(Re)discovery of the Doubly Charmed Baryon in LHCb Run-3 Data

Linnuo Zhang

University of Science & Technology of China

Supervised by Jibo He & Monica Pepe-Altarelli

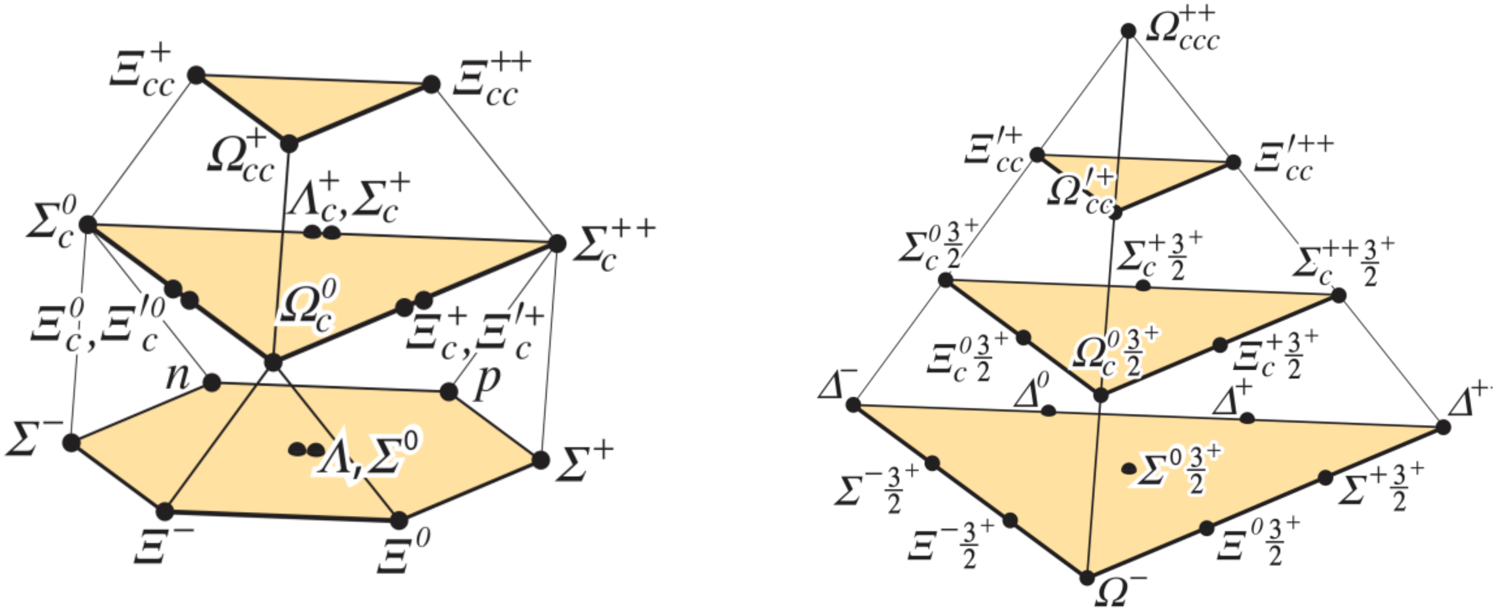
29 August 2024

Outline

- **Introduction**
- **Λ_c^+ at LHCb**
- **Rediscovery of Ξ_{cc}^{++} with 2024 data**
- **Search for $\Xi_{cc}^{+} \to \Lambda_c^+ K^- \pi^+$**
- **Summary**

Doubly charmed baryon

- **Doubly charmed baryons resemble hydrogen-like atoms.**
- **Simplify calculations and allow predictions about their characteristics.**

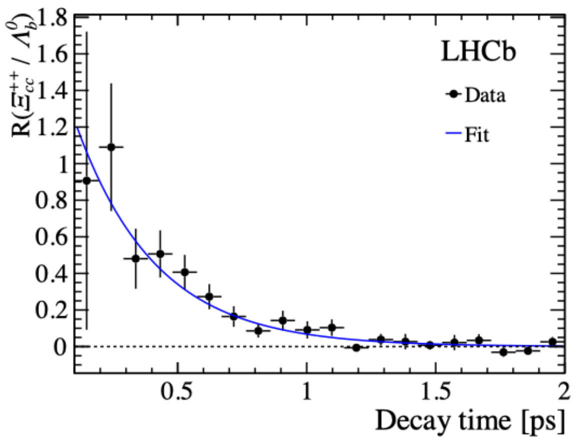


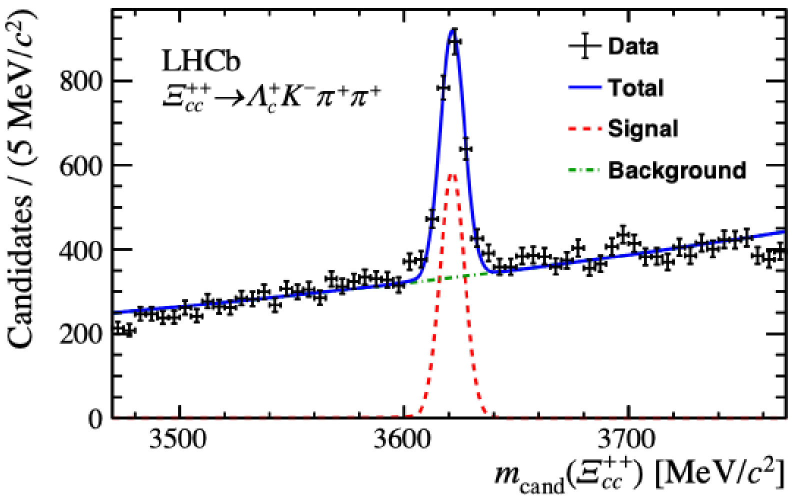
The $J^P = \frac{1}{2}^+$ (left) $J^P = \frac{3}{2}^+$ (right) flavour SU(4) ground-state charmed baryons.

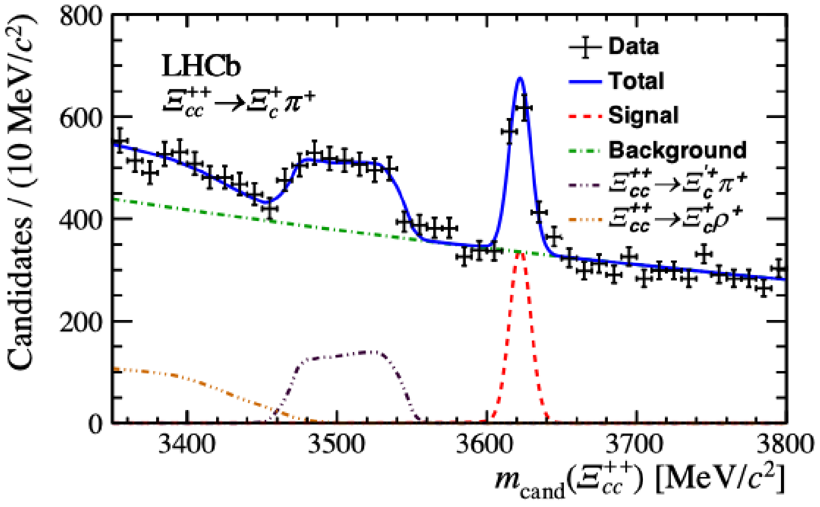
[Prog. Theor. Exp. Phys. 8 (2020) 083C01]

Ξ_{cc}^{++} at LHCb

- **Measurement of Ξ_{cc}^{++}**
  - Cross section $= (2.22 \pm 0.27 \pm 0.29) \times 10^{-4}$ [Chin.Phys. C 44 (2020) 022001]
  - $\tau(\Xi_{cc}^{++}) = 0.256^{+0.024}_{-0.022}(\text{stat}) \pm 0.014(\text{syst})$ ps [Phys. Rev. Lett. 121 (2018) no.5, 052002]
  - $m(\Xi_{cc}^{++}) = 3621.55 \pm 0.23(\text{stat}) \pm 0.30(\text{syst})\ \text{MeV}/c^2$ [Phys. Rev. Lett. 121 (2018) no.5, 052002]







N_{sig} =1598 $\qquad$ $N_{sig} = 616$

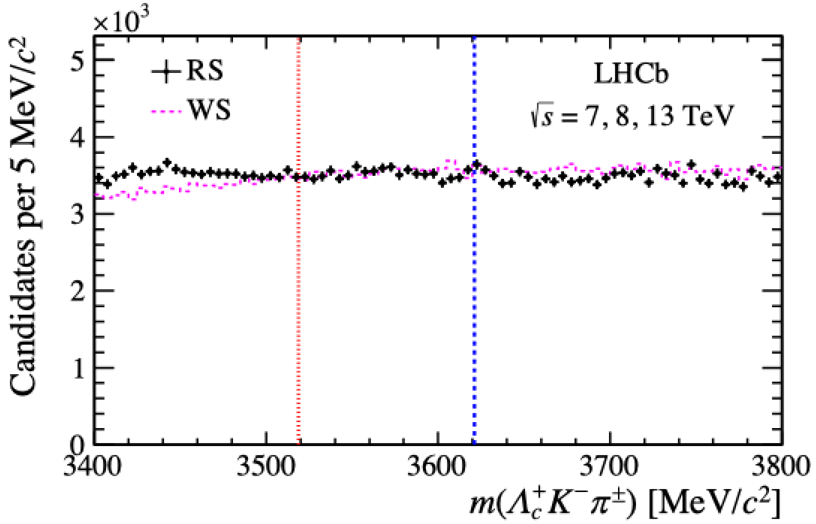
Prediction of Ξ_{cc}^{+} Lifetime

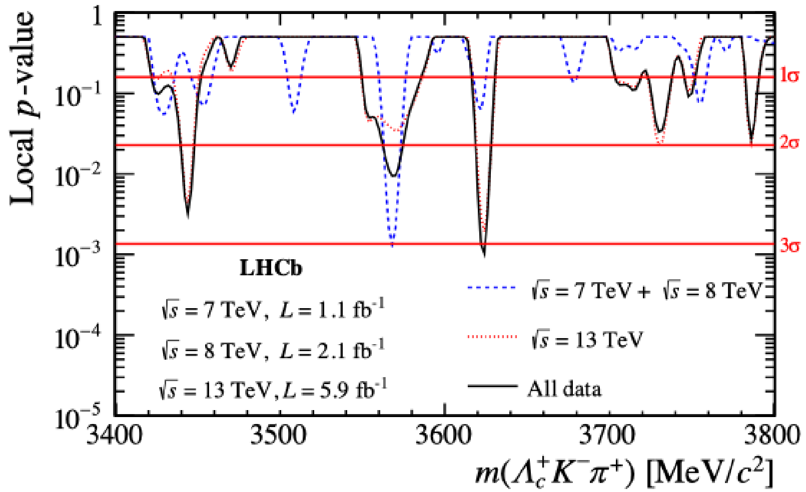
$\tau(\Xi_{cc}^{++})$ → Pauli interference effect & W-exchange effect → $\tau(\Xi_{cc}^{+}) \approx \frac{1}{3}\tau(\Xi_{cc}^{++})$

[Nucl. Phys. B 248 (1984) 261]

Ξ_{cc}^{+} Search, Present status (with Run1 & 2 data)

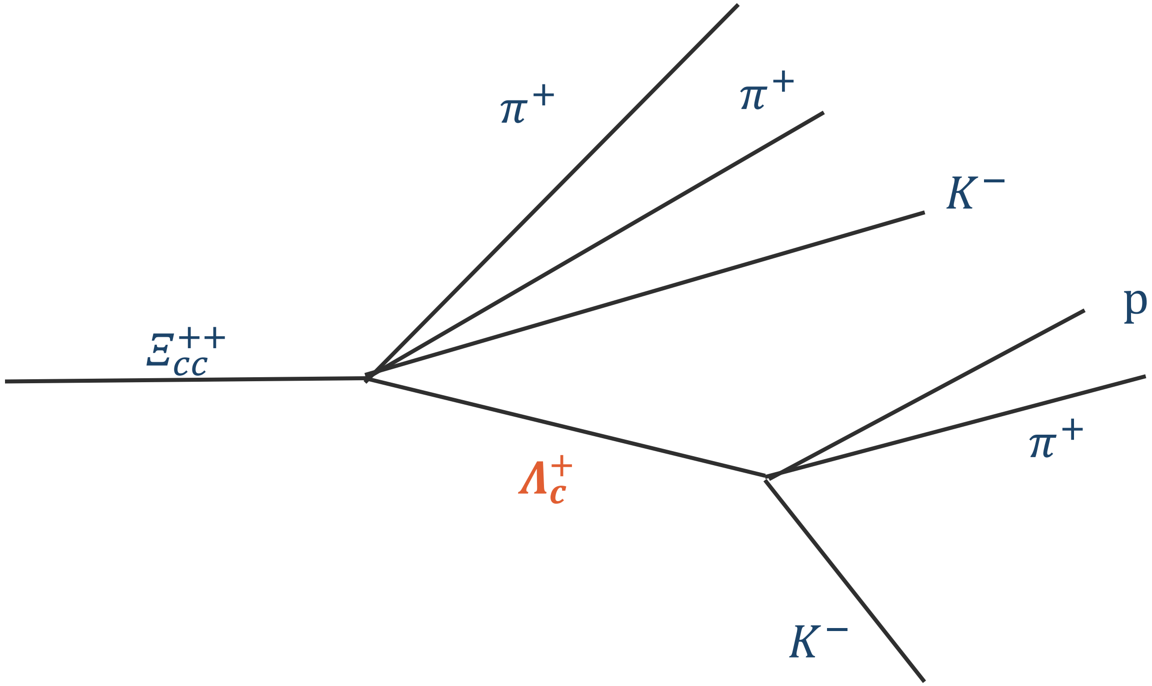
- **A bump around its isospin partner, Ξ_{cc}^{++}, with a local significance 3σ**



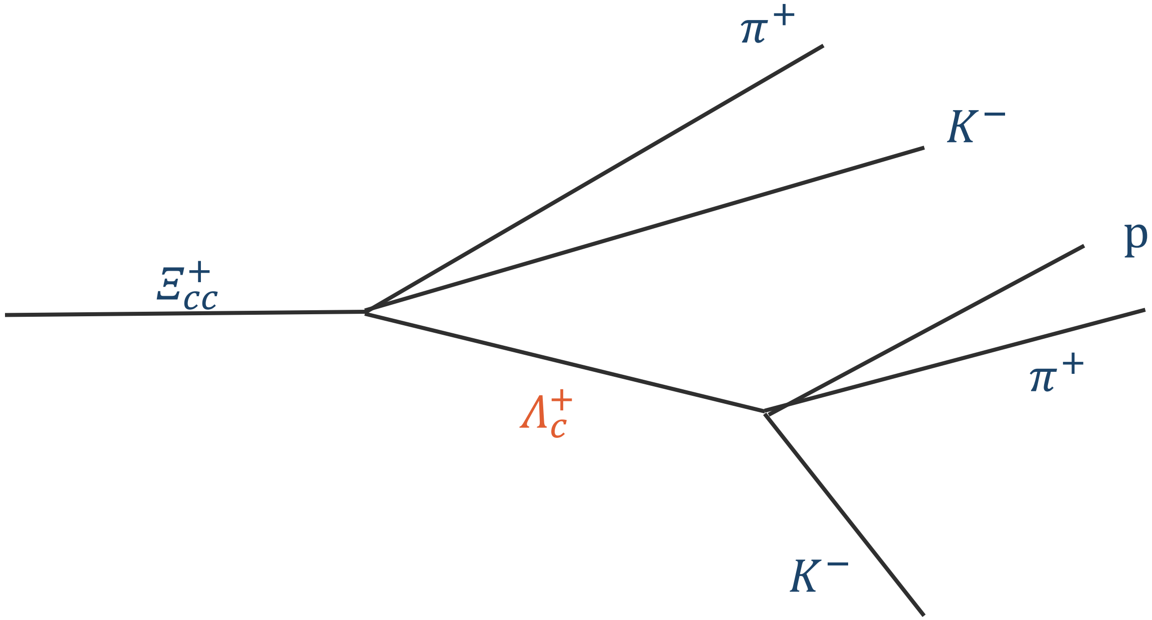


[Sci. China Phys. Mech. Astron. 63 (2020) no.2, 221062]

Why pay attention to Λ_c^+



$\Xi_{cc}^{++} \rightarrow \Lambda_c^+ K^- \pi^+ \pi^+$



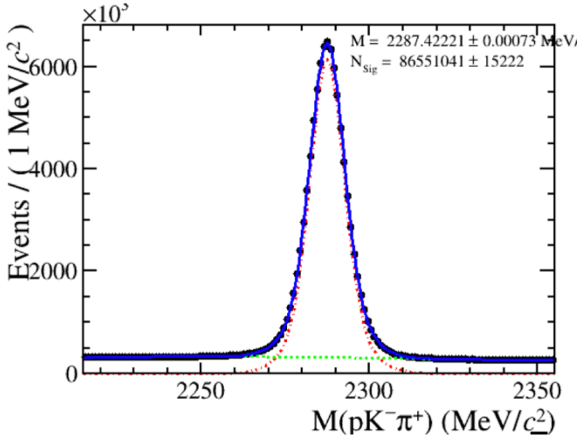
$\Xi_{cc}^{+} \rightarrow \Lambda_c^+ K^- \pi^+$

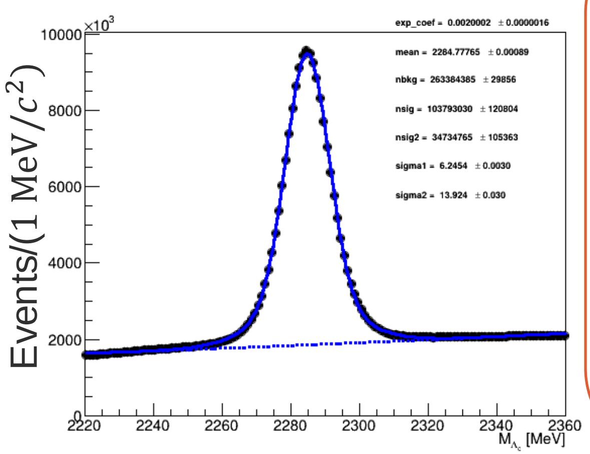
Two important channels for searching Ξ_{cc}, both including Λ_c^+

Λ_c^+ in LHCb Run3 data

- **Number of Λ_c^+ events in 2018 data: 8.6×10^7**
- **Number of Λ_c^+ events in part of 2024 data (before including UT): 1.3×10^8**

$\Lambda_c^+(\rightarrow pK^-\pi^+)$ mass distribution

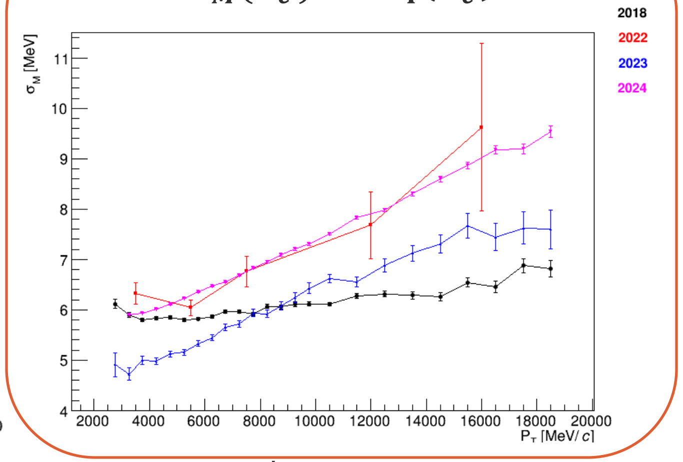




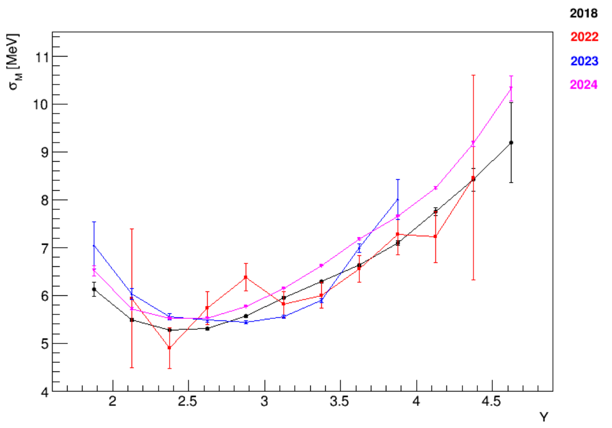
more $\Lambda_c^+ \rightarrow$ expect more Ξ_{cc}

$\Lambda_c^+(\rightarrow pK^-\pi^+)$ mass resolution (July 16)

$\sigma_M(\Lambda_c^+)$ VS $P_T(\Lambda_c^+)$

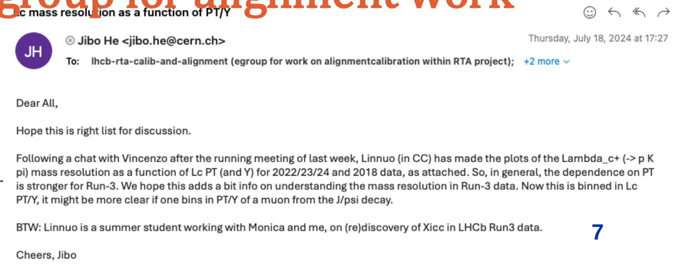


$\sigma_M(\Lambda_c^+)$ VS $y(\Lambda_c^+)$



$\sigma_M(\Lambda_c^+)$ has stronger dependence on $P_T(\Lambda_c^+)$ in Run3

Shared this result to egroup for alignment work


Lc mass resolution as a function of PT/Y

JH Jibo He <jibo.he@cern.ch> Thursday, July 18, 2024 at 17:27

To: lhcb-rta-calib-and-alignment (egroup for work on alignmentcalibration within RTA project); +2 more

Dear All,

Hope this is right list for discussion.

Following a chat with Vincenzo after the running meeting of last week, Linnuo (in CC) has made the plots of the Lambda_c+ (-> p K pi) mass resolution as a function of Lc PT (and Y) for 2022/23/24 and 2018 data, as attached. So, in general, the dependence on PT is stronger for Run-3. We hope this adds a bit info on understanding the mass resolution in Run-3 data. Now this is binned in Lc PT/Y, it might be more clear if one bins in PT/Y of a muon from the J/psi decay.

BTW: Linnuo is a summer student working with Monica and me, on (re)discovery of Xicc in LHCb Run3 data.

Cheers, Jibo

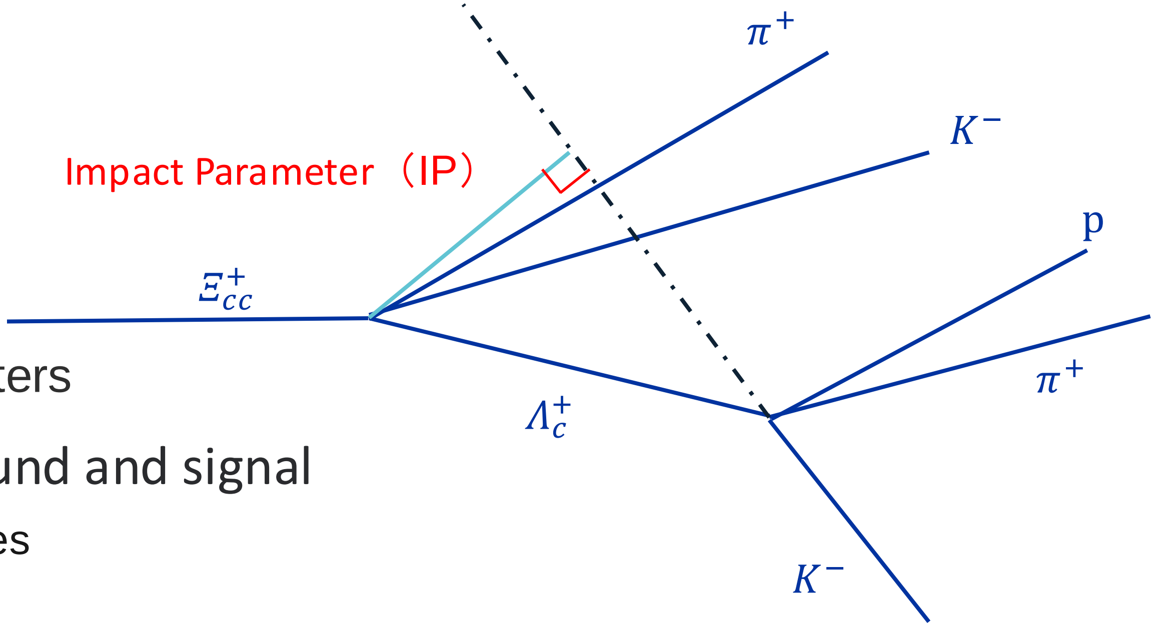
Workflow for extremely large amount of data

Without saving a new Root file

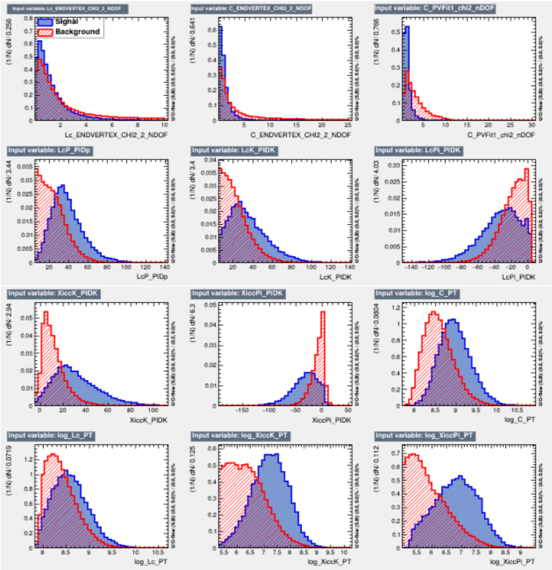
- **Use RDataFrame**
  - Deploy preselection cut
  - Define new branches
  - Get MVA response
- **NO root file saving before the final step**

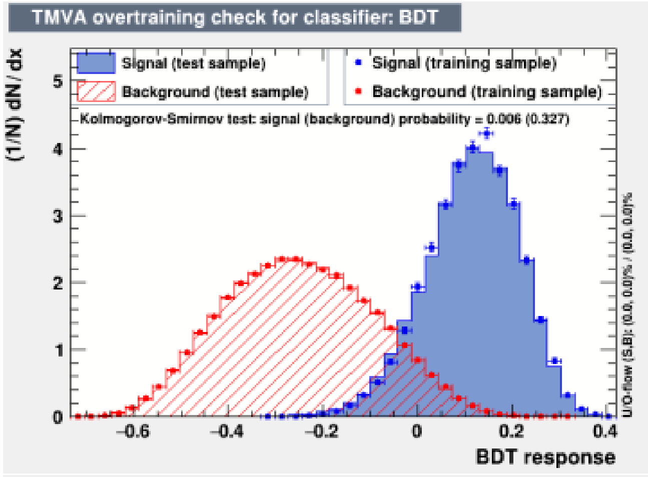
```python
import ROOT
import numpy as np
from apd import AnalysisData
import numpy

ROOT.EnableImplicitMT()

lineName = "XiccpToLcpKmPip_LcpToPpKmPip"
treeName = lineName + "/DecayTree"


# Get all Tuples from Analysis production
datasets = AnalysisData( "charm", "charm_baryon_spec_2024_data", polarity=["magdown","magup"])
paths    = datasets( eventtype=94000000, datatype=2024)
rdf      = ROOT.RDataFrame( treeName, paths )
model = ROOT.TMVA.Experimental.RReader("XiccP2LcKpi_BDTG_v1.xml")
variables = ["float_Hc_VCHI2DOF","float_Hcc_VCHI2DOF","Hcc_DTF_PV_CHI2_Hcc_DTF_PV_NDOF",\
             "HcPp_PID_P","HcKm_PID_K","HcPip_PID_K","HccKm_PID_K","HccPip_PID_K",\
             "ACos_Hcc_BPVDIRA","LcP_P_LcP_P_LcK_P_LcPi_P"]#variables used for MVA, for showing the workflow

newT = rdf.Define("Xicc_ConstM", "(Hcc_DTF_PV_MASS-Hc_M+2286.46)")
aCut = newT.Filter("HcKm_TRCHI2DOF<3 &HcPp_TRCHI2DOF<3 & HcPip_TRCHI2DOF<3 &"\
                   "HcPp_P>1000 &HcKm_P>1000&HcPip_P>1000& HcPp_PT>200&HcPip_PT>200&HcKm_PT>200 & HcPp_PT+HcKm_PT+HcPip_PT>3000 &"\
                    "TMath::Max(HcPp_PT, TMath::Max(HcKm_PT,HcPip_PT))>1000& ")#preselection just for showing the workflow
rdf_addnew = aCut.Define("float_Hc_VCHI2DOF","float(Hc_VCHI2DOF)").Define("float_Hcc_VCHI2DOF","float(Hcc_VCHI2DOF)")\
            .Define("Hcc_DTF_PV_CHI2_Hcc_DTF_PV_NDOF","float(Hcc_DTF_PV_CHI2/Hcc_DTF_PV_NDOF)")\
            .Define("log_Hcc_PT","log(Hcc_PT)").Define("log_Hc_PT","log(Hc_PT)").Define("log_HccKm_PT","log(HccKm_PT)")\
            .Define("log_HccPip_PT","log(HccPip_PT)").Define("log_HcPp_PT","log(HcPp_PT)").Define("log_HcKm_PT","log(HcKm_PT)")
df = rdf_addnew.Define('BDTG_value',ROOT.TMVA.Experimental.Compute[23, "float"](model),variables).Filter("BDTG_value[0]>0.94")
df.Snapshot("DecayTree", "afterMVAcut.root")
```

Workflow: $\Xi_{cc}^{++} \rightarrow \Lambda_c^+ K^- \pi^+ \pi^+$



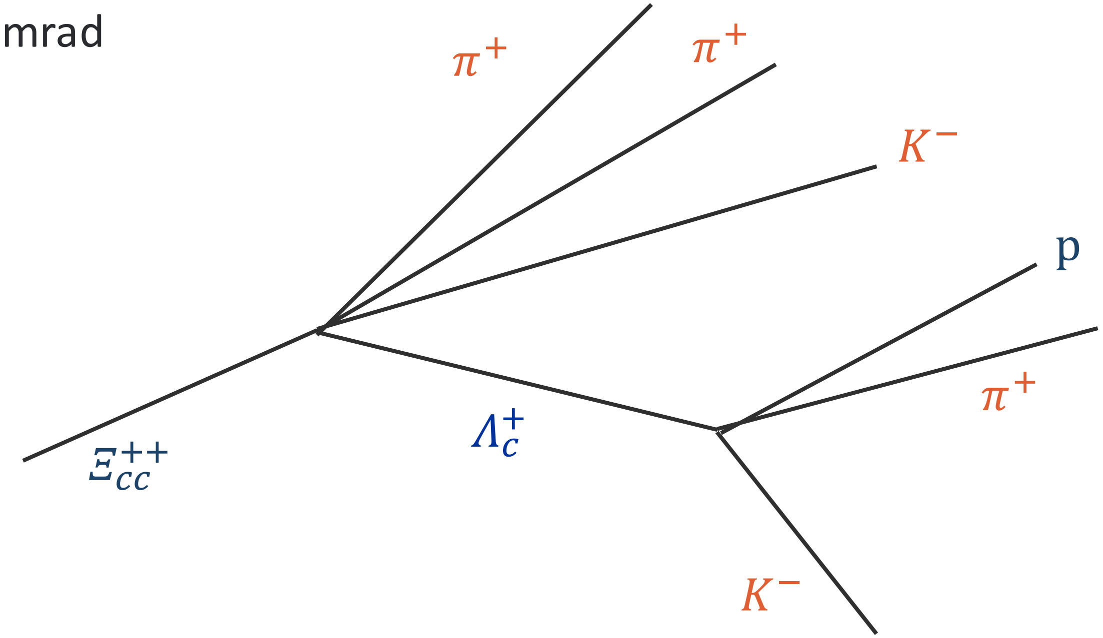
- Loose pre-selection, reduce background
  - According to features (like IP, P_T) of Ξ_{cc}^{++} and its daughters
- MVA training, use MVA response to separate background and signal
  - A ML method: exploiting the correlation between input variables
  - Get a better signal-to-background separation.





Workflow: $\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$

- Loose pre-selection, reduce background
- MVA training, use MVA response to separate background and signal
- Remove clone and duplicate candidates
  - Clone candidates
    - A track of a particle from Λ_c^+ may be reconstructed twice, and the clone one is used to reconstruct Ξ_{cc}^{++}.
    - Require the angle between the momenta of two tracks > 0.5 mrad
  - Duplicate candidates
    - 3 π^+ and 2 K^- in the whole decay process
    - π^+ / K^- from Λ_c^+ swap with that from Ξ_{cc}^{++}
    - If two (or more) Ξ_{cc}^{++} candidates share six pairs of identical final-state tracks, randomly keep one candidate.



Workflow: $\Xi_{cc}^{++} \rightarrow \Lambda_c^+ K^- \pi^+ \pi^+$

- Loose pre-selection, reduce background
- MVA training, use MVA response to separate background and signal
- Remove clone and duplicate candidates
- MVA retraining:
  - Use Ξ_{cc}^{++} signal in data as training samples
  - Use Run-3 MC samples to train a new model

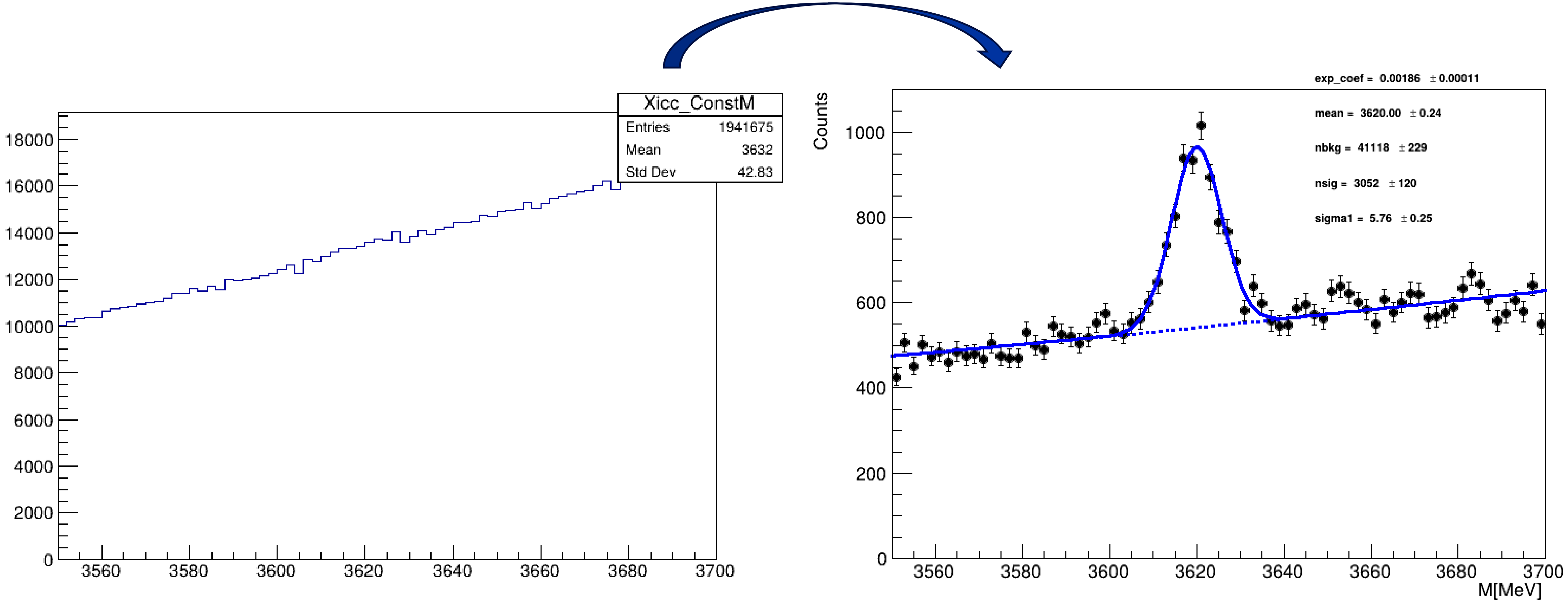
The same workflow is also used for $\Xi_{cc}^{++} \rightarrow \Xi_c^+ \pi^+, \Xi_{cc}^+ \rightarrow \Lambda_c^+ K^- \pi^+$

Expect less background, more signal, better resolution

$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$, Run-3 dataset

Apply MVA cut to preselected data

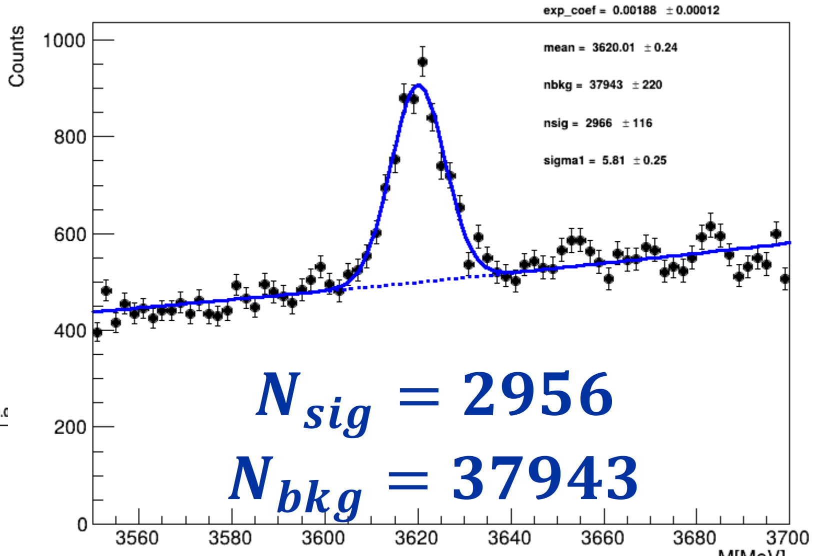
MVA model trained by Run-2 MC samples



$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$, Run-3 dataset

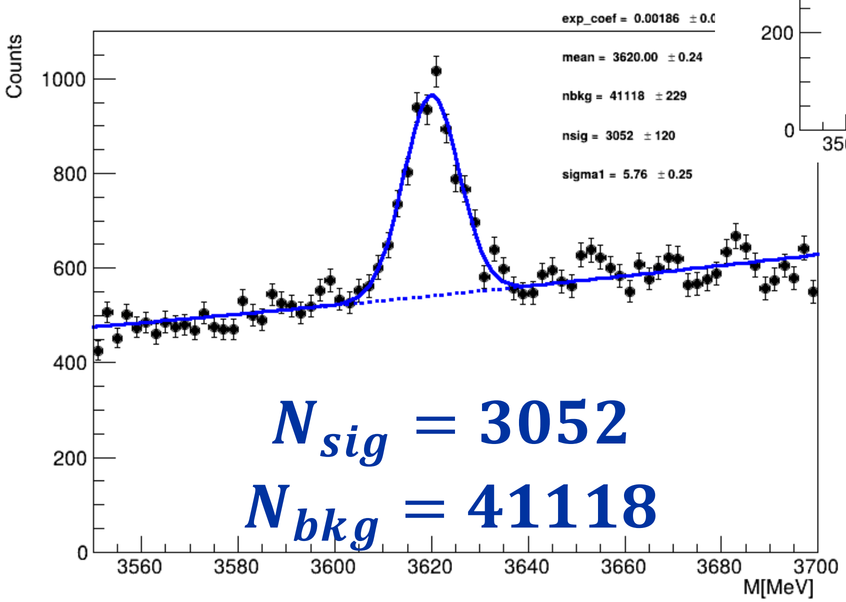
Remove duplicated and clone candidates

Remove Clones

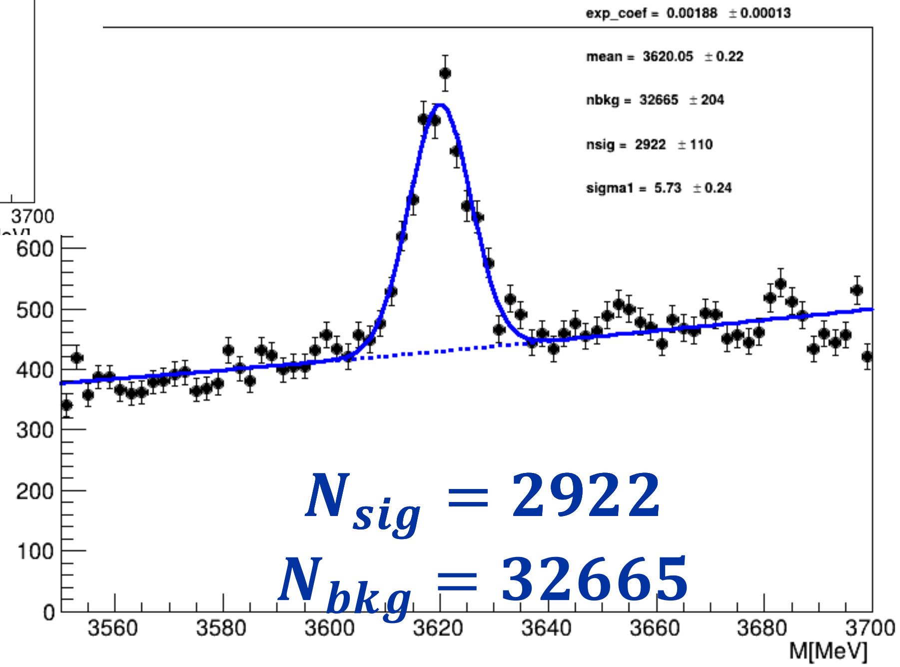


$N_{sig} = 2956$
$N_{bkg} = 37943$

Remove Duplications



$N_{sig} = 3052$
$N_{bkg} = 41118$



$N_{sig} = 2922$
$N_{bkg} = 32665$

$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$, Run-3 dataset

MVA Re-Training 1:

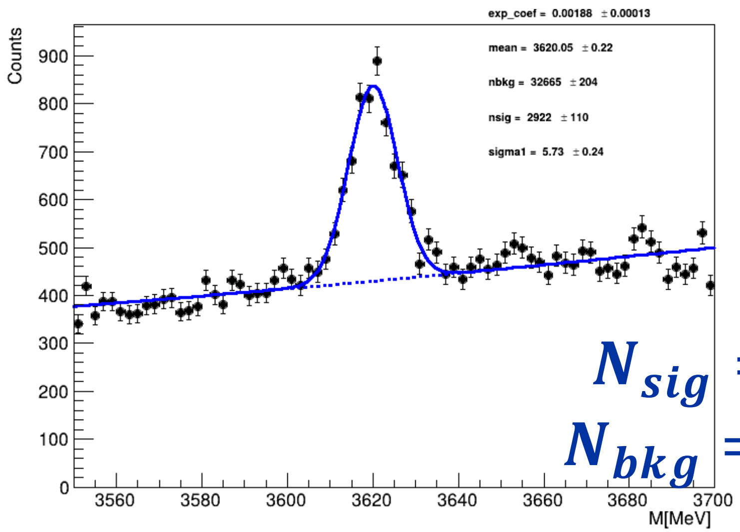
- Use signal in data as training sample
- Gradient Boosted Decision Tree (BDTG) method

Variables selection:

Ξ_{cc}^{++} : $P_{\mathrm{T}}, X_{\mathrm{FD}}^2, X_{\mathrm{IP}}^2$ ...

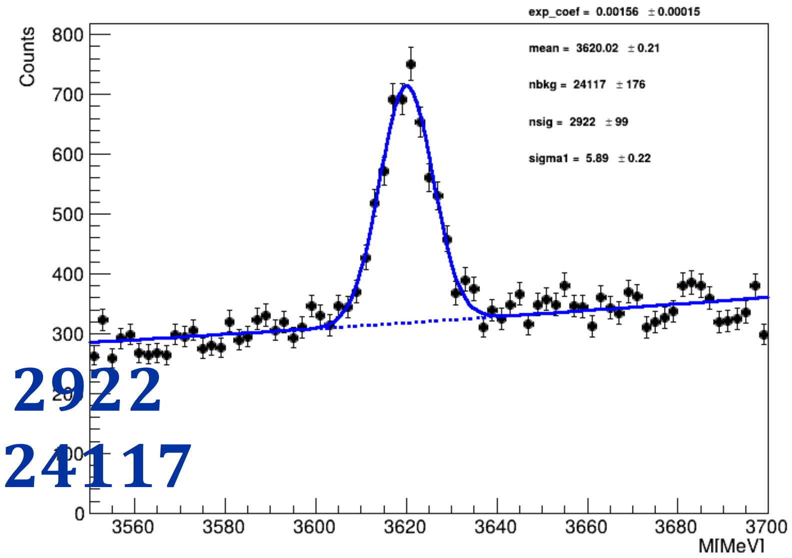
Ξ_{cc}^{++} daughters: PID, P_T ...

Before using Ξ_{cc}^{++} sweight in MVA training



$N_{sig} = 2922$
$N_{bkg} = 32665$

After using Ξ_{cc}^{++} sweight in MVA training



$N_{sig} = 2922$
$N_{bkg} = 24117$

$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$, Run-3 dataset

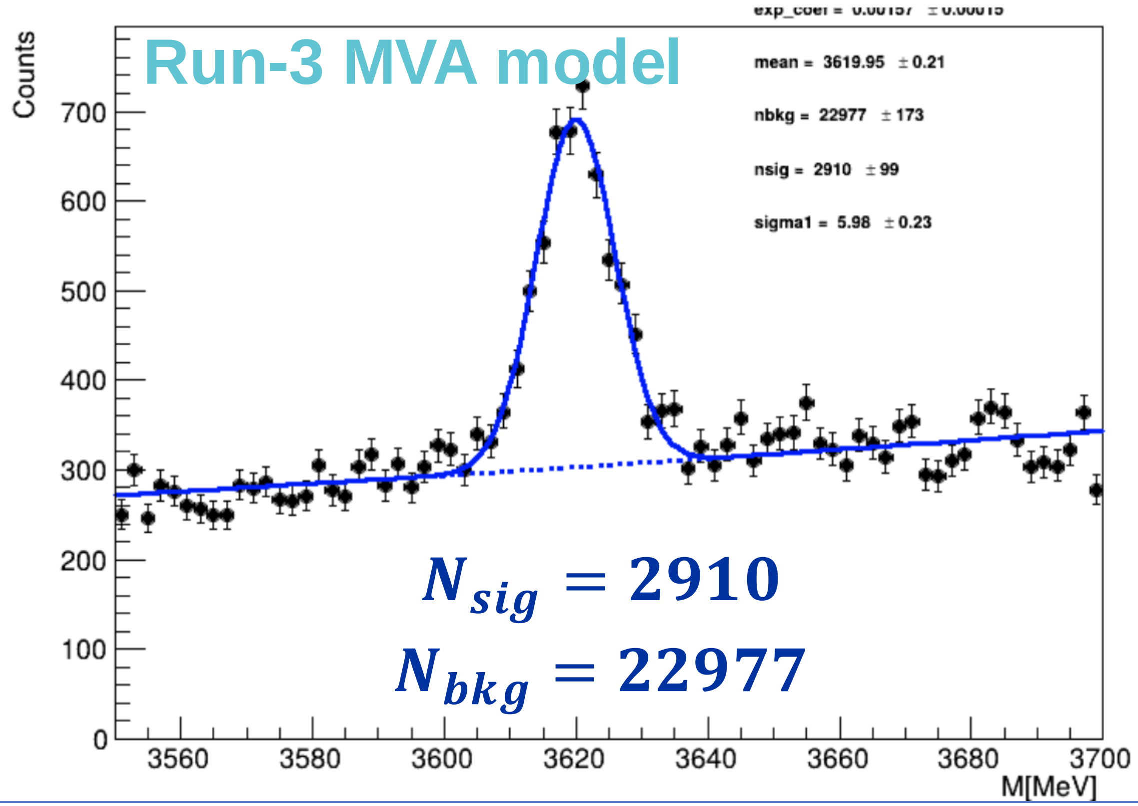
MVA Re-Training 2:

- Signal samples: Run-3 MC samples
- Background samples: WS samples
- Gradient Boosted Decision Tree (BDTG) method

Variables selection:

Ξ_{cc}^{++} : $P_{\mathrm{T}}, X_{\mathrm{FD}}^2, X_{\mathrm{IP}}^2$ ...

Ξ_{cc}^{++} daughters: PID, P_T ...

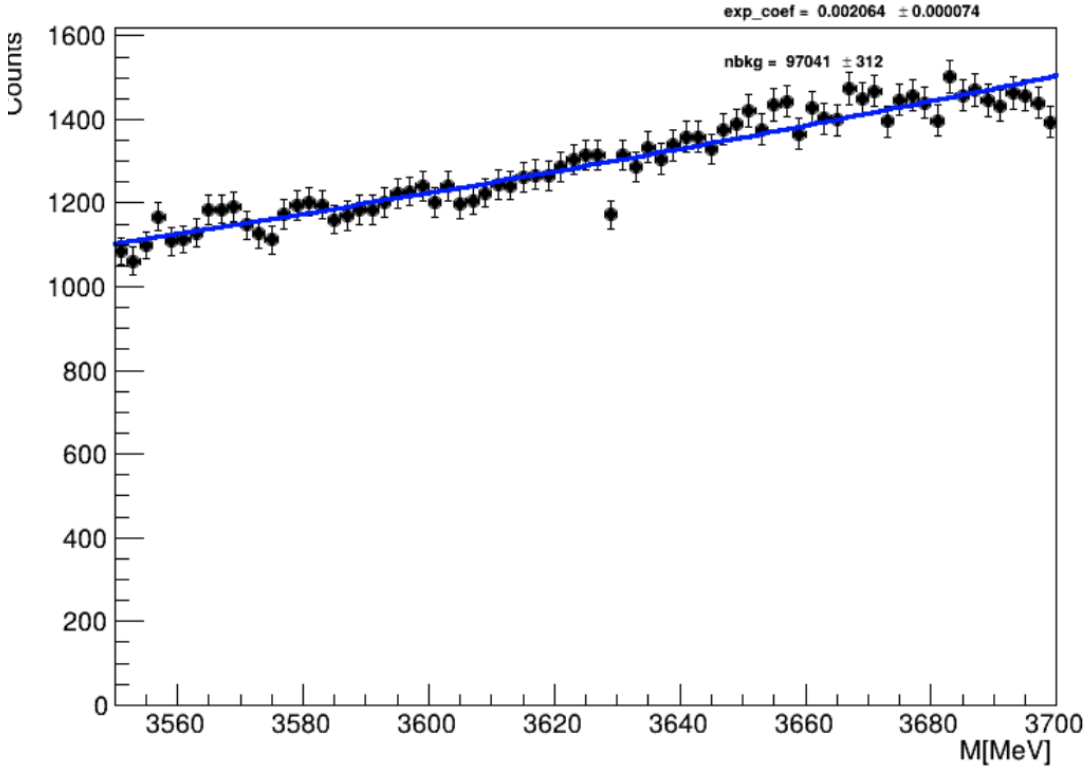


$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$, Run-3 dataset

WS channel check
MVA based:

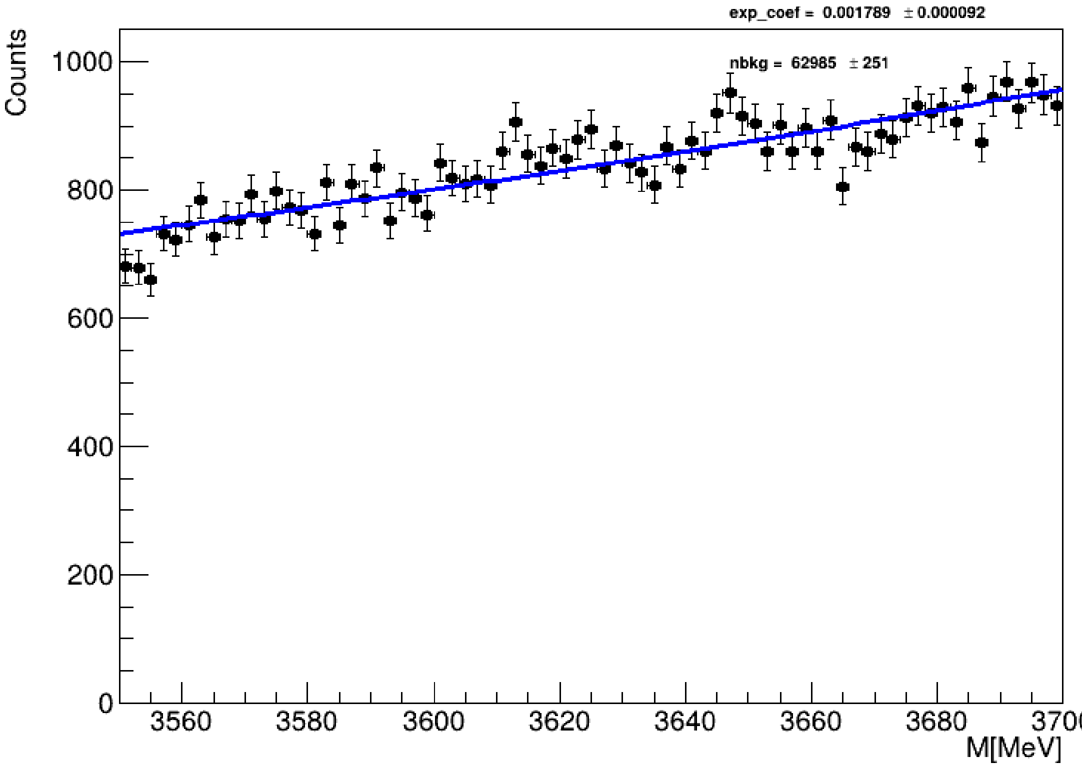
WS1:

Final state: $\Lambda_c^+ K^- \pi^- \pi^+$



WS2:

Final state: $\Lambda_c^+ K^+ \pi^- \pi^+$



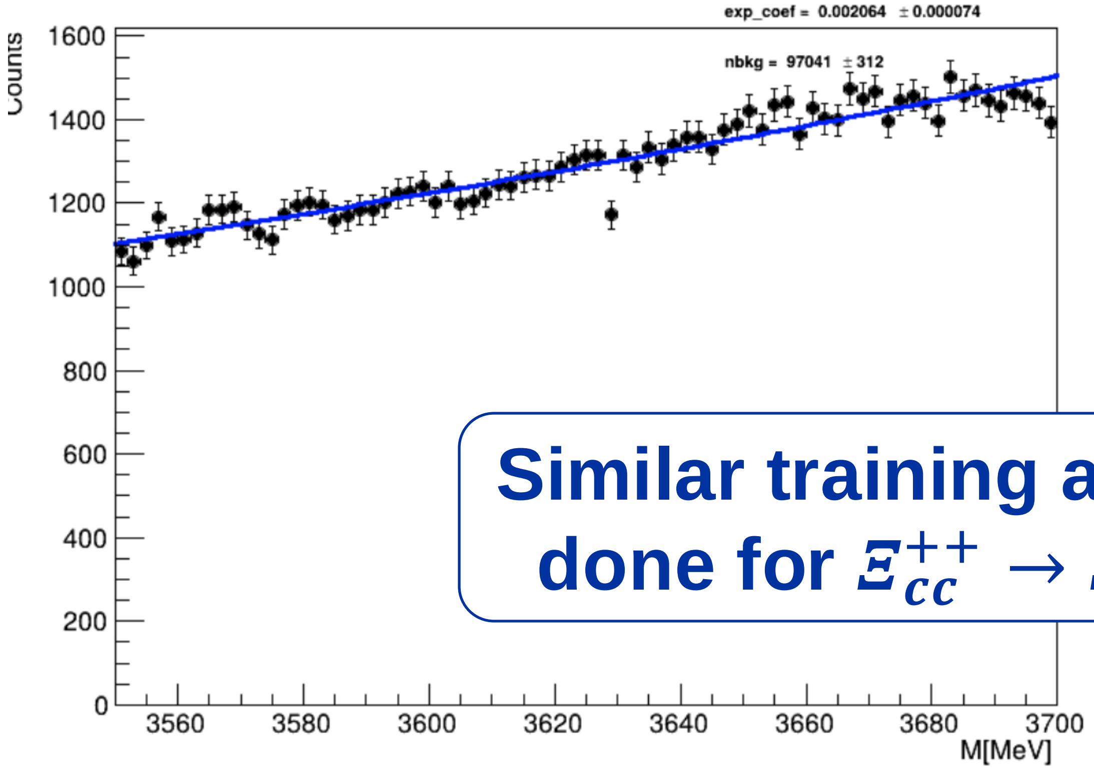
MVA does not cause fake peak in WS samples

$\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^+$, Run-3 dataset

WS channel check MVA based:

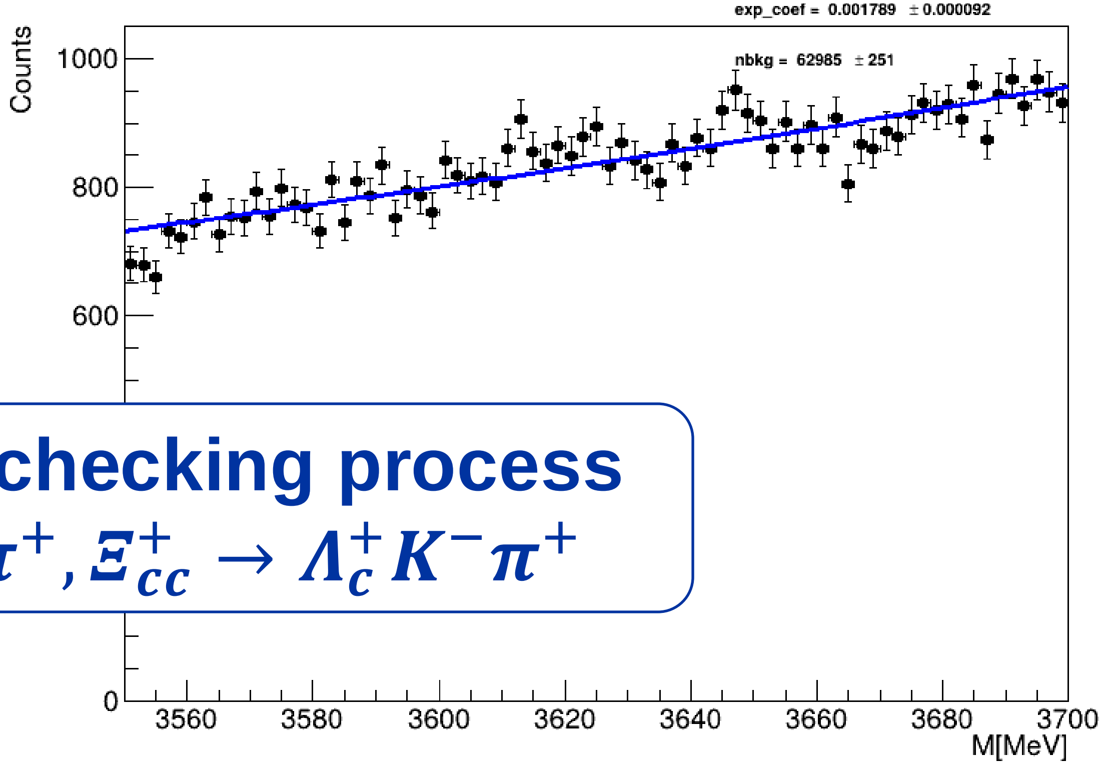
WS1:

Final state: $\Lambda_c^+ K^- \pi^- \pi^+$



WS2:

Final state: $\Lambda_c^+ K^+ \pi^- \pi^+$



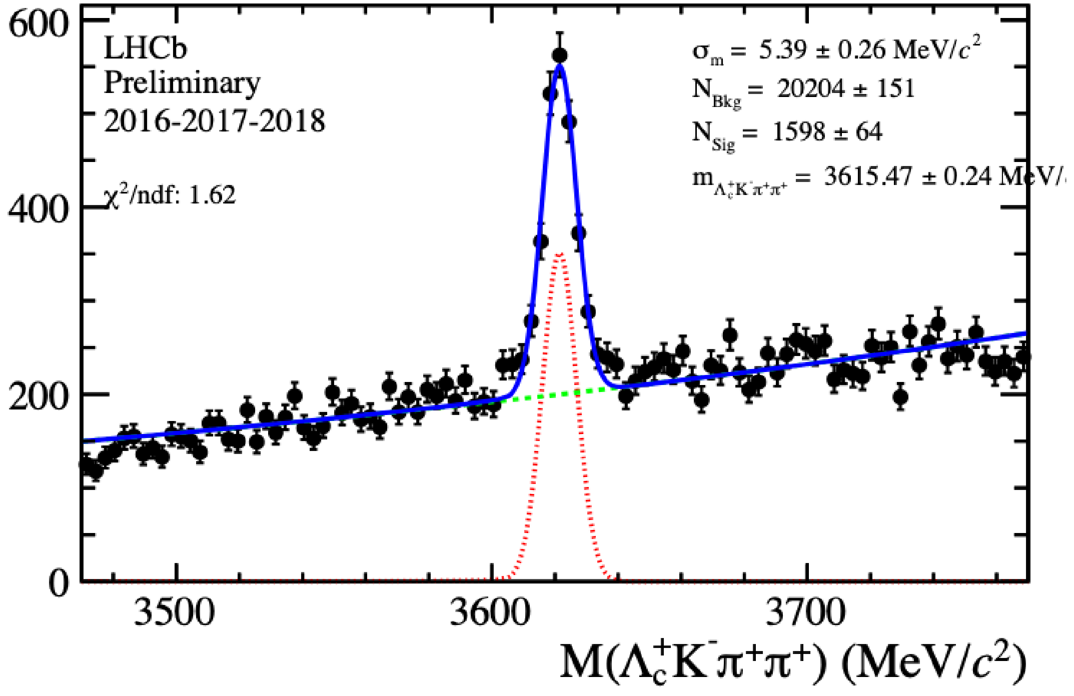
Similar training and checking process done for $\Xi_{cc}^{++} \to \Xi_c^+ \pi^+, \Xi_{cc}^+ \to \Lambda_c^+ K^- \pi^+$

MVA does not cause fake peak in WS samples

$\Xi_{cc}^{++} \to \Lambda_c^{+} K^{-} \pi^{+} \pi^{+}$

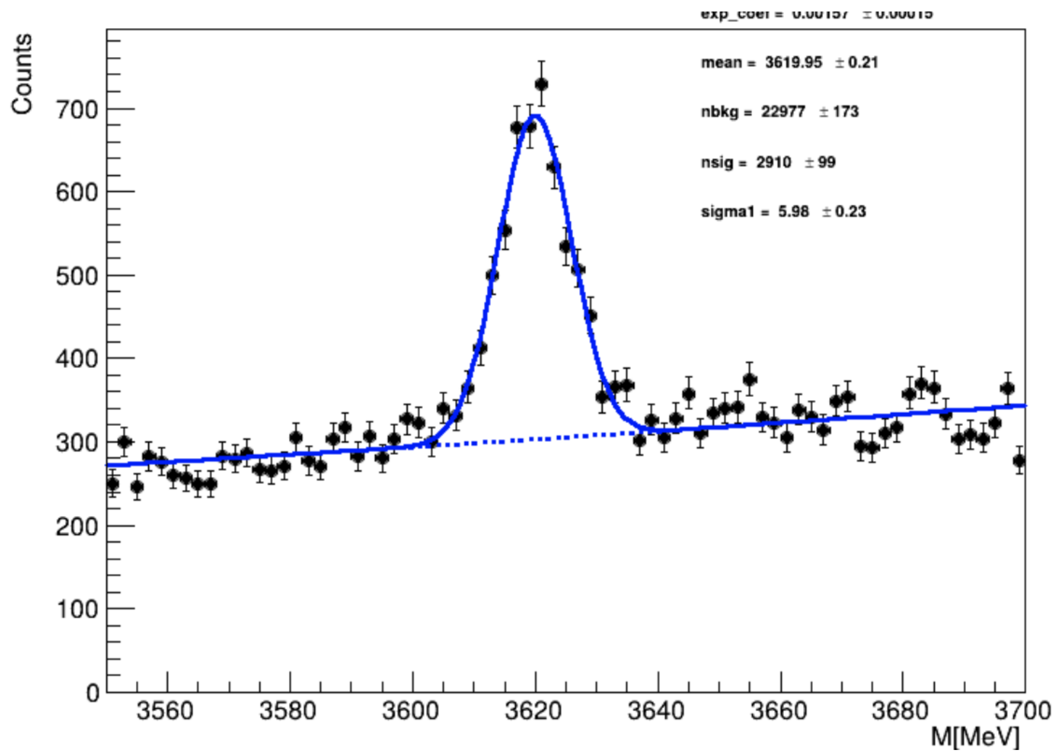
MVA based, Compared present & previous results

2016-2018 data



N_{sig} =1598

Part of 2024 data with &without UT



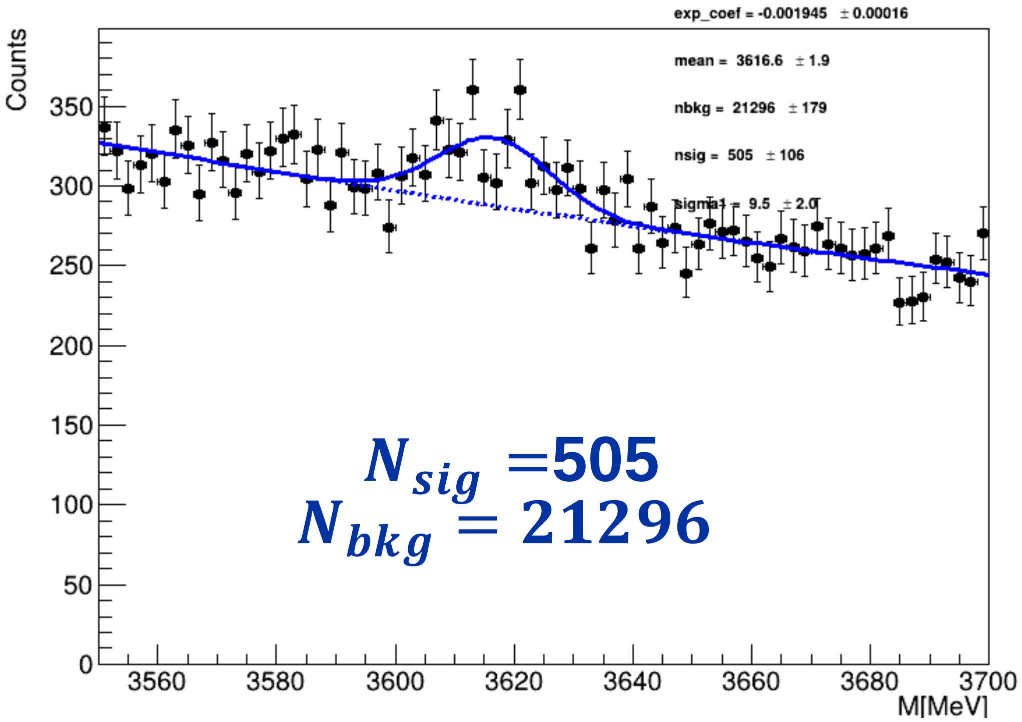
$N_{sig} = 2910$

Almost twice the signals in LHCb part 2024 Data as in 2016-2018 data

$\Xi_{cc}^{++} \to \Xi_c^{+}\pi^{+}$
MVA based, Compared present & previous results

Part of 2024 data without UT

Use Run-2 MVA cut

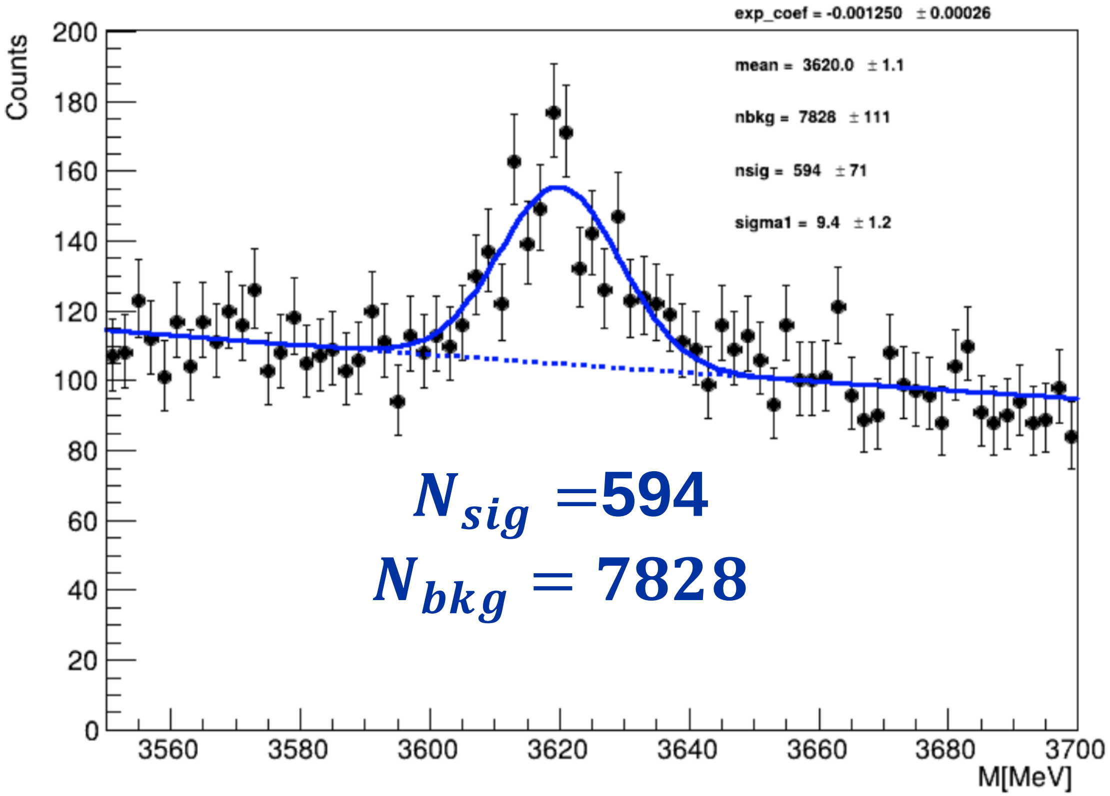


N_{sig} =505
N_{bkg} = 21296

Part of 2024 data without UT

Use MVA cut

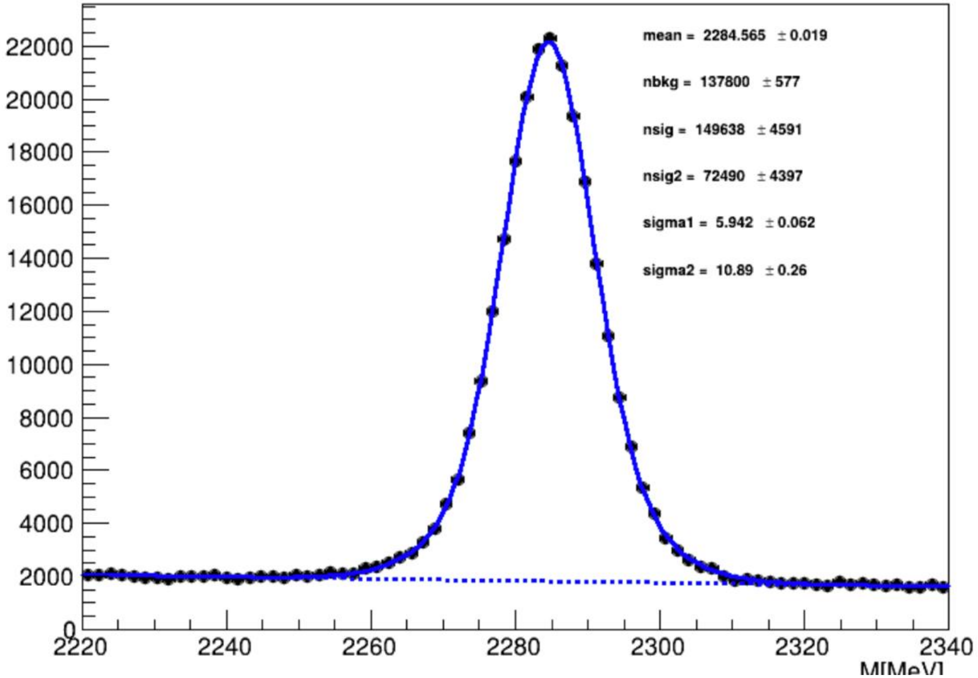
trained by 2024 MC and WS samples



N_{sig} =594
N_{bkg} = 7828

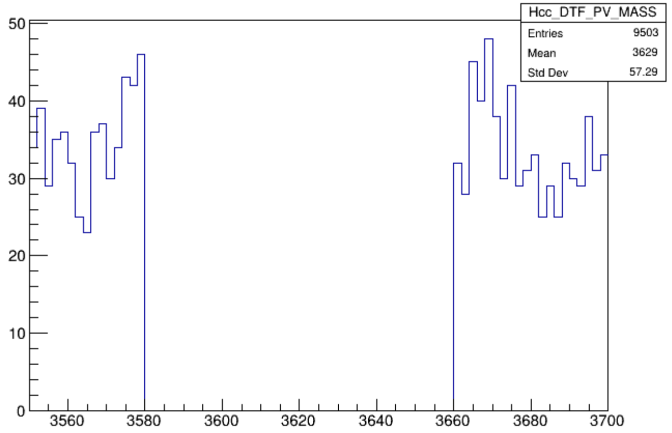
$\Xi_{cc}^{+} \to \Lambda_c^{+} K^{-} \pi^{+}$
MVA based:

Λ_c^{+} mass distribution(RS)



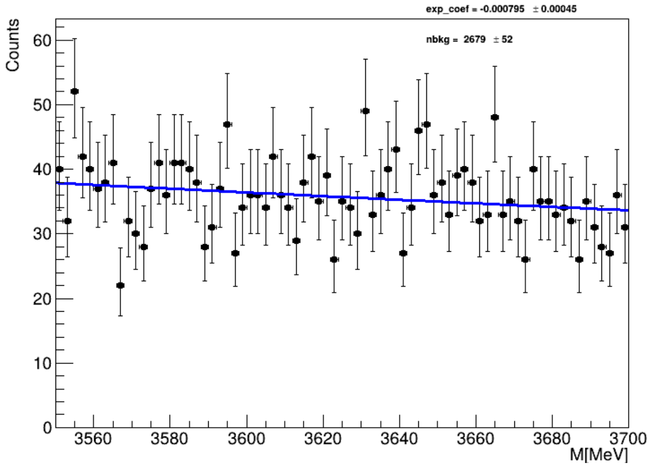
N_{sig} =235897

Blinded Signal Window (RS)



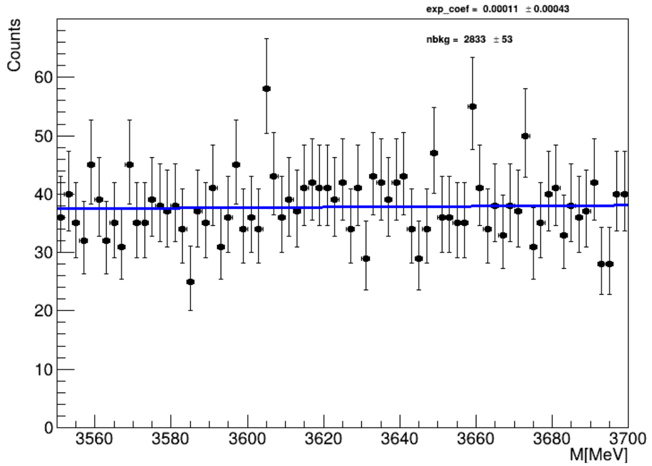
WS1:

Final state: $\Lambda_c^{+} K^{-} \pi^{-}$



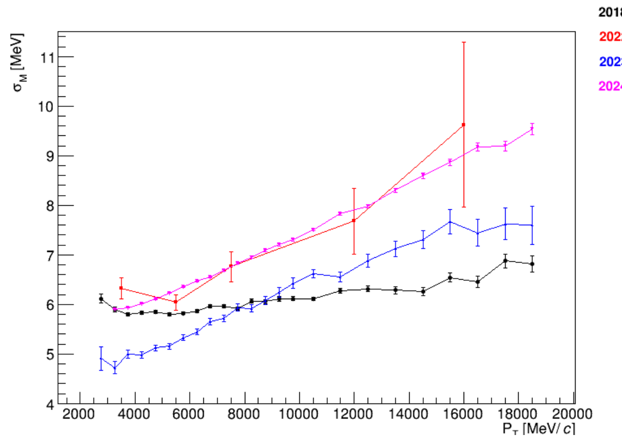
WS2:

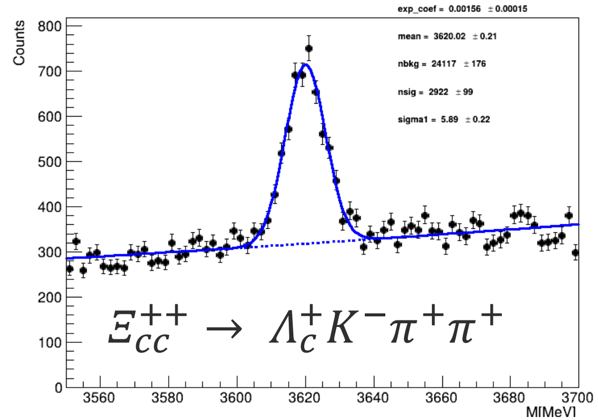
Final state: $\Lambda_c^{+} K^{+} \pi^{-}$

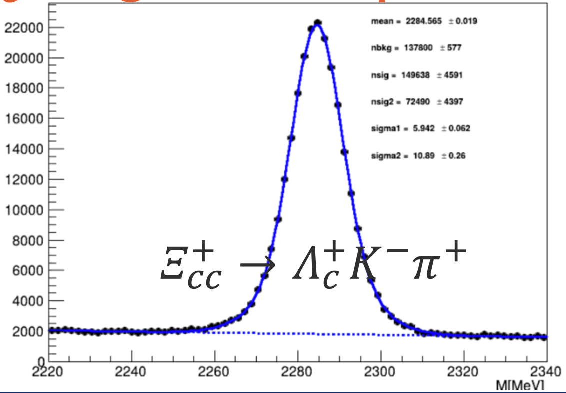


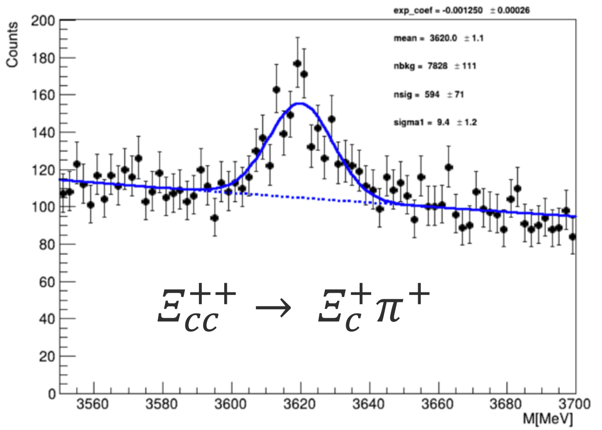
Summary

- Studied Λ_c^+ mass resolution as functions of P_T, y in 2024 sent to alignment group
- Studied and optimized event selection for:
  - $\Xi_{cc}^{++} \rightarrow \Lambda_c^+ K^- \pi^+ \pi^+$, twice signals in part of 2024 data, compared to signals in Run 2
  - $\Xi_{cc}^{++} \rightarrow \Xi_c^+ \pi^+$, better sig-to-bkg ratio with new MVA model
  - $\Xi_{cc}^{+} \rightarrow \Lambda_c^+ K^- \pi^+$, blind analysis, more Λ_c^+ signal, expect 1st observation of Ξ_{cc}^{+} w/ 2024 data



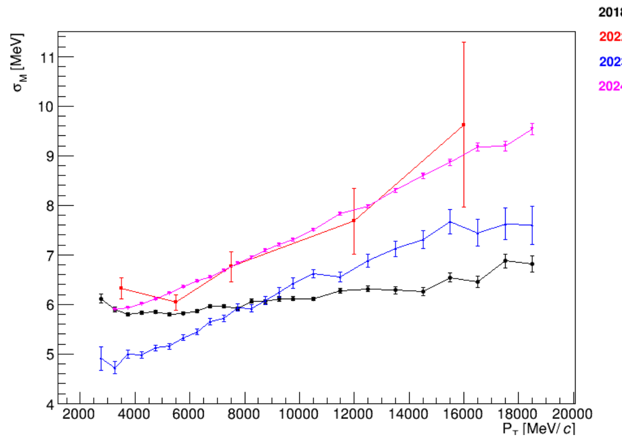




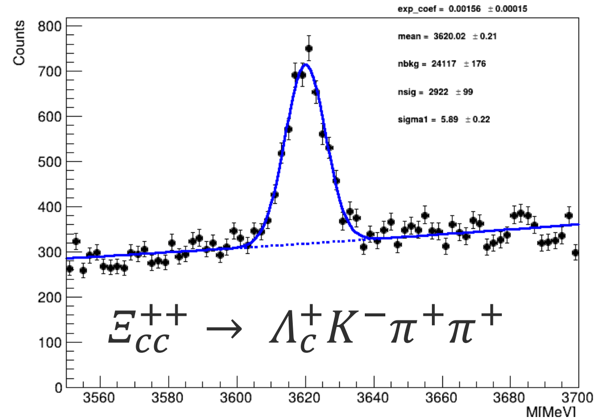


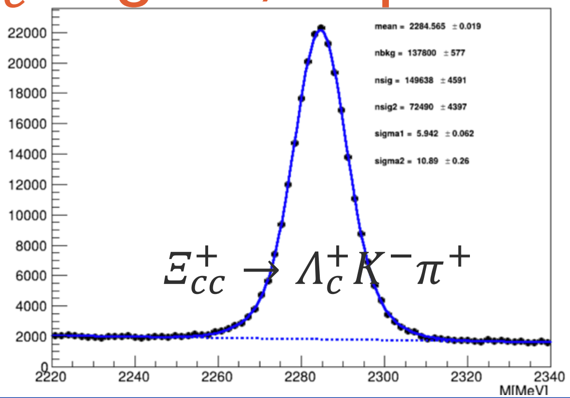
Summary

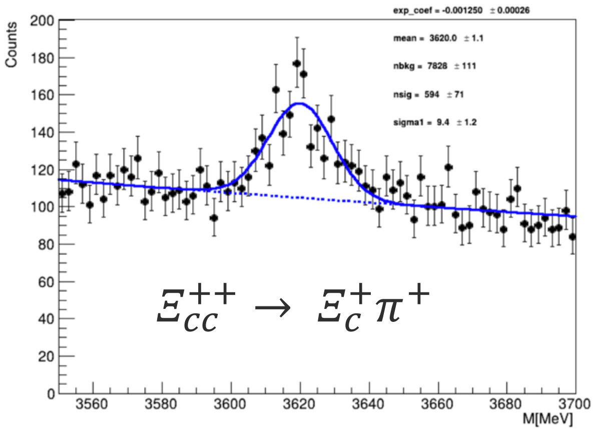
- Studied Λ_c^+ mass resolution as functions of P_T, y in 2024 sent to alignment group

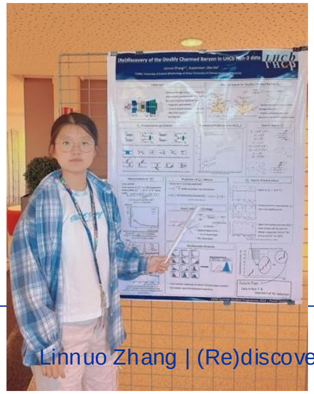


- Studied and optimized event selection for:
  - $\Xi_{cc}^{++} \rightarrow \Lambda_c^+ K^- \pi^+ \pi^+$, twice signals in part of 2024 data, compared to signals in Run 2
  - $\Xi_{cc}^{++} \rightarrow \Xi_c^+ \pi^+$, better sig-to-bkg ratio with new MVA model
  - $\Xi_{cc}^{+} \rightarrow \Lambda_c^+ K^- \pi^+$, blind analysis, more Λ_c^+ signal, expect 1st observation of Ξ_{cc}^+ w/ 2024 data









Thank you!