LHCb Analysis with Snakemake

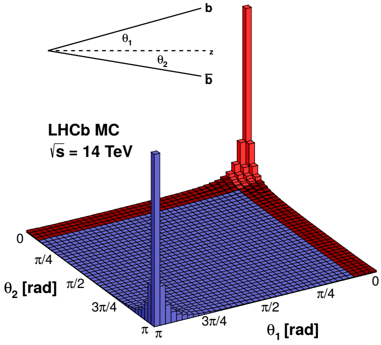
Pentaquark Search in $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$ Decays

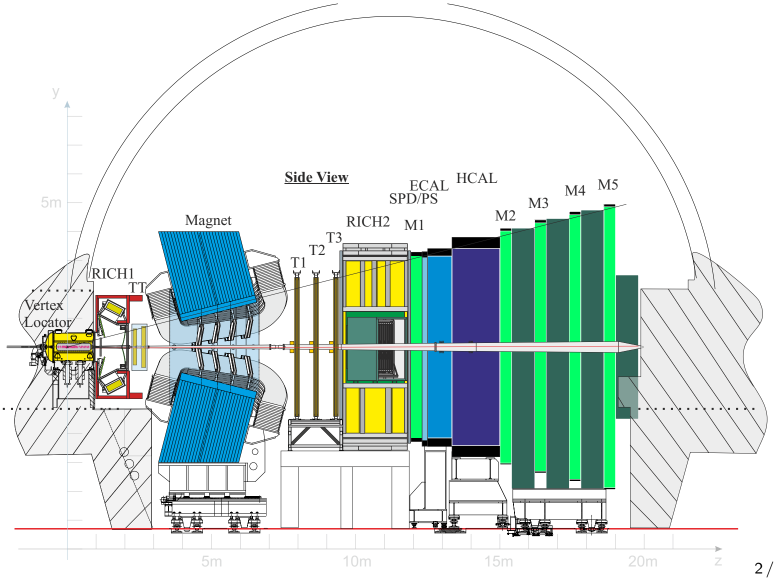
Dr. Mindaugas Šarpis

Vilnius University / The University of Manchester

April 4, 2024

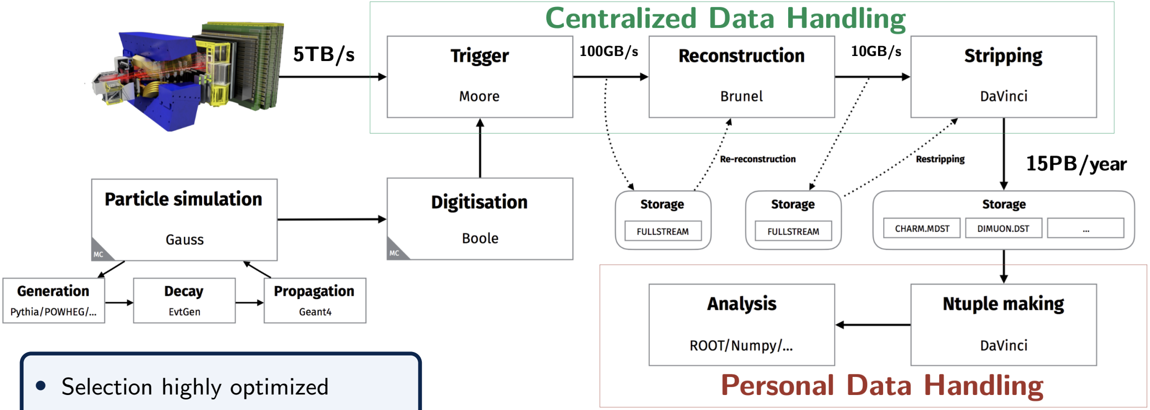
The LHCb Experiment





Acceptance:	$2 < \eta < 5$	
Vertex:	σ_{IP}	$\sim 20\ \mu$m
Time:	σ_{τ}	~ 45 fs
PID:	$\varepsilon(K)$	$\sim 95\%$

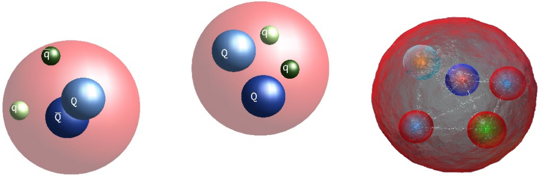
Selection of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ Decay Candidates



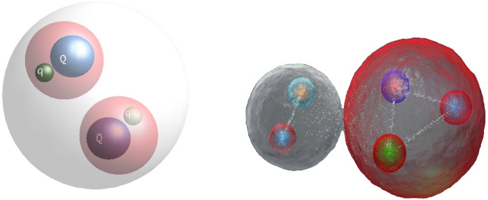
- Selection highly optimized
- Size of dataset:
  from 700 GB $\to$ 6.5 GB

The Nature of Exotic Hadrons

Tightly bound (compact) states

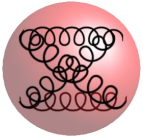


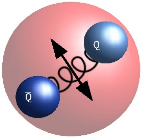
Molecular states



Other states

Glueball

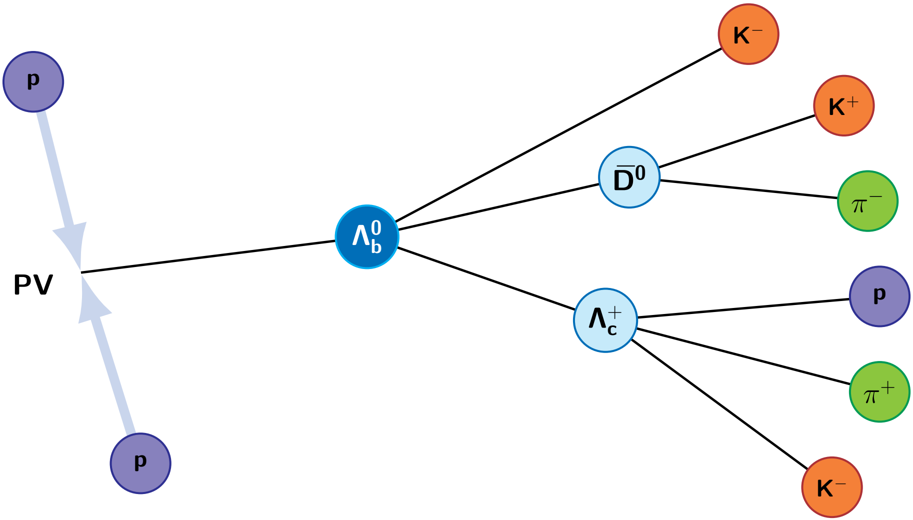




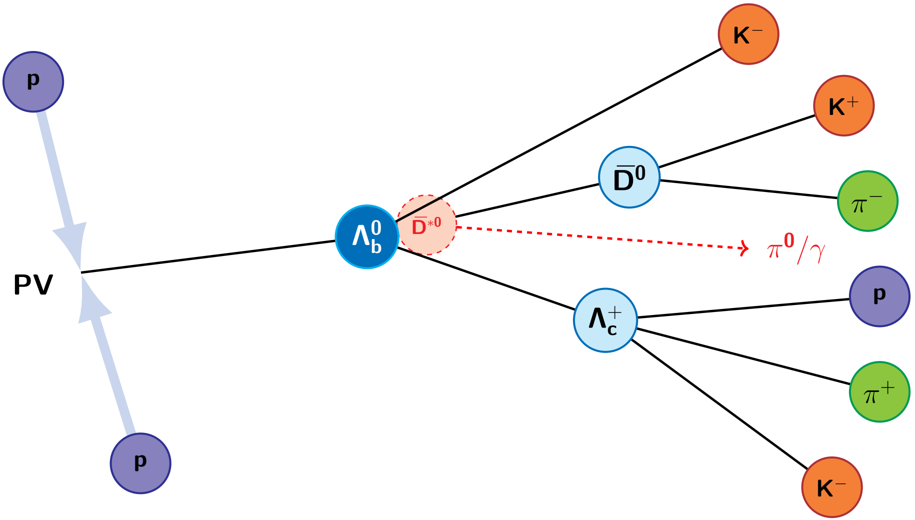
Hybrid

Soeren Lange / CERN graphics

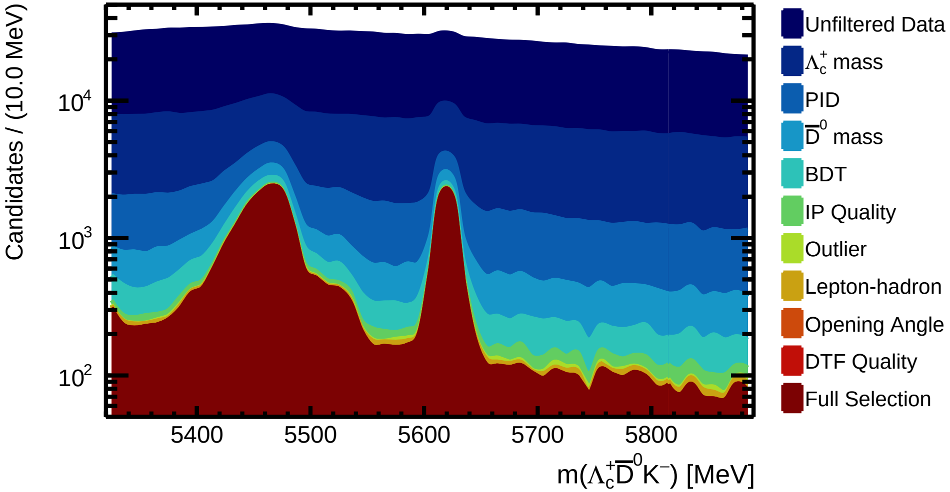
$\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ Decay



$\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$ Decay



Selection of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ Decay Candidates



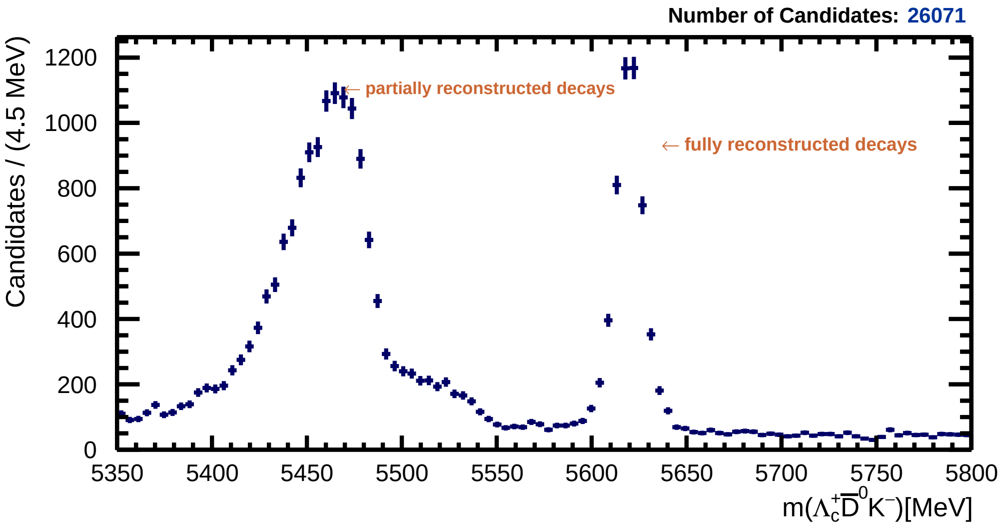
Selection of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ Decay Candidates

Different selection requirements (`cuts`) are applied to the candidates in the data set:

- Topological Cuts:
  - Number of tracks in the event
  - Impact parameter of tracks
  - Distance of closest approach of tracks
- Kinematic Cuts:
  - Momentum `P`
  - Transverse momentum $\mathtt{P_T}$
  - Mass `m`
- PID Cuts:
  - `ProbNN` response
  - `BDT` response
  - PID from RICH

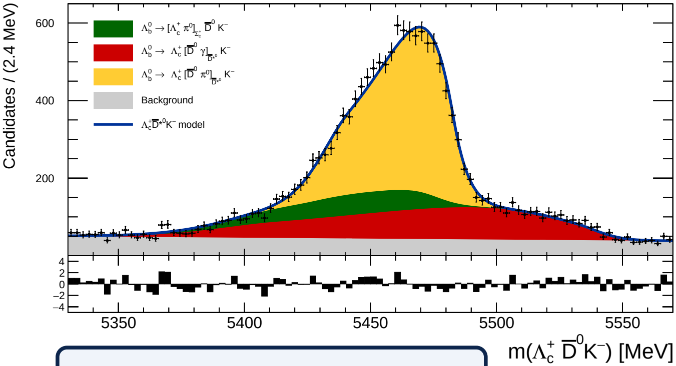
Variable	Cut	Applied to
`nLongTracks`	$<$ 500	global
`TRCHI2DOF`	$<$ 3	all tracks
`P`	$>$ 1000 MeV	all tracks
`PT`	$>$ 100 MeV	all tracks
`MIPCHI2DV(PV)`	$>$ 4	all tracks
`DOCACHI2`	$<$ 16/20/25	$\Lambda_c^+\overline{D}^0 / \Lambda_c^+K^- / \overline{D}^0K^-$
`DOCA`	$<$ 0.2/0.4/0.5 mm	$\Lambda_c^+\overline{D}^0 / \Lambda_c^+K^- / \overline{D}^0K^-$
`CHILDIP`	$<$ 0.2/0.3/0.5 mm	$\Lambda_c^+ / \overline{D}^0 / K^-$
`CHILDIPCHI2`	$<$ 16	$\Lambda_c^+, \overline{D}^0, K^-$
`CHI2VXNDF`	$<$ 32	$\Lambda_c^+\overline{D}^0K^-$ vtx.
`BPVVDZ`	$>$ 0.2 mm	$\Lambda_c^+\overline{D}^0K^-$ vtx.
`PT`	$>$ 4 GeV	Λ_b^0
`P`	$>$ 32 GeV	Λ_b^0
Λ_c^+ `BDT`	$>$ -0.15	Λ_c^+
$\overline{D}^0$ `BDT`	$>$ -0.3	$\overline{D}^0$
$\vdots$	$\vdots$	$\vdots$

$\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ Data Sample



Fit to $\Lambda_c^+ \overline{D}^0 K^-$ Invariant Mass Spectrum

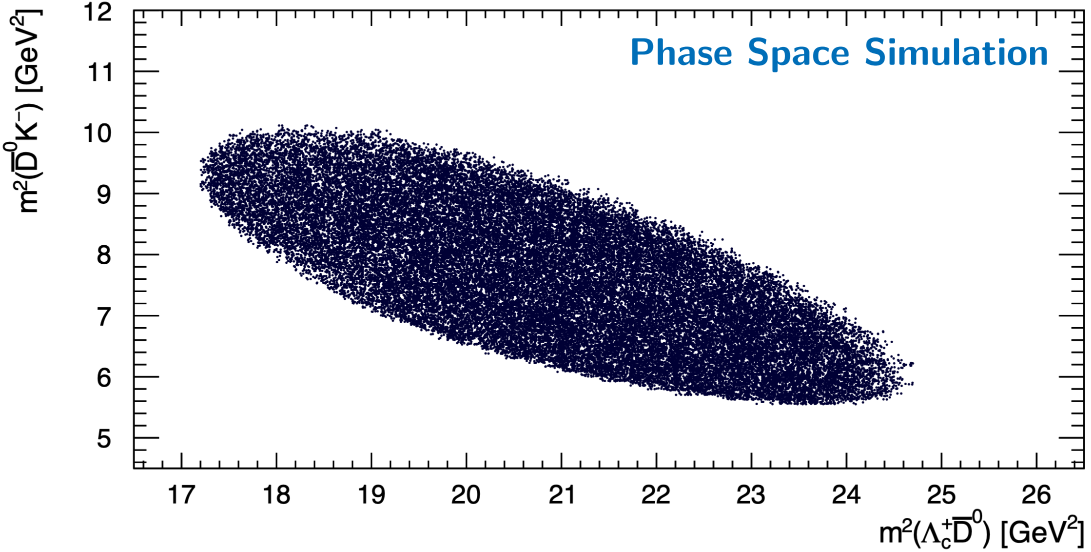
- Dominant component is $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0}(\to \overline{D}^0 \pi^0) K^-$
- $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0}(\to \overline{D}^0 \gamma) K^-$ $\sim 23\%$ in range of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0}(\to \overline{D}^0 \pi^0) K^-$
- $\Lambda_b^0 \to \Sigma_c^+(\to \Lambda_c^+ \pi^0) \overline{D}^0 K^-$ is suppressed, but indistinguishable from $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0}(\to \overline{D}^0 \pi^0) K^-$

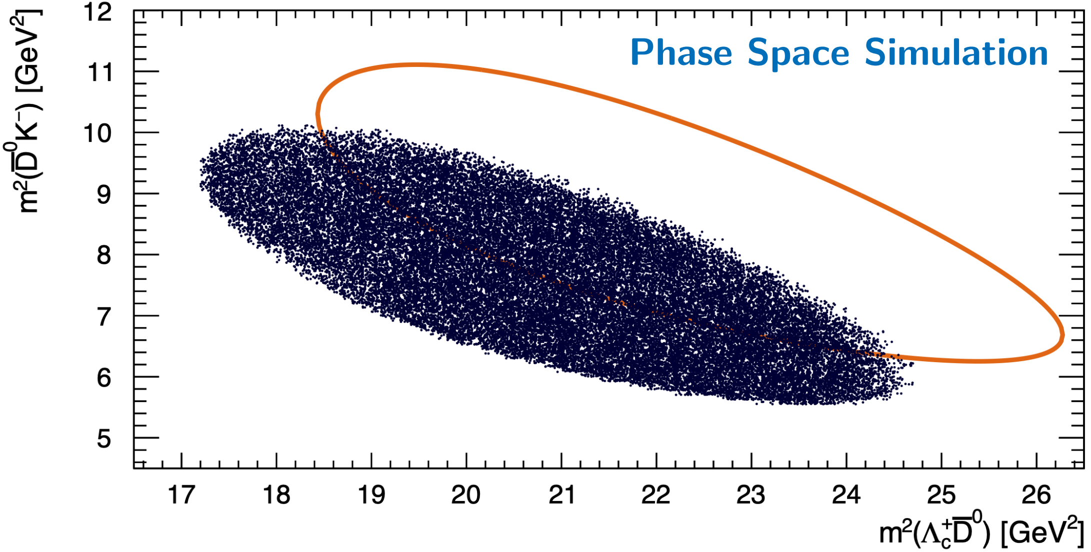


D^{*0} Decay Modes (PDG)

Γ_1	$D^0\pi^0$	$64.7 \pm 0.9\,\%$
Γ_2	$D^0\gamma$	$35.3 \pm 0.9\,\%$

$N_{\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0}(\to \overline{D}^0 \pi^0) K^-}$	7512 ± 162
$N_{\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0}(\to \overline{D}^0 \gamma) K^-}$	4071 ± 175
$N_{\Lambda_b^0 \to \Sigma_c^+(\to \Lambda_c^+ \pi^0) \overline{D}^0 K^-}$	1637 ± 178
$N_{background}$	4353 ± 131


Phase Space Simulation
$m^2(\overline{D}^0 K^-)$ [GeV^2]
$m^2(\Lambda_c^+ \overline{D}^0)$ [GeV^2]
12
11
10
9
8
7
6
5
17
18
19
20
21
22
23
24
25
26

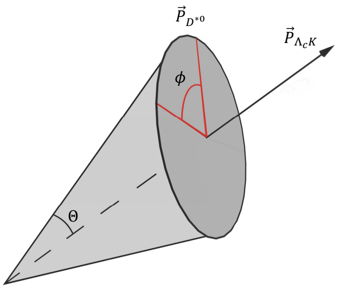


Phase Space Simulation
$m^2(\bar{D}^0 K^-)$ [GeV^2]
12
11
10
9
8
7
6
5
17
18
19
20
21
22
23
24
25
26
$m^2(\Lambda_c^+ \bar{D}^0)$ [GeV^2]

Reconstruction of $\overline{D}^{*0}$ Missing Momentum

Extended Cone Closure (ECC) Method

- Kinematic over-constraint method developed used to reconstruct missing 4-momentum of $\overline{D}^{*0}$ under γ or π^0 hypothesis
- Method is adaptable to any analysis with missing particles (especially with similar decay topology)

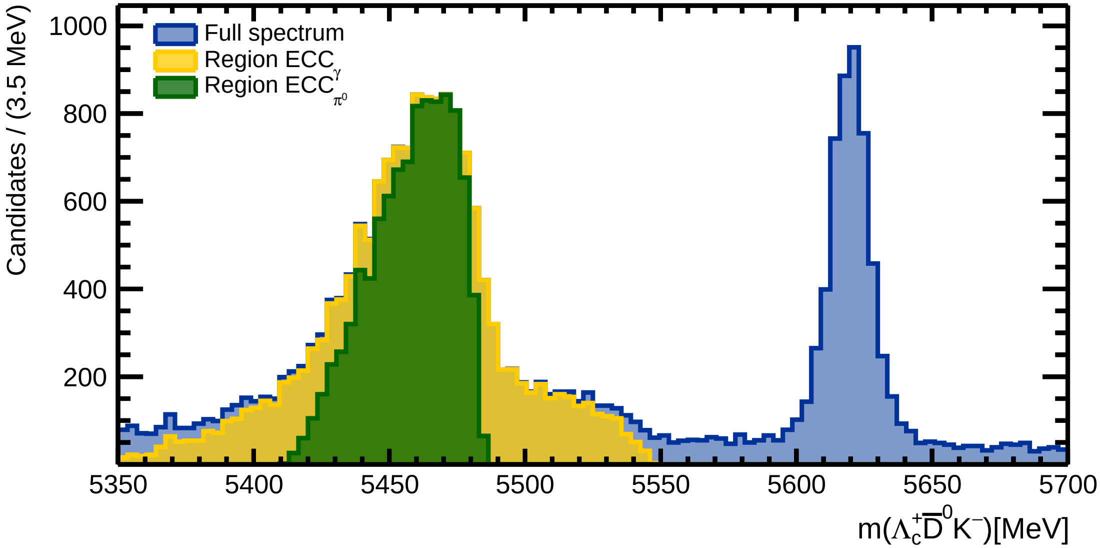


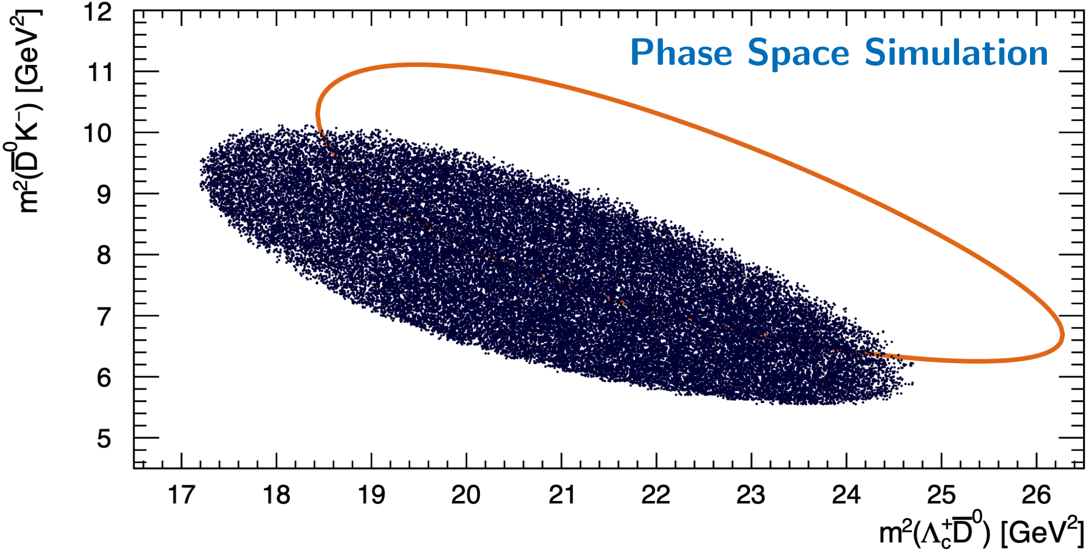
ECC Constraints

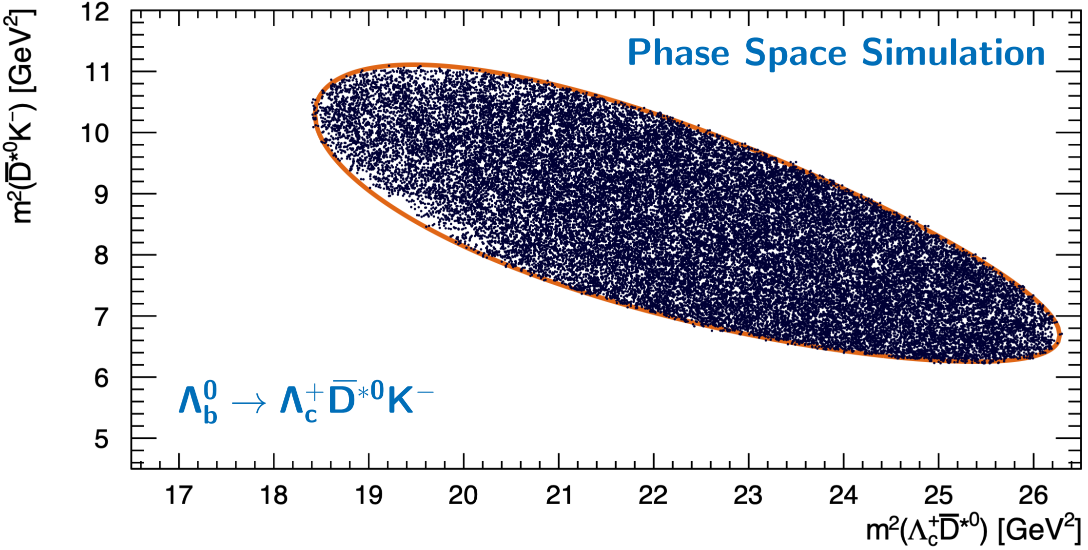
- $m_{\Lambda_b^0} \rightarrow 5619.6$ MeV
- $m_{\overline{D}^{*0}} \rightarrow 2006.8$ MeV
- m_X 0 MeV **or** 135 MeV (γ **or** π^0)
- Flight direction of Λ_b^0 points to PV

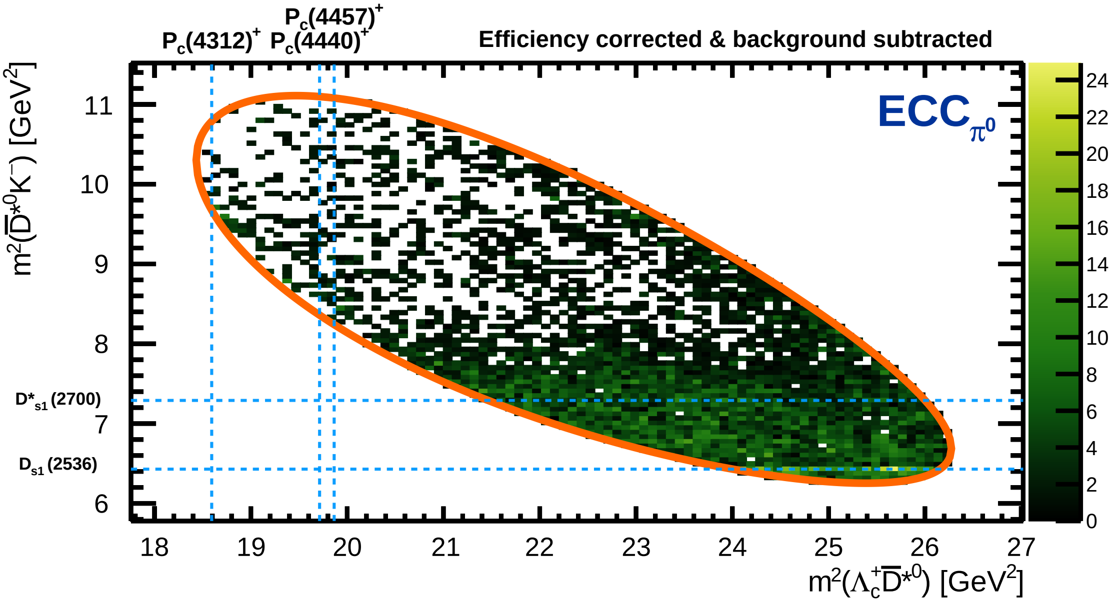
$$p_{\overline{D}^{*0}} = \sqrt{\left(\frac{m^2_{\overline{D}^{*0}} - m^2_{\overline{D}^0} - m^2_X}{2 \cdot m_{\overline{D}^0}}\right)^2 - m^2_X}$$

$$\cos\theta = \frac{m^2_{\overline{D}^{*0}} + m^2_{\Lambda_c^+ K^-} + 2 \cdot E_{\overline{D}^{*0}} \cdot E_{\Lambda_c^+ K^-} - m^2_{\Lambda_b^0}}{2 \cdot p_{\overline{D}^{*0}} \cdot p_{\Lambda_c^+ K^-}}$$

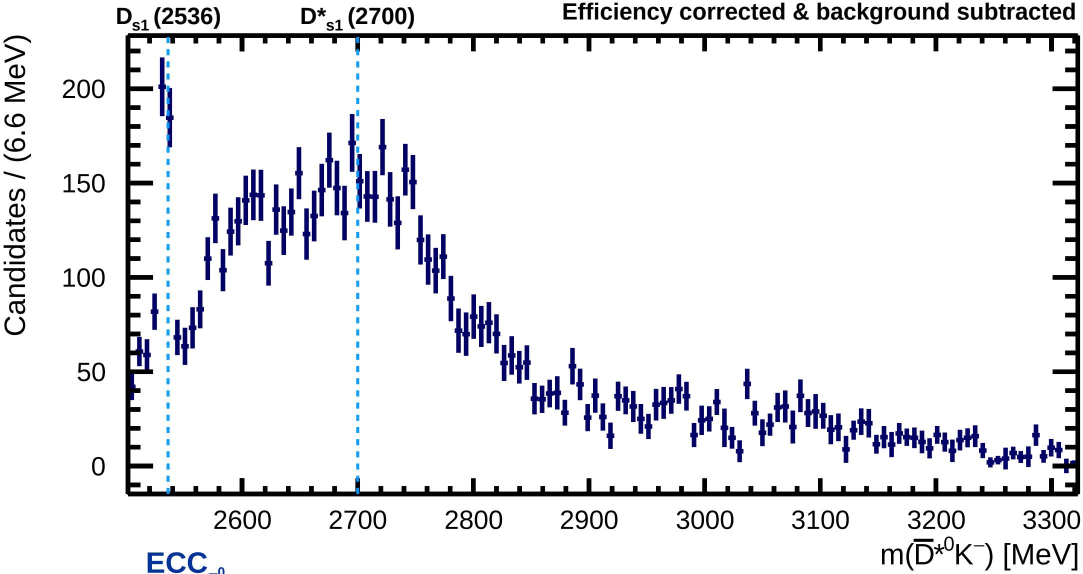

Full spectrum
Region ECC$_\gamma$
Region ECC$_{\pi^0}$
Candidates / (3.5 MeV)
1000
800
600
400
200
5350
5400
5450
5500
5550
5600
5650
5700
$m(\Lambda_c^+\overline{D}^0K^-)$[MeV]


Phase Space Simulation
$m^2(\overline{D}^0K^-)$ [GeV^2]
$m^2(\Lambda_c^+\overline{D}^0)$ [GeV^2]
12
11
10
9
8
7
6
5
17
18
19
20
21
22
23
24
25
26


Phase Space Simulation
$\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$
$m^2(\overline{D}^{*0}K^-)$ [GeV^2]
$m^2(\Lambda_c^+\overline{D}^{*0})$ [GeV^2]
12
11
10
9
8
7
6
5
17
18
19
20
21
22
23
24
25
26

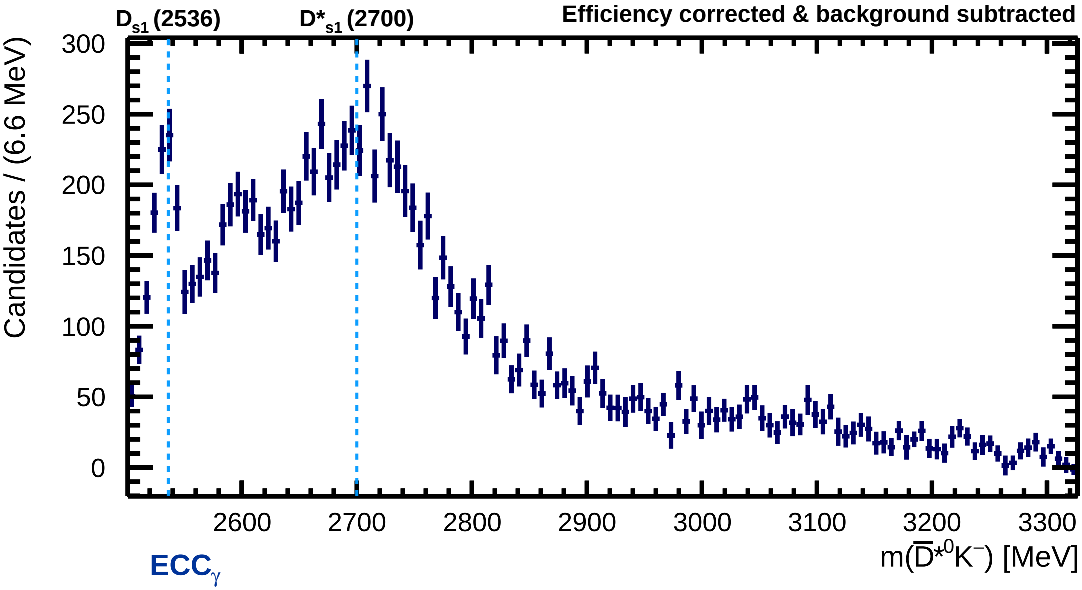


$P_c(4457)^+$
$P_c(4312)^+$ $P_c(4440)^+$
Efficiency corrected & background subtracted
ECC_{π^0}
$m^2(\bar{D}^{*0}K^-)$ [GeV^2]
$D^*_{s1}(2700)$
$D_{s1}(2536)$
$m^2(\Lambda_c^+\bar{D}^{*0})$ [GeV^2]

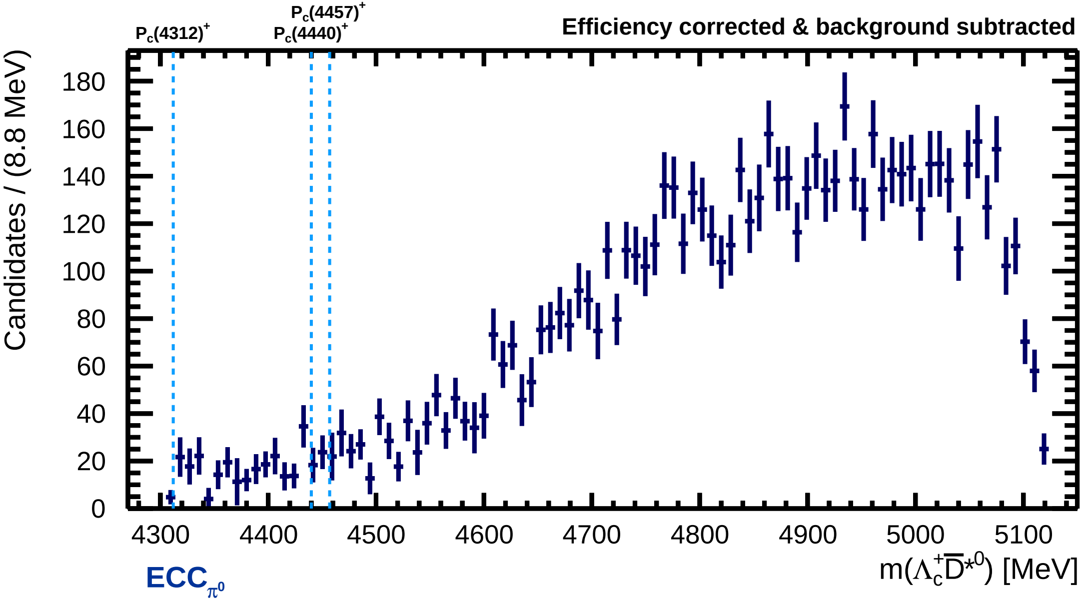


D_s1 (2536)
D*_s1 (2700)
Efficiency corrected & background subtracted
Candidates / (6.6 MeV)
200
150
100
50
0
2600
2700
2800
2900
3000
3100
3200
3300
m(D̄*0K−) [MeV]

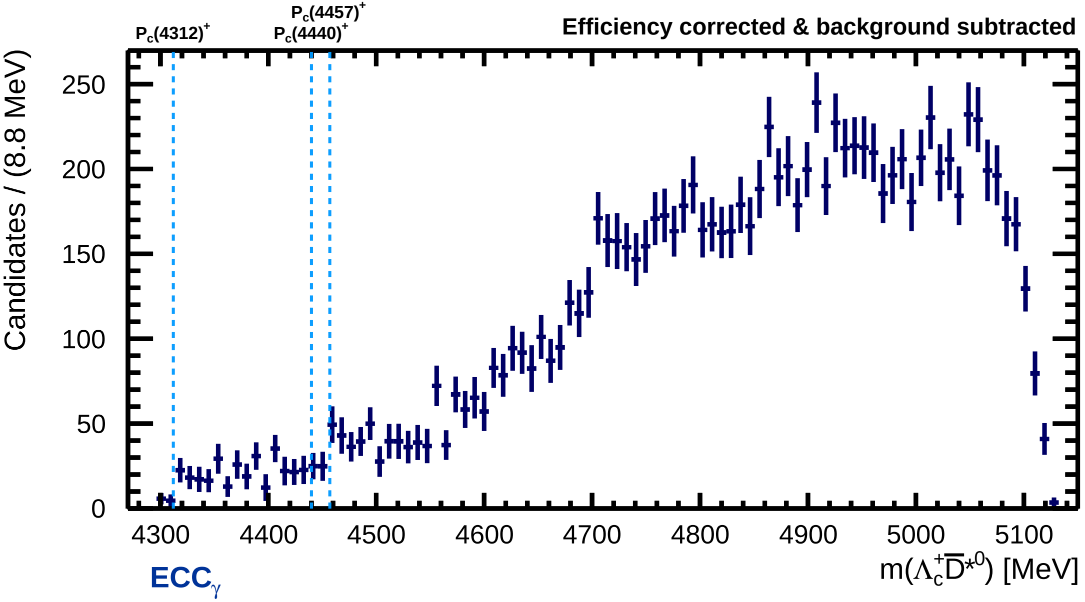
ECC_{π^0}



Efficiency corrected & background subtracted
$D_{s1}(2536)$
$D^*_{s1}(2700)$
Candidates / (6.6 MeV)
300
250
200
150
100
50
0
2600
2700
2800
2900
3000
3100
3200
3300
$m(\bar{D}^{*0}K^{-})$ [MeV]
ECC_γ



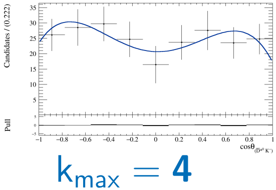
Efficiency corrected & background subtracted
$P_c(4312)^+$
$P_c(4440)^+$
$P_c(4457)^+$
Candidates / (8.8 MeV)
0
20
40
60
80
100
120
140
160
180
4300
4400
4500
4600
4700
4800
4900
5000
5100
$m(\Lambda_c^+\overline{D}^{*0})$ [MeV]
ECC_{π^0}

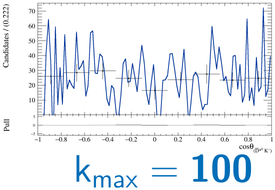


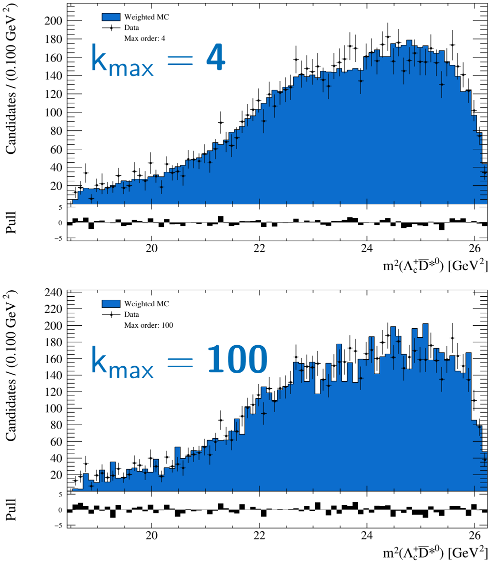
$P_c(4457)^+$
$P_c(4312)^+$
$P_c(4440)^+$
Efficiency corrected & background subtracted
Candidates / (8.8 MeV)
0
50
100
150
200
250
4300
4400
4500
4600
4700
4800
4900
5000
5100
$m(\Lambda_c^+\overline{D}^{*0})$ [MeV]
ECC_γ

Model Independent Approach

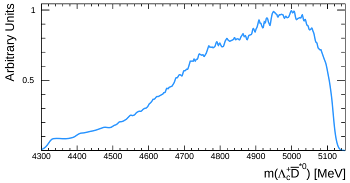
- Series of Legendre Moments up to order 4 describe the $\Lambda_c^+ \overline{D}^{*0}$ distribution well
- Any sizable fluctuations in data here would already be a hint to the presence of exotic resonances
- Interference effects from coupled channels or between resonances in $\Lambda_c^+ \overline{D}^{*0}$ itself could also cause disagreement
- The shape obtained from Legendre Moments analysis becomes a *background* model to the CLs Limit Setting procedure



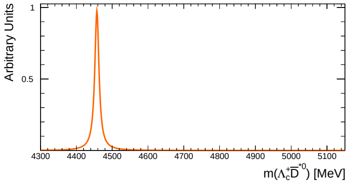




Building the Full Model

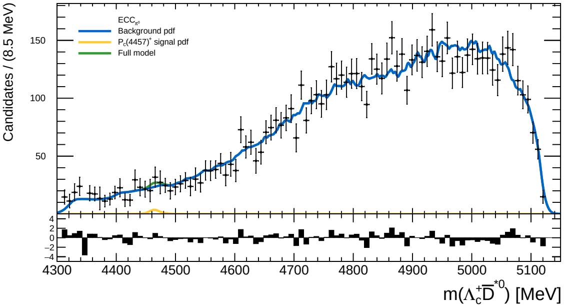


+

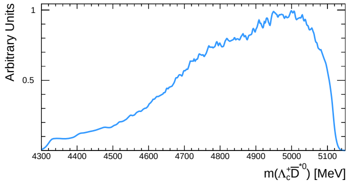


=

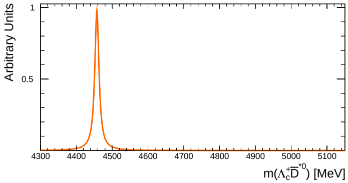
- *Background* model is the shape obtained from moments analysis
- Narrow P_c^+ signal is introduced
- Fit to data is performed
- Likelihood for *Background* only model and for Signal + *Background* model is evaluated



Building the Full Model

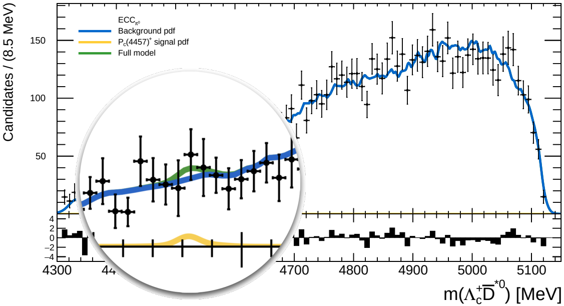


+



=

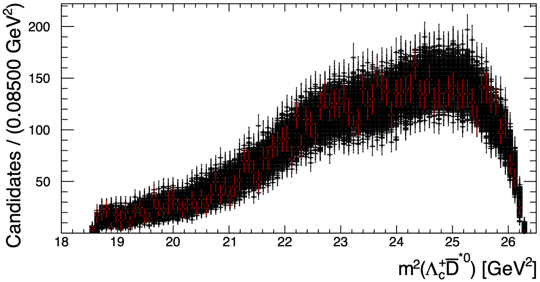
- *Background* model is the shape obtained from moments analysis
- Narrow P_c^+ signal is introduced
- Fit to data is performed
- Likelihood for *Background* only model and for Signal + *Background* model is evaluated

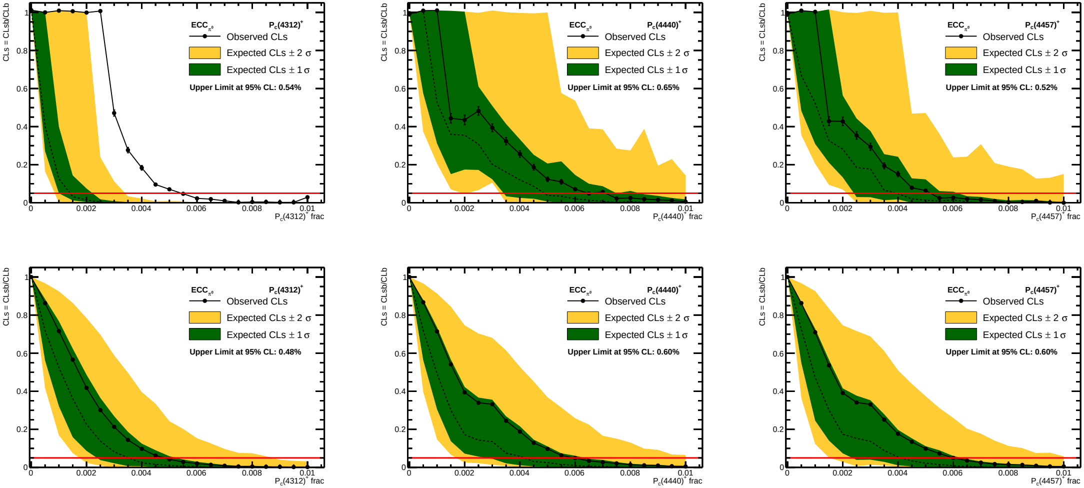


Statistical Test

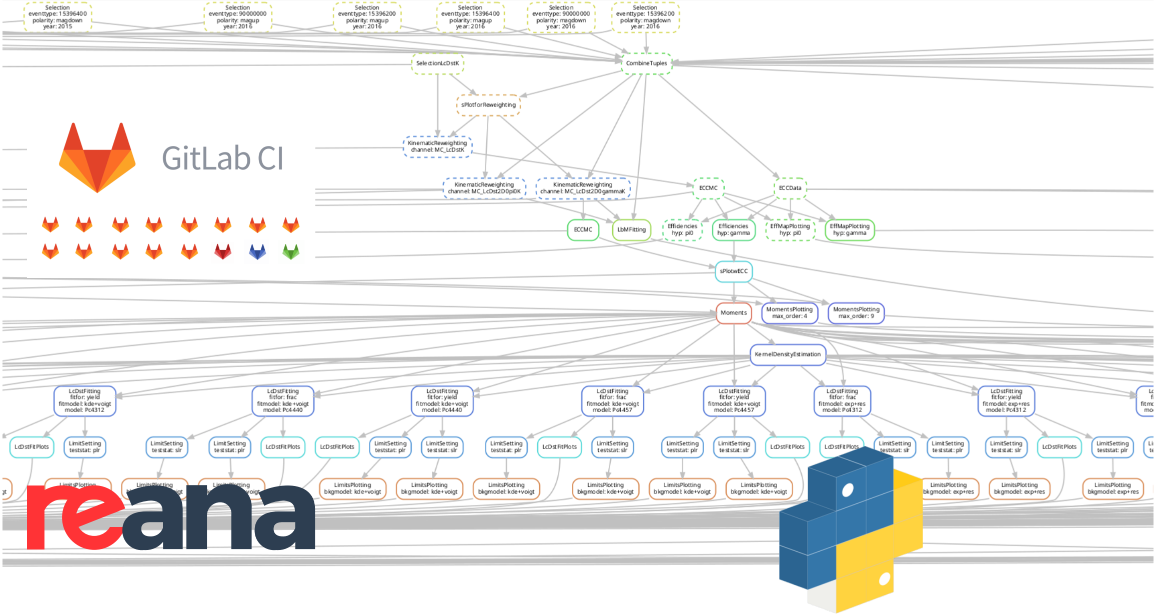
- Test statistic is defined as the Likelihood ratio
- Multiple test statistics were tried
  - ▷ Profile Likelihood Ratio (PLR) is the nominal test statistic
  - ▷ Simple Likelihood Ratio (SLR) yields very similar results but is much faster
- A number of pseudoexperiments are generated to obtain the distribution of the test statistic

Experiment	Test Statistic used
LEP	$Q_\mu = -2\,ln\frac{\mathcal{L}(data\|\mu,\tilde{\theta})}{\mathcal{L}(data\|0,\tilde{\theta})}$
Tevatron	$Q_\mu = -2\,ln\frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|0,\hat{\theta}_0)}$
LHC	$Q_\mu = -2\,ln\frac{\mathcal{L}(data\|\mu,\hat{\theta}_\mu)}{\mathcal{L}(data\|\hat{\mu},\hat{\theta})}$





Upper limits set to $\sim$ 0.4 - 0.7 %


GitLab CI
Selection
eventtype: 15396400
polarity: magdown
year: 2015
Selection
eventtype: 90000000
polarity: magup
year: 2016
Selection
eventtype: 15396200
polarity: magup
year: 2016
Selection
eventtype: 15396400
polarity: magup
year: 2016
Selection
eventtype: 90000000
polarity: magdown
year: 2016
Selection
eventtype: 15396200
polarity: magdown
year: 2016
SelectionLcDstK
CombineTuples
sPlotforReweighting
KinematicReweighting
channel: MC_LcDstK
KinematicReweighting
channel: MC_LcDst2D0pi0K
KinematicReweighting
channel: MC_LcDst2D0gammaK
ECCMC
ECCData
ECCMC
LbMFitting
Efficiencies
hyp: pi0
Efficiencies
hyp: gamma
EffMapPlotting
hyp: pi0
EffMapPlotting
hyp: gamma
sPlotwECC
Moments
MomentsPlotting
max_order: 4
MomentsPlotting
max_order: 9
KernelDensityEstimation
LcDstFitting
fitfor: yield
fitmodel: kde+voigt
model: Pc4312
LcDstFitting
fitfor: frac
fitmodel: kde+voigt
model: Pc4440
LcDstFitting
fitfor: yield
fitmodel: kde+voigt
model: Pc4440
LcDstFitting
fitfor: frac
fitmodel: kde+voigt
model: Pc4457
LcDstFitting
fitfor: yield
fitmodel: kde+voigt
model: Pc4457
LcDstFitting
fitfor: frac
fitmodel: exp+res
model: Pc4312
LcDstFitting
fitfor: yield
fitmodel: exp+res
model: Pc4312
LcDstFitPlots
LimitSetting
teststat: plr
LimitSetting
teststat: slr
LimitsPlotting
bkgmodel: kde+voigt
LimitsPlotting
bkgmodel: exp+res
reana

Points on the workflow

- Snakemake was instrumental in setting up and developing the analysis
- Analysis had to be rerun many times $\mathcal{O}$ (100s)
- Snakemake allowed for easy parallelization of the analysis
- Snakemake allowed for easy integration of the analysis into REANA
- **xrootd** is important for this analysis and we are stuck on **snakemake 7.32.4**