Uncovering Quantum Entanglement and Bell Nonlocality in $\tau^+\tau^-$ at the Large Hadron Collider through Machine Learning for Neutrino Reconstruction

Baihong Zhou (TDLI)

In collaboration with Yulei Zhang[2], Qibin Liu[1], Matthew Low[3], Tong Arthur Wu[3], Shu Li[1], Tao Han[3], Shih-Chieh Hsu[2]

1. Tsung-Dau Lee Institute, SJTU;
2. University of Washington, Seattle;
3. University of Pittsburgh, Pittsburgh;

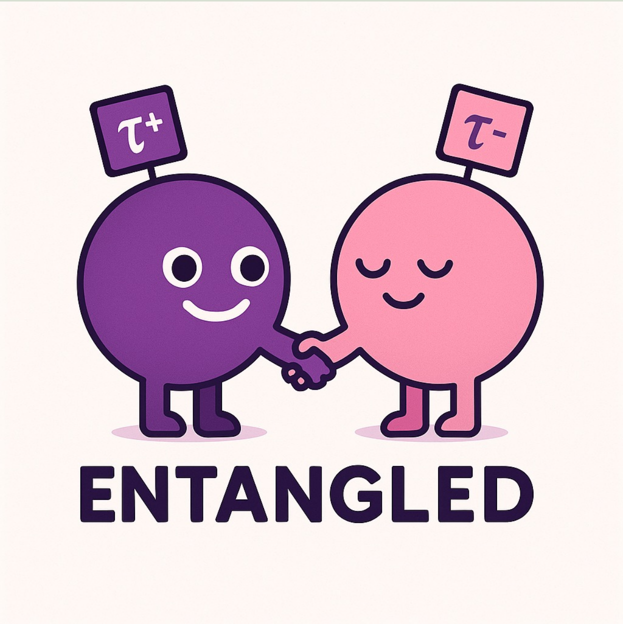
***Link*: 2504.01496**

Introduction

- Quantum entanglement is one of the **hallmarks** of **quantum mechanics**, which has been observed from the microscopic to the macroscopic, while it still remains to **be further validated** in the TeV scale;

Bosons

($H \to VV$)



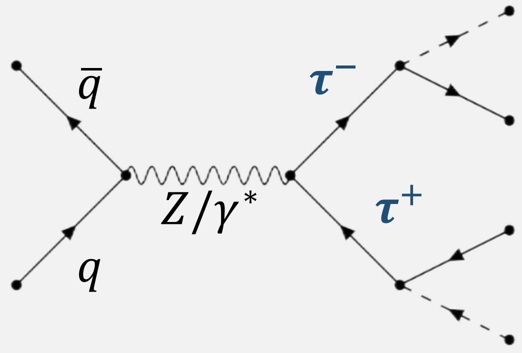
Fermions

($Z \to \tau\tau$)

Introduction

- In the ATLAS experiment, the **Quantum Entanglement** has been observed in the $pp \to t\bar{t}$ process;
- It remains comparatively **unexplored in the $pp \to \tau\tau$ system**, which

is worth taking a look using **Delphes** (with the setting of **ATLAS** detector);



$pp \to t\bar{t}$	$pp \to Z \to \tau\tau$
$\sigma_{tot} = 834\, pb$ ($\sigma_{*boost} = 20\, fb$)	$\sigma_{tot} = 1848\, pb$
$BR_{leptonic} = 5\%$	$BR_\pi = 10\%$
$\kappa_l = 1.0$	$\kappa_\pi = 1.0$
*boost: $m_{t\bar{t}} > 800\, GeV$	

Pure, entanglement

With Z dominate: [2]

$$\rho_{\tau\bar{\tau}} = \lambda\tilde{\rho}^{(+)} + (1-\lambda)\tilde{\rho}_{mix}$$

$$\lambda = \frac{(g_A^\tau)^2 - (g_V^\tau)^2}{(g_A^\tau)^2 + (g_V^\tau)^2}$$

Mix, separate

$$\rho = \frac{1}{4}\left(\mathbb{I}_2 \otimes \mathbb{I}_2 + \sum_i B_i^+ \sigma_i \otimes \mathbb{I}_2 + \sum_j B_j^- \mathbb{I}_2 \otimes \sigma_j + \sum_{ij} C_{ij} \sigma_i \otimes \sigma_j\right)$$ [1]

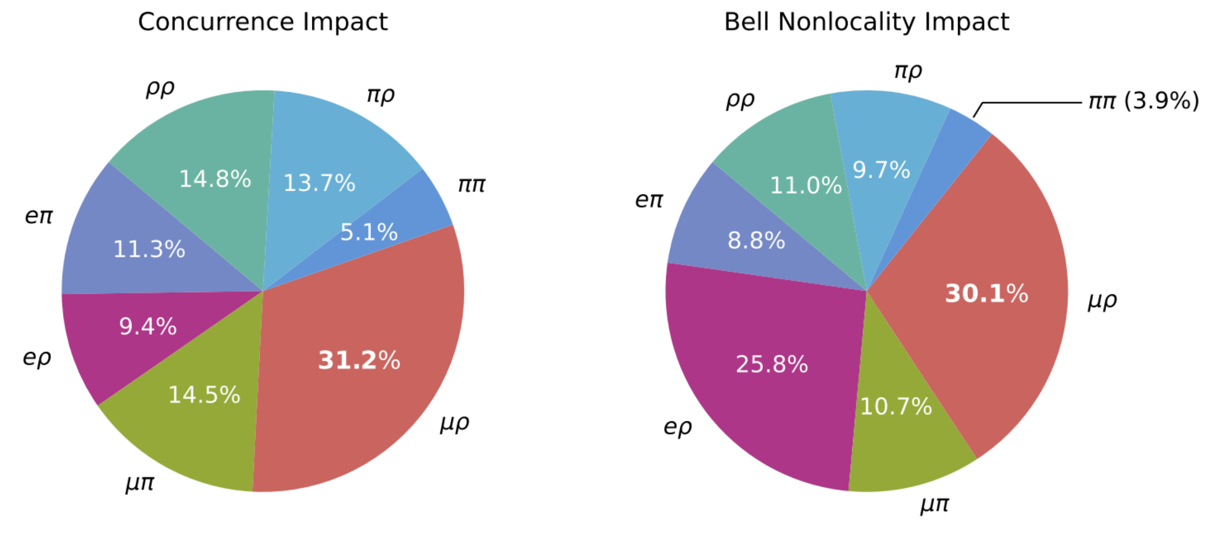
[1] Pairs of two-level systems;
[2] Constraining new physics in entangled two-qubit systems: top-quark, tau-lepton and photon pairs;

Introduction

- In the $\tau\bar{\tau}$ system, the density matrix ρ consists of a **mixture** of pure entangled and separable states, thus placing stricter constraints on the **τ reconstruction**;
- **7 distinct $\tau\tau$ decay subchannels** are taken into account for this study;

Decay	Spin Analyzing Power	Branching Ratio
$\pi\nu_\tau$	1.00	10.82%
$\rho(\pi\pi^0)\nu_\tau$	0.41	25.52%
$e\nu\nu_\tau$	-0.33	17.82%
$\mu\nu\nu_\tau$	-0.34	17.39%

Single τ decay



Overall results benefits from all channels

Introduction

- The $\tau\bar{\tau}$ system is **under constrains**, which means some **estimation techniques** are required;

$$\boldsymbol{\tau^+\tau^- \to \pi^+\pi^-}$$

$$E_{x,y}^{miss} = p_{x,y}^{\nu_1} + p_{x,y}^{\nu_2}$$

$$E_{\nu_i}^2 - \overrightarrow{p_{\nu_i}}^2 = 0$$

$$m_{\tau_i}^2 = (E_{\nu_i} + E_{\pi_i})^2 - (\overrightarrow{p_{\nu_i}} + \overrightarrow{p_{\pi_i}})^2$$

8 unknown,
6 constrains

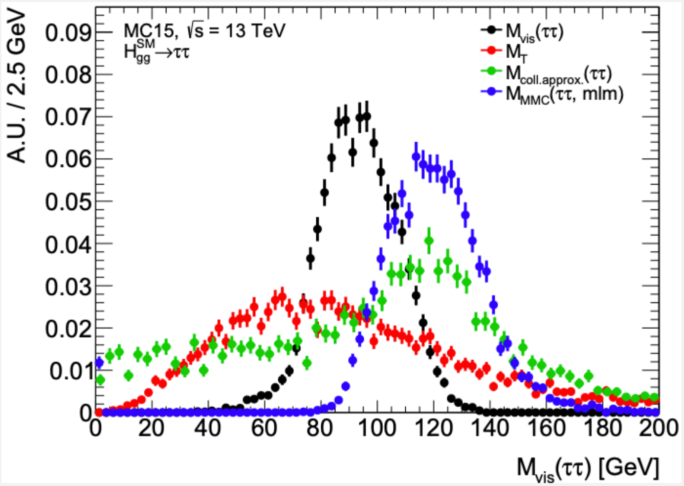
$$\boldsymbol{\tau^+\tau^- \to \ell^+\ell^-}$$

$$E_{x,y}^{miss} = p_{x,y}^{\nu_1} + p_{x,y}^{\nu_2}$$

$$m_{\tau_i}^2 = (E_{\nu_i} + E_{\pi_i})^2 - (\overrightarrow{p_{\nu_i}} + \overrightarrow{p_{\pi_i}})^2$$

8 unknown,
4 constrains

- Some techniques, like **Missing Mass Calculator** (MMC)[1] has been well designed in the ATLAS experiments;



[2]

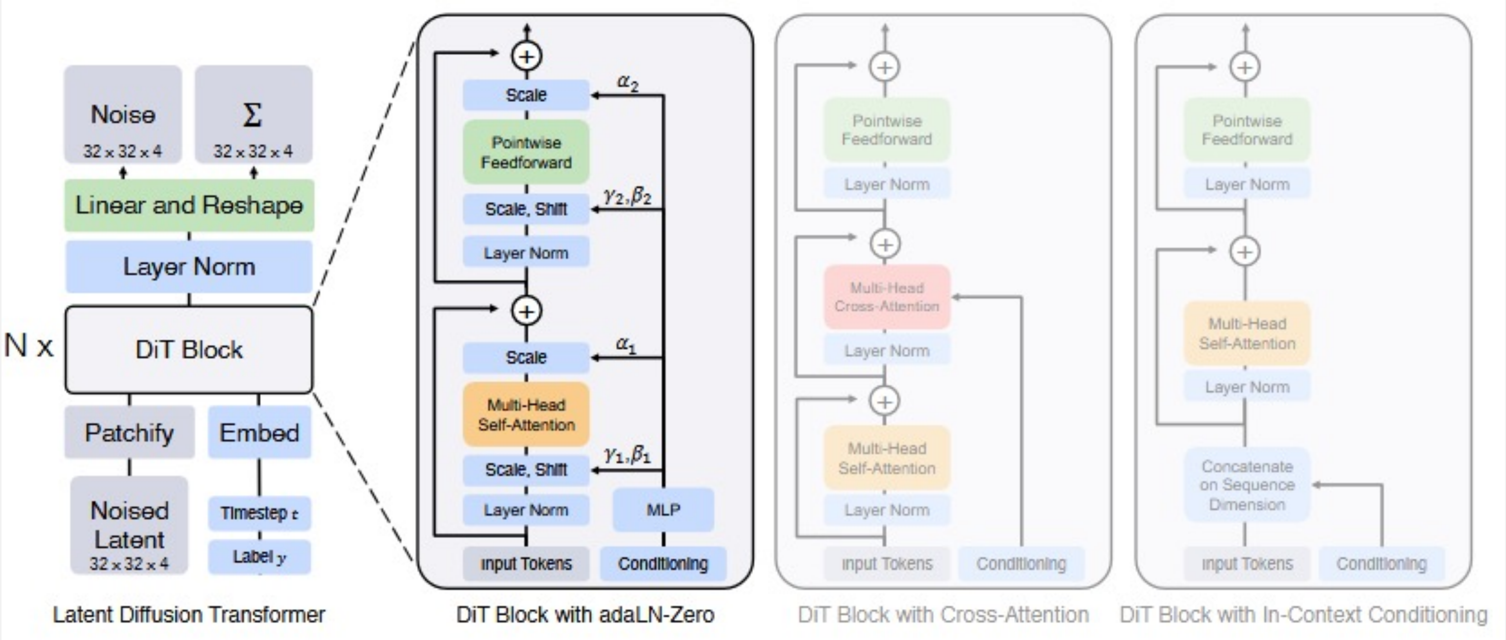
- In this study, we trained one **generative model** for ν reconstruction;

[1] A new mass reconstruction technique for resonances decaying to $\tau\tau$

[2] Effects of tau decay product reconstruction in a Higgs CP analysis with the ATLAS experiment

Introduction

- In many generation tasks, **the diffusion models** together with the **Transformer** framework have emerged at the forefront, indicating **their strong potential** in the ν reconstruction;
- The **Point-Edge Transformer (PET)** body and the **generation** head of ***OmniLearn***[1] is a good way to indicate the three momenta of ν in our $pp \rightarrow \tau\tau$ system with the **diffusion model**;



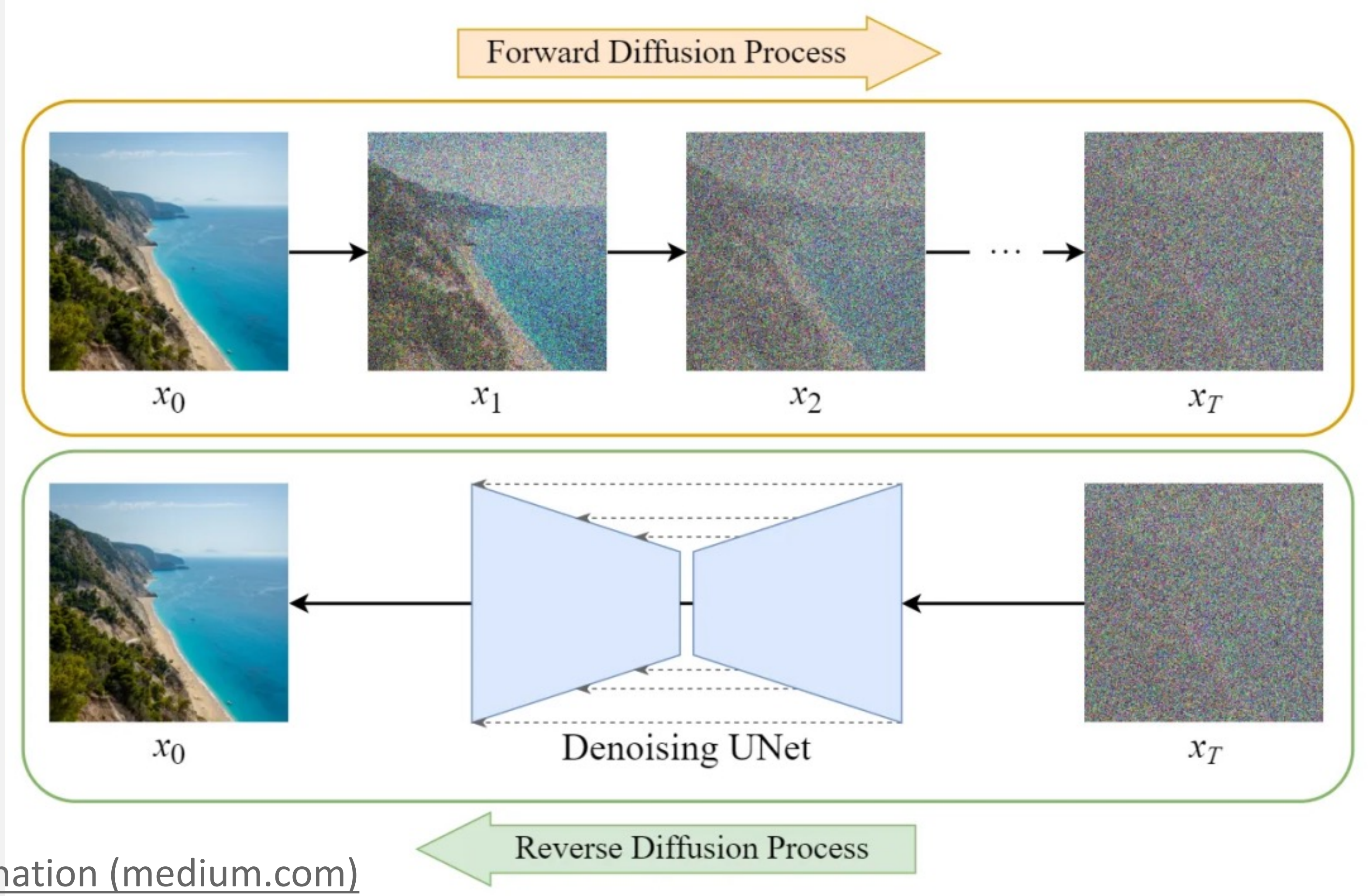
[2]

[1] Solving Key Challenges in Collider Physics with Foundation Models

[2] Scalable Diffusion Models with Transformers

Introduction

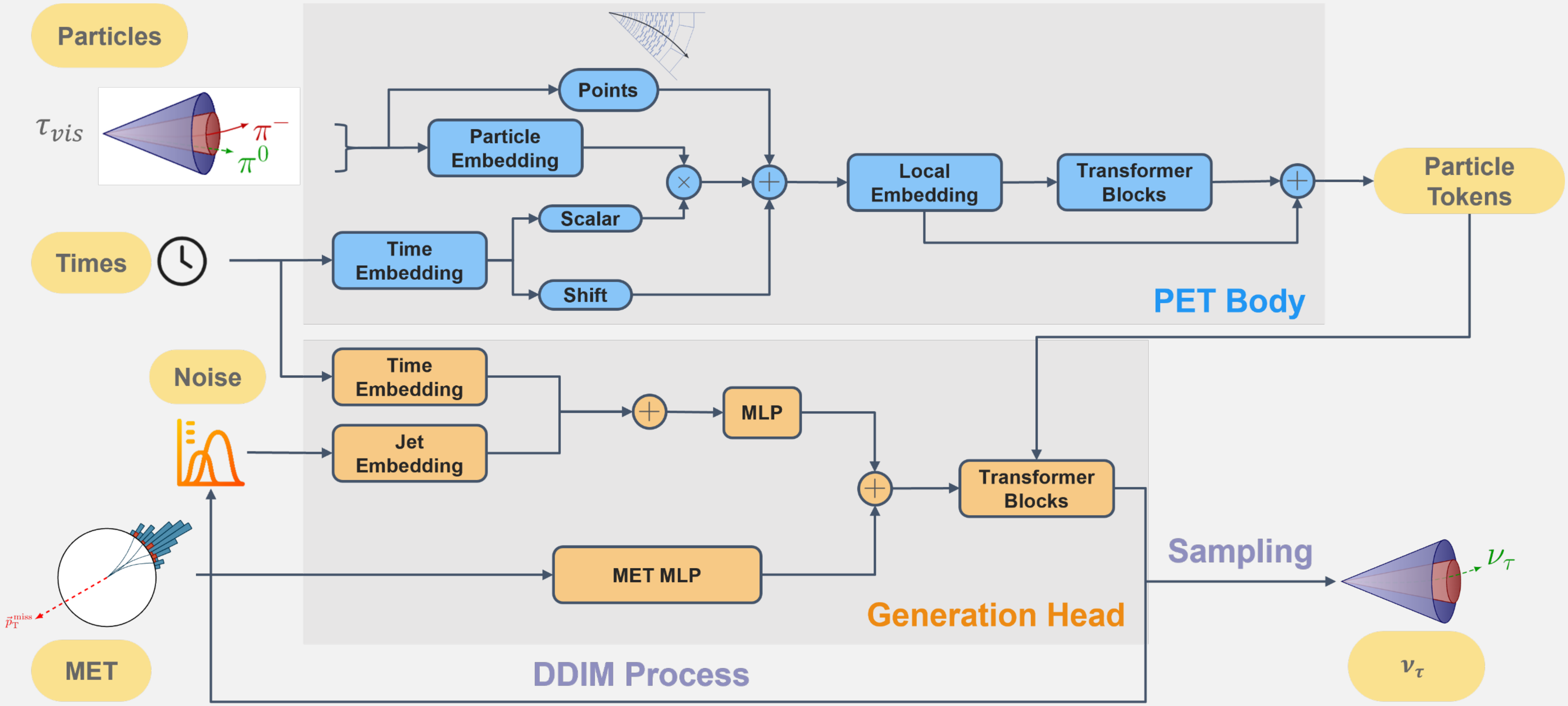
- Diffusion model is a **generative model** that adding noise in the training process and denoising in the evaluation process;



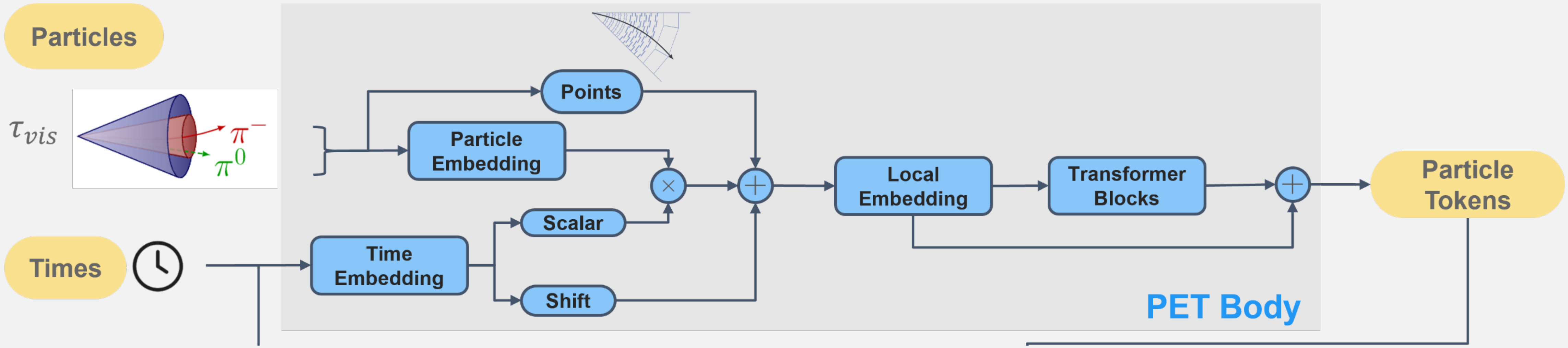
[1]

[1] Diffusion Model Explanation (medium.com)

Architectures



Architectures – PET Body



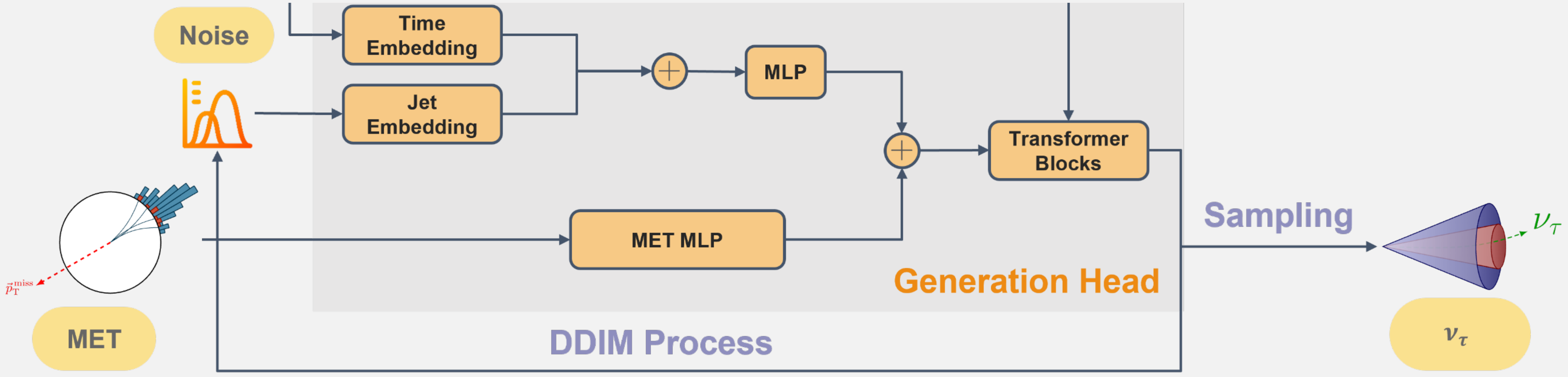
PET Body: **Tokens Generation Network**

- **Points**: Represent the (η, ϕ) coordinates of the input particles;
- **Edge features**: Combine particle features and their neighbor differences;
- **Time**: Control the process of adding noise to the original distribution;
- **Transformer** Blocks: Capture and model the **relationships** between the input features;
- **Output**: the tokens that **directs the generation head**;

Architectures – Generation Head

Generation Head: **Final Three-Momentum Generation of ν**

- **Particle tokens**: Generated from the PET body;
- **Time**: Control the process of adding noise to the original distribution;
- **MET**: Guide the generation head towards the appropriate distribution;
- **Sampling Process**: Utilize **Denoising Diffusion Implicit Models (DDIM)**[1] to produce the three-momentum of ν;



[1] Denoising Diffusion Implicit Models

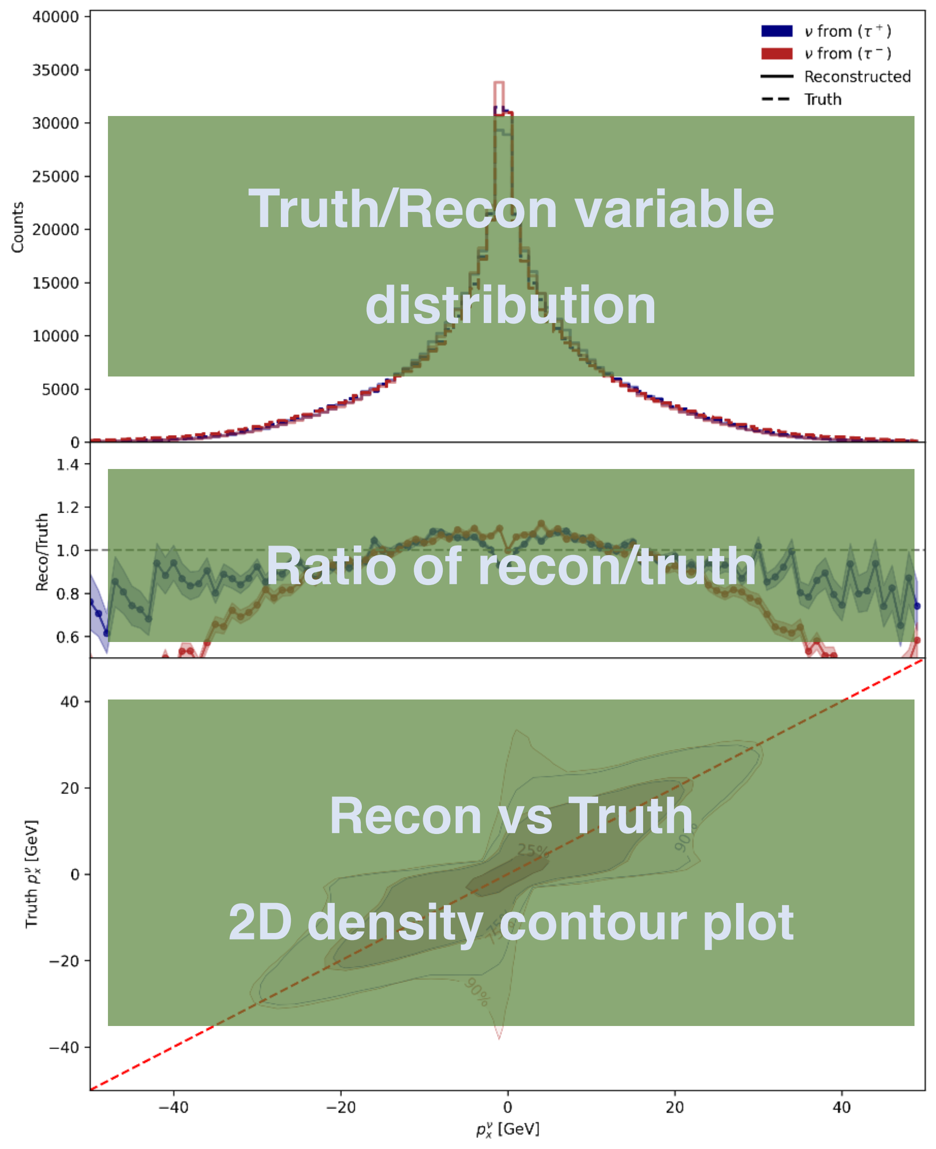
Input variables

- **7 distinct $\tau\tau$ decay subchannels** are trained separately using a consistent strategy for variable input, training and evaluation. All input variables are at the **reconstruction level**;
- To increase the number of training samples, the strategy of **randomly rotating** the system along the z-axis is used.
- The input variables are listed in the table below:

Category	Variables	Description
E_T^{miss}	$(p_T^{\text{miss}}, \phi^{\text{miss}})$	Missing transverse momentum vector
τ Visible Components	(p_T, η, ϕ, E)	Four-momentum
	Charge	Electric charge of τ-visible parts
	PID	Electron, muon, or pion identification
Small-R Jets	(p_T, η, ϕ, E)	Four-momentum
	Charge	Electric charge of the jet
	PID	Particle identification

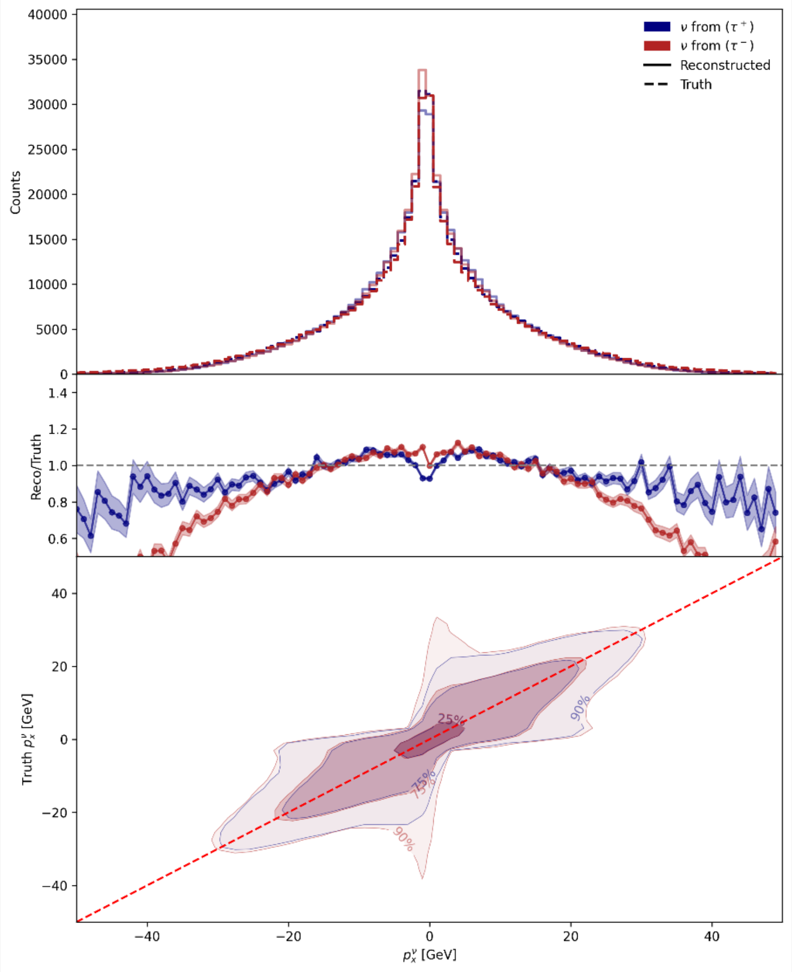
Neutrino Reconstruction

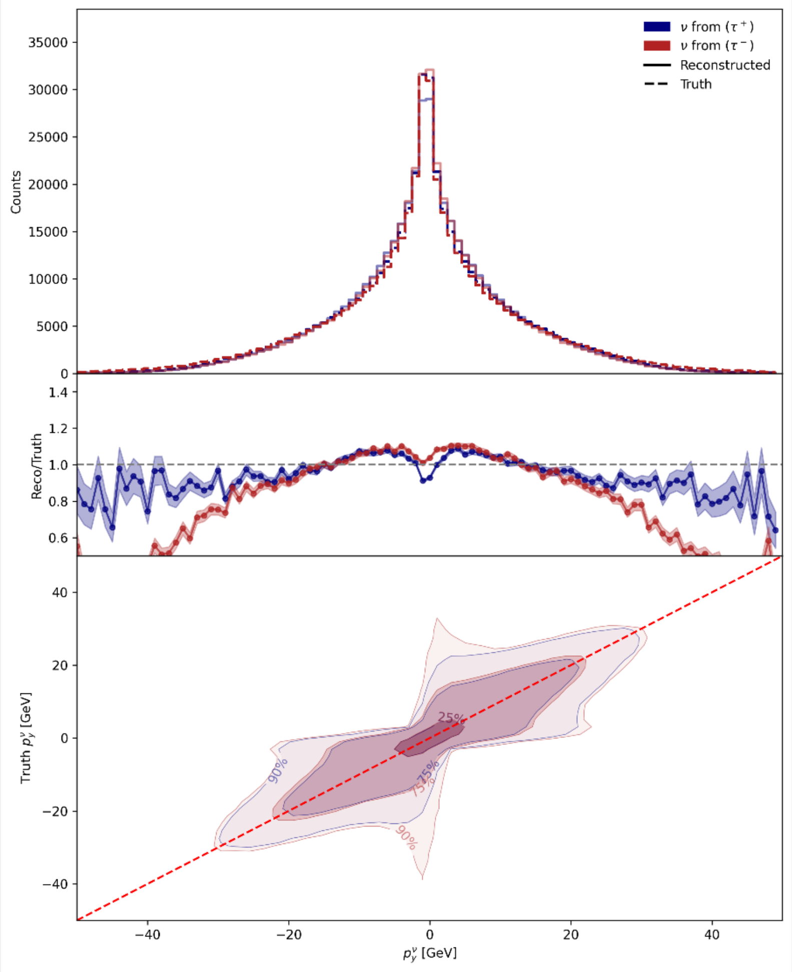
- **Purpose**: Precisely reconstruct ν to reconstruct τ **with high precision**.
- **Methods**: Use **Diffusion** model to reconstruct ν with **DDIM** sampling methods.
- **Evaluation**: In the validation dataset (**same amount** of training dataset but not used in training), each event would generate **10 candidates**.

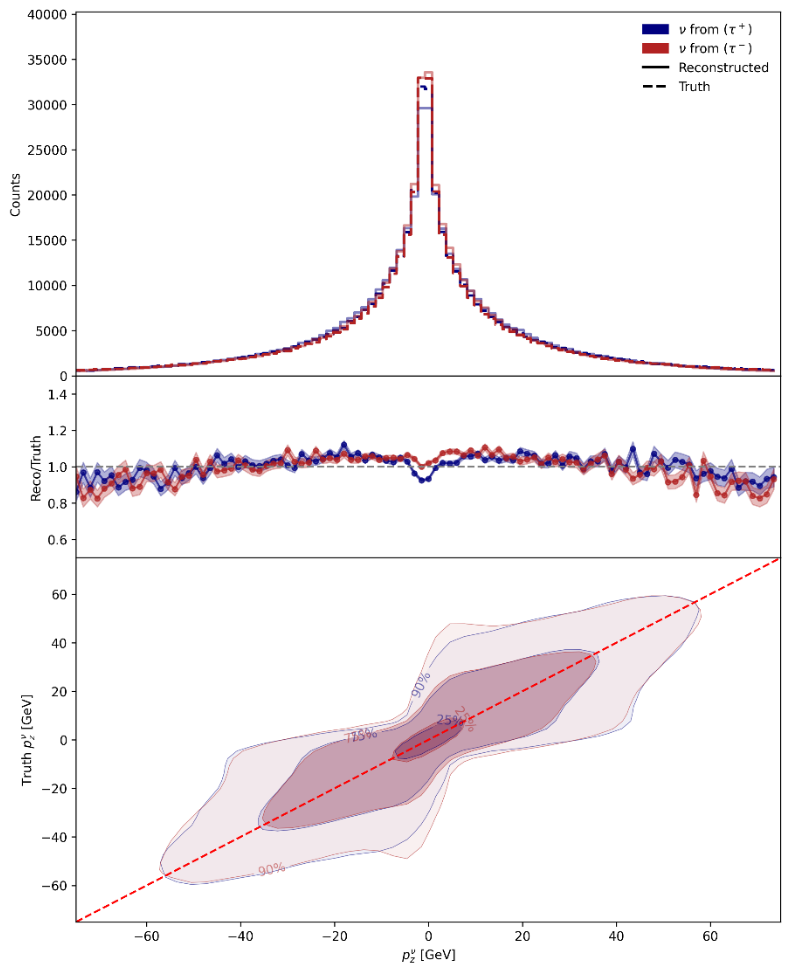


Neutrino Reconstruction - $\pi\pi$

- **Excellent reconstruction results** have been achieved for the three-momentum of ν.

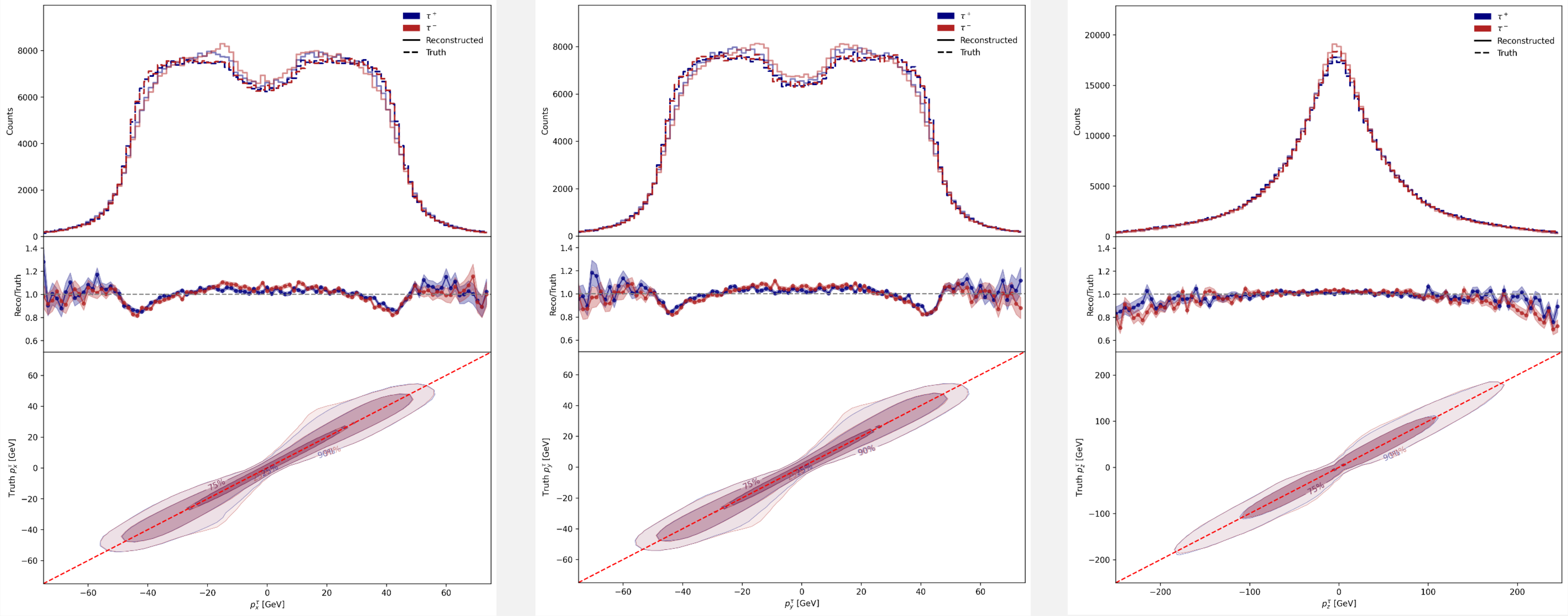






Neutrino Reconstruction - $\pi\pi$

- **Excellent reconstruction results** have been achieved for the three-momentum of τ.

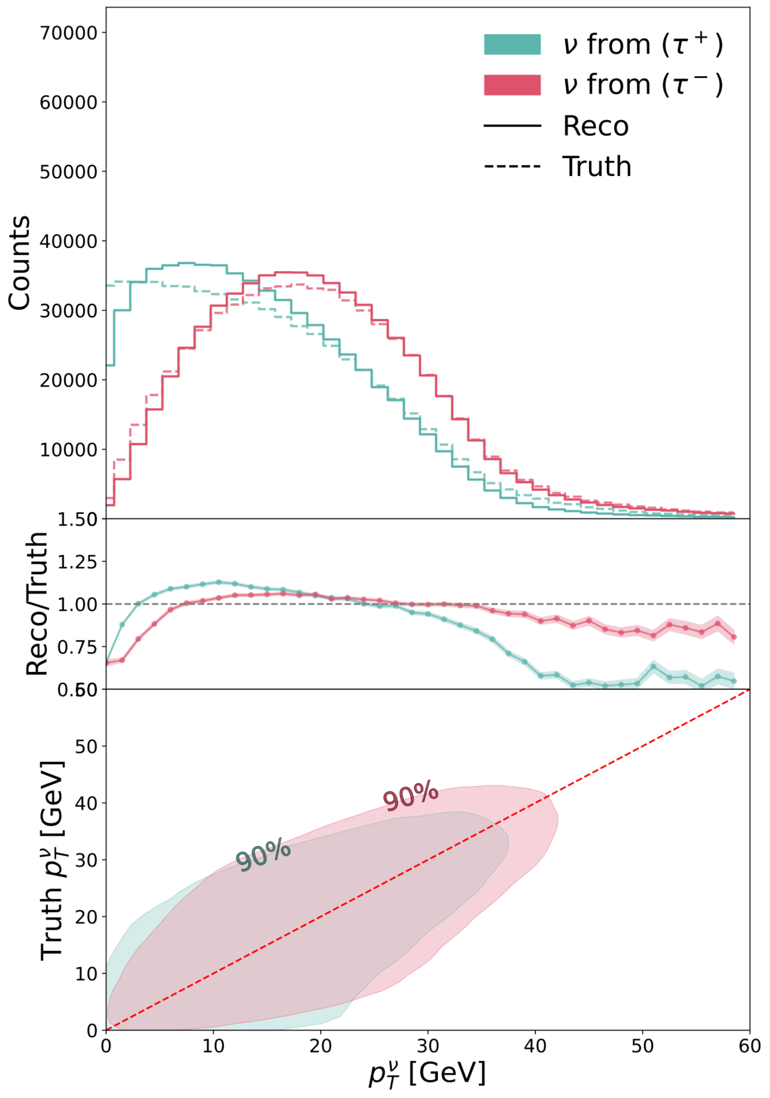


Neutrino Reconstruction - Comparison

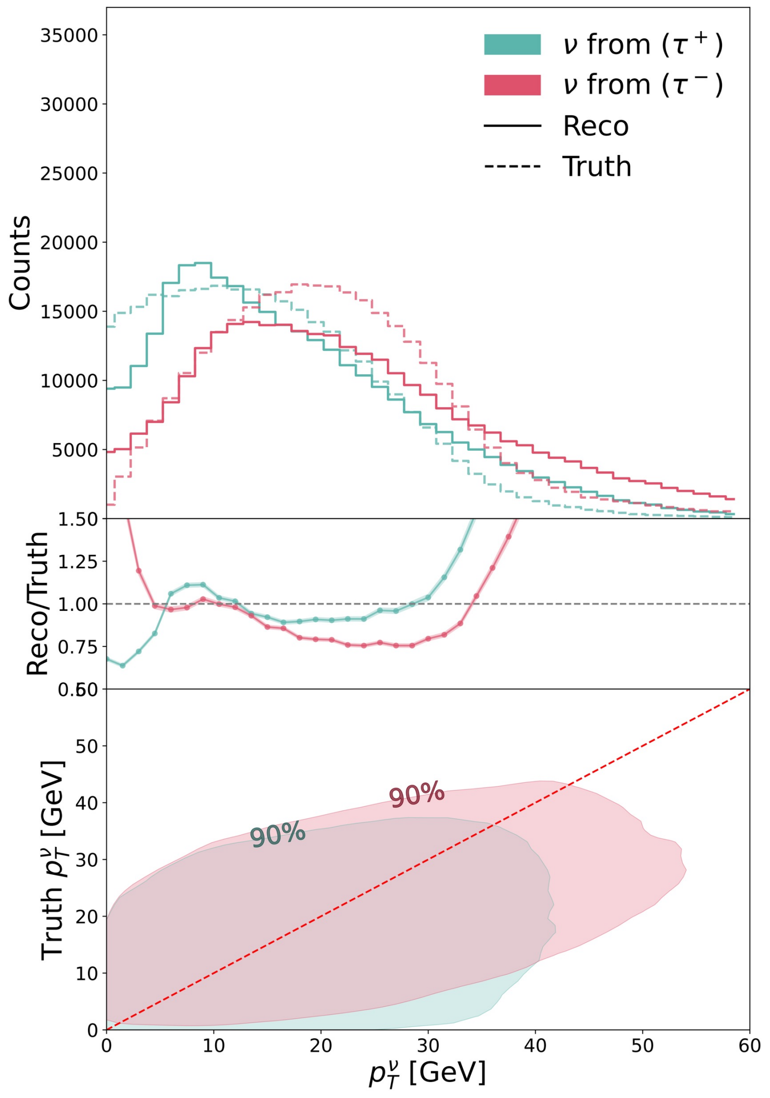
TSUNG-DAO LEE INSTITUTE

	$e\rho$ (%)		$\mu\rho$ (%)	
	ML	MMC	ML	MMC
$\Delta p^x_{\tau+}$	**15.72**	26.72	**16.03**	26.62
$\Delta p^y_{\tau+}$	**15.50**	26.00	**16.04**	27.83
$\Delta p^z_{\tau+}$	**15.70**	25.65	**16.12**	26.69
$\Delta p^x_{\tau-}$	**15.69**	26.47	**16.02**	26.85
$\Delta p^y_{\tau-}$	**15.34**	26.87	**16.15**	27.39
$\Delta p^z_{\tau-}$	**15.81**	26.26	**16.26**	25.67
$\Delta m_{\tau\tau}$	**5.81**	22.70	**5.76**	22.81

Half-Width at Half-Maximum (HWHM)



Diffusion ($l\rho - channel$)



MMC ($l\rho - channel$)

Neutrino Reconstruction - 7 subchannels

- To quantify the resolution of the reconstructed distributions, the **HWHM** is used as a metric.

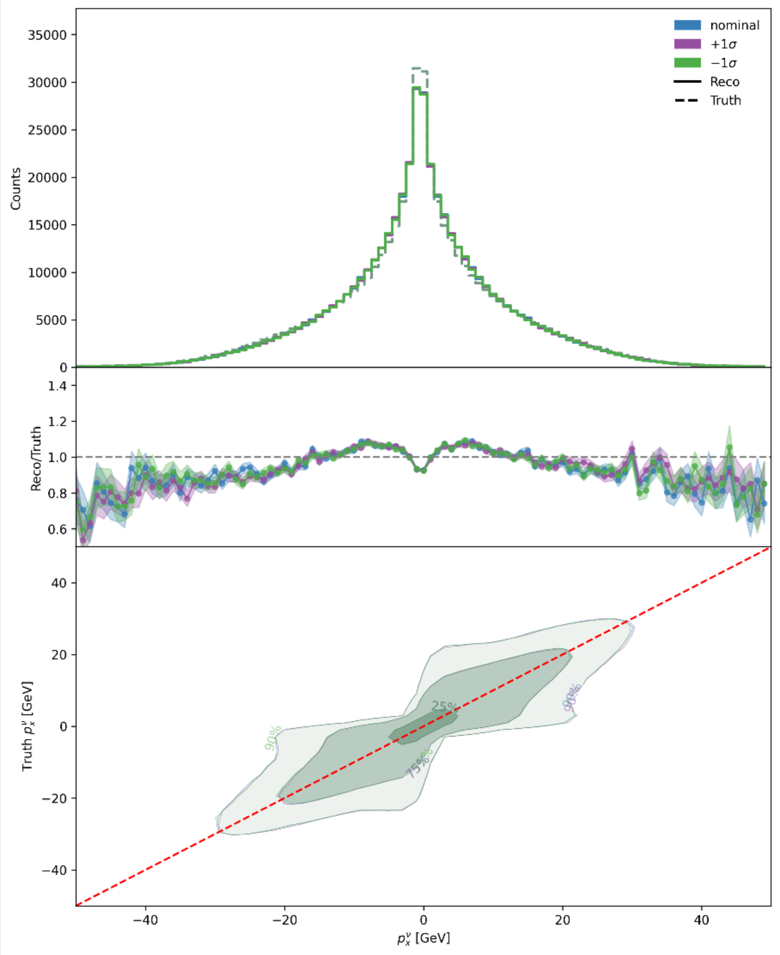
	$\pi\pi$ (%)		$e\pi$ (%)		$\mu\pi$ (%)		$\pi\rho$ (%)		$e\rho$ (%)		$\mu\rho$ (%)		$\rho\rho$ (%)	
	ML	MMC	ML	MMC	ML	MMC	ML	MMC	ML	MMC	ML	MMC	ML	MMC
$\Delta p^x_{\tau^+}$	**18.97**	25.99	**18.34**	27.91	**19.30**	28.56	**16.19**	25.45	**15.72**	26.72	**16.03**	26.62	**16.38**	25.65
$\Delta p^y_{\tau^+}$	**19.01**	26.02	**18.54**	27.26	**19.33**	28.15	**15.96**	25.26	**15.50**	26.00	**16.04**	27.83	**16.38**	25.28
$\Delta p^z_{\tau^+}$	**19.47**	25.48	**19.52**	27.46	**20.00**	27.19	**16.31**	24.85	**15.70**	25.65	**16.12**	26.69	**16.49**	25.02
$\Delta p^x_{\tau^-}$	**18.77**	25.78	**17.06**	28.38	**17.69**	27.43	**18.11**	26.13	**15.69**	26.47	**16.02**	26.85	**16.34**	25.17
$\Delta p^y_{\tau^-}$	**18.71**	25.96	**16.62**	26.22	**17.33**	28.13	**17.89**	25.79	**15.34**	26.87	**16.15**	27.39	**16.36**	25.40
$\Delta p^z_{\tau^-}$	**19.69**	25.43	**17.06**	26.27	**17.75**	27.17	**18.44**	25.49	**15.81**	26.26	**16.26**	25.67	**16.72**	24.23
$\Delta m_{\tau\tau}$	**7.94**	23.27	**6.18**	24.91	**6.41**	24.31	**7.24**	22.72	**5.81**	22.70	**5.76**	22.81	**6.27**	21.89

71 % improvements

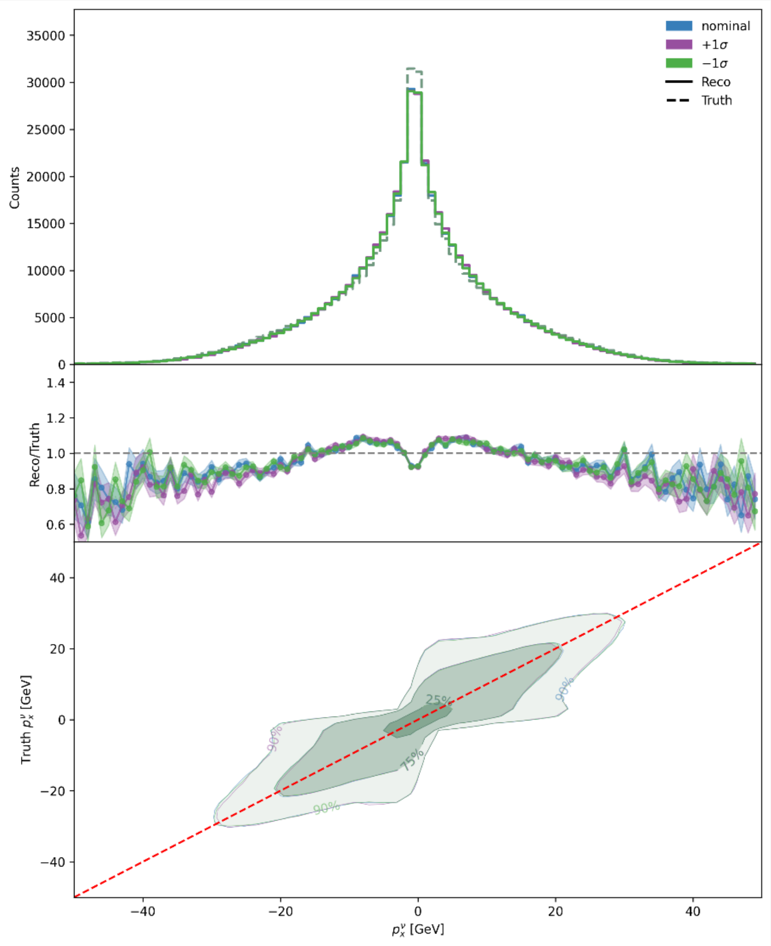
257 % improvements

Neutrino Reconstruction - Systematics

- The **network's robustness** is evaluated by: [1]



Perturbing the **MET** information by $\pm 1 GeV$



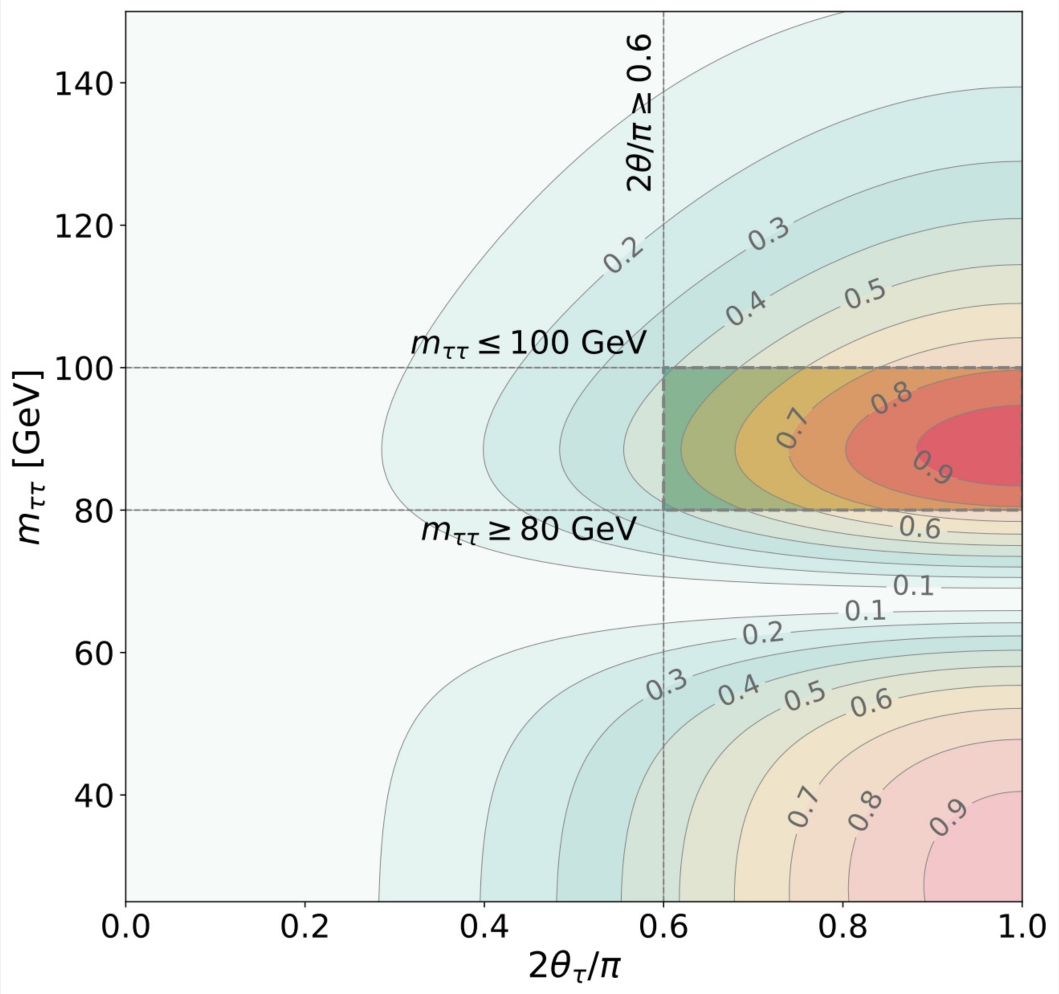
Applying a 3% shift to **all input variables**

[1] FAIR Universe HiggsML Uncertainty Challenge Competition

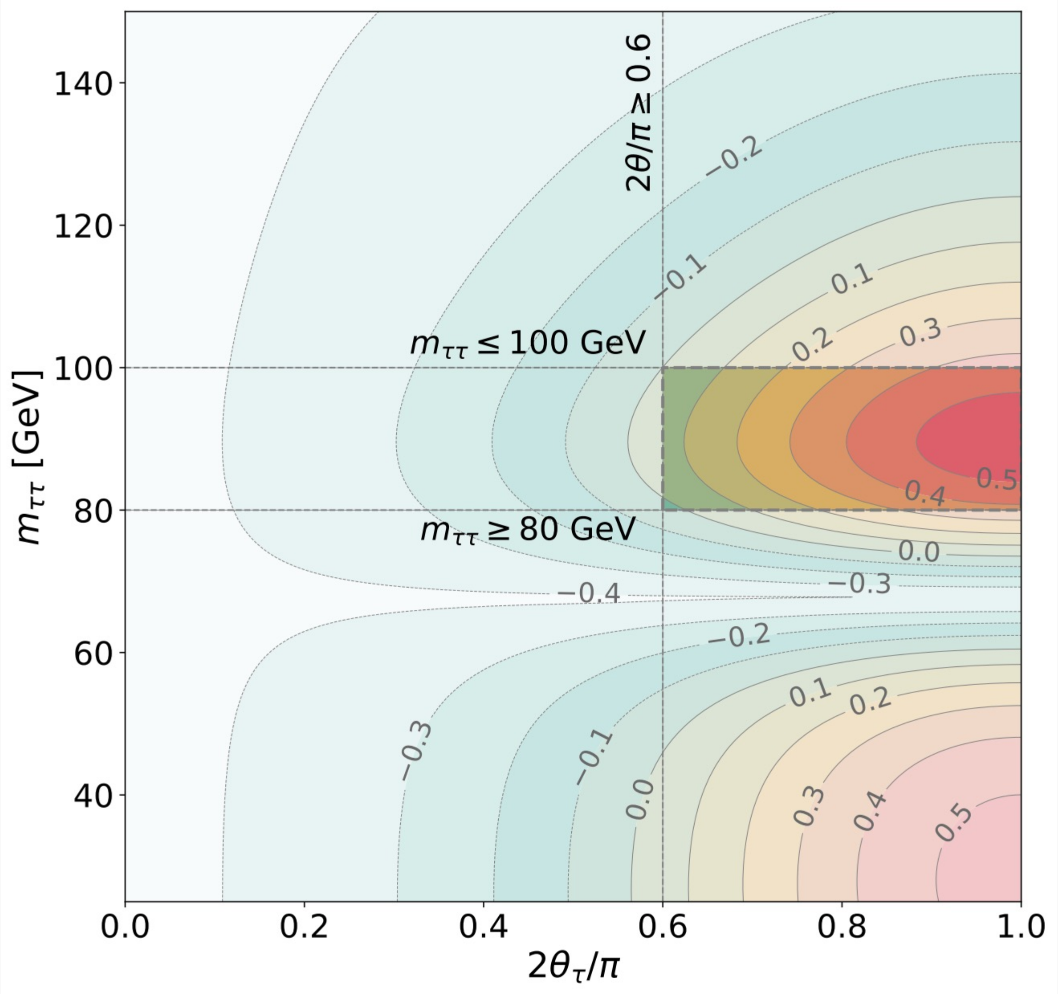
Signal Region Definition

Tsung-Dao Lee Institute

Signal Region (with triggers)



Cocurrence $(C > 0)$



Bell Nonlocality$(B > 0)$

Signal Region Yields

Subchannel	Prongness	$\lvert\kappa_A \times \kappa_B\rvert$	$\tau\tau \to \ell\pi$	$\tau\tau \to \ell\rho$	$\tau\tau \to \pi\pi$	$\tau\tau \to \pi\rho$	$\tau\tau \to \rho\rho$
SR & di-τ Trigger ($p_T^{\tau_1} > 35$ GeV & $p_T^{\tau_2} > 25$ GeV)							
$\pi\pi$			3.92 ± 0.18	0.09 ± 0.04	89.43 ± 0.48	2.90 ± 0.13	0.14 ± 0.04
$\pi\rho$			0.12 ± 0.03	22.71 ± 0.65	1.61 ± 0.06	206.39 ± 1.09	6.29 ± 0.28
$\rho\rho$			< 0.01	0.56 ± 0.10	0.06 ± 0.01	4.51 ± 0.16	629.99 ± 2.83
SR & $e+\tau$ Trigger ($p_T^e > 14$ GeV & $p_T^\tau > 25$ GeV) or single-e Trigger ($p_T^e > 26$ GeV)							
$e\pi$			378.90 ± 1.79	17.52 ± 0.57	< 0.01	0.01 ± 0.01	< 0.01
$e\rho$			8.33 ± 0.27	1233.90 ± 4.82	< 0.01	0.03 ± 0.01	0.15 ± 0.04
SR & $\mu+\tau$ Trigger ($p_T^\mu > 17$ GeV & $p_T^\tau > 25$ GeV) or single-μ Trigger ($p_T^\mu > 26$ GeV)							
$\mu\pi$			565.94 ± 2.19	25.21 ± 0.69	< 0.01	< 0.01	< 0.01
$\mu\rho$			12.63 ± 0.33	1862.06 ± 5.92	< 0.01	< 0.01	0.04 ± 0.02

Sig yields

(Total: 4996.61±8.70)

(per 1 fb^{-1})

Subchannel	$W \to \ell\nu$	$W \to \tau\nu$	$Z \to \ell\ell$	$t\bar{t}$	QCD	Total
SR & di-τ Trigger ($p_T^{\tau_1} > 35$ GeV & $p_T^{\tau_2} > 25$ GeV)						
$\pi\pi$	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
$\pi\rho$	< 0.01	< 0.01	< 0.01	0.05 ± 0.05	< 0.01	0.05 ± 0.05
$\rho\rho$	< 0.01	< 0.01	< 0.01	0.10 ± 0.07	< 0.01	0.10 ± 0.07
SR & $e+\tau$ Trigger ($p_T^e > 14$ GeV & $p_T^\tau > 25$ GeV) or single-e Trigger ($p_T^e > 26$ GeV)						
$e\pi$	< 0.01	< 0.01	< 0.01	2.87 ± 0.38	< 0.01	2.87 ± 0.38
$e\rho$	1.93 ± 0.86	0.62 ± 0.36	< 0.01	11.19 ± 0.75	< 0.01	13.74 ± 1.20
SR & $\mu+\tau$ Trigger ($p_T^\mu > 17$ GeV & $p_T^\tau > 25$ GeV) or single-μ Trigger ($p_T^\mu > 26$ GeV)						
$\mu\pi$	< 0.01	0.41 ± 0.29	< 0.01	4.13 ± 0.46	< 0.01	4.55 ± 0.54
$\mu\rho$	1.16 ± 0.67	0.62 ± 0.36	< 0.01	< 0.01	< 0.01	1.78 ± 0.76

Bkg yields

(Total: 23.10±1.57)

(per 1 fb^{-1})

Density Matrix Reconstruction

李政道研究所 TSUNG-DAO LEE INSTITUTE

Decay Approach

- Measure the **angle between τ and τ visible component** in the τ rest frame;
- Use **template fit** to extract C_{ij} components;

- $C_{ij} = \frac{9}{\kappa_a \kappa_b} < \cos\theta_i^a \cos\theta_j^b >$

[1]

Kinematic Approach

- **Parameterize** spin correlation matrix by θ (scattering angular) and β (speed of τ);
- Measuer the θ to get the C_{ij};

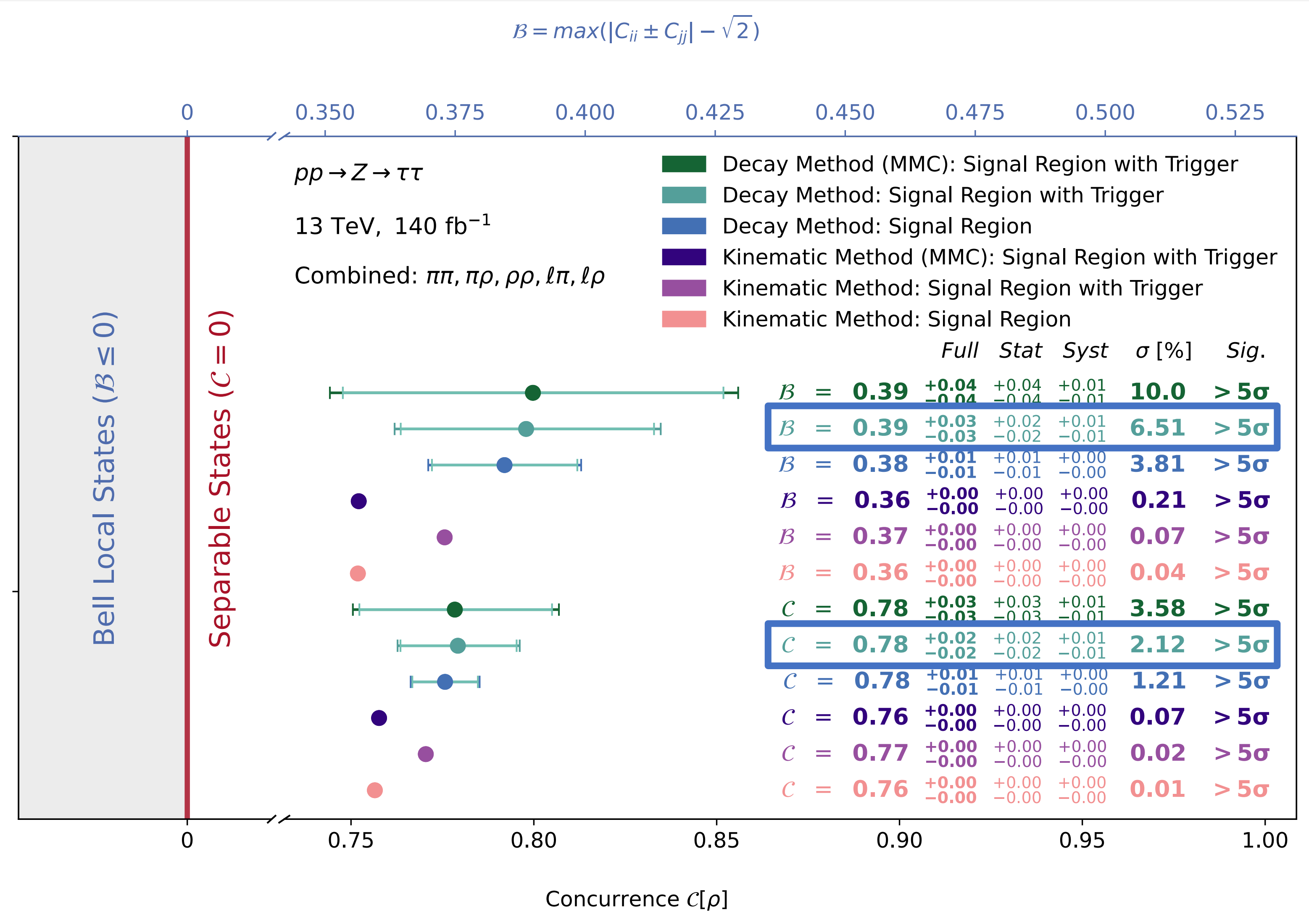
- $C_{ij} \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sin^2\theta}{1+\cos^2\theta} & 0 \\ 0 & 0 & -\frac{\sin^2\theta}{1+\cos^2\theta} \end{pmatrix}$

[1] Quantum Tomography at Colliders: With or Without Decays;

Systematics

Results

- By using ML methods to reconstruct ν, good results have been obtained;



Compare with MMC

54 %

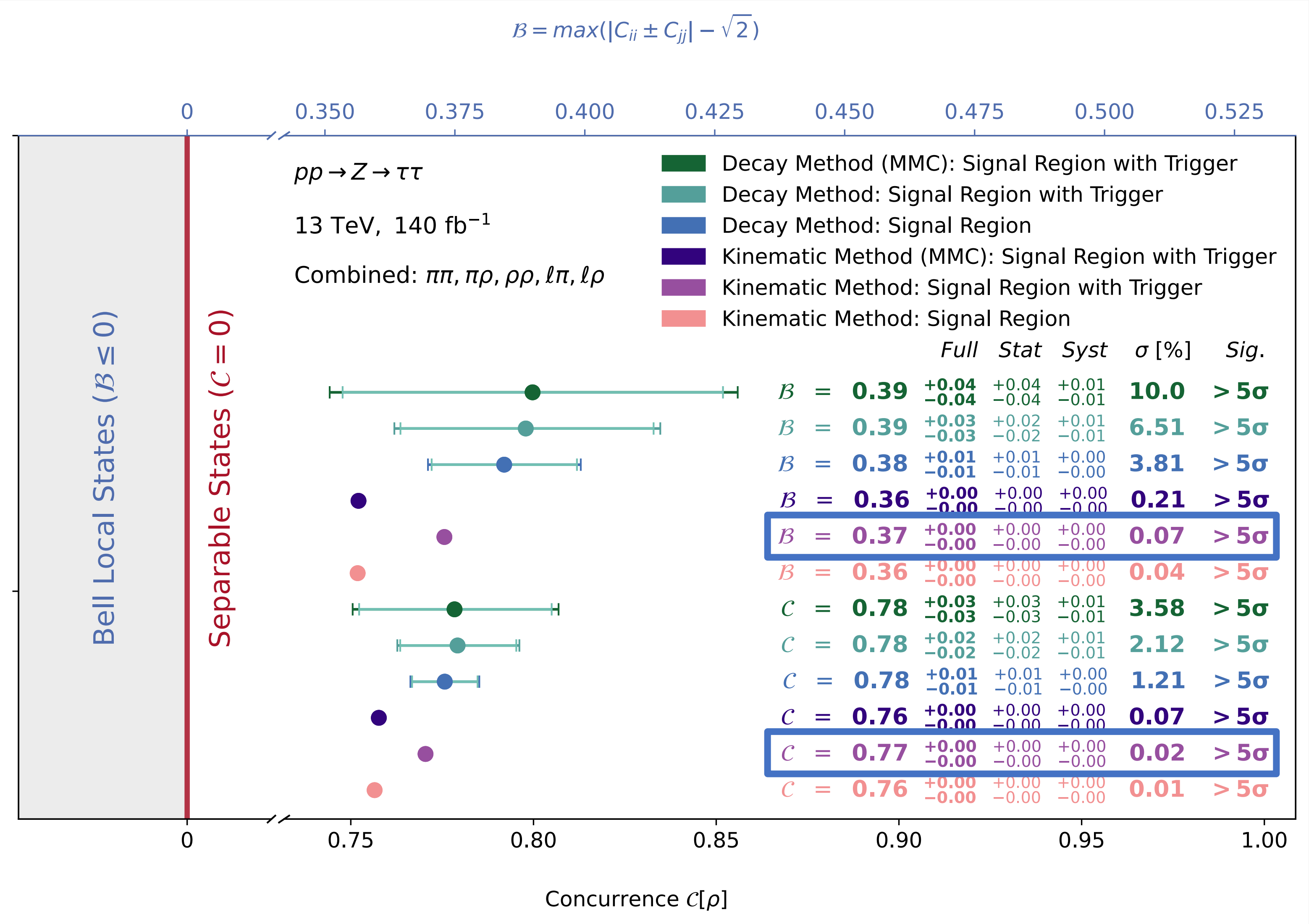
improvements

69 %

improvements

Results

- By using ML methods to reconstruct ν, good results have been obtained;



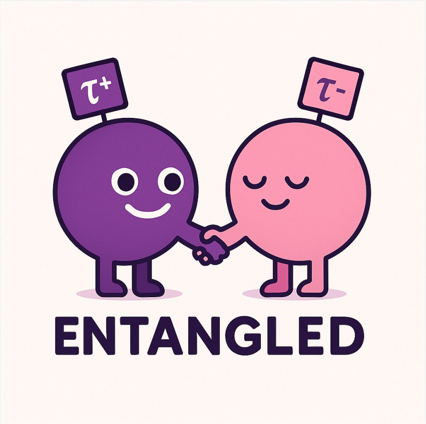
Compare with MMC

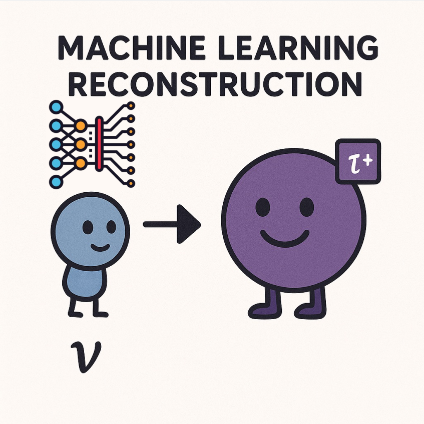
200 % improvements

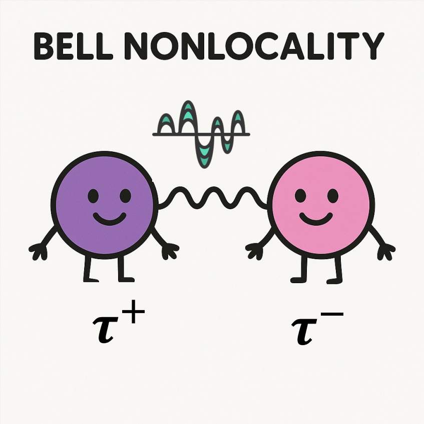
250 % improvements

Summary

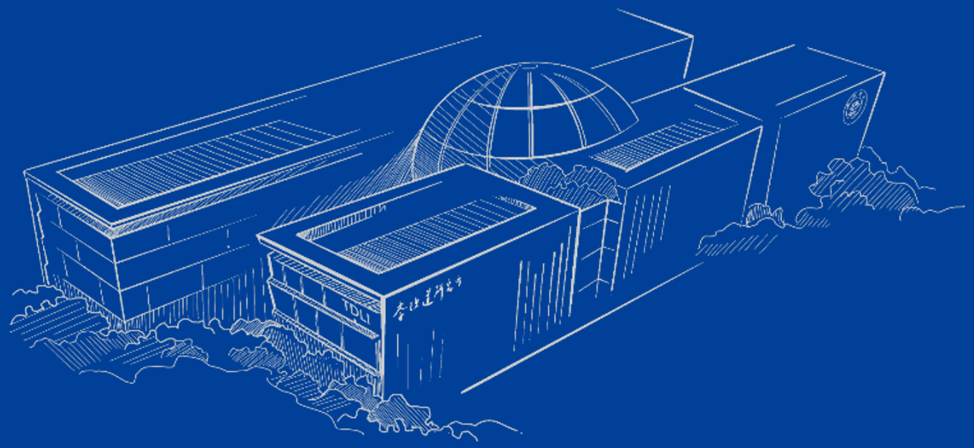
- $pp \to Z \to \tau\bar{\tau}$ **is an excellent channel for Quantum Entanglement** and **Bell Nonlocality** study:
  - **More statistics** than $t\bar{t}$ channels;
  - **Good ν reconstruction** can be achieved by using **Diffusion + PET**;
  - $> 5\sigma$ has been obtained for entanglement and Bell nonlocality only using $\mathbf{140\ fb^{-1}}$;
  - **Dr. Yulei Zhang** proposed this analysis within the **ATLAS Standard Model group** and has already initiated a **feasibility study** using official ATLAS datasets.
- Thank to **Vinicius Mikuni** and **Benjamin Nachman** for their help in using ***OmniLearn***.
- Special thanks to **Tong Arthur Wu, Matthew Low** and **Tao Han** for their detailed and expert guidance on the theory of this study.







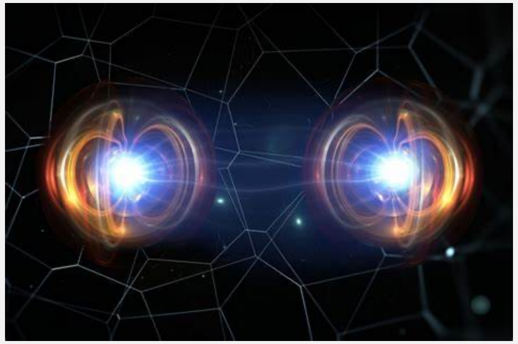
Thanks!



C and B calculation

- Consider a state ρ for a bipartite system with sub-systems i and j, it can be written as follows:

$$\rho = \frac{1}{4}\left(\mathbb{I}_2 \otimes \mathbb{I}_2 + \sum_i B_i^+ \sigma_i \otimes \mathbb{I}_2 + \sum_j B_j^- \mathbb{I}_2 \otimes \sigma_j + \sum_{ij} C_{ij} \sigma_i \otimes \sigma_j\right)$$ [1]



Here, $B_{i,j}^{\pm}$ characterizes the net polarization of the $\tau^{\pm}$, C_{ij} describers the spin correlations, and $i, j = 1,2,3$. Once ρ is reconstructed, we can calculate the concurrence C and Bell variable B.

- The definition of the C and B are as follow:
  - $C = \max(0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4), \lambda_i = \sqrt{r_i}, r_i$ is the eigenvalue, in descending magnitude of the matrix $\rho(\sigma_2 \otimes \sigma_2)\, \rho^*(\sigma_2 \otimes \sigma_2)$; **$C > 0$ for entanglement state;** [2]
  - $B = \max(\sqrt{2}\left|C_{ii} \pm C_{jj}\right| - 2)$; **$B > 0$ for Bell nonlocality;** [3]

[1] Pairs of two-level systems;
[2] Entanglement of Formation of an Arbitrary State of Two Qubits;
[3] Quantum tops at the LHC: from entanglement to Bell inequalities;

Input variables

- **7 distinct $\tau\tau$ decay subchannels** are trained separately using a consistent strategy for variable input, training and evaluation. All input variables are at the **reconstruction level**;
- To increase the number of training samples, the strategy of **randomly rotating** the system along the z-axis is used. Additionally, the τ constitutes are required the kinematics **remain consistent with** the original data.

Subchannel	Original number of samples	Number of samples after random boost
$\pi\pi$	304k	15.5M
$e\pi$	147k	14.8M
$\mu\pi$	193k	19.5M
$\pi\rho$	111k	10.9M
$e\rho$	54k	0.600M
$\mu\rho$	73k	0.817M
$\rho\rho$	38k	3.72M

Trigger

- **Di-tau trigger ($\pi\pi$, $\pi\rho$, $\rho\rho$):**
  - **Leading τ: $p_T > 35GeV$;**
  - **Sub-leading τ: $p_T > 25GeV$;**

  (HLT tau35 medium1 tracktwo tau25 medium1 tracktwo L1TAU20IM 2TAU12IM)

- **Moun+Tau trigger ($\pi\mu$, $\rho\mu$):**
  - **τ: $p_T > 25GeV$;**
  - **μ: $p_T > 14GeV$;**
  - (Or μ: $p_T > 25GeV$)

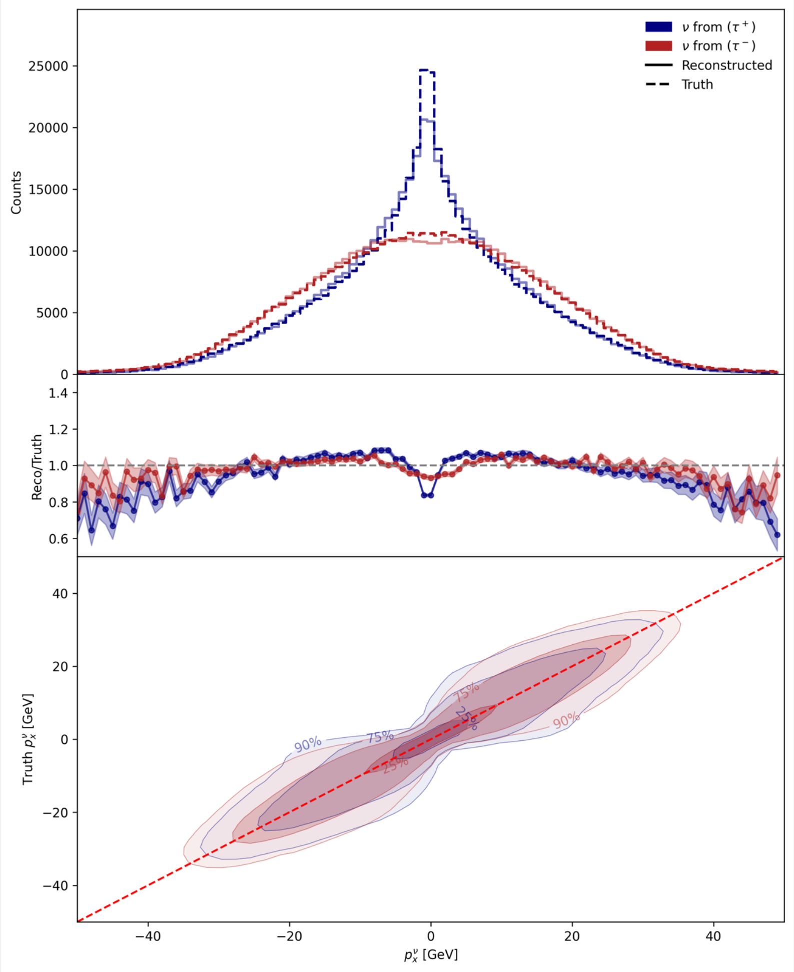
  (HLT mu14 tau25 medium1 tracktwo ||HLT mu26 ivarmedium)

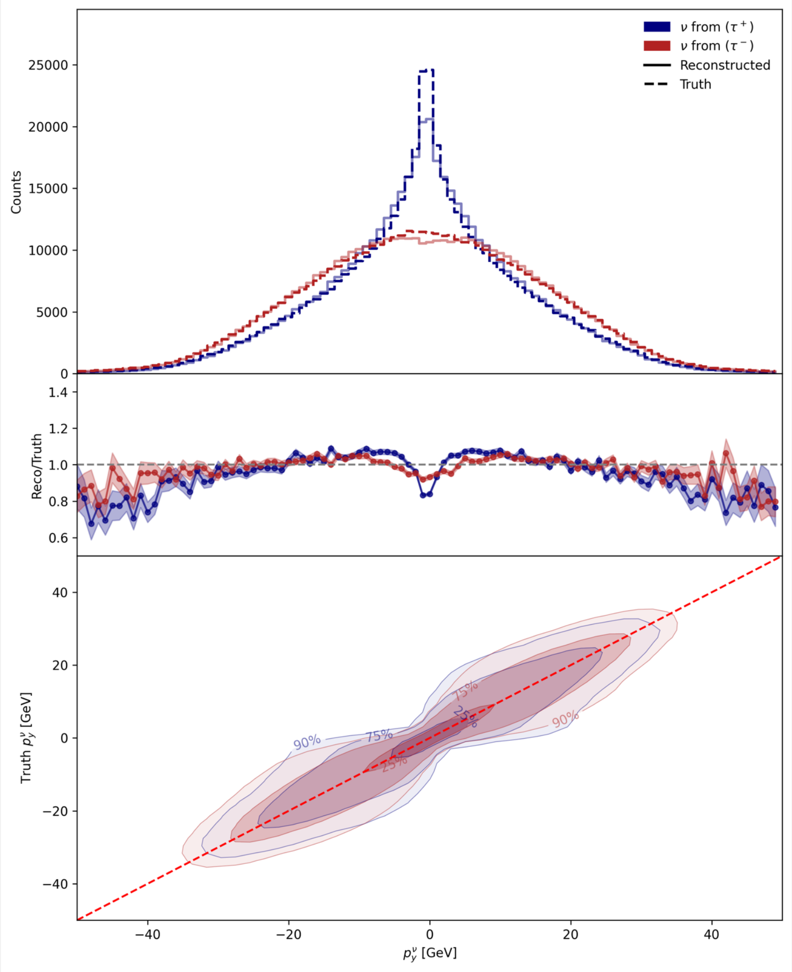
- **Electron+Tau trigger ($\pi\mu$, $\rho\mu$):**
  - **τ: $p_T > 25GeV$;**
  - **e: $p_T > 17GeV$;**
  - (Or e: $p_T > 26GeV$)

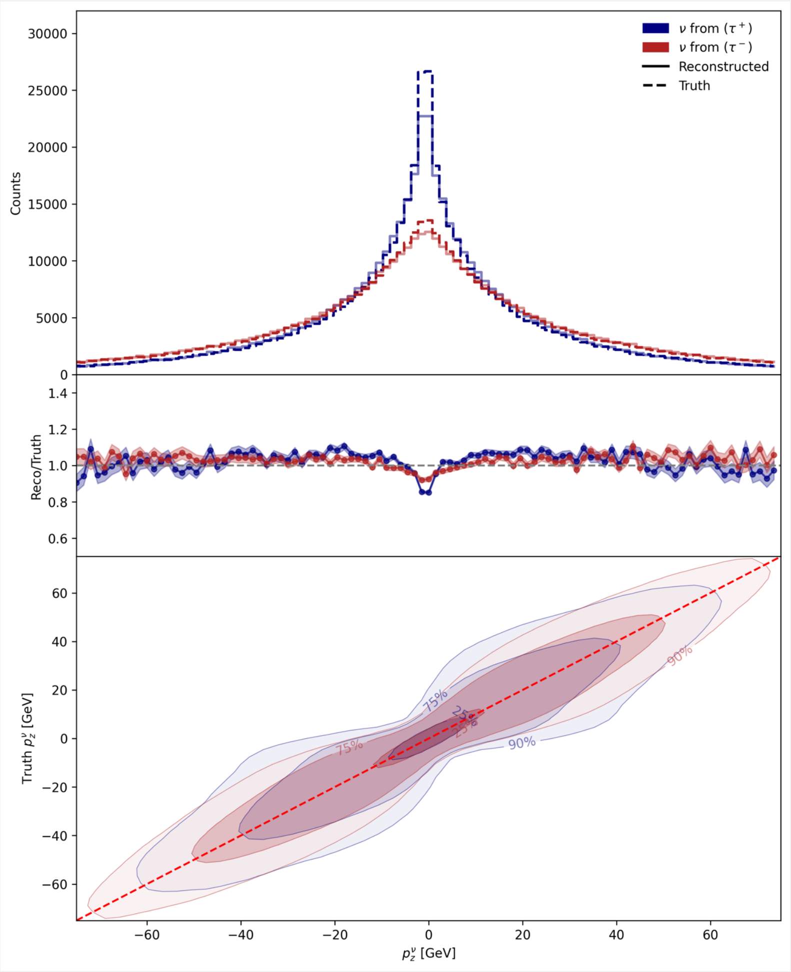
  (HLT e17 lhmedium nod0 ivarloose tau25 medium1 tracktwo||HLT e26 lhtight nod0 ivarloose)

Neutrino Reconstruction - $l\pi$

- We have achieved **excellent reconstruction results** for the three-momentum of ν.

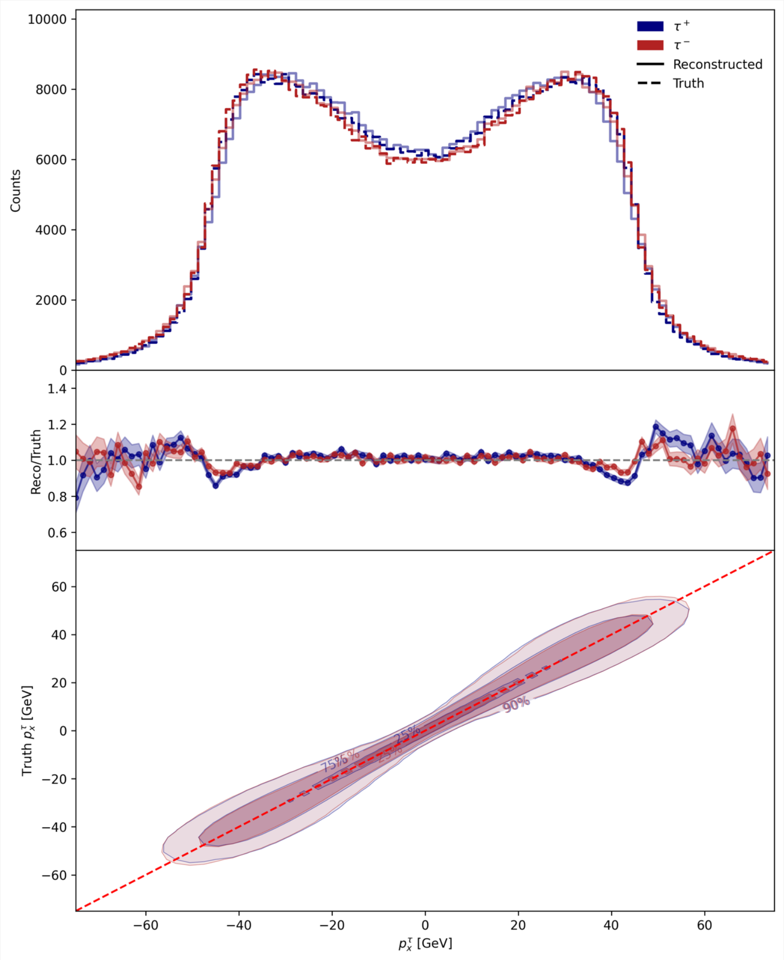


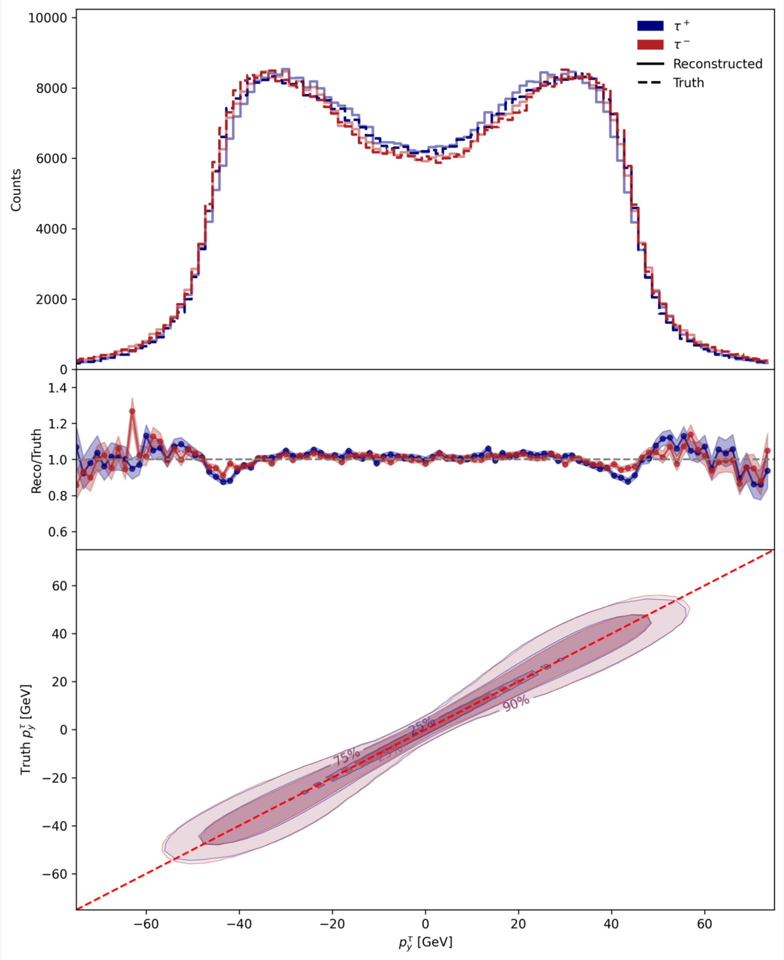


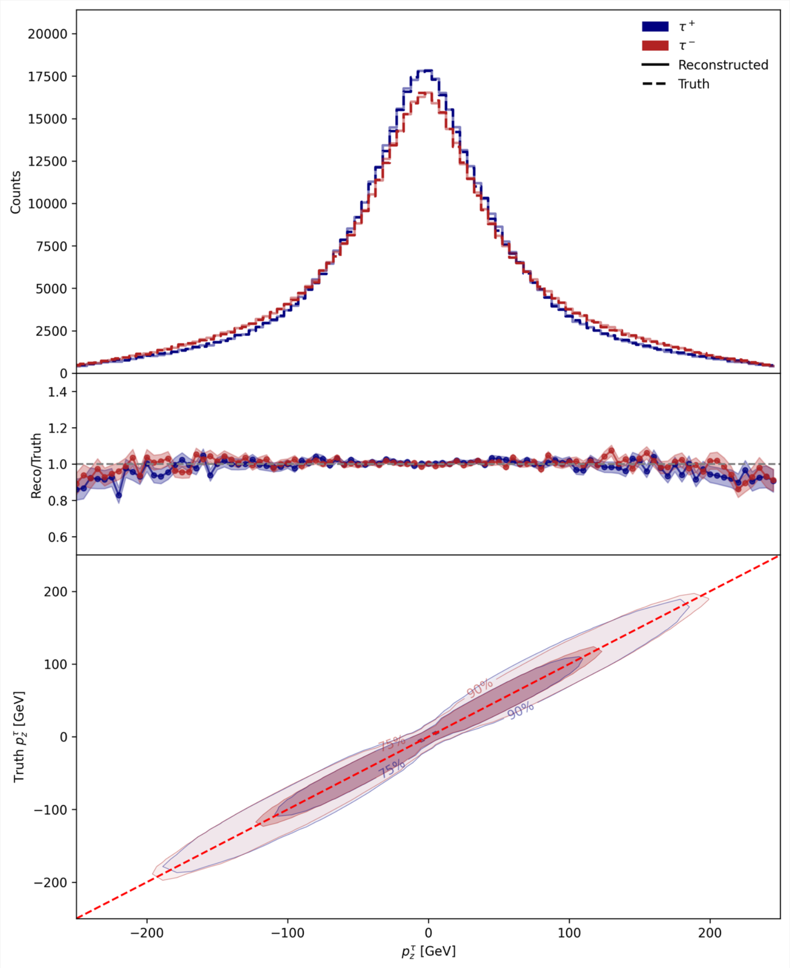


Neutrino Reconstruction - $l\pi$

- We have achieved **excellent reconstruction results** for the three-momentum of τ.

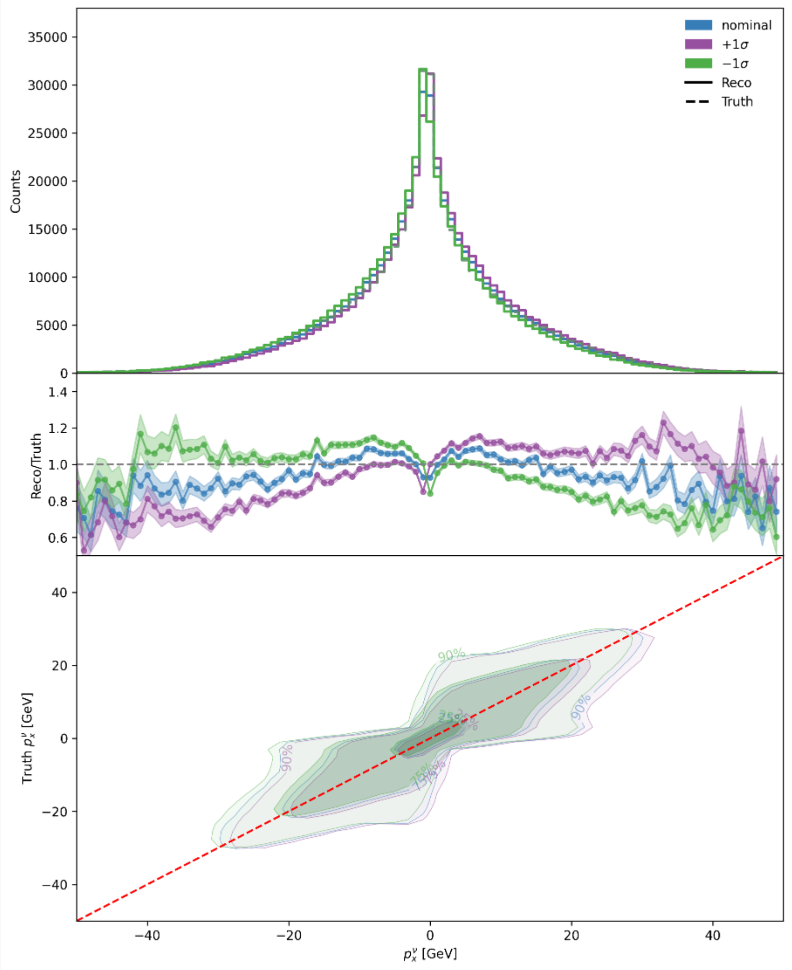


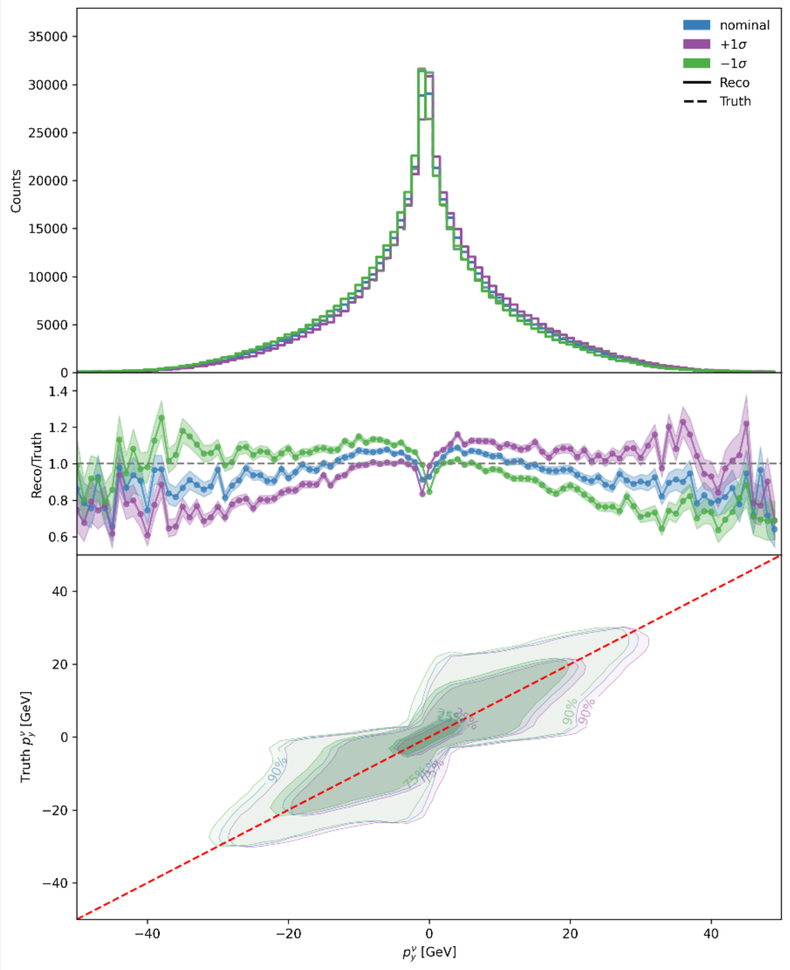


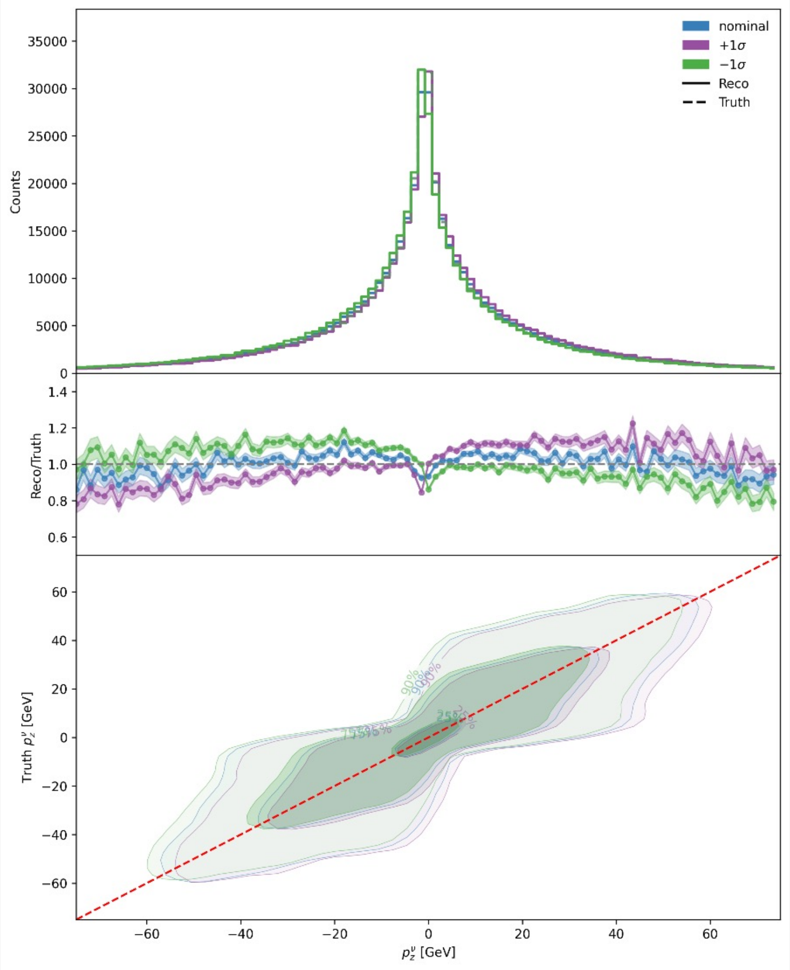


Neutrino Reconstruction - Systematics

- The stability of the **output** of the network is evaluated by **shifting the output by $\pm\sigma$**;







Systematics

	SR Only			SR & Trigger		
	$\rho\rho$	$\mu\rho$	Combined	$\rho\rho$	$\mu\rho$	Combined
All Systematics	**29.81%**	**29.76%**	**31.29%**	**31.00%**	**29.82%**	**31.35%**
MC Statistics	29.31%	29.56%	30.05%	28.93%	29.55%	28.66%
Luminosity	0.12%	0.08%	0.90%	7.73%	0.74%	6.10%
Background Cross-Section	3.39%	0.07%	2.01%	1.51%	0.05%	1.06%
Signal Cross-Section	0.23%	0.23%	1.71%	3.12%	0.52%	2.41%
Tau Energy Scale	1.47%	2.50%	2.12%	1.20%	0.89%	1.47%
Jet Enery Scale	1.67%	1.50%	4.49%	2.41%	1.72%	8.05%
Soft MET (p_x, p_y)	3.66%	1.90%	6.57%	6.68%	3.42%	7.11%
ν Sampling	0.02%	0.02%	0.03%	0.06%	0.03%	0.07%

Training and Evaluation details

李政道研究所 TSUNG-DAO LEE INSTITUTE

- The dataset is split into **training (80%)**, **validation (20%)** subsets;
- Training is conducted on 16 NVIDIA A100 GPUs using Horovod on the **Perlmutter Supercomputer**;
- The learning rate begins at 1.2×10^{-4}, following a warm-up phase and cosine decay schedule.
- We employ the **Lion optimizer** and implement an **early stopping** strategy;
- During evaluation, we generate **10 candidates** for each event and **randomly select one** as the final result to mitigate any bias introduced by the generation model;
- For DDIM, the parameters are set as $\boldsymbol{N_{steps} = 100,\ \eta = 1.0}$;

Training Times

- Below is a table presenting the **training times**:

Subchannel	Time per epoch/s	Number of training epochs
$\pi\pi$	31	613
$e\pi$	30	1000
$\mu\pi$	39	1000
$\pi\rho$	11	381
$e\rho$	9	488
$\mu\rho$	8	534
$\rho\rho$	11	900

- For evaluation, approximately **15 minutes** are required to process around **470k events**, each with **10 candidate** outputs.

Neutrino Reconstruction – $l\pi$

- We randomly designated 10% of background jets as τ for fake estimation for $l\pi$ subchannel;
- The plots demonstrate that the network **accurately reproduces** the $\tau\tau$ system mass peak, rather than simply clustering all physical processes around 90 GeV.