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 $Br(H \to \gamma\gamma)$ of the Standard Model Higgs boson via the processes $e^+e^- \to$ ZH in which $H \to \gamma\gamma$, $Z \to q\bar{q}/\mu^+\mu^-/\nu\bar{\nu}$ using the baseline conceptual detector with $\sqrt{s} = 240GeV$ at the Circular Electron Positron Collider (CEPC) . The simulated Monte Carlo events are generated and scaled to an integrated luminosity of 5.6 ab^{-1} to mimic the data. Extrapolated results to 20 ab^{-1} are also shown. The expected statistical precision of this measurement after combining 3 channels of Z boson decay is 6.9%, without any systematic uncertainties considered. The performance of CEPC electro-magnetic calorimeter (ECAL) is studied by smearing the photon energy resolution in simulated events in $e^+e^- \to$ ZH $\to q\bar{q}\gamma\gamma$ channel. In present ECAL design, the stochastic term in resolution plays the dominant role in the precision of Higgs measurements in $H \to \gamma\gamma$ channel. The impact of the resolution on the measured precision of Br(ZH $\to q\bar{q}\gamma\gamma$) 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 in 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 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 $\sqrt{s} = 240$ GeV as a Higgs factory. After 10 years running period, the CEPC will collect 5.6 ab^{-1} data, corresponding to more than 1 million Higgs boson. With this clean and large statistic 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 interact with photon through the top quark loop and massive boson loop. This mechanism gives low $H \to \gamma\gamma$ 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. This channel 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 2016 to 2018. The results are consistent with the SM prediction with the present precision. In the HL-LHC peroid the ATLAS expects to collect 3 ab^{-1} data, the projected precision of the $H \to \gamma\gamma$ signal strength is 9% [10].

A previous analysis studied the expected Higgs precision in various Higgs decay channels [7] including $H \to \gamma\gamma$. A result of 6.8% precision in σ(ZH) $\times$ $Br(H \to \gamma\gamma)$ 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 $E_{CM} =$ 240 GeV with two interaction points (IPs), accumulating an integrated luminosity of 20 ab^{-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 $H \to \gamma\gamma$ process with the latest benchmark, update the analysis method from the cut-based 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

guofangyi@ihep.ac.cn (F. Guo)
ORCID(s): 0000-0002-3864-9257 (F. Guo)

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. 6. In Sec. 7 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. 8.

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 $e^+e^- \to$ ZH, W/Z fusion $e^+e^- \to \nu\bar{\nu}H$ and $e^+e^- \to e^+e^-H$ as illustrated in Figure 1. In this analysis Higgs production via ZH process decaying to diphoton final state $e^+e^- \to ZH \to f\bar{f}\gamma\gamma$ 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 $\nu\bar{\nu}$. $Z \to e^+e^-$ channel is abandoned due to the known extremely large Bhabha background, and $Z \to \tau^+\tau^-$ channel is dropped as well because of the complexity of τ jet identification. W/Z fusion process is counted in ZH, $Z \to \nu\bar{\nu}$ sub-channel. The background process only counts the 2-fermion background $e^+e^- \to f\bar{f}$ 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]. All physical parameters are set as the SM predictions. 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 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.

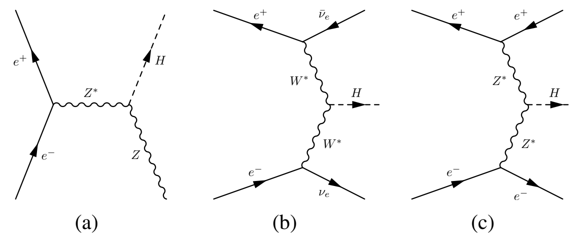


Figure 1: Feynman diagrams of the Higgs boson production processes at the CEPC: (a) $e^+e^- \to$ ZH, (b) $e^+e^- \to \nu\bar{\nu}H$ and (c) $e^+e^- \to e^+e^-H$.

Process	σ	statistics
$q\bar{q}\gamma\gamma$ sub-channel		
$e^+e^- \to \mathrm{ZH} \to q\bar{q}\gamma\gamma$	0.31 fb	100 k
$e^+e^- \to q\bar{q}$	54.1 pb	20 M
$\mu^+\mu^-\gamma\gamma$ sub-channel		
$e^+e^- \to \mathrm{ZH} \to \mu^+\mu^-\gamma\gamma$	0.15 fb	100 k
$e^+e^- \to \mu^+\mu^-$	5.3 pb	20 M
$\nu\bar{\nu}\gamma\gamma$ sub-channel		
$e^+e^- \to \mathrm{ZH} \to \nu\bar{\nu}\gamma\gamma$ $e^+e^- \to \nu\bar{\nu}H \to \nu\bar{\nu}\gamma\gamma$	0.11 fb	100 k
$e^+e^- \to \nu\bar{\nu}$	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 $Z \to inv.$ and W/Z fusion processes are considered.

3. Object reconstruction and event selection

The CEPC follows the PFA scheme in the event reconstruction, with a dedicated tookit 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, μ) 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 $W \to q\bar{q}$ and/or $Z \to q\bar{q}$ in 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 $\to \nu\bar{\nu}\gamma\gamma$ channel

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 azimuth 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_{γ} >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%
$M^{recoil}_{\gamma\gamma} \in [85, 105]$ 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. $M^{recoil}_{\gamma\gamma}$ is the recoil mass of di-photon system in CEPC $\sqrt{s} = 240$ GeV: $(M^{recoil}_{\gamma\gamma})^2 = (\sqrt{s} - E_{\gamma\gamma})^2 - p^2_{\gamma\gamma} = s - 2E_{gamgam}\sqrt{s} + m^2_{\gamma\gamma}$.

2 photons are required inclusively in the final state. In ZH $\to \mu^+\mu^-\gamma\gamma$ 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 $\to q\bar{q}\gamma\gamma$ 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) 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

Selections	Higgs signal	$\nu\bar{\nu}\gamma\gamma$ background
Inclusive 2 photons	85.51%	0.34%
E_{γ} > 30 GeV	99.81%	20.13%
$\|cos\theta_{\gamma}\|$ < 0.8	70.48%	11.56%
pT_{γ} > 20 GeV	99.97%	99.26%
$M_{missing}$ > 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.

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 $m_{\gamma\gamma}$: $|Corr_{v-m_{\gamma\gamma}}| < 30\%$.
- To reduce the redundance for the training, the linear correlation between any two variables should be small: $|Corr_{v1-v2}| < 40\%$. 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 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 $m_{\gamma\gamma}$

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
$pTt_{\gamma\gamma}$	Diphoton p_T projected perpendicular to the diphoton thrust axis	0.042
$pT_{\gamma 2}$	Transverse momentum of the leading photon	0.037

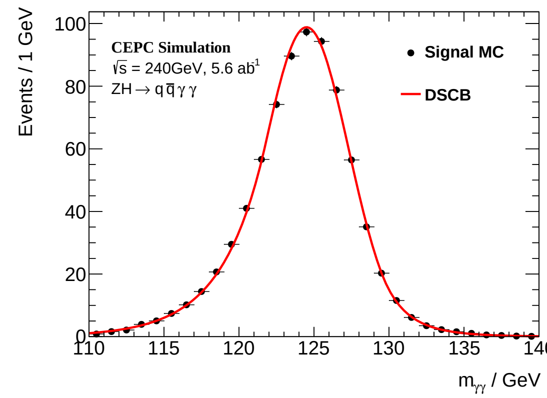
Table 7
Input variables for BDTG in $\nu\bar{\nu}\gamma\gamma$ channel.

model is described by a Double Side Crystal Ball (DSCB) function:

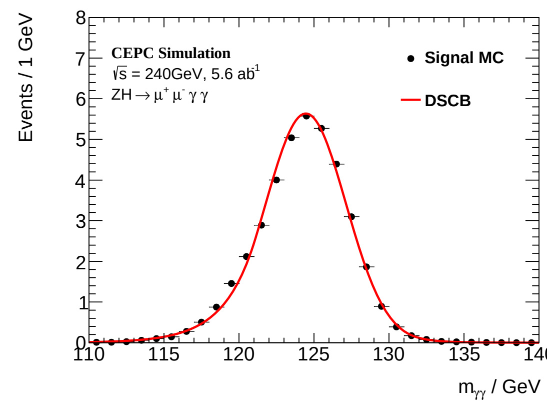
$$f(t) = N \times \begin{cases} e^{-t^2/2} & \text{if } -\alpha_{low} \le t \le \alpha_{high} \\ \dfrac{e^{-\frac{1}{2}\alpha_{low}^2}}{\left[\frac{1}{R_{low}}(R_{low}-\alpha_{low}-t)\right]^{n_{low}}} & \text{if } t < -\alpha_{low} \\ \dfrac{e^{-\frac{1}{2}\alpha_{high}^2}}{\left[\frac{1}{R_{high}}(R_{high}-\alpha_{high}+t)\right]^{n_{high}}}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

where N is a normalization factor, $t = (m_{\gamma\gamma} - \mu_{\text{CB}})/\sigma_{\text{CB}}$. 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.

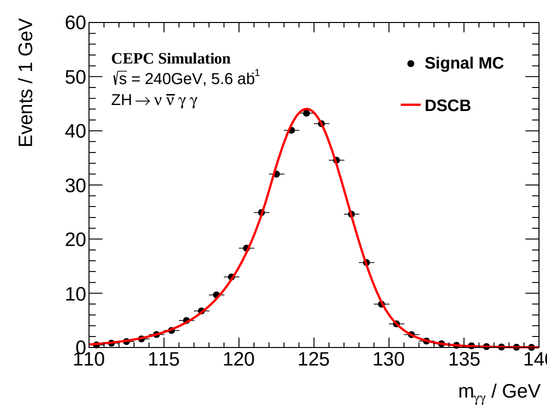
Several smooth functions (Cheybychev polynomials, exponential families and polynomial families) are tested for the background modelling. Finally the 2^{nd} order Cheybychev function is selected according to the fitted χ^2 and number of degree of freedom. The fitted results and MC $m_{\gamma\gamma}$ distribution are shown in Figure 3.



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



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

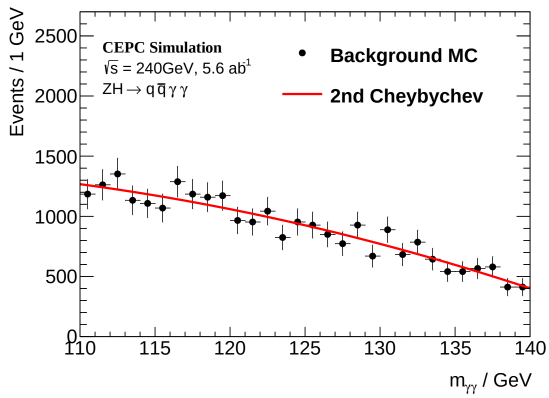


(c) $\nu\bar{\nu}\gamma\gamma$ signal

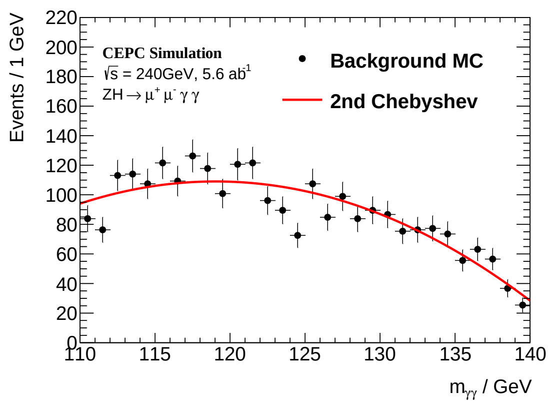
Figure 2: The signal MC and the fitted DSCB model in 3 channels.

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.

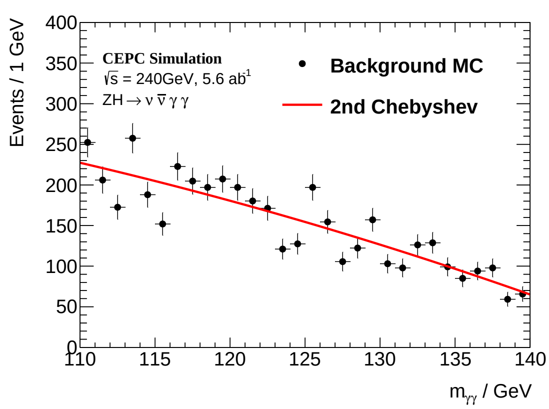
The strategies in constructing BDTG ensure the reasonable independence between the BDTG response and $m_{\gamma\gamma}$. Therefore a 2-dimension model from the multiplication of $m_{\gamma\gamma}$ and BDT models is applied to describe the signal and



(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.

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 $\nu\bar{\nu}\gamma\gamma$ chananels. The corresponding correlation coffecients for the background are 11.6%, 28.2% and 28.4% respectively.

6. Results

The signal strength $\mu = \frac{N\ (e^+e^-\to \mathrm{ZH}\to f\bar{f}\gamma\gamma)}{N_{SM}\ (e^+e^-\to \mathrm{ZH}\to f\bar{f}\gamma\gamma)}$ is extracted by a combined fit in three channels with the unbinned maximum likelihood fit method. The likelihood function is defined as:

$$\mathcal{L}(m_{\gamma\gamma}) = \prod_{c}(Pois(n|\mu\cdot S + B)\cdot \prod_{i}^{n}\frac{\mu\cdot S\times f_S(m_{\gamma\gamma}, BDT) + B\times f_B(m_{\gamma\gamma}, BDT)}{\mu S + B}), \tag{3}$$

in which:

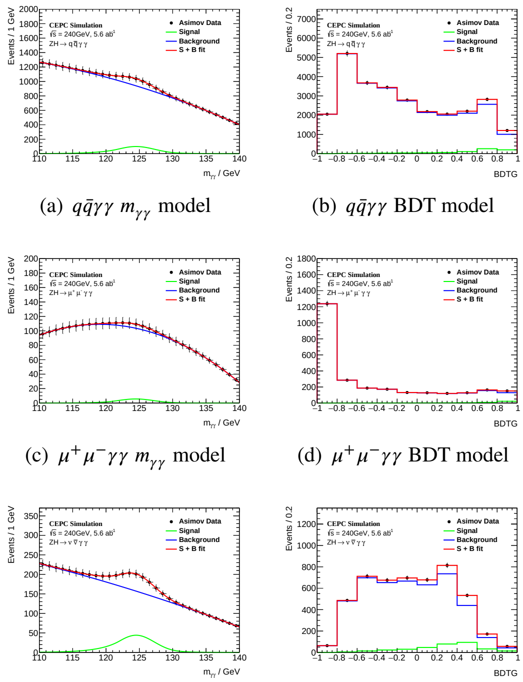
- μ is the signal strength, which is the parameter of interest (POI) in this analysis;
- n is the observed event number in the channel c;
- S and B are the expected signal and background event yields in the channel;
- $f_S(m_{\gamma\gamma}, BDT)$ and $f_B(m_{\gamma\gamma}, BDT)$ are the signal and background models in the channel c. They are derived from the MC as described in Sec. 5.

In the fitting the signal model parameters are fixed to the value in fitting the signal MC. The background yield and model parameters are floated, meaning the effect of background mis-modeling are considered and no constraints from the model-dependent cross section information is exploited.

In order to mimic the real data and avoid the statistical fluctuations of MC samples, a set of Asimov data [30] are generated from the signal + background models and are simultaneouslly 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 6.9% for $\delta(\sigma \times BR)\backslash(\sigma \times BR)$ can be reached in $H \to \gamma\gamma$ measurement in the CEPC with 5.6 ab^{-1} data. With the 20 ab^{-1} data of the updated CEPC opeartion period, this precision can reach to 3.6%. Results are summarized in Table 8. In the latest ATLAS measurement [8] $\mu = 1.04 \pm 0.06(\mathrm{stat.})^{+0.06}_{-0.05}(\mathrm{theory\ syst.})^{+0.05}_{-0.04}(\mathrm{exp.\ syst.})$. The dominant term in the theory uncertainties is the higher order QCD effects, which is expected to be well controlled in the lepton collider like CEPC. In the experimental uncertainties the photon energy resolution and the photon efficiency play an leading role, while at present CEPC study the dedicated contributions are not practicable to estimate. The background is believed to be well modeled from the sufficient sideband data so that this uncertainty is neglected.

7. $Br(H \to \gamma\gamma)$ 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



(a) $q\bar{q}\gamma\gamma$ $m_{\gamma\gamma}$ model (b) $q\bar{q}\gamma\gamma$ BDT model

(c) $\mu^+\mu^-\gamma\gamma$ $m_{\gamma\gamma}$ model (d) $\mu^+\mu^-\gamma\gamma$ BDT model

(e) $\nu\bar{\nu}\gamma\gamma$ $m_{\gamma\gamma}$ model (f) $\nu\bar{\nu}\gamma\gamma$ BDT model

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

Channel	μ @ 5.6 ab^{-1}	μ @ 20 ab^{-1}
$q\bar{q}\gamma\gamma$	1.00 ± 0.088	1.00 ± 0.046
$\mu^+\mu^-\gamma\gamma$	1.00 ± 0.357	1.00 ± 0.192
$\nu\bar{\nu}\gamma\gamma$	1.00 ± 0.114	1.00 ± 0.060
Combined	1.00 ± 0.069	1.00 ± 0.036

Table 8
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 $H \to \gamma\gamma$ 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 by a homogeneous crystal ECAL. This new ECAL is expected to have better photon resolution 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:

$$\frac{\sigma_E}{E} = A \oplus \frac{B}{\sqrt{E}} \oplus \frac{C}{E}, \tag{4}$$

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 emploied in this simplified study to focus on the photon detection only. The signal model is replaced by a simple Gaussian function since the photon energy comes from the 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 $\Delta\mu/\mu$. 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:

$$\frac{\delta\mu}{\mu} = p_0 \oplus (p_1 \times B), \tag{5}$$

where p_0 and $p_1 \times B$ represent the contributions from constant term and stochastic term respectively. A "critical point" can be defined with this relation: the two components in resolution have the same contribution to $\Delta\mu/\mu$, i.e. $p_0 = p_1 B$, also can be written as:

$$\delta\mu|_{B_C} = \sqrt{2}\delta\mu|_{B=0}. \tag{6}$$

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.

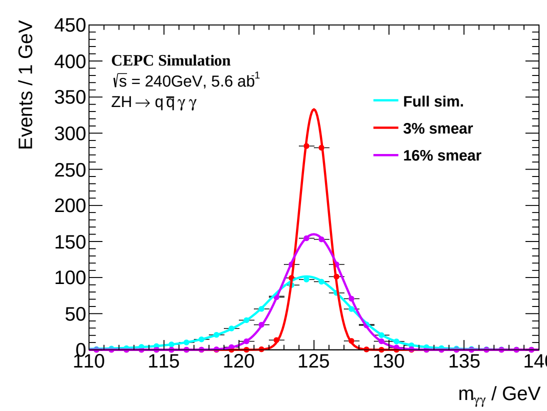


Figure 5: The signal shape for the full simulated $H \to \gamma\gamma$ sample (blue) and for two samples with smeared photon energy (3% in red and 16% in violet). The fitted signal width are 2.81 GeV, 0.94 GeV and 1.96 GeV respectively.

8. Conclusion

This paper presents the expected precision for the measurement of the $H \to \gamma\gamma$ branching ratio in the CEPC via

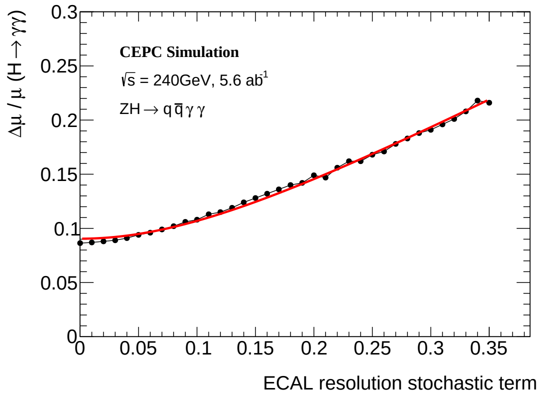


Figure 6: The $Br(H \to \gamma\gamma)$ precision in ZH $\to q\bar{q}\gamma\gamma$ channel as a function of the stochastic term in ECAL resolution from a fast analysis. The points are fitted with Eq. 5.

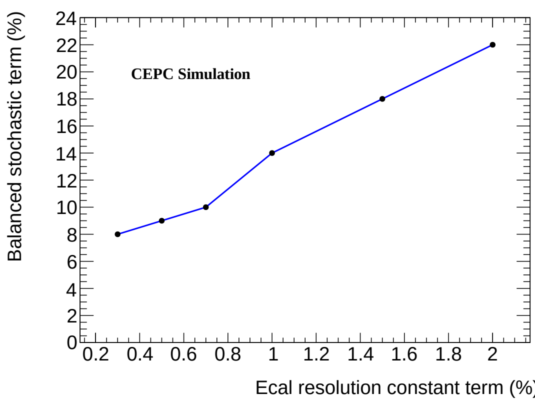


Figure 7: The balanced ECAL stochastic resolution points with different configurations of the constant term.

$ZH \to q\bar{q}\gamma\gamma$, $ZH \to \mu^+\mu^-\gamma\gamma$, $ZH \to \nu\bar{\nu}\gamma\gamma$ 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. With the scheduled integrated luminosity of 5.6 ab^{-1} a precision of 6.9% is expected to be achived at the CEPC. With 20 ab^{-1} data this precision can be 3.6%. The ECAL performance in the CEPC is further studied by smearing photon energy resolution in $q\bar{q}\gamma\gamma$ channel. A direct relationship between the ECAL resolution and the precision in $H \to \gamma\gamma$ measurement is foreseen.

9. 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] A. 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] C. 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] L. design study working group collaboration, First look at the physics case of TLEP, Journal of High Energy Physics 2014 (2014). URL: https://doi.org/10.1007%2Fjhep01%282014%29164. doi:10.1007/jhep01(2014)164.

[6] G. CERN, Cern yellow reports, vol 4 (2016): Updated baseline for a staged compact linear collider, 2016. URL: https://e-publishing.cern.ch/index.php/CYR/issue/view/24. doi:10.5170/CERN-2016-004.

[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 $h \to \gamma\gamma$ 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] Projections for measurements of Higgs boson signal strengths and coupling parameters with the ATLAS detector at a HL-LHC, Technical Report, CERN, Geneva, 2014. URL: http://cds.cern.ch/record/1956710. doi:10.17181/CERN.B5WP.VPT7, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-016.

[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.

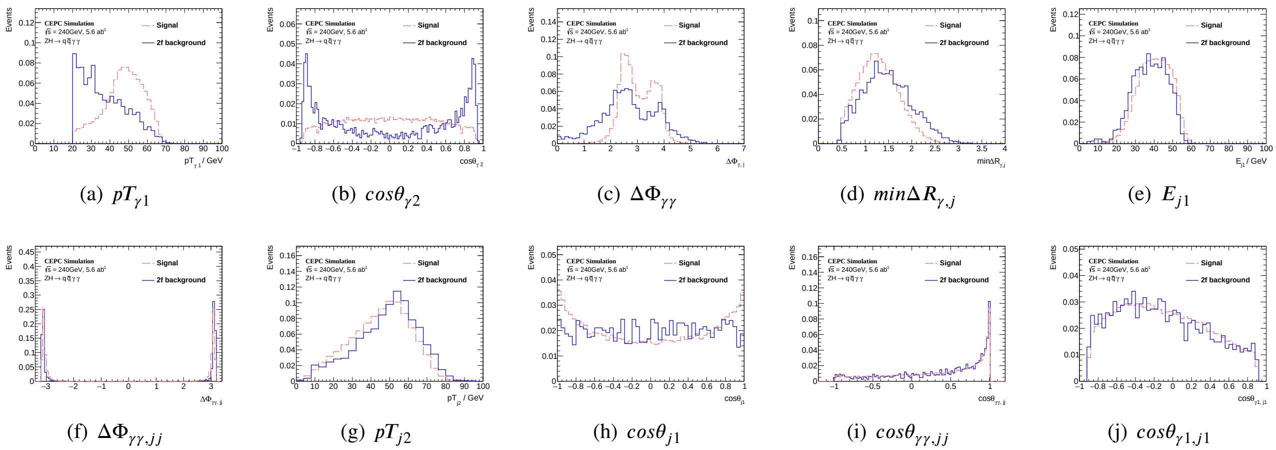
[15] M. Thomson, Particle flow calorimetry and the PandoraPFA algorithm, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 611 (2009) 25–40. doi:10.1016/j.nima.2009.09.009.

[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.



(a) $pT_{\gamma1}$ (b) $cos\theta_{\gamma2}$ (c) $\Delta\Phi_{\gamma\gamma}$ (d) $min\Delta R_{\gamma,j}$ (e) E_{j1}

(f) $\Delta\Phi_{\gamma\gamma,jj}$ (g) pT_{j2} (h) $cos\theta_{j1}$ (i) $cos\theta_{\gamma\gamma,jj}$ (j) $cos\theta_{\gamma1,j1}$

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

[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.

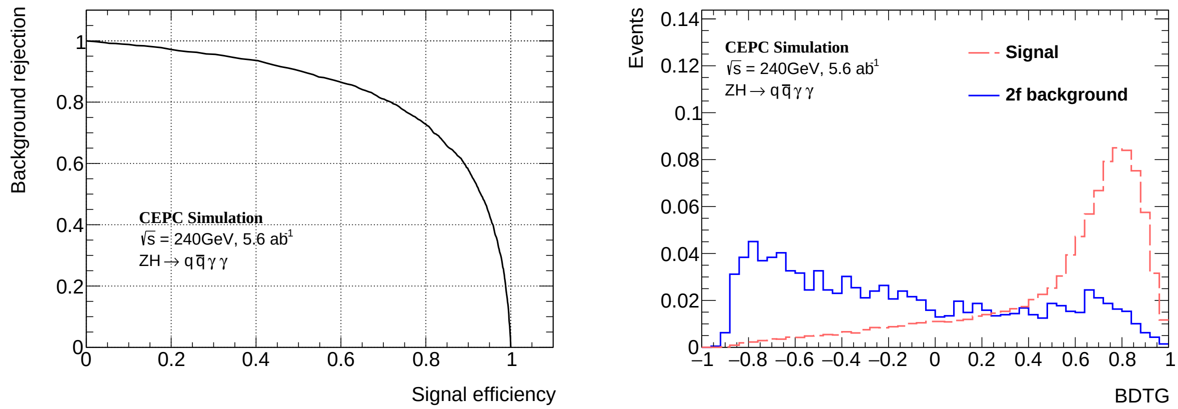
[27] S. Catani, Y. Dokshitzer, M. Olsson, G. Turnock, B. Webber, New clustering algorithm for multijet cross sections in e^+e^- annihilation, Physics Letters B 269 (1991) 432–438. doi:https://doi.org/10.1016/0370-2693(91)90196-W.

[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.

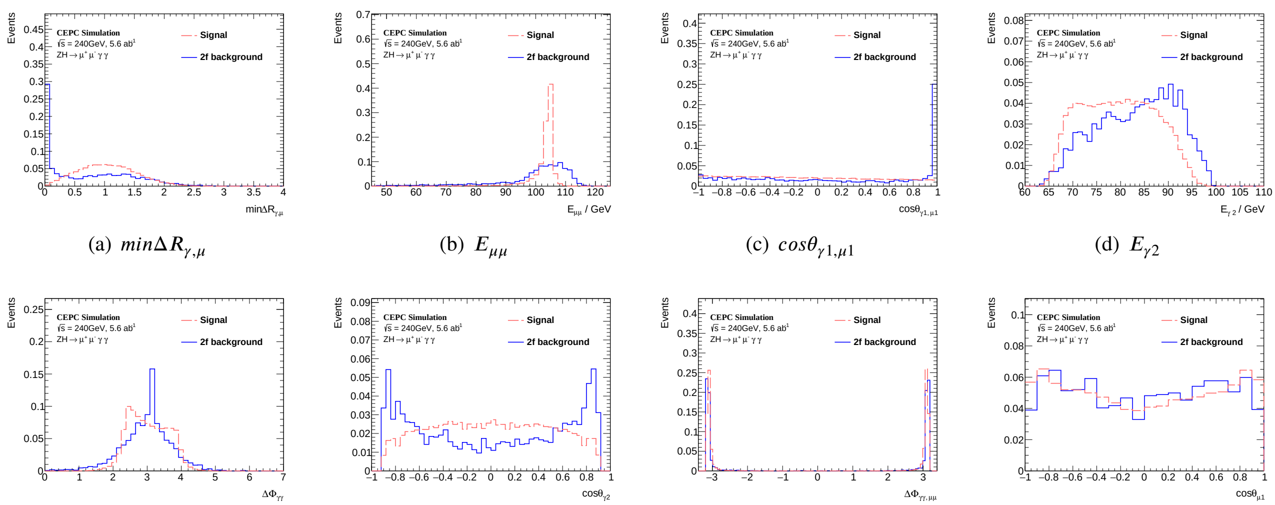
[30] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, The European Physical Journal C 71 (2011). doi:10.1140/epjc/s10052-011-1554-0.

A. Appendix



(a) (b)

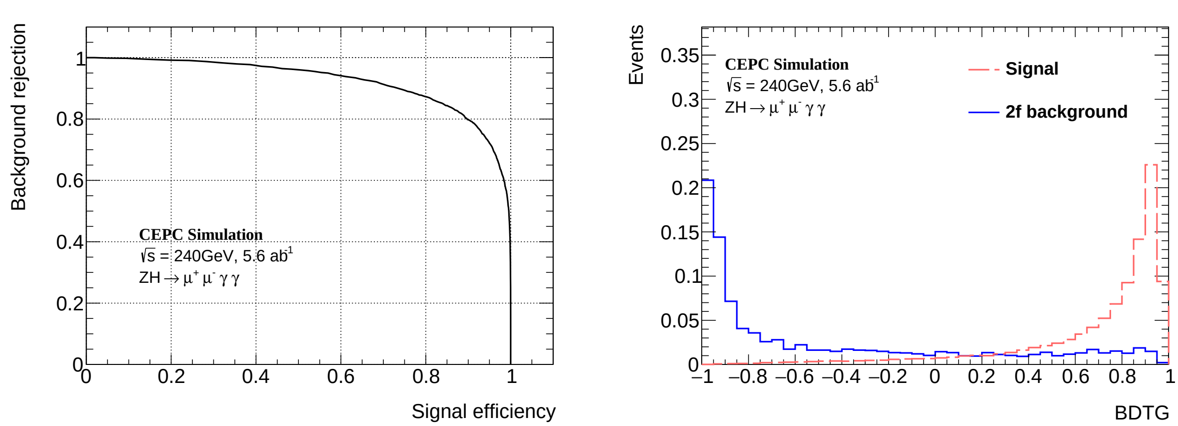
Figure 9: The ROC curve (left) and output BDTG distribution (right) in $q\bar{q}\gamma\gamma$ channel.



(a) $min\Delta R_{\gamma,\mu}$ (b) $E_{\mu\mu}$ (c) $cos\theta_{\gamma1,\mu1}$ (d) $E_{\gamma2}$

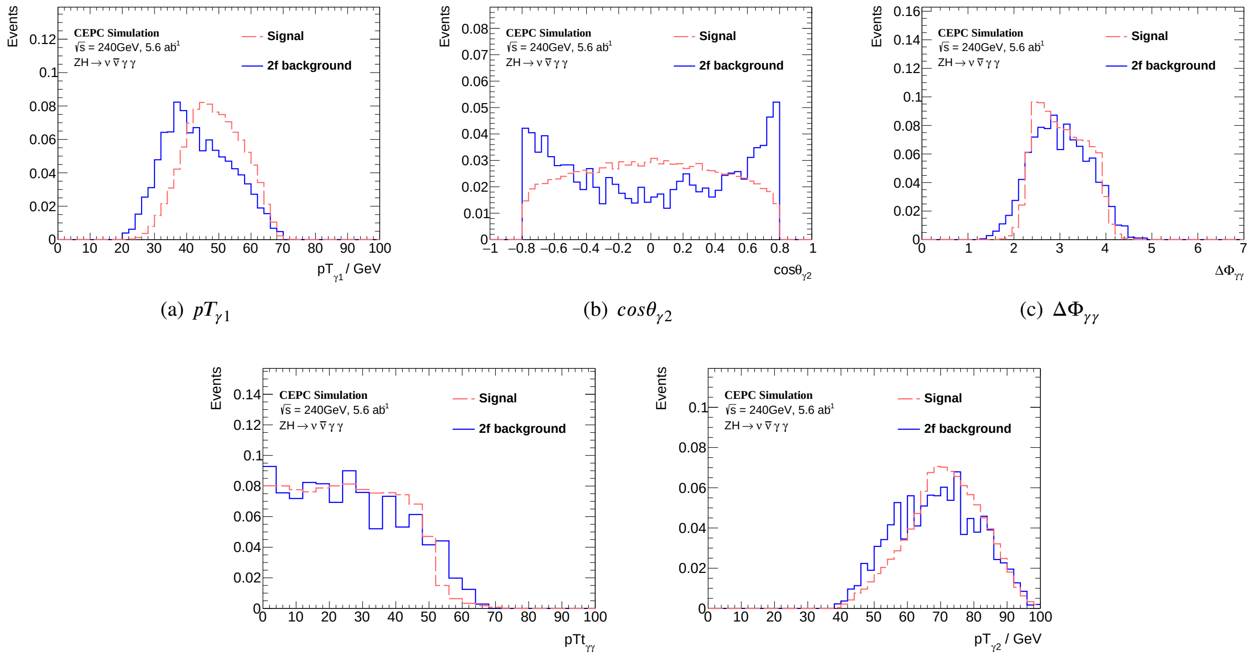
(e) $\Delta\Phi_{\gamma\gamma}$ (f) $cos\theta_{\gamma2}$ (g) $\Delta\Phi_{\gamma\gamma,\mu\mu}$ (h) $cos\theta_{\mu1}$

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



(a) (b)

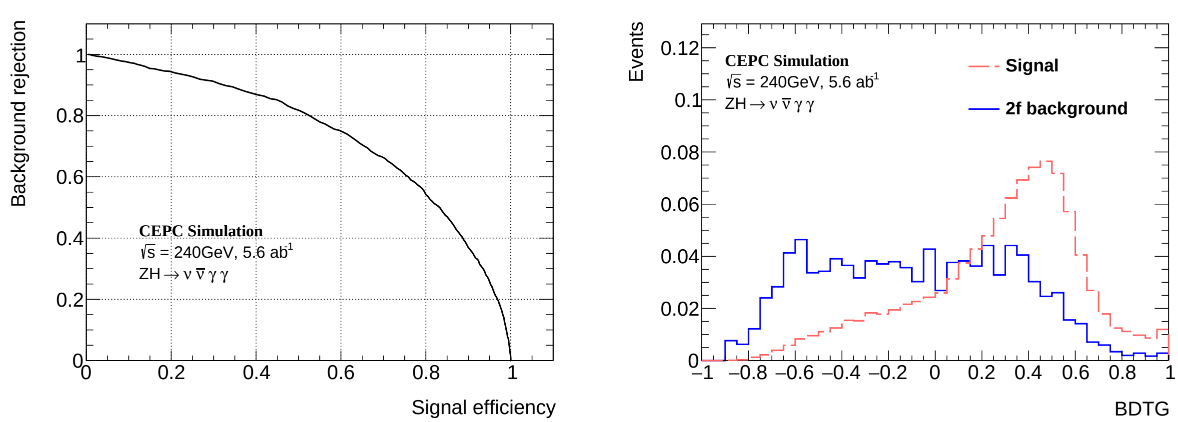
Figure 11: The ROC curve (left) and output BDTG distribution (right) in $\mu^+\mu^-\gamma\gamma$ channel.



(a) $pT_{\gamma1}$ (b) $cos\theta_{\gamma2}$ (c) $\Delta\Phi_{\gamma\gamma}$

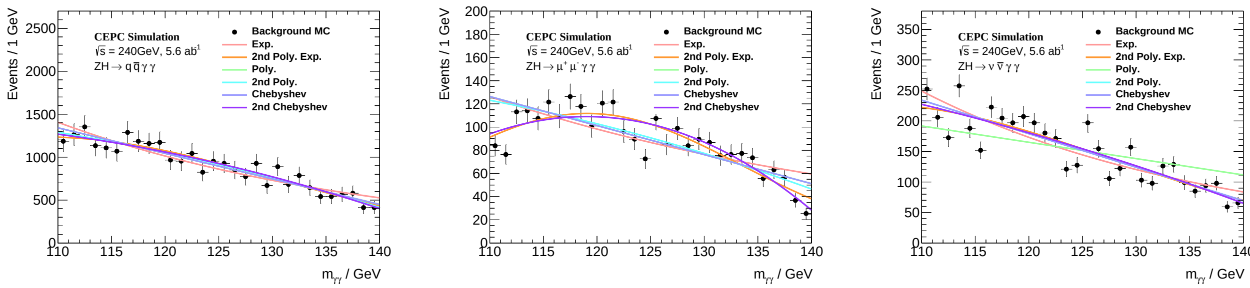
(d) $pTt_{\gamma\gamma}$ (e) $pT_{\gamma2}$

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



(a) (b)

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



(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 $\chi^2/Ndof$ value. Detailed numbers are listed in Table 9.

	$q\bar{q}\gamma\gamma$	$\mu^+\mu^-\gamma\gamma$	$\nu\bar{\nu}\gamma\gamma$
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	3.420
2nd order Chebyshev	0.596	1.789	3.411

Table 9
The χ^2/Ndof values for 6 considered models in the background modeling in 3 channels, including the first and second order exponential, polynomial and Chebyshev functions.