Based on: arXiv:2202.05082[hep-ph]

Identification of b-jets using QCD-inspired observables

Oleh Fedkevych[(1)], Charanjit K. Khosa[(2)], Simone Marzani[(1)], **Federico Sforza[(1)]**

[(1)]University & INFN Genova (IT), [(2)]University of Bristol (UK)

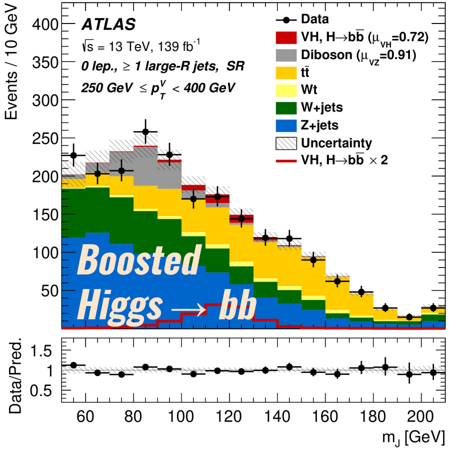
ICHEP 2022 - Bologna, Italy

The importance of beauty (quark) in HEP

Beauty (b) quarks - of m_b ~ 4.2 GeV - **are key elements when exploring fundamental particle physics at collider experiments and in pheno calculation.** Just a few examples:

Experimental window on SM & BSM:

high-p_T *b-jets* (p_T range ~20 GeV to ~TeV) from *b* quark hadronization are experimental signature of:



⇒ *Higgs* →bb (BR~58%)

⇒ *Top-quark decay*

⇒ *BSM models*

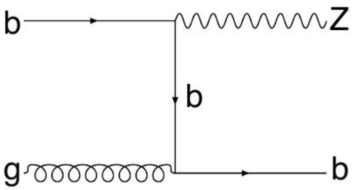
⇒ *QCD jet production*

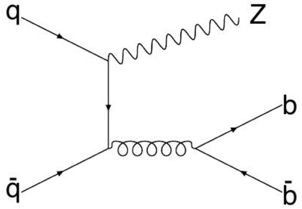
Probe of complex QCD mechanics:

- $m_b > \Lambda_{QCD} \Rightarrow$ how to include it in QCD predictions?
- Benchmark of state-of-the-art MCs and fixed-order pheno predictions (pQCD) in a problem with multiple scales, many pheno model choices, sizable unknowns

Example: QCD Flavour Schemes (FS) for Z+b@LO:

5FS: massless *b* in proton PDF **4FS:** g→*bb* splitting and $m_b \neq 0$





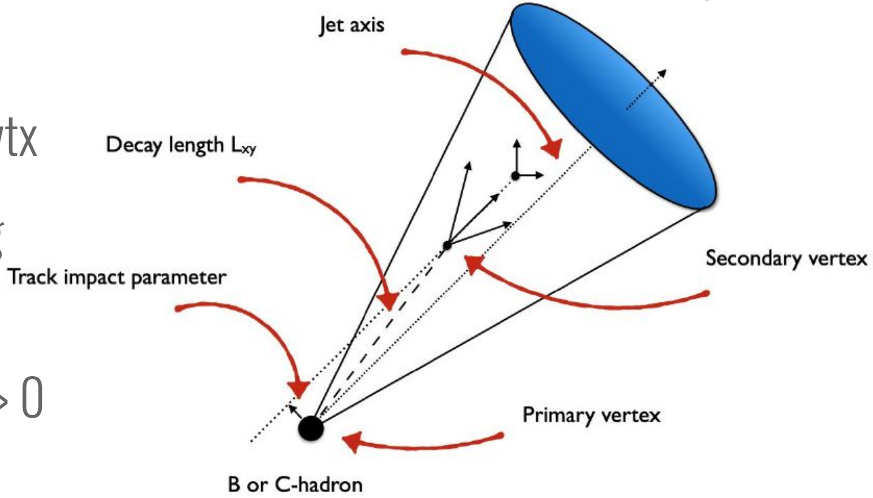
High-p_T *b-jets* at hadron colliders: *identification technique*

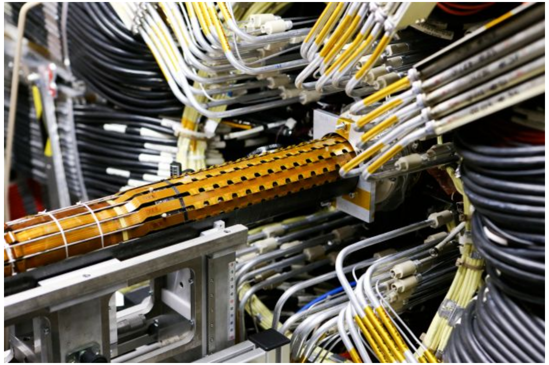
b-jet identification (*b-tagging*) relies on "detectable" characteristics of B-hadrons inside a jet:

⇒ Long lifetime, *$O(10^{-12}$ s)*, and complex decay chains give secondary/tertiary decays displaced from primary vtx

⇒ Reconstruction of B-hadron inv. mass (>5 GeV) using 4-momentum combination of multiple charged hadron

⇒ Charged hadron trajectories with *impact parameter* > 0





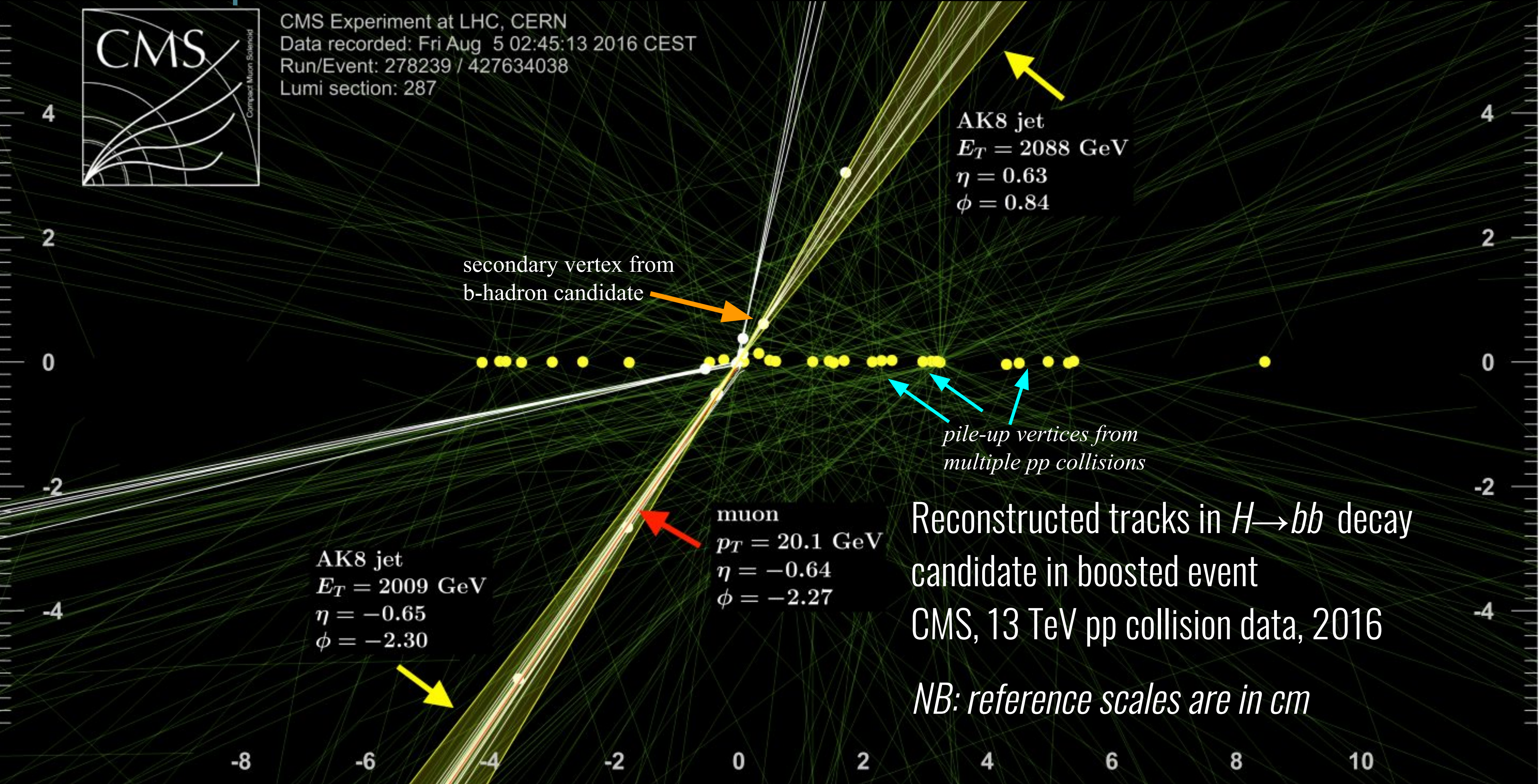
JINST 13 T05008 (2018)

⇒ All information recorded by cutting-edge tracking detectors deployed as close as possible to the beamline (e.g. ATLAS IBL at ~3 cm)

⇒ Machine-learning (ML) algorithms used to condense tracking & jet info for optimal separation ***b-jets*** **vs** ***light-jets*** *(i.e. w/o b- or c-hadrons)*

High-p_T *b-jet* at hadron colliders: *how does it look?*

[CMS-DP-2017-032 (2017)]



Reconstructed tracks in $H \to bb$ decay candidate in boosted event
CMS, 13 TeV pp collision data, 2016

NB: reference scales are in cm

Can we do better using QCD inspired observables?

QCD nature of b-jets - i.e. different radiation pattern, hadronization details, color flow, etc. - only marginally considered for b-jets identification ⇒ ***Can we exploit all this?***

- Proposing *(for first time)* use of QCD inspired observables for *b-jet* flavour identification
- Tested two jet substructure observables together with ML to maximise flavour info extraction:

  Jet Angularities [PRD **79** (2009), 074017] and ***primary Lund Plane*** [JHEP **2018**,64 (2018)]

  NB: proof of concept, other observables might also work - e.g. dead-cone? [Nature 605, 440 (2022)]

- Focus on *b-jets* for p_T > 500 GeV & boosted regime ⟶ *where tracking performance degrades*
- Evaluate performance in comparison with state-of-the-art experimental *b-tagging* algorithms
- Ultimately able to map *b-jet* substructure patterns to pQCD calculations or ME+PS simulation!

The experimental setup

Final aim is to compare performance with *b-tagging* algorithm data from LHC experiments:

- Start from Z+jets phase space used for CMS Jet Angularity measurements [JHEP **01** (2022),188]
  ⇒ $Z \to \mu\mu$ decay with $p_{T,\mu} > 26$ GeV, $|\eta_\mu| < 2.4$, 70 GeV < $M_{\mu\mu}$ < 110 GeV, $p_{T,\mu\mu} > 30$ GeV
  ⇒ Anti-k_T, R=0.4 jet, **$p_{T,jet} > 500$ GeV**, $|y_{jet}| < 1.7$, jet-Z balancing:

$$\left| \frac{p_{\mathrm{T,jet}} - p_{\mathrm{T},\mu^+\mu^-}}{p_{\mathrm{T,jet}} + p_{\mathrm{T},\mu^+\mu^-}} \right| < 0.3$$

$$|\phi_Z - \phi_{\mathrm{jet}}| > 2$$

- Simulate events pp 13 TeV collisions using Pythia v8.303 LO+PS, plus hadronization and UE (Herwig v7.2.1 used as cross check)

- *b-jet* or *light-jet* flavour assignation using experiment operative definition [EPJC **81** (2021) 3]
  ⇒ ***b-jet*** label if clustered jet has ≥ 1 *b-hadron* with $p_T \geq 5$ GeV within $\Delta R = 0.3$
  ⇒ ***light-jet*** label if no *b-hadron* nor *c-hadrons* with $p_T \geq 5$ GeV found in $\Delta R = 0.3$
  Analyze only leading jet in each event, benchmark stat. of 10^5 *b-jets* and 10^5 *light-jet*

NB: Rivet routine available upon request

Analysis of flavour dependence for Jet Angularities

Jet Angularities already proven effective in quark/gluon discrimination, also calculable in pQCD

⇒ Defined as:

$$\lambda_\alpha^\kappa = \sum_{i\in\text{jet}} \left(\frac{p_{\mathrm{T,i}}}{\sum_{j\in\text{jet}} p_{\mathrm{T,j}}} \right)^\kappa \left(\frac{\Delta_i}{R} \right)^\alpha$$

i over all jet constituents; Δ_i distance of i^{th} particle from jet axis

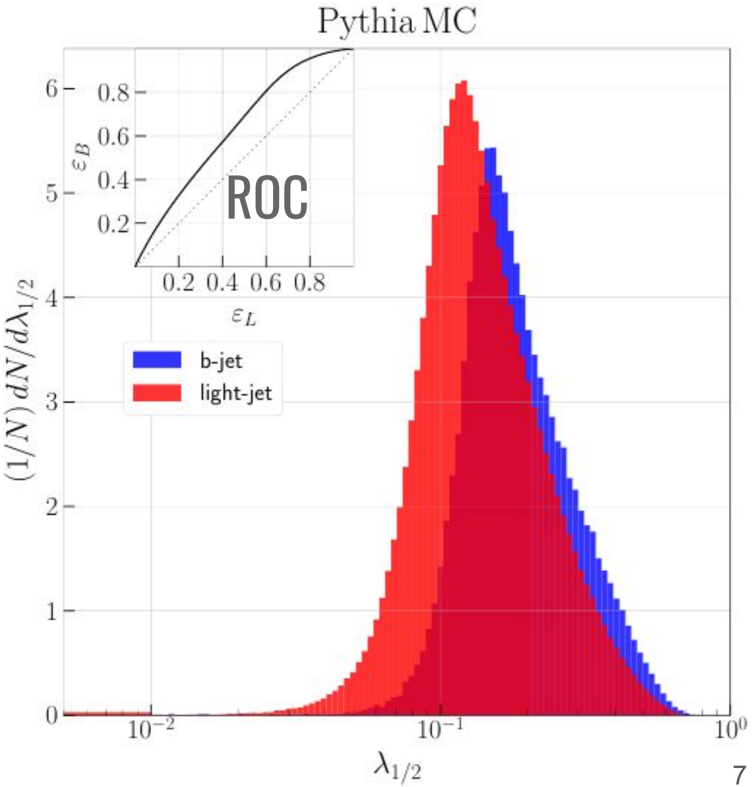
Infrared & Colinear (IRC) safety requires $\kappa = 1$, $\alpha > 0$

⇒ Considering: $\lambda^1_{1/2}$(LHA), λ^1_1(Width), λ^1_2(Thrust)

Achieved mild discrimination of b vs light jets!

Quantify it using ROC curve, i.e. efficiency beyond a cut:

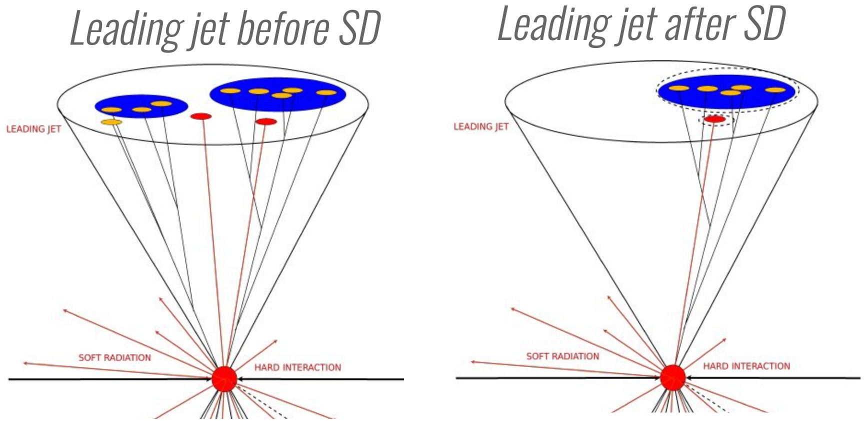
$$\varepsilon_{B/L} = \frac{1}{N_{B/L}} \int_{\lambda_{\text{cut}}}^{1} \frac{dN_{B/L}}{d\lambda} d\lambda$$

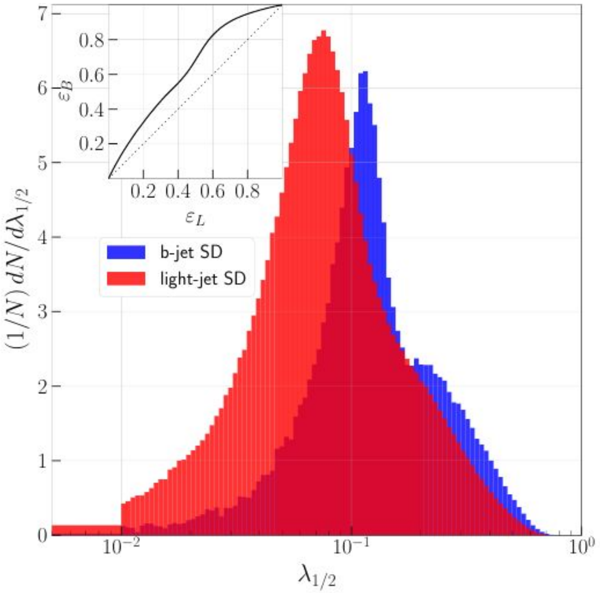


Jet grooming: b vs light jet identification against soft radiation

Problem: experiments not able to detect soft hadrons and do not have access to full set of hadron level jet constituents ⇒ *is b vs light discrimination stable removing soft radiation?*

⇒ Repeat all studies after application of SoftDrop (SD) jet grooming algorithm [[JHEP 2014, 146 (2014)]]





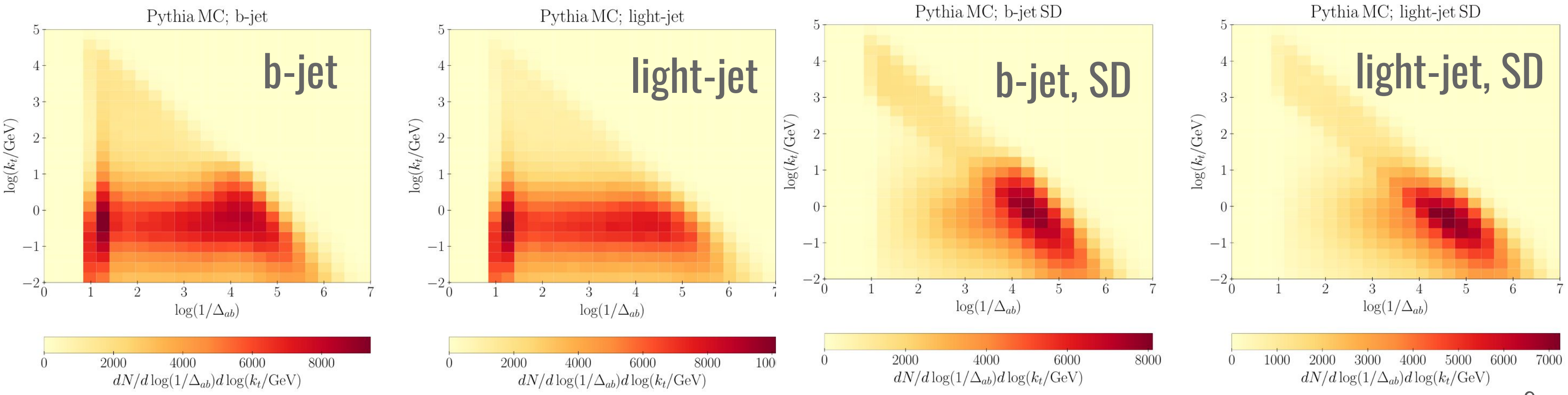
⇒ Shape of all jet angularities affected but ROC performance is stable, quantified using area under the curve (AUC):

E.g. $\lambda^1_{1/2}$ *baseline AUC = 0.639, vs* $\lambda^1_{1/2}$ *AUC with SD = 0.635*

Primary Lund-Plane: b vs light jet identification with QCD images

Single angularity distributions have some *b vs light* jet but b-tagging is often approached as a truly multi-variate problem to obtain good separation ⇒ *can QCD observables serve for this?*

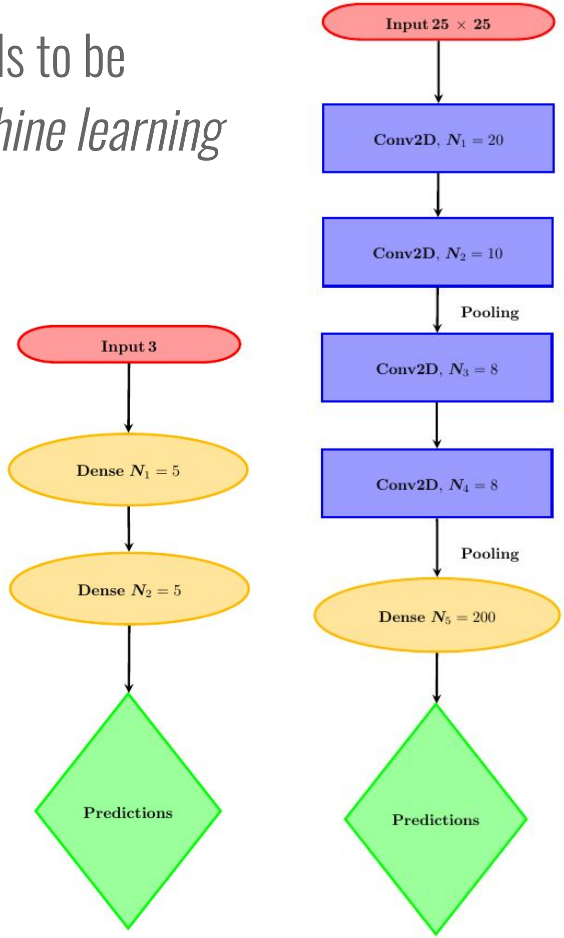
The primary Lund-Plane (pLP) builds an "image" of the QCD branching of the jet constituents in the space $\log(k_T)$ vs $\log(1/\Delta_{a,b})$ ⇒ *moving from 1D jet Angularity to images with 25x25 pixels*



Combination of multivariate information with Machine-Learning

Optimal multi-dimentional and multi-variate discrimination needs to be tackeled with advanced statistical methods ⇒ *we will use machine learning*

- Deep Neeral Network (DNN) with 3 input neurons and 5 neurons x2 hidden layers for optimal combination of the three Jet Angularities (LHE, Width, Thrust)
- Convolutional Neural Network (CNN) with 4 convolution steps and one hidden layer with 200 neurons for optimal image analysis of pLP
- Training with 60% of the simulated events, 20% for validation, 20% for performance evaluation



Benchmarking against ATLAS experiment performance

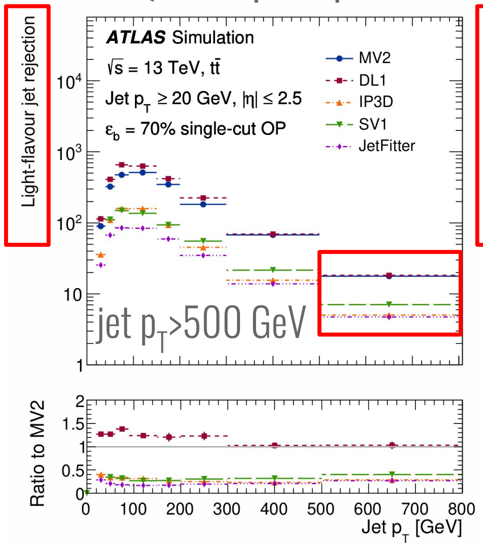
But do QCD-inspired observables provide interesting input for b-tagging at experiments?

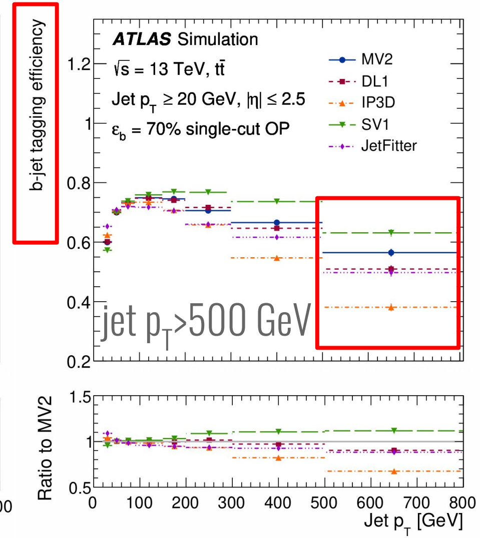
Check high-p_T *b-tagging* performance on same footing of LHC collaborations ⇒ *two stage approach used*

⇒ First set of so-called ***input-algorithms*** to obtain "low-level" quantities from reconstructed tracks: fit of secondary or tertiary decay vertices, decay displacement, invariant mass from track 4-momentum combination, distance of track minimum approach to beam line (i.e. impact parameter), etc.

⇒ *Feed to ML algorithm for optimal discrimination*

Comparing to ATLAS collaboration which provides detailed p_T and flavour dependent efficiency for fixed cut benchmarks of input taggers (SV1, IP3D JetFitter) and of final Deep Learning algorithm [EPJC **79** (2019) 970]



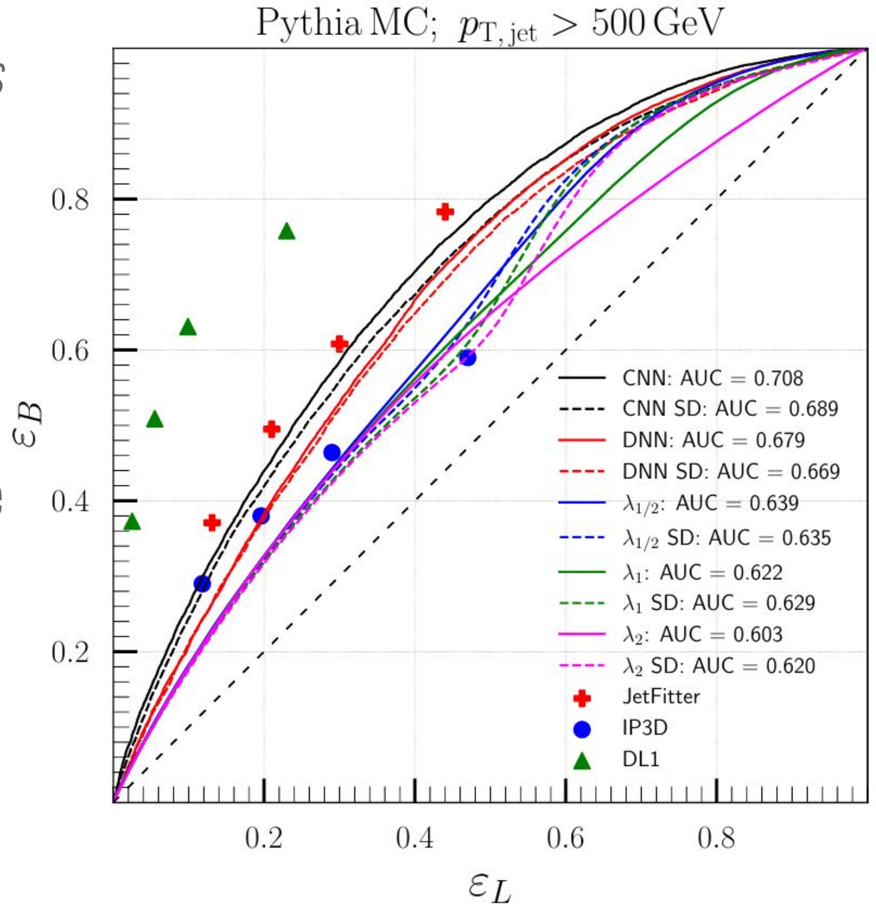


What we learnt QCD can do for b-jet flavour identification

QCD inspired observables as good as track-based input-algorithms!

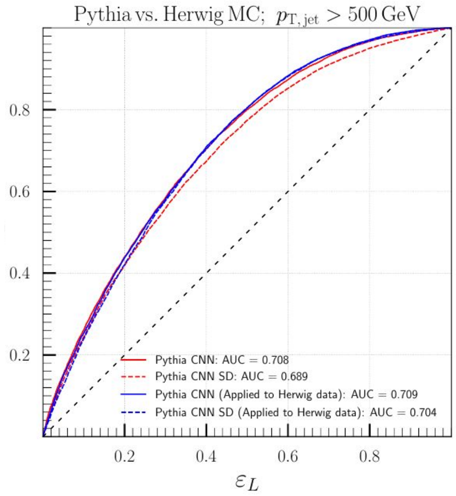
But no B-hadron info nor decay patterns are used!

⇒ QCD seems to provide mostly new & orthogonal *b-tagging* information which may be integrated for innovative *b-tagging*



Results stable against hadronization details:

⇒ Herwig for performance eval. but Py8 CNN training



Conclusions and prospects

⇒ In arXiv:2202.05082[hep-ph] we tested for the first time (at the best of our knowledge) the use of QCD inspired observables for constructing *b-jet* identification algorithms

⇒ Each individual QCD observable, namely the three Jet Angularities and the primary Lund Plane images, show some *b* vs *light jet* discrimination power ⇒ significantly enahanced when information is combined using ML algorithms, as the tested DNN and CNN classificators

⇒ Performance for *b-jet* identification using QCD inspired observables, providing orthogonal information to track-based *b-tagging* algorithms, could be a promising field with theory-informed analysis helping where current detector technology has limitations: e.g. very high-p_T

⇒ Future work could lead to measurement of QCD inspired observables in flavour-specific data samples and, ultimately, to their use as additional inputs for experimental *b-jet* tagging algorithms

Extra slides & Backup Material

b-jet identification algorithm based on track reconstruction

transverse IP d_0: distance of closest approach of track to PV in the r-ϕ projection
longitudinal IP z_0: distance between z coordinates of the PV and the track at closest approach in r-ϕ

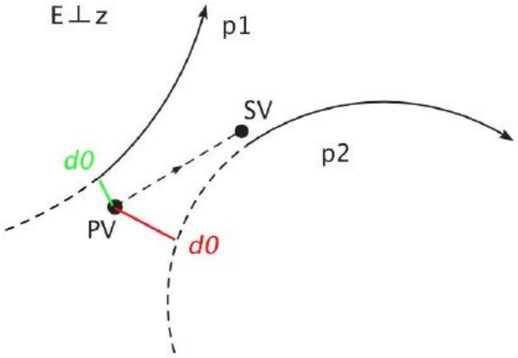
- introducing sign: "+" ("-") - track intersects the jet axis in front of (behind) the PV
- using significance, e.g. d_0/σ_{d_0}

IP3D: LLR-based, relies on d_0 and z_0 significances as well as correlations

Explicit reconstruction of the complete b-hadron decay chain.

- exploiting topological structure of weak b- and c-hadron decays inside jet
- uses Kalman filter to find common line between and position of PV, SV and TV (tertiary, c-hadron decay vertex)
  $\rightarrow$ approximating b-hadron flight path
- one track sufficient to built vertex!

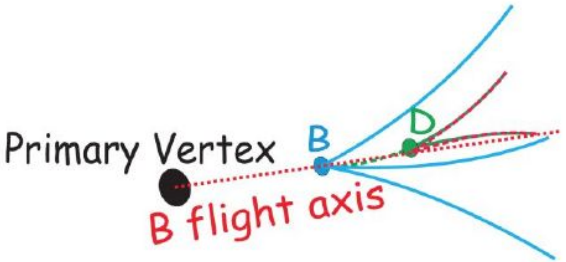
$\rightarrow$ three output nodes corresponding to the b-,c- and light



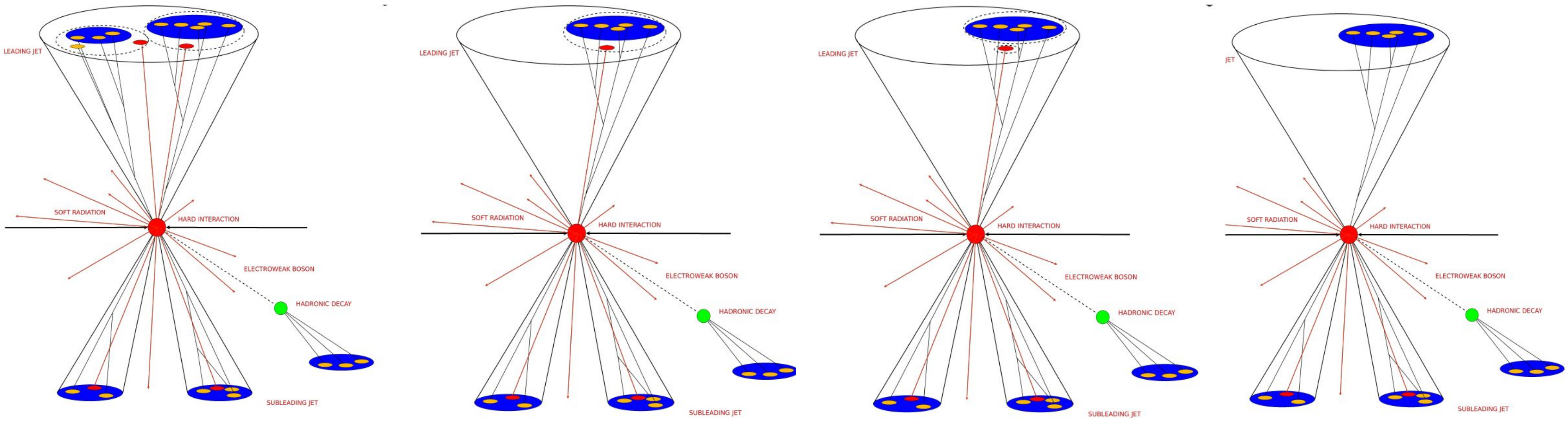
Explicit reconstruction of a *single, inclusive* secondary vertex (SV).

- using all associated, displaced tracks to form vertex candidates from track pairs (χ^2 based)
- vertices compatible with long-lived particles and material interactions are rejected
- iterative procedure to combine all tracks from 2-track-vertices into single inclusive vertex

$\rightarrow$ **SV1**: LLR-based, exploiting vertex mass, energy fraction, number of 2-track vertices, ΔR(jet,PV-SV)

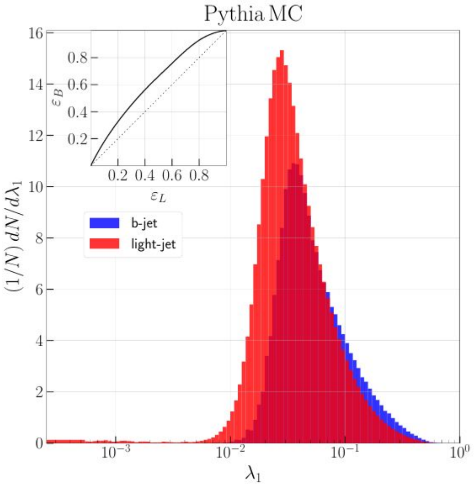


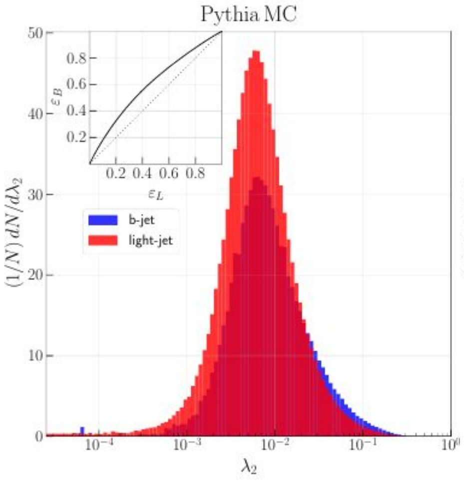
SoftDrop algorithm

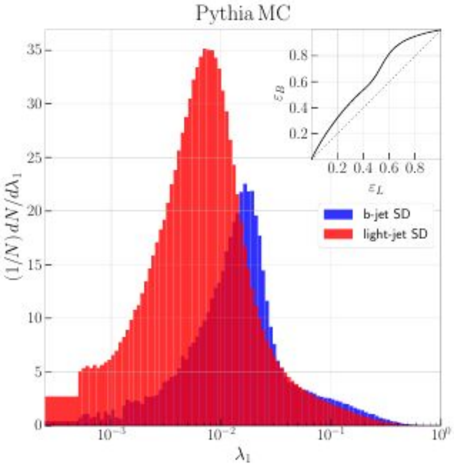


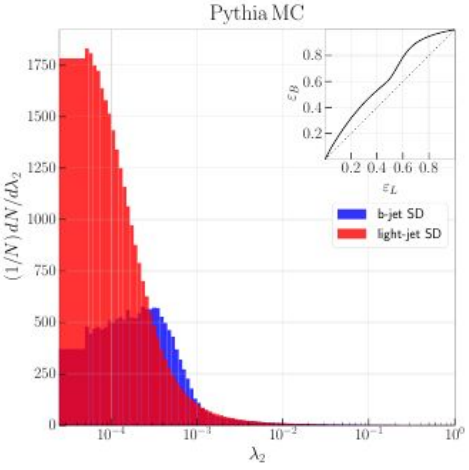
SoftDrop groomer: $\frac{\min(p_{ti},p_{tj})}{p_{ti}+p_{tj}} > z_{\mathrm{cut}}\left(\frac{\Delta R_{ij}}{R}\right)^{\beta}$

Jet Angularities for b-jets and light-jets: Width and Thrust

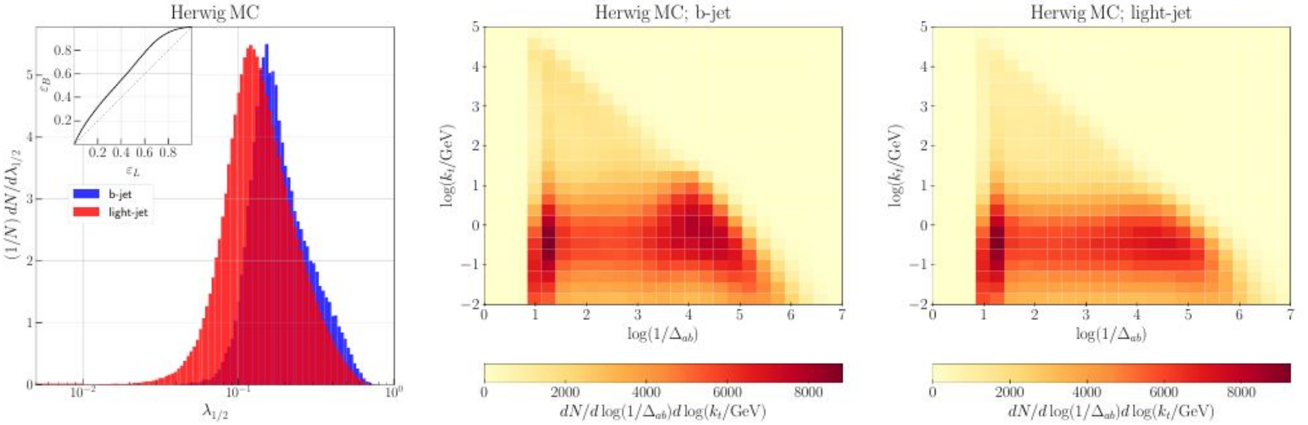








Jet Angularities (LHE) and primary Lund Plane using Herwig



(a) (b) (c)