Applications of Machine Learning to b-tagging in ATLAS

Matthew Feickert[1]
(on behalf of the ATLAS collaboration)

[1]Southern Methodist University

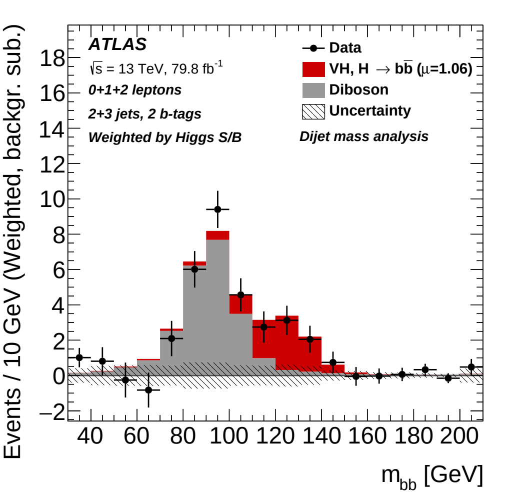
Machine Learning for Jet Physics Workshop
November 14th, 2018

Outline

Outline

Flavor Tagging in ATLAS

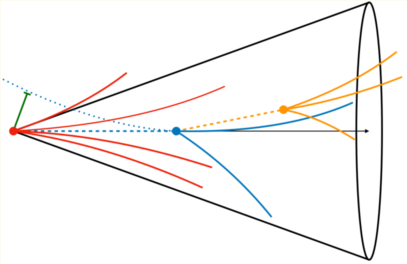
- Heavy-flavour jet identification exploits properties of the hadrons in the jet to discriminate heavy flavour (b-,c-) initiated jets from those arising from light partons (light-flavour jets)
- The scientific program at the LHC relies heavily on accurate identification of the reconstructed particles from data: b-tagging is mission critical $\longrightarrow$
- Everyone needs b-jets!
    - SM Higgs ($H \to b\bar{b}$, $HH \to 4b$)
    - top physics ($t \to Wb$)
    - BSM ($X \to b\bar{b}$)



First discovery of $VH \to b\bar{b}$ by ATLAS in 2018 [1]

Outline

b-jets contain...



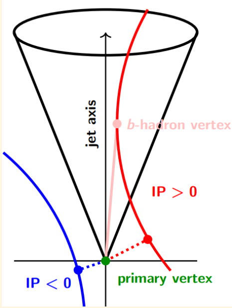
A typical *b*-hadron decay chain ($\sim$ 90%) involves a decay to a *c*-hadron [2]

- *b*-hadron decay vertex ● displaced from the primary *pp* vertex ●
- *c*-hadron decay vertex ● further displaced, often close to *b*-hadron flight axis - - -
- Tracks from secondary and tertiary vertices with large impact parameters with respect to the primary *pp* vertex

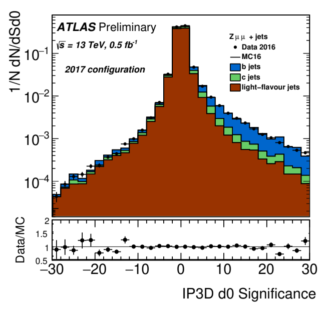
Track Impact Parameters (IP): IP2D, IP3D [4]

The signed impact parameters of tracks associated to jets are effective jet flavour discriminants

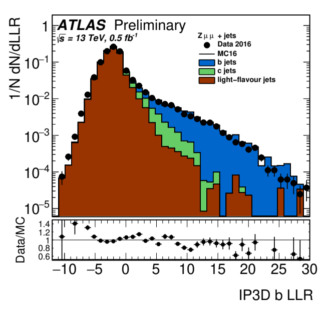
- Sign of IP: positive if track's point of closest approach to PV is downstream of plane defined by the PV and jet axis
- Tracks originating from *b*-hadrons tend to have positive IPs with high significance $(\mathrm{IP}/\sigma_{\mathrm{IP}})$
- IP3D builds up log-likelihood ratio discriminant assuming the track IPs are uncorrelated
- ● Impact parameters are highly inclusive and efficient



Sign of the IP [3]



Distribution of transverse IP significance

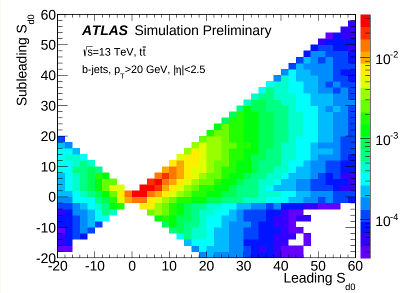


Log-likelihood ratio discriminant from 3D IP of tracks

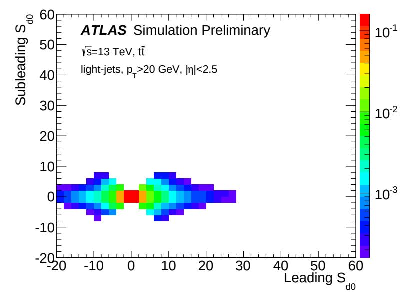
Track Impact Parameters (IP): Exploiting track multiplicity with RNNIP [4, 6]

For *b*-jets, tend to have multiple highly significant IP tracks within a jet. **Not** the case for light-flavour!

- The dimensionality of the problem of exploiting correlations grows very high with the many tracks in a *b*-jet
  - Remember: IP3D assumes IPs are uncorrelated
- Turn to **Recurrent Neural Networks** (RNN) that specialize in learning sequential correlations in arbitrary-length sequences
- Uses the same track selection as IP3D and gives additional flexibility to easily incorporate **new low level kinematic information**!
- c.f. Walter Hopkins's talk from earlier [5]



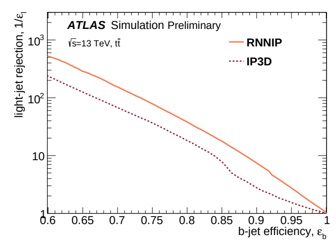
b-jets



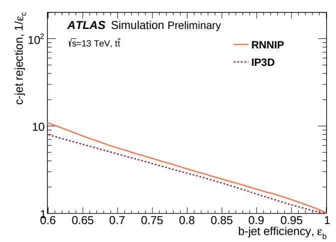
Light flavour jets

Distribution of d_0 significance for leading and subleading d_0 significance tracks

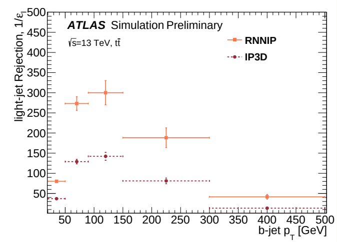
Echoing Walter [4, 5]



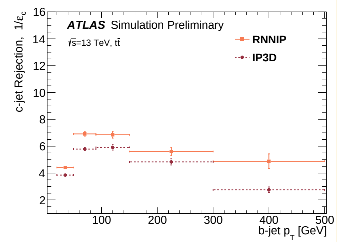
ROC for light-flavour jet rejection vs. b-jet efficiency



ROC for *c*-jet rejection vs. b-jet efficiency



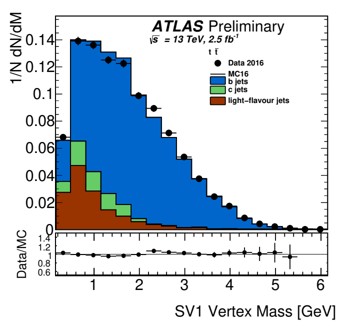
light-flavour jet rejection vs. p T for flat b-jet efficiency of 70%



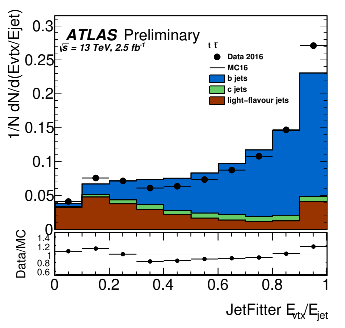
c-jet rejection vs. p T for flat b-jet efficiency of 70%

Displaced secondary vertices (SV) and Decay Chain: SV1, JetFitter [4]

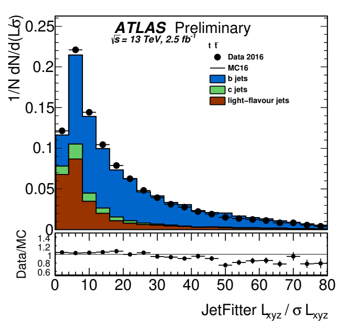
- Expectation of a secondary vertex from either *b* or *c*-hadron decays
- Most *b*-jets contain a *c*-jet, so use **Kalman filter** (JetFitter) to search for common axis for 3 vertices
  - Invariant mass of tracks at SV used to discriminate *b* or *c*-hadron decay vertices from V^0 decays or material interactions
  - hard *b*-jet fragmentation: SV carries large fraction of jet energy
  - Constraint to the decay axis further improves power of SV based discriminants



Mass of tracks at secondary vertex



Energy fraction of JetFitter secondary vertex tracks

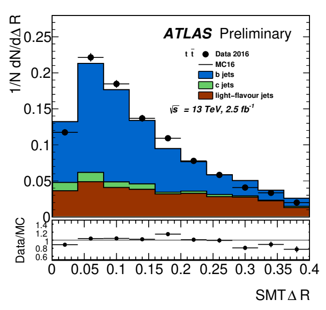


Significance of 3D decay distance

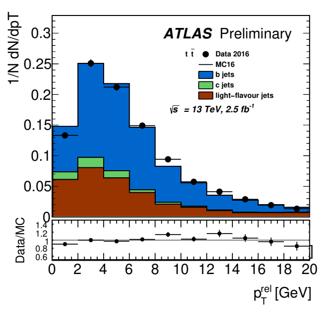
Muons: Soft Muon Tagger (SMT) [4]

Large branching fractions for the semi-leptonic c/b hadron decays and the clean "muon-in-jet" experimental signature

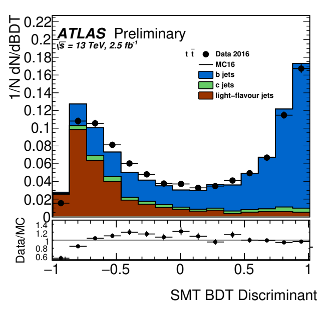
- Higher rate of muons within b/c-jets compared to light flavour jets due to the decays $B \to \mu\nu X$, $B \to DX \to \mu\nu X'$ ($\mathcal{B} \sim 10\%$)
- 🟢 Complementary to SV and IP based taggers, different c/b hadron properties exploited and ATLAS detector components employed
- 🔴 Light flavour jet backgrounds from muons produced in π/K decays in flight difficult to model in simulation



ΔR of muon w.r.t. jet axis



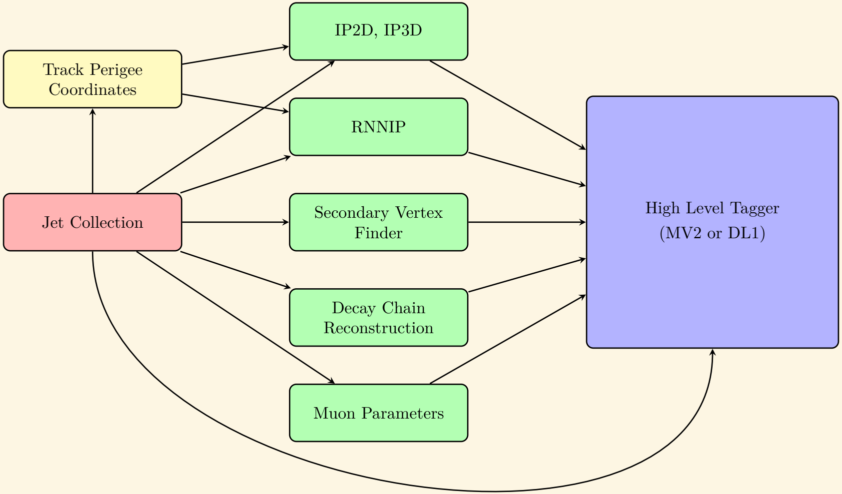
p_T of muon relative to the jet axis



BDT built from muon observables

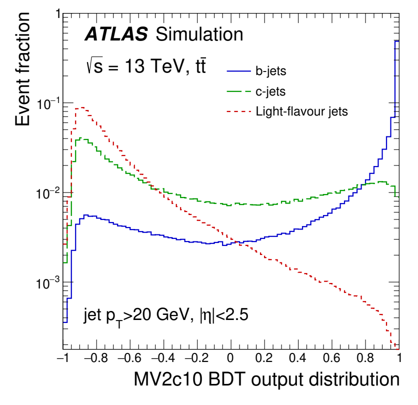
Building a High Level Tagger from low level tagger inputs

- Previous low level taggers now serve as inputs to high level taggers (BDT or DNN)

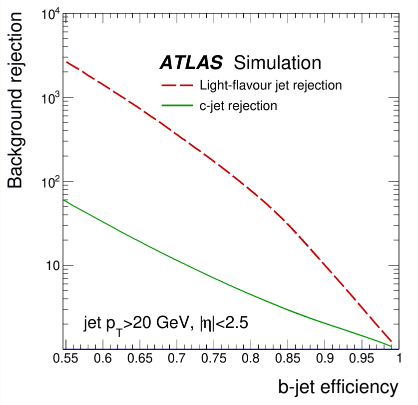


The Discriminant: High level tagger output [4, 7]

- Final discriminant is made from high level tagger output score
- Jet is *b*-tagged if score is greater than working point score defined from desired *b*-jet efficiency vs. *c*-jet and light jet rejection



Output of MV2c10 BDT

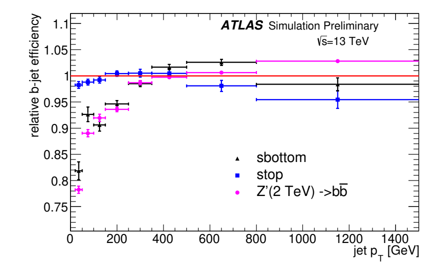


c-jet/light-jet rejection as a function of *b*jet tagging efficiency

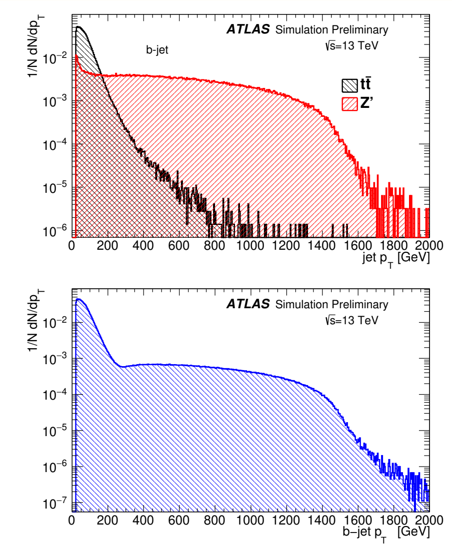
Outline

Hybrid Training [4]

- New training strategy: hybrid sample
  - $t\bar{t}$ events: low p_T ($p_T < 250$ GeV)
  - Z' events: high p_T ($p_T > 250$ GeV)
- Enriching the high p_T component of the training sample improves response in medium-to-high p_T events
- Verification of performance universality in compelling physics samples



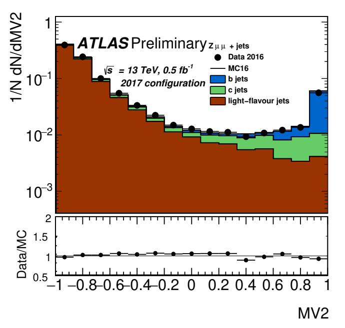
MV2MuRnn *b*-tagging efficiency on BSM samples compared to 77% working point from training on hybrid



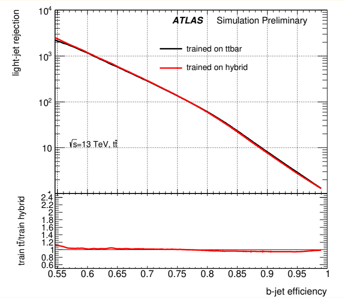
b-jet p_T distribution of the hybrid sample used for the training

MV2 (BDT)

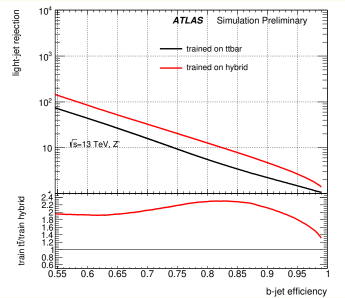
- High level BDT consumes low level tagger outputs as inputs along with p_T and η of jets
- Signal: *b*-jets, Background: *c*-jets and light-flavour jets
- *c*-jet fraction of training is 7% and light-jet fraction is 93% for suitable balance (MV2c10)



MV2c BDT output (2017 configuration)



ROC for light-flavour jet rejection vs. *b*-jet efficiency evaluated on $t\bar{t}$ and $t\bar{t} + Z'$ hybrid

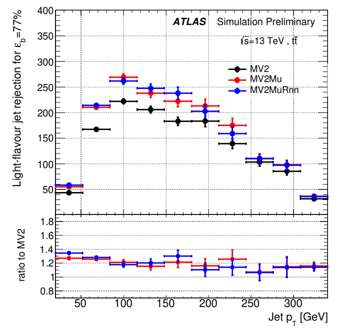


ROC for light-flavour jet rejection vs. *b*-jet efficiency evaluated on Z' and $t\bar{t} + Z'$ hybrid

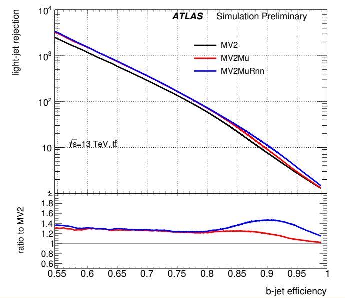
MV2 Variants

Have many low level tagger inputs, so create variants on MV2

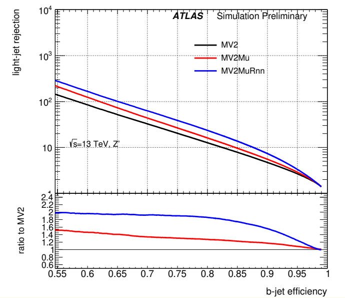
- MV2: Reference configuration of IP3D, SV1, and JetFitter
- MV2Mu: MV2 with addition of SMT
- MV2MuRnn: Kitchen sink — MV2 with SMT and RNNIP



light-flavour jet rejection vs. p_T for flat b-jet efficiency of 77%



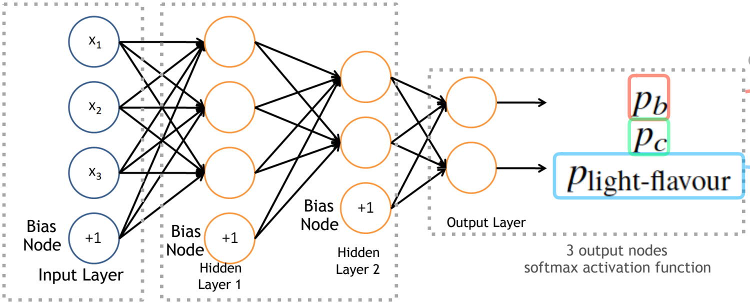
ROC for light-flavour jet rejection vs. b-jet efficiency evaluated on $t\bar{t}$



ROC for light-flavour jet rejection vs. b-jet efficiency evaluated on Z'

DL1 - General Overview

Training using
ttbar-Z' Hybrid sample



higher level variables

6-10 hidden layers
Mixtures of
MaxoutDense and Dense layers
ReLU activation function

3 output nodes
softmax activation function

Performance evaluation
on pure ttbar and Z' samples

Building the
final DL1 discriminant:
Reducing output dimensionality

b-tagging: $$DL1b f_{\text{c-jets}} = \ln\left(\frac{\boxed{p_b}}{f_{\text{c-jets}} \cdot \boxed{p_c} + (1 - f_{\text{c-jets}}) \cdot \boxed{p_{\text{light-flavour}}}}\right)$$

c-tagging: $$DL1c f_{\text{b-jets}} = \ln\left(\frac{\boxed{p_c}}{f_{\text{b-jets}} \cdot \boxed{p_b} + (1 - f_{\text{b-jets}}) \cdot \boxed{p_{\text{light-flavour}}}}\right)$$

—> **Increased Flexibility:**
+ Background weighing tuneable after training
+ Same training usable for b- and c-tagging

NN config file size ~1MB

DL1: b-tagging [4, 8, 9]

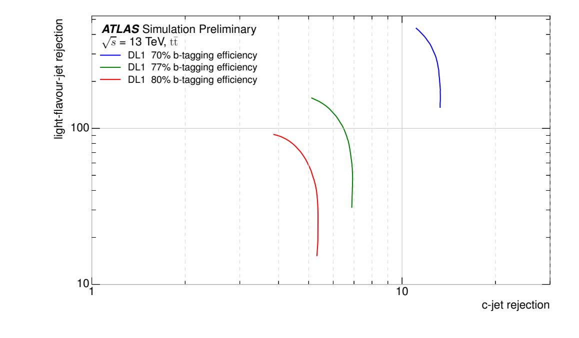
- Tuneable after training

$$DL1bf_{c\text{-jets}} = \ln\left(\frac{p_b}{f_{c\text{-jets}} \cdot p_b + (1 - f_{c\text{-jets}}) \cdot p_{\text{light-flavour}}}\right)$$

  - $f_{c\text{-jets}}$ is user tunable importance weight
  - $f_{c\text{-jets}} \to 1$ better c-jet rejection, worse light-jet rejection

- Performance of the DL1 and MV2 discriminants for tagging b-jets is found to be very similar
- DL1 provides advantages for future R&D
  - Training with Adversarial Networks, add RNNIP for end-to-end training

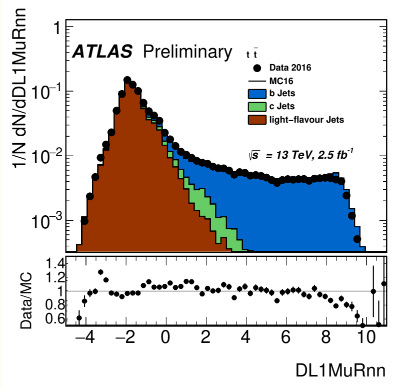
Iso-efficiency curve: Scan over full range of $f_{c\text{-jets}}$



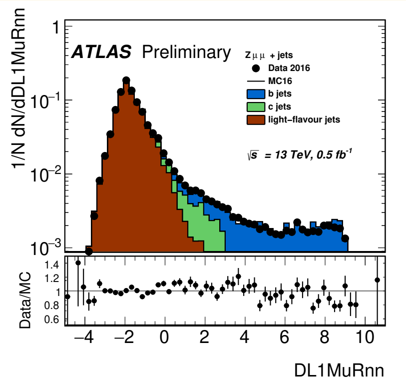
DL1 light-flavour vs c-jet rejection on $t\bar{t}$ events

DL1: Verification of modelling [4, 8, 9]

- ● Good separation of flavours
- ● Generalization
- ● Simulation describes the data within 20% with some localized differences for low and high values



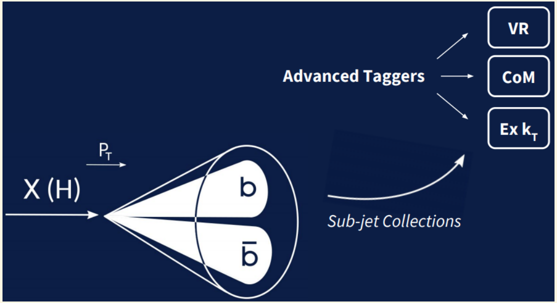
DL1MuRnn output discriminant on $t\bar{t}$-dominated $e\mu$ events



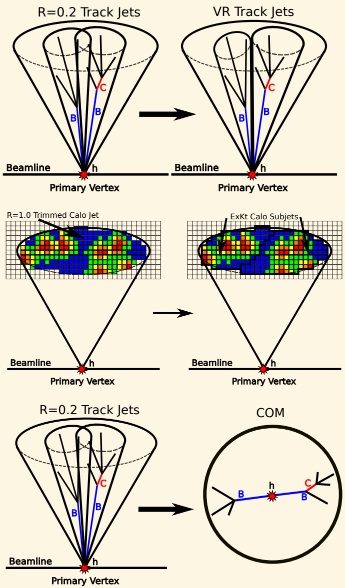
DL1MuRnn output discriminant on $Z \to \mu^+\mu^-$+jets-dominated events

$X \to b\bar{b}$ Tagger: Alternatives in the boosted regime [11]

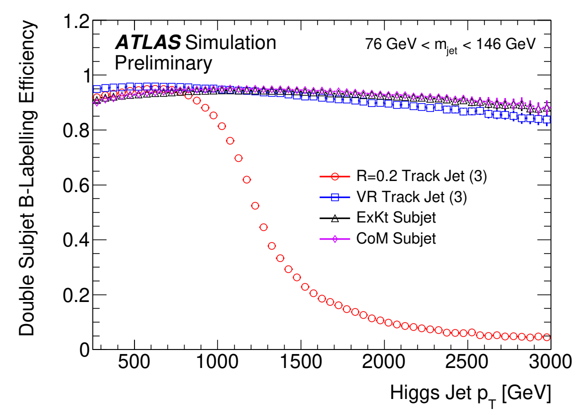
- Variable Radius Track Jets: Jet radius goes as $R \propto p_T^{-1}$
- Exclusive-k_T: k_T variant that clusters subjets until all protojet separations above threshold
- Center-of-Mass: Boost jet constituents into large-R jet rest frame and cluster with EECambridge algorithm



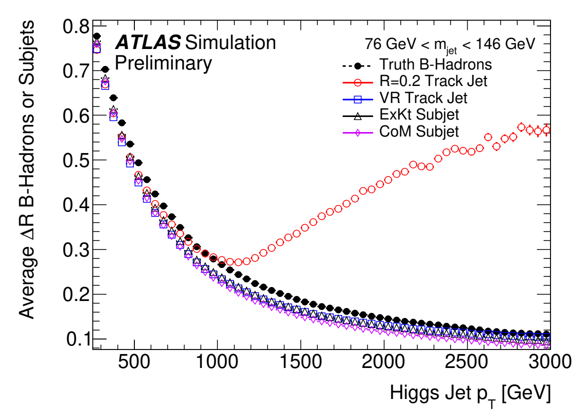
Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction techniques [10]



$X \to b\bar{b}$ Advanced Taggers: Giving ML Flavour Taggers Good Information [11]



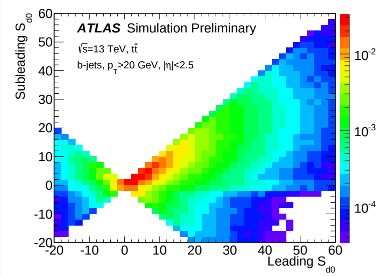
Efficiency for a Higgs jet to have 2 of the leading 3 associated subjets matched to truth b-hadrons

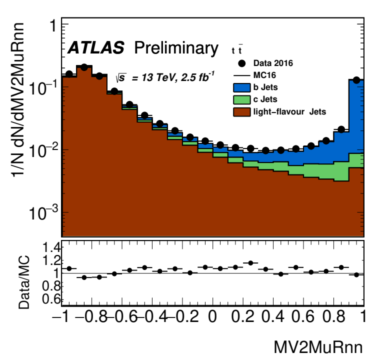


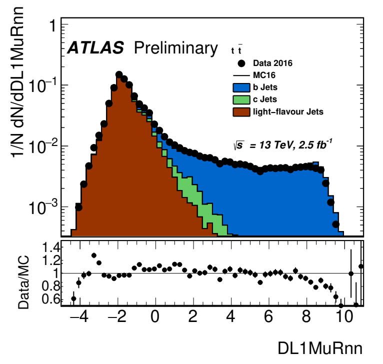
ΔR between leading two truth b-hadrons or subjets associated to Higgs jets as a function of Higgs jet p_T

Summary

- Robust low level taggers
  - IP2D, IP3D, SV1, JetFitter, RNNIP, SMT
- Powerful and versatile high level taggers
  - MV2 (BDT) and DL1 (DNN)
- Hybrid training on $t\bar{t}$ and Z' for big improvement in high p_T response
- Enabling critical physics results in Run 2 and driving forward-looking physics taggers for Run3





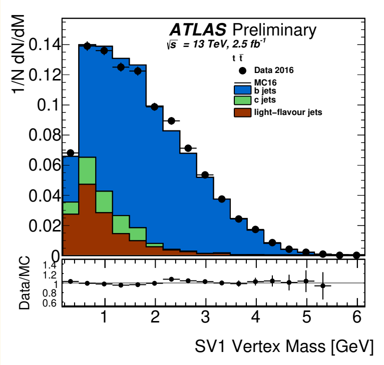


Backup

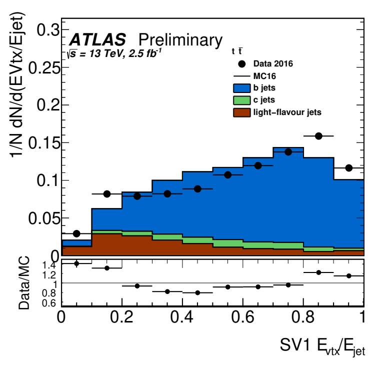
Displaced secondary vertices (SV): SV1 [4]

Expectation of a secondary vertex from either *b* or *c*-hadron decays

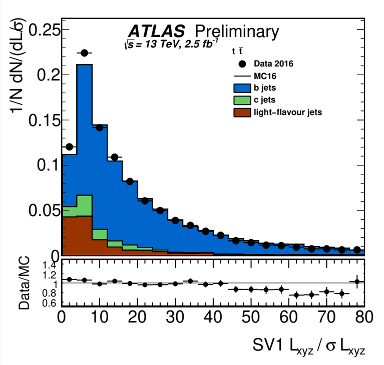
- Invariant mass of tracks at SV used to discriminate *b* or *c*-hadron decay vertices from V^0 decays or material interactions
- hard *b*-jet fragmentation: SV carries large fraction of jet energy
- SV found in most *b*-jets, but only in a few % of light flavour jets



Mass of tracks at secondary vertex



Energy fraction of secondary vertex tracks

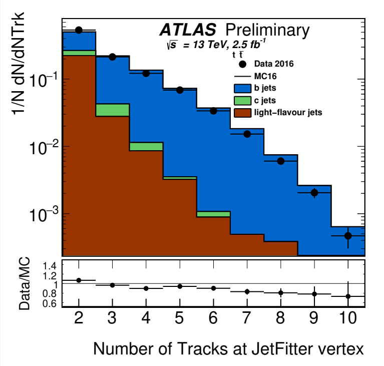


Significance of 3D decay distance

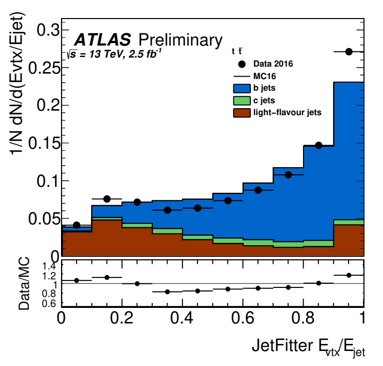
Decay Chain: JetFitter [4]

Most *b*-jets contain a *c*-jet, so use **Kalman filter** (JetFitter) to search for common axis for 3 vertices: primary *pp*, *b*-jet, *c*-jet

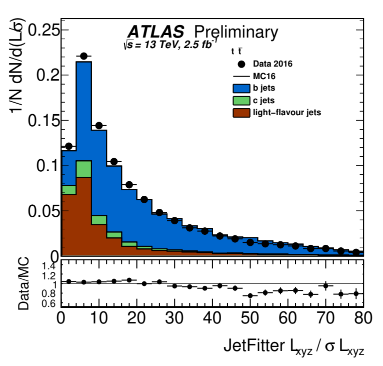
- Look for "1 track vertices" with decay chain axis, improving efficiency
- Constraint to the decay axis further improves power of SV based discriminants



N tracks associated with a JetFitter secondary vertex



Energy fraction of JetFitter secondary vertex tracks



Significance of 3D decay distance

Overview

Selection of lower level taggers:

Impact Parameter (IP) based

IP2D, IP3D:
Log-likelihood ratios using flavour hypotheses computed from summed track contributions extracted from simulation-derived templates

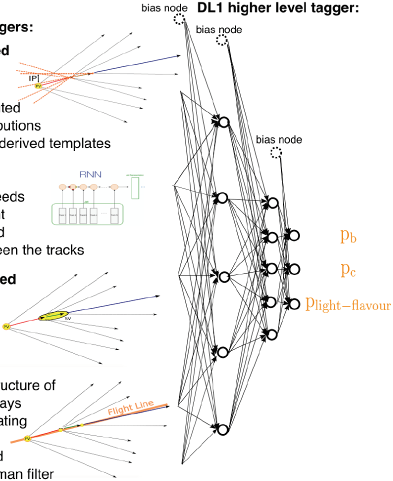
RNNIP:
Parallel approach which feeds raw tracks into a Recurrent Neural Network (RNN) and exploits correlations between the tracks

Secondary Vertex (SV) based

SV1:
Reconstructs inclusive secondary vertices

JetFitter:
Exploits the topological structure of weak b- and c-hadron decays inside the jet by approximating the b-hadron or c-hadron flight path with PV, SV and tertiary vertex using a Kalman filter

DL1 higher level tagger:



Also used:
- Jet kinematics
- Information on muons produced in b/c decays

Outline

References I

[1] **ATLAS** Collaboration, M. Aaboud *et al.*, "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector," *Phys. Lett.* **B786** (2018) 59–86, arXiv:1808.08238 [hep-ex].

[2] A. Chisholm, "Introduction to Heavy Flavour Jet Tagging with ATLAS," 2017. https://indico.cern.ch/event/655628/contributions/2670400/. ATLAS Higgs to $b\bar{b}$ / Flavor Tagging Workshop.

[3] A. Chisholm, "Calibration of light flavour jets using "Flipped" taggers and $Z+$ jet events," 2018. https://indico.cern.ch/event/734385/contributions/3028829/. Flavour Tagging Plenary Meeting.

[4] **ATLAS** Collaboration, "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run," Tech. Rep. ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul, 2017. http://cds.cern.ch/record/2273281.

[5] W. Hopkins, "Jets and Machine Learning in ATLAS," 2018. https://indico.cern.ch/event/745718/contributions/3146637/. Machine Learning for Jet Physics Workshop.

[6] **ATLAS** Collaboration, "Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment," Tech. Rep. ATL-PHYS-PUB-2017-003, CERN, Geneva, Mar, 2017. http://cds.cern.ch/record/2255226.

[7] **ATLAS** Collaboration, M. Aaboud *et al.*, "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV," *JHEP* **08** (2018) 089, arXiv:1805.01845 [hep-ex].

[8] **ATLAS** Collaboration, M. C. Lanfermann, "Deep Learning in Flavour Tagging at the ATLAS experiment,". https://cds.cern.ch/record/2274065.

References II

[9] M. C. Lanfermann, "NN Optimisation for Flavour Tagging in ATLAS,". https://erez.weizmann.ac.il/pls/htmldb/f?p=101:58:::NO:RP:P58_CODE,P58_FILE:5401,Y. Hammers & Nails - Machine Learning & HEP Workshop.

[10] M. Stankaityte, "$X \to b\bar{b}$ tagging studies in rel21," 2018. https://indico.cern.ch/event/767505/contributions/3189608/. Jet Substructure and Jet-by-Jet Tagging Meeting.

[11] **ATLAS** Collaboration, "Variable Radius, Exclusive-k$_T$, and Center-of-Mass Subjet Reconstruction for Higgs($\to b\bar{b}$) Tagging in ATLAS," Tech. Rep. ATL-PHYS-PUB-2017-010, CERN, Geneva, Jun, 2017. https://cds.cern.ch/record/2268678.