NEW ATLAS *B*-TAGGING ALGORITHM FOR Run-3

1 – *B*-TAGGING IN A NUTSHELL

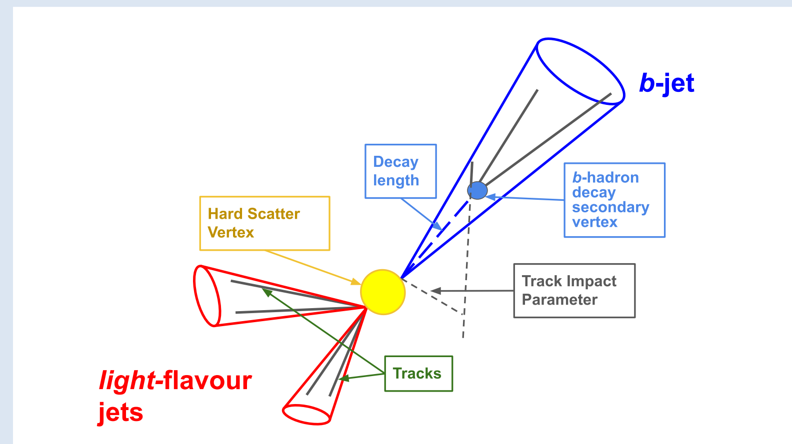
b-tagging is the process of identifying jets containing b-hadrons (*b*-jets), discriminating them from jets containing *c*-hadrons and no *b*-hadrons (*c*-jets), and from those containing neither *b* nor *c*-hadrons (light-flavour jets).

How do we identify the *b*-jets?

b-hadrons have a significant mean lifetime, large mass, and large decay multiplicities. These features allow to identify the *b*-jets:

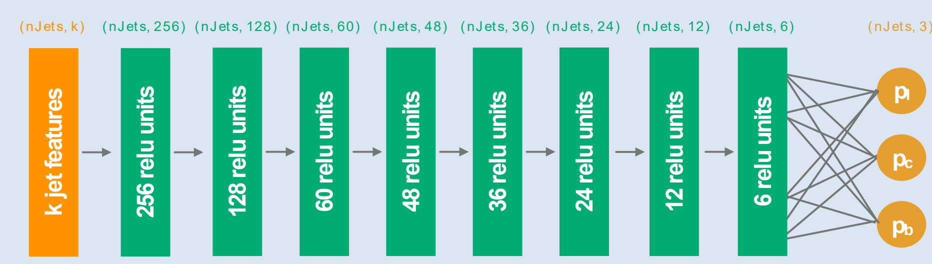
- **Tracks in *b*-jets are more likely to have large impact parameters**
- **Long-lived *b*-hadrons displaced decays lead to the presence of secondary vertices in *b*-jets**

b-tagging in ATLAS is done by training classifiers based on machine learning algorithms to explore these features. In particular, the baseline Run-2 *b*-taggers are all based on Neural Networks (NN).



2 – THE ALGORITHMS

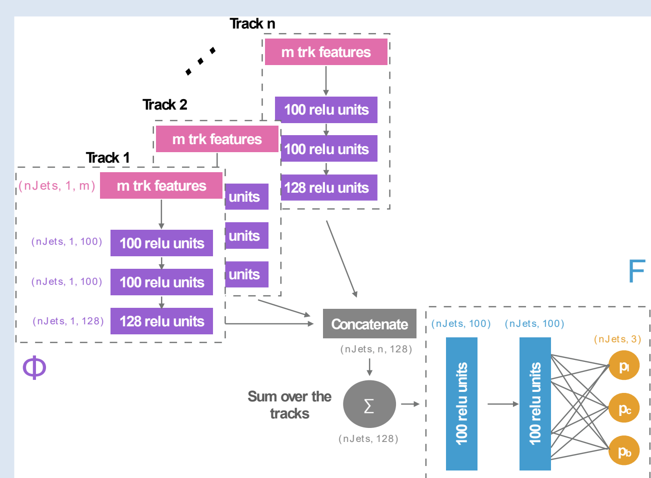
Baseline ATLAS *b*-tagging algorithms belong to the **DL1 series**, in which trained and manually optimized algorithms (the so-called Low-Level Algorithms) produce features for high-level fully-connected feed-forward NNs. State-of-the-art algorithm is **GN1**, a Graph Neural Network (GNN) that treats jets as fully connected graphs in which tracks are the nodes. GN1 is still being developed and undergoing validations.



DL1 series:

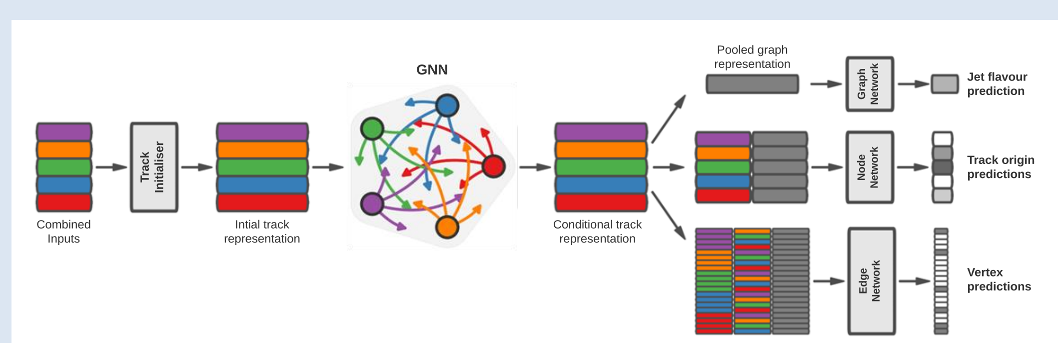
The members of the DL1 series are trained to use the features produced by the Low-Level Algorithms to classify the jets. Low-Level Algorithms use tracks to focus on specific features of *b*-jets:

- **IP2D and IP3D** use the impact parameters of tracks to identify those most likely associated to the decay products of *b* and *c* hadrons. **RNNIP** is a Recurrent Neural Network (RNN) trained to identify the origin of tracks from a set of track and jet features.
- **DIPS**, the most recent low-level algorithm, represents tracks in jets as sets of unordered data. It is based on a DeepSet architecture shown to surpass the RNN both in terms of training time and performance. Recently, an improved version of the algorithm, named **DIPS Loose**, has been developed. It is trained on a dataset containing tracks selected with looser cuts with respect to previous DIPS and RNNIP.
- **SV1 and JetFitter** use vertex-fitting techniques to reconstruct secondary vertices in jets.



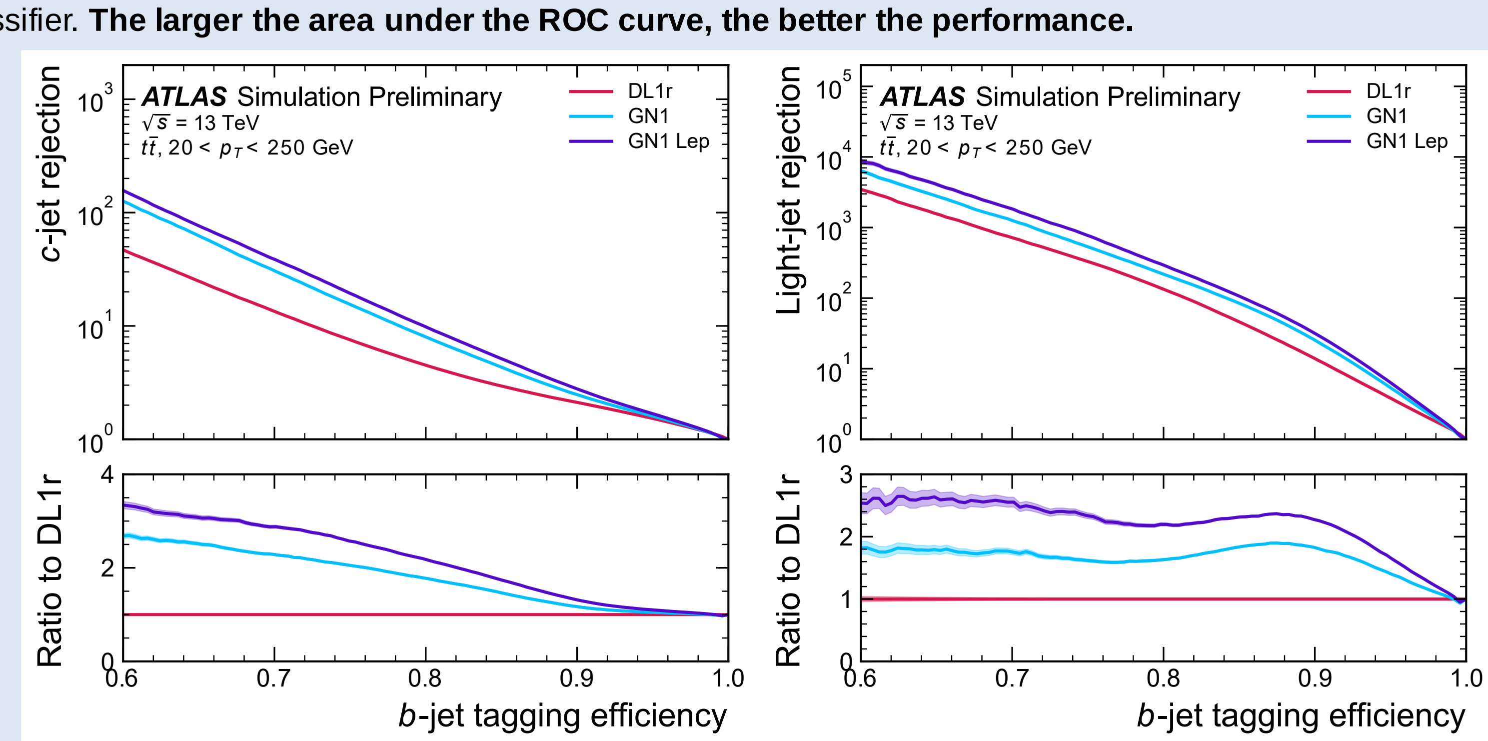
GN1:

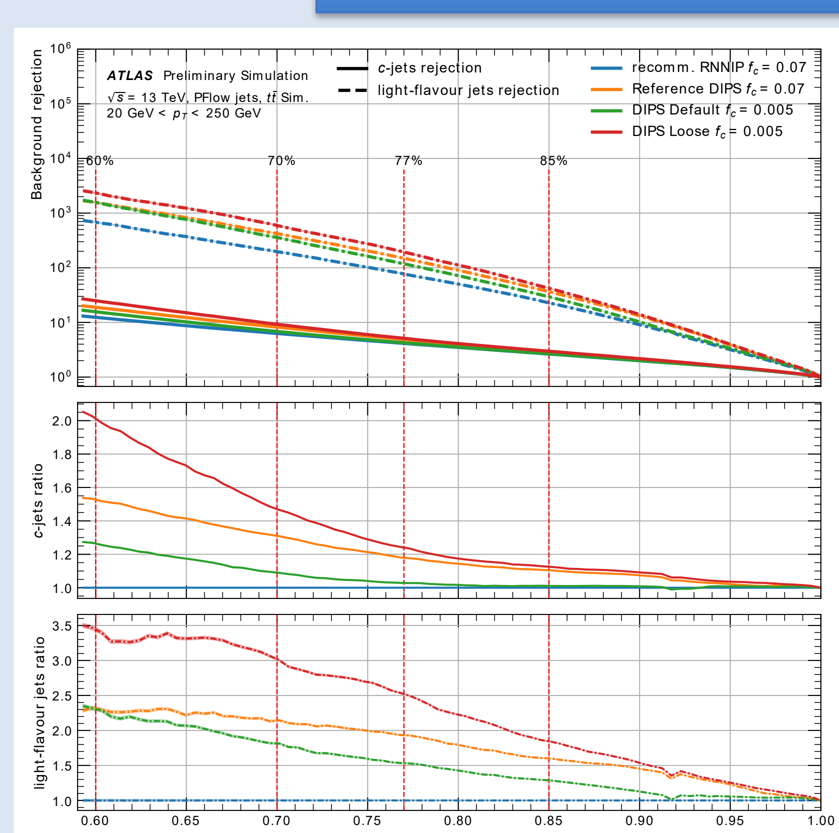
GN1 is an algorithm with a monolithic structure. It directly operates on tracks to perform *b*-tagging. At the same time, it performs vertexing and track classification, which are defined as auxiliary tasks that guide the GNN toward an understanding of the underlaying physics, hence removing the need for low-level algorithms.



3 - PERFORMANCE

Receiver Operating Characteristics (ROC) curves are a way to assess classifiers' performance. They display the fraction of *b*-jets correctly *b*-tagged vs the fraction of non *b*-jets correctly non *b*-tagged for any working point of a classifier. **The larger the area under the ROC curve, the better the performance.**





Martino Tanasini on behalf of the ATLAS Collaboration – martino.tanasini@cern.ch