

Jet flavor tagging with Particle Transformer for Higgs factories

Taikan Suehara^{1,2*}, Takahiro Kawahara², Tomohiko Tanabe³,
Risako Tagami²

¹International Center for Elementary Particle Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 133-0033, Japan.

²Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 133-0033, Japan.

³MI-6 Ltd., 8-13 Nihombashi-Kobunacho, Chuo-ku, Tokyo, 103-0024, Japan.

*Corresponding author(s). E-mail(s): suehara@icepp.s.u-tokyo.ac.jp;

Abstract

We study the performance of the Particle Transformer (ParT) for jet flavor tagging using ILD full simulation events (1M jets) as well as fast simulation samples (10M and 1M jets). We perform 3-category ($\mathbf{b/c/d}$), 6-category ($\mathbf{b/c/d/u/s/g}$), and 11-category trainings (including quark–antiquark separation), incorporating multivariate hadron particle identification information from dE/dx and time-of-flight. For $\mathbf{b/c}$ tagging, we observe a factor of 5–10 improvement over previous BDT-based taggers, and we obtain reasonable performance for strange tagging and quark/antiquark separation. The 10M-jet fast simulation study indicates that further gains are possible with higher training statistics.

Keywords: Flavor tagging, Higgs factory, Particle Transformer

1 Introduction

Event reconstruction at future e^+e^- colliders relies on extracting fine-grained information from highly segmented tracking and calorimeter systems and translating it into physics objects such as charged particles, photons, and jets. As detector granularity increases, pattern recognition challenges become more prominent, making modern

machine learning (ML) approaches a natural choice for several reconstruction and identification tasks.

Jet flavor tagging aims to identify the initiating parton flavor (e.g., b , c , or light flavor) using information carried by reconstructed particles inside a jet. A key handle for heavy flavor tagging is the finite lifetime of heavy flavor hadrons, which leads to displaced tracks and secondary vertices. Traditional approaches for Higgs Factory studies, such as LCFIPlus[1], explicitly reconstruct secondary vertices and then combine vertex- and track-level observables in multivariate classifiers.

More recently, end-to-end deep learning models have been developed that ingest low-level per-particle (or per-track) information and learn powerful representations without relying on a dedicated vertex reconstruction step. For example, ParticleNet, which operates on point-cloud representations and uses graph-based architectures, has been applied to fast simulation samples for FCC-ee and was reported to achieve substantially improved b -tag rejection compared to LCFIPlus[2]. Transformer-based models have also demonstrated strong performance in LHC jet tagging benchmarks; in particular, the Particle Transformer (ParT)[3] leverages attention mechanisms that scale well with large training samples.

In this work, we study the performance of ParT for jet flavor tagging using ILD full simulation events as well as large fast simulation samples. We focus on Higgs factory conditions and assess the impact of detailed detector information and particle identification (PID) inputs, which are expected to be especially relevant for strange jet tagging.

2 Simulation and Data samples

We use simulated events for the International Large Detector (ILD) concept[4]. The ILD tracking system comprises a silicon vertex detector, silicon inner/outer trackers, and a central Time Projection Chamber (TPC) in the barrel region, complemented by silicon based tracking in the forward region. The calorimeter system is highly granular, enabling Particle Flow reconstruction, and is surrounded by a 3.5 T solenoid; muon detectors are located outside the magnet.

Several detector features provide PID capability. In the TPC, specific ionization (dE/dx) measurements allow charged hadron separation over a broad momentum range. In addition, we assume timing information in the first ten layers of the electromagnetic calorimeter and use time-of-flight information for PID, with an indicative single hit time resolution of 100 ps.

The event processing is based on the iLCSoft[5] stack, with Geant4-based detector simulation using the ILD detector description (DDSim), digitization that models detector effects, and low-level tracking reconstruction. For the jet-based analyses, we reconstruct Particle Flow objects with PandoraPFA[6] and form jets using the Durham clustering algorithm. The reconstructed particles associated with each jet are used to build the model inputs. As part of the reconstruction chain, we apply the Comprehensive PID algorithm[7], which combines dE/dx and time-of-flight related observables in a boosted decision tree to classify charged particles. The resulting PID responses are

used as additional per-particle features for the jet taggers and are expected to provide extra discriminating power, particularly for strange jet identification.

Our baseline physics process is $e^+e^- \rightarrow ZH \rightarrow \nu\nu q\bar{q}$ at $\sqrt{s} = 250$ GeV. To evaluate flavor tagging performance across Higgs decays, we consider six Higgs-to-parton categories, $H \rightarrow b\bar{b}, c\bar{c}, s\bar{s}, u\bar{u}, d\bar{d}, gg$; in the full simulation samples, approximately 2×10^6 jets are available per category. We use the sample prepared by ILD central production (mc-2020) with iLCSoft v02-02. Event generation was done by Whizard 2.8.5[8], hadronized by Pythia 6.427[9].

To study scaling behavior and enable analyses requiring much larger statistics, we also use samples produced privately with Whizard 3.1.5 hadronized by Pythia 6 with the same configuration as the full simulation and SGV[10] fast simulation, comprising roughly 1×10^7 jets per category for $H \rightarrow b\bar{b}, c\bar{c}, d\bar{d}$ categories. The SGV samples are only used for heavy flavor (b/c) discrimination since the SGV configuration employed here does not include hadron-level PID information that is essential for strange tagging.

3 Algorithm

We employ the Particle Transformer (ParT) as the baseline jet flavor tagging model. ParT follows the standard Transformer design with an embedding stage, a stack of self-attention blocks, and a final multilayer perceptron for classification. In addition to the usual content-based attention, ParT incorporates a pairwise “interaction” term derived from the kinematics of particle pairs, which is injected as a bias into the attention weights. This interaction-aware attention is particularly suitable for jet tagging, where correlations among constituents carry essential information.

Each jet is represented as a set of reconstructed particles and is processed as a variable length sequence. Charged and neutral particles are embedded with separate input projections to account for the different feature availability. For charged particles, the input features include impact parameter related variables, track parameter uncertainties, PID-related quantities, and kinematic variables; for neutral particles, only the available kinematic and PID-related features are used. The per-particle input variables used in this study are summarized in Table 1.

We train three classification settings with different label granularities. In the 3-category training, jets are labeled as b , c , or light (denoted as d) using samples generated for $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$, and $H \rightarrow d\bar{d}$. In the 6-category training, the labels are extended to b, c, s, u, d , and g using samples generated for $H \rightarrow b\bar{b}, c\bar{c}, s\bar{s}, u\bar{u}, d\bar{d}$, and gg . For the 3- and 6-category trainings, we assign labels at the event level based on the generated sample type, without accessing per-event MC truth matching in the reconstruction.

In the 11-category training, we assign labels by matching reconstructed jets to the underlying partons using MC truth information and distinguish quarks from anti-quarks. For each reconstructed particle, we first identify the corresponding MC truth particle and group truth particles by their color-singlet system without using angular information. We then resolve the parton assignment only within each color singlet by comparing the opening angle between the truth particle and the partons (quark/gluon)

produced directly from the color singlet, and assign the truth particle to the closest parton. Using this particle-parton association, we label each reconstructed jet by the parton that receives the largest fraction of the reconstructed constituent energy.

For all trainings, we split each dataset into 80% for training, 10% for validation, and 10% for testing.

Table 1 Per-particle input variables for the Particle Transformer. (C) in Group indicates that the variables are only used for charged particles. Interaction (bias) variables are used for biasing attention weights and all others are for features input.

| Group | Variables[11] | Description |
|-----------------------|---|---|
| Impact parameters (C) | $d_0, z_0; \text{ip2D}, \text{ip3D}$ | Signed transverse/longitudinal impact parameters and their derived 2D/3D values |
| Jet distances (C) | d_j | Track displacement relative to the jet direction |
| Particle ID | Muon/Electron/Gamma flags; charged/neutral hadron flags; type | PID indicator variables and reconstructed PDG ID |
| Particle ID (C) | CPID outputs (π, K, p probabilities) | Comprehensive PID (BDT) outputs (used in 6- and 11-category trainings) |
| Kinematics | Charge; $\log(E/E_{\text{jet}}); \Delta\theta; \Delta\phi$ | Per-particle kinematic variables relative to the jet axis |
| Track errors (C) | σ (covariance elements) | Track parameter uncertainty information |
| Interaction (bias) | $\log k_t, \log z, \log \Delta R^2, \log m^2$ | Pairwise variables used to construct the interaction bias in attention |

4 Results

Figure 1 summarizes the b - and c -tagging performance for the 3-category and 11-category trainings. Compared to the conventional LCFIPlus approach, we observe a substantial reduction in background acceptance at a fixed signal efficiency. In particular, at a b -tag efficiency of 80%, the acceptance for c -jet background is reduced to the $\mathcal{O}(10^{-3})$ level, and the light-flavor background to the $\mathcal{O}(10^{-3})$ – $\mathcal{O}(10^{-4})$ level, depending on the training setup and simulation model.

The comparison between full simulation and SGV fast simulation indicates non-negligible differences in absolute performance, which we are currently investigating in more detail. In addition, within SGV we observe an improvement when increasing the training statistics from 1M to 10M jets. The 3-category and 11-category trainings yield broadly comparable b -tagging performance; however, in the high-purity regime the rejection of gluon jets becomes challenging in the 11-category setup, which is expected in part due to $g \rightarrow b\bar{b}$ splittings that can mimic heavy flavor signatures. For c -tagging, the task remains more difficult than b -tagging, yet the background acceptance at a c -tag efficiency of 50% is still significantly improved compared to LCFIPlus. Table 2(a) lists representative working points extracted from Fig. 1 at a signal efficiency of 0.8,

Table 2 (a) Background efficiencies at signal efficiency = 0.8. Category shows 3-c (3-category) or 11-c (11-category) training, followed by signal – background (eg. b - c means b -tag with c -background). (b) Background efficiencies on strange tag with 6-category training and ILD full-simulation.

| (a) | | | | (b) | |
|--------------|-----------------------|-----------------------|-----------------------|---------------------------------|------|
| Category | Full sim | SGV (1M) | SGV (10M) | Efficiency ($\times 10^{-2}$) | |
| 3-c $b - c$ | 4.53×10^{-3} | 1.48×10^{-3} | 4.64×10^{-4} | b bkg | 1.14 |
| 3-c $b - d$ | 1.42×10^{-3} | 8.05×10^{-4} | 4.93×10^{-4} | c bkg | 12.9 |
| 3-c $c - b$ | 6.12×10^{-2} | 4.24×10^{-2} | 2.94×10^{-2} | u bkg | 42.5 |
| 11-c $b - c$ | 6.47×10^{-3} | 2.83×10^{-3} | 1.26×10^{-3} | d bkg | 40.5 |
| 11-c $b - s$ | 1.08×10^{-3} | 5.81×10^{-4} | 3.99×10^{-4} | g bkg | 25.5 |
| 11-c $b - d$ | 1.08×10^{-3} | 6.20×10^{-4} | 4.01×10^{-4} | | |
| 11-c $b - u$ | 9.09×10^{-4} | 5.99×10^{-4} | 4.22×10^{-4} | | |
| 11-c $b - g$ | 1.86×10^{-2} | 1.49×10^{-2} | 1.30×10^{-2} | | |

facilitating a direct comparison across full simulation and SGV with different training statistics.

Figure 2(a) shows the strange tagging performance in the 6-category training with full simulation. The strange background efficiencies at the working point of signal efficiency $\varepsilon_s = 0.8$ (extracted from Fig. 2(a)) are summarized in Table 2(b). Strange tagging relies strongly on hadron-level PID information inside the jet; however, jets initiated by non-strange partons may still contain strange hadrons through fragmentation, including contributions from gluon splitting. Consequently, the separation is intrinsically probabilistic, and we find that discriminating among u , d , and g jets is particularly challenging. Gluon jets can nevertheless be separated from quarks to some extent due to their different kinematic and radiation patterns.

Figure 2(b) presents the agreement between the predicted labels (taken as the highest probability class) and the truth labels in the 11-category training. Beyond jet flavor identification, we observe that the model can separate quarks from antiquarks for heavy flavors (b , c , and s) with meaningful accuracy. The separation is strongest for charm, where the larger electric charge of the underlying hadrons provides additional information. In contrast, separating u and d (and their antiparticles) remains difficult, and in particular the discrimination of charge-conjugate light flavor pairs (e.g., u vs. \bar{d} and d vs. \bar{u}) is essentially consistent with random guessing.

5 Summary and Prospects

We evaluated jet flavor tagging performance with the Particle Transformer using both ILD full simulation and large statistics fast simulation samples. For heavy flavor discrimination, across the 3-, 6-, and 11-category trainings, the ParT-based taggers consistently outperform the conventional baseline, achieving substantially lower background acceptance at fixed signal efficiency (Fig. 1 and Table 2(a)). For strange tagging, we find that incorporating hadron-level PID information (CPID) provides

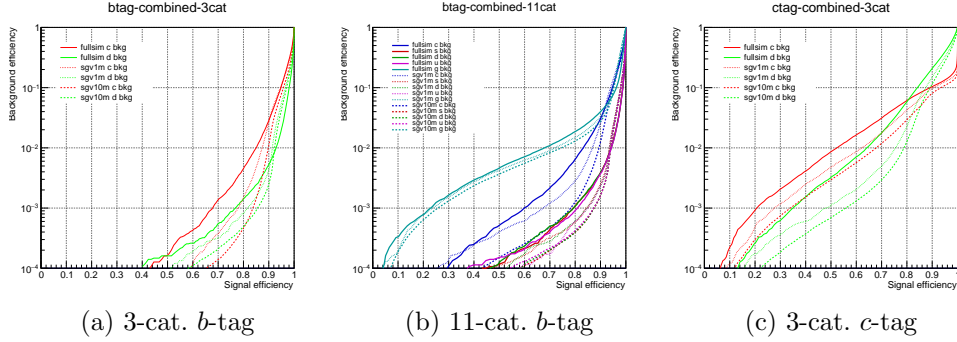


Fig. 1 Performance plots for b - and c -tagging (three panels shown side-by-side). The line colors indicate background jet categories. The solid, dotted, and dashed curves correspond to ILD full simulation, SGV fast simulation trained with 1M jets, and SGV fast simulation trained with 10M jets, respectively.

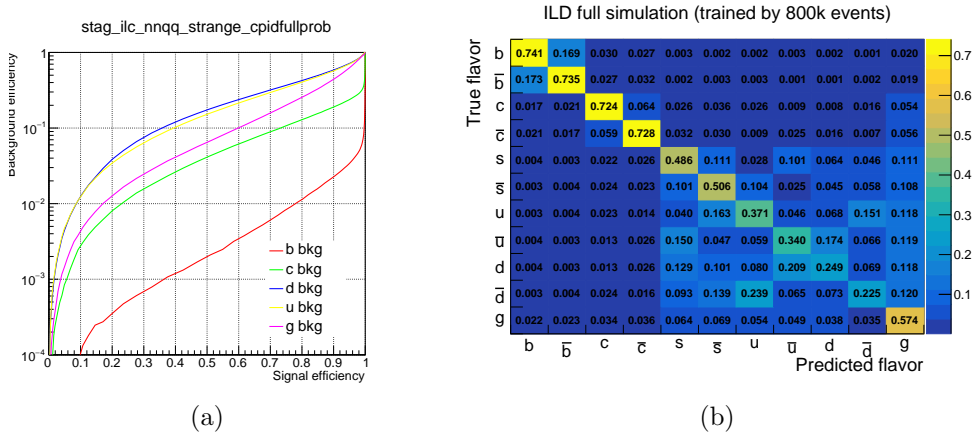


Fig. 2 (a) Signal-versus-background performance for 6-category strange tagging in ILD full simulation. (b) Confusion matrix for the 11-category classifier in ILD full simulation, where the predicted flavor is defined as the class with the highest predicted probability.

useful discrimination, although the separation among light flavors remains intrinsically limited by fragmentation (Fig. 2(a) and Table 2(b)). In the 11-category setting, the model further demonstrates sensitivity to quark–antiquark separation for heavy flavors, as reflected in the truth–prediction agreement matrix (Fig. 2(b)).

The SGV studies indicate a clear dependence on training statistics, motivating further investigations with even larger samples and a systematic validation with full simulation. The resulting taggers can be deployed in the ILD analysis workflow through an ONNX-based inference pipeline[12], and updates of physics analyses using this tool are ongoing[13][14]. Future work will focus on improving the input representation and training strategy, as well as quantifying systematic differences between full and fast simulation and their impact on physics measurements.

Acknowledgements

We would like to thank the LCC generator working group and the ILD software working group for providing the simulation and reconstruction tools and producing the Monte Carlo samples used in this study. This work has benefited from computing services provided by the ILC Virtual Organization, supported by the national resource providers of the EGI Federation and the Open Science GRID. This work is supported by JSPS KAKENHI Grant Number JP22H05113.

References

- [1] Suehara, T., Tanabe, T.: LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies. *Nucl. Instrum. Meth. A* **808**, 109–116 (2016) <https://doi.org/10.1016/j.nima.2015.11.054> [physics.ins-det]
- [2] Bedeschi Franco, S.M. Gouskos Loukas: Jet flavour tagging for future colliders with fast simulation. *The European Physical Journal C* (2022) <https://doi.org/10.1140/epjc/s10052-022-10609-1>
- [3] Qu, H., Li, C., Qian, S.: Particle Transformer for Jet Tagging (2022) [arXiv:arXiv:2202.03772](https://arxiv.org/abs/2202.03772) [hep-ph]
- [4] Abramowicz, H., et al.: International Large Detector: Interim Design Report (2020) [arXiv:2003.01116](https://arxiv.org/abs/2003.01116) [physics.ins-det]
- [5] iLCSoft. <https://github.com/ilcsoft>
- [6] Thomson, M.A.: Particle Flow Calorimetry and the PandoraPFA Algorithm. *Nucl. Instrum. Meth. A* **611**, 25–40 (2009) <https://doi.org/10.1016/j.nima.2009.09.009> [physics.ins-det]
- [7] Einhaus, U.: CPID: A Comprehensive Particle Identification Framework for Future e^+e^- Colliders (2023) [arXiv:2307.15635](https://arxiv.org/abs/2307.15635) [hep-ex]
- [8] Kilian, W., Ohl, T., Reuter, J.: WHIZARD: Simulating Multi-Particle Processes at LHC and ILC. *Eur. Phys. J. C* **71**, 1742 (2011) <https://doi.org/10.1140/epjc/s10052-011-1742-y> [arXiv:0708.4233](https://arxiv.org/abs/0708.4233) [hep-ph]
- [9] Sjostrand, T., Mrenna, S., Skands, P.Z.: PYTHIA 6.4 Physics and Manual. *JHEP* **05**, 026 (2006) <https://doi.org/10.1088/1126-6708/2006/05/026> [arXiv:hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175)
- [10] Berggren, M.: SGV 3.0 - a fast detector simulation (2012) [arXiv:1203.0217](https://arxiv.org/abs/1203.0217) [physics.ins-det]
- [11] LCFIPlus implementation adapted to ParT. <https://github.com/suehara/LCFIPlus/>. See MLInputGenerator for exact definition of input variables.

- [12] ONNX Runtime. <https://onnxruntime.ai/>
- [13] Seino, T., Sugawara, R., Suehara, T., Hosokawa, R., Ootani, W., Narita, S.: Evaluation of measurement accuracy of higgs coupling constant to strange quarks in the international linear collider. Proc. LCWS2025, in preparation. (2025)
- [14] Berggren, C.M., Bliewert, B., List, J., Ntounis, D., Suehara, T., Tian, J., Torndal, J.M., Vernieri, C.: A modern reconstruction and analysis framework for the ild zhh study. Proc. LCWS2025, in preparation. (2025)