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

Presenter: Chentao Bao[1]

[1]Zhejiang University

Workshop on Tau Physics at Belle II and STCF August 31, 2026

Outline

Outline

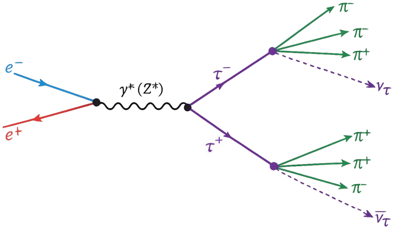
Super Tau-Charm Facility

Facility Parameters

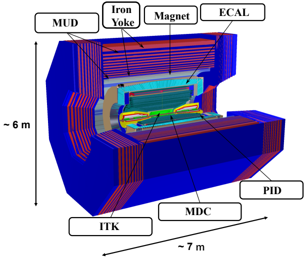
- CM energy $\sqrt{s} = 2\text{–}7$ GeV
- Luminosity $\mathcal{L} \sim 10^{35}$ cm^{-2}s^{-1}
- Beam crossing angle 2×30 mrad

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 Schematic

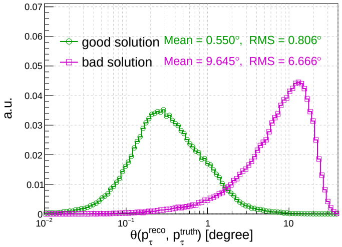


STCF Detector 3D Schematic

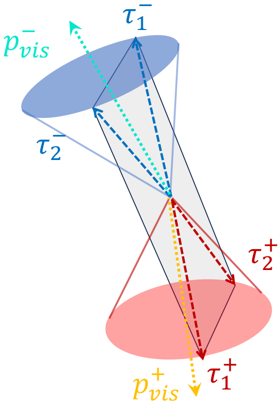
Two-Fold Ambiguity in τ Momentum

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

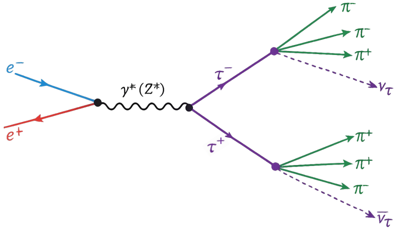


Two possible flight directions are the intersections of the two cones.

Physics Background

Handling strategy

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

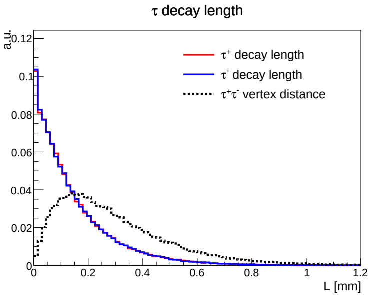


$a_1 a_1$ channel advantage:

- $\tau^\pm \to 3\pi + \nu_\tau$ 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

Motivation

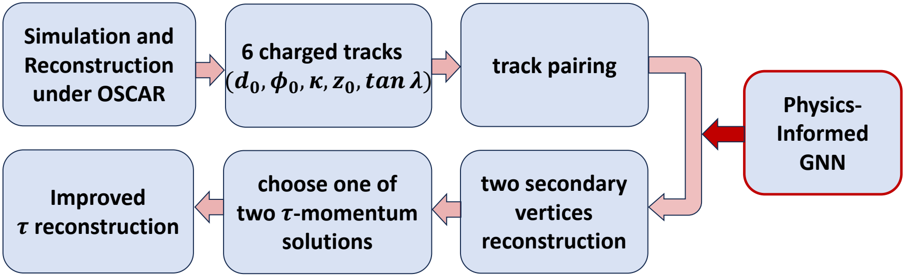
1. **Implicit global information**
   - The network learns implicit global information from track patterns (**back-to-back correlation of the τ pair**, etc.)
   - Works in the degenerate region (**short flight length** $L \lesssim$ vertex resolution), where collinearity constraints fail

2. **Uncertainty output**
   - The network uncertainty is learned from the **global empirical distribution** of prediction errors, allowing **effective selection of well-predicted events**

3. **Weakly-constrained systems**
   - For single-prong sides of $(3\pi\nu)(\pi\nu)$, $(\pi\nu)(\pi\nu)$, only one track is available, so the **secondary vertex lacks sufficient geometric constraints** for classical vertex fit

Outline

Workflow Overview



Scope

- Background channels (e.g. $q\bar{q}$) not yet included
- Systematic uncertainties not yet evaluated
- ν reconstruction & double-solution selection as downstream applications

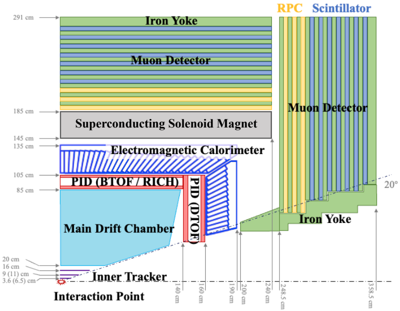
Sections

- Pairing & VertexFit $\rightarrow$ Sec. 2
- GNN and training $\rightarrow$ Sec. 3 & 4

MC Data

Simulation and Reconstruction

- OSCAR full chain simulation
- version: offline 2.6.2
- Geometry: 3 layers ITKM + MDC
- Future update: 4 layers ITKM

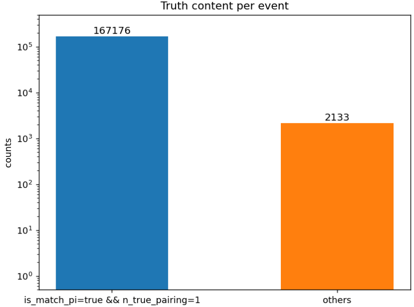


Helix Track Parameterization

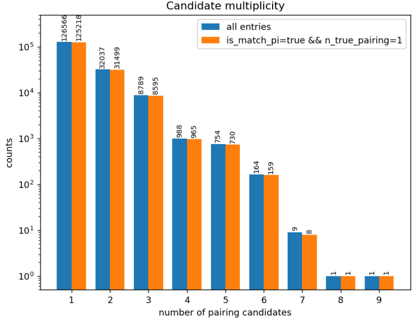
Parameter	Symbol	Meaning	Avg resolution
Impact parameter	d_0	POCA to z axis distance	$\Delta d_0 \sim 89.47\,\mu$m
Azimuthal angle	ϕ_0	Momentum direction in xy at POCA	$\Delta\phi_0 \sim 2.01$ mrad
Curvature	κ	$q/p_T \cdot \alpha B$	$\Delta\kappa \sim 0.01129\,\mathrm{GeV}^{-1}$
Longitudinal offset	z_0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

Pairing Candidates per Event

Sample: MC Data, 169,309 events of $\tau^+\tau^- \to \pi^+\pi^+\pi^-\nu + \pi^+\pi^-\pi^-\bar{\nu}$



Truth content per event



Candidate multiplicity

- is_match_pi: all six reconstructed pion tracks of the event are correctly matched to the MC-truth final-state pions.
- $n_{\text{true pairing}} = 1$: exactly one candidate in the event is the true pairing (is_true_pairing), so the correct choice exists and is unique.

Pairing Score Design

Log-likelihood ratio for a_1^+/a_1^-

$$\text{LLR}(x) = \log \frac{N_{\text{true}}(\text{bin of } x)}{N_{\text{false}}(\text{bin of } x)}$$

Coefficients: logistic regression

- Sample: one binary sample per candidate; $\mathbf{x} = [\text{LLR}_{a_1^+}, \text{LLR}_{a_1^-}, \chi^2_{\nu_1}, \chi^2_{\nu_2}, \chi^2_{km}]$, y = is_true_pairing
- Model: $P(y{=}1|\mathbf{x}) = \text{sigmoid}(\sum_i w_i \hat{x}_i + b)$, $\hat{x}_i = (x_i - \mu_i)/\sigma_i$ standardised
- Loss: cross-entropy, convex $\Rightarrow$ unique global optimum, no tuning freedom

Pairing score

$$\text{score} = 0.004 + \chi^2_{km} + 10^{-6}\,(\chi^2_{\nu_1} + \chi^2_{\nu_2}) - 10^{-3}\left[\text{LLR}_{a_1^+}(a_1^+) + \text{LLR}_{a_1^-}(a_1^-)\right]$$

- Renormalised to χ^2_{km} coefficient $= 1$; $+0.004$ keeps score positive.

Cutflow

Cutflow of $\tau \to 3\pi$ signal sample:

Selection step	Signal	Efficiency (%)
Total	1,183,495	—
6 charged tracks, total charge $= 0$	626,685	52.95
$\pi^+ = 3,\ \pi^- = 3$	278,394	44.42
passed particle pairing	165,204	59.34
τ reconstruction	157,768	95.50
Overall	157,768	**13.33**

Truth check (signal)	rate
true match	99.43%
true pairing	**98.57%**

- True pairing rate **98.57%** with overall efficiency **13.33%** on the $\tau \to 3\pi$ signal sample.

Traditional Iterative Vertex Fit

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

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

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

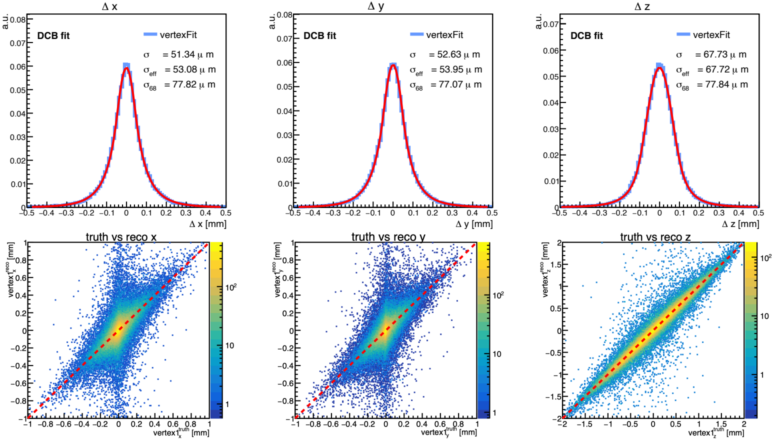
where $a = q\alpha B$, $\Delta \mathbf{x} = \mathbf{V} - \mathbf{x}_{\text{track}}$. **Iterative update:**

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_{DCB}: **Double Crystal Ball fit** to the residual distribution; Gaussian-core width σ taken as the resolution

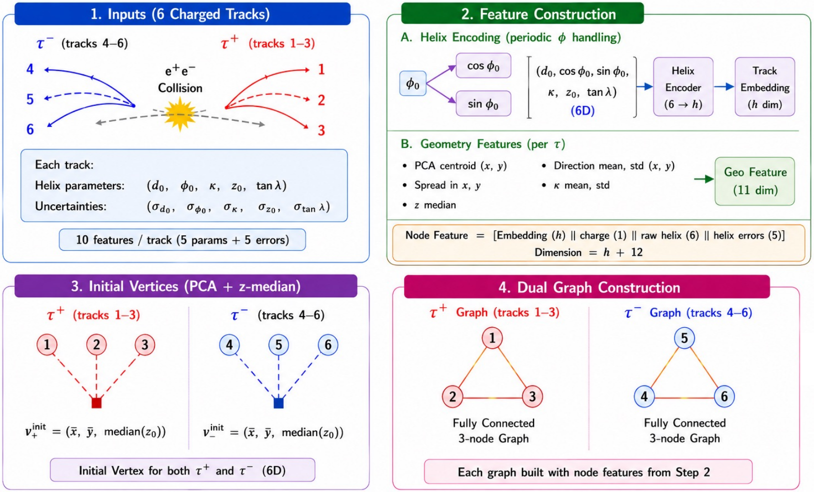
$$f(x) \propto \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \left(\left|\frac{x-\mu}{\sigma}\right| \le \alpha_{L,R}\right), \qquad A_{L,R}\left(B_{L,R} \mp \frac{x-\mu}{\sigma}\right)^{-n_{L,R}} \text{ (left / right tails)}$$



Outline

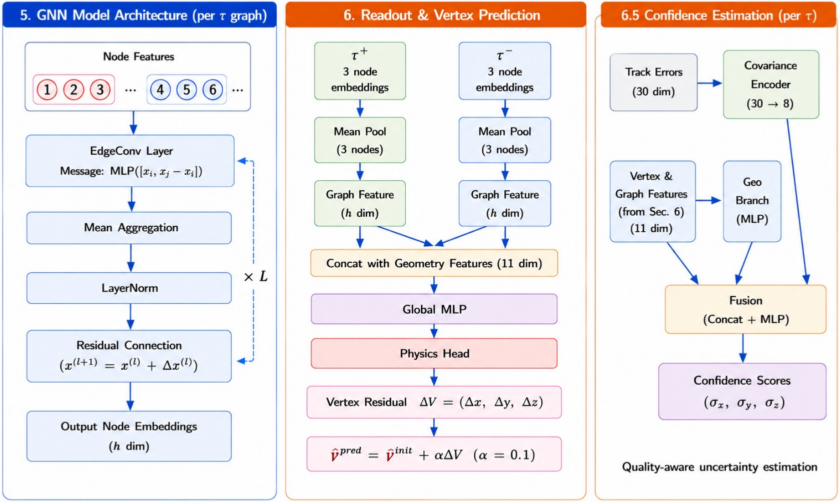
Physics-Informed GNN Architecture Overview

Chentao Bao



Physics-Informed GNN Architecture Overview

Chentao Bao



Outline

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

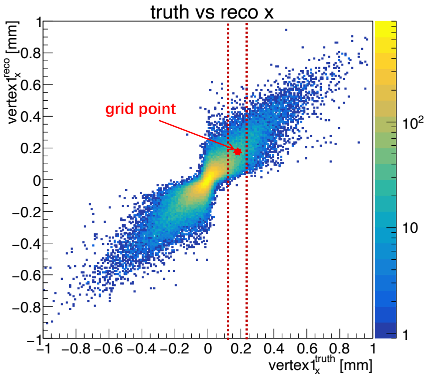
Conditional Correction

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)
- $\textbf{bias}(g) = \beta_0 - g$
- $\textbf{resolution}(g) = \sqrt{\mathrm{Var}[p_j - y(t_j)]}$



Correction Loss

- $\mathcal{L}_{\text{resolution}} = w_1 \sum_g w(g) \cdot \textbf{resolution}(g)^2$ ($w(g)$: event density at grid point g)
- $\mathcal{L}_{\text{bias}} = w_2 \sum_g w(g) \cdot (\beta_0(g) - g)^2 + w_3 \sum_g w(g) \cdot (\beta_1(g) - 1)^2$

Confidence Output

Training setup

- Backbone frozen; only the confidence head is trained

Quality: normalized error

$$Q_{\text{norm}} = |\hat{v} - v^{\text{truth}}| \,/\, \sigma(v^{\text{truth}})$$

- $\sigma(v^{\text{truth}})$: resolution function of v^{truth} from last step — **comparable across coordinates and truth positions**

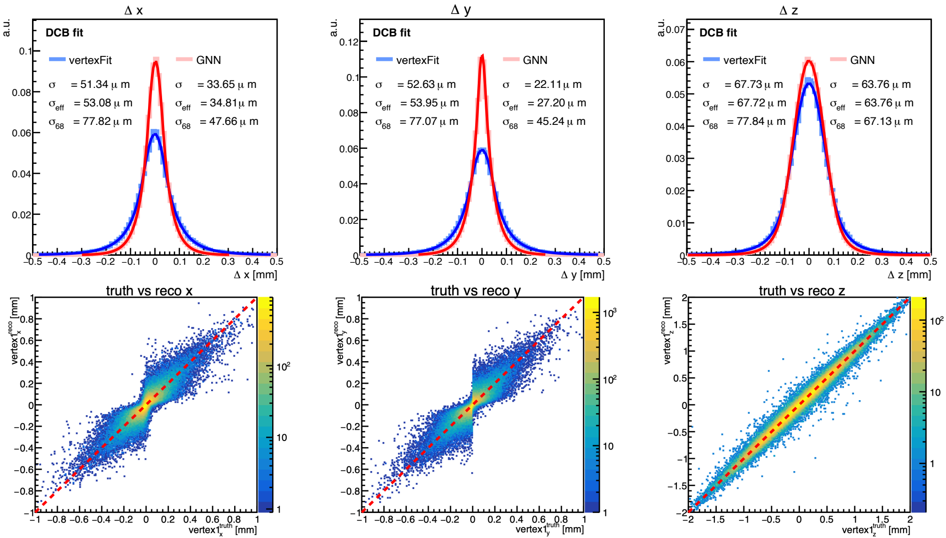
Error-magnitude-weighted ranking loss

$$\mathcal{L}_{\text{rank}} = \frac{\sum_{i<j} w_{ij}\, \text{ReLU}\big(m - \text{sign}(q_j - q_i)\,(s_i - s_j)\big)}{\sum_{i<j} w_{ij}}, \qquad w_{ij} = \tfrac{1}{2}(q_i + q_j), \quad m = 0.1$$

- s_i: confidence score; larger-error pairs get heavier penalty — **gradient pressure concentrates on the tail**, aligned with outlier rejection

GNN Result

Chentao Bao



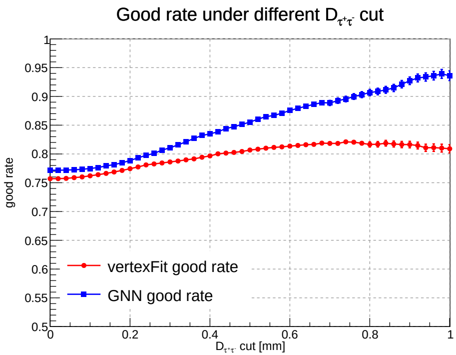
Outline

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 $\Delta R(p_\tau^{truth}, p_\tau^{selected}) < \Delta R(p_\tau^{truth}, p_\tau^{unselected})$.

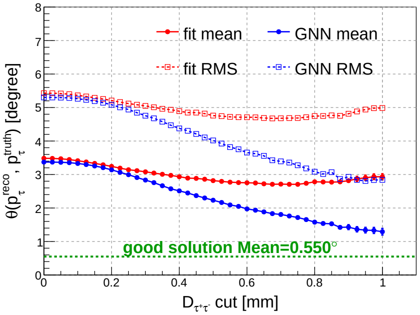
Method	$D^{\text{reco}}_{\tau^+\tau^-} > 450\ \mu$m good rate
vertexfit	79.6%
GNN	82.6%
GNN (conf>0.3)	83.9%

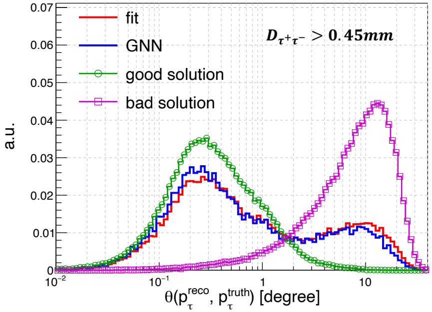
$D_{\tau^+\tau^-}$ denotes the distance between the τ^+ and τ^- decay vertices.



τ Momentum Reconstruction

- 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.
- Use $\theta(p_\tau^{\text{reco}}, p_\tau^{\text{truth}})$ to measure the performance of τ reconstruction.





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. **GNN result:**
   - The trained model substantially reduces **outlier** residuals compared with the VertexFit.
   - The network has learned the **back-to-back** correlation of the τ pair.
4. **Prospect:**
   - Apply the GNN method to weakly-constrained decay channels such as $(\pi\nu)(a_1\nu)$, $(\rho\nu)(a_1\nu)$, etc.

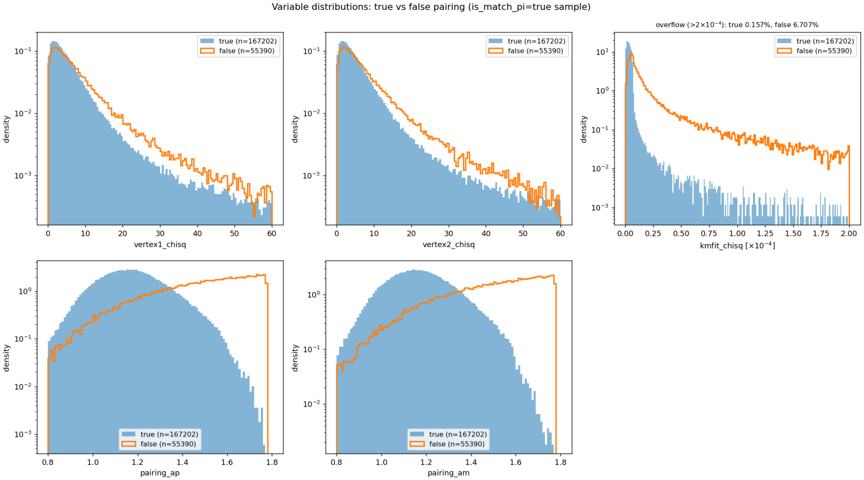
Thank You!

Questions & Discussion

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

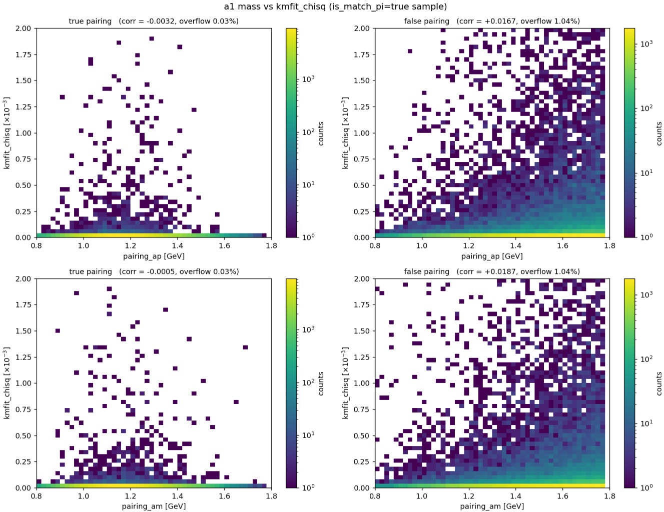
Backup

Pairing


Variable distributions: true vs false pairing (is_match_pi=true sample)
overflow (>2×10⁻⁴): true 0.157%, false 6.707%
true (n=167202)
false (n=55390)
density
vertex1_chisq
vertex2_chisq
kmfit_chisq [×10⁻⁴]
pairing_ap
pairing_am

Pairing

$$\mathrm{corr} = \frac{\mathrm{cov}\left(a_p, \chi^2_{km}\right)}{\sigma_{a_p}\,\sigma_{\chi^2_{km}}}$$

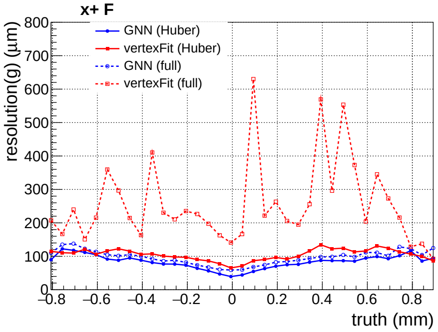


resolution(g)

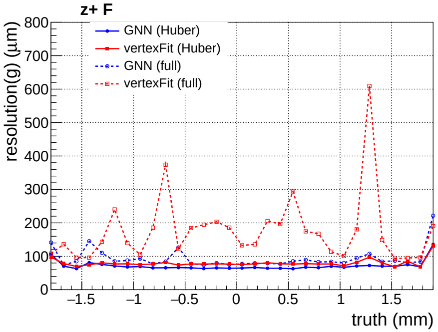
Def: resolution$(g) = \sqrt{\mathrm{Var}[\hat{r} \mid t = g]}$. Smaller **resolution**$(g) \to$ better local resolution.

σ_{Huber}: Huber IRLS robust estimation, $c = 1.345$ (**95% Gaussian efficiency**); outliers down-weighted, no trimming

$$\rho_c(r) = \begin{cases} \frac{1}{2} r^2, & |r| \le c \\ c\,|r| - \frac{1}{2} c^2, & |r| > c \end{cases}, \qquad w_i = \frac{\rho_c'(r_i)}{r_i}, \qquad \sigma^2 = \frac{\sum_i w_i r_i^2}{\sum_i w_i} \text{ (iterate)}$$



x direction



z direction

bias(g)

Chentao Bao

Physics Background

Workflow

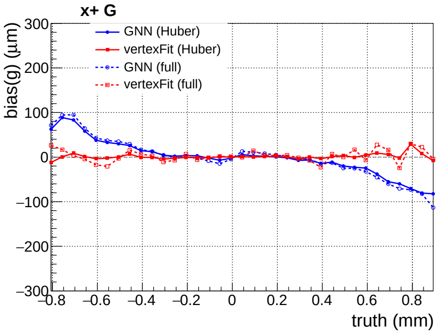
Network Architecture

Training

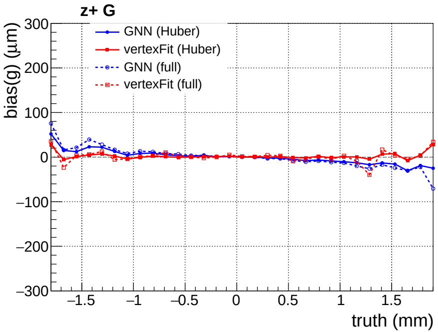
τ Double-Solution Selection

Summary

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



x direction



z direction