

Optimisation of the vertex detector and measurement of Higgs decays to second-generation quarks at the CEPC

Jialin Li ^{a,b}, Hao Liang ^c, Kaili Zhang ^{d,h}, Yifan Zhu ^f, Jun Guo ^{a,b,*},
 Haijun Yang ^{b,g} and Manqi Ruan ^{d,e,*}

^aNational Key Laboratory of Dark Matter Physics, School of Physics and Astronomy,
 Shanghai Jiao Tong University,
 Shanghai 200240, China

^bKey Laboratory for Particle Astrophysics and Cosmology (KLPPAC-MoE),
 Shanghai Key Laboratory for Particle Physics and Cosmology (SKLPPC),
 Shanghai 200240, China

^cLaboratoire Leprince-Ringuet, CNRS, Ecole polytechnique, Institut Polytechnique de Paris,
 91120 Palaiseau, France

^dInstitute of High Energy Physics, Chinese Academy of Sciences,
 Beijing 100049, China

^eSchool of Physical Sciences, University of Chinese Academy of Sciences (UCAS),
 Beijing 100049, China

^fNational Key Laboratory of Dark Matter Physics, Tsung-Dao Lee Institute and School of Physics
 and Astronomy, Shanghai Jiao Tong University,
 Shanghai 201210, China

^gNational Key Laboratory of Dark Matter Physics, School of Physics and Astronomy and Tsung-Dao
 Lee Institute, Shanghai Jiao Tong University,
 Shanghai 200240, China

^hChina Spallation Neutron Source Science Center,
 Dongguan 523803, China

E-mail: lijialin@sjtu.edu.cn, hao.liang@llr.in2p3.fr, zhangkl@ihep.ac.cn,
zyJonah@sjtu.edu.cn, jun.guo@sjtu.edu.cn, haijun.yang@sjtu.edu.cn,
ruanmq@ihep.ac.cn

ABSTRACT: The vertex detector is crucial for precision measurements of the Higgs boson at the electron-positron Higgs factory. Benchmarked with $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$ measurements in the $\nu\bar{\nu}H$ channel, we perform an optimisation study on the inner radius and spatial resolution of the vertex detector using the Jet Origin Identification (JOI) framework, which determines the parton flavour of jets using advanced Artificial Intelligence (AI) algorithm. We observe that, compared to the reference detector configuration, halving the inner radius and spatial resolution improves the transverse and longitudinal impact parameter resolution

*Corresponding author.

approximately by a factor of two, while increasing the accuracy and significance of the $H \rightarrow c\bar{c}/s\bar{s}$ measurement by 4% and 8%, respectively. Conversely, doubling these parameters results in comparable degradation, with variations in the inner radius being the dominant factor. Our results provide guidance for detector design and highlight promising prospects for identifying the $H \rightarrow s\bar{s}$ decay mode at future Higgs factories.

KEYWORDS: e^+e^- Experiments, Higgs Physics, Jets

ARXIV EPRINT: [2508.04191](https://arxiv.org/abs/2508.04191)

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | CEPC vertex detector and samples | 2 |
| 3 | Flavour tagging algorithms and performance under inner radius/spatial resolution scans | 4 |
| 3.1 | Performance of track reconstruction | 4 |
| 3.2 | Performance of jet origin identification | 6 |
| 3.3 | Performance of higgs benchmarks: $\nu\bar{\nu}H$ channel, $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$ measurements | 8 |
| 4 | Summary and discussion | 10 |
| A | Comparison between Delphes and Geant4 | 12 |
| B | Simulation study of the beam-induced background impact | 13 |
| C | Event-level distributions of the discriminant S_{ff} | 13 |

1 Introduction

Electron–positron colliders provide uniquely clean environments that facilitate precision studies of the Higgs boson properties, crucial for probing physics beyond the Standard Model (SM) [1]. Among proposed future facilities, the Circular Electron–Positron Collider (CEPC) is designed with a large circumference of 100 km and two interaction points [2], and will operate at various centre-of-mass energies. According to its Technical Design Report (TDR) [3, 4], the CEPC is planned to accumulate an integrated luminosity of 20 ab^{-1} at 240 GeV as a Higgs factory [5], yielding 4×10^6 Higgs bosons and enabling unparalleled precision in Higgs coupling measurements. In addition, it is capable of running at the Z-pole (91.2 GeV), the W^+W^- threshold (160 GeV), and could potentially be upgraded to reach the $t\bar{t}$ threshold at 360 GeV.

Measurements of the $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$ directly test the Yukawa couplings between the Higgs and the second-generation quarks. The SM predicts the $H \rightarrow c\bar{c}/s\bar{s}$ has a branching ratio of $0.029/2.3 \times 10^{-4}$ [6–8]. Notably, these correspond to about 120k $H \rightarrow c\bar{c}$ events and about 1k $H \rightarrow s\bar{s}$ events at the CEPC when considering all Higgs production modes. Among these, approximately one quarter are produced in the $\nu\bar{\nu}H$ final state from ZH ($Z \rightarrow \nu\bar{\nu}$) and WW -fusion processes. This channel offers a particularly clean environment in which all visible particles originate from the Higgs decay, and therefore provides a convenient benchmark for the simulation study used in this work. With the combination of high-performance detectors and advanced particle reconstruction algorithms, the $H \rightarrow c\bar{c}$ is expected to be measured to a percent-level of accuracy, while discovering the $H \rightarrow s\bar{s}$ decay mode is a critical and promising possibility at the future Higgs factory.

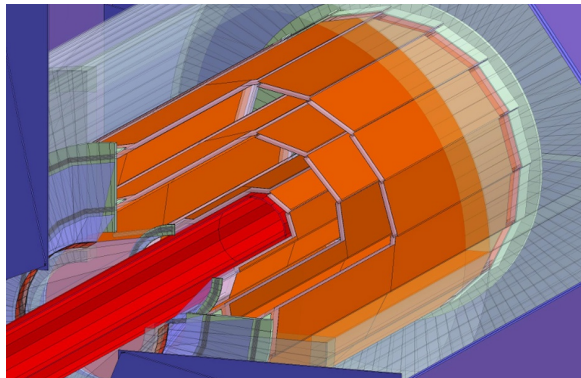


Figure 1. Schematic view of pixel detector. Two layers of silicon pixel sensors are mounted on both sides of each of the three ladders to provide six space points. Only the silicon sensor sensitive region (in orange) is depicted. The vertex detector surrounds the beam pipe (red).

The vertex detector is essential for Jet Origin Identification (JOI) [9], which determines the colored parton from which the jet originates, and thus plays a pivotal role in rare Higgs decay measurements. The characterization of the vertex detector largely depends on its geometric configuration, particularly the innermost radius and spatial resolution. In this work, we quantify the relevant detector metrics such as track impact parameter resolution and JOI performance. Furthermore, we evaluate the precision for $H \rightarrow c\bar{c}$ and significance for $H \rightarrow s\bar{s}$ as functions of these detector parameters, explicitly establishing how detector optimisation translates into physics performance.

This work presents a conceptual detector optimization study in the context of AI-enhanced reconstruction. Focusing on the vertex detector geometry, we evaluate performance at three levels — the track reconstruction, JOI, and physics benchmark processes — to provide a coherent view of how geometry optimization propagates from low-level reconstruction to physics sensitivity.

This paper is structured as follows. In section 2, we introduce the CEPC vertex detector concept, the fast simulation frameworks, the Monte Carlo samples employed. Section 3 evaluates detector-level performance: we first quantify how variations in the vertex detector’s inner radius and spatial resolution affect track impact-parameter resolution (section 3.1), then assess their influence on JOI via migration matrix metrics (section 3.2). We translate these detector and algorithmic results into physics impact by studying the statistical precision of $H \rightarrow c\bar{c}$ and the significance of $H \rightarrow s\bar{s}$ (section 3.3). Section 4 summarizes our findings, discusses limitations of the current study, and outlines avenues for further optimisation.

2 CEPC vertex detector and samples

The CEPC baseline detector adopts a Particle Flow Oriented design [11], emphasizing the separation and reconstruction of every final state particle, and measures them in the most suited sub-detectors. A crucial component of the CEPC detector is its vertex system. As referenced in the CEPC Conceptual Design Report (CDR) [10], the vertex system consists of six concentric cylindrical layers of silicon pixel sensors (figure 1). The first barrel layer is positioned merely 16 mm from the beam line, and with a per-layer material budget of only

| | R (mm) | $ z $ (mm) | $ \cos\theta $ | σ (μm) |
|---------|----------|------------|----------------|----------------------------|
| Layer 1 | 16 | 62.5 | 0.97 | 2.8 |
| Layer 2 | 18 | 62.5 | 0.96 | 6 |
| Layer 3 | 37 | 125.0 | 0.96 | 4 |
| Layer 4 | 39 | 125.0 | 0.95 | 4 |
| Layer 5 | 58 | 125.0 | 0.91 | 4 |
| Layer 6 | 60 | 125.0 | 0.90 | 4 |

Table 1. Baseline design parameters of the CEPC vertex detector. R denotes the radial position of each layer with respect to the beam line, $|z|$ the half length of the layer, $|\cos\theta|$ the corresponding polar-angle acceptance, and σ the spatial resolution of the silicon pixel sensors.

0.15% X_0 , it attains a spatial resolution of $2.8\ \mu\text{m}$ for high precision flavour tagging. The baseline geometrical parameters and spatial resolutions of the vertex detector layers used in this study are summarised in table 1. The beam pipe dominates the upstream material of the vertex detector. In the CEPC TDR design [30], the Machine Detector Interface(MDI) region is explicitly modelled, including a 0.35 mm beryllium beam pipe with a $10\ \mu\text{m}$ inner gold coating and a 0.2 mm cooling-water layer, corresponding in total to about 0.45% X_0 . Each vertex layer contributes approximately 0.06–0.28% X_0 , dominated by the sensors and local supports.

The beam-induced background is also highly relevant to the design and operation of the vertex detector, as its hits can dominate the occupancy and bandwidth of the detectors. The CEPC TDR provides a dedicated simulation of the beam-induced background, showing that the impact on the impact-parameter resolution is under control. Further details are provided in appendix B. Hence, the beam-induced background is neglected in this study.

The simulated samples correspond to the $\nu\bar{\nu}H$ final state, including both the ZH process with $Z \rightarrow \nu\bar{\nu}$ and the WW -fusion contribution ($e^+e^- \rightarrow \nu_e\bar{\nu}_eH$), corresponding to about 0.9×10^6 Higgs bosons at an integrated luminosity of $20\ \text{ab}^{-1}$. We use WHIZARD 1.95 [12] and PYTHIA 6.4 [13] as the event generators and DELPHES [14] for the detector simulation. The reconstructed final-state particles are clustered into two jets via the $e^+e^-k_t$ algorithm [15, 16], and JOI is performed to calculate likelihoods of 11 jet categories for jet-flavour tagging. The Delphes-based results are cross-checked against a GEANT4 based full simulation, demonstrating good agreement and robustness. The details can be found in the appendix A. An idealised particle identification (PID) model is adopted to study the effects of vertex detector parameters in isolation, without introducing additional dependencies from PID modelling. At the Z-pole operation, the kaon identification efficiency reach about 95–98% with the CEPC CDR baseline detector [31, 32]. The recent study further shows that, with the incorporation of AI-based techniques, a kaon identification efficiency of about 98.5% can be achieved in $\nu\bar{\nu}H$ samples at $\sqrt{s} = 240\ \text{GeV}$ [24]. This supports the use of idealised PID within the conceptual scope of the present work.

Recent international design studies for future e^+e^- colliders converged on similar requirements for the vertex detector geometry. The ECFA study [17] recommends an innermost layer radius down to 16 mm and a spatial resolution of about $3\ \mu\text{m}$, with the material

budget per layer kept around $0.1\%X_0$. Comparable specifications are adopted in the baseline designs of ILC [18], CLIC [19], and FCC-ee [20], targeting optimal flavour tagging and impact parameter resolution.

Motivated by the conclusions of the international studies, we perform parameter scans for the inner radii of the first two vertex layers and the spatial resolution, exploring scaling factors from 0.5 to 2 relative to our baseline detector configuration. This range covers both ambitious future upgrades and more conservative scenarios. The scaling factors are defined as

$$R_{\text{rad}} \equiv \frac{r_{\text{inner}}}{r_{\text{baseline}}}, \quad R_{\text{res}} \equiv \frac{\sigma_{\text{hit}}}{\sigma_{\text{baseline}}}, \quad (2.1)$$

where r_{inner} and σ_{hit} are the innermost layer radius and spatial resolution. We vary these scaling factors from 0.5 to 2 relative to the baseline detector configuration. Then we assess their impacts at multiple levels:

- At the track reconstruction level, we quantify the impact parameter resolution via the root-mean-square (RMS) of its distribution, directly reflecting the detector’s ability to localize particle trajectories;
- At the JOI performance level, we evaluate JOI using migration matrices for different configurations, and use their traces to characterize overall flavour-tagging efficiency;
- At the higgs benchmark level, we study the efficiency \times purity for $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$, and quantify the resulting statistical precision for $H \rightarrow c\bar{c}$ and signal significance for $H \rightarrow s\bar{s}$.

For each Higgs decay channel ($H \rightarrow b\bar{b}$, $c\bar{c}$, $s\bar{s}$, $u\bar{u}$, $d\bar{d}$ and $H \rightarrow gg$) of $\nu\bar{\nu}H$, we generate one million events to ensure sufficient statistical precision for jet flavour tagging and detector performance studies.

3 Flavour tagging algorithms and performance under inner radius/spatial resolution scans

This section presents a comprehensive evaluation of flavour tagging performance as a function of vertex detector configuration. We first investigate the impact parameter resolution at the track level, followed by the JOI performance under different detector setups. Finally, we connect these detector-level metrics to physics benchmarks, quantifying their influence on $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$ analyses.

3.1 Performance of track reconstruction

We quantify the sensitivity of impact parameter resolution to vertex geometry by evaluating the RMS of the Δd_0 and Δz_0 distributions, where Δd_0 and Δz_0 denote the difference between reconstructed and true transverse and longitudinal impact parameters, respectively, as shown in figure 2.

The RMS of Δd_0 and Δz_0 depends almost linearly on both the radial radius R_{rad} and the spatial resolution R_{res} , with R_{rad} playing the dominant role. Doubling R_{rad} roughly doubles the impact-parameter resolution, whereas the variations in R_{res} change it by only about 10%.

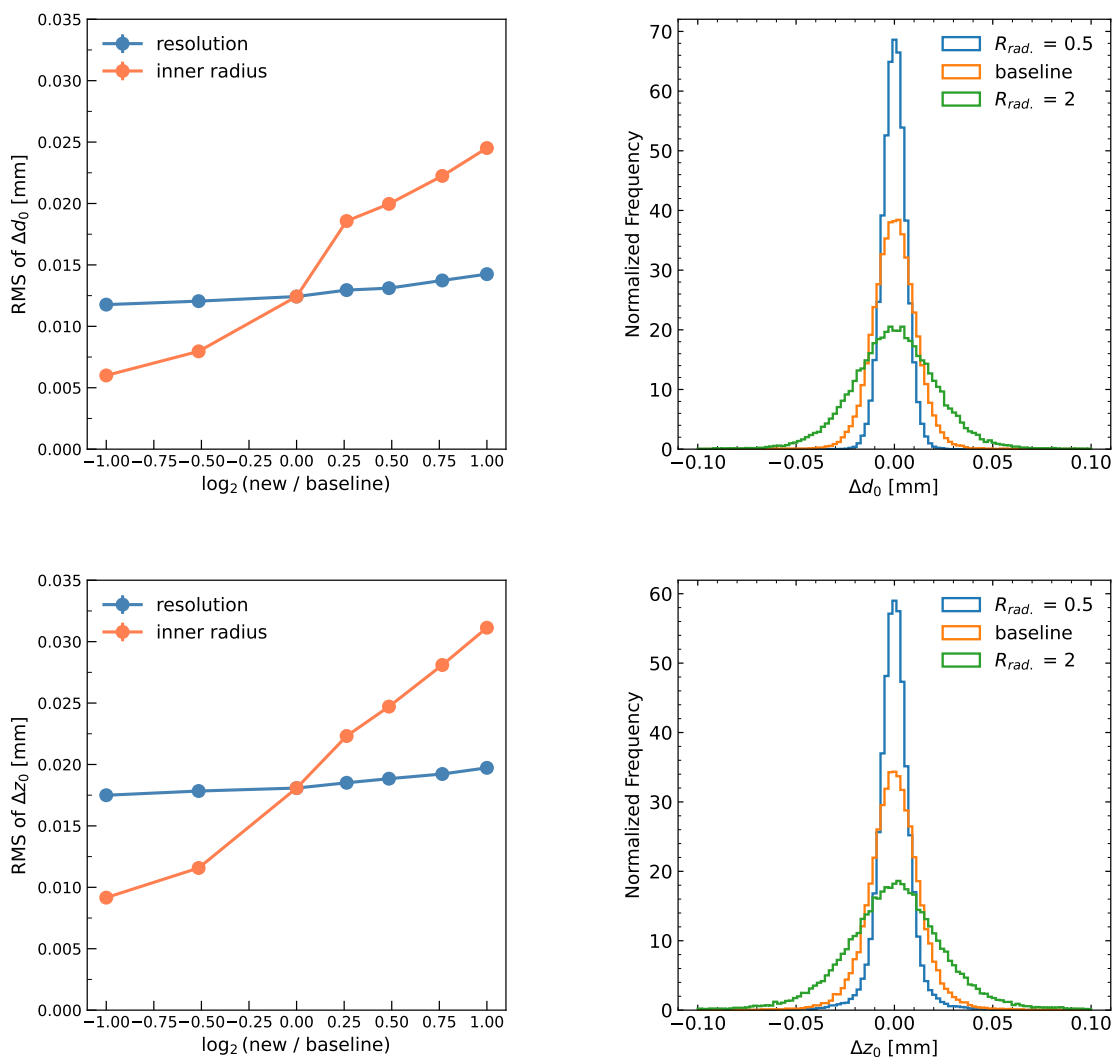


Figure 2. Left: RMS of Δd_0 (top) and Δz_0 (bottom) versus geometry scale $\log_2(\text{new}/\text{baseline})$. Right: normalised Δd_0 (top) and Δz_0 (bottom) distributions for $R_{\text{rad}} = 0.5$, baseline, and $R_{\text{rad}} = 2$, respectively. The simulated samples are $H \rightarrow b\bar{b}$.

As shown in the right panel of figure 2, reducing R_{rad} narrows the distributions of Δd_0 and Δz_0 (improving the longitudinal resolution by roughly 40%), while increasing R_{rad} broadens them. Thus, the detector’s inner radius is the principal factor in determining impact-parameter performance, and spatial resolution plays only a secondary role in the baseline design.

We also validated our fast simulation results against full GEANT4-based simulations. A detailed comparison of the impact parameter resolution is provided in appendix A. The agreement, as shown in figure 8, demonstrates the robustness of our fast simulation framework.

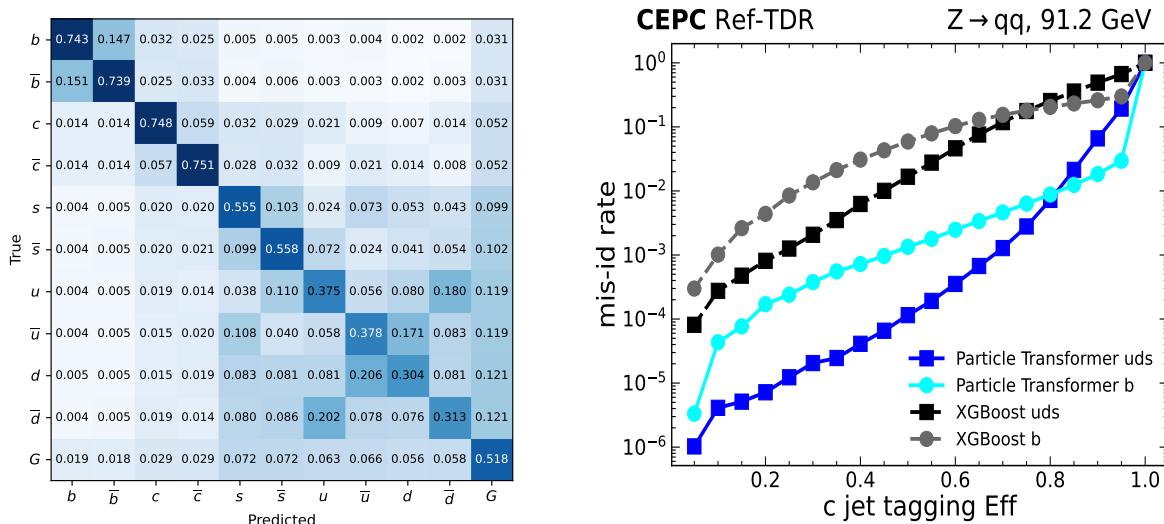


Figure 3. Left: the migration matrix of 11-dimensional jet identification of JOI under the CEPC CDR configuration. The matrix is normalized to unity for each truth label (row). Right: the comparison of JOI and XGboost of the c -jet tagging efficiency. [22]

3.2 Performance of jet origin identification

The JOI based on the ParticleNet architecture [21] is designed to determine the coloured particle from which a jet originates. For each jet, the algorithm takes the kinematic variables of the reconstructed final-state particles, their particle-type information, the impact-parameter measurements of charged tracks, their energy and momentum distributions, as inputs. It then computes likelihood scores for eleven jet-flavour categories ($b, \bar{b}, c, \bar{c}, s, \bar{s}, u, \bar{u}, d, \bar{d}$, and gluon), which are subsequently employed to construct flavour-tagging observables.

Figure 3 shows the performance of the JOI algorithm. The left-hand panel displays the normalized 11-dimensional jet-flavour migration matrix (M_{11}), obtained by assigning each jet to the flavour category with the highest score, thus giving an overview of the tagging and misidentification pattern among all flavours. The right-hand panel provides a complementary flavour-specific view by presenting the c -jet tagging performance, obtained from scanning working points with different c -jet efficiencies. For each chosen c -jet efficiency, the corresponding misidentification rates of the other flavour jets are shown. The c -tagging efficiency is defined as the fraction of true c -jets whose c -score exceeds the threshold, while the mis-ID rate is the fraction of non- c jets passing the same threshold. In the efficiency range of 0.2 to 0.8, the JOI achieves misidentification rates that are one to three orders of magnitude lower than those of the traditional multivariable analysis approach based on XGBoost [22].

From the migration matrices at varying R_{rad} as shown in figure 4, we extract key flavour-tagging metrics as shown on the left-hand side of figure 5. As R_{rad} increases from 0.5 to 2, the $b \rightarrow c, g \rightarrow c$ and $c \rightarrow s$ misidentification rate rises from about 4% to 7%. Meanwhile, the $c \rightarrow c$ tagging efficiency drops from 81% to 70%, and $s \rightarrow s$ from 60% to 54%. While the $g \rightarrow s$ rate remains stable around 14%. The trends indicate that reducing the vertex detector’s innermost radius significantly lowers the $b \rightarrow s$ and $c \rightarrow s$ misidentification rates, reflecting the crucial role of the vertex in distinguishing heavy-flavour jets from strange jets.

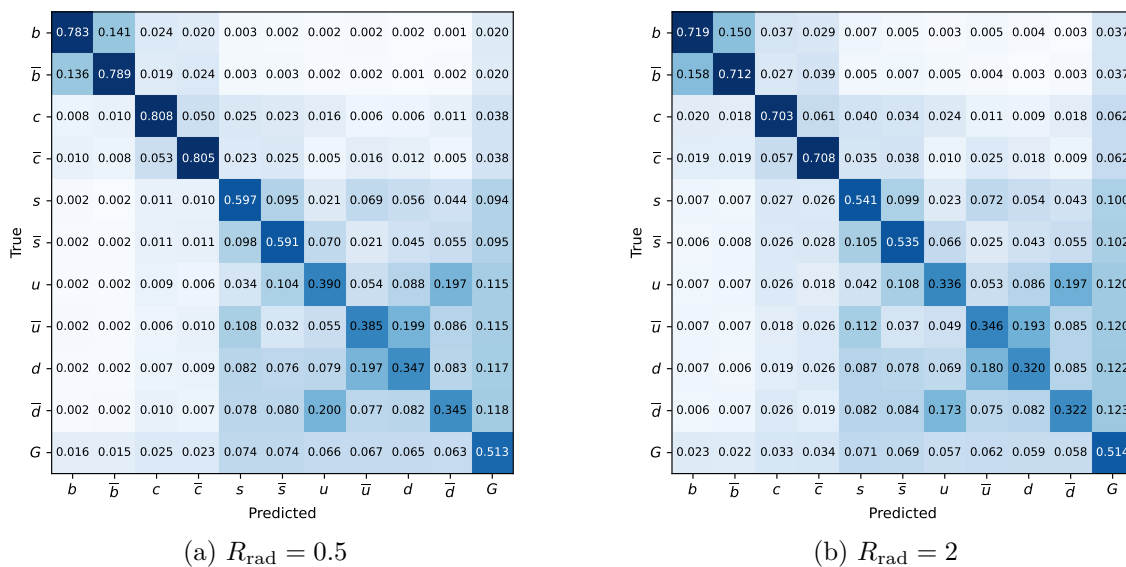


Figure 4. The migration matrix of M_{11} with $R_{rad.} = 0.5$ (a) and $R_{rad.} = 2$ (b). The matrix is normalized to unity for each truth label.

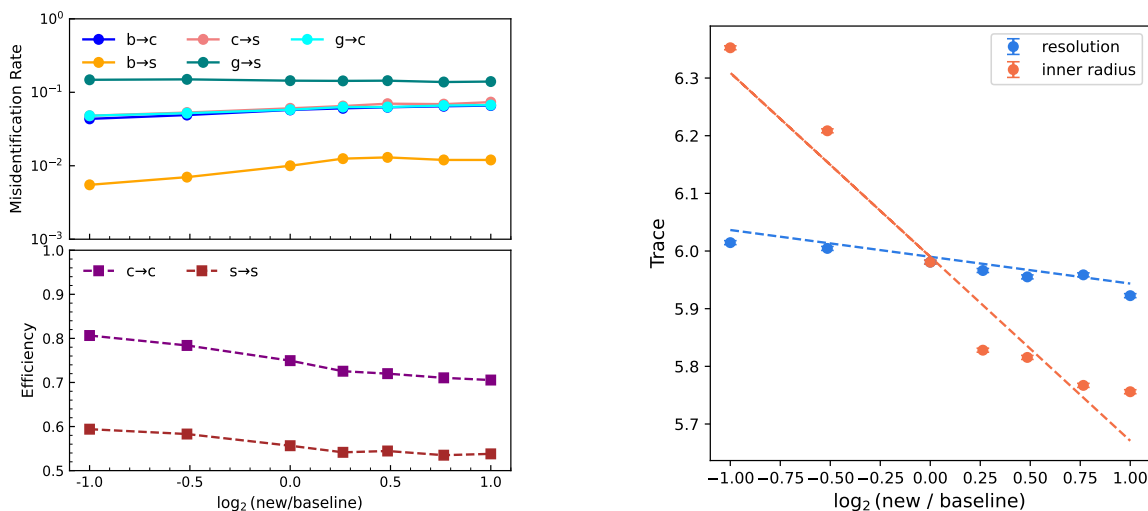


Figure 5. Left: key misidentification rates and tagging efficiencies versus $\log_2(\text{new}/\text{baseline})$ for different inner radius configurations. Right: trace of the JOI migration matrix versus $\log_2(\text{new}/\text{baseline})$ for different inner radius and resolution configurations. Orange points indicate scans of the inner radius, while blue points indicate scans of the spatial resolution. Dashed lines represent linear fits to the configurations, respectively.

In contrast, the impact of vertex detector optimisation on gluon-strange discrimination is limited as gluon jets predominantly originate from the primary vertex, and vertex detector upgrades have little effect on their identification.

To further probe the global performance of JOI, we characterize its behaviour via the trace of M_{11} , denoted as $\text{Tr}(M)$. This trace, which quantifies the total correct classification probability across all quark flavour categories, is analysed as a function of the geometry

scaling factors $\log_2(\text{new}/\text{baseline})$ for both R_{rad} and R_{res} . As shown on the right-hand side of figure 5, a linear regression to the $\text{Tr}(M)$ yields:

$$\text{Tr}(M) = 5.98 - 0.27 \log_2 R_{\text{rad}} - 0.05 \log_2 R_{\text{res}}, \quad (3.1)$$

demonstrating that variations in the inner radius impact JOI performance by a factor of five more strongly than equivalent changes in spatial resolution.

3.3 Performance of higgs benchmarks: $\nu\bar{\nu}H$ channel, $H \rightarrow c\bar{c}$ and $H \rightarrow s\bar{s}$ measurements

The JOI algorithm provides, for each jet j , a set of posterior probabilities $P_{\text{jet}}(f|j)$ that the jet originated from parton flavour f . To translate these per-jet probabilities into event-level information for $H \rightarrow jj$ decays, we construct a Bayesian likelihood of the Higgs decay hypothesis $H \rightarrow f\bar{f}$,

$$L_{f\bar{f}}(j_1, j_2) = \eta_{f\bar{f}} \mathcal{B}(H \rightarrow f\bar{f}) [P_{\text{jet}}(f|j_1)P_{\text{jet}}(\bar{f}|j_2) + P_{\text{jet}}(f|j_2)P_{\text{jet}}(\bar{f}|j_1)], \quad (3.2)$$

where $\eta_{f\bar{f}} = \frac{1}{2}$. For quark final states, the factor of $\frac{1}{2}$ arises because the correction to the branching ratio — where the first jet is f and the second jet is \bar{f} — accounts for only half of $\mathcal{B}(f\bar{f})$. This factor of $\frac{1}{2}$ also appears for gg final states, cancelling out the over-counting due to identical final states. Normalising over all decay hypotheses gives the event-level posterior

$$S_{f\bar{f}}(j_1, j_2) = \frac{L_{f\bar{f}}(j_1, j_2)}{\sum_{f'} L_{f'\bar{f}'}(j_1, j_2)}, \quad (3.3)$$

with $S_{f\bar{f}} \in [0, 1]$ and $\sum_f S_{f\bar{f}} = 1$. The corresponding $S_{f\bar{f}}$ distributions at event-level are shown in appendix C. For a given physics measurement, the candidate $H \rightarrow f\bar{f}$ signal events are defined as those satisfying $S_{f\bar{f}}$ over the threshold, where the threshold score is scanned to optimise the product $\epsilon \times P$, to balance the signal retention and background suppression, and thus the optimal measurement sensitivity. The *efficiency* ϵ is defined as the fraction of true $H \rightarrow f\bar{f}$ signal events whose $S_{f\bar{f}}$ value lies above the chosen threshold relative to the total number of simulated signal events. The *purity* P is the fraction of selected events that truly originate from the $H \rightarrow f\bar{f}$ decay, i.e. the ratio of correctly identified signal events to all selected events including background contributions.

Using the event-level posterior defined in eq. (3.3), the purity P , efficiency ϵ , and their product $\epsilon \times P$ are evaluated as functions of the discriminant score $S_{f\bar{f}}$. Since previous studies indicate that the impact of the spatial resolution on the flavour-tagging performance is small, figure 6 shows results only for three representative values of the vertex inner radius, with all other parameters fixed to the baseline values. For $H \rightarrow c\bar{c}$, the large branching ratio yields a high signal rate and good background separation over a broad range of $S_{c\bar{c}}$. Both the efficiency and purity remain high for moderately tight selections, resulting in a broad maximum of $\epsilon \times P$. In contrast, for $H \rightarrow s\bar{s}$, the maximum of $\epsilon \times P$ is reached at relatively loose selections. Although the purity improves with increasing $S_{s\bar{s}}$, the rapid loss of signal efficiency dominates the behaviour of $\epsilon \times P$, making the optimal working point primarily driven by signal efficiency.

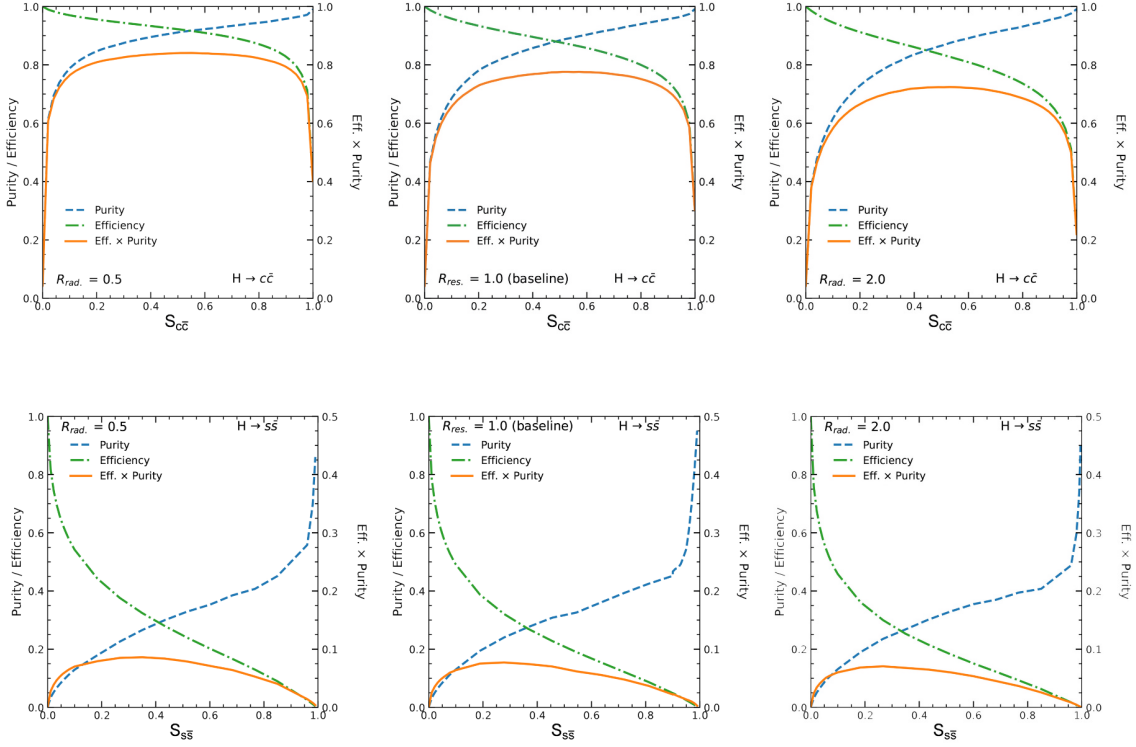


Figure 6. Purity P , efficiency ϵ , and $\epsilon \times P$ as functions of the discriminant score $S_{f\bar{f}}$ ($S_{c\bar{c}}$ for $H \rightarrow c\bar{c}$ and $S_{s\bar{s}}$ for $H \rightarrow s\bar{s}$), shown for $H \rightarrow c\bar{c}$ (top panel) and $H \rightarrow s\bar{s}$ (bottom panel). The results are presented for $R_{\text{rad}} = 0.5$ (left), baseline (middle), and $R_{\text{rad}} = 2$ (right). Only the inner radius is varied with respect to the baseline vertex detector configuration, while all other parameters remain unchanged. Note that for $H \rightarrow s\bar{s}$, the vertical axis for $\epsilon \times P$ is shown in the range $[0, 0.5]$. The curves are obtained by scanning the $S_{f\bar{f}}$ selection threshold.

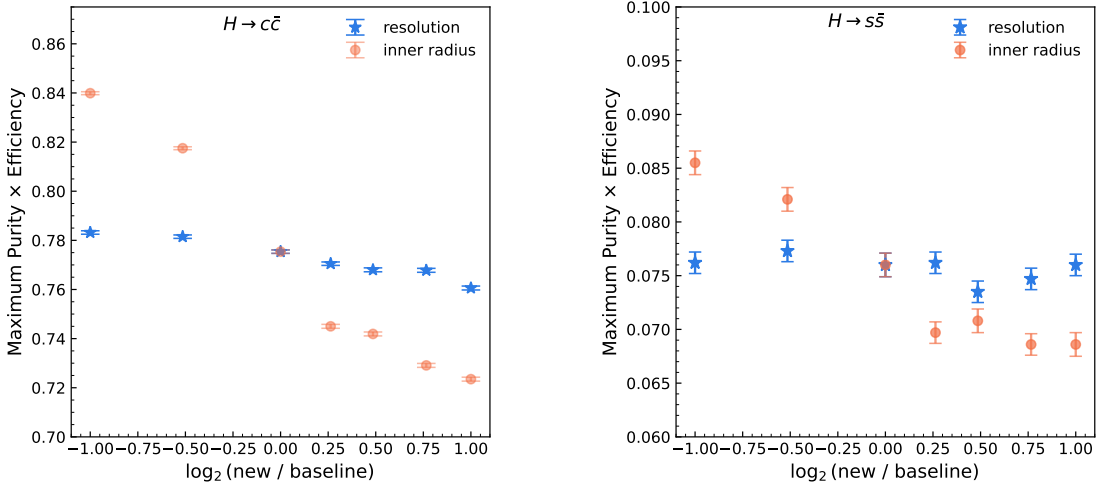


Figure 7. Maximum of $\epsilon \times P$ for $H \rightarrow c\bar{c}$ (left) and $H \rightarrow s\bar{s}$ (right). Error bars represent statistical uncertainty.

| Configuration | $c\bar{c}$ (10^3) | $b\bar{b}$ (10^3) | $s\bar{s}$ | gg (10^3) | Uncertainty (10^{-3}) |
|------------------------|-----------------------|-----------------------|------------|-----------------|---------------------------|
| baseline | 24.4 | 1.26 | 11.0 | 1.92 | 6.81 |
| $R_{\text{res}} = 0.5$ | 24.5 | 1.21 | 8.95 | 1.85 | 6.78 |
| $R_{\text{res}} = 2$ | 24.0 | 1.21 | 12.9 | 1.99 | 6.88 |
| $R_{\text{rad}} = 0.5$ | 25.6 | 0.75 | 5.54 | 1.75 | 6.55 |
| $R_{\text{rad}} = 2$ | 23.2 | 1.54 | 1.61 | 2.04 | 7.05 |

Table 2. Event yields at the optimal cut (maximizing $\varepsilon \times P$) and statistical uncertainty for $H \rightarrow c\bar{c}$, normalized to an integrated luminosity of 20 ab^{-1} in the $\nu\bar{\nu}H$ channel, corresponding to about 0.9×10^6 Higgs bosons. Columns list the signal ($c\bar{c}$), other Higgs decay backgrounds ($b\bar{b}$, $s\bar{s}$, gg), and the relative statistical uncertainty.

The maximum of $\varepsilon \times P$ quantifies analysis quality, which is summarized in figure 7. In $H \rightarrow c\bar{c}$, the maximum of $\varepsilon \times P$ scales linearly with R_{rad} , shifting by 8% per factor-of-two radius change; $H \rightarrow s\bar{s}$ exhibits similar trends but with larger statistical fluctuations (due to lower signal yield), with the maximum of $\varepsilon \times P$ varying by 12% under the same radius scans. These results reinforce the innermost vertex radius as the primary key point for optimising both decay channels, with pixel resolution showing a marginal impact.

We compute the relative statistical uncertainty for $H \rightarrow c\bar{c}$ and signal significance for $H \rightarrow s\bar{s}$ using the asymptotic approximation for Asimov data. The relative statistical uncertainty for $H \rightarrow c\bar{c}$ is estimated as eq. (3.4).

$$\frac{\delta\mathcal{B}}{\mathcal{B}} = \frac{\sqrt{s+b}}{s}. \tag{3.4}$$

where s and b denote signal and background yields (extracted from score-based selections). The significance follows eq. (3.5) [23].

$$S = \sqrt{2 \left((s+b) \ln \left(1 + \frac{s}{b} \right) - s \right)} \tag{3.5}$$

where S is the significance.

We evaluate $H \rightarrow c\bar{c}$ yields and statistical uncertainty at the optimal $\varepsilon \times P$ cut (table 2). Halving the inner radius (R_{rad}) boosts $c\bar{c}$ signal, suppresses $b\bar{b}$ background, then lowers uncertainty by approximately 4%, while spatial resolution (R_{res}) has minimal impact on precision. For $H \rightarrow s\bar{s}$ summarised in table 3, the JOI combined with normalized score cuts fully suppresses $b\bar{b}$ background ($b\bar{b}$ yields to 0), providing a critical advantage for signal isolation. Decreasing R_{rad} suppresses gg background and enhances significance by 8% compared to the baseline. Varying R_{res} causes negligible changes in signal/background yields and significance, indicating its marginal impact. These trends illustrate that the inner vertex radius dominates sensitivity for both decays, with resolution playing a minor role.

4 Summary and discussion

The vertex detector is crucial for high-precision Higgs measurements at future electron-positron Higgs factories, particularly impacting the sensitivity to rare decays such as $H \rightarrow c\bar{c}$

| Configuration | $s\bar{s}$ | $b\bar{b}$ | $c\bar{c}$ | gg | Significance |
|------------------------|------------|------------|------------|-------|---------------|
| baseline | 70.9 | 0 | 10.0 | 224.6 | 4.42 σ |
| $R_{\text{res}} = 0.5$ | 64.1 | 0 | 7.6 | 179.4 | 4.45 σ |
| $R_{\text{res}} = 2$ | 64.4 | 0 | 10.3 | 181.4 | 4.41 σ |
| $R_{\text{rad}} = 0.5$ | 70.2 | 0 | 6.2 | 189.4 | 4.76 σ |
| $R_{\text{rad}} = 2$ | 64.8 | 0 | 11.5 | 203.2 | 4.23 σ |

Table 3. Event yields at the optimal cut (maximizing $\varepsilon \times P$) and corresponding Asimov significance for $H \rightarrow s\bar{s}$, normalized to an integrated luminosity of 20 ab^{-1} in the $\nu\bar{\nu}H$ channel, corresponding to about 0.9×10^6 Higgs bosons. Columns list signal ($s\bar{s}$), other Higgs decay backgrounds ($b\bar{b}$, $c\bar{c}$, gg), and significance.

and $H \rightarrow s\bar{s}$. Precise measurements of these rare decays provide a decisive test of the Yukawa coupling predicted by the Standard Model.

This study, obtained using the AI-driven JOI framework which leverages jet momentum and impact parameters for effective signal discrimination [9], highlights the inner radius as the primary geometric parameter influencing the impact parameter resolution and subsequently flavour-tagging performance, while spatial resolution exerts a comparatively weaker influence. Benchmarking the $\nu\bar{\nu}H$ production channel, we project a statistical precision of 0.66–0.71% for $H \rightarrow c\bar{c}$, and a significance exceeding 4σ for $H \rightarrow s\bar{s}$. The measurement sensitivity is moderately affected by the vertex detector design: reducing the inner radius from the baseline leads to less than 10% improvements in both precision and significance, while increasing it has the opposite effect. In contrast, variations in spatial resolution have a much smaller impact, with only 1% changes observed across all benchmarks.

Our study is based on fast simulations and currently considers only the major backgrounds from other Higgs decay modes, representing an optimistic yet well-motivated setup. Comparisons with full (GEANT4-based) simulations confirm the reliability of our fast-simulation approach, with discrepancies at the 5% level. In a realistic environment, beam-induced and other Standard Model backgrounds would further dilute the sensitivity, and their quantitative impact, particularly the contributions related to the MDI design, still requires dedicated study. Consequently, the present results should be interpreted as upper-limit estimates achievable under idealised background conditions.

Our observations are consistent with previous studies [25, 26], which show that the inner radius of the vertex detector is the most sensitive geometric parameter for flavour tagging and Higgs precision measurements, underscoring its importance in future detector design. The search for $H \rightarrow s\bar{s}$ offers a unique probe of second-generation Yukawa couplings and is a major objective for next-generation Higgs factories. Recent global studies project measurement accuracies for this channel in the range of 17% to 140% [9, 27, 28], with the spread arising from differences in assumed luminosity, analysis strategies, background modelling, and flavour tagging performance.

Our study demonstrates that, within the scanned parameter space, the achievable sensitivity to $H \rightarrow s\bar{s}$ can vary by more than 12% due solely to vertex detector design. With

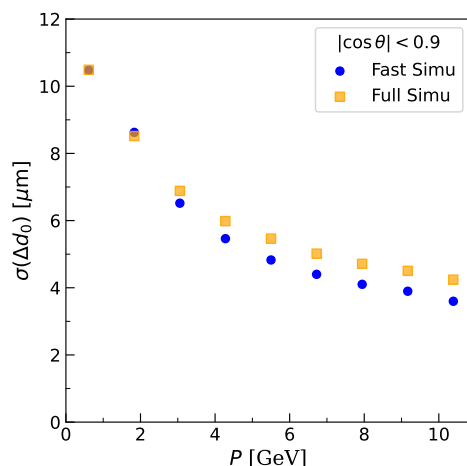


Figure 8. Comparison of the transverse impact parameter resolution $\sigma(\Delta d_0)$ as a function of track momentum P in the central region ($|\cos\theta| < 0.9$) of $H \rightarrow bb$ events between fast simulation (DELPHES) and full simulation (GEANT4).

an optimized inner radius, advances in detector design and analysis will further enhance the strange-quark Yukawa coupling reach. These improvements include better identification of hadrons containing s-quarks (such as K_S^0 and Λ [29]), optimizing operation scenarios to increase Higgs boson yields, refining and validating more precise hadronization models, enhancing the performance of flavour-tagging algorithms, leveraging advanced analysis techniques, and incorporating more Higgs production channels (l^+l^-H , $q\bar{q}H$). Collectively, these improvements could push the reach well beyond the current 4σ level, potentially enabling its first unambiguous observation and delivering a stringent test of second-generation Yukawa couplings.

Acknowledgments

We thank Yongfeng Zhu for providing the analysis tools and the useful discussions. This work was supported by National Key Program for Science and Technology Research and Development under Contract No. 2024YFA1610603 and the National Natural Science Foundation of China under Grant No.W2441004.

A Comparison between Delphes and Geant4

To validate the accuracy of our fast simulation approach, we compare its results with those from the full GEANT4-based simulation. Figure 8 shows this comparison for the transverse impact parameter (d_0) resolution dependence on track momentum (P). We observe that the difference between the fast simulation and the full simulation is less than 5%, confirming that the fast simulation provides a reliable approximation to the full simulation.

Figure 9 compares the M_{11} jet-flavour migration matrices obtained with the full GEANT4 simulation and the fast DELPHES simulation. For all quark flavours, the diagonal elements

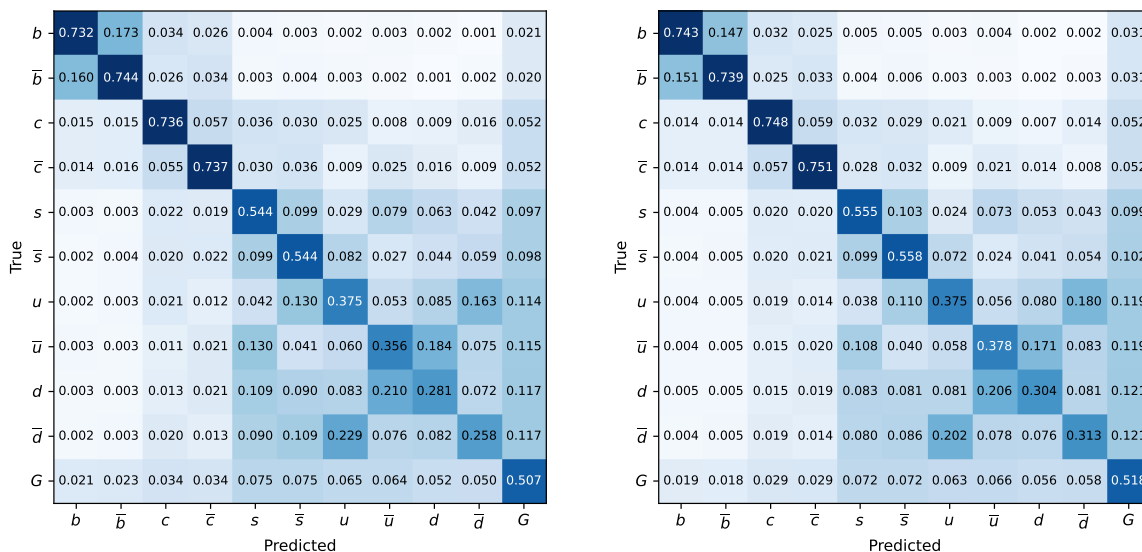


Figure 9. The migration matrix of M_{11} with full simulation by GEANT4 (left) and fast simulation DELPHES (right).

differ by less than 3%, and the off-diagonal mis-identification rates show no systematic bias. In particular, the difference of charm-jet and strange-jet tagging efficiency is within 2%. This close correspondence validates the use of the DELPHES framework for the vertex-detector optimisation presented in this work.

B Simulation study of the beam-induced background impact

The influence of beam-induced background on tracking performance has been qualitatively evaluated in the CEPC TDR [30]. As shown in figure 10, a direct comparison of the impact-parameter resolutions (d_0 and z_0) with and without beam-induced background overlay demonstrates no noticeable degradation across the momentum range. Therefore, the pixel size and spatial resolution variations studied in this work remain valid, and the overall conclusions on vertex detector optimization are not affected by beam background effects.

C Event-level distributions of the discriminant S_{ff}

To provide a more intuitive interpretation of the efficiency–purity curves shown in figure 6, we present in figure 11 the distributions of the event-level discriminant S_{ff} for the signal and the main Higgs decay backgrounds. The horizontal axis corresponds to the value of S_{ff} , while the vertical axis shows the number of events per bin. As discussed in the main text, varying the selection on S_{ff} corresponds to integrating these distributions above the chosen threshold. Tightening the selection requirement progressively shifts the selected events towards the right-hand side, while looser selections include an increasing fraction of events at small S_{ff} values.

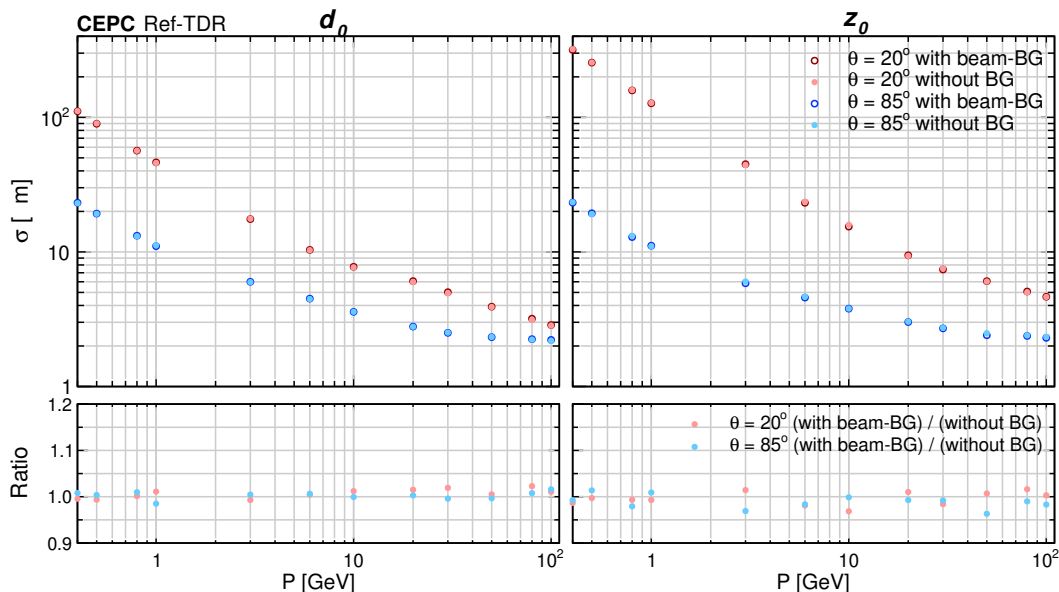


Figure 10. Comparison of the impact parameter resolutions (d_0 , z_0) with and without beam-induced background overlay at polar angles of 85° and 20° .

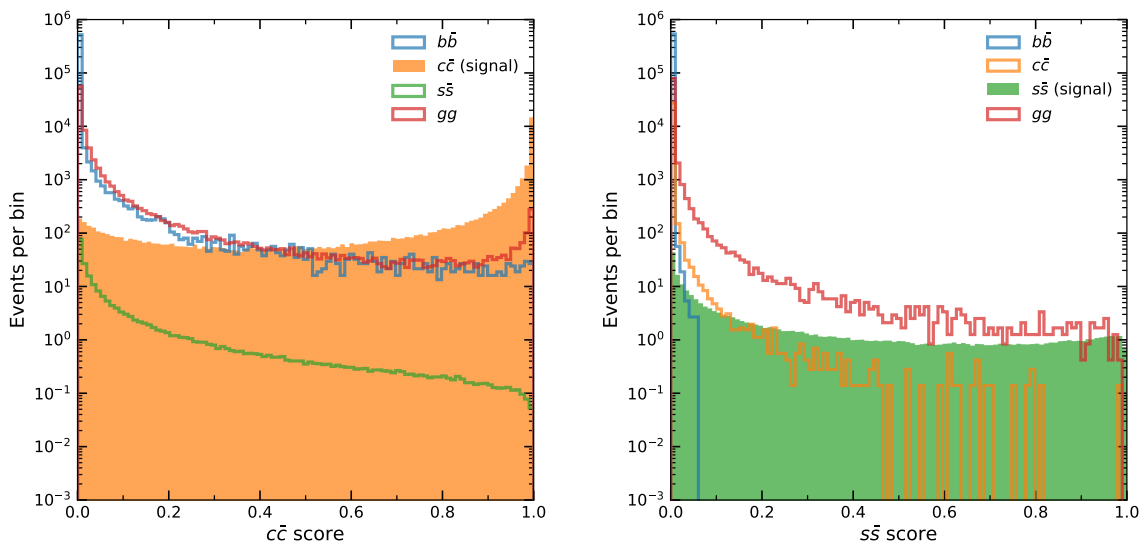


Figure 11. Distributions of the event-level discriminant S_{ff} for the baseline vertex detector configuration. Left: S_{cc} distribution for the $H \rightarrow c\bar{c}$. Right: S_{ss} distribution for the $H \rightarrow s\bar{s}$. The signal and the main Higgs decay backgrounds are shown separately. The vertical axis indicates the number of events per bin on a logarithmic scale.

Data Availability Statement. This article has no associated data or the data will not be deposited.

Code Availability Statement. This article has no associated code or the code will not be deposited.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] The European Strategy Group, *Deliberation document on the 2020 Update of the European Strategy for Particle Physics*, <https://cds.cern.ch/record/2720131>.
- [2] CEPC STUDY GROUP collaboration, *CEPC Conceptual Design Report: Volume 1 — Accelerator*, [arXiv:1809.00285](https://arxiv.org/abs/1809.00285) [[INSPIRE](#)].
- [3] CEPC STUDY GROUP collaboration, *CEPC Technical Design Report: Accelerator, Radiat. Detect. Technol. Methods* **8** (2024) 1 [[arXiv:2312.14363](https://arxiv.org/abs/2312.14363)] [[INSPIRE](#)].
- [4] CEPC PHYSICS STUDY GROUP collaboration, *The physics potential of the CEPC. Prepared for the US Snowmass Community Planning Exercise (Snowmass 2021)*, in the proceedings of the *Snowmass 2021*, Seattle, U.S.A., July 17–26 (2022) [[arXiv:2205.08553](https://arxiv.org/abs/2205.08553)] [[INSPIRE](#)].
- [5] F. An et al., *Precision Higgs physics at the CEPC*, *Chin. Phys. C* **43** (2019) 043002 [[arXiv:1810.09037](https://arxiv.org/abs/1810.09037)] [[INSPIRE](#)].
- [6] LHC HIGGS CROSS SECTION WORKING GROUP collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, *CERN Yellow Rep. Monogr.* **2** (2017) 1 [[arXiv:1610.07922](https://arxiv.org/abs/1610.07922)] [[INSPIRE](#)].
- [7] J. Duarte-Campderros, G. Perez, M. Schlaffer and A. Soffer, *Probing the Higgs-strange-quark coupling at e^+e^- colliders using light-jet flavor tagging*, *Phys. Rev. D* **101** (2020) 115005 [[arXiv:1811.09636](https://arxiv.org/abs/1811.09636)] [[INSPIRE](#)].
- [8] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01 [[INSPIRE](#)].
- [9] H. Liang et al., *Jet-Origin Identification and Its Application at an Electron-Positron Higgs Factory*, *Phys. Rev. Lett.* **132** (2024) 221802 [[arXiv:2310.03440](https://arxiv.org/abs/2310.03440)] [[INSPIRE](#)].
- [10] CEPC STUDY GROUP collaboration, *CEPC Conceptual Design Report: Volume 2 — Physics & Detector*, [arXiv:1811.10545](https://arxiv.org/abs/1811.10545) [[INSPIRE](#)].
- [11] M. Ruan et al., *Reconstruction of physics objects at the Circular Electron Positron Collider with Arbor*, *Eur. Phys. J. C* **78** (2018) 426 [[arXiv:1806.04879](https://arxiv.org/abs/1806.04879)] [[INSPIRE](#)].
- [12] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, *Eur. Phys. J. C* **71** (2011) 1742 [[arXiv:0708.4233](https://arxiv.org/abs/0708.4233)] [[INSPIRE](#)].
- [13] T. Sjostrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [[hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175)] [[INSPIRE](#)].
- [14] C. Chen et al., *Fast simulation of the CEPC detector with Delphes*, [arXiv:1712.09517](https://arxiv.org/abs/1712.09517) [[INSPIRE](#)].

- [15] T. Suehara and T. Tanabe, *LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies*, *Nucl. Instrum. Meth. A* **808** (2016) 109 [[arXiv:1506.08371](#)] [[INSPIRE](#)].
- [16] S. Catani et al., *New clustering algorithm for multi-jet cross-sections in e^+e^- annihilation*, *Phys. Lett. B* **269** (1991) 432 [[INSPIRE](#)].
- [17] T. Bergauer, *From ecfa detector roadmap to drd collaborations and beyond*, (2025), https://agenda.infn.it/event/44943/contributions/263377/attachments/137562/207083/2025-06-ESPP-Venice-Bergauer_v4b.pdf.
- [18] H. Abramowicz et al., *The International Linear Collider Technical Design Report — Volume 4: Detectors*, [arXiv:1306.6329](#) [[INSPIRE](#)].
- [19] A.C. Abusleme Hoffman et al., *Detector Technologies for CLIC*, [arXiv:1905.02520](#) [[DOI:10.23731/CYRM-2019-001](#)] [[INSPIRE](#)].
- [20] N. Barchetta, P. Collins and P. Riedler, *Tracking and vertex detectors at FCC-ee*, *Eur. Phys. J. Plus* **137** (2022) 231 [[arXiv:2112.13019](#)] [[INSPIRE](#)].
- [21] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019 [[arXiv:1902.08570](#)] [[INSPIRE](#)].
- [22] C. Zhang, *Global Physics Performance at the Ref-TDR*, (2025), <https://indico.ifae.es/event/2054/contributions/9525/attachments/2903/4524/20250617.pdf>.
- [23] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [*Erratum ibid.* **73** (2013) 2501] [[arXiv:1007.1727](#)] [[INSPIRE](#)].
- [24] Y. Wang et al., *One-to-one correspondence reconstruction at the electron-positron Higgs factory*, *Comput. Phys. Commun.* **314** (2025) 109661 [[arXiv:2411.06939](#)] [[INSPIRE](#)].
- [25] Z. Wu et al., *Study of vertex optimization at the CEPC*, [2018 JINST 13 T09002](#) [[INSPIRE](#)].
- [26] Y. Zhu et al., *ParticleNet and its application on CEPC jet flavor tagging*, *Eur. Phys. J. C* **84** (2024) 152 [[arXiv:2309.13231](#)] [[INSPIRE](#)].
- [27] Y. Zhu et al., *Holistic approach and Advanced Color Singlet Identification for physics measurements at high energy frontier*, [arXiv:2506.11783](#) [[INSPIRE](#)].
- [28] J. Altmann et al., *ECFA Higgs, electroweak, and top Factory Study*, [arXiv:2506.15390](#) [[DOI:10.23731/CYRM-2025-005](#)] [[INSPIRE](#)].
- [29] T. Zheng et al., *Reconstructing K_S^0 and Λ in the CEPC baseline detector*, *Eur. Phys. J. Plus* **135** (2020) 274 [[INSPIRE](#)].
- [30] CEPC STUDY GROUP collaboration, *CEPC Technical Design Report - Reference Detector*, [arXiv:2510.05260](#) [[INSPIRE](#)].
- [31] Y. Zhu, S. Chen, H. Cui and M. Ruan, *Requirement analysis for dE/dx measurement and PID performance at the CEPC baseline detector*, *Nucl. Instrum. Meth. A* **1047** (2023) 167835 [[arXiv:2209.14486](#)] [[INSPIRE](#)].
- [32] D. Yu et al., *Evaluation of PID Performance at CEPC and Optimization with Combined dN/dx and Time-of-Flight Data*, [arXiv:2507.18164](#) [[INSPIRE](#)].