b-jet tagging performance with ALICE at the LHC

Barbara Trzeciak (Utrecht University)
on behalf of the ALICE Collaboration

CIPANP 2018
Palm Springs, CA
May 29 – June 3 2018

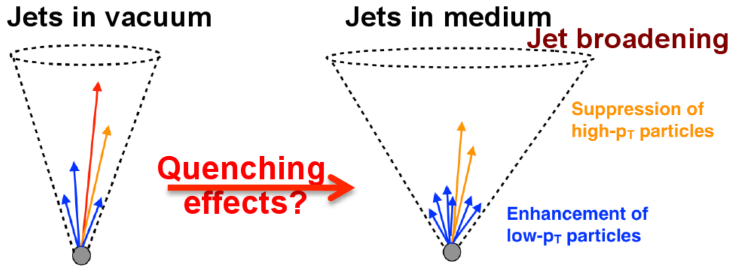
Netherlands Organisation for Scientific Research

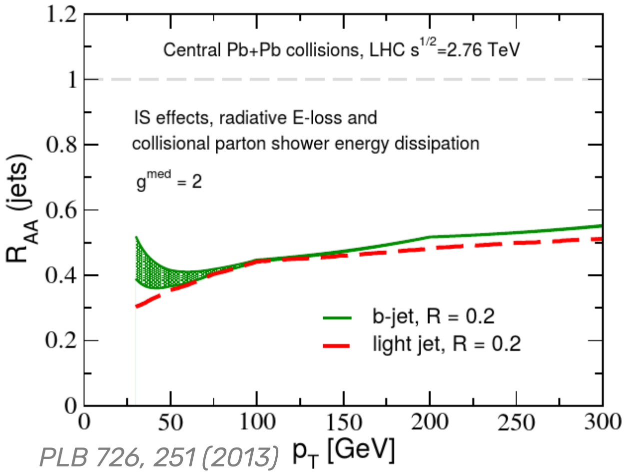
Utrecht University

b-jets: Motivation

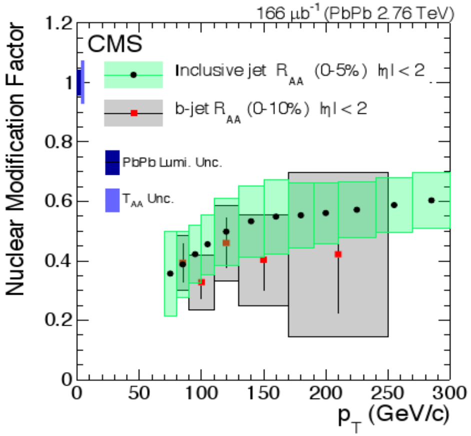
- *pp collisions:* sensitive probes of pQCD
- *pA collisions:* initial-state effects
- *AA collisions:* energy loss of hard-scattered partons via collisional and radiative processes
  - Flavour dependence of the jet quenching
  - Spatial redistribution of the lost energy

→ **b-jets: probes of the QGP transport properties**





PLB 726, 251 (2013)



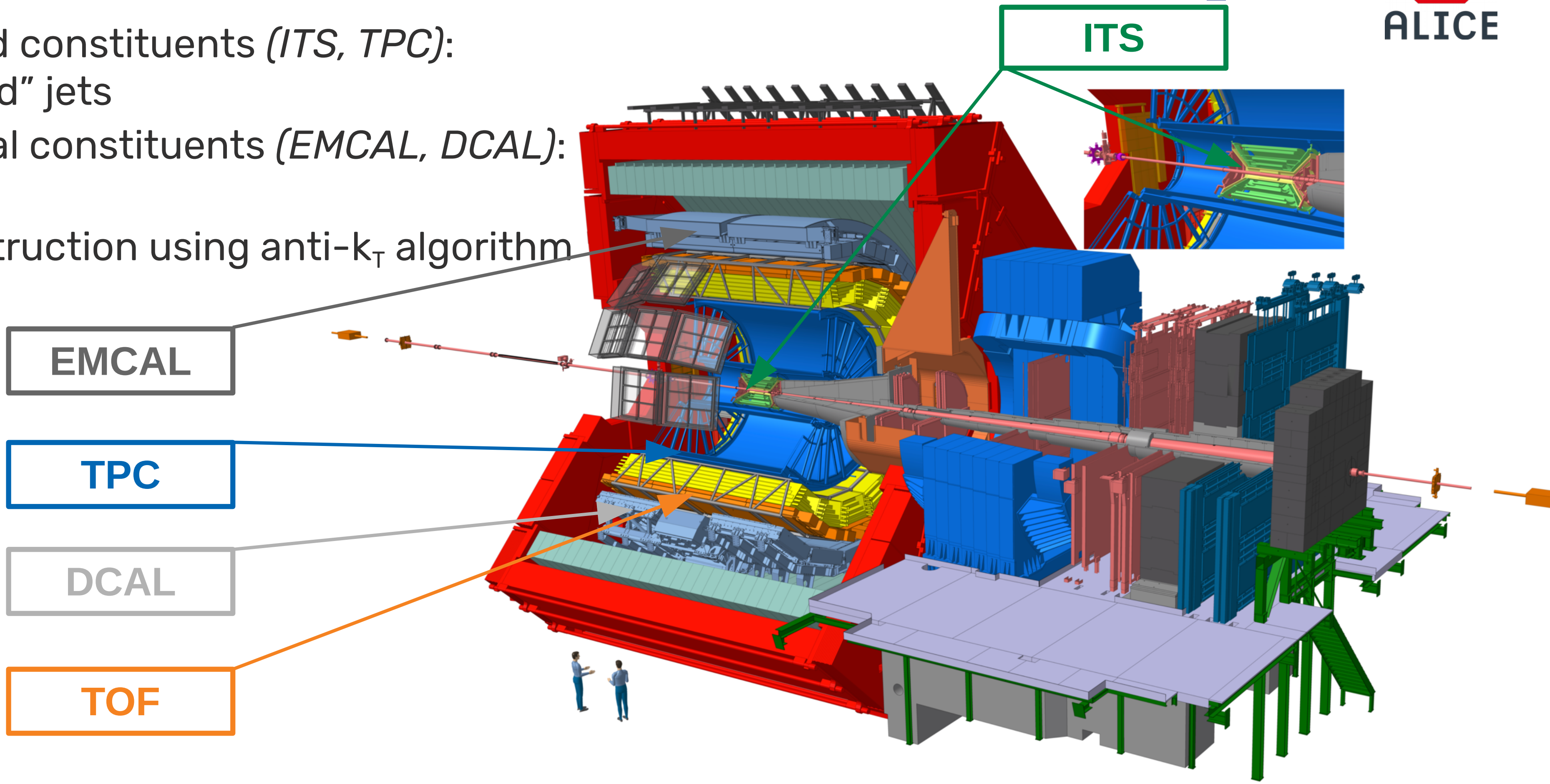
PRL 113, 132301 (2014)

- Mass effects relevant $p_T < 70$ GeV/c

  → Lower jet p_T reach with ALICE

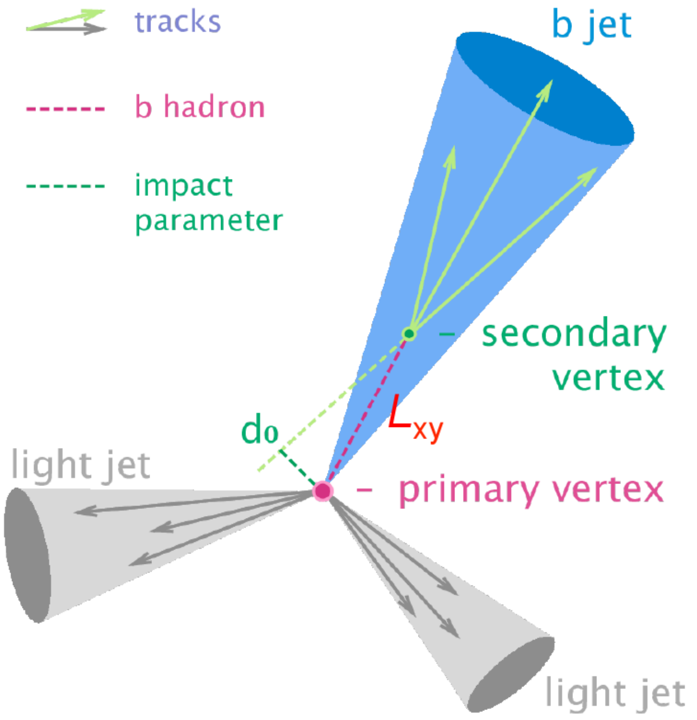
b-jets with ALICE

- Charged constituents *(ITS, TPC)*: "charged" jets
- + neutral constituents *(EMCAL, DCAL)*: full jets
- Reconstruction using anti-k_T algorithm



b-jets with ALICE

- Charged constituents *(ITS, TPC)*: "charged" jets
- + neutral constituents *(EMCAL, DCAL)*: full jets
- Reconstruction using anti-k_T algorithm
  - → b-tagging exploiting:
    - B-hadron long lifetime, $c\tau$ ~ 500 μm
      - → Displaced from primary vertex
    - Its large mass
  - → Studied b-tagging algorithms:
    - Secondary vertex: using displaced vertices
    - Track counting: based on single tracks
    - Via heavy-flavour electron identification (charm+beauty)

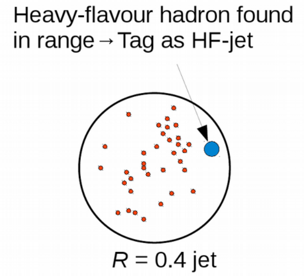


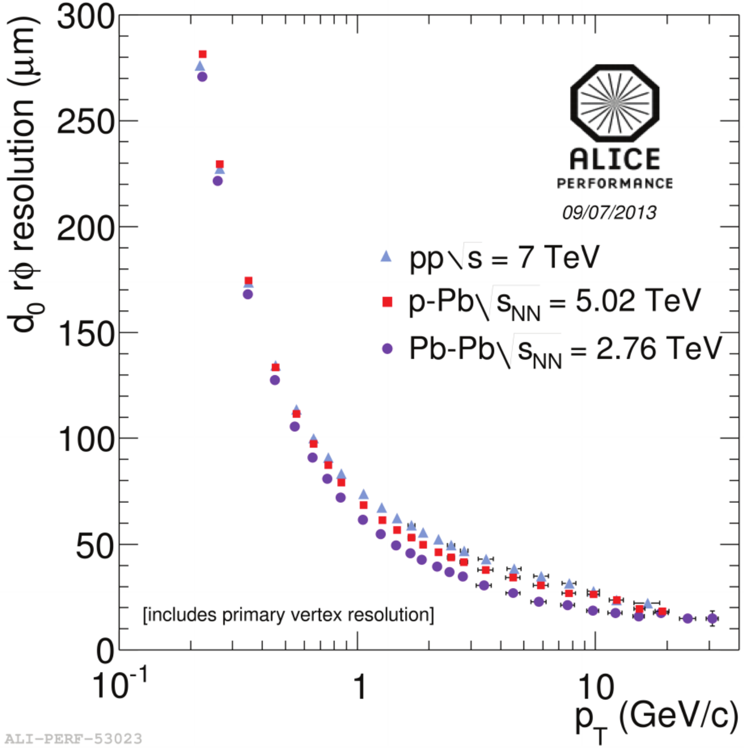
http://bartosik.pp.ua/hep_sketches/btagging

I. Secondary Vertex algorithm

- Exploiting long lifetime and large vertex mass
- TPC and ITS used for tracking and secondary vertex reconstruction
  - Track impact parameter d_0 resolution < 75 μm for p_T > 1 GeV/c
  - Secondary vertex resolution ~120 μm

  - b-jet if B-hadron within given R
  - Otherwise c-jet if charm hadron within given R
  - Otherwise light-flavour jet





SV algorithm: Simulations

- p-Pb at $\sqrt{s_{NN}}$ = 5.02 TeV, Pythia 6 + Hijing
- FastJet anti-k_T, R = 0.4, p_T^{track} > 150 MeV/c
- 3-prong vertices: vertices reconstructed from 3-track combinations, p_T^{prong} > 1 GeV/c

➔ Discriminators:

- Significance of signed secondary vertex flight distance $L_{xy}/\sigma_{L_{xy}}$:

$$L_{xy} = |\vec{L}'| \mathrm{sign}(\vec{L}' \cdot \vec{p}_{jet})$$

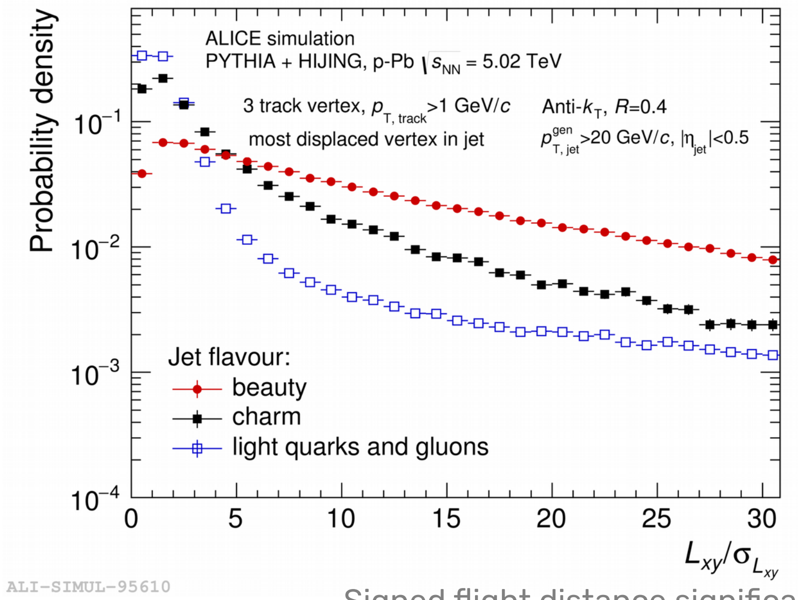
$\vec{L}'$ – vector between primary and secondary vertices

$\sigma_{L_{xy}}$ – uncertainty corresponding to L_{xy}

- SV dispersion (vertex quality measure):

$$\sigma_{vtx} = \sqrt{d_1^2 + d_2^2 + d_3^2}$$

$d_{1,2,3}$ – distances of the tracks from secondary vertex



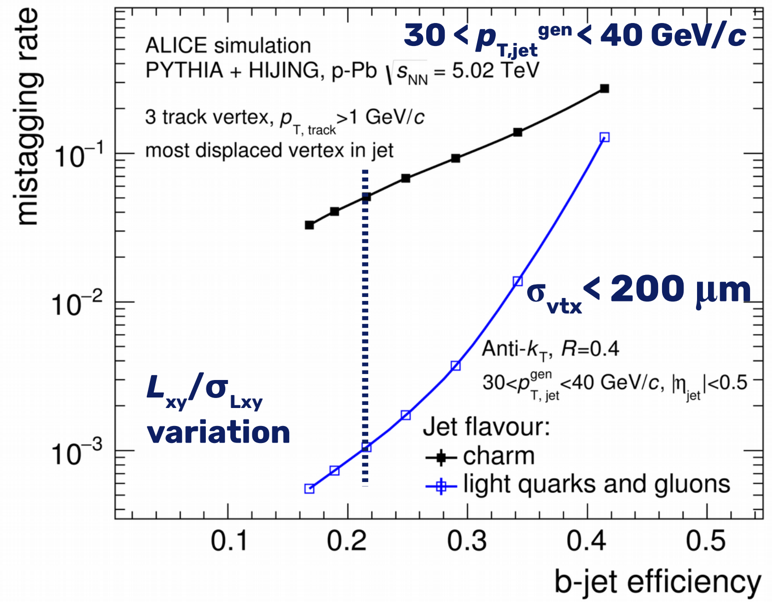
Signed flight distance significance of the most displayed secondary vertex

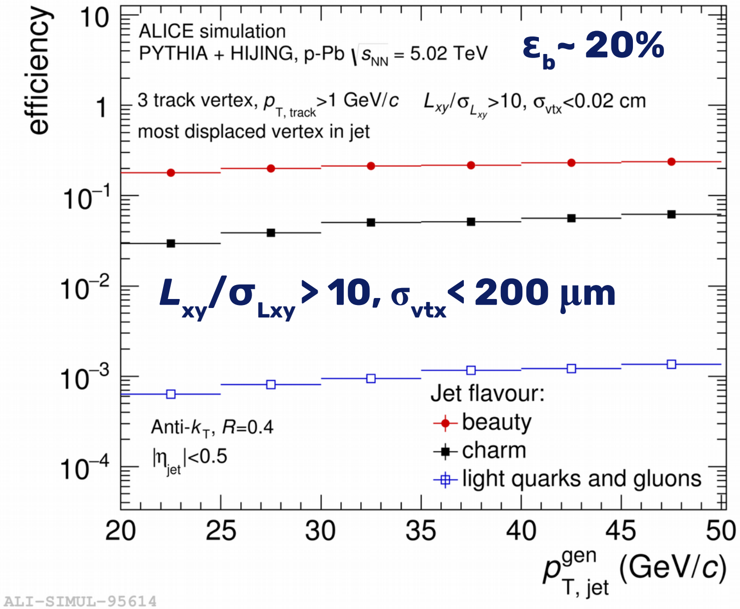
SV algorithm: Performance

- "Rectangular" cuts on the vertex properties
  - b-jet tagging efficiency and c-/udsg-jet misidentification
  - The higher b-jet efficiency, the higher the c-/udsg-jet mistagging efficiency

➔ Find condition of high purity and reasonably high efficiency

$$\epsilon_b(p_T) = \frac{dN_b^{tagged}/dp_T}{dN_b/dp_T}$$





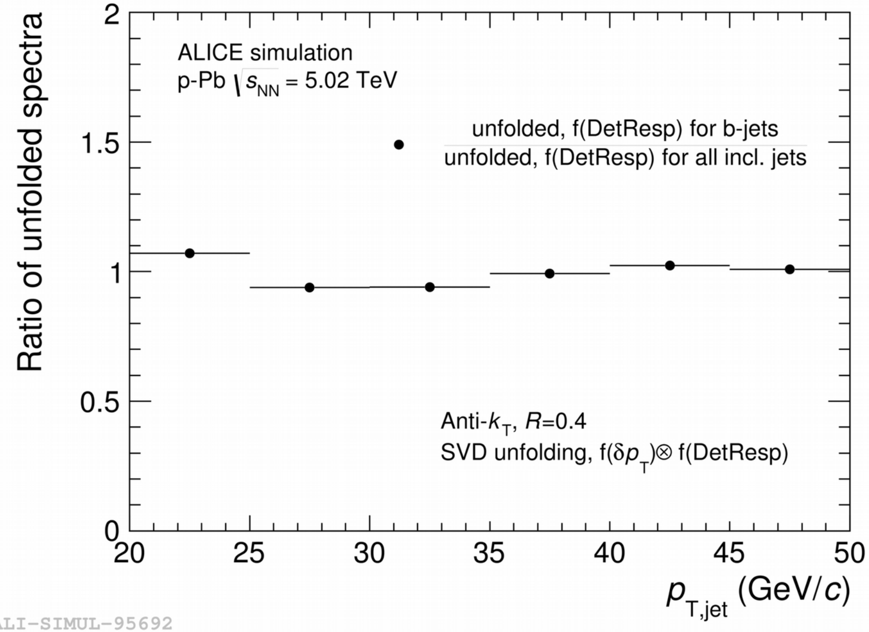
SV algorithm: Unfolding

- SVD (Singular Value Decomposition) *[1]* used for unfolding
- Background subtraction: background density calculated using CMS method *[2]*, soft clusters found using FastJet k_T with R = 0.4

  background density: $\rho_{\mathrm{CMS}} = median\{\frac{p_{\mathrm{T,i}}}{A_{\mathrm{i}}}\} \cdot \mathrm{C}$

  C: correction factor for empty clusters

- Unfolding with combined detector and background fluctuations matrix
  - Background fluctuations using Random Cone method in MC: $\delta p_T = \sum_i p_{\mathrm{T,i}} - \rho A_{\mathrm{cone}}$
- Correction stability tests performed
  - E.g.: applying efficiency and purity corrections before and after the unfolding



ALI-SIMUL-95692

[1] H. Hoecker, V.Kartverlishvili, Nucl. Instrum. Meth. A372 (1996) 469
[2] CMS collaboration, JHEP 1208 (2012) 130

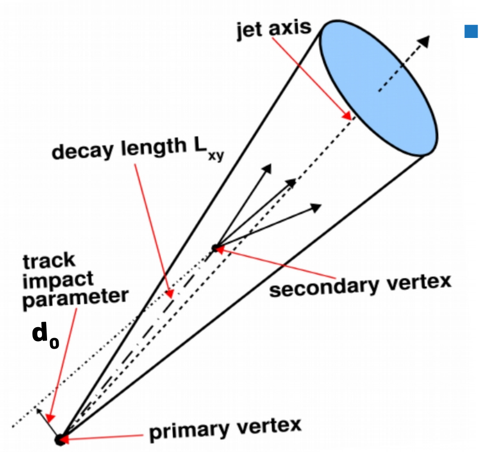
II. Track Counting algorithm

- Impact parameter d_0 in rφ for each track within a jet

$$\mathrm{sign}\left(d_0^{\mathrm{jet}}\right)=\mathrm{sign}\left(\vec{d}_0\cdot\vec{p}_{\mathrm{T,jet}}\right)$$

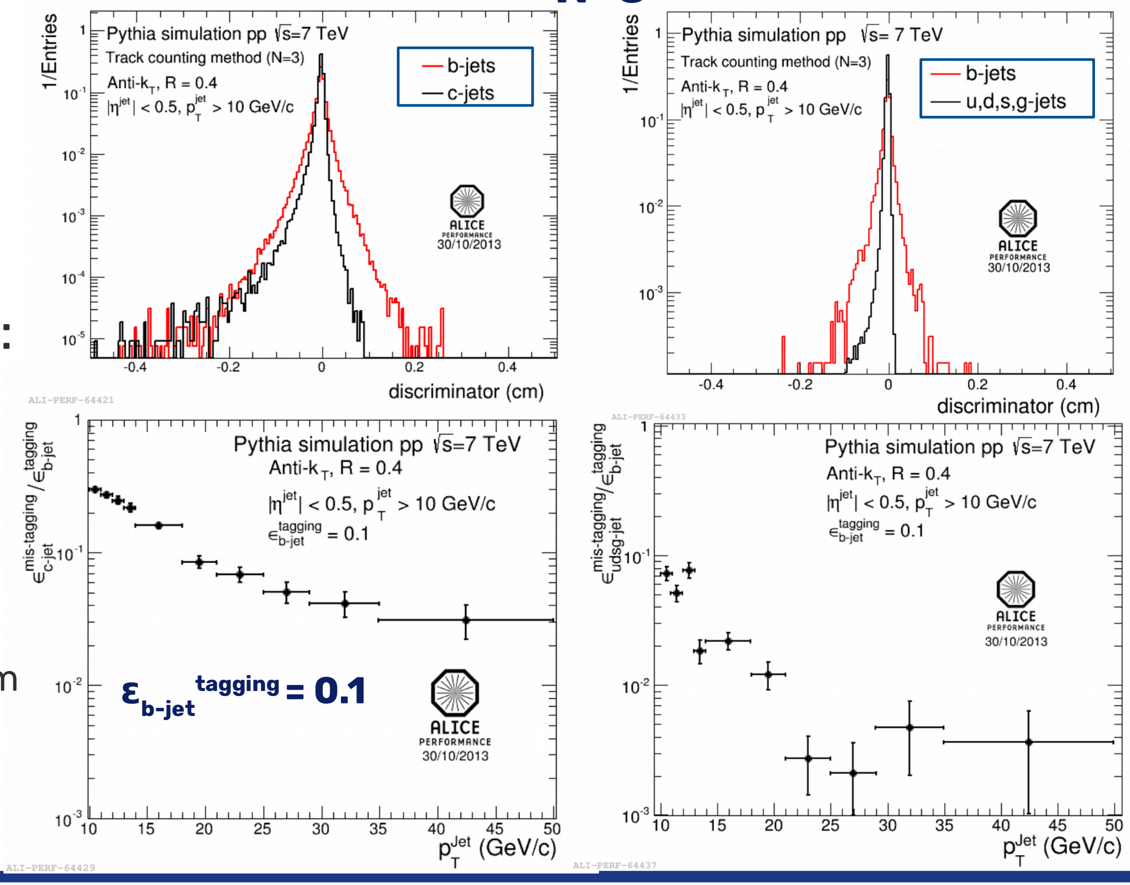
➔ Discriminator:

third, second or first (N=3,2,1) most displaced sign(d_0^{jet})



- Pythia simulation: pp at $\sqrt{s}$ = 7 TeV
  - FastJet anti-k_T, R=0.4
  - p_T^{track} > 150 MeV/c
  - p_T^{jet} > 10 GeV/c
  - d_0 rφ > 0.2 cm
  - DCA(jet,track) < 700 μm
  - ΔR(jet,track) < 0.3

N=3

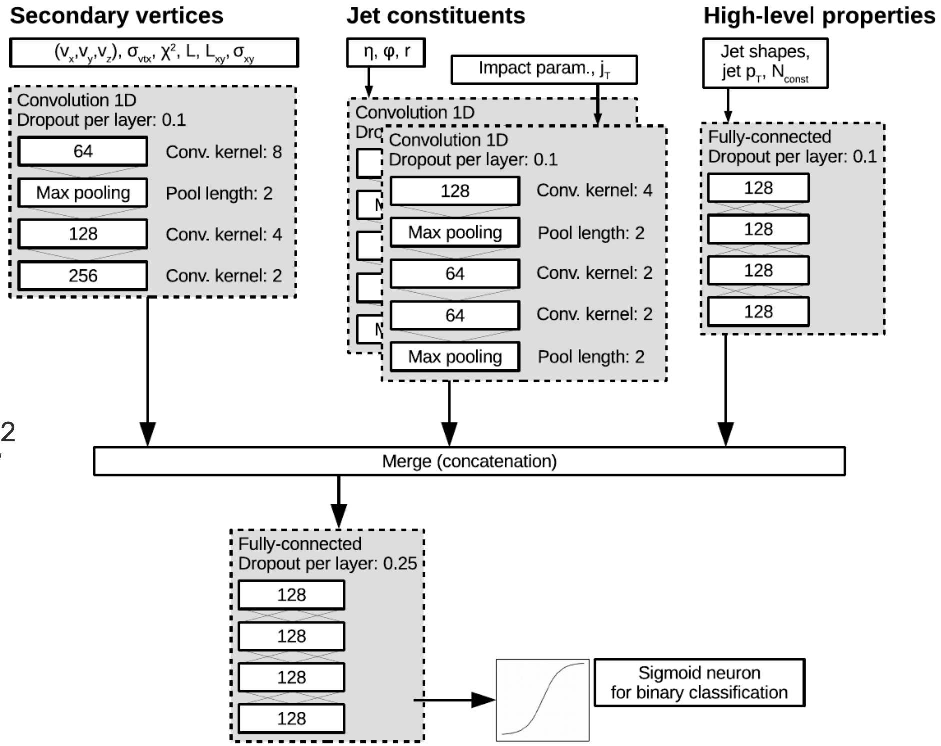


III. Machine-Learning based algorithm

- ML techniques applied to several low-level inputs: constituents, secondary vertices, track impact parameters
- General design: multibranched, multilayered neural network
- Different networks tested on different features

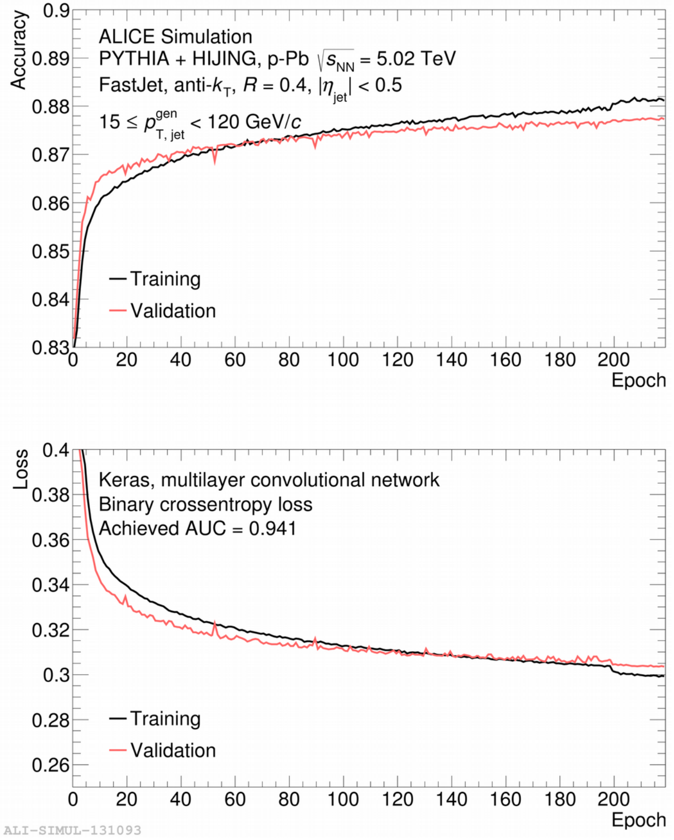
➔ Features:

- Array of secondary vertices:
  - (x,y,z) rel. to primary vertex
  - Transverse plane distance and uncertainty: L_{xy}, σ_{xy}
  - Vertex track dispersion σ_{vtx}, fit quality χ^2
- Array of constituents:
  - η, φ, r (relative to jet axis)
  - Track impact parameters D, Z and j_T



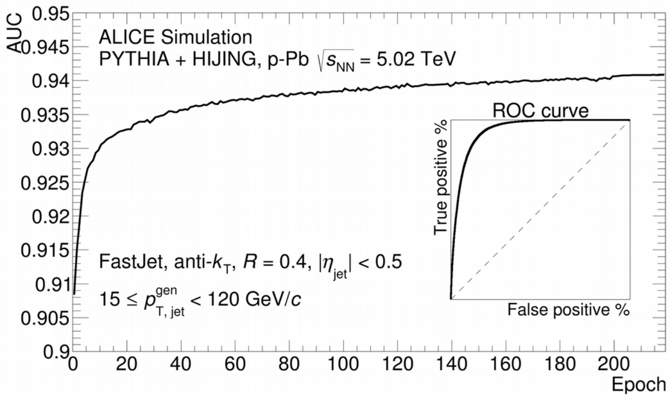
ML method: Simulations

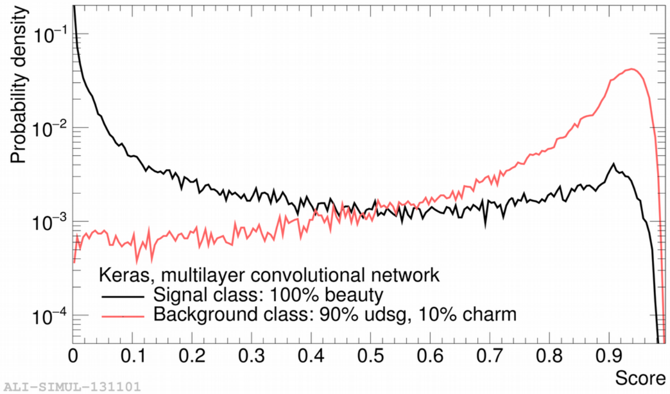
- p-Pb at $\sqrt{s_{NN}}$ = 5.02 TeV, Pythia 6 + Hijing
- FastJet anti-k_T, R = 0.4
- Underlying event corrected
- 200k training, 50k validation samples
  - Control parameters: accuracy and loss
  - Slow learning up to high epoch counts
  - Learning rate parameters lowered after 200 epoch
  - Not much to gain with longer training



ML method: Simulations

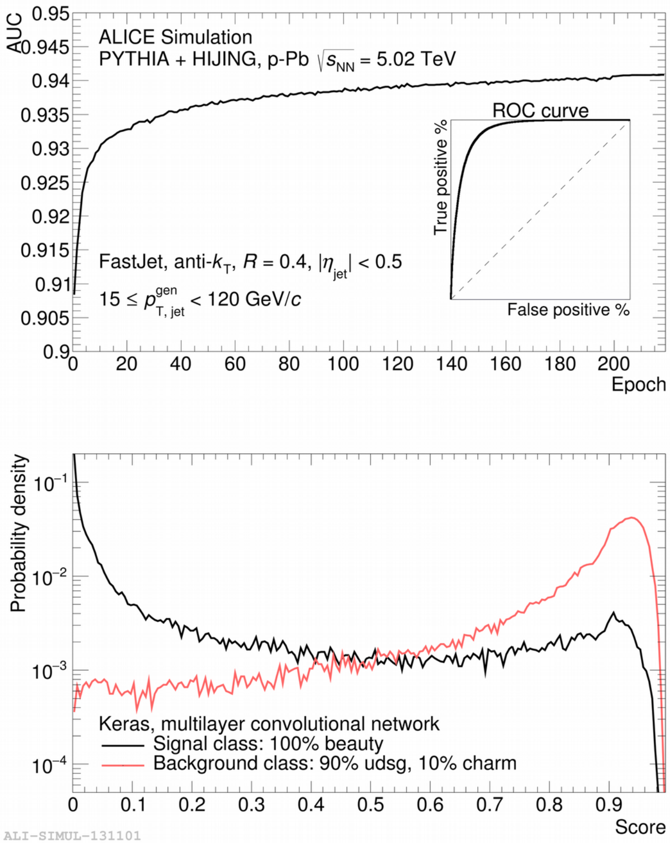
- p-Pb at $\sqrt{s_{NN}}$ = 5.02 TeV, Pythia 6 + Hijing
- FastJet anti-k_T, R = 0.4
- Underlying event corrected
- 200k training, 50k validation samples
  - Control parameters: accuracy and loss
  - Slow learning up to high epoch counts
  - Learning rate parameters lowered after 200 epoch
  - Not much to gain with longer training
  - AUC: Area Under ROC Curve, slow but constant learning up to 220 epoch
  - Clearly separated score distribution





ML method: Simulations

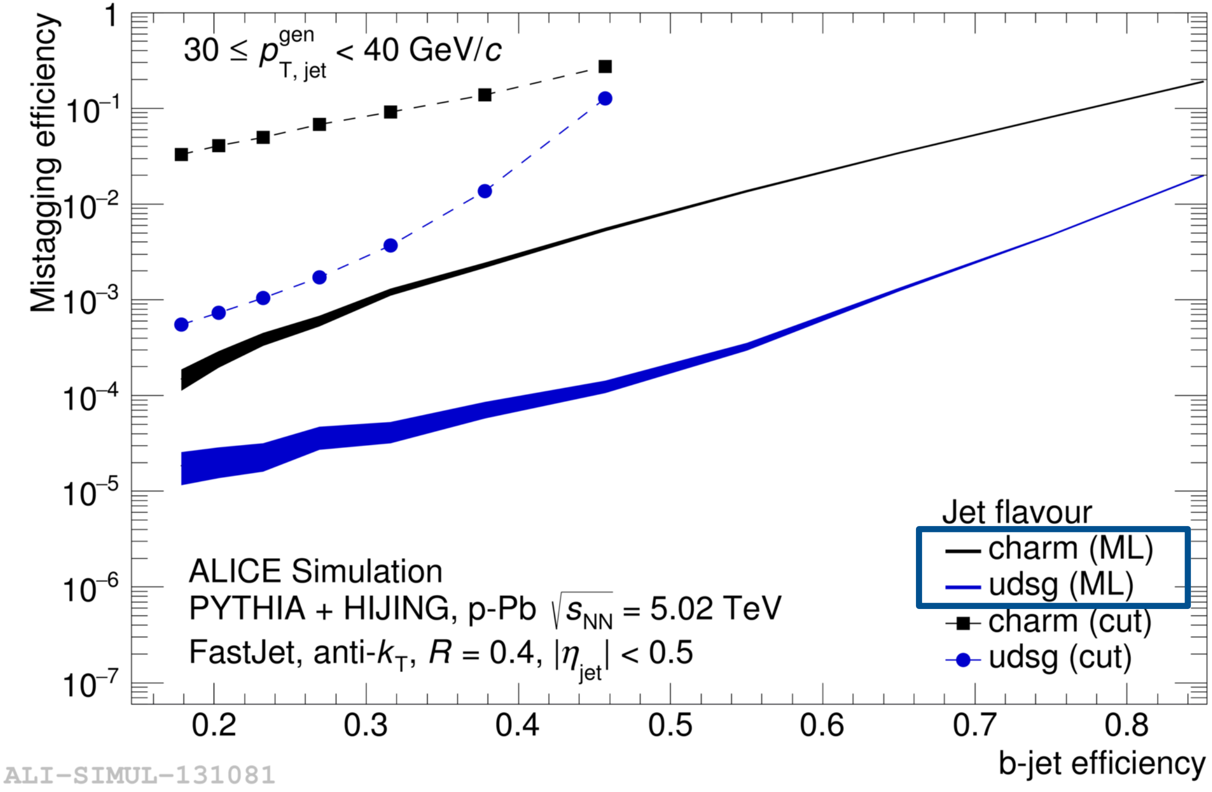
- p-Pb at $\sqrt{s_{NN}}$ = 5.02 TeV, Pythia 6 + Hijing
- FastJet anti-k_T, R = 0.4
- Underlying event corrected
- 200k training, 50k validation samples
  - Control parameters: accuracy and loss
  - Slow learning up to high epoch counts
  - Learning rate parameters lowered after 200 epoch
  - Not much to gain with longer training
  - AUC: Area Under ROC Curve, slow but constant learning up to 220 epoch
  - Clearly separated score distribution



➔ Comparison to "rectangular" cuts method

➔ Optimizing: b-jet tagging and c-/udsg-mistagging efficiency

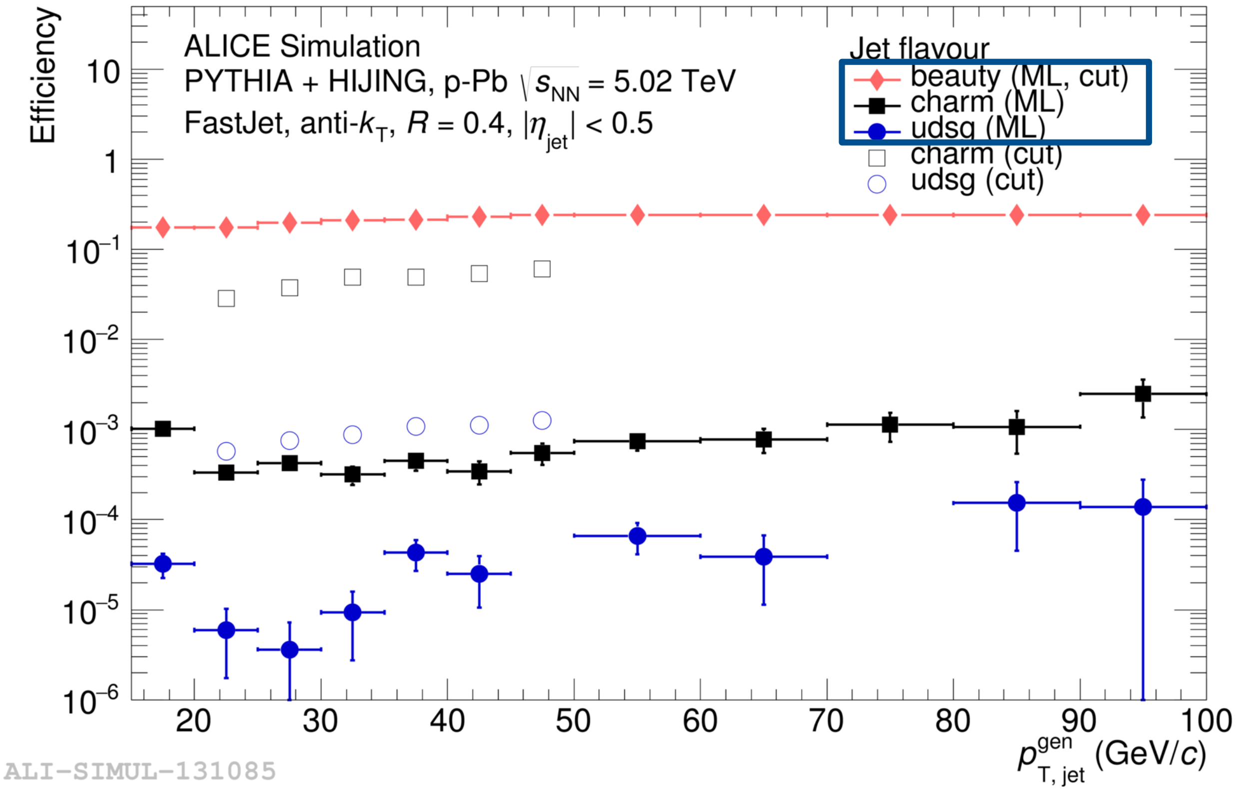
ML: b-jet vs c,udsg mistagging efficiency



- **Mistagging efficiency much lower for c-/udsg-jets**
- Very promising method

- Solid: ML-based method
- Dashed: cut-based method (previous SV slides)

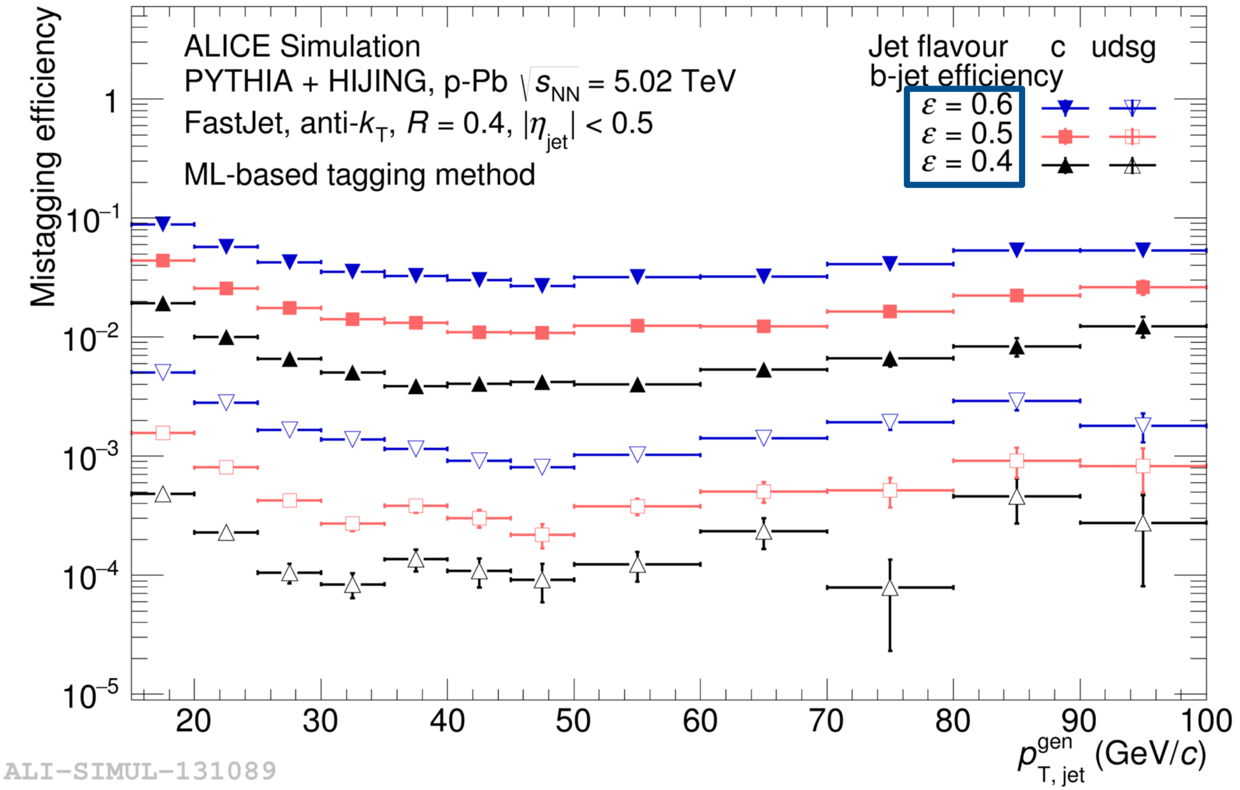
ML: Mistagging efficiency vs jet p_T



- Mistagging efficiency vs jet p_T
- Fixed (~20%) b-jet efficiency to compare to cut-based method

- Solid: ML-based method
- Open: cut-based method (previous SV slides)

ML: Mistagging efficiency



- Mistagging efficiency for higher b-jet efficiencies
  - c-jet efficiency: below 5-10%
  - udsg-jet efficiency: below 0.5-1%
- **Higher b-jet efficiency possible**

- Solid: c-jets
- Open: udsg-jets

Summary

- Performances of different b-tagging jet algorithms have been studied in pp and p-Pb MC simulations
  - Based on track counting, displaced secondary vertices and machine learning
  - **Very promising ML-based method in pp, p-Pb**
    - Allows for much higher b-jet efficiency
- Data analysis (in pp and p-Pb) being finalized
- Studies will be extended to Pb-Pb collisions: upcoming Pb-Pb run and run 3 and 4
  - Major detector upgrade with new ITS (x3 (x5) better spatial resolution on rφ (z) coordinates), improved readout (able to sustain 50 kHz Pb-Pb collisions: collect L_{int}=10 nb^{-1}, x100 gain for min. bias)
    - → **Major boost for heavy-flavour jet physics with ALICE**

Summary

- Performances of different b-tagging jet algorithms have been studied in pp and p-Pb MC simulations
  - Based on track counting, displaced secondary vertices and machine learning
  - **Very promising ML-based method in pp, p-Pb**
    - Allows for much higher b-jet efficiency
- Data analysis (in pp and p-Pb) being finalized
- Studies will be extended to Pb-Pb collisions: upcoming Pb-Pb run and run 3 and 4
  - Major detector upgrade with new ITS (x3 (x5) better spatial resolution on $r\varphi$ (z) coordinates), improved readout (able to sustain 50 kHz Pb-Pb collisions: collect L_{int}=10 nb^{-1}, x100 gain for min. bias)
    - ➔ **Major boost for heavy-flavour jet physics with ALICE**

Thank you !