Physics-Informed GNN for $\tau^+\tau^-$ Vertex Reconstruction at STCF

Presenter: Chentao Bao[1]

[1]Zhejiang University

超级陶粲装置研讨会 西安 July 2, 2026

Outline

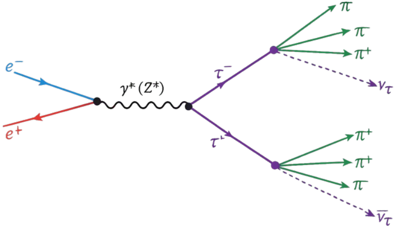
Super Tau-Charm Facility

Facility Parameters

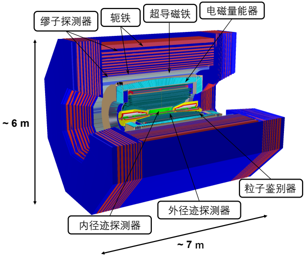
- CM energy $\sqrt{s} = 2\text{–}7$ GeV
- Luminosity $\mathcal{L} \sim 10^{35}$ cm^{-2}s^{-1}
- Beam crossing angle 2×30 mrad

Core Process

- $e^+e^- \to \tau^+\tau^-$ at 7.0 GeV
- $\tau^+ \to \pi^+\pi^+\pi^- + \bar{\nu}_\tau, \tau^- \to \pi^+\pi^-\pi^- + \nu_\tau$ ($\sim 1.2\%$)



STCF Facility Schematic

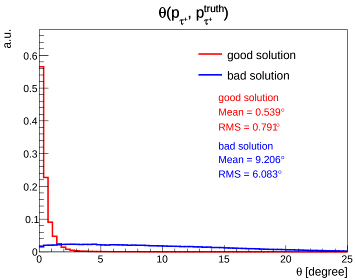


STCF Detector 3D Schematic

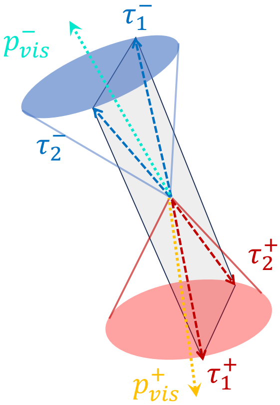
Two-Fold Ambiguity in τ Momentum

Origin of the ambiguity

- Undetected neutrinos leave the final-state four-momentum incomplete.
- In the $\tau^+\tau^-$ analysis, 6 unknown neutrino momentum components are constrained only by four-momentum conservation and the τ on-shell conditions.
- The resulting system is quadratic and generically admits **two real solutions**.



Geometric picture



Two possible flight directions are the intersections of the two cones.

Physics Background

Handling strategy

- Use **vertex information** to break the degeneracy and select the physical momentum solution.

$a_1 a_1$ channel advantage:

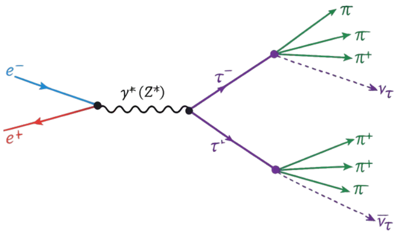
- $\tau^{\pm} \to 3\pi + \nu_\tau$ provides 6 charged tracks in total, giving strong geometric constraints for vertex reconstruction.

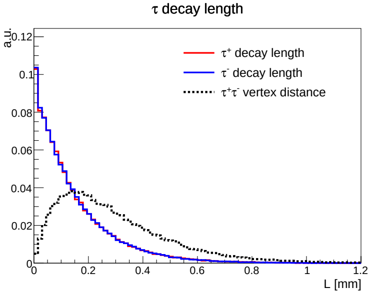
Basic Properties

- Mass $m_\tau = 1776.86 \pm 0.12$ MeV/c^2
- Lifetime $\tau = 290.3 \pm 0.5$ fs, $c\tau = 87.03\ \mu$m

Decay Vertex and Flight Length

- Boost $\beta\gamma = p_\tau / m_\tau \sim 0.5$–$2$
- Average flight distance: $L = \beta\gamma \cdot c\tau \sim 200\ \mu$m
- Vertex Distribution: $\Sigma_{x,y}^{\text{truth}} \sim 127\ \mu$m, $\Sigma_z^{\text{truth}} \sim 565\ \mu$m





Input and Dataset

Helix Track Parameterization

Parameter	Symbol	Meaning	Avg resolution
Impact parameter	d_0	POCA to z axis distance	$\Delta d_0 \sim 89.47\ \mu\text{m}$
Azimuthal angle	ϕ_0	Momentum direction in xy at POCA	$\Delta\phi_0 \sim 2.01$ mrad
Curvature	κ	$q/p_T \cdot \alpha B$	$\Delta\kappa \sim 0.01129\ \text{GeV}^{-1}$
Longitudinal offset	z_0	z coordinate at POCA	$\Delta z_0 \sim 169.32\ \mu\text{m}$
Dip angle parameter	$\tan\lambda$	p_z/p_T	$\Delta \tan\lambda \sim 0.00297$

Dataset

- Source: OSCAR full chain simulation
- Pre-Analysis: track selection and **true pairing**; 6 charged tracks have been correctly assigned to τ^+ and τ^-
- Total events: 149,610 ; Train/Val/Test: 60% / 30% / 10%

Traditional Iterative Vertex Fit

vertexfit.cc (**Gauss-Newton Weighted Least Squares**) minimizes track–vertex residuals:

$$g_{xy} = \Delta y p_x - \Delta x p_y - \frac{a}{2}(\Delta x^2 + \Delta y^2)$$

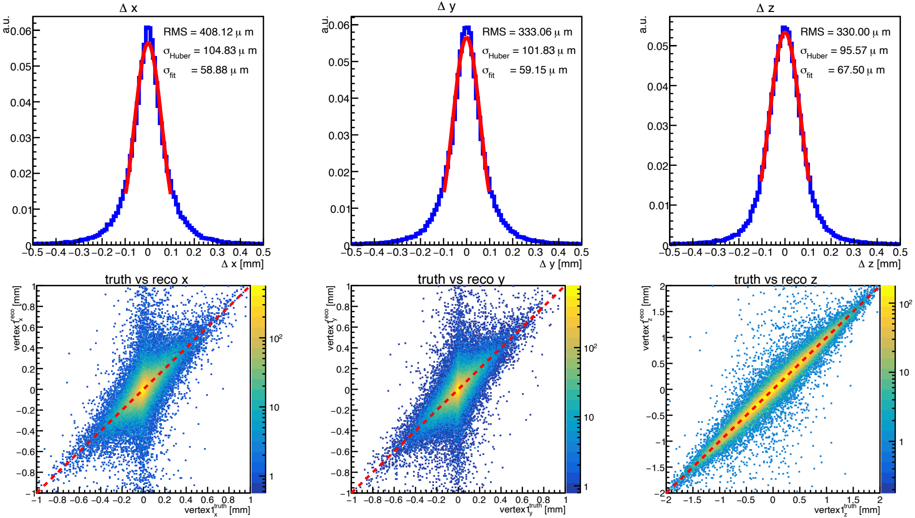
$$g_z = \Delta z - \frac{p_z}{a}\arcsin(J), \quad J = \frac{a(\Delta x p_x + \Delta y p_y)}{p_T^2}$$

where $a = q\alpha B$, $\Delta \mathbf{x} = \mathbf{V} - \mathbf{x}_{\text{track}}$. **Iterative update:**

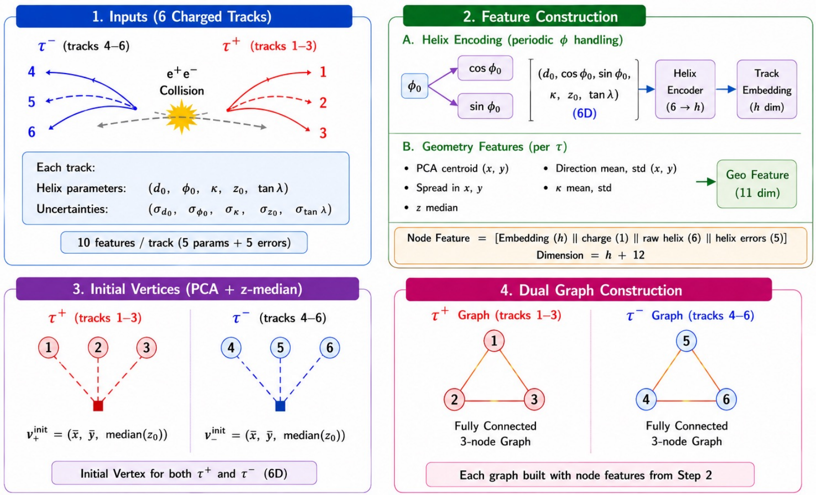
1. Initialize $\mathbf{V}_0$ from track centroid.
2. Compute residuals $g_{xy}^{(i)}, g_z^{(i)}$ and weighted vector $\mathbf{r}$.
3. Solve $\Delta \mathbf{V} = (J^T W J)^{-1} J^T W \mathbf{r}$.
4. Update $\mathbf{V} \leftarrow \mathbf{V} + \Delta \mathbf{V}$ until convergence.

VertexFit Result

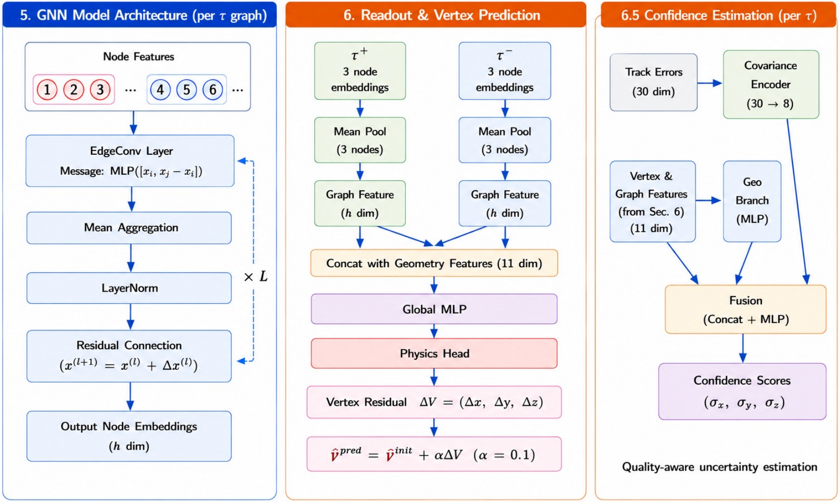
- σ_{fit}: Gaussian fit, iteratively narrowed to $\pm 1.5\sigma$
- σ_{Huber}: Huber Iteratively Reweighted Least Squares robust estimation, central 95% data



Physics-Informed GNN Architecture Overview



Physics-Informed GNN Architecture Overview



Baseline Loss

MSE Loss

$$\mathcal{L}_{\text{MSE}} = \frac{1}{B}\sum_{b=1}^{B} \|\hat{\mathbf{V}}_b - \mathbf{V}_b^{\text{true}}\|_w^2$$

- B: batch size; $\hat{\mathbf{V}}_b$, $\mathbf{V}_b^{\text{true}}$: predicted and true vertices; w: coordinate-weighted norm
- Standard regression loss measuring weighted Euclidean distance

Weighted Helix Distance Loss (WLS)

$$\mathcal{L}_{\text{WLS}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{6} w_{b,i} \cdot d_{\text{helix}}(t_{b,i}, \hat{\mathbf{V}}_b)^2$$

- $w_{b,i}$: covariance-driven weight for the i-th track in event b
- $d_{\text{helix}}(t_{b,i}, \hat{\mathbf{V}}_b)$: closest distance from predicted vertex to track helix
- Geometric consistency term; weights normalized per event

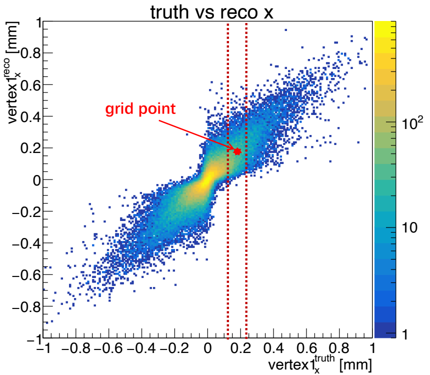
Conditional Correction

Conditional unbiased: Conditional expectation equals truth at each true position

$$\mathbb{E}[\hat{\mathbf{V}} \mid \mathbf{V}^{\text{true}} = \mathbf{v}] = \mathbf{v}, \quad \forall \mathbf{v}$$

Local Linear Regression (LLR) Core Method

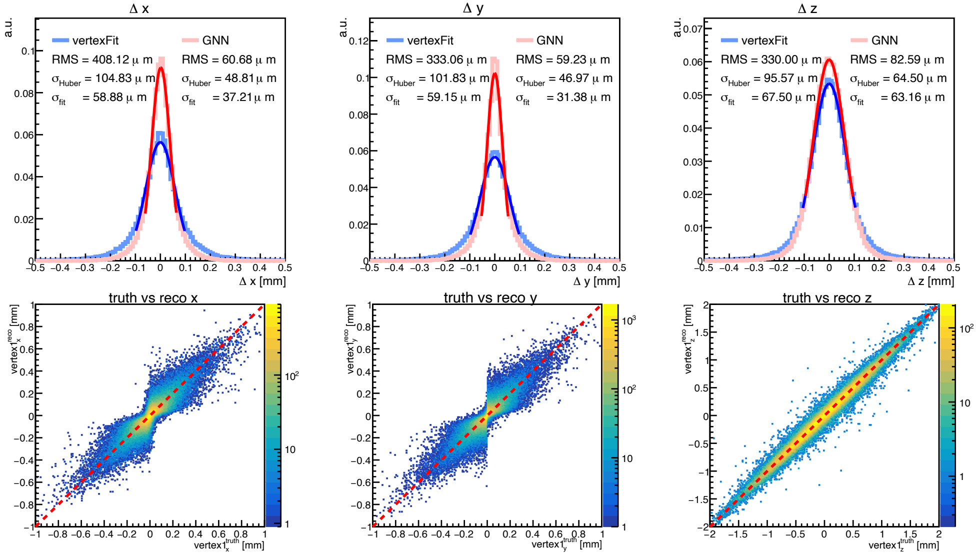
- For each **grid point**, fit local linear model with event (t_j, p_j) in window
- Local model: $y_i = \beta_0 + \beta_1(x_i - g)$ (x_i: truth, y_i: pred, g: grid center)
- $\textbf{bias}(g) = \beta_0 - g$
- $\textbf{resolution}(g) = \sqrt{\text{Var}[p_j - y(t_j)]}$



Confidence Loss

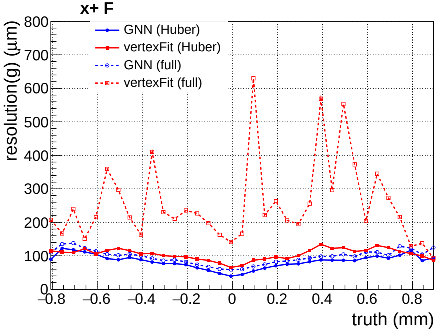
- **Quality:** $Q = |\hat{v} - v^{\text{truth}}|$
- **Ranking Loss:** order confidence scores by reconstruction error

GNN Result

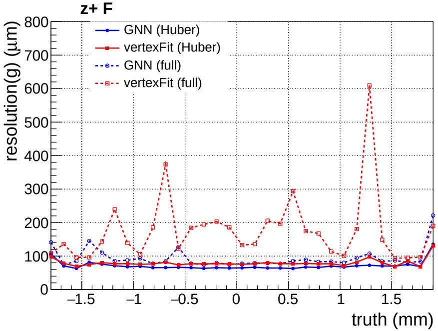

Δ x
Δ y
Δ z
a.u.
vertexFit
GNN
RMS = 408.12 μ m
σ_{Huber} = 104.83 μ m
σ_{fit} = 58.88 μ m
RMS = 60.68 μ m
σ_{Huber} = 48.81 μ m
σ_{fit} = 37.21 μ m
RMS = 333.06 μ m
σ_{Huber} = 101.83 μ m
σ_{fit} = 59.15 μ m
RMS = 59.23 μ m
σ_{Huber} = 46.97 μ m
σ_{fit} = 31.38 μ m
RMS = 330.00 μ m
σ_{Huber} = 95.57 μ m
σ_{fit} = 67.50 μ m
RMS = 82.59 μ m
σ_{Huber} = 64.50 μ m
σ_{fit} = 63.16 μ m
Δ x [mm]
Δ y [mm]
Δ z [mm]
truth vs reco x
truth vs reco y
truth vs reco z
$vertex1_x^{reco}$ [mm]
$vertex1_x^{truth}$ [mm]
$vertex1_y^{reco}$ [mm]
$vertex1_y^{truth}$ [mm]
$vertex1_z^{reco}$ [mm]
$vertex1_z^{truth}$ [mm]

resolution(g)

Def: resolution$(g) = \sqrt{\mathrm{Var}[\hat{r} \mid t = g]}$. Smaller **resolution**$(g) \rightarrow$ better local resolution.



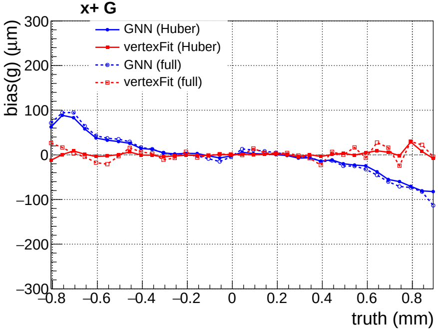
x direction



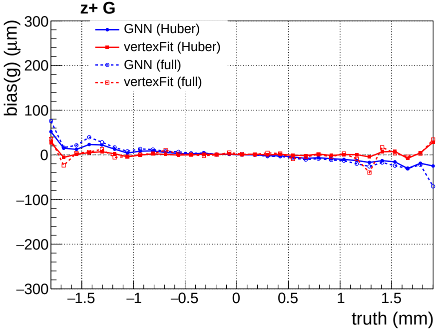
z direction

bias(g)

Def: **bias**$(g) = \mathbb{E}[\hat{r} \mid t = g] - g$. **bias**$(g) \approx 0 \rightarrow$ local unbiased.



x direction



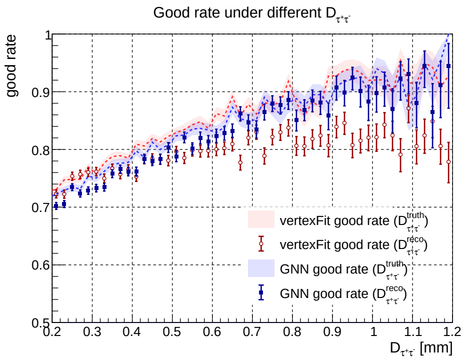
z direction

τ Double-Solution Selection Good Rate

- In the $\tau\tau$ analysis, the reconstructed vertex displacement $v_{\tau^+}^{\text{reco}} - v_{\tau^-}^{\text{reco}}$ is used to select the τ momentum solution.
- A sufficiently large $\tau^+\tau^-$ decay-length separation is a key handle for particle pairing and momentum selection.
- Define good event: $\Delta R(p_\tau^{truth}, p_\tau^{selected}) < 0.1$ or $\Delta R(p_\tau^{truth}, p_\tau^{selected}) < \Delta R(p_\tau^{truth}, p_\tau^{unselected})$.

Method	$D_{\tau^+\tau^-}^{\text{reco}} > 450\ \mu$m good rate
vertexfit	79.6%
GNN	82.6%
GNN (conf>0.3)	83.9%

$D_{\tau^+\tau^-}$ denotes the distance between the τ^+ and τ^- decay vertices.



Summary

1. Developed a **Physics-Informed GNN model** for $\tau^+\tau^-$ decay vertex reconstruction.
2. Established a **baseline training workflow** under OSCAR simulation framework.
3. The trained model substantially reduces **outlier** residuals compared with the VertexFit.
4. Future plan:
   - Augment far-end samples and impose stronger far-end constraints to address the long-flight-distance training bias;
   - Refine the Confidence network to better identify and suppress outlier events.
5. **Prospect:**
   - Extend to πa, ρa, etc. decay channels.
   - Extend to momentum reconstruction.

Chentao Bao

Thank You!

Questions & Discussion

Code Repository: /ustcfs4/STCF/STCFUser/ctbao && github

Backup

Confidence Loss: Current Practice

Quality

$$Q_d = \left|\hat{v}_d - v_d^{\text{truth}}\right|$$

- $\hat{v}_d$: predicted coordinate $d \in \{x, y, z\}$ of the vertex
- v_d^{truth}: true coordinate d
- Larger $Q_d \Rightarrow$ larger reconstruction error

Kernel-Weighted Pairwise Ranking

$$K_{ij} = \exp\left(-\frac{\|g_i - g_j\|^2}{2h^2}\right)$$

- g_i, g_j: predicted (or true) vertex positions of events i, j
- h: adaptive bandwidth via Silverman's rule
- K_{ij} gives higher weight to nearby pairs in prediction space

Margin Ranking Loss

$$\mathcal{L}_{\text{rank}} = \sum_{i,j} K_{ij}\,\text{ReLU}\big(m - \text{sign}_{ij}(s_i - s_j)\big)$$

- s_i, s_j: predicted confidence scores
- $\text{sign}_{ij} = \text{sign}(Q_i - Q_j)$
- m: fixed margin; penalizes wrong confidence ordering

Confidence Loss: Future Improvements

Squared Quality

$$Q \rightarrow Q^2$$

- Amplifies the quality gap for extreme outliers

Pair Weight by Max-Q

$$w_{ij} = 1 + \frac{\max(Q_i, Q_j)}{\sigma_Q}$$

- σ_Q: scale of quality distribution
- High-Q pairs contribute more to the loss

Adaptive Margin

$$m_{ij} = m_0 \left(1 + \frac{|Q_i - Q_j|}{\sigma_Q}\right)$$

- Larger Q-gap requires larger confidence separation

Trim Objective

$$\mathcal{L}_{\text{trim}} = \frac{1}{|S_{\text{keep}}|} \sum_{i \in S_{\text{keep}}} Q_i$$

- S_{keep}: top-confidence events kept after trimming (e.g. 90%)