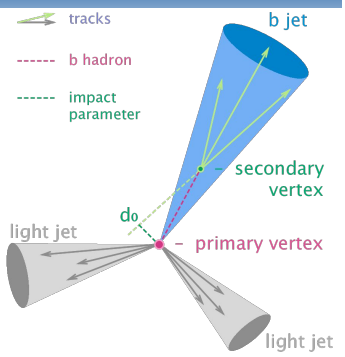
Development and Integration of a Deep-Impact-Parameter-Sets based *b*-Tagging Algorithm in ATLAS - LHCC Poster Session -

Introduction:

- Identification of the jet flavour is a crucial task for many ATLAS analyses.
- Improvements in the *b*-tagging performance bring better physics results.
- Different basic and state-of-the-art multivariate techniques are used and combined to determine the jet flavour.
- One of the new techniques is a Deep-Sets-based neural network, the Deep-Impact-Parameter-Sets (DIPS) tagger [1], which uses track information.
- All plots of this Poster can be found in [5].

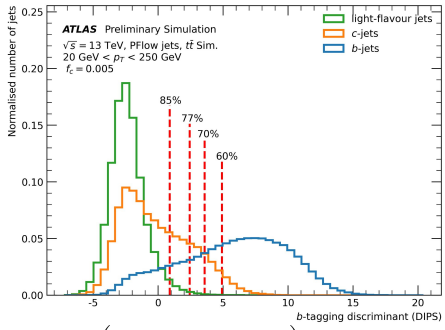
How does *b*-tagging work?

- Using unique characteristics of *b*-hadron decays
  - Lifetime of *b*-hadrons (~5mm at 50 GeV)
  - Decay topology
  - Heavier mass
- These characteristics bring interesting experimental signatures:
  - Displaced tracks
  - Displacement of secondary vertex w.r.t. primary vertex -> impact parameter
- Information provided by low-level algorithms (e.g. JetFitter[2], Secondary Vertex Finder[3])
- Here: Focus on *b*-tagging with track information



b-tagging Discriminant and Working Point Definition

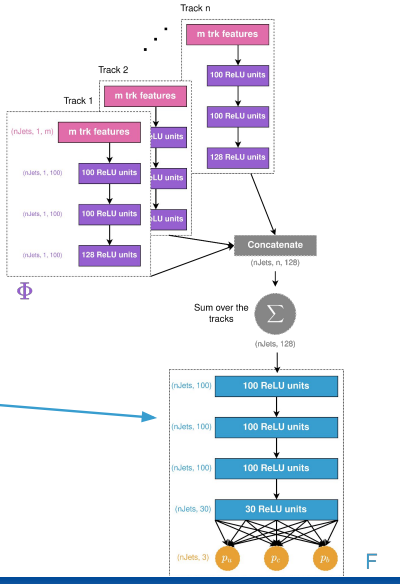
- Probability output of the networks is used to calculate a *b*-tagging discriminant D_b
- A fixed *b*-efficiency WP is defined such that a b-jet is tagged with a certain probability, for instance, 70%.
- Cut values for WP calculated by integrating over *b*-distribution from right to left, indicted by the vertical dashed lines.
- All jets above the vertical dashed line are accepted as *b*-jets, for the corresponding WPs.



$$D_b = \ln\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_u}\right)$$

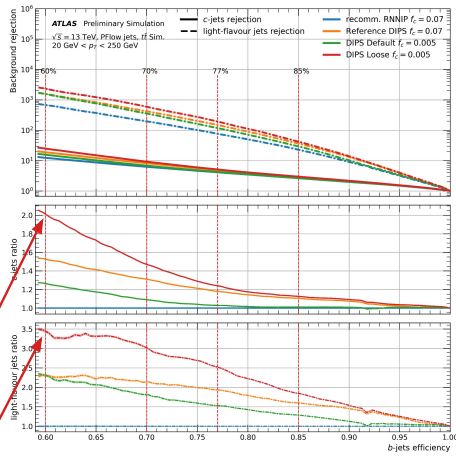
The Deep-Impact-Parameter-Sets (DIPS) Tagger

- Deep Sets[4]: Treat each element as a set without a specific order (track permutation invariance)
- Each track is processed by a neural network with shared weights between tracks (Φ network)
- Outputs of the neural network for the tracks are pooled for further processing
- Pooled outputs are used for classification by a subsequent neural network (F network)
- Will replace the Recurrent Neural Network Impact Parameter (RNNIP) tagger



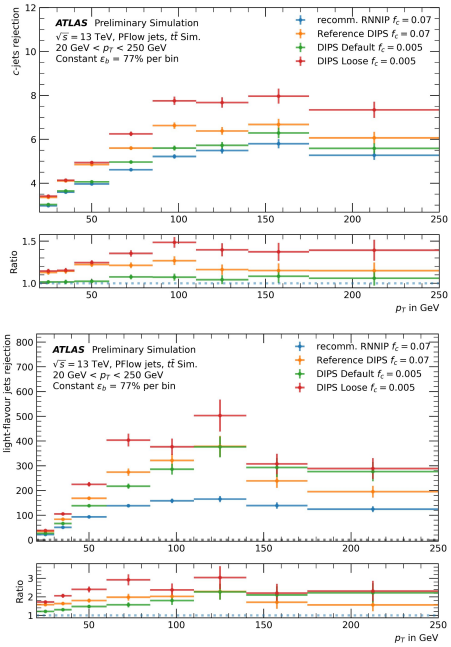
Training Results - ROC Curve

- All DIPS models outperform the recommended RNNIP
- DIPS Default worse than Reference DIPS (which was expected)
  - Reference DIPS uses the loose track selection
- New DIPS Loose has better performance compared to Reference DIPS
- DIPS Loose improvement w.r.t. recommended RNNIP:
  - *c*-jets rejection at 60% WP: 2.0
  - light-flavour jets rejection at 60% WP: 3.45



Training Results - p_T vs Rejection

- Different rejections binned by the jet p_T
- Constant *b*-efficiency per bin
- All DIPS models outperform recommended RNNIP
- DIPS Default slightly worse than Reference DIPS (expected):
  - Reference DIPS uses the loose track selection
- DIPS Loose outperform all other models
- Clear improvement in *c*-jet rejection for DIPS Loose w.r.t. recommended RNNIP
- Improvement in light-flavour jet rejection for DIPS Loose w.r.t. recommended RNNIP



Hyperparameters and Training Sample

- Two new DIPS models trained with different track selections
- Looser track selection relaxes the p_T and impact parameter requirements
- DIPS Loose capable of making use of the extra tracks!

Tagger	recomm. RNNIP	DIPS Default	DIPS Loose
Max N_{Tracks}	25		40
p_T	> 1 GeV		> 0.5 GeV
$\lvert d_0 \rvert$	< 1 mm		< 3.5 mm
$\lvert z_0 \sin(\theta) \rvert$	< 1.5 mm		< 5 mm
$\lvert \eta \rvert$	< 2.5		
$N_{\text{Pixel holes}}$	< 2		
$N_{\text{Silicon hits}}$	≥ 7		
$N_{\text{Silicon shared hits}}$	< 2		
$N_{\text{Silicon holes}}$	< 3		

- Similar hyperparameters for the three DIPS versions
- Differences between DIPS models:
  - F network architecture
  - Training sample composition
  - Number of training jets
  - Batch size
- Added jets from heavy Z´ processes to improve the high p_T tagging capabilities

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	

Alexander Froch (Albert-Ludwigs University Freiburg), 18.11.2021 for the ATLAS Collaboration

[1]: Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS, ATL-PHYS-PUB-2020-014
[2]: A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS, *J.Phys.Conf.Ser.* 119 (2008) 032032
[3]: Secondary vertex finding for jet flavour identification with the ATLAS detector, ATL-PHYS-PUB-2017-011
[4]: Deep Sets, M. Zaheer et. al., arXiv:1703.06114
[5]: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-004/