Strange Flavor Tagging with the ILD Concept at FCC-ee

Impact of V0 Reconstruction and Machine Learning-Based Approaches

Nagomi Dogishi

ICEPP, The University of Tokyo

November 4, 2026

Outline

1. Motivation
2. Physics samples & Software
3. Jet Energy Resolution
4. V0 studies
5. Summary & Future plan

Research Background and Objectives

Physical Background

- Direct measurement of the Higgs–strange Yukawa coupling ($H \to s\bar{s}$) is crucial for precision tests of the Standard Model (SM) and searches for new physics (BSM; 2HDM, Vector-like Quark models, etc.).
- At the future e^+e^- collider **FCC-ee** ($\sqrt{s} = 240$ GeV, ZH process), large-scale Higgs production is expected.

Challenges and Objectives

- Challenge: Discrimination of strange jets (s-jet) from other flavors (especially g-jet and u, d, c, b-jets) — **Strange Tagging**.
- Objective: Evaluate the jet energy resolution (JER) in FCC-ee simulations using the ILD detector concept, and conduct preliminary studies on s-jet identification using V^0 particles (K^0_S, Λ) and charged kaons ($K^{\pm}$).

ILD Detector: Configuration and Advantages

Key Components

- **VTX/VXD:** High-precision 2ndary vertexing & flavor tagging
- **TPC:** Continuous 3D tracking, PID via dE/dx
- **ECAL/HCAL:** High granularity for shower separation
- **Large-radius design:** Reduces particle overlap via PFA BR^2 optimization

Key Advantages

- **Excellent JER:** PFA yields $\sigma_E/E \approx 3$–4%
- **High-precision tracking:** $\Delta p/p^2 \sim 5 \times 10^{-5}\ (\mathrm{GeV}/c)^{-1}$

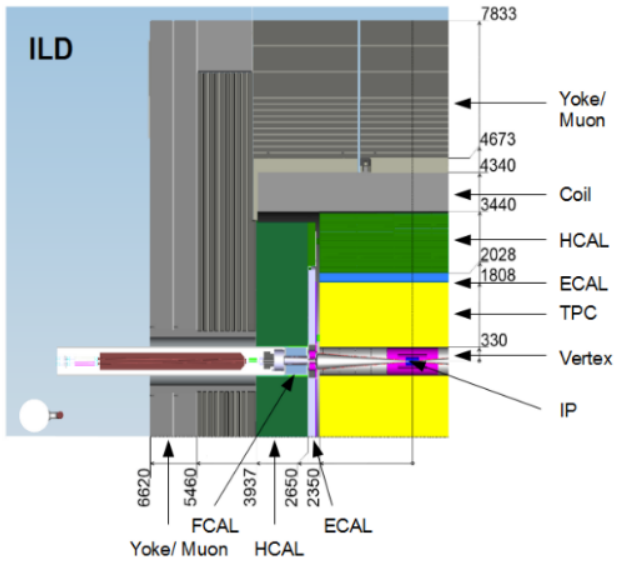


Figure: Cross-section of the ILD detector

FCC-ee Accelerator Overview and Experimental Environment

Features of FCC-ee (Future Circular Collider e^+e^-)

- A large circular e^+e^- collider with a circumference of ~ 91 km planned at CERN.
- Operates as a **Higgs Factory** ($\sqrt{s} = 240$ GeV) for precision Higgs measurements, expected to produce $\sim 10^6$ Higgs bosons via the ZH process.
- Staged operation plan: Z pole (91.2 GeV, $\sim 5 \times 10^{12}$ Z), WW threshold, ZH, $t\bar{t}$ threshold.

Advantages of e^+e^- Collisions for Flavor Tagging

- Extremely clean environment with far fewer background events (pileup, etc.) compared to hadron colliders (LHC, etc.).
- Known center-of-mass energy $\sqrt{s}$ and momenta provide an ideal environment for di-jet analyses and s-jet identification.

Physics Samples and Software Environment

Physics Samples Used ($\sqrt{s} = 240$ GeV)

- **Signal (s-jet):** $H \to s\bar{s}$ (5001 events)
- **Reference (g-jet):** $H \to gg$ (5001 events)
- **Performance evaluation:** $Z \to q\bar{q}$ (2-jets, for JER evaluation)

Software and Analysis Framework

- **Key4hep / ILCSoft (Marlin):** Full simulation and reconstruction framework.
- **PandoraPFA (PandoraPFOs):** Particle Flow reconstruction algorithm.
- **MarlinTrkTracks / LCFIPlus:** Track parameter analysis, vertex reconstruction, and flavor tagging.
- **FCCanalyses / FastJet:** Jet clustering (LCFIPlus FastJet connector).

JER Definition and Overall Evaluation ($Z \to u\bar{u}$)

Definition of JER (Jet Energy Resolution)

$$\text{JER} = \frac{\sigma_{E_{\text{jet}}}}{E_{\text{jet}}} = \frac{\text{rms}_{90}(E_{\text{jet}})}{\text{mean}_{90}(E_{\text{jet}})}$$

A robust metric using the central 90% range to suppress tail effects.

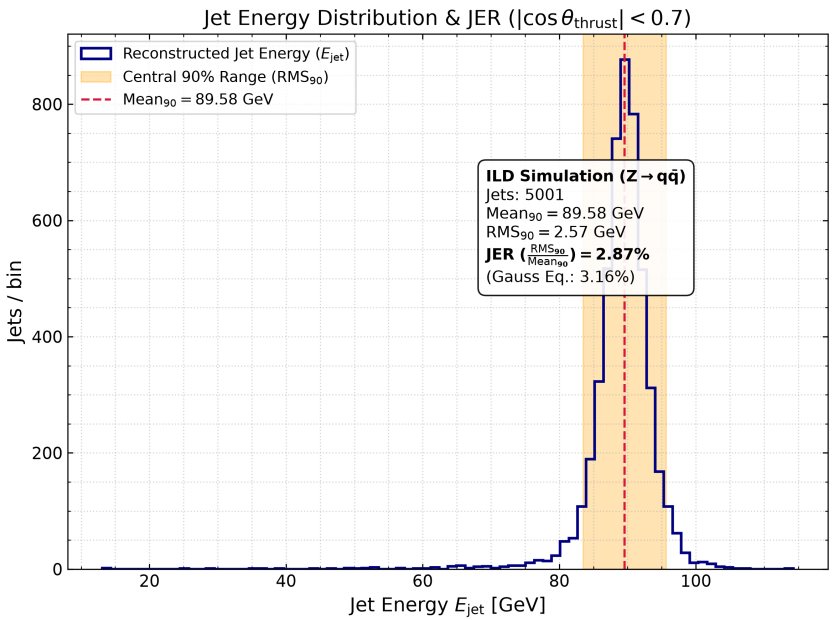


Figure: E_{vis} distribution for all events

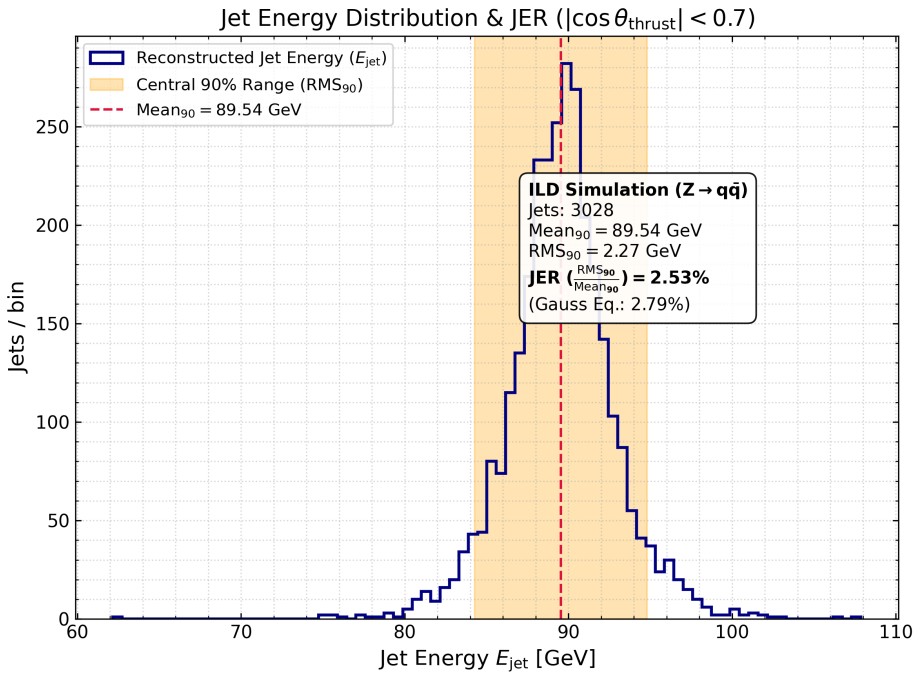
Evaluation Results (All Events)

- Sample: $Z \to u\bar{u}$ (5001 events)
- Detector: `ILD_FCCee_v01.xml`
- $\text{Mean}_{90} = 89.58$ GeV
- $\text{RMS}_{90} = 2.57$ GeV

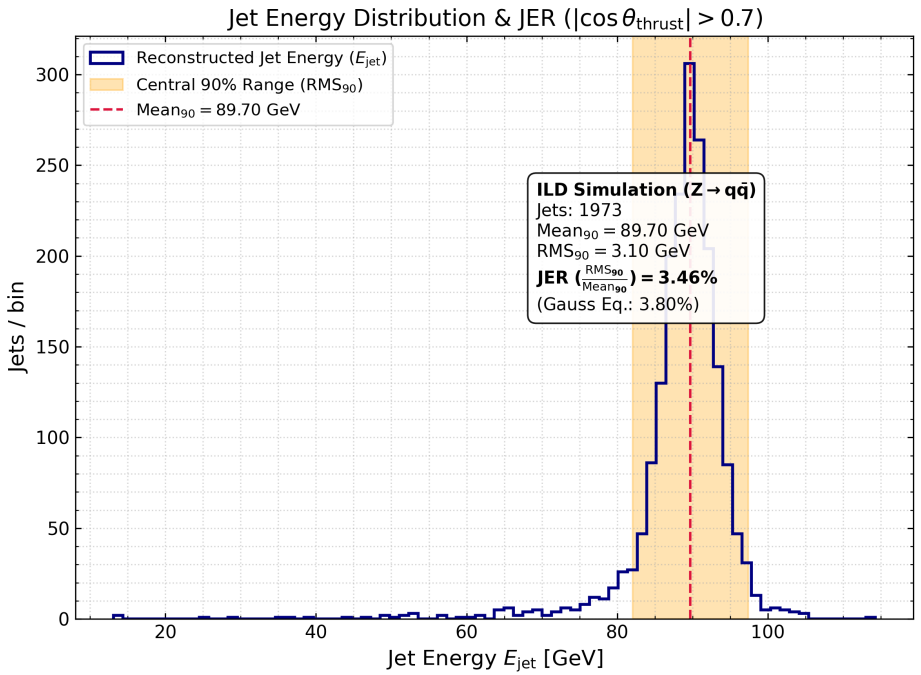
$$\textbf{JER} = \frac{\textbf{RMS}_{90}}{\textbf{Mean}_{90}} \times 1.1 = \textbf{3.16\%}$$

JER Comparison by Region (Barrel vs. Forward)

JER for the barrel region ($|\cos\theta| < 0.7$)



JER for the forward region ($|\cos\theta| > 0.7$)



Angular Correlation between V^0 Particles and Charged Strange Hadrons

Analysis Objectives, Samples, and Event Selection

Analysis Objectives

- **Objective**: Evaluate differences in V^0-$K^{\pm}$ angular correlations between $H \to s\bar{s}$ and $H \to gg$ at MC truth level
- **Key point**: Extract correlations from primary quarks using an $E > 5$ GeV cut

Observable ($\cos\theta$)

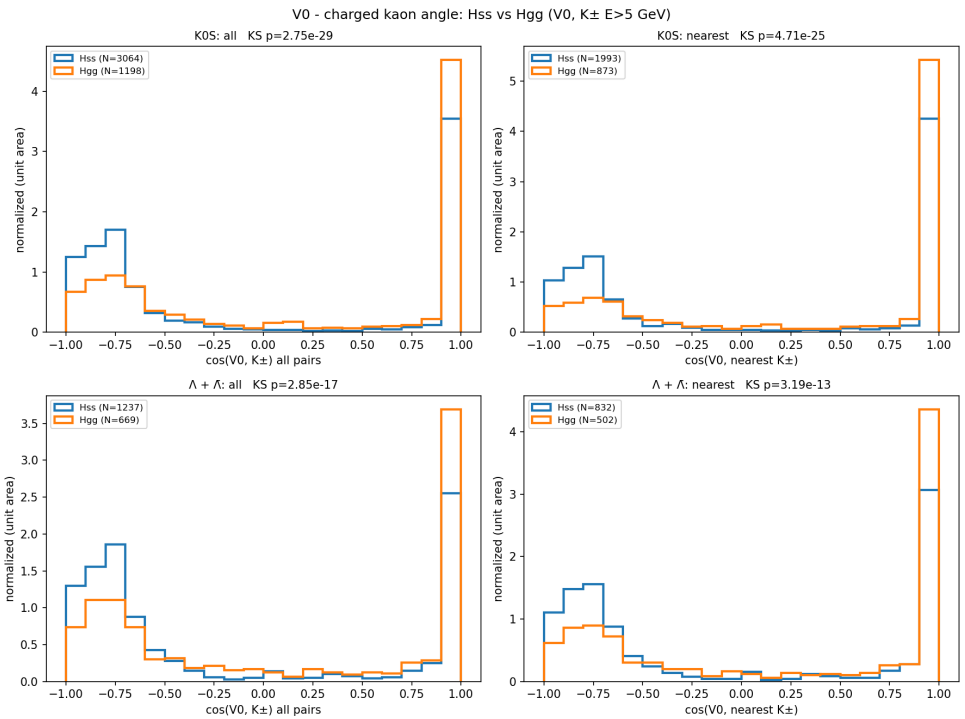
- **all**: All ($V^0, K^{\pm}$) pairs in the event
- **nearest**: Only the nearest $K^{\pm}$ to each V^0
- **opp** / **same**: Charge correlation for Λ-$K^{\pm}$ (opposite/same)

Particle Selection Criteria

1. **Higgs decay descendants**: Only `descendants` of PDG 25
2. **Secondary particle removal**: Excluded via `BIT_CREATED_IN_SIM`
3. **Selection cuts**:
   - $E > 5$ GeV (equivalent to $z > 0.08$)
   - V^0 flight radius $r_{xy} \in [1, 300]$ mm

Angular Correlation between V^0 Particles and Charged Strange Hadrons

$\cos\theta$ Distribution: $H \to s\bar{s}$ vs $H \to gg$ ($E > 5$ GeV)



Key Points ($E > 5$ GeV)

- K^0_S **(all pairs)**: $H \to s\bar{s}$ (blue) shows a strong peak on the opposite side ($\cos\theta \approx -0.8$); $H \to gg$ (orange) concentrates in the same direction ($\cos\theta \approx 1.0$).
- K^0_S **(nearest)**: $H \to s\bar{s}$ shows a significantly higher fraction of the nearest $K^{\pm}$ in the opposite jet (same-side 42.5% vs. $H \to gg$ 54.2%).
- Λ **(all pairs)**: $H \to s\bar{s}$ reflects a back-to-back structure of leading hadrons in each s, $\bar{s}$ jet.
- Λ **(nearest)**: $H \to s\bar{s}$ reaches $\sim 70\%$ on the opposite side ($\cos\theta < 0$), showing opposite behavior to $H \to gg$.

Root Cause Analysis of Low V^0 Reconstruction Efficiency

Impact of V0Finder Configuration Limits and TPC-Only Tracking

Direct Causes of Efficiency Loss

- R_{xy} **radius cut**: The `V0Finder` threshold of 30 mm eliminates nearly all V^0 PFOs with $r < 30\,\mathrm{mm}$.
- **Configuration issue**: Silicon tracking was disabled, leaving only **TPC-only tracking**.

Impact on Physics and Precision (1500 evt)

- **Daughter track loss**: No daughter tracks reconstructed (21.5%) & dropped due to DCA $> 1.5\,\mathrm{mm}$ (24.2%).
- **Resolution degradation**: Median d_0 degrades from **9 μm (with VXD)** to **0.4–2 mm (without VXD)**, adversely affecting vertex reconstruction and flavor tagging.

Silicon-Merged Tracking: Implementation and Evaluation

Workaround via `ILDReconstruction_si.py` and Performance Comparison

Implementation Strategy

- FullLDCTracking crash: The standard silicon+TPC merging module FullLDCTracking fails with the current FCC-ee geometry (missing VXD/SIT/FTD definitions), aborting with "VXD/SIT/FTD detector Not Present."
- Clupatra limitation: The silicon hit association window in Clupatra (`maxDist = 1.`) is hardcoded in C++, so the low association rate cannot be overcome via configuration parameters.

Build `MarlinTrkTracksSi` by merging existing TPC tracks with silicon (`SiTracksCT`), then run an independent `V0Finder`.

Efficiency Improvement ($R_{xy} = 1\,\mathrm{mm}$)

- K^0_S **yield**: 0.308 $\rightarrow$ **0.364** / **event** (+18%)
- **Short distance** ($r < 30\,\mathrm{mm}$): Efficiency largely recovered (1–10 mm: 0.01 $\rightarrow$ **0.19**)

Key Trade-offs

- **Purity degradation**: K^0_S purity drops from 0.81 to **0.63** (more fakes).
- **Track duplication**: Fraction of multiply-assigned tracks increases from 13% to **30%**.

Bottleneck Analysis and Future Action Plan

Limitations at Medium–Long Distances ($r = 50$–$300\,\mathrm{mm}$) and Recommended Actions

Reasons for Limited Improvement at Medium–Long Distances

- **Silicon-only limitation**: Decays at $r > 60\,\mathrm{mm}$ cannot form tracks due to insufficient hits (only 2 layers available).
- **Extrapolation failure**: `Clupatra`'s silicon association window (1 mm) is too narrow to merge with TPC tracks.

Fundamental Fixes (C++ Modifications)

- Make `Clupatra`'s 1 mm association window tunable/wider.
- Fix `FullLDCTracking` geometry definition errors for FCC-ee.

Interim Operational Guidelines

- Tighten pointing angle and mass cuts in the analysis.
- Do not feed into Pandora to prevent double-counting.

Summary and Outlook

Summary

- **JER evaluation:** Demonstrated jet energy resolution JER $\approx 3.16\%$ for $Z \to q\bar{q}$ in the ILD-FCC-ee environment.
- **V^0-based flavor discrimination:** At MC truth level, clear differences between $H \to s\bar{s}$ and $H \to gg$ are confirmed.
- **V^0 reconstruction efficiency:** Reconstruction rate of true K_S^0 is only **36.8%**; root cause is TPC-only tracking (silicon disabled), not Pandora. Adding silicon improves the $r < 30\,\mathrm{mm}$ region, but fundamental fixes require C++ modifications to `Clupatra/FullLDCTracking`.

Future Plan

1. Improve V^0 reconstruction algorithms including silicon-merged tracking.
2. Train ParT with V^0 information and investigate its effectiveness for improving strange-tagging performance.
3. Refine ML architecture and input variables.