University of Glasgow

Particle Physics Experimental

Ph.D Thesis

Author:
Stephen Ogilvy

Supervisors:
Dr. Lars Eklund
Dr. Paul Soler

December 28, 2014

Abstract

LHCb is the dedicated heavy flavour experiment on the LHC accelerator ring. An analysis measuring the relative branching fractions of $\Lambda_c^+ \to phh'$ decays, where $hh' \in \{K^-\pi^+, K^-K^+, \pi^-\pi^+, \pi^-K^+\}$, is presented using $1024.8 \pm 35.9\,\mathrm{pb}^{-1}$ of pp collisions gathered during 2011. This includes a search for the unobserved decay mode $\Lambda_c^+ \to p\pi^-K^+$. LHCb's software-based detector alignment utilises a track χ^2 minimisation approach, which is susceptible to so-called "weak modes" of misalignments. A study of the effects on physics analyses of weak mode misalignment in the LHCb's vertex locator (VELO) is presented, with a strategy for their constraint. Charged hadron discrimination in LHCb is provided by two ring imaging Cherenkov detectors (RICH detectors). A data-driven study on the performance of the aerogel Cherenkov radiator is presented. Particle identification information is used to construct variables describing the likelihood of hadronic particle hypotheses of reconstructed tracks, on which cuts are placed in typical physics analyses. The preparation of new data samples allowing for the data-driven efficiency correction of particle identification cuts on proton tracks is described.

Contents

1 $\Lambda_c^+ \to phh'$ selection efficiencies and yield extractions **2**
1.1 Efficiencies and efficiency corrections 2
1.1.1 Overview 2
1.1.2 MC efficiencies and resonance modelling 4
1.1.3 Trigger efficiencies 7
1.1.4 Generator level efficiencies 8
1.1.5 Stripping efficiencies 13
1.1.6 PID efficiencies from a fully data-driven `PIDCalib` 21
1.1.7 BDT efficiency 28
1.1.8 Full selection efficiencies and summary 28
1.2 Yield extraction 30
1.2.1 Prompt 31
1.2.2 Semileptonic 34
1.2.3 Raw and adjusted yields 37

References **38**

Chapter 1

$\Lambda_c^+ \to phh'$ selection efficiencies and yield extractions

1.1 Efficiencies and efficiency corrections

1.1.1 Overview

The correct determination of efficiencies for each stage of the selection is vital to the analysis. Starting from the Λ_c production, we define the selection and associated efficiencies as follows:

$\epsilon_{acc|gen} = \frac{N_{acc}}{N_{gen}}$
: This is the fraction of candidates which decay with all daughters produced inside the detector acceptance.

$\epsilon_{reco|acc} = \frac{N_{reco}}{N_{acc}}$
: The fraction of candidates generated within the detector acceptance which are fully reconstructible.

$\epsilon_{trig|reco} = \frac{N_{trig}}{N_{reco}}$
: The fraction of reconstructible candidates which pass the trigger requirements.

$\epsilon_{strip|trig} = \frac{N_{strip}}{N_{trig}}$
: The fraction of triggered candidates which are then stripped, pass the DTF convergence requirement and also pass our kinematic vetoes required for the `PIDCalib` calibration, excluding any PID selection in the stripping selection.

$\epsilon_{PID|strip} = \frac{N_{PID}}{N_{strip}}$

The fraction of events passing the full PID selection w.r.t the number passing the rest of the stripping selection.

$\epsilon_{BDT|PID} = \frac{N_{BDT}}{N_{PID}}$

The fraction of events passing the BDT selection w.r.t. the number passing the full PID selection.

We now make a number of assumptions to simplify this chain. The first is that the trigger selection in the prompt analysis is independent of the stripping selection - a strong assumption given the TIS trigger selection. We also assume the trigger selection in the SL analysis is also independent of the stripping selection, which is a weaker assumption and so is explicitly verified. This is a necessary assumption as when the TOS trigger requirements are applied on the SL simulation very few events survive, and tje stripping efficiency calculations become very statistically limited.

These assumptions let us remove the intermediary trigger step from the acceptance, reconstruction and stripping steps and we may now consider the following efficiency factor:

$\epsilon_{strip|acc} = \frac{N_{strip}}{N_{acc}}$

This is the fraction of Λ_c candidates decaying with all daughters in the detector accceptance which are reconstructed and stripped.

Now the final expression for the per-mode adjusted yield becomes:

$$\mathrm{M} = \frac{\mathrm{N}}{\epsilon_{\mathrm{trig|reco}} \times \epsilon_{\mathrm{acc|gen}} \times \epsilon_{\mathrm{strip|acc}} \times \epsilon_{\mathrm{PID|strip}} \times \epsilon_{\mathrm{BDT|PID}}} \tag{1.1}$$

where the terms are:

N - The extracted raw yield for the mode.

$\epsilon_{\mathrm{trig|reco}}$ - The per-mode trigger efficiency w.r.t. the reconstruction. In the prompt analysis we use a TIS chain, so the ratio of these efficiencies is expected to be 1. In the SL analysis the efficiencies are instead evaluated with simulated data. This is described in Section 1.1.3.

$\epsilon_{\mathrm{acc|gen}}$ - The per-mode acceptance of the detector geometry. This is calculated in both analyses using generator-level simulation, as described in Section 1.1.4.

$\epsilon_{\text{strip|acc}}$ - The per-mode stripping efficiency w.r.t. the candidates falling in the detector acceptance. This efficiency must be calculated from simulation. The PID cuts in the stripping are removed from the selection of simulated candidates, and the PID cut efficiency correction is treated in a separate data-driven efficiency. The no-PID stripping efficiency calculation is detailed in Section 1.1.5.

$\epsilon_{\text{PID|strip}}$ - The per-mode PID selection efficiency w.r.t. the stripped candidates. This is calculated using a novel, entirely data-driven variation of the `PIDCalib` method, and described in Section 1.1.6.

$\epsilon_{\text{BDT|PID}}$ - The per-mode BDT efficiency w.r.t. the rest of the selection. The BDT selection is only applied in the selection of prompt $\Lambda_c^+ \to p\pi^- K^+$ events, so is equal to 1 for all branching fraction ratios not involving the DCS mode. Its efficiency is extracted by fitting the signal yields of the Cabibbo-favoured distribution before and after the application of the BDT. Cross checks are performed to verify that the efficiency of the BDT selection will be uniform in the $\Lambda_c^+ \to pK^-\pi^+$ and $\Lambda_c^+ \to p\pi^- K^+$. This is described in Section 1.1.7.

The full selection efficiencies for each analysis on a per-mode basis are then given in Section 1.1.8.

1.1.2 MC efficiencies and resonance modelling

It is necessary to take efficiency corrections for several of the described stages from simulated data. Poor modelling of any variable used in the selection has the potential to bias the calculated efficiency for any given stage of the selection. The simulated data used in the prompt analysis is generated with a phase-space distribution and ignores the complex resonant structure of $\Lambda_c^+ \to phh'$ decays, while the semileptonic simulated data uses a pseudo-resonance model which attempts to account for the dominant resonant contributions (as was more fully outlined in Section **??**).

As described previously within this document in Section **??**, three-body baryonic decays to hadrons can be parameterised by 5 variables, shown in Figure 1.1. These variables are:

$M(ph_1)(h_1h_2)$ - The invariant mass of the proton and opposite sign meson, and the invariant mass of the meson pair. Decays through intermediate resonances (such as $K^*, \Lambda(1520), \phi, f^0(980)$ etc.) will result in local enhancements in these quantities not present in a phase-space generated distribution.

$\cos\theta_p$ - In Λ_c rest frame, the angle between the proton momentum vector and the polarisation axis of the Λ_c.

$\cos\phi_p$ - In Λ_c rest frame, the angle between the component of proton momentum perpendicular to the Λ_c polarisation and the direction of the lab frame Λ_c momentum vector.

$\phi_{h_1h_2}$ - In Λ_c rest frame, the angle between the plane containing the proton momentum vector and the Λ_c polarisation vector, and the plane containing the two meson momentum vectors.

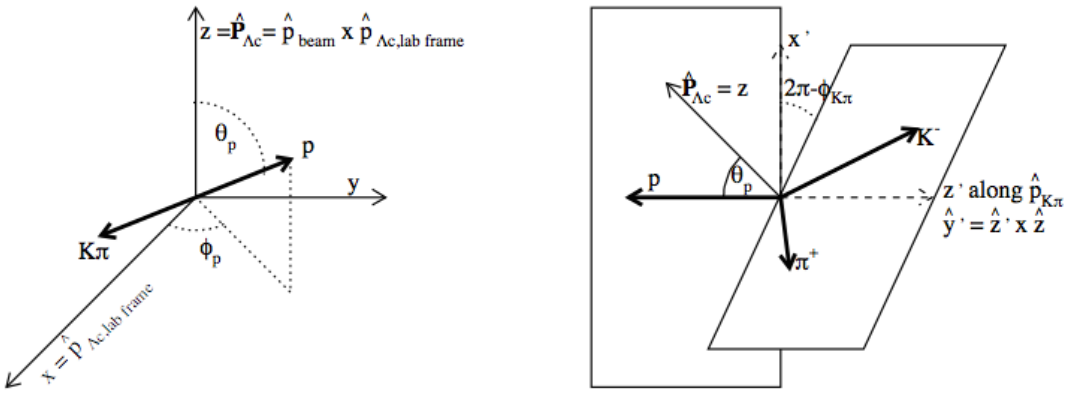
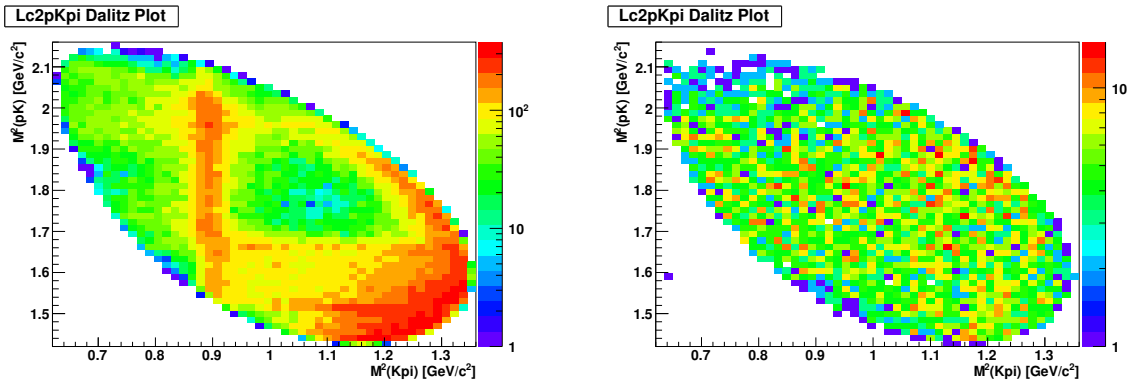


Figure 1.1: The angular variables characterising the Λ_c resonance structure. From [1].

The simulation used in this analysis does not properly account for the resonance structures in Λ_c decays. The data and MC distributions of the above variables are therefore expected to disagree to varying extents which are difficult to estimate given the poorly understood nature of $\Lambda_c^+ \to phh'$ decays. The data and MC populations of promptly selected $\Lambda_c^+ \to pK^-\pi^+$ in bins of the invariant mass resonant variables are shown in Figure 1.2 to illustrate such disagreements. It is therefore necessary to investigate each stage of the selection whereby the efficiency is taken from simulation for the following:

- Variations in acceptance across each of the 5 resonant quantities.
- Disagreements in the data/MC distributions in each of the 5 resonant quantities.

If both of these criteria are true for any variable at any given stage of the selection, phase-space averaged efficiencies from simulation may not be naively utilised for that stage of the selection.



(a) (b)

Figure 1.2: Comparison of the invariant mass of the systems of charge-opposite daughter pairs in promptly selected $\Lambda_c^+ \to pK^-\pi^+$ decays, in stripped data (a) and stripped simulation (b).

1.1.3 Trigger efficiencies

This efficiency is defined as the fraction of reconstructed candidates which pass the trigger selection. The requirement that the Λ_c decay is independent of the event triggering in the selection of prompt candidates is expected to result in efficiency ratios in the modes of 1 to first order. These efficiencies are evaluated using simulated data, however due to the low TIS efficiency of the MC this calculation is imprecise. An algorithm is run over the simulation which uses the truth information to record all reconstructible Λ_c decays. The fraction of these candidates which pass the trigger selection is equivalent to the trigger efficiency, and was investigated in each decay mode.

In the semileptonic selection a trigger requirement is utilised such that Λ_b^0 decay itself must trigger the event. While at `L0` and `Hlt1` the μ in the Λ_b^0 decay must be `TOS`, at `Hlt2` the full decay has a `TOS` requirement. As such the differing Λ_c daughter kinematics will result in different `Hlt2` efficiencies. These are calculated using simulated data via the same procedure utilised in the prompt trigger checks.

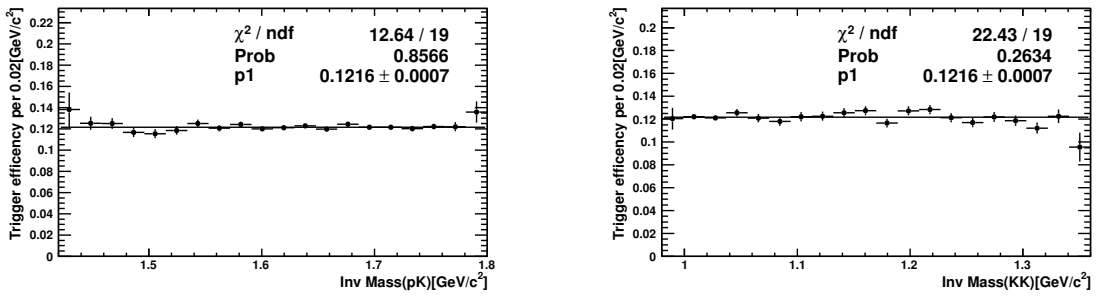
Acceptance variations in resonant quantities

The trigger efficiencies are calculated with simulated data. Given the complexity of the trigger selection algorithms it was considered possible that the SL candidate trigger acceptance may vary with respect to the poorly modelled resonant quantities. This was investigated by evaluating the trigger efficiencies in bins of the resonant variables. The resonant quantities proved to be broadly uncorrelated with the trigger efficiency for all decay modes. Some example acceptances are given in Figure 1.3 in the invariant mass systems m_{pK} and m_{KK} in semileptonically produced $\Lambda_c^+ \to pK^-K^+$.

Average full trigger efficiencies

The final results are taken with an average of the simulated efficiencies from the different magnet polarities weighted with the luminosities of the real data magnet polarities, with equal weighting given to Λ_c^+ and Λ_c^-. The final average trigger efficiencies are given in Table 1.1. The subscript "Reco" refers to the condition of being reconstructible, while "Trig" refers to the subset of these passing the trigger selection.

We emphasise that in the prompt analysis, these are not used in the final calculations of the relative branching fractions and only serve as a cross check that the TIS efficiencies of the different modes are uniform. The errors provided are the statistical uncertainty arising from limited MC statistics.



(a) (b)

Figure 1.3: The trigger acceptance in semileptonically produced $\Lambda_c^+ \to pK^-K^+$ with respect to resonant variables, demonstrating that the acceptance is broadly uncorrelated with the invariant masses m_{pK} (a) and m_{KK} (b).

Analysis	Mode	$\epsilon_{\text{Trig\|Reco}}$ [%]
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	0.47 ± 0.01
	$\Lambda_c^+ \to pK^-K^+$	0.47 ± 0.01
	$\Lambda_c^+ \to p\pi^-\pi^+$	0.44 ± 0.01
	$\Lambda_c^+ \to p\pi^-K^+$	$x.xx \pm x.xx$
SL	$\Lambda_c^+ \to pK^-\pi^+$	14.21 ± 0.03
	$\Lambda_c^+ \to pK^-K^+$	12.61 ± 0.07
	$\Lambda_c^+ \to p\pi^-\pi^+$	14.67 ± 0.08
	$\Lambda_c^+ \to p\pi^-K^+$	$x.xx \pm x.xx$

Table 1.1: The final per-mode phase-space averaged trigger efficiencies for the full trigger selection.

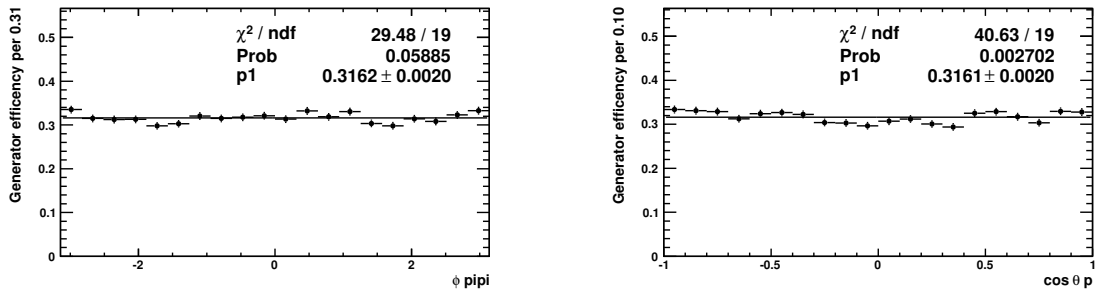
1.1.4 Generator level efficiencies

The simulation utilised in the analyses employ one of two generator-level cuts. These are:

`DaughtersInLHCb` - This requires all charged daughters in the generated decay to be produced in the range $0.01\,\text{rad} < \theta_{\text{charged}} < 0.4\,\text{rad}$ of the z-axis and all neutral daughters to be produced in the range $0.005\,\text{rad} < \theta_{\text{neutral}} < 0.4\,\text{rad}$ of the z-axis.

`LHCbAcceptance` - This requires the head particle/mother of the specified signal decay to be produced in the range $0 < \theta_{signal} < 0.4\,\text{rad}$ of the z-axis.

All generated MC samples utilise the `DaughtersInLHCb` cut, with the exception of the semileptonic CF $\Lambda_c^+ \to pK^-\pi^+$ which utilises the `LHCbAcceptance` cut. Because of the differing daughter kinematics in the $\Lambda_c^+ \to phh'$ modes it is natural to expect that the



(a) (b)

Figure 1.4: Generator cut acceptance in semileptonically produced $\Lambda_c^+ \to p\pi^-\pi^+$ with respect to θ_{hh} (a) and $\cos\phi_p$ (b).

`DaughtersInLHCb` efficiency will not cancel out when a ratio of efficiencies is taken. As such an efficiency correction must be applied.

Acceptance variations in resonant quantities

The generator level efficiencies have a weak dependence on the daughter kinematics. Therefore at this stage of the selection there exists potential for the existance of acceptance variations w.r.t. resonant quantities. This was investigated with the use of generator-level simulation with no acceptance cuts employed. 100k events were generated for each permutation of mode and polarity. By examining the decay kinematics it is determined for each simulated decay whether or not the two generator level requirement is met.

Histograms were then constructed by binning the samples in bins of the resonant variables and counting the candidates in each bin passing the cuts to construct efficiencies. We fit a first order polynomial with a gradient fixed at zero to the distribution to check for acceptance flatness. In these checks the fits to the distributions yielded acceptable p-values and reduced χ^2 values. As such it was concluded that there was no significant variation in acceptance at this stage of the selection, and that phase-space averaged efficiencies were appropriate. Example tests for the promptly produced $\Lambda_c^+ \to p\pi^-\pi^+$ magnet up efficiency distributions are shown in Figure 1.4.

Calculated efficiencies

Given the generator-level acceptances do not appear correlated to the resonant variables, we can confidently use phase-space averaged efficiencies. In the semileptonic analysis the logs of the simulation production are used to record the generator requirement efficiencies,

which are cross checked against the less precise checks calculated with the smaller samples generated with no generator requirements. In the prompt analysis the simulation has a significant non-prompt component which we remove from the samples after production. Therefore in the final results we use our smaller generator-level only samples where the Λ_c ancestry is verified as prompt. We calculate efficiencies by magnet polarity and by Λ_c charge. The prompt efficiencies are given in Table 1.2, and the semileptonic efficiencies given in Table 1.3, for both the logfile efficiencies from the sample generation and from our own generator-level samples. While we provide both, the latter are used in the final results for the prompt and the former for the semileptonic.

We note that in the prompt there is a clear trend for the generator-level efficiency of Λ_c^- to be 3 – 5 % higher depending on the decay mode than the corresponding Λ_c^+. This is attributable to the differing hadronisation of Λ_c^+ and Λ_c^- given the proton – proton collision and underlying event. The majority of prompt Λ_c production comes from $c\bar{c}$ production, which in turn predominately comes from gg interactions. As such, independently of the underlying event we would expect symmetric hardonisation of Λ_c^+ and Λ_c^-. However, the c quark can hadronise with the highly boosted proton remnants in the underlying event to form a boosted Λ_c^+, while the $\bar{c}$ quark cannot form a Λ_c^- similarly. This results in a higher proportion of Λ_c^+ being produced at high η and being lost in the beampipe, outwith the generator level acceptance. This is demonstrated in Figure 1.5 for the MagDown prompt $\Lambda_c^+ \to pK^-\pi^+$ generator-level sample. The excess in Λ_c^+ in the high η region (at around $\eta = 9$) demonstrates this effect, and it is fully expected given the LHC's production environment.

We provide the average efficiencies on a per-mode basis used in the final results. These are taken from an average of the simulated efficiencies from the different magnet polarities weighted with the luminosities of the real data magnet polarities, with equal weighting given to Λ_c^+ and Λ_c^-.

Polarity	Part/Antipart	Decay mode	Logfile Eff	True Prompt Eff
MagUp	Particle	$\Lambda_c^+ \to pK^-\pi^+$	0.208 ± 0.001	0.211 ± 0.002
		$\Lambda_c^+ \to pK^-K^+$	0.234 ± 0.001	0.237 ± 0.002
		$\Lambda_c^+ \to p\pi^-\pi^+$	0.194 ± 0.001	0.199 ± 0.003
		$\Lambda_c^+ \to p\pi^-K^+$	0.211 ± 0.001	-
	Antiparticle	$\Lambda_c^- \to \overline{p}K^+\pi^-$	0.219 ± 0.001	0.218 ± 0.002
		$\Lambda_c^- \to \overline{p}\pi^+\pi^-$	0.242 ± 0.001	0.244 ± 0.002
		$\Lambda_c^- \to \overline{p}K^+K^-$	0.200 ± 0.001	0.201 ± 0.004
		$\Lambda_c^- \to \overline{p}\pi^+K^-$	0.217 ± 0.001	-
MagDown	Particle	$\Lambda_c^+ \to pK^-\pi^+$	0.210 ± 0.001	0.210 ± 0.002
		$\Lambda_c^+ \to pK^-K^+$	0.234 ± 0.001	0.235 ± 0.002
		$\Lambda_c^+ \to p\pi^-\pi^+$	0.193 ± 0.001	0.195 ± 0.002
		$\Lambda_c^+ \to p\pi^-K^+$	0.211 ± 0.001	0.210 ± 0.002
	Antiparticle	$\Lambda_c^- \to \overline{p}K^+\pi^-$	0.214 ± 0.001	0.220 ± 0.002
		$\Lambda_c^- \to \overline{p}\pi^+\pi^-$	0.243 ± 0.001	0.242 ± 0.002
		$\Lambda_c^- \to \overline{p}K^+K^-$	0.199 ± 0.001	0.203 ± 0.002
		$\Lambda_c^- \to \overline{p}\pi^+K^-$	0.217 ± 0.001	0.217 ± 0.002

Table 1.2: The prompt generator-level efficiencies both from the logfiles of the simulation production and a small sample of generator-level MC where the Λ_c is verified as being produced promptly. All efficiencies correspond to the `DaughtersInLHCb` requirement. The latter efficiencies correspond to only those candidates which are prompt, the former includes secondary contamination and is not used but provided for comparison.

Polarity	Part/Antipart	Decay mode	Logfile Eff	Calculated Eff
MagUp	Particle	$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu,\ \Lambda_c^- \to \overline{p}K^+\pi^-$	0.334 ± 0.001	0.339 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu,\ \Lambda_c^- \to \overline{p}K^+K^-$	0.182 ± 0.001	0.182 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu,\ \Lambda_c^- \to \overline{p}\pi^+\pi^-$	0.157 ± 0.001	0.158 ± 0.002
	Antiparticle	$\Lambda_b^0 \to \Lambda_c^-\mu^+\nu_\mu,\ \Lambda_c^- \to \overline{p}K^+\pi^-$	0.332 ± 0.001	0.331 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^-\mu^+\nu_\mu,\ \Lambda_c^- \to \overline{p}K^+K^-$	0.181 ± 0.001	0.182 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^-\mu^+\nu_\mu,\ \Lambda_c^- \to \overline{p}\pi^+\pi^-$	0.158 ± 0.001	0.159 ± 0.002
MagDown	Particle	$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu,\ \Lambda_c^- \to \overline{p}K^+\pi^-$	0.334 ± 0.001	0.337 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu,\ \Lambda_c^- \to \overline{p}K^+K^-$	0.183 ± 0.001	0.181 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu,\ \Lambda_c^- \to \overline{p}\pi^+\pi^-$	0.157 ± 0.001	-
	Antiparticle	$\Lambda_b^0 \to \Lambda_c^-\mu^+\nu_\mu,\ \Lambda_c^- \to \overline{p}K^+\pi^-$	0.333 ± 0.001	0.330 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^-\mu^+\nu_\mu,\ \Lambda_c^- \to \overline{p}K^+K^-$	0.181 ± 0.001	0.182 ± 0.002
		$\Lambda_b^0 \to \Lambda_c^-\mu^+\nu_\mu,\ \Lambda_c^- \to \overline{p}\pi^+\pi^-$	0.156 ± 0.001	-

Table 1.3: The semileptonic generator-level efficiencies both from the logfiles of the simulation production and a sample of generator-level MC. The logfile efficiencies are valid in this case and use higher statistics in their calculations so are used, the "Calculated Eff" serves merely as a cross check. All $\Lambda_c^+ \to pK^-\pi^+$ modes use the `LHCbAcceptance` cut, all other modes use the `DaughtersInLHCb` cut. Some cross check efficiencies are missing due to processing errors.

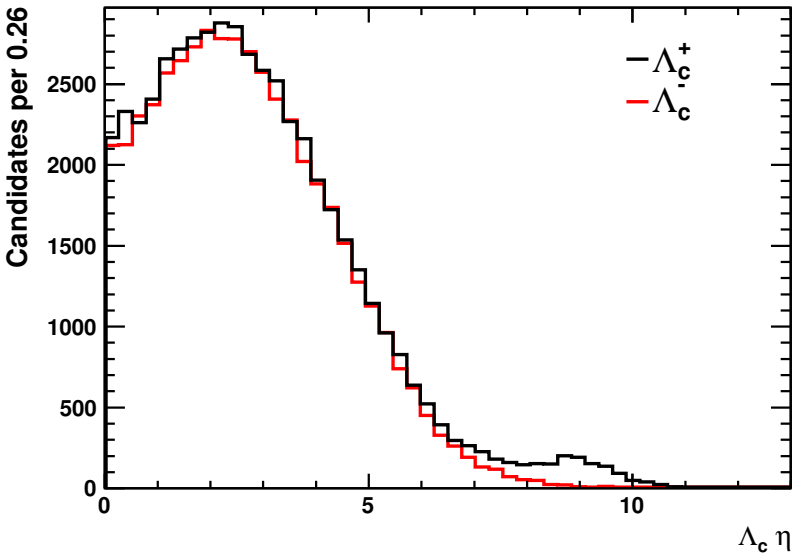


Figure 1.5: The η distributions of generator-level Λ_c^+ (black) and Λ_c^- (red). Distributions are taken from the generator-level samples of $\Lambda_c^+ \to pK^-\pi^+$ data. Shown is the excess in high η Λ_c^+ corresponding to c quarks which have hadronised with the proton remnants and fall outside the detector acceptance.

Stream	Mode	$\epsilon_{\text{Acc\|Gen}}$ [%]
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	21.27 ± 0.05
	$\Lambda_c^+ \to pK^-K^+$	23.84 ± 0.05
	$\Lambda_c^+ \to p\pi^-\pi^+$	19.66 ± 0.05
	$\Lambda_c^+ \to p\pi^-K^+$	21.40 ± 0.05
SL TIS/TOS	$\Lambda_c^+ \to pK^-\pi^+$	33.31 ± 0.07
	$\Lambda_c^+ \to pK^-K^+$	18.15 ± 0.04
	$\Lambda_c^+ \to p\pi^-\pi^+$	15.66 ± 0.04
	$\Lambda_c^+ \to p\pi^-K^+$	$xx.xx \pm x.xx$

Table 1.4: The average per-mode generator level efficiencies used in the final analysis. The subscript "Gen" refers to the condition of being generated, "Acc" refers to the condition of passing the generator level cut for that mode.

1.1.5 Stripping efficiencies

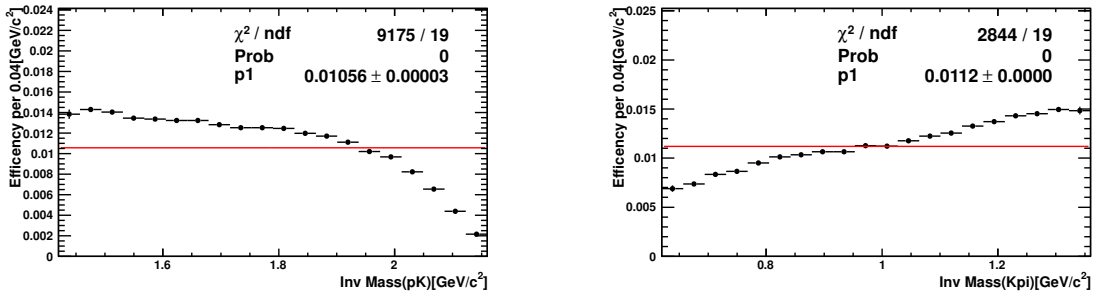
Overview

The stripping selections in prompt and semileptonic, as outlined in Section **??**, use a suite of kinematic cuts and PID information to reduce the combinatoric background. The stripping efficiency may be factorised into the efficiency of the PID cuts and the efficiency of the rest of the stripping selection. This is useful as the PID response is not well modelled in the LHCb simulation, and is better evaluated with the use of data-driven techniques.

The efficiency of the remainder of the stripping selection may be evaluated with simulated data. Further kinematic cuts on the daughter particles in the decay - specifically track p and η - are necessary for PID calibration (this was elaborated in Section **??**, and the full suite of kinematic cuts was previously given in Table **??**). As the real data candidates with daughters falling in the kinematically vetoed region do not have valid PID efficiencies, we cannot derive an efficiency correction in a data-driven way. Instead the efficiency of these cuts must be evaluated with simulated data. For convenience the evaluation of the efficiency of these kinematic cuts is performed in tandem with the no-PID stripping efficiency.

We factorise out the PID cuts by running a version of the stripping selection algorithm which has had any PID cuts removed on the simulated data. We also apply the kinematic vetoes and DTF convergence requirement on this dataset. The fraction of generated, truth-matched candidates which pass the no-PID stripping, DTF convergence requirement and kinematic vetoes is defined as the stripping efficiency. We repeat the point that the stripping efficiency in this analysis is defined as the efficiency of the no-PID stripping, DTF convergence requirement and kinematic vetoes together - we do not treat the latter separately as an "offline" efficiency as these cuts are necessary for the PID selection efficiency correction to be valid.

The selections in both prompt and semileptonic sculpt the distributions of signal kinematics. The effect this has on the stripping acceptance with respect to the resonant variables was investigated. Strong local variation in acceptance was observed in a number of variables, especially in bins of invariant masses of the daughter pairs. Such variation is displayed in Figure 1.6 for the invariant mass variables in prompt $\Lambda_c^+ \to pK^-\pi^+$.



(a) (b)

Figure 1.6: Stripping efficiency in bins of the invariant mass variables for semileptonically selected $\Lambda_c^+ \to pK^-\pi^+$. Unlike the checks for acceptance bias at the generator and trigger levels of the selection, strong variation is observed at the stripping level.

Local efficiencies and re-weighting procedure

The stripping efficiency bias w.r.t. the resonant variables necessitates a re-weighting of the simulation and of the efficiency for the stripping selection. This is achieved by dividing the simulation in multidimensional bins of the resonant variables. We count the generated candidates in each bin and the truth-matched candidates surviving the no-PID stripping selection and kinematic veto selection to acquire local efficiencies in the 5D space. We then use these local efficiencies to assign per-candidate efficiencies to the real data, extracting the adjusted yield M by summing over the bins:

$$M = \sum^{j} \frac{N_j}{\epsilon_j} \tag{1.2}$$

where N_j is the number of data candidates in bin j and ϵ_j is the efficiency in bin j. The ratio of the raw number of data candidates passing selection over the adjusted yield gives the re-weighted efficiency $\bar{\epsilon}$:

$$\bar{\epsilon} = \frac{N}{M} \tag{1.3}$$

In such a procedure the following systematic uncertainties arise:

1. The finite simulation statistics result in a binomial uncertainty on the bin efficiency. As we further subdivide our sample this uncertainty becomes larger in magnitude.

2. The bins in the schema have a finite width. As we know, the efficiency varies across the variable space. This means that for each bin in the schema, the average bin

efficiency is dependent on the distributions of the resonant variables within the bin. These are different between data and MC and so can result in biases if the binning is too coarse.

These are clearly opposing trends - to minimise the overall systematic contribution we must therefore find the optimum choice of binning.

Due to the low efficiency of both the prompt and semileptonic selections a low proportion of the generated candidates survive the stripping selection. The number of MC events which survive the selection are given in Table 1.5. This makes attaining a sufficiently fine-grained 5D binning problematic, and initial efforts to construct a full 5D binning schema yielded results with a high disagreement between the prompt and semileptonic measurements. As such we use a "reduced dimensionality" method to re-weight the efficiency within the 5D space only where necessary by discounting variables from the binning schema where the acceptance is flat or the data-MC agreement is good.

Mode	N_{Acc}	$N_{\text{Strip\|Acc}}$	$\epsilon_{\text{Strip\|Acc}}$ [%]
$\Lambda_c^+ \to pK^-\pi^+$	2308801	7225	$3.131 \pm 0.037 \times 10^{-3}$
$\Lambda_c^+ \to pK^-K^+$	2230984	5598	$2.510 \pm 0.033 \times 10^{-3}$
$\Lambda_c^+ \to p\pi^-\pi^+$	2305682	7945	$3.448 \pm 0.039 \times 10^{-3}$
$\Lambda_c^+ \to p\pi^-K^+$	-	-	-
$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu$, $\Lambda_c^+ \to pK^-\pi^+$	9585315	110290	$1.118 \pm 0.003 \times 10^{-2}$
$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu$, $\Lambda_c^+ \to pK^-K^+$	1017068	26601	$2.611 \pm 0.016 \times 10^{-2}$
$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu$, $\Lambda_c^+ \to p\pi^-\pi^+$	990964	23194	$2.342 \pm 0.015 \times 10^{-2}$
$\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu$, $\Lambda_c^+ \to p\pi^-K^+$	-	-	-

Table 1.5: The number of MC events generated and the number of MC events passing the no-PID stripping selection. The subscript "Acc" refers to the condition of passing the generator level cuts, and the subscript "Strip" refers to the condition of passing hte no-PID stripping selection. For prompt, the number generated corresponds to the number of true prompt decays, excluding those Λ_c produced from long-lived particles. The ratio of the stripped to generated candidates cannot be naively interpreted as an average stripping efficiency due to the poor data/MC agreement in resonant structure, but are given here for completeness. We also remind the reader that the SL $\Lambda_c^+ \to pK^-\pi^+$ simulation has different generator level cuts and therefore has a markedly different stripping efficiency with respect to the generator-level accepted candidates.

Propagating bin errors to errors on $\bar{\epsilon}$

In any binned efficiency schema, each bin efficiency has a binomial uncertainty associated to it. To propagate this error through to the final re-weighted efficiency, we utilise a toy

approach. In this method, we take a particular binning schema and evaluate efficiencies and binomial errors for each bin, where as usual the subscript "Acc" refers to the condition of passing generator level cuts:

$$\begin{aligned} \epsilon &= \frac{N_{\text{Strip}}}{N_{\text{Acc}}} \\ \sigma_\epsilon &= \sqrt{\frac{\epsilon(1-\epsilon)}{N_{\text{Acc}}}}. \end{aligned} \tag{1.4}$$

We then generate 1000 toy binning schemas based on the original. For every toy, in each bin we randomly resample the number of events which pass the selection from a binomial distribution, using the original number of candidates generated in said bin and the efficiency of the original bin as the n and p of the binomial distribution respectively. We then use each binning schema to assign a stripping efficiency to each real data candidate, and then from the efficiency-corrected yield extract a new stripping efficiency for the mode using the candidate sWeights:

$$\begin{aligned} M &= \sum^{n} \frac{w_i}{\epsilon_i} \\ \bar{\epsilon} &= N/M \end{aligned} \tag{1.5}$$

where w_i is the candidate sWeight, ϵ_i the candidate stripping efficiency, M the adjusted yield, N the extracted raw yield and $\bar{\epsilon}$ the re-weighted average efficiency. In such a fashion, for each toy an $\bar{\epsilon}$ is calculated. We then take the standard deviation of the $\bar{\epsilon}$ distribution for all toys over the original re-weighted efficiency as a fractional systematic uncertainty on the efficiency.

The κ binning statistic

There is no standard way to find the optimum bin granularity given a specific simulated dataset. The efficiency within a given bin should be as close to uniform as possible. If there is significant variation in efficiency across a kinematic bin then differences in kinematics between the data and simulation can result in the average efficiency in the bin being markedly different in data and simulation. However, binning too finely inflates the statistical error and eventually can cause biases.

It is difficult to visualise an efficiency space in more than 2 dimensions - as such the candidate developed a method of parameterising the efficiency variation between adjacent

bins relative to the statistical uncertainty on the adjacent bin efficiencies. This weighted difference in efficiency between bin a and bin b, κ_{ab}, may be expressed as:

$$\kappa_{ab} = \frac{|\epsilon_a - \epsilon_b|}{\sqrt{\sigma_a^2 + \sigma_b^2}}. \tag{1.6}$$

We consult the mean κ over multidimensional binning schemas. If this mean value is too large, the schema is too coarse given the simulation statistics available, and the variation of bin efficiency relative to the binomial error is too large. If the mean is too small, the schema is too fine given the statistics available, and the statistical uncertainty on the bin efficiencies is too high relative to the efficiency variation. We use the mean κ value of the bin schemas to determine which schemas provide the best description of the efficiency space - those binnings with a mean κ close to one are determined to be suitable. A full description of this test statistic and a justification of the optimum mean κ is given in Appendix **??**, which can be read as a standalone chapter.

Efficiency biases in re-weighting

A bias in the stripping efficiency re-weighting due to finite simulation statistics was also identified when using our expression for the adjusted yield. This bias results in the re-weighted efficiencies being biased downward, and is proportional to the ratio of the statistical uncertainty on the bin efficiencies divided by the bin efficiencies. For a set number of simulated events, this bias therefore becomes worse as the number of bins increases. Investigations with toy simulation determined the magnitude of this bias for each stripping efficiency. This allows an upper limit on the number of bins in a schema to be derived, which is defined as the number of bins where the magnitude of the bias is larger than the binomial uncertainty on the integrated simulation sample. The statistical motivation for this bias and the full results of our toy studies are presented in Appendix **??**, which can also be read as a standalone chapter.

Recipe for reduced dimensionality efficiencies

We describe the steps to obtain the final stripping selection efficiency and the associated systematics for each mode as follows:

1. We consult the data/MC distributions for each resonant variable. If there is good agreement, we discount the variable from the re-weighting schema.

2. We consult the stripping acceptance with respect to each resonant variable. If the acceptance is flat, we discount the variable from the re-weighting schema.

3. From steps 1 and 2 we arrive at a list of resonant variables for which we must re-weight.

4. At this stage we begin to construct local efficiencies using a variety of uniformly binned schemas, with different numbers of bins in each variable. We bin the simulated sample in the required variables and, by counting the numbers of generated and stripped candidates in each bin, assign an efficiency and binomial error on said efficiency to each bin in the schema. We construct binnings with 5 – 15 bins in each variable.

5. We then perform the following for each schema:

   (a) Ensure that the binning granularity is low enough, and the binomial errors on the bin efficiencies low enough, that the bias identified in Section 1.1.5 is not statistically significant. If it is, we discount the schema from those considered suitable.

   (b) Calculate the mean κ value for the schema. We make the ansatz that binnings with mean κ above 1.5 should be discounted because the efficiency variation across the individual bins in the schema are too high. We also discount binnings with mean κ below 0.9 because the statistical fluctuations in the individual bin efficiencies begin to dominate over genuine efficiency structure.

   (c) For each of the surviving schemas which satisfy the above stability criteria, we re-weight the real data using the efficiency bin schema to assign per-candidate efficiencies to the data. We then use these to extract average re-weighted per-mode efficiencies.

   (d) We use the toy method outlined in Section 1.1.5 to propagate the statistical uncertainty from finite simulation statistics on the individual bins (and therefore the per-candidate efficiencies) to the final re-weighted average efficiency.

6. At this point we have arrived at a set of efficiency binning schemas for the mode which satisfy the stability criteria with different binning granularities. We have used each of these schemas to assign a set of per-candidate efficiencies to the real data, and used each of these to evaluate average re-weighted efficiencies for the mode.

Each of these efficiencies has a statistical uncertainty ascribed to it due to limited MC statistics.

7. In steps 1 and 2 we have discounted variables from the schema based on data/MC compatibility or acceptance flatness. These checks are, however, 1D checks while integrating over all other resonant variables, and are insensitive to local variations with respect to the excluded variable in the multidimensional space which average out when the integration occurs. As a cross check:

   (a) If a variable has been discounted from the schema, we construct new binning schemas using the variables which have not been discounted with an arbitrary number of bins in the discounted variable under investigation. We then repeat the re-weighting procedure with this new schema to acquire an average re-weighted efficiency and associated statistical error.

   (b) We then compare this new efficiency with an efficiency determined from a binning schema and re-weighting without the discounted variable.

   (c) If the average re-weighted efficiencies are consistent within error, we take this as evidence that the discounting of the variable from the schema is well motivated. If they are not in agreement, then we take this as evidence that there is some local structure with respect to the discounted variable, and that the discounting of the variable from the schema is not valid. As such we re-include the discounted variable into the schema.

8. Now we have a final list of re-weighted efficiencies which are derived only from those schemas satisfying the stability criteria. Each of these will have different granularities. Therefore each will be subject to slightly different statistical fluctuations in each bin, but this is largely taken into account by the toy method to evaluate statistical uncertainties.

9. The foremost reason for considering an ensemble of efficiencies from valid binning schemas stems from the consideration of the systematic uncertainty which arises due to the finite bin size; this finite size means that the efficiency in each bin is not single valued. Between schemas in our ensemble of binnings, the bin boundaries have different locations. This gives us some sensitivity to the variation of efficiency across individual bins in the schemas, and its contribution to the systematic uncertainty on

the re-weighted efficiency, when we consider the variation in the extracted re-weighted efficienies in the ensemble.

10. We therefore take the median value of the ensemble of re-weighted efficiencies as the final re-weighted efficiency, with the highest and lowest efficiencies forming upper and lower bounds on the systematic uncertainty due to binning effects. We take the maximum statistical uncertainty of the efficiencies as the systematic uncertainty from finite simulation statistics.

This allows the derivation of efficiencies which are re-weighted to take account of the resonant structures in data, in such a way which allows us to estimate and minimise systematic uncertainties from limited statistics and the limitations of the binning schema.

Final re-weighted efficiencies

We herein present the final re-weighted stripping efficiencies for each analysis on a per-mode basis. We also present a comparison between the phase-space average efficiencies and the re-weighted efficiencies for each mode in the two analyses in Table 1.7. The re-weighting procedure alters most of the stripping efficiencies at statistically significant levels.

Stream	Mode	Final ϵ_{rw} [%]	Fractional systematic [%]
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	$(0.289 \pm 0.006 \pm 0.003)$ %	2.1 / 1.0 / 2.3
	$\Lambda_c^+ \to pK^-K^+$	$(0.266 \pm 0.008 \pm 0.003)$ %	3.0 / 1.1 / 3.2
	$\Lambda_c^+ \to p\pi^-\pi^+$	$(0.334 \pm 0.008 \pm 0.001)$ %	2.4 / 0.3 / 2.4
	$\Lambda_c^+ \to p\pi^-K^+$	-	-
SL	$\Lambda_c^+ \to pK^-\pi^+$	$(1.194 \pm 0.004 \pm 0.016)$ %	0.3 / 1.3 / 1.4
	$\Lambda_c^+ \to pK^-K^+$	$(2.612 \pm 0.018 \pm 0.038)$ %	0.7 / 1.5 / 1.6
	$\Lambda_c^+ \to p\pi^-\pi^+$	$(2.229 \pm 0.025 \pm 0.046)$ %	1.1 / 2.1 / 2.3
	$\Lambda_c^+ \to p\pi^-K^+$	-	-

Table 1.6: The final re-weighted efficiencies for each analysis. The first errors on the efficiencies are those from the propagated binomial uncertainties from limited simulation statistics as evaluated by our toy method. The second errors on the efficiencies are the additional systematic from the discounted variables and the finite bin width. The first fractional errors given are those from limited statistics, the second are those from the finite bin width and discounted resonant variables, the third are the sum of these in quadrature.

Table 1.7: The phase-space averaged per-mode stripping efficiencies and the re-weighted per-mode efficiencies. The errors on the phase-space averages are from finite simulation statistics, the errors on the re-weighted values are the full systematics, or the errors from finite simulation and the binning errors taken in quadrature. Also given are the differences between the phase-space averaged and re-weighted values in terms of the error in quadrature of the two efficiencies (ignoring correlations between the statistical uncertainties).

Stream	Mode	ϵ_{ave} [%]	ϵ_{rw} [%]	Difference (σ)
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	(0.313 ± 0.004) %	(0.289 ± 0.007) %	3.0
	$\Lambda_c^+ \to pK^-K^+$	(0.251 ± 0.003) %	(0.266 ± 0.009) %	1.6
	$\Lambda_c^+ \to p\pi^-\pi^+$	(0.345 ± 0.004) %	(0.334 ± 0.008) %	1.2
	$\Lambda_c^+ \to p\pi^-K^+$	-	-	-
SL TOS	$\Lambda_c^+ \to pK^-\pi^+$	(1.118 ± 0.013) %	(1.194 ± 0.016) %	3.7
	$\Lambda_c^+ \to pK^-K^+$	(2.611 ± 0.016) %	(2.612 ± 0.042) %	0.0
	$\Lambda_c^+ \to p\pi^-\pi^+$	(2.342 ± 0.015) %	(2.229 ± 0.052) %	2.1
	$\Lambda_c^+ \to p\pi^-K^+$	-	-	

1.1.6 PID efficiencies from a fully data-driven `PIDCalib`

`PIDCalib` outline

PID DLL cuts are included in all prompt and semileptonic $\Lambda_c^+ \to phh'$ stripping selections, but the PID response is known to be poorly modelled in LHCb simulation. This is caused by the lower detector occupancy in simulated events and the changing operating conditions in real data (the gas radidator temperature and pressure fluctuates over time) that are difficult to model. The `PIDCalib` package is used to assign event-by-event PID efficiencies to each candidate in a data-driven fashion.

The PID response of a given daughter is dependent on the daughter's mass, momentum and η. It is also dependent on the number of particles passing through the RICH in the complete event. The `PIDCalib` method assumes that the PID response of a given track may be characterised by a finite suite of kinematic and event variables. There exist two commonly used suites of variables within typical LHCb analyses:

- p, p_{T}, and event N_{Tracks}.
- p, η and event N_{Tracks}.

Both of these suites of variables are assumed to fully characterise the PID DLL response to a given track.

The method exploits the use of decay channels which may be cleanly reconstructed purely with kinematic constraints, and without PID cuts of any form. Reconstruction

of these modes allows for samples of p, π and K tracks to be acquired which have not been subjected to any PID cuts. The decay modes corresponding to each type of charged daughter are:

p : $\Lambda \to p\pi^-$

K/π : $D^*(2010)^+ \to D(K^-\pi^+)\pi_s^+$

By binning the data from these modes in schemas composed of one of the above sets of variables, and then applying the PID DLL cuts used in the analysis, it is possible to acquire local PID efficiencies within the variable space in an entirely data-driven way. Signal data is assigned an event-by-event PID efficiency based on its position within said variable space.

Data reference samples

Ususally in `PIDCalib`, the end-user utilises a simulated signal, or reference, sample which has had no PID cuts applied to it. This entails passing the sample to `PIDCalib`, where the application assigns each simulated candidate's tracks a PID efficiency from the performance histograms generated with the calibration data tracks. For decays with multiple daughter tracks with PID requirements, the candidate PID efficiency is then the product of the track PID efficiencies - in this way the kinematic correlations between daughter tracks are accounted for. The per-mode signal PID efficiency is therefore simply the average candidate PID efficiency, as no PID cuts have been applied to the sample.

In the case of Λ_c decays, it is known that the daughter kinematics, and especially the correlations between daughter kinematics due to intermediate resonances, are badly modelled in the LHCb simulation. In some cases where poor kinematic modelling is known to be an issue, a solution is to re-weight the simulation kinematics to match the data and then to use the simulated sample in `PIDCalib`. The inclusion of PID cuts in all stripping selections in this analysis precludes this option, as we do not have access to the unbiased PID distribution before cuts. As such, the proponents have developed an implementation of `PIDCalib` where a data-driven reference sample with PID cuts is utilised.

The calibration data is used to assign per-candidate efficiencies in the usual way to the real data reference sample. Then, to extract a per-mode average PID efficiency we calculate the PID-adjusted yield by weighting the signal candidate sWeights with the

per-candidate PID weights. The expression for the PID-adjusted yield M is

$$M = \sum^{n} \frac{w_i}{\epsilon_i} \tag{1.7}$$

where there are n candidates, each with an sWeight w_i and a `PIDCalib` determined per-candidate efficiency ϵ_i. The per-mode PID efficiency is then

$$\bar{\epsilon}_{PID} = \frac{\sum^{n} w_i}{M}. \tag{1.8}$$

This is simply the raw yield over the PID-adjusted yield, and is functionally identical to the determination of the re-weighted stripping efficiencies. As such we can correct for the PID efficiency of the Λ_c decay modes without utilising any simulation data.

Binning schema and kinematic vetoes

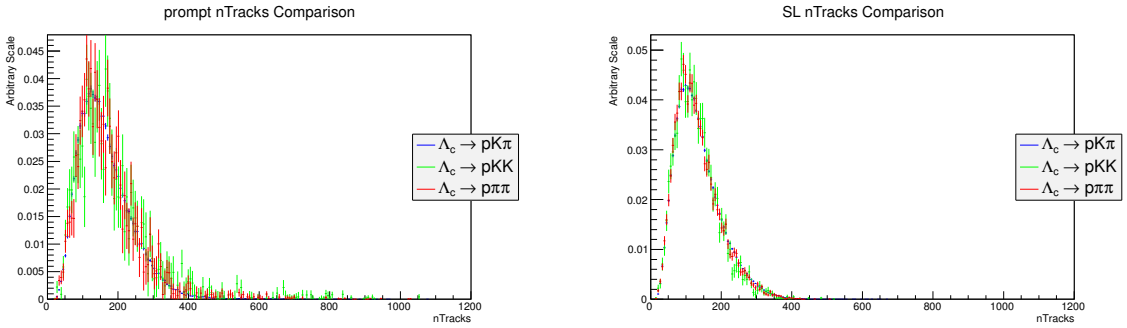
The choice of binning schema in this analysis is influenced by two factors. Firstly, the p_{T} distributions of protons from charm decays are much higher than in the calibration $\Lambda \to p\pi^-$ sample. The η distributions, however, have similar coverage. As such η is favoured over p_{T}.

A 3D binning will inflate the statistical uncertainty in each bin compared to a 2D binning. As can be seen in Figure 1.7, the distributions for event nTracks between the $\Lambda_c^+ \to phh'$ decay modes are similar. Discounting nTracks from the binning schema will result in the derived PID efficiencies becoming biased by a constant factor, which will be identical between the decay modes in the case of equal distributions of nTracks.

Therefore a 2D binning schema is chosen using p and η, detailed in full in Table 1.8. PID efficiency histograms with this binning schema are constructed using the calibration data, and are used to assign event-by-event local PID efficiencies. The histograms of PID efficiency in bins of track η and track p are shown in Figure 1.8 for the prompt selection and in Figure 1.9 for the semileptonic selection. The same plots for the semileptonically selected pions in the SL $\Lambda_c^+ \to p\pi^-\pi^+$, on which different PID cuts are placed compared to the pions in the other semileptonic modes, are shown in Figure 1.10.

Final PID efficiencies

The average per-mode PID efficiencies are given in Table 1.9. The subscript "PID" refers to the condition of passing the PID cuts used in the mode, and "Strip" refers to the condition



(a) (b)

Figure 1.7: The background subtracted event nTrack distributions in the prompt (a) and semileptonic (b) selections. The distributions are compatible in all $\Lambda_c^+ \to phh'$ decay modes.

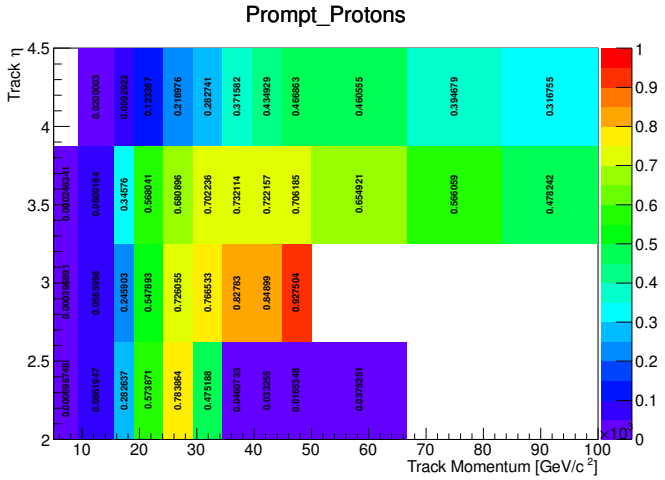
Table 1.8: The binning schemas used in the PID Calibration procedure.

Selection	Var	Bin Schema
All $\Lambda_c^+ \to phh'$	p [MeV/c^2]	5000 : 9300 : 15600 : 19000 : 24166.7 : 29333.3 : 34500 : 39666.7 : 44833.3 : 50000 : 66666.7 : 83333.3 : 100000
	η	2 : 2.625 : 3.25 : 3.875 : 4.5

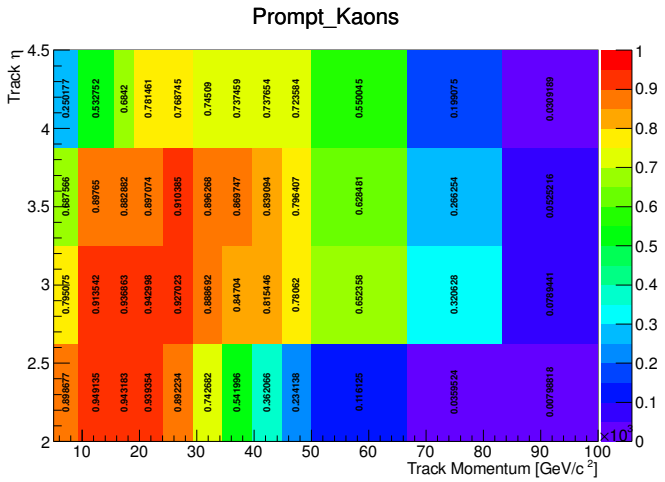
of passing the no-PID stripping selection (including the DTF convergence criteria and the kinematic vetoes).

Table 1.9: The per-mode average PID efficiencies.

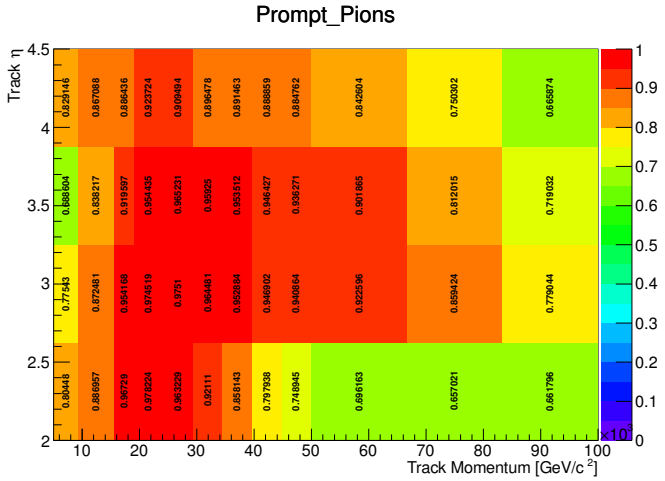
Stream	Mode	$\epsilon_{\text{PID}\|\text{Strip}}$ [%]
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	42.74 ± 0.04
	$\Lambda_c^+ \to pK^-K^+$	38.62 ± 0.04
	$\Lambda_c^+ \to p\pi^-\pi^+$	45.35 ± 0.05
	$\Lambda_c^+ \to p\pi^-K^+$	$xx.xx \pm x.xx$
SL	$\Lambda_c^+ \to pK^-\pi^+$	47.95 ± 0.05
	$\Lambda_c^+ \to pK^-K^+$	43.34 ± 0.04
	$\Lambda_c^+ \to p\pi^-\pi^+$	43.21 ± 0.04
	$\Lambda_c^+ \to p\pi^-K^+$	$xx.xx \pm x.xx$



(a) p : $\log(\mathcal{L}_p/\mathcal{L}_\pi) > 20$, $\log(\mathcal{L}_p/\mathcal{L}_K) > 12$

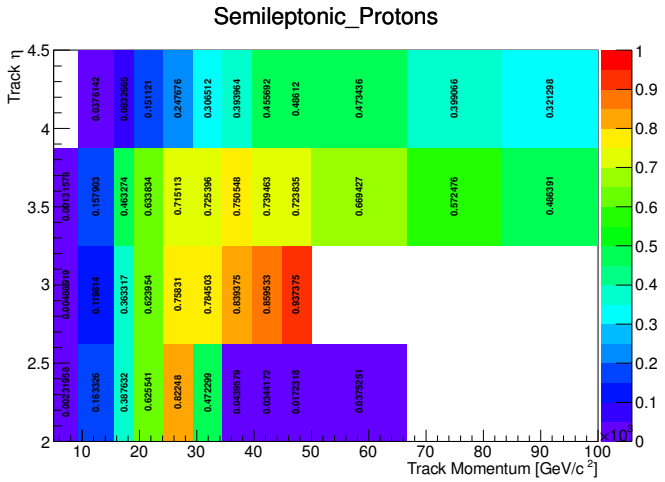


(b) K : $\log(\mathcal{L}_K/\mathcal{L}_\pi) > 10$, $\log(\mathcal{L}_K/\mathcal{L}_p) >$ -8

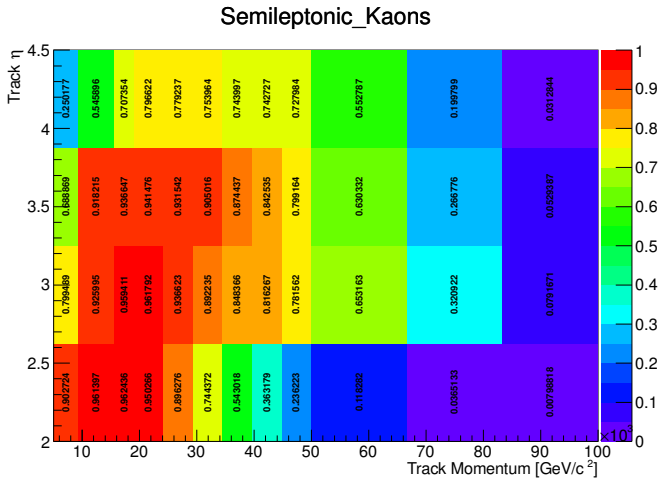


(c) π : $\log(\mathcal{L}_K/\mathcal{L}_\pi) < 0$

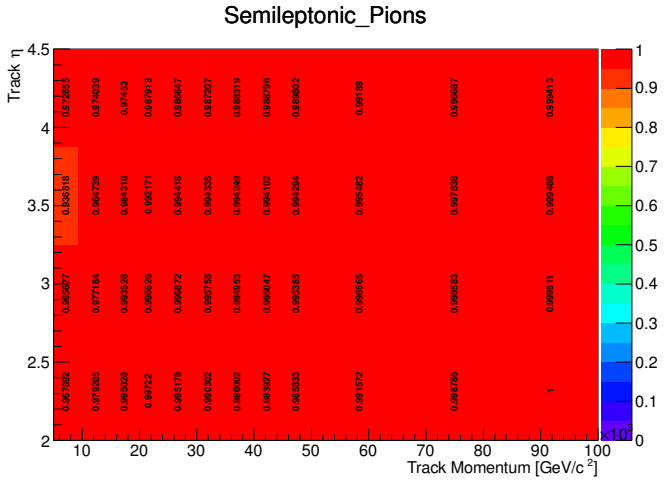
Figure 1.8: The prompt PID efficiencies in bins of track η and track p, as obtained from `PIDCalib`.



(a) p : $\log(\mathcal{L}_p/\mathcal{L}_\pi) > 20$, $\log(\mathcal{L}_p/\mathcal{L}_K) > 9$

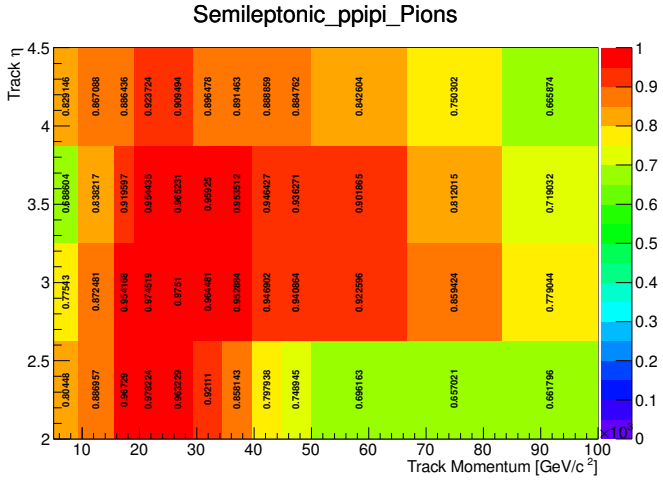


(b) K : $\log(\mathcal{L}_K/\mathcal{L}_\pi) > 10$



(c) π : $\log(\mathcal{L}_K/\mathcal{L}_\pi) < 10$

Figure 1.9: The semileptonic PID efficiencies in bins of track η and track p, as obtained from `PIDCalib`.



(a) π : $\log\left(\mathcal{L}_K/\mathcal{L}_\pi\right) < 0$

Figure 1.10: The semileptonic $\Lambda_c^+ \to p\pi^-\pi^+$ pion PID efficiencies in bins of track η and track p.

1.1.7 BDT efficiency

A BDT is applied in the offline selection of prompt $\Lambda_c^+ \to p\pi^- K^+$ events. It is likely that upon unblinding the channel that the extraction of a signal yield before the application of the BDT will be impossible. As the BDT is trained to be agnostic to the daughter kinematics in the Λ_c decay, the responses to the BDT of the CF and DCS modes should be equivalent despite the different intermediate resonances which are expected to occur in the two modes. We utilise the stripped MC to verify this, with the calculated efficiencies given in Table 1.10. The efficiencies for the $\Lambda_c^+ \to pK^-\pi^+$ and $\Lambda_c^+ \to p\pi^- K^+$ modes agree within errors. We investigate in $\Lambda_c^+ \to pK^-\pi^+$ the distributions of input variables to investigate the modelling of the BDT input variables. The Cabibbo-favoured is utilised in these checks for the higher statistics in the mode. We obeserve in all cases good agreement between data and simulation. As such we may confidently take from simulation the BDT efficiency for the DCS, verified by the identical Cabibbo-favoured response.

Table 1.10: The efficiencies of the BDT selection when applied to the prompt MC.

Stream	Mode	N_{Strip}	$N_{BDT\|Strip}$	$\epsilon_{BDT\|Strip}$ [%]
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	6507	5206	80.01 ± 0.50
	$\Lambda_c^+ \to p\pi^- K^+$	7172	5771	80.47 ± 0.47

1.1.8 Full selection efficiencies and summary

The efficiencies of the full selections for each mode are given in Table 1.11. In these figures the trigger efficiencies in the prompt analysis are included to better illustrate the efficiencies of the full selections, though the trigger efficiency correction is not included in the final prompt relative branching fraction results as the efficiencies are demonstrated to be equal across the prompt modes. These are the final efficiencies used to derive the adjusted yields in the final branching fraction results.

Table 1.11: The total selection efficiencies in the analyses. The errors are the statistical errors from finite signal simulation and `PIDCalib` calibration data only.

Stream	Mode	ϵ_{Total}
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	$(1.24 \pm 0.04) \times 10^{-6}$
	$\Lambda_c^+ \to pK^-K^+$	$(1.15 \pm 0.04) \times 10^{-6}$
	$\Lambda_c^+ \to p\pi^-\pi^+$	$(1.40 \pm 0.04) \times 10^{-6}$
	$\Lambda_c^+ \to p\pi^-K^+$	$(x.xx \pm x.xx) \times 10^{-6}$
SL	$\Lambda_c^+ \to pK^-\pi^+$	$(2.71 \pm 0.01) \times 10^{-4}$
	$\Lambda_c^+ \to pK^-K^+$	$(2.59 \pm 0.04) \times 10^{-4}$
	$\Lambda_c^+ \to p\pi^-\pi^+$	$(2.21 \pm 0.03) \times 10^{-4}$
	$\Lambda_c^+ \to p\pi^-K^+$	$(x.xx \pm x.xx) \times 10^{-4}$

1.2 Yield extraction

The analysis proedure relies on the extraction of genuine $\Lambda_c^+ \to phh'$ candidates from those candidates formed from combinatorics. This was first accomplished through a fit to the Λ_c candidate mass, however investigations of the promptly selected $\Lambda_c^+ \to phh'$ signal $IP\chi^2$ distributions revealed a significant high end tail. Such a tail is consistent with contamination from those Λ_c produced from decays of long-lived b-hadrons. We give a comparison of the data and simulation for the $\Lambda_c^+ \to pK^-\pi^+$ prompt and semileptonic $\log_e(IP\chi^2)$ in Figure 1.11, where the signal distributions of the data are extracted from fits to the Λ_c candidate mass.

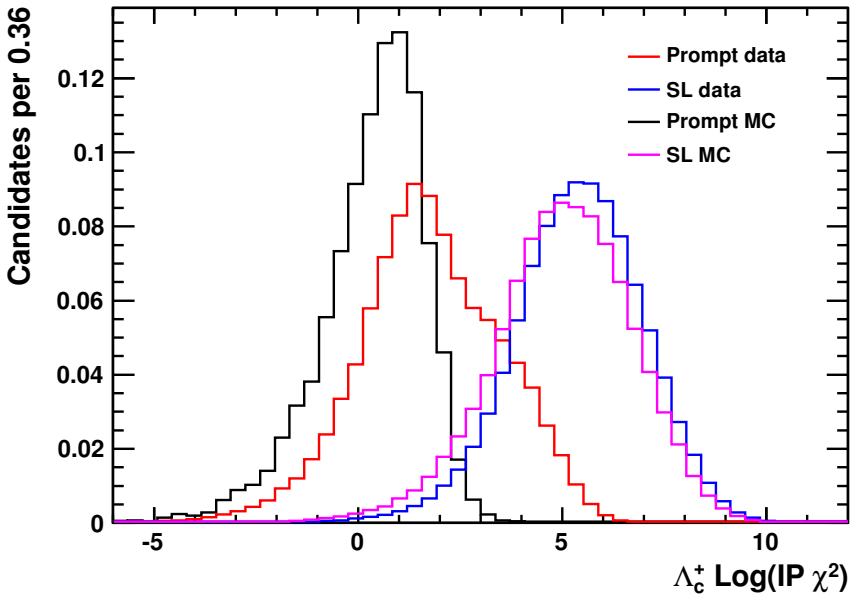


Figure 1.11: Data and simulation distributions for Λ_c $IP\chi^2$ distributions in both prompt and SL $\Lambda_c^+ \to pK^-\pi^+$. While the SL $IP\chi^2$ are generally well modelled (albeit with a slight downward shift of the distribution in simulation), the prompt distributions do not agree. The high end tail in the prompt data is strong evidence of secondary Λ_c contamination.

Simulation of secondary Λ_c candidates was used to investigate the per-mode phase-space averaged stripping acceptance. The Λ_c candidates in the sample were all verified to have a long-lived b-hadron mother. In all cases the acceptances are higher than the prompt Λ_c. Importantly, the ratios of the secondary efficiencies between the $\Lambda_c^+ \to phh'$ modes are not consistent with the ratios of the prompt efficiencies. As such, using purely prompt simulation to efficiency correct the candidate distributions extracted with only a fit to the

Λ_c mass will yield biased relative branching fraction results in the prompt analysis.

	Prompt		Secondary	
Mode	N_{Acc}	$\epsilon_{\text{Strip\|Acc}}$	N_{Acc}	$\epsilon_{\text{Strip\|Acc}}$
$\Lambda_c^+ \to pK^-\pi^+$	2308801	0.313 ± 0.004	225139	0.701 ± 0.018
$\Lambda_c^+ \to pK^-K^+$	2289265	0.251 ± 0.003	226432	0.693 ± 0.017
$\Lambda_c^+ \to p\pi^-\pi^+$	2305682	0.345 ± 0.004	223950	0.686 ± 0.017
$\Lambda_c^+ \to p\pi^-K^+$	-	-	-	

Table 1.12: The statistics and phase-space averaged stripping efficiencies for the prompt and secondary Λ_c in the "prompt" simulation. We note that the subscript "Acc" refers to the condition of passing the generator level cuts, and "Strip" refers to the condition of passing the no-PID stripping selection (including the kinematic vetoes and DTF convergance criterion).

This informs the choice of yield extraction in the analyses. The prompt yield extraction utlises a simultaneous 2D fit to the Λ_c mass and the Λ_c $\log_e(IP\chi^2)$ to extract the distribution for prompt signal candidates such that the secondary $\Lambda_c^+ \to phh'$ component may be discriminated and treated as a background. We do not attempt to conduct a measurement with the data in the prompt selections identified as secondary - the muon tagged samples in the semileptonic analysis enjoy a higher selection efficiency and greater purity, and can provide a more precise measurement of the relative branching fractions. In the semileptonic analysis no evidence for significant prompt contamination in the data samples is observed. Therefore the signal extraction in the semileptonic analysis is accomplished with a fit to the Λ_c mass.

1.2.1 Prompt

The species identified in the prompt sample are as follows:

Prompt $\Lambda_c^+ \to phh'$ - decays of $\Lambda_c^+ \to phh'$ where the Λ_c is produced either directly at the primary interaction or from a decay from a short-lived excited charm hadron. This is classed as the "signal". The mass distribution of these candidates is modelled by a Crystal Ball function [2] and a Gaussian function constrained to a shared mean in the Cabibbo-favoured mode, and a Gaussian function in all other modes. [1]

[1]The Crystal Ball function is a gaussian function with a power law low-end tail. It is of particular use in parameterising the mass distributions of candidates formed from combinations of tracks from a decaying particle, whereby unreconstructed final state radiation (FSR) results in a lower computed invariant mass of the system which manifests in a low end power law tail.

Secondary $\Lambda_c^+ \to phh'$ - decays of $\Lambda_c^+ \to phh'$ where the Λ_c is produced in the decay of a long-lived b-hadron. This is classed as a background. The mass distribution of these candidates is modelled by a Crystal Ball function and a Gaussian function constrained to a shared mean in the Cabibbo-favoured mode, and a Gaussian function in all other modes.

Combinatorics - combinations of unrelated tracks which mimic $\Lambda_c^+ \to phh'$ decays. The mass distribution of these candidates is modelled by a first order polynomial.

In all modes the $\log_e(IP\chi^2)$ candidate distributions are modelled by a Bukin function, which is a modified Novosibirsk function with tail extended tail functions. The Bukin functional form is

$$B(x;\mu;\sigma;\xi;\rho_1;\rho_2) = \begin{cases} \exp\left[\rho_1 \frac{(x-x_1)^2}{(x-x_1)^2} + \frac{(x-x_1)\xi\sqrt{\xi^2+1}.\sqrt{2\log 2}}{1} - \log 2\right] & x < x_1 \\ \exp\left[-\log 2.\left[\frac{\log\left(1+2\xi\sqrt{\xi^2}+1.\frac{x-\mu}{\sigma\sqrt{2\log 2}}\right)}{\log\left(1+2\xi\left(\xi-\sqrt{\xi^2+1}\right)\right)}\right]^2\right] & x_1 < x < x_2 \\ \exp\left[\rho_2 \frac{(x_2-x)^2}{(x_2-\mu)^2} + \frac{(x_2-x)\xi\sqrt{\xi^2+1}.\sqrt{2\log 2}}{1} - \log 2\right] & x > x_2 \end{cases} \quad (1.9)$$

where ξ is the Novosibirsk assymetry parameter, $\sqrt{2\log\sigma}$ is the full width half maximum (FWHM), μ is the position of the mode, and ρ_1 and ρ_2 are the left and right tail parameters respectively. x_1 and x_2 are the turnover points where the function has half of its maximum value - they are defined as:

$$x_1 \equiv \mu + \sigma\sqrt{2\log 2}\left(\frac{\xi}{\sqrt{\xi^2+1}} - 1\right) \quad (1.10)$$

$$x_2 \equiv \mu + \sigma\sqrt{2\log 2}\left(\frac{\xi}{\sqrt{\xi^2+1}} + 1\right) \quad (1.11)$$

The function and its first derivative are continuous at x_1 and x_2.

The fit is a simultaneous 2D fit to the Λ_c mass and the Λ_c $IP\chi^2$. We utilise an unbinned extended maximum likelihood fit to the data candidates. All parameters in the mass models are allowed to float. The fit takes place over -80 – +45 MeV/c^2 of the nominal Λ_c mass of 2286.46 MeV/c^2. In the $IP\chi^2$ models we take the shape of the combinatoric from a fit to the mass sidebands in data and fixing all Bukin parameters in the full fit to the results of this fit. For the prompt candidates we take initial values for the Bukin parameters from fits to simulated candidates, allowing all parameters to float with the

exception of the ρ_2 high-end tail, which lies in a region dominated by the secondaries and so is fixed to the value extracted from simulation. The secondary shape has its tail and assymetry parameters fixed to values determined from fits to simulated data, while its mean is allowed to float.

We await the development of the prompt fit model to be finalised before fit results can be provided in this section.

1.2.2 Semileptonic

The species identified in the semileptonic sample are as follows:

$\Lambda_b^0 \to \Lambda_c^+ \mu^- \overline{\nu}_\mu$, $\Lambda_c^+ \to phh'$ - decays of $\Lambda_c^+ \to phh'$ where the Λ_c is produced in a semileptonic decay of a Λ_b^0. This is classed as the "signal". The mass distribution of these candidates is modelled by a Crystal Ball function [2] and Gaussian function constrained to a shared mean in the Cabibbo-favoured mode, and a Gaussian function in all other modes.

Combinatorics - combinations of unrelated tracks which mimic $\Lambda_c^+ \to phh'$ decays. The mass distribution of these candidates is modelled by a first order polynomial.

We utilise an unbinned extended maximum likelihood function in the fit to the data candidates. All parameters in the mass models are allowed to float. The fit takes place over -80 – +45 MeV/c^2 of the nominal Λ_c mass of 2286.46 MeV/c^2.

The fit results after the final selection are shown for the semileptonically produced candidates in Figure 1.12 – Figure 1.14. In all plots, the total combined pdf is indicated in blue, the combined signal pdf is indicated in green, any individual signal components are indicated in cyan, and the background component is indicated in red.

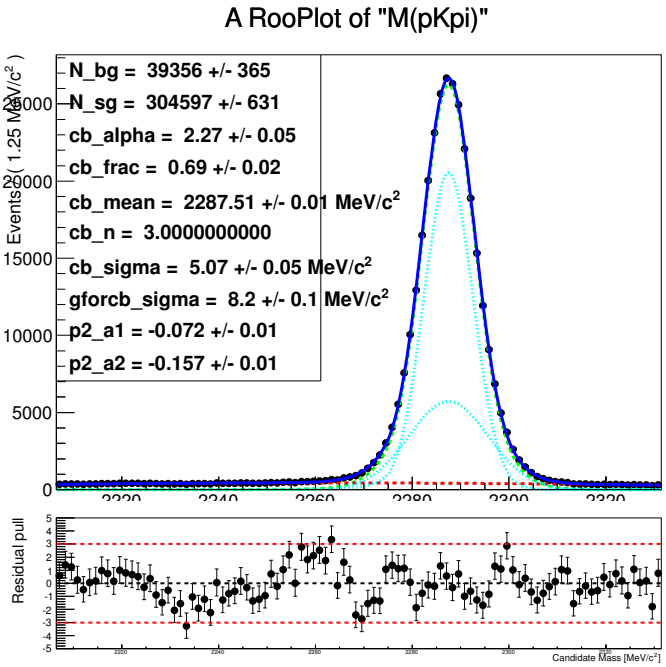


Figure 1.12: Λ_c mass fit and pull distribution for the semileptonic $\Lambda_c^+ \to pK^-\pi^+$.

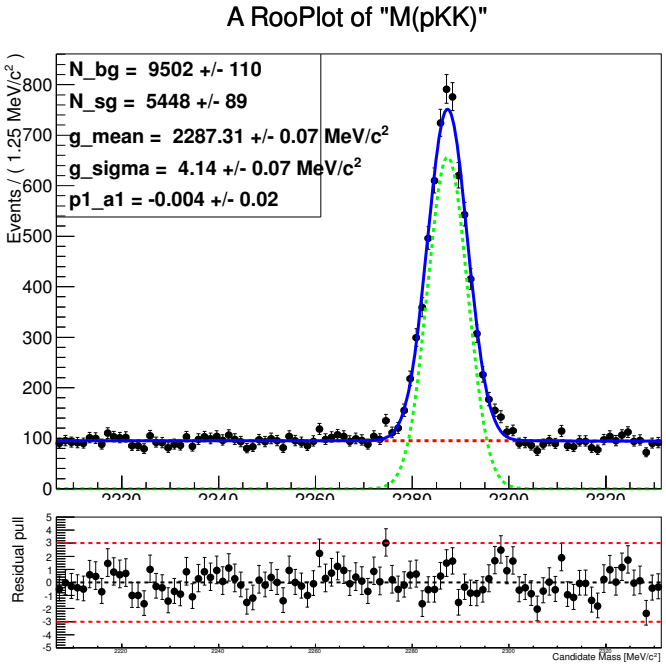


Figure 1.13: Λ_c mass fit and pull distribution for the semileptonic $\Lambda_c^+ \to pK^-K^+$.

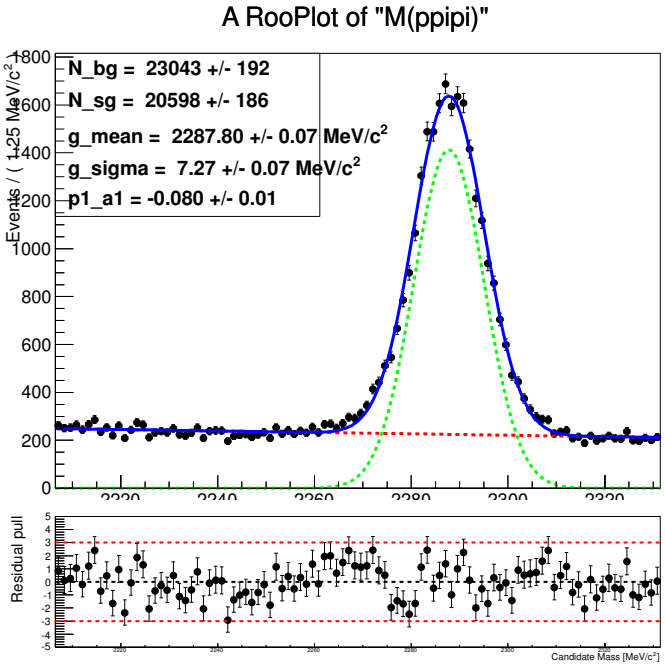


Figure 1.14: Λ_c mass fit and pull distribution for the semileptonic $\Lambda_c^+ \to p\pi^-\pi^+$.

1.2.3 Raw and adjusted yields

The raw and adjusted yields for each channel are given in Table 1.13. We do not provide errors on the adjusted yields, as these numbers when not taken as ratios fold in a number of detection and production uncertainties - their meaning in isolation is limited and they are merely given for illustration. We also note that the prompt adjusted yields fold in trigger efficiencies, which are demonstrated to cancel between the $\Lambda_c^+ \to phh'$ modes and therefore not included in the final branching fraction results.

Stream	Mode	Raw Yield	Adjusted Yield
Prompt	$\Lambda_c^+ \to pK^-\pi^+$	$xxxx \pm xxx$	$xxxx \pm xxx$
	$\Lambda_c^+ \to pK^-K^+$	$xxxx \pm xxx$	$xxxx \pm xxx$
	$\Lambda_c^+ \to p\pi^-\pi^+$	$xxxx \pm xxx$	$xxxx \pm xxx$
	$\Lambda_c^+ \to p\pi^-K^+$	$xxxx \pm xxx$	$xxxx \pm xxx$
SL	$\Lambda_c^+ \to pK^-\pi^+$	304597 ± 631	1.12×10^9
	$\Lambda_c^+ \to pK^-K^+$	5448 ± 89	2.10×10^7
	$\Lambda_c^+ \to p\pi^-\pi^+$	20598 ± 186	9.31×10^7
	$\Lambda_c^+ \to p\pi^-K^+$	$xxxx \pm xxx$	$xxxx \pm xxx$

Table 1.13: The raw yields for each of the prompt and semileptonically selected modes. The errors on the adjusted yields are the errors on the extracted yields, which accounts for the statistical uncertainty on the fitting procedure.

Bibliography

[1] E791 Collaboration, E. Aitala *et al.*, *Multidimensional resonance analysis of* $\Lambda_c^+ \to pK^-\pi^+$, Phys. Lett. **B471** (2000) 449, arXiv:hep-ex/9912003.

[2] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.