Observation of a new charmed baryon, testbeam studies for Upgrade-II, and an enhanced analysis experience for the LHCb experiment

Ryunosuke O'Neil

Doctor of Philosophy

The University of Edinburgh

2024

Abstract

A search was conducted for excited Ξ_c^+ states in the $\Xi_c^+\pi^+\pi^-$ channel using $5.6\,\text{fb}^{-1}$ of pp collisions at $\sqrt{s} = 13\,\text{TeV}$, collected by the LHCb detector during Run 2 from 2016 to 2018. The following four excited states, with their masses and widths measured, were observed

$$
\begin{aligned}
m[\Xi_c(2815)^+] &= 2816.65 \pm 0.03\,(\text{stat}) \pm 0.03\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(2815)^+] &= 2.07 \pm 0.08\,(\text{stat}) \pm 0.12\,(\text{syst})\,\text{MeV},\\
m[\Xi_c(2923)^+] &= 2922.80 \pm 0.28\,(\text{stat}) \pm 0.29\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(2923)^+] &= 5.3 \pm 0.9\,(\text{stat}) \pm 1.2\,(\text{syst})\,\text{MeV},\\
m[\Xi_c(2970)^+] &= 2968.62 \pm 0.46\,(\text{stat}) \pm 0.30\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(2970)^+] &= 31.7 \pm 1.7\,(\text{stat}) \pm 1.7\,(\text{syst})\,\text{MeV},\\
m[\Xi_c(3080)^+] &= 3076.77 \pm 0.72\,(\text{stat}) \pm 0.11\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(3080)^+] &= 6.8 \pm 2.3\,(\text{stat}) \pm 0.9\,(\text{syst})\,\text{MeV},
\end{aligned}
$$

where the uncertainties include statistical, systematic, and the PDG Ξ_c^+ mass measurement uncertainty. The $\Xi_c(2923)^+$ state is observed for the first time in any channel, and may be the isospin partner of $\Xi_c(2923)^0$ which was first observed by LHCb in the $\Lambda_c^+ K^-$ channel.

A precursor to the MightyPix series, the ATLASPix3.1, was tested at varying irradiation levels during two testbeam campaigns in June and December 2022 at the DESY-II Testbeam Facility, to evaluate the performance of the design and develop testbeam analysis methods for future campaigns. The efficiency and timing performance was evaluated for sensors over a range of fluences and operating temperatures.

A user-friendly interface, LbAPWeb, was created to automatically provide LHCb

analysts easy access to additional metrics, tools, and feedback used to validate LHCb production configurations before they are submitted and run at full scale using the Analysis Productions service. Typically, such validations would be carried out manually meaning steps are time-consuming and often skipped by analysts, necessitating re-runs which can be computationally expensive. The user interface and underlying services continue to be developed as the collaboration examines new Run 3 data, as part of the officially recommended Run 3 data processing and analysis workflow.

Lay Summary

Matter is built up from fundamental particles called quarks and leptons. The quarks are known to exist bound together generally in twos (mesons) or threes (baryons) by a fundamental force of nature called the strong force. The Standard Model of Particle Physics, based upon many years of theoretical work and experimentation, states that these quarks can only exist together according to certain rules.

Baryons containing a charm (c) quark are known as charmed baryons, and a variety of them exist for the different quark combinations at distinct masses (their "ground" state). However, particles composed of the same quarks but with higher masses, "excited" states, are also known to exist.

Our current scientific understanding of these excited states is not complete. At present, it is very difficult to predict with theoretical frameworks which excited states exist, and what mass and other properties they might have.

Experiments in particle physics are essential for filling in these gaps, by observing these particles and measuring their properties. The LHCb experiment, though not originally designed for this purpose, is very well-placed for this task due to the detector's high performance. The detector has accumulated huge samples of charmed baryons available for study to meet this challenge.

Acknowledgements

The LHCb collaboration is a wonderful thing to be a part of, and I am proud to have worked within it. It is a privilege to have been able to contribute to the effort across many areas both scientific and technical in nature, whilst working alongside many talented people. I was able to witness first collisions in the machine (the LHC) for Run 3 whilst in the LHCb control room, see the experiment with my own eyes, guide others to see the experiment, and these are things I will remember fondly.

I must thank my supervisor, Mark Williams, for providing the perfect level of supervision, for never failing to provide thoughtful answers and explanations throughout my PhD, for anticipating hurdles, helping address and organise well ahead of time, and finally for checking in with me often and keeping the ship steady when it mattered. Thank you!

Another huge thank you to Marco Pappagallo, who provided guidance from the very beginning of my PhD as I was still finding my footing in the swamp that was Run 2 LHCb analysis and baryon spectroscopy, and without whom it would not have been possible to have achieved the result that we did.

I have thoroughly enjoyed working with Chris Burr on LHCb Computing and DPA projects. Thank you for sharing your expertise, time, and valuable thoughts in these areas and for working with me to improve the quality of life in the LHCb analysis ecosystem!

A big thanks to the hard-working people I worked with at DESY on the gruelling testbeam campaigns of 2022! In particular, Jan Hammerich, Lennart Huth, Klaas Padeken, Hannah Schmitz, and others on the team whose patience with me and their willingness to collaborate, explain, and help me understand the chips, testbeam setup, and particularly difficult details of testbeam analysis helped me to apply myself in a fast-paced testbeam environment!

I would like to additionally thank my viva examination panel: Ben Wynne (internal), Davide Costanzo (external), and Yanyan Gao for an unforgettable and very well-run viva examination!

To those I worked with but was not able to mention here — many colleagues at Edinburgh, CERN, LHCb, and beyond: thanks for everything!

One of the best things about this PhD was the opportunity to spend some time out at CERN on LTA – my PhD would not have been the same without it. I think it is a wonderful thing for my funder, the Science & Technology Facilties Council (STFC) to offer, as it has added so much value to my PhD experience. It was incredibly enriching time-well-spent both at work (see the rest of my thesis for further details) and many memorable experiences out of work too. I thank the good friends I made and colleagues I worked with in Geneva and CERN. I would also thank all my wonderful friends at home and beyond who kept in touch and continued to believe in me.

Lastly, I would like to share a *haiku* written for me by my grandmother:

プロトン語り青年卒業す

—— 和田公子

Which roughly translates as

As he gives his talk
on protons(baryons), the young man
graduates.

おばあちゃんも皆さんも、ずっと応援してくれてありがとうございます。これからも頑張ります。

和田龍之介

Dedicated to my mother.

Contents

List of Figures

List of Tables

Preface

I worked on a broad variety of projects in physics, computing, and detector R&D for LHCb. On the whole, I immensely enjoyed the process of research and bringing new features and knowledge to the community. In this thesis I present the following bodies of work covering:

- **Baryon spectroscopy** (Part II): A search for charmed excited states (Ξ_c^{**+}) in the $\Xi_c^+\pi^-\pi^+$ channel over Run 2 data. I took the lead in preparing this analysis by conducting the entire data analysis with the guidance of my supervisor (Mark Williams) and another spectroscopy expert (Marco Pappagallo). I experimented freely with a range of modern tools and methods relating to the manipulation of the data and the preparation of Machine Learning classification models, while opting to stick with well-established tools for fitting. The analysis resulted in an observation of a new charmed baryon state.

- **Testbeam studies of irradiated HV-MAPS** (Chapter 9): I worked as part of a larger team of LHCb collaborators working on the MightyPix project. I participated directly in two testbeams at the DESY-II Testbeam Facility in Hamburg during June and December 2022, aiming to evaluate the performance of irradiated ATLASPix3.1 sensors. I took a leading role in setting up the Testbeam analysis chain using the Corryvreckan analysis software, and used this directly to perform data-quality cross-checks during real data-taking to validate the setup. Although very challenging at first, by collaborating with DESY and LHCb experts to gain the necessary skills and understanding to have an impact on the success of the data-taking, we were able to discover issues we could rectify before the end of the data-taking campaigns as a result. The results emerging from the testbeams are within expectations, and I hope that the knowledge and experience is successfully carried forward into the next testbeams with real MightyPix prototypes.

- **Experience enhancements in n-tupling for physics analysis** (Chapter 10): I contributed significantly towards an already established but new-and-improved *Analysis Productions* system designed to automate much of the business of creating n-tuples for physics analysis within LHCb. Much of the system was already operational at the time of my joining LHCb. However, my contribution was the replacement of an old web application,

with a completely new one that uses modern user interface frameworks and a new API backend sitting behind it. The primary benefit was a direct improvement of the entire user experience of making n-tuples, which I was also able to experience as an LHCb analyst myself! I was able to deliver these features thanks to Chris Burr, who provided a great deal of technical support and direction that made this possible.

Part I

Background

Chapter 1

Introduction

The smallest, most fundamental constituents of matter and the rules of their behaviours and interactions, to the extent that theory and experimental efforts have been able to probe, are described in the Standard Model of Particle Physics. The Standard Model is most famous for its unparalleled precision and the numerous successes of its predictions, but also the unanswered questions raised about our Universe.

This incompleteness has continually motivated the international scientific community to construct and operate colliders such as the Large Hadron Collider (LHC), currently running and delivering collisions at ever-increasing rates, to experiments that are instrumented and periodically upgraded with improved detector technologies. The physics programmes at the LHC experiments target measurements in a wide range of areas to comprehensively test and validate the Standard Model. Plans are already coalescing around a new Future Circular Collider (FCC) after the High-Luminosity LHC. These are just examples of the efforts at CERN, among many currently underway worldwide to deliver the capabilities and volume of data needed to access difficult measurements.

At the LHC, due to the structure of the colliding protons, the prevailing interactions are those described by Quantum chromodynamics (QCD), a powerful theoretical framework describing these strong interactions. At lower energies it becomes more difficult to make reliable predictions with QCD, due to the strong coupling. While there are many approaches to work around this limitation, any experimental input that can improve such approaches will provide benefits for a wide range of measurements at the LHC and beyond. Hadron spectroscopy

is a key testing ground for QCD focussing on the observation of new states, or confirming the absence of them in studied mass spectra, in addition to measuring properties, and comparing them to various theoretical approaches.

In this thesis, a search for excited Ξ_c^+ baryon states conducted by the author is covered in Part II, which led to the observation of a new state. This body of work is followed by additional studies covered in Part III regarding testbeam analyses with irradiated ATLASPix3.1 sensors, where the author took a leading role in the analysis of testbeam data and data quality checks during live data-taking. Finally, the author's efforts to improve the LHCb analysis experience with enhancements to the process of creating analysis n-tuples.

1.1 The Standard Model

The Standard Model describes matter as being made up of fundamental particles and antiparticles, comprising the quarks, leptons (listed in Table 1.1), and bosons. These particles interact with each other via the fundamental *electromagnetic*, *strong*, and *weak* forces communicated by the exchange of gauge bosons. These are: photons (γ) for electromagnetism acting between carriers of electromagnetic charge; gluons (g) for the strong force acting on colour-charge[1] carrying particles; and $W^{\pm}/Z^0$ bosons for the weak force, coupling to particles with weak isospin. The Standard Model does not provide a description of gravitation, another fundamental force of nature [2].

The atom paints a picture of these fundamental forces at work. The atomic nucleus is made up of protons (uud) and neutrons (udd) bound together by the strong force, while the orbiting electrons are kept in orbit by electromagnetism between the positively charged nucleus and negatively charged electrons. Unstable nuclei undergoing radioactive decay, are described adequately by weak interactions.

Unlike leptons such as electrons orbiting a nucleus, quarks can not exist in isolation and must be bound together in pairs (*mesons*, $q\bar{q}$) or in threes (*baryons* qqq, and *antibaryons* $\bar{q}\bar{q}\bar{q}$), due to a phenomenon of the strong interaction known as colour confinement. The resulting bound states are referred to as *hadrons*. A plethora of hadronic states have been observed owing to numerous possible

[1] No connection to visually experienced colours.

[2] From direct measurements. Refer to [1] for further details on the top quark mass.

Particle	(Antiparticle)	Charge		J^P	Isospin	Mass
u	$\overline{u}$	$+2/3$	$-2/3$	$1/2^+$	$1/2$	2.16 ± 0.07 MeV
d	$\overline{d}$	$-1/3$	$+1/3$	$1/2^+$	$1/2$	4.70 ± 0.07 MeV
e^-	e^+	-1	$+1$	$1/2^+$		0.511 MeV
ν_e	$\bar{\nu_e}$	0		$1/2^?$		$< 0.8\,\text{eV}$
s	$\overline{s}$	$-1/3$	$+1/3$	$1/2^+$	0	93.5 ± 0.8 MeV
c	$\overline{c}$	$+2/3$	$-2/3$	$1/2^+$	0	1.2730 ± 0.0046 GeV
μ^-	μ^+	-1	$+1$	$1/2^+$		105.66 MeV
ν_μ	$\bar{\nu_\mu}$	0		$1/2^?$		$< 0.19\,\text{MeV}$
b	$\overline{b}$	$+2/3$	$-2/3$	$1/2^+$	0	4.183 ± 0.007 GeV
t	$\overline{t}$	$-1/3$	$+1/3$	$1/2^+$	0	172.57 ± 0.29 GeV^2
τ^-	τ^+	-1	$+1$	$1/2^+$		1776.93 ± 0.09 MeV
ν_τ	$\bar{\nu_\tau}$	0		$1/2^?$		$< 18.2\,\text{MeV}$

Table 1.1 The twelve spin-half fermions: quarks and leptons, divided into three generations, aside their properties. Values are taken from Ref. [1].

combinations of the different quark types, and additional 'excited' states in which the hadrons can exist. As with atoms, which can exist in excited states at higher energy levels, excited states of hadrons are possible too. These excitations are defined by quantum numbers such as the radial quantum number, orbital, and spin angular momentum.

The fundamental theoretical framework describing the strong interaction is QCD. QCD is a quantum field theory describing the strong interaction between particles (i.e. quarks and gluons) which carry *colour charge*. There are three colour charges, labelled r (red), g (green), and b (blue), which are carried by quarks. Antiquarks hold the opposite charges $\bar{r}$ (anti-red), $\bar{g}$ (anti-green), and $\bar{b}$ (anti-blue). Gluons carry one colour and one anticolour charge, with eight possible colour assignments

$$r\bar{g}, r\bar{b}, g\bar{r}, g\bar{b}, b\bar{r}, b\bar{g}, \tag{1.1}$$

and

$$\frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g}), \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b}). \tag{1.2}$$

A key governing component of strong interactions is *colour confinement*. As previously mentioned, quarks can not exist in isolation and must exist in a hadronic state, as a meson, baryon, antibaryon, or the more recently known tetraquark or pentaquark, such that the total colour is neutral (i.e. the colour quantum numbers $I_3^C = Y^C = 0$), or, colour states of the quarks together form a colour singlet state. The colour states and the colour hypercharge I_3^C and isospin

Colour state	I_3^C	Y^C
r	$+1/2$	$+1/3$
g	$+1/2$	$+1/3$
b	0	$-2/3$
$\bar{r}$	$-1/2$	$-1/3$
$\bar{g}$	$+1/2$	$-1/3$
$\bar{b}$	0	$+2/3$

Table 1.2 Colour states for the quarks and antiquarks, with their corresponding colour hypercharges I_3^C and colour isospin charges Y^C.

charges Y^C are summarised in Table 1.2.

Colour confinement imposes strong constraints on the possible quark compositions of the hadronic bound states, and of those observed to date are mesons $q\bar{q}$, baryons qqq, antibaryons $\bar{q}\bar{q}\bar{q}$, tetraquarks $q\bar{q}q\bar{q}$ [3], and pentaquarks $qqqq\bar{q}$ [4–6] (a comprehensive review and list of the hadrons can be consulted in Refs. [1, 7]).

1.2 Isospin and the Quark Model

> Young man, if I could remember the names of these particles, I would have been a botanist.
>
> Enrico Fermi

In the early days of modern particle physics, experimentalists were fast running out of names for newly observed hadrons, with conspicuously similar masses but different charges. Over time it was clear that these were in fact not fundamental particles like the electron or muon, but particles with an underlying composite structure. Eventually, the quark model [8,9] was introduced. However, before the concept of colour charge and quarks, a good approximate symmetry of hadrons, isospin symmetry, was formulated.

Isospin is a quantum number related to the u- and d- quark content of the hadron, initially developed to explain how the proton and neutron could be affected by the strong interaction equally despite having different charges. From the QCD point of view, this is because the u and d quarks are almost identical in mass (but not charge), and the proton and neutron only differ by swapping a u for

a d-quark or vice-versa. The isospin interpretation is that e.g. the proton and neutron, while having approximately the same mass, were interpreted to be two different states of the same particle having a number of 'charge states' described by isospin quantum numbers.

The quark model [8,9] followed from isospin, as the existence of additional flavours (such as strangeness) became clear, which led to the formulation of the concept of quarks, initially the up, down, and strange quarks.

Currently, isospin remains a relevant concept in the quark model, and is defined in terms of the quark content of the hadrons themselves. There are five types of quark (and antiquark) that can form hadronic states, the up (u), down (d), strange (s), charm (c), and bottom (b) quarks (the t-quark is not included, as it decays too quickly to form such states.) In essence, hadrons can be organised into multiplets of equal spin J, parity P, baryon number B, strangeness S, charm C, and bottom $\tilde{B}$ quantum numbers. The members of a multiplet differ in electric charge Q [10].

The isospin quantum number is

$$I = (I_3)_{\max}, \tag{1.3}$$

which is the maximum I_3 value in a given multiplet, where I_3 is the third component of isospin,

$$I_3 = Q - \frac{Y}{2}, \tag{1.4}$$

and

$$Y = B + S + C + \tilde{B}, \tag{1.5}$$

where Y is the hypercharge. The multiplets can be visualised as in Figure 1.1, which shows the $J^P = 1/2^+$ octet for the light baryons ($C = \tilde{B} = 0$) and $J^P = 1/2^+$ nonet of charmed baryons.

1.3 Baryon spectroscopy

The baryon states listed in Table 1.3 correspond to the well-known ground state singly-charmed baryons, with assignments $J^P = 1/2^+$. In addition to these ground states, it is possible to have excited states with identical quark composition but a greater total angular momentum number J which will be

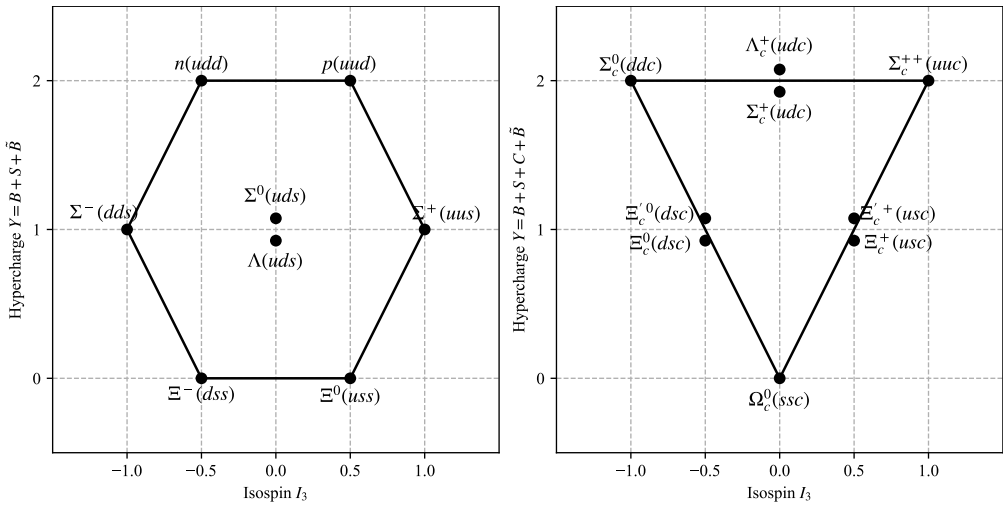

Figure 1.1 $J^P = 1/2^+$ octet of light baryons (right) and $J^P = 1/2^+$ nonet of singly-charmed baryons (left).

described in the following sections. In addition to their J^P assignments, excited states also have greater masses than the ground states, and have a measurable width if they are unstable resonances. One of the challenges of experimental baryon spectroscopy, alongside observing the states themselves, is measuring the quantum numbers in order to assign it to the correct baryon multiplet. Figure 1.2 paints a vivid picture of the excited charm baryon spectrum for Ξ_c^+, showing a multitude of mass predictions for states and their spin and parity J^P assignments.

Flavour quantum numbers

The flavour quantum numbers strangeness S, charmness C, and bottomness $\tilde{B}$ are related to the number of corresponding quarks and antiquarks,

$$S = -\left[N(s) - N(\bar{s})\right], \tag{1.6}$$

$$C = N(c) - N(\bar{c}), \tag{1.7}$$

and

$$\tilde{B} = -\left[N(b) - N(\bar{b})\right], \tag{1.8}$$

where $N(q)$ is the number of quarks (i.e. for s, c, or b), and $N(\bar{q})$ the number of antiquarks in a given hadronic state.

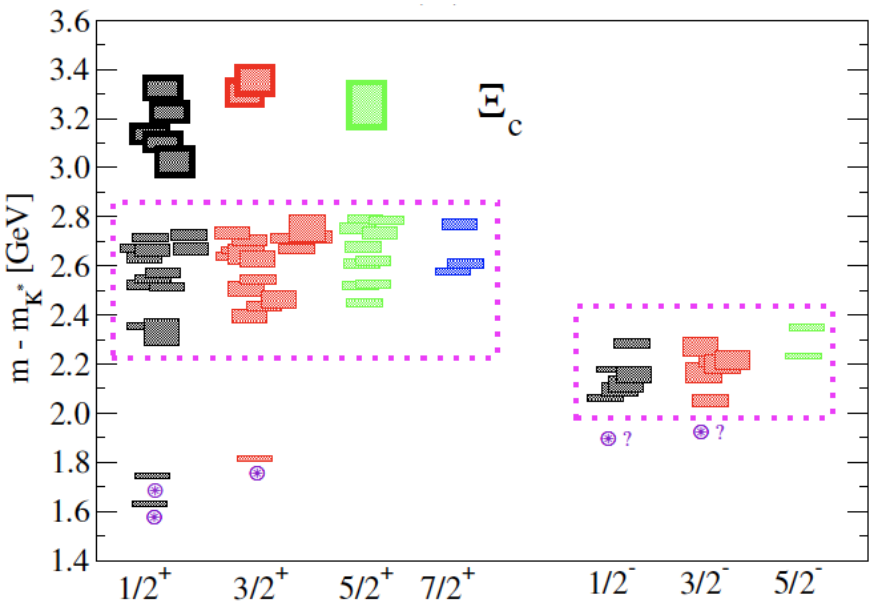

Figure 1.2 Excited spectra of singly-charmed Ξ_c^+ baryons. Reproduced from Ref. [11].

Angular momentum quantum numbers

For the baryons, the total angular momentum quantum number can take values in the range

$$|L - S| \leq J \leq |L + S|, \tag{1.9}$$

with integer increments between each allowed value, where S is the spin angular momentum, and L is the orbital angular momentum. The spin of a baryon state S is the total spin of the three quarks in the hadronic state. As all quark types are spin-$\frac{1}{2}$ particles, S may take two possible values,

$$S = 3/2, 1/2, \tag{1.10}$$

where the first value is simply the addition of the three quark spins when they are aligned ($\uparrow\uparrow\uparrow$), and the second when they are antialigned ($\uparrow\downarrow\uparrow$).

The orbital angular momentum quantum number L describes the angular moment arising from the orbiting quarks, which takes integer values, e.g. $L = 0, 1, 2, ...$ States with an orbital angular momentum $L = 0$ tend to correspond to the lightest ground state baryons.

Parity

For the baryon states, total angular momentum J is usually expressed alongside the parity of the state, for example $J^P = (1/2)^+$.

The parity P (denoted either as positive, $+$ or negative $-$) describes the symmetry of the given state under a $\mathbf{r} \to -\mathbf{r}$ transformation [2], or a reflection in spatial dimensions. For baryon states, the parity is related to the orbital angular momentum number L such that,

$$P = (-1)^L. \tag{1.11}$$

Summary

Grouping together the baryon states by their total angular momentum number and parity J^P reveals a number of 'multiplets' of baryons. Each of the baryons in the multiplets, while having the same J^P, can have different properties (mass, charge, width, and allowed decays) which can be measured. The theoretical predictions of the properties of these baryons are not always accurate, and experimental inputs can assist in improving these theoretical descriptions of the baryon multiplets.

1.4 Singly charmed baryons

The singly-charmed baryons are those containing one charm quark and two light quarks, i.e. $C = 1$. In contrast to light baryons, there is a large difference in mass between the heavier charm quark and the pair of lighter quarks, which can be described using theoretical frameworks such as e.g. Heavy-Quark Effective Theory (briefly discussed in Section 1.5). According to the PDG and in terms of the quark model picture, all of the $L = 0$ singly-charmed baryons (of which there are fifteen) have been observed [1]. Their symbols have a subscript c to denote they have a single charm quark: Λ_c, Σ_c, Ξ_c, and Ω_c. A summary of the ground state charm baryons are available in Figure 1.3, Table 1.3, and Table 1.4.

When considering higher orbital and radial excitations, the number of these baryons are much greater. There has been a wealth of recent activity by

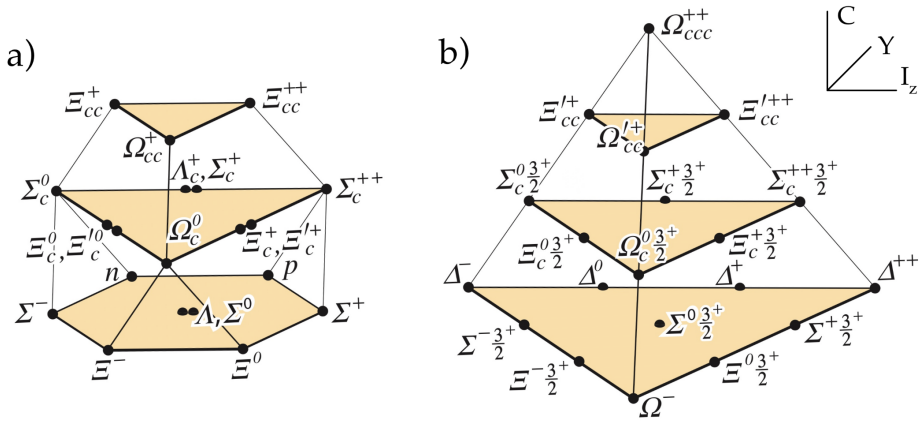

Figure 1.3 The (a) $J = 1/2$ and (b) $J = 3/2$ ground-state ($L = 0$) baryon multiplets. Reproduced from [1].

Baryon	Mass (MeV)	J^P	Y	I_3
$\Lambda_c^+(udc)$	2286.46 ± 0.14	$1/2^+$	2	0
$\Sigma_c^0(ddc)$	2453.75 ± 0.14	$1/2^+$	2	-1
$\Sigma_c^+(udc)$	$2452.65^{+0.22}_{-0.16}$	$1/2^+$	2	0
$\Sigma_c^{++}(uuc)$	2453.97 ± 0.14	$1/2^+$	2	$+1$
$\Xi_c^+(usc)$	2467.71 ± 0.23	$1/2^+$	1	$+1/2$
$\Xi_c^0(dsc)$	2470.44 ± 0.28	$1/2^+$	1	$-1/2$
$\Omega_c^0(ssc)$	2695.2 ± 1.7	$1/2^+$	0	0

Table 1.3 Table of ground state $J^P = 1/2^+$ singly-charmed baryons [1].

experiments such as LHCb, Belle, and the B-factories at the Tevatron in exploring the charmed baryon spectra. LHCb for example discovered five narrow $\Omega_c^0(css)$ states in 2017 [12] of which 4 were confirmed by Belle in e^+e^- collisions [13]. The experimental status of Ξ_c baryons is covered in Section 1.6, as it relates to the analysis presented in Part II.

Despite the observations to date, the spin and parity of most of the excited states have not been experimentally measured [1], and it is predicted that many other states are expected to exist.

1.5 Theoretical approaches

For the excited baryon states, many have been observed, however, many states yet to be discovered may exist, and of those that do, many of the properties (e.g. spin and parity) are yet to be measured. Theoretically, in contrast to the ground

Baryon	Mass (MeV)	J^P	Y	I_3
$\Sigma_c^{*0}(ddc)$	2518.48 ± 0.21	$3/2^+$	2	-1
$\Sigma_c^{*+}(udc)$	$2517.4^{+0.7}_{-0.5}$	$3/2^+$	2	0
$\Sigma_c^{*++}(uuc)$	2518.41 ± 0.22	$3/2^+$	2	$+1$
${\Xi_c}^{*+}(usc)$	2645.10 ± 0.30	$3/2^+$	1	$+1/2$
$\Xi_c^{*0}(dsc)$	2646.16 ± 0.25	$3/2^+$	1	$-1/2$
$\Omega_c^{*0}(ssc)$	2765.9 ± 2.0	$3/2^+$	0	0

Table 1.4 Table of ground state $J^P = 3/2^+$ ($L = 0$) singly-charmed baryons [1].

states it is not a simple matter to predict what excited baryons may exist, what masses they exist at, and their quantum number assignments.

A range of models exist to describe the underlying framework underpinning the excited baryons, allowing predictions to be made about which states exist, and with what properties and quantum number assignments. This section focuses on singly-charmed baryons (one heavy quark, and light diquark system). A brief coverage of some relevant theoretical models will be presented with some recent theoretical predictions for baryon masses, that may later be relevant to the results obtained in Part II.

Calculations with QCD are notoriously difficult, due to the large strong coupling at low energies, meaning that perturbation theory cannot be used. The Lattice QCD approach provides a means to solve QCD on a lattice of points in space and time. Numerical calculations with Monte Carlo methods and Lattice QCD, while very computationally intensive, can be used to make theoretical predictions about the charmed baryons. Studies into the charmed baryon spectrum with Lattice QCD have been performed, for example in Ref. [14], where mass predictions for various J^P assignments are presented including those for the Ξ_c and $\Xi_c{}'$ baryons.

Heavy Quark Effective Theory (HQET) represents another approach relevant to baryon systems consisting of two light (u-, d-, and s-) quarks and a single heavy (c- or b-) quark [15, 16]. In this approach, a system of one quark with mass much greater than the QCD energy scale, and two light quarks are considered. A recent example of an application of this method with QCD sum rules, are presented in Ref. [17], with the predictions for $\Xi_c(2815)^+$ relevant to the results of Part II.

Another theoretical model recently being used to make predictions about the singly charmed baryon spectrum is the non-relativistic harmonic oscillator model, such as the one used in Ref. [18] which lists mass predictions in Table IV.

1.6 Experimental status of excited Ξ_c^+ baryons

The observation of Ξ_c^{**0} (csd) baryons in $\Lambda_c^+ K^-$ by LHCb [19, 20], motivated by the observation of narrow Ω_c^0 (ssc) states in $\Xi_c^+ K^-$ [12, 21] (also by LHCb), hints at possible opportunities to observe the Ξ_c^{**0} isospin partners in other channels.

In the analysis presented in Part II, $\Xi_c^{**+}(csu) \to \Xi_c^+ \pi^- \pi^+$ decays are targeted to search for the isospin partners of Ξ_c^{**0} (csd). Excited Ξ_c^+ states, referred to hereafter as Ξ_c^{**+} (csu), are expected to decay strongly either resonantly via $\Xi_c(2645)^0 \pi^+$ or non-resonantly to $\Xi_c^+ \pi^- \pi^+$. The current experimental status for these excited charmed baryons, as summarised by the PDG, can be referred to in Table 1.5.

State	Mass (MeV)	Width (MeV)	J^P	Largest event sample	Observed?	Mode $\Xi_c(2645)^0\pi^+$
$\Xi_c(2815)^+$	2816.51 ± 0.25	2.43 ± 0.26	$?^?$	72.5 ± 9.6 [22]	Yes	Seen
$\Xi_c(2923)^+$	**?**	**?**	$?^?$	–	**No**	**Not seen (later observed)**
$\Xi_c(2930)^+$	2942 ± 5	**?**	$?^?$	–	No[3]	**Not seen**
$\Xi_c(2970)^+$	2964.3 ± 1.5	$20.9^{+2.4}_{-3.5}$	$1/2^+$ [23]	916 ± 55 [24]	Yes	Seen
$\Xi_c(3055)^+$	3055.9 ± 0.4	7.8 ± 1.9	$?^?$	894 [25]	Yes	**Not seen**
$\Xi_c(3080)^+$	3077.2 ± 0.4	3.6 ± 1.1	$?^?$	596 [25]	Yes	**Not seen (later observed)**

Table 1.5 Status of experimental measurements from the PDG [1] (or sources otherwise cited) for charmed Ξ_c^+ baryons.

The $\Xi_c(2815)^+$ state mass and width as recorded in the PDG currently draws from two Belle measurements [22, 24], but the spin and parity are not yet measured. The invariant mass distribution from Ref. [24] can be seen in Figure 1.4. As shown in the table, $\Xi_c(2923)^+$ has not been observed. The $\Xi_c(2923)^0$ has been observed by LHCb, with mass and width

$$\begin{aligned} m[\Xi_c(2923)^0] = &\quad 2923.04 \pm 0.25\,(\text{stat}) \pm 0.20\,(\text{syst}) \pm 0.14\,(\text{PDG})\,\text{MeV}, \\ \Gamma[\Xi_c(2923)^0] = &\quad 7.1 \pm 0.8\,(\text{stat}) \pm 1.8\,(\text{syst})\,\text{MeV}, \end{aligned}$$

as measured in Ref. [19].

The $\Xi_c(3080)^+$ mass average reported by the PDG is $3077.2 \pm 0.4\,\text{MeV}$, from largely consistent sources [25, 26]. The width average is reported as $3.6 \pm 1.1\,\text{MeV}$ from the same sources.

[3] What Belle reports as the $\Xi_c(2930)^+$ is most likely to be a superimposition of $\Xi_c(2923)^+$ and $\Xi_c(2930)^+$ states.

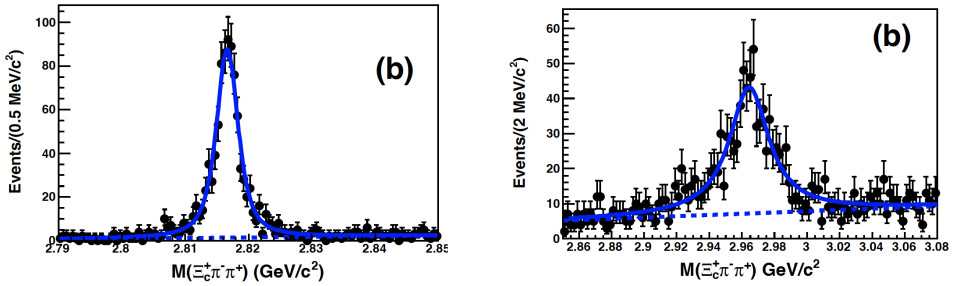

Figure 1.4 Invariant mass distributions of the $\Xi_c(2815)^+$ (left) and $\Xi_c(2970)^+$ (right) states, from recent studies in the $\Xi_c^+\pi^+\pi^-$ channel from the Belle Collaboration [24].

Width of the $\Xi_c(2970)^+$

A wide range of measurements are available regarding $\Xi_c(2970)^+$ mass and width, in particular from Belle [22, 24, 27] and with a contribution from BaBar [26]. The invariant mass plot for this channel from Ref. [24] is shown in Figure 1.4. A comparison plot of the various measurements is available in Figure 8.2.

There are conflicting width measurements in different channels, but overlapping mass measurements, raising the possibility that there are two states involved e.g. close in mass but with differing widths. The spin parity is reported to be $\frac{1}{2}^+$ by [23]. The PDG reports a mass (from fit) of $2964.3 \pm 1.5\,\text{MeV}$ but an average of $2967.1 \pm 1.4\,\text{MeV}$, and measurements up to $2975.9 \pm 1.5 \pm 2.1\,\text{MeV}$ [27].

Concerning the width, the PDG reports an average

$$\Gamma[\Xi_c(2970)^+] = 20.9^{+2.4}_{-3.5}\,\text{MeV}, \tag{1.12}$$

from Ref. [24, 27]. These measurements range from as low as $\Gamma[\Xi_c(2970)^+] = 18 \pm 6 \pm 7\,\text{MeV}$ reaching as high as $28.1 \pm 2.4^{+1.0}_{-5.0}\,\text{MeV}$ [24].

The analysis in Part II will provide new input for the mass and width (see also Figure 8.2). Further confirmation of the $\Xi_c(2970)^+$ parameters would be beneficial.

Chapter 2

The Large Hadron Collider beauty experiment

The Large Hadron Collider beauty experiment (LHCb) is one of the four large experiments of the Large Hadron Collider (LHC) at CERN, dedicated to the study of heavy flavour physics. LHCb was initially designed to search for indirect evidence for new physics, evidence of CP violation, and rare decays of beauty and charm hadrons [28], by exploiting high quantities of decays acquired from LHC pp collisions at up to 14 TeV. In Run 1–2, an integrated luminosity of 9 fb^{-1} was collected in total. In 2011–2012 alone, 10^{12} heavy flavour decays were recorded at a centre-of-mass energy of 7 TeV [29]. Owing to the successful operation and high performance of the LHCb detector, the experiment delivered a high volume of quality results. LHCb has demonstrated capabilities beyond the initial aims of the core physics programme [29], establishing itself as a General Purpose Detector (GPD) with additional successes in many areas such as hadron spectroscopy, to name one.

LHCb Run 2, a period of running of the LHC spanning years 2015–2018, saw an evolution of the trigger with the introduction of the Turbo paradigm (Section 2.2.7), adding further flexibility to the High Level Trigger and considerably expanding analysis possibilities. After the start of Long Shutdown 3 (LS3) in 2018 signalling the end of Run 2 (a long period during which the LHC is not operated for maintenance and upgrades), a wealth of analyses opportunities remained and continue to be investigated. The collaboration continues efforts in these areas, while also commissioning the detector for Run 3, and evolving the detector in preparation for challenging conditions of the High-Luminosity LHC (HL-LHC)

upgrade, expected in the 2030s.

In this chapter, the Run 1–2 era LHCb detector will be covered in Section 2.2, where the initial Detector Performance report [30] is used as a primary reference, unless otherwise cited. Additionally, the LHCb upgrade (currently in operation, at the time of writing) is briefly covered in Section 2.3, followed by a short description of LHCb Upgrade-II challenges, requirements outlined in Section 2.4.

2.1 The Large Hadron Collider and the accelerator complex at CERN

The LHC at the European Organisation for Nuclear Research (CERN) is the highest energy particle collider ever built to date. The machine delivers proton-proton collisions at centre-of-mass energies up to a design value of 14 TeV to particle physics experiments at four different locations (or 'intersection points') around the 27 km underground ring, and is also used to deliver heavy-ion collisions. The machine is part of the wider CERN accelerator complex, as displayed in Figure 2.1. The LHC delivers collisions to four large experiments: ATLAS in Meyrin (Geneva, Switzerland), ALICE in Saint-Genis (France), CMS in Cessy (France), and LHCb in Ferney-Voltaire (France).

The protons are accelerated in stages by a series of accelerators before reaching the LHC. The LHC is supplied by the Super Proton Synchrotron (SPS), an accelerator measuring approximately 7 km in circumference, providing protons at energies up to 450 GeV. The SPS is supplied with protons accelerated to 26 GeV by the Proton Synchrotron (PS), which accepts protons from the BOOSTER at 1.6 GeV. The LINAC-2 supplies protons to the BOOSTER at 50 MeV. From 2020 onwards, LINAC-2 was replaced with LINAC-4 which is now used in Run 3 beam operations.

The LHC beam comprises 2808 'bunches' of protons injected in to the machine with 1.15×10^{11} protons per bunch at the beginning of a nominal fill. There are two beams which circulate in opposite directions, directed by dipole magnets, focused by quadrupole magnets and other specialised beam optics. The spacing between proton bunches is 25 ns. A particle-free spacing ("beam abort gap") spacing of 3.0 µs is left in the beam structure to allow enough time for the field of the extraction kicker magnets to ramp up to their operating value, should a

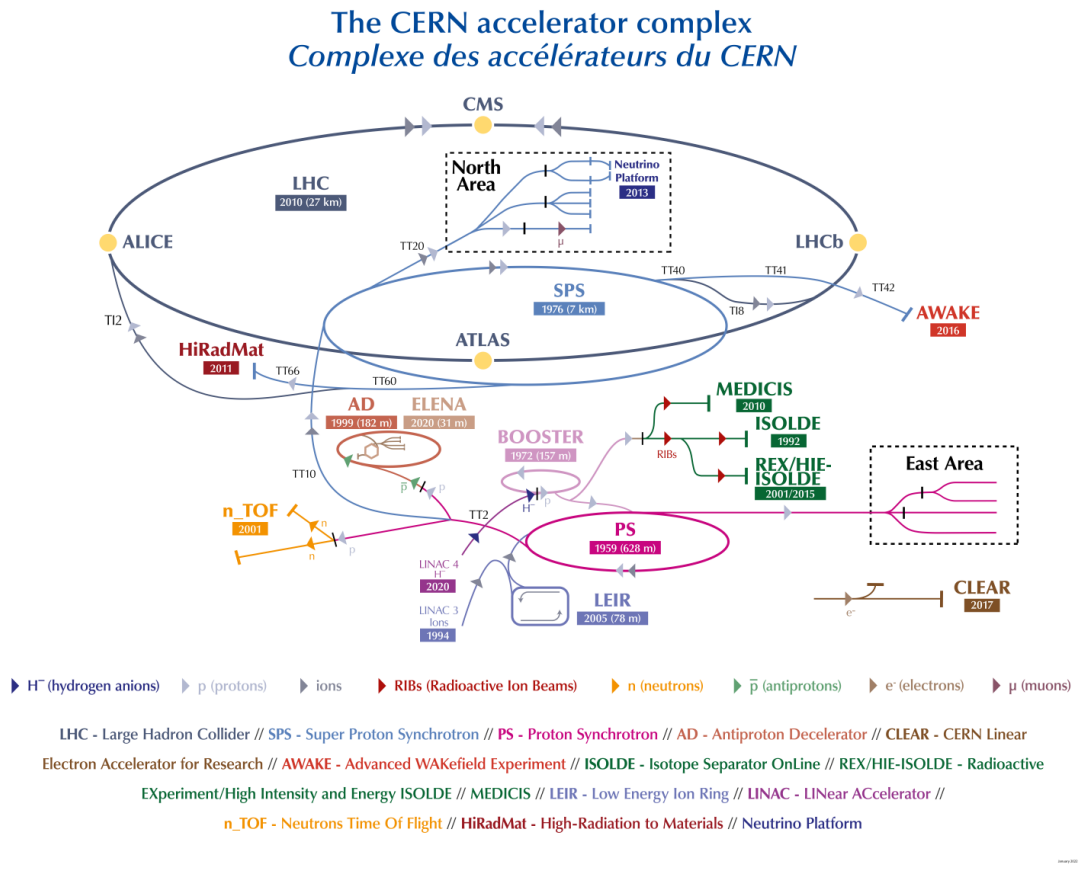

Figure 2.1 The CERN accelerator complex. The LHC is a part of a larger complex of accelerators at CERN, and is currently fed with protons via the LINAC-4, BOOSTER, PS, and SPS (before 2020, the PS was fed previously by the LINAC-2). Reproduced from Ref. [33].

beam dump be triggered [31]. The beam energy is 'ramped' after injection of all protons into the machine, up to a design value of 7 TeV. The nominal beam parameters are summarised in Table 2.1 of Ref. [32].

2.2 The LHCb detector

The LHCb detector is a single-arm spectrometer located approximately 100 m underground at LHC Point 8. The detector acceptance encompasses the forward region in the pseudorapidity range $2 < \eta < 5$ (pseudorapidity η is defined in the following subsection), as it is optimised for the detection of beauty and charm hadron decays. The LHCb geometry captures approximately half of all b and

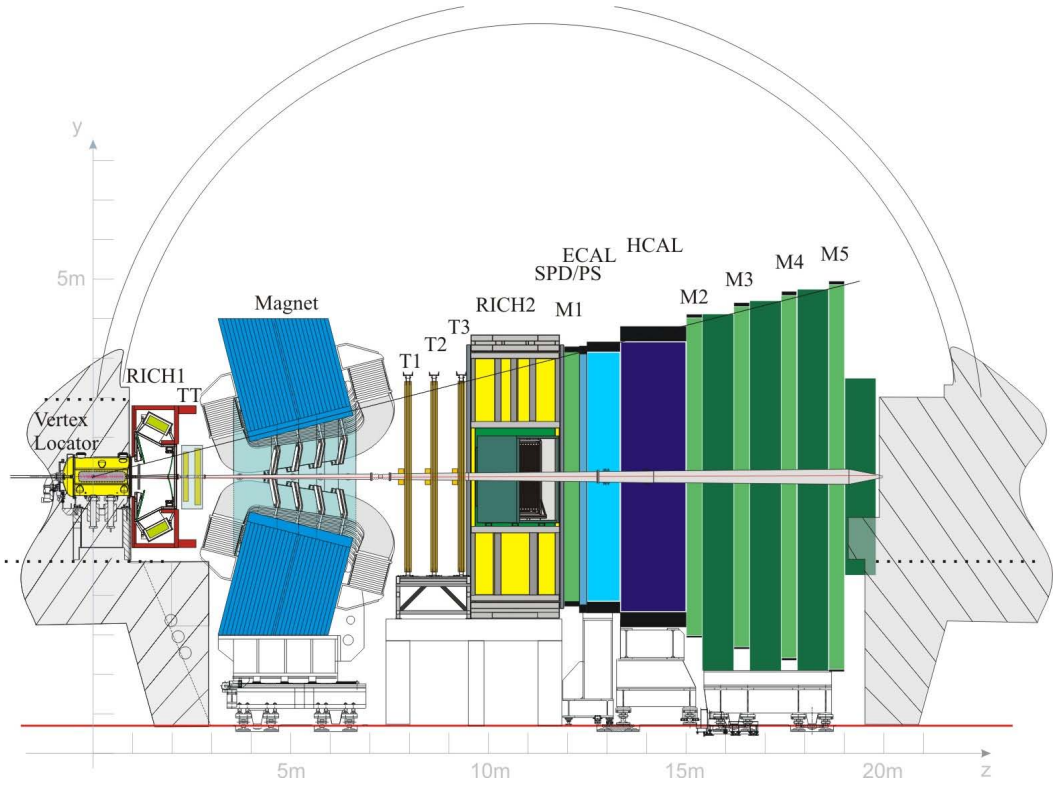

Figure 2.2 Layout of the LHCb detector in Run 1-2. Reproduced from Ref. [34].

c-hadrons emerging from the Interaction Point. The layout of the detector is summarised in Figure 2.2.

The detector is composed of seven subdetector systems: the Vertex Locator (VELO), providing crucial primary vertexing capabilities close to the Interaction Point; the tracking system (Trigger Turicensis, Inner Tracker and Outer Tracker); the Ring Imaging Cherenkov (RICH) detectors; the ECAL, HCAL calorimeters; and the MUON system.

2.2.1 Coordinate system

The LHCb coordinate system is right-handed, with the origin centred on the Interaction Point, the z-axis running along the beam axis increasing outward towards the MUON system, and the y-axis pointing upward.

The 'transverse' direction is any direction that is perpendicular to the beam line (i.e. in the x-y plane). Particle trajectories are sometimes expressed in coordinates of (p_T, η, ϕ). The transverse momentum (p_T) is the component of the

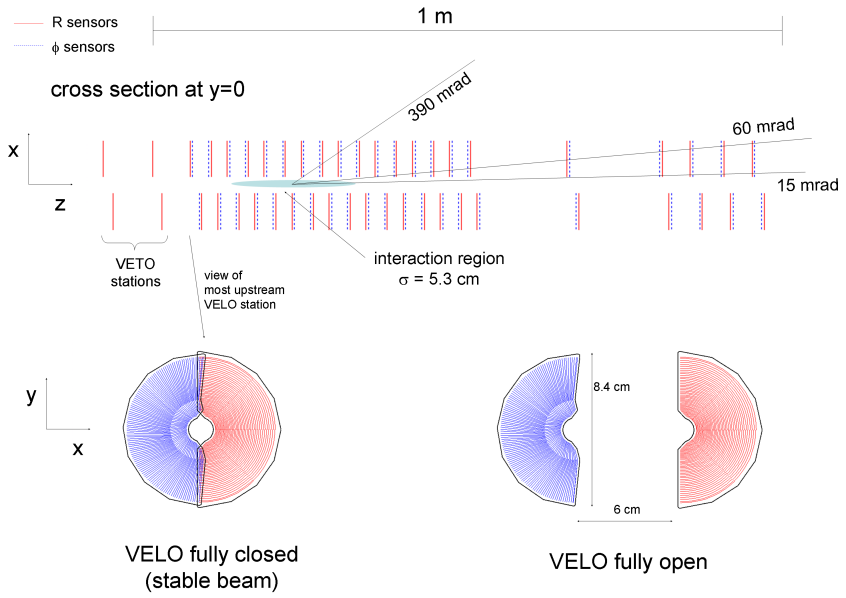

Figure 2.3 The VELO detector stations and halves in opened/closed positions. There is a slight overlap to allow for sensor misalignments. Reproduced from Ref. [30].

total momentum in the transverse direction. The pseudorapidity of a particle trajectory is defined,

$$\eta = -\ln \tan \frac{\theta}{2} \tag{2.1}$$

where θ is the angle between the total momentum and the positive direction of the beam axis. The ϕ coordinate is the polar angle of the particle trajectory in the transverse plane.

2.2.2 Vertex Locator (VELO)

For the experiment as it was in Run 1–2, the Vertex Locator (VELO) is a movable silicon strip vertex detector in two halves, comprising 21 modules per half making up 42 modules in total. The modules are oriented perpendicular to the beam axis. The detector surrounds the Interaction Point of the beam, where proton bunches collide. To avoid damage to the detector due to variations in beam position, it is kept in an 'open' position (see Figure 2.3) until stable beams are declared by the LHC. When closing, the detector can be moved such that sensors have only 7 mm separation from the beam [35].

The VELO is used to determine the location of primary interaction vertices

and secondary vertices, which are important features of beauty and charm hadron decays. Information from the VELO is used in the trigger system, track reconstruction, and can be used to determine particle decay times e.g. for particle lifetime measurements. A high vertexing resolution is a crucial component of LHCb's capabilities which are relied upon heavily in selection requirements of trigger lines and LHCb analyses. A resolution of 13 µm in the transverse plane and 71 µm in the beam direction was achieved in 2011 data [35].

2.2.3 Silicon and Outer Tracker

The Tracker Turicensis (TT) and the Inner Tracker (IT) comprise the Silicon Tracker system. The TT is a 150 cm × 130 cm plane stationed between the RICH1 and the Magnet covering the entire detector acceptance, and is illustrated in Figure 2.4. The IT covers a 120 cm × 40 cm cross-shaped region around the LHCb beam pipe in the centre of each of the T1–T3 tracking stations of the Outer Tracker. The TT and IT are instrumented with silicon microstrip sensors with a strip pitch of approximately 200 µm and spatial resolution requirement of 50 µm.

Each Silicon Tracker station is made up of four layers ($x-u-v-x$ geometry). The first and last layers are made up of vertical strips, with the two central layers rotated by angles $-5°$ and $5°$ respectively to the vertical.

The Outer Tracker planes T1–T3 are instrumented with drift straw tube technology, with 5 mm diameter straws arranged in layers — four layers per station in an $x-u-v-x$ configuration. The drift straw tubes contain a wire and

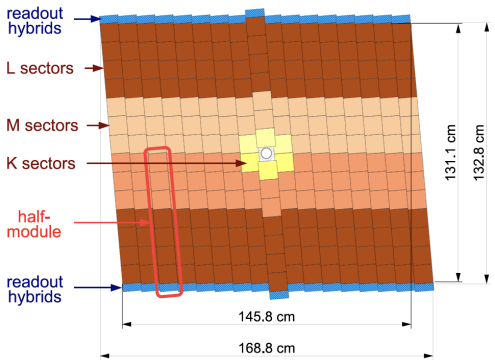

Figure 2.4 Transverse-plane view of the third layer of the TT. Reproduced from Ref. [30].

an 70-30% Argon-CO_2 gas mixture (to keep drift times below 50 ns). Incident particles liberate charge in this gas which drift to and are collected by the wire. A single hit resolution of 205 µm, close to the requirement of 200 µm, is achieved [35].

2.2.4 Ring Imaging Cherenkov (RICH) detectors

The Ring Imaging Cherenkov (RICH) detectors RICH1 (just after the VELO) and RICH2 (situated after T1–T3) are used for particle identification by making use of the Cherenkov effect. Cherenkov radiation is emitted when particles travel faster than the speed of light of the medium in which they are travelling, at an angle θ_c depending on the particle velocity and (known) material refractive index,

$$\cos\theta_c = \frac{c}{nv_p}, \tag{2.2}$$

where c is the speed of light in a vacuum, n is the material's refractive index, and v_p is the velocity of the incident particle. The known refractive indices of the detector active media and resulting ring shape from detector ring imaging can be used to determine the angle θ and used to calculate v_p. When combining the v_p prediction with momentum measurement from tracking, the particle identity can be inferred from the mass calculated as

$$m = \frac{np}{c}\cos\theta_c, \tag{2.3}$$

where m is the incident particle mass, and p is the particle momentum.

The RICH1 has a C_4F_{10} radiator, while RICH2 uses a CF_4 radiator. In Run 1 the RICH1 had an additional aerogel radiator that was eventually removed to improve the overall RICH reconstruction performance [36]. Photons produced from the radiators installed in the detectors are reflected by mirrors onto a network of Hybrid Photon Detectors (HPDs) which detect the Cherenkov photons and measure their spatial positions. The light is redirected outside of detector acceptance to preserve the low material budget. Figure 2.5 shows the momentum range coverage of the radiators, the angle-momentum dependence of the different particle species, and the reconstruction of the Cherenkov rings from the detector imaging.

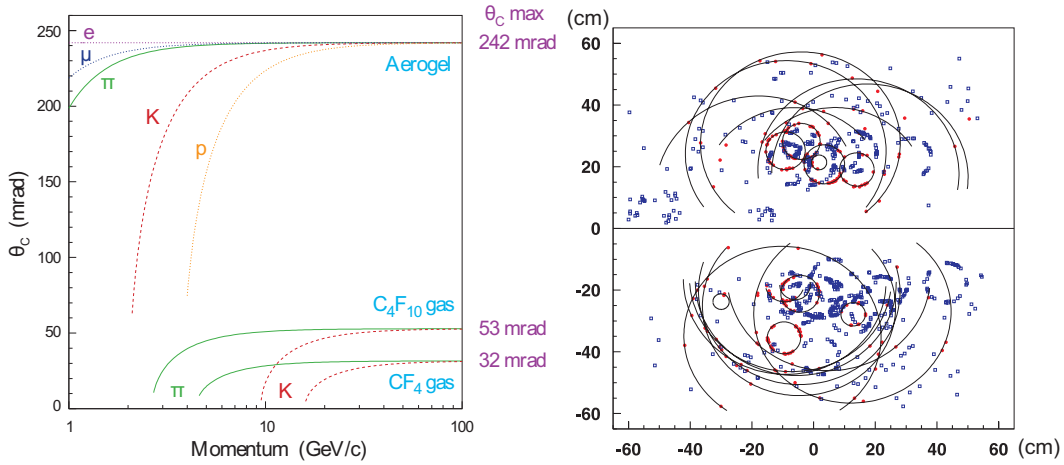

Figure 2.5 RICH Cherenkov angle vs. momentum for different particle species (left) and reconstruction of rings. The aerogel radiator in RICH1 covers the lower momentum region. Reproduced from Ref. [30].

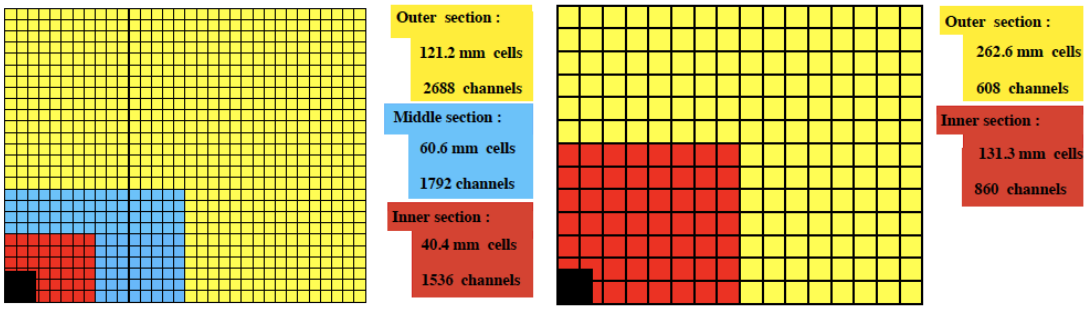

Figure 2.6 Calorimeter quadrants for the SPD/PS and ECAL on the left, and the HCAL on the right. Reproduced from Ref. [30].

2.2.5 Calorimetry

The calorimeter systems provide a crucial part of low-level fast triggering as well as being a key element of particle identification in certain cases. It is composed of the Pad/PreShower (SPD/PS), Hadronic (HCAL), and Electronic (ECAL) calorimeter systems, which detect and measure the energy and position of incident photons, electrons and hadrons. The PS/SPD uses multianode PMTs (MAPMTs), while ECAL and HCAL detectors use individual photo tubes.

The Shower Pad Detector (SPD) and PreShower (PS) help to distinguish between neutral and charged particles before entering the ECAL and HCAL detectors, rejecting key backgrounds, enabling the identification of π^0, electrons, and photons. The ECAL and HCAL both provide transverse energy measurements with the ECAL used to measure for electrons, photons, and neutral pions (π^0),

and the HCAL providing information about incident hadrons to the Level-0 (L0) trigger and enabling fast hardware triggering decisions.

All LHCb calorimeters use the same approach: light emitted by scintillation is detected by Photomultipliers (PMT) after transmission via wavelength-shifting fibres. The PS/SPD uses multianode PMTs (MAPMTs), while ECAL and HCAL detectors use individual photo tubes.

Owing to a large (two orders of magnitude) variation in hit density across calorimeter surfaces, different cell sizes are used to keep occupancy within manageable limits, with the smaller sizes closer to the beam pipe i.e. scaling transverse dimensions with increasing distance from the Interaction Point.

2.2.6 Muon chambers

The LHCb muon system comprises 1380 muon chambers over a total of five stations (labelled as M1–M5), covering a total area of $435\,\mathrm{m}^2$. The muon stations M2–M5 are situated after the ECAL and HCAL systems, with M1 situated just before them. The Muon system provides both muon identification and triggering for the experiment using multi-wire proportional chambers (MWPCs) and gas electron multipliers (Triple-GEM) chambers. The chambers close to the beam pipe are instrumented with Triple-GEM, in order to keep the occupancy within acceptable ranges, while the rest are instrumented with MWPCs. Between each of the stations M2–M5 are 80 cm thick iron absorbing material layers, and aligned hits are required in all stations for the muon trigger to fire, therefore ensuring the most penetrating muons are selected and a transverse momentum resolution of 20% is achieved.

Each station has the same number of chambers, but each station increases in size to cover the detector acceptance as distance from the Interaction Point increases. Thus, the final muon stations M4 and M5 have the worst spatial resolution, while M1–M3 have the best.

2.2.7 Trigger system

A trigger system is designed to reduce the event rate to a manageable level by only reading out the detector in the moments that meaningful indicators of physics are

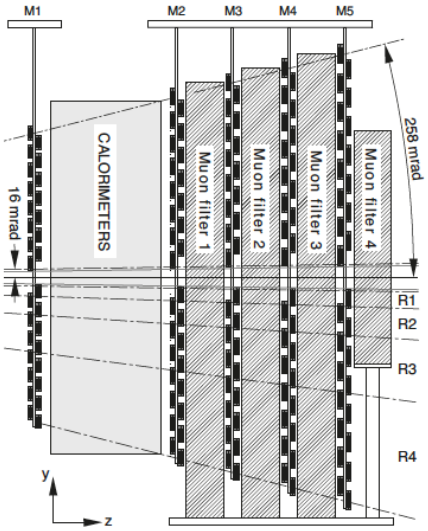

Figure 2.7 Side profile of MUON system. Reproduced from Ref. [30].

observed. At Point 8, there is a bunch crossing rate of 40 MHz, with a 25 ns time period between each LHC bunch crossing. With one event per bunch crossing, the 40 MHz event rate is reduced in steps using a combination of hardware and software-based systems that decide which events should be saved for later study. The complexity of the selection requirements are carefully balanced at each layer to ensure decisions are completed in time, avoiding the possibility of event buffer overflow and consequent data loss while maximising physics output.

- **Level-0 (L0) trigger**: A hardware-based trigger that reduces the event rate from 40 MHz to 1 MHz. Must be able to make decisions very quickly to deliver this 40x reduction in rate.

- **High Level Trigger 1 (HLT1)**: The first stage of the software trigger in which a fast event reconstruction is performed, and the event passed along to HLT2 if one or more selection lines, which are broad in their scope, have a positive decision.

- **High Level Trigger 2 (HLT2)**: A full reconstruction is performed at this step, with events stored offline should any of the configurable selection lines pass. This system has evolved to the point that a full calibration and detector alignment is performed preceding the reconstruction, making the trigger reconstruction equal in quality to the offline reconstruction. Portions of the full event can be selected to be saved, in order to significantly preserve storage and increase the available bandwidth.

The L0 trigger reads out from the Calorimeter, Muon, and VELO to ensure that the full detector is read out at a manageable rate. This is achieved by both: rejecting events in which combinatorial backgrounds are expected to be too high and would occupy too much bandwidth in the trigger; and triggering events with meaningful signatures of physics by using the calorimeter and muon systems. In the latter case, the L0 trigger searches for hadron, electron, or photon clusters with the highest transverse energy in the calorimeter system. The trigger also searches for two muons with the highest transverse momenta. To reject events that would consume too much bandwidth in the trigger system, the VELO estimates the number of pp interactions encountered in the bunch crossing, and the SPD also provides an estimate of the total number of tracks in the event.

Should the L0 trigger return a positive decision, the full detector is read out, and the event passed to the High Level Trigger. In HLT1, if high-quality tracks with high transverse momentum and/or large Impact Parameter (IP) are observed, the event is passed to HLT2. In HLT2, a wider and more diverse collection of selection lines performing inclusive and exclusive physics selections are checked using information from the full track reconstruction. If any decisions are positive, the event is saved to offline storage.

In Run 2, the turbo stream was introduced to HLT2, allowing full reconstruction of candidates in real-time at the cost of saving less information about the event, i.e. tracks and detector outputs that are not relevant to the decays of interest. This was made possible by significant improvements to the reconstruction software, an expansion of the computing resources available for HLT1 and HLT2 at the Event Filter Farm[1], and the introduction of alignment and calibration routines performed during online data-taking [38]. For a summary of the flow of data through the trigger in Run 2, refer to Figure 2.8.

2.3 LHCb Upgrade-I

The LHCb detector has been upgraded to meet demands of running at a higher instantaneous luminosity in Run 3 [40]. The layout of the detector is illustrated in Figure 2.9. This section is included in this thesis for completeness, as a very brief coverage of the changes, and the reader is encouraged to refer to Ref. [40]

[1] 800 new computing nodes were added in Run 2 making a total of ≈1700 nodes (≈27,000 cores). See the performance paper for more details [37].

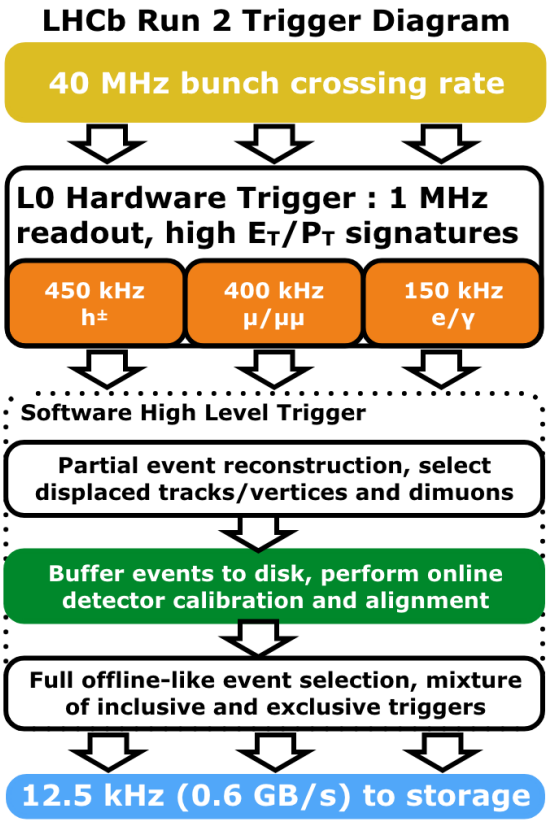

Figure 2.8 Run 2 trigger data flow diagram. Reproduced from Ref. [39].

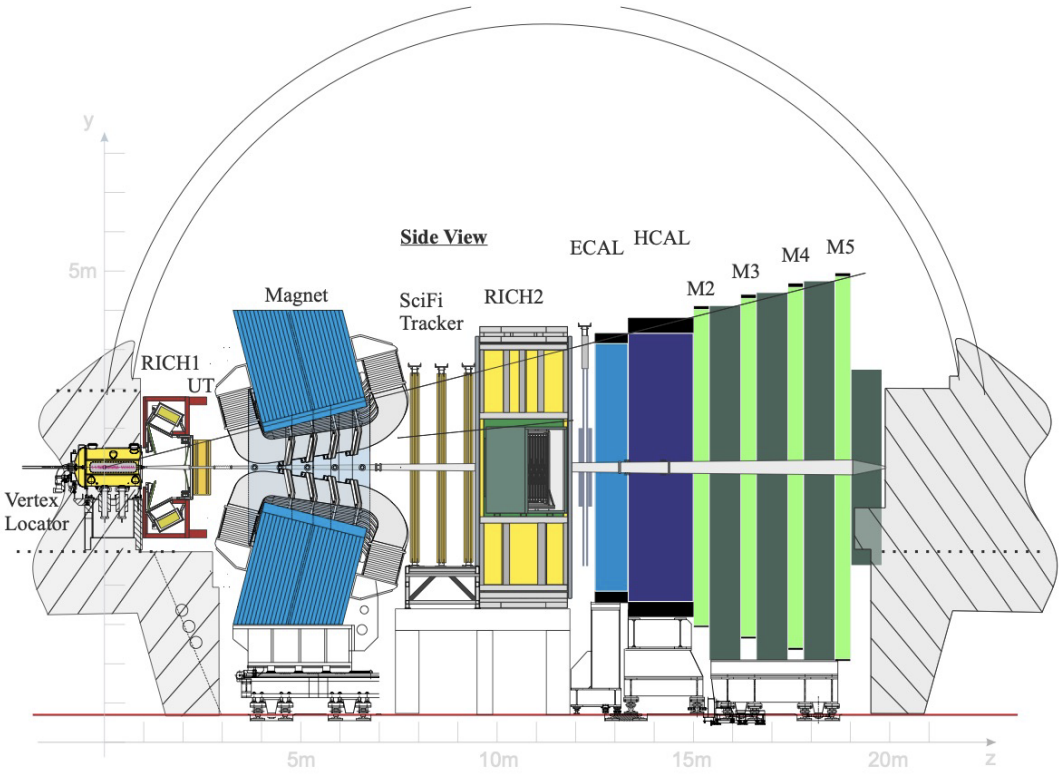

Figure 2.9 LHCb Upgrade detector. Reproduced from Ref. [40].

for a more comprehensive description of Upgrade-I. The author's involvements in Run 3 software development and service tasks are described in Part III.

Some key changes of note in this upgrade are the complete removal of the hardware (L0) trigger. The trigger system is replaced with a fully software-based trigger with HLT1 running fully on GPUs, capable of a full 40 MHz input event rate. Additionally, many of the subdetector systems have undergone a complete replacement. Notably, the tracking (VELO, UT, and Scintillating Fibre (SciFi)) and RICH systems.

2.4 LHCb Upgrade-II

During Long Shutdown 4 (LS4), the LHC will be upgraded to run at significantly higher instantaneous luminosity. This upgraded machine will be called the High-Luminosity Large Hadron Collider (HL-LHC), and for LHCb an order of magnitude increase in the instantaneous luminosities up to a maximum of

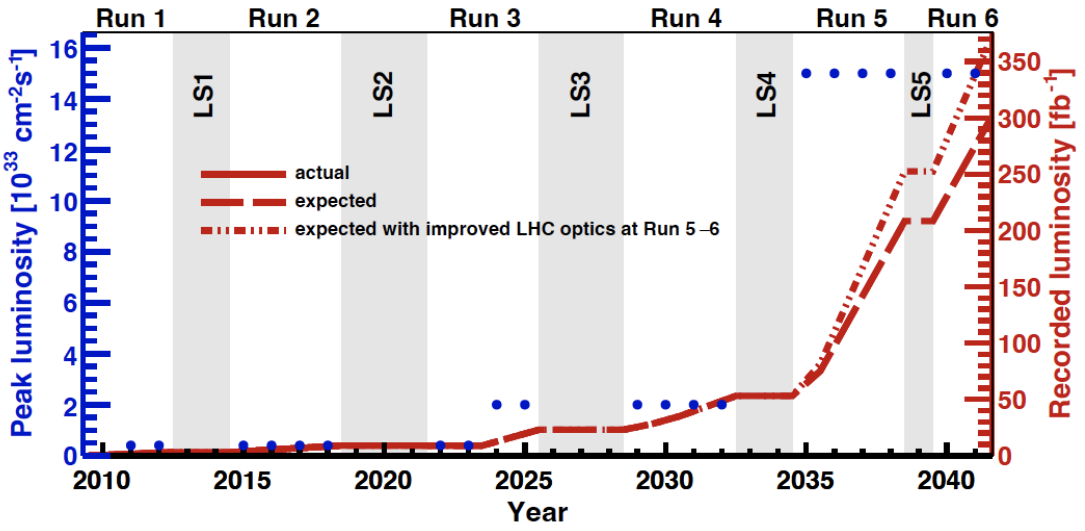

Figure 2.10 Expected increase in both integrated and instantaneous luminosity up to and beyond 2030. Based upon the official LHC schedule of 2021. Reproduced from Ref. [42].

$1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [41] is expected to increase the total integrated luminosity collected to 300 fb^{-1} or more, as shown in Figure 2.10.

These ambitious upgrades will bring about new opportunities for physics through vastly increased volumes of data, and new challenges for instrumentation:

- **More interactions per bunch crossing**: Approximately $\approx$ 40 interactions per bunch crossing are expected. To fully exploit this, existing detectors will need to be upgraded with higher granularity sensors and fast timing capabilities in order to deal with the higher track multiplicities and rates.

- **Higher radiation environment**: detectors will need to be sufficiently radiation-hard (resistant to radiation damage) to remain performant over the full run period and to reduce the cost of replacements.

The new detector layout is shown in Figure 2.11. The most relevant part of this upgrade to the author's work is the Mighty Tracker (MT), the new tracking system, which will be instrumented with silicon and the SciFi technology of Upgrade-I.

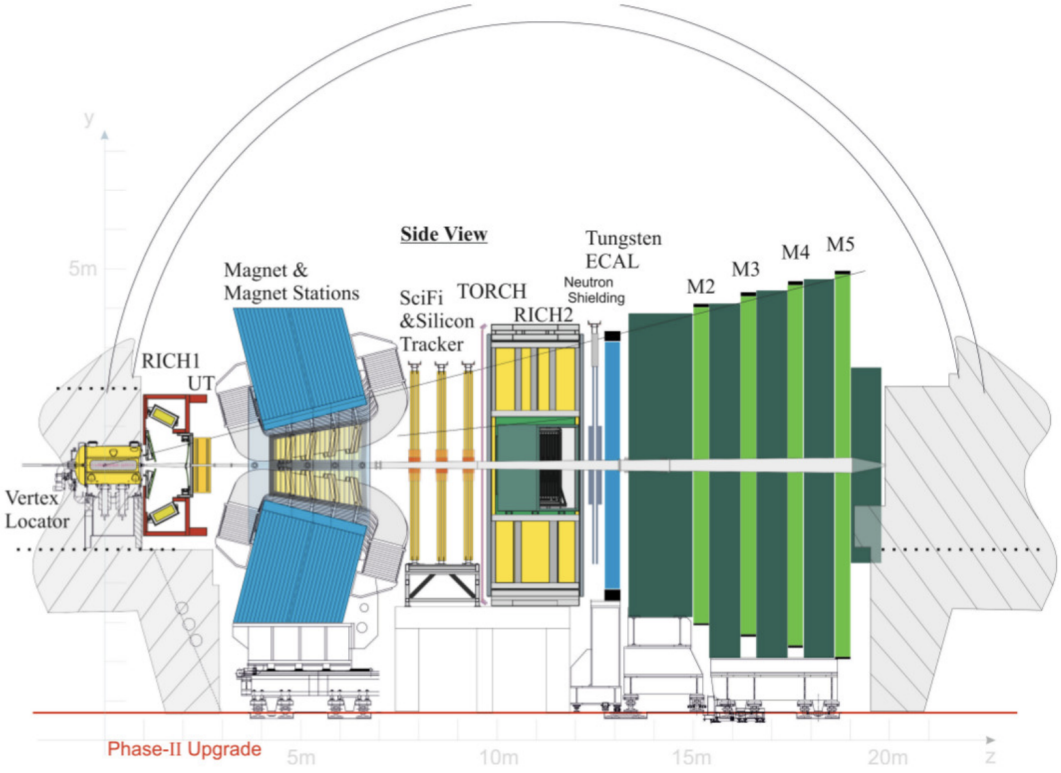

Figure 2.11 LHCb Upgrade-II detector layout. Reproduced from Ref. [41].

2.4.1 A Hybrid Tracking Solution: The Mighty Tracker

The currently installed tracking system of the Upgrade-I detector is the SciFi (scintillating fibre detectors read out by SiPMs). In an Upgrade-II scenario, the SciFi would require frequent replacements, due to accumulated radiation damage and degraded reconstruction performance over the data-taking period. Additionally, detector occupancy and ghost rates (Figure 2.12) estimated from simulations [41] demonstrate issues with pattern recognition and subsequently track reconstruction to be expected in the region close to the beam pipe. The ghost rate represents the fraction of reconstructed tracks (fake tracks or 'ghosts') that do not correspond to an actual particle traversing the detector. A ghost track can be determined by attempting to match a reconstructed track trajectory to a generated Monte Carlo particle from the simulation, if no match is possible, the track is designated a ghost. Moreover, high radiation doses are expected in this inner region, which would necessitate frequent and expensive replacements of the active detector material. As such, this has led towards the adoption of a hybrid approach in which this lower occupancy outer region is proposed to be instrumented with the existing SciFi detector concept, and the region close to the beam pipe instrumented with a higher-granularity, radiation-tolerant silicon technology. This tracking system for LHCb Upgrade-II is named the Mighty Tracker (MT).

The baseline design for the MT is three stations with four planes of scintillating fibre detector (SciFi) planes per station, with the SciFi planes arranged in an $x - u - v - x$ geometry with the first and last planes oriented vertically, and the middle planes rotated $-5°$ and $5°$ respectively [42]. To maintain Upgrade-I performance in light of HL-LHC upgrades, the central (inner, middle) region surrounding the beam pipe is to be instrumented with silicon modules. The SciFi planes are interleaved with these silicon modules, such that they are placed at each end of the station and between each plane, making 6 modules in total. A diagram of the Mighty Tracker stations, showing inner (red, silicon), middle (blue, silicon), and outer (beige, SciFi) regions is shown in Figure 2.13. A variety of MT designs are still under consideration, which may affect the final detector design in Ref. [42].

The silicon modules, represented by the rectangles in Figure 2.13, are instrumented using High Voltage Monolithic Active Pixel Sensor (HV-MAPS) sensors, referred to as the MightyPix, which are to cover the active area in the layout

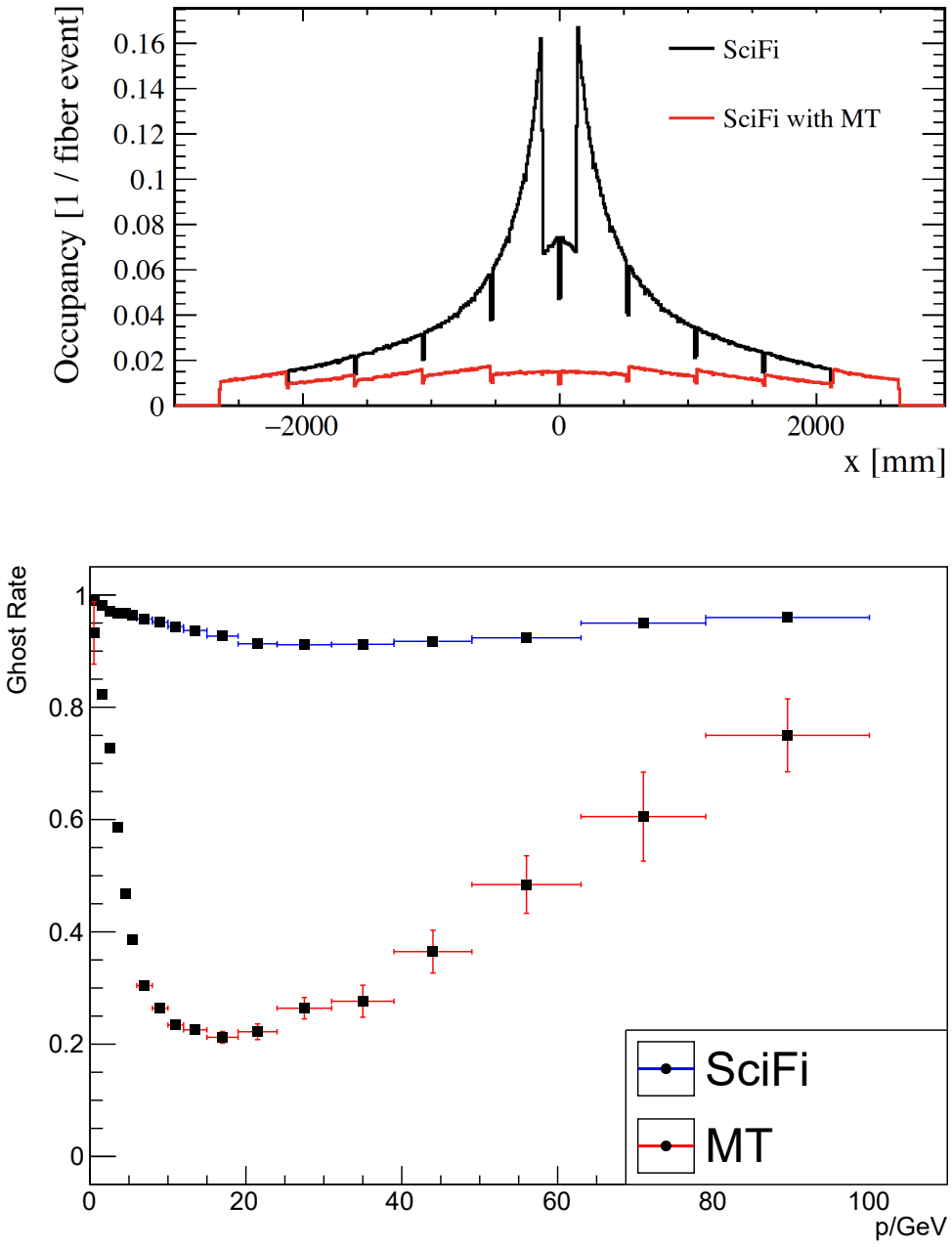

Figure 2.12 Simulated occupancies (top) and ghost rates (bottom) for Upgrade-II Mighty Tracker vs. SciFi scenarios. Reproduced from Ref. [41].

described in Figure 2.14. The design choices and requirements to be met by the MightyPix, which is still under development, will be described in the following section, and in the introductory text of Chapter 9, which covers the author's contributions in this area.

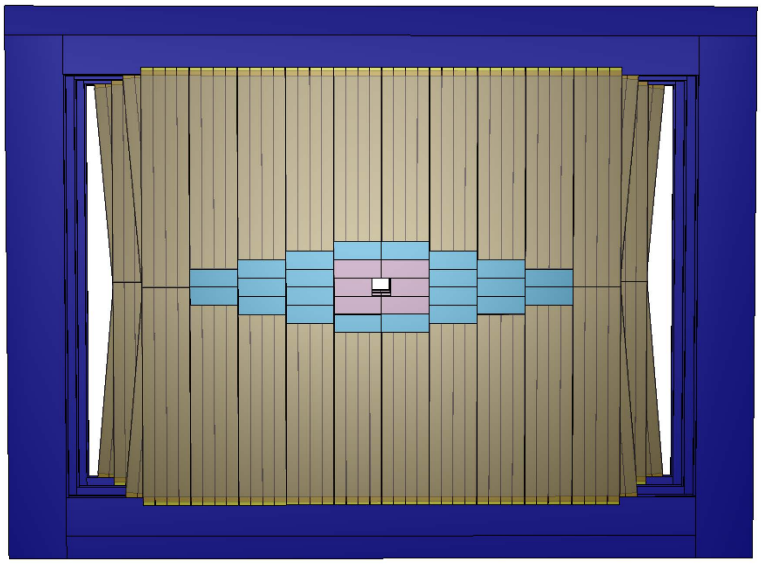

Figure 2.13 Transverse-plane view of the Mighty Tracker stations showing both SciFi and HV-MAPS components in the inner region. Reproduced from Ref. [41].

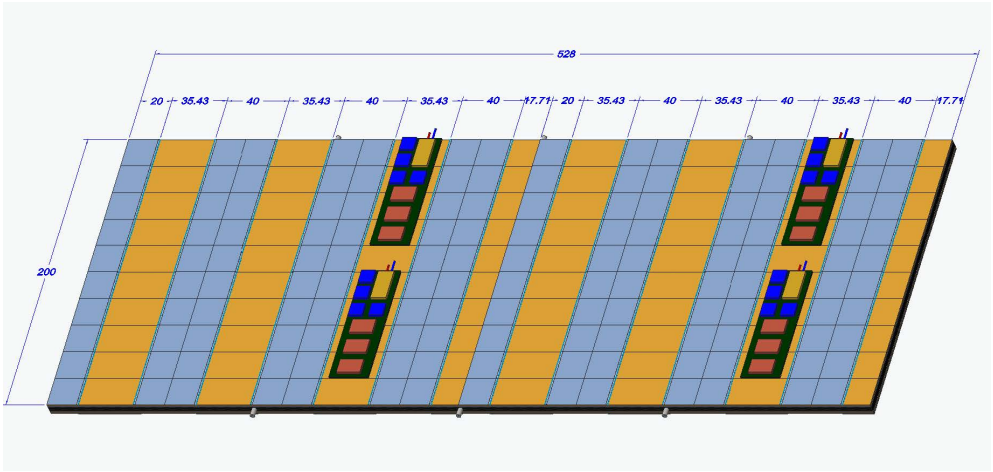

Figure 2.14 Diagram of a single HV-MAPS module of the Mighty Tracker. The blue areas show the active detector material, with the orange areas housing the electronics. The reverse side of the displayed module is arranged in an offset arrangement such that the full plane is instrumented with HV-MAPS. Reproduced from Ref. [41].

2.4.2 HV-MAPS

A particle depositing a charge in some semiconductor material (i.e. silicon) liberates a number of electron-hole pairs (charge carriers) proportional to the charge deposited in the material. Applying an electric field across the material causes these charge carriers to move to their respective electrode once liberated, inducing an electric current which can be amplified and measured. The semiconductor material (with atomic number Z) may behave in this way due to *doping*, a process wherein impurities are introduced to the material. Doping with a material of $Z + 1$ atomic number provides an availability of mobile negative charge carriers (electrons), while in the case of $Z - 1$, mobile positive charge carriers (holes). For example, starting with p-doped silicon and introducing an n-dopant from the surface creates a 'pn-junction' [43].

In silicon pixel detectors, silicon semiconductor materials are arranged in a segmented pixel configuration, providing 2D (x, y) position coordinates over the area covered by these pixels. The readout electronics would need to collect the current from each individual pixel, amplify, measure the signal, and record the event with coordinates, timestamp, et cetera. One way of achieving this is to bond the electrodes of each silicon pixel to the readout electronics keeping the silicon and electronic components initially separate, the approach of hybrid pixel sensors. However, a key drawback of this approach is the number of bonds and the extra material required, which could increase multiple scattering effects in the full detector considerably. If the sensors are fabricated such that the readout electronics are integrated into the pixels themselves, the total material and the number of required bonds can be reduced, which is the approach of monolithic pixel sensors [43].

For LHCb Upgrade-II, it has been proposed in the baseline design to instrument with depleted silicon pixel sensors, or Depleted Monolithic Active Pixel Sensors (DMAPS) [42]. Depletion refers to the property where electrons and holes in the active silicon are separated due to an applied electric field (by applying a bias voltage), creating a wider depletion region between the p- and n-doped material. Depleted active silicon will have a faster charge collection (an improved time resolution) and improved radiation tolerance compared to non-depleted sensors. Non-depleted active silicon does not have this depletion region, and thus charge carriers diffuse to their respective electrodes, resulting in a slower response time [44].

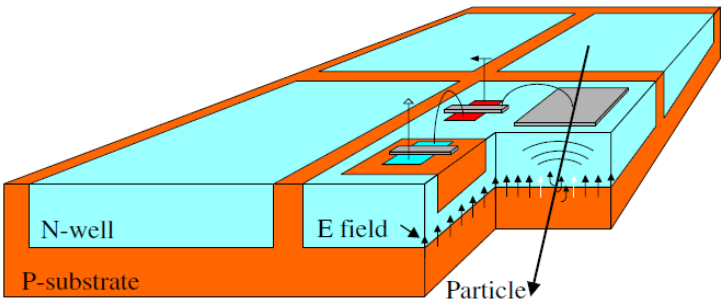

Figure 2.15 Sketch showing four silicon pixels in the HV-MAPS concept. An incident particle liberates a charge within the depleted active material, which induces a current signal through the readout electronics sitting in a deep n-well, which can amplify the signal and perform signal processing. Reproduced from Ref. [46].

Commerical High Voltage Complimentary Metal-Oxide-Semiconductor (HV-CMOS) processes have been used to design and create HV-MAPS such as the ATLASPix3 [44] leading to the ATLASPix3.1 and MuPix series of sensors [45]. A demonstration of the pixel layout for this sensor technology is described in Figure 2.15.

The ATLASPix3.1 is used for the studies described in Chapter 9, as the MightyPix1 was not available at the time. The requirements for a final MightyPix product for the Upgrade-II detector will also be described.

Part II

Observation of Ξ_c^+ baryons decaying to $\Xi_c^+\pi^-\pi^+$

Chapter 3

Introduction and strategy

Baryon spectroscopy, the study of the masses and decays of baryons, presents a valuable experimental testing ground for low energy quantum chromodynamics (QCD) which describes baryon properties such as mass, width, and quantum numbers. Experimental searches bearing fresh discoveries and measurements of excited baryon states can provide additional constraining power when developing and refining QCD calculations, assisting theoretical efforts in these areas.

LHCb has demonstrated over Run 2 the ability of the detector to carry out high precision studies of excited baryon spectra, further enhanced by availability of high statistics. Measurement of useful parameters from newly discovered and existing excited states can be attempted in a large variety of channels. A host of these decay channels have already been studied by LHCb.

3.1 Analysis approach

The analysis presented in this part of the thesis aims to search for excited Ξ_c^+ states in the $\Xi_c^+\pi^-\pi^+$ channel,

$$\begin{aligned} \Xi_c^{**+} \;&\longrightarrow\; \Xi_c(2645)^0\pi^+ \\ &\hookrightarrow \Xi_c^+\left(\to pK^-\pi^+\right)\;\pi^-, \end{aligned} \tag{3.1}$$

where the $\Xi_c(2645)^0$ is an intermediate resonance. Charge-conjugate decays are always implicitly considered, unless otherwise specified.

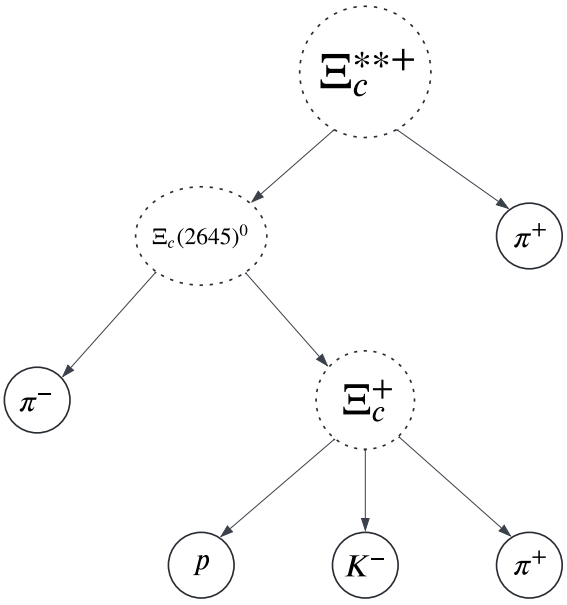

Figure 3.1 Tree representation of reconstructed decay. Nodes with solid lines represent tracks of final state particles, and dashed lines are candidates created from 2-body or 3-body combinations of child nodes.

This decay can be represented in the form of a 'decay tree', as shown in Figure 3.1. The first step in the selection is to select Ξ_c^+ baryon candidates from the relevant inclusive HLT2 line $\Xi_c^+ \to pK^-\pi^+$. The line is considered 'inclusive' as it intends to allow a wide selection of decay channels to be reconstructed by saving additional information within the event, such as extra pion or kaon tracks, in addition to the candidate that triggered the line. A Ξ_c^+ candidate is combined at the production vertex with a single negative pion (π^-) to form a $\Xi_c(2645)^0$ candidate, followed by another 2-body combination of $\Xi_c(2645)^0$ with a positive pion (π^+). These two pions are referred to as *bachelor pions* later. As relatively large numbers of pions are expected to be saved in each event, it is necessary to reduce candidate multiplicity and some basic kinematic cuts are introduced in Section 4.2. Until further selection requirements are introduced, Ξ_c^{**+} candidates may include both direct decays $\Xi_c^{**+} \to \Xi_c^+\pi^-\pi^+$, resonant decays $\Xi_c^{**+} \to \Xi_c(2645)^0\pi^+$, and combinatorial backgrounds (which are expected to dominate).

Each of the reconstructed Ξ_c^{**+} candidates then undergo three stages of offline selection requirements, including the use of multivariate analyses (MVAs).

- **Stage I — Ξ_c^+ optimisation:** rectangular cuts, MVA (optimising for maximum significance), and Ξ_c^+ mass window (Sections 4.1 and 4.3).
- **Stage II — $\Xi_c(2645)^0$ optimisation:** $\Xi_c(2645)^0$ MVA (optimising for maximum significance) (Section 4.4).
- **Stage III — Ξ_c^{**+} optimisation:** $\Xi_c(2645)^0$ mass window, MVA (optimising for Punzi Figure of Merit (PFoM)) & clone removal (Section 4.5).

See Figure 3.2 for an illustrative diagram. For each MVA a variety of classifiers and frameworks were investigated, and the one giving the best performance was chosen. In some cases, some libraries performed better than others as they were either able to provide more customisation of the models, were better suited to input sample type (e.g. sWeighted data vs. non-sWeighted MC and WS data), or were generally more performant with regard to implementation details. This means that different algorithms are used across the three stages.

Three MVAs are trained and used to maximise the significance of genuine Ξ_c^+, $\Xi_c(2645)^0$, and Ξ_c^{**+} candidates respectively in stages I, II, and III. As there are large Ξ_c^+ and $\Xi_c(2645)^0$ signals present in the data, stage one and two MVAs are trained using data with the *sPlot* [47] technique, where the resulting "sWeights" are used to train MVAs to distinguish between signal and background candidates. *sPlot* is a statistical technique enabling the analysis of variable distributions for a data sample comprising events originating from different processes e.g. a "signal" process and "background" process, on the basis of a discriminating variable, such as the invariant mass. Effectively, with this method the e.g. signal and background components of the data sample can be reconstructed separately. A crucial rule is that the variables studied via this method must not be correlated with the discriminating variable (i.e. they are statistically independent). Employing the *sPlot* technique requires a good description of the data sample by way of a fit model that appropriately describes the processes in the given data sample, a well-behaving fit to the discriminating variable, and fit covariance matrix. The results of this fit are used to generate a set of "sWeights" for each component, which can be used to reconstruct variable distributions for the e.g. signal and background components e.g. by filling a weighted histogram. Each set of sWeights when summed equal the corresponding fit component area, and may additionally contain negative weight values. The sWeight is related to the probability that the corresponding entry originates from e.g. the signal or background component. When using the signal sWeights to

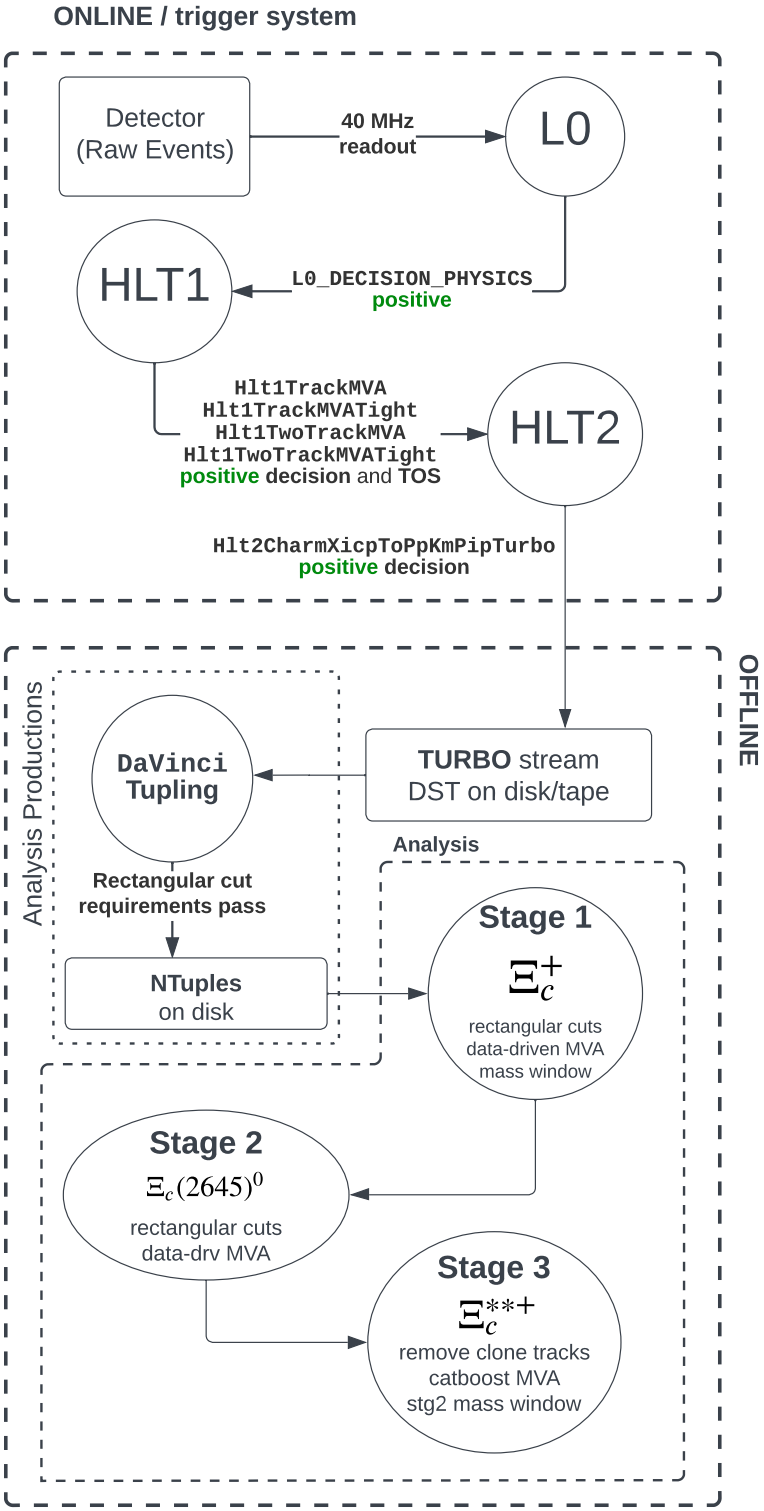

Figure 3.2 An illustration of the flow of events through the relevant processing steps and selection requirements.

construct the signal distribution, the "background-like" entries in the sample tend to be downweighted, while "signal-like" entries tend to have positive weights close to or greater than one. The main takeaway is that plotting a histogram having weighted the data sample by the given set of sWeights for a component, will reconstruct the variable distribution for that component. In this analysis, *sPlot* is used in the training of MVAs by including the weights in the training process. For example, in Section 4.3 the signal sWeight is used directly in the loss function Equation (4.4).

The final state (stage III) MVA is trained using a combination of simulated signal samples detailed in Section 3.2, and wrong-sign pion data as proxy for background. The term "wrong-sign" refers to the use of particle combinations that do not conserve charge, hence the 'wrong' sign is used. Samples of wrong-sign data are useful for understanding combinatorial background arising from random particle combinations that do not correspond to the genuine signal process of interest. Wrong-sign candidates are important for understanding the impact of selections on the background shape, as the expected dominant background contribution in the right-sign channel originates from combinatorics (background arising from random particle combinations). Each selection algorithm is optimised according to the appropriate selection performance metrics, as detailed in their respective sections.

The right-sign channel $\Xi_c^+\pi^-\pi^+$ was blinded during development of all selection requirements. Signal shapes were determined from simulation samples (Chapter 5) and background shapes from wrong-sign channels $\Xi_c^+\pi^-\pi^-$ and $\Xi_c^+\pi^+\pi^+$. After unblinding, a fit model is determined in Section 6.1. Systematic uncertainties (Chapter 7) are evaluated on all measurements, based on choices and/or limitations of the analysis method.

3.2 Details of utilised data

3.2.1 Real Data

Run 2 data spanning the years 2016–2018 were used, merging samples of all years and magnet polarities. The amount of data used corresponds to $5.6\,\text{fb}^{-1}$ integrated luminosity. The data samples were produced using the DIRAC [48] production system and LHCb Analysis Productions [49,50] (Section 10.2), which

ensures that all configuration metadata and produced n-tuples are preserved. Run 2 data from the year 2015 do not contain the Turbo line streams needed for the analysis, in part due to non-finalised line configurations and selections during that data taking period.

The Ξ_c^+ candidates were reconstructed online in the High Level Trigger (HLT2) after passing L0, HLT1, and the requirements of an inclusive *Turbo* selection line `Hlt2CharmHadXicpToPpKmPipTurbo`. A momentum scaling calibration [51] is additionally applied to the Ξ_c^+ candidates, and the systematic contributions of this calibration are discussed in Section 7.1.4.

Further details of the selection are outlined in Chapter 4.

3.2.2 Simulation samples

Four Monte Carlo simulation samples, summarised in Table 3.1, were created to facilitate estimation of the $\Xi_c(2645)^0$ (Section 5.2) mass resolution, and Ξ_c^{**+} mass resolution at various masses (Section 5.3). The samples were filtered through a realistic emulation of the HLT2 trigger requirements (later discussed in Section 4.1). The samples were also used as proxy for signal when training the stage III selection (Section 4.5).

Events with a Ξ_c^{**+} resonance were generated at masses of currently known states for both *MagUp* and *MagDown* dipole magnet polarities, and the event generator was configured to allow two possible decay paths:

- $\Xi_c^{**+} \to \Xi_c^+ \pi^- \pi^+$ (non-resonant, 50% fraction)
- $\Xi_c^{**+} \to (\Xi_c(2645)^0 \to \Xi_c^+ \pi^-)\, \pi^+$ (resonant, 50% fraction)

In the simulated event, *pp*-collisions are generated by Pythia [52] configured for LHCb [53]. Unstable particle decays are handled by EvtGen [54], and final-state radiation is simulated with Photos [55]. The Geant4 framework [56] is used to simulate the interactions of particles with a model of the LHCb detector, and the detector response [57].

Each generated event is selected according to the generator-level requirements outlined in Table 3.2, these were chosen specifically to ensure that generated MC events have all tracks within LHCb detector acceptance, and have a high

enough p_T to pass HLT2 requirements. Due to the high CPU time per generated event, a side effect of imposing these generator-level requirements, the ReDecay fast-simulation option was used [58].

Additionally, as this was a HLT2-filtered MC production, generated events were retained in the final output only if $\Xi_c^+ \to pK^-\pi^+$ HLT2 line requirements (shown in Table 4.1) were satisfied after running the full reconstruction chain. From the information provided in Table 3.1, approximately 3 % of generated events are retained after the HLT2 filtering step.

3.2.3 Analysis Preservation

Care was taken to ensure that all steps of the analysis are well-defined and preserved. The n-tuples used in the analysis to develop the offline selections discussed in Chapter 4 were produced using the LHCb DaVinci application. The DPA-WP2 Analysis Productions [49, 50] service was used to configure the large-scale data production, which was run by the Dirac Transformation System [48]. The produced n-tuples are stored on LHCb EOS storage, which is accessible to the entire collaboration and will remain preserved alongside the configuration that produced the data. Further information on the Analysis Productions system and n-tuple production is available in Chapter 10.

The main offline analysis workflow is defined in the `snakemake` [59] workflow definition language, with additional ad-hoc analysis scripts available for further studies. Software dependencies are installed from the `conda-forge` package repository, and managed by the `conda` package manager. The environment has been preserved using `lb-conda`, an LHCb service which preserves the conda environment in a git repository, making it available on CVMFS for future use

State	Event Type	m (MeV)	Γ (MeV)	# generated	# after filtering (MU/MD)
$\Xi_c(2815)^+$	`26265072`	2816.74	2.43	3,000,000	89,430 / 88,440
$\Xi_c(2923)^+$	`26265073`	2923.04	7.1	3,000,000	82,568 / 85,007
$\Xi_c(2970)^+$	`26265074`	2966.34	20.9	3,000,000	85,450 / 89,245
$\Xi_c(3055)^+$	`26265075`	3055.90	7.8	3,000,000	85,826 / 85,513

Table 3.1 Details of generated HLT2-filtered simulation samples, including the number of generated events, and events remaining after HLT2 filtering. The *Event Type* is a number used within LHCb encapsulating the details of the generated decay. The generated mass (m) and width (Γ) values used are taken from the PDG [1].

Particle (and charge-conj.)	TightCut Requirements
Ξ_c^+	$p_{\mathrm{T}} > 0.9\,\mathrm{GeV}$
p	$0.005 < \theta < 0.4$ $1.95 < \eta < 5.05$ $p_{\mathrm{T}} > 220\,\mathrm{MeV}$ $p > 9\,\mathrm{GeV}$
K^-	$0.005 < \theta < 0.4$ $1.95 < \eta < 5.05$ $p_{\mathrm{T}} > 220\,\mathrm{MeV}$ $p > 3\,\mathrm{GeV}$
π^+	$0.005 < \theta < 0.4$ $1.95 < \eta < 5.05$ $p_{\mathrm{T}} > 180\,\mathrm{MeV}$ $p > 3\,\mathrm{GeV}$

Table 3.2 Generator-level requirements used for all simulation samples. All variables referred to in the table are generator-level quantities.

and aiding analysis reproducibility. Analysis code is versioned in a GIT repository, on the CERN GitLab system [60].

Chapter 4

Selection requirements

This chapter will describe the blinded selection strategy for the $\Xi_c^{**+} \to \Xi_c^+ \pi^- \pi^+$ analysis. The selection occurs in stages, beginning in the LHCb trigger system (to follow) where Ξ_c^+ candidates are created and selected in real time. This is followed by creation of offline n-tuples, which involve additional selections when creating Ξ_c^{**+} candidates, with the aim to suppress combinatorial backgrounds. The n-tuples are later reduced in a three-stage offline selection process (labelled stage I, stage II, and stage III) leveraging Machine Learning models each time to have the best chance of observing new states.

4.1 Trigger

This analysis makes use of the large number of candidates collected in the Run 2 period 2016–2018 by a HLT2 trigger line specifically designed to collect $\Xi_c^+ \to pK^-\pi^+$ decays.

The selection requirements that must be met for a Ξ_c^+ candidate to be saved by the HLT2 selection line are summarised in Table 4.1. Additionally, L0 and HLT1 requirements are implicitly part of this selection, however there are no requirements for specific L0 or HLT1 requirements to pass as the analysis selection takes an inclusive blanket approach to selecting Ξ_c^+ candidates wherever they might emerge in the LHCb trigger.

Three good quality tracks with a high transverse momentum (p_T) that have not originated from close to any pp interaction primary vertex are combined to form

Ξ_c^+ baryon candidates. This is achieved in effect with a transverse momentum cut, $p_T > 200$ MeV, and $\min(\chi^2_{IP}) > 6$ cut on the Ξ_c^+ daughter tracks (p, K^-, and π^+). The latter refers to the minimum impact parameter χ^2 (χ^2_{IP}) value with respect to all primary vertices, which must be greater than 6. The impact parameter χ^2 in this context is a measure of how likely the daughter track originated from a given primary vertex, and is defined as the increase in the primary vertex fit χ^2 with the corresponding track included [61].

Following the combination, a series of additional requirements are imposed such that the candidate falls within an invariant mass window $2392 - 2543\,\text{MeV}$, the sum of the daughter transverse momenta is greater than 3 GeV, the DIRA cosine is greater than cos 1.57°, and minimum decay lifetime requirement of 0.15 ps is met. The most detailed summary of requirements is in Table 4.1.

Particle(s)	**Selection Criteria**
K^-, p, π^+	$p_\text{T} > 200\,\text{MeV}$ $\chi^2_\text{MINIP}(\text{PVs}) > 6.0$
$(\Xi_c^+$ OR $\Lambda_c^+) \to pK^-\pi^+$ combiner	$2201 < m(pK^-\pi^+) < 2553\,\text{MeV}$ $\sum_{dau} p_\text{T} > 3000\,\text{MeV}$ $N(p_\text{T} > 400) > 2$ $N(\chi^2_\text{MINIP}(\text{PVs}) > 16.0) > 0$ $N(\chi^2_\text{MINIP}(\text{PVs}) > 9.0) > 1$ $\chi^2_{endvertex}/N_{dof} < 10.0$ BPVDIRA $> \cos(1.570796°)$ BPVLTIME > 0.15 ps
Ξ_c^+ only (filter)	$2392 < m < 2543\,\text{MeV}$ $\chi^2_{endvertex}/N_{dof} < 10.0$ BPVDIRA $> \cos(1.570796°)$ BPVLTIME > 0.15 ps **Daughter particles** (p, K^-, π^+) $\sum p_\text{T} > 3000\,\text{MeV}$ $\max(p_\text{T}) > 1000\,\text{MeV}$ $N(p_\text{T} > 9.0\,\text{MeV}) > 1$ $\min(p_\text{T}) > 200\,\text{MeV}$ $\min(\chi^2_\text{MINIP}(\text{PVs})) > 6.0$ $\max(\chi^2_\text{MINIP}(\text{PVs})) > 16.0$ $N(\chi^2_\text{MINIP}(\text{PVs}) > 9.0) > 1$

Table 4.1 Summary of Ξ_c^+ selection requirements in the HLT2 trigger. The second row summarises the shared set of selections both utilised by Ξ_c^+ and Λ_c^+ turbo lines. Variables definitions are available in Appendix A.

Using the available information from generated simulation samples discussed in

Section 3.2.2, particularly Table 3.1, the HLT2 selection line is estimated to retain around 3 % of signal (of which pass the generator cuts).

4.2 Offline combiner cuts

The LHCb DaVinci application produces the n-tuples to be used in the analysis. The application is configured such that Ξ_c^{**+} candidates are built using two stages of two-body particle combiners, firstly combining the Ξ_c^+ candidate with a bachelor π^- track to form a $\Xi_c(2645)^0$ candidate, followed by the combination of this $\Xi_c(2645)^0$ candidate with a bachelor π^+ track. Without appropriate selection requirements, the number of candidates produced by these routines (the candidate multiplicity) are highly numerous due to the large number of pions saved in each event and the large number of possible combinations.

To reduce candidate multiplicity (or combinatorial backgrounds) and therefore the n-tuple output size, cuts on kinematic and particle identification (PID) variables are used in order to remove candidates outside the range of interest with respect to our studies. They are summarised in this section for clarity.

4.2.1 Bachelor pion selection

As discussed, in order to reduce candidate multiplicity it is required that all bachelor pion tracks, which are used to reconstruct the $\Xi_c(2645)^0$ and Ξ_c^{**+} candidates, satisfy the following selection requirements

$$\begin{aligned} \chi^2_{IP,BPV} &< 9, \\ \text{ProbNN(ghost)} &< 0.4, \\ \text{ProbNN}(\pi^\pm) &> 0.1, \\ p_\text{T} &> 200\,\text{MeV}, \end{aligned}$$

where ProbNN refers to the output of the standard particle identification MVA algorithm used in LHCb for a respective hypothesis i.e. a pion, or ghost (a reconstructed artefact that does not correspond to an actual particle that traversed the detector) [29], p_T is the transverse momentum variable, and $\chi^2_{IP,BPV}$ is as defined in Section 4.1.

These requirements ensure that a high-p_T track, which is both unlikely to be a ghost track, likely to be a pion, and consistent with emerging from close to the primary vertex, is used to reconstruct the $\Xi_c(2645)^0$ and Ξ_c^{**+} candidates.

4.2.2 Selection of $\Xi_c(2645)^0 \to \Xi_c^+\pi^-$ candidates

Since the studies target primarily resonant decays we cut on the quantity

$$m[\Xi_c^+\pi^-] - m[\Xi_c^+] - m[\pi^+] < 150\,\text{MeV}, \tag{4.1}$$

to select a broad area around the $\Xi_c(2645)^0$ resonance, a decision based upon study of this variable in a subset of the 2016 data. This cut significantly reduces the candidate multiplicity and therefore the dataset size when running over the entire 2016–2018 data, but may limit the usefulness of the resulting n-tuples regarding any studies into the non-resonant region. There is an additional requirement that $\chi^2_{vertex} < 15$ on the $\Xi_c(2645)^0$ candidate, where χ^2_{vertex} is a measure of the vertex fit quality. This requirement ensures that tracks composing the $\Xi_c(2645)^0$ candidate are consistent with originating from a common vertex.

4.2.3 Selection of $\Xi_c^{**+} \to \Xi_c(2645)^0\pi^+$ candidates

All Ξ_c^{**+} candidates are required to satisfy the following conditions,

$$\begin{gathered}\text{Valid Best Primary Vertex (BPV)},\\ m[\Xi_c^+\pi^-\pi^+] < 3300\,\text{MeV},\end{gathered}$$

where the candidate is checked to see if a valid Primary Vertex (PV) was assigned, and the candidate is required to have a mass under 3300 MeV, in order to avoid unblinding the higher mass region under study by another LHCb analysis.

4.3 Stage I: $\Xi_c^+ \to pK^-\pi^+$ selection

4.3.1 Preselection

The offline selection begins with a rectangular cut study on $\Xi_c^+ \to pK^-\pi^+$ candidates and their daughter tracks, where selection requirements were determined using a smaller subset of the 2016–2018 data and the *sPlot* [47] technique, by way of a fit to the $m(pK^-\pi^+)$ variable. The resulting sPlots, which were created for each individual variable, were used to determine the ideal thresholds for signal efficient cuts. Then once again used to estimate expected background and signal efficiencies, considering one variable and corresponding selection requirement at a time. The primary purpose of these rectangular cut requirements (listed in Table 4.2) is to reduce the size of datasets produced, by focussing on any background candidates that can be trivially removed.

A Ξ_c^+ mass histogram, shown in Figure 4.1 demonstrates how the finalised rectangular cut requirements perform on the small subsample of $\Xi_c^+ \to pK^-\pi^+$ baryons (specifically, 2016 *MagUp*). Considering only candidates within the mass region $[2432.5, 2502.5]$ MeV (approximately 51% of candidates in this subsample), fitting each of the histograms separately using a double Gaussian and first-order polynomial model, we find that the signal and background rejection of all of the combined preselection requirements are 27.9% and 85.0% respectively. This is consistent with the reported values in Table 4.2, calculated via sWeights. This mass window is further tightened later on in the selection, after sWeights and fully trained MVAs are imposed using a larger data sample.

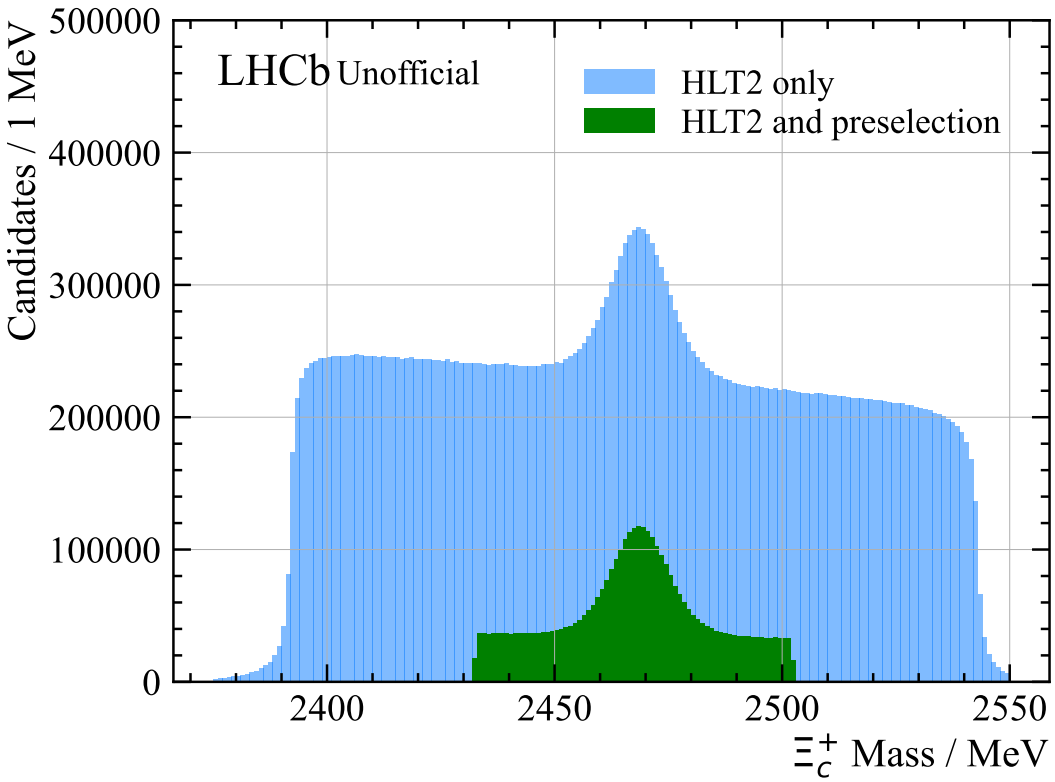

Figure 4.1 The $m[pK^-\pi^+]$ (Ξ_c^+ mass) histogram for the 2016 *MagUp* data sample before (HLT2 only) and after a mass window requirement and rectangular cuts are applied (HLT2 and preselection). By quick visual inspection, signal purity is vastly improved with a simple mass window cut and the determined preselection cuts.

Particle	Requirement	rej%	Sig. rej%	Bkg rej%
Ξ_c^+	$0 < \chi^2_{IP} < 20$	45.17 %	6.63 %	49.67 %
Ξ_c^+	$0 < \chi^2_{FD} < 2000$	27.24 %	8.22 %	29.46 %
Ξ_c^+	DIRA > 0.99995	0.16 %	0.05 %	0.18 %
Ξ_c^+	$0 < \chi^2_{ENDVERTEX}/N_{DOF} < 10$	0.00 %	0.00 %	0.00 %
K^- from Ξ_c^+	ProbNNk > 0.4	17.76 %	3.06 %	19.48 %
K^- from Ξ_c^+	ProbNNghost < 0.35	11.14 %	3.37 %	12.04 %
K^- from Ξ_c^+	TRACK_CloneDist < 0	0.67 %	0.07 %	0.74 %
π^+ from Ξ_c^+	ProbNNpi > 0.4	14.19 %	2.16 %	15.60 %
π^+ from Ξ_c^+	ProbNNghost < 0.35	11.06 %	2.97 %	12.00 %
π^+ from Ξ_c^+	isMuon == false	7.57 %	3.84 %	8.01 %
π^+ from Ξ_c^+	TRACK_CloneDist < 0	1.19 %	0.13 %	1.31 %
p from Ξ_c^+	ProbNNp > 0.4	32.37 %	2.43 %	35.87 %
p from Ξ_c^+	ProbNNghost < 0.35	19.91 %	3.67 %	21.81 %
p from Ξ_c^+	TRACK_CloneDist < 0	1.36 %	0.04 %	1.52 %
p from Ξ_c^+	Total momentum $p > 10$ GeV	0.00 %	0.00 %	0.00 %
–	**Altogether**	78.99 %	27.04 %	85.06 %

Table 4.2 Rectangular cut requirements on Ξ_c^+ candidates within the mass region $2432.5 < m(pK^-\pi^+) < 2502.5\,\text{MeV}$ and their percentage rejection on signal and background candidates when applying each cut on its own. Rejection fractions are determined by using the sum of signal and background sWeights to estimate the number of respective candidates before and after applying a cut. It is assumed that all candidates outside this mass region are combinatorial backgrounds. Variable definitions are available in Appendix A.

In addition to the above method, we may use *sPlot* [47] on the result of the fit on the 'HLT2 only' sample to obtain sWeights. These weights may be used to produce proxy signal and background distributions for variables (as long as they are not significantly correlated with the mass), and estimates of the number of signal and background events given some requirement, by summation of the relevant sWeights i.e. $N_S = \sum w_S$, $N_B = \sum w_B$ before and after a cut. These are used to obtain the % rejection columns in Table 4.2.

A 27.4% rejection of signal is additionally observed, largely caused by tightened ProbNN and χ^2_{IP} requirements that were introduced to reduce production size. However, from a retrospective point of view the storage constraints as perceived were stricter than they actually were, and nearly a third of Ξ_c^+ candidates were discarded as a result. Future analyses adopting a similar approach should ideally target the highest possible signal efficiency in the preselection approach.

4.3.2 Data-driven neural network (MVA) selection

Following the application of these preselection requirements and obtaining n-tuples for the full Run 2 data sample, the Ξ_c^+ selection is further optimised by a MVA, i.e. training a machine learning (ML) algorithm. A fully data-driven approach using the *sPlot* technique to obtain proxy signal and background distributions was preferred, as it does not rely on experiment Monte Carlo (MC) simulation where disagreements with data are common and expected. A representative sub-sample of all available data was used in the study, where the sampling fraction (without replacement) also known as a *prescale* was chosen as 5%. This sub-sample of Ξ_c^+ candidates was divided into two folds by `eventNumber`, with the fold number assigned using the formula

$$\text{fold} = \text{eventNumber} \;\; mod \;\; K, \tag{4.2}$$

where $K = 2$, the number of folds.

For each fold obtained by this procedure, an unbinned maximum-likelihood fit to the Ξ_c^+ mass variable is performed using the RooFit [62] package. The signal peak is modelled by a double Gaussian, which is the sum of two Gaussian Probability Distribution Functions (PDFs) with a shared mean parameter (M), two width parameters σ_1, σ_2, and a fraction $G_{1,frac}$. The background is modelled by a Chebyshev polynomial with one free parameter p_1 (a straight line). The

results can be seen in Figure 4.2, parameters are additionally summarised in Table 4.4. The fit model as described is shown to describe the data adequately, based on the χ^2/N_{dof} values and pull distributions.

Using the total PDFs resulting from these fits, the *sPlot* technique is employed in order to produce a set of signal and background sWeights for each fold. The sample and sWeights were used to train and test two classifiers. The first classifier is trained with fold zero, tested and evaluated with fold one, and the second classifier trained with fold one, then tested and evaluated with fold zero. This is referred to as k-folding (where $k = 2$), and is used to reduce biases in the MVA-based selection, such that a classifier is not used to select any candidates that were used to train it. It is also beneficial in the respect that events used in training are not required to be discarded. See Figure 4.3 for an illustration of the k-folding procedure.

The type of classifier used (library implementation, configuration, and classifier hyper-parameters) was decided by a process of experimentation and trial-and-error. A number of algorithms (Boosted Decision Tree, Neural Net, Decision Tree Forest) with implementations from a range of libraries (TMVA [63], scikit-learn [64], catboost [65], tensorflow [66] and keras [67]) were tested. The focus will remain only on the classifier with the highest performance, which was chosen for the final version of the selection - in this case, a neural network.

A feed-forward neural network was configured using the Keras component of the TensorFlow [66, 67] library, used for building, training, and evaluation of the neural network. The neural-network topology was chosen as $[n, 32, 16, 4, 1]$ by trial-and-error, where n is the number of input features (variables). A leaky rectified linear unit (LeakyReLU) activation is used for the hidden layers $[32, 16, 4]$, and a sigmoid activation function is used for the final output layer such that the classifier response is a value between 0 and 1. The Leaky ReLU is similar to a rectified linear unit (ReLU) activation, with a constant non-zero negative gradient in the negative x region,

$$y = \begin{cases} x & \text{for } x \geq 0 \\ -\alpha x & \text{for } x < 0 \end{cases} \tag{4.3}$$

where $\alpha = 0.05$ in the configured model for the purposes of this analysis. A rectified linear unit (ReLU) always has $\alpha = 0$; a leaky ReLU is considered to be more robust against the vanishing gradient problem due to the non-zero negative

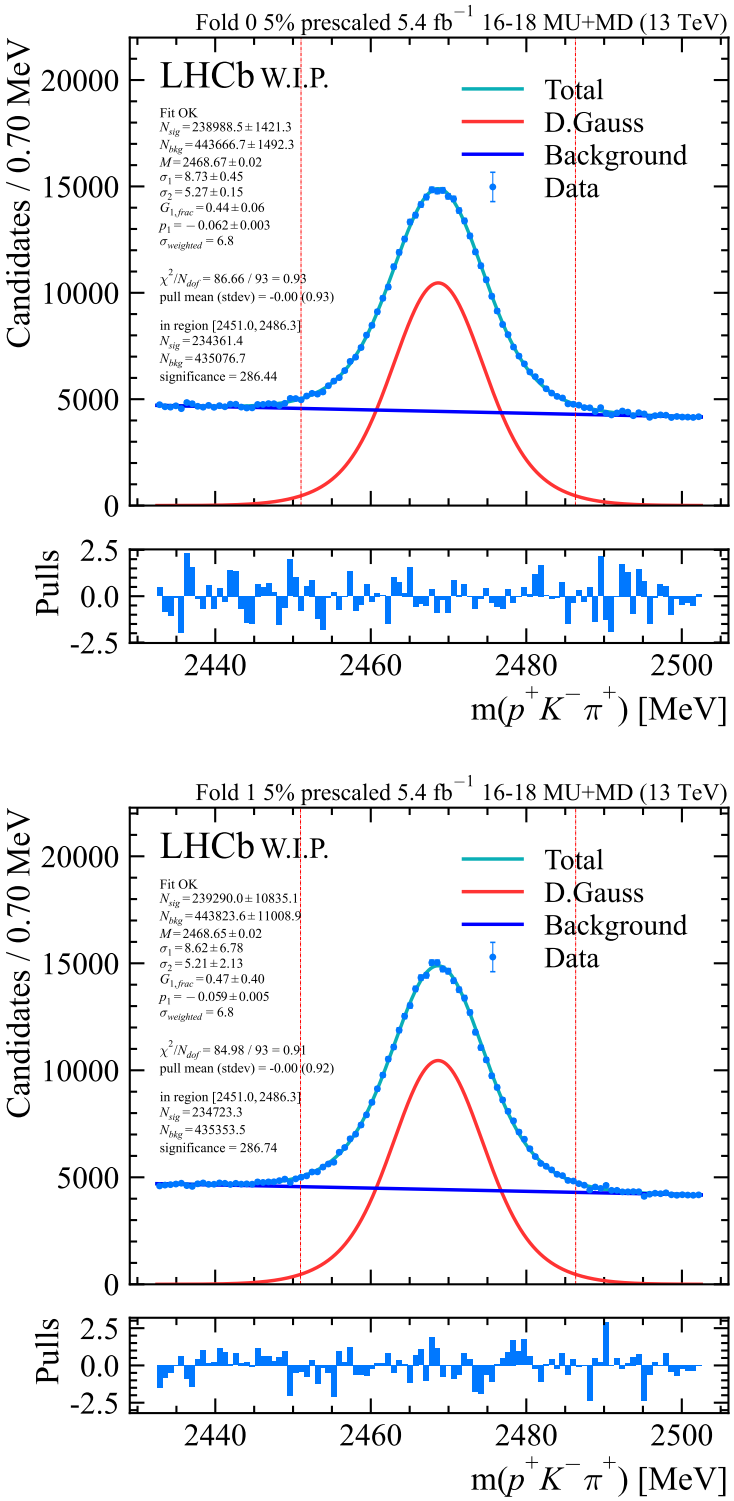

Figure 4.2 An *sFit* to the Ξ_c^+ invariant mass performed separately for each fold. A 5% prescaling is applied to the input samples used in these fits. At this stage only trigger and stage I rectangular cuts are applied.

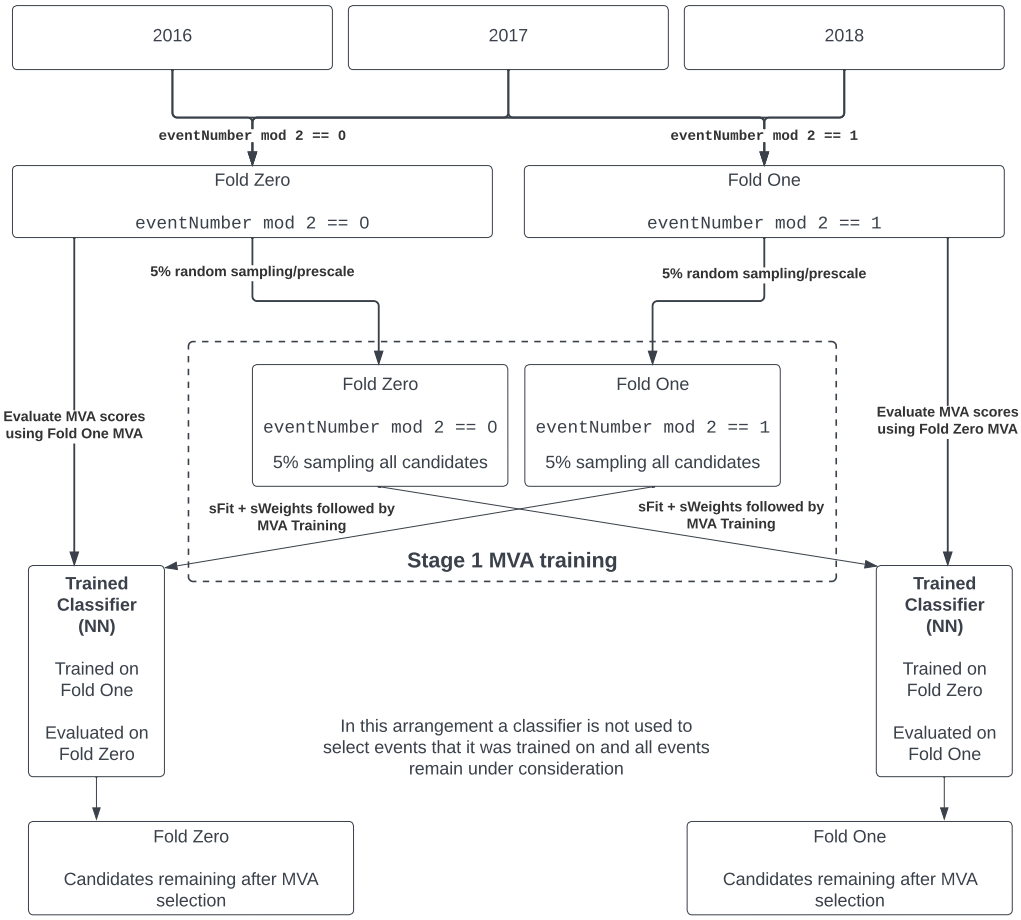

Figure 4.3 An illustration of how the data are k-folded, how the folds are used to train two separate classifiers, and determine MVA scores without discarding any part of the sample.

gradient (α) in the otherwise "dead" region $x < 0$ [68].

A loss function specifically designed for sWeighted samples is used to train this network, making use of the "Constrained MSE" loss function [69],

$$L_{mse} = \sum_i (y_i - w_i)^2 , \tag{4.4}$$

where y_i is the classifier prediction, and w_i is the signal sWeight for the corresponding candidate. Before entering the network, input variables are pre-processed using an `IronTransformer` from the `hep_ml` [70] library, which attempts to 'flatten' the input variables into uniform-like distributions with values in the $[0, 1]$ range. Pre-processing is a common step in Machine Learning, as it often improves classifier training performance by e.g. standardising, normalisation, or flattening the distributions of input features.

MVA Input Variable			
Ξ_c^+	p	K^-	π^+
$\chi^2_{endvertex}$	PIDK	PIDK	PIDK
χ^2_{IP}	PIDp	PIDp	PIDp
χ^2_{FD}	p_{T}	ProbNNk	ProbNNpi
FD	IP	MC15TuneV1_ProbNNk	MC15TuneV1_ProbNNpi
ETA	MC15TuneV1_ProbNNk	MC15TuneV1_ProbNNp	ProbNNghost
DIRA	MC15TuneV1_ProbNNp	ProbNNghost	
DOCAMIN	MC15TuneV1_ProbNNpi		
DOCA12	ProbNNghost		
DOCA13			
DOCA23			

Table 4.3 Ξ_c^+ MVA input variables. Definitions for variables are detailed in Appendix A.

The $n = 29$ input features (see Table 4.3) were chosen by a decision tree ensemble-based feature selection method, where features were ranked according to an impurity-based importance (see the `ExtraTreesClassifier` documentation of the `scikit-learn` [64] library). As a sanity-check, this was followed by a manual inspection of unfolded signal and background sPlots, and a sample of these are shown in Figure 4.4 and 4.5. The chosen variables are not seen to be highly correlated with the fit variable, as shown in Figure 4.6.

After training both classifiers, the output of the neural network is evaluated over test and train datasets and the signal predictions (or classifier responses) for both folds are binned into histograms and normalised for comparison. A plot exhibiting

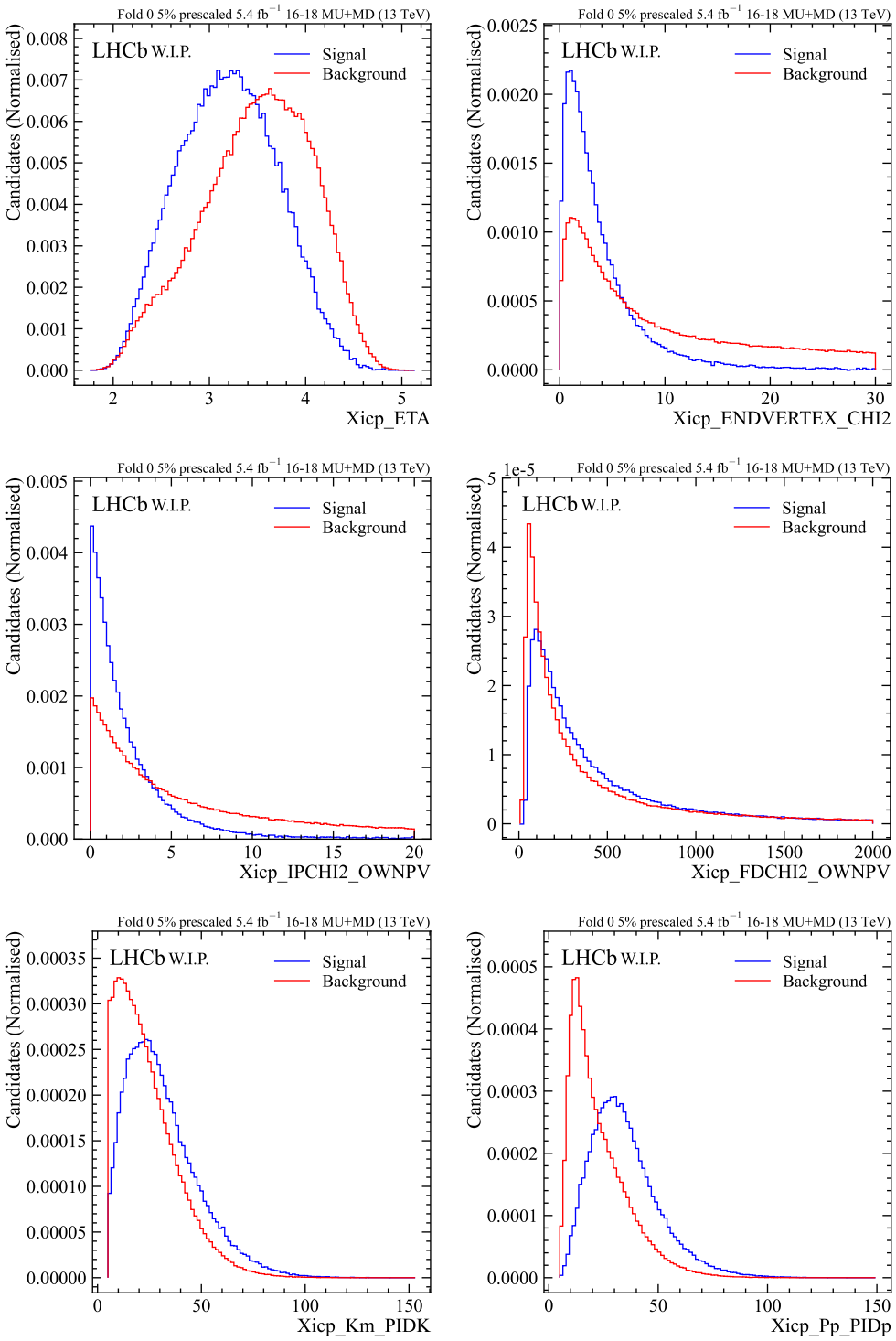

Figure 4.4 Ξ_c^+ variable sPlots (Part 1), taken from fold zero sample. Refer to Appendix A for variable definitions.

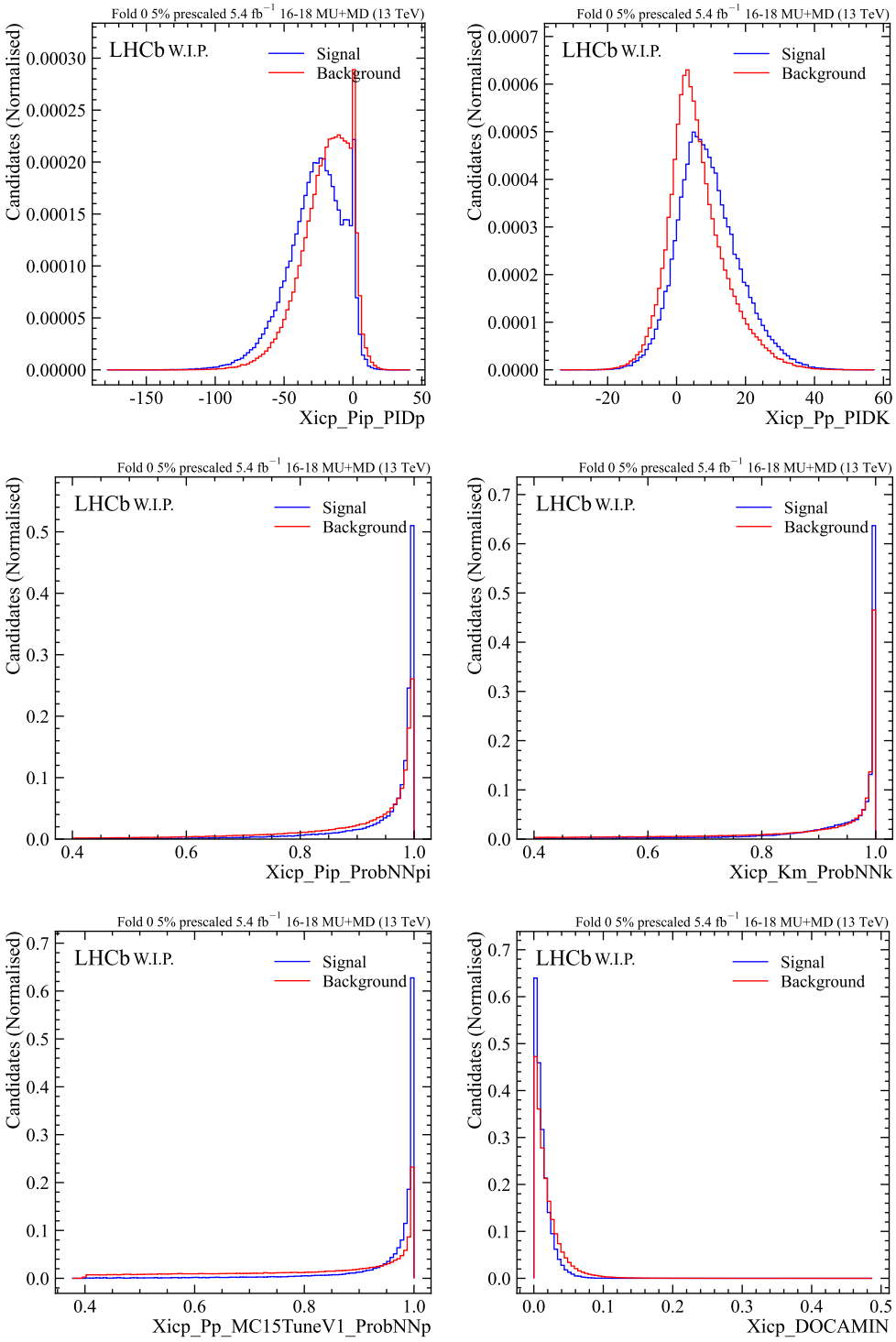

Figure 4.5 Ξ_c^+ variable sPlots (Part 2), taken from fold zero sample. Refer to Appendix A for variable definitions.

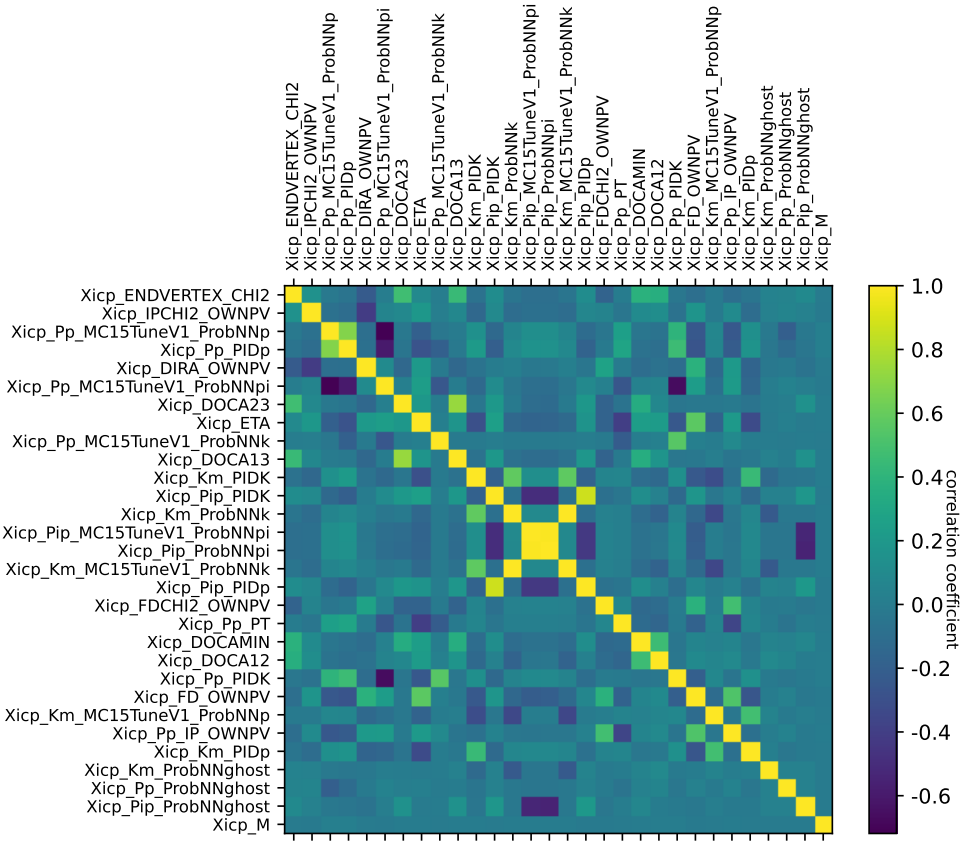

Figure 4.6 Correlation matrix of all MVA variables for the full (unfolded) training sample, and additionally the invariant mass $m(pK^-\pi^+)$, which is the fit variable. It should be noted that the correlation with the Ξ_c^+ invariant mass is very close to zero for all chosen variables. When using the *sPlot* technique it is important to confirm that variables are not significantly correlated with the relevant fit variable, if they are to be used in *sPlot* studies.

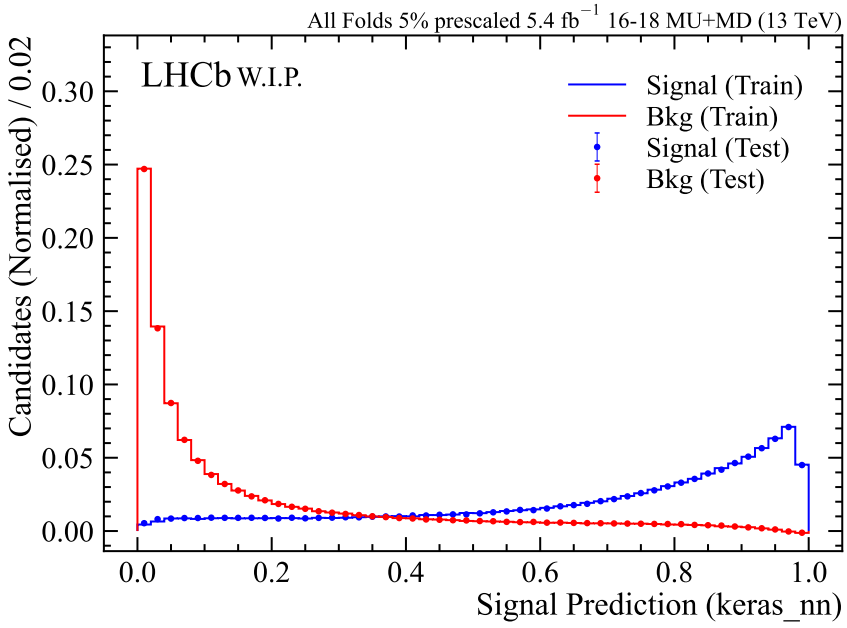

Figure 4.7 A normalised histogram of the neural network response (both folds are combined) for the training and test samples.

the MVA performance can be seen in Figure 4.7, showing that the training and testing samples are consistent for both signal and background proxy samples, and indicating that overtraining is under control.

To determine the optimal cut value, signal and background efficiencies ϵ_{sig}, ϵ_{bkg} are calculated with respect to classifier response trial cut $P(sig) > \delta$ values with δ in the range $[0, 1]$. The efficiencies are used to estimate the significance after applying the cut. An optimal cut on the classifier response was then obtained with respect to the significance metric

$$\text{significance}(\delta) = \frac{N_{sig}(\delta)}{\sqrt{N_{sig}(\delta) + N_{bkg}(\delta)}} = \frac{\epsilon_{sig}(\delta) n_{sig}}{\sqrt{\epsilon_{sig}(\delta) n_{sig} + \epsilon_{bkg}(\delta) n_{bkg}}} \tag{4.5}$$

where n_{sig}, n_{bkg} are the number of signal and background candidates before applying any MVA cut, obtained by integrating the respective PDFs within a region of interest $[2451.0, 2486.3]$ MeV, $N_{sig}(\delta)$, $N_{bkg}(\delta)$ are the remaining number of signal and background candidates after an MVA cut at some value δ, estimated using the estimated signal and background efficiencies $\epsilon_{sig}(\delta), \epsilon_{bkg}(\delta)$. The value of $\delta = 0.37$ is determined by the scan as shown in Figure 4.8, such that the MVA cut requirement $P(sig) > 0.37$ maximises the signal significance.

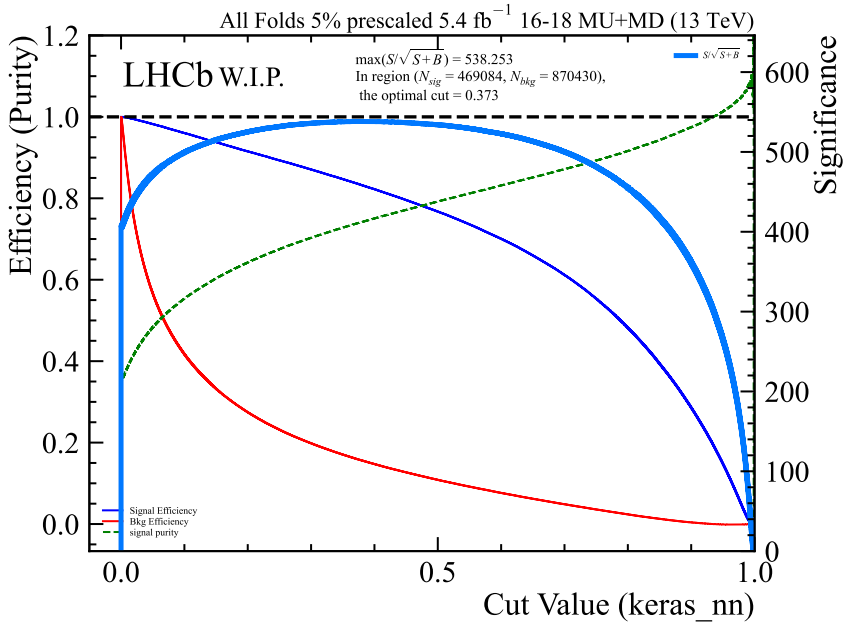

Figure 4.8 A scan of lower bound cut value (x-axis) vs. signal, background efficiencies and significance. The significance-optimised cut on the neural network response is $P(sig) > 0.37$, where $P(sig)$ is the response value of the neural network.

4.3.3 Mass window

Following the application of the $P(sig) > 0.37$ requirement, a mass window is additionally applied such that $2451.3 < m[\Xi_c^+] < 2486.0\,\text{MeV}$. The mass window is chosen such that at least 95% of the signal distribution (as studied in previous fit studies) is retained.

4.3.4 Summary

The Ξ_c^+ mass distribution after the stage I selection can be seen in Figure 4.9. The significance of genuine Ξ_c^+ candidates have considerably increased from 409 to 538 (calculated using the 5% pre-scaled sample) as a result of the classifier's high background rejection. A table summarising the relevant optimised parameters of the stage I selection can be referred to in Table 4.4.

To improve the performance of this selection method even further for any future analyses following a similar path, it may be prudent to relax some of the rectangular cuts that were applied at the start, to recover some of the signal

candidates rejected (in this case, $\approx$ 30%) and rely more on the multivariate discrimination power of the neural network.

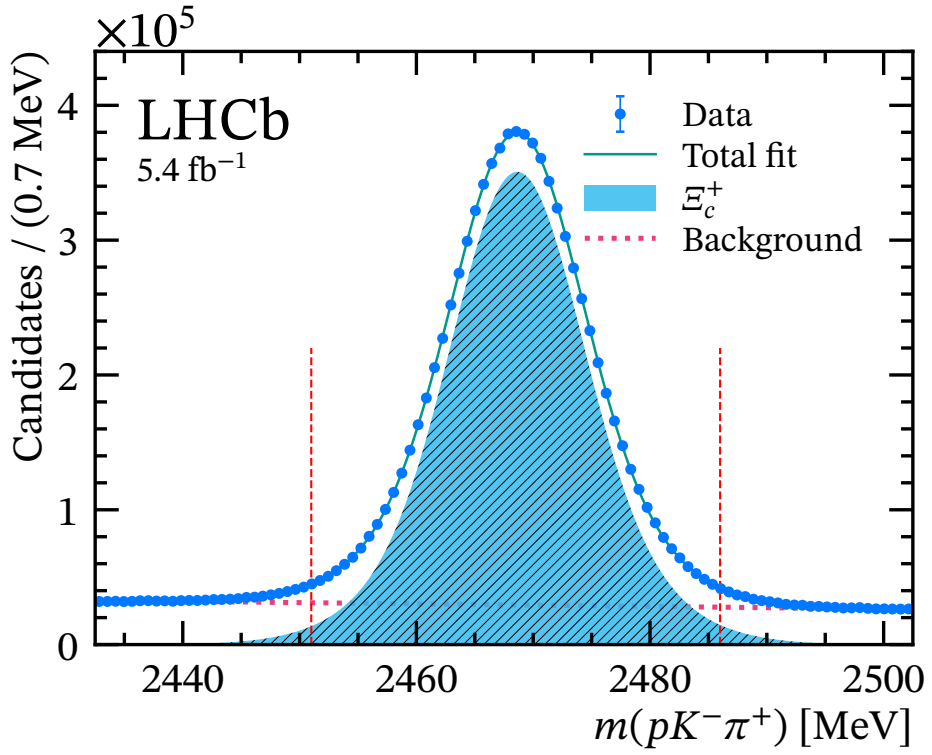

Figure 4.9 A plot of the $m(pK^-\pi^+)$ distribution (Ξ_c^+ mass) for all remaining Ξ_c^+ candidates, after applying stage I requirements.

4.4 Stage II: $\Xi_c(2645)^0 \to \Xi_c^+\pi^-$

As for Ξ_c^+, a dedicated data-driven MVA is configured and trained to improve the purity of the $\Xi_c(2645)^0$ candidate sample. Following a very similar methodology to stage I, the $\Xi_c(2645)^0$ candidates are split into two folds, and maximum-likelihood fits to

$$m_{corr}[\Xi_c(2645)^0] = m[\Xi_c^+\pi^-] - m(\Xi_c^+) + m_{PDG}(\Xi_c^+), \tag{4.6}$$

are obtained for fold zero and fold one assigned candidates which remain after all Ξ_c^+ selections, including the stage I (Ξ_c^+) multivariate classifier, have passed. This fit variable is chosen for the improved mass resolution, since the resolution effects from the Ξ_c^+ mass have been subtracted, and the $\Xi_c(2645)^0$ peak has a

Selection step	Relevant Parameters
Ξ_c^+ Fit – Fold Zero	$N_{sig} = 238988.5 \pm 1421.3$ $N_{bkg} = 443666.7 \pm 1492.3$ $M[\Xi_c^+] = 2468.7 \pm 0.02\,\text{MeV}$ $\sigma_1 = 8.73 \pm 0.45$ $\sigma_2 = 5.27 \pm 0.15$ $G_{1,frac} = 0.44 \pm 0.06$ $p_1 = -0.062 \pm 0.003$ $\chi^2/N_{dof} = 0.93$ pulls mean (stdev) = 0.00 (0.93)
Ξ_c^+ Fit – Fold One	$N_{sig} = 239290.0 \pm 10835.1$ $N_{bkg} = 443823.6 \pm 11008.9$ $M[\Xi_c^+] = 2468.65 \pm 0.02\,\text{MeV}$ $\sigma_1 = 8.62 \pm 6.78$ $\sigma_2 = 5.21 \pm 2.13$ $G_{1,frac} = 0.47 \pm 0.40$ $p_1 = -0.059 \pm 0.005$ $\chi^2/N_{dof} = 0.91$ pulls mean (stdev) = 0.00 (0.92)
MVA Significance Optimisation	**Mass Region of Interest** $[2451, 2486]\,\text{MeV}$ **Significance** $\max(N_{sig}/\sqrt{N_{sig} + N_{bkg}}) = 538.253$ **Optimal cut** MVA$P(sig) > 0.373$
Ξ_c^+ Fit (Fold Zero + One) – After MVA	$N_{sig} = 395654.8 \pm 1052.2$ $N_{bkg} = 145258.1 \pm 925.5$ Ξ_c^+ $M = 2468.62 \pm 0.01\,\text{MeV}$ $\sigma_1 = 8.80 \pm 0.23$ $\sigma_2 = 5.33 \pm 0.08$ $G_{1,frac} = 0.39 \pm 0.03$ $p_1 = -0.111 \pm 0.005$ $\chi^2/N_{dof} = 1.24$ pulls mean (stdev) = 0.00 (1.07)
Ξ_c^+ mass window requirement	$2451.3 < m[\Xi_c^+] < 2486.0\,\text{MeV}$

Table 4.4 Summary of parameters determined during development of the stage I selection.

higher purity concentrated within a smaller mass range.

4.4.1 Fit model and result

The fit result for both folds can be seen in Figure 4.10, along with the corresponding distribution for the wrong-sign $\Xi_c^+\pi^+$ sample. This wrong-sign distribution was plotted to check if it could be used to model the background shape, however some peaking backgrounds can be observed in both wrong-sign and signal channels. In the wrong-sign channel, Figure 4.10 (magenta), one peaking background is observed and expected to originate from $\Xi_c(2815)^+ \to (\Xi_c(2645)^0 \to \Xi_c^+\pi^-)\pi^+$ where the π^- is not reconstructed. This was confirmed by generating a simulation sample with the fast simulation software package RapidSim for this partially reconstructed decay and comparing the location and shape of the resulting distribution.

The signal component is modelled by the convolution of a Relativistic Breit–Wigner (RBW) with mass M and width Γ to model the resonance, and a double Gaussian to model the detector resolution. A RBW has the form

$$f_{RBW}(m; M, \Gamma) = \frac{q\,(q/q_0)^{2L}\,B'_L(q,q_0)^2}{(m^2 - M^2) + M^2\Gamma^2(m)}, \tag{4.7}$$

where the mass-dependent width is defined as

$$\Gamma(m; M, \Gamma) = \Gamma\frac{M}{m}\left(\frac{q}{q_0}\right)^{2L+1} B'_L(q,q_0)^2, \tag{4.8}$$

and

$$q = \frac{1}{2m}\sqrt{\left(m^2 - (m_1+m_2)^2\right)(m^2 - (m_1-m_2)^2)}, \tag{4.9}$$

additionally, for $L = 1$ (as expected for $\Xi_c(2645)^0 \to \Xi_c^+\pi^-$) the Blatt–Weisskopf barrier [71] factors are

$$B_1(q) = \sqrt{\frac{2z}{1+z}}, \quad B'_1(q,q_0)^2 = \sqrt{\frac{1+z_0}{1+z}}, \quad z = (\|q\|\,R)^2\,,\ z_0 = (\|q_0\|\,R)^2\,, \tag{4.10}$$

where L is the angular momentum change for the decaying resonance.

The Blatt–Weisskopf radius R [71] is fixed to $R = 3.1\,\text{GeV}^{-1}$, and the daughter particle masses $m_1 = m[\Xi_c(2645)^0]$ and $m_2 = m[\pi^+]$ are fixed to known measured values [1]. A range of choices for R are tested in Section 7.1.1 which did not find

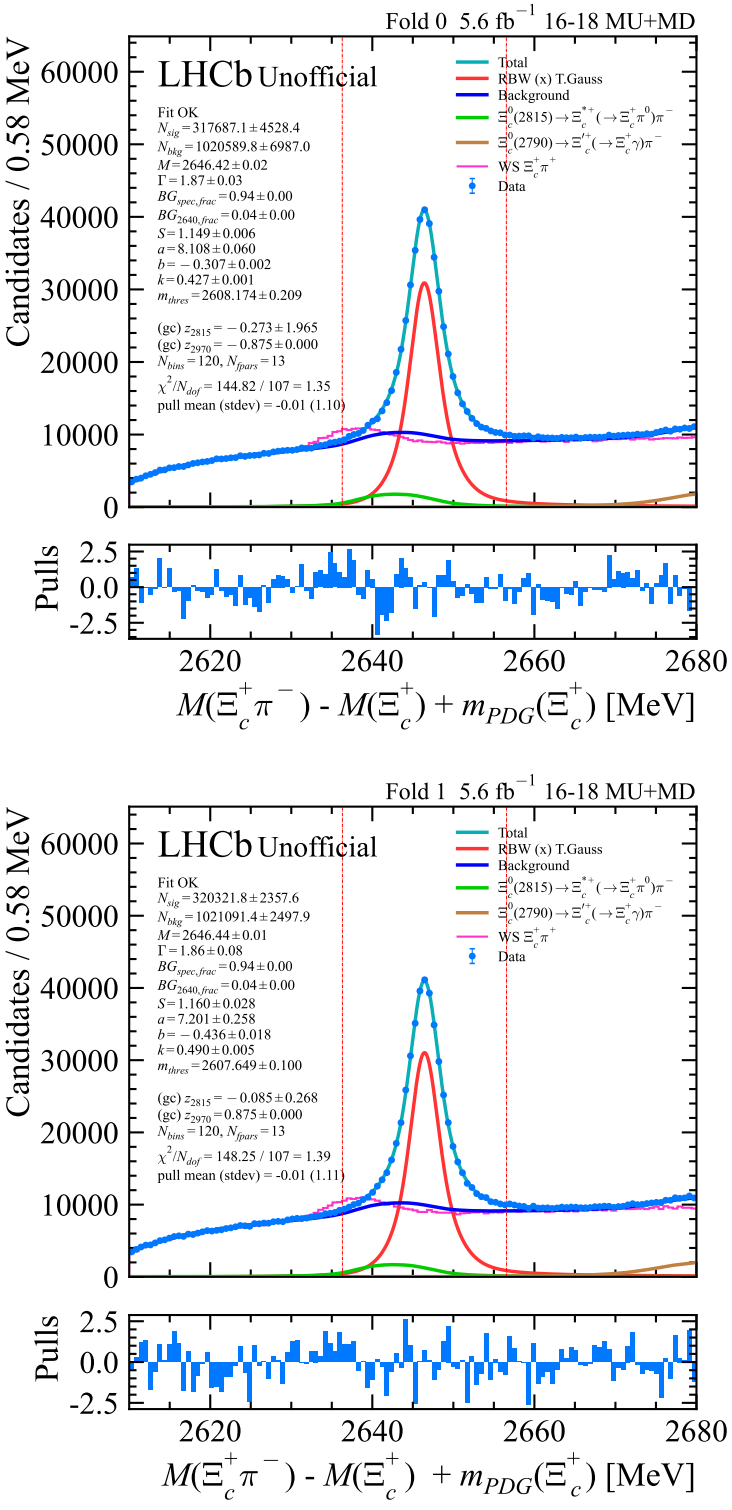

Figure 4.10 A fit to the reconstructed invariant mass $m(\Xi_c^+\pi^-)$ subtracted by the mass difference $m(\Xi_c^+)-m_{PDG}(\Xi_c^+)$ where m_{PDG} is the PDG measured value 2467.94 MeV. Fold 0 and 1 are shown at the top and bottom respectively. The wrong-sign (WS) overlay (magenta) is shown for comparison and is not part of the fit model. The $\chi^2/N_{dof}\approx 1$, and the pulls are centred on zero with unit standard deviation indicating a good fit. The fit model is described in the text.

significant systematic variations.

The double Gaussian parameters σ_1, σ_2, $G_{1,frac}$ were fixed to values obtained by simulation (see Section 5.2 and Figure 5.1), and the Gaussian widths were permitted to scale coherently by a free parameter S to allow for data-simulation disagreements.

The dominant background originates from random combinations of Ξ_c^+ candidates and pions π^- not originating from a single genuine singly-excited charmed baryon $\Xi_c(2645)^0$. The combinatorial background shape was chosen empirically and can be modelled by a PDF of the form

$$f(x; a, b, k, m_{thres}) = a(x - m_{thres})^k + b(x - m_{thres}), \tag{4.11}$$

with free parameters a, b, k, m_{thres}, and observable x. The fit model was not complete without considering additional background components:

- Close to the PDG reference value in the expected signal region, a partially reconstructed background caused by the decay $\Xi_c(2815)^0 \to \Xi_c(2645)^+(\to \Xi_c^+\pi^0)\,\pi^-$ where the π^0 is not reconstructed.
- Another partially-reconstructed background component from $\Xi_c(2790)^0 \to \Xi_c{}'^+\,(\to \Xi_c^+\gamma)\,\pi^-$ in the region close to 2680 MeV.

The partially-reconstructed decays are modelled by a histogram PDF where the histogram is created using 1 million decays generated by the RapidSim [72] simulation software, and the histogram PDF was allowed to shift according to the free parameter z, with a Gaussian constraint centred on zero with a width of 0.3 MeV for $\Xi_c(2815)^0$, and 0.5 MeV for $\Xi_c(2790)^0$ respectively. The Gaussian constraints were chosen as the PDG uncertainties on the $\Xi_c(2815)^0$ and $\Xi_c(2790)^0$ measured masses, which were also set as the generated masses in the RapidSim configuration.

4.4.2 Data-driven MVA selection

The parameters resulting from the fit are listed in Table 4.6, and the PDFs were used to extract sWeights. A gradient boosting classifier implemented in the `sklearn` library [64] (a `HistGradientBoostingClassifier`) was trained (with K-folding, as before) using the sWeighted signal and background distributions, and

Variable	Description
logXic0St_DTF_PV_XicpPDG_CHI2NDOF	$\ln \chi^2_{DTF}/N_{dof}$ of the DecayTreeFitter fit with PV and Ξ_c^+ mass constraint.
logXic0St_ENDVERTEX_CHI2NDOF	$\ln \chi^2_{endvertex}/N_{dof}$.
Xic0St_OWNPV_CHI2NDOF	χ^2_{IP}/N_{dof} of the $\Xi_c(2645)^0$ candidate.
Xic0St_P	Total momentum of the $\Xi_c(2645)^0$ candidate.
XicpIPCHI2_div_Xic0St_dau_IPCHI2SUM	$\chi^2_{IP}[\Xi_c^+]/\left(\chi^2_{IP}[\Xi_c^+] + \chi^2_{IP}[\pi^-]\right)$.
logXicp_ENDVERTEX_CHI2NDOF	$\ln \chi^2_{IP}/N_{dof}$ of the Ξ_c^+ candidate.
xicp_sig_proba	Stage I (Ξ_c^+) MVA classifier score.
1mSqrtPi1_ProbNNpi	$1 - \sqrt{\text{ProbNNpi}}$ of the bachelor pion.
1mSqrtPi1_ProbNNghost	$1 - \sqrt{\text{ProbNNghost}}$ of the bachelor pion.
logPi1_IPCHI2_OWNPV	$\ln \chi^2_{IP}/N_{dof}$ of the bachelor pion.
Pi1_P	Total momentum p of the bachelor pion.
Pi1_PT	Transverse momentum p_T of the bachelor pion.
Pi1_MC15TuneV1_ProbNNpi	PID variable: ProbNNpi for the bachelor pion.
Pi1_MC15TuneV1_ProbNNghost	Ghost probability of the pion track.

Table 4.5 $\Xi_c(2645)^0 \to \Xi_c^+\pi^-$ MVA input variables. Refer to Appendix A for variable definitions where unclear.

optimised by using the signal significance in the region encompassing a majority of the signal PDF area. The input variables are displayed in Table 4.5.

The current separation plot for fold 0 and 1 can be seen in Figure 4.11. The optimal cut is chosen as in stage I by using the signal significance in the mass region of interest, with the final cut requirement chosen as $P(sig) > 0.20$. This is a loose cut corresponding to the maximally obtainable significance. However, some background is still removed while avoiding a significant number of signal candidates being affected by the cut.

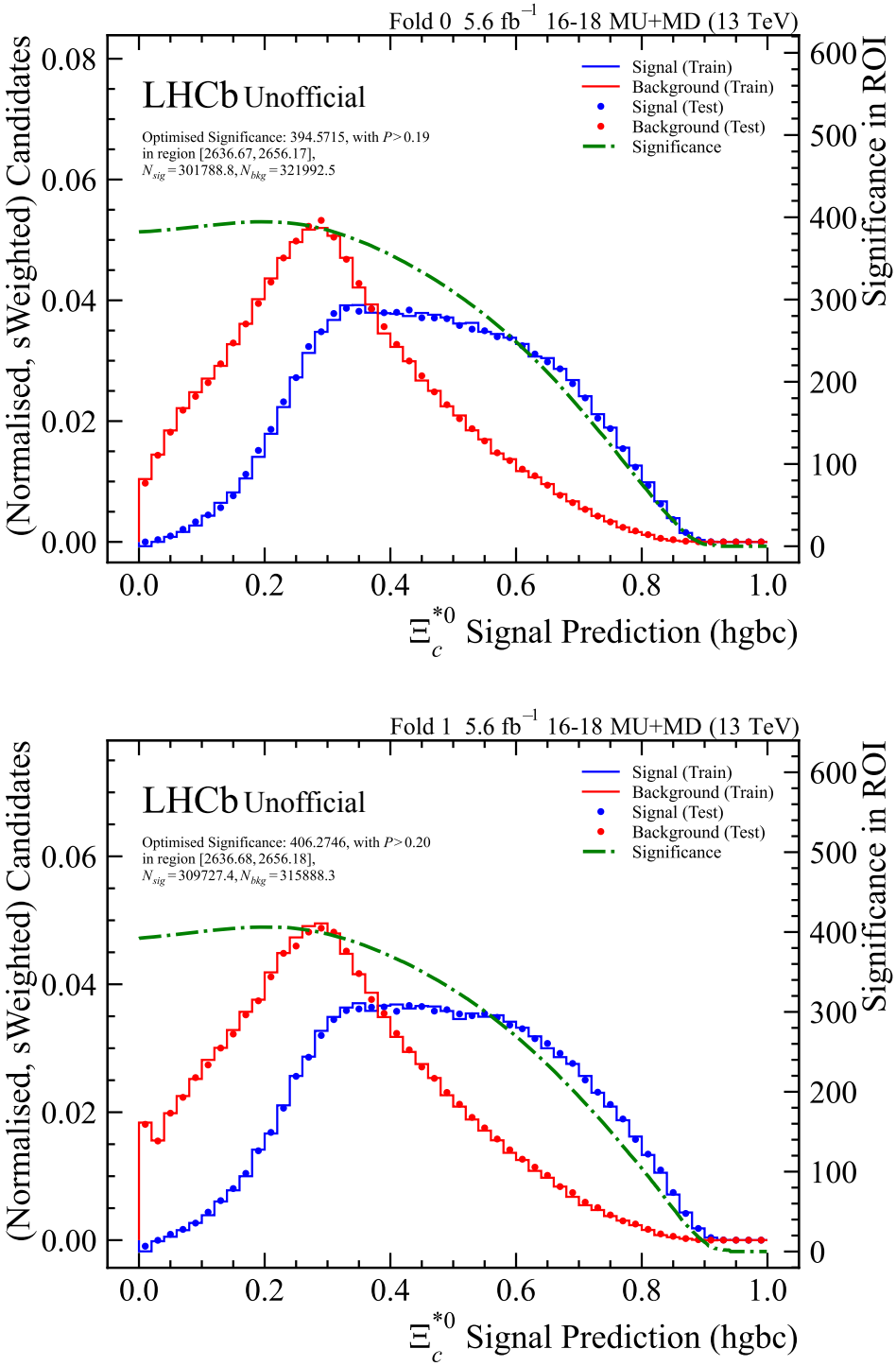

Figure 4.11 Stage II $\Xi_c(2645)^0$ Gradient BDT classifier response sPlots for Fold 0 and Fold 1, including an overlay of the significance corresponding to a classifier response cut. The N_{sig}, N_{bkg} labels on each figure represent the number of candidates before any classifier response cut with respect to the signal and background PDFs integrated in the specified mass region of interest i.e. $[2636, 2656]$ MeV.

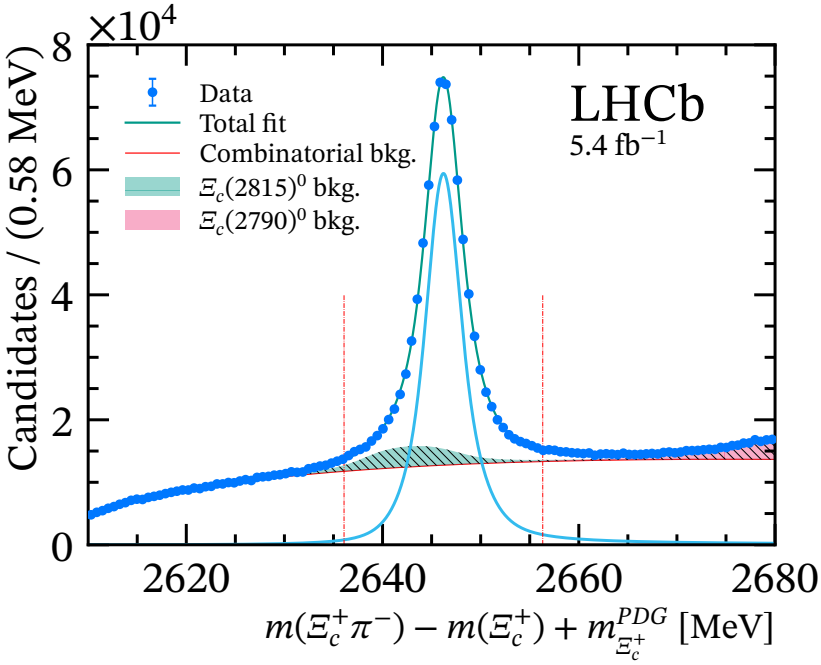

Figure 4.12 The $m_{corr}[\Xi_c(2645)^0]$ distribution for $\Xi_c(2645)^0$ candidates remaining after stage II requirements.

4.4.3 Mass window

The mass region of interest, and mass window requirement is chosen as

$$2635 < m_{corr}[\Xi_c(2645)^0] < 2656\,\text{MeV}. \tag{4.12}$$

As in the previous selection stage, this choice of mass window is taken as the range in which 95% of the signal PDF area is enclosed.

4.4.4 Summary

A summary of all relevant optimised parameters and cuts is shown in Table 4.6. A plot of the $m_{corr}[\Xi_c(2645)^0]$ distribution for remaining $\Xi_c(2645)^0$ candidates after stage II can be referred to in Figure 4.12. For an overview of the signal efficiency and background rejection of the requirements described in this chapter, please refer forward to Table 4.10.

Selection step	Relevant Parameters
$\Xi_c(2645)^0$ Fit – Fold Zero	$N_{sig} = 317687 \pm 4528$ $N_{bkg} = 1020589 \pm 6987$ $M = 2646.42 \pm 0.02\,\text{MeV}$ $\Gamma = 1.87 \pm 0.03\,\text{MeV}$ $S = 1.15 \pm 0.01$ $a = 8.11 \pm 0.06$ $b = -0.307 \pm 0.002$ $k = 0.427 \pm 0.001$ $m_{thres} = 2608.174 \pm 0.21\,\text{MeV}$ $\chi^2/N_{dof} = 1.35$ pulls mean (stdev) $= -0.01$ (1.10)
$\Xi_c(2645)^0$ Fit – Fold One	$N_{sig} = 320322 \pm 2358$ $N_{bkg} = 1021091 \pm 2498$ $M = 2646.44 \pm 0.01\,\text{MeV}$ $\Gamma = 1.86 \pm 0.08$ MeV $S = 1.16 \pm 0.03$ $a = 7.20 \pm 0.26$ $b = -0.436 \pm 0.018$ $k = 0.490 \pm 0.005$ $m_{thres} = 2607.6 \pm 0.1\,\text{MeV}$ $\chi^2/N_{dof} = 1.39$ pulls mean (stdev) $= -0.01$ (1.11)
MVA Significance Optimisation	**Mass Region of Interest** $[2635, 2656]\,\text{MeV}$ **Significance** $\max(N_{sig}/\sqrt{N_{sig} + N_{bkg}}) = 406$ **Optimal cut** MVA > 0.2

Table 4.6 Summary of parameters determined during development of the stage II selection.

4.5 Stage III: $\Xi_c^{**+} \to \Xi_c(2645)^0\pi^+$

Following the application of stage I and stage II selections, the final state Ξ_c^{**+} candidates are selected using a partially data-driven approach. An MVA is trained using the truth-matched Monte Carlo samples outlined in Section 3.2.2 as signal, and wrong-sign $\Xi_c^+\pi^-\pi^-$ (*WSPMM*), $\Xi_c^+\pi^+\pi^+$ (*WSPPP*) samples as proxy for background. Wrong-sign samples were chosen as proxy for background, since combinatorial backgrounds due to incorrect combinations of pions with a Ξ_c^+ candidate are considered the dominant contribution. There are two wrong-sign channels of relevance:

1. $(\Xi_c(2645)^0 \to \Xi_c^+\pi^-)\pi^-$ referred to as $\Xi_c^+\pi^-\pi^-$ or *WSPMM* — not possible to reconstruct Ξ_c^{**+}, but genuine $\Xi_c(2645)^0$ decays included.

2. $\Xi_c^+\pi^+\pi^+$ referred to as $\Xi_c^+\pi^+\pi^+$ or *WSPPP* — not possible to reconstruct Ξ_c^{**+} nor $\Xi_c(2645)^0$.

Since this is a blind analysis searching for new Ξ_c^{**+} states, the *sPlot* and significance-based optimisation approach cannot be used as the right-sign distribution is blinded until a full selection has been decided and fixed, and right-sign data cannot be directly used. Instead, the PFoM [73] is used with wrong-sign data (background proxy) and simulation samples (signal proxy) as described later in this section to choose an appropriate cut on the MVA score.

4.5.1 Preselection and multivariate analysis

Some basic preselection requirements are applied to the sample listed in Table 4.7, which includes the stage II mass window. These cuts are highly efficient for signal, as shown in Table 4.10 (99.6% when applied to the $\Xi_c(2815)^+$ simulation sample). The mass window is not applied until this stage, mainly for technical reasons and flexibility, in order not to remove all non-resonant decays from the data sample files until necessary. As in previous sections, such requirements are chosen to target high background rejection with close to 100% signal efficiency resulting from any given cut. The χ^2_{DTF}/N_{dof} is particularly effective discriminator as it represents the quality of a re-fit of the decay tree having constrained the decay to originate from the PV and fixed the Ξ_c^+ mass to the known PDG measured value [1].

Ξ_c^{**+} **selection requirements before MVA**
Ξ_c^{**+} $\chi^2_{endvertex} > 0$
Ξ_c^{**+} $\chi^2_{IP} > 0$
Ξ_c^{**+} $\chi^2_{FD} > 0$
Ξ_c^{**+} $0 < \chi^2_{DTF}/N_{dof} < 2.0$
Ξ_c^{+} $\chi^2_{endvertex} > 0$
Ξ_c^{+} $\chi^2_{IP} > 0$
$\pi^{\pm}$ $\chi^2_{IP} > 0$
$2635 < m_{corr}[\Xi_c(2645)^0] < 2656\,\text{MeV}$

Table 4.7 Ξ_c^{**+} preselection requirements. The last requirement is particularly important as it imposes a mass window on the $\Xi_c(2645)^0$ intermediate resonance, removing a significant fraction of background, and nonresonant signal candidates. The m_{corr} variable is defined in Equation (4.6). See Appendix A for other variable definitions where not clear.

A gradient boosted decision tree implementation from the `catboost` library [65] is trained with 15 input features/variables, summarised in Table 4.8, and Figures 4.13, 4.14. The variables are inspected further in Appendix B. The classifier performance is optimised using the PFoM [73], which allows one to choose an appropriate threshold corresponding to a specific region of interest and the desired target significance, using proxy samples for signal and background, thus no knowledge of the unblinded right-sign channel is required. This is calculated with respect to a specific MC sample as proxy for signal, such that

$$N_{gen} \cdot \text{PFoM} = \frac{N_{MC}}{(\sigma/2) + \sqrt{N_{bkg}}}, \tag{4.13}$$

where N_{gen} is the number of generated events in the MC sample, N_{MC} is the number of MC events within the mass region of interest (ROI) enclosing the simulated resonance of the MC sample, σ is the target significance (a value of $\sigma = 5$ is used), and N_{bkg} is the number of background events within the mass ROI. The N_{bkg} value is calculated by counting the number of candidates in the wrong-sign data $\Xi_c^+\pi^+\pi^+$ channel inside the mass ROI.

The variables chosen as listed in Table 4.8 comprise primarily: bachelor pion PID variables, flight distance and impact parameter χ^2, combinations of the transverse momentum, and the aforementioned χ^2_{DTF}/N_{dof}. Some variables have been log-transformed to improve the classifier performance, as is standard practice.

Approximately equal numbers of signal and background candidates were provided

to train the classifier, after applying the preselection cuts of Table 4.7. It is important to note that these requirements include the mass window cut on the $\Xi_c(2645)^0$ m_{corr} mass variable determined in stage II (Section 4.4.3).

The signal sample was created from a truth-matched sample of MC signal candidates merged across all generated states, where the truth-matching required a genuine Ξ_c^{**+}, Ξ_c^+, $\Xi_c(2645)^0$, and pair of bachelor pions. A background proxy sample was constructed by merging both $\Xi_c{}^+\pi^-\pi^-$ and $\Xi_c{}^+\pi^+\pi^+$ wrong-sign channel samples and randomly sampling from the merged sample without replacement, such that the number of entries in the background proxy sample was approximately equal to the signal proxy sample for the training process. This meant sampling a 1.8% fraction of the merged wrong-sign candidates. By balancing the number of entries in signal and background training samples, classifier performance metrics such as the AUC improved. The AUC is defined as the area under the ROC curve, which is obtained by plotting the signal efficiency vs. the background efficiency while varying the classifier response cut. While this signal proxy sample (combination of all simulation) is used in the training, only one sample corresponding to the $\Xi_c(2923)^+$ state is used later when calculating the PFoM (Equation (4.13)).

Classifier train-test separation histograms are displayed in Figure 4.15. It was additionally possible to rank the input variables (see Figure 4.17). Following the calculation of the PFoM as shown in Figure 4.16, the classifier response requirement was chosen as $P(sig) > 0.91$. This choice was motivated by finding the cut value corresponding to the maximum value of the metric as calculated in the mass region of the $\Xi_c(2923)^+$ simulation sample, defined as $[2901.74, 2944]$ MeV.

The performance of the full selection after applying all algorithms are evaluated in Section 4.6. In the following subsections 4.5.3 and 4.5.2, clone tracks and multiple candidates are introduced. Refer forward to the cut flow table in Table 4.10 for an overview of signal efficiency and background rejection for the requirements described in this chapter.

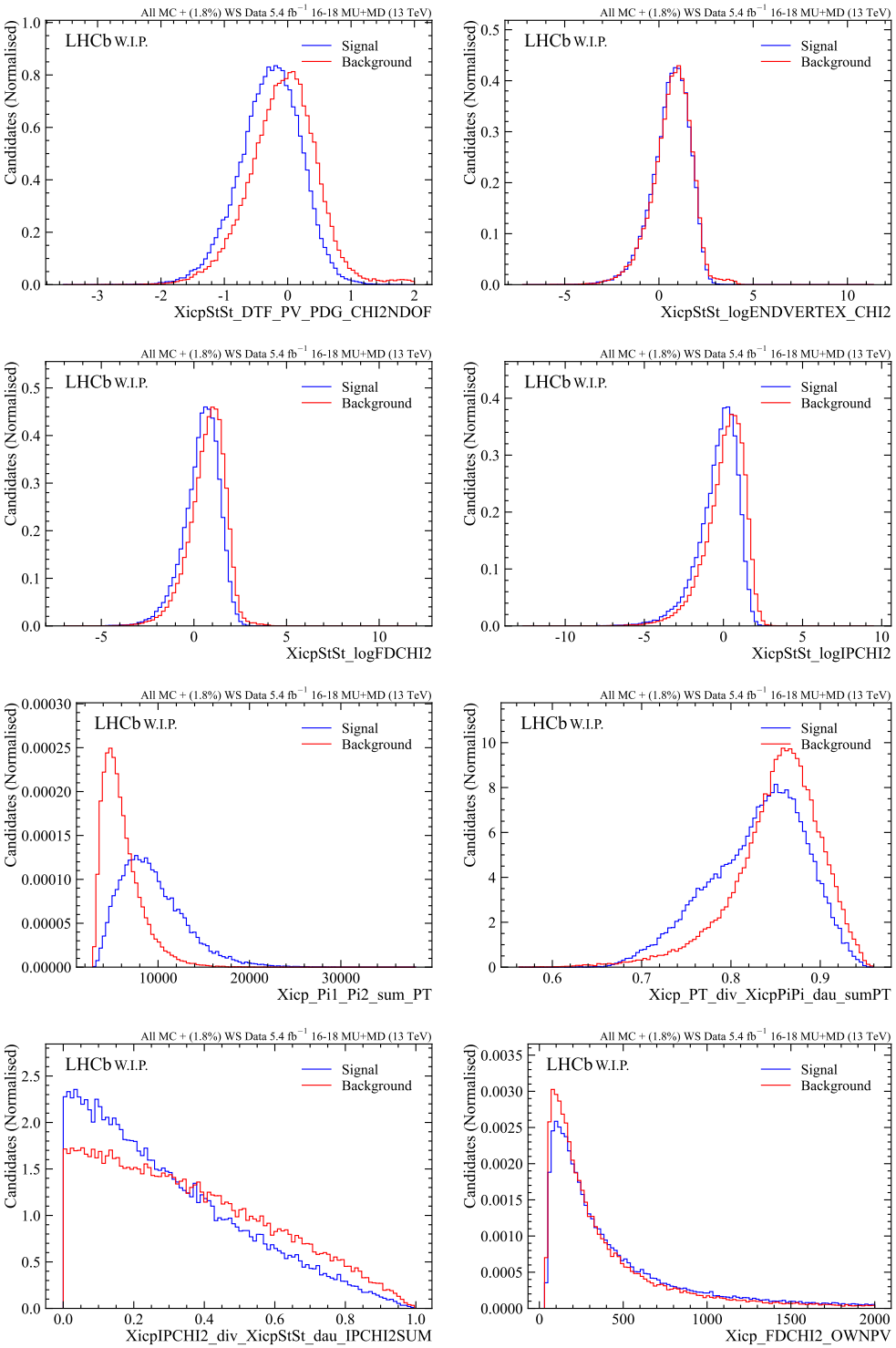

Figure 4.13 Ξ_c^{**+} MVA input variables (Part 1)

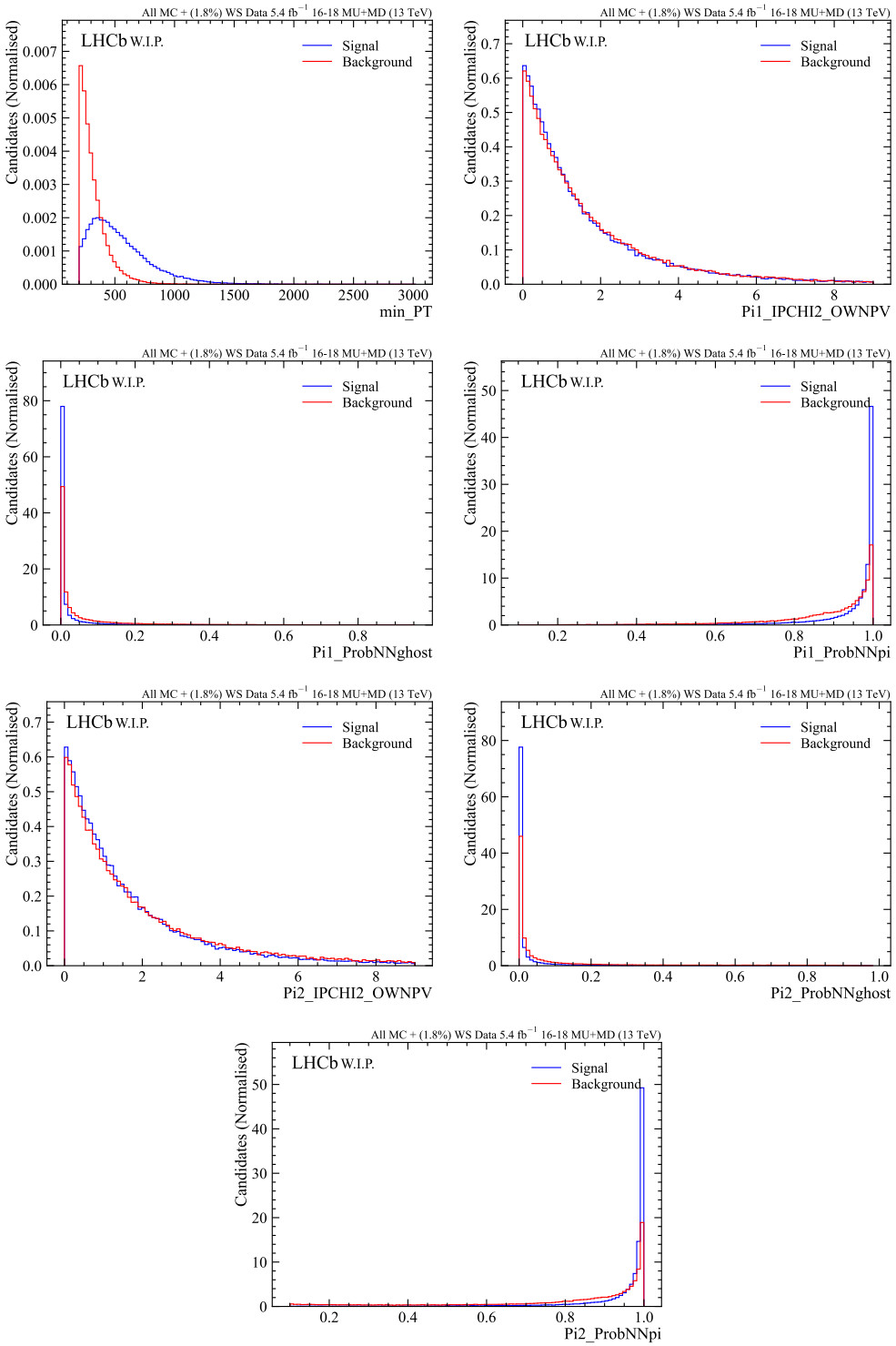

Figure 4.14 Ξ_c^{**+} MVA input variables (Part 2)

Ξ_c^{**+} MVA Input Variable
$\log(\chi^2_{endvertex}[\Xi_c^{**+}])$
$\log(\chi^2_{IP}[\Xi_c^{**+}])$
$\log(\chi^2_{FD}[\Xi_c^{**+}])$
χ^2_{DTF}/N_{dof}
$\chi^2_{FD}[\Xi_c^+]$
π^+, π^- `ProbNNpi`
π^+, π^- χ^2_{IP}
π^+, π^- `ProbNNghost`
$p_{\mathrm{T}}(\Xi_c^+) + p_{\mathrm{T}}(\pi^+) + p_{\mathrm{T}}(\pi^-)$
$\min(p_{\mathrm{T}}(\pi^+), p_{\mathrm{T}}(\pi^-))$
$p_{\mathrm{T}}(\Xi_c^+)/\left(p_{\mathrm{T}}(\Xi_c^+) + p_{\mathrm{T}}(\pi^+) + p^-(\pi^-)\right)$
$\chi^2_{IP}(\Xi_c^+)/\left(\chi^2_{IP}(\Xi_c^+) + \chi^2_{IP}(\pi^+) + \chi^2_{IP}(\pi^-)\right)$

Table 4.8 Ξ_c^{**+} MVA input variables calculated with respect to the Ξ_c^{**+}, Ξ_c^+, and bachelor pions.

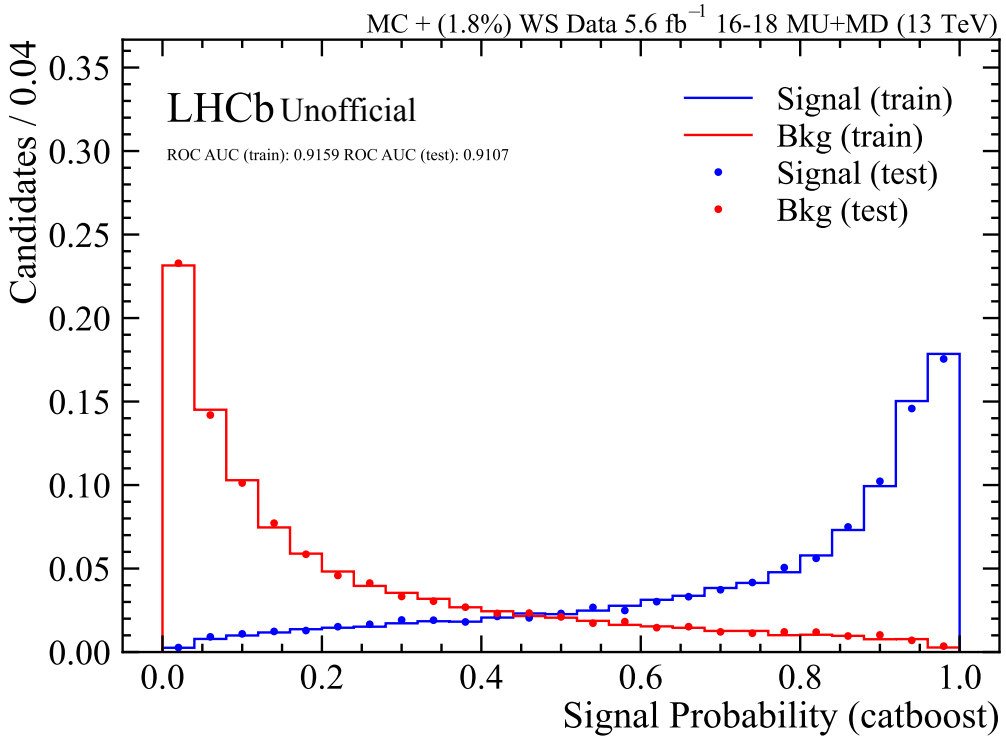

Figure 4.15 The Gradient BDT model response after training for both training and test samples, used as an overtraining check.

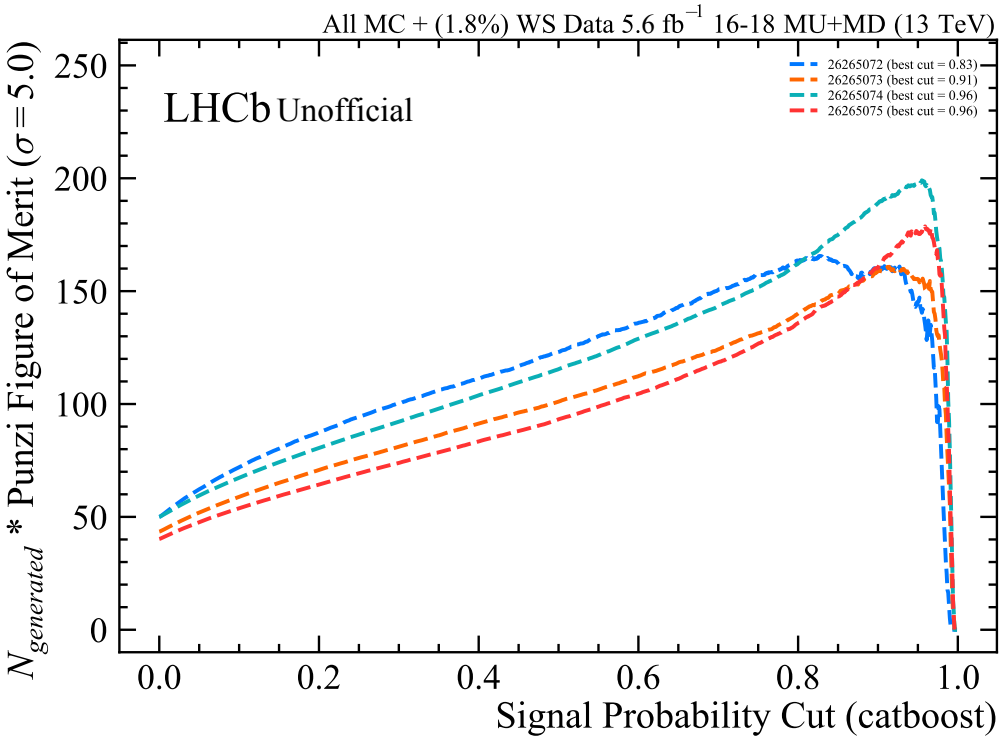

Figure 4.16 A plot of the PFoM for all available Monte Carlo simulation samples calculated within the corresponding signal mass regions. The optimal cut is taken to be the maximum of each distribution. From the curves shown on the plot, cut value from `26265073` (EventNumber of the MC sample) corresponding to $\Xi_c(2923)^+$ was selected to be used as the final MVA cut requirement.

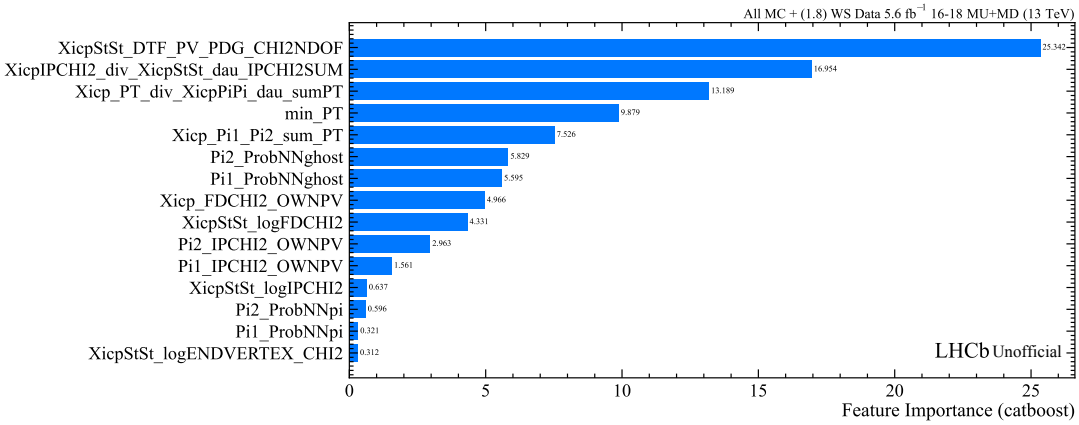

Figure 4.17 The feature or variable importances as determined by the `catboost` [65] library.

# candidates / event	# events	# candidates
1	34,708	34,708
2	5,540	11,080
3	778	2,334
4 or more	154	648
Total	41,180	48,770

Table 4.9 A table of the number of events and candidates with respect to the number of candidates per event, after all selections are applied.

4.5.2 Multiple candidates

An event has a candidate multiplicity greater than one if there are multiple ways to combine the particles (i.e. π^+, π^-, Ξ_c^+) in a given event to reconstruct the physics of interest, in particular pions, which can be particularly numerous even after kinematic preselection cuts. The candidate multiplicity is evaluated by grouping candidates by their event number and counting the size of each group. Before stage I requirements are fully applied, there as many as 10 candidates per event. After all selection stages, as can be seen in Figure 4.18 and Table 4.9, a majority (84%) of events yield only a single candidate. It is expected that each event will have at most one genuine Ξ_c^{**+}, except in rare circumstances. In reality, it is common that a single Ξ_c^+ candidate will be combined with multiple sets of bachelor pions which ultimately pass all requirements. At most, one will be signal with the remainder as background. Any procedure to pick one candidate in events with multiplicity greater than one will unnecessarily remove signal. As such, it has been decided to keep all candidates after the selection. There is no multiple candidate treatment and all candidates are kept. Cross-checks to understand the effect of multiple candidates on our measurements are discussed in Section 7.2.5.

4.5.3 Clone removal

During particle combination in HLT2 and DaVinci, it is possible for tracks that are similar, in that they share information (e.g. detector hits), to be utilised more than once per reconstructed Ξ_c^+ or Ξ_c^{**+} candidate, these are called 'clones'. A track can sometimes be described as a clone if it shares a majority of hits with another track, as one example. In this analysis, a candidate is defined as containing clone tracks by the degree of angular separation (between the total momentum vectors) between any two tracks used to reconstruct the decay

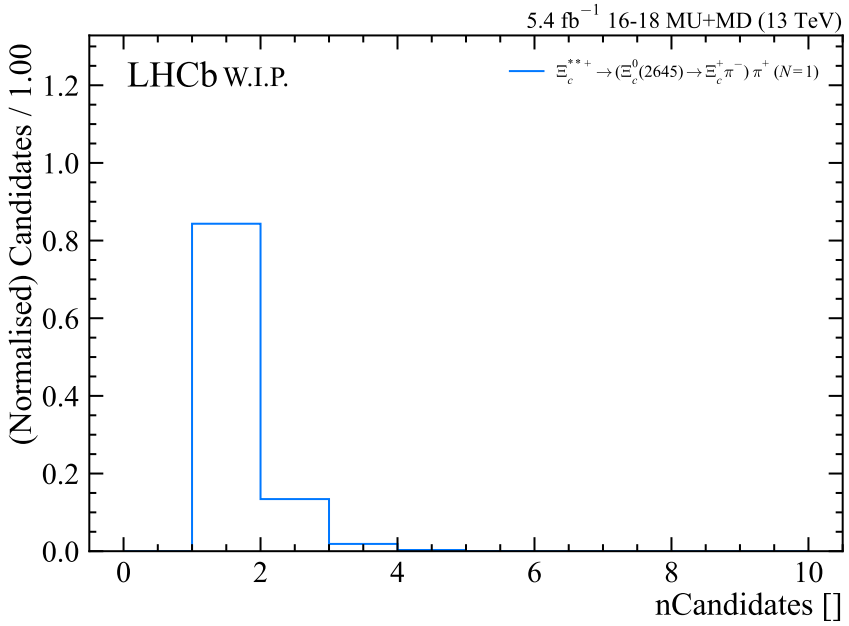

Figure 4.18 Number of Candidates per event remaining after all selections, as a fraction of the full sample. A majority (82%) of events contain one candidate only.

candidate.

Clones are removed by imposing a minimum threshold requirement on the angular separation between any two p, K^-, π^+ particle momentum vectors per candidate, such as the angular seperation variable shown in Figure 4.19. A very small or zero angular separation indicates that the tracks are clones and that the candidate should be removed. The threshold is determined by studying all possible angular separation plots between Ξ_c^+ daughter and the bachelor pion tracks, and the final selection requirement is set at $\Delta\theta > 0.08$ mrad.

The clone cuts remove a small fraction of candidates, see Table 4.10.

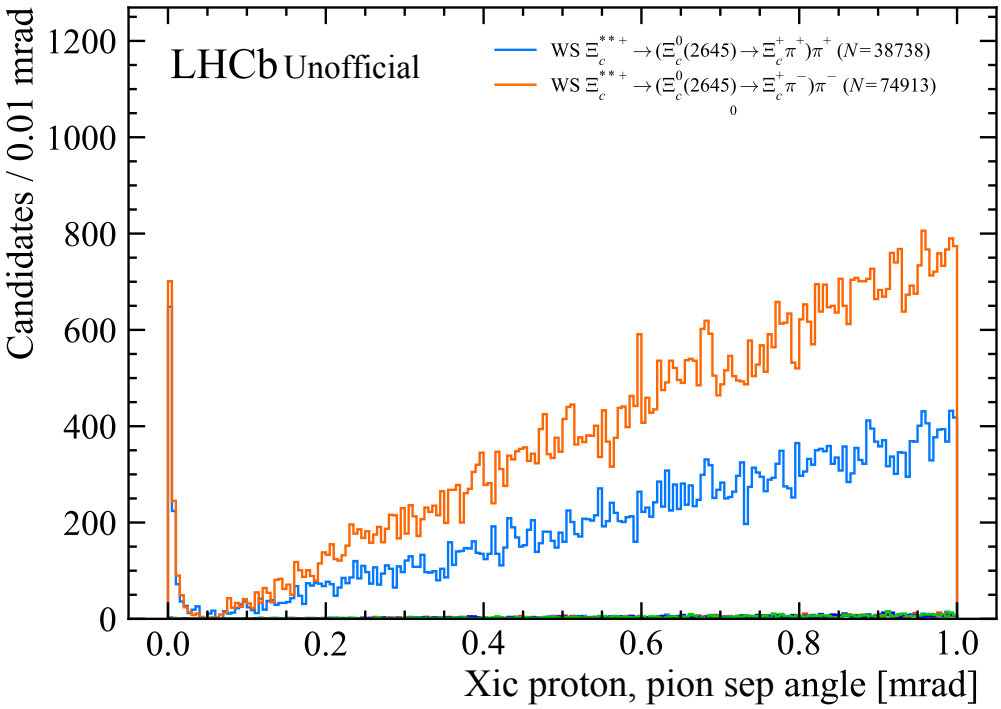

Figure 4.19 Angular separation between the Ξ_c^+ proton and pion daughter tracks. As the angular separation decreases, the number of candidates also falls except for a sharp increase closer to zero. This sharp increase is an indicator that the two tracks in question are clones.

4.6 Unblinded spectrum and discussion

The full selection stages I-III are applied to the unblinded right-sign $\Xi_c^+\pi^-\pi^+$ data channel. See Table 4.10 for full summary of cut efficiencies broken down by stage and sub-stages.

Stage	**Cut applied**	**Candidates remaining**	**retention%**	$\Xi_c(2815)^+$ **MC sig. efficiency%**
I	Tupling Preselection	183,011,955	–	–
	Rectangular Cuts	63,307,255	34.6%	85.7 %
	Ξ_c^+ mass window	39,232,340	62.0%	96.4 %
	Ξ_c^+ MVA	15,922,392	40.6%	44.2 %
II	Preselection	15,497,515	97.3%	100 %
	$\Xi_c(2645)^0$ MVA	10,835,487	69.9%	96.2 %
III	Presel. (inc. $\Xi_c(2645)^0$ window)	2,535,025	23.4%	99.6%
	Ξ_c^{**+} MVA	48,932	1.93%	27.1%
	Clone cuts	48,770	99.7%	99.7%

Table 4.10 Cut efficiencies from start to end on Ξ_c^{**+} candidates. The last column shows efficiencies when cuts are applied to a truth-matched $\Xi_c(2815)^+$ simulation sample.

Figure 4.20 shows the right-sign ΔM distribution after all selections are applied, where

$$\Delta M = m[\Xi_c^{**+}] - m[pK^-\pi^+] - m[\pi^-] - m[\pi^+]. \tag{4.14}$$

At least four prominent peaks are observed, and known PDG values for similar and known states or their isospin partners are overlaid on the plot. There is a generally good agreement of the background shape between the right-sign channel, wrong-sign $\Xi_c^+\pi^+\pi^+$, and $\Xi_c^+\pi^-\pi^-$ channels. In Figure 4.21, the effect of applying each stage of selections on simulated signal and wrong-sign data is demonstrated by plotting histograms of ΔM after each stage. In these figures, there is a peaking background that is prominent in the wrong-sign (*WSPMM*) sample. This is further studied in the following subsection to ensure that there are no peaks in the right-sign channel as an artefact of the selections described in this chapter.

Peaking background source in wrong-sign sample

For the wrong-sign $\Xi_c^+\pi^-\pi^-$ (*WSPMM*) channel (see Figure 4.21), a peaking background at ≈ 80 MeV is noticeable, and is more prominent in Figure 4.21(middle) after Stage II requirements, and especially Figure B.1 just before the data are used as background proxy in training the Stage III MVA.

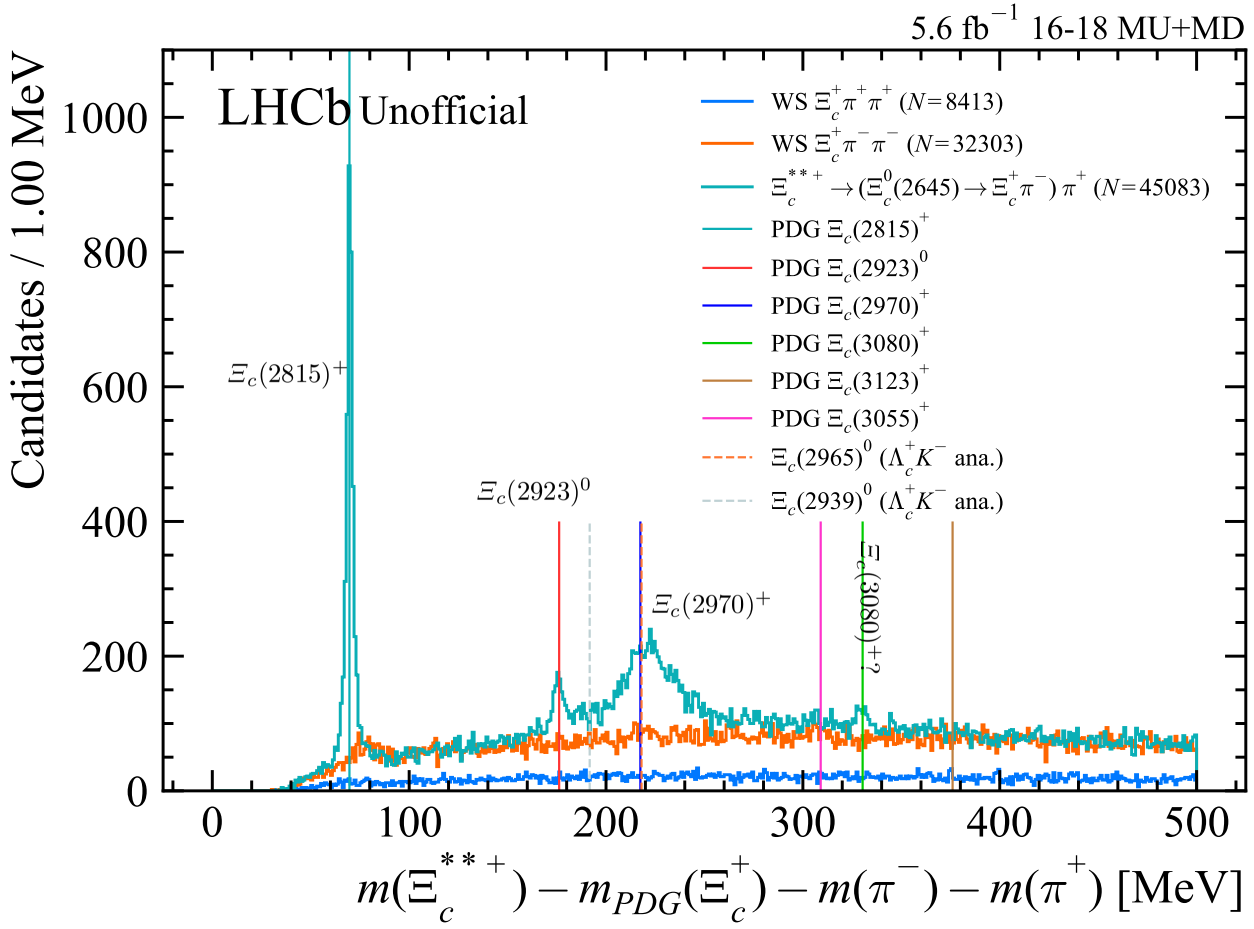

Figure 4.20 The unblinded Ξ_c^{**+} ΔM mass distribution. Each resonance has an overlaid line at the known PDG mass values of identical or similar (isospin partner) states.

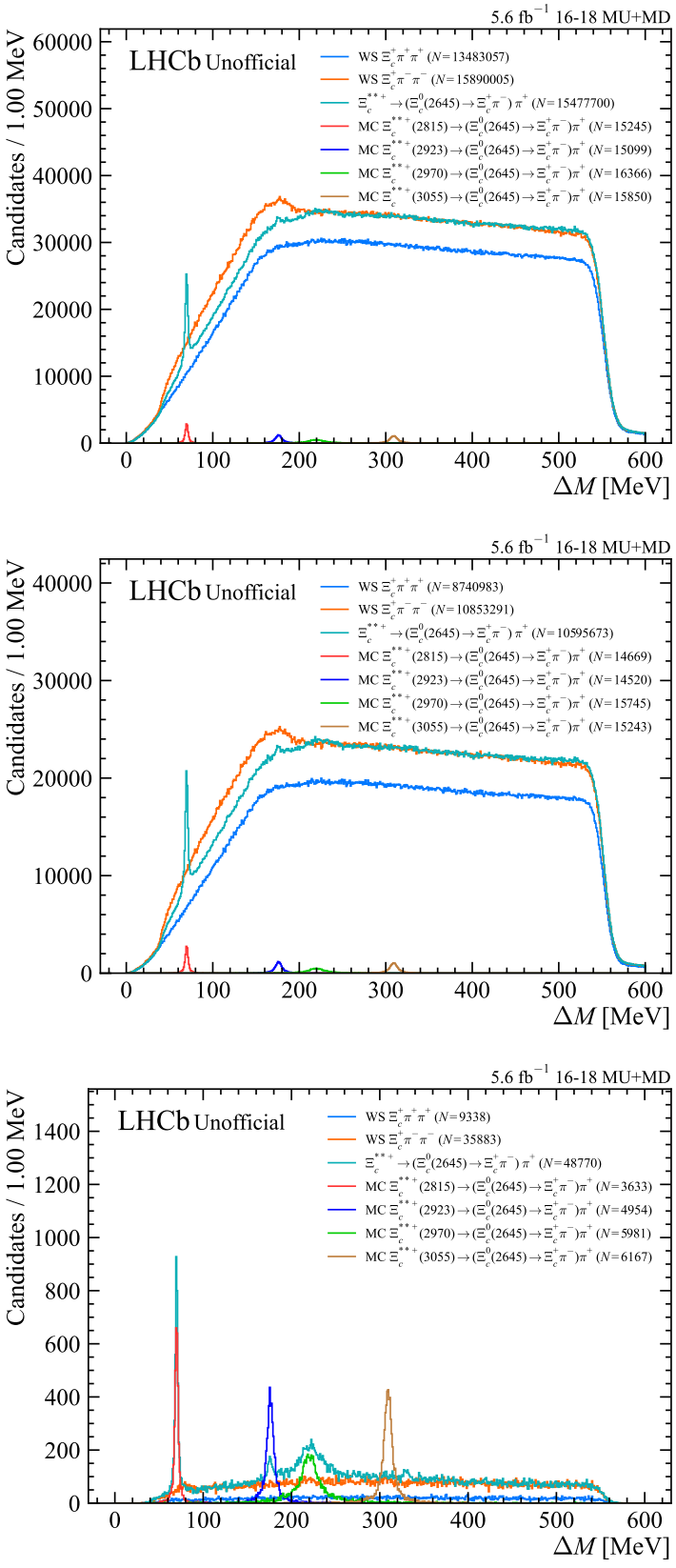

Figure 4.21 Mass distribution after stage I (first), stage II (middle), stage III (last) selections. Right-sign distribution shown for comparison.

This peaking background can be explained as a kinematic sculpting effect that can only be present in the wrong-sign $\Xi_c^+\pi^-\pi^-$ channel, primarily driven by a cut on the invariant mass $m[\Xi_c^+\pi^-]$. There is initially a cut imposed on the $m[\Xi_c^+\pi^-] - m[\Xi_c^+] - m[\pi^-] < 150\,\text{MeV}$, as described in Section 4.2, from which the background component in the region $[40, 200]\,\text{MeV}$ in Figure 4.21(middle) arises. This component is 'sculpted' into a peak after applying the stage II mass window ($[2635, 2656]\,\text{MeV}$), which leads to the feature shown in Figure B.1 at ≈ 80 MeV.

The source of this background component begins with combinations of $\Xi_c^+\pi^-$, observed to have a large fraction of $\Xi_c(2645)^0$ passing the stage II mass window requirement. Expanding this to the wrong-sign $\Xi_c^+\pi_1^-\pi_2^-$ system, it is possible to obtain candidates with $\Xi_c(2645)^0 \to \Xi_c^+\pi_2^-$, and/or $\Xi_c(2645)^0 \to \Xi_c^+\pi_1^-$. In this system, it becomes possible to reconstruct candidates where both $\Xi_c^+\pi_1^-$ and $\Xi_c^+\pi_2^-$ combinations pass the stage II mass window requirement resulting in a peak-like structure at approximately 80 MeV. See Figure 4.22. The position and width of the peak itself is affected by the choice of the mass window requirement, and this effect cannot be present in the right-sign $\Xi_c^+\pi^-\pi^+$ because there is no $\Xi_c(2645)^0$ in the $\Xi_c^+\pi_2^+$ system.

This hypothesis can be further verified using available simulation samples (Section 3.2.2). For a wrong-sign $\Xi_c^+\pi_1^-\pi_2^-$ MC sample, where selection requirements up to and including stage II are applied, a ΔM histogram is created for each of the following truth-matching requirements:

1. Require $\Xi_c^+\pi_1^-$ to originate from a $\Xi_c(2645)^0$. No matching for π_2^-.
2. (1) with mass window $[2635, 2656]\,\text{MeV}$ applied to the $\Xi_c^+\pi_1^-$ system.
3. Require $\Xi_c^+\pi_2^-$ to originate from a $\Xi_c(2645)^0$. No matching for π_1^-.
4. (3) with mass window $[2635, 2656]\,\text{MeV}$ applied to the $\Xi_c^+\pi_1^-$ system.

See Figure 4.23 for the results, and Figure 4.24 for an alternative view of the same distributions.

In both figures, a bump at the same position as in the WS data is observed after applying the stage II mass window requirement. It is concluded that this feature appears only in the wrong-sign channel because there is a possibility of the real $\Xi_c(2645)^0$ daughter pion being swapped with another random same-sign pion from the event, and this does not happen in the right-sign channel.

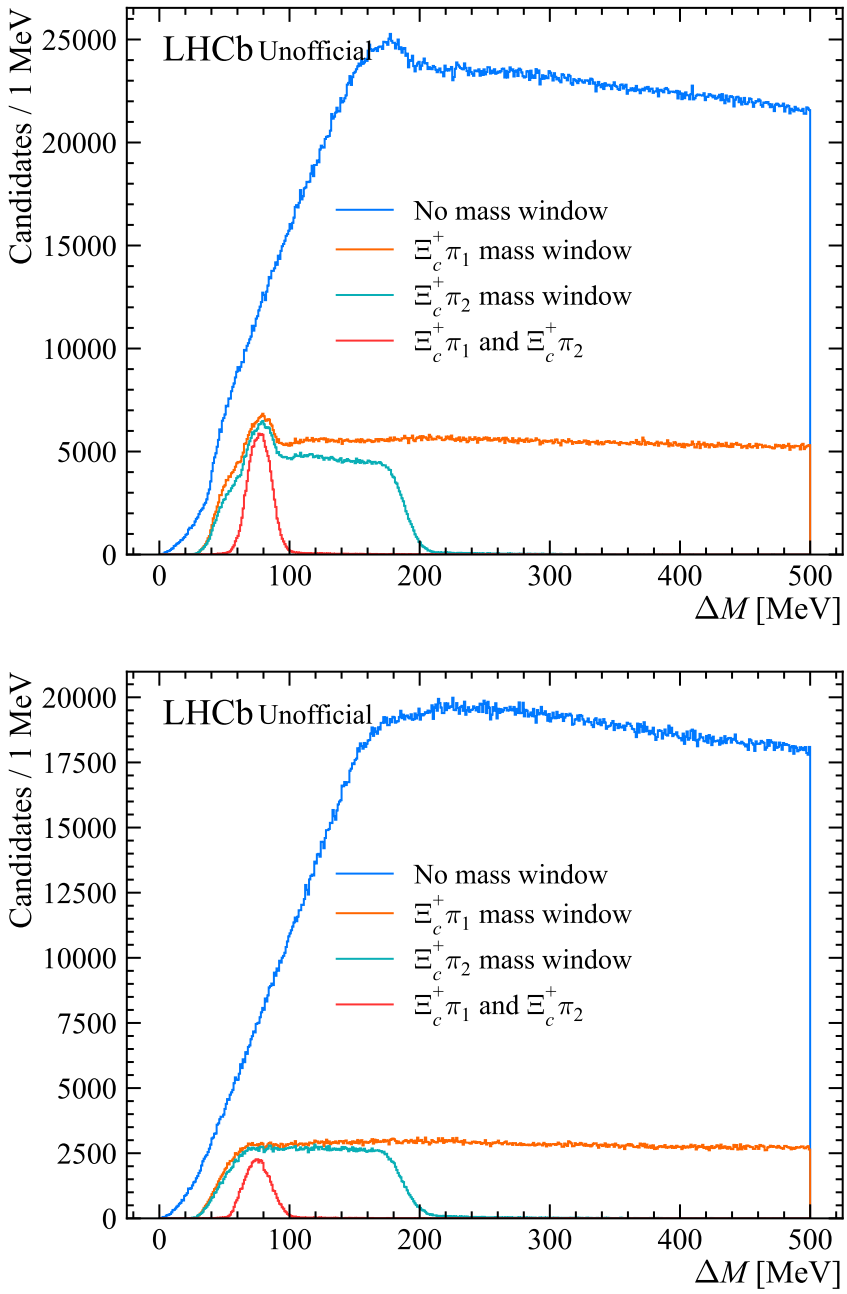

Figure 4.22 Depiction of kinematic sculpting effect leading to a bump at $\approx 80\,\text{MeV}$. A mass window $[2635, 2656]\,\text{MeV}$ is applied to the system $\Xi_c^+\pi_1\pi_2$ for $\Xi_c^+\pi_1$ only, $\Xi_c^+\pi_2$ only, and both combinations together. Plots are created for both WS $\Xi_c^+\pi^-\pi^-$ (first), and $\Xi_c^+\pi^+\pi^+$ (last).

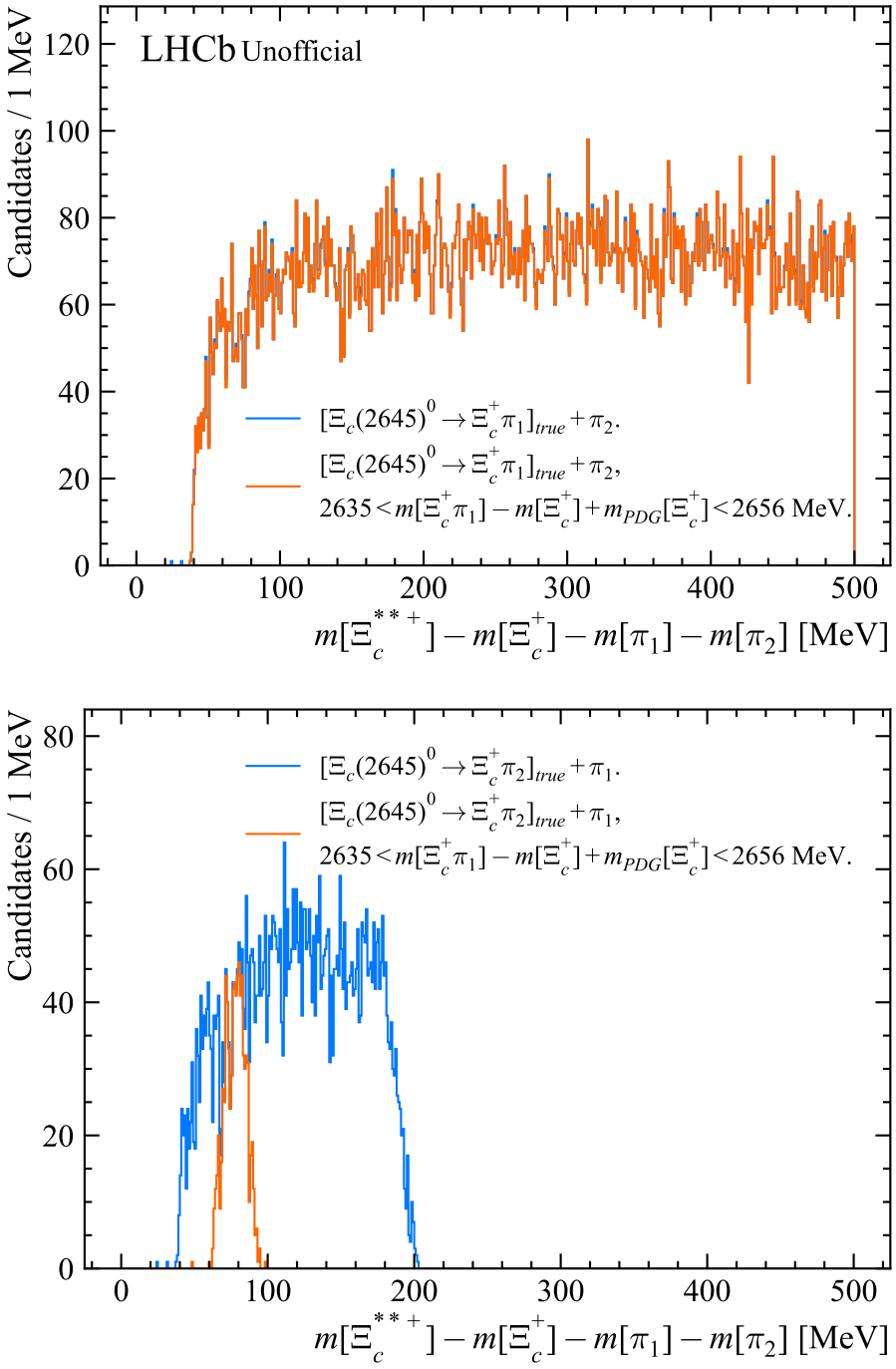

Figure 4.23 Plots showing ΔM distribution for truth-matched candidates with $\Xi_c(2645)^0 \to \Xi_c^+\pi_1^-$ (top) and $\Xi_c(2645)^0 \to \Xi_c^+\pi_2^-$ (bottom) in the wrong-sign $\Xi_c^+\pi^-\pi^-$ channel.

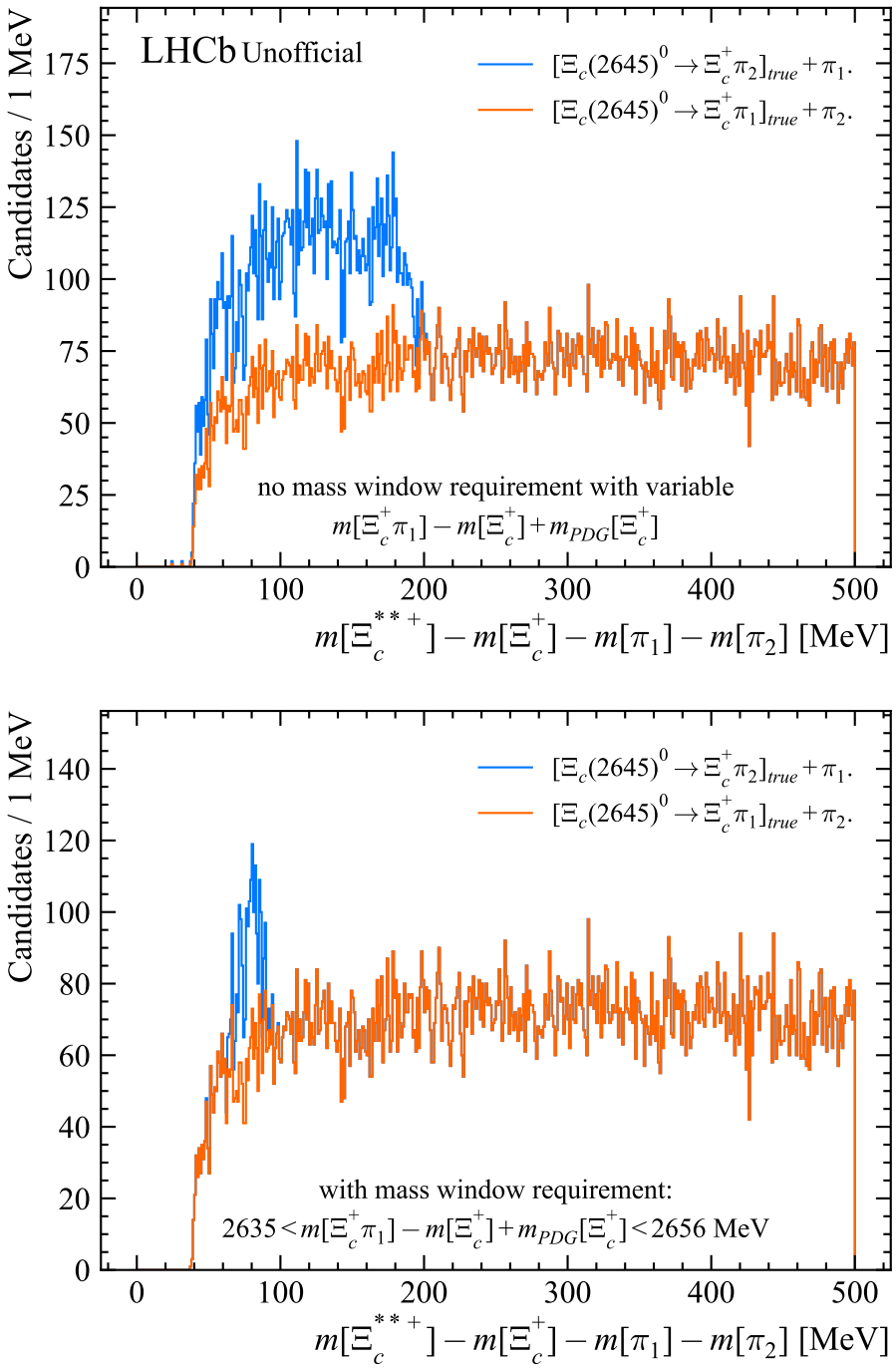

Figure 4.24 A histogram stack plot which more closely resembles the two components of wrong-sign background observed in wrong-sign $\Xi_c^+\pi^-\pi^-$ data. The top plot shows the two components before the stage II $\Xi_c^+\pi^-$ mass window, and again (lower) having applied the mass window.

Chapter 5

Simulation studies

This section is dedicated to outlining the details of studies involving only the simulation samples, which are relevant to estimates of the mass resolutions for the states discussed in later chapters, and also inform elements of the event selection described in the previous section. The simulation samples used are those described in Section 3.2.2.

5.1 Truth matching

Simulated events undergoing reconstruction do not always reflect accurately what a genuine simulated decay would look like after reconstruction, for a multitude of reasons (multiple ways to reconstruct the decay, among other inefficiencies). This can be mitigated by matching the reconstructed simulated decay with the underlying generated objects to ensure that the reconstructed decays being studied correspond to the genuine generated decays of interest in a process called truth matching. Truth matching is performed by requiring that the particle IDs of Ξ_c^{**+}, $\Xi_c(2645)^0$, Ξ_c^+, and two bachelor pions $\pi^+\pi^-$ match the expected particle IDs. In practice, these 'IDs' are integers assigned to particles by the Particle Data Group which encapsulate the particle identity, signed by their charge. The 'bachelor' pion IDs are matched where applicable.

5.2 $\Xi_c(2645)^0$ mass resolution

In order to improve the quality of a fit to the $\Xi_c(2645)^0$ mass distribution, Monte Carlo (MC) simulation are used to determine a model for the detector mass resolution that may be used to appropriately constrain a fit to data. Such a model is obtained by an unbinned maximum likelihood fit of a double Gaussian function to the following variable

$$x = M_{reco}\left[\Xi_c^+\pi^-\right] - M_{reco}[\Xi_c^+] + M_{PDG}[\Xi_c^+] - M_{TRUE}\left[\Xi_c(2645)^0\right], \qquad (5.1)$$

where M_{reco} refers to the reconstructed mass, and M_{TRUE} refers to the particle "truth" mass in the generated MC event, where reconstructed candidates are required to have a genuine matched $\Xi_c(2645)^0$, and Ξ_c^+, π^- daughter particles.

Since the results of this fit are to be used in the stage II selection to appropriately constrain the expected mass resolution contribution to the $\Xi_c(2645)^0$ mass resonance shape in data, stage I selection requirements (defined in the previous chapter) are applied, and the MC samples have all been reconstructed and processed in the same way as the data would be. When using the resulting parameters to fit real data, the width ($\sigma_{1,2}$) and fraction parameter of the double Gaussian are used and kept fixed. However, an additional floating scale parameter S is introduced which scales both widths up or down, to account for potential data-MC mismatches. The parameter S as introduced in this context is specific to this fit model.

The fit and resulting parameters are summarised in Figure 5.1.

5.3 Ξ_c^{**+} mass resolution

Monte Carlo simulated samples for simulated Ξ_c^{**+} states at various masses, specifically $M[\Xi_c^{**+}] = 2816.74, 2923.04, 2966.34, 3055.90\,\text{MeV}$ (as listed in Table 3.1) are also used to determine the Ξ_c^{**+} mass resolution and its dependence on the Ξ_c^{**+} mass in terms of the mass difference quantity ΔM (defined in Equation (4.14)).

To estimate the mass resolution as a function of the mass difference ΔM, a

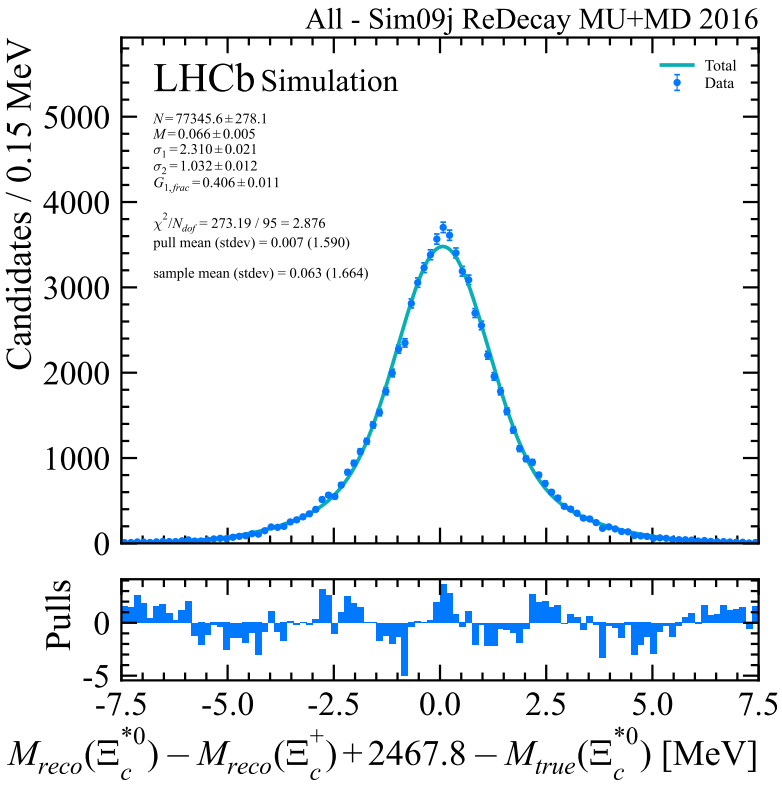

Figure 5.1 A fit to the difference between reconstructed and truth $\Xi_c(2645)^0$ mass. The fit model deviates close to the peak position, however this is not a crucial requirement as the primary goal for this study is to model the detector resolution width, which is achieved.

simultaneous fit to the variable

$$M_{res} = M_{\text{reco}} - M_{\text{true}}, \tag{5.2}$$

is performed to the four simulation samples (one per Ξ_c^{**+} mass hypothesis), where M_{reco} is the simulation reconstructed mass and M_{true} the simulation truth mass without reconstruction or resolution effects. A double Gaussian model with common parameters mean μ, σ_1, σ_2, and G_{frac} are fit to each simulation sample, a ΔM dependence is allowed for the overall width of the resolution distribution by introducing a scale factor S to each sample, multiplying both Gaussian widths (thus becoming $S_i\sigma_1$ and $S_i\sigma_2$ respectively). The scale factor S_{2815} for $\Xi_c(2815)^+$ is constrained to unity in order to fix the overall scale.

The results of this simultaneous fit are scale factors S_i ($i = 2923, 2970, 3055$) corresponding to each simulated sample and one set of double Gaussian parameters

$$\begin{aligned}
S_{2815} &= 1, \\
S_{2923} &= 1.63 \pm 0.03, \\
S_{2970} &= 1.90 \pm 0.03, \\
S_{3055} &= 2.30 \pm 0.04, \\
\mu &= 0.13 \pm 0.01\,\text{MeV}, \\
\sigma_1 &= 0.79 \pm 0.01\,\text{MeV}, \\
\sigma_2 &= 1.83 \pm 0.07\,\text{MeV}, \\
G_{frac} &= 0.89 \pm 0.01.
\end{aligned}$$

The fit model plotted over the individual simulation samples are also displayed in Figure 5.2. To facilitate a fit to ΔM in data, the fit model will need to have a well-defined resolution for some given value of ΔM, this is achieved by interpolating between the scale factor points by fitting a simple curve to the points to describe S as function of ΔM. Each scale factor is plotted against the true ΔM value corresponding to the simulated resonance, and a further fit performed with PDF

$$S(\Delta M) = 1 + a\left(\Delta M - \Delta M_{2815,\text{true}}\right)^b, \tag{5.3}$$

where $\Delta M_{2815,\text{true}}$ is the ΔM value corresponding to the simulated $\Xi_c(2815)^+$ resonance. This effectively provides a function which describes how the double

Gaussian, which models the mass resolution, scales in width as a function of ΔM, eliminating the need for free parameters relating to the mass resolution in the overall mass fit. The result of this study is used in Chapter 6.

The result of the fit is displayed in Figure 5.2 (top four plots), and the resulting function with the parameters determined from the fit are used directly in subsequent fits to the full unblinded ΔM distribution. The determined parameters are

$$a = 0.011 \pm 0.004,$$
$$b = 0.87 \pm 0.07,$$

where the fit $\chi^2/N_{dof} = 0.95$ indicates a good fit.

With good description of the Ξ_c^{**+} mass resolution, it is possible to proceed with fitting the data distribution.

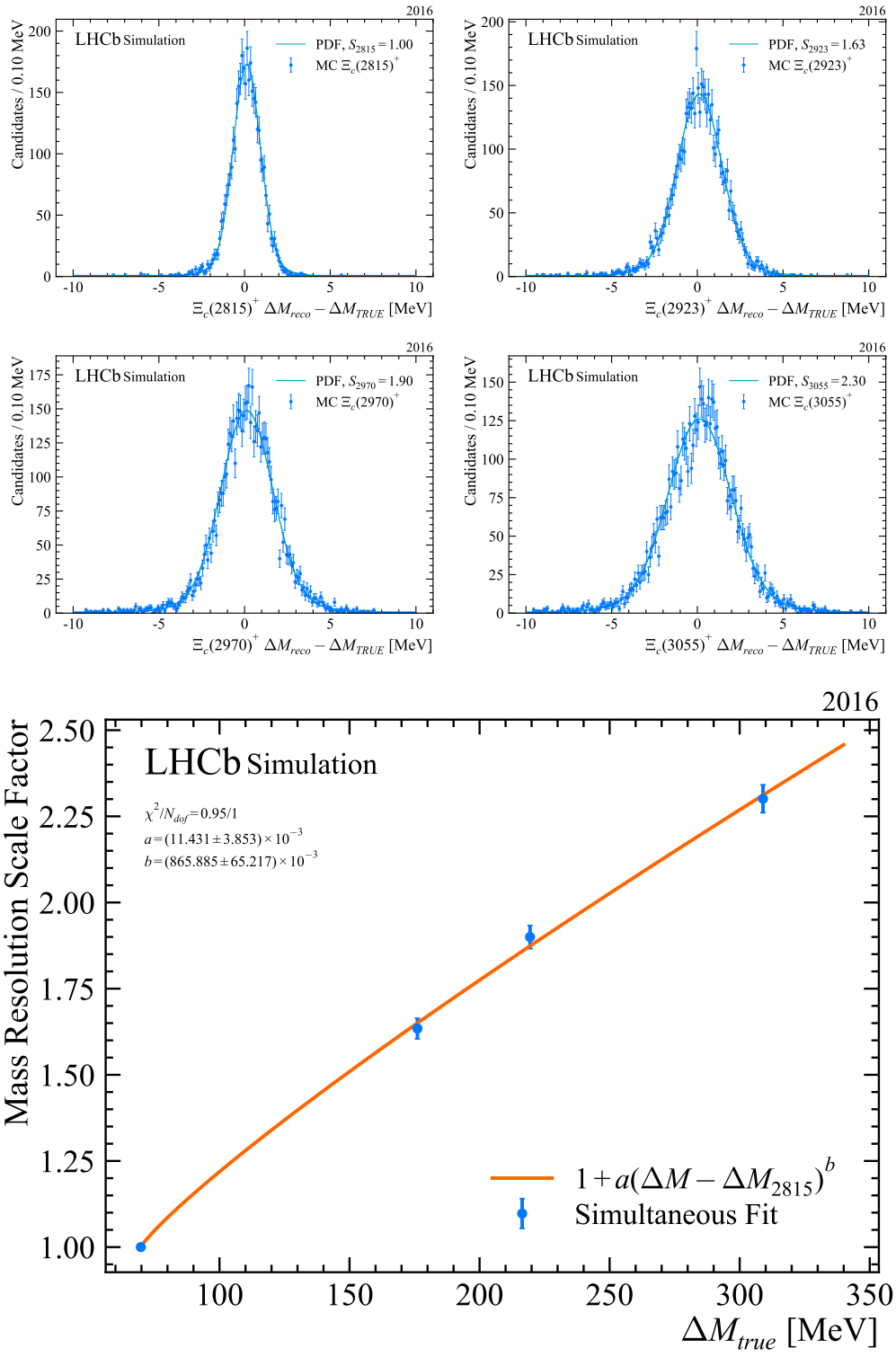

Figure 5.2 Simultaneous fit to Ξ_c^{**+} Monte Carlo samples (above) used to estimate the detector mass resolution vs. ΔM, and fit to determined scale factors S_i as a function of ΔM_{true} (below).

Chapter 6

Studying the Ξ_c^{**+} Mass Spectrum

This chapter covers the steps taken to understand the right-sign $\Xi_c^+\pi^-\pi^+$ mass spectrum after unblinding. The process begins with a development of an appropriate fit model that adequately describes the background shape and peaks. This is described in the section to follow. This is followed by a fit and a study of the resulting parameters, peak significances where necessary, and later on studies of the systematic errors where variations to the developed fit model are trialled to understand how the measured parameters are affected (Chapter 7).

6.1 Developing a fit model

The mass difference variable ΔM (Equation (4.14)) is calculated after using the DecayTreeFitter (DTF) [74] configured such that there is a PV and Ξ_c^+ mass constraint, therefore mass terms for the particles Ξ_c^+, π^+, π^- are fixed to the known PDG values [1]. Figure 6.1 is a plotted histogram of this variable, exhibiting at most four structures each of which may correspond to actual Ξ_c^{**+} excited states, and a low background throughout. It is hypothesised that these resonances from left to right are $\Xi_c(2815)^+$, $\Xi_c(2923)^+$, $\Xi_c(2970)^+$, and $\Xi_c(3080)^+$.

Each mass resonance is modelled by the convolution of a RBW (Equation (4.7)), and a double Gaussian

$$\mathrm{RBW}(\Delta M; M, \Gamma, L, m_1, m_2) \circledast \mathrm{DG}(\Delta M; \sigma_1, \sigma_2, G_{frac}), \tag{6.1}$$

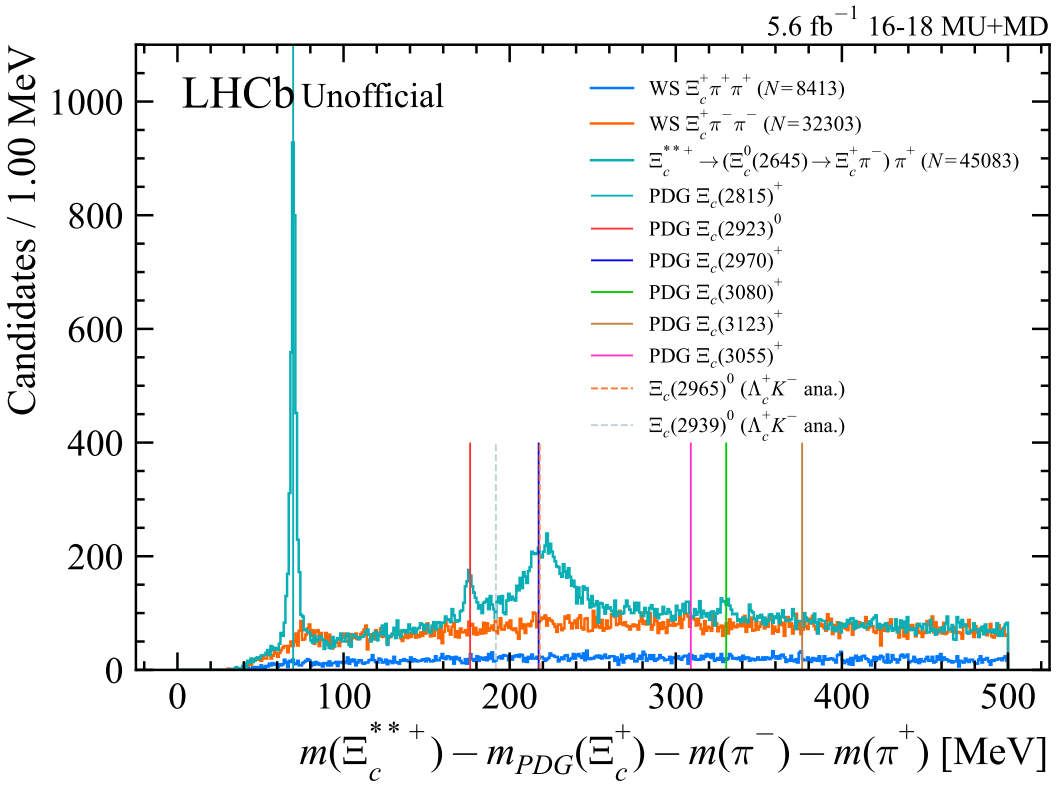

Figure 6.1 The unblinded Ξ_c^{**+} ΔM mass distribution. This figure is a partial duplicate of Figure 4.20, repeated for convenience.

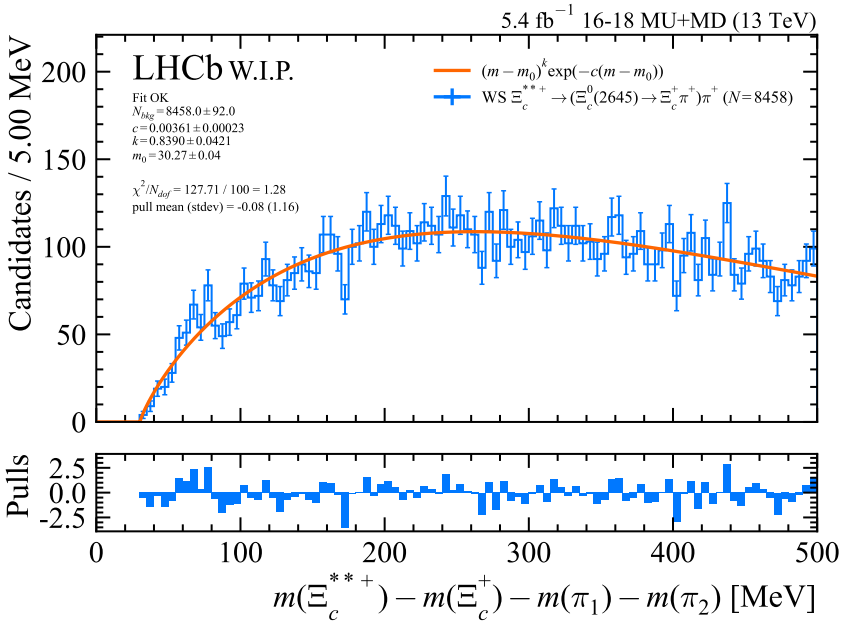

Figure 6.2 A fit of the background model, Equation (6.2), to the wrong sign ($\Xi_c^+\pi^+\pi^+$) background distribution.

where the parameters M (mass), Γ (width) are floating and $\sigma_1 = S(\Delta M)\sigma_{1,s}$, $\sigma_2 = S(\Delta M)\sigma_{2,s}$ where $\sigma_{i,s}$ come from those determined in Section 5.3 and $S(\Delta M)$ is defined in Equation (5.3). Hence, the double Gaussian function increases in width according to Equation (5.3) with increasing mass M. The angular momentum change L is chosen and fixed for each state as appropriate, and daughter masses are set as $m_1 = m[\Xi_c(2645)^0]$, $m_2 = m[\pi^+]$. These choices were accounted for in the systematic uncertainties (see Table 7.1).

The background model is determined empirically by studying wrong-sign background shapes such as the $\Xi_c^+\pi^+\pi^+$ channel, as shown in Figure 6.2. The chosen background model, also displayed in the figure fitted to the data, is defined by the function

$$f(x) = x^k \exp(-cx), \tag{6.2}$$

where $x = \Delta M - m_0$, and k (power), c (slope), m_0 (cutoff) are fit parameters.

These descriptions of the signal and background components are used to carry out an unbinned maximum likelihood fit to the mass spectrum in data, which is displayed in Figure 6.3. The goodness of fit is indicated from the χ^2/N_{dof} and pull distribution which indicates a good fit, with $\chi^2/N_{dof} \approx 1$, and generally well-behaved pull distribution throughout (mean zero and standard deviation of one). Parameters extracted from this fit are discussed in the sections to follow.

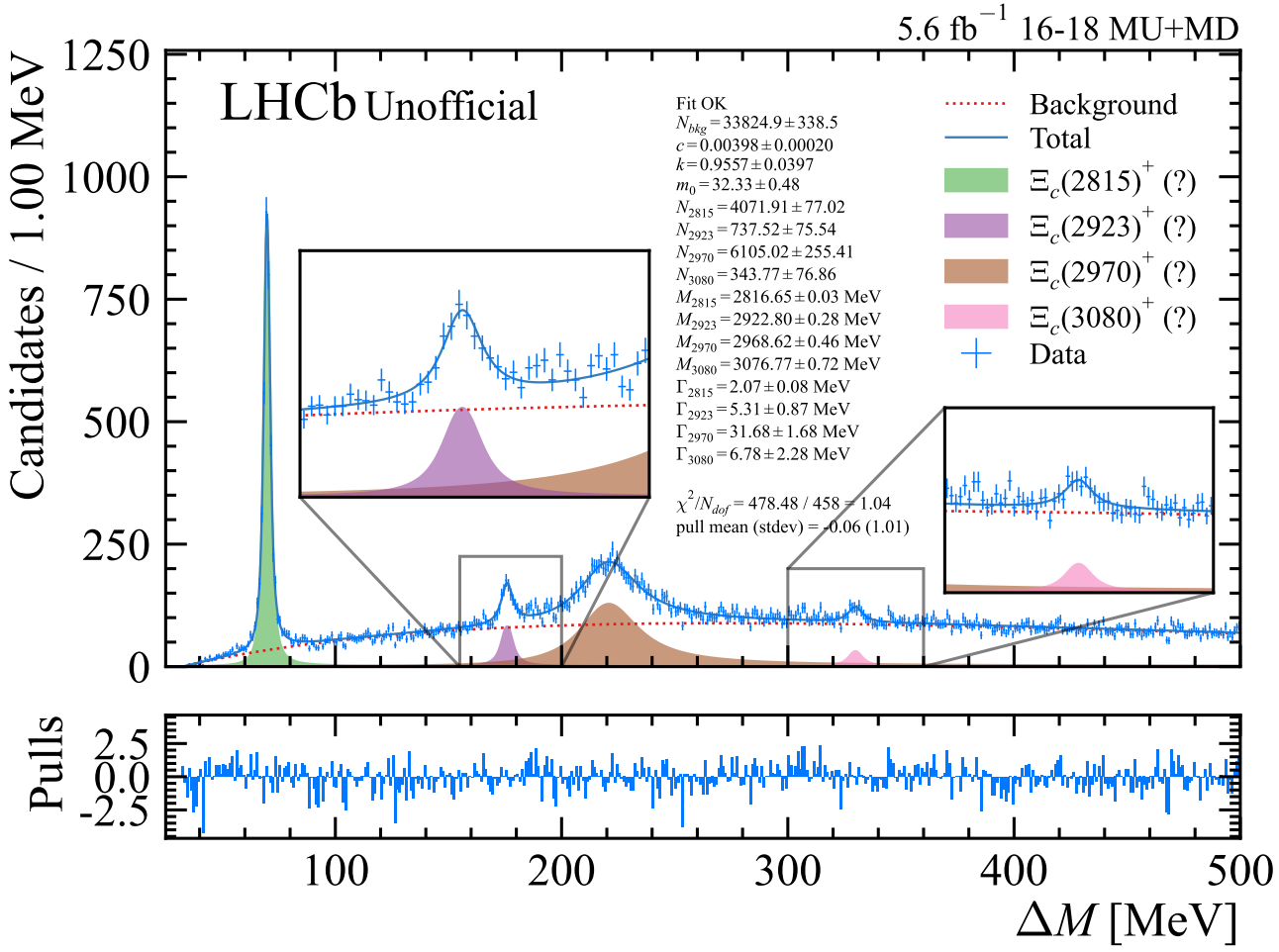

Figure 6.3 A plot of the full fit to the ΔM spectrum including all measured parameters and performance quantities. The regions close to the $\Xi_c(2923)^+$ and $\Xi_c(3080)^+$ resonances have been enlarged.

6.2 Fit results

To summarise all measured parameters from the fit for each observed state,

$$
\begin{aligned}
N[\Xi_c(2815)^+] &= 4072 \pm 77 \text{ candidates,} \\
\Delta M[\Xi_c(2815)^+] &= 69.80 \pm 0.03\,(\text{stat}) \pm 0.03\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(2815)^+] &= 2816.65 \pm 0.03\,(\text{stat}) \pm 0.03\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2815)^+] &= 2.07 \pm 0.08\,(\text{stat}) \pm 0.12\,(\text{syst})\,\text{MeV}, \\
N[\Xi_c(2923)^+] &= 738 \pm 76 \text{ candidates,} \\
\Delta M[\Xi_c(2923)^+] &= 175.95 \pm 0.28\,(\text{stat}) \pm 0.29\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(2923)^+] &= 2922.80 \pm 0.28\,(\text{stat}) \pm 0.29\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2923)^+] &= 5.3 \pm 0.9\,(\text{stat}) \pm 1.2\,(\text{syst})\,\text{MeV}, \\
N[\Xi_c(2970)^+] &= 6105 \pm 255 \text{ candidates,} \\
\Delta M[\Xi_c(2970)^+] &= 221.77 \pm 0.46\,(\text{stat}) \pm 0.30\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(2970)^+] &= 2968.62 \pm 0.46\,(\text{stat}) \pm 0.30\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2970)^+] &= 31.7 \pm 1.7\,(\text{stat}) \pm 1.7\,(\text{syst})\,\text{MeV}, \\
N[\Xi_c(3080)^+] &= 344 \pm 77 \text{ candidates,} \\
\Delta M[\Xi_c(3080)^+] &= 329.92 \pm 0.72\,(\text{stat}) \pm 0.11\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(3080)^+] &= 3076.77 \pm 0.72\,(\text{stat}) \pm 0.11\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(3080)^+] &= 6.8 \pm 2.3\,(\text{stat}) \pm 0.9\,(\text{syst})\,\text{MeV},
\end{aligned}
$$

where additionally the systematic uncertainties are shown. The details of their assignment and evaluation are described in Chapter 7. The uncertainties on the yields are statistical only, since the yields are not used to measure any reported properties of these particles. In the subsection following, significances will be presented for $\Xi_c(2923)^+$ and $\Xi_c(3080)^+$.

6.2.1 Significances of observed peaks

The significances are evaluated for the states $\Xi_c(3080)^+$ and $\Xi_c(2923)^+$ with Wilks' theorem [75], using the negative log-likelihood (NLL) values of the fit

where for each resonance the model is included and excluded (null hypothesis). The p-value for each resonance is calculated as

$$p = D_3 \left[\sqrt{2 \left(\mathrm{NLL}_{null} - \mathrm{NLL} \right)} \right] \tag{6.3}$$

where $D_3(x)$ is the complement of the cumulative distribution function (CDF) of the χ^2 distribution with three degrees of freedom (calculated by the function `ROOT::Math::chisquared_cdf_c`). The p-value is converted into a significance using the `RooStats::HypoTestResult` utility in ROOT [76].

The two significances are calculated as,

$$\begin{aligned} &\text{significance}[\Xi_c(2923)^+] = 14.6, \\ &\text{significance}[\Xi_c(3080)^+] = 5.4, \end{aligned}$$

and significances for the other two resonances are considered to be much greater than 5σ. For the $\Xi_c(3080)^+$ state, the significance resulting from using the alternative background model, as described in Section 7.1.2, is quoted as is a more conservative estimate. Following the same logic, the significance of the $\Xi_c(2923)^+$ is quoted from the fit using the nominal background model.

Chapter 7

Systematic uncertainties and cross checks

In the section to follow, the individual studies carried out to determine systematic assignments to the final measurements are discussed. In addition, a variety of cross-check studies are covered in Section 7.2, which did not contribute to the final systematic assignments as the variations of the measurements were considered negligible. The overall systematic assignments are summarised in Table 7.6.

7.1 Systematic sources

7.1.1 Fixed parameters of the signal fit model

Resonances in the mass spectrum are modelled by the RBW (Equation (4.7)). The parameter r is the Blatt–Weisskopf radius [71], fixed at $r = 3.1\,\text{GeV}^{-1}$, and L is the change in angular momentum relating to the resonance, nominally fixed to $L = 0$ for $\Xi_c(2815)^+$, $\Xi_c(2923)^+$, $\Xi_c(3080)^+$ and $L = 1$ for $\Xi_c(2970)^+$. A number of variations of the parameter r about the nominal value were tested to understand the change in resonance masses and widths resulting from the fit to ΔM. Table 7.1 shows the resulting change in masses and widths for all resonances with respect to the variations of L and r, and three approaches to assign a systematic error were tested. Disallowed choices of L are not considered when varying the parameter to evaluate systematic contributions.

Changed parameters	$\Xi_c(2815)^+$		$\Xi_c(2923)^+$		$\Xi_c(2970)^+$		$\Xi_c(3080)^+$	
	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 0, r = 0.002$	−0.0005	+0.0146	−0.0345	−0.5539	−0.0442	−0.2563	−0.0255	+0.5149
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 0, r = 0.003$	−0.0005	+0.0146	−0.0345	−0.5539	−0.0442	−0.2563	−0.0255	+0.5149
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 0, r = 0.004$	−0.0005	+0.0146	−0.0345	−0.5539	−0.0442	−0.2563	−0.0255	+0.5149
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 1, r = 0.002$	−0.0005	+0.0147	−0.0343	−0.5566	−0.0430	−0.2761	−0.0234	+0.3512
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 1, r = 0.003$	−0.0005	+0.0146	−0.0343	−0.5547	−0.0433	−0.2772	−0.0236	+0.3930
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 1, r = 0.004$	−0.0005	+0.0146	−0.0342	−0.5533	−0.0436	−0.2780	−0.0238	+0.4217
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 2, r = 0.002$	−0.0005	+0.0150	−0.0351	−0.5677	−0.0419	−0.2721	−0.0258	+0.1210
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 2, r = 0.003$	−0.0005	+0.0149	−0.0350	−0.5660	−0.0426	−0.2719	−0.0250	+0.1769
$L2815 = 0, L2923 = 0, L2970 = 0, L3080 = 2, r = 0.004$	−0.0005	+0.0149	−0.0345	−0.5617	−0.0422	−0.2743	−0.0234	+0.2398
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 0, r = 0.002$	+0.0006	−0.0206	+0.1315	+1.6233	+0.2525	+1.0499	+0.0271	−0.9349
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 0, r = 0.003$	+0.0005	−0.0186	+0.1098	+1.3823	+0.2007	+0.8547	+0.0266	−0.8254
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 0, r = 0.004$	+0.0005	−0.0162	+0.0894	+1.1443	+0.1369	+0.5686	+0.0239	−0.6932
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 1, r = 0.002$	+0.0007	−0.0206	+0.1305	+1.6178	+0.2546	+1.0584	+0.0190	−0.9697
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 1, r = 0.003$	+0.0005	−0.0186	+0.1097	+1.3773	+0.2023	+0.8709	+0.0220	−0.8580
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 1, r = 0.004$	+0.0005	−0.0162	+0.0889	+1.1397	+0.1376	+0.5750	+0.0210	−0.7022
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 2, r = 0.002$	+0.0006	−0.0200	+0.1308	+1.6148	+0.2553	+1.0781	+0.0064	−1.0302
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 2, r = 0.003$	+0.0005	−0.0186	+0.1091	+1.3726	+0.2051	+0.8849	+0.0102	−0.9194
$L2815 = 0, L2923 = 0, L2970 = 2, L3080 = 2, r = 0.004$	+0.0004	−0.0154	+0.0888	+1.1310	+0.1366	+0.5901	+0.0122	−0.7700
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 0, r = 0.002$	−0.0005	+0.0130	−0.0475	−0.3806	−0.0104	−0.5687	−0.0234	+0.4317
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 0, r = 0.003$	−0.0005	+0.0130	−0.0472	−0.3878	−0.0123	−0.5524	−0.0235	+0.4359
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 0, r = 0.004$	−0.0005	+0.0133	−0.0473	−0.4035	−0.0159	−0.5339	−0.0241	+0.4448
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 1, r = 0.002$	−0.0005	+0.0130	−0.0484	−0.3867	−0.0094	−0.5593	−0.0229	+0.3042
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 1, r = 0.003$	−0.0005	+0.0132	−0.0472	−0.3916	−0.0118	−0.5569	−0.0226	+0.3361
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 1, r = 0.004$	−0.0005	+0.0135	−0.0475	−0.4060	−0.0155	−0.5376	−0.0234	+0.3732
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 2, r = 0.002$	−0.0005	+0.0135	−0.0487	−0.3951	−0.0081	−0.5690	−0.0248	+0.0671
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 2, r = 0.003$	−0.0005	+0.0135	−0.0479	−0.4003	−0.0102	−0.5528	−0.0239	+0.1244
$L2815 = 0, L2923 = 2, L2970 = 0, L3080 = 2, r = 0.004$	−0.0005	+0.0137	−0.0475	−0.4125	−0.0137	−0.5353	−0.0234	+0.1933
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 0, r = 0.002$	+0.0006	−0.0224	+0.1491	+1.8351	+0.2655	+0.4270	+0.0265	−0.9850
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 0, r = 0.003$	+0.0005	−0.0204	+0.1271	+1.6196	+0.2174	+0.2719	+0.0262	−0.8673
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 0, r = 0.004$	+0.0004	−0.0177	+0.0976	+1.3272	+0.1568	+0.0951	+0.0235	−0.7343
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 1, r = 0.002$	+0.0006	−0.0224	+0.1490	+1.8339	+0.2665	+0.4349	+0.0194	−1.0199
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 1, r = 0.003$	+0.0005	−0.0204	+0.1270	+1.6185	+0.2188	+0.2771	+0.0213	−0.8945
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 1, r = 0.004$	+0.0004	−0.0178	+0.0993	+1.3413	+0.1565	+0.0800	+0.0204	−0.7528
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 2, r = 0.002$	+0.0005	−0.0227	+0.1513	+1.8513	+0.2621	+0.4093	+0.0055	−1.0988
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 2, r = 0.003$	+0.0005	−0.0203	+0.1258	+1.6071	+0.2219	+0.2989	+0.0093	−0.9602
$L2815 = 0, L2923 = 2, L2970 = 2, L3080 = 2, r = 0.004$	+0.0004	−0.0175	+0.0984	+1.3352	+0.1572	+0.0837	+0.0110	−0.8196
RMS	0.00	0.02	0.09	1.12	0.15	0.55	0.02	0.68
Half Largest	0.00	0.01	0.08	0.93	0.13	0.54	0.01	0.55
Largest	0.00	0.02	0.15	1.85	0.27	1.08	0.03	1.10

Table 7.1 The parameters L and r (GeV^{-1}) are fixed parameters in the final mass fit. Each row represents the results of an alternative fit using a different combination of values for these fixed parameters, such that the systematic contribution can be determined. Each cell represents the shift from the nominally measured value.

It was decided to pick the **RMS** method over assigning the maximum absolute (Largest) shifts as systematics to each measurement, as the latter is too conservative.

7.1.2 Background model

An alternative background composed of a third-order `RooPolynomial` PDF, with coefficients $a = 50$ (fixed), b, and c (floating), was tested as an alternative background model using the signal candidate sample, checking that the fit quality was of comparable quality to the nominal fit. In Figure 7.1, the fit with this alternative model is shown.

A comparison of the resulting parameters are tabulated in terms of parameter shifts from the nominal fit result, with associated systematic assignments, in Table 7.2. The systematic uncertainty is assigned as half of the shift value.

Changed parameters	$\Xi_c(2815)^+$		$\Xi_c(2923)^+$		$\Xi_c(2970)^+$		$\Xi_c(3080)^+$	
	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)
Alternative Model (shift)	-0.0017	$+0.0632$	-0.0052	-0.5207	-0.2429	-3.1027	$+0.0403$	-1.2119
Assigned	0.00	0.03	0.00	0.26	0.12	1.55	0.02	0.61

Table 7.2 Table of shifts and assigned systematic uncertainty values with respect to the nominal fit and alternative background fit. The assignment to each parameter is taken as half of the absolute value of the parameter shift.

7.1.3 MC-data disagreement

The systematic contribution arising from disagreements between detector resolution effects in simulation and data are studied by applying a scale factor to all resolution function width parameters of 0.9 and 1.1 ($\pm 10\%$). This choice of $\pm 10\%$ is motivated by the fit in Figure 4.10 that test the Data-MC agreement in resolution via the parameter S, showing $\approx 10\%$ variation. The shifts of measured parameters and assigned uncertainties are noted in Table 7.3, with assignments chosen as the values determined by the RMS method.

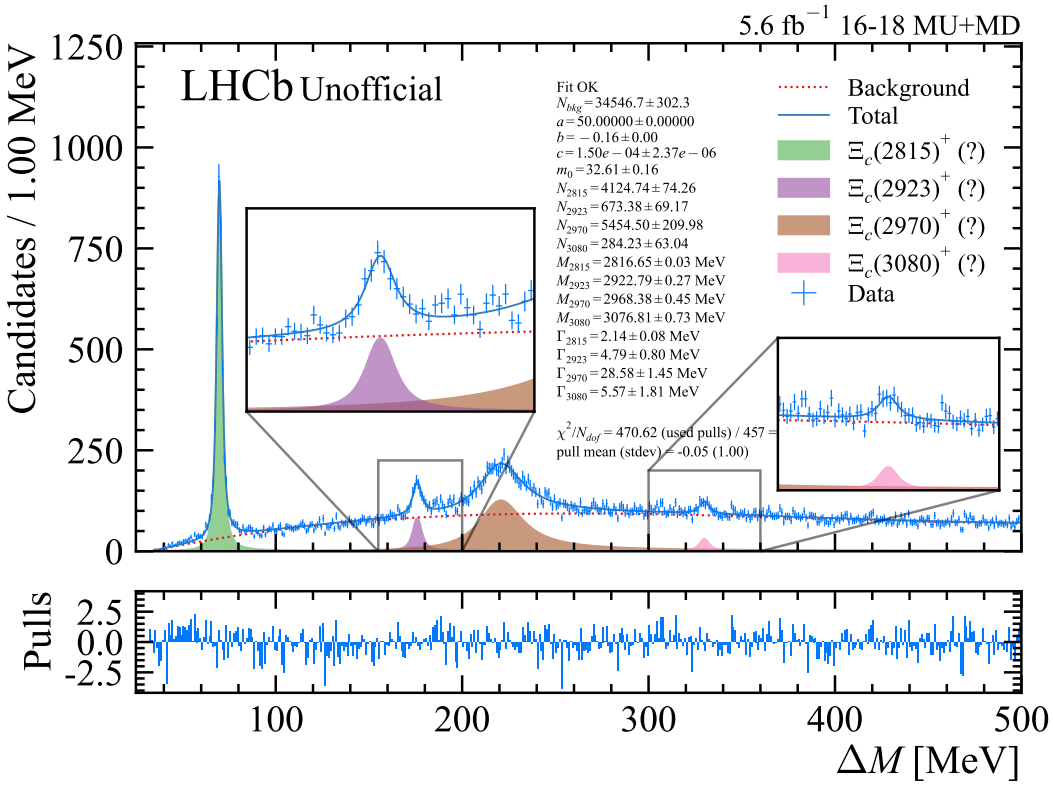

Figure 7.1 A version of the fit using an alternative background model. The relevant background model parameters are $N_{bkg} = 34546 \pm 302$, $b = -0.161 \pm 0.001$, $c = (1.50 \pm 0.02) \cdot 10^{-4}$.

Changed parameters	$\Xi_c(2815)^+$		$\Xi_c(2923)^+$		$\Xi_c(2970)^+$		$\Xi_c(3080)^+$	
	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)
$S = 0.9$	+0.0005	+0.1150	−0.0009	+0.0957	−0.0098	−0.0444	−0.0036	+0.0838
$S = 1.1$	−0.0005	−0.1203	+0.0015	−0.1083	+0.0002	−0.0265	+0.0012	−0.0003
RMS	0.00	0.12	0.00	0.10	0.01	0.04	0.00	0.06
Half Largest	0.00	0.06	0.00	0.05	0.00	0.02	0.00	0.04
Largest	0.00	0.12	0.00	0.11	0.01	0.04	0.00	0.08

Table 7.3 Data-MC discrepancy systematic assignments (RMS values are assigned as systematics).

Contribution	$\Xi_c(2815)^+$		$\Xi_c(2923)^+$		$\Xi_c(2970)^+$		$\Xi_c(3080)^+$	
	m (MeV)	Γ (MeV)	m (MeV)	Γ (MeV)	m (MeV)	Γ (MeV)	m (MeV)	Γ (MeV)
Momentum scaling	0.0209	–	0.0527	–	0.0664	–	0.0989	–
Energy loss	0.02	–	0.02	–	0.02	–	0.02	–

Table 7.4 Momentum scaling and energy loss systematic uncertainty assignments.

7.1.4 Momentum scaling uncertainty and energy loss

These two systematic contributions originate from the momentum scaling calibration [51], and detector material effects (referred to as 'energy loss' systematic), which affect all mass measurements as summarised in Table 7.4.

Mass measurements can be affected by systematic mismeasurement of particle momenta, which can arise from e.g. imperfect material scattering or the overall momentum scale of the detector (imperfect detector alignment, or imperfect modelling of the magnetic field).

The momentum scaling calibration, applied as discussed in Section 3.2.2, introduces a contribution to systematic uncertainty corresponding to 0.03% of the Q-value (ΔM) [51]. The energy loss systematic originates from imperfect modelling of the detector material which results in an assignment of $2 * 0.01 = 0.02\,\text{MeV}$ to all mass measurements (i.e. two pion tracks, as the Ξ_c^+ mass is subtracted). This is motivated by the energy loss correction systematic for $D^0 \to K^+K^-K^-\pi^+$ from Ref. [51] which is listed as 0.03 MeV, this means a contribution of about 0.01 MeV per track.

7.1.5 Selection-induced mass bias

To confirm to what degree these mass measurements are biased by the selection, the available simulation samples were used in a study involving fits to the variable

$$\Delta M_{reco} - \Delta M_{true}, \tag{7.1}$$

i.e. the mass resolution, in order to determine the mean parameter of a double or triple Gaussian function, whichever is most appropriate. For each Ξ_c^{**+} simulation sample, the difference (shift) in mean parameter μ was determined where (a) no selections other than those in Section 4.2 are applied and (b) all selections are applied. It is understood that a significant difference in μ between sample (a) and

(b) may indicate a selection-induced bias. The results of the study are presented and discussed in Figure 7.2 and 7.3.

As a result of this study, it can be argued that the selection-induced bias is not significant (consistent with zero). Nonetheless, a conservative systematic error on all mass measurements is assigned as the mean of all the calculated μ-shifts, as 0.016 ± 0.012 MeV.

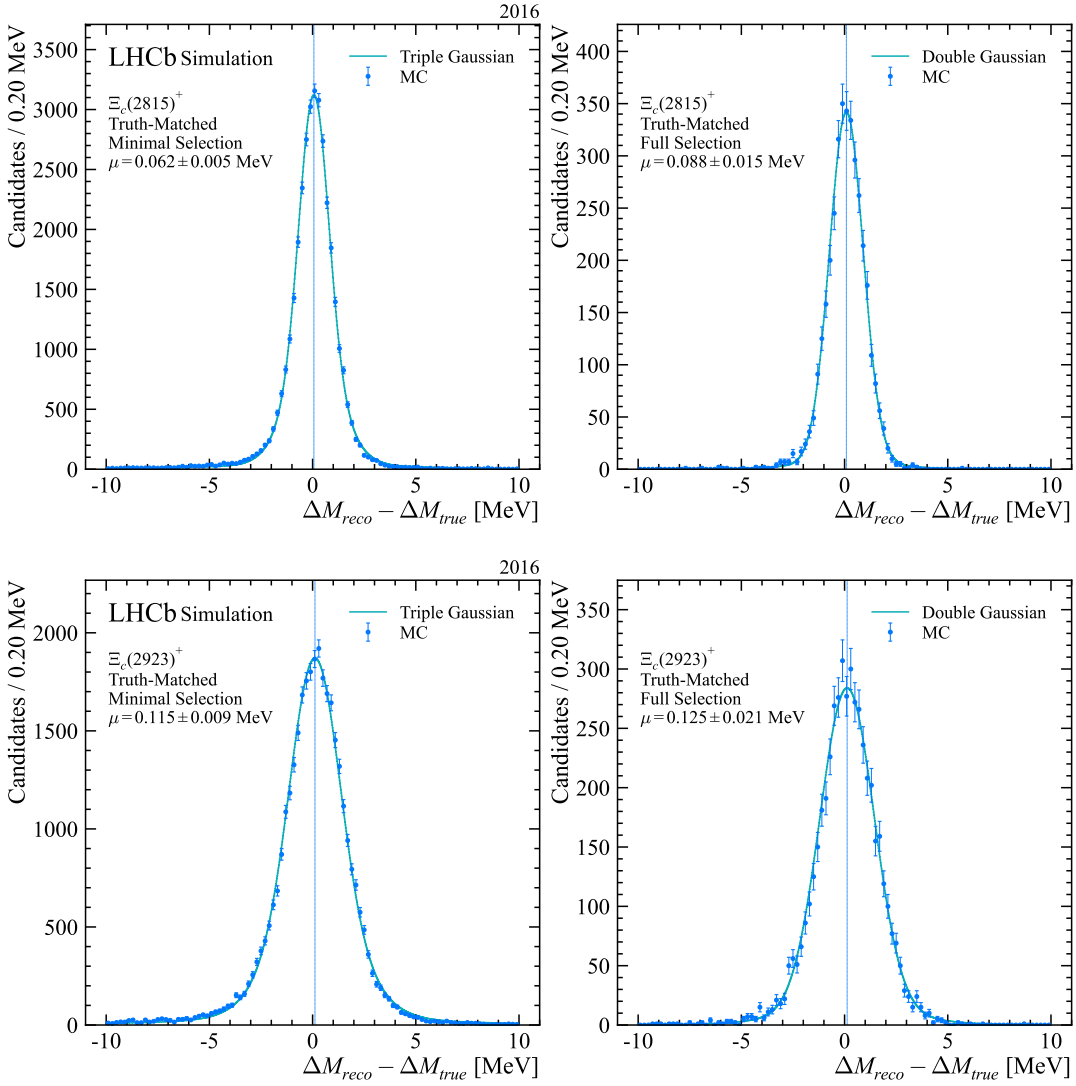

Figure 7.2 Plots showing unbinned negative log-likelihood (NLL) fits to $\Xi_c(2815)^+$ (first row) and $\Xi_c(2923)^+$ (second row). The first column shows fits to the "minimal" selection samples (a triple Gaussian function has been used in the fit model), and second column to "full" selection samples (a double Gaussian function is used) where all selections have been applied. All simulation samples shown are truth-matched. The fit is used to extract the mean of the Gaussian μ, which is used to determine a μ-shift, defined as the difference in μ determined for the "minimal" sample and "full" selected samples, which quantifies a mass bias caused by the selection.

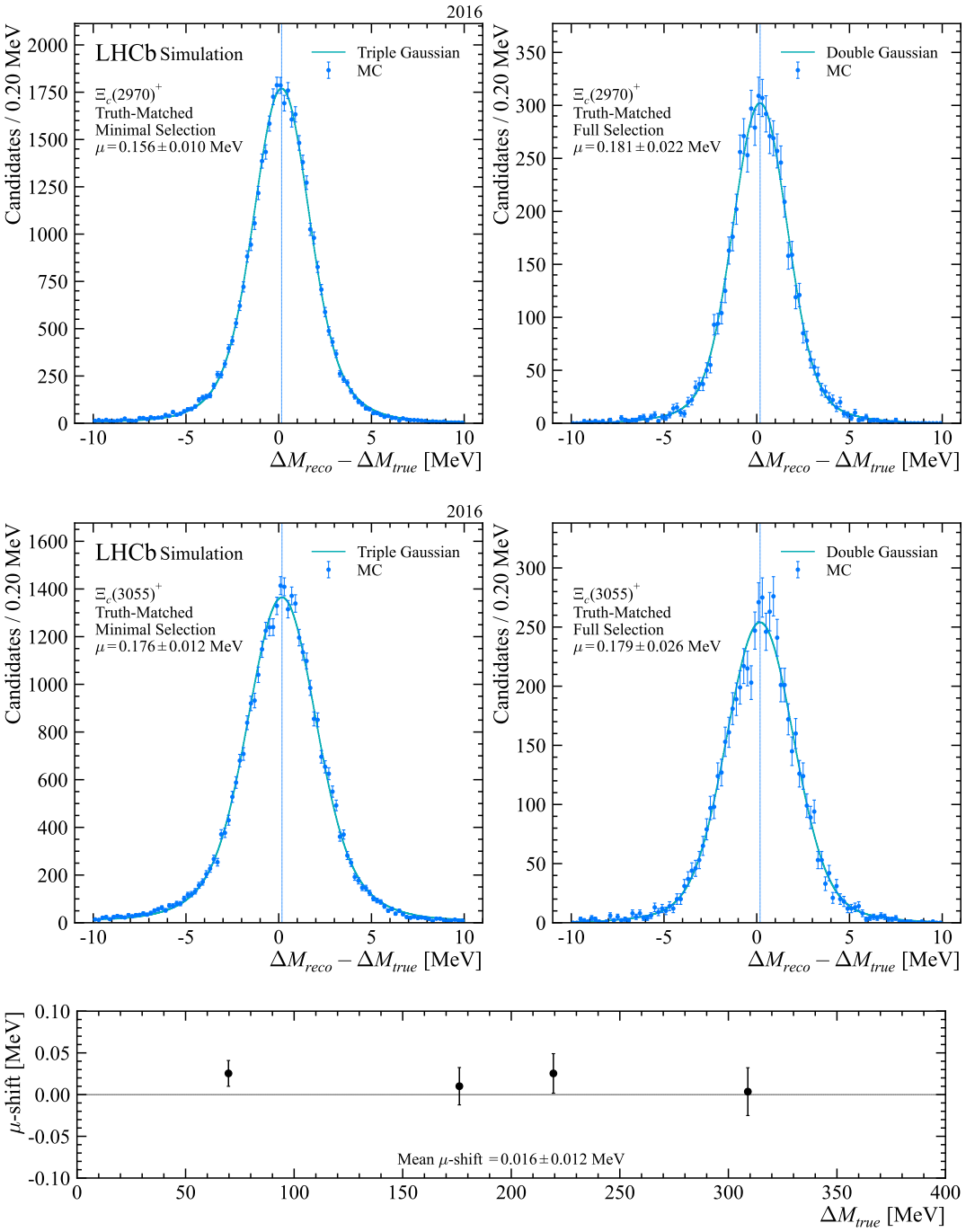

Figure 7.3 Plots showing unbinned NLL fits to $\Xi_c(2970)^+$ (first row) and $\Xi_c(3055)^+$ (second row). The third row shows a summary of all μ-shifts against the corresponding ΔM_{true} value for the MC sample.

7.1.6 Interference effects between the $\Xi_c(2923)^+$ and $\Xi_c(2970)^+$ resonances

The possibility of interference effects between the resonances $\Xi_c(2923)^+$ and $\Xi_c(2970)^+$ are explored by introducing an interference term into the signal model for these resonances, where the degree of the interference effect is parameterised by a phase angle ϕ. The double Relativistic Breit–Wigner PDF is defined

$$f_{DBW}(m; M_1, \Gamma_1, M_2, \Gamma_2, n_0, \phi) = \\ T_1(m; M_1, \Gamma_1) + n_0^2 T_2(m; M_2, \Gamma_2) + n_0 T_{12}(m; M_1, \Gamma_1, M_2, \Gamma_2, \phi), \quad (7.2)$$

where m is the Ξ_c^{**+} invariant mass variable, n_0 parameterises the normalisation of the second resonance, M_i, Γ_i are mass and width parameters for the resonances, $T_1(m)$, $T_2(m)$, $T_{12}(m)$ refer to the first resonance term, second resonance term, and the interference term. The first and second terms are defined

$$T_i(m; M_i, \Gamma_i) = \frac{C_i^2}{A_i^2 + B_i^2}, \quad (7.3)$$

with

$$A_i(m, M_i) = m^2 - M_i^2, \quad (7.4)$$

$$B_i(m, M_i, \Gamma_i) = M_i \Gamma(m; M_i, \Gamma_i), \quad (7.5)$$

$$C_i(m, M_i, \Gamma_i) = \sqrt{M_i \Gamma(m; M_i, \Gamma_i)}, \quad (7.6)$$

and $\Gamma(m; M_i, \Gamma_i)$ is the mass-dependent width, as previously defined in Equation (4.8). The interference term is

$$T_{12}(m; M_1, \Gamma_1, M_2, \Gamma_2, \phi) = \frac{2C_1C_2}{(A_1^2 + B_1^2)(A_2^2 + B_2^2)} \cdot \\ [(A_1A_2 + B_1B_2)\cos\phi + (A_1B_2 - B_1A_2)\sin\phi], \quad (7.7)$$

where the phase angle ϕ parameter appears.

The result of the fit is shown in Figure 7.4. Table 7.5 summarises differences in mass, width, and background parameters. Systematic uncertainties are assigned for mass and width measurements, taking half of the shift value from the table.

State	Parameter	Change	Change / Stat. err.
Background	Δm_0	-0.3013	0.63
	Δk	$+0.0397$	1.00
	Δc	$+0.0001$	0.68
$\Xi_c(2815)^+$	Δm	-0.0007 MeV	0.02
	$\Delta\Gamma$	$+0.0134$ MeV	0.16
$\Xi_c(2923)^+$	Δm	-0.5316 MeV	1.91
	$\Delta\Gamma$	-0.6957 MeV	0.80
$\Xi_c(2970)^+$	Δm	$+0.4387$ MeV	0.96
	$\Delta\Gamma$	-1.0541 MeV	0.63
$\Xi_c(3080)^+$	Δm	-0.0262 MeV	0.04
	$\Delta\Gamma$	$+0.3162$ MeV	0.31

Table 7.5 Table of parameter difference (and uncertainty) between the nominal fit and a fit with interference effects included in the model for $\Xi_c(2923)^+$ and $\Xi_c(2970)^+$. The fourth column shows the change divided by the statistical precision of the nominal fit parameter.

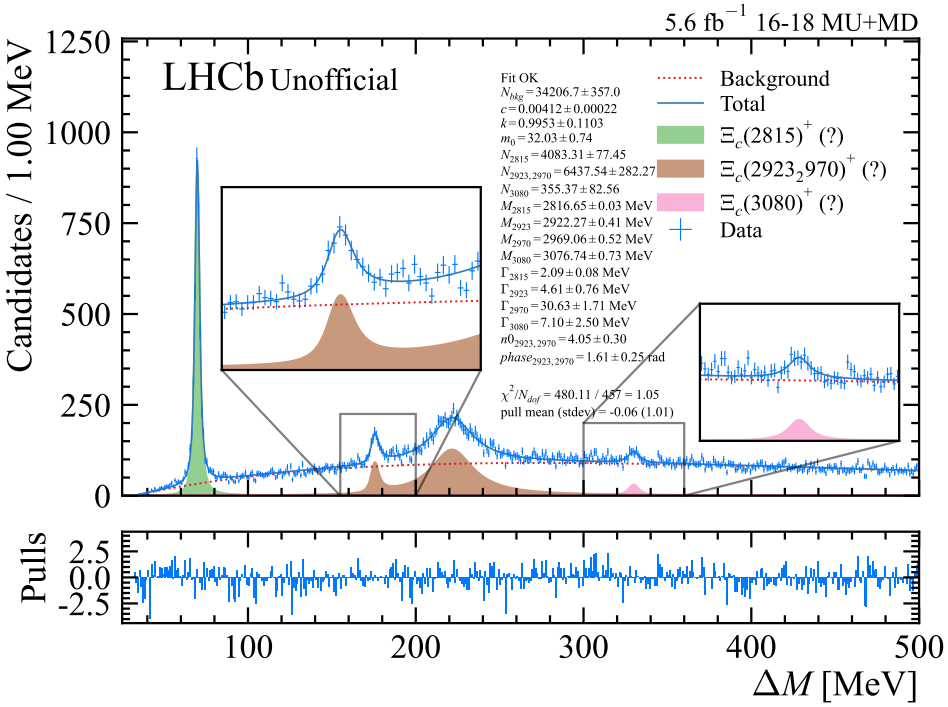

Figure 7.4 Re-running the fit with interference effects included in the signal model.

Contribution	$\Xi_c(2815)^+$		$\Xi_c(2923)^+$		$\Xi_c(2970)^+$		$\Xi_c(3080)^+$	
	m (MeV)	Γ (MeV)	m (MeV)	Γ (MeV)	m (MeV)	Γ (MeV)	m (MeV)	Γ (MeV)
RBW parameters	0.00	0.02	0.09	1.12	0.15	0.55	0.02	0.68
MC-Data disagreement	0.00	0.12	0.00	0.10	0.01	0.04	0.00	0.06
Alternative background	0.00	0.03	0.00	0.26	0.12	1.55	0.02	0.61
Momentum scaling	0.02	–	0.05	–	0.07	–	0.10	–
Energy loss	0.02	–	0.02	–	0.02	–	0.02	–
Selection-induced mass bias	0.02	–	0.02	–	0.02	–	0.02	–
2923-2970 interference model	0.00	0.01	0.27	0.35	0.22	0.53	0.01	0.16
Quadrature sum	0.03	0.12	0.29	1.20	0.30	1.73	0.11	0.92
Statistical uncertainty	0.03	0.08	0.28	0.87	0.46	1.68	0.72	2.28
PDG uncertainty	0.23	–	0.23	–	0.23	–	0.23	–

Table 7.6 A summary of all assigned systematic uncertainties.

7.1.7 Ξ_c^+ mass uncertainty

The latest fit from the Particle Data Group (PDG) for Ξ_c^+ was used [1]. The relevant uncertainty is 0.23 MeV, which is assigned to all mass measurements,

$$m_{PDG}[\Xi_c^+] = 2467.71 \pm 0.23 \text{ MeV}. \tag{7.8}$$

This uncertainty is introduced since the mass is calculated using the measured ΔM and the PDG measurement such that

$$m = \Delta M + m_{PDG}[\Xi_c^+], \tag{7.9}$$

where ΔM (Equation (4.14)) and $m_{PDG}[\Xi_c^+]$ are as previously defined.

7.1.8 Summary of uncertainties

The systematic uncertainties discussed in this section and their individual contributions to the total systematic uncertainty per measured parameter are summarised in Table 7.6.

For the $\Xi_c(2815)^+$ state, the uncertainty arising from the use of the PDG mass (Section 7.1.7) dominates the mass measurement, while for the width measurement the systematic uncertainty dominates, originating from MC-Data disagreement (Section 7.1.3).

For $\Xi_c(2923)^+$ state, the mass measurement is affected the most by the systematic uncertainty contribution from the alternative interference model (Section 7.1.6). The width is also dominated by systematic uncertainty arising from the varied

parameters of the RBW function (Table 7.1).

The $\Xi_c(2970)^+$ mass measurement is dominated by statistical uncertainty, but also has a sizeable systematic contribution from the interference model (Section 7.1.6). The width is affected considerably by the systematic contribution from the alternative background model (Section 7.1.2).

Finally, the largest uncertainties pertaining to the $\Xi_c(3080)^+$ mass and width measurements are primarily statistical. The width however does have a sizeable systematic contribution from the alternative background (Section 7.1.2).

7.2 Cross-checks

A variety of additional checks were carried out by fitting subsets of the data and inspecting the change in measured parameters. Some of these studies uncovered variations that were significant enough to be assigned as systematic error and were covered in the previous section. However, the rest of these studies that do not contribute to systematic uncertainties are summarised in this section for completeness.

7.2.1 Fits to separate charges

The fit was repeated using charge-split samples, as shown in Figure 7.5. It is found that the change with respect to the these fits (not including extended PDF N parameters) are consistent with zero for all parameters (see Table 7.7).

State	Parameter	Value
Background	Δm_0	1.7 ± 1.2
	Δk	-0.07 ± 0.1
	Δc	-0.0003 ± 0.0004
$\Xi_c(2815)^+$	Δm	$+0.08 \pm 0.07$ MeV
	$\Delta\Gamma$	-0.05 ± 0.17 MeV
$\Xi_c(2923)^+$	Δm	-1.1 ± 0.5 MeV
	$\Delta\Gamma$	-0.2 ± 1.7 MeV
$\Xi_c(2970)^+$	Δm	-0.7 ± 0.9 MeV
	$\Delta\Gamma$	-0.7 ± 3.4 MeV
$\Xi_c(3080)^+$	Δm	$+2.0 \pm 1.2$ MeV
	$\Delta\Gamma$	$+5 \pm 4$ MeV
Total	χ^2/N_{dof}	13.7 / 11

Table 7.7 Table of parameter difference (and uncertainty) between fits to Ξ_c^{**-} and Ξ_c^{**+} candidate samples. The signal masses and widths are consistent for all states between the charge-split sub-samples.

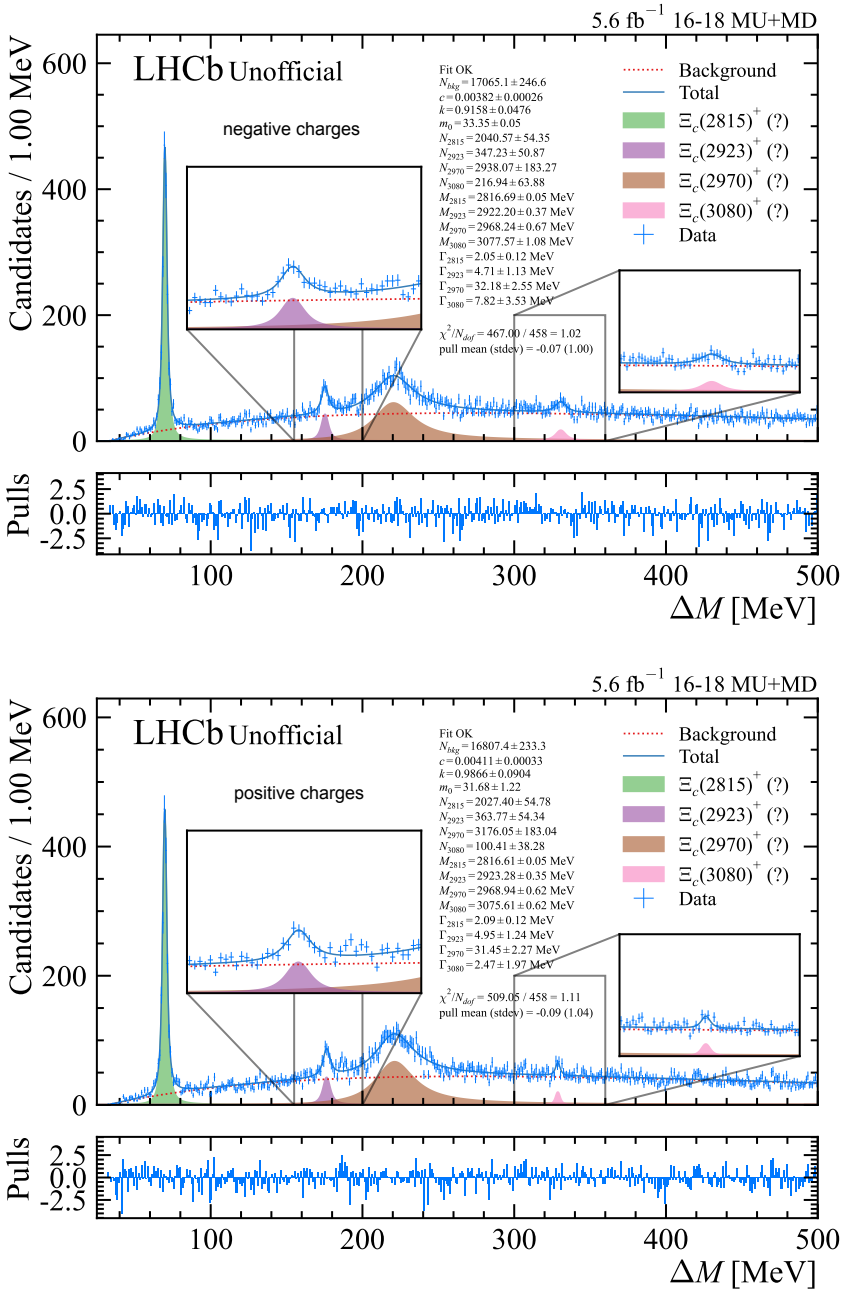

Figure 7.5 Running the fit separately on negatively (Ξ_c^{**-}) and positively (Ξ_c^{**+}) charged candidates.

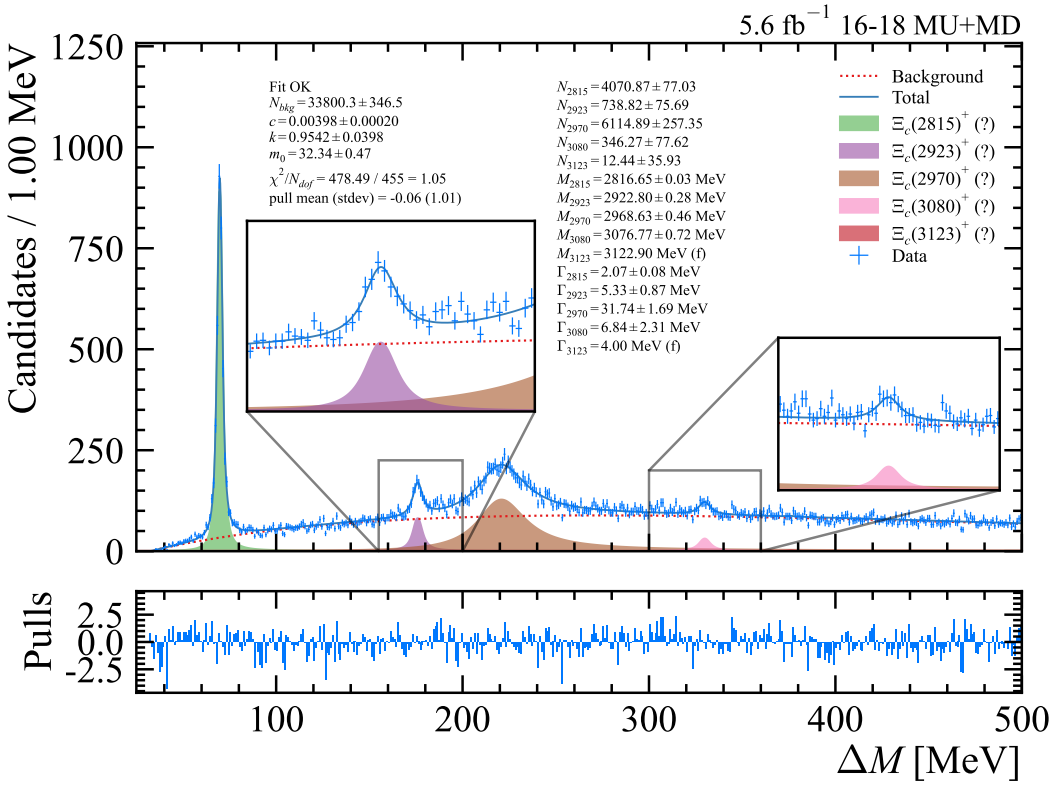

Figure 7.6 A repeat of the fit including the $\Xi_c(3123)^+$ resonance additionally considered. However, it is shown to have a very small yield of 12.44. The mass and width is fixed to PDG values.

7.2.2 Consideration of additional resonances in the signal model

Additional Ξ_c^{**+} resonances such as those listed in the PDG can be tested by adding them individually into the signal model and studying how the result of the fit is affected. Since the yields are expected to be small, some parameters such as the mass or width can be fixed to sensible values such as those measured by the PDG or in previous analyses such as $\Xi_c^{**0} \to \Lambda_c^+ K^-$ [19]. The results of the fits are displayed in Figure 7.6 and 7.7. The overall change in mass and width parameters are summarised in Table 7.8.

The significance of the $\Xi_c(3055)^+$ is calculated from the negative log likelihood difference between the nominal fit and fit including the extra $\Xi_c(3055)^+$ resonance, with a calculated significance value of 2.55σ.

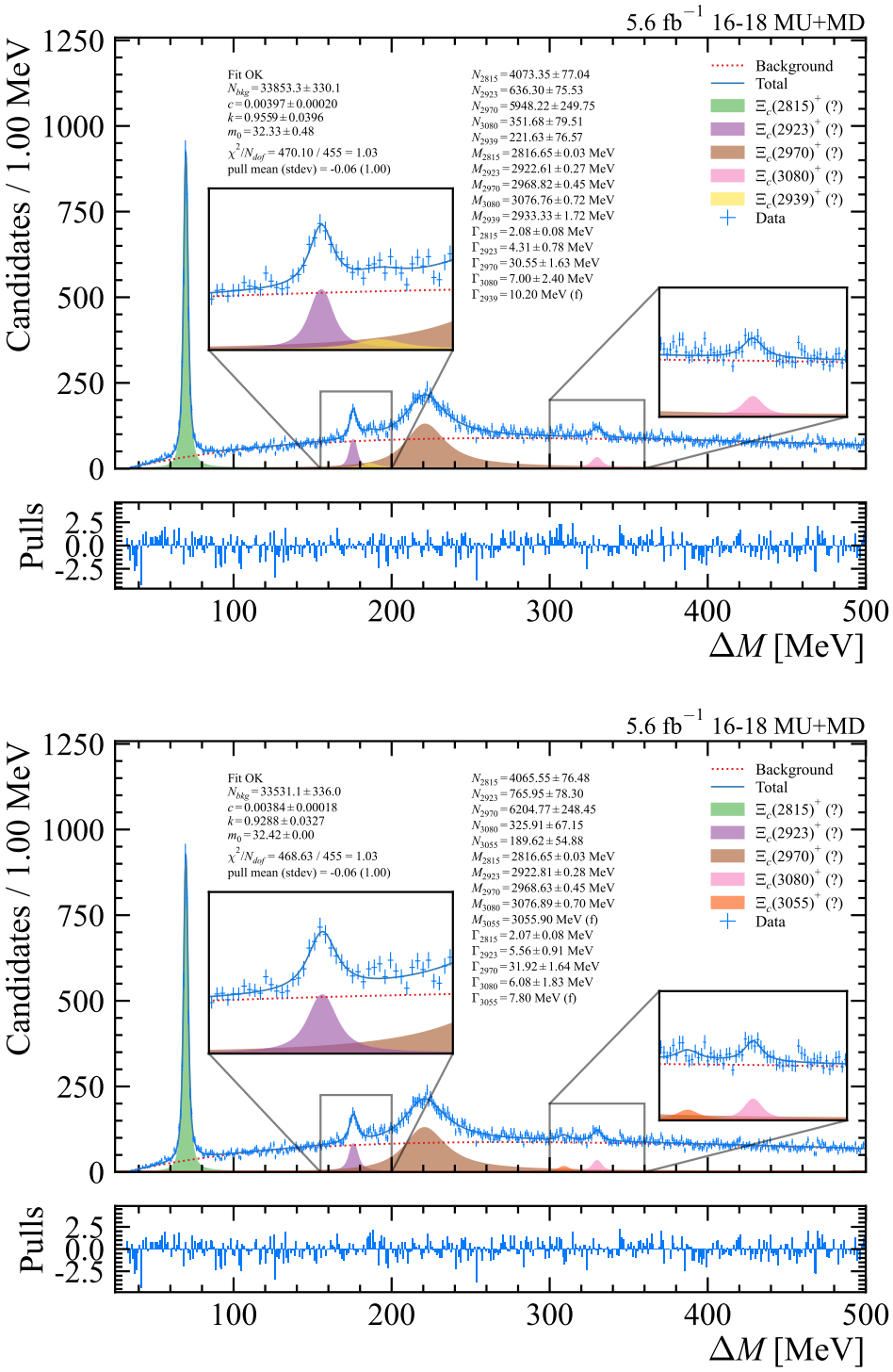

Figure 7.7 Fitting the mass spectrum after including resonances $\Xi_c(2939)^+$ (top) and $\Xi_c(3055)^+$ (bottom) in the signal model. Mass and width parameters of $\Xi_c(3055)^+$ were fixed to PDG values, with a floating yield. The $\Xi_c(2939)^+$ width parameter was fixed to known value of the $\Xi_c(2939)^0$ measured in Ref. [19], with floating mass and yield parameters.

Additional resonance	$\Xi_c(2815)^+$		$\Xi_c(2923)^+$		$\Xi_c(2970)^+$		$\Xi_c(3080)^+$	
	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)	Δm (MeV)	$\Delta\Gamma$ (MeV)
$\Xi_c(2939)^+$	-0.00009	0.00192	-0.18604	-1.00006	0.19771	-1.13722	-0.00764	0.21575
	(-0.003)	(+0.02)	(-0.67)	(-1.15)	(+0.43)	(-0.68)	(-0.01)	(+0.1)
$\Xi_c(3055)^+$	0.00051	-0.00719	0.00794	0.24950	0.00508	0.23507	0.12043	-0.69997
	(+0.02)	(-0.09)	(+0.03)	(+0.29)	(+0.01)	(+0.14)	(+0.17)	(-0.31)
$\Xi_c(3123)^+$	-0.00004	-0.00113	0.00026	0.02061	0.00496	0.05242	-0.00083	0.06043
	(-0.001)	(-0.01)	(0.001)	(+0.02)	(+0.01)	(+0.03)	(-0.001)	(+0.03)

Table 7.8 For each considered additional resonance: the change in mass and width parameters from the nominal fit results, with standard score below each shift (the change divided by the statistical error).

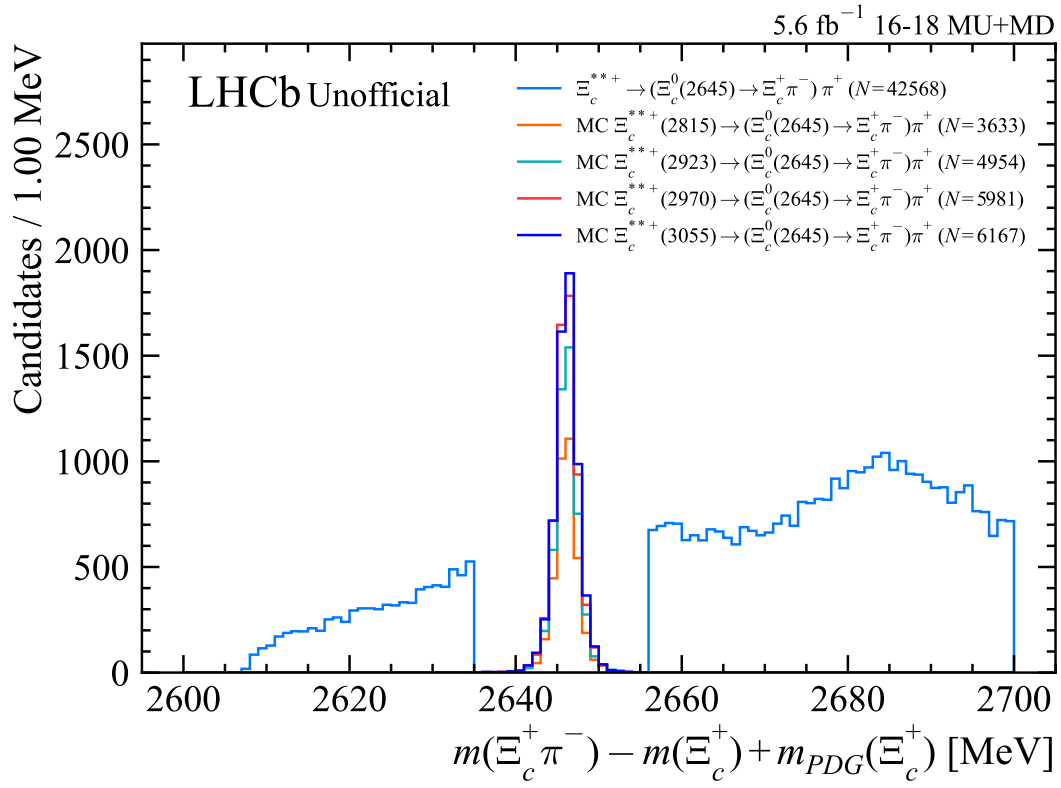

Figure 7.8 Nonresonant $\Xi_c(2645)^0$ "corrected" mass distribution. The light blue histogram shows the variable to which the requirement is applied. In this cross-check, the requirement is reversed such that Ξ_c^{**+} candidates consistent with decaying resonantly are removed. All other parts of the selection are unchanged. If one plots the MC without this requirement one can see the $\Xi_c(2645)^0$ resonance sits inside this window as shown by the simulated samples overlaid on the plot.

7.2.3 Peeking into the non-resonant mass spectrum

The selection currently targets resonant decays via $\Xi_c(2645)^0$, and the non-resonant region in the $m[\Xi_c(2645)^0] - m[\Xi_c^+]$ region is cut out in stage II-III via a mass window requirement targeting the $\Xi_c(2645)^0$ resonance. Additionally, the stage III MVA is optimised for resonant decays and may perform suboptimally in the non-resonant case. However, to explore the resonant region we can reverse the stage II-III $\Xi_c(2645)^0$ mass window requirement and explore the Ξ_c^{**+} ΔM distribution as illustrated in Figure 7.8. The ΔM distribution for non-resonant Ξ_c^{**+} decays resulting from the modified requirement is shown in Figure 7.9, where it can be seen that the distribution shows no indication of mass peaks, except for a small excess close to threshold. Otherwise, background largely dominates.

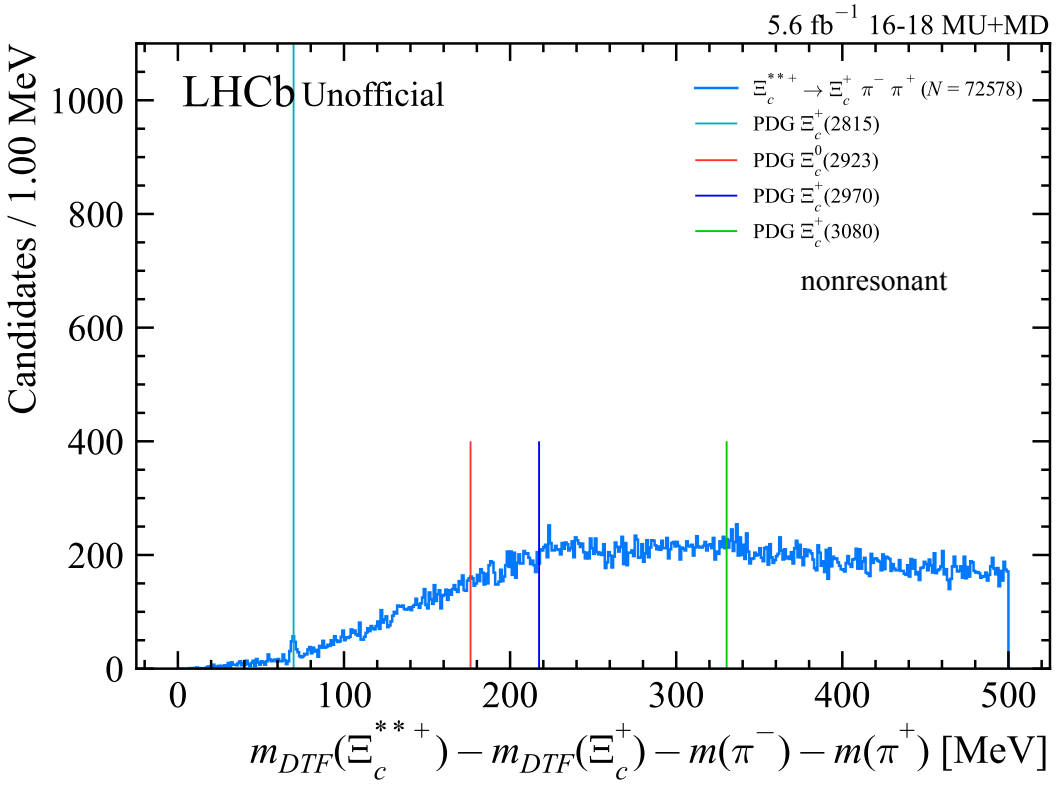

Figure 7.9 The non-resonant mass distribution is heavily dominated by background. Although some very small peaks are visible, it is difficult to draw any conclusions about them.

State	Parameter	2016	2017	2018
Background	Δm_0	-0.2014 (-0.33)	**fixed**	$+0.1209$ ($+0.08$)
	Δk	$+0.0385$ ($+1.50$)	$+0.0049$ ($+0.08$)	-0.0732 (-0.87)
	Δc	-0.0000 (-0.04)	$+0.0000$ ($+0.08$)	-0.0002 (-0.62)
$\Xi_c(2815)^+$	Δm	-0.0633 (-0.98)	$+0.0579$ ($+0.98$)	-0.0024 (-0.05)
	$\Delta\Gamma$	$+0.0720$ ($+0.47$)	$+0.1562$ ($+1.08$)	-0.2284 (-1.75)
$\Xi_c(2923)^+$	Δm	$+0.1492$ ($+0.28$)	-0.2912 (-0.73)	$+0.3783$ ($+0.59$)
	$\Delta\Gamma$	-0.5356 (-0.28)	-0.4853 (-0.45)	$+1.4603$ ($+0.66$)
$\Xi_c(2970)^+$	Δm	-1.0503 (-1.43)	$+0.5363$ ($+0.66$)	$+0.6156$ ($+0.76$)
	$\Delta\Gamma$	-4.0401 (-1.63)	$+2.1202$ ($+0.69$)	$+2.1771$ ($+0.72$)
$\Xi_c(3080)^+$	Δm	$+1.9679$ ($+1.54$)	-0.9981 (-1.09)	$+0.2142$ ($+0.19$)
	$\Delta\Gamma$	-1.4037 (-0.33)	-0.6755 (-0.16)	-2.0190 (-0.62)

Table 7.9 Comparison of fit results with nominal fit for samples split by data taking year. The parameter changes are tabulated alongside the standard deviation (in brackets).

State	Parameter	*MagUp*	*MagDown*
Background	Δm_0	$+0.8155$ ($+0.63$)	-0.1235 (-0.21)
	Δk	-0.0893 (-1.08)	$+0.0361$ ($+0.19$)
	Δc	-0.0004 (-0.95)	$+0.0001$ ($+0.61$)
$\Xi_c(2815)^+$	Δm	$+0.0455$ ($+0.96$)	-0.0441 (-0.94)
	$\Delta\Gamma$	-0.0183 (-0.15)	-0.0023 (-0.02)
$\Xi_c(2923)^+$	Δm	-0.0214 (-0.06)	-0.0020 (-0.00)
	$\Delta\Gamma$	-0.7226 (-0.62)	$+0.8321$ ($+0.64$)
$\Xi_c(2970)^+$	Δm	$+0.1954$ ($+0.27$)	-0.1062 (-0.18)
	$\Delta\Gamma$	$+4.8158$ ($+1.67$)	-3.6115 (-1.78)
$\Xi_c(3080)^+$	Δm	$+1.4093$ ($+0.79$)	-0.4412 (-0.48)
	$\Delta\Gamma$	$+8.3150$ ($+3.39$)	-1.3487 (-0.69)

Table 7.10 Comparison of fit results with nominal fit for samples split by magnet polarity.

7.2.4 Fits to different data-taking years and magnet polarities

In LHCb the data-taking conditions can change each year, and additionally the reconstruction can also be affected by the polarity of the magnet. To cross-check that no such effects are relevant to these measurements, the samples are split by data-taking year and magnet polarity and fit separately. The shifts from the nominal values are studied in Tables 7.9 and 7.10. Any such effects if present are negligible.

State	Parameter	Shift (std. dev)
Background	Δm_0	$+0.6223$ $(+0.62)$
	Δk	$+0.0105$ $(+0.17)$
	Δc	-0.0000 (-0.07)
$\Xi_c(2815)^+$	Δm	$+0.0162$ $(+0.46)$
	$\Delta\Gamma$	$+0.0311$ $(+0.35)$
$\Xi_c(2923)^+$	Δm	-0.0069 (-0.03)
	$\Delta\Gamma$	-0.4299 (-0.52)
$\Xi_c(2970)^+$	Δm	$+0.0766$ $(+0.16)$
	$\Delta\Gamma$	$+1.7458$ $(+0.91)$
$\Xi_c(3080)^+$	Δm	$+0.1529$ $(+0.19)$
	$\Delta\Gamma$	$+1.2459$ $(+0.42)$

Table 7.11 Comparison of fit results with nominal fit for samples split by multiple candidate handling strategy. The parameter changes as a result of this strategy are tabulated alongside the standard deviation (in brackets).

7.2.5 Multiple candidate strategy effect on measurements

A simple approach to handling multiple candidates was considered, where for each event with more than candidate, one candidate is picked at random and the unpicked candidates discarded. See table of results at Table 7.11. The effect is concluded to be minimal.

Chapter 8

Interpretation and Conclusions

We observe four structures each of which may correspond to actual Ξ_c^{**+} excited states. From the final mass fit to ΔM (the Q-value), it was hypothesised that these resonances from smaller to larger mass (Figure 6.3) are $\Xi_c(2815)^+$, $\Xi_c(2923)^+$, $\Xi_c(2970)^+$, and $\Xi_c(3080)^+$. All resonances are observed with significances greater than 5σ. The results of the fit yield measurements of the masses and widths of four excited states,

$$
\begin{aligned}
m[\Xi_c(2815)^+] &= 2816.65 \pm 0.03\,(\text{stat}) \pm 0.03\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(2815)^+] &= 2.07 \pm 0.08\,(\text{stat}) \pm 0.12\,(\text{syst})\,\text{MeV},\\
m[\Xi_c(2923)^+] &= 2922.80 \pm 0.28\,(\text{stat}) \pm 0.29\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(2923)^+] &= 5.3 \pm 0.9\,(\text{stat}) \pm 1.2\,(\text{syst})\,\text{MeV},\\
m[\Xi_c(2970)^+] &= 2968.62 \pm 0.46\,(\text{stat}) \pm 0.30\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(2970)^+] &= 31.7 \pm 1.7\,(\text{stat}) \pm 1.7\,(\text{syst})\,\text{MeV},\\
m[\Xi_c(3080)^+] &= 3076.77 \pm 0.72\,(\text{stat}) \pm 0.11\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV},\\
\Gamma[\Xi_c(3080)^+] &= 6.8 \pm 2.3\,(\text{stat}) \pm 0.9\,(\text{syst})\,\text{MeV},
\end{aligned}
$$

where the $\Xi_c(2923)^+$ state is observed for the first time in any channel. The $\Xi_c(3080)^+$ state, although previously observed in other channels by Belle [77], is seen for the first time in this particular channel.

For comparison with our measurement of $\Xi_c(2923)^+$, the $\Xi_c(2923)^0$ observed by

LHCb in $\Lambda_c^+ K^-$, was measured with mass and width

$$
\begin{aligned}
m[\Xi_c(2923)^0] = &\quad 2923.04 \pm 0.25\,(\text{stat}) \pm 0.20\,(\text{syst}) \pm 0.14\,(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2923)^0] = &\quad 7.1 \pm 0.8\,(\text{stat}) \pm 1.8\,(\text{syst})\,\text{MeV},
\end{aligned}
$$

as measured in Ref. [19], which is fully consistent with the $\Xi_c(2923)^+$ mass as measured here. This is a positive indication that the two states are isospin partners.

The states $\Xi_c(2815)^+$ and $\Xi_c(2970)^+$ are well known, both initially observed and studied by CLEO and Belle [22, 77, 78]. Our sample of $\Xi_c(2815)^+$ and $\Xi_c(2970)^+$ candidates is considerably larger than those of previous measurements (ranging between 10–10^3 events), consisting of 4080±77 and 5905±245 candidates respectively. Table 8.1 is an updated version of Table 1.5 with our measurements side-by-side. Figures 8.1, 8.2, and 8.3 also show comparisons with measurements from experiments Belle and BaBar. The measurements for $\Xi_c(2970)^+$ appear to differ (though not considerably) from those reported in the PDG, and our measured value for $\Xi_c(3080)^+$ width is evidently limited by statistics. Our

State	Mass (MeV)	Width (MeV)	J^P	Observed?	Mode $\Xi_c(2645)^0\pi^+$
$\Xi_c(2815)^+$	(PDG) 2816.51 ± 0.25 (ours) $2816.65 \pm 0.03 \pm 0.03 \pm 0.23$	(PDG) 2.43 ± 0.26 (ours) $2.07 \pm 0.08 \pm 0.12$	$?^?$	Yes	Seen
$\boldsymbol{\Xi_c(2923)^+}$	(PDG) – (ours) $2922.80 \pm 0.28 \pm 0.29 \pm 0.23$	(PDG) – (ours) $5.3 \pm 0.9 \pm 1.2$	$?^?$	**Now yes**	**Now seen**
$\Xi_c(2970)^+$	(PDG) 2964.3 ± 1.5 (ours) $2968.62 \pm 0.46 \pm 0.30 \pm 0.23$	(PDG) $20.9^{+2.4}_{-3.5}$ (ours) $31.7 \pm 1.7 \pm 1.6$	$1/2^+$ [23]	Yes	Seen
$\Xi_c(3080)^+$	(PDG) 3077.2 ± 0.4 (ours) $3076.77 \pm 0.72 \pm 0.11 \pm 0.23$	(PDG) 3.6 ± 1.1 (ours) $6.8 \pm 2.3 \pm 0.9$	$?^?$	Yes	**Now seen**

Table 8.1 Status of experimental measurements from the PDG [1] (or sources otherwise cited) updated with our measurements.

measurement of the $\Xi_c(2970)^+$ width shown in Figure 8.2, although quite far from the PDG average, is consistent with the most recent Belle measurement [24].

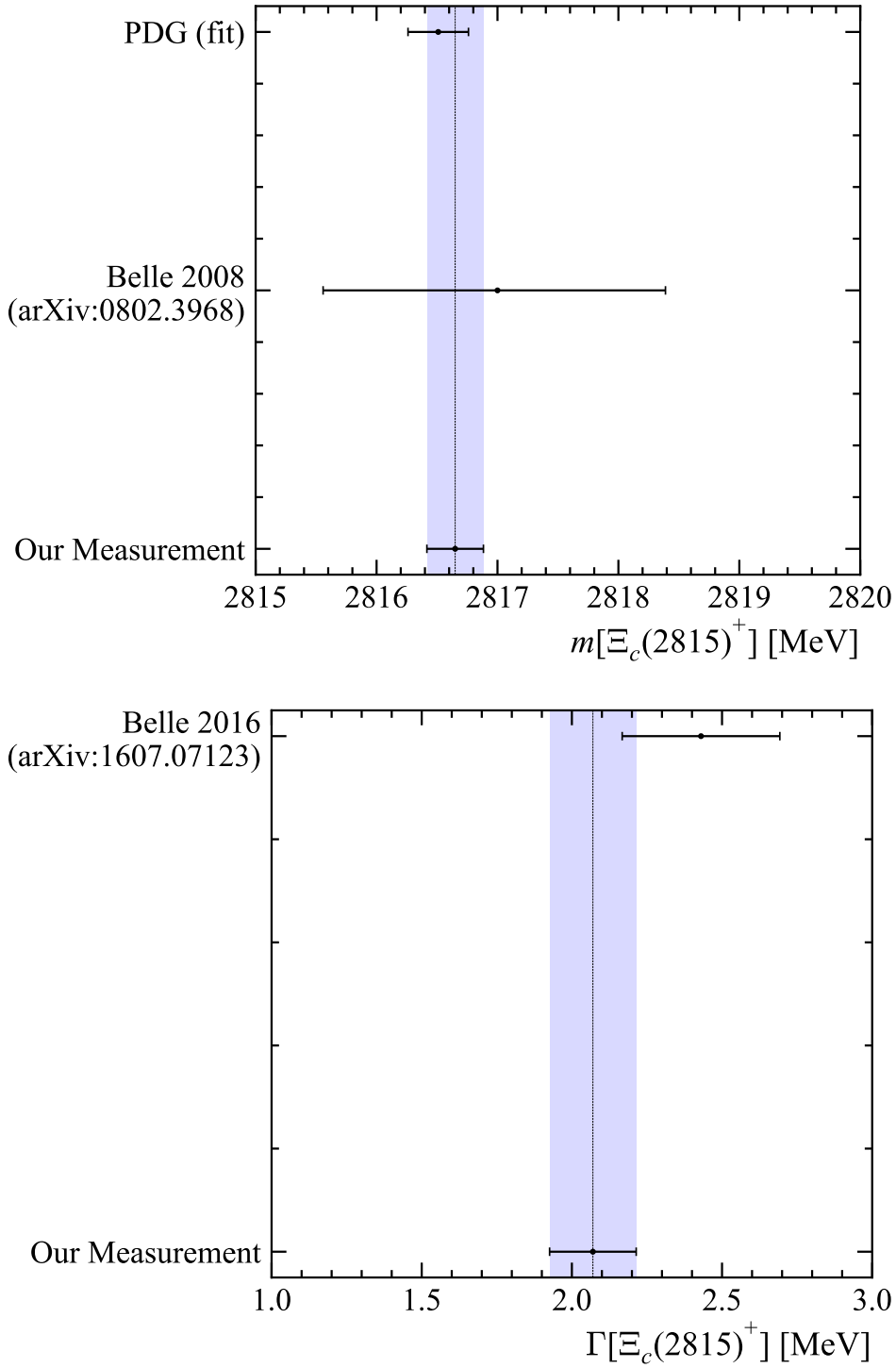

Figure 8.1 Measurement comparison plots for mass and width of $\Xi_c(2815)^+$. The blue shaded area represents our measurement.

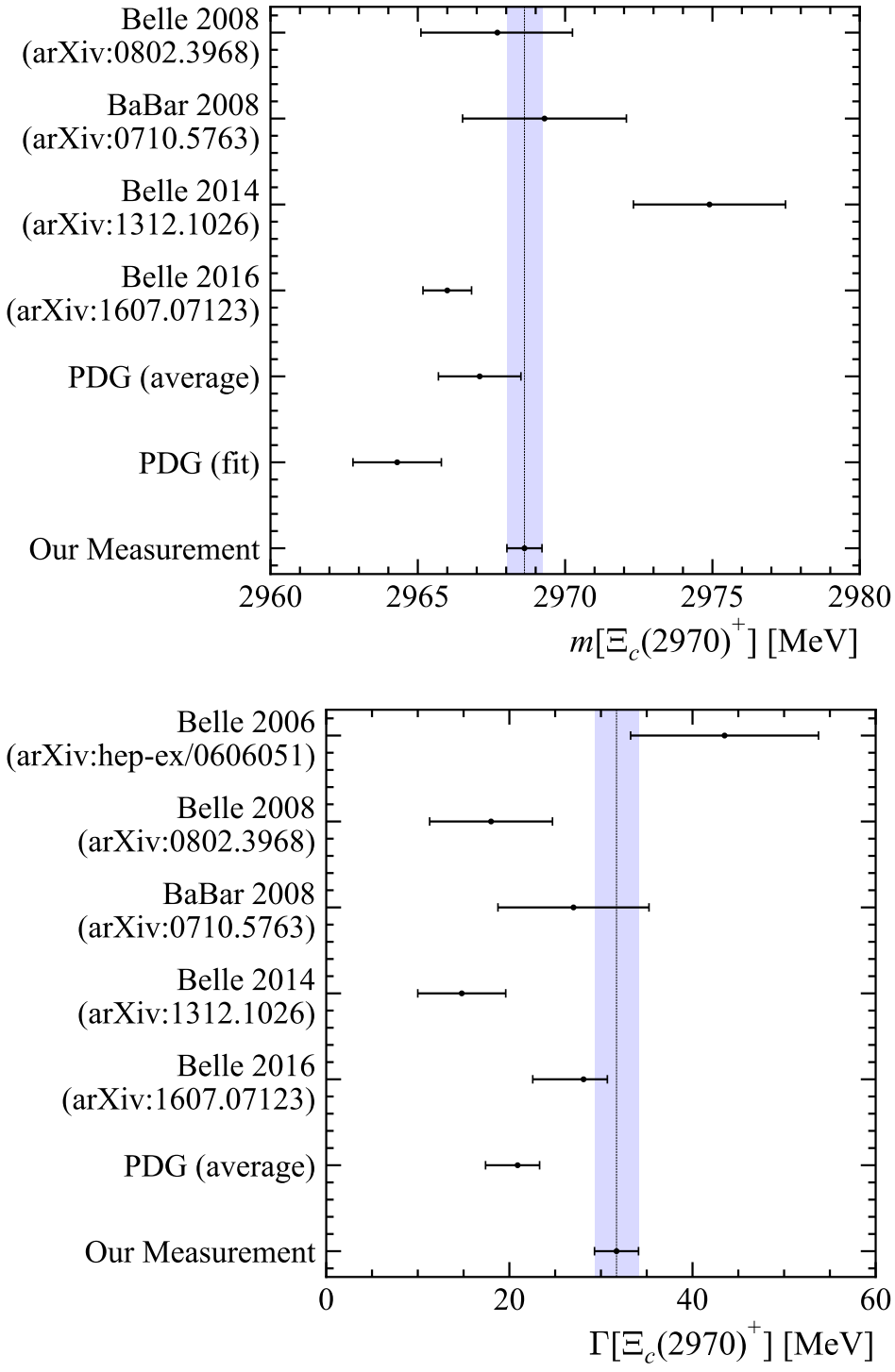

Figure 8.2 Measurement comparison plots for mass and width of $\Xi_c(2970)^+$.

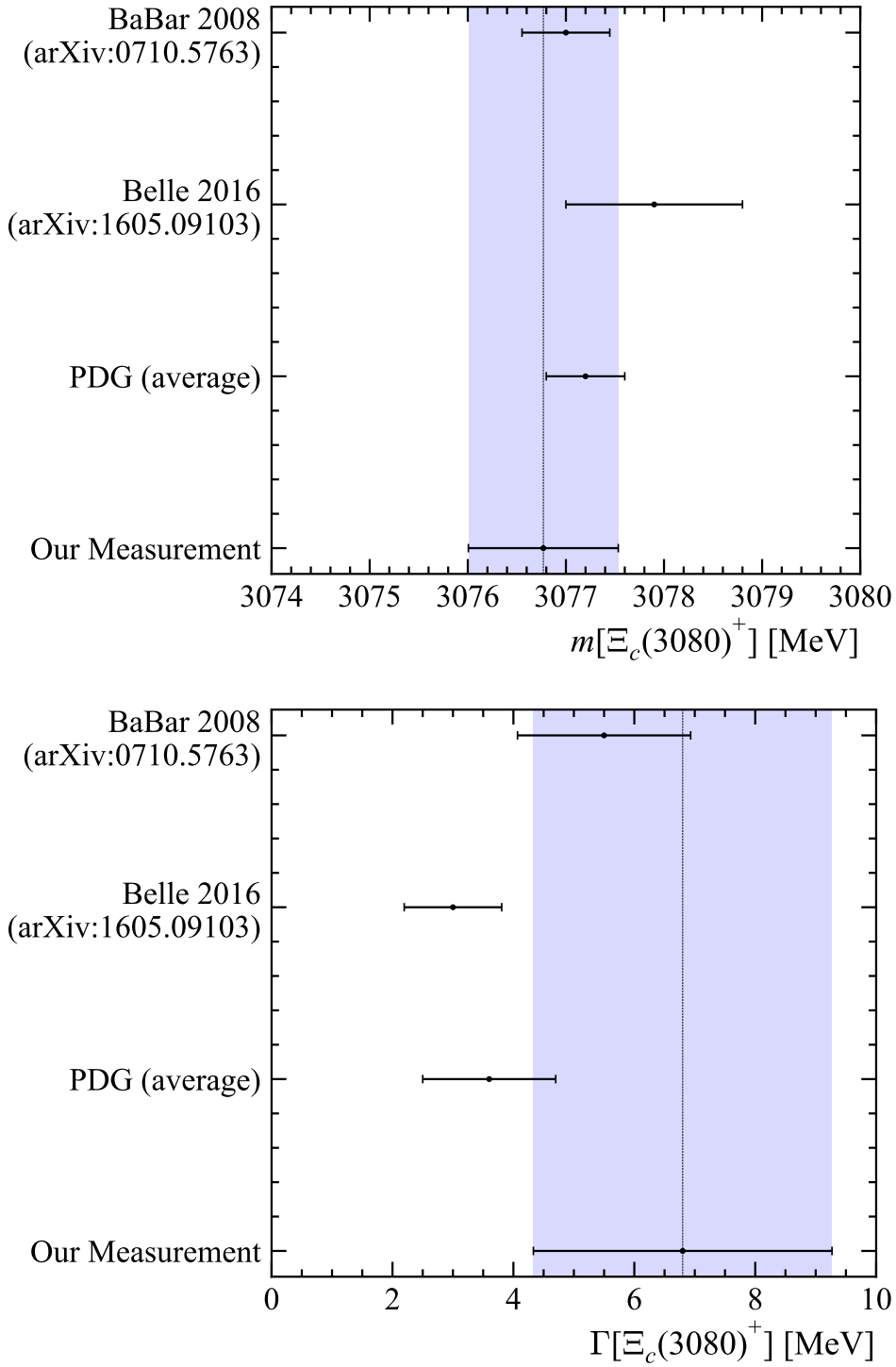

Figure 8.3 Measurement comparison plots for mass and width of $\Xi_c(3080)^+$.

Part III

Detector R&D and service work

Chapter 9

Testbeam studies of the irradiated ATLASPix3.1

The MightyPix R&D project will develop and test HV-MAPS technology for use in the LHCb Mighty Tracker detector of LHCb Upgrade-II (Section 2.4). This upgrade is necessary in order to resolve significantly higher interactions per bunch crossing of the HL-LHC that would lead to high detector occupancy, unmanageable ghost rates in tracking, and frequent replacements of active detector material due to low radiation hardness. These issues would present in the case that the existing Upgrade-I SciFi technology is not adapted. A new subdetector must meet this challenge of operating in harsher conditions when the HL-LHC starts running from 2030 onwards, where experiments will see an order of magnitude increase in instantaneous luminosity.

Due to the high radiation fluence expected to be accumulated over the expected HL-LHC data-taking period in the "inner" region close to the LHC beam pipe, the higher radiation-tolerance of monolithic silicon HV-CMOS technologies available in the present day are preferred to the existing SciFi technology currently utilised by the SciFi detector (Upgrade-I). As discussed in Section 2.4.1, a fully SciFi based Mighty Tracker would require frequent replacements of the active detector material in this inner region, and the high occupancy would still lead to a problematic ghost rate when reconstructing charged particles.

As such, a hybrid approach consisting of HV-MAPS active region close to the beamline and well-understood scintillating fibre technology in the outer regions is considered the most cost-balanced and functional approach for the Mighty

Property	Requirement
Sensor size	$2\,\text{cm} \times 2\,\text{cm}$
Sensor thickness	$\leq 200\,\mu\text{m}$
Pixel size	$\leq 100\,\mu\text{m} \times 300\,\mu\text{m}$
Timing resolution	$\approx 3\,\text{ns}$ within 25 ns window
In-time efficiency	$> 99\%$ within 25 ns window
Radiation tolerance	$3 \times 10^{14}\,1\text{MeVn}_{\text{eq}}/\text{cm}^2$
Noise limit	5 Hz/pixel
Power consumption	< 150 mW cm^{-2}
Data word length	32 bits
Compatible with LHCb Readout system	

Table 9.1 MightyPix requirements from Ref. [79] and Table 23 of Ref. [42].

Pixel Size	$50\,\mu\text{m} \times 150\,\mu\text{m}$
Pixel Matrix	372 rows × 132 cols
Chip Size	$20.2\,\text{mm} \times 21\,\text{mm}$

Table 9.2 ATLASPix3.1: Table of specifications.

Tracker design.

As MightyPix1 prototype chips were not available to test in 2022, and accounting for planning constraints for the mechanical and electrical designs of the Mighty Tracker, as well as the need to understand the radiation tolerance of the MightyPix, an alternative device was needed to act as proxy for the studies required. The ATLASPix3.1 was chosen on the basis of being a full-size HV-CMOS chip and predecessor to the MightyPix series of chips.

The measurements leading to the results in this chapter have been performed at the Test Beam Facility at DESY Hamburg (Germany), a member of the Helmholtz Association (HGF).

9.1 The ATLASPix3.1

The ATLASPix3.1 is a full-size HV-MAPS chip fabricated by a TSI 180 nm process, the same HV-CMOS technology used to fabricate the MightyPix1. It is a resubmitted version of its precursor chip, the ATLASPix3.0 [80], with minor modifications to the design covered in Ref. [79]. Specifications are summarised in Table 9.2.

9.2 Testbeam motivations and setup

Besides tests in tabletop lab environments, sensor performance may be evaluated in a live particle beam environment, a "testbeam". This section will discuss the testbeam and analysis strategy followed in the testbeam campaigns of Section 9.3.

A testbeam generally aims to replicate an environment close to real data-taking and conditions of a particle physics experiment. It tends to involve the setup of a sensor, the Device Under Test (DUT), of which performance is to be evaluated, and some reference sensors arranged in a telescopic geometry around the DUT. Additionally, the setup includes all necessary surrounding equipment required for Data Acquisition (DAQ) and operation of the sensors to facilitate running in a range of conditions. The sensors are situated along a beam line of, for example, 4.8 GeV electrons which induce activity in the devices which are recorded for later study and analysis.

Typically, testbeam analysis studies tend to measure DUT characteristics such as

1. **Timing resolution:** how quickly does the sensor respond, and what are the variations in the timing response?

2. **Pixel/sensor efficiency:** do the pixels/sensors tend to miss incident particles, and/or do the noise levels affect reconstruction performance?

3. **Trends in performance:** How is performance affected when the sensors have been subject to different operating temperatures, settings, and radiation exposure (measured in 1 MeV $n_{\mathrm{eq}}/\mathrm{cm}^2$ units).

The efficiency of the sensors as a measure of performance is particularly important with regard to determining the optimal working conditions of the chip as e.g. irradiation level increases.

Finding the answers to the above questions can provide crucial input towards sensor design and the detector project in understanding whether related designs or prototypes meet requirements, and whether adjustments or mitigations are necessary going forward. Additionally, the campaigns help to build up crucial expertise, analysis experience, and analysis methods.

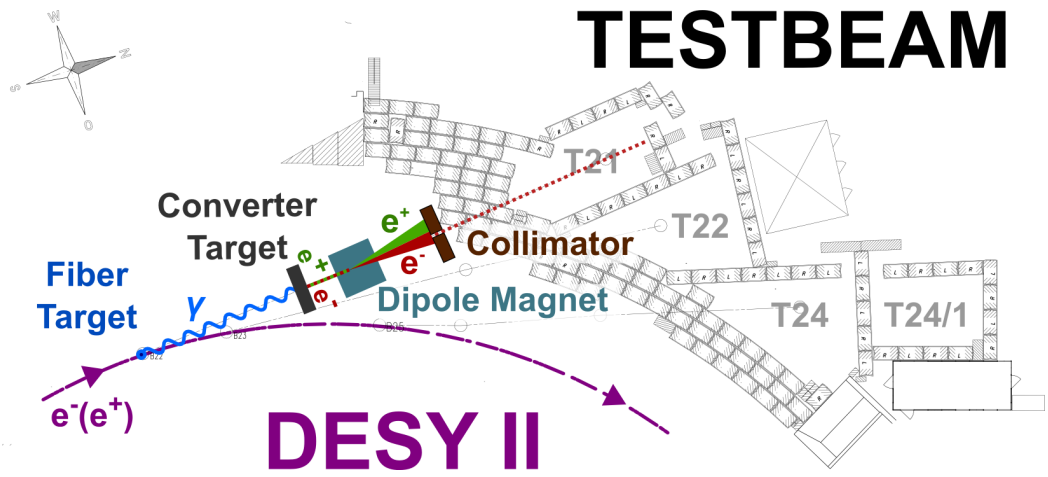

Figure 9.1 Diagram of DESY-II Testbeam Facility. An electron-positron beam is incident upon a fibre target moved into the beam line, which produces bremsstrahlung photons. These photons are incident upon a secondary target to produce electron positron pairs which are charge and momentum-selected before proceeding to the relevant beam area via a collimator. The testbeam setup and experiments are housed in the beam area. Reproduced from Ref. [84] with permission.

The rest of the chapter will discuss the work of two testbeam campaigns carried out in June and December 2022. The testbeam setup, operation, debugging, offline analysis, discussions and interpretations of results were carried out by a team of members of the LHCb Upgrade-II Mighty Tracker project from a range of institutes. Results have been presented at conferences [81], discussed in an LHCb internal note [79], and are or will be published [82,83]. This thesis chapter will focus primarily on the relevant analysis contributions made by the author throughout the campaign.

9.2.1 Setup and analysis principles

The campaigns discussed in Section 9.3 were both conducted at the DESY-II Testbeam Facility in Hamburg, Germany [84]. The facility layout is displayed in Figure 9.1. The ADENIUM telescope with EUDAQ2 [85, 86] was used in these campaigns, with the ATLASPix3.1 as the DUT. A multitude of sensors operated at different temperatures and irradiated to various levels were at hand and could be installed as the DUT.

The ADENIUM telescope [85] has six planes of ALPIDE [87] sensors (Table 9.3),

Figure 9.2 Adenium telescope and DUT set up in the DESY testbeam area (June).

Pixel Size	28 µm × 28 µm
Pixel Matrix	512 rows × 1024 cols
Chip Size	1.5 cm × 3.0 cm
Spatial Resolution	5 µm × 5 µm

Table 9.3 ALPIDE: Table of specifications [87].

developed for use by the ALICE collaboration. The DUT is connected to a MUDAQ readout system [88], configured as an untriggered readout, and is situated between the ALPIDE sensor planes. The area of the DUT covered by the telescope planes is smaller than the total active area of the DUT sensor. A photograph of the telescope and DUT mounted in the testbeam area is shown in Figure 9.2.

An AIDA Trigger Logic Unit (TLU) [89] synchronises both the readout of the DUT sensor and the telescope sensors by: providing a central clock signal to the sensors; a trigger signal to instruct the telescope sensors to read data; and busy signals to prevent reading prematurely before sensors are ready to take data again. At the beginning of a run, the TLU provides a reset signal to the MUDAQ system to ensure the start of both EUDAQ2 and MUDAQ runs are synchronised. Trigger scintillators situated just after the telescope drive a trigger signal which is then distributed by the TLU, providing also an event timestamp which is saved. Run control is handled by the EUDAQ2 software [90] in which the MUDAQ system is integrated. In this setup, a single 'run' of the DAQ system configures

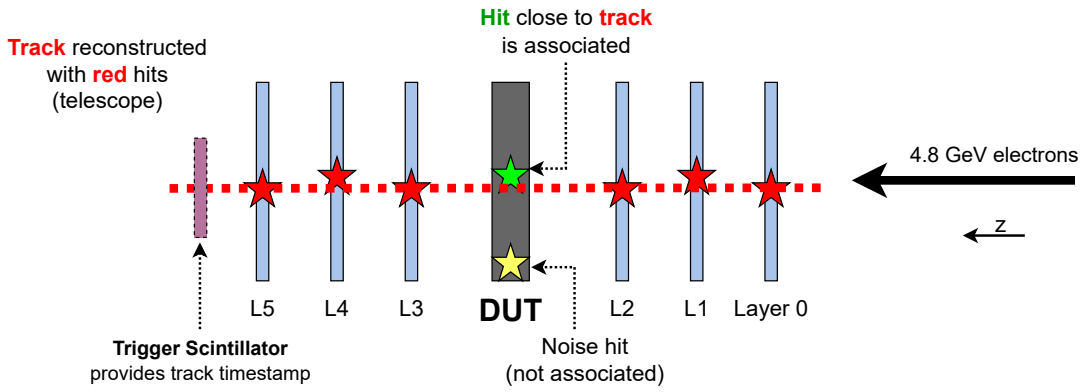

Figure 9.3 Demonstration of telescope setup, tracking, and association of DUT hits to a track. Reproduced from Ref. [81].

the sensors and produces two data files, one per readout system.

The configuration step at the beginning of the run will tend to involve varying of settings such as the `ThPix` (pixel threshold), varied in steps within a defined range to perform a scan. The pixel threshold is a set value above which a pixel registers a hit, usually expressed in Digital to Analogue Converter (DAC) units. DAC units can be converted to multiples of electron charges according to measured values in Table 5.1 of Ref. [82].

Files produced by each run from the DAQ systems are stored and used later for offline analysis using the Corryvreckan analysis framework [91]. Hits originating from the telescope sensors are organised into clusters and used to reconstruct tracks, as depicted in Figure 9.3. The DUT is not included in the tracking algorithm. Instead, DUT clusters are associated to reconstructed tracks according to a spatial cut. The fraction of tracks with associated hit clusters can be used to calculate overall sensor efficiency, pixel efficiency, among other useful sensor performance indicators. The total efficiency is defined as

$$\epsilon_{total} = \frac{N_{tracks,assoc}}{N_{tracks}}, \tag{9.1}$$

where $N_{tracks,assoc}$ is the number of tracks with an associated DUT hit, and N_{tracks} is the total number of reconstructed tracks in a given event sample.

Timing performance is additionally considered, as a sensor should be able to register hits quickly enough such that data is acquired on-time within the 25 ns window between LHC bunch crossings. Timing performance is studied by

histograms of the Time of Arrival (ToA) or the time residual,

$$t_{res} = t_{track} - t_{cluster}, \tag{9.2}$$

which is calculated from the track timestamp t_{track} (the trigger timestamp) and earliest hit timestamp within the DUT $t_{cluster}$. Another metric, known as the in-time efficiency, measures the proportion of the DUT clusters that arrived 'on-time' within a 25 ns window, which is calculated from the t_{res} histogram by taking the 25 ns wide integral centred on the peak position, and dividing by the total area of the histogram.

Corrections to pixel hit timestamps are required due to 'delay' and 'time walk' characteristics that must be calculated offline, as they are not accounted for in the data acquisition hardware or chip design. These corrections are described in Section 9.2.5.

While the ATLASPix3.1 is a well characterised chip, it is not expected to meet MightyPix requirements, due to choice of pixel electronics. It is generally expected that the MightyPix will have adequately improved design to meet requirements.

9.2.2 Event reconstruction software

The Corryvreckan [91] framework is an extendable open-source modular analysis package that is capable of analysing testbeam data from a variety of setups, making it an increasingly popular choice among the testbeam analysis community.

Data acquired from a testbeam campaign are analysed one run at a time by a single invocation of the application executable, according to a user-specified configuration file. Following application initialisation, each of the configured steps of the event processing are carried out by customisable modules, configured via user-written TOML files. In the event loop, and for each event, modules are called upon in the order that they appear in the configuration file, to process that event. Module configuration parameters can be set, allowing for fine control over e.g. spatial and timing cuts, features of the tracking, and other parameters that can improve reconstruction results. Each module creates plots and histograms which are saved in their own directory inside a single output ROOT file for later debugging and/or analysis.

In this testbeam setup, the reconstruction begins with an event loading from both DAQ output files for the current run where pixel hits are read and assigned to sensors organised in a geometry. An event is defined by the time bounds of the telescope event. Hits are clustered spatially and in time, followed by tracking, association of DUT hits to reconstructed tracks, and the calculation of a range of variables used in characterisation and performance studies.

For a successful reconstruction of a run, the run must be 'good': no crashes during data-taking, no configuration missteps, no data corruption, and all hardware functioning as intended.

9.2.3 Alignment procedure

An 'aligned' geometry file accurately describing the sensor planes, specifications, positions and rotations is needed to reconstruct the events of each run to high quality. A good alignment means the geometry is a faithful representation of the sensors as they were in their real-world positions and orientations. A combination of manual alignment by-hand, followed by track-based global alignment by total-χ^2 minimisation was used. Good alignment quality and functioning offline event reconstruction is not a dependency of online data taking, so data taking is able to continue in parallel, however finding a good alignment early in the campaign rather than after, is crucial for validating the successful operation of the full testbeam setup and the resolution of potentially show-stopping issues.

Measurements of z-separations between each sensor plane were taken for each testbeam setup, as alignment algorithms cannot easily be used to align sensor planes along the beam direction (z). This is due to the fact that the change in track parameters is typically insensitive to a change in the z-position of a sensor plane as the track x-y intersect is not sensitive to the sensor z-position, meaning the track-χ^2 is also insensitive to this degree of freedom. Therefore, introducing alignment degrees of freedom for z is not likely to affect the global χ^2 of the alignment track sample, and would destabilise the global alignment fit. In both testbeam campaigns, these positions were measured by hand and recorded in logbooks.

A Corryvreckan geometry file is constructed by hand. Initially, telescope planes are added with x, y positions set to zero while z is fixed to hand measured values. Rotations start at zero unless there are special considerations for the testbeam

setup. The alignment procedure outlined in the Corryvreckan manual [92] is followed until a full telescope alignment is obtained. The geometry is thereafter extended to include the DUT, and a good DUT alignment is obtained as per the procedure in the manual.

This alignment procedure is repeated for any run where the setup changed i.e. changed DUT sensor, an access was required which caused any planes to be nudged, or any sufficient increase in sensor misalignment over time requiring re-alignment. This tended to mean that the alignment procedure was carried out at the start of any threshold scan or change of DUT sensor. It was sufficient to use an automated alignment procedure as the first alignment at the beginning of data taking tended to resolve the largest misalignments from the base geometry, and further misalignments introduced tended to be small or only affect the DUT.

9.2.4 Wider analysis workflow

After successfully analysing single runs and obtaining aligned geometries covering all runs, the analysis can be expanded to cover all acquired testbeam data. The results obtained from the reconstruction and analysis modules may be merged for measurement runs conducted under identical experimental conditions and settings i.e. same temperature, device, threshold settings, et cetera.

An automated workflow was devised using the `snakemake` [59] application for both testbeam campaigns to run the Corryvreckan application over all runs, calculate timing corrections, and then group and merge the results where possible to ensure each measured datapoint is a combination of multiple repeats. Due to the sheer volume of data and bookkeeping required, automating the workflow and ensuring consistent treatment of all runs saved significant time on analysis bookkeeping and helped focus priorities on understanding DUT characteristics and achieving measurement goals.

9.2.5 Timing corrections

Due to delays in signal being registered according to factors such as a delay dependent on the metal layer of a registering pixel, and due to a phenomenon known as time-walk, ad-hoc corrections for each run were necessary in order to achieve the best time resolution.

The corrections were calculated in steps, in the subsections to follow. The `AnalysisTimingATLASPIX` module of the Corryvreckan software package was used to apply the calculated timing corrections to each pixel hit [91, 93].

Delay correction

A pixel hit generates a signal which travels along a metal line to the periphery of the chip, where the hit is registered. A delay in hit registration is encountered depending on the length of the metal lines travelled by the signal, presenting as a shift in the time residual distribution. The delay is defined as

$$t_{delay}(r) = \frac{1}{N} \sum_i t^r_{res,i}, \tag{9.3}$$

as the mean of the time residual distribution for pixel hits with a given row number r, where $t^r_{res,i}$ is the i^{th} t_{res} value in the sub sample of N pixel hits under row number r. The distance travelled by the signal depends on the pixel row number and metal layer. There are three metal layers corresponding to the row number ranges

$$\begin{aligned} \text{Layer M5: } 0 \leq \text{row} \leq 124, \\ \text{Layer M4: } 125 \leq \text{row} \leq 247, \\ \text{Layer MT: } 248 \leq \text{row} \leq 371. \end{aligned}$$

Plotting the delay against row number yields Figure 9.4 where an overall offset, and linear dependences on the row number can be observed. A correction will recentre the delay for each row about zero, by subtracting from a pixel hit timestamp the relevant correction value for the pixel row based on some empirically determined linear fit model.

For each run, correction of these delays are carried out by first determining a time offset from the time correlations, such that the *time correlation* distribution is centred on zero after applying the time offset. Secondly, a linear fit to the delay

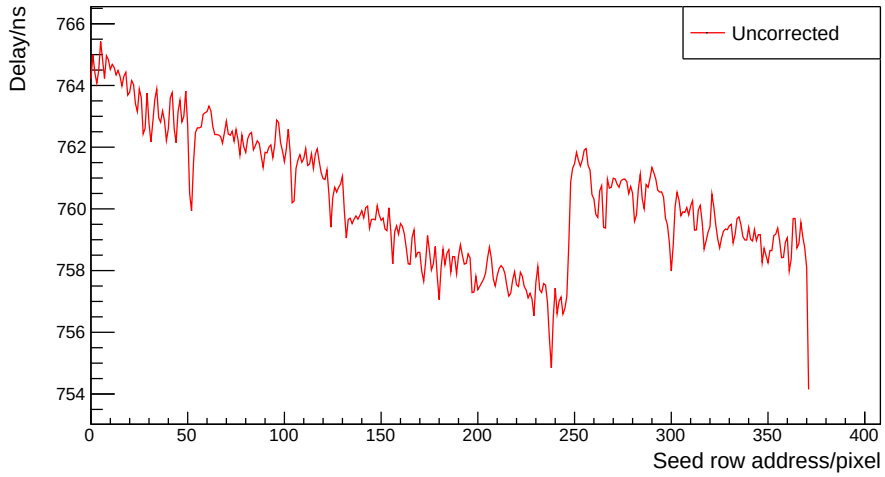

Figure 9.4 Pixel hit delay vs. row number. There is an overall offset from zero and a linear dependence in the three row number ranges corresponding to the distinct metal layers. Reproduced from Ref. [79].

time vs. row distribution is carried out in each metal layer row range such that

$$t_{delay}^{fit}(r) = \begin{cases} a_{\mathrm{M5}} r + b_{\mathrm{M5}} & \text{if } 0 \leq r \leq 124, \\ a_{\mathrm{M4}} r + b_{\mathrm{M4}} & \text{if } 125 \leq r \leq 247, \\ a_{\mathrm{MT}} r + b_{\mathrm{MT}} & \text{if } 248 \leq r \leq 371, \end{cases} \tag{9.4}$$

where a_{M5}, b_{M5}, a_{M4}, b_{M4}, a_{MT}, b_{MT}, are fit parameters to be determined for the corresponding ranges.

The resulting $t_{delay}^{fit}(r)$ values are calculated for each row number after fitting and obtaining the a and b parameters for each metal layer, and are saved and used in subsequent reconstruction to correct the delay effects. The time residual distribution itself is corrected by subtracting the overall time offset and relevant delay value for the row number of each pixel hit such that

$$t_{res}^{corr} = t_{track} - t_{hit} - t_{offset} - t_{delay}^{fit}(r), \tag{9.5}$$

and considering the hit timestamp only

$$t_{hit}^{corr} = t_{hit} + t_{offset} + t_{delay}^{fit}(r), \tag{9.6}$$

which is the correction applied to each pixel timestamp in a row r.

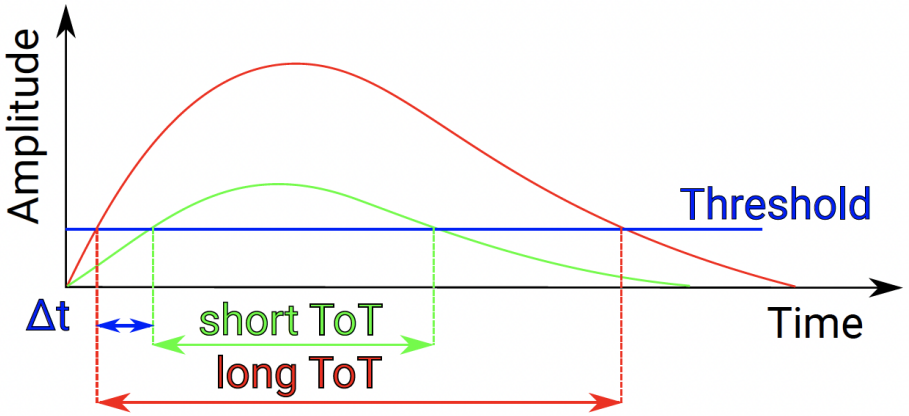

Figure 9.5 Depiction of time walk effects when considering larger vs. smaller signal shapes. Reproduced from Ref. [94].

Time walk correction

Time walk refers to the extra time taken by smaller signals to reach the comparator threshold, in addition to spending a low amount of Time over Threshold (ToT). This is in contrast to larger signals, which reach the threshold faster and have a higher ToT. This effect causes a delay in hit registration which increases with lower charges. Figure 9.5 demonstrates how this delay arises from the signal shapes and Figure 9.6 shows the correlation between the raw ToT value (y-axis) and the time residual (x-axis).

A time walk correction value is calculated for each ToT bin by fitting the time resolution variable in a ±250 ns slice centred on the bin and shifting the distribution such that the peak is at zero by using the obtained fit parameters. A Gaussian fit would be typically used to determine the correction value for each slice, however owing to asymmetries in the tails in some cases and fit instabilities the central value of the maximum sized bin is chosen as the more robust method.

After calculating correction values for each bin, as with the delay correction, the result is saved in a ROOT file and used in subsequent Corryvreckan reconstruction jobs to correct the pixel timestamps, where the correction value for the corresponding ToT bin is used. An example of the resulting ToT vs time residual histogram is shown in Figure 9.7, where the peaks in t_{res} slices are now centred on zero. Some hits are still delayed after this correction. However, this represents an overall small fraction of the total sample of pixel hits.

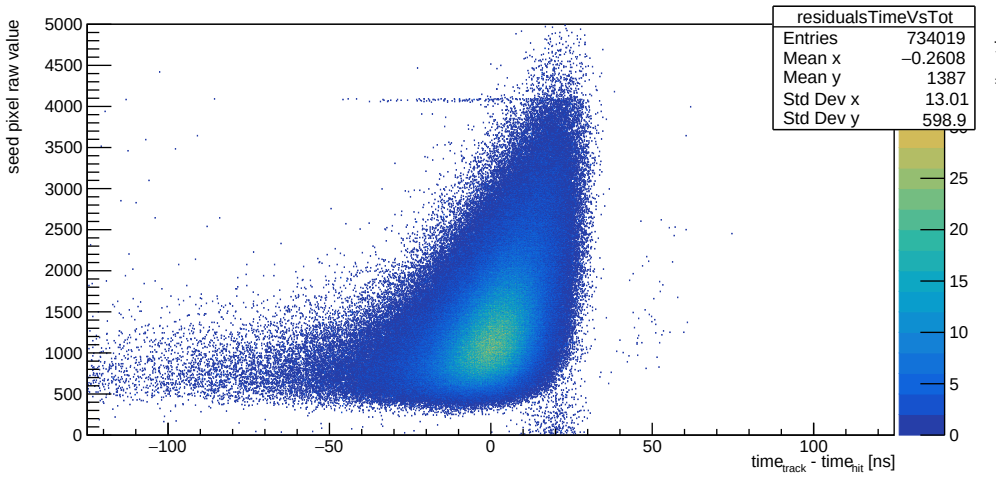

Figure 9.6 Two-dimensional histogram of ToT (y-axis) versus time residual without time walk correction (June 2022).

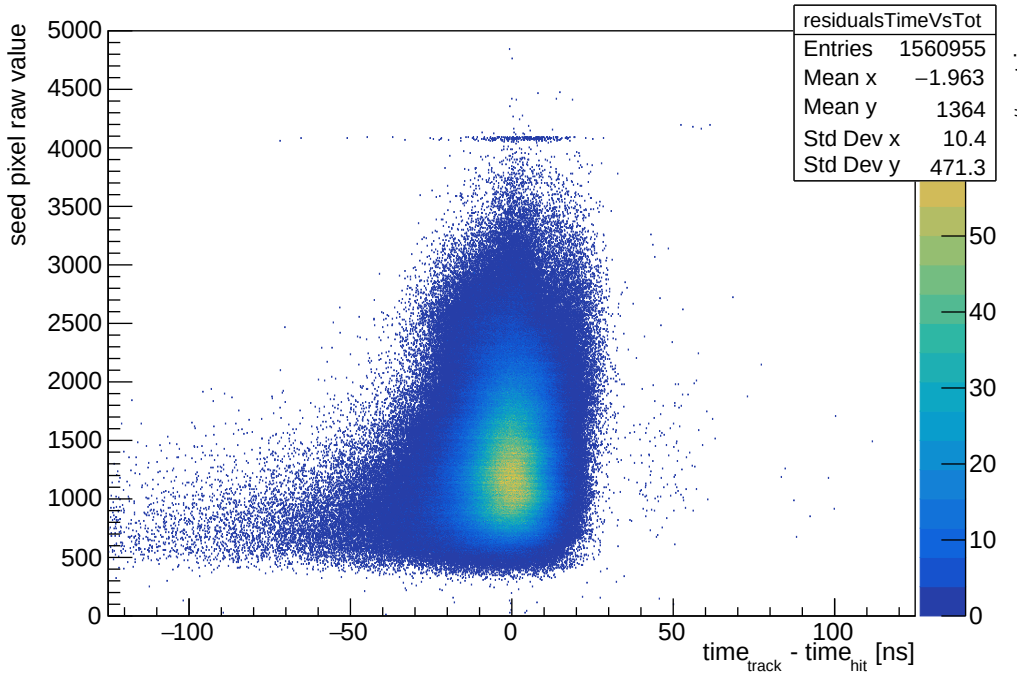

Figure 9.7 ToT (y-axis) versus time residual **after** time walk correction (June 2022).

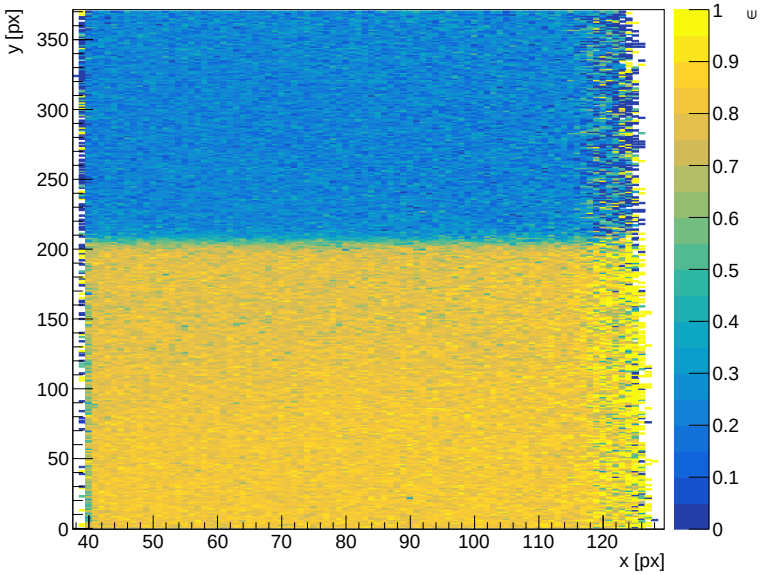

Figure 9.8 June 2022: Inefficiency covering approximately half of the DUT, caused by a misaligned trigger scintillator. Tracks intersecting the blue area on the DUT do not intersect the out-of-place scintillator downstream, causing the telescope not to trigger.

Once the reconstruction is reconfigured to run with both delay and time-walk corrections, the analysis may proceed.

9.3 Testbeam results and discussion

The June 2022 testbeam was carried out with the aforementioned setup in the T22 area (Figure 9.1) with an electron beam energy of 4.8 GeV, collecting data successfully for the following DUT sensors in Table 9.4. This testbeam managed to successfully cover measurements at all operating temperatures for the unirradiated ATLASPix3.1, and the 1×10^{14} $1\text{MeVn}_{\text{eq}}/\text{cm}^2$ irradiated sensor. The 3×10^{14} $1\text{MeVn}_{\text{eq}}/\text{cm}^2$ sensor was additionally measured, however collected statistics were comparatively low due to some issues discovered in the position of the trigger scintillator. The misalignment of the scintillator caused tracks intersecting approximately half of the active DUT area not to trigger the readout of the telescope. The primary indicator of this effect was the plot displayed in Figure 9.8, which was produced in near-real time and was instrumental in promptly identifying and resolving this issue, which would have affected all data taken during the campaign.

Additionally, some bugs in the DAQ hardware which were addressed in the December Testbeam as described in Ref. [82] were causing runs to be cut short in the June testbeam data, which additionally caused reductions in collected statistics as the runs affected by this problem were excluded altogether from the analysis.

The measurements covered in the following subsections will focus on total efficiency measurements (Section 9.3.1), time resolution (Section 9.3.2), and in-time efficiency (Section 9.3.3).

Sensor	**Irradiation Level**	**Beam Time**		
Type	($1\text{MeVn}_{\text{eq}}/\text{cm}^2$)	$-10\,^{\circ}\text{C}$	$0\,^{\circ}\text{C}$	$5\,^{\circ}\text{C}$
ATLASPix3.1	None	1.7 hr	1.3 hr	2.4 hr
ATLASPix3.1	1×10^{14}	1.6 hr	2.3 hr	–
ATLASPix3.1	3×10^{14}	52 min	5 min	47 min
ATLASPix3.1	1×10^{15}	–	–	–
ATLASPix3.1	3×10^{15}	–	–	–
ATLASPix3.0	None	–	–	–

Table 9.4 June 2022: Total run time in beam for each DUT device and operating temperature where data was actively being collected and track reconstruction was possible, as a measure of collected statistics and time in active beam. The beam was operated at a rate of 10 kHz (approx). The sensors with no operation time were disregarded due to misconfigured settings.

Sensor	**Irradiation Level**	**Beam Time**		
Type	($1\text{MeVn}_{\text{eq}}/\text{cm}^2$)	$-10\,^{\circ}\text{C}$	$0\,^{\circ}\text{C}$	$5\,^{\circ}\text{C}$
ATLASPix3.1	None	2.8 hr	1.5 hr	5.6 hr
ATLASPix3.1	1×10^{14}	1.8 hr	3.9 hr	1.1 hr
ATLASPix3.1	3×10^{14}	1.9 hr	2.0 hr	2.1 hr
ATLASPix3.1	1×10^{15}	–	–	–
ATLASPix3.1	3×10^{15}	–	–	–
ATLASPix3.0	None	1.5 hr	3.6 hr	2.2 hr

Table 9.5 December 2022: Total run time for each DUT device and operating temperature. The 1×10^{15} and 3×10^{15} $1\text{MeVn}_{\text{eq}}/\text{cm}^2$ were operated, however the data could not be used due to misconfigured settings and were disregarded.

9.3.1 Total efficiency vs. Pixel threshold

As previously described, the Corryvreckan [91] software package was used to reconstruct the events, clusters, and tracks. The General Broken Lines [95] track

model was used for tracking with exactly 6 hits required per track (one hit per telescope plane), a spatial cut of 300 µm × 300 µm and timing cut of 200 µs. The spatial cut defines an elliptical area in the x-y telescope plane within which a cluster must lie relative to the track intersect, to be associated to the track. The timing cut rejects clusters that arrived too early or late relative to the track time [92]. To avoid ambiguity, events with more than a single track were discarded.

The Corryvreckan module `AnalysisEfficiency` was configured to consider only tracks with a χ^2/N_{dof} less than 5, to ensure high quality tracks were used in the calculation of the pixel efficiency. Furthermore, tracks occurring within 200 ns of the telescope frame (event) edges were discarded from consideration. Additionally, some cuts are applied by default, such as: tracks intersecting outer edge pixels are removed; tracks intersecting close to masked pixels are removed. Tracks that do not intersect the DUT are also not considered.

The total efficiency is calculated as in Equation (9.1) for the entire sensor on a per-run basis. Runs were grouped by temperature, device ID, and pixel threshold setting and merged in order to determine total efficiency values per threshold setting value. The results are summarised in Figures 9.9 and 9.10. Finally, an additional noise estimate was calculated by looking at the DUT hits arriving outside of beam spills, as described in Ref. [79]. This is made possible by the fact that the DUT is read out untriggered, in contrast to the telescope which has a triggered readout. The noise estimate is plotted alongside the efficiency in Figure 9.11.

A general conclusion to be made is that the June and December results are consistent, excluding the 3×10^{14} $1\text{MeVn}_{\text{eq}}/\text{cm}^2$ sensor. In the December results, efficiency measurement quality is improved for the 3×10^{14} $1\text{MeVn}_{\text{eq}}/\text{cm}^2$ sensor owing to a larger available data sample, and the efficiency appears to be much higher with an efficiency of 75% at a setting of 150 DAC, compared to 60% in June. However, the noise rate per pixel appears to be significantly worse in that it does not reach the same sub-0.03 Hz/pixel level as quickly with increasing pixel threshold as the sensor did in June. This degradation has been attributed to annealing effects, caused by long periods spent unrefrigerated while in transit [79, 82].

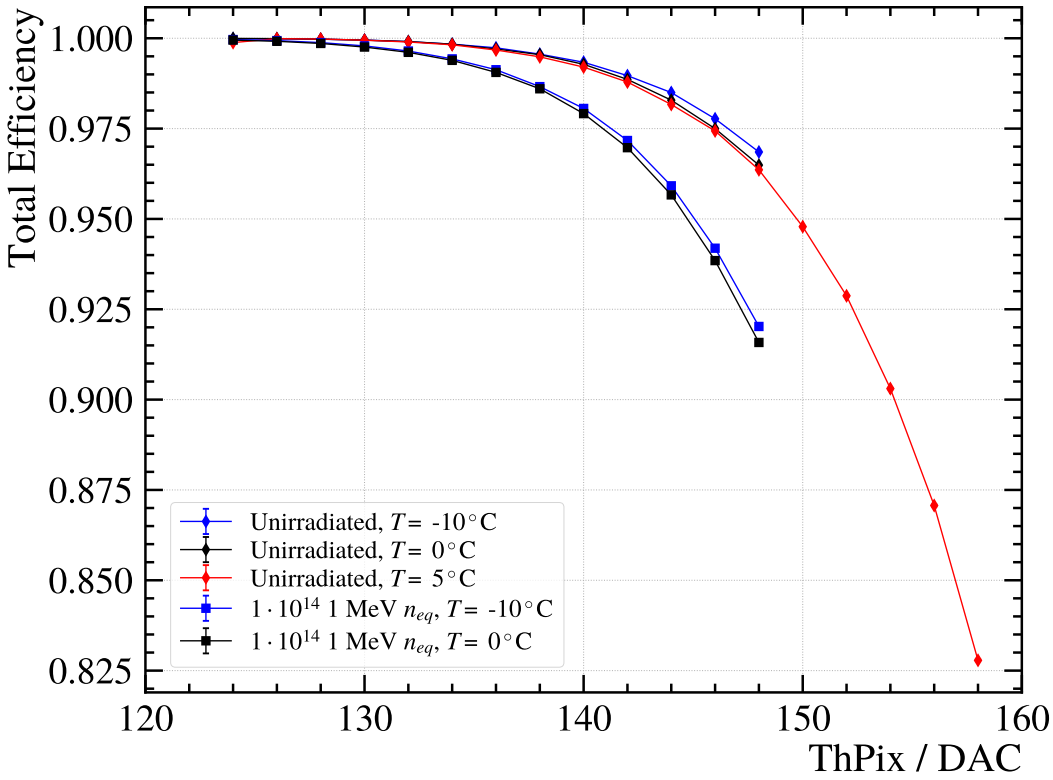

Figure 9.9 June 2022: total sensor efficiency versus pixel threshold setting in DAC units (excluding the $3 \times 10^{14}\,1\mathrm{MeVn_{eq}/cm^2}$ device). A high efficiency is observed for the unirradiated sensor decreasing with increased threshold (ThPix) value and temperature. A faster decrease is observed for the irradiated sensor.

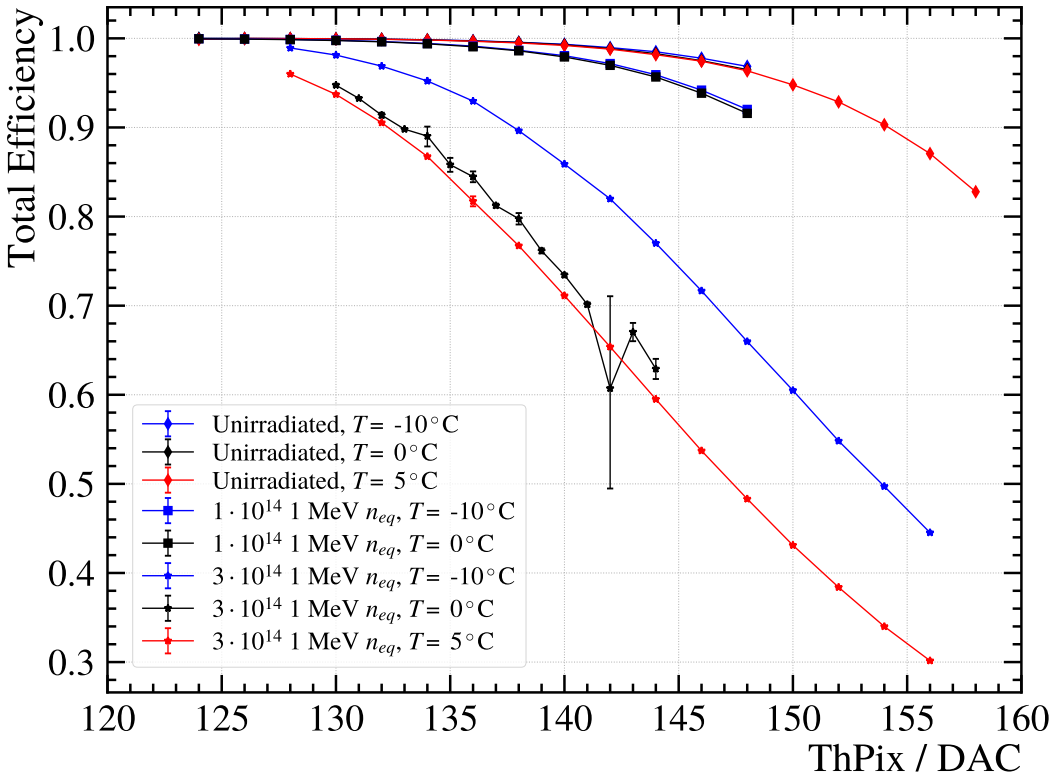

Figure 9.10 June 2022: total sensor efficiency versus pixel threshold setting in DAC units. The y-axis range is now expanded and all available measurements are now included. There is a significant drop in efficiency for the $3 \times 10^{14}\,1\mathrm{MeVn_{eq}/cm^2}$ sensor, with some higher uncertainty data points due to low availability of statistics.

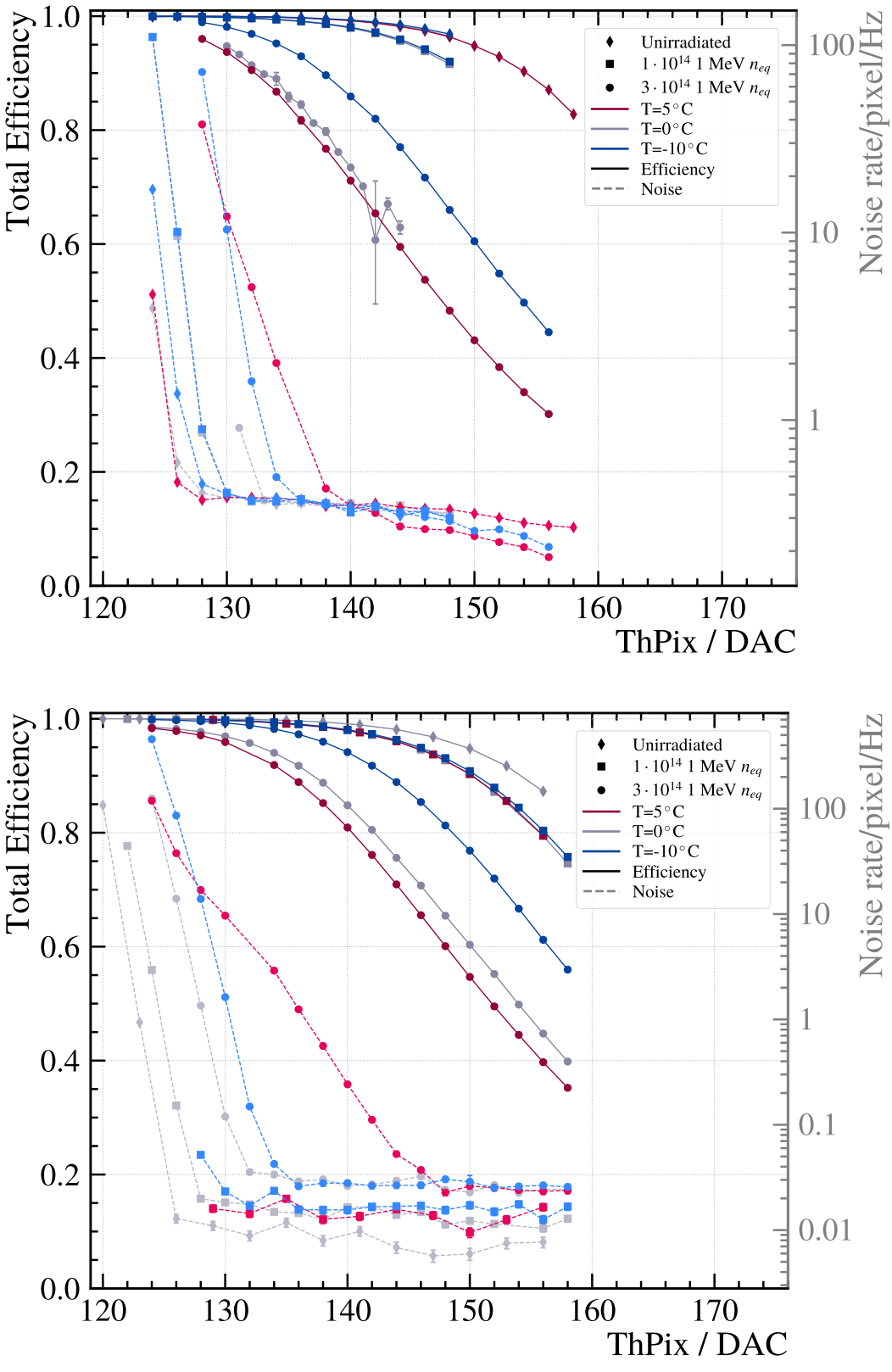

Figure 9.11 Total sensor efficiency versus pixel threshold setting in DAC units with an additional noise estimates (dashed connected data points) showing both June (top) and December (bottom) results. Reproduced from Ref. [79].

9.3.2 Time of Arrival resolution vs. pixel threshold

The ToA resolution is defined in Equation (9.2). As in the previous subsection, the same track $\chi^2/N_{dof} < 5$ requirement and frame edge cuts are imposed.

The t_{res} value calculated for the DUT sensor relative to the track is filled into a histogram, yielding a distribution such as the one in Figure 9.12. A single-sided crystal ball function from the RooFit [62] package adequately describes this distribution, with a symmetric Gaussian core and left side tail. In concrete terms the model is defined

$$f(x; x_0, \sigma, \alpha, n) = \begin{cases} A \cdot (B - \frac{x-x_0}{\sigma})^{-n} & \text{for } \frac{x-x_0}{\sigma} < -\alpha, \\ \exp\left(-\frac{1}{2} \cdot \left[\frac{x-x_0}{\sigma}\right]^2\right) & \text{otherwise,} \end{cases} \tag{9.7}$$

with normalisation factors

$$A = \left(\frac{n}{|\sigma|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right), \qquad B = \frac{n}{|\alpha|} - |\alpha|, \tag{9.8}$$

where x is the fit variable, x_0 is the mean of the Gaussian function, σ is the width of the Gaussian core, α is the distance in standard deviations from the mean after which there is a transition to a power law tail, and n is the exponent of the left-side power law tail [62].

Following the determination of the σ variable from the fit representing the time resolution, a further correction is applied due to the binning of the pixel timestamp which is provided in 8 ns bins, such that

$$\sigma_{\text{corr}} = \sqrt{\sigma^2 - \left(\frac{8\,\text{ns}}{\sqrt{12}},\right)^2}, \tag{9.9}$$

where σ is time resolution obtained from the fit.

The resulting ToA resolution values after correction plotted against the Pixel Threshold are shown in Figure 9.13. The ToA resolution approaches 5 ns towards the lowest threshold settings, however does not drop below it. This is in line with expectations that the ATLASPix3.1 would not be able to meet the requirements outlined in Table 9.1. However, the ATLASPix3.1 is used only as a proxy for measurements, and improvements in pixel electronics for the MightyPix are expected to yield improved timing performance in future studies.

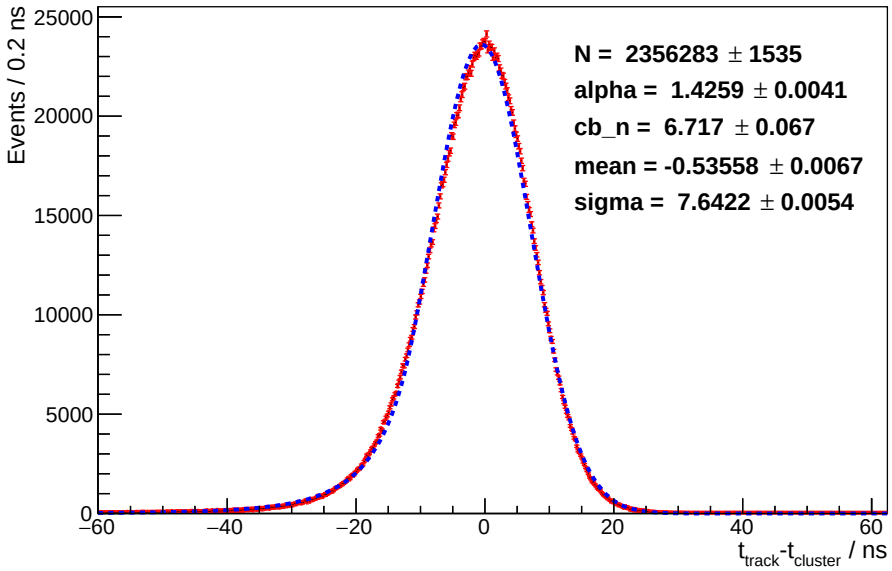

Figure 9.12 June 2022: Time resolution histogram from pixel threshold setting 132 for unirradiated ATLASPix3.1 operated at −10 °C, fit with crystal ball function as defined in the text.

For the 3×10^{14} 1 MeV n_{eq}/cm^2 sensor in both June and December results, a significant degradation in timing performance is observed. In December the ToA resolution increases at a higher rate to above 20 ns. The December testbeam data had a wider binning for the pixel timestamp of 10 ns, which does affect the reported ToA resolution.

It is not obvious from the ToA plots shown whether the in-time efficiency requirement is met. Additionally, there are some evident instabilities in the fitting process affecting data points. To resolve this, a more robust measure of timing performance that is more of interest can be used, the in-time efficiency, which will be discussed in the next section.

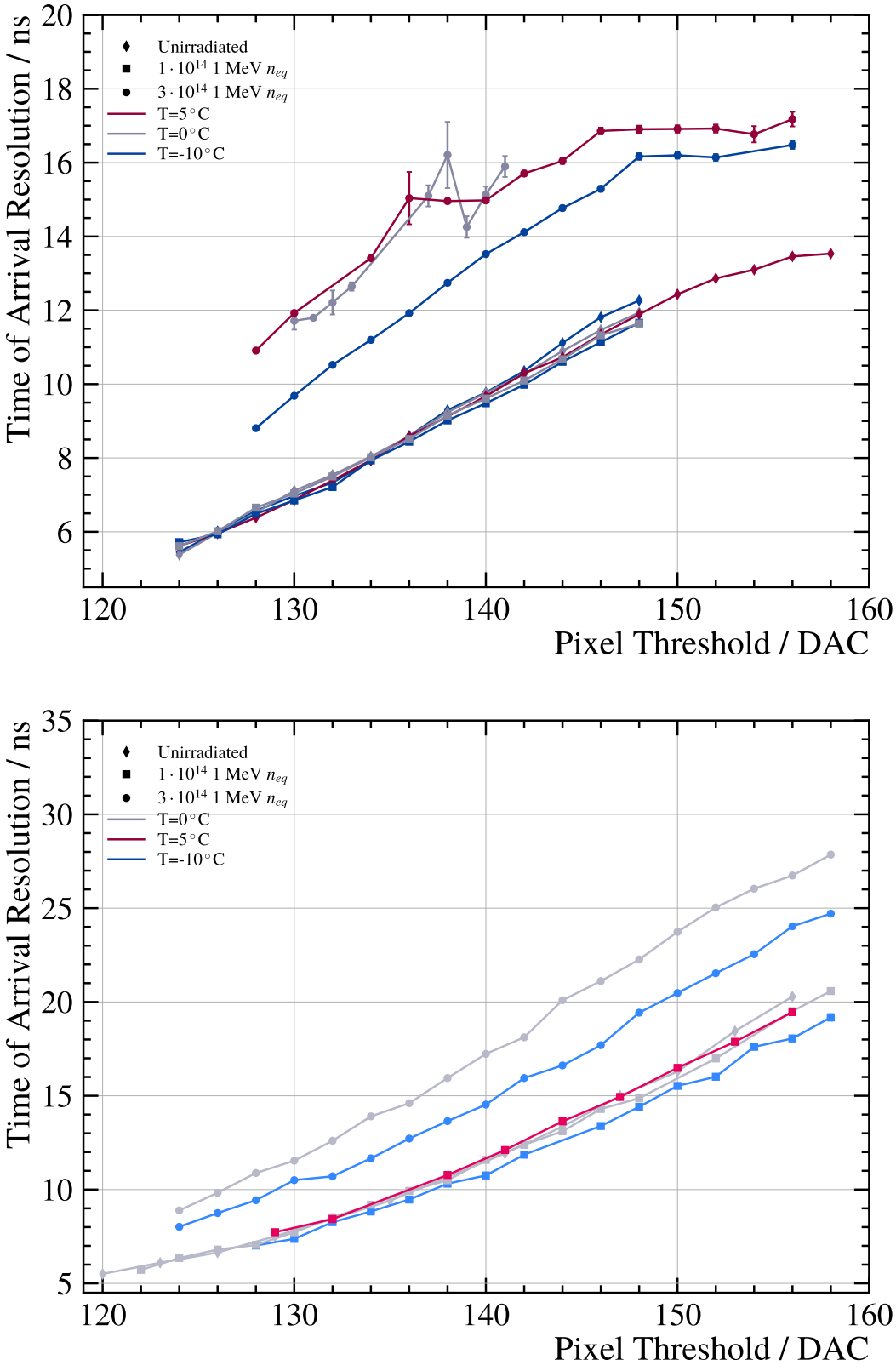

Figure 9.13 ToA versus pixel threshold setting in DAC units for June (top) and December (bottom) testbeam campaigns.

9.3.3 In-time efficiency vs. pixel threshold

The in-time efficiency is measured as the proportion of DUT hits that arrived inside a 25 ns window centred on the peak of the time residual distribution, as briefly discussed in Section 9.2.1. This does not rely on fitting methods and as such is considered a more robust measure of timing performance, and more directly relevant to MightyPix requirements and demands of the HL-LHC. This is because the most relevant deciding factor is whether an HV-MAPS used in the Mighty Tracker design is able to acquire all hits in the space of a 25 ns bunch crossing.

Figure 9.14 shows results for the June 2022 testbeam and shows a more clear dependence on irradiation of the sensors, which is particularly dramatic for the most highly irradiated sensor shown in the plot. In keeping with the conclusions of the previous section, none of the sensors fulfil the $> 99\%$ in-time requirement in Table 9.1.

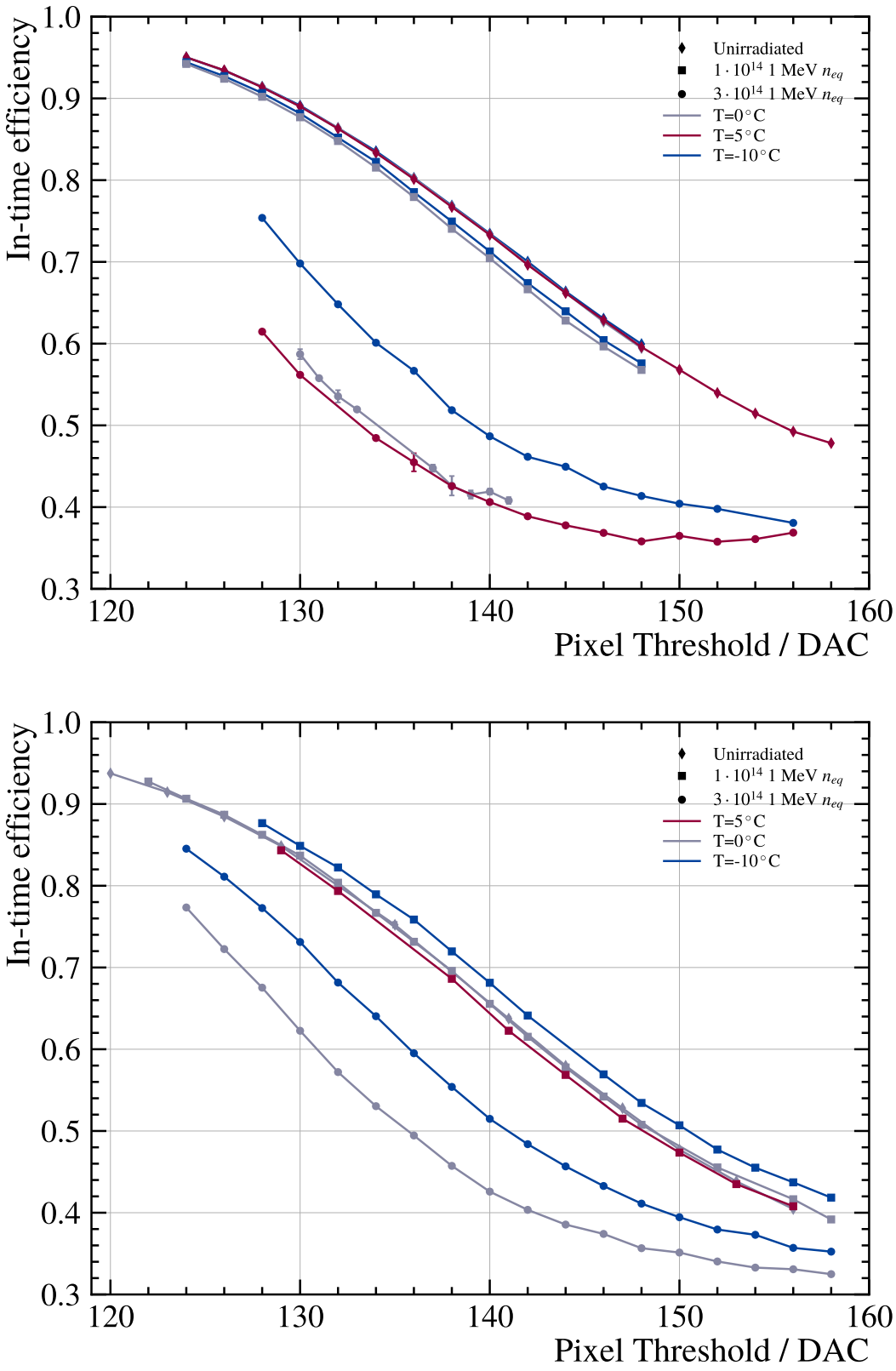

Figure 9.14 In-time efficiency versus pixel threshold setting in DAC units for June (top) and December (bottom) testbeam campaigns.

9.4 Conclusion and looking ahead

Two testbeam campaigns with the ATLASPix3.1 were conducted in June and December 2022 yielding successful measurements of the efficiency and timing performance for unirradiated and irradiated sensors. Both campaigns yielded results consistent with each other, with the December testbeam generally preferred due to the improved stability of the data-taking. In the December campaign, issues with the internal clock of the chips were worked around by introducing an external clock from the AIDA TLU [82], which improved the data-taking stability and ultimately increased collected statistics. The December testbeam results also included measurements of an unirradiated ATLASPix3.0 device, which could be compared with previously determined results elsewhere and the results of the ATLASPix3.1. In both campaigns, the most highly irradiated sensors, both 1×10^{15} and 3×10^{15} 1 MeV n_{eq}/cm^2, were not operated at the optimal tuning settings and were therefore unable to be studied.

The results of both campaigns generally indicate that:

1. **Lower temperatures are better**: The sensors operated better (higher efficiency, better timing) at temperatures 0 °C or lower.

2. **Higher irradiation levels lead to overall decreased total efficiency**: With increasing irradiation, the performance suffered as indicated by the degraded total efficiency of the sensors and increased noise hit estimates. This could still be mitigated by operating at lower temperatures, choice of electronics, tuning the pixel threshold values and other chip settings.

3. **Acceptable noise level achievable**: At low pixel thresholds, the noisy pixel hit rate is high, but decreases to a sub-0.03 Hz/pixel plateau at a threshold of between 130–140 DAC units, even for the more highly irradiated sensors. The total pixel efficiency for the lesser irradiated and unirradiated sensors remained high. The pixel thresholds should be tuned in order to control this noise rate, while preserving the total efficiency.

4. **ATLASPix3.1 timing performance does not meet MightyPix requirements**: The expected 5 ns time-of-arrival resolution was achieved at the lower pixel thresholds, which does not meet the 3 ns requirement listed in Table 9.1. This is however expected, and may improve in later MightyPix prototype iterations.

5. **DUTs may have suffered annealing**: Due to an unrefrigerated period while in transit after the June campaign, the sensors as they were operated in December may have undergone some annealing, affecting the overall performance of the irradiated chips. This can be observed in a wider spread of noise rate estimates in e.g. Figure 9.11(bottom).

By carrying out both campaigns, a foundation of knowledge, setups, and tools has been developed which may carry forward to future planned campaigns with real MightyPix sensor prototypes. This involves an analysis framework and method that is highly re-usable (as demonstrated by its use in two of the discussed campaigns in this thesis), and automated to reduce redundant manual analysis of the data. The data-taking setup used in these testbeams has also been demonstrated to work in a stable manner, and may also be used in future campaigns with the upcoming MightyPix prototypes. It is hoped that this will lead to a fast turnaround of results, and ultimately quicker feedback towards the design of the MightyPix.

Chapter 10

Analysis user experience enhancements

An analyst in High Energy Physics often faces a range of technical obstacles, learning curves, and challenges during their data analysis — often these are not directly related to any physics problem under investigation and can involve unrelated operational issues, or application bugs, falling out of the scope of regular physics analysis tasks. While many established workflows and tools exist, as data volumes, the demands on limited infrastructure, and expert time escalate, newer and more robust approaches are becoming increasingly necessary.

Some areas where this is particularly evident is the creation of n-tuples for offline analysis, and user jobs submitted to the Worldwide LHC Computing Grid (WLCG) (managed by DIRAC [48]). This section will outline some of the challenges that have emerged since Run 2, the author's contributions to the Data Processing & Analysis (DPA) project such as the creation of the LbAPWeb application (Section 10.3) and the evolution of the Analysis Productions service from pre- to post- Run 3. This work is primarily motivated from the analyst's experience while working on the analysis in Part II with particular relevance to Section 3.2.3.

10.1 The LHCb analyst experience from Run 2 to the Upgrade

Before Run 3, LHCb analysts created n-tuples by submitting their own computing jobs (referred to as user jobs), monitoring the status of the jobs as they execute, then collecting and dealing with the output of the jobs after completion. This was typically carried out by interacting directly with the WLCG, the Grid, using the Ganga [96] application running on the user's computer or, more commonly, via the LXPLUS service. Ganga handles all necessary interactions with the Dirac service to configure a standalone set of jobs and processing of datasets indicated by the user, with the user-provided application configuration options. Job statuses and outcomes are submitted, tracked by the Ganga application, and data stored at storage elements at Grid sites. If these jobs failed, the user would need to manually inspect each failing job to determine the cause of the problem and either resubmit the job as-is, re-submit the entire set of jobs with a modified application configuration, or consult the community for further assistance. This manual monitoring aspect is by far the most time-consuming and error-prone aspect of the user job model.

The user will either keep output files at the Grid sites that they were produced at, or download them to a storage area under their own management. Data preservation is usually handled by manual processes agreed within Physics Working Groups (WGs), typically only once an analysis has reached late stages of review. All metadata pertaining to the job configuration, application configuration, and application logs is stored temporarily by the Grid or on the user's computer, and likely to be lost if not manually preserved.

Since Run 2, due to the increased luminosity of the LHC and higher event rates for LHCb, the expected size of datasets that an LHCb analyst would need to run over will increase considerably. With the increased size of datasets, the demands on user time increase with more data to handle, more jobs to monitor, and a higher occurrence of failures to handle.

To address this issue, the Analysis Productions system which originated from and was extended from the WG productions system [97], was introduced to alleviate the anticipated pressures on users. This was achieved by providing a simple method to configure n-tuple productions to run in a centrally-managed system. The user provides only the necessary information required to run an n-tupling

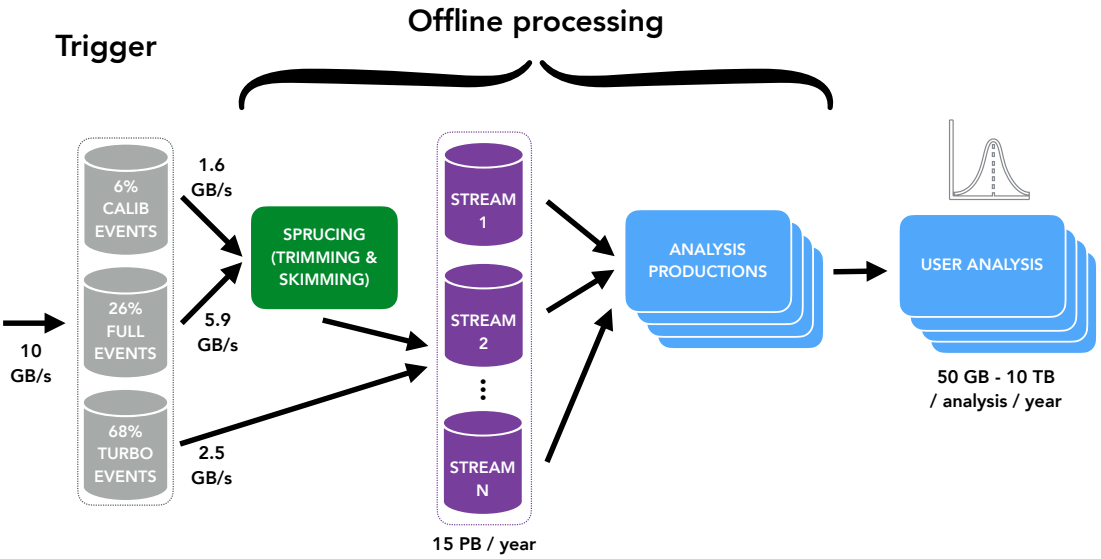

Figure 10.1 Run 3 offline processing data flow diagram. Reproduced from Ref. [39].

workflow within the DIRAC Transformation System (as with other centralised offline processing and simulation productions), which handles job submission, monitoring, and resubmission in place of the end-user. The operational aspect of running the n-tupling workflow is managed by LHCb Grid experts.

10.2 Analysis Productions

The Analysis Productions service represents a new approach to n-tupling workflows with a centrally-managed approach, completely removing the job monitoring, user data preservation and management aspect from user and WG responsibilities. The service is developed as part of the DPA project [49, 50], and is considered part of the official offline processing workflow as the final step before user analysis (see Figure 10.1). As such, it is a tool that many LHCb users will interact with when creating the n-tuples for their analysis. In 2024, this tool was used extensively, with an 30-day input throughput average of $\approx 25\,\mathrm{GB\,s^{-1}}$ with spikes of $50\,\mathrm{GB\,s^{-1}}$ to $180\,\mathrm{GB\,s^{-1}}$ in the latter half of the year, taking note that the LHC and LHCb began operations in March (see Figure 10.3). In November that year, approximately 2,500 samples [98] were created from 158 requests (see Figure 10.2).

A simple interface is provided to configure a productions via a YAML format

```yaml
2018_MagUp:
  wg: Charm
  inform:
    - roneil
  application: DaVinci/v45r6
  options:
    - "XicpStStToXicpPiPi.py"
  automatically_configure: yes
  turbo: yes
  input:
    bk_query: "/LHCb/Collision18/../../../CHARMMULTIBODY.MDST"
  output: XICPSTSTTOXICPPIPI.root
```

Listing 1 An example YAML specification for an Analysis Production job. This configuration is fully described in the text.

configuration file and accompanying application option files, stored in a GIT repository accessed via the CERN GitLab service. GitLab naturally facilitates review and testing via continuous integration workflow for each production before submission, which happens automatically after a merge request is approved and merged by a liaison representing their WG.

Upon merging, submission of the production is an automatic process. The YAML configuration file is used to create DIRAC Transforms and a Production Request for each n-tuple to be produced. A user-written YAML configuration file specifies all details of input/output data, application(s) to run, where to find the application option files, who should be notified of the status of the production, and where the outputs should be stored. All configuration files and metadata used to run each production are preserved by DIRAC and Analysis Productions, ensuring provenance and traceability. Should any data files be archived or lost, they can always be recreated from this preserved metadata in DIRAC. Output data are also tracked in the LHCb Bookkeeping system [99].

Listing 1 shows an example of a valid YAML configuration file which was used to create n-tuples for the analysis described in Part II, specifically the processing of the HLT2 (Section 2.2.7) collision data collected in the year 2018. For organisational purposes a responsible WG and user is specified. This is followed by the application name and version to be used for the processing, and path(s) to the user-provided `options` file(s) which will configure the application (i.e. what decay(s) should be considered, what variables should be calculated and stored in the output n-tuple, et cetera). In this case, DAVINCI is used, as it is the standard

Figure 10.2 A summary of opened and closed productions in 2024.

application maintained by the collaboration to create n-tuples for any data stored offline for further processing. The input data to the application is specified as a path in the LHCb Bookkeeping System describing a dataset of Run 2 HLT2 output, and the output filename. A specific output location is not required to be specified by the user, as this is handled automatically.

Application errors are caught by small-scale automated testing before submission via a continuous integration pipeline. This is carried out by submitting a single DIRAC job, per Analysis Production job, that runs the configuration with a small subset of the full input dataset. Application log files and other debugging tools are made available to the user via a web application developed by the author in Section 10.3. If tests are successful, the productions are permitted to be submitted. Operational errors after submission are handled by LHCb Grid experts, either manually or (increasingly common) by automated recovery procedures.

The presentation of feedback generated by the continuous integration testing process to the user is a particularly important feature. In particular, ensuring analysts are able to validate the data produced by their proposed application configuration as efficiently as possible, while also ensuring the workflow will not adversely burden the Grid, and highlighting and debugging application errors or unintended application behaviours efficiently. The main focus of this chapter will be the developments in this area, which led to the creation of the web application, LbAPWeb, and the backend linking it to DIRAC and the Analysis Productions system, LbAPI.

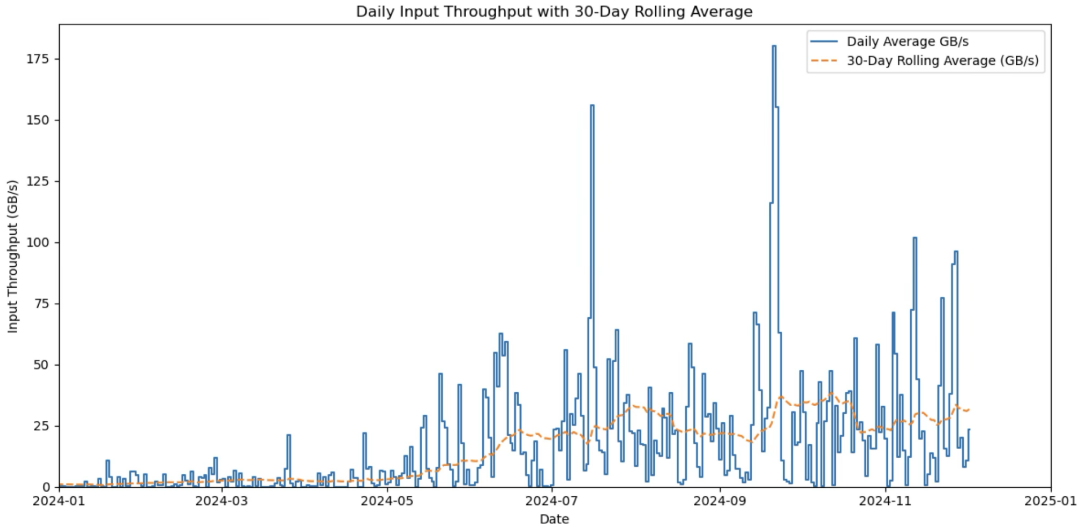

Figure 10.3 Analysis Productions' input throughput in 2024 with 30-day rolling average [98].

10.2.1 LbAPI: RESTful access to Analysis Productions and Dirac

To adequately extend the Analysis Productions service, a Representational State Transfer (REST) API was created to provide a convenient interface between Analysis Productions services and a variety of potential new tools, utilities, and external services. Figure 10.4 demonstrates how LbAPI is linked to other services, and the applications available to users that provide an interface to these linked services.

As shown in Figure 10.4, a range of other services are linked to LbAPI. The Analysis Productions continuous integration (CI) pipeline, is run inside a `celery` task queue which validates the configuration, submits jobs to the LHCbDIRAC service, and monitors these jobs until they have completed. The CI pipeline is triggered via webhooks sent to LbAPI when a user pushes a commit to the GIT repository, which causes a new task to be created and sent to `celery`. The jobs sent to DIRAC continually report on the test status and upload logs and output files to LbAPI, which are stored in databases and OpenSearch.

Testing outcomes and feedback are reported on GitLab in comments, the LbAPWeb tool in Section 10.3, and additionally by directly integrating with the GitLab CI system.

The LbAPI tool additionally provides an interface for the Analysis Productions

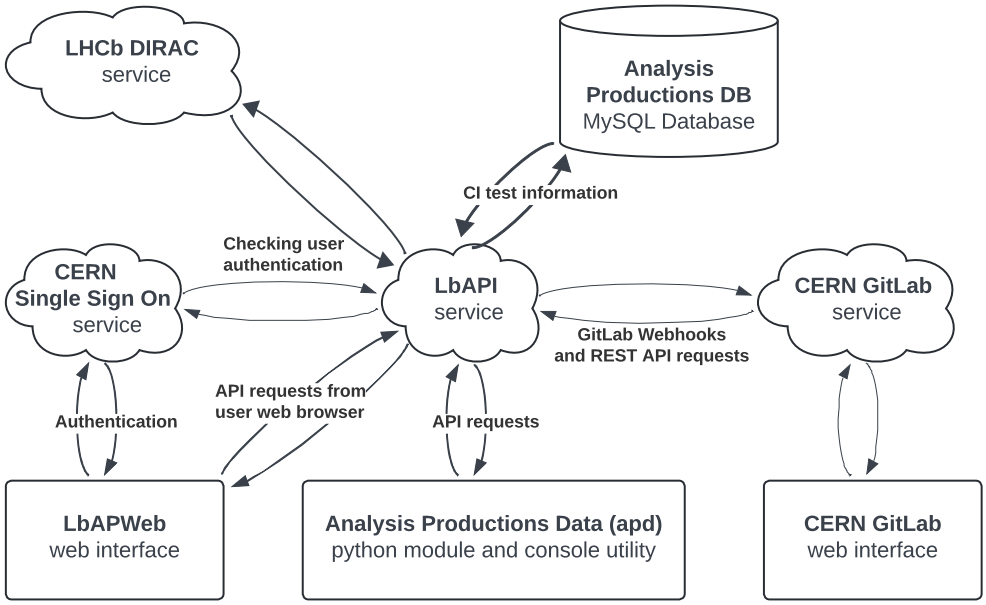

Figure 10.4 Diagram of services linked to LbAPI.

Data (apd) tool [100] which is a python library and console utility that can be trivially installed by analysts and used in conjunction with their analysis scripts. The tool is used to query the location of produced n-tuples, via a unique set of identifiers (tags, names, version numbers) to identify the sample. The tool is able to authenticate with CERN Single Sign-On (SSO), send queries to LbAPI (which in turn queries DIRAC), and returns a list of paths to all constituent files of the requested n-tuple. They can either be downloaded, or accessed remotely via the XRootD protocol, dependent on the location.

The apd tool is a crucial component of analysis preservation, in that users only need to refer to their data by their identifying metadata (name, version) and/or tags, which are not dependent on where the data is stored (as this may change over time). N-tuples produced in the Analysis Production system are automatically tagged by their data type (data or simulation), data taking year, and magnet polarity. Additional tags may be added manually by the user through the LbAPIWeb interface, further aiding organisation of the data. Further addition of automated tags may significantly increase the searchability of all LHCb n-tuples, potentially encouraging the re-use of existing datasets.

Finally, the LbAPI service provides all information to the LbAPIWeb web application, which runs entirely in the user's browser. A browser queries

the REST endpoints to fetch all the necessary information about Analysis Productions tests and available datasets from submitted productions. LbAPI has all necessary links to the Analysis Productions database, OpenSearch instance, and DIRAC to provide this information.

10.3 LbAPWeb: an intuitive interface to creation and management of LHCb n-tuples

A web application, `LbAPWeb`, was created to replace the pre-existing `flask` web application of Analysis Productions. LbAPWeb was developed initially with a focus on improving the user experience of the Analysis Productions service, interpretability of CI test results, and in addition an interface to browse and search datasets produced using the Analysis Production system and display their metadata. Since its creation, a number of additional features have been added, and another service (LbMCSubmit) now operates through this web application.

Validation of productions and debugging of issues is often viewed as an optional open-ended task, usually involving a degree of manual work by the user to download a file to their computer, open it and create some kind of figure of merit, or run a portion of their analysis steps to verify their production is ready to submit in its entirety.

If an analyst runs into problems, it is not always obvious what counts as an error, or what observations, if any, should be reported to the wider community and/or referred directly to experts. Problems may be missed or ignored, due to the time, manual work required, or expert knowledge requirements, causing production output to be incomplete due to failures or the failure of all submitted jobs (and potentially high wastage of CPU time on Grid infrastructure). Additionally, job failures may occasionally be caused by operational problems at Grid sites, not directly related to the user's production. From a user perspective, it can be challenging to distinguish these from application errors arising from genuine application bugs, or user errors originating from application misconfiguration.

With the wide adoption of the Analysis Productions system in Run 3, the model has shifted towards a more centralised approach with standard testing, provenance, and preservation baked in. Once submitted, a production is run automatically and investigated by Grid experts in the event of job failures. A

CI testing pipeline improves the chance of application errors or misconfigurations being noticed before submission.

This new web application, LbAPWeb, was developed as a vehicle to present a wide range of feedback, metrics, tests, and even interactive output file browsing functionality (with JSRoot) from the CI results with little additional efforts from analysts, provided that they use the service. Modern web technologies are leveraged to provide an easy to parse interface and to visualise results in various ways by taking advantage of the large range of packages from the Node Package Manager (NPM) [101] ecosystem. Providing the results in this way, 'for free', saves analysts from having to perform basic validations manually by themselves, an action performed by many LHCb analysts daily. Considering the frequency of such actions, the time savings can be considered significant.

10.3.1 The application

The web application, named LbAPWeb, is a static frontend Single-Page Application (SPA) created using NextJS [102] and React [103] frameworks in the JavaScript and Typescript languages. The frontend handles all user-facing interface aspects, while querying a REST Application Programming Interface (API) backend in order to access relevant information about production testing information and available data from DIRAC and Analysis Productions. The user interface is created using the Bootstrap and React-Bootstrap libraries.

The data provided from the API is protected behind a layer of authentication. User authentication is handled via CERN SSO and the issuing of bearer tokens stored temporarily in the browser.

The choice of frameworks and the use of the NPM ecosystem has facilitated a smooth initial application development and various extensions with new features and improvements. It has been extended by other projects in the LHCb collaboration significantly since creation, including the development of the LbMCSubmit system frontend, providing an analogous experience with the validation and submission of LHCb Simulation productions.

The application is deployed automatically via GitLab Continuous Integration and Deployment (CI/CD) to CERN OpenShift.

10.3.2 Executive summary of implemented features

This section will provide a general overview of the features available in the application:

- **User-friendly interface to Analysis Productions**: easily navigable, and fast-loading – users can quickly locate the information they are looking for to validate that their production will work as intended.
- **Easy-to-parse layout of configured productions and test results**: information is displayed in tables (Figure 10.6), with colour-coding, icons, and with clear headings and user interface elements.
- **Interactive display of test results**: Test results are displayed through interactive log viewers with line numbers, search, highlighting functionality, and plots aiding the debugging process.
- **Facilitation of fast debugging**: failing jobs are clearly signposted, with a quick navigation route to information (log files, memory usage) that can identify the problem (Figure 10.7).
- **Interactive browsing of n-tuples via JSRoot**: Users can quickly check and inspect histograms of their n-tuple variables using the JSRoot [104] browser provided, without downloading any files or writing scripts.
- **Quick access to Grid impact metrics**: Memory monitoring results via the `prmon` [105] tool are visualised using a plot of the RAM usage vs. wall clock time (Figure 10.5), as well as additional timing and overall metrics to quickly indicate whether the production will run smoothly on the Grid.
- **An index of all n-tuples productions and metadata**: All submitted productions and the n-tuples produced can be browsed by members of the collaboration. Metadata pertaining to storage use, links to CERN GitLab merge requests and issues, provenance information, DIRAC metadata, and other information can be easily accessed (Figure 10.9). Data management tasks are additionally possible. Old or superseded datasets can be marked as archived, indicating to LHCb data management teams that the data can be cleaned to preserve storage. Datasets may also be added to and shared between different analyses, should it be re-used (Figure 10.8).

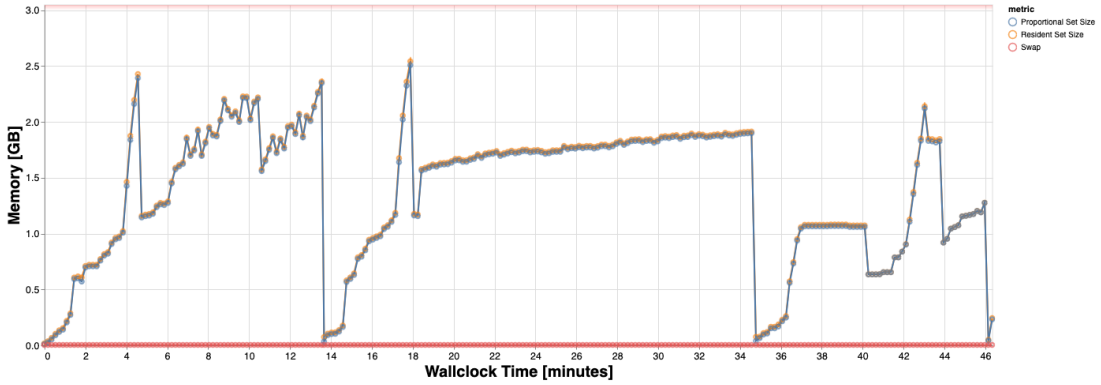

Figure 10.5 Job resource utilisation plot as rendered and shown on LbAPWeb.

- **Comprehensively documented**: Dedicated `Sphinx` documentation pages for using the systems described are available to users and linked from this web app.

10.3.3 Conclusion and further work

A new web application, LbAPWeb, has been delivered with a view to improving the offline analysis experience and organisation of LHCb datasets within the collaboration. Through the delivery of this application, and the evolution of the various Analysis Productions services to support it, the debugging process for productions have become simpler, and multiple benefits have been realised. For example, awareness has increased within the collaboration regarding memory usage on the Grid, through the introduction of `prmon` [105] to the testing process, with a visualisation of the job resource usage in LbAPWeb as shown in Figure 10.5. This has lead to multiple discoveries of memory leaks, and other resource utilisation issues. The testing workflow and presentation of results via the web application has also provided a simpler way for analysts to share reproducible demonstrators for application bugs and issues. If a user discovers an application error that needs to be reported, they need only share a link to demonstrate the problem and the steps needed to reproduce it. These are all in addition to the direct features and benefits discussed in Section 10.3.2.

Ultimately, a quicker turnaround in creating, testing, and validating n-tupling workflows mean that LHCb physicists are able to progress their analyses with a quicker turnaround on their n-tuples, without the need to adapt to changes

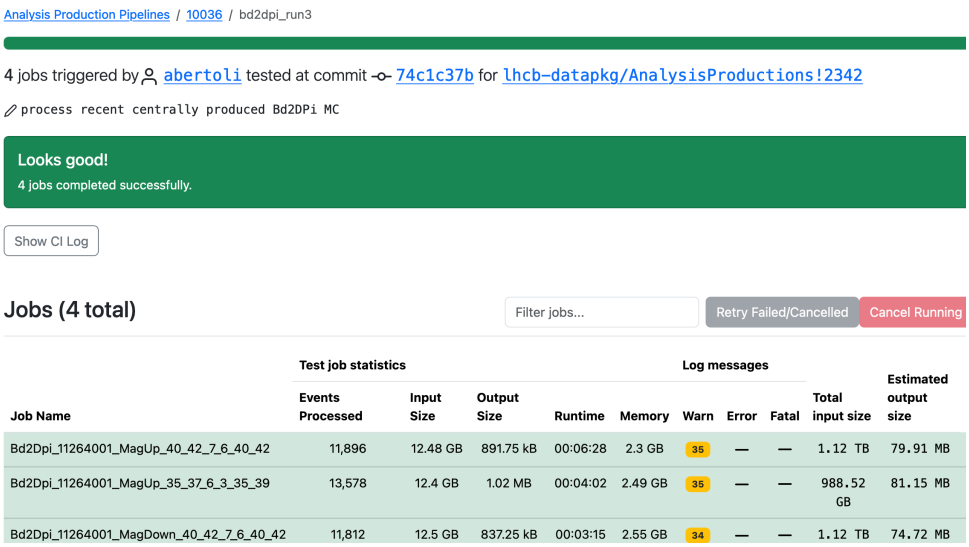

Figure 10.6 LbAPWeb: Overview page for a single instance of an Analysis Productions CI test. The individual test jobs and associated metadata are displayed alongside results and progress information.

in the complex Grid environment, and the challenges of processing ever-larger volumes of collision data. This is particularly likely to benefit newer members of the collaboration. Additionally, the uptake within the collaboration has been encouraging. For example, in 2024 an average of 83.7 productions were submitted per month, with as many as 158 submitted in November corresponding to approximately 2,500 created samples [98], see Figure 10.2.

The application has been significantly extended to support other projects within the collaboration, most notably the LbMCSubmit project. The LbMCSubmit project provides a similar experience to Analysis Productions with regard to prototyping, testing, and submitting LHCb Monte Carlo simulation production workflows. The LbAPWeb application is used to display the results of testing, to aid the user in debugging and validating their Simulation production, providing a similar experience to Analysis Productions.

The web application is likely to be expanded to meet a host of additional needs and provide more features as Run 3 data-taking continues, the amount of available collision data increases to greater volumes, and efficient utilisation of LHCb resources becomes of greater importance, all balanced against ensuring that physics analyses can proceed smoothly. This will likely involve the expansion of automated ‘standard’ tests and validations, expansion of log file scanning and analysis, and a host of automated recommendations to analysts to improve the

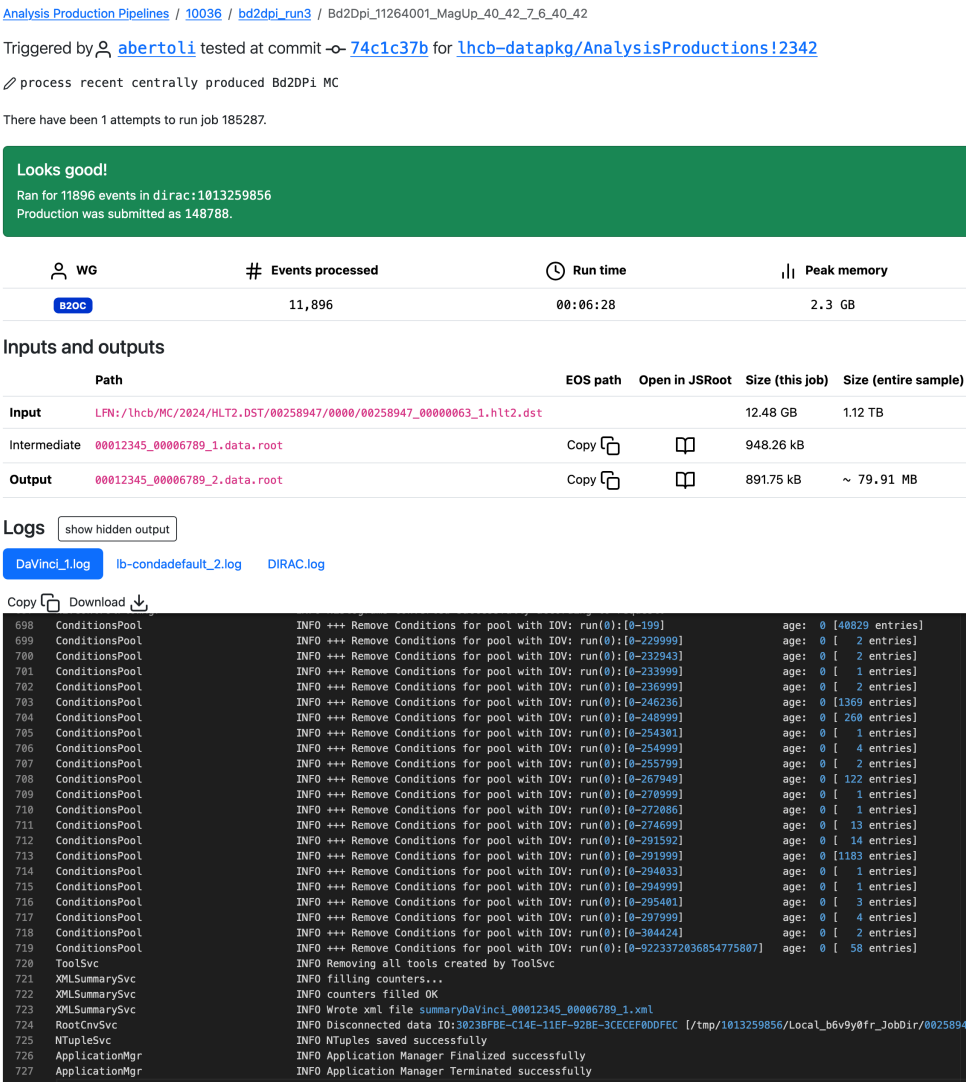

Figure 10.7 LbAPWeb: A page showing a single test job within the Analysis Productions CI test. More details about the job configuration, test status, inputs and outputs, and log files for debugging purposes are displayed.

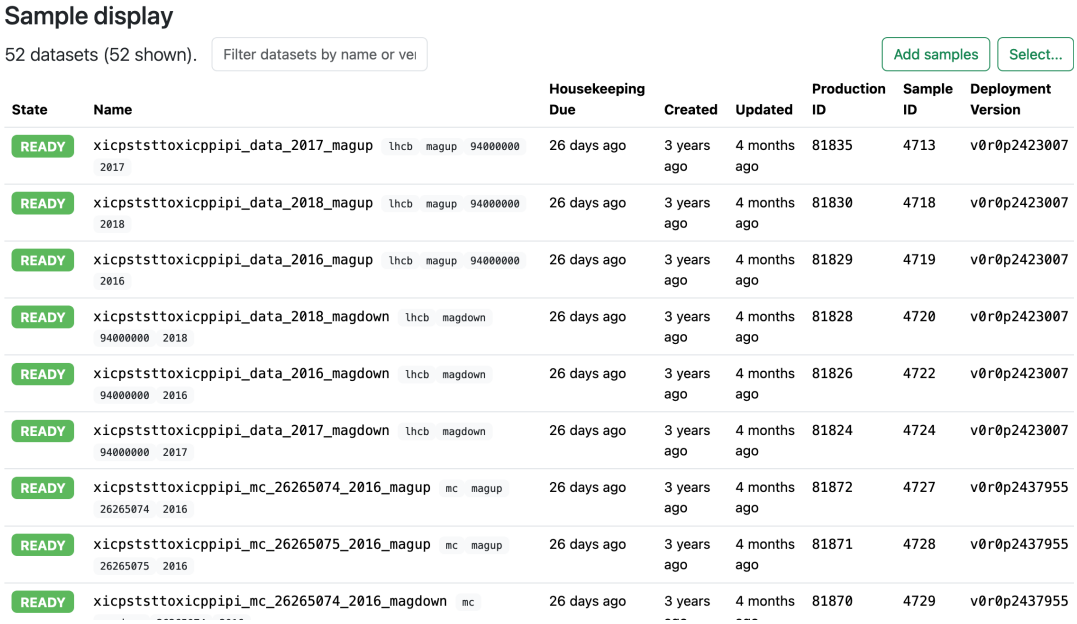

Figure 10.8 LbAPWeb: A overview table of samples grouped under a single analysis. There are options to add samples from another analysis, and to select and manage one or more samples.

efficiency of their workflows, maximising the overall benefit for the collaboration and valuable time savings.

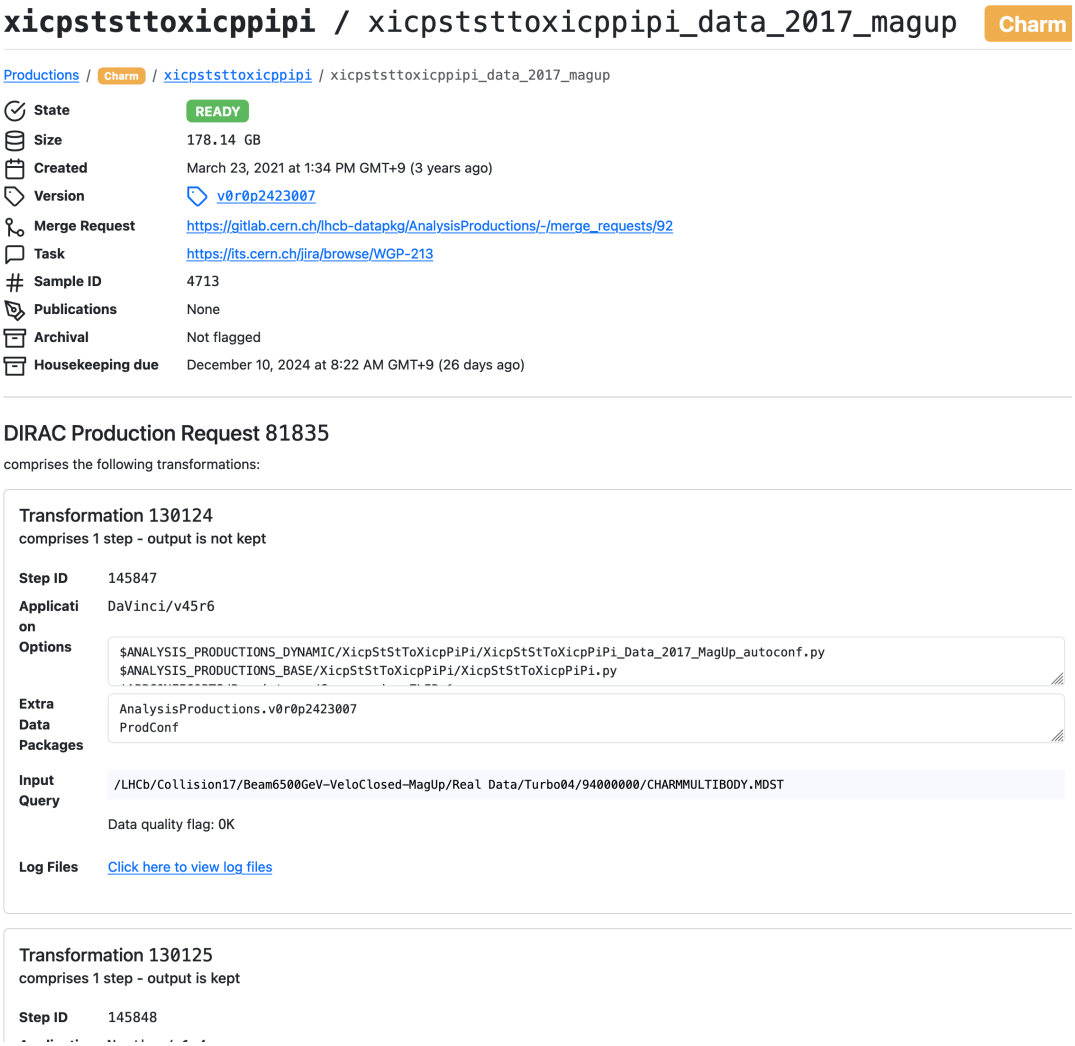

Figure 10.9 LbAPWeb: A "single sample" view of the details pertaining to an analysis sample.

Part IV

Conclusion

Chapter 11

Conclusion

This thesis covered in four parts: the introductory material; charm spectroscopy analysis plus results; testbeam campaigns and studies following; and finally work performed throughout toward efforts to improve the LHCb analysis user experience, motivated by accumulated experience while performing the work. This chapter will present a broad summary of the results presented in each area and offer ideas on future directions of the work covered.

11.1 Observation of new Ξ_c^+ baryons decaying to $\Xi_c^+\pi^-\pi^+$

A search for excited states in the $\Xi_c^+\pi^-\pi^+$ channel was conducted using 5.6 fb^{-1} of data collected by the LHCb detector in 2016–2018. Candidates for excited charm baryon states were reconstructed using a series of three selection stages, starting from Ξ_c^+ candidates picked by a specialised selection line in the LHCb HLT2 trigger.

After unblinding, the mass spectrum ΔM revealed four peaks at $\Delta M = 69.80, 175.95, 221.77, 329.92$ MeV. These peaks are hypothesised to correspond to the $\Xi_c(2815)^+$, $\Xi_c(2923)^+$, $\Xi_c(2970)^+$, and $\Xi_c(3080)^+$ resonances, which are all known to the PDG except $\Xi_c(2923)^+$ although the $\Xi_c(2923)^0$ is a known state. Upon development of an appropriate fit model, an unbinned maximum likelihood fit was carried out, and is displayed in Figure 11.1.

The parameters determined from the fit alongside their uncertainties are

$$
\begin{aligned}
N[\Xi_c(2815)^+] &= 4072 \pm 77 \text{ candidates}, \\
\Delta M[\Xi_c(2815)^+] &= 69.80 \pm 0.03\,(\text{stat}) \pm 0.03\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(2815)^+] &= 2816.65 \pm 0.03\,(\text{stat}) \pm 0.03\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2815)^+] &= 2.07 \pm 0.08\,(\text{stat}) \pm 0.12\,(\text{syst})\,\text{MeV}, \\
N[\Xi_c(2923)^+] &= 738 \pm 76 \text{ candidates}, \\
\Delta M[\Xi_c(2923)^+] &= 175.95 \pm 0.28\,(\text{stat}) \pm 0.29\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(2923)^+] &= 2922.80 \pm 0.28\,(\text{stat}) \pm 0.29\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2923)^+] &= 5.3 \pm 0.9\,(\text{stat}) \pm 1.2\,(\text{syst})\,\text{MeV}, \\
N[\Xi_c(2970)^+] &= 6105 \pm 255 \text{ candidates}, \\
\Delta M[\Xi_c(2970)^+] &= 221.77 \pm 0.46\,(\text{stat}) \pm 0.30\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(2970)^+] &= 2968.62 \pm 0.46\,(\text{stat}) \pm 0.30\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2970)^+] &= 31.7 \pm 1.7\,(\text{stat}) \pm 1.7\,(\text{syst})\,\text{MeV}, \\
N[\Xi_c(3080)^+] &= 344 \pm 77 \text{ candidates}, \\
\Delta M[\Xi_c(3080)^+] &= 329.92 \pm 0.72\,(\text{stat}) \pm 0.11\,(\text{syst})\,\text{MeV}, \\
m[\Xi_c(3080)^+] &= 3076.77 \pm 0.72\,(\text{stat}) \pm 0.11\,(\text{syst}) \pm 0.23(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(3080)^+] &= 6.8 \pm 2.3\,(\text{stat}) \pm 0.9\,(\text{syst})\,\text{MeV},
\end{aligned}
$$

where N is the number of candidates calculated from the signal PDF area for the corresponding state, ΔM is as defined in Equation (4.14), m is the measured mass calculated from ΔM, and Γ is the measured width of the resonance.

The $\Xi_c(2923)^+$ is observed for the first time, and the $\Xi_c(3080)^+$ is a known state but observed in $\Xi_c(2645)^0\pi^+$ for the first time. Focussing on the $\Xi_c(2923)^+$, the measured mass and width can now be compared to that of the $\Xi_c(2923)^0$ which was observed by LHCb,

$$
\begin{aligned}
m[\Xi_c(2923)^0] = &\quad 2923.04 \pm 0.25\,(\text{stat}) \pm 0.20\,(\text{syst}) \pm 0.14\,(\text{PDG})\,\text{MeV}, \\
\Gamma[\Xi_c(2923)^0] = &\quad 7.1 \pm 0.8\,(\text{stat}) \pm 1.8\,(\text{syst})\,\text{MeV},
\end{aligned}
$$

as measured in Ref. [19].

Comparisons with the measurements as listed in the PDG and PDGLive [1, 7] demonstrates broad agreement with existing experimental measurements (see:

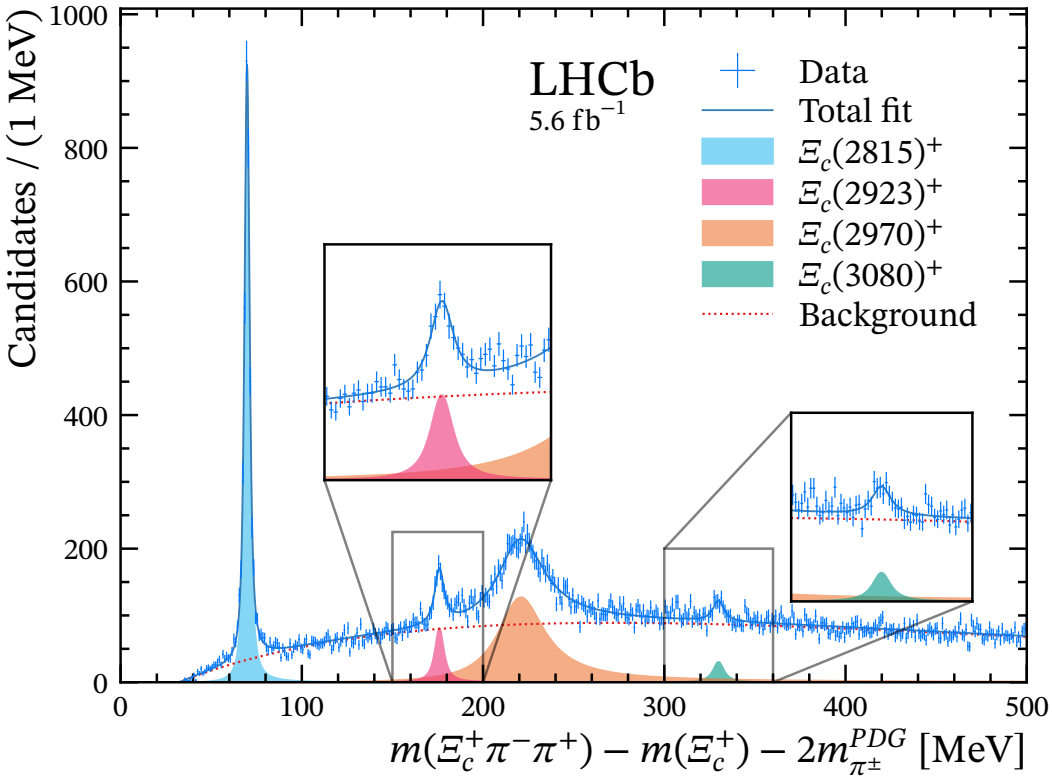

Figure 11.1 Plot of the $\Xi_c^+\pi^-\pi^+$ mass spectrum ΔM histogram, with overlaid fit model.

Figures 8.1 to 8.3).

Additional efforts will be necessary to determine the spin and parity quantum number assignments J^P, which were not determined in this instance. This will require a further understanding of the $\cos\theta_c$ helicity angle distribution for this decay in the LHCb context, in order to make a J^P assignment with reasonable confidence.

At the time of writing, this analysis has passed internal review stages and a paper draft has entered LHCb collaboration-wide review pending submission to a journal (Physical Review Letters). Depending on the outcome of the remaining review processes, the analysis as it is presented in this thesis may differ from the published result.

11.2 Testbeam studies of irradiated ATLASPix3.1

Two testbeam campaigns were carried out at the DESY-II Testbeam Facility in June and December 2022, with multiple ATLASPix3.1 sensors tested at varying irradiation levels and temperatures $(-10,\ 0,\ 5)$ °C. The primary focus of the campaigns were to measure the efficiencies of the sensors as a function of operating temperature, pixel threshold, and irradiation level.

The operation of the data-taking setup in December tended to be more stable than the June setup, which demonstrated instabilities leading to shorter runs and less collected data overall. The stability issues were resolved in December by the use of an external clock as described in Ref. [82]. These improvements, coupled with the knowledge accumulated from the June testbeam enabled a larger range of sensors to be tested, including the ATLASPix3.0.

It was generally concluded that: the sensors operated best at temperatures below 0 °C; a higher irradiation level causes a significant degradation in performance that is worse at increased temperatures; acceptable noise levels are achievable, after some tuning of the pixel thresholds; sub-5 ns time-of-arrival resolution was not achieved, but this was as expected.

Carrying out both campaigns resulted in an accumulation of valuable experience and tools that can be carried forward to future campaigns aiming to test upcoming MightyPix sensor prototypes. In particular, the reconstruction chain using the

Corryvreckan application may be adapted, alongside the tools for automation and analysis scripts to calculate and measure important quantites for sensor characterisation.

11.3 Analysis experience enhancements

The Analysis Productions service is the officially recommended and centrally maintained service for creating n-tuples within LHCb, which processed on average $\approx 25\,\mathrm{GB\,s^{-1}}$ of LHCb data in the latter half of 2024 (with frequent spikes to $50\,\mathrm{GB\,s^{-1}}$ to $180\,\mathrm{GB\,s^{-1}}$). In November 2024, 158 requests were submitted corresponding to 2,500 created samples [98]. It has been extended with a new web application, LbAPWeb, to support LHCb analysts with the creation of their n-tuples, the management of them after their creation, and improved discoverability of other n-tuples within the collaboration that have been created using this service. In addition to achieving these primary aims, it has enabled other services such as LbMCSubmit to provide similar features, and also led to increased awareness of the importance of monitoring resource metrics when testing and running data processing workflows at scale.

The web application and the services underpinning it have enabled a wealth of developments in LHCb Data Processing and Analysis activities to date, resulting in significant overall time savings and a more efficient use of resources (both human and computing). Extending the services to facilitate the smooth operation of LHCb's activities on the grid and in managing the wealth of n-tuples that will be created by analysts during Run 3, will likely bring significant benefits to collaborators. This may take the form of extensions in automated testing and feedback to users, including feedback specific to the application being used. Additionally, extending the capabilities of the system to support custom analysis steps (filtering, transformation of n-tuples) in addition to currently supported LHCb applications for creating n-tuples. It would also be beneficial to improve the indexing and search capabilities of the service to facilitate a more sophisticated search functionality.

Additionally, improved integrations with the DIRAC service, which is typically not interacted with by the user, could provide a helpful bridge to additional information about stored samples, the provenance information (i.e. what stages of processing were needed to produce the sample, and what was the input data),

among other valuable metadata — which the user would typically not have easy access to. Analysis Productions and connected services will continue to mature over the course of Run 3 and beyond.

Appendix A

Variable glossary

IP: Impact Parameter

The shortest distance between (usually) the primary vertex (PV) and vertex of a given particle.

FD: Flight Distance

Refers to the origin or primary vertex to decay vertex distance.

χ^2_{IP} **(IPCHI2): Impact Parameter** χ^2

A measure of how likely it is that a given particle originated from (usually) the primary vertex (PV).

$\chi^2_{MINIP}(PVs)$**: Minimum Impact Parameter** χ^2 **with respect to all PVs**

The minimum value within the container of χ^2_{IP} values calculated with respect to all primary vertices (PVs).

χ^2_{FD} **(FDCHI2): Flight Distance** χ^2

A measure of how well the given particle decay vertex can be separated from a vertex (usually primary or origin).

$\chi^2_{endvertex}$**: End Vertex** χ^2

The composite particle vertex fit χ^2.

(BPV)DIRA: Direction Angle

The cosine of the angle between the particle momentum and direction-of-flight vector from the Best Primary Vertex (BPV) to the particle decay vertex.

ProbNN variables

Output from ProbNN PID multivariate algorithm used in LHCb [29].

ProbNNghost: Track Ghost Probability

Interpreted as the track ghost probability.

MC15TuneV1_ProbNN

ProbNN variables with the `MC15TuneV1` PIDCalib configuration

BPVLTIME: Lifetime

The proper lifetime of the particle with respect to the BPV.

TRACK_CloneDist

A variable used to remove a small number of candidates with clone tracks from the sample.

DOCA: Distance of Closest Approach

Refers to the Distance of Closest Approach between two particle tracks.

ETA: Pseudorapidity η

Refers to the pseudorapidity η component of the momentum vector expressed in coordinates of (p_T, η, ϕ).

isMuon: Muon ID

Refers to whether there was any information in the Muon system corresponding to a the track.

Appendix B

Stage III MVA input variables in separate WS channels

Histogram density plots were created to show any differences between the wrong sign channels used as proxy for background when training the Stage III MVA in Section 4.5. See Figures B.2 and B.3. It was found that the wrong sign channels when regarding each variable separately were very similar. In Figure B.1, the shapes in ΔM can be compared between all channels before they were used to train the stage III MVA (this plot was created after unblinding as a cross-check).

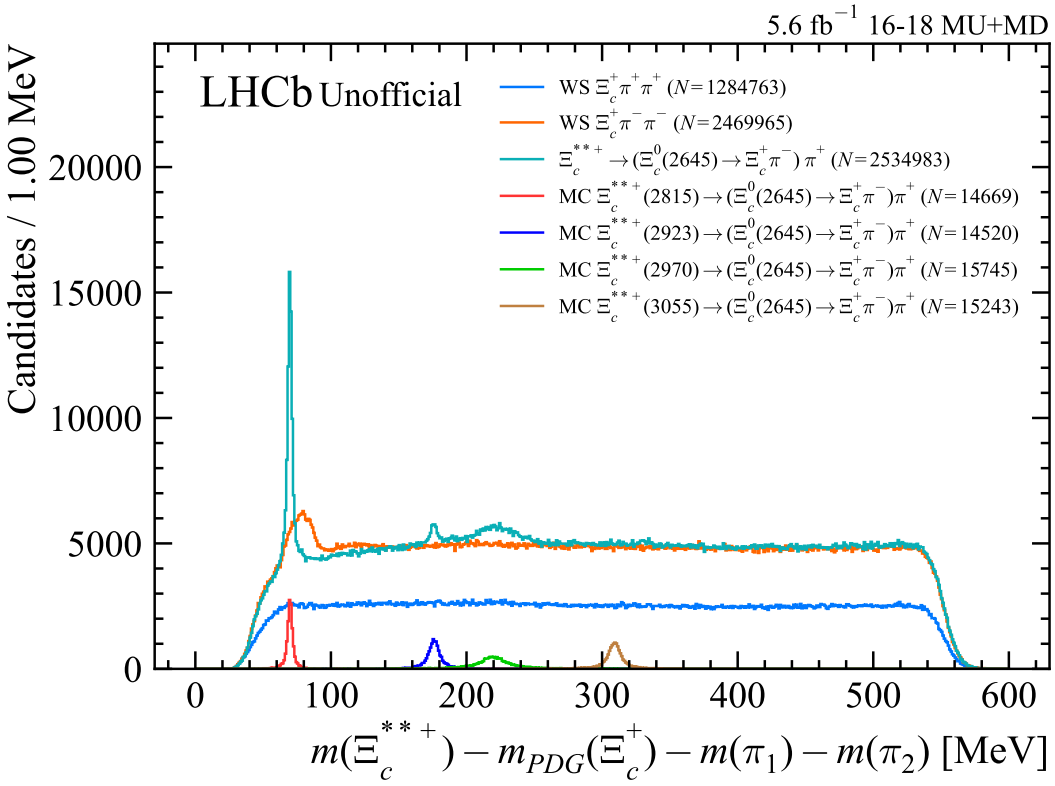

Figure B.1 Wrong-Sign vs. Right-Sign vs. Monte Carlo samples looking in ΔM having applied stage II selections and stage III preselections only.

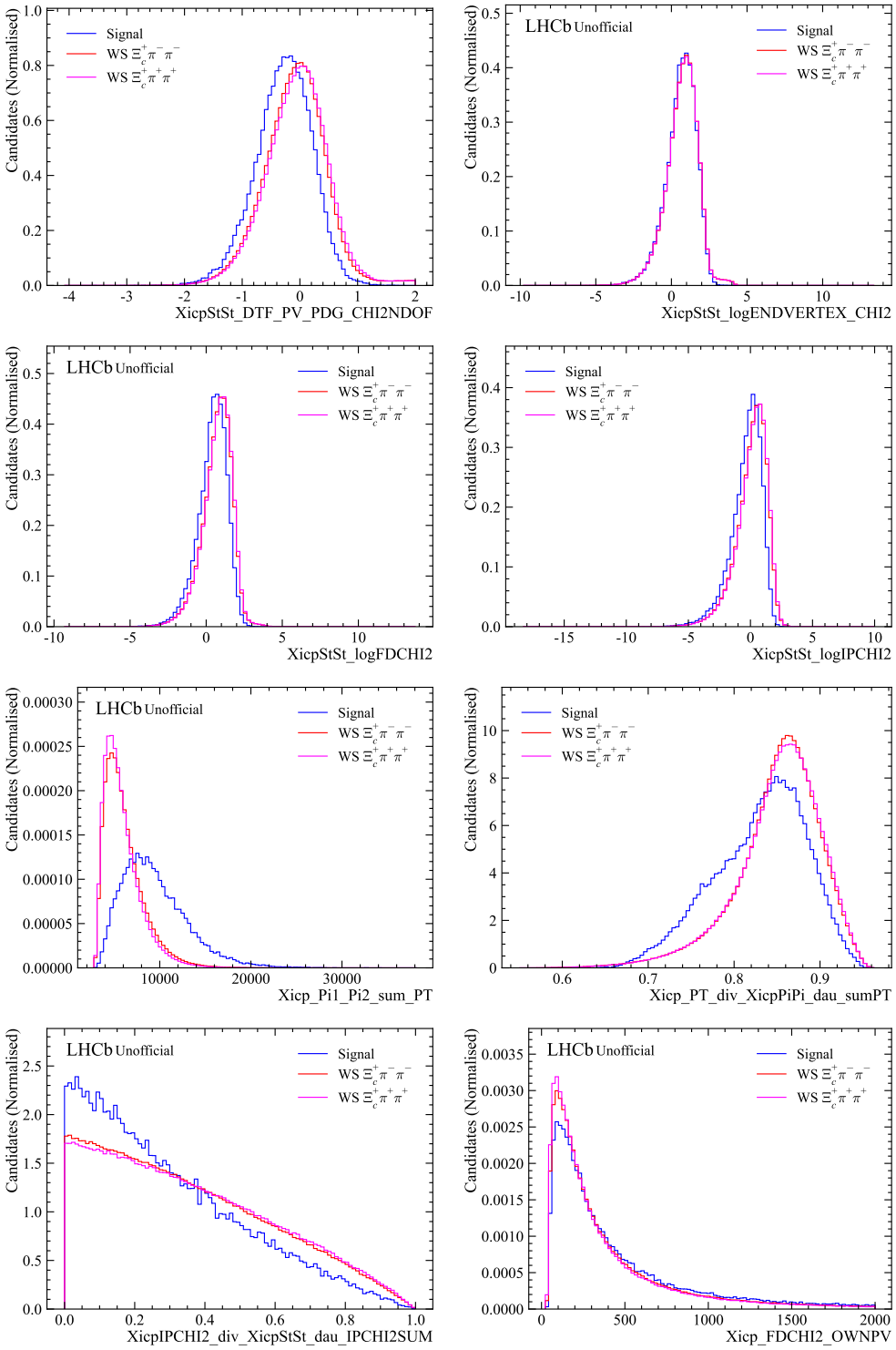

Figure B.2 Ξ_c^{**+} MVA input variables with separate histograms created for each wrong sign channel (Part 1)

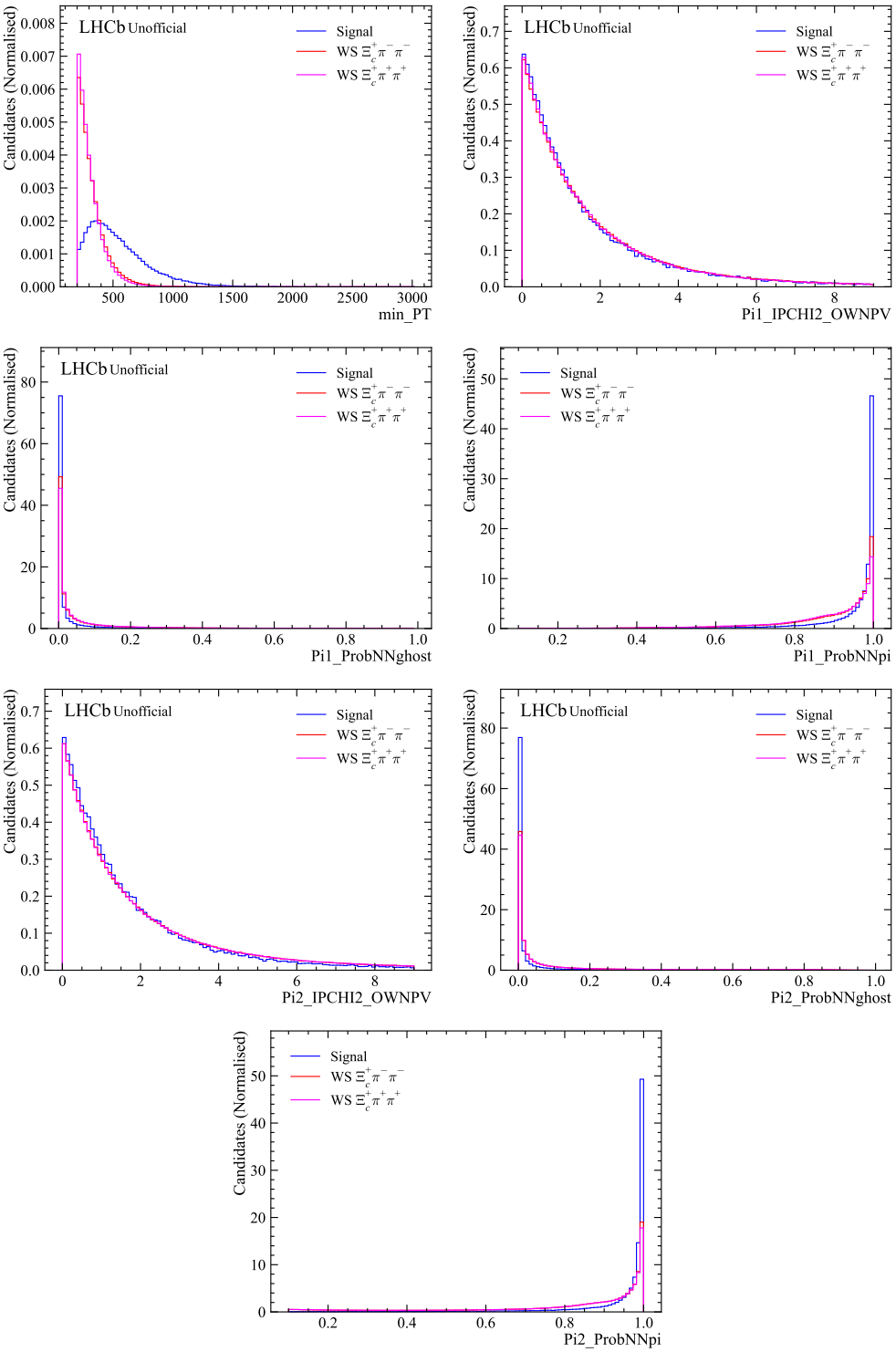

Figure B.3 Ξ_c^{**+} MVA input variables (Part 2)

Glossary

MagDown Refers to collision data taken while the LHCb magnet polarity was Down. 44, 122

MagUp Refers to collision data taken while the LHCb magnet polarity was Up. 44, 51, 52, 122

Area Under Curve (AUC) Defined as the integrated area under the ROC curve. 75

Direction Angle (DIRA) The angle between the 4-momenta sum of a composite particle's decay products and a line starting at the primary vertex pointing towards the decay vertex 48

Receiver operating characteristic (ROC) The resulting curve after plotting the signal efficiency (true positive rate) versus the background efficency (false positive rate). 75, 189

CVMFS CernVM FS. A system used to distribute software, typically used by WLCG grid nodes, but also useful for development and using software while avoiding the manual steps needed to install it. 45

Impact Parameter (IP) The shortest distance between (usually) the primary vertex (PV) and vertex of a given particle. 25

Interaction Point The crossing point of LHC proton bunches. Where proton bunches collide. 18, 19, 23

LXPLUS A remote logon (SSH) computing cluster provided by CERN to users holding a CERN computing account 160

n-tuple Datasets of a typically rectangular record-based structure, created usually from a production, usually stored in the ROOT format. In LHCb,

each record typically contains information about a reconstructed candidate, with variables describing some physics of interest to an analyst xix, xx, 4, 44, 45, 47, 49, 50, 54, 159–163, 165, 168, 169, 181, 190

occupancy A measure of the proportion of detector channels which have hits, averaged over events. A detector with regions of high occupancy may cause issues in downstream pattern recognition and reconstruction algorithms 23, 30, 133

production In the context of distributed computing in LHCb, describes a workflow (usually defined and run centrally) that produces some output data, such as an n-tuple. 160–163, 166–170, 189

selection line A group of selection requirements targeting a specific decay (exclusive selection) or a range of them (inclusive selection). When a selection line returns a positive decision i.e. to select the targeted decay(s), it is additionally possible to specify which parts of the event to save (Run 3). 24, 25, 44, 47

webhook An API request made to a specified URL, configured to be sent upon meeting some criteria e.g. an event occurring, such as a user pushing a commit to GIT repository 164

Acronyms

API Application Programming Interface 167

AUC Area Under Curve 75, 189

BPV Best Primary Vertex 183, 184

CI continuous integration 164, 166, 167

CI/CD Continuous Integration and Deployment 167

DAC Digital to Analogue Converter xv, xvi, 138, 149–151, 154, 156

DAQ Data Acquisition 135, 137, 138, 140, 147

DIRA Direction Angle 48, 189

DPA Data Processing & Analysis 159, 161

DUT Device Under Test xv, 135–141, 146–148, 152, 155, 158

HL-LHC High-Luminosity Large Hadron Collider 27, 30

HV-CMOS High Voltage Complimentary Metal-Oxide-Semiconductor 35, 133, 134

HV-MAPS High Voltage Monolithic Active Pixel Sensor 30, 32, 33, 35, 133, 134, 155

LHC Large Hadron Collider 3, 15, 16, 19, 27, 159, 161, 189, 192

MVA multivariate analysis xiii, xiv, xvii, 40, 41, 43, 49, 51, 54, 55, 57, 58, 61, 62, 64, 65, 69, 73, 74, 76–79, 83, 120

NPM Node Package Manager 167

PFoM Punzi Figure of Merit 41, 73–75, 79

PS Proton Synchrotron 16, 17

PV Primary Vertex 50, 73, 97

RBW Relativistic Breit–Wigner 66, 97, 103, 114

REST Representational State Transfer 164, 166, 167

RICH Ring Imaging Cherenkov ix, 21, 27

ROC Receiver operating characteristic 75, 189

SciFi Scintillating Fibre 27, 28, 30, 32, 133

SPA Single-Page Application 167

SPS Super Proton Synchrotron 16, 17

SSO Single Sign-On 165, 167

TLU Trigger Logic Unit 137, 157

ToA Time of Arrival xvi, 139, 152–154

ToT Time over Threshold xv, 144, 145

VELO Vertex Locator ix, 18, 19

WG Working Group 160–162

WLCG Worldwide LHC Computing Grid 159, 160, 189

Bibliography

[1] Particle Data Group Collaboration, S. Navas *et al.*, *Review of particle physics*, Phys. Rev. D **110** (2024) 030001.

[2] M. Thomson, *Modern particle physics*, Cambridge University Press, Cambridge, 2013.

[3] LHCb collaboration, R. Aaij *et al.*, *First observation of a doubly charged tetraquark candidate and its neutral partner*, Phys. Rev. Lett. **131** (2023) 041902, arXiv:2212.02716.

[4] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi p K^-$ decays*, Phys. Rev. Lett. **115** (2015) 072001, arXiv:1507.03414.

[5] LHCb collaboration, R. Aaij *et al.*, *Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$*, Phys. Rev. Lett. **122** (2019) 222001, arXiv:1904.03947.

[6] LHCb collaboration, R. Aaij *et al.*, *Observation of a $J/\psi\Lambda$ resonance consistent with a strange pentaquark candidate in $B^- \to J/\psi\Lambda\overline{p}$ decays*, Phys. Rev. Lett. **131** (2023) 031901, arXiv:2210.10346.

[7] Particle Data Group, *pdgLive: Baryon Summary Table*, https://pdglive.lbl.gov/ParticleGroup.action?sumtab&type=B. Accessed: 18/08/2024.

[8] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964) 214.

[9] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1* CERN-TH-401, CERN, Geneva, 1964.

[10] B. R. Martin and G. Shaw, *Particle Physics*, Manchester Physics Series, Wiley, 2017.

[11] M. Padmanath and N. Mathur, *Charmed baryons on the lattice*, in *7th International Workshop on Charm Physics*, 2015, arXiv:1508.07168.

[12] LHCb collaboration, R. Aaij *et al.*, *Observation of five new narrow Ω_c^0 states decaying to $\Xi_c^+ K^-$*, Phys. Rev. Lett. **118** (2017) 182001, arXiv:1703.04639.

[13] Belle collaboration, J. Yelton *et al.*, *Observation of Excited Ω_c Charmed Baryons in e^+e^- Collisions*, Phys. Rev. D **97** (2018) 051102, arXiv:1711.07927.

[14] H. Bahtiyar *et al.*, *Charmed baryon spectrum from lattice QCD near the physical point*, Phys. Rev. D **102** (2020) 054513, arXiv:2004.08999.

[15] H. Georgi, *An effective field theory for heavy quarks at low energies*, Physics Letters B **240** (1990) 447.

[16] E. Eichten and B. Hill, *An effective field theory for the calculation of matrix elements involving heavy quarks*, Physics Letters B **234** (1990) 511.

[17] H.-M. Yang and H.-X. Chen, *2p-wave charmed baryons from qcd sum rules*, Phys. Rev. D **109** (2024) 036032.

[18] E. Ortiz-Pacheco and R. Bijker, *Masses and radiative decay widths of S- and P-wave singly, doubly, and triply heavy charm and bottom baryons*, Phys. Rev. D **108** (2023) 054014, arXiv:2307.04939.

[19] LHCb collaboration, R. Aaij *et al.*, *Observation of new Ξ_c^0 baryons decaying to $\Lambda_c^+ K^-$*, Phys. Rev. Lett. **124** (2020) 222001, arXiv:2003.13649.

[20] LHCb collaboration, R. Aaij *et al.*, *Study of the $B^+ \to \Lambda_c^+ \overline{\Lambda}_c^- K^+$ decay*, Phys. Rev. **D108** (2023) 012020, arXiv:2211.00812.

[21] LHCb collaboration, R. Aaij *et al.*, *Observation of excited Ω_c^0 baryons in $\Omega_b^- \to {\Xi_c}^+ K^- \pi^+$ decays*, Phys. Rev. **D104** (2021) L091102, arXiv:2107.03419.

[22] Belle collaboration, T. Lesiak *et al.*, *Measurement of masses of the $\Xi_c(2645)$ and $\Xi_c(2815)$ baryons and observation of $\Xi_c(2980) \to \Xi_c(2645)\pi$*, Phys. Lett. B **665** (2008) 9, arXiv:0802.3968.

[23] Belle collaboration, T. J. Moon *et al.*, *First Determination of the Spin and Parity of a Charmed-Strange Baryon, $\Xi_c(2970)^+$*, arXiv:2007.14700.

[24] Belle collaboration, J. Yelton *et al.*, *Study of Excited Ξ_c States Decaying into Ξ_c^0 and Ξ_c^+ Baryons*, Phys. Rev. D **94** (2016) 052011, arXiv:1607.07123.

[25] Belle collaboration, Y. Kato *et al.*, *Studies of charmed strange baryons in the ΛD final state at Belle*, Phys. Rev. D **94** (2016) 032002, arXiv:1605.09103.

[26] BaBar collaboration, B. Aubert *et al.*, *A Study of Excited Charm-Strange Baryons with Evidence for new Baryons $\Xi_c(3055)^+$ and $\Xi_c(3123)^+$*, Phys. Rev. D **77** (2008) 012002, arXiv:0710.5763.

[27] Belle collaboration, Y. Kato *et al.*, *Search for doubly charmed baryons and study of charmed strange baryons at Belle*, Phys. Rev. D **89** (2014) 052003, arXiv:1312.1026.

[28] *LHCb: Technical Proposal*, CERN-LHCC-98-004, 1998.

[29] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[30] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[31] O. S. Brüning *et al.*, *LHC Design Report (Volume 1)*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004.

[32] M. Benedikt *et al.*, *LHC Design Report (Volume 3)*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004.

[33] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, CERN-GRAPHICS-2022-001, 2022, General Photo.

[34] R. Lindner, *LHCb detector layout*, Available: http://cds.cern.ch/record/1087860, 2008. LHCb Collection.

[35] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[36] LHCb RICH collaboration, A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, Nucl. Instrum. Meth. A **876** (2017) 221, arXiv:1703.08152.

[37] S. Borghi *et al.*, *First spatial alignment of the LHCb VELO and analysis of beam absorber collision data*, Nucl. Instrum. Meth. **A618** (2010) 108 LHCb-DP-2010-001.

[38] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[39] LHCb collaboration, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector* , LHCb-FIGURE-2020-016, 2020.

[40] LHCb collaboration, R. Aaij *et al.*, *The LHCb Upgrade I*, arXiv:2305.10515, to appear in JINST.

[41] LHCb collaboration, *LHCb Framework TDR for the LHCb Upgrade II Opportunities in flavour physics, and beyond, in the HL-LHC era*, CERN-LHCC-2021-012, 2022.

[42] LHCb collaboration, *LHCb Upgrade II Scoping Document*, CERN-LHCC-2024-010, 2024, Manuscript in preparation.

[43] H. Spieler, *Semiconductor Detector Systems*, vol. v.12 of *Semiconductor Science and Technology*, Oxford University Press, Oxford, 2005.

[44] I. Perić *et al.*, *High-voltage cmos active pixel sensor*, IEEE Journal of Solid-State Circuits **56** (2021) 2488.

[45] A. Schöning *et al.*, *MuPix and ATLASPix – Architectures and Results*, PoS **Vertex2019** (2020) 024, arXiv:2002.07253.

[46] I. Peric, *A novel monolithic pixelated particle detector implemented in high-voltage CMOS technology*, Nucl. Instrum. Meth. A **582** (2007) 876.

[47] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[48] A. Tsaregorodtsev *et al.*, *DIRAC3: The new generation of the LHCb grid software*, J. Phys. Conf. Ser. **219** (2010) 062029; R. Graciani Diaz *et al.*, *Belle-DIRAC setup for using Amazon Elastic Compute Cloud*, Journal of Grid Computing **9** (2011) 65.

[49] S. Borghi *et al.*, *LHCb Offline Analysis in Run 3 OATF recommendations*, LHCb-INT-2020-003, 2020.

[50] C. Burr *et al.*, *Data Processing & Analysis (DPA) Project*, LHCb-INT-2020-009, 2020.

[51] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of D meson mass differences*, JHEP **06** (2013) 065, arXiv:1304.6865.

[52] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820; T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[53] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[54] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[55] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199** (2016) 86, arXiv:1011.0937.

[56] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[57] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[58] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, *ReDecay: A novel approach to speed up the simulation at LHCb*, Eur. Phys. J. **C78** (2018) 1009, arXiv:1810.10362.

[59] F. Mölder *et al.*, *Sustainable data analysis with snakemake [version 2; peer review: 2 approved]*, F1000Research **10** (2021) .

[60] Ryunosuke O'Neil, *lhcb-charm / XicpStToXicpPiPi - GitLab*, https://gitlab.cern.ch/lhcb-charm/XicpStToXicpPiPi. Accessed 18/07/2024.

[61] M. Kucharczyk, P. Morawski, and M. Witek, *Primary Vertex Reconstruction at LHCb*, LHCb-PUB-2014-044, 2014.

[62] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, arXiv:physics/0306116.

[63] A. Hoecker *et al.*, *TMVA 4 — Toolkit for Multivariate Data Analysis with ROOT. Users Guide.*, arXiv:physics/0703039.

[64] F. Pedregosa *et al.*, *Scikit-learn: Machine learning in Python*, J. Machine Learning Res. **12** (2011) 2825, arXiv:1201.0490, and online at http://scikit-learn.org/stable/.

[65] L. Prokhorenkova *et al.*, *Catboost: unbiased boosting with categorical features*, 2017. doi: 10.48550/ARXIV.1706.09516.

[66] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, 2015. Software available from tensorflow.org.

[67] F. Chollet *et al.*, *Keras*, https://keras.io, 2015.

[68] A. L. Maas, A. Y. Hannun, and A. Y. Ng, *Rectifier nonlinearities improve neural network acoustic models*, in *Proceedings of the International Conference on Machine Learning*, **28**, (Atlanta, Georgia), 2013.

[69] M. Borisyak and N. Kazeev, *Machine Learning on data with sPlot background subtraction*, JINST **14** (2019) P08020, arXiv:1905.11719.

[70] A. Rogozhnikov, *Reweighting with boosted decision trees*, J. Phys. Conf. Ser. **762** (2016) 012036, arXiv:1608.05806, https://github.com/arogozhnikov/hep_ml.

[71] J. M. Blatt and V. F. Weisskopf, *Theoretical nuclear physics*, Springer, New York, 1952.

[72] G. A. Cowan, D. C. Craik, and M. D. Needham, *Rapidsim: An application for the fast simulation of heavy-quark hadron decays*, Computer Physics Communications **214** (2017) 239.

[73] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, arXiv:physics/0308063.

[74] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, arXiv:physics/0503191.

[75] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[76] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81, See also "ROOT" [software], Release v6.24/00, 29/09/2022.

[77] Belle collaboration, R. Chistov *et al.*, *Observation of new states decaying into* $\Lambda_c^+ K^- \pi^+$ *and* $\Lambda_c^+ K_S^0 \pi^-$, Phys. Rev. Lett. **97** (2006) 162001, arXiv:hep-ex/0606051.

[78] CLEO collaboration, J. P. Alexander *et al.*, *Evidence of new states decaying into* $\Xi_c^* \pi$, Phys. Rev. Lett. **83** (1999) 3390, arXiv:hep-ex/9906013.

[79] C.-D. Arslan *et al.*, *Test Beam with Irradiated ATLASPix3.1*, LHCb-INT-2023-007, 2023.

[80] D. Kim, *Timing Study and Optimization of ATLASPix3 a full-scale HV-MAPS Prototype*, Master's thesis, University of Heidelberg, 2020.

[81] Ryunosuke O'Neil, *HV-MAPS for the LHCb Upgrade-II Mighty Tracker*, https://indico.cern.ch/event/1140707/contributions/5086206/. The 31st International Workshop on Vertex Detectors. Accessed 18/07/2024.

[82] J. Hammerich, *Evaluation of an HV-CMOS Pixel Sensor Prototype for the LHCb Mighty Tracker and Development of a Novel Circuit for Improved Time Resolution*, PhD thesis, University of Liverpool, 2024.

[83] H. Schmitz, *Thesis in preparation*, PhD thesis, University of Bonn.

[84] R. Diener *et al.*, *The DESY II test beam facility*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **922** (2019) 265.

[85] Y. Liu *et al.*, *ADENIUM — A demonstrator for a next-generation beam telescope at DESY*, JINST **18** (2023) P06025, arXiv:2301.05909.

[86] Y. Liu *et al.*, *EUDAQ2—A flexible data acquisition software framework for common test beams*, JINST **14** (2019) P10033, arXiv:1907.10600.

[87] M. Mager, *ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **824** (2016) 434, Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[88] L. Huth, *A High Rate Testbeam Data Acquisition System and Characterization of High Voltage Monolithic Active Pixel Sensors"*., PhD thesis, University of Heidelberg, 2018.

[89] P. Baesso, D. Cussans, and J. Goldstein, *The AIDA-2020 TLU: a flexible trigger logic unit for test beam facilities*, JINST **14** (2019) P09019, arXiv:2005.00310.

[90] Y. Liu *et al.*, *EUDAQ2—A flexible data acquisition software framework for common test beams*, JINST **14** (2019) P10033, arXiv:1907.10600.

[91] D. Dannheim *et al.*, *Corryvreckan: A Modular 4D Track Reconstruction and Analysis Software for Test Beam Data*, JINST **16** (2021) P03008, arXiv:2011.12730.

[92] CLICdp collaboration, J. Kröger, S. Spannagel, and M. Williams, *User Manual for the Corryvreckan Test Beam Data Reconstruction Framework, Version 1.0*, arXiv:1912.00856.

[93] J. Kroeger, *Characterisation of a High-Voltage Monolithic Active Pixel Sensor Prototype for Future Collider Detectors*, PhD thesis, Heidelberg U., 2021, Presented 13 Oct 2021.

[94] J. Hammerich, *Analog Characterization and Time Resolution of a large scale HV-MAPS Prototype*, Master's thesis, University of Heidelberg, 2018.

[95] C. Kleinwort, *General Broken Lines as advanced track fitting method*, Nucl. Instrum. Meth. A **673** (2012) 107, arXiv:1201.4320.

[96] J. T. Mościcki *et al.*, *Ganga: A tool for computational-task management and easy access to grid resources*, Computer Physics Communications **180** (2009) 2303–2316.

[97] C. Burr, *Searching for rare charm decays, performing alignment studies and improving the analysis ecosystem in HEP*, PhD thesis, University of Manchester, 2019.

[98] C. Burr. Private Communication, 2024.

[99] Z. Mathe, *Feicim: A browser and analysis tool for distributed data in particle physics*, PhD thesis, U. Coll., Dublin, 2012, Presented 01 Jun 2012.

[100] C. Burr, B. Couturier, and R. O'Neil, *Facilitating the preservation of LHCb Analyses with APD*, EPJ Web Conf. **295** (2024) 08008.

[101] GitHub, *Node package manager*, https://www.npmjs.com.

[102] Vercel, *Next.js framework*, https://nextjs.org/.

[103] Meta, Inc. , *React framework*, https://react.dev/.

[104] JSROOT Authors, *JSRoot*, Available: https://root.cern/js/. Accessed: 25/08/2024.

[105] G. A. Stewart and A. S. Mete, *prmon: process monitor*, 2022. doi: 10.5281/zenodo.6606168.