中国科学院高能物理研究所
Institute of High Energy Physics Chinese Academy of Sciences

Measuring the Electron Yukawa Coupling at CEPC

Siya Feng, Hongbo Liao, Gang Li, Junyuan Chen

23 October 2026

Outline

1/ Physics Motivation

2/ Research Status

3/ Main Content

4/ Polarization Study

5/ Analysis Framework:
- Sample Generation
- Detector Simulation and Object Reconstruction
- Event Preselection
- Particle Transformer
- Simultaneous Fit

6/ Comparison with FCC-ee Results

7/ Summary and Next Steps

8/ References

Physics Motivation

- Higgs production mediated by first-generation fermion Yukawa couplings (u、d、e^-、ν_e) remains experimentally unverified.
- The exceedingly small electron Yukawa coupling leads to a dramatically low resonant Higgs production cross section, making its measurement extremely challenging.
- Hadron colliders cannot effectively measure the electron Yukawa coupling, as the signal is overwhelmed by an enormous background. A new experimental approach is therefore required.
- CEPC's clean collision environment and precisely controlled initial state provide an ideal experimental platform for precision electroweak measurements.

Physics Motivation

- This study probes Higgs production mediated by first-generation fundamental fermion Yukawa couplings.
- It provides a rigorous test of the electron Yukawa coupling — **the first such investigation at CEPC**.
- It explores electron-associated scalar physics beyond the Standard Model, with sensitivity to energy scales of approximately 100 TeV and beyond.
- It searches for the potential existence of new particles nearly degenerate in mass with the Higgs boson.

Research Status

Reference study (FCC-ee-based)	CMS link	ATLAS link	HL-LHC link	FCC-hh link
$\gamma_e \leq 1.6\gamma_e^{SM}$	$\gamma_e \leq 240\gamma_e^{SM}$	$\gamma_e \leq 260\gamma_e^{SM}$	$\gamma_e \leq 200\gamma_e^{SM}$	$\gamma_e \leq 50\gamma_e^{SM}$

Table 6 Individual significances (in std. deviations σ) expected per decay channel for s-channel Higgs boson production in e^+e^- collisions at FCC-ee for $\mathcal{L}_{\text{int}} = 10\,\text{ab}^{-1}$ and $\delta_{\sqrt{s}} = 4.1\,\text{MeV}$. The last column quotes the combined significance

$H \to gg$	$H \to WW^* \to$ $\ell\nu\,2j$; $2\ell\,2\nu$; $4j$	$H \to ZZ^* \to$ $2j\,2\nu$; $2\ell\,2j$; $2\ell\,2\nu$	$H \to b\bar{b}$	$H \to \tau_{\text{had}}\tau_{\text{had}}$; $c\bar{c}$; $\gamma\,\gamma$	Combined
1.1σ	$(0.53 \otimes 0.34 \otimes 0.13)\sigma$	$(0.32 \otimes 0.18 \otimes 0.05)\sigma$	0.13σ	$< 0.02\sigma$	1.3σ

Main Content

- Measure the s-channel resonant Higgs production cross section at the CEPC, and extract y_e via $\sigma \propto y_e^2$.

- Focus on the most promising signal channel, $e^+e^- \to H \to gg$, and its corresponding background, $e^+e^- \to q\bar{q}$.

- Since Reference study (FCC-ee-based assumptions) only relies on simple event pre-selection and BDTs to improve the S/B, we can build upon their approach by employing more advanced methods to further enhance the significance:
  - Exploit beam polarization to enhance the signal cross section.
  - Employ machine learning (Particle Transformer) for event selection optimization, combined with a simultaneous fit to further enhance the signal sensitivity.

Polarization Study

Assess the feasibility of enhancing the S/B via polarization:

1. **Monte Carlo simulation — longitudinal polarization:**

- Keep the degree of polarization the same

  - electron —— left-handed polarization, positron —— left-handed ~
  - electron —— left-handed polarization, positron —— right-handed ~
  - electron —— right-handed polarization, positron —— left-handed ~
  - electron —— right-handed polarization, positron —— right-handed ~

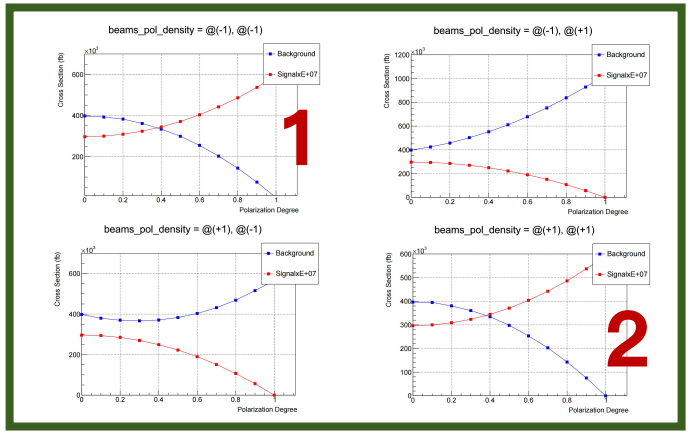
  Only change the electron's degree of polarization, keep $P_{e^+} = 0$.

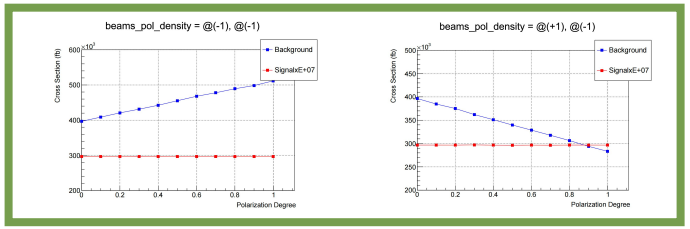
  - electron —— left-handed polarization
  - electron —— right-handed polarization

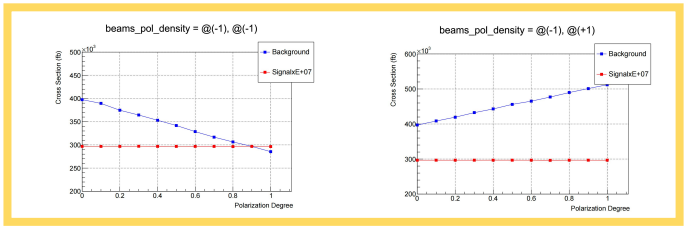
  Only change the positron's degree of polarization, keep $P_{e^-} = 0$.

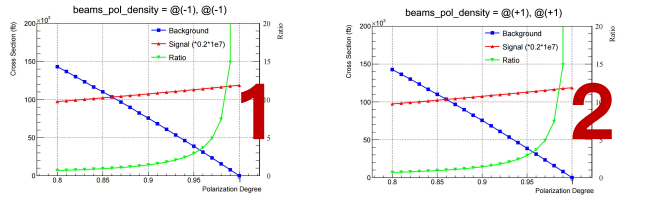
  - positron —— left-handed polarization
  - positron —— right-handed polarization

- The scan results show that longitudinal polarization requires an excessively high polarization degree to reach the desired S/B. This approach is therefore **discarded**.









Polarization Study

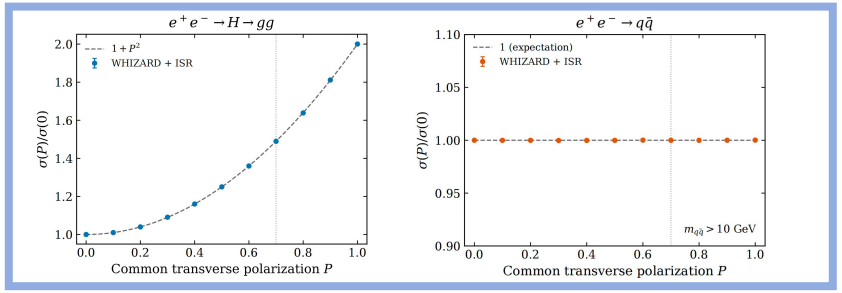
2. Monte Carlo simulation — transverse polarization

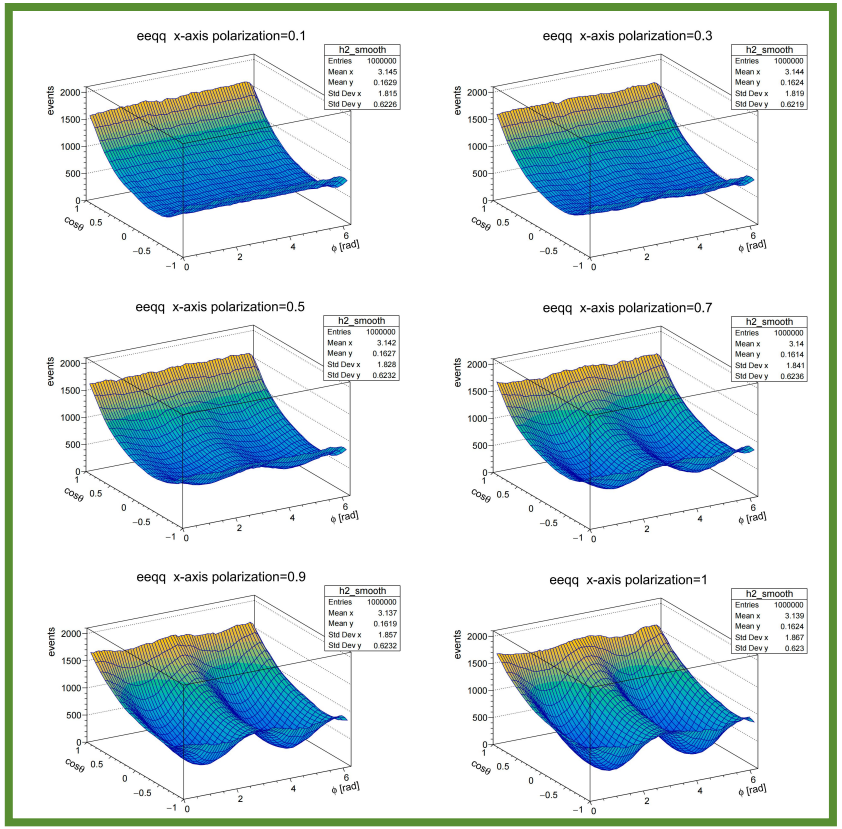
- The polarization direction serves as the primary discriminant: the signal cross section varies with it, whereas the background remains constant. Maximal signal enhancement is achieved with electron polarization along +x and positron polarization along −x.

Transverse polarization introduces two additional discriminants beyond the longitudinal case:

- ϕ: Signal and background distributions are examined at different polarization degrees. Higher polarization → better separation.
- $\cos\theta$: Signal and background are well separated regardless of polarization degree. (Signal remains uniform in both ϕ and cos θ; only background plots are shown.)

These variables can be fed into ParT or a simultaneous fit to further boost the S/B ratio.



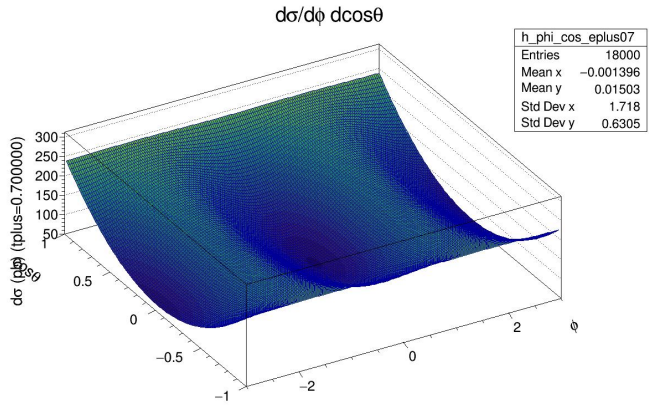


Polarization Study

3. Mathematical formula verification and fitting:

- **Formula verification:** The differential cross section formula for $e^+e^- \to q\bar{q}$ under transverse polarization was derived (**link**) based on the relevant literature. Evaluated at $s = (91.1880\ GeV)^2$, the calculated total cross section is 39.69 nb, in good agreement with the existing experimental result of 41.48 nb (**link**).

$$\frac{d\sigma}{d\Omega} = \frac{3\alpha^2}{4s}\left[\left(1 + \frac{2v_e v_q}{\sin^2\theta_W \cos^2\theta_W}\cdot\frac{s(s-M_Z^2)}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2} + \frac{(v_e^2+a_e^2)(v_q^2+a_q^2)}{\sin^4\theta_W\cos^4\theta_W}\cdot\frac{s^2}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2}\right)(1+\cos^2\theta)\right.$$
$$+\left(\frac{4a_e a_q}{\sin^2\theta_W\cos^2\theta_W}\cdot\frac{s(s-M_Z^2)}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2} + \frac{8v_e a_e v_q a_q}{\sin^4\theta_W\cos^4\theta_W}\cdot\frac{s^2}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2}\right)\cos\theta$$
$$-P^2\left(1 + \frac{2v_e v_q}{\sin^2\theta_W\cos^2\theta_W}\cdot\frac{s(s-M_Z^2)}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2} + \frac{(v_e^2-a_e^2)(v_q^2-a_q^2)}{\sin^4\theta_W\cos^4\theta_W}\cdot\frac{s^2}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2}\right)\sin^2\theta\cos 2\phi$$
$$\left.+P^2\cdot\frac{2a_e v_q}{\sin^2\theta_W\cos^2\theta_W}\cdot\frac{sM_Z\Gamma_Z}{(s-M_Z^2)^2+M_Z^2\Gamma_Z^2}\cdot\sin^2\theta\sin 2\phi\right]$$



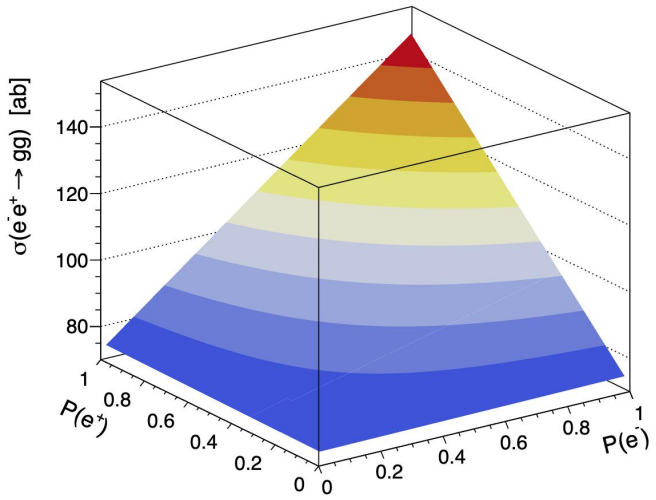
- **Fitting and plotting with the theoretical formula:** The resulting images are consistent with those obtained from Monte Carlo simulation, providing a cross-validation that the polarization-based S/B enhancement method is viable.

Analysis Framework/Sample Generation

Signal: $e^+e^- \to H \to gg$ **Background:** $e^+e^- \to q\bar{q}$ **Tool:** Whizard3

Variation of signal cross-section under different polarization conditions (ISR is turned off here to more intuitively observe the impact of polarization on cross-section):

P_{e^+}/P_{e^-}	0	60%	70%	80%	90%	100%
0	73.8	73.8	73.8	73.8	73.8	73.8
40%	73.8	91.6	94.5	97.5	100.4	103.4
70%	73.8	104.8	110.0	115.2	120.4	125.5
100%	73.8	118.1	125.5	132.9	140.3	147.7



Condition Settings:

- Set the collision energy to 125.0 GeV and enable ISR. (For the impacts of ISR and beam energy spread on the signal cross-section, see the appendix A for details.)
- Set both electrons and positrons to be transversely polarized with P=70%, where electrons are polarized along the positive x-axis and positrons along the negative x-axis.

Analysis Framework/ Detector Simulation and Object Reconstruction

Signal: $e^+e^- \to H \to gg$ **Background:** $e^+e^- \to q\bar{q}$ **Tool:** Delphes

We use Delphes to simulate the CEPC detector response and reconstruct leptons, photons, and jets from final-state particles.

		b jets (CEPC)	c jets (CEPC)	Gluon jets (CEPC)	Gluon jets (FCC-ee-based)
Reco/tagging efficiency		92.3%	80.3%	65.0%	70%
Mistagging rates	b jets	/	4.1%	2.7%	0.001%
	c jets	2.8%	/	6.6%	0.01%
	u jets	0.3%	3.0%	18.4%	1%
	d jets	0.3%	3.0%	18.4%	1%
	s jets	0.3%	4.0%	16.1%	1%
	Gluon jets	2.5%	5.2%	/	/

- The CEPC performance, based on full simulation, is provided by Zhang Kaili.
- Reference study (FCC-ee-based assumptions) performance is based on idealized assumptions of a perfect detector.

Analysis Framework/Event Preselection

	CEPC	Reference study (FCC-ee-based assumptions)
$\sigma_{ee\to H\to gg}$	34.3 ab (with polarization)	23 ab (without polarization)
$\sigma_{ee\to qq}$	396.3 pb (with isr)	61 pb (without isr)
L	$10\ ab^{-1}/year$	$10\ ab^{-1}/year$
cut	2 gluon – tagged jets	2 gluon – tagged jets
	0 isolated $l^{\pm}$	0 isolated $l^{\pm}$
	visible mass > 114GeV	
S/B (before cut)	8.6×10^{-8} (with isr)	3.8×10^{-7} (without isr)
S/B (after cut)	4.3×10^{-6}	1.8×10^{-3}
Signal efficiency	24.91%	50%
Background efficiency	0.51%	0.006%

- CEPC preselection is performed programmatically via ROOT tree branches.
- Reference study (FCC-ee-based assumptions) relies on direct numerical calculation to obtain the idealized result.

Analysis Framework/Particle Transformer

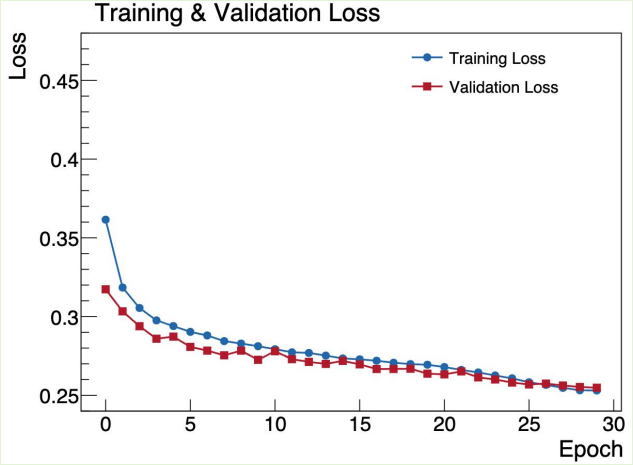
Input Features (see the appendix B for plots)

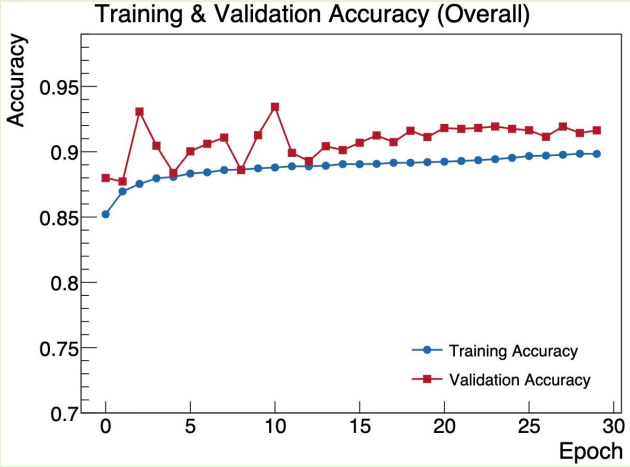
- Particle-level Features：

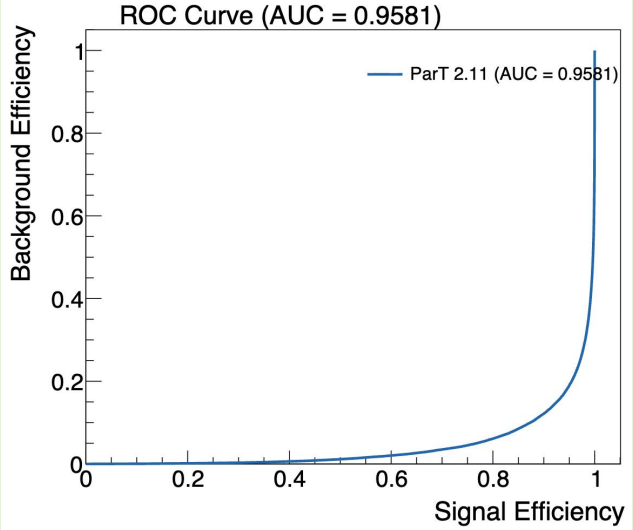
    1. Kinematics：part_pt_log, part_e_log, part_logptrel, part_logerel, part_deltaR, part_deta, part_dphi
    2. Particle Identification：part_charge, part_isChargedHadron, part_isNeutralHadron, part_isPhoton, part_isElectron, part_isMuon
    3. Tracking：part_d0, part_d0err, part_dz, part_dzerr
    4. Calorimeter：part_em_fraction, part_had_fraction
    5. dE/dx PID Probabilities：part_prob_pi, part_prob_k, part_prob_p

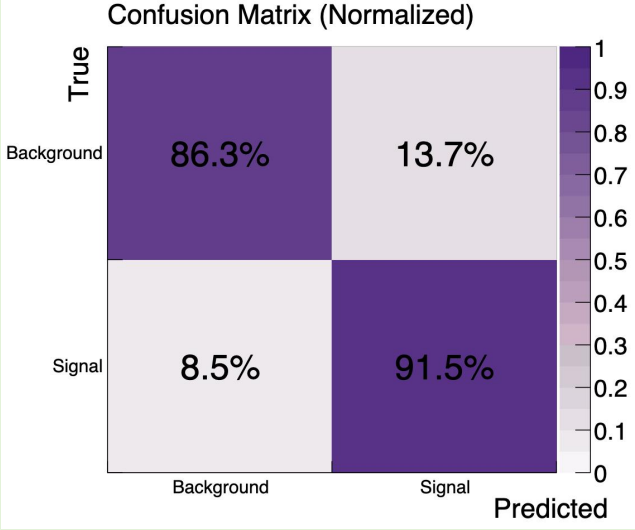
- Lorentz Vectors：part_px, part_py, part_pz, part_energy

Analysis Framework/Particle Transformer



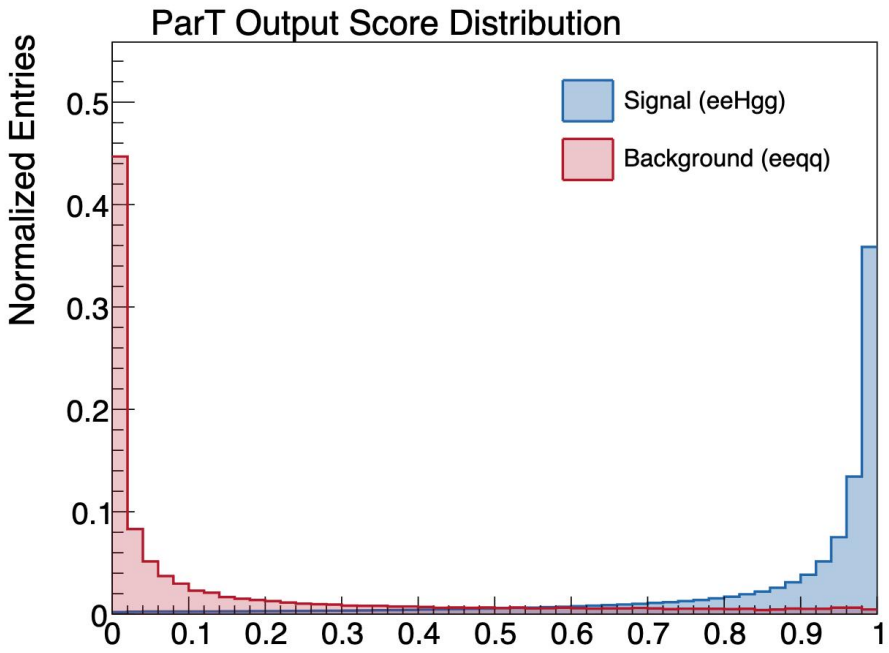


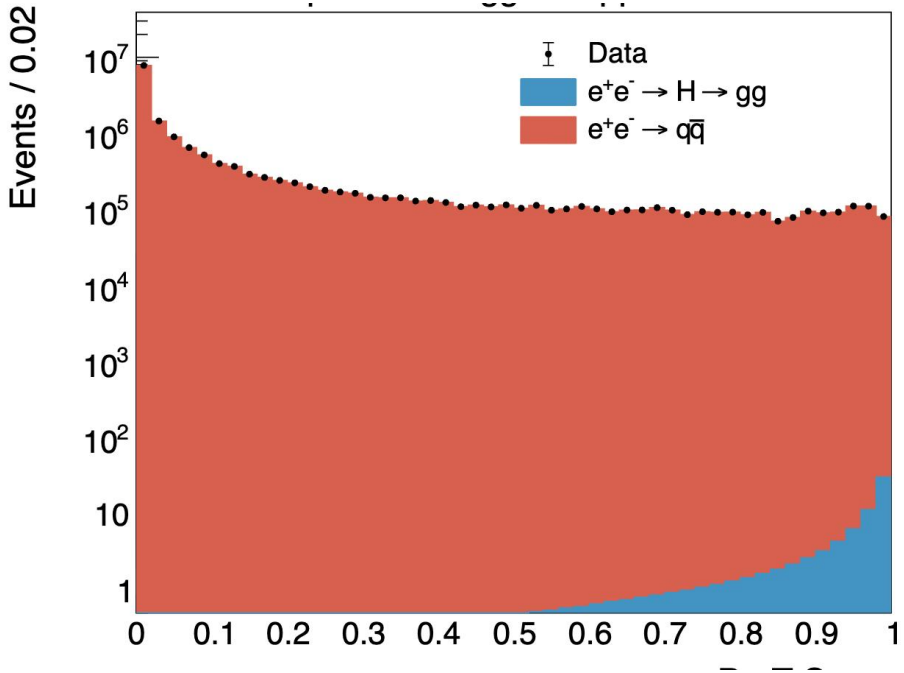




- **Loss:** Training loss decreases steadily from 0.36 to ~0.25, with no rise in validation loss — no overfitting.
- **Accuracy:** Validation Balanced Accuracy fluctuates within a narrow range of 0.1, indicating stable training.
- **ROC curve:** AUC = 0.96, demonstrating excellent signal–background score separation.
- **Confusion matrix:** Signal efficiency 91.5%, background efficiency 86.3%.

Analysis Framework/Particle Transformer

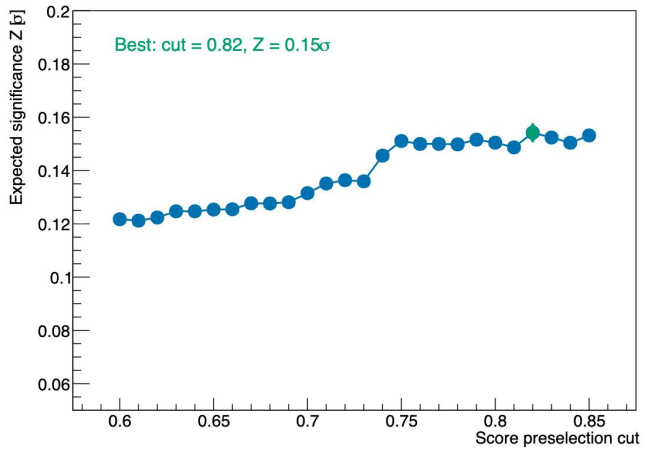


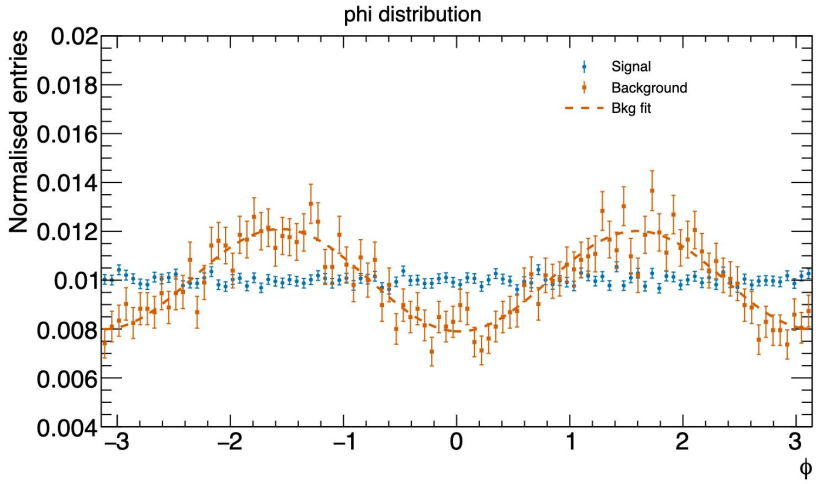


- **Score distribution:** Signal peaks near 1 and background near 0, with clear separation — the model achieves strong discriminative power.
- **Cross-section weighted score distribution:** After applying the true cross sections at an integrated luminosity of 10 ab^{-1}/year, the background outnumbers the signal by ~7 orders of magnitude. Nevertheless, a stringent high-score cut can suppress the background dramatically while retaining most of the signal.

Analysis Framework/Simultaneous Fit

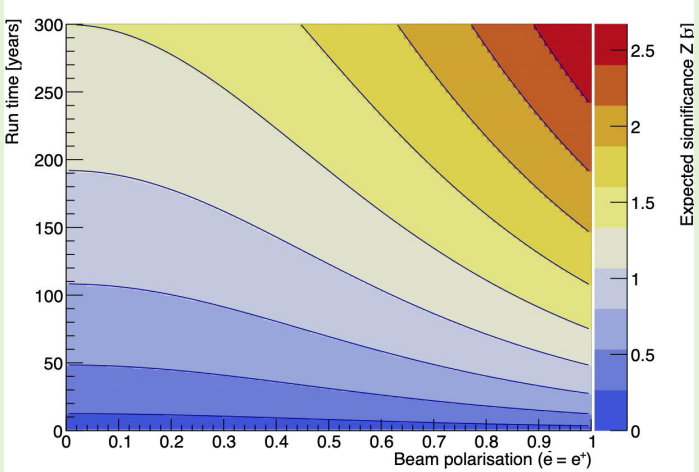
- Unlike hard cuts that discard events, the simultaneous fit uses all data — high–S/B regions are naturally weighted more, while information from the rest is retained rather than thrown away.
- Before the fit, a score preselection is applied (score > 0.82), with the optimal threshold determined by a significance scan.
- A simultaneous binned likelihood fit is performed in the 2D (ParT score, φ) plane.

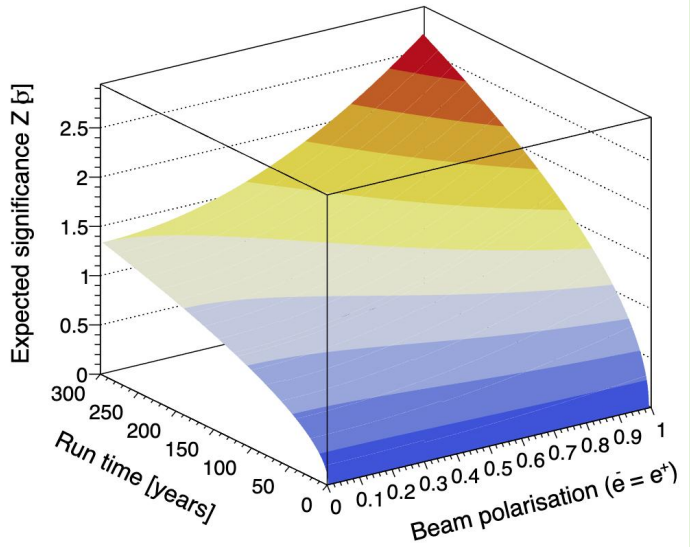




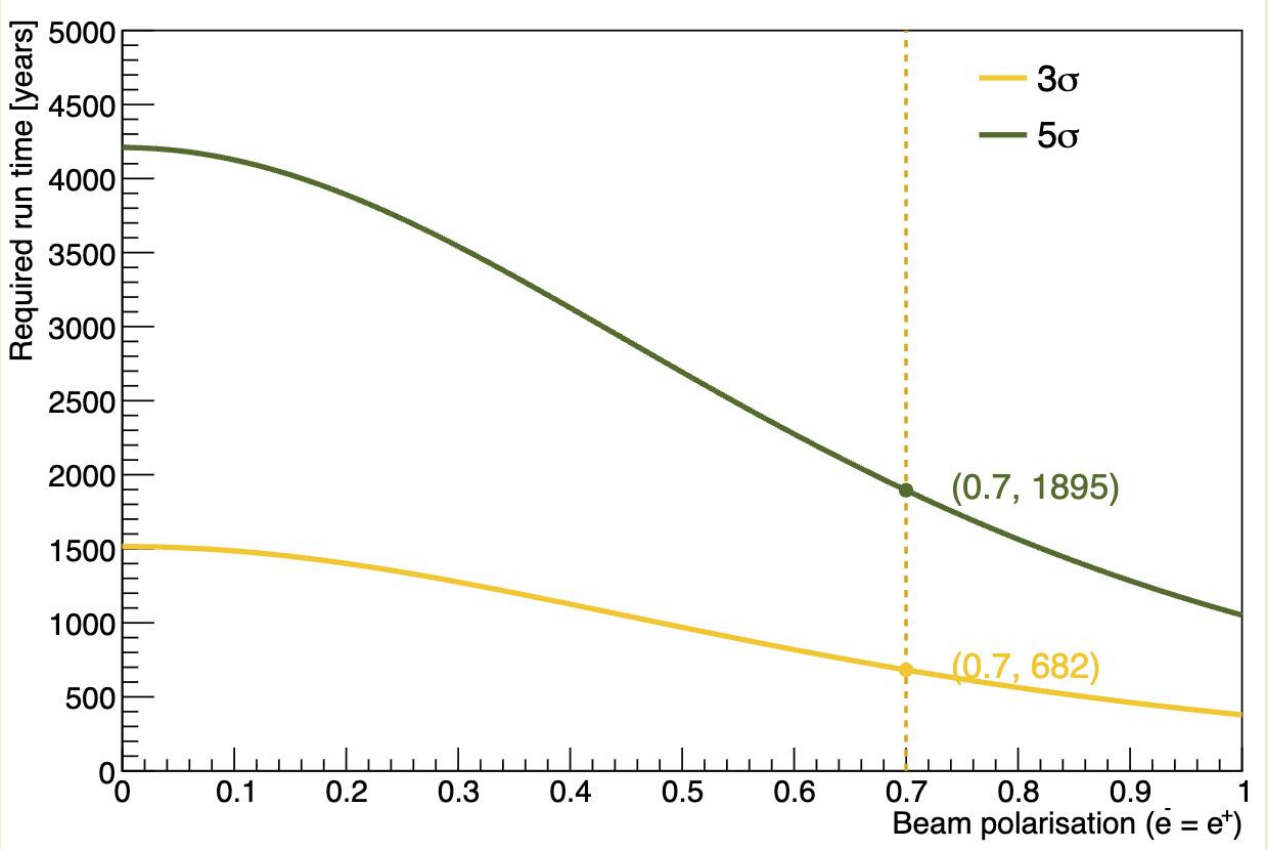
Method	Significance
Hard cut (score>0.98)	0.1σ
2D simultaneous fit	0.2σ

Analysis Framework/Simultaneous Fit





$L = 10\ ab^{-1}/year$



Comparison with FCC-ee Results

	CEPC	Reference study (FCC–ee–based)	Combined	
Polarization	70%	0	0	70%
Gluon tagging/mis–tagging rates	CEPC version	FCC–ee version	FCC–ee version	
Event Preselection	2 gluon – tagged jets 0 isolated $l^{\pm}$ visible mass > 114GeV	2 gluon – tagged jets 0 isolated $l^{\pm}$	2 gluon–tagged jets 0 isolated $l^{\pm}$ visible mass>114GeV	
Event Selection Optimization	ParT Simultaneous Fit	BDT	ParT Simultaneous Fit	
Significance	0.2σ (with isr)	1.1σ (sig with isr, bkg without isr) 0.4σ (with isr)	3.1σ (with isr)	4.5σ (with isr)

- Here, “combined” refers to adopting the complete CEPC analysis workflow, with only the parameters associated with the gluon-tagging component replaced by those used in the reference study (FCC-ee-based assumptions).
- Event preselection for this component is likewise performed through numerical calculations.

Comparison with FCC-ee Results

- The differences between the results of CEPC and Reference study (FCC-ee-based assumptions) are mainly caused by Gluon tagging. The gluon mistagging rate of Reference study is approximately 20 times better than ours. The following lists the significant improvements that can be achieved when CEPC uses different Gluon tags.
- Of course, a good Gluon tag performance will impose higher requirements on the detector. Exactly which aspects these requirements cover we are still investigating.

		Gluon jets	Gluon jets	Gluon jets	Gluon jets	Gluon jets
Reco/tagging efficiency		65.0%	65.0%	65.0%	65.0%	65.0%
Mistagging rates	b jets	2.7%	Reduce by 5 times	Reduce by 10 times	Reduce by 15 times	Reduce by 20 times
	c jets	6.6%				
	u jets	18.4%				
	d jets	18.4%				
	s jets	16.1%				
	Gluon jets	/	/	/	/	/
Significance (with isr)		0.2σ	1.2σ	2.3σ	3.4σ	4.3σ

Summary and Next Steps

Summary:

Feasibility study of measuring the electron Yukawa coupling at the CEPC:

- Completed the background survey and defined the research direction: enhancing the signal-to-background ratio via beam polarization and machine learning.
- Verified the feasibility of this approach through Monte Carlo simulation and analytical cross-section calculations.
- Generated signal and background samples with Whizard3, and performed detector simulation and object reconstruction with Delphes.
- Implemented event preselection, yielding a preliminary improvement in the S/B ratio.
- Performed event selection optimization using the Particle Transformer and a simultaneous fit, achieving an expected significance of 0.2σ.
- Compared with the Reference study (FCC-ee-based assumptions) result, and evaluated the maximum achievable significance when adopting their gluon-tagging approach, yielding 4.5σ.

Outlook:

- Study gluon tagging and estimate systematic uncertainties using toy Monte Carlo.
- Write and publish the paper.

References

[1] F. An, et al., "Measuring the electron Yukawa coupling via resonant s-channel Higgs production at FCC-ee," Eur. Phys. J. Plus 137, 201 (2022). DOI: 10.1140/epjp/s13360-021-02204-2

[2] CMS Collaboration, "Search for the Higgs boson decay to a pair of electrons in proton–proton collisions at $\sqrt{s} = 13 TeV$," Phys. Lett. B 846, 137783 (2023). DOI: 10.1016/j.physletb.2023.137783

[3] ATLAS Collaboration, "Search for the Higgs boson decays $H \to ee$ and $H \to e\mu$ in pp collisions at $\sqrt{s} = 13 TeV$ with the ATLAS detector," Phys. Lett. B 801, 135148 (2020). DOI: 10.1016/j.physletb.2019.135148

[4] A. Blondel, et al., "FCC-ee: Your Questions Answered," arXiv:1906.02693 [hep-ph] (2019). DOI: 10.48550/arXiv.1906.02693

[5] A. Abada, et al. (FCC Collaboration), "FCC-hh: The Hadron Collider — Future Circular Collider Conceptual Design Report Volume 3," Eur. Phys. J. Spec. Top. 228, 755–1107 (2019). DOI: 10.1140/epjst/e2019-900087-0

[6] S. Jadach and R. A. Kycia, "Lineshape of the Higgs boson in future lepton colliders," Phys. Lett. B 755, 58–63 (2016). DOI: 10.1016/j.physletb.2016.01.065

References

[7] J. Gao, et al. (CEPC Study Group), "CEPC Technical Design Report — Accelerator," Radiat. Detect. Technol. Methods 8, 1–1105 (2024). DOI: 10.1007/s41605-024-00445-3

[8] CEPC Study Group, "CEPC Technical Design Report — Reference Detector," arXiv:2510.05260 [hep-ex] (2025). DOI: 10.48550/arXiv.2510.05260

[9] M. Dobbs and J.B. Hansen, "The HepMC2 User Manual," CERN, 2009. Available online: https://hepmc.web.cern.ch/hepmc/releases/HepMC2_user_manual.pdf

[10] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.3 Physics and Manual," CERN, 2003. Available online: https://www.pythia.org/download/pythia6/lutp0613man2.pdf

[11] K. Hagiwara, "Transverse-polarization effects in e^+e^- collisions: The role of chiral symmetry," Phys. Rev. D 33, 3203 (1986). DOI: 10.1103/PhysRevD.33.3203

[12] H. Qu, C. Li, and S. Qian, "Particle Transformer for Jet Tagging," in Proceedings of the 39th International Conference on Machine Learning (ICML 2022), PMLR 162, 18281–18292 (2022). arXiv:2202.03772 [hep-ph]. DOI: 10.48550/arXiv.2202.03772

Appendices

Appendix A

$$\sigma_{ee\to H} = \frac{4\pi\Gamma(H\to e^+e^-)}{(s-m_H^2)^2 + m_H^2\Gamma_H}$$

$$s = (125\ GeV)^2$$

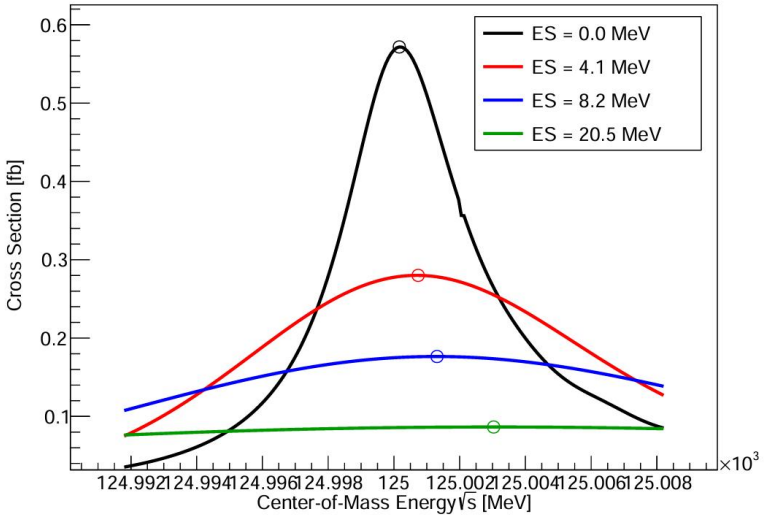
$$\sigma_{ee\to H} = \frac{4\pi\Gamma(H\to e^+e^-)}{m_H^2\Gamma_H} = 1.64\ fb$$

ISR

$$\sigma_{res}(s) = \int_0^{1-s_m/s} dx\sigma(s(1-x))F(x,s) = 0.57\ \text{fb}$$

Beam Energy Spread

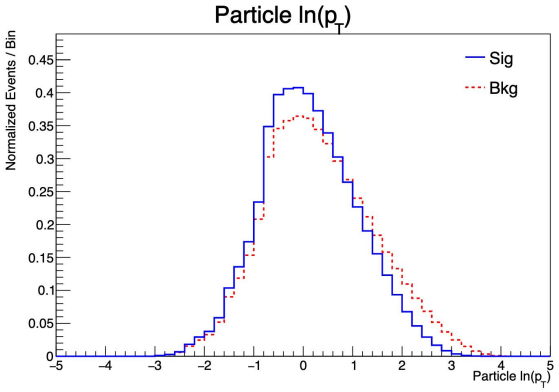
$$\sigma_{obs}(s) = \frac{R}{s} + \int_0^{\infty} G(s,s')\sigma_{res}(s')\,ds' = 0.28\ \text{fb}$$

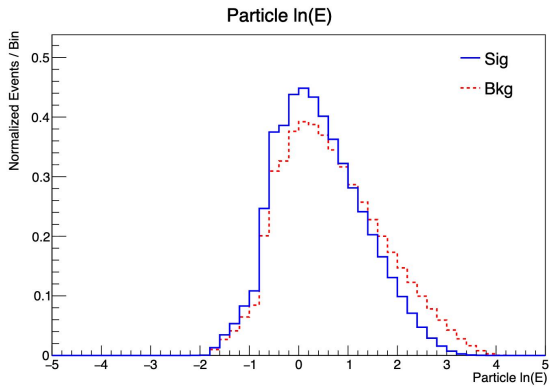


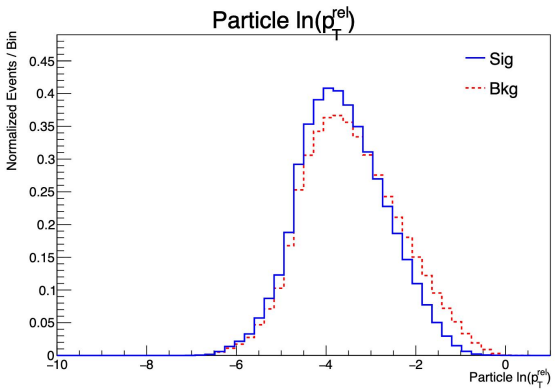
ES(MeV)	σ_{max}(fb)
0	0.57
4.1	0.28
8.2	0.18
20.5	0.09

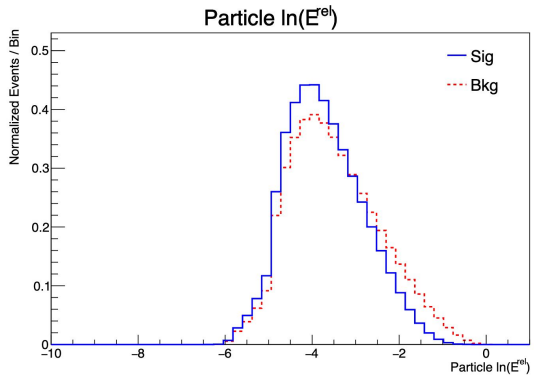
Appendix B

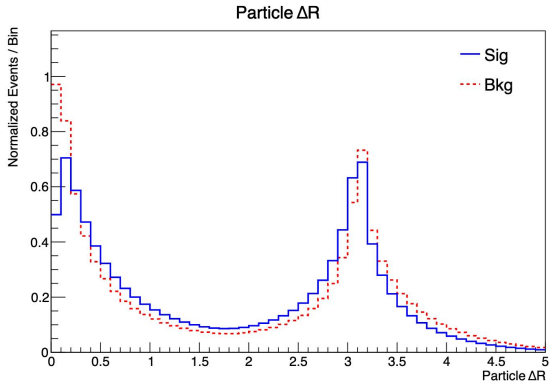
Kinematics:

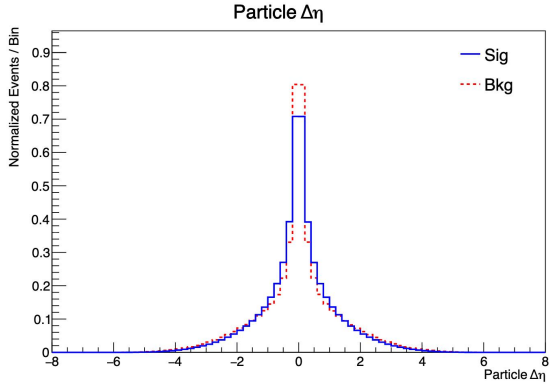


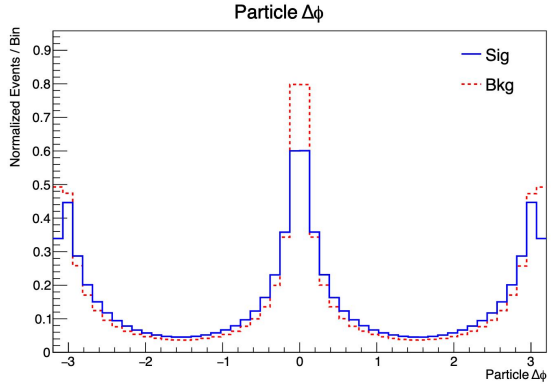






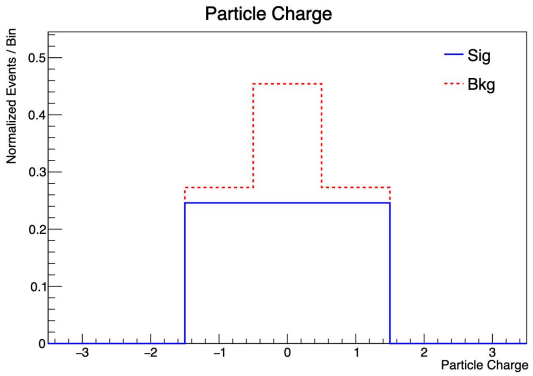


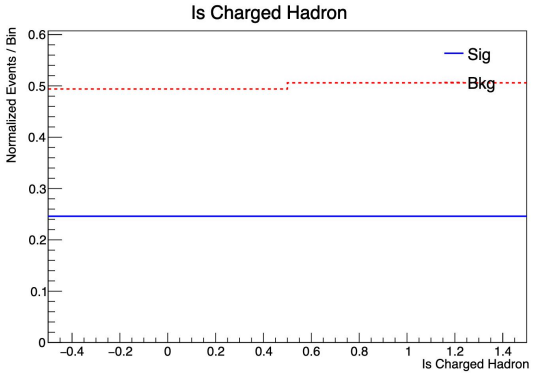


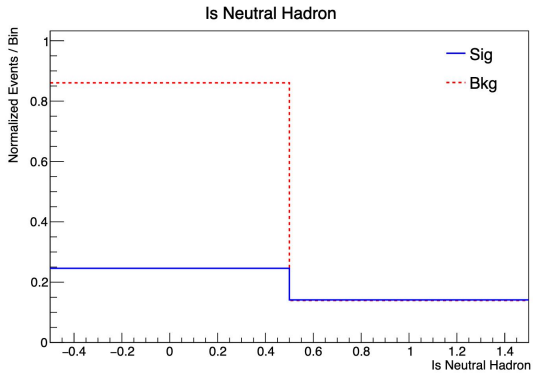


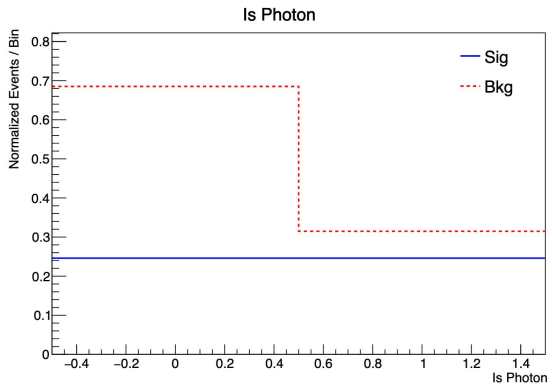
Appendix B

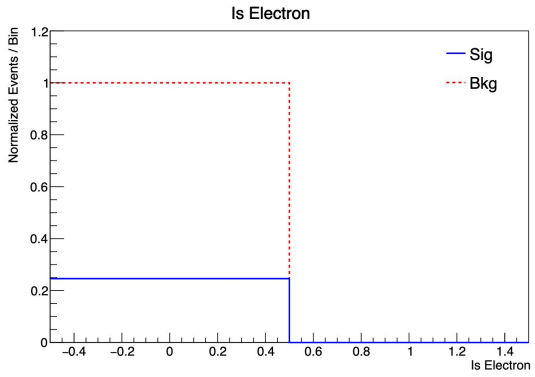
Particle Identification:

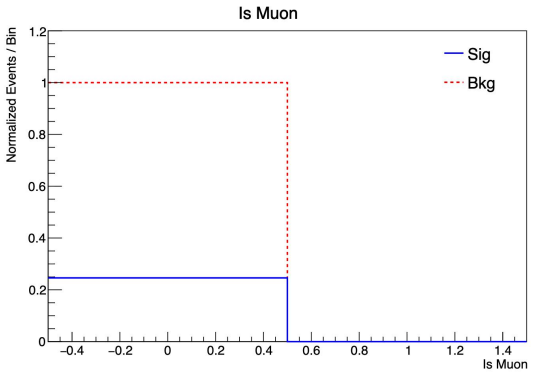






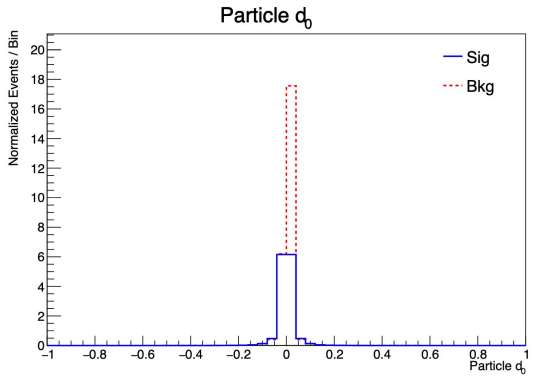


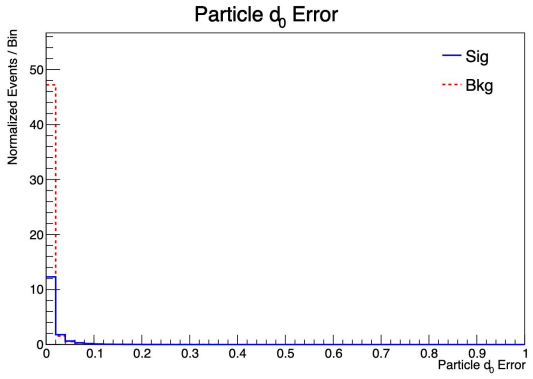


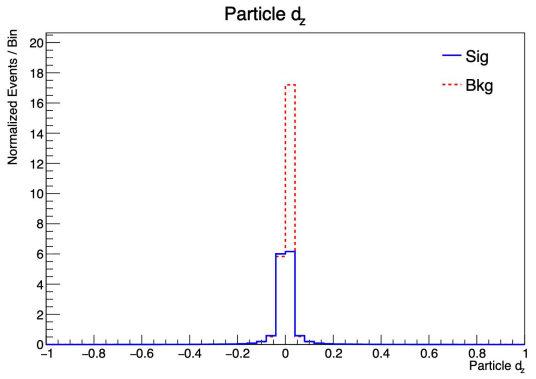


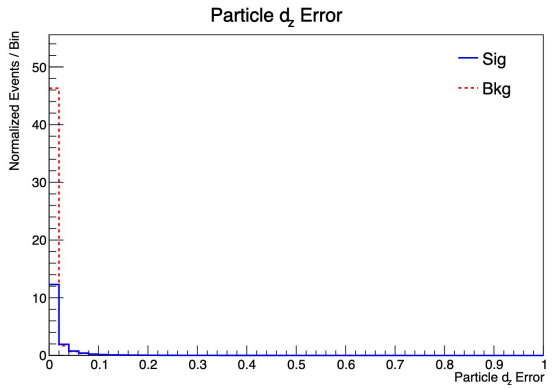
Appendix B

Tracking:

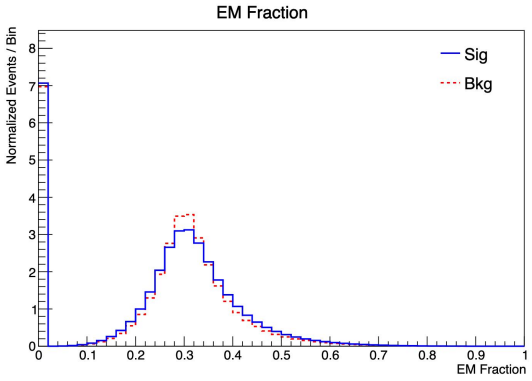


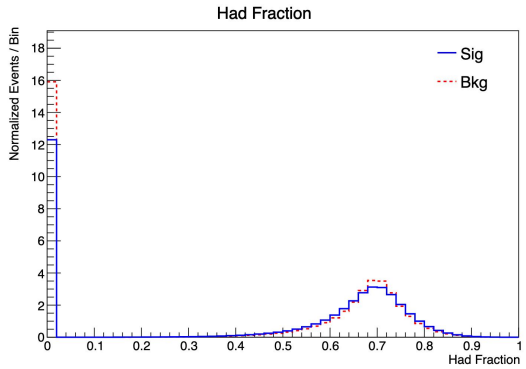






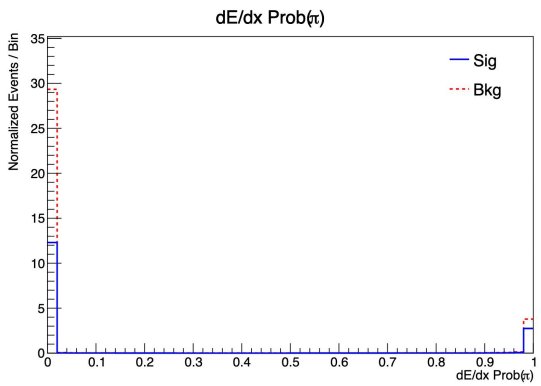
Calorimeter:

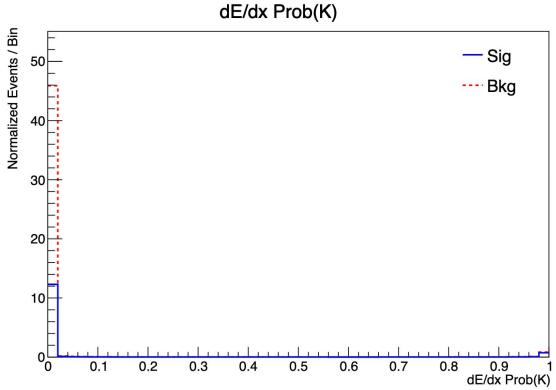


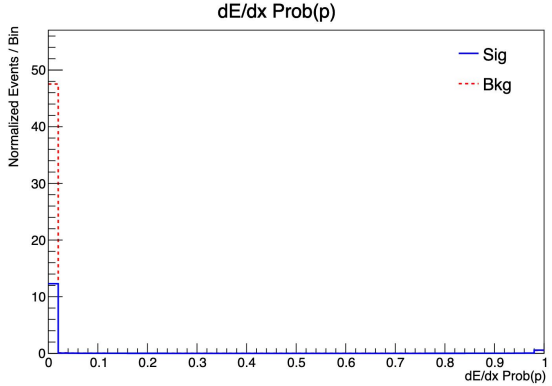


Appendix B

dE/dx PID Probabilities:







Lorentz Vectors:

