Multivariate Analysis Techniques for charm reconstruction with ALICE

Chiara Zampolli for ALICE

2nd IML Machine Learning Workshop

9-12 April 2018

Outline

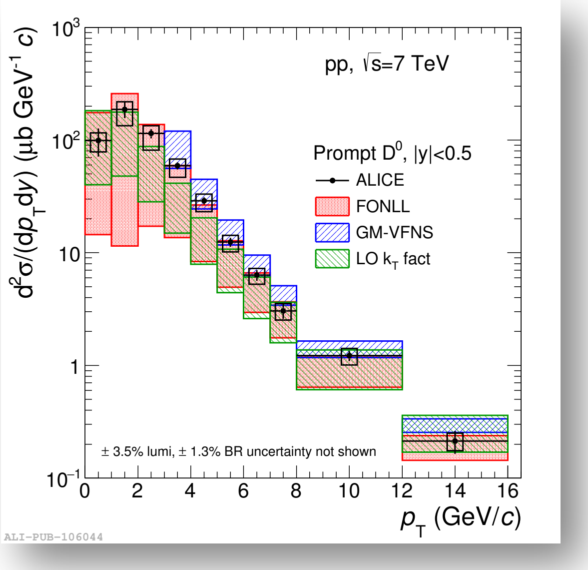
- Physics motivation
- The ALICE detector
  - Vertexing, tracking and particle identification
- Λ_c analysis in ALICE
- Boosted Decision Trees for Λ_c studies
- Results
  - Comparison with standard analysis approach
- Summary and conclusions

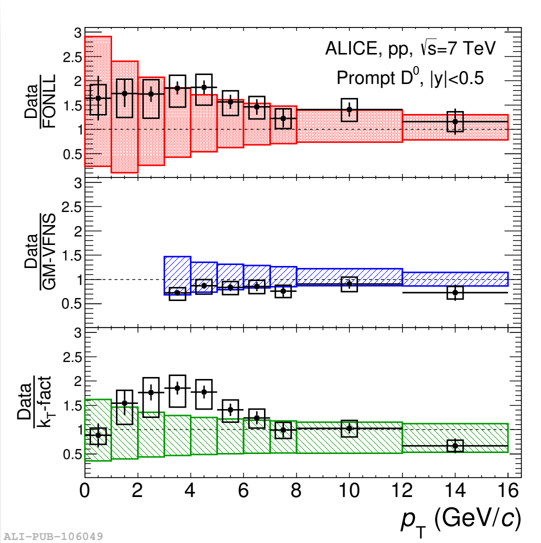
The results shown here are from the first ALICE published analysis using MVA techniques

Disclaimer
The focus of this talk is on methods and techniques more than physics results

Why heavy flavours?

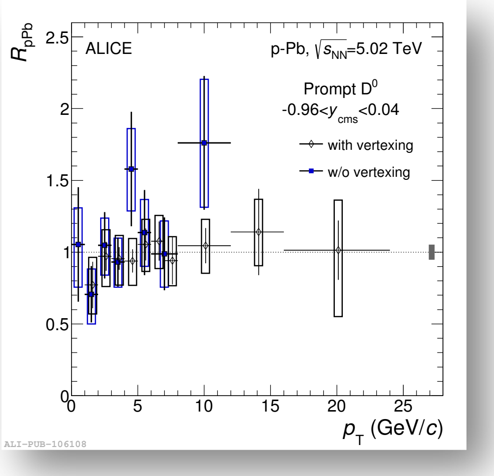
- **pp collisions**: Heavy quarks (HF: c, $m_c \approx 1.3$ GeV/c^2, b, $m_b \approx 4.5$ GeV/c^2) pairs are produced in the hard scattering processes at the initial stages of the collisions with large Q^2
  - Important test of perturbative QCD calculations and Monte Carlo predictions
    - How does hadronization occur? (colour reconnection, strings, ropes, multi parton interactions) Is it different from e^+e^-?
  - Necessary baseline to interpret heavy ion





Phys. Rev. C 94, 054908 (2016)

Why heavy flavours?

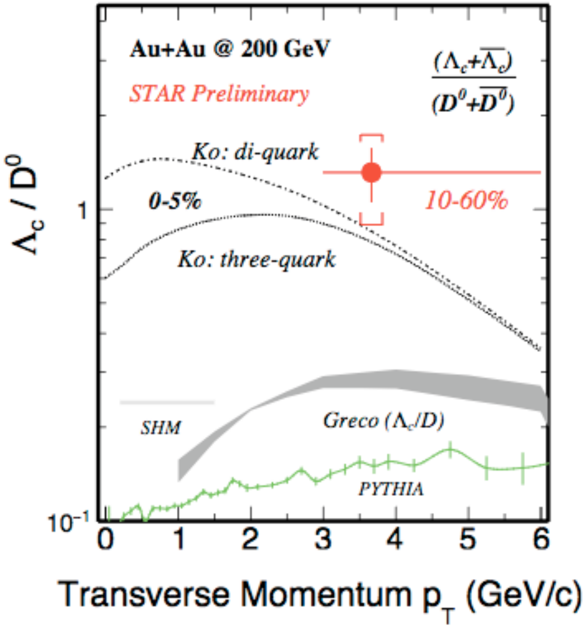


$$R_{\mathrm{pPb}} = \frac{1}{A}\,\frac{\mathrm{d}\sigma^{\mathrm{prompt\,D}}_{\mathrm{pPb}}/\mathrm{d}p_{\mathrm{T}}}{\mathrm{d}\sigma^{\mathrm{prompt\,D}}_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}}$$

Phys. Rev. C 94, 054908 (2016)

- **p-Pb collisions**:
    - Allow to disentangle “hot” medium effects (related to the QGP) from “cold nuclear matter” effects
        - Is a QGP-like state formed in p-Pb collisions?
        - What is just the effect of the presence of a nucleus in the collision?

Why heavy flavours?



Nucl. Phys. A967 (2017) 928–931

- **Pb-Pb collisions**: HF pairs have a formation time (τ = O(0.1) fm) shorter than the life time of the Quark Gluon Plasma (QGP)
  - Experience the evolution of the medium and interact with it
    - Is the hadronization mechanism modified in the presence of the QGP?

Why heavy flavours? Why Λ_c?

- **pp collisions**: Heavy-Flavour quark (HF: c, m_c = 1.275 GeV/c^2, b, m_b = 4.18 GeV/c^2) pairs are produced in the hard scattering processes at the initial stages of the collisions with large Q^2
  - Important test of perturbative QCD calculations and Monte Carlo predictions
    - How does **hadronization** occur? (colour reconnection, strings, ropes, multiparton interactions) Is it different from e^+e^-?
  - Necessary baseline to interpret Heavy-Ion results
- **p-Pb collisions**:
  - Allow to **disentangle "hot"** medium effects (related to the QGP) **from "cold nuclear matter"** effects
    - Is a QGP-like state formed in
    - What is just the effect of the
- **Pb-Pb collisions**: HF pairs have a fo life time of the Quark Gluon Plasm
  - Experience the evolution of th
    - Is the **hadronization** mechan

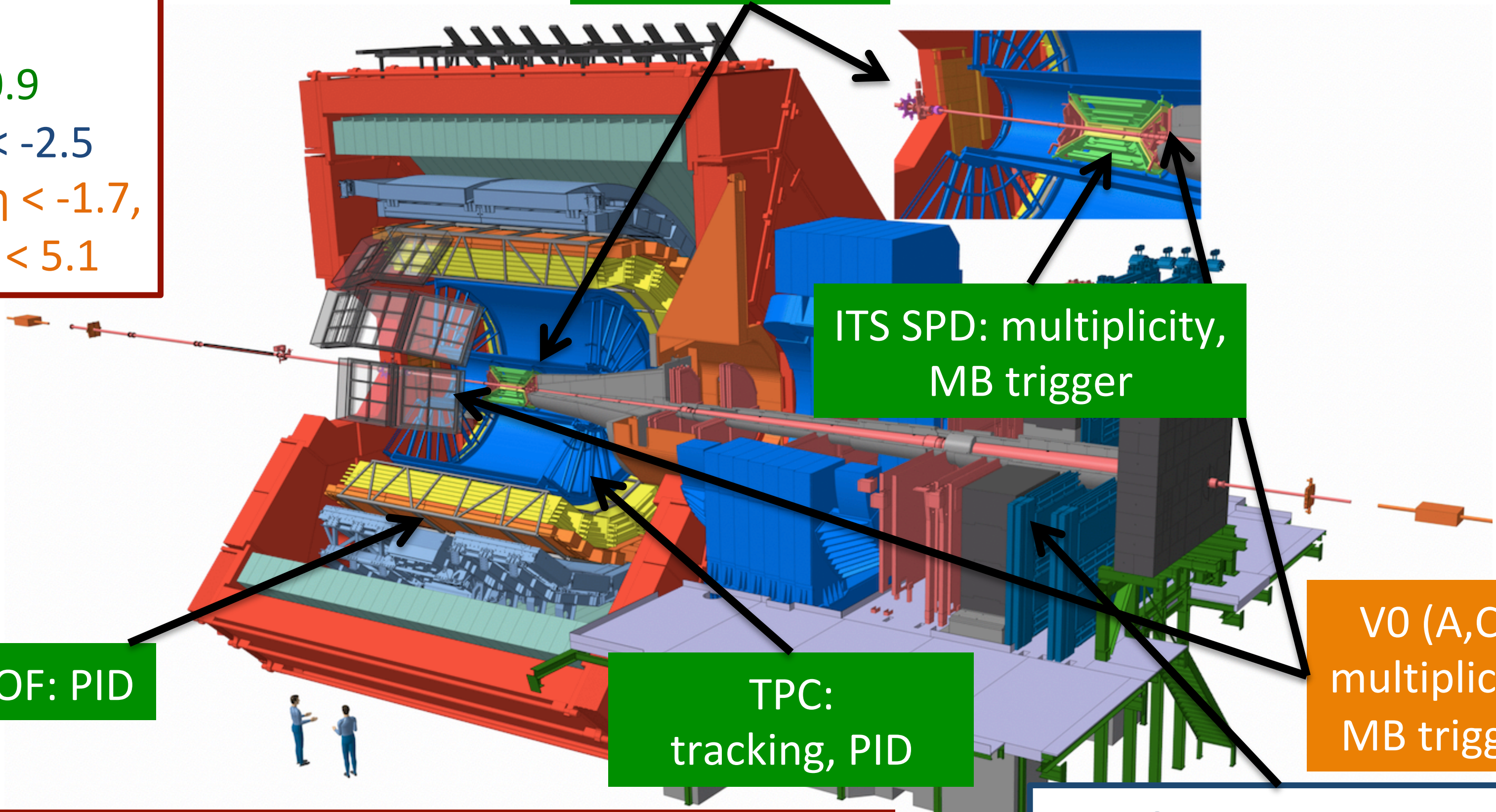
Both in pp and Pb-Pb, different hadronization mechanisms influence the baryon to meson ratios, with p-Pb collisions needed "in between"

➡ Importance of Λ_c^+ cross-section, Λ_c^+/D^0, R_{pPb} and R_{PbPb} measurements

The ALICE detector

Heavy-Ion experiment
Coverage:

- $|\eta| < 0.9$
- $-4 < \eta < -2.5$
- $-3.7 < \eta < -1.7$, $2.8 < \eta < 5.1$



ITS: vertexing, tracking, PID

ITS SPD: multiplicity, MB trigger

TOF: PID

TPC: tracking, PID

V0 (A,C): multiplicity, MB trigger

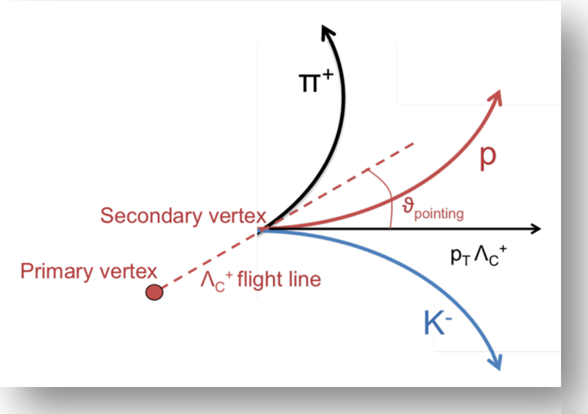
MUON spectrometer: μ reconstruction, trigger

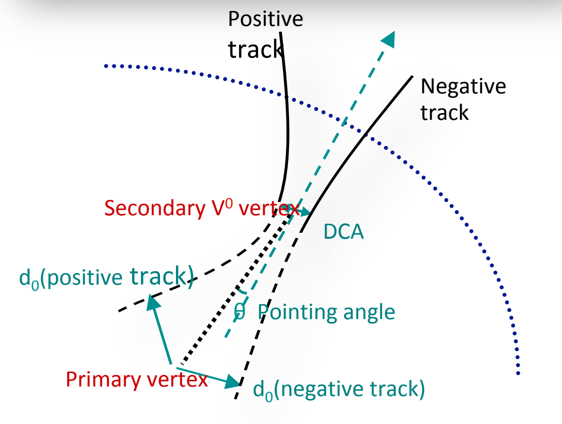
p-Pb Min Bias collisions (V0 trigger) at $\sqrt{s_{NN}} = 5.02$ TeV
~100 x 10^6 collisions, $\mathcal{L}_{int} = 48.6\ \mu b^{-1}$

Λ_c reconstruction in ALICE

Λ_c^+ (udc): mass = 2284 MeV/c^2
$c\tau \approx 60$ µm

Standard analysis based on rectangular cuts approach using topological and kinematical variables





$\Lambda_c^+ \rightarrow pK^-\pi^+$, BR = (6.35±0.33)%

- Candidates built from triplets of reconstructed tracks with the correct sign combination. Secondary vertex reconstructed from the triplet
- Reconstruction quality selections, topological and geometrical cuts, and Bayesian PID applied to the three daughter tracks/Λ_c candidate

$\Lambda_c^+ \rightarrow pK^0_s$, BR = (1.58±0.08)% [$K^0_s \rightarrow \pi^+\pi^-$ BR = (69.20±0.05)%]

- Candidates built from pairing a reconstructed track with a K^0_s candidate
- Reconstruction quality selections, topological and geometrical cuts, and n_σ PID applied to the daughter track/V^0/Λ_c candidate

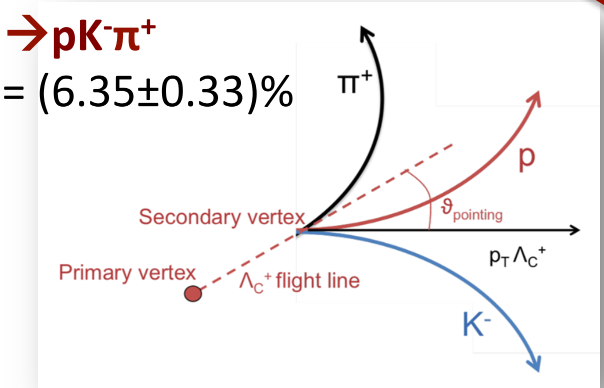
Λ_c reconstruction in ALICE

Λ_c^+ (udc): mass = 2284 MeV/c^2

$c\tau \approx 60$ µm

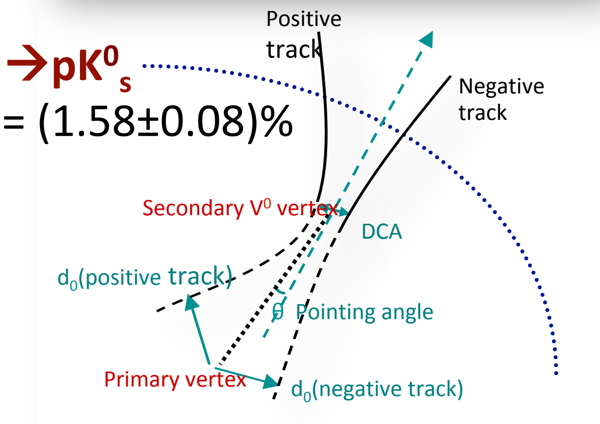
$\Lambda_c^+ \rightarrow pK^-\pi^+$

BR = (6.35±0.33)%

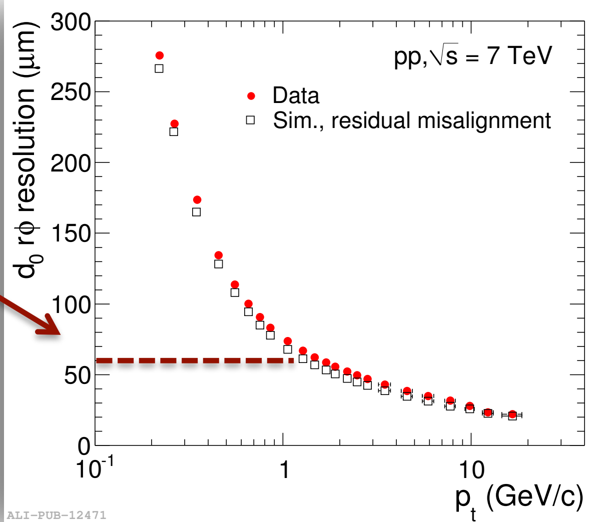


$\Lambda_c^+ \rightarrow pK^0_s$

BR = (1.58±0.08)%



Challenging reconstruction of the secondary vertex



JHEP 01 (2012) 128

In addition: very high background from uncorrelated tracks (even higher in Pb-Pb)

Require very precise tracking (high spatial and momentum track resolution) and excellent Particle Identification

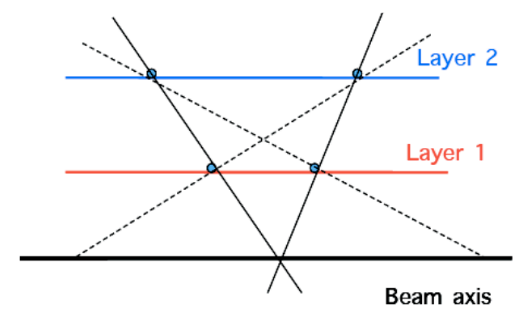
Vertexing and Tracking in ALICE

- The **Inner Tracking System** ($|\eta| < 0.9$)
  - Six silicon layers, three technologies: SPD, SDD, SSD
  - Primary and secondary vertex reconstruction
  - Tracking + standalone reconstruction
  - PID via d*E*/dx from SDD and SSD analog read-out
    - In the low p_T region, down to 100 MeV/*c* for π standalone tracking

Layer	Technology	R (cm)	±z (cm)	Spatial resolution (μm)	
				rϕ	z
1	SPD	3.9	14.1	12	100
2	SPD	7.6	14.1	12	100
3	SDD	15.0	22.2	35	25
4	SDD	23.9	29.7	35	25
5	SSD	38.0	43.1	20	830
6	SSD	43.6	48.9	20	830

Vertexing and Tracking in ALICE

- The **Inner Tracking System** (|η| < 0.9)
  - Six silicon layers, three technologies: SPD, SDD, SSD
  - Primary and secondary vertex reconstruction
  - Tracking + standalone reconstruction
  - PID via d*E*/dx from SDD
    - In the low p_T region ndalone tracking

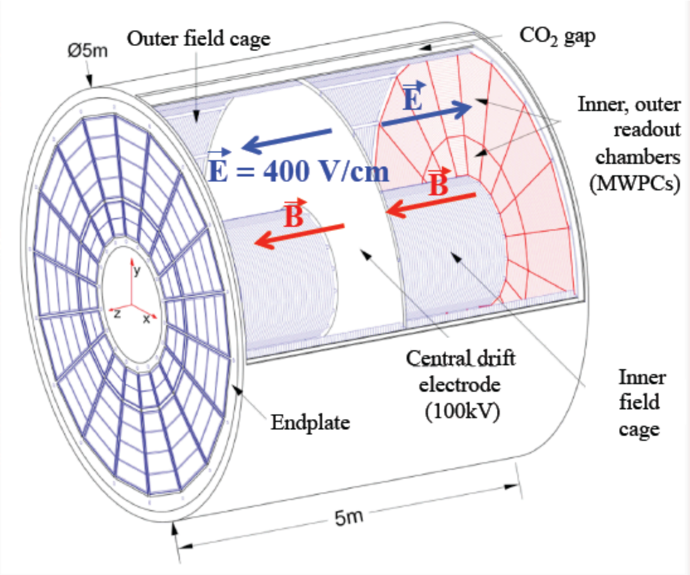


Layer	Technology			tial resolution (μm)	
				ϕ	z
1	SPD			12	100
2	SPD			12	100
3	SDD	15.0	22.2	35	25
4	SDD	23.9	29.7	35	25
5	SSD	38.0	43.1	20	830
6	SSD	43.6	48.9	20	830

SPD tracklets are used to build the first estimate of the primary vertex

Vertexing and Tracking in ALICE

- Large **Time Projection Chamber** optimized for high-multiplicity environment

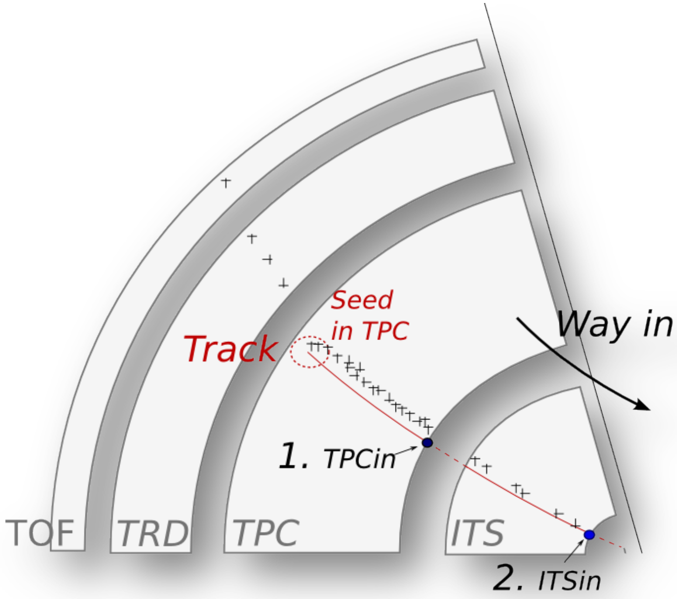


- *L*=5 m, Ø = 5 m, 92 m^3 (inner radius ~80 cm)
- Material (η=0): 3% X_0
- Drift time: 92 μs
- ~ 560000 pads

- Efficient tracking (~80%) in |η| < 0.8
- Momentum resolution (TPC+ITS) $\sigma(p_T)/p_T$ < 2% up to 10 GeV/*c*
- PID (truncated mean over a max. of 159 signals) with $\sigma_{dE/dx}$~5.5% and 7% in pp and Pb-Pb collisions, respectively
- Space-point resolution 0.8 (1.2) mm in xy (z)

Vertexing and Tracking in ALICE

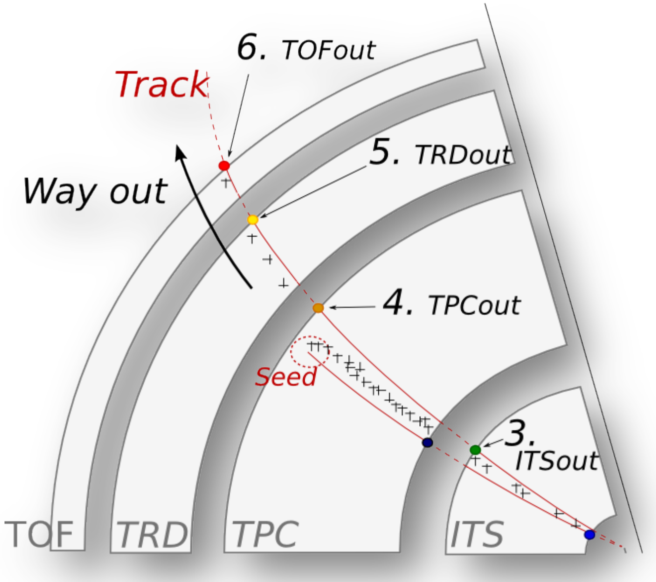
- Large **Time Projection Chamber** optimized for high-multiplicity environment



- Track seeds built using SPD primary vertex and pairs of TPC points in adjacent pads in the TPC outer region
- **Kalman filter** used to propagate the tracks **inwards**

Vertexing and Tracking in ALICE

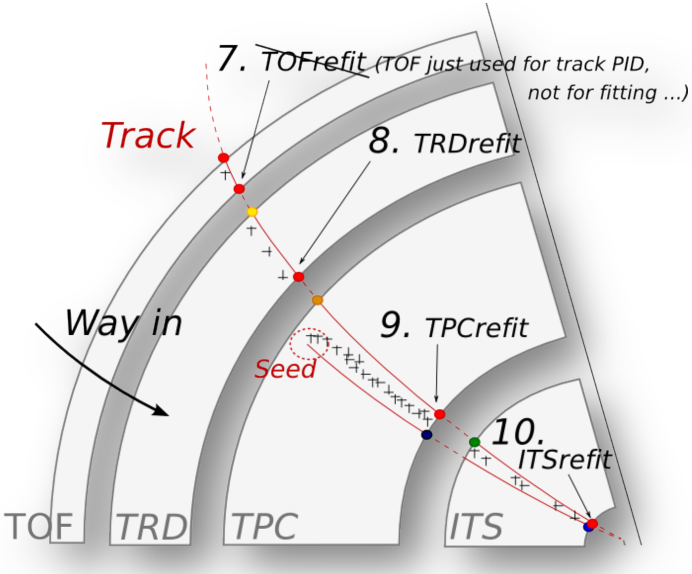
- Large **Time Projection Chamber** optimized for high-multiplicity environment



- Track seeds built using SPD primary vertex and pairs of TPC points in adjacent pads in the TPC outer region
- **Kalman filter** used to propagate the tracks inwards, **outwards**

Vertexing and Tracking in ALICE

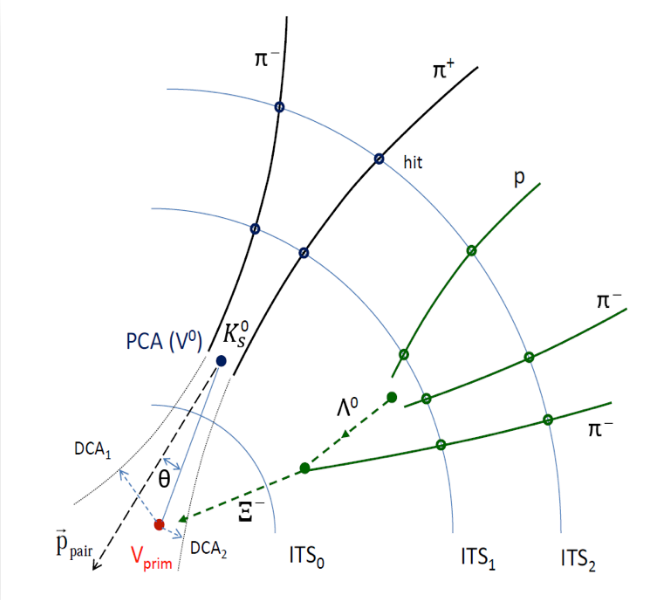
- Large **Time Projection Chamber** optimized for high-multiplicity environment



- Track seeds built using SPD primary vertex and pairs of TPC points in adjacent pads in the TPC outer region
- **Kalman filter** used to propagate the tracks inwards, outwards, and **to refit them finally inwards**

Vertexing and Tracking in ALICE

- Large **Time Projection Chamber** optimized for high-multiplicity environment



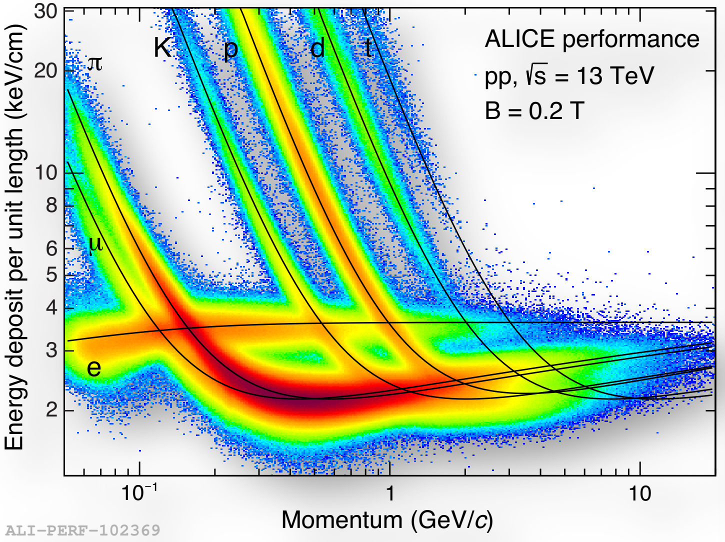
- Track seeds built using SPD primary vertex and pairs of TPC points in adjacent pads in the TPC outer region
- **Kalman filter** used to propagate the tracks inwards, outwards, and **to refit them finally inwards**

- Primary vertex determined from TPC+ITS tracks
- Secondary vertices from weak decays (e.g. $K^0_s \rightarrow \pi^+\pi^-$) built from tracks with large Distance of Closest Approach (DCA) to primary vertex

Particle Identification in ALICE

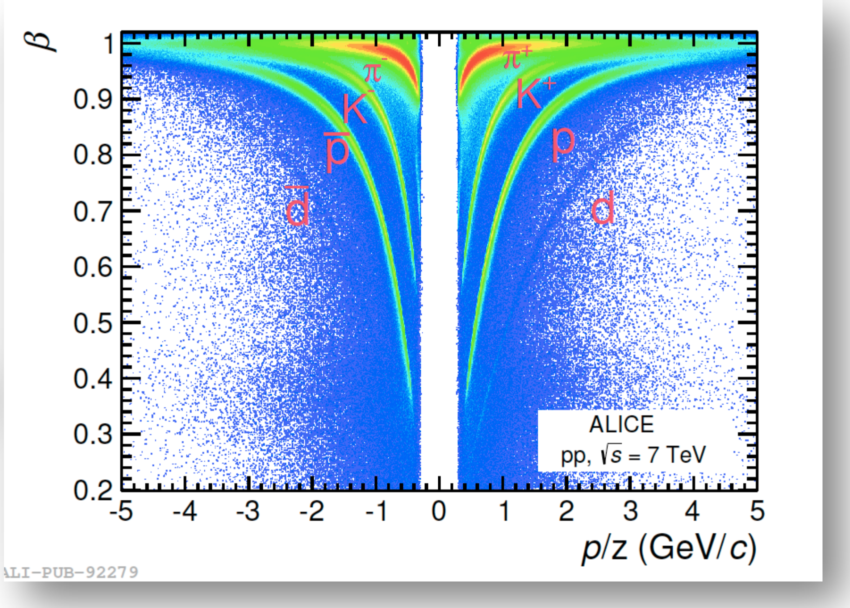
- Particle Identification of charged hadrons and leptons (e, μ) provided in ALICE by ITS, **TPC**, Time Of Flight (**TOF**), Transition Radiation (TRD), High-Momentum PID (HMPID), Muon Spectrometer

TPC dE/dx



Chin. Phys. C, 40, 100001 (2016) and 2017 update

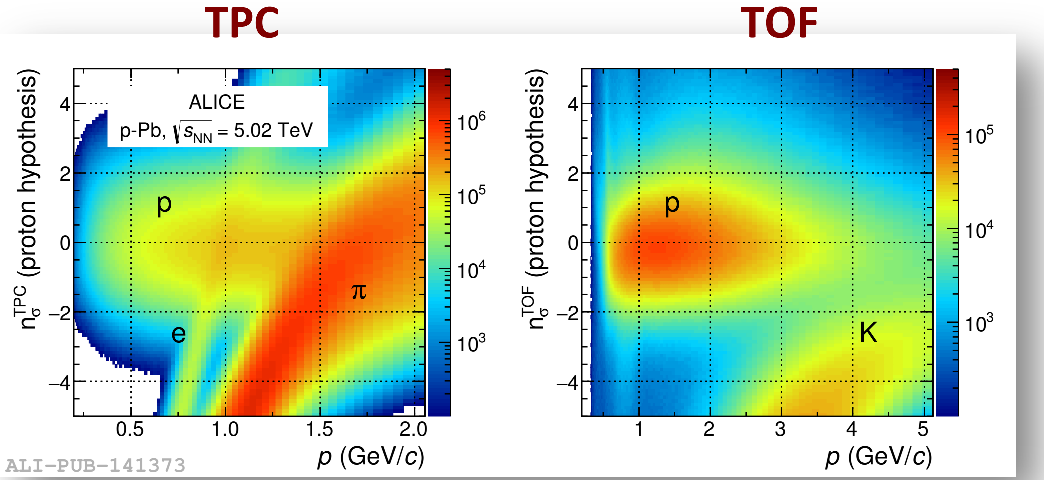
TOF time of flight



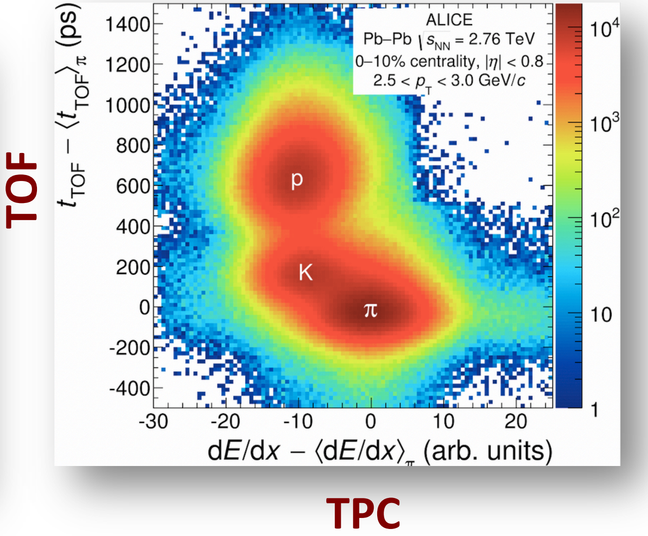
EPJC 75 (2015) 226

Particle Identification in ALICE

- Particle Identification of charged hadrons and leptons (e, μ) provided in ALICE by ITS, **TPC**, Time Of Flight (**TOF**), Transition Radiation (TRD), High-Momentum PID (HMPID), Muon Spectrometer



arXiv:1712.09581



Eur. Phys. J. Plus 131 (2016) 168

Here: PID applied as a cut on deviation in terms of number of σ (detector resolution) of the measured signal from the expected one at a given momentum and for a given particle (n_σ cut); or combining the signals from different detectors with a Bayesian approach (Bayesian PID)

Λ_c^+ analysis in ALICE

Decay channel	**Collision system**	**$\sqrt{s_{NN}}$ [TeV]**	**strategy**	
			Method	PID
$\Lambda_c^+ \rightarrow pK^-\pi^+$	pp	7	STD	Bayes
$\Lambda_c^+ \rightarrow pK^0_s$			STD	n_σ
$\Lambda_c^+ \rightarrow e^+\nu_e\Lambda$			Pair combination	n_σ
$\Lambda_c^+ \rightarrow pK^-\pi^+$	p-Pb	5.02	STD	Bayes
			MVA	n_σ, Bayes
$\Lambda_c^+ \rightarrow pK^0_s$			STD	n_σ
			MVA	N_σ, Bayes

- Analysis performed in p_T intervals
- Λ_c raw yield extracted fitting the invariant mass distributions of the candidates selected as described in the following slides
  - Gaussian (signal) + 1st or 2nd order polynomial (background)
- Corrections for acceptance, efficiency, B feed-down applied
- **Final result: Λ_c^+ cross-section (charge conjugate implicitly included)**

Boosted Decision Trees for Λ_c reconstruction

- Investigated at first as an alternative to the rather long manual optimization of cuts used to reduce the background
  - **TMVA package**, PoS ACAT (2007) 040
- Resulted as a powerful, automatic and easy-to-use tool with results comparable or slightly better than the standard approach

BDT variables

$\Lambda_c^+ \rightarrow pK^-\pi^+$

p_T of the daughter tracks
DCA of the daughter tracks
decay length
d_{12}, d_{23} (*)
Resolution on secondary vertex
$\cos\theta_{pointing}$
Normalized decay length
Bayes probability for daughter tracks

$\Lambda_c^+ \rightarrow pK^0_s$

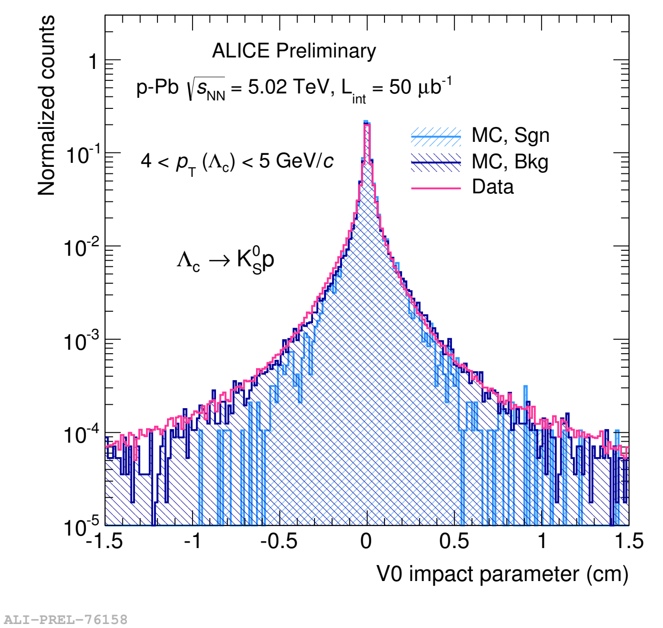
p_T of the bachelor track (p)
d_0 of the bachelor track
V^0 invariant mass
d_0 of the V^0
Lifetime f the V^0
Bayes probability for the bachelor track

d_{xy} = distance between a pair of unlike-sign tracks and the primary vertex

Boosted Decision Trees for Λ_c reconstruction

- Investigated at first as an alternative to the rather long manual optimization of cuts used to reduce the background
  - **TMVA package**, PoS ACAT (2007) 040
- Resulted as a powerful, automatic and easy-to-use tool with results comparable or slightly better than the standard approach

BDT variables



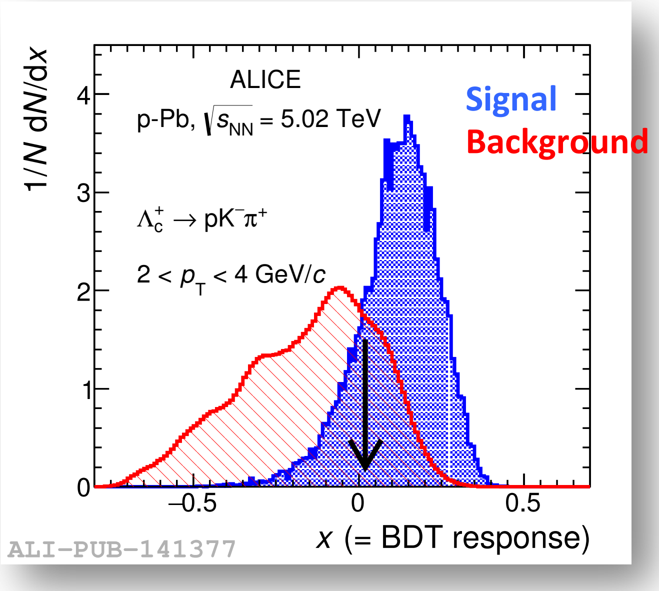
$\Lambda_c^+ \rightarrow pK^0_s$

p_T of the bachelor track (p)
d_0 of the bachelor track
V^0 invariant mass
d_0 of the V^0
Lifetime f the V^0
Bayes probability for the bachelor track

Signal trained on dedicated Monte Carlo simulations specific to the decay under study. Background built either from data or from MC

BDT details

- AdaBoost with β = 0.5
  - No significant difference in final results when using GradientBoost (systematic uncertainty)
- Gini index as separation type
  - No significant difference in final results when using Cross-Entropy (systematic uncertainty)
- Cut on the BDT was chosen to optimize the statistical significance S/√(S+B)



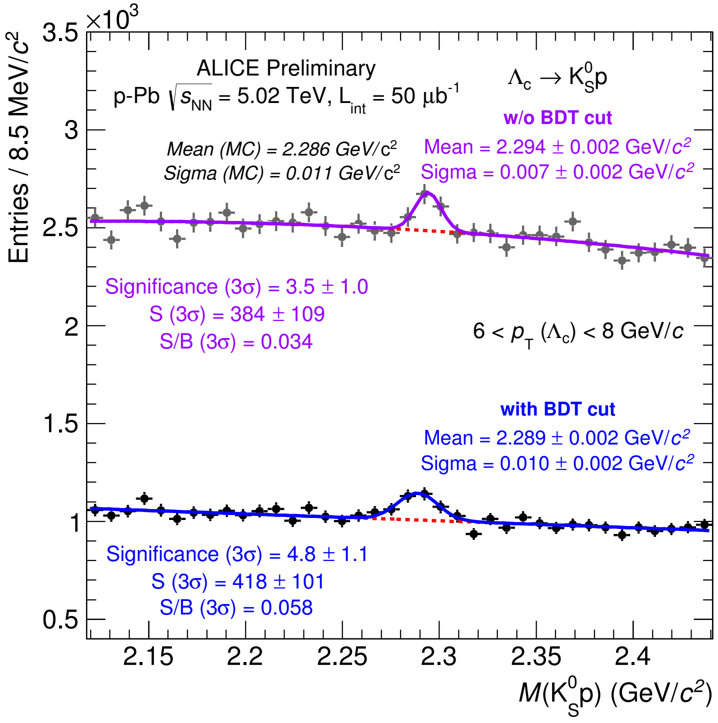
arXiv:1712.09581

Efficiency of BDT cut typically >=80%

- Systematic uncertainty from BDT:
  - Configuration (boost, separation, number of trees, input variables…): negligible
  - BDT cut:

$\Lambda_c^+ \rightarrow pK^-\pi^+$		**$\Lambda_c^+ \rightarrow pK^0_s$**	
Lowest p_T	Highest p_T	Lowest p_T	Highest p_T
8%	6%	5%	8%

Λ_c yield extraction with BDT cut



ALI-PREL-76142

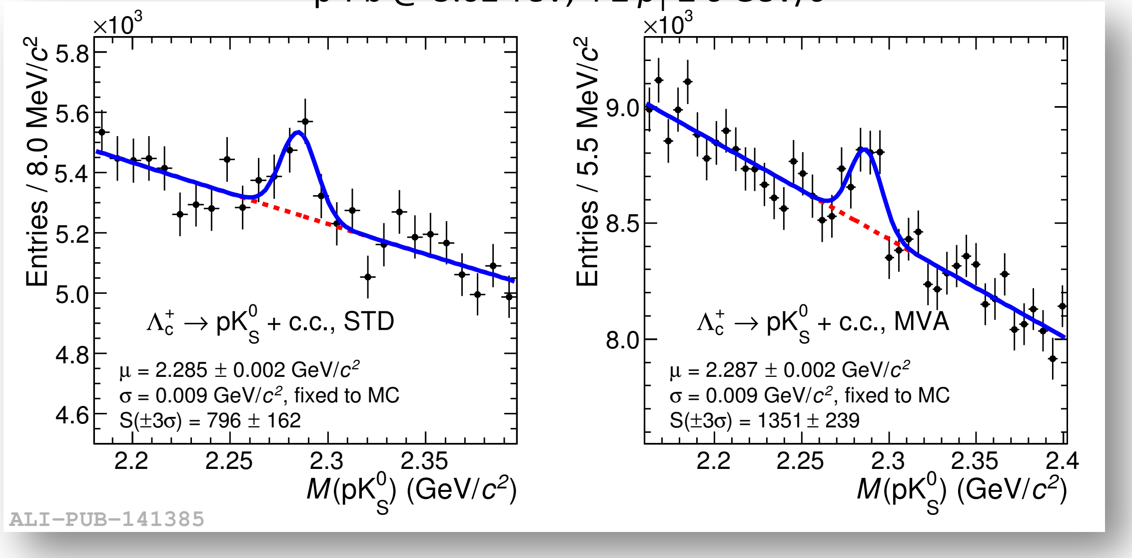
Including a cut on the BDT response:

- The signal over background ratio increases;
- The statistical significance of the signal increases;
- The Λ_c^+ peak parameters from the Gaussian fit (mean, sigma) are closer to the Monte Carlo values.

Λ_c yield extraction comparison to standard approach

p-Pb @ 5.02 TeV, 4 ≤ p_T ≤ 6 GeV/c

arXiv:1712.09581

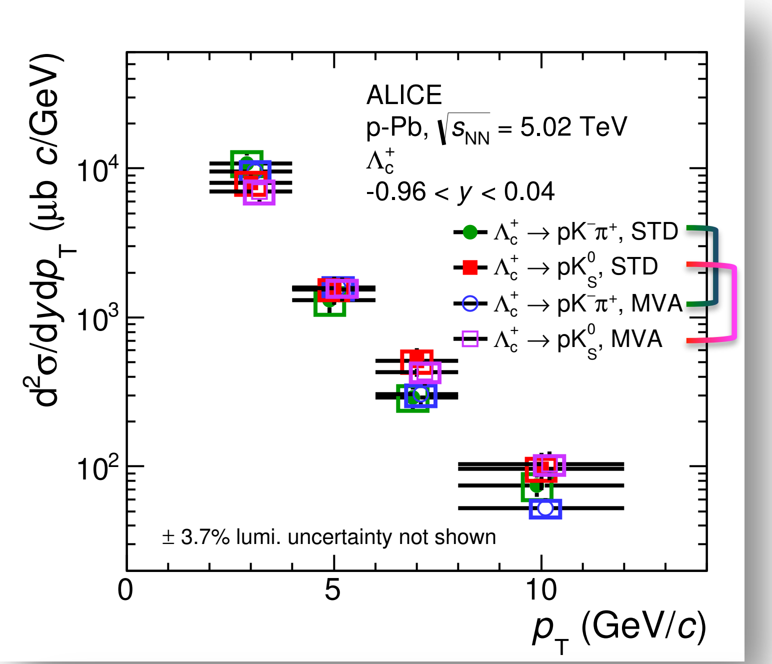


Slight improve in statistical precision and significance when using MVA approach

p_T (GeV/c)	Signal				Significance			
	$\Lambda_c^+ \to pK^-\pi^+$		$\Lambda_c^+ \to pK^0_s$		$\Lambda_c^+ \to pK^-\pi^+$		$\Lambda_c^+ \to pK^0_s$	
	STD	BDT	STD	BDT	STD	BDT	STD	BDT
2-4	5756±17%	7429±15%	2239±17%	2910±17%	6.1	5.3	4.9	4.9
4-6	1287±19%	1704±12%	796±20%	1350±18%	4.9	6.6	4.1	4.7
6-8	470±20%	886±18%	333±21%	346±18%	4.7	5.2	4.2	4.6
8-12	266±16%	203±16%	140±27%	255±21%	6.4	6.3	3.5	3.9

Results: cross section

- Cross-section for Λ_c^+ production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV obtained through an averaging procedure of the two decay channels with the two analyses approaches (standard and MVA-based)



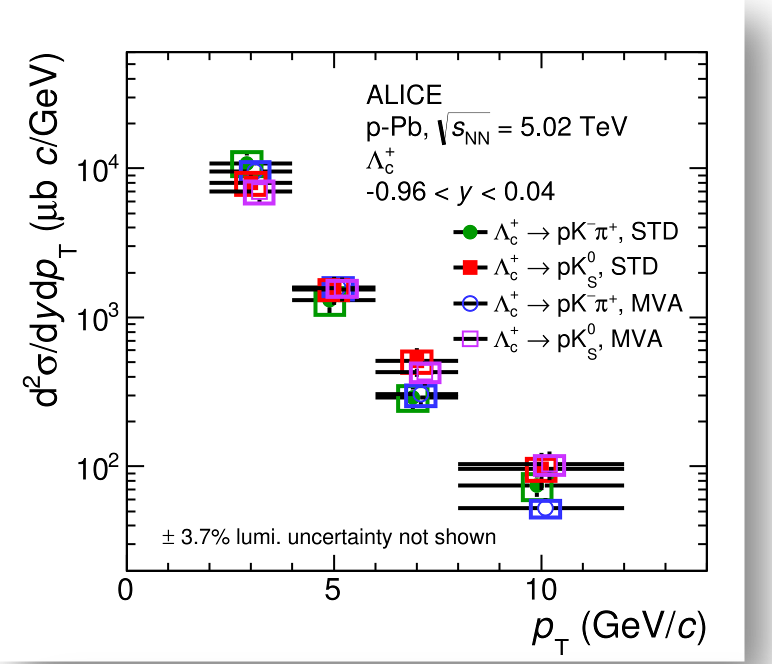
arXiv:1712.09581

Compatible results obtained with standard and MVA approach
→ BDT approach under control

$$\frac{d^2\sigma^{\Lambda_c^+}}{dp_T dy} = \frac{1}{2c_{\Delta y}\Delta p_T} \frac{1}{BR} \frac{f_{prompt} \cdot N^{\Lambda_c}_{|y|<y_{fid}}}{(A \times \varepsilon)_{prompt}} \frac{1}{\mathscr{L}_{int}}$$

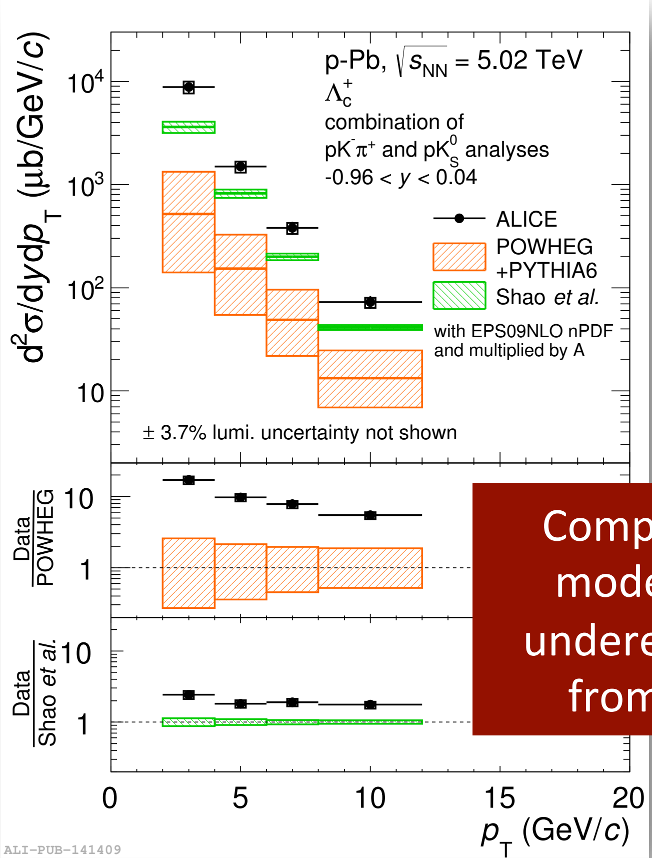
Results: cross section

- Cross-section for **Λ_c^+** production in p-Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV obtained through an averaging procedure of the two decay channels with the two analyses approaches (standard and MVA-based)



arXiv:1712.09581

$$\frac{d^2\sigma^{\Lambda_c^+}}{dp_T dy} = \frac{1}{2c_{\Delta y}\Delta p_T}\frac{1}{BR}\frac{f_{prompt}\cdot N^{\Lambda_c}_{|y|<y_{fid}}}{(A\times\varepsilon)_{prompt}}\frac{1}{\mathscr{L}_{int}}$$



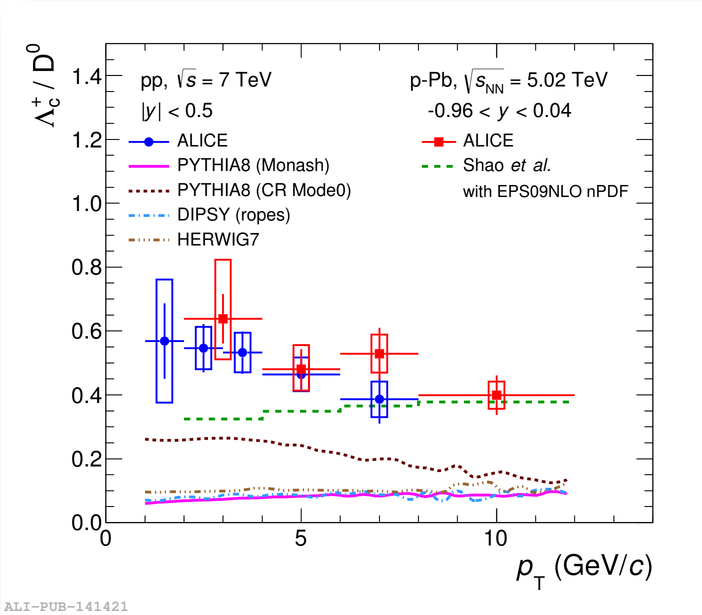
Comparison to models: clear underestimation from theory

Results: Λ_c^+/D^0

- Λ_c^+/D^0 ratio – sensitive to the hadronization mechanism

Comput. Phys. Commun. 178 (2008) 852–867,
JHEP 08 (2015) 003,
Phys. Rev. D92 no. 9, (2015) 094010,
Eur. Phys. J. C58 (2008) 639–707,
Eur. Phys. J. C77 no. 1, (2017) 1

arXiv:1712.09581



Very similar behaviour in pp and p-Pb collisions

Comparison to previous measurement (in different collision systems, kinematic regimes), and to LHCb also done

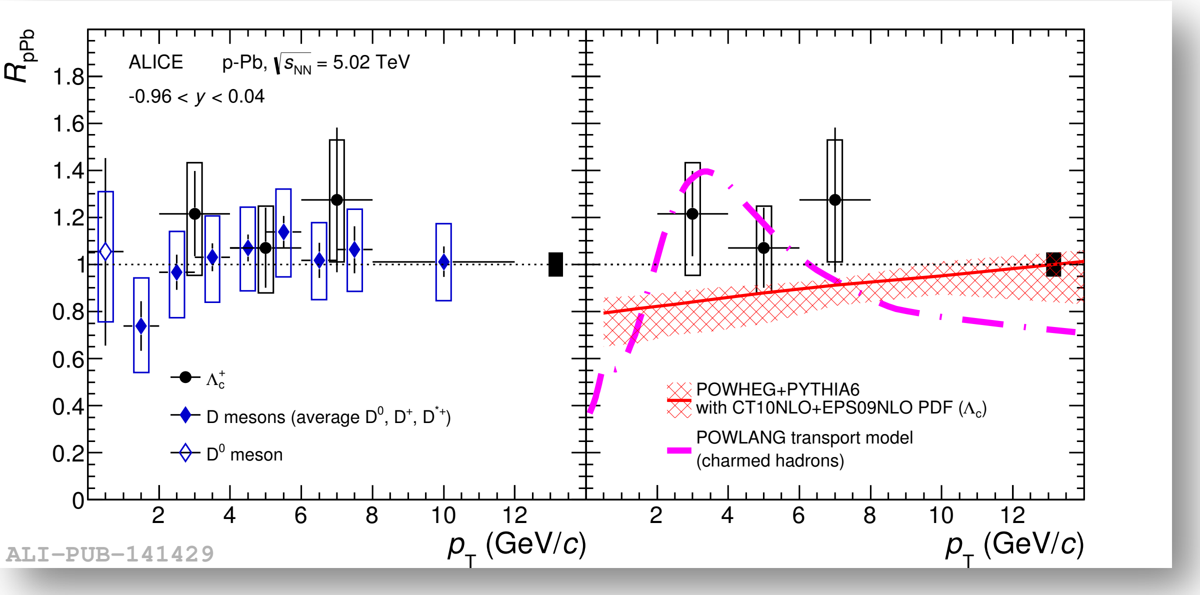
- Predictions underestimate measurements both in pp and p-Pb collisions
  - In pp: predictions based on fragmentation parameters from e^+e^-; PYTHIA8 with enhanced colour reconnection is the closest to data
  - In p-Pb: calculations by Shao et al. based on parameterization of LHCb pp data using EPS09NLO nuclear modification factor close to data

Further results

- Nuclear modification factor R_{pPb}

$$R_{\mathrm{pPb}} = \frac{1}{A} \frac{\mathrm{d}\sigma^{5\mathrm{TeV}}_{\mathrm{pPb}}/\mathrm{d}p_{\mathrm{T}}}{f^{\sqrt{s},y}_{\mathrm{FONLL}}(p_{\mathrm{T}}) \cdot \mathrm{d}\sigma^{7\mathrm{TeV}}_{\mathrm{pp}}/\mathrm{d}p_{\mathrm{T}}} \qquad \left(f^{\sqrt{s},y}_{\mathrm{FONLL}}(p_{\mathrm{T}}) = \frac{\mathrm{d}\sigma^{5\mathrm{TeV}}_{\mathrm{FONLL}}/\mathrm{d}p_{\mathrm{T}}}{\mathrm{d}\sigma^{7\mathrm{TeV}}_{\mathrm{FONLL}}/\mathrm{d}p_{\mathrm{T}}} \right)$$

  - Compatible with unity → no relevant difference w.r.t. pp collisions



arXiv:1712.09581

Summary and Conclusions

- **Multivariate Analysis** techniques have been used in a challenging ALICE analysis with success
  - Easy to use, they provided a useful alternative analysis to the standard approach based on manually optimized rectangular cuts
  - The **performance** comparison to the standard approach showed a slightly better statistical precision and significance
- The same approach is being used for **Run2** data, where a bigger sample should allow for more precise and differential results
- Run2 **Pb-Pb** collisions are another challenge on the table, with their high background underneath the Λ_c signal
- In **Run3/4**, the ALICE upgraded ITS will offer an even better detector performance (material budget, impact parameter resolution). The much larger Λ_c sample in HI collisions (continuous readout) will allow to investigate in detail the charm hadronization mechanisms in the QGP
  - **MVA techniques might be crucial to resolve efficiently signal from background down to $p_T = 0$**
  - **And MVA also for: b-jet tagging, Λ_b reconstruction, etc..**

"You don't know much," said the Duchess, "And that's a fact."
Alice's Adventures in Wonderland, L. Carroll

Being able to go from idea to result with the least possible delay is key to doing good research.
S. Wunsch, yesterday

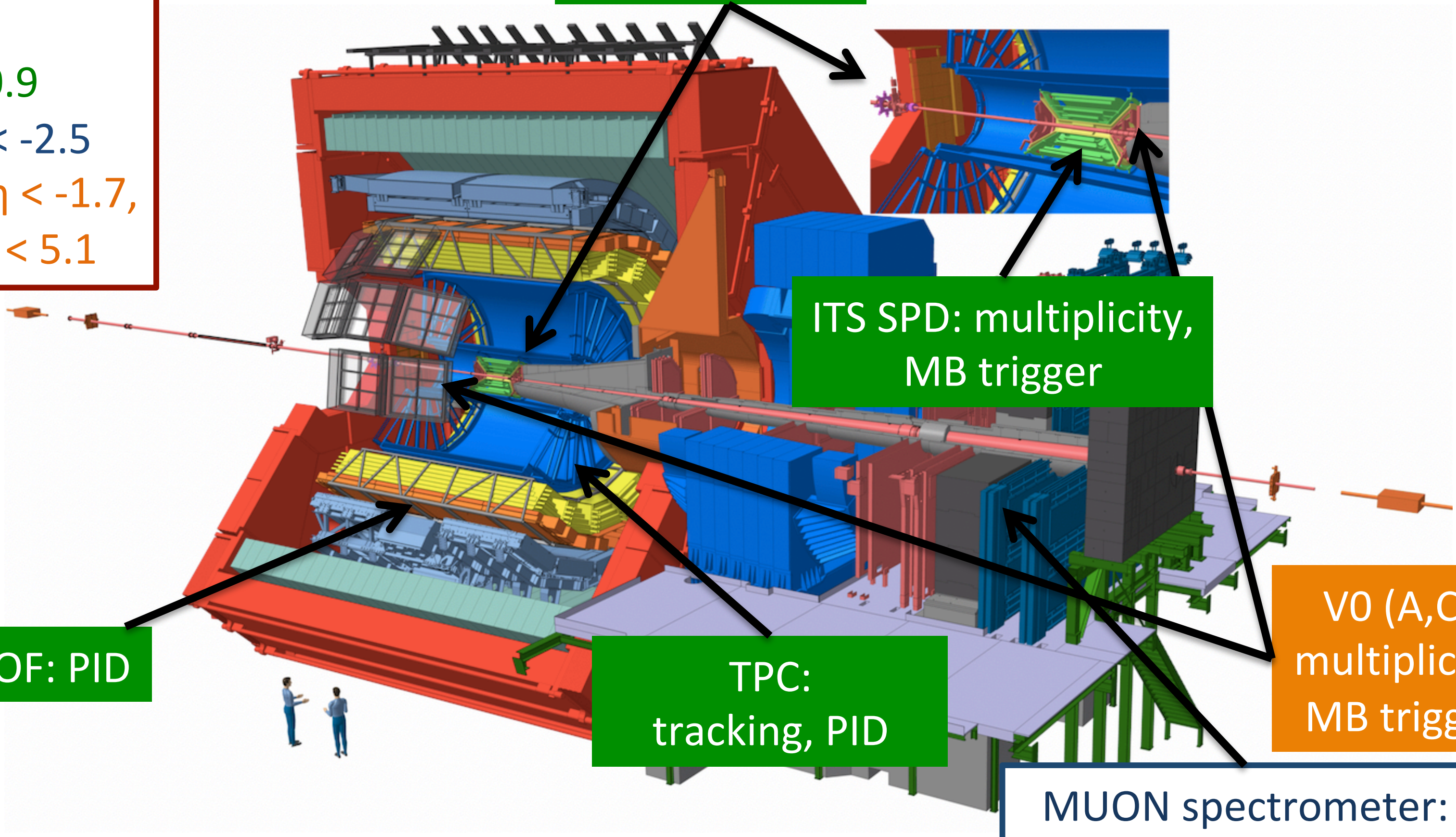
BACKUP

The ALICE detector

Heavy-Ion experiment

Coverage:

- $|\eta| < 0.9$
- $-4 < \eta < -2.5$
- $-3.7 < \eta < -1.7$, $2.8 < \eta < 5.1$



ITS: vertexing, tracking, PID

ITS SPD: multiplicity, MB trigger

TOF: PID

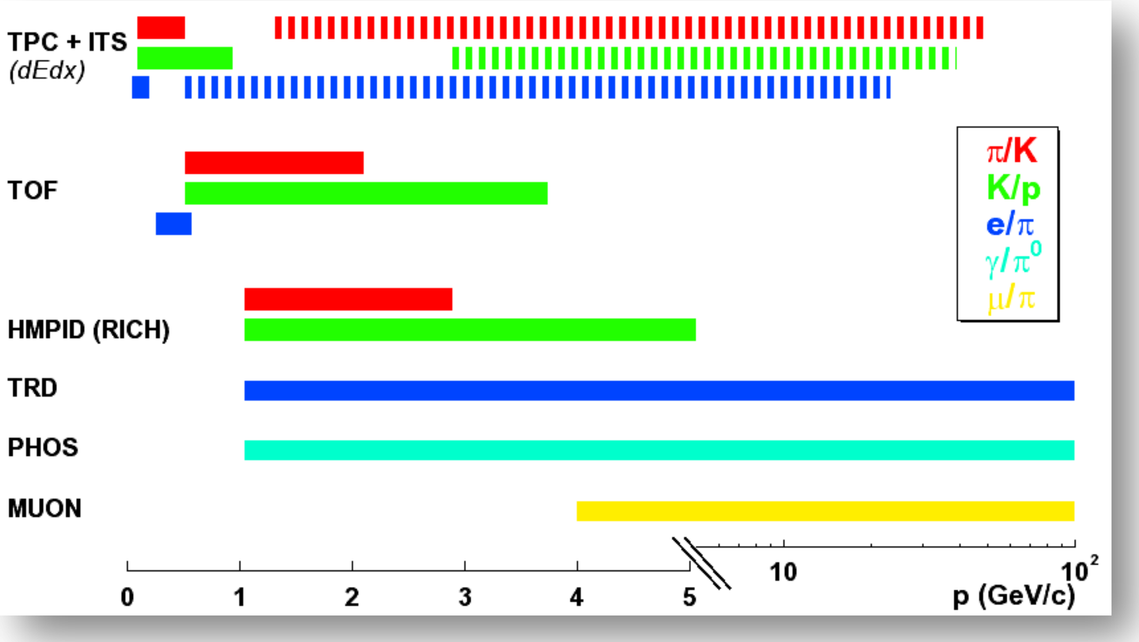
TPC: tracking, PID

V0 (A,C): multiplicity, MB trigger

MUON spectrometer: μ reconstruction, trigger

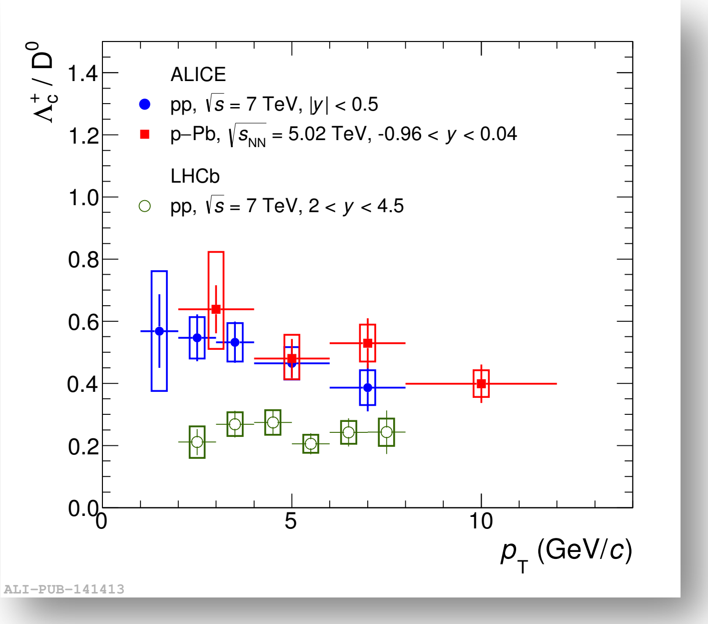
Particle Identification in ALICE

- Particle Identification of charged hadrons and leptons (e, μ) provided in ALICE by ITS, TPC, Time Of Flight (TOF), Transition Radiation (TRD), High-Momentum PID (HMPID), Muon Spectrometer

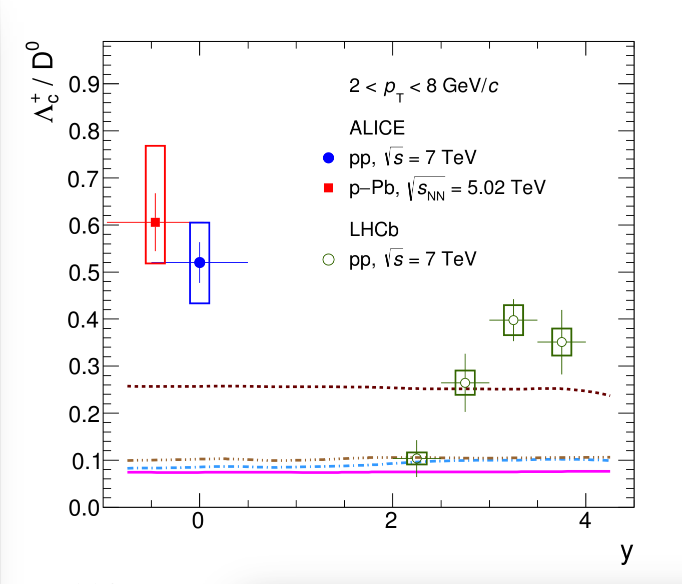


Results: Λ_c^+/D^0

- Comparison to LHCb



ALI-PUB-141413



ALI-PUB-141425

LHCb-CONF-2017-005