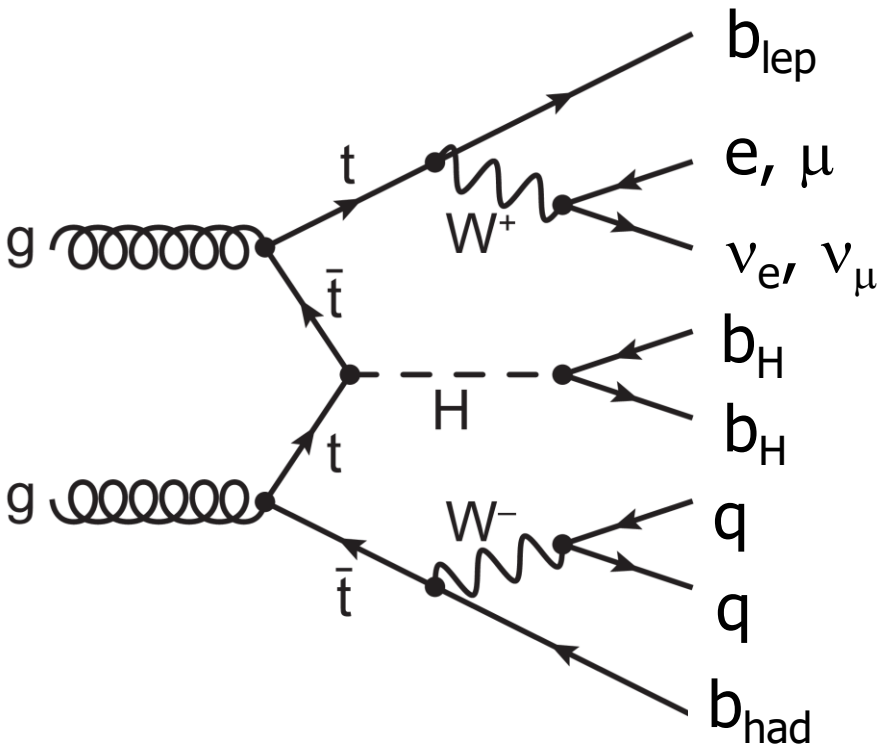
Top-Quark Reconstruction using Symmetry Preserving Attention Networks

IHEP ML Innovation Group Meeting, February 2, 2023

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

UC Irvine[1], IHEP[2], Fudan[3], National Tsing Hua[4], Washington[5]

Jet-Parton Matching in Top Analyses



- Jet-parton assignment (i.e. top reconstruction) is a crucial component in top-quark 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

$b_{lep}, b_H, b_H, \varnothing, b_{had}, \varnothing, q, q$

Attention for Top Reconstruction

- Attention mechanisms are superceding RNNs & LSTMs in neuro linguistic programming.
  - Permutation invariant & can handle variable-length lists
- **Tensor Attention: generalization of attention to encode symmetries** ($t\leftrightarrow\bar{t}$, $b\leftrightarrow\bar{b}$ in H, $q\leftrightarrow\bar{q'}$ in W)

e.g. Two-body decay symmetries (W→$q\bar{q'}$, H→$b\bar{b'}$)

$$S^{ijk} = \frac{1}{2}\left(\theta^{ijk} + \theta^{jik}\right)$$

$$O^{ijk} = X^i_n X^j_m X^k_l S^{nml}.$$

$i\leftrightarrow j$ symmetry

- Allow us to test every possible permutation in a single pass.

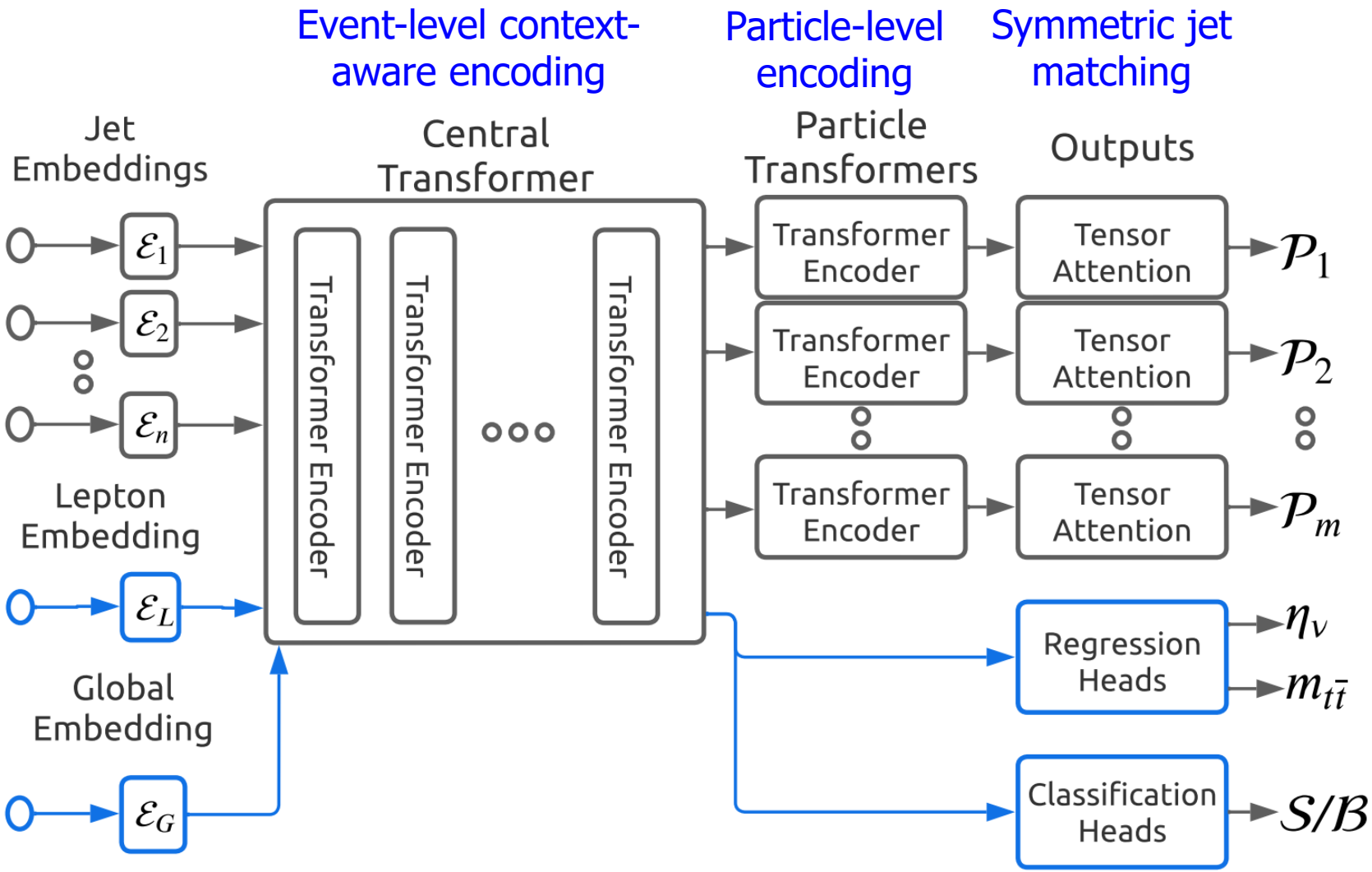
Previous Works & What's New

- Previous studies have concentrated on all-hadronic final states ($t\bar{t}$, $t\bar{t}H$, 4top):
  - M. Fenton, A. Shmakov et al., Phys. Rev. D 105, 112008 (2022)
  - A. Shmakov, M.Fenton et al., SciPost Phys. 12, 178 (2022)
- In our latest studies, the framework has been extended to be able to handle multiple physics objects and event-level features.
- Other new functionalities will also be presented in this talk.

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
- Top mass = 173 GeV

Object selection (This work considers semi-leptonic final states):

- 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)
- Object overlap removal done

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

Baseline: Existing Methods

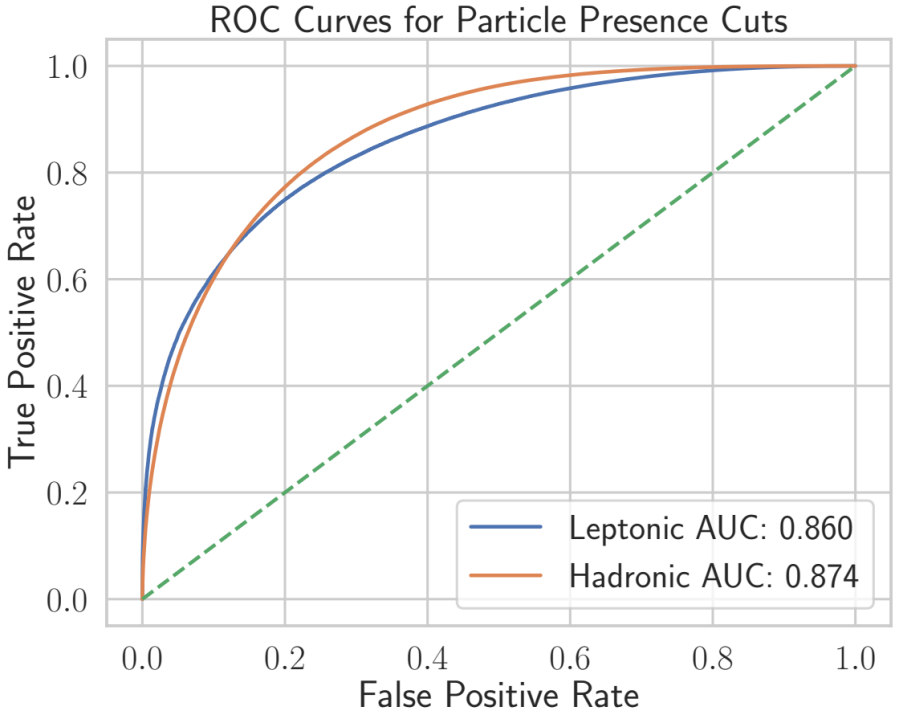
1. (χ^2 minimization: The simplest approach, considered in the previous all-hadronic studies; not considered in this talk)

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 $t\bar{t}$ & $t\bar{t}H$

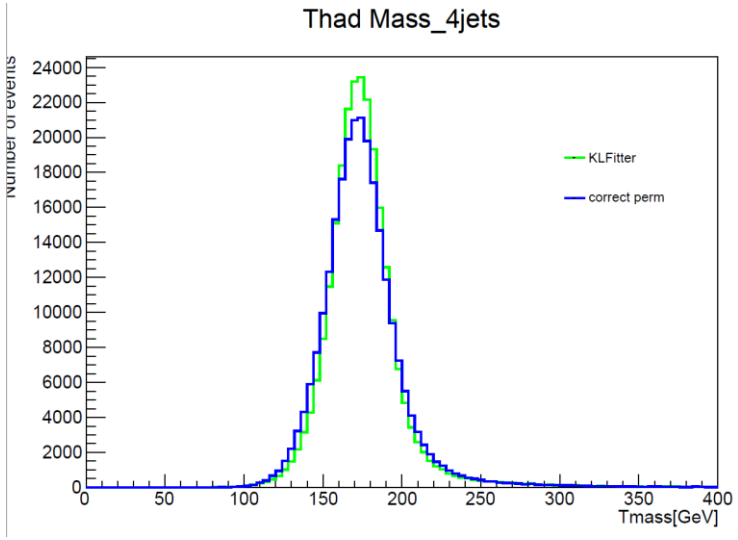
- New version of SPA-Net can handle different types of physics objects (jets, leptons, etc.)
- Can add event-level variables (MET, MET ϕ, etc.)
- $t\bar{t}$
  - SPA-Net: **75.6%** full event reconstruction (**85.5-59.8%** vs NJets)
  - Permutation DNN: **64.9% (80.3-48.8%)**
  - KLFitter: **52.1% (77.2-23.7%)**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

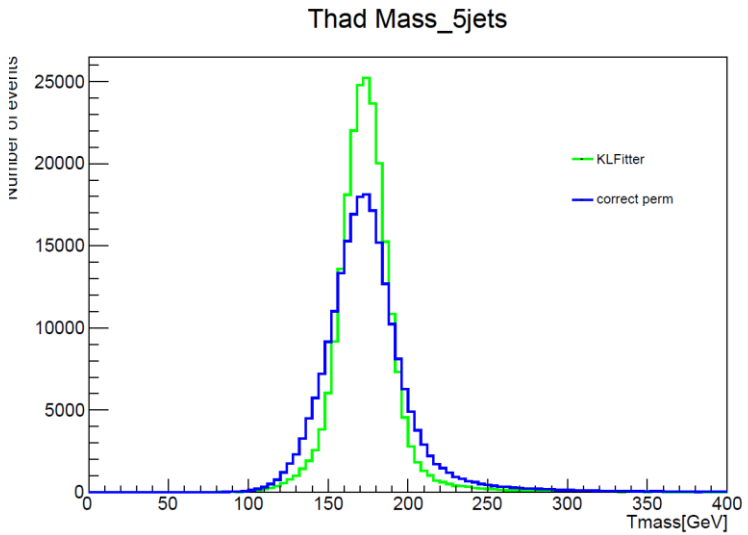


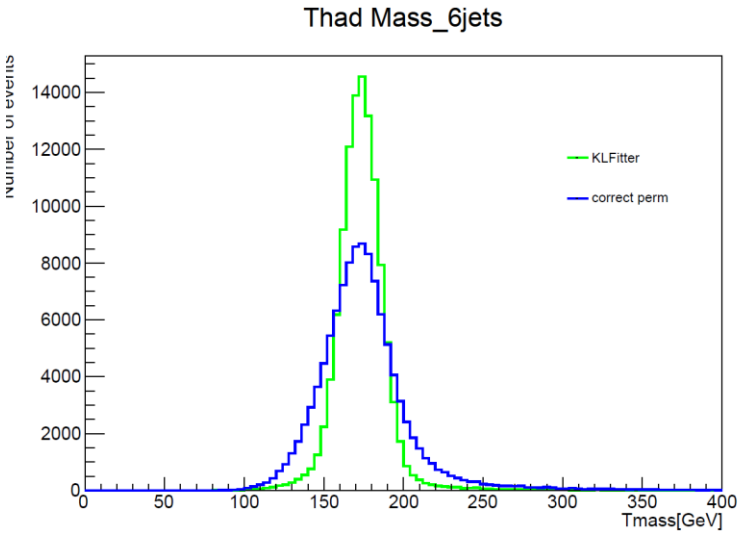
ROC from detection probability (explained later)
X: false ID in unreconstructable events
Y: Eff. in reconstructable events

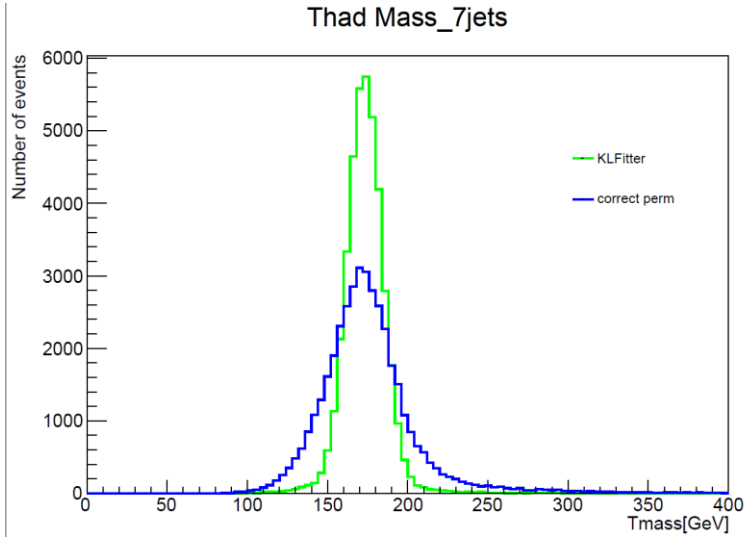
Reconstructed Hadronic Top Mass (KLFitter)

- KLFitter assumes top mass for jet assignment & obtains incorrectly narrow mass peak.
- This tendency is enhanced with the jet multiplicity.

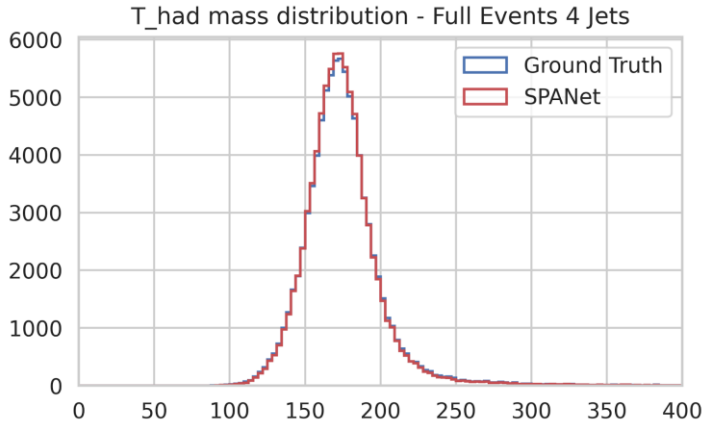


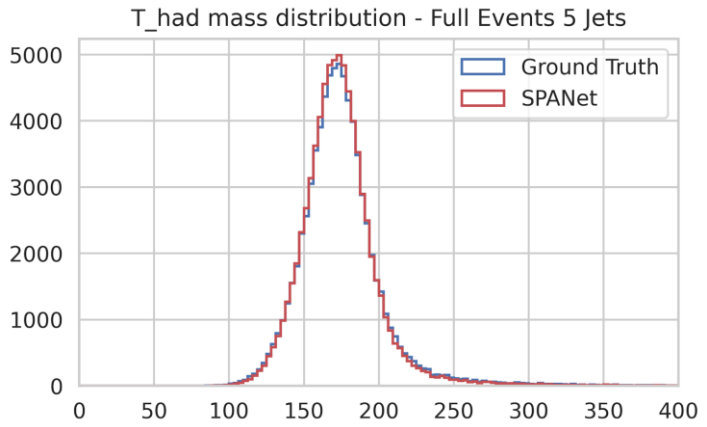


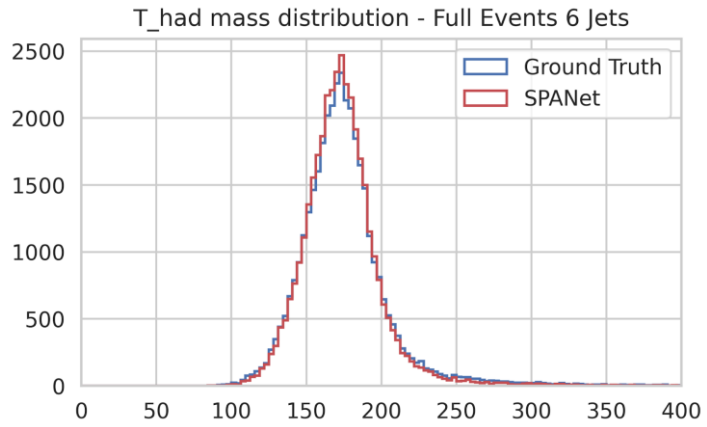


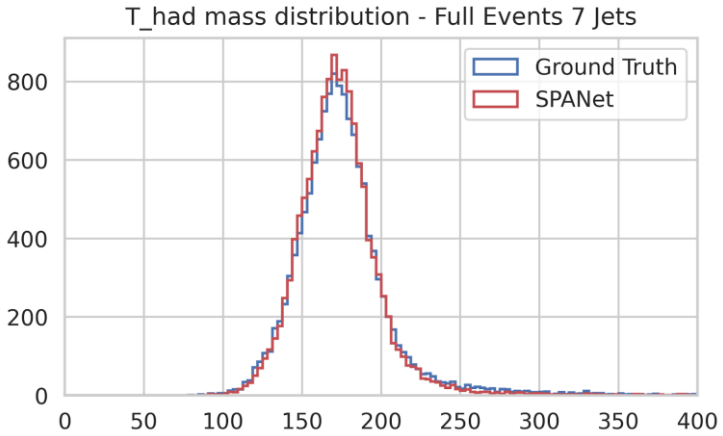


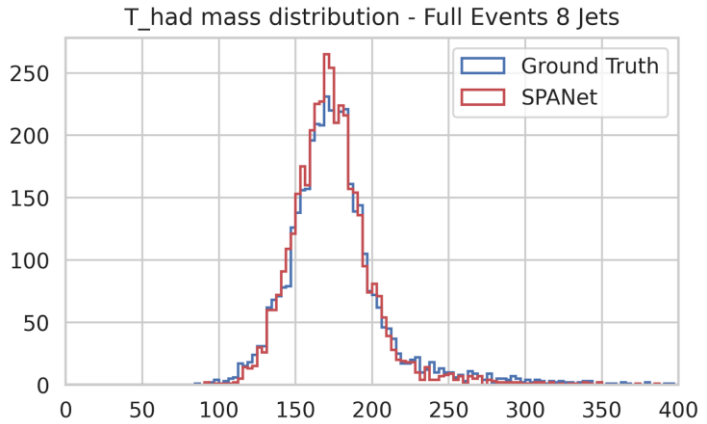
Reconstructed Hadronic Top Mass (SPA-Net)

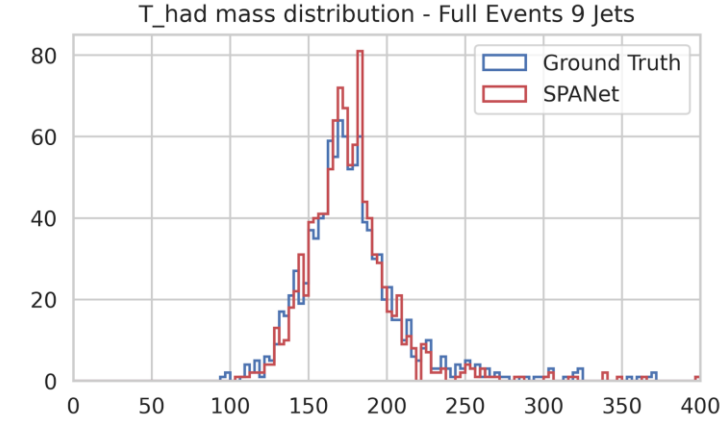






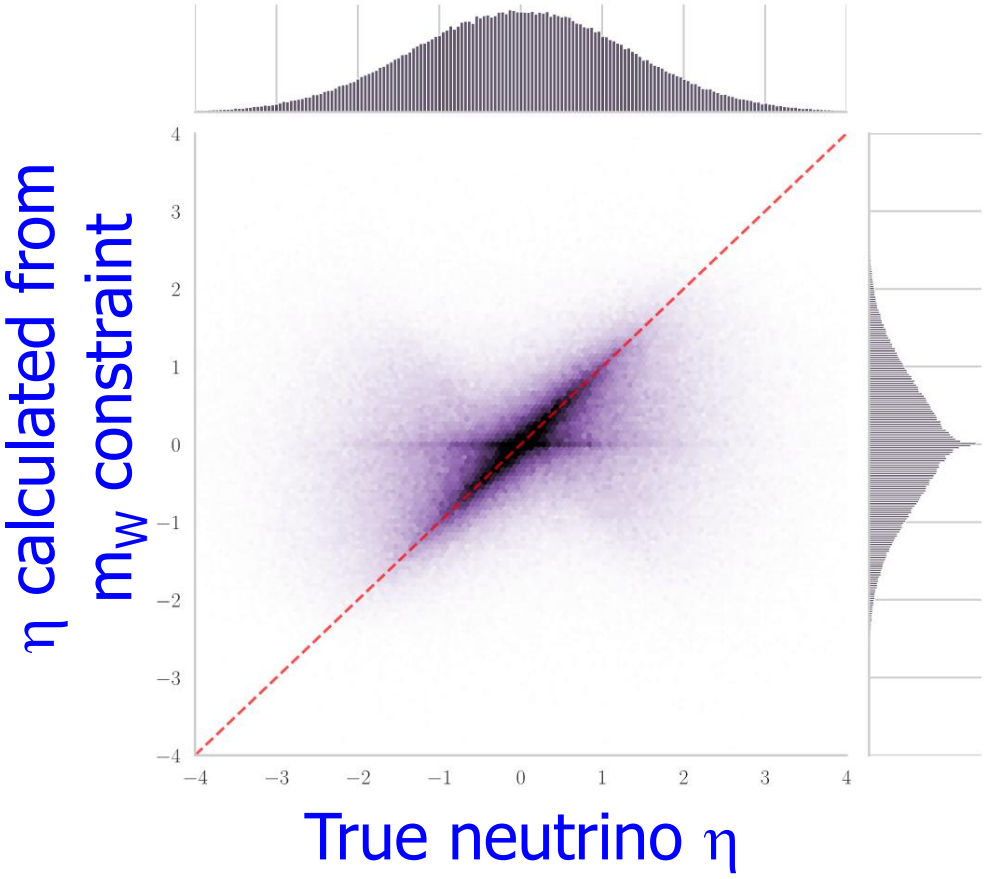


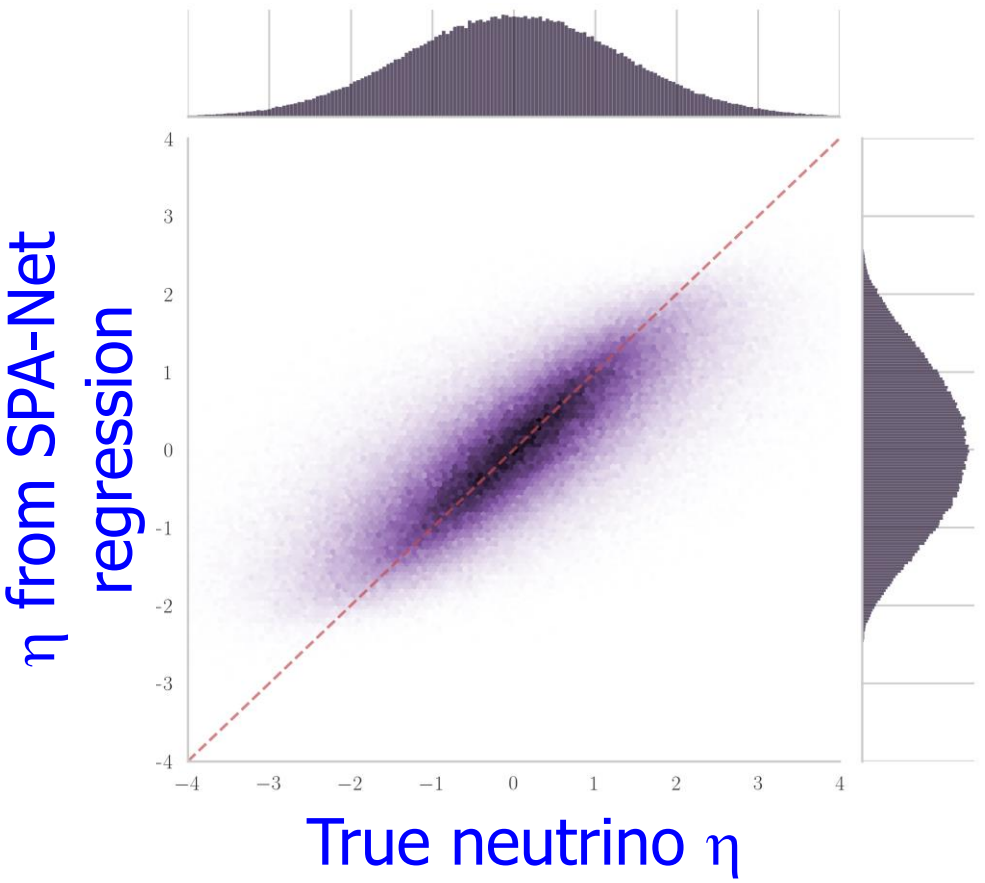




Excellent mass reconstruction performance; compatible w/ the true jet combination even for high jet multiplicities

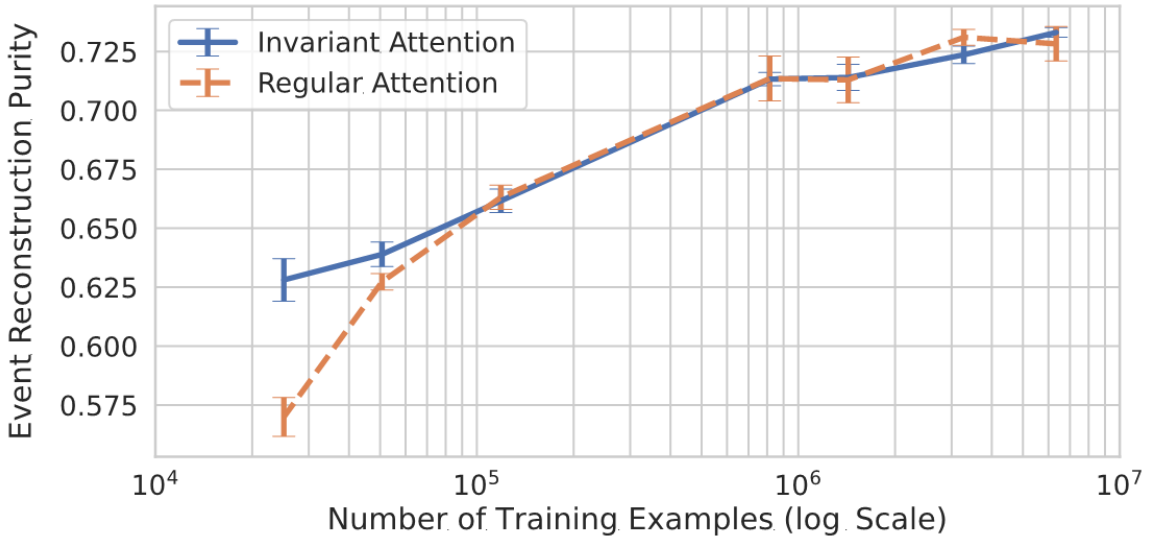
New Features: Regression of Kinematics

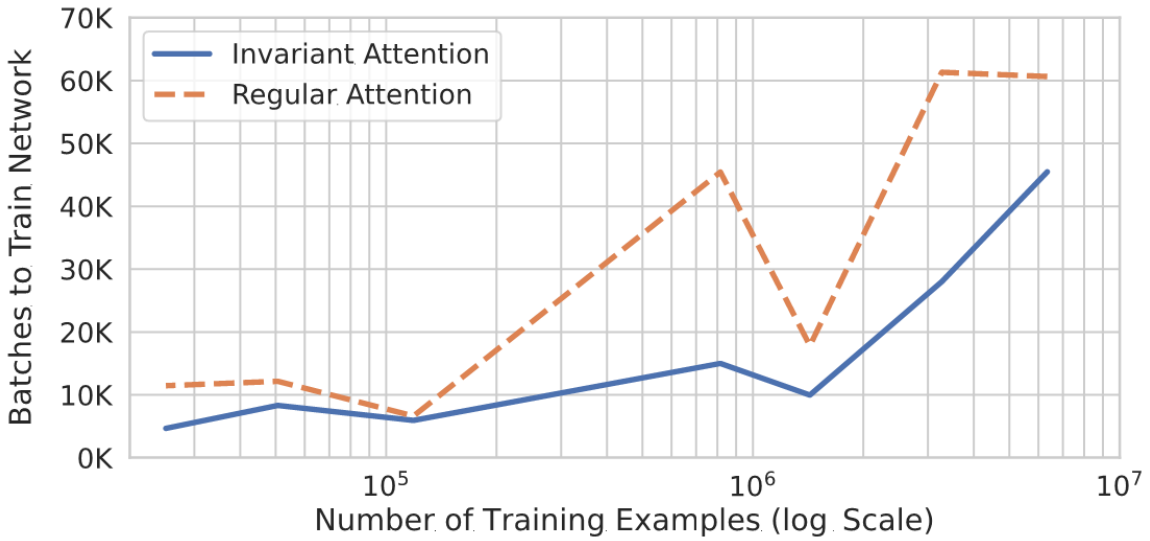




- SPA-Net can now train to reconstruct $t\bar{t}$ & simultaneously train to regress a variable (e.g. neutrino η or p_z, $t\bar{t}$ invariant mass)
- Neutrino η is more diagonal than the traditional method w/ improved RMS (1.39→0.9).

Considering Symmetries - Lorentz Invariance

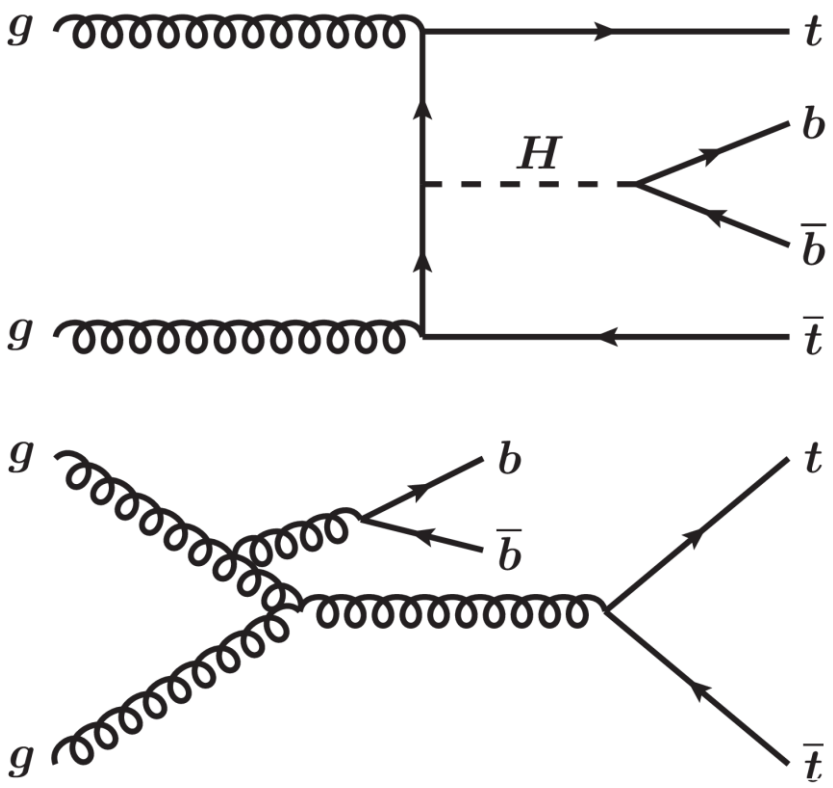


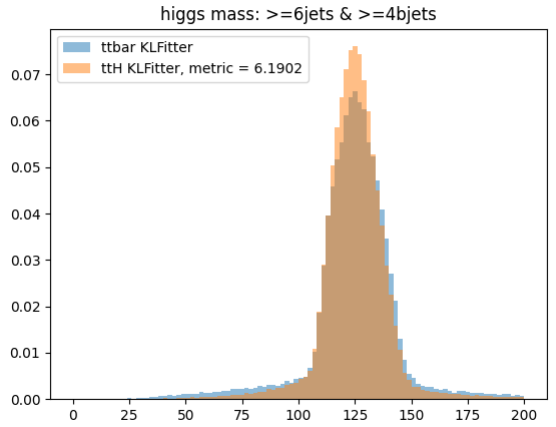


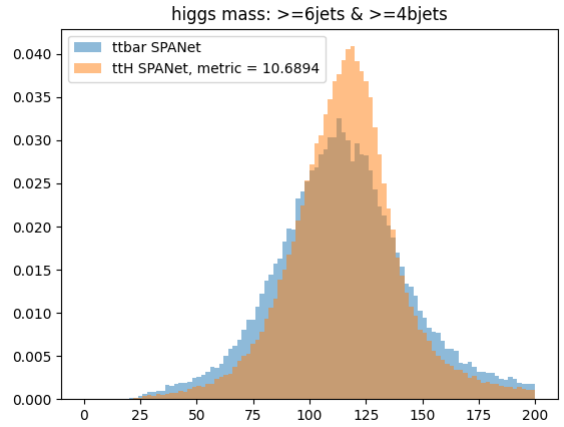
- Adding Lorentz invariance to the network does not change the jet-parton assignment accuracy for most cases, but improves the performance for small datasets.
- Lorentz invariance brings visible improvement in speed: i.e. significant reduction of batches needed to train the network.

Studies inspired by: *Congqiao Li, Huilin Qu, Sitian Qian, Qi Meng, Shiqi Gong, Jue Zhang, Tie-Yan Liu, and Qiang Li, arXiv.2208.07814 (2022)*

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







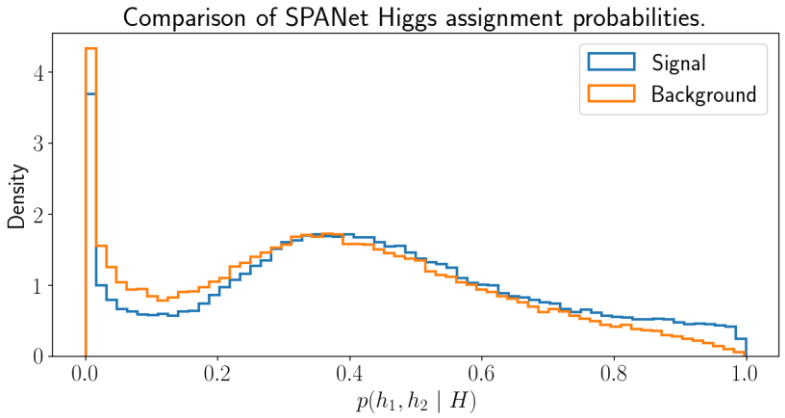
- $t\bar{t}b\bar{b}$ background is rather large with very similar kinematics to $t\bar{t}$H.
- t & H kinematics are main inputs to the BDT.
- However, the fraction of reconstructable events is only 35% in $t\bar{t}$H semi-lep. events.

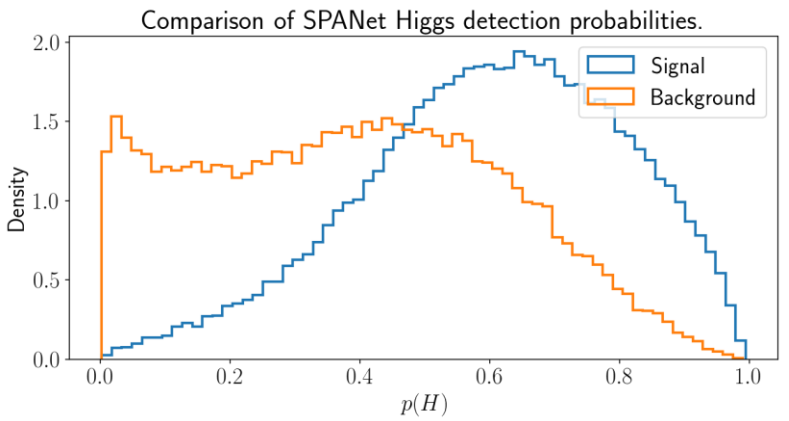
- **"Goodness" of the jet-parton assignment is also important to remove unreconstructable events.** → i.e. likelihood for KLFitter, a score for permutation DNN

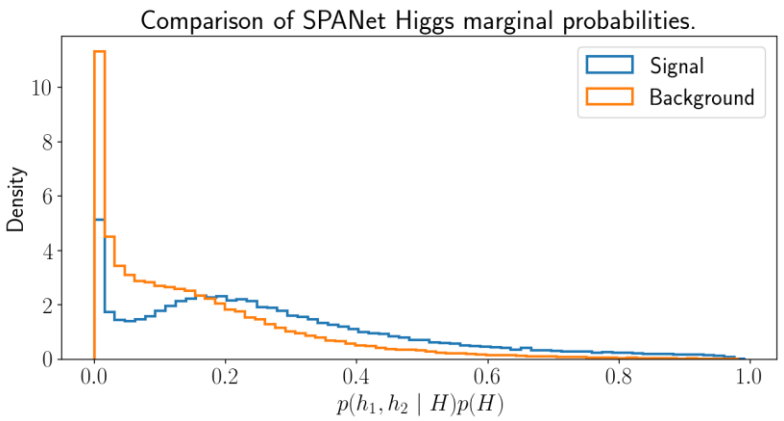
Updates: SPA-Net Probabilities

- Defined for each particle to reconstruct (t, H). **9 probabilities 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:** the product of the two above

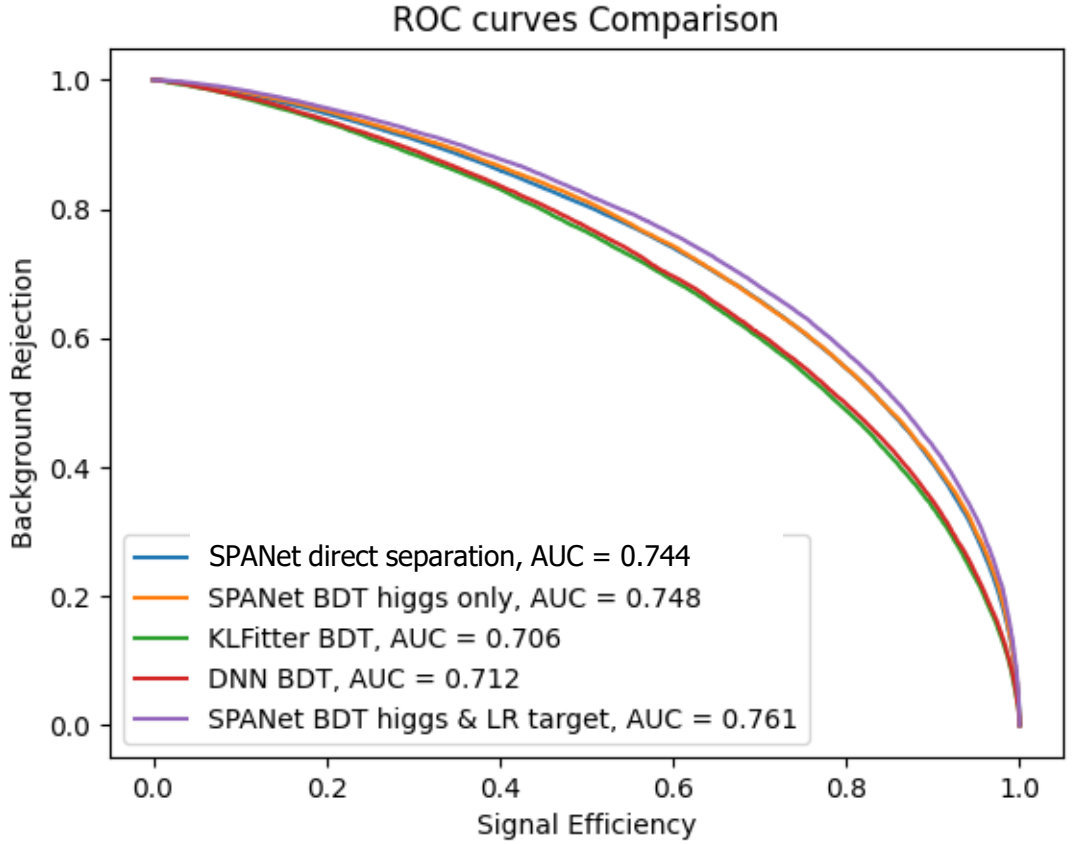






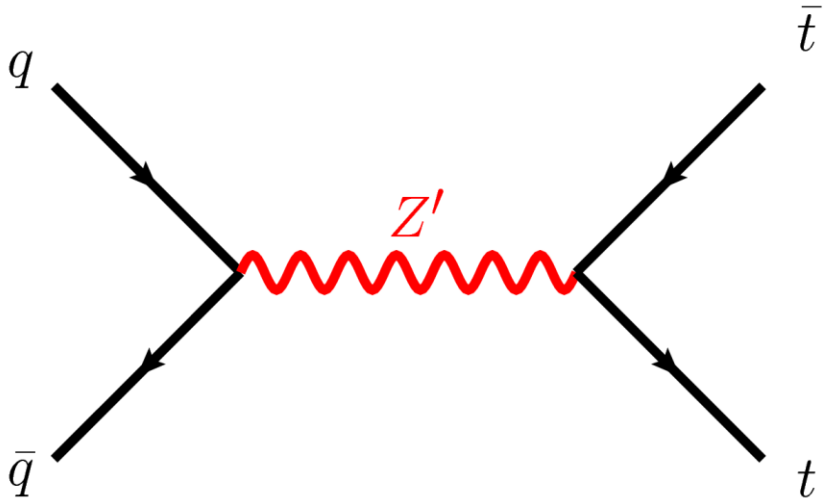
New Features: Signal/BG Discrimination

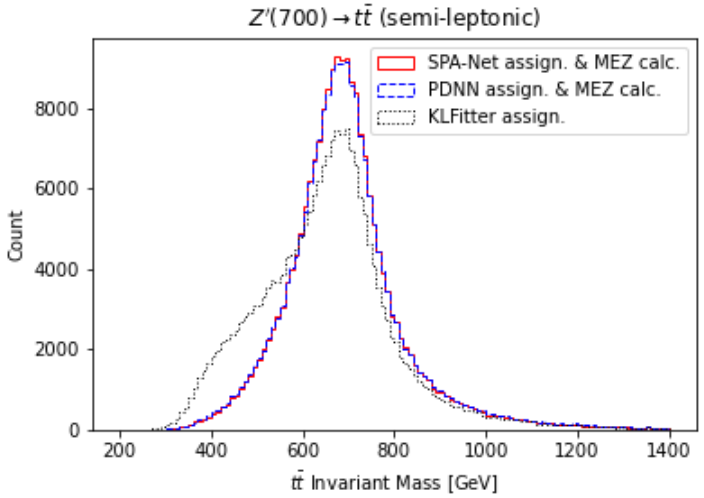
- SPA-Net probabilities alone already provide good separation: AUC~0.75 (not in the figure)
- A new feature in SPA-Net can provide ttH/ttbb discrimination directly. → ~same performance (0.744 vs 0.748) as the full BDT w/ kinematics & Higgs probabilities.
  - **This indicates that SPA-Net is learning the kinematics as well!**
- **SPA-Net for jet-assignment & with all 9 probabilities included in the BDT gives the best results.**

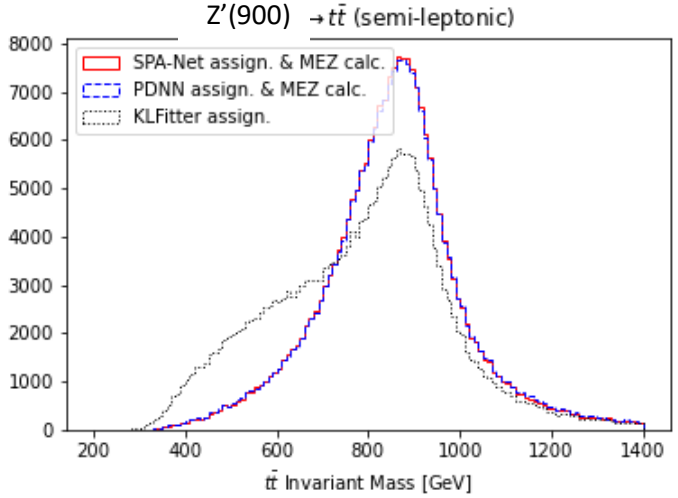


Transformer architecture provides us with meaningful embeddings for every jet, particle, and event: a big benefit over permutation-based models

Z'→$t\bar{t}$ Searches

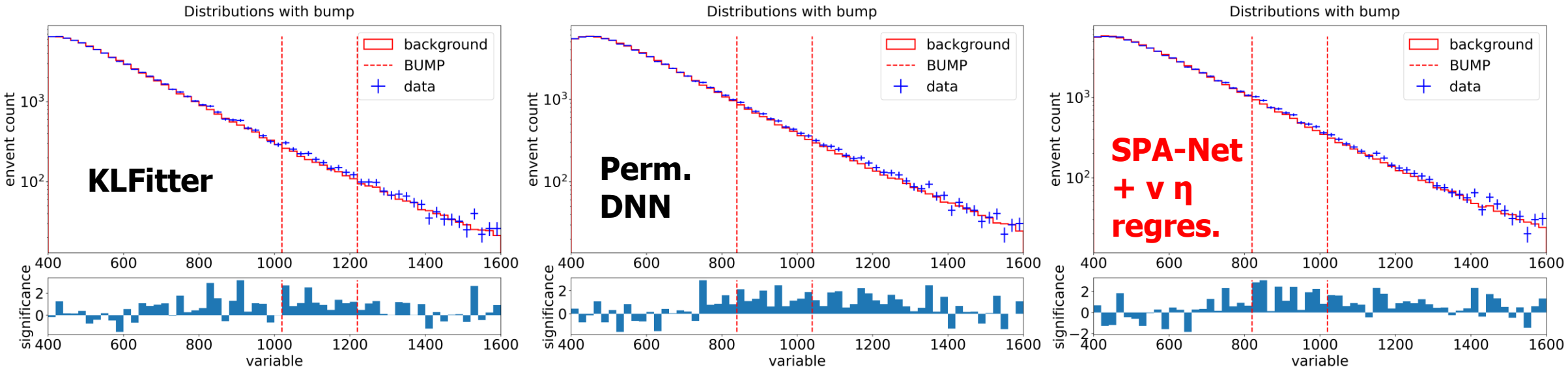






- Successful reconstruction of top quarks is crucial for the $t\bar{t}$ resonance searches (e.g. Z', RS Graviton, Heavy Higgs A/H→$t\bar{t}$).
- More jets are radiated for higher mass resonance. Significantly improved mass reconstruction from KLFitter.

WIP: Impact on Z'→$t\bar{t}$ Searches



- Visible improvement in signal significance: e.g. global significance = **2.96σ (KLFitter)**, 3.19σ (p.DNN), **3.87σ (SPA-Net), 4.41σ (SPA-Net+v η regression)** for m(Z')=900 GeV [integ. lumi.=3 fb^{-1} of pseudo-data assumed here]
- Removing unreconstructable events with SPA-Net probabilities should further enhance the sensitivity. (studies under way)

Thanks to Louis Vaslin for feedback on pyBumpHunter

SPA-Net Package (new version!)

https://github.com/Alexanders101/SPANet

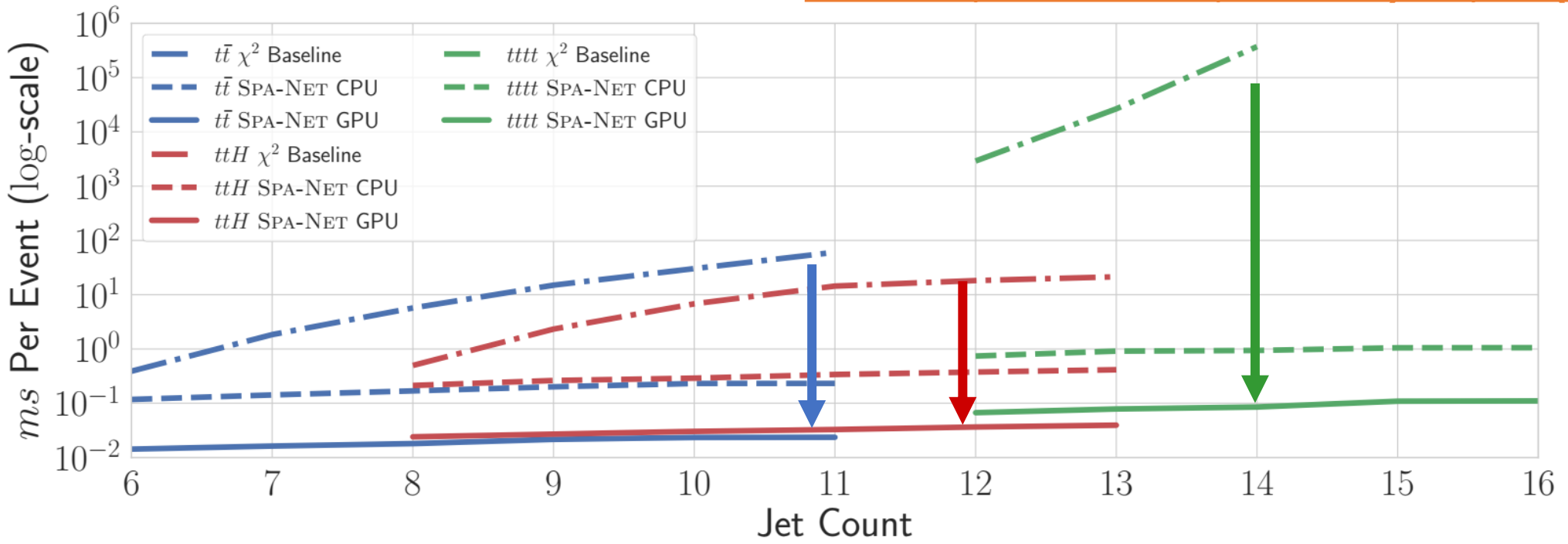
SPA-Net is not limited to top physics! Could be considered for SUSY RPV & HH→4b, for example

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.

CPU Time (prev. studies from allhad)

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



- **A few orders of magnitude improvement w/ SPA-Net compared to χ^2. A further acceleration w/ GPU.**
- Similar improvement expected for semi-leptonic final states & with much more visible improvement compared to KLFitter.

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.