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

LCWS 2026

1. 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 ($K^\pm$).

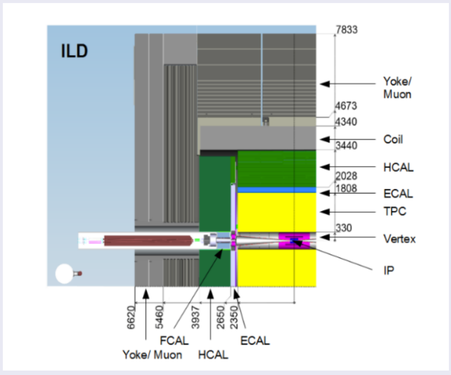
2. ILD Detector and FCC-ee

Key Components and Advantages

- **VTX/VXD:** High-precision secondary vertexing and flavor tagging
- **TPC:** Continuous 3D tracking with PID via dE/dx
- **Excellent JER:** $\sigma_E/E \approx 3$–4% via PFA
- **Tracking:** $\Delta p/p^2 \sim 5 \times 10^{-5}\ (\mathrm{GeV}/c)^{-1}$

FCC-ee (Future Circular Collider)

- Circular e^+e^- collider, circumference ~ 91 km (CERN)
- Higgs Factory at $\sqrt{s} = 240$ GeV: $\sim 10^6$ ZH events
- Staged runs: Z pole, WW, ZH, $t\bar{t}$ threshold
- Clean environment (low pileup); known $\sqrt{s}$ ideal for s-jet ID



Cross-section of the ILD detector

3. Physics Samples and Software

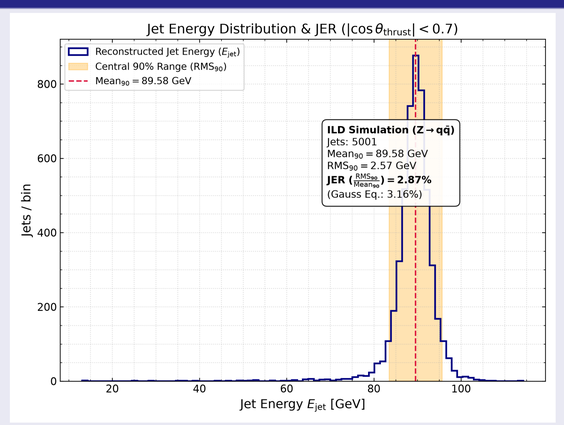
Physics Samples ($\sqrt{s} = 240$ GeV)

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

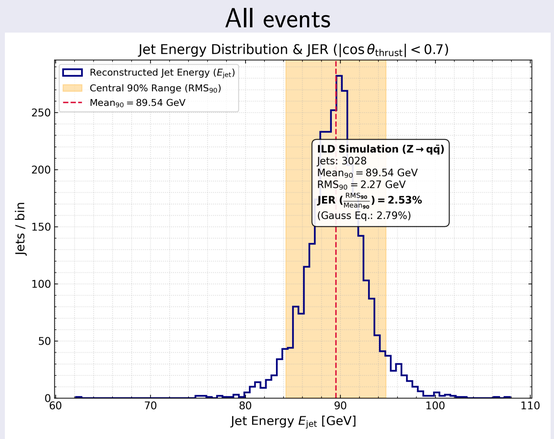
Software and Framework

- **Key4hep / ILCSoft (Marlin):** Full simulation & reconstruction
- **PandoraPFA:** Particle Flow reconstruction
- **MarlinTrkTracks / LCFIPlus:** Tracking, vertexing & flavor tagging
- **FCCanalyses / FastJet:** Jet clustering

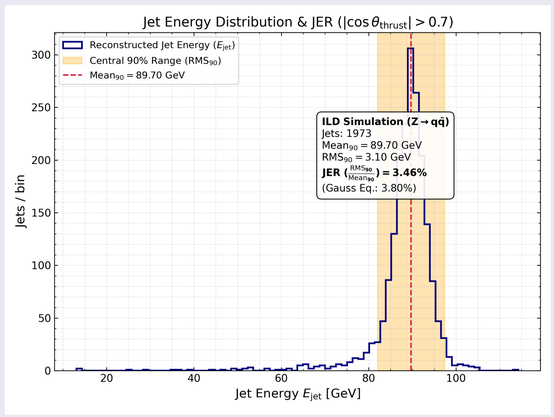
4. Jet Energy Resolution (JER)



All events



Barrel ($|\cos\theta| < 0.7$)



Forward ($|\cos\theta| > 0.7$)

Definition of JER(Jet Energy Resolution)

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

A robust metric using the central 90% of the jet energy distribution to suppress tail effects.

Evaluation Results (All Events)

- Sample: $Z \to u\bar{u}$, 5001 evt
- Detector: `ILD_FCCee_v01`
- $\mathrm{Mean}_{90} = 89.58$ GeV
- $\mathrm{RMS}_{90} = 2.57$ GeV

$$\mathbf{JER = 3.16\%}$$

Both overall and region-specific JER are consistent with PFA expectations (~ 3–4%). **Detector model and simulation validated.**

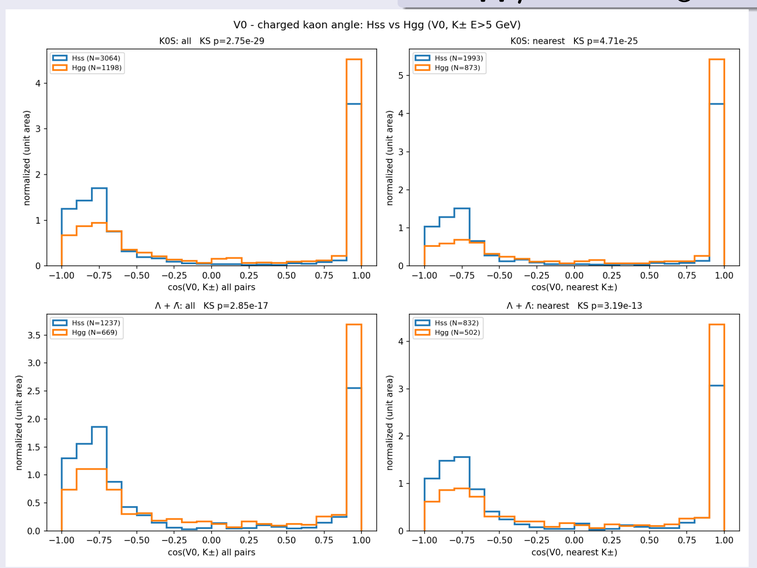
5. V^0–Charged Kaon Angular Correlation

Analysis Objectives

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

Particle Selection

- Higgs decay descendants (PDG 25) only
- Remove secondaries via `BIT_CREATED_IN_SIM`
- $E > 5$ GeV, $r_{xy} \in [1, 300]$ mm
- Observables: **all** pairs, **nearest** $K^\pm$, **opp/same** charge



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

- K^0_S **(all)**: $H \to s\bar{s}$ peaks opposite ($\cos\theta \approx -0.8$); $H \to gg$ peaks same-side ($\cos\theta \approx 1.0$)
- K^0_S **(nearest)**: same-side fraction 42.5% vs. 54.2% — clear separation
- Λ **(nearest)**: $H \to s\bar{s}$ reaches $\sim 70\%$ opposite side ($\cos\theta < 0$), opposite to $H \to gg$

Clear $H \to s\bar{s}$ / $H \to gg$ discrimination via angular correlations confirmed at truth level

6. Root Cause Analysis of Low V^0 Efficiency

Direct Causes

- R_{xy} **cut**: `VOFinder` threshold 30 mm eliminates nearly all V^0 PFOs with $r < 30$ mm
- **TPC-only tracking**: silicon information disabled — *root cause*

True K^0_S breakdown (1222, 1500 evt)

- No daughter tracks reconstructed: **21.5%**
- DCA > 1.5 mm: **24.2%**
- R_{xy} cut: $\approx 4\%$
- Reconstructed: **36.8%**

- Median d_0: **9 μm** (with VXD) vs. **0.4–2 mm** (without) — adversely affects vertex reconstruction and flavor tagging
- Only **38%** of primary-vertex tracks have VXD hits (PR #191 configuration issue)

7. Silicon-Merged Tracking: Implementation and Evaluation

Implementation Strategy

- **FullLDCTracking**: crashes with FCC-ee geometry ("VXD/SIT/FTD detector Not Present")
- **Clupatra**: silicon association window (`maxDist`=1 mm) hardcoded in C++ — cannot be overridden via config
- **Workaround**: TPC + `SiTracksCT` → `MarlinTrkTracksSi`; run independent `V0Finder`

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

- K^0_S yield: $0.308 \to$ **0.364 / event** ($+18\%$)
- $r < 30$ mm efficiency: $0.01 \to$ **0.19** at 1–10 mm

Key Trade-offs

- K^0_S purity: $0.81 \to$ **0.63** (more fakes)
- Track duplication: $13\% \to$ **30%**
- Must *not* feed into Pandora (double-counting)

Limitations at $r = 50$–300 mm

- Decays at $r > 60$ mm: $\leq$2 Si layers → no standalone track
- Clupatra 1 mm window too narrow to associate TPC+Si hits

Fundamental Fixes (C++)

- Parameterize/widen `Clupatra`'s association window
- Fix `FullLDCTracking` geometry for FCC-ee

8. 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^0_S is only **36.8%**; root cause is TPC-only tracking (silicon disabled), not Pandora. Adding silicon improves the $r < 30$ 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.