Dissertation
submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg. Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Marian Stahl
born in: Wetzlar
Oral examination: 11.07.2018

First observation of the decay $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ in preparation of a pentaquark search in the $\Lambda_c^+ \overline{D}^{(*)0}$ system at the LHCb experiment

Referees:
Prof. Dr. Stephanie Hansmann-Menzemer
Prof. Dr. Klaus Reygers

Abstract

This thesis reports the first observation of the decays $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^*(2007)^0 K^-$ using data corresponding to an integrated luminosity of 3 fb^{-1} collected at 7 and 8 TeV center-of-mass energies in proton-proton collisions with the LHCb detector. A future amplitude analysis of these Λ_b^0 decays allows to search the $\Lambda_c^+ \overline{D}^{(*)0}$ system for resonant contributions from P_c^+ pentaquarks. Two of which have been discovered in 2015 by the LHCb experiment in the $J/\psi\, p$ system. Observing P_c^+ pentaquarks in the $\Lambda_c^+ \overline{D}^{(*)0}$ system would allow for further insight to the nature of these resonances.

In addition, the measurement of the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^*(2007)^0 K^-$ branching fractions relative to the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ is made, resulting in

$$\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = (14.04 \pm 0.58 \pm 0.33 \pm 0.45)\,\%$$
$$\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{0*}(2007)^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = \left(43.5 \pm 1.4\,^{+1.2}_{-0.8} \pm 1.4\right)\,\% \,.$$

The first uncertainty is statistical, the second systematic and the third is due to the knowledge of the branching fractions of $D_s^- \to K^- K^+ \pi^-$ and $\overline{D}^0 \to K^+ \pi^-$. It is shown that the obtained result agrees with previously measured $B \to D \overline{D}^{(*)0} K^-$ branching fractions, as expected from the lowest order approximation of QCD factorisation.

The developed methods are integral components of the future amplitude analysis of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$. These and further algorithms presented here were developed in a modularised way to be applied to a variety of analyses at LHCb, in particular pentaquark searches.

Kurzfassung

Diese Arbeit dokumentiert die erste Endeckung der Zerfälle $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ und $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^*(2007)^0 K^-$ unter Verwendung von Daten die mithilfe des LHCb Detektors bei Schwerpunktsenergien von 7 und 8 TeV in proton-proton Kollisionen aufgezeichnet wurden und einer integrierten Luminosität von 3 fb^{-1} entsprechen. Eine zukünftige Amplitudenanalyse dieser Zerfälle erlaubt die Untersuchung des $\Lambda_c^+ \overline{D}^{(*)0}$ Systems auf resonante Beiträge von P_c^+ Pentaquarks. Zwei solcher Resonanzen wurden 2015 beim LHCb Experiment im $J/\psi\, p$ System entdeckt. Eine mögliche Beobachtung der P_c^+ Pentaquarks im $\Lambda_c^+ \overline{D}^{(*)0}$ System erlaubt Rückschlüsse auf die Natur dieser Resonanzen.

Zusätzlich wurden die $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ und $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^*(2007)^0 K^-$ Verzweigungsverhältnisse relativ zum Zerfall $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ gemessen. Diese betragen

$$\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = (14.04 \pm 0.58 \pm 0.33 \pm 0.45)\,\%$$
$$\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{0*}(2007)^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = \left(43.5 \pm 1.4\,^{+1.2}_{-0.8} \pm 1.4\right)\,\% \,,$$

wobei die erste Unsicherheit statistisch, die zweite systematisch und die dritte aufgrund der gemessenen Unsicherheiten der $D_s^- \to K^- K^+ \pi^-$ und $\overline{D}^0 \to K^+ \pi^-$ Verzweigungsverhältnisse ist. Die Messung stimmt mit bereits gemessenen $B \to D \overline{D}^{(*)0} K^-$ Verzweigungsverhältnissen überein, wie von QCD-Faktorisierung in niedrigster Ordnung erwartet.

Die hier entwickelten Methoden zur Signalselektion und Effizienzkorrektur sind ein integraler Bestandteil der $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ Amplitudenanalyse. Diese und weitere der hier vorgestellten Algorithmen wurden modular entwickelt und sind auf andere Analysen in LHCb übertragbar, insbesondere Pentaquark-Suchen.

Introduction

Protons and neutrons can be thought of as consisting of three quarks, bound by the strong interaction. Such strongly interacting particles are called hadrons, which the quark model allowed to classify into baryons, such as protons and neutrons, and mesons, which are bound states of quark- and antiquark. However, further combinations like a system made of four quarks and one antiquark were anticipated, known today as pentaquark. Pentaquarks have a loaded history with inconclusive experimental hints from the 70s and 80s, hyped evidence from about a dozen experiments in the 2000s which were then overwhelmed by contradicting findings, leaving strong scepticism about the existence of pentaquarks. That scepticism dictated a cautious analysis of an unusual structure in $\Lambda_b^0 \to J/\psi\, pK^-$ decays[1] observed by the LHCb experiment during a Λ_b^0 mass measurement. It turned out that the structure is due to two pentaquark-resonances, named $P_c^+(4380)$ and $P_c^+(4450)$, in the subsystem of J/ψ $(c\bar{c})$ and p (uud).

Former studies of the $c\bar{c}$ system already established another type of hadron that differed from mesons and baryons, called tetraquark and made of two quarks and two antiquarks. Both, tetra- and pentaquarks, allow to probe the quark model and the underlying theory of strong interactions from a new perspective, since their production, decay and binding mechanism is expected to differ from that of conventional hadrons. Yet all hadrons eventually have to be described in a coherent framework that allows to project the phenomenological picture to the fundamental theory of QCD.

The LHCb experiment provides a unique dataset for hadron spectroscopy, in particular pentaquarks. The discovered P_c^+ $(c\bar{c}uud)$ pentaquarks are expected to decay to Λ_c^+ (cud) $\overline{D}^{(*)0}$ $(u\bar{c})$ as well, but the nature of their binding will affect the decay rate to this system. The $\Lambda_c^+\overline{D}^{(*)0}$ system can be studied in $\Lambda_b^0 \to \Lambda_c^+\overline{D}^{(*)0}K^-$ decays, providing access to the full kinematic information. It is thus an important step to discover the $\Lambda_b^0 \to \Lambda_c^+\overline{D}^{(*)0}K^-$ decay, and establish methods for selection and efficiency correction for a subsequent amplitude analysis.

This step, together with a measurement of the $\Lambda_b^0 \to \Lambda_c^+\overline{D}^{(*)0}K^-$ branching fraction relative to the known $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay is presented here. The branching fraction is of interest in its own right, since the comparison to $B \to D\overline{D}^{(*)0}K^-$ branching fractions allows to probe the lowest order approximation of QCD factorisation. Further, the branching fraction contributes to the total inclusive $b \to c\bar{c}s$ rate, which is an important ingredient in model-independent searches for physics beyond the standard model in B meson decays. On the experimental side, their high yield suggests their presence as background in other b-hadron decays, which can be estimated with the branching fraction measurement. The decay $\Lambda_b^0 \to \Lambda_c^+\overline{D}^{(*)0}K^-$ and the reference $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay are reconstructed with the same final state particles. This choice ensures that many systematic uncertainties cancel, in particular in the correction for efficiencies due to reconstruction and selection.

This thesis is structured as follows: The first section introduces basic concepts of QCD, assess and motives pentaquark searches. The LHCb experiment is subject to the second section, introducing the detector components and outlining the flow from raw data to data for offline analyses. The analysis strategy is defined in section three, followed by a discussion on software development tools and statistical methods. Section five describes the development and calibration of classification variables for non-prompt Λ_c^+, D^0 and D_s^+ hadrons. The selection of $\Lambda_b^0 \to \Lambda_c^+\overline{D}^{(*)0}K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays and fits to their invariant mass spectra follow. The fits are used for statistical unfolding of the signal components, which are input to the efficiency correction detailed in section eight. Systematic uncertainties are assessed in section nine; all of which are combined to the results of section ten and subject to a closing discussion.

An additional bibliography for referencing online documentation and code is used. Such references begin with a C. Further, some of the references are internal to LHCb members. These begin with an I.

[1] The use of charge conjugation and natural units ($\hbar = c = 1$) is implied unless otherwise stated.

Table of contents

1 Motivation and theoretical background

This section starts with a "folkloristic"[1] view of the standard model of particle physics with a focus on the quark model. A more formal view on the topic is presented afterwards. The section closes by discussing exotic hadron spectroscopy and its impact on the understanding of quantum chromodynamics (QCD).

1.1 Historical introduction

A cornerstone of hadron physics was set in 1909 by Rutherford, Geiger and Mardsen, who conducted scattering experiments of α particles incident on gold-foil to discover the nuclear structure within atoms [2]. That nuclei themselves are composite objects was determined in 1917 using the nuclear reaction of pure Nitrogen with α-particles, where Rutherford found hydrogen-like particles. He postulated that the hydrogen nucleus is a fundamental particle [3], today known as proton (p). To compensate for the electric repulsion force of protons in nuclei neutrons (n) were hypothesised.

The existence of neutrons was established by Chadwick in 1932 [4], who re-examined and re-interpreted the known reaction of Beryllium with α-particles. Shortly after, Heisenberg proposed a new quantum number to explain the observed symmetries between protons and neutrons, like their mass and interaction strength [5]. This was termed isospin, which marked an important step towards the quark model.

Until this point, only ordinary matter was experimentally established, and the antimatter solutions that emerged from the Dirac-equation were considered a theoretical artifact [6]. Surprisingly, antimatter could be observed in a cloud chamber experiment which detected cosmic rays in 1932. The antimatter particles left an ionisation trail with a curvature matching the mass to charge ratio of an electron, but which deflected in the opposite direction [7]. Consequently, antiprotons ($\overline{p}$) and antineutrons ($\overline{n}$) were anticipated. It took another 20 years, until they were discovered in fixed target experiments at the Bevatron – one of the first proton synchrotrons [8, 9].

Another postulated particle was similarly approached with scepticism; the (electron) neutrino (ν_e). Pauli postulated its existence in 1930 to solve the conundrum of the radioactive beta decay. In that decay, the energy spectrum of the electron, emitted by a decaying neutron, is measured to be continuous. This stood in contrast to the expectation of a two body decay to the visible system of proton and electron. However, the neutrino, which Pauli claimed to be undetectable, was discovered in a reactor experiment in 1956; antineutrinos from beta-decays scattered off protons to produce neutrons and positrons [10].

Even though nuclear matter had been established, little was known about the interaction of protons and neutrons. In a model formulated by Yukawa in 1935, internuclear interactions are described by the exchange of massive bosons, known as mesons [11]. The first particles consistent with Yukawas description were discovered in 1947 in cosmic radiation [12]; those were the charged pions ($\pi^{\pm}$).

This set of hadrons and leptons could have completed the picture of (elementary) particle physics, but there were unforeseen discoveries in the meantime. First, the muon (μ) had been discovered in cosmic radiation in 1936 [13]. Because of its mass, it was thought to be the meson predicted by Yukawa, But later studies with "μ-mesons" showed that they did not interact via the (strong) nuclear force. A different approach to probe Yukawa's theory was to study nuclear interaction at short distances, *i.e.* higher energies. The attraction between nucleons was larger than expected, which led to the prediction of further mesons: an iso-scalar scalar($f_0(500)$)[2] and iso-scalar and iso-vector vector mesons (ω and $\rho^{\pm,0}$). The zoo of elementary particle began to grow further with the observation of unstable particles.

[1] That term, and the general idea of the historical introduction, are taken from Ref. [1].

[2] The nature of the $f_0(500)$ (or σ) is still elusive, see *e.g.* the review on scalar mesons in [14] and references therein.

These were identified by their V-shaped (Λ or K^0_S) or kinked ($K^\pm$) decay patterns in cloud chamber photographies [15, 16]. Further studies have shown that these particles were produced in pairs following certain rules: *e.g.* a Λ was produced together with a K^+, but never with a K^-. To explain this behaviour, a new quantum number was introduced, termed strangeness. Strangeness is conserved in the production processes of the strong interaction, but violated in the weak decay. It was realised that such weak processes are similar to β-decays, where isospin and parity conservation is violated.

More discoveries followed, namely an iso-quadruplet originating from πN interactions, known as Δ-resonances, a strange iso-triplet, known as Σ-baryons, and a doubly strange iso-doublet, known as Ξ-baryons along with excitations of these states. The number of elementary particles appeared to explode, as shown in Fig. 1.1. In attempts to find a scheme for all the particles, the ancient concept of elementary particles eventually led to a breakthrough. In a first model, Sakata extended the isospin formalism by strangeness, such that the proton, neutron and Λ appeared as building blocks of matter [17]. Five years later, in 1961, Gell-Mann and independently Ne'eman recycled the ideas of Sakata and formulated an abstract approach where elementary baryons and mesons did not exist [18, 19]. In his paper, Gell-Mann established the term "eightfold way" for the group theoretical octet representation of baryons, along with the term gluon to describe the coupling of the strong force. In addition, he anticipated the existence of the Ω^- baryon – a particle with strangeness -3 and a mass ~ 1685 MeV which was needed to complete the spin 3/2 decuplet of baryons [20].

That missing piece, the Ω^-, was discovered in 1964 [22]. However, the question *why* baryons and mesons fit into the multiplets of the "eightfold way" had still not been answered. From the perspective of group theory, the fundamental representation of the observed multiplets is a unitary triplet (**3**). Hence, Gell-Mann and Zweig independently postulated three new particles: quarks or aces [23, 24, 25].

Despite the elegant description of hadrons by the quark model, the existence of quarks was doubted by many physicists until the mid 1970s. One objection was that individual isolated quarks could not be discovered, a phenomenon known as the confinement. The consequences of confinement are now well understood, but its origin remains elusive (for a comprehensive summary, see Ref. [26]). A second problem was identified in the "edges" of the baryon decuplet, which seemingly violated the Pauli exclusion principle. To solve this, Greenberg proposed colour charge [27], initially perceived as a technical sleight.

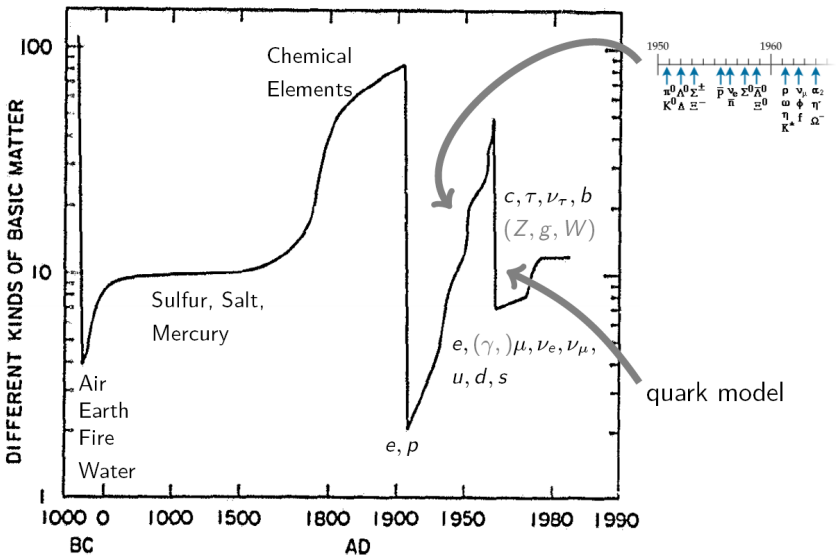


Figure 1.1: Number of particles considered elementary through time. Taken from Ref. [21] with modifications.

During this time of uncertainty regarding the quark model, high energy experiments at the Stanford Linear Accelerator Center (SLAC) indicated that protons have substructure [28, 29]. This substructure has been described by the parton model [30, 31, 32], which avoided to introduce any hypothesis about the nature of the constituents, *e.g.* that they are quarks. This was further motivated by the measured fact that electrically charged partons only contribute about half of the total fraction of the proton momentum. This meant that the remaining half is carried by neutral constituents, and was first circumstantial evidence for gluons. About ten years later, gluons were established by the observation of exclusive planar three-jet events in electron-positron collisions at all four experiments of the PETRA electron-positron collider at DESY [33, 34, 35, 36]. A Feynman diagram and event display of such a process is shown in Fig. 1.2. Today, quarks and gluons have been identified as the constituents that were found in experiments by the principle of asymptotic freedom [37, 38].

Even before the discovery of the gluon the quark model had been successfully used to predict a fourth type of quark, called charm quark. It was needed to explain the observed suppression of flavour-changing neutral currents in loop diagrams, called the GIM mechanism [40]. Experimental confirmation came with the discovery of the J/ψ meson in 1974 [41, 42]. The quark model was thus commonly accepted going forward.

A third generation of quarks was first proposed in 1973 (before the J/ψ discovery). The concept was one of a few possible solutions by Kobayashi and Maskawa [43] who extended the GIM mechanism to explain the observed *CP* violation in kaon decays [44]. Only two years later, in 1975, observation of the τ lepton [45] marked the first sighting of a member of the third generation of elementary particles.

Just another two years later, owing to ever larger collider experiments, particles consisting of a third generation down-type quarks, named bottom or beauty (b) quarks, were discovered [46]. The up-type equivalent, the top quark (t) and the missing τ-neutrino (ν_τ), were expected to be just around the corner, but it took more than 20 years to complete the family: with the top quark discovery in 1995 [47, 48] and the first observation of ν_τ interactions in 2000 [49]. It is known today that there are three, and only three, light generations of particles due to precision measurements on the Z boson resonance [50].

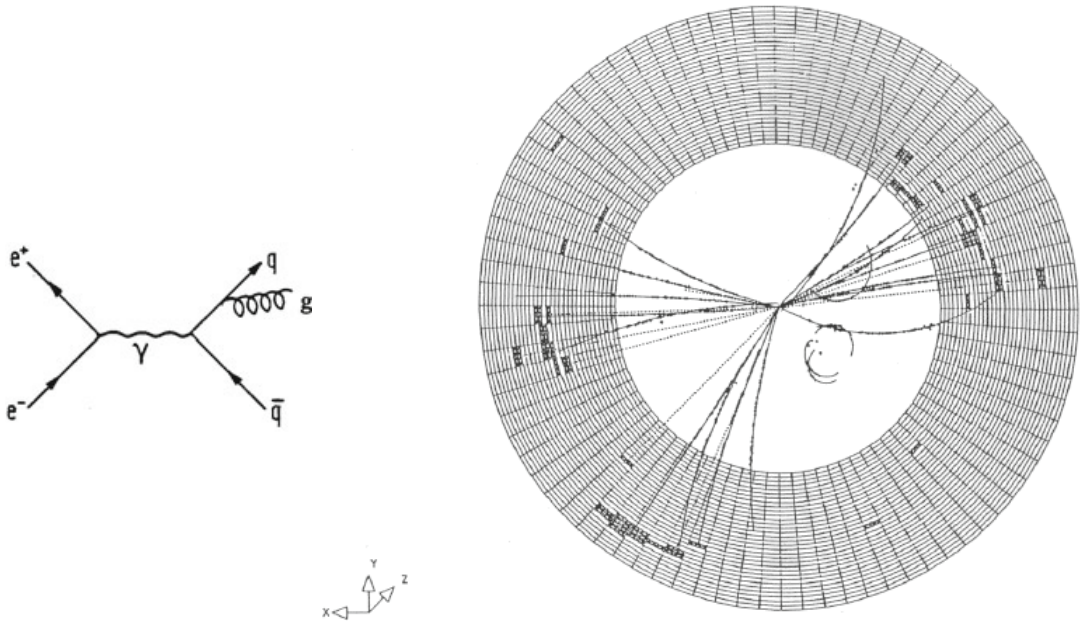


Figure 1.2: A leading order Feynman diagram of gluon bremsstrahlung (left), which is experimentally seen as a planar three-jet event. Such an event is shown in an event display from the JADE experiment (right). Taken from Ref. [39]

The Z boson, together with the $W^{\pm}$ bosons, are the mediators of the weak interaction. They were discovered in 1983 at the UA1 and UA2 experiments of the proton-antiproton collider at CERN [51, 52, 53, 54]. The UA1 and UA2 experiments were designed for this discovery, since properties of $W^{\pm}$ and Z were firmly predicted by a unified theory of electromagnetic and weak interactions. The theory developed from the attempt to write down a self-consistent gauge theory for weak interactions, in analogy with quantum electrodynamics (QED). It was realised that this is only possible if the theory includes QED [55, 56, 57].

Due to the short range of weak interactions, the weak exchange bosons were expected to be massive. Thus, a unified electroweak theory required a mechanism such that the bosons acquired mass and the photon remained massless, a mechanism now known as electroweak symmetry breaking [58, 59, 60]. This involves an additional interaction with a new fundamental scalar particle known as the Higgs boson (H^0). The long anticipated discovery of the Higgs boson was announced in 2012 by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) experiment at CERN [61, 62].

The mechanism of electroweak symmetry breaking allows for Yukawa-like couplings of chiral fermion-antifermion pairs to the Higgs field. This coupling gives mass to elementary fermions. However, elementary quarks are confined in hadrons, and there is another mechanism that drives the generation of mass at low energies. That mechanism is known since the early 1960s and is at the core of QCD. It is known as the dynamical chiral symmetry breaking (DCSB) [63, 64, 65, 66]. As a consequence almost all of the mass in the visible universe, that is, light baryonic matter – protons and neutrons – does not originate from the coupling to the Higgs field. DCSB is a nonperturbative feature of QCD, generating mass from nothing, *i.e.* from QCD itself. Still, QCD is linked to the Higgs sector by the fact that pions have mass. The pion would be massless in the absence of a mechanism that can generate a current-mass for at least one light-quark, as *e.g.* explained in Ref. [67]. In the following, the underlying theoretical framework of QCD is sketched out.

1.2 QCD and the quark model

QCD shall briefly be introduced in a more formal setting. Its Lagrangian is derived from Yang-Mills theory, a gauge theory with a non-Abelian symmetry group. Fundamental degrees of freedom of the QCD Lagrangian are identified and phenomenological concepts are outlined. As an example of the latter, flavour symmetry is discussed which allows to group hadrons into flavour-multiplets. Such phenomenological tools provide a basic formalism in the description of (exotic) hadrons.

1.2.1 Yang-Mills theory

In 1954, Yang and Mills formulated a gauge theory based on the non-Abelian SU(2) isospin-symmetry group to explain strong interactions [68]. The formalism can be extended to any compact semi-simple Lie group, such as SU(N). The theory describes transformation properties of field operators as follows: let $\{\phi_r(x)|r = 1, ...f\}$ be a set field operators, denoted as f-dimensional vector $\phi(x)$, which transforms under a $N^2 - 1$-dimensional gauge group SU(N) like

$$\phi(x) \rightarrow \exp\left[i \sum_{a=1}^{N^2-1} \theta^a(x) T_a \right] \phi(x) \,, \tag{1.1}$$

with the generators of the gauge group T_a. Note that the gauge shift $\theta^a(x)$ in a local symmetry depends on space-time, which is not the case in global symmetries. In order to construct a gauge invariant

Lagrangian, the derivatives of the field operators must transform like the field operators themselves. Using Einstein-summation convention henceforth, the gauge covariant derivative transforms like

$$\mathscr{D}_\mu \phi(x) \to \exp[i\theta^a(x)T_a]\left[\mathscr{D}_\mu \phi(x)\right] . \tag{1.2}$$

It is known from differential geometry that this condition is fulfilled by

$$\mathscr{D}_\mu \phi(x) = \left[\partial_\mu + igA^a_\mu(x)T_a\right]\phi(x) . \tag{1.3}$$

Thus the condition of local gauge invariance of the Lagrangian naturally introduces a coupling of $\phi(x)$ to $A_\mu(x)$ with strength g and N^2-1 gauge vector fields $A^a_\mu(x)$, whose quanta are the gauge bosons of the theory. These gauge fields transform like

$$A^a_\mu(x)T_a \to A^a_\mu(x)\left[e^{i\theta^b(x)T_b}T_a e^{-i\theta^b(x)T_b}\right] + \frac{i}{g}\left[\partial_\mu e^{i\theta^b(x)T_b}\right]e^{-i\theta^b(x)T_b} . \tag{1.4}$$

To arrive at a locally gauge invariant Lagrangian, the gauge fields are set in relation to the field strength tensor $F^a_{\mu\nu}(x)$. In a Yang-Mills theory, the field strength tensor is defined by the commutator of the gauge covariant derivatives

$$\left[\mathscr{D}_\mu, \mathscr{D}_\nu\right] = -igT_a F^a_{\mu\nu}(x) . \tag{1.5}$$

To solve this equation for $F^a_{\mu\nu}(x)$, the property

$$[T_a, T_b] = if_{ab}{}^c T_c \tag{1.6}$$

of the Lie-algebra is used, where $f_{ab}{}^c$ are the groups structure constants. With the help of Eq. (1.3), the field strength tensor is written as

$$F^a_{\mu\nu}(x) = \partial_\mu A^a_\nu(x) - \partial_\nu A^a_\mu(x) + g f^a{}_{bc} A^b_\mu(x) A^c_\nu(x) . \tag{1.7}$$

It transforms as a gauge vector – is thereby no observable quantity – but the product is the demanded gauge invariant quantity known as the gauge field Lagrangian:

$$\mathscr{L}_{\text{gf}}(x) = -\frac{1}{4}F^{\mu\nu}_a(x)F^a_{\mu\nu}(x) . \tag{1.8}$$

In the presence of a Dirac field $\phi(x)$ the Lagrangian is given by

$$\mathscr{L}(x) = \mathscr{L}_{\text{gf}}(x) + \mathscr{L}_0(x) = \mathscr{L}_{\text{gf}}(x) + \overline{\phi}(x)(i\not{\mathscr{D}} - M)\phi(x) , \tag{1.9}$$

i.e. the gauge field Lagrangian plus the Lagrangian for a free Dirac fermion. In Eq. (1.9) $\overline{\phi}(x) = \phi^\dagger(x)\gamma^0$ is the Pauli adjoint spinor and $\mathscr{D}_\mu$ is the gauge covariant derivative, using the Feynman slash notation $\not{\mathscr{D}} = \gamma^\mu \mathscr{D}_\mu$. The derivation so far is valid for any N^2-1-dimensional gauge group SU(N). QCD will be discussed the following.

1.2.2 QCD in the context of a Yang-Mills theory

QCD is described by a Yang-Mills theory with an underlying $\text{SU}(3)_\text{c}$ colour charge group. Experiments verified the colour charge hypothesis and determined that the number of colours is $N_\text{c} = 3$. The only compact semi-simple Lie group having 3-dimensional irreducible real and complex representations is SU(3). The complex representations are needed to account for anticolour, carried by antiquarks and gluons.

By applying the $SU(3)_c$ colour charge group of QCD to the Lagrangian of Yang-Mills theory, Eq. (1.9), the Lagrangian of QCD can be written as

$$\begin{aligned}\mathcal{L}_{\mathrm{QCD}}(x) = &-\frac{1}{4}\left(\partial_\mu G^a_\nu(x) - \partial_\nu G^a_\mu(x)\right)\left(\partial^\mu G^\nu_a(x) - \partial^\nu G^\mu_a(x)\right) + \overline{q}^\alpha_f\left(i\partial\!\!\!/ - M_f\right)q^\alpha_f \\ &+ g_s \overline{q}^\alpha_f \not{G}^a(x)\left(\frac{\lambda_a}{2}\right)_{\alpha\beta} q^\beta_f \\ &+ \frac{g_s}{2} f_a{}^{bc}\left(\partial_\mu G^a_\nu(x) - \partial_\nu G^a_\mu(x)\right) G^\mu_b(x) G^\nu_c(x) \\ &- \frac{g_s^2}{4} f^{abe} f_{cde} G^\mu_a(x) G^\nu_b(x) A^c_\mu(x) G^d_\nu(x)\,.\end{aligned} \tag{1.10}$$

Here, α, β are the colour indices, a are the 8 gauge field indices, q_f denotes a quark field of flavour f and μ, ν the Dirac indices. Those were omitted on the fermion spinors for readability. Further, a notation is chosen in which the flavour quantum number is explicitly conserved. The generators $T_a = \frac{\lambda_a}{2}$ of $SU(3)$ are given by the Gell-Mann matrices λ_a [18]. To link the formal gauge field A and the field operator ϕ to QCD, the notation G for the gluon field and q for the quark field was chosen.

The coupling g_s is the strong gauge coupling. In the quark mass term $-\overline{q}^\alpha_f M_f q^\alpha_f$ the "bare" quark masses enter. These are the ones generated by the Higgs mechanism. Moreover, it is seen that a term $\sim m^2 G^\mu_a G^a_\mu$ does not appear, since it would violate gauge invariance (*cf.* Eq. (1.4)). Thus, gluons, as any gauge vector bosons in a Yang-Mills theory, must be massless.

The first line in Eq. (1.10) contains the kinetic terms for quark and gluon fields; the second line describes the colour interaction between quarks and gluons and line 3 and 4 give rise to self-interaction of 3$^{\mathrm{rd}}$ and 4$^{\mathrm{th}}$ order of the gluons. These terms come from the non-vanishing commutator of the Lie-algebra in Eq. (1.7) and appear even in the absence of other fields.

This is in contrast to QED, described by a U(1) Abelian gauge theory, where the force-carrier, the photon, is not charged and no self-interaction occurs at leading order. The gluon self interaction in QCD has fundamental consequences: the property of asymptotic freedom at small distances/high energies; and confinement of colour charges at large distances/low energies. An analytic solution at all distances is strived for. However, several methods exist that allow to describe observations effectively. Some of their basic consequences are discussed in the following.

1.2.3 Quark Properties

The description of quantised quark- and gluon fields from first principles is highly non-trivial. The usage of perturbative methods similar to QED breaks down at low energies. That is because the coupling strength $\alpha_s(\mu^2_{\mathrm{R}})$ is a function of an (unphysical) renormalisation scale μ^2_{R}, which is needed to apply perturbative methods. Below a scale of $\mu^2_{\mathrm{R}} = \Lambda^2_{\mathrm{QCD}}$, called the Landau pole, the perturbative expansion diverges and QCD becomes a strongly coupled gauge theory. The value of this scale depends on several parameters and methods, and ranges from $\Lambda_{\mathrm{QCD}} \approx 300 - 1100\,\mathrm{MeV}$ [69].

It is conventional to call quarks heavy if their masses are above Λ_{QCD}, and light below that. But unlike leptons, quarks are confined inside hadrons and are not observed as physical particles. Their masses can therefore not be measured directly, such that any quantitative statement must make careful reference to the particular theoretical framework that is used to define it. Commonly used frameworks rely on perturbative approaches to calculate quark masses, such as the ones shown in Tab. 1.1.

The table also shows flavour quantum numbers, which were the key to establish the quark model, as discussed earlier. These quantum numbers are the third component of the isospin I_3, strangeness S, charm C, bottomness/beauty B' and topness/truth T, and are eigenstates of the strong interaction.

They are rotated in flavour-space for electroweak interactions, so that transitions from up-type (u, c, t) to down-type (d, s, b) quarks and vice versa are allowed.

Because QCD conserves the flavour quantum number (flavour-blindness), hadrons can be labelled by their minimum (valence) quark content, but are dynamically "dressed" with quarks and gluons. QCD, in the picture of the quark model, does not states to be qqq-baryons or $q\overline{q}$-mesons. So combinations like tetraquarks $(qq)(\overline{qq})$, mesonic molecules $(q\overline{q})(q\overline{q})$, pentaquarks $qqqq\overline{q}$, hybrid mesons $qg\overline{q}$, glueballs ggg, etc. are not forbidden.

I		II		III	
u	up $I_3 = +\frac{1}{2}$ $2.2^{+0.6}_{-0.4}$ MeV	c	charm $C = +1$ 1.28 ± 0.03 GeV	t	top $T = +1$ 173.1 ± 0.6 GeV
d	down $I_3 = -\frac{1}{2}$ $4.7^{+0.5}_{-0.4}$ MeV	s	strange $S = -1$ 96^{+8}_{-4} MeV	b	bottom $B' = -1$ $4.18^{+0.04}_{-0.03}$ GeV

Table 1.1: Quark Properties. The generation number is given on top. Name, flavour quantum number and the quark mass are given next to the symbol of the quark. The up-type (u, c, t) quarks carry an electric charge of $+\frac{2}{3}e$ and the down-type (d, s, b) quarks $-\frac{1}{3}e$. The masses were taken from ref. [14] and obtained by calculations using the modified minimal subtraction ($\overline{\text{MS}}$) renormalisation scheme [70]

1.2.4 A phenomenological approach to QCD: flavour SU(3) and the quark model

Historically, QCD was approached in a simpler fashion. There, hadrons were grouped into multiplets, relying on the approximate mass-degeneration of observed states, and treating quark fields as flavour symmetric. The flavour group of up, down and strange quarks, $\text{SU}(3)_\text{f}$, has two independent fundamental representations $\mathbf{3}$ and $\overline{\mathbf{3}}$. These representations can be displayed as weight diagrams in the (I_3, Y)-plane (fig. 1.3), where I_3 is the third component of the isospin and Y, the sum of strangeness and baryon number, is the (strong) hypercharge. The nodes in the $\mathbf{3}$ ($\overline{\mathbf{3}}$) diagram correspond to the three quark flavours u, d and s ($\overline{u}, \overline{d}$ and $\overline{s}$). The black solid lines between the nodes correspond to the lowering and raising operators which move between the different weight vectors, each forming a SU(2) subalgebra.

The laws of representation theory are then used to construct hadron multiplets. The hadronic flavour wave functions are obtained by performing the direct sum decomposition of the tensor product of irreducible representations. This is commonly referred to as *coupling* the quark flavours. For $\text{SU}(3)_\text{f}$-symmetric $q\overline{q}$ states, this is

$$\mathbf{3} \otimes \overline{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \,, \tag{1.11}$$

for qqq states the decomposition reads:

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = (\mathbf{6}_\text{S} \oplus \overline{\mathbf{3}}_\text{A}) \otimes \mathbf{3} = \mathbf{10}_\text{S} \oplus \mathbf{8}^{(2)}_\text{M} \oplus \mathbf{1}_\text{A} \,. \tag{1.12}$$

Here, the subscripts A, S and M denote antisymmetric, symmetric and mixed symmetric flavour representations respectively, and the superscript denotes the multiplicity of a multiplet. The weight diagrams of the decomposed irreducible representations are shown in Fig. 1.4.

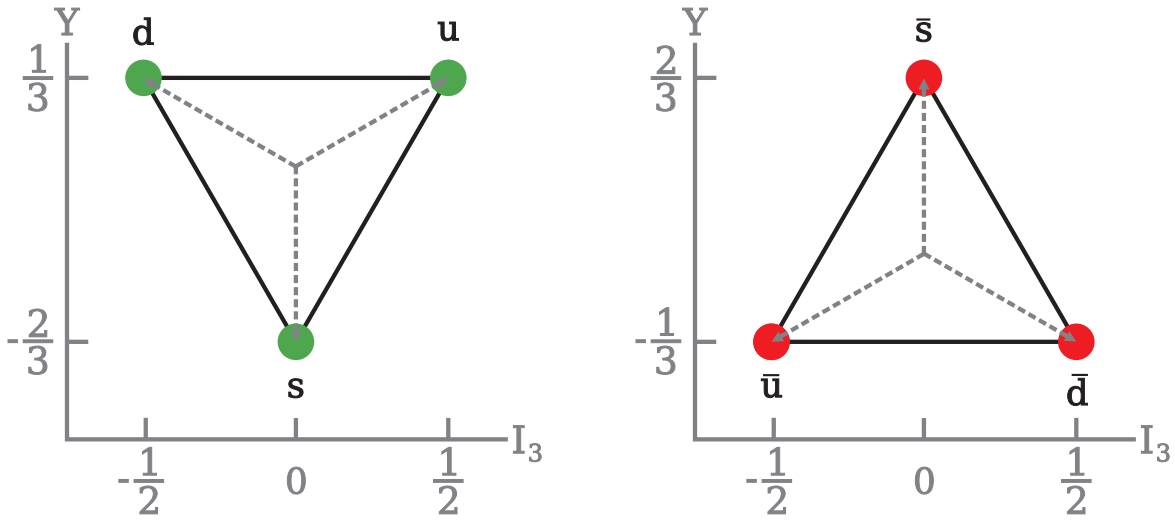


Figure 1.3: Weight diagram of quarks and antiquarks. (a) The fundamental **3** representation of the $SU(3)_f$ group. (b) Weight diagram for the $\bar{\mathbf{3}}$ representation of $SU(3)_f$.

To relate observed states of mesons and baryons to the $SU(3)_f$-multiplets, their flavour wave function needs to be coupled to possible spin, spatial and colour wave functions which define the quantum numbers of the system. In addition, radial excitations emerge when treating hadrons as (effective) two-body systems in a spatially dependent potential. Due to confinement, the colour wave function has to be an antisymmetric singlet. Since colour and flavour posses the same group structure, the decompositions (1.11) and (1.12) apply to colour as well, but only the $\mathbf{1}_A$ representation is realised in nature. The spatial wave function depends on the orbital angular momentum $\ell \in \mathbb{N}^0$, and is either symmetric (positive parity $P = (-1)^\ell$) or antisymmetric (negative P). Quark spin has a SU(2) group structure and needs to be coupled to the flavour state according to representation theory. Since baryons are fermions, they need to obey Fermi-Dirac statistics, which requires an antisymmetric wave function.

With similar arguments, flavour symmetry can be extended to all quark flavours – apart from the top quark, which decays before it can form a hadron due to its mass – and the lowest lying states follow the (qualitative) expectations from the quark model. However, the quark model predicts a large number of excited states, which are not seen in experiments, *cf.* the review on the quark model in [14]. Especially if four- and five-quark configurations are taken into account, the number of expected states seems to explode.

On the other hand, only relatively few candidates for these states exist. Thus, understanding the underlying mechanisms which allow or forbid the existence of these hadrons will provide valuable insights to QCD phenomenology. Of course hadronic properties which are derived from flavour symmetry only treat valence quarks, and any predictive model needs to account for the descriptions of underlying dynamic processes, either through models or directly from QCD itself.

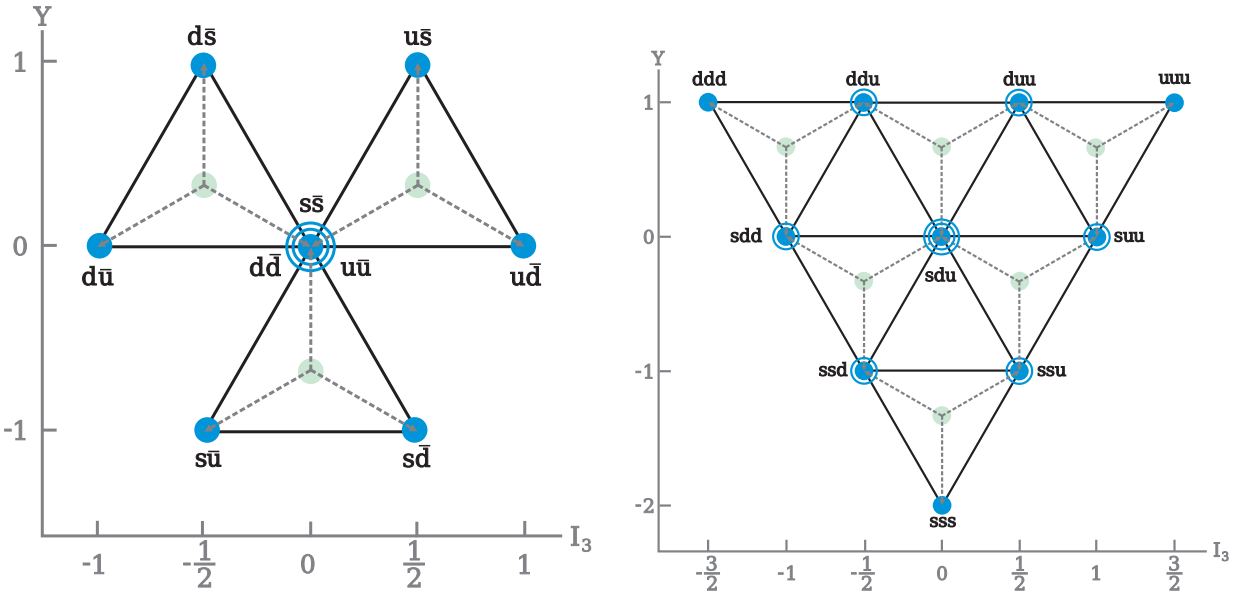


Figure 1.4: Multiplets of $SU(3)_f$-symmetric combinations of $q\bar{q}$ (a) and qqq (b). These are the weight diagrams corresponding to the flavour wave functions of the states.

1.3 Pentaquarks

This section sketches the field of exotic hadron spectroscopy with a focus on experimental efforts, notably pentaquark searches. Eventually, implications in the search for $uudc\bar{c}$ pentaquarks in the Λ_c^+ (udc) $\bar{D}^{(*)0}$ $(\bar{c}u)$ subsystem of the decay $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ are discussed.

The search for hadrons that manifestly contain more than three quarks is as old as the quark model itself – both Gell-Mann and Zweig anticipated the existence of tetra- penta- and even higher multi-quark states in their ground-breaking publications [23, 24, 25]. First searches for pentaquarks – called Z^+ or Z^0 baryons at the time – were carried out in kaon nucleon scattering experiments, where resonant structures in partial waves with exotic quantum numbers, *i.e.* baryon number +1 and positive strangeness, were searched for. Even though six candidates with weak evidence have been listed in the 1986 edition of the review of particle physics [71], the claims were seen with scepticism: *"the standards of proof must simply be much more severe here than in a channel in which many resonances are already known to exist."*.

The subject disappeared off the radar until 2002, when the LEPS collaboration claimed evidence for a light and narrow $uudd\bar{s}$ resonance, called Θ^+ [72]. The search was motivated by the prediction of a light (1530 MeV) and narrow (< 15 MeV) pentaquark in the framework of a chiral quark soliton model five years earlier [73]. The LEPS paper triggered an avalanche of phenomenological studies, most of which required considerable fine-tuning to accommodate the Θ^+, but it also triggered many experimental searches. About a dozen of experiments found evidence for the Θ^+, while an overwhelming majority did not see the resonance. In retrospective, the claimed Θ^+ signals are commonly assumed to be caused by either statistical fluctuations, kinematic cuts, reflections, experimental artefacts, or a combination thereof. The history and fate of the Θ^+ is comprehensively summarised in Ref. [74].

As pentaquarks were about to disappear a second time, the Belle collaboration observed a puzzling state in the study of $B^\pm \to J/\psi\, \pi^+\pi^- K^\pm$ decays [75], termed $X(3872)$. It is a narrow resonance, consistent with the experimental resolution, in the $J/\psi\, \pi^+\pi^-$ channel, directly at the $D^{*0}\bar{D}^0$ threshold.

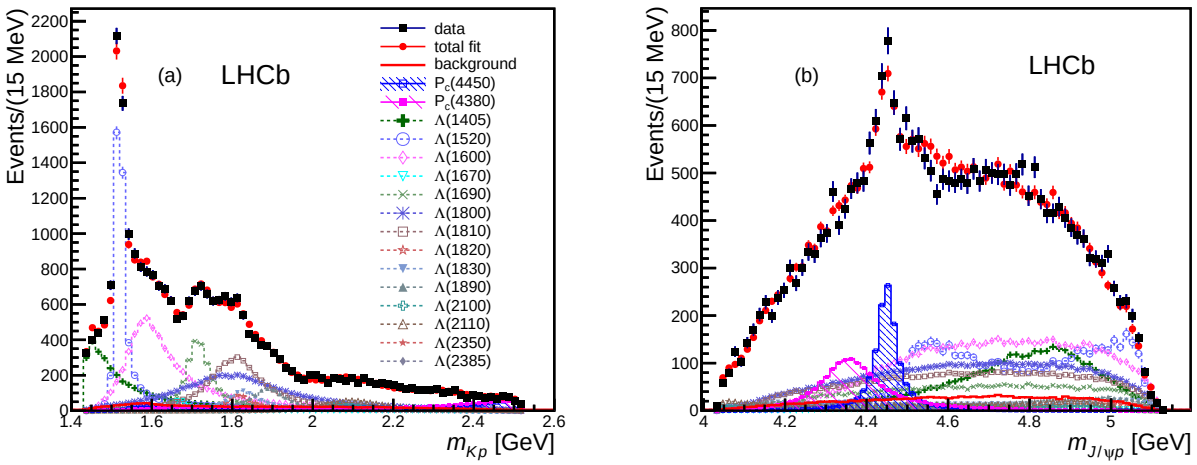


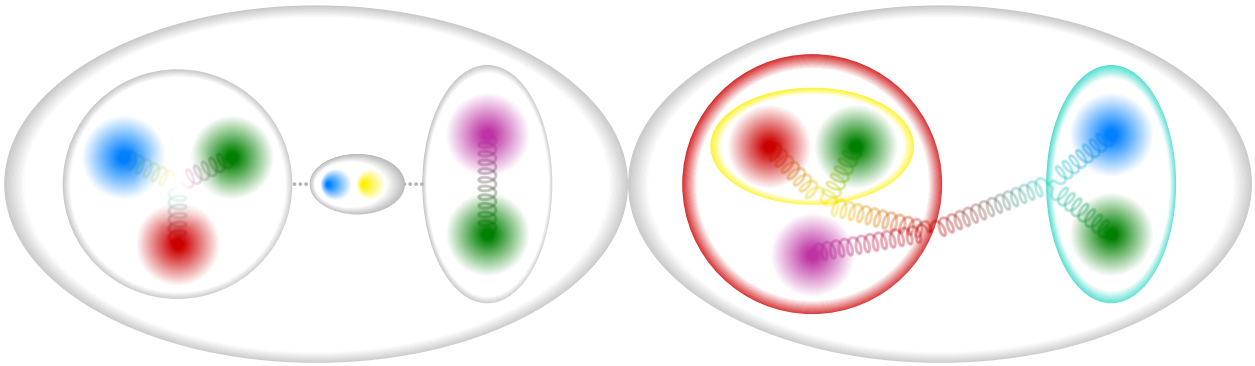
Figure 1.5: Fit projections for the (a) pK^- and (b) $J/\psi p$ invariant mass systems of a six-dimensional helicity amplitude analysis including the two P_c^+ states. The uncertainties on the fit results are due to simulation statistics.

Due to this peculiar mass and the narrow width, the state did not fit into the conventional spectrum of charmonium ($c\bar{c}$) resonances. In the following years, the Belle observation has been confirmed, and further charmonium-like resonances, that are incompatible with quark model expectations have been observed. Among them, charged resonances with a $c\bar{c}$ component – an unambiguous indication of a state consisting of at least four quarks.

Even though tetraquarks were more or less established, pentaquarks were still approached with scepticism. Hence, the observation of two resonances in the $J/\psi p$ system by LHCb in 2015 came as a big surprise [76]. The resonances, denoted as $P_c(4380)^+$ and $P_c(4450)^+$, have been discovered in the exclusive $\Lambda_b^0 \to J/\psi pK^-$ decay by a six-dimensional helicity amplitude analysis, describing the full kinematics and the angular structure of the Λ_b^0 decay. The analysis used efficiency corrected and background subtracted data, to which the helicity amplitude model was fit. The fit parameters of the model are the helicity couplings of the resonances, which were at first only Λ^* resonances in the pK^- subsystem. They were described by relativistic Breit-Wigner shapes with fixed masses and widths. However, fits with only Λ^* resonances did not describe the data well. Also a fit with a single exotic component was not satisfactory. Eventually, a fit with two exotic components in the $J/\psi p$ subsystem lead to a good result, for which the pK^- and $J/\psi p$ mass projections are shown in Fig. 1.5.

The extracted parameters from the helicitly amplitude analysis were the following: The $P_c(4380)^+$ has a mass of $4380 \pm 8 \pm 29\,\text{MeV}$ and a width of $205 \pm 18 \pm 86\,\text{MeV}$, while the $P_c(4450)^+$ is narrower, with a mass of $4449.8 \pm 1.7 \pm 2.5\,\text{MeV}$ and a width of $39 \pm 5 \pm 19\,\text{MeV}$. The preferred spin-parity (J^P) assignments are of opposite parity, with one state having spin $3/2$ and the other $5/2$. The fit fractions were measured to be $8.4 \pm 0.7 \pm 4.2\,\%$ for the $P_c(4380)^+$ and $4.1 \pm 0.5 \pm 1.1\,\%$ for the $P_c(4450)^+$.

Further, a model independent analysis of the $\Lambda_b^0 \to J/\psi pK^-$ decay was carried out [77]. The study supports the need for at least one exotic component to describe the data, and measured the significance of that component, consistent with the narrow $P_c(4450)^+$, to be grater than $9\,\sigma$. The $P_c(4380)^+$, due to it's large width, could not be confirmed by the model independent approach. A result consistent with the $\Lambda_b^0 \to J/\psi pK^-$ helicity amplitude analysis has been obtained in it's Cabibbo suppressed analogue, the $\Lambda_b^0 \to J/\psi p\pi^-$ channel [78]. However, due to fewer statistics, the P_c^+ components could not be singled out. Still, evidence was found for exotic contributions: either the P_c^+s or the $Z(4200)^-$ in $J/\psi\pi^-$, or both. The $Z(4200)^-$ is one of the charged charmonium-like resonances mentioned earlier and has been discovered by Belle in $B^0 \to J/\psi K^+\pi^-$ decays. [79].



(a) (b)

Figure 1.6: Illustration of a pentaquark as bound system of diquark and triquark (a) or as meson-baryon molecule (b).

In summary, the P_c^+'s decay channel, measured masses, widths and fit fractions clearly indicate that the states have a minimal quark content of $uudc\bar{c}$ in the quark model.

The community responded lively to the observation of the P_c^+s. In contrast to the light and narrow Θ^+, the P_c^+ states fit into the models in a natural way. The most vividly discussed interpretation of the P_c^+'s nature are baryon-meson molecules or strongly bound systems, like genuine pentaquarks, diquark-diquark antiquark or diquark-triquark combinations. Two of these interpretations are illustrated in Fig. 1.6. The quark content suggests a decay to Λ_c^+ $(udc)\bar{D}^{(*)0}$ $(\bar{c}u)$ if they are in fact real resonances and not be due to kinematic effects like cusps or anomalous triangle singularities, as *e.g.* hypothesised in Ref. [80]. Firm predictions whether the P_c^+ states are expected to decay into $\Lambda_c^+\bar{D}^{(*)0}$ exist for the molecular interpretation, depending on the molecular composition as shown in Tab. 1.2. The molecular compositions have characteristically different decay patterns, so that the observation of P_c^+ states in different channels is crucial to distinguish them.

$P_c(4380)^+$ decays/partial widths (MeV)				
Mode \ Composition	$\Sigma_c(2520)\bar{D}$		$\Sigma_c\bar{D}^*$	$J/\psi N(1440)$
$\Lambda_c\bar{D}$	[✗]	1.2	13.7	✗
$\Lambda_c\bar{D}^*$	✓	110.4	28.6	✓
$J/\psi N$	✓	2.7	19.8	✓
$\Sigma_c(2455)\bar{D}$	[✗]	0.01	0.09	✗
$\Lambda_c\bar{D}\pi$	✓	7.5	–	✗
$\eta_c N$	✗	0.2	0.05	✗
$\eta_c\Delta$	✓	–	–	✗
$J/\psi\Delta$	✗	–	–	✗
$J/\psi N\pi$	✓	–	–	✓
ωp	–	6.1	1.3	–

$P_c(4450)^+$ decays/partial widths (MeV)					
$\chi_{c1}p$	$\Sigma_c\bar{D}^*$		$\Lambda_c(2595)\bar{D}$	$J/\psi N(1520)$	
✓	[✗]	14.9	[✓]	✗	
✓	✓	16.3	[✓]	✓	
✓	✓	2.6	✓	✓	
✓	[✗]	0.2	✓	✗	
✓	✓	0.5	[✗]	✓	
✗	✗	0.02	✓	✗	
✗	✓	–	✗	✗	
✗	✓	–	✗	✗	
✗	✓	–	✗	✓	
–	–	0.4	–	–	

Table 1.2: (Qualitative) predictions of decay widths for a hypothesised molecular composition of the P_c^+s. Composed from Refs. [81] and [82]. Particle charges are omitted to retain compatibility with the publications, which anticipate isospin parters of the P_c^+s. Hence, $\bar{D}^*$ means $\bar{D}^*(2007)^0$ or $D^*(2010)^-$.

Apart from the decay to $\Lambda_c^+ \overline{D}^{(*)0}$ the P_c^+ pentaquarks can be searched for in further decays including the $J/\psi\, p$ subsystem, such as $B_s^0 \to J/\psi\, p\overline{p}$, $\Xi_b^0 \to J/\psi\, pK^-$ or $\Xi_b^- \to J/\psi\, pK^-K^-$. To probe properties of the P_c^+ states, decays to different final states, *e.g.* $\Lambda_b^0 \to \chi_{c1} pK^-$, $\Lambda_b^0 \to \eta_c pK^-$, $\Lambda_b^0 \to \Sigma_c^{++} D^- K^-$ are of interest. And even more: the discovery implies a tower of pentaquarks with different quark content. All of these searches are feasible with the LHCb experiment and will be carried out in a broad programme of exotic hadron spectroscopy.

Recent reviews put the observed P_c^+ pentaquarks into context of exotic hadron spectroscopy by discussing experimental findings, theoretical techniques and possible interpretations [83, 84]. Some of the striking open questions that have been identified in these reviews are the following:

- No picture/model is able to explain all observed exotic hadrons. Does it exist[1]?
- All unambiguously identified exotic hadrons contain $c\overline{c}$ or $b\overline{b}$. Do light, or open flavour exotic hadrons exist? If not, why are they forbidden?
- Is it a coincidence that most states are observed close to thresholds?

A community white paper has been set up in order to approach these and many other open questions systematically from both, experimental and theoretical side [86].

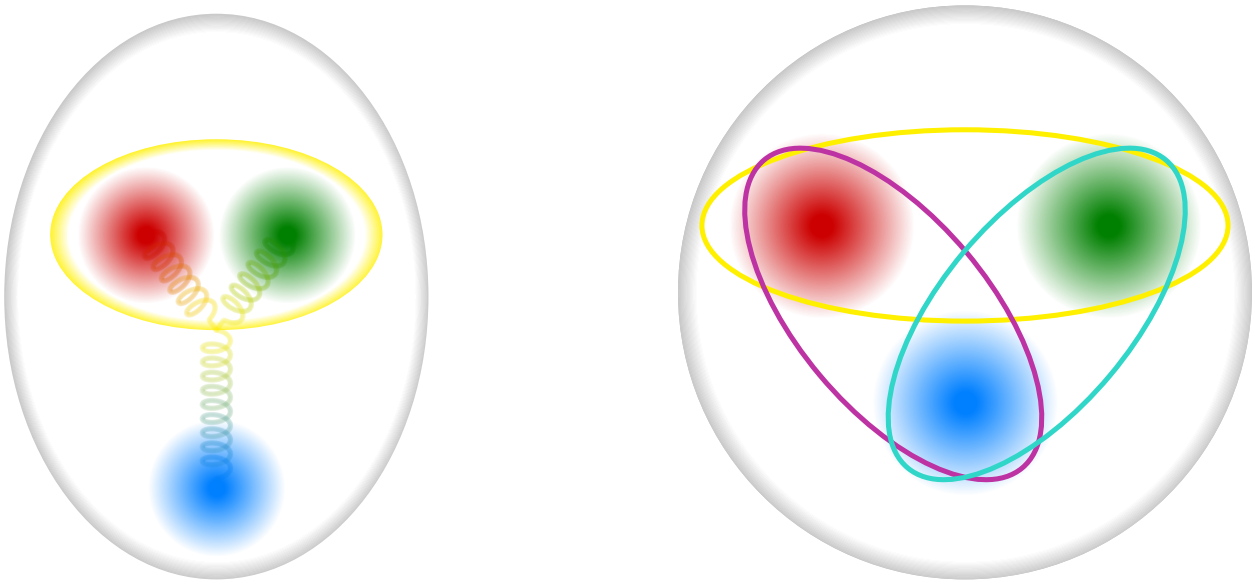
1.4 Pentaquark searches in a broader context

Fitting pentaquarks into the hadronic spectrum is a major challenge. It eventually requires a common understanding of mesons, baryons and QCD exotica in a framework that connects the effective pictures described above to the pure Yang-Mills sector of QCD. Finding this connection is an outstanding theoretical challenge for which the extension of the hadronic spectrum to exotic states can be valuable input. The essential questions that need to answered in this process will be those of confinement, dynamical chiral symmetry breaking (DCSB), and their relation [67]. If these concepts are understood, QCD could be a self-contained *true* theory, *i.e.* a theory that is valid over all scales by confinement and asymptotic freedom, and which naturally generates a mass scale by DCSB. This paradigm could also be applied to beyond standard model physics, in which the electroweak sector is generated dynamically; an example of such a model is extended technicolor [87].

Back to the hadronic spectrum: The best understood part of it is certainly the heavy quarkonium spectrum – $b\overline{b}$ and $c\overline{c}$. It is well described by non-relativistic quarks in a static potential [88]; usually pictured as a flux-tube forming colour field between quark and antiquark. However, the picture of flux tubes and linear potentials in the light meson spectrum, *e.g.* in the form of Regge trajectories [89], breaks down [90]. It is however debatable how to interpret Regge trajectories in the light spectrum, since they work for the majority of the spectrum; see *e.g.* Ref. [91]. Finding a coherent picture for light mesons is challenging; a possibility is that quark and antiquark are dynamically dressed by gluons and the Goldstone modes of the chiral quark condensate. In any case, light quarks move at relativistic velocities and can therefore not act as static sources of potentials. Moreover would light-particle creation and annihilation effects disturb the propagation of gluonic fields, such that a flux-tube would dissolve well within a hadron's interior.

If the simple constituent quark picture of light mesons does not hold true, which impact does it have on baryons? In the constituent quark plus flux-tube picture, baryons may be bound by a Y-shaped flux-tube as shown in Fig. 1.7(a). The illustration also shows a static diquark, a coloured, but confined qq constituent for which evidence has been found in lattice QCD simulations [92, 93].

[1] Gell-Mann's totalitarian principle states: "Everything not forbidden is compulsory.", *i.e.* models are expected to mix. Such hybrid models are *e.g.* currently discussed for the $X(3872)$ [84] or the Roper resonance [85].



(a) (b)

Figure 1.7: Illustration of baryons composed of a static diquark and a single quark (a) and an entangled set of dynamic diquarks (b) in the fashion of Borromean rings. Diquarks are continuously broken up and recreated, and so contributing to binding.

The static diquark is shown to distinguish it from a modern notion of dynamic diquarks, which are to be seen as extended objects with internal structure [94]. Recently the picture of the nucleon as a Borromean bound-state (*cf.* Fig. 1.7(b)) has been proposed [95]. Whichever picture prevails in the end, it has to be embedded into the searched for framework linking to the Yang-Mills sector of QCD.

In that regard, tetraquarks and pentaquarks with both, heavy and light constituents bridge the gap between effective theories of the charmonium regime and the approaches to nonperturbative continuum QCD. For instance, diquark correlations and flux tubes both play a key role in a picture in which exotic hadrons are described as composed of dynamically produced diquarks [96, 97]. It seems natural that such a model predicts a different production cross-section compared to a molecular picture. Can such models be used to explain why resonances like the $Z_c(3900)^+ \to J/\psi\,\pi^+$ [98] or the $Z_c(4055)^+ \to \psi(2S)\pi^+$ [99] were only be observed in e^+e^- production and not in b decays, where they should have been seen in high statistics analyses of $\bar{B}^0 \to J/\psi\,K^-\pi^+$ [79] or $\bar{B}^0 \to \psi(2S)K^-\pi^+$ [100]? Will the P_c^+ pentaquarks be seen in the proposed photoproduction experiment at JLab [101]?

The main points are briefly summarised. Exotic hadrons allow to study the effective degrees of freedom in hadrons from a new perspective. This will in first instance lead to a better effective description of the hadronic spectrum and consequently QCD. This is desirable by itself, since understanding QCD effects often limits measurements in which they are a nuisance. The more fundamental points, such as confinement and DCSB, can however only be addressed with advances in the non-perturbative regime of QCD from theoretical side. A better understanding of this regime, and eventually the link to the pure Yang-Mills sector, might be gained from effective models. From the experimental side, it is thus crucial to provide broad – finding as many states as possible – and detailed – examining properties of the states – measurements of the hadronic spectrum. The work presented in this thesis is part of a programme of exotic hadron spectroscopy at LHCb, and contributes a small amount to the understanding of QCD in this context.

2 The LHCb experiment

First the hardware of the LHCb spectrometer is introduced in a generic way. Afterwards, the flow of information is followed from proton-proton collisions to a dataset ready for analysis. The focus for that part lies on four key detector performance figures driving the sensitivity of the analysis shown later. These are: trigger efficiency, spatial resolution for good reconstruction of vertices, momentum resolution for clean mass reconstruction and particle identification.

Note that the analysis presented in this thesis uses a dataset corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$ recoded during Run I of the LHC in 2011 and 2012. Therefore, the descriptions given here match the conditions during Run I.

The LHCb experiment is one of currently seven experiments at the Large Hadron Collider (LHC) experiment hosted at the European Council for Nuclear Research (CERN, from "Conseil Européen pour la Recherche Nucléaire") astride the Franco-Swiss border near Geneva. LHCb is a dedicated heavy flavour physics experiment, whose main goal is to search for indirect evidence of processes beyond the standard model of particle physics (SM) in *CP* violation and rare decays of beauty and charm hadrons. Given the experiments success, it is often also referred to as general purpose detector in the forward region. This work, being part of a programme of exotic hadron spectroscopy, and many other unprecedented analyses support this notion.

2.1 The LHCb spectrometer

The spectrometer design suits the predominant production mechanism of heavy flavour hadrons at the LHC via gluon fusion, which boosts the produced system in either forward or backward direction. It has a forward angular coverage from approximately 15 to 300 (250) mrad in the bending (non-bending) plane. The spectrometer layout is shown in Fig. 2.1. Most detector subsystems are assembled in two halves, which can be moved out horizontally for maintenance purposes, as well as to provide access to the beam-pipe. They are referred to as the detector A- and C-sides. Further, the spectrometer is conceptually separated into an upstream part from the vertex locator (VELO) to the magnet, and a downstream part from the magnet to the muon stations. A right-handed coordinate system is defined with z along the beam axis from the interaction point into the detector and y vertical upward. Positive (negative) z-direction is also referred to as downstream (upstream).

The core physics programme of LHCb [102] requires excellent vertex, decay time and momentum resolution, precise particle identification, and a versatile trigger scheme providing a fast and efficient event selection with low misclassification rate. These requirements are reflected in the design of the various subdetectors which are briefly introduced in the following. A detailed comprehensive description of the spectrometer can be found in [103], while its performance during Run I is reviewed in [104].

2.1.1 Magnet

The spectrometer magnet, described in Ref. [105], is needed for a precise measurement of the momenta of charged particles. It is a warm dipole magnet with saddle-shaped coils in a window-frame yoke with sloping poles to match the detector acceptance. It provides an integrated field of about 4 Tm for tracks of 10 m, accommodating the contrasting needs of low magnetic fields in the RICHs envelope and a field as high as possible between the VELO and the Trigger Tracker (TT). The magnetic field deflects charged particles in the horizontal ($x-z$) plane, and also has an impact on the trajectory of the LHC beams; this is compensated for with three dipole magnets [106].

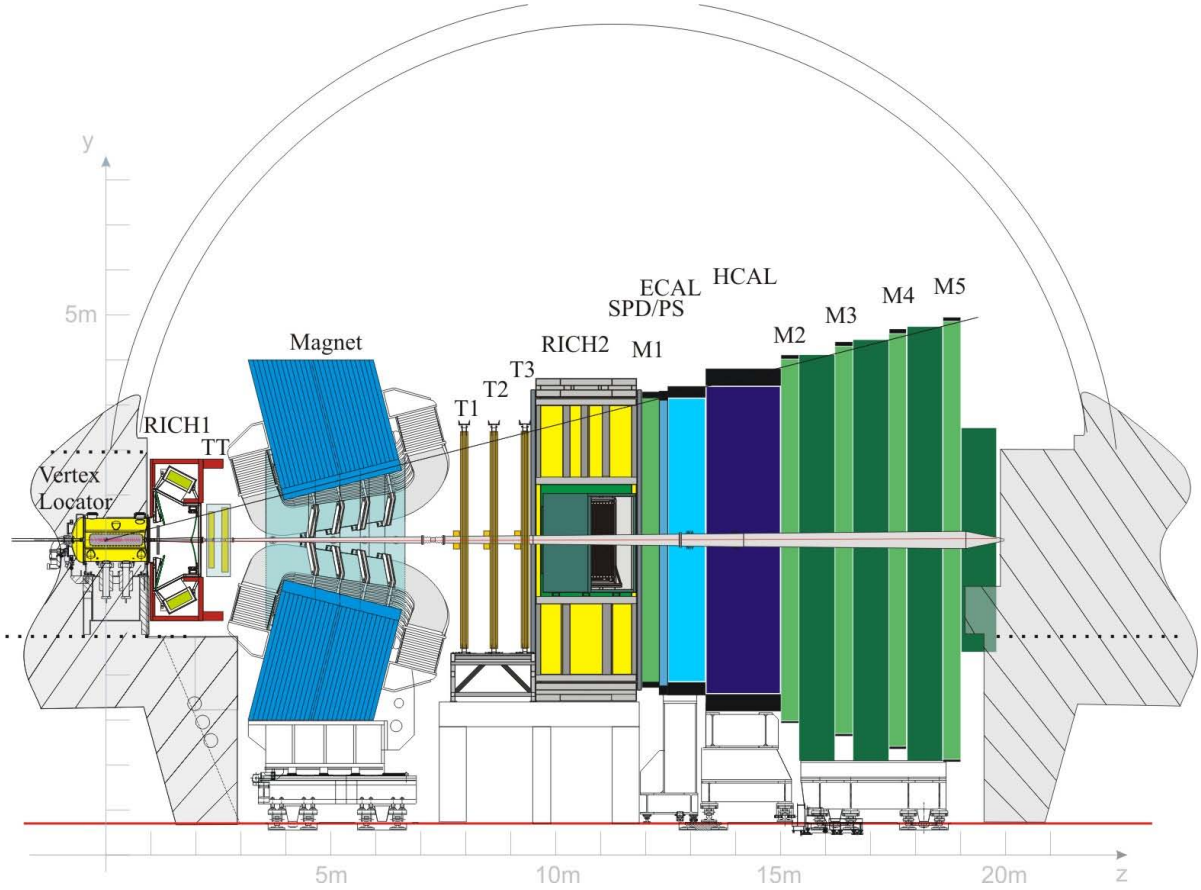


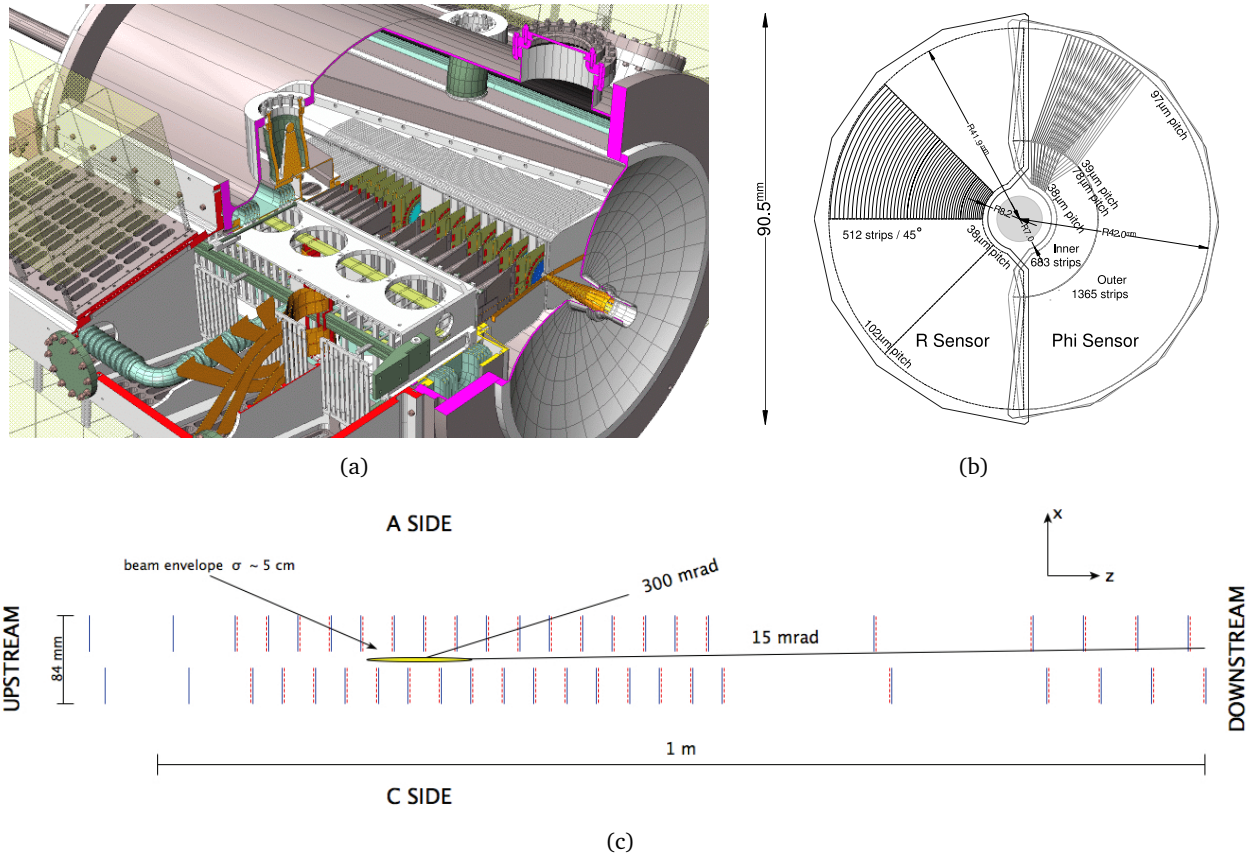
Figure 2.1: Schematic side view of the LHCb spectrometer.

To achieve precise track reconstruction, the magnetic field integral is determined with permille-level precision using Hall probes [I1]. Beside advantages in track reconstruction, the magnet is used to achieve higher precision in measurements sensitive to detection asymmetries, such as *CP* violation measurements. This is achieved by inverting the magnet polarity, which was done about every second month during Run I.

2.1.2 Vertex locator

The VELO measures track coordinates close to the interaction region, and is used to identify primary interaction vertices and secondary vertices from heavy flavour hadrons. Their average flight distance is about 1 cm due to the heavy quark production mechanism and the strong boost at LHC energies.

The VELO is a silicon microstrip [107] detector operated in vacuum with sensors moving as close as 7 mm from the LHC beam. This distance to the beam is smaller than required by LHC operations during beam injection. Therefore the detector is constructed such, that its halves are moveable in x and y and the VELO is closed at the beginning of each fill. Due to the proximity to the beam, there is no beam pipe in the interaction region. In order to maintain the LHC machine vacuum, it is separated from the detectors by an RF box. The detectors are thus located in a secondary vacuum, mounted within a vacuum vessel (*cf.* Fig. 2.2(a)). The RF box surfaces facing the beam are 0.3 mm thick corrugated sheets, known as the RF foil. The VELO is operated in an extreme radiation environment, requiring radiation tolerant technology [108, 109], and cooling of the sensors to $(-7 \pm 2)^\circ$C [110].



(a) (b)

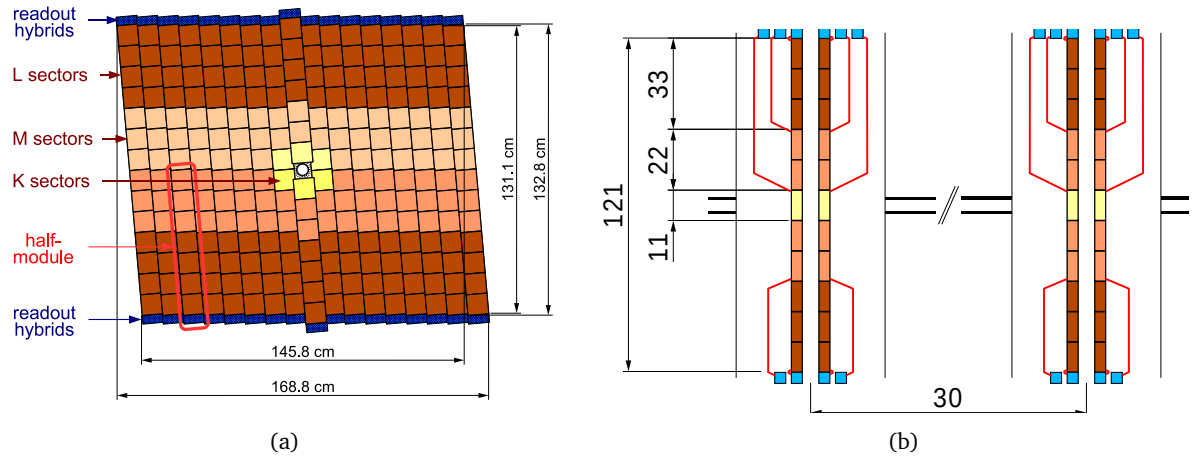
(c)

Figure 2.2: (a) Schematic view of the interaction region with the VELO vacuum tank being cut open, showing the VELO sensors, hybrids and module support [110]. (b) Sketch of a VELO sensor in the $x-y$ plane illustrating its $r-\phi$ geometry [103]. In the ϕ sensor, two adjacent modules are shown to highlight the stereo angle between them. (c) Sketch of the arrangement of sensors in the $x-z$ plane. $R(\Phi)$ sensors are shown with solid blue (dashed red) lines [110].

The VELO contains 42 silicon modules arranged along the beam, as shown in Fig. 2.2(c), each providing a measurement of the r (R sensors) and ϕ (Φ sensors) coordinates. An R sensor is subdivided into four segments of approximately 45°, while Φ sensors have two segments. The inter-strip pitch varies from approximately 40 to 100 μm across the sensor (*cf.* Fig. 2.2(b)). depending on the strip pitch and the projected angle, the measured hit resolution ranges from 4 μm to approximately 30 μm. The resolution is correlated to the number of fired strips by a traversing particle. Further details on the VELO performance during Run I are given in [110].

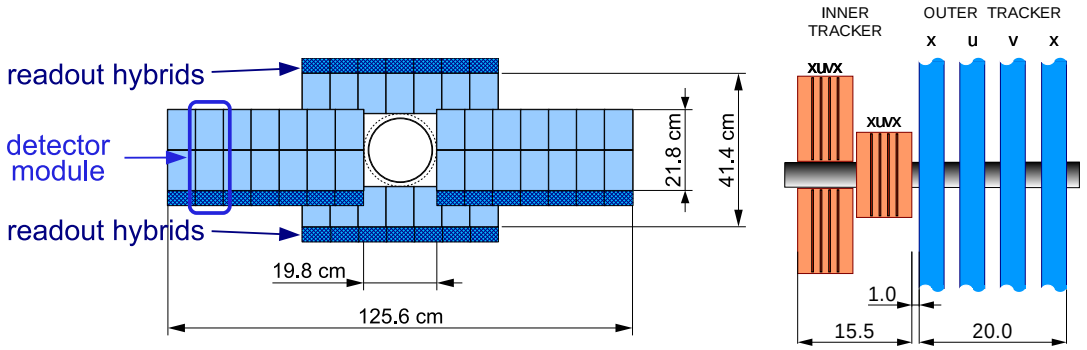
2.1.3 Silicon trackers

The silicon trackers comprise the Trigger Tracker (TT) and the Inner Tracker (IT), which are located upstream and downstream of the LHCb dipole magnet respectively. Both detector systems use silicon microstrip sensors with a pitches of 183 μm in the TT and 198 μm in the IT. The TT consists of one, the IT of three tracking stations, in which the individual detection layers are arranged in an x–u–v–x layout to increase the momentum resolution of reconstructed tracks. The strips of the first (x) layer are vertically aligned, while the u and v stereo layers are rotated by $-/+5°$, respectively.



(a) (b)

Figure 2.3: Schematic front (a) and side (b) views of the TT [103, 112].



(a) (b)

Figure 2.4: Schematic front (a) and top (b) views of the IT [103, 112].

The TT covers the full acceptance upstream of the magnet, while the IT is situated in a 120 cm by 40 cm cross-shaped region around the beam pipe with the highest density of tracks corresponding to around 1.2 % of the total acceptance. The total sensitive area in the TT and IT is 8 m^2 and 4.2 m^2 respectively. Each of the TT and IT tracking stations are housed in a light tight and thermally and electrically insulated detector volume, in which a temperature below 5°C is maintained.

The half-modules of the TT are subdivided into different readout sectors – shown for the *v* layer in Fig. 2.3(a) – to compensate for higher particle fluxes close to the beam pipe. Each of the sectors is read out by a front-end hybrid [111]. These hybrids are stacked at the outer end of a half-module and connected by Kapton cables to the inner sectors as illustrated in Fig. 2.3(b).

The IT modules consist of either one or two silicon sensors as shown in Fig. 2.4(a) and are mounted directly in front of the outer tracker modules in each tracking station (*cf.* Fig. 2.4(b)). The hit resolution is measured to be approximately 50 µm using Run I data [104], depending on the occupancy of the region. The performance of the silicon trackers during Run I is reviewed in Ref. [113].

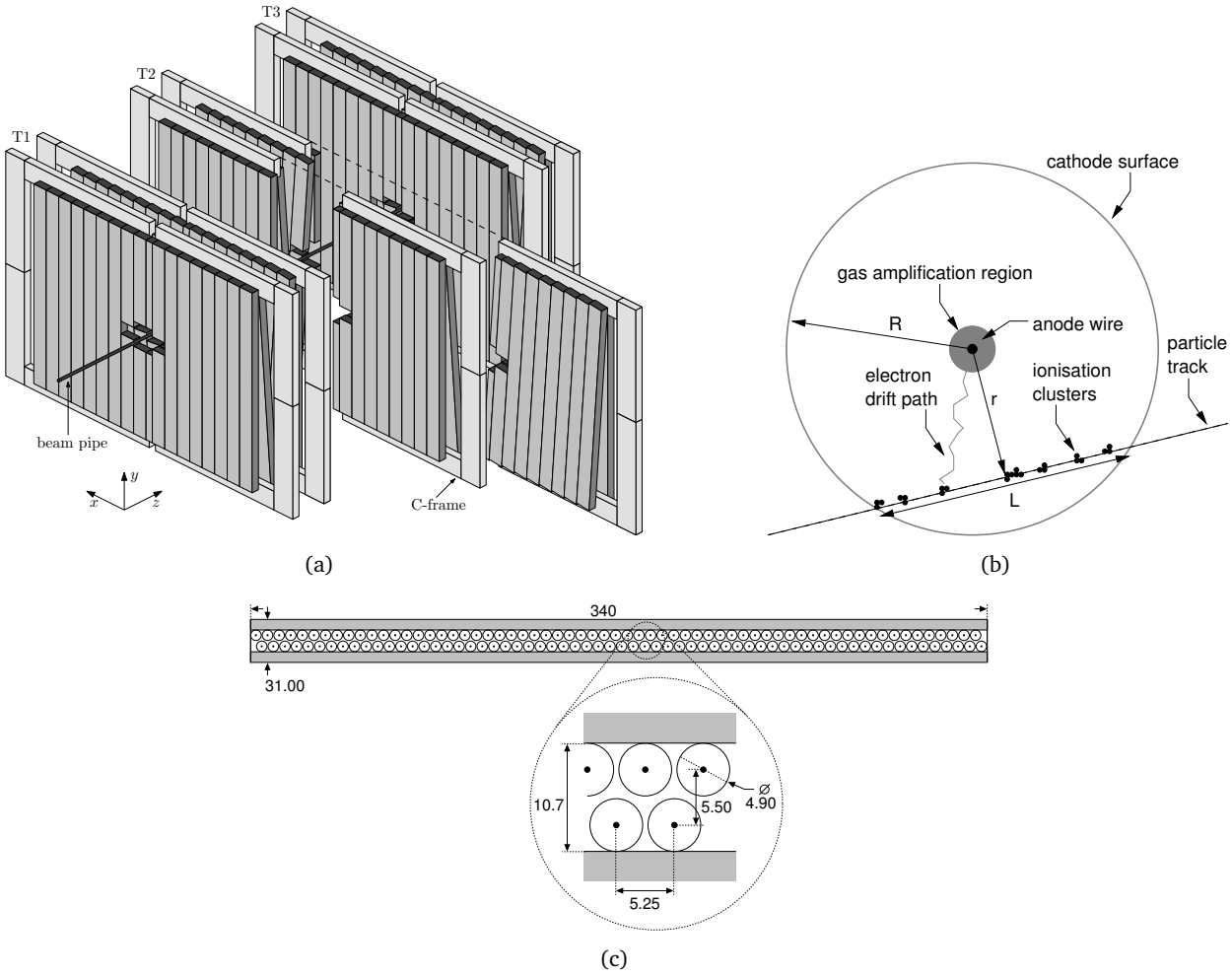


Figure 2.5: (a) Arrangement of OT modules in layers and stations. (b) Sketch of a charged particle that traverses a straw. (c) Cross section of a straw-tube module. All taken from Ref. [116]

2.1.4 Outer tracker

The largest area of the tracking stations T1–T3 downstream of the LHCb magnet is covered by modules of the outer tracker (OT). The OT is a drift-tube gas detector, described in Ref. [114]. As for the IT, the OT has four layers arranged in an x–u–v–x geometry. The total active area of a station is 597 cm × 485 cm. Each detector layer is built from 14 long and 8 short gas-tight straw-tube modules with drift-time readout, shown in Fig. 2.5(a). Each module contains two staggered layers of drift-tubes with an inner diameter of 4.9 mm, shown in Fig. 2.5(c). The drift-tubes are used to measure the arrival time of the OT readout signals with respect to the LHC clock to infer the drift distance r, and hence the hit position, as illustrated in Fig. 2.5(b). The signal is read out once per 25 ns on a positive hardware trigger signal, such that a counting gas that guarantees fast drift times and prevents ageing effects on the anode wires was chosen [115].

The occupancy of the OT, especially in the inner regions, was larger than originally anticipated, due to twice larger instantaneous luminosity at LHCb (*cf.* Sec. 2.2.2). Despite these challenging conditions, the design single hit resolution of about 200 µm could be reached. And it can even be further improved by taking horizontal displacement per half monolayer during alignment into account. A detailed review of the OT performance is given in [116].

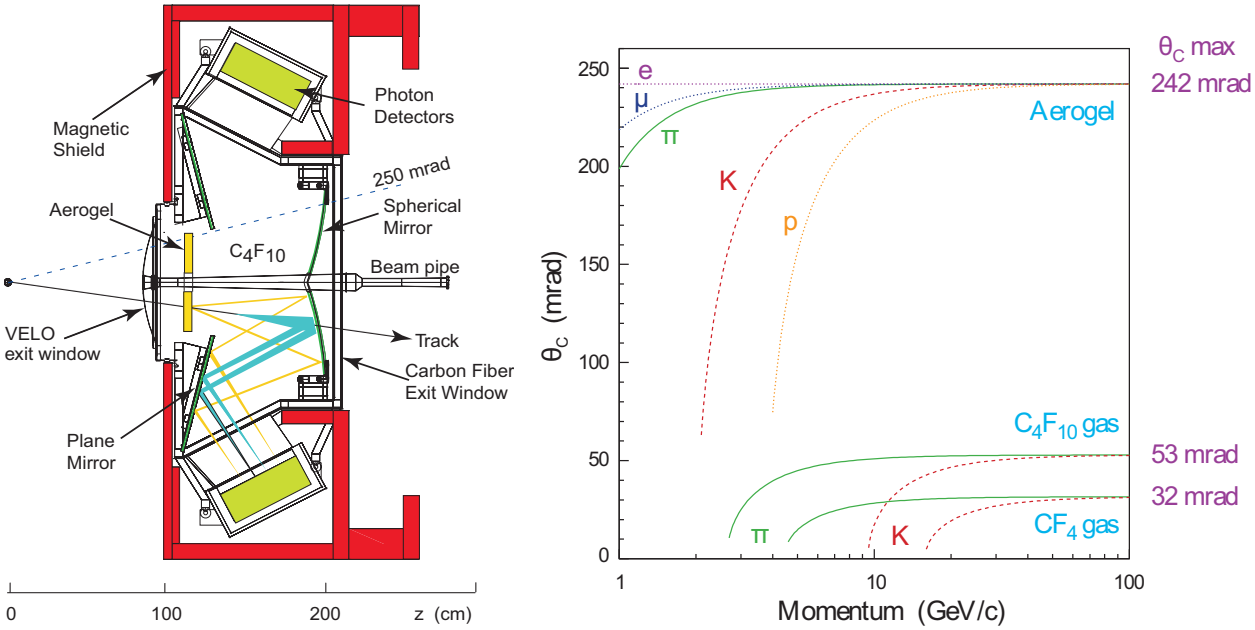
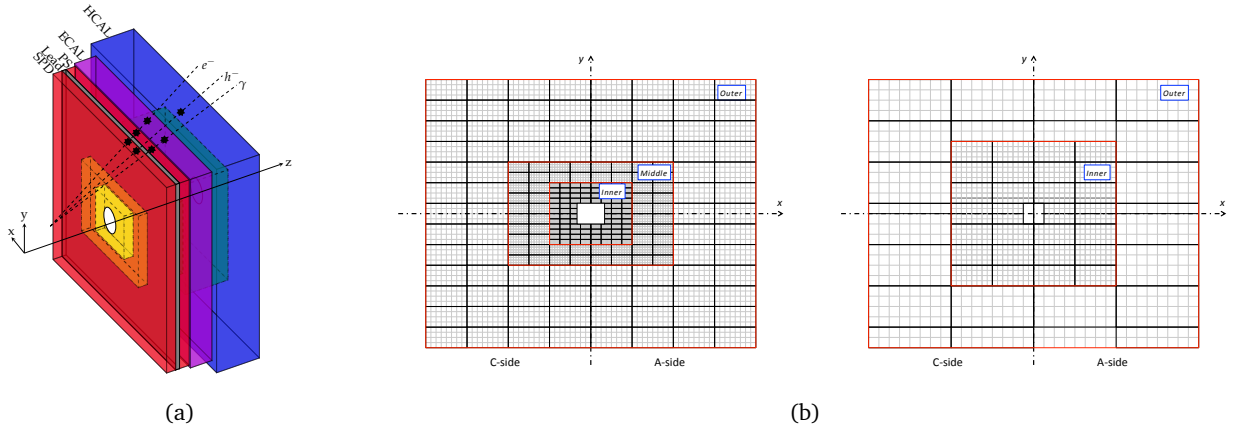


Figure 2.6: (a) Schematic side view of the RICH1 detector. Taken from [103]. (b) Cherenkov angle as a function of the momentum of different particles for the RICH radiators.

2.1.5 RICH

Identification of charged hadrons (π, K, p) in the momentum range from 2 to 100 GeV is primarily achieved by two ring imaging Cherenkov detectors (RICH1 and RICH2). In addition, the RICH system contributes to the identification of charged leptons (e, μ), complementing information from the calorimeter and muon systems. Its information is crucial in rejecting combinatorial physics backgrounds.

RICH1, located immediately downstream of the VELO, covers the low and intermediate momentum region from 2 to 40 GeV over the full spectrometer angular acceptance. While RICH2, located downstream of the tracking stations, covers the high momentum region from 15 to 100 GeV over an 15 to 120 mrad range of acceptance. The RICH detectors use low dispersion fluorcarbon gases at room temperature – C_4F_{10} in RICH1, CF_4 in RICH2– in which high energetic particles emit Cherenkov radiation in a cone around their trajectory. Both detectors have a similar optical system to project the Cherenkov photons onto an array of hybrid photon detectors [117]. It consists of a tilted spherical focusing primary mirror and a flat secondary mirror to limit the size of the detector along z. The layout of RICH1 is shown in Fig. 2.6(a). An additional radiator made of 16 tiles of silica aerogel was used in RICH1 during Run I to be able to separate kaons and protons at low momentum. The momentum-dependence of expected Cherenkov angles for given particle types and radiators are shown in Fig. 2.6(b). In order to distinguish between different types of hadrons, the Cherenkov angle needs to be measured precisely. Its resolution was determined to be 1.618 ± 0.002 mrad for C_4F_{10}, 0.68 ± 0.02 mrad for CF_4 and about 5.6 mrad (FWHM) for aerogel. After RICH reconstruction, the resolutions translate into separation power to distinguish particle types, as shown later in Sec. 2.2.5 More details on the RICH detector and reconstruction performance during Run I is given in [118].



(a) (b)

Figure 2.7: (a) Arrangement of calorimeters. The z scale of SPD/PS is exaggerated with respect to ECAL/HCAL [122]. (b) Lateral segmentation of the SPD/PS and ECAL (left) and the HCAL (right). Grey lines indicate boundaries of individual cells, black lines correspond to readout segments. [123]

2.1.6 Calorimeters

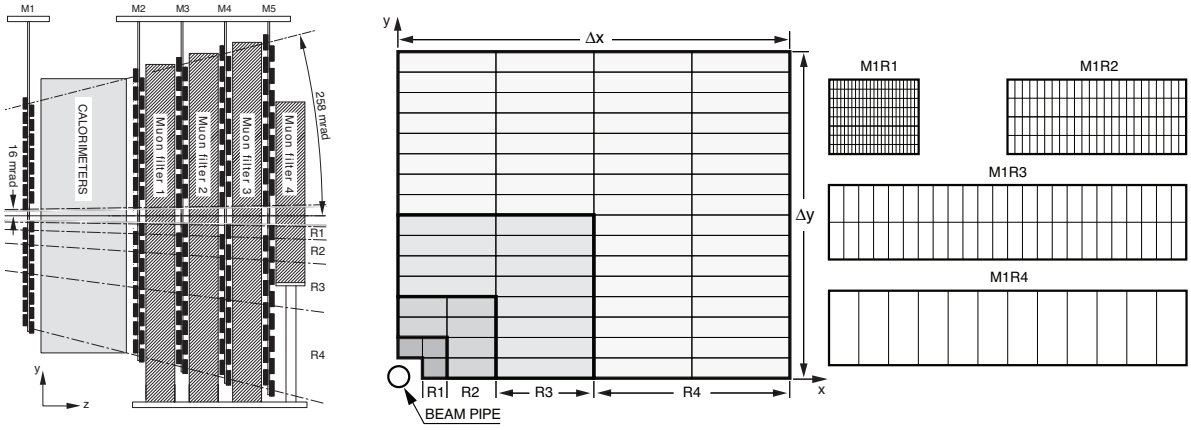
The calorimeter system selects hadrons, electrons and photons for the hardware trigger (L0) according to their deposited transverse energy. Furthermore, it provides energy and position measurements and allows for discrimination between hadrons, electrons and photons.

The calorimeter system, whose arrangement is shown in Fig. 2.7(a), is composed of a scintillating pad detector (SPD), a preshower (PS), a shashlik type electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) [119]. A 15 mm lead converter with a thickness of 2.5 radiation lengths is placed between the planes of rectangular scintillating pads of the SPD and the PS. All subcomponents follow the same basic principle: scintillation light is transmitted to photomultipliers by wavelength-shifting fibres. The single fibres for the SPD/PS cells are read out using multianode photomultiplier tubes, while the fibre bunches in the ECAL and HCAL modules require individual phototubes.

The SPD is designed to identify charged particles; most of which are pions. Those are a background in the PS and ECAL, which is reduced by a measurement of the longitudinal partitioning of the electromagnetic shower in the PS detector and the main section of ECAL. The ECAL was designed to reach optimal energy resolution of high energy photon showers, which are required to be fully contained in the ECAL's fiducial volume of 25 radiation lengths. To account for higher occupancy in the inner region, a segmentation into three different sections has been chosen for the ECAL, with a corresponding projective geometry for the SPD and PS detectors. The layout is shown in Fig. 2.7(b). The outer dimensions match projectively those of the tracking system, while the square hole around the beam-pipe approximately limits the inner acceptance to 25 mrad.

The HCAL is a sampling device made from iron as absorber and scintillating tiles as active material, which follow an unconventional arrangement along the beam axis. Given the dimensions of the hadronic showers, the HCAL is segmented into two zones with larger cell sizes. The thickness of the HCAL is limited to 5.6 nuclear interaction lengths (λ_i) due to space constraints.

Performance and resolutions have been determined in Ref. [120] for the electromagnetic calorimeter and [121] for the hadronic calorimeter.



(a) (b)

Figure 2.8: (a) Arrangement of the muon system. (b) Left: front view of a quadrant of a muon station. Each thin framed rectangle represents one of the 276 chambers, which is divided into logical pads. Right: Pads of the four M1 regions. Finer (coarser) pad granularity is chosen in M2 and M3 (M4 and M5), where the number of pad columns is in-(de-)creased by a factor 2. [103]

2.1.7 Muon system

The muon detection system, described in Ref. [124], provides muon identification and constitutes the hardware trigger together with the calorimeters. The system is composed of five stations (M1–M5) of rectangular shape, placed along the beam axis with projective geometry (*cf.* Fig. 2.8(a)). Each of the stations is equipped with multi-wire proportional chambers (MWPC), except in the highest rate region of M1, where triple gas electron multiplier (GEM) [125] detectors are used. Stations M2 to M5 are placed downstream of the calorimeters and are interleaved with 80 cm thick iron absorbers to discriminate penetrating muons from high energetic hadrons which have not been absorbed in the HCAL.

The hardware trigger requires excellent time resolution in combination with high efficiency, realised by optimised charge collection geometry and gas mixtures in the MWPCs and GEM chambers. The minimum momentum that a muon must have to traverse the five stations is approximately 6 GeV. The total absorber thickness, including the calorimeters, is approximately 20 nuclear interaction lengths. Station M1 is used to improve the p_{T} measurement in the hardware trigger, but does not contribute significantly when the track is matched with the measurements from the tracking detectors in the software trigger or offline. The stations are divided into four regions, R1 to R4, with increasing distance from the beam axis, to obtain comparable occupancies in each region (*cf.* Fig. 2.8(b)). The detector performance is reviewed in [126].

2.2 Data taking and reconstruction

This section follows the data flow in LHCb, from proton-proton collisions to an analysis-ready dataset. First, a brief overview of the LHC accelerator complex is given, highlighting parts which are relevant for LHCb. Then, the data taking conditions are discussed, followed by an overview of the Run I trigger system, the reconstruction and simulation chains.

2.2.1 LHC operations

The large hadron collider (LHC) [106, 127, 128] is a proton-proton collider designed to reach centre-of-mass energies of 14 TeV and 30 million proton-proton collisions per second. The two counter rotating proton beams are guided by strong magnetic fields through two separate accelerator rings with 27 km circumference. To make use of existing infrastructure at CERN, the LHC has been built in the in the existing tunnel of the precessing large electron positron (LEP) collider. approximately 22 km of the LEP tunnel consist of curved sections, in which bending dipole magnets and focussing magnets are installed. The remaining 5 km consist of eight interaction regions that provide space for the experiments, injection and extraction elements, acceleration and collimation devices.

The length of the existing tunnel and the desired beam energy of 7 TeV per proton beam imply magnetic bending fields of 8.4 T. Additional constraints come from the relatively small 3.76 m diameter of the LEP tunnel into which two separate magnet apertures with opposite field orientations had to be squeezed. Each of the eight arcs of the LHC consists of 46 repeating series of one quadrupole and three dipole magnets. Each magnet uses niobium-titanium (NbTi)-based superconducting cables, which are operated at a temperature of 1.9 K. In combination with the high magnetic field, this means that there is a small margin before the superconducting state is lost. Even if only a section of the NbTi cable becomes a normal conductor, ohmic losses increase the operating temperature still further. This effect is known as magnet quench.

The beam of the LHC starts off in the LINAC2 linear accelerator, is passed at 50 MeV to a multi-ring booster synchrotron for acceleration to 1.4 GeV, and then to the 628 m-circumference proton synchrotron (PS). In the PS, the beam reaches 26 GeV and is partitioned into the desired bunch pattern for the LHC. At full intensity, each beam contains about 1.15×10^{11} protons, and consist of 2808, about 30 cm long bunches, which are separated by about 7.5 m or 25 ns. Another transfer from the PS into the 7 km-circumference super proton synchrotron (SPS) is made, where the beam is further accelerated to 450 GeV. From the SPS, the beam is injected into the LHC by fast kicker magnets which abruptly change the beam trajectory to move it down the 3 km transfer line. The acceleration chain is schematically shown as a part of the CERN accelerator complex in Fig. 2.9. An injection into the LHC takes about 20 minutes, due to the high impedance in the magnet circuits, the process of increasing the magnet current/the beam energy. During this operational phase, the transverse beam dimensions shrink. Once stable beam conditions are met, the LHCb VELO is closed and data can be recorded.

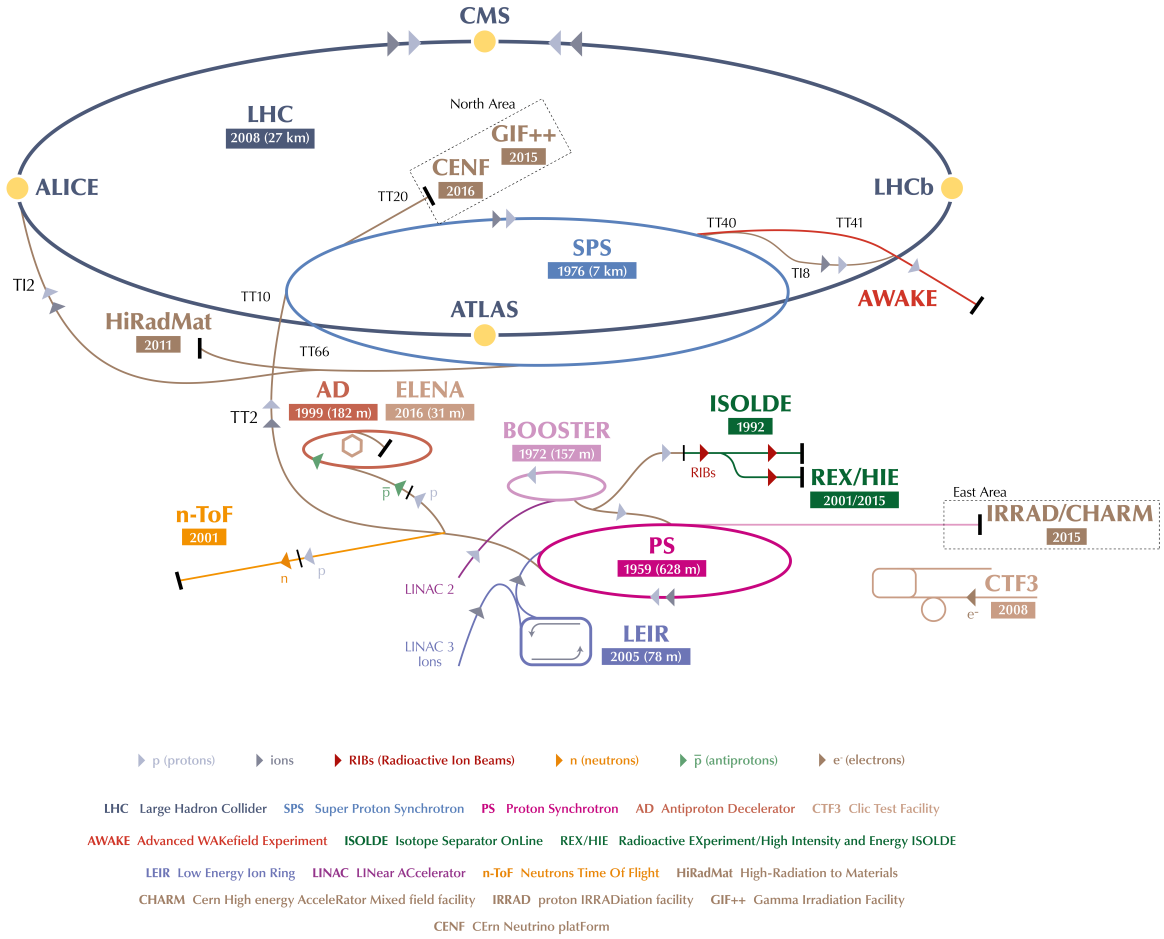


Figure 2.9: Schematic view of the CERN accelerator complex as of 2016 to 2018 [129].

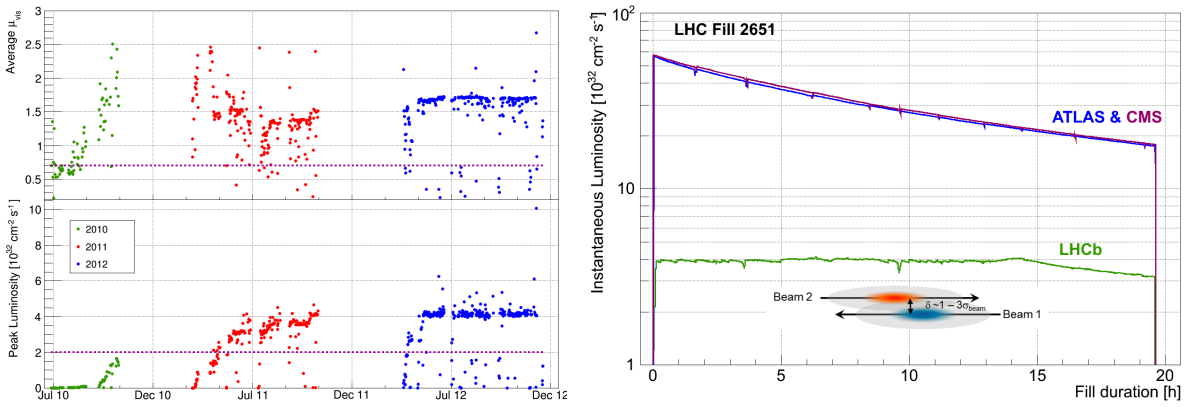
2.2.2 Data taking at LHCb

A crucial parameter for collider and experiment is the event rate, which depends the number of collisions, the product of the machine luminosity and the energy-dependent collision cross section. The luminosity is entirely determined by the beam parameters and given by

$$L = \frac{f_{\text{rev}} n_{\text{b}} N^2}{\sigma_x \sigma_y} F(\Phi, \sigma_x, \sigma_y, \sigma_s) \, .$$

Here, f_{rev} is the revolution frequency, n_{b} is the number of bunches, N is the number of protons per bunch, and F is a geometric function that depends on the crossing angle Φ of the beams, the transverse r.m.s. beam sizes σ_x, σ_y at the interaction points and the r.m.s. bunch length σ_s.

To ensure good detector performance for the targeted precision measurements in LHCb, it is crucial to control the visible pile-up μ_{vis}. Pile-up is defined as the average number of visible interactions per bunch-crossing [130]. The luminosity and pile-up as a function of time during Run I is shown in Fig. 2.10(a). Due to the lower number of bunches in 2010, the visible pile-up clearly exceeded the LHCb design value, while the luminosity was still below that. It was demonstrated that the trigger and reconstruction work efficiently under such harsh conditions and that the physics output was not compromised, such that, at an increased number of bunches, LHCb took data at up to twice its design luminosity in 2011 and 2012.



(a) (b)

Figure 2.10: (a) Pile-up and instantaneous luminosity as a function of time. The dotted lines show the design values. (b) instantaneous luminosity for ATLAS, CMS and LHCb during a long fill. After ramping to the desired luminosity at the LHCb interaction point, the luminosity is kept stable by adjusting the transversal beam overlap. After about 15 hours, the luminosity decreases for LHCb as well, due to the difference in the final focusing at the collision points. Taken from [104].

In 2011 a luminosity levelling procedure was introduced at the LHCb interaction point. By adjusting the transverse overlap of the beams at LHCb, the instantaneous luminosity could be kept stable to within about 5 % during a fill, as illustrated in Figure 2.10(b). The luminosity levelling procedure minimises the effects of luminosity decay, allowing to maintain the same trigger configuration during a fill and to reduce systematic uncertainties due to changes in the detector occupancy.

The integrated luminosity recorded by LHCb was 1.11 fb^{-1} in 2011 at a centre-of-mass energy of 7 TeV and 2.08 fb^{-1} in 2012 at 8 TeV. Luminosity calibrations were carried out with the LHCb detector for the various centre-of-mass energies at which data has been taken. A statistical combination of a van der Meer scan [131] and beam-gas imaging [132] luminosity calibration methods were employed [133]. The combined method is the most precisely measured luminosity at the LHC.

2.2.3 Trigger system

The LHCb trigger, described in Ref. [134], is a two-level system. The first is implemented in hardware (level zero – L0), followed by the two-stage software level trigger, called high level triggers HLT1 and HLT2. The L0 trigger is designed to reduce the event rate from the nominal LHC bunch crossing rate of 40 MHz to a maximum of 1.1 MHz. During Run I the rate of visible interactions was about 13 MHz [135]. The complete detector is read out on a positive L0 decision and the data is sent to the software trigger, that is implemented on the event filter farm (EFF), which had about 30000 processing cores in 2012.

In Run I, the bulk of data is selected by five L0 selection streams, selecting events with either high p_T muons or large transverse energy deposits in the calorimeter. On a positive L0 decision, the readout boards (trigger electronics level 1 board – TELL1) of the individual detectors send data through a switching network to the EFF nodes [136]. There, dedicated algorithms, referred to as trigger lines, compute a HLT1 decision, which decrease the event rate to about 70 kHz. Due to limited computing resources the algorithms rely on a *partial* event reconstruction. The HLT1 trigger line used in the analysis presented in this thesis is the inclusive beauty and charm line, `Hlt1TrackAllL0`. It selects good quality track candidates based on their transverse momentum and displacement from the primary vertex. This trigger line gets the dominant part of the HLT1 bandwidth allocated: about 58 kHz.

The HLT2 decision is then based on several inclusive or exclusive selection algorithms, relying on an offline-like event reconstruction. Fully hadronic trigger lines are of particular interest for the analysis presented in this thesis. These lines make use of multivariate classification algorithms, constructing two-, three- and four-track vertices [137, 138]. These so-called *topological trigger lines* are based on a modified Boosted Decision Tree with discretised input variables [139].

In 2012 the trigger system was improved further by introducing the concept of a *deferred trigger* [140]. This trigger stores about 20 % of all L0 accepted events on the local disks of the HLT farm nodes. The buffered events are processed if farm nodes are idle, *e.g.* during interfill gaps of the LHC. The increased effective computing power allowed for higher quality reconstruction and looser selection criteria. Not only such improvements, but also the data taking environment, computational resources and physics goals cause changes in trigger lines. These have been summarised and studied in [141] for 2011, and [142] for 2012 data taking.

When combining the output of trigger lines, the following terms, defining the trigger logic in LHCb, are introduced:

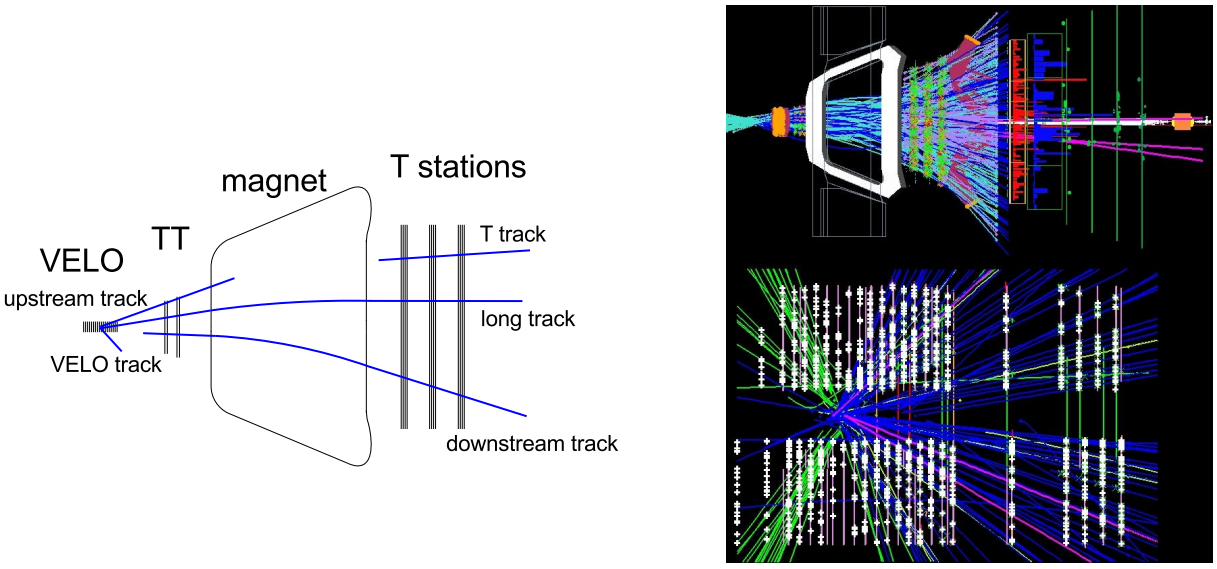
- **TIS** (triggered on signal) : particles of a given signal decay chain caused a positive decision on a given (combination of) trigger line(s)
- **TOS** (triggered independent of signal) : other particles than those of the signal decay chain in the event caused a positive decision on a given (combination of) trigger line(s)
- **TOB** (triggered on both) : neither TIS nor TOS. Only a combination of signal and non-signal particles caused a positive decision on a given (combination of) trigger line(s)
- **TISTOS** : the logical `AND` of TIS and TOS

2.2.4 Track reconstruction

The LHCb reconstruction algorithms define different track types, shown in Fig. 2.11(a). These are VELO tracks, which have hits in the VELO; upstream tracks, which have hits in the VELO and TT; T tracks, which have hits in the T stations; downstream tracks, which have hits in TT and the T stations; and long tracks, which have hits in the VELO and the T stations. The latter tracks can additionally have hits in TT. Long tracks are the highest quality tracks comprising all available information from the trackers and are therefore used in most physics analyses. Downstream tracks mainly play a role in the reconstruction of daughters from long-lived particles which have decayed after the VELO (usually weakly decaying strange hadrons, such as Λ or K^0_S). Track reconstruction can be subdivided into a track finding/pattern recognition part and a track fitting part.

The basic track finding algorithms, called VELO tracking [143] and T seeding [144], reconstruct VELO and T track candidates which are used as seeds for upstream, long and downstream tracks. Long track candidates are found by two dedicated algorithms. The first, called forward tracking [145], starts with VELO tracks and searches for corresponding hits in the T stations. The second, called track matching [146, 147], uses both VELO and T tracks as input and matches them in the magnet region. Downstream tracks use T tracks as seed and searches for corresponding hits in the TT [148]. The outputs of all algorithms are merged, eliminating candidates that were found twice.

The track fit is done by a Kalman filter [149], taking into account multiple scattering and corrects for energy loss due to ionisation [150, 151]. In a typical $b\bar{b}$ event, such as the one shown in Fig. 2.11(b), the algorithms reconstruct a sizeable fraction (up to 20 % in high occupancy events) of fake tracks, *i.e.* charged tracks that do not correspond to a real particle which passed through the detector. Fake long tracks can originate from falsely reconstructed track segments in the VELO or the T stations, from a mismatch of VELO and T station segments or from hadronic interaction of particles with the detector.



(a) (b)

Figure 2.11: (a) Track types in LHCb [152]. (b) Display of reconstructed tracks and assigned hits from a 2011 $B_s^0 \to \mu^+\mu^-$ candidate event. The insert shows the corresponding VELO region [C14]

The rate of fake tracks is significantly reduced in a dedicated offline filtering algorithm based on a neural network classifier [I2]. Mismatched track segments are due to the long lever arm between the tracking stations up- and downstream of the magnet and remain to be the most abundant category after fake track rejection.

Detector alignment

To provide an optimal momentum resolution, the position and orientation of the LHCb tracking detector elements in the global reference frame must be known with an accuracy significantly better than the single hit resolution: the detectors need to be well *aligned*. The alignment of the LHCb tracking detectors uses information from optical and mechanical surveys and from reconstructed charged particle trajectories. On software level, alignment is based on the Millipede algorithm [153] and a track fit-based algorithm [154]. The former uses a simplified track model, converges within one iteration, but can align only one sub-detector at once. The latter includes the alignment parameters in the track fit model and finds the optimal alignment parameters by minimizing the χ^2 calculated on all the tracks considered as input. The alignment proceeds in four steps, in accordance to the aligned detectors: VELO [155, 156], tracking stations [157], RICH mirrors [118], and muon stations [126], with each step detailed in the given references.

Performance of track reconstruction algorithms

Several measures characterise the performance of the LHCb track reconstruction. A short summary of performance parameters of interest for this work is given. In the analysis described later, almost the entire topological selection[1] is done directly after reconstruction in an LHCb-wide pre-processing campaign. This is opposed to an offline selection, and demonstrates how clean a topological selection based on good track reconstruction is.

- The *track reconstruction efficiency* measures the probability that the trajectory of a charged particle that has passed through the full tracking system is reconstructed. A data driven tag-and-probe method, using $J/\psi \to \mu^+\mu^-$ is used, which determined track reconstruction efficiencies above 96 % for largest parts of the experimentally covered kinematic phase space [158].

- *Primary vertex (PV)* and *impact parameter (IP) resolutions* are important measures in the reconstruction of the decay topology, *i.e.* the spatial resolution of the reconstructed tracks. As such they are driven by the detector performance and the reconstruction algorithms of the VELO.

  The PV resolution is measured by randomly dividing an event into two samples of tracks of equal size. The PV is reconstructed independently in both samples, which is then used to infer the resolution. The x and y resolutions are shown in Fig. 2.12(c). Due to the boost in z–direction and the arrangement of VELO layers, the PV resolution is worse in z. For a PV with 25 tracks, a z resolution of 71 µm was measured.

  The IP of a track is defined as its distance from the PV at its point of closest approach. Transverse IP_x and IP_y distributions of tracks originating from a PV that is made up of at least 25 tracks are measured as a function of a variable of interest. Here, the variable of interest is $1/p_\mathrm{T}$, and the linear dependence of the IP resolution is a consequence of multiple scattering and the geometry of the vertex detector. In each of the $1/p_\mathrm{T}$ bins, the IP_x and IP_y distributions are approximately Gaussian, and their width is consequently plotted as resolution shown in Fig. 2.12(d).

- Good *mass* and *momentum resolution* are essential parameters, providing high discriminating power between signal and background processes. The momentum resolution is measured with $J/\psi \to \mu^+\mu^-$ and is related to the mass resolution, in the limit of massless muons which have similar momentum, via the approximation

$$\left(\frac{\delta p}{p}\right)^2 = 2\left(\frac{\sigma_m}{m}\right)^2 - 2\left(\frac{p\sigma_\theta}{m\theta}\right)^2 .$$

  Here, m is the invariant mass of the J/ψ candidate and σ_m is the Gaussian width obtained from a fit to the mass distribution. The second term is a correction for the opening angle, θ, between the two muons, where σ_θ is the per-event error on θ which is obtained from the track fits of the two muons. The relative momentum resolution is shown in Fig. 2.12(a). The relative mass resolution has been measured with dimuon resonances and is shown in Fig. 2.12(b).

The momentum resolution of a mother particle in decay chains involving intermediate metastable particles, can further be improved by a kinematic fit [159], known as *decay tree fit* (DTF) within LHCb. The algorithm parametrises a decay chain in terms of vertex positions, decay lengths and momentum parameters, and then fits these parameters simultaneously using a Kalman filter. In the fit, the decay vertex of a particle is obtained by including its hypothesised origin and mass. Kinematic constraints such as the measured parameters of the final state tracks, 4-momentum conservation at each vertex etc. are taken into account.

[1] Some variables related to the decay topology and reconstruction are input to the *D-from-B* BDTs, but those are dominated by PID variables.

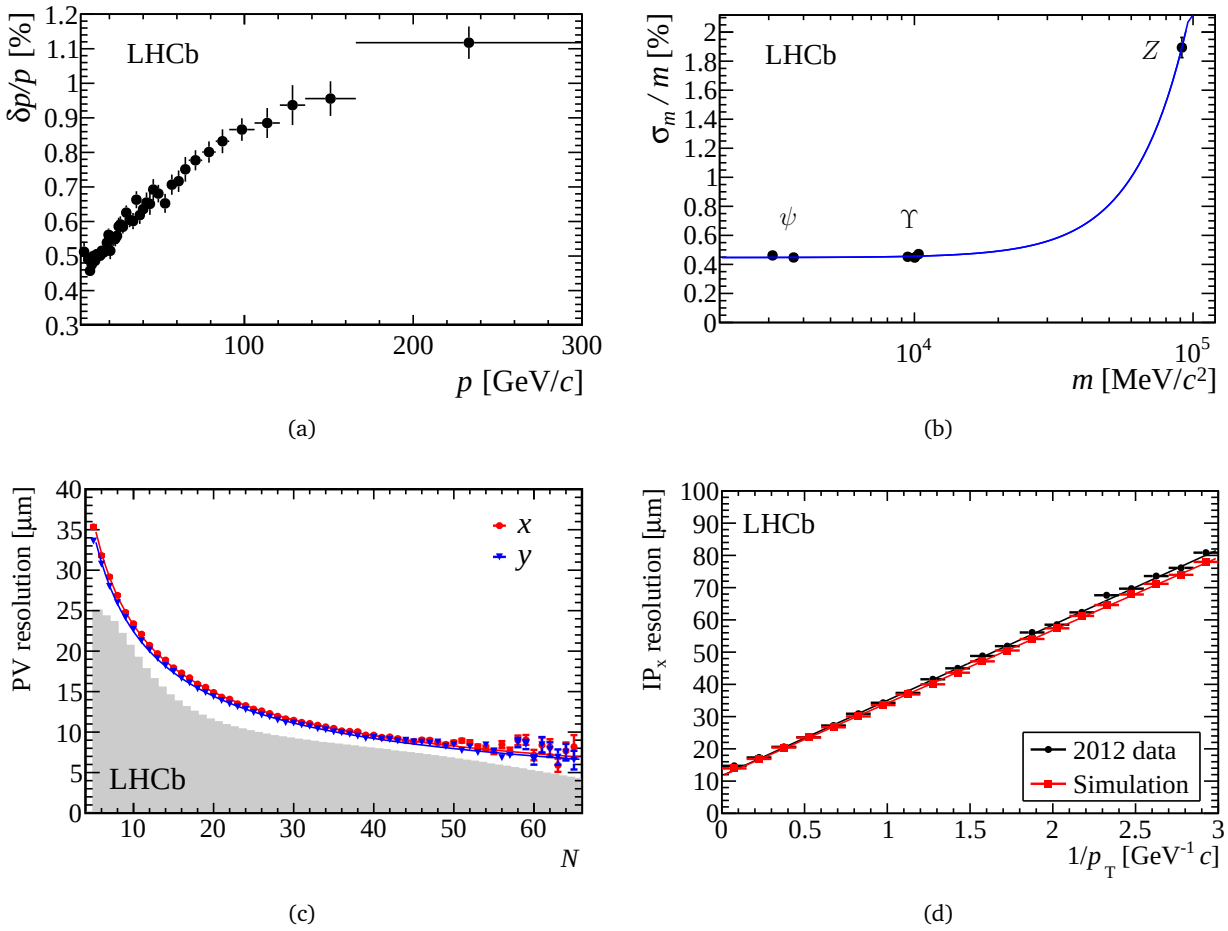
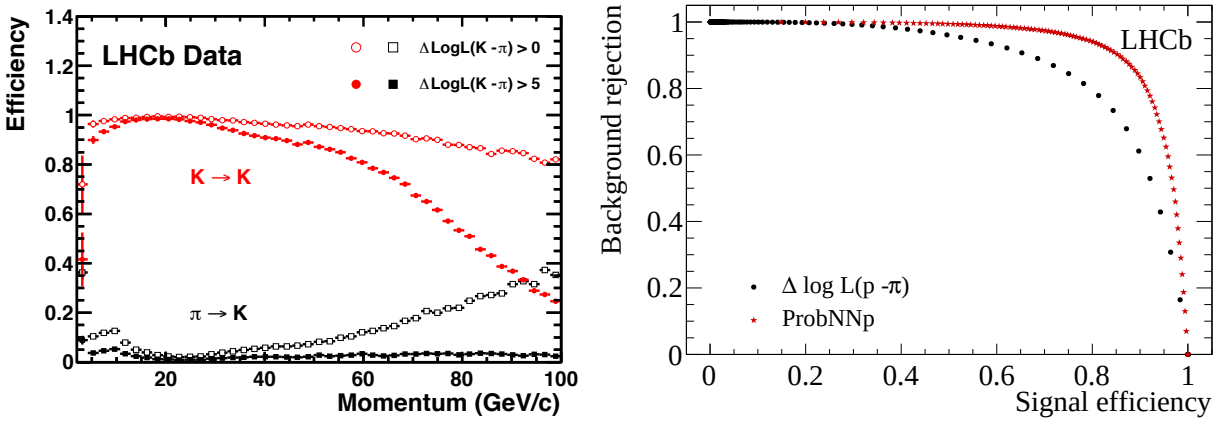


Figure 2.12: Performance of the LHCb track reconstruction: (a) Momentum resolution of long tracks from $J/\psi \rightarrow \mu^+\mu^-$ decays [104]. (b) mass resolution obtained from invariant mass fits to dimuon resonances [104]. (c) primary vertex (PV) resolution for events with one reconstructed PV as a function of the track multiplicity N. The grey histogram shows the distribution of N per PV in arbitrary units for all events that pass the HLT [110]. (d) Impact parameter resolution in x-direction [110].

2.2.5 Particle identification

In this work, decays with six hadrons in the final state will be studied, for which most of the separation power against combinatorial background in the offline analysis came from informations of the particle identification (PID) detectors.

The PID for hadrons is obtained by combining the Cherenkov angle information of the RICH detectors with the track momentum provided by the track reconstruction. All tracks in the event are considered simultaneously by an overall event log-likelihood, allowing for optimal treatment of tracks where Cherenkov cones overlap. The minimisation procedure of that event negative log-likelihood (NLL) starts with the assumption, that all particles are pions, since they are the most abundant species in high energy pp collisions. The event NLL is then computed under this assumption, using the distribution of photon hits in the HPDs, the associated tracks and uncertainties as input. Then, the NLL is recomputed, changing the mass hypothesis of a single track to e, μ, π, K, p, leaving all other hypotheses unchanged. The particle hypothesis of the track which caused the largest descent of the event NLL is then fixed, and the procedure is repeated until no further improvement in the event NLL is found.



(a) (b)

Figure 2.13: (a) Kaon identification efficiency and pion misidentification rate for two different requirements of the difference in log-likelihood [118]. (b) Comparison of difference in log-likelihood and neural network response for protons [104].

In practice, some modifications to this algorithm have been implemented to increase its speed [118]. The final result of the PID algorithm is the difference in log-likelihood ($\Delta \log L(X-\pi)$) when a track is assigned the pion mass hypothesis compared to another mass hypothesis X. The performance of the difference in log-likelihood is measured with data driven tag-and-probe methods in which large samples of genuine π, K, p tracks are exclusively reconstructed using kinematic selections only. The decays providing this information were chosen to be $K^0_S \to \pi^+\pi^-$, $D^{*+} \to \left[K^-\pi^+\right]_{D^0}\pi^+$ and $\Lambda^0 \to p\pi^-$. The performance of the difference in log-likelihood of kaon and pion using the D^{*+} calibration sample is shown in Fig. 2.13(a). It should be noted that the performance not only depends on kinematic measures, such as the momentum, but also on the track multiplicity of the event, correlated to the RICH occupancy.

Further information on the PID come from the muon stations and the calorimeters. Their informations are used to calculate a combined log-likelihood. Moreover, using additional information from the tracking detectors, an artificial neural network has been developed, whose response for a particle type `X` is referred to as `ProbNNX` [I3][104]. The performance of the neural network response compared to the difference in log-likelihood for protons is shown in Fig. 2.13(b). It performs significantly better than the log-likelihood difference.

Efficiencies of PID variables

Simulating responses of PID detectors is non-trivial, as it requires well modelled kinematics of traversing particles, detector occupancy, and changing experimental conditions such as alignment, temperature and gas pressure. This holds true in particular for the simulation of variables like `ProbNNX`, which combine many informations from the PID detectors. This motivates the use of data driven techniques to measure efficiencies of PID response variables. The most common tool in LHCb providing such a data driven efficiency measurement is called PIDCalib [160]. It allows to extract the PID efficiency as a function of a set of calibration variables, usually p, η and `nTracks` (the number of tracks in the event), by estimating yields of clean calibration sample signals before and after PID selection. The calibration sample in the analysis presented in this work is $D^{*+} \to \left[K^-\pi^+\right]_{D^0}\pi^+$, which is used to determine the efficiency of a cut on `ProbNNk` > 0.18. Hence, neglecting technicalities described in detail in Ref. [160], the ratio of D^{*+} yield after and before the cut on `ProbNNk` > 0.18 gives the efficiency.

Another approach is the re-sampling or correction of PID variables in simulation samples. These approaches, described in Ref. [I4], are known as PIDGen and PIDCorr respectively, and have been integrated in PIDCalib. In essence, again omitting technical details, the probability density function (p.d.f.) of set of calibration and PID variables, *e.g.* p, η, `nTracks` and `ProbNNk`, is obtained on a clean calibration sample with the help of kernel density estimators from the Meerkat package [161]. In PIDGen, a new PID variable is now sampled according to the calibration sample p.d.f., replacing the PID variable of the simulation sample. PIDCorr on the other hand first needs the p.d.f. of the simulation sample corresponding to the one obtained on the calibration sample. Then PIDCorr corrects the PID variable in the simulation sample with a variable transformation, which uses the p.d.f.s obtained from simulation and calibration samples. The transformation shifts the value of the PID variable, and hence preserves all simulated relations to other variables in the sample. Both approaches allow to extract PID efficiencies from the transformed simulated samples.

This work will introduce further PID-like variables for decaying particles, for which efficiencies need to be obtained as well. This is described in detail in Sec. 5.

2.2.6 Data processing and simulation

The data flow, described in the LHCb computing model [162], from detector data to the trigger system has been discussed in Sec. 2.2.3. The output of the trigger, the so called `RAW` data, is transferred to the CERN Tier 0 centre, archived on tape, copied and distributed to the six LHCb Tier 1 sites. There, and at CERN, a first pass reconstruction is executed shortly after the data is taken. Another full reconstruction is run at the end of each year's data taking. In 2012 the reconstruction procedure has been optimised to also use computing power of the Tier 2 sites [163]. Each reconstruction pass, labelled by a `Reco` version number [C24], is followed by a *stripping* pass [C43], where several hundreds of different physics group selections are executed. The selected events are written to one or more of about ten output streams.

The reconstruction output (`FULL.DST` – DST stands for data summary "tape") contains a copy of the `RAW` data. The DST after the stripping pass contains the same information as FULL.DST for events selected by the *stipping line*, as well as additional information stored by the stripping algorithms. Many exclusive analyses however only require information relating the few reconstructed particles of the selected decay tree. A microDST format has been developed containing this reduced information, reducing the required storage space from ∼120 kB/event on DST to ∼13 kB/event on microDST [163]. These stripped datasets are the input to user analysis and to further centralized processing by physics working groups, such that analysts do not need to access `RAW` data or the non-stripped reconstruction output. The described data flow is illustrated in Fig. 2.14.

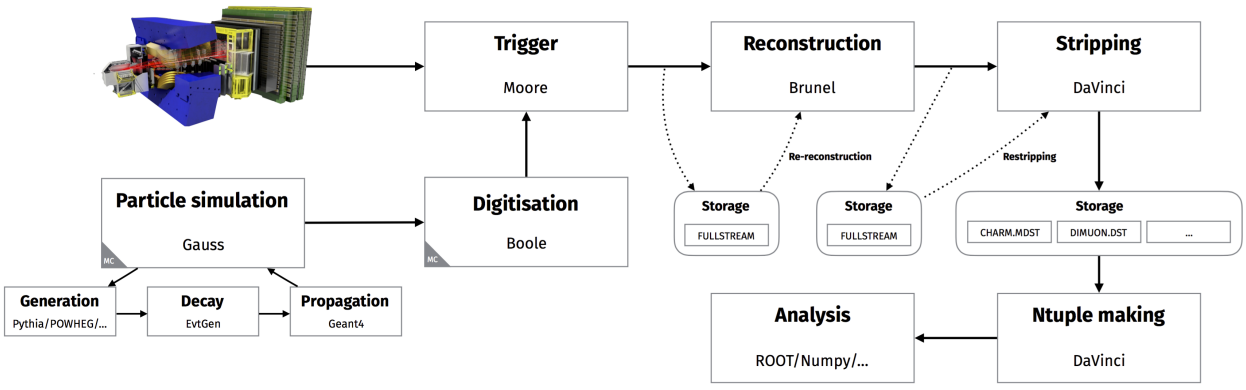


Figure 2.14: Data flow and associated applications in LHCb [164].

The applications associated to this data flow, MOORE [C42], BRUNEL [C39] and DAVINCI [C40], are based on the GAUDI architecture [165] and the LHCb event model [I5].

Simulation in LHCb is performed with the Gauss package [166][C41]. It interfaces PYTHIA [167, 168] with a specific LHCb configuration [169]. Decays are described by EVTGEN [170], in which final state radiation is generated by PHOTOS [171]. The interaction of generated particles with detector material, hits and deposits in the detector are modelled with GEANT4 [172, 173]. After digitisation [C41], *i.e.* the application of the detector response to previously created hits and deposits, simulated and real data use the same software.

3 Analysis strategy

This work reports the first observation of the decays $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^*(2007)^0 K^-$, and the measurement of their branching fraction relative to the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. That decay is chosen as reference channel because it has been measured before with relatively large signal yield [174] and contains the same set of particles in the final state, so that many systematic uncertainties cancel. The reconstruction chain for the decays read:

$$
\begin{aligned}
&\text{(signal) } \Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \text{ with } \Lambda_c^+ \to p K^- \pi^+ \text{ and } \bar{D}^0 \to K^+ \pi^- \,, \\
&\qquad\quad \Lambda_b^0 \to \Lambda_c^+ \bar{D}^*(2007)^0 K^- \text{ with } \Lambda_c^+ \to p K^- \pi^+ \,,\ \bar{D}^*(2007)^0 \to \bar{D}^0 \pi^0 \text{ or } \bar{D}^0 \gamma \text{ and } \bar{D}^0 \to K^+ \pi^- \,, \\
&\text{(reference) } \Lambda_b^0 \to \Lambda_c^+ D_s^- \text{ with } \Lambda_c^+ \to p K^- \pi^+ \text{ and } D_s^- \to K^- K^+ \pi^- \,.
\end{aligned}
$$

Here, final state particles are highlighted in green, and particles which are not reconstructed are marked in red. Throughout the document, the terms signal and reference channel are used according to the name given in parenthesis. The $\bar{D}^*(2007)^0$ decay chain is treated as signal, but due to the missing particles in the reconstruction chain sometimes referred to as *partially reconstructed* decay. Throughout the documentation of the analysis in this work $\bar{D}^*(2007)^0$ is denoted as $\bar{D}^{*0}$ for brevity.

Both, signal and reference channel require the reconstruction of six hadrons for which particle identification plays a crucial role. For this reason the selection after the LHCb-wide stripping, introduced in Sec. 2.2.6, is essentially driven by dedicated PID-like variables for charm hadrons, termed *D-from-B* BDTs. These were first developed in Refs. [I6][175, 174], and it has been shown in $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays that their discriminating power is far superior compared to combining standard PID variables for each final state particle [176]. The *D-from-B* BDTs have been re-trained, embedded into a software package and calibrated in the course of this work. This is detailed later in Sec. 5.

After optimisation of the selection (Sec. 6), the signal yields of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ are determined (Sec. 7). The next step towards the relative branching fraction, $\mathscr{R}^{(*)}_{\Lambda_b^0}$, is to correct for detector- and reconstruction-induced effects. These can all be conflated into efficiency corrected yields N_{corr} of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$, such that the relative branching fraction reads

$$
\mathscr{R}^{(*)}_{\Lambda_b^0} = \frac{N_{\text{corr}}^{\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-}}{N_{\text{corr}}^{\Lambda_b^0 \to \Lambda_c^+ D_s^-}} \frac{\mathscr{B}(D_s^- \to K^- K^+ \pi^-)}{\mathscr{B}(\bar{D}^0 \to K^+ \pi^-)} \,. \tag{3.1}
$$

It is hence a simple ratio of two numbers, the efficiency corrected yields N_{corr}, and a correction for the branching fractions of $D_s^- \to K^- K^+ \pi^-$ and $\bar{D}^0 \to K^+ \pi^-$, which is taken from the PDG [14].

The efficiency correction compensates for various acceptances, selections and inefficiencies in the reconstruction. This does however not mean that N_{corr} can be interpreted in a straightforward manner as the total number of produced $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ or $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays. Estimating this number is more involved, since efficiencies which cancel in the ratio of Eq. (3.1) would have to be taken into account as well. Here, the efficiency corrected yield of a decay d, is computed on an event-by-event basis as

$$
N_{\text{corr}}^{\text{d}} = \sum_i \frac{{}_s\mathscr{P}_{\text{d}}(\boldsymbol{\mu}_i)}{\prod_k \varepsilon_k(\boldsymbol{\omega}_{i,k})} \qquad \forall \{\boldsymbol{\mu}_i, \boldsymbol{\omega}_{i,k}\} \subset \boldsymbol{x}_i \left(\cap_k c_k^{\text{d}}\right). \tag{3.2}
$$

The formula is understood as follows: the sum runs over all reconstructed and selected events i; the reconstruction and selection procedure is put into formula on the right: the event i is part of a sample $\boldsymbol{x}$ which depends on the conjunction of reconstruction and selection steps for decay d, c_k^{d}. Each of these k steps have to be corrected for with a corresponding efficiency ε_k.

The efficiencies are measured as a function of calibration variables $\boldsymbol{\omega}$. In practice, a look-up table of efficiencies as function of $\boldsymbol{\omega}_k$ is created with the help of a calibration sample. An example of this is the creation of a look-up table from a simulation sample in $\boldsymbol{\omega}_{\text{Trigger}} = (p_{\text{T}}, y)^T$, *i.e.* transverse momentum and rapidity, for the efficiency correction of the trigger selection. Then, for each event i in the signal sample, the measured value for $\boldsymbol{\omega}_k$ determines where the value of ε_k is read off from the look-up table.

But not each of the events in the sample $\boldsymbol{x}$ is a signal event. This is taken into account by removing the background with an unfolding algorithm, that allows to give each event a weight ${}_s\mathscr{P}_{\text{d}}(\boldsymbol{\mu}_i)$: the sWeight (*cf.* Sec. 4.2.4). It depends on the set of discriminating variables $\boldsymbol{\mu}$. For example, in the case of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ these are the invariant masses of the decay products of Λ_b^0, Λ_c^+ and D_s^-. Note that both, efficiency and sWeight implicitly depend on selection and reconstruction steps c_k^{d}, but it is the analysts obligation to ensure that the efficiency corrected yield does not.

After accounting for systematic effects in the computation of N_{corr}, the relative branching fraction $\mathscr{R}_{\Lambda_b^0}$ is obtained according to Eq.(3.1). The computation of $\mathscr{R}^*_{\Lambda_b^0}$, *i.e.* the relative branching fraction of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$, requires another step in which the $\bar{D}^{*0}$ decay channels, $\bar{D}^{*0} \to \bar{D}^0 \pi^0$ and $\bar{D}^{*0} \to \bar{D}^0 \gamma$, are combined. This is necessary for kinematic reasons, which lead to different reconstruction and selection efficiencies.

4 Analysis tools

Much effort of this work went into scalable software development to provide a framework for future analyses. The concepts and developed tools are introduced in the first part of this section. The second part is then dedicated to statistical tools used throughout this work.

4.1 Software development

Software development in high energy physics often escapes focus, even though a large portion of time is spent on it. This paradigm gradually shifts, but much more effort is needed to provide efficient, well tested, portable and easy to use tools suiting the needs of analysts. To be able to develop such tools, common steps in typical analysis workflows need to be identified. The tool would then provide a flexible and well documented solution to that problem, in a way that the analyst would only have to change a minimal amount of settings to be able to integrate it into a different analysis.

An example for this would be a package to perform fits, or one to make plots which can be fully configured at runtime, *i.e.* by making changes in a configuration file, and not in the source code. Flexible in this context then means, that most configurable parameters are optional and pre-set to the most common use-case. The ideal case of such modularised software would be, that the analyst would only have to provide a file which defines the workflow and interfaces the modules, and another file providing individual configurations for the modules, like fit model, settings for plots etc.

For this work, a large part of the software is already released in a modularised manner. All of the modules described in the following have been developed in the course of this work. In fact, most modules were already used in the bachelor thesis of Christoph Otte [177], the master thesis of Florian Reiß [176] and diploma thesis of Christopher Vahl [178], leading to the first observation of $\Lambda_b^0 \to J/\psi\,\Lambda\phi$ [I7]. All of theses tools are part of repositories hosted by the CERN GitLab platform [C9], offering code review, collaborative features and services for reproducibility which are detailed later. The repositories are accessible by all members of the LHCb collaboration.

Some of the modularised, portable tools based on the C++ standard library [C28], ROOT [C30] and boost [C7] are briefly introduced here. The scripts of the tools are written to be configurable at runtime with the help of the `boost property tree INFO` parser [C8].

IOjuggler [C17] is a header-only library at the base of most tools. It parses command line options and facilitates the usage of ROOT objects.

ntuple-gizmo [C22] includes scripts to

- make histograms from various ROOT storage formats
- make plots from various objects
- manipulate ROOT `TTrees` (a specific storage format)
- automatically produce plots for studying misidentification backgrounds (*cf.* Sec. 5.1.3)
- optimise cuts (*cf.* appendix B.2)
- train and apply a classifier or regressor with TMVA [179]

EffDalitz [C12] is a library to create, apply and plot efficiency maps (*cf.* Sec. 8)

2DAdaptiveBinning [C2, C1] includes a simple algorithm to generate an adaptive binning in two dimensions, and provides a custom histogram class that reproduces the essential functions of the standard implementation of a two-dimensional histogram in ROOT. It is fully supported in the EffDalitz and ntuple-gizmo packages.

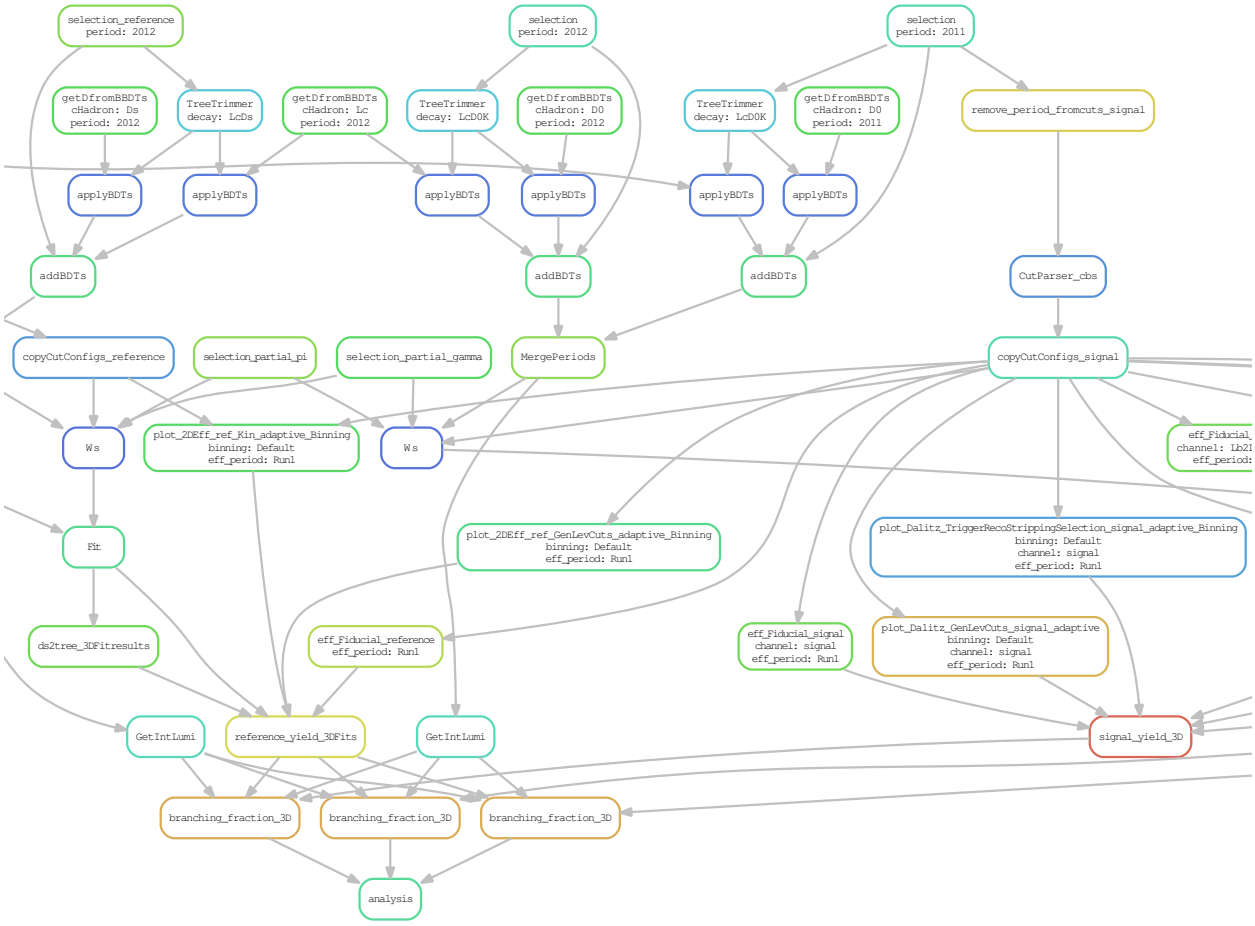


Figure 4.1: Excerpt of the directed acyclic graph of rules (steps in the analysis) executed by Snakemake to produce the branching ratio measurement. Note that systematic uncertainties, producing input datasets, cross-checks and the *D-from-B* BDT workflow are not included. Each coloured box corresponds to a rule, the arrows illustrate the dependencies between rules. Scalability is handled by the use of wildcards, *e.g.* the data taking period in the top nodes.

beef [C5] is a wrapper package around RooFit, the Root fitting framework, and is used for fitting and plotting. To ensure good fit quality, fits presented throughout the document are checked to return with a good status code from the Hesse minimisation, with accurate covariance matrix, and the best fit value of every fit parameter had to be more than 3 naïve σ away from its domain limit. To facilitate these checks, beef prints warning messages, can highlight fit results, large (anti-)correlations and examine the pull distribution of a plot with a Wald-Wolfowitz runs test or a simple check whether a configurable number of bins are above a configurable threshold.

The modules and dedicated scripts of the analysis are interfaced by the snakemake workflow management system [180]. Snakemake is used to create the workflow of the analysis in a reproducible and scalable way via a Python based language. The workflow of this analysis is shown as a directed acyclic graph in Fig. 4.1.

A related, more widely discussed topic is **reproducibility**, see *e.g.* [181, 182, 183, 184, 185, 186]. Within LHCb, there is an effort to provide reproducibility of analyses, summarised in the "LHCb Analysis Preservation Roadmap" [187]. The analysis presented in this thesis is fulfilling the requirements labelled "best practices" in the LHCb Analysis Preservation Roadmap, and is a pioneering analysis in this regard [188].

4.2 Statistical methods

From a statistics point of view, the main objective of the analysis presented here will be the following: given the observed data, provide an estimate $\widehat{\mathscr{R}}^{(*)}_{\Lambda_b^0}$ of the relative branching fraction $\mathscr{R}^{(*)}_{\Lambda_b^0}$, that has the so-called best properties. The term is defined in the context of parameter estimation in Sec. 4.2.1. Furthermore, a well defined *statistical uncertainty* of $\widehat{\mathscr{R}}^{(*)}_{\Lambda_b^0}$ needs to be estimated. This is done in terms of a *frequentist interval estimation*, which needs to have good *coverage* properties (*cf.* Sec. 4.2.3). Another uncertainty, the *systematic uncertainty*, is introduced by inaccuracies in the process, strategy or method of a measurement. This uncertainty is often difficult to estimate due to its complexity or limited amount of possibilities to probe the process/strategy/method. It should therefore be understood in the context of *Bayesian statistics* as the analysts best (subjective) estimate. While observer bias should in principle be avoided, *e.g.* by doing a blind analysis [189], minimisation of the uncertainty – or equivalently, maximisation of the *statistical power* – is desired. Beside a careful evaluation of systematic effects to reduce the systematic uncertainty, methods to increase the statistical power by an optimised selection, and by making optimal use of the selected data will be utilised. The later employed methods are introduced in Secs. 4.2.5 and 4.2.4.

4.2.1 Parameter estimation

The procedure of estimating a parameter's value given the **sample** (the observed data) $\boldsymbol{x}$ is called parameter fitting. The best fit value, or estimate $\hat{\theta}$, of the true parameter θ^* is inferred from a function of the data: the **estimator**. There is no fundamental rule dictating how an estimator must be constructed. Therefore, in the frequentist notion, it is tried to choose the estimator which has the best properties, *i.e.* is consistent, unbiased, efficient, and robust. Where consistency means that the estimate $\hat{\theta}$ converges to θ^* in the limit of large sample sizes. An unbiased estimator returns an estimate that does not deviate from θ^* if an experiment with limited sample size would be repeated asymptotically many times, *i.e.* in the limit of a large number of independent and identically distributed (*i.i.d.*) experiments. The efficiency of an estimator is given by the ratio of its own variance over the variance calculated by the Cramér-Rao lower bound. Since the variance of a biased estimator might be lower than that of a less (or un-) biased one, a combination of bias and variance of the estimator is commonly used to measure its quality. Robust estimators are not affected by outliers or other small departures from model assumptions. A more formal definition is *e.g.* given in [190, 14, 191]. Next, an estimator with good properties is introduced: the maximum likelihood estimator.

4.2.2 The method of (extended) maximum likelihood

The method of maximum likelihood is a technique for estimating a parameter's value given a finite sample of data. Under certain conditions, detailed *e.g.* in Ref. [190], it can be shown that the maximum likelihood estimator is consistent, unbiased, efficient and robust. The likelihood function of n measured values $x_1 \dots x_n$ is given by the product of the individual conditional probability density functions (p.d.f.s)

$$L(\theta) = f(x_1|\theta) \cdot \ldots \cdot f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)\,, \tag{4.1}$$

that is the *joint* p.d.f. of i.i.d. measurements x_i. The condition, that the measurements are independent allows to write the joint p.d.f. as a product – *i.e.* the individual p.d.f.s *factorise*.

For mainly technical reasons, it is more convenient to use the negative log-likelihood

$$\ell(\theta) = -\log L(\theta) = -\sum_{i=1}^{n} \log f(x_i|\theta) \,. \tag{4.2}$$

The maximum likelihood estimate $\hat{\theta}$ for the parameter(s) θ is the value that maximises $L(\theta)$, or minimises ℓ. This means it is solution to

$$\partial_\theta \ell(\theta) = 0 \,. \tag{4.3}$$

Since $L(\theta)$ is the product of probability measures given the data, it is intuitively comprehensible that a maximum of this product selects the best estimate.

In this work, the minimisation (4.3) is done with the C++ equivalent of the MIGRAD subroutine within MINUIT [192], based on Newton's method. Furthermore, the sample size n itself is a Poisson distributed random variable in all fits carried out in this work. To account for the sample size, the likelihood is *extended*:

$$L(\nu, \theta) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i|\theta) \,. \tag{4.4}$$

The extended likelihood (4.4) shares many of the properties and advantages of the standard maximum likelihood method, and can lead to improved estimators $\hat{\theta}$ if the parameter of the Poisson distribution ν is given as a function of θ [193].

4.2.3 Uncertainty estimates

This section introduces methods to obtain statistical uncertainties and confidence intervals in this work. In the frequentist sense, the variance of an estimator is a measure of how widely the estimates would be distributed if the experiment were to be repeated many times with the same sample size. This should not be confused with a probability that the true value of the parameter is found within the interval – in the frequentist notion, it either is, or it's not. In most cases, the square root of the variance – the standard deviation σ – is reported as statistical uncertainty. The limitations of this approach are discussed and the statistically cleaner notion of a confidence interval is introduced.

Statistical uncertainty

In the large sample limit, it can be shown (under certain conditions), that the maximum likelihood estimator is normal [190], *i.e.* described by a Gaussian p.d.f.. And in fact, for most practical problems, the maximum likelihood estimates are Gaussian distributed. As a consequence, the standard deviation (covariance matrix) of a (multidimensional) estimator effectively summarises all of the information about how repeated experiments would be distributed.

Assuming an efficient and unbiased estimator of a scalar parameter θ, the variance $\mathbb{V}$ of the estimator is given by the endpoint of the Cramér-Rao lower bound, *i.e.*

$$\mathbb{V}\left[\hat{\theta}\right] \geq \mathbb{E}\left[\partial_\theta^2 \ell\right]^{-1} \,. \tag{4.5}$$

The right hand side of Eq. (4.5) can be solved analytically, *i.e.* by carrying out the integrals to get the expected value ($\mathbb{E}[X] = \int x f(x) \mathrm{d}x$). Here however, extending the problem to a multidimensional parameter $\boldsymbol{\theta}$ and using the invariance under change of parameter of the maximum likelihood solution, the entries i, j of the inverse of the estimated covariance matrix $\hat{\boldsymbol{V}}$ are given by

$$\left(\hat{\boldsymbol{V}}^{-1}\right)_{ij} = \partial^2_{\theta_i,\theta_j} \ell \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \,. \tag{4.6}$$

In this work, the second derivative is calculated numerically using the `HESSE` subroutine of `MINUIT` [192].

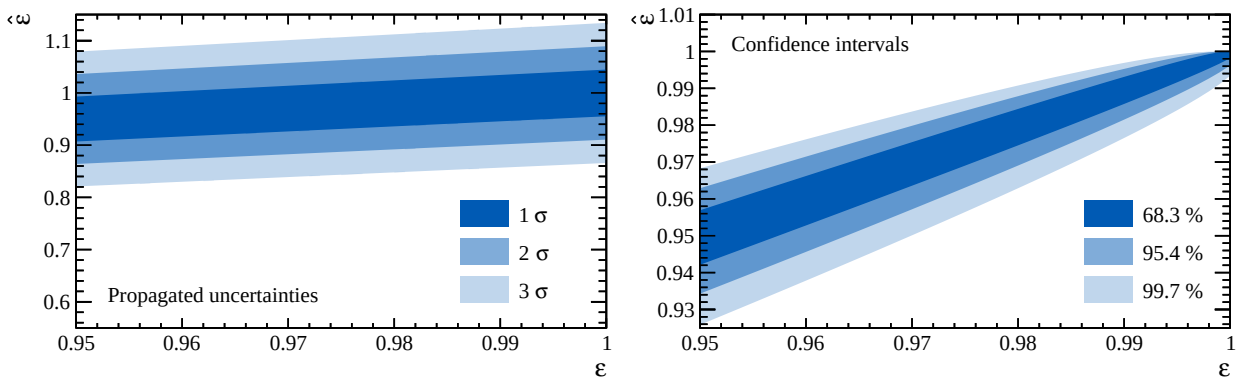


Figure 4.2: Confidence intervals for a ratio of two numbers $\varepsilon = \frac{k}{n}$, where $n = 1000$ and $k \in [950, 1000]$.
(a) Propagation of uncertainty assuming k and n are independent Poisson distributed variables, with estimated standard deviation $\hat{\sigma}(k) = \sqrt{k}$; n analogous.(b) Confidence interval[1] of ε, assuming that k is binomial distributed.

[1]The plot shows a Clopper-Pearson interval [195] as implemented in ROOT [C37].

Confidence intervals

In cases, where the conditions for asymptotic normality of the maximum likelihood estimator are not fulfilled, a confidence interval that is defined by Eq. (4.6) will have bad properties. The interval $[a, b]$ (which depends on the estimator) is the $(1-\alpha)\cdot 100\,\%$ (two-sided) confidence interval for θ if it contains the true value θ^*, *i.e.* if

$$\mathbb{P}\left(a(\hat{\theta}) \leq \theta^* \leq b(\hat{\theta})\right) = 1 - \alpha \,. \tag{4.7}$$

Estimated intervals are called **valid**, if the coverage probability $\mathbb{P}\left(a(\hat{\theta}) \leq \theta^* \leq b(\hat{\theta})\right)$ is equal to, or slightly larger than the confidence level $1-\alpha$ as the number of experiments goes to infinity. The case $\mathbb{P}\left(a(\hat{\theta}) \leq \theta^* \leq b(\hat{\theta})\right) > 1-\alpha$ is called *overcoverage*, the contrary $\mathbb{P}\left(a(\hat{\theta}) \leq \theta^* \leq b(\hat{\theta})\right) < 1-\alpha$ is called *undercoverage*.

If the p.d.f. of the estimator is known, *e.g.* by means of a Monte Carlo study, "exact" confidence intervals – in the sense that they do not undercover – can be constructed; this is known as *Neyman construction* [194, 191]. In practice, most methods to construct confidence intervals are derived from approximations of estimator properties, which can be validated by pseudo-experiments.

In this work, confidence intervals will be used in the context of efficiencies. This means if the term uncertainty is related to efficiencies later, it actually means a 68.27 % confidence interval, *i.e.* the equivalence of the standard deviation ($1\,\sigma$) of a Gaussian estimator. However, estimators for efficiencies should be binomial and not Gaussian. In the exaggerated example shown in Fig. 4.2, an efficiency is once treated as ratio of two random variables, and the other time as parameter in a binomial process. By adding this information, it is seen that the estimator is much more efficient. Moreover, the domain of the variable is respected ($\varepsilon > 1$ is not defined). Finally, it is seen that the Gaussian estimators produce straight parallel limiting lines, while the estimator derived from the binomial distribution forms a curved confidence interval at the domain limit. A more technical discussion of estimators for efficiencies is diverted to appendix A.

4.2.4 sPlot

A tool to unfold data distributions called ${}_s\mathscr{P}lot$ [196] is presented. In this work it is used to subtract signal from background components in a statistically optimal way. The tool makes use of a *discriminating variable* in which the distribution of all *sources of events* is known. The discriminating variable allows to infer *control variable* distributions of the individual sources of events. An example for this is the following: signal and background are defined by a fit to an invariant mass spectrum; that fit can then be used to infer the signal- or background p_{T} distribution. In this case signal/background are the sources of events (denoted as d in the following), the invariant mass is the discriminating variable ($\boldsymbol{\mu}$), and p_{T} is the control variable ($\boldsymbol{\omega}$).

The formalism is now introduced along the lines of the original publication [196], with notation adapted to this work. There is an important limitation to the method, which needs to be made clear first: Unfolded distributions can only be obtained if the control variable p.d.f. $M(\boldsymbol{\omega})$ and the p.d.f. of the discriminating variable(s) $f(\boldsymbol{\mu})$ factorise. That is, the joint p.d.f. $f(\boldsymbol{\omega},\boldsymbol{\mu})$ can be written as product of two p.d.f.s $M(\boldsymbol{\omega})\cdot f(\boldsymbol{\mu})$. If this requirement is not fulfilled, the resulting unfolded distributions will likely be wrong. A way to test this, is to plot the discriminating variable in intervals/bins of the control variable and see if the *shape* (not the normalisation) of all sources is the same.

The task of unfolding for a specific source of events d, is to estimate the distribution of a control variable $\widehat{M}_{\mathrm{d}}(\boldsymbol{\omega})$. The ansatz here is to find a weight $\mathscr{P}_{\mathrm{d}}$ that projects out that distribution of source d from the distribution in which all n_{d} sources are mixed $M(\boldsymbol{\omega}) = \sum_{l=1}^{n_{\mathrm{d}}} M_l(\boldsymbol{\omega})$. A first guess to such a weight would be the fraction of events of type d in a small region $\boldsymbol{\delta\omega}$ with mean value $\overline{\boldsymbol{\omega}}$. That is nothing else than choosing (small) intervals in $\boldsymbol{\omega}$, and making a fit in each of these intervals to get the fit fractions of source d. If the fit is an extended maximum likelihood fit, the fit fraction of d in $\boldsymbol{\delta\omega}$ is the estimated number of events $N_{\mathrm{d}}^{\boldsymbol{\delta\omega}}$. Putting this into a formula, the searched weight of event i in the interval $\boldsymbol{\delta\omega}$ is

$$\mathscr{P}_{\mathrm{d}}(\boldsymbol{\mu}_i) = \frac{N_{\mathrm{d}}^{\boldsymbol{\delta\omega}} f_{\mathrm{d}}(\boldsymbol{\mu}_i)}{\sum_{l=1}^{n_{\mathrm{d}}} N_l^{\boldsymbol{\delta\omega}} f_l(\boldsymbol{\mu}_i)} \qquad \forall \left\{\boldsymbol{\mu}_i | \boldsymbol{\omega}_i \in \boldsymbol{\delta\omega}\right\} .$$

Here, the expression $\left\{\boldsymbol{\mu}_i | \boldsymbol{\omega}_i \in \boldsymbol{\delta\omega}\right\}$ states, that the measured variables $\boldsymbol{\mu}_i$ and $\boldsymbol{\omega}_i$ of the event i is contained in the interval $\boldsymbol{\delta\omega}$. This weight $\mathscr{P}_{\mathrm{d}}\ (\boldsymbol{\mu}_i)$ now allows to estimate the projected (normalised) distribution of the control variable in $\boldsymbol{\delta\omega}$:

$$N_d \widehat{M}_{\mathrm{d}}(\overline{\boldsymbol{\omega}})\boldsymbol{\delta\omega} = \sum_i \mathscr{P}_{\mathrm{d}}(\boldsymbol{\mu}_i) \qquad \forall \left\{\boldsymbol{\mu}_i | \boldsymbol{\omega}_i \in \boldsymbol{\delta\omega}\right\} .$$

By letting the size of the intervals $\boldsymbol{\delta\omega}$ become infinitesimal, and taking the expected value of N, the true distribution $M^*(\overline{\boldsymbol{\omega}})$ is retained:

$$\begin{aligned} \mathbb{E}\left[N_{\mathrm{d}}\widehat{M}_{\mathrm{d}}(\overline{\boldsymbol{\omega}})\right] &= \int \mathrm{d}\boldsymbol{\mu} \sum_{l=1}^{n_d} N_l f_l(\boldsymbol{\mu}) \boldsymbol{\delta}(\boldsymbol{\omega}(\boldsymbol{\mu}) - \overline{\boldsymbol{\omega}}) \mathscr{P}_{\mathrm{d}}(\boldsymbol{\mu}) \\ &= N_d \int \mathrm{d}\boldsymbol{\mu} f_d(\boldsymbol{\mu}) \boldsymbol{\delta}(\boldsymbol{\omega}(\boldsymbol{\mu}) - \overline{\boldsymbol{\omega}}) = N_{\mathrm{d}} M^*_{\mathrm{d}}(\overline{\boldsymbol{\omega}}) . \end{aligned} \tag{4.8}$$

From the first to the second line, it is used that the weight $\mathscr{P}_{\mathrm{d}}(\boldsymbol{\mu})$ projects out the contribution of species d. However, the so defined weights suffer from the implicit dependency of the control variable $\boldsymbol{\omega}$ on the discriminating variable $\boldsymbol{\mu}$, which enter via the delta-function $\boldsymbol{\delta}(\boldsymbol{\omega}(\boldsymbol{\mu}) - \overline{\boldsymbol{\omega}})$. This dependency forces to stick to the paradigm of performing fits in (infinitesimally) small intervals of $\boldsymbol{\delta\omega}$. The ${}_s\mathscr{P}lot$ formalism provides a solution to overcome this problem.

The line of thought is as follows: assume that the (unknown) true distribution $M^*(\boldsymbol{x})$ can be factored out from Eq. (4.8), as

$$\begin{aligned}\mathbb{E}\left[N_\mathrm{d}\widehat{M}_\mathrm{d}(\overline{\boldsymbol{\omega}})\right] &= \int \mathrm{d}\boldsymbol{\mu} \int \mathrm{d}\boldsymbol{\omega} \sum_{l=1}^{n_d} N_l f_l(\boldsymbol{\mu}) M_l^*(\boldsymbol{\omega}) \boldsymbol{\delta}(\boldsymbol{\omega} - \overline{\boldsymbol{\omega}}) \mathscr{P}_\mathrm{d}(\boldsymbol{\mu}) \\ &= N_\mathrm{d} \sum_{l=1}^{n_d} M_l^*(\overline{\boldsymbol{\omega}}) \left(N_l \int \mathrm{d}\boldsymbol{\mu} \frac{f_l(\boldsymbol{\mu}) f_\mathrm{d}(\boldsymbol{\mu})}{\sum_{l'=1}^{n_d} N_{l'} f_{l'}(\boldsymbol{\mu}_i)} \right) . \end{aligned} \tag{4.9}$$

It can be seen from this expression that the true control variable distribution of species d ($N_\mathrm{d} M_\mathrm{d}^*(\overline{\boldsymbol{\omega}})$) can only be projected out if $f_\mathrm{d}(\boldsymbol{\mu})$ is fully discriminating (*i.e.* a simple cut can be applied which only selects species d). However, there is a statistical solution that can correct for the term in parenthesis in Eq. (4.9). It is given by the expectation value of the inverse of the covariance matrix of the extended maximum likelihood estimator. Having found this relation is at the core of ${}_s\mathscr{P}lot$.

So after inserting back that solution to the equations above and solving for the searched weight, it is seen that this weight – which projects out the control variable distribution $N_d \widehat{M}_\mathrm{d}(\overline{\boldsymbol{\omega}})$ – is given by:

$${}_s\mathscr{P}_\mathrm{d}(\boldsymbol{\mu}_i) = \frac{\sum_{l=1}^{n_d} \widehat{\boldsymbol{V}}_{\mathrm{d}l} f_l(\boldsymbol{\mu}_i)}{\sum_{l'=1}^{n_d} N_{l'} f_{l'}(\boldsymbol{\mu}_i)} . \tag{4.10}$$

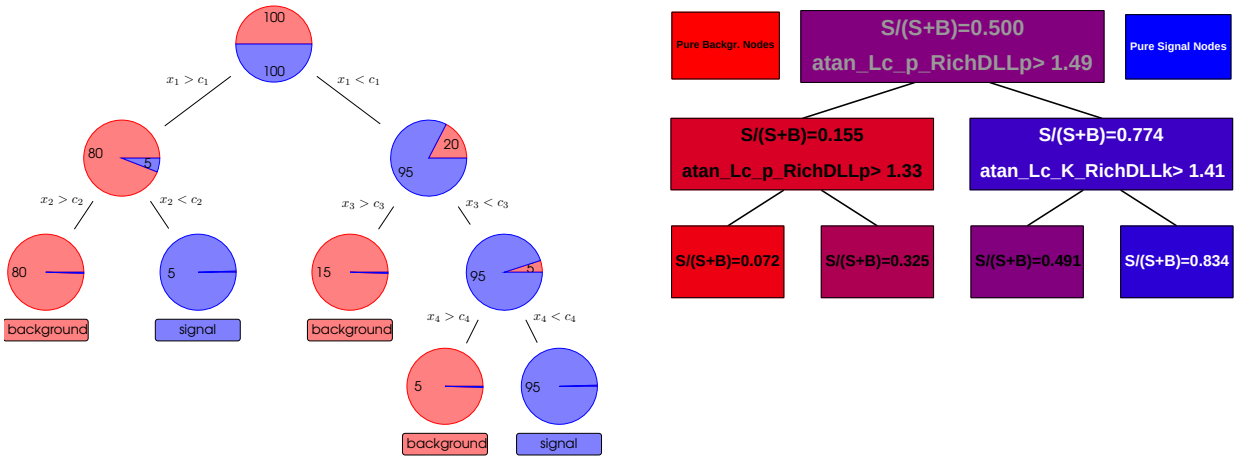
This is the **sWeight**. Here, $\widehat{\boldsymbol{V}}$ is the estimated covariance matrix found by an extended maximum likelihood fit. Statistical uncertainties of *sWeighted* distributions are given by the sum of squared sWeights $\sqrt{\sum_i ({}_s\mathscr{P}_d(\boldsymbol{\mu}_i))^2}$ in the asymptotic limit.

4.2.5 Gradient boosted decision trees

Suppose, as in case of ${}_s\mathscr{P}lot$, a sample composed of multiple sources of events is given, *e.g.* signal and background. The task is to predict whether an event is signal or background based on a set of input variables $\boldsymbol{x}$. This task is called **classification**, and there exist a variety of tools which do this classification in a statistically optimised way. A widely used classification tool are *boosted decision trees*, which are literally decision trees that are boosted. Both terms, the decision tree and the boosting will be discussed in this section. In this work, it has been found that *gradient boosting* algorithms yield superior results compared to other types of boosting, such that this specific boosting algorithm is detailed.

Decision trees

Up to now, only the task, a sample with multiple sources of events and a set of input variables have been introduced. It is not clear and *how* the classification is done. In principle, there are two options to this *how*: they go by the name of *supervised* and *unsupervised machine learning*. Here, supervised learning is used. That means that a **training** sample in which the inputs variables and their corresponding class (signal or background) are known. With the help of this sample, a set of rules is created, which in the case of a decision tree (DT) is just a series of simple cuts, like the ones shown in Fig. 4.3. This part is usually called *growing* a DT. The way how a DT is grown depends on the boosting algorithm and will thus be explained later.



(a) (b)

Figure 4.3: (a) Schematic view of a decision tree. The numbers are to be understood as the sum of weights in signal and background categories. The constants c_i, which are obtained in the training, determine if an event with parameter values $x_i \in \boldsymbol{x}$ will be classified as signal or background in this DT. Taken from [197].
(b) A DT obtained with TMVA during studies described in Sec. 5. The sums of weights in signal (S) and background categories (B) were normalised the root node of the DT. In contrast to the schematic tree, this tree is not fully grown, *i.e.* not every event of training sample has been classified correctly.

The DT shall now be examined by taking the path of an input sample. Starting at the top, the *root node*[1], both classes, signal (S) and background (B), are evenly distributed during training. Note that this is not a mandatory requirement of the training sample; if the number of signal and background events is not the same, the samples can be weighted to contain the same sum of weights. From the root node, the sample is split into subsamples that pass or fail a given cut/selection rule. In the example of Fig. 4.3(b) all events with a value of `atan_Lc_p_RichDLLp` greater than 1.49 take the path to the right, while the rest of the events take the left path. In both examples of Fig. 4.3 the right path has a higher signal purity, but this is not a strict rule. Splitting of the sample continues until certain stopping criteria are fulfilled. These can usually be configured by the user. If no stopping criteria would be defined, the splitting would continue until every single event of the training sample is classified correctly. In such a case, the tree is *fully grown*, as the one shown in Fig. 4.3(a). In other words: all *terminal nodes* are either classified as signal or background.

After training, which *grows* the DT, the DT is applied to real or test data. Assume, that the test data contains an event with `atan_Lc_p_RichDLLp` $= 1.53$, `atan_Lc_K_RichDLLk` $= 1.40$ (and many other variables) to which the DT in Fig. 4.3(b) should be applied. It does not matter what these variables mean right now, but they are explained in Sec. 5.2. Here they are used to deduce what the output of the DT will be. On the root node, the right path would be taken, since the measured `atan_Lc_p_RichDLLp` value is in fact greater than 1.49. At the next node, the DT checks if `atan_Lc_K_RichDLLk` is greater than 1.41. In the example this is not the case, so the left path is taken. The purity $S/(S+B)$ in the training sample at this stage is 0.491. This is a terminal node, where the DT has to return its decision. Given the default settings of the software that produced the plot, that decision would be “background”. In practice, the output signal or background is translated to $+1$ or -1 respectively.

[1] Unlike trees in nature, DTs grow from top to bottom.

The example implies that a single decision tree, returning a binary output, has several drawbacks. One is that the example event will be classified as signal, since `atan_Lc_K_RichDLLk` $= 1.40$ pushes it into a signal-like terminal node. But it is clear from the purity figures that this decision will more likely be wrong than that of a signal or background event in the leftmost or rightmost terminal nodes respectively. Another issue is that of *overtraining*. In the example case, the value of `atan_Lc_K_RichDLLk` is very close to the cut value which decides whether the event is classified signal or background. A statistical fluctuation, either in the training sample, leading to a "wrong" rule, or in the test sample by *e.g.* a resolution effect, can easily push the event to the signal- or background side of the DT.

These problems are overcome by using many DTs, and combining their outputs by boosting. When their outputs are combined, each DT still returns $+1$ or -1, but the boosting algorithm will give them a weight, such that the *response* from a training or test sample will be a distribution between -1 and $+1$. A boosted decision tree (BDT) is thus not a tree in the literal sense, but rather a *decision forest*. The algorithm of gradient boosting requires the trees to be grown as regression trees, which are introduced next.

Regression trees

Even though it is the goal to train a BDT for a classification task, gradient boosting relies on regression trees in the minimisation that is detailed later. A regression tree has the same basic structure as the DTs introduced above. The difference is, that regression trees will return a real valued rather than a boolean response.

The open question of how to grow a DT is now addressed in the context of regression trees. In supervised learning, the system under study is consisting of a multi-dimensional vector of input variables $\boldsymbol{x}$ and a *response* variable y. In the training phase, the pairs of inputs $\boldsymbol{x}_i$ and responses y_i are known. This knowledge allows to grow the tree. The growing algorithm now needs to decide automatically on the splitting variables an the split points. The resulting tree will be fully described by its terminal nodes, since they partition the input space into disjoint regions R_j, each with response b_j. This can be expressed as

$$T(\boldsymbol{x}) = \sum_{j=1}^{J} b_j \mathbf{1}_{R_j}(\boldsymbol{x}) \,, \tag{4.11}$$

where $\mathbf{1}_{R_j}(\boldsymbol{x})$ is 1 if the input $\boldsymbol{x}$ is in region R_j and 0 otherwise. An example is shown in Fig. 4.4, for which the regression model might be written as

$$\hat{T}(\boldsymbol{X}) = \sum_{j=1}^{5} b_j \mathbf{1}_{R_j}((X_1, X_2)) \,.$$

When growing a regression tree that should predict a response in a region R_j, it makes sense to estimate

$$\hat{b}_j = \text{ave}(y_i | \boldsymbol{x}_i \in R_j) \,, \tag{4.12}$$

i.e. that the response of the regression tree is the average of the training responses in each region. This also minimises the sum of squares $\sum_i (y_i - T(\boldsymbol{x}_i))^2$. For finding the regions, the least squares method is computationally not feasible and binary splits considering only one input variable, as in Fig. 4.4, rather than the whole input space are performed. The regions are searched by a *greedy* algorithm that probes all $\dim(\boldsymbol{x})$ input variables at split points t using the whole training sample. For simplicity, take a single split into regions R_1 and R_2, which should solve

$$\min_{l,t} \left[\min_{b_1} \sum_{\boldsymbol{x}_i \in R_1(l,t)} (y_i - b_1)^2 + \min_{b_2} \sum_{\boldsymbol{x}_i \in R_2(l,t)} (y_i - b_2)^2 \right] . \tag{4.13}$$

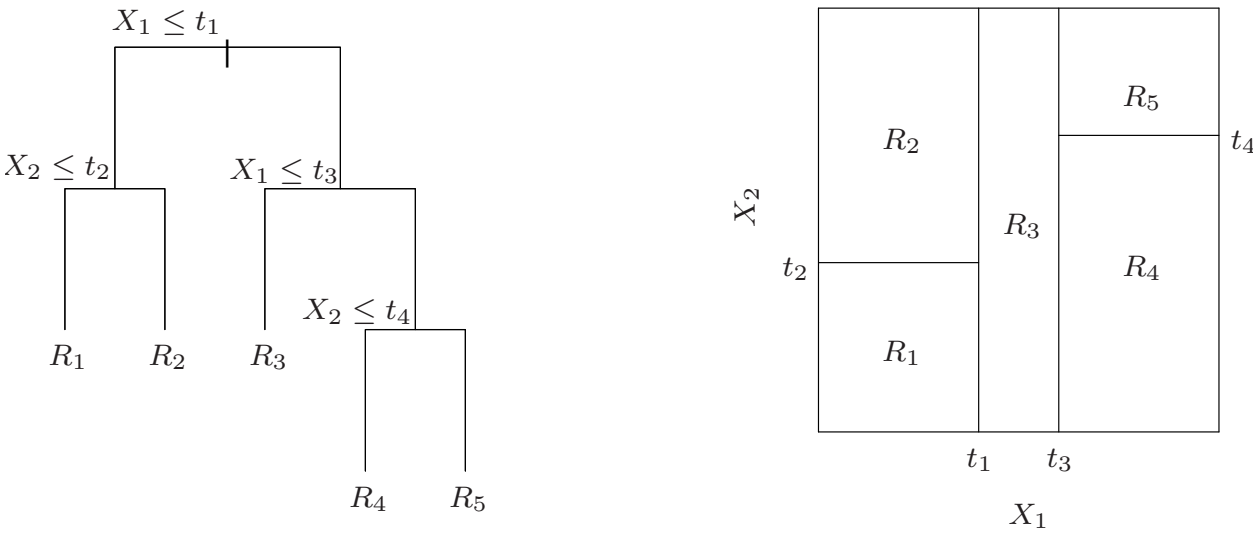


Figure 4.4: A DT (a) partitions an input space in two dimensions X_1 and X_2 by four splits $t_1 \ldots t_4$ into five regions $R_1 \ldots R_5$ (b). Taken from Ref. [198].

Here, l labels one of the input parameters, *i.e.* $l \in 1 \ldots \dim(\boldsymbol{x})$. The inner minimisation is solved by the averages of the training responses (4.12), and the outer minimisation is solved by scanning a predefined number of splits in each input variable[1]. It is straight forward to expand the minimisation to arbitrary trees (4.11).

To summarise: regression trees are grown by a greedy algorithm that scans for splits in the input parameter space that minimise the squared difference between training response and tree output. Several of such trees shall be used in the following. Experience has shown that the combination of the order of a thousand such trees, each relatively small and therefore called *weak learner*, leads to better results than any combination of larger trees. How this combination is done is subject to the following paragraph.

Gradient boosting

The boosting algorithm defines the parameters of a BDT during training. It is briefly summarised here how this is done in TMVA [179], the software used in this work. Their algorithm is based on the original publication on gradient boosting, Ref. [199].

Gradient boosting stars from a problem that is common to any algorithm in supervised machine learning. The goal is to obtain an estimate $\hat{F}(\boldsymbol{x})$ of a function $F^*(\boldsymbol{x})$ that maps the input $\boldsymbol{x}$ to the response y in a statistically optimised way. To find the optimal estimate for $F^*(\boldsymbol{x})$, a **loss function** $L(y, F(\boldsymbol{x}))$ is introduced, which penalises misclassification of events during training. That loss function can *e.g.* be the squared error $L(y, F(\boldsymbol{x})) = (y - F(\boldsymbol{x}))^2$. The statistically optimal function $F^*(\boldsymbol{x})$ thus minimises the expected value of that loss function:

$$F^*(\boldsymbol{x}) = \arg\min_F \mathbb{E}_{\boldsymbol{x},y}\left[L(y, F(\boldsymbol{x}))\right] \ . \tag{4.14}$$

The function optimisation problem, Eq. (4.14), shall be solved numerically by incremental functions (*boosts*).

[1] In TMVA, this configurable parameter is called `nCuts`.

That numerical solution is **gradient-descent**, which iteratively follows the steepest local gradient to arrive at a global minimum. Gradient-descent introduces recursively defined local gradients g_m at step m, with step size ρ_m. Using gradient-descent, the searched optimal function can be written as

$$F^*(\boldsymbol{x}) = \sum_{m=0}^{M} -\rho_m g_m(\boldsymbol{x}) = -\rho_m \, \partial_{F(\boldsymbol{x})} \mathbb{E}_{\boldsymbol{x},y} \left[L(y, F(\boldsymbol{x})) \right] \Big|_{F(\boldsymbol{x})=F_{m-1}(\boldsymbol{x})} \, .$$

Note how the function of the previous step $F_{m-1}(\boldsymbol{x})$ enters in the loss function.

It is worthwhile to back-pedal at this point to get a global view. The function that optimally maps the classifier inputs to the response is searched in a function minimisation problem (4.14). The iterative method of gradient descent allows to search that minimum by stepping towards the direction of steepest descent. This direction is given by the local gradients $-g_m$, but the step sizes, given by the ρ_ms need to be known as well. They are the solution to *line search*:

$$\rho_m = \arg\min_{\rho} \mathbb{E}_{\boldsymbol{x},y} \left[L(y, F_{m-1}(\boldsymbol{x}) - \rho \, g_m(\boldsymbol{x})) \right] \, . \tag{4.15}$$

Thus, the function minimisation became a iterative parameter minimisation problem.

The solution to the line search problem (4.15) can be approximated by *fitting* regression trees to the negative gradients. That is because the loss function is maximally reduced by the gradients, each pointing from the prediction of the previous estimate $F_{m-1}(\boldsymbol{x})$ toward the direction of the true response. The parameters $\Theta = (b_j, R_j)$ of a regression tree fitting the negative gradients are given by

$$\hat{\Theta}_m = \arg\min_{\Theta} \sum_{i=1}^{N} (-g_{im} - T(\boldsymbol{x}_i; \Theta))^2 \, . \tag{4.16}$$

The gradients at each training event g_{im} are also called residuals or pseudo-responses, and are given by

$$\tilde{y}_{i,m} = -g_m(\boldsymbol{x}_i) = -\partial_{F(\boldsymbol{x}_i)} \left[L(y_i, F(\boldsymbol{x}_i)) \right] \Big|_{F(\boldsymbol{x}_i)=F_{m-1}(\boldsymbol{x}_i)} \, .$$

The derivative of the loss function is simple to calculate, such that, choosing a binomial log-likelihood loss function as in TMVA, the residuals are

$$\tilde{y}_{i,m} = \frac{2y_i}{1 + e^{2y_i F_{m-1}(\boldsymbol{x}_i)}} \, .$$

Note that this loss function to determine the step sizes ρ_m is different from the loss function that has been used to grow the regression trees, which has been chosen as squared error loss in (4.16). In TMVA, the standard settings for the loss functions are binomial log-likelihood and Huber loss [200], a hybrid of squared and absolute loss to be less sensitive to outliers.

In summary, classification with gradient boosting estimates a function that maps a set of input parameters to a real valued response. The function is obtained using the numerical method of gradient-descent. At each iteration, a regression tree is fitted to minimise the residuals of the previous step. The resulting classifier response, the estimate $\hat{F}(\boldsymbol{x})$, is a weighted sum of shallow decision trees.

Hyperparameters and Regularisation

So far, the classifier response is fully determined by the training sample, the number of boosting steps/DTs M, and the loss functions. There are however several settings, also known as **hyperparameters**, that TMVA and other packages provide to steer the training. Apart from the number of trees, the size of trees and nodes can be configured. Further options concern the **regularisation**: the stability of the algorithms performance and the prevention of overfitting. One of the hyperparameters to regularise gradient BDTs is the *shrinkage* ν, sometimes also called learning rate, which damps the weight of each boost

$$F_m(\boldsymbol{x}_i) = F_{m-1}(\boldsymbol{x}_i) + \nu\rho_m T(\boldsymbol{x}_i;\Theta_m)\,, \qquad 0 \leq \nu \leq 1\,. \tag{4.17}$$

There is a trade-off between shrinkage and the number of boosting steps (and thus also with computational power), so that these should be monitored simultaneously in hyperparameter optimisation. An additional method to prevent overfitting, is called *bagging* – short for bootstrap aggregating. Bagging uses a resampled training dataset at each boosting step. In TMVA this is done by assigning a weight according to a Poissonian random number with the bagging-fraction (0.6 by default in TMVA) as mean parameter to each training event.

5 Identification of non-prompt charm hadrons

This section describes the development and calibration of PID-like variables for the identification of non-prompt $\Lambda_c^+ \to pK^-\pi^+$, $D^0 \to K^-\pi^+$ and $D_s^+ \to K^+K^-\pi^+$ decays. The variable can thus be understood as a portable classification tool for a decay, contrary to the `ProbNN` PID variables for single particles (*cf.* Sec. 2.2.5). The new variables play a crucial role in the offline selection of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays.

In order to classify non-prompt charm hadrons, a multivariate selection based on supervised machine learning algorithms is used. Those algorithms are known to produce better background rejection at a given signal efficiency when compared to standard cut-based or likelihood-based selections. Machine learning tools are widely used in LHCb, the prime example being the `ProbNN` PID variables to classify π, K, p, e and μ candidates.

The idea for a generic non-prompt charm hadron selection, termed "*D-from-B* BDTs", has been developed in Ref. [I6] and applied to the analysis of beauty hadron decays into pairs of open charm particles in Refs. [175, 174]. It could be shown in these analyses, that a selection based on such *D-from-B* BDTs outperforms a selection using the `ProbNN` PID variables and kinematic variables in a standard way. The *D-from-B* BDTs not only draw strength from combining PID and kinematic variables in an optimised way by exploiting the correlations of the charm hadron daughters; but also from using information from especially the RICH, but also muon and tracking detectors directly.

After their first application in Ref. [175], the *D-from-B* BDTs were updated for Ref. [174]. For this work, they are updated once more[1], and they need to be applicable to Run II data in the near future. Moreover, an update of the $D^+ \to K^-\pi^+\pi^+$ BDT is planned. Also, the concept can be expanded to further charm hadrons, like Ξ_c and η_c, but also to other decay channels, excited states, or other secondary intermediate states, like the ϕ meson or other strange hadrons. To minimise the amount of effort needed to develop and calibrate such BDTs, an efficient and scalable software framework is needed.

The task of developing such a software has been picked up by the Heidelberg LHCb group, and a successful implementation was done in the bachelor thesis of Christoph Otte [177], which has been tested in the master thesis of Florian Reiß [176]. However, a dependence on the year of data taking has been found, as well as imperfections in the efficiency correction for tight selections. The workflow is revisited and improved in this work. The data is split by data taking period and it is shown that the efficiency correction of the revisited *D-from-B* BDTs works over a wide range of cuts.

During the efficiency correction procedure, a known problem has been encountered, namely that common binomial confidence intervals are not valid for data-driven efficiencies. That is because the signals which are needed to measure the efficiencies have to be separated from backgrounds, which distort the commonly used intervals. A solution to this problem is proposed in appendix A, and has been integrated into `beef`, the custom-build fitting framework (*cf.* Sec. 4.1).

The *D-from-B* are trained and evaluated on calibration data, namely the decays $\Lambda_b^0 \to \Lambda_c^+\pi^-$, $B^- \to D^0\pi^-$ and $\overline{B}_s^0 \to D_s^+\pi^-$. The procedure is entirely data-driven, *i.e.* no information from simulation is used. Calibration channel selection before the BDT training procedure is detailed in Sec. 5.1. Input variables and hyperparameter settings of the BDT training are described in Sec. 5.2, followed by details on the individual $\Lambda_c^+ \to pK^-\pi^+$, $D^0 \to K^-\pi^+$ and $D_s^+ \to K^+K^-\pi^+$ BDTs. The most challenging part is the efficiency calibration in Sec. 5.6, followed by a closure test on the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channel. The code related to this section is available in a GitLab repository [C11].

[1] The earlier BDTs have been trained on a different stripping version (`v20`, now: `v21`). A global cut on a figure correlated with the RICH occupancy has been relaxed (`nLongTracks` from 250 to 500). Since the PID efficiencies depend on the occupancy, the BDTs needed to be re-trained.

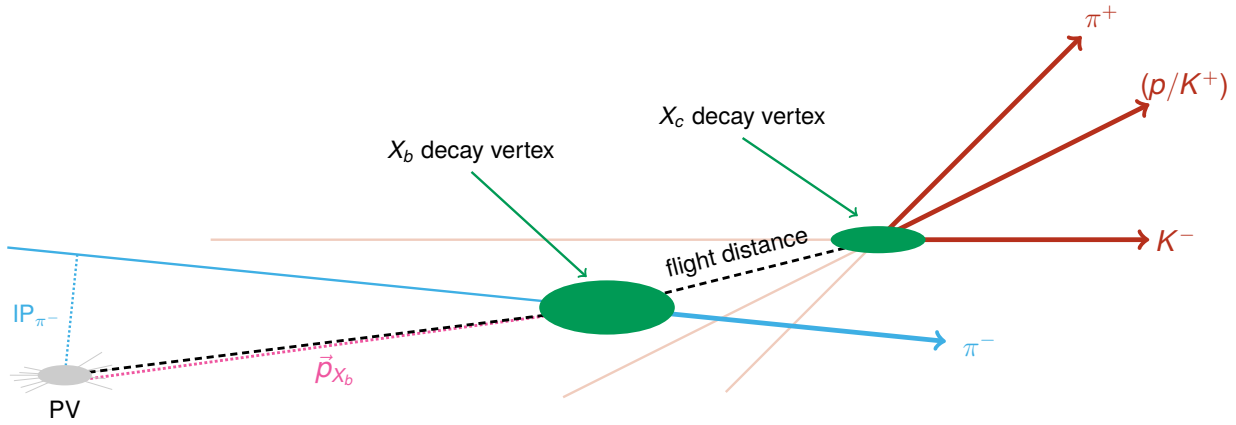


Figure 5.1: Decay topology of $X_b \to X_c\pi^-$ decays used as BDT calibration channels. Primary vertex (PV), X_b and X_c decay vertices are typically measured to be well separated due to dilated decay lengths of the boosted beauty and charm hadrons. The size of the ellipse at the vertices implies different spatial resolutions, which become more precise with a larger number of particles emerging from a vertex.
Final state particles are drawn as thick lines with arrows; their tracks are reconstructed as approximately straight lines in the interaction region as indicated by the thinner back-extended lines. The spatial trajectory of intermediate particles is drawn as black dashed line, while their momenta, as combined from the respective daughter particles, might not point into exactly the same direction (the dotted magenta line). The cosine of the angle between the two vectors is known as direction angle (DIRA).
Impact parameters (IPs) of all final state and intermediate particles help to further discriminate against combinatorial backgrounds; here, only the IP of the bachelor pion is drawn.

5.1 Calibration channel selection

The selection of the *D-from-B* BDT calibration channels is described in detail in the following. The calibration decays $\Lambda_b^0 \to \Lambda_c^+\pi^-$, $B^- \to D^0\pi^-$ and $\bar{B}_s^0 \to D_s^+\pi^-$ are collectively referred to as $X_b \to X_c\pi^-$ decays in the following. The trigger and stripping selection of the $X_b \to X_c\pi^-$ and the $\Lambda_b^0 \to \Lambda_c^+\bar{D}^{(*)0}K^-$ and $\Lambda_b^0 \to \Lambda_c^+D_s^-$ decays are based on a common module which is described here.

The section is subdivided into a brief discussion of the decay topology of b hadrons (Sec. 5.1.1), the LHCb-wide stripping selection (Sec. 5.1.2), an offline selection strategy for all calibration channels (Sec. 5.1.3), and the strategy to define signal- and background samples for BDT training (Sec. 5.1.4).

5.1.1 Event topology of *b*-hadron decays

Some basic considerations of the decay topology of b hadrons in LHCb are put in front of the detailed description, since the selection of $X_b \to X_c\pi^-$ decays is almost exclusively based on geometrical and kinematic variables describing the topology. The $X_b \to X_c\pi^-$ decay topology is illustrated in Fig. 5.1.

Through the predominant b-hadron production mechanism at the LHC via gluon fusion, the system is strongly boosted in either forward or backward direction. Consequently, typical b-hadrons decay lengths are dilated to the order of a centimetre and can be measured as well separated from the primary vertex (PV) by the VELO. Typical impact parameters (IPs) of stable final state particles from secondary vertices are in the order of a millimetre, which is large compared to the IP resolution of the order of 50 µm (*cf.* Sec. 2.2.4), and can therefore be well separated from particles that emerge from the PV. Further suppression of backgrounds come from requiring that particle momenta and track trajectories point to their decay vertex. If this concerns a weakly decaying particle, the cosine of the angle between the reconstructed momentum vector and the vector between production and decay vertex, known as direction angle (DIRA), should be close to 1.

5.1.2 Trigger and Stripping selection

At the trigger stage, candidates are required to be either triggered on signal (TOS) by the L0 hadron trigger line or triggered independent of signal (TIS) by any other of the physics trigger lines at L0. The HLT1 selection is implicitly encoded in the HLT2 selection required by the stripping module and draws from the inclusive charm and beauty line, `HLT1TrackAllL0` (*cf.* Sec. 2.2.3). At the HLT2 stage, candidates are required to be TOS or TIS on the logical `OR` of the topological and inclusive ϕ trigger lines, constituting about 40 to 45 % of the overall HLT2 output rate [C26]. In summary, the trigger lines are chosen to select inclusive hadronic b-hadron decays.

Next, the stripping selection is summarised, beginning with the LHCb specific condensed information, which is explained in the following. All selection steps, or cuts, are numbered and listed in Tabs. 5.1–5.3. The tables are divided into global and track selection, selection of intermediate particles and finally selection of the b-hadron candidate. The stripping is based on the `Beauty2Charm` module using the following lines:

$\Lambda_b^0 \to \Lambda_c^+ \pi^-$ `StrippingLb2LcPiNoIPLc2PKPiBeauty2CharmLine` [C35]

$B^- \to D^0 \pi^-$ `StrippingB2D0PiD2HHBeauty2CharmLine` [C33]

$\bar{B}_s^0 \to D_s^+ \pi^-$ `StrippingB02DPiD2HHHBeauty2CharmLine` [C32]

The lines are part of the `BhadronCompleteEvent` stream of the 2014 legacy stripping of Run I data (`Stripping21(r1)`) [C43], based on the official production version `Reco14` [C27, C24].

Track selection

#	Cut	Applied to	Purpose
0	`Hlt2Topo.*Decision OR Hlt2IncPhi.*Decision`	global	HLT2 selection stream
1	`CloneDist > 5000`	all tracks	Clone track rejection
2	`TRCHI2DOF < 3`	all tracks	Good track quality
3	`nLongTracks < 500`	global	Remove outliers, limit combinatorics
4	`TRGHP < 0.4`	all tracks	Fake track rejection
5	`P > 1000 MeV`	all tracks	Remove low momentum tracks
6	`PT > 100 MeV`	all tracks	Remove low transverse momentum tracks
7	`MIPCHI2DV(PV) > 4.0`	all tracks	Reject tracks from any primary vertex

Table 5.1: Global and track based selection used at stripping level.

Using the triggered events as input (#0), the stripping selection starts from proto-particles; these are objects summarising the reconstruction information. Further objects of importance are tracks, used as a synonym for a reconstructed trajectory of a charged particle, in contrast to particles, which can be charged tracks with a mass hypothesis, but also neutral and/or intermediate particles. In case of the calibration channels the proto-particles are required to be reconstructed as long tracks, passing a clone track rejection (`CloneDist > 5000`, #1) and a track quality cut (`TRCHI2DOF < 3`, #2). The clone track rejection is based on the Kullback-Liebler divergence which measures the difference in information content between two track candidates [201].

The track quality cut is already applied in the reconstruction of long tracks and requires tracks to have a good χ^2/nDoF, which is calculated by the Kalman filter track fit [202, 203]. The total number of long tracks in the event is limited to 500 (`nLongTracks < 500`, #3), removing outliers which cause large combinatorics.

The fake track rate is reduced by a cut on the track ghost-probability (`TRGHP < 0.4`, #4). The track ghost-probability is the response of an artificial neural network. The version used in Run I is documented in Ref. [I2]. The cuts on momentum and transverse momentum can be seen as loosest selection possible given the reconstruction chain, since multi-body beauty and charm decays tend to produce at least one soft particle. The last track related cut removes tracks from primary vertices (PVs) by a cut on the minimal impact parameter (IP) χ^2 of a track projection with respect to all PVs (`MIPCHI2DV(PV) > 4.0`, #7). The IP is the distance of closest approach of a track projection to a PV calculated by a vertex fit [C21]. The corresponding χ^2 may be understood as increase of χ^2 of the PV vertex fit when the track is added to the vertex [C19]. It behaves similar to IP/(Δ IP) and is calculated using a Kalman filter. After the generic selection of tracks, the proto-particles are promoted to particles by assigning the mass hypothesis of the respective candidate.

Charm hadron selection

#	Cut	Applied to	Purpose
8	`ASUM(PT) > 1800 MeV`	c daughters	Fast soft background rejection
9	`TRCHI2DOF < 2.5 & PT > 500 MeV & P > 5000 MeV`	at least one c daughter	Require leading track
10	`ACUTDOCA < 0.5 mm`	c daughters	Save CPU time
11	`PIDp > -10`	p from X_c	Reduce combinatorics
12	`PIDk > -10`	K from X_c	Reduce combinatorics
13	`PIDk < 20`	π from X_c	Reduce combinatorics
14	`VCHI2/VDOF < 10`	X_c	c daughters emerge from same space point
15	`BPVVDCHI2 > 36`	X_c	Reject prompt charm hadrons
16	`BPVDIRA > 0`	X_c	Decay vertex downstream of PV

Table 5.2: Stripping selection of charm hadron candidates. As detailed in the text, there is a subtle difference between *combination* cuts, such as 8, and *mother* cuts (14-16). That is that the combination cut is applied to the combination of daughter particles, whereas the mother cut requires a vertex fit before the cut can be applied [C44].

Charm hadron candidates are built in the stripping and shared among all of the `Beauty2Charm` lines. A common set of cuts is used that is summarised in Tab. 5.2. The first cut listed there is the sum of scalar transverse momenta, which is applied before the vertex fit. This allowed to relax the cuts on the (transverse) track momenta (#5 and #6), since much background is rejected before the CPU intensive vertex fit. The physical reason to cut on the sum of scalar transverse momenta is the following: beam and flight direction of the beauty and charm particles are strongly correlated due to the b production mechanism at the LHC. Therefore, the sum of track p_{T} is correlated to the masses of the mother particles. In fact, if the beauty and charm hadrons have low p_{T}, the scalar sum of daughter p_{T} is close to the mothers mass. For this reason $\sum p_{\mathrm{T}} > 5(1.8)\,\mathrm{GeV}$ for beauty (X_b) and charm candidates is chosen (`ASUM(...PT) > 5000 MeV`, #17; `ASUM(PT) > 1800 MeV`, #8).

Further soft background is reduced by selecting a leading track, produced in the decay of the heavy hadron. This is done by tighter selections on track χ^2, p, p_T – and IP in case of a b hadron – (`TRCHI2DOF < 2.5 & PT > 500 MeV & P > 5000 MeV`, #9; `TRCHI2DOF < 2.5 & PT > 1700 MeV & P > 10000 MeV & MIPDV(PV) > 0.1 mm`, #18). To save computing resources, a cut on the maximum distance of closest approach between tracks forming a charm candidate is placed (`ACUTDOCA < 0.5 mm`, #10). The last cuts before performing the vertex fit of the charm candidate are PID cuts on the combined DLL variables (*cf.* Sec. 2.2.5), which are chosen as loose as possible (`PIDp > -10`, #11 for a proton candidate; `PIDk > -10`, #12 for a kaon candidate; `PIDk < 20`, #13 for a pion candidate).

The vertex fit [C21] is done with a loose vertex quality requirement (`VCHI2/VDOF < 10`, #14). The resulting vertex is required to be inconsistent with (`BPVVDCHI2 > 36`,#15) and downstream (`BPVDIRA > 0`, #16) of the PV. The former of the cuts is the χ^2 of the distance of PV to the decay vertex of the charm candidate [C20]. The latter is known as direction angle: here, it is the cosine of the angle between the c hadron momentum and the spatial vector from PV to decay vertex. Values close to 1 are desired; they indicate that both vectors point in the same direction.

Beauty hadron selection

#	Cut	Applied to	Purpose
17	`ASUM(PT) > 5000 MeV`	b daughters	Fast soft background rejection
18	`TRCHI2DOF < 2.5 & PT > 1700 MeV & P > 10000 MeV & MIPDV(PV) > 0.1 mm`	at least one b daughter	Require leading track
19	`TRCHI2DOF < 2.5 & PT > 500 MeV & P > 5000 MeV`	bachelor track	Loose kinematic selection
20	`(5200 < AM < 6000) MeV`	Λ_b^0 combination	Fast invariant mass cut
21	`(4750 < AM < 7000) MeV`	$B^-/\bar{B}_s^0$ comb.	Fast invariant mass cut
22	`VCHI2/VDOF < 10`	X_b	b daughters emerge from same space point
23	`BPVIPCHI2 < 25`	$B^-/\bar{B}_s^0$	Candidate produced in PV
24	`BPVDIRA > 0.999`	$B^-/\bar{B}_s^0$	Candidate points back to best PV
25	`BPVLTIME > 0.2 ps`	X_b	Reject combinatorial background

Table 5.3: Stripping selection on the beauty hadron and bachelor pion candidates.

The beauty hadron candidate is formed from a charm hadron candidate of the previous step and a pion candidate: the bachelor track. It is selected with loose kinematic cuts only (`TRCHI2DOF < 2.5 & PT > 500 MeV & P > 5000 MeV`, #19). An invariant mass cut before the b vertex fit is carried out to save CPU time (`(5200 < AM < 6000) MeV`, #20 for Λ_b^0 candidates; `(4750 < AM < 7000) MeV`, #21 for $B^-/\bar{B}_s^0$ candidates).

The b vertex fit is done with a loose vertex quality requirement (`VCHI2/VDOF < 10`, #22). The beauty hadron candidate is further required to be consistent with emerging from (`BPVIPCHI2 < 25`, #23) and pointing back (`BPVDIRA > 0.999`, #24) to the PV. Note that these cuts are not applied to the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ candidates to be able to use them as intermediate particles from other weak decays, *e.g.* $\Xi_b^- \to \Lambda_b^0 \pi^-$. Finally, the decay time of the b hadron candidate is required to be larger than 0.2 ps (`BPVLTIME > 0.2 ps`, #25). The decay time is computed by a dedicated decay time fitter [C25]

5.1.3 Offline selection of $X_b \to X_c\pi^-$ decays

An offline selection to reject combinatorial, Cabibbo suppressed, misidentification and prompt backgrounds is carried out. It is required to not apply too restrictive cuts during offline selection, in particular on the charm subsystem. The reasons for this strategy are the following: The BDT should do the selection work – any cut applied before training will have to be made on the sample where the BDTs are going to be used. A similar argument is to keep as much signal as possible, so that later analyses can profit from high signal efficiency. The last argument is practical: each cut introduces a new selection efficiency that needs to be understood.

The strategy particularly requires cautious studies of misidentification backgrounds, which were done in a generic way described in a paragraph below. These backgrounds mainly come from single misidentified[1] particles in the decay chain. If the misidentified particles come from another decay chain and form narrow resonances, they can be vetoed. Such vetos can however be inefficient, or shape the kinematic phase space by introducing kinks and edges. This means that vetos, especially on the charm daughter system, should be avoided if possible. On the other side they have to be taken into account when defining signal and background samples, *i.e.* in the fitting procedure detailed later.

From a different point of view it is plausible to argue that the BDT itself should be able to classify misidentified particles as background. This holds true for the charm system, as the BDT only classifies displaced X_cs. If the charm system is not affected, a veto is applied if a resonance due to misidentification persists after applying the nominal selection cuts and an additional signal invariant mass cut.

The offline selection is carried out with the above guidelines. The pool of cuts for all three calibration channels can be summarised in a compact table 5.4 and the effect of all cuts is shown in Fig. 5.2. The columns of the table are understood as follows:

π^- PID Cabbibo suppressed backgrounds are rejected by a PID selection of the bachelor π^- candidate, using a combination of pion and kaon PID variables (`ProbNNpi(1-ProbNNk) > 0.05`). Note that the Cabibbo suppressed b decays could be kept in principle and also trained on, but they complicate the fitting. With more rigorous descriptions of the backgrounds, it might be feasible to include them in future versions of the package.

ψ veto Vetos on J/ψ (and $\psi(2S)$ in $B^- \to D^0\pi^-$) from double misidentification of pion candidates are required in all channels. The J/ψ peak is seen in the combination of the bachelor pion and an oppositely charged pion from the charm system. The vetos are very efficient (*cf. e.g.* Fig. 5.10(b)), due to excellent μ–π separation with PID variables. Hence, only a selection on the bachelor pion candidate within J/ψ and $\psi(2S)$ mass windows is needed (`ProbNNpi*(1-ProbNNmu) > 0.2 OR ProbNNmu < 0`).

c veto In the $\bar{B}^0_s \to D^+_s\pi^-$ calibration channel, misidentification backgrounds from $\Lambda^+_c \to \{K^+ \to p\}K^-\pi^+$ and $D^+ \to \{K^+ \to \pi^+\}K^-\pi^+$ are vetoed. The notation $\{K^+ \to p\}$ implies that the K^+ candidate is in fact a proton. The vetos had to be applied, since the $\bar{B}^0_s \to D^+_s\pi^-$ invariant mass spectrum is swamped by $\Lambda^0_b \to \Lambda^+_c\pi^-$ and $\bar{B}^0 \to D^+\pi^-$ decays, which make fitting infeasible; in particular with regard to the determination of efficiencies. The veto again made use of PID variables within a 25 MeV mass window around the misidentified Λ^+_c and D^+ peaks. In addition, signal efficiency is retained by requiring a tight cut on the invariant mass of the K^+K^- subsystem of the D^+_s candidate (`ProbNNk*(1-ProbNNp(i)) > 0.05 OR abs(mKK-1019.416) < 5 MeV`, where `ProbNNp` vetoes Λ^+_c and `ProbNNpi` vetoes D^+)

[1] Given the very loose PID selection in the stripping, the term *misassigned* would be more appropriate. However, *misidentified* is kept to avoid confusion.

X_b–PV IP χ^2 The cut on the IP χ^2 of X_b w.r.t. the best PV is further reduced compared to the stripping selection (#23). The cut reduces prompt backgrounds in the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ and $\bar{B}_s^0 \to D_s^+ \pi^-$ calibration channels (`BPVIPCHI2 < 12`). It is not surprising that the cut removes much combinatorial background in the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ channel, since no IP χ^2 cut has been applied in the stripping.

X_b–PV DIRA To further reduce combinatorial background in the $\bar{B}_s^0 \to D_s^+ \pi^-$ decay, it was decided to tighten the DIRA cut (`BPVDIRA > 0.9999`, *cf.* #24).

Channel/Cut	π^- PID	ψ veto	c veto	X_b–PV IP χ^2	X_b–PV DIRA
$\Lambda_b^0 \to \Lambda_c^+ \pi^-$	✓	✓	✗	✓	✗
$B^- \to \bar{D}^0 \pi^-$	✓	✓	✗	✗	✗
$\bar{B}_s^0 \to D_s^+ \pi^-$	✓	✓	✓	✓	✓

Table 5.4: Offline selection of *D-from-B* BDT calibration channels. The cut values are listed in the text or the code [C18]

Generic method to spot misidentification backgrounds

A method to search for misidentification backgrounds in multi-body decays is described. It makes use of the fact, that swapping the mass hypotheses of a single particle in the decay chain can be fully described by the invariant mass of the combined system with original mass hypothesis M and a single particle momentum asymmetry β [I8]. Resonances consisting of a misidentified particle will emerge as curved band in a 2D plot of M and β. Similar plots are known from two-body systems. These are Armenteros-Podolanski plots [204], where the longitudinal momentum asymmetry is plotted against the transverse momentum of either decay product; and a plot of the signed momentum imbalance [205] $\alpha = q_1(1 - p_1/p_2)$ against the invariant mass of the system, where p_1 (p_2) is the lower (higher) of the particle momenta and q the respective charge. Examples of both plots are shown in Fig. 5.3.

The generalised single particle momentum asymmetries which allow to produce similar plots for multi-body decays can be written as

$$\beta_{\boldsymbol{i}} = \frac{-p_{i_1} + \sum_{j=2}^{\dim(\boldsymbol{i})} p_{i_j}}{\sum_{l=1}^{\dim(\boldsymbol{i})} p_{i_l}} \,, \tag{5.1}$$

with index vector $\boldsymbol{i} = (i_1, ..., i_{\dim(i)})^T$. The computation and plotting have been released in a custom-built easy to use tool [C10] which defines and evaluates all possible invariant mass and single particle momentum asymmetry-combinations $(M_{\boldsymbol{I}}, \beta_{\boldsymbol{i}})$ in an n-body final state. The index vectors $\boldsymbol{I}$ and $\boldsymbol{i}$ are 2 to n-dimensional and always contain the same set of (respectively different) indices. The difference between $\boldsymbol{I}$ and $\boldsymbol{i}$ comes from the fact, that the first index in $\boldsymbol{i}$ labels the misidentified particle, *i.e.* permutations in the indices play a role for β, but not for M.
To give an example, consider the three-body combination $\boldsymbol{I} = (2,3,6)^T$ of a six-body final state. It has M-β combinations (abbreviated for readability): $(M_{236}, \beta_{236}), (M_{236}, \beta_{362}), (M_{236}, \beta_{623})$. Applying this example to the signal channel $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$, with the final state $\overset{1}{p}\overset{2}{K^-}\overset{3}{\pi^+}\overset{4}{K^+}\overset{5}{\pi^-}\overset{6}{K^-}$ the combination 236 would label the Λ_c^+ daughters K^- and π^+, and the bachelor K^-. That example illustrates the numbering scheme, but as no structure is seen in this particular case, the combinatorially dull case (M_{123}, β_{123}) of the decay $\bar{B}_s^0 \to D_s^+ \pi^-$ is given as showcase example in Fig. 5.4. The plot clearly illustrates that the D_s^+ signal is swamped by D^+ and Λ_c^+ as stated before. Note that the plot shows data directly after the stripping selection.

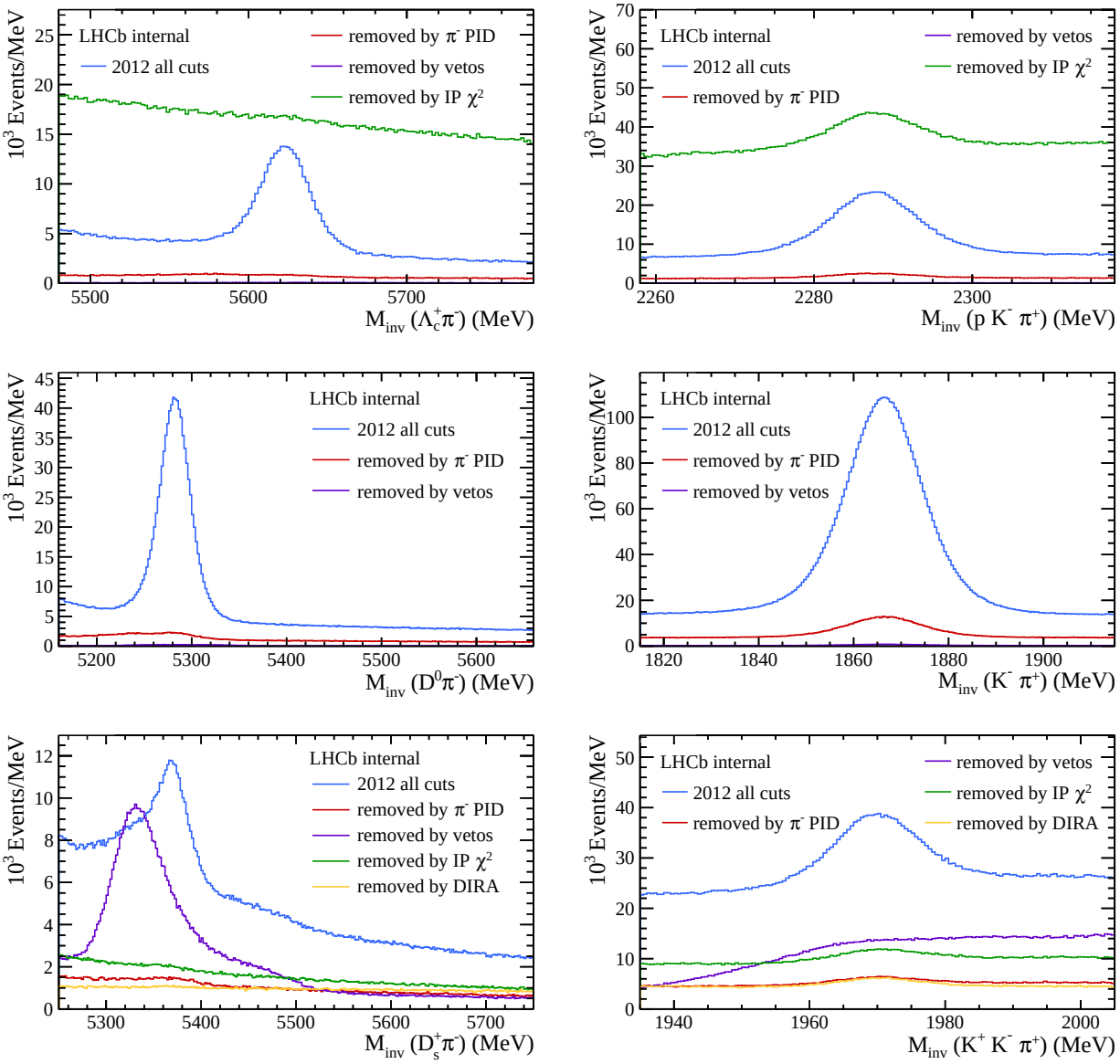
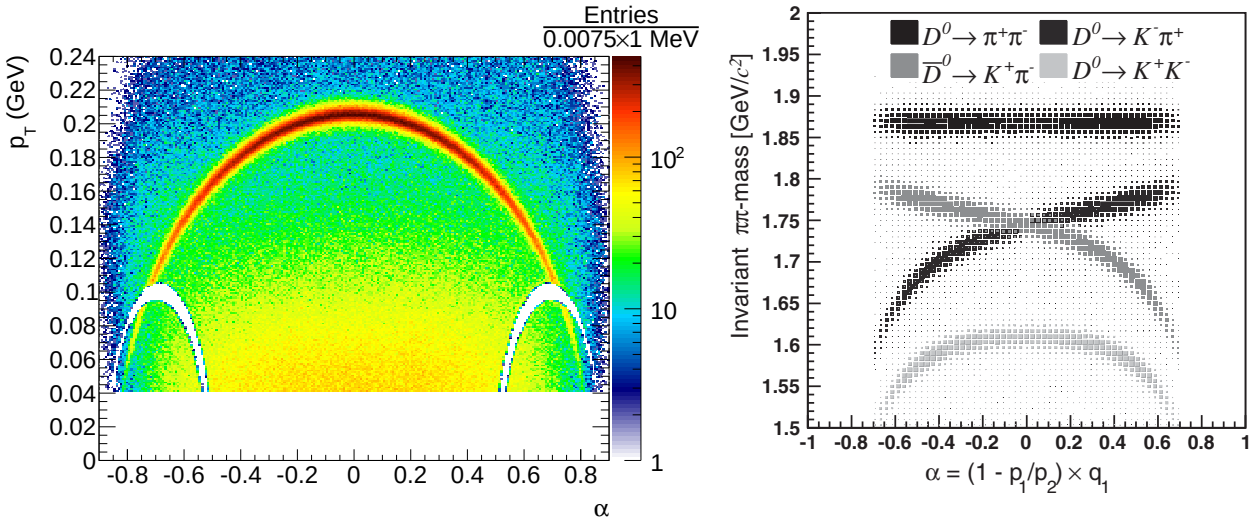


Figure 5.2: Offline selection of *D-from-B* BDT calibration channels for the 2012 data-taking period. The plots show the beauty candidate invariant mass systems on the left, the charm candidates on the right. Events passing all cuts are plotted in blue. Events which have been removed by a specific cut, but pass all other selection criteria are plotted in various colours as indicated by the plot-labels.



(a) (b)

Figure 5.3: (a) Armenteros-Podolanski plot from two charged tracks. The big arc is due to $K^0_S \to \pi^+\pi^-$ decays; The small arcs due to $\Lambda \to p\pi^-$ and $\overline{\Lambda} \to \overline{p}\pi^+$ decays have been vetoed. Here, $\alpha = \frac{p_L^+ - p_L^-}{p_L^+ + p_L^-}$, and a p_T cut has been applied. (b) Mass of $D^0 \to h^+h'^-$ candidates, with pion assignment for both hadrons h, as a function of the signed momentum imbalance α. Plots taken from Refs. [206, 207] respectively.

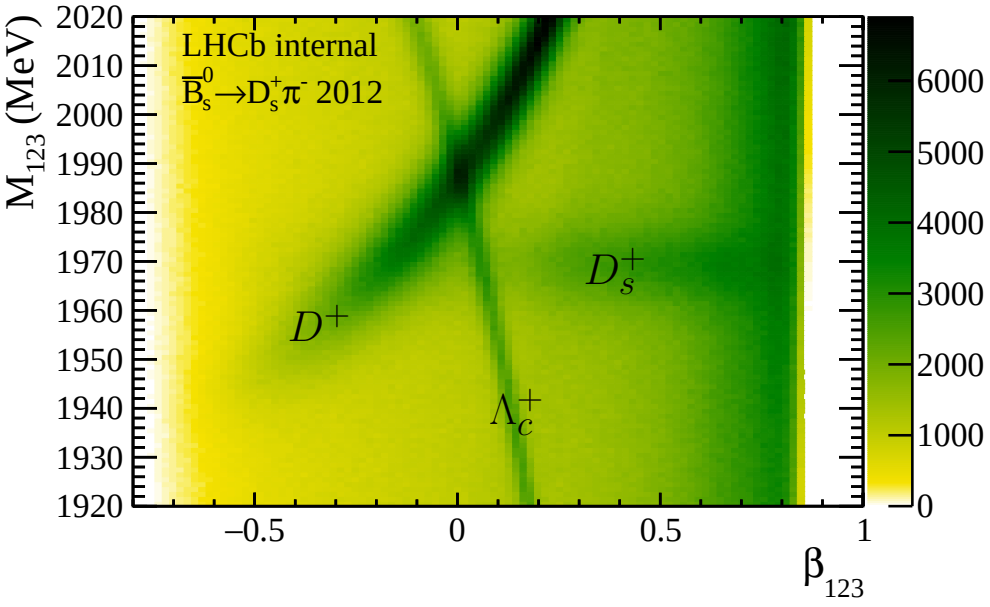


Figure 5.4: Mass *vs.* momentum asymmetry of the 3-body charm system, where the K^+ candidate is the potentially misidentified track. The broader $D^+ \to \pi^+K^-\pi^+$ band, extending from negative asymmetries at low mass to positive asymmetries at high mass, and the narrower $\Lambda^+_c \to pK^-\pi^+$ band clearly dominate the faint horizontal $D^+_s \to K^+K^-\pi^+$ band at 1970 MeV.

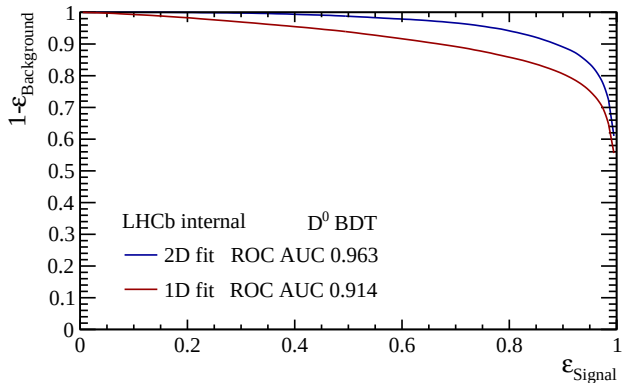


Figure 5.5: Background rejection as a function of signal efficiency obtained on a test sample with a BDT which has been trained using sWeights from 1D and 2D fits.

Sample	signal		background	
	2011	2012	2011	2012
$\Lambda_b^0 \to \Lambda_c^+ \pi^-$	87294	208415	62784	161456
$B^- \to D^0 \pi^-$	308316	749964	134855	374451
$\bar{B}_s^0 \to D_s^+ \pi^-$	41023	102435	118294	326226

Table 5.5: Yields per category of TMVA input samples. Given for orientation.

5.1.4 Strategy to define BDT training samples

Both signal and background samples for BDT training are defined from the exact same events, but are weighted differently, using sWeights (*cf.* Sec. 4.2.4 and Ref. [196]). The strategy to obtain the weights is discussed here, the individual fits are carried out later in Secs. 5.3– 5.5.

All three calibration channels are clean enough to define a clear beauty hadron signal with almost fully efficient and well understood PID information used on the charm daughters (cuts #11 – #13 in the stripping). The corresponding invariant mass distributions of beauty and charm candidates after offline selection are used as discriminating variables to obtain sWeights. This choice differs from previous versions of the *D-from-B* BDTs, which used only the beauty candidate invariant mass as discriminating variable. That this choice leads to superior performance of the BDTs is shown by comparing sWeights from one- *vs.* two-dimensional fits in the $\bar{D}^0$ BDT training in Fig. 5.5. The performance is measured by the receiver operating characteristic (ROC) curve, which relates signal efficiency to background rejection. The plot is interpreted as follows: In a 1D fit, there is a invariant mass cut selecting the X_c, and that dimension is integrated out. In the 2D case however, events left and right of the X_c peak get smaller weight to account for the fact that they are more likely to be background.

The further strategy is to fit in a wide invariant mass range of the X_b first to define shape parameters of the heuristically chosen functions. Which functions are chosen is detailed later in the individual sections of the BDTs. Subsequently a narrow range about the X_b signal with fixed shapes is fitted to define the sWeights. This staged procedure facilitates to disentangle various background contributions and to eventually get a smaller spread of weights, stabilising the BDT training.

The signal sWeights are extracted from the $X_b \to X_c \pi^-$ signals only, while the background sWeights are computed from the sum of combinatorial and misidentified backgrounds, meaning that the small portions of partially reconstructed and Cabbibo suppressed backgrounds are omitted from the training. In fact, changing the definitions for signal and background could lead to further improvements and will be studied in the future. One of the changes was mentioned already: that is the addition of the Cabibbo suppressed modes to the signal. Another change concerns the background, where the combinatorial background from real charm hadrons is added implicitly. These could however emerge from other b decays and should thus not be treated as background.

For orientation, the fitted yields are summarised here in Tab. 5.5. The sum of sWeights used in the training is slightly lower than that, since outliers and long tails in input parameter-space with very low statistics are removed to stabilise the BDT training.

5.2 BDT training

The BDTs are trained with the toolkit for multivariate analysis with ROOT (TMVA) [179]. Here, input and hyperparameter settings of the BDT training are specified. Both, parameters and settings are used in the same way in all three BDTs.

5.2.1 BDT input variables

Input variables to the BDTs are summarised in Tab. 5.6. Some of these variables have already been used in the selection, others need further explanation given in the following. Selected input variables, which are among the three variables of highest importance to the Λ_c^+, D^0 and D_s^+ BDTs are shown in Fig. 5.6.

BDT input variables from the charm candidate concern its kinematics and topology. Basic kinematic, topologic, raw PID and track reconstruction information is then provided by the charm daughter candidates. It has been found in the first report on the D-from-B BDTs, Ref. [I6], that the BDT performance strongly profits from using some of the input variables to the `ProbNN` classifier directly, rather than the `ProbNN` response itself.

Variable	Particle
log(ENDVERTEX_VCHI2)	X_c
log(FDCHI2_OWNPV)	X_c
log(IPCHI2_OWNPV)	X_c, h
log(P)	X_c
log(PT)	X_c, h
beta	X_c
atan(RichDLL{e,mu,k,p,bt})	h
RichAbove{Ka,Pr}Thres	h
UsedRich{1,2}Gas	h

Variable	Particle
UsedRichAerogel	h
isMuon	h
atan(MuonLL{mu,bg})	h
MuonNShared	h
VeloCharge	h
log(TRACK_GhostProb)	h
log(TRACK_VeloCHI2NDOF)	h
log(TRACK_TCHI2NDOF)	h
log(TRACK_MatchCHI2)	h

Table 5.6: Input variables to the D-from-B BDTs. There are 6 kinematic variables for the charmed hadron X_c and 21 variables for each daughter track $h = p, K, \pi$.

No variables involving the beauty mother are put into the BDT training to be able to use the classifiers in kinematically different b decays. The first variables listed as input are therefore kinematic variables of the charm candidate. The variables `ENDVERTEX_VCHI2` and `IPCHI2_OWNPV` have been introduced in the selection using a different notation. That is, because the stripping and the nTuple production interface different algorithms in which the variable names differ. Thus, `ENDVERTEX_VCHI2` is the X_c decay vertex χ^2 similar to stripping cut #14, and `IPCHI2_OWNPV` is the X_c IP χ^2 w.r.t. its own best PV, similar to #15. The flight distance χ^2 (`FDCHI2_OWNPV`) is the geometrical distance between primary and X_c decay vertex in units of $\chi^2 = (\boldsymbol{V}_{\mathrm{PV}} - \boldsymbol{V}_{X_c})^T (\boldsymbol{\Sigma}_{\mathrm{PV}} + \boldsymbol{\Sigma}_{X_c})^{-1} (\boldsymbol{V}_{\mathrm{PV}} - \boldsymbol{V}_{X_c})$, with vertices $\boldsymbol{V}$ and their covariances $\boldsymbol{\Sigma}$ [C20].

Further X_c input variables are the (transverse) momentum and the momentum asymmetry β, Eq. (5.1), of the X_c daughters, with the p, K^-, K^+ candidates being the potentially misidentified particle for the Λ_c^+, D^0 and D_s^+ BDTs respectively. Using this new variable instead of the momentum of each daughter increased the performance of each BDT. This is remarkable, since two, respectively three variables could be replaced by a single one without loosing information, but rather increasing the BDTs separation power.

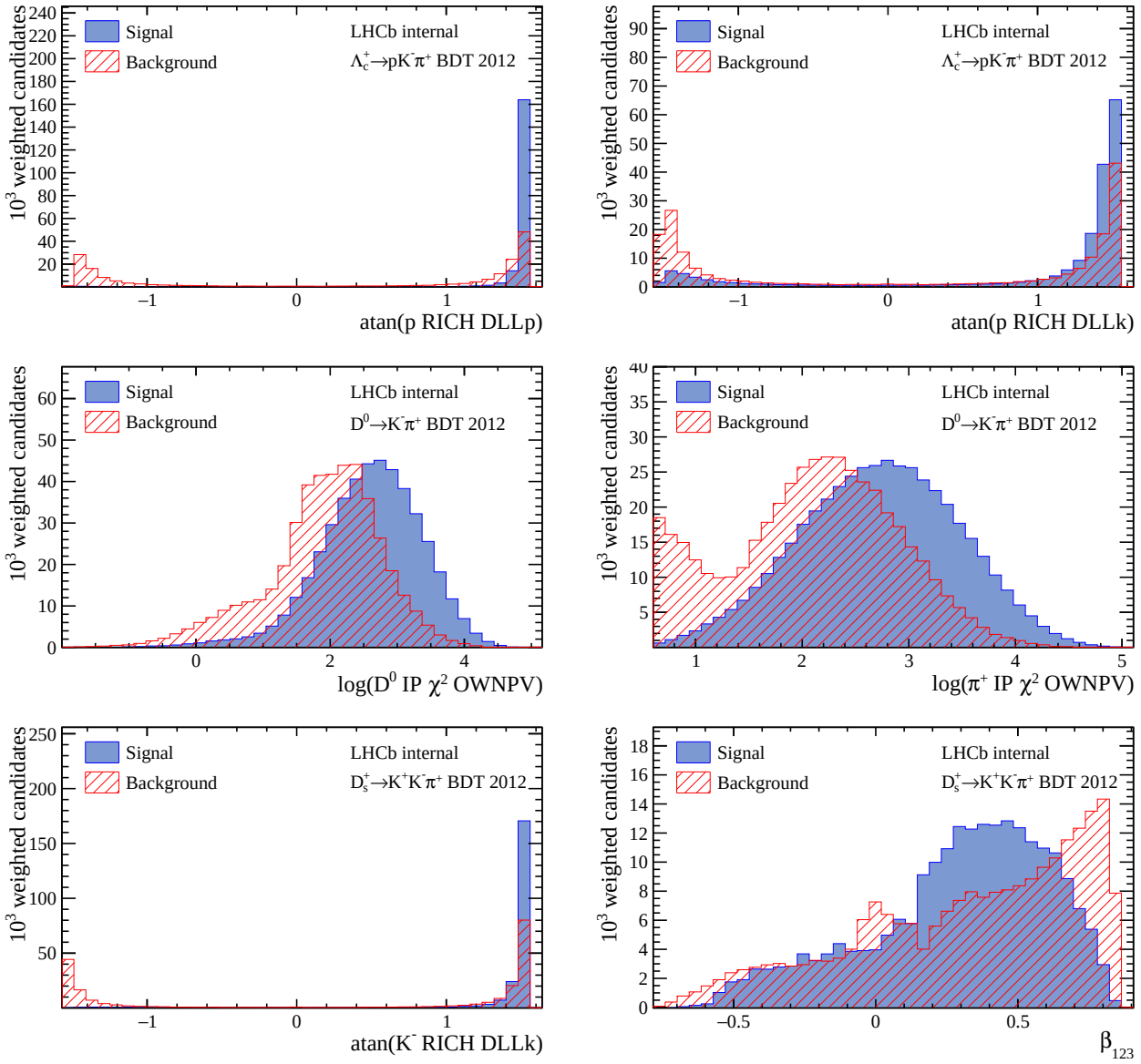


Figure 5.6: Weighted input variable distributions from the 2012 training samples; distributions from 2011 are not shown, but similar. Each of the variables is among the three variables of highest importance to the classifier. The top row shows the transformed proton and kaon DLL distributions for the proton candidate of the Λ_c^+. The IP χ^2 distributions of D^0 and π^+ candidates are plotted in the middle row. Note the cut at log(4)≈0.6 in the distribution of the π^+ candidate from the stripping selection. The last row shows the transformed kaon DLL distribution of the K^- candidate in the D_s^+, and the momentum asymmetry β_{123} of the D_s^+ system (*cf.* Fig. 5.4)

The following variables all concern single final state particles. First are the delta log-likelihood (DLL) variables from the RICH reconstruction (*cf.* Sec. 2.2.5). It is clear that *e.g.* the `RichDLLp` of the proton candidate in the Λ_c^+ BDT will be among the most important training variables. But also the `RichDLLk` of the proton candidate is of importance and is even, after the `RichDLLp`, the second most important input for the Λ_c^+ BDT. This example shows how powerful the DLL variables are, even if they do not directly relate to the hypothesised ID of the candidate. Following the recommendation of Ref. [I6], all considered hypotheses of the RICH reconstruction, e, μ, K, p,bt, are used during training. Here, bt denotes the hypothesis of the track being below threshold velocity to emit Cherenkov radiation.

Boolean variables of the RICH reconstruction follow; `RichAbove{Ka,Pr}Thres` is true if K or p hypotheses were above Cherenkov radiation threshold and the associated track traversed at least one active radiator; `UsedRich{1,2}Gas` and `UsedRichAerogel` are true if the candidate has information from RICH1, RICH2 gas radiators (C_4F_{10},CF_4) and from the RICH1 aerogel radiator, respectively.

Four input variables are related to the muon detector. These are `isMuon`, a boolean for muon identification using hit information in the muon stations [208]; `MuonLL{mu,bg}`, the likelihood of muon or non-muon (bg) hypothesis using information of the muon stations only and `MuonNShared`, which is the number of hits in the muon stations that the candidate shares with others in the event.

The last set of input variables come from track reconstruction algorithms. The `VeloCharge` variable estimates the number of particles that contributed to the energy loss measured in the VELO associated to a VELO track candidate. For a standard track, this variable is close to 1, while it is close to 2 for photons that created an electron-positron pair. The track ghost-probability has been introduced earlier as measure to reject fake tracks (#4). The remaining variables, `TRACK_VeloCHI2NDOF`, `TRACK_TCHI2NDOF` and `TRACK_MatchCHI2` are χ^2 values, which are computed by a track Kalman-fit [202, 209] of the VELO and T station segments, and the breakpoint between T and TT stations respectively.

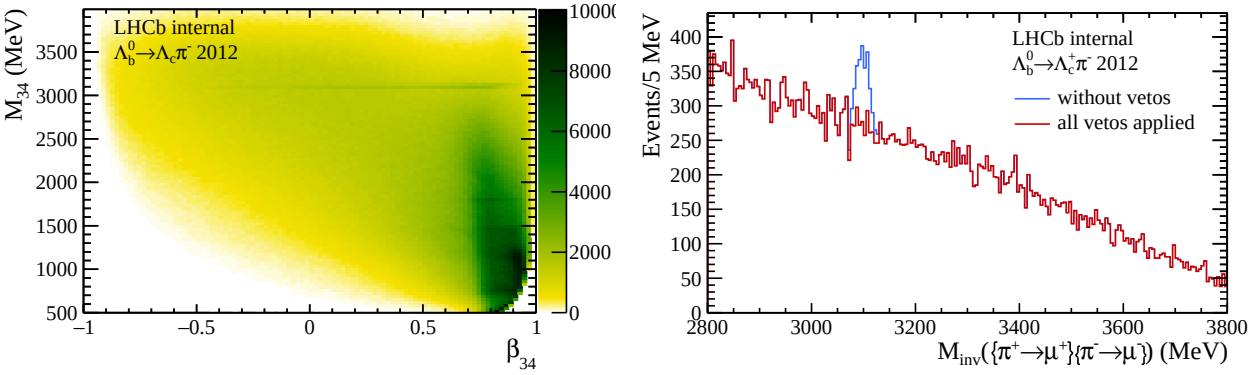
Compared to earlier version of the BDTs, it was found that an arcus-tangens or logarithmic preprocessing on some of the (strongly peaking) variables helps with the numerical stability of the training. For the same reason, outliers in input variable distributions are removed from the training.

5.2.2 Tuning of BDT hyperparameters

All classifiers are gradient boosted decision trees, that have been trained with the same hyperparameter settings, summarised here in Tab. 5.7. The settings have been chosen to perform best with large input samples as given by the calibration samples. This is especially reflected in the small learning rate, `Shrinkage` introduced in Sec. 4.2.5, and the relatively large number of trees and cuts. The number of cuts, `nCuts`, determines the scanned values for determining the splits of a regression tree according to Eq.4.13. Due to the large number of trees and the low learning rate, very shallow trees have been chosen. This is reflected in the `MaxDepth = 2` parameter, which allows the root node to split, and the tree is terminated after splitting the resulting nodes another time. The node size, or the size of the region, is not allowed to contain less than 2 % of the input events. Overtraining is avoided by using the bootstrap aggregation method mentioned in Sec. 4.2.5. Despite its name, the option to treat negative weights has been well tested and found to work as expected. As per default, the training samples are split in half for training and performance evaluation in order to avoid biases.

Property	Value
BoostType	Grad
Shrinkage	0.05
NTrees	1600
nCuts	80
MaxDepth	2
MinNodeSize	2%
UseBaggedBoost	true
NegWeightTreatment	Pray

Table 5.7: BDT training parameters deviating from the default settings given in [179]



(a) (b)

Figure 5.7: (a) Mass *vs.* momentum asymmetry of the two pion candidates. A J/ψ signal is visible as horizontal band in the 3 GeV region. This plot is produced directly after the stripping selection. Other structures seen in this plot vanish after the offline selection.
(b) Di-muon invariant mass of the pion candidates. The J/ψ signal is rejected by the applied veto. Events shown in this plot must pass all remaining selection criteria and lie in a 15(25) MeV window around the X_c (X_b) candidate.

5.3 Λ_c^+ BDT

The Λ_c^+ BDT is trained on data using $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ decays with $\Lambda_c^+ \to pK^-\pi^+$. The data is preselected as described in Sec. 5.1. Beside the bachelor PID cut, which is the same for all calibration channels, a J/ψ veto in the invariant mass of the two oppositely charged pion candidates, which have been assigned the muon mass, has been applied. The mass *vs.* momentum asymmetry plot in which this peaking background has been identified and the respective invariant mass plot are shown in Fig. 5.7.

As outlined in Sec. 5.1.4, sWeights are used to label events for BDT training. They are extracted from two-dimensional fits in the charm and beauty candidate invariant mass systems. The fits are carried out in two stages, in which shape parameters are extracted first, and are fixed in a fit to a narrower X_b mass range, leaving the X_c mass range unchanged. The BDT training is stabilised by this procedure since the sWeights are less scattered.

The fit uses pre-defined probability density functions (p.d.f.s) in RooFit to describe all shapes in an heuristic manner. This approach is opposed to extract p.d.f. shapes from simulation as it is done for *e.g.* the partially reconstructed $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ decay in the signal channel fit. Due to this, the fitted spectra should be taken with a grain of salt, especially when it comes to the yields or fit-fractions of partially reconstructed, misidentified or Cabibbo suppressed decays. For the purpose of defining sWeights for BDT training, this approach is good enough and leads to valid results (*cf.* Sec. 5.7). In the future, it is planned to model these backgrounds by centrally produced or fast simulation techniques, like RapidSim [210].

Decays that feed into the $\Lambda_c^+ \pi^-$ invariant mass spectrum have been taken from a previous LHCb analysis [211]. Those are $\bar{B}^0 \to D^+ \pi^-$, $\bar{B}_s^0 \to D_s^+ \pi^-$, $\Lambda_b^0 \to \Lambda_c^+ K^-$ and a partially reconstructed component dominated by $\Lambda_b^0 \to \Lambda_c^+ \rho^-$. The shapes of $\bar{B}^0 \to D^+ \pi^-$, $\bar{B}_s^0 \to D_s^+ \pi^-$ and $\Lambda_b^0 \to \Lambda_c^+ K^-$ decays are modelled as Gaussian p.d.f.s

$$G(M;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(M-\mu)^2}{2\sigma^2}} \tag{5.2}$$

whose shape parameters (μ, σ) have been extracted from fits to data in the $\Lambda_c^+ \pi^-$ invariant mass $M = M_{\text{inv}}(\Lambda_c^+ \pi^-)$ with a selection suiting the misidentified or Cabibbo suppressed channel.

Partially reconstructed background is modelled by a Novosibirsk function [212] with positive tail/asymmetry parameter η

$$N(M; M_{\mathrm{p}}, \eta, \sigma_{\mathrm{E}}) = \exp\left[-\frac{1}{2\sigma_0^2}\ln^2\left(1-\frac{M-M_{\mathrm{p}}}{\sigma_{\mathrm{E}}}\eta\right)-\frac{\sigma_0^2}{2}\right] . \tag{5.3}$$

The remaining parameters of the function are the peak value M_{p} and the resolution or width σ_{E}; σ_0 is given by $(2/\xi)\sinh^{-1}(\eta\xi/2)$ with $\xi = 2\sqrt{\ln 4} \approx 2.36$. The original use of the function was the Compton photon energy spectrum, and convolutes the theoretical Compton spectrum with a logarithmic Gaussian.

The $\Lambda_b^0 \to \Lambda_c^+\pi^-$ signal and the $\Lambda_c^+ \to pK^-\pi^+$ signal in the charm candidate mass projection are respectively modelled by the sum of two Gaussian p.d.f.s with common mean

$$DG(M; \mu, \sigma, f_\sigma, f_G) = f_G G(M; \mu, \sigma) + (1-f_G)G(M; \mu, f_\sigma \cdot \sigma) . \tag{5.4}$$

Here, the previous definition of the Gaussian p.d.f., Eq. (5.2), has been used. The combinatorial background in both mass representations is modelled by exponential functions

$$E(M; \tau) = e^{M\tau} .$$

These building blocks, Eq. (5.2)–(5.3), are now used to construct the two-dimensional p.d.f. It is convenient to define

$$\begin{aligned}
\mathscr{F}_{\mathrm{sig},\Lambda_b^0} &:= DG\left(M_{\mathrm{inv}}(\Lambda_c^+\pi^-); \mu_{\Lambda_b^0}, \sigma_{\Lambda_b^0}, f_{\sigma,\Lambda_b^0}, f_{G,\Lambda_b^0}\right) \\
\mathscr{F}_{\mathrm{sig},\Lambda_c^+} &:= DG\left(M_{\mathrm{inv}}(pK^-\pi^+); \mu_{\Lambda_c^+}, \sigma_{\Lambda_c^+}, f_{\sigma,\Lambda_c^+}, f_{G,\Lambda_c^+}\right) \\
\mathscr{F}_{\mathrm{bkg},\Lambda_b^0} &:= E\left(M_{\mathrm{inv}}(\Lambda_c^+\pi^-); \tau_{\Lambda_b^0}\right) \\
\mathscr{F}_{\mathrm{bkg},\Lambda_c^+} &:= E\left(M_{\mathrm{inv}}(pK^-\pi^+); \tau_{\Lambda_c^+}\right) \\
\mathscr{F}_{\mathrm{PR},\Lambda_b^0} &:= N\left(M_{\mathrm{inv}}(\Lambda_c^+\pi^-); M_{\mathrm{p,PR}}, \eta_{\mathrm{PR}}, \sigma_{\mathrm{E,PR}}\right) \\
\mathscr{F}_{\mathrm{misID},\bar{B}^0} &:= G\left(M_{\mathrm{inv}}(\Lambda_c^+\pi^-); \tilde{\mu}_{\bar{B}^0}, \tilde{\sigma}_{\bar{B}^0}\right) \\
\mathscr{F}_{\mathrm{misID},\bar{B}_s^0} &:= G\left(M_{\mathrm{inv}}(\Lambda_c^+\pi^-); \tilde{\mu}_{\bar{B}_s^0}, \tilde{\sigma}_{\bar{B}_s^0}\right) \\
\mathscr{F}_{\mathrm{CS},\Lambda_b^0} &:= G\left(M_{\mathrm{inv}}(\Lambda_c^+\pi^-); \tilde{\mu}_{\mathrm{CS}}, \tilde{\sigma}_{\mathrm{CS}}\right) ,
\end{aligned} \tag{5.5}$$

where parameters with a tilde have been fixed in the fits. With that, the full fit model in the $\Lambda_c^+\pi^-$ invariant mass is given by

$$\begin{aligned}
\mathscr{F}_{\Lambda_c^+\pi^-} = {} & N_{\Lambda_b^0\to\Lambda_c^+\pi^-}\mathscr{F}_{\mathrm{sig},\Lambda_b^0}\mathscr{F}_{\mathrm{sig},\Lambda_c^+} + N_{\Lambda_b^0\to\Lambda_c^+X}\left[f_{\Lambda_b^0\to\Lambda_c^+K^-}\mathscr{F}_{\mathrm{CS},\Lambda_b^0}\mathscr{F}_{\mathrm{sig},\Lambda_c^+} + (1-f_{\Lambda_b^0\to\Lambda_c^+K^-})\mathscr{F}_{\mathrm{PR},\Lambda_b^0}\mathscr{F}_{\mathrm{sig},\Lambda_c^+}\right] \\
& + N_{\mathrm{bkg}}\Big[f_{\bar{B}_s^0}\mathscr{F}_{\mathrm{misID},\bar{B}_s^0}\mathscr{F}_{\mathrm{bkg},\Lambda_c^+} + r_{\bar{B}_s^0/\bar{B}^0}f_{\bar{B}_s^0}\mathscr{F}_{\mathrm{misID},\bar{B}^0}\mathscr{F}_{\mathrm{bkg},\Lambda_c^+} \\
& \quad + f_{\mathrm{comb.}\ \Lambda_c^+}\mathscr{F}_{\mathrm{bkg},\Lambda_b^0}\mathscr{F}_{\mathrm{sig},\Lambda_c^+} + \left(1-f_{\bar{B}_s^0}(1+r_{\bar{B}_s^0/\bar{B}^0})-f_{\mathrm{comb.}\ \Lambda_c^+}\right)\mathscr{F}_{\mathrm{bkg},\Lambda_b^0}\mathscr{F}_{\mathrm{bkg},\Lambda_c^+}\Big] .
\end{aligned} \tag{5.6}$$

Fit projections of this model to data in the full invariant mass range are shown in Fig. 5.8. As indicated in Eq. (5.6), the Cabibbo suppressed and partially reconstructed decays define a separate category beside the signal (only the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ component) and background (everything else) categories for extracting the sWeights.

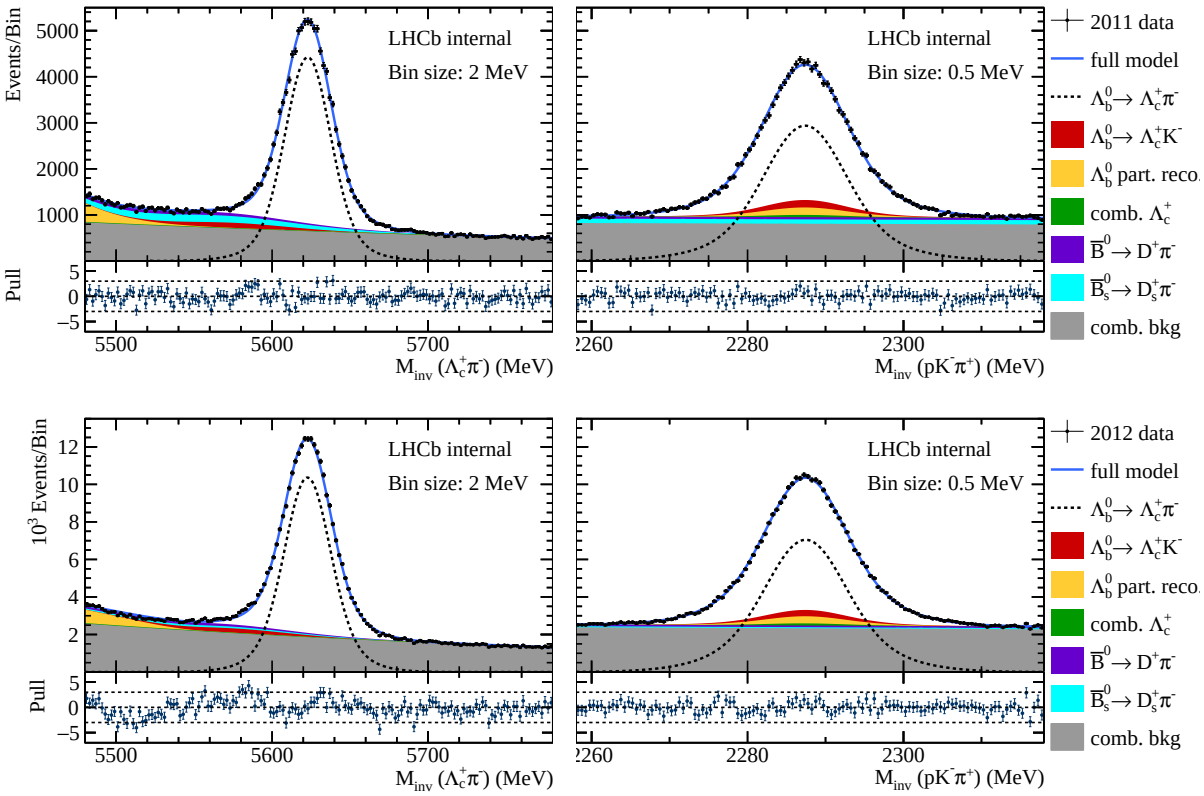


Figure 5.8: Fit projections of the two-dimensional fit to the $\Lambda_c^+\pi^-$ and $pK^-\pi^+$ invariant mass for 2011 (top) and 2012 data (bottom).

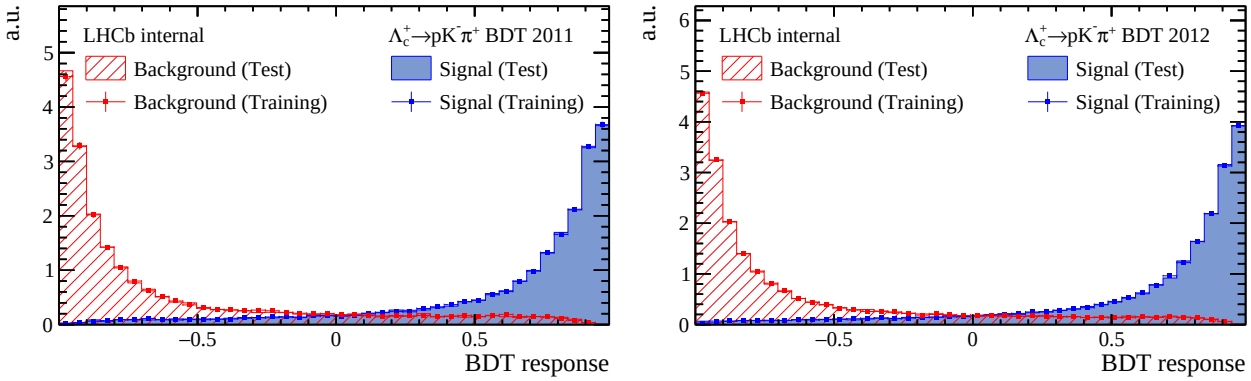
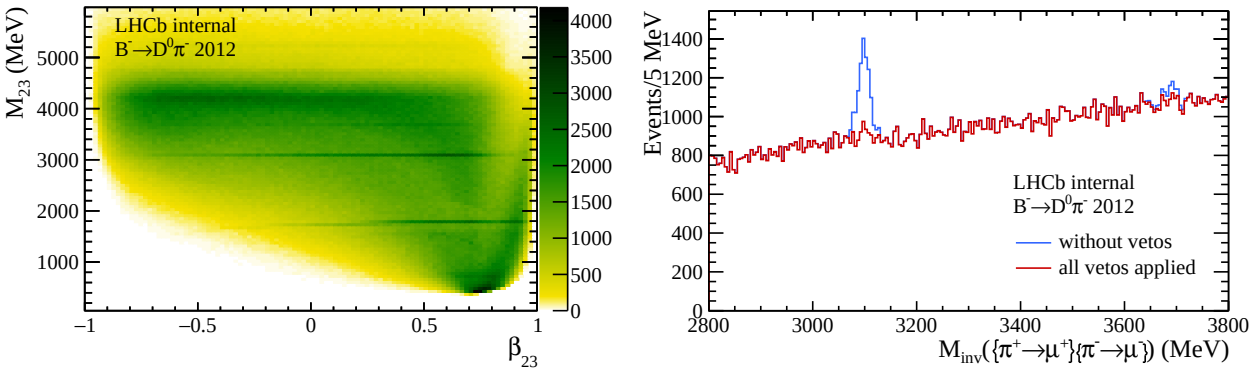


Figure 5.9: Distributions of Λ_c^+ BDT responses in 2011 (left) and 2012 (right) test and training samples.

After the fits, the sWeighted data is passed to the BDT training. There, the input variables with highest importance to the Λ_c^+ BDT are related to proton identification, namely `RichDLLp` and `RichDLLk` of the proton candidate, followed by the Λ_c^+ to PV IP χ^2 and the three-body momentum asymmetry β_{123}. The variables ranked below differ in 2011 and 2012 samples and both contain `RichDLLk` of the kaon candidate, the Λ_c^+ flight distance χ^2, the proton track ghost-probability and the pion p_T among the 10 highest ranked variables. The superimposed BDT response distributions of training and testing samples is shown in Fig. 5.9, where no signs of overtraining are seen. The BDT is thus ready to be tested on $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ data.



(a) (b)

Figure 5.10: (a) Mass *vs.* momentum asymmetry of the two pion candidates. The J/ψ and $\psi(2S)$ signals are visible as horizontal bands in the 3-4 GeV region. This plot is produced directly after the stripping selection. Other structures seen in this plot vanish after the offline preselection.
(b) Di-muon invariant mass of the pion candidates. Signals of J/ψ and $\psi(2S)$ are rejected by the applied veto. Events shown in this plot must pass all remaining selection criteria and lie in a 15(25) MeV window around the $X_c(X_b)$ candidate.

5.4 D^0 BDT

The D^0 BDT is trained on data using $B^- \to D^0\pi^-$ decays with $D^0 \to K^-\pi^+$. The data is preselected as described in Sec. 5.1. The stripping selection for this calibration channel is already very clean, such that, apart from the bachelor PID cut, only a J/ψ and $\psi(2S)$ veto in the di-muon invariant mass of the two oppositely charged pion candidates has been applied. The corresponding mass *vs.* momentum asymmetry plot in which this peaking background has been identified and the respective invariant mass plot is shown in Fig. 5.10.

The signals in both mass representations are modelled by the sum of two Gaussian p.d.f.s with common mean, Eq. (5.4). The combinatorial background in either mass is described by exponential p.d.f.s, (5.3). Partially reconstructed decays are modelled by a Novosibirsk function with floating parameters, Eq. (5.3). The Cabibbo suppressed decay $B^- \to D^0K^-$ has a asymmetric, but peaking shape (c.f. [213]) and is thus modelled by a Crystal Ball function [214]. The Crystal Ball function is a Gaussian with mean μ connected to an exponential tail at $\alpha\cdot\sigma$ of the Gaussian. The sign of α determines if this happens on the left (+) or right (−) side. The slope of the exponential is given by n. The functional form used in the fit is

$$CB(M;\mu,\sigma,\alpha,n) = \frac{1}{\sigma\tilde{N}_{\mathrm{CB}}} \begin{cases} \exp\left(-\frac{(M-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{M-\mu}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\alpha^2/2} \cdot \left(\frac{n}{|\alpha|} - |\alpha| - \frac{M-\mu}{\sigma}\right)^{-n}, & \text{for } \frac{M-\mu}{\sigma} \leq -\alpha\,, \end{cases} \tag{5.7}$$

with a normalisation factor

$$\tilde{N}_{\mathrm{CB}} = \frac{ne^{-\alpha^2/2}}{|\alpha|(n-1)} + \sqrt{\frac{\pi}{2}}\left[1+\mathrm{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right].$$

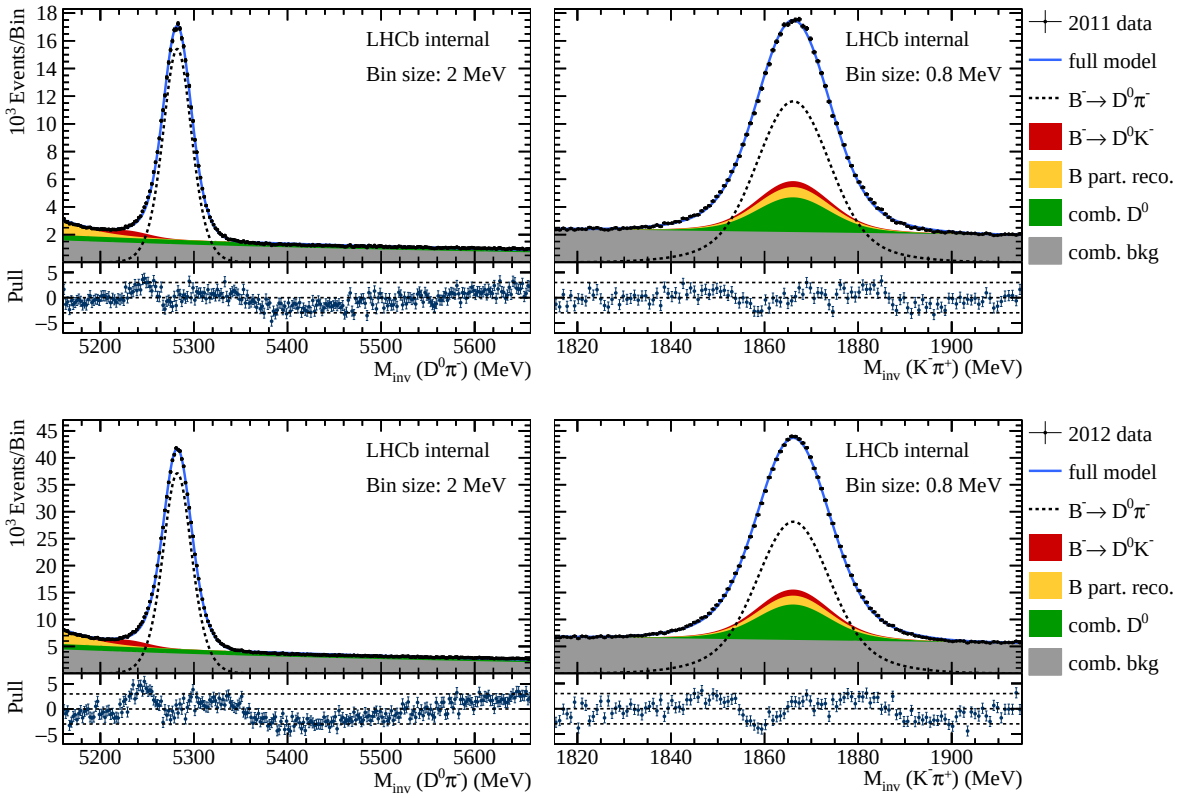


Figure 5.11: Fit projections of the two-dimensional fit to the $D^0\pi^-$ and $K^-\pi^+$ invariant mass for 2011 (top) and 2012 data (bottom).

Analogous to $\Lambda_b^0 \to \Lambda_c^+\pi^-$, Eq. 5.5, the individual components of the fit are redefined as

$$\begin{aligned}
\mathscr{F}_{\text{sig},B^-} &:= DG\left(M_{\text{inv}}(D^0\pi^-); \mu_{B^-}, \sigma_{B^-}, f_{\sigma,B^-}, f_{G,B^-}\right)\\
\mathscr{F}_{\text{sig},D^0} &:= DG\left(M_{\text{inv}}(K^-\pi^+); \mu_{D^0}, \sigma_{D^0}, f_{\sigma,D^0}, f_{G,D^0}\right)\\
\mathscr{F}_{\text{bkg},B^-} &:= E\left(M_{\text{inv}}(D^0\pi^-); \tau_{B^-}\right)\\
\mathscr{F}_{\text{bkg},D^0} &:= E\left(M_{\text{inv}}(K^-\pi^+); \tau_{D^0}\right)\\
\mathscr{F}_{\text{PR},B^-} &:= N\left(M_{\text{inv}}(D^0\pi^-); M_{\text{p,PR}}, \eta_{\text{PR}}, \sigma_{\text{E,PR}}\right)\\
\mathscr{F}_{\text{CS},B^-} &:= CB\left(M_{\text{inv}}(D^0\pi^-); \tilde{\mu}_{\text{CS}}, \tilde{\sigma}_{\text{CS}}, \tilde{\alpha}_{\text{CS}}, \tilde{n}_{\text{CS}}\right)\,,
\end{aligned}$$

where fixed parameters are again accentuated by a tilde. The fit model in the $\bar{D}^0\pi^-$ invariant mass reads

$$\begin{aligned}
\mathscr{F}_{\bar{D}^0\pi^-} = {} & N_{B^-\to D^0\pi^-}\mathscr{F}_{\text{sig},B^-}\mathscr{F}_{\text{sig},D^0} + N_{B^-\to D^0X}\left[f_{B^-\to D^0K^-}\mathscr{F}_{\text{CS},B^-}\mathscr{F}_{\text{sig},D^0} + (1-f_{B^-\to D^0K^-})\mathscr{F}_{\text{PR},B^-}\mathscr{F}_{\text{sig},D^0}\right]\\
& + N_{\text{bkg}}\left[f_{\text{comb. }D^0}\mathscr{F}_{\text{bkg},B^-}\mathscr{F}_{\text{sig},D^0} + (1-f_{\text{comb. }D^0})\mathscr{F}_{\text{bkg},B^-}\mathscr{F}_{\text{bkg},D^0}\right]\,. \qquad (5.8)
\end{aligned}$$

Although the model is improvable, as already discussed for the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ fit, it provides robust results for defining sWeights and evaluating the efficiencies. Fit projections of this model to data in the full invariant mass range are shown in Fig. 5.11. As indicated in Eq. (5.8), the Cabibbo suppressed and partially reconstructed decays define a separate category beside the signal and background categories for extracting the sWeights.

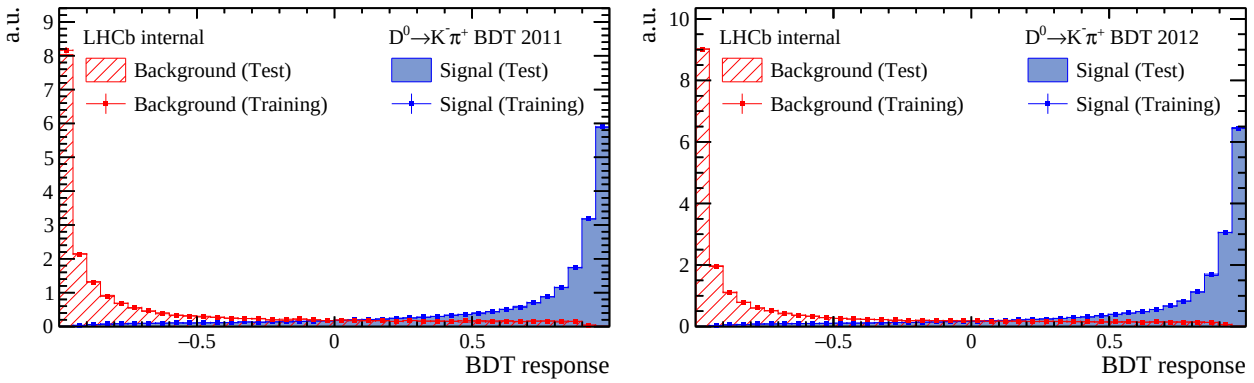


Figure 5.12: Distributions of D^0 BDT responses in 2011 (left) and 2012 (right) test and training samples.

After the fits, the sWeighted data is passed to the BDT training. There, the input variables with highest importance to the D^0 BDT are the `RichDLLk` of the kaon candidate, followed by the D^0 to PV and π^+ to PV impact parameter χ^2s, the `isMuon` decision of the pion and he `RichDLLk` of the pion candidate. The variables ranked below differ in 2011 and 2012 samples and both contain the $K^- - \pi^+$ momentum asymmetry, the kaon track ghost-probability, K^- to PV IP χ^2 and the pion p_{T} among the 10 highest ranked variables. The superimposed BDT responses of training and testing samples is shown in Fig. 5.12.

5.5 D_s^+ BDT

The D_s^+ BDT is trained on data using $\bar{B}_s^0 \to D_s^+\pi^-$ decays with $D_s^+ \to K^+K^-\pi^+$. The data is selected as described in Sec. 5.1. The $\bar{B}_s^0 \to D_s^+\pi^-$ channel is the most challenging of the calibration channels described here, because of large backgrounds from $\bar{B}^0 \to D^+\pi^-$ and $\Lambda_b^0 \to \Lambda_c^+\pi^-$ which feed into the D_s^+ and $\bar{B}_s^0$ signals. In contrast to the situation for the Λ_c^+ BDT, where the level of misidentified background is moderate, the $\bar{B}^0 \to D^+\pi^-$ and $\Lambda_b^0 \to \Lambda_c^+\pi^-$ decays make up a large fraction of events in the signal region. A veto to reject D^+ and Λ_c^+ decays is used to reduce these backgrounds. Another veto on the bachelor and charm daughter pion candidates, similar to the ψ vetos for the Λ_c^+ and D^0 BDTs, is applied as well. One mass *vs.* momentum asymmetry plot of this calibration channel has already been shown in Fig. 5.4. Further plots of the offline selection are shown in Fig. 5.13.

The signals in both mass representations are modelled by the sum of two Gaussian p.d.f.s with common mean, Eq. (5.4). The combinatorial backgrounds are modelled as double and single exponential p.d.f.s in the beauty and charm candidate masses respectively. The double exponential function is given by

$$DE(M;\tau_1,\tau_2,f_E) = f_E E(M;\tau_1) + (1-f_E)E(M;\tau_2)\,.$$

Partially reconstructed decays, like $\bar{B}_s^0 \to D_s^{*+}\pi^-$ and $\bar{B}^0 \to D^{*+}\pi^-$, are again modelled by a Novosibirsk function, Eq. (5.3). Crystal Ball functions are used to model Cabibbo suppressed and misidentified decays, Eq. (5.7). The shapes are motivated by MC simulated shapes shown in Figs. 8.3 and 8.4 of Ref. [215]. The $\bar{B}^0 \to D^+\pi^-$ and $\Lambda_b^0 \to \Lambda_c^+\pi^-$ decays are modelled by a Novosibirsk function with fixed parameters and an exponential function in the $K^+K^-\pi^+$ invariant mass system. A small CKM suppressed $B^0 \to D_s^+\pi^-$ component is modelled by a Gaussian signal with fixed parameters.

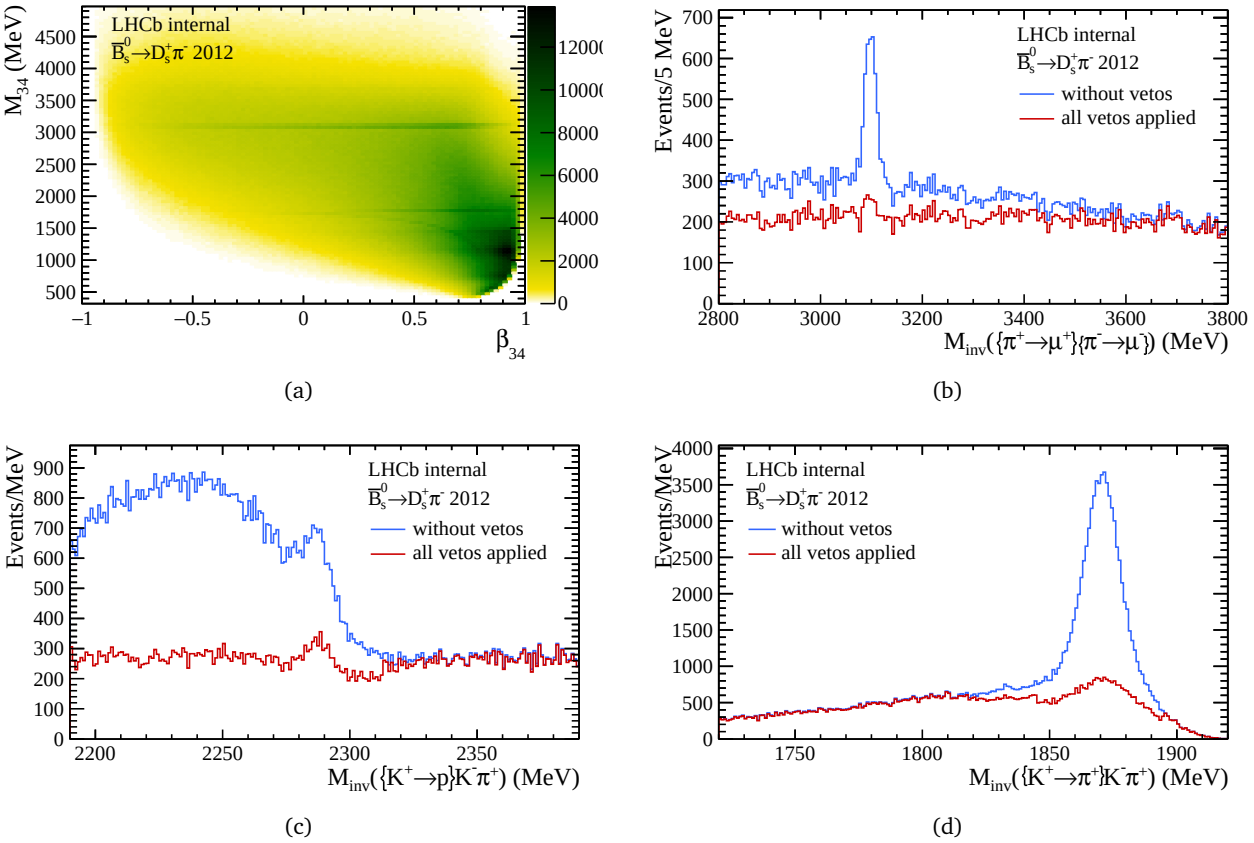


Figure 5.13: (a) Mass *vs.* momentum asymmetry of the pion candidates. The J/ψ signal is visible as horizontal band in the 3 GeV region. (b) Di-muon invariant mass of the pion candidates. The J/ψ signal is rejected by the applied veto. (c)/(d) Invariant mass of the charm system where the proton/pion mass is assigned to the K^+ candidate. Events shown in the veto plots must pass all remaining selection criteria and lie in a 15(25) MeV window around the $X_c(X_b)$ candidate.

Analogous to $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $B^- \to D^0\pi^-$, the individual components of the fit are redefined as

$$
\begin{aligned}
\mathscr{F}_{\text{sig},\bar{B}^0_s} &:= DG\left(M_{\text{inv}}(D_s^+\pi^-); \mu_{\bar{B}^0_s}, \sigma_{\bar{B}^0_s}, f_{\sigma,\bar{B}^0_s}, f_{G,\bar{B}^0_s}\right)\\
\mathscr{F}_{\text{sig},D_s^+} &:= DG\left(M_{\text{inv}}(K^+K^-\pi^+); \mu_{D_s^+}, \sigma_{D_s^+}, f_{\sigma,D_s^+}, f_{G,D_s^+}\right)\\
\mathscr{F}_{\text{bkg},\bar{B}^0_s} &:= DE\left(M_{\text{inv}}(D_s^+\pi^-); \tau_{1,D_s^+}, \tau_{2,D_s^+}, f_{E,D_s^+}\right)\\
\mathscr{F}_{\text{bkg},D_s^+} &:= E\left(M_{\text{inv}}(K^+K^-\pi^+); \tau_{D_s^+}\right)\\
\mathscr{F}_{\text{PR},\bar{B}^0_s} &:= N\left(M_{\text{inv}}(D_s^+\pi^-); M_{\text{p,PR}}, \eta_{\text{PR}}, \sigma_{\text{E,PR}}\right)\\
\mathscr{F}_{\text{CS},\bar{B}^0_s} &:= CB\left(M_{\text{inv}}(D_s^+\pi^-); \tilde{\mu}_{\text{CS}}, \tilde{\sigma}_{\text{CS}}, \tilde{\alpha}_{\text{CS}}, \tilde{n}_{\text{CS}}\right)\\
\mathscr{F}_{\text{sig},B^0} &:= G\left(M_{\text{inv}}(D_s^+\pi^-); \tilde{\mu}_{\text{sig},B^0}, \tilde{\sigma}_{\text{sig},B^0}\right)\\
\mathscr{F}_{\text{misID},\bar{B}^0_s} &:= CB\left(M_{\text{inv}}(D_s^+\pi^-); \tilde{\mu}_{\text{misID},\bar{B}^0_s}, \tilde{\sigma}_{\text{misID},\bar{B}^0_s}, \tilde{\alpha}_{\text{misID},\bar{B}^0_s}, \tilde{n}_{\text{misID},\bar{B}^0_s}\right)\\
\mathscr{F}_{\text{misID},D^+} &:= N\left(M_{\text{inv}}(K^+K^-\pi^+); \tilde{M}_{\text{p,misID},D^+}, \tilde{\eta}_{\text{misID},D^+}, \tilde{\sigma}_{\text{E,misID},D^+}\right)\\
\mathscr{F}_{\text{misID},\Lambda_b^0} &:= CB\left(M_{\text{inv}}(D_s^+\pi^-); \tilde{\mu}_{\text{CS},\Lambda_b^0}, \tilde{\sigma}_{\text{CS},\Lambda_b^0}, \tilde{\alpha}_{\text{CS},\Lambda_b^0}, \tilde{n}_{\text{CS},\Lambda_b^0}\right)\\
\mathscr{F}_{\text{misID},\Lambda_c^+} &:= E\left(M_{\text{inv}}(K^+K^-\pi^+); \tau_{\text{misID},\Lambda_c^+}\right)\,.
\end{aligned}
$$

Fixed parameters are again accentuated by a tilde. The fit model in the $D_s^+\pi^-$ invariant mass reads

$$\begin{aligned}\mathscr{F}_{\bar{D}^0\pi^-} = & N_{\bar{B}_s^0\to D_s^+\pi^-}\mathscr{F}_{\text{sig},\bar{B}_s^0}\mathscr{F}_{\text{sig},D_s^+} \\ & + N_{X_b\to X_c X}\Big[f_{\bar{B}_s^0\to D_s^+K^-}\mathscr{F}_{\text{CS},\bar{B}_s^0}\mathscr{F}_{\text{sig},D_s^+} + f_{B^0\to D_s^+\pi^-}\mathscr{F}_{\text{sig},B^0}\mathscr{F}_{\text{sig},D_s^+} \\ & \quad + \left(1 - f_{B^0\to D_s^+\pi^-} - f_{\bar{B}_s^0\to D_s^+K^-}\right)\mathscr{F}_{\text{PR},\bar{B}_s^0}\mathscr{F}_{\text{sig},D_s^+}\Big] \\ & + N_{\text{bkg}}\Big[f_{\bar{B}^0}\mathscr{F}_{\text{misID},\bar{B}^0}\mathscr{F}_{\text{misID},D^+} + r_{\Lambda_b^0}\mathscr{F}_{\text{misID},\Lambda_b^0}\mathscr{F}_{\text{misID},\Lambda_c^+} \\ & \quad + f_{\text{comb. }D_s^+}\mathscr{F}_{\text{bkg},\bar{B}_s^0}\mathscr{F}_{\text{sig},D_s^+} + \left(1 - f_{\bar{B}_s^0} - f_{\Lambda_b^0} - f_{\text{comb. }D_s^+}\right)\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{bkg},D_s^+}\Big].\end{aligned} \tag{5.9}$$

Fit projections of this model to data in the full invariant mass range are shown in Fig. 5.14. As indicated in Eq. (5.9), the CKM suppressed and partially reconstructed decays define a separate category beside the signal and background categories for extracting the sWeights.
After the fits, the sWeighted data is passed to the BDT training. There, the input variables with highest importance to the D_s^+ BDT are the `RichDLLk` of the K^+ candidate, followed by the 3 body momentum asymmetry β_{123} and the `RichDLLk` of the K^- candidate. The variables ranked below differ in 2011 and 2012 samples and both contain the D^0 to PV and π^+ to PV impact parameter χ^2s, the D_s^+ flight distance χ^2, the K^+ and π^+ track ghost-probabilities, and the K^- p_T among the 10 highest ranked variables. The superimposed BDT responses of training and testing samples is shown in Fig. 5.15.

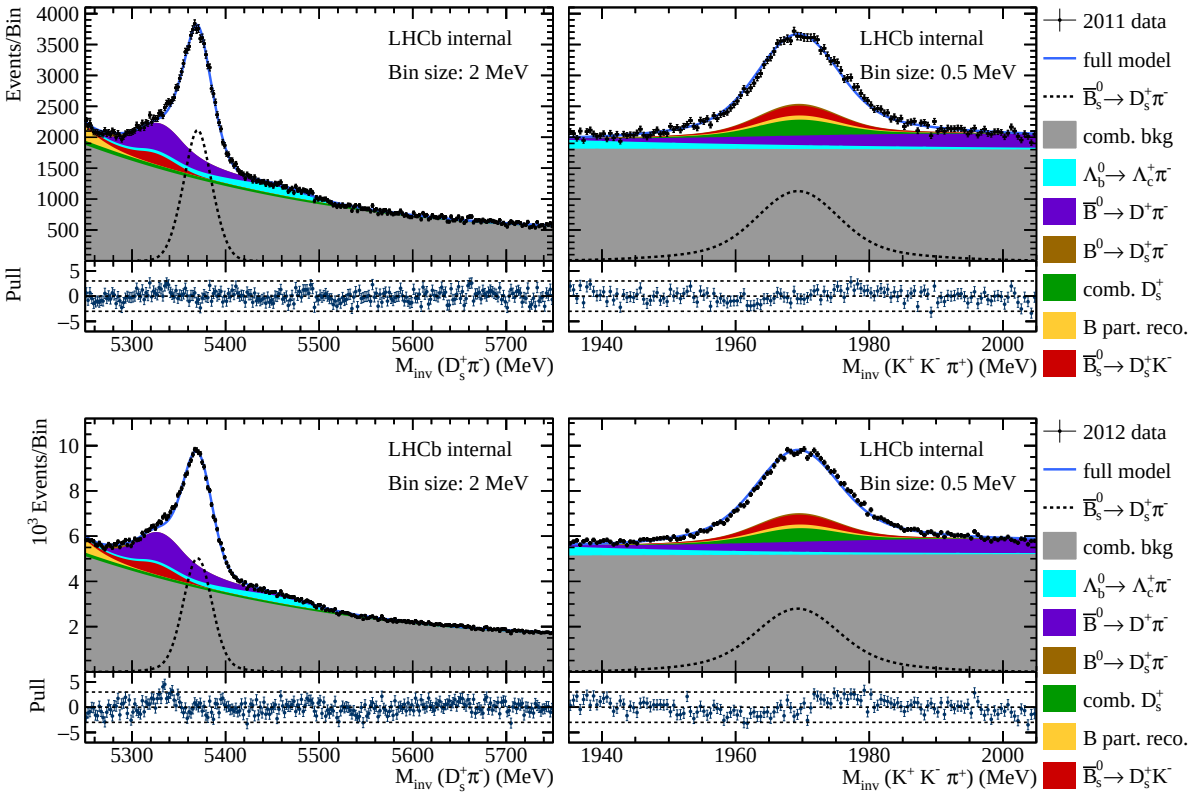


Figure 5.14: Fit projections of the two-dimensional fit to the $D_s^+\pi^-$ and $K^+K^-\pi^+$ invariant mass for 2011 (top) and 2012 data (bottom).

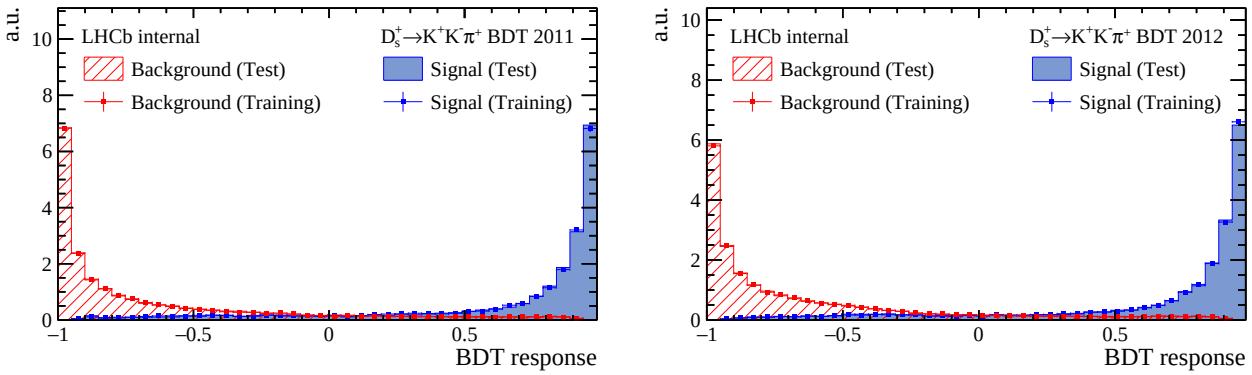


Figure 5.15: Distributions of D_s^+ BDT responses in 2011 (left) and 2012 (right) test and training samples.

5.6 BDT efficiencies

To be able to apply a certain BDT response cut and correct for it, as it is done later for the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channels, the efficiency of such a cut needs to be evaluated. Here, this is done by creating two-dimensional adaptively binned histograms, which store the efficiency in the form of a lookup table as a function of calibration variables ($\boldsymbol{\omega}$ in the formula for efficiency correction, Eq. (3.2)). The efficiency in each bin is estimated in a data-driven manner by fits to the calibration channels. Since the calibration channel fits include backgrounds, efficiency confidence intervals can not be estimated by common approaches. An effective method to estimate binomial confidence intervals in the presence of backgrounds is presented here for the first time in appendix A.

The efficiency lookup tables presented in the following are evaluated for BDT response cuts of -0.15 and -0.25 for Λ_c^+ and D^0/D_s^+ BDTs respectively. This set of cuts corresponds to the optimised $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ selection, which is detailed later in Sec. 6. The calibration variables $\boldsymbol{\omega}$ were chosen to be the logarithm of the flight-distance χ^2 from the X_c candidate to its own best primary vertex and the number of tracks in the event (`nTracks`). The former is chosen, to account for the fact that the lifetime acceptance of the X_b hadron can differ for different decay modes, and is properly accounted for in X_b decays to double open charm [I9]. The `nTracks` variable is chosen since the occupancy in the RICH detectors has an impact on the number of detected photoelectrons and hence the reconstruction quality in the RICH [118]. A kinematic variable like the X_b hadron (transverse) momentum is not needed, since the kinematics of the calibration samples are similar to those of the decays under study [176]. This is due to the fact that the lab-frame kinematics are mostly due to the boost from the b-hadron and not the b-hadron rest frame kinematics of the c-hadron [I9]. The choice of variables is validated in Sec. 5.7.

An adaptive binning in these variables is chosen. The algorithm has been developed by the author for studies of the track reconstruction asymmetry [I10], where a common optimal binning for positive and negative tracks was simultaneously searched for. The algorithm [C2] recursively splits a fine-binned two-dimensional histogram in its median along one axis, alternating the axis in every step. The parameter that determines the "resolution" of the binning is the minimal number of events per bin. The minimal number of events in a signal region around the nominal X_b- and X_c masses passing the BDT cut chosen here is 1920. This rather odd number ensures that all six combinations of data-taking period and calibration channel have a fluctuation-free binning. Fluctuations in the binning scheme can occur if certain bins contain just a little less than two times the minimal number of events whereas other bins contain more than that and will therefore be split again. These fluctuations can easily be seen when plotting the resulting binning (*cf.* Fig. 5.16).

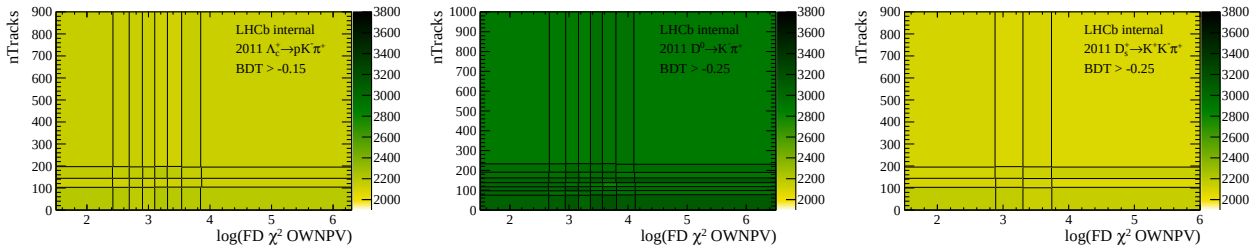


Figure 5.16: Binning in log(FD χ^2 OWNPV) and `nTracks`. Each bin has roughly the same number of signal events.

Once a suitable binning is found, the efficiency is measured by two-dimensional simultaneous unbinned extended maximum likelihood (EML) fits to the orthogonal samples of candidates which have been accepted ("pass") or rejected ("fail") by the BDT cut in every log(FD χ^2 OWNPV)–`nTracks` bin. This rather costly procedure is chosen for two reasons: First, the efficiencies are defined as lookup tables, so that the calibration samples do not have to be reprocessed by the user. This is opposed to re-weighting strategies, where the user would re-weight variables in the calibration sample, so that they match the distribution of the signal sample they should be applied to. In the future, the lookup tables are foreseen to be fitted to provide a continuous description of the efficiencies.

The second reason concerns the fact that a fit in every of the log(FD χ^2 OWNPV)–`nTracks` bins is needed. This approach is different from the one taken *e.g.* in PIDCalib [160], where calibration variable-integrated fits are carried out, and the efficiency in each bin is extracted from the corresponding sWeights. This can not be done in the present case, since the *shape* of background p.d.f.s changes as a function of the calibration variables, violating the requirement of p.d.f. factorisation in ${}_s\mathscr{P}lot$, *cf.* Eq. (4.9). This is demonstrated for the $\overline{B}{}^0_s \to D_s^+\pi^-$ case in Fig. 5.17.

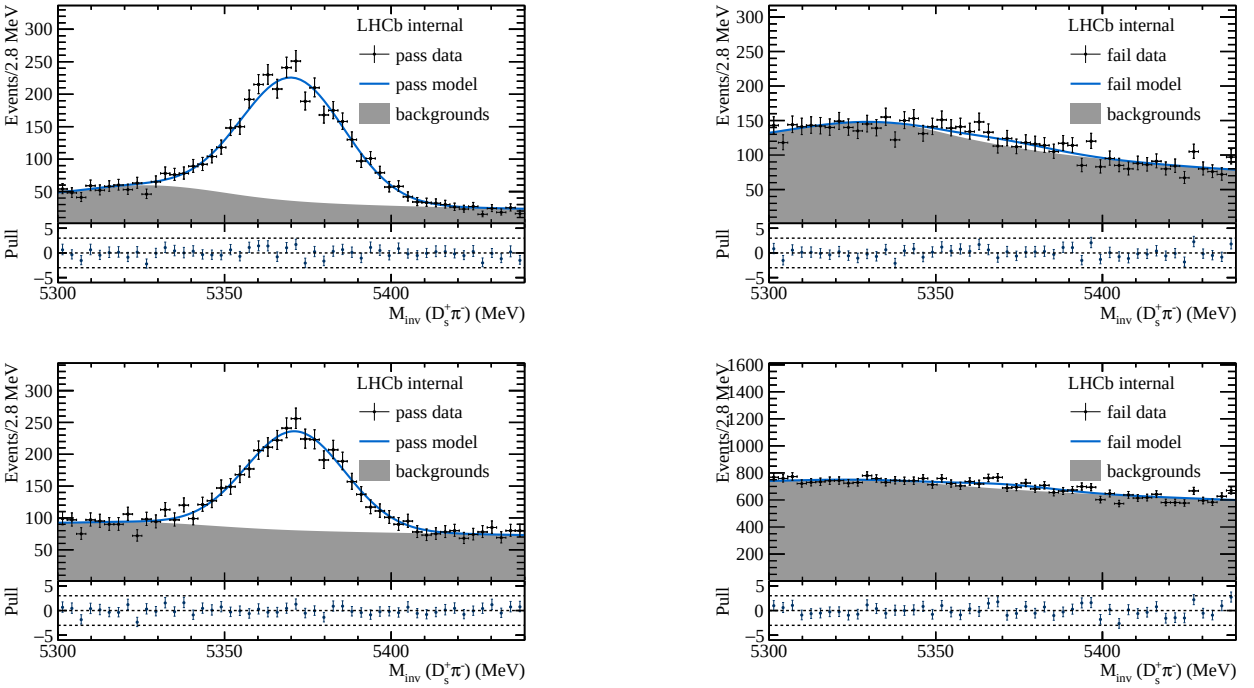


Figure 5.17: Fit projections in the $D_s^+\pi^-$ invariant mass of the 2D simultaneous unbinned EML fit to the "pass" (left) and "fail" (right) categories in two different log(FD χ^2 OWNPV)–`nTracks` bins (top and bottom).

Some adjustment to the fit model had to be made to account for the lower statistics in each bin. The fits are again carried out in two dimensions, using the narrow $X_c\pi^-$ invariant mass range described above. The sum of two Gaussian signals is replaced by a single Gaussian shape. Beside the signal and background yields, the fractions of peaking and combinatorial X_c backgrounds are split between the "pass" and "fail" categories. That is, because these components are part of the background model, but do not scale as the efficiency of the combinatorial background. Some components were omitted: the $\bar{B}^0 \to D^+\pi^-$ in the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ calibration channel, and the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $B^0 \to D_s^+\pi^-$ components in the $B_s^0 \to D_s^+\pi^-$ calibration channel. Since the Cabibbo-suppressed components fluctuated around 0 in the in the "fail" category of the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $B_s^0 \to D_s^+\pi^-$ channels and caused unstable fits, these components were set to zero.

Another detail of the fitting procedure concerns the validity of the efficiency confidence interval. Since the signals in "pass" and "fail" categories are affected by nuisances like backgrounds it is not straight-forward to calculate a valid confidence interval for this binomial process. The result of this work is only marginally affected by the size of the efficiency confidence interval. Thus, the discussion of the proposed solution is diverted to appendix A.

The resulting efficiency lookup tables are shown in Fig. 5.18, which show a clear trend of lower efficiency in the region of low log(FD χ^2 OWNPV) and large `nTracks`. The efficiency drop for low log(FD χ^2 OWNPV) is attributed to the fact that the BDTs make use of the topological separation between X_c signal and background. The signal, due to it's longer lifetime compared to most backgrounds, is therefore easier to select in the long-lived region at large log(FD χ^2 OWNPV). The efficiency drop towards a larger number of tracks is less pronounced, and affects the RICH reconstruction, which is less efficient for higher detector occupancies.

5.7 BDT efficiency validation

The BDT efficiency calibration obtained in the previous section from $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $\bar{B}_s^0 \to D_s^+\pi^-$ will be used to estimate efficiency-corrected yields of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays. The doubly open charm decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ is kinematically different from the calibration channel decays and allows to probe both Λ_c^+ and D_s^+ BDT response cuts simultaneously. This is feasible due to the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay topology, allowing to determine a signal yield even if no BDT response cuts have been applied. The validation will probe if the efficiency corrected yield is constant over a wide range of applied cuts, and hence if the calibration is valid.

The event-by-event efficiency correction as introduced in Eq. (3.2) is used. On condition that the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ signal candidates pass all prior selection and reconstruction steps, the efficiency corrected yield reads

$$N_{\text{corr}}^{\Lambda_b^0 \to \Lambda_c^+ D_s^-} = \sum_i \frac{{}^s\mathscr{P}_{\Lambda_b^0 \to \Lambda_c^+ D_s^-}\left(\left(M_{\text{inv}}(\Lambda_c^+ D_s^-), M_{\text{inv}}(pK^-\pi^+), M_{\text{inv}}(K^-K^+\pi^-)\right)_i^T\right)}{\prod_{k=\{\Lambda_c^+, D_s^+ \text{ BDTs}\}} \varepsilon_k\left(\left(\log(\text{FD}\chi^2), \texttt{nTracks}\right)_{i,k}^T\right)} \tag{5.10}$$

$$\forall \left\{\left(M_{\text{inv}}(\Lambda_c^+ D_s^-), M_{\text{inv}}(pK^-\pi^+), M_{\text{inv}}(K^-K^+\pi^-), \log(\text{FD}\chi^2), \texttt{nTracks}\right)_i^T\right\} \subset \boldsymbol{x}_i^{\Lambda_b^0 \to \Lambda_c^+ D_s^-}\left(\cap_k c_k^{\Lambda_b^0 \to \Lambda_c^+ D_s^-}\right).$$

As already stated in Sec. 3, it is the analysts obligation to ensure that the efficiency corrected yield does not depend on the applied selection $c_k^{\Lambda_b^0 \to \Lambda_c^+ D_s^-}$. The selection enters implicitly in the sWeights and the measured calibration variables via the events i. It is probed here, if the efficiency corrected yields are truly independent of the selection. This is done by applying many combinations of Λ_c^+ and D_s^- BDT response cuts and correcting by efficiencies measured with the $\Lambda_b^0 \to \Lambda_c^+\pi^-$ and $B_s^0 \to D_s^-\pi^+$ calibration channels (*cf.* Sec. 5.6). For brevity this procedure is called "scan" in the following.

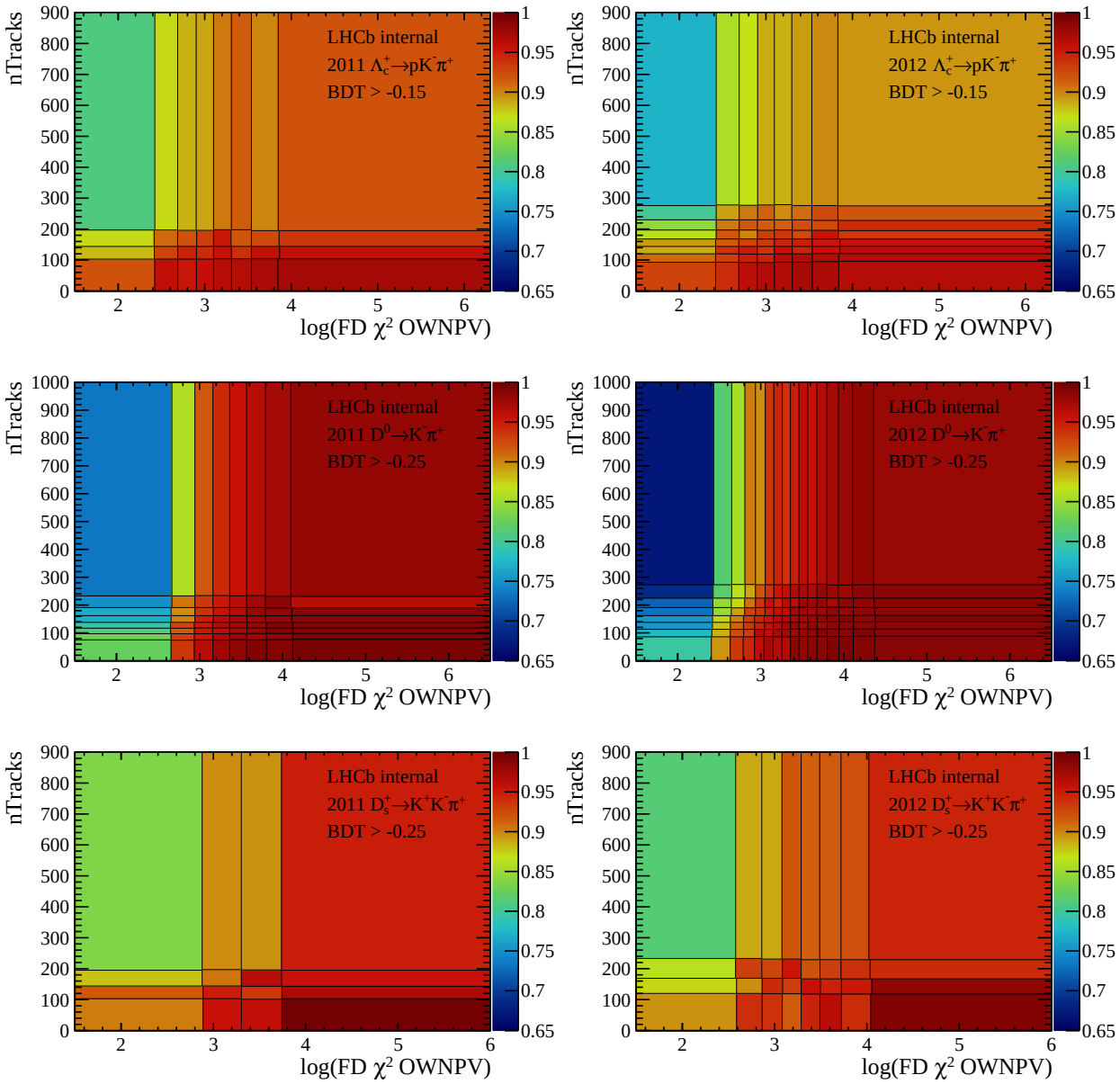


Figure 5.18: *D-from-B* BDT efficiency lookup tables for 2011 (left) and 2012 data (right). The top row shows efficiencies of Λ_c^+ BDT response cut larger than −0.15, the middle and lower rows show efficiencies of D^0 and D_s^+ response cuts larger than −0.25 respectively.

The first step is to measure the yield without applying BDT response cuts. The corresponding three-dimensional unbinned EML fits in the $\Lambda_c^+ D_s^-$, $pK^-\pi^+$ and $K^-K^+\pi^-$ invariant masses are described in detail later (Sec. 7.4). Fit projections of the $\Lambda_c^+ D_s^-$ invariant mass for 2011 and 2012 data are shown in Fig. 5.19. Due to the much larger background compared to the data with nominal Λ_c^+ and D_s^- BDT selection, it can be challenging to estimate nuisance parameters like the background shape and yield.

If nuisance parameters are furthermore (anti-)correlated with the signal yield, efficiency corrected yields might appear biased w.r.t. the reference yield. That is because the bulk of background is rejected with relatively loose cuts, leaving samples which contain the same candidates over a wide range of cuts. In these samples, the fit will measure similar nuisance parameters values.

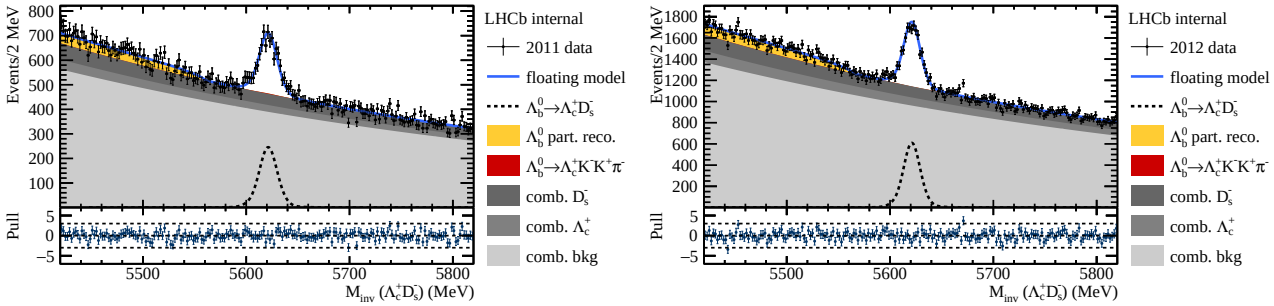


Figure 5.19: Fit projections of the 3D unbinned EML fit to 2011 (top) and 2012 (bottom) $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ test channel candidates without *D-from-B* BDT response cuts. The fit with floating signal shape is shown. No significant difference to the other systematic studies (fixed or constrained signal shape) was seen by eye.

Since these parameters are best estimated on a sample with large statistical power, the nominal set of BDT cuts, Λ_c^+ BDT > -0.15 and D_s^+ BDT > -0.25, is used to measure nuisance parameters. The scan will thus be performed by fixing all signal shape parameters to this reference-measurement, by using a Gaussian-constraint on these parameters, or by letting them float.

The results of the *D-from-B* BDT efficiency validation are shown in Figs. 5.20 and 5.21. The statistical uncertainty on the reference yield is drawn as a blue band. Since the scanned points are statistically fully correlated to the reference yield, only systematic uncertainties are shown. Those are merely visible, since most markers are larger. The systematic uncertainties come from propagating upper and lower limits of the 68.27 % confidence intervals of the lookup tables to the efficiency corrected yield[1].

It can be derived from Fig. 5.20, that the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ sample contains a larger number of false Λ_c^+ candidates, since the Λ_c^+ BDT rejects up to twice as much background as the D_s^- BDT. This is expected, since no strict proton selection is in place and protons are less abundant compared to kaons. The Λ_c^+ BDT shows a slight downward trend in the efficiency corrected yield, meaning that the efficiencies tend to be over-estimated with tighter cuts. This could be due to under-estimated partially reconstructed or misidentified backgrounds in the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ calibration channel. The effect is however too small to justify further systematic studies. Moreover, the same Λ_c^+ BDT cuts are applied in $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ selections, such that systematic effects cancel to first order. Higher order effects depend on relative differences in signal unfolded calibration variable distributions of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays.

The plots show that the proposed method of efficiency correction leads to results which are well in agreement with the reference yield. It is concluded that the *D-from-B* BDT efficiencies have been properly calibrated and that *D-from-B* BDT response cuts can be applied in the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ selection. However, this does not mean that the *D-from-B* BDTs can be blindly applied in every selection. It should be noted, that all $X_b \to X_c \pi^-$, $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channels are based on the same trigger and a similar stripping selection, based on the `Beauty2Charm` module.

[1] The corresponding formula, Eq. (9.1), is shown later in a more general context, since the effect here is marginal.

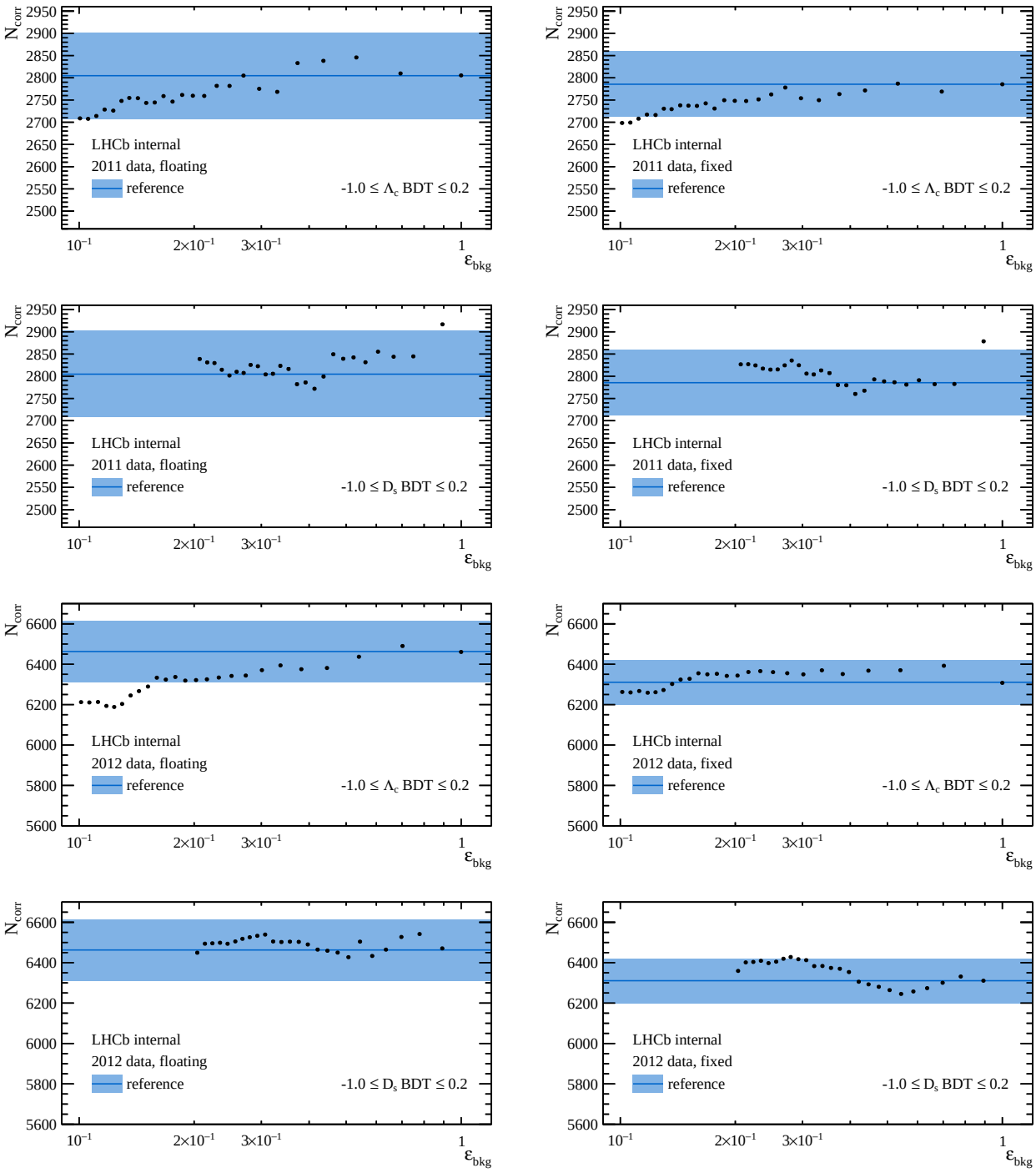


Figure 5.20: Efficiency corrected yields as a function of the background efficiency from a scan through equidistant Λ_c^+ and D_s^- BDT cut values on the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ test sample. The step-size was chosen to be 0.05. The Λ_b^0 signal becomes purer from right to left. Here, either a Λ_c^+ or D_s^- BDT response cut has been applied. The reference yield from a sample without BDT cuts is shown as blue line, the corresponding statistical uncertainty as blue band. As the statistical uncertainties on the scanned points would be 100 % correlated to the reference point, only uncertainties due to limited calibration channel statistics have been propagated. Those are usually smaller than the markers of the scanned points. The Λ_c^+ (D_s^+) BDT scan for 2011 and 2012 data are shown in the top (second) and third (bottom) row respectively. The left column shows the scan where nuisance parameters were left floating, while the right column shows fits with fixed nuisance parameters from the nominal signal fit.

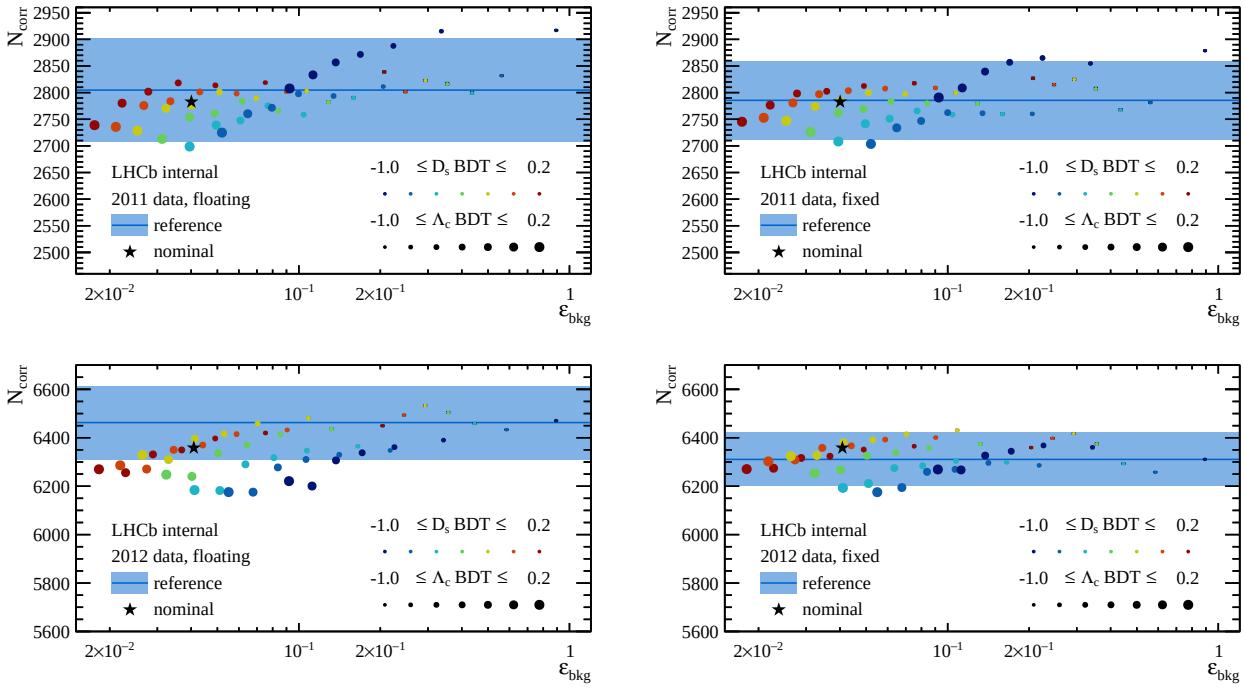


Figure 5.21: Efficiency corrected yields as a function of the background efficiency when both, Λ_c^+ and D_s^- BDT cuts have been applied to the test sample. The step-size was chosen to be 0.2. The Λ_b^0 signal becomes purer from right to left. The reference yield from a sample without BDT cuts is shown as blue line, the corresponding statistical uncertainty as blue band. As the statistical uncertainties on the scanned points would be 100 % correlated to the reference point, only uncertainties due to limited calibration channel statistics have been propagated. Those are usually smaller than the markers of the scanned points. The BDT scan for 2011 data is shown on the top two plots, that for 2012 data on the bottom two. The left column shows the scan where nuisance parameters were left floating, while the right column shows fits with fixed nuisance parameters from the nominal signal fit.

6 Selection of signal and reference channel

The selection for both $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays for the branching fraction measurement is described. The trigger and stripping selection is mostly covered by that of the *D-from-B* BDT calibration channels in Sec. 5.1.2, because the same trigger and a similar stripping selection for $X_b \to X_c \pi^-$, $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ [C36] and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ [C34] based on the `Beauty2Charm` module are used.

Differences are introduced in the Λ_b^0 candidate selection, where no additional cuts on the bachelor kaon are required, so that it must only pass the common track selection in Tab. 5.1. The invariant mass combination of the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ daughters before the vertex fit is required to be `(5200 < AM < 9000) MeV` (*cf.* cut #20 in Tab. 5.3); for $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ it is `(5200 < AM < 7000) MeV`. Moreover, both Λ_b^0 candidates need to pass IP χ^2 and direction angle cuts (`BPVIPCHI2 < 25`, #23 and `BPVDIRA > 0.999`, #24 in Tab. 5.3), which were not required for the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ calibration channel. A set of fiducial cuts in $p_{\mathrm{T},\Lambda_b^0}$ and $y_{\Lambda_b^0}$, to facilitate the efficiency correction, is applied offline as shown in Fig. 6.1.

Up to this point, only loose PID information (cuts #11–#14 in Tab. 5.2) is used. Cuts on the *D-from-B* BDTs and the PID variable `ProbNNk` of the bachelor kaon have thus been studied in a cut optimisation scan using a custom-built generic algorithm described in appendix B.1. The values found there have been rounded to the values in Tabs. 6.1 and 6.2. The rounding effect is negligible, since the gradients of signal- and background efficiencies in the cut parameter space are small around the chosen cuts, resulting in a small variation in the number of selected candidates. This is qualitatively understandable, since the classifier responses in Figs. 5.9, 5.12, 5.15 and page 29 of Ref. [I3] are relatively flat and sparsely populated around the cut values. Initially, alternative selection strategies including more variables were considered. Those are summarised in appendix B.2

The $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays involve six final state particles and are therefore prone to misidentification backgrounds. More than 50 combinations [C38] that were likely to exhibit narrow mass peaks have been studied with the help of single particle momentum-asymmetries, *cf.* Sec. 5.1.3. Surprisingly, also due to the excellent performance of the *D-from-B* BDTs in rejecting these backgrounds, only four misidentified peaks remained in the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ sample, and two in the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ sample. Three of the six peaks were due to ϕ mesons where a K^+ has been misidentified as proton; two peaks were due to $D^*(2010)^-$ decays where the bachelor π^- has been misidentified as K^-; and one peak in the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ sample is due to D_s^+ decays, where the K^+ is misidentified as a proton. These combinations have been vetoed with the cuts shown in Tab. 6.1 and 6.2. The corresponding plots are shown appendix B.3. To avoid confusion, the offline invariant mass ranges used for fitting are tabulated as well. The cuts are also used to create Fig. 6.1 and the plots in appendix B.3.

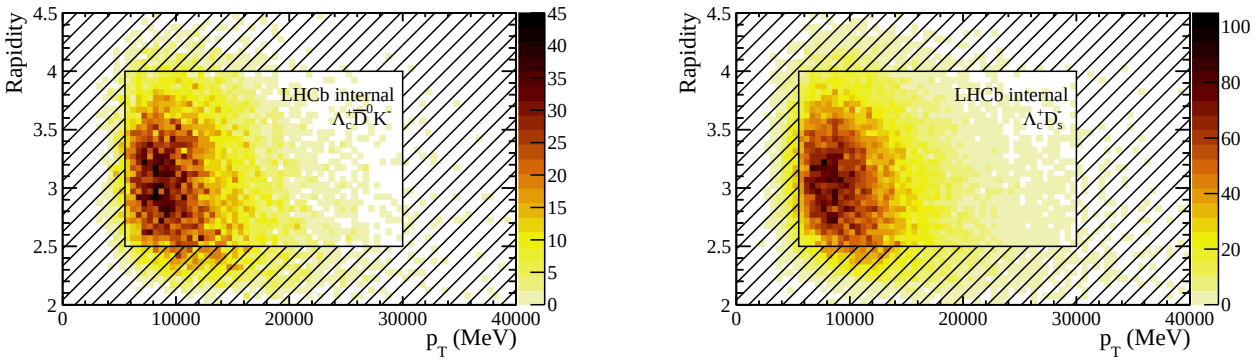


Figure 6.1: Candidates passing all but the fiducial selection in $p_{\mathrm{T},\Lambda_b^0}$ and $y_{\Lambda_b^0}$ for $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ (left) and $\Lambda_c^+ D_s^-$ decays (right). The fiducial cut removes candidates in the hatched region.

Type	Variable	Range/Cut
Fiducial	$p_{\mathrm{T},\Lambda_b^0}$	$\in [5.5, 30]\,\mathrm{GeV}$
	$y_{\Lambda_b^0}$	$\in [2.5, 4.0]$
Veto	$M_{\mathrm{inv}}(K^-_{\Lambda_c^+}\{p \to K^+\})$	$\notin [1011.7, 1027.7]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(K^-_{\mathrm{bachelor}}\{p \to K^+\})$	
	$M_{\mathrm{inv}}(\bar{D}^0\{K^-_{\mathrm{bachelor}} \to \pi^-\})$	$\notin [2010.4, 2013.4]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(\bar{D}^0\{K^-_{\Lambda_c^+} \to \pi^-\})$	
PID	Λ_c^+ BDT	> -0.15
	$\bar{D}^0$ BDT	> -0.25
	K^- `ProbNNk`	$> \ 0.18$
Mass	$M_{\mathrm{inv}}(pK^-\pi^+)$	$\in [2258, 2318]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(K^+\pi^-)$	$\in [1822, 1912]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(\Lambda_c^+\bar{D}^0K^-)$	$\in [5350, 5850]\,\mathrm{MeV}$

Table 6.1: Summary of $\Lambda_c^+\bar{D}^0K^-$ selection cuts applied on top of the stripping selection.

Type	Variable	Range/Cut
Fiducial	$p_{\mathrm{T},\Lambda_b^0}$	$\in [5.5, 30]\,\mathrm{GeV}$
	$y_{\Lambda_b^0}$	$\in [2.5, 4.0]$
Veto	$M_{\mathrm{inv}}(K^-_{\Lambda_c^+}\{p \to K^+\})$	$\notin [1011.7, 1027.7]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(K^-_{\Lambda_c^+}\pi^+_{\Lambda_c^+}\{p \to K^+\})$	$\notin [1938.3, 1998.3]\,\mathrm{MeV}$
		OR p`ProbNNp`$\cdot(1-p$`ProbNNk`$) > 0.05$
PID	Λ_c^+ BDT	> -0.15
	D_s^+ BDT	> -0.25
Mass	$M_{\mathrm{inv}}(pK^-\pi^+)$	$\in [2258, 2318]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(K^+K^-\pi^-)$	$\in [1935, 2005]\,\mathrm{MeV}$
	$M_{\mathrm{inv}}(\Lambda_c^+D_s^-)$	$\in [5420, 5820]\,\mathrm{MeV}$

Table 6.2: Summary of $\Lambda_c^+D_s^-$ selection cuts applied on top of the stripping selection.

7 Mass fits

Fits to invariant mass spectra associated to the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays are detailed. The challenging part of the fits is to determine which components *feed down* into the b-hadron mass spectra. Considerations to these components are presented in Secs. 7.1–7.3 and in appendix C.1. Fit results are presented in Sec. 7.4.

7.1 Feed-down backgrounds in the $\Lambda_c^+ \bar{D}^0 K^-$ invariant mass

In addition to combinatorial and misidentification backgrounds, which were subject to the previous section, there are partially reconstructed decays which populate the invariant mass distribution of $\Lambda_c^+ \bar{D}^0 K^-$. Both, the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ branching ratio measurements are affected by these feed-down backgrounds. Recall that the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ decay is only partially reconstructed itself, but it dominates the spectrum by far. In the selected invariant mass range, real feed-down, *i.e.* without misidentified particles in the decay chain, is expected from Λ_b^0 decays with missing photons, a single missing pion or Ξ_b decays with missing photons and one or two missing pions. The only feed-down decay considered in the nominal fit is $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^0\right]_{\Sigma_c(2455)^+} \bar{D}^0 K^-$. For systematic studies, the decays $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 \left[K^- \pi^0\right]_{K^*(892)^-}$, $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^0\right]_{\Sigma_c(2520)^+} \bar{D}^0 K^-$ and $\Xi_b^- \to \Lambda_c^+ \left[\bar{D}^0 \pi^-\right]_{D^*(2010)^-} K^-$ were considered. These decays are discussed in the following. A more detailed estimation of the level of feed-down from these and other channels can be found in appendix C.1.

Since the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ signal is a feed-down decay itself, it has to be disentangled from other resonant and non-resonant components in the four-body $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0$ decay. The largest contributions beside $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0 \pi^0\right]_{\bar{D}^{*0}} K^-$ are expected from $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 \left[K^- \pi^0\right]_{K^*(892)^-}$ and $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^0\right]_{\Sigma_c(2455)^+} \bar{D}^0 K^-$ decays. Even though the $\Lambda_b^0 \to \Sigma_c(2455)^+ \bar{D}^0 K^-$ component is similar to that of $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0 \pi^0\right]_{\bar{D}^{*0}} K^-$, as shown in Fig. 7.1, the fit is able to separate them. However, sizeable anti-correlations (down to -0.5) are measured between their fit fractions. Such large anticorrelations can affect the fit stability, which is tested in appendix C.2 and found to give consistent results.

The $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^0\right]_{\Sigma(2520)_c^+} \bar{D}^0 K^-$ component, shown in Fig. 7.1, is omitted due to an insignificant fit fraction. Further, only the high mass tail of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 \left[K^- \pi^0\right]_{K^*(892)^-}$ contributes and is not separable from combinatorial backgrounds. In fact, adding the component to the nominal fit yields a negative fit fraction compatible with a statistical fluctuation. However, both decays are taken into account for evaluation of the systematic uncertainty.

Possible feed-down background into the Λ_b^0 signal region comes from Ξ_b decays missing a pion. These backgrounds are estimated to be up to ∼2 % relative to the signal yield (*cf.* appendix C.1). The decay $\Xi_b^- \to \Lambda_c^+ \left[\bar{D}^0 \pi^-\right]_{D^*(2010)^-} K^-$ (Fig. 7.1(f)) is expected to contribute most. Since the fit finds negative fit fractions, the component is only taken into account for evaluation of the systematic uncertainty.

In a $\Lambda_c^+ \bar{D}^0 K^-$ invariant mass range below ∼5350 MeV, decays with two missing pions can contribute. The spectrum is expected to be dominated by $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^- \pi^+\right]_{\Lambda_c^{*+}} \bar{D}^0 K^-$ decays, where Λ_c^{*+} essentially means $\Lambda_c(2595)^+$ and $\Lambda_c(2625)^+$. The data in this region is available, but shows a smooth turn-on around 5200 MeV. This is due to a mass combination cut in the stripping line and intermediate steps like the decay tree fit. The turn-on is not reproduced in generator level simulations of $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^- \pi^+\right]_{\Lambda_c^{*+}} \bar{D}^0 K^-$ decays, due to the missing simulation of detector response, reconstruction and stripping. Consequently, these models can not be used for fitting, such that the $\Lambda_c^+ \bar{D}^0 K^-$ invariant mass range is chosen to be [5350,5850] MeV.

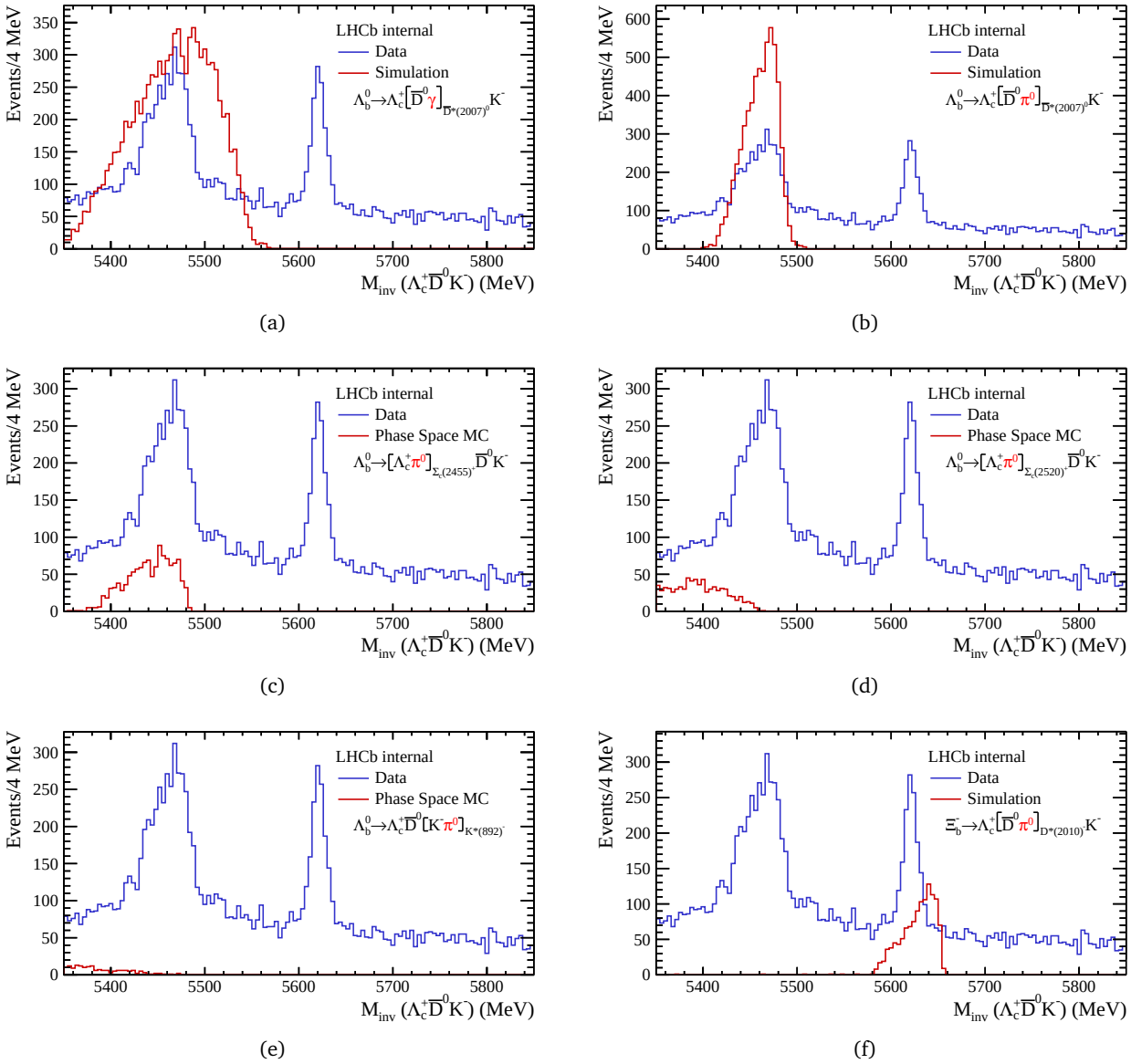


Figure 7.1: Simulated feed-down decays in the $\Lambda_c^+ \overline{D}^0 K^-$ invariant mass. The raw simulated sample size and the selected data is shown. The distributions are thus not to scale. (a) $\Lambda_b^0 \to \Lambda_c^+ \left[\overline{D}^0 \gamma\right]_{\overline{D}^{*0}} K^-$. (b-e) Resonant modes in the four-body decay of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^- \pi^0$. (f) The largest expected Ξ_b feed-down: $\Xi_b^- \to \Lambda_c^+ \left[\overline{D}^0 \pi^-\right]_{D^*(2010)^-} K^-$. Both $\Lambda_b^0 \to \Lambda_c^+ \left[\overline{D}^0 \pi^0\right]_{\overline{D}^{*0}} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \left[\overline{D}^0 \gamma\right]_{\overline{D}^{*0}} K^-$ simulation samples were produced in the centralised LHCb simulation chain. Samples for the other decays are generator level phase space simulation, *i.e.* the output of PYTHIA and EVTGEN without simulation of the detector response is taken and the decay matrix element is constant.

7.2 Feed-down backgrounds in the $\Lambda_c^+ D_s^-$ invariant mass

The feed-down background dominating the $\Lambda_c^+ D_s^-$ invariant mass is $\Lambda_b^0 \to \Lambda_c^+ \left[D_s^- \gamma\right]_{D_s^{*-}}$, where the photon is not reconstructed. This is the only background which is taken into account, but further considerations are briefly sketched.

Decays to higher excited D_s^- states are expected to be suppressed by roughly a factor 5 with respect to the $\Lambda_b^0 \to \Lambda_c^+ \left[D_s^- \gamma\right]_{D_s^{*-}}$ decay[1].

Decays with a single missing pion are only expected from excited D_s^- states, and not from $\Lambda_b^0 \to \left[\Lambda_c^+ \pi^0\right]_{\Sigma_c^{(*)+}} D_s^-$ due to isospin conservation of the spectator quark system [216]. Still, modelling and composing the components with correct proportions would require a dedicated thorough study, since Λ_b^0 polarisation effects are expected to play a role [217].

Contrary to the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ channel, Ξ_b feed-down into the signal mass region of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays is expected to be negligible, because of CKM-, colour- and fragmentation fraction-suppression ($f_{\Xi_b^0}/f_{\Lambda_b^0} \approx 0.2$ [218]). Accounting for selection efficiencies in addition, the most likely decays of the diagram in Fig. 7.2 are expected to be as low as $\approx 0.1\,\%$ compared to the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ yield.

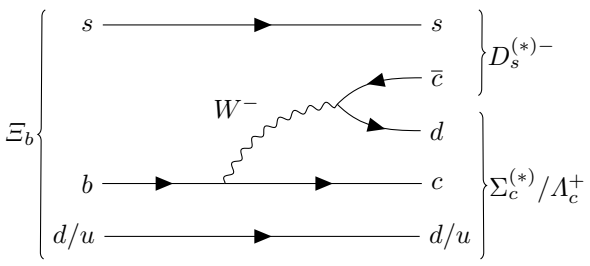


Figure 7.2: Feynman diagram of $\Xi_b^0 \to \Lambda_c^+/\Sigma_c^{(*)+} D_s^{(*)-}$ decays.

7.3 Λ_b^0 decays into single- and no charm

Decays that peak at the Λ_b^0 mass in the $\Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_c^+ D_s^-$ invariant mass spectra, but not in one or both X_c invariant mass systems are examined. Namely, these are the decays $\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-$, $\Lambda_b^0 \to p K^- \pi^+ \bar{D}^0 K^-$, $\Lambda_b^0 \to p K^- \pi^+ D_s^-$ and $\Lambda_b^0 \to p K^- \pi^+ \pi^- K^+ K^-$. Their yields can be measured directly in three-dimensional fits to the X_b and X_c systems. It is argued here, that only $\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-$ decays need to be taken into account.

Some general thoughts concerning suppression of single- and no charm rates from the imposed selection and enhancement due to charm branching fractions are discussed first. Foremost, the non-resonant decay products of single- and no charm decays have to be in their respectively selected invariant mass region (*cf.* Tab. 6.1). This means for instance, that the invariant mass of the $K^+\pi^-$ pair in a single charm decay $\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-$ has to be within [1822,1912] MeV. This suppresses these background substantially, but the rate of suppression is difficult to estimate without resorting to simulation.

Further suppression is expected from other charm hadron selection criteria from the *D-from-B* BDTs and in the stripping. However, the variables used there do not explicitly require a geometrical separation of the X_c w.r.t. the X_b vertices. The selection effects are counter-balanced by the charm hadron branching fractions which are typically in the order of 5 % and thus increase the yield of non-charmed decays by a factor 20. It will be shown later, that the suppression from the selection efficiency is an order of magnitude larger than the enhancement from charm branching fractions.

The individual decays shall be discussed now in brevity. The decay $\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-$ is similar to the *D-from-B* control channel decay $\Lambda_b^0 \to \Lambda_c^+ \pi^-$, where a K^+K^- pair is created from vacuum. High total rates, up to factors of 50 (5) compared to $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ ($\Lambda_b^0 \to \Lambda_c^+ D_s^-$) are expected, including the mentioned factor 20 from the charm branching fractions. So, even if the selection efficiency is as low as a few per-mill (few per-cent) in the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ ($\Lambda_b^0 \to \Lambda_c^+ D_s^-$) channel, a rate could be measured. The decays are thus taken into account in the nominal fits.

The decays $\Lambda_b^0 \to p K^- \pi^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to p K^- \pi^+ D_s^-$ are neither tree-level nor penguin-processes and thus expected to be highly suppressed. The no-charm decay $\Lambda_b^0 \to p K^- \pi^+ \pi^- K^+ K^-$ is doubly CKM suppressed and further suppressed by the invariant mass range of the charm system. Both decays are thus not taken into account.

[1] For instance $B^0 \to D^- D_s^+$ compared to $B^0 \to D^- D_{s0}(2317)^+/D_{s1}(2460)^+$

7.4 Fit model and results

The fits are carried out in three dimensions, spanned by the X_c and X_b invariant mass systems. It has the disadvantage that the relatively wide sidebands in the X_c invariant mass distributions increase the amount of combinatorial background. On the other hand, there are no mass cuts which introduce inefficiencies, and the $\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-$ single charm contribution discussed in the previous section, as well as the type of combinatorial background, can be extracted directly. The *D-from-B* BDTs improved the rejection of combinatorial background to a level which makes three-dimensional fits feasible.

Signal contributions in all dimensions are modelled by the sum of two Crystal Ball functions [214], Eq. (5.7), with a common mean and symmetric tail parameters. The functional form used in the fit is

$$DCB(M;\mu,\sigma,\alpha,n,f_{CB},f_\sigma) = f_{CB}CB(M;\mu,\sigma,\alpha,n) + (1-f_{CB})CB(M;\mu,f_\sigma\cdot\sigma,-\alpha,n)\,. \tag{7.1}$$

The parameters $f_{\mathrm{CB}}, \alpha, n$ and f_σ, listed in Tab. 7.1, are determined from fits to the signal simulation samples, shown in Fig. 7.3.

Parameter	Signal			Reference		
	Λ_b^0	Λ_c^+	$\bar{D}^0$	Λ_b^0	Λ_c^+	D_s^-
f_{CB}	0.62 ± 0.03	0.59 ± 0.04	0.76 ± 0.02	0.55 ± 0.05	0.45 ± 0.05	0.52 ± 0.06
α	2.33 ± 0.08	2.26 ± 0.09	2.94 ± 0.12	2.08 ± 0.14	2.48 ± 0.20	2.39 ± 0.19
n	1.76 ± 0.18	1.09 ± 0.29	0.22 ± 0.16	2.34 ± 0.43	0.47 ± 0.33	0.57 ± 0.33
f_σ	1.67 ± 0.03	1.87 ± 0.04	2.00 ± 0.04	1.46 ± 0.09	1.92 ± 0.08	1.77 ± 0.06

Table 7.1: Signal shape parameters determined by fits to simulated signal samples.

Combinatorial backgrounds are modelled with exponential p.d.f.s, $E(M;\tau) = e^{M\tau}$, in the nominal fit. The partially reconstructed background in the reference channel, which is dominated by $\Lambda_b^0 \to \Lambda_c^+ \left[D_s^- \gamma\right]_{D_s^{*-}}$ decays, is well separated from the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ signal. Therefore it only has a minor effect on the branching fraction measurement and is chosen to be modelled by a heuristically chosen Novosibirsk p.d.f. [212], Eq. (5.3). The feed-down backgrounds in the signal channel have been discussed above and in appendix C.1. They are modelled with kernel density p.d.f.s, $KDE(M|x_{\mathrm{d}})$, from the `RooKeysPdf` class [C29], whose shapes are determined from the simulation samples x_{d} corresponding to decay d, as shown in Fig. 7.1.

Similar to the fit models of the *D-from-B* BDT control channels, Secs. 5.3–5.5, the mentioned building blocks are now used to construct the three-dimensional p.d.f. It is convenient to define

$$\begin{aligned}
\mathscr{F}_{\mathrm{sig},\Lambda_b^0} &:= DCB\left(M_{\mathrm{inv}}(\Lambda_c^+\bar{D}^0K^-);\mu_{\Lambda_b^0},\sigma_{\Lambda_b^0},\tilde{\alpha}_{\Lambda_b^0},\tilde{n}_{\Lambda_b^0},\tilde{f}_{CB,\Lambda_b^0},\tilde{f}_{\sigma,\Lambda_b^0}\right)\\
\mathscr{F}_{\mathrm{bkg},\Lambda_b^0} &:= E\left(M_{\mathrm{inv}}(\Lambda_c^+\bar{D}^0K^-);\tau_{\Lambda_b^0}\right)\,,
\end{aligned}$$

with analogous definitions for the Λ_c^+ and $\bar{D}^0$ components, as well as the Λ_b^0,Λ_c^+ and D_s^- components in the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay. The parameters with a tilde have been fixed to their best fit estimates from simulation, listed in Tab. 7.1. Further, the Novosibirsk- and kernel density p.d.f.s are abbreviated as

$$\begin{aligned}
\mathscr{F}_{\mathrm{PR},\Lambda_b^0} &:= N\left(M_{\mathrm{inv}}(\Lambda_c^+D_s^-);M_{\mathrm{p,PR}},\tilde{\eta}_{\mathrm{PR}},\tilde{\sigma}_{\mathrm{E,PR}}\right)\\
\mathscr{F}_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\pi^0]_{\bar{D}^{*0}}K^-} &:= KDE\left(M_{\mathrm{inv}}(\Lambda_c^+\bar{D}^0K^-)\middle|x_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\pi^0]_{\bar{D}^{*0}}K^-}\right)\,,
\end{aligned}$$

with analogous definitions for the $\Lambda_b^0 \to \Lambda_c^+[\bar{D}^0\gamma]_{\bar{D}^{*0}}K^-$ and $\Lambda_b^0 \to \Sigma_c(2455)^+\bar{D}^0K^-$ components.

Two parameters of the Novosibirsk p.d.f. had to be fixed and are accentuated by a tilde. This is done because the three parameters, $M_{\mathrm{p,PR}}$, η_{PR} and $\sigma_{\mathrm{E,PR}}$ are highly correlated in floating fits ($\rho_{M_{\mathrm{p,PR}},\eta_{\mathrm{PR}}} = 0.993$, $\rho_{M_{\mathrm{p,PR}},\sigma_{\mathrm{E,PR}}} = 0.980$, $\rho_{\eta_{\mathrm{PR}},\sigma_{\mathrm{E,PR}}} = 0.993$). The high correlation leads to a faulty estimate of the covariance matrix. As the uncertainty on the signal yield, despite a globally bad covariance matrix in some fits, did not change when either combination of these parameters are fixed, it is concluded that fixing the two parameters is legitimate.

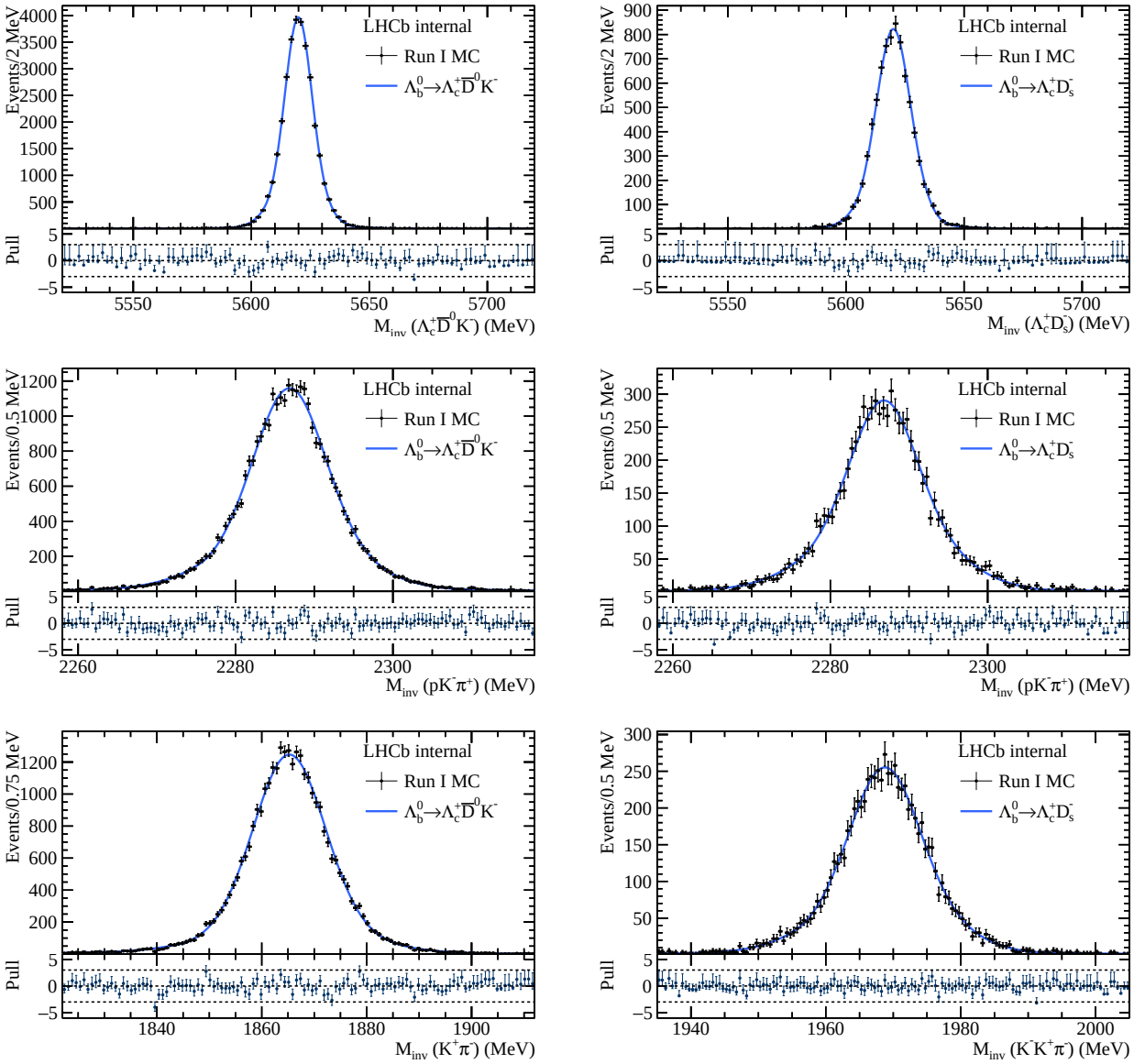


Figure 7.3: Projections of the three-dimensional $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ (left) and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ fits (right) to the respective signal simulation samples.

With the above definitions, the full fit model in the $\Lambda_c^+ \bar{D}^0 K^-$ invariant mass is given by

$$
\begin{aligned}
\mathscr{F}_{\Lambda_c^+\bar{D}^0K^-} = & N_{\Lambda_b^0\to\Lambda_c^+\bar{D}^0K^-}\mathscr{F}_{\text{sig},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\bar{D}^0} \\
& + f_{\Lambda_b^0\to\Lambda_c^+K^+\pi^-K^-}N_{\Lambda_b^0\to\Lambda_c^+\bar{D}^0K^-}\mathscr{F}_{\text{sig},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{bkg},\bar{D}^0} \\
& + N_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\pi^0]_{\bar{D}^{*0}}K^-}\mathscr{F}_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\pi^0]_{\bar{D}^{*0}}K^-}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\bar{D}^0} \\
& + N_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\gamma]_{\bar{D}^{*0}}K^-}\mathscr{F}_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\gamma]_{\bar{D}^{*0}}K^-}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\bar{D}^0} \\
& + f_{\Lambda_b^0\to\Sigma_c(2455)^+\bar{D}^0K^-}N_{\Lambda_b^0\to\Lambda_c^+\bar{D}^0K^-}\mathscr{F}_{\Lambda_b^0\to\Sigma_c(2455)^+\bar{D}^0K^-}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\bar{D}^0} \\
& + N_{\text{comb. bkgs}}\Big[f_{\text{comb. }\bar{D}^0}\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{bkg},\Lambda_c^+}\mathscr{F}_{\text{sig},\bar{D}^0} \\
& \qquad + f_{\text{comb. }\Lambda_c^+}\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{bkg},\bar{D}^0} \\
& \qquad + f_{\text{comb. }\Lambda_c^+\bar{D}^0}\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\bar{D}^0} \\
& \qquad + (1 - f_{\text{comb. }\bar{D}^0} - f_{\text{comb. }\Lambda_c^+} - f_{\text{comb. }\Lambda_c^+\bar{D}^0})\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{bkg},\Lambda_c^+}\mathscr{F}_{\text{bkg},\bar{D}^0}\Big] .
\end{aligned}
\tag{7.2}
$$

In a similar way, the nominal $\Lambda_c^+ D_s^-$ model is given by

$$
\begin{aligned}
\mathscr{F}_{\Lambda_c^+D_s^-} = & N_{\Lambda_b^0\to\Lambda_c^+D_s^-}\mathscr{F}_{\text{sig},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\text{D}_\text{s}^-} \\
& + f_{\Lambda_b^0\to\Lambda_c^+K^+\pi^-K^-}N_{\Lambda_b^0\to\Lambda_c^+D_s^-}\mathscr{F}_{\text{sig},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{bkg},\text{D}_\text{s}^-} \\
& + N_{\text{PR}}\mathscr{F}_{\text{PR}}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{sig},\text{D}_\text{s}^-} \\
& + N_{\text{comb. bkgs}}\Big[f_{\text{comb. }\text{D}_\text{s}^-}\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{bkg},\Lambda_c^+}\mathscr{F}_{\text{sig},\text{D}_\text{s}^-} \\
& \qquad + f_{\text{comb. }\Lambda_c^+}\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{sig},\Lambda_c^+}\mathscr{F}_{\text{bkg},D_s^-} \\
& \qquad + (1 - f_{\text{comb. }\bar{D}^0} - f_{\text{comb. }\Lambda_c^+})\mathscr{F}_{\text{bkg},\Lambda_b^0}\mathscr{F}_{\text{bkg},\Lambda_c^+}\mathscr{F}_{\text{bkg},\text{D}_\text{s}^-}\Big] .
\end{aligned}
\tag{7.3}
$$

The combinatorial $\Lambda_c^+ D_s^-$ background is not included in the nominal fit, since its fit fraction came out negative. This component is accounted for in the fit stability tests in appendix C.2 and the determination of the systematic uncertainty (*cf.* Sec. 9.1). Resulting fits to signal and reference channel are shown in Figs. 7.4 and 7.5; and the best fit estimates are tabulated in Tabs. 7.2 and 7.3. These fits are used directly for the computation of sWeights which are needed to determine efficiency corrected yields.

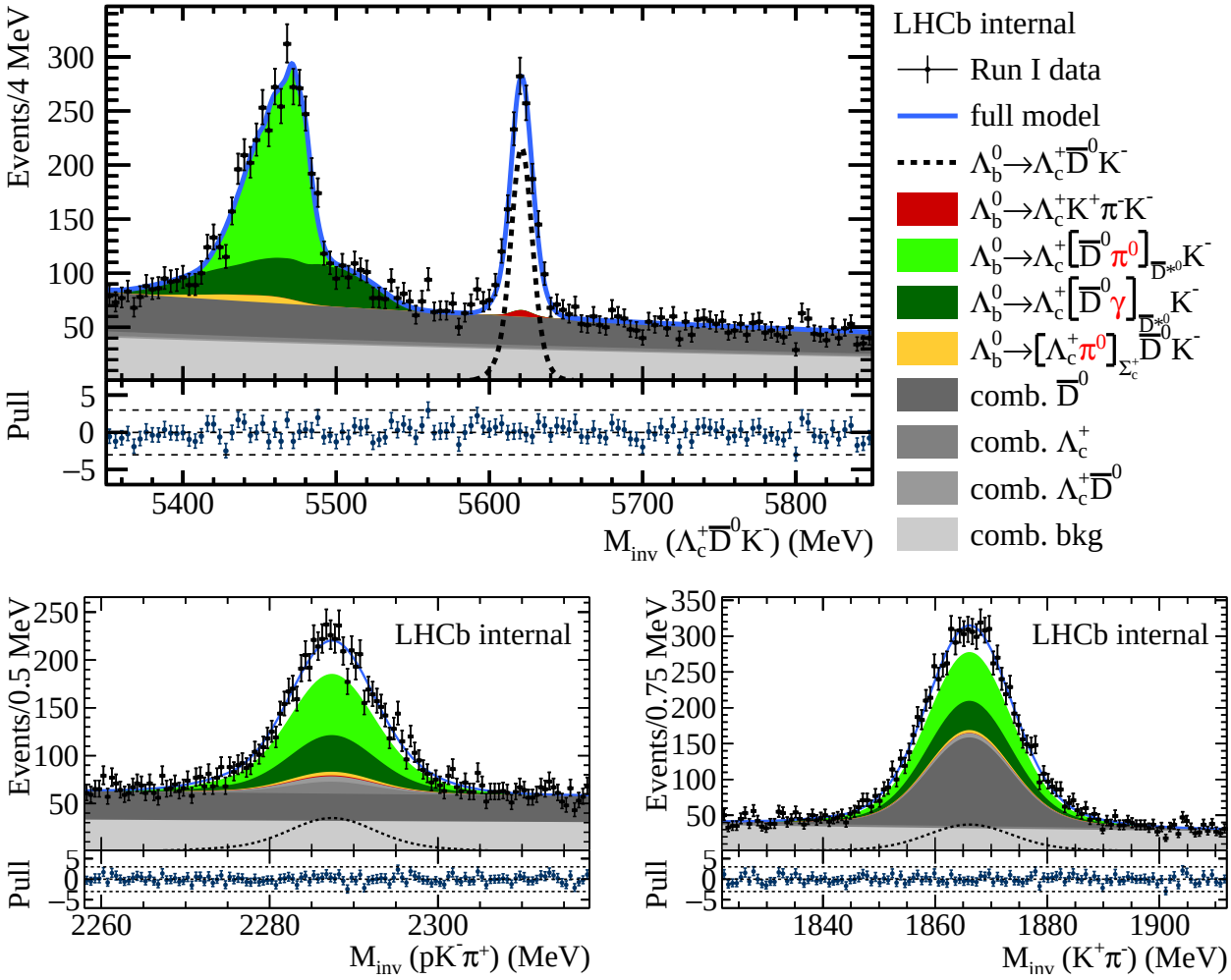


Figure 7.4: Fit projections of the three-dimensional $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ fit. Note that most of the combinatorial background stems from the charm candidate sidebands.

Paramter	Λ_b^0	Λ_c^+	$\bar{D}^0$
μ	5621.01 ± 0.30 MeV	2287.42 ± 0.12 MeV	1866.15 ± 0.12 MeV
σ	6.29 ± 0.25 MeV	4.67 ± 0.10 MeV	7.20 ± 0.11 MeV
τ	$(-1.19 \pm 0.09) \cdot 10^{-3}$ MeV $^{-1}$	$(-1.4 \pm 0.7) \cdot 10^{-3}$ MeV $^{-1}$	$(-2.9 \pm 0.6) \cdot 10^{-3}$ MeV $^{-1}$
$N_{\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-}$		1020 ± 40	
$f_{\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-}$		0.029 ± 0.022	
$N_{\Lambda_b^0 \to \Lambda_c^+ [\bar{D}^0 \pi^0]_{\bar{D}^{*0}} K^-}$		1870 ± 80	
$N_{\Lambda_b^0 \to \Lambda_c^+ [\bar{D}^0 \gamma]_{\bar{D}^{*0}} K^-}$		1120 ± 90	
$f_{\Lambda_b^0 \to \Sigma_c(2455)^+ \bar{D}^0 K^-}$		0.11 ± 0.08	
$N_{\text{comb. bkgs}}$		7770 ± 110	
$f_{\text{comb. } \Lambda_c^+}$		0.047 ± 0.010	
$f_{\text{comb. } \bar{D}^0}$		0.439 ± 0.014	
$f_{\text{comb. } \Lambda_c^+ \bar{D}^0}$		0.021 ± 0.011	

Table 7.2: Best estimates of floating parameters in the $\Lambda_c^+ \bar{D}^0 K^-$ model.

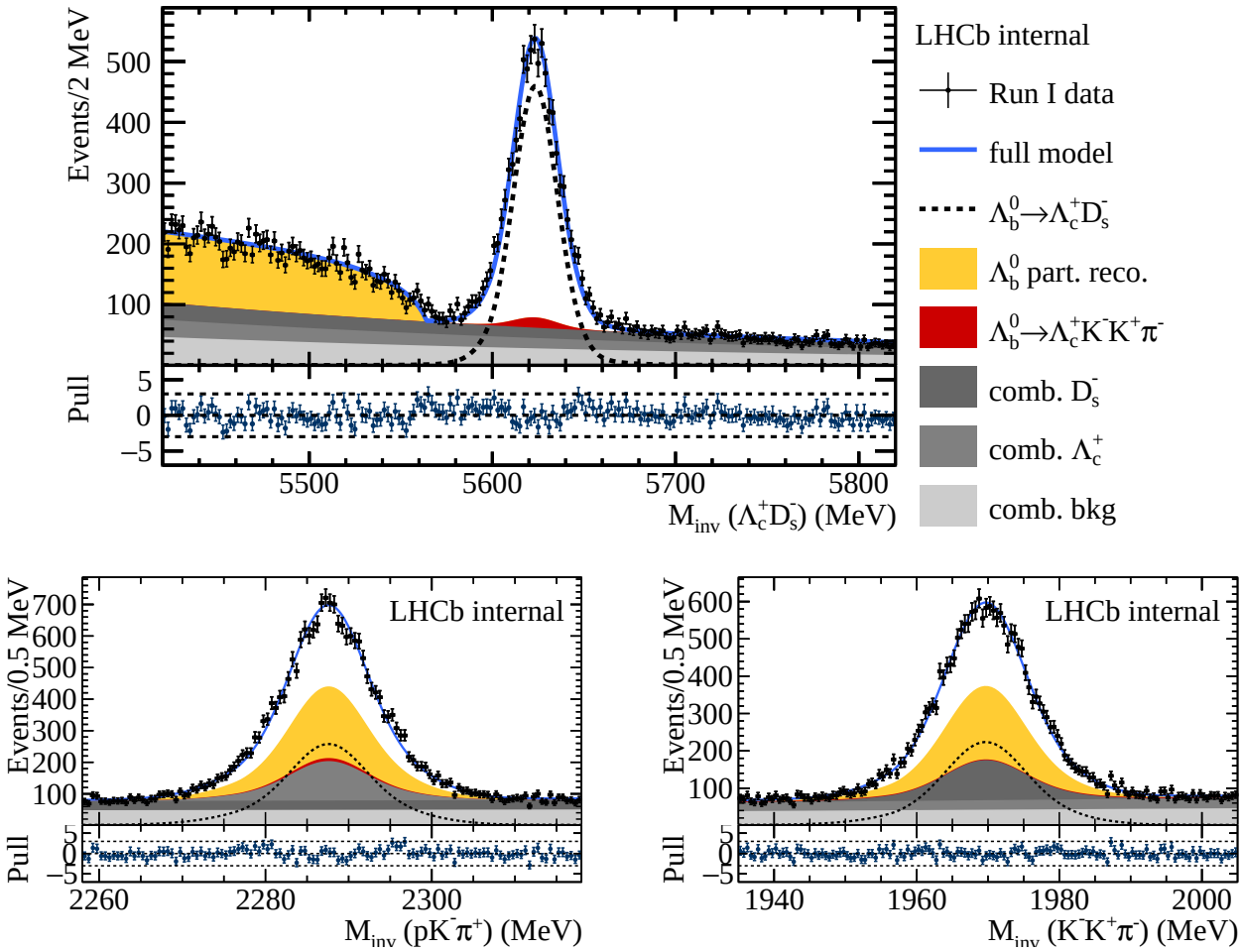


Figure 7.5: Fit projections of the three-dimensional $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ fit

Paramter	Λ_b^0	Λ_c^+	D_s^-
μ	5623.43 ± 0.18 MeV	2287.59 ± 0.06 MeV	1969.66 ± 0.07 MeV
σ	10.81 ± 0.14 MeV	4.14 ± 0.04 MeV	5.12 ± 0.05 MeV
τ	$(-2.51 \pm 0.09) \cdot 10^{-3}$ MeV $^{-1}$	$(8 \pm 6) \cdot 10^{-4}$ MeV $^{-1}$	$(2.5 \pm 0.5) \cdot 10^{-3}$ MeV $^{-1}$
$N_{\Lambda_b^0 \to \Lambda_c^+ D_s^-}$		7400 ± 110	
$f_{\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-}$		0.037 ± 0.008	
N_{PR}		6520 ± 110	
$N_{\text{comb. bkgs}}$		$(1.306 \pm 0.016) \cdot 10^4$	
$f_{\text{comb. } D_s^-}$		0.272 ± 0.009	
$f_{\text{comb. } \Lambda_c^+}$		0.274 ± 0.009	

Table 7.3: Best estimates of floating parameters in the $\Lambda_c^+ D_s^-$ model.

8 Efficiencies

This section details the generation of efficiency lookup tables to be able to compute efficiency corrected yields for the branching ratio measurement (*cf.* Eq. (3.2)). The lookup tables are a crucial input for the amplitude analyses of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$.

Several types of efficiencies must be considered, to account for detector acceptance, trigger, reconstruction, stripping and offline selection. The first selection and reconstruction steps up to the offline selection are well modelled by simulation. These steps are mainly based on geometrical and kinematic selections and well understood properties of the detector to which the simulation is adapted to [169]. This part of the efficiency correction is thus called *kinematic* efficiency and also includes fiducial and veto cuts listed in Tabs. 6.1 and 6.2. The mass cuts in these tables are included in the procedure as well, but are 100 % efficient.

The main discriminating power of the following offline selection steps come from cuts on PID variables. These are the *D-from-B* BDTs and the `ProbNNk` cut on the bachelor kaon of the signal channel. Both, *D-from-B* BDTs and `ProbNNk` rely on a multitude of input variables and also exploit correlations between them. Although the simulation is reasonably good at simulating these (*cf.* Fig. B.5), it is not good enough for the targeted precision. For this reason, the efficiency of these selection steps is evaluated in a data-driven manner. All calibration samples and calibration variables are listed in Tab. 8.1.

Step	calibration sample	calibration variables	
		$\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$	$\Lambda_b^0 \to \Lambda_c^+ D_s^-$
Fiducial cuts Acceptance Trigger Reconstruction Stripping Vetos	simulation	Dalitz plot	p_T, rapidity
X_c BDTs	$X_b \to X_c \pi^-$ data	$\log(\mathrm{FD}\chi^2)$, `nTracks`	
PID K^-	$D^{*+} \to \left[K^-\pi^+\right]_{D^0} \pi^+$ data	p, η, `nTracks`	–

Table 8.1: Calibration samples and variables for signal and reference channel for each selection or reconstruction step.

Note that efficiency corrected yields of signal and reference channel by themselves can *not* be interpreted as the number of such decays produced in pp collisions at the LHCb interaction point. That is, because efficiencies and systematic effects which are expected to cancel in the master formula of the relative branching fraction, Eq. (3.1) are not taken into account.

Still, order of magnitude estimates of integrated signal weighted efficiencies are given. The acceptance efficiency ranges from about 92 % in $\Lambda_b^0 \to \Lambda_c^+ \left[\overline{D}^0 \pi^0\right]_{\overline{D}^{*0}} K^-$ decays to 99 % in $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays within the p_T–y fiducial region, which selects about 8 to 9 % of the candidates in both, signal and reference channel. Within the LHCb acceptance and the fiducial region, the trigger efficiency is about 20 % for $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and slightly lower for $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ (18 %). Reconstruction and stripping efficiencies are about to 19 % for $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ and 14 % for $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. The vetos are about 98 % efficient in all channels. Finally, the combined *D-from-B* BDT selection efficiencies are about 93 % for $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ channels and 88 % for the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ channel, while the bachelor kaon PID selection is also about 93 % efficient for the $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ channels.

8.1 Kinematic efficiencies

Lookup tables for the kinematic efficiencies are created from simulation samples. In the process, fiducial and acceptance cut efficiencies are separated from the combined trigger to vetos efficiency.

Efficiencies from simulated samples are computed as the ratio of *truth-matched* reconstructed and selected candidates over the number of generated signal decays within the fiducial region as a function of either the Dalitz plot or p_T and y. The term truth-matching refers to the procedure that associates simulated to reconstructed particles. In this work, this information is obtained with the background classification tool [C3] described in [219]. Note that truth-matching introduces a selection on simulated samples, whose systematic effects are expected to cancel in the ratio of efficiency corrected yields.

The efficiencies of the partially reconstructed $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$ decays are evaluated separately for the $\overline{D}^{*0} \to \overline{D}^0 \pi^0$ and $\overline{D}^{*0} \to \overline{D}^0 \gamma$ final states. It turned out that both integrated efficiencies are compatible within uncertainties. However, this is partly only due to compensation of individual components of the efficiencies, as shown later.

Evaluating efficiencies as a function of certain calibration variables makes implicit assumptions. Either the efficiency is constant as a function of any other variable; Or the efficiency corrected p.d.f. of the phase space that has been integrated out has the same shape in signal and calibration sample; Or both effects lead to a balanced integrated efficiency, which factorises as a function of the chosen calibration variables. Integrating out p_T and y of the Λ_b^0 candidate in the signal channel, is justified by a good description of the variables in simulation (*cf.* Fig. 8.1).

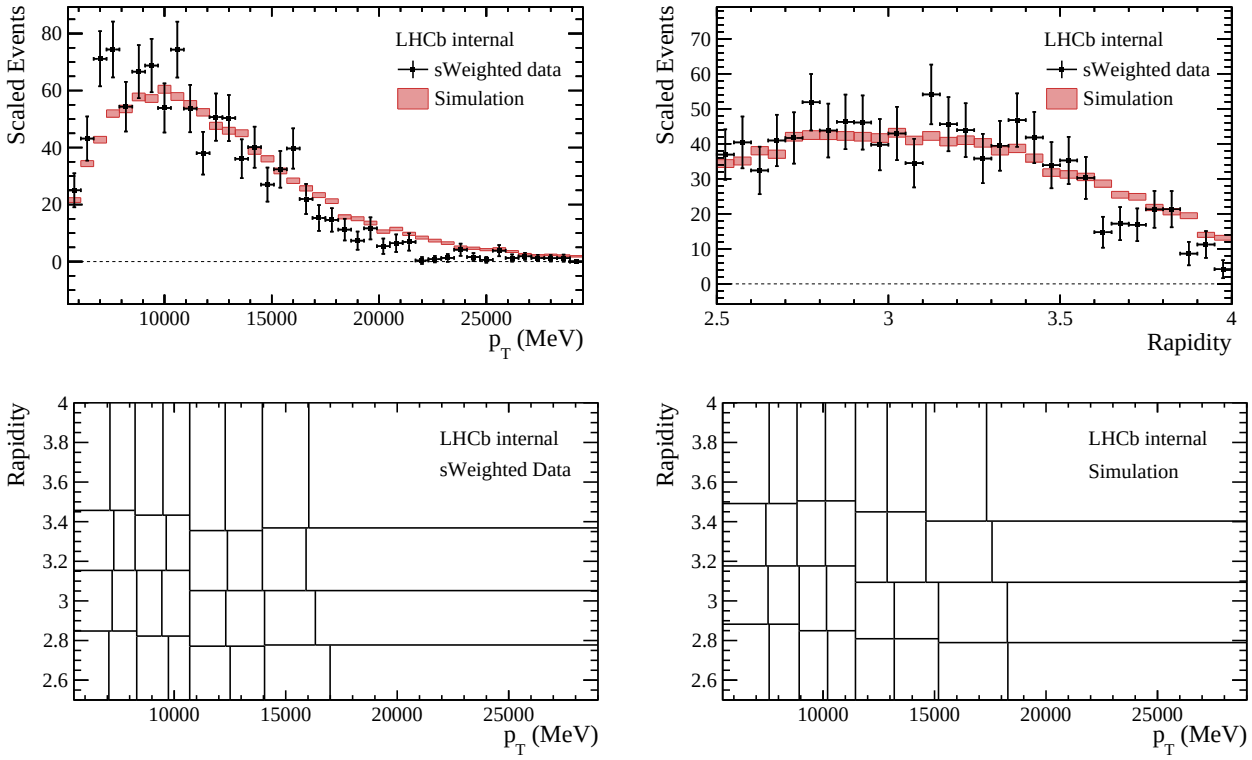


Figure 8.1: Comparison of data and simulation in p_T and y (top). An adaptive binning with the same number of bins is used to compare the two-dimensional distribution (bottom). This is done because the comparison of two conventionally binned two-dimensional distributions by eye is difficult. Especially in this case, where the data-distribution has few entries. The adaptive binning is to be interpreted as a density profile, since all bins contain approximately the same number of events. For data the sum of weights is about 31 per bin, while for simulation there are about 840 events per bin.

The computation of the efficiencies as a function of calibration variables is done by choosing an optimised binning scheme in the two-dimensional calibration variable space. The binning relies on the same algorithm that has been used to generate binning schemes for the *D-from-B* BDT efficiency lookup tables in Sec. 5.6. In each bin, a 68.27 % confidence interval of the efficiency is computed, using a binomial interval estimator, namely the Wilson confidence interval. The interval is detailed in appendix A and computed with the `TEfficiency` class in ROOT [C37].

When applied to data, the calibration variables of the exclusive $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays are calculated after a dedicated decay tree fit (*cf.* Sec. 2.2.4), which now includes the world-average Λ_b^0 mass. Such a fit can not be used for the calibration variables of the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ decays because it would try to constrain the four-momentum of the visible system to that of a Λ_b^0.

8.1.1 Fiducial and acceptance cuts

The acceptance cut is applied at generator level in simulation; *i.e.* after an event is generated and the decays are simulated, but before propagation through the spectrometer (*cf.* Sec. 2.2.6). There, the spectrometer acceptance is a cut on the opening angle θ, namely $10\,\text{mrad} < \theta < 400\,\text{mrad}$ for charged and $5\,\text{mrad} < \theta < 400\,\text{mrad}$ for neutral particles. Since θ is highly correlated with p_T and y, the efficiency of the fiducial cut is evaluated as single integrated number before the acceptance cut.

From kinematic considerations it is likely that the efficiency of this cut varies across the Dalitz plot. Especially in its corners, where one daughter is produced almost at rest. Hence, the LHCb simulation package, GAUSS, is manipulated such that a sample is written to disk before and after the acceptance cut is applied. Now, the acceptance efficiency can be evaluated as a function of the Dalitz plot, as shown in Fig. 8.2. However, a flat acceptance cut efficiency is observed in all cases, meaning that the kinematics is driven by the boost of the Λ_b^0. On the other hand, a clear difference between $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0 \pi^0\right]_{\bar{D}^{*0}} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0 \gamma\right]_{\bar{D}^{*0}} K^-$ is seen supporting separated efficiency correction of these decays.

8.1.2 Trigger to offline selection

The least amount of simulation statistics is available for the trigger to offline selection steps. It is thus crucial to select a binning which optimally reproduces the efficiency profile. Former studies have shown that the choice of equally populated over equidistant binning is far superior – especially if the variables show any kind of dependence or correlation [I10, I11][213]. The adaptive binning algorithm introduced in Secs. 4.1 and 5.6, performs binary splits and uses one parameter to steer the binning granularity. Finer binning schemes lead to larger confidence intervals, while coarser binning schemes tend to be biased, since the true efficiency distribution cannot be resolved; think of a steep slope that gets averaged out in a bin where calibration and signal channel distributions are different. Finding an optimal binning scheme, even with the adaptive algorithm at hand, is a non-trivial task, leading to the systematic study detailed in Sec. 9.2.

It has been found that a good minimum number of reconstructed, selected and truth-matched events per bin is 64 for all $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ decay modes, leading to 128 or 256 bins across the Dalitz plot depending on period and mode. Less simulated events are available in the sample of the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ reference decay channel. The minimum number of events is reduced to 16 and 32 for the 2011 and 2012 simulation samples, resulting in 64 and 128 bins respectively.

Efficiency lookup tables for the trigger to offline selection are shown in Fig. 8.3. Note that the fiducial cut efficiency is computed in a separate step. The PIDCorr technique, *cf.* Sec. 2.2.5 and Ref. [I4], based on kernel density p.d.f.s from the Meerkat package [161] has been used to transform the `ProbNN` variables used in the highly efficient D_s^+ veto in the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ selection (*cf.* Fig. **??**, Tab. 6.2).

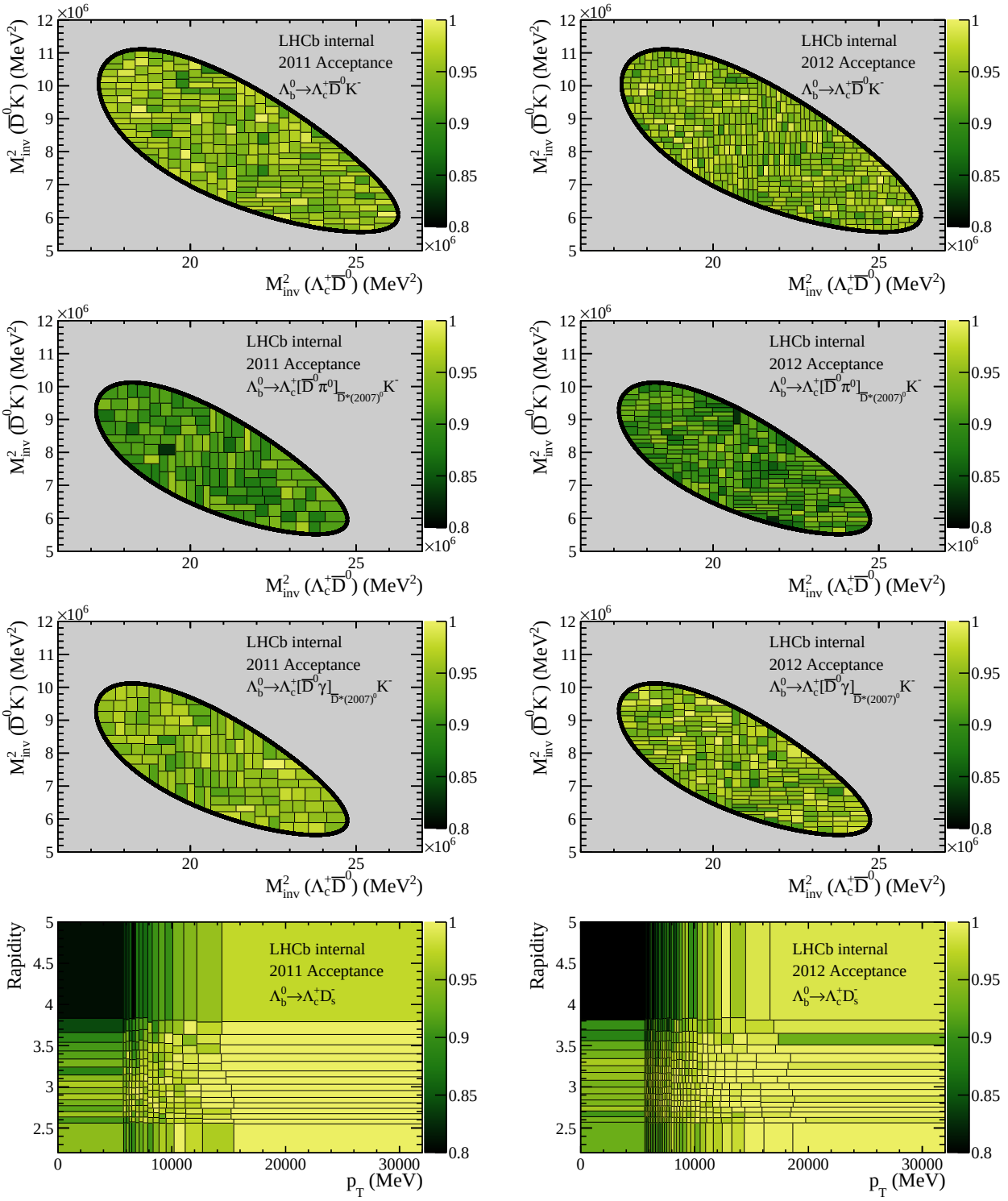


Figure 8.2: Simulated acceptance cut efficiencies for 2011 (left) and 2012 (right) data taking periods, given the fiducial selection ($5.5 < p_T < 30$ GeV and $2.5 < y < 4.0$). The top row shows $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^0 K^-$, the second row $\Lambda_b^0 \to \Lambda_c^+ \left[\overline{D}^0 \pi^0\right]_{\overline{D}^{*0}} K^-$, the third row $\Lambda_b^0 \to \Lambda_c^+ \left[\overline{D}^0 \gamma\right]_{\overline{D}^{*0}} K^-$ and the last row $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decays.
The superimposed Dalitz plot boundaries for the two $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$ decay modes are calculated heuristically. This means that the mother mass has been reduced by (slightly more than) the $\overline{D}^{*0} - \overline{D}^0$ mass difference (due to resolution effects). The width of the simulated $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{*0} K^-$ distribution washes out the Dalitz plot boundary; therefore the $\overline{D}^0$ mass is reduced by 10 MeV in the calculation of the boundary.

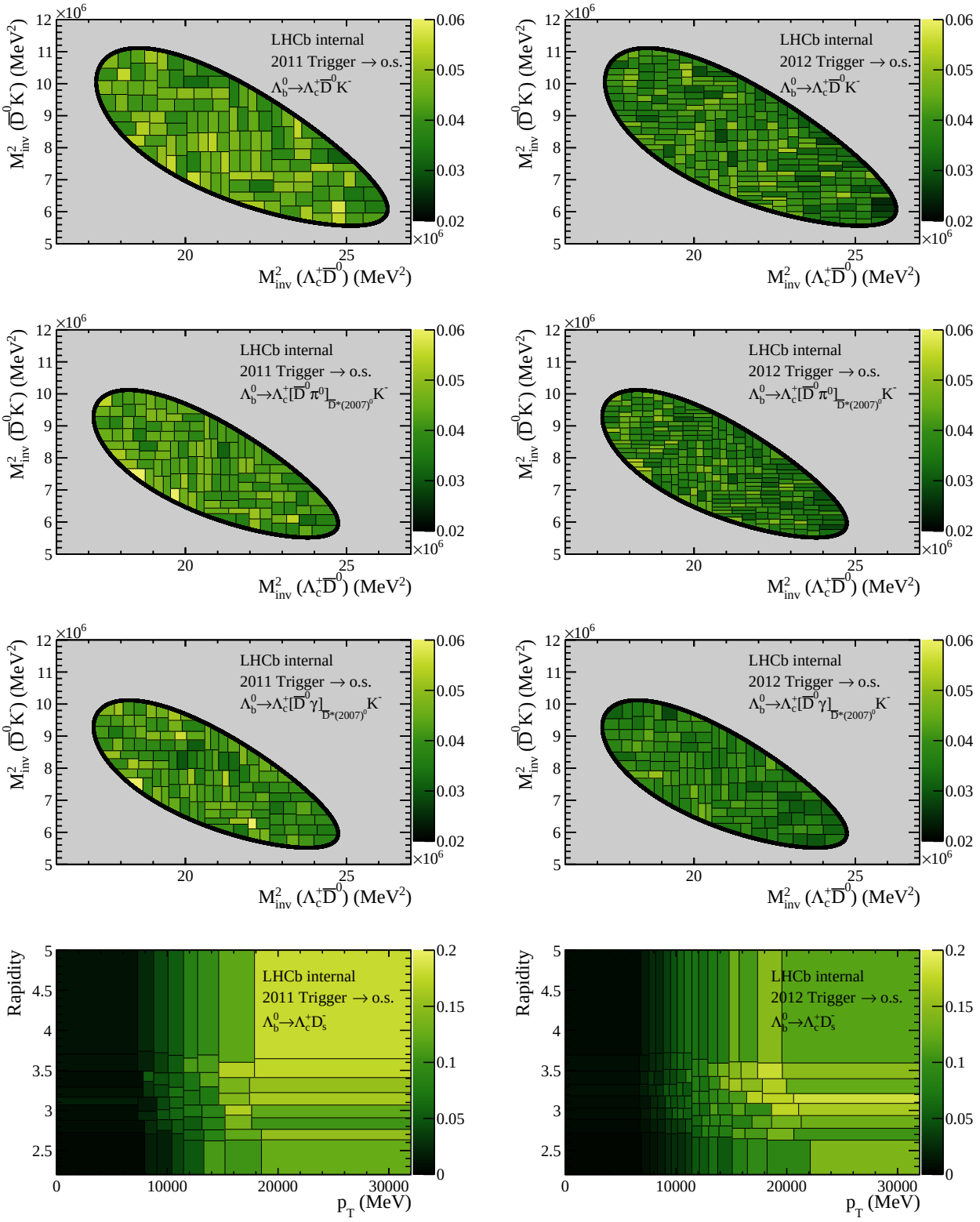


Figure 8.3: Combined efficiency for trigger, reconstruction, stripping and offline selection stages evaluated within the LHCb acceptance. The efficiency of the fiducial cuts is not included.

8.2 PID efficiencies

The efficiencies of the *D-from-B* BDTs have been discussed in Sec. 5.6 and validated in Sec. 5.7. The remaining selection step is on the `ProbNNk` variable of the bachelor kaon. The efficiency lookup tables are obtained from data-driven calibration samples from $D^{*+} \to \left[K^-\pi^+\right]_{D^0} \pi^+$ decays provided by LHCb PID working group in the PIDCalib package, *cf.* Sec. 2.2.5 and Ref. [160].

Here, the PID selection efficiency is evaluated as a function of the number of tracks in the event, momentum and pseudo-rapidity of the bachelor kaon. The efficiency lookup tables are produced in five bins of the number of tracks in the event, with bin-edges 0, 100, 200, 300, 450, 650. Within each of these bins, a two-dimensional adaptive binning in p_T and pseudo-rapidity (η) with at least 512 events per bin that pass the `ProbNNk` selection is created.

This is a much larger number compared to the previously chosen one for the trigger to offline selection efficiency lookup tables. Still, due to the huge statistics of the calibration sample, the adaptive binning created efficiency lookup tables with more than 10,000 bins in the `nTracks` = 100–200 and 200–300 bins of the 2012 sample. On the other hand, choosing a larger number of events per bin compared to the case where simulation samples are used is necessary, because the efficiency lookup tables are created from sWeighted quantities and subject to statistical fluctuations. These fluctuations can even be observed with at least 512 events per bin. In regions, where the efficiency is close to 1, rare ($< 0.1\,\%$ of the cases) fluctuations above 1 occur. They are corrected for by setting the numerator to the denominator value in the efficiency calculation to still have a reasonable estimate of the uncertainty in the bin.

Figure 8.4 shows the highest occupancy bin, which is least populated. In other bins, the binning is too fine to be able to see the bin contents without zooming in. Not surprisingly, the area of regions with high efficiency grow larger with smaller occupancy.

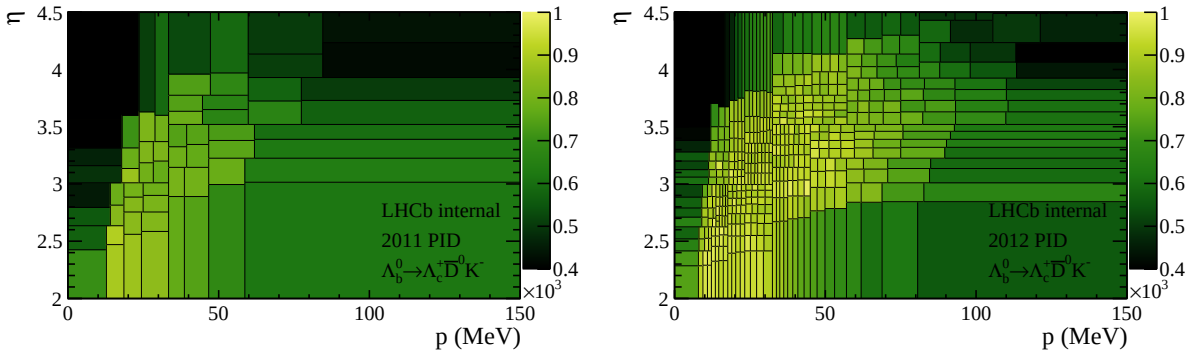


Figure 8.4: PID efficiency lookup tables for the least populated `nTracks` bin from 450 to 650 for 2011 (left) and 2012 data (right).

9 Systematic uncertainties

Due to the simplicity of the master formula (Eq. (3.1)), and the carefully chosen offline signal selection, the list of systematic uncertainties is relatively short. The dominant systematic uncertainties are the choice of fit model and the binning of calibration variables for the efficiency correction. Furthermore, the limited simulation and calibration sample size are taken into account. It is demonstrated that these effects are negligible in the overall uncertainty-budget.

9.1 Fit model

Several choices concerning the fit model have been made in Sec. 7, all of which are tested here systematically. This includes not only variation of the signal and background p.d.f.s, but also adding statistically insignificant components, which are expected to be seen with more data (c.f. Secs. 7 and C.1). Differences of nominal and alternative models are summarised in Tabs. 9.1 and 9.2. The resulting systematic uncertainties are negligible for the measurement of the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ branching fraction, but sizeable for $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$.

One of the alternative models replaces the nominal signal shape with the sum of two Gaussian distributions with a common mean, Eq. (5.4). Similar to the nominal procedure, the simulation samples are fitted to determine fit fraction f_{G} and resolution-factor f_σ.

To test a more flexible – but physically less motivated – background distribution, the fits are carried out with a 3$^{\mathrm{rd}}$ order Chebychev polynomial. Relatively large (up to +0.725) correlations between the second Chebychev coefficient in the X_c mass and the fraction of Λ_c^+ and $\bar{D}^0/D_s^-$ combinatorial background have been observed. That is because the best estimate of that coefficient is negative, describing an inverse parabola, such that it resembles the signal shape.

Next, components with small fit fractions are removed in the signal channel fit. These are labelled minor components in Tab. 9.1, and comprise the $\Lambda_b^0 \to \Sigma_c(2455)^+ \bar{D}^0 K^-$, the combinatorial $\Lambda_c^+ \bar{D}^0$ background and the single charm $\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-$ component. Since the single charm component shares the p.d.f.s in the $\Lambda_c^+ \bar{D}^0 K^-$ and $pK^-\pi^+$ invariant masses with the signal component, there is a sizeable anticorrelation of -0.369, and a larger signal yield when the component is excluded. Similar holds true for the $\Lambda_b^0 \to \Sigma_c(2455)^+ \bar{D}^0 K^-$ component, whose fit fraction is anticorrelated to the partially reconstructed $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ components: -0.332 ($\bar{D}^{*0} \to \bar{D}^0 \gamma$) -0.532 ($\bar{D}^{*0} \to \bar{D}^0 \pi^0$).

On the contrary, partially reconstructed components which have been omitted from the nominal fit are added back for systematic studies. These components (*cf.* Tab. 9.1) have been described in Sec. 7 and are $\Xi_b^- \to \Lambda_c^+ D^{*-} K^-$, $\Lambda_b^0 \to \Sigma_c(2520)^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^*(892)^-$. Their inclusion leads to sizeable anticorrelations between the fit fraction of the added component and the yield of similarly shaped components in the same mass range. Especially the fit fraction of the $\Xi_b^- \to \Lambda_c^+ D^{*-} K^-$ component is strongly anticorrelated to the signal yield (-0.658), and the nominally omitted $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^*(892)^-$ component is anticorrelated to the $\Lambda_b^0 \to \Sigma_c(2455)^+ \bar{D}^0 K^-$ component (-0.614).

The problem of large anticorrelations with the $\Lambda_b^0 \to \Lambda_c^+ [\bar{D}^0 \pi^0]_{\bar{D}^{*0}} K^-$ yield was initially tackled by fixing the ratio of $\bar{D}^{*0}$ partial widths to $R_{D^*} = \Gamma(D^{*0} \to D^0 \pi^0)/\Gamma(D^{*0} \to D^0 \gamma) = 1.83$ [14]. This is now included as a systematic study and leads to a larger fit fraction of the $\bar{D}^{*0} \to \bar{D}^0 \pi^0$ component and simultaneously a lower fit fraction of the $\bar{D}^{*0} \to \bar{D}^0 \gamma$ component. That is because the ratio is estimated to be lower in the nominal fit: $R_{D^*} = 1.67 \pm 0.17$. This corresponds to $\Gamma(D^{*0} \to D^0 \gamma) = 0.375 \pm 0.023$, assuming π^0 and γ decay modes sum up to unity and is in better agreement with former world averages listed in the PDG database (*e.g.* [220]).

Model variation $\Lambda_c^+\bar{D}^0K^-$	ΔN_{sig}	$\Delta N_{\text{PR}\pi^0}$	$\Delta N_{\text{PR}\gamma}$
Double Gaussian signal	−9	+4	−7
Chebychev background	**−22**	**−29**	−61
Excluding minor components	**+26**	+40	**+99**
Including all omitted components	+10	+5	−10
Fix R_{D^*}	−3	**+44**	**−75**

Table 9.1: Difference of signal yields obtained in systematic studies and nominal yields $\Delta N = N_{\text{syst}} - N_{\text{nominal}}$. The abbreviations N_{sig}, $N_{\text{PR}\pi^0}$, $N_{\text{PR}\gamma}$ and R_{D^*} are used for $N_{\Lambda_b^0\to\Lambda_c^+\bar{D}^0K^-}$, $N_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\pi^0]_{\bar{D}^{*0}}K^-}$, $N_{\Lambda_b^0\to\Lambda_c^+[\bar{D}^0\gamma]_{\bar{D}^{*0}}K^-}$ and $\Gamma(D^{*0}\to D^0\pi^0)/\Gamma(D^{*0}\to D^0\gamma) = 1.83$ respectively. The terms minor and omitted components are detailed in the text. The largest deviations are printed in boldface.

Model variation $\Lambda_c^+D_s^-$	$\Delta N_{\Lambda_b^0\to\Lambda_c^+D_s^-}$
Double Gaussian signal	−9
Chebychev background	−144
Excluding $\Lambda_b^0\to\Lambda_c^+K^+K^-\pi^-$	+176
Including combinatorial $\Lambda_c^+D_s^-$ background	+38

Table 9.2: Difference of $\Lambda_b^0\to\Lambda_c^+D_s^-$ yields from systematic studies compared to nominal yields $\Delta N = N_{\text{syst}} - N_{\text{nominal}}$.

In order to estimate systematic uncertainties, models with large deviations from the nominal yield – printed bold in Tab. 9.1 – are propagated through the analysis chain with a "matching" $\Lambda_c^+D_s^-$ model, as tabulated in Tab. 9.3. Not surprisingly, the largest systematic uncertainty comes from excluding insignificant components. In particular for the $\Lambda_b^0\to\Lambda_c^+\bar{D}^{*0}K^-$ decay because there are no components in the $\Lambda_c^+D_s^-$ fit model which could compensate the effect. For the $\Lambda_b^0\to\Lambda_c^+\bar{D}^0K^-$ signal such a compensation is observed for the exclusion of the single charm component.

The deviations from the nominal yields mostly cancel in the ratio of efficiency corrected yields for the Chebychev background model. The effect is larger for the $\Lambda_b^0\to\Lambda_c^+\bar{D}^{*0}K^-$ branching fraction, because the partially reconstructed components cover a wider invariant mass range, in which a flexible background has a larger impact. Fixing R_{D^*} in the fit model yields large deviations in the individual partially reconstructed yields; but these cancel in the combination of the $\Lambda_b^0\to\Lambda_c^+\bar{D}^{*0}K^-$ branching fraction. The largest positive and negative deviations are taken as upper and lower systematic uncertainty.

$\Lambda_c^+\bar{D}^0K^-$ model	$\Lambda_c^+D_s^-$ model	$\Delta\mathscr{R}_{\Lambda_b^0}$ [%]	$\Delta\mathscr{R}^*_{\Lambda_b^0}$ [%]
Excluding minor components	Excluding $\Lambda_b^0\to\Lambda_c^+K^+K^-\pi^-$	+0.054	+1.027
Fix R_{D^*}	nominal	−0.016	−0.331
Chebychev background	Chebychev background	−0.013	−0.450

Table 9.3: Impact of fit systematics on the ratio of branching fractions $\mathscr{R}$ (*cf.* Eq. (3.1)).

9.2 Binning of calibration variables

In this work, efficiencies are computed in bins[1] of calibration variables, *cf.* Sec. 8. The outcome of the analysis under systematic variation of the binning schemes of the efficiency lookup tables with the least statistical power is studied here. The study shows that the choice of binning in the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ branching fraction measurement is the single dominating systematic uncertainty. It could be reduced with larger simulation samples and a thorough study of the involved binning schemes. This is not done here, since the measurement is dominated by the statistical uncertainty, and will in future be dominated by the limited knowledge of D^0 and D_s^+ branching fractions.

Efficiency lookup tables with different binning schemes are propagated through the entire analysis pipeline. The procedure is similar to the fit model uncertainties; meaning only *compatible* binning schemes of signal and reference channel lookup tables are evaluated. Compatible implies that signal and reference lookup tables are required to contain a similar amount of events per bin, or a similar number of bins. This is done to emulate the analysts choice. It would for example not make sense to evaluate the branching fraction with a coarse binning scheme in the signal channel and a fine binning scheme in the reference channel or vice versa.

Since the adaptive binning algorithm is fairly primitive, *i.e.* only performs binary splits, the number of "reasonable" binning schemes, balancing systematic bias against statistical uncertainty, is limited. The efficiency lookup table with the fewest "reasonable" binning schemes is the 2011 $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ trigger to offline selection lookup table, with only 1393 truth-matched events passing all reconstruction and selection steps. Since there is no strong correlation between p_T and rapidity in $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and since binary splits are performed, the number of bins are powers of two. In this specific case, binning schemes with 64 and 32 bins are considered "reasonable", containing about 22 or 44 reconstructed, selected and truth-matched events per bin. Clearly, a lower number of events per bin would limit the statistical precision of the efficiency estimate, while a coarser binning can lead to a bias, since the efficiency gradient (mainly from low to high p_T, *cf.* Fig. 8.3(g)) is not reproduced.

Here, two "reasonable" binning schemes for 2011, four for 2012 $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ data and five schemes for each of 2011 and 2012 $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ data are found manually. The numbers differ because of the available simulation statistics. In the future it is planned to automate this procedure by monitoring a figure of merit which would indicate if a binning scheme produces "unreasonable" results. This figure of merit could be the relative uncertainty on the weighted mean of efficiencies in the lookup table.

Even though the number of "reasonable" binning schemes is small for single samples, the fact that efficiency lookup tables are created per year and channel make for larger combinatorics. Eventually, 31 combinations of binning schemes have been chosen. This number is doubled by inverting the axes of calibration variables when creating the binning scheme; such that the adaptive binning algorithm starts to split the input histogram in the median of the y- instead of the x-axis.

Since none of the binning scheme combinations is clearly superior – as opposed to the choice of fit model driven by a physics motivation – the mean of the resulting distribution of relative branching ratios is taken as final result. The standard deviation of this distribution is taken as systematic uncertainty.

[1] Further commonly used methods are re-weighting of simulation/calibration sample, or finding a parametrisation of the efficiency lookup table. In the case of re-weighting, the choice of a binning scheme implicitly enters in the actual method of re-weighting; and can be complex in case of re-weighting with gradient boosted decision trees, where the binning enters in the growing of regression trees, *cf.* Sec. 4.2.5. A parametrisation of the efficiency lookup table is the most desirable, but also most complex case, since a robust parametrisation, whose choice introduces a further uncertainty, has to be found.

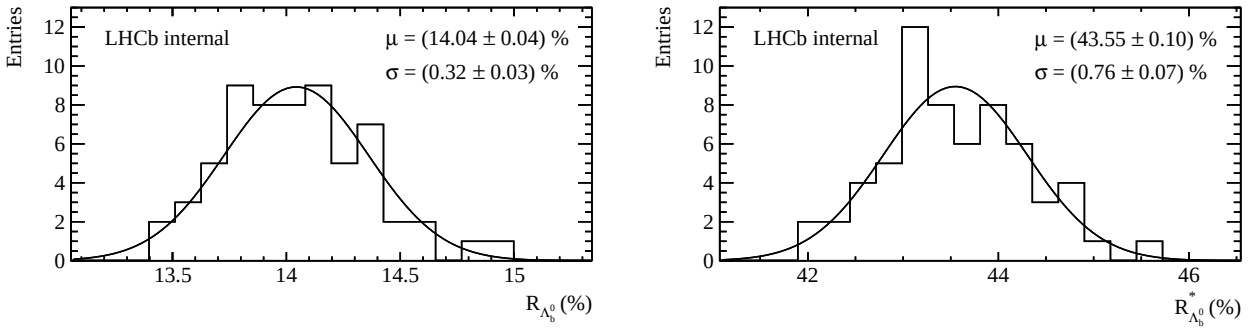


Figure 9.1: Result of the systematic study of binning scheme combinations for the relative branching ratios.

9.3 Calibration and signal sample sizes

Uncertainties arising from the limited size of simulated, D-from-B BDT and PIDCalib samples are evaluated. Further, marginal sample sizes differences of signal and reference channel, due to processing errors from distributed computing when in the production of datasets for offline analysis, is taken into account. These uncertainties are propagated to the efficiency corrected yields, Eq. (3.2), as

$$\Delta N^{\mathrm{d}}_{\mathrm{corr}} = \sqrt{\sum_i \left(\frac{{}_s\mathscr{P}_{\mathrm{d}}(\boldsymbol{\mu}_i)}{\prod_k \varepsilon_k(\boldsymbol{\omega}_{i,k})} \right)^2 \left[1 + \sum_{k'} \left(\frac{\Delta\varepsilon_{k'}(\boldsymbol{\omega}_{i,k'})}{\varepsilon_{k'}(\boldsymbol{\omega}_{i,k'})} \right)^2 \right]}. \tag{9.1}$$

The first term in this formula is the statistical uncertainty using the asymptotic properties of sWeights $\Delta N_{\mathrm{d,asymptotic}} = \sqrt{\sum_i ({}_s\mathscr{P}_{\mathrm{d}}(\boldsymbol{\mu}_{i,\mathrm{d}}))^2}$, and the second term contains the uncertainties on the efficiencies in each bin of lookup table k'. These are the statistical uncertainties of simulation and calibration sample sizes, which are evaluated here. The uncertainty attributed to the efficiencies, $\Delta\varepsilon_{k'}(\boldsymbol{\omega}_{i,k'})$, is evaluated per efficiency lookup table-bin using the 68.27 % Wilson confidence interval as described in Sec. 8. Differences in signal and reference channel sample sizes are corrected for by evaluating the ratio of integrated luminosities. Their uncertainties are propagated.

The results from calibration and signal sample size uncertainties are shown in Tab. 9.4, which summarises all systematic uncertainties affecting the relative branching fraction measurement of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$. The uncertainties from the efficiency lookup table binning and the fit model uncertainty for the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ branching fraction measurement dominate.

Type of uncertainty	$\Delta\mathscr{R}_{\Lambda_b^0}$ [%]	$\Delta\mathscr{R}^*_{\Lambda_b^0}$ [%]
Fit model	$^{+0.535}_{-0.159}$	$^{+10.265}_{-4.495}$
Binning	±3.173	±5.682
Simulation sample size	±0.794	±1.958
BDT calibration sample size	±0.082	±0.311
PID calibration sample size	±0.061	±0.215
Relative data sample sizes	±0.116	±0.361

Table 9.4: Systematic uncertainties on the relative branching ratios $\mathscr{R}^{(*)}_{\Lambda_b^0}$.

10 Results

The last steps to obtain the results for the relative branching fractions $\mathscr{R}^{(*)}_{\Lambda_b^0} = \frac{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-)}{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}$ are detailed. Further, implications of the branching ratio measurement are discussed and a comparison with similar decays from b-mesons is made.

10.1 Correction of statistical uncertainties

Statistical uncertainties are propagated from the invariant mass fits to the master formula (3.1) using the asymptotic properties of sWeights. The problem with this procedure is, that the effect of nuisance parameters on the signal yield is not propagated. That is, because these are fixed to their best fit estimates when using the RooStats implementation of the ${}_s\mathscr{P}lot$ technique, which follows the recommendation given in the original ${}_s\mathscr{P}lot$ publication [196].

The correction uses the difference of yield uncertainties from the nominal and a subsequent fit, where shape parameters have been fixed. This difference is added in quadrature to the uncertainty of the sWeights in the asymptotic limit, $\Delta N_{\mathrm{d,asymptotic}} = \sqrt{\sum_i ({}_s\mathscr{P}_\mathrm{d}(\boldsymbol{\mu}_{i,\mathrm{d}}))^2}$, such that

$$\Delta N_\mathrm{d} = \sqrt{\sum_i \left({}_s\mathscr{P}_\mathrm{d}(\boldsymbol{\mu}_{i,\mathrm{d}})\right)^2 + \left(\Delta N_\mathrm{d}^2 - \Delta N^2_{\mathrm{d,fixed\ shapes}}\right)}\,. \tag{10.1}$$

The formula has been derived in Ref. [221] and is validated with a toy Monte Carlo study in appendix D. Propagating ΔN_d to the uncertainty on the efficiency corrected yield (9.1) leaves

$$\Delta N_{\mathrm{d,corr}} = \sqrt{\sum_i \left(\frac{{}_s\mathscr{P}_\mathrm{d}(\boldsymbol{\mu}_{i,\mathrm{d}})}{\prod_k \varepsilon_k(\boldsymbol{\omega}_{i,k})}\right)^2 \left[1 + \sum_{k'}\left(\frac{\Delta\varepsilon_{k'}(\boldsymbol{\omega}_{i,k'})}{\varepsilon_{k'}(\boldsymbol{\omega}_{i,k'})}\right)^2\right] + \frac{N^2_{\mathrm{d,corr}}}{N^2_\mathrm{d}}\left(\Delta N_\mathrm{d}^2 - \Delta N^2_{\mathrm{d,fixed\ shapes}}\right)}\,. \tag{10.2}$$

This formula allows for separate treatment of the uncertainties due to the ${}_s\mathscr{P}lot$ formalism, the here introduced correctionterm and the simulation and calibration channel statistics, which enter as uncertainties in each bin of the respective efficiency lookup table. The latter uncertainty has already been accounted for in Sec. 9.3. So the statistical uncertainty on the efficiency corrected yields, and thus $\mathscr{R}^{(*)}_{\Lambda_b^0}$, is given by the ${}_s\mathscr{P}lot$ formalism- and correction-terms.

10.2 Combination of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ decay modes

Since different efficiencies for $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\pi^0\right]_{\bar{D}^{*0}} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\gamma\right]_{\bar{D}^{*0}} K^-$ have been observed (*cf.* sec. 8.1), it was decided to treat the two $\bar{D}^{*0}$ decay modes separately, and only combine them in the final step of the analysis workflow. This means that individual relative branching fractions for $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\pi^0\right]_{\bar{D}^{*0}} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\gamma\right]_{\bar{D}^{*0}} K^-$ with respect to $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ are computed and added for the final result, assuming that both $\bar{D}^{*0}$ decay modes sum up to 1.

Several exceptions have to be taken into account in the combination. First, the correlation of $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\pi^0\right]_{\bar{D}^{*0}} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\gamma\right]_{\bar{D}^{*0}} K^-$ fit fractions have to be taken into account. This correlation is -0.239, and leads to a smaller combined uncertainty compared to an uncorrelated propagation of uncertainty. Next, systematic uncertainties have to be combined according to method and calibration sample.

Systematic uncertainties due to the fit model are propagated through the whole analysis chain, and the difference between each studied combined relative branching fraction ($\mathscr{R}^*_{\Lambda_b^0,\text{fitsyst}}$) and the nominal combined relative branching fraction ($\mathscr{R}^*_{\Lambda_b^0,\text{nominal}}$) is computed in a dedicated step before the actual combination step. The largest deviation is taken as systematic uncertainty.

Combined binning scheme uncertainties can be obtained in two ways. One would be similar to the combination of fit model systematic uncertainties. This is realised in Sec. 9.2, where the combined relative branching fractions are filled into a histogram and fitted. The other method leads to the same result and fills the individual relative $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\pi^0\right]_{\bar{D}^{*0}} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \left[\bar{D}^0\gamma\right]_{\bar{D}^{*0}} K^-$ branching fractions into a histogram and combines mean values and widths by propagation of uncertainty without correlation; since separate simulation samples are used. For the same reason, the simulation sample size uncertainty is added in quadrature as well.

The combination of *D-from-B* BDT systematic uncertainties is done assuming 100 % correlation. This conservative assumption holds to first order; it would be reduced by the normalised integral of non-overlapping distributions in the BDT calibration variables log(FD χ^2) and `nTracks`. The same consideration is true for the systematic uncertainty due to the PID calibration sample size.

Eventually, the uncertainty due to the $D_s^-/\bar{D}^0$ branching fraction ratio and the uncertainty due to the luminosities of signal and reference channels are treated as fully correlated, *i.e.* the relative uncertainties, which are equal for both $\bar{D}^{*0}$ channels, are conserved.

10.3 Results for the relative branching fractions

The resulting relative branching fractions amount to

$$\mathscr{R}_{\Lambda_b^0} = \frac{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-)}{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = (14.04 \pm 0.58 \pm 0.33 \pm 0.45)\,\%$$

$$\mathscr{R}^*_{\Lambda_b^0} = \frac{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-)}{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = \left(43.5 \pm 1.4\,^{+1.2}_{-0.8} \pm 1.4\right)\,\%\,,$$

where the first uncertainty is statistical, the second uncertainty is systematic and the third is due to the knowledge of the branching fractions of $D_s^- \to K^-K^+\pi^-$ and $\bar{D}^0 \to K^+\pi^-$.

10.4 Implications and comparison to similar *b*-hadron decays

The decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ is a two body tree-level decay, shown in Fig. 10.1(a), that has been observed by LHCb [174]. The decays $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ are three-body decays and are colour suppressed with respect to $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ (*cf.* Fig. 10.1(b)), but can as well proceed via a doubly colour suppressed diagram, as shown in Fig. 10.1(c) which is not possible for the $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ decay. The naïve expectation from colour factors of the diagrams in Fig. 10.1, $\frac{1}{9} + \frac{1}{81} \approx 12.3\,\%$, is close to above measured value of $\mathscr{R}_{\Lambda_b^0}$. However, it could not be expected that this crude approximation holds true to this level of precision, since $\mathscr{R}_{\Lambda_b^0}$ depends in addition on the available phase-space and the hadronisation fractions of the virtual W to D_s^- or $\bar{D}^0 K^-$ [222].

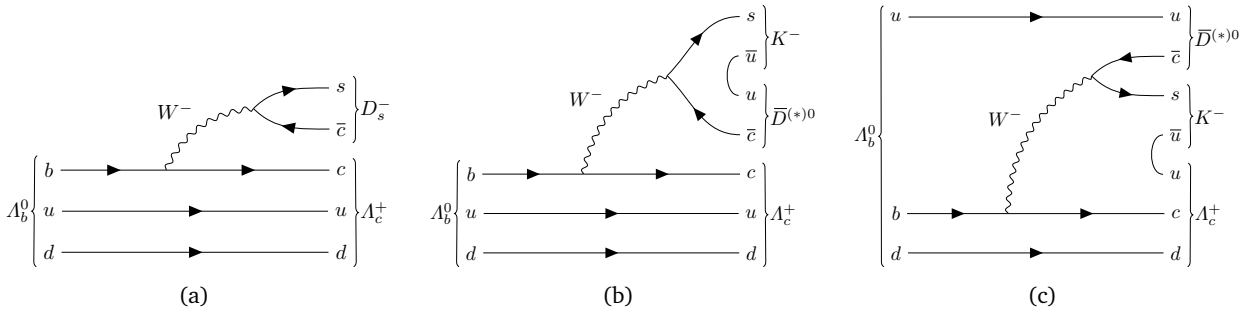


Figure 10.1: Tree-level Feynman diagrams of (a) $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ and (b),(c) $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$.

The results obtained are compared to similar mesonic $X_b \to X_c \bar{D}^{(*)0} K^-$ decays. In such decays, the ud quark pair of the Λ_b^0 are replaced by one antiquark: $\bar{u}$, $\bar{d}$, $\bar{s}$ or $\bar{c}$. It is expected that the double ratio $\mathcal{R}^{(*)}_{\Lambda_b^0} / \mathcal{R}^{(*)}_{X_b} = \left[\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} \right] \Big/ \left[\frac{\mathcal{B}(X_b \to X_c \bar{D}^{(*)0} K^-)}{\mathcal{B}(X_b \to X_c D_s^-)} \right]$ is close to unity, with small deviations due to the available phase space. The results of this comparison are shown in Tab. 10.1 with the results of this work included. It is seen that the assumption holds within uncertainties.

The values for $\mathcal{R}^{(*)}$ have been calculated taking the ratio of the respective branching fractions from the PDG database [14]. Their uncertainties are obtained by error-propagation using the uncertainties on the world averages. The world-averages of the reference decays $\bar{B}^0 \to D^+ D_s^-$ and $B^- \to D^0 D_s^-$ are dominated by measurements of the B-factories. The branching fraction of the decays $\bar{B}^0 \to D^+ \bar{D}^{(*)0} K^-$ is determined in a BaBar measurement [223], while the same measurement dominates the world average of the $B^- \to D^0 \bar{D}^{(*)0} K^-$ branching fraction. For the B_c^- decays, only the LHCb measurements of $\frac{\mathcal{B}(B_c^- \to J/\psi D_s^-)}{\mathcal{B}(B_c^- \to J/\psi \pi^-)}$ [224], $\frac{\mathcal{B}(B_c^- \to J/\psi \bar{D}^0 K^-)}{\mathcal{B}(B_c^- \to J/\psi \pi^-)}$ and $\frac{\mathcal{B}(B_c^- \to J/\psi \bar{D}^{*0} K^-)}{\mathcal{B}(B_c^- \to J/\psi \bar{D}^0 K^-)}$ [225] were taken into account.

Decay	Q value	$\mathcal{R}_{X_b}$	$\mathcal{R}^*_{X_b}$
$\bar{B}^0 \to D^+ \bar{D}^{(*)0} K^-$	3410 MeV	$14.9 \pm 2.3\,\%$	$48 \pm 7\,\%$
$B^- \to D^0 \bar{D}^{(*)0} K^-$	3414 MeV	$16 \pm 4\,\%$	$70 \pm 9\,\%$
$\bar{B}_s^0 \to D_s^+ \bar{D}^{(*)0} K^-$	3398 MeV	no data	
$\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$	3333 MeV	$14.0 \pm 0.8\,\%$	$43.5^{+2.3}_{-2.1}\,\%$
$B_c^- \to J/\psi \bar{D}^{(*)0} K^-$	3179 MeV	$15 \pm 6\,\%$	$80 \pm 40\,\%$

Table 10.1: Summary of $X_b \to X_c \bar{D}^{(*)0} K^-$ decays. The energy release (Q value) being available for the virtual W decay is given by $M(X_b) - M(X_c)$, in accordance to [222].

More general implications of the here presented measurement are discussed in the following. Both $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ are tree-level $b \to c\bar{c}s$ transitions. For Λ_b^0 decays, these transitions are dominated by the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^- X$, where X should be read as additional pions or excitations of Λ_c^+ or D_s^-. The inclusive decay rate $b \to c\bar{c}s$ is an essential ingredient for model- and decay channel independent search for new physics and was estimated to be $23 \pm 2\,\%$ [226].

A lower rate could be expected in Λ_b^0 decays, due to the internal W capture mechanism of the form $bu \to cd$. These contributions would be non-factorising [227] and lead to an enhancement of DN final states [222]. On the other hand, if u-quark and d-quark in the Λ_b^0 are treated as (ud) diquark, the Λ_b^0 could be treated similar to a meson. This implies that the diquark can be treated as spectator in lowest order approximation of factorisation approaches [228, 229].

This can be tested here, using the estimated absolute $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ branching fraction [174]

$$\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-) = 1.1 \pm 0.1\,\% \,.$$

The absolute branching fractions of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ would then be given by

$$\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-) = (1.5 \pm 0.2) \cdot 10^{-3} \quad \text{and} \quad \mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-) = (4.8 \pm 0.5) \cdot 10^{-3} \,.$$

These can be compared to

$$\mathscr{B}(\bar{B}^0 \to D^+ \bar{D}^0 K^-) = (1.07 \pm 0.07 \pm 0.09) \cdot 10^{-3}, \mathscr{B}(\bar{B}^0 \to D^+ \bar{D}^{*0} K^-) = (3.46 \pm 0.18 \pm 0.37) \cdot 10^{-3}$$
$$\text{and} \mathscr{B}(B^- \to D^0 \bar{D}^0 K^-) = (1.45 \pm 0.33) \cdot 10^{-3}, \qquad \mathscr{B}(B^- \to D^0 \bar{D}^{*0} K^-) = (6.32 \pm 0.19 \pm 0.45) \cdot 10^{-3} \,.$$

Hence, both $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-$ branching fractions agree with their mesonic equivalent under the assumption of the lowest order approximation of factorisation.

11 Summary and Outlook

This thesis reports first observation of the decays $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^*(2007)^0 K^-$ and presents the measurement of their branching fraction relative to the decay $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. A dataset corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$ collected at 7 and 8 TeV center-of-mass energies in proton-proton collisions with the LHCb detector has been used. The results obtained are

$$\frac{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-)}{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = (14.04 \pm 0.58 \pm 0.33 \pm 0.45)\,\%$$
$$\frac{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{0*}(2007)^0 K^-)}{\mathscr{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)} = \left(43.5 \pm 1.4\,^{+1.2}_{-0.8} \pm 1.4\right)\,\%\,,$$

where the first uncertainty is statistical, the second uncertainty is systematic and the third is due to the knowledge of the branching fractions of $D_s^- \to K^- K^+ \pi^-$ and $\bar{D}^0 \to K^+ \pi^-$. About 1000 and 3000 $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^*(2007)^0 K^-$ decays have been reconstructed respectively. These signals have been established with an overwhelming statistical significance.

The obtained result agrees with previously measured $B \to D\bar{D}^{(*)0}K^-$ branching fractions, as expected by the lowest order approximation of QCD factorisation, which implies that quarks not participating in the weak process can be treated as spectators. In a wider context, the $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ branching fraction contributes to the total inclusive $b \to c\bar{c}s$ rate. The precise measurement of this rate is an important ingredient in model-independent searches for physics beyond the standard model in B meson decays.

This thesis additionally documents an important stepping stone towards the amplitude analyses of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^*(2007)^0 K^-$, which enables the search for P_c^+ pentaquarks in the $\Lambda_c^+ \bar{D}^0$ and $\Lambda_c^+ \bar{D}^*(2007)^0$ systems. The amplitude analyses profit from the algorithms developed here for signal selection and efficiency correction. On the selection side, a powerful tool to select non-prompt decays of charm hadrons has been developed and calibrated (*cf.* sec. 5). It effectively provides a variable to identify charm-hadron decays and can be applied to any analysis involving secondary charm hadrons. On the efficiency correction side, a flexible algorithm to compute efficiencies as a function of calibration variables has been developed. It interfaces a custom-built class that creates and handles two-dimensional adaptively binned histograms to harness more of the sample's statistical power.

The tools of selection, efficiency correction, and others used throughout the work, have been made available as portable modules. Currently, many of these modules are used in the amplitude analysis of $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$ and the analyses of $\Lambda_b^0 \to J/\psi\,\Lambda\phi$ and $\Lambda_b^0 \to \Lambda_c^+ D^- K^- \pi^+$ decays. In essence, the methods developed here are expected to be integrated and improved in many future analyses and so help to contribute to the programme of LHCb, in particular in the search for pentaquarks. Pursuing such spectroscopic searches will lead to a better effective description of the hadronic spectrum. In the long term, hadron spectroscopy will help to strengthen the link between effective models and the non-perturbative regime of QCD. A better understanding of this sector enables to address fundamental points, such as dynamical chiral symmetry breaking and confinement.

A Binomial confidence intervals in the presence of background

This section discusses the derivation of an estimator for data-driven efficiencies, *i.e.* binomial processes. The problem faced here is the following: A valid binomial confidence interval has to be inferred from data, which contains background. First, well established concepts of binomial confidence intervals in the signal-only case are introduced. The simplest one, the Wald approximation, will then be used to derive an estimator that yields better coverage in the presence of background. It is then shown that the new estimator and the signal-only Wald approximation can be used to scale the input parameters of signal-only estimators with better properties, to be able to use them in the presence of background.

The so derived effective intervals are tested with toy studies that sample from data of the *D-from-B* BDT efficiency lookup tables. Two approaches to generate toy data are introduced and compared. It could be shown that the modified confidence intervals provide better coverage than the signal-only intervals, but that both toy-generation approaches have shortcomings which are qualitatively understood. However, small undercoverages were still observed, so that a conservative approach is taken and the measured uncertainties on data will be artificially inflated by a factor corresponding to the largest measured deviation for each *D-from-B* BDT efficiency lookup table.

A.1 Confidence intervals for binomial processes

The process of selecting signal events from the sample of recorded data is a series of Bernoulli trials, whose probability measure is modelled by the Binomial distribution. Because the binomial distribution is a discrete probability distribution and difficult to calculate for large numbers of trials, a variety of approximations to estimate confidence intervals exist.

The simplest approximation is the *normal* – because the estimator is approximated by a normal distribution – or *Wald approximation*. It is given by

$$\hat{\varepsilon} \pm z \sqrt{\frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{n}} \,, \tag{A.1}$$

where $\hat{\varepsilon}$ is the estimate of the binomial probability (the anticipated efficiency), z is the $1-\alpha/2$ quantile of the standard normal distribution and n the number of trials. The interval can be derived from the central limit theorem, which is used later to derive the new estimator. But even in the large sample limit is known to have invalid coverage.

Another approximation that is derived from the central limit theorem is the Wilson interval [230]. The difference to the Wald interval is that Wilson "inverted the hypothesis test". This means, that he took the range of values θ_0 as confidence interval for which the hypothesis $\theta_0 = \theta^*$ would not be rejected at significance level α. For the upper endpoint, one uses the largest value ε_u such that $\varepsilon_\mathrm{u} - z\sqrt{\varepsilon_\mathrm{u}(1-\varepsilon_\mathrm{u})/n}$ contains $\hat{\varepsilon}$. The analogous done for the lower limit, such that the solutions for the resulting quadratic equations give the confidence interval

$$\frac{\hat{\varepsilon} + \frac{z^2}{2n}}{1+\frac{z^2}{n}} \pm \frac{z}{1+\frac{z^2}{n}} \sqrt{\frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{n} + \frac{z^2}{4n^2}} \,. \tag{A.2}$$

This interval is known to have asymptotically good, but invalid coverage properties.

A valid (for that reason recommended by the PDG) but mostly overcovering interval has been derived by Clopper and Pearson [195]. The interval does not use the asymptotic approximation but is rather based on the cumulative probabilities of the binomial distribution. In practice, the Clopper-Pearson

interval is calculated using quantiles of the beta-distribution in ROOT [C37]. Properties of the intervals have been discussed *e.g.* in Refs. [231, 232].

A.2 Modification of the Wald approximation

Extracting data-driven efficiencies is a challenging task. There are several sources of potential systematic uncertainties. The presence of combinatorial background, non-factorisation of signal and background p.d.f.s as a function of calibration variables, statistical fluctuations and nuisance parameters due to unknown p.d.f. shapes.

This section introduces an effective modification of the Wald approximation, and its implementation into the custom-built wrapper package for ROOFIT used in this thesis. Initially, the efficiency fitter has been a standalone tool called `effi` [C13], which has now been integrated into the `beef` package [C16].

A.2.1 Model setup and notation

The algorithm uses simultaneous extended maximum likelihood fits to "pass" and "fail" categories of a sample. The yields of these categories are estimates of the number of candidates passing or failing the probed selection criterion. In order to extract signal and background efficiencies, the fit has to be parametrised in the form

$$f(\boldsymbol{\mu};\boldsymbol{\theta}) = S f_s(\boldsymbol{\mu};\boldsymbol{\theta}_s) + B f_b(\boldsymbol{\mu};\boldsymbol{\theta}_b)\,, \tag{A.3}$$

where f denotes a p.d.f., $\boldsymbol{\mu}$ the observable(s), $\boldsymbol{\theta} = (S, B, \boldsymbol{\theta}_s, \boldsymbol{\theta}_b)$ the fit parameters with the number of signal and background events S, B. The model is defined in a config-file that is read in by the algorithm at runtime. The simultaneous p.d.f. is built automatically taking into account a configurable set of parameters which are split between "pass" and "fail" categories.

With a model in this form, the Wald approximation is modified. The 68.27 % confidence interval of the Wald estimator is given by

$$\hat{\sigma}(\hat{\varepsilon}) = \sqrt{\frac{\hat{\varepsilon}(1-\hat{\varepsilon})}{n}}\,, \tag{A.4}$$

where $\hat{\varepsilon}$ is the estimated efficiency, and n is the total number of independent trials in a binomial process. The Wald approximation makes use of the central limit theorem in which n is the mean and variance of a Poisson-distributed random variable. In the outlined case of data-driven efficiency-estimation, the random variable will not be Poisson-distributed due to nuisance parameters.

The fit will however provide estimators of mean and variance of all relevant random variables. For measuring a signal efficiency, the model in Eq. (A.3) is split into "pass" and "fail" categories, where the parameter S becomes S_{pass} and S_{fail}. For simplifying the notation, the random variables corresponding to S_{pass} and S_{fail} are denoted by P and F with true expected values $\mathbb{E}(P) := \mu_P^*$, $\mathbb{E}(F) := \mu_F^*$. Best estimates returned by the fit are denoted by $\widehat{\mathbb{E}}(P) := \hat{p}$, $\widehat{\mathbb{E}}(F) := \hat{f}$ and $\widehat{\mathbb{V}}(P) := \widehat{\Delta p}^2$, $\widehat{\mathbb{V}}(F) := \widehat{\Delta f}^2$. The fit does not know that it should measure a binomial process, so that the estimated efficiency is written as a function of two independent random variables

$$\varepsilon(P,F) = \frac{P}{P+F}\,. \tag{A.5}$$

This is the implementation in the code, and has the advantage, that P and F are orthogonal, *i.e.* do not share events. Equivalently, "total" T and "pass" P categories could be defined

$$\varepsilon(T,P) = \frac{P}{T}\,, \tag{A.6}$$

This form of the efficiency will now be used to derive the Wald approximation.

A.2.2 Derivation of the Wald approximation

In the following, the Wald approximation will be derived as limiting case from the propagation of uncertainties. The definition of variance for any bivariate function $f(X,Y)$ is

$$\mathbb{V}[f(X,Y)] = \mathbb{E}\left[(f(X,Y) - \mathbb{E}[f(X,Y)])^2\right] .$$

The standard way of propagating uncertainties is a Taylor expansion of $f(X,Y)$ around the expected values $\theta^* := (\mu_X^*, \mu_Y^*)$ of X, Y. Up to first order, and using $\mathbb{E}[f(X,Y)] = f(\theta^*)$, this is

$$\begin{aligned}\mathbb{V}[f(X,Y)] &= \mathbb{E}\left[\left(f(\theta^*) + \partial_X f(\theta^*)(X-\mu_X^*) + \partial_Y f(\theta^*)(Y-\mu_Y^*) - f(\theta^*)\right)^2\right] \\ &= \mathbb{E}\left[(\partial_X f(\theta^*))^2 (X-\mu_X^*)^2 + (\partial_Y f(\theta^*))^2 (Y-\mu_Y^*)^2 + 2\partial_X f(\theta^*)\partial_Y f(\theta^*)(X-\mu_X^*)(Y-\mu_Y^*)\right] .\end{aligned}$$

Using variance $\mathbb{V}[X] = \mathbb{E}\left[(X-\mu_X^*)^2\right]$ and covariance $\mathbb{C}[X,Y] = \mathbb{E}\left[(X-\mu_X^*)(Y-\mu_Y^*)\right]$ leads to

$$\mathbb{V}[f(X,Y)] = (\partial_X f(\theta^*))^2 \mathbb{V}[X] + (\partial_Y f(\theta))^2 \mathbb{V}[Y] + 2\,\partial_X f(\theta^*)\partial_Y f(\theta^*)\,\mathbb{C}[X,Y] . \tag{A.7}$$

Up to this point, this is the standard formula for propagation of uncertainty of a bivariate function.

The efficiency defined in Eq. (A.6) will now be plugged into Eq. (A.7) to derive the Wald approximation. Thus, $f(X,Y)$ becomes $\varepsilon(T,P) = P/T$, $\theta^* = (\mu_T^*, \mu_P^*)$, $\partial_T \varepsilon(\theta^*) = -\frac{\mu_P^*}{\mu_T^{*2}}$, $\partial_P \varepsilon(\theta^*) = \frac{1}{\mu_T^*}$ and Eq. (A.7) reads

$$\mathbb{V}[\varepsilon(T,P)] = \frac{\mu_P^{*2}}{\mu_T^{*4}}\mathbb{V}[T] + \frac{1}{\mu_T^{*2}}\mathbb{V}[P] - 2\,\frac{\mu_P^*}{\mu_T^{*3}}\mathbb{C}[T,P] . \tag{A.8}$$

The covariance term can be re-written as

$$\mathbb{C}[T,P] = \mathbb{E}[TP] - \mathbb{E}[T]\mathbb{E}[P] ,$$

where $\mathbb{E}[P] = \mu_\varepsilon^* \mu_T^*$ is the expected value from a binomial distribution, or equivalently a binomial hierarchy model with random variable T. This implies that ε and T are uncorrelated, because $\mathbb{E}[P] = \mathbb{E}[\varepsilon T] = \mathbb{E}[T]\mathbb{E}[T] + \mathbb{C}[\varepsilon, T] = \mu_\varepsilon^* \mu_T^*$. Hence, the covariance of T and P is

$$\mathbb{C}[T,P] = \mu_\varepsilon \mathbb{E}[T^2] - \mathbb{E}[T]\mu_\varepsilon \mathbb{E}[P] = \mu_\varepsilon \mathbb{V}[T] .$$

Inserting back into Eq. (A.8) yields

$$\begin{aligned}\mathbb{V}[\varepsilon(T,P)] &= \frac{\mu_P^{*2}}{\mu_T^{*4}}\mathbb{V}[T] + \frac{1}{\mu_T^{*2}}\mathbb{V}[P] - 2\frac{\mu_P^*}{\mu_T^{*3}}\mu_\varepsilon^* \mathbb{V}[T] \\ &= \frac{1}{\mu_T^{*2}}\mathbb{V}[P] - \frac{\mu_P^{*2}}{\mu_T^{*4}}\mathbb{V}[T] \\ &= \frac{1}{\mu_T^{*2}}\left(\mathbb{V}[P] - \mu_\varepsilon^{*2}\mathbb{V}[T]\right) ,\end{aligned}$$

where $\mu_\varepsilon^* = \mu_P^*/\mu_T^*$ has been used. Letting P and T be Poisson-distributed with expected value and variance of $\hat{t} = \widehat{\Delta t}^2 \to \mu_T^* = \mathbb{V}[T]$, $\hat{p} = \widehat{\Delta p}^2 \to \mu_P^* = \mathbb{V}[P]$ in the asymptotic limit yields the Wald approximation:

$$\begin{aligned}\widehat{\mathbb{V}}[\varepsilon(T,P)] &= \frac{1}{\hat{t}^2}\left(\hat{p} - \hat{\varepsilon}^2 \hat{t}\right) \\ &= \frac{1}{\hat{t}}\left(\hat{\varepsilon} - \hat{\varepsilon}^2\right) .\end{aligned} \tag{A.9}$$

A.2.3 Using the fit estimates

Now, the same strategy will be applied, but taking the best estimates from the fit. This means that P and F are not Poisson-distributed, but where the estimates of expected values $\hat{p}$, $\hat{f}$ and variance $\widehat{\Delta p}^2$, $\widehat{\Delta f}^2$ have been measured.

The actual model is set up to use "pass" and "fail", Eq. (A.5), instead of "total" and "pass". The ingredients for the variance are thus $f(X,Y) = \varepsilon(P,F)$, $\theta = (\hat{p},\hat{f})$, $\partial_P \varepsilon(\theta) = \frac{\hat{f}}{(\hat{p}+\hat{f})^2}$, $\partial_F \varepsilon(\theta) = -\frac{\hat{p}}{(\hat{p}+\hat{f})^2}$, leading to

$$\widehat{\mathbb{V}}[\varepsilon(P,F)] = \frac{1}{(\hat{p}+\hat{f})^4}\left(\hat{f}^2\widehat{\Delta p}^2 + \hat{p}^2\widehat{\Delta f}^2 - 2\,\hat{p}\hat{f}\,\widehat{\mathbb{C}}(P,F)\right) .$$

For the covariance term, P and F are defined as $P = \varepsilon T$ and $F = (1-\varepsilon)T$ using the auxiliary random variable $T \equiv t(P,F) = P + F$. The covariance is approximated as

$$\begin{aligned}\widehat{\mathbb{C}}(P,F) &= \widehat{\mathbb{E}}(PF) - \widehat{\mathbb{E}}(P)\widehat{\mathbb{E}}(F) \\ &= \widehat{\mathbb{E}}[\varepsilon(1-\varepsilon)T^2] - \widehat{\mathbb{E}}(\varepsilon T]\widehat{\mathbb{E}}[(1-\varepsilon)T] \\ &\approx \hat{\varepsilon}(1-\hat{\varepsilon})\left(\widehat{\mathbb{E}}(T^2) - \widehat{\mathbb{E}}(T)^2\right) \\ &= \hat{\varepsilon}(1-\hat{\varepsilon})\widehat{\mathbb{V}}(T) .\end{aligned} \tag{A.10}$$

With this definition of the efficiency, the covariance term unfortunately contains another bivariate variance, leading to a infinite (but convergent) recursive term:

$$\widehat{\mathbb{V}}[t(P,F)] = \widehat{\Delta p}^2 + \widehat{\Delta f}^2 + 2\,\widehat{\mathbb{C}}(P,F) .$$

Here, the propagation of uncertainty (Eq. (A.7)) has been used again.
In the algorithm, the recursion is stopped by an approximation at second order where the asymptotic Poissonian approximation $\widehat{\mathbb{V}}(T) \approx \hat{p} + \hat{f}$ is used, leaving the modified Wald approximation:

$$\widehat{\mathbb{V}}[\varepsilon(P,F)] \approx \frac{\hat{f}^2\widehat{\Delta p}^2 + \hat{p}^2\widehat{\Delta f}^2 - 2\,\hat{p}\hat{f}\left(\hat{\varepsilon}(1-\hat{\varepsilon})\left(\widehat{\Delta p}^2 + \widehat{\Delta f}^2 + 2\,\hat{\varepsilon}(1-\hat{\varepsilon})(\hat{p}+\hat{f})\right)\right)}{(\hat{p}+\hat{f})^4} . \tag{A.11}$$

Measured yields can then be scaled by a ratio of the Wald- and the modified Wald-approximation

$$r_{\text{eff}} = \frac{\hat{\varepsilon}(1-\hat{\varepsilon})/(\hat{p}+\hat{f})}{\widehat{\mathbb{V}}[\varepsilon(P,F)]} \tag{A.12}$$

in order to compute a confidence interval with better coverage properties provided by the `TEfficiency` class in the algorithm. Due to the influence of the nuisance parameters, the scaling typically acts as a damping of the measured yields and can be understood as representation of their effective statistical power.

For limited statistics and efficiencies near the domain endpoints, the Wald approximation is known to strongly undercover. This results in a very small scaling parameter ($r_{\text{eff}} < 0.05$) or strongly skewed confidence intervals ($\frac{\hat{\varepsilon}-\varepsilon_{\text{lower}}}{\varepsilon_{\text{upper}}-\hat{\varepsilon}} \notin \left[\frac{2}{3}, \frac{3}{2}\right]$). In the algorithm this is effectively handled by scaling directly to the lower limit if $\hat{\varepsilon} > 0.5$ or upper limit if $\hat{\varepsilon} < 0.5$ of the desired type of interval[1]. The coverage of these limits has been tested by dedicated standalone [C31] and project-integrated [C15] toy studies, which are detailed in the following.

[1] In a planned update, the scaling would be done to the Wilson interval [230] equivalence of Eq. (A.11)

A.3 Standalone toy Monte Carlo

The derived modification is tested for potential biases and coverage properties. A standalone script that generates toy datasets and passes them to the algorithm has been set up [C31].

The script starts by reading in a model. After the model has been loaded, a fit to a sample that was generated using the starting-parameters of the model is carried out. The covariance matrix of this fit is then used to sample sets of all floating fit-parameters for the individual toy experiments by eigendecomposition. Signal and background datasets are then generated separately by their individual p.d.f.s (*cf.* Eq. (A.3)) using the respective set of parameters from the preceding step. Entries of the so obtained signal and background samples are then rejected by importance sampling. The sampling rate is with the respective true signal or background efficiency. The decision whether an event has been accepted or rejected is stored in a variable that is added to the generated dataset.

Signal and background samples are then merged and passed to the fitting algorithm. The procedure is repeated several hundred times to be able to study bias and coverage. The quantity of interest in these studies is the pull, defined by

$$\text{Pull} = \begin{cases} \frac{\hat{\varepsilon}-\varepsilon^*}{\hat{\varepsilon}-\varepsilon_{\text{lower}}}, & \text{if } \hat{\varepsilon} < \varepsilon^* \\ \frac{\hat{\varepsilon}-\varepsilon^*}{\varepsilon_{\text{upper}}-\hat{\varepsilon}}, & \text{otherwise} \end{cases}$$

The true efficiency ε^* is the aforementioned sampling rate. On the other hand, for each toy experiment, there is a unique sampled efficiency. It will be argued in the following that the sampled efficiency leads to distorted pull distributions which are explicable by the method of sampling, rather than the modification of the interval. The unique sampled efficiency is a random variable that depends on the number of generated signal events and the importance sampling. In the presence of nuisance parameters, the number of generated signal events is not Poissonian any more, effectively increasing the variance of the sampled efficiency.

This happened on the sampling side; on the side of the fit, the estimated (symmetric) variance of the efficiency using the modified Wald approximation, Eq. (A.11), only depends on the central values and measured uncertainties of "pass" and "fail" yields, which are sampled accurately. The effect will therefore be seen in the asymmetry of the confidence interval limits with respect to the central value, or - in terms of Bayesian statistics - higher moments of the efficiency posterior distribution. In the toy study, this is seen as left-right asymmetry around 0 in Fig. A.1, in particular for the scaled Clopper-Pearson and Wilson intervals, where the effect is overcompensated due to the effective scaling, but also less pronounced and sign-flipped in the Wald pull-distributions due to the symmetry of the interval.

In the following, the results of a scan through 15 different sets of signal and background efficiencies ($\varepsilon_{\text{sig}} > \varepsilon_{\text{bkg}}$) will be discussed. The model used for generating toy data and fitting is sum of two Gaussian p.d.f.s with shared mean as signal and a first order Chebychev polynomial as background. The parameters are chosen to be similar to those obtained by a one-dimensional fit to the $\Lambda_c^+ D_s^-$ spectrum in a narrow range around the Λ_b^0 signal. The sample size has been decreased compared to the $\Lambda_c^+ D_s^-$ data to 2000 signal and 3000 background events in order to save CPU time and pronounce the effects of limited statistics.

The results of this study, and also from a variation of the yields, can be summarised by the four pull histograms shown in Fig. A.1. The working point of $\varepsilon_{\text{sig}} = 97\,\%$ and $\varepsilon_{\text{bkg}} = 75\,\%$ was chosen from the scan, because these values are representative for regions of the BDT efficiency lookup tables (*cf.* Sec. 5.6). The pulls show that the Wald approximation drastically undercovers. The bias ($\mu = -0.18 \pm 0.09$) in this plot is likely to be caused by entries out of the plotting range and the asymmetry in the pull's numerator discussed above. This asymmetry is less pronounced in the approximated uncertainty-propagation (Eq. (A.11)) due to larger measured uncertainties. It is overcompensated by the asymmetry of the

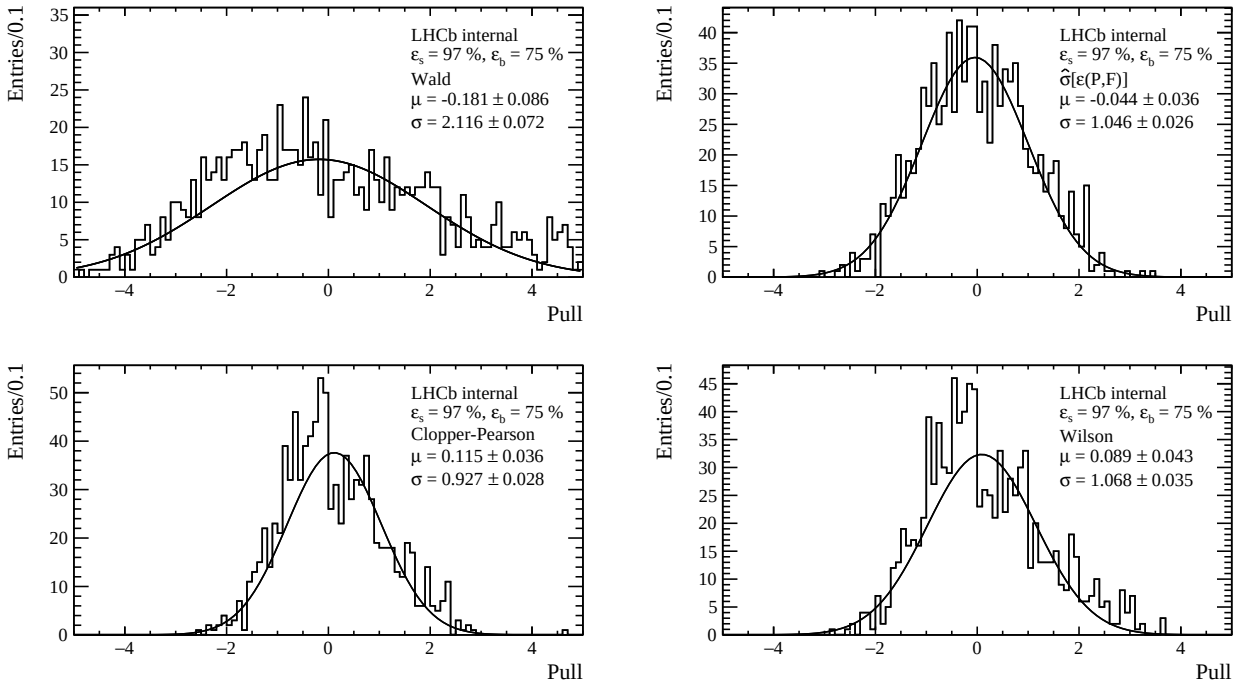


Figure A.1: Comparison of pull distributions obtained by the toy Monte Carlo studies described in the text. The top row shows Wald and modified Wald approximations. The bottom plots show pulls from Clopper-Pearson and Wilson intervals, whose input parameters have been scaled by the effective factor r_{eff} Eq. (A.12).

Clopper-Pearson and Wilson limits due to the scaling, which effectively degrades the input statistics to the calculation. It is seen that the distribution of lower limits (pulls < 0) is more narrow than the distribution for upper limits, i.e. the lower limits are larger than the sampled ones, while the upper limits show good coverage in the Clopper-Pearson case and slightly undercover when using Wilson-limits. To overcome the imperfect sampling, another method, embedded in the *D-from-B* package is studied in the next section.

A.4 Toy Monte Carlo studies for the D-from-B package

The toy Monte Carlo studies described in the previous paragraph will now be used on a fit-result obtained on data from the *D-from-B* efficiency evaluation. This means that the toys discussed below are generated from the measurement directly – by generating "pass" and "fail" samples based on the nuisance parameter sampling – as opposed to generating a "total" signal- and background-dataset and adding "pass" and "fail" categories by importance sampling. Both methods have advantages and disadvantages, which are easiest explained with the help the results from this study.

The toy studies have been carried out in certain bins of the efficiency lookup tables standing out by *e.g.* low statistics, large efficiency uncertainties or large background yields. The studied bins undercovered by about 30 % up to at most 66 % using the above described scaling factor on Clopper-Pearson intervals and the integrated sampling method.

The largest potential source of undercoverage stems from the implementation of the model, because "pass" and "fail" yields are treated as random variables with a covariance that is measured by the fit. In most cases this covariance is much smaller than the expectation from the underlying statistical model, *i.e.* from using the knowledge that "pass" and "fail" datasets are orthogonal.

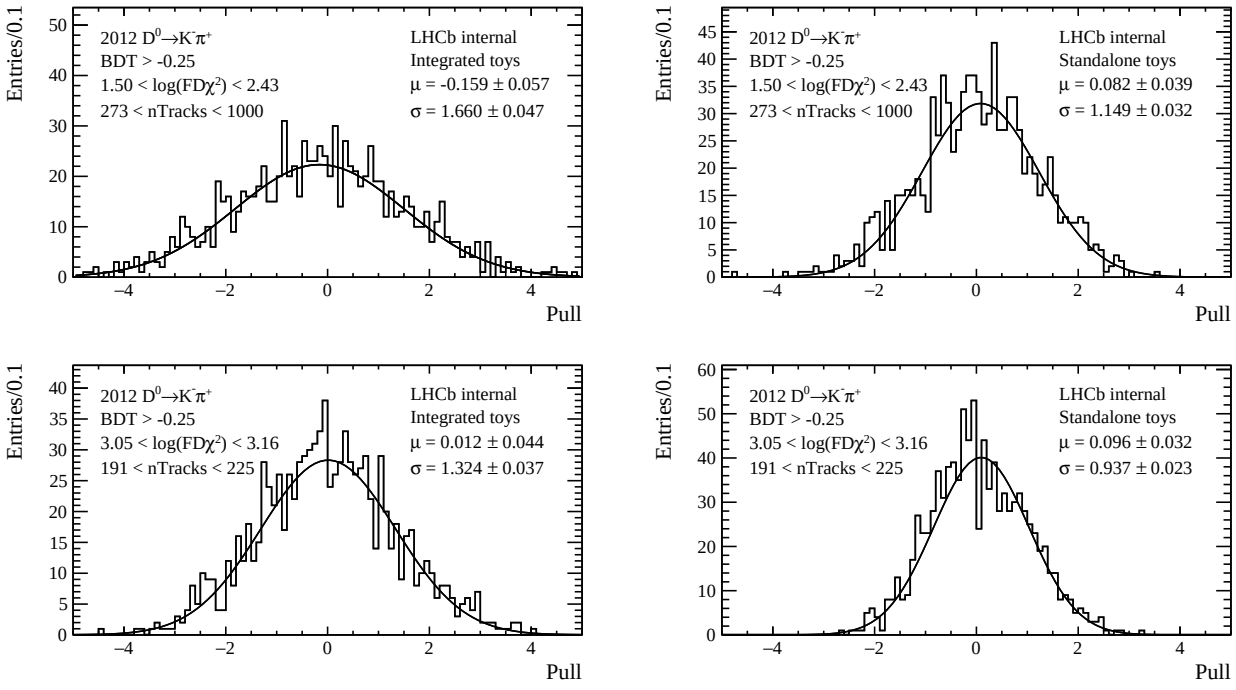


Figure A.2: Pull distributions of the two different toy Monte Carlo approaches studies described in the text.

The situation is further complicated by the complexity of the model, which introduces additional efficiencies for the partially reconstructed and Cabibbo-suppressed backgrounds. The imperfect modelling not only enters in the fit itself, but also when sampling nuisance parameters and generating toy samples. For this reason, it could well be, that the observed undercoverage is due to improperly generated toy samples, and that the effect in the fit itself is not as striking.

The undercovering *D-from-B* toys have been cross-checked with toys from the standalone script, essentially supporting the above statements. A exemplary comparison is shown in Fig. A.2. There, the bins of the efficiency lookup tables, for which the largest undercoverage was measured is shown in the top row. The bottom row shows the bin with the least signal events from the 2012 D^0 efficiency lookup table. A relatively large signal efficiency of $95.61^{+0.56}_{-0.63}$ has been measured in the latter bin, where the asymmetry in the numerator becomes evident again in the standalone toy study.

However, a conservative approach is taken and the measured uncertainties will be artificially inflated by a factor corresponding to the largest measured factor for each efficiency lookup table.

A.5 Approaches to use the efficiency as fit-parameter

In an earlier version of the fitting algorithm it was tried to include the efficiency as fit-parameter. Investigated methods and their shortcomings will be discussed in the following. The method is not used, but the lessons learned here are worthwhile to be discussed briefly.

The strategy was to perform sequential fits to ”total“ and ”pass“ datasets. When fitting the ”pass“ dataset, (additional) constraints to fit-parameters can be applied. One strategy was to constrain shape parameters by a Gaussian p.d.f. using results from the fit to the ”total“ sample and by additionally adding a dependency on the signal and background efficiency. Constraints like

$$G(\theta_p; \mu = \hat{\theta}_t, \sigma = \delta\hat{\theta}_t(1-\hat{\varepsilon})) \tag{A.13}$$

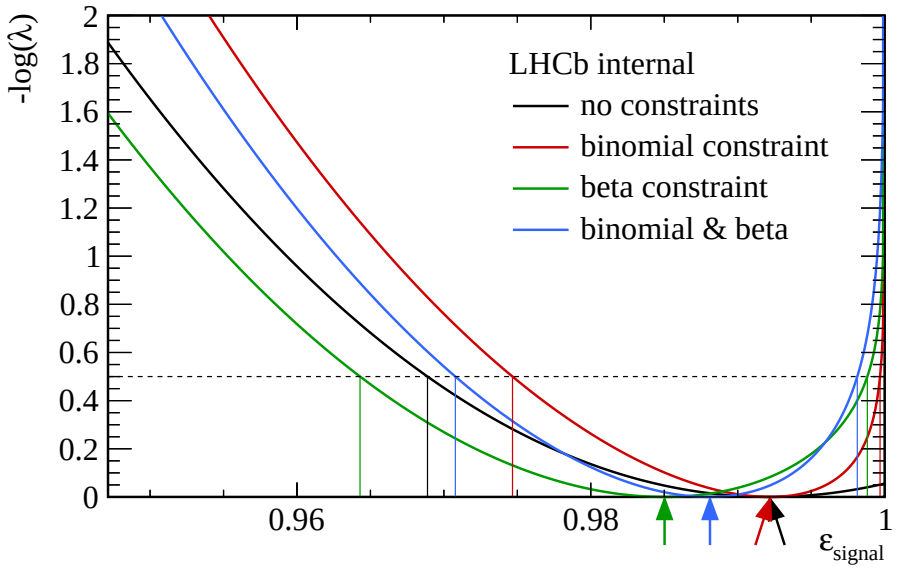


Figure A.3: Profile log-likelihood as function of the signal efficiency. The likelihoods were generated from a fit to the 2012 $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ calibration data in the lowest rightmost bin of the efficiency lookup table. The best fit values for ε are highlighted with arrows. Vertical lines show the estimated uncertainty from the MINOS algorithm [192], nicely in line with the expectation from the profile log-likelihood indicated by the horizontal dashed line.

improved the sensitivity on the resulting efficiency. Here G is a Gaussian p.d.f. in a shape parameter of the "pass" fit θ_p. The mean μ is fixed to the best fit value of the fit to the "total" sample. The width σ penalises values of $\hat{\theta}_p$ which are far from $\hat{\theta}_t$ by a heuristic factor δ. Another potentially superior approach, which has not been realised for technical reasons, would be to directly manipulate or constrain the covariance matrix in the "pass" fit.

Moreover, it was tried to constrain signal and background efficiency parameters. The effect of these constraints on the profile log-likelihood[1] is shown in Fig. A.3. The binomial constraint works as follows: first, a fit to the "pass" sample is carried out and the best fit value of the number of "pass" events $\hat{s}_{\text{pass}} = \hat{\varepsilon} \cdot \hat{s}_{\text{tot}}$ is obtained. In a subsequent fit, again to the "pass" sample, a binomial p.d.f. with fixed yields, at $\hat{p}$, $\hat{t}$, and floating ε is used as external constraint. Fixing the yields at their best fit value results in a too optimistic constraint and therefore undercovers.

It was tried to scale the yields in the binomial p.d.f., which did not result in the desired coverage. If this scaling factor approaches 0, the likelihood tends to the likelihood from the unconstrained fit (the red approaches the black line in Fig. A.3) except near physical boundaries. In Fig. A.3 the scaling factor was set manually to 0.01. Constraints with a beta distribution tend to introduce a bias as shown in Fig. A.3, where Beta$(\varepsilon; \alpha = 1.2, \beta = 1.2)$ was used. On the other hand, beta distribution-constraints can stabilise fits with efficiencies close to 0 or 1.

The uncertainties returned by the MINOS algorithm coincide with the expectations from the profile likelihood function (which is considerably slower to evaluate) in every tested scenario, but it was not possible to obtain valid intervals from this in a range of tests. Also more involved statistical methods from the RooStats package have been tried without success. At an early stage, simultaneous fits including the efficiency as parameter have been examined and tended to be very unstable.

[1] Because of its asymptotic properties, the likelihood function can be used to estimate confidence intervals [191]. Being close to the parameter domain spoils asymptotic normality of the maximum likelihood, leading to wrong predictions for the variance of the estimator (*cf.* Eq. 4.6). A likelihood that is independent of all other parameters of the model but ε, is given by the profile likelihood $L_p = L(\varepsilon, \hat{\hat{\boldsymbol{\nu}}}(\varepsilon))$. The double-hat notation indicates the profiled values of the nuisance parameters $\boldsymbol{\nu}$, defined as the values, which maximise the likelihood for the given ε. The profile likelihood ratio is given by $\lambda(\varepsilon) = \frac{L_p(\varepsilon)}{L(\hat{\varepsilon}, \hat{\boldsymbol{\nu}})}$

B Selection of signal and reference channel

This appendix supplements Sec. 6 with details on the cut optimisation procedure, a discussion on alternative selection strategies and plots of veto cuts.

B.1 Optimisation of binary cuts

The optimisation procedure is done with a custom-built script that automatically scans a range of simple cuts with a defined step-size in the phase space spanned by the set of variables [C6]. A simple cut is defined by a relational operator $>$, $<$, $==$, $!=$, $>=$, $<=$ and a value. The script interfaces the custom-built `beef` wrapper package for fitting. To avoid observer bias, a k-fold cross-validation can be chosen.

The fits are used to calculate a figure of merit at the current position of the scan, which defines a cut-value for each variable (referred to as cutset). The scanning procedure is entirely automated and configurable; its runtime options are provided in a configuration file. The whole scan ends after a given amount of steps. A step is finished if the figure of merit of all possible combinations of cut-values in a given phase space range has been evaluated and a maximum was found. If this maximum has been found at the a boundary of the configured range, it is automatically extended by one step-size if possible.

The phase space range of the next step is then given by the neighbouring points of the current optimum. In addition, the number of additional intermediate cut-values between the new boundaries and the optimal point are configurable. An illustration of this procedure is shown in Fig. B.1.

Each variable can be scanned separately (referred to as 1D scan), as part of the phase space spanned by the set of variables (full scan), or as combination of both. During the 1D scan, the variables are processed successively, such that the order of processing might change the result. Additionally, the cuts of all variables except the one under study are fixed at pre-defined values. These are either given in the configuration file for the first step ($\tilde{x}$), or are set to the optimum that was found by the previous step.

The full scan simply considers every combination of cuts in the phase space range given the step size of each variable, and can thus be pictured as a multidimensional grid with each node representing a cut-set. A full scan requires more fits and suffers from the curse of dimensionality, but is not affected by possible correlations among cuts with respect to the figure of merit, because the cut values are scanned simultaneously.

Signal and background yields for the calculation of the figure of merit are obtained by a properly normalised integral of the signal and background p.d.f.s in a fixed range - the (fixed) signal region.

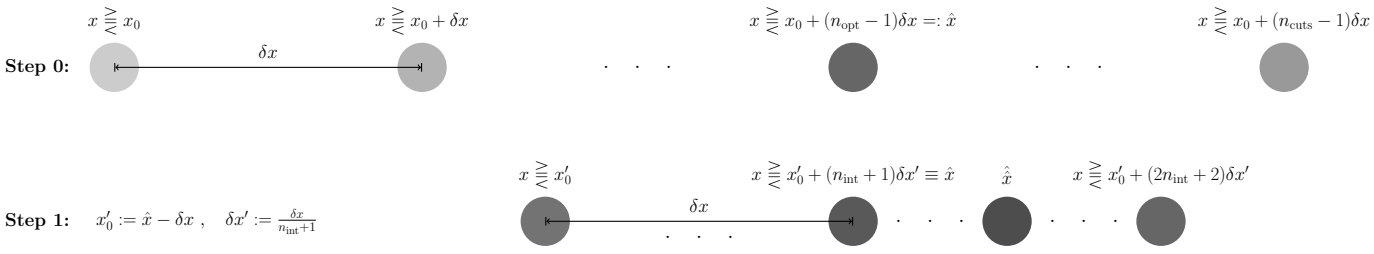


Figure B.1: Sketch of the cut optimisation for a variable x in 1D. The figure of merit is illustrated by the brightness of the points and gets darker for a larger figure of merit. The convention of notation is in-line with the code. The hyperparameters n_{cuts}, n_{int}, δx, x_0, for the number of cuts, the number of intermediate cuts in subsequent steps, the step-size and the loosest cut respectively, as well as the relational operator are configurable. Optimal cut values for step 0 and 1 are represented by $\hat{x}$ and $\hat{\hat{x}}$ respectively.

Alternatively, the range

$$[\hat{\mu} - n_\sigma \cdot \hat{\sigma}, \hat{\mu} + n_\sigma \cdot \hat{\sigma}] \ , \tag{B.1}$$

with signal mean $\hat{\mu}$, width/resolution $\hat{\sigma}$ and a configurable parameter n_σ, defines the (floating) signal region of each fit individually. The hyperparameter settings for the signal channel optimisation can be looked up in Ref. [C23]. It was found that the figure of merit is almost constant over a relatively wide range of cuts, resulting in a negligible effect from the choice of the final set of cuts.

B.2 Discussion on offline selection strategies

In order to achieve the best possible sensitivity, the final selection of events has to be optimised. The optimal working point is obtained by maximising a suitable figure of merit, which for all studies described here was chosen to be $S/\sqrt{(S+B)}$. Additionally, the optimisation must not bias the final result, which could be the case if the optimisation is conducted directly on data. The risk in such a scenario is that the miss-classification rate of background events in the signal region as signal (type I error) is larger than what is estimated by the efficiency correction, leading to an over-estimated branching fraction. The signal optimisation strategy is now carefully chosen such that both criteria – high sensitivity and an unbiased result – are fulfilled.

Several techniques to find this optimal working point have been examined and are summarised here. Eventually, a relatively simple k-fold cross-validated optimisation of binary cuts on the *D-from-B* BDT responses and `ProbNNk` of the bachelor kaon was chosen. The gain in sensitivity from more sophisticated selection methods involving multivariate classification tools has been too low to justify its use. Especially in spite of additional uncertainties introduced by the computation of the classifier response cut efficiency and the overhead of systematic studies, contributing additional uncertainties.

Separating the final selection step into three PID-like variables and the kinematic MVA has not been the prime choice, since a multivariate analysis including all inputs would have potentially been superior to the described rectangular cut scan. When training a classification algorithm, several challenges have to be met to be able to successfully apply it to data. Concerning the performance of the classifier, hyperparameter-settings and input variables (including potentially performance-enhancing transformations) have to be chosen carefully. More critical challenges, in the sense of whether a classifier is eventually applicable to real physics data, are the choices of (labelled) input data to the multivariate algorithm. Probably the most common choice is the use of simulation as signal and sideband data as background input. Further options that have been examined are discussed in the following.

- Training a classifier on signal data – avoiding bias by cross-validation techniques – is ruled out; mainly because the statistics for training is too low. The problem is further complicated by the fact that the distributions of the input variables would have been obtained using the ${}_s\mathscr{P}lot$ technique, for which the p.d.f. factorisation discussed in Sec. 4.2.4 has to be studied. After any reasonable pre-selection to calculate sWeights, there is still a large number of background events, which enhances small non-factorisable effects, manifesting itself in distributions with negative bin contents. For this analysis this would practically mean to train a classifier on maximally 1200[1] events, and using input variables with regions where the distribution becomes negative. Despite the fact that the ${}_s\mathscr{P}lot$ technique breaks down in such a scenario, it was found that no classifier was able to produce a convincing result with such input.

- It was tried to generate the classifier input with the help of a (conditional) generative adversarial network [233, 234] on sWeighted data. Similar issues as described for training directly on data forced the use of a different strategy.

[1] with 1200 being the sum of sweights

- A simpler way to generate signal input data is to sample the variables from sWeighted data. If all considered variables would have been taken into account, the sampling would have corresponded to a 26-dimensional Monte Carlo integration – a computationally too expensive task with unknown chances of success. Instead, all variables have been sampled from spline interpolated data individually. The splines have been obtained from smoothed sWeighted histograms. The smoothing became necessary to flatten out statistical fluctuations in sparsely populated regions, which eventually gets picked up in the training of the classifier. The compatibility of the sampled data with its sWeighted input was tested with Kolmogorov-Smirnov tests for each variable. The sum of Kolmogorov-Smirnov test-scores served as a qualitative measure for the 1D compatibility of sampled and real data and was used to adjust parameters in the used algorithms.

- The best results have been achieved by training a classifier using only kinematic variables from simulation and background data from the Λ_b^0 sidebands.
  The sideband regions, $5560 < M_{\text{inv}}(\Lambda_c^+ \overline{D}^0 K^-) < 5595$ and $5645 < M_{\text{inv}}(\Lambda_c^+ \overline{D}^0 K^-) < 5680$ are shown in Fig. B.2. A cut on the resulting classifier-response was then chosen together with the remaining *D-from-B* BDT and kaon PID variables in a 4D cut optimisation procedure as described above. The variables used for training are listed in table B.1. Kolmogorov-Smirnov test-scores were evaluated for simulated data w.r.t. sWeighted data. The resulting scores were significantly lower compared to spline sampled data described above. It was tried to improve the compatibility of simulation and data by re-weighting variables that deviated most. A re-weighting in one and two dimension resulted in better accumulated Kolmogorov-Smirnov scores, but only marginal differences were seen in the final result (compatible with uncertainties from the following optimisation scan). The receiver operator characteristic (ROC) curves from the here described training are shown in Fig. B.5(a). Note that due to differences in data and MC, a higher area under curve does not necessarily correspond to a better figure of merit when applying the classifier to data. An interesting observation is made in Fig. B.5(b), showing a training where a classifier was trained to distinguish MC and spline sampled from sWeighted data. It shows that the re-weighting only leads to a marginal improvement. In contrast, it clearly seen how large the impact of correlations is for this data. The classifier can easily distinguish between sWeighted data and 1D spline sampled distributions, which agree better when judging from projections and the Kolmogorov-Smirnov test-scores only.

The strategy described in the last bullet-point has been studied in greater detail, such that also hyperparameters and input variable transformations were examined. Contrary to the *D-from-B* package, it was found that the recently integrated class of deep neural networks into TMVA resulted in slightly higher ROC area-under-curves than a BDT. Still, bagged gradient BDTs with a large number of trees and a small learning rate have been chosen here as baseline solution.

The 1D projections of sweighted data and simulation of the input parameters as well as the resulting classifier response have been monitored throughout the studies. Some of the 22 classifier input variables, as well as the classifier output clearly deviated from data. So a more conservative solution, including only 10 input variables with good data-MC agreement and good separation power was tested as well. The comparison of one dimensional projections of these variables between sWeighted data and simulation are shown in Fig. B.3. The BDT responses in simulation and data are shown in Fig. B.4.

Before proceeding with this ansatz the comparison to a fully cutbased analysis was carried out. The results of this comparison is summarised in Tab. B.2. The table shows that the cut optimisation favoured a higher background rejection over a larger signal efficiency when maximising the figure of merit. But most importantly, the comparison shows, that the kinematic classifier will not improve the sensitivity of the final measurement to an extent, where it would compensate for additionally introduced systematic uncertainties. In conclusion, a relatively simple high level selection is favoured, using only binary cuts.

Variable	Particle(s)	Description
log(BPV IPχ^2)	Λ_b^0,Λ_c^+,$\bar{D}^0$,K^-	Impact parameter χ^2 w.r.t the best PV computed by a Kalman-filter based vertex fit [C21]
log(BPVPDS)	Λ_b^0	Path distance significance w.r.t. the best PV
log(BPVτ)	Λ_b^0	Lifetime of the particle in ns
log(ΔM)	Λ_b^0	Dispersion of M^2 from the particle 4-vector and the covarinace matrix
log(p_T)	Λ_b^0,Λ_c^+,$\bar{D}^0$,K^-	Transverse momentum
log(Δp_T)	Λ_b^0,Λ_c^+,$\bar{D}^0$,K^-	Transverse-momentum dispersion
log(1 − DIRA)	Λ_b^0	Direction angle (i.e. the cosine of the angle between momemtum and vector from origin- to end-vertex)
arctan($c\tau$ sign)	Λ_c^+,$\bar{D}^0$	$c\tau$ sign $= \frac{c\tau}{\Delta c\tau}$ with lifetime distance $c\tau$ and its uncertainty $\Delta c\tau$
log(DIRAQ)	Λ_c^+,$\bar{D}^0$	DIRAQ $= (1 - \mathrm{DIRA}) \cdot \Delta\mathrm{DIRA}$ with ΔDIRA from the position-momentum covariance matrices
α_AP	$\bar{D}^0$	Longitudinal momentum asymmetry
$\log\left(\frac{\mathrm{ProbNNghost}}{1-\mathrm{ProbNNghost}}\right)$	K^-	ANN (TMVA MLP) output using PID variables and track ghost probability [I2] to reject fake tracks, described here

Table B.1: Summary of variables used in the kinematic MVA. All variables except the ghost probability use variables after a kinematic fit, in which the Λ_c^+ and $\bar{D}^0$ have been fixed to their nominal masses and the decay chain points back to the best primary vertex. Variables coloured in green are used in the reduced MVA.

Selection	N_sig	N_bkg (full range)
Cutbased	978 ± 46	3761 ± 124
Full MVA	945 ± 42	2799 ± 110
Reduced MVA	939 ± 43	3150 ± 114

Table B.2: Comparison of signal and background yields after cut optimisation between a high level selection based on *D-from-B* BDTs and kaon PID with and without kinematic MVA.

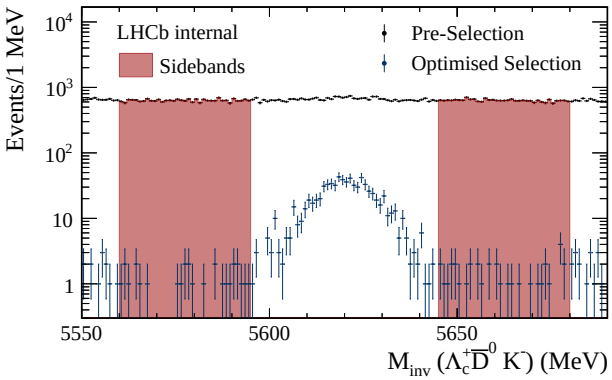


Figure B.2: Definition of background data used to train kinematic MVAs.

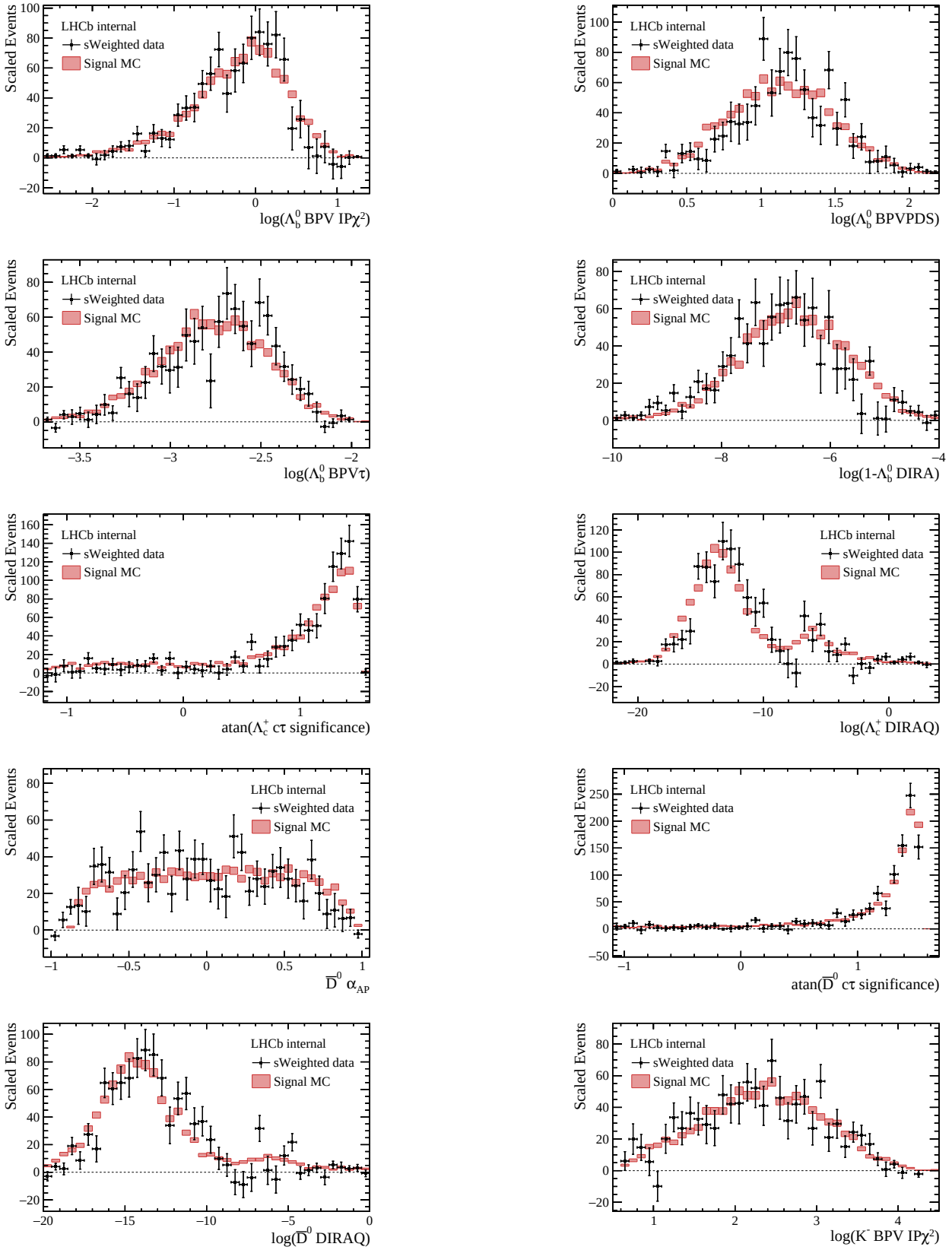


Figure B.3: Comparison of signal MC and sweighted data. The sweights have been obtained from a one dimensional fit to a cut-based selection in a narrow range around the Λ_b^0 mass.

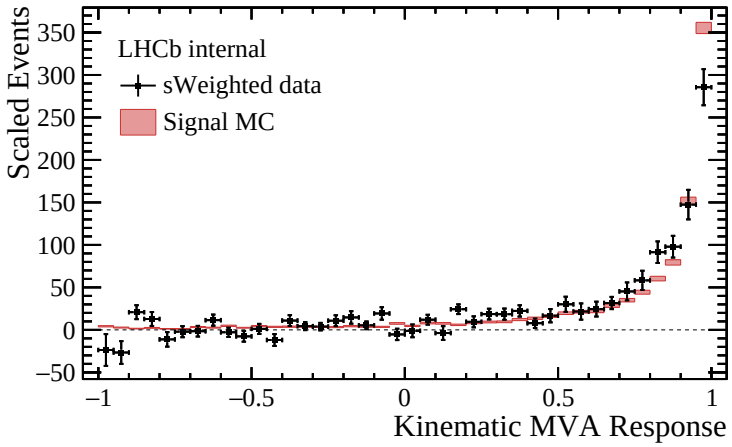
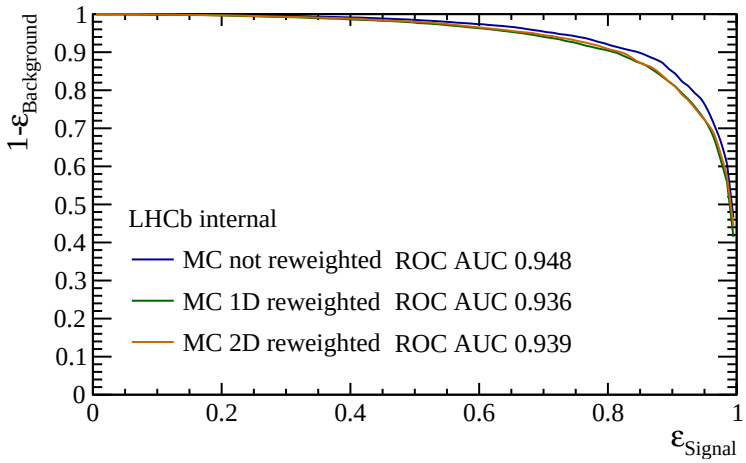
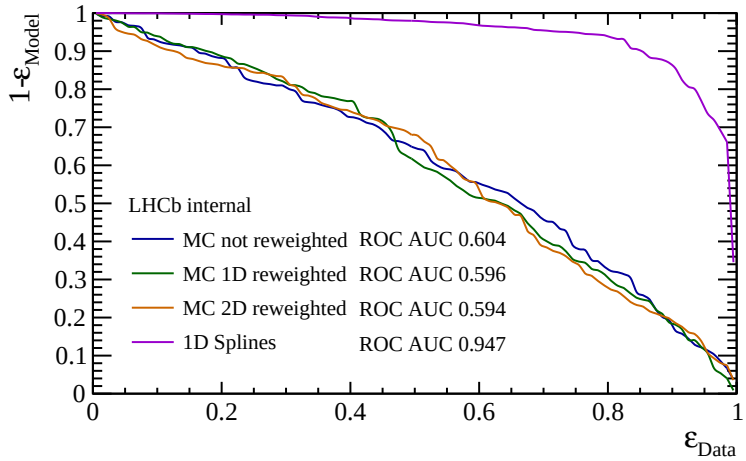


Figure B.4: Comparison between the BDT responses in data and simulation. The BDT was trained on the reduced set of input variables as described in the text.



(a)



(b)

Figure B.5: (a) ROC curves of a BDT using (rewighted) Monte Carlo simulation data as input for signal and data from the Λ_b^0 sidebands as background.
(b) ROC curves of a classifier being trained to discriminate sweighted data from the plotted signal data definitions. The area under curve is 0.5 in the asymptotic limit if the classifier can not distinguish the samples.

B.3 Vetos in signal selection

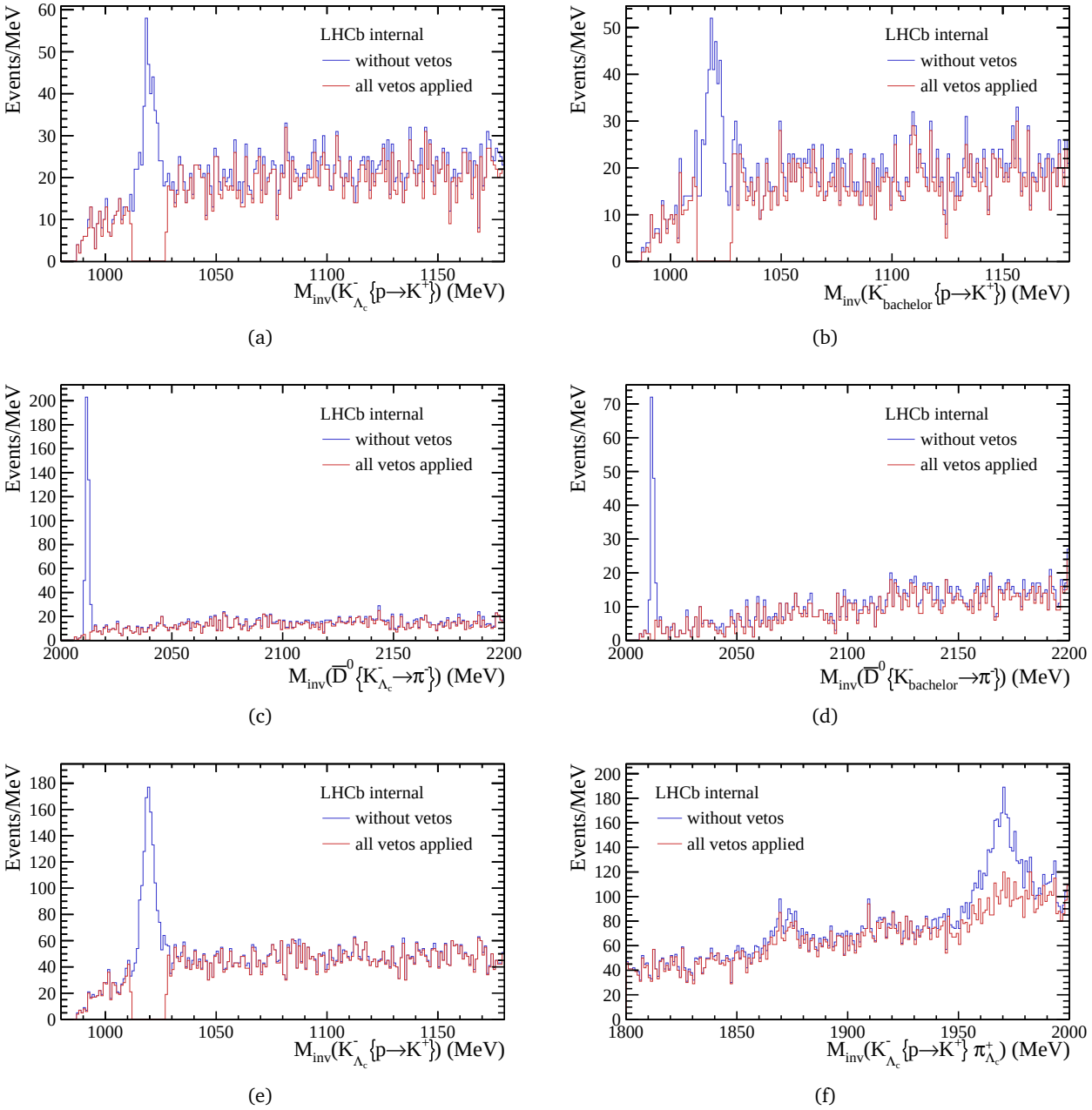


Figure B.6: Vetos applied in the selection of $\Lambda_b^0 \to \Lambda_c^+ \overline{D}^{(*)0} K^-$ (a)–(d) and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$ candidates (e) and (f). The x-axis labels the misidentified particles.
Plots (a),(b),(e) show invariant mass distributions of K^+K^- in which the $\phi(1020)$ signal is vetoed.
Plots (c) and (d) show invariant mass distributions of $\overline{D}^0\pi^-$ in which the $D^*(2010)^-$ signal is vetoed.
Plot (f) shows the invariant mass distribution of $K^+K^-\pi^+$ in which the D_s^+ signal is vetoed.

C Mass fits

This appendix supplements Sec. 7 with details on feed-down backgrounds, and a stability test of the nominal fits to signal and reference channel.

C.1 Feed-down estimates

Decays feeding down into the invariant mass spectrum of $\Lambda_c^+\bar{D}^0K^-$ are discussed, expected feed-down fractions are estimated and summarised in Tab. C.1. The feed-down fraction defined over the feed-down and signal yields as

$$f_{\text{fd}} = \frac{N_{\text{fd}}}{N_{\Lambda_b^0\to\Lambda_c^+\bar{D}^0K^-}} \,. \tag{C.1}$$

The calculations are to be understood as a guess to whether a feed-down background should be accounted for or not. It should be stated that the observed feed-down will not only depend on the physics of the decay, but also on detector acceptance and selection efficiencies. As the selection has been optimised to select $\Lambda_b^0 \to \Lambda_c^+\bar{D}^0K^-$ decays, it is anticipated that the selection efficiencies of feed-down decays is lower, leading to a lower observed feed-down.

Decay	feeds into Λ_b^0	expected f_{fd}
$\Lambda_b^0 \to \Lambda_c^+\bar{D}^0\left[K^-\pi^0\right]_{K^*(892)^-}$	✗	0.2–0.4
$\Lambda_b^0 \to \left[\Lambda_c^+\pi^0\right]_{\Sigma_c^{(*)+}}\bar{D}^0K^-$	✗	0.12
$\Lambda_b^0 \to \left[\Lambda_c^+K^-\pi^0\right]_{\Xi_c^{*0}}\bar{D}^0$	✗	< 0.01
$\Xi_b^0 \to \Lambda_c^+\bar{D}^0K^-\pi^0$	✓	< 0.01
$\Xi_b^- \to \Lambda_c^+\left[\bar{D}^0\pi^-\right]_{D^*(2010)^-}K^-$	✓	0.025
$\Xi_b \to \Xi_c\bar{D}^0h^-$	✓	< 0.01

Table C.1: Potential Λ_b^0 and Ξ_b decays which feed into the $\Lambda_c^+\bar{D}^0K^-$ invariant mass distribution. The second column indicates if the decay feeds into the $\Lambda_b^0 \to \Lambda_c^+\bar{D}^0K^-$ signal region (✓), or not (✗). Particles in red are not reconstructed.

$\Lambda_b^0 \to \Lambda_c^+\bar{D}^0\left[K^-\pi^0\right]_{K^*(892)^-}$ Unfortunately there is no measurement on corresponding doubly open charm channels from b-mesons to make an estimate based on factorisation. Also channels involving the same quark structure, but hadronising into hidden charm have not been measured yet. However, several measurements of b-mesons to single open charm allow for an order of magnitude comparison between K^- and $K^*(892)^-$ modes, which range between 1.3 ($B^0 \to D_s^-K^{(*)-}$) and 2.4 ($B^0 \to D^-K^{(*)-}$). Only the upper tails of the $\Lambda_c^+\bar{D}^0K^-$ invariant mass distribution of the $\Lambda_b^0 \to \Lambda_c^+\bar{D}^0\left[K^-\pi^0\right]_{K^*(892)^-}$ decay feed down into the fit-region starting at 5350 MeV. The fraction of events in this region is about 0.15 (in phase-space MC without detector simulation). This means that the expected feed-down fraction roughly lies between 0.2 and 0.4.

$\Lambda_b^0 \to \left[\Lambda_c^+\pi^0\right]_{\Sigma_c^{(*)+}}\bar{D}^0K^-$ These decays proceed via the same tree-level diagrams as the signal decay, as shown in Fig. C.1. But unlike the signal decay, the digram in Fig. C.1(a) is forbidden since the isospin of the Λ_b^0 spectator quarks/diquark would not be conserved [216]. This statement is based on the factorisation of electroweak and strong contributions in internal and external W-exchange diagrams [235].

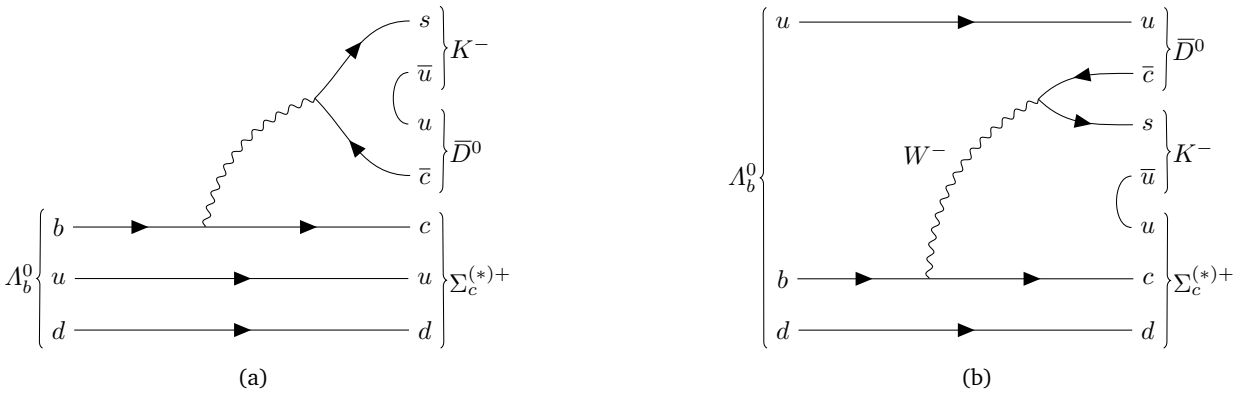


Figure C.1: Leading order Feynman-diagrams of $\Lambda_b^0 \to \Sigma_c^{(*)+}\overline{D}^0 K^-$. Diagram (a) is forbidden by isospin conservation of the spectator quarks, while diagram (b) is doubly colour suppressed.

The hypothesised *ud* diquark in the Λ_b^0 has to be split up and hadronise in different decay products. Such a diquark-splitting process is measured in the charm-sector in $\Lambda_c^+ \to \Sigma^+\pi^0$ decays. The branching ratio of this decay is – within uncertainties – equal to the branching ratios of $\Lambda_c^+ \to \Sigma^0\pi^+$ and $\Lambda_c^+ \to \Lambda\pi^+$ [236]. This means that a suppression from splitting the diquark is not expected. In addition, these decays show that the decays to $\Sigma(I=1)$ and $\Lambda(I=0)$ happen at a similar rates. For this reason, it can be assumed that $\Lambda_b^0 \to \left[\Lambda_c^+\pi^0\right]_{\Sigma_c^{(*)+}}\overline{D}^0 K^-$ happens at the same rate as $\Lambda_b^0 \to \Lambda_c^+\overline{D}^0 K^-$ in the subleading, doubly colour suppressed diagram C.1(b). Based on these considerations a feed-down fraction of $\frac{1+1/9}{9} \approx 12\,\%$ can be estimated.
The contribution of $\Sigma_c(2520)^+$ compared to $\Sigma_c(2455)^+$ is estimated to be less than a third, based on an LHCb measurement of $\Lambda_b^0 \to \Lambda_c^+\pi^+\pi^-\pi^-$ [237].
The $\Sigma_c^{(*)+}$ modes can however be enhanced by $\Lambda_b^0 \to \left[\left[\Lambda_c^+\pi^0\right]_{\Sigma_c^{(*)+}} K^-\right]_{\Xi_c^{*0}}\overline{D}^0$ decays, which are discussed in the following point.

$\Lambda_b^0 \to \left[\Lambda_c^+ K^-\pi^0\right]_{\Xi_c^{*0}}\overline{D}^0$ The decay $\Lambda_b^0 \to \Xi_c^0\overline{D}^0$ is the internal-W-emission-equivalent of $\Lambda_b^0 \to \Lambda_c^+ D_s^-$, thus suppressed by one colour factor. Further suppression comes from the requirement to decay into the excited $\Xi_c(2980)$ or $\Xi_c(3080)$ baryons. Quantifying this suppression is challenging and leads to widely varying results. Hints towards the order of magnitude come from analyses of the semileptonic Λ_b^0 and Ξ_b decays.
In Ref. [238] $\frac{N(\Lambda_b^0 \to \left[\Lambda_c^+\pi^-\pi^+\right]_{\Lambda_c(2765/2880)^+}\mu^-\overline{\nu}_\mu)}{N(\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu)}$ in the order of $\lesssim 0.1\,\%$ has been measured. This measurement can be related to Ξ_c^{*0}s as follows: The $\Xi_c(2980)$ could, according to [239], be the first radial excitation of the Ξ_c baryon, and have similar properties as the $\Lambda_c(2765)^+$, while the $\Xi_c(3080)$ is proposed to be a $J^P = \frac{5}{2}^+$ state whose partner would be the $\Lambda_c(2880)^+$.
An internal study including decays to excited Ξ_c baryons [I12], suggests that the yield of both exclusive $X_b \to \left[\Lambda_c^+ K^-\pi^+\right]_{\Xi_c(2980/3080)^{*0}}\mu^-\overline{\nu}_\mu$ decays together is in the order of 1 % compared to the inclusive $X_b \to \Xi_c\mu^-\overline{\nu}_\mu$ mode. With the colour suppression factor from internal W-exchange, $\frac{\mathcal{B}(\Lambda_b^0 \to \Xi_c^0\overline{D}^0)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}$, and the counteracting colour suppression in the ratio of signal and reference-channel $\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+\overline{D}^0 K^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ D_s^-)}$, feed-down fractions from the permill region up to the order of 1 % are anticipated. This is too small to be measured in the given sample. Note that $\Xi_c(2980)$ and $\Xi_c(3080)$ are likely to decay via $\left[\Lambda_c^+\pi^0\right]_{\Sigma_c^{(*)+}} K^-$, thus being a resonant mode in the $\Lambda_b^0 \to \Lambda_c^+ K^-\pi^0\overline{D}^0$ four-body Dalitz decay that potentially enhance the Σ_c modes.

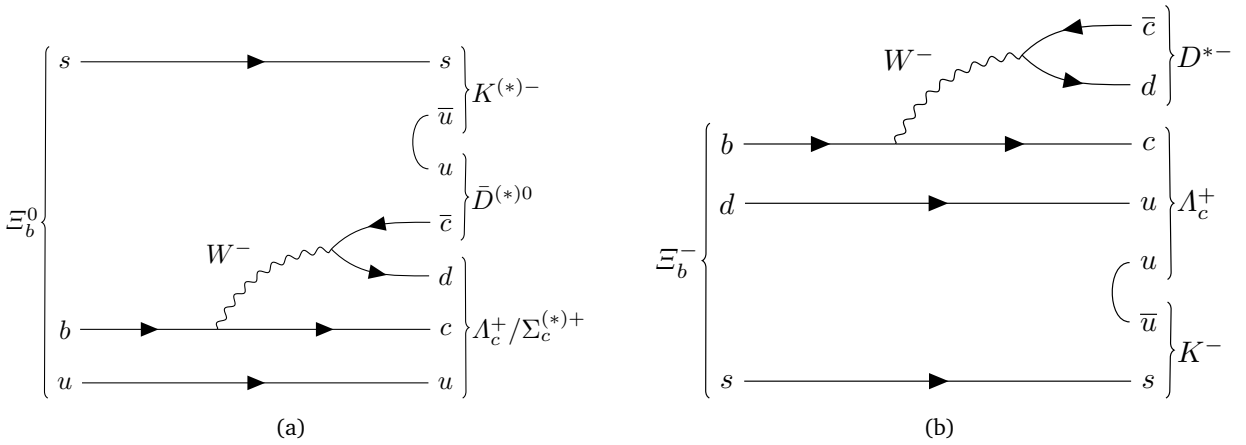


Figure C.2: Leading order Feynman-diagrams of resonant $\Xi_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0$ decays (a) and $\Xi_b^- \to \Lambda_c^+ \left[\bar{D}^0 \pi^-\right]_{D^*(2010)^-} K^-$ (b).

$\Xi_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0$ For all Ξ_b modes discussed in this and the following points, it is assumed that a π^0 or π^- has not been reconstructed. Ξ_b decays with missing pion are problematic, since their phase-space threshold is around 5655 MeV, such that resonances with a small energy-release, *e.g.* D^*, create peaking structures in the Λ_b^0 signal region.
The $\Xi_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0$ decay is very similar to the already discussed $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0$ four-body decay, and should be dominated by the $\bar{D}^{*0} \to \bar{D}^0 \pi^0$ component, even though the Σ_c mode does not suffer from isospin conservation of the spectator quarks. The lowest-order Feynman-diagram is shown in Fig. C.2(a), which is suppressed by a colour- and CKM-factor compared to the signal decay. Furthermore, the fragmentation fraction ratio $f_{\Xi_b^0}/f_{\Lambda_b^0} \approx \frac{1}{5}$ [218] and $\frac{\mathscr{B}(\Xi_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-)}{\mathscr{B}(\Xi_b^0 \to \Lambda_c^+ \bar{D}^0 K^-)} = f_{\Lambda_c^+ \bar{D}^{*0} K^-} \approx 3$ (including the D^* mode with a missing photon) are taken into account, leading to

$$f_{\Xi_b^0 \to \Lambda_c^+ \bar{D}^{*0} K^-} \approx \frac{3}{9 \cdot 20 \cdot 5} \approx 0.3\,\% \,. \tag{C.2}$$

Even with $K^*(892)^-$ and unsuppressed Σ_c modes, $f_{\Xi_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0}$ should be well below 1 %.

$\Xi_b^- \to \Lambda_c^+ \left[\bar{D}^0 \pi^-\right]_{D^*(2010)^-} K^-$ This decay proceeds via an external W emission and therefore does not suffer the additional colour-suppression of the $\Xi_b^0 \to \Lambda_c^+ \bar{D}^0 K^- \pi^0$ decays. The estimation follows the one above, but the branching ratio of $D^*(2010)^- \to \bar{D}^0 \pi^-$ ($0.677 \approx 2/3$) has to be accounted for in addition. This leads to

$$f_{\Xi_b^- \to \Lambda_c^+ D^*(2010)^- K^-} \approx \frac{3 \cdot 2/3}{20 \cdot 5} \approx 2\,\% \,. \tag{C.3}$$

Because this background peaks in the Λ_b^0 signal region, it has a direct influence on the branching ratio measurement and needs to be studied. The shapes of this decay and $\Xi_b^0 \to \Lambda_c^+ \left[\bar{D}^0 \pi^0\right]_{\bar{D}^{*0}} K^-$ are statistically not separable, but the observed yield should be dominated by the charged mode.

$\Xi_b \to \Xi_c \bar{D}^0 h^-$ These decays proceed via singly colour suppressed Feynman-diagrams, analogous to $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$, but are CKM-suppressed The $\Xi_b \to \Xi_c \bar{D}^0 h^-$ decay with highest similarity to the signal decay is $\Xi_b^0 \to \left[p K^- \pi^+ \pi^0\right]_{\Xi_c^+} \bar{D}^0 K^-$. It is suppressed by the fragmentation fraction ratio

$f_{\Xi_b^0}/f_{\Lambda_b^0} \approx 0.2$ [218], and the $pK^-\pi^+$ invariant mass selection, which, using phase-space simulation, is ≈ 0.1. The Ξ_c^+ decay to $pK^-\pi^+\pi^0$ has not been observed yet, but should be on the same level as, or slightly enhanced, compared to the Cabibbo suppressed decay $\Xi_c^+ \to pK^-\pi^+$. To close the loop back to $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^0 K^-$, a conservative estimate of the relative branching fractions $\frac{\mathcal{B}(\Xi_c^+ \to pK^-\pi^+\pi^0)}{\mathcal{B}(\Lambda_c^+ \to pK^-\pi^+)} \lesssim 0.5$[1] lead to a feed-down fraction of $\lesssim 0.01$.
Moreover, the selection efficiency of the $pK^-\pi^+$ system should be significantly smaller when coming from a Ξ_c^+ decay with a missing π^0 compared to the signal case, where the $pK^-\pi^+$ system comes from the Λ_c^+. The decrease in efficiency should mainly be due to the topological selection from vertex-related quantities.
Other $\Xi_b \to \Xi_c \bar{D}^0 h^-$ decays like $\Xi_b^- \to \left[pK^-\pi^+\pi^-\right]_{\Xi_c^0} \bar{D}^0 K^-$ or $\Xi_b^- \to \left[pK^-K^-\pi^+\right]_{\Xi_c^0} \bar{D}^0 \pi^-$ can be excluded for similar reasons, and have an additional colour suppression factor.

C.2 Fit stability test

Several components with low fit fractions are included in the nominal fit. These are the single charm $\Lambda_b^0 \to \Lambda_c^+ K^+\pi^-K^-$ decays in both signal and reference channel, the combinatorial $\Lambda_c^+\bar{D}^0$ background and the partially reconstructed $\Lambda_b^0 \to \Sigma_c(2455)^+\bar{D}^0K^-$ decay in the signal channel. Even though these components are compatible with no contribution to the overall fit, it was decided to include them to provide an unbiased estimate on the signal yields. The term biased in this context anticipates the following scenario: If the components are added back in a systematic study on the signal yields, such as described in Sec. 9.1, the signal yields would be lower, and the resulting uncertainty would be unreasonably asymmetric towards lower yields. On the other hand, the fit fraction of theses small components were expected to be this low, and there is good physical motivation to include them.

The fit stability is tested with a variant of bootstrap aggregation (*bagging*). In the conducted bagging test [C4], 20 % of selected events are rejected randomly, and a fit is carried out on the remaining sample. This is repeated 500 times and the resulting estimates of floating fit parameters are histogrammed. Bad fits are sorted out automatically according to the status code or the quality of the covariance matrix returned by the minimisation algorithm. If the fit is stable, the resulting histograms of estimates of floating fit parameters should be asymptotically Gaussian. The mean of this Gaussian would be the best estimate from the fit to data; and the width would be the estimated fit uncertainty times a correlation factor of $\sqrt{0.2} \approx 0.45$, to account for the fraction of events that are thrown away by the bagging method. The results of this test are shown in Fig. C.3 using the nominal signal fit model to produce the plots in the top and lower left plots, and the reference fit with an additional combinatorial $\Lambda_c^+ D_s^-$ background component to produce the lower right plot.

It is concluded that the inclusion of components with low fit fractions lead to stable fits with an unbiased estimator for signal and reference yields. The reduction and "symmetrisation" of the systematic uncertainty, mentioned in the beginning of this section, comes at a price of a larger statistical uncertainty. That is because the additional $\Lambda_b^0 \to \Lambda_c^+ K^+\pi^-K^-$ and $\Lambda_b^0 \to \Sigma_c(2455)^+\bar{D}^0K^-$ components are anticorrelated to the signal components and the partially reconstructed $\Lambda_b^0 \to \Lambda_c^+[\bar{D}^0\pi^0]_{\bar{D}^{*0}}K^-$ component respectively, and therefore increase the statistical uncertainty on their yields.

[1] Cabibbo suppressed Ξ_c^+ decays could be enhanced by the fact that some Cabibbo-allowed decays of the Ξ_c^+ seem to be forbidden [240].

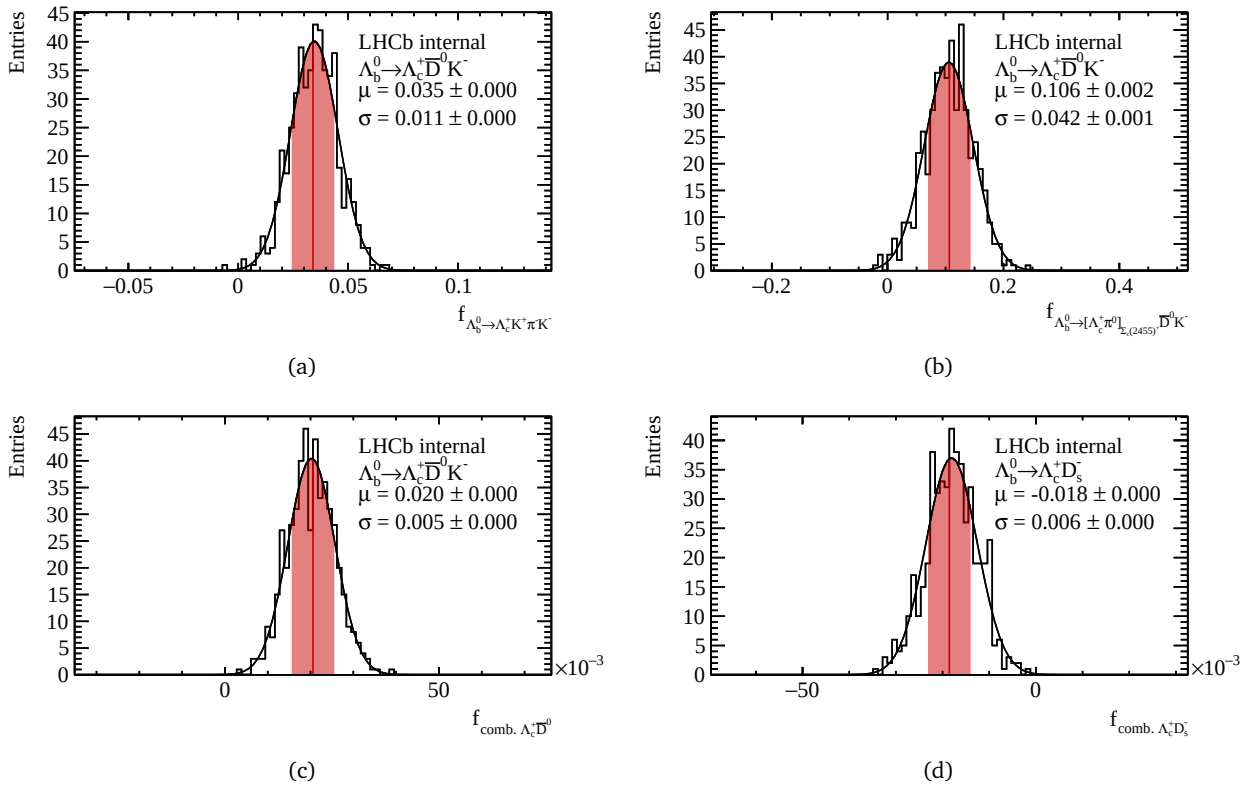


Figure C.3: Fit stability tested with a variant of bagging for the parameters (a) $f_{\Lambda_b^0 \to \Lambda_c^+ K^+ \pi^- K^-}$ (b) $f_{\Lambda_b^0 \to \Sigma_c(2455)^+ \bar{D}^0 K^-}$ and (c) $f_{\text{comb. } \Lambda_c^+ \bar{D}^0}$ of the nominal signal channel mass fit. The parameter $f_{\text{comb. } \Lambda_c^+ \bar{D}^0}$ of a dedicated fit to the reference channel is shown in (d). The solid red lines mark the best estimate from a fit to the full sample, while the filled area corresponds to the expected asymptotic standard deviation of the histogram, as described in the text.

D Toy Monte Carlo studies for the signal-fit

A Monte Carlo study has been set up in order to validate the method of correcting the uncertainty on the efficiency corrected yield when using sWeights, Eq. (10.2). The fit result of the nominal 3D signal fit, as described in Sec. 7.4, is used to sample toy datasets. These datasets are then input to the nominal fitting procedure, where a first fit is executed to get sWeights and compute pulls for a cross-check of this fit itself. These pulls are defined by

$$\text{Pull}_{\text{fit}} = \frac{N_{\text{sig,toy}} - N_{\text{sig,data}}}{\Delta N_{\text{sig,toy}}} \,, \tag{D.1}$$

with the best fit values of the signal yield on data $N_{\text{sig,data}}$, or $N_{\text{sig,toy}}$ and the respective uncertainty $\Delta N_{\text{sig,toy}}$ determined by fits to the generated datasets.

In a subsequent fit, all shape-parameters – *i.e.* floating fit parameters which are not defined as yield of a species d in the ${}_s\mathscr{P}lot$ computation – are fixed to their best fit values. The uncertainty $\Delta N_{\text{sig,toy,fixed shapes}}$ on the signal yield from that fit are used to compute the pulls of the uncertainty correction

$$\text{Pull}_{\text{corr}} = \frac{N_{\text{sig,toy}} - N_{\text{sig,data}}}{\sqrt{\sum_{i\in\text{toy-dataset}} \left({}_s\mathscr{P}_{\text{sig}}(\boldsymbol{\mu}_{i,\text{sig}})\right)^2 + \left(\Delta N^2_{\text{sig,toy}} - \Delta N^2_{\text{sig,toy,fixed shapes}}\right)}} \,. \tag{D.2}$$

The signal sWeights are denoted by ${}_s\mathscr{P}_{\text{sig}}(\boldsymbol{\mu}_{i,\text{sig}})$, where $\boldsymbol{\mu}$ are the three invariant masses in which the fit is carried out.

The result of this study is shown in Fig. D.1 and validates the method. In fact, only a minor bin-migration is observed just left of Pull $= +1$, showing that both definitions of the uncertainty are almost equivalent in this analysis.

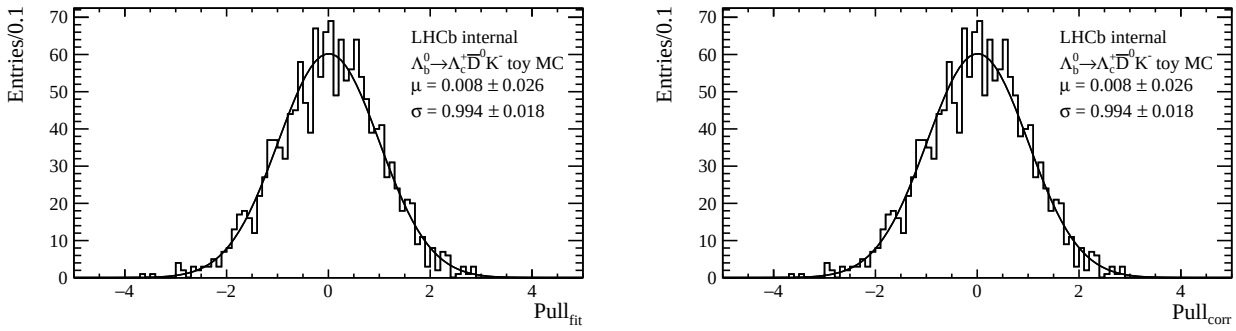


Figure D.1: Pulls from a toy Monte Carlo study of the nominal signal fit (left) and pulls from the uncertainty correction when using sWeights (right). Both methods are statistically equivalent.

References

[1] D. J. Griffiths, *Introduction to elementary particles*, Physics textbook, Wiley, New York, NY, 2nd ed., 2008. doi: 10.1002/9783527618460.

[2] E. Rutherford, *The scattering of α and β particles by matter and the structure of the atom*, Philos. Mag. Series 6 **21** (1911), no. 125 669.

[3] E. Rutherford, *Collision of α particles with light atoms. IV. An anomalous effect in nitrogen*, Philos. Mag. Series 6 **37** (1919), no. 222 581.

[4] J. Chadwick, *Possible Existence of a Neutron*, Nature **129** (1932) 312.

[5] W. Heisenberg, *Über den Bau der Atomkerne. I*, Z. Phys. **77** (1932) 1.

[6] P. A. M. Dirac, *The quantum theory of the electron*, Proc. Royal Soc. A **117** (1928), no. 778 610.

[7] C. D. Anderson, *The positive electron*, Phys. Rev. **43** (1933) 491.

[8] O. Chamberlain, E. Segrè, C. Wiegand, and T. Ypsilantis, *Observation of antiprotons*, Phys. Rev. **100** (1955) 947.

[9] B. Cork, G. R. Lambertson, O. Piccioni, and W. A. Wenzel, *Antineutrons produced from antiprotons in charge-exchange collisions*, Phys. Rev. **104** (1956) 1193.

[10] C. L. Cowan *et al.*, *Detection of the free neutrino: a confirmation*, Science **124** (1956), no. 3212 103.

[11] H. Yukawa, *On the interaction of elementary particles. I*, J. Phys. Soc. Jpn. 3rd Series **17** (1935) 48.

[12] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, *Processes involving charged mesons*, Nature **159** (1947) 694.

[13] C. D. Anderson and S. H. Neddermeyer, *Cloud chamber observations of cosmic rays at 4300 meters elevation and near sea-level*, Phys. Rev. **50** (1936) 263.

[14] Particle Data Group, C. Patrignani *et al.*, *Review of particle physics*, Chin. Phys. **C40** (2016) 100001, and 2017 update.

[15] G. D. Rochester and C. C. Butler, *Evidence for the Existence of New Unstable Elementary Particles*, Nature **160** (1947) 855.

[16] A. J. Seriff *et al.*, *Cloud-chamber observations of the new unstable cosmic-ray particles*, Phys. Rev. **78** (1950) 290.

[17] S. Sakata, *On a composite model for the new particles*, Prog. Theor. Phys. **16** (1956), no. 6 686.

[18] M. Gell-Mann, *Symmetries of baryons and mesons*, Phys. Rev. **125** (1962) 1067.

[19] Y. Ne'eman, *Derivation of strong interactions from a gauge invariance*, Nucl. Phys. **26** (1961), no. 2 222 .

[20] M. Gell-Mann, *Strong interactions of strange particles - Discussion*, Proceedings of the 11th International Conference on High-energy Physics (1962) 805.

[21] National Research Council, *Elementary-Particle Physics*, The National Academies Press, Washington, DC, 1986. doi: 10.17226/629.

[22] V. E. Barnes *et al.*, *Observation of a hyperon with strangeness minus three*, Phys. Rev. Lett. **12** (1964) 204.

[23] M. Gell-Mann, *A schematic model of baryons and mesons*, Phys. Lett. **8** (1964), no. 3 214 .

[24] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 1*, CERN-TH-401 (1964).

[25] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, CERN-TH-412 (1964).

[26] J. Greensite, *An introduction to the confinement problem*, Lect. Notes Phys. **821** (2011) 1.

[27] O. W. Greenberg, *Spin and unitary-spin independence in a paraquark model of baryons and mesons*, Phys. Rev. Lett. **13** (1964) 598.

[28] E. D. Bloom *et al.*, *High-energy inelastic $e-p$ scattering at 6° and 10°*, Phys. Rev. Lett. **23** (1969) 930.

[29] M. Breidenbach *et al.*, *Observed behavior of highly inelastic electron-proton scattering*, Phys. Rev. Lett. **23** (1969) 935.

[30] J. D. Bjorken, *Asymptotic sum rules at infinite momentum*, Phys. Rev. **179** (1969) 1547.

[31] R. P. Feynman, *The behavior of hadron collisions at extreme energies*, Conf. Proc. **C690905** (1969) 237.

[32] J. D. Bjorken and E. A. Paschos, *Inelastic electron-proton and γ-proton scattering and the structure of the nucleon*, Phys. Rev. **185** (1969) 1975.

[33] TASSO collaboration, R. Brandelik *et al.*, *Evidence for planar events in e^+e^- annihilation at high-energies*, Phys. Lett. **86B** (1979) 243.

[34] MARK-J collaboration, D. P. Barber *et al.*, *Discovery of three-jet events and a test of quantum chromodynamics at PETRA*, Phys. Rev. Lett. **43** (1979) 830.

[35] PLUTO collaboration, C. Berger *et al.*, *Evidence for gluon bremsstrahlung in e^+e^- annihilations at high-energies*, Phys. Lett. **86B** (1979) 418.

[36] JADE collaboration, W. Bartel *et al.*, *Observation of planar three jet events in e^+e^- annihilation and evidence for gluon bremsstrahlung*, Phys. Lett. **91B** (1980) 142.

[37] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343.

[38] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973) 1346.

[39] P. Söding, *On the discovery of the gluon*, Eur. Phys. J. **H35** (2010) 3.

[40] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D **2** (1970) 1285.

[41] E598 collaboration, J. J. Aubert *et al.*, *Experimental observation of a heavy particle J*, Phys. Rev. Lett. **33** (1974) 1404.

[42] SLAC-SP-017 collaboration, J.-E. Augustin *et al.*, *Discovery of a narrow resonance in e^+e^- annihilation*, Phys. Rev. Lett. **33** (1974) 1406.

[43] M. Kobayashi and T. Maskawa, *CP-violation in the renormalizable theory of weak interaction*, Prog. Theor. Phys. **49** (1973), no. 2 652.

[44] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π decay of the K_2^0 meson*, Phys. Rev. Lett. **13** (1964) 138.

[45] M. L. Perl *et al.*, *Evidence for anomalous lepton production in e^+–e^- annihilation*, Phys. Rev. Lett. **35** (1975) 1489.

[46] S. W. Herb *et al.*, *Observation of a dimuon resonance at 9.5 GeV in 400-GeV proton-nucleus collisions*, Phys. Rev. Lett. **39** (1977) 252.

[47] CDF collaboration, F. Abe *et al.*, *Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab*, Phys. Rev. Lett. **74** (1995) 2626, arXiv:hep-ex/9503002.

[48] D0 collaboration, S. Abachi *et al.*, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632, arXiv:hep-ex/9503003.

[49] DONUT collaboration, K. Kodama *et al.*, *Observation of tau neutrino interactions*, Phys. Lett. **B504** (2001) 218, arXiv:hep-ex/0012035.

[50] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3, S. Schael *et al.*, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, arXiv:hep-ex/0509008.

[51] UA1 collaboration, G. Arnison *et al.*, *Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV*, Phys. Lett. **122B** (1983) 103.

[52] UA2 collaboration, M. Banner *et al.*, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\overline{p}p$ collider*, Phys. Lett. **122B** (1983) 476.

[53] UA1 collaboration, G. Arnison *et al.*, *Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider*, Phys. Lett. **126B** (1983) 398.

[54] UA2 collaboration, P. Bagnaia *et al.*, *Evidence for $Z^0 \to e^+e^-$ at the CERN $\overline{p}p$ collider*, Phys. Lett. **129B** (1983) 130.

[55] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.

[56] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[57] A. Salam, *Weak and electromagnetic interactions*, 8th Nobel Symposium Lerum, Sweden, 367 (1968).

[58] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.

[59] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

[60] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585.

[61] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv:1207.7214.

[62] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv:1207.7235.

[63] Y. Nambu, *Axial vector current conservation in weak interactions*, Phys. Rev. Lett. **4** (1960) 380.

[64] J. Goldstone, *Field theories with superconductor solutions*, Nuovo Cim. **19** (1961) 154.

[65] Y. Nambu and G. Jona-Lasinio, *Dynamical model of elementary particles based on an analogy with superconductivity. I*, Phys. Rev. **122** (1961) 345.

[66] Y. Nambu and G. Jona-Lasinio, *Dynamical model of elementary particles based on an analogy with superconductivity. II*, Phys. Rev. **124** (1961) 246.

[67] I. C. Cloet and C. D. Roberts, *Explanation and prediction of observables using continuum strong qcd*, Prog. Part. Nucl. Phys. **77** (2014) 1, arXiv:1310.2651.

[68] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191.

[69] A. Deur, S. J. Brodsky, and G. F. de Teramond, *The QCD running coupling*, Prog. Part. Nucl. Phys. **90** (2016) 1, arXiv:1604.08082.

[70] W. A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta, *Deep-inelastic scattering beyond the leading order in asymptotically free gauge theories*, Phys. Rev. D **18** (1978) 3998.

[71] Particle Data Group, M. Aguilar-Benitez *et al.*, *Review of particle physics – Z baryons*, Phys. Lett. **B170** (1986) 289.

[72] LEPS collaboration, T. Nakano *et al.*, *Evidence for a narrow $S = +1$ baryon resonance in photoproduction from the neutron*, Phys. Rev. Lett. **91** (2003) 012002, arXiv:hep-ex/0301020.

[73] D. Diakonov, V. Petrov, and M. V. Polyakov, *Exotic anti-decuplet of baryons: Prediction from chiral solitons*, Z. Phys. **A359** (1997) 305, arXiv:hep-ph/9703373.

[74] K. H. Hicks, *On the conundrum of the pentaquark*, Eur. Phys. J. H **37** (2012) 1.

[75] Belle collaboration, S.-K. Choi *et al.*, *Observation of a narrow charmoniumlike state in exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ decays*, Phys. Rev. Lett. **91** (2003) 262001, arXiv:hep-ex/0309032.

[76] 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.

[77] LHCb collaboration, R. Aaij *et al.*, *Model-independent evidence for $J/\psi p$ contributions to $\Lambda_b^0 \to J/\psi p K^-$ decays*, Phys. Rev. Lett. **117** (2016) 082002, arXiv:1604.05708.

[78] LHCb collaboration, R. Aaij *et al.*, *Evidence for exotic hadron contributions to $\Lambda_b^0 \to J/\psi p \pi^-$ decays*, Phys. Rev. Lett. **117** (2016) 082003, arXiv:1606.06999.

[79] Belle collaboration, K. Chilikin *et al.*, *Observation of a new charged charmoniumlike state in $\bar{B}^0 \to J/\psi K^-\pi^+$ decays*, Phys. Rev. **D90** (2014), no. 11 112009, arXiv:1408.6457.

[80] F.-K. Guo, U.-G. Meißner, W. Wang, and Z. Yang, *How to reveal the exotic nature of the $P_c(4450)$*, Phys. Rev. **D92** (2015), no. 7 071502, arXiv:1507.04950.

[81] T. J. Burns, *Phenomenology of $p_c(4380)^+$, $p_c(4450)^+$ and related states*, Eur. Phys. J. A **51** (2015) 152, arXiv:1509.02460.

[82] C. W. Shen and Y. H. Lin, *Decay behaviors of the P_c hadronic molecules*, arXiv:1710.09037.

[83] R. F. Lebed, R. E. Mitchell, and E. S. Swanson, *Heavy-quark QCD exotica*, Prog. Part. Nucl. Phys. **93** (2017) 143 , arXiv:1610.04528.

[84] H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, *The hidden-charm pentaquark and tetraquark states*, Physics Reports **639** (2016) 1 , arXiv:1601.02092, The hidden-charm pentaquark and tetraquark states.

[85] V. D. Burkert and C. D. Roberts, *Roper resonance – solution to the fifty year puzzle*, arXiv:1710.02549.

[86] R. A. Briceno *et al.*, *Issues and opportunities in exotic hadrons*, Chin. Phys. **C40** (2016), no. 4 042001, arXiv:1511.06779.

[87] J. R. Andersen *et al.*, *Discovering Technicolor*, Eur. Phys. J. Plus **126** (2011) 81, arXiv:1104.1255.

[88] E. Eichten *et al.*, *Charmonium: the model*, Phys. Rev. **D17** (1978) 3090, [Erratum: Phys. Rev.D21,313(1980)].

[89] T. Regge, *Introduction to complex orbital momenta*, Nuovo Cim. **14** (1959) 951.

[90] A. Tang and J. W. Norbury, *Properties of Regge trajectories*, Phys. Rev. **D62** (2000) 016006, arXiv:hep-ph/0004078.

[91] C. S. Fischer, S. Kubrak, and R. Williams, *Mass spectra and Regge trajectories of light mesons in the Bethe-Salpeter approach*, Eur. Phys. J. **A50** (2014) 126, arXiv:1406.4370.

[92] C. Alexandrou, P. de Forcrand, and B. Lucini, *Evidence for diquarks in lattice QCD*, Phys. Rev. Lett. **97** (2006) 222002, arXiv:hep-lat/0609004.

[93] R. Babich *et al.*, *Diquark correlations in baryons on the lattice with overlap quarks*, Phys. Rev. **D76** (2007) 074021, arXiv:hep-lat/0701023.

[94] G. Eichmann *et al.*, *Baryons as relativistic three-quark bound states*, Prog. Part. Nucl. Phys. **91** (2016) 1, arXiv:1606.09602.

[95] J. Segovia, C. D. Roberts, and S. M. Schmidt, *Understanding the nucleon as a Borromean bound-state*, Phys. Lett. **B750** (2015) 100, arXiv:1506.05112.

[96] S. J. Brodsky, D. S. Hwang, and R. F. Lebed, *Dynamical picture for the formation and decay of the exotic XYZ mesons*, Phys. Rev. Lett. **113** (2014), no. 11 112001, arXiv:1406.7281.

[97] R. F. Lebed, *The pentaquark candidates in the dynamical diquark picture*, Phys. Lett. **B749** (2015) 454, arXiv:1507.05867.

[98] BESIII collaboration, M. Ablikim *et al.*, *Observation of a charged charmoniumlike structure in* $e^+e^- \to \pi^+\pi^- J/\psi$ *at* $\sqrt{s} =$ 4.26 *GeV*, Phys. Rev. Lett. **110** (2013) 252001, arXiv:1303.5949.

[99] Belle collaboration, X. L. Wang *et al.*, *Measurement of* $e^+e^- \to \pi^+\pi^-\psi(2S)$ *via initial state radiation at Belle*, Phys. Rev. **D91** (2015) 112007, arXiv:1410.7641.

[100] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the* $Z(4430)^-$ *state*, Phys. Rev. Lett. **112** (2014) 222002, arXiv:1404.1903.

[101] Z. E. Meziani *et al.*, *A search for the LHCb charmed 'pentaquark' using photo-production of* J/ψ *at threshold in Hall C at Jefferson Lab*, arXiv:1609.00676.

[102] LHCb Collaboration, B. Adeva *et al.*, *Road map for selected key measurements of LHCb*, Tech. Rep. arXiv:0912.4179, CERN-LHCb-PUB-2009-029, CERN, Geneva, May, 2010.

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

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

[105] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007. LHCb-TDR-001.

[106] L. Evans and P. Bryant, *LHC machine*, J. Instrum. **3** (2008), no. 08 S08001.

[107] E. H. M. Heijne *et al.*, *A silicon surface barrier microstrip detector designed for high energy physics*, Nuclear Instruments and Methods **178** (1980), no. 2 331 .

[108] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011. LHCb-TDR-005.

[109] A. Affolder *et al.*, *Radiation damage in the LHCb vertex locator*, JINST **8** (2013) P08002, arXiv:1302.5259.

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

[111] A. Bay *et al.*, *Hybrid design, procurement and testing for the LHCb silicon tracker*, Tech. Rep. CERN-LHCB-2005-061, 2005. doi: 10.5170/CERN-2005-011.67.

[112] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029. LHCb-TDR-008.

[113] M. Tobin, *Performance of the LHCb tracking detectors*, PoS Vertex2012 (2013) 047.

[114] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024. LHCb-TDR-006.

[115] S. Bachmann *et al.*, *Ageing in the lhcb outer tracker: Phenomenon, culprit and effect of oxygen*, Nucl. Instr. Meth. Phys. Res. A **617** (2010), no. 1 202 .

[116] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[117] M. Alemi *et al.*, *First operation of a hybrid photon detector prototype with electrostatic cross-focussing and integrated silicon pixel readout*, Nucl. Instr. Meth. Phys. Res. A **449** (2000), no. 1 48 .

[118] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[119] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036. LHCb-TDR-002.

[120] A. Arefev *et al.*, *Beam Test Results of the LHCb Electromagnetic Calorimeter.*, Tech. Rep. CERN-LHCb-2007-149, CERN, Geneva, May, 2008. revised version submitted on 2008-05-15 09:09:53.

[121] P. Perret and X. Vilasís-Cardona, *Performance of the lhcb calorimeters during the period 2010-2012*, J. Phys. : Conf. Ser **587** (2015) 012012.

[122] J. R. Harrison, *Radiation damage studies in the LHCb VELO detector and searches for lepton flavour and baryon number violating tau decays*, PhD thesis, University of Manchester, 2014, CERN-THESIS-2014-068.

[123] A. Martin Sanchez, P. Robbe, and M.-H. Schune, *Performances of the LHCb L0 Calorimeter Trigger*, Tech. Rep. CERN-LHCb-PUB-2011-026, CERN, Geneva, Jun, 2012.

[124] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010. LHCb-TDR-004.

[125] G. Bencivenni *et al.*, *Advances in triple-gem detector operation for high-rate particle triggering*, Nucl. Instr. Meth. Phys. Res. A **513** (2003), no. 1 264 , Proceedings of the 6th International Conference on Position-Sensitive Detectors.

[126] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346.

[127] O. S. Brüning and P. Collier, *Building a behemoth*, Nature **448** (2007) 285.

[128] O. S. Brüning *et al.*, *LHC design report, v.1 : the LHC main ring*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004. doi: 10.5170/CERN-2004-003-V-1; O. S. Brüning *et al.*, *LHC design report, v.2 : the LHC infrastructure and general services*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004. doi: 10.5170/CERN-2004-003-V-2; M. Benedikt *et al.*, *LHC design report, v.3 : the LHC injector chain*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004. doi: 10.5170/CERN-2004-003-V-3.

[129] E. A. Mobs, *The CERN accelerator complex*, , OPEN-PHO-ACCEL-2016-013.

[130] LHCb collaboration, R. Aaij *et al.*, *Absolute luminosity measurements with the LHCb detector at the LHC*, JINST **7** (2012) P01010, arXiv:1110.2866.

[131] S. van der Meer, *Calibration of the effective beam height in the ISR*, , CERN-ISR-PO-68-31.

[132] M. Ferro-Luzzi, *Proposal for an absolute luminosity determination in colliding beam experiments using vertex detection of beam-gas interactions*, Nucl. Instr. Meth. Phys. Res. A **553** (2005), no. 3 388.

[133] LHCb collaboration, R. Aaij *et al.*, *Precision luminosity measurements at LHCb*, JINST **9** (2014) P12005, arXiv:1410.0149.

[134] LHCb collaboration, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031. LHCb-TDR-010.

[135] T. Head, *The LHCb trigger system*, JINST **9** (2014) C09015.

[136] LHCb collaboration, *LHCb online system, data acquisition and experiment control: Technical Design Report*, CERN-LHCC-2001-040. LHCb-TDR-007.

[137] M. Williams *et al.*, *The HLT2 topological lines*, Tech. Rep. CERN-LHCb-PUB-2011-002, CERN, Geneva, Jan, 2011.

[138] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers*, Tech. Rep. CERN-LHCb-PUB-2011-016, CERN, Geneva, Sep, 2011. LHCb-INT-2011-030.

[139] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, arXiv:1210.6861.

[140] M. Frank *et al.*, *Deferred High Level Trigger in LHCb: a boost to CPU resource utilization*, J. Phys. Conf. Ser. **513** (2014) 012006.

[141] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[142] J. Albrecht *et al.*, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. Conf. Ser. **513** (2014) 012001, arXiv:1310.8544.

[143] O. Callot, *FastVelo, a fast and efficient pattern recognition package for the Velo*, Tech. Rep. CERN-LHCb-PUB-2011-001, CERN, Geneva, Jan, 2011. LHCb.

[144] O. Callot and M. Schiller, *PatSeeding: A Standalone Track Reconstruction Algorithm*, Tech. Rep. CERN-LHCb-2008-042, CERN, Geneva, Aug, 2008.

[145] O. Callot and S. Hansmann-Menzemer, *The Forward Tracking: Algorithm and Performance Studies*, Tech. Rep. CERN-LHCb-2007-015, CERN, Geneva, May, 2007.

[146] M. Needham and J. Van Tilburg, *Performance of the track matching*, Tech. Rep. CERN-LHCb-2007-020, CERN, Geneva, Mar, 2007.

[147] M. Needham, *Performance of the Track Matching*, Tech. Rep. LHCb-2007-129. CERN-LHCb-2007-129, CERN, Geneva, Oct, 2007.

[148] O. Callot, *Downstream Pattern Recognition*, Tech. Rep. CERN-LHCb-2007-026, CERN, Geneva, Mar, 2007.

[149] R. E. Kálmán, *A new approach to linear filtering and prediction problems*, J. Basic Eng. (1960) 35.

[150] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.

[151] J. Van Tilburg and M. Merk, *Track simulation and reconstruction in LHCb*, PhD thesis, Vrije Universiteit Amsterdam, 2005, CERN-THESIS-2005-040.

[152] M. De Cian, *Track Reconstruction Efficiency and Analysis of $B^0 \to K^{*0}\mu^+\mu^-$ at the LHCb Experiment*, PhD thesis, Universität Zürich, Sep, 2013, CERN-THESIS-2013-145.

[153] V. Blobel, *Software alignment for tracking detectors*, Nucl. Instrum. Meth. **A566** (2006) 5.

[154] W. Hulsbergen, *The Global covariance matrix of tracks fitted with a Kalman filter and an application in detector alignment*, Nucl. Instrum. Meth. **A600** (2009) 471, arXiv:0810.2241.

[155] S. Viret, C. Parkes, and M. Gersabeck, *Alignment procedure of the LHCb Vertex Detector*, Nucl. Instrum. Meth. **A596** (2008) 157, arXiv:0807.5067.

[156] M. Gersabeck, S. Viret, and C. Parkes, *Performance of the LHCb Vertex Detector Alignment Algorithm determined with Beam Test Data*, Nucl. Instrum. Meth. **A596** (2008) 164, arXiv:0807.5069.

[157] J. Amoraal *et al.*, *Application of vertex and mass constraints in track-based alignment*, Nucl. Instrum. Meth. **A712** (2013) 48, arXiv:1207.4756.

[158] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

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

[160] L. Anderlini *et al.*, *The PIDCalib package*, Tech. Rep. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016.

[161] A. Poluektov, *Kernel density estimation of a multidimensional efficiency profile*, JINST **10** (2015), no. 02 P02011, arXiv:1411.5528.

[162] LHCb collaboration, *LHCb computing: Technical Design Report*, CERN-LHCC-2005-019. LHCb-TDR-011.

[163] M. Cattaneo, P. Charpentier, P. Clarke, and S. Roiser, *Recent and planned changes to the LHCb computing model*, J. Phys. Conf. Ser. **513** (2014) 032017.

[164] L. Bel, *The LHCb Starterkit*, PoS ICHEP2016 (2017) 334, image from https://lhcb.github.io/starterkit-lessons/first-analysis-steps/dataflow.html.

[165] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

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

[167] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[168] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[169] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, pp. 1155–1161, 2010. doi: 10.1109/NSSMIC.2010.5873949.

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

[171] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv:hep-ph/0506026.

[172] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[173] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[174] LHCb collaboration, R. Aaij *et al.*, *Study of beauty hadron decays into pairs of charm hadrons*, Phys. Rev. Lett. **112** (2014) 202001, arXiv:1403.3606.

[175] LHCb collaboration, R. Aaij *et al.*, *First observations of $\bar{B}^0_s \to D^+D^-$, $D^+_s D^-$ and $D^0\bar{D}^0$ decays*, Phys. Rev. **D87** (2013) 092007, arXiv:1302.5854.

[176] F. Reiß, *Measurement of efficiency corrected yields of the decay $\Lambda^0_b \to \Lambda^+_c D^-_s$– a first step towards a pentaquark search in the decay $\Lambda^0_b \to \Lambda^+_c \bar{D}^0 K^-$*, Master thesis, Physikalisches Institut, Heidelberg, 2017, https://www.physi.uni-heidelberg.de//Publications/main.pdf.

[177] C. Otte, *Identifying charmed hadrons from b hadron decays at LHCb*, Bachelor thesis, Physikalisches Institut, Heidelberg, 2017, https://www.physi.uni-heidelberg.de//Publications/BachelorThesisOtte.pdf.

[178] C. O. A. Vahl, *First observation of the decay $\Lambda^0_b \to J/\psi\Lambda\phi$ with the LHCb experiment*, Diploma thesis, Physikalisches Institut, Heidelberg, 2017, https://www.physi.uni-heidelberg.de//Publications/thesis_vahl.pdf.

[179] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039, latest version under http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf (dated October 4, 2013).

[180] J. Köster and S. Rahmann, *Snakemake – a scalable bioinformatics workflow engine*, Bioinformatics **28** (2012), no. 19 2520.

[181] M. McNutt, *Reproducibility*, Science **343** (2014), no. 6168 229.

[182] J. Conrad, *Reproducibility: Don't cry wolf*, Nature **523** (2015) 27.

[183] B. A. Nosek *et al.*, *Promoting an open research culture*, Science **348** (2015), no. 6242 1422.

[184] D. B. Allison, A. W. Brown, B. J. George, and K. A. Kaiser, *Reproducibility: A tragedy of errors*, Nature **530** (2016) 27.

[185] M. Baker, *1,500 scientists lift the lid on reproducibility*, Nature **533** (2016) 452.

[186] ALLEA - All European Academies, *The european code of conduct for research integrity, revised edition*, , ISBN 978-3-00-055767-5.

[187] S. Neubert *et al.*, *LHCb analysis preservation roadmap*, , `https://www.authorea.com/users/42472/articles/140021-lhcb-analysis-preservation-roadmap/_show_article`.

[188] A. Silver, *Collaborative software development made easy*, Nature **550** (2017) 143.

[189] R. MacCoun and S. Perlmutter, *Blind analysis: Hide results to seek the truth*, Nature **526** (2015) 187.

[190] R. V. Hogg, J. W. McKean, and A. T. Craig, *Introduction to Mathematical Statistics*, Pearson Higher Education, 7th ed., 2013.

[191] G. Cowan, *Statistical data analysis*, Oxford University Press, 1998.

[192] F. James and M. Roos, *Minuit: A system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun. **10** (1975) 343.

[193] R. J. Barlow, *Extended maximum likelihood*, Nucl. Instrum. Meth. **A297** (1990) 496.

[194] *Outline of a theory of statistical estimation based on the classical theory of probability*, Philos. Trans. Royal Soc. A **236** (1937), no. 767 333.

[195] C. J. Clopper and E. S. Pearson, *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika **26** (1934), no. 4 404.

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

[197] P. Seyfert, *The search for $\tau \to \mu\mu\mu$ at LHCb*, PhD thesis, Physikalisches Institut, Heidelberg, Jan, 2015, CERN-THESIS-2015-021.

[198] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer Series in Statistics, Springer, Dordrecht, 2nd ed., 2009. doi: 10.1007/978-0-387-84858-7.

[199] J. H. Friedman, *Greedy function approximation: A gradient boosting machine*, Ann. Stat. **29** (2001), no. 5 1189.

[200] P. J. Huber, *Robust estimation of a location parameter*, Ann. Math. Statist. **35** (1964) 73.

[201] M. Needham, *Clone Track Identification using the Kullback-Liebler Distance*, Tech. Rep. CERN-LHCb-2008-002, CERN, Geneva, Jan, 2008.

[202] R. Hierk, M. Merk, M. Needham, and R. Van der Eijk, *Performance of the LHCb 00 track fitting software*, Tech. Rep. LHCb-2000-086, CERN, Geneva, Aug, 2000.

[203] E. Bos *et al.*, *The Trajectory Model for Track Fitting and Alignment*, Tech. Rep. CERN-LHCb-2007-008, CERN, Geneva, Mar, 2007.

[204] J. Podolanski and R. Armenteros, *III. Analysis of V-events*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **45** (1954), no. 360 13, `arXiv:http://dx.doi.org/10.1080/14786440108520416`.

[205] CDF collaboration, T. Aaltonen *et al.*, *Observation of new charmless decays of bottom hadrons*, Phys. Rev. Lett. **103** (2009) 031801, `arXiv:0812.4271`.

[206] M. Stahl, *Analysis of resonant structures in the pK^0_S-channel at HERMES*, Master's thesis, Giessen U., 2013. DESY-HERA-HERMES-13-008.

[207] CDF collaboration, T. Aaltonen *et al.*, *Measurements of direct cp violating asymmetries in charmless decays of strange bottom mesons and bottom baryons*, Phys. Rev. Lett. **106** (2011) 181802, `arXiv:1103.5762`.

[208] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, `arXiv:1306.0249`.

[209] R. Van der Eijk, R. Hierck, M. Merk, and M. Needham, *Addition of the vertex detector measurements to the track fit*, .

[210] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, `arXiv:1612.07489`.

[211] LHCb collaboration, R. Aaij *et al.*, *Study of the kinematic dependences of Λ_b^0 production in pp collisions and a measurement of the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ branching fraction*, JHEP **08** (2014) 143, `arXiv:1405.6842`.

[212] Belle collaboration, H. Ikeda *et al.*, *A detailed test of the CsI(Tl) calorimeter for BELLE with photon beams of energy between 20-MeV and 5.4-GeV*, Nucl. Instrum. Meth. **A441** (2000) 401.

[213] LHCb collaboration, R. Aaij *et al.*, *Measurement of the $B^\pm$ production asymmetry and the CP asymmetry in $B^\pm \to J/\psi K^\pm$ decays*, Phys. Rev. **D95** (2017) 052005, arXiv:1701.05501.

[214] 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.

[215] S. Wandernoth, *Measurement of the $B_s^0 - \bar{B}_s^0$ Oscillation Frequency at LHCb using 1 fb^{-1} of data taken in 2011*, PhD thesis, Physikalisches Institut, Heidelberg, 2013, CERN-THESIS-2013-317.

[216] A. Datta, H. J. Lipkin, and P. J. O'Donnell, *Implications of isospin conservation in Λ_b^0 decays and lifetime*, Phys. Lett. **B450** (1999) 250, arXiv:hep-ph/9809294.

[217] S. Groote and J. G. Korner, *Theory of heavy baryon decay*, pp. 185–191, 2000. arXiv:hep-ph/0003116. doi: 10.1142/9789812791870_0028.

[218] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the mass and lifetime of the Ξ_b^0 baryon*, Phys. Rev. Lett. **113** (2014) 032001, arXiv:1405.7223.

[219] V. V. Gligorov, *Reconstruction of the Channel $B_d^0 \to D^+\pi^-$ and Background Classification at LHCb (revised)*, Tech. Rep. CERN-LHCb-2007-044, CERN, Geneva, Jun, 2007. revised version submitted on 2008-01-24 12:46:44.

[220] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38** (2014) 090001, and 2015 update.

[221] LHCb collaboration, R. Aaij *et al.*, *Observation of the decays $\Lambda_b^0 \to \chi_{c1} p K^-$ and $\Lambda_b^0 \to \chi_{c2} p K^-$*, Phys. Rev. Lett. **119** (2017) 062001, arXiv:1704.07900.

[222] J. D. Bjorken, *Estimates of decay branching ratios for hadrons containing charm and bottom quarks.*, , FERMILAB-PUB-86-189-T.

[223] BaBar collaboration, P. del Amo Sanchez *et al.*, *Measurement of the $B \to \bar{D}^{(*)0} D^{(*)0} K$ branching fractions*, Phys. Rev. **D83** (2011) 032004, arXiv:1011.3929.

[224] LHCb collaboration, R. Aaij *et al.*, *Observation of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays*, Phys. Rev. **D87** (2013) 112012, arXiv:1304.4530.

[225] LHCb collaboration, R. Aaij *et al.*, *Observation of $B_c^+ \to J/\psi D^{(*)} K^{(*)}$ decays*, Phys. Rev. **D95** (2017) 032005, arXiv:1612.07421.

[226] F. Krinner, A. Lenz, and T. Rauh, *The inclusive decay $b \to c\bar{c}s$ revisited*, Nucl. Phys. **B876** (2013) 31, arXiv:1305.5390.

[227] J. G. Korner, M. Kramer, and D. Pirjol, *Heavy baryons*, Prog. Part. Nucl. Phys. **33** (1994) 787, arXiv:hep-ph/9406359.

[228] J. Zhu, Z.-T. Wei, and H.-W. Ke, *The semi-leptonic and non-leptonic weak decays of Λ_b^0*, arXiv:1803.01297.

[229] M. Bauer, B. Stech, and M. Wirbel, *Exclusive Nonleptonic Decays of D, D(s), and B Mesons*, Z. Phys. **C34** (1987) 103.

[230] E. B. Wilson, *Probable inference, the law of succession, and statistical inference*, J. Am. Stat. Assoc. **22** (1927), no. 158 209.

[231] L. D. Brown, T. T. Cai, and A. DasGupta, *Interval estimation for a binomial proportion*, Statist. Sci. **16** (2001) 101.

[232] R. D. Cousins, K. E. Hymes, and J. Tucker, *Frequentist evaluation of intervals estimated for a binomial parameter and for the ratio of poisson means*, Nucl. Instr. Meth. Phys. Res. A **612** (2010), no. 2 388 , arXiv:0905.3831.

[233] M. Mirza and S. Osindero, *Conditional Generative Adversarial Nets*, CoRR **abs/1411.1784** (2014).

[234] I. Goodfellow *et al.*, *Generative Adversarial Nets*, in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani *et al.*, eds.), pp. 2672–2680. Curran Associates, Inc., 2014.

[235] L.-L. Chau, *Quark mixing in weak interactions*, Physics Reports **95** (1983), no. 1 1 .

[236] BESIII collaboration, M. Ablikim *et al.*, *Measurements of absolute hadronic branching fractions of Λ_c^+ baryon*, Phys. Rev. Lett. **116** (2016), no. 5 052001, arXiv:1511.08380.

[237] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fractions for $B_{(s)} \to D_{(s)}\pi\pi\pi$ and $\Lambda_b^0 \to \Lambda_c^+\pi\pi\pi$*, Phys. Rev. **D84** (2011) 092001, Erratum ibid. **D85** (2012) 039904, arXiv:1109.6831.

[238] LHCb collaboration, R. Aaij *et al.*, *Measurement of the shape of the $\Lambda_b^0 \to \Lambda_c^+ \mu^- \overline{\nu}_\mu$ differential decay rate*, arXiv:1709.01920, submitted to Phys. Rev. D.

[239] H.-Y. Cheng and C.-K. Chua, *Strong Decays of Charmed Baryons in Heavy Hadron Chiral Perturbation Theory: An Update*, Phys. Rev. **D92** (2015), no. 7 074014, arXiv:1508.05653.

[240] H.-Y. Cheng, *Charmed baryons circa 2015*, arXiv:1508.07233.

References to LHCb internal documentation

[I1] J. Prisciandaro, F. Blanc, and T. Nakada, *Improved magnetic field map with 2011 measurements for the LHCb dipole magnet*, Tech. Rep. LHCb-INT-2012-012, CERN, Geneva, Mar, 2012.

[I2] J. Brehmer, J. Albrecht, and P. Seyfert, *Ghost probability: an efficient tool to remove background tracks*, Tech. Rep. CERN-LHCb-INT-2012-025, CERN, Geneva, Sep, 2012.

[I3] C. Jones, *ANN PID Retuning for Reco14 Data*, https://indico.cern.ch/event/226062/contributions/475644/attachments/371741/517276/ANNPIDRetuning-Reco14-06052013.pdf. Accessed: 2018-04-29.

[I4] A. Poluektov, *Correction of simulated particle identification response in LHCb using transformation of variables*, Tech. Rep. CERN-LHCb-INT-2017-007, CERN, Geneva, Apr, 2017. in preparation, draft at https://svnweb.cern.ch/cern/wsvn/lhcbdocs/Notes/INT/2017/007/drafts/LHCb-INT-2017-007_v0r0.pdf.

[I5] M. Cattaneo, *The new LHCb event data model*, Tech. Rep. LHCb-2001-142, CERN, Geneva, Nov, 2001.

[I6] M. Williams, *Generic D from B Selections*, Tech. Rep. CERN-LHCb-INT-2012-002, CERN, Geneva, Jan, 2012.

[I7] C. Vahl, S. Neubert, A. Piucci, and M. Stahl, *Observation of the decay $\Lambda_b^0 \to J/\psi \Lambda \phi$*, Tech. Rep. LHCb-ANA-2017-053, Sep, 2017.

[I8] M. Artuso *et al.*, *Measurement of the CP violating asymmetry a_{sl}^s*, Tech. Rep. LHCb-ANA-2016-004, May, 2016.

[I9] A. Cukierman, C. Dorothy, and M. Williams, *Baryonic double open charm decays*, Tech. Rep. LHCb-ANA-2013-078, Mar, 2014.

[I10] M. Stahl and M. Vesterinen, *Tracking asymmetry measurement with $J/\psi \to \mu^+\mu^-$ decays*, Tech. Rep. CERN-LHCb-INT-2015-012, CERN, Geneva, Feb, 2016.

[I11] R. Borchardt *et al.*, *Measurement of the $B^\pm$ production asymmetry and the CP asymmetry in $B^\pm \to J/\psi K^\pm$ decays*, Tech. Rep. LHCb-ANA-2015-030, May, 2015.

[I12] M. Stahl, *Charm spectroscopy from semileptonic b-baryon decays*, https://indico.cern.ch/event/361854/contributions/858011/attachments/720452/988964/2015-06-17-M_Stahl-CharmSpectroscopy_in_SLb_decays.pdf. Accessed: 2018-04-29.

References to code and public websites

[C1] 2D Adaptive Binning Doxygen documentation. https://www.physi.uni-heidelberg.de/~mstahl/2DAdaptiveBinning/html/annotated.html.

[C2] 2D Adaptive Binning repository. https://gitlab.cern.ch/mstahl/2DAdaptiveBinning. 2708cd213c2be44eeb816fdc5df9f7857798ee1c.

[C3] BackgroundCategory Class Reference. http://lhcb-doxygen.web.cern.ch/lhcb-doxygen/davinci/latest/db/d23/class_background_category.html#details. Accessed: 2018-04-29.

[C4] Bagging script for beef. https://gitlab.cern.ch/lhcb-bandq-exotics/Lb2LcD0K/blob/master/src/beef_bagging.cpp. 8467dafe28a1b7ce6e1dd0fb86972dc48dc84e0f.

[C5] beef. https://gitlab.cern.ch/sneubert/beef. 828514712c8e80443512caf4692b4b91ac5c8456.

[C6] Binary cut optimisation. https://gitlab.cern.ch/sneubert/ntuple-gizmo/blob/1e85a275611b8c236542627ac3950b39447b1d76/doc/CutOptScan.md. e734011557df2626a1016cdfab64c94ebec3e4a4.

[C7] Boost C++ libraries. http://www.boost.org/. Accessed: 2018-04-29.

[C8] boost property tree library, INFO parser. http://www.boost.org/doc/libs/1_65_1/doc/html/property_tree/parsers.html#property_tree.parsers.info_parser.

[C9] CERN GitLab. https://gitlab.cern.ch/. Accessed: 2018-04-29.

[C10] Code documentation of momentum asymmetry tool. https://gitlab.cern.ch/sneubert/ntuple-gizmo/blob/master/doc/misID_beta.md. 1e85a275611b8c236542627ac3950b39447b1d76.

[C11] D-from-B BDT repository. https://gitlab.cern.ch/sneubert/DfromBBDTs. f6a041a2cbe50082000ddcfdf7dd48a21098bf77.

[C12] EffDalitz. https://gitlab.cern.ch/apiucci/EffDalitz. a1b220751b86ed16cbe37ae2ae904d0d38ddae00.

[C13] effi Documentation. https://gitlab.cern.ch/sneubert/beef/blob/master/config/README_effi.md. b7330eee4481fdc718798a5c0948eef887274038.

[C14] $B_s^0 \to \mu^+\mu^-$ event displays 2011. https://lhcb-reconstruction.web.cern.ch/lhcb-reconstruction/Panoramix/PRplots/2011/event_displays_for_bs_mm.htm. Accessed: 2018-04-29.

[C15] Integrated effi toy MC. https://gitlab.cern.ch/sneubert/DfromBBDTs/blob/master/src/binned_eff_uncertainties.cpp. f7459c95d448cae056b4bb45e19a84a0f414f9de.

[C16] Integration of effi into beef. https://gitlab.cern.ch/sneubert/beef/merge_requests/82. 885572d795dc7a4fb3a070782e25bc4ad863802e.

[C17] IOjuggler. https://gitlab.cern.ch/mstahl/IOjuggler. f0e4b343a151f9c612186389785a0af5dba4206f.

[C18] Λ_c^+ cuts, D^0 cuts, D_s^+ cuts. https://gitlab.cern.ch/sneubert/DfromBBDTs/blob/master/config/fits/cutLcPi.info, https://gitlab.cern.ch/sneubert/DfromBBDTs/blob/master/config/fits/cutD0Pi.info,https://gitlab.cern.ch/sneubert/DfromBBDTs/blob/master/config/fits/cutDsPi.info. 5d68c078ac18740db264af806c1e459aa2e25f54.

[C19] LoKi::DistanceCalculator in LHCb Software Documentation. http://lhcb-doxygen.web.cern.ch/lhcb-doxygen/lhcb/latest/db/db5/class_lo_ki_1_1_distance_calculator.html. Accessed: 2018-04-29.

[C20] LoKi::Geometry in LHCb Software Documentation. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/latest_doxygen/d3/d1a/namespace_lo_ki_1_1_geometry.html. Accessed: 2018-04-29.

[C21] LoKi::KalmanFilter in LHCb Software Documentation. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/latest_doxygen/d3/dcd/namespace_lo_ki_1_1_kalman_filter.html#details. Accessed: 2018-04-29.

[C22] ntuple-gizmo. https://gitlab.cern.ch/sneubert/ntuple-gizmo. 76949b913a958e66fed5bef37d8aa998955235f2.

[C23] Parameters of the cut optimisation scan. https://gitlab.cern.ch/lhcb-bandq-exotics/Lb2LcD0K/blob/master/config/MVA/CutOptScan_cbs.info. afd4e655b344e901d3463cee2e84adf6526eda9a.

[C24] ProcessingPasses in LHCb TWiki. https://twiki.cern.ch/twiki/bin/view/Main/ProcessingPasses. r196 - 2017-12-26.

[C25] PropertimeFitter in LHCb Software Documentation. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/latest_doxygen/da/d05/class_propertime_fitter.html. Accessed: 2018-04-29.

[C26] Rates of Hlt2 lines in Hz. https://twiki.cern.ch/twiki/bin/view/LHCb/Run1HltRatesPage. Accessed: 2018-04-29.

[C27] RecoXX in LHCb TWiki. https://twiki.cern.ch/twiki/bin/view/LHCb/RecoXX. r23 - 2016-07-18.

[C28] C++ reference. http://en.cppreference.com/w/cpp. Accessed: 2018-04-29.

[C29] RooKeysPdf Documentation. https://root.cern.ch/doc/master/classRooKeysPdf.html. ROOT 6.06/02.

[C30] ROOT Data Analysis Framework. https://root.cern.ch/. Accessed: 2018-04-29.

[C31] Standalone effi toy MC. https://gitlab.cern.ch/mstahl/beef/blob/eff_errprop/src/effi_ep_toyMC.cpp. 8ffa5c50a0e3171041c9b6d1c617bafb84e304c5.

[C32] StrippingB02DPiD2HHHBeauty2CharmLine. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/config/stripping21/bhadroncompleteevent/strippingb02dpid2hhhbeauty2charmline.html. Accessed: 2018-04-29.

[C33] StrippingB2D0PiD2HHBeauty2CharmLine. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/config/stripping21/bhadroncompleteevent/strippingb2d0pid2hhbeauty2charmline.html. Accessed: 2018-04-29.

[C34] StrippingLb2LcDD2HHHPIDBeauty2CharmLine. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/config/stripping21/bhadron/strippinglb2lcdd2hhhpidbeauty2charmline.html. Accessed: 2018-04-29.

[C35] StrippingLb2LcPiNoIPLc2PKPiBeauty2CharmLine. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/config/stripping21/bhadroncompleteevent/strippinglb2lcpinoiplc2pkpibeauty2charmline.html. Accessed: 2018-04-29.

[C36] StrippingX2LcD0KD02KPiBeauty2CharmLine. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/config/stripping21/bhadron/strippingx2lcd0kd02kpibeauty2charmline.html. Accessed: 2018-04-29.

[C37] TEfficiency Documentation. https://root.cern.ch/doc/master/classTEfficiency.html. ROOT 6.06/02.

[C38] Tested misID combinations in $\Lambda_b^0 \to \Lambda_c^+ \bar{D}^{(*)0} K^-$ and $\Lambda_b^0 \to \Lambda_c^+ D_s^-$. https://gitlab.cern.ch/lhcb-bandq-exotics/Lb2LcD0K/blob/master/scripts/Selection/LcD0KSelector.C#L304, https://gitlab.cern.ch/lhcb-bandq-exotics/Lb2LcD0K/blob/master/scripts/Reference/LcDsSelector.C#L146. Accessed: 2018-04-29.

[C39] The Brunel project. http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/. Accessed: 2018-04-29.

[C40] The DaVinci project. http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/. Accessed: 2018-04-29.

[C41] The Gauss project. http://lhcbdoc.web.cern.ch/lhcbdoc/gauss/. Accessed: 2018-04-29.

[C42] The Moore project. http://lhcbdoc.web.cern.ch/lhcbdoc/moore/. Accessed: 2018-04-29.

[C43] The Stripping Project. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/. Accessed: 2018-04-29.

[C44] Types of LoKi cuts. https://twiki.cern.ch/twiki/bin/view/LHCb/LoKiHybridFilters#Mother_Cut_Daughter_Cut_Combinat. Accessed: 2018-04-29.