COFUND. A project supported by the European Union

Introduction to Heavy Flavour Jet Tagging with ATLAS

ATLAS $H \to b\bar{b}$ and Flavour Tagging Workshop

5th September 2017

Andy Chisholm (CERN)
on behalf of the ATLAS collaboration

Overview

An introduction to *b*-tagging with ATLAS and associated experimental and theory issues...

Overview

- ***b*-hadrons:** *b*-tagging relies on the unique properties of the *b*-hadrons, why are they special?
- **Experimental Signature:** How do *b*-jets look to the ATLAS detector?
- **Algorithms:** How can we exploit these features to "tag" *b*-jets?
- **Simulation:** To what extent do we rely on simulation to understand *b*-tagging and how accurate is it?
- **Calibration:** How do we understand the performance of a *b*-tagging algorithm in data?

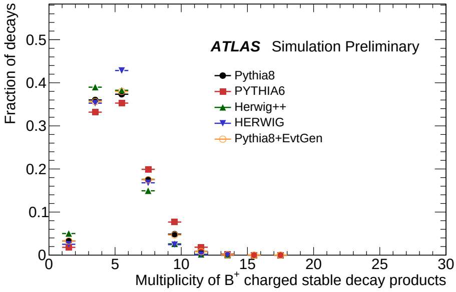
Jet Labelling Conventions

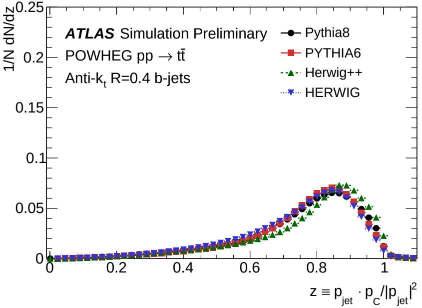
- ***b*-jet:** Jets containing a *b*-hadron
- ***c*-jet:** Jets containing a *c*-hadron but no *b*-hadron
- **Light flavour jet:** Jets containing no *b* or *c*-hadrons (originating from u, d, s quark and gluon fragmentation)

Properties of *b*-hadrons

The tagging of *b*-jets relies upon the unique properties of the *b*-hadrons:

- **Lifetime:** Long enough to lead to a measureable decay length (around 5mm for a 50 GeV boost)
- **Mass:** Largest masses of any hadrons, leading to high decay product multiplicities (average of 5 charged particles per decay)
- **Fragmentation:** Much harder than jets initiated by other parton species (*b*-hadrons carry around 75% of jet energy, on average)





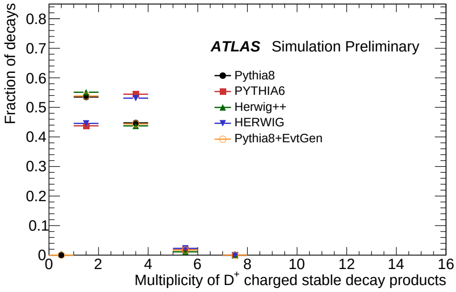
Left: Mean charged multiplicity in B^+ mesons decays

Right: *b*-quark fragmentation function

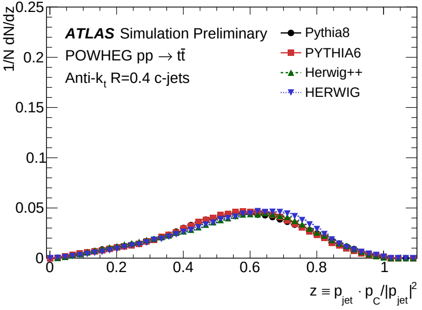
Properties of c-hadrons

Properties of the c-hadrons allow "c-tagging", but also very relevent for b-tagging given $B \rightarrow D$ decay chains and the rejection of c-initiated jets:

- **Lifetime:** Shorter than the b-hadrons by around a factor of 2-3, still enough for measureable decay length (around 1-3mm for a 50 GeV boost)
- **Mass:** Around 3× lower than b-hadrons (mean of ≈ 2 charged particles per decay)
- **Fragmentation:** Softer than b-jets, but still harder than jets initiated by light parton species (c-hadrons carry around 55% of jet energy, on average)



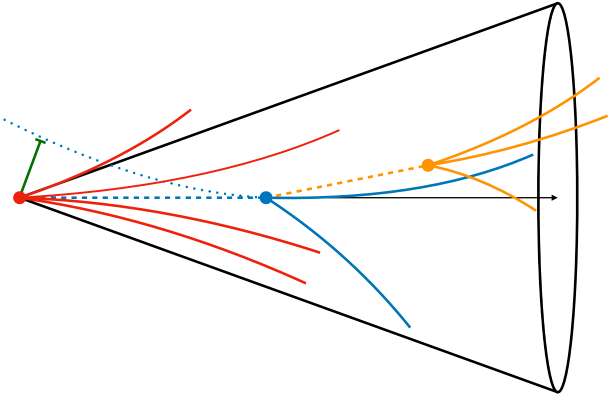
Left: Mean charged multiplicity in D^+ mesons decays



Right: c-quark fragmentation function

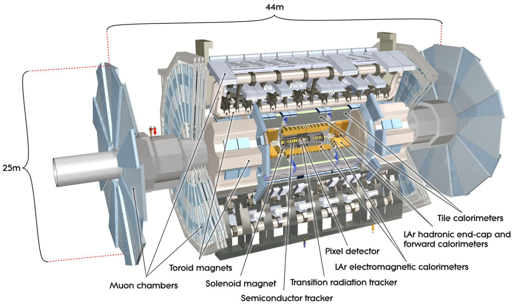
Anatomy of a "typical" *b*-jet

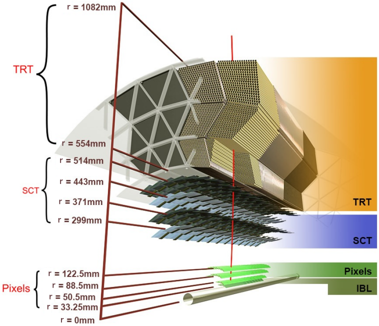
A typical *b*-hadron decay chain ($\sim$ 90%) involves a decay to a *c*-hadron



- ***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**

General purpose detector, well suited to studying heavy flavour jets

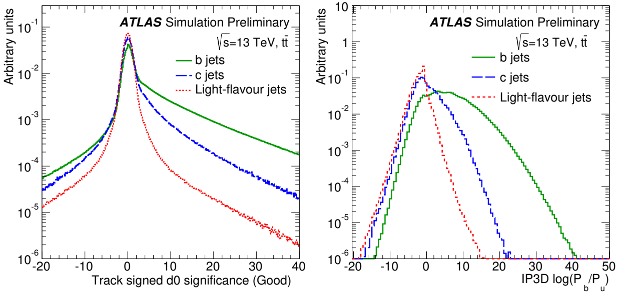




- **Inner Detector (ID):** Silicon Pixels and Strips (SCT) with Transition Radiation Tracker (TRT) $|\eta| < 2.5$ and (new for Run 2) Insertable B-Layer (IBL)
- **LAr EM Calorimeter:** Highly granular + longitudinally segmented (3-4 layers)
- **Had. Calorimeter:** Plastic scintillator tiles with iron absorber (LAr in fwd. region)
- **Muon Spectrometer (MS):** Triggering $|\eta| < 2.4$ and Precision Tracking $|\eta| < 2.7$
- **Jet Energy Resolution:** Typically $\sigma_E/E \approx 50\%/\sqrt{E(\text{ GeV})} \oplus 3\%$
- **Track IP Resolution:** $\sigma_{d_0} \approx 60\,\mu\text{m}$ and $\sigma_{z_0} \approx 140\,\mu\text{m}$ for $p_T = 1$ GeV (with IBL)

The signed IPs of tracks associated to jets are powerful jet flavour distriminants:

- Exploit “sign” of impact parameter: positive if track point of closest approach to PV is downstream of plane defined by the PV and jet axis
- Tracks from b-hadrons tend to have highly significant (IP/σ_{IP}) positive IPs, while most tracks from the PV have a narrow, symmetric distribution
- ✓ Very inclusive and highly efficient
- ✗ Relies upon accurate measurement of jet axis, sensitive to “mis-tag” high IP tracks from V^0 or material interactions, IP/σ_{IP} difficult to model in detector simulation

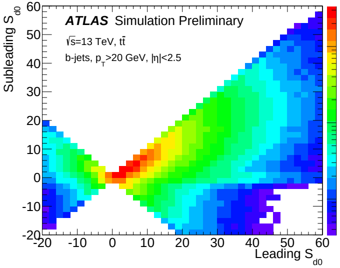


Left: Transverse IP significance distribution **Right: likelihood ratio discriminant based on 3D IPs of tracks**

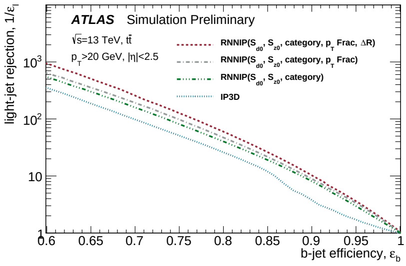
Exploiting *b*-hadron properties: 1 - Track Impact Parameters (IP)

For a *b*-jets, if you find one high IP track within a jet, you're more likely to find another... BUT this is not the case for light jets! How can we exploit this?

- Given the many tracks in a jet, exploiting these correlations is problem of very high dimensionality! Most IP taggers typically ignore such correlations...
- Machine learning techniques such as "Recurrent Neural Networks" (RNN) can learn sequential dependencies of arbitrary-length sequences, well suited to this problem
- New (for 2017) ATLAS RNN IP algorithms outperform baseline IP taggers, having (expected to have) learnt to exploit these correlations



Left: IP correlations for tracks from *b*-hadron decays

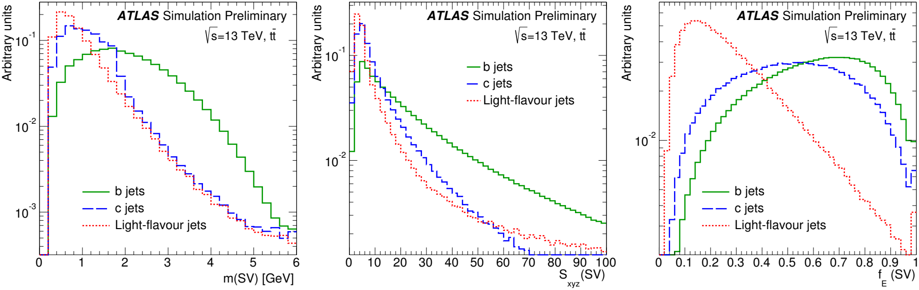


Right: Performance of standard (IP3D) and RNN IP taggers

Exploiting b-hadron properties: 2 - Secondary Vertices (SV)

Exploit expectation of a secondary vertex from either *b* or *c*-hadron decays:

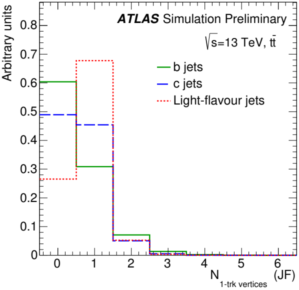
- Attempt to reconstruct a secondary vertex from high IP tracks associated with jet
- Use invariant mass of tracks at SV to discriminate b or c-hadron decay vertices from V^0 decays or material interations
- Further exploit hard b-jet fragmentation, SV should carry a large fraction of jet energy
- ✓ SV found in up to ≈ 80% of b-jets but only a few % of light flavour jets
- ✗ Degraded light jet rejection as jet p_T increases, careful considerations to mitigate "tagging" of material interactions required

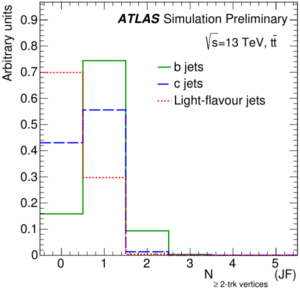


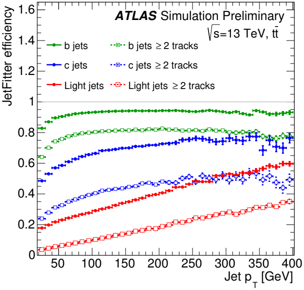
Left: Inv. mass of tracks at SV **Centre: 3D SV decay length significance** **Right: Energy fraction of SV tracks**

Exploit common occurance of cascade decay chain; *b*-hadron $\rightarrow$ *c*-hadron:

- Use Kalman filter to search for common axis on which three vertices lie: primary (*pp*) $\rightarrow$ secondary (*b*-hadron) $\rightarrow$ tertiary (*c*-hadron)
- Can then look for “1 track vertices” with decay chain axis
- ✓ Addition of 1 track vertices improves efficiency, constraint to decay chain axis improves separation power of SV based discriminants
- ✗ Degraded performance for *c*/*b*-hadron vertices as jet p_T increases, high fake rate for 1 track vertices (increases light jet “mis-tag” rate)





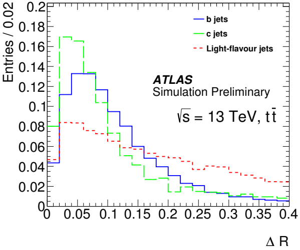


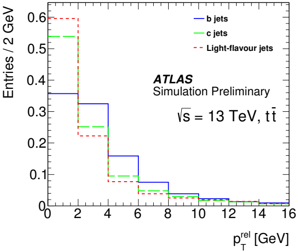
Left: Multiplicity of 1 track vertices **Centre: Multiplicity of 2+ track vertices** **Right: Chain reco. efficiency vs. jet p_T**

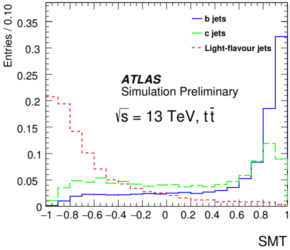
Exploiting b-hadron properties: 4 - Muons (Soft Muon Tagger)

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

- Expect much higher rate of muons within b/c-jets, relative to light flavour jets, due to the decays $B \to \mu\nu X$ and $B \to DX \to \mu\nu X'$ ($\mathcal{B}$ of around 10% each)
- ✓ 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



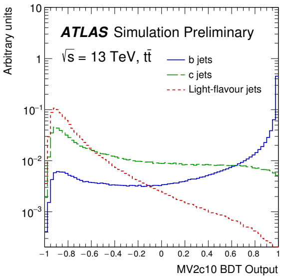




Left: ΔR of muon w.r.t. jet axis **Centre: p_T of muon relative to the jet axis** **Right: BDT built from muon observables**

Combine approaches together to exploit the full topology of a *b*-jet and mitigate the shortcomings of the individual methods:

- Use the output of the three basic approaches as input to a boosted decision tree (BDT) to build a single multivariate discriminant
- ✓ Benefit from the advantages of all basic techniques/algorithms to build a single very powerful discriminant
- ✗ Complex sensitivity to convolution of all detector and physics modelling issues (see later), relies strongly on "calibration" in data (as do "basic" algorithms)



Combined BDT discriminant ↑

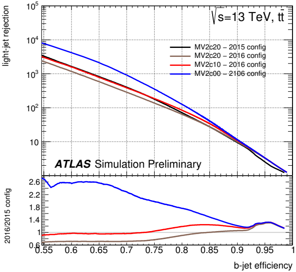
BDT Cut Value	b-jet Efficiency [%]	c-jet Rejection	Light-jet Rejection	τ Rejection
0.9349	60	34	1538	184
0.8244	70	12	381	55
0.6459	77	6	134	22
0.1758	85	3.1	33	8.2

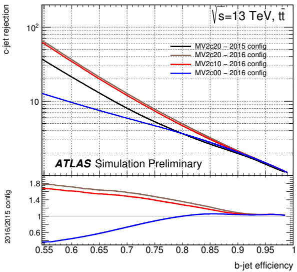
- Simplest way to define a "*b*-tagged" jet is through definition of "fixed cut" working points, with characteristic *b*-jet eff. vs. *c* and light flavour jet rej. shown above
- Choice of working point chosen driven by needs of individual physics analyses...

Exploiting *b*-hadron properties: 5 - Bring Everything Together

The training of a multivariate discriminant offers great flexibility for final application to physics analyses:

- For example, the light vs. *c*-jet rejection can be easily tuned to suit analysis needs through varying flavour composition of the training sample





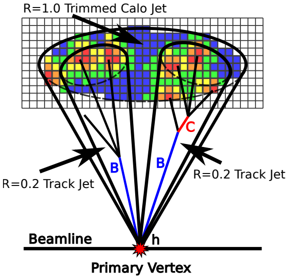
What about more sophisticated machine learning techniques?

- Modern “Deep Learning” neural network techniques offer more flexibilty and can potentially better exploit input correlations compared to a BDT
- ATLAS *b*-taggers based on deep learning available for LHC 2017 run (see **ATL-PHYS-PUB-2017-013** and **ATL-PHYS-PUB-2017-003**)

b-tagging for boosted $H \to b\bar{b}$: Introduction

What about b-tagging large-R (~ 1.0) jets with expected substructure such as a boosted $H, Z, X \to b\bar{b}$ decay?

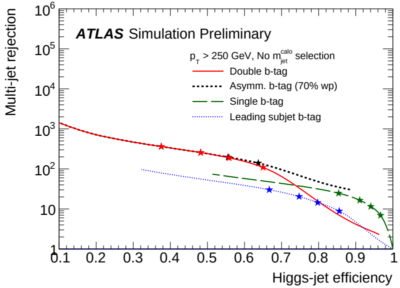
- In a boosted $H \to b\bar{b}$ system, as the angular separation of the b-jets tends to the small-R ($R = 0.4$) jet radius, the reconstruction and b-tagging of two "isolated" small-R jets becomes inefficient
- The $H \to b\bar{b}$ system can be reconstructed as a large-R ($R = 1.0$) jet with smaller radius (e.g. $R = 0.2$) "subjets" (track jets) identified within
- Standard b-tagging can then be performed on tracks associated with these subjets, ask for a single or double b-tag from the system to purify $X \to b\bar{b}$ large-R jet sample

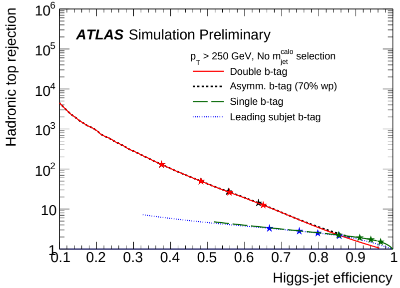


Double b-tagging performance depends on technique used to identify the subjets...

b-tagging for boosted $H \rightarrow b\bar{b}$: Baseline ATLAS Method

- Reconstruct anti-k_t $R = 1.0$ calorimeter jet and apply "jet trimming"[†] to discard soft component with a subjet radius of $R_{\text{subjet}} = 0.2$ and a minimum p_T fraction of 5%
- Associate anti-k_t $R = 0.2$ track jets with $R = 1.0$ calorimeter jets using "ghost association"[‡] and perform "standard" b-tagging on track jets (see below)
- $H \rightarrow b\bar{b}$ "tagger" then built by requiring additional selection based on the invariant mass of the "trimmed" $R = 1.0$ jet





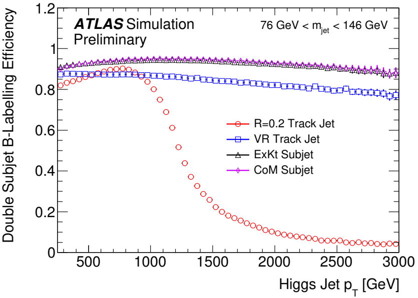
Left: $H \rightarrow b\bar{b}$ tag efficiency vs. multi-jet rej. **Right: $H \rightarrow b\bar{b}$ tag efficiency vs. hadronic top decay rej.**

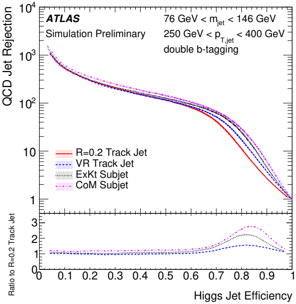
† D. Krohn, J. Thaler and L. T. Wang, "Jet Trimming", JHEP **1002** (2010) 084 (arXiv:0912.1342)

‡ M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," Phys. Lett. B **659** (2008) 119 (arXiv:0707.1378)

Performance of baseline method degrades for very high p_T^H, alternative methods to identify subjets have also been investigated:

- **Variable Radius Track Jets:** jet radius varies as a function of jet p_T as $R \propto p_T^{-1}$
- **Exclusive-k_t:** Varient of k_t algorithm, will cluster subjets until all protojet separations are above a given threshold or fixed number of jets is obtained
- **Calo. subjets in C.O.M. frame:** Boost jet constituents into to the large-R jet rest frame, then cluster subjets with EECambridge algorithm†





Left: Double b-tag efficiency vs. p_T^H **Right: $H \to b\bar{b}$ tag efficiency vs. multi-jet rej.**

† Y. L. Dokshitzer et al., "Better jet clustering algorithms", JHEP **9708** (1997) 001 (hep-ph/9707323)

Modelling of heavy flavour jets: Introduction

Performance of *b*-tagging algorithms in MC simulation depends on modelling of heavy flavour hadron production jets, particularly the following common aspects/issues:

- **Hadron Masses and Lifetimes:** Values in MC generators do not always represent current "best knowledge" or PDG/HFLAV combinations
- **Hadron Decays:** Choice of decay form factors, branching fraction values for exclusive decays (as above), re-production of inclusive branching fractions (e.g. $B \to \mu X$)
- **Fragmentation:** Relative fractions of hadron species and modelling of fragmentation function
- **Gluon Splitting:** Modelling at low $\Delta R_{q\bar{q}}$ and fraction of $g \to b\bar{b}$ relative to $g \to c\bar{c}$

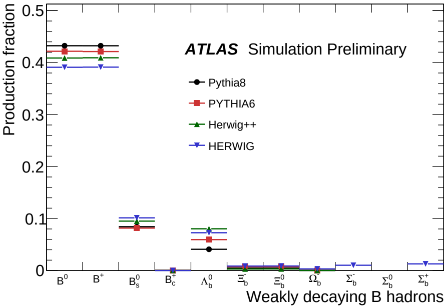
Detailed survey of heavy flavour hadron production modelling in common MC generators performed by ATLAS in ATL-PHYS-PUB-2014-008

- Compare a number of b/c-hadron observables in $t\bar{t}$ events generated with four common MC event generators:

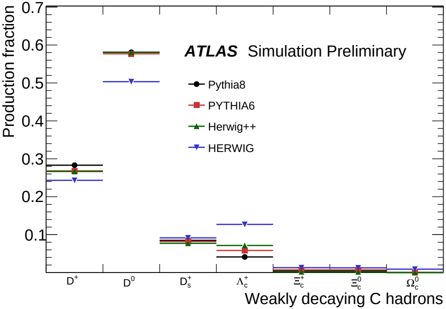
Pythia8(.175) (w./w.o. EvtGen), **Pythia6(.427)**, **Herwig++(2.6.3a)**, **HERWIG(6.520)**

Modelling of heavy flavour jets: Fragmentation Fractions

Properties exploited by *b*-tagging (e.g. lifetime and ch. mult.) vary significantly among *c*/*b*-hadron species, crucial to accurately model the relative production rates:



b-hadron fragmentation fractions



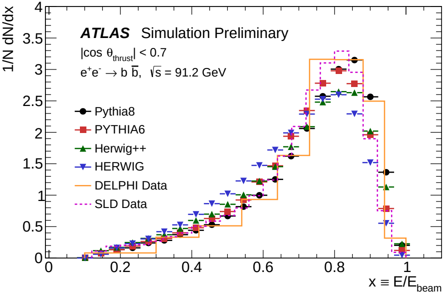
c-hadron fragmentation fractions

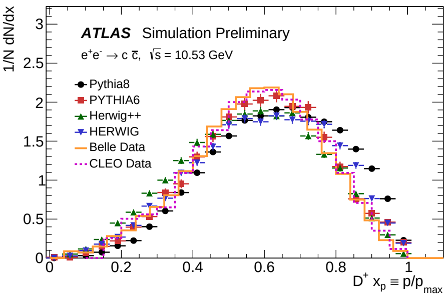
- Composition of *b*-baryon population rather varied and large spread in Λ_b fraction
- Λ_c^+ fraction most varied (most difficult to measure), generally good agreement in D^0 fraction (excl. HERWIG)

General agreement to within 10% for most copiously produced *b*/*c*-hadrons, though variations at this level can certainly noticeably affect *b*-tagging efficiencies

Modelling of heavy flavour jets: Fragmentation Functions

The "hardness" of heavy flavour quark fragmentation is exploited by *b*-tagging algorithms (e.g. SV energy fraction variables), accurate modelling crucial:





b-hadron fragmentation fractions

- b-quark fragmentation in $e^+e^- \rightarrow b\bar{b}$ at $\sqrt{s} = 91.2$ GeV, comparing LEP and SLC measurements with generator predictions

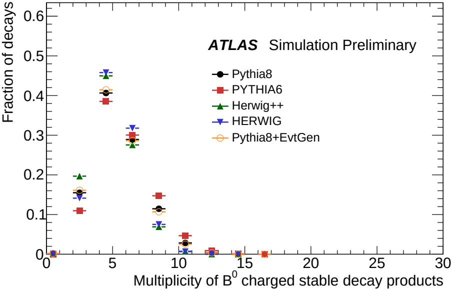
c-hadron fragmentation fractions

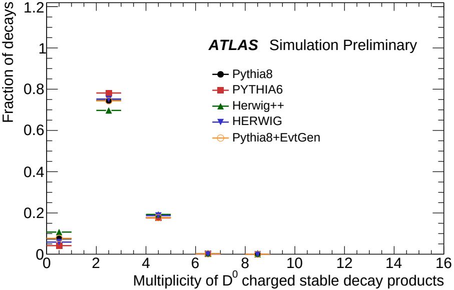
- c-quark fragmentation into D^+ in $e^+e^- \rightarrow c\bar{c}$ at $\sqrt{s} = 10.53$ GeV, comparing B-factory measurements with generator predictions

Several noticable discrepencies, how reliable are conclusions based on data/MC comparisons made with e^+e^- data/observables transfered to the LHC environment?

Modelling of heavy flavour jets: Hadron Decays

b-tagging algorithms are sensitive to "inclusive" modelling of decays, such as the number of charged particles produced in c/b-hadron decays:



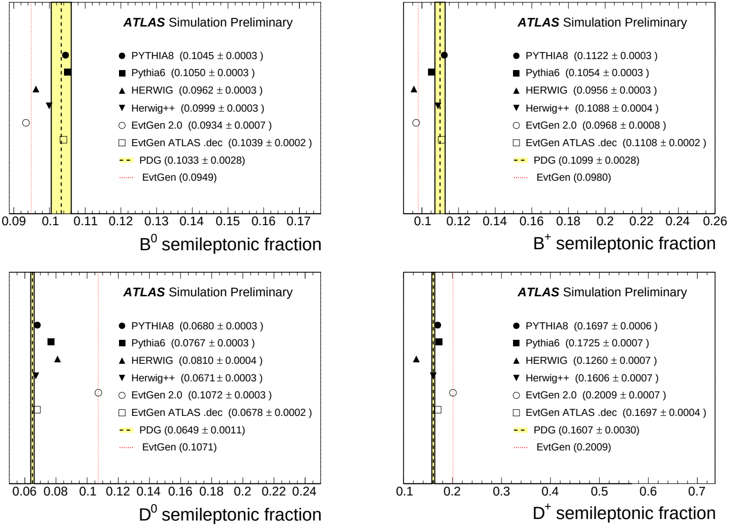


B^0 meson charged decay multiplicity **D^0 meson charged decay multiplicity**

- Can impact *b*-tagging efficiency by modifying the rate of SV finding and JetFitter decay chain identification

The multiplicities vary between generators by as much as $\sim 20\%$ for *b*-hadrons and $\sim 15\%$ for *c*-hadrons!

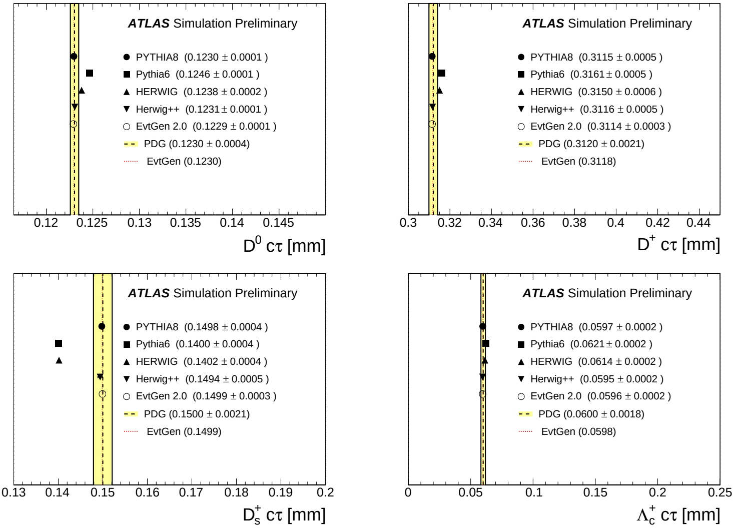
Muon-in-jet (SMT) based tagging sensitive to inclusive rate of $B/D \to e^{-}\bar{\nu}_e X$



EvtGen values particularly different to those from the PDG (2012)

Modelling of heavy flavour jets: *c*-hadron lifetimes

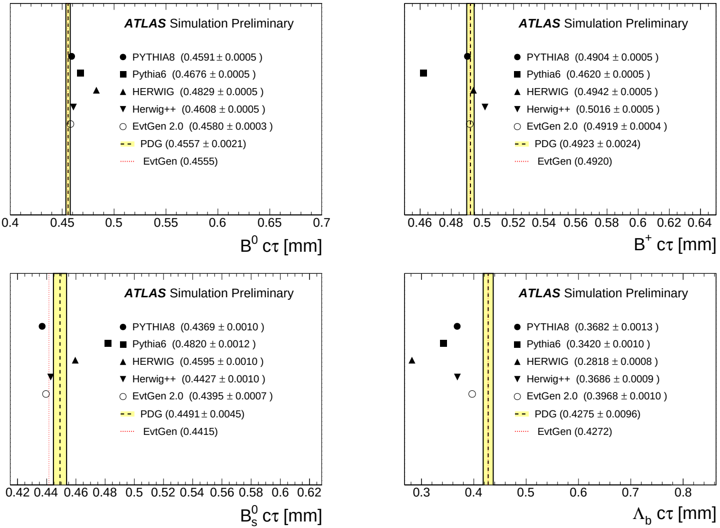
General agreement among generator and PDG values for weakly decaying *c*-hadrons



Only notable discrepency in D_s^+ value used by (Fortran) HERWIG and Pythia6

Modelling of heavy flavour jets: *b*-hadron lifetimes

Much more discrepency in generator and PDG values for weakly decaying *b*-hadrons

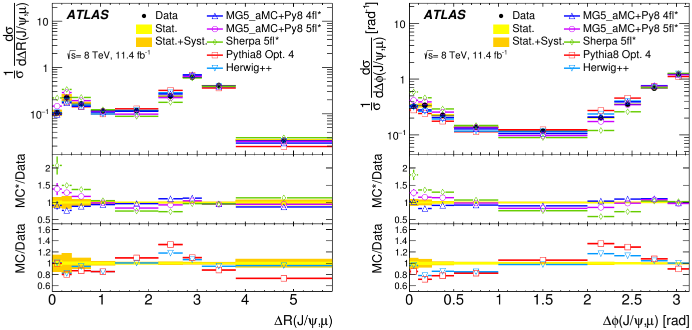


Λ_b lifetime particularly problematic in "plain" generators (without EvtGen)

Modelling of heavy flavour jets: $g \to b\bar{b}$ splitting

The region of small-angle $b\bar{b}$ pair production (dominated by $g \to b\bar{b}$) is particularly sensitive to the details of the various calculations...

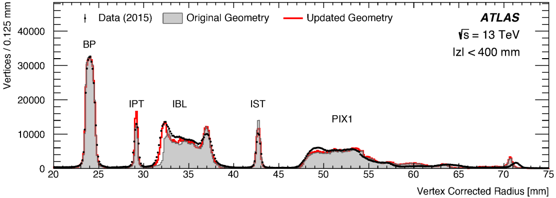
- Important for modelling of b-tagging since tagging efficiency for jets containing one or two $b(c)$-hadrons is significantly different!
- Highlighted in measurements† of proxy observables (ang. separation between $b \to J/\psi(\mu^+\mu^-)X$ and $\bar{b} \to \mu X'$) for the $b\bar{b}$ ang. separation in inclusive production



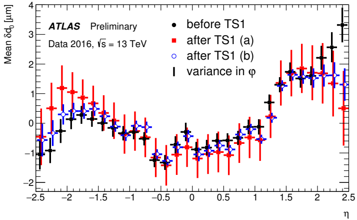
Much variation among MC generators with different approaches to model $g \to b\bar{b}$

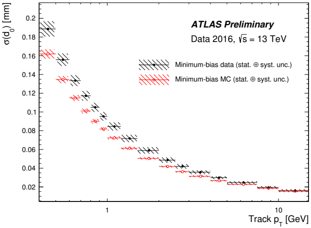
† ATLAS Collaboration, arXiv:1705.03374

In addition to b/c-hadron simulation, performance of b-tagging in MC also depends on the simulated performance of the ATLAS detector, which is also not perfect...



Example 1: Material - **Radius of hadronic interaction vertices reconstructed in the inner tracker detectors in data and MC simulation, sensitive to missing or mis-modelled material in the detector simulation**





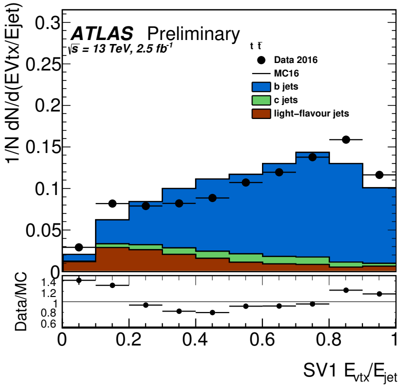
Example 2: Alignment - **Mean residual in transverse IP**

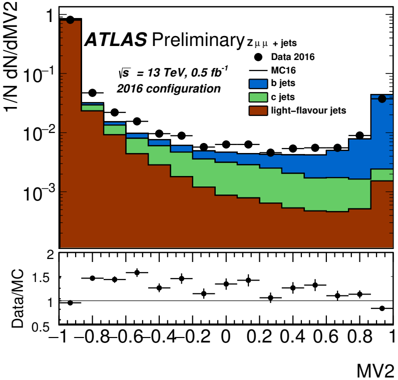
Consequence - **Difference in track transverse IP resolution in data**

How can these modelling issues affect *b*-tagging?

Two (worst I could find) examples of disciminant distributions data and simulation, where both MC generator and detector simulation effects can contribute:

Right: Final MVA discriminant, highly non-trivial sensitivty to convolution of all modelling issues





Left: SV energy fraction, sensitive to fragmentation modelling

If one were to make a "cut" to select a purified sample of *b*-jets, the efficiency and *c* and light jet "mis-tag" rates would differ in data and simulation! So how can we reliablly use *b*-tagging with MC events?

Calibration - Introduction

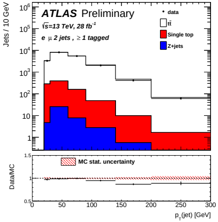
Solution is to perform a "calibration" of the *b*-tagging efficiency; measure the efficiency in data and correct MC efficiency accordingly

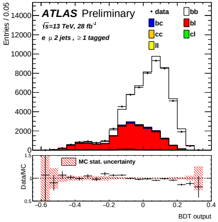
- Crucial to "calibrate" performance not only for *b*-jets, but also *c* and light flavour jets (which are often more affected by detector modelling issues)
- Calibration analyses typically focus on the selection of pure and well understood sample of jets of a given flavour in data with which to measure the efficiency
- *b*-jet calibrations often rely on $t\bar{t}$ events while pure samples of *c* and light flavour jets are typically more difficult to isolate

As a demonstration of the concept, how is this calibration performed for *b*-jets?

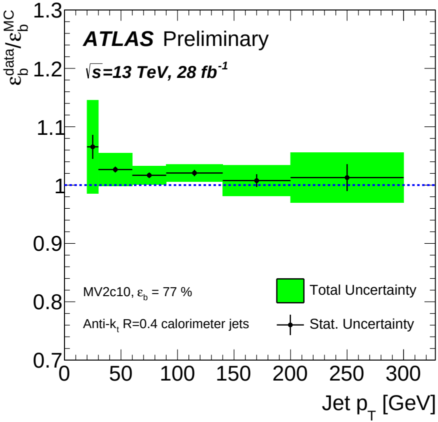
Exploit copious *b*-jet production in top quark decays:

- Select sample of di-leptonic $t\bar{t}$ events $t\bar{t} \to \ell^+\ell^- b\bar{b} + E_{\mathrm{T}}^{\mathrm{miss}}$, very low backgrounds in $e + \mu$ channel
- With multivariate analysis based on kinematic varibles, can enrich sample of jets with $b-$jets from top quark decays





Leads to precise understanding of efficiency in data, total relative uncertainty typically less than 10%



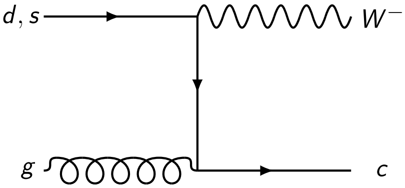
- Combination of all b-jet and detector modelling issues manifest as slightly lower b-tagging efficiency in simulation
- Correction applied to MC event weights

Calibration - Summary of c-jet and light flavour jet calibrations

Generally more difficult to isolate sufficienctly pure samples of c and light jets, given very high c and light jet rejection factors of b-tagging algorithms...

c-jets: The following processes are exploited to isolate high purity c-jet samples

- Semi-leptonic $t\bar{t}$ events (with $W \to cs, cd$ decays)
- $W(\ell\nu) + c(\mu\nu X)$ events, exploit charge correlation between leptons (example diagram $\to$)
- Multi-jet events with a reconstructed $D^{\star\pm} \to D^0\pi_s^\pm$ (with $D^0 \to K\pi$)

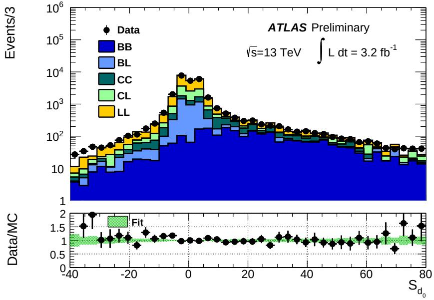


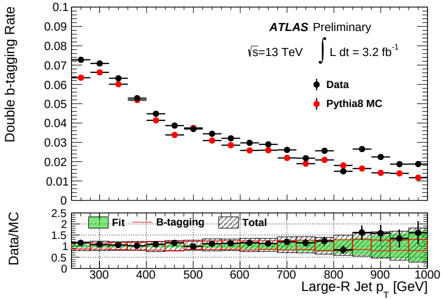
Light flavour jets: Typically use multi-jet event samples with the following techniques

- For light jets, differences in performance between data and simulation tend to be dominated by detector modelling effects (e.g. track IP resolution)
- "Negative tag" - reversing the IP sign of tracks or decay length significance sign of SVs can provide a good approximation of the mistag rate due to resolution effects
- "Adjusted MC" - Modify track parameters in simulated events to reflect resolution in data, repeat b-tagging with "adjusted" tracks to estimate change in mistag rate

Difficult to isolate sample of $X \to b\bar{b}$ events for calibration, so turn to $g \to b\bar{b}$ splitting in multi-jet events as a control sample:

- From a multi-jet event sample, identify large $R = 1.0$ jets with two associated small $R = 0.2$ track jets
- Enrich $g \to b\bar{b}$ content by requiring a reconstructed muon to be associated to at least one of the track jets (preferentilly selecting $b \to B \to \mu\nu X$)
- Perform fit to transverse IP significance (S_{d_0}) distribution of the tracks with highest $|S_{d_0}|$ to determine $g \to b\bar{b}$ component

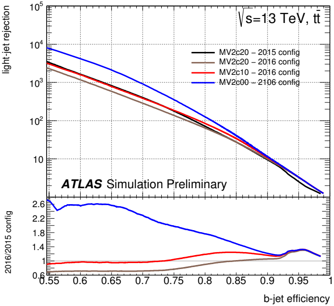


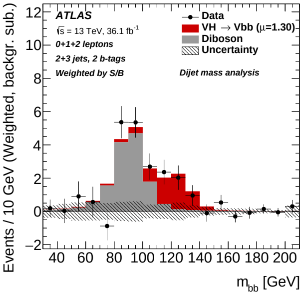


Left: Leading track IP significance distribution **Right: Double *b*-tagging efficiency in data and simulation**

Summary

- *b*-tagging exploits the unique properties of the *b*-hadrons to identify *b*-jets containing distinctive high impact parameter tracks and secondary vertex signatures
- As a consequence, simulated *b*-tagging performance is sensitive to many issues surrounding the modelling of c/b hadron production/decay in MC generators
- Well established calibration techniques developed to account for such issues, allowing MC simulation to remain a powerful tool to aid measurements





End result is highly performant and well understood *b*-tagging algorithms which enable to explore heavy flavour final states towards advancing our understanding of the SM and perhaps elucidate what may lie beyond it!

Additional Slides

References

"Expected performance of the ATLAS b-tagging algorithms in Run-2"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-022/

"Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2016-012/

"Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-013/

"Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-003/

"Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-008/

"Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2016-039/

"Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs ($\rightarrow b\bar{b}$) Tagging in ATLAS"

http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-010/

Measurements of c-hadron fragmentation fractions at LEP

H_c	ALEPH $f(c \to H_c)$ [%]	DELPHI $f(c \to H_c)$ [%]	OPAL $f(c \to H_c)$ [%]
D^0	$55.3 \pm 1.6 \pm 3.4$	$53.4 \pm 1.5 \pm 2.2$	$58.2 \pm 4.0\ ^{+3.9}_{-3.6}$
D^+	$23.4 \pm 0.8 \pm 1.3$	$22.2 \pm 0.7 \pm 0.9$	$22.8 \pm 2.9\ ^{+1.6}_{-2.0}$
D_s^+	$9.1 \pm 1.5 \pm 0.5$	$9.7 \pm 0.9 \pm 0.5$	$7.1 \pm 1.9 \pm 0.9$
Λ_c^+	$5.8 \pm 0.6 \pm 0.3$	$6.4 \pm 1.3 \pm 0.7$	$3.5 \pm 1.6 \pm 0.6$
$D^{\star +}$	$23.3 \pm 1.0 \pm 0.9$	$24.1 \pm 0.6 \pm 0.9$	$23.0 \pm 0.4 \pm 0.9$

Table from: Eur. Phys. J. C75 (2015) 19 (arXiv:1404.3888)

Table 1: Fragmentation fractions of b quarks into weakly-decaying b-hadron species in $Z \to b\bar{b}$ decay, in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV.

b hadron	Fraction at Z [%]	Fraction at $\bar{p}p$[%]
B^+, B^0	40.4 ± 0.9	33.9 ± 3.9
B_s	10.3 ± 0.9	11.1 ± 1.4
b baryons	8.9 ± 1.5	21.2 ± 6.9

Table from: K.A. Olive et al. (Particle Data Group), Chin. Phys. C, 38, 090001 (2014)

“Ghost Association”

M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas,” Phys. Lett. B 659 (2008) 119 (arXiv:0707.1378)

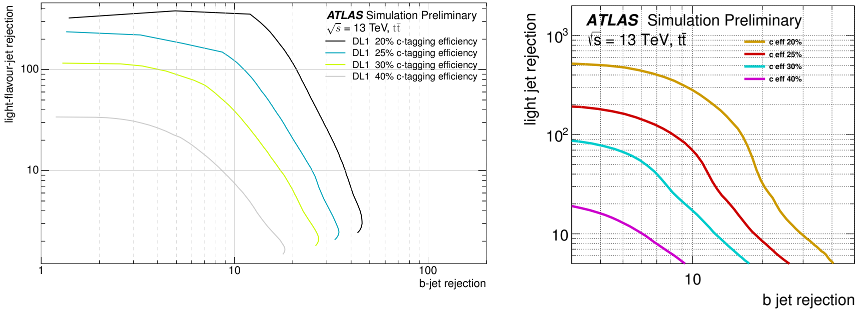
“The jet area is a non-trivial (and novel) concept insofar as a jet consists of pointlike particles, which themselves have no intrinsic area. To define a sensible area one therefore adds additional, infinitely soft particles (ghosts) and identifies the region in y, ϕ where those ghosts are clustered with a given jet. The extent of this region gives a measure of the (dimensionless) jet area.”

"Ghost Association" of $R = 0.2$ track jets to $R = 1.0$ calo. jets

ATLAS Collaboration, "Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV," ATLAS-CONF-2016-039

"In this procedure, the ghosts are the track jet 4-vectors in the event, with the track jet p_T set to an infinitesimal amount, essentially only retaining the direction of the track jets. This ensures that jet reconstruction is not altered by the ghosts when the calorimeter clusters plus ghosts are reclustered. The reclustering is then performed using the anti-k_T algorithm with $R = 1.0$. The calorimeter jets after reclustering are identical to the ungroomed parents of the trimmed jets used in this analysis, with the addition of the associated track jets retained as constituents. In this analysis, track jets ghost-associated to the large-R jet refer to track jets found this way within the catchment area of the ungroomed parent jet, but the kinematics of the large-R jet are still measured using the trimmed jet."

c-jet tagging



Two examples of the performance of MVA based *c*-jet taggers, using "Deep Learning" (left) and BDT (right) MVA training techniques