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

Zhen Wang,[1,2,3] Xuliang Zhu,[1,2] Elham E Khoda,[4] Shih-Chieh Hsu,[4] Nikolaos Konstantinidis,[5] Ke Li,[4] Shu Li,[1,2,6,7] Michael J. Ramsey-Musolf,[1,2,8,9] Yanda Wu,[1,2] and Yuwen E. Zhang[5]

[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, $h \to ss$, 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 fb^{-1} e^+e^- collision data at $\sqrt{s} = 240$ 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 ration with 95% CL. The ratio of these limits over the SM production cross-section is estimated to be around 5×10^{-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 $\sqrt{s} = 240$ 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]

$$\begin{aligned} V = & -\mu^2|H|^2 + \lambda|H|^4 + \frac{1}{2}a_1|H|^2S + \frac{1}{2}a_2|H|^2S^2 \\ & + b_1S + \frac{1}{2}b_2S^2 + \frac{1}{3}b_3S^3 + \frac{1}{4}b_4S^4. \end{aligned} \tag{1}$$

After EWSB, the two scalar fields can be parametrized as

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

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:

$$\begin{aligned} h_1 &= h\cos\theta + s\sin\theta, \\ h_2 &= -h\sin\theta + s\cos\theta. \end{aligned} \tag{3}$$

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

$$V \supset \frac{1}{6}\lambda_{111}h_1^3 + \frac{1}{2}\lambda_{211}h_2h_1^2 + \frac{1}{2}\lambda_{221}h_2^2h_1 + \frac{1}{6}\lambda_{222}h_2^3, \tag{4}$$

where λ_{211} governs the exotic Higgs decay. This cubic coupling's value is

$$\lambda_{211} = 2s^2cb_3 + \frac{a_1}{2}c\left(c^2 - 2s^2\right) + \left(2c^2 - s^2\right)sva_2 - 6\lambda sc^2v, \tag{5}$$

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

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

$$\begin{aligned} \Gamma(h_1) &= \cos^2\theta\,\Gamma(h_2), \\ \Gamma(h_2) &= \sin^2\theta\,\Gamma^{\text{SM}}\Big|_{m_2} + \Gamma(h_2 \to h_1h_1), \end{aligned} \tag{6}$$

where the exotic decay partial width is

$$\Gamma(h_2 \to h_1h_1) = \frac{1}{32\pi^2 m_2}\lambda_{211}^2\sqrt{1 - \frac{4m_1^2}{m_2^2}}. \tag{7}$$

As shown in [13], only m_1, $\cos\theta$, a_2, b_3, and b_4 are free parameters in this model. We take $m_1 \in [5, 60]$ 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 $\mathcal{O}(0.01)$ 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 GeV during an SFOEWPT scnario [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 exotics 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@NLOat 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 $E_{\text{cone}}^2 < 4E_\ell + 12.2\,\text{GeV}$, where E_ℓ is the lepton energy, and E_{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^- or $\mu^+\mu^-$). Each electron and muon must have energy higher than 20 GeV. Furthermore, the two leptons are required to have $|\cos\theta_{e^+e^-}| < 0.71$ or $|\cos\theta_{\mu^+\mu^-}| < 0.81$. 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 $\cos\phi_{e^+e^-} > -0.93$ $\cos\phi_{\mu^+\mu^-} > -0.93$. To cut more backgrounds, we require the invariant mass of the lepton system ($M_{\ell\bar{\ell}}$) 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 ($M_{\ell\bar{\ell}}^{\text{recoil}}$) is required to be in the range of $M_{\ell\bar{\ell}^{\text{recoil}}} \in [124, 140]$ GeV, where the recoil mass is defined in equation (8). The selections are summarized in Table 1.

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

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.

Variable	Selection Criteria
Lepton p_T	$p_T^{\text{lep}} > 20\,\text{GeV}$
Lepton angle	$\lvert\cos\theta_{e^+e^-(\mu^+\mu^-)}\rvert < 0.71\ (0.81)$
Track angle	$\cos\phi > -0.74$
Dilepton mass	$M_{\ell\bar{\ell}} \in [77.5, 104.5]$ GeV
Recoil mass	$M_{\text{recoil}}^{\ell\bar{\ell}} \in [124, 140]$ GeV

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

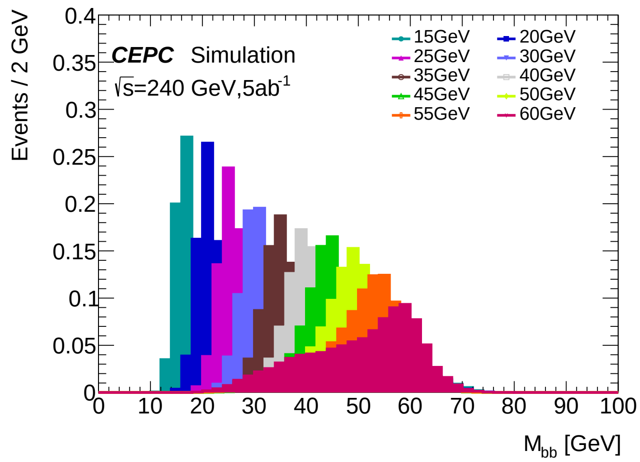

FIGURE 1: Reconstructed invariant mass distributions (M_{bb}) of the di b-jet system after pairing the jets. Histograms in different colors correspond to different signal samples whose mass range from 15 GeV to 60 GeV.

The background contributions are divided into three categories: Higgs processes with $b\bar{b}$ final state ($\ell\ell Hbb$), other Higgs processes ($\ell\ell H$), and non-Higgs processes. Since the $H \to b\bar{b}$ 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_s. 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-likeness value for each jet. The b-likeness (L_b) value represents the probability of a jet being a b-jet and the values lie between 0 and 1. The b-likenesses (L_{bi}) of the individual jets are used to compute a combined b-likeness or b-jet efficiency (f_b), defined as

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

We further define the b-jet inefficiency factor as

$$b_{\text{ineff}} = \log(1 - f_b). \quad (10)$$

In this analysis, b_{ineff} is used to enhance the rejection of non-b-jet backgrounds. The signal process has four b-quarks in final

M_s [GeV]	N_{leaves}	$\text{Data}_{\text{min}}^{\text{leaf}}$	f_{bagging}	f_{feature}	λ_{L1}	λ_{L2}
15	11	52	$4.62E-01$	$7.32E-01$	$2.10E-02$	$1.13E-07$
20	250	23	$5.68E-01$	$6.21E-01$	$1.02E-01$	$3.74E-01$
25	13	59	$7.06E-01$	$6.95E-01$	$9.24E-06$	$1.21E-04$
30	3	11	$5.21E-01$	$5.33E-01$	$2.54E-08$	$1.11E+00$
35	4	53	$8.79E-01$	$8.85E-01$	$1.59E-04$	$2.00E-05$
40	215	17	$8.75E-01$	$6.60E-01$	$2.56E-02$	$3.02E-04$
45	98	39	$8.11E-01$	$6.03E-01$	$1.29E-08$	$5.41E-05$
50	185	77	$6.80E-01$	$6.34E-01$	$7.62E-06$	$6.78E-05$
55	110	92	$9.81E-01$	$5.62E-01$	$7.18E-07$	$4.65E-01$
60	12	78	$7.02E-01$	$8.18E-01$	$3.68E-05$	$5.19E-04$

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

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 approach by training a BDT to enhance signal sensitivity. The BDT is implemented using the LIGHTGBM [31] python package 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 several widely used tools rely on depth-wise tree growth, LIGHTGBM 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 sensitivity of signal samples at every mass point, 10 models are trained with different signal samples with M_s in the range of 15 GeV to 60 GeV; a separate model is trained for each mass point. The model hyperparameters are optimized using Optuna [32] hyperparameter optimization framework. The best configuration for each model is different and they are summarized in Table 2. A total of 100 trials are used to tune the hyperparameters 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 descrbied in Table 3. Kinematic variables like p_T of leading 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 training. 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_{diff}^{ss}) is also used as a training variable. Apart from the other kinematic observable, the number of particles used to reconstruct the four jets of an event N_{particle}^{4j} 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 $|\cos\theta_{\text{j0,j2}}^{\text{Helicity}}|$ are used because the other two jets have the same value. The BDT also used the b_{ineff} to further increase the power of the model to separate signal and backgrounds.

Variable	Description
p_T^{l1}, p_T^{l2}	Lepton p_T
$M_{\ell\ell}^{\text{recoil}}$	Recoil mass of the dilepton system
$E^{\text{j0}}, E^{\text{j1}}, E^{\text{j2}}, E^{\text{j3}}$	Energy of each reconstructed jet
M^{4j}	Invariant mass of the four-jet system
M_{4j}^{recoil}	Recoil mass of the four-jet system
M_{bb}	Average mass of the two reconstructed s candidates
M_{diff}^{ss}	Difference between the two reconstructed s particles
N_{particle}^{4j}	Number of particles inside the four jets
$y_{12}, y_{23}, y_{34}, y_{45}, y_{56}, y_{67}$	Distance between jet constituents
$\cos\theta_{\text{boost}}^{s_1}, \cos\theta_{\text{boost}}^{s_2}$	Boosted angle between two jets from each s candidate
$\cos\theta_s^{\text{open}}$	Opening angle between the reconstructed s particles
$\lvert\cos\theta_{\text{j0}}^{\text{Helicity}}\rvert, \lvert\cos\theta_{\text{j2}}^{\text{Helicity}}\rvert$	Helicity of the first and the third jets
b_{ineff}	b-jet inefficiency

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.

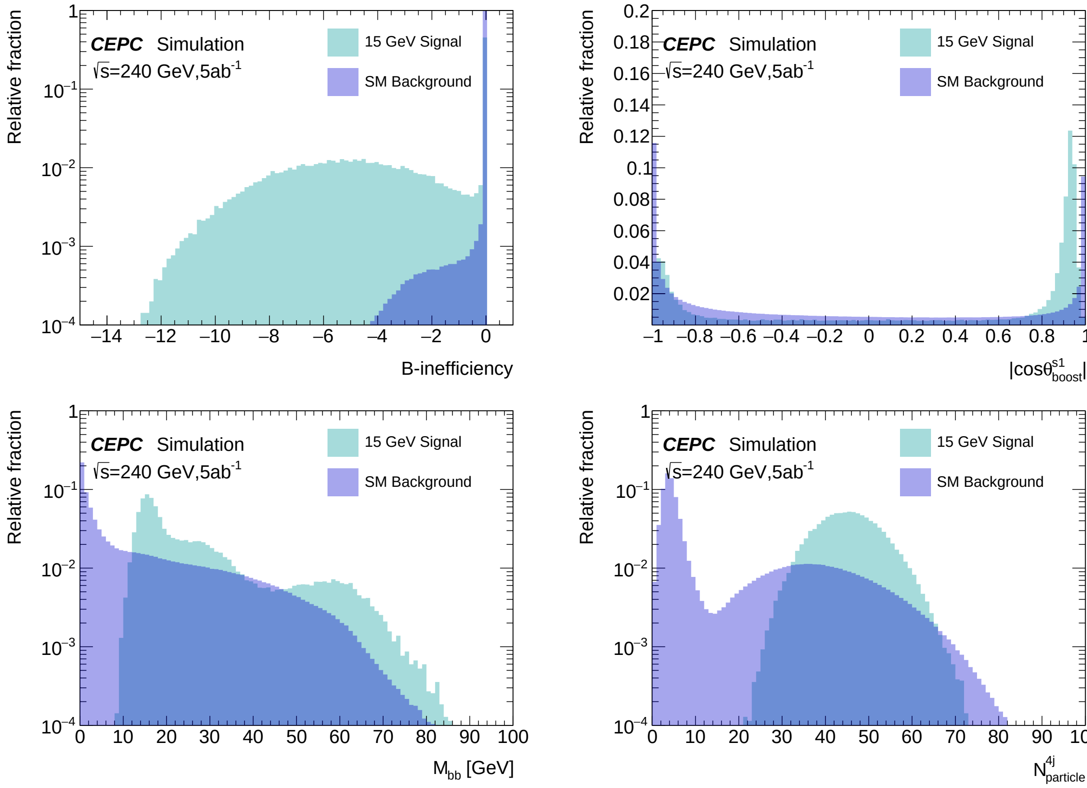

FIGURE 2: Signal and background distributions of four (out of 24) variables used for the BDT training. The signal histogram is from 15 GeV singlet MC samples.

Figure 2 shows four variables, with large signal to background discrimination power, used for the BDT training

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 $ZZ \to q\bar{q} + \mu^+\mu^-$ 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:

$$\mathrm{JER} = \sqrt{\sigma^2_{\mathrm{FastJet-TruthJet}} + \sigma^2_{\mathrm{FastJet-LCFIPlus}}}, \quad (11)$$

where σ is the jet resolution uncertainty. The estimated JER uncertainty is shown in Figure 3 as a function of truth jet energy.

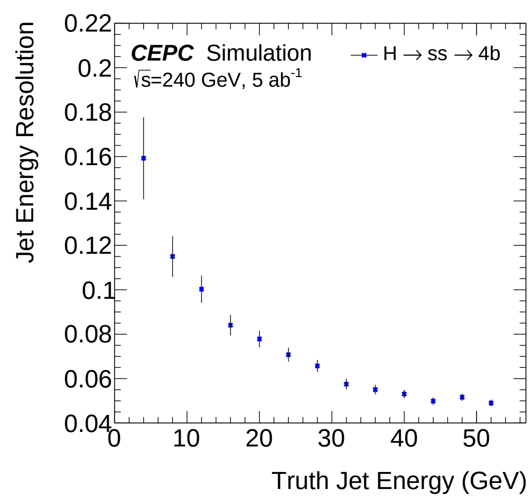

FIGURE 3: The jet energy resolution (equation (11)) as a function of truth jet energy. The error bars represent statistical errors.

Out of all the three different sources of uncertainties, the jet energy resolution uncertainty tends to be negligible after the statistical fit. The post-ft flavor tagging uncertainty is $\sim$1%. The background modeling uncertainty due to limited yields of Higgs process is around $\sim$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.

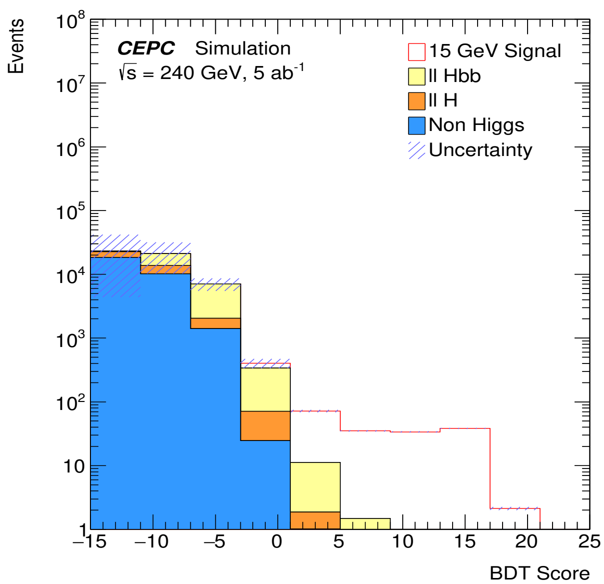

FIGURE 4: BDT score distribution for the combined electron and muon channel, along with an uncertainty band that includes both statistical and systematic uncertainties. The signal scores are shown in red histogram. All the histograms are scaled to $5\,\text{ab}^{-1}$ luminosity.

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 $\times$ branching ration ($\sigma_{ZH} \times \text{B}(H \to ss)$) at 95% confidence level (CL) using the CL$_s$ 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 ($\sigma_{ZH} \times \text{B}(H \to ss)$) and SM production cross section (σ_{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 15 GeV. Due to limited signal simulations, our results are not extended below 15 GeV.

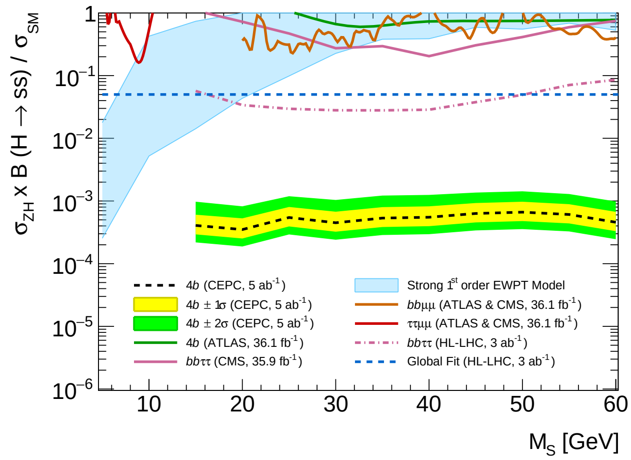

FIGURE 5: The ratio of expected upper limits at 95% CL on the $\sigma_{ZH} \times \text{B}(H \to ss)$ and SM cross section (σ_{SM}) as a function of singlet mass (M_s). The expected limits obtained from the BDT analysis are represented by the black dotted line with 1σ (yellow) and 2σ (green) uncertainty bands. The blue-shaded region in the plot indicates points that predict a strong first-order electroweak phase transition with successful tunneling, as determined by numerical scans [13]. The plot also displays the current (solid) and projected (dashed) sensitivities of the LHC in the $\tau\tau\mu\mu$ (red) [40, 41], $bb\mu\mu$ (orange) [42, 43], $bb\tau\tau$ (pink) [44, 45], and $4b$ (green) [46] channels. For the $bb\mu\mu$ and $\tau\tau\mu\mu$ channels, the best limit between ATLAS and CMS is shown at each mass point. Additionally, the HL-LHC projection for the indirect limit on the total exotic branching fraction is displayed as the blue dashed line [47].

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 $\sim$0.002 using a cut-and-count analysis. Cut-based approach is studied and optimized with "$n-1$" method but it is unnecessary to develop selections based on 10 signal samples. We decide to release selections and use 10 BDT classifiers to deal with 10 signal samples. BDT has explored more power in variables such as b_{ineff} and therefore improves the limits further by a factor ranging from 2$\sim$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, $H \to ss$, where each s-boson 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-

Expected 95% CL upper limits	
M_s [GeV]	$\sigma_{ZH} \times \mathrm{B}(H \to ss)/\sigma_{\mathrm{SM}}$
15	4.07×10^{-4}
20	3.50×10^{-4}
25	5.45×10^{-4}
30	4.49×10^{-4}
35	5.34×10^{-4}
40	5.50×10^{-4}
45	6.32×10^{-4}
50	6.63×10^{-4}
55	6.09×10^{-4}
60	4.58×10^{-4}

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^-H and muon $\mu^+\mu^-H$ channels using the multivariate BDT approach.

ing 5000 fb^{-1} e^+e^- collision data at $\sqrt{s} = 240$ GeV in CEPC. An LGBM-based BDT is used to discriminate signal events from the backgrounds, and the BDT score is used as the final discriminant for the statistical analysis. We improved the sensitivity by a factor ranging from 2$\sim$4 by using the BDT-based approach compared to the standard cut-and-count analysis. In the end, we set the upper limit on the ZH production cross-section times branching ratio of the decay $H \to ss$ with 95% CL using the 4b final state. The estimated limits rangess from 4.07×10^{-4} for $M_s = 15$ GeV to 4.58×10^{-4} for $M_s = 60$ GeV. We want to highlight that this realistic study yields a weaker exclusion limit compared to the projected values in [13]. Hence, it demonstrates the importance of carrying out such studies with a dedicated simulation using realistic detector characteristics and experimental uncertainties. The study with 4b final state at the CEPC could conclusively test the possibility of an SFOEWPT in the extended SM with a light singlet of mass as low as 15 GeV. This CEPC simulation study shows significantly higher sensitivity than that can be achieved at the LHC at this low mass range.

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. 19Z103010239. Ke Li, Elham E. Khoda, and Shih-Chieh Hsu are supported by the National Science Foundation (NSF) under grant no. 2110963.

References

[1] David E. Morrissey and Michael J Ramsey-Musolf. Electroweak baryogenesis. *New Journal of Physics*, 14(12):125003, Dec 2012.

[2] K. Kajantie, M. Laine, K. Rummukainen, and M. Shaposhnikov. The electroweak phase transition: a non-perturbative analysis. *Nuclear Physics B*, 466(1):189–258, 1996.

[3] K. Kajantie, M. Laine, K. Rummukainen, and M. Shaposhnikov. A non-perturbative analysis of the finite-t phase transition in $SU(2) \times U(1)$ electroweak theory. *Nuclear Physics B*, 493(1):413–438, 1997.

[4] K. Kajantie, M. Laine, K. Rummukainen, and M. Shaposhnikov. Is there a hot electroweak phase transition at $m_H \gtrsim m_W$? *Phys. Rev. Lett.*, 77:2887–2890, Sep 1996.

[5] F. Csikor, Z. Fodor, and J. Heitger. End point of the hot electroweak phase transition. *Phys. Rev. Lett.*, 82:21–24, Jan 1999.

[6] Dietrich Bodeker and Wilfried Buchmuller. Baryogenesis from the weak scale to the grand unification scale. *Rev. Mod. Phys.*, 93(3):035004, 2021.

[7] Qing-Hong Cao, Katsuya Hashino, Xu-Xiang Li, and Jiang-Hao Yue. Multi-step phase transition and gravitational wave from general $\mathbb{Z}_2$ scalar extensions. *arXiv preprint arXiv:2212.07756*, 12, 2022.

[8] Chiara Caprini, Mark Hindmarsh, Stephan Huber, Thomas Konstandin, Jonathan Kozaczuk, Germano Nardini, Jose Miguel No, Antoine Petiteau, Pedro Schwaller, Géraldine Servant, and David J. Weir. Science with the space-based interferometer eLISA. II: gravitational waves from cosmological phase transitions. *Journal of Cosmology and Astroparticle Physics*, 2016(04):001, Apr 2016.

[9] Stefano Profumo, Michael J. Ramsey-Musolf, and Gabe Shaughnessy. Singlet Higgs phenomenology and the electroweak phase transition. *JHEP*, 08:010, 2007.

[10] Cheng-Wei Chiang and Bo-Qiang Lu. First-order electroweak phase transition in a complex singlet model with $\mathbb{Z}_3$ symmetry. *JHEP*, 07:082, 2020.

[11] Wei Chao, Gui-Jun Ding, Xiao-Gang He, and Michael Ramsey-Musolf. Scalar Electroweak Multiplet Dark Matter. *JHEP*, 08:058, 2019.

[12] Michael J. Ramsey-Musolf. The electroweak phase transition: a collider target. *JHEP*, 09:179, 2020.

[13] Jonathan Kozaczuk, Michael J. Ramsey-Musolf, and Jessie Shelton. Exotic higgs boson decays and the electroweak phase transition. *Physical Review D*, 101(11), Jun 2020.

[14] Marcela Carena, Zhen Liu, and Yikun Wang. Electroweak phase transition with spontaneous Z_2-breaking. *JHEP*, 08:107, 2020.

[15] Stefano Profumo, Michael J Ramsey-Musolf, and Gabe Shaughnessy. Singlet higgs phenomenology and the electroweak phase transition. *Journal of High Energy Physics*, 2007(08):010, Aug 2007.

[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.

[17] Marcela Carena, Zhen Liu, and Yikun Wang. Electroweak phase transition with spontaneous z2-breaking. *Journal of High Energy Physics*, 2020(8), Aug 2020.

[18] The ILD Concept Group. The international large detector: Letter of intent, 2010.

[19] Ties Behnke, James E. Brau, Philip N. Burrows, Juan Fuster, Michael Peskin, Marcel Stanitzki, Yasuhiro Sugimoto, Sakue Yamada, and Hitoshi Yamamoto. The international linear collider technical design report—volume 4: Detectors, 2013.

[20] The CEPC Study Group. CEPC conceptual design report: Volume 2 - physics & amp; detector, 2018.

[21] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[22] Wolfgang Kilian, Thorsten Ohl, and Jurgen Reuter. WHIZARD: Simulating Multi-Particle Processes at LHC and ILC. *Eur. Phys. J. C*, 71:1742, 2011.

[23] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. Feynrules 2.0—A complete toolbox for tree-level phenomenology. *Computer Physics Communications*, 185(8):2250–2300, Aug 2014.

[24] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[25] P. Mora de Freitas and H. Videau. Detector simulation with MOKKA / GEANT4: Present and future. In *International Workshop on Linear Colliders (LCWS 2002)*, pages 623–627, 8, 2002.

[26] GEANT4 Collaboration. GEANT4—A simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[27] Manqi Ruan et al. Reconstruction of physics objects at the Circular Electron Positron Collider with Arbor. *Eur. Phys. J. C*, 78(5):426, 2018.

[28] Dan Yu, Manqi Ruan, Vincent Boudry, and Henri Videau. Lepton identification at particle flow oriented detector for the future e^+e^- Higgs factories. *Eur. Phys. J. C*, 77(9):591, 2017.

[29] Taikan Suehara and Tomohiko Tanabe. LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies. *Nucl. Instrum. Meth. A*, 808:109–116, 2016.

[30] S. Catani, Yuri L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber. New clustering algorithm for multi - jet cross-sections in e^+e^- annihilation. *Phys. Lett. B*, 269:432–438, 1991.

[31] Guolin Ke, Qiwei Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weiye Ma, and Qiwei Ye. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30:3146–3154, 2017.

[32] Takuya Akiba, Shotaro Sano, Takeru Yanase, Daitetsu Ohta, Shin Koyama, Toshihiko Nakashima, and Yukino Yanagisawa. Optuna: A next-generation hyperparameter optimization framework, 2019.

[33] Ivan Smiljanić, Ivanka Božović-Jelisavčić, Goran Kačarević, Nataša Vukašinović, Tatjana Agatonović-Jovin, Gordana Milutinović-Dumbelović, Jasna Stevanović, and Mirko Radulović. Integrated luminosity measurement at CEPC. In *International Workshop on Future Linear Colliders*, 5, 2021.

[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 $h \to b\bar{b}/c\bar{c}/gg$ in associated $(e^+e^-/\mu^+\mu^-)h$ production at the CEPC. *Chinese Physics C*, 44(1):013001, Jan 2020.

[35] Matteo Cacciari and Gavin P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett. B*, 641:57–61, 2006.

[36] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2):1554, 2011.

[37] A. L. Read. Presentation of search results: the cls technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, Sep 2002.

[38] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 434(2):435–443, 1999.

[39] 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. Study of electroweak phase transition in exotic higgs decays at the CEPC, 2022.

[40] ATLAS Collaboration. Search for Higgs bosons decaying to aa in the $\mu\mu\tau\tau$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS experiment. *Phys. Rev. D*, 92(5):052002, 2015.

[41] CMS Collaboration. Search for light pseudoscalar boson pairs produced from decays of the 125 GeV Higgs boson in final states with two muons and two nearby tracks in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 800:135087, 2020.

[42] ATLAS Collaboration. Search for Higgs boson decays into a pair of light bosons in the $bb\mu\mu$ final state in pp collision at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B*, 790:1–21, 2019.

[43] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two muons and two b quarks in pp collisions at 13 TeV. *Phys. Lett. B*, 795:398–423, 2019.

[44] CMS Collaboration. Projection of searches for exotic Higgs boson decays to light pseudoscalars for the High-Luminosity LHC. Technical report, CERN, Geneva, 2019.

[45] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two b quarks and two τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 785:462, 2018.

[46] ATLAS Collaboration. Search for the Higgs boson produced in association with a vector boson and decaying into two spin-zero particles in the $H \to aa \to 4b$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:031, 2018.

[47] HL/HE WG2 Collaboration. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:221–584, 2019.