The use of the full set of input variables from the SMT algorithm and the output of the SMT multivariate in the training is set to 7% and that of light-flavoured jets to 93%. These values achieve a suitable balance.

The first high-level tagger is a BDT discriminant that combines the output of the low-level taggers.

Training Samples
- Using hybrid sample composed of SM $t\bar{t}$ and $Z \rightarrow q\bar{q}$ events
- More statistics in higher $p_T$ region
- Undersampling approach applied to match $p_T$ and $m_T$ distributions for all 3 flavour categories
- Ensure independency of tagging from kinematics
- Using 23M jets for training

Deep Neural Network Architecture
- Deep neural network requires also
  - Preprocessing, feature selection...
  - Network with fully connected layers
  - Multi-class output $\Rightarrow$ allows also $c$-tagging
  - $D_{L1 \text{b-tag}} = b$ ($z = (1 + P_{	ext{light-flavour}})$)
- Using Keras (2.2.4) framework with tensorflow backend
- Full training procedure relies on HDF5
- Application can be run in ATLAS reconstruction software, relying on the LWTNN C++ interface

Hyper Parameter (HP) Optimisation with GRID GPUs
- Using docker image (built by Gitlab CI) for jobs
- Configurable amount of HP combinations (config)
- Workflow optimised for GRID-submission
- 800 combinations over 5 HP dimensions
- Optimisation provides good results

Final Training Results
- Dedicated trainings of Particle-Flow jets and Variable Radius Track jets
  - New $b$-tagging recommendations for ATLAS
- Performance gain similar for Particle-Flow and Variable Radius Track jets
  - Gain in light-jet rejection higher than in $c$-jet rejection
  - Increase in performance up to 100% for light-jet rejection & up to 50% for $c$-jet rejection

Why $b$-Tagging?
- Several interesting physics processes have $b$-quarks in their final state
- Or a veto on $b$-quarks can suppress the background
- Heavy-flavour tagging important tool for physics analyses
- Precision measurements
- Search for new physics

Network Output
- Each jet gets probability for being a $b$, $c$, or light flavour jet
- Good separation of $b$- & light-jets
- $b$- & $c$-jets have more similar physics behaviour

Deep-learning Structure in ATLAS
- Baseline taggers deploy specific heavy flavour jet properties
  - Long lifetime ($\sim 1.9$ ps $\Rightarrow$ 3mm track in detector)
  - High mass ($\sim 5$ GeV)
  - High decay product multiplicity
  - $b$-hadron decays to a $c$-hadron ($V_{ub}/V_{cb}$)
- High-level taggers (MV2 & DL1) combine these information (40-50 variables)