New approach for searching for CP violation at the LHCb experiment

Author: Jakub Ryżka
Supervisors: dr hab. inż. Tomasz Szumlak, prof. AGH (AGH UST)
dr Artur Ukleja (NCBJ NCNR)
04.12.2020 Kraków
Białasówka

Outline

- LHCb experiment & collaboration
- CP violation
- Motivation & recent CP discovery in charmed particles
- S_{CP} binned method
- Kernel Density Estimation (KDE)
- Results for simulated data
- Data
- Summary

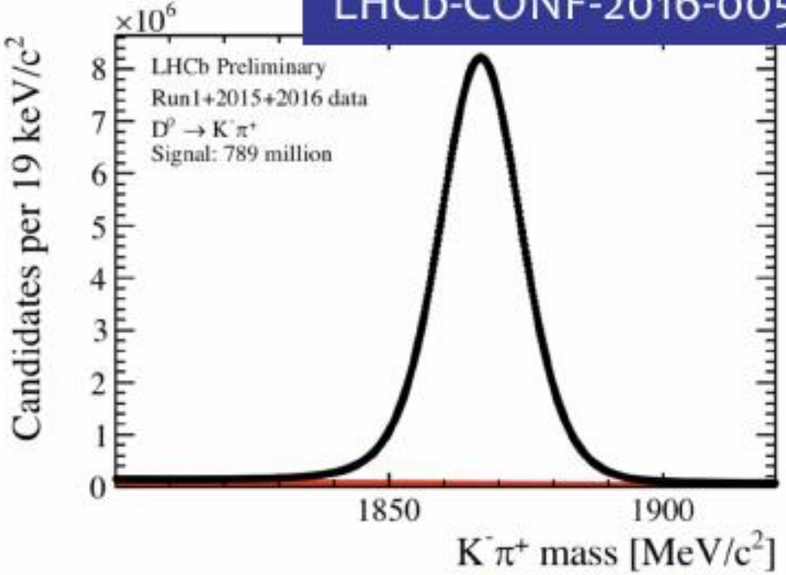
LHC: a charm factory

- At the LHC, the production cross-section of charm is ~ 20 times larger than the beauty one

$$\sigma(pp \to c\bar{c}X) = 1419 \pm 134\ \mu b\ @\ \sqrt{s} = 7\ TeV^{*}$$
$$\sigma(pp \to c\bar{c}X) = 2840 \pm 226\ \mu b\ @\ \sqrt{s} = 13\ TeV^{**}$$

* Nucl. Phys. B871 (2013) 1-20
** J. High Energ. Phys. (2017) 74

LHCb-CONF-2016-005



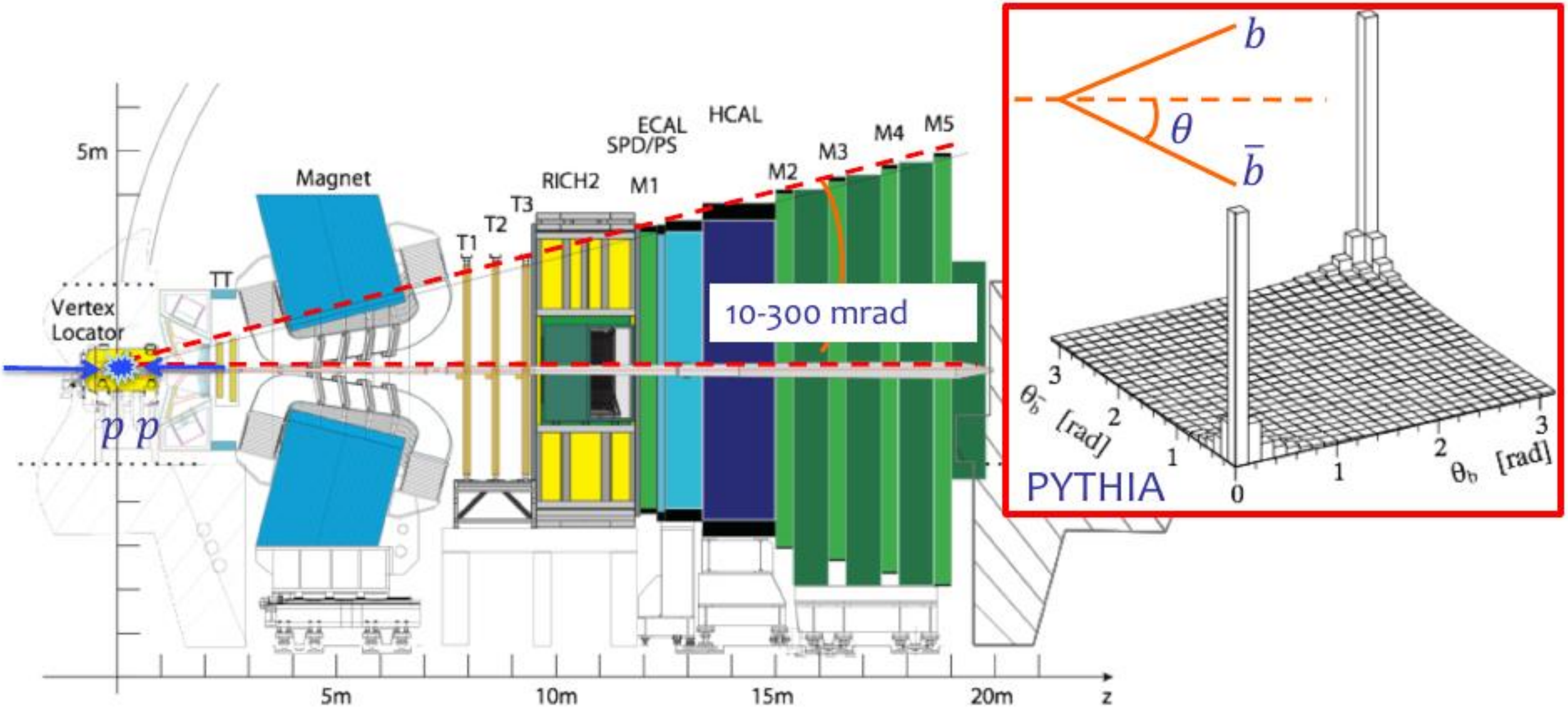
More than 1 billion of $D^0 \to K^-\pi^+$ events reconstructed with the full LHCb data sample

The LHCb detector

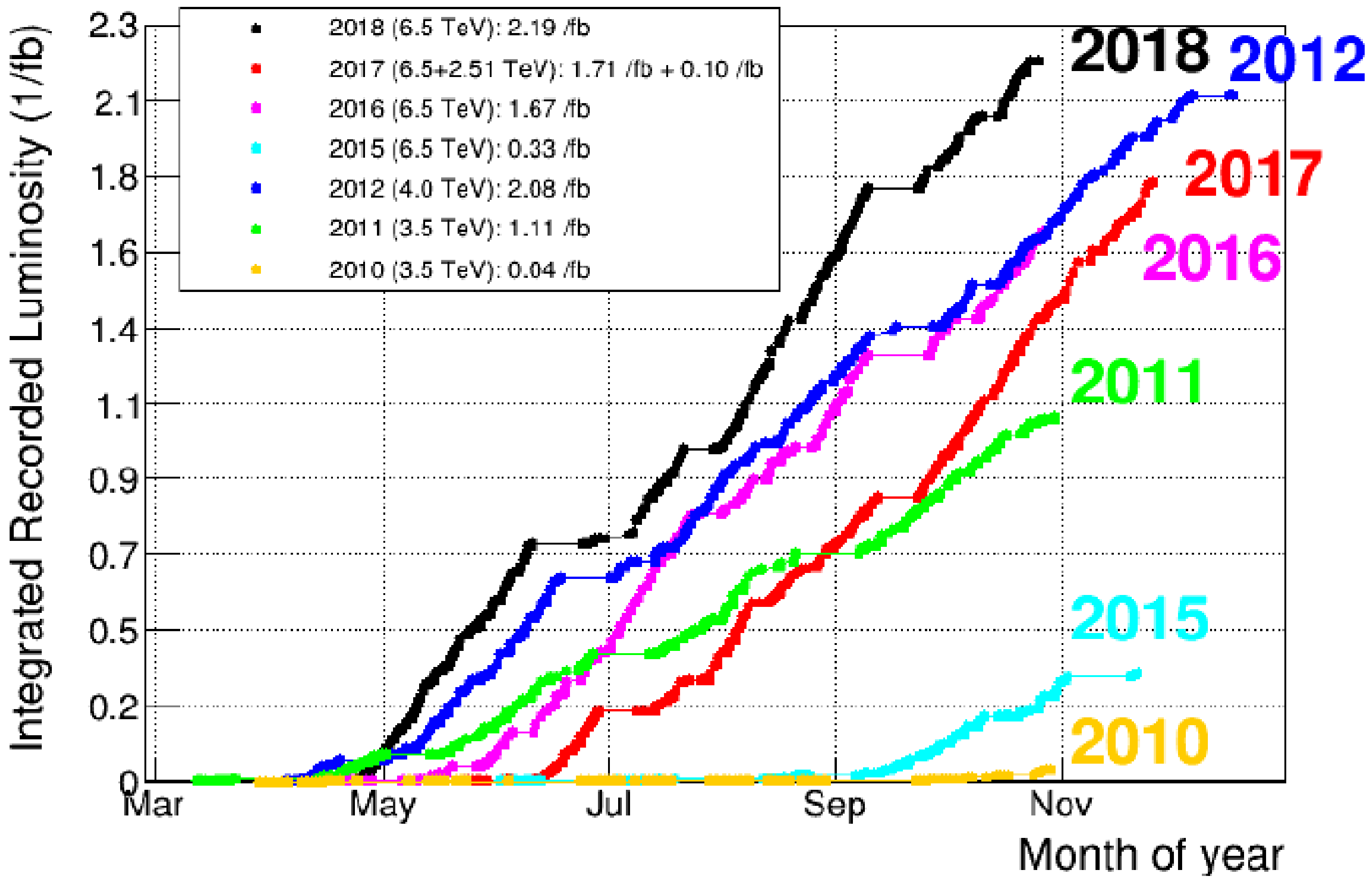
Int. J. Mod. Phys. A30 (2015) 07

FORWARD-PEAKED PRODUCTION

- LHCb designed as forward spectrometer (operating in collider mode) covering the pseudorapidity range $2 < \eta < 5$



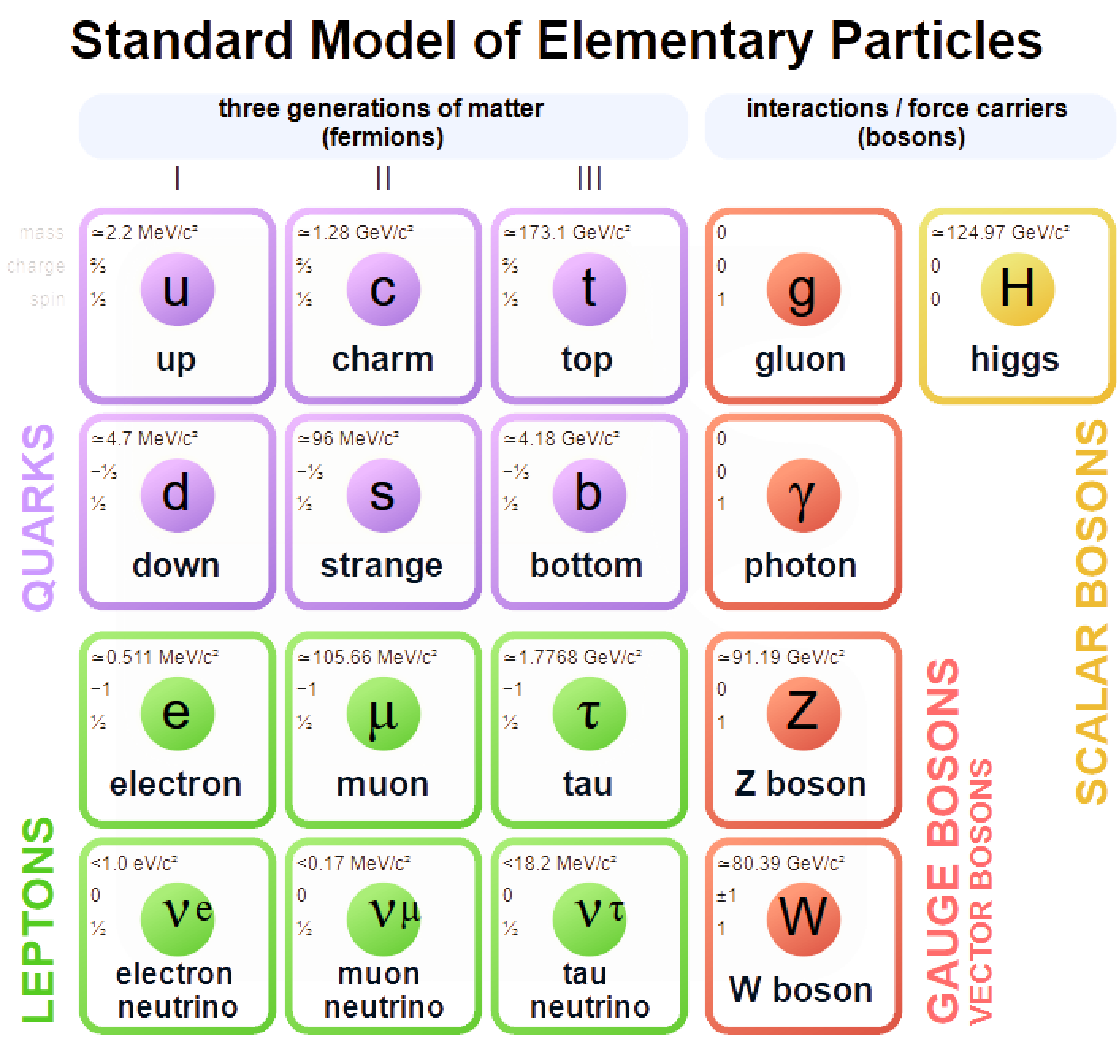
Integrated recorded luminosity



THANKS LHC!!

The full LHCb data set is about 9 fb^{-1}

What's the matter with the matter?

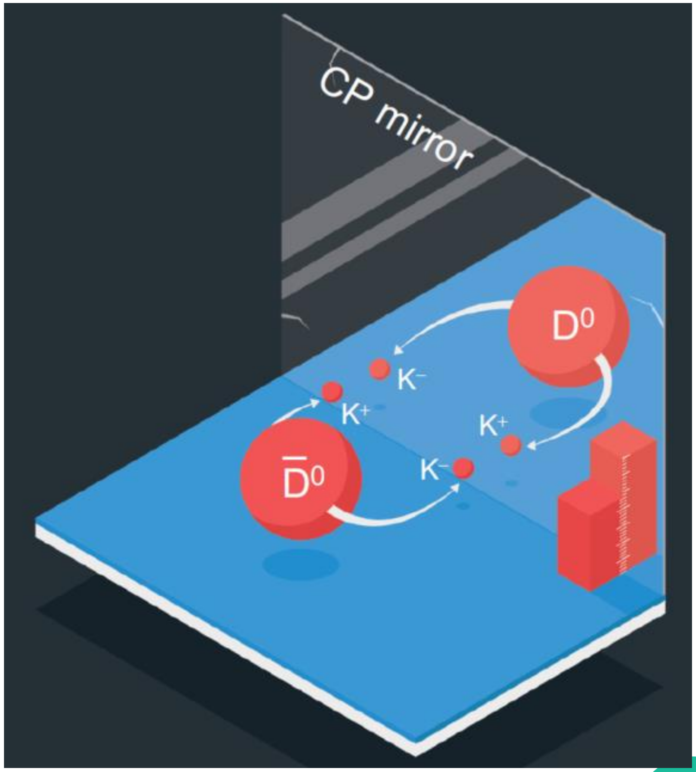


CP-violation

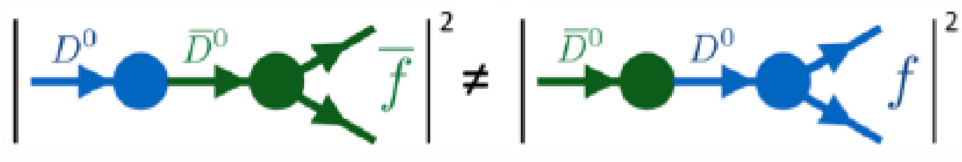
- **It is violation of CP-symmetry**
  - C – charge conjugation symmetry
  - P – parity symmetry
- **Three types of CP-violation**
  a) Mixing
  b) Direct
  c) Interference (mixing + direct)

$$A_{symCP} \sim |A_1||A_2|\sin(\varphi_1 - \varphi_2)\sin(\delta_1 - \delta_2)$$

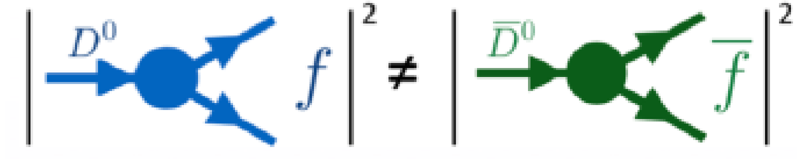
- **CP-symmetry states that laws of physics should be the same if a particle is interchanged with its antiparticle and under spacial inversion**



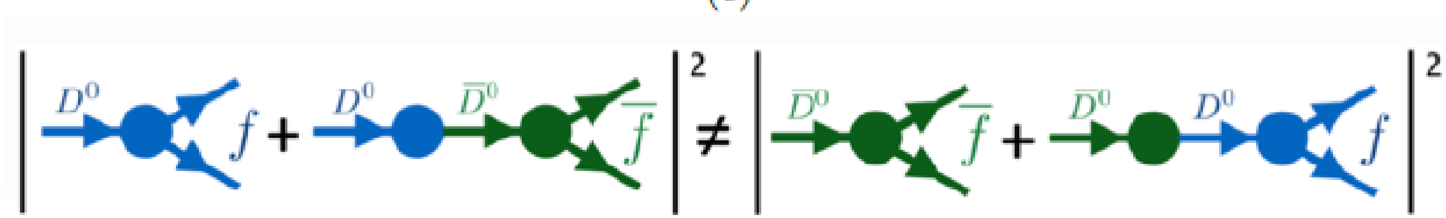
Three types of CP-violation



(b)



(c)



Cabibbo-Kobayashi-Maskawa Matrix (CKM)

V_{CKM}:

$$\begin{array}{c} \\ u \\ c \\ t \end{array} \begin{array}{c} \begin{array}{ccc} d & s & b \end{array} \\ \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \end{array}$$

- Wolfenstein parametrisation: $V_{CKM} =$

$$\begin{pmatrix} 1 - {}^{1}/_{2}\,\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - {}^{1}/_{2}\,\lambda^2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}$$

➔ phase that breaks CP symmetry

V_{CKM}:

$$\begin{array}{c} \\ u \\ c \end{array} \begin{array}{c} \begin{array}{cc} d & s \end{array} \\ \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} \end{array}$$

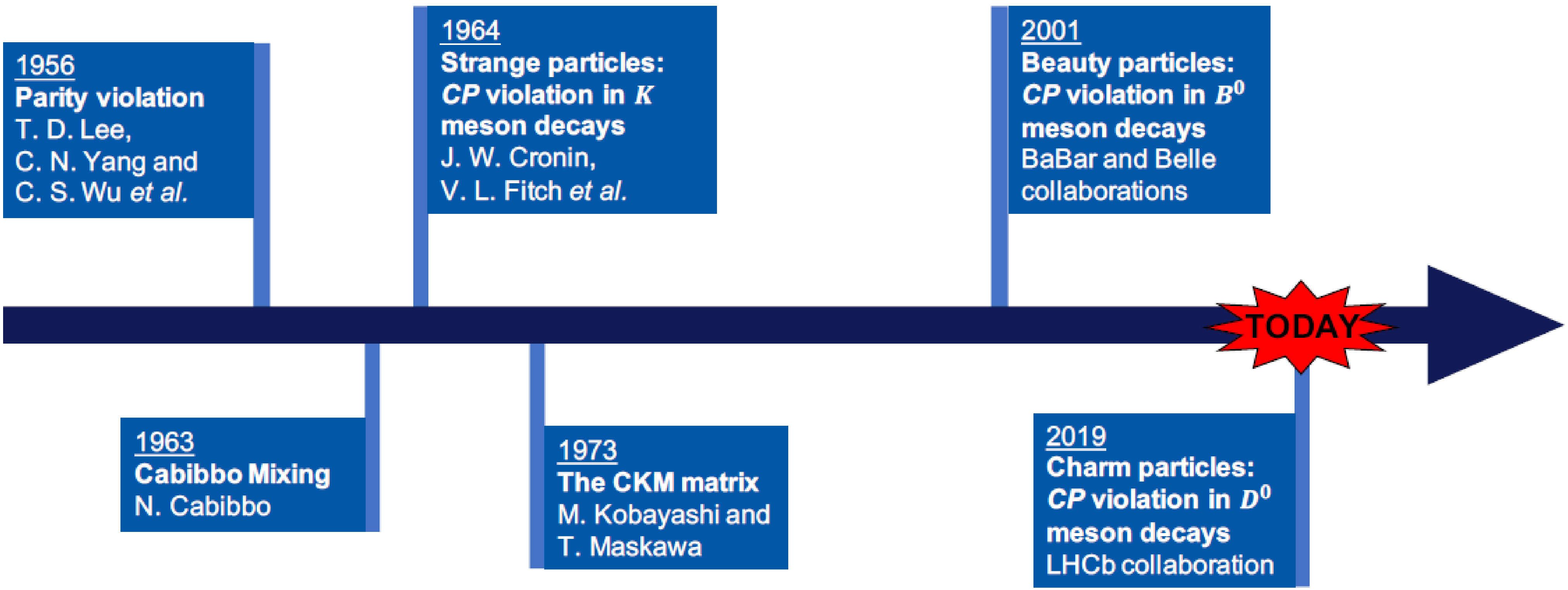
$V_{CKM} =$

$$\begin{pmatrix} 1 - {}^{1}/_{2}\,\lambda^2 & \lambda \\ -\lambda & 1 - {}^{1}/_{2}\,\lambda^2 \end{pmatrix}$$

➔ CP is conserved

- *At least 3 generations are needed to make it possible for the CP symmetry to be broken*

CP-violation key dates

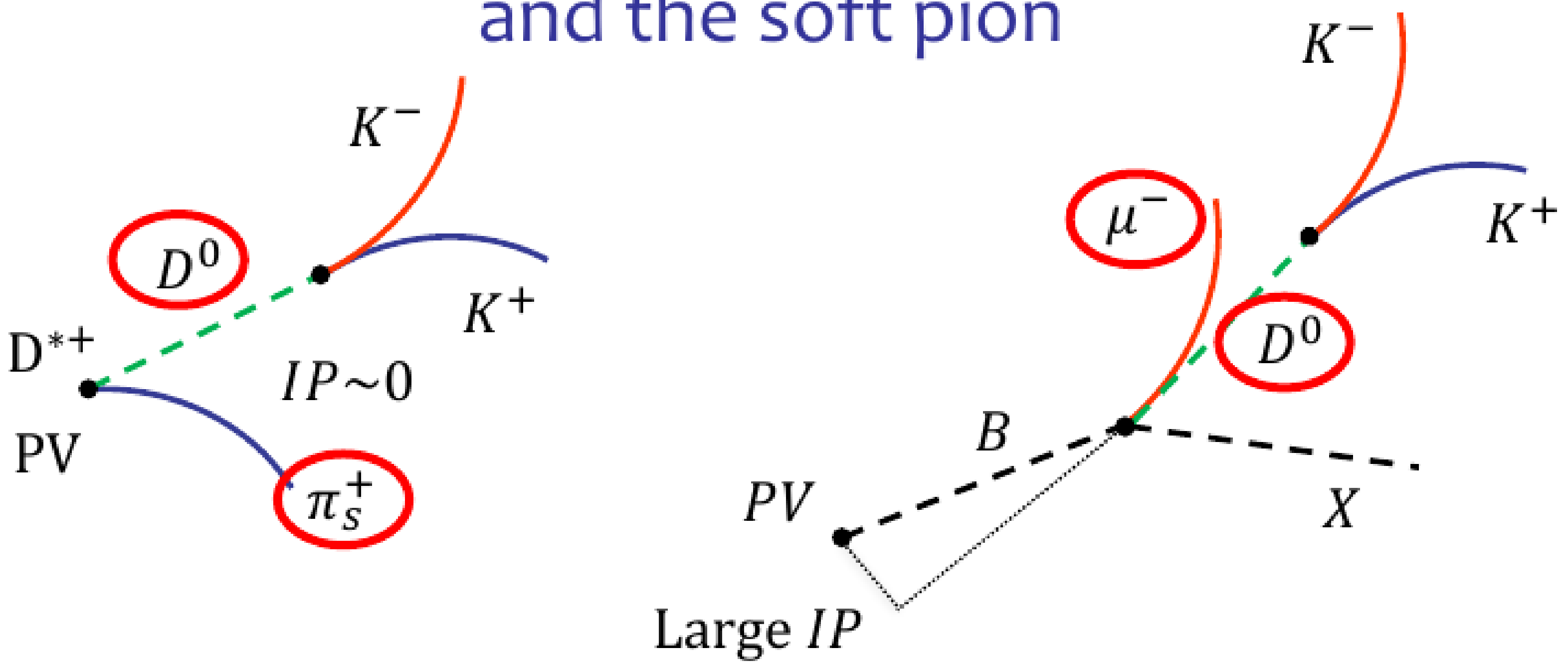


Motivation

- **First observation of CP-violation in charmed particles decays in 2018 (but not in charmed baryons!)**
- **Huge sample of charmed hadrons from Run I and Run II**
- **New model-independent approach**

D^0 production exploits in the analysis presented today

Experimentally we can tag D^0 flavour at production by means of the charge of the muon and the soft pion



Time-integrated CP asymmetry

CP asymmetry is defined as

$$A_{CP}(f) = \frac{\Gamma(D^0 \to f) - \Gamma(\overline{D}^0 \to f)}{\Gamma(D^0 \to f) + \Gamma(\overline{D}^0 \to f)}$$

with $f = K^-K^+$ and $f = \pi^-\pi^+$

The flavour of the initial state (D^0 or $\overline{D}^0$) is tagged by the charge of the slow pion from $D^{*\pm} \to D^0\pi^+$ or muon from $B \to D^0(\to f)\mu^-X$

The raw asymmetry for tagged D^0 decays to a final state f is given by

$$A_{\mathrm{raw}}(f) = \frac{N(D^0 \to f) - N(\overline{D}^0 \to f)}{N(D^0 \to f) + N(\overline{D}^0 \to f)}$$

where N refers to the number of reconstructed events of decay after background subtraction

ΔA_{CP} π-tagged

What we measure is the physical asymmetry plus asymmetries due both to production and detector effects

$$A_{\text{raw}}(f) = A_{CP}(f) + \cancel{A_D(f)} + A_D(\pi_s^+) + A_P(D^{*+})$$

CP asymmetry

Any charge-dependent asymmetry in slow pion reconstruction

$D^{*\pm}$ production asymmetry

- No detection asymmetry for D^0 decays to K^-K^+ or $\pi^-\pi^+$
- … if we take the raw asymmetry difference

$$\Delta A_{CP} \equiv A_{raw}(KK) - A_{raw}(\pi\pi) = A_{CP}(KK) - A_{CP}(\pi\pi)$$

- the D^{*+} production and the slow pion detection asymmetries will cancel

Results with full LHCb data sample [9 fb^{-1}]

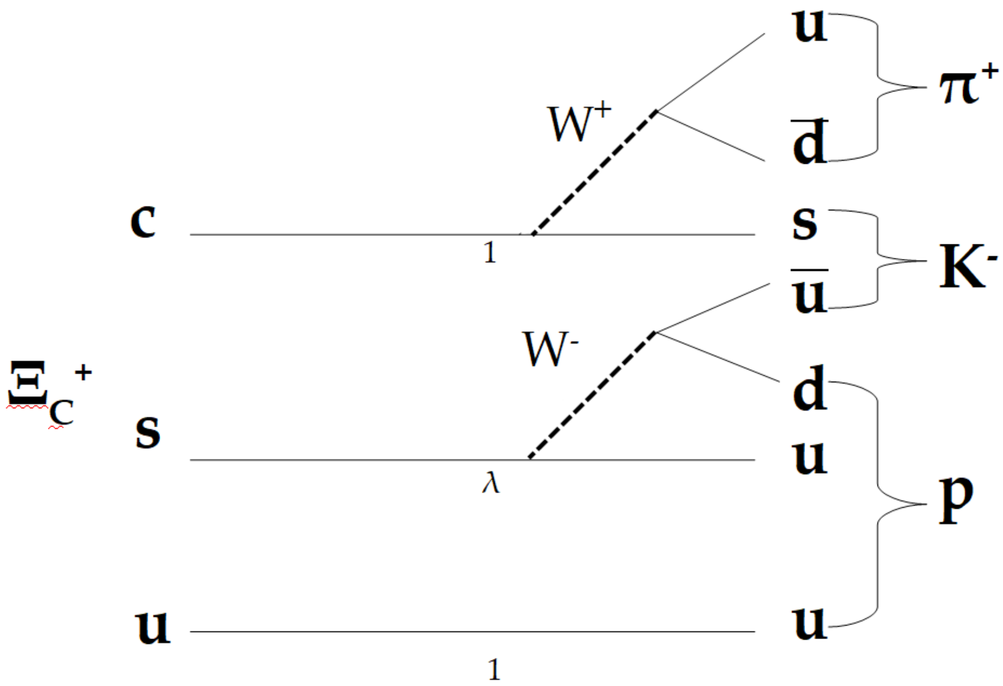
LHCb-PAPER-2019-006

$$\Delta A_{CP} = (-15.4 \pm 2.9)\times 10^{-4}$$

5.3 standard deviations from zero

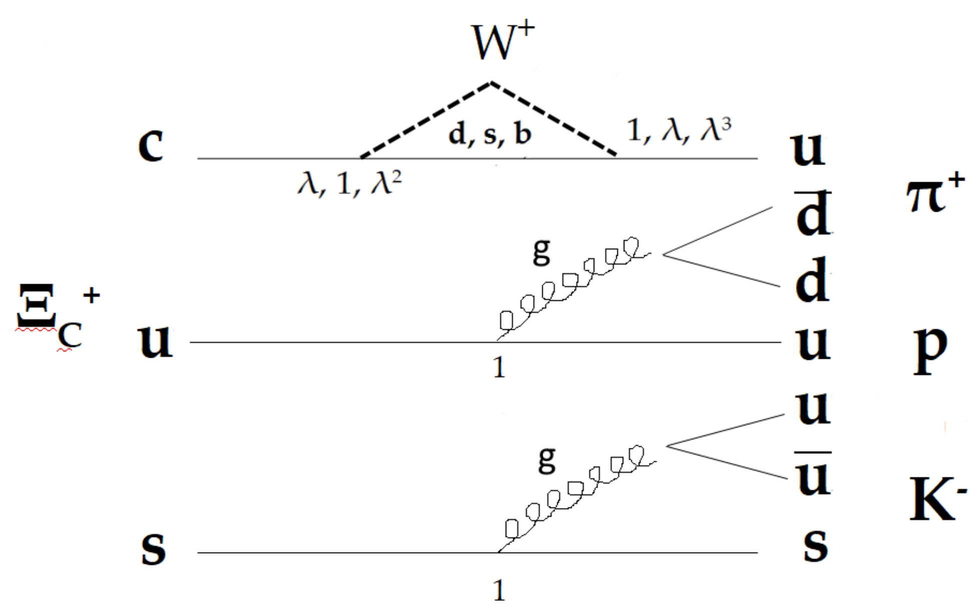
This is the first **observation** of *CP* violation in the decay of **charm** hadrons

Signal decays – charmed baryon Ξ_C

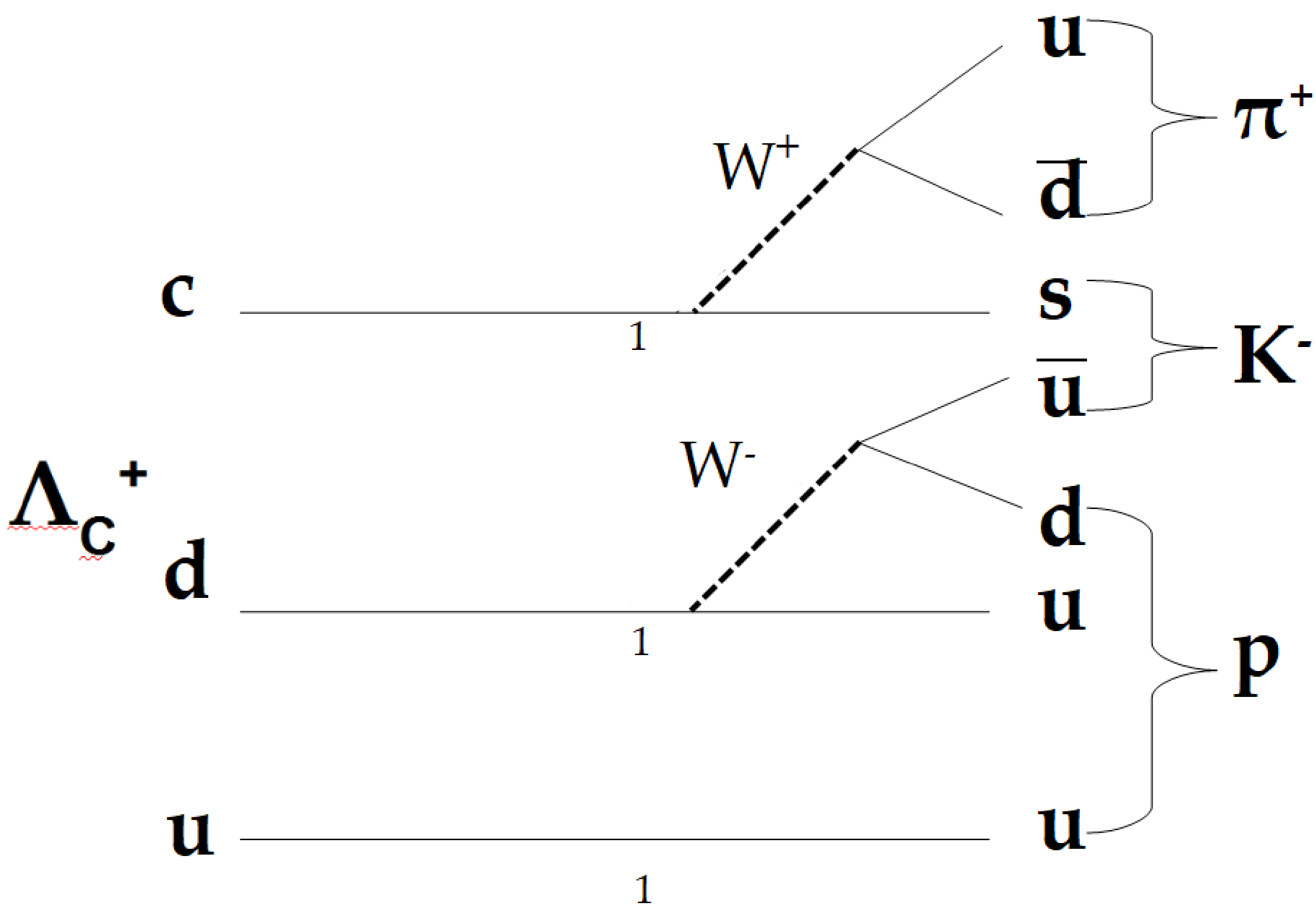


$$A_T \sim \lambda T$$

$$P \sim \lambda P_d + \lambda P_s + \lambda^5 P_b$$



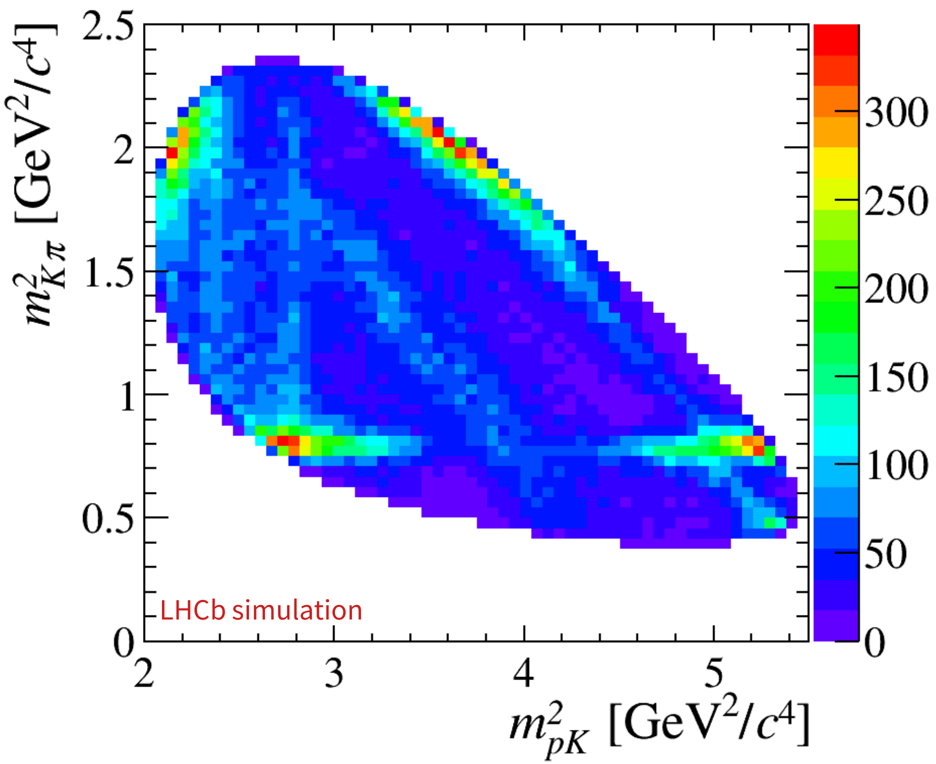
Control decay – charmed baryon Λ_C



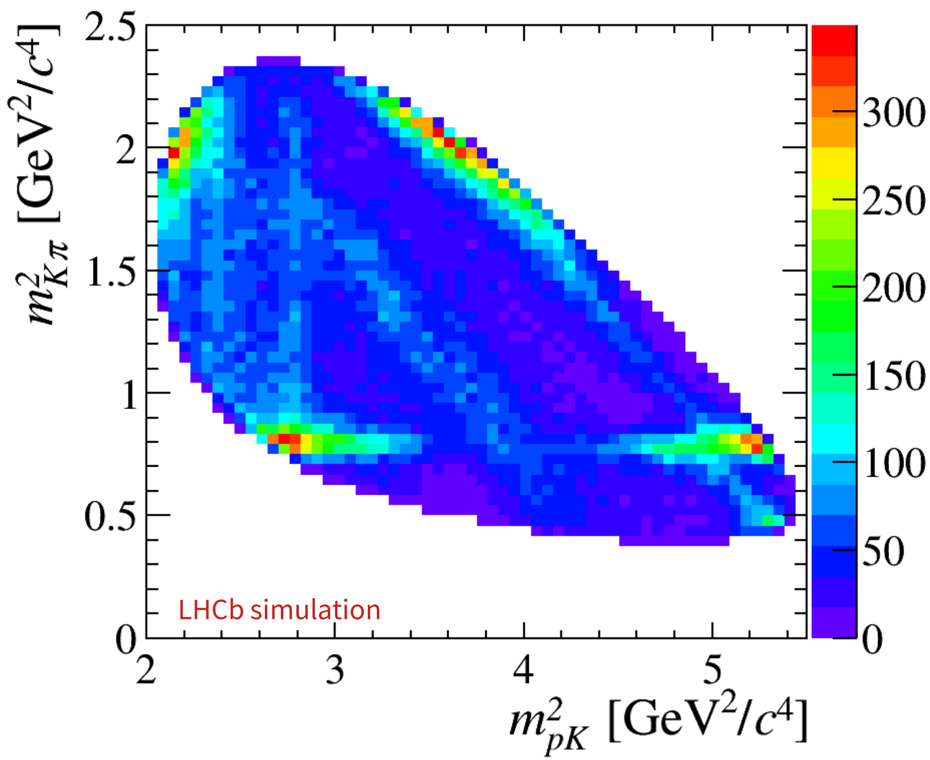
$A_T \sim 1T$

Input sample

Two simulated Monte Carlo data sets, containing 200 000 events from $\Xi_c \rightarrow pK\pi$ decays have been considered. The first one was generated according to a CP-conserving model and the other with a moderate amount of CP-violation (20%) in the vector resonance K^*.



Dalitz plot for sample with no CP-violation



Dalitz plot for sample with 20% CP-violation

Towards new approach

- To test the methods and study their sensitivity and statistical rejection power when estimating the CPV a dedicated Monte-Carlo toy experiments are generated
- The core idea is based on the "isobar ansatz" where the Dalitz phase-space for a selected 3-body decay (may be extended to n-body) is described by a coherent sum of N isobar (2-body systems, resonances) amplitudes with a small admixture of continuum (non-resonant) event:

$$\mathcal{M}(\vec{x}) = a_{\mathrm{nr}} e^{i\phi_{\mathrm{nr}}} + \sum_r a_r a^{i\phi_r} \mathcal{A}_r(\vec{x}).$$

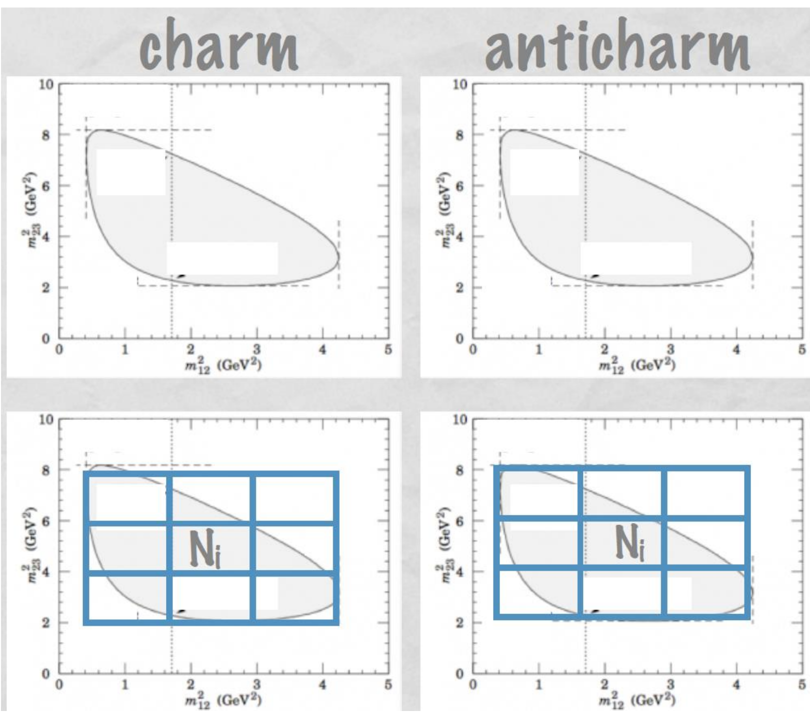
- Usually, the number of resonances that enter the fit (model) is motivated by the concrete experiment (observed decay modes) The asymmetry is introduced in an "effective" way by reducing the fit fraction of a dominating resonance
- The final "observed" probability distribution defined over the Dalitz phase space is calculated using the formula:

$$f(\vec{x}) = |\mathcal{M}(\vec{x})|^2 / \int |\mathcal{M}(\vec{x})|^2 d\vec{x}$$

Jakub Ryżka

S_{CP} binned method

The binned method is based on dividing the phase space into n bins. For each bin, comparison between Dalitz plots for particles and antiparticles is performed.



Artur Ukleja, EB Meeting 2018

Significance of the difference between number of particles (N^+) and antiparticles (N^-) is computed, using the following expression:

$$S^i_{CP} = \frac{N_i^+ - \alpha N_i^-}{\sqrt{\alpha(N_i^+ + N_i^-)}}$$

$\alpha = N^+/N^-$ Accounts for global asymmetries

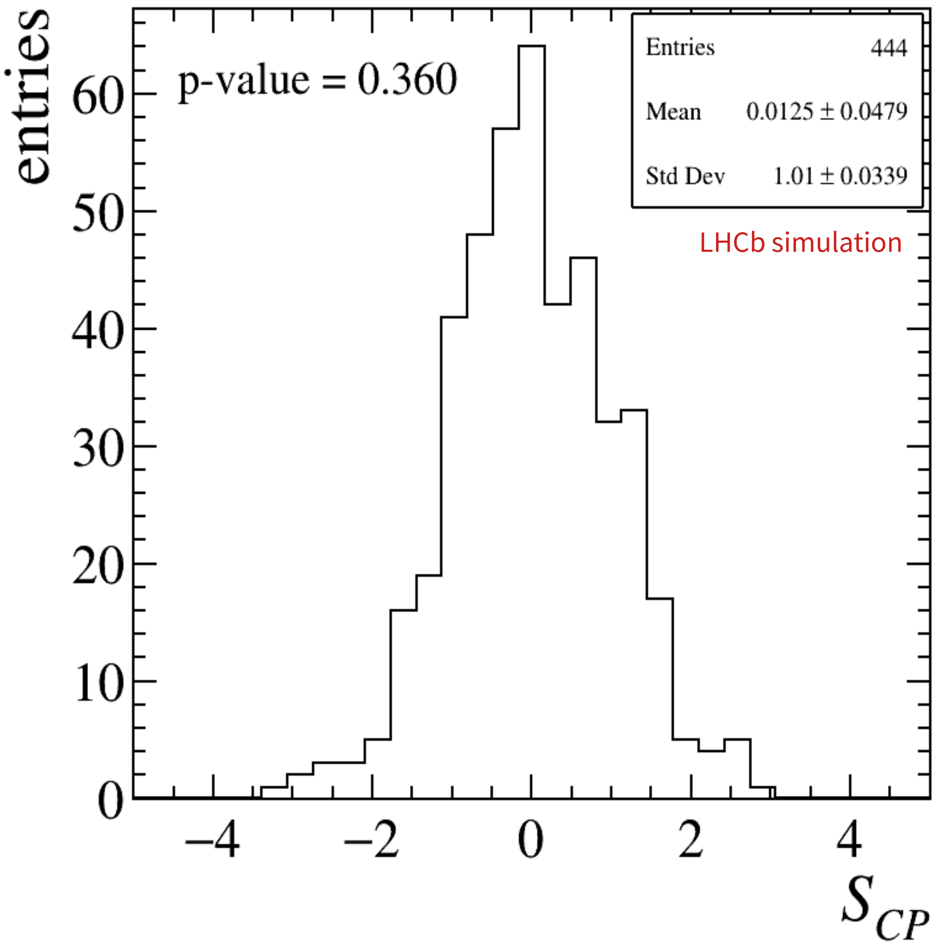
Without local asymmetries S_{CP} is Gaussian distribution with $\mu = 0$ and $\sigma = 1$.

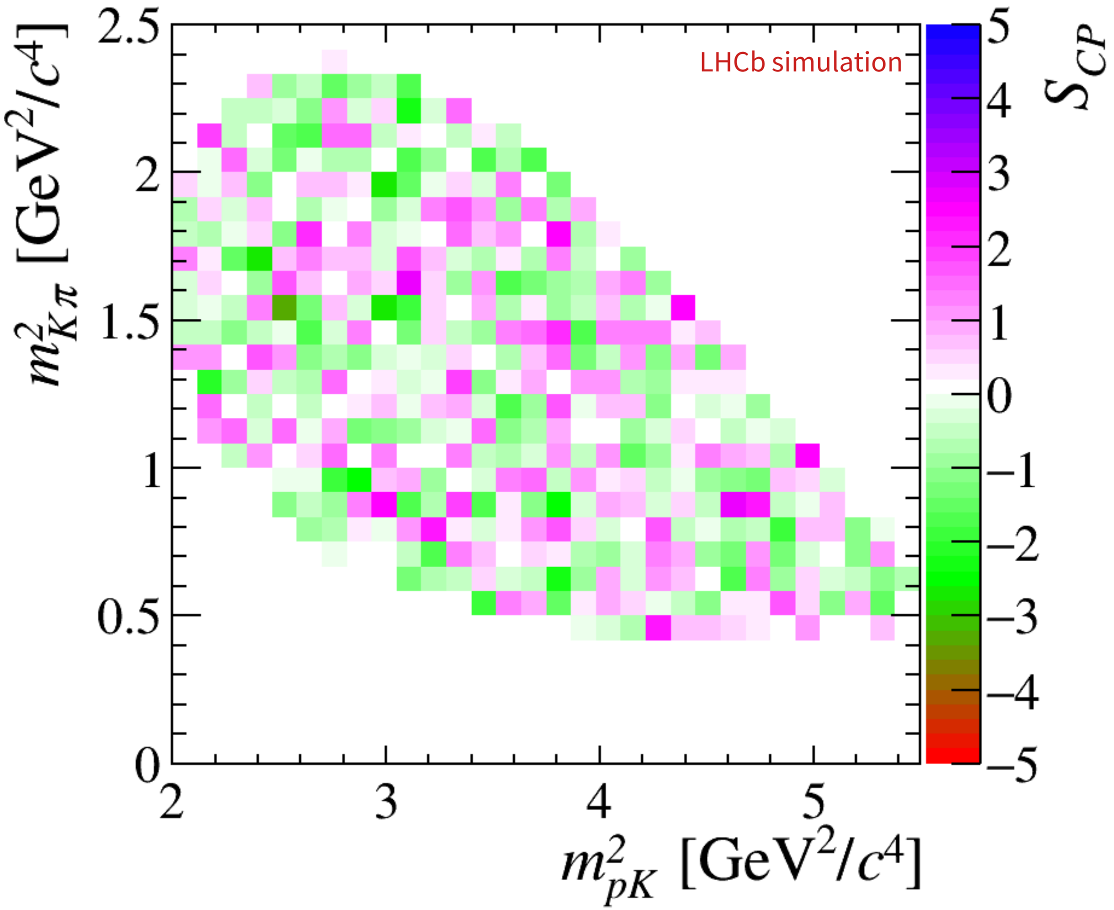
Calculate a $\chi^2/ndf \equiv \Sigma_i(S^i_{CP})^2 / (nbins-1)$

Obtain p-value   p-value $\ll 1$ in case of CPV

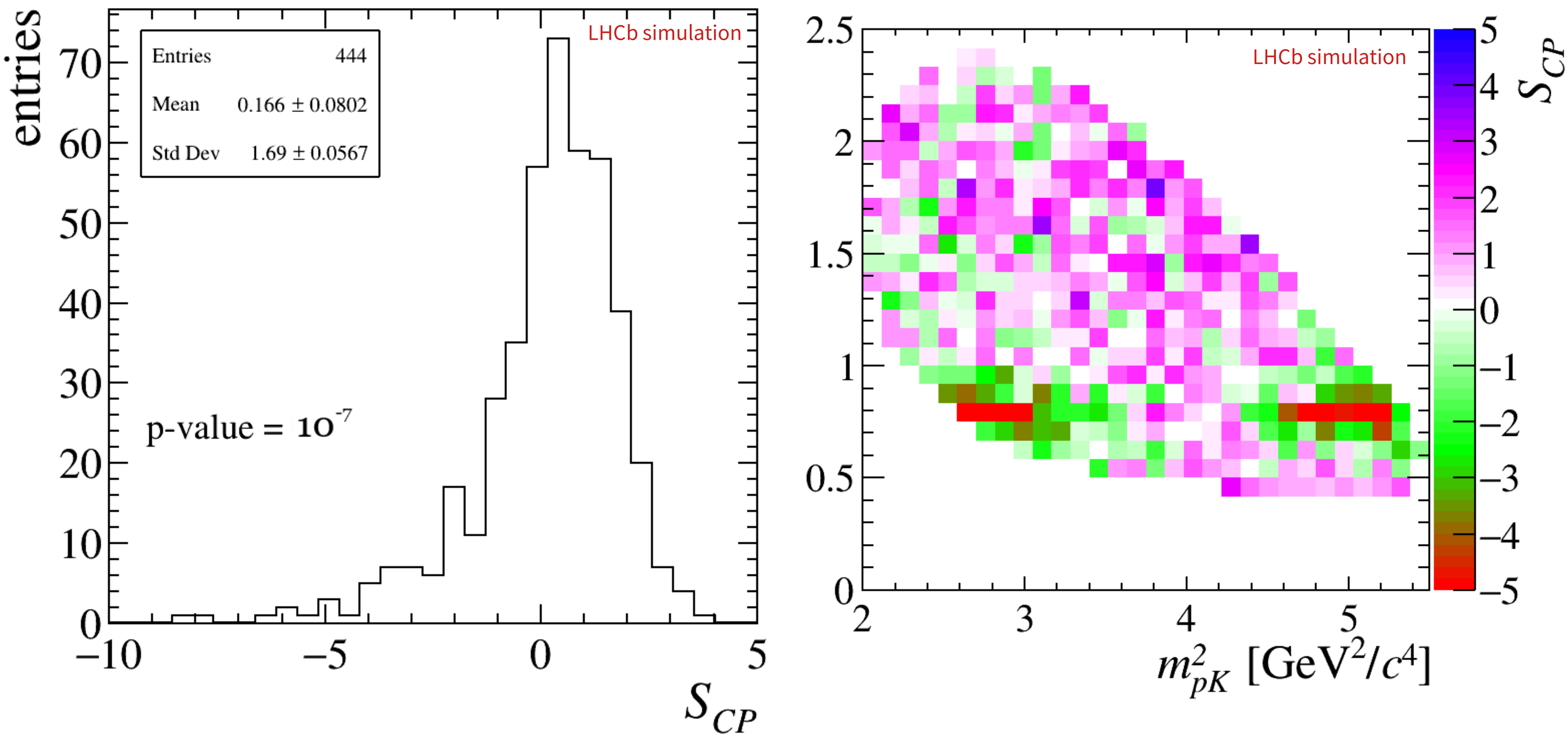
p-value < 0.0000003 (5σ)

S_{CP} for sample with 0% CP-violation

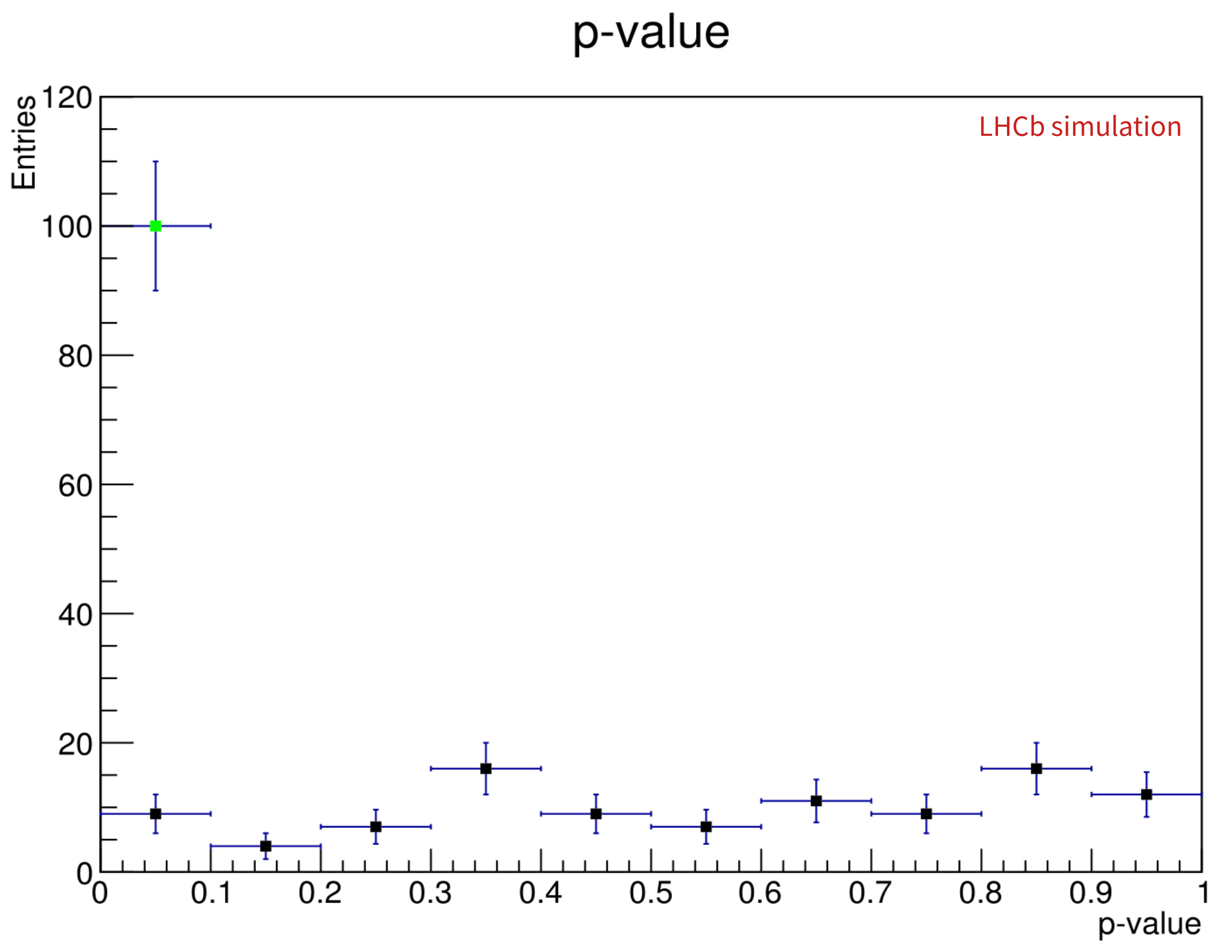




S_{CP} for sample with 20% CP-violation



S_{CP} – results for 100 pseudoexperiments



Black squares: p-value for samples with no CP-violation
Green square: p-value for samples with 20% CP-violation

Kernel Density Estimation

Kernel Density Estimation is a non-parametric way to estimate the probability density function *f* of a random variable.

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\omega(x - x_i, h)$$

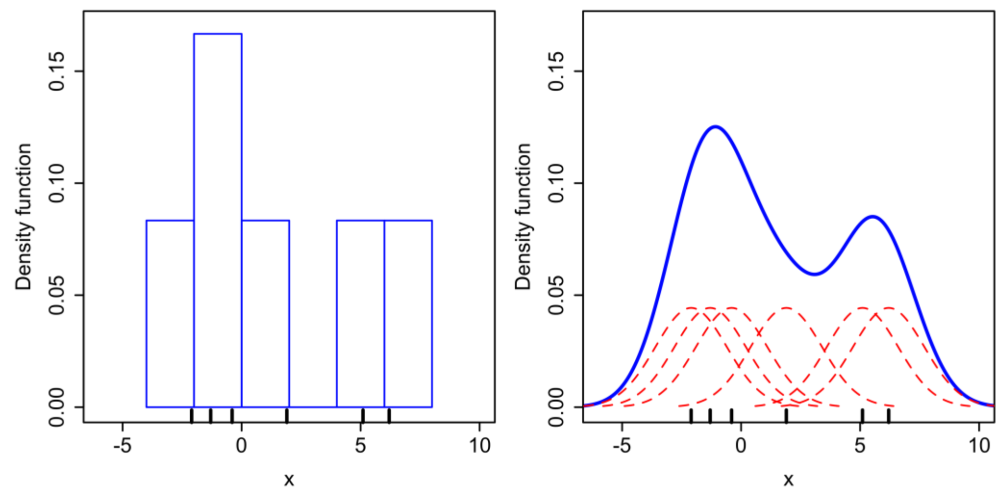
where:

$$\omega(t,h) = \frac{1}{h}K\left(\frac{t}{h}\right)$$

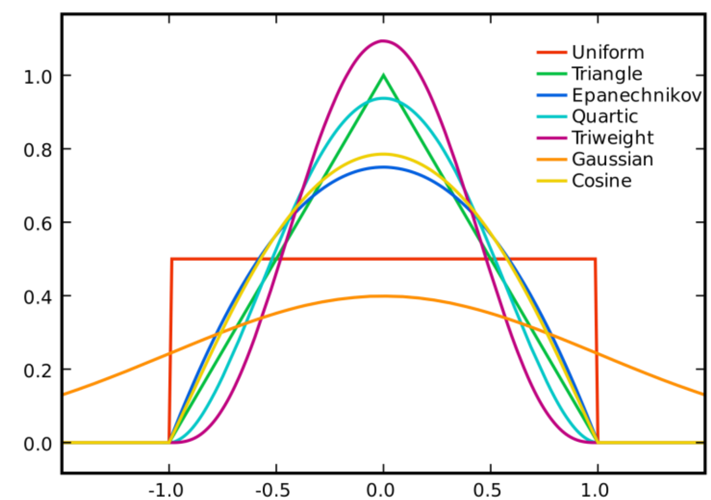
is the weighting function. ***K*** is the kernel, which determines the shape of the weighting function and ***h*** is the smoothing parameter.

In this analysis triangle kernel was used:

$$w(t,h) = \begin{cases} \frac{1}{h}(1 - |t|/h) & \text{for } |t| < h \\ 0 & \text{otherwise} \end{cases}$$

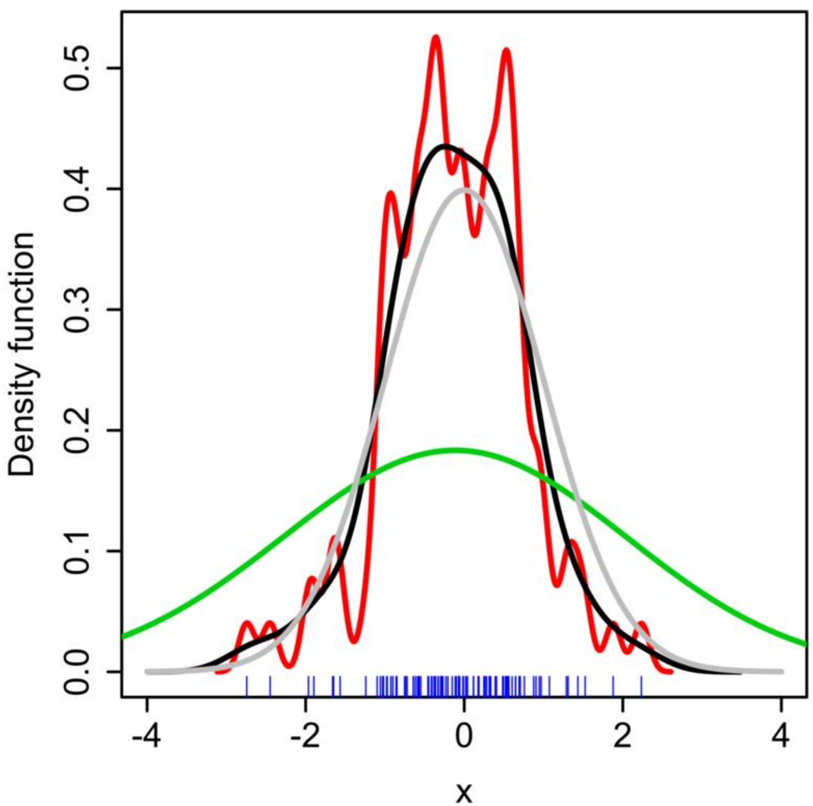


KDE example with Gaussian Kernel



Commonly used kernels

Bandwidth optimalization



KDE with different bandwidths of
a random sample

The bandwidth parameter ***h*** has a significant impact on the KDE method performance. For invariant functions one can use a globally determined bandwidth:

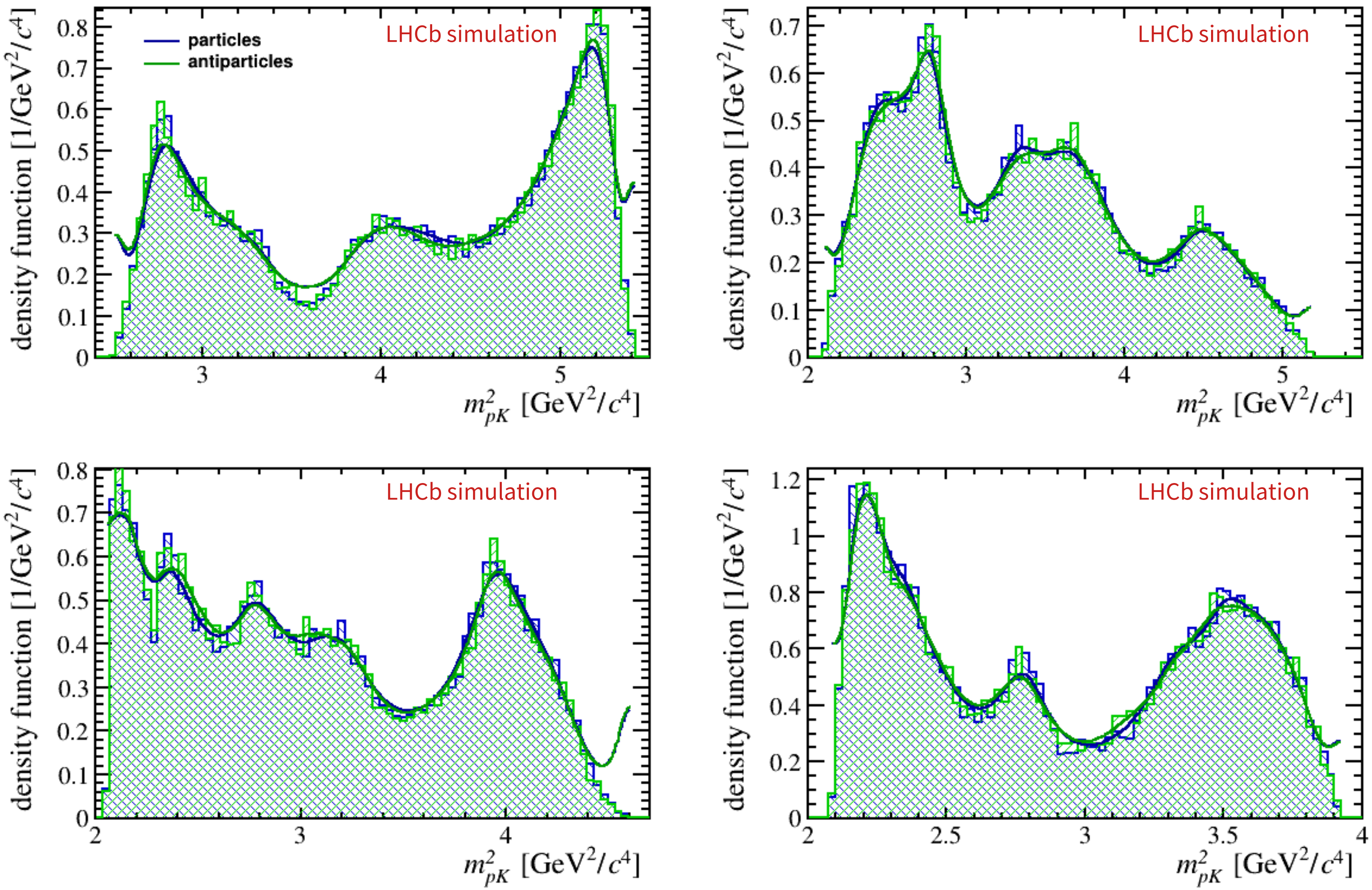
$$h = k\hat{S}N^{-\frac{1}{5}}$$

For distribution with more complicated shape a bandwidth parameter should depend on local features of the data. Hence, an adaptive h_{opt}, which takes into account properties of the analysed data should be considered:

$$h_{opt}^{i} = \frac{h}{\sqrt{\hat{f}(x_i)}}$$

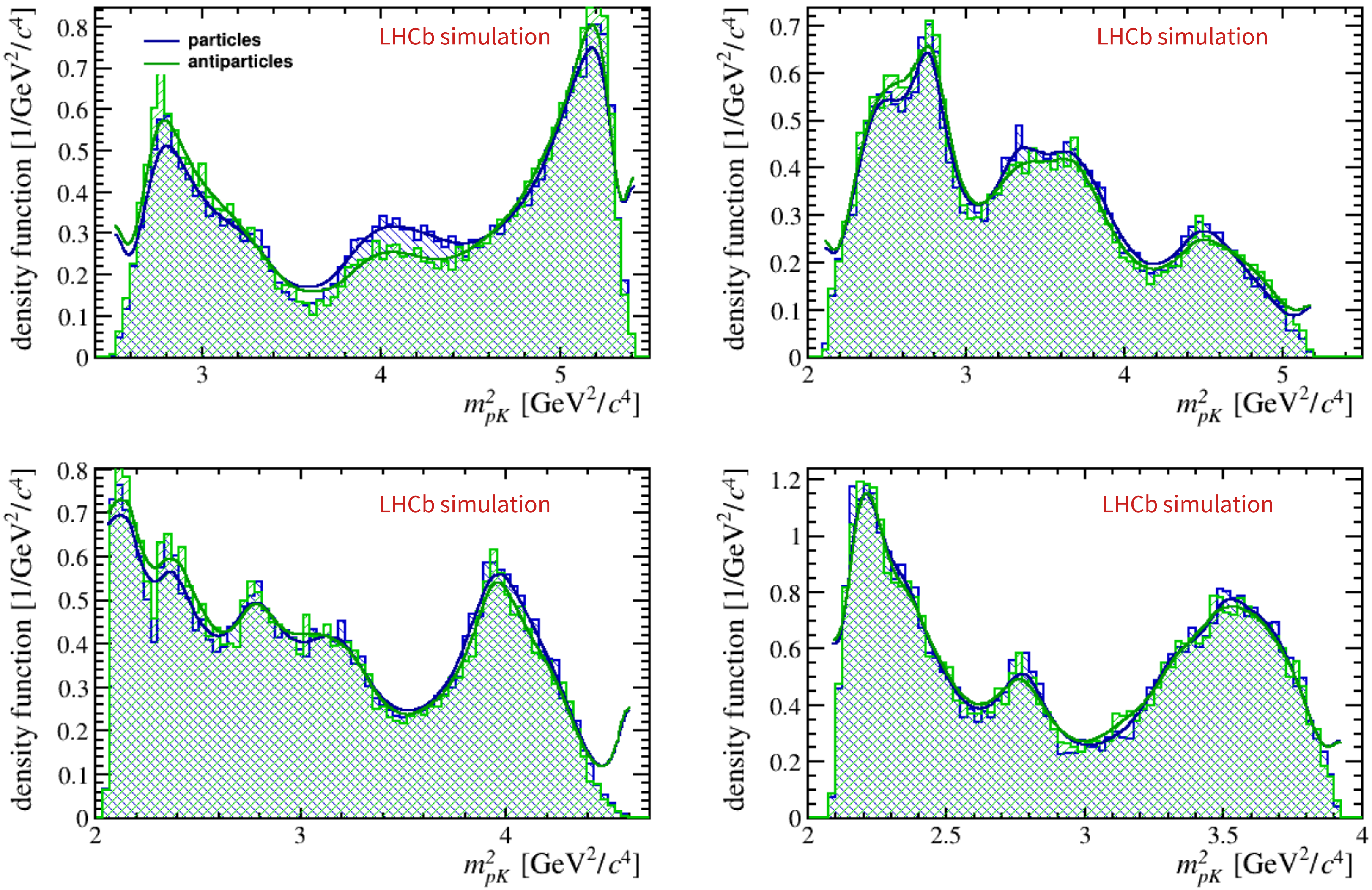
KDE – 1D results

Preliminary



Density function with triangle kernel for chosen regions for the sample with no CP-violation

KDE – 1D results



Density function with triangle kernel for chosen regions for the sample with 20% CP-violation

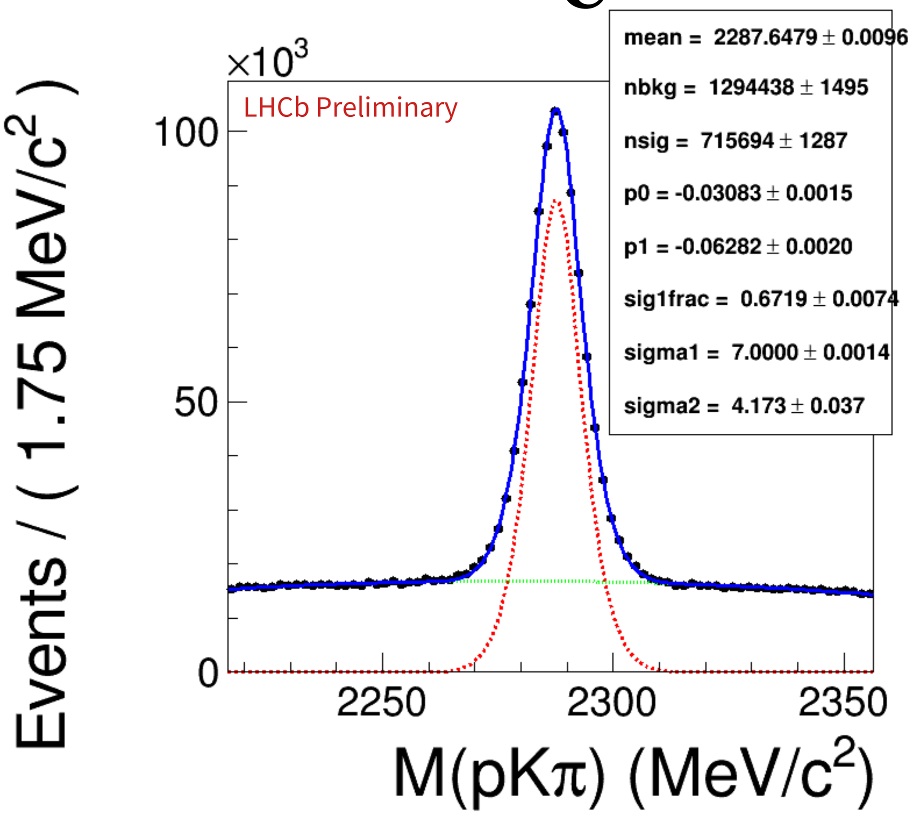
Data

- **2018 sample Run II**
- **First estimations based on 10 Ntuples – MagUp and MagDown**
- **~ 300 mln candidates with Ξ_c decays in 2016 - 2018 samples**

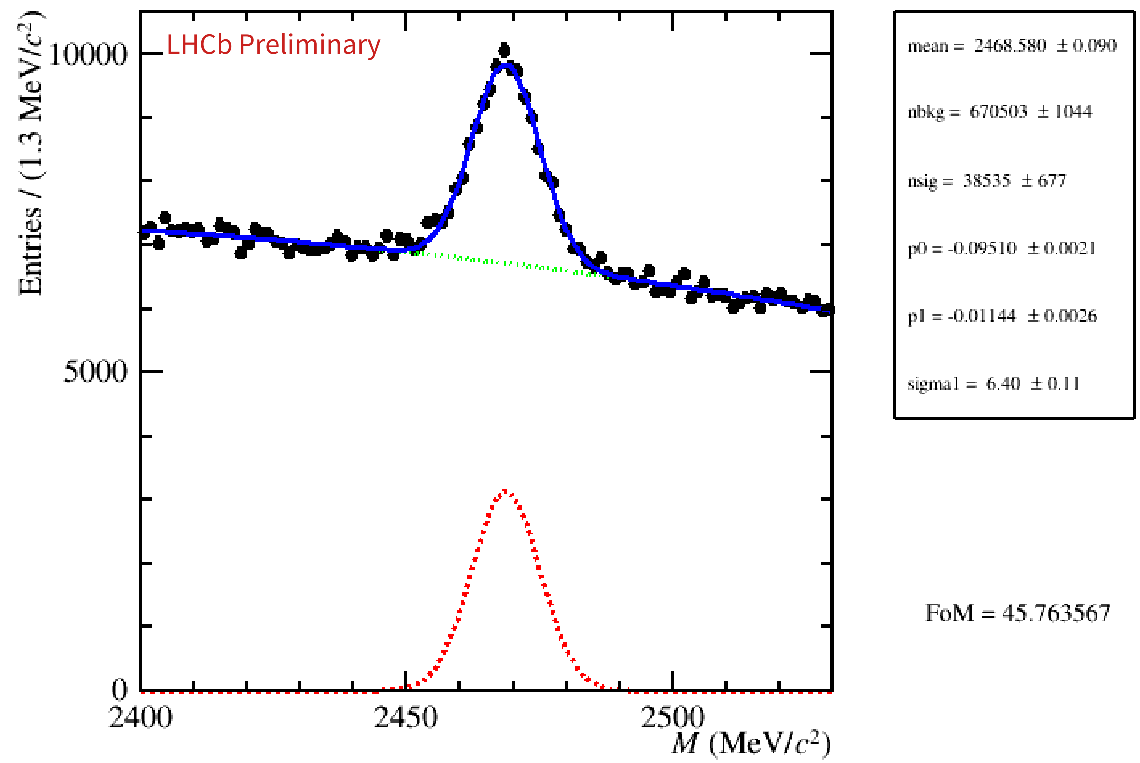
- **Cuts:**
  - **proton/kaon/pion:**
    - PID, ProbNN, IPX^2, TRACK_GhostProb, momentum,
  - **charm baryon**
    - Vertex X^2/ndof, IPX^2, transverse momentum, DIRA, η

Mass distributions – before cuts

Λ_c



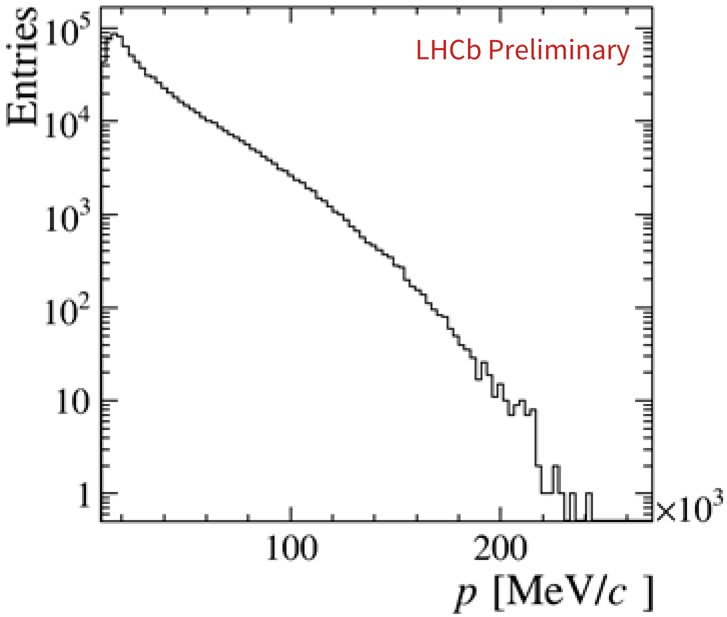
Ξ_c

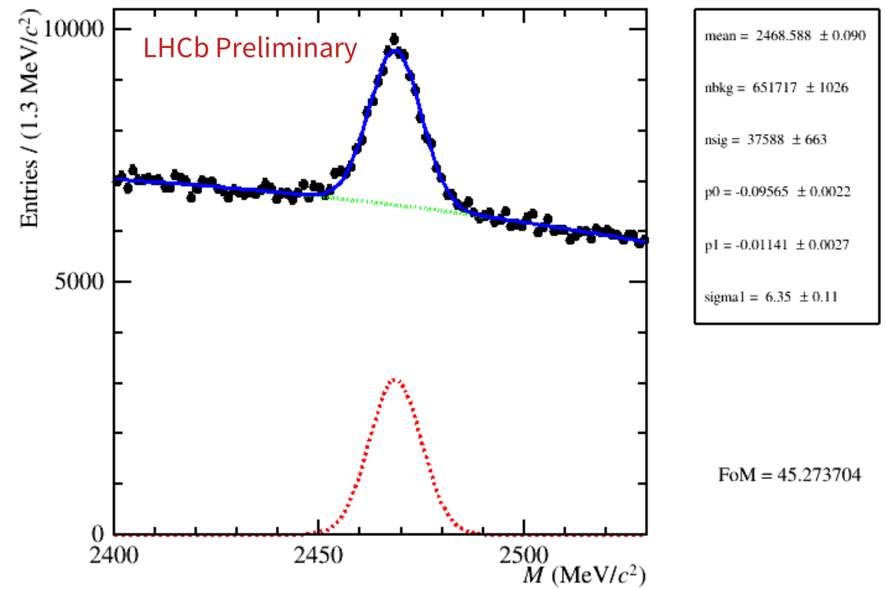


FoM = 46

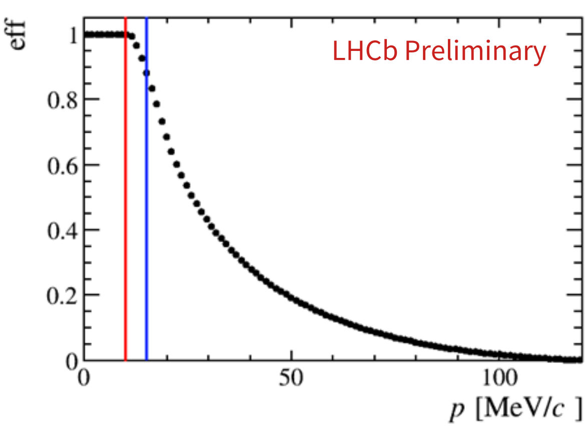
Figure of Merit (FoM): $S/\sqrt{(S+B)}$

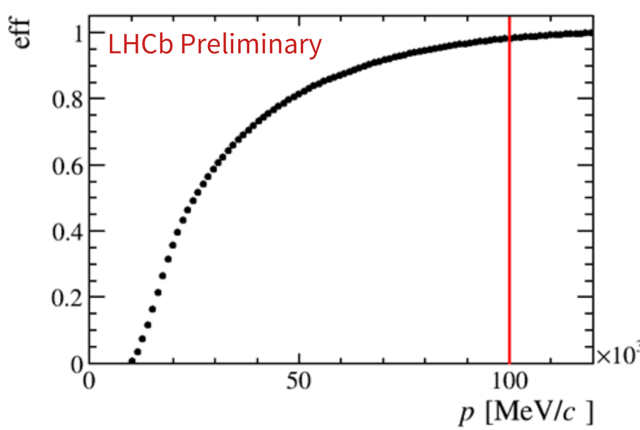
Proton momentum

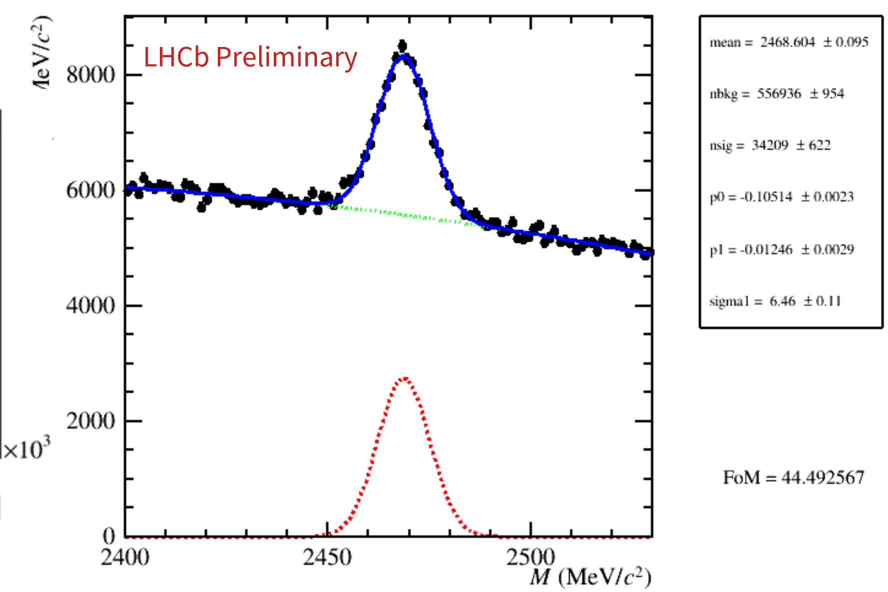




$10 < p < 100$ [GeV]
FoM = 45







10 [GeV] < p
15 [GeV] < p

p < 100 [GeV]

$$Eff = \frac{\#entries\ left\ after\ cut}{\#all\ entries}$$

$15 < p < 100$ [GeV]

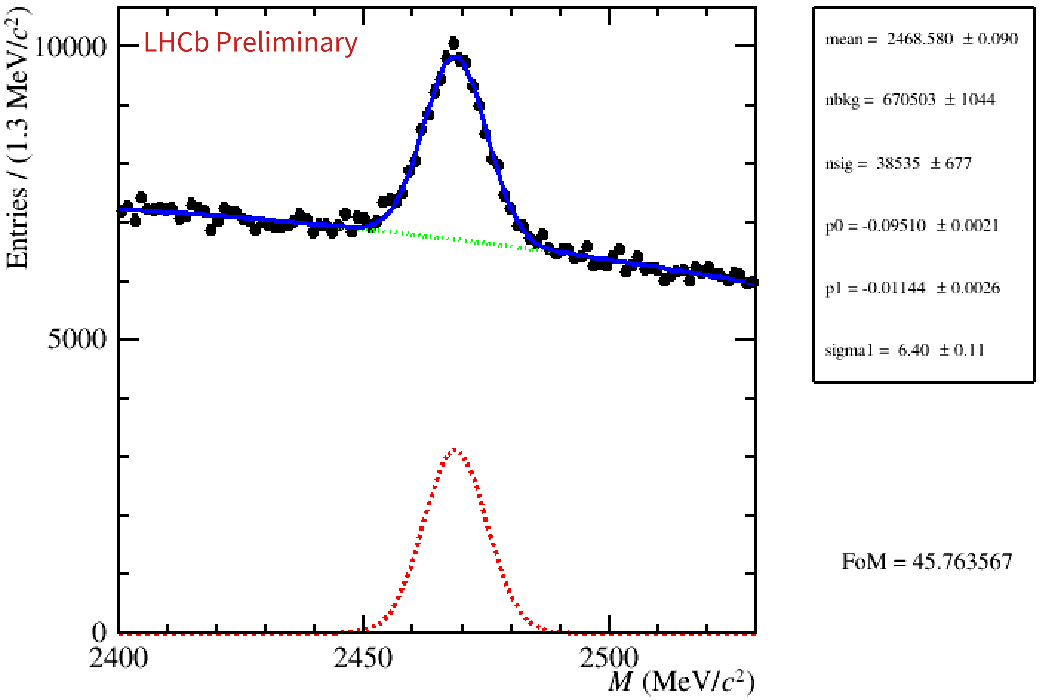
FoM = 44.5

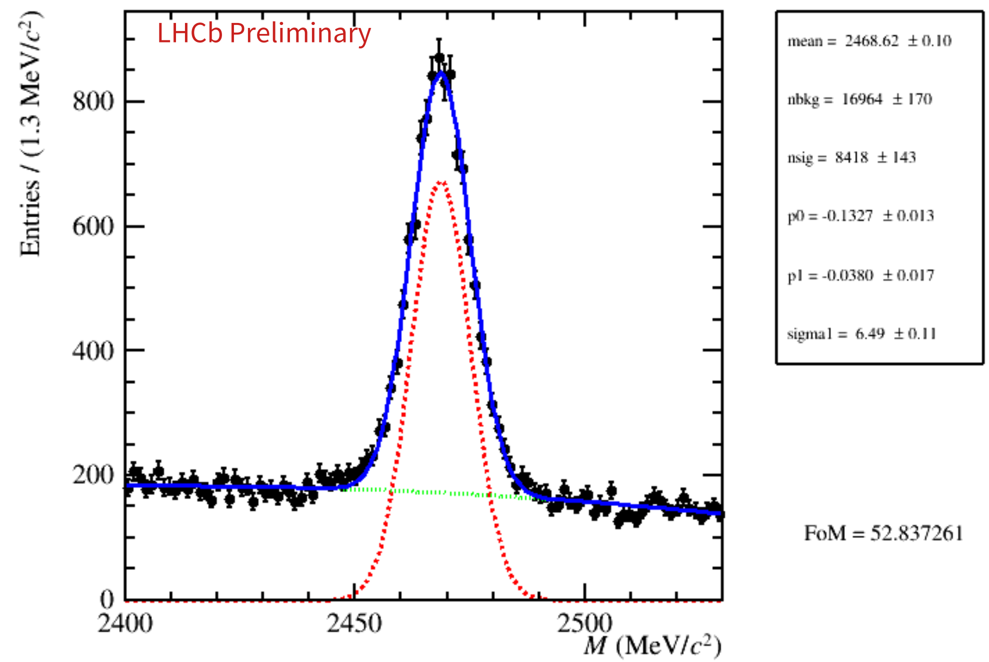
Tested cuts

proton, momentum

cut	10 & 80 [GeV]	15 & 80 [GeV]	10 &100 [GeV]	15 & 100 [GeV]
S	36k	33k	37.5k	34k
B	627k	533k	651k	556k
FoM	44.3	43.5	45	44.5

Ξ_c mass distribution - before & after cuts





Before cuts:
S = 38.5k
B = 670.5k
FoM = 46

After cuts:
S = 8.4k
B = 17k
FoM = 53

Analysis – other plans

- **Krakow – Warsaw Collaboration**
  - Run I completed (dr. Artur)
  - K-Nearest Neighbours unbinned method (kNN)
- **Krakow – UK Cooperation**
  - Developing and applying energy test

Conclusion

- The preliminary results show that the Kernel Density Estimation technique may potentially be sensitive to CPV and works properly when there is no asymmetry to be found
- Further studies will be carried out in order to improve performance of KDE method
- First computations on real data were performed
- Selection cuts are chosen:
  - FoM (before) = 46, FoM (after) = 53
- Expectation for whole real data sample:
  - ~1.3 mln Ξc canditates and ~100 mln Λc canditates in Run II data

Thank you for your attention!