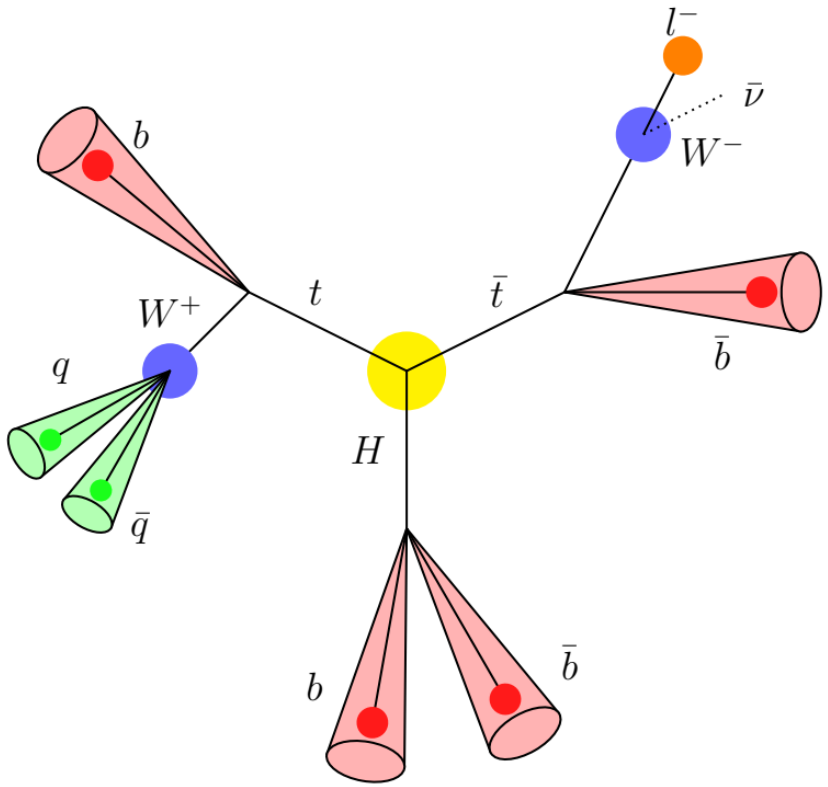
Symmetry Preserving Attention Networks (Spa-Net) for Resolved Top & Higgs Reconstruction at the LHC

Higgs 2023, November 27 – December 2, 2023, IHEP

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 for Heavy Particles



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

Unsorted list of jets

$j_1, j_2, j_3, j_4, j_5, j_6, j_7, j_8$

$\longleftrightarrow$

Target partons

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

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, PMLR 162:18281-18292 (2022)
  - Attention-Based Cloud Network (ABCNet): V. Mikuni, F. Canelli, EPJ Plus 135, 463 (2020); Lukas Gouskos, Fabio Iemmi, Sascha Liechti, Benedikt Maier, Vinicius Mikuni , Huilin Qu, PRD 108, 096003 (2023)
  - etc.

Symmetric Tensor Attentions

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

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!**

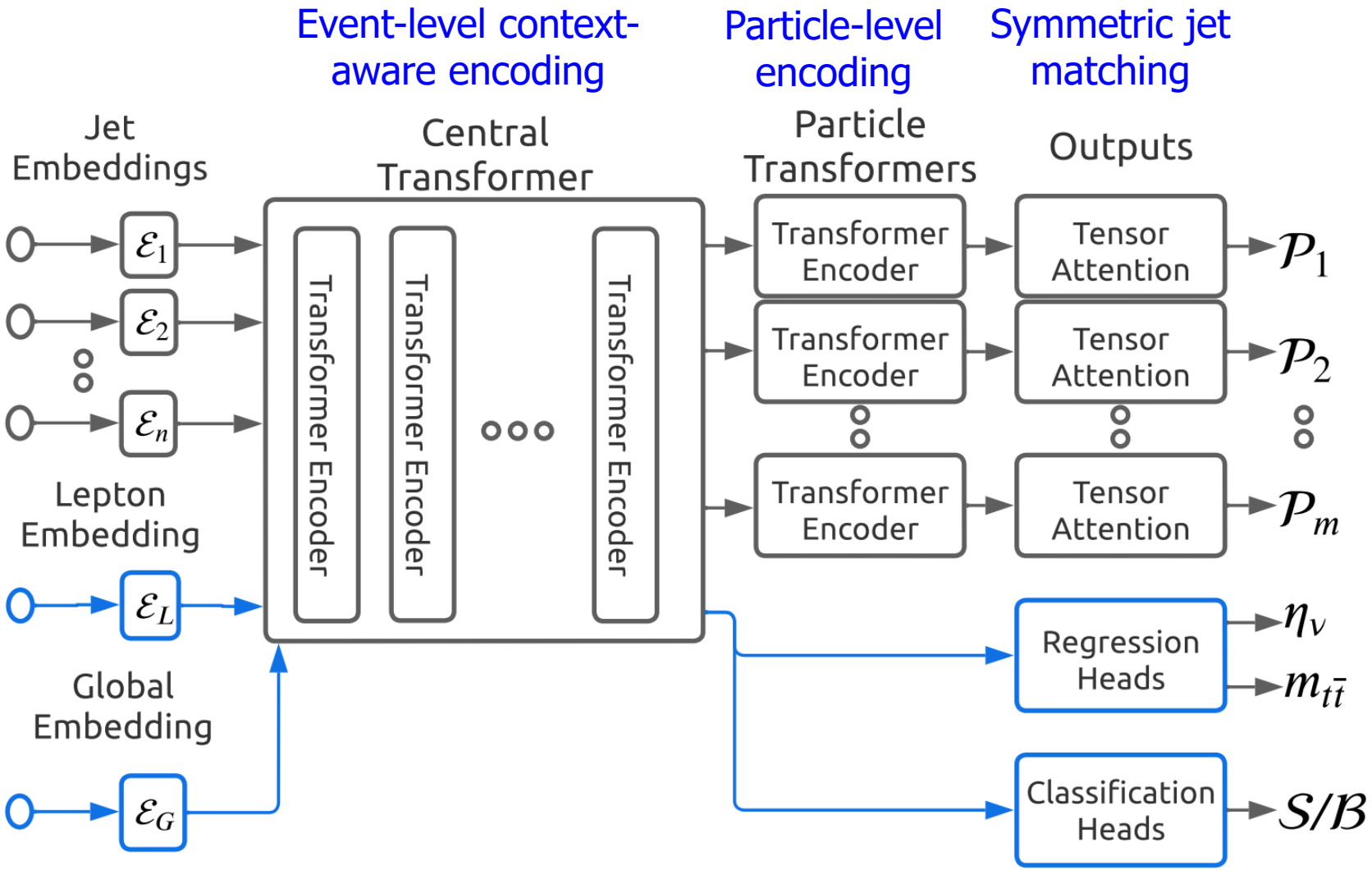
$$\mathcal{L}_{min}^{masked} = \min_{\sigma \in G_E} \left(\sum_{i=1}^{m} \frac{\mathcal{M}_{\sigma(i)} CE(\mathcal{P}_i, \mathcal{T}_{\sigma(i)})}{CB\left(\mathcal{M}_{\sigma(1)}, \mathcal{M}_{\sigma(2)}, \ldots, \mathcal{M}_{\sigma(m)}\right)} \right)$$

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*

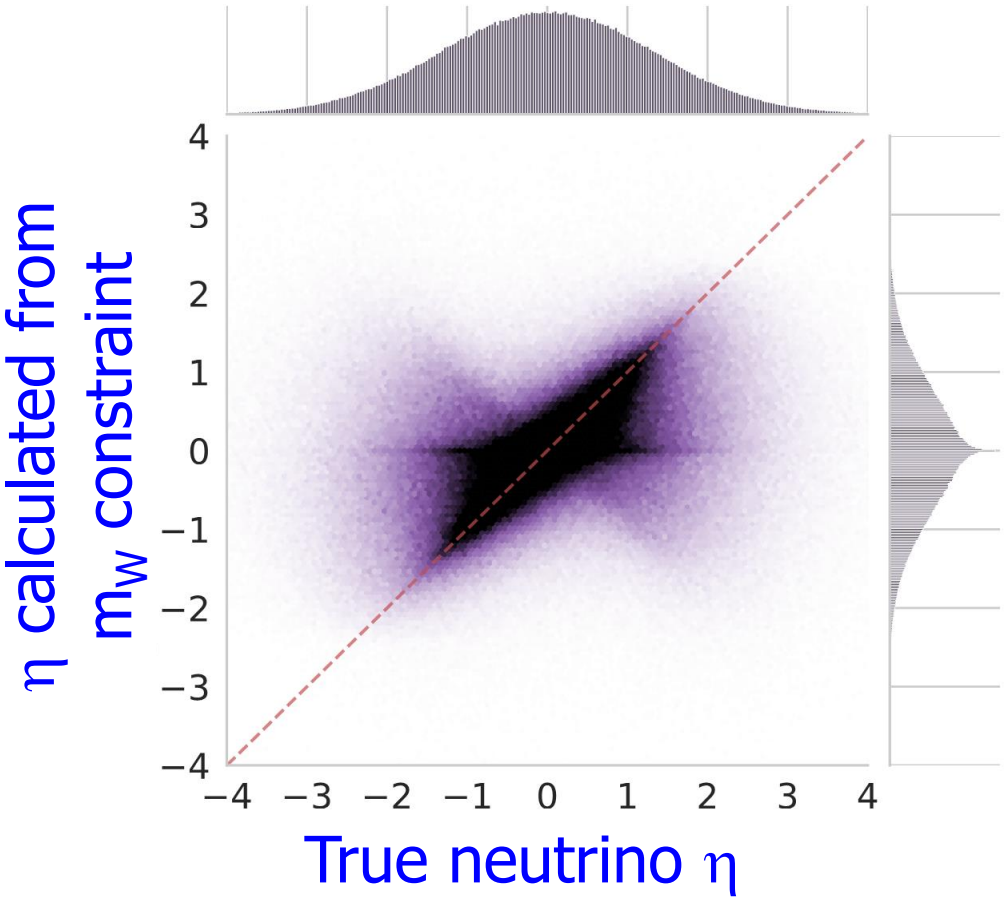
Reconstruction in semi-leptonic $t\bar{t}$ & $t\bar{t}H$

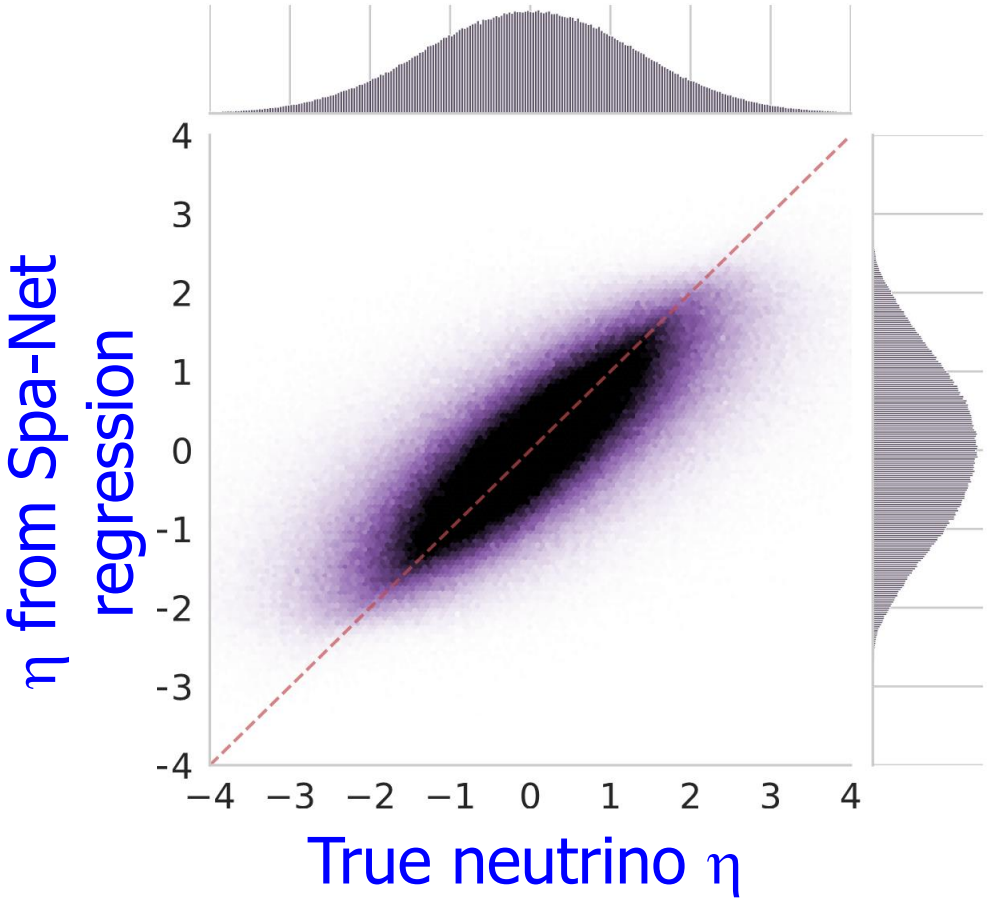
- New version of Spa-Net can handle different types of physics objects (jets, leptons, etc.)
- Can add event-level variables (MET, MET ϕ, etc.)

- $t\bar{t}$
  - Spa-Net: **75%** full event reconstruction (84-60% vs Njets)
  - Permutation DNN: **70%** (83-51%)
  - KLFitter: **41%** (71-12%)
- $t\bar{t}H(\rightarrow b\bar{b})$
  - Spa-Net: **45%** full event (54-33% vs Nj)
  - Permutation DNN: **39%** (49-28%)
  - KLFitter: **19%** (38-5%)

	N_{jets}	SPANet Eff. (%)				PDNN Eff. (%)				KLFitter Eff. (%)			
		Ev.	t_{H}	t_{L}	H	Ev.	t_{H}	t_{L}	H	Ev.	t_{H}	t_{L}	H
All $t\bar{t}$ Events	$= 4$	81	80	86	–	74	80	78	–	60	66	65	–
	$= 5$	74	72	84	–	68	69	79	–	32	37	47	–
	≥ 6	66	61	82	–	57	53	75	–	18	20	35	–
	All	**76**	**73**	**85**	–	**68**	**69**	**78**	–	**42**	**44**	**53**	–
Full $t\bar{t}$ Events	$= 4$	84	84	90	–	83	83	89	–	71	71	77	–
	$= 5$	73	74	87	–	69	71	84	–	28	39	52	–
	≥ 6	60	63	84	–	51	55	79	–	12	21	37	–
	All	**75**	**76**	**87**	–	**70**	**72**	**85**	–	**41**	**47**	**58**	–
All $t\bar{t}H$ Events	$= 6$	43	54	69	50	33	48	57	43	31	36	55	43
	$= 7$	39	48	68	49	31	43	58	43	16	25	47	29
	≥ 8	34	42	68	47	27	36	56	40	11	15	46	24
	All	**40**	**49**	**69**	**49**	**31**	**43**	**57**	**43**	**21**	**26**	**50**	**34**
Full $t\bar{t}H$ Events	$= 6$	54	65	73	62	49	60	68	58	38	49	60	48
	$= 7$	42	55	70	56	36	50	64	51	13	31	48	31
	≥ 8	33	47	69	52	28	42	60	46	05	17	46	24
	All	**45**	**57**	**71**	**57**	**39**	**52**	**64**	**52**	**19**	**33**	**52**	**35**

New Features: Regression of Kinematics

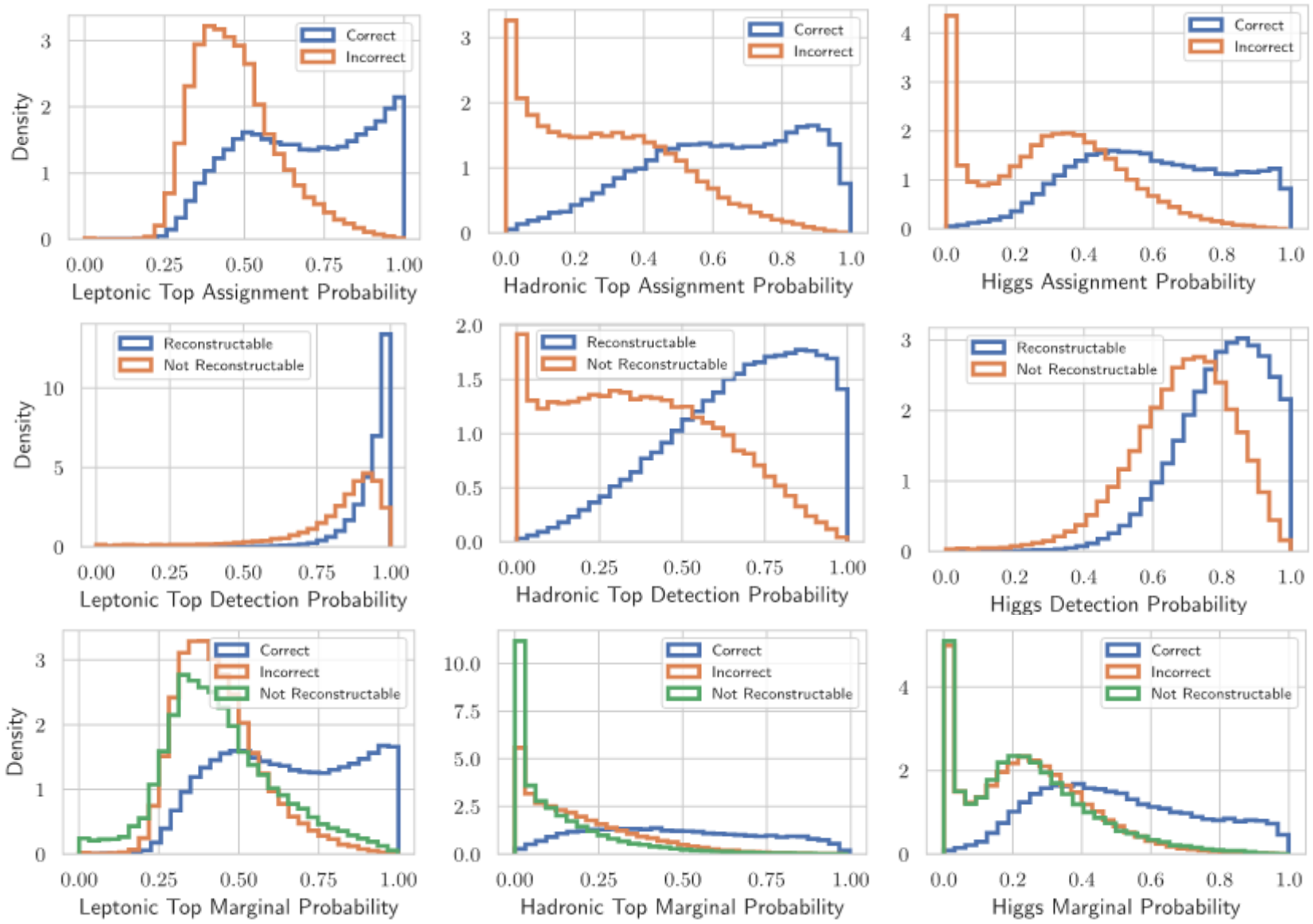




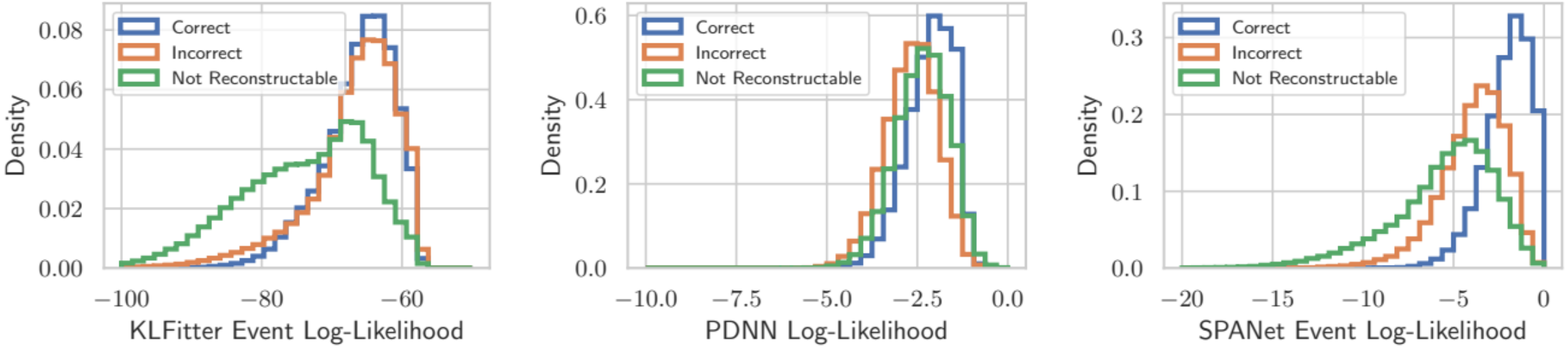
- Spa-Net can now train to reconstruct $t\bar{t}$ & simultaneously train to regress a variable (e.g. neutrino η or p_z, $t\bar{t}$ invariant mass)
- Neutrino η is **more diagonal** than the traditional method w/ **improved RMS** (1.39→0.9).

Particle Presence Outputs

- **Assignment probability:** Given *t* or *H* is present, are the predicted jets correct?
- **Detection probability:** Is *t* or *H* reconstructable?
- **Marginal probability:** essentially the product of the two above

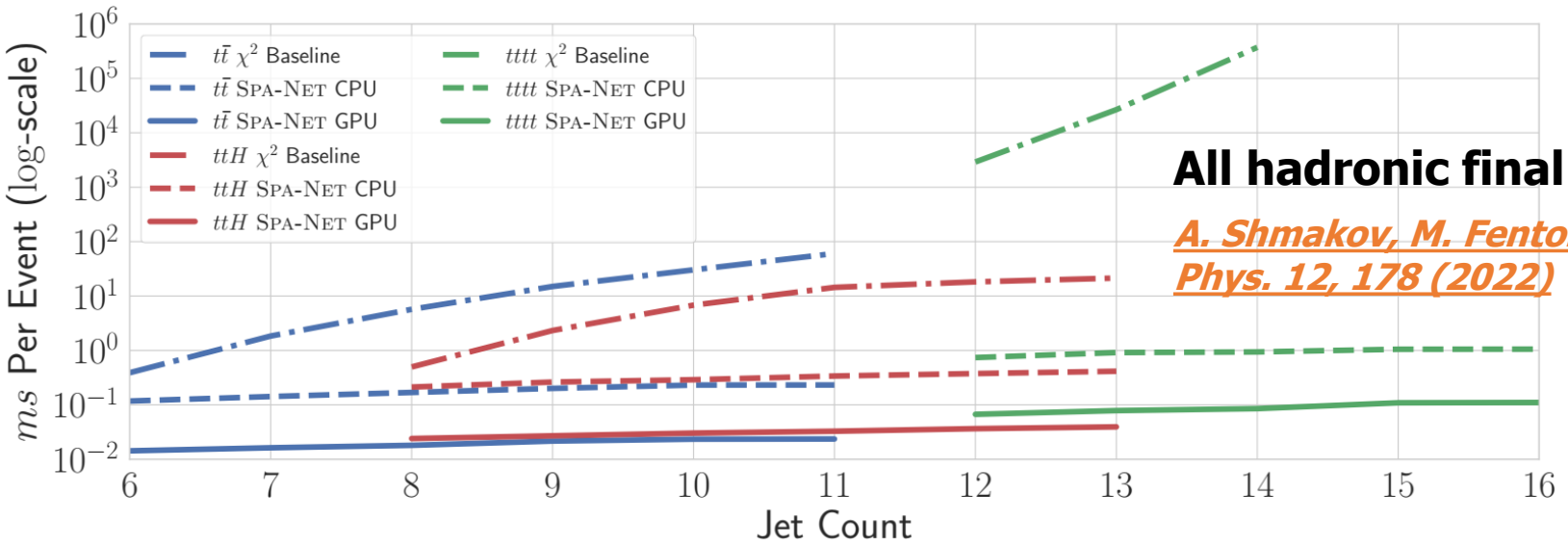


Likelihood Comparison



- Event-level likelihood is evaluated for KLFitter, PDNN & Spa-Net
- Permutation methods (KLFitter & PDNN) can only provide event-level scores w/ little separation for correct/incorrect assignments

Computational Overhead



All hadronic final states

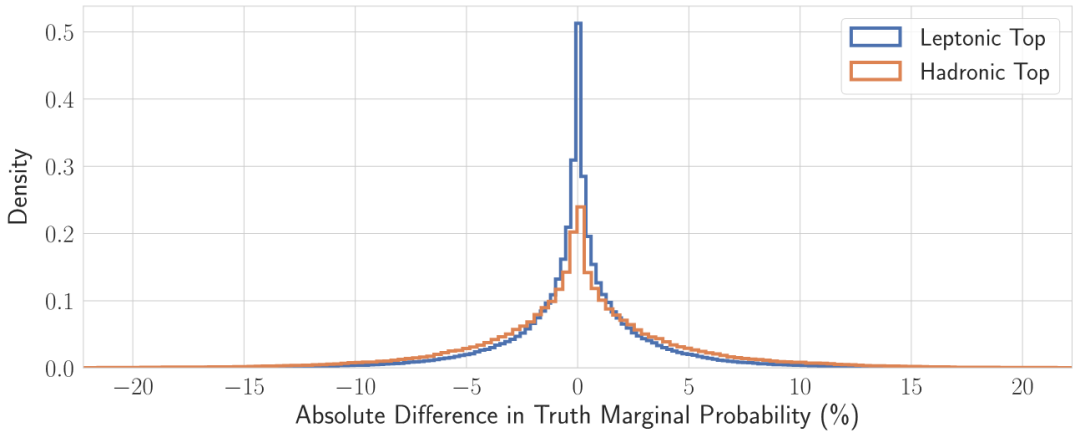
A. Shmakov, M. Fenton et al., SciPost Phys. 12, 178 (2022)

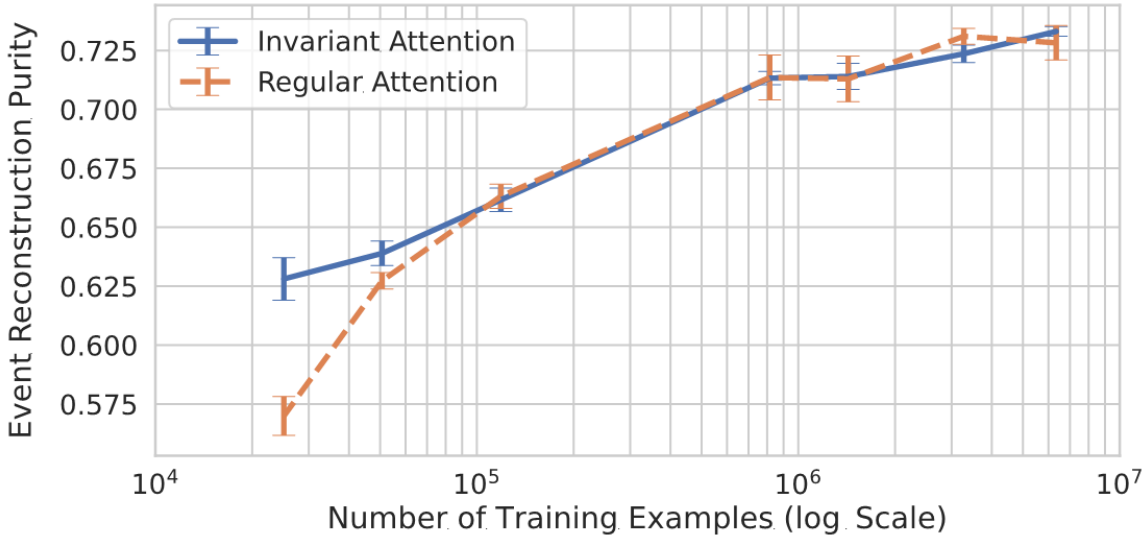
- **A few orders of magnitude improvement** w/ Spa-Net from χ^2 or KLFitter. A further acceleration w/ GPU.
- One order of magnitude faster than permutation DNN for ttH (i.e. high multiplicity events)!

Semi-leptonic final states

	$t\bar{t}$	$t\bar{t}H$
KLFitter	24 events per second	2 events per second
PDNN CPU	2626 events per second	51 events per second
PDNN GPU	3034 events per second	101 events per second
SPANet CPU	705 events per second	852 events per second
SPANet GPU	4407 events per second	3534 events per second

Ablation Studies: Symmetries





- **Network approximately learns Rotation & Lorentz symmetries!**
- Randomly rotate in ϕ & mirrored across the beam axis → negligible difference in performance
- Adding Lorentz invariance to the network does not change the jet-parton assignment accuracy for most cases, but improves the performance for small datasets & visible improvement in speed (see backup).

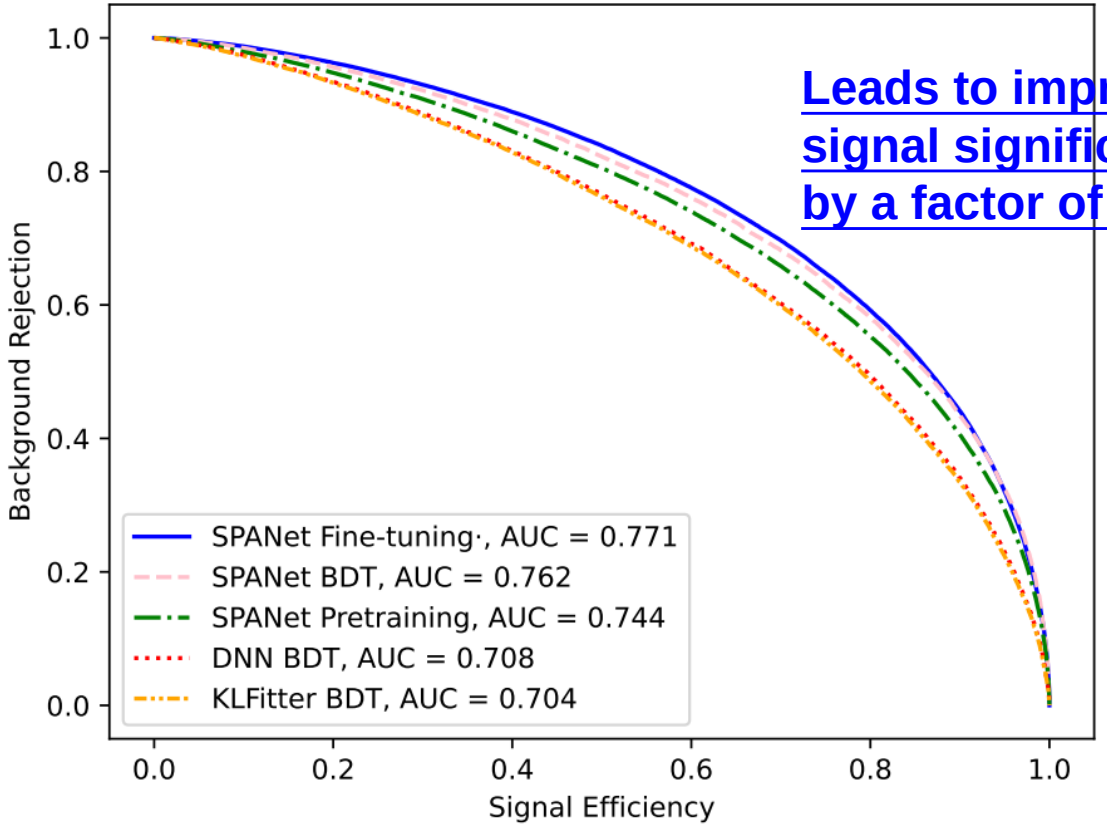
Consistent results as C. Li et al., arXiv:2208.07814 & S. Qiu et al., Phys. Rev. D 107, 114029 (2023)

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

- Spa-Net outperforms existing methods! (**AUC=0.704, 0.708 → 0.771**)
  - Full Spa-Net implementation w/ fine tuning is the best
  - Spa-Net jet assignment+BDT comes next
- **Spa-Net reaches 3(5)σ w/ Run-2(3)** whereas benchmark methods cannot

	Signal significance	Upper cross section limit [pb]	Upper signal strength limit
KLFitter BDT	2.4σ (4.1σ)	0.426 (0.248)	0.840 (0.489)
PDNN BDT	2.4σ (4.1σ)	0.421 (0.246)	0.831 (0.486)
SPA-NET BDT	3.0σ (5.2σ)	0.340 (0.196)	0.671 (0.387)
SPA-NET pre-training	2.7σ (4.8σ)	0.371 (0.214)	0.732 (0.423)
SPA-NET fine-tuning	3.1σ (5.7σ)	0.332 (0.179)	0.655 (0.353)

Numbers: Run 2 (Run 3)

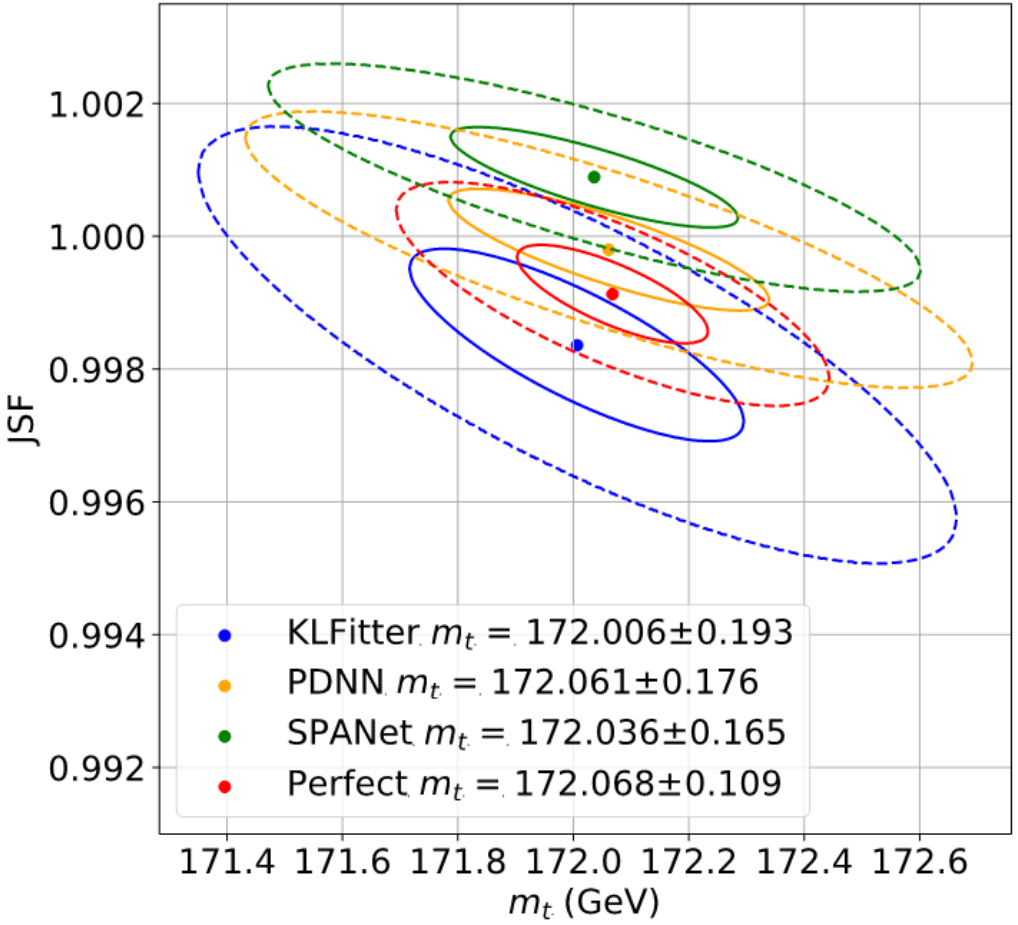


Leads to improved signal significance by a factor of ~1.4.

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

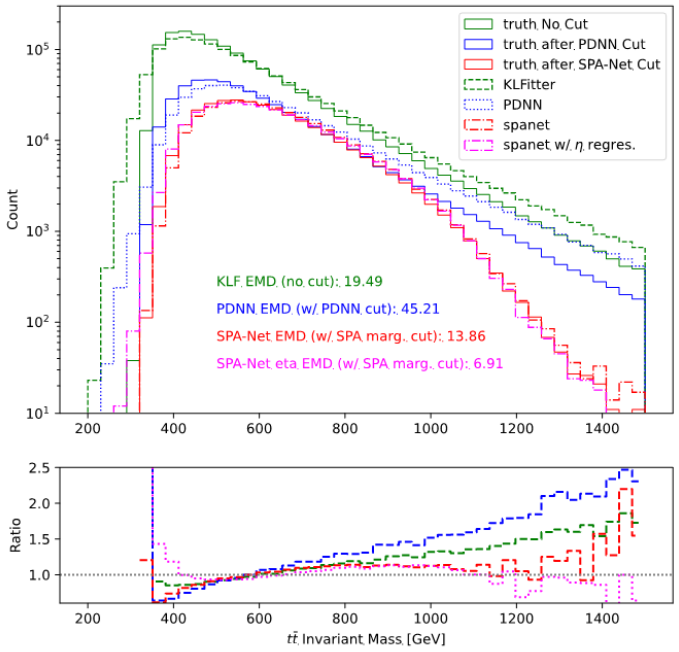
Top Quark Mass Measurement

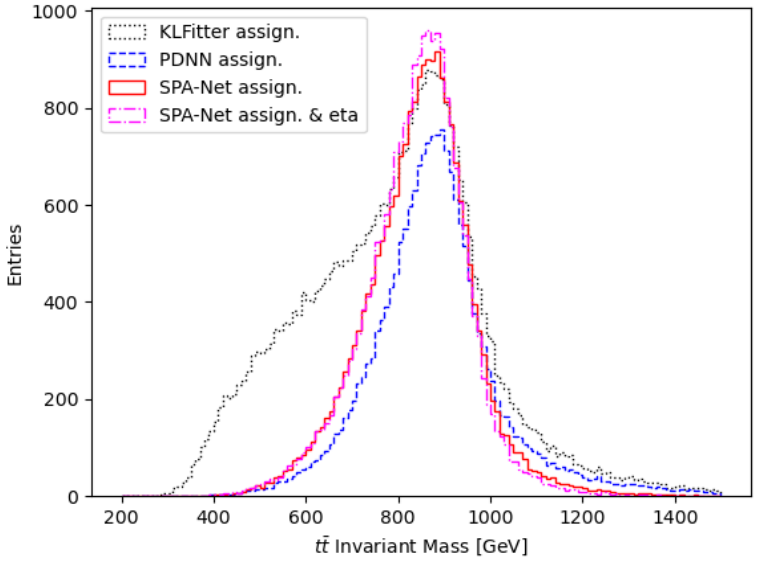
- 2D fit to the invariant mass distributions hadronic top quark & W performed.
- Only jet scale factor systematics (±4%) is considered.
- Top mass & JSF are extracted by a template fit from MC.
- **15% improvement in uncertainty from benchmark methods.** → still room for further improvement compared to the perfect jet assignment



Impact on Z′→$t\bar{t}$ Searches

- Z′→$t\bar{t}$ searches are already dominated by systematic uncertainty.
- **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. 2.5σ →5.5σ for Z′(500 GeV))





	KLFitter	PDNN	SPA-NET	SPA-NET w/ η^{ν}
$m_{Z'} = 500$ GeV	1.2σ (2.5σ)	1.8σ (3.5σ)	2.8σ (5.5σ)	2.7σ (5.4σ)
$m_{Z'} = 700$ GeV	1.6σ (3.3σ)	2.5σ (4.9σ)	3.1σ (6.1σ)	2.9σ (5.7σ)
$m_{Z'} = 900$ GeV	1.9σ (3.9σ)	2.8σ (5.5σ)	4.3σ (8.5σ)	4.1σ (8.2σ)

Numbers:
Run 2 (Run 3)

Assumes improved syst. in Run 3

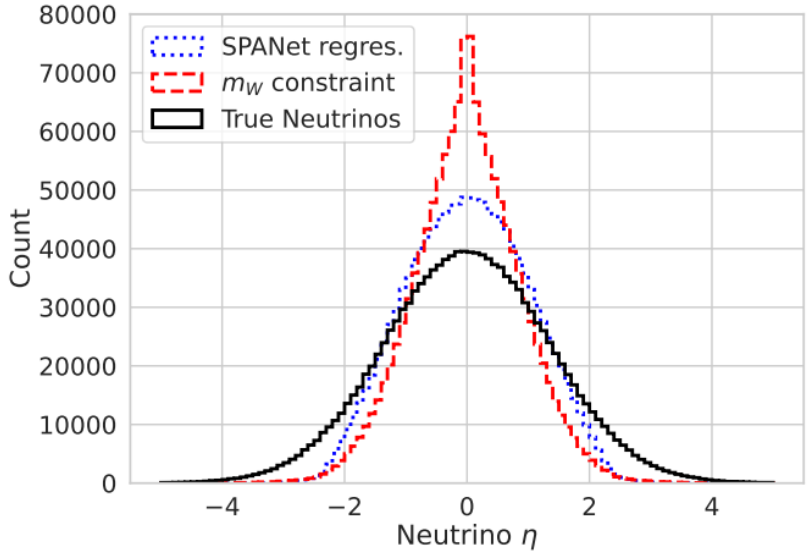
Summary

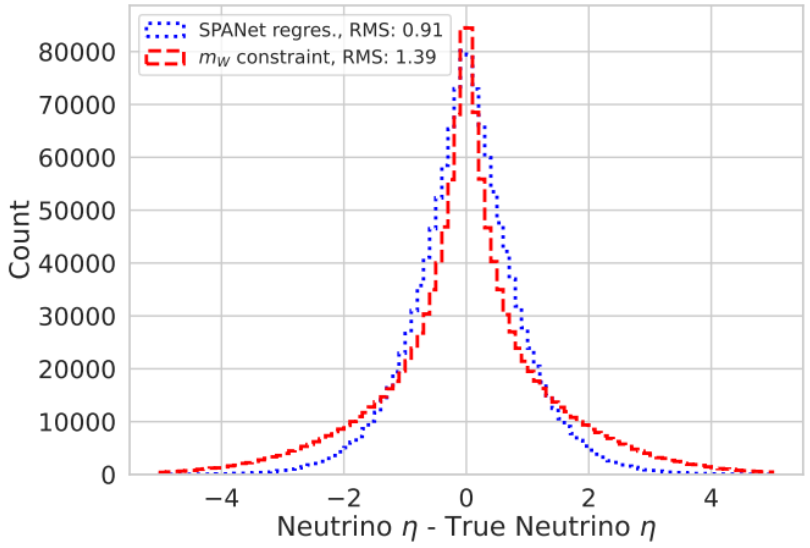
- Spa-Net provides efficient & excellent performance for event reconstruction in complex final states from multi-objects.
  - Superb CPU/GPU time, no limitation on object/jet multiplicity
  - Possible application to any jet-parton or even any "X"-"Y" assignment problem
- Transformer architecture provides 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:
  - **[newest paper!] M. Fenton, A. Shmakov, H. Okawa et al., arXiv:2309.01886 (2023)**
  - A. Shmakov, M.Fenton et al., SciPost Phys. 12, 178 (2022)
  - M. Fenton, A. Shmakov et al., Phys. Rev. D 105, 112008 (2022)

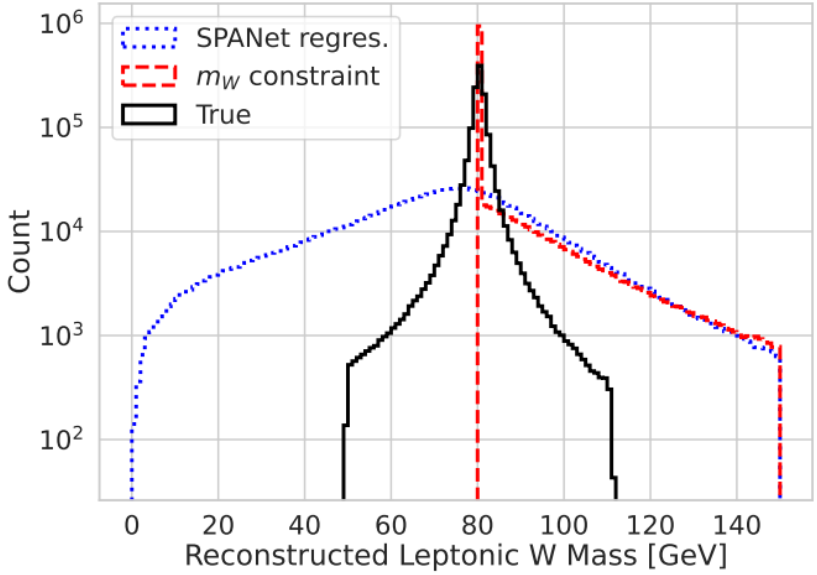
Code available at:
https://github.com/Alexanders101/SPANet

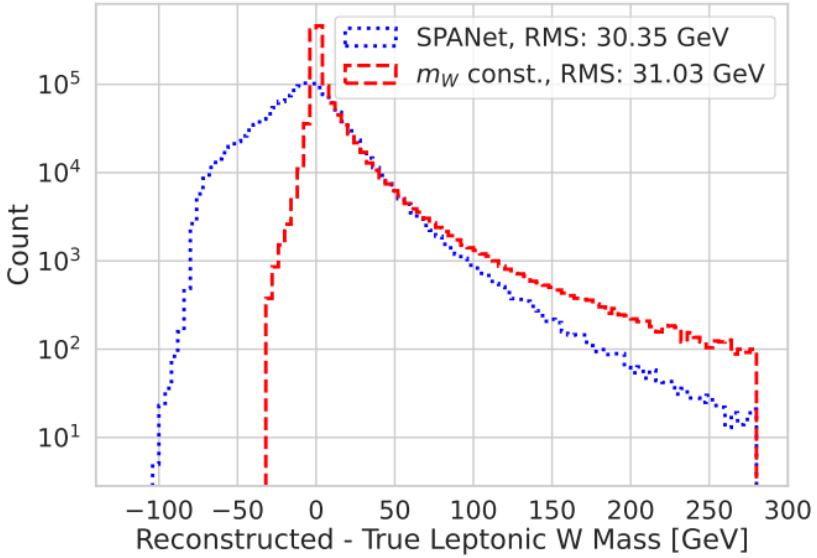
backup

New Features: Regression of Kinematics

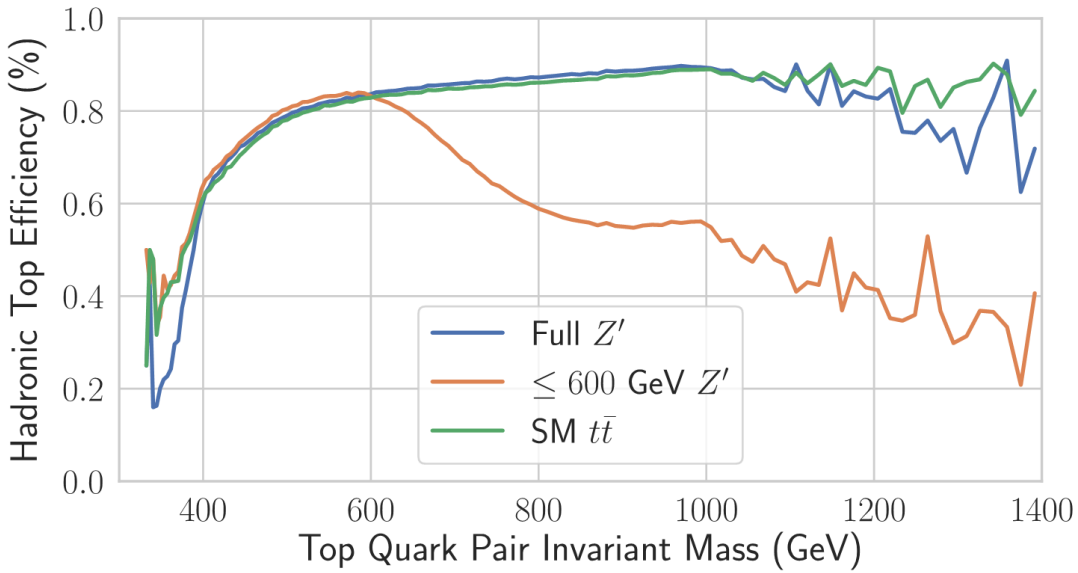






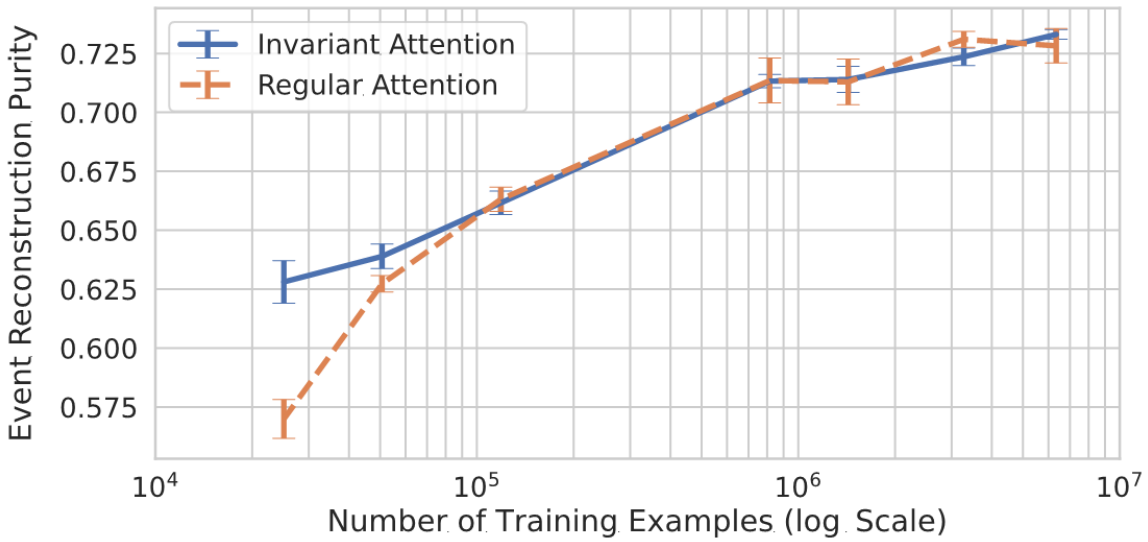


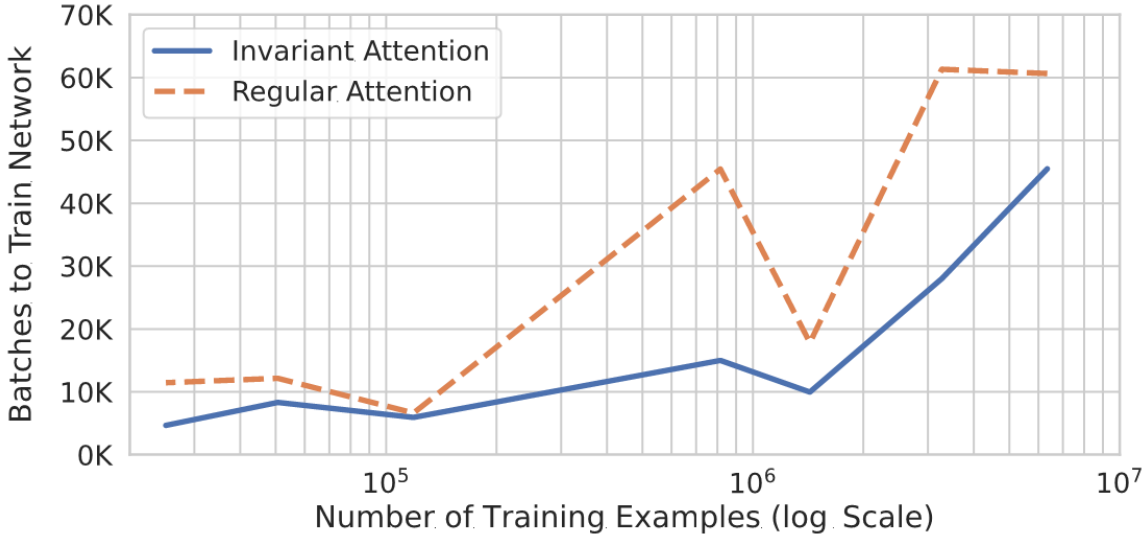
Ablation Studies: Kinematic Range



- **Kinematic range of the training is important!** → If trained on events w/ m_{tt}<600 GeV, reco eff. degrades in higher mass region
- No dependence on the process (SM $t\bar{t}$ or Z')

Ablation Studies: Lorentz Symmetry

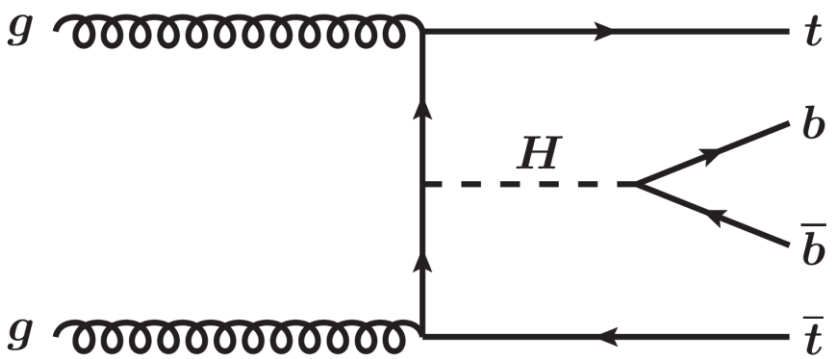


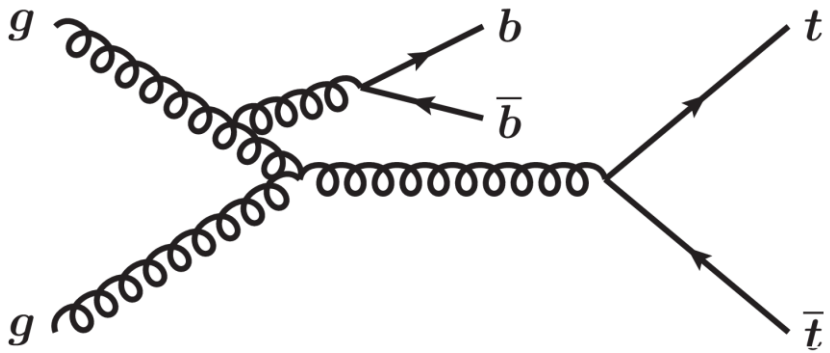


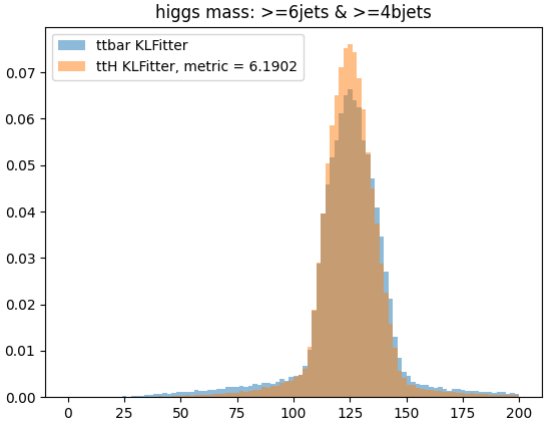
- Adding Lorentz invariance to the network does not change the jet-parton assignment accuracy for most cases, but improves the performance for small datasets.
- Lorentz invariance brings visible improvement in speed: i.e. significant reduction of batches needed to train the network.

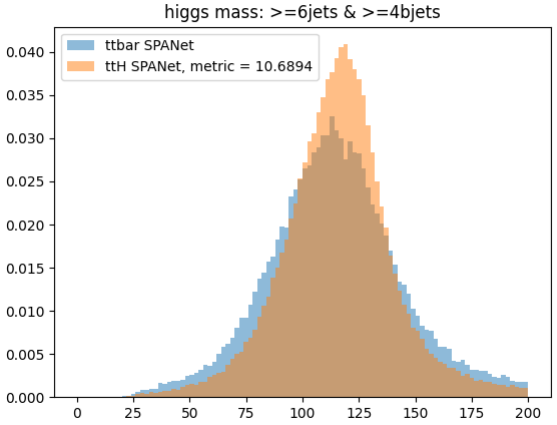
Consistent results as C. Li et al., arXiv:2208.07814 & S. Qiu et al., Phys. Rev. D 107, 114029 (2023)

$t\bar{t}$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.

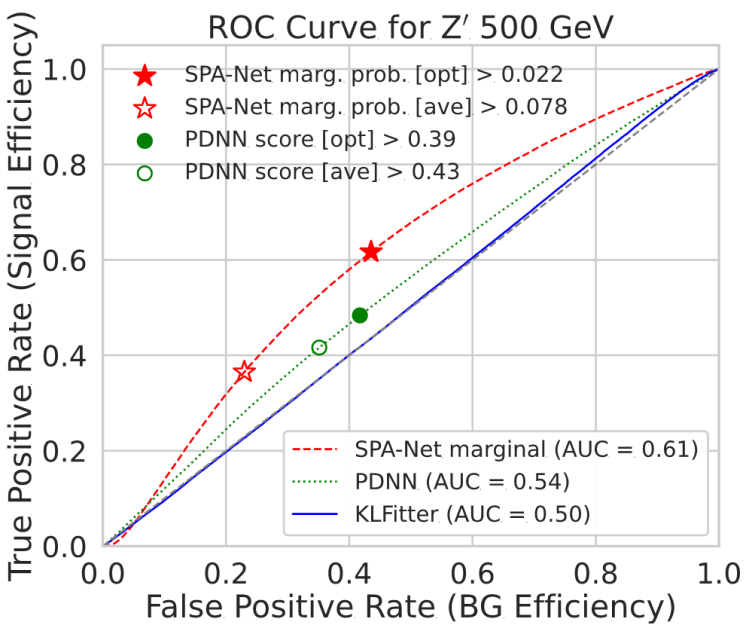
- **"Goodness" of the jet-parton assignment is also important to remove unreconstructable events.** → i.e. likelihood for KLFitter, a score for permutation DNN

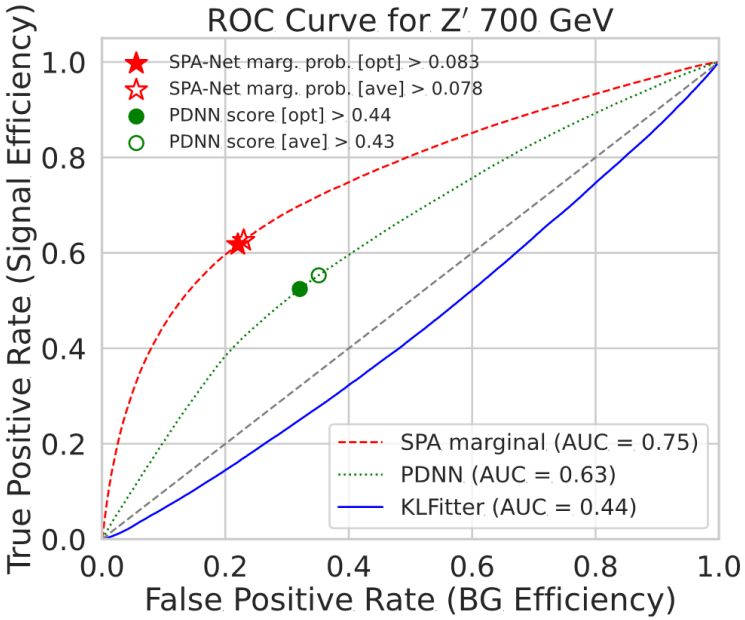
$t\bar{t}$H(→$b\bar{b}$) semi-leptonic

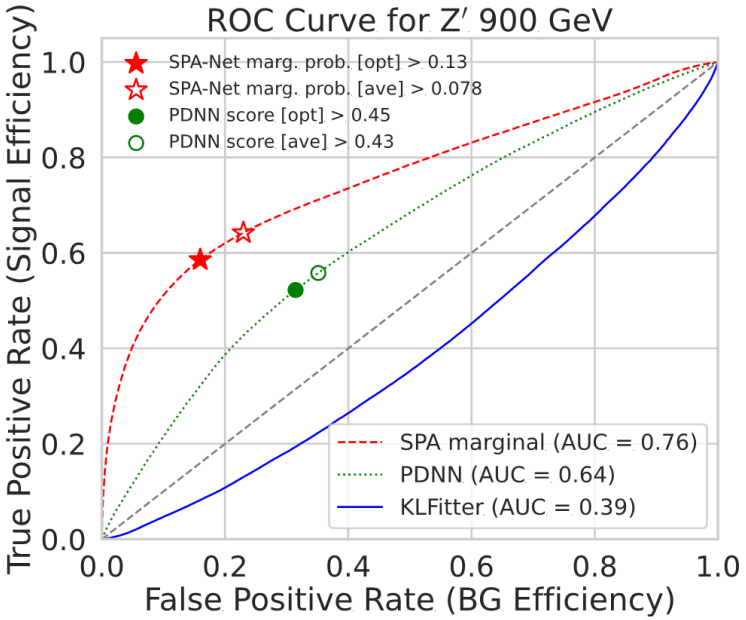
Variable	Definition
General kinematic variables	
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b-jet pairs
$\Delta R_{bb}^{\text{max } p_{\text{T}}}$	ΔR between the two b-jets with the largest vector sum p_{T}
$\Delta\eta_{jj}^{\text{max}}$	Maximum $\Delta\eta$ between any jet pairs
$H_{\text{T}}^{\text{had}}$	Scalar sum of jet p_{T}
$m_{bb}^{\text{min } \Delta R}$	Mass of two b-jets with the smallest ΔR
$m_{\text{jj}}^{\text{min } \Delta R}$	Mass of any jet pair with the smallest ΔR
$N_{bb}^{\text{Higgs 30}}$	Number of b-jet pairs with invariant mass within 30 GeV of the Higgs mass
$\Delta R_{l,bb}^{\text{min}}$	Smallest ΔR between the lepton and the combination of the two b-jets
Variables with Higgs boson and top quark reconstruction	
m_{bb}^{Higgs}	Mass of the Higgs boson candidate
$m_{H,b_{\text{lep top}}}$	Mass of the Higgs boson candidate and b-jet from leptonic top quark candidate
$\Delta R_{bb}^{\text{Higgs}}$	ΔR between b-jets from the Higgs boson candidate
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs boson candidate and $t\bar{t}$ candidate system
$\Delta R_{H,\text{leptop}}$	ΔR between Higgs boson candidate and leptonic top quark candidate
$\Delta R_{H,b_{\text{hadtop}}}$	ΔR between Higgs boson candidate and b-jet from hadronic top candidate decay
Scores from jet-parton assignment	
LHD	Log-likelihood discriminant from the KLFitter
A_{PDNN}	Assignment score from Permutation DNN
A_{higgs}	Assignment probability of Higgs boson target from SPA-NET
D_{higgs}	Detection probability of Higgs boson target from SPA-NET
M_{higgs}	Marginal probability of Higgs boson target from SPA-NET
E_{higgs}	Assignment entropy of Higgs boson target from SPA-NET
A_{leptop}	Assignment probability of leptonic top quark candidate from SPA-NET
D_{leptop}	Detection probability of leptonic top quark candidate from SPA-NET
M_{leptop}	Marginal probability of leptonic top quark candidate from SPA-NET
E_{leptop}	Assignment entropy of leptonic top quark candidate from SPA-NET
A_{hadtop}	Assignment probability of hadronic top quark candidate from SPA-NET
D_{hadtop}	Detection probability of hadronic top quark candidate from SPA-NET
M_{hadtop}	Marginal probability of hadronic top quark candidate from SPA-NET
E_{hadtop}	Assignment entropy of hadronic top quark candidate from SPA-NET

- Most kinematic variables are taken from the latest ATLAS results.

$Z' \rightarrow t\bar{t}$







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 & ttH 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.