Machine Learning in Top Physics in the ATLAS & CMS Collaborations

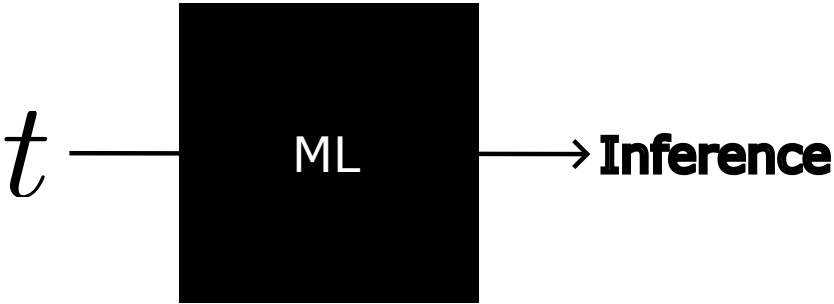
Philip Keicher on behalf of the ATLAS & CMS Collaborations
Durham, September 9th, 2022

Special thanks to Marcel Vos, Michael Fenton, Kai-Feng Chen and Valentina Guglielmi for providing material!

Introduction

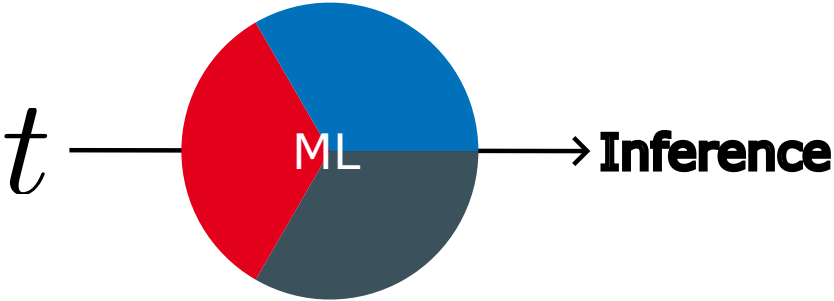
$$t \longrightarrow \textbf{Inference}$$

Introduction



- ML-techniques standard for many aspects of top-related physics

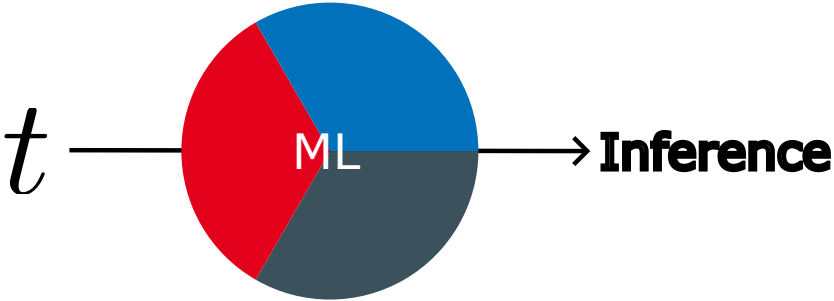
Introduction



- ML-techniques standard for many aspects of top-related physics
  - Tagging [b-tagging (CMS)] [b-tagging (ATLAS)] [top-tagging landscape paper]
  - Reconstruction [TOP22, M. Quinnan] [TOP22, S. Wuchterl] [TOP22, J. Park] [TOP22, R. Kogler]
  - Construction of final observable [TOP22, J. Lambert] [TOP22, J. v.d.Linden] [TOP22, T. Tran] [TOP22, J. S. Wilson] [TOP22, S. Wuchterl] [TOP22, L. Cerrito] [TOP22, J. Park] [TOP22, R. Kogler] [TOP22, D. Scheirich] [A. Smolkovic] [TOP22, B. Lopes] [TOP22, A. Rey] [TOP22, N. Manganelli]

$\rightarrow$ Multitude of different strategies and architectures

Introduction



- ML-techniques standard for many aspects of top-related physics
  - Tagging [b-tagging (CMS)] [b-tagging (ATLAS)] [top-tagging landscape paper]
  - Reconstruction [TOP22, M. Quinnan] [TOP22, S. Wuchterl] [TOP22, J. Park] [TOP22, R. Kogler]
  - Construction of final observable [TOP22, J. Lambert] [TOP22, J. v.d.Linden] [TOP22, T. Tran] [TOP22, J. S. Wilson] [TOP22, S. Wuchterl] [TOP22, L. Cerrito] [TOP22, J. Park] [TOP22, R. Kogler] [TOP22, D. Scheirich] [A. Smolkovic] [TOP22, B. Lopes] [TOP22, A. Rey] [TOP22, N. Manganelli]

→ Multitude of different strategies and architectures

- Today's talk: DNN as final observable and future prospects

EFT Analysis in associated Z production

[CMS-TOP-21-001]

- Search for BSM physics in the scope of Effective Field Theory (EFT) [TOP22, J. S. Wilson]

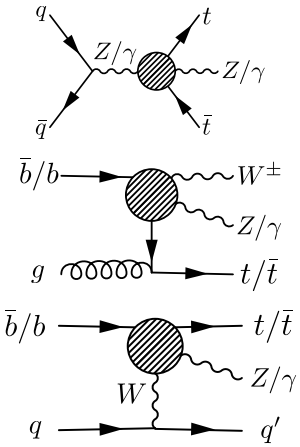
$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_{i}^{\dim(6)} \frac{c_i}{\Lambda^2} O_i$$

- Targeting $t\bar{t} + Z$, tZW, tZq processes
- Considering EFT operators

O_{tZ}, O_{tW}	EWK dipole moments
$O^{3}_{\varphi Q}$	left-handed SU(2) triplet current
$O^{-}_{\varphi Q}$, $O_{\varphi t}$	neutral current operators for left-(right-) handed top quarks

- Events distributed into five orthogonal regions: SR-3l, SR-4l, 2 control regions, 1 data-driven region

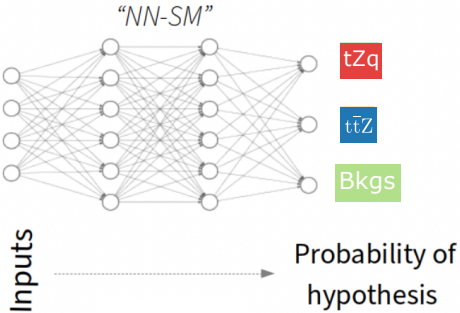
→ **New**: Considers interference effects **during** ML training!

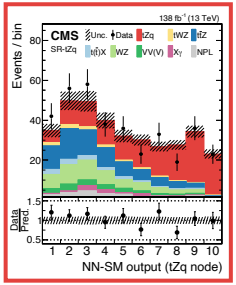


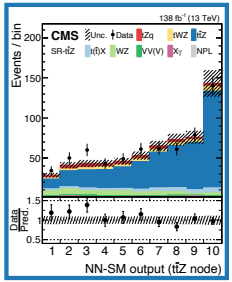
[arXiv:1904.05637]

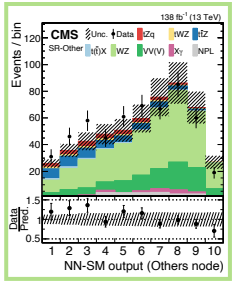
ML in CMS EFT analysis

[CMS-TOP-21-001]



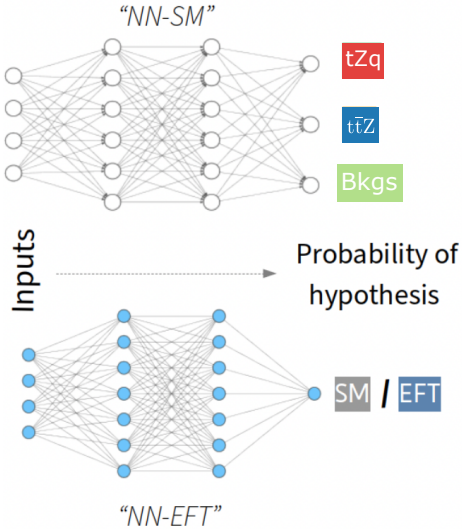






ML in CMS EFT analysis

[CMS-TOP-21-001]



- Further trainings with events classified as tZq and $t\bar{t} + Z$

→ **Binary classification** of SM vs. EFT with following setups

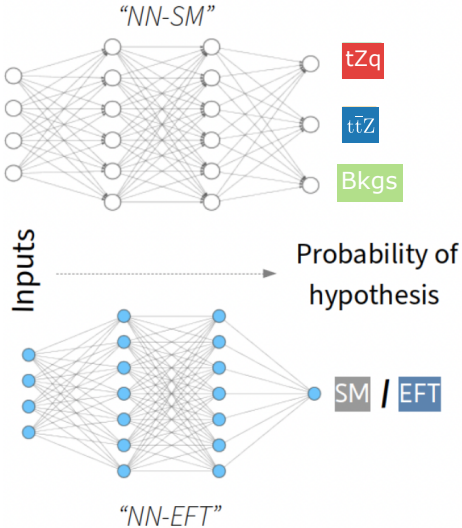
1D: Consider only one operator at a time

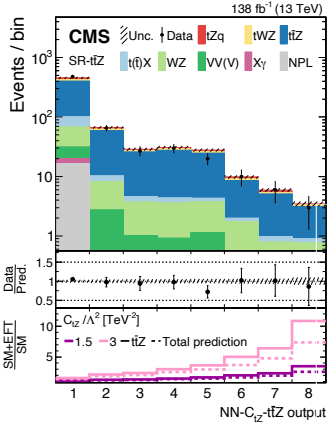
5D: Consider effects from all operators simultaneously

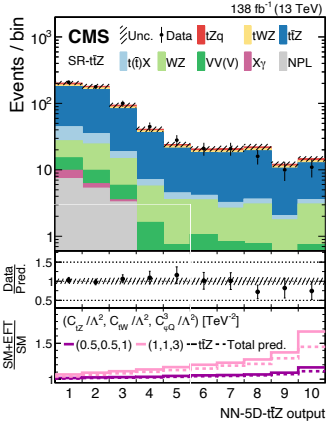
⇒ Total of **8 EFT NNs**

ML in CMS EFT analysis

[CMS-TOP-21-001]







Results

[CMS-TOP-21-001]

WC/Λ^2 $[\mathrm{TeV}^{-2}]$	95% CL confidence intervals			
	Other WCs fixed to SM		5D fit	
	Expected	Observed	Expected	Observed
c_{tZ}	$[-0.97, 0.96]$	$[-0.76, 0.71]$	$[-1.24, 1.17]$	$[-0.85, 0.76]$
c_{tW}	$[-0.76, 0.74]$	$[-0.52, 0.52]$	$[-0.96, 0.93]$	$[-0.69, 0.70]$
$c^{3}_{\varphi Q}$	$[-1.39, 1.25]$	$[-1.10, 1.41]$	$[-1.91, 1.36]$	$[-1.26, 1.43]$
$c^{-}_{\varphi Q}$	$[-2.86, 2.33]$	$[-3.00, 2.29]$	$[-6.06, 14.09]$	$[-7.09, 14.76]$
$c_{\varphi t}$	$[-3.70, 3.71]$	$[-21.65, -14.61] \bigcup [-2.06, 2.69]$	$[-16.18, 10.46]$	$[-19.15, 10.34]$

- Increase in sensitivity by usage of ML between 20-70 %

→ **ML crucial for this analysis**

Charge asymmetry in $t\bar{t}\gamma$ production at ATLAS

[ATLAS-CONF-2022-049] [TOP22, J. v.d.Linden] [TOP22, A. Rey] [TOP22, J. Keaveney]

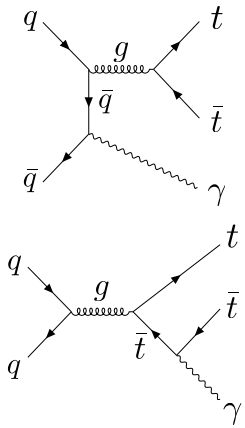
- Goal: Measure asymmetry in quark-initiated $t\bar{t}$ production

$$A_C = \frac{(N(|y_t|) > N(|y_{\bar{t}}|)) - (N(|y_t|) < N(|y_{\bar{t}}|))}{(N(|y_t|) > N(|y_{\bar{t}}|)) + (N(|y_t|) < N(|y_{\bar{t}}|))}$$

- Challenge: A only visible in quark-initiated production → diluted at LHC

→ Analyze $t\bar{t}\gamma$ production in semi-leptonic decay channel
  - γ for tagging
  - **Background** contribution: ~ **70 %** in final analysis phase space

→ Use ML to **enhance signal purity**

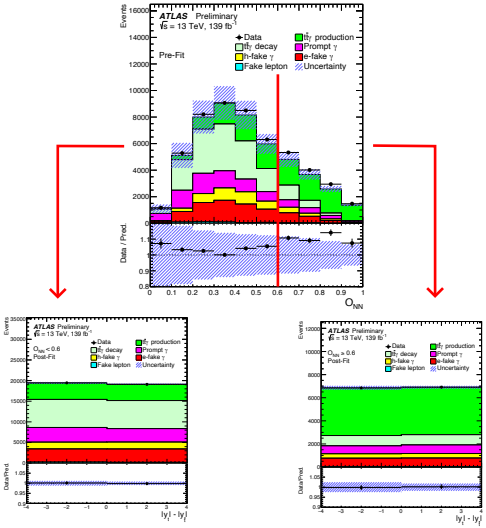


ML in ATLAS $t\bar{t}\gamma$ analysis

[ATLAS-CONF-2022-049] [TOP22, A. Rey]

- Feed-forward, **three hidden layers** with 96 (16) nodes
- **5-fold cross validation**
- Binary cross-entropy loss
- 21 input features (kinematic/angular variables of objects, b-tagging information, event shape information)
- Use **NN output to define signal & background regions**

→ **Background** contribution in SR: ~ **45 %**



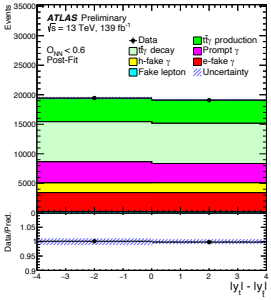
Results

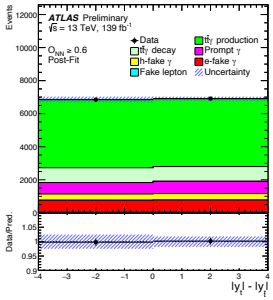
[ATLAS-CONF-2022-049] [TOP22, A. Rey]

- Final strategy: **Unfolding**
- Result: $\boldsymbol{A_C} = \mathbf{-0.006 \pm 0.030}$

→ **Compatible with SM or null** ($\boldsymbol{A_C^{\mathrm{MC}}} = \mathbf{-0.014 \pm 0.001}$)

- Dominated by statistical uncertainties, handling of (some) **systematics challenging**



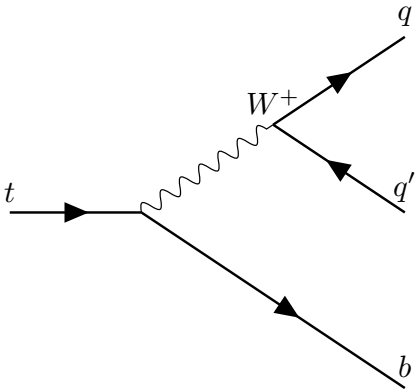


ML in Future Applications

Current solutions: boosted-top tagger in ATLAS with open data

[ATL-PHYS-PUB-2022-039]

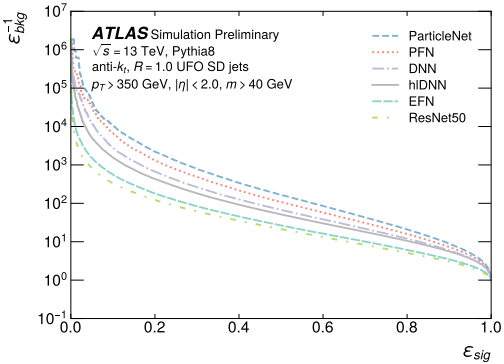
- Goal: Tag systems from boosted top quarks
- Update on [top-tagging landscape paper] (Delphes sim)



Current solutions: boosted-top tagger in ATLAS with open data

[ATL-PHYS-PUB-2022-039]

- Goal: Tag systems from boosted top quarks
- Update on [top-tagging landscape paper] (Delphes sim)
- Test different network architectures with **full sim**
  - Best performance by graph-NN
  - CNN architectures worse than expected
- Data with full detector simulation available at [CERN OpenData] for further study



Looking forward: Future applications of ML

- Modeling of underlying physics crucial for many top-related analyses

→ Parameters in simulation programs (e.g. Pythia) crucial

- Many parameters not derivable from first principles, obtained with ambiguities
- Solution: Generate dedicated simulations with varied parameter values
- **Problem**: dedicated samples often have **limited statistics**, challenging to obtain meaningful uncertainty estimate

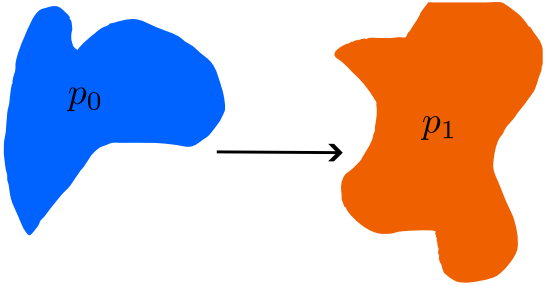
→ Question: Can ML techniques help here?

Possible solution: DCTR approach

[PhysRevD.101.091901]

- Underlying problem: find transformation $p_0(x) \rightarrow p_1(x)$

$\rightarrow$ find $w(x) = p_0(x)/p_1(x)$



Possible solution: DCTR approach

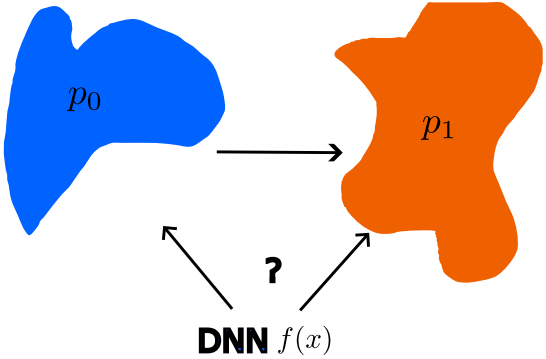
[PhysRevD.101.091901]

- Underlying problem: find transformation $p_0(x) \rightarrow p_1(x)$

$\rightarrow$ find $w(x) = p_0(x)/p_1(x)$

- Problem can be solved with ML

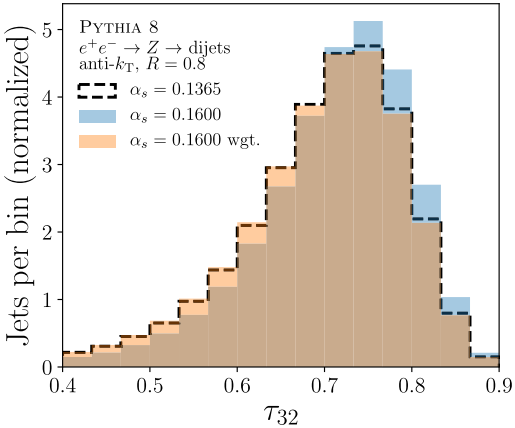
$$w(x) \sim \frac{f(x)}{1 - f(x)}$$



Advantages of DCTR

[PhysRevD.101.091901]

- Use **full, multidimensional** phase space for derivation of weights
- Enables **continuous weights** for whole phase space
- Can also be used for **derivation of the parameters** directly from data
- Approach already used in GAN application [JINST 15 P11004] and top-related analyses [JHEP09(2021)058]
- On-going study within the collaborations to check applicability



Conclusions

- ML has **significant role** in top physics!
- Wide array of strategies and applications, very **active field of research**
  - CMS example [CMS-TOP-21-001]
  - ATLAS example [ATLAS-CONF-2022-049]
- Many **new developments** on-going
  - **DCTR** [PhysRevD.101.091901]
  - But also **much more**! E.g. [TOP22, M. Fenton]

→ **Stay tuned!**

Backup

EFT coefficients

[CMS-TOP-21-001]

Operator	WC	Mapping to Warsaw-basis coefficients
$\mathcal{O}_{\mathrm{tZ}}$	c_{tZ}	$\mathrm{Re}\{-s_{\mathrm{W}}c_{\mathrm{uB}}^{(33)}+c_{\mathrm{W}}c_{\mathrm{uW}}^{(33)}\}$
$\mathcal{O}_{\mathrm{tW}}$	c_{tW}	$\mathrm{Re}\{c_{\mathrm{uW}}^{(33)}\}$
$\mathcal{O}_{\varphi\mathrm{Q}}^{3}$	$c_{\varphi\mathrm{Q}}^{3}$	$c_{\varphi\mathrm{q}}^{3(33)}$
$\mathcal{O}_{\varphi\mathrm{Q}}^{-}$	$c_{\varphi\mathrm{Q}}^{-}$	$c_{\varphi\mathrm{q}}^{1(33)}-c_{\varphi\mathrm{q}}^{3(33)}$
$\mathcal{O}_{\varphi\mathrm{t}}$	$c_{\varphi\mathrm{t}}$	$c_{\varphi\mathrm{u}}^{(33)}$

Analysis selections

[CMS-TOP-21-001]

Selection requirement	SR-3ℓ	SR-$t\bar{t}Z$-4ℓ	WZ CR	ZZ CR
Lepton multiplicity	$=3$	$=4$	$=3$	$=4$
$m_{3\ell} - m_{Z}$	—	—	>15 GeV	—
Z boson candidates multiplicity	$=1$	$=1$	$=1$	$=2$
Jet multiplicity	≥ 2	≥ 2	—	—
b jet multiplicity	≥ 1	≥ 1	$=0$	—
p_{T}^{miss}	—	—	>50 GeV	—

Input features

[CMS-TOP-21-001]

Variable	NN-SM	NN-c_{tZ}-tZq	NN-c_{tZ}-ttZ	NN-c_{tW}-tZq	NN-c_{tW}-ttZ	NN-$c_{\varphi Q}^{3}$-tZq	NN-$c_{\varphi Q}^{3}$-ttZ	NN-5D-tZq	NN-5D-ttZ
p_T^Z	—	✓	✓	✓	✓	✓	✓	✓	✓
$\eta(Z)$	✓	✓	✓	—	—	✓	—	—	✓
$\Delta\phi(\ell_1^Z \ell_2^Z)$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$p_T(t)$	✓	✓	✓	—	✓	✓	—	✓	✓
$\eta(t)$	—	✓	✓	✓	✓	✓	—	—	✓
$m(t, Z)$	—	—	—	—	—	—	—	—	—
$\lvert\eta(j')\rvert$	✓	—	—	—	—	—	—	✓	—
$p_T(j')$	✓	✓	—	✓	—	—	—	—	—
$\Delta R(b, \ell_t)$	—	✓	—	✓	—	—	—	—	—
$\Delta R(j', \ell_t)$	✓	—	—	—	—	—	—	—	—
$\Delta R(t, Z)$	—	✓	✓	✓	—	✓	—	—	✓
$\Delta\eta(Z, j')$	—	✓	—	—	—	—	—	✓	—
ΔR between t and the closest lepton	—	✓	—	✓	—	—	—	—	—
ΔR between j' and the closest lepton	—	—	—	—	—	—	—	✓	—
$m_{3\ell}$	✓	—	—	—	✓	—	✓	—	✓
m_T^W	✓	✓	✓	—	—	—	—	—	✓
p_T^{miss}	✓	—	—	—	—	—	—	—	—
Lepton asymmetry	✓	—	—	✓	✓	—	—	✓	—
$\cos\theta_Z^*$	—	—	✓	—	—	✓	—	—	✓
Max. p_T among jet pairs	—	—	—	—	—	—	✓	—	✓
Max. DEEPJET discriminant	✓	—	—	—	—	—	—	—	—
b jet multiplicity	✓	—	—	—	—	—	—	—	—
Three-momenta of the three leading leptons	✓	—	—	—	—	—	—	—	—
Three-momenta of the three leading jets	✓	—	—	—	—	—	—	—	—
DEEPJET discriminants of the three leading jets	✓	—	—	—	—	—	—	—	—
Number of variables	33	11	8	8	6	7	4	7	10

Backup

Systematic uncertainties

[CMS-TOP-21-001]

	Source	Type	Correlation
Experimental	Integrated luminosity	Yield	Partial
	Trigger efficiency	Yield	—
	Pileup	Both	✓
	Lepton identification and isolation	Both	✓
	b tagging	Both	Partial
	Jet energy scale	Both	Partial
	Jet energy resolution	Both	—
	Missing transverse momentum	Both	—
	L1 ECAL inefficiency	Both	✓
Theoretical	PDF	Both	✓
	α_S	Both	✓
	ME scales μ_R, μ_F	Both	✓
	Signal SM cross sections	Yield	✓
	ISR and FSR	Both	✓
	Additional radiation	Shape	✓
Backgrounds	WZ normalization	Yield	✓
	VV(V) normalization	Yield	✓
	$t(\bar{t})X$ normalization	Yield	✓
	$X\gamma$ normalization	Yield	✓
	NPL normalization	Yield	✓
	NPL misidentification probabilities	Both	✓

Observables for inference

[CMS-TOP-21-001]

Fit configuration	Region					
	SR-tZq	SR-$t\bar{t}Z$	SR-Others	SR-$t\bar{t}Z$-4ℓ	CR WZ	CR ZZ
1D c_{tZ}	NN-c_{tZ}-tZq	NN-c_{tZ}-$t\bar{t}Z$				
1D c_{tW}	NN-c_{tW}-tZq	NN-c_{tW}-$t\bar{t}Z$				
1D $c^3_{\varphi Q}$	NN-$c^3_{\varphi Q}$-tZq	NN-$c^3_{\varphi Q}$-$t\bar{t}Z$	m_T^W	Counting experiments		
1D $c^-_{\varphi Q}$	NN-SM (tZq node)	NN-SM ($t\bar{t}Z$ node)				
1D $c_{\varphi t}$	NN-SM (tZq node)	NN-SM ($t\bar{t}Z$ node)				
2D and 5D	NN-5D-tZq	NN-5D-$t\bar{t}Z$				

Contribution of uncertainty groups

[CMS-TOP-21-001]

Source	c_{tZ}	c_{tW}	$c^{3}_{\varphi Q}$	$c^{-}_{\varphi Q}$	$c_{\varphi t}$
tZq normalization	<0.1	<0.1	1.2	0.1	0.8
$t\bar{t}Z$ normalization	0.6	<0.1	0.4	37	38
tWZ normalization	0.1	0.1	<0.1	0.7	2.1
Background normalizations	<0.1	<0.1	6.9	3.6	6.8
NPL background estimation	1.4	0.2	5.6	0.3	3.8
Jet energy scale	<0.1	<0.1	0.8	0.7	2.3
Jet energy resolution	<0.1	<0.1	<0.1	<0.1	1.4
p_T^{miss}	<0.1	<0.1	<0.1	<0.1	0.2
b tagging	<0.1	<0.1	0.9	2.0	0.3
Other (experimental)	<0.1	<0.1	1.6	0.8	0.6
Lepton identification and isolation	0.4	0.4	1.2	2.2	0.8
Theory	2.1	1.1	0.4	0.9	0.9

ATLAS Tagger: Hyperparameters 1

[ATL-PHYS-PUB-2022-039]

Model	Hyper-parameters	
hlDNN	Hidden Layers: 5 Nodes per Layer: 180 *Activation Functions: ReLU* *Kernel Initialization: glorot uniform* Learning Rate: 4×10^{-5} Batch Size: 250 Batch Normalization: not used	$\Rightarrow$ **Feed-forward NN**
DNN	Hidden Layers: 5 Nodes per Layer: 400 *Activation Functions: ReLU* *Kernel Initialization: glorot uniform* L1 Regularization: 2×10^{-4}, applied to all layers Learning Rate: 1.2×10^{-5} Batch Size: 250 Batch Normalization: applied before activation function for all layers except output layer	$\Rightarrow$ **Feed-forward NN**
EFN	Φ Hidden Layers: 5 Φ Nodes per Layer: 350 Latent Dropout: 0.084 F Hidden Layers: 5 F Nodes per Layer: 300 F Dropout: 0.036 *Activation Functions: ReLU* *Kernel Initialization: glorot normal* Learning Rate: 6.3×10^{-5} Batch Size: 350	$\Rightarrow$ **DeepSet NN**
PFN	Φ Hidden Layers: 5 Φ Nodes per Layer: 250 Latent Dropout: 0.072 F Hidden Layers: 5 F Nodes per Layer: 500 F Dropout: 0.022 *Acitvation Functions: ReLU* *Kernel Initialization: glorot normal* Learning Rate: 7.9×10^{-5} Batch Size: 250	$\Rightarrow$ **DeepSet NN**

ATLAS Tagger: Hyperparameters 2

[ATL-PHYS-PUB-2022-039]

Model	Hyper-parameters
ResNet 50	Bottom Layer: 7x7 2D convolution with strides (2, 2) and zero padding Number of Stages: 4 Blocks per Stage: (3, 4, 6, 3) *Block Type: bottleneck* Block Output Filters: (64, 128, 256, 512) *Activation Functions: ReLU* *Kernel Initialization: he uniform* Batch Normalization Momentum: 0.1 *Global Pooling: average* Initial Learning Rate: 1×10^{-2} *Scheduler: decrease learning rate by factor of 0.1 every 10 epochs* Batch Size: 256
ParticleNet	Φ Number of Stages: 3 Blocks per Stage: (3, 3, 3) Block Output Features: (64, 224, 384) k Nearest Neighbors: 18 Top Layer Nodes: 125 *Activation Functions: ReLU* *Kernel Initialization: glorot normal* Batch Normalization Momentum: 0.7 Global Pooling: max Learning Rate: 4.2×10^{-4} Batch Size: 250

$\Rightarrow$ **CNN**

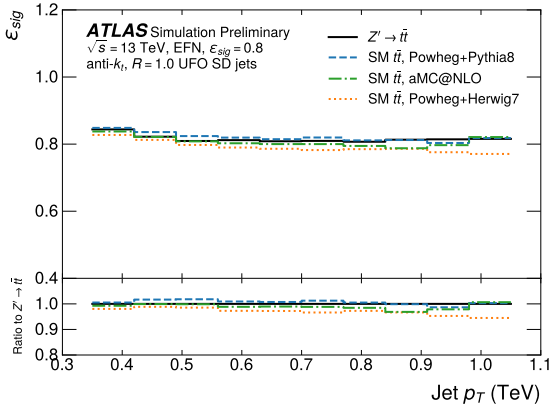
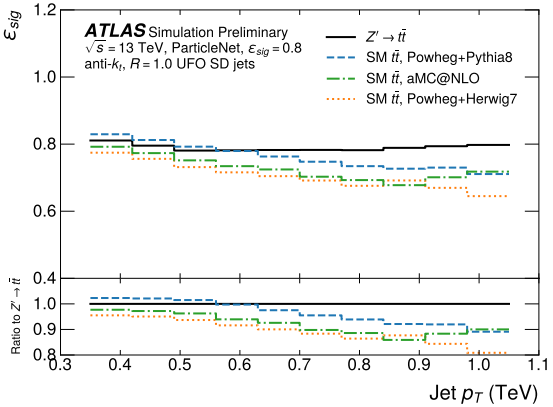
$\Rightarrow$ **Graph NN**

ATLAS Tagger: hIDNN input features

[ATL-PHYS-PUB-2022-039]

Quantity Type	Symbols	References
N-subjettiness	τ_1, τ_2, τ_3, τ_4	[66] [67]
k_t Splitting Scales	$\sqrt{d_{12}}$, $\sqrt{d_{23}}$	[15]
Generalized Energy Correlation Functions	ECF_1, ECF_2, ECF_3, C_2, D_2, L_2, L_3	[68] [69] [70]
Minimum Pair-wise Invariant Mass	Q_w	[15]
Thrust Major	T_m	[71]

Table 1: A listing of the 15 quantities used to train the baseline high level quantity tagger

ATLAS Tagger: MC dependency

[ATL-PHYS-PUB-2022-039]