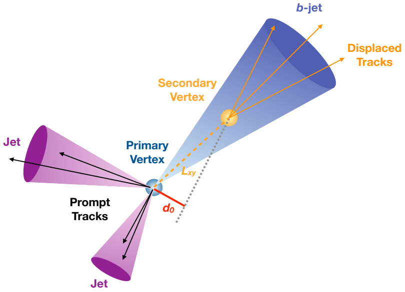
Recent Developments in Jet Flavour Tagging at ATLAS

ICHEP 2022

8th July 2022

Andy Chisholm (University of Birmingham)
on behalf of the ATLAS collaboration

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 714893 (ExclusiveHiggs)

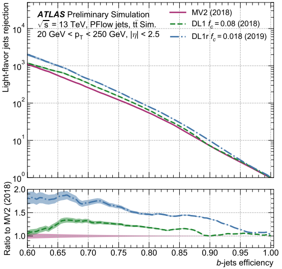


Tracking information → "Low level" taggers

- **Impact Parameter** - IP2D/3D [1] (likelihood), RNNIP [2] (recurrent neural network)
- **Secondary / Tertiary Vertices** - SV1 [1,3] (multi-track vertex finder)
- **Topological Multi-vertex** - JetFitter [4] ($b \to c$ decay chain, Kalman filter)

"Low level" tagger outputs → "High level" taggers

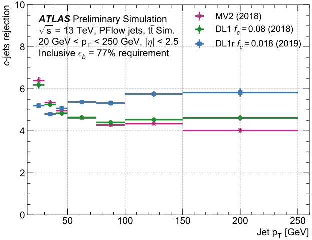
- Single discriminant for physics analysis, classifies jet flavour based on heavy flavour hadron content
- **MV2** [1] - Based on a boosted decision tree, extensively used in earlier Run 2 analyses
- **DL1** [1] - Based on deep neural network, **DL1r** variant (includes RNNIP input) baseline for "legacy" Run 2 and early Run 3 analyses

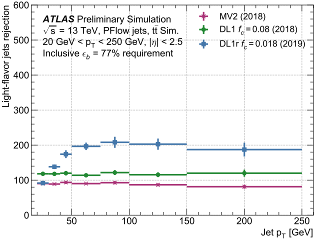


[1] EPJC 79 (2019) 970, [2] ATL-PHYS-PUB-2017-003, [3] ATL-PHYS-PUB-2017-011, [4] ATL-PHYS-PUB-2018-025

"High level" tagger discriminants can be defined for both *b*-tagging and *c*-tagging, operating points (OP) are defined for use in physics analyses (e.g. fixed *b*-jet eff.)

Example → 77% *b*-jet efficiency OP





Performance of OPs measured in data to derive corrections for simulated events (efficiency scale factors), baseline methods include:

***b*-jet efficiency** EPJC 79 (2019) 970

- Likelihood-based method with a sample of di-leptonic $t\bar{t}$ events

***c*-jet mis-tag rate (efficiency)** EPJC 82 (2022) 95

- Likelihood-based method with a sample of semi-leptonic $t\bar{t}$ events $(W \to c\{s, d\})$

Light flavour jet mis-tag rate FTAG-2021-002 and ATLAS-CONF-2018-006

- "Negative tag" method with a sample of $Z +$ jet events

ATLAS is pursuing a broad programme of new developments in flavour tagging, spanning new algorithms, calibration methods, simulation improvements and analysis tools, a selection of the most recent will be highlighted in this talk:

Alternative (non-$t\bar{t}$) calibration methods with multi-jet events

- Measurements of b-tagging efficiencies with the $p_{\mathrm{T}}^{\mathrm{rel}}$ method (ATL-PHYS-PUB-2022-025)
- Measurements of b-tagging efficiencies for very high p_{T} jets (ATL-PHYS-PUB-2022-010)

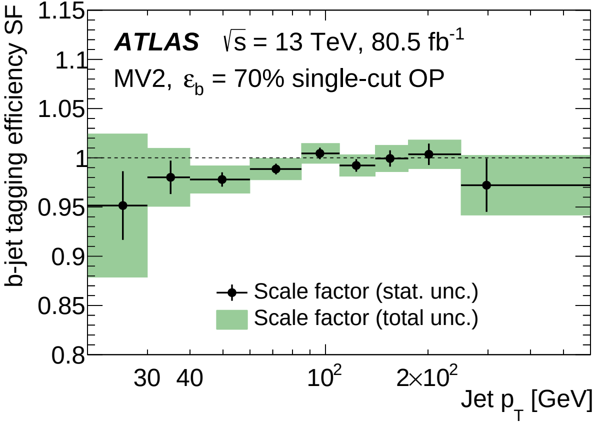
New algorithm developments

- Improved impact parameter-based tagging with Deep Sets (FTAG-2021-004)
- Flavour tagging with graph neural networks (ATL-PHYS-PUB-2022-027)

New analysis tools (not discussed directly in this talk, due to time constraints)

- Novel eigenvector decomposition approach to correlate flavour tagging uncertainties among physics analyses (ATL-PHYS-PUB-2022-024)
- Simulation-based extrapolation of b-tagging uncertainties for variable radius track jets at high p_{T} (FTAG-2022-001)
- New procedure for the re-weighting of heavy flavour hadron production fractions in ATLAS simulated event samples

Further information in the five ATLAS flavour tagging contributions to the ICHEP 2022 poster session this evening, please consider visiting!

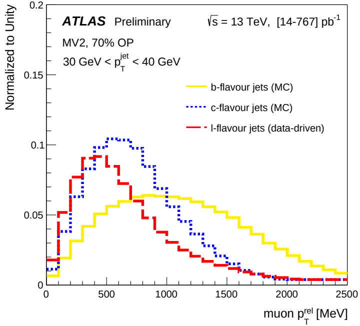


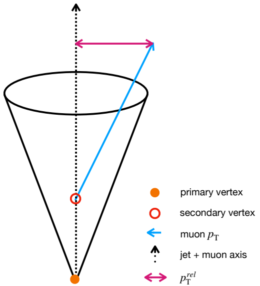
The well established likelihood-based approach to measure *b*-jet tagging efficiencies in data using $t\bar{t}$ events suffers from two main limitations:

- **1** - Large systematic uncertainties at very low jet p_{T} (from jet energy scale uncertainties)
- **2** - Large statistical uncertainties at high jet p_{T} and limited reach, with no measurements beyond $p_{\mathrm{T}} > 600$ GeV

The p_T^{rel} method, applied to multi-jet events, offers a valuable alternative approach to the baseline b-tagging efficiency measurement based on $t\bar{t}$ events

- Method **exploits the large masses of the b-hadrons**, relative to the c- and light hadrons, to identify b-jets





- Based on sample of jets geometrically matched to muons to **target semi-leptonic b-hadron decays**

p_T^{rel} strongly correlated with muon momentum in decaying hadron's rest frame $\rightarrow$ sensitive to the hadron's mass $\rightarrow$ jet flavour classification

Measuring b-tag efficiencies with the p_T^{rel} method (ATL-PHYS-PUB-2022-025)

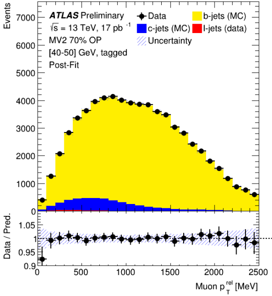
Event Selection

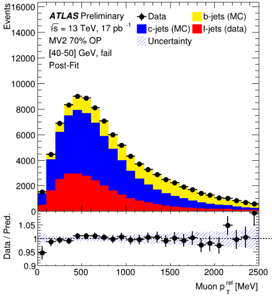
- **b-jet enhanced region** - Require at least one b-tagged (85% OP) jet, in addition to a muon-containing jet
- **Light flavour jet enhanced region** - Require no b-tagged (85% OP) jets in the event

Binned Likelihood Template Fit

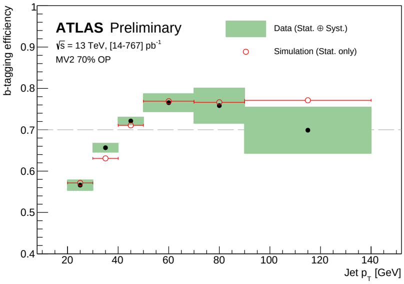
- Perform combined template fit to p_T^{rel} distributions of events from **b-jet enhanced region** which separately pass or fail the b-tagging requirement being studied
- b-jet and c-jet templates built from simulated events, light flavour template determined from data events in **light flavour jet enhanced region** (small b/c-jet contamination accounted for in likelihood)
- Three free parameters extracted: **b-tagging efficiency data/simulation scale factor (POI)**, b-jet and non-b-jet event normalisation factors

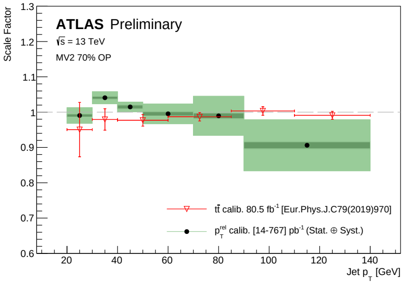
Fits performed for six jet p_T bins × four operating points (OP) of MV2 high level b-tagging algorithm





b-jet tagging efficiency in data and simulation found to be generally compatible with few % level precision, offers precision superior to $t\bar{t}$ approach† at low jet p_T



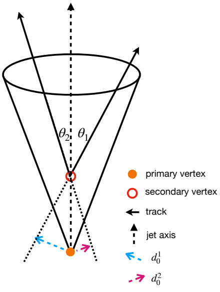


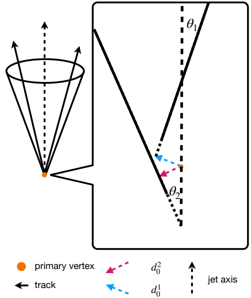
- Subject to different sources of uncertainty than $t\bar{t}$ likelihood method†, including:
  - Uncertainties associated with experimental performance of jet and muon identification and energy/momentum scales
  - Modelling of $g \to b\bar{b}$ contribution, b-hadron production fractions, $b \to \mu X$ and $b \to c \to \mu X$ decay rates and p^* distributions

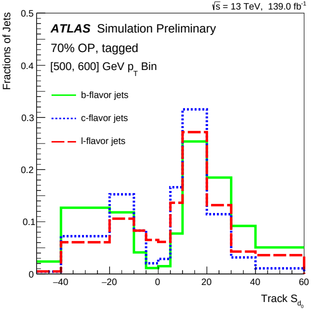
p_T^{rel} offers a valuable opportunity to cross-check a long established approach with a very different methodology, good consistency between measurements observed ✓

† Eur. Phys. J. C 79 (2019) 970

The rapidly falling jet p_T spectrum of $t\bar{t}$ events precludes the study of b-tagging efficiencies at high p_T, large samples of multi-jet events can offer a solution...







b-jet **Light flavour jet**

- High p_T multi-jet processes offer less topological information than $t\bar{t}$ events to understand the flavour composition of the jet sample $\rightarrow$ a discriminant based on the characteristics of the jets themselves is required to exploit the data sample
- The **signed transverse impact parameter significance S_{d_0}** for tracks associated with jets provides sensitivity to the jet flavour classification

$S_{d_0} = \text{sign} \cdot d_0/\sigma_{d_0}$, with sign $= d_0 \cdot \sin\theta / |d_0 \cdot \sin\theta|$ and σ_{d_0} the measured uncertainty on d_0

Measuring b-tag efficiencies for high p_T jets (ATL-PHYS-PUB-2022-010)

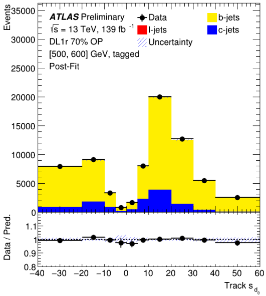
Event Selection - "Tag-and-Probe" Approach

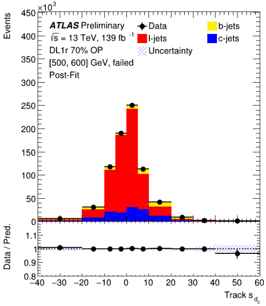
- Exploit dominant $pp \rightarrow b\bar{b}$ topology and target muon-in-jet signature to enrich b-jet fraction to $\approx 10\%$
- **Tag Jet** - $p_T > 20$ GeV and b-tagged (85% OP)
- **Probe Jet** - $p_T > 500$ GeV, matched to a muon, well separated in transverse plane from tag jet

Binned Likelihood Template Fit

- Consider $\boldsymbol{S_{d_0}}$ **of track with second largest** $|\boldsymbol{S_{d_0}}|$ **as primary discriminant** (largest is more sensitive to mis-reconstruction effects)
- Perform combined fit to events which separately pass or fail the b-tagging requirement being studied, extract number of b-tagged b-, c- and light flavour jets in data, relative to simulation
- Templates from simulated events, separately for "pass" and "fail" (S_{d_0} correlated with b-tagging discriminant)

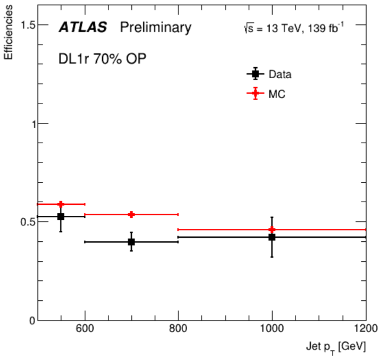
Fits performed for three jet p_T bins $\times$ four operating points (OP) of DL1r high level b-tagging algorithm

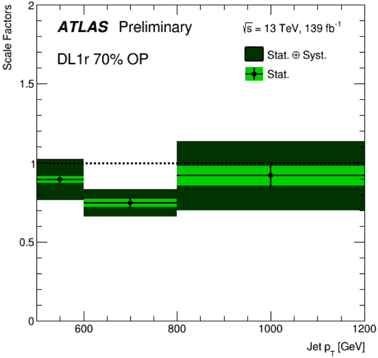




Measuring b-tag efficiencies for high p_T jets (ATL-PHYS-PUB-2022-010)

b-jet tagging efficiency in data and simulation found to be generally compatible for jets with p_T in the region of 1 TeV, with $\mathcal{O}(10\%)$ compatibility

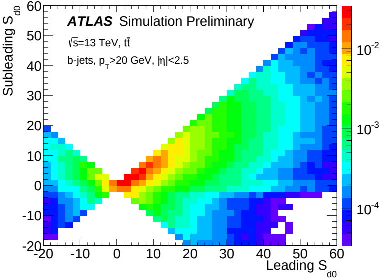


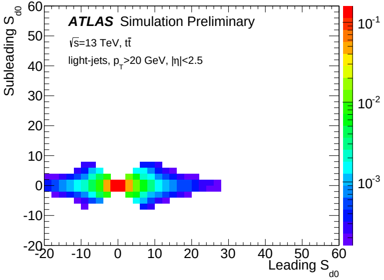


- Precision limited by systematic uncertainties, important effects include modelling of charged particle production within jets, description of tracking performance in simulation and limited size of simulated event samples

Important to constrain uncertainties associated with b-tagging performance at high p_T, where extrapolations from lower p_T data were previously relied upon

The impact parameters (IP) of tracks from b/c-hadron decays are intrinsically correlated, since they originate from a common decay vertex...





This feature was not exploited in early ATLAS IP-based taggers (IP3D), but forms a central motivation for more recent approaches:

- **RNNIP** [ATL-PHYS-PUB-2017-003] - Based on recurrent neural networks, considers tracks in a jet as a variable-length sequence, ordered by S_{d_0}
- **DIPS** [ATL-PHYS-PUB-2020-014] - Based on Deep Sets formalism, considers tracks in a jet as an unordered variable-length set $\rightarrow$ more physically motivated

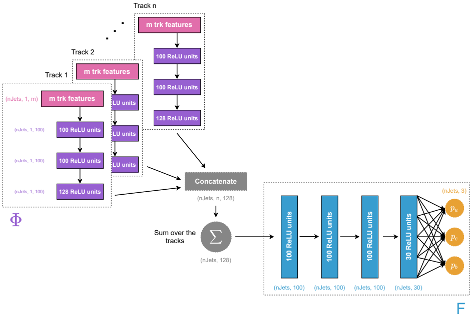
RNNIP already provided substantial performance improvements over IP3D, DIPS matches this performance but is faster to train and evaluate, offering more scope for optimisation and exploration of new track features!

Several improvements considered relative to original version, described in [ATL-PHYS-PUB-2020-014]

- Revised architecture, training procedure and hyperparameter optimisation
- Extended training sample (8× and enhanced at high p_T)
- Alternative looser track selection investigated ↓

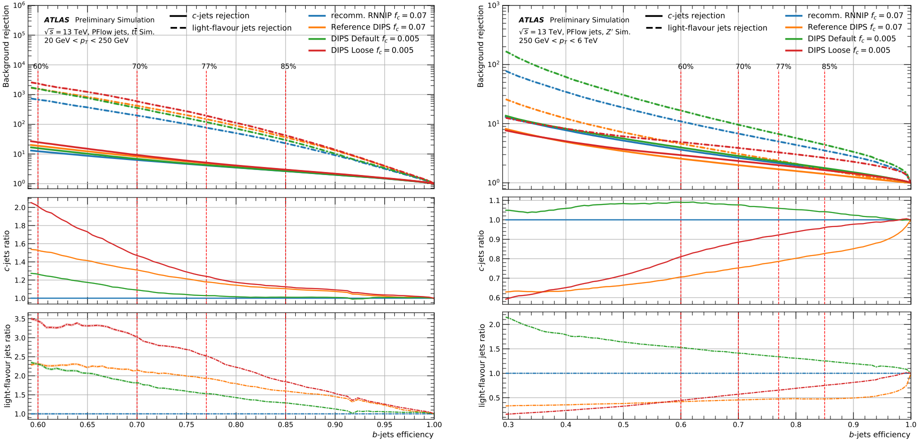
Track Selection			
	RNNIP	DIPS (default)	DIPS (loose)
Max. N_{tracks}	24		40
Min. p_T	1 GeV		0.5 GeV
Max. $\|d_0\|$	1 mm		3.5 mm
Max. $\|z_0 \cdot sin(\theta)\|$	1.5 mm		5 mm

DIPS Track Input Features
Impact parameter significances (S_{d_0} and S_{z_0})
Transverse momentum fraction ($\log\left[p_T^{track}/p_T^{jet}\right]$)
Opening angle w.r.t. jet axis ($\log[\Delta R(\text{track}, \text{jet})]$)
Hit information for each tracking layer



Architecture of DIPS neural network revised with two additional hidden layers and training batch size increased (256 → 15000)

Improved DIPS, with looser track selection, offers higher rejection at low jet p_T, while revised training improves high p_T performance with default track selection

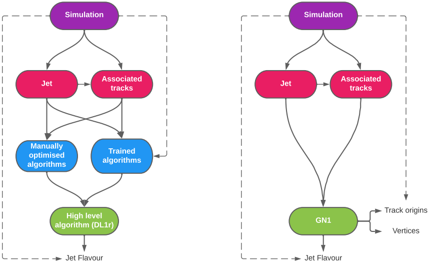


The improved DIPS will form the baseline "low level" IP tagger for Run 3, forming one input to dedicated variant of DL1

Jet Flavour Tagging with Graph Neural Networks (ATL-PHYS-PUB-2022-027)

GN1 is a novel flavour tagging algorithm, based on Graph Neural Networks (GNN)†

- The current ATLAS baseline "high level" tagger, DL1, is based on input variables from independently optimised "low level" track-based algorithms
- Conversely, GN1 utilises a single neural network, using tracks and some jet information directly as input



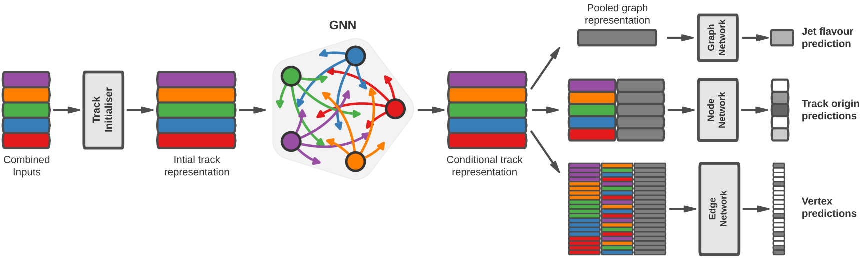
GN1 is trained with two auxiliary objectives to aid the primary classification goal:

- **1** - Grouping of tracks originating from a common vertex
- **2** - Prediction of the underlying physics process from which each track originated (e.g. primary interaction, b/c-hadron decay, material interaction, pileup or fake)
- In addition to the primary goal of classification, these objectives guide the network towards a more complete "understanding" of the underlying physics processes

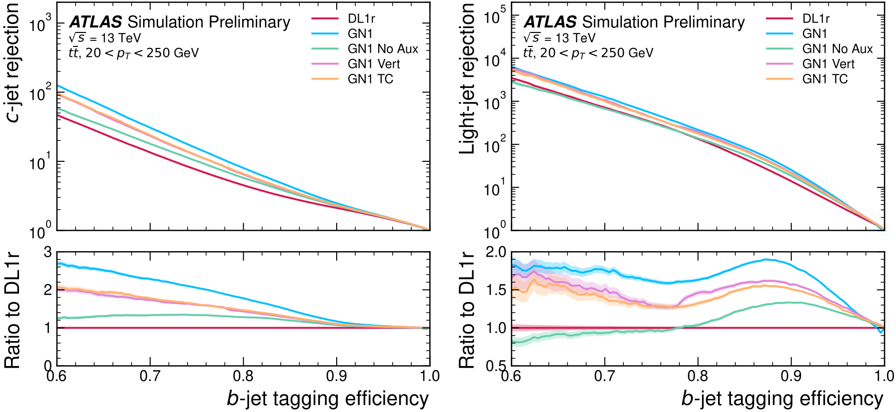
Potentially easier to maintain (fewer algorithms / simpler to retune) and auxiliary vertex and track origin predictions could also be used in other applications

† For more information see: EPJC 81 (2021) 6 and arXiv:2002.08772

GN1 Input Features	
Jet	transverse momentum and pseudorapidity
Tracks	q/p, η and ϕ relative to jet axis impact parameters and uncertainties hit information for each tracking layer label if used for lepton reconstruction (GL1Lep variant)



Output node representations of graph network used to predict the flavour class, track origins and the track-pair vertex compatibilities for each jet



- ***b*-tagging** - Roughly 2.1× (1.8×) improvement in *c*-jet (light flavour jet) rejection relative to DL1r, for a *b*-jet efficiency of 70% **(shown above)**
- ***c*-tagging** - Roughly 2.0× (1.6×) improvement in *b*-jet (light flavour jet) rejection relative to DL1r, for a *c*-jet efficiencies of 20 - 40% **(figures in backup slide)**

Clear improvement in classification performance from inclusion of auxiliary objectives in the training and model, similar gains from each auxiliary objective

Summary

ATLAS is pursuing a broad programme of new developments in flavour tagging to further exploit the important Run 2 dataset and prepare for the forthcoming Run 3 data taking at $\sqrt{s} = 13.6$ TeV

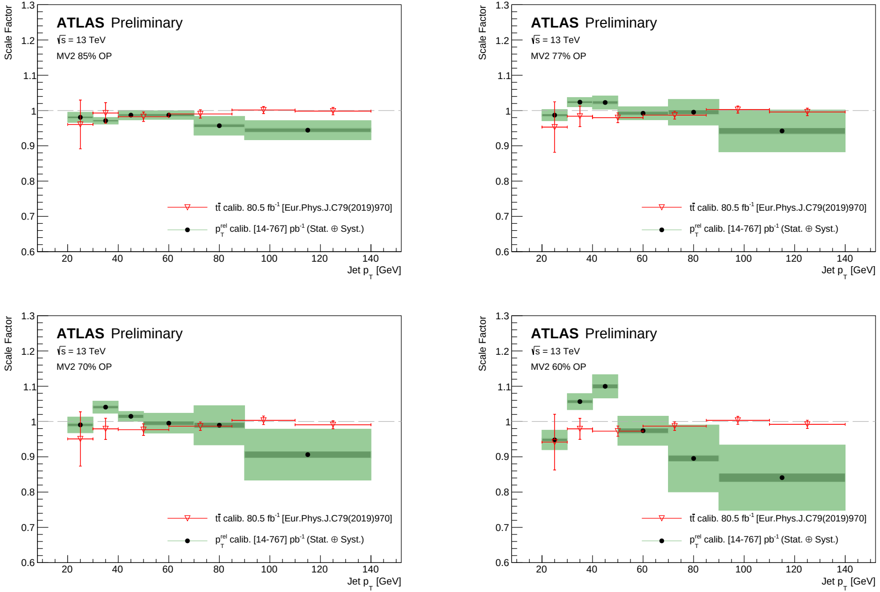
- **New alternative calibrations with multi-jet events** - two methods reduce the reliance on $t\bar{t}$ events and improve the understanding of flavour tagging performance in data, both in terms of precision and kinematic reach
- **Improved DIPS algorithm** - Thorough revision and performance improvements for the baseline impact parameter based tagger, ready for start of Run 3 data taking
- **GNN Tagger** - Promising performance in simulation, behaviour in data and various alternative MC simulation configurations now under study

For many more public ATLAS results on flavour tagging, please see:

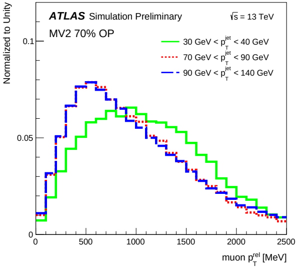
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/FlavourTaggingPublicResultsCollisionData

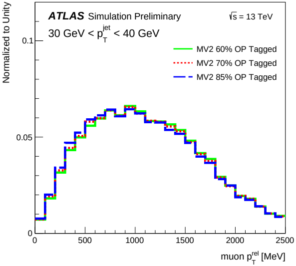
Additional Slides

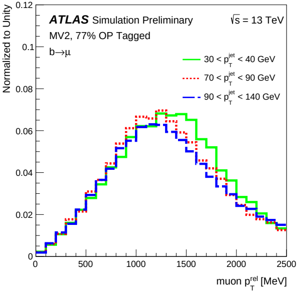
Hyperparameter	Reference DIPS	DIPS Default	DIPS Loose
Aggregation function	Summation		
Loss function	Categorical Crossentropy		
Optimiser	ADAM (Adaptive Moment Estimation)		
Activation function	ReLU (Rectified Linear Unit)		
Output activation function	Softmax		
Regularisation	Batch Normalisation		
ϕ $N_{\text{Hidden layer}}$	3		
ϕ $N_{\text{Nodes/layer}}$	[100, 100, 128]		
F $N_{\text{Hidden layer}}$	2	4	
F $N_{\text{Nodes/layer}}$	[100, 100]	[100, 100, 100, 30]	
Training sample composition	$t\bar{t}$	$70\,\%$ $t\bar{t}$, $30\,\%$ Z'	
Number of training jets	3 M	22.8 M	
Batch size	256	15000	
Free (trainable) parameters	48987	62167	
Fixed parameter	1056	1316	

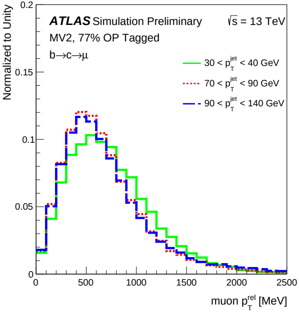

ATLAS Preliminary
√s = 13 TeV
MV2 85% OP
MV2 77% OP
MV2 70% OP
MV2 60% OP
Scale Factor
Jet p_T [GeV]
tt̄ calib. 80.5 fb⁻¹ [Eur.Phys.J.C79(2019)970]
p_T^rel calib. [14-767] pb⁻¹ (Stat. ⊕ Syst.)

Measuring b-tag efficiencies with the p_T^{rel} method (ATL-PHYS-PUB-2022-025)









c-tagging performance