Probing Electroweak Phase Transition at CEPC via Exotic Higgs Decays with $4b$ Final States

Zhen Wang, Xuliang Zhu, Elham E Khoda, Shih-Chieh Hsu, Nikolaos Konstantinidis, Ke Li, Shu Li, Michael J. Ramsey-Musolf, Yanda Wu, and Yuwen E. Zhang

1. Tsung-Dao Lee Institute, Shanghai Jiao Tong University, 520 Shengrong Road, Shanghai 201210, China
2. Institute of Nuclear and Particle Physics, School of Physics and Astronomy, Key Laboratory for Particle Astrophysics and Cosmology (MOE), Shanghai Key Laboratory for Particle Physics and Cosmology (SKLPPC), Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China
3. Department of Physics, Duke University, Durham, North Carolina 27708, USA
4. Department of Physics, University of Washington, Seattle 98195-1560, USA
5. Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
6. Center for High Energy Physics, Peking University, 5 Yiheyuan Road, Beijing 100871, China
7. School of Mechanical and Electronic Engineering, Suzhou University, Suzhou 234000, Anhui, China
8. Amherst Center for Fundamental Interactions, Department of Physics, University of Massachusetts, Amherst, MA 01003, USA
9. Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

A strong first-order electroweak phase transition (EWPT) can be induced by light new physics weakly coupled to the Higgs. This study focuses on a scenario in which the first-order EWPT is driven by a light scalar $s$ with a mass between 15 and 60 GeV. A search for exotic decays of the Higgs boson into a pair of spin-zero particles, $h \rightarrow ss$, where the $s$-boson decays into $b$-quarks promptly is presented. The search is performed in events where the Higgs is produced in association with a $Z$ boson, giving rise to a signature of two charged leptons (electrons or muons) and multiple jets from $b$-quark decays. The analysis is considering a scenario of analyzing $5000 \text{fb}^{-1}$ $e^+e^-$ collision data at $\sqrt{s} = 240 \text{ GeV}$ from the Circular Electron Positron Collider (CEPC). This study with $4b$ final state conclusively tests the expected sensitivity of probing the light scalars in the CEPC experiment. Upper limits are set on the Higgs to double singlet cross-section times branching ratio with 95% CL. The ratio of these limits over the SM production cross-section is estimated to be around $5 \times 10^{-4}$ for the mass range of 15–60 GeV. The sensitivity reach is significantly higher than that can be achieved at the LHC.

Keywords: electroweak phase transition, exotic Higgs decay, Lepton collider, CEPC

DOI: 10.31526/LHEP.2023.436

1. INTRODUCTION

With the discovery of the Higgs boson, particle physics has entered a new phase. Accurately measuring the properties of the Higgs boson will be an essential aspect of any strategic plan for high-energy physics in the forthcoming decades. The Higgs boson’s properties and interactions also dictate the character of the ElectroWeak Phase Transition (EWPT). Hence, the investigation of the Higgs boson not only enhances our comprehension of the natural world but also propels us toward the discovery of novel physical phenomena.

The CEPC is a proposed next-generation electron-positron collider and will be operating at around $\sqrt{s} = 240 \text{ GeV}$, which is expected to yield over one million Higgs bosons which provides excellent opportunities for people to perform studies on Higgs boson. At CEPC, the Higgs boson is produced associated with a $Z$ boson via electron-positron annihilation, it has unique advantages over the study in Large Hadron Collider (LHC), and the environment is much cleaner and can provide excellent signal-to-noise ratio which is the key to precise measurements and discovery potential for Beyond Standard Model (BSM) physics. This study will help to further strengthen the BSM aspect of the physics motivation for CEPC in addition to its important role in the SM Higgs factory for precision physics.

A key question in the high energy physics and cosmology frontier is to investigate the thermal history of ElectroWeak Symmetry Breaking (EWSB) in the early universe. Suitable types of EWSB could provide an answer to many mysteries of our universe, such as the matter-antimatter asymmetry [1]. The nonperturbative lattice simulations have shown that our universe underwent a smooth crossover phase transition within the standard model (SM) [2, 3, 4, 5]. This type of EWSB catches little interest for us, since electroweak baryogenesis [1, 6] and stochastic gravitational wave background [7] are both obtained from a strong first-order electroweak phase transition (SFOEWPT), and the gravitational wave could be observed at future space-based experiment facilities such as LISA [8]. An SFOEWPT is expected within the BSM scenarios, including the Higgs field portal interactions with gauge singlet–real [9] or complex [10]—or with electroweak multiplets [11], etc. Any new particles involved in such SFOEWPT cannot interact too weakly with the SM Higgs boson, nor can they have masses too heavy with respect to the electroweak temperature scale [12]. In this work, we consider a lighter BSM real singlet particle with a mass lighter than half of the SM Higgs particle. This new scalar $s$ is coupled to the SM Higgs $h$ and it could catalyze an SFOEWPT. The theory analysis of this BSM model has shown that there exists a lower bound on the corresponding exotic Higgs decay branching ratio as a function of new scalar...
mass [13]. We will briefly describe the model in the next section; see [14] for a complementary study of such exotic Higgs decays as a signature of an SFOEWPT.

The relevant potential involving the exotic scalar field $S$ and the Higgs field, $H$, can be written as [15, 13]

$$V = -\mu^2 |H|^2 + \lambda |H|^4 + \frac{1}{2} a_1 |H|^2 S + \frac{1}{2} a_2 |H|^2 S^2$$
$$+ b_1 S + \frac{1}{2} b_2 S^2 + \frac{1}{3} b_3 S^3 + \frac{1}{4} b_4 S^4.$$  (1)

After EWSB, the two scalar fields can be parametrized as

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}, \quad S = v_s + s,$$  (2)

where $v = 246\text{GeV}$ is the zero temperature vacuum expectation value (VEV) for the Higgs field and $v_s$ is the VEV for the singlet field. One can add a tadpole term in the potential and set $v_s = 0$ or remove the tadpole term while keeping $v_s$. We choose the former in this study.

The mass eigenstates can be produced by the mixing of two scalar fields $h$ and $s$:

$$h_1 = h \cos \theta + s \sin \theta,$$
$$h_2 = -h \sin \theta + s \cos \theta.$$  (3)

We take $h_1$ to be singlet-like particle with mass $m_1$, and $h_2$ to be SM-like Higgs particle with $m_2 = 125\text{GeV}$. To see a visible decay of the SM-like Higgs $h_2$, we require the mixing angle $\cos \theta \neq 0$. Constrained by experiment limit, $|\cos \theta|$ cannot be too large and the EWPT region with successful tunneling is insensitive to the precise value. We take $\cos \theta = 0.01$, and this value is large enough for $h_1$ to decay promptly [13]. The trilinear scalar interactions in terms of mass eigenstates read

$$V \supset \frac{1}{6} \lambda_{111} h_1^3 + \frac{1}{2} \lambda_{211} h_2 h_1^2 + \frac{1}{2} \lambda_{212} h_2^2 h_1 + \frac{1}{6} \lambda_{222} h_2^3,$$  (4)

where $\lambda_{211}$ governs the exotic Higgs decay. This cubic coupling’s value is

$$\lambda_{211} = 2 \sqrt{s} c h_3 + \frac{d_1}{\sqrt{s}} \left( c^2 - 2 s^2 \right) + \left( 2 c^2 - 2 s^2 \right) s v a_2 - 6 \lambda \alpha c^2 v,$$  (5)

where $s \equiv \sin \theta$ and $c \equiv \cos \theta$.

The total widths of the singlet-like scalar and SM-like Higgs are

$$\Gamma (h_1) = \cos^2 \theta \Gamma (h_2),$$
$$\Gamma (h_2) = \sin^2 \theta \Gamma_{SM} |_{m_2} + \Gamma (h_2 \to h_1 h_1),$$  (6)

where the exotic decay partial width is

$$\Gamma (h_2 \to h_1 h_1) = \frac{1}{32 \pi m_2^2} \lambda_{211}^2 \sqrt{1 - \frac{4 m_1^2}{m_2^2}}.$$  (7)

As shown in [13], only $m_1$, $\cos \theta$, $a_2$, $b_3$, and $b_4$ are free parameters in this model. We take $m_1 \in [5, 60] \text{GeV}$ to necessitate the occurrence of the exotic Higgs decay and $\cos \theta = 0.01$ to fit the experimental constraint. The exotic Higgs decay opens a powerful probe for the small mixing angle $\cos \theta$, which can be detected up to $O(0.01)$ in the future colliders, see [13, 16] for more details. Under a given $m_1$ and $\cos \theta$, we scan over $a_2, b_3, v \in [10^{-4}, 1]$ and $b_4 \in [10^{-5}, 1]$ in the numerical simulation. The parameter region that satisfies an SFOEWPT had been worked out in [13], whose results we have reproduced for this study.

The search for such new singlet-like scalar particles could be performed in many high-energy physics experimental facilities, including the ongoing LHC and potential future large lepton colliders [13, 14]. Among these facilities, future $e^+ e^-$ colliders could have sufficient sensitivity to probe the new scalar masses down to at least $\sim 10 \text{GeV}$ during an SFOEWPT scenario [13, 14, 17]. Under such attractive motivation, we propose performing a detailed study of a future exotic Higgs decay search with the reference detector simulation CEPC.

2. DETECTOR

The baseline design of the CEPC detector is developed from the ILD concept [18, 19], and it is further optimized for the collision situation for CEPC beams. It is guided by the particle flow principle of the final state particle reconstruction-oriented algorithm. The Particle Flow Algorithm (PFA) reconstructs particles and takes advantage of hits within these detectors. The baseline concept contains the inner part and outer part from the vertex detector to the muon detectors, the default design option for the tracking system is a hybrid of a silicon tracker and a Time Projection Chamber which is usually referred to as TPC. The benchmark concept benefits from high granular sampling ECAL and HCAL, which provides 3D spatial and energy information. The calorimeter system provides energy measurements for photons and neutral hadrons. The jet energy resolution of 3%-5% is expected for jets between 20 and 100 GeV. What comes after the HCAL is an iron yoke instrumented with the muon detector [20].

3. SIGNAL AND BACKGROUND MODELING

MADGRAPH5_aMC@NLO [21] and WHIZARD [22] Monte Carlo event generators are used in this analysis to simulate the signal and background process. The singlet-like exotic Higgs signal model described in Section 1 is implemented using FeynRules [23] and imported to MADGRAPH5_aMC@NLO. The signal events are generated using MadGraph5_aMC@NLO at leading order (LO), and the parton showering and hadronization modeling is done with PYTHIA8 [24]. Signal samples are generated for different mass points starting from 15 GeV to 60 GeV with an interval of 5 GeV. The other SM background events are generated using WHIZARD, with PYTHIA8 to simulate parton showering and hadronization. All of the samples used in the analysis are generated at nonpolarized electron-positron collision at $\sqrt{s} = 240 \text{GeV}$. The detector simulation is performed by Mokka [25], a Geant4 [26] based detector simulation software. The simulated hits are digitized and reconstructed with ArborPFA [27]. All the backgrounds are modeled using the samples simulated by CEPC.
4. OBJECT RECONSTRUCTION

The charged leptons like electrons and muons are identified using the Lepton Identification in Calorimeter with High Granularity (LICH) algorithm [28]. LICH is a multivariate lepton identification technique developed for future detectors using high-granularity calorimeters. Electrons and muons are identified with 99.7% and 99.9% efficiencies, respectively, where the misidentification rates are smaller than 0.07%. We also apply lepton isolation in the analysis. It is required to have \( E_{\text{cone}}^2 < 4E_l + 12.2\text{GeV}, \) where \( E_l \) is the lepton energy, and \( E_{\text{cone}} \) is the energy within a cone of \( \cos \theta_{\text{cone}} > 0.98 \) around the lepton. Jets in the event are reconstructed from the particle flow object (PFO) using the LCFIPlus software package [29], which integrates vertex finding and flavor tagging with jet reconstruction. So, LCFIPlus package also provided the flavor tagging information for each jet. Before jet reconstruction, leptons are eliminated from the ArborPFO. Jets are clustered by grouping particles based on their momentum and spatial distribution using the Durham algorithm [30]. In this analysis, we performed an exclusive clustering requiring exactly four jets in the final state.

5. EVENT SELECTION

We want to keep events with four b-jets coming from singlet-like exotic scalar and two same-flavor opposite-sign leptons coming from the Z-boson decay. The events are selected in two stages using a set of loose preselections and a more sophisticated multivariate approach to classify signal and background events using a Boosted Decision Tree (BDT).

5.1. Preselection

At the preselection stage, a set of loose criteria is applied to remove background events as much as possible. As the analysis studies both electron and muon final states, each event is required to have two same-flavor isolated leptons with opposite charges \((e^+e^- \text{ or } \mu^+\mu^-)\). Each electron and muon must have energy higher than 20 GeV. Furthermore, the two leptons are required to have \( |\cos \theta_{\text{cone}}| < 0.71 \text{ or } |\cos \theta_{\text{cone}}| < 0.81 \). To remove most of the background events with lepton pair production, we require the angle between the two isolated tracks corresponding to the leptons to satisfy \( \cos \phi_{e^+e^-} > -0.93 \text{ or } \cos \phi_{\mu^+\mu^-} > -0.93 \). To cut more backgrounds, we require the invariant mass of the lepton system \( (M_{\ell\ell}) \) to be within the Z-mass window of 77.5–104.5 GeV. In signal-like events, the recoil mass should be around the Higgs boson mass. So, to further suppress the non-Higgs backgrounds, the recoil mass of the dilepton system \( M_{\ell\ell}^{\text{recoil}} \) is required to be in the range of \( M_{\ell\ell}^{\text{recoil}} \in [124, 140] \text{ GeV} \), where the recoil mass is defined in equation (8). The selections are summarized in Table 1.

\[
M_{\ell\ell}^{\text{recoil}} = \sqrt{\left(\sqrt{s} - E_{\ell} - E_{\ell}\right)^2 - \left(\vec{P}_{\ell} + \vec{P}_{\ell}\right) \cdot \left(\vec{P}_{\ell} + \vec{P}_{\ell}\right)}.
\] (8)

After selecting events with two leptons and four jets, the singlet-like exotic scalar is reconstructed by combining the jets. The reconstruction process involves checking all possible combinations of jet pairs and selecting the pair with the smallest mass difference. Reconstructed signal mass distributions are shown in Figure 1. As we go higher in mass, the distribution becomes wider.

### Table 1: List of criteria applied in the pre-selection.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Selection Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lepton ( p_T )</td>
<td>( p_T^{\text{lep}} &gt; 20 \text{ GeV} )</td>
</tr>
<tr>
<td>Lepton angle</td>
<td>(</td>
</tr>
<tr>
<td>Track angle</td>
<td>( \cos \phi &gt; -0.74 )</td>
</tr>
<tr>
<td>Dilepton mass</td>
<td>( M_{\ell\ell} \in [77.5, 104.5] \text{ GeV} )</td>
</tr>
<tr>
<td>Recoil mass</td>
<td>( M_{\ell\ell}^{\text{recoil}} \leq [124, 140] \text{ GeV} )</td>
</tr>
</tbody>
</table>

The background contributions are divided into three categories: Higgs processes with \( b\bar{b} \) final state \((\ell\ell H bb)\), other Higgs processes \((\ell\ell H)\), and non-Higgs processes. Since the \( H \rightarrow b\bar{b} \) process alone forms a large background, it is considered a separate group; all other Higgs decay modes are grouped into a different category.

5.2. Flavor Tagging

Flavor tagging has a significant impact in suppressing the backgrounds further and improving the reconstruction of \( M_h \). We used the flavor tagging toolkit in LCFIPLUS, which relies on a multivariate approach. The training for the \( b\)-tagging algorithm is accomplished using the gradient-boosted decision trees, utilizing variables such as jet kinematics, track impact parameters, and secondary vertex parameters. The \( b\)-tagging model returns a \( b\)-likelihood value for each jet. The \( b\)-likelihood \( (L_b) \) value represents the probability of a jet being a \( b\)-jet and the values lie between 0 and 1. The \( b\)-likelihoods \( (L_b) \) of the individual jets are used to compute a combined \( b\)-likeliness or \( b\)-jet efficiency \( (f_b) \), defined as

\[
f_b = \frac{L_{b1}L_{b2}L_{b3}L_{b4} + (1 - L_{b1})(1 - L_{b2})(1 - L_{b3})(1 - L_{b4})}{L_{b1}L_{b2}L_{b3}L_{b4}}.
\] (9)

We further define the \( b\)-jet inefficiency factor as

\[
b_{\text{inj}} = \log (1 - f_b).
\] (10)

In this analysis, \( b_{\text{inj}} \) is used to enhance the rejection of non-\( b\)-jet backgrounds. The signal process has four \( b\)-quarks in final...
briefly described in Table 3. Kinematic variables like the curve (AUC) of $\sim$ parameters for each model. It helped to achieve an area under configuration for each model is different and they are summa-
point. The model hyperparameters are optimized using Op-
$^{15}$ GeV to 60 GeV; a separate model is trained for each mass
the singlet-like exotic scalar. Since there are two
$M_{s}$ states; therefore, the possibility for such events to be missed by
the b-tagging tool is smaller than other processes.

### 5.3. Gradient-Boosted Decision Tree

After preselections, this analysis follows a multivariate ap-
proach by training a BDT to enhance signal sensitivity. The
BDT is implemented using the LIGHTGBM [31] python pack-
age and trained to classify the signal events from the rest of
the standard model background events. In particular, we are
using a gradient-boosted decision tree for our study. While sev-
eral widely used tools rely on depth-wise tree growth, LIGHT-
GBM grows the trees leaf-wise. The leaf-wise growth algorithm
can lead to faster convergence compared to depth-wise growth.
However, if not accompanied by appropriate parameters, leaf-
wise growth might lead to overfitting. To maximize the sen-
itivity of signal samples at every mass point, 10 models are
trained to classify the signal events from the rest of the
background.

The models are optimized using Optuna [32] hyperparameter optimization framework. The best
configuration for each model is different and they are summa-
ized in Table 2. A total of 100 trials are used to tune the hyper-
parameters for each model. It helped to achieve an area under
the curve (AUC) of $\sim 0.99$ for every BDT.

The BDTs are trained using 24 input variables which are
briefly described in Table 3. Kinematic variables like $p_{T}$ of lead-
ing and subleading leptons, recoil mass of the dilepton system,
the energy of each jet, invariant mass of the four-jet system,
and recoil mass of the four-jet system are used for the BDT train-
ing. The mass of the di-b jet system reconstructs the mass of
the singlet-like exotic scalar. Since there are two s particles in
the process, an average of the two reconstructed masses ($M_{bb}$)
is used for BDT training. In addition, the difference between
the reconstructed masses of the two s particles ($M_{\text{diff}}^{s}$) is also
used as a training variable. Apart from the other kinematic
observables, the number of particles used to reconstruct the four
jets of an event $N_{\text{particle}}^{s}$ plays a crucial role to suppress fake
jets. We have considered only jet constituent particles with energy higher than 0.4 GeV. Selected distances between pairs of the top 7 jet constituents ranked by transverse momentum are
used. Furthermore, we use the boosted angle and opening angle
between the reconstructed s particles along with the helicity
angles of two jets. The helicity angle of only the first jet and
third jet $|\cos \theta^{\text{Helicity}}_{1/2}|$ are used because the other two jets have
the same value. The BDT also used the $b_{\text{ineff}}$ to further increase
the power of the model to separate signal and backgrounds.

<table>
<thead>
<tr>
<th>$M_{s}$ [GeV]</th>
<th>$N_{\text{leaves}}$</th>
<th>Data$^{\text{leaf}}_{\min}$</th>
<th>$f_{\text{bagging}}$</th>
<th>$f_{\text{feature}}$</th>
<th>$\lambda_{11}$</th>
<th>$\lambda_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>11</td>
<td>52</td>
<td>$4.62E - 01$</td>
<td>$7.32E - 01$</td>
<td>$2.10E - 02$</td>
<td>$1.13E - 07$</td>
</tr>
<tr>
<td>20</td>
<td>250</td>
<td>23</td>
<td>$5.68E - 01$</td>
<td>$6.21E - 01$</td>
<td>$1.02E - 01$</td>
<td>$3.74E - 01$</td>
</tr>
<tr>
<td>25</td>
<td>13</td>
<td>59</td>
<td>$7.06E - 01$</td>
<td>$6.95E - 01$</td>
<td>$9.24E - 06$</td>
<td>$1.21E - 04$</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>11</td>
<td>$5.21E - 01$</td>
<td>$5.33E - 01$</td>
<td>$2.54E - 08$</td>
<td>$1.11E + 00$</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
<td>53</td>
<td>$8.79E - 01$</td>
<td>$8.85E - 01$</td>
<td>$1.59E - 04$</td>
<td>$2.00E - 05$</td>
</tr>
<tr>
<td>40</td>
<td>215</td>
<td>17</td>
<td>$8.75E - 01$</td>
<td>$6.60E - 01$</td>
<td>$2.56E - 02$</td>
<td>$3.02E - 04$</td>
</tr>
<tr>
<td>45</td>
<td>98</td>
<td>39</td>
<td>$8.11E - 01$</td>
<td>$6.03E - 01$</td>
<td>$1.29E - 08$</td>
<td>$5.41E - 05$</td>
</tr>
<tr>
<td>50</td>
<td>185</td>
<td>77</td>
<td>$6.80E - 01$</td>
<td>$6.34E - 01$</td>
<td>$7.62E - 06$</td>
<td>$6.78E - 05$</td>
</tr>
<tr>
<td>55</td>
<td>110</td>
<td>92</td>
<td>$9.81E - 01$</td>
<td>$5.62E - 01$</td>
<td>$7.18E - 07$</td>
<td>$4.65E - 01$</td>
</tr>
<tr>
<td>60</td>
<td>12</td>
<td>78</td>
<td>$7.02E - 01$</td>
<td>$1.81E - 01$</td>
<td>$3.68E - 05$</td>
<td>$5.19E - 04$</td>
</tr>
</tbody>
</table>

**Table 2:** Hyperparameters of 10 different LIGHTGBM models after optimization with Optuna. The first column refers to signal mass. $N_{\text{leaves}}$ is maximum number of leaves (or terminal nodes) in a decision tree. Data$^{\text{leaf}}_{\min}$ controls the minimum number of data samples required to be present in a leaf (or terminal node) of a decision tree. $f_{\text{bagging}}$ and $f_{\text{feature}}$ refers to fraction of data samples (rows) and features (columns) used for training, respectively. $\lambda_{11}$ and $\lambda_{12}$ controls the L1 and L2 regularization term in the objective function of LIGHTGBM.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{T}^{\ell}$, $E_{T}$</td>
<td>Lepton $p_{T}$</td>
</tr>
<tr>
<td>$M_{\text{recoil}}^{l}$</td>
<td>Recoil mass of the dilepton system</td>
</tr>
<tr>
<td>$E^{0}, E^{1}, E^{2}, E^{3}$</td>
<td>Energy of each reconstructed jet</td>
</tr>
<tr>
<td>$M_{4j}$</td>
<td>Invariant mass of the four-jet system</td>
</tr>
<tr>
<td>$N_{\text{particle}}^{s}$</td>
<td>Number of particles inside the four jets</td>
</tr>
<tr>
<td>$y_{12}, y_{23}, y_{34}, y_{45}, y_{56}, y_{67}$</td>
<td>Distance between jet constituents</td>
</tr>
<tr>
<td>$\cos \theta_{\text{boost}}^{s}$, $\cos \theta_{\text{boost}}^{s}$</td>
<td>Boosted angle between two jets from each s candidate</td>
</tr>
<tr>
<td>$\cos \theta_{\text{open}}^{s}$</td>
<td>Opening angle between the reconstructed s particles</td>
</tr>
<tr>
<td>$</td>
<td>\cos \theta_{0}^{\text{Helicity}}</td>
</tr>
<tr>
<td>$b_{\text{ineff}}$</td>
<td>b-jet inefficiency</td>
</tr>
</tbody>
</table>

**Table 3:** Variables used for the BDT training. Helicity angle of the first jet and the third jet is used because the other two jets have the same value.
6. UNCERTAINTY ESTIMATION

The major source of systematic uncertainty comes from event yields of fixed background. In order to estimate uncertainties from background modeling, event yields of primary $\ell\ell Hbb$ background are varied up and down by 5% while other background processes are varied by 100% [28]. The systematic uncertainties from luminosity [33] and lepton identification [28] are considered to be small and therefore ignored in this analysis.

Furthermore, the analysis relies on flavor tagging; so corresponding uncertainties are considered. We are following the method from [34], where the uncertainty was estimated to be 0.78% using the $Z\rightarrow q\bar{q} + \mu^+\mu^-$ control sample. However, given that the final state of this analysis includes more jets, a more conservative approach is adopted, and a 1% flat uncertainty is assigned as the flavor tagging uncertainty.

Jet energy resolution (JER) also plays an important role in this analysis. The jets in our signal process are soft, and the CEPC detector is expected to perform worse while reconstructing these soft jets. We estimated this uncertainty for the low-energy signal-like jets by calculating the energy difference between a truth jet and a reconstructed jet using the $e^+e^- k_t$-algorithm from the FastJet package [35]. An additional correction factor is added to account for the difference between the two jet clustering packages, FastJet and LCFIPlus. The LCFIPlus jet clustering algorithm is reading the vertex detector information from the particle-flow objects (PFO) in CEPC simulation. The total uncertainty is estimated by combining these two independent factors:

$$\text{JER} = \sqrt{\sigma_{\text{FastJet-TruthJet}}^2 + \sigma_{\text{FastJet-LCFIPlus}}^2}$$

where $\sigma$ is the jet resolution uncertainty. The estimated JER uncertainty is shown in Figure 3 as a function of truth jet energy.
Out of all the three different sources of uncertainties, the jet energy resolution uncertainty tends to be negligible after the statistical fit. The post-fit flavor tagging uncertainty is ∼1%. The background modeling uncertainty due to limited yields of Higgs process is around ∼4% post fit. The contribution from non-Higgs processes is between 2 and 5% above $M_s = 20\text{GeV}$, whereas it is around 10% at $M_s = 15\text{GeV}$. For the non-Higgs processes, it ranges between 10 and 30%.

7. RESULTS
In this analysis, we used the BDT classifier’s raw output as the main discriminant. The BDT score is much higher for a signal event compared to that of the background events as shown in Figure 4. Both systematic and statistical uncertainties discussed in Section 6 are shown as the error band in Figure 4.

The bins of the BDT score distribution are used to define a test statistic based on profile likelihood ratio. The test statistic is used to set an expected upper limit on the signal cross section × branching ration ($\sigma_{ZH} \times B(H \rightarrow ss)$) at 95% confidence level (CL) using the CL$_e$ method [36, 37, 38]. All the systematic uncertainties discussed in Section 6 and the statistical uncertainty are considered in the fit. The systematic uncertainties are included to the likelihood using nuisance parameters. Gaussian, log-normal, or Poisson priors are used in the likelihood to include the nuisance parameters.

Figure 5 and Table 4 show the ratio of 95% CL upper limit on ($\sigma_{ZH} \times B(H \rightarrow ss)$) and SM production cross section ($\sigma_{SM}$) as a function of singlet mass ($M_s$). The expected limit curve is almost flat, and we can probe the singlet scalar with mass as low as 15GeV. Due to limited signal simulations, our results are not extended below 15GeV.

The upper limits of the existing results from ATLAS and CMS using different channels and the projections for future High Luminosity LHC (HL-LHC) are also overlaid for comparison. The HL-LHC limit can reach up to 0.1 with the full dataset while CEPC can push the limit to ∼0.002 using a cut-and-count analysis. Cut-based approach is studied and optimized with “n−1” method but it is unnecessary to develop selections based on 10 signal samples. We decide to release selections and use 10 BDT classifiers to deal with 10 signal samples. BDT has explored more power in variables such as $b_{finef}$ and therefore improves the limits further by a factor ranging from 2~4 depending on signal masses.

The CEPC can have an unprecedented sensitivity in the search for such scalar particles coming from Higgs decay compared to the HL-LHC. A similar study was conducted and submitted as a Snowmass white paper [39]. This paper performs a complete study of jet energy resolution which is crucial for this analysis as the final contains several soft jets. The paper also focuses on multivariate analysis to improve signal sensitivity, and the input variables are carefully optimized.

8. CONCLUSION
This paper presents a search for exotic decays of the Higgs boson into a pair of spin-zero singlet-like scalar particles, $H \rightarrow ss$, where each s-quark decays into two $b$-quarks. We particularly studied the Higgs produced in association with a $Z$ boson that decays leptonically. The study is done in a scenario of analyz-
Experimental uncertainties. The study with 4b tagged simulation using realistic detector characteristics and extracts the importance of carrying out such studies with a dedicated limit compared to the projected values in [13]. Hence, it demonstrates that this realistic study yields a weaker exclusion for $M_H$.

<table>
<thead>
<tr>
<th>$M_t$ [GeV]</th>
<th>$\sigma_{ZH} \times B(H \rightarrow ss)/\sigma_{SM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>$4.07 \times 10^{-4}$</td>
</tr>
<tr>
<td>20</td>
<td>$3.50 \times 10^{-4}$</td>
</tr>
<tr>
<td>25</td>
<td>$5.45 \times 10^{-4}$</td>
</tr>
<tr>
<td>30</td>
<td>$4.49 \times 10^{-4}$</td>
</tr>
<tr>
<td>35</td>
<td>$5.34 \times 10^{-4}$</td>
</tr>
<tr>
<td>40</td>
<td>$5.50 \times 10^{-4}$</td>
</tr>
<tr>
<td>45</td>
<td>$6.32 \times 10^{-4}$</td>
</tr>
<tr>
<td>50</td>
<td>$6.63 \times 10^{-4}$</td>
</tr>
<tr>
<td>55</td>
<td>$6.09 \times 10^{-4}$</td>
</tr>
<tr>
<td>60</td>
<td>$4.58 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

**TABLE 4:** The expected 95% CL upper limit for the product of the signal’s cross-section times branching ratio relative to the Standard Model cross-section. The results are presented for the combined electron $e^+e^-$ and muon $\mu^+\mu^-$ $H$ channels using the multivariate BDT approach.

Expected 95% CL upper limits

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

ACKNOWLEDGMENTS

During the course of this work, we had several helpful conversations with Yu Bai, Gang Li, and Manqi Ruan, for which we are thankful. We thank Yu Bai for his help with event reconstruction and suggestions on baseline selections. We thank Gang Li and Manqi Ruan for their help with estimating the jet energy resolution uncertainties. Michael Ramsey-Musolf and Yanda Wu are partially supported by the National Natural Science Foundation of China under grant no. 19Z10301029. Ke Li, Elham E. Khoda, and Shih-Chieh Hsu are supported by the National Science Foundation (NSF) under grant no. 2110963.

References


[16] Patrick Draper, Jonathan Kozaczuk, and Scott Thomas. Precision inclusive Higgs physics at $e^+e^-$ colliders with tracking detectors and without calorimetry. *JHEP*, 09:174,
2020.


[34] Yu Bai, Chun-Hui Chen, Ya-Quan Fang, Gang Li, Man-Qi Ruan, Jing-Yuan Shi, Bo Wang, Pan-Yu Kong, Bo-Yang Lan, and Zhan-Feng Liu. Measurements of decay branching fractions of $h \rightarrow b\bar{b}/c\bar{c}/gg$ in associated $(e^+e^-/\mu^+\mu^-)h$ production at the CEPC. *Chinese Physics C*, 44(1):013001, Jan 2020.


