\gamma\gamma}$ distribution for the background at around Higgs mass (125 GeV) can be reconstructed through the excellent calorimeter energy resolution in CEPC. The signal $m_{\gamma\gamma}$ model is described by a Double Side Crystal Ball (DSCB) function:

$$f(t) = N \times \begin{cases} e^{-t^2/2} & \text{if } -\alpha_{low} \leq t \leq \alpha_{high} \\ \frac{e^{-\frac{1}{2}\alpha_{low}^2}}{\left[\frac{1}{R_{low}}(R_{low}-\alpha_{low}-t)\right]^{n_{low}}} & \text{if } t < -\alpha_{low} \\ \frac{e^{-\frac{1}{2}\alpha_{high}^2}}{\left[\frac{1}{R_{high}}(R_{high}-\alpha_{high}+t)\right]^{n_{high}}} & \text{if } t > \alpha_{high} \end{cases} \quad (2)$$

where N is a normalization factor, $t = (m_{\gamma\gamma} - \mu_{CB})/\sigma_{CB}$. Figure 2 shows the fitted $m_{\gamma\gamma}$ signal shape in 3 channels. They are well described by the DSCB function. The resolution is estimated to be 2.81 / 2.68 / 2.74 GeV in $q\bar{q}\gamma\gamma/\mu^+\mu^-\gamma\gamma/\nu\bar{\nu}\gamma\gamma$ channel.

Several smooth functions (Chebychev polynomials, exponential families and polynomial families) are tested for the background modelling, and the one with the least χ^2/N_{dof}

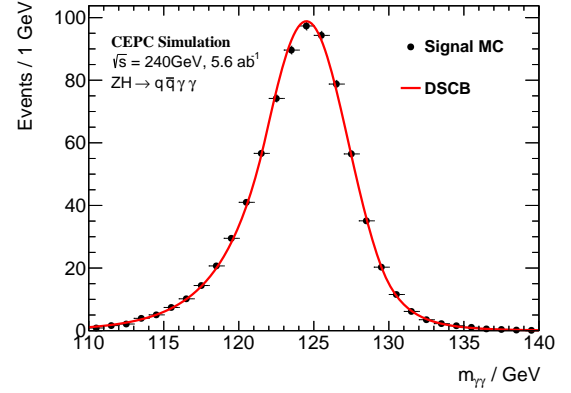
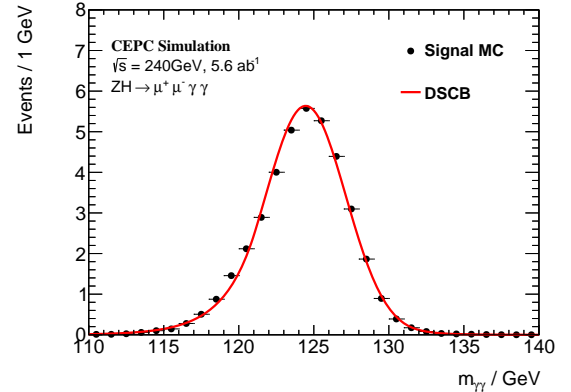
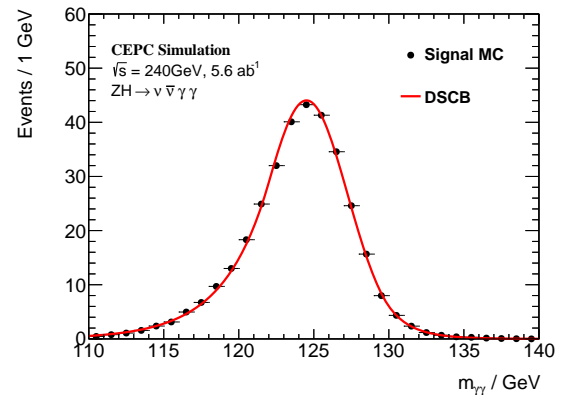

 (a) $q\bar{q}\gamma\gamma$ signal

 (b) $\mu^+\mu^-\gamma\gamma$ signal

 (c) $\nu\bar{\nu}\gamma\gamma$ signal

Figure 2: The signal MC and the fitted DSCB model in 3 channels.

value is finally selected. The results are listed in Table 8 and shown in Figure 3.

There is no expectation on the BDTG response distributions, so the histograms from the MC of signal and background are used to build the binned Probability Density Function (PDF), which is used as the model of BDTG distributions.

The strategies in constructing BDTG ensure the reasonable independence between the BDTG response and $m_{\gamma\gamma}$.

Channel	Selected function	χ^2/N_{dof}
$q\bar{q}\gamma\gamma$	2nd order Chebyshev	0.60
$\mu^+\mu^-\gamma\gamma$	2nd order Chebyshev	1.79
$\nu\bar{\nu}\gamma\gamma$	1st order Chebyshev	3.32

Table 8

The decided background model in 3 channels. Tested functions include the exponential, 2nd order exponential polynomial, 1st and 2nd order polynomials, 1st and 2nd order Chebyshev polynomials.

Therefore a 2-dimension model from the multiplication of $m_{\gamma\gamma}$ and BDT models is applied to describe the signal and background. A high correlation can introduce the mismodeling of the signal and/or background process. As a check, the linear correlation coefficients between $m_{\gamma\gamma}$ and BDT are -3.45%, -11.6%, 8.33% for signal in $q\bar{q}\gamma\gamma$, $\mu^+\mu^-\gamma\gamma$ and $\nu\bar{\nu}\gamma\gamma$ channels. The corresponding correlation coefficients for the background are 11.6%, 28.2% and 28.4% respectively.

6. Systematic uncertainties

The systematic uncertainties relevant to this measurement can be caused by many sources, e.g. the theoretical calculation, the measurement of the integrated luminosity, the object reconstruction and identification. However, at this stage most of them are not specifically studied yet for the CEPC. So in this paper we only present the methodology of analyzing the systematic uncertainties in CEPC and take the leading terms into account. Further quantified analysis awaits updates on the theoretical calculation, more comprehensive detector performance optimization and real data.

Based on the strategy of event modeling in Sec. 5, the systematic uncertainties can be grounded into two types: uncertainties in the expected signal yields in each channel, and the uncertainties in the modeling of the signal $m_{\gamma\gamma}$ distribution. The background yields and $m_{\gamma\gamma}$ model parameters are floated to consider the background mismodeling effect and the contributions from model-dependent background process cross section calculation. The uncertainty of BDT modeling for both signal and background is contained by an envelope, which is counted into the signal event yield uncertainty.

These systematic terms are incorporated into the likelihood model as nuisance parameters. For each of these nuisance parameters, a Gaussian or log-normal constraint PDF is included in the likelihood function, for the symmetric term like the $m_{\gamma\gamma}$ shape peak position, or non-negative term like the event yield. The construction of likelihood with the nuisance parameters are in Sec 7.

6.1. Theoretical uncertainties

In contrast to hadron collider, the theoretical calculations are less dependent on higher order QCD radiative corrections and have no influence from the Parton Distribution Functions. The most significant part the QCD can enter is the loop calculation in $H \rightarrow \gamma\gamma$ branching ratio. This contributes a 5% uncertainty on the signal yield [4]. The

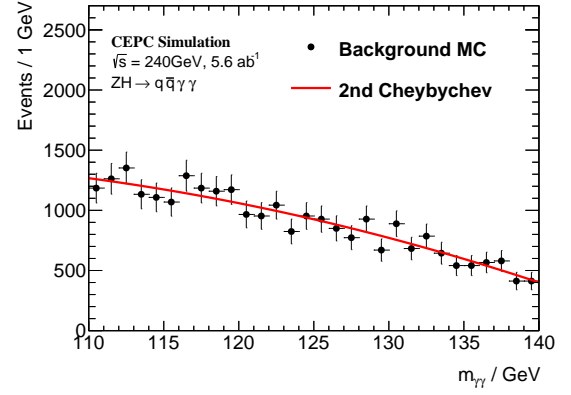
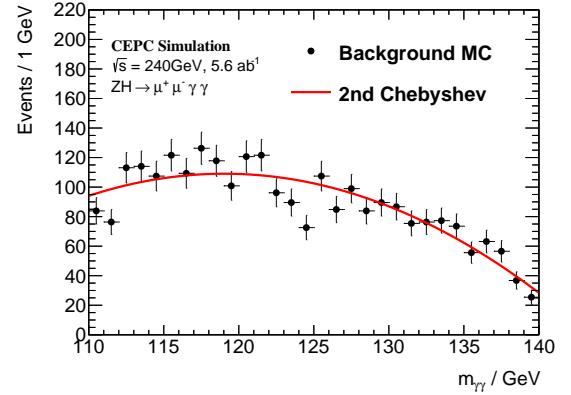
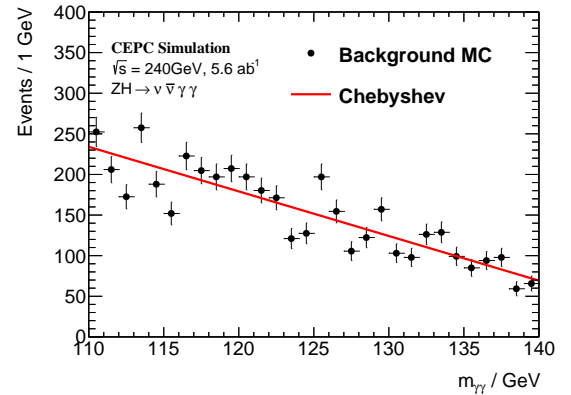
(a) $q\bar{q}\gamma\gamma$ background(b) $\mu^+\mu^-\gamma\gamma$ background(c) $\nu\bar{\nu}\gamma\gamma$ background

Figure 3: The background MC and the fitted $m_{\gamma\gamma}$ models in 3 channels.

uncertainty from parton shower only exist in the channel with hadronic final state, and can be described by the MC sample difference from a set of generators. Due to the lack of calculations, we assume a totally 5% uncertainty from theoretical part on the signal yield in this paper.

6.2. Experimental uncertainties

The experimental systematic uncertainties affecting this measurement can include: integrated luminosity, detector

acceptance, trigger efficiency, object reconstruction and identification efficiency, object energy scale and resolution. In CEPC the luminosity can be monitored by the Lumi-Cal with the high statistics Bhabha process, thus a relative accuracy of 0.1% is expected to be achieved [4]. Pile-up effects and underlying events should be negligible. A well-described detector geometry in the simulation is able to provide precise model of the detector acceptance and response. The possible mismodeling can be further fixed with some data-driven methods. So the uncertainties should be very small. The photon reconstruction, identification and energy rely on the dedicated algorithms and the real data. In CEPC CDR these uncertainties are studied to be controlled with sub-percent level. Furthermore, the known physics processes can be used as standard candles for the calibration, like $Z \rightarrow e^+e^- + \gamma$, $\pi^0 \rightarrow \gamma\gamma$, etc. Similarly electrons, muons and jets can in principle be described well. In this di-photon channel study, the photon related uncertainties should be dominant. So we assume a 1% uncertainty on the photon efficiency and 0.05% uncertainties on the photon energy scale (PES) and resolution (PER). Others terms remains to be added with better understanding about the experiments.

Among above experimental uncertainties, the signal yield is affected by the luminosity, the photon efficiency and the impact of the photon energy scale and resolution uncertainties on the selection efficiency. A set of alternative simulation samples are generated, randomly rejecting 1% photons, or scaling the energy up/down for 0.05% or smearing the photon energy with 0.05%. The expected signal yields are counted after all the selections, and a relative variation $\delta n^i = \frac{|n_{var}^i - n_{nom}^i|}{n_{nom}^i}$ is used to represent the influence from each term. This value from photon efficiency is around 2%, and from photon energy scale and resolution is at the order of 0.01%. Only symmetric uncertainty is considered so there is no up/down variation.

The signal $m_{\gamma\gamma}$ distribution is described with the double-side crystal ball function. The photon energy scale uncertainty is propagated to the peak position of the signal peak, and the photon energy resolution uncertainty is propagated to the signal width. They are estimated by refitting the signal shape in the variation samples and comparing with the nominal one: $\delta\mu_{CB} = \frac{\mu_{CB,var} - \mu_{CB,nom}}{\mu_{CB,nom}}$, $\delta\sigma_{CB} = \frac{\sigma_{CB,var} - \sigma_{CB,nom}}{\sigma_{CB,nom}}$. The impact from PES to the signal peak is between 0.04% to 0.10%, in different channels, and the impact from PER to the signal width is 0.004% to 0.02%. The Higgs mass measurement uncertainty is also considered. The latest result is 125.38 ± 0.14 GeV by CMS in diphoton channel [30] and 124.94 ± 0.18 GeV by ATLAS in 4-lepton channel [31]. So we can assume a 0.1 GeV uncertainty in the future. This is added on the signal peak position.

The influence from these aforementioned uncertainties on the BDT modeling is studied by comparing the BDT distribution bin by bin between the nominal and variation MC sample. The maximum variation value $\delta n = \frac{|n_{var} - n_{nom}|}{n_{nom}}$ in all BDT bins and systematic terms is applied on the signal yield as the uncertainty from BDT, except the bin with

low statistics (bin content less than 5% of total yield). The uncertainty from BDT itself is believed to be included in this envelope value. In three channels this term is 0.5% to 0.7%.

7. Results

The signal strength $\mu = \frac{N(e^+e^- \rightarrow ZH \rightarrow f\bar{f}\gamma\gamma)}{N_{SM}(e^+e^- \rightarrow ZH \rightarrow f\bar{f}\gamma\gamma)}$ is extracted by a combined fit in three channels with the unbinned maximum likelihood fit method. The likelihood function is built with the models in Sec. 5 and the constraints coming from systematic uncertainties in Sec. 6:

$$\mathcal{L}(\mu, \theta; (m_{\gamma\gamma}, \text{BDT})) = \prod_c \text{Pois}(n_c | N_c(\mu, \theta)) \cdot \prod_i^n f_c((m_{\gamma\gamma}, \text{BDT})^i; \theta) \cdot \prod_j G(\theta_j), \quad (3)$$

in which:

- μ is the signal strength, which is the parameter of interest (POI) in this analysis;
- θ are nuisance parameters defined for each systematic term.
- n_c is the observed event number in the channel c ;
- $N_c(\mu, \theta) = \mu S_{SM,c}(\theta_{yield}) + B_c$. $S_{SM,c}(\theta_{yield})$ is the expected signal yield in the channel, including the relevant nuisance parameters. B_c is the background yield;
- $f_c((m_{\gamma\gamma}, \text{BDT})^i; \theta)$ is the probability density function built with the signal and background model in Sec. 5:

$$f_c((m_{\gamma\gamma}, \text{BDT})^i; \theta) = \frac{1}{N_c} \times [\mu S_{SM,c}(\theta_{yield}) f_{c,sig}((m_{\gamma\gamma}, \text{BDT})^i; \theta) + B_c f_{c,bkg}((m_{\gamma\gamma}, \text{BDT})^i; \theta)]. \quad (4)$$

- The signal yield $S_{SM,c}$, shape peak μ_{CB} and width σ_{CB} are affected by the systematic uncertainties with a response function, so

$$\begin{aligned} S_{SM,c}(\theta_{yield}) &= S_{SM,c} \prod_j e^{\theta_j \sqrt{\ln(1+\delta_j^2)}}, \\ \mu_{CB}(\theta_{peak}) &= \mu_{CB}^{nom} \prod_j (1 + \delta_j \theta_j), \\ \sigma_{CB}(\theta_{width}) &= \sigma_{CB}^{nom} \prod_j e^{\theta_j \sqrt{\ln(1+\delta_j^2)}} \end{aligned} \quad (5)$$

- $G(\theta_j)$ is the unitary Gaussian constraint PDF for the nuisance parameter j , with mean 0 and width 1.

	$(\frac{\delta\mu}{\mu})_{q\bar{q}\gamma\gamma}$	$(\frac{\delta\mu}{\mu})_{\mu^+\mu^-\gamma\gamma}$	$(\frac{\delta\mu}{\mu})_{\nu\bar{\nu}\gamma\gamma}$
Stat. only	0.098	0.371	0.127
Theo 5%	0.110	0.376	0.138
Lumi 0.1%	0.098	0.372	0.127
photon eff 1%	0.100	0.372	0.128
PES 0.05%	0.098	0.372	0.128
PER 0.05%	0.098	0.372	0.128
mH 0.08	0.098	0.371	0.127
BDT	0.099	0.372	0.128
Total	0.112	0.376	0.139

Table 9

The decoupled contribution from considered systematic uncertainties in 3 channels. Numbers in the table show the combination of statistical uncertainty and the corresponding systematic term.

Channel	$\mu \pm \text{unc. (stat.)} @ 5.6\text{ab}^{-1}$	$\mu \pm \text{unc. (stat.)} @ 20\text{ab}^{-1}$
$q\bar{q}\gamma\gamma$	$1.00 \pm 0.112(0.098)$	$1.00 \pm 0.075 (0.052)$
$\mu^+\mu^-\gamma\gamma$	$1.00 \pm 0.376(0.371)$	$1.00 \pm 0.208 (0.200)$
$\nu\bar{\nu}\gamma\gamma$	$1.00 \pm 0.139(0.127)$	$1.00 \pm 0.087 (0.067)$
Combined	$1.00 \pm 0.094(0.077)$	$1.00 \pm 0.068 (0.040)$

Table 10

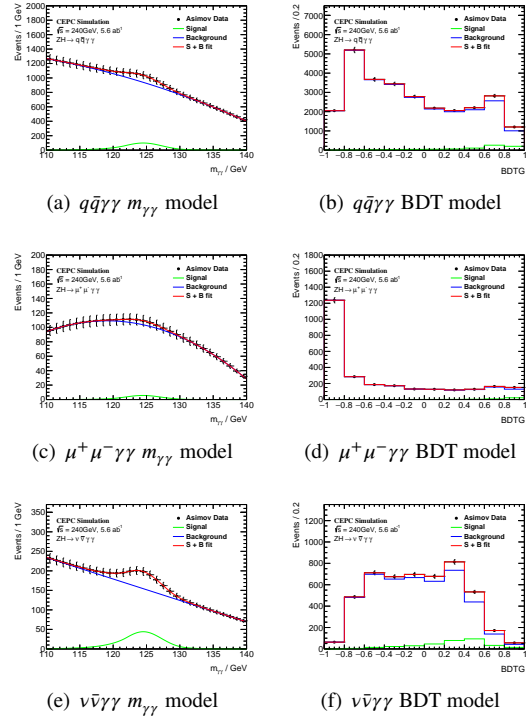
Expected signal strengths from Asimov data fit and the corresponding precision in 3 channels and the combination. Results in 20 ab^{-1} are obtained by re-fitting the workspace with the scaled signal and background yields.

In the fitting the signal model parameters are fixed to the value in fitting the signal MC. The background yields, model parameters and all nuisance parameters are floated as mentioned in Sec. 6.

In order to mimic the real data and avoid the statistical fluctuations of MC samples, a set of Asimov data [32] are generated from the signal + background models and are simultaneously fitted to obtain the expected precision and significance. Figure 4 shows the $m_{\gamma\gamma}$ and BDTG distributions of the Asimov data and the models in 3 channels. A final precision of 9.4% for $\delta(\sigma \times Br) \backslash (\sigma \times Br)$ can be reached in $H \rightarrow \gamma\gamma$ measurement in the CEPC with 5.6 ab^{-1} data. With the 20 ab^{-1} data of the updated CEPC operation period, this precision can reach to 6.8%. The statistical only results are 7.7% and 4.0% for 2 data taking schemes. Results are summarized in Table 9 and 10. With our preliminary assumption this measurement will be still statistics dominant in 5.6 ab^{-1} , and the theoretical uncertainty plays a leading role among all systematic terms.

8. $Br(H \rightarrow \gamma\gamma)$ precision with ECAL resolution

While fitting the $m_{\gamma\gamma}$ shape, the width of the signal peak is a direct connection between the measurement precision in $H \rightarrow \gamma\gamma$ channel and the ECAL resolution. Currently a new detector design for CEPC is under development [12, 13, 14], in which the present Si-W sampling ECAL will be replaced


Figure 4: The Combined fit to the Asimov data in 3 channels.

by a homogeneous crystal ECAL. This new ECAL is expected to have better photon resolution ($\sigma_E/E = 3\%/\sqrt{E}$) and more efficient neutral meson (π^0) reconstruction. This can benefit the jet reconstruction and flavor physics study at the CEPC. A rough estimation in the $q\bar{q}\gamma\gamma$ channel is performed within the strategy of this work.

In the estimation the selected photon is replaced by the truth photon with a smearing in its energy. Normally the ECAL energy is approximated as:

$$\frac{\sigma_E}{E} = A \oplus \frac{B}{\sqrt{E}} \oplus \frac{C}{E}, \quad (6)$$

where A stands for the constant term like energy leakage, readout threshold, etc. B represents the stochastic term from photoelectron statistics and depends on the sensitive material. C comes from the electronic noises. Presently the noise term C is expected to be 0, and the constant term A is expected to be at the level of 1%. The photon energy is smeared with the stochastic term B varying from 1% to 35%. Figure 5 shows a comparison between the $m_{\gamma\gamma}$ shape from the full simulation and 2 smearing points 3% and 16%. The same selection criteria are applied as in Sec. 3, while the BDT is not employed in this simplified study to focus on the photon detection only, which is expected to have a roughly 30% decrease comparing with the result in Sec. 7. A Gaussian function is used to describe the signal model from the energy smearing. The 2-dimension model is replaced with a 1-dimension $m_{\gamma\gamma}$ model, and a similar unbinned maximum likelihood fit is performed to extract the signal strength precision $\delta\mu/\mu$, without systematic uncertainties.

Figure 6 shows the relationship between energy resolution B and the fitted precision $\delta\mu/\mu$. These points can be fitted with the following function:

$$\frac{\delta\mu}{\mu} = p_0 \oplus (p_1 \times B), \quad (7)$$

where p_0 and $p_1 \times B$ represent the contributions from constant term and stochastic term respectively. From this relation the homogeneous ECAL is able to bring a 28% improvement in the statistic precision of signal strength measurement. Also a "critical point" can be defined: the two components in resolution have the same contribution to $\delta\mu/\mu$, i.e. $p_0 = p_1 \times B$. When the constant term A is fixed to 1%, the critical point for B , within this definition, is 14%. This indicates the constant term in resolution would become the dominant contribution at new ECAL design point with $B = 3\%$. A scanning for a series of constant terms and the corresponding balanced stochastic terms is shown in Figure 7.

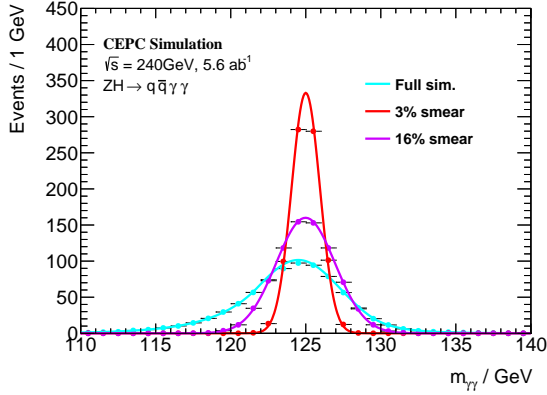


Figure 5: The signal shape for the full simulated $H \rightarrow \gamma\gamma$ sample (blue) and for two samples with smeared photon energy (3% in red and 16% in violet). The fitted signal width are 2.81 GeV, 0.94 GeV and 1.96 GeV respectively.

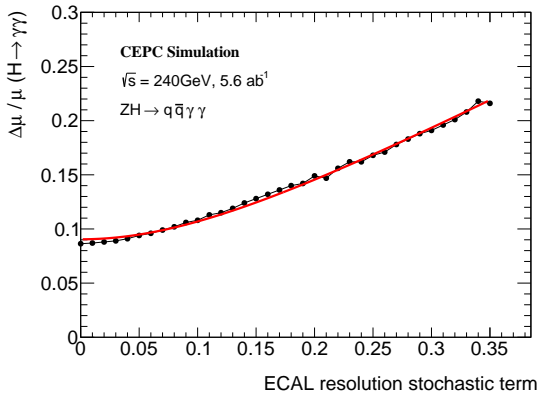


Figure 6: The $Br(H \rightarrow \gamma\gamma)$ precision in $ZH \rightarrow q\bar{q}\gamma\gamma$ channel as a function of the stochastic term in ECAL resolution from a fast analysis. The points are fitted with Eq. 7.

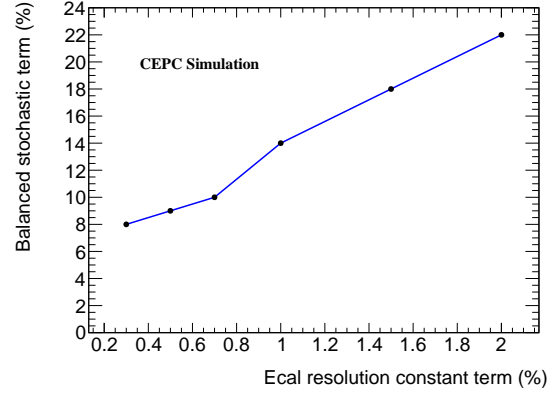


Figure 7: The balanced ECAL stochastic resolution points with different configurations of the constant term.

9. Conclusion

This paper presents the expected precision for the measurement of the $H \rightarrow \gamma\gamma$ branching ratio in the CEPC via $ZH \rightarrow q\bar{q}\gamma\gamma$, $ZH \rightarrow \mu^+\mu^-\gamma\gamma$, $ZH \rightarrow \nu\bar{\nu}\gamma\gamma$ channels. The physics events are reconstructed with the CEPC-v4 detector simulation, and selected by a set of criteria. A BDTG is developed for further signal/background separation, and is used along with $m_{\gamma\gamma}$ as discriminating variables in the maximum likelihood fit when extracting the signal strength. We build a preliminary framework for the systematic uncertainty analysis in CEPC with the nuisance parameters, and take several leading terms into account. With the scheduled integrated luminosity of 5.6 ab^{-1} a precision of 9.4% (7.7% stat.) is expected to be achieved at the CEPC, With 20 ab^{-1} data this precision can be 6.8% (5.5% stat.). More mature results await further development of this framework and better knowledge of systematic terms in the CEPC. Meanwhile the ECAL performance is studied by smearing photon energy resolution in $q\bar{q}\gamma\gamma$ channel. A direct relationship between the ECAL resolution and the precision in $H \rightarrow \gamma\gamma$ measurement is foreseen.

10. Acknowledgments

The authors would like to thank the CEPC software group for the technical supports of simulation, reconstruction packages, as well as the CEPC physics group for the valuable discussions. This study is supported by the IHEP innovative project on sciences and technologies under Project No. E2545AU210.

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, Physics Letters B 716 (2012) 1–29. URL: <https://doi.org/10.1016%2Fj.physletb.2012.08.020>. doi:10.1016/j.physletb.2012.08.020.
- [2] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Physics Letters B 716 (2012) 30–61. URL: <https://doi.org/10.1016%2Fj.physletb.2012.08.021>. doi:10.1016/j.physletb.2012.08.021.

- [3] T. Behnke, J. E. Brau, P. N. Burrows, J. Fuster, M. Peskin, M. Stanitzki, Y. Sugimoto, S. Yamada, H. Yamamoto, The international linear collider technical design report - volume 4: Detectors, 2013. URL: <https://arxiv.org/abs/1306.6329>. doi:10.48550/ARXIV.1306.6329.
- [4] M. Dong, et al. (CEPC Study Group), CEPC Conceptual Design Report: Volume 2 - Physics & Detector (2018). arXiv:1811.10545.
- [5] The FCC Collaboration, Fcc-ee: The lepton collider, The European Physical Journal Special Topics 228 (2019) 261–623. URL: <https://doi.org/10.1140/epjst/e2019-900045-4>. doi:10.1140/epjst/e2019-900045-4.
- [6] A. Robson, P. N. Burrows, N. C. Lasheras, L. Linssen, M. Petric, D. Schulte, E. Sicking, S. Stapnes, W. Wuensch, The compact linear e^+e^- collider (cllc): Accelerator and detector, 2018. URL: <https://arxiv.org/abs/1812.07987>. doi:10.48550/ARXIV.1812.07987.
- [7] F. An, et al., Precision Higgs physics at the CEPC, Chin. Phys. C 43 (2019) 043002. doi:10.1088/1674-1137/43/4/043002. arXiv:1810.09037.
- [8] ATLAS Collaboration, Measurement of the properties of higgs boson production at $\sqrt{s} = 13$ tev in the $h \rightarrow \gamma\gamma$ channel using 139 fb $^{-1}$ of pp collision data with the atlas experiment, 2022. URL: <https://arxiv.org/abs/2207.00348>. doi:10.48550/ARXIV.2207.00348.
- [9] Measurements of Higgs boson properties in the diphoton decay channel at $\sqrt{s} = 13$ TeV, Technical Report, CERN, Geneva, 2020. URL: <https://cds.cern.ch/record/2725142>.
- [10] M. Cepeda, et al., Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC, CERN Yellow Rep. Monogr. 7 (2019) 221–584. doi:10.23731/CYRM-2019-007.221. arXiv:1902.00134.
- [11] CEPC Accelerator Study Group, Snowmass2021 white paper af3-cepc, 2022. URL: <https://arxiv.org/abs/2203.09451>. doi:10.48550/ARXIV.2203.09451.
- [12] H. Cheng, et al. (CEPC Physics Study Group), The Physics potential of the CEPC. Prepared for the US Snowmass Community Planning Exercise (Snowmass 2021), in: 2022 Snowmass Summer Study, 2022. arXiv:2205.08553.
- [13] B. Qi, Y. Liu, R & d of a novel high granularity crystal electromagnetic calorimeter, Instruments 6 (2022). URL: <https://www.mdpi.com/2410-390X/6/3/40>. doi:10.3390/instruments6030040.
- [14] Y. Liu, J. Jiang, Y. Wang, High-granularity crystal calorimetry: conceptual designs and first studies, Journal of Instrumentation 15 (2020) C04056–C04056. URL: <https://doi.org/10.1088/1748-0221/15/04/c04056>. doi:10.1088/1748-0221/15/04/c04056.
- [15] M. Thomson, Particle flow calorimetry and the PandoraPFA algorithm, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 611 (2009) 25–40. doi:10.1016/j.nima.2009.09.009.
- [16] W. Kilian, T. Ohl, J. Reuter, WHIZARD: Simulating Multi-Particle Processes at LHC and ILC, Eur. Phys. J. C 71 (2011) 1742. doi:10.1140/epjc/s10052-011-1742-y. arXiv:0708.4233.
- [17] T. Sjostrand, L. Lonnblad, S. Mrenna, P. Z. Skands, Pythia 6.3 physics and manual (2003). arXiv:hep-ph/0308153.
- [18] T. Taylor, D. Treille, The Large Electron Positron Collider (LEP): Probing the Standard Model, Adv. Ser. Direct. High Energy Phys. 27 (2017) 217–261. 45 p. URL: <https://cds.cern.ch/record/2312570>. doi:10.1142/9789814749145_0007.
- [19] CEPC Conceptual Design Report: Volume 1 - Accelerator (2018). arXiv:1809.00285.
- [20] X. Mo, G. Li, M.-Q. Ruan, X.-C. Lou, Physics cross sections and event generation of e^+e^- annihilations at the CEPC, Chin. Phys. C 40 (2016) 033001. doi:10.1088/1674-1137/40/3/033001. arXiv:1505.01008.
- [21] P. Mora de Freitas, H. Videau, Detector simulation with MOKKA / GEANT4: Present and future (2002) 623–627.
- [22] S. Agostinelli, et al. (GEANT4), GEANT4—a simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.
- [23] M. Ruan, Arbor, a new approach of the particle flow algorithm, 2014. doi:10.48550/ARXIV.1403.4784.
- [24] M. Ruan, et al., Reconstruction of physics objects at the Circular Electron Positron Collider with Arbor, Eur. Phys. J. C 78 (2018) 426. doi:10.1140/epjc/s10052-018-5876-z. arXiv:1806.04879.
- [25] F. Gaede, S. Aplin, R. Glattauer, C. Rosemann, G. Voutsinas, Track reconstruction at the ILC: the ILD tracking software, Journal of Physics: Conference Series 513 (2014) 022011. doi:10.1088/1742-6596/513/2/022011.
- [26] D. Yu, M. Ruan, V. Boudry, H. Videau, Lepton identification at particle flow oriented detector for the future e^+e^- higgs factories, The European Physical Journal C 77 (2017). doi:10.1140/epjc/s10052-017-5146-5.
- [27] S. Catani, Y. Dokshitzer, M. Olsson, G. Turnock, B. Webber, New clustering algorithm for multijet cross sections in e^+e^- annihilation, Physics Letters B 269 (1991) 432–438. doi:https://doi.org/10.1016/0370-2693(91)90196-W.
- [28] A. Hoecker, et al., TMVA - Toolkit for Multivariate Data Analysis, 2007. arXiv:physics/0703039.
- [29] M. Stone, Cross-validatory choice and assessment of statistical predictions, Journal of the Royal Statistical Society. Series B (Methodological) 36 (1974) 111–147. URL: <http://www.jstor.org/stable/2984809>.
- [30] A. M. Sirunyan, et al. (CMS), A measurement of the Higgs boson mass in the diphoton decay channel, Phys. Lett. B 805 (2020) 135425. doi:10.1016/j.physletb.2020.135425. arXiv:2002.06398.
- [31] Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel using 139 fb $^{-1}$ of $\sqrt{s} = 13$ TeV pp collisions recorded by the ATLAS detector at the LHC (2022). arXiv:2207.00320.
- [32] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, The European Physical Journal C 71 (2011). doi:10.1140/epjc/s10052-011-1554-0.

A. Appendix

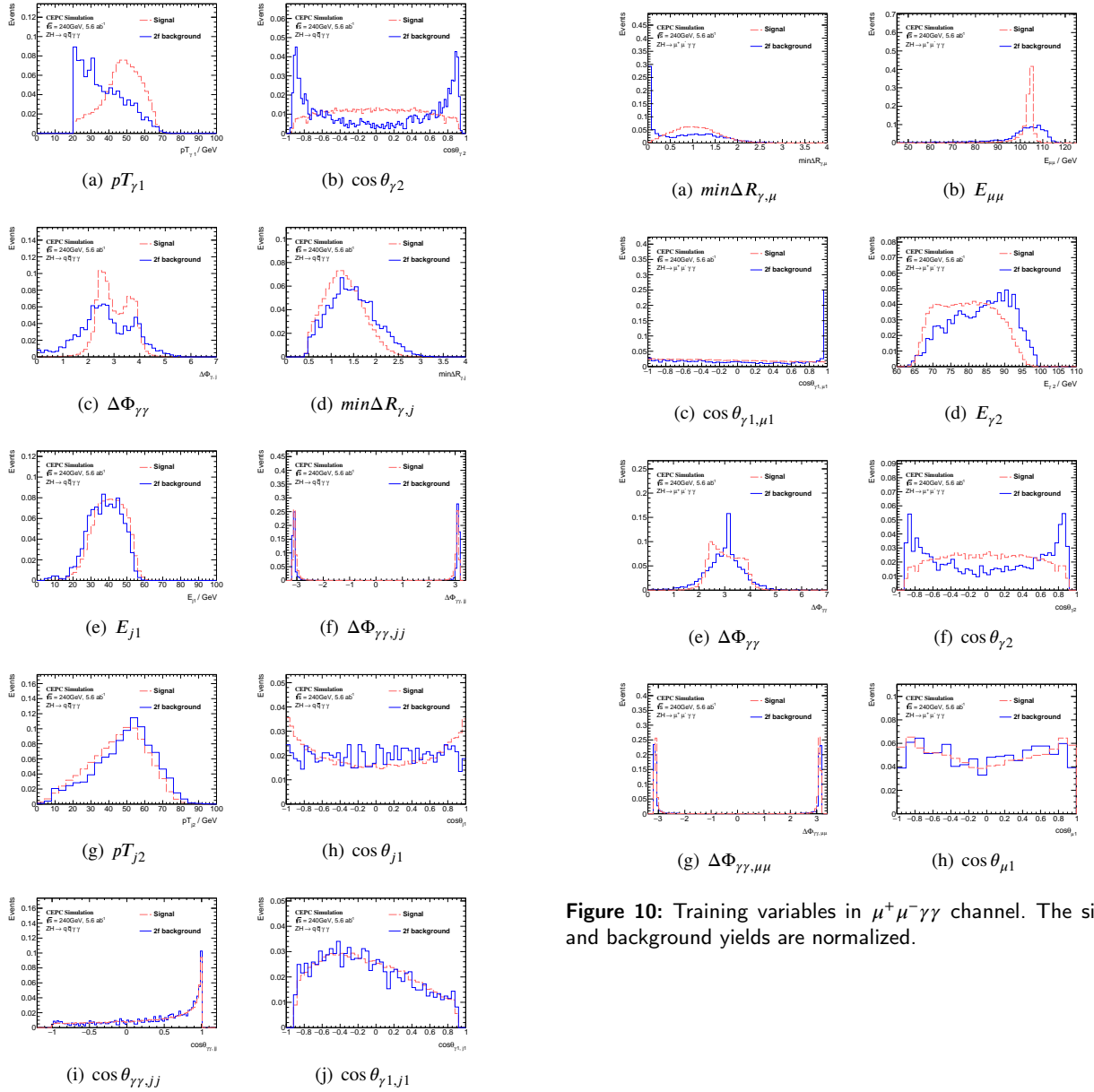


Figure 8: Training variables in $q\bar{q}\gamma\gamma$ channel. The signal and background yields are normalized.

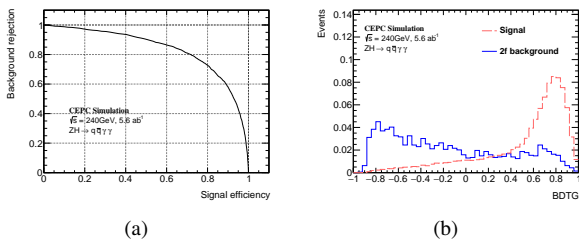


Figure 9: The ROC curve (left) and output BDTG distribution (right) in $q\bar{q}\gamma\gamma$ channel.

Figure 10: Training variables in $\mu^+\mu^-\gamma\gamma$ channel. The signal and background yields are normalized.

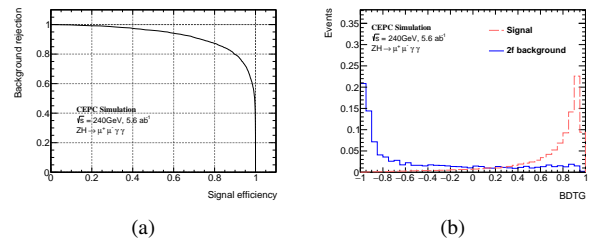


Figure 11: The ROC curve (left) and output BDTG distribution (right) in $\mu^+\mu^-\gamma\gamma$ channel.

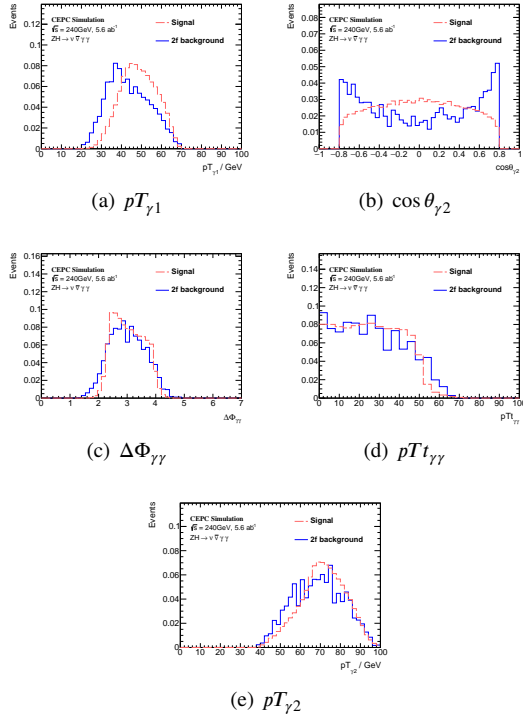


Figure 12: Training variables in $\nu\bar{\nu}\gamma\gamma$ channel. The signal and background yields are normalized.

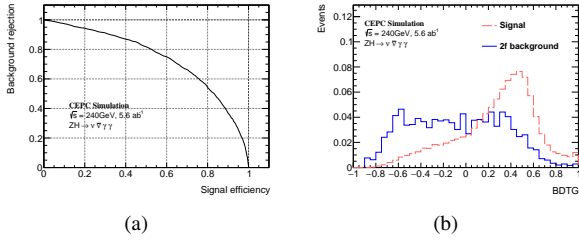
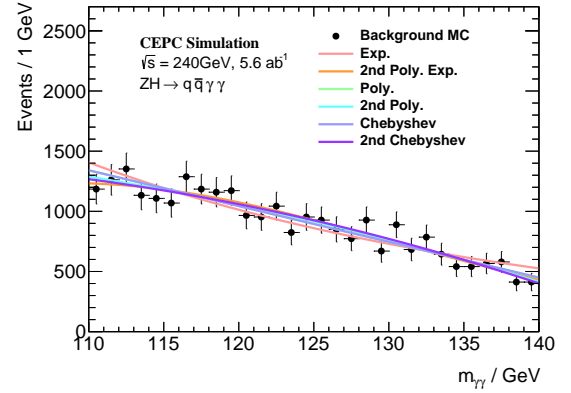


Figure 13: The ROC curve (left) and output BDTG distribution (right) in $\nu\bar{\nu}\gamma\gamma$ channel.

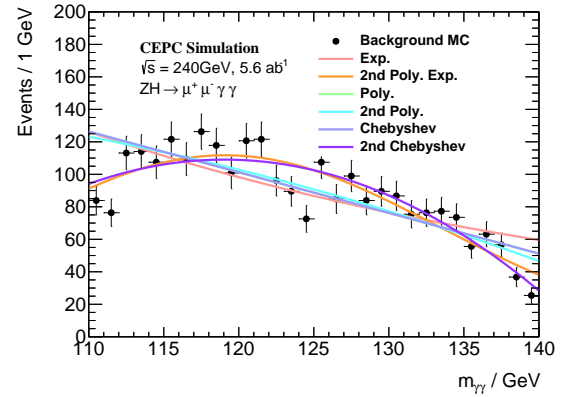
	$q\bar{q}\gamma\gamma$	$\mu^+\mu^-\gamma\gamma$	$\nu\bar{\nu}\gamma\gamma$
1st order Exp.	0.941	5.423	3.786
2nd order Exp.	0.610	2.035	3.435
1st order Poly.	0.644	4.321	7.399
2nd order Poly.	0.600	3.758	3.439
1st order Chebyshev	0.644	4.321	3.320
2nd order Chebyshev	0.596	1.789	3.411

Table 11

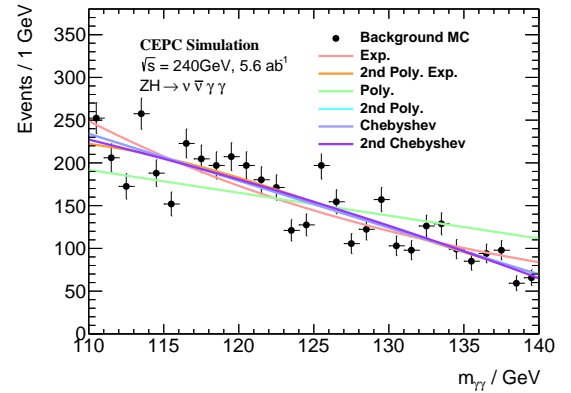
The χ^2/N_{dof} values for 6 considered models in the background modeling in 3 channels, including the first and second order exponential, polynomial and Chebyshev functions.



(a) $q\bar{q}\gamma\gamma$ background



(b) $\mu^+\mu^-\gamma\gamma$ background



(c) $\nu\bar{\nu}\gamma\gamma$ background

Figure 14: Tested functions for the background modeling. In All 3 channels the second order Chebyshev function gives the smallest χ^2/N_{dof} value. Detailed numbers are listed in Table 11.