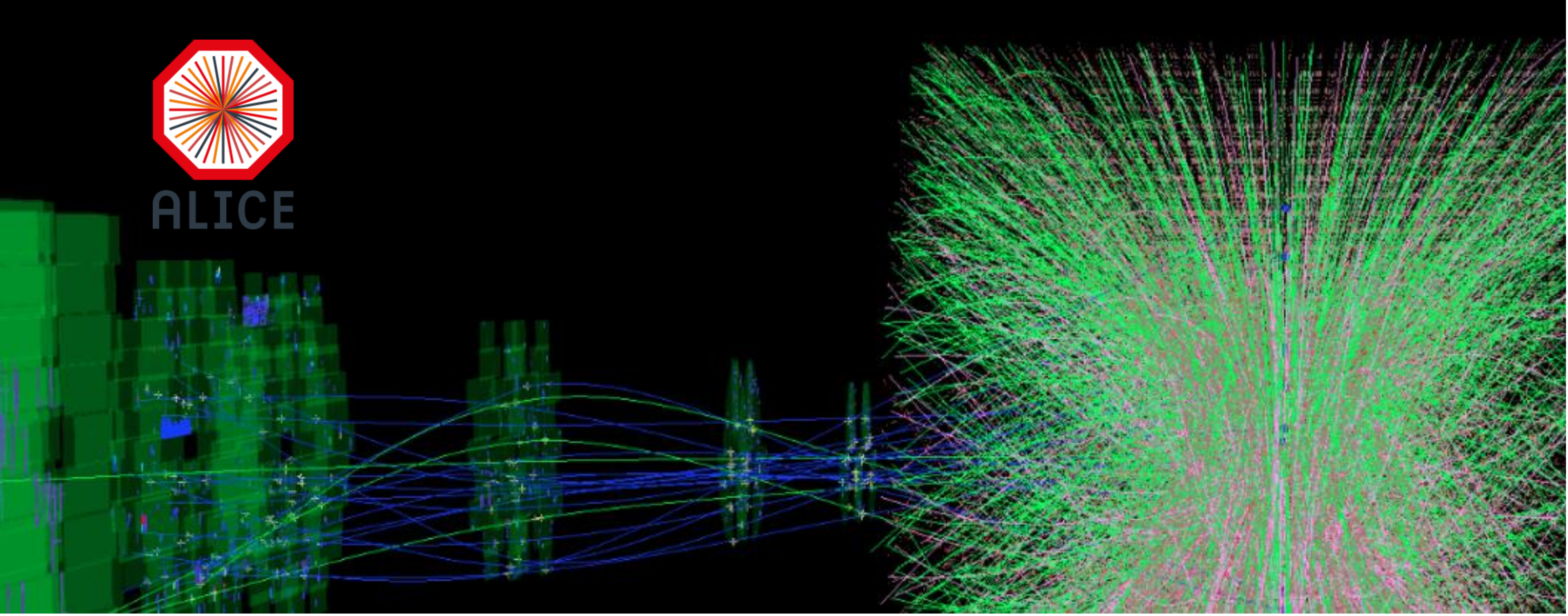


PERFORMANCE OF THE ALICE SECONDARY VERTEX B-TAGGING ALGORITHM

Lukáš Kramárik

on behalf of the ALICE collaboration

Faculty of Nuclear Sciences and Physical Engineering

Czech Technical University in Prague

Zimányi school 2015

7.12. – 11.12.2015

CONTENT

- Motivation for b-jet studies
- Secondary vertex reconstruction
- b-jet tagging algorithm performance studies
- Unfolding corrections in p-Pb collisions
- Outlook

PROPERTIES OF b-JETS

b quarks

- $m \sim 4.2$ GeV/c^2
- High energy needed for their creation, produced in the collision, not as an effect of medium
- Expected to lose energy via elastic scatterings and radiative processes
- Their energy loss in medium is expected to be smaller than for lighter quarks and gluons (dead-cone effect)

(Yuri L. Dokshitzer and D.E. Kharzeev. *Heavy quark colorimetry of QCD matter.* Phys.Lett., B519:199–206, 2001.)

Jets

- Sprays of collimated particles coming from the fragmentation of hard-scattering partons
- Jet finding algorithms reconstruct the full jet starting from charged tracks and neutral energy

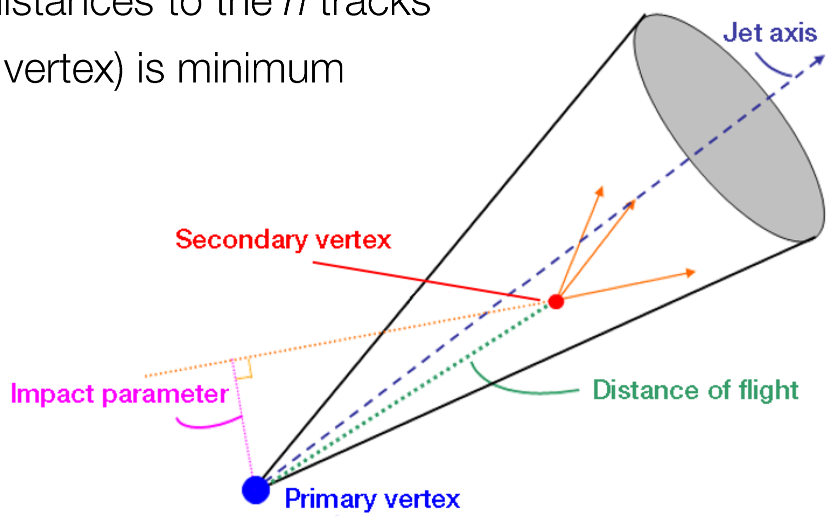
PROPERTIES OF b-JETS

<u>b-jets</u>

- Come from fragmentation of beauty quarks
- Allow studies of redistribution of the lost energy and possible modifications to b-quark fragmentation in the medium
- b-tagging algorithms exploit long lifetime ($c\tau$ ~500 μm) and large mass (~5 GeV/c^2) of B hadrons

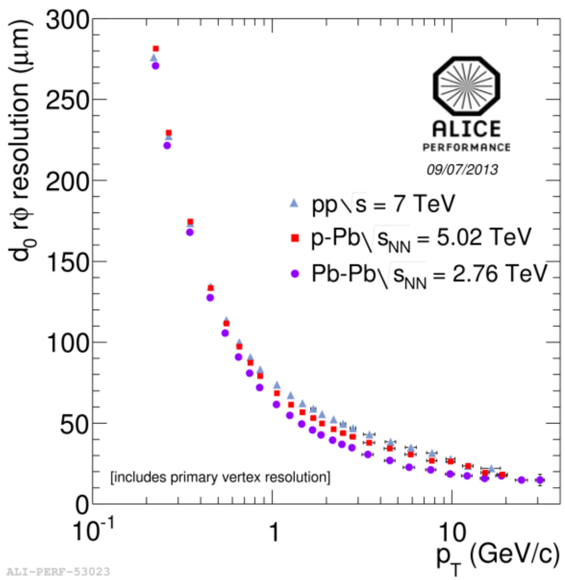
<u>Secondary vertices</u>

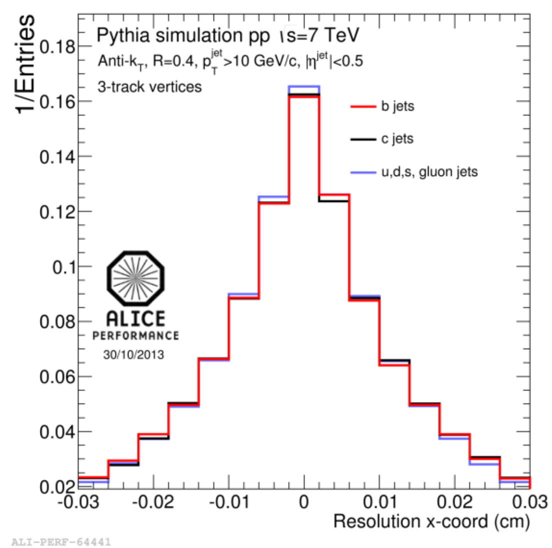
- Calculated as the point for which the sum of the distances to the *n* tracks (approximated to linear in proximity of the primary vertex) is minimum
- Secondary vertices from beauty-hadron decays:
  - Large invariant mass
  - Large secondary vertex distance of flight L_{xy}

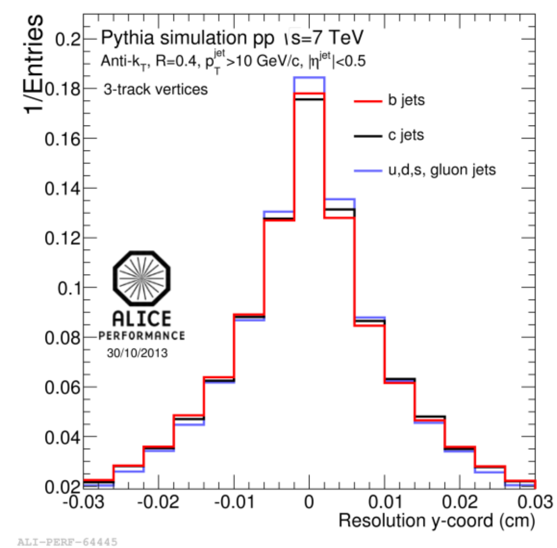


SECONDARY VERTEX RECONSTRUCTION

- Inner Tracking System (ITS) and Time Projection Chamber (TPC) are used for tracking and secondary vertex reconstruction
- Track impact parameter resolution better than 70 μm for $p_T > 1$ GeV/*c*, for all collision systems
- Resulting secondary vertex resolution ~120 μm







Track impact parameter resolution in the transverse plane vs p_T

3-prong vertex resolution (in the xy plane)

SECONDARY VERTEX RECONSTRUCTION

Jet reconstruction:

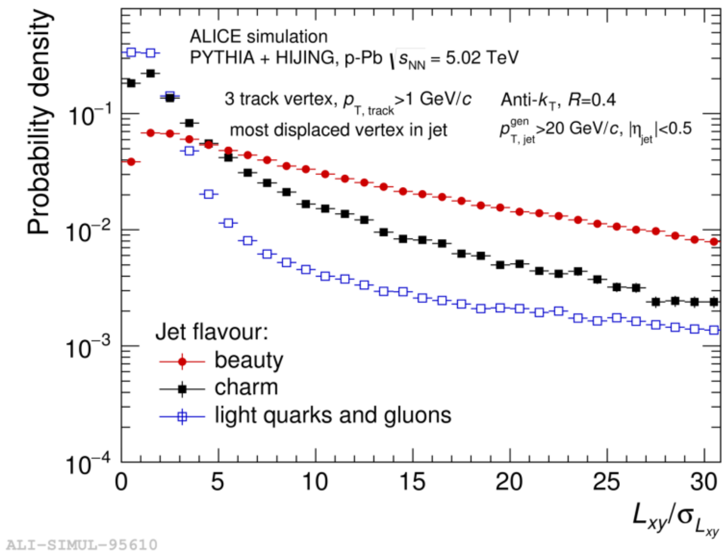
- FastJet Anti-k_T, $R = 0.4$
- $p_{T,\ track} > 0.15$ GeV/c

Secondary vertex reconstruction:

- Using 3 tracks
- Tracks in a jet $p_{T,\ prong} > 1$ GeV/c

Discriminating variables for b-tagging:

- Significance of signed secondary vertex distance of flight L_{xy}/σ_{Lxy} in the transverse plane
- Secondary vertex dispersion, $\sigma_{vtx} = \sqrt{d_1^2 + d_2^2 + d_3^2}$ where $d_{1,2,3}$ are the distances of tracks from secondary vertex



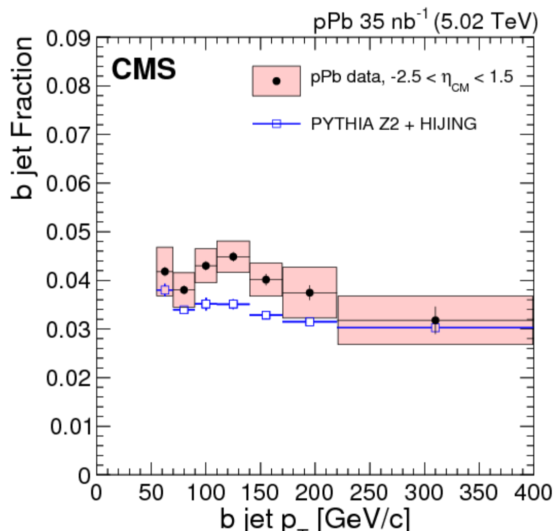
Signed flight distance significance of secondary vertex. Jets of different flavour defined in MC based in presence of B- (D-) hadron.

SECONDARY VERTEX ALGORITHM PERFORMANCE IN p-Pb COLLISIONS

- Tagging efficiency:

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

- The goal is to find the working point (cuts used in further analysis) with high purity and reasonably high efficiency at the same time
- b-jet fraction from CMS ~2-4% ⇒ cuts with mistagging rate 100 times smaller than tagging efficiency are needed to reduce background



Transverse momentum spectra of b-jets in pPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV,
CMS Collaboration (Khachatryan, Vardan et al.), arXiv:1510.03373, CMS-HIN-14-007, CERN-PH-EP-2015-205.

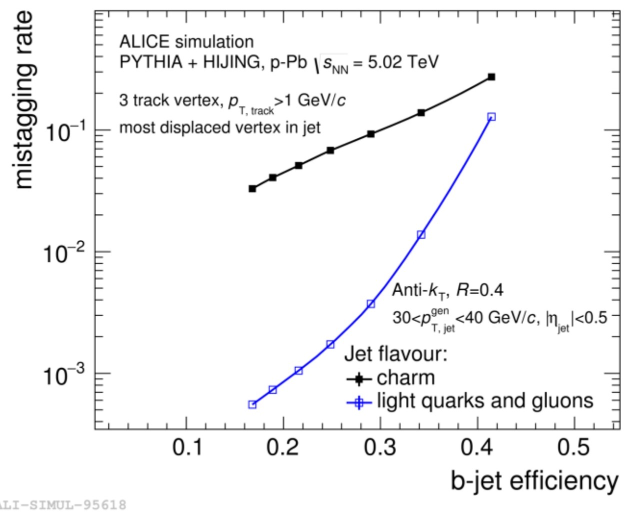
b-jet fraction in p-Pb collisions for different rapidity ranges for $50 < p_{T,\,jet} < 200$ GeV/c from CMS experiment

SECONDARY VERTEX ALGORITHM PERFORMANCE IN p-Pb COLLISIONS

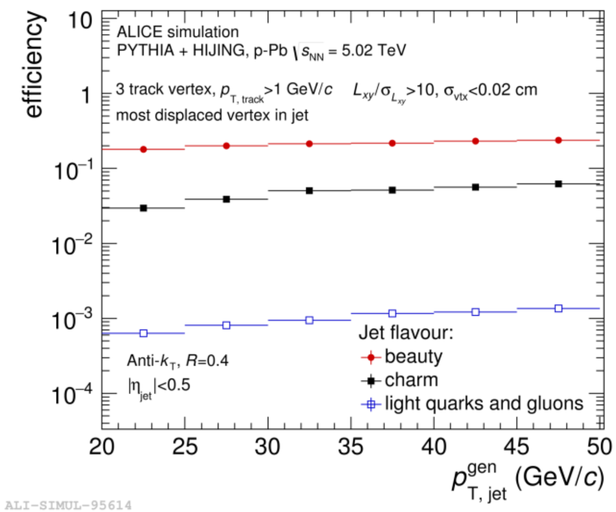
- Tagging efficiency:

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

- The goal is to find the working point (cuts used in further analysis) with high purity and reasonably high efficiency at the same time
- Cut variation ⇒ different tagging and mistagging efficiencies



Tagging efficiency and mistagging rate with different cuts on L_{xy}/σ_{Lxy} of secondary vertex for $30 < p_{T,\ jet} < 40$ GeV/c.



(Mis)Tagging efficiencies of secondary vertex algorithm for $L_{xy}/\sigma_{Lxy} > 10$ and $\sigma_{vtx} < 0.02$ cm.

UNFOLDING CORRECTIONS IN p-Pb COLLISIONS

- Measured spectrum is a convolution of a true spectrum and detector response function
- True jet spectrum is found via unfolding procedure

- Singular Value Decomposition (SVD) unfolding was used

(Hocker, Andreas et al., *SVD approach to data unfolding*, Nucl.Instrum.Meth. A372 (1996) 469-481.)

- Constant background density ρ_{CMS} is subtracted from $p_{\mathrm{T,\,jet}}$:

$$\rho_{CMS} = \mathbf{median}\left\{\frac{p_{T,i}}{A_i}\right\}\cdot C$$

  - $p_{\mathrm{T},i}$ are soft clusters, found by FastJet k_{T}, A_i is the jet area, C is the correction factor for the empty clusters.

- Background fluctuations δp_{T} are calculated with MC (HIJING) via random cone method:

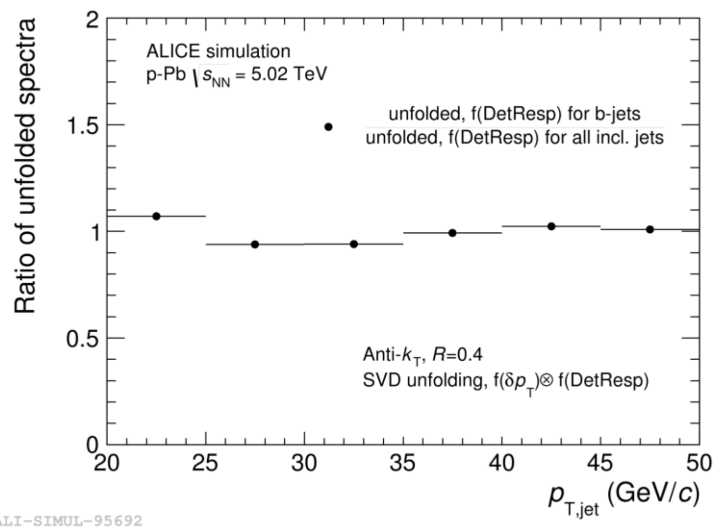
$$\delta p_T = \sum_i p_{T,i} - \rho \cdot A_{cone}$$

  - sum is over track $p_{\mathrm{T},i}$ in a cone.

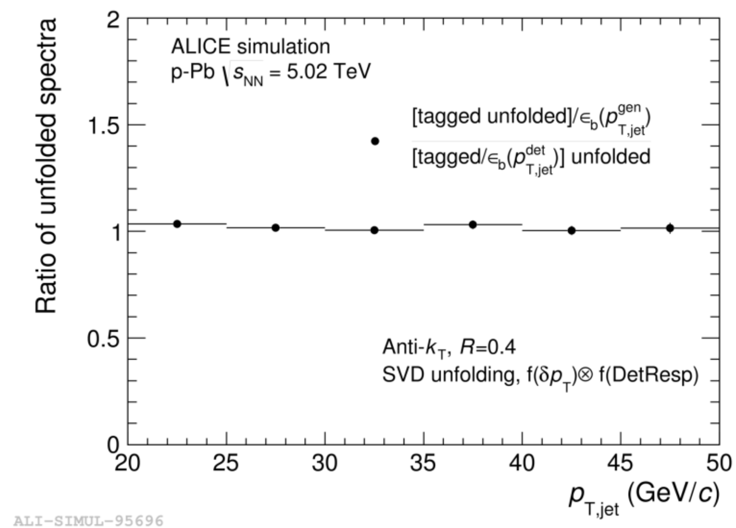
UNFOLDING CORRECTIONS IN p-Pb COLLISIONS

- MC b-jet p_T spectrum is unfolded with the **combined matrix**:
  - detector response
  - background fluctuations of jets
- Comparison of unfolding with **response matrix** for b-jets and for inclusive jets

- Tagged spectrum should be:
  - **Unfolded**
  - Corrected for **tagging efficiency** and purity
- Estimation of systematic influence of the order of corrections



Same results when unfolding spectrum with detector response matrix of inclusive and b-jets



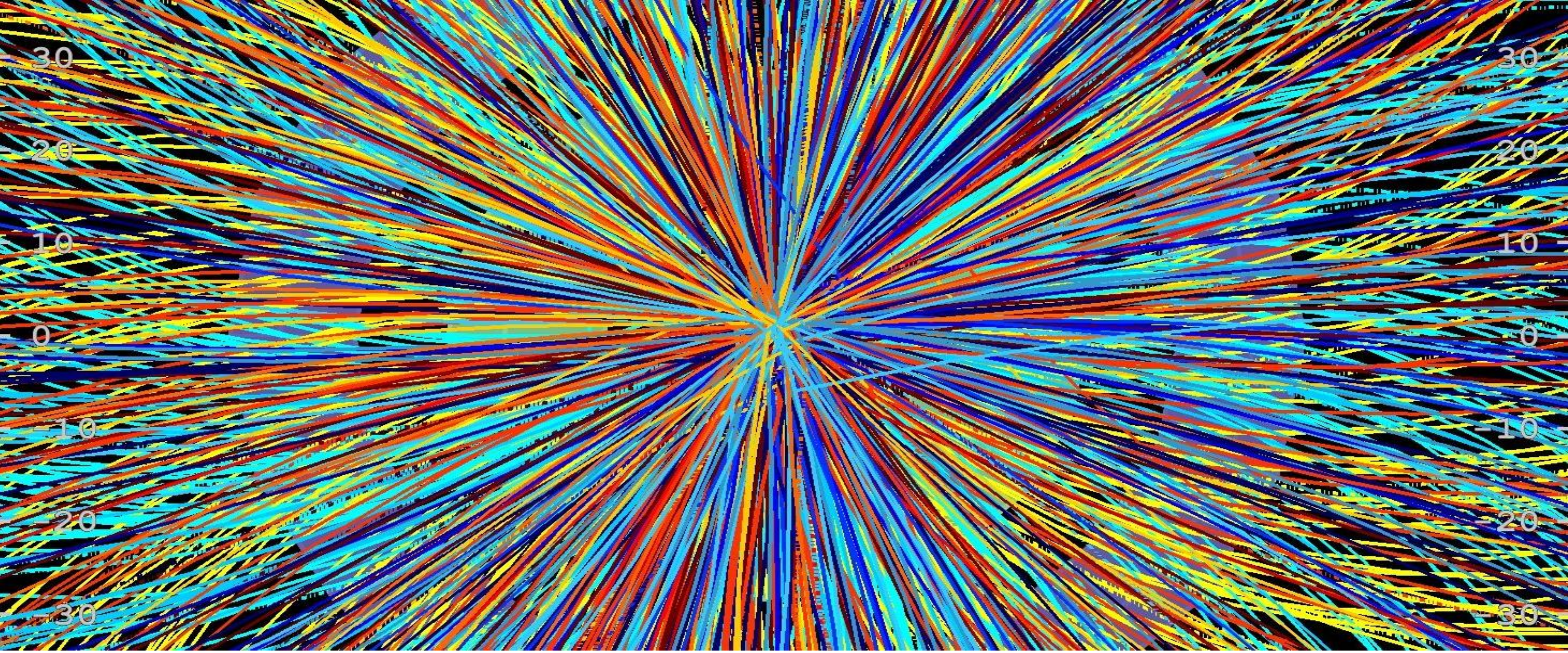
It was verified that the correction order does not give significant differences in the resulting spectrum

OUTLOOK

- Tagging cuts for secondary vertex algorithms are being optimized:
  - beauty efficiencies as large as possible
  - keeping charm and light-jet contamination small
    - b-jet fraction ~2-4% ⇒ cuts which have 100 times smaller contamination of light flavors than b-jet efficiency are needed
- Corrections for background, background fluctuations and for detector response are implemented
- b-jet spectrum can be unfolded using a detector response matrix for all inclusive jets
- Order of corrections (tagging efficiency vs unfolding) gives compatible results

Ongoing studies:

- MC and data-driven estimation of tagging purity
- Track cuts and selection for secondary vertex reconstruction in order to obtain higher purity



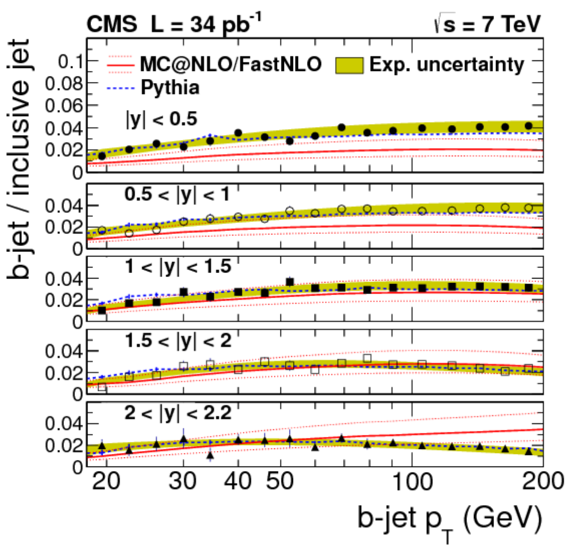
THANK YOU FOR YOUR ATTENTION

ACKNOWLEDGMENT

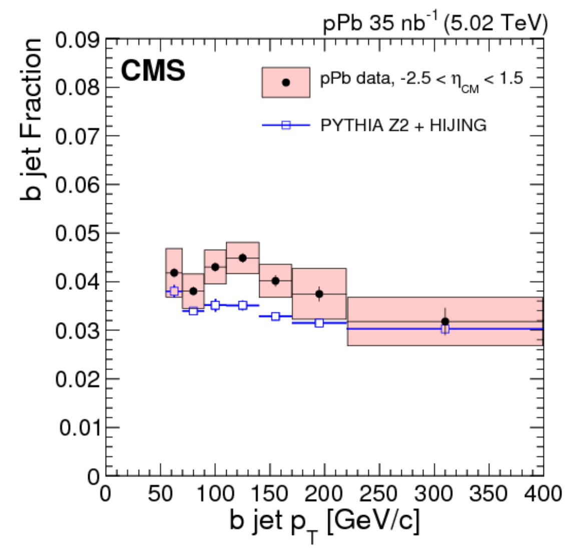
This work was supported by the European social fund within the framework of realizing the project „Support of inter-sectoral mobility and quality enhancement of research teams at Czech Technical University in Prague“, CZ.1.07/2.3.00/30.0034 and by Grant Agency of the Czech Technical University in Prague, grant No. SGS13/215/OHK4/35/14.

BACKUP SLIDES

b-JET FRACTION IN p-Pb AND pp FROM CMS



b-jet fraction in pp collisions for different rapidity ranges for $20 < p_{T,\, jet} < 200$ GeV/c from CMS experiment

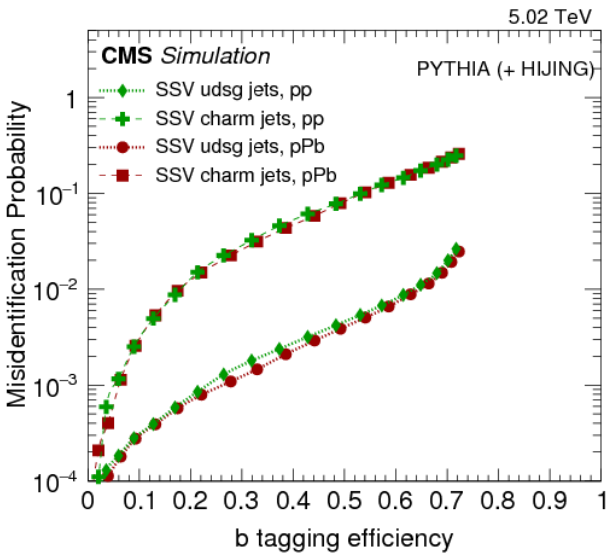


b-jet fraction in p-Pb collisions for different rapidity ranges for $50 < p_{T,\, jet} < 200$ GeV/c from CMS experiment

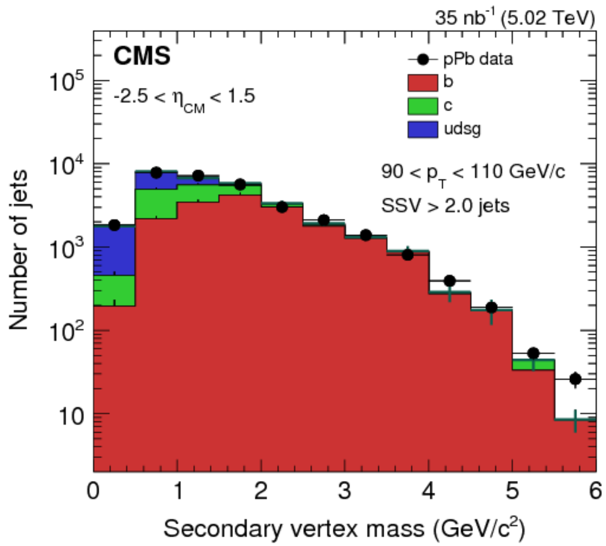
Inclusive b-jet production in pp collisions at $\sqrt{s}$ = 7 TeV, CMS Collaboration (Serguei Chatrchyan et al.) arXiv:1202.4617, JHEP 1204 (2012) 084.

Transverse momentum spectra of b-jets in pPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV, CMS Collaboration (Khachatryan, Vardan et al.), arXiv:1510.03373, CMS-HIN-14-007, CERN-PH-EP-2015-205.

PERFORMANCE OF b-TAGGING IN p-Pb AT CMS



Tagging efficiency and mistagging rate with different cuts on L_{xy}/σ_{Lxy} of secondary vertex in pp and p-Pb collisions.



Template fit to SSV tagged invariant mass distribution for $90 < p_{\mathrm{T,\,jet}} < 110$ GeV/*c*.

Transverse momentum spectra of b-jets in pPb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV,
CMS Collaboration (Khachatryan, Vardan et al.),
arXiv:1510.03373, CMS-HIN-14-007, CERN-PH-EP-2015-205.

UNFOLDING CORRECTIONS, SVD

- Measured spectrum *m(x')* is a convolution of a true spectrum *t(x)* and detector response function *A(x;x')*:

$$m(x') = \int dx A(x;x')t(x)$$

- In practice one solves the discrete equation with the spectra as histograms and the response function as a matrix: $M_m = A_{m,t} \cdot T_t$

- Given a detector response and background fluctuation matrix, the unfolding can be done with the combined matrix $C_{m,t}$. The measured spectrum M_m is the "detector" spectrum D_d distorted by background fluctuations $B_{m,d}$: $M_m = B_{m,d} \cdot D_d$
- The "detector" spectrum D_d is the true spectrum distorted by detector effects $A_{d,t}$: $D_d = A_{d,t} \cdot T_t$
- Thus, measured spectrum is:

$$M_m = B_{m,d} \cdot A_{d,t} \cdot T_t = C_{m,t} \cdot T_t$$

- Finding and inverse matrix is an ill-posed problem, since it is very sensitive to small perturbations of the data.
- Here we used SVD (singular value decomposition) as a main method for our results