

Particle-level transformers for 95 GeV Higgs boson searches at future e^+e^- Higgs factories

Yabo Dong¹, Manqi Ruan², Kun Wang^{3,*}, Haijun Yang⁴, and Jingya Zhu^{1,†}

¹*School of Physics and Electronics, Henan University, Kaifeng 475004, China*

²*Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China*

³*College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China*

⁴*State Key Laboratory of Dark Matter Physics, Key Laboratory for Particle Astrophysics and Cosmology (MOE), Shanghai Key Laboratory for Particle Physics and Cosmology (SKLPPC), School of Physics and Astronomy & Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai 200240*



(Received 8 December 2025; accepted 24 February 2026; published 16 March 2026)

Motivated by several mild excesses around 95 GeV, we investigate the prospects for a light scalar S produced via Higgsstrahlung, $e^+e^- \rightarrow Z(\mu^+\mu^-)S$, at future e^+e^- Higgs factories. We take the CEPC as a benchmark, with a center-of-mass energy of $\sqrt{s} = 240$ GeV and an integrated luminosity of $L = 20 \text{ ab}^{-1}$. We focus on the decay modes $S \rightarrow \tau^+\tau^-$ and $S \rightarrow b\bar{b}$. To maximize sensitivity, we employ the particle-level transformer networks particle transformer (ParT) and its more-interactive variant MIParT, which exploit the features of all reconstructed objects and their correlations. For a representative signal benchmark, this approach improves the expected statistical precision on the signal strength by factors of 2.3 in the $\tau^+\tau^-$ channel and 1.4 in the $b\bar{b}$ channel compared to a cut-based analysis. Within the flipped next-to-two-Higgs-doublet model (N2HDM-F), the CEPC can measure the signal strength with a statistical precision down to 1.0% in the $\tau^+\tau^-$ channel and 0.69% in the $b\bar{b}$ channel using MIParT. It can achieve a 5σ discovery for $\mu_{\tau\tau}^{ZS} > 1.6 \times 10^{-2}$ or $\mu_{b\bar{b}}^{ZS} > 5.0 \times 10^{-3}$, and reach 1% precision for $\mu_{\tau\tau}^{ZS} > 0.93$ or $\mu_{b\bar{b}}^{ZS} > 0.14$. These gains are expected to qualitatively carry over to other future lepton colliders such as FCC-ee and the ILC. Our results demonstrate the potential of particle-level machine-learning techniques to strengthen light Higgs searches at future e^+e^- Higgs factories.

DOI: 10.1103/13h1-zdmz

I. INTRODUCTION

The discovery of the Higgs boson at the LHC in 2012 [1,2] completed the last missing piece of the Standard Model (SM) puzzle. However, several unresolved questions continue to cast shadows over the SM, among which the origin of the baryon asymmetry of the Universe remains one of the most compelling [3–5]. According to the Sakharov conditions [6], three criteria must be fulfilled to generate the observed matter-antimatter asymmetry: (i) baryon number violation, (ii) sufficient C and CP violation, and (iii) departure from thermal equilibrium. In the SM, the C and CP violations are too weak, and the electroweak phase transition (EWPT) associated with the 125 GeV Higgs boson is merely a smooth crossover rather

than a strong first-order transition [7,8]. These shortcomings strongly suggest that new physics beyond the SM is required to account for the baryon asymmetry. A minimal yet well-motivated approach is to extend the Higgs sector by introducing additional scalar fields. Such extensions can naturally provide new sources of C and CP violation and can also induce a strong first-order EWPT, thereby realizing the necessary out-of-equilibrium conditions [9].

Several experimental excesses near 95 GeV have been reported across different search channels, suggesting intriguing hints of a light scalar resonance. The earliest indication dates back to 2003, when the LEP experiment observed a local significance of 2.3σ for a hypothetical lighter Higgs boson S with a mass around 98 GeV in the process $e^+e^- \rightarrow ZS(\rightarrow b\bar{b})$ [10]. In 2018, the CMS Collaboration reported a local excess of 2.8σ in the $gg \rightarrow S \rightarrow \gamma\gamma$ channel at a mass of 95.3 GeV, based on the combined 8 TeV (19.7 fb^{-1}) and 13 TeV (35.9 fb^{-1}) datasets [11]. Additional evidence emerged in 2022, when CMS reported a 3.1σ excess in the $pp \rightarrow S \rightarrow \tau^+\tau^-$ channel near 100 GeV [12]. More recently, in 2024, both major LHC Collaborations released updated diphoton results: CMS observed a 2.9σ excess at 95.4 GeV [13],

*Contact author: kwang@usst.edu.cn

†Contact author: zhuji@henu.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

while ATLAS reported a milder 1.7σ excess at 95.3 GeV in $pp \rightarrow S \rightarrow \gamma\gamma$ [14]. Taken together, these observations may point toward the existence of a light Higgs boson with a mass around 95.5 GeV.

The collider phenomenology of the hypothetical light Higgs boson has been extensively explored in recent studies through Monte Carlo (MC) simulations [15–19]. Hadron colliders, such as the Large Hadron Collider (LHC), operate at high center-of-mass energies, enabling the exploration of a wide range of physical processes [20,21]. However, the complex QCD background at hadron colliders poses a major challenge for precision measurements and for probing subtle new-physics signals. Our previous analysis indicates that achieving a 5σ discovery significance in the top-pair-associated diphoton channel, $pp \rightarrow t\bar{t}S(\rightarrow \gamma\gamma)$, at the high-luminosity LHC (HL-LHC) requires a production cross section exceeding approximately 0.3 fb [18]. Signals with smaller cross sections may instead be probed through alternative decay channels or at future collider facilities.

Future Higgs factories, such as the Future Circular Collider (FCC-ee) [22,23], the International Linear Collider (ILC) [24–26], and the Circular Electron-Positron Collider (CEPC) [27–30], provide a promising avenue for probing a possible 95 GeV light Higgs boson and measuring its couplings with high precision. This enhanced sensitivity stems from the much cleaner environment of lepton colliders compared to hadron colliders. Our study focuses on the CEPC, motivated by its suitable center-of-mass energy and high integrated luminosity. The scenarios of the other two colliders will be discussed at the end of the paper. The CEPC aims to study the properties of the 125 GeV Higgs boson with unprecedented accuracy and is expected to deliver over 20 ab^{-1} at 240 GeV [27–30]. Given that the dominant production mechanism for a 95 GeV scalar at the CEPC is the Higgsstrahlung process ($e^+e^- \rightarrow ZS$), this work is dedicated to its detailed investigation.

Our previous work [19] demonstrated that machine learning (ML) algorithms, including extreme gradient boosting (XGBoost), gradient boosting decision trees (GBDT), and deep neural networks (DNN), can enhance the sensitivity of searches for a light Higgs boson around 95 GeV. Furthermore, in recent years, several advanced ML architectures have shown considerable promise in jet tagging tasks [31–37]. Among these, transformer-based models such as the particle transformer (ParT) [36] and the more-interaction particle transformer (MIParT) [37] have demonstrated outstanding performance. Their strong performance stems primarily from the attention mechanism, which effectively captures complex correlations among final-state particles and yields a more expressive representation of jet substructure. This capability precisely addresses a key shortcoming of simpler network architectures, such as DNN. Overall, particle-level ML techniques

offer substantial potential for new-physics searches and are expected to significantly improve both sensitivity and measurement precision at present and future colliders.

The persistent experimental anomalies near 95 GeV have motivated theoretical interpretations. These interpretations can be broadly classified according to the decay channels considered: focus solely on the $\gamma\gamma$ channel [18,38–49], analyze both $\gamma\gamma$ and $b\bar{b}$ channels [50–71] or $\gamma\gamma$ and $\tau^+\tau^-$ channels [72–74], while consider all three channels, $\gamma\gamma$, $b\bar{b}$, and $\tau^+\tau^-$ [75–93]. These anomalies have been interpreted through a twin-peak scenario in the Georgi-Machacek model [87,93,94]. The type-II and flipped next-to-two-Higgs-doublet model scenarios (N2HDM-II and N2HDM-F) provide particularly elegant solutions. They offer additional scalar degrees of freedom that can simultaneously account for the observed $b\bar{b}$ and $\tau^+\tau^-$ anomalies [54,56,76–78,95]. In this study, we perform a comprehensive scan of the N2HDM parameter space to examine the model’s compatibility with the observed 95 GeV excess. Subsequently, an MC simulation and a cut-based jet-level analysis are carried out for the Higgsstrahlung process, $e^+e^- \rightarrow ZS$, in various final states at the 240 GeV CEPC. To further enhance the search sensitivity and measurement precision for the 95 GeV light Higgs boson, we extract features from all final-state particles and employ them to train the ParT and MIParT networks. Finally, we present a detailed discussion of the achievable parameter coverage and measurement precision at the CEPC, both with and without the inclusion of machine learning, and demonstrate the model-independent discovery potential of the CEPC.

The remainder of this paper is organized as follows. In Sec. II, we briefly introduce the N2HDM-F and present the results of the parameter-space scan. Section III describes the MC simulation and the cut-based analysis strategy. The implementation of the ParT and MIParT networks, together with the corresponding performance studies and discussions, is detailed in Sec. IV. Finally, the conclusions are summarized in Sec. V.

II. THE BROKEN-PHASE FLIPPED N2HDM

The next-to-two-Higgs-doublet model (N2HDM) extends the Standard Model scalar sector through the introduction of a second $SU(2)_L$ Higgs doublet Φ_2 and an additional real scalar singlet Φ_S [96–99]. After electroweak symmetry breaking, the model yields three neutral CP -even Higgs states, one of which corresponds to the observed SM-like Higgs boson at 125 GeV. A lighter state can naturally arise around 95 GeV, providing a potential explanation for the mild excesses reported in several experiments. In the following, we identify the light CP -even state H_1 with the putative 95 GeV resonance and denote it by $S \equiv H_1$ whenever convenient.

A. Model setup

The scalar potential of the N2HDM can be written as [99]

$$\begin{aligned}
 V = & m_{11}^2 |\Phi_1|^2 + m_{22}^2 |\Phi_2|^2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \text{h.c.}) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 \\
 & + \lambda_3 (\Phi_1^\dagger \Phi_1) (\Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_2) (\Phi_2^\dagger \Phi_1) + \frac{\lambda_5}{2} [(\Phi_1^\dagger \Phi_2)^2 + \text{h.c.}] \\
 & + \frac{1}{2} m_{\Phi_S}^2 \Phi_S^2 + \frac{\lambda_6}{8} \Phi_S^4 + \frac{\lambda_7}{2} (\Phi_1^\dagger \Phi_1) \Phi_S^2 + \frac{\lambda_8}{2} (\Phi_2^\dagger \Phi_2) \Phi_S^2.
 \end{aligned} \quad (1)$$

The potential respects a softly broken Z_2 symmetry ($\Phi_2 \rightarrow -\Phi_2$) and an exact Z_2' symmetry ($\Phi_S \rightarrow -\Phi_S$), which forbids linear terms in Φ_S and explains the absence of certain couplings.

After electroweak symmetry breaking, the N2HDM particle spectrum contains three CP -even neutral scalars $H_{1,2,3}$ (the masses of the three $H_{1,2,3}$ increase in sequence), one pseudoscalar A , and a pair of charged scalars $H^\pm$. The electroweak sector is in the broken phase when both doublets and the singlet acquire nonzero vacuum expectation values (VEVs). In this scenario, the three CP -even scalar fields mix through a 3×3 rotation matrix, which have the following form [99]:

$$R = \begin{pmatrix} c_{\alpha_1} c_{\alpha_2} & s_{\alpha_1} c_{\alpha_2} & s_{\alpha_2} \\ -(c_{\alpha_1} s_{\alpha_2} s_{\alpha_3} + s_{\alpha_1} c_{\alpha_3}) & c_{\alpha_1} c_{\alpha_3} - s_{\alpha_1} s_{\alpha_2} s_{\alpha_3} & c_{\alpha_2} s_{\alpha_3} \\ -c_{\alpha_1} s_{\alpha_2} c_{\alpha_3} + s_{\alpha_1} s_{\alpha_3} & -(c_{\alpha_1} s_{\alpha_3} + s_{\alpha_1} s_{\alpha_2} c_{\alpha_3}) & c_{\alpha_2} c_{\alpha_3} \end{pmatrix}, \quad (2)$$

where $c_\alpha \equiv \cos \alpha$ and $s_\alpha \equiv \sin \alpha$. The parameters of N2HDM can then be parametrized as

$$\begin{aligned}
 & m_{H_1}, \quad m_{H_2}, \quad m_{H_3}, \quad m_{H_A}, \quad m_{H^\pm}, \\
 & \alpha_1, \quad \alpha_2, \quad \alpha_3, \quad m_{12}^2, \quad v, \quad v_S, \quad \tan \beta,
 \end{aligned} \quad (3)$$

where $m_{H_1}, m_{H_2}, m_{H_3}, m_{H_A}$, and $m_{H^\pm}$ denote the masses of H_1, H_2, H_3, A , and $H^\pm$, respectively; α_1, α_2 , and α_3 are the mixing angles of the three CP -even neutral scalars; m_{12}^2 is the soft Z_2 -breaking parameter; $v = 246$ GeV and v_S are the VEVs of the electroweak doublet and the singlet field, respectively; and $\tan \beta$ is the ratio of the two doublet VEVs.

The most interesting part of the couplings are the CP -even Higgs boson H_i to the massive gauge bosons, the quarks, and the leptons. The reduced coupling of Higgs boson H_i to the massive gauge bosons $V = W, Z$ ($C_{H_i VV}$) can be given by

$$C_{H_i VV} = \begin{cases} c_{\alpha_2} c_{\beta - \alpha_1}, & i = 1, \\ -c_{\beta - \alpha_1} s_{\alpha_2} s_{\alpha_3} + c_{\alpha_3} s_{\beta - \alpha_1}, & i = 2, \\ -c_{\beta - \alpha_1} s_{\alpha_2} c_{\alpha_3} - s_{\alpha_3} s_{\beta - \alpha_1}, & i = 3, \end{cases} \quad (4)$$

where c_x denotes $\cos(x)$ and s_x denotes $\sin(x)$ for the angles $x = \alpha_1, \alpha_2, \alpha_3$, and $\beta - \alpha_1$.

In the N2HDM-F scenario, the reduced coupling of Higgs boson H_i to up-type quarks and leptons f ($C_{H_i ff}$) can be given by

$$C_{H_i ff} = \begin{cases} s_{\alpha_1} c_{\alpha_2} / s_{\beta}, & i = 1, \\ (c_{\alpha_1} c_{\alpha_3} - s_{\alpha_1} s_{\alpha_2} s_{\alpha_3}) / s_{\beta}, & i = 2, \\ -(c_{\alpha_1} s_{\alpha_3} + s_{\alpha_1} s_{\alpha_2} c_{\alpha_3}) / s_{\beta}, & i = 3. \end{cases} \quad (5)$$

The reduced coupling of Higgs boson H_i to the down-type quark q_d ($C_{H_i q_d q_d}$) can be given by

$$C_{H_i q_d q_d} = \begin{cases} c_{\alpha_1} c_{\alpha_2} / c_{\beta}, & i = 1, \\ -(c_{\alpha_1} s_{\alpha_2} s_{\alpha_3} + c_{\alpha_3} s_{\alpha_1}) / c_{\beta}, & i = 2, \\ -(c_{\alpha_1} s_{\alpha_2} c_{\alpha_3} - s_{\alpha_1} s_{\alpha_3}) / c_{\beta}, & i = 3. \end{cases} \quad (6)$$

In this scenario, the couplings of H_i to b quarks and to τ leptons exhibit distinct dependences on the model parameters, which can naturally account for the observed excesses in the $b\bar{b}$ and $\tau^+ \tau^-$ channels. These coupling patterns illustrate how the N2HDM-F permits independent control of quark and lepton Yukawa couplings, thereby providing the flexibility required to simultaneously accommodate both $b\bar{b}$ and $\tau^+ \tau^-$ excesses.

If H_1 is assumed to be the candidate for the possible experimental excesses and H_2 is assumed to be the SM-like Higgs boson with a mass of 125.09 GeV, one can reparametrize the parameters in terms of effective couplings and mixing matrix elements,

$$\begin{aligned}
 & m_{H_1}, \quad m_{H_2}, \quad m_{H_3}, \quad m_{H_A}, \quad m_{H^\pm}, \quad m_{12}^2, \quad v_S \\
 & C_{H_2 VV}^2, \quad C_{H_2 i\bar{i}}, \quad \text{sign}(R_{23}), \quad R_{13}, \quad \tan \beta,
 \end{aligned} \quad (7)$$

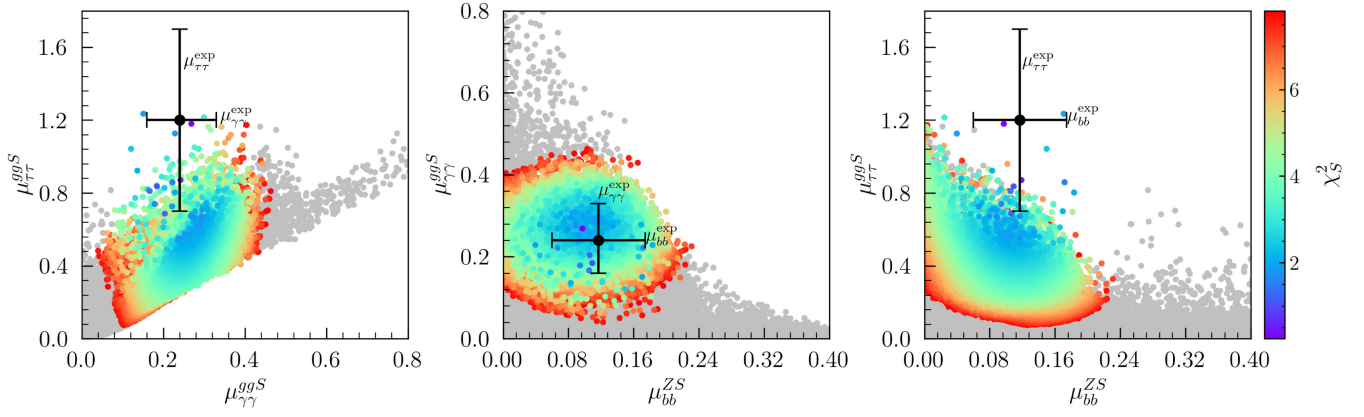


FIG. 1. Surviving samples under the above constraints in the $\mu_{\tau\tau}^{ggS}$ versus $\mu_{\gamma\gamma}^{ggS}$ plane (left panel), the $\mu_{\gamma\gamma}^{ggS}$ versus μ_{bb}^{ZS} plane (middle panel), and $\mu_{\tau\tau}^{ggS}$ versus μ_{bb}^{ZS} plane (right panel) with colors indicating the fit result of χ_S^2 and the gray samples having $\chi_S^2 > 7.82$.

where $C_{H_2VV}^2$ and $C_{H_2\bar{t}t}^2$ are the effective couplings of H_2 to massive gauge bosons and top-quark, $R_{13,23}$ are the mixing matrix elements between $H_{1,2}$ and the singlet field.

B. Parameter scan and constraints

We performed a random scan over the parameter space of the N2HDM-F using the package `ScannerS_v2.0.0` [100,101]. With the mass of m_{H_2} fixed at 125.09 GeV, the remaining 11 parameters were varied within the following ranges:

$$\begin{aligned}
 95 < m_{H_1} < 96 \text{ GeV}, \quad \text{sign}(R_{23}) = \pm 1, \\
 300 < m_{H_{3,A}} < 1500 \text{ GeV}, \quad 580 < m_{H^\pm} < 1500 \text{ GeV}, \\
 0.8 < \tan\beta < 10, \quad -1 < R_{13} < 1, \\
 0.7 < C_{H_2VV}^2 < 1.0, \quad 0.7 < C_{H_2\bar{t}t}^2 < 1.2, \\
 10^{-3} < m_{12}^2 < 5 \times 10^5 \text{ GeV}^2, \quad 1 < v_S < 3000 \text{ GeV}.
 \end{aligned} \tag{8}$$

The following constraints are considered:

- (i) Constraints from perturbative unitarity: The constraints require that the eigenvalues $\mathcal{M}_{2 \rightarrow 2}^i$ of the $2 \rightarrow 2$ scattering matrix $\mathcal{M}_{2 \rightarrow 2}$ satisfy

$$|\mathcal{M}_{2 \rightarrow 2}^i| < 8\pi, \tag{9}$$

at tree level [102].

- (ii) Constraints from vacuum stability: The package `EVADe` [103,104] is used to test the stability at tree level.
- (iii) Constraints from B physics and electroweak precision observables: The global fit result of Ref. [105] on oblique parameters and flavor constraints at 2σ level are used.
- (iv) Constraints from Higgs searches and SM-like Higgs measurements: The `HiggsBounds` – 5.10.0

package [106] and `HiggsSignal` – 2.6.0 package [107] are used to employ the direct searches for additional Higgs and SM-Higgs measurements constraints.

C. Features of the N2HDM-F scenario

Figure 1 shows the surviving samples under the above constraints in the $\mu_{\tau\tau}^{ggS}$ versus $\mu_{\gamma\gamma}^{ggS}$ plane (left panel), the $\mu_{\gamma\gamma}^{ggS}$ versus μ_{bb}^{ZS} plane (middle panel), and $\mu_{\tau\tau}^{ggS}$ versus μ_{bb}^{ZS} plane (right panel), respectively. The colors indicate the fit results of χ_S^2 ,¹ defined as

$$\chi_S^2 = \frac{(\mu_{\tau\tau}^{ggS} - \mu_{\tau\tau}^{\text{exp}})^2}{(\Delta\mu_{\tau\tau}^{\text{exp}})^2} + \frac{(\mu_{\gamma\gamma}^{ggS} - \mu_{\gamma\gamma}^{\text{exp}})^2}{(\Delta\mu_{\gamma\gamma}^{\text{exp}})^2} + \frac{(\mu_{bb}^{ZS} - \mu_{bb}^{\text{exp}})^2}{(\Delta\mu_{bb}^{\text{exp}})^2}, \tag{10}$$

where μ_{xx}^{exp} and $\Delta\mu_{xx}^{\text{exp}}$ are the central values and their corresponding 1σ uncertainties for the experimental excesses in the $\tau^+\tau^-$, $b\bar{b}$, and $\gamma\gamma$ channels, respectively. These are parametrized as $\mu_{\tau\tau}^{\text{exp}} = 1.2_{-0.5}^{+0.5}$ [12], $\mu_{bb}^{\text{exp}} = 0.117 \pm 0.057$ [10], and $\mu_{\gamma\gamma}^{\text{exp}} = 0.24_{-0.08}^{+0.09}$ [13,14], and are marked in the Fig. 1. The μ_{xx}^{ggS} and μ_{xx}^{ZS} are the signal strength in the gluon fusion channel ($gg \rightarrow S \rightarrow xx$) and Higgsstrahlung channel ($ee \rightarrow ZS(\rightarrow xx)$) and can be parametrized as

$$\mu_{xx}^{ggS} = \frac{\sigma(gg \rightarrow S) \times \text{Br}(S \rightarrow xx)}{\sigma_{\text{SM}}(gg \rightarrow h_{95}) \times \text{Br}_{\text{SM}}(h_{95} \rightarrow xx)} \tag{11}$$

and

$$\mu_{xx}^{ZS} = \frac{\sigma(e^+e^- \rightarrow ZS) \times \text{Br}(S \rightarrow xx)}{\sigma_{\text{SM}}(e^+e^- \rightarrow Zh_{95}) \times \text{Br}_{\text{SM}}(h_{95} \rightarrow xx)}, \tag{12}$$

¹In constructing χ_S^2 , we treat the three excesses as independent Gaussian measurements and neglect possible correlations among the experimental systematics. This is sufficient for our purpose of identifying representative benchmark scenarios and should not be interpreted as a rigorous global statistical combination.

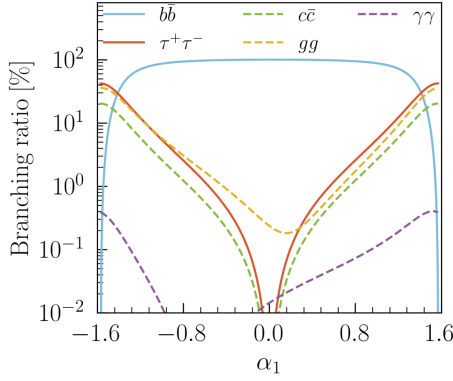


FIG. 2. The branching ratios of the light Higgs boson S to $b\bar{b}$, $c\bar{c}$, $\gamma\gamma$, $\tau^+\tau^-$, and gg in the N2HDM-F as a function of α_1 , for the benchmark points with $\cos\alpha_2 = \sqrt{2}/2$ and $\tan\beta = 2$.

where σ and Br denote the production cross section and branching ratio in the beyond the Standard Model (BSM) scenario, respectively, while σ_{SM} and Br_{SM} represent the corresponding SM values. All three excesses can be elegantly explained within the broken phase N2HDM-F scenario. The best-fit point has a minimum χ^2_S value of 0.24.

Figure 2 shows the branching ratios of the light Higgs boson S to $b\bar{b}$, $c\bar{c}$, $\gamma\gamma$, $\tau^+\tau^-$, and gg as a function of α_1 in the N2HDM-F, for benchmark points with $\cos\alpha_2 = \sqrt{2}/2$ and $\tan\beta = 2$. These branching ratios are computed using the package N2HDECY [108]. It can be seen that S decays predominantly to $b\bar{b}$ over most of the α_1 range, except near $\alpha_1 \approx \pm\pi/2$ (where $\cos\alpha_1 \approx 0$). In this region, the dominant decay channels become $\tau^+\tau^-$, $c\bar{c}$, and gg . However, the $c\bar{c}$ and gg channels are associated with a larger SM background from Z boson hadronic decays [109]. Considering the channels where experimental signal excesses have been observed, we focus on the $b\bar{b}$ and $\tau^+\tau^-$ channels in the subsequent analysis. The combination of these two channels ensures good coverage across the entire α_1 parameter space.

III. MONTE CARLO SIMULATION

A. Signal and background modeling

We focus on the Higgsstrahlung process $e^+e^- \rightarrow Z(\rightarrow \mu^+\mu^-)S(\rightarrow \tau^+\tau^-(b\bar{b}))$ at the CEPC with a center-of-mass

energy of 240 GeV. In this signal process, the hypothetical scalar S is produced in association with a Z boson, where S subsequently decays into a pair of τ leptons or b quarks, and the Z boson decays into a muon pair. Although the dominant decay channel of the Z boson is hadronic and invisible [109], we instead select the dimuon final state to utilize the recoil mass technique for reconstructing the mass of S . The recoil mass M_{recoil} is defined as [110]

$$M_{\text{recoil}} \equiv \sqrt{s + M_{\mu^+\mu^-}^2 - 2\sqrt{s}(E_{\mu^+} + E_{\mu^-})}, \quad (13)$$

where $\sqrt{s}$ is the collision energy, $M_{\mu^+\mu^-}^2$ is the square of invariant mass of dimuon, and E_{μ} is the energy of muon. Other useful observables can be found in Ref. [111]. One can also choose Z boson decays into a pair of electrons to calculate the recoil mass, and this situation will be discussed later in the paper.

For the $\tau^+\tau^-$ decay channel, the dominant irreducible SM backgrounds originate from the process $e^+e^- \rightarrow \mu^+\mu^-\tau^+\tau^-$, in which the $\mu^+\mu^-$ and $\tau^+\tau^-$ pairs can decay from a Z boson or a virtual photon. Additionally, the $\tau^+\tau^-$ pair may also come from the decay of the SM-like Higgs boson. The leading Feynman diagrams for the signal and the irreducible backgrounds in the $\tau^+\tau^-$ channel are shown in Fig. 3, where panel (a) corresponds to the signal, and panels (b), (c), and (d) represent the irreducible backgrounds. We also consider the reducible SM background from $e^+e^- \rightarrow \mu^+\mu^-xx$, where x denotes a final-state particle from a light-flavor quark, a b quark, an electron, or a muon, which could be misidentified as a tau lepton in the detector.

In the case of the $b\bar{b}$ decay channel, the background sources are analogous to those in the $\tau^+\tau^-$ channel, with the roles of $b\bar{b}$ and $\tau^+\tau^-$ are interchanged. Therefore, we adopt a common background process for both channels: $e^+e^- \rightarrow \mu^+\mu^-xx$, where $x = u, d, c, s, b, e, \mu, \tau$. These background processes comprise two components: a resonant part, in which the muon pair comes from a Z boson, and a nonresonant part, where the muon pair originates from a virtual photon.

Figure 4 shows the surviving samples in the $\tan\beta$ versus α_1 planes with the colors indicating the square of the

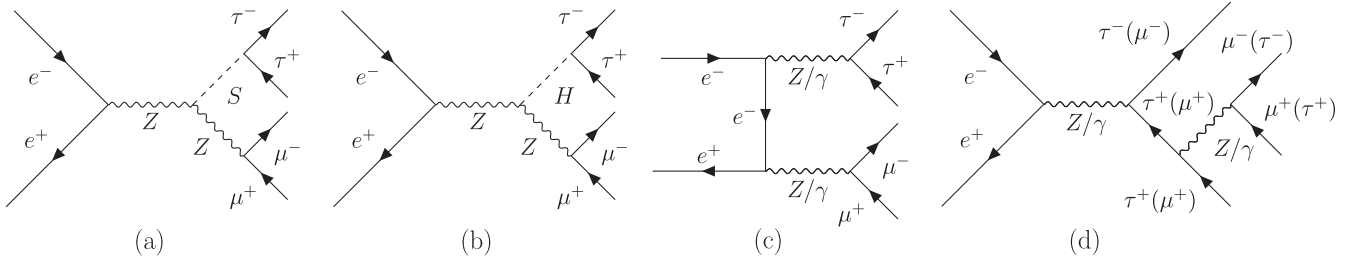


FIG. 3. Representative Feynman diagrams of the signal and irreducible backgrounds in the $\tau^+\tau^-$ decay channel, where panel (a) corresponds to the signal, and panels (b)–(d) represent the irreducible backgrounds. Diagrams in panels (c) and (d) also imply those related by fermion-antifermion exchange, $\mu \leftrightarrow \tau$ interchange, and $Z \leftrightarrow \gamma$ exchange.

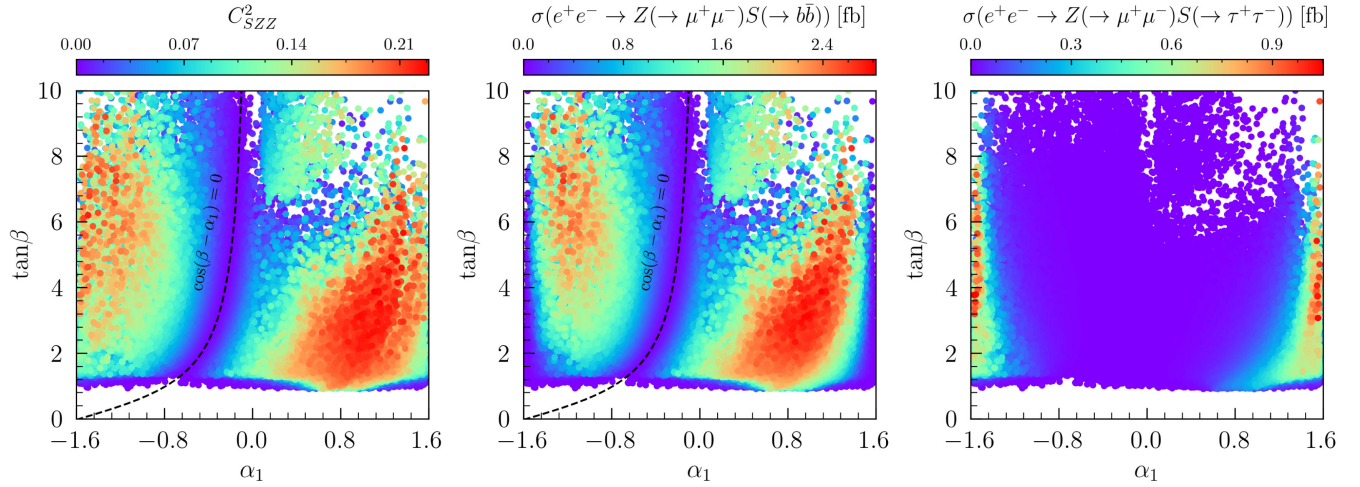


FIG. 4. Surviving samples in the $\tan\beta$ versus α_1 plane with the colors indicating the square of the reduced coupling of S to Z boson (left panel), the cross section of the $b\bar{b}$ channel (middle panel), and the cross section of the $\tau^+\tau^-$ channel (right panel).

reduced coupling of S to the Z boson C_{SZ}^2 (left panel), the cross section of the $b\bar{b}$ channel (middle panel), and the cross section of the $\tau^+\tau^-$ channel (right panel). The cross section for the $b\bar{b}$ and $\tau^+\tau^-$ channels is computed as

$$\sigma = \sigma_{\text{SM}}(e^+e^- \rightarrow ZS)|_{m_h=95.5 \text{ GeV}} \times C_{SZ}^2 \times \text{Br}(Z \rightarrow \mu^+\mu^-) \cdot \text{Br}(S \rightarrow \tau^+\tau^-(b\bar{b})), \quad (14)$$

where $\sigma_{\text{SM}}(e^+e^- \rightarrow ZS)|_{m_h=95.5 \text{ GeV}}$ is the SM cross section for this process with a Higgs mass of 95.5 GeV, calculated at leading order using the `MadGraph5_aMC@NLO_v3.6.2` package [112,113]. The branching ratio of $\text{Br}(Z \rightarrow \mu^+\mu^-)$ takes the SM value of 3.3662% as given by the PDG [109]. The C_{SZ}^2 and $\text{Br}(S \rightarrow \tau^+\tau^-(b\bar{b}))$ provided by the N2HDM-F.

As shown in the left panel, the squared reduced coupling C_{SZ}^2 can reach up to 0.23 for the samples near $\alpha_1 \approx 0.8$. In contrast, C_{SZ}^2 approaches zero for samples with $\cos(\beta - \alpha_1) \approx 0$, where the light Higgs boson S couples weakly to massive gauge bosons and could escape detection at the CEPC. The middle and right panels indicate that the cross section for the $b\bar{b}$ channel can reach 2.8 fb for the samples with $\alpha_1 \approx 0.8$, while that for the $\tau^+\tau^-$ channel reaches 1.1 fb near $\alpha_1 \approx \pm\pi/2$. As anticipated, except for the samples near $\cos(\beta - \alpha_1) \approx 0$, where the cross section of the $b\bar{b}$ channel is suppressed, the cross section of the $\tau^+\tau^-$ channel becomes comparatively larger, as can be inferred from a combined interpretation of all three panels. In the following MC simulations, the cross sections for the $\tau^+\tau^-$ and $b\bar{b}$ decay channels are assumed to be 1/5 of their SM values as a benchmark point.

B. Event generation and selection in the $\tau^+\tau^-$ channel

The `MadGraph5_aMC@NLO_v3.6.2` package [112,113] is used to perform the MC simulation. During the simulation,

jet clustering is performed using the `ee $k_T$` algorithm through `FastJet` [114], and the tau jet tagging efficiency adopted the lower limit 80% as given in Ref. [115]. The particle decays are handled by `PYTHIA_v8.2` [116] through the `MG5a-MC_PY8_interface`. The CEPC detector response is simulated using the CEPC fourth-card with `Delphes_v3.5.0` [117,118]. More than one million events of signal and background are generated through the MC simulation.

In the event selection, two muons in the final state are required to calculate the recoil mass M_{recoil} to reconstruct the mass of S . We apply the following basic cut to the events:

$$N(\mu) \geq 2, |\eta(\mu)| < 2.5, \quad p_T(\mu) > 10 \text{ GeV}, \\ \Delta R(\mu_1, \mu_2) > 0.4, \quad 66 < M_{\mu\mu} < 116 \text{ GeV}, \quad (15)$$

where η , ϕ , and p_T denote the pseudorapidity, azimuthal angle, and transverse momentum, respectively; $\Delta R(\mu_1, \mu_2) = \sqrt{(\eta_{\mu_1} - \eta_{\mu_2})^2 + (\phi_{\mu_1} - \phi_{\mu_2})^2}$ gives the angular separation between the two muons; $M_{\mu\mu} = \sqrt{(E_{\mu_1} + E_{\mu_2})^2 - (\vec{p}_{\mu_1} + \vec{p}_{\mu_2})^2}$ is the invariant mass of the muon pair.

In the signal process, the muon pair originates from the Z boson decay, and its invariant mass distribution is expected to peak near the Z boson mass. In contrast, a significant fraction of the background events involve muon pairs not produced from a Z boson resonance and thus, do not exhibit such a peak in their invariant mass distribution. To suppress this nonresonant background component and ensure robust signal retention, we require that $66 < M_{\mu\mu} < 116 \text{ GeV}$.

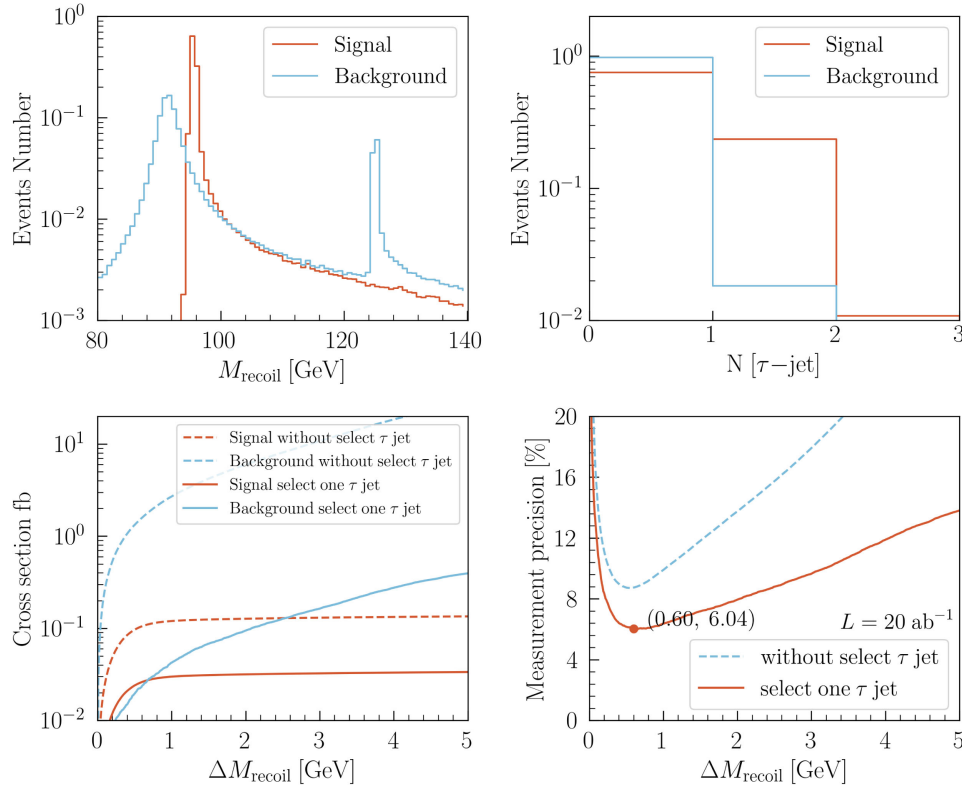


FIG. 5. Normalized event distributions for the signal and background in the recoil mass M_{recoil} (upper left panel) and the τ -jet number $N[\tau\text{-jet}]$ (upper right panel), along with the signal and background cross sections under different selections on these variables (lower left panel), as well as the corresponding measurement precision of the signal strength after applying these selections (lower right panel) in the $\tau^+\tau^-$ channel.

The normalized event distributions of the recoil mass (M_{recoil}) and the number of τ -jet ($N[\tau\text{-jet}]$) for signal and background after the Basic cut are shown in the upper left and upper right panels of Fig. 5, respectively. The upper left panel shows that the signal M_{recoil} distribution peaks sharply around 95.5 GeV, consistent with the mass of the light Higgs boson S . In contrast, the background distribution exhibits peaks near 91 GeV and 125 GeV, corresponding to the masses of the Z boson and the SM-like Higgs boson, respectively, with the peak at 91 GeV being considerably broader. The upper right panel indicates that only about 2% of background events contain at least one τ jet, whereas for the signal, this fraction reaches approximately 25%. Therefore, applying selections on M_{recoil} and the number of τ jets can effectively suppress the background.

A fine-grained scan of the ΔM_{recoil} window from 0 to 5 GeV is performed. The signal region is defined as $|M_{\text{recoil}} - 95.5 \text{ GeV}| < \Delta M_{\text{recoil}}$, and the cross sections of signal and background are calculated based on the number of surviving events in each window. The lower-left panel of Fig. 5 shows the signal and background cross sections after applying different selections on the recoil-mass window ΔM_{recoil} and on the number of τ jets ($N[\tau\text{-jet}]$). The

corresponding statistical precision of the signal strength measurement is shown in the lower right panel, calculated using [119],

$$P = \frac{\sigma(\mu)}{\mu} = \frac{\sqrt{S+B}}{S}, \quad (16)$$

where S and B denote the number of expected signal and background events for different ΔM_{recoil} windows, and both are calculated by their cross sections, assuming an integrated luminosity of 20 ab^{-1} at the CEPC.

As shown in the lower left panel of Fig. 5, the background cross section exceeds that of the signal by more than 1 order of magnitude if no selection is applied on the number of τ jets, regardless of the M_{recoil} cut. However, by selecting events with at least one τ jet and applying an appropriate M_{recoil} cut, the signal and background cross sections can be brought to the same order of magnitude. Furthermore, for any given M_{recoil} cut, the measurement precision is significantly improved when requiring a τ jet in the final state. The optimal precision of 6.04% is achieved for events containing a τ jet and satisfying $\Delta M_{\text{recoil}} \leq 0.60 \text{ GeV}$. The choice reflects a balance between background suppression and signal retention: while a narrower window reduces

TABLE I. Cut flow of signal and background in the $\tau^+\tau^-$ channel at 240 GeV CEPC with $L = 20 \text{ ab}^{-1}$.

Cuts	Cross section (fb)		Precision (%)
	Signal	Background	
Initial	0.218	84.250	29.81
Basic	0.159	53.373	32.56
τ -jet	0.039	1.000	18.41
M_{recoil}	0.026	0.024	6.04

background more effectively, it also excludes a significant fraction of signal events, thereby degrading the overall precision. Based on this, we apply the following τ -jet cut and M_{recoil} cut that require

$$N[\tau - \text{jet}] \geq 1 \quad (17)$$

and

$$\Delta M_{\text{recoil}} = 0.60 \text{ GeV}. \quad (18)$$

Table I shows the cut flow for the $\tau^+\tau^-$ decay channel of the signal and background at the 240 GeV CEPC with an integrated luminosity of $L = 20 \text{ ab}^{-1}$. After the τ -jet cut,

the background is suppressed by approximately 2 orders of magnitude while about 25% of the signal is retained, improving the signal strength measurement precision from 32.56% to 18.41%. A subsequent M_{recoil} cut further suppresses the background by over an order of magnitude and preserves about 67% of the signal, leading to a final precision of 6.04%.

After applying the basic cut, Fig. 6 shows the normalized distributions of the transverse momenta, pseudorapidities, and azimuthal angles of the two muons and the leading jet, as well as the angular separations ΔR between the two muons and between the leading muon and the leading jet. As shown in this figure, the significant overlap between the signal and background means that further cuts on these kinematic observables do not lead to a significant improvement in signal significance or measurement precision.

C. Event generation and selection in the $b\bar{b}$ channel

The event generation for the $b\bar{b}$ channel follows the same procedure as that for the $\tau^+\tau^-$ channel, with the b -jet tagging efficiency also set to 80%. In the event selection, the basic cut is also the same as that in the $\tau^+\tau^-$ channel. The only difference is that selecting a b jet in the final state can reduce the reducible background. Figure 7 shows the cross sections for signal and background processes (left

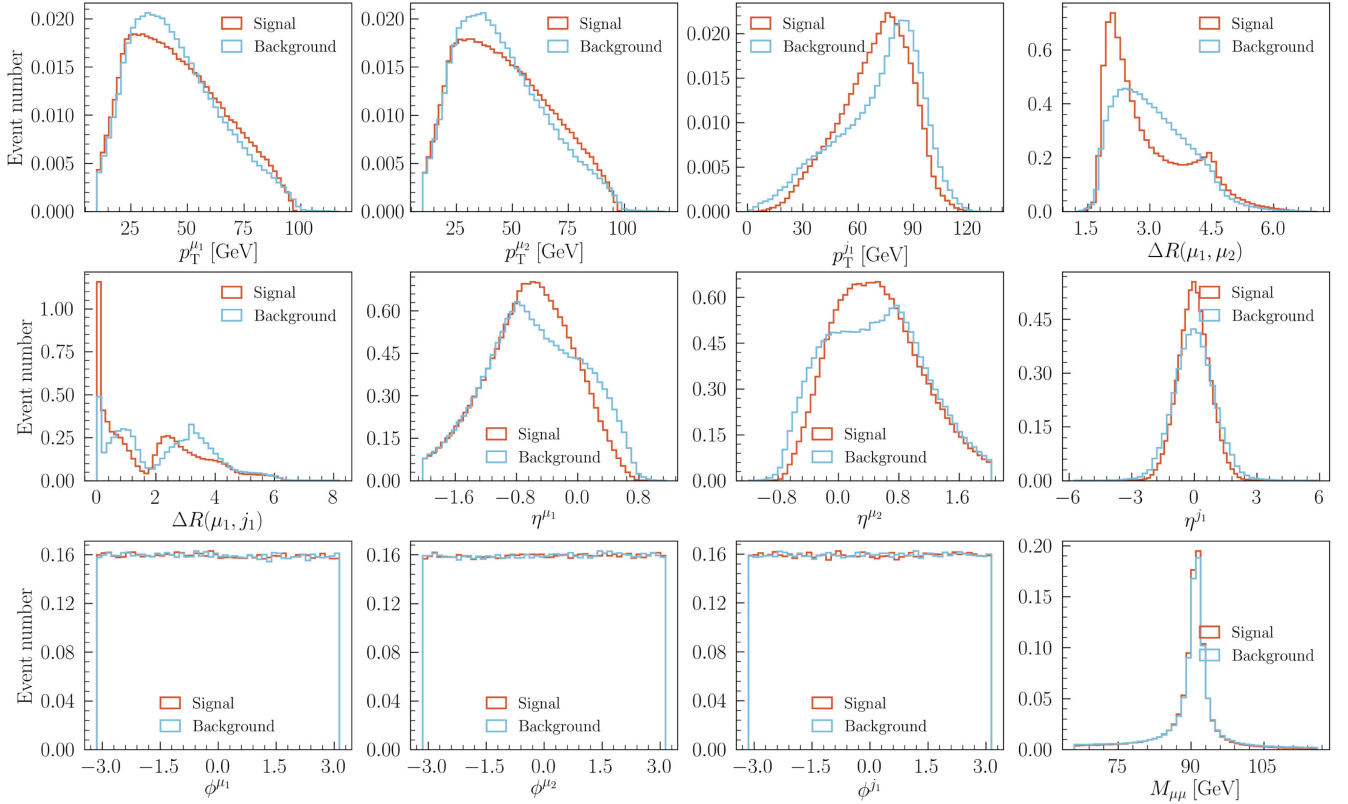


FIG. 6. The normalized distributions of the transverse momenta, pseudorapidities, and azimuthal angles of the two muons and the leading jet, as well as the angular separations ΔR between the two muons and between the leading muon and the leading jet, after applying the basic cut.

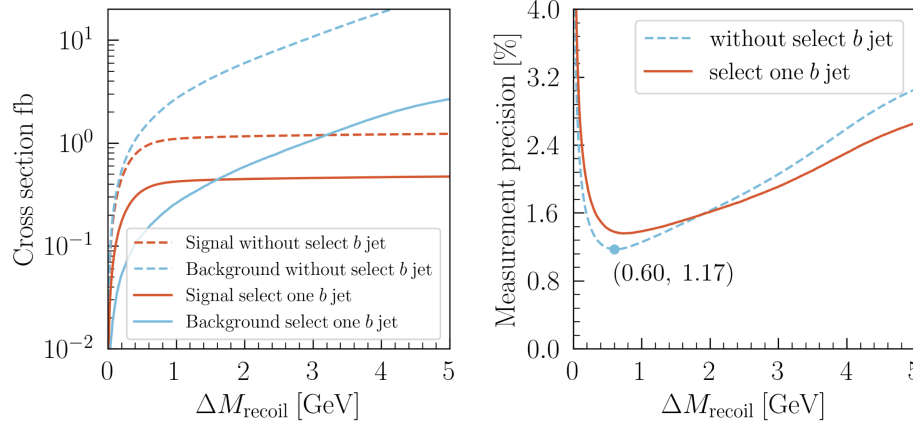


FIG. 7. The cross sections of the signal and background (left panel), and the signal strength measurement precision (right panel) under different selections on the M_{recoil} and $N[b \text{ jet}]$ in the $b\bar{b}$ channel.

panel) and the corresponding precision on the signal strength (right panel) after applying different selections on the recoil-mass window ΔM_{recoil} and on the number of b -jets ($N[b\text{-jet}]$). Unlike the $\tau^+\tau^-$ decay channel, the signal and background cross sections become comparable by applying only the M_{recoil} selection. Selecting events that contain a b jet suppresses the background by approximately 1 order of magnitude, but also reduces the signal efficiency. Consequently, this combined selection fails to achieve the maximum measurement precision attainable with the M_{recoil} selection alone. The signal strength can be measured with a precision of 1.17% by selecting events with $\Delta M_{\text{recoil}} = 0.60$ GeV and without requiring a b jet in the final state. This choice of window is optimal to maximize the measurement precision because a narrower window would discard too much signal, while a wider one would introduce excessive background. Then the M_{recoil} cut are performed that require the event be satisfied for

$$|M_{\text{recoil}} - 95.5| \leq 0.60 \text{ GeV}. \quad (19)$$

The cut flow for the $b\bar{b}$ decay channel of the signal and background at the 240 GeV CEPC with an integrated luminosity of $L = 20 \text{ ab}^{-1}$ is shown in Table II. It can be seen that the M_{recoil} cut suppresses the background more than 1 order of magnitude while retaining about 69% of the signal. The corresponding measurement precision is improved from 3.76% to 1.17%. The $b\bar{b}$ channel is also

TABLE II. Cut flow of signal and background in the $b\bar{b}$ channel at 240 GeV CEPC with $L = 20 \text{ ab}^{-1}$.

Cuts	Cross section (fb)		Precision (%)
	Signal	Background	
Initial	2.099	84.250	3.13
Basic	1.391	53.373	3.76
M_{recoil}	0.963	1.585	1.17

plagued by a limitation similar to that in the $\tau^+\tau^-$ channel. Despite successive cuts on jet-level kinematic observables, the measurement precision does not improve significantly.

In fact, through a systematic exploration of additional selection criteria for $\tau^+\tau^-$ and $b\bar{b}$ channels, we find that imposing additional cuts on fundamental jet- or lepton-level observables does not translate into a noticeable improvement in precision over the presented results. This provides strong evidence that the current cut-based analysis is close to optimally configured. The further improvement of measurement precision using the cut-based method in the $\tau^+\tau^-$ and ($b\bar{b}$) decay channels is primarily constrained by two technical limitations:

- (i) First, despite a general τ -tagging efficiency of up to 80%, only about 25% of hadronically decaying τ leptons are successfully identified as τ jets in the $\tau^+\tau^-$ channel, leading to significant signal loss. While the situation in the $b\bar{b}$ channel is better, it similarly suffers from a substantial loss of signal events.
- (ii) Second, event selection that depends on reconstructed jet information inevitably discards a considerable amount of information from the final-state particles, some of which may be critical. Furthermore, the high multiplicity and diversity of final-state particles make it challenging for the cut-based method to fully exploit all available information and capture the underlying correlations among them.

To enhance the search sensitivity for the 95 GeV light Higgs boson and improve the measurement precision, we employ deep neural networks (DNNs), specifically the ParT [36] and MIParT [37] architectures. Both architectures are based on the transformer model, which excels at capturing complex correlations within the data.

IV. APPLICATION OF TRANSFORMER NETWORKS

ParT [36] incorporates pairwise particle interactions into the attention mechanism, achieving superior particle tagging

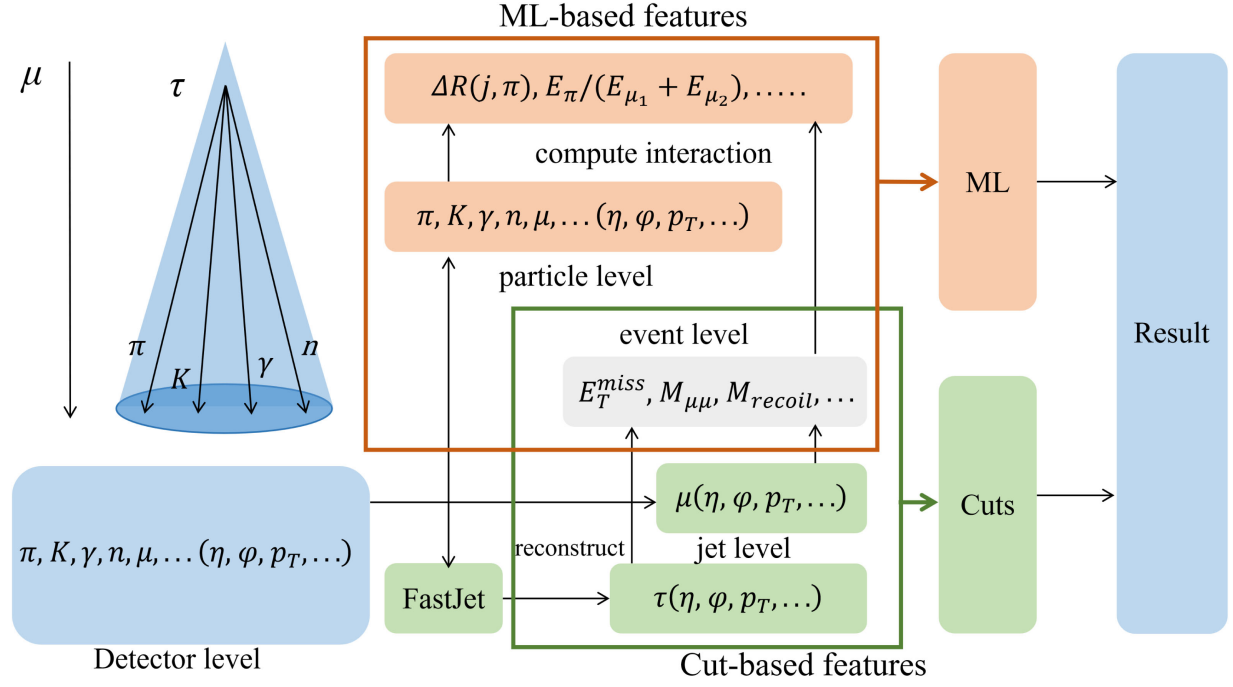


FIG. 8. Schematic comparison of the available features used in jet-level cut-based and particle-level ML analysis.

performance compared to standard transformer architectures and previous ML models. Building upon ParT, MIParT [37] introduces a more-interaction attention mechanism that increases the dimensionality of particle interaction embeddings. This enhancement not only significantly reduces the model size but also leads to further performance gains. These methods can exploit a richer and more complex set of information from the event data compared to cut-based analysis.

For the cut-based analysis, the available information comprises jet-level observables (e.g., jet four-momenta), lepton-level observables (e.g., muon four-momenta), and event-level variables (e.g., the invariant mass and recoil mass of the dimuon system). ParT and MIParT can both train directly on final-state particle-level features.

In our ML-based analysis, events are first preselected according to the basic cut of Eq. (15), and the ParT and MIParT networks are subsequently trained on these events using the following event-level features:

- (i) The recoil mass of two muons: M_{recoil} .
- (ii) The invariant mass of two muons: $M_{\mu\mu}$.
- (iii) The average azimuthal angle of two muons²: $\phi = (\phi_{\mu_1} + \phi_{\mu_2})/2$.

²The reason why we use averaged quantities such as $\phi = (\phi_1 + \phi_2)/2$, is that the architecture of the original ParT and MIParT networks needs a reference direction. We therefore define the reference direction as the average of the two muon directions. We have tested that the choice of reference direction has almost no impact on the network's performance.

- (iv) The average pseudorapidity of two muons: $\eta = (\eta_{\mu_1} + \eta_{\mu_2})/2$.

The following features at the particle-level are also used:

- (i) The transverse momentum of the final state particles: P_T^{part} .
- (ii) The energy of the final state particles: E^{part} .
- (iii) The renormalized transverse momentum of the final state particles: $R_{P_T^{\text{part}}} = P_T^{\text{part}} / (P_T^{\mu_1} + P_T^{\mu_2})$.
- (iv) The renormalized energy of the final state particles: $R_{E^{\text{part}}} = E^{\text{part}} / (E^{\mu_1} + E^{\mu_2})$.
- (v) The difference between the azimuthal angle of the final state particles and the average azimuthal angle of two muons ϕ : $\Delta\phi^{\text{part}} = \phi^{\text{part}} - \phi$.
- (vi) The difference between the pseudorapidity angle of the final state particles and the average pseudorapidity of two muons η : $\Delta\eta^{\text{part}} = \eta^{\text{part}} - \eta$.
- (vii) The angular separations between the final state particles to the event η and ϕ : $\Delta R^{\text{part}} = \sqrt{(\eta^{\text{part}} - \eta)^2 + (\phi^{\text{part}} - \phi)^2}$.
- (viii) The charge of the final state particles: $C^{\text{part}} = \pm 1$ for charged particle and $C^{\text{part}} = 0$ for neutral particle.
- (ix) The particle identification of the final-state particles: charged hadron, neutral hadron, photon, muon, or electron.

A schematic diagram comparing the available features used in the jet-level cut-based analysis versus our particle-level ML-based analysis is presented in Fig. 8. Compared to the cut-based analysis, which only utilizes information from jets, muons, and event-level variables, our ML-based analysis

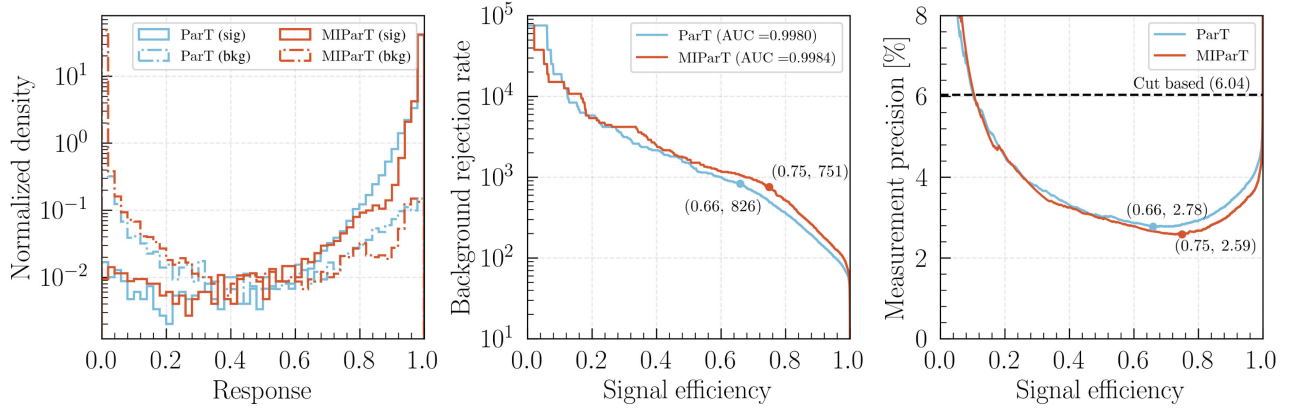


FIG. 9. The normalized distributions of the classifier response scores for the signal and background events in the testing dataset (left panel), the background rejection rate as a function of signal efficiency (middle panel), and the signal strength measurement precision versus signal efficiency (right panel) in the $\tau^+\tau^-$ channel.

incorporates a wider set of features at the particle level. These features are difficult to leverage effectively with cut-based methods but can be fully exploited by models such as ParT and MIParT, which possess strong information processing capabilities.

During data preprocessing, the following features were standardized after applying a logarithmic transformation: p_T^{part} , E^{part} , $R_{p_T^{\text{part}}}$, and $R_{E^{\text{part}}}$. To reduce computational cost, the maximum number of particles per event was set to 50. We constructed a balanced dataset of over 300,000 events (with 150,000 signal and 150,000 background events) for training and evaluating the ParT and MIParT models. The dataset was split into a training set (90% of the data) and a validation set (10%). Other architectural hyperparameters, such as the embedding dimension, number of transformer blocks, number of attention heads, and hidden dimension in the feedforward network, followed the default configurations used in ParT [36] and MIParT [37]. The models were trained with a batch size of 256 and an initial learning rate of 0.001. After 20 epochs, the model achieving the best validation performance was selected, and its final performance was evaluated on a separate, held-out test set consisting of 100,000 events.

A. Performance in the $\tau^+\tau^-$ channel

Figure 9 presents the normalized distributions of the classifier response scores for signal and background events in the testing dataset (left panel), the background rejection rate as a function of signal efficiency (middle panel), and the signal strength measurement precision versus signal efficiency (right panel) in the $\tau^+\tau^-$ channel. The signal efficiency corresponds to the fraction of true signal events correctly identified, while the background rejection rate indicates the fraction of background events successfully

rejected.³ The AUC of ParT and MIParT can reach 0.9980 and 0.9984, respectively. Physically, the signal efficiency represents the fraction of true signals correctly identified, while the background rejection rate indicates the fraction of background events successfully rejected.

For the benchmark point, the $\tau^+\tau^-$ channel signal strength measurement precision reaches 2.78% using ParT at a signal efficiency of 0.66, which corresponds to a background rejection rate of 826, as shown in the right panels of Fig. 9. A marginal improvement is observed with MIParT, which delivers a precision of 2.59% at a higher signal efficiency of 0.75 and a higher background rejection rate of 751, indicating its superior performance in this scenario. Relative to the 6.04% measurement precision obtained with the cut-based analysis, the precision is improved by factors of 2.2 and 2.3 when using ParT and MIParT, respectively. This improvement is consistent with the findings of Ref. [120], which demonstrated that a holistic event-level approach combined with advanced color singlet identification (ACSI) can enhance the expected precision by factors of two to six in processes where the SM-like Higgs boson is produced in association with neutrino or quark pairs and decays into light-flavor quarks or gluons.

A detailed comparison between the cut-based and ML-based analysis for the $\tau^+\tau^-$ channel is presented in Table III. The signal efficiency and background rejection

³The signal efficiency is equivalent to the true positive rate (TPR), $\epsilon_{\text{sig}} = N_{\text{TP}}/(N_{\text{TP}} + N_{\text{FN}})$, and the background rejection corresponds to $1/\text{FPR}$, where $\text{FPR} = N_{\text{FP}}/(N_{\text{FP}} + N_{\text{TN}})$. Here, N_{TP} , N_{FP} , N_{FN} , and N_{TN} denote the number of true positive, false positive, false negative, and true negative events, respectively. A background rejection rate of 200 at signal efficiency is 50% means that for every 200 true background events, approximately one is misclassified as signal, while half of the true signal events are correctly identified.

TABLE III. Comparison of the cut-based and the ML-based analysis in the $\tau^+\tau^-$ decay channel.

Method	Signal efficiency	Background rejection rate	Precision (%)
Cut-based	0.16	2224	6.04
ParT	0.66	826	2.78
MIParT	0.75	751	2.59

rate here are evaluated relative to the events after the basic cut. The substantial improvement in measurement precision from the ML-based analysis arises from a markedly higher signal efficiency, achieved with only a modest compromise in background rejection rate. It is worth noting that, compared to the improvements offered by jet-level ML algorithms such as extreme gradient boosting, gradient boosting decision trees, and deep neural networks in our previous work [19], the enhancement from particle-level transformers here is even more pronounced. This enhancement can be attributed to the efficient use of particle-level features and the superior ability of the ParT and MIParT architectures to extract underlying correlations among them.

B. Performance in the $b\bar{b}$ channel

The same ML methodology is applied to the $b\bar{b}$ decay channel. Figure 10 shows the normalized distributions of classifier response scores for signal and background events in the testing dataset (left panel), the background rejection rate as a function of signal efficiency (middle panel), and the signal strength measurement precision versus signal efficiency (right panel) in the $b\bar{b}$ channel. The AUC values for MIParT and ParT reach 0.9914 and 0.9909, respectively, with MIParT slightly outperforming ParT.

As shown in the middle and right panels of Fig. 10, the ParT model in the $b\bar{b}$ channel achieves a signal strength

TABLE IV. Comparison of the cut-based and the ML-based analysis in the $b\bar{b}$ decay channel.

Method	Signal efficiency	Background rejection rate	Precision (%)
Cut-based	0.69	33.7	1.17
ParT	0.71	126.3	0.85
MIParT	0.78	89.9	0.84

measurement precision of 0.85% at a signal efficiency of 0.71, while maintaining a background rejection of 126.3. In comparison, MIParT achieves a precision of 0.84% at a slightly higher signal efficiency of 0.78, with a background rejection rate of 89.9. Although MIParT exhibits a marginally higher AUC, the difference in measurement precision between the two models is negligible in this case. Compared to the cut-based analysis, which yields a precision of 1.17%, the ML-based analysis improves the measurement precision by a factor of 1.4 in the $b\bar{b}$ channel. A detailed comparison of the cut-based and the ML-based analysis in the $b\bar{b}$ decay channel is provided in Table IV. The ML-based analysis enhances both the signal efficiency and background rejection rate, leading to an overall improvement in measurement precision.

It is worth noting that the improvement achieved by MIParT over ParT is more pronounced in the $\tau^+\tau^-$ channel than in the $b\bar{b}$ channel. This observation can be understood from the underlying event topology. The $\tau^+\tau^-$ final state involves missing transverse energy and multiple decay modes, resulting in more diverse kinematic and substructure patterns. The additional interaction layers in MIParT enable the network to capture these complex interparticle correlations more effectively. In contrast, the $b\bar{b}$ final state carries relatively complete and less ambiguous information; hence, the performance gain from enhanced interaction modeling in MIParT is naturally smaller.

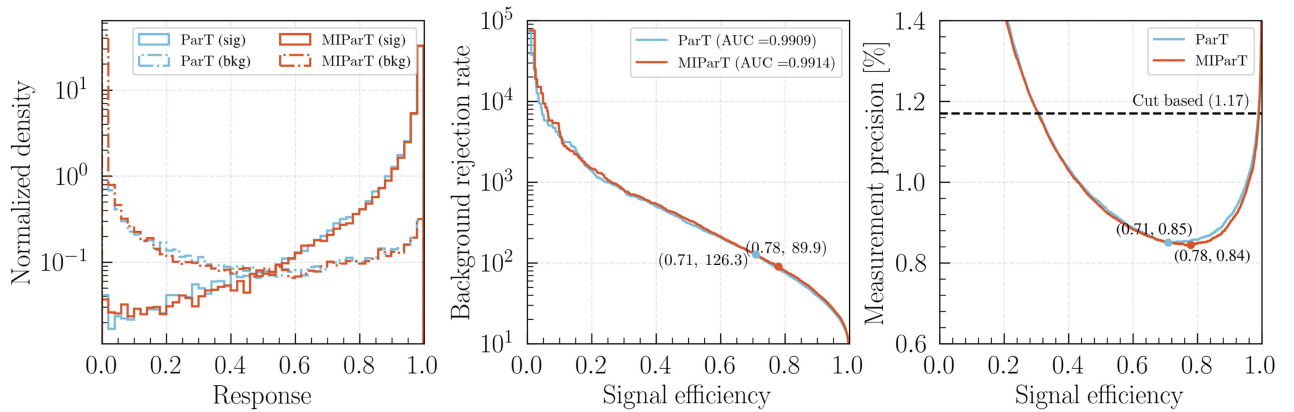


FIG. 10. The normalized distributions of the classifier response scores for the signal and background events in the testing dataset (left panel), the background rejection rate as a function of signal efficiency (middle panel), and the signal strength measurement precision versus signal efficiency (right panel) in the $b\bar{b}$ channel.

C. Comparative discussion

To assess the coverage of the N2HDM-F parameter space at the 240 GeV CEPC by the cut-based and the ML-based analysis, Fig. 11 shows the surviving samples in the $\tan\beta$ versus α_1 plane. The color scale indicates the signal strength measurement precision in the $\tau^+\tau^-$ channel without ML (upper left), with MIParT (upper right), in the $b\bar{b}$ channel without ML (lower left), and with MIParT (lower right), assuming an integrated luminosity of $L = 20\text{ab}^{-1}$ at the CEPC. Gray points represent samples that cannot be covered at the 5σ signal significance level in the corresponding scenario. The signal significance is computed using the Poisson formula [121],

$$\mathcal{Z} = \sqrt{2[(S+B)\ln(1+S/B) - S]}, \quad (20)$$

where S and B denote the expected numbers of signal and background events, respectively.

The following conclusions can be drawn from the Fig. 11:

- (i) As shown in the upper two panels, for a given α_1 , samples with larger $\tan\beta$ are more challenging to detect than those with smaller $\tan\beta$. Meanwhile, samples with $|\alpha|$ close to $\pi/2$, which corresponds to $|\sin\alpha_1|$ near 1, can be measured with higher precision. This observation is consistent with Eq. (5), which indicates that the coupling of S to τ is proportional to $\sin\alpha_1$ and inversely proportional to $\sin\beta$. The surviving samples with the maximum signal strength in the $\tau^+\tau^-$ channel can be detected with 2.2% precision without ML, and with 1.0% precision when using MIParT.
- (ii) From the lower two panels, one can see that almost all the surviving samples can be covered at the 5σ level in the $b\bar{b}$ decay channel except the surviving samples with $\cos(\beta - \alpha_1) \approx 0$, as it is the dominant decay channel for most parameter space. In addition, the use of MIParT can further improve the signal strength measurement precision; the maximum can

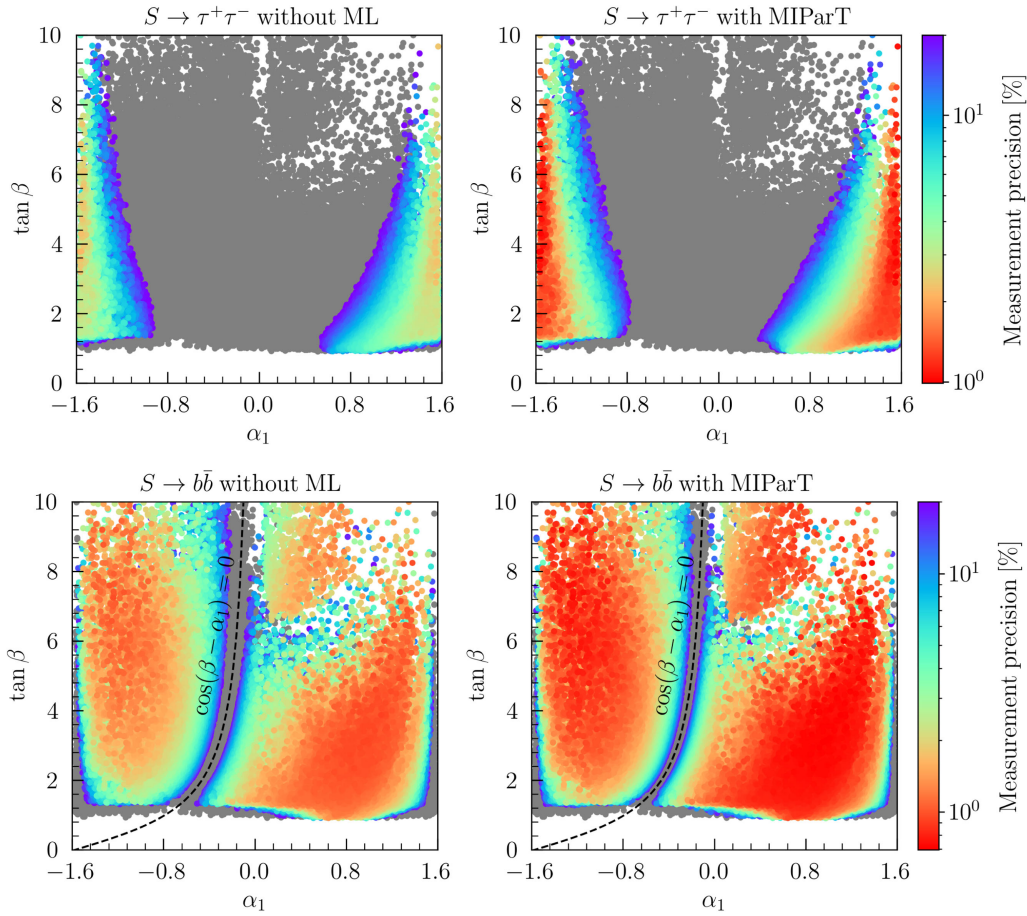


FIG. 11. The surviving samples in the $\tan\beta$ versus α_1 plane. The color represents the signal strength measurement precision in the $\tau^+\tau^-$ channel without ML (upper left), with MIParT (upper right), in the $b\bar{b}$ channel without ML (lower left), and with MIParT (lower right), based on an integrated luminosity of $L = 20\text{ab}^{-1}$ at CEPC. Gray points indicate samples that cannot be covered at the 5σ signal significance level under the corresponding scenario.

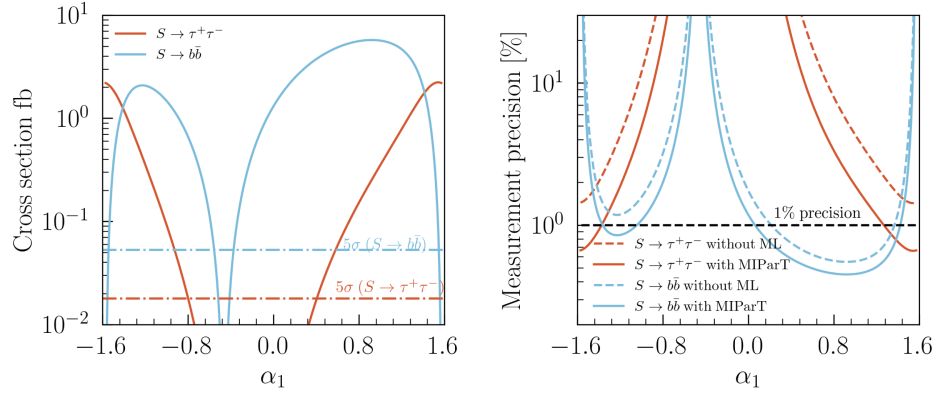


FIG. 12. The cross section (left panel) and the signal strength measurement precision (right panel) as a function of the model parameter α_1 in the $\tau^+\tau^-$ decay channel (red lines) and the $b\bar{b}$ decay channel (blue lines) for the benchmark points with $\cos\alpha_2 = \sqrt{2}/2$ and $\tan\beta = 2$.

be improved from 0.93% without ML to 0.69% with MIParT.

- (iii) As we expect, the samples with small $\tan\beta$ and $|\alpha|$ close to $\pi/2$ are hardly be detected in $b\bar{b}$ channel, but can be detected in $\tau^+\tau^-$ channel at high measurement precision.

For a detailed discussion on the improvement in CEPC coverage capacity provided by the ML-based analysis over the cut-based analysis, Fig. 12 shows the cross section (left panel) and the signal strength measurement precision (right panel) as a function of the model parameter α_1 in the $\tau^+\tau^-$ decay channel (red lines) and the $b\bar{b}$ decay channel (blue lines) for the benchmark points with $\cos\alpha_2 = \sqrt{2}/2$ and $\tan\beta = 2$. Without ML, only samples in the range $0.2 < \alpha_1 < 1.4$ can be measured with 1% precision in the $b\bar{b}$ decay channel, while no samples satisfy the 1% precision requirement in the $\tau^+\tau^-$ channel. After applying MIParT, samples with $\alpha_1 < -1.4$ or $\alpha_1 > 1.3$ can be measured at 1%

precision in the $\tau^+\tau^-$ decay channel. Only those in the region $-1 < \alpha_1 < 0$ remain undetectable at this precision level in both the $b\bar{b}$ and $\tau^+\tau^-$ channels. In terms of 5σ discovery potential at the 240 GeV CEPC with an integrated luminosity of $L = 20 \text{ ab}^{-1}$, signal regions are considered covered for the cross section exceeds 0.018 fb in the $\tau^+\tau^-$ channel or 0.051 fb in the $b\bar{b}$ channel. The only parameter region inaccessible at this significance is $-0.55 < \alpha_1 < -0.37$.

The CEPC signal strength measurement precisions of 5%, 2%, and 1% for the $\tau^+\tau^-$ and $b\bar{b}$ channels are presented in Fig. 13 on the $\text{Br}(S \rightarrow \tau^+\tau^-)$ versus C_{SZZ} plane (left panel) and the $\text{Br}(S \rightarrow b\bar{b})$ versus C_{SZZ} plane (right panel), respectively. The shadow area can be covered at the 5σ level, and the regions above the contours correspond to parameter spaces accessible at the corresponding precision. The signal strength in the $\tau^+\tau^-$ or $b\bar{b}$ decay channel from the Higgsstrahlung process is defined as

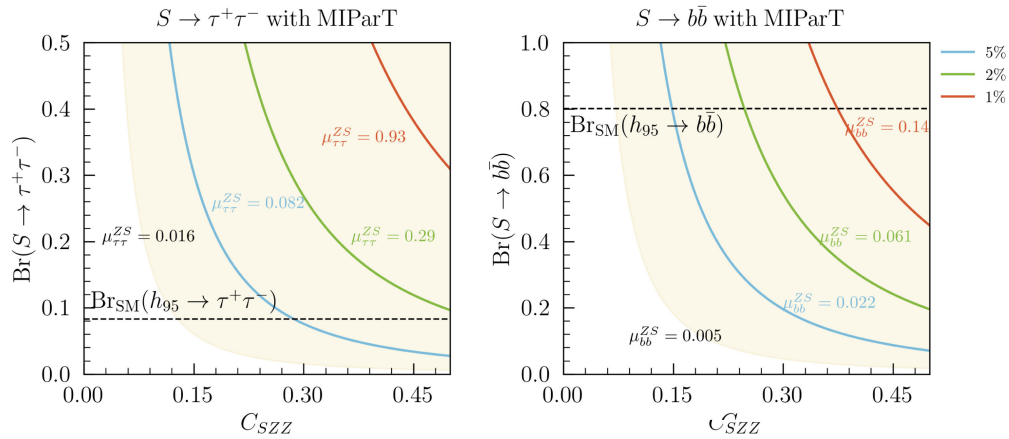


FIG. 13. The CEPC sensitivity in the $\tau^+\tau^-$ and $b\bar{b}$ channels on the $\text{Br}(S \rightarrow \tau^+\tau^-)$ versus C_{SZZ} plane (left panel) and $\text{Br}(S \rightarrow b\bar{b})$ versus C_{SZZ} plane (right panel). The shadow area can be covered at the 5σ level and the range upper the lines can be detected at the corresponding precision.

$$\mu_{\tau\tau(bb)}^{ZS} = C_{SZZ}^2 \times \frac{\text{Br}(S \rightarrow \tau^+\tau^-(b\bar{b}))}{\text{Br}_{\text{SM}}(h_{95} \rightarrow \tau^+\tau^-(b\bar{b}))}. \quad (21)$$

Based on this definition, the 5σ discovery thresholds are $\mu_{\tau\tau}^{ZS} > 0.016$ and $\mu_{bb}^{ZS} > 0.005$ for the $\tau^+\tau^-$ and $b\bar{b}$ channels, respectively. Furthermore, the minimum signal strengths required to achieve specific measurement precisions are outlined below:

- (i) For a 5%, 2%, or 1% precision in the $\tau^+\tau^-$ channel, $\mu_{\tau\tau}^{ZS}$ should exceed 0.082, 0.29, or 0.93, respectively.
- (ii) Correspondingly, in the $b\bar{b}$ channel, μ_{bb}^{ZS} should be greater than 0.022, 0.061, or 0.14.

Throughout this work, we report purely statistical uncertainties on the signal strength. A realistic CEPC analysis will, however, also be subject to experimental and theoretical systematic uncertainties. Primary sources of systematic error include jet reconstruction, jet energy scale and resolution, as well as theoretical uncertainties in the predictions for cross sections and spectral shapes. Based on studies from the CEPC Technical Design Report [122], the total systematic uncertainty is estimated to be on the order of 2%. A shift from statistical to systematic uncertainty dominance is expected to occur with our ML method, especially in the $\tau^+\tau^-$ channel.

In the process of $e^+e^- \rightarrow ZS$, as long as the mass of S is fixed, the observables are almost entirely determined by its mass. The coupling constants primarily affect the cross sections and branching ratios, whereas they do not significantly alter the event shapes. Consequently, their impact on the simulation results is relatively limited. These results then are also applicable to other multi-Higgs models featuring a light Higgs boson with a mass around 95.5 GeV.

The signal channel with the Z boson decaying to electrons has also been investigated using the same MC simulation. Due to the comparable tracking and calorimetric performance for electrons and muons in the CEPC baseline detector, the resulting sensitivity is similar. When the electron and muon channels of the Z boson are combined, the measurement precision can be further improved. A detailed discussion of the electron channel is omitted here for brevity, but it does not change our main conclusions. The hadronic $Z \rightarrow q\bar{q}$ decays, which have a much larger branching ratio, are in principle also promising, but require a dedicated analysis with more challenging QCD backgrounds and jet reconstruction. We therefore leave the exploration of hadronic recoil channels for future work.

V. CONCLUSIONS

In this work, we have explored the prospects for detecting a light Higgs boson with a mass of 95 GeV at future e^+e^- Higgs factories. We take the CEPC as a benchmark, with a center-of-mass energy of $\sqrt{s} = 240$ GeV and an integrated luminosity of $L = 20 \text{ ab}^{-1}$. The analysis leverages the ParT architecture and its more-interactive variant, MIParT, which

are specifically designed to process particle-level observables and extract their complex correlations. This advanced capability translates directly into superior measurement precision: on the same MC samples, our ML-based analysis outperforms the cut-based analysis, improving the expected statistical precision by factors of 2.3 and 1.4 in the $\tau^+\tau^-$ and $b\bar{b}$ channels, respectively. A shift from statistical to systematic uncertainty dominance is expected to occur with our ML method, especially in the $\tau^+\tau^-$ channel. Consequently, the region of the N2HDM-F parameter space accessible to high-precision measurement (1%) at the CEPC is significantly expanded. This improvement originates from the efficient exploitation of final-state particle information and the superior capability of transformer architectures to capture global event features and interparticle dependencies.

Most of the parameter space in the N2HDM-F can be probed at the 5σ significance level when the $\tau^+\tau^-$ and $b\bar{b}$ channels are combined, except in regions near $\cos(\beta - \alpha_1) \approx 0$, where the coupling of the 95 GeV scalar S to the Z boson is highly suppressed, leading to small production rates. For the surviving points with the largest signal strength, a measurement precision of 2.2% (0.93%) can be achieved in the $\tau^+\tau^- (b\bar{b})$ channel using a cut-based analysis, which improves to 1.0% (0.69%) when the MIParT architecture is employed.

To present the results in a model-independent way, we express the sensitivities of the 20 ab^{-1} CEPC in terms of the signal strength $\mu_{\tau\tau}^{ZS}$ and μ_{bb}^{ZS} . The ML-based analysis yields the following sensitivity reaches:

- (i) $\tau^+\tau^-$ channel: A 5σ discovery can be achieved for $\mu_{\tau\tau}^{ZS} > 1.6 \times 10^{-2}$, while a 1% precision measurement is attainable for $\mu_{\tau\tau}^{ZS} > 0.93$.
- (ii) $b\bar{b}$ channel: A 5σ discovery can be achieved for $\mu_{bb}^{ZS} > 5.0 \times 10^{-3}$, and a 1% precision measurement for $\mu_{bb}^{ZS} > 0.14$.

These results demonstrate that particle-level transformer networks can significantly enhance the sensitivity to light Higgs states, thereby improving the measurement precision at future lepton colliders. Although this study is performed using the CEPC baseline detector and luminosity assumptions, the conclusions are broadly applicable to other future e^+e^- Higgs factories such as FCC-ee [22,23] and ILC [24–26]. Since the ParT/MIParT classifiers rely mainly on kinematic features that depend on the collider energy and on m_S , rather than on machine-specific details, we expect the relative improvements over cut-based analysis demonstrated here to hold quite generally for other e^+e^- Higgs factories. Our work offers a compelling blueprint for applying ML-based analysis to advance the precision Higgs physics program and explore BSM Higgs sectors.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant No. 12275066 and by the startup research funds of Henan University.

The work of H. Y. was also supported by the National Natural Science Foundation of China under Grant No. W2441004. The work of M. R. was also supported by the National Key Program for SVT Research and Development under Contract No. 2024YFA1610603. And the work of K. W. was also supported by the Open Project of the Shanghai Key Laboratory of Particle Physics and Cosmology under Grant No. 22DZ2229013-3.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX: DETAILS OF THE ParT AND MIParT IMPLEMENTATIONS

We briefly summarize the main hyperparameters and setup for our ParT and MIParT implementations. Both architectures follow the original designs presented in Ref. [36] (ParT) and [37] (MIParT), with the following specific choices:

- (i) *Embedding dimension* d_{model} : 128 for ParT, 64 for MIParT.
- (ii) *Transformer blocks*: ParT uses eight particle- and two class-attention layers; MIParT uses five MI-particle, five particle, and two class-attention layers.
- (iii) *Attention heads*: $N_{\text{heads}} = 8$ for all attention layers.
- (iv) *Feed-forward hidden dimension* d_{ff} : 512 for ParT, 256 for MIParT.
- (v) *Dropout rate*: 0.1.
- (vi) *Optimiser*: RAdam with an initial learning rate of 1×10^{-3} .
- (vii) *Loss function*: Cross entropy with equal class weighting to mitigate class imbalance.
- (viii) *Batch size*: 126.
- (ix) *Training epochs*: 20.

The dataset was randomly split into 90% for training and 10% for validation. An independent sample of 100k events was held out for testing. No signs of overfitting were observed; the performance on the test set was consistent with that on the validation set.

Further architectural details, along with the complete training code, can be found in the original references and the associated GitHub home page: weaver [123], ParT [124], and MIParT [125].

-
- [1] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Lett. B* **716**, 30 (2012).
 - [2] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **716**, 1 (2012).
 - [3] K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, *Nucl. Phys. B* **466**, 189 (1996).
 - [4] G. R. Farrar and M. E. Shaposhnikov, *Phys. Rev. Lett.* **70**, 2833 (1993); **71**, 210(E) (1993).
 - [5] M. B. Gavela, P. Hernandez, J. Orloff, and O. Pene, *Mod. Phys. Lett. A* **9**, 795 (1994).
 - [6] A. D. Sakharov, *Pis'ma Zh. Eksp. Teor. Fiz.* **5**, 32 (1967).
 - [7] P. Huet and E. Sather, *Phys. Rev. D* **51**, 379 (1995).
 - [8] K. Rummukainen, K. Kajantie, M. Laine, M. E. Shaposhnikov, and M. Tsypin, in *Proceedings of the 5th International Workshop on Thermal Field Theories and Their Applications* (1998), [arXiv:hep-ph/9809435](https://arxiv.org/abs/hep-ph/9809435).
 - [9] Anisha, L. Biermann, C. Englert, and M. Mühlleitner, *J. High Energy Phys.* **08** (2022) 091.
 - [10] R. Barate *et al.* (LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, and OPAL Collaborations), *Phys. Lett. B* **565**, 61 (2003).
 - [11] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **793**, 320 (2019).
 - [12] A. Tumasyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **07** (2023) 073.
 - [13] A. Hayrapetyan *et al.* (CMS Collaboration), *Phys. Lett. B* **860**, 139067 (2025).
 - [14] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **01** (2025) 053.
 - [15] P. Sharma, A.-T. Mulaudzi, K. Mosala, T. Mathaha, M. Kumar, B. Mellado, A. Crivellin, M. Titov, M. Ruan, and Y. Fang, *Phys. Lett. B* **870**, 139953 (2025).
 - [16] K. Wang and J. Zhu, *Chin. Phys. C* **48**, 073105 (2024).
 - [17] J. Dutta, J. Lahiri, C. Li, G. Moortgat-Pick, S. F. Tabira, and J. A. Ziegler, *Eur. Phys. J. C* **84**, 926 (2024).
 - [18] Y. Dong, K. Wang, and J. Zhu, *Phys. Rev. D* **112**, 055013 (2025).
 - [19] Y. Dong, M. Ruan, K. Wang, H. Yang, and J. Zhu, *Chin. Phys. C* **50**, 031001 (2026).
 - [20] G. L. Bayatian *et al.* (CMS Collaboration), *J. Phys. G* **34**, 995 (2007).
 - [21] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet (ATLAS Collaboration), ATLAS Insertable B-Layer, CERN Technical Design Report No. CERN-LHCC-2010-013, ATLAS-TDR-19, 2010, <https://cds.cern.ch/record/1291633>.
 - [22] A. Abada *et al.* (FCC Collaboration), *Eur. Phys. J. Spec. Top.* **228**, 261 (2019).
 - [23] A. Abada *et al.* (FCC Collaboration), *Eur. Phys. J. C* **79**, 474 (2019).
 - [24] H. Baer *et al.* (ILC Collaboration), [arXiv:1306.6352](https://arxiv.org/abs/1306.6352).
 - [25] T. Behnke *et al.*, [arXiv:1306.6327](https://arxiv.org/abs/1306.6327).

- [26] D. M. Asner *et al.*, in Snowmass 2013: Snowmass on the Mississippi (2013), [arXiv:1310.0763](#).
- [27] M. Dong *et al.* (CEPC Study Group), [arXiv:1811.10545](#).
- [28] CEPC Study Group, [arXiv:1809.00285](#).
- [29] F. An *et al.*, *Chin. Phys. C* **43**, 043002 (2019).
- [30] W. Abdallah *et al.* (CEPC Study Group), *Radiat. Detect. Technol. Methods* **8**, 1 (2024); **9**, 184(E) (2025).
- [31] P. T. Komiske, E. M. Metodiev, and J. Thaler, *J. High Energy Phys.* **01** (2019) 121.
- [32] H. Qu and L. Gouskos, *Phys. Rev. D* **101**, 056019 (2020).
- [33] V. Mikuni and F. Canelli, *Eur. Phys. J. Plus* **135**, 463 (2020).
- [34] V. Mikuni and F. Canelli, *Mach. Learn. Sci. Tech.* **2**, 035027 (2021).
- [35] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, and T.-Y. Liu, *J. High Energy Phys.* **07** (2022) 030.
- [36] H. Qu, C. Li, and S. Qian, [arXiv:2202.03772](#).
- [37] Y. Wu, K. Wang, C. Li, H. Qu, and J. Zhu, *Chin. Phys. C* **49**, 013110 (2025).
- [38] A. Kundu, S. Maharana, and P. Mondal, *Nucl. Phys.* **B955**, 115057 (2020).
- [39] L. Liu, H. Qiao, K. Wang, and J. Zhu, *Chin. Phys. C* **43**, 023104 (2019).
- [40] D. Borah, S. Mahapatra, P. K. Paul, and N. Sahu, *Phys. Rev. D* **109**, 055021 (2024).
- [41] A. A. Abdelalim, B. Das, S. Khalil, and S. Moretti, *Nucl. Phys.* **B985**, 116013 (2022).
- [42] S. Ashanujjaman, S. Banik, G. Coloretti, A. Crivellin, B. Mellado, and A.-T. Mulaudzi, *Phys. Rev. D* **108**, L091704 (2023).
- [43] S. Banik, A. Crivellin, S. Iguro, and T. Kitahara, *Phys. Rev. D* **108**, 075011 (2023).
- [44] G. Coloretti, A. Crivellin, and B. Mellado, *Phys. Rev. D* **110**, 073001 (2024).
- [45] M. Maniatis and O. Nachtmann, *Phys. Rev. D* **111**, 055009 (2025).
- [46] K. Wang, F. Wang, J. Zhu, and Q. Jie, *Chin. Phys. C* **42**, 103109 (2018).
- [47] J. Cao, X. Jia, Y. Yue, H. Zhou, and P. Zhu, *Phys. Rev. D* **101**, 055008 (2020).
- [48] W. Li, H. Qiao, K. Wang, and J. Zhu, [arXiv:2312.17599](#).
- [49] C. Wang, J.-Q. Tao, M. A. Shahzad, G.-M. Chen, and S. Gascon-Shotkin, *Chin. Phys. C* **46**, 083107 (2022).
- [50] J. A. Aguilar-Saavedra and F. R. Joaquim, *Eur. Phys. J. C* **80**, 403 (2020).
- [51] P. Escribano, V. M. Lozano, and A. Vicente, *Phys. Rev. D* **108**, 115001 (2023).
- [52] P. S. B. Dev, R. N. Mohapatra, and Y. Zhang, *Phys. Lett. B* **849**, 138481 (2024).
- [53] A. Ahrich, M. L. Bellile, M. O. Khojali, M. Kumar, and A.-T. Mulaudzi, *Phys. Rev. D* **110**, 015025 (2024).
- [54] T. Biekötter, M. Chakraborti, and S. Heinemeyer, *Eur. Phys. J. C* **80**, 2 (2020).
- [55] R. Benbrik, M. Boukidi, K. Kahime, S. Moretti, L. Rahili, and B. Taki, *Phys. Lett. B* **868**, 139688 (2025).
- [56] S. Heinemeyer, C. Li, F. Lika, G. Moortgat-Pick, and S. Paasch, *Phys. Rev. D* **106**, 075003 (2022).
- [57] T. Biekötter and M. O. Olea-Romacho, *J. High Energy Phys.* **10** (2021) 215.
- [58] S. Banik, G. Coloretti, A. Crivellin, and H. E. Haber, *Phys. Rev. D* **111**, 075021 (2025).
- [59] J. Dutta, J. Lahiri, C. Li, G. Moortgat-Pick, S. F. Tabira, and J. A. Ziegler, [arXiv:2504.14529](#).
- [60] H. Xu, Y. Wang, X.-F. Han, and L. Wang, *Chin. Phys.* **50**, 013108 (2026).
- [61] R. Benbrik, M. Boukidi, S. Moretti, and S. Semlali, *Phys. Lett. B* **832**, 137245 (2022).
- [62] T. Biekötter, M. Chakraborti, and S. Heinemeyer, *Int. J. Mod. Phys. A* **36**, 2142018 (2021).
- [63] J. Cao, X. Guo, Y. He, P. Wu, and Y. Zhang, *Phys. Rev. D* **95**, 116001 (2017).
- [64] W. Li, H. Qiao, and J. Zhu, *Chin. Phys. C* **47**, 123102 (2023).
- [65] J. Cao, X. Jia, J. Lian, and L. Meng, *Phys. Rev. D* **109**, 075001 (2024).
- [66] J. Lian, *Phys. Rev. D* **110**, 115018 (2024).
- [67] J. Cao, X. Jia, and J. Lian, *Phys. Rev. D* **110**, 115039 (2024).
- [68] U. Ellwanger and C. Hugonie, *Eur. Phys. J. C* **83**, 1138 (2023).
- [69] U. Ellwanger, C. Hugonie, S. F. King, and S. Moretti, *Eur. Phys. J. C* **84**, 788 (2024).
- [70] K. Choi, S. H. Im, K. S. Jeong, and C. B. Park, *Eur. Phys. J. C* **79**, 956 (2019).
- [71] U. Ellwanger and C. Hugonie, *Eur. Phys. J. C* **84**, 526 (2024).
- [72] Z.-f. Ge, F.-Y. Niu, and J.-L. Yang, *Eur. Phys. J. C* **84**, 548 (2024).
- [73] S. Yaser Ayazi, M. Hosseini, S. Paktinat Mehdiabadi, and R. Rouzbehi, *Phys. Rev. D* **110**, 055004 (2024).
- [74] Q. Chang, X. Du, and P. Zhu, [arXiv:2509.26155](#).
- [75] A. Khanna, S. Moretti, and A. Sarkar, *Nucl. Phys.* **B1022**, 117229 (2026).
- [76] T. Biekötter, A. Grohsjean, S. Heinemeyer, C. Schwanenberger, and G. Weiglein, *Eur. Phys. J. C* **82**, 178 (2022).
- [77] J. A. Aguilar-Saavedra, H. B. Câmara, F. R. Joaquim, and J. F. Seabra, *Phys. Rev. D* **108**, 075020 (2023).
- [78] T. Biekötter, S. Heinemeyer, and G. Weiglein, *J. High Energy Phys.* **08** (2022) 201.
- [79] T. Biekötter, S. Heinemeyer, and G. Weiglein, *Eur. Phys. J. C* **83**, 450 (2023).
- [80] T. Biekötter, S. Heinemeyer, and G. Weiglein, *Phys. Lett. B* **846**, 138217 (2023).
- [81] T. Biekötter, S. Heinemeyer, and G. Weiglein, *Phys. Rev. D* **109**, 035005 (2024).
- [82] G. Arcadi, G. Busoni, D. Cabo-Almeida, and N. Krishnan, *Phys. Rev. D* **110**, 115028 (2024).
- [83] R. Benbrik, M. Boukidi, and S. Moretti, *Phys. Rev. D* **110**, 115030 (2024).
- [84] S. Iguro, T. Kitahara, Y. Omura, and H. Zhang, *Phys. Rev. D* **107**, 075017 (2023).
- [85] D. Azevedo, T. Biekötter, and P. M. Ferreira, *J. High Energy Phys.* **11** (2023) 017.
- [86] A. Belyaev, R. Benbrik, M. Boukidi, M. Chakraborti, S. Moretti, and S. Semlali, *J. High Energy Phys.* **05** (2024) 209.
- [87] T.-K. Chen, C.-W. Chiang, S. Heinemeyer, and G. Weiglein, *Phys. Rev. D* **109**, 075043 (2024).
- [88] P. Mondal and S. Samanta, [arXiv:2506.06427](#).

- [89] A. Kundu, P. Mondal, and G. Moulataka, [arXiv:2411.14126](#).
- [90] A. Hmissou, S. Moretti, and L. Rahili, *Phys. Rev. D* **113**, 015024 (2026).
- [91] A. Hmissou, S. Moretti, and L. Rahili, *Phys. Rev. D* **112**, 075049 (2025).
- [92] R. Benbrik, M. Boukidi, K. Kahime, S. Moretti, L. Rahili, and B. Taki, [arXiv:2510.19605](#).
- [93] A. Ahriche, *Phys. Rev. D* **110**, 035010 (2024).
- [94] X. Du, H. Liu, and Q. Chang, *Phys. Rev. D* **112**, 015019 (2025).
- [95] M. Y. Sassi and G. Moortgat-Pick, *Eur. Phys. J. C* **85**, 1230 (2025).
- [96] C.-Y. Chen, M. Freid, and M. Sher, *Phys. Rev. D* **89**, 075009 (2014).
- [97] A. Drozd, B. Grzadkowski, J. F. Gunion, and Y. Jiang, *J. High Energy Phys.* **11** (2014) 105.
- [98] S. von Buddenbrock, N. Chakrabarty, A. S. Cornell, D. Kar, M. Kumar, T. Mandal, B. Mellado, B. Mukhopadhyaya, R. G. Reed, and X. Ruan, *Eur. Phys. J. C* **76**, 580 (2016).
- [99] M. Muhlleitner, M. O. P. Sampaio, R. Santos, and J. Wittbrodt, *J. High Energy Phys.* **03** (2017) 094.
- [100] R. Coimbra, M. O. P. Sampaio, and R. Santos, *Eur. Phys. J. C* **73**, 2428 (2013).
- [101] M. Muhlleitner, M. O. P. Sampaio, R. Santos, and J. Wittbrodt, *Eur. Phys. J. C* **82**, 198 (2022).
- [102] S. Kanemura, T. Kubota, and E. Takasugi, *Phys. Lett. B* **313**, 155 (1993).
- [103] W. G. Hollik, G. Weiglein, and J. Wittbrodt, *J. High Energy Phys.* **03** (2019) 109.
- [104] P. M. Ferreira, M. Muhlleitner, R. Santos, G. Weiglein, and J. Wittbrodt, *J. High Energy Phys.* **09** (2019) 006.
- [105] J. Haller, A. Hoecker, R. Kogler, K. Mönig, T. Peiffer, and J. Stelzer, *Eur. Phys. J. C* **78**, 675 (2018).
- [106] P. Bechtle, D. Dercks, S. Heinemeyer, T. Klingl, T. Stefaniak, G. Weiglein, and J. Wittbrodt, *Eur. Phys. J. C* **80**, 1211 (2020).
- [107] P. Bechtle, S. Heinemeyer, T. Klingl, T. Stefaniak, G. Weiglein, and J. Wittbrodt, *Eur. Phys. J. C* **81**, 145 (2021).
- [108] I. Engeln, M. Muhlleitner, and J. Wittbrodt, *Comput. Phys. Commun.* **234**, 256 (2019).
- [109] S. Navas *et al.* (Particle Data Group), *Phys. Rev. D* **110**, 030001 (2024).
- [110] Z. Chen, Y. Yang, M. Ruan, D. Wang, G. Li, S. Jin, and Y. Ban, *Chin. Phys. C* **41**, 023003 (2017).
- [111] J. M. Yang, Y. Zhang, P. Zhu, and R. Zhu, *Phys. Rev. D* **108**, 075015 (2023).
- [112] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *J. High Energy Phys.* **06** (2011) 128.
- [113] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [114] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [115] D. Yu, M. Ruan, V. Boudry, H. Videau, J.-C. Brient, Z. Wu, Q. Ouyang, Y. Xu, and X. Chen, *Eur. Phys. J. C* **80**, 7 (2020).
- [116] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015).
- [117] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.
- [118] M. Selvaggi, *J. Phys. Conf. Ser.* **523**, 012033 (2014).
- [119] T. Zheng, J. Xu, L. Cao, D. Yu, W. Wang, S. Prell, Y.-K. E. Cheung, and M. Ruan, *Chin. Phys. C* **45**, 023001 (2021).
- [120] Y. Zhu, H. Liang, Y. Wang, Y. Che, H. Wang, C. Zhou, H. Qu, and M. Ruan, [arXiv:2506.11783](#).
- [121] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011); **73**, 2501(E) (2013).
- [122] S. P. Adhya *et al.* (CEPC Study Group), [arXiv:2510.05260](#).
- [123] <https://github.com/hqucms/weaver-core>.
- [124] https://github.com/jet-universe/particle_transformer.
- [125] <https://github.com/USST-HEP/MIParT>.