



# Sexaquark search in $\Xi_b^0$ decays

CERN Summer Student Report

**Sam Deacon**

Supervisors: María Vieites Díaz, Patrick Koppenburg

Abstract

The nature of dark matter is an unresolved question in particle physics, with the lack of candidates remaining one of the most intriguing mysteries of modern physics. It has been proposed that the triply singlet colour-flavour-spin *uuddss* state of six quarks (sexaquark) could be very tightly bound, possibly even stable or incredibly long-lived. Decays of $\Xi_b^0$ baryons have been identified as decay channels in which the sexaquark could be produced and influence the missing mass distribution in the decay. Here, we search for $\Xi_b^0\ (usb) \to 2\ \Lambda^0\ (uds) + K^+ + \bar{p}$ decays where the two $\Lambda^0\ (uds)$ baryons in the final state have the same quark content as the sexaquark candidate.

# Contents

# 1 Introduction

This section gives some context for the project and provides motivation for studying $\Xi_b^0$ decays to search for sexaquarks. It also outlines the primary objectives of the analysis.

## 1.1 The LHCb experiment

The Large Hadron Collider (LHC) is the world's largest particle accelerator, located on the border between France and Switzerland in an underground tunnel 27km in circumference. The Large Hadron Collider beauty (LHCb) experiment is a detector designed specifically for detection of decays of b- and c-hadrons.

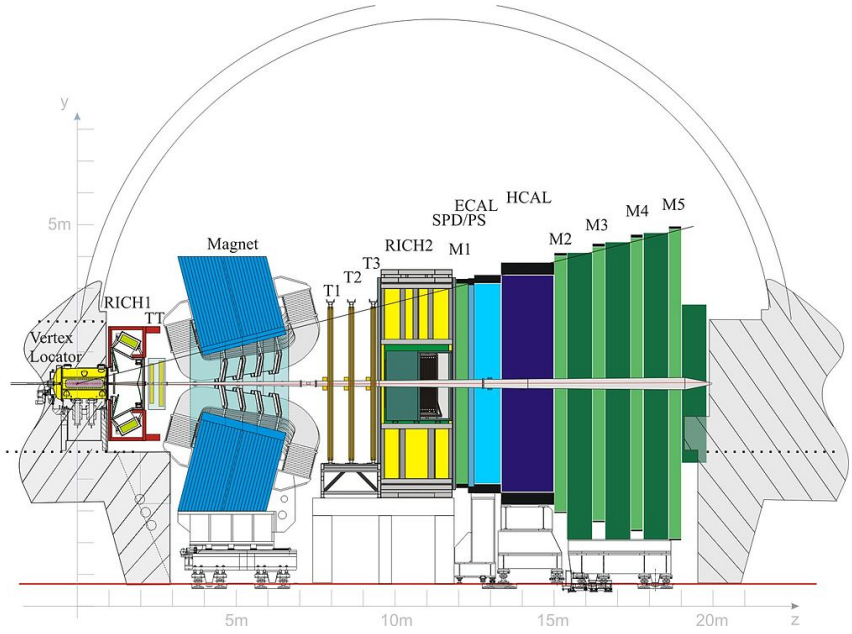


Figure 1: Illustration of the LHCb detector. Image taken from [1].

Fig. 1 shows a side view of the detector. Proton-proton collisions take place inside the VELO where the particle trajectories close to the interaction point are measured. The particles then travel into the Ring Imaging Cherenkov detector (RICH I) where hadron particle identification (PID) occurs. After this the particles pass through the Tracker Turicensis (TT) a strong electromagnet and more tracking stations before reaching the second RICH detector and the calorimeters and finally the muon chambers. Here, we are largely concerned with the measurements made in the VELO, RICH and

the trackers as we use kinematic variables such as the transverse momentum (PT) of the particles to train a multivariate analysis algorithm and make cuts on PID variables to help discriminate between signal and background.

Data collected at LHCb is being used to tackle many longstanding questions about the nature of our universe, one of which is the nature of dark matter. At LHCb, we can search for decay channels in which the final state contains pairs of baryons which match the quark content of an interesting dark matter candidate: the sexaquark.

## 1.2 Sexaquarks

It has been proposed that a group of six quarks (in particular $uuddss$, a neutral, spin-0 dibaryon) could be a very tightly bound, possibly stable or long-lived state. This is due to its triply singlet configuration in flavour, spin and colour. If discovered, this sexaquark would be a good standard model candidate for dark matter, due to its predicted stability [2].

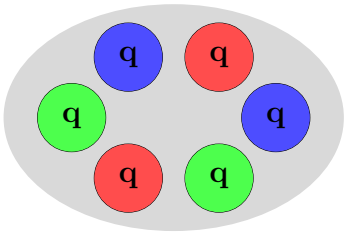


Figure 2: Illustration of a sexaquark.

Sexaquarks may be produced by a $\Xi_b^0$ decay

$$\Xi_b^0(usb) \to \mathrm{S}(uuddss) + K^+ + \bar{p}.$$

Here, we study the analogous decay

$$\Xi_b^0(usb) \to 2\ \Lambda^0(uds) + K^+ + \bar{p},$$

where the two $\Lambda^0$ baryons in the final state have the same quark content as the sexaquark. The Feynman diagram for this decay is shown in Fig. 3

Fig. 4 shows the full decay tree for this particular decay. The $\Lambda^0$ baryons fly and decay to a proton and a $\pi^-$ at tertiary vertices away from the $\Xi_b^0$ decay vertex. As outlined in Section 2, the kinematics, vertex quality and other properties of the particles in this decay tree can be used to train a multivariate analysis algorithm to differentiate between background and signal events in the data sample.

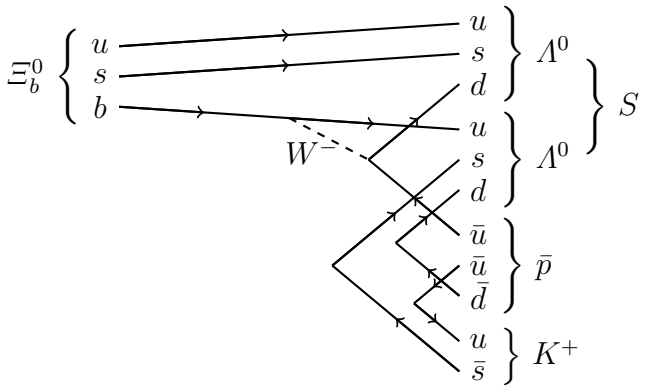


Figure 3: Feynman diagram for $\Xi_b^0(usb) \to 2\ \Lambda^0(uds) + K^+ + \bar{p}$.

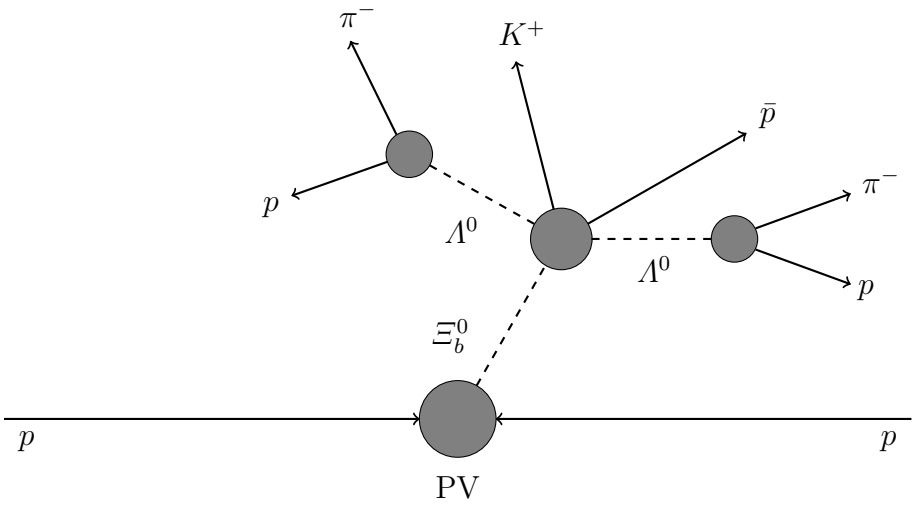


Figure 4: Decay tree for $\Xi_b^0(usb) \to 2\ \Lambda^0(uds) + K^+ + \bar{p}$.

## 1.3 Objectives and Datasets

Primarily, this analysis aims to conduct a feasibility study of the search for

$$\Xi_b^0(usb) \to 2\ \Lambda^0(uds) + K^+ + \bar{p}$$

decays. By training an algorithm to isolate signal events from the background, we look for mass peaks for each $\Lambda^0$ baryon and the $\Xi_b^0$ in the data with the main objective being the observation of the $\Xi_b^0$ peak as this would confirm the existence of the decay. One would then proceed to search for the sexaquark as missing energy in the decay. The analysis is carried out using data from LHCb Run 2 (2016, 2017, 2018) and Monte Carlo simulation.

# 2 Multivariate Analysis

To reject combinatorial background, a multivariate analysis (MVA) algorithm is trained using a portion of the data from the Run 2 sample and the full Monte Carlo (MC), combining samples from both magnet polarities. The variables[1] used to train the MVA are given in Table 1 and Fig. 5 shows the normalised distributions (MC and data) for the six variables with the highest importance. A full set of distributions for all the BDT features can be found in Appendix A.

Table 1: BDT features

| Particles | $\chi^2_{\text{endvtx}}$ | $\text{IP}_{\text{OWNPV}}$ | $\chi^2_{\text{IP}}$ | PT | P | cpt(0.70)/PT |
|---|---|---|---|---|---|---|
| $\Xi_b^0$ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $\Lambda^0(1)$ | | ✓ | ✓ | ✓ | ✓ | |
| $\Lambda^0(2)$ | | ✓ | ✓ | ✓ | ✓ | |
| $K^+$ | | | | ✓ | | ✓ |
| $\bar{p}$ | | | | ✓ | | ✓ |
| $p(\Lambda_1)$ | | ✓ | | ✓ | ✓ | ✓ |
| $\pi^-(\Lambda_1)$ | | | | ✓ | | ✓ |
| $p(\Lambda_2)$ | | ✓ | | ✓ | ✓ | ✓ |
| $\pi^-(\Lambda_2)$ | | | | ✓ | | ✓ |

The algorithm used is a gradient boosted decision tree (BDT) and is implemented using the XGBoost and scikit-learn libraries in python. The values of the hyperparameters are optimised to reduce overtraining while maintaining a high BDT performance by scanning over a range of values for each parameter. We scan over the `learning_rate` with a resolution of 0.05 between 0.05 and 0.20, the `max_depth` for integer values between 2 and 4 and `n_estimators` with a resolution of 100 between 300 and 500. The Kolmogorov-Smirnov test is used as an indicator of overtraining. This scan yields scores (area under ROC curve) between 0.82 and 0.93 with evidence of overtraining in configurations leading to scores above 0.86. As such, a set of hyperparameters giving a score of 0.86 is used for the training. The set of hyperparameters used is given in Table 2.

The full training sample is formed from the MC sample and the upper sideband of the background sample ($m(\Xi_b^0) > 5900$ MeV). We split this sample into two subsamples, one for training (67%), one for testing (33%) such that we may test on a sample where we know whether events are signal

[1]These include particle momentum (P) and transverse momentum (PT), end vertex $\chi^2$ ($\chi^2_{\text{endvtx}}$), impact parameter (associated with its own primary vertex) ($\text{IP}_{\text{OWNPV}}$) and its $\chi^2$ ($\chi^2_{\text{IP}}$) and the ratio of the cone PT isolation variable to PT (cpt(0.70)/PT).

or background and can also check for overtraining.

Fig. 6 shows the performance of the MVA, in particular the response of the MVA from signal and background samples. The MVA is evaluated on the train and test samples where we look for signs of overtraining; none is observed. Also shown in the figure is a ROC curve for the classifier.

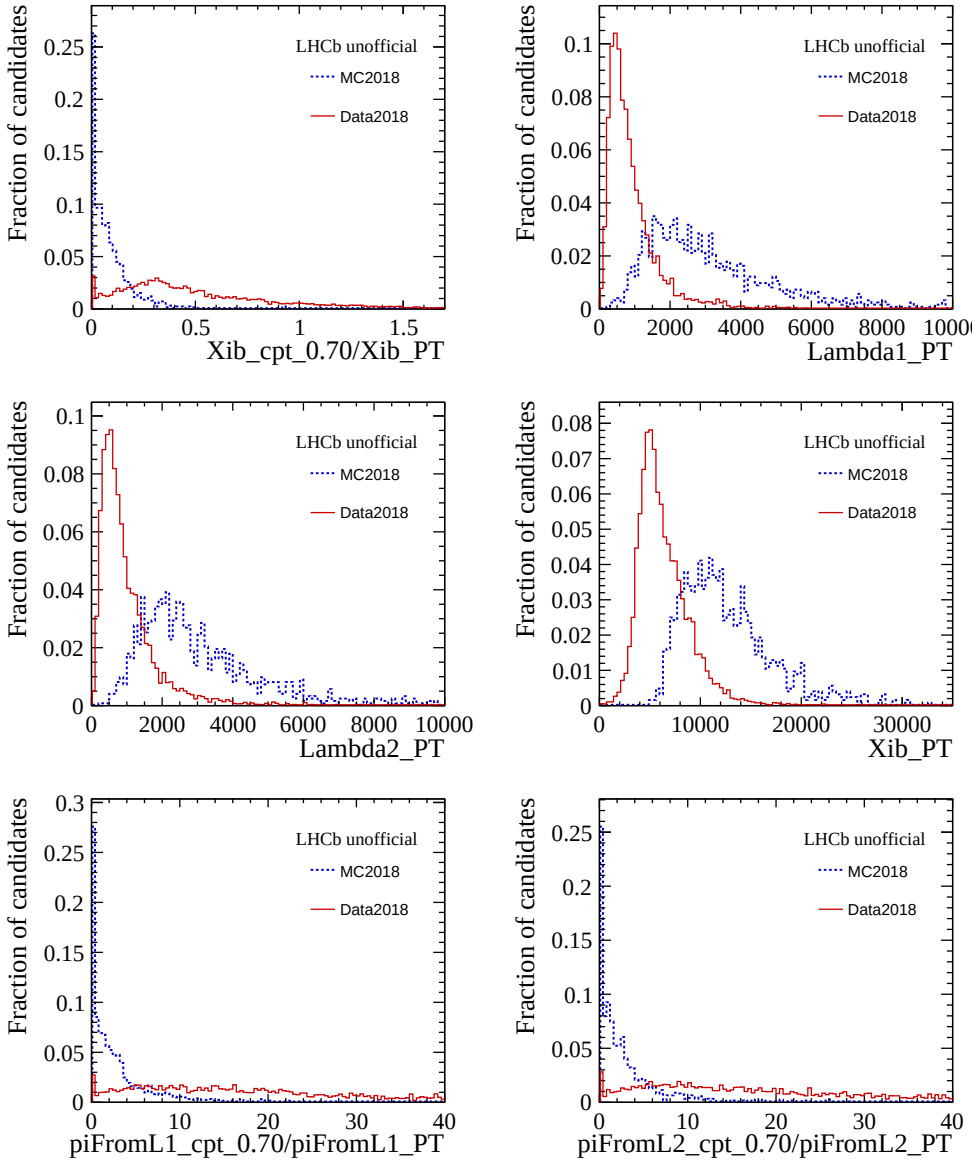


Figure 5: Normalised distributions of variables used as features of the BDT. MC distributions are depicted by dashed blue lines and data distributions are depicted by solid red lines.

Table 2: Hyperparameters used for BDT training

| Hyperparameter | Setting |
|---|---|
| `learning_rate` | 0.15 |
| `max_depth` | 2 |
| `n_estimators` | 500 |

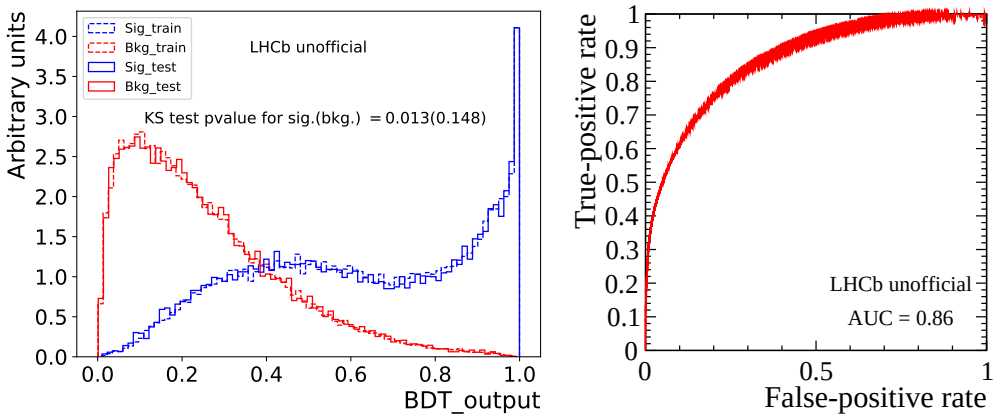


Figure 6: Performance plots for the MVA. Left: BDT output for the test and training samples of the signal and background samples. Right: ROC curve.

After training and testing, the MVA is applied to the rest of the background data sample ($m(\Xi_b^0) < 5900$ MeV) to isolate the peaks in the mass distributions of the $\Xi_b^0$ baryon and the $\Lambda^0$ baryons. Further details of these results are outlined in Section 3.

# 3 Results

The main results of the analysis are mass distributions for the initial state particle $\Xi_b^0$ and the daughters of this which constitute the quark content of the sexaquark, namely the two $\Lambda^0$ baryons. We look for peaks in these distributions as an indication of the decay we search for, applying cuts in the BDT output and on the particle identification (PID) variables associated with the daughters of the $\Lambda^0$ baryons to isolate the signal from the background.

We plot the distributions from the portion of the data sample not used for training the MVA ($m(\Xi_b^0) < 5900$ MeV) and use the MC distributions as a visual indicator of the expected location for peaks in the background.

## 3.1 $\Lambda^0$ mass distributions

We first look at the distributions for the daughters of the $\Xi_b^0$ which constitute the quark content of the sexaquark. Fig. 7 shows the distributions with and without cuts on the BDT output and proton PID. No effect was observed from varying the $\pi^-$ PID cut. The BDT manages to select the $\Lambda^0$ baryons well, with tighter cuts accentuating a peak in each of the background distributions which matches the expected peak value and shape of the distribution in Monte Carlo.

## 3.2 $\Xi_b^0$ mass distributions

In contrast to what is seen in the $\Lambda^0$ distributions, applying the same cuts to the $\Xi_b^0$ does not yield a clear peak in the background distribution. Fig. 8 shows the distributions, again with no cuts applied and then with the same cuts as above. The cuts lower the number of events but do not give any clear visual indication of a peak and by extension the decay we search for.

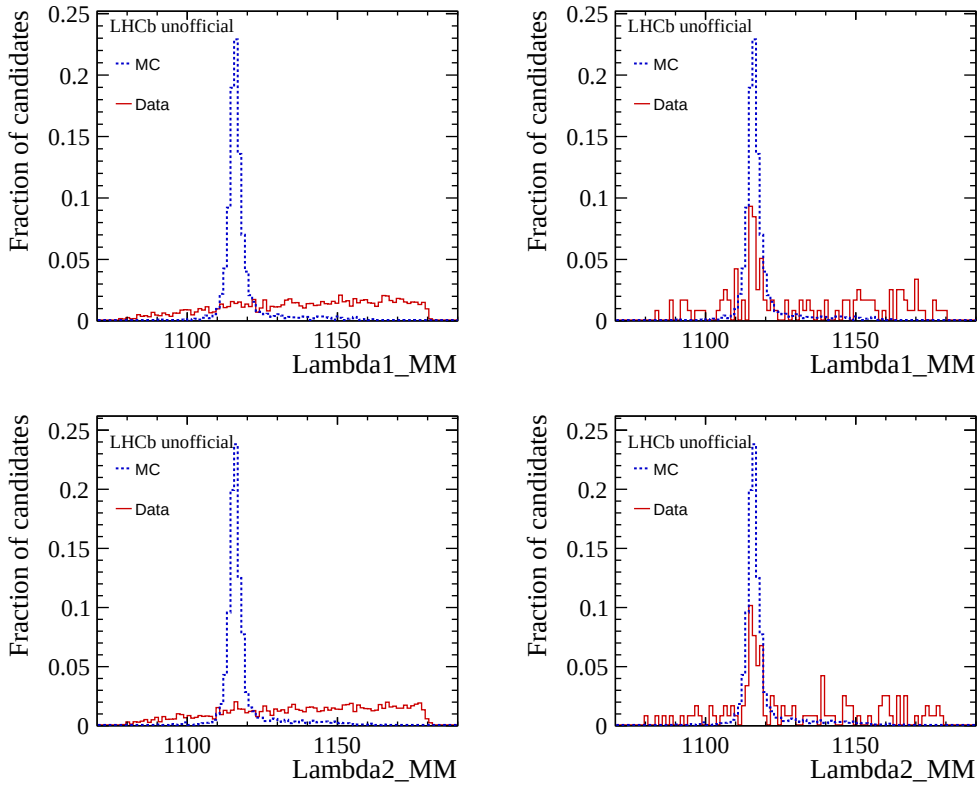


Figure 7: $\Lambda^0$ mass distributions in simulation (blue) and data (red). Left: No cuts applied. Right: BDT $> 0.4$, ProbNNp $> 0.4$.

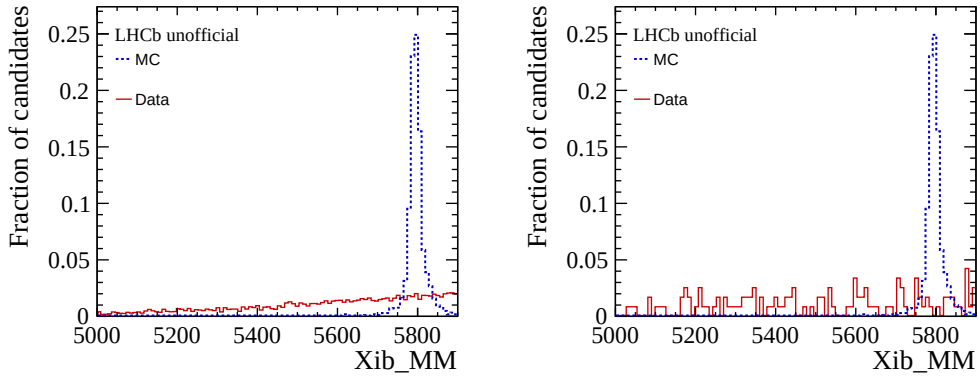


Figure 8: $\Xi_b^0$ mass distributions in simulation (blue) and data (red). Left: No cuts applied. Right: BDT $> 0.4$, ProbNNp $> 0.4$.

### 3.3 Recasting the analysis: $\Lambda_b^0$ decay

A similar decay to that which we search for has a $\Lambda_b^0$ baryon in the initial state in place of the $\Xi_b^0$. The Feynman diagram is shown in Fig. 9. We can recast our search using the same data and Monte Carlo to look for this decay due to the proximity of the masses of the $\Xi_b^0$ (5792 MeV) and $\Lambda_b^0$ (5620 MeV) baryons as well as the similarities in the kinematics and topology of the decays we search for.

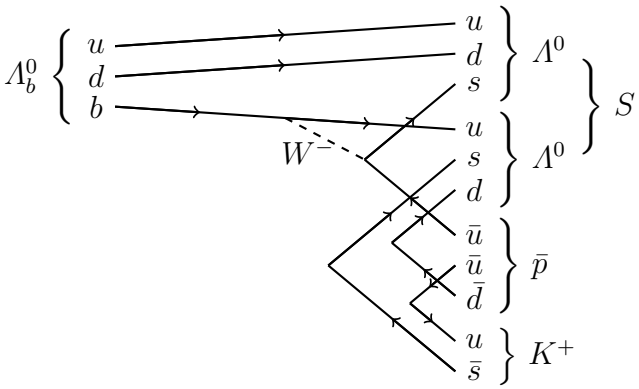


Figure 9: Feynman diagram for $\Lambda_b^0(udb) \to 2\ \Lambda^0(uds) + K^+ + \bar{p}$

Fig. 10 shows the $\Xi_b^0$ mass distribution again with cuts on the BDT and ProbNNp applied. The $\Lambda_b^0$ mass value (5620 MeV) is indicated, from which we see that there is no peak in the region we would expect, leading us to draw similar conclusions as those made about the $\Xi_b^0$.

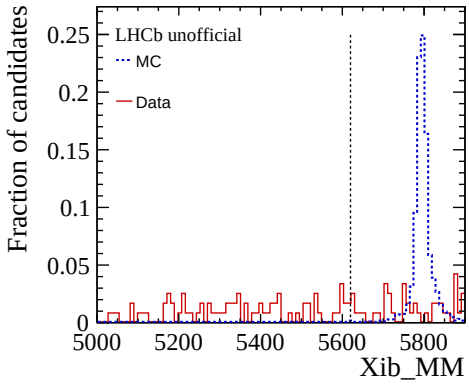


Figure 10: $\Xi_b^0$ mass distribution in simulation (blue) and data (red). The black dashed line indicates the position of the expected mass peak for $\Lambda_b^0$.

# 4 Conclusions and Outlook

## 4.1 Conclusions

The primary goal of this analysis was to perform a feasibility study of the search for

$$\Xi_b^0(usb) \to 2\ \Lambda^0(uds) + K^+ + \bar{p}.$$

The results in Section 3 show that by applying a PID cut on the proton daughter of the $\Lambda^0$ and for a sufficiently hard cut on the BDT output, we can observe a peak in the background distributions for either $\Lambda^0$ at the expected invariant mass value. However, with the same cuts applied no peak is seen around the $\Xi_b^0$ mass. As such, we conclude that the decay we search for is not detectable in the data using our methods. If the decay was seen, we could search for the sexaquark as missing energy in it.

## 4.2 Outlook

Combining this analysis with other works in the field gives a clear indication that we will not see the decay of $\Xi_b^0$ to the sexaquark state, at least not in Run 2 data. Future analyses searching for this decay and others like it will likely make use of data from Run 3 and beyond, where one might hope to see different results, leading to new understanding of the structure and properties of dark matter. In particular, replacing the L0 hardware trigger with a fully software based trigger for Run 3 will allow for softer cuts on kinematic variables (e.g. PT) of the decay daughters which should improve selection efficiency.

Further work can be done with this analysis to fully characterise the background, which, along with some theoretical work, could give insight into other possible decay modes in which the sexaquark could appear in the final state.

## 4.3 Acknowledgements

It goes without saying that this project would not have been anywhere near as enjoyable or useful as a learning experience had I not had the invaluable input of my two supervisors, María Vieites Díaz and Patrick Koppenburg. My thanks to them and to all who offered advice, delivered lectures or travelled with me around Switzerland at the weekends over the course of my 8 weeks at CERN.

# Appendices

# A BDT features

Distributions in Monte Carlo (MC) and data for each of the training variables used for training the MVA are shown in the figures below.

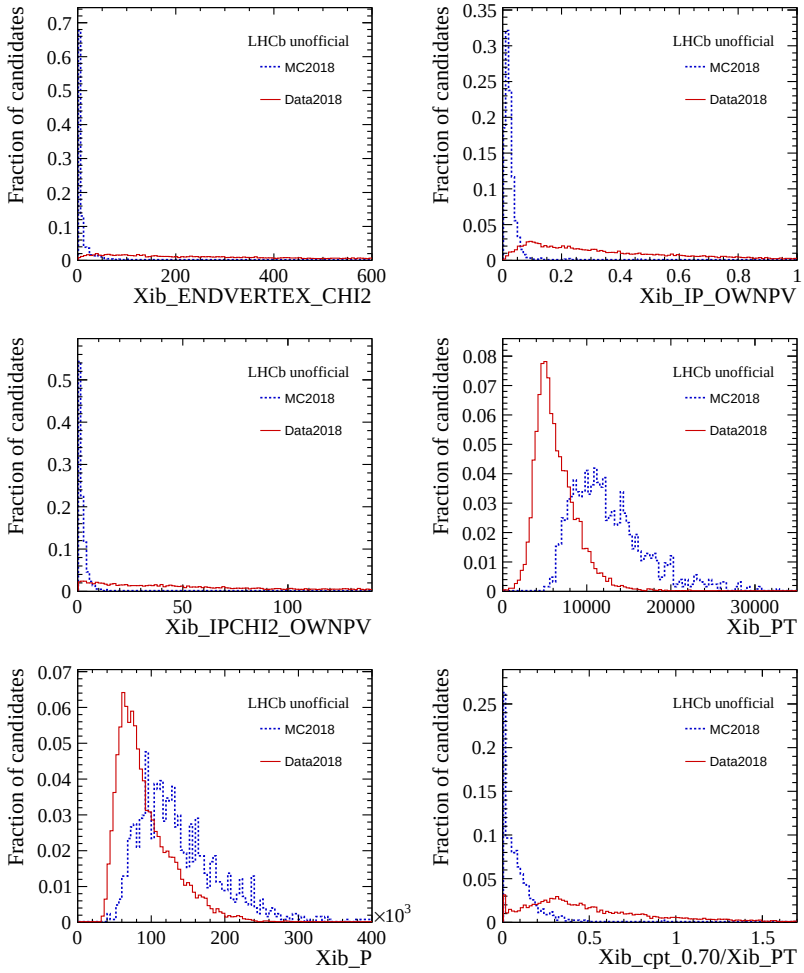


Figure 11: Normalised MC (dashed blue) and data (solid red) distributions of $\Xi_b^0$ variables used to train the BDT.

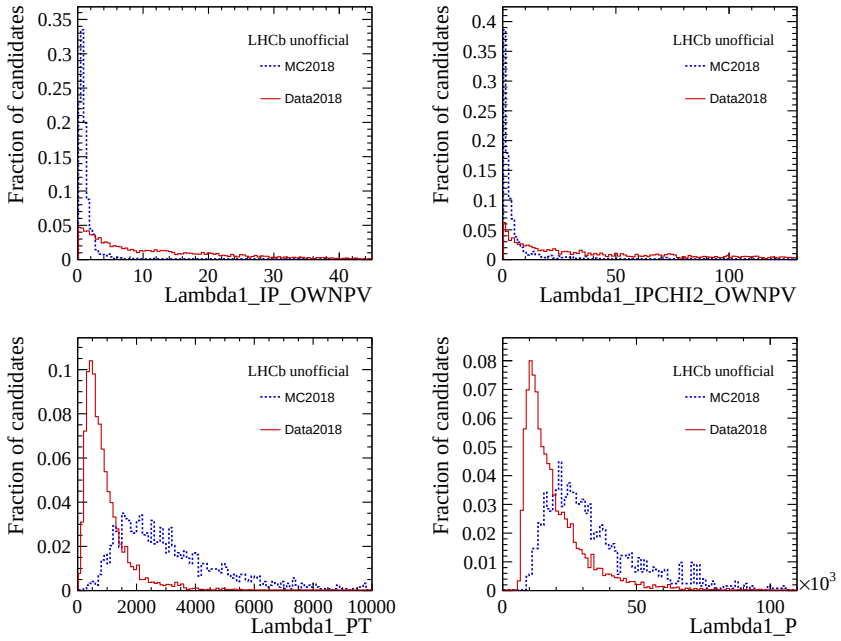


Figure 12: Normalised MC (dashed blue) and data (solid red) distributions of $\Lambda^0(1)$ variables used to train the BDT.

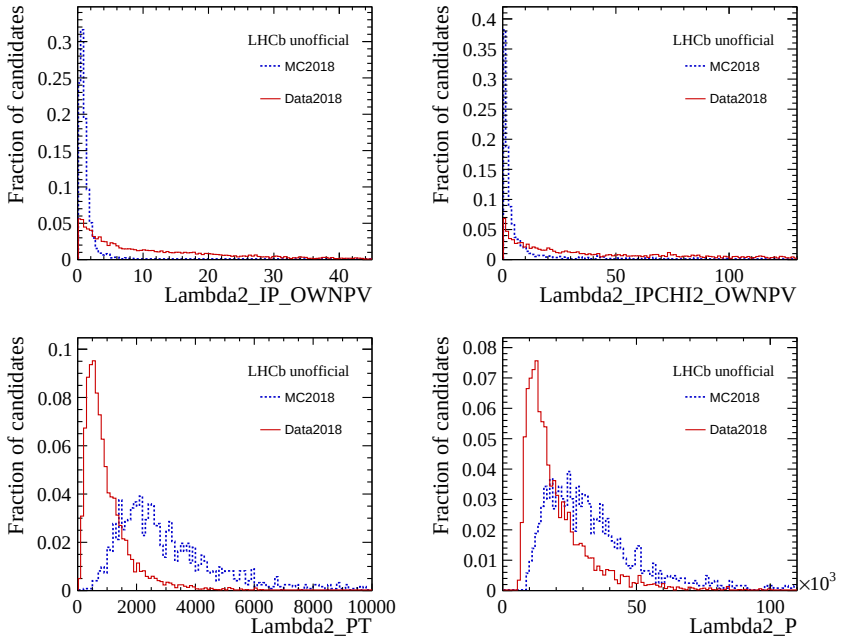


Figure 13: Normalised MC (dashed blue) and data (solid red) distributions of $\Lambda^0(2)$ variables used to train the BDT.

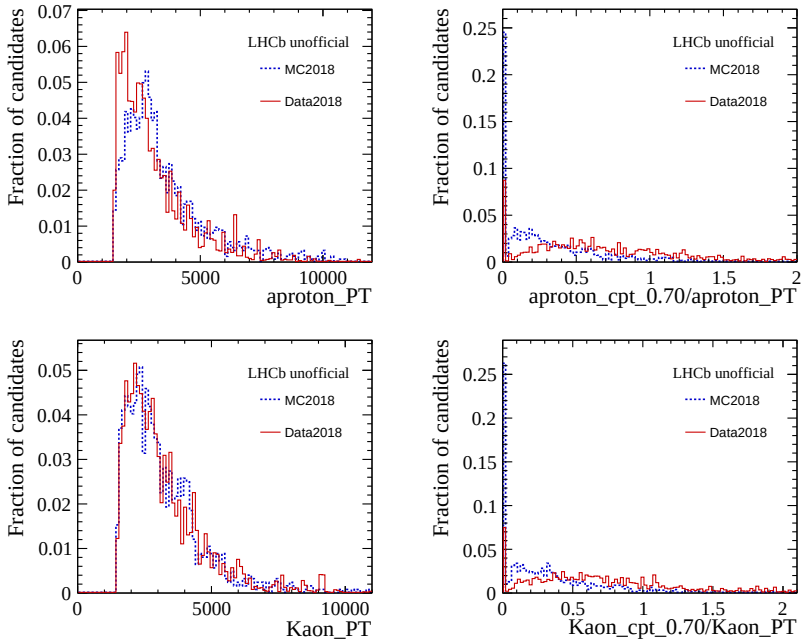


Figure 14: Normalised MC (dashed blue) and data (solid red) distributions of $\bar{p}$ and $K^+$ variables used to train the BDT.

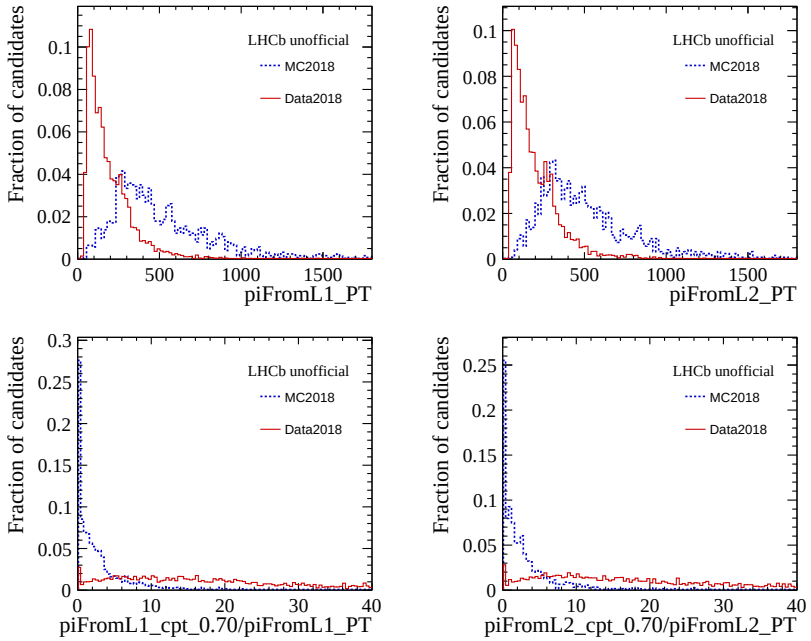


Figure 15: Normalised MC (dashed blue) and data (solid red) distributions of $\pi^-$ variables used to train the BDT.

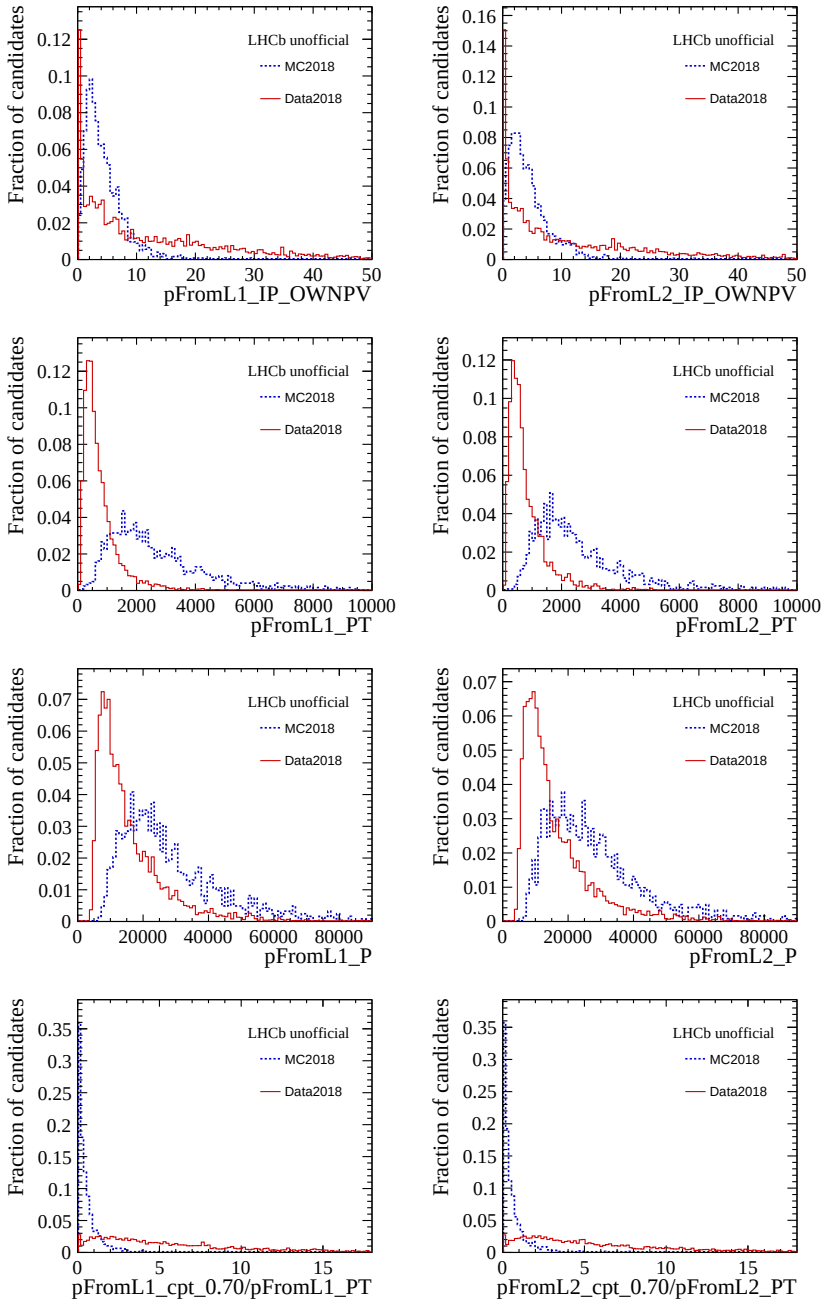


Figure 16: Normalised MC (dashed blue) and data (solid red) distributions of proton variables used to train the BDT.

# References

[1] "The LHCb Detector at the LHC", The LHCb Collaboration et al 2008 JINST 3 S08005.

[2] "Stable Sexaquark", G. R. Farrar, 1708.08951