Jet Substructure/Tagging in ATLAS

Robert Les,
On behalf of the ATLAS collaboration

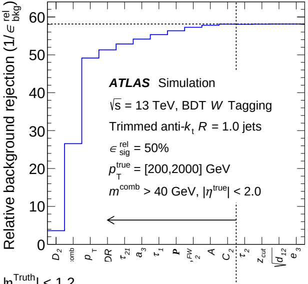
COMETA WG3
April 22, 2023

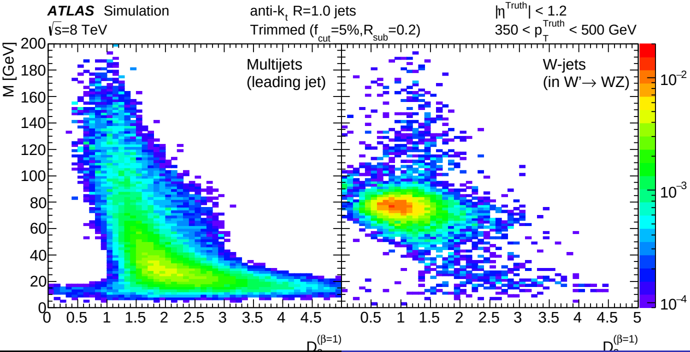
Substructure Basics

Substructure of large-R jets is pivotal to studying $W/Z/t/h$ hadronic jets against the huge QCD background at the LHC

Studied historically in ATLAS:

- Dedicated measurement observables
- "Taggers" to improve general measurements



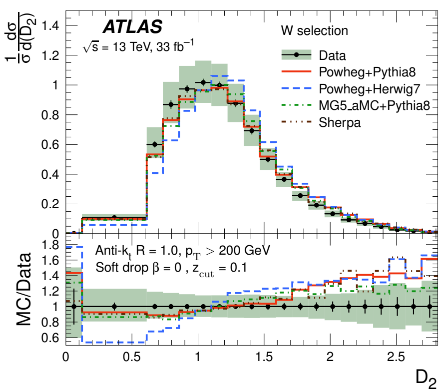


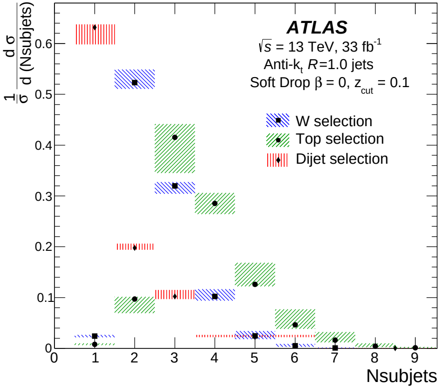
Substructure measurements in boson topologies

Several substructure measurements in ATLAS in QCD jets. Some examples:

- Softdrop mass and substructure
- Lund-plane measurement
- Energy-energy correlations

Only one result specifically in W/t enriched jet selections at 13 TeV

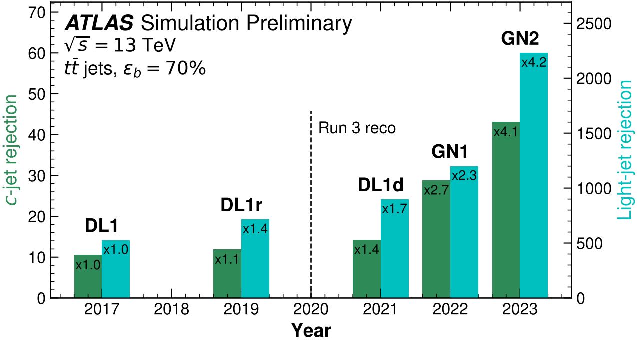




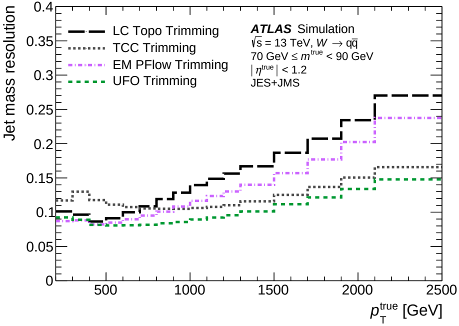
Tagger Progress

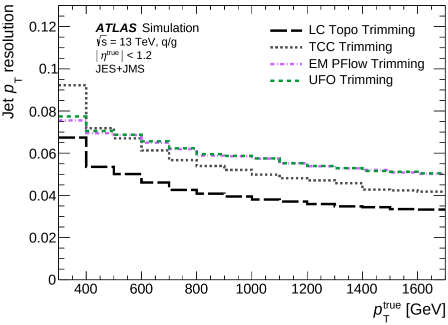
Historically taggers have:

1) High-level taggers on jet substructure observables
   - Robust and easy to interpret at analysis/theory level
2) Machine learning taggers on jet-substructure
   - Non-trivially combine several observables for better discrimination
3) Machine learning taggers on low-level inputs
   - Lose interpretativeness for performance



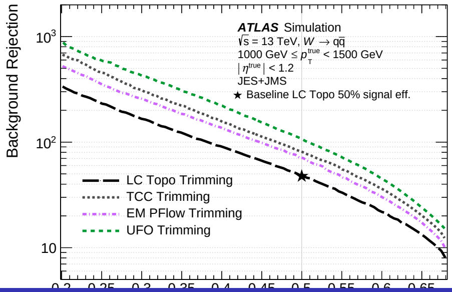
Small aside: Jet collections



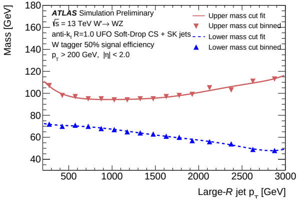


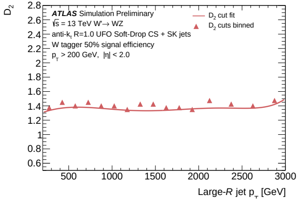
Tagger performance deeply connected to jet definition!

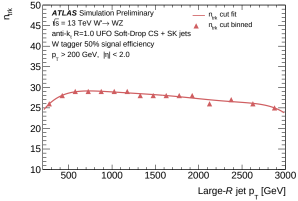
- Over Run2 moved from LCTopo $\rightarrow$ TCC $\rightarrow$ UFO jets
- Better mass+substructure resolution
- Factor 2 gains from this alone



Historic taggers: W/Z-tagger

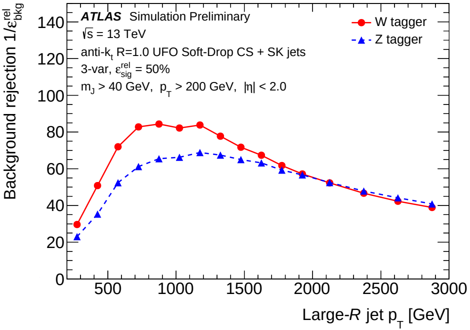






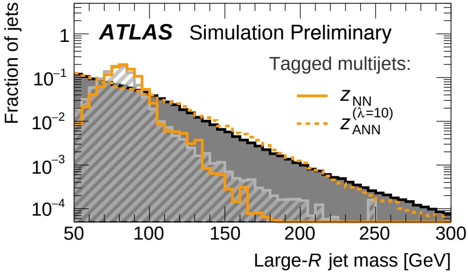
Baseline W/Z-taggers provided at fixed $50/80\%$ signal efficiency

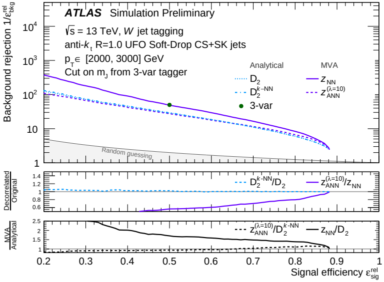
- Fixed-cut 3-variable tagger: Jet mass, $D_2^{\beta=1}$, $N(tracks)$
- Flat efficiency as a function of jet p_{T}
- Different performance for transverse/longitudinal bosons



Similar to top-tagger experience, developed a DNN W-tagger over 10 substructure variables

- But obviously the DNN learns that the mass is very good discriminant
  - Sculpts background to look like signal!
  - Difficult to use sideband regions in analysis
- Train against a second "adversarial" network to force network to decorrelate the feature: Loss function $L = L_{classifier} - \lambda L_{adv}$
- Decorrelated tagger similar to cut-based. Correlated almost 10x better

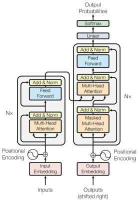


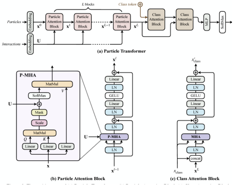


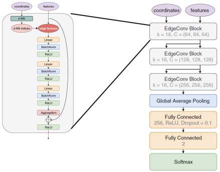
ML-era: Architectures

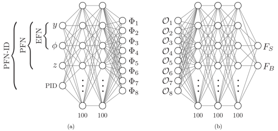
Machine-learning/pheno community is developing faster then we can test on data!

- Train directly on lowest level objects: jet constituents themselves
- Various advanced architectures suggested
  - Normal dense neural networks
  - Energy/Particle Flow networks: General decomposition of IRC-safe observables
  - ResNet50: CNN Architecture representing jet as image
  - ParticleNet: Point-cloud represented data
  - ParticleTransformer: Transformer based architecture
  - GN2X: Graph network with auxiliary tasks
  - LundNet: Graph on declustering history
  - PELICAN: Lorentz invariant network









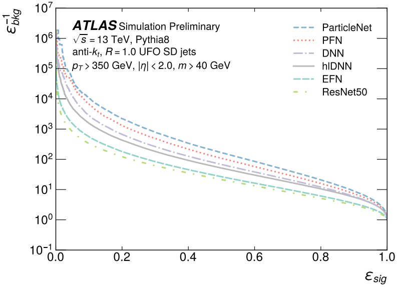
*Nice semi-recent summary: arXiv:1902.09914

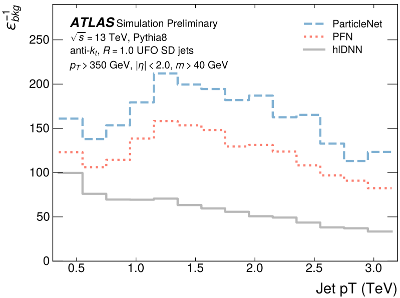
Constituent based top tagger outperform high-level ones:

- Another factor 2/3 improvement!
- Unclear why ResNet/EFN under-perform w.r.t pheno
  - Real simulation studies important!

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

Open-data and documentation available!





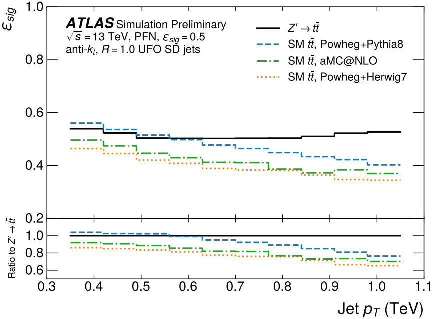
Aside: Constituent top-tagging modelling

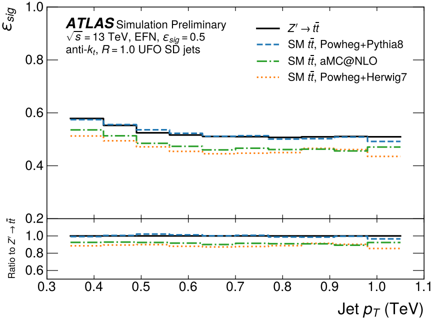
But as we provide lower-level information the taggers can become more generator dependent

- Best performant network would have double the modeling uncertainty!
- Also in-progress looking into effect of tracking/cluster uncertainties

Theory-driven taggers may help us though

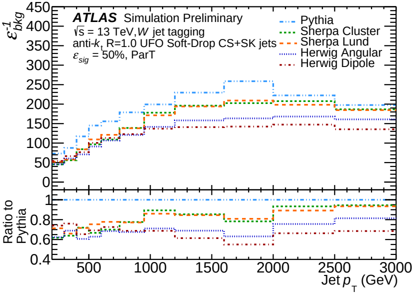
- EFN has less model dependency then baseline tagger since focus on IRC-safe observables





Similar results for constituent based W-tagger

- Better performance
  - Transformers at top
- But large modeling uncertainty



Model	AUC	ACC	ε_{bkg}^{-1} @ ε_{sig} = 0.5	ε_{bkg}^{-1} @ ε_{sig} = 0.8	# Params	Inference Time
EFN	0.920	0.835	35.1	7.95	56.73k	0.065 ms
PFN	0.931	0.853	44.7	9.50	57.13k	0.11 ms
ParticleNet	0.933	0.826	46.2	9.76	366.16k	0.36 ms
ParticleTransformer	0.951	0.880	77.9	14.6	2.14M	0.28 ms

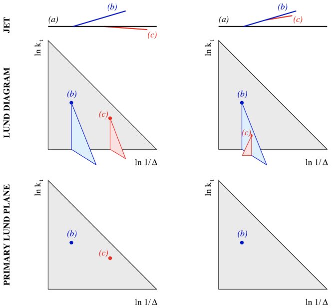
Lund Jet Plane

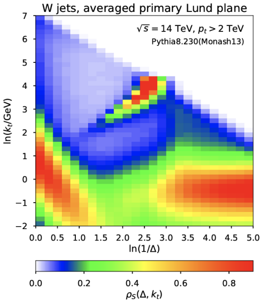
Can recluster jet with C/A algorithm and each split can be represented on lund-plane based on:

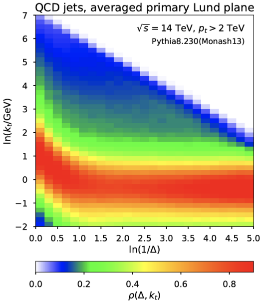
transverse momentum k_T, angle Δ, and momentum fraction z

Variables can distinguish splitting type:

- ISR: low $\ln(1/\Delta)$
- Non-perturbative: low $\ln(k_T)$
- hard colinear: high $z \rightarrow$ top graph edge





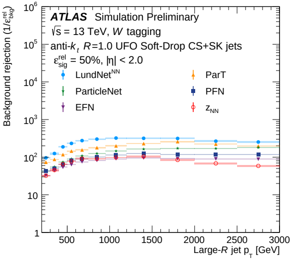


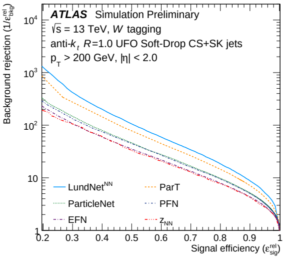
Trained a graph-neural network inspired by ParticleNet on the Lund-plane points:

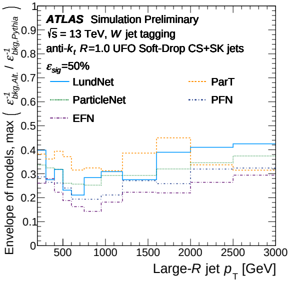
- Include mass decorrelation component
- This physics-driven tagger is most performant!

But also see large modelling dependency

- Can clip out regions of Lund-plane to remove IRC-effects?





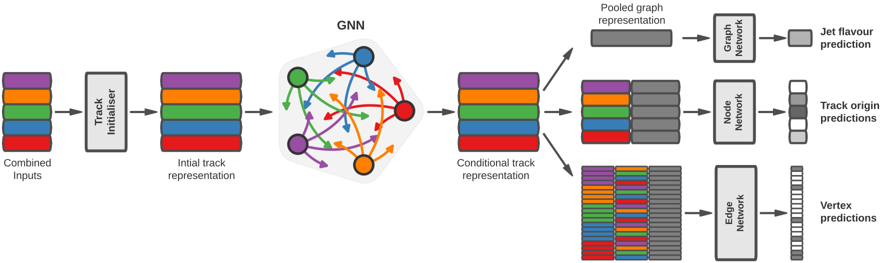


$h/Z \to bb$ taggers

The development of h-taggers started from a different point

- Early studies on best way to count number of b-hadrons in reclustered subjets
- Now using transformer architectures with flavour sensitive information of tracks within jet
- GN2X network is top-of-the-line
  - Has auxillary tasks to also classify individual tracks and match vertices

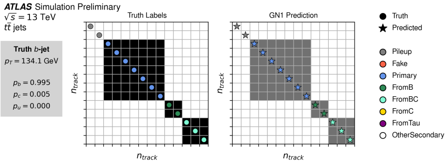
Jet Input	Description
p_{T}	Large-R jet transverse momentum
η	Signed large-R jet pseudorapidity
mass	Large-R jet mass
Track Input	**Description**
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of track relative to the large-R jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the large-R jet ϕ
d_0	Closest distance from track to primary vertex (PV) in the transverse plane
$z_0 \sin\theta$	Closest distance from track to PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0 \sin\theta)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
subjetIndex	Integer label of which subjet track is associated to (GN2X + Subjets only)
Subjet Input	**Description** (Used only in GN2X + Subjets)
p_{T}	Subjet transverse momentum
η	Subjet signed pseudorapidity
mass	Subjet mass
energy	Subjet energy
$\mathrm{d}\eta$	Pseudorapidity of subjet relative to the large-R jet η
$\mathrm{d}\phi$	Azimuthal angle of subjet relative to the large-R jet ϕ
GN2 p_b	b-jet probability of subjet tagged using GN2
GN2 p_c	c-jet probability of subjet tagged using GN2
GN2 p_u	light flavour jet probability of subjet tagged using GN2
Flow Input	**Description** (Used only in GN2X + Flow)
p_{T}	Transverse momentum of flow constituent
energy	Energy of flow constituent
$\mathrm{d}\eta$	Pseudorapidity of flow constituent relative to the large-R jet η
$\mathrm{d}\phi$	Azimuthal angle of flow constituent relative to the large-R jet ϕ

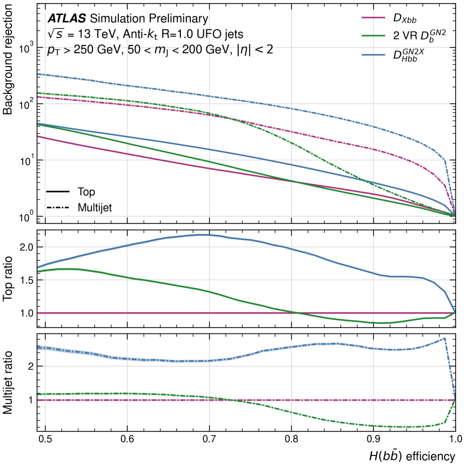


Huge gains with GN2 using the full tracking information (right)

Auxillary task perfromance very good (left)

- Example from small-R b-tagging

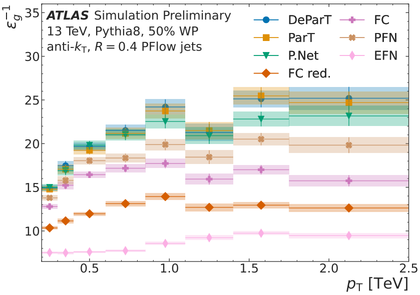


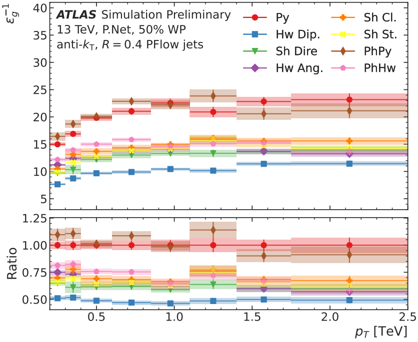


Constituent q/g-tagging

Similar studies also in $R = 0.4$ q/g-tagging!

- Extremely interesting in the context of VBS/VBF measurements
- Harder problem then $W/t/h$-tagging
- All previous conclusions follow here
- Currently limited by capability to calibrate





The tagger approach to substructure limited by capability to calibrate

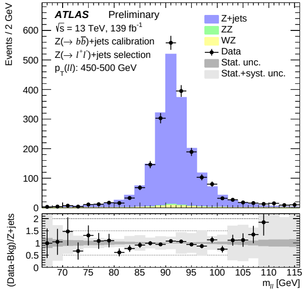
- Train on simulation, but also apply to data

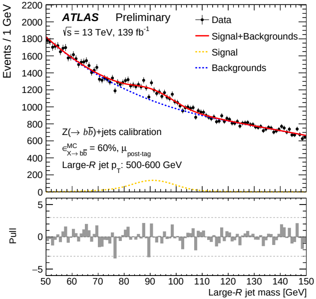
Often use top-down approach, where calibrate to a well known resonance

- $t\bar{t}$, $V+\text{jets}$, $V+\gamma$
- Can measure QCD jet rejection in multijet topologies

Difficulties can arise:

- No easy SM resonance: e.x. calibrating to h
- Limited probe statistics at high-p_{T}
- Limited trigger/selection capabilities at low-p_{T}



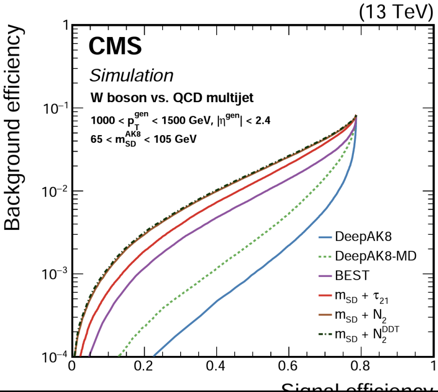


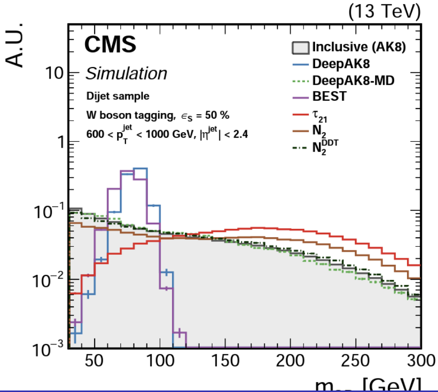
Future Directions: Multi-classification tagger

CMS is ahead of us in many ways

Their DeepAK8-MD tagger is a constituent-based, multi-class, mass-decorrelated tagger

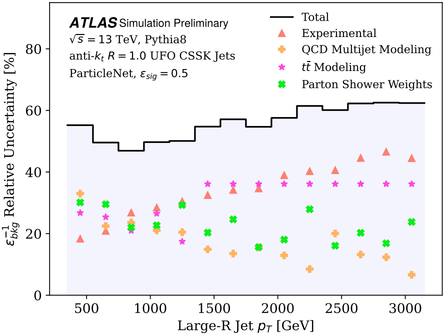
- Particle-net only looked at recently by ATLAS
- ATLAS has not looked at simultaneous multi-classification
- Deployed/calibrated in many analyses

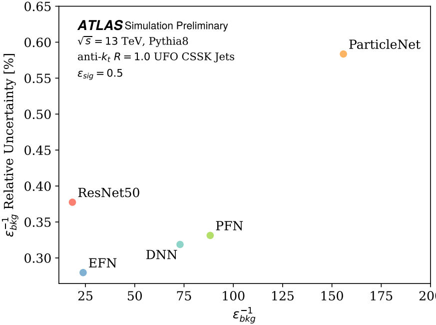




New results evaluating approximate bottom-up uncertainties on top-taggers

- Better taggers have higher uncertainties
- Want to develop want ways to break this trend





Summary

Current substructure studies:

- Rich history of $W/t/h$-taggers and now q/g-taggers
  - Trend of exponential gains
- Bottleneck for deployment is often the calibration
- Not currently large literature of ATLAS substructure within resonant jets

Future taggers:

- Lot's of R&D in jet-constituent level taggers
  - General trend: Better performance but more model dependency
- Investigating ways of breaking this model dependency
  - Hope open data programs to wider community will help with this

BACKUP