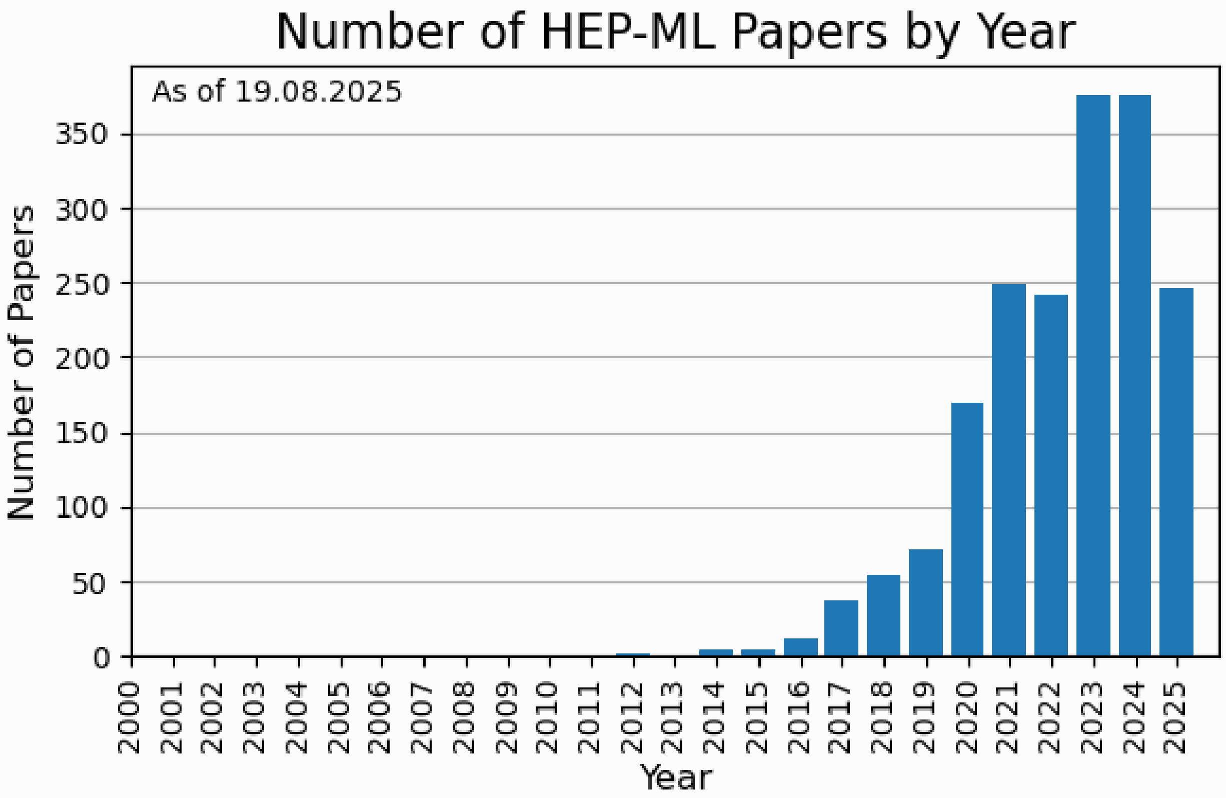
The 11th China LHC Physics Conference (CLHCP2025)

AI and Machine Learning Application in High Energy Physics

Liang Li 李亮

Shanghai Jiao Tong University

AI and ML in High Energy Physics



[Living Review of Machine Learning for Particle Physics]

- **638 papers on the first page!**
- **Read them all, or, let AI do it for you ☺**

Impossible to cover everything in 20 minutes

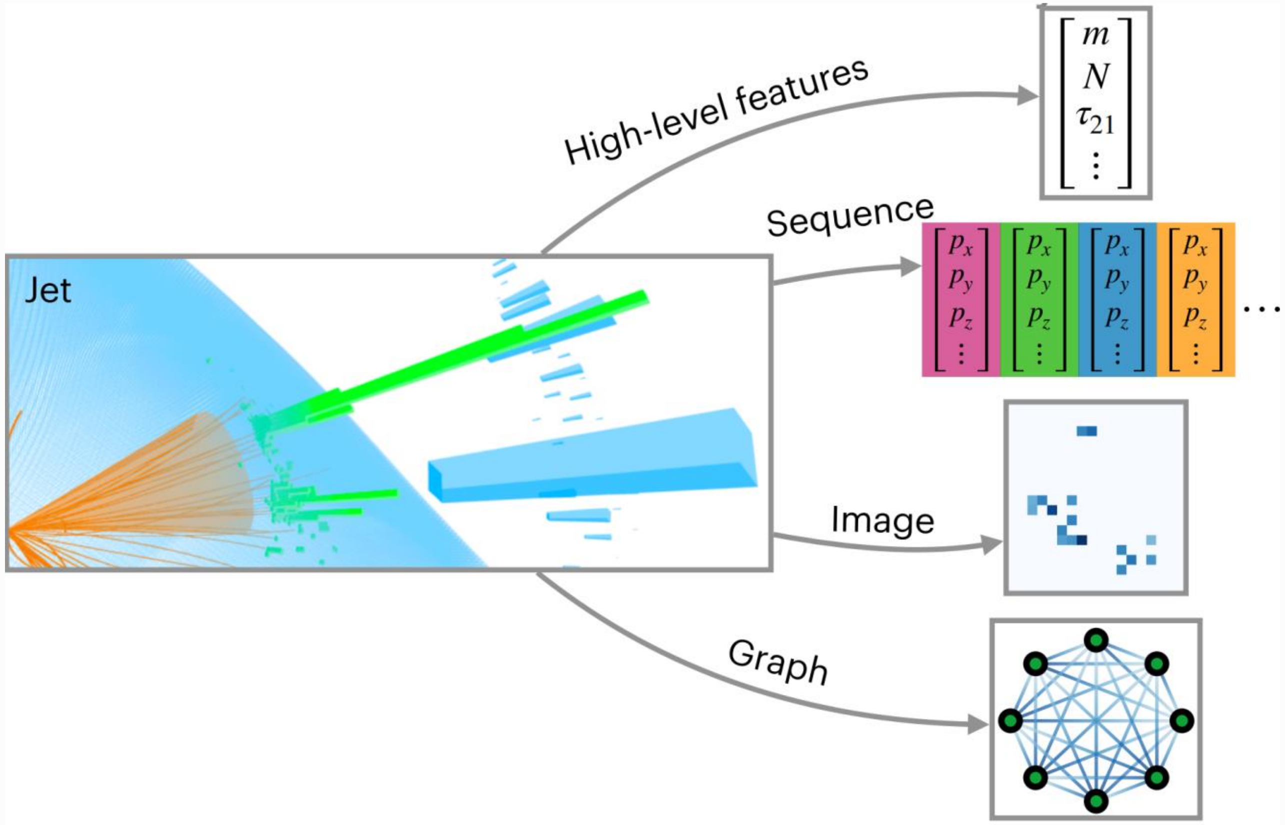
- **Highly selective and apologize for missing some important work, especially to our theory friends**

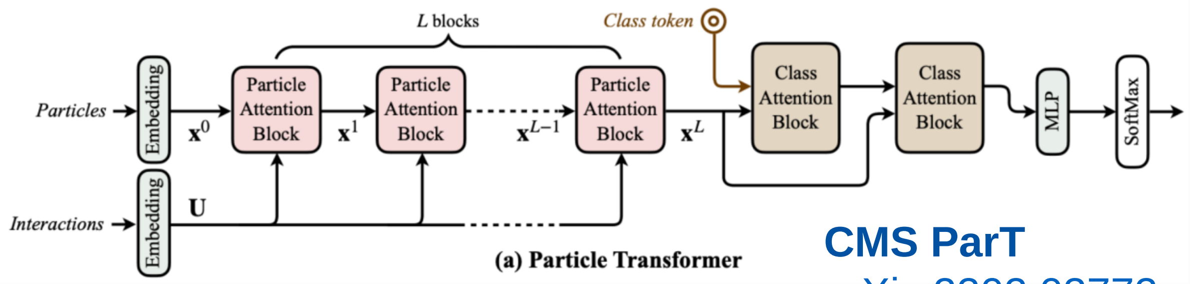
AI and ML in High Energy Physics

SJTUPA 上海交通大学物理与天文学院

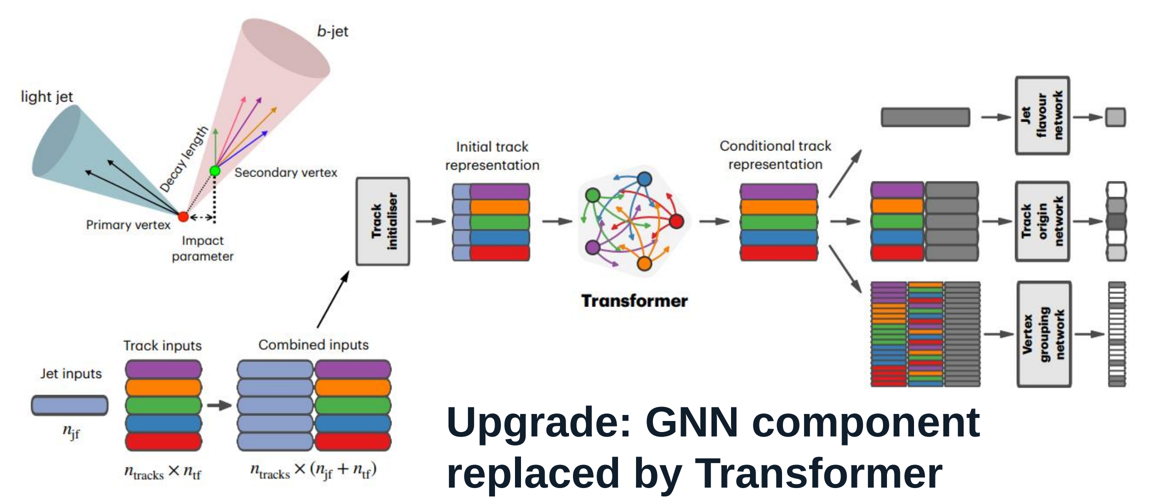
- ✓Classifier (Supervised)
- ✓Self-guided Detection/Search (Weakly Supervised/Unsupervised)
- ✓Reconstruction
- ✓Simulation
- ✓Unfolding (Not covered)

Smarter and More Sophisticated Classifier





CMS ParT
arXiv:2202.03772



Upgrade: GNN component replaced by Transformer

ATLAS General Network 2
arXiv:2505.19689

Smarter and More Sophisticated Classifier

Recap: Series of GloParT tagger (v1-v3)

Sep 2021

Initiated project: **ParticleNet for H→VV tagging** for novel boosted H→WW sea

JME presentations[22.01.11 by Cristina/Zichun][22.03.08 by Dawei][22.08.09 by Dawe

ParticleNet for HVV

Process	Final state/ prongness	heavy flavour	# of classes
H→WW (full-hadronic)	qqqq	0c/1c/2c	3
	qqq		3
H→WW (semi-leptonic)	eνqq	0c/1c	2
	μνqq		2
	τ_eνqq		2
	τ_μνqq		2
	τ_hνqq		2
H→qq		bb	1
		cc	1
		ss	1
		qq (q=u/d)	1
H→ττ	$\tau_e\tau_h$		1
	$\tau_\mu\tau_h$		1
	$\tau_h\tau_h$		1
t→bW (hadronic)	bqq	1b + 0c/1c	2
	bq		2
t→bW (leptonic)	beν	1b	1
	bμν		1
	bτ_eν		1
	bτ_μν		1
	bτ_hν		1
QCD		b	1
		bb	1
		c	1
		cc	1
		others (light)	1

Sep 2022

GloParT v1

[22.12.13 by Congqiao (JME)]

- upgrade to **ParT architecture** (so-called GloParT v
- propose: “large model for large-scale classification” & subsequent fine-tuning capabilities (concept of Global ParT)

GloParT v1:
used in the following analysis
- Boosted bbWW search: **CMS-PAS-HIG-23-012**
- HWW (0l/1l/VH): **HIG-24-008**
- X→H(bb)Y(WW): **B2G-23-007**

[23.02.22 by Congqiao (ML Forum)]

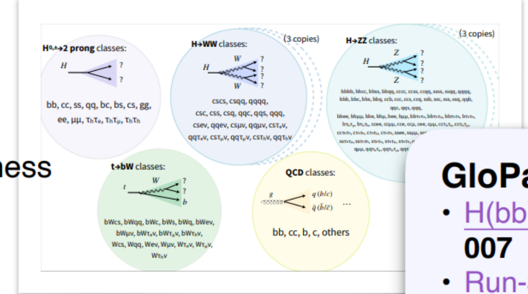
the study of fine-tuning (f.t.) capabilities - **first demonstration**

May 2023

GloParT v2

[23.07.24 by Congqiao (BTV)]

significantly improve pre-training comprehensiveness (extend to 314+2 nodes)



[23.09.13 ML Town Hall] [23.12.05 ML Forum]

- re-studied f.t. capability with the stronger model
- study of f.t. for anomaly detection

GloParT v2:
- H(bb)+γ: **B2G-24-007**
- Run-3 VH(bb/cc) (AK15): **HIG-25-001**
- boosted W→cb
- Run-3 HH(4b)

July 2024

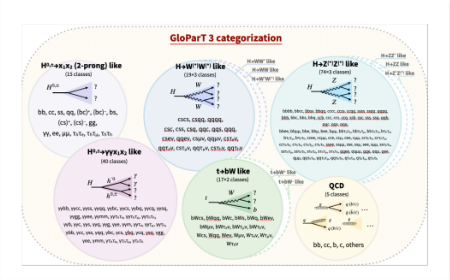
GloParT v3

[24.09.20 by Congqiao (BTV)]

- v3 enhencement + extend to 374+374+2 nodes

[24.10.18 Cross-POG]

A comprehensive review prepared for GloParTv3’s integration into cmssw



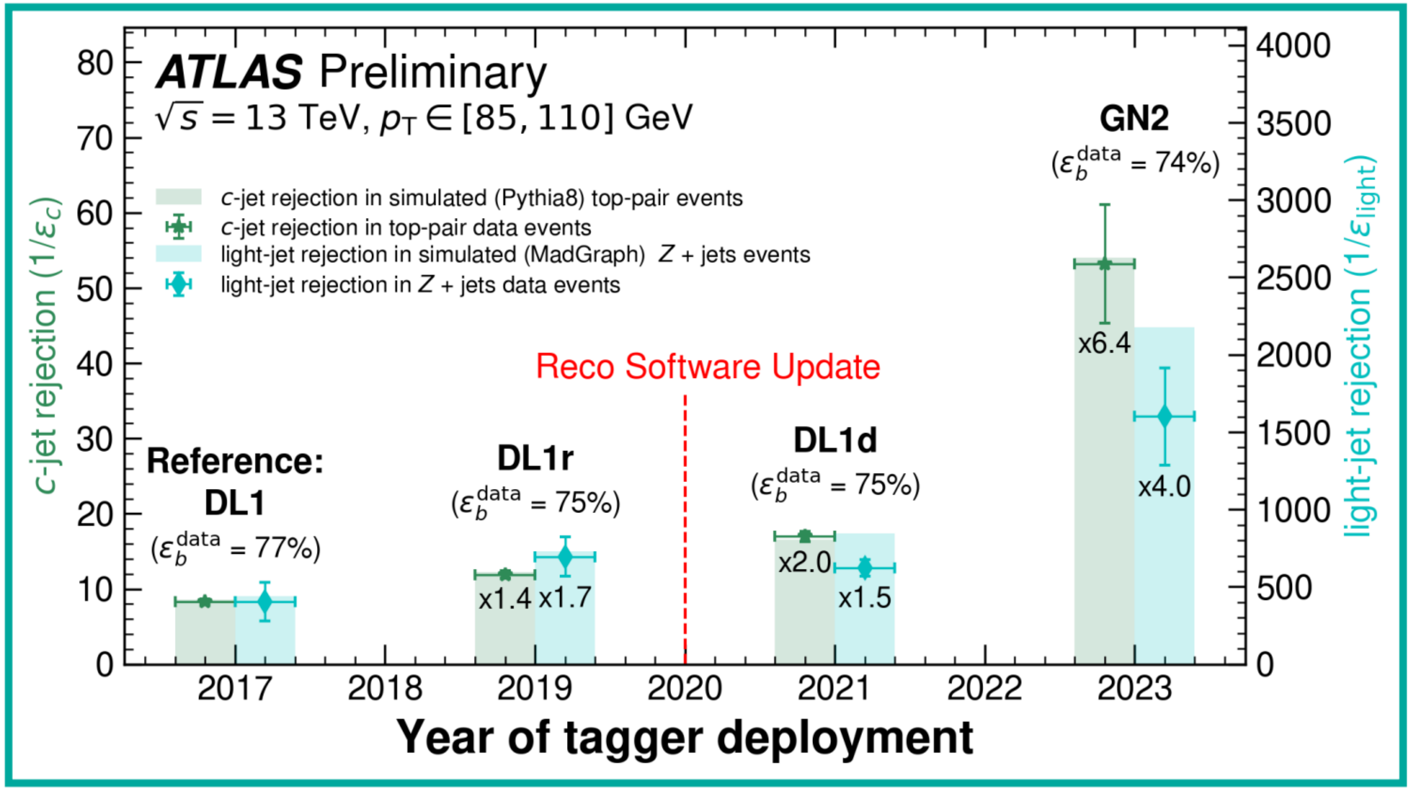
GloParT v3:
- Run-3 HH(4b)
- (more to join)

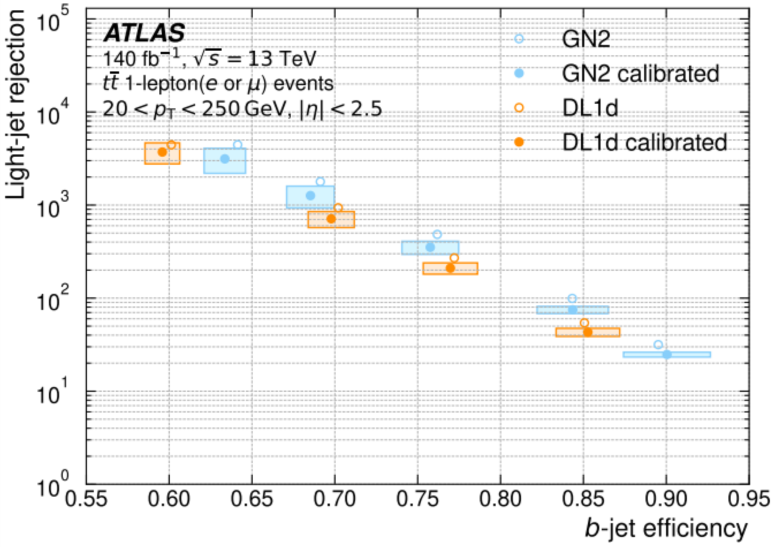
...

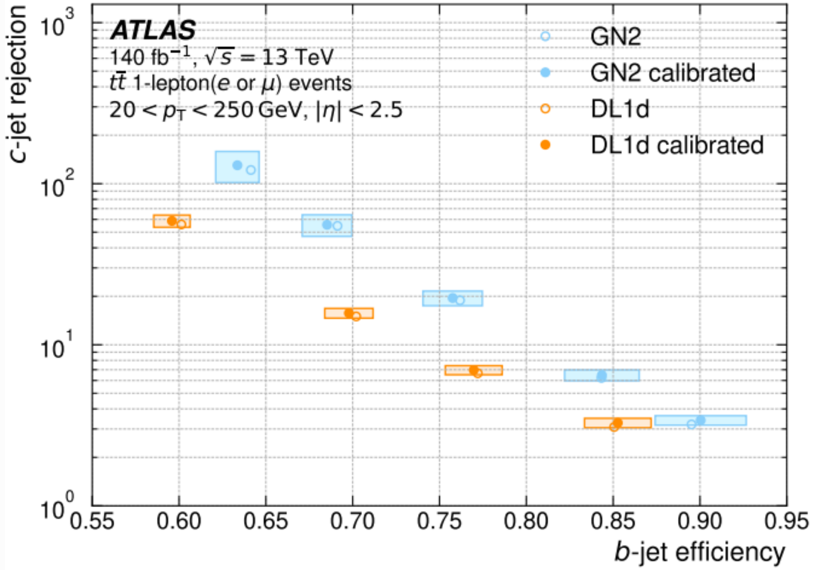
***Global Particle Transformer* (GloParT) algorithm**

Transformer model is based on “self-attention” mechanism: **Transfer model can focus on certain parts of the input data**, giving more weight to crucial features and disregarding unimportant ones.

Better and More Robust Performance



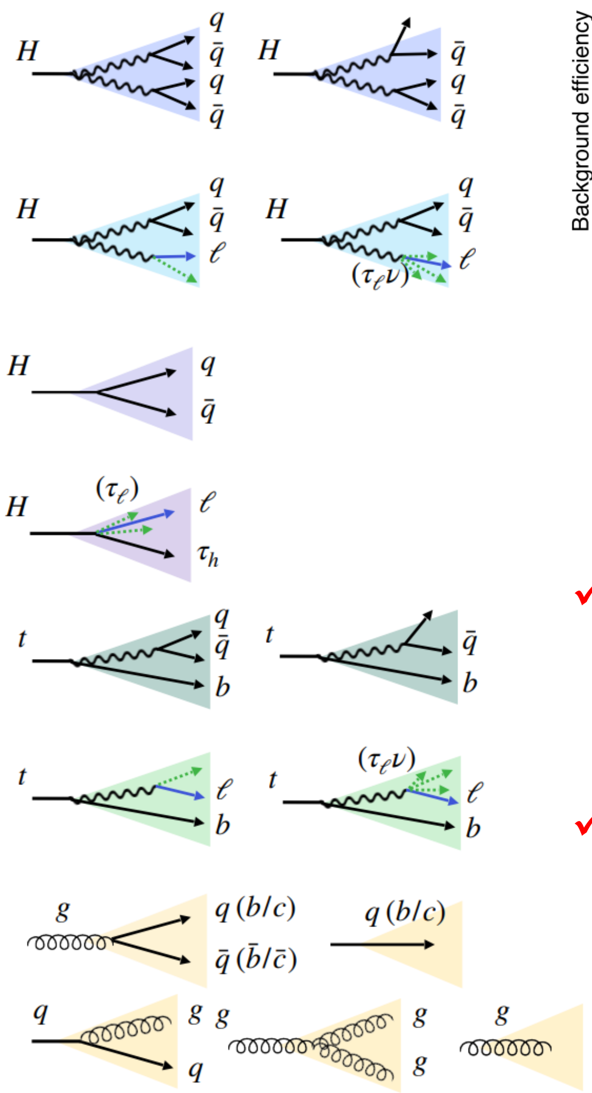




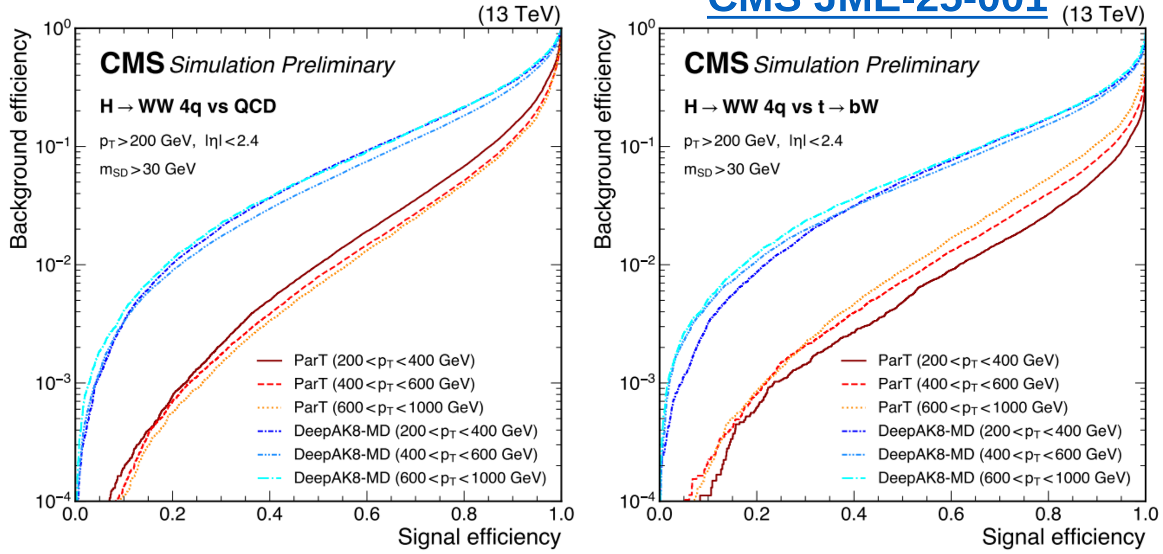
- ✓ **Significant improvement after years of development**
- ✓ **Essential calibrations done for b-/c-jet and light jet flavors**
- ✓ **Performance in data matches simulation after calibration**

Better and More Robust Performance

Process	Final state		Flavor	# of classes
H→WW (full-hadronic)	qqqq	⊗	0c / 1c / 2c	3
	qqq			3
H→WW (semi-leptonic)	eνqq	⊗	0c / 1c	2
	μνqq			2
	$\tau_e\nu qq$			2
	$\tau_\mu\nu qq$			2
	$\tau_h\nu qq$			2
H→qq		⊗	bb	1
			cc	1
			ss	1
			qq (q=u/d)	1
H→ττ	$\tau_e\tau_h$			1
	$\tau_\mu\tau_h$			1
	$\tau_h\tau_h$			1
t→bW (hadronic)	bqq	⊗	1b + 0c / 1c	2
	bq			2
t→bW (leptonic)	beν	⊗	1b	1
	bμν			1
	$b\tau_e\nu$			1
	$b\tau_\mu\nu$			1
	$b\tau_h\nu$			1
QCD			b	1
			bb	1
			c	1
			cc	1
			others (light)	1

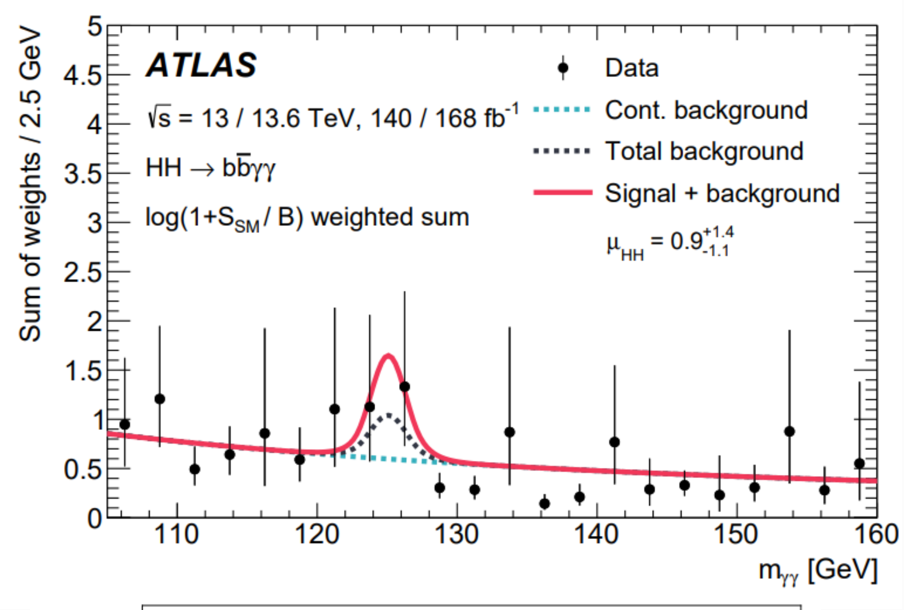


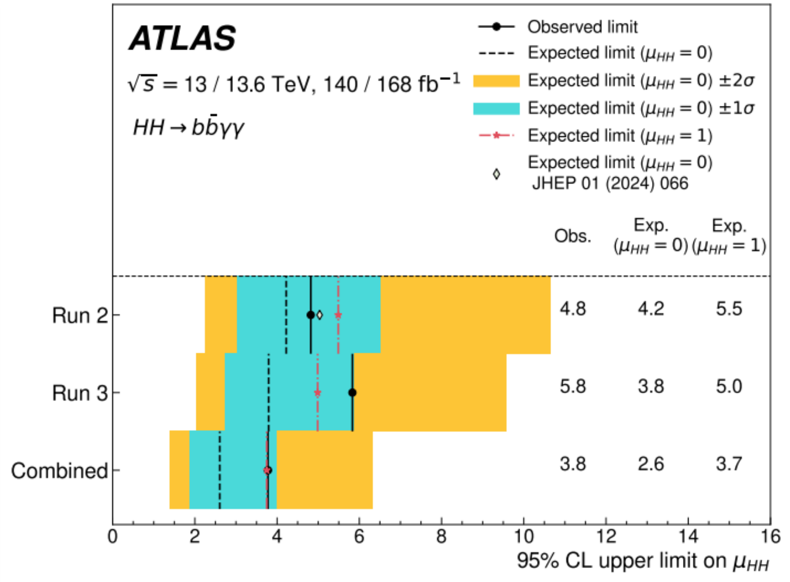
[CMS JME-25-001](#)



- ✓ **Highly granular multi-classifier gives 6-20 fold improvement in background rejection rate on H→WW* →4j vs. QCD/top jets**
  - **Compared with early DeepAK8-MD tagger**
- ✓ **Challenge for tagger calibrations**
  - **Hard to find SM events in similar topology**
  - **New technique uses Lund jet plane**
  - **Effectively measure scale factors per quark sub-jet**

Impact on Physics

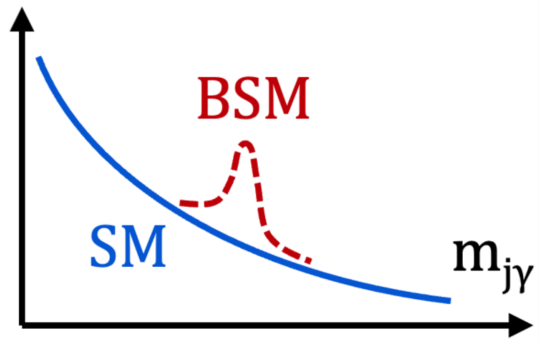


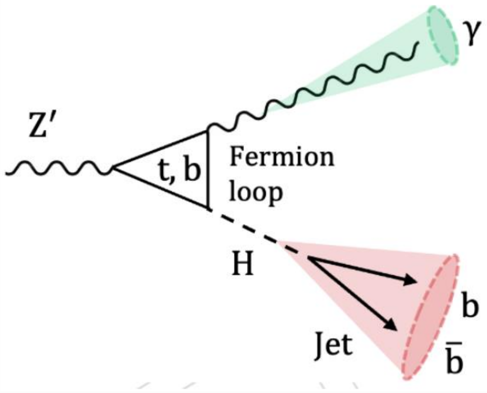


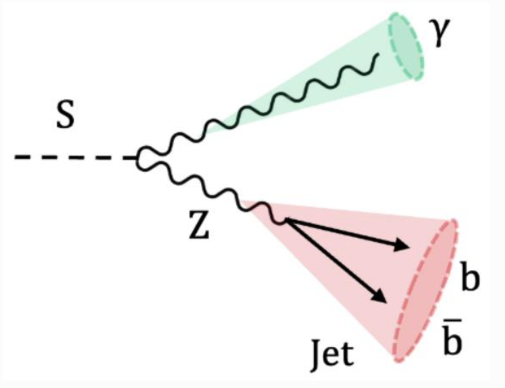
GN2 alone brings 20% improvement for HH → bbγγ analysis

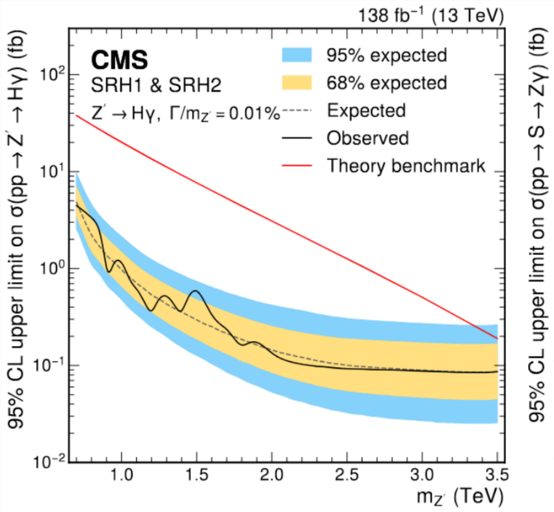
[arXiv:2507.03495](https://arxiv.org/abs/2507.03495)

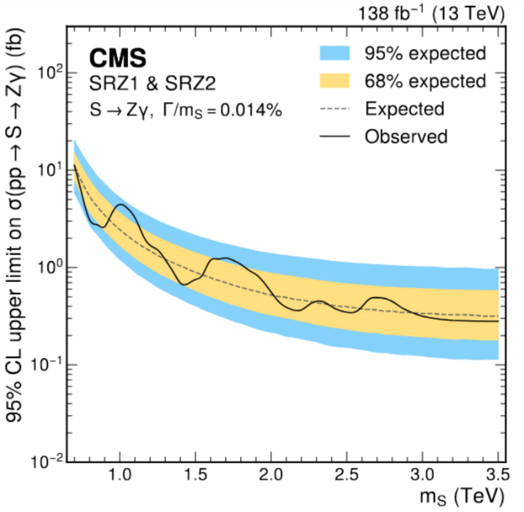
High-mass resonances in H/Z(bb)+γ final state





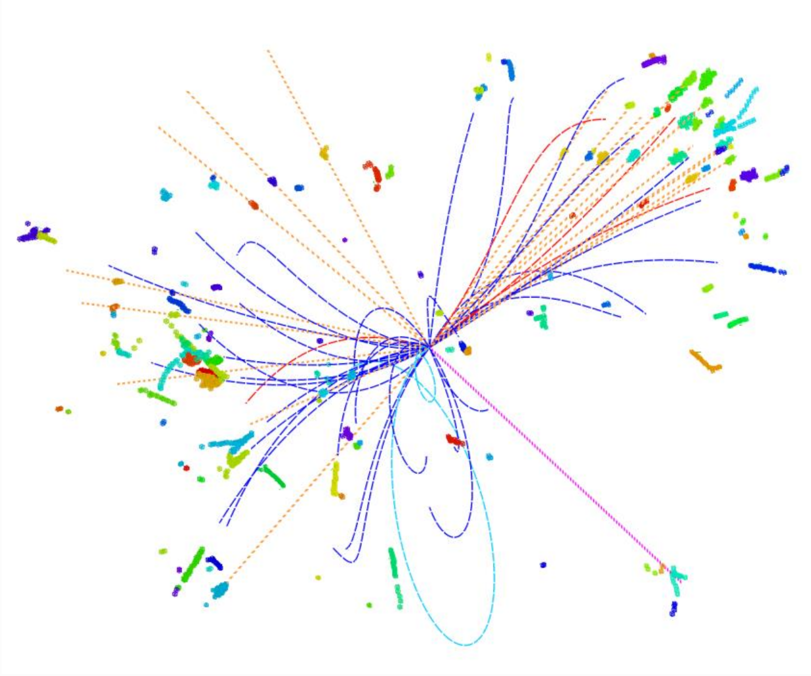




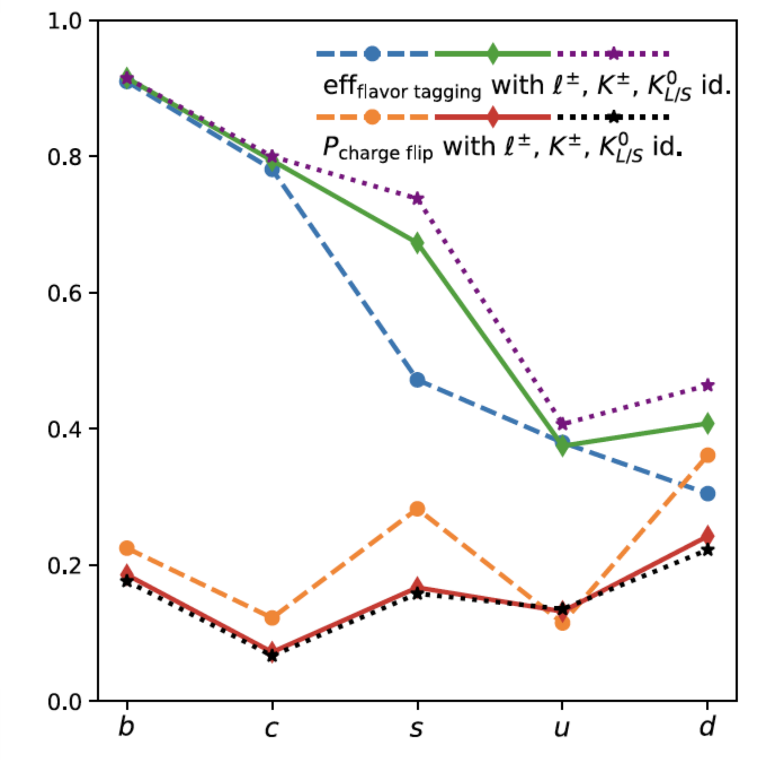


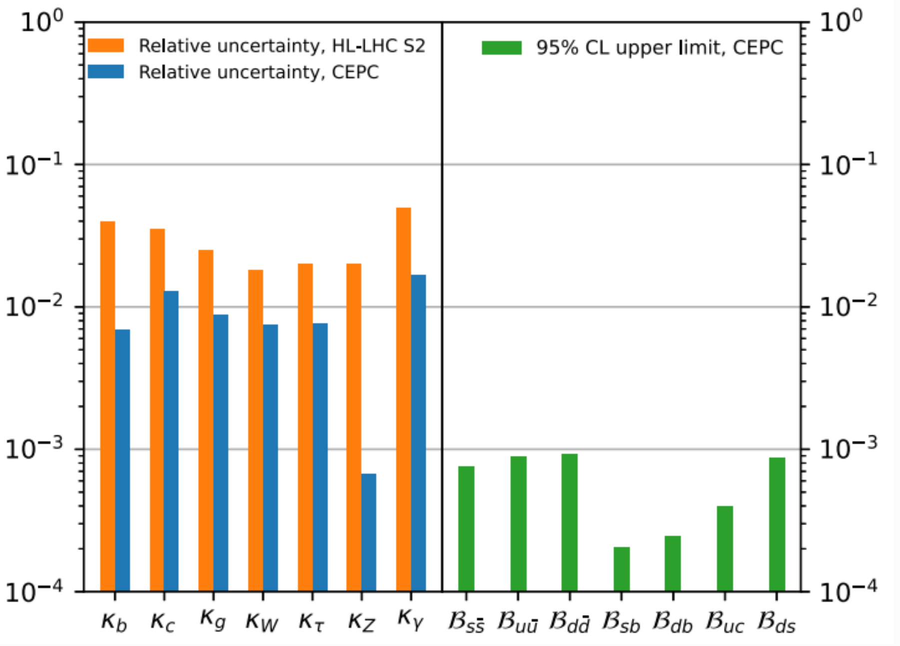
- ✓ **GloParT V2 used for X→bb tagger**
  - **H/Z→bb vs. QCD jets**
- ✓ **Most stringent limits for both channels**

Impact on Physics



$e^+e^- \to \nu\bar{\nu}H \to \nu\bar{\nu}gg$





PRL 132, 221802 (2024)

Jet Origin ID:

- **11 categories (5 quarks + 5 anti quarks + gluon) identification, realized at Full Simulated di-jet events at CEPC CDR baseline with Arbor + ParticleNet (GNN).**
- **Jet flavor tagging efficiencies ranging from 67% to 92% for b-, c-, and s-quarks and jet charge flip rates of 7%–24% for all quark species. Higgs decay BRs range from 2×10^{-4} to 1×10^{-3} (95% C.L.).**

Let’s Go Even Bigger: Event-Level Foundation Model

Core Idea: One strong body + many small heads

🧠 Decoder – Discriminative Heads：

Segmentation

- Inspired by Meta AI’s segmentation networks
  - The model performs set prediction (queries → predict class & mask), preserving permutation symmetry.
  - Naturally extendable from objects to substituents without changing the model design.

Classification

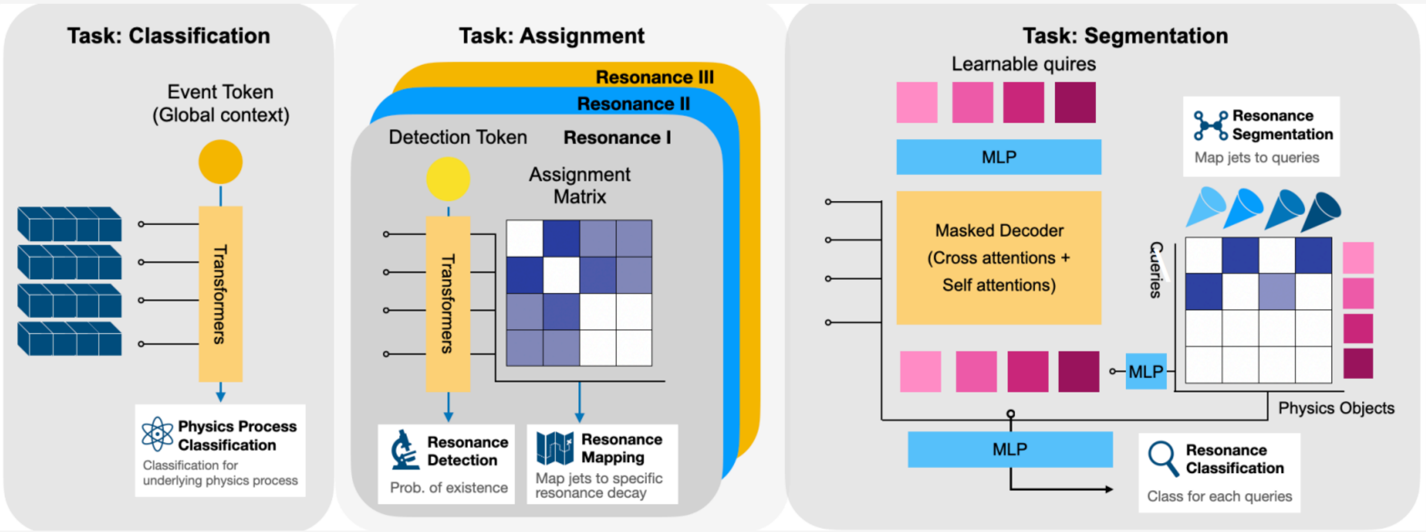
- Multi-Class event classifiers (with regression)

Assignment

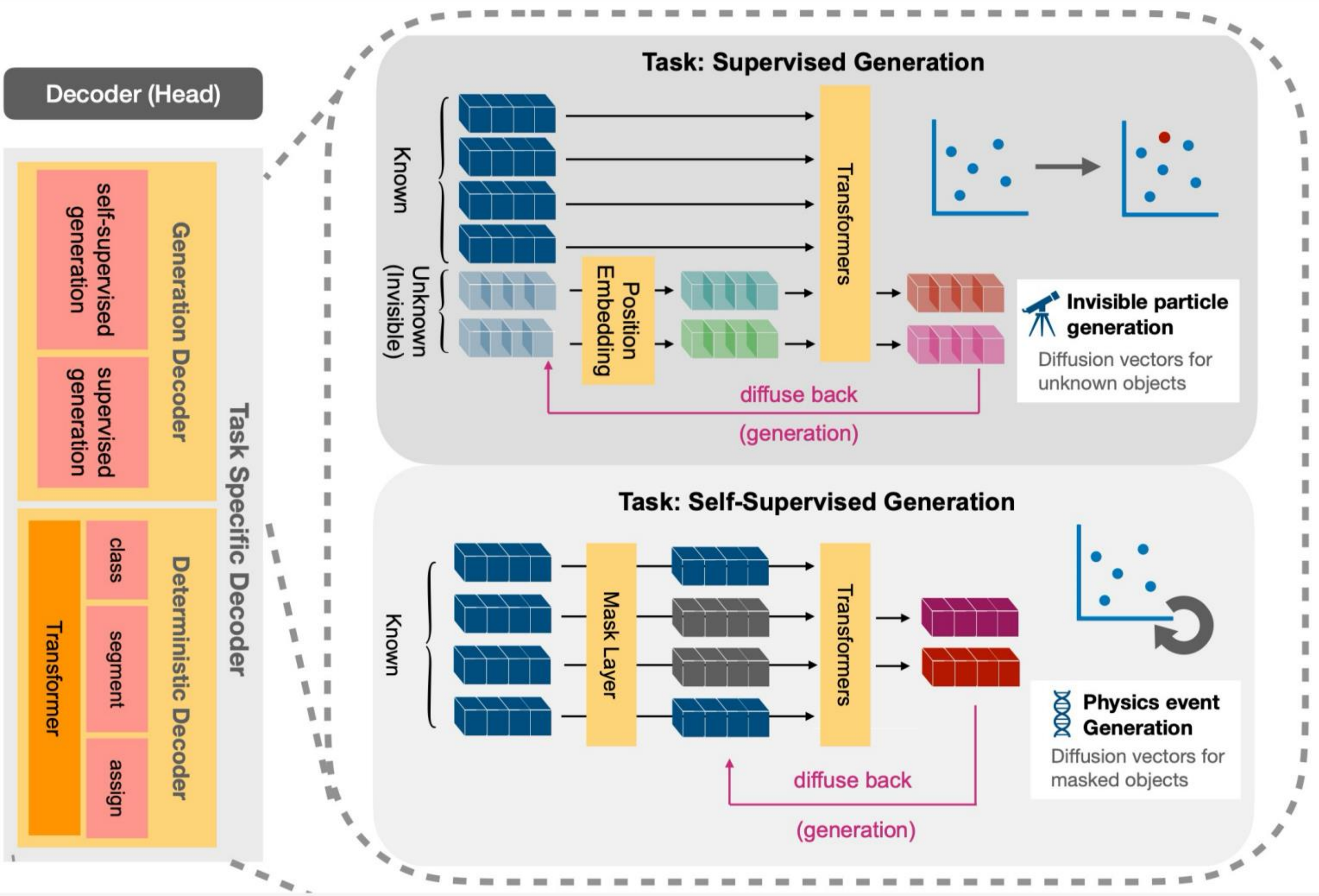
- Symmetry-aware mapping of objects to truth partons (requires known decay topology).
- High accuracy for well-defined processes, but rigid, costly, not generalizable.

🔢 Input Representation

- 📌 **Particle Cloud (Up to 18 Particles per Event):**
  - Each particle is encoded with 7 features: **4-momentum, isbJet, isLepton**, and **charge.**
- 🌐 **Global Features / Event Observables:**
  - Missing transverse energy
  - Number of leptons, number of jets
  - Invariant mass of visible objects
  - Scalar sums like **HT, ST**, etc.



EveNet: Event-Level Foundation Model



Core Idea: One strong body + many small heads

🧠 Decoder – Generation Head：

Supervised Generation

- Use known objects as input to predict missing ones (e.g., neutrinos).
- Diffusion models capture high-dimensional probability densities → predict the most likely kinematics.

Self-supervised Generation

- Mask part/all of the inputs and reconstruct them with a diffusion model.
- Learns underlying event structure without requiring labels.

Weakly Supervised: Search for Exotic Higgs Decay

✓ **Jointly trained on the Assignment and Classification tasks**

2-15x improvement on bkgd. rejection

- **Signal:** $H \rightarrow aa \rightarrow bbbb$

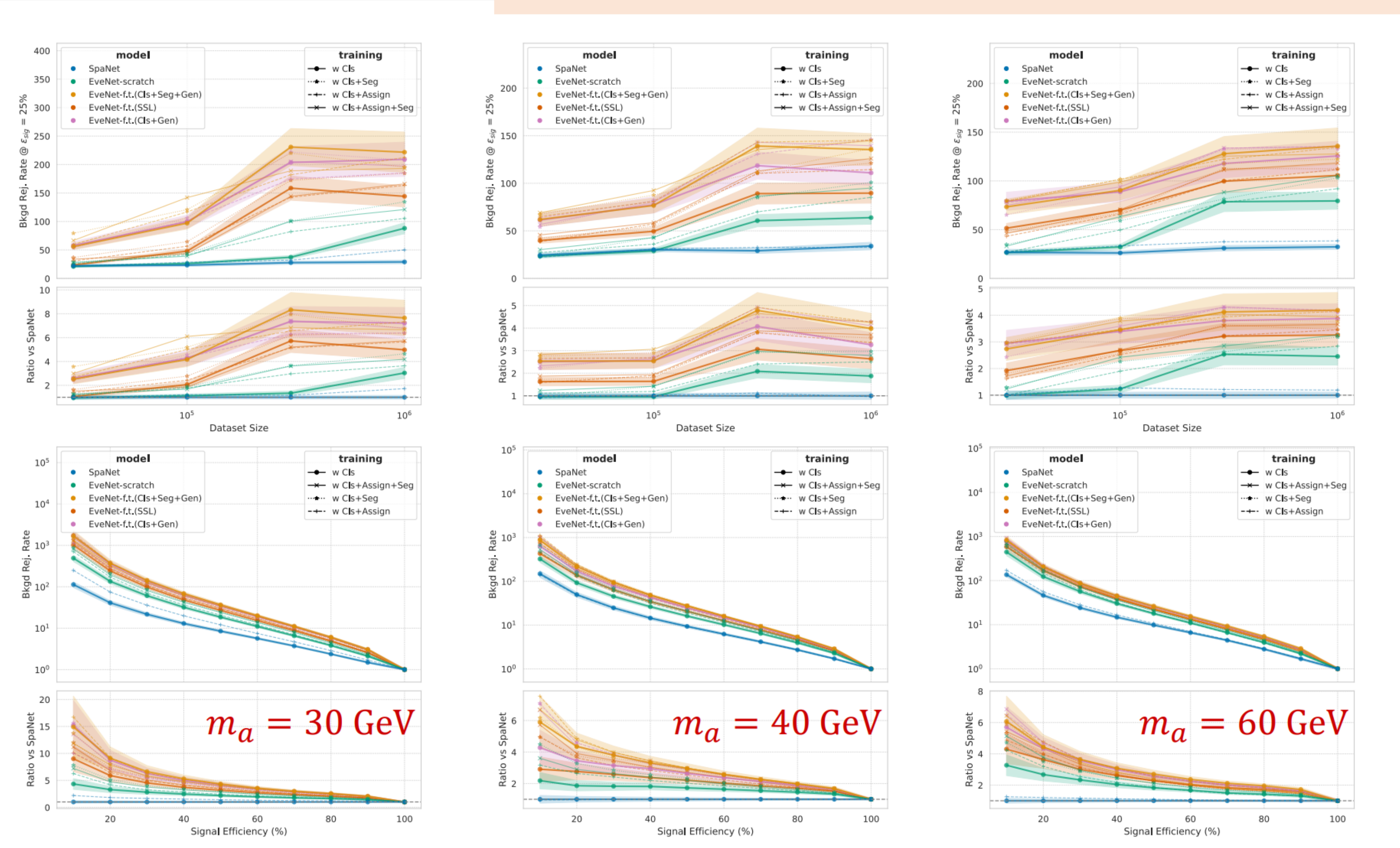
$(m_a = 30, 40, 60\ \text{GeV})$

- **QCD:** $bbbb,\ bbbj,\ bbjj$
- **Reference Network:** SPANet

(same hidden dim)

Classification

1. **Inversed ROC**
2. **Bkgd. Rejection rate @ signal efficiency of 25%**



Self-Supervised: Anomaly Detection

SJTUPA 上海交通大学物理与天文学院

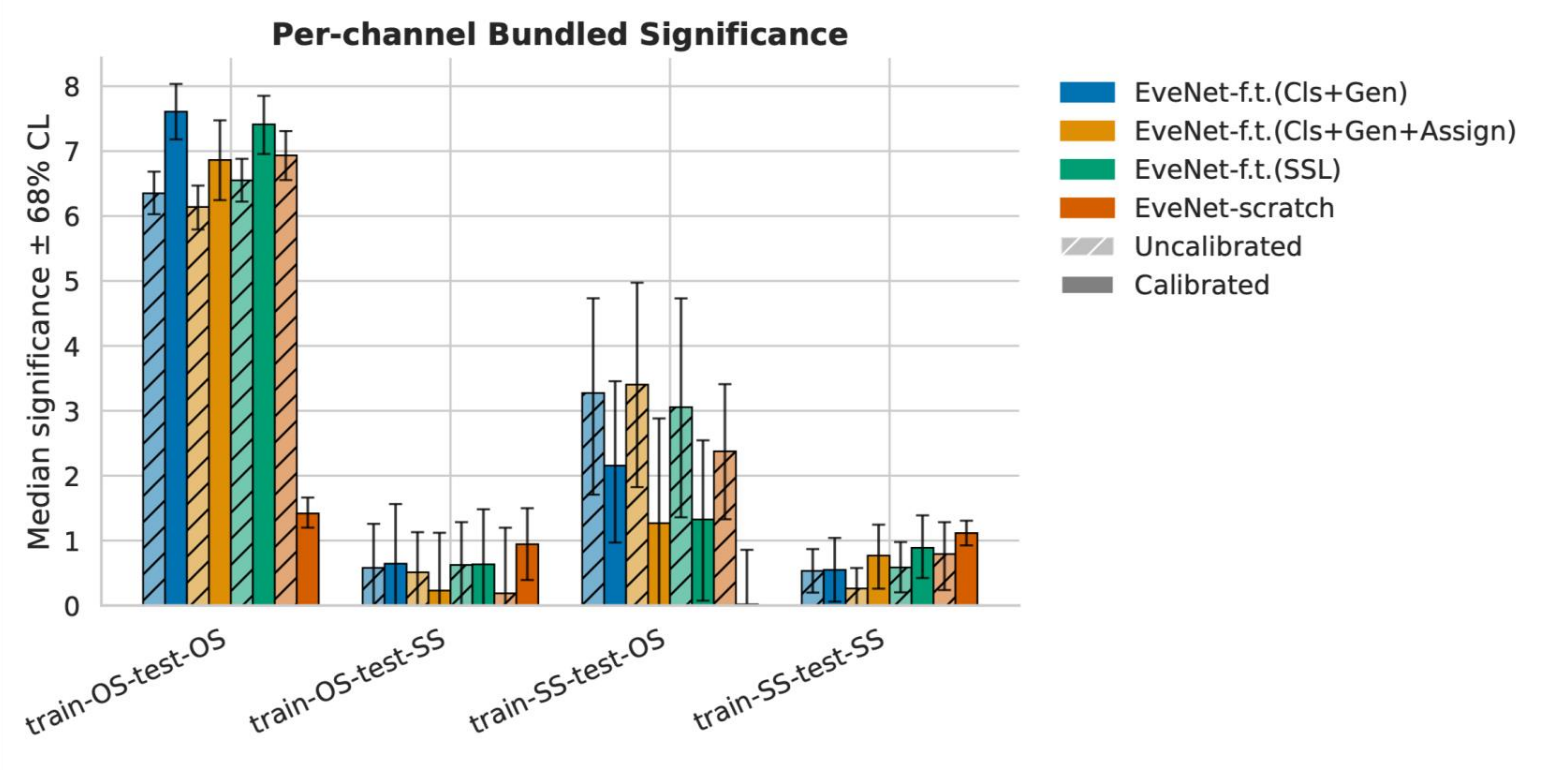
- ***Reference paper***: 2502.14036 (*To test EveNet's generative capability, we extend an existing anomaly detection method* ***using normalizing flows*** *by replacing it* ***with diffusion-based generation*** *of full 4-momentum)*
- **Dataset**: CMS Open Data (2016 DoubleMu primary dataset) targeting Υ resonances in di-muon final states.

Final Significance (ℓ-reweighting)

- **paper: 6.4σ**
- **EveNet-Pretrain: 7.5σ**
- ~~**EveNet-Scratch**: $?\sigma$~~ (**mass sculpting** ✕)

Note: *the energy regime here is even different from the main samples in pretrain*

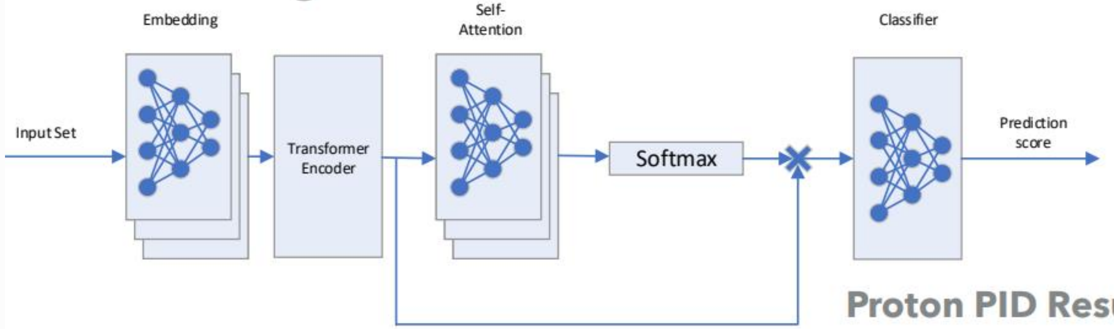
1σ improvement on significance

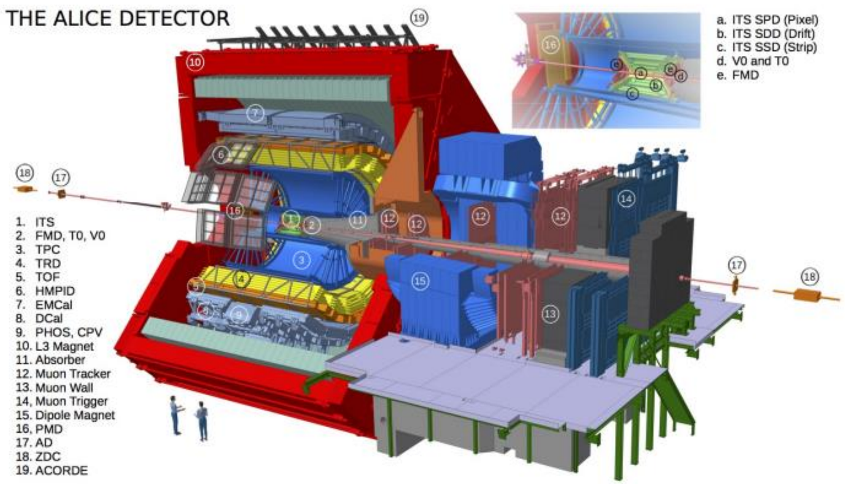


✓ **Trained on Self-Supervised Generation task**

Reconstruction: PID

- **Transformer** for particle ID in ALICE can result in higher purity and efficiency than standard methods
- Use **domain adversarial neural networks** to mitigate data-simulation differences

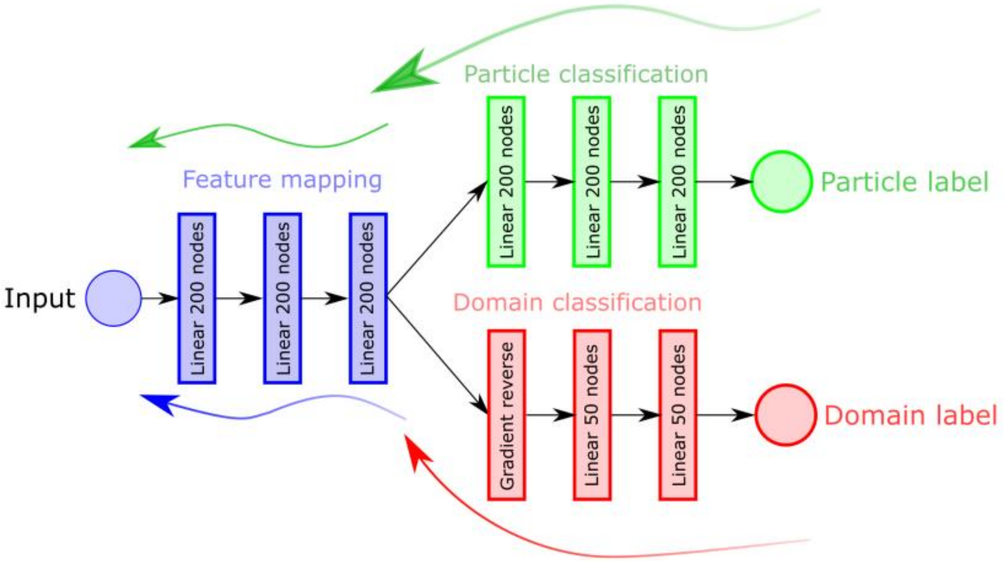


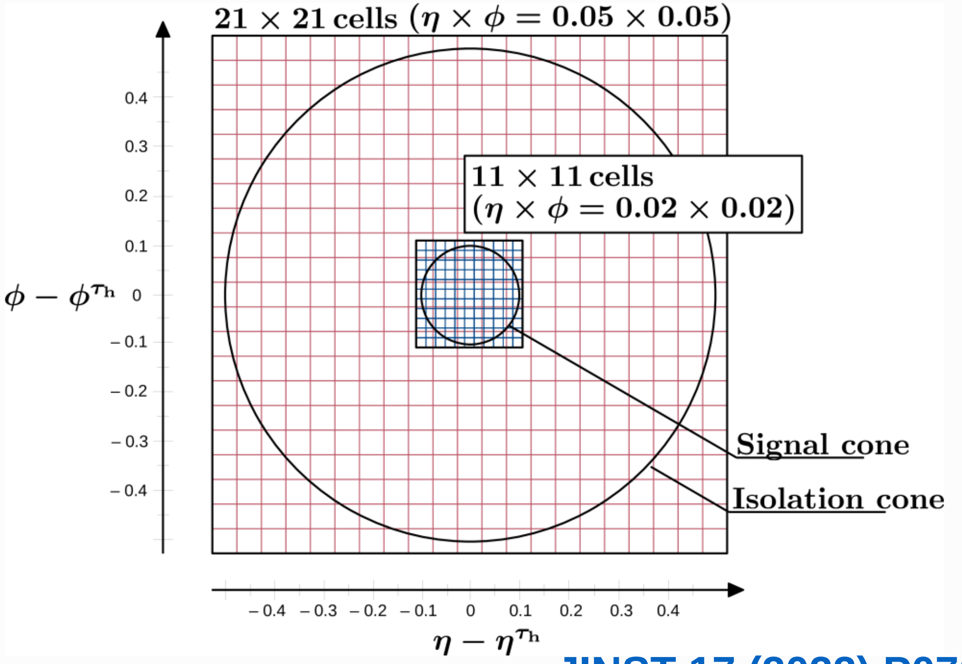


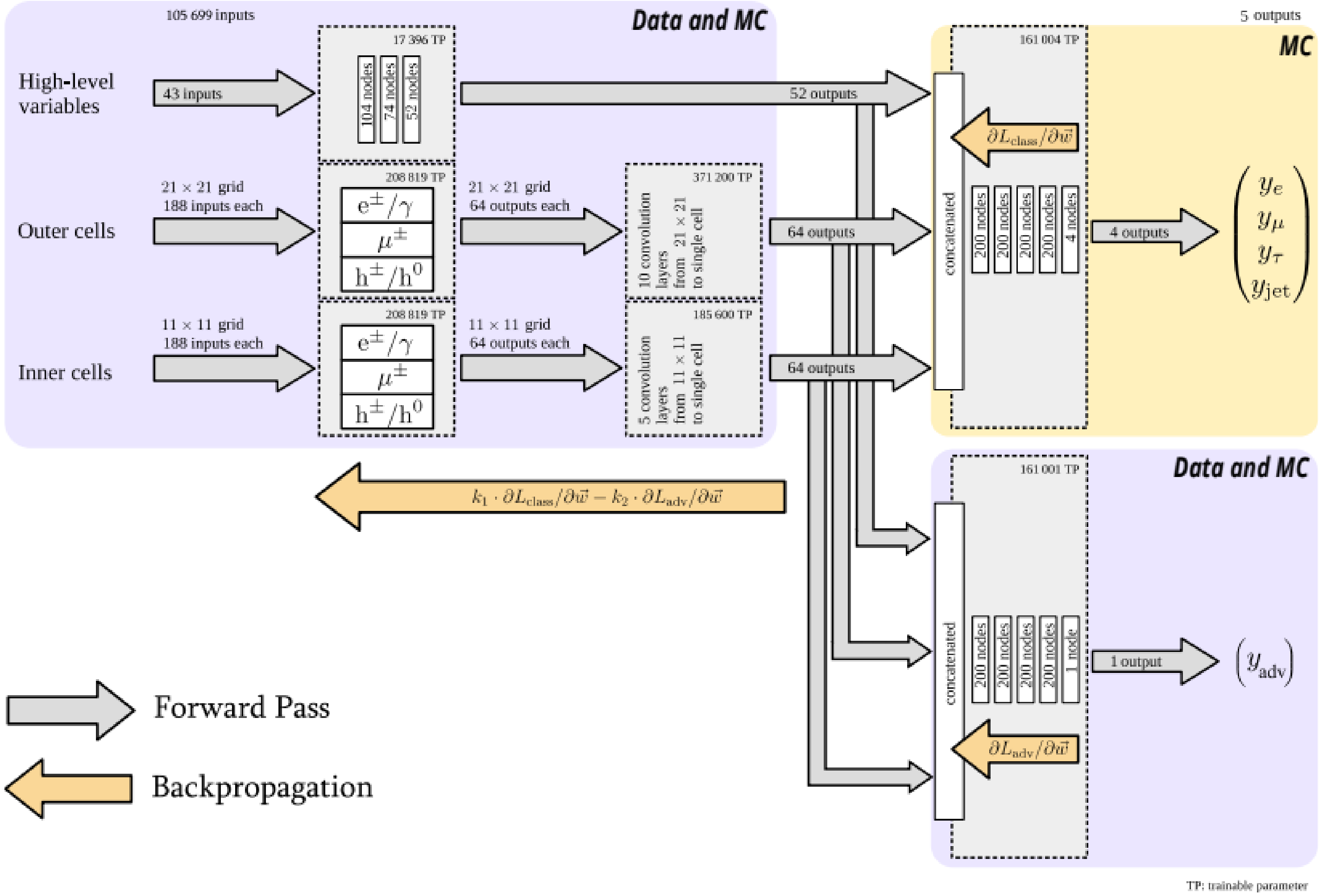
Proton PID Results

Model	Precision	Recall	F_1
Standard	**99.40 ± 0.01**	59.72 ± 0.03	74.61 ± 1.88
Ensemble	97.16 ± 0.46	93.74 ± 0.30	95.42 ± 0.12
Mean	97.85 ± 0.41	93.34 ± 0.32	95.54 ± 0.06
Proposed	97.80 ± 0.44	**93.86 ± 0.27**	**95.79 ± 0.07**
Regression	97.38 ± 0.40	93.67 ± 0.38	95.49 ± 0.15

Transformer





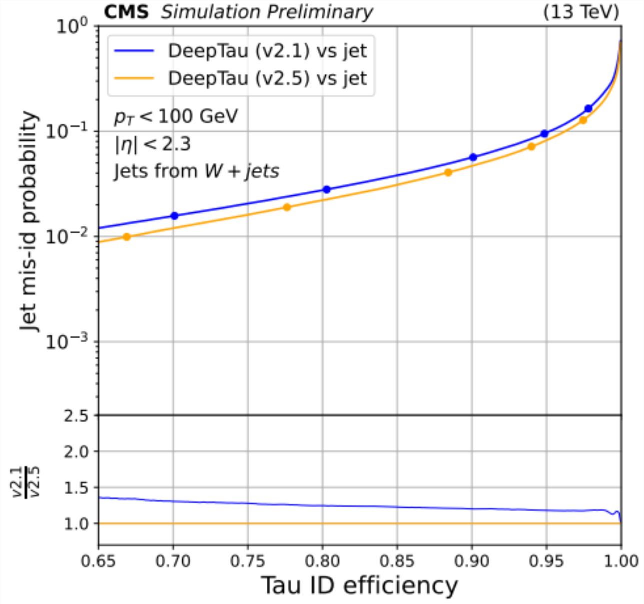


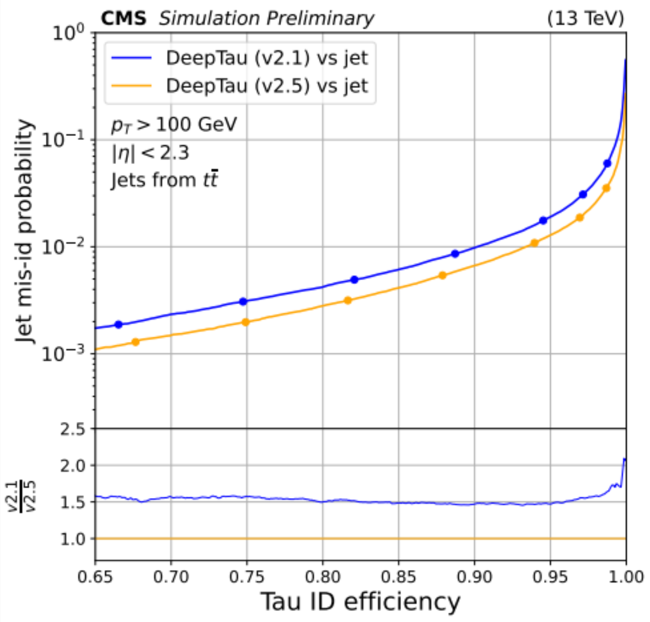
[JINST 17 (2022) P07023]

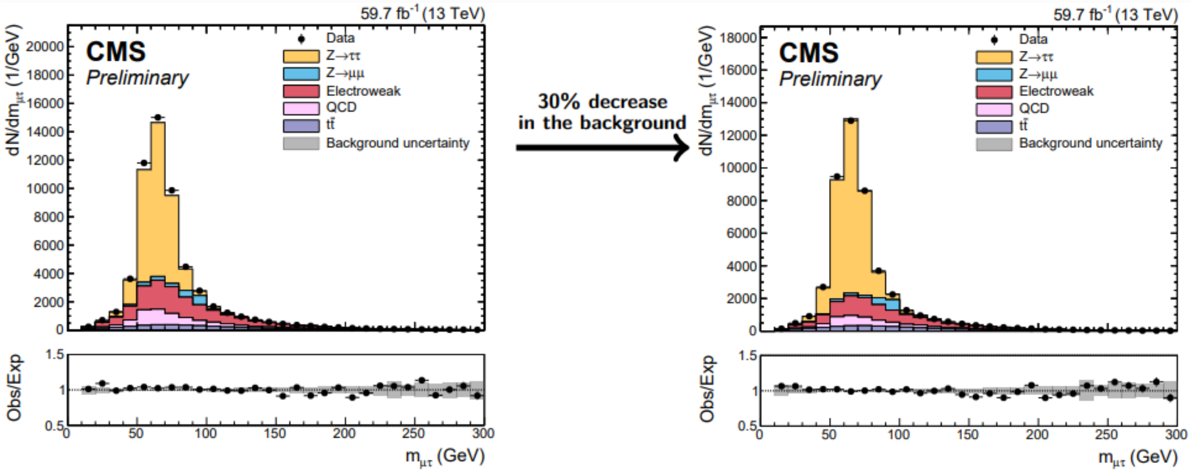
CMS DeepTau: multi-class tau identification algorithm based on CNN

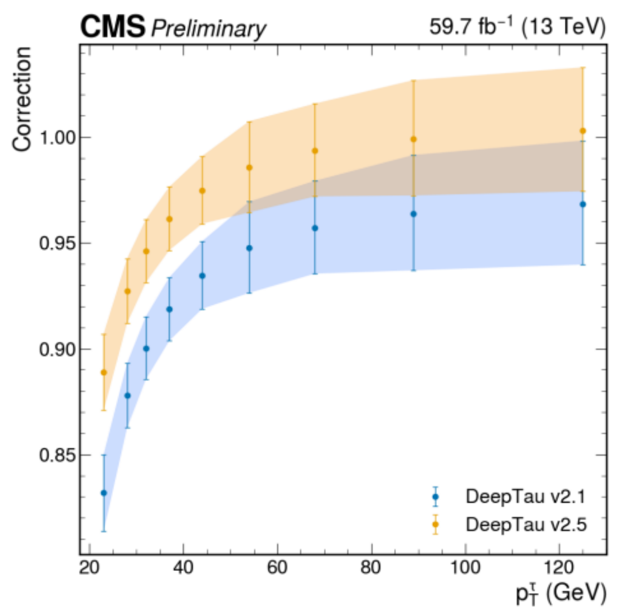
- **v2.5 adds domain adaptation subnetwork with adversarial training to deal with MC mismodelling in the high-purity region**
- **Better data handling**
- **Feature standardization, hyperparameter optimization**

Reconstruction: Tau ID









[CMS-DP-2024-063](CMS-DP-2024-063)

DeepTau v2.5 significant improvement compared to v2.1

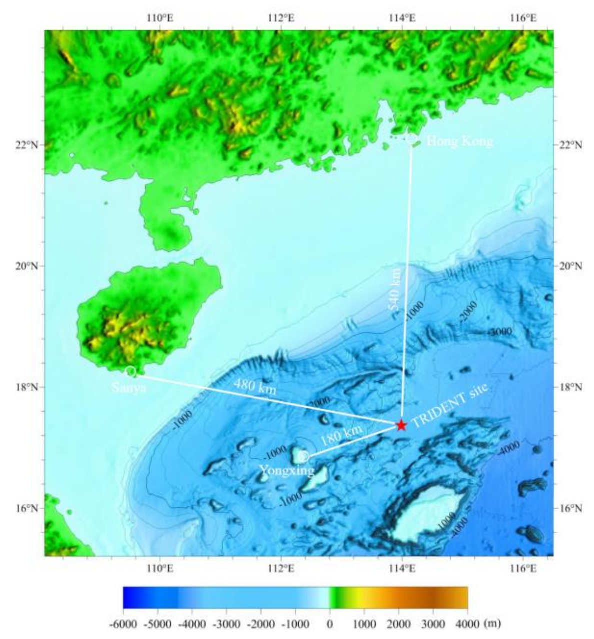
- **Jet misidentification reduced by ≈ 50%**
- **30% decrease in the background**
- **Data vs. MC scale corrections are closer to 1**
- **Minimizing dependence on MC mismodelling**

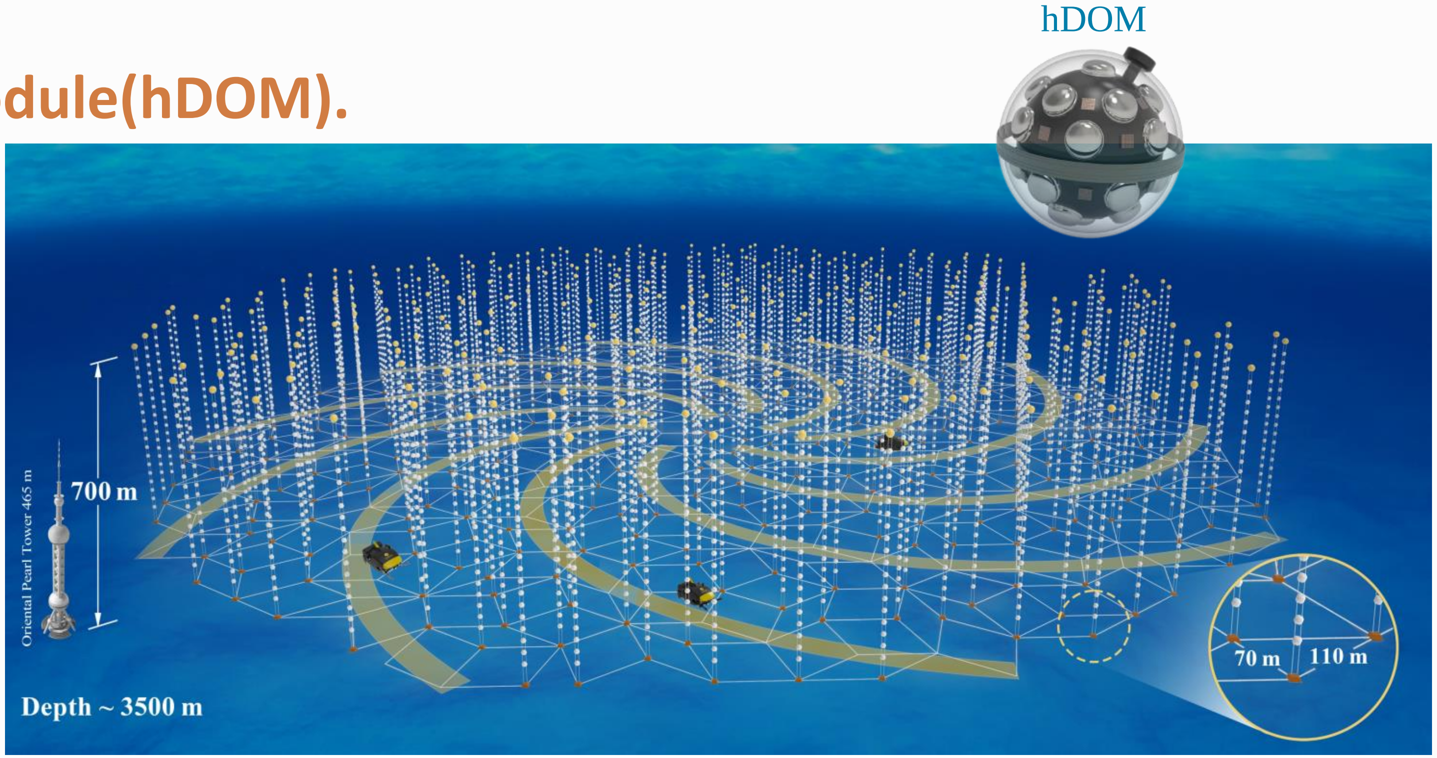
TRIDENT Experiment

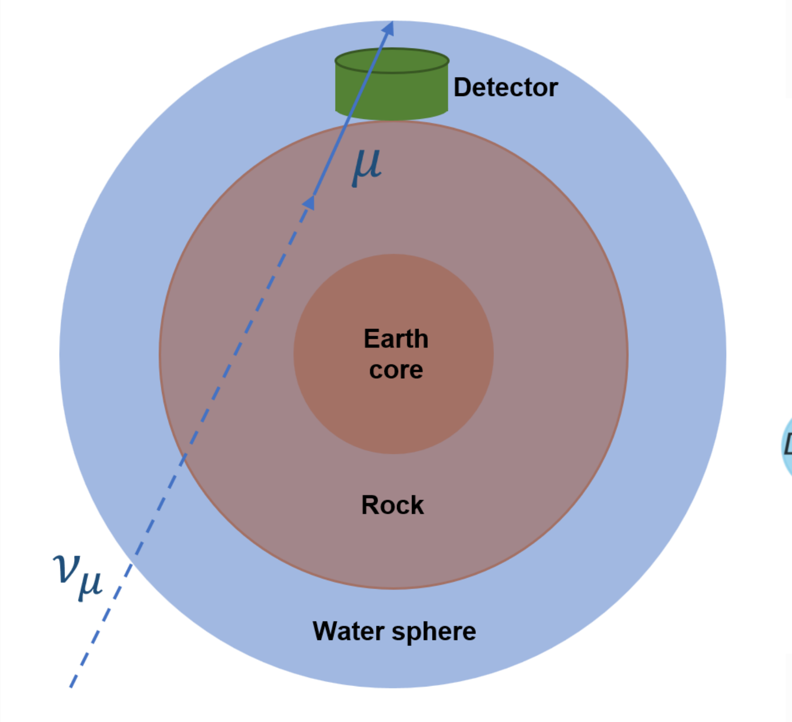
- **TRIDENT: TRoplcal DEep-sea Neutrino Telescope.**

  A multi-cubic-kilometre neutrino telescope in the western Pacific Ocean. *Nat Astron* (2023).

- **To be located in the South China Sea.**
- **Penrose tiling structure with 2000m radius, 700m height (8.7 km³). 3500m deep under sea level.**
- **24220 hybrid Digital Optical Module(hDOM).**

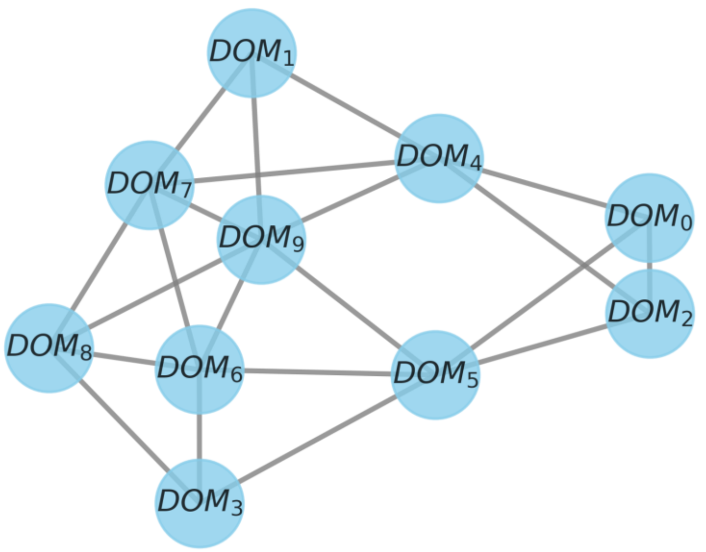


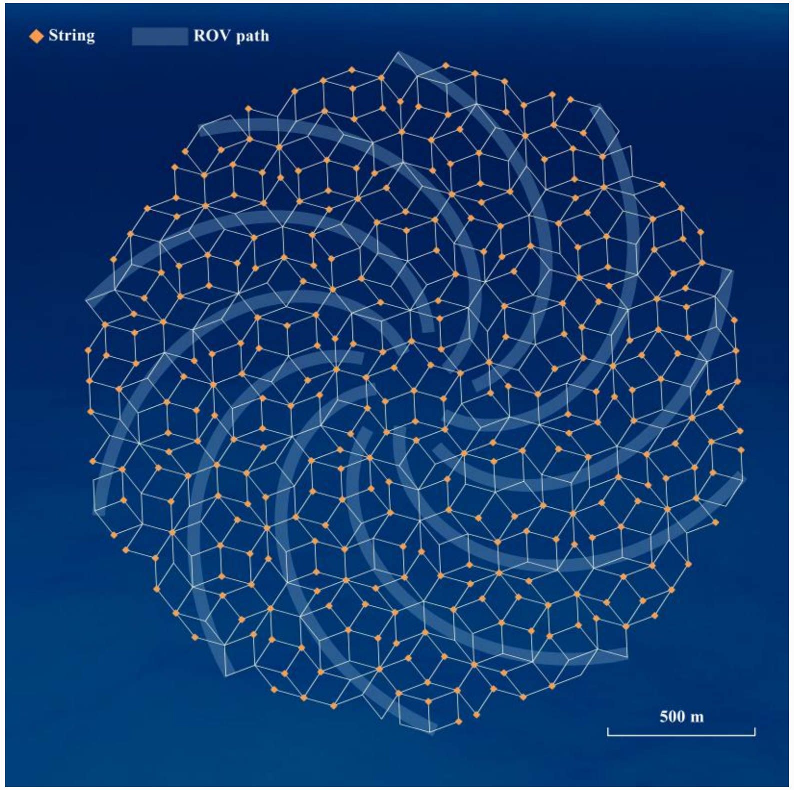




Preliminary earth model

Top view of TRIDENT





Neutrino event generator Based on CORSIKA8

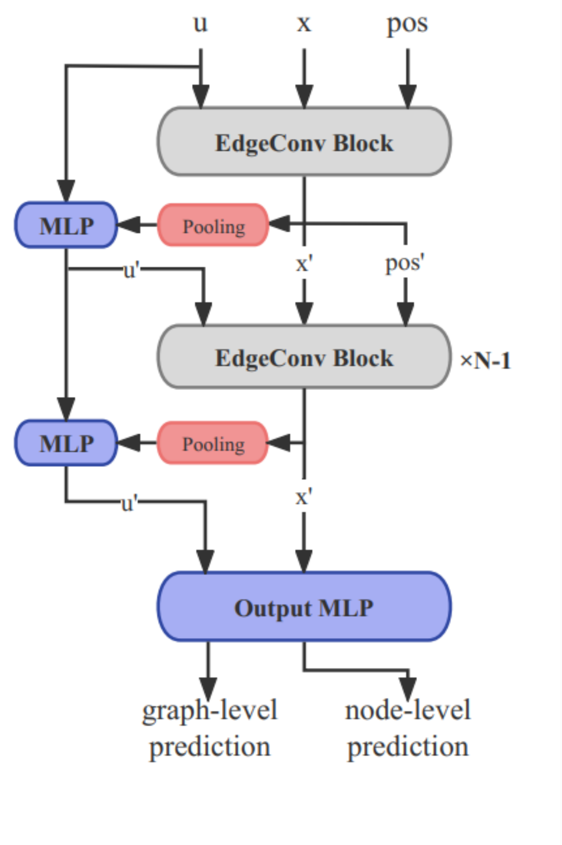
Detector simulation based on Geant4

Use point cloud to represent neutrino events:

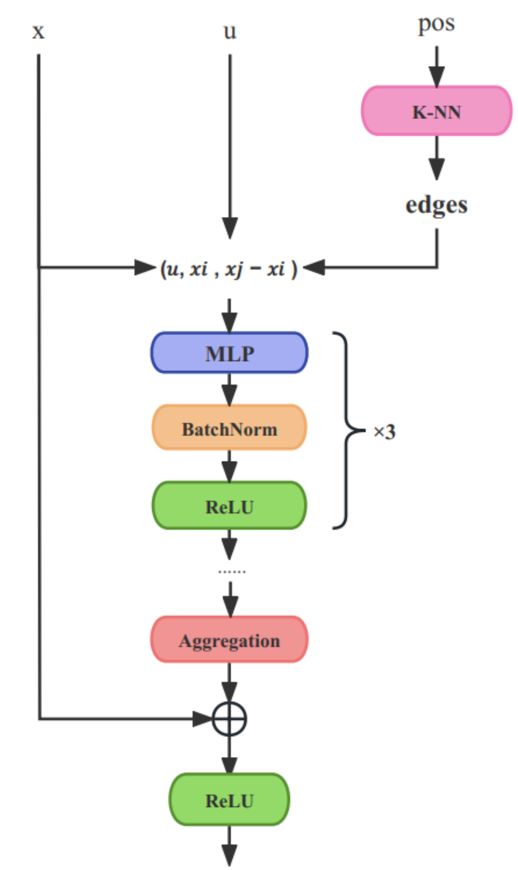
- Triggered DOMs → Nodes of point cloud
- Location of DOMs → Coordinate of nodes, pos_i.
- DOM-measured time → Features of nodes, x_i.

TRIDENT: Neutrino Reconstruction

- GNN is built based on EdgeConv block: modified block as in ParticleNet
- Both graph-level and node-level target can be predicted.



Network Structure



EdgeConv Block

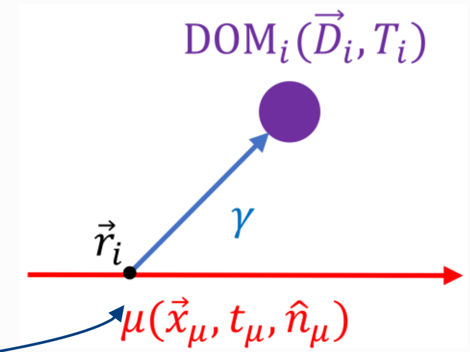
ν_μ Direction reconstruction

$train : validation : test = 900k : 70k : 100k$

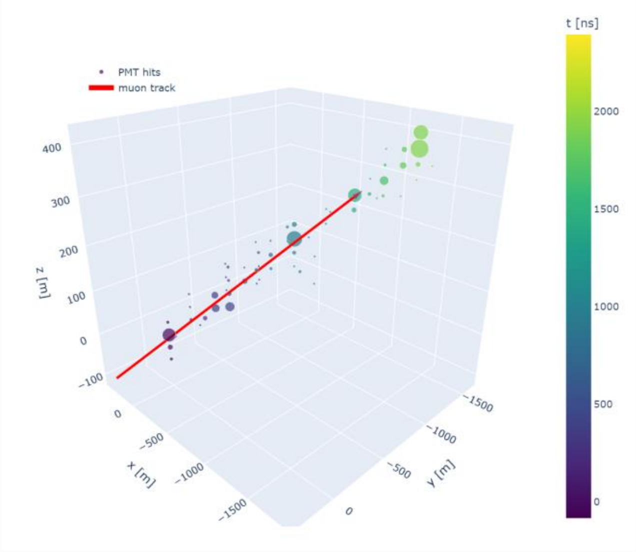
- **Input features**: location $\vec{D}_i$, first photon arrival time T_i and number of photo hits n_i.
- To make full use of the geometric feature of track-like events, the network is trained to predict $\vec{r}_i$ for each DOM_i.
- **Linear fit** on the predicted $\vec{r}'_i$ to reconstructs $\hat{n}_\mu$.
- **Loss function**: mean square error (MSE) with weight proportional to n_i:

$$Loss = \Sigma_i n_i \times \left|\overrightarrow{output}_i - \vec{r}_i\right|^2 / \Sigma_i n_i$$

- **Hybrid-GNN models: LITE, LARGE**

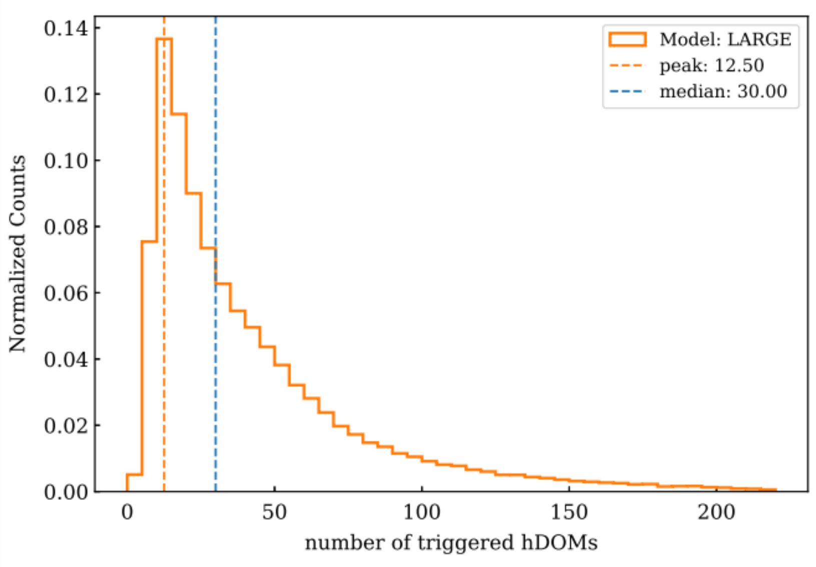


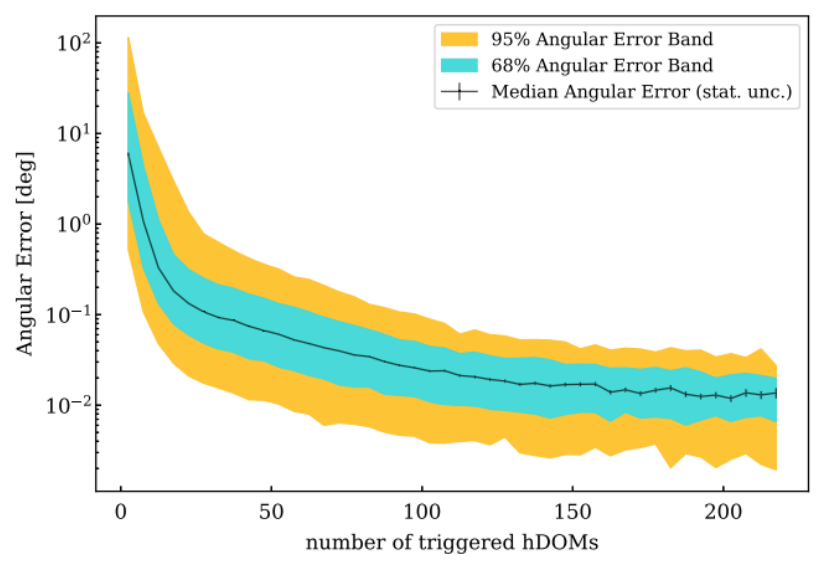
μ emits photon and triggers $\mathrm{DOM_i}$



Track-like event display

TRIDENT: Neutrino Reconstruction





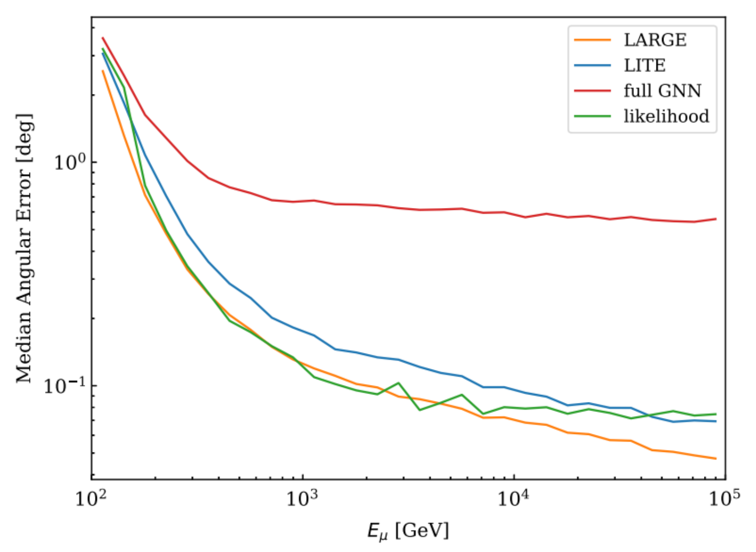


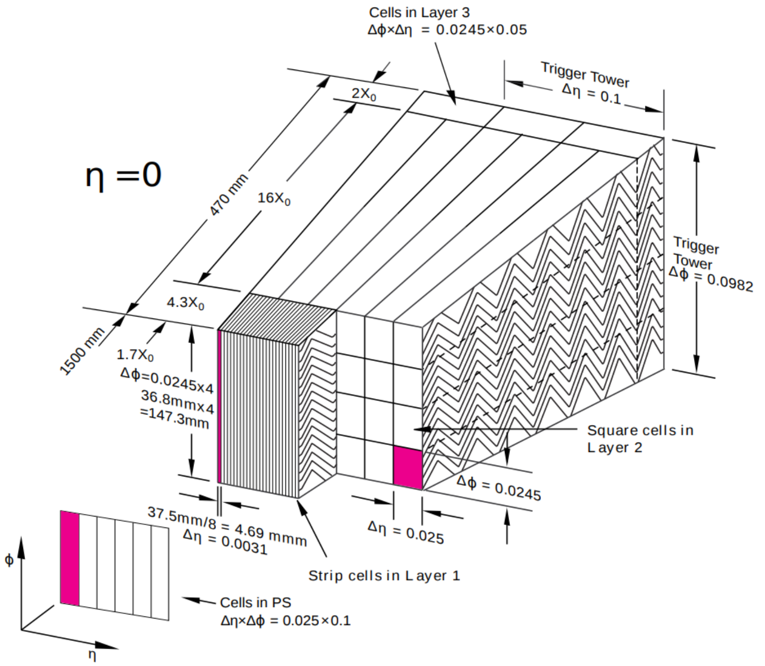
TABLE I. Mean run-time cost per inference.

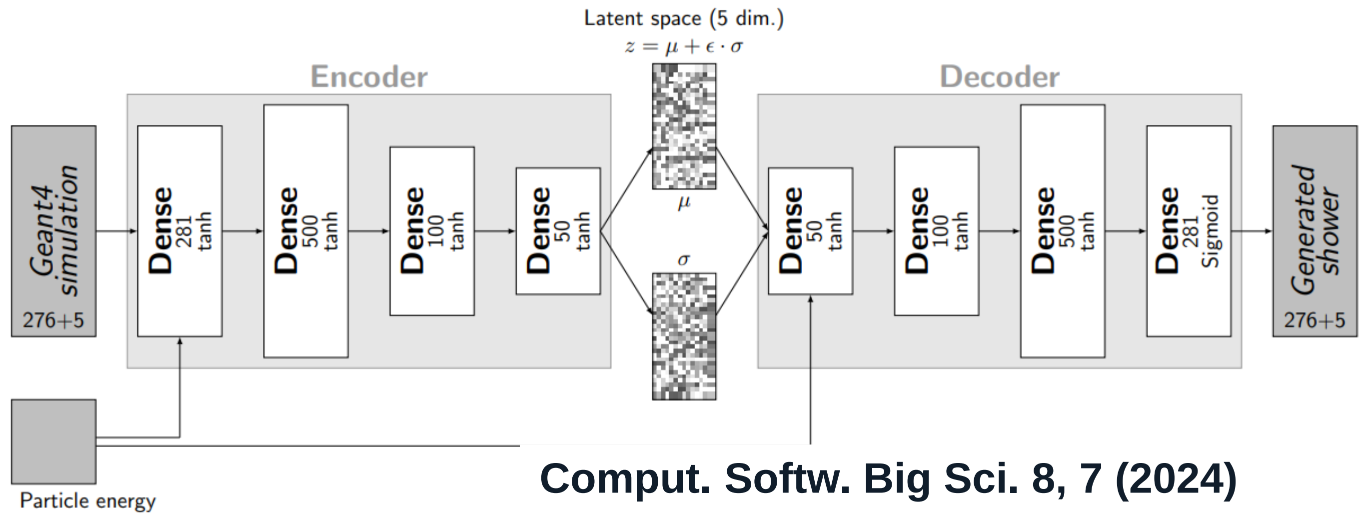
Method	Time (0.1–1 TeV) (ms)	Time (1–10 TeV) (ms)	Time (10–100 TeV) (ms)
Likelihood	1552.30	1259.86	919.14
GNN light (GPU)	0.19	0.21	0.29
GNN large (GPU)	0.38	0.78	2.37
GNN light (CPU)	5.05	12.53	30.44
GNN large (CPU)	54.71	152.48	181.80

Phys. Rev. D 112, 072012

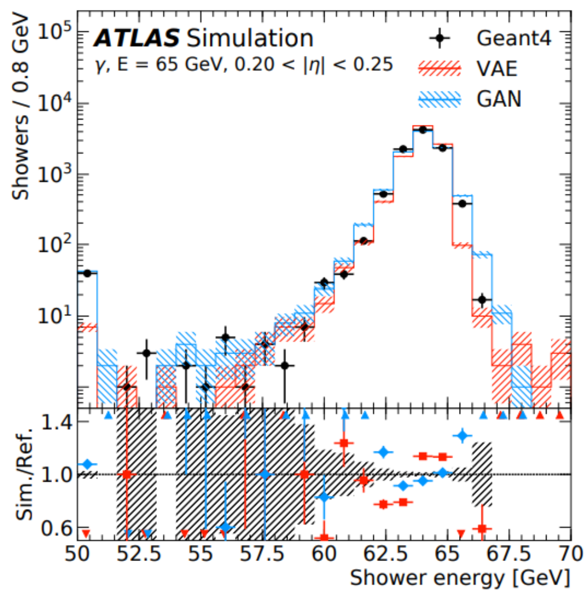
- **Median angular error decreases from 1 degree to 0.1 degree as the energy of ν_μ increases**
- **Light hybrid-GNN model (LITE) runs 0.19–0.29 ms per event on GPUs, 1000 times faster than traditional likelihood fitting method --- real time processing**
- **Large hybrid-GNN model (LARGE) takes longer but with more precision --- offline processing**

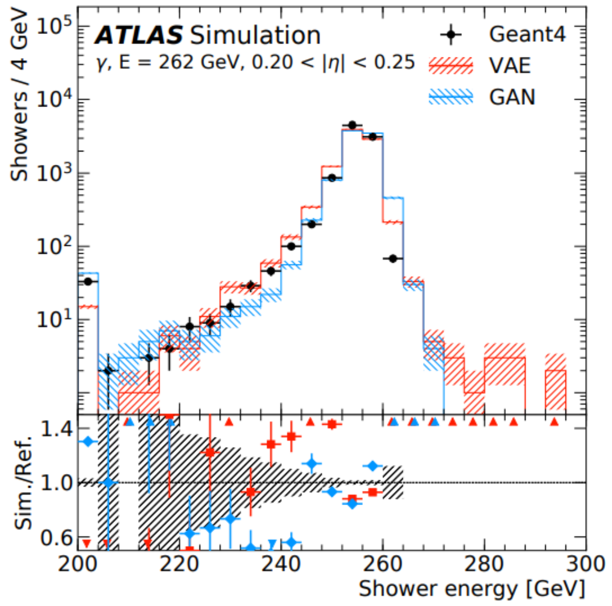
Faster Simulation: Photon Showers





Comput. Softw. Big Sci. 8, 7 (2024)





Use Variational autoencoders (VAE) and generative adversarial networks (GAN) to simulation ATLAS photon showers

- **VAE/GAN: x100 faster than GEANT4 full simulation**
- **Good agreement between GAN/VAE and Geant4 for EM showers of different energies**
- **GAN needs improvement in the longitudinal shower development**

Conclusion and Outlook

- ✓ The field of high energy physics is rapidly adopting ML and AI techniques leading to real impact on physics results
  - ✓ Diversified use case: classification with both supervised and weakly/self-supervised scheme, reconstruction, simulation and more
  - ✓ Trend: bigger and more sophisticated, more generalized
  - ✓ Foundation model + fine tuning seems to be the next direction
  - ✓ Human insight and expertise remain essential to success in the foreseeable future