Constituent Based Top-Quark Tagging with the ATLAS Detector

Kevin Greif, for the ATLAS Collaboration

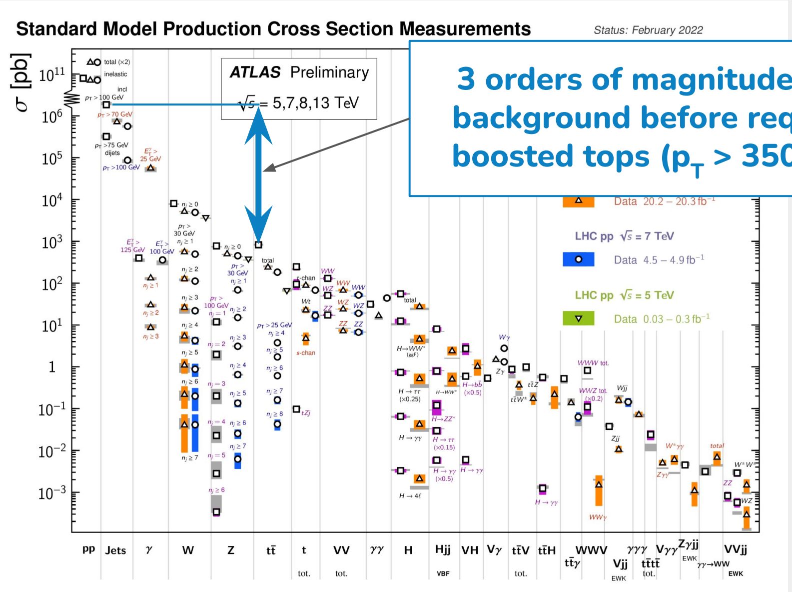
kgreif@uci.edu

Boosted Top Quark Tagging

Isolate boosted top jets -> Interesting Physics!

- **Precision SM measurements**
- **Heavy resonance searches**

But difficult to separate tops from light quark / gluon background



3 orders of magnitude more background before requiring boosted tops (p_T > 350 GeV)

"Easy" solution: Use ML to solve classification task

- **Exploit differences in substructure**
- **Powerful classification algorithms have already been developed**

"A robot looking at particle collisions, oil painting"

A New Class of Top Taggers

ATLAS approach in Run 2:

1. Calculate a set of *jet level quantities*
   a. Splitting scales, subjettiness, etc.
2. Combine using ML classifier

New approach*:

1. Combine *constituent level information* with (larger and more complex) ML classifier

Jet level quantity information ⊂ Constituent information

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

Pheno studies (1902.09914) show great performance...but do these results translate to a realistic contexts?

- DELPHES vs. realistic detector sim.
- Complex jet reconstruction

*CMS has already put the new approach to effective use (2004.08262)

Simulation Samples and Jet Reconstruction

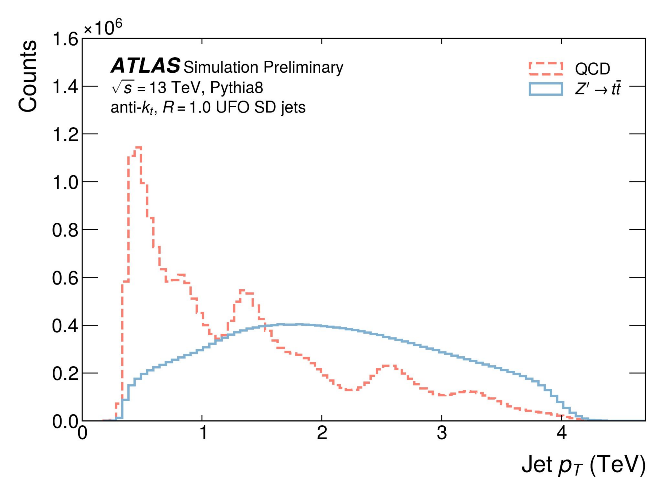
- Signal boosted tops obtained from simulated Z'→ttbar events
  - Use both leading and sub-leading jets
  - Cross section reweighted to populate kinematic region of interest p_T ~ 0.35-4 TeV
- Background light quark / gluon initiated jets obtained from dijet events
- All events simulated at LO with Pythia8
- Jets reconstructed from UFO inputs, using anti-kt algorithm w/ R=1.0
  - Soft-drop grooming
  - CS+SK pileup mitigation
- "Contained" top jets obtained by placing requirements on matched truth jet

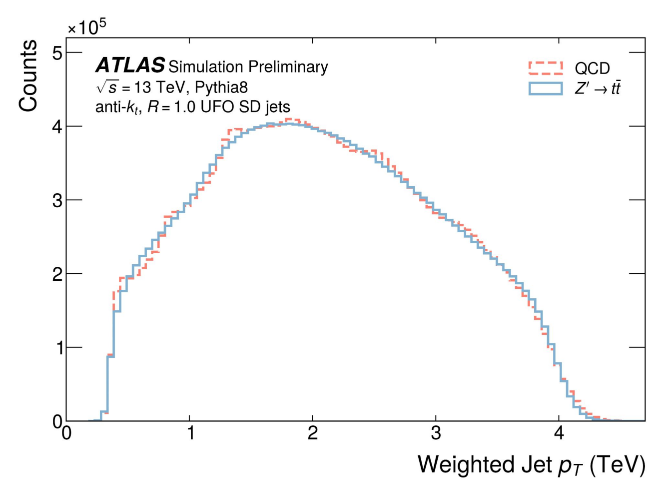
Jet Requirements	Top quark jet requirements
Jet $\|\eta\| < 2.0$	$dR(\text{jet, truth jet}) < 0.75$
Jet $p_{\mathrm{T,truth}} > 350$ GeV	$dR(\text{truth jet, top parton}) < 0.75$
Number of constituents ≥ 3	Ungroomed truth jet mass > 140 GeV
Jet mass > 40 GeV	Number ghost associated b-hadrons ≥ 1
	Truth jet $\sqrt{d_{23}} > \exp(3.3 - 6.98 \times 10^{-4} \times \text{truth jet } p_{\mathrm{T}})$

"A group of UFOs riding a sunbeam as a Renaissance oil painting"

Jet p_T and Training Weights

- Without event weights, background jet sample contains unphysical p_T spectrum
- Derive weights that match background p_T spectrum to signal
  - First order measure to prevent background-sculpting
  - Produces reasonable distribution of weights → stable training
- Weights are applied to loss function in tagger training

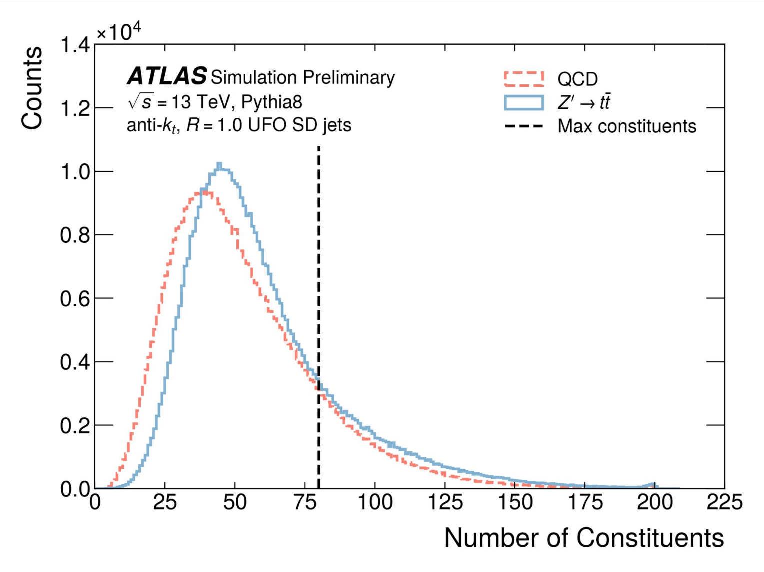




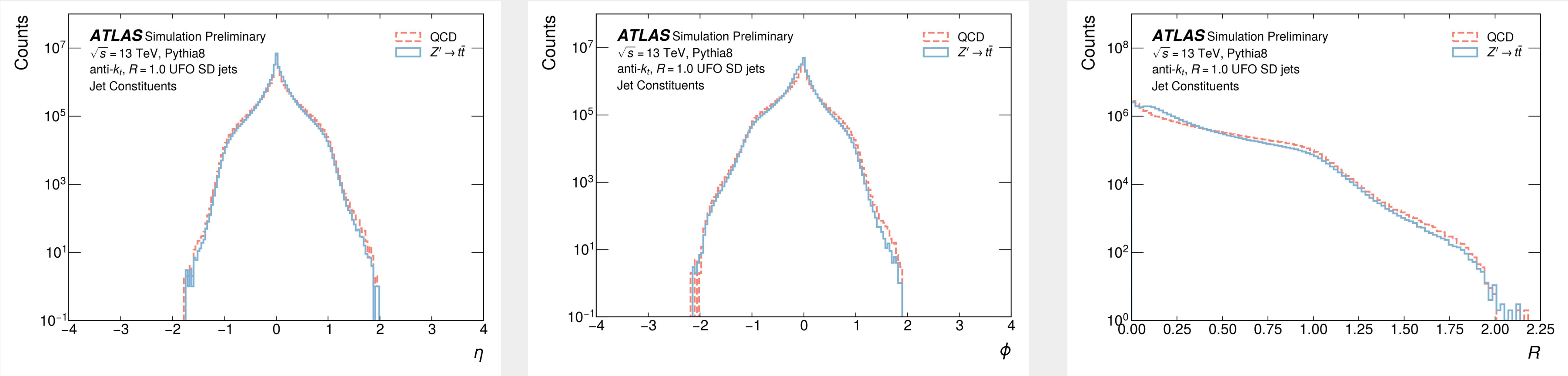
Constituent Level Pre-Processing

- Pre-process constituent level information to exploit known symmetries
- Consider at most the 80 leading constituents in each jet
- Then build 7 constituent level inputs

"An overflowing bowl containing UFOs, oil painting"



Constituent Level Pre-Processing

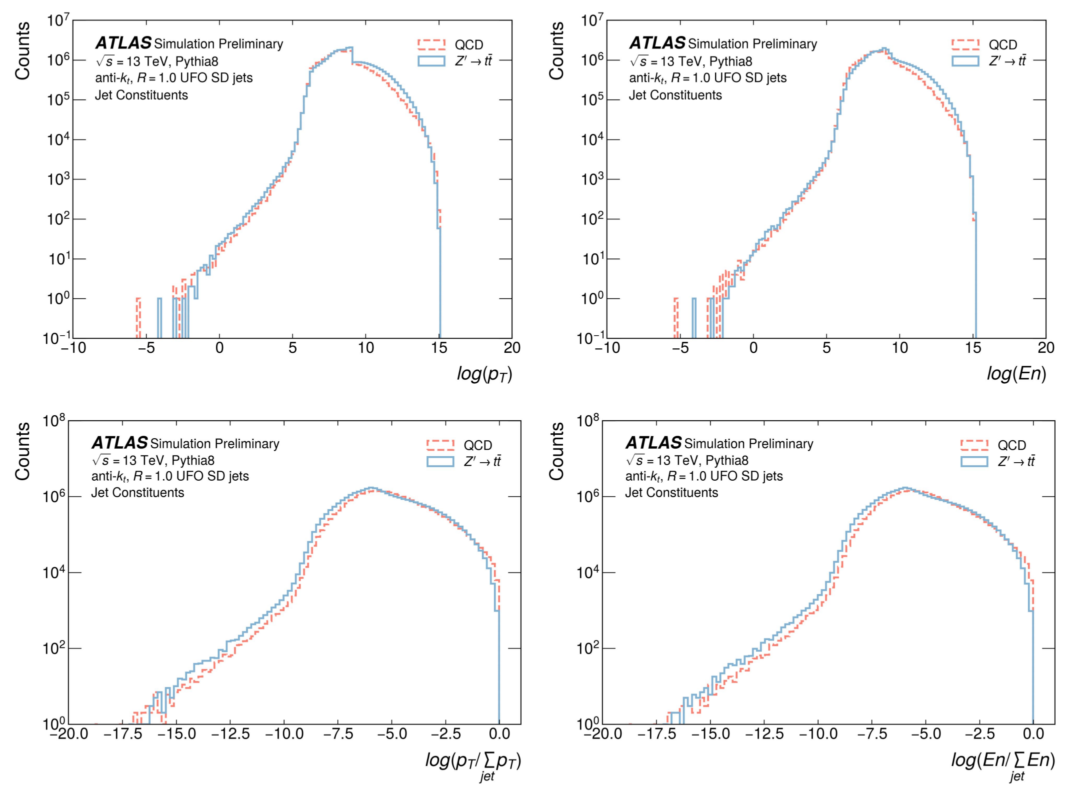


$$R = \sqrt{\eta^2 + \phi^2}$$

Angular Pre-Processing:

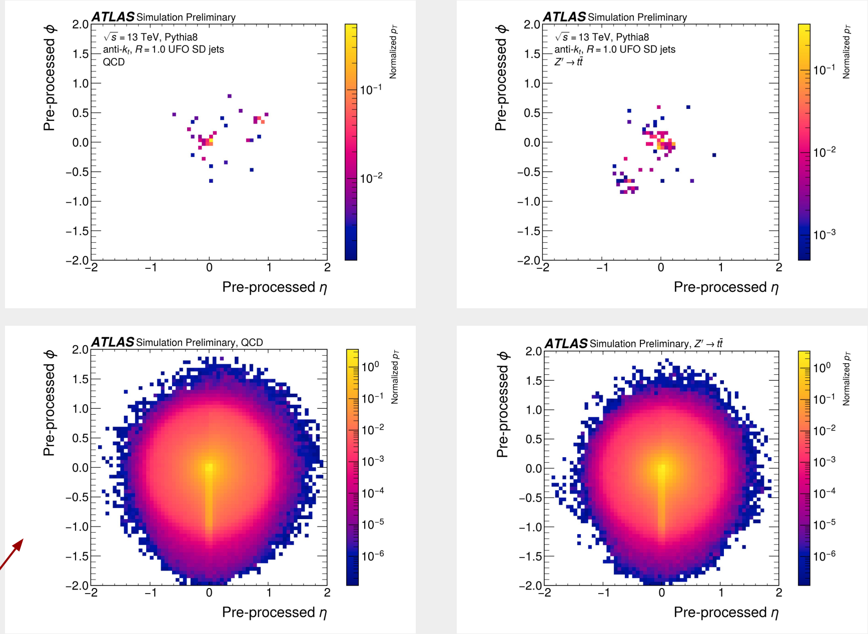
1. Center highest p_T constituent
2. Rotate so that 2nd highest p_T constituent sits on the negative ϕ-axis
3. If needed, reflect jet about ϕ-axis such that 3rd highest p_T constituent sits in the positive η half-plane

Constituent Level Pre-Processing



Top-Quark Taggers

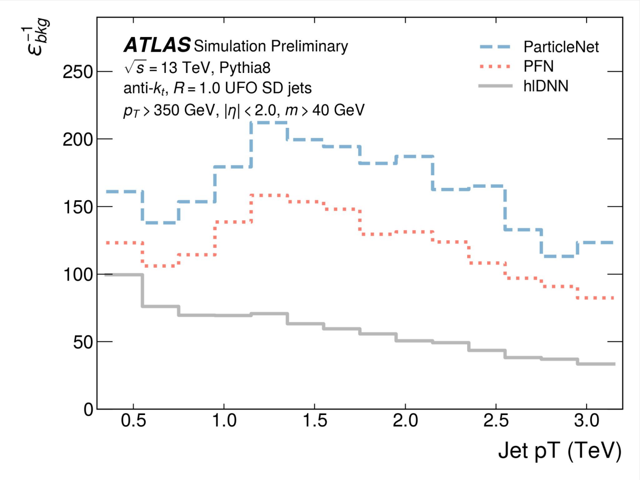
1. High Level Quantity Baseline
   a. Tagger used by [ATLAS for Run 2]
   b. High level quantities used are shown in backup
2. Densely Connected Neural Network
   a. Flatten jet constituents into vector
   b. Fully connected layers, ReLu activations, etc.
3. Energy Flow / Particle Flow Networks
   a. https://arxiv.org/abs/1810.05165
   b. Deep sets networks
4. ResNet 50
   a. https://arxiv.org/abs/1512.03385
   b. CNN architecture interprets data as an image
5. ParticleNet
   a. https://arxiv.org/abs/1902.08570
   b. Graph network used extensively by CMS



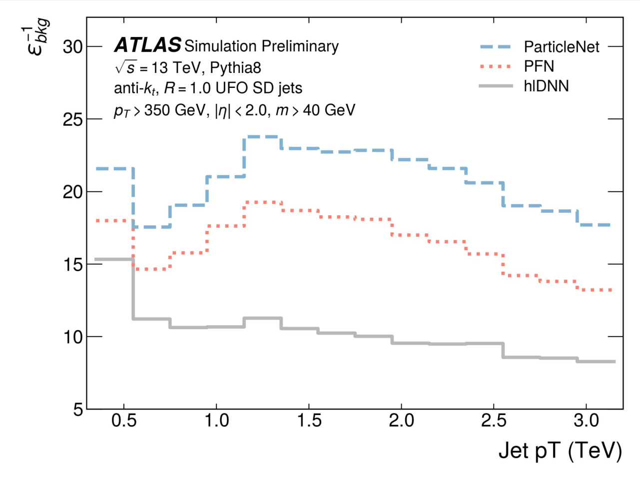
Background Signal

All models trained to minimize cross entropy loss w/ Adam optimizer

Some constituent based taggers work very well…



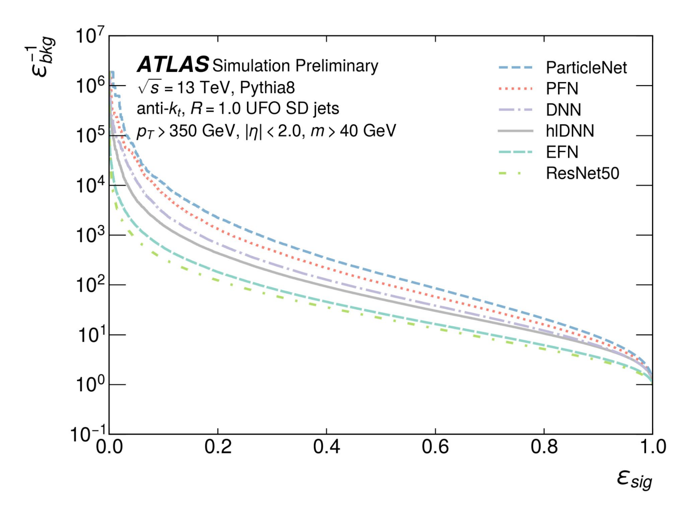
$\epsilon_{sig} = 0.5$



$\epsilon_{sig} = 0.8$

- Particle Net / PFN achieve ~2-3x improvement in background rejection across kinematic range
- Constituent based tagger’s performance peaks in the 1-2 TeV range

...but others underperform on an ATLAS data set!



ResNet 50 and EFN underperform relative to expectation from pheno studies

Wouldn't it be nice if we could just train taggers in a realistic context from the beginning??

Model	AUC	ACC	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.5$	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.8$	# Params	Inference Time
ResNet 50	0.885	0.803	21.4	5.13	1,486,209	9 ms
EFN	0.901	0.819	26.6	6.12	1,670,451	4 ms
hlDNN	0.938	0.863	51.5	10.5	93,151	3 ms
DNN	0.942	0.868	67.7	12.0	876,641	3 ms
PFN	0.954	0.882	108.0	15.9	689,801	4 ms
ParticleNet	0.961	0.894	153.7	20.4	764,887	38 ms

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$)			#Param
			single	mean	median	
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

https://arxiv.org/abs/1902.09914

All MC samples used for training / testing of taggers in this study are now **publicly available!**

Data: http://opendata.cern.ch/record/15013
Documentation: https://gitlab.cern.ch/atlas/ATLAS-top-tagging-open-data

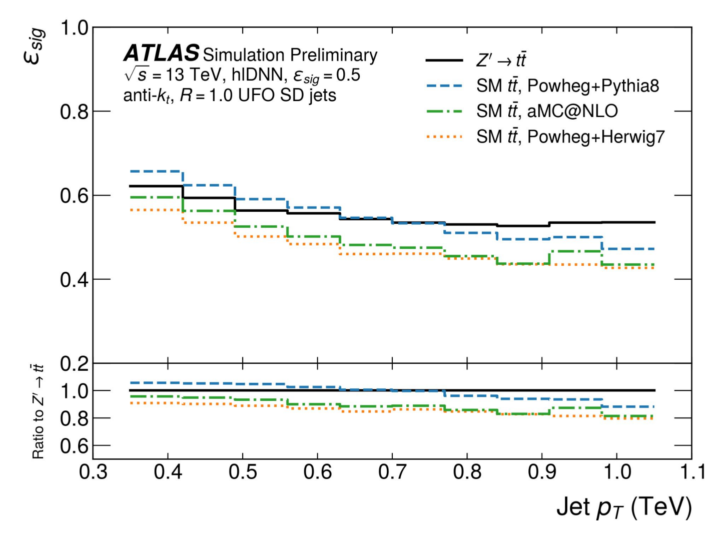
QCD Modeling Dependence

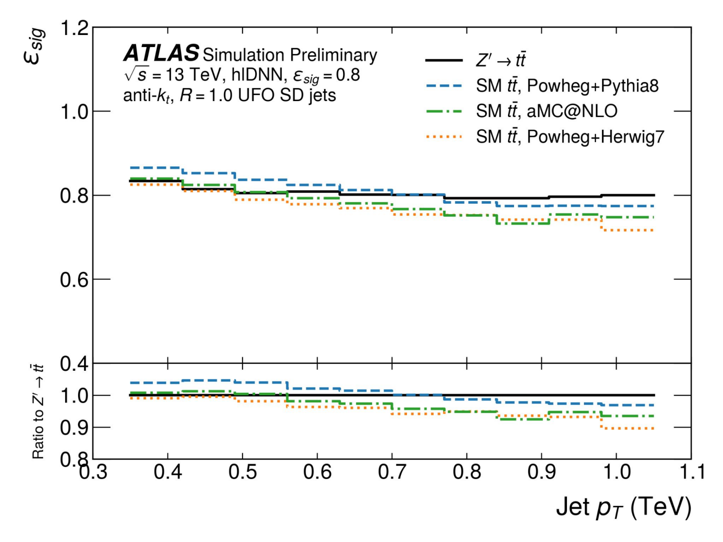
- Evaluate trained taggers on events simulated with alternative MC generators to assess model dependence of constituent based taggers
  - SM ttbar production, semi-leptonic decay mode
  - Note taggers are trained using Z' -> ttbar samples
- Matrix elements evaluated at NLO
- Attempt to isolate top jets by requiring dR(top, anti-top) > 2.0 at parton level
- Comparing signal efficiency between alternative and nominal samples shows dependence of tagger performance on QCD modeling

3 alternative samples:

1. Powheg + Pythia 8
2. Powheg + Herwig 7
3. MadGraph_aMC@NLO + Pythia 8

hlDNN Model Dependence



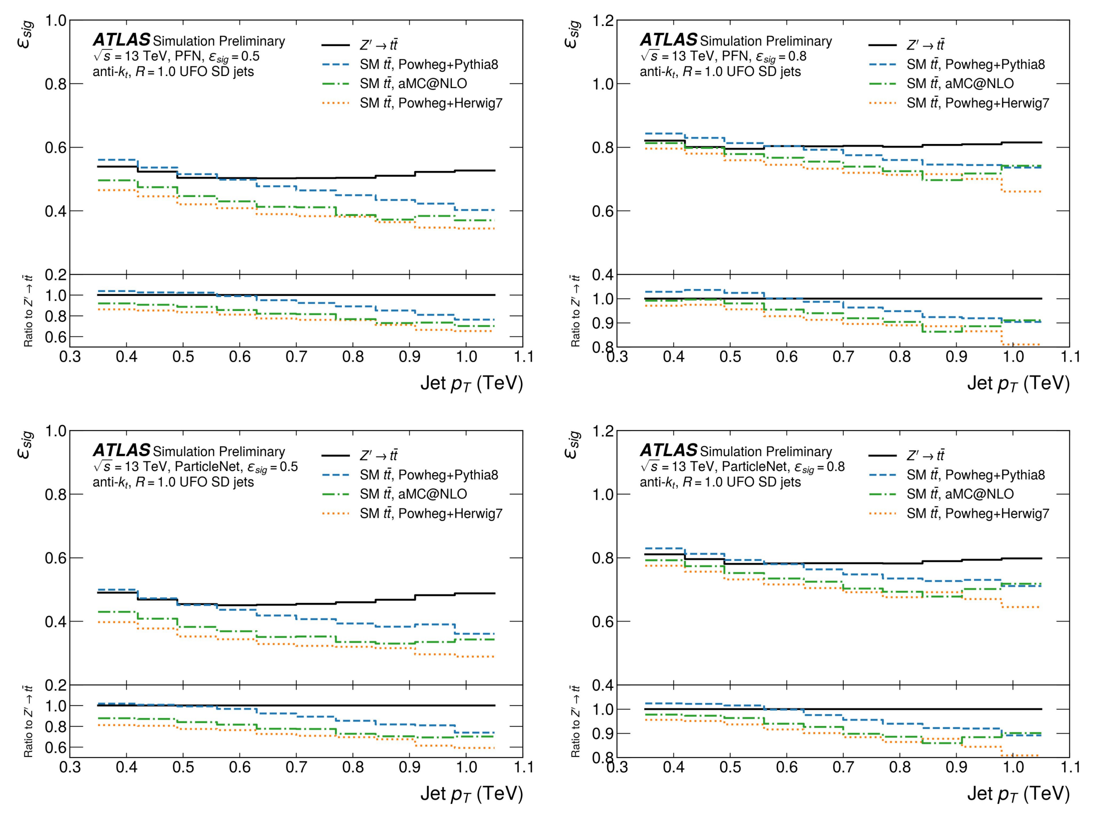


Decreasing performance w/ pT is common feature of all SM ttbar samples regardless of parton shower model. Suggests this results from ISR/FSR differences

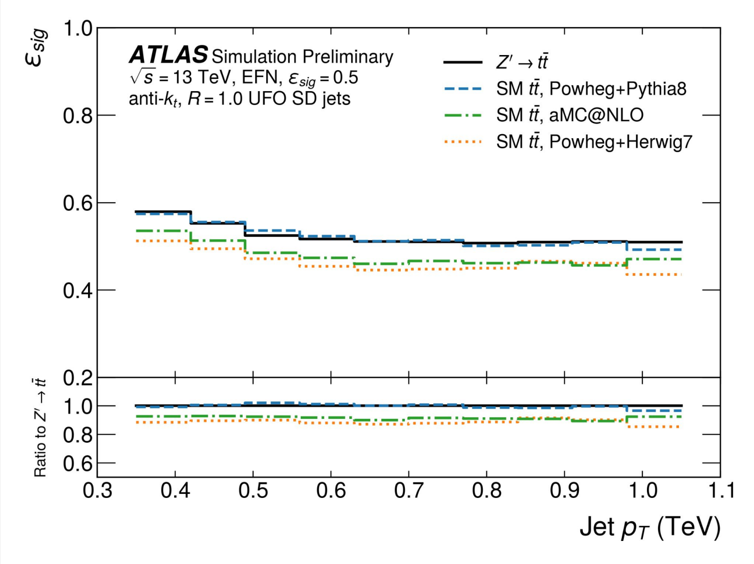
Constituent Based Tagger Model Dependence

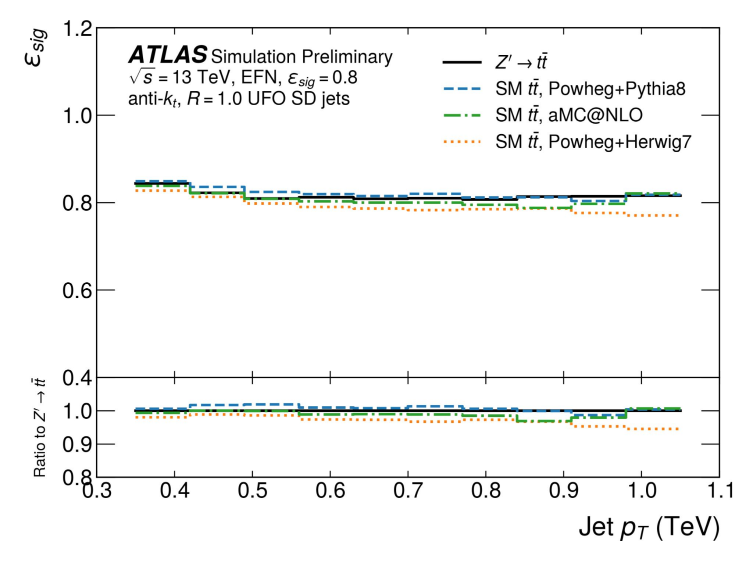
- PFN and Particle Net show increased model dependence
- Exact cause is an open question!
- QCD modeling dependence will contribute to systematic uncertainties in SF measurements and physics analyses!
- Also implies a suboptimal tagger

AUC is not the only performance metric!



EFN Model Dependence





EFN is less sensitive to soft radiation by definition. Results in less model dependence than even the high level quantity baseline!

Conclusions

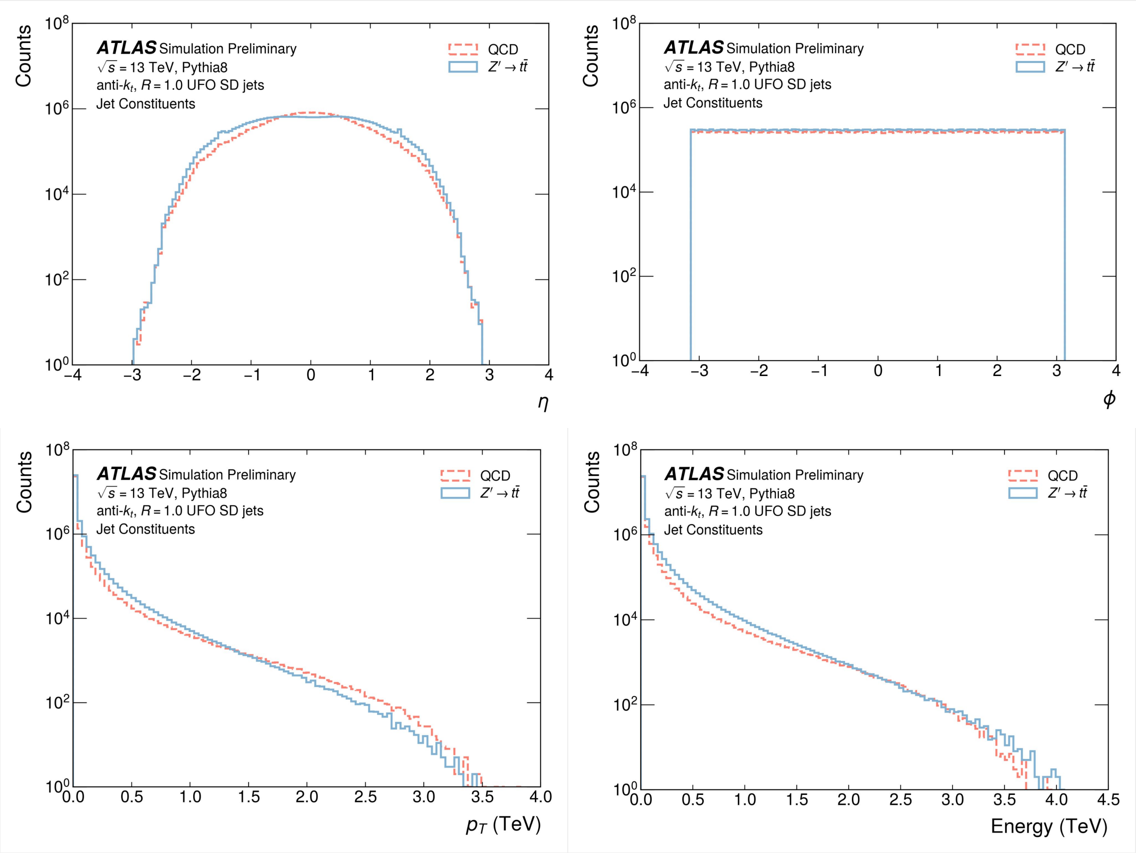
- Some constituent-based top taggers (PFN, ParticleNet) show significantly stronger performance than the baseline
- Others (ResNet 50 and EFN) underperform relative to pheno studies
  - **Realistic detector simulation and jet reconstruction matters!**
- Taggers differ in their sensitivity to QCD modeling.
  - **AUC is not the only performance metric that should be considered!**
- Datasets publicly available!
  - Data: http://opendata.cern.ch/record/15013
  - Documentation: https://gitlab.cern.ch/atlas/ATLAS-top-tagging-open-data
- Results in ATLAS public note

Public Data:

“A robot performing experiments in a mountainous landscape, oil painting”

Backup

Raw Constituent Level Distributions



High Level Quantities

Quantity Type	Symbols
N-subjettiness	$\tau_1, \tau_2, \tau_3, \tau_4$
k_t Splitting Scales	$\sqrt{d_{12}}, \sqrt{d_{23}}$
Generalized Energy Correlation Functions	$ECF_1, ECF_2, ECF_3, C_2, D_2, L_2, L_3$
Minimum Pair-wise Invariant Mass	Q_w
Thrust Major	T_m

Hyper-parameter Tuning Procedure

- Hyperopt tree of parzen estimators algorithm used to suggest new hyper-parameter configurations
    - For hlDNN, DNN, EFN, PFN: O(100) trainings run, model which achieved lowest validation loss selected as optimal
    - ResNet 50 and ParticleNet training times limited tuning to ~30 trainings
- For ResNet50 and ParticleNet, Asynchronous Hyperband Scheduler (ASHA) used to terminate poorly performing configurations