Measurement of $H\to \mu^+\mu^-$ production in association with a Z boson at the CEPC

Zhen-Wei Cui(崔震巍)\textsuperscript{1,1) } Qiang Li(李强)\textsuperscript{1,2,2) } Gang Li(李刚)\textsuperscript{2,3 } Man-Qi Ruan(阮曼奇)\textsuperscript{2,3 } 
Lei Wang(王晓)\textsuperscript{1 } Da-Neng Yang(杨大能)\textsuperscript{1} 
\textsuperscript{1} Department of Physics and State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing 100871, China 
\textsuperscript{2} CAS Center for Excellence in Particle Physics, Beijing 100049, China 
\textsuperscript{3} Institute of High Energy Physics, Beijing 100049, China

Abstract: The Circular Electron-Positron Collider (CEPC) is a future Higgs factory proposed by the Chinese high energy physics community. It is planned to operate at a center-of-mass energy of 240–250 GeV and is expected to accumulate an integrated luminosity of 5 ab$^{-1}$ over ten years of operation. At the CEPC, Higgs bosons will be dominantly produced from the ZH associated process. The vast number of Higgs events collected will enable precise studies of its properties, including Yukawa couplings to massive particles. With GEANT4-based simulation of detector effects, we study the feasibility of measuring the Higgs boson decaying into a pair of muons at the CEPC. The results with and without information from the Z boson decay products are provided, showing that a signal significance of over 10 standard deviations can be achieved and the H-$\mu$-$\mu$ coupling can be measured within 10% accuracy.

Keywords: Higgs, CEPC, Yukawa coupling

PACS: 13.66.Fg, 14.80.Bn, 13.66.Jn

DOI: 10.1088/1674-1137/42/5/053001

1 Introduction

The discovery of the Higgs-like boson completes the particle table of the Standard Model (SM) of particle physics. Up-to-date LHC measurements all indicate that the Higgs boson is indeed highly SM-like [1–6]. In the SM, Higgs couplings to massive particles are proportional to their mass (squared). Hence, the event rate of Higgs bosons to the first and second generation of massive fermions can be very small, making them difficult to measure at the LHC. The Circular Electron-Positron Collider (CEPC) [7], however, is designed to run at around 240–250 GeV with a combined instantaneous luminosity of $2 \times 10^{34}$ cm$^{-2}$ s$^{-1}$ for the two planned experiments, and will deliver 5 ab$^{-1}$ of integrated luminosity over ten years of running. The huge amount of data will enable precise measurement of the branching ratios of the Higgs to light fermions and determine the associated Yukawa couplings, including H-$\mu$-$\mu$, which is crucial to validate the consistency of the SM Higgs mechanism, since any deviation would indicate the existence of new physics.

Searches for $H\to \mu^+\mu^-$ production have been performed at the ATLAS and CMS experiments with Run-I and Run-II data [8–10]. The most stringent observed (expected) upper limit on the cross-section times branching ratio is found to be 2.8 (2.9) times the SM prediction [10]. Projections have also been made for the High Luminosity-LHC assuming an integrated luminosity of 3000 fb$^{-1}$ collected by the ATLAS or CMS detector, which can lead to a signal significance of about 7 $\sigma$ [11] with an accuracy of around 20% [12]. Studies have also been performed for the International Linear Collider (ILC). Considering a center mass energy of 250 GeV and an integrated luminosity of 250 fb$^{-1}$, the signal is dominated by the Higgs-strahlung from a Z boson and the signal significances for the sub-processes with a Z boson decaying into $\nu\bar{\nu}$ and $q\bar{q}$ are found to be 1.8 and 1.1 $\sigma$, respectively [13]. Then, accumulating 2000 fb$^{-1}$ at 250 GeV and 4000 fb$^{-1}$ at 500 GeV with actual beam polarization, the precision of signal strength becomes 20.5% and 15.4% respectively [14]. At a center-of-mass energy of 1 TeV with an integrated luminosity of 500 fb$^{-1}$, the...
signal is dominated by the WW-fusion process and a sensitivity of 2.75 \( \sigma \) can be achieved [15].

At the CEPC, the signal \( H \to \mu^+\mu^- \) production is dominated by the Higgs-strahlung from a Z boson. We perform a feasibility study based on events generated at leading order accuracy with initial state radiation (ISR), parton shower, hadronization and detector effects simulated.

Considering that 70% of the Z bosons decay hadronically and 20% decay invisibly, we focus on two scenarios: one for Z boson inclusive decay and the other for hadronic decay. The first case maximally exploits the statistics of the produced H\( \to \mu^+\mu^- \) events and the second takes advantage of the major part of the decay kinematics. For both cases, we first perform a cut-based analysis and then improve the measurement using a Boosted Decision Tree (BDT) technique.

This paper is organized as follows. Section 2 describes event generation and simulation. Section 3 presents results for the inclusive measurement. Section 4 presents results for the Z\( \to \ell\ell \) decay channel. Section 5 summarizes the paper.

## 2 Monte Carlo simulation

At a 250 GeV CEPC, Higgs bosons will mainly be produced through Higgs-strahlung, i.e. \( e^+e^- \to ZH \). With an integrated luminosity of 5000 fb\(^{-1}\), about 230 of our signal events \( H \to \mu^+\mu^- \) can be produced. The expected background to the signal production includes 2-fermion processes \( e^+e^- \to ff \), where \( f \) can be any SM fermion other than the top quark, and 4-fermion processes, which can be mediated through associated ZZ, WW, ZZ, WW production and a single Z boson production. All Monte Carlo (MC) events are generated with the WHIZARD V1.9.5 [16] event generator at parton level with ISR and interference effects included. The generated events are interfaced to PYTHIA 6 [17] for parton shower and hadronization simulation. Detector effects are simulated with the CEPC detector implemented with Mokka/GEANT4 [7, 18, 19]. The detector is assumed to have a similar structure to the International Large Detector (ILD) [20, 21] at the ILC [22]. At the CEPC, the muon identification efficiency is expected to be over 99.5% for \( P_T \) larger than 10 GeV, and with excellent \( P_T \) resolution of \( \sigma_{1/P_T} = 2 \times 10^{-5} \pm 1 \times 10^{-3} / (P_T \sin \theta) \). The fully simulated events are reconstructed with a particle-flow algorithm ArborPFA [23]. More details about the CEPC sample set can be found in Ref. [24].

The major SM backgrounds, including all the 2-fermion processes \( e^+e^- \to ff \), where \( ff \) refers to all lepton and quark pairs except \( tt \) and 4-fermion processes (ZZ, WW, ZZ or WW, single Z). The initial state radiation (ISR) and all possible interference effects are automatically taken into account in the generation. The classification for four fermion production follows that of LEP [25], depending crucially on the final state. For example, if the final states consist of two mutually charge conjugated fermion pairs that could decay from both WW and ZZ intermediate states, such as \( e^+e^- + \nu_\mu \bar{\nu}_\mu \), this is classified as a “ZZ or WW” process. If there is \( e^\pm \) together with its partner neutrino and an on-shell W boson in the final state, this is called a “single W” process. Meanwhile, if there is an electron-positron pair and an on-shell Z boson in the final state, it is called a “single Z” process. Detailed information on the 2-fermion and 4-fermion samples used in our analyses are listed in Tables A1 and A2.

## 3 Inclusive analysis

A recoil mass method enables a measurement of the \( H \to \mu^+\mu^- \) production without measuring the associated Z boson decay. We define the recoil mass as

\[
M_{\text{recoil}}^2 = s + M_H^2 - 2E_H\sqrt{s},
\]

where \( \sqrt{s} \) is the center-of-mass energy, and \( M_H \) and \( E_H \) correspond to the reconstructed mass and energy of the Higgs boson. The ZH (H\( \to \mu^+\mu^- \)) events form a peak in the \( M_{\text{recoil}} \) distribution at the Z boson mass window.

We select two muons with the largest transverse momenta and consider selections on the following kinematic variables: invariant mass of the di-muon system \( M_{\mu^+\mu^-} \), recoil mass of the di-muon system \( M_{\text{recoil}}^\mu \), transverse momentum of the di-muon system \( P_T^{\mu^+\mu^-} \), third component of the di-muon momentum \( P_{z_{\mu^+\mu^-}} \), energy of di-muon system \( E_{\mu^+\mu^-} \), and angular variables in the laboratory system frame \( \cos \theta_{\mu^+} \), \( \cos \theta_{\mu^-} \), \( \cos \theta_{\mu^+\mu^-} \), and \( \cos \theta_{\mu^+\mu^-} \), where \( \theta_{\mu^+} \) and \( \theta_{\mu^-} \) means the polar angle of \( \mu^- \), \( \mu^+ \); \( \theta_{\mu^+\mu^-} \) means the angle between \( \mu^- \) and \( \mu^+ \); and \( \theta_{\mu^+\mu^-} \) represents the angle between the Z boson and muon leptons.

### 3.1 Cut-count analysis

The event numbers under selection flow, which are determined by maximizing \( s/\sqrt{s+b} \), with \( s \) and \( b \) represent signal and background yields, are summarized in Table 1. The two mass windows \( M_{\mu^+\mu^-} \) and \( M_{\text{recoil}}^\mu \) are set in accordance with the signal signature. \( P_T^{\mu^+\mu^-} \) and \( P_{z_{\mu^+\mu^-}} \) are set to reduce the ZZ events, where one of the Z bosons decays to \( \mu^- \mu^+ \), and Drell-Yan Z\( \to \mu^+\mu^- \) background. The Higgs and Z boson decays can lead to different \( \cos \theta_{\mu^+} \) and \( \cos \theta_{\mu^-} \) distributions due to the spin-dependence of the couplings and the parity violation of the weak interaction. \( \cos \theta_{\mu^+\mu^-} \) selection is chosen to suppress the 2f background.
An unbinned maximum likelihood fit is performed on the $M_{\mu^+\mu^-}$ distribution. The signal is parameterized by a crystal ball function, with parameters fixed by simulated events. The background is parametrized by a second order Chebychev function chosen by F-test [28].

Figure 2 shows the post-fit result of the invariant mass distribution of the di-muon system. The fitted number of signal events is $\pm 13$. At 68% confidence
level, an accuracy from $-17\%$ to $18\%$ can be achieved for the signal strength based on a likelihood scan. The signal under the peak at 124.4–125.2 GeV leads to a high significance of $8.8\sigma$, via simple counting $\sqrt{2(s+b)\ln(1+s/b)}$.

3.2 BDT optimization

We have also exploited the Toolkit for Multivariate Analysis (TMVA) [26] for further background rejection, using the method of Gradient Boosted Decision Trees (BDTG). After fixing the range of the invariant mass and the recoil mass as mentioned above, 5 variables are taken as inputs to TMVA, including $\cos\theta_{\mu\pm}$, $\cos\theta_{\mu\mp}$ and $P_{Z_{\mu\pm\mu\mp}}$. The choice of these variables is based on many tests and importance ranking. The resulted BDT response distribution can be seen in Figure 3, where the agreement between training and testing samples shows no obvious overtraining. We then take the final event selections as: BDTG response $>0.369$, $20<P_{Z_{\mu\pm\mu\mp}}<64$ GeV and $\cos\theta_{\mu\pm\mu\mp}<-0.996$. A maximum likelihood fit is performed on the resulting invariant mass of the di-muon system. The signal and background probability functions are parametrized in the same form as in the previous cut-count study.

Figure 3 shows the BDT response distribution and the post-fit result of $M_{\mu\pm\mu\mp}$. The fitted number of signal events is $62\pm3$ at 68% confidence level. As in the inclusive channel, we perform a likelihood fit to extract the signal yield and strength parameter. The quality of the fit is demonstrated in Fig. 5. The signal yield from the fit is $75\pm5$. The signal strength can be determined with an uncertainty from $-16\%$ to $17\%$, at 68% confidence level. The signal significance under the peak at 124.3–125.2 GeV is found to be $10.8\sigma$.

4 $Z(q\bar{q})H(\mu\mu)$ analysis

Of all the $Z$ boson decay modes, the hadronic channel is most promising, due to its large branching fraction ($\sim70\%$). The exclusive method of the $k_t$ algorithm for $e^+e^-$ collisions in Fastjet [27] is used to reconstruct two jets with the particles expect the chosen $\mu^-$ and $\mu^+$, and the jets are sorted by energy. We perform an analysis on the $Z(q\bar{q})H(\mu\mu)$ production. Apart from the previously mentioned variables related to the $H(\mu\mu)$ system, we further exploit the following selections on jets: the third component of di-jet system momentum $P_{Z_{jj}}$, the recoil mass of the di-jet system $M_{jj}^{\text{recoil}}$ mass of jets $M_{j1,2}$, and the invariant mass of the di-jet system $M_{jj}$.

4.1 Cut-count analysis

A cut-count analysis is performed for the exclusive analysis. The event flow under selections are summarized in Table 2. Selections on single and di-jet masses eliminate most background without hard jets. The recoil mass cut further reduces the $Z(ll)/Z(q\bar{q})$ background.

As in the inclusive channel, we perform a likelihood fit to extract the signal yield and strength parameter. The quality of the fit is demonstrated in Fig. 5. The signal yield from the fit is $75.5\pm12.5$. The signal strength can be determined with an uncertainty from $-16\%$ to $17\%$, at 68% confidence level. The signal significance under the peak at 124.3–125.2 GeV is found to be $10.8\sigma$.
4.2 BDT improvement

In order to achieve the highest significance, we perform a two-step multivariate analysis. The first step exploits a MLP (multi layer perceptron) [26] method to suppress the fully leptonic WW and ZZ backgrounds. After applying $M_{\text{rec}}^{\mu^+\mu^-} > 90$ GeV, 4 variables including $M_{j1,2}$, $M_{jj}$ and $M_{\text{rec}}^{\mu^+\mu^-}$ are considered as inputs for the MLP. The effectiveness of this MLP is shown in Fig. 6. After requiring the MLP response to be greater than 0.71, we exploit a BDTG to further reduce the backgrounds from semileptonic ZZ and WW. In this second step, the variables $\cos\theta_{\mu_1\mu_2}$, $\cos\theta_{\mu_1\mu_2}$, $P_{Z_{1\mu_1\mu_2}}$, $P_{Z_{1\mu_1\mu_2}}$, $M_{jj}$, and $M_{j1,2}$ are taken as inputs.

After the two-step multivariate analysis, we require a BDTG response $> -0.13$, $90.4 < M_{\text{rec}}^{\mu^+\mu^-} < 93$ GeV, and $28 < P_{Z_{\mu_1\mu_2}} < 64$ GeV. Finally, we perform a likelihood fit to extract the signal yield and strength parameter, as
shown in Fig. 7. The signal yield from the fit is 73.4±12.4. Based on a likelihood scan, the signal strength can be determined with an uncertainty from -16% to 17%, at 68% confidence level. The significance of the signal in the peak region 124.3–125.2 GeV is found to be 10.8σ, which means the best boundary to distinguish signal and background is nearly an N-dimensional rectangle in the parameter phase space.

5 Summary

The feasibility of measuring \( H \rightarrow \mu^+ \mu^- \) at the CEPC has been studied considering center-of-mass energy 250 GeV collisions and 5000 fb\(^{-1}\) integrated luminosity. The measurement was performed in two complementary channels: ZH production without measuring the Z boson decay, and ZH production with the Z boson decaying hadronically. For each decay channel, a cut-count analysis was tested and followed with an improvement using multivariate techniques. Similar results are obtained from the two channels.

Finally, we want to mention in brief possible systematic factors affecting this future analysis. By referring mostly to Ref. [19], which shares most of the common factors, the systematics uncertainty should be under control, while the statistical uncertainty will dominate for this analysis at the CEPC.

Over 10 \( \sigma \) significance can be reached for the signal \( H \rightarrow \mu^+ \mu^- \) process. The accuracy of the signal strength can be measured with ±17% uncertainty and the associated \( H-\mu-\mu \) coupling can be restricted to 10% level. The results are comparable to the High-Luminosity LHC.

The authors would like to thank Xin Mo, Dan Yu and Yuqian Wei for useful discussions.

Appendix A

Table A1. Details of the two-fermion background samples.

| process | final states | \( \sigma/\text{fb} \) | events expected |
|---------|--------------|----------------|----------------|
| uu      | u, \( \bar{u} \) | 9995.35 | 50476527 |
| dd      | d, \( \bar{d} \) | 9808.71 | 49533965 |
| cc      | c, \( \bar{c} \) | 9974.20 | 50369725 |
| ss      | s, \( \bar{s} \) | 9805.39 | 49517234 |
| bb      | b, \( \bar{b} \) | 9803.04 | 49505372 |
| qq      | q, \( \bar{q} \) | 49561.30 | 250284565 |
| e2e2    | \( \mu^- \mu^+ \) | 4967.58 | 25084566 |
| e3e3    | \( e^-e^+ \) | 4374.94 | 22093447 |
| bhabha  | e^-e^+\gamma | 24992.21 | 126210660 |

Table A2. Details of the four-fermion background samples.

| process | final states | \( \sigma/\text{fb} \) | events expected |
|---------|--------------|----------------|----------------|
| ZZ(h)utut | up, up, up, up | 83.09 | 419604 |
| ZZ(h)ddtt | down, down, down, down | 226.2 | 1142310 |
| ZZ(h)uqnotd | uq, uq, (sq, bq), (sq, bq) | 95.65 | 483032 |
| ZZ(h)cqnots | cq, cq, (dq, bq), (dq, bq) | 96.04 | 485092 |
| ZZ(al)nuup | nu_\mu,\tau, nu_\mu,\tau, up, up | 81.72 | 412686 |
| ZZ(al)nudn | nu_\mu,\tau, nu_\mu,\tau, down, down | 134.86 | 681043 |

Continued on next page
