Symmetry Preserving Attention Networks (SPA-Net) for Resolved Top & Higgs Reconstruction at the LHC

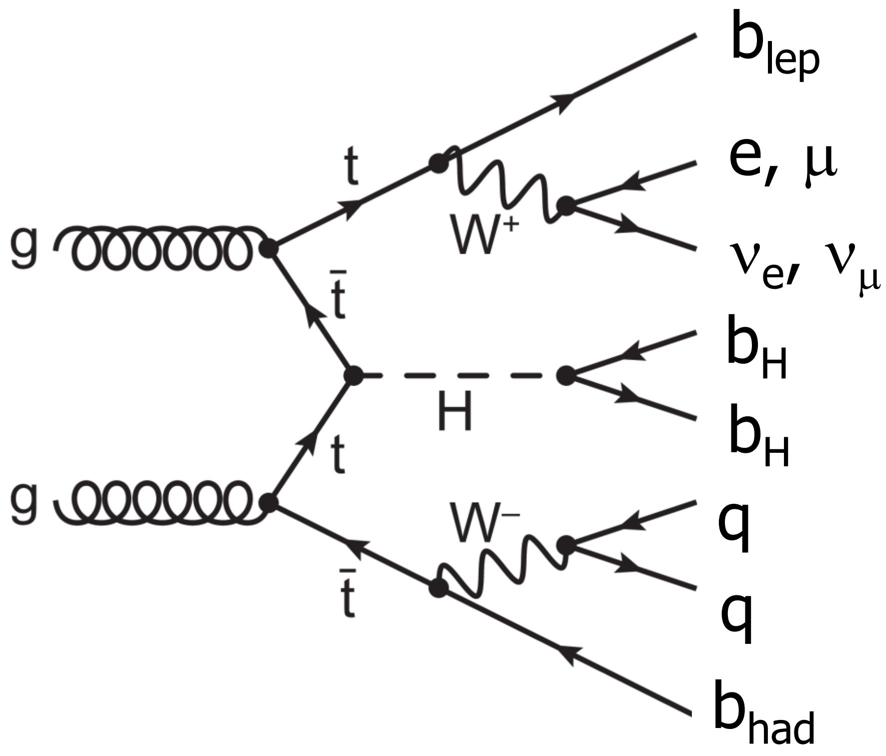
Quantum Computing & Machine Learning Workshop, August 11-14, 2023

Shandong University

Michael Fenton[1], Alexander Shmakov[1], **大川(Okawa)英希(Hideki)[2]**, Yuji Li[3], Ko-Yang Hsiao[4], Shih-Chieh Hsu[5], Daniel Whiteson[1], Pierre Baldi[1]

UC Irvine[1], **高能所(IHEP)[2]**, Fudan[3], National Tsing Hua[4], Washington[5]

Jet-Parton Matching for Top, etc.



- Jet-parton assignment (e.g. top, Higgs reconstruction) is a crucial component in $t\bar{t}$, $t\bar{t}H$, 4top & other high-multiplicity analyses.
- Standard algorithms compare all possible permutations of jets per event & systematics
  - Combinatoric diverges with jet multiplicity.

Unsorted list of jets
$j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8$

↔

Target partons
b_{lep}, b_H, b_H, ∅, b_{had}, ∅, q, q

Attention Transformers

- **Attention mechanisms are superceding RNNs & LSTMs in neuro linguistic programming.**
  - Permutation invariant & can handle variable-length lists
- A paradigm shifting impact as seen in Chat GPT, etc.
- **Why not use it in particle physics? → Yes, we already do!**
  - **SPA-Net (this & previous works)**
  - Particle Transformer: Huilin Qu, Congqiao Li, Sitian Qian, arXiv:2202.03772
    → Application to ATLAS & CEPC (王书栋's talks)
  - Attention-Based Cloud Network (ABCNet): Lukas Gouskos, **Fabio Iemmi (IHEP),** Sascha Liechti, Benedikt Maier, Vinicius Mikuni , Huilin Qu, arXiv:2211.02029
  - etc.

Symmetric Tensor Attentions

- **Symmetric Tensor Attentions:** generalization of attention to encode symmetries ($t \leftrightarrow \bar{t}$, $b \leftrightarrow \bar{b}$ in H, $q \leftrightarrow \overline{q'}$ in W)
- Natural permutation invariance from attention: no arbitrary p_T-ordering
- Symmetry from particle decays encoded: e.g. Two-body decay symmetries (W→$q\overline{q'}$, H→$b\overline{b'}$)

Symmetryic Weight Tensor (for t→bqq case;
Θ: learnable weights)

$$S^{i_1 i_2 i_3} = \Theta^{i_1 i_2 i_3} + \Theta^{i_2 i_1 i_3}$$

Attention Weights (X: list of particle vectors)

$$O^{j_1 j_2 j_3} = X^{j_1}_{i_1} X^{j_2}_{i_2} X^{j_3}_{i_3} S^{i_1 i_2 i_3}$$

Joint distributions for particles

$$\mathcal{P}^{j_1 j_2 j_3} = \frac{\exp\left(O^{j_1 j_2 j_3}\right)}{\sum_{j_1, j_2, j_3} \exp\left(O^{j_1 j_2 j_3}\right)}$$

Training

- One output per final-state particle & can embed symmetries in the loss function if needed (e.g. $t \leftrightarrow \bar{t}$ in all hadronic final state)

$$\mathcal{L}_{min} = \min_{\sigma \in G_E} \sum_{i=1}^{m} CE(\mathcal{P}_i, \mathcal{T}_{\sigma(i)})$$

CE: cross entropy
T: δ-distributions containing one possible valid jet assignment
σ: symmetries b/w particles

- **Partial Event Training**
  - Jets are often lost due to the detector acceptance. → e.g. As much as 65% in $t\bar{t}$H semi-lep. events cannot be fully reconstructed.
  - Instead of fully discarding such events, we keep particles in the training if they are reconstructable. **Highly efficient usage of training data statistics!**

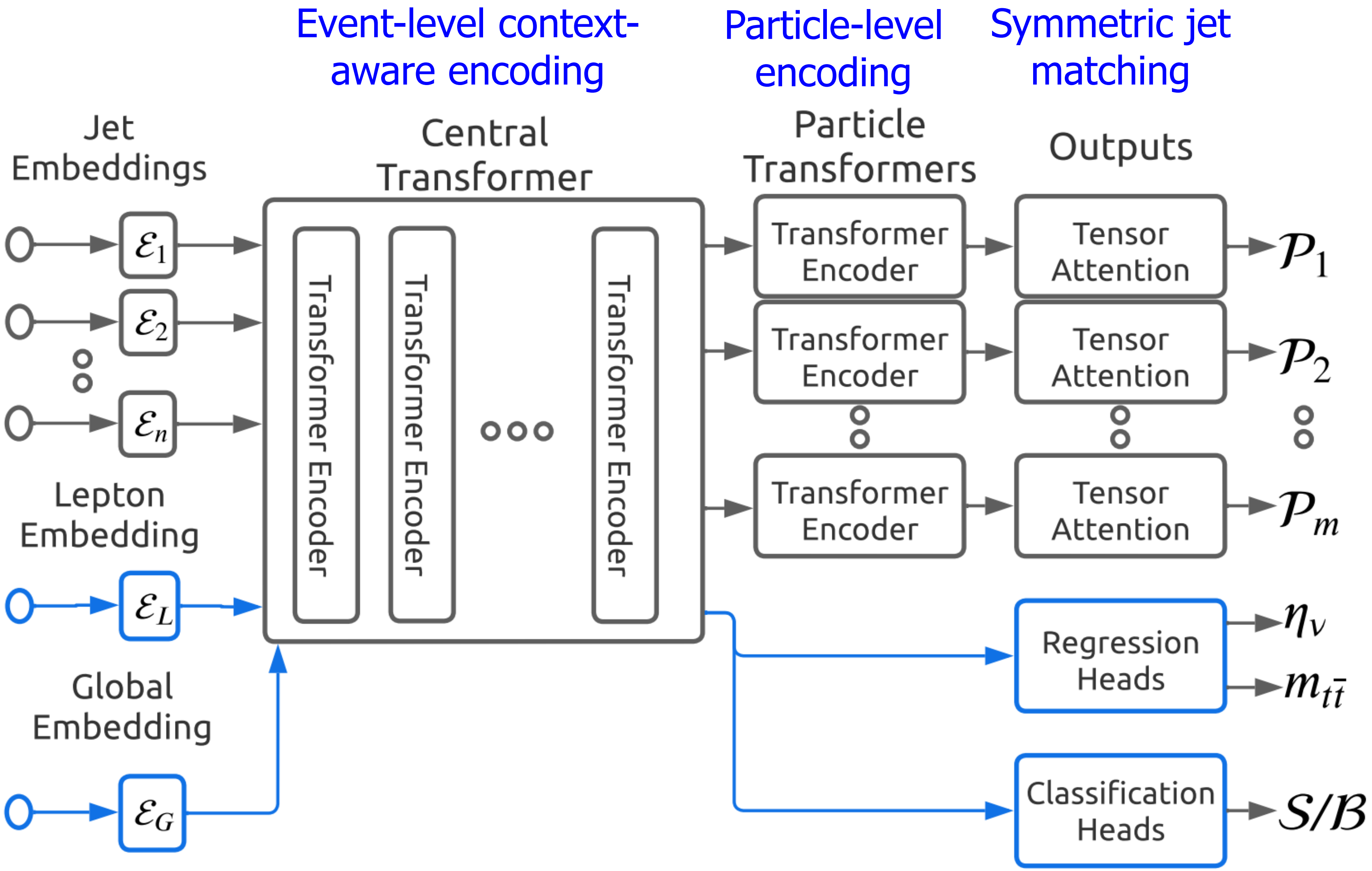
$$\mathcal{L}_{min}^{masked} = \min_{\sigma \in G_E} \left(\sum_{i=1}^{m} \frac{\mathcal{M}_{\sigma(i)} CE(\mathcal{P}_i, \mathcal{T}_{\sigma(i)})}{CB\left(\mathcal{M}_{\sigma(1)}, \mathcal{M}_{\sigma(2)}, \ldots, \mathcal{M}_{\sigma(m)}\right)} \right)$$

M: mask term
CB: Normalization factor to balance the class of particle presence

Symmetry Preserving Attention Networks (SPA-Net)

New Version!

Unordered list of object four-momenta + additional info (e.g. btag) & event-level variables (e.g. MET)



Full jet/doublet/triplet assignment distributions for every particle targe (e.g. t, H)

Dataset & Selection

- Generated MadGraph 5 interfaced with Pythia8 for showering & hadronization
- Detector response with Delphes v3.4.2 using the CMS card
- Top mass = 173 GeV

Object selection:

- Object overlap removal done
- Electron, muon p_T>25 GeV, $|\eta|$<2.5.
- Jet p_T>25 GeV, $|\eta|$<2.5 (dR matching considerd for truth jet-parton assignment)

Preselection: =1-lepton, ≥4 jets & ≥2 b-jets (for both $t\bar{t}$ & $t\bar{t}H$)

Baseline: Existing Methods

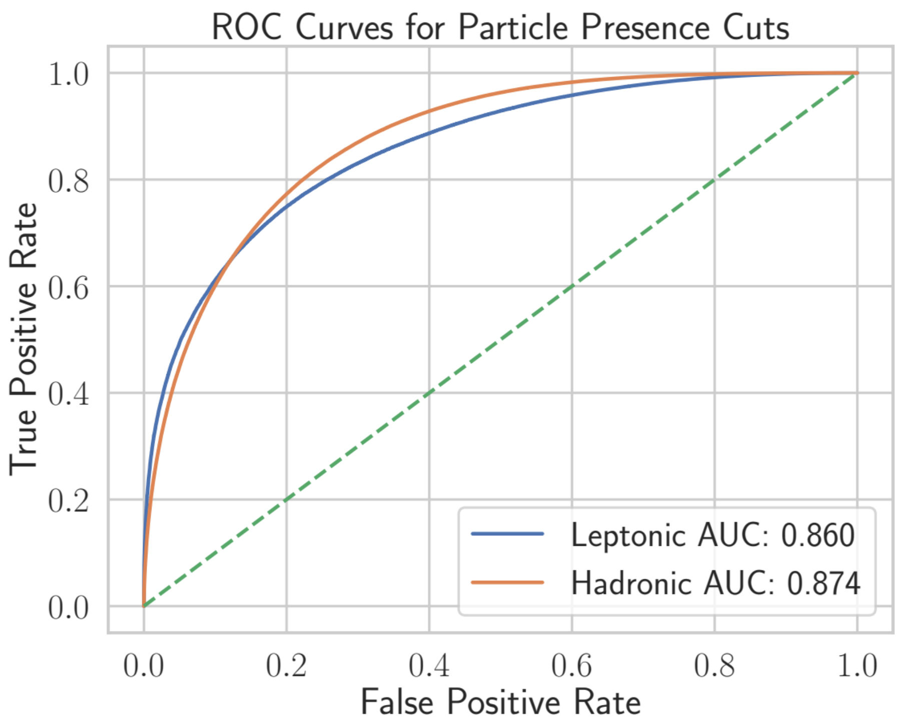
1. (χ^2 minimization: The simplest approach, considered in the previous all-hadronic studies; not considered in this talk)

2. KLFitter: likelihood-based kinematic fitting, assuming or not assuming a specific top mass (in this talk, the former); *J. Erdmann et al., NIM A 748 (2014) 18*

3. Permutation DNN: DNN considering all possible permutations; *J. Erdmann et al., JINST 14 (2019) P11015*

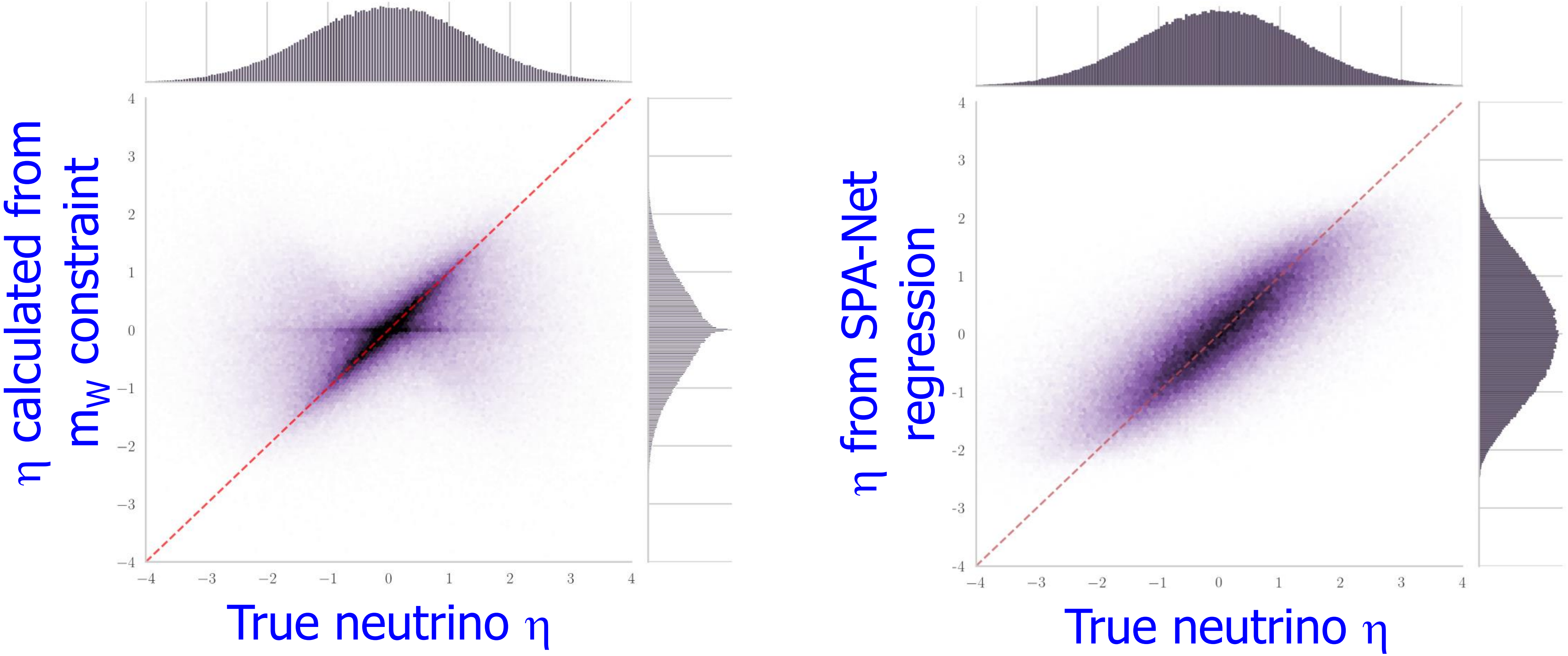
Reconstruction in semi-leptonic $t\bar{t}$ & $t\bar{t}H$

- New version of SPA-Net can handle different types of physics objects (jets, leptons, etc.)
- Can add event-level variables (MET, MET ϕ, etc.)
- $t\bar{t}$
  - SPA-Net: **75.6%** full event reconstruction (85.5-59.8% vs NJets)
  - Permutation DNN: **64.9%** (80.3-48.8%)
  - KLFitter: **52.1%** (77.2-23.7%)
- $t\bar{t}H(\rightarrow b\bar{b})$
  - SPA-Net: **54.2%** in 6j, 42.6% in 7j
  - Permutation DNN: **48.8%** in 6j, 36.4% in 7j
  - KLFitter: **31.4%** in 6j, 17.7% in 7j



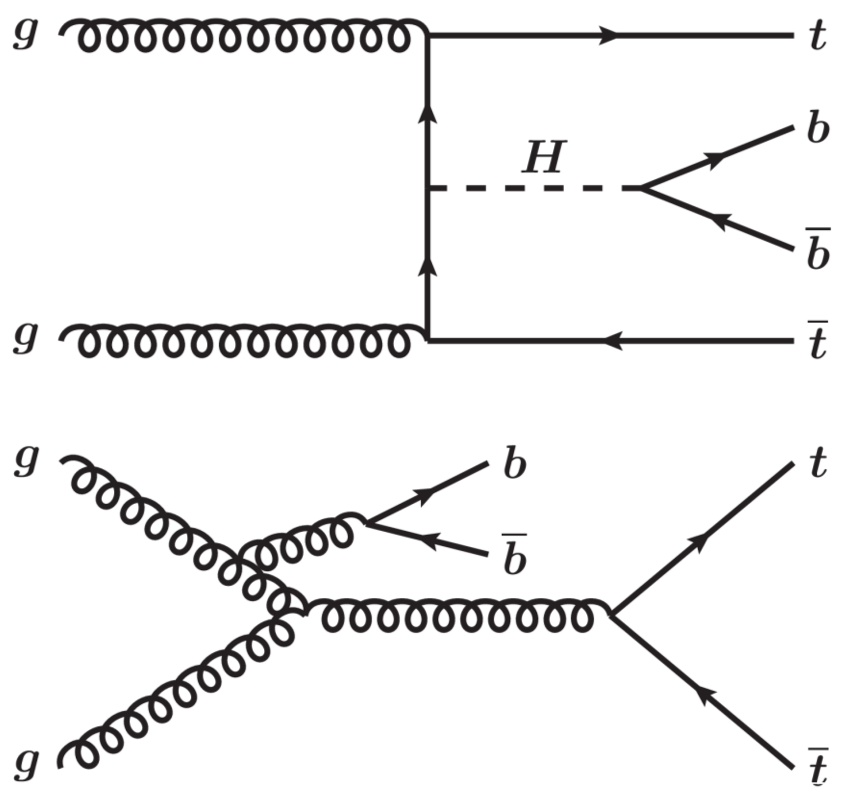
ROC from detection probability (explained later)
X: false ID in unreconstructable events
Y: efficiency in reconstructable events

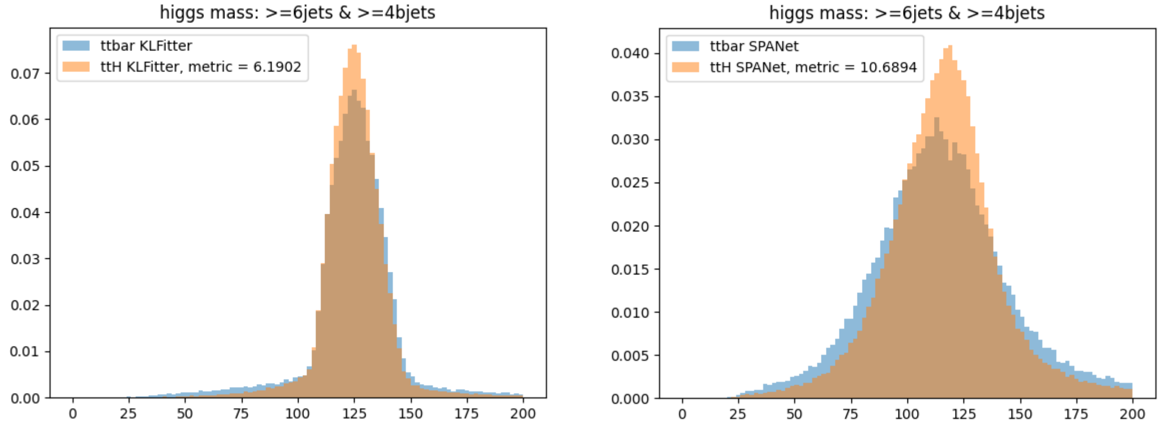
New Features: Regression of Kinematics



- SPA-Net can now train to reconstruct $t\bar{t}$ & simultaneously train to regress a variable (e.g. neutrino η or p_z, $t\bar{t}$ invariant mass)
- Neutrino η is more diagonal than the traditional method w/ improved RMS (1.39→0.9).

$t\bar{t}$H($\rightarrow b\bar{b}$) semi-leptonic





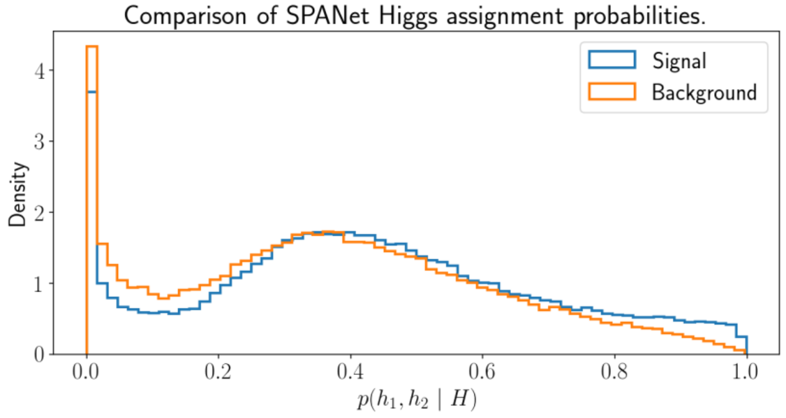
- $t\bar{t}b\bar{b}$ background is rather large with very similar kinematics to $t\bar{t}$H.
- t & H kinematics are main inputs to the BDT.
- However, the fraction of reconstructable events is only 35% in $t\bar{t}$H semi-lep. events.

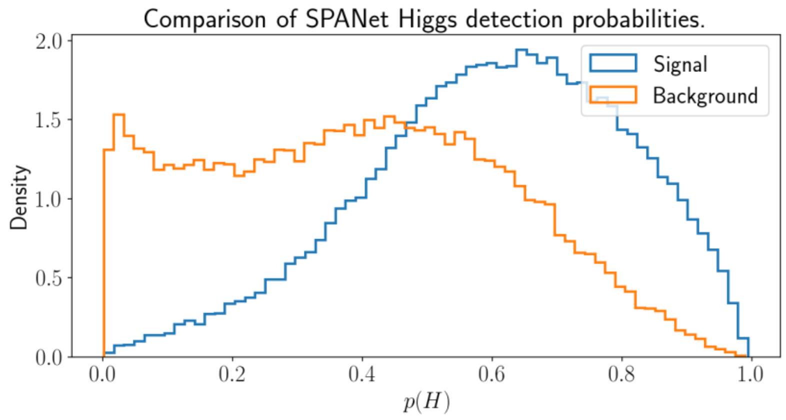
- **"Goodness" of the jet-parton assignment is also important to remove unreconstructable events.** → i.e. likelihood for KLFitter, a score for permutation DNN

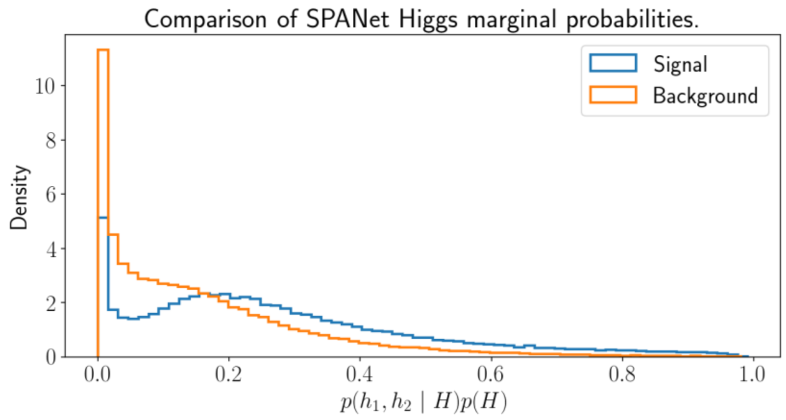
Updates: SPA-Net Probabilities & Entropies

- Defined for each particle to reconstruct (t, H). **9 probabilities & 3 entropies in total for ttH**. **SPA-Net can provide detailed information on the goodness of jet-parton assignment.**

- **Detection probability:** Is t or H reconstructable?
- **Assignment probability:** Given t or H is present, are the predicted jets correct?
- **(Pseudo-)marginal probability:** essentially the product of the two above
- **Entropy:** -Σ(P x log P) over singlets/doublets/triplets represented as matrix P=NxNx…N







$t\bar{t}$H($\rightarrow b\bar{b}$) semi-leptonic

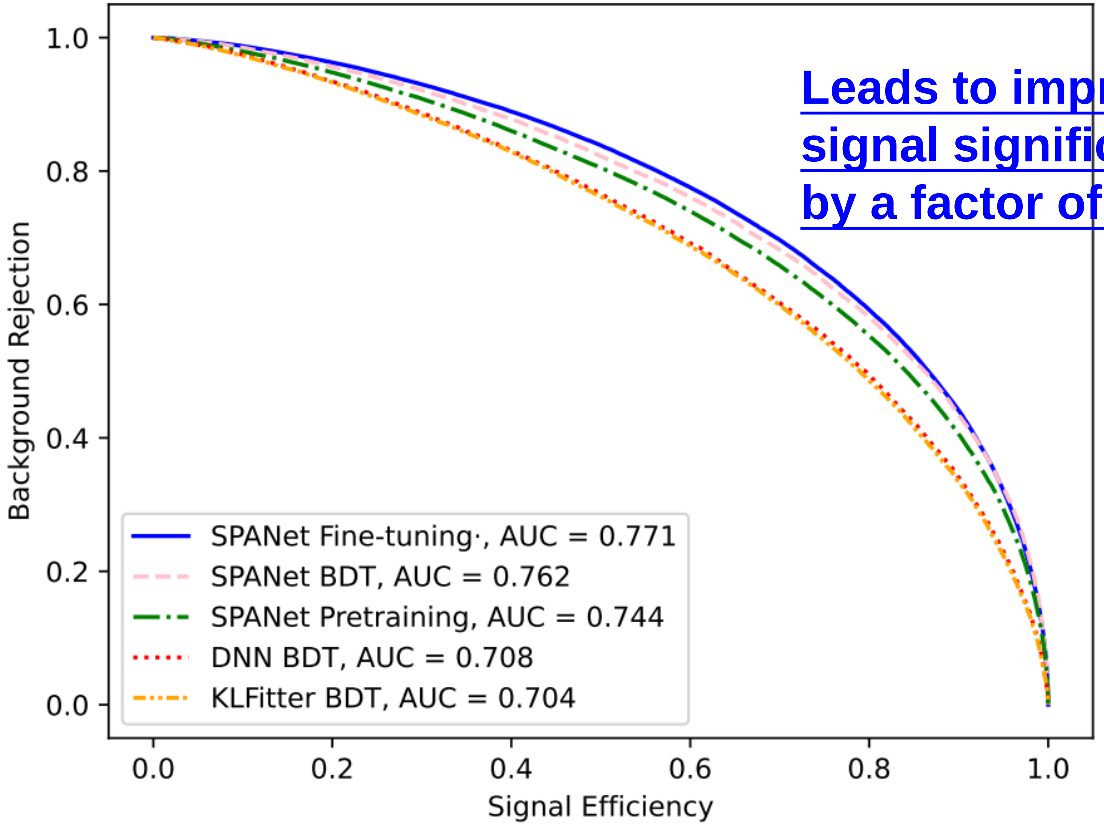
- Two analysis styles compared in this study:

1. **Traditional analysis style using BDT** using reconstructed kinematic variables **after the jet-parton assignment** (KLFitter, permutation DNN, SPA-Net).

2. **Pure SPA-Net-driven method**

New Features: Signal/BG Discrimination

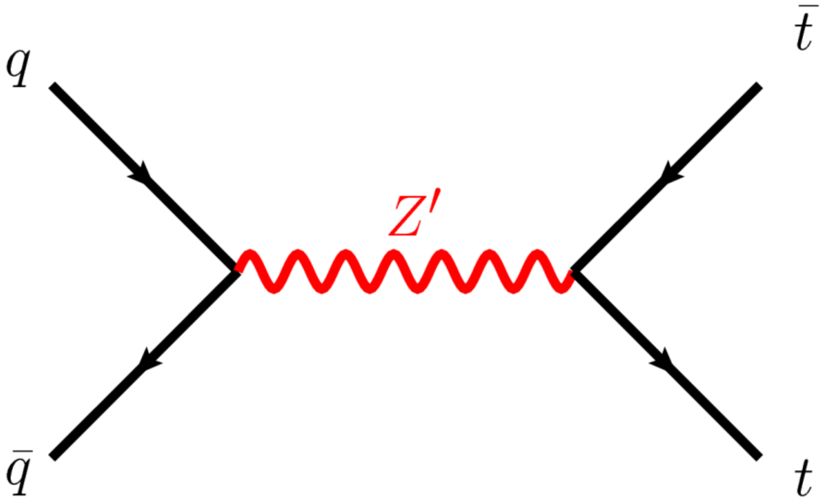
- SPA-Net outperforms existing methods! (**AUC=0.704, 0.708**)
- A new feature in SPA-Net can provide ttH/ttbb discrimination directly. → Best performance w/ fine-tuning! (**AUC=**0.744→**0.771**)
- SPA-Net jet-parton assignment and BDT w/ kinematics & probabilities also has excellent performance. (**AUC=0.762**)
  - Including particle probabilities in BDT is VERY important

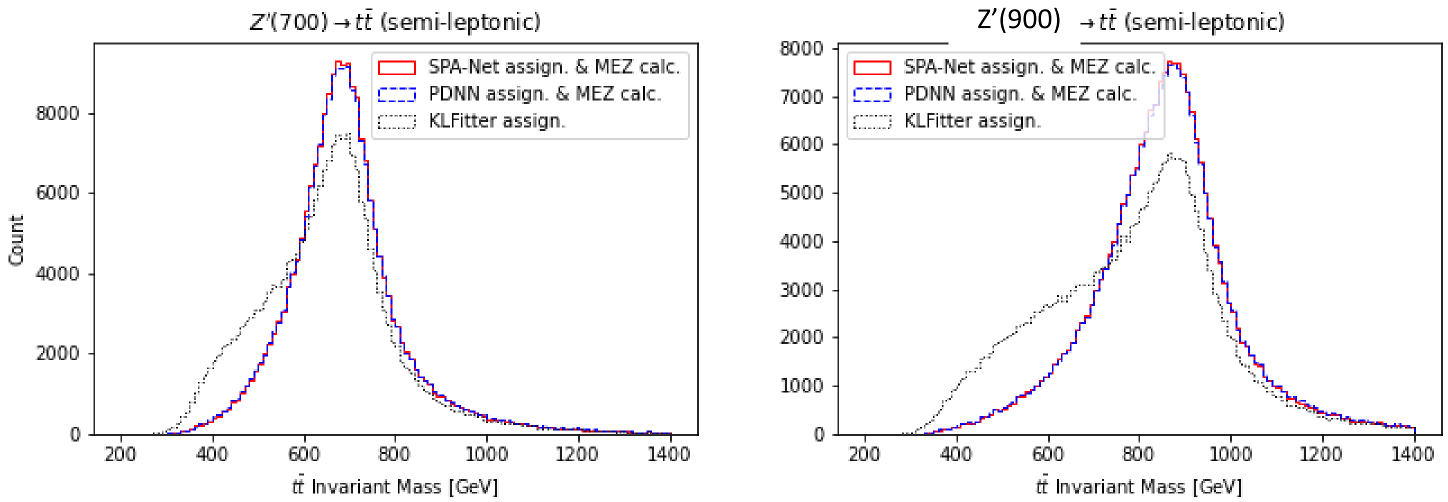


Leads to improved signal significance by a factor of ~1.4.

Transformer architecture provides us with meaningful embeddings for every jet, particle, and event: a big benefit over permutation-based models

Z’→$t\bar{t}$ Searches

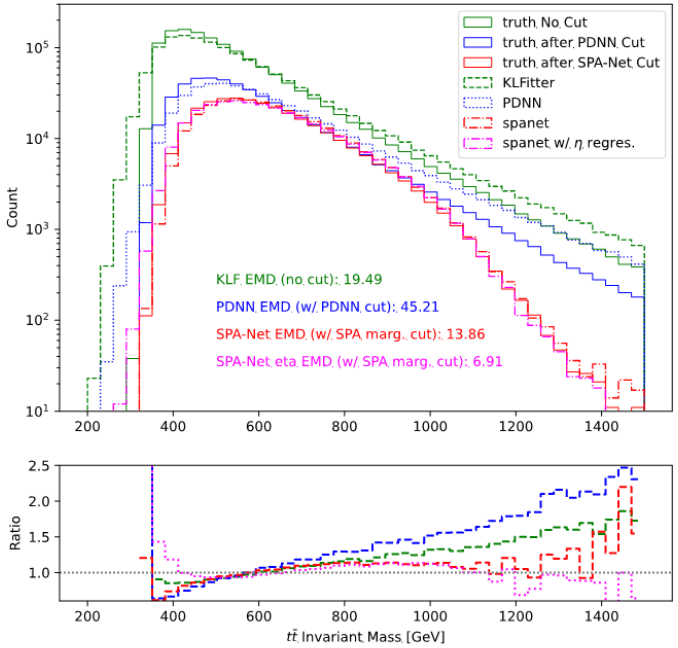


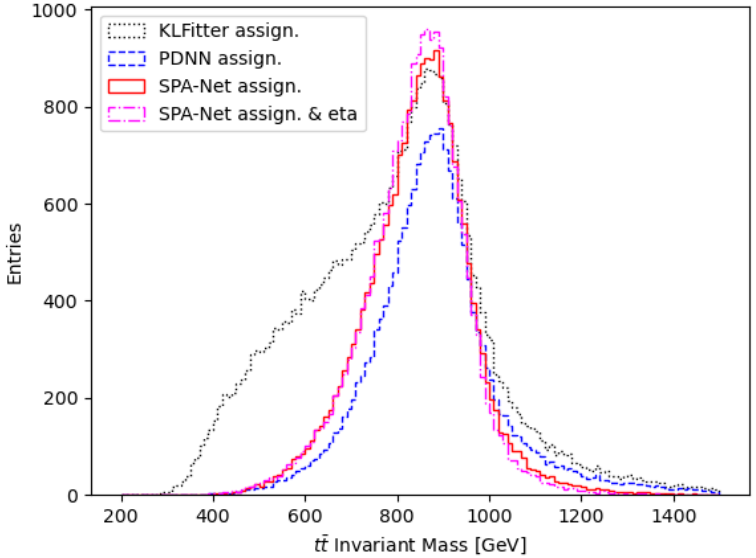


- Successful reconstruction of top quarks is crucial for the $t\bar{t}$ resonance searches (e.g. Z’, RS Graviton, Heavy Higgs A/H→$t\bar{t}$).
- Z’→$t\bar{t}$ searches are already dominated by systematic uncertainty. Adding more data would not help much & reduction of systematics and/or dramatic improvement in analysis strategies is necessary.
- SPA-Net provides significantly improved mass reconstruction from KLFitter.

Impact on Z’→$t\bar{t}$ Searches

- Also, **SPA-Net probabilities allow us to select signals while significantly suppressing the BG.**
- Evidence or discovery could happen even if not doable w/ traditional methods. (e.g. 1.9σ →4.3σ for Z’(900 GeV))

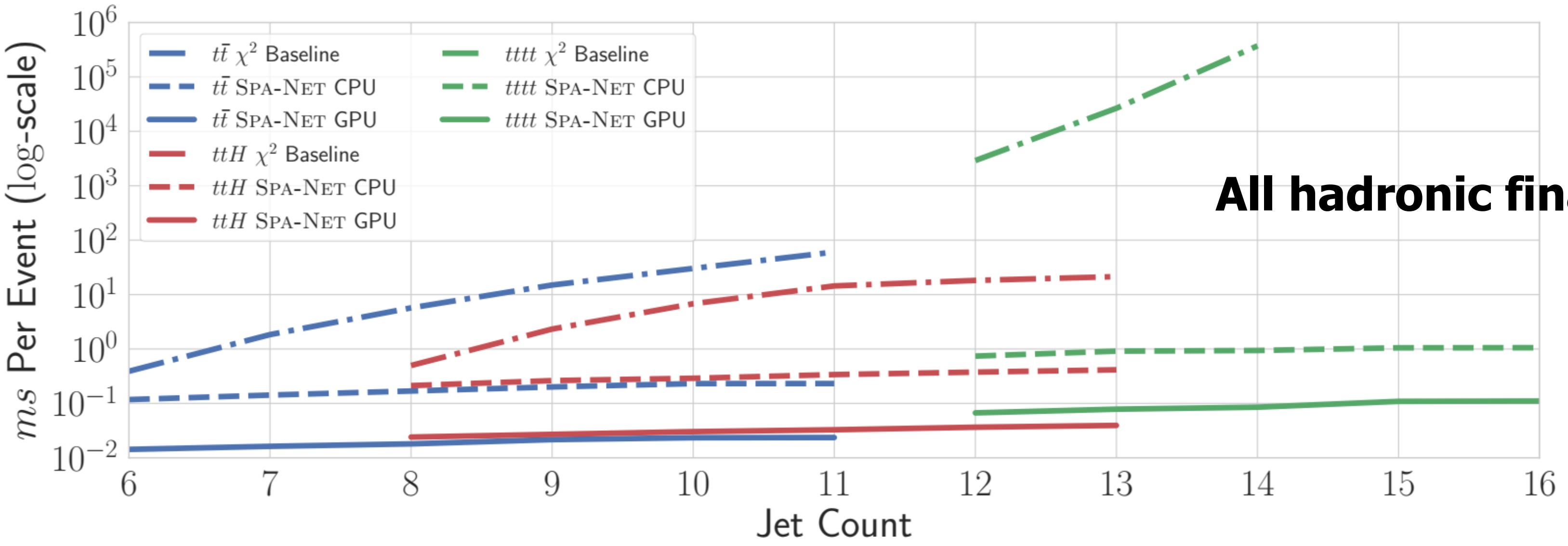




	KLFitter	PDNN	SPA-NET	SPA-NET w/ ν η
$m_{Z'} = 500$ GeV	1.24 σ	1.75 σ	2.75 σ	2.71 σ
$m_{Z'} = 700$ GeV	1.63 σ	2.45 σ	3.06 σ	2.87 σ
$m_{Z'} = 900$ GeV	1.94 σ	2.77 σ	4.30 σ	4.13 σ

CPU/GPU Time

A. Shmakov, M. Fenton et al., SciPost Phys. 12, 178 (2022)



- A few orders of magnitude improvement w/ SPA-Net compared to χ^2 or KLFitter. A further acceleration w/ GPU.
- One order of magnitude faster than permutation DNN for ttH (i.e. high multiplicity events)!

	ttH semilep	Z'→ttbar semilep
SPA-Net	3534 events/s [GPU] 852 events/s [CPU]	4407 events/s [GPU] 705 events/s [CPU]
Perm. DNN	101 events/s [GPU] 51.4 events/s [CPU]	3034 events/s [GPU] 2626 events/s [CPU]
KLFitter	1.95 events/s	24.4 events/s

Selling Points of SPA-Net over Perm. DNN

1. Better reconstruction efficiency for top quarks (or any other particle of interest)!
2. Provides detailed quality metrics (particle-level scores) to remove unreconstructable events (3 probabilities & entropy for each particle)
3. Can run regression/classification in parallel.
4. Less hyperparameter optimization needed
5. It's MUCH faster!

SPA-Net Package (new version!)

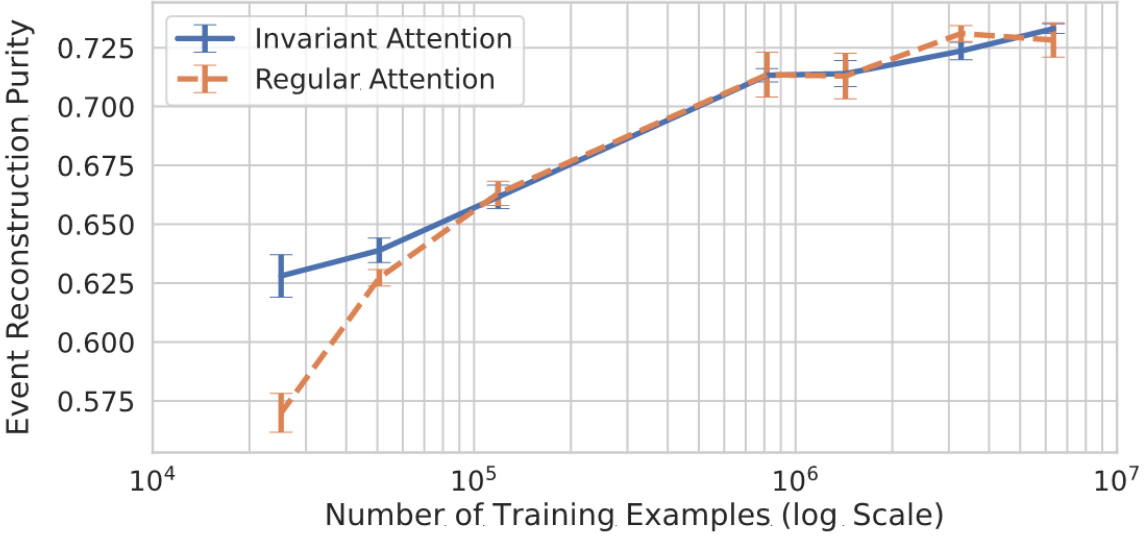
https://github.com/Alexanders101/SPANet

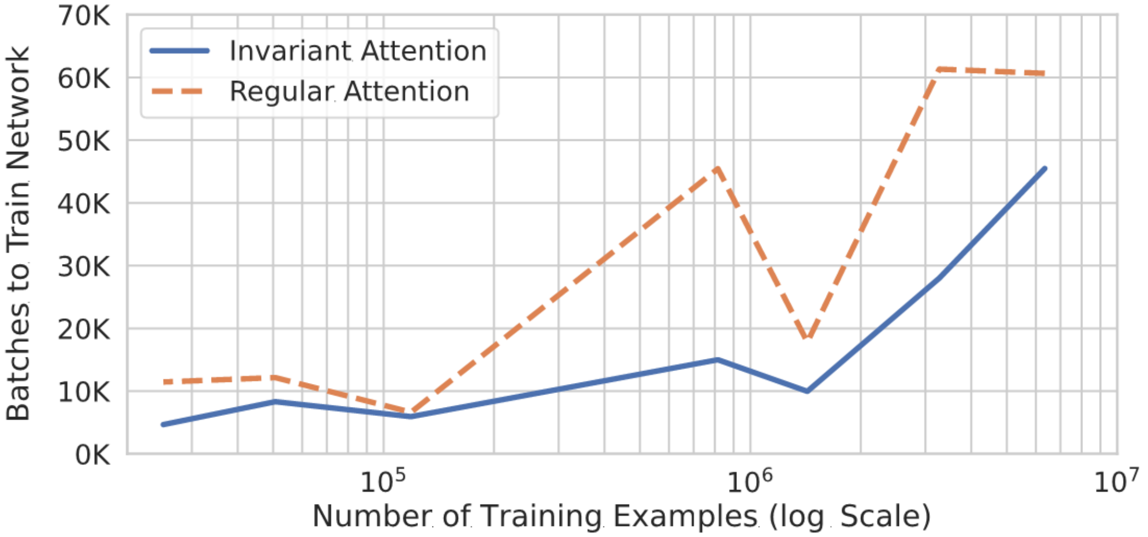
SPA-Net is not limited to top physics!

New features in v2:

1. New configuration file format with more options on inputs and event topology.
2. Allow for several different inputs, including global inputs (e.g. MET, MET ϕ) for additional context.
3. New Regression and Classification output heads for performing per-event or per-particle predictions.
4. Gated transformers and linear layers for more robust networks. Less hyperparameter optimization.

Considering Symmetries - Lorentz Invariance





- Adding Lorentz invariance to the network does not change the jet-parton assignment accuracy for most cases, but improves the performance for small datasets.

- Lorentz invariance brings visible improvement in speed: i.e. significant reduction of batches needed to train the network.

Investigations motivated by C. Li et al., arXiv:2208.07814

Summary

- SPA-Net provides efficient & excellent performance for event reconstruction in complex final states from multi-objects.
  - Superb CPU/GPU time, no limitation on object/jet multiplicity
  - Possible application to any jet-parton or even any “X”-”Y” assignment problem
- Transformer architecture provides us with meaningful embeddings for every jet, particle, and event: a big benefit over permutation-based models.
  - Reconstruction of missing components (e.g. neutrino η),
  - Direct signal/background discrimination,
  - Quality metrics to reject unreconstructable events.
- References:
  - M. Fenton, A. Shmakov et al., Phys. Rev. D 105, 112008 (2022)
  - A. Shmakov, M.Fenton et al., SciPost Phys. 12, 178 (2022)
  - Studies presented today, paper in preparation.