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-hadrons decays:
  - Lifetime of b-hadrons (~5mm at 50GeV)
  - 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

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 (Ω-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 (∆Ω-network)
  - Will replace the Recurrent Neural Network Impact Parameter (RNNIP) tagger

Hyperparameters and Training Sample

- Two new DIPS models trained with different track selections
- Looser track selection relaxes the pT and impact parameter requirements
- DIPS Loose capable of making use of the extra tracks!
- Similar hyperparameters for the three DIPS versions
- Differences between DIPS models:
  - ∆Ω network architecture
  - Training sample composition
  - Number of training jets
  - Batch size
- Added jets from heavy Z processes to improve the pT tagging capabilities

Hyperparameters and Training Sample

- Eight hyperparameters for training each model:
  - Number of training jets
  - Training jets sample composition
  - Batch size
  - Additional track selection

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, indicated by the vertical dashed lines.
- All jets above the vertical dashed line are accepted as b-jets, for the corresponding WPs.

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 the 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. the corresponding WPs.
  - Improvement in light-flavour jet rejection for DIPS Loose w.r.t. the corresponding WPs.