January 22, 2018

Machine Learning for flavour-tagging

In the ATLAS experiment

Geoffrey GILLES on behalf of the ATLAS collaboration

Bergische Universität Wuppertal

Introduction

BERGISCHE UNIVERSITÄT WUPPERTAL

- Heavy-flavour tagging (*b*-tagging) is an important tool for physics analysis
  - Intensively used for signal identification, background suppression
    → Measurements in top-quark or Higgs-boson sectors, searches for New Physics, etc.

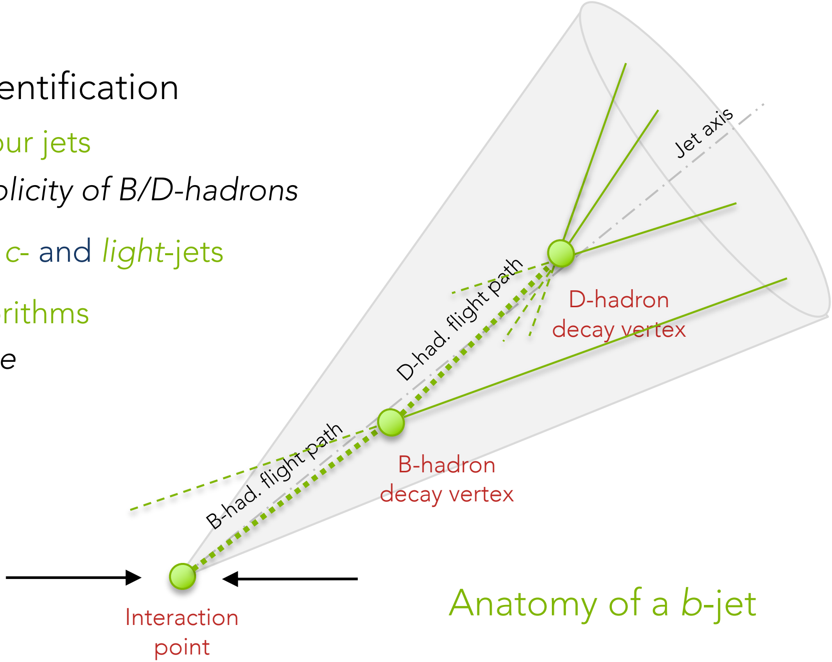
"Where Machine Learning techniques play a central role"

- Outline of this presentation
  - Illustrate some lessons learned from using Machine Learning for flavour-tagging in ATLAS
  - Through three examples

    1. Extracting more information from data illustrated by impact parameter based taggers
    2. The power of combining information illustrated by secondary vertex finding algorithms
    3. State of the art of high-level flavour-tagging algorithms used in ATLAS

Identifying heavy-flavour jets

- Inclusive approach for heavy-flavour jet identification
  - Exploiting specific topology of heavy-flavour jets
    → *Long lifetime, high mass & decay multiplicity of B/D-hadrons*
  - Performance led by power to separate *b*-, *c*- and *light*-jets
  - Using dedicated and complementary algorithms
    → *With different complexity & performance*



Anatomy of a *b*-jet

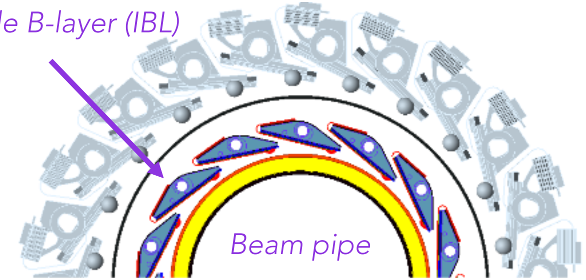
Ingredients for *b*-tagging

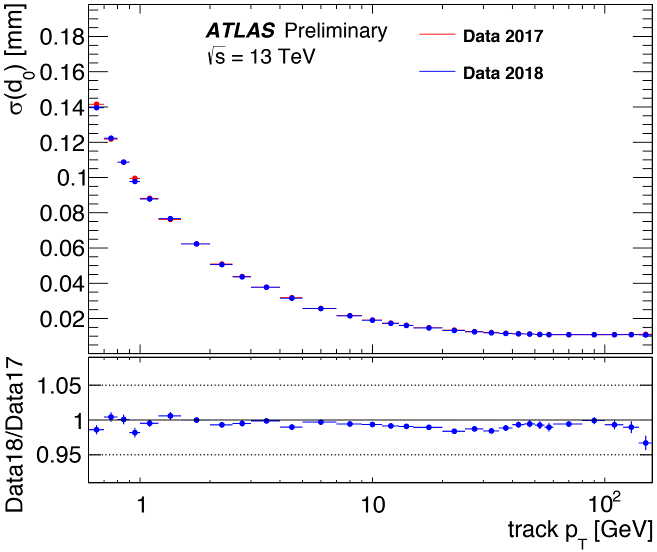
- Tracks
  - Tracks-to-jets association based on p_T dependent ΔR selection
  - IP resolution determined by first two pixel detector layers
    → Crucial to distinguish *B*-hadron decay from fragmentation tracks
  - Primary and secondary vertex reconstruction also a key ingredient
    → Main challenge due high pile up condition at pp collider

    "The cleaner environment at e^+e^- collider is an advantage"

- Jets
  - Reconstruction algorithm → *To deal with different jet environments*
  - Direction → *To assign a "lifetime sign" to tracks*
  - Exploit physics properties & detector resolutions dependencies

- Leptons
  - Used to identify semi-leptonic *B*-hadron decays





Extracting more information from data

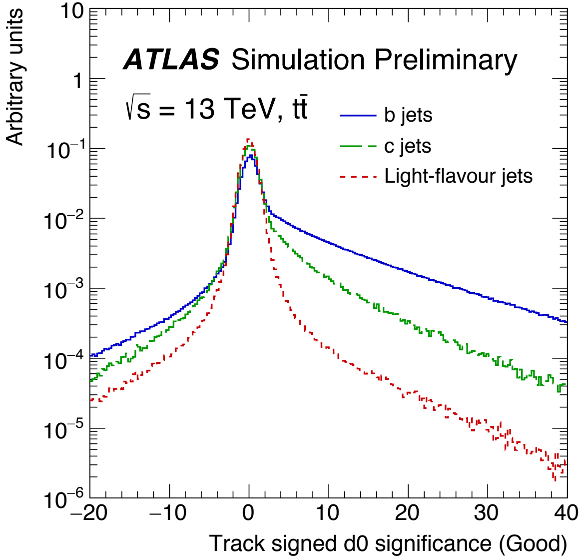
Illustrated by Impact Parameter (IP) based taggers

Impact parameter based tagger

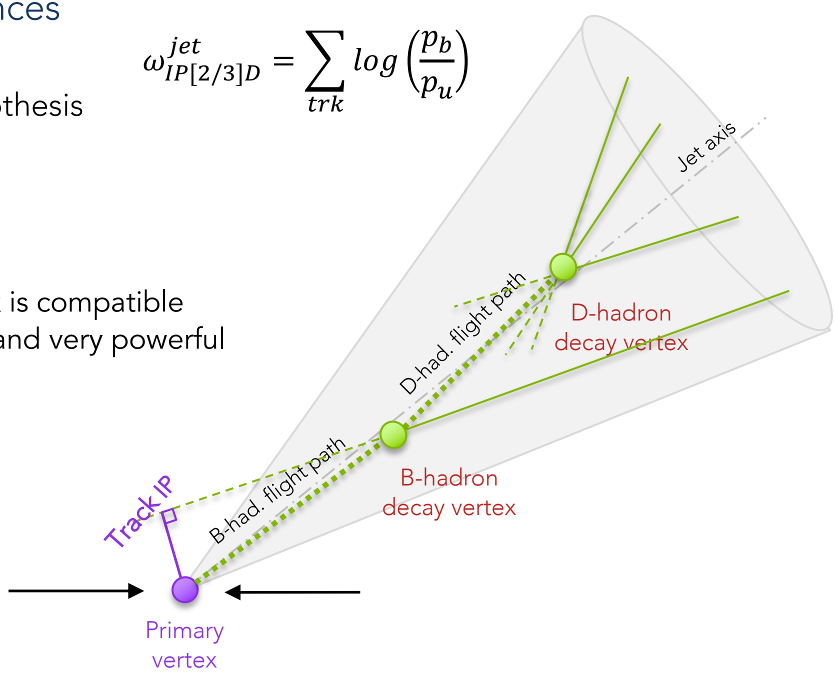
ATL-PHYS-PUB-2016-012

- IP[2/3]D: Uses d_0/z_0 impact parameter significances
  - Track categorization based on pixel hit patterns
  - Likelihood ratios between *b*-, *c*- and *light*-jet hypothesis computed as sum of per-track log probabilities

$$\omega^{jet}_{IP[2/3]D} = \sum_{trk} log\left(\frac{p_b}{p_u}\right)$$



IP significances
Tell you if a track is compatible with PV. Simple and very powerful

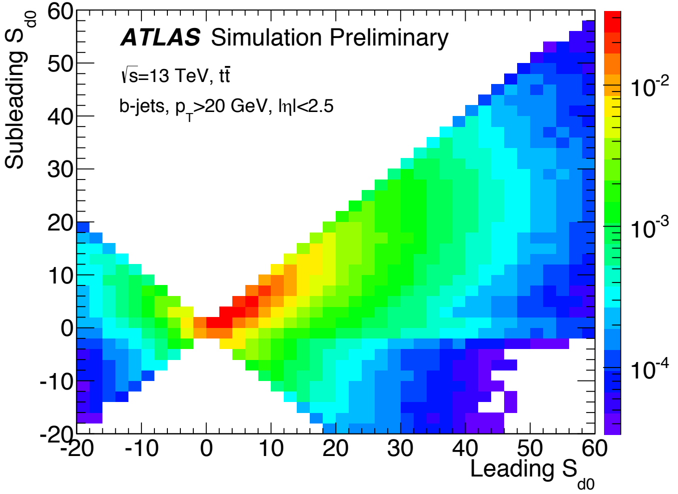


NB: Template probability density function obtained form reference histograms and derived from MC simulation

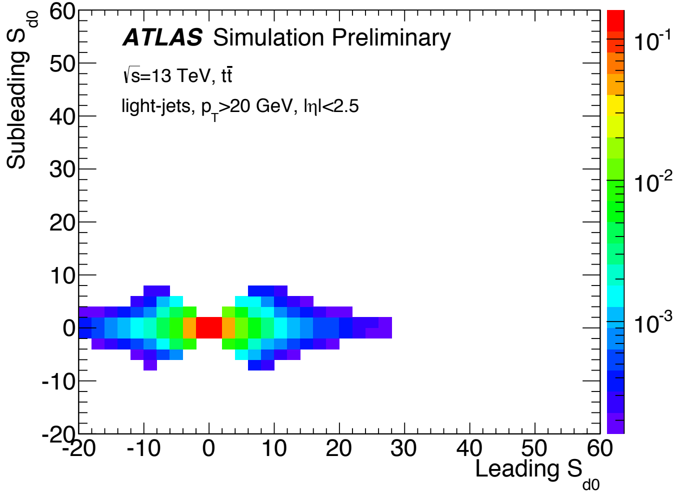
Impact parameter based tagger

"IPTag relies on track compatibility with PV but not sensitive to track correlations"

b-jets



Light-jets



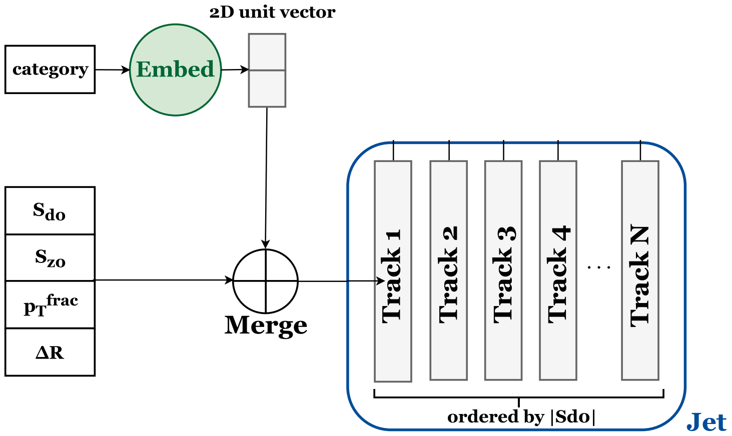
Possibility to capture track correlations using sequence classification in Machine Learning

Sequence Classification & Recurrent Neural Network

ATL-PHYS-PUB-2017-003

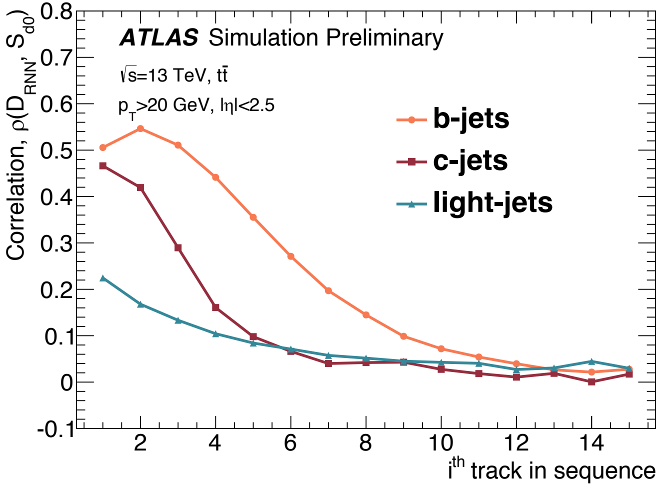
Build sequence of track, imposing physics-inspired ordering

Exploit kinematic information



Track sequence feeding RNN

Correlations of per track S_{d0} with NN output

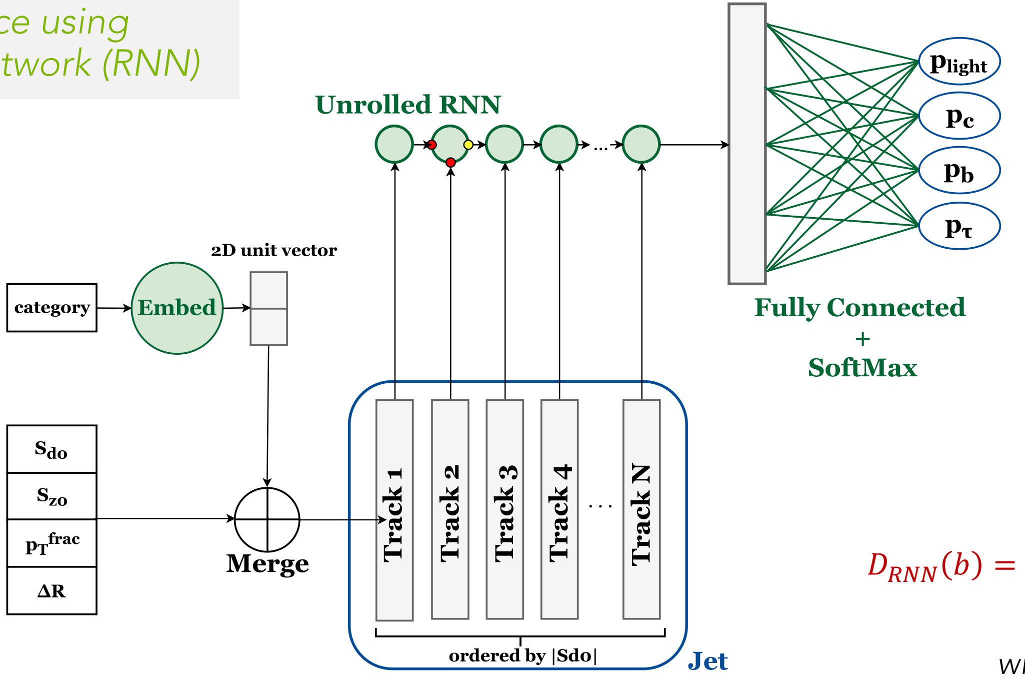


Sequence Classification & Recurrent Neural Network

ATL-PHYS-PUB-2017-003

Analyse sequence using Recurrent Neural Network (RNN)

Exploit kinematic information



Track sequence feeding RNN

Flavour-tagging probability outputs combined into discriminating function

$$D_{RNN}(b) = \ln \frac{P_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_{light}}$$

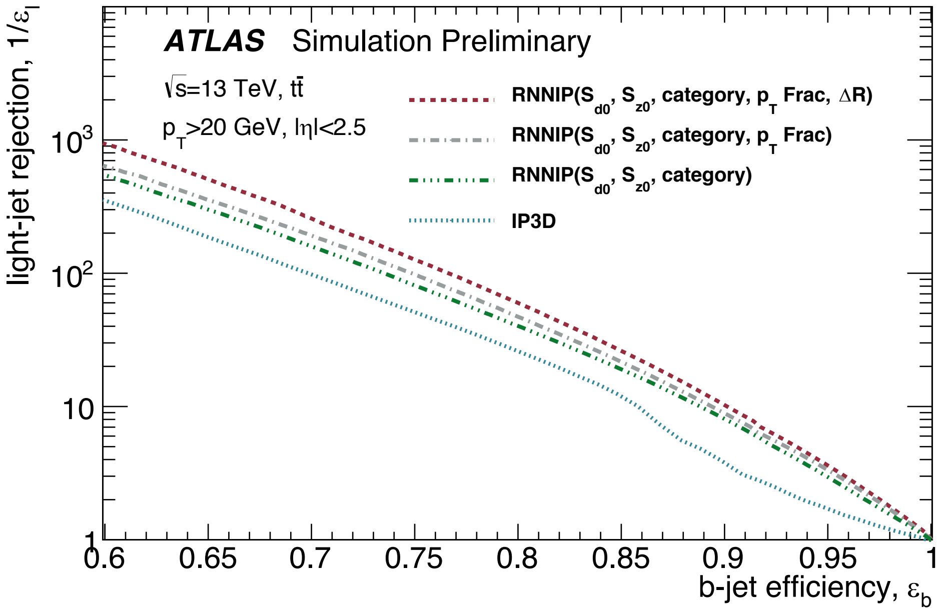
with $f_c = 0.07$ *and* $f_\tau = 0$

Comparing *b*-tagging performance

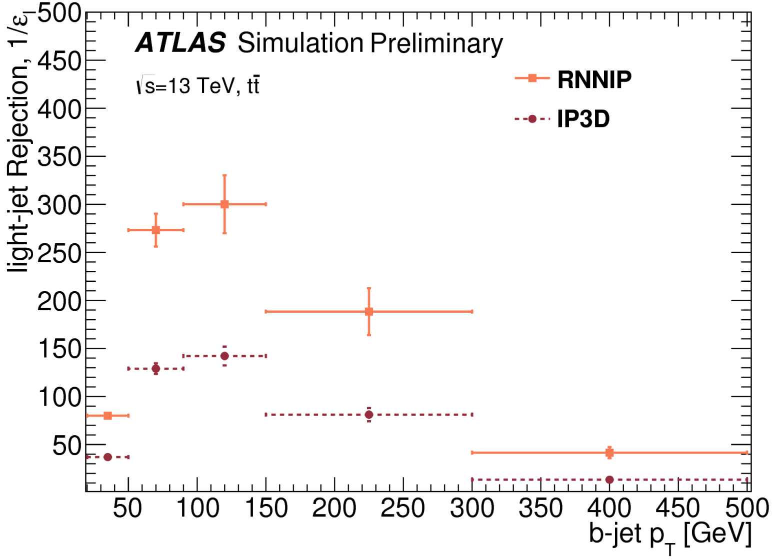
ATL-PHYS-PUB-2017-003

"RNNIP more powerful than naïve Bayes model especially for intermediate to high-p_T jets"

ROC curves: e.g. light-jet rejection vs. b-jet efficiency



Light-jet rejection vs. p_T
(at 70% b-jet efficiency)



The power of combining information

Illustrated by secondary vertex finding algorithms

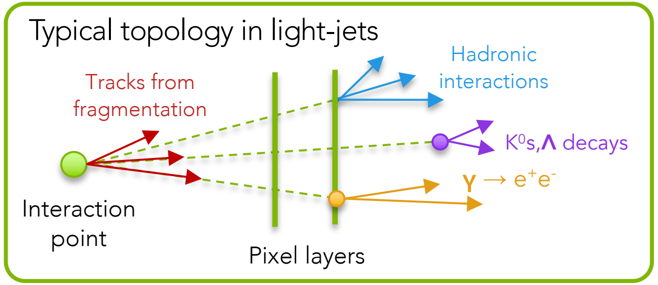
Towards displaced vertex reconstruction

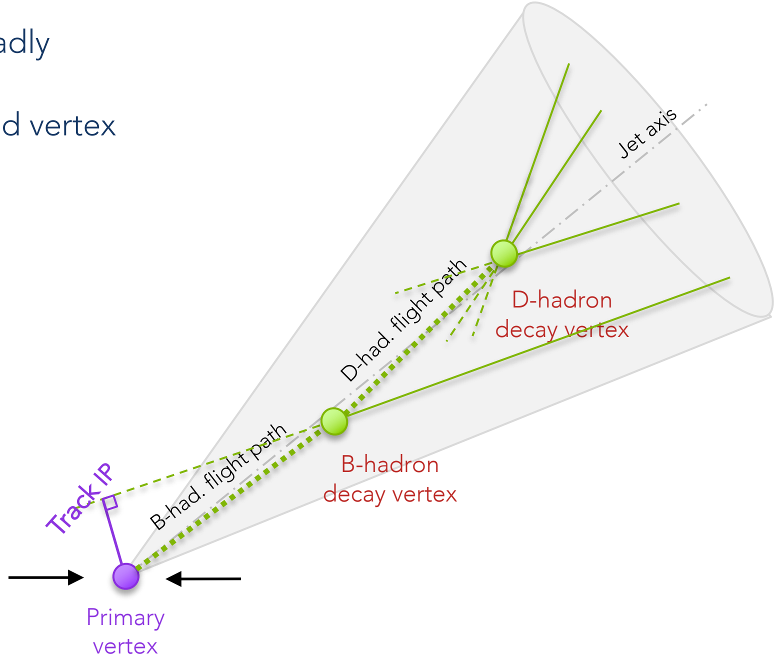
BERGISCHE
UNIVERSITÄT
WUPPERTAL

- What to expect from secondary vertex finding in Jets?
  - Disentangle high-IP tracks belonging to real verteces from badly reconstructed fragmentation tracks
  - Tracks with low IP parameter can still contribute to a displaced vertex
  - A way reduce number of light-jets faking b-jets considerably

"Complementary to IP-based algorithms"

⚠ *There are also real vertices in light-jets !*



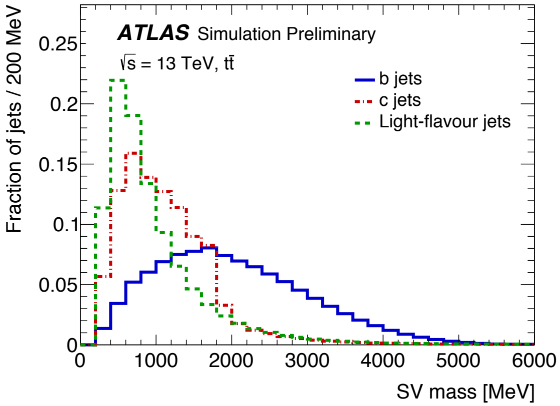


1^{st} step in secondary vertex reconstruction

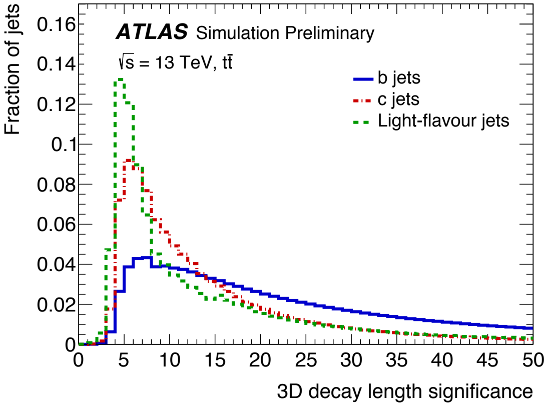
ATL-PHYS-PUB-2017-011

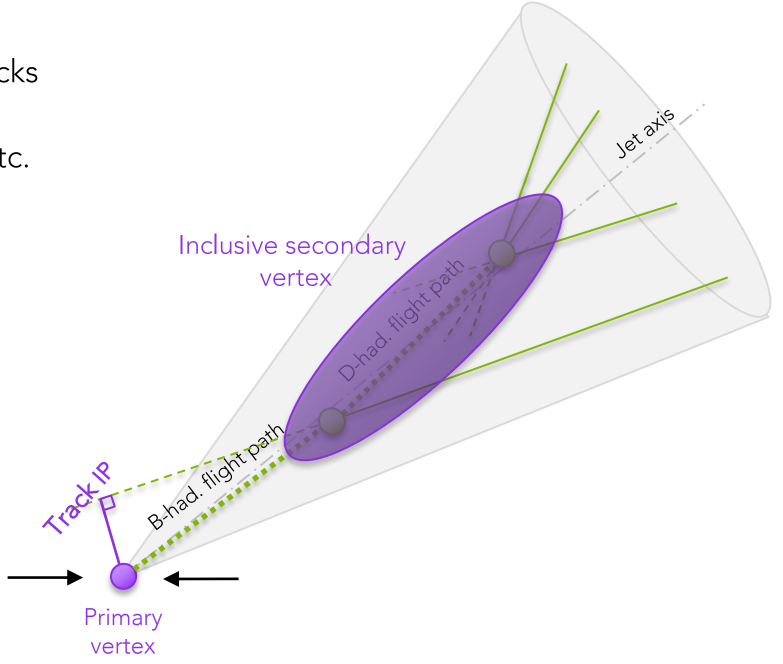
- SV1: Inclusive displaced 2^{nd} vertex reconstruction
  - Fit tracks from 2-track vertices candidates, discarding tracks from Λ/K^0s decays, conversions and material interactions
  - Summarize important vertex infos: decay length, mass, etc.

2^{nd} vertex invariant mass



2^{nd} vertex decay length



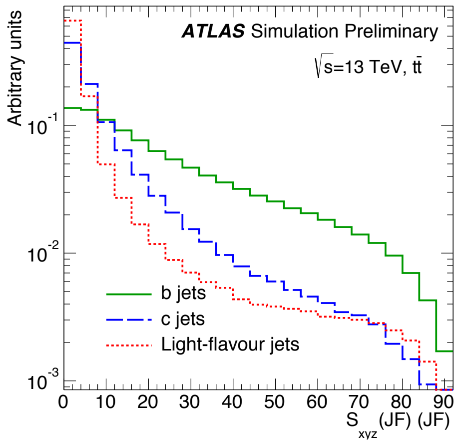


… incomplete topologies like only 1 track from B or D hard to reconstruct

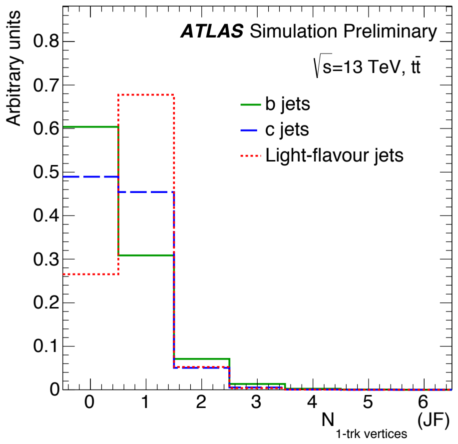
Decay chain multi-vertex reconstruction

ATL-PHYS-PUB-2018-025

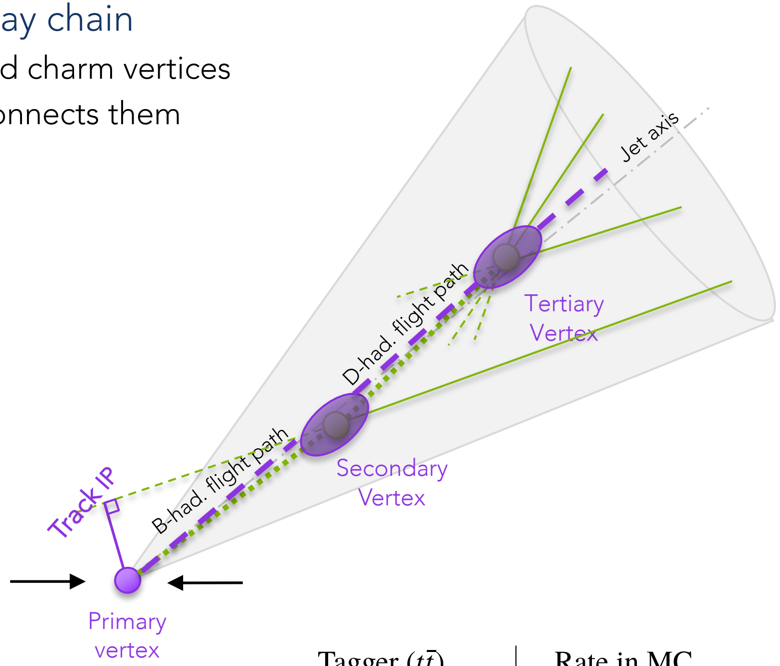
- JetFitter : Exploits topological structure of B/D-hadron decay chain
  - Kalman filter used to find common line through PV, bottom and charm vertices
  - Ability to reconstruct vertices even when only a single track connects them
  - Summarize important vertex infos: decay length, mass, efc …



Improve discrimination between b- and c-jets



1-trk vertices very sensitive to pile up
A difficulty at pp colliders but less at e^+e^- colliders



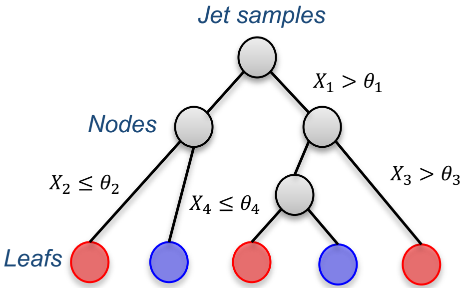
Tagger ($t\bar{t}$)	Rate in MC
SV1	0.187±0.001
JF secondary vertex	0.217±0.001
JF tertiary vertex	0.0504±0.0005

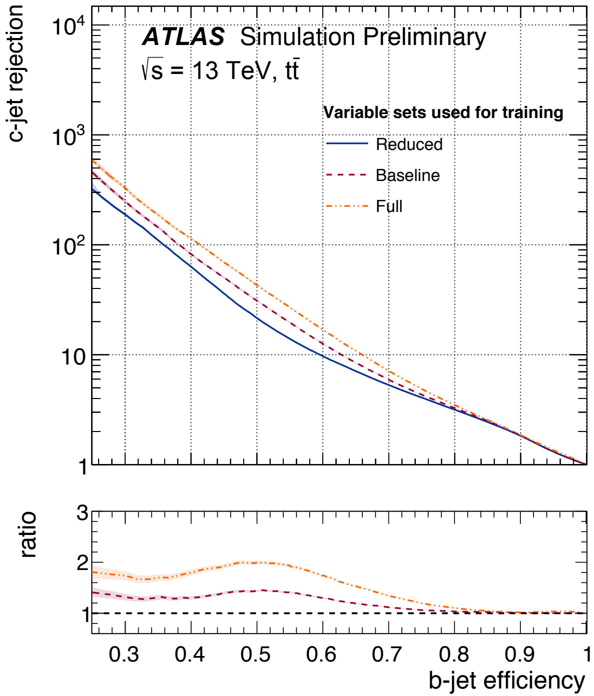
Building a multivariate *b*-tagging algorithm

ATL-PHYS-PUB-2018-025

e.g. Combining JetFitter algorithm outputs using TMVA Boosted Decision Trees (BDT)

Input	Variable	
Kinematics	p_T η	*Inclusive properties*
Reduced	m f_E $\Delta R(\vec{p}_{\rm jet}, \vec{p}_{\rm vtx})$ $S_{\rm xyz}$ $N_{\rm TrkAtVtx}$	
Baseline	above variables + $N_{\rm 2TrkVtx}$ $N_{\rm 1\text{-}trk\ vertices}$ $N_{\rm \geq 2\text{-}trk\ vertices}$	*Multi-vertex topology*
Full	above variables + $L_{xyz}(2^{\rm nd}/3^{\rm rd}{\rm vtx})$ $L_{xy}(2^{\rm nd}/3^{\rm rd}{\rm vtx})$ $m_{\rm Trk}(2^{\rm nd}/3^{\rm rd}{\rm vtx})$ $E_{\rm Trk}(2^{\rm nd}/3^{\rm rd}{\rm vtx})$ $f_E(2^{\rm nd}/3^{\rm rd}{\rm vtx})$ $N_{\rm TrkAtVtx}(2^{\rm nd}/3^{\rm rd}{\rm vtx})$	*Full topological decay chain information*





"The amount of combined information is a key ingredient"

NB: Even better improvement on light-jet rejection (c.f. back-up slides)

State of the art

High-level flavour tagging algorithms

High-level flavour-tagging algorithms in ATLAS

☞ Using MVA

ATL-PHYS-PUB-2017-013

IP2D/IP3D
IP based taggers

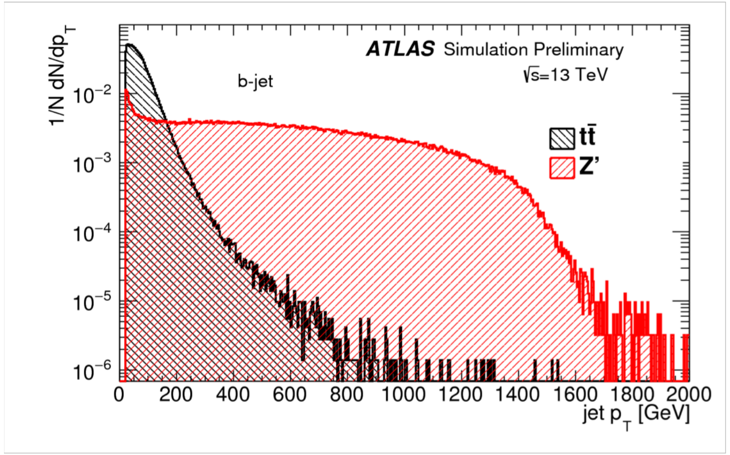
☞ **SMT**
Semi-leptonic decays to muons

SV1
Inclusive secondary vertex reconstruction

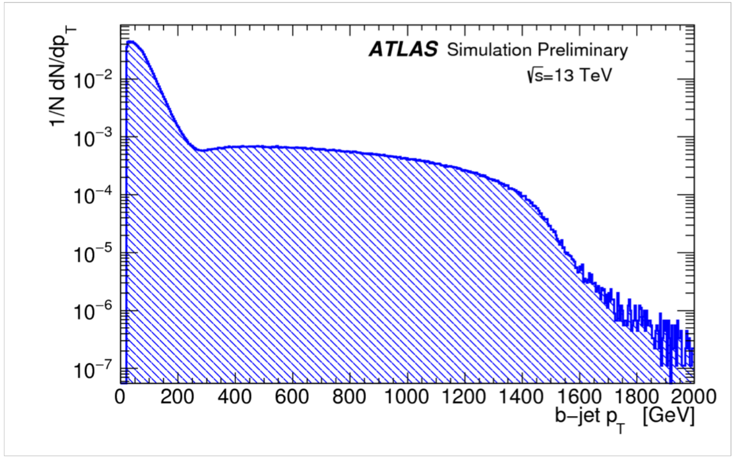
☞ **IPRNN**
Recurrent NN on track parameters

JetFitter
Topological *b*-hadron decay reconstruction

☞ **MV2** / ☞ **DL1**
High level combinations using BDT / Deep NN
for *b*- and *c*-tagging



Hybrid training
Exploit jets from $t\bar{t}$ and broad $Z' \rightarrow q\bar{q}, c\bar{c}, b\bar{b}$

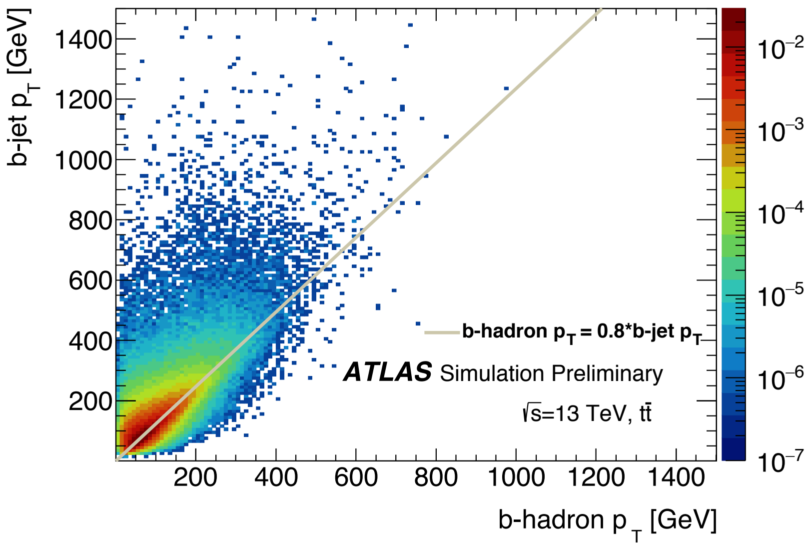


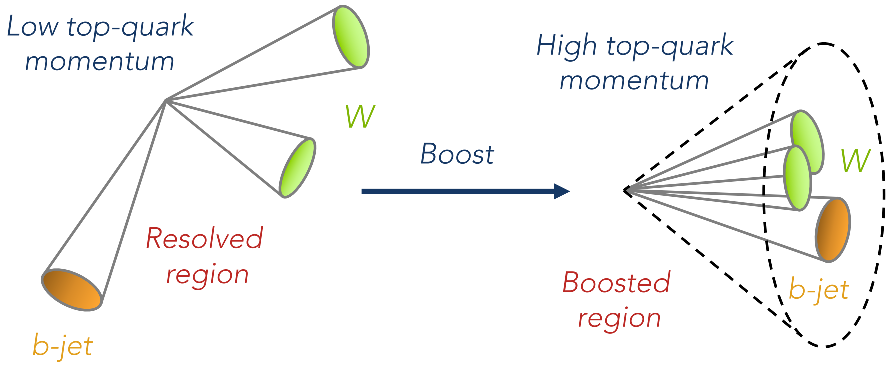
Hybrid training sample

ATL-PHYS-PUB-2017-013

"Correlation between jet p_T & heavy-hadron p_T is crucial to flavour-tagging performance"

$t\bar{t}$ events





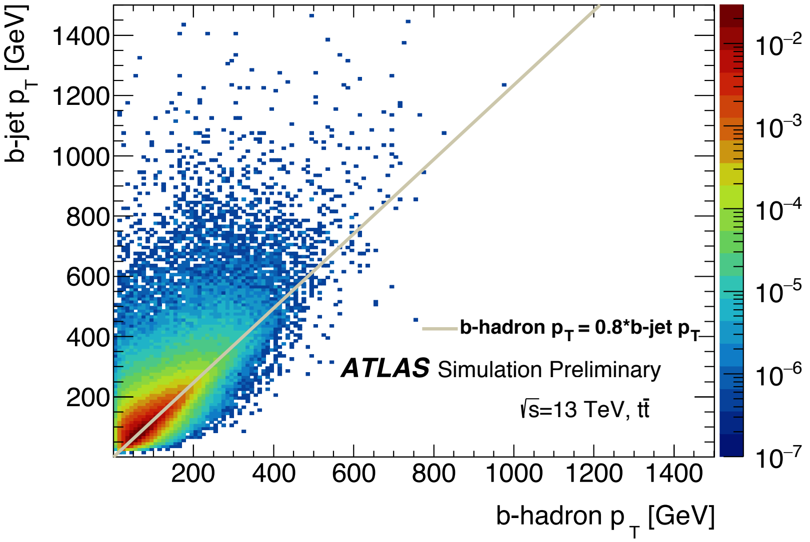
Above m_t, jet p_T determined by nearby hadronic activity
→ reducing correlation with b-hadron p_T

Hybrid training sample

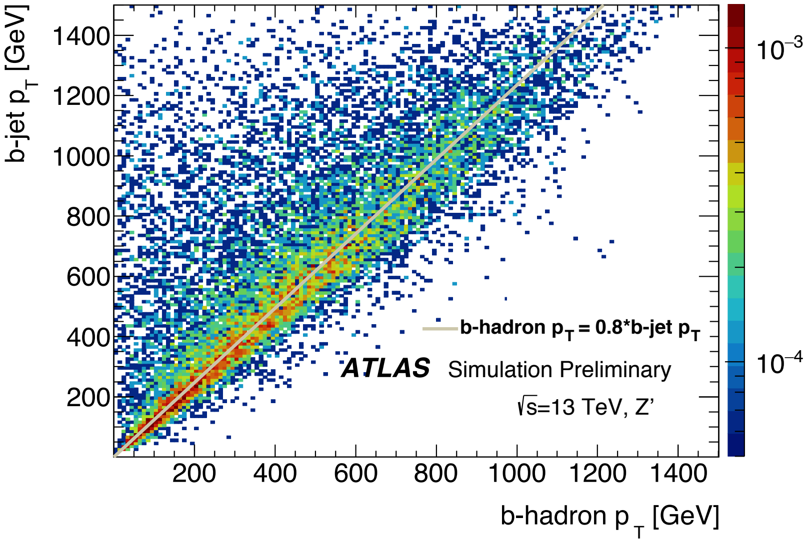
ATL-PHYS-PUB-2017-013

"Correlation between jet p_T & heavy-hadron p_T is crucial to flavour-tagging performance"

$t\bar{t}$ events



Z' events



Above m_t, jet p_T determined by nearby hadronic activity
→ reducing correlation with b-hadron p_T

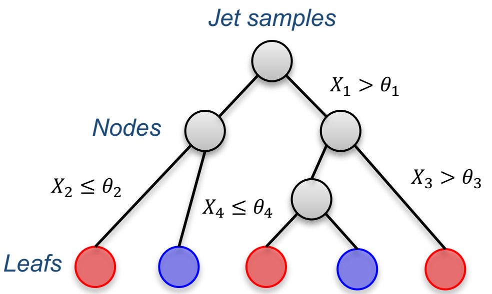
New strategy developed to probe high-p_T regime using broad Z' events

The MV2 Algorithm

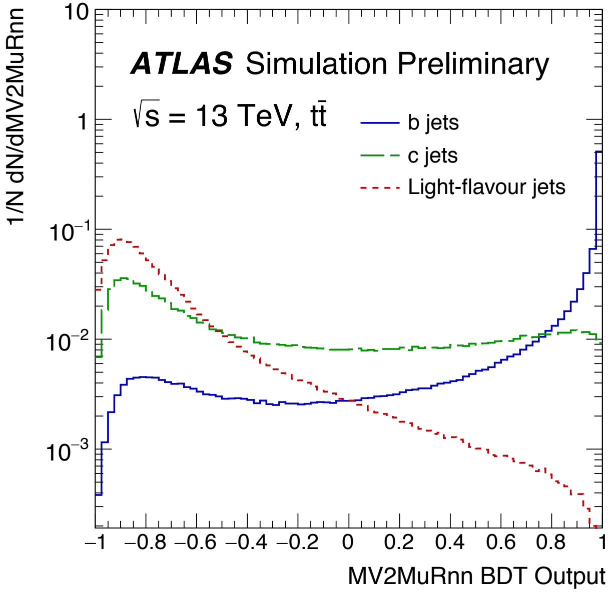
ATL-PHYS-PUB-2017-013

"ATLAS Run II reference b-tagging algorithm"

Combines baseline tagger outputs using TMVA Boosted Decision Tree (BDT)



e.g. MV2MuRNN output discriminant



Including Jet p_T / η to take advantage of correlations between input variables

→ Signal (*b*) reweighted to match background (*c+light*) spectra

Three main scenarios for *b*-tagging

MV2 = IPTag + SV1 + JetFitter

MV2Mu = IPTag + SV1 + JetFitter + SMT

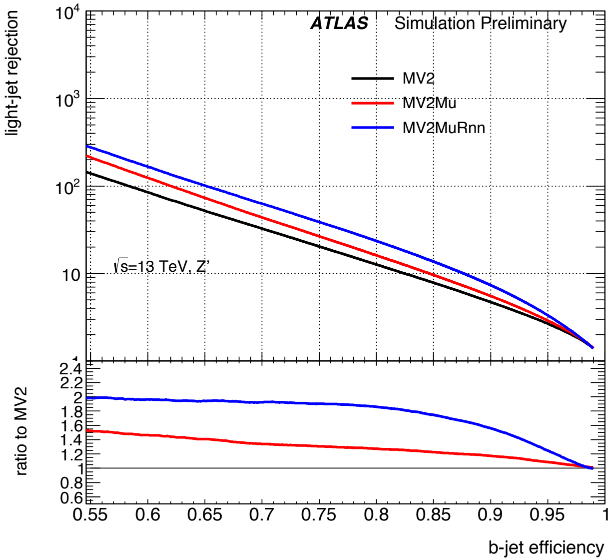
MV2MuRNN = IPTag + SV1 + JetFitter + SMT + RNNIP

Optimized BDT parameterization & c-jet fraction to obtain best performance

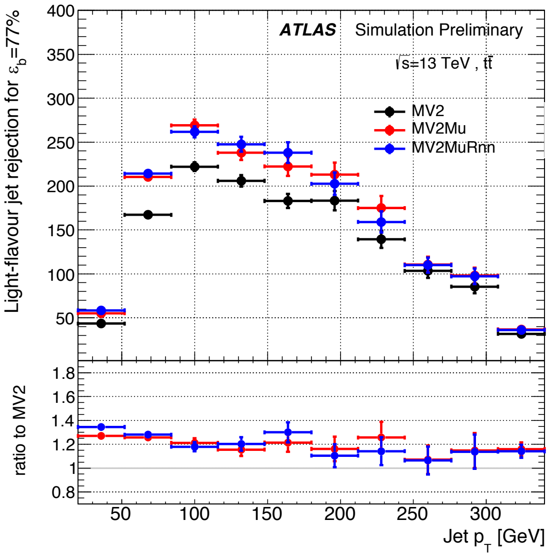
The MV2 Algorithm

ATL-PHYS-PUB-2017-013

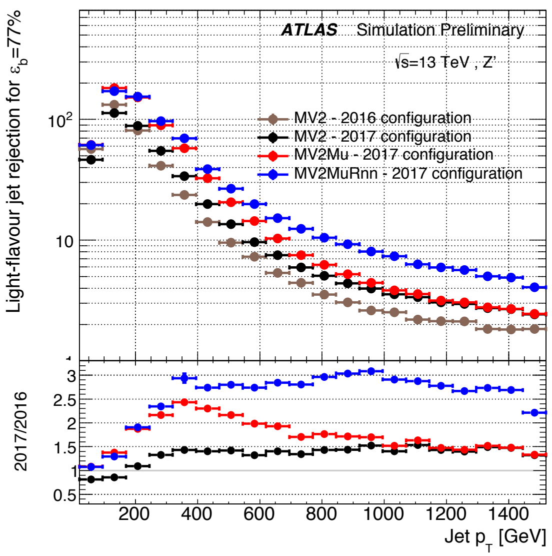
"Machine Learning exploits complementarity of baseline taggers to extract better performance "



Illustrate how combining information maximalises algorithm performance



Soft-Muon Tagger improves performance at low p_T



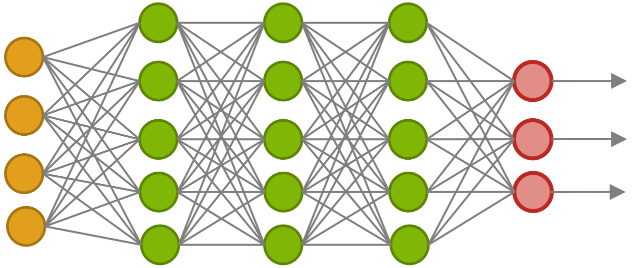
RNNIP improve performance at high p_T

High level tagger using Deep Learning

ATL-PHYS-PUB-2017-013

"DL1: Deep Feed Forward Neural Network"

Factorises learning of structures in data across many layers



Same input information as in MV2

↓

Results in comparable performance

Multidimensional outputs combined into discriminating functions

↓

All flavours treated equally during training offering large flexibility →

$$D_{RNN}(b) = \ln \frac{P_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{light}} \quad \text{b-tagging}$$

$$D_{RNN}(c) = \ln \frac{P_c}{f_b \cdot p_b + (1 - f_b) \cdot p_{light}} \quad \text{c-tagging}$$

NB: Trained with Keras with Theano backend & Adam Optimizer

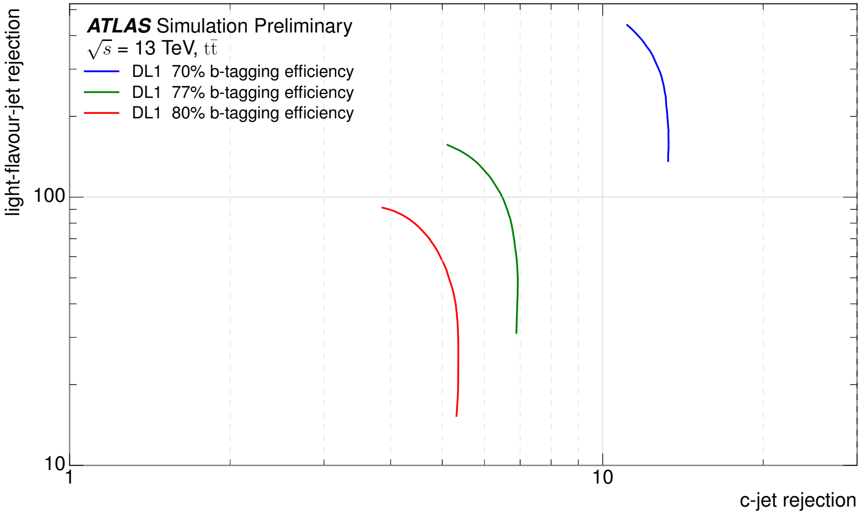
High level tagger using Deep Learning

ATL-PHYS-PUB-2017-013

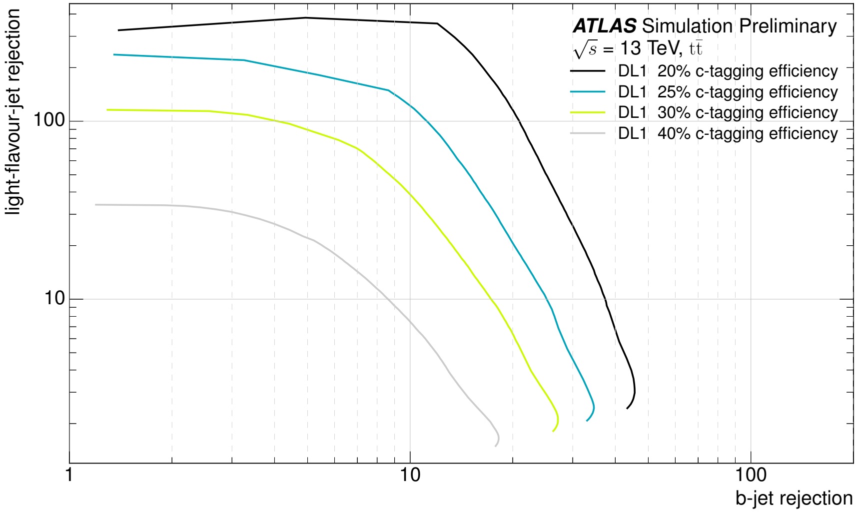
"Powerful flexibility of applications and optimisation"

Final c- and b-jets fractions chosen a posteriori according to desired performance

b-tagging



charm-tagging



Conclusions

Conclusions

- Over the years, Machine Learning became essential for flavour tagging
  - Developing a general strategy based on combining complementary information
  - To cope with the complexity of the exercise, going further in improving the performance

- ATLAS exploits five baseline approaches to extract jet flavour information
  - Making use of impact parameter based taggers, 2nd vertex finders, or exploiting soft muons
  - Combined in powerful high-level tagging algorithms used at physics analysis level

  → e.g. BDT based MV2 algorithm
  - Algorithms trained on hybrid $t\bar{t} + Z'$ event sample

  → Topology and kinematics of training samples are important

- Deep Learning technics offer new & powerful paradigm for Machine Learning
  - To extract further information from data as illustrated by RNNIP
  - Offer a large flexibility of applications and optimisation like DL1 compared to MV2

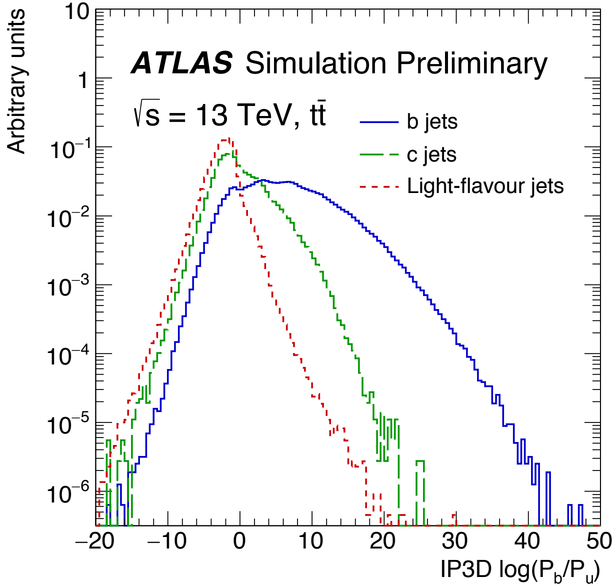
"Keeping in mind that the amount of information exploited drives the performance"

Back up

IP track categories & IP3D log-likelihood ratio

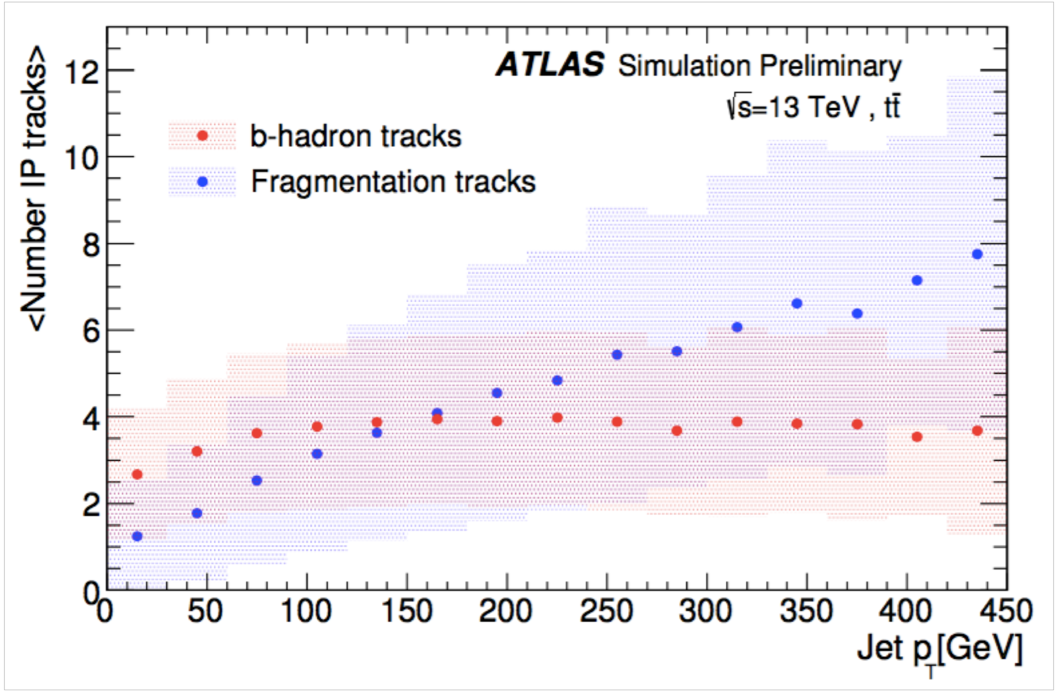
ATL-PHYS-PUB-2016-012

#	Category	Fractional contribution [%] *b*-jets	*c*-jets	light-jets
0	No hits in first two layers; expected hit in IBL and b-layer	1.9	2.0	1.9
1	No hits in first two layers; expected hit in IBL and no expected hit in b-layer	0.1	0.1	0.1
2	No hits in first two layers; no expected hit in IBL and expected hit in b-layer	0.04	0.04	0.04
3	No hits in first two layers; no expected hit in IBL and b-layer	0.03	0.03	0.03
4	No hit in IBL; expected hit in IBL	2.4	2.3	2.1
5	No hit in IBL; no expected hit in IBL	1.0	1.0	0.9
6	No hit in b-layer; expected hit in b-layer	0.5	0.5	0.5
7	No hit in b-layer; no expected hit in b-layer	2.4	2.4	2.2
8	*Shared* hit in both IBL and b-layer	0.01	0.01	0.03
9	At least one *shared* pixel hits	2.0	1.7	1.5
10	Two or more *shared* SCT hits	3.2	3.0	2.7
11	*Split* hits in both IBL and b-layer	1.0	0.87	0.6
12	*Split* pixel hit	1.8	1.4	0.9
13	*Good*	83.6	84.8	86.4



Impact of fragmentation tracks

ATL-PHYS-PUB-2016-012



The JetFitter Algorithm

ATL-PHYS-SLIDE-2017-559

1 – Track pre-selection

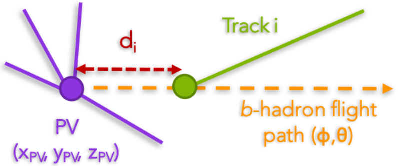
- Minimise particles not originating from b- or c-hadron decays.

→ *Reject tracks compatible with PV, K^0s/Λ^0 decays, γ conversions or material interaction*

→ *p_T>500 MeV, $|IP_{r\phi}|$<3.5mm & $|IP_z|$<5mm*

2 – Initialisation

- First fit performed under the hypothesis that each selected track represents a single vertex along *b*–hadron flight path, initialised to the jet axis.

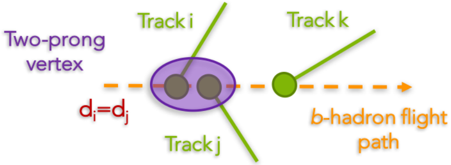


$$\vec{d} = (x_{PV}, y_{PV}, z_{PV}, \phi, \theta, d_1, d_2, \dots, d_N)$$

3 – Clustering

- Two vertices with highest probability to form a common vertex which lies on the *b*-hadron flight axis are merged.

→ *All two-vertices combinations considered*



- Prior to decay chain fit.
- Additional track selection applied.

→ *p_T>750 MeV, $|IP_{r\phi}|$<1.5mm & $|IP_z|$<3mm*

→ *$|IP_{r\phi}|/\sigma(IP_{r\phi})$<5 and $|IP_z|/\sigma(IP_z)$>2*

Against pileup

4 – Vertex fitting

- Track parameters of individual tracks iteratively added as input to the measurement equation, specifying the vertex number to be updated.

→ *KF formalism extension*

$$\vec{q} = \hat{\vec{C}} + \hat{A}\vec{d} + \hat{B}\vec{p}$$

- Updated decay chain $\vec{d}$ provided after each iteration.

5 – Vertex finding

- Highest compatible vertices merged

→ *Similarly to the first clustering step*

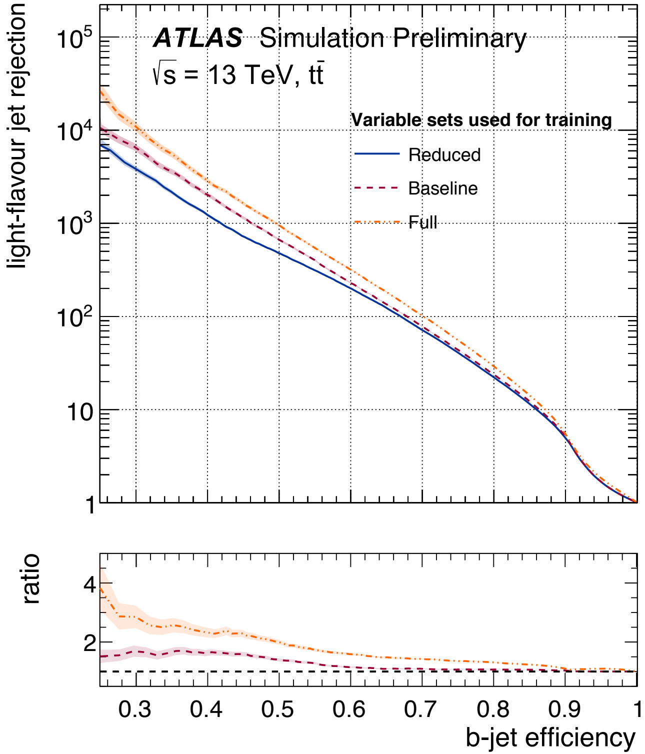
- Then a new complete fit performed
- Procedure iterates until no vertex pairs with a probability above a certain threshold exist anymore

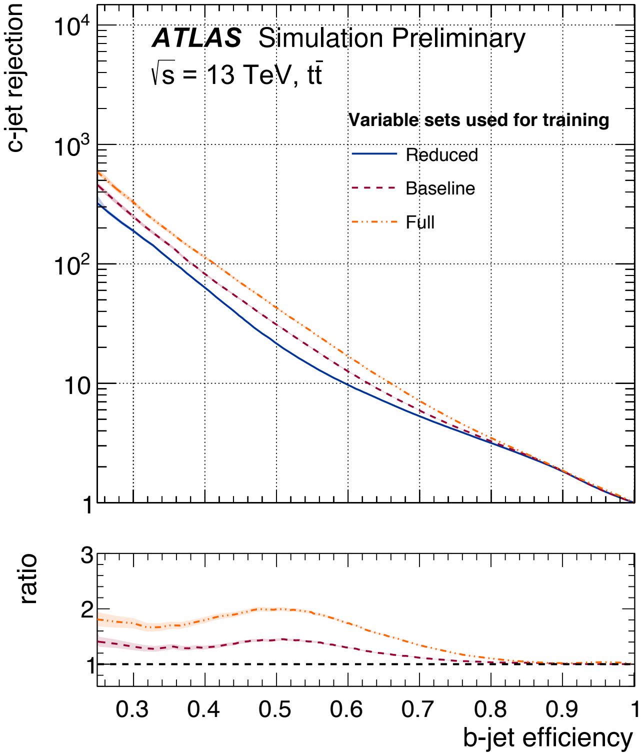
→ *Well defined association of tracks to vertices, with at least one track each*

Performance of JetFitter BDT discriminant

BERGISCHE UNIVERSITÄT WUPPERTAL

ATL-PHYS-PUB-2018-025





MV2 input variables

BERGISCHE UNIVERSITÄT WUPPERTAL

ATL-PHYS-PUB-2016-012

Input	Variable	Description
Kinematics	$p_T(jet)$	Jet transverse momentum
	$\eta(jet)$	Jet pseudo-rapidity
IP2D, IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio between the b- and light jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{\text{light}})$	Likelihood ratio between the c- and light jet hypotheses
SV	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion masses
	$f_{\text{E}}(\text{SV})$	Fraction of the charged jet energy in the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{\text{2TrkVtx}}(\text{SV})$	Number of two track vertex candidates
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertices
	$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertices
	$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
	$\Delta R(\text{jet}, \text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
Jet Fitter	$N_{\text{2TrkVtx}}(\text{JF})$	Number of 2-track vertex candidates (prior to decay chain fit)
	$m(\text{JF})$	Invariant mass of tracks from displaced vertices assuming pion masses
	$S_{\text{xyz}}(\text{JF})$	Significance of the average distance between the primary and displaced vertices
	$f_{\text{E}}(\text{JF})$	Fraction of the charged jet energy in the secondary vertices
	$N_{\text{1-trk vertices}}(\text{JF})$	Number of displaced vertices with one track
	$N_{\geq\text{2-trk vertices}}(\text{JF})$	Number of displaced vertices with more than one track
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices