

Search for Rare b to Open-Charm Two-Body Decays of Baryons at LHCb



Dissertation

for the degree of
doctor rerum naturalium (Dr. rer. nat.)

submitted by
M.Sc. Nis Meinert
to the Faculty of Mathematics and Natural Sciences
of the University of Rostock, Germany

Rostock, Germany
Juni 2020

Referees:

Privatdozent Dr. Roland Waldi

Institut für Physik, Universität Rostock
Albert-Einstein-Str. 23
D-18059 Rostock, Germany

Prof. Dr. Bernhard Spaan

Experimentelle Physik 5, Technische Universität Dortmund
Otto-Hahn-Str. 4a
D-44227 Dortmund, Germany

Opening of the doctoral degree procedure: 24.06.2020

Date of scientific colloquium: 09.10.2020

“Everything should be built top-down, except the first time.”
— Alan Perlis, first recipient of the Turing Award.

Abstract

A search for the rare two-body decays $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$ is performed with proton-proton collision data, corresponding to an integrated luminosity of 6 fb^{-1} , collected by the LHCb experiment at a center-of-mass energy of 13 TeV. The decay $\Lambda_b \rightarrow D^0 \Lambda$ is seen with a statistical significance of 5.5 standard deviations, and constitutes the discovery for this decay. The branching fraction, measured using the $\Lambda_b \rightarrow D^0 p\pi^-$ decay for normalization, is

$$\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda) = (9.9 \pm 2.3 \pm 1.6 \pm 1.1) \times 10^{-6},$$

where the uncertainties are statistical, systematic, and external, respectively. An excess of $\Xi_b^0 \rightarrow D^0 \Lambda$ candidates w.r.t. the background is observed with a statistical significance of 1.8 standard deviations and is used to estimate the upper limit

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)} < 0.5 \quad (\text{CL} = 95\%),$$

where $f_{\Xi_b^0}/f_{\Lambda_b}$ is the ratio of the fragmentation fractions of b -quarks into Ξ_b^0 and Λ_b baryons.

Kurzfassung

In dieser Arbeit wird eine Suche nach den seltenen Zweikörper-Zerfällen $\Lambda_b \rightarrow D^0 \Lambda$ und $\Xi_b^0 \rightarrow D^0 \Lambda$ mit Proton-Proton Kollisionen präsentiert. Der analysierte Datensatz entspricht einer integrierten Luminosität von 6 fb^{-1} und wurde durch das LHCb Experiment bei einer Schwerpunktsenergie von 13 TeV aufgezeichnet. Der Zerfall $\Lambda_b \rightarrow D^0 \Lambda$ wird mit einer statistischen Signifikanz von 5,5 Standardabweichungen beobachtet und ist somit als Neuentdeckung einzustufen. Das gemessene Verzweigungsverhältnis mit statistischem, systematischem und externem Fehler, normiert mit gemessenen $\Lambda_b \rightarrow D^0 p\pi^-$ Zerfällen, ist

$$\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda) = (9.9 \pm 2.3 \pm 1.6 \pm 1.1) \times 10^{-6}.$$

Eine Anhäufung von $\Xi_b^0 \rightarrow D^0 \Lambda$ Kandidaten gegenüber dem Untergrund wird mit einer statistischen Signifikanz von 1,8 Standardabweichungen beobachtet und zur Berechnung einer Obergrenze benutzt:

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)} < 0.5 \quad (\text{CL} = 95\%).$$

Dabei bezeichnet $f_{\Xi_b^0}/f_{\Lambda_b}$ das Verhältnis der Fragmentierungsanteile von b Quarks in Ξ_b^0 und Λ_b Baryonen.

Introduction

The idea of simplifying the description of different types of matter by introducing fundamental substructures is old. In Ancient Greece the *elements* earth, water, air and fire were assumed to be fundamental. In the early 1800s, John Dalton used the concept of atoms to explain why elements always react in ratios of small whole numbers. Later, in the early 1900s, Ernest Rutherford discovered the nucleus. Today, the concept of protons, neutrons and electrons building different types of atoms is well established and taught in school. The standard model of particle physics (SM) introduces a new layer of fundamental particles. Particles like the proton and the neutron are hadrons, composed of quarks. The electron stays fundamental and joins the group of the fundamental leptons. There are more quarks than one finds within protons and neutrons, and there are more fundamental leptons than the electron. However, these particles are not stable and decay until only the lightest particles are left, *i.e.*, in nature we only observe protons, neutrons and electrons most of the time. The proton is a baryon and the only stable hadron. The neutron is a long-living particle but eventually decays via the β -decay,

$$n \rightarrow p e^- \bar{\nu}_e.$$

Mesons are another possible combination of quarks. Instead of consisting of three quarks (baryon), they are made of one quark and one anti-quark. Whereas there is one stable baryon, all mesons are unstable. The lightest mesons are the pion triplet (π^+ , π^0 , π^-) and their dominant decay modes are:

$$\begin{aligned}\pi^0 &\rightarrow 2\gamma, \\ \pi^+ &\rightarrow \mu^+ \bar{\nu}_\mu, \\ \pi^- &\rightarrow \mu^- \nu_\mu,\end{aligned}$$

where the final states of the charged (neutral) pions are leptons (bosons).

Classical atomic models do not need mesons, since their physics is dominated by the interaction between the protons, neutrons and electrons. The lion's share of the visible matter in our universe is baryonic and not mesonic. On the one hand, it is the CP violation in meson decays that is well studied at colliders but, on the other hand, one of the unsolved mysteries of our universe is its large baryon anti-baryon asymmetry (and not meson anti-meson asymmetry) which can be considered as the largest known macroscopic CP violation. The SM of particle physics (as well as the SM of cosmology) yet have failed explaining the order of magnitude of this asymmetry, whereas at the same time results of CP violation in meson decays are in great agreement with the SM predictions.

In the present analysis we contribute to the experimental foundation for studies of CP violations in baryon decays by searching for two-body decays of the Λ_b and the Ξ_b^0 baryon which can be used at future experiments to estimate the CP violation parameter γ .

Conventions

After spending years in the field of experimental high energy physics, adopting a certain kind of jargon and unspoken conventions is unavoidable. The author tried his best to avoid jargon when possible and summarize abbreviations in a glossary on page 103.

Throughout this work we will frequently refer to invariant masses, *e.g.*, as $m(\Lambda_b)$ or $m(D^0\Lambda)$. Whereas the former typically refers to the measured mass of the given hadron (the Λ_b baryon in this example), the latter notation is meant as an abbreviation: It should be read as the result of summing the invariant masses of a D^0 candidate and a Λ candidate, according to four-momentum addition where *candidates* are themselves recursively obtained by such four-momentum additions. We note that in particular for recorded data, this does not only include the desired decay channel ($\Lambda_b \rightarrow D^0\Lambda$ in the above example), but potentially also include random track combinations and other kind of background contributions. On top of this, our notation of invariant masses also always implicitly includes the respective CP conjugated particles, *e.g.*, the invariant mass distribution $m(D^0\Lambda)$ is the result of a four-momentum addition of D^0 and Λ candidates (*e.g.*, $\Lambda_b \rightarrow D^0\Lambda$), as well as \bar{D}^0 and $\bar{\Lambda}$ candidates (*e.g.*, $\bar{\Lambda}_b \rightarrow \bar{D}^0\bar{\Lambda}$).

Contents

1 Theory	1
1.1 CP Violation	4
1.2 Quark Model	5
1.3 Strong Isospin	7
1.4 $\Sigma^0 \rightarrow \Lambda\gamma$ Background	9
1.4.1 Limitations of naïve SU(2) Arguments	9
1.4.2 Background Contamination by $\Sigma^0 \rightarrow \Lambda\gamma$ at LHCb	10
1.4.3 Summary	11
1.5 CP Violation in b -Baryon Decays	13
2 The LHCb Detector at the LHC	15
2.1 Experimental Setup	15
2.1.1 Tracking	16
2.1.2 Ring Imaging Cherenkov Detectors	17
2.1.3 Calorimeters	17
2.1.4 Muon System	17
2.1.5 Dipole Magnet	18
2.1.6 Trigger System	18
2.1.7 PID	19
2.2 Variables and Acronyms used as Selection Criteria	20
3 Reconstruction and Stripping Efficiencies	23
4 Tuning MC Simulated Events by Determining Weights	27
4.1 The Decay $\Lambda_b \rightarrow J/\psi \Lambda$	28
4.1.1 Preselection	28
4.1.2 Loose Selection	28
4.1.3 Tight Selection	29
4.2 Extraction of Weights	33
4.2.1 Truth Matched vs. Sideband Subtracted MC Simulated Events	34
4.2.2 Scheme 1	35
4.2.3 Scheme 2	40
5 MVA of the Decay $\Lambda_b \rightarrow D^0 \Lambda$	45
5.1 Preprocessing	46
5.1.1 Preselection	46
5.1.2 Loose Selection	46
5.2 Tight Selection using MVA Techniques	50
5.2.1 The Data Pipeline	50
5.2.2 The Λ Classifier	51
5.2.3 The Λ_b - D^0 Classifier	67
5.2.4 The Tier 2 Classifier	70

6 The Normalization Channel $\Lambda_b \rightarrow D^0 p \pi^-$	77
6.1 The Decay $\Lambda_b \rightarrow D^0 p \pi^-$	78
6.1.1 Preselection	78
6.1.2 Loose Selection	78
6.1.3 Calibration	78
6.2 Yield Extraction	79
7 Physical Backgrounds	83
7.1 Non-Resonant Background	83
7.2 Partially Reconstructed Backgrounds	85
7.3 Reflections	86
7.3.1 Charmless Decays	86
7.3.2 K_s Reflections from b -Meson Decays	88
8 Yield Estimation	91
8.1 The Fit Model	91
8.2 Yield Extraction	94
8.3 Validation of Yield Significances with Pseudo-Experiments	96
9 Estimation of Branching Ratios	97
9.1 Branching Ratio $\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)/\mathcal{B}(\Lambda_b \rightarrow D^0 p \pi^-)$	97
9.2 Branching Ratio $\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)/\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)$	100
9.3 Summary and Outlook	102
Glossary	103
Bibliography	107
A Reconstruction and Stripping Efficiencies	115
B Stripping Selections	119
C Decay Tree Fit	123
D Fit Probability	127
E Cubic Splines and their derivatives	129
F Weighting Scheme 1 Supplementaries	133
G Confusion Matrix	137
H Cross-Validation	139
I Principal Component Analysis	141
J Support Vector Machines	143
K Ensemble Learning	145
L MVA Supplementaries	147
M Cross-Checking the Efficiency Estimates of MVA Features	155

N Kinematics of the Charmless Three-Body Decay $\Lambda_b \rightarrow \Lambda K^+ K^-$	161
O Partially Reconstructed Backgrounds	167
P Fit Supplementaries	173
Q The Exponential Function as a PDF	183
R Clipped Gaussian Distribution	185
S Estimations of a CI for the Ξ_b^0-Λ_b Ratio	187

Chapter 1

Theory

When working toward the solution of a problem, it always helps if you know the answer.

— Rule of Accuracy, from Arthur Bloch's book *Murphy's Law*.

The standard model of particle physics (SM) is a theory which describes three of the four known fundamental interactions, *i.e.*, the strong, the weak and the electromagnetic interaction. Gravity is not yet included in this theory, but since its coupling is weak w.r.t. the other fundamental interactions at scales of particle accelerator energies, the impact can be neglected for all processes within the present analysis. The standard model uses the framework of a quantum field theory to describe the dynamics of all known fundamental particles. The description is Lorentz invariant, obeys locality¹ and probability conservation, and is thus considered a *healthy* theory.

The (fundamental) particles are quarks, like the u - and d -quark forming the hadronic matter (*e.g.*, protons and neutrons), leptons (*e.g.*, electrons and electron neutrinos), and bosons (*e.g.*, photons). Within the SM, these particles are the fundamental excitations of respective fields.

The gauge bosons of this theory arise from Yang-Mills fields. In total, there are eight massless bosons of the strong interaction (gluons), three massive bosons (W^\pm and Z) and one massless boson (photon) of the electro-weak interaction. One additional boson arises from the Higgs mechanism and brings masses to all fundamental particles which do interact with this boson. The fundamental fermions of the standard model are arranged in multiplets based on the local symmetries given by the respective Yang-Mills theory:

$$\begin{aligned} \ell_L^i &= (\mathbf{1}, \mathbf{2})_{-1} & i &= 1, 2, 3, \\ \ell_R^i &= (\mathbf{1}, \mathbf{1})_{-2} & i &= e^-, \mu^-, \tau^-, \\ q_L^i &= (\mathbf{3}, \mathbf{2})_{\frac{1}{3}} & i &= 1, 2, 3, \\ u_R &= (\mathbf{3}, \mathbf{1})_{\frac{4}{3}} & u &= u, c, t, \\ d'_R &= (\mathbf{3}, \mathbf{1})_{-\frac{2}{3}} & d &= d, s, b, \end{aligned}$$

where the subscripts L and R indicate whether the particle is left (L) or right (R) handed, and the bold faced numbers are the representation of the gauge group of the strong interaction and the weak isospin, respectively. The lower number indicates the weak hypercharge. The electric charge q of each particle can be determined by the Gell-Mann-Nishijima formula [1, 2], yielding a charge of $q = 2/3|e|$ for the up-type quarks u, c, t and $q = -1/3|e|$ for the down-type quarks d', s', b' . The leptons e^- , μ^- and τ^- are charged equally with $q = -|e|$, whereas the neutrinos are neutral and therefore do not couple electro-magnetically. Every fermion comes with one anti-fermion. From a group theoretical point of view, these particles originate from the conjugated fermion representation, whereas from a field theoretical point of view they are obtained by mirroring the charge (C) and the parity (P) of the respective fermion.

¹The action only contains terms in which the fields and their derivatives are evaluated at the same space-time point.

In total, there are six left- and three right-handed leptons, where the former are arranged in weak isospin doublets, *i.e.*, (e_L^-, ν_e) , (μ_L^-, ν_μ) and (τ_L^-, ν_τ) and the latter appear in singlets (e_R^-) , (μ_R^-) , (τ_R^-) .

The six known quarks u, c, t, d', s', b' also appear left and right handed. Again, the latter appears in weak iso-singlets, whereas the former are arranged in weak iso-doublets, *i.e.*, (u, d') , (c, s') and (t, b') . The weak interaction describes the transitions of the left-handed up-type quarks (u, c, t) to the left-handed down-type quarks (d', s', b') . This conversion within the weak iso-doublets is mediated via the coupling between the W^\pm bosons and the charged weak current

$$\mathcal{J}_\mu^{\text{CC}} = (\bar{u}, \bar{c}, \bar{t}) \gamma_\mu (1 - \gamma_5) \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}.$$

In the next section we will elaborate that (d', s', b') are the weak flavor eigenstates and differ from the mass eigenstates of the quarks. Taking this into account leads to the introduction of an unitary matrix that connects flavor and mass eigenstates. This matrix is called the CKM matrix and brings four additional degrees of freedom, *i.e.*, three mixing angles and one CP violating phase.

The gauge groups are described within the framework of semisimple Lie algebras such that the total gauge symmetry of the standard model is:

$$\mathcal{F} = \text{SU}(3) \times \text{SU}(2) \times \text{U}(1),$$

where $\text{SU}(3)$ is the gauge group of the strong interaction and $\text{SU}(2) \times \text{U}(1)$ is the respective gauge group of the electro-weak interaction. The Higgs mechanism (spontaneously) breaks this symmetry down to

$$\mathcal{F} = \text{SU}(3) \times \text{SU}(2) \times \text{U}(1) \xrightarrow{\text{Higgs}} \text{SU}(3) \times \text{U}(1)_{\text{em}}.$$

The Lagrangian of the standard model reads:

$$\mathcal{L} = -\frac{1}{2} \text{tr}(\mathbf{F}_{\mu\nu} \mathbf{F}^{\mu\nu}) + \bar{\psi} i \not{D} \psi + \mathcal{L}_{\text{Yuk}} + (\mathbf{D}_\mu \phi)^2 - V(\phi),$$

where we abbreviated

$$\mathbf{F}_{\mu\nu} \mathbf{F}^{\mu\nu} = \mathbf{G}_{\mu\nu} \mathbf{G}^{\mu\nu} + \mathbf{W}_{\mu\nu} \mathbf{W}^{\mu\nu} + \frac{1}{2} \mathbf{B}_{\mu\nu} \mathbf{B}^{\mu\nu}$$

and introduced the covariant derivative

$$\mathbf{D}_\mu \psi = (\partial_\mu + i g_s \mathbf{G}_\mu + i g \mathbf{W}_\mu + i g' \mathbf{B}_\mu) \psi.$$

All fermions fields are gathered in ψ , whereas ϕ is the Higgs field, coupling left and right handed fermions via a Yukawa coupling \mathcal{L}_{Yuk} with an according potential $V(\phi)$. The tensors $\mathbf{G}_{\mu\nu}$, $\mathbf{W}_{\mu\nu}$ and $\mathbf{B}_{\mu\nu}$ are the field-strength tensors of the gauge bosons. The former tensor corresponds to the strong interaction, whereas the latter tensors correspond to the electro-weak interaction. All three tensors are functions of the corresponding gauge fields and of the generator of their gauge groups, *e.g.*,

$$\begin{aligned} \mathbf{G}_{\mu\nu} &= \partial_\mu \mathbf{G}_\nu - \partial_\nu \mathbf{G}_\mu + i g_s [\mathbf{G}_\mu, \mathbf{G}_\nu] \\ &= (\partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c) \frac{\lambda_a}{2}, \end{aligned}$$

where G_μ^a are the gluon fields (index a takes values $1 \dots 8$), f_{abc} is the structure constant and λ_a are the generator of the SU(3) group. The fields G_μ^a are chosen such that the transformation of the field strength tensor is:

$$\mathbf{G}_{\mu\nu}(x) \xrightarrow{\text{SU}(3)} \mathbf{U}(x) \mathbf{G}_{\mu\nu}(x) \mathbf{U}^\dagger(x),$$

with $\mathbf{U}(x) \equiv U(x)$ being an arbitrary transformation of SU(3), thus leaving the total Lagrangian invariant. We note that applying the trace operation to $\mathbf{G}_\mu \equiv \mathbf{G}_\mu^\dagger = G_\mu^a(x) \frac{\lambda_a}{2}$ only affects the generator space

$$\text{tr } \mathbf{G}_\mu \equiv G_\mu^a(x) \text{tr } \frac{\lambda_a}{2},$$

thus yielding

$$\text{tr} (\mathbf{G}_{\mu\nu} \mathbf{G}^{\mu\nu}) = G_{\mu\nu}^a G_b^{\mu\nu} \text{tr} \frac{\lambda_a \lambda_b}{4} = \frac{1}{2} G_{\mu\nu}^a G_a^{\mu\nu}.$$

The same holds for the electro-weak case where the generators of the electro-weak $\text{SU}(2) \times \text{U}(1)$ group are the weak isospin $\sigma_a/2$ (Pauli matrices) and the weak hypercharge Y .

While the gauge symmetry of the strong interaction is unbroken and all gluons are indistinguishable, the electro-weak sector of the standard model is broken and the respective field excitations become distinguishable. The physical fields are the charge eigenstates \mathbf{W}_μ^+ , \mathbf{W}_μ^- , \mathbf{Z}_μ , and the photon field \mathbf{A}_μ . They are obtained via the transformations

$$\begin{aligned} \mathbf{W}_\mu^\pm &= \mathbf{W}_\mu^1 \pm i \mathbf{W}_\mu^2, \\ \begin{pmatrix} \mathbf{Z}_\mu \\ \mathbf{A}_\mu \end{pmatrix} &= \begin{pmatrix} \cos \vartheta_W & -\sin \vartheta_W \\ \sin \vartheta_W & \cos \vartheta_W \end{pmatrix} \begin{pmatrix} \mathbf{W}_\mu^3 \\ \mathbf{B}_\mu \end{pmatrix}. \end{aligned}$$

This coupling also affects their coupling constants g and g' , hence it is sensible to introduce the weak mixing angle ϑ_W as

$$e = g \sin \vartheta_W = \frac{gg'}{\sqrt{g^2 + g'^2}},$$

where e is the electric charge. With this, the (classical²) standard model of particle physics has 18 free parameters in total,

- 3 couplings: g_s , e , $\sin \vartheta_w$,
- 2 boson masses: m_W , m_H ,
- 3 lepton masses: m_e , m_μ , m_τ ,
- 6 quark masses: m_u , m_d , m_s , m_c , m_t , m_b ,
- 4 parameters of the CKM matrix.

For the sake of completeness we note that we did not account for any affects from renormalization in the Lagrangians shown above, although the standard model is a renormalizable theory. Renormalization changes the behavior of charges and couplings and make these quantities effectively momentum dependent. When doing actual calculations it is mandatory to take this into account, not least because of observable properties such as the running of coupling constants. However, the topic of renormalization is complex and will neither significantly enrich this brief overview substantively, nor yield deeper insights into the presented analysis.

²Without neutrino masses.

1.1 CP Violation

Above, we have seen that the charged current $\mathcal{J}_\mu^{\text{CC}}$ couples the weak eigenstates. This coupling is diagonal such that there is no mixing of the eigenstates. The flavor eigenstates, however, do not correspond to the mass eigenstates, but there is one arbitrary rotation between flavor and mass eigenstates of the up-type quarks and the down-type quarks each. These two rotations cancel for all terms of the Lagrangian, except for the charged current $\mathcal{J}_\mu^{\text{CC}}$, since this is the only term connecting up- and down-type quarks. In $\mathcal{J}_\mu^{\text{CC}}$ both rotations appear in product, thus collapsing to one unitary matrix V :

$$\mathcal{J}_\mu^{\text{CC}} = (\bar{u}', \bar{c}', \bar{t}') \gamma_\mu (1 - \gamma_5) \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = (\bar{u}, \bar{c}, \bar{t}) \gamma_\mu (1 - \gamma_5) V \begin{pmatrix} d \\ s \\ b \end{pmatrix},$$

where the prime indicates mass eigenstates.

The matrix V is called the CKM matrix and was first introduced by Kobayashi and Maskawa [3] as an extension of the two-dimensional Cabibbo matrix [4]. The CKM matrix is parametrized by three mixing angles and one CP violating phase. Conventionally, one multiplies V to the right, thus mixing down-type mass eigenstates. When referring to mass eigenstates, each up-type quark now decays to a superposition of three down-type quarks:

$$\begin{pmatrix} u' \\ c' \\ t' \end{pmatrix} = \begin{pmatrix} u \\ c \\ t \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.$$

The probability of finding a down-type quark after the decay of an up-type quark u_i in a specific mass eigenstate d_j is proportional to the square of the respective matrix element V_{ij} . Since these matrix elements are complex they differ from the decay of the respective anti-quarks in a complex phase, *i.e.*, the transition $\bar{u}_i \rightarrow \bar{d}_j$ is proportional to the square of V_{ij}^* . In the absence of multiple decay channels, this phase is unphysical and vanishes after squaring the matrix element. In the case of multiple decay channels, all joining the same final state but bringing different strong phases, this phase yields a physical effect in different real valued decay amplitudes between particles and the respective anti-particles. This effect is called direct CP violation.

The CKM matrix exhibits a clear hierarchy, often expressed in the Wolfenstein parameterization [5]:

$$\begin{aligned} V &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\ &= \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \end{aligned}$$

where the four parameters (A, λ, ρ, η) are real numbers. Here, $\lambda \approx 0.23$ is the sine of the Cabibbo angle and plays the role of an expansion parameter, whereas η represents the CP violating phase. Note, that this representation only assigns complex phases to the matrix elements V_{ub} and V_{td} . Expanding the CKM matrix in higher orders of λ , brings additional phases to three other matrix elements.

The CKM matrix is unitary, thus yielding twelve distinct complex relations among the matrix elements:

$$(VV^\dagger)_{ij} = (V^\dagger V)_{ij} = \delta_{ij}.$$

Six of these twelve relations ($i \neq j$) can be represented geometrically as triangles in the complex plane, *e.g.*,

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0,$$

where the length of the edges are given by the absolute values of each addend. The three angles are given by the argument of the ratio of neighboring edges, *i.e.*,

$$\begin{aligned}\alpha &= \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right) \approx \arg\left(-\frac{1-\rho-i\eta}{\rho+i\eta}\right), \\ \beta &= \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right) \approx \arg\left(-\frac{1}{1-\rho-i\eta}\right), \\ \gamma &= \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right) \approx \arg(\rho+i\eta).\end{aligned}$$

The unitary of the CKM matrix implies $\alpha + \beta + \gamma = \pi$, as well as equal areas of all triangles. Any discrepancy of these relations would indicate a non unitary CKM matrix and thus corresponding to physics beyond the standard model. At the time of writing, the CKM angle γ is among the least well known constraints, therefore, being an excellent probe of new physics [6, 7].

At this point it seems appropriate to mention, that neutrinos are set massless within in the classical standard model, thus excluding lepton mixing. Ever since the discovery of neutrino oscillation [8, 9, 10] we yet know that neutrinos must have finite masses and therefore leptons mix also. The corresponding mixing matrix is called the PMNS matrix [11] and allows CP violation for leptons. (See *e.g.*, Ref. [12] for some exciting results for CP violation measurements in neutrino oscillations.) There are at least two (conflicting) extensions to the classical standard model that explain the finite neutrino masses but since there is no clear preference from experimental data which of these extensions can be excluded, we did not account for them in our previous parameter counting. Anyhow, finite neutrino masses will not influence the results of the present analysis since we are dealing with pure hadronic states.

Another source of CP violation could occurs in strong decays. It is easy to extend the standard model accordingly and this extension is occasionally considered *more natural*. Experiments show, however, that the strong interaction seems to conserve CP . This conservation is not yet fully understood and is known as the *Strong CP Problem* in literature [7, 13, 14], but appears in perfect agreement with all experimental results so far. Therefore, the only source of CP violation in the classical standard model are weak decays.

1.2 Quark Model

Quantum Chromodynamics (QCD) is the quantum field description of the strong interaction that enlarges the pool of quantum numbers of the standard model. All quarks now possess the new quantum number *color* which can take on three different values. QCD is an unbroken $SU(3)$ gauge symmetry, acting on the color quantum number. The eight generators of this Lie-Algebra induce eight massless gauge bosons, called *gluons*.

QCD is a non-abelian gauge theory, hence gluons are charged and interact with each other. Similar to Quantum Electrodynamics (QED), there exist quark loops which screen the (color) charge at large distances. In marked contrast to QED the gauge bosons of QCD carry charge themselves, therefore allowing gluon loops at the same level. The gluon loops anti-screen and dominate the quark loops, hence QCD becomes stronger at large distances. This feature of QCD confines quarks and gluons such that none of them can be isolated. Only states forming color singlets can be observed at large distance scales, *i.e.*, the possible combinations of quarks q and anti-quarks \bar{q} are mesons ($q\bar{q}$), baryons (qqq), tetraquarks ($qq\bar{q}\bar{q}$) and pentaquarks ($qqqq\bar{q}$), as well as their respective anti-particles. These combinations are called *hadrons*. Furthermore, combinations of two gluons forming a color singlet (glueballs) are possible.

The masses of quarks forming a hadron are much lighter than the masses of the respective hadron. The reason for that can be found in the two different mechanisms providing masses to quarks and to hadrons. The former originates directly from the spontaneous symmetry breaking by the Higgs mechanism, the latter originates from the energy of quark and gluon condensates of the QCD vacuum. The QCD vacuum includes virtual $q\bar{q}$ pairs that also contribute to the absolute mass of the hadron. In order to distinguish between these and the valence quarks, we call these virtual quarks *sea quarks*.

The quark model of particle physics is a classification scheme for hadrons in terms of their valence quarks. The quantum numbers of the quarks are added, obeying the rules of spin algebras, such that each hadron is uniquely defined by their sum. Besides spin, quarks have another charge of a SU(2) symmetry, *isospin*. The two lightest quarks u and d are arranged in iso-doublets, all other quarks are iso-singlets. This quantum number turns out to be useful, since QCD is unbroken in terms of isospin, thus giving rise to a conserved quantity. (Different masses and electrical charges break this symmetry in the electro-weak sector of the standard model.) Adding the isospin of quarks, one finds the isospin of hadrons by analyzing the outer product, *e.g.*, combining a u - and d -quark (taken from the (u, d) doublet **2**), and a s -quark (singlet **1**) yields,

$$\mathbf{2} \otimes \mathbf{2} \otimes \mathbf{1} = \mathbf{3} \oplus \mathbf{1},$$

a triplet and a singlet. The triplet (isospin $I = 1$) consists of the three baryons Σ^+ (uus), Σ^0 (uds) and Σ^- (dds), the singlet ($I = 0$) is the Λ baryon (uds) where we have given the valence quark contents in parentheses. Note, that the Σ^0 and the Λ have the same valence quark content, but different isospin. Therefore they are distinguishable. Another prominent example of an isospin pair is the nucleon, having isospin $I = 1/2$. The two states, that differ in their valence quark content but are anyhow indistinguishable in strong interactions, are the proton (uud) and the neutron (udd).

Below, we give a short overview over the intermediate particles Λ_b , Ξ_b^0 , Λ and D^0 which play a major role in the present analysis. All figures, even though not explicitly marked, are taken from the Particle Data Group [15]:

Λ_b Baryon In the quark model, the Λ_b baryon is an iso-singlet state, consisting of valence quarks u , d and b . The lightest Λ_b baryon has $I(J^P) = 0(1/2^+)$. The mass is measured as

$$m(\Lambda_b) = 5.61960(17) \text{ GeV}/c^2,$$

thus it is the lightest baryon with one b - or \bar{b} -quark. Energy conservation does not allow for any strong decays, therefore the Λ_b decays weakly which explains its relatively long lifetime $\tau(\Lambda_b)$,

$$\tau(\Lambda_b) = 1.471(9) \text{ ps},$$

corresponding to a narrow natural width of roughly $\hbar/\tau \approx 0.4 \text{ meV}$. The dominant decay modes are $\Lambda_b \rightarrow \Lambda_c \ell^- \bar{\nu}_\ell$ and $\Lambda_b \rightarrow \Lambda_c D_s^-$ with a relative decay rate of $10.3(2.2)\%$ and $1.10(0.10)\%$, respectively. The decays $\Lambda_b \rightarrow D\Lambda$ and $\Lambda_b \rightarrow D^*\Lambda$ that we study in the present analysis have not yet been observed.

Ξ_b^0 Baryon In the quark model the Ξ_b baryons, Ξ_b^0 (bsu) and Ξ_b^- (bsd), are the constituents of an iso-doublet with $I(J^P) = 1/2(1/2^+)$. Both resonances were discovered at the Tevatron collider in the years 2011 and 2007 [16, 17, 18] (more than a quarter of a century after the Λ_b discovery [19]). The mass of the Ξ_b^0 baryon is measured as

$$m(\Xi_b^0) = 5.7919(5) \text{ GeV}/c^2,$$

which makes it $172.5(4) \text{ MeV}/c^2$ heavier than the Λ_b baryon and thus forbids any strong decays such as $\Xi_b^0 \rightarrow \Lambda_b \bar{K}^0$. In the framework of heavy quark expansion all weakly decaying charmless b -hadrons have the same lifetime up to corrections at the order $(\Lambda_{\text{QCD}}/m_b)^2$,

where Λ_{QCD} is the energy scale at which the strong coupling becomes large and m_b is the mass of the b -quark [20]. The measured mean lifetime ratio of the Ξ_b^0 and the Λ_b hadron, $\tau(\Xi_b^0)/\tau(\Lambda_b) = 1.006(21)$ [21], confirms this expectation. None of the (weak) decays into the final states $D^0\Lambda$, $D^*\Lambda$, $D^0\Sigma^0$ or $D^0\Xi^0$, that we consider as signal or (potential) background contributions in the present analysis, have yet been observed.

Λ Baryon We have discussed the Λ baryon and its role in the quark model previously. The mass is measured as

$$m(\Lambda) = 1.115683(6) \text{ GeV}/c^2.$$

The dominant decay modes of the $I(J^P) = 0(1/2^+)$ baryon are $\Lambda \rightarrow p\pi^-$ and $\Lambda \rightarrow n\pi^0$ with branching fractions of 63.9(5)% and 35.8(5)%, respectively. Whereas the latter mode is almost impossible to reconstruct at LHCb, the reconstruction of the former is feasible, albeit with the caveat that due to the long lifetime of the V^0 particle,

$$\tau(\Lambda) = 263.2(20) \text{ ps},$$

most decays will take place outside of the VELO (*cf.* Chap. 2).

D^0 Meson The D^0 is a $I(J^P) = 1/2(0^-)$ meson, consisting of valence quarks c and \bar{u} with a mass of

$$m(D^0) = 1.86483(5) \text{ GeV}/c^2.$$

The PDG lists more than 300 measured (weak) decay modes of the D^0 meson which explains its relatively short lifetime (in comparison with other weakly decaying hadrons),

$$\tau(D^0) = 0.4101(15) \text{ ps}.$$

Modes of particular interest for the present analysis (and at future prospects) are the $D^0 \rightarrow hh'$ modes, where h and h' refer to pions or kaons, and $D^0 \rightarrow K^\mp\pi^\pm\pi^+\pi^-$ modes. The mode $D^0 \rightarrow K^-\pi^+$ (*i.e.*, $h = K^-$ and $h' = \pi^+$) is referred to as the Cabibbo favored mode due to its $c \rightarrow s u \bar{d}$ transition (proportional to $1 - \lambda^2$ in the Wolfenstein parameterization). Similarly, the modes $D^0 \rightarrow K^+K^-$ and $D^0 \rightarrow \pi^+\pi^-$ are referred to as Cabibbo suppressed (proportional to λ) and finally $D^0 \rightarrow K^+\pi^-$ is referred to as the Cabibbo doubly suppressed mode (proportional to λ^2).

1.3 Strong Isospin

In contrast to the electro-weak interaction that breaks various symmetries by allowing flavor transitions, QCD appears to be invariant under several symmetries. These symmetries can be categorized into gauge symmetries, *e.g.*, color SU(3), accidental symmetries, such as the baryon number conservation, and approximate symmetries. The (strong) isospin is an approximate symmetry of the SM, but can be considered a conserved symmetry for QCD. Quarks multiplets as shown above are broken in the context of the complete SM due to finite mass differences of the u - and d -quarks, introduced by the Yukawa couplings (Higgs mechanism), and different electric charges of up and down type quarks. Nevertheless, isospin turned out to be useful as an approximately conserved quantum number in certain contexts, for example in nucleon nucleon scattering processes, in predicting states of the quark model as part of the Eightfold Way or as starting point of various lattice QCD calculations [22, 23, 24, 25].

In general, assuming isospin invariance of QCD with finite quark masses is a good approximation, since the quark masses of the u - and d -quark are a small perturbation on the hadronic scales of QCD, making the corresponding baryons proton and neutron almost degenerated. Isospin invariance makes the Hamiltonian of the strong interaction a scalar in iso-space and

thus gives rise to a special case of the Wigner-Eckart theorem which states that strong decay ratios are governed by Clebsch-Gordan (GC) coefficients.

The CG coefficients can be used as a proxy for testing the precision of the isospin approximation either qualitatively for explaining strong suppression of decays (corresponding to vanishing CG coefficients) or even quantitatively for predicting branching fractions. A canonical examples for the former is the decay $\rho^0 \rightarrow \pi^0\pi^0$ (not yet seen³) and (more recently) $\Lambda_b \rightarrow J/\psi \Sigma^0$ [26]. Examples for the latter are the production of Δ^{++} and Δ^0 resonances in $p\pi^-$ and $p\pi^+$ collisions (*cf.* Fig. 1.1 with data taken from Ref. [27]) and K^* decays into $K\rho$ and $K\omega$. In both cases CG coefficients predict a relative branching fraction of a factor of three which is in good agreement with experimental results as shown in Fig. 1.1 for the production rate of Δ resonances and with the results of Ref. [28] for $K_1(1270)$ decays:

$$\frac{\mathcal{B}(K_1(1270) \rightarrow K^+\rho^0)}{\mathcal{B}(K_1(1270) \rightarrow K^+\omega)} = \frac{0.42 \pm 0.06}{0.11 \pm 0.02} = 3.8 \pm 0.9.$$

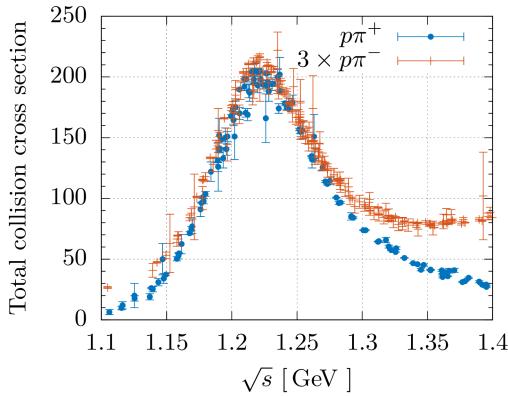


Fig. 1.1. Production rate of $p\pi^+ \rightarrow \Delta(1232)^{++}$ and $p\pi^- \rightarrow \Delta(1232)^0$ with data taken from the Particle Data Group [27]. (Data files are courtesy of the COMPAS Group, IHEP, Protvino, Russia). Isospin conservation predicts an exact ratio of 3 which is in good agreement with the measurements shown above (note the scaling of the $p\pi^-$ data).

An interesting case of isospin violation in strong decays is the decay $\eta \rightarrow 3\pi$. This strong decay of a $I^G(J^{PC}) = 0^+(0^{-+})$ state into $1^-(0^{-+})$ states is forbidden by G -parity. Since G -parity is a combination of isospin and C -parity and the latter is conserved, this decay indeed violates isospin symmetry. Interestingly, $\eta \rightarrow 2\pi$ decays are forbidden by C -parity which makes the 3π decays, albeit (approximately) forbidden, the dominant decay modes⁴ [29, 30, 31]

$$\begin{aligned}\mathcal{B}(\eta \rightarrow \pi^+\pi^-\pi^0) &= 22.6 \pm 0.5\%, \\ \mathcal{B}(\eta \rightarrow \pi^0\pi^0\pi^0) &= 34.0 \pm 0.8\%, \\ \mathcal{B}(\eta \rightarrow \pi^+\pi^-) &< 1.3 \times 10^{-5} \quad (\text{CL} = 90\%), \\ \mathcal{B}(\eta \rightarrow \pi^0\pi^0) &< 3.5 \times 10^{-4} \quad (\text{CL} = 90\%),\end{aligned}$$

and the branching fractions similar to the (isospin violating) electro-magnetic decay $\eta \rightarrow \gamma\gamma$ [29]

$$\mathcal{B}(\eta \rightarrow \gamma\gamma) = 38.5 \pm 0.5\%.$$

³Challenging analysis due to the irreducible background mode $\rho^0 \rightarrow \pi^0\pi^0\gamma$ which explain the absence of measured upper limits.

⁴This also explains the exceptionally narrow width of η .

These measurements shine light onto the quantitative power of the isospin argument by showing, that strong, isospin forbidden decays are on the same level as QED transitions ($\mathcal{O}(\alpha^2)$ in this case). In general, isospin breaking effects can be large in electro-weak decays but are suppressed when strong decays can contribute. Furthermore, the suppression of isospin violation is considered largely independent of the momentum transfer (Q -value), *i.e.*, isospin mixes two mass eigenstates, for example

$$\Lambda_b \rightarrow J/\psi (\cos(\theta)\Lambda + \sin(\theta)\Sigma^0),$$

where the mixing angle θ is largely independent of the momentum transfer [32] and thus allows inferring the isospin suppression between decays with different Q -values.

1.4 $\Sigma^0 \rightarrow \Lambda\gamma$ Background

Λ and Σ baryons are part of the baryon octet. The former is an isospin singlet and the latter three resonances Σ^- , Σ^0 , and Σ^+ form an isospin triplet. The quantum numbers of the Λ and the Σ^0 baryons are identical, including the common valence quarks (uds), except for the isospin which separates the former as a singlet and the latter as part of the triplet. The Σ^0 baryon can thus decay fast without quark transitions into the singlet state Λ via emission of a soft photon. Transitions between the Σ^0 and Λ via strong interaction are forbidden due to the small isospin splitting, $m(\Sigma^0) - m(\Lambda) = 76.959(23) \text{ MeV}/c^2$ [15], which is lighter than any meson mass. Other decay channels besides $\Sigma^0 \rightarrow \Lambda\gamma$ are $\mathcal{B}(\Sigma^0 \rightarrow \Lambda\gamma\gamma) < 3\%$ (experimentally measured at a 90 % CL) and $\mathcal{B}(\Sigma^0 \rightarrow \Lambda e^+ e^-) = 5 \times 10^{-3}$ (theoretical QED calculation) [33, 34].

Due to the small isospin splitting, the photon of the radiative transition will almost always escape undetected at LHCb and is therefore not included in the reconstruction step of the present analysis which makes the Σ^0 resonance a partially reconstructed background in most analyses with intermediate Λ baryons.

1.4.1 Limitations of naïve SU(2) Arguments

Since both, Λ and Σ^0 share the same valence quarks, the decision which of both actually hadronizes in decays is imposed by QCD. The production of Λ and Σ^0 baryons in QCD was studied at lepton colliders, *e.g.*, by the BESIII collaboration in J/ψ and $\psi(2S)$ decays [35]. The results of Ref. [35] are summarized in Tab. 1.1.

Tab. 1.1. Results for branching fractions \mathcal{B} of Λ and Σ^0 production in charmonia decays taken from Ref. [35]. Statistical and systematic uncertainties are added in quadrature. Additionally, we calculate the ratio of Σ^0 and Λ branching fractions, based on the published results of Ref. [35] and assign uncertainties obtained from linear error propagation while ignoring correlations.

Channel	$\mathcal{B} (\times 10^{-4})$	Ratio
$J/\psi \rightarrow \Lambda\bar{\Lambda}$	19.43 ± 0.33	
$J/\psi \rightarrow \Sigma^0\bar{\Sigma}^0$	11.64 ± 0.23	0.599 ± 0.016
$\psi(2S) \rightarrow \Lambda\bar{\Lambda}$	3.97 ± 0.12	
$\psi(2S) \rightarrow \Sigma^0\bar{\Sigma}^0$	2.44 ± 0.11	0.615 ± 0.033

In the case of Λ and Σ^0 production from charmonia resonances, hadrons are not only produced via QCD (OZI suppressed decays), but also via QED, *e.g.*, $J/\psi \rightarrow \gamma \rightarrow \Lambda\bar{\Lambda}$. Assuming isospin conservation for QCD, only the latter process gives rise to isospin breaking decays such as

$J/\psi \rightarrow \Lambda\bar{\Sigma}^0$ which is a transition from an isospin singlet state $|I = 0, I_3 = 0\rangle$ into $|I = 1, I_3 = 0\rangle$. Even though QED breaks isospin conservation, the single photon allows only one quark anti-quark pair to form a $I = 0$ or $I = 1$ state. All other pairs still obey QCD and are thus isospin singlets, *i.e.*, only final states of at most $\Delta I = 1$ are possible. The Clebsch-Gordan (CG) coefficients for a $\Sigma^0\bar{\Sigma}^0$ pair are

$$\underbrace{|1,0\rangle}_{\Sigma^0} \otimes \underbrace{|1,0\rangle}_{\bar{\Sigma}^0} = \left\{ \frac{1}{3} |0,0\rangle, \frac{2}{3} |2,0\rangle \right\}.$$

Thus, even in the case of isospin violating $\Delta I = 1$ processes, $\Lambda\bar{\Lambda}$ and $\Sigma^0\bar{\Sigma}^0$ pairs can only form a $I = 0$ state. Consequently, the branching ratio of $\Lambda\bar{\Lambda}$ and $\Sigma^0\bar{\Sigma}^0$ pairs should be governed by the corresponding CG coefficients, and thus be $1 : 1/3$. Experimentally, a ratio of 60% is measured, thus pointing towards an additional isospin breaking contribution, such as final state interactions and possible corrections of a full SU(3) consideration.

When created in (OZI suppressed) strong decays, the initial isospin state of the u -, d - and s -quarks is well known due to the isospin conservation of QCD, whereas the situation is more complicated in weak decays, since isospin conservation is not guaranteed anymore and thus the initial isospin state of the quarks is typically unknown. $K \rightarrow \pi\pi$ decays have a long history of revealing counter intuitive selection rules between different isospin states, *i.e.*, the amplitudes of $K_s \rightarrow \pi^0\pi^0$, $K_s \rightarrow \pi^+\pi^-$ and $K^+ \rightarrow \pi^+\pi^0$ can be used to unfold amplitudes corresponding to isospin states $I = 0$ and $I = 2$.⁵ Experimentally, the ratio of the amplitudes for $I = 0$ and $I = 2$ is measured to be larger than expected from the CG coefficients, thus hinting towards a suppression of large isospins beyond a naïve angular momentum coupling description. In this particular case, this deviation is known as the so-called $\Delta I = 1/2$ rule for kaons and is well described in theory, for example in the framework of chiral perturbation theory [36].

1.4.2 Background Contamination by $\Sigma^0 \rightarrow \Lambda\gamma$ at LHCb

Since the soft photon in $\Sigma^0 \rightarrow \Lambda\gamma$ cannot be reconstructed at LHCb, most analysis with intermediate Λ resonances will be contaminated with the partially reconstructed background events of Σ^0 decays which are irreducible.

For example, this is the case in the charmless two-body decay $B^+ \rightarrow p\bar{\Lambda}$, analyzed at LHCb [37]. Penguin contributions with a $\bar{b} \rightarrow \bar{s}$ loop are expected to dominate, but tree-level and annihilation diagrams also contribute. Electro-weak penguins and tree diagrams create a $u\bar{u}$ or $d\bar{d}$ pair that couples with the spectator quark either into an isospin $I = 1/2$ or $I = 3/2$ state, corresponding to $\Delta I = 0$ and $\Delta I = 1$, respectively. The other $q\bar{q}$ pair is created from the QCD vacuum and thus does not change the total isospin. In case of the annihilation diagram and gluon penguins, two $q\bar{q}$ pairs are created from the QCD vacuum and only $\Delta I = 0$ is possible.

The quark states can hadronize as either $p\bar{\Lambda}$ or $p\bar{\Sigma}^0$, *i.e.*,

$$p\bar{\Lambda} : \quad \left| \frac{1}{2}, \frac{1}{2} \right\rangle \otimes |0,0\rangle = \left| \frac{1}{2}, \frac{1}{2} \right\rangle,$$

$$p\bar{\Sigma}^0 : \quad \left| \frac{1}{2}, \frac{1}{2} \right\rangle \otimes |1,0\rangle = \left\{ \frac{1}{3} \left| \frac{1}{2}, \frac{1}{2} \right\rangle, \frac{2}{3} \left| \frac{3}{2}, \frac{1}{2} \right\rangle \right\}.$$

Inferring a suppression of $\Delta I = 1$ from the $\Delta I = 1/2$ selection rule for kaons, Σ^0 resonances would be suppressed by a factor of $1/3$. Vice-versa, if $\Delta I = 0$ would be suppressed, then electro-weak penguins and tree diagrams would induce an increase of Σ^0 hadronization. The fit to

⁵The $I = 1$ final state is unreachable for pairs of pions due to their bosonic nature, *i.e.*, bosons are described by symmetric wave functions, whereas combinations of odd angular momentum L or odd isospin I quantum numbers are antisymmetric. Hence L and I must either be both odd or both even. For kaon decays, $L = 0$ and therefore I can have only even values.

recorded $B^+ \rightarrow p\bar{\Lambda}$ events at LHCb prefers the former option by finding $N(B^+ \rightarrow p\bar{\Sigma}^0)$ being compatible with zero but $N(B^+ \rightarrow p\bar{\Lambda}) = 13.0^{+5.1}_{-4.3}$ [37]. We note, that the amount of available statistics in this channel is similar to $\Lambda_b \rightarrow D^0\Lambda$ where we expect the very same suppression factor. (Even without relying on “mysterious ΔI rule[s]”, *cf.* Ref. [36].)

In Tab. 1.2 we show a selection of Λ_b decays with an intermediate Λ baryon together with the corresponding Λ - Σ^0 asymmetry

$$a(\Lambda : \Sigma^0) := \frac{f(\Lambda) - f(\Sigma^0)}{f(\Lambda) + f(\Sigma^0)}, \quad (1.1)$$

where $f(\Lambda)$ and $f(\Sigma^0)$ refer to the amounts of Λ and Σ^0 , respectively. An asymmetry of -1 , 0 , $+1$ thus corresponds to a pure Σ^0 , balanced Λ and Σ^0 , and pure Λ decay, respectively.

For some of the decays listed in Tab. 1.2 W -exchange diagrams are also possible (*e.g.*, Ref. [38]) which are often considered suppressed in the literature. We note, that a strong suppression is only given for mesons due to the required spin alignment of the quark and anti-quark pair, but that this is not necessarily the case for baryons.

The decays $\Lambda_b \rightarrow \Lambda hh'$, $\Lambda_b \rightarrow \Lambda\phi$ and $\Lambda_b \rightarrow J/\psi\Lambda$ were analyzed at LHCb [39, 40, 41], whereas the other decays are subject of the present analysis. The ranges for the values of the Λ - Σ^0 asymmetry arises from the yet unknown ΔI selection rules we mentioned above. For example, the u -quark produced via W -exchange in the $\Lambda_b \rightarrow \Lambda\phi$ decay can either end within a $|0, 0\rangle$ or a $|1, 0\rangle$ state together with the spectator, corresponding to a Λ - Σ^0 asymmetry of either 1 or -1 , respectively. In the case of tree and penguin decays of the Λ_b baryon, we can further constrain the possible combinations by observing, that the spectator diquark (ud) is an isospin singlet state $|0, 0\rangle$ which is, since not involved in the decay in leading order, conserved. In decays via W -exchange, this initial singlet state is broken, though, and the spectator quark can contribute in the subsequent isospin coupling.

Most interesting for the present analysis is the ratio of the absolute amount of reconstructed decays with intermediate Λ and Σ^0 in order to estimate the background contamination of $\Lambda_b \rightarrow D^0\Sigma^0$ in $\Lambda_b \rightarrow D^0\Lambda$. Unfortunately, there are a couple of caveats with naïvely interpreting the yields of the fitted modes. For example in the analysis of $\Lambda_b \rightarrow \Lambda hh'$ decays, the authors could only account for partially reconstructed background components with a missing soft photon in general, due to the lack of experimental data of further Λ_b background candidates [39]. Contributions of $\Sigma^0 \rightarrow \Lambda\gamma$ are, however, most physically motivated, but only one reasonable instance. Unfortunately, the partially reconstructed background $\Sigma^0 \rightarrow \Lambda\gamma$ peak in roughly the same place as the cross-feed contribution, which is quite well understood, but as the corresponding signal yields are small, this results in a reasonable uncertainty on which contributions are which. Additionally, the same problem occurs in the $\Lambda_b \rightarrow \Lambda ch$ control mode where the Λ_c can either decay into $\Lambda\pi^+$ or $\Sigma^0\pi^+$. Both of these amplitudes are measured by the BESIII collaboration and found to be significantly greater than zero [42] and thus pollute the interpretation of the fitted yields as the true charmless decays, without Λ_c component. Luckily for the present analysis, this effect is most pronounced in the $\Lambda\pi^+\pi^-$ final state (*cf.* Chap. 7). The branching fractions for corresponding modes with kaons are considerably more suppressed.

1.4.3 Summary

Isospin (non-)conservation is helpful for deciding whether or not a decay is suppressed. Finding exact branching ratios, though, can be error prone as we saw in the case of $\Lambda\bar{\Lambda}$ vs. $\Sigma^0\bar{\Sigma}^0$ production or due to yet unknown ΔI selection rules. However, in the case of $\Lambda_b \rightarrow D^0\Lambda$ the situation is much easier and we do not expect large corrections neither to the tree, nor to the W -exchange diagram.

Regarding the available statistics, we will not be able to constrain or extract the relative branching ratio but we find it noteworthy to mention, that the Σ^0 modes are of great interest

Tab. 1.2. Selection of Λ_b decays with an intermediate Λ baryon, primary quark transitions, and Λ - Σ^0 asymmetry $a(\Lambda : \Sigma^0)$ as defined in Eq. (1.1). For decays via W -exchange the spectator quark has to be included into the isospin coupling (flavor is shown in parentheses).

Channel	Quark transition		Λ - Σ^0 asymmetry
$\Lambda_b \rightarrow \Lambda \pi^+ \pi^-$	$b \rightarrow u s \bar{u}$	(tree & penguin)	$1/9 \dots 1/3$
	$b \rightarrow s d \bar{d}$	(penguin)	$1/9 \dots 1/3$
	$b u(d) \rightarrow u s(d)$	(exchange)	$1/9 \dots 1/3$
$\Lambda_b \rightarrow \Lambda K^+ \pi^-$	$b \rightarrow u d \bar{u}$	(tree & penguin)	$-1/6 \dots 1/3$
	$b \rightarrow d s \bar{s}$	(penguin)	$1/3$
	$b u(d) \rightarrow u d(d)$	(exchange)	$-1/6 \dots 1/3$
$\Lambda_b \rightarrow \Lambda K^+ K^-$	$b \rightarrow s u \bar{u}$	(tree & penguin)	$0 \dots 1/2$
	$b \rightarrow s s \bar{s}$	(penguin)	$1/2$
	$b u(d) \rightarrow u s(d)$	(exchange)	$0 \dots 1/2$
$\Lambda_b \rightarrow \Lambda \phi$	$b \rightarrow s s \bar{s}$	(penguin)	1
	$b u(d) \rightarrow u s(d)$	(exchange)	$-1 \dots 1$
$\Lambda_b \rightarrow J/\psi \Lambda$	$b \rightarrow c \bar{c} s$	(tree & penguin)	1
	$b u(d) \rightarrow u s(d)$	(exchange)	$-1 \dots 1$
$\Lambda_b \rightarrow D^0 \Lambda$	$b \rightarrow c \bar{u} s$	(tree)	$1/2$
	$b u(d) \rightarrow c s(d)$	(exchange)	$1/2$
$\Lambda_b \rightarrow \bar{D}^0 \Lambda$	$b \rightarrow u s \bar{c}$	(tree)	$1/2$
$\Xi_b^0 \rightarrow D^0 \Lambda$	$b \rightarrow c d \bar{u}$	(tree)	$-1 \dots 1/2$
	$b u(s) \rightarrow c d(s)$	(exchange)	$1/2$
$\Xi_b^0 \rightarrow \bar{D}^0 \Lambda$	$b \rightarrow u d \bar{c}$	(tree)	$-1 \dots 1/2$

for future experiments for at least two reasons: First, for the same reasons *CP* violation is expected in $\Lambda_b \rightarrow D\Lambda$, it is also expected for $\Lambda_b \rightarrow D\Sigma^0$ with different strong phases. Secondly, SU(3) considerations yield relations among the amplitudes of $\Lambda_b \rightarrow D^0X$ and $\Xi_b^0 \rightarrow D^0X$ decays [43]:

$$\begin{aligned} 1 - \sqrt{3} \frac{\mathcal{A}(\Lambda_b \rightarrow D^0\Lambda)}{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)} + \sqrt{2} \frac{\mathcal{A}(\Xi_b^0 \rightarrow D^0\Xi^0)}{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)} &= 0, \\ 1 + \sqrt{3} \frac{\mathcal{A}(\Lambda_b \rightarrow D^0\Lambda)}{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)} - 2 \frac{\lambda_s}{\lambda_d} \frac{\mathcal{A}(\Xi_b^0 \rightarrow D^0\Sigma^0)}{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)} &= 0, \\ 1 - 2 \frac{\lambda_d}{\lambda_s} \frac{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)}{\mathcal{A}(\Xi_b^0 \rightarrow D^0\Sigma^0)} + \sqrt{3} \frac{\mathcal{A}(\Xi_b^0 \rightarrow D^0\Lambda)}{\mathcal{A}(\Xi_b^0 \rightarrow D^0\Sigma^0)} &= 0, \\ 1 - \sqrt{3} \frac{\mathcal{A}(\Lambda_b \rightarrow D^0\Lambda)}{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)} + \sqrt{2} \frac{\lambda_s}{\lambda_d} \frac{\mathcal{A}(\Lambda_b \rightarrow D^0n)}{\mathcal{A}(\Lambda_b \rightarrow D^0\Sigma^0)} &= 0, \end{aligned}$$

with

$$\begin{aligned} \lambda_d &:= V_{ud}^* V_{cb}, \\ \lambda_s &:= V_{us}^* V_{cb}. \end{aligned}$$

These should be tested and used to constrain decays that are unavailable for current experiments such as $\Lambda_b \rightarrow D^0n$.

1.5 *CP* Violation in *b*-Baryon Decays

CP violation is an interference effect and originates from the complex phases of the CKM matrix. In order to get a measurable physical quantity in amplitudes the given observable has to come from the superposition of two (or more) decay modes that contribute a *CP* even and a *CP* odd term (CKM phases are *CP* odd) such that the interference term does not cancel in the magnitude. In general, such decay modes are categorized into three classes, direct *CP* violation, *CP* violation in mixing, and *CP* violation in interference of mixing and decays (*cf.* any good text book about flavor physics for more details). Mesons offer a rich laboratory for measuring *CP* violation among all three classes, *e.g.*, Refs. [44, 45]. In baryonic systems, and in particular in decays of *b*-baryons such as the Λ_b or Ξ_b^0 baryon, there can be no mixing due to conservation of baryon number and therefore no indirect *CP* violation. Studies of direct *CP* violations are often limited to measuring asymmetries, since the *CP* even phases have to be taken from theory which makes the result model dependent and typically come with large uncertainties [46, 47]. Lately, two methods have been proposed to overcome this issue [48, 49]. They reinterpret methods originally proposed by Atwood, Dunietz and Soni [50, 51], referred to as ADS, and a method first proposed by Gronau, London and Wyler [52, 53], referred to as GLW, that are already successfully carried out in decays of mesons [54], now for decays of baryons. As proposed, these methods allow a clean extraction of the CKM phase γ in a model independent way⁶ and at the same time neither require tagging nor time-dependence. In particular, in the ADS method γ can be extracted by the study of the six decays $\Lambda_b \rightarrow D^0\Lambda$, $\Lambda_b \rightarrow \bar{D}^0\Lambda$, $\Lambda_b \rightarrow D_{CP}\Lambda$, and the respective $\bar{\Lambda}_b$ modes, where

$$D_{CP} \equiv \begin{cases} \frac{|D^0\rangle + |\bar{D}^0\rangle}{\sqrt{2}} & \text{if } CP\text{-even modes are considered,} \\ \frac{|D^0\rangle - |\bar{D}^0\rangle}{\sqrt{2}} & \text{else.} \end{cases}$$

⁶In particular the proposed modes do not suffer from penguin pollution.

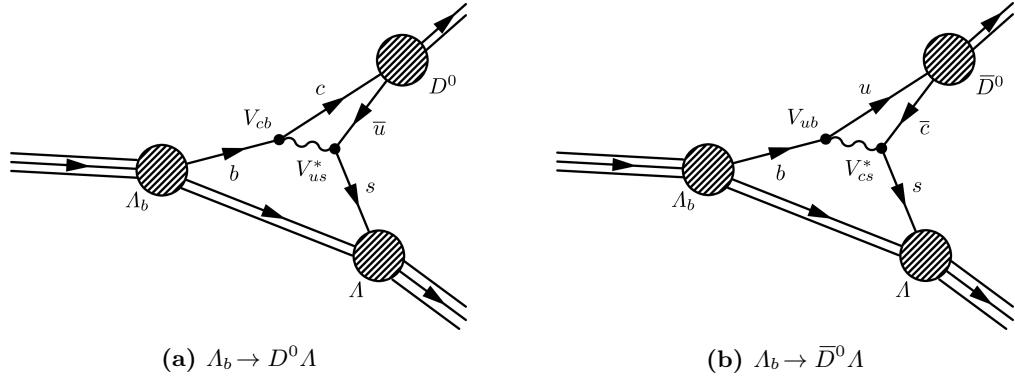


Fig. 1.2. Feynman diagrams of the decays $\Lambda_b \rightarrow D^0\Lambda$ and $\Lambda_b \rightarrow \bar{D}^0\Lambda$. The latter decay is strongly suppressed w.r.t. to the former due to the $b \rightarrow u$ transition. The suppression can be reduced by reconstructing $D \rightarrow K^+\pi^-$ which is Cabibbo doubly suppressed for the former but Cabibbo favored for the latter.

Key to the present analysis is the idea, to enhance the CP violating contribution by reconstructing $\Lambda_b \rightarrow [K^+\pi^-]_D\Lambda$ which compensates the suppression of $\Lambda_b \rightarrow \bar{D}^0\Lambda$ by favoring the Cabibbo (doubly) suppressed $D^0 \rightarrow K^+\pi^-$ mode in the $\Lambda_b \rightarrow D^0\Lambda$ decay (*cf.* Fig. 1.2). Further, the statistics of the studies can be enriched by also including the four-body modes $D \rightarrow K^\pm\pi^\mp\pi^+\pi^-$. Similarly, the same technique is applicable to other two-body decays, such as $\Xi_b^{0/-} \rightarrow D\Xi^{0/-}$ and $\Omega_b^- \rightarrow D\Omega^0$ but also to the various backgrounds of the present analysis (*cf.* Chap. 7) and also three-body decays such as $\Lambda_b \rightarrow D^0 p K^-$ or $\Lambda_b \rightarrow \Lambda\pi^+\pi^-$ [55]. (The latter is also well suited for measurement of T violation with triple products [56].) We note further that in all these modes, the CP violation is encoded in the entire decay chain, do not rely on CP violation in the D meson system and thus genuinely leverages the observation of CP violation in baryon decays. The same holds for the cited GLW method where the D^0 and \bar{D}^0 are reconstructed in CP eigenstates K^+K^- and $\pi^+\pi^-$.

In contrast to the two-body decay modes, the three-body decay modes have to be studied in Dalitz plots that require a meticulous description of the various background and resonance contributions. On the contrary, two-body decays come with significant smaller sample sizes at LHCb, due to the lower branching fraction, and reconstruction and trigger inefficiencies of the long living Λ baryons, but offer a much easier to analyze decay topology. Since none of the $\Lambda_b \rightarrow D\Lambda$ decays have been observed at the time of writing, the main focus of the present analysis is to establish a branching ratio for $\Lambda_b \rightarrow D^0\Lambda$ by reconstructing $D^0 \rightarrow K^-\pi^+$ (and thus neglecting Cabibbo doubly suppressed contributions from $\Lambda_b \rightarrow \bar{D}^0\Lambda$). We find the available dataset also sensitive to the decay $\Xi_b^0 \rightarrow D^0\Lambda$ which is on its own a candidate for measuring CP violation. We note that a CP violation study of the Ξ_b^0 decay using the ADS method does not suffer from a contamination of charmless backgrounds⁷ as it is the case for CP violation studies of the corresponding Λ_b decay, whereas the expected amount CP violation is much lower.

⁷Due to the absence of $\Xi_b^0 \rightarrow \Lambda K^+\pi^-$ decays, *cf.* Sec. 7.3.

Chapter 2

The LHCb Detector at the LHC

The LHCb experiment is one of the four major HEP experiments at the Large Hadron Collider (LHC) and understands its primary focus in the realm of b - and c -physics. The LHC is a particle accelerator located at the CERN facility. Its main component is a storage ring with a total length of 26.7 km in which protons are collided symmetrically. At the time of writing the LHC is the most powerful particle accelerator in the world.¹ Since the year 2011, the proton beam energy has increased from 3.5 TeV to 4 TeV in the year 2012 and to 6.5 TeV in the year 2015. The periods of data taking are divided into the so-called Run 1 and Run 2 where the former refers to the years 2011 and 2012 and the latter to the years 2015 until 2018, corresponding to an integrated luminosity of roughly 3 fb^{-1} and 6 fb^{-1} , respectively.

Due to the large center-of-mass energy at the LHC the highly correlated b - and \bar{b} -hadrons are predominately produced in the same forward or backward cone. The LHCb detector is thus designed as a (single arm) forward spectrometer, covering a forward cone from 15 mrad to 300 mrad in the bending plane and 15 mrad to 250 mrad in the non-bending plane. The detector configuration consists of several tracking stations and calorimeters to reconstruct charged and neutral particles. An effective particle identification is provided by a large dipole magnet and two Cherenkov radiators. Different stages of hard- and software triggers reduce the event rate to a frequency at which events can be stored to disks.

In the present analysis we analyze the full available Run 2 dataset, recorded in the years 2015 until 2018. Data recorded during the years 2011 and 2012 (Run 1) are not taken into account for several reasons: First, from a technical point of view the experimental setup was frequently changed during Run 1. This includes changes to the trigger configuration which would require a separate analysis of data recorded during 2011 and the first and second half of 2012. Secondly, only the second half of 2012 includes an efficient dedicated Λ trigger, diminishing the total selection efficiency of the previous parts. Thirdly, even though the Λ_b production fraction is larger at small energies, this advantage is overcompensated by far due to the smaller $b\bar{b}$ cross-section, making Run 2 significantly more efficient in terms of the overall signal efficiency. Taking into account the larger data sample in terms of luminosity we conclude that adding Run 1 data could not help to significantly improve our results, but would imply disproportionate larger complexity and is therefore disfavored.

2.1 Experimental Setup

In the following we give a short overview about a selection of detector components that are relevant for the present analysis. More detailed information, illustrations and thorough descriptions of the entire detector can be found in Refs. [59, 60].

¹Although powerful, it will most likely not destroy earth [57, 58].

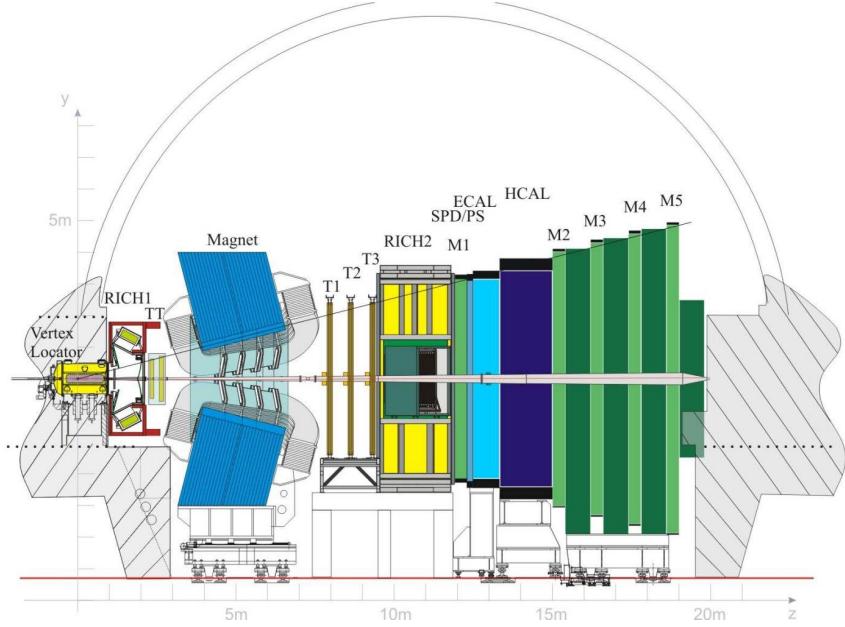


Fig. 2.1. Schematic view of the LHCb detector by R. Lindner (2008).

2.1.1 Tracking

A vertex locator (VELO) provides the precise measurements of tracks near of the pp interaction point. Besides track reconstruction this information is also useful to distinguish primary vertices (PV), *e.g.*, from particles originating directly from the interaction point, and secondary vertices from long living b - and c -hadrons. The VELO consists of 42 silicon modules, each of them equipped with radial and azimuthal strips. The modules are semicircularly shaped and arranged in pairs such that they surround the beam pipe perpendicularly. The pitches of the silicon sensors increase linearly from $38\text{ }\mu\text{m}$ at the inner radius ($r = 8.2\text{ mm}$) to $102\text{ }\mu\text{m}$ at the outer radius ($r = 42\text{ mm}$). The spatial design of the modules was chosen such that charged particles with a pseudo rapidity $1.6 < \eta < 4.9$ cause at least three hits inside the VELO. The total length of the VELO is not sufficient to cover all end vertices of long living V^0 particles such as the Λ baryon or the K_s meson. If both daughters of a V^0 two-body decay are reconstructed within the VELO we refer to the reconstructed tracks as long tracks (LL). Otherwise, if the reconstruction is only based on hits in the tracking stations TT and T1-T3, we refer to them as downstream tracks (DD). For the sake of brevity, we categorize entire decay chains, such as $\Lambda_b \rightarrow D^0 \Lambda$, also as LL (DD) if both Λ daughters are reconstructed as long (downstream) tracks.

Besides the VELO, four more tracking systems are placed, referred to as the TT, located upstream, and the modules T1, T2 and T3, located downstream of the spectrometer magnet. The TT and the inner parts of the T1-T3 stations are constructed from p-on-n silicon microstrips detectors with a hit efficiency above 99 % and a hit resolution of approximately $50\text{ }\mu\text{m}$. The silicon microstrips are arranged in four layers, corresponding to an active area of approximately 8 m^2 . The resolution of the T1-T3 stations varies from the innermost towards the outer parts. Whereas silicon strips are used for the inner parts, the outer parts of the T1-T3 stations are straw-tubes. The straw-tubes measure the drift time of ionised charges with a total delay below 75 ns.

2.1.2 Ring Imaging Cherenkov Detectors

Two Ring Imaging Cherenkov detectors (RICH), referred to as RICH1 and RICH2, detect Cherenkov light. Cherenkov radiation is an electromagnetic radiation that gets emitted by charged particles in a dielectric medium if the velocity v of the particles is greater than the (group) velocity c_m of the given medium. For constant velocities the emitted light cone has a constant opening angle ϕ ,

$$\sin \frac{\phi}{2} = \frac{c_m}{v} = \frac{c_0}{nv},$$

and thus directly allows the inference of v if the refractive index n (and the speed of light in vacuum c_0) is known. Inside the RICH units, Cherenkov cones are mapped to spherical coordinates (r, ϑ) , where r is a function of the opening angle ϕ and ϑ is the azimuthal part of the Cherenkov photons trajectory. This projection is achieved via spherical mirrors which reflect the photons to Hybrid Photon Detectors where they are detected. By combining the measurements of the opening angle ϕ as a function of v and the deflection in the magnetic field, a hypothesis for the invariant mass can be set and thus allow to identify the particle (*cf.* Sec. 2.1.7). Each RICH unit is filled with a different gas admixture, corresponding to different refractive indices to lower the Cherenkov thresholds and to allow a diverse momentum resolution. Since Run 2, RICH1 is filled with C_4F_{10} , whereas CF_4 is used as the radiator in RICH2, making the units sensitive to the momentum ranges 2 to 60 GeV/c and 15 to 100 GeV/c , respectively. (Particle dependent thresholds are listed in Tab. 2.1.)

2.1.3 Calorimeters

Neutral particles like photons are neither deflected in magnetic fields, nor do they emit Cherenkov light, nor are they detected in the VELO, TT and T1-T3, hence a calorimeter system is necessary. The calorimeters bring a large amount of material into the detector as they have to stop the passing particles completely to measure the deposited energy. In order to not affect the measurements in the tracking stations, they are placed downstream of them. In total, four types of calorimeters are used in the LHCb detector. Together they provide the identification of electrons, protons and other hadrons, as well as the energies and positions of photons and neutral hadrons. Furthermore, the calorimeter measurements are used to select candidates with high transverse momentum for the first trigger level L0. The four calorimeter systems are a scintillating pad detector (SPD), a preshower calorimeter (PS), an electromagnetic calorimeter (ECAL) and an hadronic calorimeter (HCAL). The first two systems consist of plain scintillator tiles, separated from each other by a thin lead layer (2.5 radiation lengths), and the ECAL (25 radiation lengths) and HCAL systems have a shashlik² and a sampling construction³, respectively. During the traversal, the particles are stopped in the lead and iron layers and deposit their energy. This energy produces light in the (organic) scintillators which is transmitted via optical fibers to photo multipliers where the photons are detected. Since the hit density varies over the calorimeter surface perpendicular to the beam axis, a variable lateral segmentation is adopted. The outer dimension is chosen such that it projectively matches those of the tracking system and the inner dimension is limited by the beam pipe. During Run 2, the ECAL and HCAL are predominantly used for contributing to the trigger decisions, *cf.* Sec. 2.1.6.

2.1.4 Muon System

Although muons are charged particles their interaction is small compared to hadrons like pions or kaons due to their leptonic nature, and also small to the electron due their large invariant mass. Muons thus do not deposit their entire energies in the calorimeters and dedicated muon

²A sampling scintillator and lead structure perpendicular to the beam axis.

³A sampling scintillator and iron structure parallel to the beam axis.

systems are necessary. In total, five rectangular shaped muon stations M1-M5 are placed upstream and downstream of the calorimeter systems. M1 is a triple gas electron multiplier and is placed upstream in front of the calorimeter stations in order to improve measurements of the transverse momentum within the trigger logic. The stations M2-M5 are multi wire proportional chambers and are placed downstream behind all other detector units. In order to discriminate muons against the abundant hadronic background and cosmic rays, muons are required to produce aligned hits in all five stations, corresponding to a minimum momentum above 6 GeV/c. The stations are interleaved by thick iron absorbers of 20 nuclear interaction lengths. The chambers cover an active area of 435 m², have a time resolution below 4.5 ns and differ in their lateral segmentation similar to the tracking stations and the calorimeter systems.

2.1.5 Dipole Magnet

The spectrometer magnet is a warm dipole magnet with a saddle-shaped coil design in a window frame yoke with sloping poles. It provides an integrated magnet field of roughly 4 T m for particles with a track length of 10 m. Three smaller dipole magnets inside this magnet compensate the impact on the LHC beam. In order to analyze and cancel asymmetries of the detector units between oppositely charged tracks, the polarity of the magnetic field can be switched. In the following we refer to these states as *mag. up* and *mag. down*. Unfortunately, the magnetic field is not known exactly throughout the entire detector (neither temporal, nor spatial), thus leading to an inexact calibration of the particle momenta. Furthermore, charged particles loose energy by ionization or by emitting photons when they traverse material or are deflected in a magnet field. The emitted particles (typically electrons or photons) tend to have low energies and therefore remain undetected. Such losses are only reproduced partially in simulations, hence the reconstructed energy of particles is too low, resulting in asymmetries in the inferred momenta. Additionally, misalignment effects of the detector units lower the resolution.

Software tools are available that mitigate these effects by correcting the momentum scale phenomenological. During the present analysis we use a tool which calibrates the scale with the two-body decays $J/\psi \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow J/\psi K^+$ where in the former the dimuon pair and in the latter the K^+ is used for the calibration. The samples are split w.r.t. the product of particle charge and magnet polarity and thus removes a potential bias due to whether the particles are deflected inwards or outwards by the magnet field.

2.1.6 Trigger System

The LHC machine operates with a bunch crossing frequency of 40 MHz which is unfeasible for any kind of unfiltered reconstruction or data storing. A dedicated trigger system thus select bunch crossings with at least one inelastic pp interaction and further reduces its rate below 12.5 kHz by applying filter criteria which, in general, aim to select b - and c -quark decays. The trigger system is arranged in two different tiers, referred to as the low level trigger L0 and the high level trigger HLT.

The L0 trigger is directly implemented in hardware and reduces the event rate below 1.1 MHz by combining information of the calorimeters and the muon systems. The complexity of L0 decisions are strictly limited by the LHC bunch crossing frequency and latency constraints, and allows only read-outs of the transverse energy deposited by showers in clusters in the calorimeter systems SPD, PS, ECAL and HCAL, as well as transverse momenta measured in the muon chambers.

The trigger decision in the calorimeter is based on the transverse energy E_T of single clusters where a cluster consists of 2×2 calorimeter cells. The transverse energy of each cell $E_{T,i}$ is

corrected by an inclination angle θ_i of a hypothetical neutral particle, accumulated,

$$E_T = \sum_{i=1}^4 E_{T,i} \sin \theta_i ,$$

and tested against thresholds. In particular, the L0 decision for hadrons is based on single tracks and thus is insensitive to different decay topologies, *e.g.*, $\Lambda_b \rightarrow D^0 p \pi^-$ and $\Lambda_b \rightarrow D^0 (\Lambda \rightarrow p \pi^-)$ have a similar L0 responses if the invariant mass of $p \pi^-$ is close to $m(\Lambda)$, whereas for $\Lambda_b \rightarrow D^0 p \pi^-$ and $\Lambda_b \rightarrow D^0 p K^-$ the L0 response can vary due to different hadronic showers signatures of π^- and K^- . Unlike the E_T based decisions of showers in the calorimeters, the transverse momenta measured in the muon chambers does take combinations of up to two tracks into account and thus leverages an effective triggering of dimuon pairs for example in high statistics modes such as $\Lambda_b \rightarrow J/\psi \Lambda$.

The L0 trigger system rejects collisions for further processing if the transverse energy of clusters (calorimeter) or transverse momenta (muon chambers) are below certain thresholds. The nominal value of each threshold is the objective of an online maximization of the signal efficiency under different LHC conditions and are tuned during data taking. MC simulated events typically scale these thresholds and do not reflect temporal changes since this would imply an unnecessary waste of computing intensive detector simulations. The fidelity of the L0 trigger response in MC simulated events thus has to be studied carefully if needed.

The latency of an individual L0 decision is 4 μs which allows a subsequent read-out of the entire detector for the HLT trigger which include a full off-line event reconstruction. The HLT was completely redesigned for the Run 2 data taking period due to changes of the bunch crossing frequency and the amount of visible interactions per bunch crossing, and now differs strongly from its previous design outlined in Ref. [61]. During Run 2 the HLT performs partial and full event reconstructions in software applications running on computing farms and stores the events at a rate of 12.5 kHz. An in-depth description, as well as efficiency studies of the LHCb trigger system can be found in Refs. [62, 63] (L0 and HLT studies for Run 1) and Ref. [64] (L0 and HLT studies for Run 2).

Terminology-wise, events of a decay directly involved in setting a trigger (line) are called TOS (trigger on signal) events. If set, triggers typically cause the entire event to be stored to disk, including the decay tree initiated by the other associated b - or c -quark, even if not necessarily involved in the trigger decision. These events are called TIS events (trigger independent of signal). We note that both decay trees of a $b\bar{b}$ or $c\bar{c}$ pair can set trigger (lines). In this case they are both considered TOS and TIS.

2.1.7 PID

The particle identification (PID) at LHCb is built upon information of four different detector parts. The calorimeters mainly contribute to the recognition and identification of neutral particles (γ, π^0) or electrons, muon chambers identify muons and the two RICH systems primarily identify charged hadrons ($\pi^+, \pi^-, K^+, K^-, p, \bar{p}$) and contribute to the identification of charged leptons (e, μ) as well.

This information is gathered by two different approaches. In the first method the likelihood information of each of the four subsystems is accumulated in a combined log-likelihood difference. During analysis we refer to it as DLL, for example $\text{DLL}(X - Y)$, where X and Y are two different mass hypotheses. The second method utilizes a shallow neural network with the information of the four subsystems as the input layer, one hidden layer and six output neurons which are trained w.r.t. the particle hypotheses e, μ, K, π, p and *ghost*, where the latter refers to trajectories that do not correspond to a single particle (*cf.* Ghost Prob.). These neural networks thus also encode correlations between the detector units and typically outperform the more canonical approach of linear adding log-likelihoods differences [60, 61, 65]. In the present

analysis we refer to the numerical value of a neuron (output layer) corresponding to a particle hypothesis X for a given particle Y as $\text{ProbNN}_X(Y)$. The neurons are trained to yield large values if a given hypothesis is met, *e.g.*, requiring large values for $\text{ProbNN}_k(p)$ could be used as a veto against kaons for proton candidates. The numerical value is normalized to the interval $[0, 1]$ and is thus said to be interpretable as probabilities for the given particle hypothesis (hence the name). We adopt the name but avoid the probability interpretation in the following.⁴

The PID value of the RICH systems are based on $\text{DLL}(X - Y)$ values where X and Y are two different mass hypotheses. For the RICH to yield a meaningful number for this (*i.e.* not zero) it requires:

- The track is in the acceptance of at least one of the two RICH radiator volumes.
- The expected signature in the RICH is different for X and Y .

The former criterion does not change with the mass hypothesis assigned to a track, but depends dominantly on the path length of particles that traverse the RICH volumes. The latter means that at least one of the two hypothesis X or Y (the lightest) needs to be above threshold, in at least one RICH radiator. If one of X or Y is below threshold, that does not matter, as long as the other is above. The momentum thresholds for different (charged) particles, calculated from the (measured) RICH refractive indices are given in Tab. 2.1 below.

Tab. 2.1. Momentum thresholds in GeV/c of the RICH systems for different charged particles, calculated from the measured refractive indices of the RICH systems and the respective particle masses.

Particle	RICH1 [GeV/c]	RICH2 [GeV/c]
e^\pm	0.00978	0.0170
μ^\pm	2.02	3.52
π^\pm	2.67	4.65
K^\pm	9.45	16.4
p, \bar{p}	17.9	31.3
d, \bar{d}	35.9	62.6

2.2 Variables and Acronyms used as Selection Criteria

During analysis we frequently use abbreviations and acronyms which refer to variables that are used for discriminating background events. Below, we give a short overview over the most common ones and occasionally refer to this list in the subsequent chapters.

$\Delta\chi^2_{\text{IP}}$ Difference between the χ^2 value of the PV reconstructed with and without the track under consideration. Clearly, by adding an additional DoF the χ^2 value will always improve, however the improvement is less pronounced for spurious than for genuine tracks on average.

DLL($X - Y$) Delta log-likelihood of PID values of the RICH systems w.r.t. the different mass hypotheses X and Y .

DIRA Cosine of the angle between the momentum of the particle and the direction vector from some reference vertex or 3D-point to the end-vertex of the particle.

⁴This clearly depends on the given definition of *probability*. Recent results show that the output of certain neural network can in fact be interpreted as probabilities via Bayesian inference [66, 67]. However, the shallow networks used for PID in the present analysis do not use such techniques.

DOCA Distance of closest approach for pairs of tracks. This quantity is evaluated before applying a DTF since the latter forces this value to zero for all daughters of a common mother. In general, the DOCA of the daughters of a (common) decay is small (albeit non-zero due to finite resolution effects) and large for random track combinations on average.

$|m(X) - \text{PDG}|$ The absolute difference between the invariant mass of particle X as obtained from four momentum addition and the respective nominal value (typically provided by the PDG).

Best fit probability If multiple PVs are reconstructed for a single event the matching is ambiguous in DTFs with constraint PV. We disambiguate by choosing the PV corresponding to the best fit in terms of the evaluated χ^2 value and refer to it as the *best fit*.

Ghost Prob. Output of a neural network based algorithm to identify tracks which do not correspond to the trajectory of a (single) true particle but rather originates from detector noise or multiple particles due to mismatching [68].

Chapter 3

Reconstruction and Stripping Efficiencies

The combined reconstruction and stripping efficiency is determined with MC simulated events. A full MC simulation of a single event includes a time consuming simulation of the interactions of the particle shower with the detector material, as well as the response of the detector itself. The reconstruction efficiency of the LHCb detector is well below 100 %, *e.g.*, particles with a transverse momentum below a certain threshold will almost never be stored to disk due to the associated noisy detector response, unmet trigger requirements, or geometric cut-offs. During simulations, decays which only produce such particles are skipped to save the time intensive simulation of the detector traversal. This so-called *Generator Cut* has no counterpart in recorded data but will not change the overall efficiency when the latter is defined wide enough such that it covers selection requirements which would filter out these events for good. For the present analysis the stripping phase obeys this requirement such that, from a technical point of view, the product of reconstruction and stripping efficiency is given by¹

$$\text{rec.} \times \text{strip. eff.} = \varepsilon_{\text{gen}} \times \frac{\#\text{DTT}}{\#\text{DST}}, \quad (3.1)$$

where ε_{gen} is the generator cut efficiency, and $\#\text{DST}$ and $\#\text{DTT}$ are technical abbreviations for the total amount of events after the detector simulation and the number of events after the respective stripping phase.

These efficiencies are not stable during Run 2, since (high) level triggers are under permanent changes as well as the simulation algorithms. In Appx. A we list the trigger and simulation versions for the decays under consideration (*cf.* Tab. A.1) and show graphical representations of the respective generator cut efficiencies (*cf.* Fig. A.1). The combined reconstruction and stripping efficiency is shown in Fig. 3.1. The corresponding values are listed in Tab. A.2 and Tab. A.3. The efficiencies shown in Fig. 3.1 are normalized to the respective weighted mean of each decay for the full available data set (still separated w.r.t. the track types LL and DD) where we find the weighted mean ε of N efficiencies $p_i = n_i/N_i$ by calculating²

$$\varepsilon = \frac{\sum_i^N w_i p_i \varepsilon_{\text{gen},i}}{\sum_i^N w_i},$$

where w_i is the inverse sum in quadrature of the uncertainty of the generator cut efficiency $\sigma_{\text{gen},i}$ and the respective binomial uncertainty

$$w_i = \sigma_i^{-2} = \left(\sigma_{\text{gen},i}^2 + \frac{p_i(1-p_i)}{N_i} \right)^{-1}.$$

Several things in Fig. 3.1 are striking:

¹Trigger decisions are only set as *trigger bits* at this point and are thus not part of any selection requirement yet.

²In order to unfold relative deviations among different MC simulated decays from biases introduced by fit models, we use truth-matched events for this task.

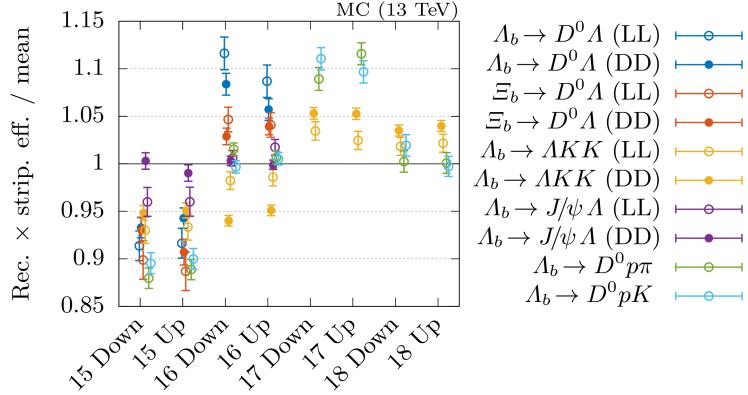


Fig. 3.1. Combined reconstruction and stripping efficiency for the decays under consideration. For the sake of brevity, magnet polarities are referred to as *Down* and *Up* for mag. down and mag. up, respectively. In order to compensate for their wide spread, each value is normalized to the respective weighted mean of each decay for the full available data set. (Not all decays are simulated for the years 2017 and 2018.) For example $y \approx 0.95$ for $\Lambda_b \rightarrow \Lambda KK$ (DD) at $x = 15$ Down reads as a 5 % deviation from the weighted mean of all simulated $\Lambda_b \rightarrow \Lambda KK$ decays with DD tracks, where x and y refer to the abscissa and ordinate, respectively.

- The discrepancy between simulated $\Lambda_b \rightarrow D^0\Lambda$ and $\Xi_b^0 \rightarrow D^0\Lambda$ decays is larger than anticipated: The kinematics of both topologically identical decays should be very similar such that the deviation was more likely introduced by the various updates to the simulation framework (`Sim09c` → `Sim09h/g`). Therefore, we take simulated $\Xi_b^0 \rightarrow D^0\Lambda$ decays as the better proxy of genuine $\Lambda_b \rightarrow D^0\Lambda$ decays than the dedicated $\Lambda_b \rightarrow D^0\Lambda$ simulation for determining the combined generator cut and stripping efficiency.
- The efficiency drop for $\Lambda_b \rightarrow D^0ph^-$ for the year 2018: For the present analysis this drop affects $\Lambda_b \rightarrow D^0p\pi^-$ and stays unclear. For comparison we added the efficiency of MC simulated $\Lambda_b \rightarrow D^0pK^-$ and $\Lambda_b \rightarrow \Lambda K^+K^-$ where the drop is only visible for the former, hinting towards a correlation with the D^0 meson.
- The difference between LL and DD tracks is compatible for $\Lambda_b \rightarrow J/\psi\Lambda$ and $\Lambda_b \rightarrow \Lambda K^+K^-$: Even though the detector response for the former can be very different due to the dimuon pair in the final state, the difference between the track types of the Λ daughters should be similar for the former and the latter. Further, we see that the variation of DD tracks is smaller than for LL tracks in both cases for the years 2015 and 2016.
- Similarly, the double ratio of $\Lambda_b \rightarrow D^0\Lambda$ and $\Xi_b^0 \rightarrow D^0\Lambda$ and both track types is compatible with one. Hence, the ratio of the products of reconstruction and stripping efficiency for $\Lambda_b \rightarrow D^0\Lambda$ and $\Xi_b^0 \rightarrow D^0\Lambda$ does not depend on the track type in good approximation.

The observed deviations do come with uncertainties and pinpointing exact causes is not possible. When reliable values of the efficiencies are needed, we will therefore add a 10 % uncertainty to compensate for deviations that were introduced in the MC simulated events but do not have any counterpart in recorded data. This is a conservative approximation and will likely, based on the presented study, cover the true deviation.

Taking the ratios of the combined reconstruction and stripping efficiency of two Λ_b decay modes gives access to suppression factors up to the stripping process. There are two relevant

cases, first the relative suppression of $\Lambda_b \rightarrow D^0 p\pi^-$ and $\Lambda_b \rightarrow D^0 \Lambda$, needed for estimating the relative branching fraction of both, and secondly, the suppression of physical background contributions of $\Lambda_b \rightarrow D^0 p\pi^-$ and $\Lambda_b \rightarrow \Lambda K^+ K^-$ in the invariant mass of D^0 and Λ candidates.

For estimating the former, we use MC simulated $\Lambda_b \rightarrow D^0 p\pi^-$ events, reconstructed as $\Lambda_b \rightarrow D^0 p\pi^-$, and MC simulated $\Xi_b \rightarrow D^0 \Lambda$ events, reconstructed as $\Lambda_b \rightarrow D^0 \Lambda$ (the difference between reconstructed Λ_b and Ξ_b is only the name tag). The ratio as a function of different simulation conditions and track types of the Λ daughters is shown in Fig. 3.2. The weighted

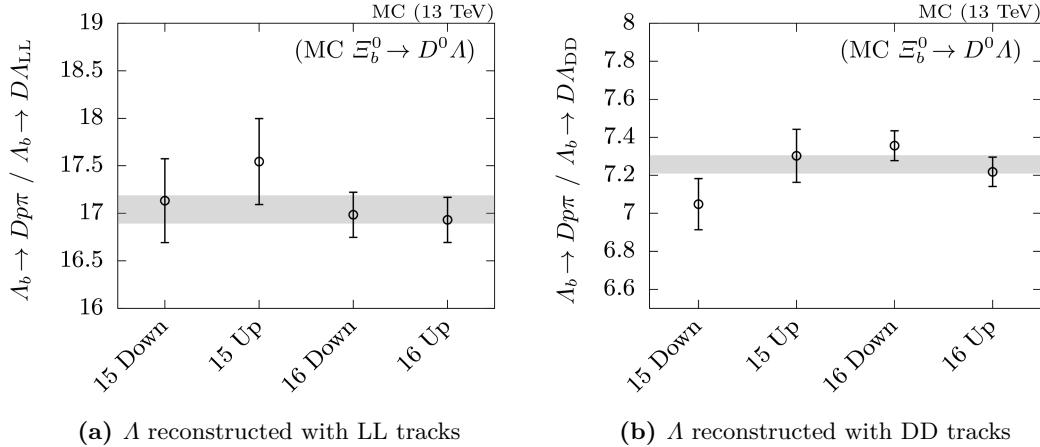


Fig. 3.2. Ratio of the combined reconstruction and stripping efficiency of $\Lambda_b \rightarrow D^0 p\pi^-$ and $\Lambda_b \rightarrow D^0 \Lambda$ where we used $\Xi_b \rightarrow D^0 \Lambda$ simulated decays as a proxy for the latter and differentiate between the track type of the Λ daughters. The weighted means (grey box) are 17.04(15) and 7.26(5) for LL tracks (left) and DD tracks (right), respectively.

means are 17.04(15) and 7.26(5) for LL and DD tracks, respectively.

The suppression factor s of the background contribution of $\Lambda_b \rightarrow D^0 p\pi^-$ in the invariant mass $m(D^0 \Lambda)$ is only relevant for LL tracks and is given by the ratio of the combined reconstruction and stripping efficiency ε of simulated $\Lambda_b \rightarrow D^0 p\pi^-$ decays when reconstructed as $D^0 p\pi^-$ and $D^0 \Lambda$. This suppression factor s reduces the amount n of reconstructed $\Lambda_b \rightarrow D^0 p\pi^-$ decays, determined by an appropriate fitting technique in recorded data³,

$$n/s = n \sqrt{\frac{\varepsilon(\Lambda_b \rightarrow D^0 p\pi^- \rightsquigarrow D^0 p\pi^-)}{\varepsilon(\Lambda_b \rightarrow D^0 p\pi^- \rightsquigarrow D^0 \Lambda)}} ,$$

where \rightsquigarrow indicates the reconstruction state, hence n/s is the expected amount of background contributions in $m(D^0 \Lambda)$ up to the stripping stage.

Since the recorded $\Lambda_b \rightarrow D^0 p\pi^-$ sample allows a clean extraction of n , whereas the available samples of $\Lambda_b \rightarrow \Lambda K^+ K^-$ is noisier due to the smaller branching fraction and more pronounced background contributions (*cf.* Ref. [39]), we also use n for estimating the background contributions of charmless backgrounds. We therefore take the results of the PDG,

$$\begin{aligned} \frac{\mathcal{B}(\Lambda_b \rightarrow \Lambda K^+ K^-)}{\mathcal{B}(\Lambda_b \rightarrow \Lambda_c \pi^-)} &= (3.29 \pm 0.39) \times 10^{-3} , \\ \frac{\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)}{\mathcal{B}(\Lambda_b \rightarrow \Lambda_c \pi^-)} &= 0.13 \pm 0.01 , \end{aligned}$$

³In particular this means $n \neq \#DTT$ as used in Eq. (3.1).

which are derived from the reported results of Refs. [39, 69], and find

$$\kappa := \frac{\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)}{\mathcal{B}(\Lambda_b \rightarrow \Lambda K^+ K^-)} = 40 \pm 6,$$

assuming uncorrelated errors. The latter assumption only holds for the statistical uncertainty strictly since both analyses are using different decay channels. The systematic uncertainties though, do include a non-vanishing correlation, in particular because both analyses were carried out using data from the same detector. Unfolding of correlated and uncorrelated fractions, however, is non-obvious and we will therefore use the conservative assumption of purely uncorrelated contributions which will slightly overestimate the total uncertainty.

Using κ as the correction factor, the amount of reconstructed $\Lambda_b \rightarrow \Lambda K^- K^+$ decays thus reads

$$n/s' = n \left/ \left(\kappa \times \frac{\varepsilon(\Lambda_b \rightarrow D^0 p\pi^- \rightsquigarrow D^0 p\pi^-)}{\varepsilon(\Lambda_b \rightarrow \Lambda K^- K^+ \rightsquigarrow D^0 \Lambda)} \times \frac{\mathcal{B}(D^0 \rightarrow K^- \pi^+)}{\mathcal{B}(\Lambda \rightarrow p\pi^-)} \right) \right).$$

The values of s and s' are shown in Fig. 3.3. The weighted mean values are

$$s = 24.4 \pm 0.4, \quad (3.2a)$$

$$s' = \begin{cases} 530 \pm 80 & (\text{LL}), \\ 210 \pm 30 & (\text{DD}). \end{cases} \quad (3.2b)$$

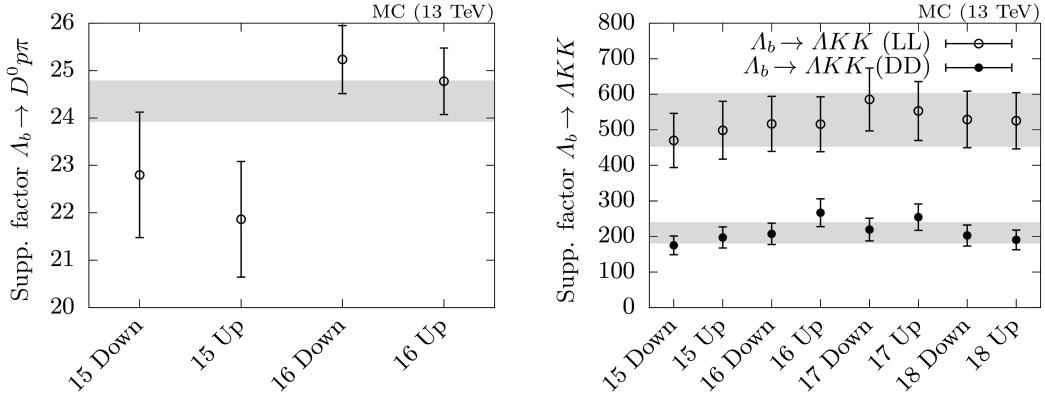


Fig. 3.3. Suppression factors s (left) and s' (right) as defined in Eqs. (3.2) and respective weighted mean values (box).

Chapter 4

Tuning MC Simulated Events by Determining Weights

Light weight, baby!

— Ronnie Coleman.

Monte Carlo (MC) simulated events are a product from chaining various different software frameworks that use MC methods to simulate the interaction and passage of particles through matter. In particular the simulation of the passage of particles through matter highly relies on a high fidelity of the description of the detector assembly. At the same time, a meticulous detector description slows down the simulation significantly and neither geometric, nor material specific properties of all sub-components can be known exactly. For example, it is only possible to measure the magnetic field when the detector is partially disassembled. The spatial distribution of the magnetic field of the assembled detector, which is a crucial input parameter for the particle simulation, thus already relies on error-prone estimations. Further shortcomings are inaccurate alignments of detector units or imprecise physical input parameters, such as polarizations of initial state particles.

Some of these effects also limit the resolution of recorded data, but typically affect recorded data and MC simulated events differently. In the end, recorded data and MC simulated events will never match exactly. In particular, it is known that the distributions of transverse momentum p_T and pseudorapidity η of simulated Λ_b baryons can deviate significantly.

The idea of the following section is to weight simulated Λ_b decays w.r.t. $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$ in such a way that they match the distribution of recorded data. When using the same set of selection requirements for simulated and recorded data, the extracted weights are selection independent scale factors (weights) that transform feature distributions of simulated Λ_b decays to the corresponding distributions of recorded data. In particular this means that those weights should not depend on kinematic properties, such as lifetime (and thus track type) of the respective daughters and grand-daughters of the Λ_b , and thus are applicable also for different decay modes of the Λ_b baryon.¹ In order to minimize trigger dependent deviations we only use L0 TIS triggered $\Lambda_b \rightarrow J/\psi \Lambda$ events to extract weights and use them in a subsequent step for weighting MC simulated $\Lambda_b/\Xi_b^0 \rightarrow D^0 \Lambda$ events.

Properties that deviate between recorded and simulated data due to correlations with $p_T(\Lambda_b)$ or $\eta(\Lambda_b)$ will automatically improve with this technique. We check this by comparing the momentum distribution $p(\Lambda_b)$ of unweighted and weighted MC simulated $\Lambda_b \rightarrow J/\psi \Lambda$ events.

We try two different strategies for extracting weights. The former minimizes deviations for both $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$, whereas the latter uses $p_T(\Lambda_b)$ only. In the subsequent analysis we use the latter approach but incorporate deviations when using the weights obtained with the former strategy as a systematic uncertainty. A description of the selection criteria used to increase the purity of the $\Lambda_b \rightarrow J/\psi \Lambda$ data samples is given in Sec. 4.1. A detailed explanation

¹We further use these weights to correct simulated Ξ_b^0 decays and incorporate deviations as a systematic uncertainty.

and discussion of the two different weighting schemes is given in Sec. 4.2.2 and Sec. 4.2.3. We further briefly motivate the use of sideband subtraction rather than relying on truth-matching in Sec. 4.2.1.

4.1 The Decay $\Lambda_b \rightarrow J/\psi \Lambda$

The decay $\Lambda_b \rightarrow J/\psi \Lambda$ is a high statistics channel at LHCb and was used there in the past to measure for example polarization effects [70]. Due to its large branching fraction this decay was one of the first discovered Λ_b decay channels [71] and measurements of the product of production fraction and branching fraction $f(b \rightarrow \Lambda_b) \times \mathcal{B}(\Lambda_b \rightarrow J/\psi \Lambda)$ at the D0 and the CDF collaborations [72, 73] are still used today for determining branching fractions from measurements of Λ_b branching ratios.

In the following we use $\Lambda_b \rightarrow (J/\psi \rightarrow \mu^-\mu^+)(\Lambda \rightarrow p\pi^-)$ to extract weights for calibrating the MC samples for $\Lambda_b \rightarrow D^0\Lambda$. In order to minimize systematic uncertainties introduced by imprecise simulated trigger responses, only L0 TIS triggered events are used. In the following we will outline the selection steps which we divide into a preselection, a loose, and a tight selection.

4.1.1 Preselection

We use the full recorded data set of Run 2 and the stripping versions listed in Tab. 4.1. Despite their different naming, there are no major differences between different stripping versions for the involved stripping lines. Mother particles are reconstructed from daughter particles that passed

Tab. 4.1. Stripping and Reco versions used for reconstructing $\Lambda_b \rightarrow J/\psi \Lambda$.

Year	Stripping	Reco
2015	24r1	15a
2016	28r1	16
2017	29r2	17
2018	34	18

dedicated *combination* selections. Properties of the reconstructed mother particle are refined through the vertex fit procedure (no advanced constraints, such as mass or PV constraints, are applied) and are subject to dedicated *mother* selections, whereas properties of the respective daughter particles are not updated for subsequent selection steps. All selection criteria of the preselection step are listed in Tab. B.1.

4.1.2 Loose Selection

A decay tree fit is applied and the corresponding χ^2_{DTF} distribution² is used for discriminating combinatorial background. The χ^2_{DTF} distribution has 8 degrees of freedom (*cf.* Sec. C for a more detailed discussion):

- Mass constraint of Λ and J/ψ : 2 DoF
- Λ vertex constraint: 1 DoF
- $\Lambda_b \rightarrow \mu^+\mu^-\Lambda$ vertex constraint: 3 DoF

²We will later find that this distribution does not exactly follow the distribution of a *true* χ^2 -distribution. For the sake of brevity we will nevertheless refer to it as a χ^2_{DTF} .

- Λ_b PV constraint: 2 DoF

The PV constraint leverages an ordering of different PV hypotheses (if available) w.r.t. the goodness of a respective DTF (in terms of χ^2_{DTF}). We use this to select only candidates corresponding to the best PV hypothesis for the following steps.

The selection criteria of the loose selection are shown in Tab. 4.2 and are grouped into five categories: Category 1 reduces combinatorial background in the signal region, category 2 ensures disjunct samples w.r.t. the track types LL and DD, and category 3 reduces disk consumption. The momentum requirements for the final state particles in category 1 are motivated by the fact that particles do have to have a minimal velocity to emit Cherenkov radiation. Cherenkov light is instrumented in the RICH detectors for the particle identification at LHCb. Particle identification (PID) below this threshold and above $\gtrsim 150$ GeV/c is ineffective. (These selection requirements supersede the implicit cut-off of $p > 1.4$ GeV/c due to magnet banding.) The fiducial selection criteria (category 4) are a common choice for initial state particles at LHCb. These selections help to avoid known issues with the fidelity of the detector geometry description in the inner- and outermost regions and turned out to be a conservative choice when the overall statistic is sufficient. Category 5 minimizes the effect of poorly described trigger in simulated data which could potentially introduce a systematic difference between recorded and simulated events.

Tab. 4.2. Selection criteria of loose selection used for reconstructing $\Lambda_b \rightarrow J/\psi \Lambda$. The selections are grouped into five categories which are explained in Sec. 4.1.2.

Particle	Selection	Category
p	$9 \leq p \leq 150$ GeV/c	1
π	$3 \leq p \leq 150$ GeV/c	1
Λ	decay length ≥ 0	1
Λ (LL)	z -pos. of decay vertex < 0.5 m	2
Λ (DD)	z -pos. of decay vertex ≥ 0.5 m	2
Λ_b	$5.47 \leq m \leq 5.77$ GeV/ c^2	3
Λ_b	$2 \leq \eta \leq 4.5$	4
Λ_b	$p_T \leq 20$ GeV/c	4
Λ_b	L0 TIS events only	5

The data samples are split w.r.t. the different track types of the Λ daughters which are referred to as LL and DD. In order to optimize a FoM, we define a signal region spanning $5.58 \leq m(J/\psi \Lambda) \leq 5.66$ GeV/ c^2 . Events outside this region (upper and lower sideband) are considered pure combinatorial background, whereas events inside the signal region are considered to be an admixture of signal and (combinatorial) background events. Physical background processes are neglected in this part of the analysis since no visible contributions are visible in the invariant mass distributions.

4.1.3 Tight Selection

The objective of the tight selection is to maximize the signal significance as the FoM for n_{sig} signal events and n_{bkg} background events in the defined signal region,

$$\text{FoM} := \text{FoM}(n_{\text{sig}}, n_{\text{bkg}}) = \frac{n_{\text{sig}}}{\sqrt{n_{\text{sig}} + n_{\text{bkg}}}}. \quad (4.1)$$

The signal significance is maximized for selection requirements w.r.t. the χ^2_{DTF} distribution of the DTF and the flight distance significance of the Λ baryon, where the latter is defined as

the flight distance $\text{FD}(\Lambda)$ over the corresponding standard deviation $u_{\text{FD}}(\Lambda)$ of the Λ baryon,

$$\Lambda \text{ flight dist. sig.} := \frac{\text{FD}(\Lambda)}{u_{\text{FD}}(\Lambda)}.$$

The cumulative distributions for both of these features are shown in Fig. 4.1 and Fig. 4.2 for recorded data in the defined signal region and MC simulated events. The cumulative distribution of χ^2_{DTF} indicates a strong separation power between signal and background events, whereas the cumulative distributions of the Λ flight distance only show minor differences between recorded data and simulated events, hinting towards a low background contamination for the Λ baryon, *i.e.*, the background in $m(J/\psi \Lambda)$ predominantly consists of combinatorial background events with genuine Λ baryons.

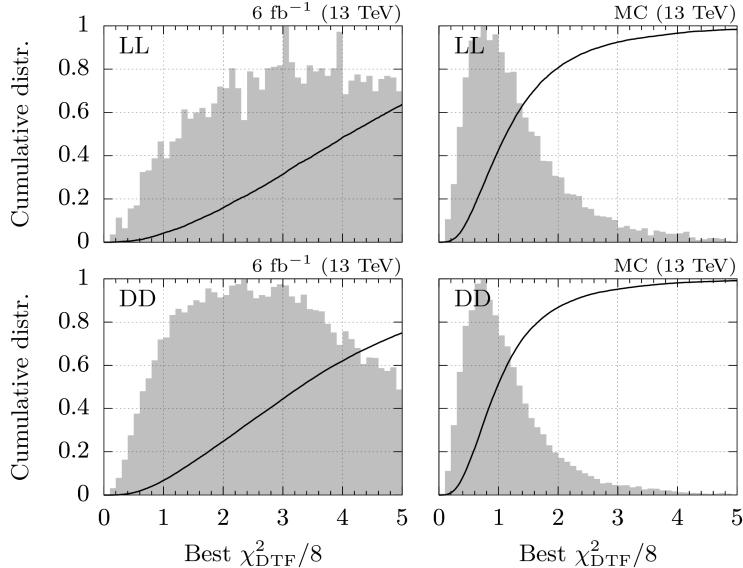


Fig. 4.1. Cumulative distribution (solid line) of the χ^2_{DTF} distribution over DoF for different track types (top and bottom), and for recorded data in the defined signal regions (left) and truth-matched simulated events in the signal region (right). The gray shaded areas indicate the corresponding distributions of $\chi^2_{\text{DTF}}/\text{DoF}$.

The combinatorial background in $m(J/\psi \Lambda)$ (rec. data) is sufficiently linear such that the FOM as defined in Eq. (4.1) can be evaluated with a sideband-subtraction, *cf.* Fig. 4.3. (Definition of the signal and background regions are given in Sec. 4.1.2.) The signal significance for a selection w.r.t. χ^2_{DTF} has a maximum for LL and DD tracks, implying that signal events prefer smaller values of χ^2_{DTF} , *i.e.*, stronger support for the assumed hypothesis of the DTF, and thus motivates the selection criterion for the tight selection

$$\chi^2_{\text{DTF}} \stackrel{!}{\leq} \begin{cases} 3 & (\text{LL}), \\ 2 & (\text{DD}), \end{cases} \quad (4.2)$$

whereas the FOM of the Λ flight distance significance is monotonic for requirements that either prefer large or low values and thus discourage an additional selection w.r.t. this feature. In Fig. 4.4 we show the invariant mass $m(J/\psi \Lambda)$ after applying this and all previously mentioned selection requirements.

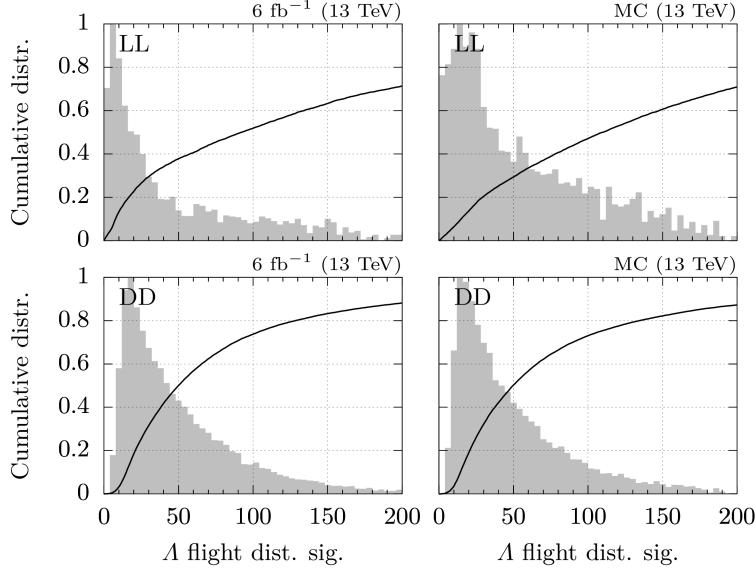


Fig. 4.2. Cumulative distribution (solid line) of the Λ flight distance over the corresponding standard deviation for different track types (top and bottom), and for recorded data in the defined signal regions (left) and truth-matched simulated events in the signal region (right). The gray shaded areas indicate the corresponding distributions of the respective significance of the flight distance itself.

Efficiency Determination of the Tight Selection

The technique of sideband subtraction also leverages the estimation of a data driven efficiency determination of the selections given in Eq. (4.2) by evaluating

$$\varepsilon := \frac{n_{\text{sig}}}{n_{\text{sig}} + \bar{n}_{\text{sig}}} ,$$

where n_{sig} and \bar{n}_{sig} are the amount of signal events that pass the selection and the amount of events that are rejected, respectively. The uncertainty u_ε of each of these figures is given by the uncertainty of the mean background ($f\sqrt{n_{\text{sideband}}} = \sqrt{fn_{\text{bkg}}}$), the fluctuation of the true background around the mean background in the signal region ($\sqrt{n_{\text{bkg}}}$) and the fluctuation of the true signal around the mean signal ($\sqrt{n_{\text{sig}}}$)

$$u_\varepsilon = \sqrt{n_{\text{sig}} + (1 + f)n_{\text{bkg}}} ,$$

where f is the scale factor that translates the number of observed background events in the sideband region n_{sideband} to the estimated number of background events in the signal region $n_{\text{bkg}} = f \times n_{\text{sideband}}$. The amounts n_{sig} and \bar{n}_{sig} are statistically independent, hence the uncertainty of the cut efficiency ε can be extracted by ordinary error propagation. In Tab. 4.3 we show n_{sig} with its associated uncertainty $\sqrt{(1 + f)n_{\text{bkg}}}$, as well as the selection efficiencies ε .

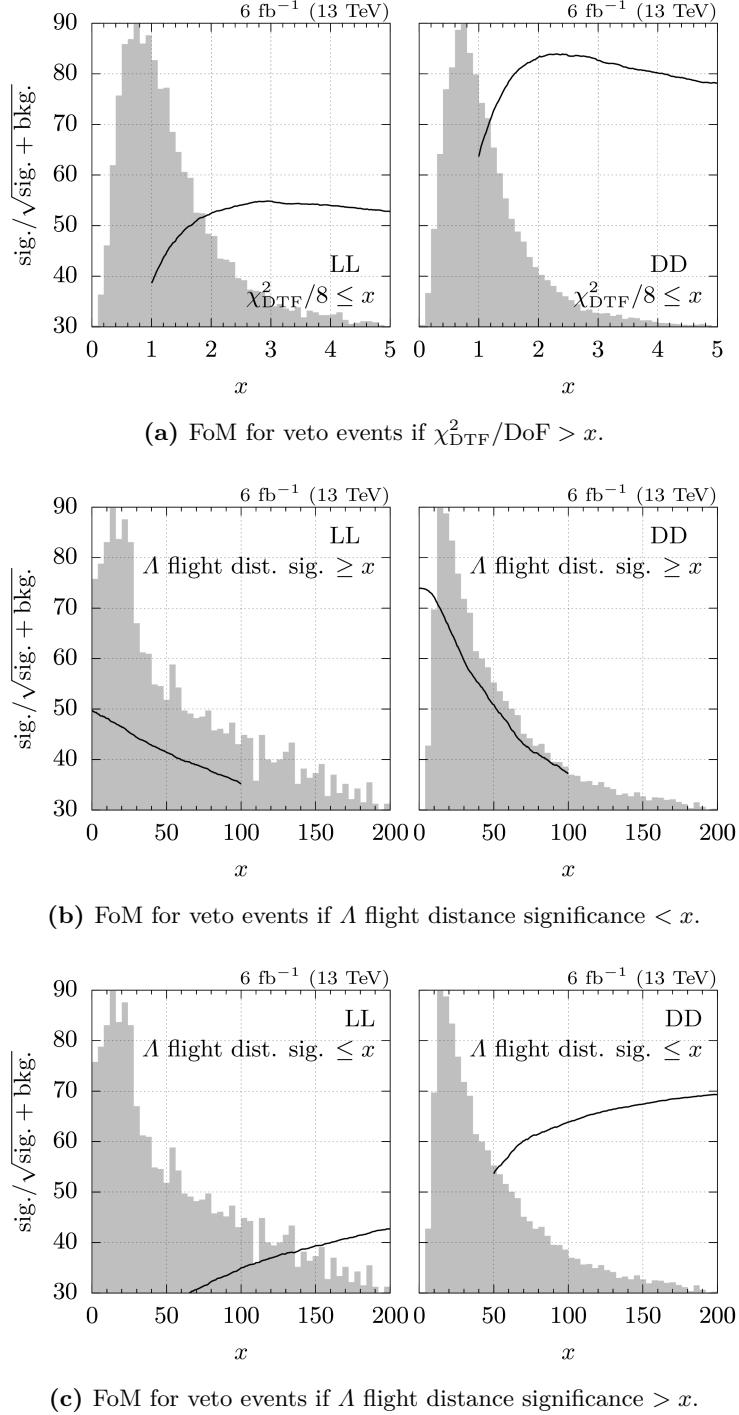


Fig. 4.3. FoM defined as signal efficiency (solid black line) obtained by sideband-subtraction in recorded data and the corresponding distribution of truth-matched simulated events (grey shaded area). The FoM of the Λ flight distance significance is monotonic for requirements that either prefer large or low values and thus discourage an additional selection w.r.t. this feature.

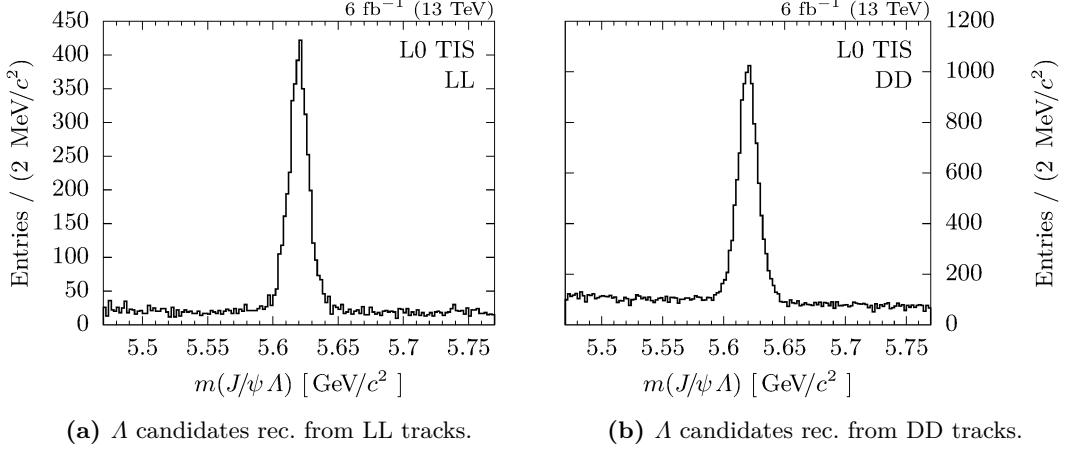


Fig. 4.4. Combined invariant mass of J/ψ and Λ candidates after tight selection from recorded data. (Including L0 TIS only requirement.) This selection is used for determining the weights for tuning MC simulated Λ_b decays.

Tab. 4.3. Total amount of signal events n_{sig} after applying the tight selection to recorded $\Lambda_b \rightarrow J/\psi \Lambda$ data, as well as the respective selection efficiency ε .

	n_{sig}	ε
LL	3653 ± 33	$(91.4 \pm 1.2)\%$
DD	9590 ± 70	$(79.1 \pm 0.9)\%$

4.2 Extraction of Weights

The features transverse momentum $p_T(\Lambda_b)$ and pseudorapidity $\eta(\Lambda_b)$ of the Λ_b are not uncorrelated, but related via the identity relation

$$\eta = \text{artanh} \frac{p_z}{p} = \text{artanh} \sqrt{1 - \left(\frac{p_T}{p}\right)^2},$$

where artanh is the inverse hyperbolic tangent (aka area hyperbolic tangent), and p_z and p the z -component and magnitude of the momentum, respectively.

Due to this non-negligible correlation, altering one distribution will also affect the other and vice versa. Ideally, the extraction of weight factors should be performed in the 2d-plane of both variables. Reliable weights, though, also require decent statistics which turned out to be problematic, especially for LL tracks.

We try two different schemes for extracting weights. First, we only consider the respective marginal distributions and find the weights in an iterative approach. The underlying assumption is a factorization of the weights in the 2d-plane $w(p_T, \eta) = w_1(p_T) \times w_2(\eta)$ and thus, per definition, ignores all correlation contributions. We cross-check this assumption with a third quantity $p(\Lambda_b)$ which is another marginal distribution in the p_T - η space.

In this first scheme we discover unexpected deviations between the distributions of $\eta(\Lambda_b)$ for different track types but at the same time a decent compatibility with one for all $\eta(\Lambda_b)$ weights w.r.t. the given statistical uncertainties. This motivates our second scheme where weights are extracted based on $p_T(\Lambda_b)$ only.

4.2.1 Truth Matched vs. Sideband Subtracted MC Simulated Events

Similar to recorded data, MC simulated events contain not only signal, but also background events coming either from true physical background processes or are combinatoric remnants. A marked difference between recorded and simulated events is that the latter could be attached with a label and thus, in theory, can be unambiguously identified as signal (truth-matched) or background event (*unmatched*). However, in practice this approach suffers from a non-zero mistag probability during reconstruction which introduces an unphysical error. This error has no counter part in recorded data which is critical when correlated with the variables that are the objective of the weighting scheme.

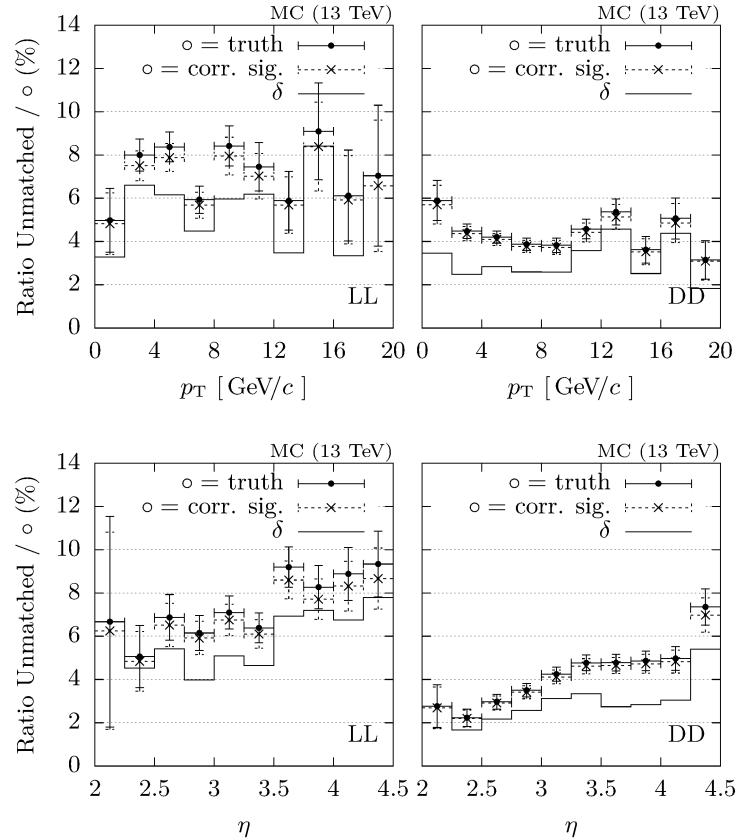


Fig. 4.5. Comparison of the ratio of unmatched and truth-matched (referred to as *truth*) with the ratio of unmatched and sideband corrected events (referred to as *corr. sig.*) in bins of $p_T(\Lambda_b)$ (top) and $\eta(\Lambda_b)$ (bottom) for track types LL (left) and DD (right). The values of δ , as defined in Eq. (4.3), are shown at the same y -axis. The given error bars do not account for the strong correlations between unmatched and truth-matched events.

In the Fig. 4.5 we compare the ratio of unmatched and truth-matched events with the ratio of unmatched and sideband corrected events as a function of $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$. The samples used for determining the ratios are pairwise uncorrelated, whereas the ratios themselves are strongly correlated which has to be taken into account when comparing the distribution of the ratios. We define δ as the correction when using truth-matched events during the weighting

procedure instead of sideband corrected simulated events,

$$1 + \delta := \frac{b'}{b} = 1 + \frac{\Delta}{a/b'}, \quad (4.3)$$

where a , b and b' are the amount of unmatched, truth-matched and sideband corrected simulated events, respectively and Δ is defined as the difference

$$\Delta := \frac{a}{b} - \frac{a}{b'}.$$

We note that in the limit of empty sidebands $b' = b + a$ and thus $\delta = a/b$, as expected. From Fig. 4.5 we infer $\mathcal{O}(\delta) = 5\%$. Non uniform contributions of δ will skew the distributions of the weights. Further, if δ is correlated differently for LL and DD, this introduces an unphysical difference between the track types. (An absolute difference of δ introduces a uniformly distributed difference and is thus less critical.) From these figures, such a correlation cannot be excluded with high confidence and we will therefore not use truth-matched, but sideband corrected MC simulated events for the weighting process.

4.2.2 Scheme 1

In this first scheme we extract weights by iteratively improving the $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$ dependent weight w ,

$$w(p_T, \eta) = w_1(p_T) \times w_2(\eta),$$

and validate its values in the $p(\Lambda_b)$ distribution.

In Fig. 4.6 and Fig. 4.7 we show the distributions of $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$ for recorded data, as well as for simulated events as obtained from sideband subtractions. These distributions are used pairwise to get the binned ratio of recorded data and simulated events for each of these features as shown in Fig. 4.8, where each bin corresponds to the ratio of the respective bin in the histogram of reconstructed and MC simulated data.

After taking the ratios, the histograms are normalized to unity such that in case of a common underlying distribution every bin entry should be one (within uncertainty). The distributions of the ratios show that this is neither the case for $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$, nor for $p(\Lambda_b)$. This deviation is expected and motivates the recalibration of the MC simulated events with weights. In the following we will use the distribution of the three-momentum magnitude $p(\Lambda_b)$ to benchmark the performance of the calibration. If recorded data and simulated events follow the same underlying distribution, the sum

$$\chi^2 \equiv \sum_{i=1}^n \chi_i^2$$

of the normalized deviation from one χ_i^2 for each bin i ,

$$\chi_i^2 \equiv \left(\frac{r_i - 1}{u(r_i)} \right)^2,$$

where r_i and $u(r_i)$ is the central value and its uncertainty of the i -th bin, respectively, is then χ^2 -distributed with n DoF (number of bins). For LL (DD) we find $\chi^2 \approx 21$ ($\chi^2 \approx 109$) and thus reject the hypothesis of a common underlying distribution for recorded data and simulated events on a $> 98\%$ confidence level according to Eq. (D.1).

Weights are calculated by taking the binned, normalized ratio of the marginal distributions of recorded and simulated events for p_T or η in subsequent steps. The data set is split w.r.t. the track types LL and DD. The resulting histograms of ratios $w_1(p_T)$ and $w_2(\eta)$, binned for the given quantity p_T and η , are then used to calculate the p_T and η dependent weight $w(p_T, \eta) := w_1(p_T) \times w_2(\eta)$ for a given simulated event. The iteration procedure is structured as following:

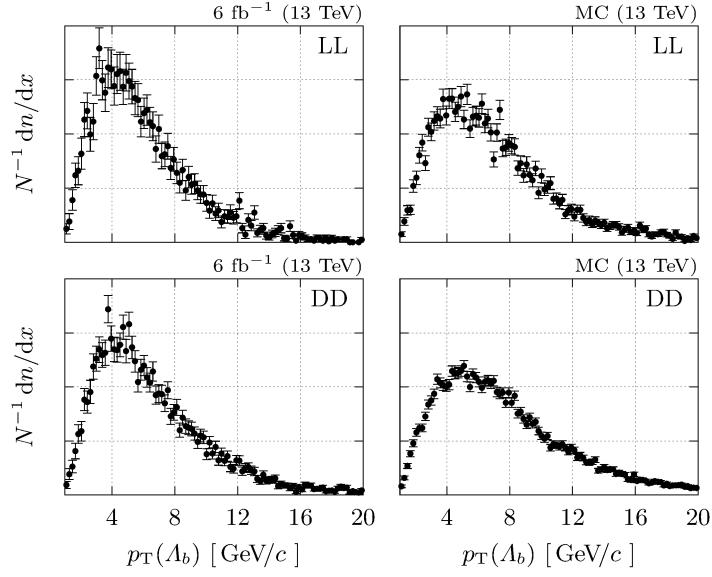


Fig. 4.6. Distributions of the transverse momentum of the Λ_b baryon for recorded data (left), simulated events (right) and different track types LL and DD (top and bottom) as obtained from sideband subtractions.

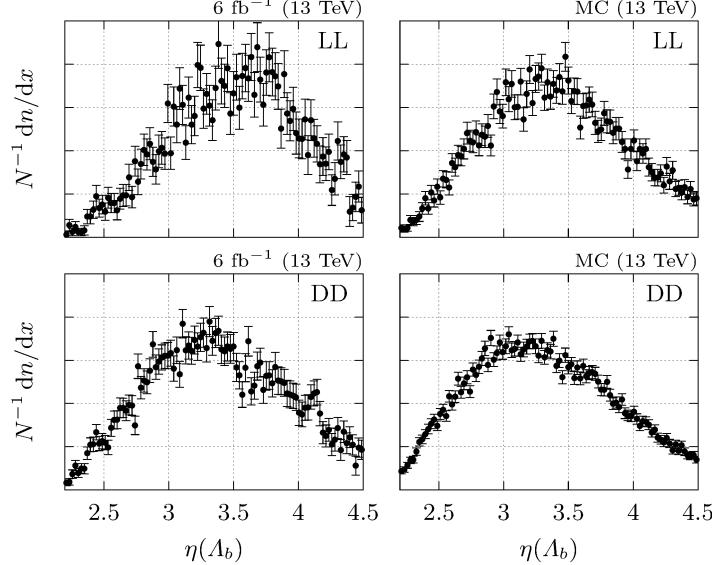


Fig. 4.7. Distributions of the pseudorapidity of the Λ_b baryon for recorded data (left), simulated events (right) and different track types LL and DD (top and bottom) as obtained from sideband subtractions.

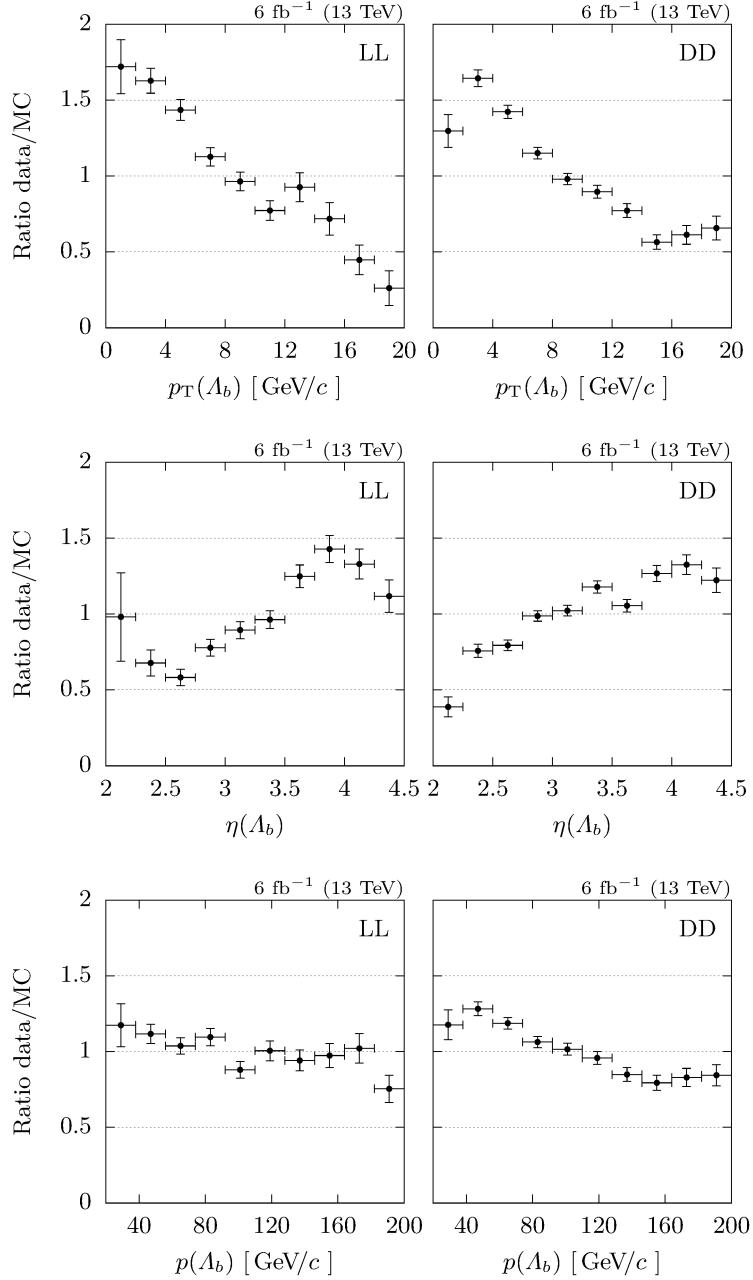


Fig. 4.8. Ratio of the (binned) distributions of the transverse momentum (top), pseudorapidity (middle) and three-momentum magnitude (bottom) of the Λ_b baryon for recorded data and MC simulated events. The ratios are split w.r.t. the track types LL (left) and DD (right).

1. Initialize all weights with one.
2. Update $w_1(p_T)$ using weight factors from the previous iteration.
3. Update $w(p_T, \eta) = w_1(p_T) \times w_2(\eta)$.
4. Update $w_2(\eta)$ using the updated $w_1(p_T)$ and $w_2(\eta)$ from the previous iteration.
5. Update $w(p_T, \eta) = w_1(p_T) \times w_2(\eta)$.
6. Continue with step 2 until convergence is reached.

Each iteration yields a factor $w_1(p_T)$ and $w_2(\eta)$ for a given p_T and η bin. The final weights are their product. The convergence of this approach is measured in the weight update for each bin, separately.

Starting from the first iteration the histograms for (weighted) simulated events are filled with tuples of the particle event with an associated weight (x_i, w_i) . After filling, the content of a bin j is the sum of its weights $w_i^{(j)}$ and the associated uncertainty $u^{(j)}$ is

$$u^{(j)} = \sqrt{\sum_i (w_i^{(j)})^2}.$$

We note that for $w_i^{(j)} = 1 \forall i, j$ this scheme is equivalent to unweighted events where the uncertainty of each bin with bin content n is given by \sqrt{n} .

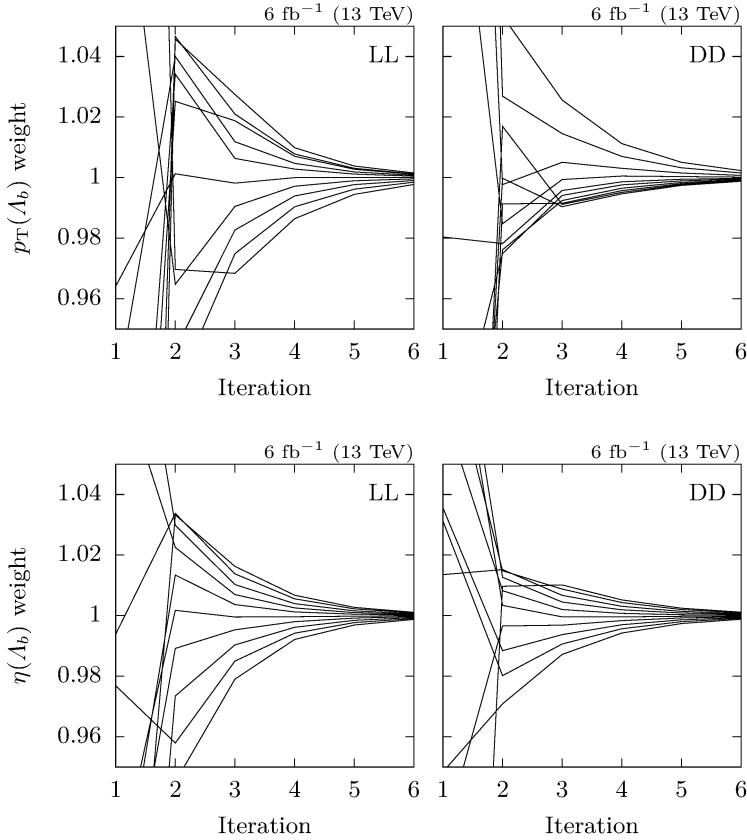


Fig. 4.9. Convergence of $w_1(p_T)$ (top) and $w_2(\eta)$ (bottom) during the iterative weighting process for LL and DD tracks (left and right).

In Fig. 4.9 we show the convergence of the iterative weighting process for LL and DD tracks. Each solid line is the weight update of a bin as a function of the iteration number. Convergence is achieved when all multiplicative updates have approached the value one. In the above case, we stop the iteration after six iterations and consider the product of all weight updates (starting with the value one of the zeroth iteration) as the converged final weight for each bin.

The significance of the obtained weights is quantized in the p -values corresponding to the hypothesis of a common underlying distribution for recorded and simulated events, *i.e.*, $w_1(p_T) = 1$ and $w_2(\eta) = 1$, respectively. These p -values are listed in Tab. 4.4 after six successive iterations. Incompatibilities with these hypotheses are expected and indicate the necessity of the scaling process, whereas the p -value for $r(p) = 1$, where $r(p)$ is the (binned) ratio of the three-momentum magnitude distribution of recorded data and simulated events, exhibits the improvement of features that are scaled implicitly due to correlations with p_T and η . The change of $r(p)$ during the iterations is also listed in Tab. 4.4 and shows the improvement of the fidelity of the MC simulated events due to the scaling with $w(p_T, \eta)$.

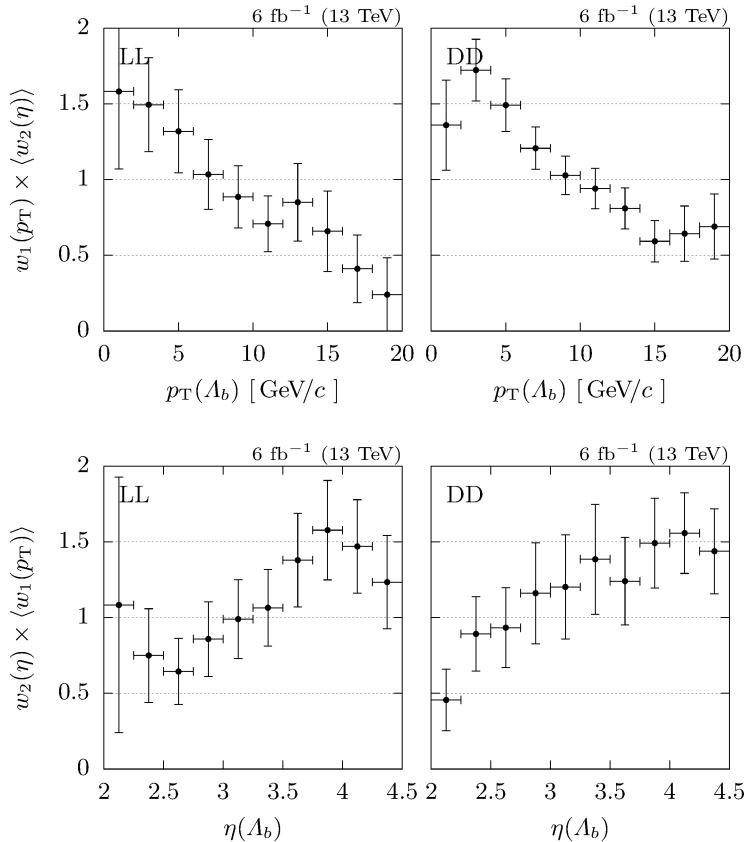


Fig. 4.10. Product of $w_1(p_T)$ and the mean of $w_2(\eta)$ (top) and vice versa (bottom) for LL (left) and DD (right) tracks.

It is worthwhile to mention that the final weights do differ from the initial ratio of recorded and yet unweighted simulated events due to the correlation between p_T and η . It is the product of $w_1(p_T)$ and $w_2(\eta)$ that will eventually reproduce the initially observed ratios, not the marginal distributions themselves. A visualization of this is given in the Fig. 4.10 that shows the product of $w_1(p_T)$ and the mean of $w_2(\eta)$ and vice versa for LL and DD tracks.

A priori, the distribution of weights cannot be expected to be smooth, since each bin is corrected

Tab. 4.4. p -values w.r.t. the hypotheses $w_1(p_T) = 1$ and $w_2(\eta) = 1$ after six successive iterations, as well as the change of the p -value for $r(p) = 1$, where $r(p)$ is the (binned) ratio of the three-momentum magnitude distribution of recorded data and simulated events, during the iterations.

	LL	DD
$w_1(p_T)$	3 %	0 %
$w_2(\eta)$	71 %	59 %
$r(p)$	2 % \mapsto 53 %	0 % \mapsto 24 %

individually. However, due to correlations, each p_T bin is linked to an entire set of η bins and vice versa. A correction of one bin will therefore also influence the weights of neighboring bins, hence a smoothing algorithm can be applied to reduce DoF and compensate uncertainties partially of each bin.

The separate weighting of LL and DD tracks was chosen, because doing so takes different selection criteria and heterogeneous sample sizes trivially into account. However, separate weighting is not physically motivated, since the genuine p_T and η distributions of Λ_b particles should be the same and independent of a specific decay mode or the track type of a (grand-)daughter. In order to include this physical constraint back in, smoothing by fitting (natural) equidistant cubic splines with four DoF (*cf.* Appx. E) is performed separately for $w_1(p_T)$ and $w_2(\eta)$, but simultaneously for the different track types. The resulting fits are shown in Fig. 4.11 together with the distributions of $w_1(p_T)$ and $w_2(\eta)$ and show an unphysical discrepancy between LL and DD tracks for $w_2(\eta)$.

Since the weights $w_1(p_T)$ and $w_2(\eta)$ should neither depend on kinematic properties of Λ_b daughters, nor on the decay channel itself, sensitivity is increased by combining our weights with weights extracted in a $\Lambda_b \rightarrow D^0 p \pi^-$ analysis [74]. (All final state particles are long tracks in this decay.) We find a common set of weights by determining the weighted average of our simultaneously fitted $w(p_T, \eta)$ distribution and the corresponding one of the $\Lambda_b \rightarrow D^0 p \pi^-$ analysis. The fit results are shown in Fig. 4.12 and exhibit a good agreement for $w_1(p_T)$, but large deviations for $w_2(\eta)$. The combination reduces the statistical uncertainty and hints towards a difference between LL and DD that was already visible previously, but yet insignificant. In comparison with the distributions of the non-normalized ratios of LL and DD, this difference is mostly promoted as an accumulation in the ratio of recorded and simulated data for LL tracks and $\eta(\Lambda_b) \gtrsim 3.25$.

In Appx. F we summarize some investigations that exclude various possible explanations for the observed deviations in $w_2(\eta)$. Eventually, the reason for this discrepancy stays unclear and motivates the second weighting scheme, outlined in the next section.

4.2.3 Scheme 2

In the considered sample of $\Lambda_b \rightarrow J/\psi \Lambda$ events, the weights $w_2(\eta)$ are compatible with one in good approximation, *cf.* Tab. 4.4. Hence, a conservative approximation is to use $w(p_T, \eta) = 1 \times w'_2(p_T)$. Since $p_T(\Lambda_b)$ and $\eta(\Lambda_b)$ are correlated, it is not sufficient to simply set $w_2(p_T) \equiv w'_2(p_T)$ in $w(p_T, \eta) = w_1(p_T) \times w_2(\eta)$. Instead the ratio of recorded data and unweighted simulated events is taken as weights. (Since there is only one variable, there is no need to perform this in an iterative approach.)

The ratios are smoothed using a (natural) cubic spline with four DoF (*cf.* Appx. E). Again, the spline is fitted simultaneously to LL and DD tracks. Subsequently, the smoothed weights are used to weight the MC simulated events. The resulting distributions, as well as the corre-

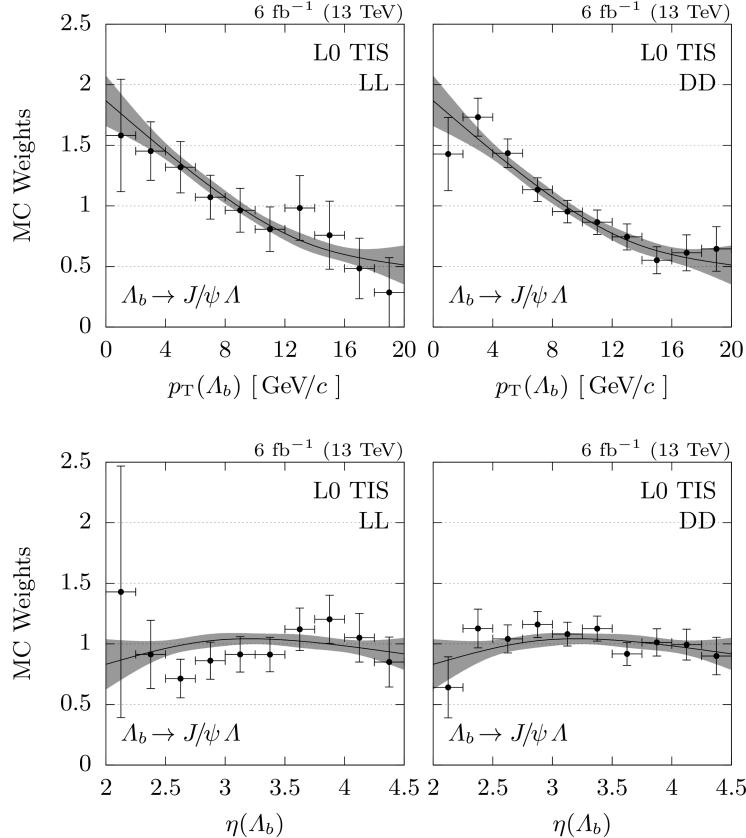


Fig. 4.11. Weights $w_1(p_{\text{T}})$ (top) and $w_2(\eta)$ (bottom) for LL (left) and DD (right) tracks, as well as two equidistant (natural) cubic spline fits, each with four DoF. The distributions for LL and DD tracks are fitted simultaneously.

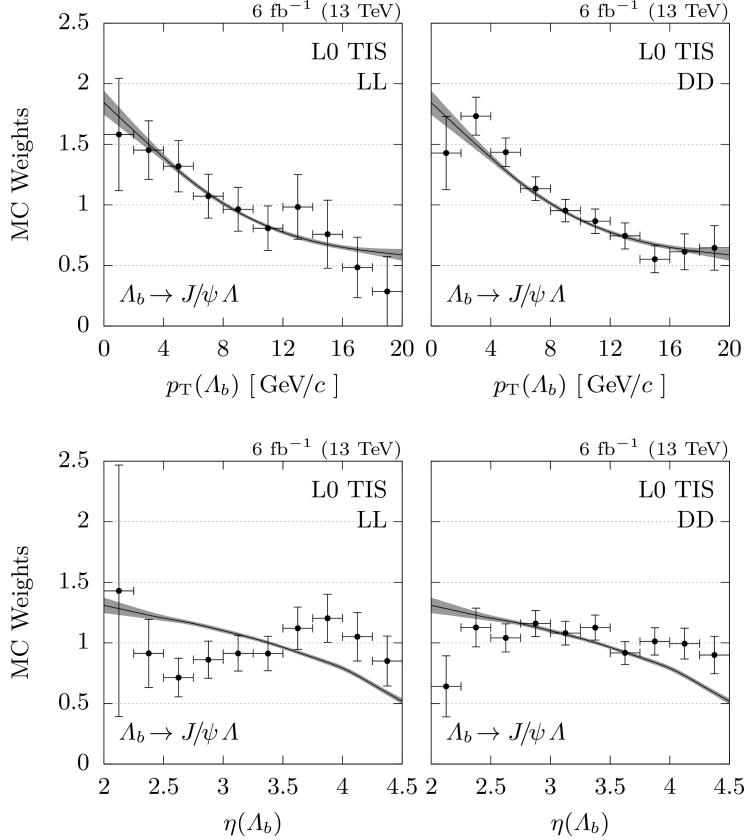


Fig. 4.12. Averaged fit results of weights for $\Lambda_b \rightarrow J/\psi \Lambda$ (simultaneously for LL and DD tracks) and $\Lambda_b \rightarrow D^0 p K^-$ events (taken from Ref. [74]).

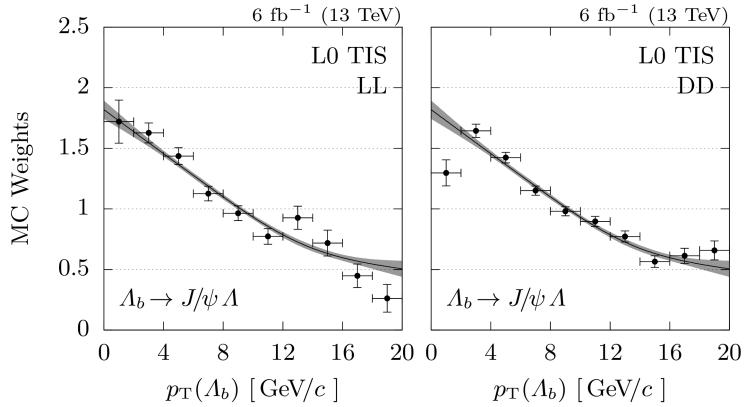


Fig. 4.13. p_T -dependent weights obtained by weighting scheme 2, as well as spline fit, evaluated simultaneously for LL and DD tracks.

sponding spline fits for $p_T(\Lambda_b)$ and both track types are shown in Fig. 4.13.

In Fig. 4.14 we show the ratio of recorded data and weighted MC simulated events for $p_T(\Lambda_b)$ and $p(\Lambda_b)$. The combined p -values for both track types w.r.t. the hypothesis of a common underlying distribution for recorded and simulated events are approximately 8% ($\chi^2 \approx 30$) and 0.1% ($\chi^2 \approx 45$) for the $p(\Lambda_b)$ and $p_T(\Lambda_b)$ distributions, respectively. These values do not include uncertainties of the spline fit and thus reflect only the p -value for a specific choice of weighting function.

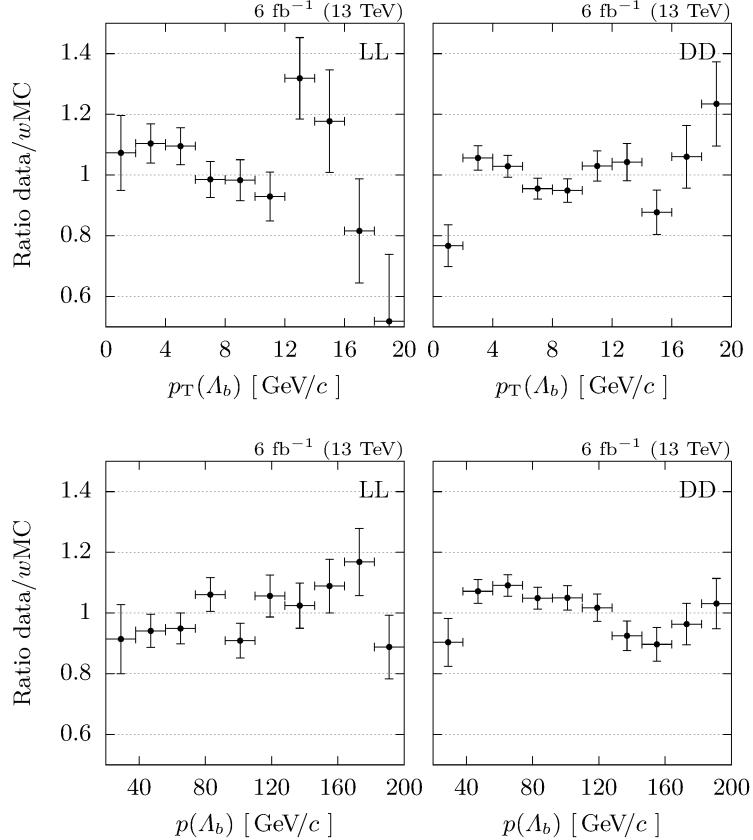


Fig. 4.14. Ratio of recorded data and weighted MC simulated events according to weighting scheme 2 for the transverse momentum of the Λ_b baryon $p_T(\Lambda_b)$ (top) and its three-momentum magnitude $p(\Lambda_b)$ (bottom).

Chapter 5

MVA of the Decay $\Lambda_b \rightarrow D^0 \Lambda$

You wanted a banana but what you got was a gorilla holding the banana and the entire jungle.

— Joe Armstrong, creator of Erlang, on software reusability.

Multivariate analysis (MVA) techniques using machine learning algorithms have a long tradition in particle physics. Ever since the year 2014 (deep) neural networks enrich the pool of utilized algorithms [75]. Over time, their popularity in applications of high energy physics has increased. Moreover, image and four-vector-based neural networks, as well as taggers relying on additional considerations from relativistic kinematics or theory expectations, have outperformed classical approaches [76, 77].

On the one hand, classical approaches that dominated analyses of Run 1 data, such as BDTs or SVMs are still used for classifying *signal* and *background* (*e.g.*, a recent analysis uses a BDT to set an upper limit on $B^+ \rightarrow K^+ \mu^\pm e^\mp$ [78]). The reason for this sustained success of canonical techniques over neural networks is based on the fact that decay reconstructions such as $B^+ \rightarrow K^+ \mu^\pm e^\mp$ or $\Lambda_b \rightarrow D^0 \Lambda$ are characterized by a small set of high level features and classifiers are boosted by specific domain knowledge of these features. Encoding domain knowledge into neural networks, which typically are build upon low level features, is challenging and networks have to spend many iterations until they learn high level features up to a comparable degree to canonical MVA approaches. A large number of iterations requires a large training set which is typically limited by the number of available MC simulated events. In practice, neural networks therefore often cannot catch up with canonical approaches which benefit from their fast convergence.¹ On the other hand, neural networks really shine when the feature set is large and domain knowledge is either unimportant or error prone, for example due to imperfect (MC) simulations. An example for this case is the problem of jet tagging which can be carried out on calorimeter images entirely where the feature space is spanned by each pixel. Here, neural networks have outperformed canonical approaches and pushed authors to even consider dropping theory input from MC simulations completely (*e.g.*, Ref. [76]).

The present analysis clearly falls into the former class. The feature set that we use to describe the decay $\Lambda_b \rightarrow D^0 \Lambda$ comprises 18 features and the data set is limited by the amount of available MC simulated events. We therefore concentrate on training classical MVA algorithms, thoroughly compare their performance and eventually use support vector machines (SVMs) to separate genuine $\Lambda_b \rightarrow D^0 \Lambda$ decays from combinatorial background.

The outline of this chapter is twofold: In Sec. 5.1 we describe the preprocessing step which cleans and reduces the data set up to a point where it becomes feasible to apply MVA algorithms. In Sec. 5.2 we then describe and compare various MVA algorithms which combine into a strong learner that we subsequently use to classify the recorded data of Run 2.

¹Given that the problem can be characterized with a small set of high level features.

5.1 Preprocessing

Before starting with the MVA of the decay $\Lambda_b \rightarrow D^0 \Lambda$ we apply a set of (rectangular) selection criteria grouped in a *preselection* step (*cf.* Sec. 5.1.1) and a *loose selection* (*cf.* Sec. 5.1.2). The preselection is part of the mandatory stripping phase and can be considered immutable for the present analysis. The main focus of the loose selection is to further remove obvious background and outlier candidates from the samples. These filtering steps help against overfitting in general and, from a technical point of view, makes the usage of SVMs as classifiers feasible (*cf.* the $\mathcal{O}(m^{2\cdots 3} \times n)$ scaling as explained in Appx. J).

5.1.1 Preselection

We use the full recorded data set of Run 2 and the output of the stripping line `StrippingLb2D0Lambda0{LL,DD}D02HHBeauty2CharmLine`. The stripping versions are listed in Tab. 5.1. Despite their different naming, the selection requirements vary slightly between versions. These differences are compensated by adopting the tightest selection requirement among conflicting values if necessary.

Tab. 5.1. Stripping and Reco versions used for reconstructing $\Lambda_b \rightarrow D^0 \Lambda$.

Year	Stripping	Reco
2015	24r1	15a
2016	28r1	16
2017	29r2	17
2018	34	18

All selection criteria of the preselection step are listed in Tab. B.2 where we use the same nomenclature that we introduced in Sec. 4.1.1. Additionally, at least one HLT trigger flag among `Hlt2.*IncPhi.*Decision` and `Hlt2Topo.*Decision` is required, *cf.* Refs. [64, 79] for more detailed information.

5.1.2 Loose Selection

Similar to the previous analysis of $\Lambda_b \rightarrow J/\psi \Lambda$ we apply two DTFs to improve the resolution of kinematic features, such as the flight distances of intermediate particles. Both DTFs fit the entire decay chain $\Lambda_b \rightarrow D^0 \Lambda$ and constrain the PV, as well as the mass of the D^0 candidates, but only the second DTF further constrains the mass of Λ candidates. This approach is motivated by the fact that we train a dedicated Λ classifier against combinatorial background in the invariant mass $m(p\pi^-)$, where we extract the distribution of the latter in sidebands of Λ candidates. Since we apply a selection w.r.t. the DTF probability before, a $m(\Lambda)$ constraint would suppress the sidebands completely. The fit probability of the second DTF is used as a feature in the MVA described in Sec. 5.2.2.

We again select only events corresponding to the best PV hypothesis for the following steps. All other selection criteria of the loose selection, as well as their approximated efficiencies are shown in Tab. 5.2, Tab. 5.3 and Tab. 5.4, and are grouped into five categories. Categories 1 to 4 are the same that we introduced previously in Sec. 4.1.2. The purpose of selection requirements of category 5 are to reject physical backgrounds coming from $\Lambda_b \rightarrow D^0 p\pi^-$ and charmless decays. Requiring a finite flight distance reduces these otherwise irreducible backgrounds effectively as shown and discussed in dedicated chapters.

The efficiencies listed in Tab. 5.3 and Tab. 5.4 are approximations of the signal and background efficiencies, where we use truth-matched MC simulated events for the former and 50k

randomly drawn events from the recorded data set for the latter. We cannot use calibrated events for the former since most of the rejected events lie outside the definition range of the calculated weights. We consider this a minor flaw since the estimated efficiencies are well above 90 % for each of the selection requirements individually and these efficiencies only contribute in second order to our ratio measurement with $\Lambda_b \rightarrow D^0 p\pi^-$ decays where similar deviations are expected and common fidelity issues thus cancel. The background efficiency is approximated by counting the total amount of recorded data before and after requiring a specific selection criterion. This approach is motivated by the fact that at this stage the overwhelming majority of the recorded data are not genuine $\Lambda_b \rightarrow D^0 \Lambda$ decays and can thus be considered as background only events in good approximation.

Besides the physically motivated selection requirements w.r.t. the flight distance of the Λ baryon, the most critical selection in terms of possible signal efficiency loss, as shown in Tab. 5.3 and Tab. 5.4, is the requirement

$$3 \leq p(\pi) \leq 150 \text{ GeV}/c,$$

where $p(\pi)$ refers to the three-momentum magnitude of the π^- meson (Λ decay). This selection is motivated if reliable PID information for this particle is necessary, *e.g.*, for rejecting physical background processes. This is not the case for the $\Lambda \rightarrow p\pi^-$, hence we skip this criterion and will not use PID information for this meson in the following.

The total efficiencies of the loose selections for the MC simulated events are 72.90(35) % and 75.52(23) % for LL and DD tracks, respectively. The uncertainties are found by evaluating the variance $\sqrt{\sigma^2(p)} = \sqrt{p(1-p)/N}$ or equivalently by ordinary error propagation of pN and $(1-p)N$, assuming Poisson uncertainties, where p is the efficiency and N the total amount of MC simulated events.

Tab. 5.2. Selection criteria of the loose selection used for reconstructing $\Lambda_b \rightarrow D^0 \Lambda$. The selections are grouped into five categories which are explained in Sec. 5.1.2. To avoid ambiguity we refer to the daughters of the $D^0 \rightarrow K^-\pi^+$ decay as h . The selection requirement w.r.t. the three-momentum magnitude of the pion from the decay $\Lambda_b \rightarrow p\pi^-$ (marked with †) is skipped for the subsequent analysis.

Particle	Selection	Category
p	$9 \leq p \leq 150 \text{ GeV}/c$	1
π^\dagger	$3 \leq p \leq 150 \text{ GeV}/c$	1
h	$3 \leq p \leq 150 \text{ GeV}/c$	1
Λ (LL)	z -pos. of decay vertex $< 0.5 \text{ m}$	2
Λ (DD)	z -pos. of decay vertex $\geq 0.5 \text{ m}$	2
Λ (LL)	$10 \leq \text{FD sig.} \leq 500$	5
Λ (DD)	$0 \leq \text{FD sig.} \leq 500$	5
D^0	$0 \leq \text{FD sig.} \leq 100$	5
D^0	$\Delta\chi_{\text{IP}}^2$ w.r.t. best PV ≥ 5	1
Λ_b	$5 \leq m \leq 6.1 \text{ GeV}/c^2$	3
Λ_b	$2 \leq \eta \leq 4.5$	4
Λ_b	$p_T \leq 20 \text{ GeV}/c$	4
Λ_b	$\Delta\chi_{\text{IP}}^2$ w.r.t. best PV ≤ 25	1
Λ_b	\exists converged DTF	1
Λ_b	$\chi_{\text{DTF}}^2/\text{DoF} \leq 10$ (DTF w/o $m(\Lambda)$ constraint)	1

Tab. 5.3. Approximations of the signal and background efficiencies for LL tracks where we use truth-matched MC simulated events for the former (MC) and 50k randomly drawn events from the recorded data set for the latter (rec. data). Besides the single selection efficiency (*i.e.*, reduction if only the corresponding criterion is used) we also list the efficiency drop, when using all but the corresponding criterion vs. the efficiency of the full selection criteria and refer to it as the *difference*. To avoid ambiguity we refer to the daughters of the $D^0 \rightarrow K^- \pi^+$ decay as h^- and h^+ , respectively.

Particle	Selection	single		difference	
		rec. data [%]	MC [%]	rec. data [%]	MC [%]
p	$9 \leq p \leq 150 \text{ GeV}/c$	96.94(8)	98.97(8)	0	0
π^a	$3 \leq p \leq 150 \text{ GeV}/c$	89.23(14)	90.92(23)	-0.49	-6.53
h^-	$3 \leq p \leq 150 \text{ GeV}/c$	94.27(10)	98.86(8)	-0.26	-0.83
h^+	$3 \leq p \leq 150 \text{ GeV}/c$	94.20(10)	98.82(9)	-0.27	-0.79
Λ	z -pos. of decay vertex $\geq 0.5 \text{ m}$	93.28(11)	95.13(17)	-0.47	-3.69
A	$0 \leq \text{FD sig.} \leq 500$	40.28(22)	90.67(23)	-6.53	-6.78
D^0	$0 \leq \text{FD sig.} \leq 100$	54.92(22)	94.09(19)	-3.90	-4.24
D^0	$\Delta\chi_{\text{IP}}^2$ w.r.t. best PV ≥ 5	93.80(11)	99.02(8)	-0.57	-0.52
Λ_b	$5 \leq m \leq 6.1 \text{ GeV}/c^2$	71.72(20)	100	-2.10	0
Λ_b	$2 \leq \eta \leq 4.5$	94.45(10)	98.08(11)	-0.31	-1.21
Λ_b	$p_{\text{T}} \leq 20 \text{ GeV}/c$	93.06(11)	96.59(14)	-0.13	-2.47
Λ_b	$\Delta\chi_{\text{IP}}^2$ w.r.t. best PV ≤ 5	99.870(16)	100	0	0
Λ_b	\exists converged DTF	87.18(15)	99.70(4)	-0.32	-0.04
Λ_b	$\chi_{\text{DTF}}^2/\text{DoF} \leq 10$ (DTF w/o $m(\Lambda)$ constraint)	56.36(22)	99.70(4)	-0.32	-0.04
	Combination	5.02(10)	72.90(35)		

^anot used

Tab. 5.4. Approximations of the signal and background efficiencies for DD tracks where we use truth-matched MC simulated events for the former (MC) and 50k randomly drawn events from the recorded data set for the latter (rec. data). Besides the single selection efficiency (*i.e.*, reduction if only the corresponding criterion is used) we also list the efficiency drop, when using all but the corresponding criterion vs. the efficiency of the full selection criteria and refer to it as the *difference*. To avoid ambiguity we refer to the daughters of the $D^0 \rightarrow K^- \pi^+$ decay as h^- and h^+ , respectively.

Particle	Selection	single			difference		
		rec. data [%]	MC [%]	rec. data [%]	MC [%]	MC [%]	
p	$9 \leq p \leq 150 \text{ GeV}/c$	92.81(12)	99.680(30)	-0.03	0		
	$3 \leq p \leq 150 \text{ GeV}/c$	75.15(19)	93.14(13)	-1.51	-4.92		
π^a	$3 \leq p \leq 150 \text{ GeV}/c$	95.64(9)	99.10(5)	-0.43	-0.74		
h^-	$3 \leq p \leq 150 \text{ GeV}/c$	95.67(9)	99.09(5)	-0.46	-0.79		
h^+	$3 \leq p \leq 150 \text{ GeV}/c$						
Λ	z -pos. of decay vertex $< 0.5 \text{ m}$	76.16(19)	94.14(12)	-0.66	-4.22		
Λ	$10 \leq \text{FD sig.} \leq 500$	82.47(17)	97.45(8)	-0.24	-1.86		
D^0	$0 \leq \text{FD sig.} \leq 100$	55.64(22)	93.46(13)	-7.63	-4.80		
D^0	$\Delta\chi^2_{\text{IP}}$ w.r.t. best PV ≥ 5	95.27(9)	99.51(4)	-1.42	-0.33		
Λ_b	$5 \leq m \leq 6.1 \text{ GeV}/c^2$	64.47(21)	100	-5.96	0		
Λ_b	$2 \leq \eta \leq 4.5$	95.64(9)	99.24(5)	-0.32	-0.54		
Λ_b	$p_T \leq 20 \text{ GeV}/c$	98.85(5)	92.94(14)	-0.10	-5.73		
Λ_b	$\Delta\chi^2_{\text{IP}}$ w.r.t. best PV ≤ 25	99.564(29)	100	-0.02	0		
Λ_b	\exists converged DTF	94.38(10)	99.32(4)	-0.25	-0.28		
Λ_b	$\chi^2_{\text{DTF}}/\text{DoF} \leq 10$ (DTF w/o $m(\Lambda)$ constraint)	60.54(22)	99.49(4)	-6.67	-0.31		
	Combination	11.01(14)	75.52(23)				

^anot used

5.2 Tight Selection using MVA Techniques

In this section we study different MVA techniques to maximize the signal efficiency of $\Lambda_b \rightarrow D^0 \Lambda$, while at the same time optimizing the suppression of the combinatorial background. The suppression of physical background processes, such as charmless decays $\Lambda_b \rightarrow \Lambda K^- K^+$, are not the main focus of this selection step, albeit their suppression will clearly benefit from the outlined techniques, and will be studied later. The related decay $\Xi_b^0 \rightarrow D^0 \Lambda$ has the same topological signature as $\Lambda_b \rightarrow D^0 \Lambda$ and a very similar kinematic behavior is expected. Optimizing the signal efficiency of $\Lambda_b \rightarrow D^0 \Lambda$ thus also increases the sensitivity to this decay.

We train a classifier which itself is built upon both trivial and complex sub-classifier, to distinguish between *signal* and (combinatorial) *background* events on labeled training data. The training data set $X \in \mathbb{R}^{m \times n}$ is an admixture of calibrated MC simulated events for the former (label *signal*) and recorded data for the latter (label *background*). The events of class *signal* are calibrated truth-matched MC simulated events where the calibration factors are the weights that we established in Chap. 4. These factors are taken into account during the training of our classifiers² and during the evaluation. Technically, the calibration is encoded into a vector of event based weights $w \in \mathbb{R}^m$ which is passed to the classifier together with the associated labels $y \in \mathbb{R}^m$. In total we use $n = 18$ features for the classification where one is the fit probability of a DTF of the entire decay chain $\Lambda_b \rightarrow D^0 \Lambda$ including a PV constraint and mass constraints for the D^0 and Λ candidates (seven DoF in total), and two of them, $\text{ProbNNp}(p)$ and $\text{ProbNNk}(K)$, are themselves the output of (pre-trained) neural networks that fuse PID information as described in Sec. 2.1.7. The remaining 15 features are the following five kinematic properties of the Λ , D^0 and Λ_b candidates:

1. Transverse momentum p_T .
2. DIRA w.r.t. the best PV, transformed as $-\lg(1 - \text{DIRA})$.
3. $\Delta\chi_{\text{IP}}^2$, transformed as $\log_{10}(\Delta\chi_{\text{IP}}^2)$.
4. Fit probability of the end vertex.
5. Significance of the flight distance. (Refined by a DTF which constrains the PV as well as the masses of the Λ , D^0 and Λ_b candidates.)

The kinematic properties are split into two disjunct groups (5+10) and are used as the features of two separate classifiers, a Λ classifier using only Λ properties and a Λ_b - D^0 classifier which uses the remaining 10 properties as features. Eventually, the output of the classifiers, the fit probability of the DTF, as well as $\text{ProbNNp}(p)$ and $\text{ProbNNk}(K)$ responses are fed into a second classifier tier, as shown in Fig. 5.1. The reason behind the outlined design decision, as well as detailed explanations of the classifiers are given in the following sections and in Appx. M.

5.2.1 The Data Pipeline

Before the data are fed into the different MVA algorithms, certain transformation and filtering steps have to be applied. Most steps are necessary for the training and evaluation step (*e.g.*, standardization), whereas others (*e.g.*, balancing) should only be applied during training. Formally, we split the training and evaluation phase into the two separate steps *fit* and *transform* and implement them as steps of a pipeline as shown in Fig. 5.2. During the *fit* step only certain rows (*i.e.*, instances) of X , y and w are selected which reflects balancing and cross-validation. (The technique of cross-validation as used in the present analysis is explained in Appx. H.) Features that are used for the training of one of the classifiers are

1. standardized,

²SVMs as well as decision trees are capable of taking instance based weights into account.

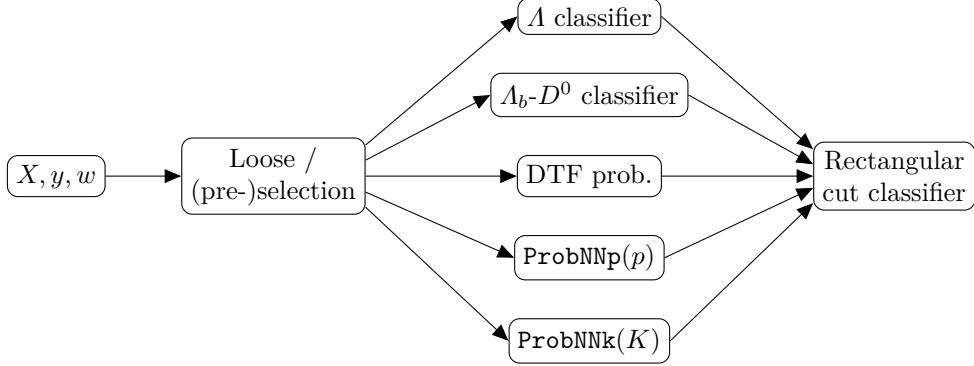


Fig. 5.1. Data flow of the MVA. The data matrix $X \in \mathbb{R}^{m \times n}$ (m instances and n features), respective labels $y \in \mathbb{R}^m$ and weights $w \in \mathbb{R}^m$ are filtered by the preselection and loose selection step and then pass through two tiers of MVAs.

2. decorrelated according to the respective signal distributions via a PCA transformation, and
3. subsequently ordered w.r.t. the Wasserstein distance (signal vs. background) of the respective PCA components,

where the last two steps are skipped for SVMs. Each of these steps involves a trivial fitting step (*e.g.*, finding the mean and standard deviation for the standardization) which then yields parameters that are used to transform the data. These parameters are obtained during the training step (*fit* pipeline) and are kept constant during the evaluation step. Fixing these values during the latter step is important, since, for instance, the mean of a given feature can deviate between the training and testing sample (*i.e.*, $(X, y, w) \neq (X', y', w')$ in Fig. 5.2).

The standardization of features helps to improve the performance of classifiers that are sensitive to the (different) scales of the features (*e.g.*, SVMs), whereas the decorrelation step increases the performance of classifiers that are based on rectangular selection requirements (*e.g.*, decision trees). We note that the evaluation of the mean and standard deviation, as well as the correlation matrix is carried out on a balanced subset of the available data (*fit* pipeline) and will therefore not necessarily standardize or decorrelate the entire data set. The ordering of the PCA components allows a first order approximation of the importance hierarchy of the features. The decorrelation via PCA transformation, as well as the subsequent ordering via the Wasserstein metric is elaborated in Appx. I.

5.2.2 The Λ Classifier

The Λ classifier is designed to increase the purity of the Λ candidates, *i.e.*, the trained classifier should have learned to identify genuine $\Lambda \rightarrow p\pi^-$ decays and reject random combinations, as well as reflections. It is therefore trained on the lower and upper sidebands of the invariant mass distribution $m(p\pi^-)$ (background) and with calibrated truth-matched MC simulated events (signal).

The detector response critically depends on the lifetime of the Λ baryon (*i.e.*, the track type of the Λ daughters) as discussed previously. We therefore split the samples w.r.t. to the track types LL and DD and train two separate classifiers. The invariant mass distribution $m(p\pi^-)$ as well as the sideband boundaries are shown in Fig. 5.3. The numerical values for the sideband boundaries are listed in Tab. 5.5.

In order to find the efficiency of the final classifier, we split the data sample into a training and a test sample. Subsequently, the classifier is optimized and trained on the former and

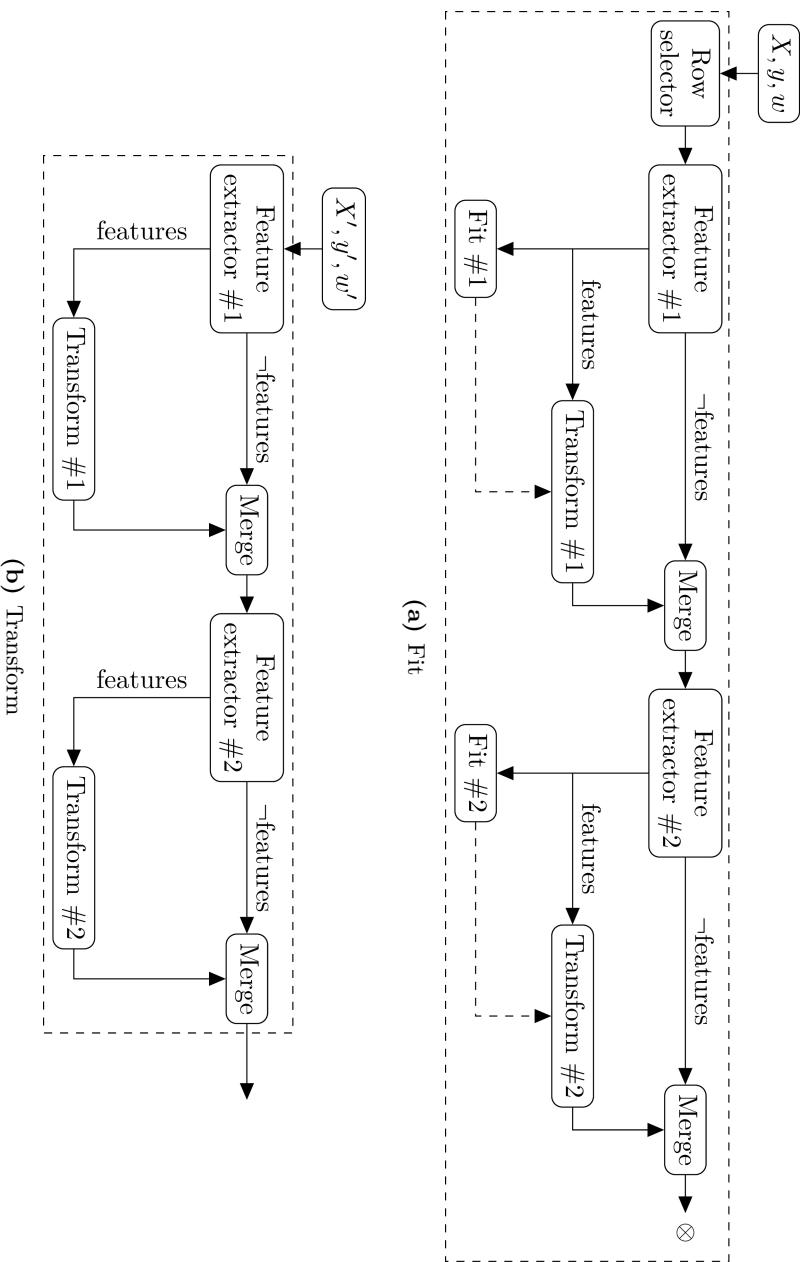


Fig. 5.2. Fit and transform step as described in Sec. 5.2.1. The former fits parameters of various transform steps (*e.g.*, *Transform #1*) which are fixed during evaluation of the latter.

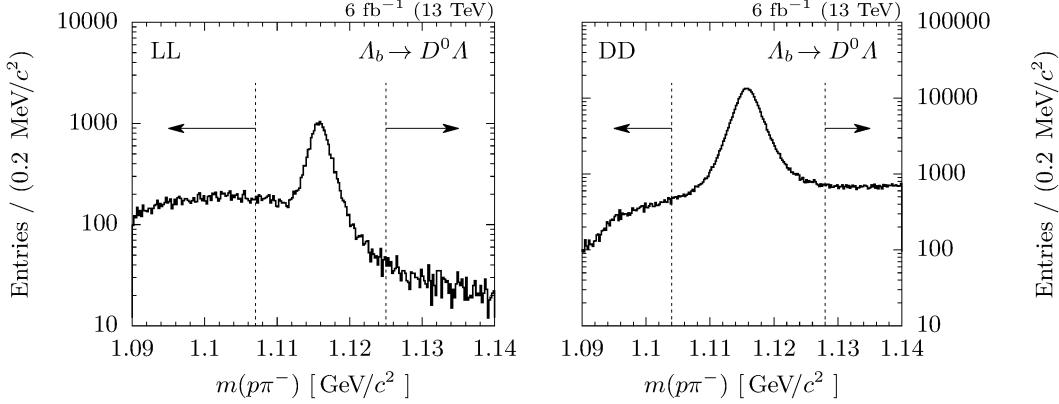


Fig. 5.3. Distribution of the invariant mass of Λ candidates after the loose selection for the different track types LL (left) and DD (right). The dashed lines indicate the sidebands used for training the Λ classifier.

Tab. 5.5. Definition of the low and high edge of the lower and upper sidebands of $m(p\pi^-)$. The different sizes for LL and DD tracks are motivated by the different signal peak resolution and the amount of available data.

	lower sideband		upper sideband	
	low [GeV/ c^2]	high [GeV/ c^2]	low [GeV/ c^2]	high [GeV/ c^2]
LL	1.090	1.107	1.125	1.140
DD	1.090	1.104	1.128	1.140

eventually evaluated on the latter. The ratio of the sizes of the training and test samples is the objective of a min-max problem: A large training sample allows low-biased, complex classifier but the evaluation of the efficiency on the testing sample also comes with a large uncertainty due to the low statistics of this sample. Let p be the true signal efficiency of the classifier that categorizes n_{sig} signal and n_{bkg} background events from $n = n_{\text{sig}} + n_{\text{bkg}}$ given instances:

$$p := \frac{n_{\text{sig}}}{n}.$$

Its uncertainty u_p is then given by the standard deviation of a binomial distribution

$$\frac{u_p}{p} = \sqrt{\frac{1-p}{p} \frac{1}{n}}.$$

For estimation, n is the number of (MC simulated) signal events in the testing sample and n_{sig} (n_{bkg}) is the number of TP (FN). In Fig. 5.4 we show the amount of (MC simulated) signal events that corresponds to 1 %, 5 % and 10 % uncertainty of the classifier efficiency as a function of the efficiency itself. We choose $n = 1000$ which corresponds to a sub 5 % uncertainty for classifier efficiencies above 30 %.

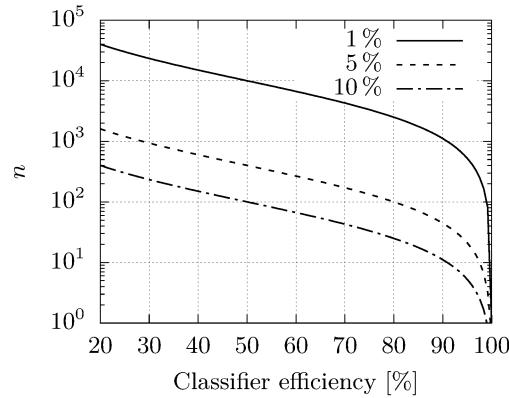


Fig. 5.4. Amount of events n that corresponds to 1 %, 5 % and 10 % uncertainty of the classifier efficiency as a function of the efficiency itself.

It is well known that imbalanced data in classification tasks can reduce classification power of classifiers dramatically by having a bias towards the majority classes in the data set [80, 81]. We therefore balance our dataset each time before training one of our classifiers and since the accuracy is limited by the amount of available (MC simulated) signal events anyhow, we also prune our dataset with DD tracks for the evaluation steps for the sake of computational ease. In Tab. 5.6 we list the sizes of the pruned trainings and testing sets for LL and DD tracks. An additional pruning to balance the data set, as described in Sec. 5.2.1, happens during the training step and is not reflected in Tab. 5.6.

We train five classifiers (twice for LL and DD tracks) using different machine learning algorithms for each (tier 1.1), as well as a stacking classifier (tier 1.2), which combines the output of the former five classifiers. In Tab. 5.7 we list the classifiers, as well as their respective hyper-parameters and whether or not they were optimized or fixed during the training phase.

Tier 1.1

The five tier 1.1 classifiers are

Tab. 5.6. Sizes of training and testing samples for LL and DD tracks used for the MVA. (The background sample (rec. data) was pruned in a previous step.)

Label	LL		DD	
	Training	Testing	Training	Testing
Signal	10 491	1000	26 176	1000
Background	10 151	10 000	50 000	10 000

Tab. 5.7. Hyper-parameters of the Λ classifiers and $\Lambda_b\text{-}D^0$ classifiers. Parameter values are optimized via a 5-fold cross-validation if not marked with \dagger .

Classifier	Hyper-parameter	Value	
		Λ Clf.	$\Lambda_b\text{-}D^0$ Clf.
SVM	Kernel	RBF †	RBF †
	γ (kernel coefficient)	0.1	0.1
	C (regularization)	50 (LL), 1 (DD)	1 (LL), 5 (DD)
Extra Trees	Criterion	Gini †	Gini †
	Number of trees	100	100
	Max. depth	10	20
Random Forest	Criterion	Gini †	Gini †
	Number of trees	100	100
	Max. depth	10	20
BDT	Loss	Deviance †	Deviance †
	Criterion	Friedman MSE †	Friedman MSE †
	Number of trees	100	100
	Max. depth	5	5
Ada. BDT	Loss	Exponential †	Exponential †
	Criterion	Friedman MSE †	Friedman MSE †
	Number of trees	100	100
	Max. depth	5	5
Stacking	Kernel	RBF †	—
	γ (kernel coefficient)	0.1	—
	C (regularization)	10	—

1. a SVM,
2. Extra Trees,
3. a Random Forest,
4. a BDT using gradient boosting, and
5. a BDT using adaptive boosting.

We give a short introduction, as well as an explanation of the hyper-parameters listed in Tab. 5.7 for SVMs in Appx. J, as well as for the ensemble learning algorithms (2. - 5.) in Appx. K. The hyper-parameters are optimized in a 5-fold cross-validation scheme (*cf.* Appx. H) where the optimization objective is the ROC-AUC in the aggregation of the testing folds (*i.e.*, not the previously separated testing set). In Fig. 5.5, Fig. 5.6 and Fig. 5.7 we show the ROC-AUC on the testing and the training folds for different hyper-parameter configurations. Plotting the latter has the advantage to easily identify configurations that suffer from overfitting (large difference between ROC-AUC values), and low variance but high bias configurations (small ROC-AUC values). The desired maximum is typically located at the point in configuration space where the ROC-AUC values between testing and training folds start to diverge. The evaluation of the SVMs are computationally intense (*cf.* Appx. J). We therefore only scan two slices of the two-dimensional $C\text{-}\gamma$ configuration space, whereas we use full grid-search for optimizing the decision tree ensembles. The points $(C, \gamma) = (50, 0.1)$ and $(C, \gamma) = (1, 0.1)$ for LL and DD tracks, respectively, are scanned twice by the former technique and yield slightly different ROC-AUC values due to the randomized partitioning.³

Ever since the advent of neural networks, early stopping (*e.g.*, Ref. [82]) has become a commonly used regularization technique approach to prevent models to perform badly in the testing set after training a certain amount of iterations⁴ (*e.g.*, Ref. [83]). The very same behavior due to overfitting is observed when using boosting techniques in particular, or gradient descent techniques in general, and early stopping techniques have proven to be an effective antidote [84, 85]. Whereas additional regularization is required for these kind of learners, a key advantage of SVMs is their intrinsic regularization such that additional regularization by applying early stopping is not necessary here, albeit occasionally applied in order to shortcut training time [86]. We therefore use early stopping techniques for training the ensemble learners (at each point in the hyper-parameter configuration space) to prevent overfitting and rely on intrinsic regularization of SVMs. A consistency check for the latter assumption is given by the convergence behavior of the SVMs shown in Fig. 5.8.

Tier 1.2 (Stacking)

In the second tier we aggregate the output of the classifiers of tier 1.1 and use them as the feature set of a next classifier. Again, the features are transformed before the training phase of the tier 1.2 classifier begins. The transformation involves a standardization and subsequent decorrelation via a PCA transformation, followed by a sorting of the PCA components w.r.t. their separation power between signal and background, measured in the Wasserstein l_1 metric. The feature set before and after decorrelation is shown in Fig. L.1, Fig. L.2, and Fig. L.3, Fig. L.4, respectively. A strong correlation among the features is apparent, pointing towards a redundancy of the classifiers due to an exhaustion of the available information which resembles the findings of Refs. [87, 88]. We account for this degeneration by only using the first and second PCA components in the tier 1.2 classifier. We use a SVM with an RBF kernel for classifying the two-dimensional feature set and optimize the corresponding hyper-parameters C and γ via

³The estimation of the performance of the final classifier is not obtained via cross-validation but on the separated testing set. Hence, no uncertainty due to fluctuation in the cross-validation processes is taken into account for subsequent steps.

⁴In the context of neural networks another commonly used technique is *Dropout*.

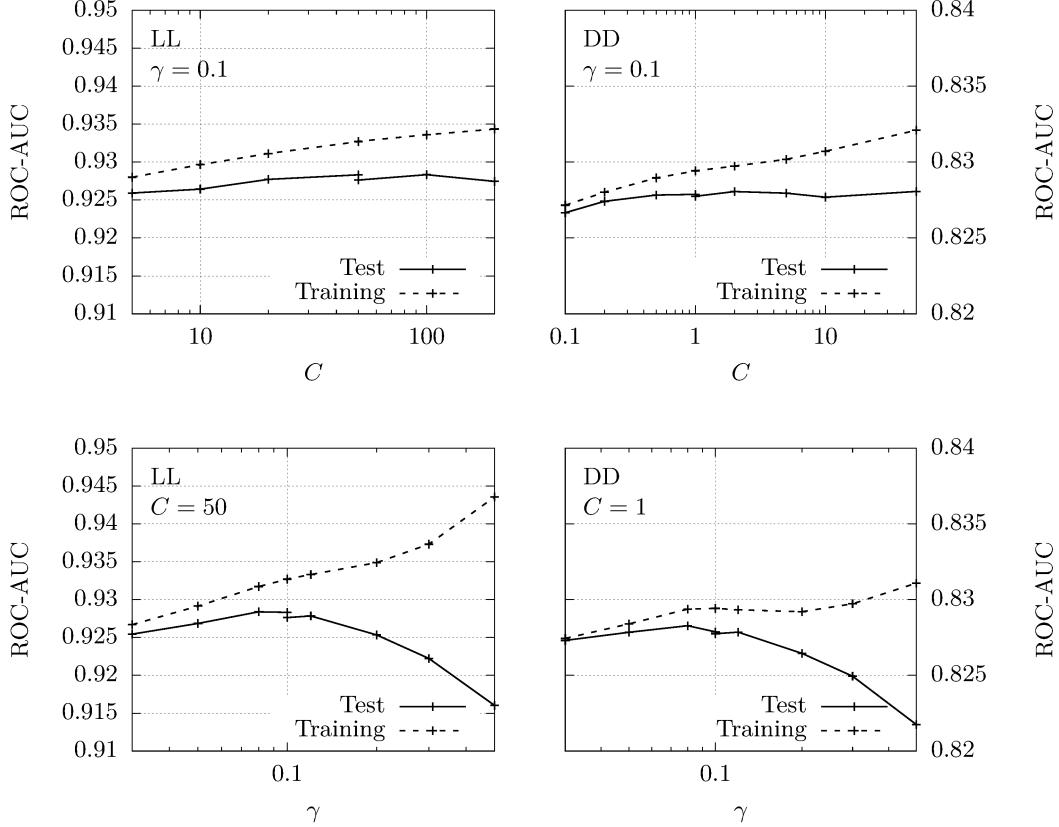


Fig. 5.5. Performance of the SVM (Λ classifier) measured with the ROC-AUC for different values of the hyper-parameter C (top) and γ (bottom) and track types LL (left) and DD (right). The performance is evaluated via a 5-fold cross-validation which explains the minor fluctuation of configurations that are evaluated twice.

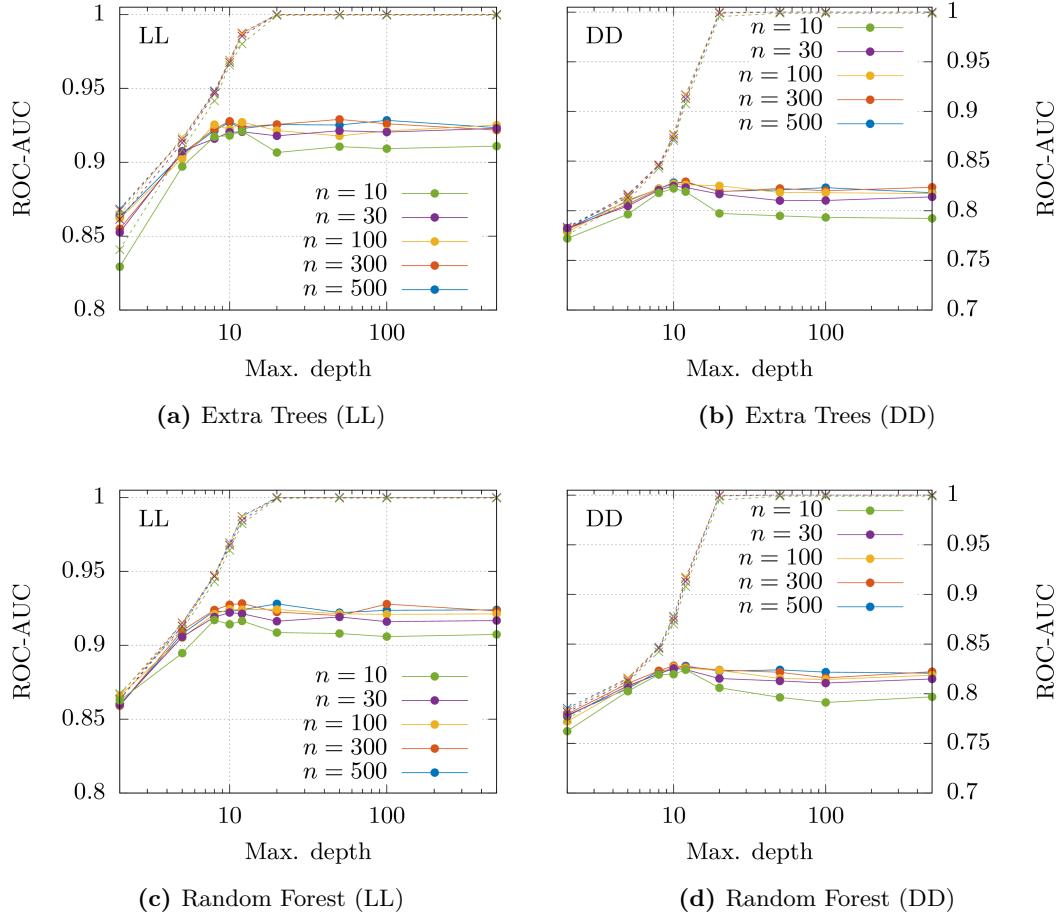


Fig. 5.6. Performance of Extra Trees (top) and a Random Forest (bottom) measured with the ROC-AUC for different values for the maximal allowed tree depth (x -axis) and number of trees (n), evaluated on the testing folds (solid lines) and training folds (dashed lines).

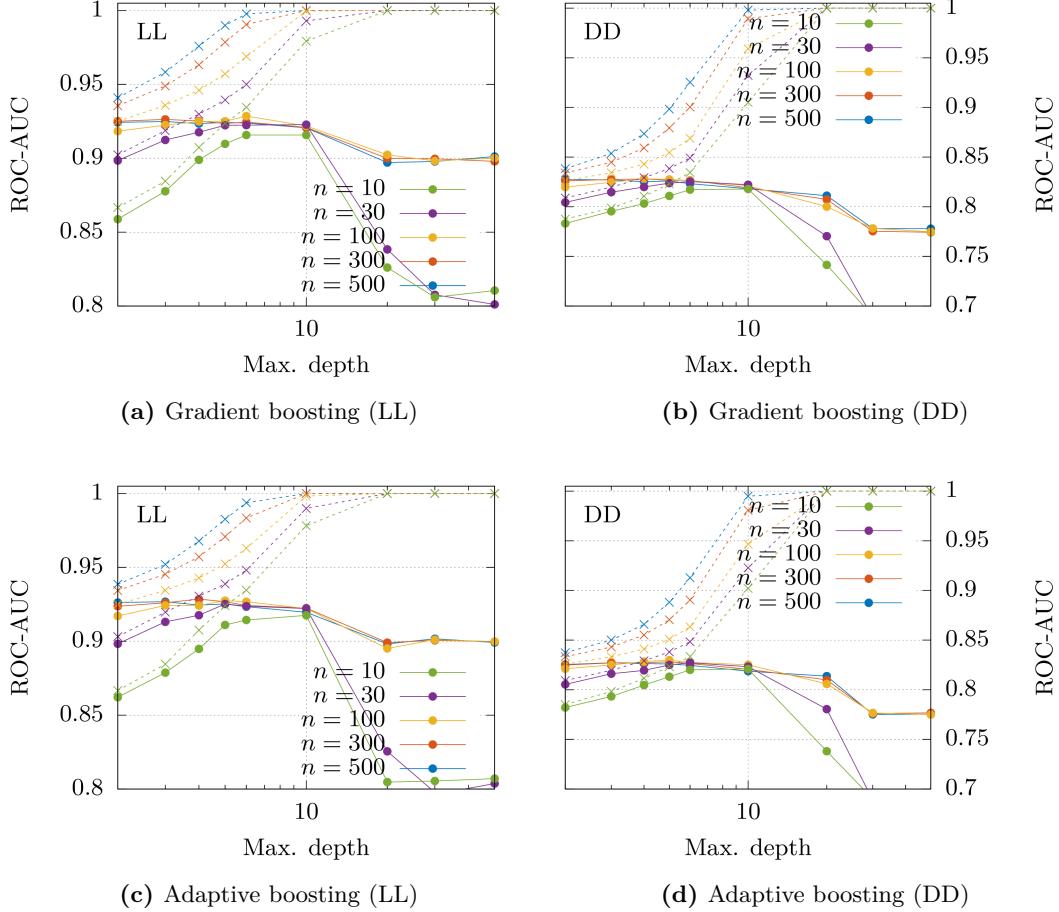


Fig. 5.7. Performance of BDT classifiers (using gradient (top) and adaptive (bottom) boosting) measured with the ROC-AUC for different values for the maximal allowed tree depth (x -axis) and number of trees (n), evaluated on the testing folds (solid lines) and training folds (dashed lines).

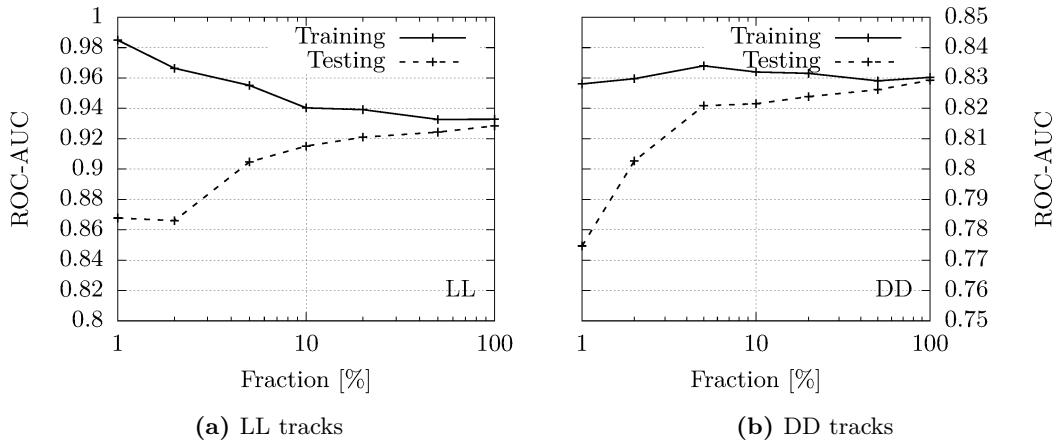


Fig. 5.8. Convergence of the SVMs (\mathcal{A} classifier) given by the ROC-AUC values for different sample sizes where a sample size of 100 % corresponds to the size of the full training set. The ROC-AUC is evaluated via a 5-fold cross-validation scheme on the training (solid line) and testing folds (dashed line).

a 5-fold cross-validation scheme. The results are listed in Tab. 5.7 (referred to as *Stacking*) and shown visually in Fig. 5.9. Additionally, the limited size of the feature set allows for a simple graphical representation of the decision boundary of the optimized classifier which is given in Fig. 5.10.

Evaluation

The performance of all trained classifiers are judged w.r.t. their ROC-AUC values, evaluated on the separated testing set. The corresponding ROC curves are shown in Fig. 5.11 for LL and DD tracks. The training of the stacking classifier (tier 1.2) does not show any improvement over

Tab. 5.8. ROC-AUC values in percent for tier 1.1 and tier 1.2 *A* classifiers (separately trained for LL and DD tracks), evaluated on the training and testing set.

Classifier	Training		Testing	
	LL [%]	DD [%]	LL [%]	DD [%]
SVM	93.25	82.92	92.86	83.25
Extra Tree	93.02	82.69	91.45	82.44
Random Forest	96.50	87.07	92.34	83.38
BDT	95.30	85.16	92.60	83.47
Ada. BDT	94.92	84.86	92.51	83.37
Stacking	95.46	87.57	90.87	80.42

the tier 1.1 classifiers. For the sake of consistency (also with the A_b - D^0 classifier), we therefore use the tier 1 SVM as our final classifier, both for LL and DD tracks. The response, as well as the efficiency for given thresholds are shown graphically in Fig. 5.12. Further, we show the distribution of the marginal and linear correlation distributions of the five-dimensional feature space before and after the PCA transformation in Fig. 5.13, Fig. 5.14 and Fig. 5.15, Fig. 5.16, where we color-coded misclassified events in red (signal events classified as background) and blue (background events classified as signal) when using a threshold of zero.

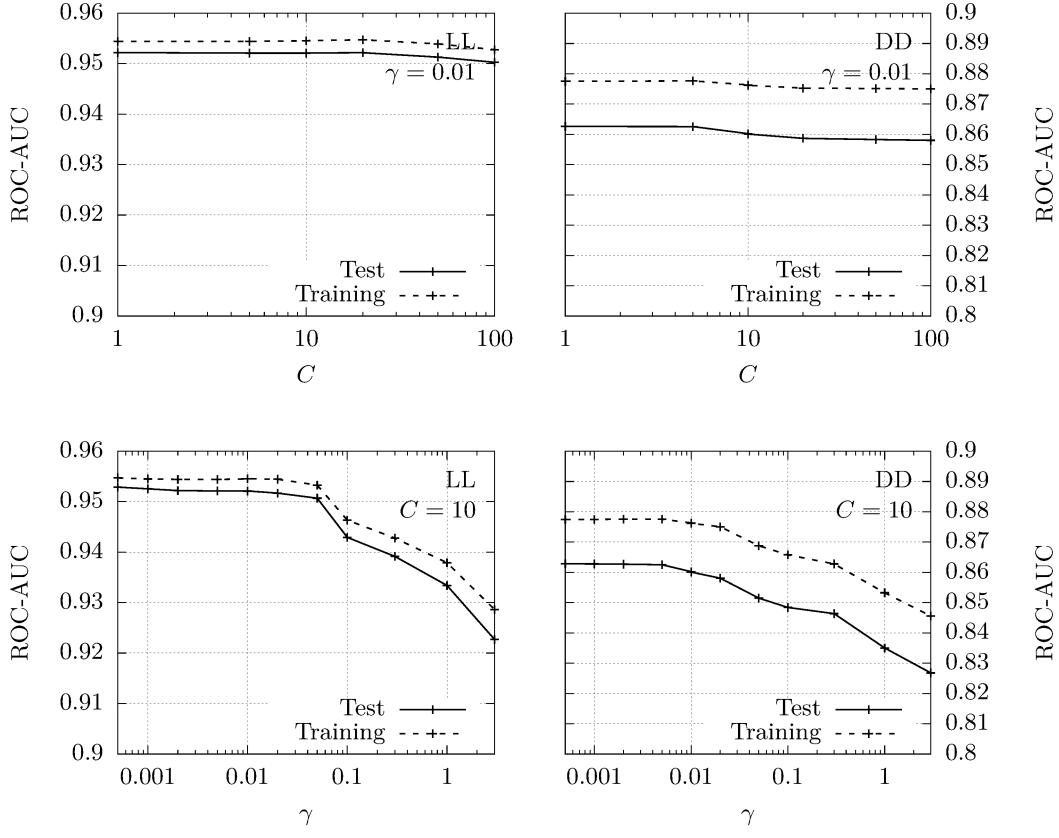


Fig. 5.9. Performance of the stacking classifier (SVM with an RBF kernel) measured with the ROC-AUC for different values of the hyper-parameter C and γ , evaluated via a 5-fold cross-validation scheme.

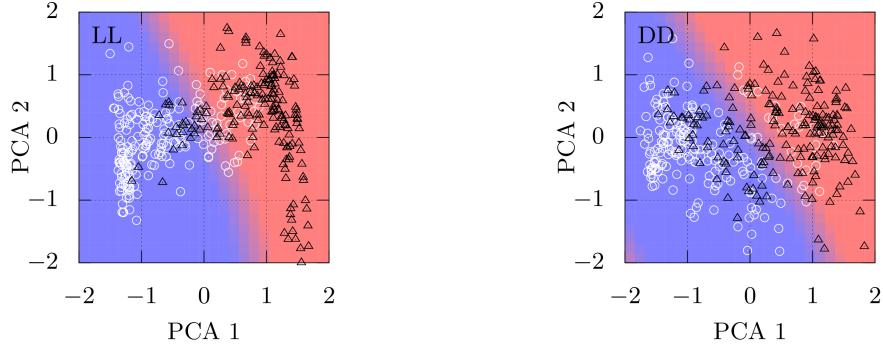


Fig. 5.10. Decision boundary for LL (left) and DD (right) tracks of the stacking classifier (SVM with an RBF kernel) overlaid with 400 instances of the signal and background class (200 each). The signal (background) class is depicted with circles (triangles) and the corresponding classifier response in blue (red).

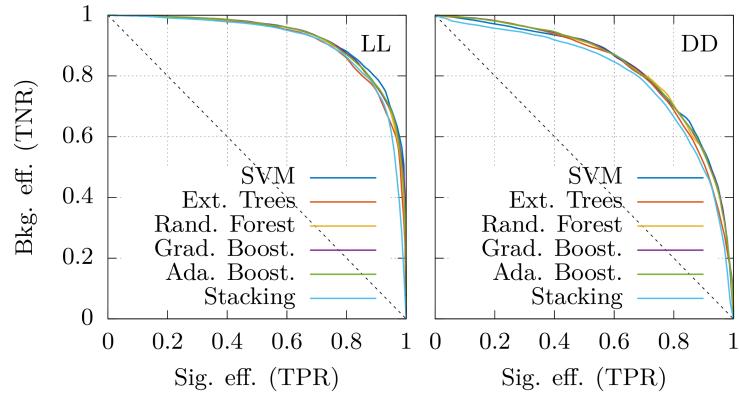


Fig. 5.11. ROC curves of the optimized Λ classifiers, evaluated on the testing set.

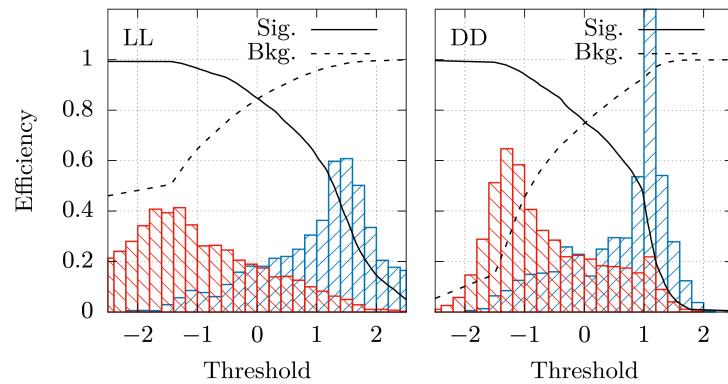


Fig. 5.12. Efficiency (solid and dashed line) when using a given threshold (x -axis) for classifying signal and background, and response function (histogram) of the tier 1 SVM (Λ classifier), trained separately for LL (left) and DD (right) tracks.

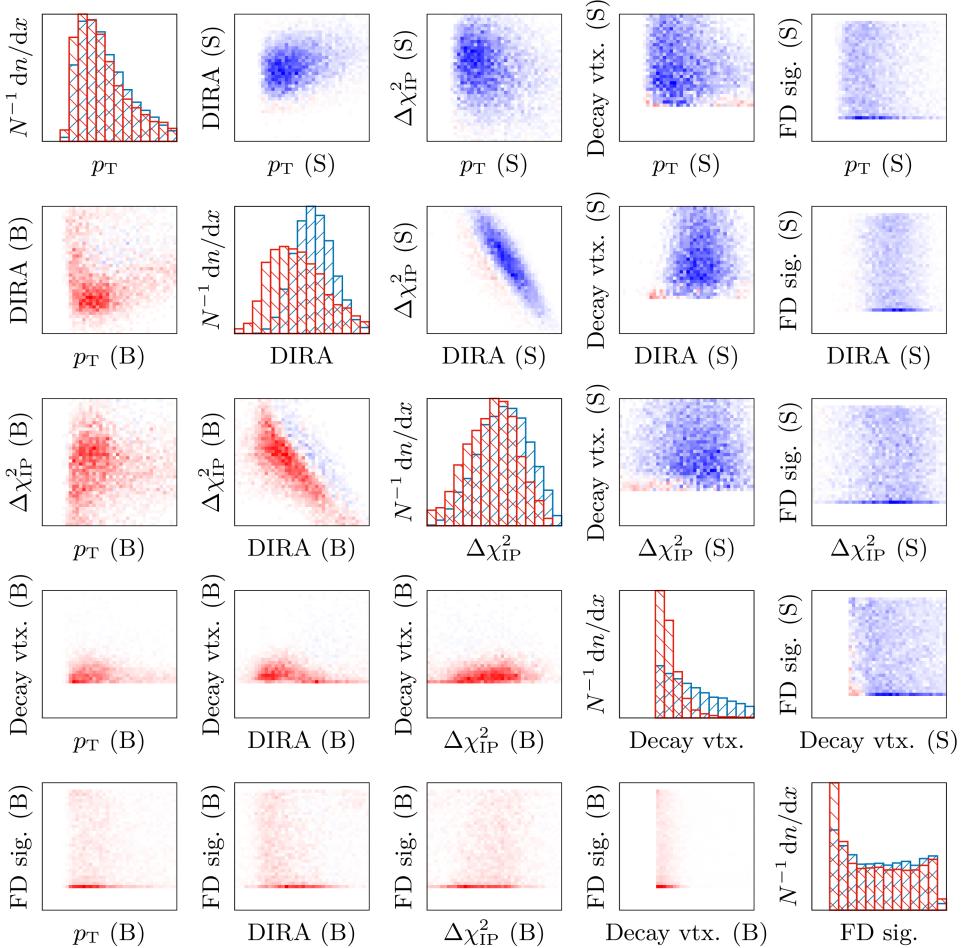


Fig. 5.13. Marginal and linear correlation distributions of the feature space (Λ classifier) before applying PCA transformation for LL tracks. In the correlation distributions misclassified events are color-coded in red (signal events classified as background) and blue (background events classified as signal) when using a threshold of zero.

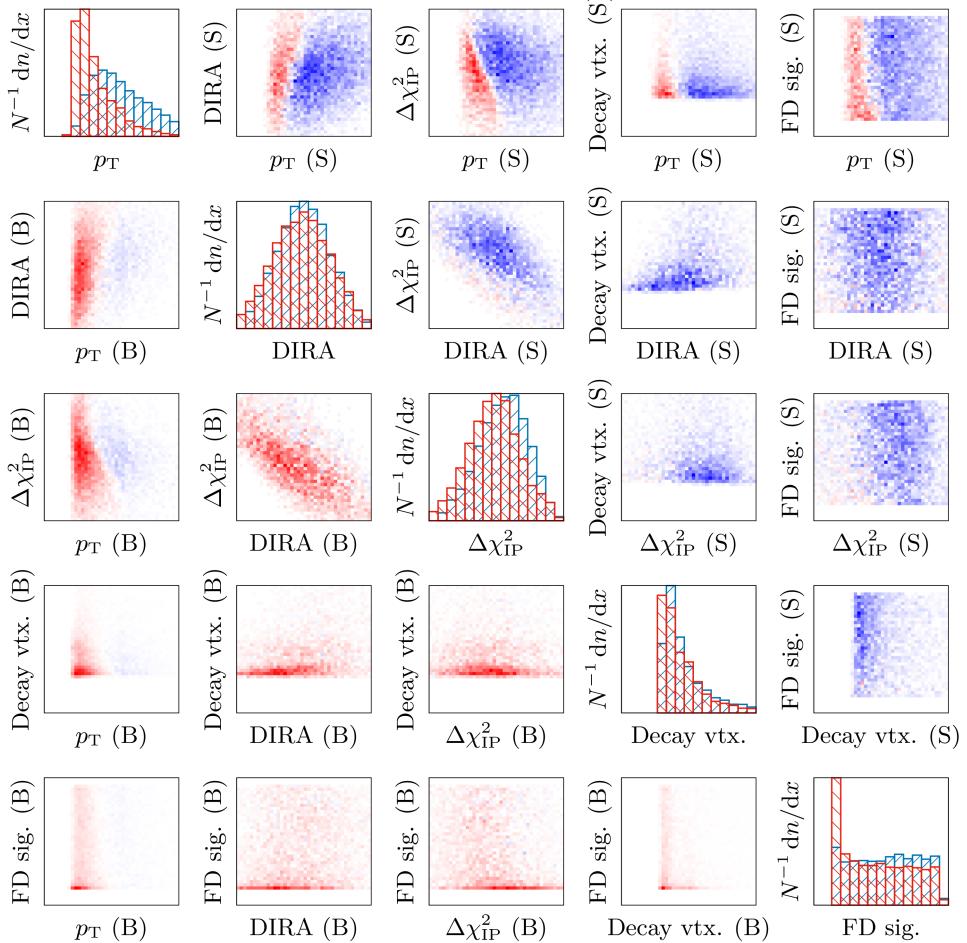


Fig. 5.14. Marginal and linear correlation distributions of the feature space (Λ classifier) before applying PCA transformation for DD tracks. In the correlation distributions misclassified events are color-coded in red (signal events classified as background) and blue (background events classified as signal) when using a threshold of zero.

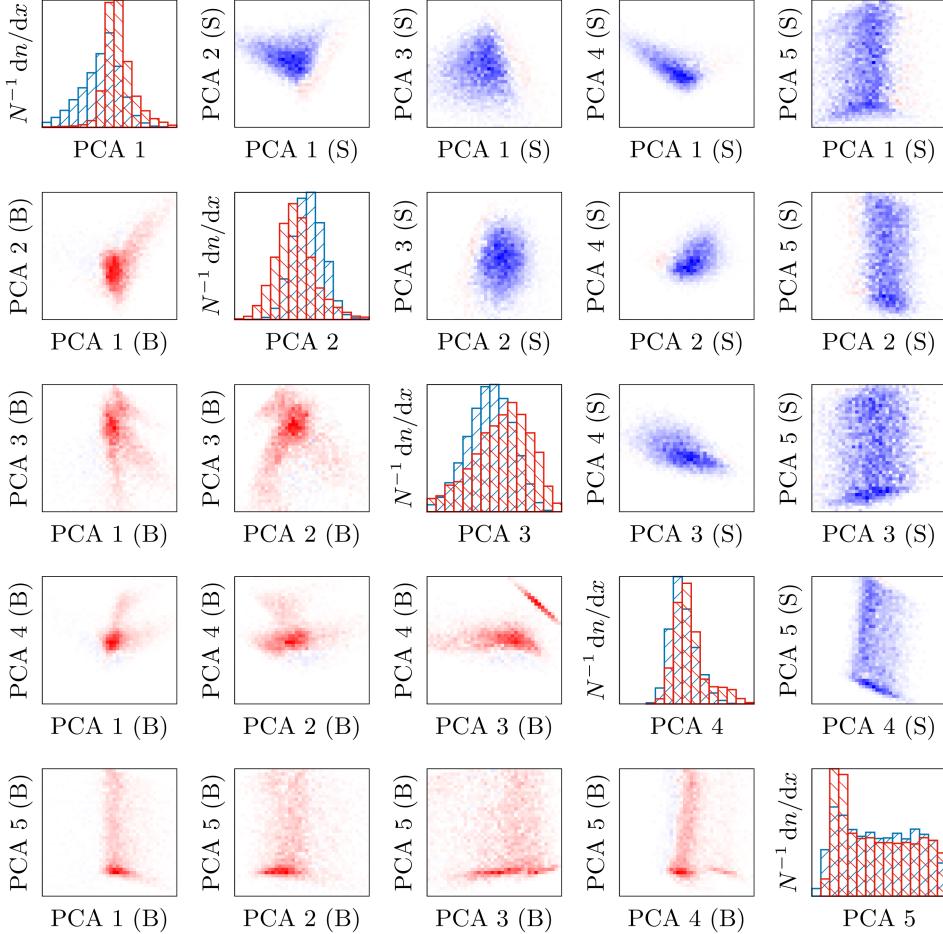


Fig. 5.15. Marginal and linear correlation distributions of the PCA components (Λ classifier) for LL tracks. In the correlation distributions misclassified events are color-coded in red (signal events classified as background) and blue (background events classified as signal) when using a threshold of zero.

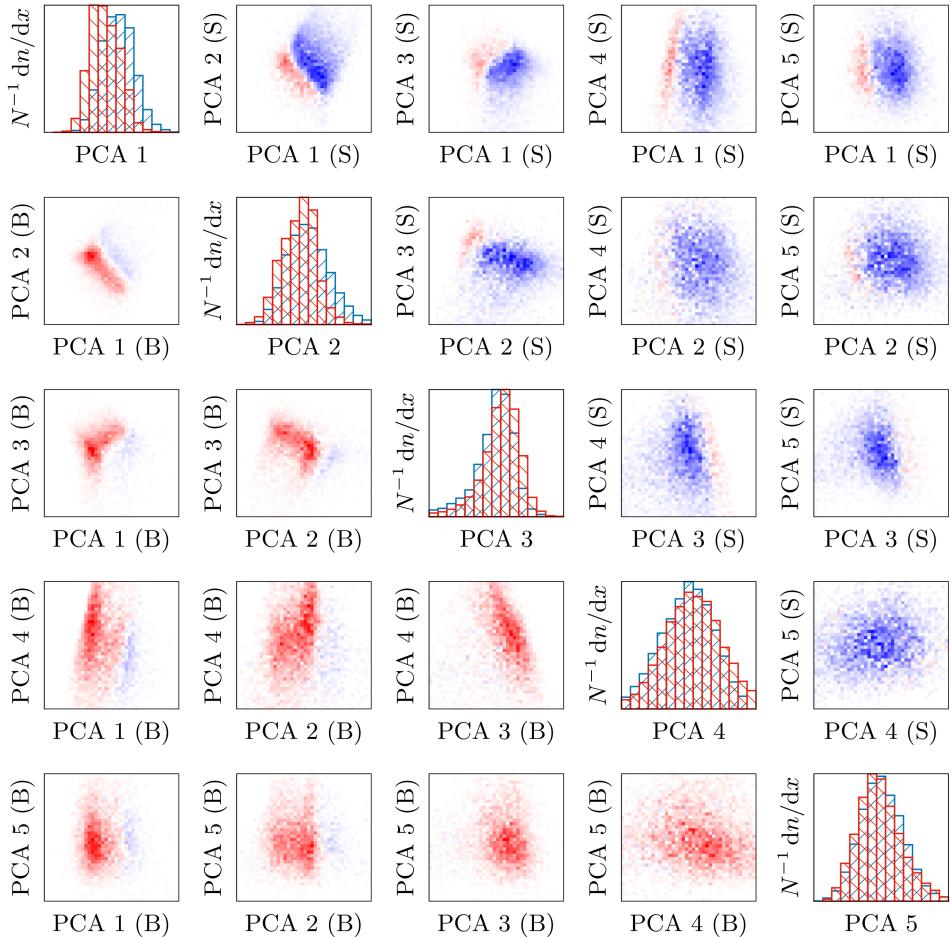


Fig. 5.16. Marginal and linear correlation distributions of the PCA components (Λ classifier) for DD tracks. In the correlation distributions misclassified events are color-coded in red (signal events classified as background) and blue (background events classified as signal) when using a threshold of zero.

5.2.3 The Λ_b - D^0 Classifier

The Λ_b - D^0 classifier is trained and optimized very similar to the Λ classifier that we described in the previous section. Differences are the increased feature set (10 features instead of 5) and the fact that the background sample for the Λ_b - D^0 classifier is the full recorded $\Lambda_b \rightarrow D^0\Lambda$ data set, rather than sidebands as used for training the Λ classifier. The decision for the latter is motivated by the fact that classification algorithms typically tolerate impure labels (*e.g.*, Ref. [89, 90]) and that the expected impurity in the case $m(D^0\Lambda)$ is extremely low. At the same time, this approach minimizes the risk of a selection bias by a non-trivial correlation of the invariant mass $m(D^0\Lambda)$ and the feature set. The invariant mass distribution $m(D^0\Lambda)$ is shown in Fig. L.7 and is apparently dominated by combinatorial background. Separation of the latter and genuine $\Lambda_b \rightarrow D^0\Lambda$ decays is thus the main objective of the Λ_b - D^0 classifier.

The approximation of the performance of the classifiers is limited by the amount of available MC simulated events (signal) and, for the sake of computational ease, the data set is pruned, yielding the same sample sizes as listed in Tab. 5.6, except for a larger training set of LL tracks (pruned to 20 000 instances). We test and optimize the same type of classifiers as we did in the previous section and thereby find, that the SVM performs best and that a stacking classifier does not help to improve the overall performance. For the sake of brevity we thus suppress the separation into tier 1.1 and tier 1.2 and only list the results of tier 1.1 in the following. The optimized hyper-parameter of the classifiers are listed in Tab. 5.7.

A key feature of a PCA analysis, especially when paired with ordering w.r.t. the separation power (as proposed in Appx. I) is the possibility to reduce the feature set and thus evade the *curse of dimensionality*⁵ up to a certain degree. In Fig. 5.17 we show the ROC-AUC value (evaluated on the test set) for the different classifiers, when only the first $n = 2$ to 10 PCA components are used for the training. Even though the beginning of saturation is visible after $n \gtrsim 7$, the performance gain still justifies the increasing of the dimensionality of the feature space. Hence, we use the full available feature set and use the PCA transformation to decorrelate the feature set (without reduction) in order to increase the performance of the decision tree based classifiers.

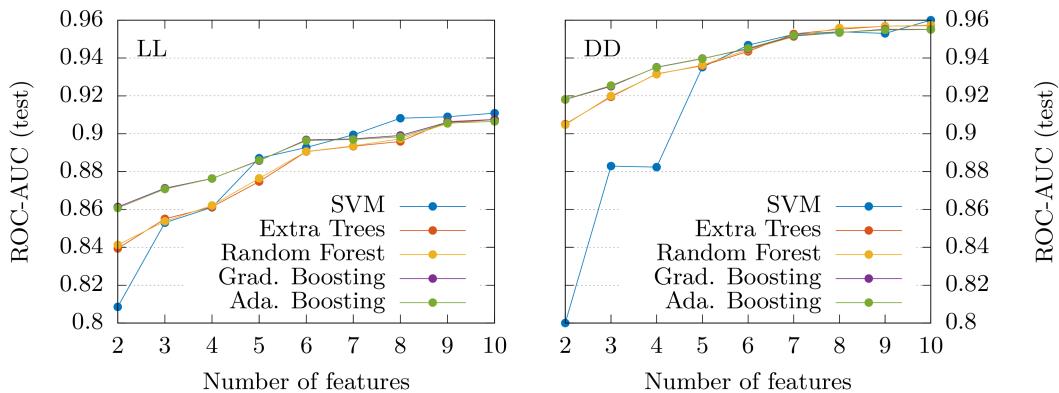


Fig. 5.17. ROC-AUC values (evaluated on the test set) for the different Λ_b - D^0 classifiers, when only the first $n = 2$ to 10 PCA components (ordered w.r.t. the separation power) are used for the training.

⁵In order to train learner effectively and avoid overfitting, the number of training instances should increase with the number of features. The *Curse of dimensionality* [91] describes the highly non-linear relation between both which can lead to a reduced performance if the amount of the former is limited. A common mitigation is to reduce the number of features and only concentrate on the most important ones, *e.g.*, via PCA transformation. Other and more complex dimensionality reduction schemes are Kernel PCA [92] or LLE [93] (which both come with an additional set of hyper-parameters that have to be learned).

The output, as well as the correlation of the classifiers are shown in Fig. 5.18 and Fig. 5.19. The corresponding ROC-AUC values are listed in Tab. 5.9. By not only evaluating the ROC-AUC values on the testing set of corresponding track type, but also on the other set (as listed in Tab. 5.9), it is possible to verify the necessity of splitting the Λ_b - D^0 classifiers w.r.t. the track type of Λ (candidate) daughters. An impact of the latter cannot be assumed a priori, but seemingly influences the performance of the classifiers (for instance via resolution dependent correlations) such that it makes sense (a posteriori) to split the available data set and train into two, rather than one, specialized classifiers.

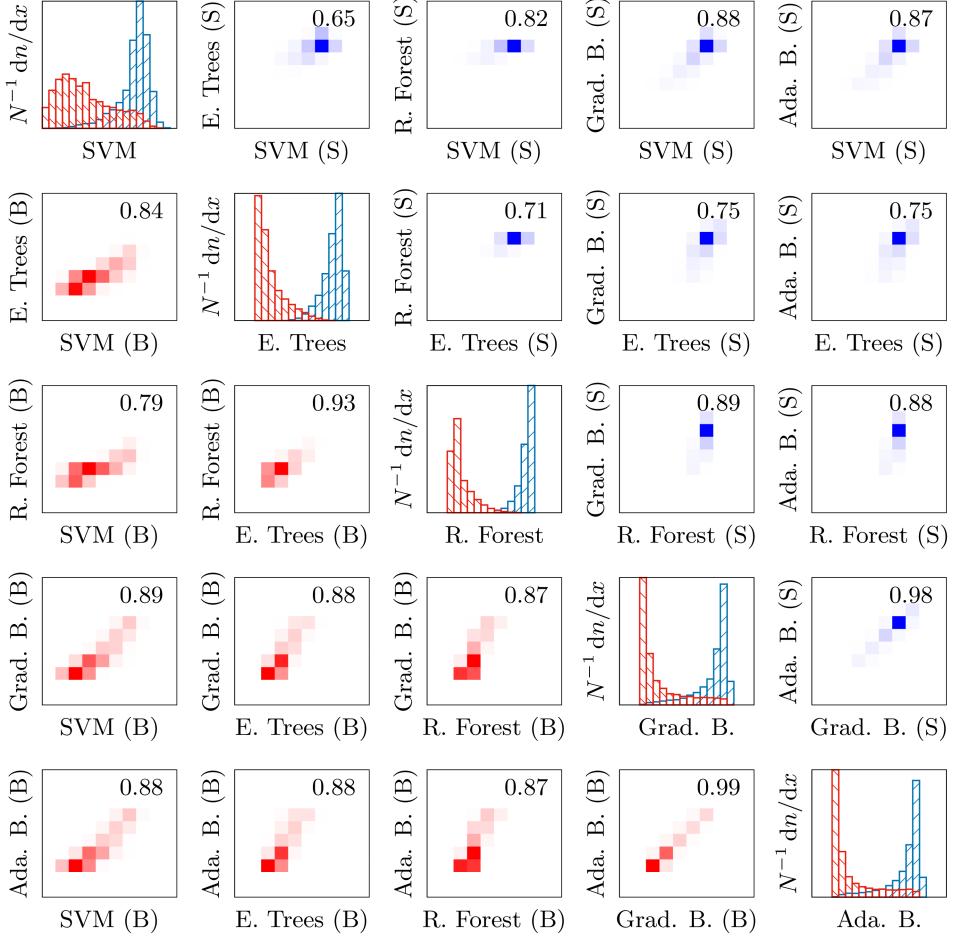


Fig. 5.18. Distributions of the responses and correlations of the Λ_b - D^0 classifiers for LL tracks, separated for signal (S) and background (B). On top of the correlation distributions, we show the PCC as a measure for the linear correlation.

Resting on the ROC-AUC values of the classifiers, we again chose the SVM as the Λ_b - D^0 classifier. The respective ROC curves for LL and DD tracks, as well as the response functions are shown in Fig. L.5 and Fig. 5.20, respectively. Similar to Sec. 5.2.2, we investigate the convergence of the SVM and show the result in Fig. L.6.

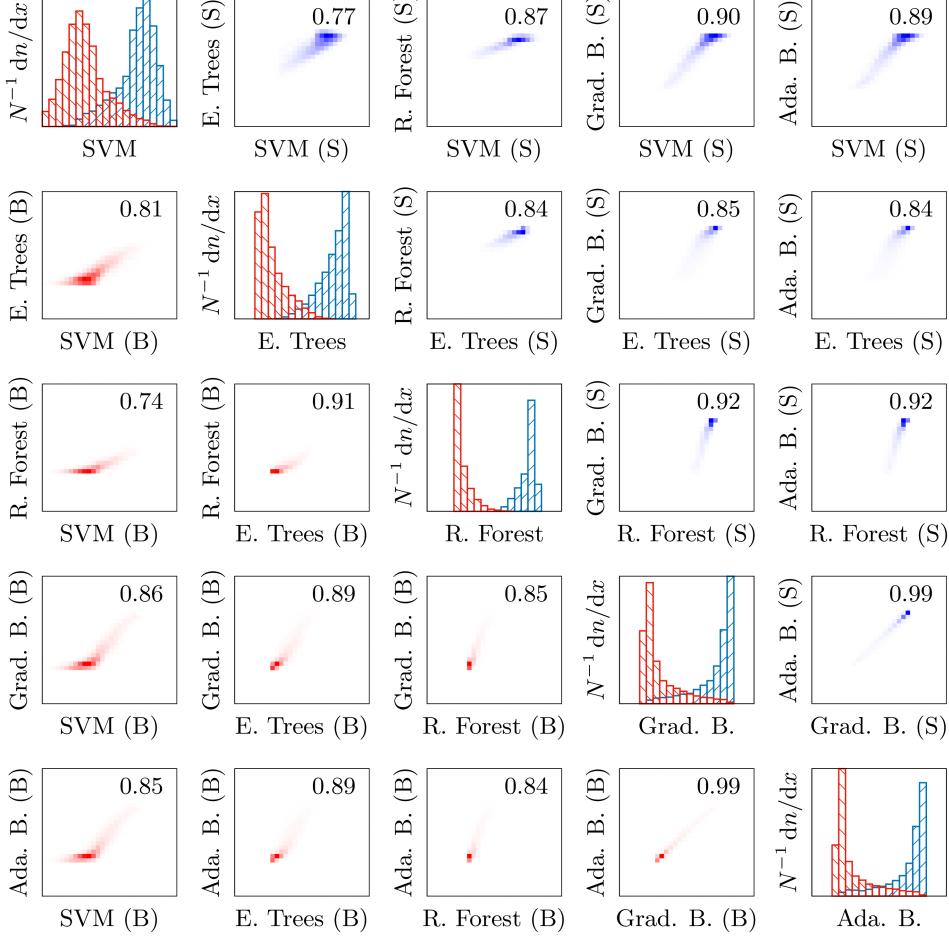


Fig. 5.19. Distributions of the responses and correlations of the Λ_b - D^0 classifiers for DD tracks, separated for signal (S) and background (B). On top of the correlation distributions, we show the PCC as a measure for the linear correlation.

Tab. 5.9. ROC-AUC values in percent for tier 1.1 and tier 1.2 Λ_b - D^0 classifiers (separately trained for LL and DD tracks). ROC-AUC values are evaluated on the testing sets of both track types, *e.g.*, *LL trained* and *DD tested* refers to the ROC-AUC value of a classifier that was trained with LL tracks, but evaluated on the testing set with DD track candidates.

Classifier	LL trained		DD trained	
	LL tested [%]	DD tested [%]	LL tested [%]	DD tested [%]
SVM	91.09	89.34	83.71	94.51
Extra Tree	90.87	79.05	84.16	94.00
Random Forest	90.68	82.29	83.15	94.17
BDT	91.33	82.93	84.19	94.44
Ada. BDT	91.33	82.93	84.19	94.44
Stacking	91.32	82.25	84.12	94.43

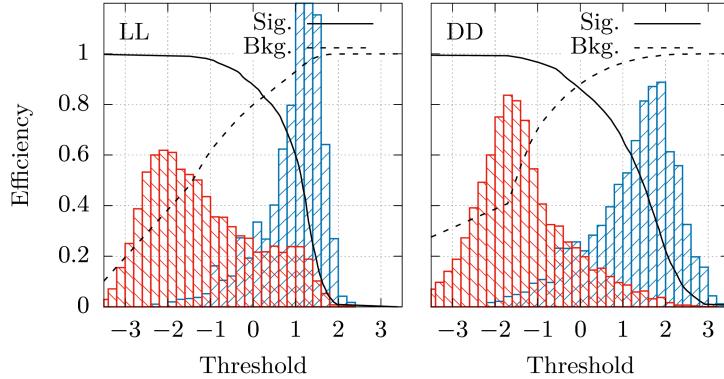


Fig. 5.20. Efficiency (solid and dashed line) when using a given threshold (x -axis) for classifying signal and background, and response function (histogram) of the SVM (Λ_b - D^0 classifier), trained separately for LL (left) and DD (right) tracks.

5.2.4 The Tier 2 Classifier

After having trained optimized Λ and Λ_b - D^0 classifiers, we fuse the normalized output of these classifiers with the DTF probability, as well as the output of PID providing neural networks `ProbNNp(p)` and `ProbNNk(K)`, and use these five variables as the feature set of the Tier 2 classifier. The (marginal) distribution, as well as the correlations of the feature set are shown in Fig. 5.21 (LL tracks) and Fig. 5.22 (DD tracks).

The Tier 2 classifier is intentionally kept rudimentary and applies only a single threshold requirement (rectangular cut) to each feature. By design, each feature is expected to have values close to one for signal-like events, and values close to zero else. The threshold requirement therefore reads as a required lower bound t_i for each feature f_i , *i.e.*,

$$\text{signal} : \Leftrightarrow (f_1 > t_1) \wedge (f_2 > t_2) \wedge \dots \wedge (f_5 > t_5).$$

Doing so makes the Tier 2 classifier unable to learn any higher level features beyond the marginal distributions (such as correlation) of the features. In particular this renders the need to recalibrate the `ProbNNp` and `ProbNNk` correlations in MC simulated events unnecessary which is an error prone business and requires decent statistics. In Appx. M we find that selection requirements for the last three features introduce deviations between recorded and simulated data for thresholds close to one, whereas the fidelity of the first two features (Λ and Λ_b - D^0 classifier) is sufficient. We will therefore not aggressively optimize thresholds requirements with the Tier 2 classifier, but choose loose selection requirements and tighten them in a subsequent step using only the output of the Λ and Λ_b - D^0 classifiers.

As a cross-check we compare the results with the decisions of a canonically trained binary decision tree (Gini criterion and max. depth of 5). The ROCs of all classifiers (*i.e.*, feature set and both Tier 2 variants) are shown in Fig. 5.23 and the respective ROC-AUC values are listed in Tab. 5.10.

The Rectangular Cut Classifier

For uncorrelated features, the relations between TPR, FPR, FNR and TNR (*cf.* Appx. G for our definitions of *true*, *false*, *positive* and *negative*) in the presence of rectangular selection

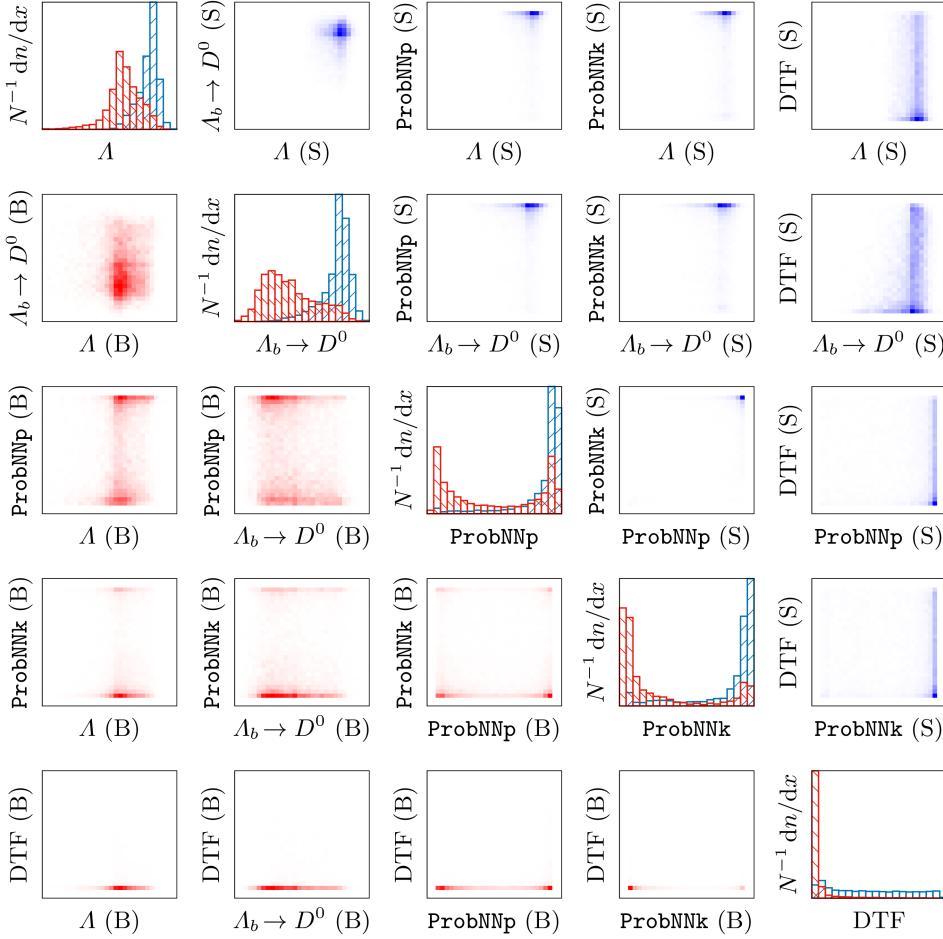


Fig. 5.21. Marginal distributions and correlations of the feature set of the Tier 2 classifier for LL tracks. Instances of the *signal* (S) and the *background* (B) class are color-coded in blue and red, respectively.

Tab. 5.10. ROC-AUC values in percent for Tier 2 Λ classifiers (separately trained for LL and DD tracks), evaluated on the training and testing set. For numerical reasons, the ROC curve of the rectangular cut classifier (Rec. Clf.) is only evaluated for $TPR < 90\%$ (*cf.* Fig. 5.23) which explains the inhibited ROC-AUC values for this classifier.

Classifier	Training		Testing	
	LL [%]	DD [%]	LL [%]	DD [%]
Λ Clf.	93.46	82.93	89.48	82.09
$\Lambda_b \rightarrow D^0$ Clf.	92.69	94.90	90.76	94.62
DTF prob.	94.32	91.60	93.72	91.86
ProbNN(p)	79.82	74.95	78.72	74.31
ProbNN(K)	86.32	84.82	85.96	84.61
Decision Tree	97.74	96.96	96.77	96.78
Rec. Clf.	94.98	93.67	94.86	93.06

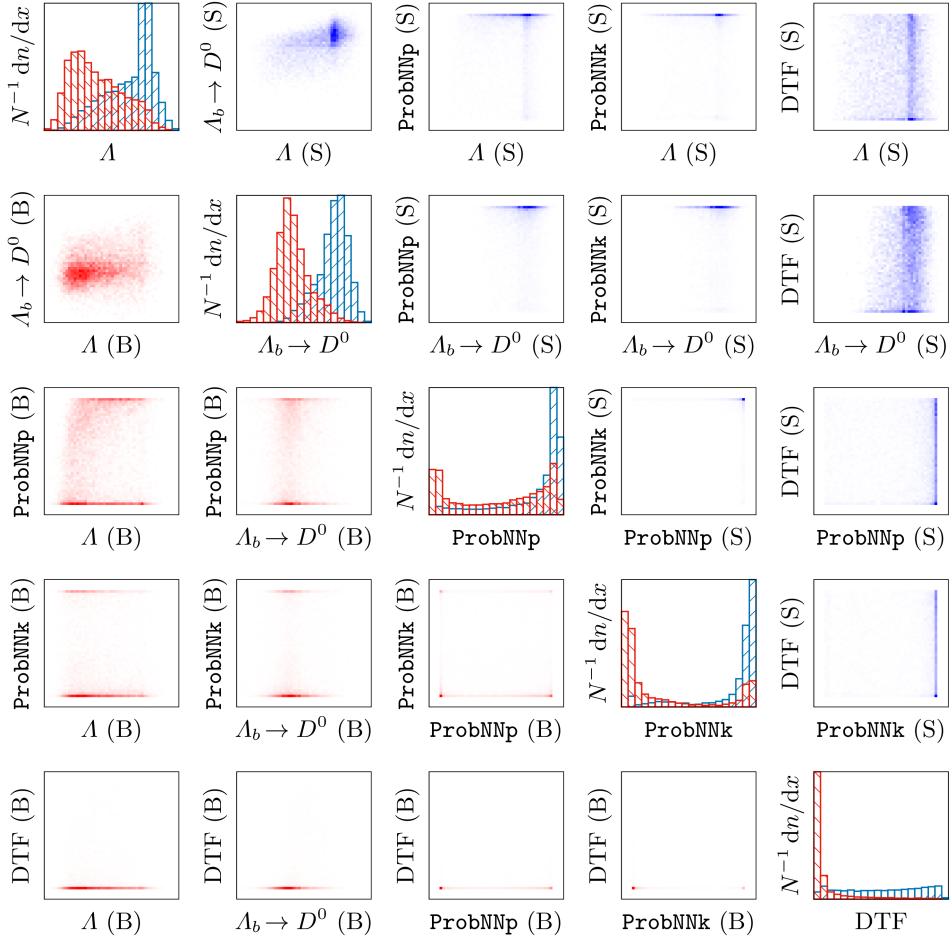


Fig. 5.22. Marginal distributions and correlations of the feature set of the Tier 2 classifier for DD tracks. Instances of the *signal* (S) and the *background* (B) class are color-coded in blue and red, respectively.

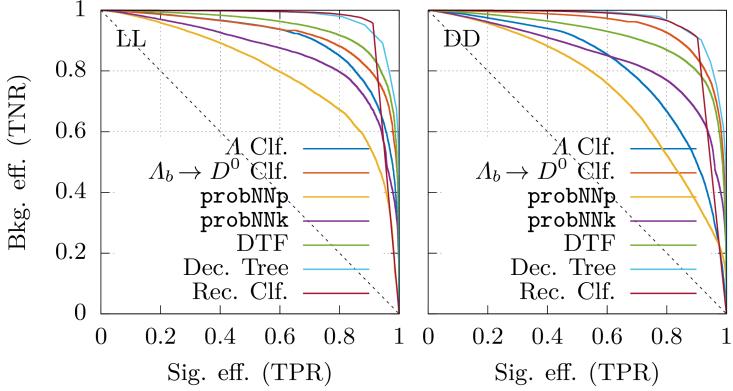


Fig. 5.23. ROC curves of the feature set used to train Tier 2, *i.e.*, the decision tree (Dec. Tree) and the rectangular cut classifier (Rec. Clf.), as well the respective ROC curves of the Tier 2 classifiers themselves. Note that we evaluated the rectangular classifier only for $\text{TPR} < 90\%$ due to numerical instability.

requirements read

$$\begin{aligned} \text{TPR} &= \text{TPR}_1 \times \text{TPR}_2 \times \dots, \\ \text{FPR} &= \text{FPR}_1 \times \text{FPR}_2 \times \dots, \\ \text{FNR} &= 1 - \text{TPR} = 1 - \text{FNR}_1 \times \text{FNR}_2 \times \dots, \\ \text{TNR} &= 1 - \text{FPR} = 1 - (1 - \text{TNR}_1) \times (1 - \text{TNR}_2) \times \dots, \end{aligned}$$

and in particular $\text{TNR} \neq \prod_i \text{TNR}_i$. In our case we are striving to minimize the overall FPR for a given target TPR \hat{p}_{TPR} by requiring thresholds \vec{t} :

$$\vec{\nabla}_{\vec{t}, \lambda} \mathcal{L} \stackrel{!}{=} 0 \quad \text{with} \quad \mathcal{L}(\vec{t}, \lambda) = \text{FPR}(\vec{t}) - \lambda [\text{TPR}(\vec{t}) - \hat{p}_{\text{TPR}}]. \quad (5.1)$$

In case of uncorrelated features

$$\text{TPR}_i = (\hat{p}_{\text{TPR}})^{1/n} \quad (5.2)$$

solves this optimization problem. However, a certain degree of correlation is expected for the feature set of the Tier 2 classifier and we therefore have to solve Eq. (5.1) numerically. The Lagrange multiplier λ in Eq. (5.1) is unbound which makes numerical approaches difficult. Hence, we rewrite the problem as

$$\vec{\nabla}_{\vec{t}} \mathcal{L} \stackrel{!}{=} 0 \quad \text{with} \quad \mathcal{L}(\vec{t}) = \text{FPR}(\vec{t}) + \underbrace{\log(|\text{TPR}(\vec{t}) - \hat{p}_{\text{TPR}}| + \epsilon)}_{\text{penalty}}. \quad (5.3)$$

where we replaced the latter part of the Lagrangian function with a penalty term. Due to the discontinuities of the derivatives of \mathcal{L} , gradient descent techniques will not work sufficiently. We therefore use simulated annealing [94] with initial values taken from Eq. (5.2) to find thresholds \vec{t} for a given TPR \hat{p}_{TPR} while keeping the overall FPR minimal. This method becomes numerically unstable for $\hat{p}_{\text{TPR}} < 0.1$ and $\hat{p}_{\text{TPR}} > 0.9$, hence we limit the TPR scanning to $0.1 \leq \hat{p}_{\text{TPR}} \leq 0.9$. The optimized thresholds for a given TPR are shown for each feature in Fig. 5.24a. In Fig. 5.24b we show the TNR_i of feature i as a function of the overall TPR.

As the objective of the loose optimization of the Tier 2 classifier we choose a FoM,

$$\text{FoM}(\text{TPR}) = \frac{\text{TPR}}{\sqrt{\text{FPR}}} \sim \frac{f_{\text{sig}}}{\sqrt{f_{\text{bkg}}}}, \quad (5.4)$$

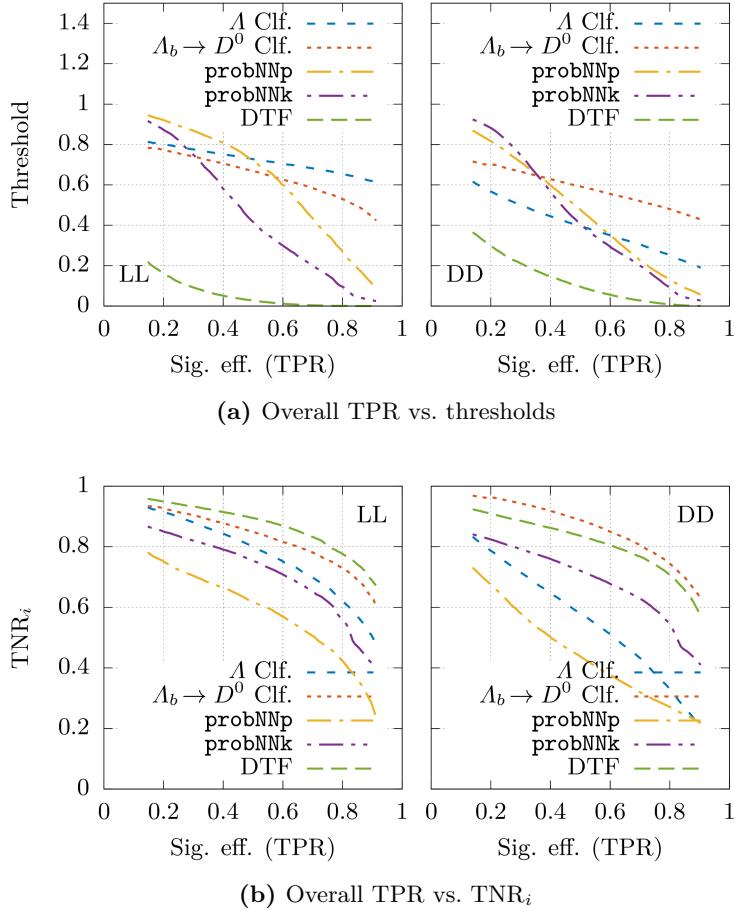


Fig. 5.24. Thresholds (top) and TNR_i of each feature i of the Tier 2 classifier as function of the overall TPR. Note that the latter is not equivalent to the ROC curve (as shown Fig. 5.23) since we use the overall TPR as the x -axis, as opposed to the TPR_i of the respective feature i .

where f_{sig} and f_{bkg} are the signal efficiency and the fraction of background events in the signal channel, respectively, which makes Eq. (5.4) an approximation of the signal efficiency in the limit of large background. In particular this allows an evaluation of the FoM without extracting a signal yield from recorded data,⁶ which would require a fit of the signal component $\Lambda_b \rightarrow D^0 \Lambda$ and does not befit an unbiased search for this decay. The maximum of Eq. (5.4) thus is not necessarily the true maximum of the significance, but allows for an unbiased optimization.

In Fig. 5.25 we show the inverted ROC curve, as well as the FoM as defined in Eq. (5.4) as a function of the overall TPR. The thresholds are calculated w.r.t. Eq. (5.3) and fixed for the subsequent steps at values that correspond to a TPR of 60 % (40 %) for LL (DD) tracks in Fig. 5.25. The subsequent (tight) maximization of the FoM w.r.t. the responses of the Λ and Λ_b - D^0 classifiers is shown in Fig. 5.26 and the final thresholds are listed in Tab. 5.11. In Appx. L we show the invariant mass distributions of the D^0 and Λ candidates from recorded data and truth-matched simulated events after applying the thresholds. After refining the thresholds of

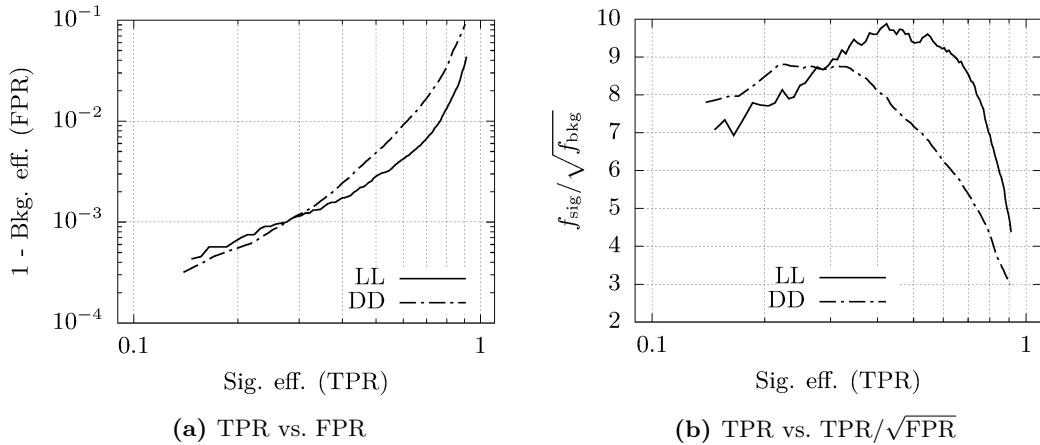


Fig. 5.25. FPR and signal significance approximation of the Tier 2 classifier for track types LL and DD as a function of the overall TPR. The corresponding thresholds for each feature are optimized w.r.t. Eq. (5.3) and refined in an subsequent step (not shown).

the Λ and Λ_b - D^0 classifiers, the overall TPR (test set) is 55.3(16) % and 18.0(12) % for LL and DD tracks, respectively.

From the ROC curves shown in Fig. 5.23 we see that the rectangular classifier performs similar to the decision tree (and even outperforms it slightly for LL tracks and $\text{TPR} < 90\%$). The depth of the decision tree is limited to 5 which is still enough to pick up correlations of the five dimensional feature space. The fact that this additional information does not help to increase the overall classification power, justifies the outlined usage of a rectangular classifiers and reduces the source of additional uncertainties.

⁶See Ref. [95] for a comprehensive discussion of different FoMs in the context of high energy physics.

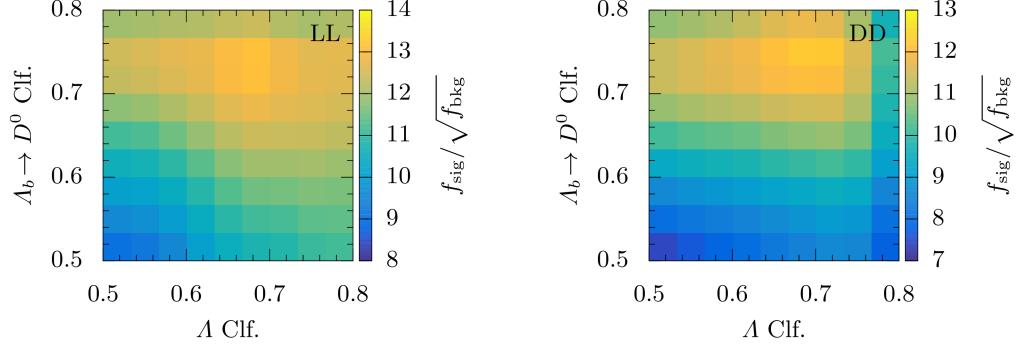


Fig. 5.26. Approximation of the signal efficiency as a function of the responses of the Λ and Λ_b - D^0 classifiers. (The other three feature thresholds are kept fix.)

Tab. 5.11. Optimized thresholds (required lower bounds) for each feature of the Tier 2 classifier and the overall TPR for LL and DD tracks. The TPR is evaluated on the training and test set, and is given w.r.t. the events that are left after the loose selection.

	LL	DD
Λ Clf.	0.67	0.70
Λ_b - D^0 Clf.	0.73	0.77
DTF prob.	0.012	0.15
ProbNNp(p)	0.60	0.60
ProbNNk(K)	0.30	0.57
TPR (training)	55.8(5) %	18.33(23) %
TPR (test)	55.3(16) %	18.0(12) %

Chapter 6

The Normalization Channel $\Lambda_b \rightarrow D^0 p\pi^-$

A normalization channel should comprise two key characteristics: First, the decay topology should be as close as possible to the primary decay channel such that common uncertainties and biases cancel. Secondly, the normalization channel should be well established, clean to extract and contribute to the error budget as little as possible.

Obvious candidates for the present analysis are $\Lambda_b \rightarrow D^0 p\pi^-$, $\Lambda_b \rightarrow \Lambda K^+\pi^-$, $\Lambda_b \rightarrow J/\psi \Lambda$ and $\bar{B}_{(s)}^0 \rightarrow \bar{D}^0 K_s$. Despite its large branching fraction and clean reconstruction, the $J/\psi \Lambda$ mode, in particular if the J/ψ is reconstructed in its dimuon mode, will differ strongly in its detector response and trigger signature and thus requires a careful study of the corresponding simulation fidelity of both modes and therefore is not an ideal candidate. The branching fractions of the other modes, as reported by the PDG, are

$$\begin{aligned}\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-) &= (6.3 \pm 0.7) \times 10^{-4}, \\ \mathcal{B}(\Lambda_b \rightarrow \Lambda K^+\pi^-) &= (5.7 \pm 1.3) \times 10^{-6}, \\ \mathcal{B}(B^0 \rightarrow \bar{D}^0 K^0) &= (5.2 \pm 0.7) \times 10^{-5}, \\ \mathcal{B}(B_s^0 \rightarrow \bar{D}^0 \bar{K}^0) &= (4.3 \pm 0.9) \times 10^{-4},\end{aligned}$$

which do not include the b -hadron production fractions which themselves depend on the kinematics of the b quark [96, 97]. Further, only roughly half of the K^0 mesons decay as K_s and within a similar detector acceptance as Λ baryons.¹ The other part will most likely escape undetected. Taking both effects into account qualifies $\Lambda_b \rightarrow D^0 p\pi^-$ as the most efficient candidate.

The final states of the charmless three-body decay approximate the signature of $\Lambda_b \rightarrow D^0 \Lambda$ best, but this advantage is compensated by its small branching fraction and the noisy background (*cf.* Ref. [39]). We note that this valuation will likely change in the future and future experiments might rank charmless three-body decays as their most suitable normalization candidate if provided with a sufficiently large set of Λ_b decays.

The listed b -meson decays also have a long living V^0 particle in their decay chain. During analysis, $K_s \rightarrow \pi^+\pi^-$ decays have to be grouped w.r.t. their track types LL and DD, similar to $\Lambda \rightarrow p\pi^-$. On the one hand, this leverages the access to track type specific properties which could cancel partially in the branching ratio, on the other hand, splitting the data sample hampers the analysis and, even though K_s and Λ are both V^0 particles, the final states $\pi^+\pi^-$ and $p\pi^-$ can cause different signatures in the LHCb detector which make $D^0 p\pi^-$ the better proxy for $D^0 \Lambda$, especially for LL tracks. At the same time, the fidelity of MC simulated Λ_b decays is known for being imperfect, especially in terms of kinematic distributions. Using Λ_b decays in the primary decay, as well as the normalization mode minimizes this uncertainty source in the branching fraction.

Considering the listed arguments, we use $\Lambda_b \rightarrow D^0 p\pi^-$ as the normalization mode. We will briefly discuss the decay in Sec. 6.1 and will then outline the selection strategy. In Sec. 6.2 we will then extract signal yields which we will use later for determining the branching fraction.

¹Neglecting small CP violating corrections and assuming similar lifetimes of the K_S state and the Λ baryon.

6.1 The Decay $\Lambda_b \rightarrow D^0 p \pi^-$

The decay $\Lambda_b \rightarrow D^0 p \pi^-$ is a high statistics channel and leveraged the discovery of the Λ_b baryon in 1981 [19]. The final states are the same as in $\Lambda_b \rightarrow D^0 \Lambda$ but due to the different quark transitions, it is much less susceptible for CP violation.² The large branching fraction gives a good signal to background ratio after applying simple rectangular selection requirements. Rather than optimizing these selection requirements for $\Lambda_b \rightarrow D^0 p \pi^-$, the opulent signal to background ratio allows the use of a subset of the optimized requirements for the primary decay $\Lambda_b \rightarrow D^0 \Lambda$ with some minor corrections. These selections are split into a preselection and a loose selection part which are outlined in the following sections.

6.1.1 Preselection

We use the full recorded data set of Run 2 and the stripping versions listed in Tab. 5.1. Despite their different naming, the selection requirements vary slightly. These differences are compensated by adopting the tightest selection requirement among conflicting values if necessary.

Tab. 6.1. Stripping and Reco versions used for reconstructing $\Lambda_b \rightarrow D^0 p \pi^-$.

Year	Stripping	Reco
2015	24r1	15a
2016	28r1	16
2017	29r2	17
2018	34	18

All selection criteria of the preselection step are listed in Tab. B.3 where we use the same nomenclature that we introduced in Sec. 4.1.1. Additionally, at least one HLT trigger flag among `Hlt2.*IncPhi.*Decision` or `Hlt2Topo.*Decision` is required, *cf.* Refs. [64, 79] for more detailed information.

6.1.2 Loose Selection

Similar to the previous analyses of $\Lambda_b \rightarrow J/\psi \Lambda$ and $\Lambda_b \rightarrow D^0 \Lambda$ decays, we select only events corresponding to the best PV hypothesis for the following steps. All selection criteria of the loose selection are shown in Tab. 6.2 and are grouped into five categories. Again, categories 1 to 4 are the same that we used and described previously in Sec. 4.1.2 and Sec. 5.1.2. The purpose of selection requirements of category 5 is to reject physical backgrounds such as $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$, $B^0 \rightarrow \bar{D}^0 K^- \pi^+$ or $B_s^0 \rightarrow \bar{D}^0 K^- \pi^+$ and $\Lambda_b \rightarrow D^0 p K^-$ by requiring a minimal threshold w.r.t. `ProbNNp(p)` and a veto against large values of `ProbNNk(π)`, respectively. Since most of the charge tracks in an event are caused by genuine pions, the former selection requirement also helps to suppress combinatorial background (category 1). The combined efficiency for MC simulated events is 71.10(25) % where the uncertainty only takes statistical fluctuations into account (*cf.* Sec. 5.1.2).

6.1.3 Calibration

Three-body decays typically have resonances among their final states, whereas in two-body decays the only possible resonance is the mother of the decay chain itself. In particular, $\Lambda_b \rightarrow D^0 p \pi^-$ decays have a rich set of resonances among $D^0 p$ and $D^0 \pi^-$ pairs [69]. These resonances

²Changing the pion to a kaon reduces the branching fraction by a factor of λ^2 (Wolfenstein parameterization) but now allows an effective extraction of the CKM phase γ (*cf.* Sec. 1.5).

Tab. 6.2. Selection criteria of loose selection used for reconstructing $\Lambda_b \rightarrow D^0 p \pi^-$. The selections are grouped into four categories which are explained in Sec. 6.1.2.

Particle	Selection	Category
π	$3 \leq p \leq 150 \text{ GeV}/c$	1
K	$3 \leq p \leq 150 \text{ GeV}/c$	1
p	$9 \leq p \leq 150 \text{ GeV}/c$	1
p	$\text{ProbNNp} \geq 0.2$	1, 5
π	$\text{ProbNNk} \leq 0.3$	5
D^0	$0 \leq \text{FD sig.} \leq 100$	1
D^0	$\Delta\chi_{\text{IP}}^2 \text{ w.r.t. best PV} \geq 5$	1
Λ_b	$\Delta\chi_{\text{IP}}^2 \text{ w.r.t. best PV} \leq 25$	1
Λ_b	$5 \leq m \leq 6 \text{ GeV}/c^2$	3
Λ_b	$2 \leq \eta \leq 4.5$	4
Λ_b	$p_T \leq 20 \text{ GeV}/c$	4
Λ_b	$\exists \text{ converged DTF}$	1
Λ_b	$\chi_{\text{DTF}}^2/\text{DoF} \leq 10$	1

are not taken into account during the generation of the MC simulated decays. Instead, the decays are simulated flat in the so-called *square Dalitz plot* (*cf.* Refs. [98, 99] and in particular Ref. [100] for the definition of a square Dalitz plot). In order to make recorded data and MC simulated event comparable we calibrate the former by assigning weights to each event.³ The weights are the inverted values of the (normalized) smoothed square Dalitz plot profile, where the smoothing is performed in bins using a 5×5 kernel function

$$\mathcal{K} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 2 & 2 & 0 \\ 1 & 2 & 5 & 2 & 1 \\ 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

The smoothed profile is the result of a convolution with \mathcal{K} and a subsequent normalization where kernel pixels that extends past the boundary are set to zero (Kernel Crop). Below, we apply this efficiency correction to recorded data after the respective selection steps but suppress an explicit mention for the sake of brevity. As a consequence, we also use (binned) weighted minimum χ^2 -fits in the following instead of a (single entry) maximum likelihood method. Calibrating the recorded data changes the shapes of the physical background, whereas the impact on the fitted signal yields is less than 14%.

6.2 Yield Extraction

The distribution of the invariant mass $m(D^0 p \pi^-)$ after applying the loose selection requirements is shown in Fig. 6.1. An exhaustive fit model that comprises a precise signal model, as well as all background components such as partially reconstructed $\Lambda_b \rightarrow D^{*0} p \pi^-$ (large enhancement between $5.3 \text{ GeV}/c^2$ and $5.5 \text{ GeV}/c^2$) or reflections from $\Lambda_b \rightarrow D^0 p K^-$ decays or other B^0 and B_s^0 decays requires a thorough analysis of all these components, *cf.* Ref. [69]. For the

³Technically, the parameters m' and θ' (*cf.* Ref. [100]), are evaluated with a DTF where all intermediate particle masses (*i.e.*, including $m(\Lambda_b)$ itself) are constrained to increase the resolution of the Dalitz distribution.

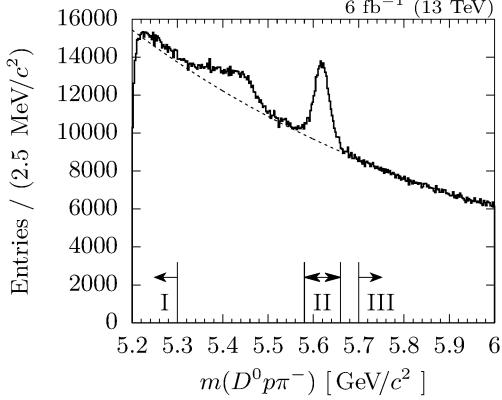


Fig. 6.1. Invariant mass distribution of D^0 , p and π^- candidates after applying the loose selection requirements. The regions I and III are used for extracting the shape of the combinatorial background. Region II indicates the signal region that is used for extracting the signal yield. The dashed line indicate the fitted shape of the combinatorial background.

present analysis we find that a simple yet flexible fit model suffices, while not contributing to the overall error budget excessively. Within this fit model, the signal region is limited tightly to $5.58 \leq m(D^0 p \pi^-) \leq 5.66 \text{ GeV}/c^2$ (region II in Fig. 6.1) and thus reduces the impact of physical background contribution, in particular of reflections below $m(\Lambda_b)$. The shape of the combinatorial background is parametrized with the exponential function given in Eq. (Q.3) and is extracted from a lower sideband $5.2 \leq m(D^0 p \pi^-) \leq 5.3 \text{ GeV}/c^2$ and an upper sideband $5.7 \leq m(D^0 p \pi^-) \leq 6 \text{ GeV}/c^2$, referred to as region I and III in Fig. 6.1. The signal yield is obtained by extrapolating the fitted background PDF into region II, scaling with the number of events in regions I and III, and subtracting the result from the number of events in region II. Applying this fitting strategy to the distribution of the invariant mass $m(D^0 p \pi^-)$ after the loose selection, as shown in Fig. 6.1, yields $7.04(6) \times 10^4$ signal events where the uncertainty only takes statistical fluctuations into account. In order to estimate the systematic uncertainty we use the same fitting strategy to extract the yields from two further reduced samples, where the former (latter) sample is obtained by requiring $\text{ProbNNk} \geq 0.3$ (0.57) for the kaon and a DTF probability above 0.01. The thresholds of the selection requirements w.r.t. the ProbNNk responses are the optimized values for LL and DD tracks that we found for the $\Lambda_b \rightarrow D^0 \Lambda$ decay in Chap. 5, whereas the threshold w.r.t. the DTF probability is chosen ad hoc and proven to significantly improve the signal to background ratio. Both samples are shown in Fig. 6.2.

Requiring both of these selection criteria results in a cleaner data sample but also unveils a bias of the fit model at the lower tail of the Λ_b signal peak. The fitted yields of these two fits are shown in Tab. 6.3. Applying the very same fitting strategy to MC simulated events gives access to the (MC predicted) efficiency of the required selection and allows the extrapolation to the total amount of events before requiring the selection. Both of these extrapolated numbers should match the number we found with our first fit. The deviation is an admixture of deviations due to fidelity issues of the MC simulated events and an overestimation of the signal yield due to a bias of the fitting model. A conservative approach to approximate the systematic uncertainty of the fit is to ignore the former part and take the entire deviation between the three fitted yields as an upper limit of the uncertainty interval,

$$n = 7.0(5) \times 10^4.$$

The uncertainty corresponds to a relative uncertainty of 7% which is sufficient for the sake

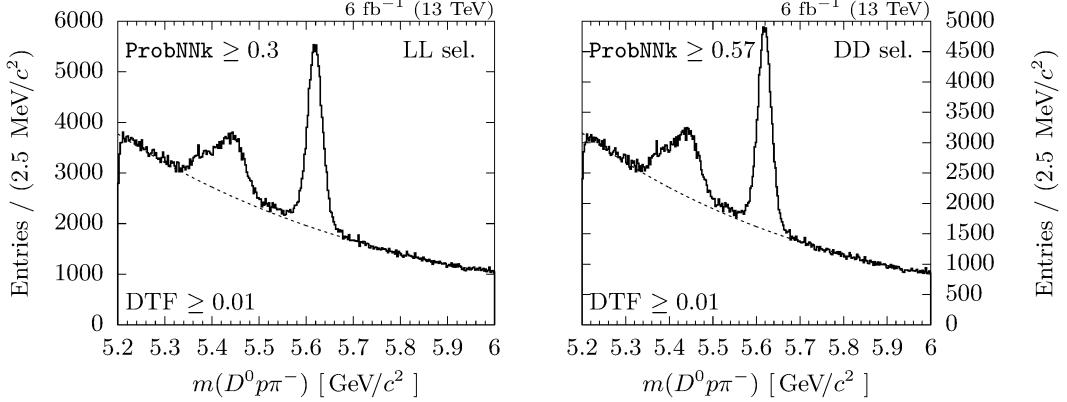


Fig. 6.2. Combined invariant mass distributions of D^0 , p and π^- candidates, as well as the fit of the combinatorial background (dashed line). The distributions are the result of a set of selection criteria that resembles part of the tight selection for LL (left) and DD (right) $\Lambda_b \rightarrow D^0 \Lambda$ decays.

of the present analysis. As stated, this approximation is conservative and renders additional studies, such as the unfolding of the strongly correlated systematic uncertainties of the three samples, unnecessary.

Tab. 6.3. Efficiencies of applying additional selection criteria (LL sel.) and (DD sel.) on top of the loose selection of $\Lambda_b \rightarrow D^0 p \pi^-$ decays, evaluated by fitting MC simulated events. These efficiencies are then used to extrapolate (ext.) the fitted yields (fit 2) of recorded data and compared with the fitted yield (fit 1) of the respective distribution after requiring only the loose selection.

	LL sel.	DD sel.
MC efficiency	77.020(34) %	71.70(4) %
Rec. data (fit 2)	$5.746(34) \times 10^4$	$5.301(32) \times 10^4$
Rec. data (ext.)	$7.45(4) \times 10^4$	$7.39(4) \times 10^4$
Rec. data (fit 1)		$7.04(6) \times 10^4$

In Chap. 9 we argue that only events with a positive L0 TIS trigger decision are needed for the normalization of the branching ratio. The corresponding distributions of the invariant mass $m(D^0 p \pi^-)$ and the respective fits for recorded data are shown in Fig. 6.3. The fits yield (including our 7% uncertainty estimation) $n = 3.93(28) \times 10^4$ and $n = 3.60(25) \times 10^4$ for the LL and DD selection, respectively.

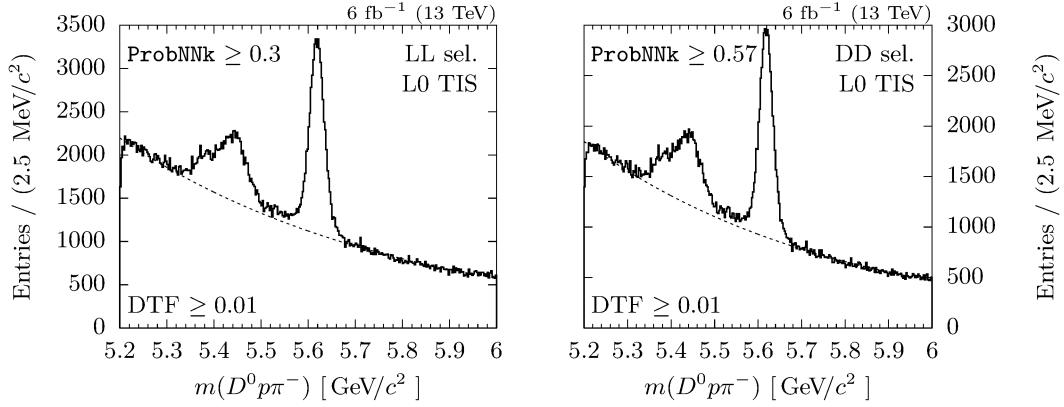


Fig. 6.3. Combined invariant mass of D^0 , p and π^- candidates from recorded data. The selection criteria resembles parts of the tight selection for LL (left) and DD $\Lambda_b \rightarrow D^0 \Lambda$ decays. Additionally, the L0 TIS trigger is required.

Chapter 7

Physical Backgrounds

Today's prediction is tomorrow's prior.

— Glen Cowan, during TAE summer school 2017

Regarding G. Cowan's quote about tomorrow's prior, decays that were first observed during Run 1, such as $\Lambda_b \rightarrow \Lambda hh'$ and $\bar{B}_s^0 \rightarrow D^0 K_s$ not only have become today's prior, they already have to be considered as background candidates in Run 2 analyses. In this section we discuss those and other physical background contributions which we separate into non-resonant backgrounds (Sec. 7.1), partially reconstructed backgrounds (Sec. 7.2), and reflections (Sec. 7.3). While we find that most background contributions can be neglected due to a sufficiently strong suppression, partially reconstructed decays including D^{*0} resonances or Σ^0 baryons, stay a nuisance and enter the fit model of the subsequent $m(D^0\Lambda)$ fit.

7.1 Non-Resonant Background

$\Lambda_b \rightarrow D^0 p\pi^-$ decays have the potential of being a dangerous (non-resonant) background for the $\Lambda_b \rightarrow D^0\Lambda$ mode, if the Λ daughters are reconstructed as LL tracks. If not sufficiently suppressed, this background is irreducible in the $m(D^0\Lambda)$ distribution and very hard to unfold from genuine $\Lambda_b \rightarrow D^0\Lambda$ decays by conventional fitting approaches. Luckily for the present analysis, the kinematic suppression of the stripping phase already gives a strong suppression,¹ as shown in Chap. 3. Applying the dedicated tight $\Lambda_b \rightarrow D^0\Lambda$ selection (*cf.* Sec. 5.2) gives an additional suppression and accepts 65 of 46 740 (weighted) events, where the latter is the amount of events that already pass the dedicated pre- and loose $\Lambda_b \rightarrow D^0\Lambda$ selection steps.² A 95 % confidence interval of the corresponding efficiency is $[1.1 \dots 1.8] \times 10^{-3}$. The efficiencies of each selection step individually, without requiring any of the other criteria, are shown in Tab. 7.1. Combining this efficiency with the suppression factor of the stripping phase leverages the estimation of an upper limit of the accumulated probability P of seeing at most n genuine $\Lambda_b \rightarrow D^0 p\pi^-$ events in the invariant mass of D^0 and Λ candidates:

$$P = \sum_{k=0}^n \binom{N}{k} p^k (1-p)^{N-k}.$$

In Fig. 7.1 we show the accumulated probability for $p = 9 \times 10^{-5}$ (dashed line) which corresponds to a conservative approximation of both factors. We argue that this suppression on its own is not sufficient. In the following we will therefore tighten the mass window of $m(p\pi^-)$ around $m(\Lambda)$ and establish a veto against small values of the Λ flight distance significance. These additional requirements decrease the combined efficiency to at most $p = 8 \times 10^{-6}$ at a 95 % CL (solid line in Fig. 7.1), corresponding to a probability of seeing not more than three events above 99 %.

¹Mostly due to the required finite mass window around $m(\Lambda)$.

²Technically, we use so-called *filtered* events for this task. Filtered events give larger yields but estimating the combined reconstruction and stripping efficiency is more challenging than for ordinary simulated events (which we used for the latter task).

Tab. 7.1. Efficiencies of applying the tight selection criteria as found in Sec. 5.2 of the $\Lambda_b \rightarrow D^0 \Lambda$ selection to genuine $\Lambda_b \rightarrow D^0 p\pi^-$ decays, evaluated with MC simulated events.

Selection criterion	Efficiency [%]
Λ Clf. ≥ 0.67	4.68(10)
$\Lambda_b - D^0$ Clf. ≥ 0.73	19.26(18)
ProbNNp ≥ 0.6	22.87(19)
ProbNNk ≥ 0.3	23.35(20)
DTF prob. ≥ 0.012	0.61(4)
Combination	0.139(17)

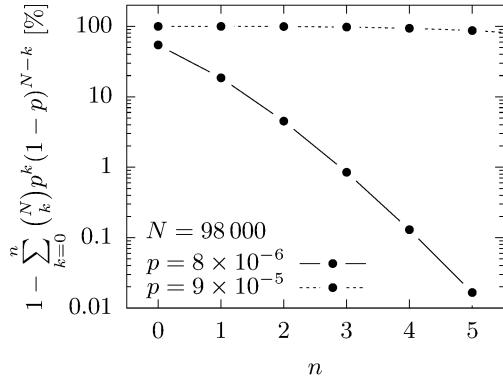


Fig. 7.1. Probability of seeing more than n genuine $\Lambda_b \rightarrow D^0 p\pi^-$ decays in $m(D^0 \Lambda)$. The solid (dashed) line corresponds to our choice of selection steps, including (excluding) the additional selection requirements w.r.t. $m(p\pi^-)$ and the Λ flight distance significance.

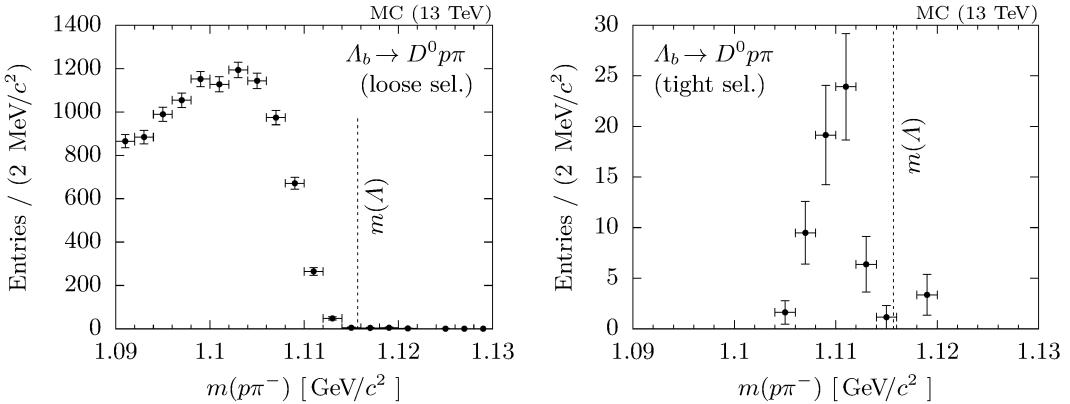


Fig. 7.2. Combined invariant mass of p and π^- candidates of simulated $\Lambda_b \rightarrow D^0 p\pi^-$ decays, reconstructed and fitted via a DTF as $\Lambda_b \rightarrow D^0 \Lambda$. The loose selection (left) and tight selection (right) are the dedicated $\Lambda_b \rightarrow D^0 \Lambda$ selections as described in Sec. 5.1.

In Fig. 7.2 we show the invariant mass distribution of $m(p\pi^-)$ before and after applying the dedicated $\Lambda_b \rightarrow D^0\Lambda$ tight selection. Apparently, it is not flat but strongly (asymmetrically) suppressed for $m(p\pi^-) > m(\Lambda)$. We find that this structure is introduced by requiring a minimum flight distance (*cf.* Sec. 5.1.2) and simultaneously constraining the Λ mass in a DTF (*cf.* Appx. C for a more detailed description of this correlation.) We choose

$$\begin{aligned} m(p\pi^-) - m(\Lambda_{\text{PDG}}) &> -4 \text{ MeV}, \\ \Lambda \text{ FD sig.} &> 25, \end{aligned}$$

where $m(\Lambda_{\text{PDG}})$ is taken from Ref. [15], as thresholds for $m(p\pi^-)$ and the flight distance significance of the Λ baryon which leaves, when applied in combination, only one event left (weight 1.12). In Fig. 7.3 we show the cumulative distribution of the latter in combination with and without the tight selection. With $n < 2$ and a preselection suppression factor of 20 this corresponds to an 95 % confidence interval of $[0.26 \dots 8] \times 10^{-6}$.

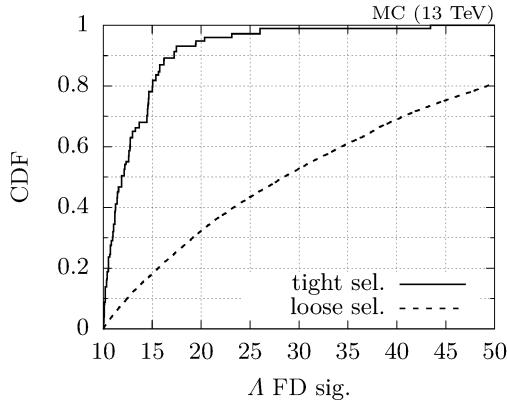


Fig. 7.3. Cumulative distribution of selections w.r.t. the flight distance significance of the (spurious) Λ candidates from $\Lambda_b \rightarrow D^0 p\pi^-$ decays after (solid line) and before (dashed line) applying the dedicated tight $\Lambda_b \rightarrow D^0\Lambda$ selection.

We refer to Ref. [69] for studies of Dalitz plots of $\Lambda_b \rightarrow D^0 p\pi^-$ decays and note that the Λ baryon is located at low values of $m_{p\pi}^2$ and large values of m_{Dp}^2 . Both areas show no large resonance structures and thus support the conservative trait of our approximation.

7.2 Partially Reconstructed Backgrounds

In the present analysis partially reconstructed backgrounds come from intermediate states that decay to either a soft photon or a soft neutral pion. These low energetic particles typically escape undetected at LHCb and thus making the respective backgrounds irreducible. In particular, we will consider $\Lambda_b/\Xi_b^0 \rightarrow D^0\Sigma^0$ and $\Xi_b^0 \rightarrow D^0\Xi^0$ decays, as well as $\Lambda_b/\Xi_b^0 \rightarrow D^{*0}\Lambda$, where in the former cases the γ (π^0) from $\Sigma^0 \rightarrow \Lambda\gamma$ ($\Xi^0 \rightarrow \Lambda\pi^0$), and in the latter case the γ (π^0) from $D^{*0} \rightarrow D^0\gamma$ ($D^{*0} \rightarrow D^0\pi^0$) is lost. We devoted the entire Sec. 1.3 of our theory part to the discussion of the irreducible $\Sigma^0 \rightarrow \Lambda\gamma$ background. In addition we analyze its kinematics, as well as the kinematics of the other partially reconstructed background sources in depth in Appx. O. Since these backgrounds are irreducible we will encounter them in our fitting model.

7.3 Reflections

Apart from random track combinations, non-resonant and partially reconstructed decays, so-called *reflections* potentially contaminate the signal region in the invariant mass distribution $m(D^0\Lambda)$. Reflections are fully reconstructed decays where a spurious mass hypothesis was assigned to at least one of the particles within the respective decay chain. Since reconstruction is typically performed in the upstream direction of the respective decay chain, this wrong assignment also propagates as a shift and dilution of the mass distributions. Decays, such as $\bar{B}_s^0 \rightarrow D^0 K_s$, can contaminate the signal region if the kaon decay $K_s \rightarrow \pi^+ \pi^-$ becomes a reflection by misidentifying the π^+ as a proton and thus fakes a $\Lambda \rightarrow p \pi^-$ decay. This smears the invariant mass distribution of the K_s , as well as the B_s^0 meson mass distribution and consequently introduces background contributions to the invariant mass distribution $m(D^0\Lambda)$.

For a given n -body decay the invariant mass of the mother M is given by the invariant masses m_i and three-momenta \vec{p}_i of the daughters via the relation

$$M^2 = \left(\sum_{i=1}^n \sqrt{m_i^2 + \vec{p}_i^2} \right)^2 - \left(\sum_{i=1}^n \vec{p}_i \right)^2.$$

The shift of M in case of a spurious mass hypothesis $m_j \rightarrow m_j + \delta m_j$ is non-linear in δm_j and depends on the respective momenta, hence the transformation of the invariant mass distributions due to reflections is non-trivial and depend on the momentum distributions.

In the following we discuss the two main sources of reflections in the present analysis. In Sec. 7.3.1 we study charmless three-body decays and then discuss the possible contamination of genuine K_s decays in Sec. 7.3.2. Both background candidates are eventually found to be negligible given the amount of available statistics after applying the dedicated tight $\Lambda_b \rightarrow D^0\Lambda$ selection and are thus not included in the subsequent $m(D^0\Lambda)$ fit model.

7.3.1 Charmless Decays

Charmless three-body decays $\Lambda_b/\Xi_b^0 \rightarrow \Lambda h h'$ are a nuisance in searches for CP violation in $\Lambda_b \rightarrow D^0\Lambda$ decays when hh' are spuriously reconstructed as $D^0 \rightarrow hh'$. The reported branching fractions are (*cf.* Refs. [15, 39])

$$\begin{aligned} \mathcal{B}(\Lambda_b \rightarrow \Lambda K^+ K^-) &= (1.62 \pm 0.23) \times 10^{-5}, \\ \mathcal{B}(\Lambda_b \rightarrow \Lambda K^+ \pi^-) &= (5.7 \pm 1.3) \times 10^{-6}, \\ \mathcal{B}(\Lambda_b \rightarrow \Lambda \pi^+ \pi^-) &= (4.7 \pm 1.9) \times 10^{-6}, \end{aligned}$$

and thus will appear as physical background in the relevant ADS and GLW modes. Similarly, $\Xi_b^0 \rightarrow \Lambda K^+ K^-$, $\Xi_b^0 \rightarrow \Lambda K^- \pi^+$ and $\Xi_b^0 \rightarrow \pi^+ \pi^-$ decays are background candidates in $\Xi_b^0 \rightarrow D^0\Lambda$ analyses. All these modes are hard to control and require decent statistics to, for example, estimate their contribution to the signal yields by analyzing the combined invariant masses of D^0 and Λ candidates in slices of the D^0 flight distance. Beneficially for the present analysis, the $\Lambda_b \rightarrow \Lambda K^- \pi^+$ decay is suppressed w.r.t. $\Lambda_b \rightarrow \Lambda K^+ \pi^-$ with at least a factor of a box diagram and neither $\Lambda_b \rightarrow \Lambda K^- \pi^+$, nor $\Xi_b^0 \rightarrow \Lambda K^- \pi^+$ were observed experimentally at the time of writing.

For the present analysis where we reconstruct the D^0 meson as $D^0 \rightarrow K^- \pi^+$, the $\Lambda_b \rightarrow \Lambda K^+ K^-$ background only enters indirectly via reflection. Below, we show that we can suppress the contribution coming from $\Lambda_b \rightarrow \Lambda K^+ K^-$ decays (largest branching fraction) sufficiently and thus renders dedicated analyses of the remaining $\Lambda_b \rightarrow \Lambda h h'$ modes unnecessary. We use weighted, truth-matched MC simulated events³ for this task where we rely on the established weights without adding acceptance corrections due to the limited dataset.

³Generated flat in phase space.

First, we apply the dedicated $\Lambda_b \rightarrow D^0\Lambda$ preselection step with the tightened requirements we established in Sec. 7.1 which already greatly reduces the $\Lambda_b \rightarrow \Lambda K^+ K^-$ data set from 2875 to 570 and 7055 to 1785 events⁴ for LL and DD tracks, respectively.⁵ Secondly, we apply the selection steps w.r.t. the DTF probability, and the responses of the ProbNNp and ProbNNk classifiers for the p and K^- , respectively. The corresponding efficiencies are listed in Tab. 7.2. The efficiencies for the selection requirements w.r.t. ProbNNp and ProbNNk are compatible with those we found for genuine $\Lambda_b \rightarrow D^0\Lambda$ decays, whereas the DTF reveals its strong separation power due to the D^0 mass constraint.

Tab. 7.2. Efficiencies of the dedicated $\Lambda_b \rightarrow D^0\Lambda$ tight selection requirements when applied to $\Lambda_b \rightarrow \Lambda K^+ K^-$, reconstructed as $\Lambda_b \rightarrow D^0\Lambda$, and genuine $\Lambda_b \rightarrow D^0\Lambda$ decays. The efficiencies are obtained from weighted MC simulated events.

	$\Lambda_b \rightarrow \Lambda K^+ K^-$		$\Lambda_b \rightarrow D^0\Lambda$	
	LL [%]	DD [%]	LL [%]	DD [%]
DTF prob.	43.5(21)	28.1(11)	89.16(31)	82.07(24)
ProbNNp(p)	94.2(10)	86.6(8)	92.87(26)	81.54(25)
ProbNNk(K)	91.3(12)	84.7(9)	92.51(26)	84.50(23)
Combination	26.4(18)	24.9(10)	88.51(32)	88.67(20)

The efficiencies of applying the Λ and Λ_b - D^0 classifier selection requirements are estimated after applying all previous discussed selections and are listed in Tab. 7.3. Interestingly, both classifiers show a strong separation power, except for the Λ classifier when applied to LL tracks. In order to understand this behavior we train SVMs on the reduced data set using all pairwise combinations of two features to estimate each feature importance. For the Λ_b - D^0 classifier we find that almost the entire separation power stems from the D^0 flight distance feature. This is expected, since in the case of genuine $\Lambda_b \rightarrow \Lambda K^+ K^-$ decays, the fitted flight distance of the $K^+ K^-$ pair is smaller on average than for genuine D^0 meson. More surprisingly, the Λ classifier is also capable to separate between genuine $\Lambda_b \rightarrow D^0\Lambda$ and $\Lambda_b \rightarrow \Lambda K^+ K^-$ decays. The rather large deviation for LL and DD tracks is rooted in the mild thresholds of the former but adjusts when the thresholds are raised. Using the same technique we used previously, we find that the separation power is driven by deviations in the p_T distributions which we discuss in detail in Appx. N.2.

Tab. 7.3. Efficiencies of the optimized $\Lambda_b \rightarrow D^0\Lambda$ tight selection requirements when applied to $\Lambda_b \rightarrow \Lambda K^+ K^-$, reconstructed as $\Lambda_b \rightarrow D^0\Lambda$, and genuine $\Lambda_b \rightarrow D^0\Lambda$ decays. The efficiencies are obtained from weighted MC simulated events after applying the selection steps listed in Tab. 7.2.

	$\Lambda_b \rightarrow \Lambda K^+ K^-$		$\Lambda_b \rightarrow D^0\Lambda$	
	LL [%]	DD [%]	LL [%]	DD [%]
Λ Clf.	97.5(11)	34.3(27)	96.54(21)	50.3(4)
Λ_b - D^0 Clf.	68(4)	31.0(26)	80.0(5)	47.0(4)
Combination	67(4)	18.9(21)	77.9(5)	33.5(4)

Applying the entire dedicated $\Lambda_b \rightarrow D^0\Lambda$ tight selection leaves 122 and 72 weighted events left

⁴The reduced numbers are the accumulated weights of the MC simulated events.

⁵This strong suppression is driven by the fact that the dedicated $\Lambda_b \rightarrow D^0\Lambda$ preselection step already includes requirements w.r.t. the fit probability, D^0 flight distance, etc.

for LL and DD tracks, respectively. Their corresponding invariant mass distributions $m(D^0\Lambda)$ are shown in Fig. 7.4. (We refer to Appx. N.1 for a brief discussion of these mass distributions.) Using the suppression factor for LL and DD tracks that we established in Sec. 3, we expect 7.8(15) and 4.8(9) genuine $\Lambda_b \rightarrow \Lambda K^+ K^-$ events after the dedicated $\Lambda_b \rightarrow D^0\Lambda$ tight selection steps.

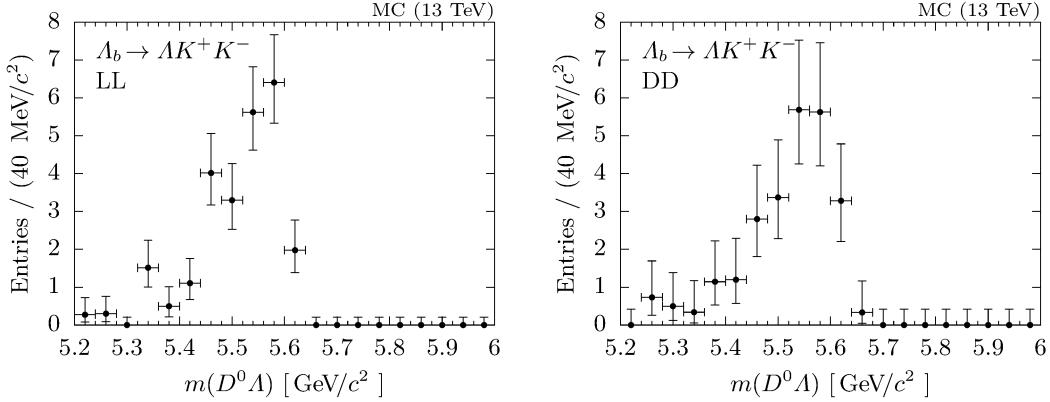


Fig. 7.4. Combined invariant mass of $D^0 \rightarrow K^- \pi^+$ and $\Lambda \rightarrow p \pi^-$ candidates from MC simulated $\Lambda_b \rightarrow \Lambda K^+ K^-$ decays for LL (left) and DD (right) tracks.

In theory, these amounts could be suppressed further by requiring a minimal flight distance of the D^0 candidates. For example, requiring

$$D^0 \text{ FD sig.} > 2, \quad (7.1)$$

reduces the expectation of genuine $\Lambda_b \rightarrow \Lambda K^+ K^-$ events to 2.6(6) and 0.79(26), but also suppresses $\Lambda_b \rightarrow D^0\Lambda$ to 83.6(5)% and 80.7(5)% for LL and DD, respectively (*cf.* Fig. N.4). Alternatively, the reflection shape can be parametrized and fitted in recorded data. However, regarding the limited data sample, we find in Chap. 8 that an unfolding from $\Xi_b^0 \rightarrow D^{*0}\Lambda$ contributions is not possible and since the latter component is disfavored by our fit model, we argue that $\Lambda_b \rightarrow \Lambda K^+ K^-$ contributions in the signal region are negligible without further suppression. We cross-check this assumption by requiring Eq. (7.1) and fitting the invariant mass of D^0 and Λ candidates from recorded data in configuration 1. (See Chap. 8 for a description of the fit model and its configurations, as well as Fig. P.9 for a projection of the fit to recorded data.) The fitted signal yield of $\Lambda_b \rightarrow D^0\Lambda$ decays reduces from 31(7) to 23(6). Assuming a reduction of 82% (combination of both track types), the expected yield is 25.4(21) and thus includes the fitted yield within two standard deviations. We thus conclude that $\Lambda_b \rightarrow \Lambda K^+ K^-$ events contribute negligible in the signal region. Since $\Xi_b^0 \rightarrow \Lambda K^- \pi^+$ decays are expected to behave similarly and due to the additional CKM suppression w.r.t. $\Xi_b^0 \rightarrow D^0\Lambda$ decays, we argue further that the contribution of charmless backgrounds in the Ξ_b^0 mode is negligible, too.

7.3.2 K_s Reflections from b -Meson Decays

In the decay $\Lambda_b \rightarrow D^0\Lambda$ we expect reflections coming from B^0 and B_s^0 meson decays to $D^0 K_s$, caused by a misidentification of the π^+ in $K_s \rightarrow \pi^+ \pi^-$. The leading contribution to the decays $\bar{B}^0 \rightarrow D^0 K_s$ and $\bar{B}_s^0 \rightarrow D^0 K_s$ are internal, color-suppressed tree diagrams, where the former includes the Cabibbo suppressed $b \rightarrow c s \bar{u}$ transition and the latter the Cabibbo favored $b \rightarrow c d \bar{u}$ transition. The respective B^0 branching fraction is well established, measured by the Babar and

Belle collaborations [101, 102], and the B_s^0 mode was first observed and measured at LHCb [103]:

$$\begin{aligned}\mathcal{B}(\bar{B}^0 \rightarrow D^0 \bar{K}^0) &= (5.2 \pm 0.7) \times 10^{-5}, \\ \mathcal{B}(\bar{B}_s^0 \rightarrow D^0 K^0) &= (4.3 \pm 0.9) \times 10^{-4}.\end{aligned}$$

Besides being very useful for leveraging access to the CKM phase γ and even ϕ_s in the latter case, both branching fractions are significantly larger than the expected branching fraction of $\Lambda_b \rightarrow D^0 \Lambda$ [49], making them also background candidates for the present analysis. However, it is the fraction κ of misidentified B^0 and B_s^0 mesons in the signal region of $\Lambda_b \rightarrow D^0 \Lambda$ that has to be considered. This fraction can be estimated by the product of ratios of the b -hadron productions (I), the branching fractions of the respective decay modes (II), the reconstruction efficiencies (III), and the selection efficiencies (IV):

$$\kappa(B_{(s)}^0) := \underbrace{\frac{f_{B_{(s)}^0}}{f_{\Lambda_b}}}_{\text{I}} \times \underbrace{\frac{\mathcal{B}(\bar{B}_{(s)}^0 \rightarrow D^0 K_s)}{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)}}_{\text{II}} \times \underbrace{\frac{\mathcal{B}(K_s \rightarrow \pi^+ \pi^-)}{\mathcal{B}(\Lambda \rightarrow p \pi^-)}}_{\text{III}} \times \underbrace{\frac{\varepsilon_{\text{rec}}(\bar{B}_{(s)}^0 \rightarrow D^0 K_s)}{\varepsilon_{\text{rec}}(\Lambda_b \rightarrow D^0 \Lambda)}}_{\text{IV}} \times \underbrace{\frac{\varepsilon_{\text{sel}}(\bar{B}_{(s)}^0 \rightarrow D^0 K_s)}{\varepsilon_{\text{sel}}(\Lambda_b \rightarrow D^0 \Lambda)}}_{\text{IV}}.$$

The relative production rates (I) are measured by LHCb [96, 97]. Factor II is found by inserting the theory prediction from Ref. [49] and the respective established branching ratios,⁶:

$$\frac{\mathcal{B}(\bar{B}_{(s)}^0 \rightarrow D^0 K_s)}{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)} \times \frac{\mathcal{B}(K_s \rightarrow \pi^+ \pi^-)}{\mathcal{B}(\Lambda \rightarrow p \pi^-)} = \begin{cases} 6.2(8) & \text{for } \bar{B}^0 \rightarrow D^0 K_s, \\ 51(11) & \text{for } \bar{B}_s^0 \rightarrow D^0 K_s. \end{cases}$$

The final approximation will be dominated by the vague uncertainty of the theory prediction for $\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)$, hence we aim for an upper limit of κ . In this context we assume the ratio of the reconstruction efficiencies (III) to be one in good approximation. Factor IV for $\bar{B}_s^0 \rightarrow D^0 K_s$ is found using weighted simulated events and used as an upper limit of $\bar{B}^0 \rightarrow D^0 K_s$. Due to the lack of a sufficient amount of MC simulated events we use the simulation framework **RapidSim** [104] instead. These simulated events are calibrated with weights obtained from $B^0 \rightarrow J/\psi K_s$ rather than $\Lambda_b \rightarrow J/\psi \Lambda$ decays due to the initial b -meson, using the same techniques outlined in Sec. 4.2.2. Similar to the previous sections,⁷ we find a sufficient suppression, due to the very broad $m(p \pi^-)$ distribution for misidentified $K_s \rightarrow \pi^+ \pi^-$ decays, on its own allowing a suppression (factor IV) below 3% without loosing any genuine $\Lambda \rightarrow p \pi^-$ event. Applying the entire $\Lambda_b \rightarrow D^0 \Lambda$ tight selection eventually reduces $\kappa(\bar{B}_s^0)$ below 4% at a 95% confidence level, rendering an additional suppression, *e.g.*, via a $m(K^0)$ veto unnecessary.

⁶The kaon in the decay of the B^0 (B_s^0) meson is produced in an unmixed flavor eigenstate \bar{K}^0 (K^0), whereas the time-evolution propagates the mass eigenstates K_L and K_S . Here, we assume $|\langle K_S | K^0 \rangle|^2 = |\langle K_S | \bar{K}^0 \rangle|^2 = 1/2$ and thus neglecting CP violating effects in the kaon system which are known to be at a level of 10^{-3} [15].

⁷For the sake of brevity we abstain from a detailed discussion. The chosen techniques are equivalent to the ones that were outlined previously, and the remarkable strong suppression of rectangular selections w.r.t. mass and **ProbNNp** tolerate large uncertainties. **RapidSim**'s fidelity issues concerning the resolution of K_S particles is found to be negligible in this context.

Chapter 8

Yield Estimation

With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.

— Discovered by John von Neumann and exercised in Ref. [105].

In this chapter we establish a fit model to extract signal yields from 260 signal candidates which are left after applying the previously outlined selection criteria to the Run 2 data sample. This sparse data sample forces us to assume certain constraints and work with approximations. We outline our assumptions and meticulously evaluate our fit model which eventually allows us to obtain signal yields and yield significances for the decay modes $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$.

8.1 The Fit Model

Due to the small amount of available recorded signal candidates that passed the dedicated filtering steps and the absence of a precise background model in terms of statistics and polarization, we base our fit model onto the following assumptions:

- The kinematic properties of the final states of signal and background events are following similar distributions, since they would have been filtered out by the preceding filtering steps otherwise. Their resolution is thus given by the signal shape of the Λ_b or Ξ_b^0 baryon which are accessible through the respective MC simulated decays in good approximation. This assumption is based on the short lifetime of the b -hadrons w.r.t. the detector resolution, corresponding to a natural width of roughly $0.4 \text{ meV}/c^2$.
- Due to the absence of a priori knowledge concerning the polarization of the D^{*0} and Σ^0 states in $\Lambda_b \rightarrow D^{*0} \Lambda$ and $\Lambda_b \rightarrow D^0 \Sigma^0$, it is not possible to unfold the contribution of the latter from the broad $D^0 \pi^0$ and $D^0 \gamma$ decay modes of the former (*cf.* Tab. O.1). We verify this assumption with pseudo-experiments. The same assumption holds for the very broad (potential) background of $\Xi_b^0 \rightarrow D^0 \Xi^0$ decays.
- For similar reasons we fix the yield ratio of the fit components of $D^{*0} \rightarrow D^0 \pi^0$ and $D^{*0} \rightarrow D^0 \gamma$ to its central value of 1.83 measured by the Babar and BESIII collaborations and fitted by the PDG [106, 107, 15]. Further, neither the ratio of the fitted yields of $\Lambda_b \rightarrow D^0 \Lambda$ and $\Lambda_b \rightarrow D^{*0} \Lambda$, nor of $\Xi_b^0 \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^{*0} \Lambda$, nor of $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$ itself, should depend on the track type of the daughters of the Λ baryon and are therefore fitted simultaneously for both track types. For $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$, we cross-check this assumption with MC simulated events. The double ratio (*cf.* Tab. 8.1) is 1.026(23) and barely significantly deviates from one. The central value is less than 3% which is negligible w.r.t. the total systematic uncertainties.
- The combinatorial background follows an exponential function which we parametrize according to Eq. (Q.2). Our choice of parameter k 's sign is physically motivated but not

required during the fits. We note that this implicitly includes the linear model $1 \pm kx$ for small values of k and in particular the uniformly distributed (combinatorial) background for $k = 0$.

The fit is applied to the combined invariant mass of D^0 and Λ candidates and evaluated simultaneously on the unbinned datasets of simulated and recorded events of both track types. The total amount of available events for each of these six samples is listed in Tab. 8.1. The fit

Tab. 8.1. Amount of MC simulated and recorded events available for fitting. The fit is evaluated simultaneously in all six samples. Further, we give the (weighted) fraction for the MC simulated events w.r.t. the amount of events N , after the preselection. The double ratio of these fractions is 1.025(24) and thus barely deviate from one significantly.

	LL			DD
	n	n/N [%]	n	n/N [%]
MC sim. $\Lambda_b \rightarrow D^0 \Lambda$	5682	36.0(4)	5548	13.81(19)
MC sim. $\Xi_b^0 \rightarrow D^0 \Lambda$	6244	35.0(4)	6432	13.27(18)
rec. data	113	—	147	—

model is a normalized, weighted sum of the three components \mathcal{S} , \mathcal{D} and \mathcal{B} , describing the Λ_b/Ξ_b^0 signal, the D^{*0} background and the combinatorial background, respectively,

$$\mathcal{L} := \left\| \begin{pmatrix} \mathcal{S} \\ \mathcal{D} \\ \mathcal{B} \end{pmatrix} \right\|_{\vec{f}} \equiv (1 - f_1) \times \mathcal{S} + (1 - f_2) f_1 \times \mathcal{D} + f_1 f_2 \times \mathcal{B}. \quad (8.1)$$

For brevity we introduced the shorthand notation

$$\left\| \vec{\mathcal{V}} \right\|_{\vec{f}} := \sum_{j=1}^n p_j(\vec{f}) \times \mathcal{V}_j,$$

which wraps the (normalized) weighted sum of the components $\vec{\mathcal{V}} = (\mathcal{V}_1, \dots, \mathcal{V}_n)$ with weights $\vec{f} = (f_1, \dots, f_{n-1})$, by defining p_j according to

$$\begin{aligned} p_j &:= (1 - f_j) \prod_{i=1}^{j-1} f_i \quad \text{for } j < n, \\ p_n &:= \prod_{i=1}^{n-1} f_i. \end{aligned}$$

Using this notation we define \mathcal{S} and \mathcal{D} for Λ_b and Ξ_b :

$$\begin{aligned} \mathcal{S} &:= \left\| \begin{pmatrix} \mathcal{G}^{(2)}(\Xi_b) \\ \mathcal{G}^{(2)}(\Lambda_b) \end{pmatrix} \right\|_{f_s}, \\ \mathcal{D} &:= \left\| \begin{pmatrix} \mathcal{K}(\Xi_b) * \mathcal{G}_c^{(2)}(\Lambda_b) \\ \mathcal{K}(\Lambda_b) * \mathcal{G}_c^{(2)}(\Lambda_b) \end{pmatrix} \right\|_{f_{D^*}}, \end{aligned}$$

where $\mathcal{G}^{(2)}$ ($\mathcal{G}_c^{(2)}$) refers to a double Gaussian with shared (zero) mean

$$\mathcal{G}^{(2)} := \left\| \begin{pmatrix} \mathcal{G}_1 \\ \mathcal{G}_2 \end{pmatrix} \right\|_{f_G},$$

and \mathcal{K} are the kernel functions for $\Lambda_b/\Xi_b^0 \rightarrow D^{*0}\Lambda$ decays as introduced in Appx. O. More precisely, the latter is composed of the fixed superposition of the D^{*0} modes $D^{*0} \rightarrow D^0\pi^0$ and $D^{*0} \rightarrow D^0\gamma$, and is convoluted with the resolution function as motivated previously. During evaluation of the likelihood for the MC simulated samples, the corresponding signal component is separated by enforcing $(f_1, f_s) = (0, 0)$ or $(f_1, f_s) = (0, 1)$. Besides that, f_2 and f_s are constrained to not depend on the track type, whereas f_1 is allowed to vary among different track types. Similarly, the Gaussian shapes

$$\begin{aligned}\mathcal{G}_1 &:= \begin{cases} \mathcal{G}(x|\mu, \sigma_1^{(\text{LL})}) & \text{for LL,} \\ \mathcal{G}(x|\mu + \Delta\mu, \sigma_1^{(\text{DD})}) & \text{for DD,} \end{cases} \\ \mathcal{G}_2 &:= \begin{cases} \mathcal{G}(x|\mu, \sigma_2^{(\text{LL})}) & \text{for LL,} \\ \mathcal{G}(x|\mu + \Delta\mu, \sigma_2^{(\text{DD})}) & \text{for DD,} \end{cases}\end{aligned}$$

are allowed to deviate among different track types. This includes $f_{\mathcal{G}}$ and the shared mean value, technically encoded by the shared shift $\Delta\mu$.

The unbinned, single entry fit w.r.t. Eq. (8.1) is performed with different configurations of fixed and floating parameters and within the two different mass ranges $5.5 \leq m(D^0\Lambda) \leq 6 \text{ GeV}/c^2$ and $5.2 \leq m(D^0\Lambda) \leq 6 \text{ GeV}/c^2$. We tried different combinations of floating and fixed polarizations in the D^* kernel functions \mathcal{K} but could not obtain reasonable results. Thus we stick to an unpolarized description, leaving with six configurations which yield converging fits:

1. Using the narrow mass range $5.5 \leq m(D^0\Lambda) \leq 6 \text{ GeV}/c^2$ cuts off most of the $\Lambda_b \rightarrow D^{*0}\Lambda$ background. The value of the fraction f_{D^*} is thus fixed to zero, making f_2 the yield of $\Xi_b^0 \rightarrow D^{*0}\Lambda$ and leaves 22 floating parameters that are optimized by the fit. The fit converges successfully, but prefers $f_2 = 1$ (upper boundary). The fitted values are listed in Tab. P.1. In Fig. 8.1 we show the accumulated projection of recorded data and the fit function. Other projections are shown in Appx. P.
2. In order to extract the significance of the signal yields observed with the previous fit, we fix all contributions coming from MC simulated events, keep f_{D^*} fixed to zero and disable D^{*0} contributions in accordance with the previous results by enforcing f_2 to one. The likelihood Eq. (8.1) is then maximized with f_s fixed to one and zero, corresponding to the removal of the Ξ_b^0 and Λ_b component, respectively, and then compared with the results when f_s is a floating parameter. The projection of recorded data of the former two fits is shown in Fig. P.6 together with respective fit projections. Since the fits with f_s set to one and zero have exactly one DoF less than the fit which allows floating values for f_s , the difference of their respective log-likelihoods $\Delta \ln \mathcal{L}$ can be used to approximate the (statistical) significance $S = \sqrt{-2\Delta \ln \mathcal{L}}$. Doing so, we estimate a signal significance of $S = 5.5$ and $S = 1.8$ for the decays $\Lambda_b \rightarrow D^0\Lambda$ and $\Xi_b^0 \rightarrow D^0\Lambda$, respectively. This approximation is based on Wilks theorem [108] whose applicability is tested with a pseudo-experiment in Sec. 8.3.
3. In order to estimate the significance of the $\Lambda_b \rightarrow D^{*0}\Lambda$ background, we reevaluate the fit using the broad mass range $5.2 \leq m(D^0\Lambda) \leq 6 \text{ GeV}/c^2$ and keep all fit parameters floating. The invariant mass distributions and fit projections of recorded data are shown in Fig. P.7. We note that f_2 does not depend on the track type, presumably causing the combinatorial background for LL tracks to overshoot at low $m(D^0\Lambda)$ values.
4. Extending the fit model and allowing f_2 to take on different values for different track types fixes the issues that we observed during the previous fit (*cf.* Fig. P.8). The fitted values for f_2 are 0.66(12) and 0.98(9) for LL and DD tracks, respectively, differ strongly and cannot be physically motivated.

5. We repeat the fit in configuration 3 but this time replace the resolution function used for smearing the D^{*0} kernel functions \mathcal{K} with the centralized Ξ_b^0 signal shape.
6. We artificially enrich the data set by drawing n instances from a smeared, unpolarized $\Lambda_b \rightarrow D^0 \Sigma^0$ distribution, as derived in Appx. O, to the recorded data sample. The amount of drawn instances n is fixed to $1/3$ of the fitted Λ_b signal yield in accordance with our estimations in Sec. 1.3. Since we expect genuine $\Lambda_b \rightarrow D^0 \Sigma^0$ in the recorded data, we consider this is a conservative approximation of this background. We abstain from a similar cross-check of $\Xi_b^0 \rightarrow D^0 \Sigma^0$ decays due to the low $\Xi_b^0 \rightarrow D^0 \Lambda$ yields and the unclear branching ratio of the Σ^0 mode.
7. In the previous configurations the mass difference $m(\Xi_b^0) - m(\Lambda_b)$ was fixed to its nominal value $172.5 \text{ MeV}/c^2$. The PDG states an uncertainty of $0.4 \text{ MeV}/c^2$ which we use to test the sensitivity of our model to this value by setting $m(\Xi_b^0) - m(\Lambda_b)$ to $172.1 \text{ MeV}/c^2$ (7a) and $172.9 \text{ MeV}/c^2$ (7b).

Regarding the yields of the decays $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$, the results of the outlined fits are compatible and give almost identical values for both modes as listed in Tab. 8.2.

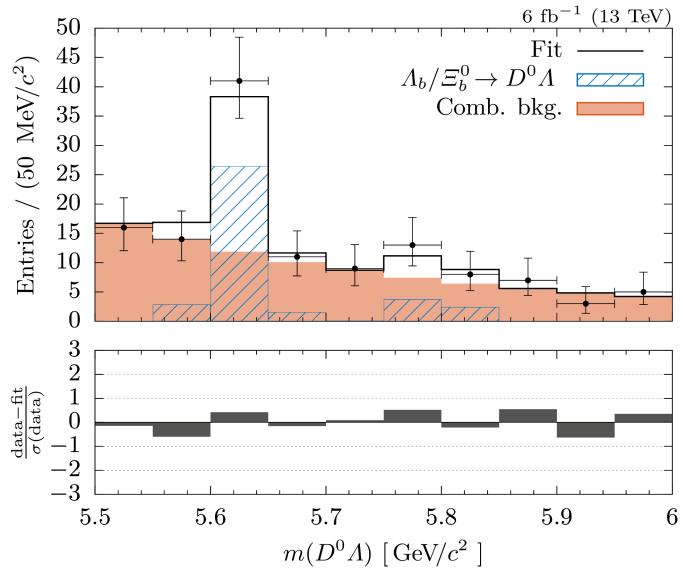


Fig. 8.1. Combined invariant mass of D^0 and Λ candidates (of both track types), as well as the accumulated projection of the fit in configuration 1. Since f_2 is compatible with one (at upper limit), the corresponding D^{*0} background contribution is suppressed (graphically).

8.2 Yield Extraction

According to our definition of the likelihood in Eq. (8.1) the expanded fractions of $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$ are $(1 - f_1) \times f_s$ and $(1 - f_1) \times (1 - f_s)$, respectively. The yields for each track type are found by multiplication with the amount of recorded events of the given track type n_{LL} and n_{DD} . Consequently, the accumulated expanded fraction for both track types is given by the weighted sum of the track type dependent fractions where n_{LL} and n_{DD} are the respective weights. We note that for estimating the yields by multiplication with $N = n_{\text{LL}} + n_{\text{DD}}$, the

Tab. 8.2. Corrected yields as obtained from (unbinned, single-entry) likelihood maximization in the configurations 1 and 3 - 7. The extraction and correction of the yields is discussed in Sec. 8.2.

Configuration	$\Lambda_b \rightarrow D^0 \Lambda$			$\Xi_b^0 \rightarrow D^0 \Lambda$		
	LL & DD	LL	DD	LL & DD	LL	DD
Fit 1	31 ± 7	16 ± 5	15 ± 5	6 ± 4	3.2 ± 2.2	3.0 ± 2.2
Fit 3	32 ± 7	16 ± 5	16 ± 5	6 ± 4	2.8 ± 2.1	2.9 ± 2.3
Fit 4	32 ± 7	16 ± 5	16 ± 5	6 ± 4	3.1 ± 2.2	2.9 ± 2.2
Fit 5	32 ± 7	16 ± 5	16 ± 5	6 ± 4	2.9 ± 2.2	2.9 ± 2.3
Fit 6	32 ± 7	16 ± 5	16 ± 5	6 ± 4	3.0 ± 2.2	3.0 ± 2.3
Fit 7a	32 ± 7	16 ± 5	16 ± 5	6 ± 4	3.1 ± 2.3	2.8 ± 2.2
Fit 7b	32 ± 7	16 ± 5	16 ± 5	6 ± 4	3.0 ± 2.4	3.2 ± 2.3

statistical uncertainty of n_{LL} and n_{DD} only contributes in N , not in the weights since they are part of the chosen (exact) projection of the fit.¹ Besides that, we find that ordinary error propagation with the fitted correlations of f_1 and f_s and symmetric uncertainties is sufficient due to the low asymmetry of each of the respective asymmetric uncertainties of 3% or less.

In order to investigate a possible bias of the yields and to test the validity of the error estimates we run a pseudo-experiment where we draw $N = n_{\text{LL}} + n_{\text{DD}}$ random events from the fitted PDF (configuration 1) and apply the very same unbinned maximum likelihood fit to these generated data that was used to obtain the parameters of the PDF itself. In Fig. 8.2 we show the distribution of the fitted yields of the $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$ components after 1000 consecutive runs of the outlined technique. (Also see Appx. P for more figures of the pseudo-experiments.) For an unbiased fit with valid error estimates, both distributions should

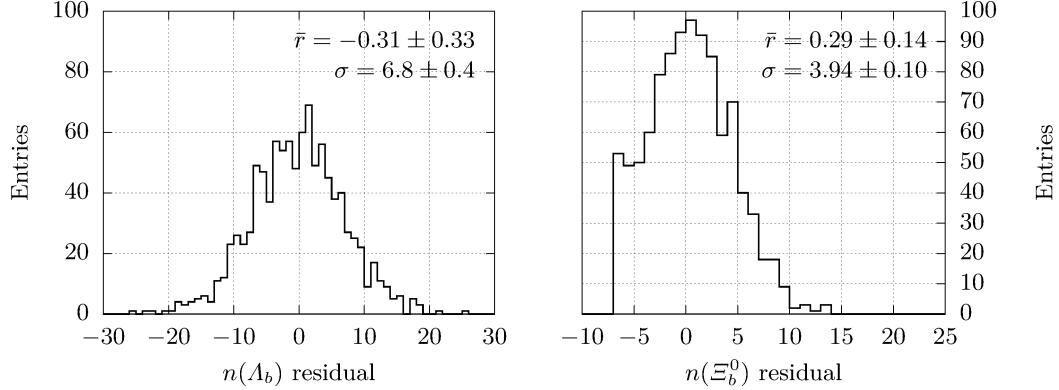


Fig. 8.2. Difference of the fitted signal yields of pseudo-experiments and recorded data (residual) where the latter was used during generation of the former, as well as the difference of the respective expected and actual sample mean (\bar{r}) and standard deviation (σ). If $\bar{r} = 0$ and if the standard deviation is consistent with those found by the fit to recorded data, then the fit is unbiased and has valid error estimations, respectively.

be distributed according to a clipped Gaussian distribution (*cf.* Appx. R). If so, the sample mean and the standard deviation would follow Eq. (R.1) and Eq. (R.2), respectively, and the difference

¹This Poisson part contributes less than 7% to the total uncertainty of the estimated yields. The rest is the multinomial error.

of the expected and actual sample mean (referred to as \bar{r} in Fig. 8.2) should be compatible with zero. From Fig. 8.2 we see that the expectations are met for both standard deviations and the mean value of the pseudo-experiments of the $\Lambda_b \rightarrow D^0 \Lambda$ fit, within the respective uncertainties, corresponding to a valid error estimation and an unbiased fit, respectively. For fitted $\Xi_b^0 \rightarrow D^0 \Lambda$ yields though, the results are larger by an offset of 0.29(14) on average, which is less than 5% of the nominal value. In a first order approximation we correct for this bias by subtracting fitted yields with the offset. The corrected signal yields are listed in Tab. 8.2.

8.3 Validation of Yield Significances with Pseudo-Experiments

When calculating the signal yield significances we made use of Wilks theorem [108] that states, as the sample size tends to infinity, twice the log-likelihood ratio

$$2 \log \frac{\mathcal{L}_1}{\mathcal{L}_2} \equiv 2 \Delta \ln \mathcal{L},$$

tends to a χ^2 -distribution with k DoF, where k is the absolute difference of the DoF of \mathcal{L}_1 and \mathcal{L}_2 . This approximation works best if all k DoF are uncorrelated (trivially given for $k = 1$), but quickly becomes worse otherwise. In order to verify the quality of the approximation we again use a pseudo-experiment. We generate three sample sets where the first consists of instances drawn from the full fitted PDF, and the latter two are drawn from a modified PDF where f_s is set to zero and one to disable the Λ_b and Ξ_b^0 signal component, respectively, and f_1 is adjusted to retain the same ratio between the remaining signal mode and the combinatorial background \mathcal{B} (*cf.* Appx. P and in particular Fig. P.1 and Fig. P.11 for more details). In total we generate 1000 samples for each set and track type, and each sample consists of 50 (70) instances which corresponds to the amount of events of LL (DD) tracks within the fit range. Each sample is fitted (in configuration 2, *cf.* Sec. 8.1) and the log-likelihood ratios are calculated. The resulting distribution of twice these ratios is shown in Fig. 8.3, as well as the expected (clipped) χ^2 -distribution as derived in Appx. R. The distributions appear to be in

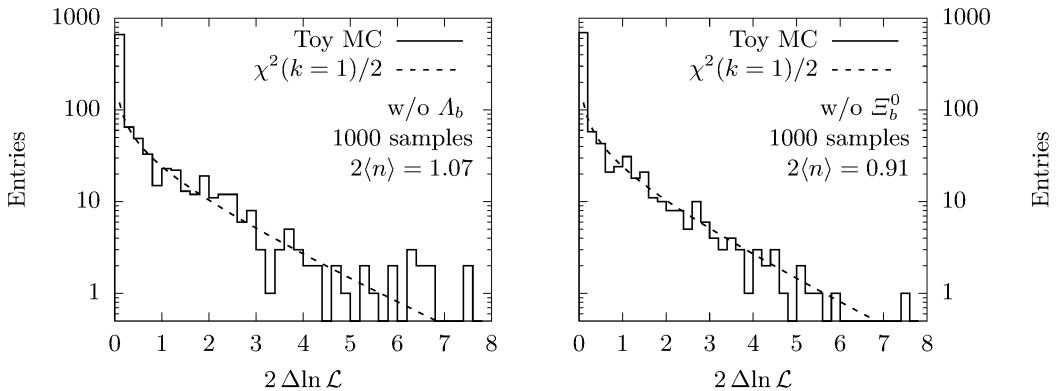


Fig. 8.3. Distribution of twice the log-likelihood ratio for 1000 samples of a pseudo-experiment (Toy MC) used for validating the estimated Λ_b (left) and Ξ_b^0 (right) yield significances. The samples are generated under the null hypothesis (no signal) and can thus benchmark our actual observation of twice the log-likelihood ratio of roughly 31 (left) and 3 (right). The validity is based on Wilks theorem and is ensured, when the distribution follows a (clipped) χ^2 -distribution (dashed line).

good agreement with the expected (clipped) χ^2 -distributions and thus validate the estimated yield significance.

Chapter 9

Estimation of Branching Ratios

If enough data is collected, anything may be proven by statistical methods.

— Williams and Holland’s Law, from Arthur Bloch’s book *Murphy’s Law*.

With the available dataset and the established fit model we transform the measured yields of $\Lambda_b \rightarrow D^0\Lambda$ and $\Xi_b^0 \rightarrow D^0\Lambda$ decays into branching ratios. In Sec. 9.1 we determine the branching fraction of $\Lambda_b \rightarrow D^0\Lambda$ w.r.t. the three-body decay $\Lambda_b \rightarrow D^0 p\pi^-$ and discuss systematic uncertainties. In Sec. 9.2 we then determine the branching ratio of the Ξ_b^0 decay w.r.t. its Λ_b counterpart which can be directly extracted from the fit. Since the Ξ_b^0 yield has a low significance we estimate two-sided confidence intervals and upper limits for this ratio.

9.1 Branching Ratio $\mathcal{B}(\Lambda_b \rightarrow D^0\Lambda)/\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)$

The data sample used for extracting the signal significances in the previous chapter was an admixture of L0 TIS and L0 TOS triggered events. Especially for the former, MC techniques cannot simulate the efficiency reliably. For the given dataset the amount of events with a negative L0 TIS trigger decision (*i.e.*, only L0 TOS trigger is set) is very low (*cf.* Fig. P.2) and only four such events are found in the signal bin ($5.60 < m(D^0\Lambda) \leq 5.64 \text{ GeV}/c^2$). We argue that excluding those four events by requiring a positive L0 TIS trigger decision does not significantly impact the overall significance, but renders dedicated studies of L0 TOS trigger decisions, that would come with their own set of statistical and systematic uncertainties, needless. In contrast to L0 TOS triggered events, the L0 TIS efficiency cancels in good approximation in ratios of decays with common final states, such as $\Lambda_b \rightarrow D^0\Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$. The quality of the approximation even improves, if contributions of non-trivial correlations among the b -hadrons, that were produced during the pp interaction, cancel in the ratio because they are of the same type (*e.g.*, both are Λ_b baryons). Since both conditions are met for $\Lambda_b \rightarrow D^0\Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$, we discard events with a negative L0 TIS trigger decision for the estimation of the branching fraction.

Applying the fit in configuration 1 to the reduced data sample yields

$$n_{\text{TIS}}(\Lambda_b \rightarrow D^0\Lambda) = 28 \pm 7 = \begin{cases} 14 \pm 5 & \text{LL}, \\ 14 \pm 5 & \text{DD}. \end{cases}$$

The projections of this fit for LL and DD tracks are shown in Fig. 9.1.

The branching ratio is determined by correcting the ratio of the fitted yields by their combined reconstruction and stripping efficiency $\varepsilon_{\text{rec \& strip}}$ (*cf.* Chap. 3), the efficiency of the tight selection ε_{sel} and the branching fraction of the $\Lambda \rightarrow p\pi^-$ decay:

$$\begin{aligned} \frac{\mathcal{B}(\Lambda_b \rightarrow D^0\Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)} &= \frac{n_{\text{TIS}}(\Lambda_b \rightarrow D^0\Lambda)}{n_{\text{TIS}}(\Lambda_b \rightarrow D^0 p\pi^-)} \times \frac{\varepsilon_{\text{rec \& strip}}(\Lambda_b \rightarrow D^0 p\pi^-)}{\varepsilon_{\text{rec \& strip}}(\Lambda_b \rightarrow D^0\Lambda)} \\ &\times \frac{\varepsilon_{\text{sel}}(\Lambda_b \rightarrow D^0 p\pi^-)}{\varepsilon_{\text{sel}}(\Lambda_b \rightarrow D^0\Lambda)} \times \frac{1}{\mathcal{B}(\Lambda \rightarrow p\pi^-)} = \begin{cases} 0.015 \pm 0.005 \pm 0.003 & (\text{LL}), \\ 0.017 \pm 0.005 \pm 0.004 & (\text{DD}), \end{cases} \end{aligned}$$

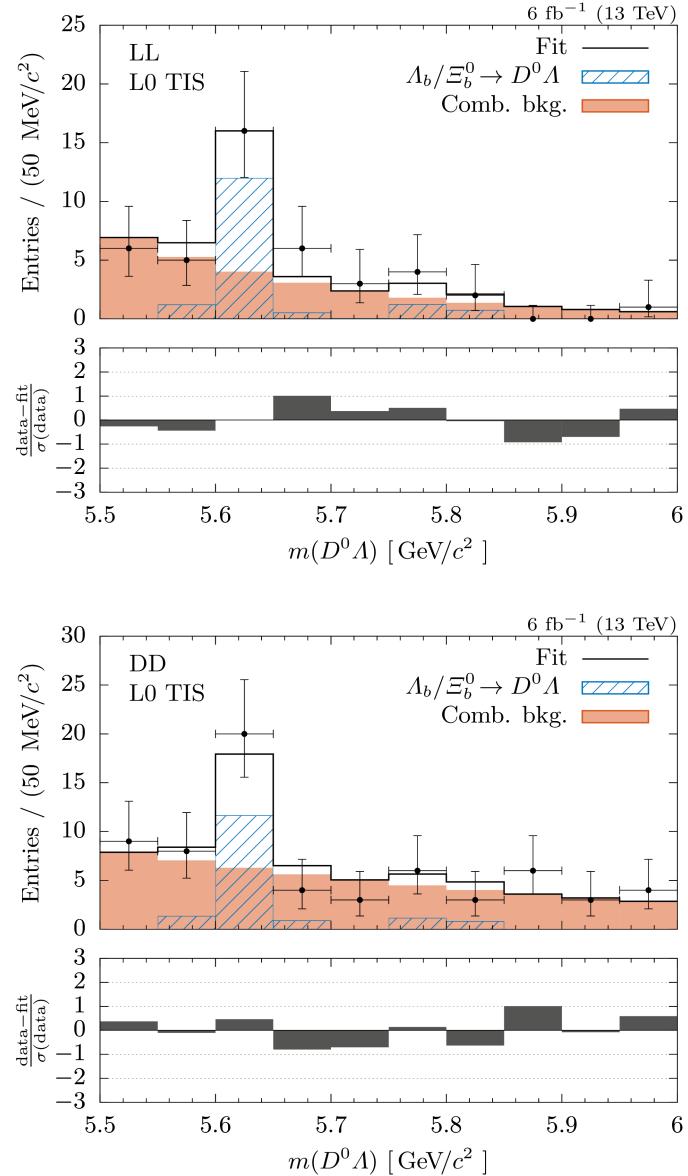


Fig. 9.1. Combined invariant mass of D^0 and Λ candidates of track type LL (top) and DD (bottom) from recorded data, as well as the corresponding projections of the simultaneous fit in configuration 1. For recorded data a positive L0 TIS trigger decision is required.

where the first (second) error is the total statistical (systematic) uncertainty as the result of an ordinary error propagation. The components of this propagation are listed in Tab. 9.1 and are discussed below:

$\Lambda_b \rightarrow D^0 \Lambda/p\pi^-$ fit Fitted yields of L0 TIS triggered $\Lambda_b \rightarrow D^0 \Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$ events.

The statistical error of the former is dominated by the multinomial part of the fit model.

Different variations of this fit model (*cf.* Sec. 8.1) did not reveal major deviations and we therefore approximate the systematic uncertainty being less than 1 : 14 (< 10 %). The total error of the latter is dominated by the systematic uncertainty from the respective fit model (*cf.* Sec. 6.2).

Rec. \times strip. ratio Ratio of the combined reconstruction and stripping efficiency as derived in Chap. 3. The figures are given relative to the total amount of generated $\Lambda_b \rightarrow D^0 \Lambda$ decays and thus also account for the splitting ratio of $\Lambda \rightarrow p\pi^-$ into LL and DD tracks. The assumed systematic uncertainty of 10 % is discussed in Chap. 3. We note that the statistical uncertainty is correlated among $\Lambda_b \rightarrow D^0 \Lambda$ decays of both track types since both of them use the same sample of $\Lambda_b \rightarrow D^0 p\pi^-$ candidates for normalization. However, the magnitude of the statistical uncertainties is small compared to the other contributions. Thus they can be neglected.

$\Lambda_b \rightarrow D^0 \Lambda/p\pi^-$ tight sel. Efficiency of the tight selection and $\Lambda_b \rightarrow D^0 p\pi^-$ veto for the former, and efficiency of a subset of the tight selection for the latter, estimated with MC simulated events. In order to avoid a potential bias, we use the subset of the available MC simulated $\Lambda_b \rightarrow D^0 \Lambda$ decays not used during training (and hyper-parameters optimization) of the classifier at the cost of a slightly larger statistical uncertainty (*cf.* Sec. 5.2). This restriction does not affect the efficiency estimation with MC simulated $\Lambda_b \rightarrow D^0 p\pi^-$ decays since this sample was not involved in the training procedure of the classifiers. The origin of the systematic uncertainties is discussed in Appx. M and considered a conservative approximation, since parts might cancel in the branching ratio due to the common set of thresholds. The large amount of available MC simulated $\Lambda_b \rightarrow D^0 p\pi^-$ decays yields a small systematic uncertainty which can be neglected (including its correlation) in the context of the full error analysis.

MC calibration In Chap. 4 we established two different calibration schemes. We use Scheme 2 for the calibration of MC simulated events and take the deviation to Scheme 1 as a systematic uncertainty.

The results for LL and DD tracks can be combined and averaged to improve the overall significance

$$\frac{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)} = 0.016 \pm 0.004 \pm 0.003.$$

Combining the ratio with the measured branching fraction

$$\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-) = (6.3 \pm 0.7) \times 10^{-4},$$

as reported by the PDG, yields

$$\frac{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)} \times \mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-) = (9.9 \pm \underbrace{2.3 \pm 1.6 \pm 1.1}_{3.0}) \times 10^{-6}, \quad (9.1)$$

where the first (second) uncertainty is statistical (systematic) and the third error is the (external) uncertainty of the normalization. The total uncertainty is 3.0×10^{-6} . Our result thus is compatible with the theory prediction from Ref. [49]

$$\mathcal{B}_{\text{pred}}(\Lambda_b \rightarrow D^0 \Lambda) = 4.56 \times 10^{-6}$$

Tab. 9.1. Nominal values (n), as well as statistical (u_{stat}) and systematic (u_{sys}) uncertainties involved in the estimation of the branching ratio of $\Lambda_b \rightarrow D^0 \Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$ for both track types. Uncertainties which are (partially) correlated among the estimated branching ratios of LL and DD tracks are marked with \dagger .

	LL			DD		
	n	u_{stat}	u_{sys}	n	u_{stat}	u_{sys}
$\Lambda_b \rightarrow D^0 \Lambda$ fit (L0 TIS)	14	31 %	< 10 %	14	33 %	< 10 %
$\Lambda_b \rightarrow D^0 p\pi^-$ fit (L0 TIS)	39 300	< 1 %	7 %	36 000	< 1 %	7 %
Rec. \times strip. ratio \dagger	17.04	< 1 %	10 %	7.26	< 1 %	10 %
$\Lambda_b \rightarrow D^0 \Lambda$ tight sel.	0.348	2.9 %	7 %	0.136	6.6 %	19 %
$\Lambda_b \rightarrow D^0 p\pi^-$ tight sel. \dagger	0.5476	< 1 %	4.3 %	0.5098	< 1 %	4.9 %
MC calibration	—	—	1.9 %	—	—	5.5 %
$1/\mathcal{B}(\Lambda \rightarrow p\pi^-)^\dagger$ [15]	1.565	—	< 1 %	1.565	—	< 1 %
		31 %	18 %		33 %	26 %

within two standard deviations. We note that the measurement of $\mathcal{B}(\Lambda_b \rightarrow D^0 p\pi^-)$, although dominated by an LHCb measurement, is not correlated with our measurement since the LHCb analysis uses data recorded during Run 1 rather than Run 2 as we do in the present analysis [15, 69].

9.2 Branching Ratio $\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)/\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)$

The constraint fit to the combined invariant mass of D^0 and Λ candidates of both track types allows a clean extraction of the branching ratio

$$\frac{\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)} = \frac{f_{\Lambda_b}}{f_{\Xi_b^0}} \times \frac{1 - f_s}{f_s}, \quad (9.2)$$

where $f_{\Lambda_b}/f_{\Xi_b^0}$ is the ratio of the fragmentation fractions of b -quarks into Λ_b and Ξ_b^0 baryons, and f_s is a fit parameter that we established in Sec. 8.1. (Not to be confused with the fragmentation fractions of b -quarks into B_s^0 mesons.) The systematic uncertainty of this technique is very low in comparison to the large statistical uncertainty. Fidelity issues in MC simulated events play a minor role and enter only if they induce a non-trivial correction to the ratio Eq. (9.2) which is disfavored by the MC simulated $\Lambda_b \rightarrow D^0 \Lambda$ and $\Xi_b^0 \rightarrow D^0 \Lambda$ decays (*cf.* Tab. 8.1). In particular we find that the nominal value of the branching ratio is largely independent of our choice of the MC calibration.

We estimate two confidence intervals (CI) following a frequentist interpretation by using pseudo-experiments, and two CIs using a Bayesian interpretation by scanning the fitted likelihood profiles of f_s and assuming a uniformly distributed prior $(1 - f_s)/f_s$. All these intervals are shown in Fig. 9.2 and Tab. 9.2 with 68 % and 90 % CL. (We note that the upper boundary of the two-sided *shortest* CIs (in both interpretations) with a 90 % CL are equivalent to 95 % CL upper limits.) We also estimate a Bayesian upper limit when requiring a minimal flight distance significance of D^0 candidates to suppress charmless Ξ_b^0 backgrounds (*cf.* Sec. 7.3.1) and append the result to Tab. 9.2. In Appx. S we give a more detailed overview about the estimation of these CIs.

The measurements of absolute branching fractions are difficult to perform at hadron colliders without using external input. In particular, measurements of absolute Ξ_b branching fractions are limited by the amount of available data, thus it is still common to report branching ratios

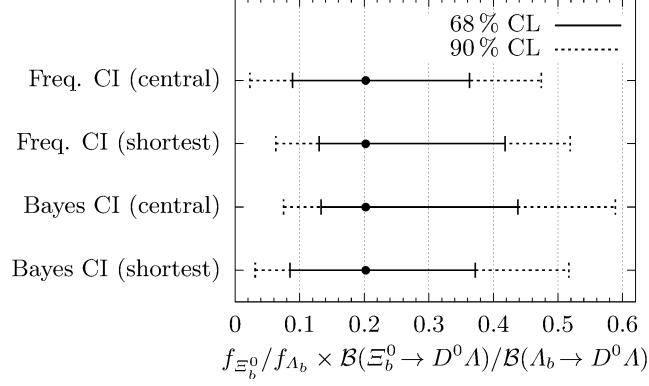


Fig. 9.2. Two-sided CIs with 68 % and 90 % coverage for $(1 - f_s)/f_s$, estimated using different frequentist and Bayesian approaches. We note that the upper boundaries of the (two-sided) central CIs with 90 % CL are identical with 95 % CL upper limits.

Tab. 9.2. Two-sided CIs with 68 % and 90 % coverage, as well as 95 % CL upper limits for $(1 - f_s)/f_s$, estimated using different frequentist and Bayesian approaches. The limits shown in the last row (\dagger) are the results of a fit when an additional veto against charmless Ξ_b^0 backgrounds is required.

Method	68 % CL	90 % CL	95 % CL
Freq. CI (central)	[0.089 ... 0.363]	[0.023 ... 0.474]	[0 ... 0.474]
Freq. CI (shortest)	[0.130 ... 0.418]	[0.063 ... 0.519]	—
Bayes CI (central)	[0.133 ... 0.438]	[0.075 ... 0.589]	[0 ... 0.589]
Bayes CI (shortest)	[0.085 ... 0.372]	[0.031 ... 0.517]	—
Bayes CI (upper limit \dagger)	—	[0 ... 0.437]	[0 ... 0.537]

involving Ξ_b baryons in product with the respective b -fragmentation ratios. Lately, first indirect measurements of $f_{\Xi_b^0}/f_{\Lambda_b}$ were carried out by the LHCb collaboration [21] and leveraged the predictions $f_{\Xi_b^0}/f_{\Lambda_b} = 0.065(20)$ [109] and $f_{\Xi_b^0}/f_{\Lambda_b} = 0.054(20)$ [110]. Still, it is known from precise measurements of b -fragmentations into different b -mesons and Λ_b baryons that these values depend on transverse momentum and pseudo-rapidity [96, 97]. A first measurement of $f_{\Xi_b^-}/f_{\Lambda_b}$ at different center-of-mass energies confirms this dependence for Ξ_b baryons [111]. Since measurements and predictions of $f_{\Xi_b^0}/f_{\Lambda_b}$ are still rare and kinematic dependencies are not yet included in the predictions, we stick to the common practice and only report the product of the branching fraction $\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)/\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)$ and the fragmentation ratio $f_{\Xi_b^0}/f_{\Lambda_b}$.

9.3 Summary and Outlook

Using the full available Run 2 data set we find the decay $\Lambda_b \rightarrow D^0 \Lambda$ with a statistical significance of 5.5 standard deviations and estimate the branching ratio

$$\frac{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 p \pi^-)} = 0.016 \pm 0.004 \pm 0.003, \quad (9.3)$$

where the first uncertainty is statistical and the second is systematic. An excess of $\Xi_b^0 \rightarrow D^0 \Lambda$ candidates is observed with a statistical significance of 1.8 standard deviations and used to estimate upper limits, *e.g.*,

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0 \Lambda)} < 0.5 \quad (\text{CL} = 95\%). \quad (9.4)$$

Let us recap how these results were obtained: In Chap. 4 we calibrated MC simulated events and used them together with recorded data to train binary classifiers in Chap. 5 to separate signal and combinatorial background. Contributions from physical backgrounds were analyzed in Chap. 7 and were either found negligible after requiring dedicated selections or were included in the fit model described in Chap. 8. Studies of the normalization mode were presented in Chap. 6.

The results reported in Eq. (9.1), Eq. (9.3), and Eq. (9.4) are limited statistically, even though the intermediate particles are reconstructed in their dominant modes, *i.e.*, $D^0 \rightarrow K^- \pi^+$ and $\Lambda \rightarrow p \pi^-$. The inclusion of more decay modes, for example $D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$, can increase the significances slightly and allow CP measurements. However, to actually measure CKM parameters, at least the Cabibbo suppressed decays $D^0 \rightarrow K^- K^+$ and $D^0 \rightarrow \pi^- \pi^+$ have to be reconstructed, corresponding to the requirement of roughly $1/\lambda^2 \approx 20$ times more data. This is in reach within the next runs of the LHC. Additionally, charmless Λ_b decays will enter into these modes as non-resonant background and require rich statistics to control, for example by analyzing the lifetime of D^0 candidates.

A sufficiently large data sample which would allow the reconstruction of the Cabibbo doubly suppressed modes $D^0 \rightarrow K^+ \pi^-$, would not just leverages an extraction of the CKM parameter γ using ADS and GLW methods in decays of baryons, but would also allow a clean extraction of $\mathcal{B}(\Xi_b^0 \rightarrow D^0 \Lambda)$ due to the absence of charmless backgrounds in this mode. On top, an extraction of Σ^0 modes (both in Λ_b and Ξ_b^0) similar to Ref. [26] would become feasible and thus would allow inference of yet unseen modes such as $\Xi_b^0 \rightarrow D^0 \Xi^0$ or $\Lambda_b \rightarrow D^0 n$.

Glossary

- $\Delta\chi^2_{\text{IP}}$** Difference between the χ^2 value of the PV reconstructed with and without the track under consideration. 20, 47–50, 79, 120–122
- ADS** Method for extracting the CKM angle γ in decays such as $A_b \rightarrow D\Lambda$, first proposed by Atwood, Dunietz and Soni [50, 51], *cf.* Sec. 1.5. 13, 14, 86, 102
- Babar** Collaboration and HEP experiment at PEP-II. 88, 91
- BDT** (Gradient) *Boosted Decision Tree*, supervised (machine) learning model. 45, 55, 56, 59, 60, 69, 145
- Belle** Collaboration and HEP experiment at the KEKB accelerator. 89
- BEPC II** Circular e^+e^- Collider, operating at $\sqrt{s} = 2$ to 4.63 GeV. 103
- BESIII** Collaboration and HEP experiment at the Beijing e^+e^- Collider II (BEPC II). 9, 11, 91
- CKM** Unitary matrix which contains information on the magnitudes and complex phases for flavor-changing weak interactions. 2–5, 13, 78, 88, 89, 102–104, 106
- DD** Both final state particles of a V^0 decay (*e.g.*, $\Lambda \rightarrow p\pi^-$ or $K_s \rightarrow \pi^+\pi^-$) are reconstructed as *downstream tracks*, *cf.* Sec. 2.1.1 for more details. 16, 23–26, 29, 30, 33–42, 47, 49, 51, 53–62, 64, 66, 68–72, 75–77, 80–82, 87, 88, 92, 93, 95–100, 117, 118, 120, 121, 127, 128, 133–135, 148, 150, 151, 155–159, 163, 164, 173–177, 179, 180
- DIRA** *Direction angle*, cosine of the angle between the momentum of the particle and the direction vector from some reference vertex or 3D-point to the end-vertex of the particle. 20, 50, 120–122
- DLL** *Delta log-likelihood*. 19, 20, 120–122
- DOCA** *Distance of closest approach*. 21, 120–122
- DoF** *Degrees of Freedom*. 20, 28–30, 32, 35, 40, 41, 47–50, 79, 93, 96, 120–124, 127, 129, 133, 134, 158, 178, 185
- DTF** *Decay tree fit*. 21, 29, 30, 46–50, 70, 71, 76, 79, 80, 84, 85, 87, 123–125, 127, 128, 155–159, 163, 165
- FN** *False Negative*, *cf.* Appx. G. 54, 137
- FNR** *False Negative Rate*, *cf.* Appx. G. 70
- FoM** *Figure of merit*. 29, 30, 32, 73, 75
- FP** *False Positive*, *cf.* Appx. G. 137

FPR *False Positive Rate*, cf. Appx. G. 70, 73, 75

Ghost Prob. Output of an algorithm to identify tracks which do not correspond to the trajectory of a (single) true particle but rather originate from detector noise or multiple particles due to mismatching. 19, 21, 121, 122

GLW Method for extracting the CKM angle γ in decays such as $A_b \rightarrow D\Lambda$, first proposed by Gronau, London and Wyler [52, 53], cf. Sec. 1.5. 13, 14, 86, 102

HEP *High Energy Physics*. 15, 103, 155

HLT *High level trigger*, includes a full off-line reconstruction. 18, 19, 46, 78

KEKB Circular e^+e^- collider, operating at $\sqrt{s} = m(\Upsilon(4S)) = 10.57 \text{ GeV}/c^2$. 103

L0 Level 0 trigger of the LHCb trigger system, implemented in hardware, followed by the high level triggers. 17–19, 27–29, 33, 81, 82, 97–100, 136, 174

LL Both final state particles of a V^0 decay (*e.g.*, $\Lambda \rightarrow p\pi^-$ or $K_s \rightarrow \pi^+\pi^-$) are reconstructed as *long tracks*, cf. Sec. 2.1.1 for more details. 16, 23–26, 29, 30, 33–42, 47, 48, 51, 53–63, 65, 67–71, 75–77, 80–83, 87, 88, 92, 93, 95–100, 117, 118, 120, 121, 127, 128, 133–135, 147, 149, 151, 156–159, 163, 164, 173–177, 179, 180

MC *Monte Carlo*, refers to the used simulation strategy. The pp collisions are generated using PYTHIA [112] with a specific LHCb configuration [113]. Decays of unstable particles are described by EvtGen [114], where final-state radiations is generated using PHOTOS [115]. The interaction of the particles with the detector (and its response) are implemented using the Geant4 toolkit [116, 117] as described in Ref. [118]. 19, 23–25, 27, 28, 30, 33–35, 37, 39, 40, 43, 45–51, 54, 67, 70, 77–81, 84, 86–89, 91–93, 97, 99, 100, 102, 133, 134, 153, 155, 156, 163–165, 175, 176

MVA *Multi Variate Analysis*. 45, 46, 50, 51, 55, 152, 153

PCA *Principal Component Analysis*, unsupervised (machine) learning technique, cf. Appdx. I. 51, 56, 60, 63–67, 141, 149, 150

PCC *Pearson correlation coefficient*, measure of the linear correlation between two variables. For a given sample of paired data $\{(x_1, y_1), \dots, (x_n, y_n)\}$ the PCC is given by the estimates of the covariance and variances:

$$\text{PCC} = \frac{\sum_{i=0}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^n (y_i - \bar{y})^2}},$$

where \bar{x} (\bar{y}) is the sample mean of x_i (y_i). The denominator normalizes this expression, according to the Cauchy-Schwarz inequality, to the interval $[-1, +1]$, where $+1$ indicates a total positive linear correlation, 0 no linear correlation and -1 a total negative linear correlation. Higher order correlations are not reflected in the PCC. 68, 69, 147–150

PDF *Probability Density Function*. 80, 95, 96, 174, 187, 188

PDG *Particle Data Group* [15], collaboration of particle physicists that compiles and reanalyzes published results. 7, 21, 25, 77, 91, 94, 99, 120, 121, 161

PEP-II Circular asymmetric-energy e^+e^- collider, operating at $\sqrt{s} = 10.58 \text{ GeV}/c^2$. 103

PID *Particle identification*. 19, 20, 29, 47, 50, 70

PV *Primary Vertex*, origin vertex of first particle in most decay chains, typically created by pp -scattering close to the beam-pipe. 16, 20, 21, 28, 29, 46–50, 78, 79, 103, 120–124

reflection Background contribution due to misidentification of at least one particle. 51, 79, 80, 83, 86, 88, 162

RICH *Ring Imaging Cherenkov*, system of the LHCb detector providing identification of charged particles. 17, 19, 20, 29

ROC *Receiver Operating Characteristic* (curve), graphical plot of TPR vs. TNR for varying threshold. 60, 62, 68, 70, 71, 73–75, 151

ROC-AUC *Area Under ROC Curve*, probability that a given classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. 56–61, 67–71, 151

Run 1 First major periode of data taking at the LHC, corresponding to $\mathcal{L} \approx 1 \text{ fb}^{-1}$ at $\sqrt{s} = 7 \text{ TeV}$ (2011) and $\mathcal{L} \approx 2 \text{ fb}^{-1}$ at $\sqrt{s} = 8 \text{ TeV}$ (2012). 15, 19, 45, 83, 100

Run 2 Second major periode of data taking at the LHC, corresponding to $\mathcal{L} \approx 6 \text{ fb}^{-1}$ and $\sqrt{s} = 13 \text{ TeV}$. 15, 17, 19, 23, 28, 45, 46, 78, 83, 91, 100, 102, 134, 135

stripping Set of selection criteria applied to data directly after data taking. 23–25, 28, 46, 78, 83, 119

SVM *Support-Vector Machine*, supervised (machine) learning model. 45, 46, 50, 51, 55–57, 60–62, 67–70, 87, 139, 143, 144, 151

Tevatron Circular $p\bar{p}$ collider at Fermilab operating at a beam energy of 1 TeV. 6

TIS *trigger independent of signal*. 19, 27–29, 33, 81, 82, 97–100, 136, 174

TN *True Negative*, cf. Appx. G. 137

TNR *True Negative Rate*, cf. Appx. G. 70, 74, 105

TOS *trigger on signal*. 19, 97, 136

TP *True Positive*, cf. Appx. G. 54, 137

TPR *True Positive Rate*, cf. Appx. G. 70, 71, 73–76, 105

truth-matched Each simulated track is attached with a label that allows a matching of tracks and simulated decays. If the reconstruction hypothesis of a candidate matches the *true* generated one (including all tracks), the given candidate is referred to as *truth-matched*. 23, 30–32, 34, 35, 46, 48–51, 75, 86, 105, 127, 128, 153

truth-matching Process of deciding whether a given simulated event is truth-matched or *unmatched*. 28

VELO *Vertex Locator*, silicon vertex detector surrounding the interaction region. 7, 16, 17, 133, 158

Wolfenstein parameterization Parametrization of the CKM matrix, introduced by Lincoln Wolfenstein with the four real parameters λ , A , ρ and η . 4, 7, 78

Bibliography

- [1] K. Nishijima, “Charge Independence Theory of V Particles”, *Prog. Theo. Phys.* **13**, 285 (1955).
- [2] M. Gell-Mann, “The interpretation of the new particles as displaced charge multiplets”, *Il Nuovo Cimento* **4**, 848 (1956).
- [3] M. Kobayashi *et al.*, “*CP*-Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theo. Phys.* **49**, 652 (1973).
- [4] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10**, 531 (1963).
- [5] L. Wolfenstein, “Parametrization of the Kobayashi-Maskawa Matrix”, *Phys. Rev. Lett.* **51**, 1945 (1983).
- [6] T. Browder *et al.*, “On the physics case of a super flavour factory”, *J. High Energ. Phys.* **02**, 110 (2008).
- [7] M. Ciuchini and A. Stocchi, “Physics Opportunities at the Next Generation of Precision Flavor Physics”, *Annu. Rev. Nucl. Part. Sci.* **61** (2011).
- [8] Super-Kamiokande Collaboration, “Evidence for Oscillation of Atmospheric Neutrinos”, *Phys. Rev. Lett.* **81**, 1562 (1998).
- [9] SNO Collaboration, “Measurement of the Rate of $\nu_e + d \rightarrow ppe^-$ Interactions Produced by 8B Solar Neutrinos at the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **87**, 071301 (2001).
- [10] SNO Collaboration, “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **89**, 011301 (2002).
- [11] Z. Maki *et al.*, “Remarks on the Unified Model of Elementary Particles”, *Prog. Theo. Phys.* **28**, 870 (1962).
- [12] T2K Collaboration, “Constraint on the matter-antimatter symmetry-violating phase in neutrino oscillations”, *Nature* **580**, 339 (2020).
- [13] T. Mannel, “Theory and Phenomenology of *CP* Violation”, *Nucl. Phys.* **B167**, 170 (2007).
- [14] LHCb Collaboration, “Search for the *CP*-violating strong decays $\eta \rightarrow \pi^+ \pi^-$ and $\eta'(958) \rightarrow \pi^+ \pi^-$ ”, *Phys. Lett.* **B764**, 233 (2017).
- [15] Particle Data Group, *Phys. Rev.* **D98**, 030001 (2018).
- [16] CDF Collaboration, “Observation of the Ξ_b^0 Baryon”, *Phys. Rev. Lett.* **107**, 102001 (2011).
- [17] D0 Collaboration, “Direct Observation of the Strange b Baryon Ξ_b^- ”, *Phys. Rev. Lett.* **99**, 052001 (2007).

- [18] D0 Collaboration, “Observation and Mass Measurement of the Baryon Ξ_b^- ”, Phys. Rev. Lett. **99**, 052002 (2007).
- [19] M. Basile *et al.*, “Evidence for a new particle with naked *beauty* and for its associated production in high-energy (pp) interactions”, Lett. Nuovo Cimento **31**, 97 (1981).
- [20] I.I. Bigi, “The QCD Perspective on Lifetimes of Heavy-Flavour Hadrons”, arXiv:hep-ph/9508408 (1995).
- [21] LHCb Collaboration, “Precision Measurement of the Mass and Lifetime of the Ξ_b^0 Baryon”, Phys. Rev. Lett. **113**, 032001 (2014).
- [22] R. Machleidt and H. Müther, “Charge symmetry breaking of the nucleon-nucleon interaction: ρ - ω mixing versus nucleon mass splitting”, Phys. Rev. **C63**, 034005 (2001).
- [23] M. Gell-Mann, “A schematic model of baryons and mesons”, Phys. Lett. **8**, 214 (1964).
- [24] R. Lewis *et al.*, “Charmed baryons in lattice QCD”, Phys. Rev. **D64**, 094509 (2001).
- [25] PACS-CS Collaboration, “Charmed baryons at the physical point in 2 + 1 flavor lattice QCD”, Phys. Rev. **D87**, 094512 (2013).
- [26] LHCb Collaboration, “Isospin Amplitudes in $\Lambda_b^0 \rightarrow J/\psi/\Lambda(\Sigma^0)$ and $\Xi_b^0 \rightarrow J/\psi/\Xi^0(\Lambda)$ Decays”, Phys. Rev. Lett. **124**, 111802 (2020).
- [27] Particle Data Group, Phys. Rev. **D98**, 030001 (2018). Data files for total collision cross sections of $p\pi^-$ and $p\pi^+$, “Hadron(photon)-hadron(photon) collisions”, courtesy of the COMPAS Group (IHEP, Protvino, Russia)
<http://pdg.lbl.gov/2019/hadronic-xsections/>
- [28] C. Daum *et al.*, “Diffractive production of strange mesons at 63 GeV”, Nucl. Phys. **B187**, 1 (1981).
- [29] CLEO Collaboration, “Measurement of Prominent η Decay Branching Fractions”, Phys. Rev. Lett. **99**, 122001 (2007).
- [30] A.M. Blik *et al.*, “Searches for rare and forbidden neutral decays of η mesons at the GAMS-4 π facility”, Phys. At. Nucl. **70**, 693 (2007).
- [31] KLOE Collaboration, “Upper limit on the $\eta \rightarrow \pi^+\pi^-$ branching ratio with the KLOE detector”, Phys. Lett. **B606**, 276 (2005).
- [32] Private communication with G. Isidori (Oct. 2019).
- [33] J. Colas *et al.*, “Search for Electromagnetic Decays of Neutral Hyperons in the 1210-1500 MeV Mass Range and of Neutral Mesons in the 150-600 MeV Mass Range”, Nucl. Phys. **B91**, 253 (1975).
- [34] G. Feinberg, “Internal Pair Creation in Σ^0 Decay”, Phys. Rev. **109**, 1019 (1958).
- [35] BESIII Collaboration, “Study of J/ψ and $\psi(3686)$ decay to $\Lambda\bar{\Lambda}$ and $\Sigma^0\bar{\Sigma}^0$ final states”, Phys. Rev. **D95**, 052003 (2017).
- [36] J. Kambor *et al.*, “ $K \rightarrow 2\pi$ and $K \rightarrow 3\pi$ decays in next-to-leading order chiral perturbation theory”, Phys. Lett. **B261**, 496 (1991).
- [37] LHCb Collaboration, “Evidence for the two-body charmless baryonic decay $B^+ \rightarrow p\bar{\Lambda}$ ”, J. High Energ. Phys. **04**, 162 (2017).

- [38] ARGUS Collaboration, “Evidence for W-exchange in charmed baryon decays”, Phys. Lett. **B342**, 397 (1995).
- [39] LHCb Collaboration, “Observations of $\Lambda_b^0 \rightarrow \Lambda K^+ \pi^-$ and $\Lambda_b^0 \rightarrow \Lambda K^+ K^-$ decays and searches for other Λ_b^0 and Ξ_b^0 decays to $\Lambda h^+ h^-$ final states”, J. High Energ. Phys. **05**, 81 (2016).
- [40] LHCb Collaboration, “Observation of the $\Lambda_b^0 \rightarrow \Lambda \phi$ decay”, Phys. Lett. **B759**, 282 (2016).
- [41] LHCb Collaboration, “Measurement of the ratio of branching fractions of the decays $\Lambda_b^0 \rightarrow \psi(2S) \Lambda$ and $\Lambda_b^0 \rightarrow J/\psi \Lambda$ ”, J. High Energ. Phys. **03**, 126 (2019).
- [42] BESIII Collaboration, “Measurements of absolute hadronic branching fractions of Λ_c^+ baryon”, Phys. Rev. Lett. **116**, 052001 (2016).
- [43] Private communication with A. Dery, M. Ghosh, Y. Grossman and S. Schacht (Jan. 2020).
- [44] LHCb Collaboration, “Measurement of the CP Asymmetry in B_s^0 - \bar{B}_s^0 Mixing”, Phys. Rev. Lett. **117**, 061803 (2016).
- [45] LHCb Collaboration, “Measurement of CP asymmetries in two-body $B_{(s)}^0$ -meson decays to charged pions and kaons”, Phys. Rev. **D98**, 032004 (2018).
- [46] LHCb Collaboration, “Measurement of matter-antimatter differences in beauty baryon decays”, Nature Phys. **13**, 391 (2017).
- [47] LHCb Collaboration, “Search for CP violation in $\Lambda_b^0 \rightarrow p K^-$ and $\Lambda_b^0 \rightarrow p \pi^-$ decays”, Phys. Lett. **B787**, 124 (2018).
- [48] I. Dunietz, “ CP violation with beautiful baryons”, Z. Phys. **C56**, 129 (1992).
- [49] A.K. Giri *et al.*, “Possibility of extracting the weak phase γ from $\Lambda_b \rightarrow \Lambda D^0$ decays”, Phys. Rev. **D65**, 073029 (2002).
- [50] D. Atwood *et al.*, “Enhanced CP Violation with $B \rightarrow K D^0 (\bar{D}^0)$ Modes and Extraction of the Cabibbo-Kobayashi-Maskawa Angle γ ”, Phys. Rev. Lett. **78**, 3257 (1997).
- [51] D. Atwood *et al.*, “Improved methods for observing CP violation in $B^\pm \rightarrow K D$ and measuring the CKM angle γ ”, Phys. Rev. **D63**, 036005 (2001).
- [52] M. Gronau *et al.*, “On determining a weak phase from charged B decay asymmetries”, Phys. Lett. **B265**, 172 (1991).
- [53] M. Gronau *et al.*, “How to determine all the angles of the unitarity triangle from $B_d^0 \rightarrow D K_S$ and $B_s^0 \rightarrow D \phi$ ”, Phys. Lett. **B253**, 483 (1991).
- [54] LHCb Collaboration, “Measurement of CP observables in the process $B^0 \rightarrow D K^{*0}$ with two- and four-body D decays”, J. High Energ. Phys. **08**, 041 (2019).
- [55] X.H. Guo and A.W. Thomas, “Direct CP violation in $\Lambda_b \rightarrow n(\Lambda) \pi^+ \pi^-$ decays via ρ - ω mixing”, Phys. Rev. **D58**, 096013 (1998).
- [56] S. Arunagiri and C.Q. Geng, “ T -violating triple product asymmetries in $\Lambda_b \rightarrow \Lambda \pi^+ \pi^-$ decay”, Phys. Rev. **D69**, 017901 (2004).
- [57] J.-P. Blaizot *et al.*, “Study of potentially dangerous events during heavy-ion collisions at the LHC: report of the LHC safety study group”, CERN-2003-001 (2003).

- [58] J. Ellis *et al.* (LHC Safety Assessment Group), “Review of the safety of LHC collisions”, *J. Phys. G. Nucl. Partic.* **35**, 11 (2008).
- [59] LHCb Collaboration, “The LHCb Detector at the LHC”, *J. Instrum.* **3**, S08005 (2008).
- [60] LHCb Collaboration, “LHCb Detector Performance”, *Int. J. Mod. Phys. A* **30**, 1530022 (2015).
- [61] LHCb Collaboration, “The LHCb Detector at the LHC”, *J. Instrum.* **3** (2008).
- [62] LHCb Collaboration, “The LHCb trigger and its performance in 2011”, *J. Instrum.* **8** (2013).
- [63] J. Albrecht *et al.* (LHCb HLT project), “Performance of the LHCb High Level Trigger in 2012”, *J. Phys. Conf. Ser.* **513**, 012001 (2014).
- [64] LHCb Collaboration, “Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC”, *J. Instrum.* **14** (2019).
- [65] D. Derkach *et al.* (on behalf of the LHCb Collaboration), “Machine-Learning-based global particle-identification algorithms at the LHCb experiment”, *J. Phys. Conf. Ser.* **1085** (2018).
- [66] A. Kendall and Y. Gal, “What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?”, arXiv:1703.04977 (2017).
- [67] K. Shridhar *et al.*, “A Comprehensive guide to Bayesian Convolutional Neural Network with Variational Inference”, arXiv:1901.02731 (2019).
- [68] M.D. Cian *et al.*, “Fast neural-net based fake track rejection in the LHCb reconstruction”, LHCb-PUB-2017-011 (2017).
- [69] LHCb Collaboration, “Studies of beauty baryon decays to D^0ph^- and $\Lambda_c^+h^-$ final states”, *Phys. Rev. D* **89**, 032001 (2014).
- [70] LHCb Collaboration, “Measurement of the $\Lambda_b \rightarrow J/\psi \Lambda$ decay amplitudes and the Λ_b polarisation in pp collisions at $\sqrt{s} = 7\text{TeV}$ ”, *Phys. Lett. B* **724**, 27 (2013).
- [71] UA1 Collaboration, “First observation of the beauty baryon Λ_b in the decay channel $\Lambda_b \rightarrow J/\psi \Lambda$ at the CERN proton-antiproton collider”, *Phys. Lett. B* **273**, 540 (1991).
- [72] D0 Collaboration, “Measurement of the production fraction times branching fraction $f(b \rightarrow \Lambda_b \cdot \mathcal{B}(\Lambda_b \rightarrow J/\psi \Lambda))$ ”, *Phys. Rev. D* **84**, 031102 (2011).
- [73] CDF Collaboration, “Observation of $\Lambda_b \rightarrow J/\psi \Lambda$ at the Fermilab proton-antiproton collider”, *Phys. Rev. D* **55**, 1142 (1997).
- [74] Private communication with H. Viemann (Aug. 2019) about his (at that time) ongoing analysis of the decay $\Lambda_b \rightarrow D^0 p K^-$.
- [75] P. Baldi *et al.*, “Searching for exotic particles in high-energy physics with deep learning”, *Nat. Commun.* **5**, 4308 (2014).
- [76] G. Kasieczka *et al.*, “Deep-learning top taggers or the end of QCD?”, *J. High Energ. Phys.* **05**, 6 (2017).
- [77] G. Kasieczka *et al.*, “The Machine Learning landscape of top taggers”, *SciPost Phys.* **7**, 014 (2019).

- [78] LHCb Collaboration, “Search for Lepton-Flavor Violating Decays $B^+ \rightarrow K^+ \mu^\pm e^\mp$ ”, Phys. Rev. Lett. **123**, 241802 (2019).
- [79] M. Williams *et al.*, “The HTL2 Topological Lines”, LHCb-PUB-2011-002 (2011).
- [80] R.A. Bauder and T.M. Khoshgoftaar, “The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data”, Health Inf. Sci. Syst. **6**, 9 (2018).
- [81] B. Krawczyk, “Learning from imbalanced data: open challenges and future directions”, Prog. Artif. Intell. **5**, 221 (2016).
- [82] Y. Yao *et al.*, “On Early Stopping in Gradient Descent Learning”, Constr. Approx. **26**, 289 (2007).
- [83] L. Prechelt, “Early Stopping – But When?”, ISBN:978-3642352898 (2012).
- [84] W. Jiang, “Process consistency for AdaBoost”, Ann. Statist. **32**, 13 (2004).
- [85] T. Zhang and B. Yu, “Boosting with early stopping: Convergence and consistency”, Ann. Statist. **33**, 1538 (2005).
- [86] T.V. Bandos *et al.*, “Statistical criteria for early-stopping of support vector machines”, Neurocomputing **70**, 13 (2007).
- [87] M. Banko and E. Brill, “Scaling to Very Very Large Corpora for Natural Language Disambiguation”, Proceedings of ACL 2001, 26 (2001).
- [88] A. Halvey *et al.*, “The Unreasonable Effectiveness of Data”, IEEE Intell. Syst. **24**, 2 (2009).
- [89] E.M. Metodiev *et al.*, “Classification without labels: learning from mixed samples in high energy physics”, J. High Energy. Phys **10**, 174 (2017).
- [90] L.M. Dery *et al.*, “Weakly supervised classification in high energy physics”, J. High Energ. Phys. **05**, 145 (2017).
- [91] R.E. Bellman, “Adaptive Control Processes”, ISBN:978-0691652214 (1961).
- [92] B. Schölkopf *et al.*, “Kernel principal component analysis”, Artificial Neural Networks, ICANN (1997).
- [93] S.T. Roweis and L. K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding”, Science **290**, 5500 (2000).
- [94] S. Kirkpatrick *et al.*, “Optimization by Simulated Annealing”, Science **220**, 4598 (1983).
- [95] G. Punzi, “Sensitivity of searches for new signals and its optimization”, arXiv:physics/0308063 (2003).
- [96] LHCb Collaboration, “Measurement of b hadron production fractions in 7 TeV pp collisions”, Phys. Rev. **D85**, 032008 (2012).
- [97] LHCb Collaboration, “Measurement of b hadron production fractions in 13 TeV pp collisions”, Phys. Rev. **D100**, 031102 (2019).
- [98] R.H. Dalitz, “CXII. On the analysis of τ -meson data and the nature of the τ -meson”, Philos. Mag. **44**, 357 (1953).

- [99] R.H. Dalitz, “Decay of τ Mesons of Known Charge”, Phys. Rev. **94**, 1046 (1954).
- [100] LHCb Collaboration, “Dalitz plot analysis of $\bar{B}_s^0 \rightarrow \bar{D}^0 K^- \pi^+$ decays”, Phys. Rev. **D90**, 072003 (2014).
- [101] BABAR Collaboration, “Measurement of $\bar{B}^0 \rightarrow D^{(*)0} \bar{K}^{(*)0}$ branching fractions”, Phys. Rev. **D74**, 031101 (2006).
- [102] Belle Collaboration, “Observation of $\bar{B}^0 \rightarrow D^0 \bar{K}^0$ and $\bar{B}^0 \rightarrow D^0 \bar{K}^{*0}$ Decays”, Phys. Rev. Lett. **90**, 141802 (2003).
- [103] LHCb Collaboration, “Observation of $B_s^0 \rightarrow \bar{D}^0 K_s^0$ and Evidence for $B_s^0 \rightarrow \bar{D}^0 K_s^0$ Decays”, Phys. Rev. Lett. **116**, 161802 (2016).
- [104] G.A. Cowan *et al.*, “RapidSim: An application for the fast simulation of heavy-quark hadron decays”, Comput. Phys. Commun. **214**, 239 (2017).
- [105] J. Mayer *et al.*, “Drawing an elephant with four complex parameters”, Am. J. Phys. **78**, 648 (2010).
- [106] BABAR Collaboration, “Measurement of the branching ratios $\Gamma(D_s^{*+} \rightarrow D_s^+ \pi^0)/\Gamma(D_s^{*+} \rightarrow D_s^+ \gamma)$ and $\Gamma(D_s^{*0} \rightarrow D^0 \pi^0)/\Gamma(D_s^{*0} \rightarrow D^0 \gamma)$ ”, Phys. Rev. **D72**, 091101 (2005).
- [107] BESIII Collaboration, “Precision measurement of the D^{*0} decay branching fractions”, Phys. Rev. **D91**, 031101 (2015).
- [108] S.S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, Ann. Math. Statist. **9**, 60 (1938).
- [109] D. Wang, “Sum rules for CP asymmetries of charmed baryon decays in the $SU(3)_F$ limit”, Eur. Phys. J. **C79**, 429 (2019).
- [110] H. Jiang, F. Yu, “Fragmentation-fraction ratio f_{Ξ_b}/f_{Λ_b} in b - and c -baryon decays”, Eur. Phys. J. **C78**, 224 (2018).
- [111] LHCb Collaboration, “Measurement of the mass and production rate of Ξ_b^- baryons”, Phys. Rev. **D99**, 052006 (2019).
- [112] T. Sjostrand, S. Mrenna, P. Skands, “PYTHIA 6.4 Physics and Manual”, J. High Energ. Phys. **05**, 26 (2006).
- [113] I. Belyaev *et al.*, “Handling of the generation of primary events in Gauss, the LHCb simulation framework”, J. Phys. Conf. Ser. **331** (2011).
- [114] D.J. Lange, “The EvtGen particle decay simulation package”, Nucl. Instrum. Meth. **A462**, 152 (2001).
- [115] P. Golonka, Z. Was, “PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays”, Eur. Phys. J. **C45**, 97 (2006).
- [116] S. Agostinelli *et al.*, “GEANT4—a simulation toolkit”, Nucl. Instrum. Meth. **A506**, 250 (2003).
- [117] J. Allison *et al.*, “Geant4 Developments and Applications”, IEEE Trans. Nucl. Sci. **53**, 270 (2006).
- [118] M. Clemencic *et al.*, “The LHCb Simulation Application, Gauss: Design, Evolution and Experience”, J. Phys. Conf. Ser. **331** (2011).

- [119] W.D. Hulsbergen, “Decay chain fitting with a Kalman filter”, Nucl. Instrum Meth. **A552**, 566 (2005).
- [120] R.H. Bartels *et al.*, “An Introduction to Splines for Computer Graphics and Geometric Modeling”, ISBN:978-1558604001 (1998).
- [121] C.M. da Fonseca, “On the eigenvalues of some tridiagonal matrices”, J. Comput. Appl. Math. **200**, 283 (2007).
- [122] R.A. Usmani, “Inversion of a tridiagonal Jacobi matrix”, Linear Algebra Appl. **212-213**, 413 (1994).
- [123] T. Gutsche *et al.*, “Polarization effects in the cascade decay $\Lambda_b \rightarrow \Lambda(\rightarrow p\pi^-) + J/\psi(\rightarrow \ell^+\ell^-)$ in the covariant confined quark model”, Phys. Rev. **D88**, 114018 (2013).
- [124] E. Gross and O. Vitells, “Trial factors for the look elsewhere effect in high energy physics”, Eur. Phys. J. **C70**, 525 (2010).
- [125] M. Stone, “Cross-Validatory Choice and Assessment of Statistical Predictions”, J. R. Stat. Soc. **B36**, 111 (1974).
- [126] G.C. Cawley and N.L.C. Talbot, “On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation”, J. Mach. Learn. Res. **11**, 2079 (2010).
- [127] A. Ramdas *et al.*, “On Wasserstein Two Sample Testing and Related Families of Non-parametric Tests”, arXiv:1509.02237 (2015).
- [128] C. Hsu and C. Lin, “A comparison of methods for multiclass support vector machines”, IEEE T. Neural Netw. **13**, 415 (2002).
- [129] C. Cortes and V. Vapnik, “Support-vector networks”, Mach. Learn. **20**, 273 (1995).
- [130] D.H. Wolpert, “Stacked generalization”, Neural Networks **5**, 241 (1992).
- [131] L. Breiman, “Bagging predictors”, Mach. Learn. **24**, 123 (1996).
- [132] T.K. Ho, “Random decision forests”, Proceedings of 3rd International Conference on Document Analysis and Recognition (Montreal), **1**, 278 (1995).
- [133] P. Geurts *et al.*, “Extremely randomized trees”, Mach. Learn. **63**, 3 (2006).
- [134] L. Breiman, “Arcing the Edge” Technical Report 486, University of California at Berkeley (1997).
- [135] Y. Freund and R.E.Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting”, J. Comput. Syst. Sci. **55**, 119 (1997).
- [136] A. Geron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems”, ISBN:978-1492032649 (2017).
- [137] S. Bach *et al.*, “On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation”, PLoS One **10** (2015).
- [138] R.C. Fong and A. Vedaldi, “Interpretable Explanations of Black Boxes by Meaningful Perturbation”, ICCV, 3449 (2017).
- [139] G. Montavon *et al.*, “Methods for interpreting and understanding deep neural networks”, Digit. Signal Process. **73** (2018).

Bibliography

- [140] G. Louppe *et al.*, “Learning to Pivot with Adversarial Networks”, arXiv:1611.01046 (2017).
- [141] S. Wunsch *et al.*, “Identifying the Relevant Dependencies of the Neural Network Response on Characteristics of the Input Space”, Comput. Softw. Big Sci. **2**, 5 (2018).
- [142] Z.J. Ajaltouni *et al.*, “ Λ_b decays into Λ -vector”, Phys. Lett. **B614**, 3 (2005).
- [143] J. Neyman, E.S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses”, Proc. R. Soc. Lond. **A231**, 289 (1933).

Appendix A

Reconstruction and Stripping Efficiencies

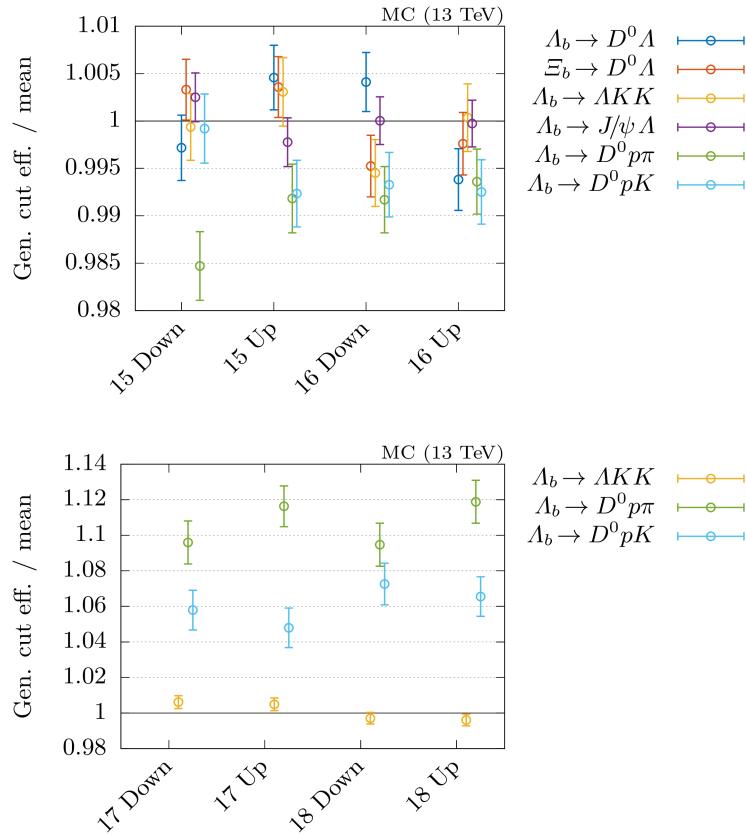


Fig. A.1. Generator cut efficiencies of various decays for different simulation conditions. For the sake of brevity magnet polarities are referred to as *Down* and *Up* for mag. down and mag. up, respectively. In order to compensate for their wide numerical spread (*cf.* Tab. A.1), each value is normalized to the respective mean of each decay for the full available data set. (Not all decays are simulated for the years 2017 and 2018.)

Tab. A.1. Simulation and trigger versions, as well as generator cut efficiencies (Gen. Cut) and amounts (#DST) for different decays and simulation conditions. For the sake of brevity, magnet configurations mag. down and mag. up are abbreviated with \downarrow and \uparrow , respectively.

Decay	Year	Simulation	Trigger	Gen. Cut [%]	#DST
$\Lambda_b \rightarrow D^0 \Lambda$	2015 \downarrow	Sim09c	0x411400a2	21.12(7)	2 009 840
	2015 \uparrow	Sim09c	0x411400a2	21.28(7)	2 000 794
	2016 \downarrow	Sim09c	0x6138160F	21.27(7)	2 040 356
	2016 \uparrow	Sim09c	0x6138160F	21.05(7)	2 004 574
$\Xi_b^0 \rightarrow D^0 \Lambda$	2015 \downarrow	Sim09h	0x411400a2	21.29(7)	1 000 273
	2015 \uparrow	Sim09h	0x411400a2	21.29(7)	1 000 140
	2016 \downarrow	Sim09g	0x6139160F	21.12(7)	3 005 698
	2016 \uparrow	Sim09g	0x6139160F	21.17(7)	3 001 062
$\Lambda_b \rightarrow D^0 p\pi^-$	2015 \downarrow	Sim09d	0x411400a2	15.50(6)	151 253
	2015 \uparrow	Sim09d	0x411400a2	15.61(6)	156 266
	2016 \downarrow	Sim09d	0x6139160F	15.61(6)	606 368
	2016 \uparrow	Sim09d	0x6139160F	15.64(5)	604 536
	2017 \downarrow	Sim09f	0x62661709	17.25(19)	13 356 991
	2017 \uparrow	Sim09f	0x62661709	17.57(18)	13 370 991
	2018 \downarrow	Sim09f	0x617d18a4	17.23(19)	7 100 997
	2018 \uparrow	Sim09f	0x617d18a4	17.61(19)	7 621 996
$\Lambda_b \rightarrow D^0 pK^-$	2015 \downarrow	Sim09d	0x411400a2	17.04(6)	154 815
	2015 \uparrow	Sim09d	0x411400a2	16.92(6)	154 293
	2016 \downarrow	Sim09d	0x6139160F	16.94(6)	607 837
	2016 \uparrow	Sim09d	0x6139160F	16.93(6)	604 865
	2017 \downarrow	Sim09f	0x62661709	18.04(19)	13 316 384
	2017 \uparrow	Sim09f	0x62661709	17.87(19)	13 857 088
	2018 \downarrow	Sim09f	0x617d18a4	18.29(20)	7 382 992
	2018 \uparrow	Sim09f	0x617d18a4	18.17(19)	7 487 998
$\Lambda_b \rightarrow \Lambda K^+ K^-$	2015 \downarrow	Sim09c	0x411400a2	22.42(8)	1 002 336
	2015 \uparrow	Sim09c	0x411400a2	22.51(8)	1 004 828
	2016 \downarrow	Sim09c	0x6138160F	22.31(8)	2 502 437
	2016 \uparrow	Sim09c	0x6138160F	22.45(8)	2 500 360
	2017 \downarrow	Sim09f	0x62661709	22.57(8)	2 508 342
	2017 \uparrow	Sim09f	0x62661709	22.55(8)	2 503 107
	2018 \downarrow	Sim09f	0x617d18a4	22.37(7)	2 501 960
	2018 \uparrow	Sim09f	0x617d18a4	22.35(7)	2 504 022
$\Lambda_b \rightarrow J/\psi \Lambda$	2015 \downarrow	Sim09c	0x411400a2	19.89(5)	519 498
	2015 \uparrow	Sim09c	0x411400a2	19.79(5)	506 398
	2016 \downarrow	Sim09c	0x6139160F	19.84(5)	2 008 494
	2016 \uparrow	Sim09c	0x6139160F	19.83(5)	2 008 422

Tab. A.2. Total amounts (#DTT) and product of generator cut efficiency and fractions (#DTT over #DST) in percent for different decays (Gen. Decay) and data taking condition when reconstructed in the given decay mode (Rec. Decay). For the sake of brevity, magnet configurations mag. down and mag. up are abbreviated with \downarrow and \uparrow , respectively. (Part 1/2)

				Gen. Decay	Rec. Decay	Year	#DTT	LL	% DTT	#DTT	DD	% DTT
$\Lambda_b \rightarrow D^0 \Lambda$	$\Lambda_b \rightarrow D^0 \Lambda$	2015	\uparrow	3572		0.0375(6)	8361		0.0879(10)			
		2015	\downarrow	3540		0.0377(6)	8350		0.0888(10)			
		2016	\uparrow	4399		0.0459(7)	9793		0.1021(11)			
		2016	\downarrow	4252		0.0447(7)	9482		0.0996(11)			
$\Xi_b^0 \rightarrow D^0 \Lambda$	$\Lambda_b \rightarrow D^0 \Lambda$	2015	\uparrow	1980		0.0421(10)	4813		0.1024(15)			
		2015	\downarrow	1953		0.0416(9)	4692		0.0999(15)			
		2016	\uparrow	6984		0.0491(6)	16126		0.1133(10)			
		2016	\downarrow	6918		0.0488(6)	16227		0.1144(10)			
$\Lambda_b \rightarrow D^0 p\pi^-$	$\Lambda_b \rightarrow D^0 p\pi^-$	2015	\uparrow	7045		0.722(9)	—		—			
		2015	\downarrow	7302		0.729(9)	—		—			
		2016	\uparrow	32373		0.833(5)	—		—			
		2016	\downarrow	31933		0.826(5)	—		—			
		2017	\uparrow	692296		0.894(10)	—		—			
		2017	\downarrow	696910		0.916(9)	—		—			
		2018	\uparrow	339065		0.823(9)	—		—			
		2018	\downarrow	355522		0.821(9)	—		—			
		2015	\uparrow	6850		0.754(9)	—		—			
		2015	\downarrow	6908		0.758(9)	—		—			
$\Lambda_b \rightarrow D^0 pK^-$	$\Lambda_b \rightarrow D^0 pK^-$	2016	\uparrow	30117		0.839(6)	—		—			
		2016	\downarrow	30247		0.846(6)	—		—			
		2017	\uparrow	690195		0.935(10)	—		—			
		2017	\downarrow	716056		0.923(10)	—		—			
		2018	\uparrow	346474		0.858(9)	—		—			
		2018	\downarrow	346040		0.840(9)	—		—			

Tab. A.3. Total amounts (#DTT) and product of generator cut efficiency and fractions (#DTT over #DST) in percent for different decays (Gen. Decay) and data taking condition when reconstructed in the given decay mode (Rec. Decay). For the sake of brevity, magnet configurations mag. down and mag. up are abbreviated with \downarrow and \uparrow , respectively. (Part 2/2)

Gen. Decay	Rec. Decay	Year	#DTT	LL		DD	
				% DTT	#DTT	% DTT	#DTT
$\Lambda_b \rightarrow D^0 p\pi^-$	$\Lambda_b \rightarrow D^0 \Lambda$	2015 \uparrow	309	0.0317(18)	—	—	—
		2015 \downarrow	334	0.0334(18)	—	—	—
		2016 \uparrow	1283	0.0330(9)	—	—	—
		2016 \downarrow	1289	0.0333(9)	—	—	—
$\Lambda_b \rightarrow D^0 K^+ K^-$	$\Lambda_b \rightarrow D^0 K^+ K^-$	2015 \uparrow	4870	0.1089(16)	17 622	0.3942(33)	—
		2015 \downarrow	4881	0.1093(16)	17 648	0.3953(33)	—
		2016 \uparrow	12 903	0.1151(11)	43 827	0.3908(23)	—
		2016 \downarrow	12 865	0.1155(11)	44 038	0.3953(23)	—
		2017 \uparrow	13 463	0.1212(11)	48 639	0.4377(25)	—
		2017 \downarrow	13 322	0.1200(11)	48 581	0.4376(25)	—
		2018 \uparrow	13 339	0.1193(11)	48 118	0.4303(24)	—
		2018 \downarrow	13 407	0.1197(11)	48 428	0.4323(24)	—
$\Lambda_b \rightarrow D^0 K^+ K^-$	$\Lambda_b \rightarrow D^0 \Lambda$	2015 \uparrow	165	0.003 69(29)	443	0.0099(5)	—
		2015 \downarrow	157	0.003 52(28)	397	0.0089(4)	—
		2016 \uparrow	435	0.003 88(19)	1084	0.009 67(30)	—
		2016 \downarrow	429	0.003 85(19)	829	0.007 44(26)	—
		2017 \uparrow	408	0.003 67(18)	1088	0.009 79(30)	—
		2017 \downarrow	442	0.003 98(19)	961	0.008 66(28)	—
		2018 \uparrow	418	0.003 74(18)	1091	0.009 76(30)	—
		2018 \downarrow	421	0.003 76(18)	1162	0.010 37(31)	—
$\Lambda_b \rightarrow J/\psi \Lambda$	$\Lambda_b \rightarrow J/\psi \Lambda$	2015 \uparrow	4072	0.1559(25)	14 344	0.549(5)	—
		2015 \downarrow	3989	0.1559(25)	13 870	0.542(5)	—
		2016 \uparrow	16 537	0.1633(13)	55 604	0.5492(27)	—
		2016 \downarrow	16 731	0.1652(13)	55 387	0.5469(27)	—

Appendix B

Stripping Selections

In this appendix we list preselection steps of the so-called *stripping* phase during reconstruction of $\Lambda_b \rightarrow J/\psi \Lambda$, $\Lambda_b \rightarrow D^0 \Lambda$, and $\Lambda_b \rightarrow D^0 p \pi^-$ decays in Tab. B.1, Tab. B.2 and Tab. B.3, respectively. The optimization of these selection requirements was not objective of the present analysis and, due to the deep encoding in the data storing process during data taking, should be considered immutable in the context of the present analysis. Most requirements are considered *loose* and are superseded by thresholds in the subsequent selection steps. We refer to Sec. 2.2 for an explanation of the acronyms and abbreviations.

Tab. B.1. Selection criteria of the preselection step used for reconstructing $\Lambda_b \rightarrow J/\psi \Lambda$. *Combination* selections are evaluated before applying a vertex fit, whereas *mother* selections are refined by such a fit. Variables and acronyms are explained Sec. 2.2.

Particle	Type	Selection
μ	—	$p_T > 0.5 \text{ GeV}/c$
μ	—	$\text{DLL}(\mu - \pi) > 0$
p	—	$p > 2 \text{ GeV}/c$
p	—	$p_T > 0.25 \text{ GeV}/c$
p (L)	—	$\text{DLL}(p - \pi) > -5$
p (L)	—	min. χ^2 dist. to any PV > 9
p (D)	—	min. χ^2 dist. to any PV > 4
π	—	$p > 2 \text{ GeV}/c$
π	—	$p_T > 0.1 \text{ GeV}/c$
π (L)	—	$\text{DLL}(p - \pi) > -5$
π (L)	—	min. χ^2 dist. to any PV > 9
π (D)	—	min. χ^2 dist. to any PV > 4
at least one final state	—	$p > 10 \text{ GeV}/c$
at least one final state	—	$p_T > 1.7 \text{ GeV}/c$
at least one final state	—	χ^2/DoF of track fit < 4
at least one final state	—	min. dist. to any PV $> 4\sigma$ and $> 0.1 \text{ mm}$
J/ψ	combination	$ m(J/\psi) - \text{PDG} < 150 \text{ MeV}/c^2$
J/ψ	combination	χ^2 of DOCA(μ^+, μ^-) < 20
J/ψ	mother	$ m(J/\psi) - \text{PDG} < 80 \text{ MeV}/c^2$
J/ψ	mother	χ^2 of end vertex < 16
Λ (LL)	combination	$ m(\Lambda) - \text{PDG} < 50 \text{ MeV}/c^2$
Λ (DD)	combination	$ m(\Lambda) - \text{PDG} < 80 \text{ MeV}/c^2$
Λ (LL)	combination	χ^2 of DOCA(p, π^-) < 30
Λ (DD)	combination	χ^2 of DOCA(p, π^-) < 25
Λ	mother	$ m(\Lambda) - \text{PDG} < 15 \text{ MeV}/c^2$
Λ	mother	χ^2 of end vertex < 20
Λ_b	combination	$5.02 \leq m(\Lambda_b) \leq 6.22 \text{ GeV}/c^2$
Λ_b	mother	$5.12 \leq m(\Lambda_b) \leq 6.12 \text{ GeV}/c^2$
Λ_b	mother	χ^2/DoF of end vertex < 10
Λ_b	mother	lifetime w.r.t. best PV $> 0.2 \text{ ps}$
Λ_b	mother	$\Delta\chi^2_{\text{IP}}$ w.r.t. best PV < 25
Λ_b	mother	DIRA w.r.t. best PV > 0.999

Tab. B.2. Selection criteria of the preselection step used for reconstructing $\Lambda_b \rightarrow D^0 \Lambda$. *Combination* selections are evaluated before applying a vertex fit, whereas *mother* selections are refined by such a fit. Variables and acronyms are explained Sec. 2.2. At this early stage of the analysis, the final states of D^0 are not treated differently and in order to avoid ambiguity with the π from $\Lambda \rightarrow p\pi^-$ we refer to them as h for the sake of brevity.

Particle	Type	Selection
p	—	$p > 2 \text{ GeV}/c$
p	—	$p_T > 0.25 \text{ GeV}/c$
p (L)	—	$\text{DLL}(p - \pi) > -5$
p (L)	—	min. χ^2 dist. to any PV > 9
p (D)	—	min. χ^2 dist. to any PV > 4
π	—	$p > 2 \text{ GeV}/c$
π	—	$p_T > 0.1 \text{ GeV}/c$
π (L)	—	$\text{DLL}(p - \pi) > -5$
π (L)	—	min. χ^2 dist. to any PV > 9
π (D)	—	min. χ^2 dist. to any PV > 4
h	—	$p > 1 \text{ GeV}/c$
h	—	$p_T > 0.1 \text{ GeV}/c$
h	—	min. χ^2 dist. to any PV > 4
h	—	χ^2/DoF of track fit < 4
h	—	Ghost Prob. < 0.4
at least one final state	—	$p > 10 \text{ GeV}/c$
at least one final state	—	$p_T > 1.7 \text{ GeV}/c$
at least one final state	—	χ^2/DoF of track fit < 4
at least one final state	—	min. dist. to any PV $> 4\sigma$ and $> 0.1 \text{ mm}$
all final states i	—	$\sum_i p_T^{(i)} > 5 \text{ GeV}/c$
any h	—	$p_T > 0.5 \text{ GeV}/c$
any h	—	$p > 5 \text{ GeV}/c$
D^0	combination	$1764.84 \leq m(K^- \pi^+) \leq 1964.84 \text{ MeV}/c^2$
D^0	combination	$\text{DOCA}(K^-, \pi^+) < 0.5 \text{ mm}$
D^0	mother	χ^2/DoF of end vertex < 10
D^0	mother	χ^2 distance to best PV > 36
Λ (LL)	combination	$ m(\Lambda) - \text{PDG} < 50 \text{ MeV}/c^2$
Λ (DD)	combination	$ m(\Lambda) - \text{PDG} < 80 \text{ MeV}/c^2$
Λ (LL)	combination	χ^2 of DOCA(p, π^-) < 30
Λ (DD)	combination	χ^2 of DOCA(p, π^-) < 25
Λ	mother	$ m(\Lambda) - \text{PDG} < 15 \text{ MeV}/c^2$
Λ	mother	χ^2 of end vertex < 20
Λ_b	combination	$5.02 \leq m(\Lambda_b) \leq 6.22 \text{ GeV}/c^2$
Λ_b	mother	$5.12 \leq m(\Lambda_b) \leq 6.12 \text{ GeV}/c^2$
Λ_b	mother	χ^2/DoF of end vertex < 10
Λ_b	mother	lifetime w.r.t. best PV $> 0.2 \text{ ps}$
Λ_b	mother	$\Delta\chi_{\text{IP}}^2$ w.r.t. best PV < 25
Λ_b	mother	DIRA w.r.t. best PV > 0.999

Tab. B.3. Selection criteria of preselection used for reconstructing $\Lambda_b \rightarrow D^0 p\pi^-$. *Combination* selections are evaluated before applying a vertex fit, whereas *mother* selections are refined by such a fit. Variables and acronyms are explained Sec. 2.2. At this early stage of the analysis, the final states of D^0 are not treated differently and in order to avoid ambiguity with the π^- from $\Lambda_b \rightarrow D^0 p\pi^-$ we refer to them as h for the sake of brevity.

Particle	Type	Selection
h	—	$p > 1 \text{ GeV}/c$
p	—	$p > 2 \text{ GeV}/c$
π	—	$p > 2 \text{ GeV}/c$
p	—	$\text{DLL}(p - \pi) > -10$
π	—	$\text{DLL}(K - \pi) < 20$
at least one final state	—	$p > 10 \text{ GeV}/c$
at least one final state	—	$p_T > 1.7 \text{ GeV}/c$
at least one final state	—	χ^2/DoF of track fit < 4
at least one final state	—	min. dist. to any PV $> 4\sigma$ and $> 0.1 \text{ mm}$
all final states	—	$p_T > 0.1 \text{ GeV}/c$
all final states	—	min. χ^2 dist. to any PV > 4
all final states	—	χ^2/DoF of track fit < 4
all final states	—	Ghost Prob. < 0.4
all final states i	—	$\sum_i p_T^{(i)} > 5 \text{ GeV}/c$
any among $\{p, \pi\}$	—	$p_T > 0.5 \text{ GeV}/c$
any among $\{p, \pi\}$	—	$p > 5 \text{ GeV}/c$
any h	—	$p_T > 0.5 \text{ GeV}/c$
any h	—	$p > 5 \text{ GeV}/c$
D^0	combination	$1764.84 \leq m(K^-\pi^+) \leq 1964.84 \text{ MeV}/c^2$
D^0	combination	$\text{DOCA}(K^-, \pi^+) < 0.5 \text{ mm}$
D^0	mother	χ^2/DoF of end vertex < 10
D^0	mother	χ^2 distance to best PV > 36
$p\pi^-$	combination	$p_T > 1 \text{ GeV}/c$
$p\pi^-$	combination	$m(p\pi) < 5 \text{ GeV}/c^2$
$p\pi^-$	combination	$\text{DOCA}(p, \pi^-) < 0.5 \text{ mm}$
$p\pi^-$	mother	χ^2 of end vertex < 16
$p\pi^-$	mother	χ^2 dist. to best PV > 16
Λ_b	combination	$5.02 \leq m(\Lambda_b) \leq 6.22 \text{ GeV}/c^2$
Λ_b	mother	$5.12 \leq m(\Lambda_b) \leq 6.12 \text{ GeV}/c^2$
Λ_b	mother	χ^2/DoF of end vertex < 10
Λ_b	mother	lifetime w.r.t. best PV $> 0.2 \text{ ps}$
Λ_b	mother	$\Delta\chi_{\text{IP}}^2$ w.r.t. best PV < 25
Λ_b	mother	DIRA w.r.t. best PV > 0.999

Appendix C

Decay Tree Fit

In complex decay chains, such as $\Lambda_b \rightarrow (D^0 \rightarrow K^-\pi^+)(\Lambda \rightarrow p\pi^-)$, typically only the very first particle in the decay chain (Λ_b) and the final state particles (K^- , π^+ , p and π^-) produce signatures in the detector. In general, the initial particle is produced during the high energetic pp -scattering which causes a plethora of detector signatures. Its position is therefore directly reconstructible and referred to as the primary vertex (PV), whereas the latter are reconstructed as (charged) tracks from multiple vertices, measured by various detector modules. The matching of both is ambiguous in general and is additionally impeded by measurement inaccuracies. A decay tree fit (DTF) refines those measurements in one global fit w.r.t. a set of constraints and allows an ordering of the different track and PV combinations by their respective χ^2 value. A DTF is a fit of constraints¹ and therefore a correction of the measured track parameters w.r.t. given hypotheses. The measurements are assumed as external constraints with uncertainties, whereas the assumed hypotheses are encoded as strict internal constraints. A DTF without any constraints is pointless and has zero DoF. For n measured points and $0 < k < n$ constraints the number of parameters is $n - k$ since each constraint can be used to eliminate one measured point by giving it in terms of the remaining measurements. Hence the DoF of the χ^2 of the DTF is $n - (n - k) = k$.

A DTF is used to fit an entire decay chain which can include intermediate states. Each intermediate state can either be a short-living resonance or an intermediate state with a finite decay length. Whether or not an intermediate state is considered short- or long-living depends on the detector resolution. For the LHCb detector, J/ψ mesons are considered to be short-living, whereas D^0 mesons or Λ baryons are categorized as long-living intermediate states. Each intermediate state contributes to the set of constraints by forcing origin- and decay-vertex to be fixed points. For $\Lambda_b \rightarrow (D^0 \rightarrow K^-\pi^+)(\Lambda \rightarrow p\pi^-)$ this leads to a set of three constraints, *i.e.*, three DoF in total (*cf.* Fig. C.1):

- The p and π^- tracks should not be skewed. Their common intersection is the decay vertex of the Λ baryon.
- Similarly, both, K^- and π^+ , should intersect at the decay vertex of the D^0 meson.
- The tracks of the intermediate states Λ and D^0 that are inferred by four-momentum conservation of the final state particles, should not be skewed lines either, but form the decay vertex of the Λ_b baryon. Consequently, this vertex is also the origin vertex of the intermediate states and can be used to determine the lifetime of the respective states.

For long-living intermediate states that decay via a two-body channel, each vertex constraint contributes one additional DoF to the χ^2 of the respective DTF. For short-living intermediate states or decay channels with more than two final states the contribution is larger. For example, the total DoF of the vertex constraints of a DTF of a $\Lambda_b \rightarrow (J/\psi \rightarrow \ell^+\ell^-)(\Lambda \rightarrow p\pi^-)$ decay sums up to four, since the J/ψ meson is a resonance and thus effectively makes the decay vertex of the Λ_b a three point vertex $\Lambda_b \rightarrow \ell^+\ell^-\Lambda$ that comes with three DoF.

¹Technically, at LHCb the DTF is implemented as a Kalman filter, *cf.* Ref. [119] for more details.

Additional constraints can be applied, *e.g.*, in most cases it makes sense to constrain the origin vertex of the very first particle of a given decay chain to the PV and since this is a fixed point this constraint adds two DoF to the fit. Another useful constraint is a mass constraint which forces the invariant mass, given by four-momentum addition of an intermediate state, to match a given particle hypothesis. Each mass constraint adds one additional DoF to the fit.

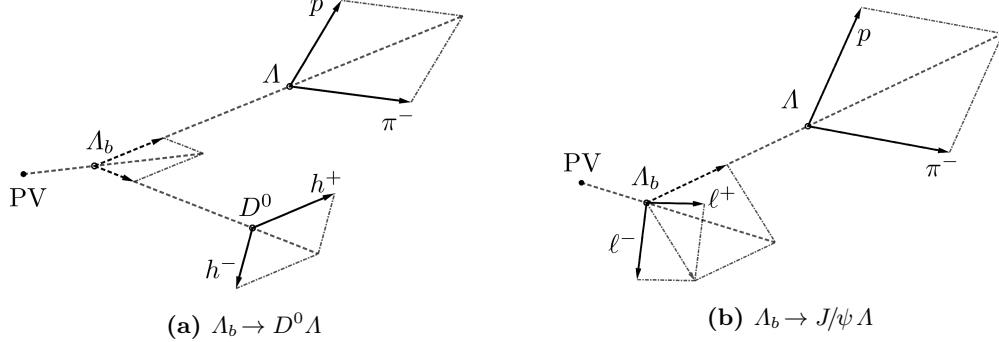


Fig. C.1. Decay topologies as seen by a DTF of the decays $A_b \rightarrow D^0 \Lambda$ (left) and $A_b \rightarrow J/\psi \Lambda$ (right). At LHCb the J/ψ is a (short-living) resonance and the decay vertex of the A_b baryon appears as a three-point vertex $A_b \rightarrow \ell^+ \ell^- \Lambda$ during reconstruction (four DoF w/o PV or mass constraints), whereas the flight lengths of the D^0 meson and Λ baryon can be measured significantly and finite decay lengths can be accommodated in a DTF (three DoF w/o PV or mass constraints).

Constraining the mass of resonances, in general improves the resolution of related kinematic features, but does also introduce non-trivial distortions of background components. For example, during the analysis of the background contribution of $A_b \rightarrow D^0 p \pi^-$ decays in the invariant mass of D^0 and Λ candidates, we discovered a dilution of the flight distance of Λ candidates. In case of genuine $A_b \rightarrow D^0 p \pi^-$ decays, the distribution of the flight distance of $p \pi^-$ combinations, spuriously forming a Λ candidate, is symmetrically smeared around zero due to resolution effects in presence of no $m(\Lambda)$ mass constraint and is uncorrelated with the invariant mass $m(p \pi^-)$ as shown in Fig. C.2a. Constraining $m(\Lambda)$ to its nominal value significantly increases the smearing and introduces a strong correlation between the flight distance and the invariant mass $m(p \pi^-)$ as shown in Fig. C.2b. (Clearly, the latter has to be evaluated before applying the DTF and forcing its value to the nominal value.) The reason for this behavior is that a mass constraint (similar to a PV constraint) affects the opening angle of the daughters of the Λ candidate: If the combined invariant mass of the p and π^- is below the constrained value of $m(\Lambda)$, the DTF pushes the three-momentum magnitudes towards larger values during the optimization. A correlation with the decay tree of the corresponding D^0 is evaded by simultaneously increasing the opening angle between the three-momentum vectors which corresponds to an increase of the reconstructed flight distance. Similarly, invariant masses above $m(\Lambda)$ pushes the opening angle towards smaller values and thus decrease the reconstructed flight distance.

In Fig. C.3 we compare the effects of requiring a minimal flight distance and a minimal DTF probability for the quoted example of the non-resonant background $A_b \rightarrow D^0 p \pi^-$ in $m(D^0 \Lambda)$. The data are simulated $A_b \rightarrow D^0 p \pi^-$ decays after a preselection step. Before applying neither of both criteria, the distribution of the invariant mass is flat in good approximation. Requiring a minimal DTF probability symmetrically rejects events which are too far off the fixed $m(\Lambda)$ value, whereas the flight distance criterion rejects only events above this value, resulting in the strongly asymmetric distribution we witnessed in Sec. 7.1.

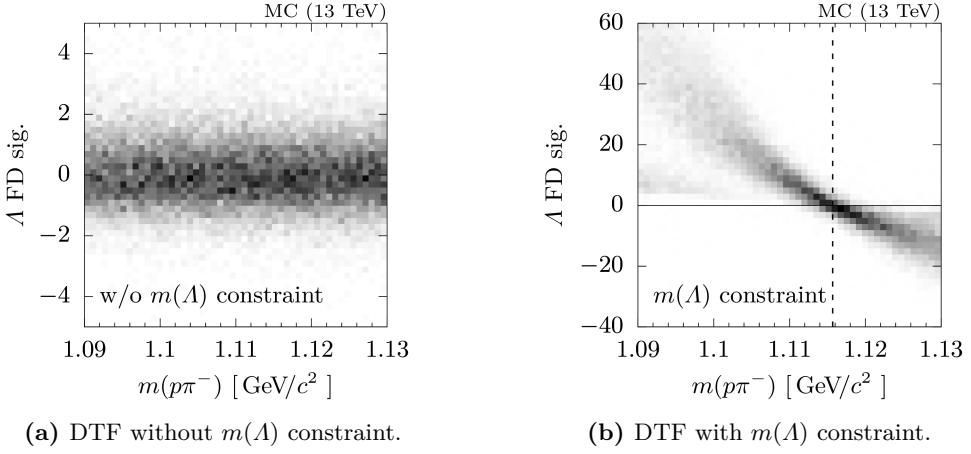


Fig. C.2. Correlation of the combined invariant mass of p and π^- candidates of genuine $\Lambda_b \rightarrow D^0 p \pi^-$ decays, reconstructed as $\Lambda_b \rightarrow D^0 \Lambda$ and the flight distance significance in presence of a DTF. Constraining the invariant mass $m(p\pi^-)$ to $m(\Lambda)$ (dashed line) introduces a significant correlation and increases the smearing of the flight distance distribution compared to the case of a DTF without a $m(\Lambda)$ constraint (left). (Note the different scaling of the y -axis.) The invariant mass shown on the x -axes (left and right) is the result of a four-momentum addition before applying the respective DTFs.

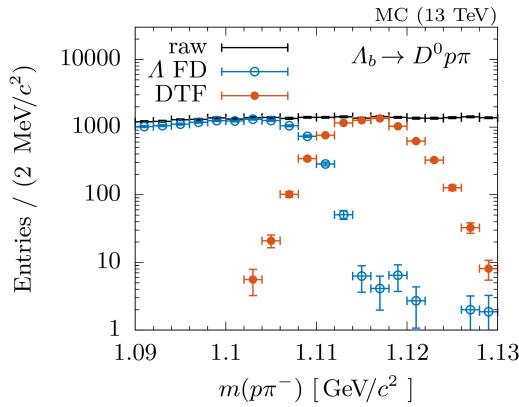


Fig. C.3. Effect of requiring a minimal flight distance (Λ FD) and a minimal DTF probability (DTF) to simulated $\Lambda_b \rightarrow D^0 p \pi^-$ decays that are reconstructed and fitted as $\Lambda_b \rightarrow D^0 \Lambda$ to the combined invariant mass of p and π^- candidates.

Appendix D

Fit Probability

In Sec. 4.1.3 we use the distribution of χ^2_{DTF} of a DTF to suppress combinatorial background. An interesting transformation of the χ^2_{DTF} distribution is achieved via the regularized upper incomplete gamma function

$$\chi^2 \mapsto \frac{1}{\Gamma(\text{DoF}/2)} \int_{\chi^2/2}^{\infty} dt t^{\text{DoF}/2-1} e^{-t}, \quad (\text{D.1})$$

where $\Gamma(s)$ is the complete gamma function. The result of this transformation, which we refer to as the *fit probability* $\text{Prob}(\chi^2, n)$, is the probability that an observed χ^2_{obs} exceeds the value χ^2 by chance, even for a correct model. The fit probability itself is distributed uniformly for χ^2 -distributed events with DoF degrees of freedom.

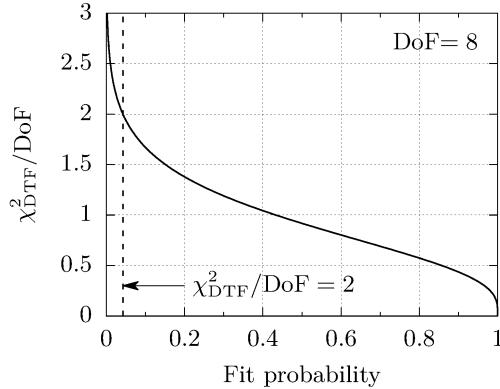


Fig. D.1. Fit probability as a function of χ^2 with 8 DoF as defined in Eq. (D.1).

For the selection requirements of the established tight selection of $\Lambda_b \rightarrow J/\psi \Lambda$ decays, $\chi^2_{\text{DTF}}/\text{DoF} = 3$ (2) for LL (DD) tracks with DoF = 8 (*cf.* Fig. D.1), the inverted fit probability is

$$1 - \text{Prob}(3 \times 8, 8) = 99.8\%, \\ 1 - \text{Prob}(2 \times 8, 8) = 95.8\%.$$

These numbers do not reproduce those found in data (*cf.* Tab. 4.3) and thus $\chi^2_{\text{DTF}}/\text{DoF}$ is not exactly χ^2 -distributed with 8 DoF. In Fig. D.2 we show the fit probability distribution of LL and DD tracks, and for recorded data and (truth-matched) simulated events. Although none of these distributions are flat as expected from a χ^2 -distribution, there is no significant discrepancy

between recorded data and simulated events visible.¹ The reason for the deviation from the expected uniform distribution at small fit probability values is unclear. Small fit probability values mean large χ^2 values such that the accumulation at small fit probability values translates into too many DTFs with large χ^2 values, *i.e.*, deviations of a Gaussian shape that raises χ^2 -tail contributions.

There are two main reasons which could potentially cause accumulations at small fit probability values: Non-Gaussian effects and an incomplete uncertainty budget of the applied track fit. Due to the central limit theorem, the former effect will get diluted for many different contributions, eventually, whereas the latter remains, even for many different effects. Examples that typically contribute to both of these categories are instrumental misalignment and imprecise material calculations which both introduce additional scattering and other kinds of interactions. Additionally, both of these could potentially impair the spatial and temporal resolution of the magnetic field. On top of that, scattering at large angles of charged particles (all final state particles in the present analysis are charged) with atomic electrons or nuclei, bend tracks that are assumed to be straight lines in the DTF approach. All these effects increases the χ^2 value of the DTF and will affect the fit probability calculation if the corresponding uncertainties are incompletely described during the track fit.

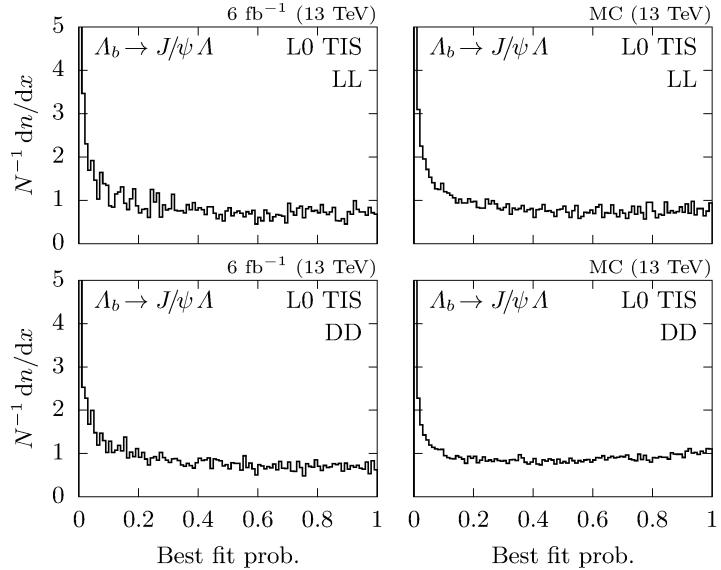


Fig. D.2. Fit probability of LL and DD tracks (top and bottom), and for recorded data (left) and truth-matched, simulated events (right).

¹In Appx. M we discuss the effect of deviations for large χ^2 values. These deviations are a nuisance for efficiency estimations, but not for the estimation of the calibration factors.

Appendix E

Cubic Splines and their derivatives

Throughout the present analysis we occasionally use cubic splines for smoothing distributions, by fitting y_i for a fixed set of (x_i, y_i) tuples and calculating a spline between these points. Typically, each y_i can be associated with an uncertainty and in order to propagate this uncertainty to the resulting spline fit, the derivatives of a spline function are needed. Due to the lack of available implementations for this specific problem, we roll out our own implementation of cubic splines, based on the theory described below.

Splines are a model independent way to interpolate between grid points. Cubic splines realize this interpolation with polynomial pieces of the form

$$f_i(t) = a_i + b_i t + c_i t^2 + d_i t^3 \quad \text{for } i = 1 \dots k,$$

where t is a parameter $t \in [0, 1]$. For $n = k + 1$ grid points $(x_0, y_0), \dots, (x_k, y_k)$ there are k polynomial pieces f_i . Each parameter t maps for the polynomial piece f_i from the interval $t \in [0, 1]$ to $x \in [x_i, x_{i+1}]$

$$t = t(x) = \frac{x - x_i}{x_{i+1} - x_i}$$

such that

$$f(x) = \begin{cases} f_1(t) & \text{for } x \in [x_0, x_1] \\ \vdots \\ f_k(t) & \text{for } x \in [x_{k-1}, x_k] \end{cases}$$

has a gapless coverage on $[x_0, x_k]$. Each polynomial piece brings 4 DoF. Following Ref. [120] these DoF are fixed by simultaneously satisfying the following set of constraints:

- $f_i(0) = y_i$ for $i = 1 \dots k$,
- $f_i(1) = y_{i+1}$ for $i = 1 \dots k$,
- $f'_i(1) = f'_{i+1}(0)$ for $i = 1 \dots (k - 1)$,
- $f''_i(1) = f''_{i+1}(0)$ for $i = 1 \dots (k - 1)$ and,
- $f''_1(0) = f''_k(1) = 0$ (natural splines).

These are $k + k + (k - 1) + (k - 1) + 2 = 4k$ relations that constrain all $4k$ coefficients a_i, b_i, c_i and d_i . The latter relation defines *natural* (cubic) splines and is a boundary condition. Other common boundary conditions are:

- *not-a-knot*: The third derivative is also continuous at (x_1, y_1) and (x_k, y_k) .
- *periodic*: The interpolated functions is assumed to be periodic, i.e. $(x_0, y_0) \equiv (x_n, y_n)$.
- *clamped*: The first derivative at curve ends are zero.

The coefficients a_i, b_i, c_i and d_i are unambiguously defined by the given set of grid points y_0, \dots, y_k :

$$\begin{aligned} a_i &= y_i, \\ b_i &= D_i, \\ c_i &= 3(y_{i+1} - y_i) - 2D_i - D_{i+1}, \\ d_i &= 2(y_i - y_{i+1}) + D_i + D_{i+1}, \end{aligned}$$

with $D_i := f'_i(0)$. This set of equations translates into solving the matrix equation

$$\underbrace{\begin{pmatrix} 2 & 1 & & & \\ 1 & 4 & 1 & & \\ & 1 & 4 & 1 & \\ & & 1 & 4 & 1 \\ & & & \ddots & \ddots & \ddots \\ & & & & 1 & 4 & 1 \\ & & & & & 1 & 2 \end{pmatrix}}_{=:T} \begin{pmatrix} D_0 \\ D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_{k-1} \\ D_k \end{pmatrix} = 3 \begin{pmatrix} y_1 - y_0 \\ y_2 - y_0 \\ y_3 - y_1 \\ \vdots \\ y_{k-1} - y_{k-3} \\ y_k - y_{k-2} \\ y_k - y_{k-1} \end{pmatrix}.$$

The matrix T as defined above is a special case of the more general class of non-singular tridiagonal matrices

$$\tilde{T} = \begin{pmatrix} a_1 & b_1 & & & \\ c_1 & a_2 & b_2 & & \\ & c_2 & a_3 & b_3 & \\ & & \ddots & \ddots & \ddots \\ & & & c_{n-2} & a_{n-1} & b_{n-1} \\ & & & & c_{n-1} & a_n \end{pmatrix}$$

(missing entries correspond to nil entries). The solution for \tilde{T}^{-1} can be found in Refs. [121, 122] and reads for the case $\tilde{T} = T$:

$$(T^{-1})_{ij} = (T^{-1})_{ji} = (-1)^{i+j} \theta_{n-j} \frac{\theta_{i-1}}{\theta_n} \quad \text{for } i \leq j,$$

where θ_i is defined recursively by

$$\theta_i = a_i \theta_{i-1} - \theta_{i-2} \quad \text{for } i = 2, 3, \dots, n$$

with initial conditions $\theta_0 = 1, \theta_1 = a_1 = 2$.

From the construction of the coefficients we see that they are linear in y_i (in particular there are no cross combinations such as $y_i y_j$) and there are no constant offsets. Therefore, derivation w.r.t. y_j of f_i is the same as replacing $y_r = \delta_{rj}$ for $r = 0 \dots k$. Derivatives of f_i w.r.t. y_i are again cubed polynomials in t ,

$$\frac{\partial f_i}{\partial y_j} = a_{ij} + b_{ij}t + c_{ij}t^2 + d_{ij}t^3. \quad (\text{E.1})$$

For $n = 4$ the numerical values of the coefficients a_{ij} , b_{ij} , c_{ij} and d_{ij} are

$$a_{i,j+1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}_{i,j+1},$$

$$b_{i,j+1} = 780 \times \begin{pmatrix} -989 & 1254 & -336 & 90 \\ -362 & -168 & 672 & -180 \\ 97 & -582 & -12 & 630 \end{pmatrix}_{i,j+1},$$

$$c_{i,j+1} = 780 \times \begin{pmatrix} 0 & 0 & 0 & 0 \\ 627 & -1422 & 1008 & -270 \\ -168 & 1008 & -1692 & 1080 \end{pmatrix}_{i,j+1},$$

$$d_{i,j+1} = 780 \times \begin{pmatrix} 209 & -474 & 336 & -90 \\ -265 & 810 & -900 & 450 \\ 71 & -426 & 924 & -930 \end{pmatrix}_{i,j+1}.$$

These derivatives do no longer depend on the particular choice of y_i and they are symmetric under $i \leftrightarrow (n-i)$, $j \leftrightarrow (k-j)$ and $t \leftrightarrow (1-t)$ transformations:

$$\frac{\partial f_i(t)}{\partial y_j} = \frac{\partial f_{n-i}(1-t)}{\partial y_{k-j}},$$

for k splines and $n = k + 1$ grid points. In particular, these equations read for $t = 0$ or $t = 1$

$$a_{i,j+1} = (a + b + c + d)_{n-i,k-j+1} = \delta_{i,j+1}.$$

Since these derivatives are the coefficients of a f_i expansion w.r.t. the grid points y_i , i.e.,

$$f_i(t) = \sum_{j=0}^k q_{ij} y_j = \sum_{j=0}^k \frac{\partial f_i(t)}{\partial y_j} y_j \quad \text{for } i = 1 \dots k,$$

they obey

$$\frac{\partial f_i(t=0)}{\partial y_j} = \frac{\partial f_{i-1}(t=1)}{\partial y_j} = \delta_{ij},$$

that appear as maxima and periodic knots in Fig. E.1.

We note once more that q_{ij} is not a function of y_i and is unambiguously defined by n . This is also given for other expansions of f_i w.r.t. the grid points, for instance in the *sinc-interpolation*

$$f_i(t) = \sum_{j=0}^k \text{sinc}((t-j)\pi) y_j$$

which resembles the shape of the derivatives shown in Fig. E.1 in good approximation. The expansion coefficients q_{ij} of the sinc-interpolation do not even depend on n anymore and there is no need to invert a $n \times n$ band matrix. However, this computational benefit is compensated by the fact that sinc-interpolations tend to overshoot compared to the cubic spline approach, even if damped with a Gaussian function, as shown in Fig. E.2.

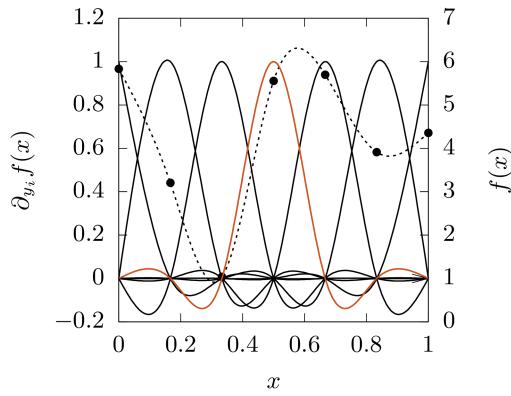


Fig. E.1. Natural cubic spline (dashed line, right axis) of $n = 7$ equidistant (x_i, y_i) tuples (dots, right axis), as well as its derivatives w.r.t. y_i (solid lines, left axis).

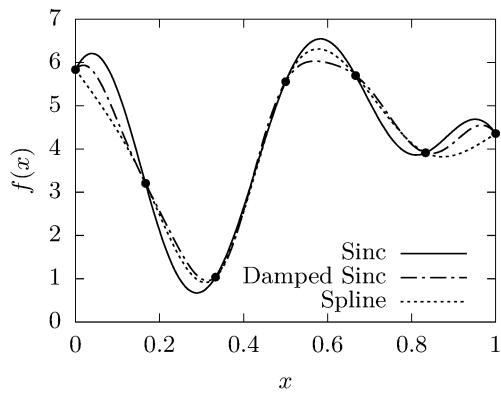


Fig. E.2. Sinc-interpolation w/o and with Gaussian damping ($\sigma = 1$), as well as a natural cubic spline for $n = 7$ equidistant (x_i, y_i) tuples (dots).

Appendix F

Weighting Scheme 1 Supplementaries

Below, we summarize some investigations that exclude various possible explanations for the observed deviations in $w_2(\eta)$ (Scheme 1) that we established in Chap. 4.

F.1 Influence of Polarizations

Theoretically, polarizations could cause deviations, when not properly reproduced in simulated events. The significant polarization of the Λ particle in $\Lambda_b \rightarrow J/\psi \Lambda$ was predicted [123] and measured by the LHCb collaboration [70]. However, assuming CP conservation, polarization effects should cancel exactly for Λ_b and $\bar{\Lambda}_b$. Since the weighting procedure outlined above considers combined data sets of Λ_b and $\bar{\Lambda}_b$ (raw asymmetry in data $1 - n(\Lambda_b)/n(\bar{\Lambda}_b) = -3.9(1.9)\%$ for LL and $-8.4(1.6)\%$ for DD) and no enhancement is observed in $\eta(\Lambda_b)$ distributions where samples are explicitly split w.r.t. the proton charge, we exclude that polarization effects introduce such an effect.

F.2 Influence of Misalignment

A major difference to the final state particles $p\pi^-$ in $\Lambda_b \rightarrow D^0 p\pi^-$ is that the origin vertex of $p\pi^-$ in $\Lambda_b \rightarrow J/\psi \Lambda$ is well displaced, even for LL tracks. Large values of the z -position of the Λ decay vertex correspond to the case that less layers of the VELO are passed by the final state particles and the sensitivity to misalignment effects become stronger. We therefore consider the double ratio of recorded data and MC simulated events for $\eta < 3.25$ and $\eta \geq 3.25$:

$$\frac{d}{dz} \left(\frac{\text{data, } \eta(\Lambda_b) < 3.25}{\text{MC, } \eta(\Lambda_b) < 3.25} / \frac{\text{data, } \eta(\Lambda_b) \geq 3.25}{\text{MC, } \eta(\Lambda_b) \geq 3.25} \right) < 0 \\ \Leftrightarrow \text{missing MC events for } \eta > 3.25 \text{ and large } z.$$

The normalized double ratio and a fit of a linear function $f(x|c_1, c_2) = c_1 + c_2 x$ is shown in Fig. F.1. The uncertainty band of the fit is calculated by ordinary error propagation

$$u_f = \sqrt{\sum_{ij} \left(\frac{\partial f}{\partial c_i} \frac{\partial f}{\partial c_j} \right)^2 \text{cov}(c_i, c_j)} = \sqrt{u_{c_1}^2 + x^2 u_{c_2}^2 + 2x \text{cov}(c_1, c_2)}.$$

No tension for a preferred slope is visible w.r.t. the given uncertainties.

F.3 Influence of χ^2_{DTF}

In order to exclude that the difference between LL and DD tracks is introduced by the χ^2_{DTF} selection criterion, we compare the distribution of the ratio of recorded data and simulated events, parametrized in $\eta(\Lambda_b)$, for $\chi^2_{\text{DTF}}/\text{DoF}$ below and above the respective 50 % quantiles of

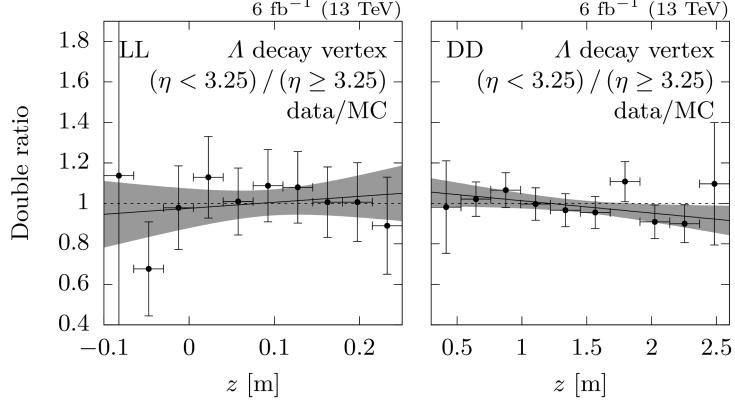


Fig. F.1. Normalized double ratio of $(\eta < 3.25)/(\eta \geq 3.25)$ and rec. data / MC sim. events as a function of the z -position of the Λ decay vertex for LL (left) and DD (right) tracks.

LL and DD tracks. The ratios are shown in Fig. F.2 and unveil the very same deviation that was observed previously between LL and DD tracks for small values of η in both 50 % quantiles. We therefore exclude the hypothesis that χ^2_{DTF} selections introduced the observed deviation.

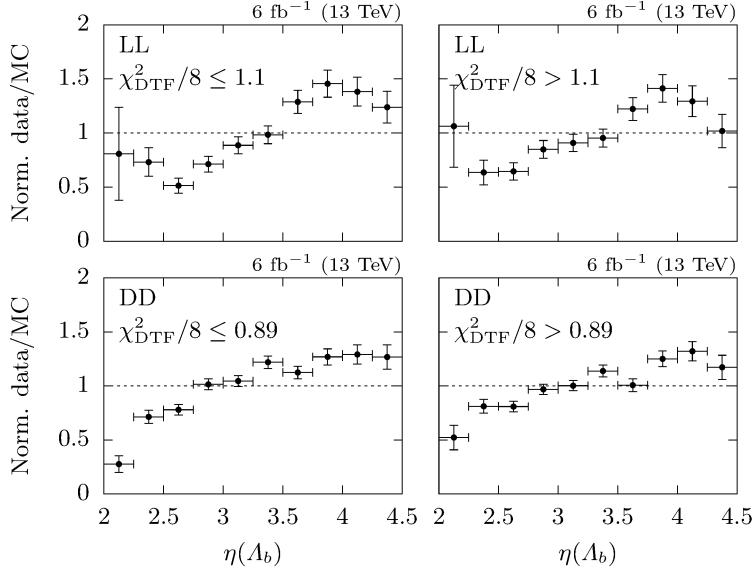


Fig. F.2. Ratios of recorded data and simulated events, parametrized in $\eta(\Lambda_b)$ for lower and upper 50 % quantiles of $\chi^2_{\text{DTF}}/\text{DoF}$.

F.4 Influence of temporal changes during data taking

MC simulated events are only available for the years 2015 and 2016, whereas we use the full data set of recorded data of Run 2 in order to increase the significance of the extracted weights. Since neither the center of mass energy, nor major changes to the apparatus separate the data recorded in the years 2015 and 2016 from those recorded in 2017 and 2018, the same set of

weights is expected when using recorded data of the years 2015 and 2016 only, albeit with larger uncertainties. In Fig. F.3 we show $w_2(\eta)$ only using recorded data of the years 2015 and 2016.

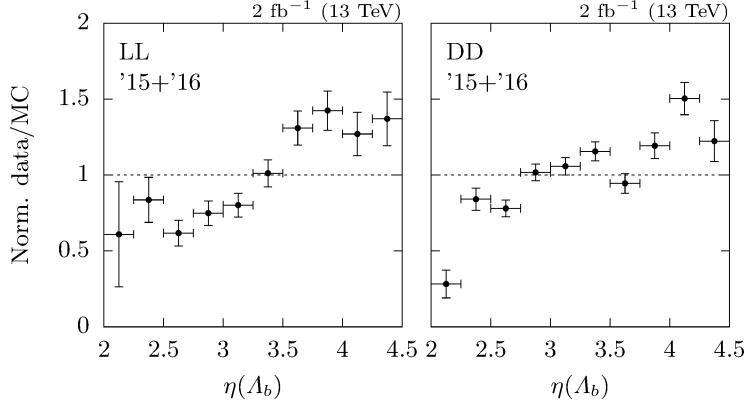


Fig. F.3. Weighting factor $w_2(\eta)$ for LL (left) and DD (right) tracks found by using recorded data of the years 2015 and 2016 instead of full Run 2 data.

Further, we check the impact of magnet polarity and splitting w.r.t. the sign of magnet polarity and proton charge (+1 for p and -1 for \bar{p}) on the observed accumulation of events for $\eta(\Lambda_b) \gtrsim 3.25$ and LL tracks, and show the corresponding distributions of $w_2(\eta)$ in Fig. F.4. The accumulation in the subsample for a negative product of proton charge and magnet polarity, $Q(p) \times \text{pol.} < 0$, appears pronounced in comparison with $Q(p) \times \text{pol.} > 0$, indicating an misalignment or dead pixels in the left detector hemisphere in a up- to downstream orientation. (Mag. up and mag. down are associated with +1 and -1 , respectively.) However, these effects are insignificant and it is unclear why they seem to be absent in $\Lambda_b \rightarrow D^0 p\pi^-$ but are not correlated with the z -position of the Λ decay vertex.

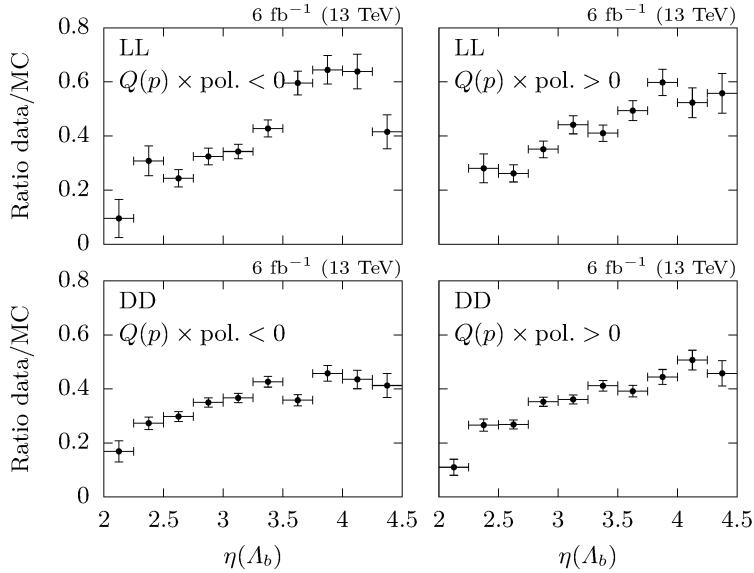


Fig. F.4. Samples split w.r.t. the sign of the product of proton charge and magnet polarity.

F.5 Summary

The reason for the deviation stays unclear. For the decay $\Lambda_b \rightarrow J/\psi \Lambda$ (L0 TIS triggered), this effect could be a statistical fluctuation, due to misalignment effects that are not reproduced well in the simulated events, or other reasons. The deviation w.r.t. $\Lambda_b \rightarrow D^0 p\pi^-$ decays (L0 TOS triggered) is significant but the reason also stays unclear. Further it is unclear, whether TOS triggered $\Lambda_b \rightarrow D^0 \Lambda$ will be distributed similar to $\Lambda_b \rightarrow J/\psi \Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$ or deviate from both. (It is worthwhile to mention that for perfect simulations there should not be any deviation in any of these samples.)

Appendix G

Confusion Matrix

Our definitions of *true* and *false*, as well as *positive* and *negative* in the context of classification throughout the present analysis are listed in the confusion matrix, shown in Fig. G.1. The matrix reads as following: a row listed the predicted class label and a column shows the genuine class label, *e.g.*, we refer to events that are predicted being of class *background*, but are actually genuine *signal* events as FN, whereas we refer to genuine background events, that are spuriously classified as *signal*, as FP.

		Actual	
		Sig.	Bkg.
Predicted	Sig.	TP	FP
	Bkg.	FN	TN

Fig. G.1. Confusion matrix as used in the present analysis for the classes *signal* (Sig.) and *background* (Bkg.). The abbreviations TP, FP, FN and TN refer to *true positive*, *false positive*, *false negative*, and *true negative*, respectively.

Appendix H

Cross-Validation

Advanced machine learning algorithms typically come with a set of hyper-parameters, *e.g.*, the C and γ parameter of SVMs or the number of trees in a random forest that are fixed before the learning process begins and are not objective of the training process itself. The choice of these values can impact the performance of the trained model and help convergence, yet optimal values are unknown *a priori*. In a canonical approach one splits a second test set and uses it to optimize the hyper-parameters since optimizing on the training (test) set could seed overfitting¹ (selection bias) which clearly comes by the costs of a reduced training set. Cross-validation (*e.g.*, Refs. [125, 126]) is an alternative technique that allows to optimize hyper-parameters on the training set without suffering from overfitting. In the present analysis we use a 5-fold cross-validation scheme where the training set is partitioned into five, equally sized folds (*cf.* Fig. H.1). Hyper-parameters are optimized on each of these folds by training the classifier on four and evaluated on the last fold (4+1). The loss of accuracy in terms of a large uncertainty due to the small sample size of each of the folds used for the evaluation is compensated by combining the results of the five independent folds.

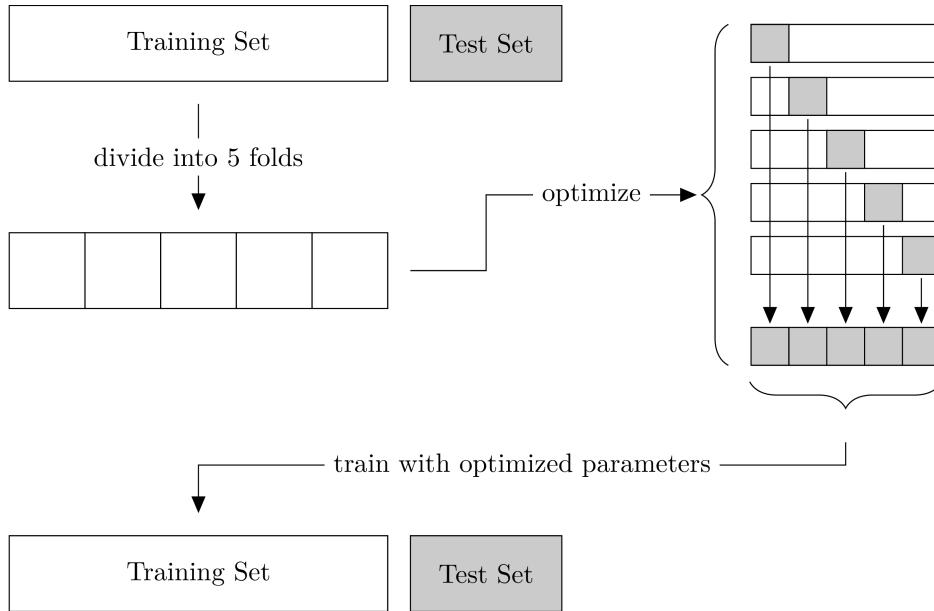


Fig. H.1. Outline of a 5-fold cross-validation where hyper-parameters are optimized on folds of the training set instead of the test set.

¹Especially when scanning the hyper-parameter space exhaustively, this problem is similar to the so-called *look elsewhere effect* [124].

Appendix I

Principal Component Analysis

The Principal Component Analysis (PCA) is an unsupervised machine learning technique which is often used to reduce the dimensionality of a given feature set. A data set is decorrelated by diagonalization of the respective covariance matrix and subsequently ordered by the standard deviation of each feature. A reduction is achieved by pruning features with a low standard deviation (after being decorrelated).

The latter step is driven by the motivation that features with low standard deviations carry less information than is needed to describe the distribution of the data. We note, however, that in classification tasks the information needed to distinguish between a given set of categories can be dominantly encoded in the features with a low standard deviation. We show such an example in Fig. I.1.

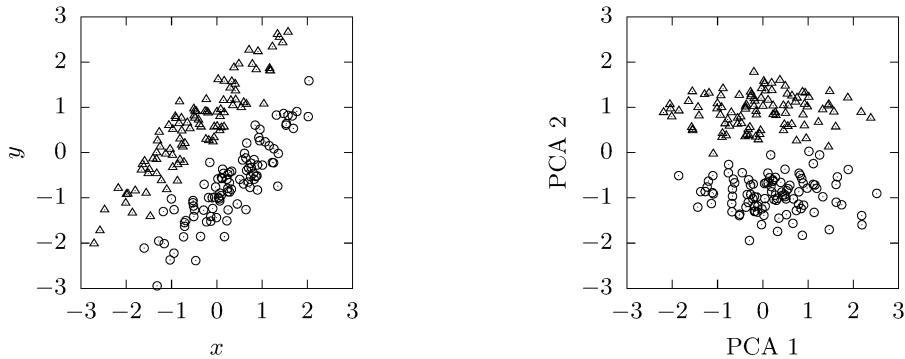


Fig. I.1. Distribution of the two features x, y of a joint data sample of two categories (circle and triangle). The canonical PCA transformation decorrelates the distribution by introducing rotated features PCA 1 and PCA 2, ordered by their respective standard deviations.

In the given example, PCA 1 clearly is not helpful at all for distinguishing between the two categories, whereas PCA 2, albeit having a lower standard deviation, separates both distributions significantly. We propose to use the l_1 Wasserstein distance for ordering the PCA components instead of the standard deviation. As opposed to the more general p -th Wasserstein distance, the calculation of the l_1 distance is computational feasible, *i.e.*,

$$l_1(u, v) = \int dx |U(x) - V(x)|,$$

where U and V are the cumulative distributions of u and v , respectively (*cf.* Ref. [127]). In Fig. I.2 we show the l_1 distance for our pseudo-experiment. The l_1 distance appear to be a much better criterion for sorting PCA components than the standard deviation in the canonical

approach. We note that in direct comparison with other metrics, such as the Kolmogorov-Smirnov test, the Wasserstein l_1 distance is also much more stable (numerically) for distributions that differ strongly.

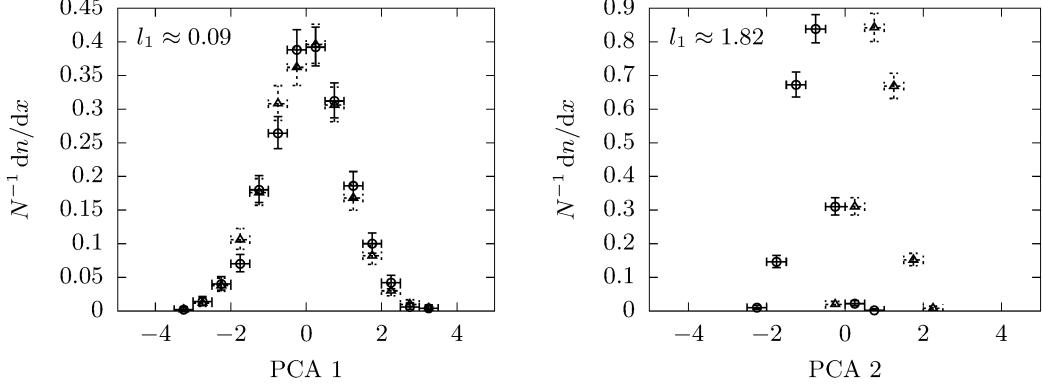


Fig. I.2. Distribution of PCA 1 (left) and PCA 2 (right). Even though PCA 1 has a larger standard deviation, it is not helpful at all for distinguishing between the two categories (circle and triangle). In our proposed solution the Wasserstein metric l_1 is used which rank PCA 2 well above PCA 1 in this case.

Appendix J

Support Vector Machines

Support Vector Machines (SVMs) are supervised machine learning models that are used in their canonical form for linear (binary¹) classification and regression, as well as for anomaly detection (*e.g.*, Ref. [129]). In its simplest form (hard margin), the training objective of an SVM for a binary classification task is the parametrization of a hyperplane with normal vector \vec{w} and offset b (*i.e.*, \vec{y} in plane $\Leftrightarrow \vec{w} \cdot \vec{y} + b = 0$) such that $|\vec{w}|_2$ becomes minimal and all instances of different classes are separated by the hyperplane, *i.e.*,

$$t_i (\vec{w} \cdot \vec{x}^{(i)} + b) \geq 1 \quad \forall i,$$

where $x^{(i)} \in \mathbb{R}^n$ is the feature vector of the i -th instance and t_i is given by the respective label vector $y_i \in \mathbb{R}^n$,

$$t_i := \begin{cases} +1 & \text{if } y_i \sim \text{signal}, \\ -1 & \text{if } y_i \sim \text{background}. \end{cases}$$

Instances with the minimal distance $t_i (\vec{w} \cdot \vec{x}^{(i)} + b) = 1$ are considered to lie on the *margin* and are thus referred to as the *supporting vectors* of the classifier (hence the name). Obviously, the hard margin problem is only solvable for linearly separable data. By introducing slack variables $\zeta_i \geq 0$ for each instance i , this constraint is relaxed

$$\underset{\vec{w}, b, \vec{\zeta}}{\operatorname{argmin}} \frac{1}{2} |\vec{w}|_2^2 + C |\vec{\zeta}|_1 \quad \text{subject to } t_i (\vec{w} \cdot \vec{x}^{(i)} + b) \geq 1 - \zeta_i \text{ and } \zeta_i \geq 0 \quad \forall i. \quad (\text{J.1})$$

This relation makes the slack vector ζ_i ($\vec{\zeta} \in \mathbb{R}^m$) interpretable as a measure of the margin violation of the i -th instance. The objective of the optimization thus reads as the simultaneous maximization of the margin, $\operatorname{argmin}_{\vec{w}} |\vec{w}|_2 \equiv \operatorname{argmin} |\vec{w}|_2^2$, and minimization of the margin violations, $\operatorname{argmin} |\vec{\zeta}|_1$. The relative weight between those contrary optimization goals is given by the regularization parameter C .

Practically, C is a hyper-parameter that controls the influence of outliers and regularizes the decision boundary of an SVM. In Fig. J.1 we show the decision boundaries of (linear) SVMs that were trained on randomly generated data with $C = 1$ and $C = 10$. The generated data set is partitioned w.r.t. two classes (*circle* and *triangle*) and appended by one outlier instance of class *circle* at $(x, y) = (1.5, -0.5)$. The influence of this outlier is tested by training one classifier with the full dataset and the other one on a reduced data set where the outlier was removed.

By their very nature, SVMs are linear classifiers for a given feature space \vec{x} . The dimensionality can yet be increased by encoding higher order combinations such as x_i^2 or $x_i x_j$ in \vec{x} itself,

¹Multiclass classification can be achieved via techniques such as *one-against-all* or *one-against-one* [128].

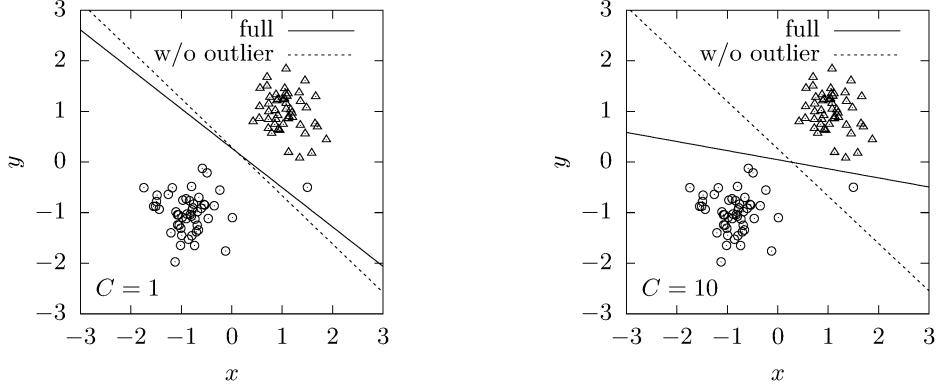


Fig. J.1. Decision boundaries of linear SVMs with different values for their respective regularization parameter C , trained on either the full data set (solid line) or a data set where the outlier at $(x, y) = (1.5, -0.5)$ was removed (dashed line).

e.g.,

$$\vec{x} \equiv \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{\vec{\phi}} \vec{\phi}(\vec{x}) = \begin{pmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix}.$$

The optimization problem Eq. (J.1), as well its corresponding Lagrangian dual are quadratic programming (QP) problems. The latter QP however, only depends on inner products $\vec{x}^{(i)} \cdot \vec{x}^{(j)}$ as opposed to Eq. (J.1) that depends on $\vec{x}^{(i)}$ itself. This allows the application of the Mercer's theorem that guarantees the existence of a kernel function $K(\vec{x}^{(i)}, \vec{x}^{(j)})$ such that

$$\vec{x}^{(i)} \cdot \vec{x}^{(j)} \xrightarrow{\vec{\phi}} \vec{\phi}(\vec{x}^{(i)}) \cdot \vec{\phi}(\vec{x}^{(j)}) \equiv K(\vec{x}^{(i)}, \vec{x}^{(j)}),$$

if $\vec{\phi}$ respects the Mercer's condition. In the present analysis we frequently use the (Gaussian) RBF kernel,

$$K(\vec{x}^{(i)}, \vec{x}^{(j)} | \gamma) \equiv \exp \left\{ -\gamma \left| \vec{x}^{(i)} - \vec{x}^{(j)} \right|_2^2 \right\},$$

which corresponds to a function $\vec{\phi}(\vec{x})$ that maps \vec{x} into an infinite-dimensional space and thus leverages the training of complex, non-linear classifier using SVMs.

For m training instances and n features, the complexity of training SVMs scales between $\mathcal{O}(m^2 \times n)$ and $\mathcal{O}(m^3 \times n)$, due to the inversion of the kernel matrix which makes SVMs, even if they are versatile machine learning models on small to medium sized data sets, impractical to use on large data sets.

Appendix K

Ensemble Learning

Trees sprout up just about everywhere in computer science.

— Donald Knuth, in his book *Combinatorial Algorithms* (4A).

Ensemble learning techniques build upon the observation that an ensemble of weak learners, *i.e.*, classifiers that were not provided with the complete set of available information during training (*e.g.*, in terms of number of training instances or number of features), can outperform single classifiers that were provided with the entire available information, when aggregated properly.

A straightforward realization of an ensemble learner is called *stacking* [130] where the output of trained classifiers (tier 1) is fed into another classifier (tier 2). This tier 2 classifier aggregates the outputs of tier 1, uses them as its feature set and outputs the final decision.

Another kind of ensemble learning is achieved by bagging (short for bootstrap aggregation) [131] where the same classifier is trained on different random subsets of the training set. Additional randomness (and thus a greater diversity) is achieved by only using certain random subsets of the available feature set. This again increases the diversity among the classifiers which typically increase the bias of each classifier but decreases the overall bias and variance of the aggregation. When bagging and feature sampling is used for growing decision trees, the aggregation (typically referred to as the *forest*), is called a *Random Forest* [132]. Each tree of such a forest is trained by optimizing thresholds which partition the data set at each node. The thresholds are optimized w.r.t. a certain criterion (*e.g.*, Gini impurity or entropy¹) where first, the best thresholds for each of the given features is found and secondly, the overall best threshold among the optimized thresholds is used for partitioning. If the first step is replaced by drawing a random threshold, the forest is called (a forest of) *Extra Trees* (short for *Extremely Randomized Trees*) [133].

Boosting is yet another ensemble learning technique where sequentially trained classifiers try to correct their predecessor. When using decision trees, two major kinds of boosting are common, *i.e.*, *Gradient Boosting* and *Adaptive Boosting* [134, 135]. For the former (typically referred to as *BDT*) regression trees are fitted to the residual errors of the predecessor which makes the aggregation the linear sum of all trained classifiers. The latter (typically referred to as *Ada. BDT*) uses weights instead of the residuals in order to increase the focus on misclassified events in subsequent evaluations. Note that in order to deal with residuals, regression trees have to be used rather than ordinary binary decision trees and thus mean squared errors (MSE) are used as a criterion. Technically, the aggregation is achieved by summing the decisions of the sequentially trained classifiers in a loss function. For BDTs using gradient and adaptive boosting the loss function is referred to as *deviance* and *exponential*, respectively. More information is given in the literature, in particular in Ref. [136].

¹Gini impurity $G = 1 - \sum_k p_k^2$ and entropy $H = -\sum_k p_k \ln p_k$, where $p_k \in [0, 1]$ is the fraction of class k in a given set.

Appendix L

MVA Supplementaries

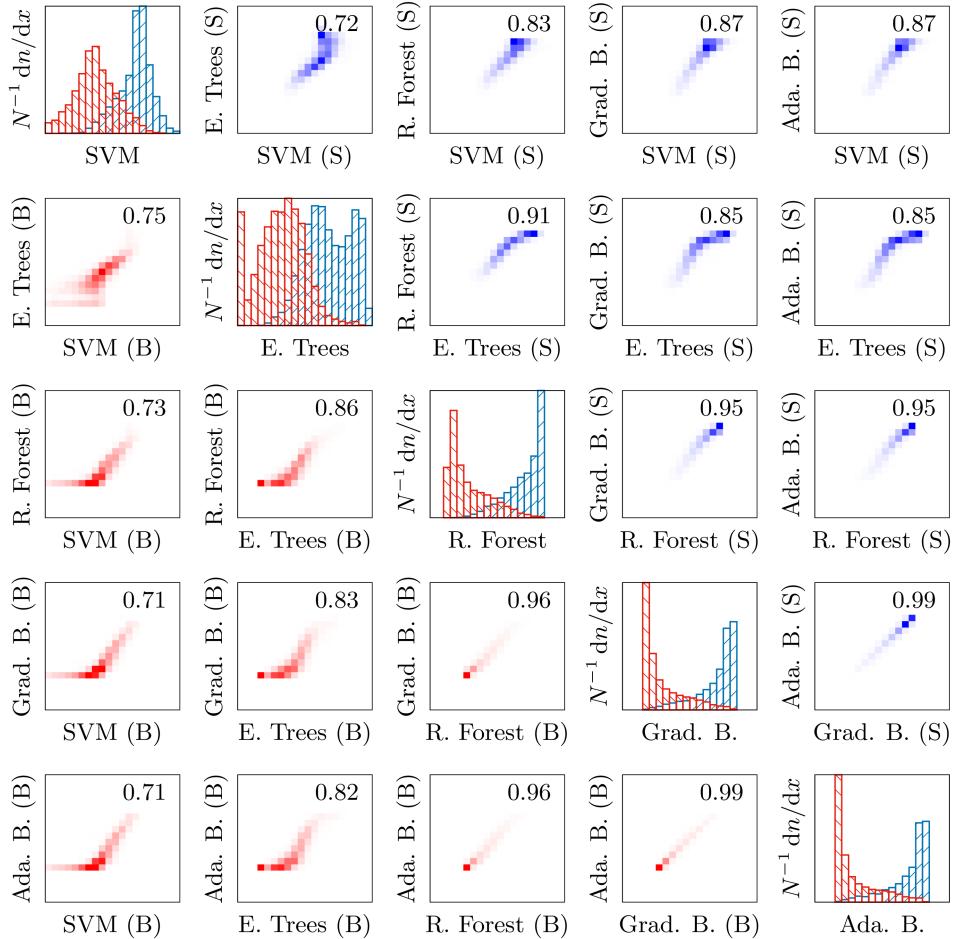


Fig. L.1. Distributions of the responses and correlations of the tier 1.1 Λ classifiers for LL tracks, separated for signal (S) and background (B). On top of the correlation distributions we show the PCC as a measure for the linear correlation.

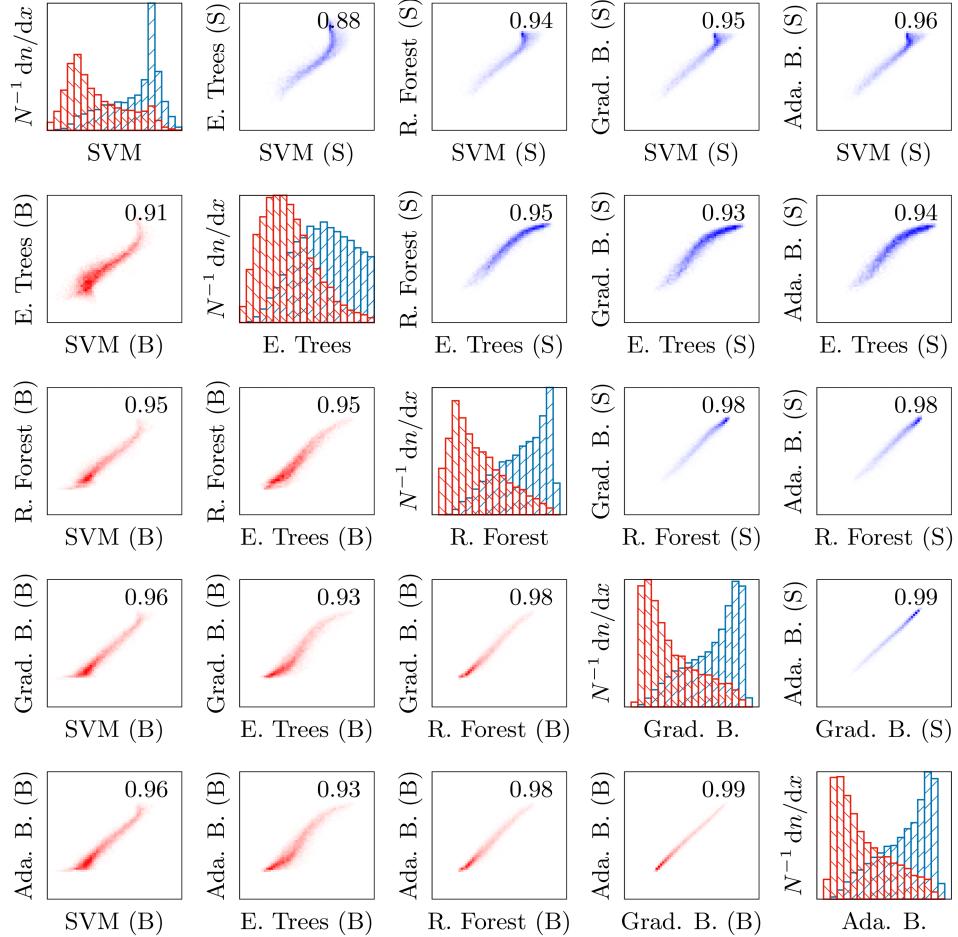


Fig. L.2. Distributions of the responses and correlations of the tier 1.1 Λ classifiers for DD tracks, separated for signal (S) and background (B). On top of the correlation distributions we show the PCC as a measure for the linear correlation.

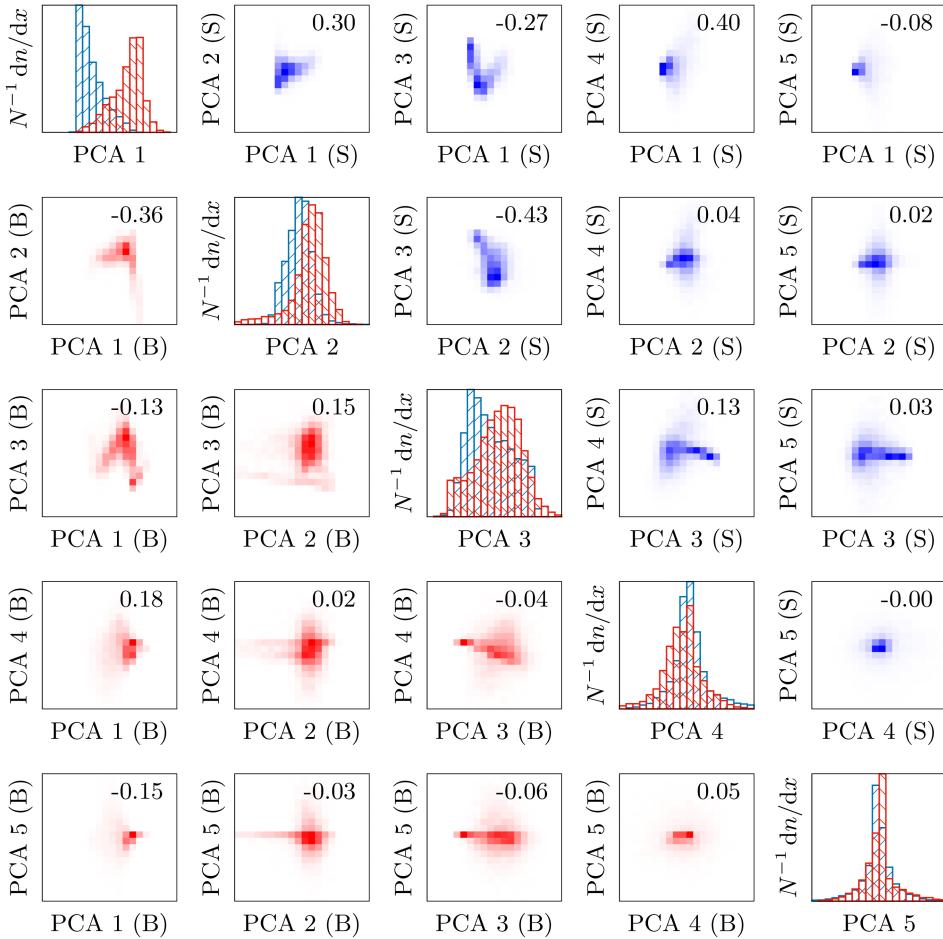


Fig. L.3. Distributions of the responses and correlations of the tier 1.1 Λ classifiers for LL tracks after PCA transformation, separated for signal (S) and background (B). On top of the correlation distributions we show the PCC as a measure for the linear correlation.

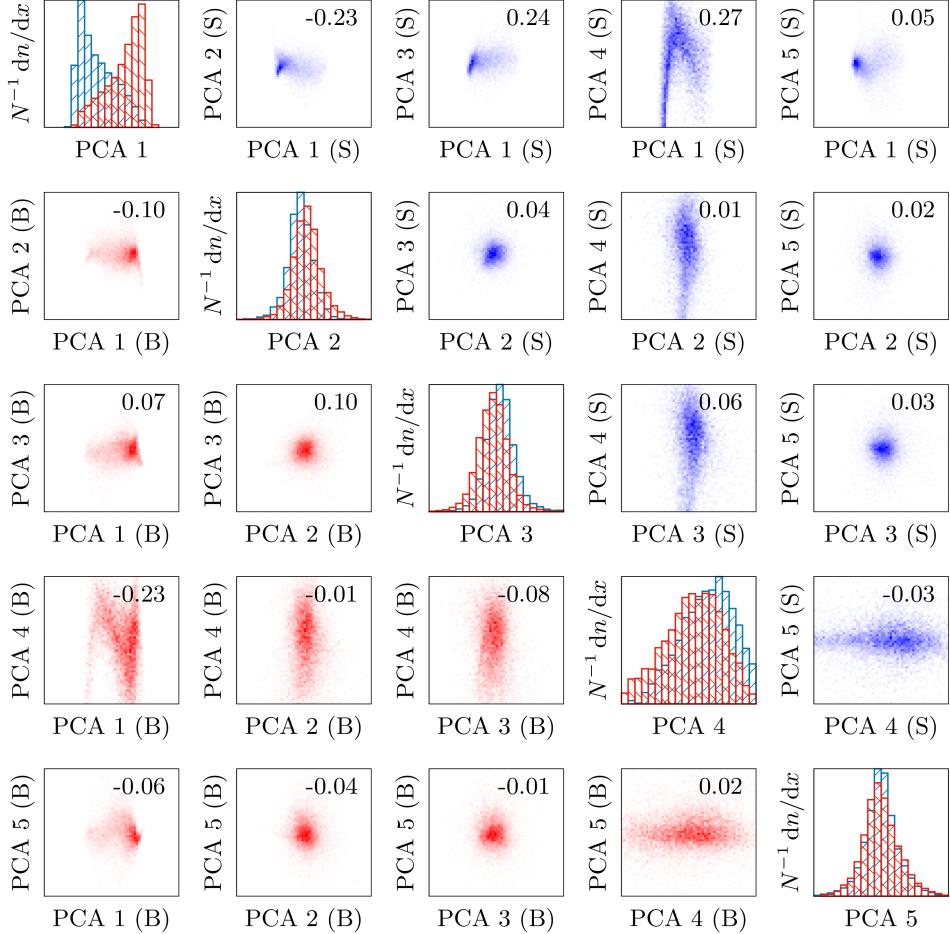


Fig. L.4. Distributions of the responses and correlations of the tier 1.1 Λ classifiers for DD tracks after PCA transformation, separated for signal (S) and background (B). On top of the correlation distributions we show the PCC as a measure for the linear correlation.

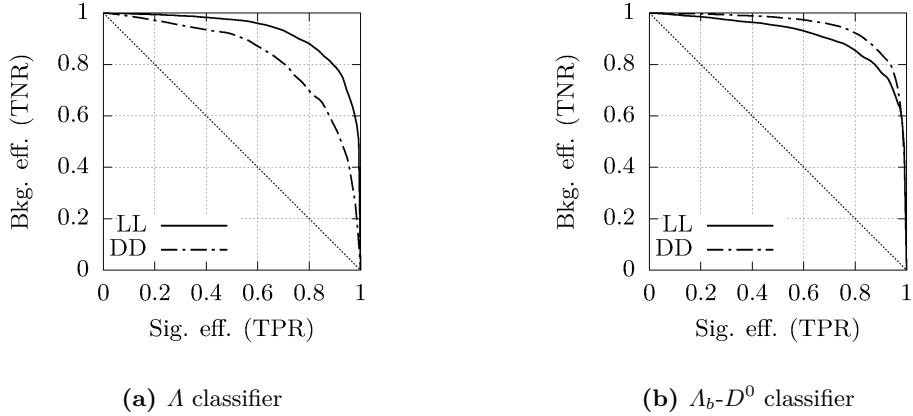


Fig. L.5. ROC curves of the SVMs used as the Λ classifier (left) and Λ_b - D^0 classifier (right) for LL and DD tracks.

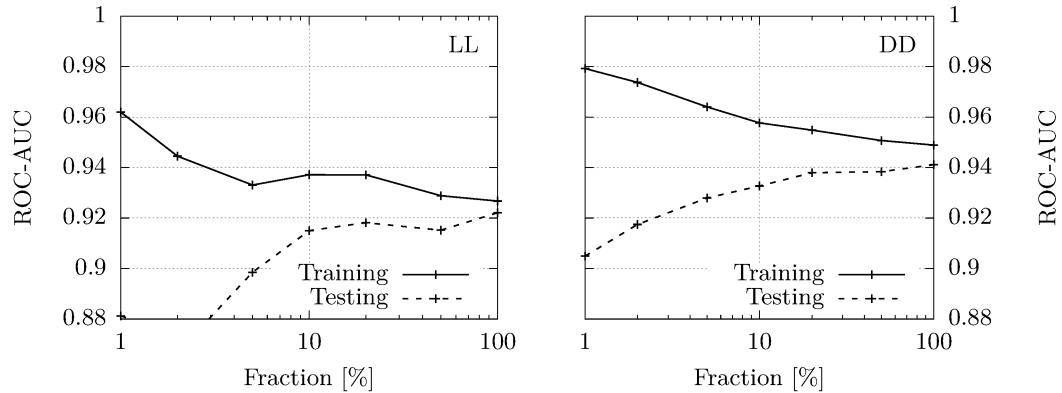


Fig. L.6. Convergence of the SVMs (Λ_b - D^0 classifier) given by the ROC-AUC values for different sample sizes where a sample size of 100 % corresponds to the size of the full trainings set. The ROC-AUC is evaluated via a 5-fold cross-validation scheme on the training (solid line) and testing folds (dashed line).

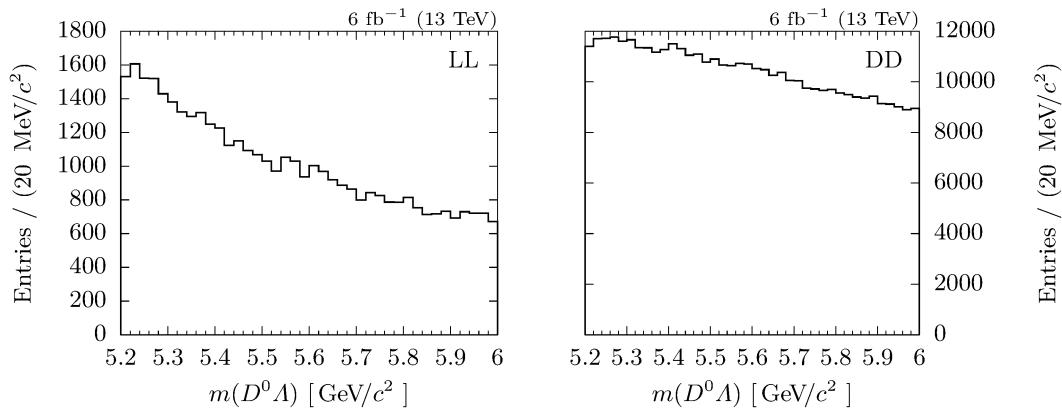


Fig. L.7. Combined invariant mass of D^0 and Λ candidates after loose selection from recorded data, as used for training the Λ_b - D^0 classifier (background class).

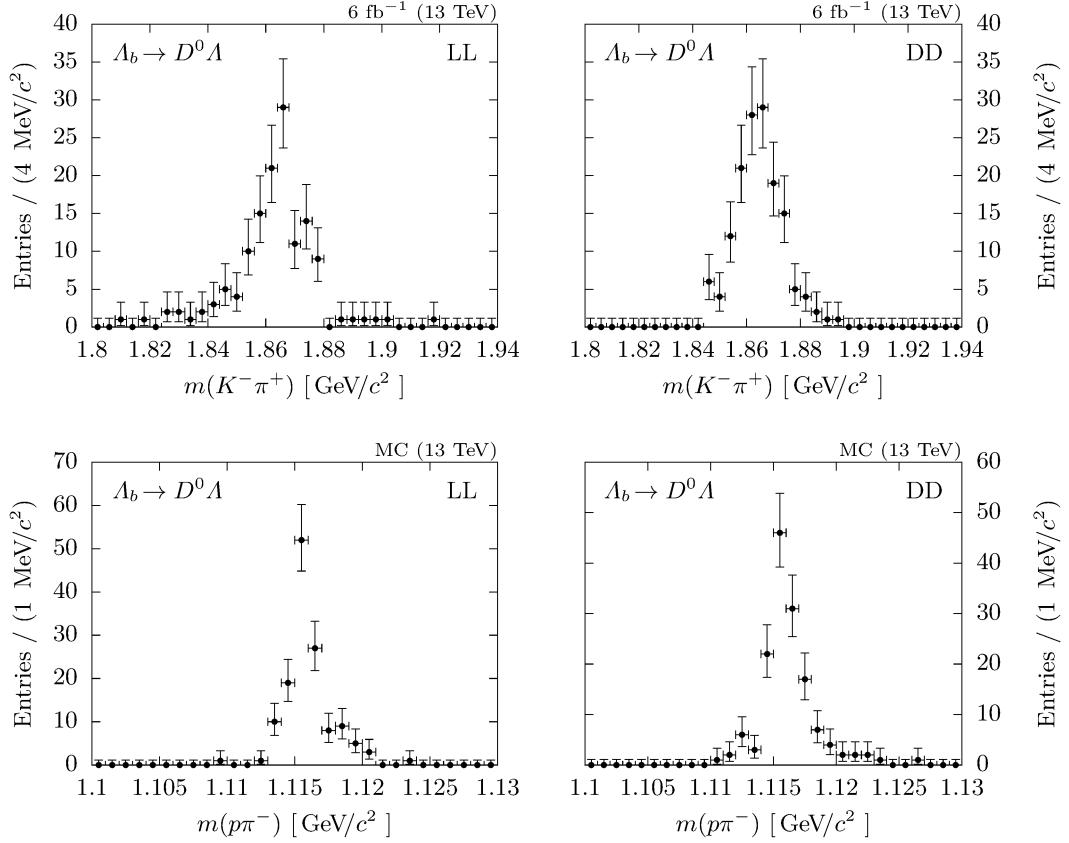


Fig. L.8. Invariant masses of the D^0 candidates (top) and A candidates (bottom), reconstructed from recorded data that are classified as *signal* $A_b \rightarrow D^0 A$ decays by the final MVA classifier.

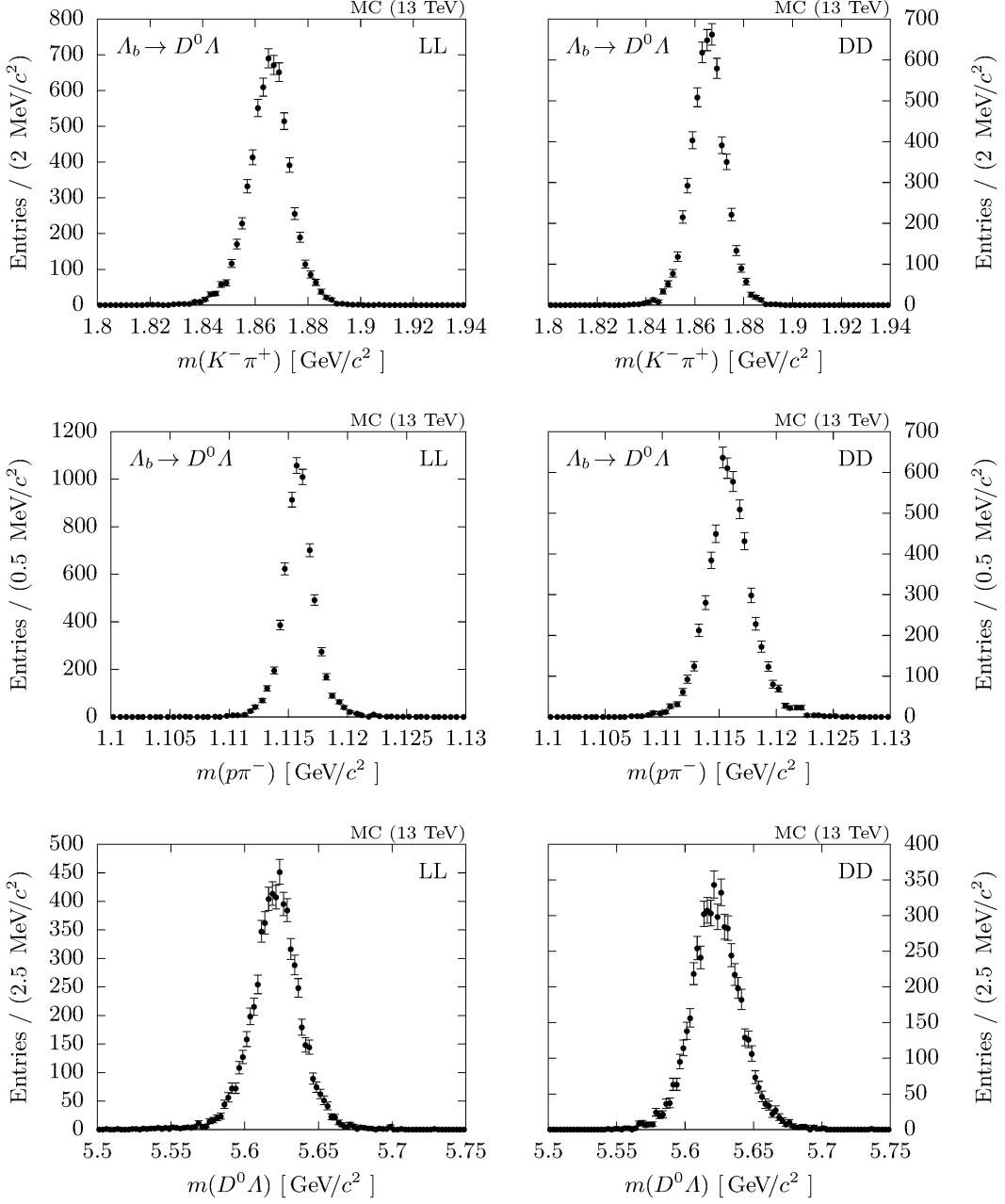


Fig. L.9. Invariant masses of MC simulated D^0 , Λ and Λ_b candidates (top to bottom), truth-matched as genuine $\Lambda_b \rightarrow D^0 \Lambda$ decays that are classified as *signal* $\Lambda_b \rightarrow D^0 \Lambda$ decays by the final MVA classifier.

Appendix M

Cross-Checking the Efficiency Estimates of MVA Features

Dowerjai, no prowerjai.

— Vladimir Lenin, allegedly.

In the realm of machine learning, interpretability of trained models has become an important topic and new explanation methods pop up frequently, *e.g.*, Refs. [137, 138, 139]. In HEP applications and in particular in the context of (binary) classification problems where training instances of the labels *signal* and *background* are drawn from separate sources, *i.e.*, from MC simulated events (signal) and recorded data (background), the question arises to what degree these training data are representative of the real data. Again, various approaches came up lately to tackle this question when the training data are not representative of the real data, *e.g.*, Refs. [140, 141]. In the context of the present analysis, we limit ourself to the study of the fidelity of the classifier responses that we evaluate with MC simulated events in order to obtain the efficiencies of our classification. If certain nuisance parameters would be present, which would help our classifiers to learn to distinguish recorded data from simulated events rather than signal and background signatures (in recorded data), the classifiers, as well as estimated efficiencies would be worthless.

To keep our model verifiable w.r.t. the estimated efficiencies without using sophisticated yet complex approaches as proposed in Ref. [140], we split our classifiers into two disjoint sub-classifiers (*cf.* Sec. 5.2) which can now be verified using a data driven approach. These two classifiers, the Λ classifier and the Λ_b - D^0 classifier, are applied to recorded $\Lambda_b \rightarrow J/\psi \Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$ candidates which we used previously for estimating the calibration factors (*cf.* Chap. 4) and the normalization (*cf.* Chap. 6), respectively. The rich statistics and the clean data samples allow an efficiency estimation using recorded data, leveraging the direct comparison of these figures with the predictions from simulated events.¹

In Sec. M.1 we outline our strategy for extracting the efficiencies. We will find a sufficient fidelity of all used classifiers. The proxy modes also allow an estimation of the fidelity of the DTF probability distribution where we witness a large deviation for DD tracks. The deviation, as well as the implication are discussed in Sec. M.2.

M.1 Efficiency Estimation

The efficiency ε_f of a feature f is defined as the ratio

$$\varepsilon_f := \frac{n_F}{n_{F \setminus f}},$$

¹This approach implicitly assumes common fidelity issues of the respective features among the simulated Λ_b decays.

where n_F is the amount of events that are left after requiring the full selection F (see below) and $n_{F \setminus f}$ is the corresponding amount if feature f is left out. The definition of the full selection F depends on the proxy mode. Whether or not a feature is included in F of a given proxy mode is listed in Tab. M.1, as well as the estimated efficiencies. If a given feature f is included in F , it obeys the selection thresholds of the dedicated $\Lambda_b \rightarrow D^0 \Lambda$ tight selection (*cf.* Sec. 5.2), except for the DTF probability of the $\Lambda_b \rightarrow D^0 p\pi^-$ proxy. (The performance of the DTF probability is discussed separately in Sec. M.2.) The amounts n_F and $n_{F \setminus f}$ are obtained using the outlined

Tab. M.1. Efficiency estimates of the tier 2 features, obtained from recorded data and MC simulated events, as well as the corresponding (daughters of the) proxy modes. The efficiencies ε_f are defined as the ratio of the amounts n_F and $n_{F \setminus f}$, where n_F is the amount of events that are left after requiring features to obey the dedicated $\Lambda_b \rightarrow D^0 \Lambda$ tight selection (if having a counter part in the proxy mode) and $n_{F \setminus f}$ is the corresponding amount if feature f is left out. In the last column we list the residuals of the ratios which are taken as systematic deviations due to an imperfect simulation fidelity. In the bottom row we give the sum in quadrature of these uncertainties.

Feature	Proxy	Rec. data [%]		MC sim. [%]		1 – ratio [%]	
		LL	DD	LL	DD	LL	DD
Λ Clf.	$J/\psi \Lambda$	85.5(13)	23.4(6)	85.8(7)	23.82(19)	0.4(18)	1.6(26)
Λ_b - D^0 Clf.	$Dp\pi$	80.6(7)	38.8(4)	78.35(6)	37.79(5)	-2.9(8)	-2.8(11)
DTF	$J/\psi \Lambda$	85.4(14)	59.3(17)	87.9(7)	72.6(7)	2.9(18)	18.3(25)
DTF	$Dp\pi$	94.1(7)	94.4(11)	97.96(9)	97.67(18)	3.9(8)	3.3(12)
ProbNNp(p)	$J/\psi \Lambda$	90.5(13)	81.9(32)	94.4(8)	83.7(8)	4.1(16)	2(4)
ProbNNk(K)	$Dp\pi$	89.8(7)	81.2(10)	91.23(8)	84.26(15)	1.8(8)	3.6(12)
						7 %	19 %

fit approaches in Sec. 4.1.3 and Sec. 6.2. In order to test the sensitivity of the calibration of recorded $\Lambda_b \rightarrow D^0 p\pi^-$ data (*cf.* Sec. 6.1.3) we estimate the efficiency with and without calibration and find deviations below 5 %, *i.e.*, our feature selection only depends weakly on resonance structures. In Fig. M.1 we show the invariant mass distributions used for extracting the efficiency of the ProbNNk classifier as an example. Since the efficiency is obtained by taking the ratio of fitted yields, the bias of the fit model as discussed in Sec. 6.2 only plays a minor role in the ratio. In the case of $\Lambda_b \rightarrow J/\psi \Lambda$, the subset of the dedicated $\Lambda_b \rightarrow D^0 \Lambda$ tight selection, even though not optimized for the proxy mode, suppresses almost all background events as shown in Fig. M.2, again allowing a low-bias extraction of the feature efficiencies.

Deviations from one in the ratios of ε_f of recorded data and simulated events are considered as systematic uncertainties due to fidelity issues of simulated events. The sum in quadrature of the residuals (as listed in Tab. M.1) is used as a conservative approximation of the overall uncertainty.

M.2 Fidelity of the DTF Probability Distribution

Splitting our handcrafted classifier into two parts, *i.e.*, the Λ and Λ_b - D^0 classifier, leverages the outlined cross checking of the estimated efficiency in the proxy modes $\Lambda_b \rightarrow J/\psi \Lambda$ and $\Lambda_b \rightarrow D^0 p\pi^-$, respectively. This assumption of similar distributions among our primary mode $\Lambda_b \rightarrow D^0 \Lambda$ and the proxy modes does not hold for the χ^2_{DTF} distribution of the DTF if they are

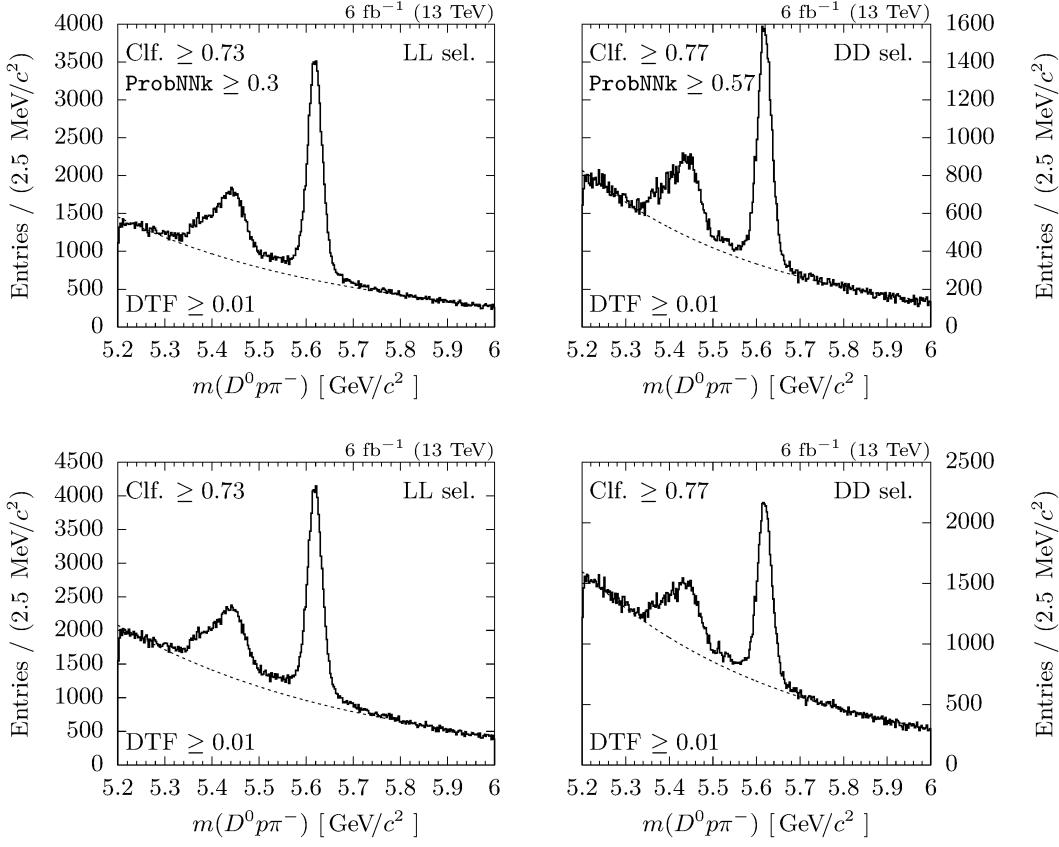


Fig. M.1. Background fits to the invariant mass $m(D^0 p \pi^-)$ used for extracting the signal yield of $\Lambda_b \rightarrow D^0 p \pi^-$ decays. The abbreviations *LL sel.* and *DD sel.* refer to the choice of thresholds which obey the dedicated $\Lambda_b \rightarrow D^0 \Lambda$ tight selection for LL and DD tracks, respectively. The yields are used to estimate the efficiency of the **ProbNNk** classifier by taking the ratio of the fitted yields n_F (top) and $n_{F \setminus f}$ (bottom).

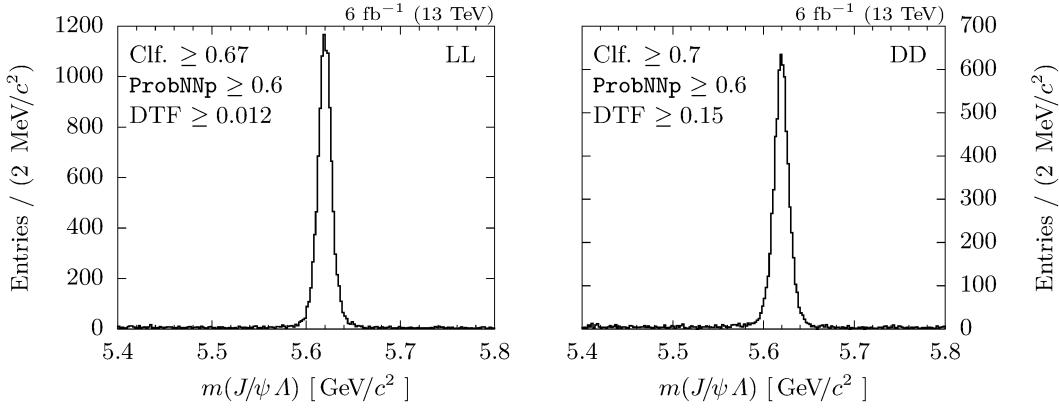


Fig. M.2. Invariant mass distribution of J/ψ and Λ candidates, refined via a DTF, as used for estimating n_F ($\Lambda_b \rightarrow J/\psi \Lambda$ proxy mode).

not genuinely χ^2 -distributed with the correct DoF.² On the one hand, the DTF probability is not uniformly distributed as discussed in Appx. D. On the other hand we have no other way to cross-check the efficiency of the DTF probability and therefore will stick to this approach anyhow and consider the results as approximations.

As listed in Tab. M.1 we determined ε_f for the DTF probability in both proxy modes where we took the thresholds of the dedicated $\Lambda_b \rightarrow D^0 \Lambda$ tight selection for the $\Lambda_b \rightarrow J/\psi \Lambda$ proxy and 0.01 for the $\Lambda_b \rightarrow D^0 p\pi^-$ proxy. The choice of the latter is rooted in a correlation between the DTF probability and the bias of the respective fit model. We consider this a minor issue due to redundancies in the $\Lambda_b \rightarrow J/\psi \Lambda$ mode. Further, we find that the large deviation for DD tracks dominantly is caused by large χ^2_{DTF} values (see below) and thus would likely have shown up also for this lowered threshold, if present.

In Fig. M.3 we show the fit that we used to extract $n_{F \setminus f}$ for the DTF probability (DD tracks). Regarding the logarithmic y -axis we exclude that the deviation is introduced by a bias of the fit model. In Fig. M.4 we show the cumulative distribution of the DTF probabilities for LL and DD tracks. We see that the deviation for DD tracks comes from an initial offset at large χ^2_{DTF} values and then keeps this difference. Based on this we conclude that this effect has to be attributed to the Λ baryon itself, due to the good agreement for LL tracks in both proxy modes. Since daughters of track type DD of the Λ baryons are the only particles in our consideration whose momentum information is not taken from the VELO (by definition), this issue is likely to be rooted in an incomplete error matrix for DD tracks. Correcting this error matrix is out of scope of the present analysis and we will therefore take the full deviation as an systematic uncertainty.

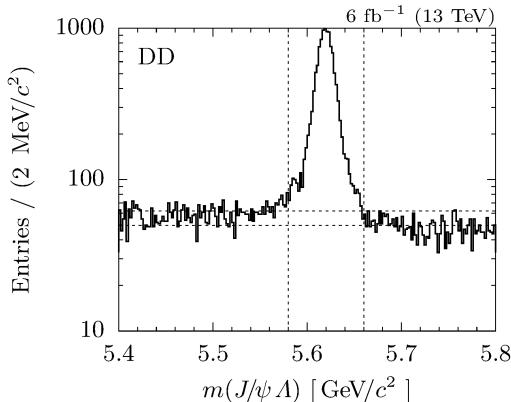


Fig. M.3. Combined invariant mass of J/ψ and Λ candidates used to extract the signal yield $n_{F \setminus f}$ for the DTF probability (DD tracks) via sideband subtraction. The sideband subtraction is evaluated twice, once using only the lower sideband, and a second time using only the upper sideband. The difference between both evaluations is taken as a systematic uncertainty. We note that regarding the large signal to background ratio, biases of the fitting technique affect the numerical value of the yield only slightly.

²If χ^2 distributed with the correct DoF the distributions transform under Eq. (D.1) to uniform distributions and are thus equal.

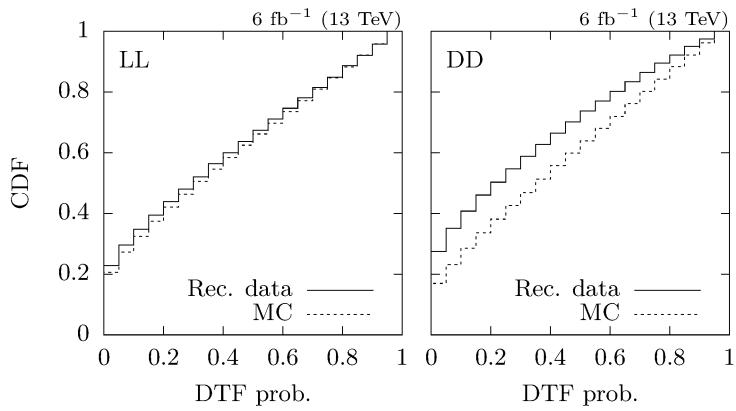


Fig. M.4. Cumulative distribution function (CDF) of the DTF probability distributions for LL and DD tracks. The CDFs are extracted via sideband subtraction and are thus binned. The solid and dashed line refer to recorded data and simulated events, respectively. The y -value of a bin with high-edge x thus corresponds to the sum of all events with DTF probability $\leq x$. In particular, the first bin includes the effects of χ^2_{DTF} cut-offs at large values.

Appendix N

Kinematics of the Charmless Three-Body Decay $\Lambda_b \rightarrow \Lambda K^+ K^-$

The charmless three-body decay $\Lambda_b \rightarrow \Lambda K^+ K^-$ appears as a peaking background in $m(D^0\Lambda)$ if not sufficiently suppressed. In Sec. 7.3.1 we establish such a suppression and benefit from some particular kinematic features we want to discuss in this chapter. First, we discuss the distribution of the reflected mass $m(\Lambda K^-\pi^+)$ in Sec. N.1 and introduce some notation. We then study the momentum distribution of the Λ baryon in Sec. N.2 which leveraged the strong separation power of the Λ classifier.

N.1 Mass Distributions

For describing the relativistic kinematic of three-body decays $M \rightarrow 1 2 3$ one often uses combined squared masses of two daughters, *e.g.*, m_{12}^2 . Following the notation of the PDG, these squared masses are given by the four momenta p_i of the daughters

$$m_{ij}^2 := (p_i + p_j)^2 ,$$

and obey useful relations, such as

$$m_{12}^2 + m_{23}^2 + m_{13}^2 = M^2 + m_1^2 + m_2^2 + m_3^2 , \quad (\text{N.1})$$

where M (m_i) refers to the mass of the decaying mother (daughter i). Eq. (N.1) can be rewritten

$$m_{12}^2 = M^2 + m_3^2 - 2ME_3 ,$$

where E_3 is the energy of particle 3 in the rest frame (CMS) of M and thus allows the determination of the kinematic boundaries of m_{ij}^2 , *e.g.*,

$$\begin{aligned} \max m_{23}^2 &= (E_2^* + E_3^*)^2 - \left(\sqrt{(E_2^*)^2 - m_2^2} - \sqrt{(E_3^*)^2 - m_3^2} \right)^2 , \\ \min m_{23}^2 &= (E_2^* + E_3^*)^2 - \left(\sqrt{(E_2^*)^2 - m_2^2} + \sqrt{(E_3^*)^2 - m_3^2} \right)^2 , \end{aligned}$$

where

$$\begin{aligned} E_2^* &= \frac{m_{12}^2 - m_1^2 + m_2^2}{2m_{12}} \quad \text{and} \\ E_3^* &= \frac{M^2 - m_{12}^2 - m_3^2}{2m_{12}} \end{aligned}$$

are the energies of particle 2 and 3 in the rest frame of m_{12} .

In the context of, the present analysis the three-body decay $\Lambda_b \rightarrow \Lambda K^- K^+$ appears as a (non-resonant) reflection in the $m(D^0\Lambda)$ distribution. More generally, a reflection occurs if w.l.o.g. particle 3 with genuine mass m_3 is reconstructed using the spurious mass hypothesis m_x . If so, the reconstructed invariant mass of the mother in the CMS \tilde{M}_{cms} is given by

$$\begin{aligned}\tilde{M}_{\text{cms}} &= \sqrt{m_1^2 + \vec{p}_1^2} + \sqrt{m_2^2 + \vec{p}_2^2} + \sqrt{m_x^2 + \vec{p}_3^2} \\ &= \frac{M^2 + m_1^2 - m_{23}^2}{2M} + \frac{M^2 + m_2^2 - m_{13}^2}{2M} + \sqrt{\left(\frac{M^2 + m_3^2 - m_{12}^2}{2M}\right)^2 - m_3^2 + m_x^2}\end{aligned}$$

and thus

$$\tilde{M}_{\text{cms}}(m_{12}^2|m_x) = \frac{M^2 - m_3^2 + m_{12}^2}{2M} + \sqrt{\left(\frac{M^2 + m_3^2 - m_{12}^2}{2M}\right)^2 - m_3^2 + m_x^2}. \quad (\text{N.2})$$

We note two things: First, only if $m_x = m_3$ then $\tilde{M}_{\text{cms}} = M$ for any frame of reference. Otherwise, Eq. (N.2) only holds in the CMS of M . Secondly, if the invariant mass is determined in the laboratory system whilst already assuming the spurious mass hypothesis, then the energy and thus the boost of the respective particle is also wrong. In this case, this will result in a smearing due to the inaccurate boosting of \tilde{M}_{cms} , degrading Eq. (N.2) to an approximation rather than an exact solution.

In Fig. N.1 we show the Dalitz plot [98, 99] of $m_{23}^2 = m_{KK}^2$ and $m_{12}^2 = m_{\Lambda K^-}^2$, as well as the correlation of m_{12}^2 with the reflected mass $m(\Lambda K^- \pi^+)$ together with the approximate solution Eq. (N.2). The data points are calculated from unsmeared values taken from simulations with

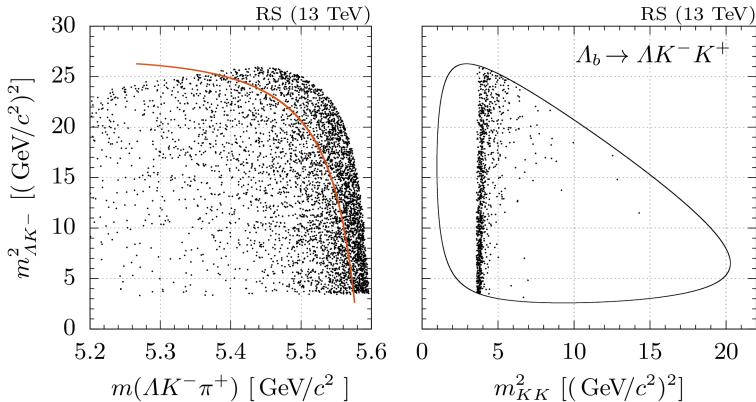


Fig. N.1. Dalitz plot (right) of $m_{23}^2 = m_{KK}^2$ and $m_{12}^2 = m_{\Lambda K^-}^2$, as well as the correlation of m_{12}^2 with the reflected mass $m(\Lambda K^- \pi^+)$ using simulated data (left). The solid, orange line on the left indicates the reflected invariant \tilde{M}_{cms} in the CMS of M , according to Eq. (N.2). The data are filtered w.r.t. the (smeared) vicinity of $m_{2x}^2 = m_{K-\pi^+}^2$ to the D^0 mass.

the framework **RapidSim** [104] and filtered w.r.t. the (smeared) vicinity of $m_{2x}^2 = m_{K-\pi^+}^2$ to the D^0 mass. (See Ref. [39] for Dalitz plots from recorded data. Given the limited data sample, resonance structures are not obvious to identify, however, known Λ_b decays, such as $\Lambda_b \rightarrow \Lambda \phi$ [40], are well separated in m_{KK}^2 .) The distribution of the invariant mass $m(\Lambda K^- \pi^+)$ in the CMS is (approximately) proportional to the product of the inverted derivative of Eq. (N.2) (solid line in the left part of Fig. N.1) and the density of $m_{12}^2 \equiv m_{\Lambda K^-}^2$, where the latter is flat for the central part of $m_{\Lambda K^-}^2$, due to the narrow mass range centered around the $m(D^0)$ mass.

At the sharp edges of the Dalitz distribution (*i.e.*, the kinematic boundaries) $m_{\Lambda K^-}^2$ gets sliced, resulting in a smearing of $m(\Lambda K^- \pi^+)$. The peak in the $m(\Lambda K^- \pi^+)$ distribution (*cf.* Fig. 7.4), caused by the steepness of the inverted derivative of Eq. (N.2), is thus naturally smeared. We note that this smearing is often exclusively and imprecisely attributed to *resolution effects* of the apparatus but in fact is an admixture of the outlined natural smearing in the CMS, a broadening due to boosting and eventually due to the quoted resolution effects.

N.2 Kinematics of the Λ Baryon in Two and Three-Body Decays

Naïvely, one might not expect major deviations between the kinematics of the Λ baryon in the $\Lambda_b \rightarrow D^0 \Lambda$ decay and its charmless counterpart. However, we saw deviations in the respective p_T distributions in Sec. 7.3.1 and saw the Λ classifier picking up this deviation to effectively separate between both decays.

In Fig. N.2 we show the three-momentum magnitudes of Λ baryons in the Λ_b rest frame. For the two-body decays $\Lambda_b \rightarrow D^0 \Lambda$, this value is a δ -spike (dashed line in Fig. N.2), sitting at

$$p^* = \sqrt{\frac{[m_{\Lambda_b}^2 - (m_{D^0} + m_\Lambda)^2][m_{\Lambda_b}^2 - (m_{D^0} - m_\Lambda)^2]}{4m_{\Lambda_b}^2}},$$

whereas in three-body decays the distribution is broad. The maximal value in the latter case corresponds to the configuration where the Λ goes back-to-back with the hh' pair of the respective $\Lambda_b \rightarrow \Lambda hh'$ pair, allowing slightly larger values if $m_h + m_{h'} < m_{D^0}$, but the majority of configurations results in values below p^* as shown in Fig. N.2. In Fig. N.3 we show the

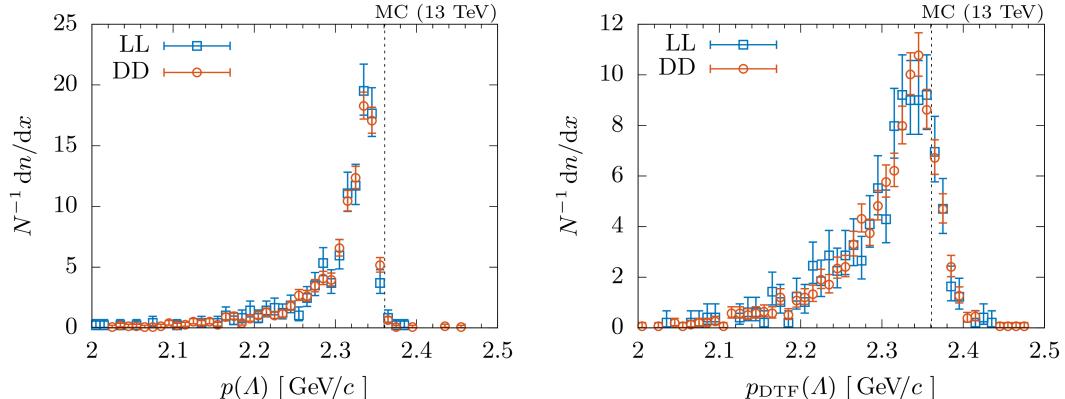


Fig. N.2. Distribution of three-momentum magnitudes of Λ baryons in the Λ_b rest frame for LL and DD tracks for MC simulated $\Lambda_b \rightarrow \Lambda K^- K^+$ decays (data points) and $\Lambda_b \rightarrow D^0 \Lambda$ decays (dashed line). On the left we show the unsmeared values and on the right the refined values after applying a DTF assuming $\Lambda_b \rightarrow D^0 \Lambda$ decays.

normalized p_T distributions of the two and three-body decays. The deviation between both is not as pronounced as the deviation between the flight distance significances of the D^0 meson (*cf.* Fig. N.4) that we utilize in the Λ_b - D^0 classifier but still holds a certain separation power that explains the suppression capability of the Λ classifier against charmless decays.

In the present analysis, $\Lambda_b \rightarrow \Lambda K^- K^+$ is considered a background, hence the implication of applying a DTF assuming a decay tree $\Lambda_b \rightarrow D^0 \Lambda$ is of interest. In the right part of Fig. N.2 we show the three-momentum magnitude of the Λ baryon in the Λ_b rest frame after applying

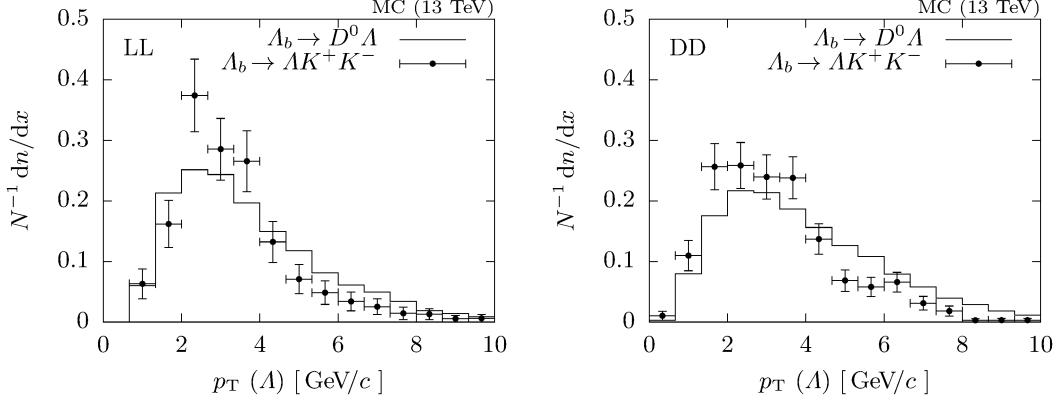


Fig. N.3. Transverse momentum distribution of Λ baryons from MC simulated $\Lambda_b \rightarrow D^0 \Lambda$ and $\Lambda_b \rightarrow \Lambda K^- K^+$ decays. The distributions are normalized in order to compensate their largely different yields for the sake of comparison. The transverse momentum of the charmless three-body decay seemingly prefer smaller values (both for LL tracks on the left, and DD tracks on the right) for reasons we elaborate in Sec. N.2.

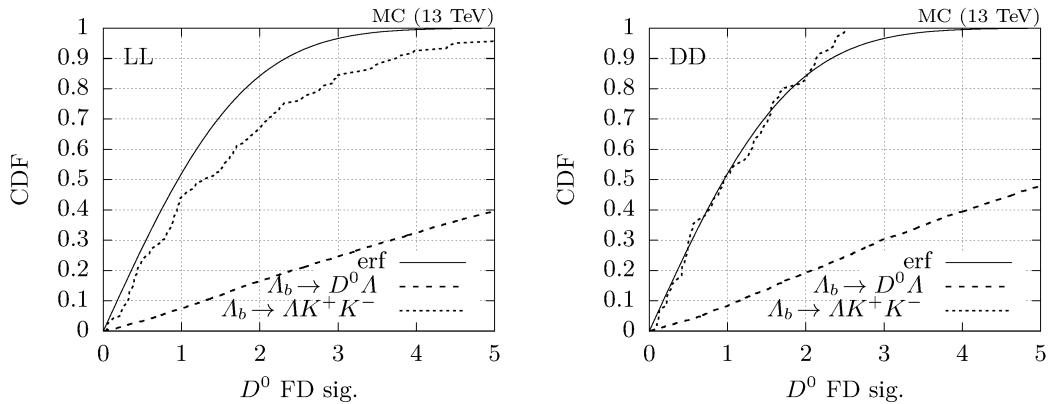


Fig. N.4. CDF of flight distance significance of D^0 mesons from MC simulated $\Lambda_b \rightarrow D^0 \Lambda$ and $\Lambda_b \rightarrow \Lambda K^- K^+$ decays where the latter is reflected as $\Lambda_b \rightarrow \Lambda K^- \pi^+$ and spuriously reconstructed as $\Lambda_b \rightarrow D^0 \Lambda$. For comparison reason we also show the CDF of a Gaussian function (Error function).

such a DTF. Unsurprisingly, the values are smeared and shifted towards larger values. As a consequence, the very same behavior is observed when comparing the invariant masses $m(D^0\Lambda)$ before and after applying the DTF, shown in Fig. 7.4 and Fig. N.5, respectively.

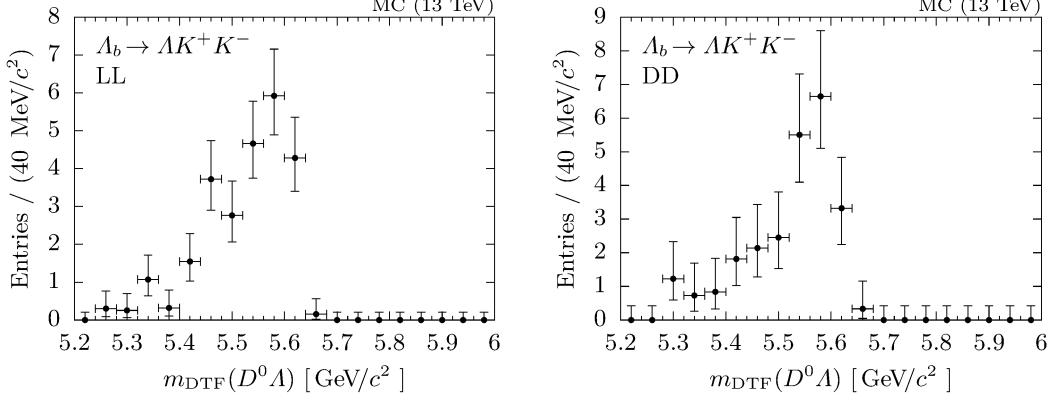


Fig. N.5. Combined invariant mass of $D^0 \rightarrow K^-\pi^+$ and $\Lambda \rightarrow p\pi^-$ candidates from MC simulated $\Lambda_b \rightarrow \Lambda K^-K^+$ decays after applying a DTF assuming the decay tree $\Lambda_b \rightarrow D^0\Lambda$.

Appendix O

Partially Reconstructed Backgrounds

Decays where at least one particle is lost during reconstruction are called *partially reconstructed* decays. In the following we discuss decays akin to $\Lambda_b \rightarrow D^{*0} \Lambda$ with $D^{*0} \rightarrow D^0 X$ where the X is lost during reconstruction. In particular, this description is applicable for the decays $\Lambda_b/\Xi_b^0 \rightarrow D^{*0} \Lambda$ with $D^{*0} \rightarrow D^0 \pi^0$ and $D^{*0} \rightarrow D^0 \gamma$, as well as $\Lambda_b/\Xi_b^0 \rightarrow D^0 \Sigma^0$ with $\Sigma^0 \rightarrow \Lambda \gamma$, and $\Xi_b^0 \rightarrow D^0 \Xi^0$ with $\Xi^0 \rightarrow \Lambda \pi^0$ which are considered as critical (partially reconstructed) background events in the invariant mass distribution of $\Lambda_b/\Xi_b^0 \rightarrow D^0 \Lambda$ events. In the following

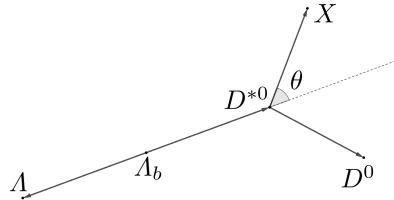


Fig. O.1. Decay topology of $\Lambda_b \rightarrow D^{*0} \Lambda$ with $D^{*0} \rightarrow D^0 X$ in the Λ_b rest frame.

we briefly introduce the notation and then find an explicit solution for the contribution of the various partially reconstructed backgrounds to the invariant mass $m(D^0 \Lambda)$ as a function of the opening angle θ (*cf.* Fig. O.1). We then discuss the implications of polarization effects and establish a universal fit model.

O.1 Kinematics of Partially Reconstructed Decays

The invariant mass is a Lorentz scalar and can thus be calculated in any frame of reference. For the sake of simplicity, we choose the CMS of the Λ_b particle and determine the spurious, reconstructed invariant mass of the Λ_b , denoted as M , here, without contributions from the left out particle X as

$$M = \sqrt{(E_A + E_D)^2 - (p - p_{\parallel,D})^2 - (p_{T,D})^2}, \quad (\text{O.1})$$

where E_A and E_D are the energies of the Λ and D^0 , p is the momentum of the Λ_b and D^{*0} , and $p_{\parallel,D}$ and $p_{T,D}$ are the parallel and transversal components of the D^0 momentum vector w.r.t. the flight direction of the D^{*0} particle, respectively.

The energy of the Λ particle is given by its invariant mass m_Λ and its momentum p

$$E_\Lambda = \sqrt{m_\Lambda^2 + p^2},$$

where p is given by the invariant masses of the Λ_b , D^{*0} and Λ particle

$$p = \frac{\sqrt{\left(m_{\Lambda_b}^2 - (m_{D^*} + m_\Lambda)^2\right)\left(m_{\Lambda_b}^2 - (m_{D^*} - m_{\Lambda_b})^2\right)}}{2m_{\Lambda_b}^2}.$$

The energy and momentum of the D^0 particle is found by boosting the corresponding energy E'_D from the CMS of its mother D^{*0} into the CMS of the Λ_b :

$$\begin{aligned} E_D &= \gamma \sqrt{m_D^2 + p'^2} + \sqrt{\gamma^2 - 1} p' \cos \theta, \\ p_{\parallel,D} &= \sqrt{(\gamma^2 - 1)(m_D^2 + p'^2)} + \gamma p' \cos \theta, \\ p_{T,D} &= p'_{T,D} = p' \sin \theta, \end{aligned}$$

where we defined the angle θ as shown in Fig. O.1. The variable p' is the momentum of the D^0 and X particle in the CMS of the D^{*0} particle and is given by the invariant masses of D^{*0} , D^0 and x particle

$$p' = \frac{\sqrt{(m_{D^*}^2 - (m_D + m_X)^2)(m_{D^*}^2 - (m_D - m_X)^2)}}{2m_{D^*}^2},$$

whereas the Lorentz factor γ is given by the kinematics of the D^{*0} in the Λ_b CMS

$$\gamma = \frac{E_{D^*}}{m_{D^*}} = \frac{\sqrt{m_{D^*}^2 + p'^2}}{m_{D^*}} = \sqrt{1 + \left(\frac{p}{m_{D^*}}\right)^2}.$$

Insertion into Eq. (O.1) yields

$$\begin{aligned} M^2 &= \left(\sqrt{m_A^2 + p^2} + \gamma \sqrt{m_D^2 + (p')^2} + \sqrt{\gamma^2 - 1} p' \cos \theta \right)^2 \\ &\quad - \left(p - \sqrt{(\gamma^2 - 1)(m_D^2 + (p')^2)} - \gamma p' \cos \theta \right)^2 - (p' \sin \theta)^2 \\ &= 2 \left[\left(\sqrt{m_A^2 + p^2} + \gamma \sqrt{m_D^2 + (p')^2} \right) \sqrt{\gamma^2 - 1} \right. \\ &\quad \left. + \left(p - \sqrt{(\gamma^2 - 1)(m_D^2 + (p')^2)} \right) \gamma \right] p' \cos \theta \\ &\quad + \left(\sqrt{m_A^2 + p^2} + \gamma \sqrt{m_D^2 + (p')^2} \right)^2 - \left(p - \sqrt{(\gamma^2 - 1)(m_D^2 + (p')^2)} \right)^2 - (p')^2 \\ &= 2 \left[\sqrt{(\gamma^2 - 1)(m_A^2 + p^2)} + p\gamma \right] p' \cos \theta \\ &\quad + m_A^2 + m_D^2 + 2\sqrt{m_D^2 + (p')^2} \left(\gamma \sqrt{m_A^2 + p^2} + p\sqrt{\gamma^2 - 1} \right) \\ &= 2 \frac{m_{\Lambda_b}}{m_{D^*}} pp' \cos \theta + m_A^2 + m_D^2 + (m_{\Lambda_b}^2 - m_A^2 - m_{D^*}^2) \sqrt{\frac{m_D^2 + (p')^2}{m_{D^*}^2}}. \end{aligned}$$

In the case of $D^{*0} \rightarrow D^0\gamma$ the expression can be simplified further by using $m_X = 0$:

$$\begin{aligned} (M(\theta))^2 &= \sqrt{(m_{\Lambda_b}^2 - (m_A + m_{D^*})^2)(m_{\Lambda_b}^2 - (m_A - m_{D^*})^2)} \times \frac{m_{D^*}^2 - m_D^2}{2m_{\Lambda_b}m_{D^*}^3} \cos \theta \\ &\quad \times \frac{m_{\Lambda_b}^2(m_{D^*}^2 + m_D^2) + (m_A^2 - m_{D^*}^2)(m_{D^*}^2 - m_D^2)}{2m_{D^*}^2}. \end{aligned}$$

The spurious, reconstructed invariant mass $m(\Lambda_b) = M(\theta)$ is thus distributed as

$$M(\theta) = \sqrt{a^2 + b^2 \cos(\theta)},$$

with positive constants a and b that are given by the invariant masses of the decaying daughter and grand-daughter particles. The kinematic boundaries are given by $M(0)$ and $M(\pi)$. In Tab. O.1 we list the kinematic boundaries for $\Lambda_b/\Xi_b^0 \rightarrow D^{*0}\Lambda$ with $D^{*0} \rightarrow D^0\pi^0$ and $D^{*0} \rightarrow D^0\gamma$, as well as $\Lambda_b/\Xi_b^0 \rightarrow D^0\Sigma^0$ with $\Sigma^0 \rightarrow \Lambda\gamma$, and $\Xi_b^0 \rightarrow D^0\Xi^0$ with $\Xi^0 \rightarrow \Lambda\pi^0$.

Tab. O.1. Kinematic boundaries of the invariant mass $m(D^0\Lambda)$, as well as parameters a and b (*cf.* Eq. (O.1)) for various Λ_b and Ξ_b^0 decays.

Decay channel	min [MeV/ c^2]	max [MeV/ c^2]	a [MeV/ c^2]	b MeV/ c^2
$\Lambda_b \rightarrow D^*\Lambda, D^* \rightarrow D\pi$	5350.2(4)	5452.06(23)	5401.36(19)	741.8(19)
$\Lambda_b \rightarrow D^*\Lambda, D^* \rightarrow D\gamma$	5240.66(24)	5568.11(17)	5406.87(19)	1330.28(31)
$\Lambda_b \rightarrow D\Sigma, \Sigma \rightarrow \Lambda\gamma$	5300.33(18)	5601.69(17)	5453.09(17)	1281.69(19)
$\Xi_b \rightarrow D^*\Lambda, D^* \rightarrow D\pi$	5514.5(6)	5620.9(5)	5568.0(5)	769.8(20)
$\Xi_b \rightarrow D^*\Lambda, D^* \rightarrow D\gamma$	5400.0(5)	5742.1(5)	5573.7(5)	1380.48(35)
$\Xi_b \rightarrow D\Sigma, \Sigma \rightarrow \Lambda\gamma$	5460.1(5)	5774.7(5)	5619.6(5)	1329.34(24)
$\Xi_b \rightarrow D\Xi, \Xi \rightarrow \Lambda\pi$	5106.4(9)	5640.4(5)	5380.0(6)	1693.9(14)

O.2 Implication of Polarization

In general, $\cos\theta$ in Eq. (O.1) will not be distributed uniformly for Λ_b/Ξ_b^0 decays, even for unpolarized Λ_b/Ξ_b^0 particles, since all involved particles in the respective decay chain carry non-zero spin themselves and therefore will be polarized. These polarizations are distributed by the weak interaction in a non-trivial manner. For a given distribution of $f(\cos\theta)$ in the Λ_b rest frame, the function $M(\cos\theta) = \sqrt{a^2 + b^2 \cos\theta}$ is distributed according to

$$g(m) = \frac{f(M^{-1}(m))}{M'(M^{-1}(m))} = \frac{2m}{b^2} f\left(\frac{m^2 - a^2}{b^2}\right).$$

For unpolarized decays, $f(x) = \text{const.}$, this is a linear function $g(m) \propto mb^{-2}$. Below, we list $g(m)$ for a few simple polarization assumptions:

$$\begin{aligned} f_1(x) &= \frac{1}{2} \quad \rightarrow \quad g(m) = \frac{m}{b^2}, \\ f_{2,3}(x) &= \frac{1 \mp x}{2} \quad \rightarrow \quad g(m) = \frac{m}{b^2} \times \left[1 \mp \left(\frac{m^2 - a^2}{b^2} \right) \right], \\ f_4(x) &= \frac{3(1 - x^2)}{4} \quad \rightarrow \quad g(m) = \frac{m}{b^2} \times \frac{3}{2} \left[1 - \left(\frac{m^2 - a^2}{b^2} \right)^2 \right], \\ f_5(x) &= \frac{3}{2}x^2 \quad \rightarrow \quad g(m) = \frac{m}{b^2} \times 3 \left(\frac{m^2 - a^2}{b^2} \right)^2. \end{aligned}$$

Due to the lack of experimental measurements of the polarization¹ of the decays under consideration none of these models can be favored a priori. Instead, the polarization could be measured

¹More precisely, we use the term *polarization* to refer to the transverse polarization of particles since longitudinal contributions are expected to vanish in pp collisions due to parity conservation in strong interactions [142].

by fitting the distribution of recorded data with a generic, Taylor expansion based approach

$$g^{\text{fit}}(m|\alpha, \beta) = \frac{m}{b^2} \left[1 + \alpha \left(\frac{m^2 - a^2}{b^2} \right) + \beta \left(\frac{m^2 - a^2}{b^2} \right)^2 \right].$$

In Fig. O.2 we show the distributions of the polarization assumptions f_i , as well as an example for the corresponding distributions of $g(m)$ for the decay $D^{*0} \rightarrow D^0\pi^0$ (Λ_b decay). We note

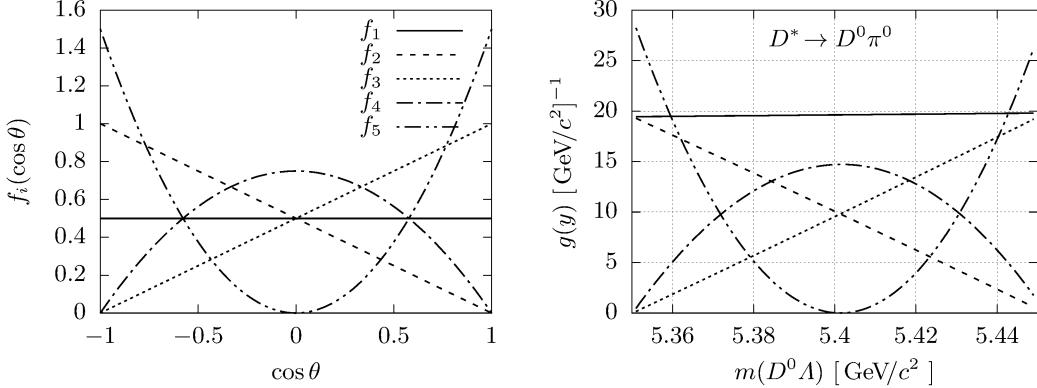


Fig. O.2. Distribution $g(m)$ for the decay $\Lambda_b \rightarrow D^{*0}\Lambda$ with $D^{*0} \rightarrow D^0\pi^0$ (right) assuming different polarizations f_i (left).

that the presence of polarization can change the shape significantly, *e.g.*, f_4 suppresses $g(m)$ up to zero, whereas the enhancement of f_5 is maximal at the same point. Furthermore, we note that the shape of $g(m)$ is also imposed by the values of a and b which makes the unpolarized shape of $g(m)$ very flat, whereas in three-body decays, such as $\Lambda_b \rightarrow \Sigma^0 hh'$, $g(m)$ by itself, is already steep [39]. Another example for the implication of polarization is the decay of the pseudo-scalar particle B^0 into two vector particles, $B^0 \rightarrow DK^{*0}$, where the very same partially reconstructed background $D^{*0} \rightarrow D^0\pi^0$ and $D^{*0} \rightarrow D^0\gamma$ had to be described and found to differ significantly from being unpolarized in both cases [54].

In recorded data this shape will typically be smeared out due to various factors, such as the limited resolution of the experimental setup. A first order approximation is motivated by the central limit theorem and state that the smearing can be parametrized by a convolution with a centered Gaussian function $\mathcal{G}_c \equiv \mathcal{G}(x|\mu = 0, \sigma)$. For simple functions such as the partial quadratic polynomial \mathcal{K} ,

$$\mathcal{K}(x|a, b) = \begin{cases} 1 + ax + bx^2 & \text{for } x_1 < x < x_2, \\ 0 & \text{else,} \end{cases}$$

the convolution can be evaluated analytically:

$$(\mathcal{K} * \mathcal{G}_c)(x) = -\frac{1 + ax + b(\sigma^2 + x^2)}{2} \operatorname{erf} \left(\frac{x-y}{\sqrt{2}\sigma} \right) - \sigma^2(a + b(x+y)) \mathcal{G}(x|y, \sigma) \Big|_{y=x_1}^{y=x_2}, \quad (\text{O.2})$$

with

$$\mathcal{G}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right).$$

The normalization of $(\mathcal{K} * \mathcal{G}_c)(x)$ can also be found analytically by integration:

$$\begin{aligned} \int dx (\mathcal{K} * \mathcal{G})(x) &= - \left[\frac{x-y}{2} + a \frac{x^2 - y^2 - 3\sigma^2}{4} + b \frac{x^3 + 3\sigma^2 x - y^3}{6} \right] \operatorname{erf} \left(\frac{x-y}{\sqrt{2}\sigma} \right) \\ &\quad - \left[1 + a \frac{x+y}{2} + b \frac{2\sigma^2 + x^2 + xy + y^2}{3} \right] \sigma^2 \mathcal{G}(x|y, \sigma) \Big|_{y=x_1}^{y=x_2}. \end{aligned}$$

If instead of using one Gaussian, a weighted sum centered Gaussian shapes $\mathcal{G}_{c,i}$ with zero mean but different widths σ_i is used for the smearing, Eq. (O.2) can be generalized due to distributivity and associativity with scalar multiplication of convolutions:

$$\begin{aligned} (\mathcal{K} * \sum_i w_i \mathcal{G}_{c,i})(x) &= \sum_i w_i (\mathcal{K} * \mathcal{G}_{c,i})(x) \\ &= - \sum_i \frac{1 + ax + b(\sigma_i^2 + x^2)}{2} w_i \operatorname{erf} \left(\frac{x-y}{\sqrt{2}\sigma_i} \right) \Big|_{y=x_1}^{y=x_2} \\ &\quad - \sum_i \sigma_i^2 (a + b(x+y)) w_i \mathcal{G}(x|y, \sigma_i) \Big|_{y=x_1}^{y=x_2}. \end{aligned}$$

Appendix P

Fit Supplementaries

Tab. P.1. Fit results of configuration 1, using the narrow mass range $5.5 \leq m(D^0\Lambda) \leq 6 \text{ GeV}/c^2$. The value of f_{D^*} (marked with \dagger) is fixed to zero. All other values are floating parameters during the likelihood maximization.

Parameter	Value		
	LL	DD	Unit
$\mu(\Lambda_b)$	5.621 14(21)		GeV/c^2
$\mu(\Xi_b)$	5.794 95(21)		GeV/c^2
$\Delta\mu(\Lambda_b)$	0.001 62(32)		GeV/c^2
$\Delta\mu(\Xi_b)$	0.001 31(31)		GeV/c^2
$\sigma_1(\Lambda_b)$	0.014 35(28)	0.0151(7)	GeV/c^2
$\sigma_1(\Xi_b)$	0.014 42(35)	0.0158(5)	GeV/c^2
$\sigma_2(\Lambda_b)$	0.0346(27)	0.0248(16)	GeV/c^2
$\sigma_2(\Xi_b)$	0.0283(17)	0.0280(18)	GeV/c^2
$f_{\mathcal{G}}(\Lambda_b)$	0.903(21)	0.72(9)	
$f_{\mathcal{G}}(\Xi_b)$	0.838(35)	0.81(5)	
f_s		0.83(10)	
k	5.2(16)	2.1(10)	$1/(\text{GeV}/c^2)$
f_{D^*}		0^\dagger	
f_1	0.62(10)	0.77(8)	
f_2		1.0(7)	

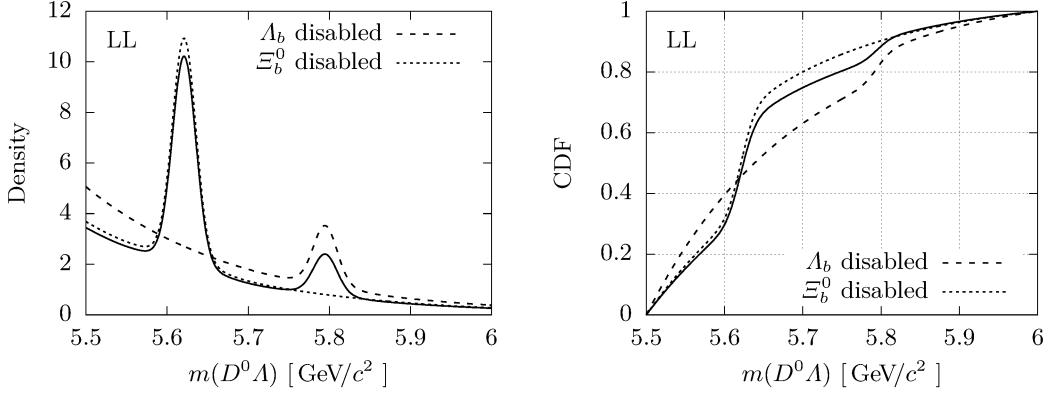


Fig. P.1. PDFs (left) and their cumulative distributions (right) as fitted in the projection of LL tracks in configuration 1 (solid line), as well as the shapes with disabled $\Lambda_b \rightarrow D^0\Lambda$ or $\Xi_b^0 \rightarrow D^0\Lambda$ component (dashed lines). The cumulative distributions (CDF) are used for generating the pseudo-experiments.

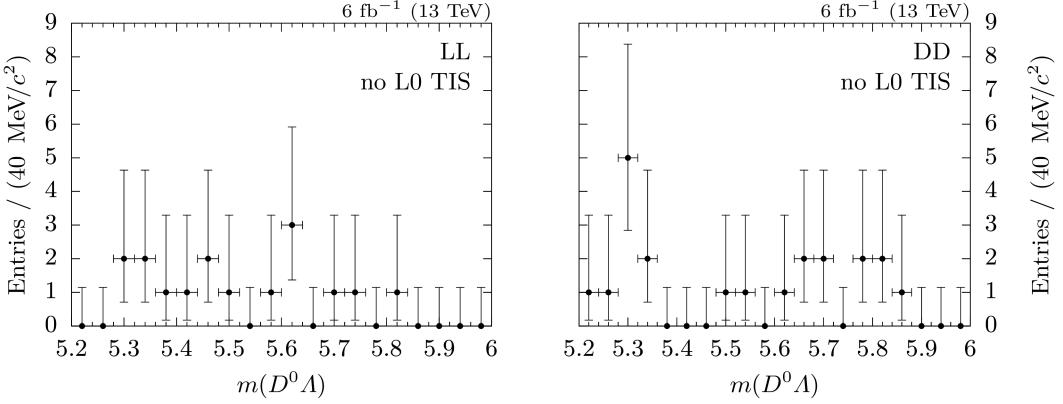


Fig. P.2. Combined invariant mass of D^0 and A candidates of track type LL (left) and DD (right) from recorded data with a negative L0 TIS trigger decision.

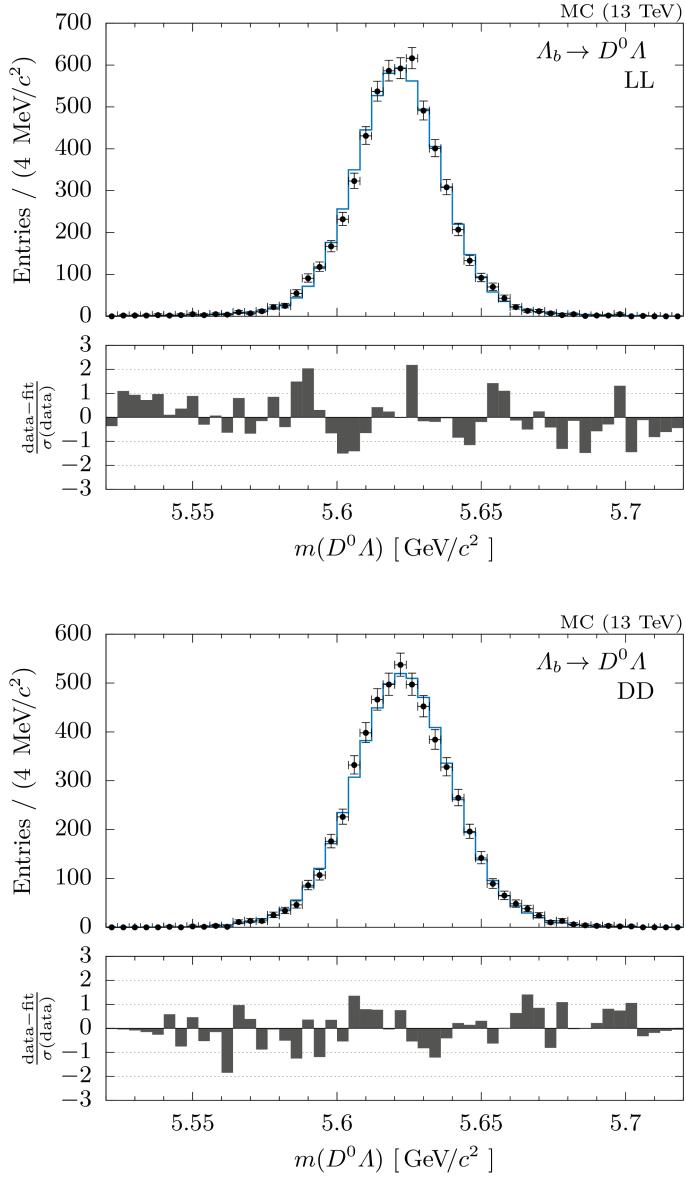


Fig. P.3. Combined invariant mass of D^0 and Λ candidates of track type LL (top) and DD (bottom) from MC simulated $\Lambda_b \rightarrow D^0\Lambda$ decays, as well as the corresponding projections of the simultaneous fit in configuration 1.

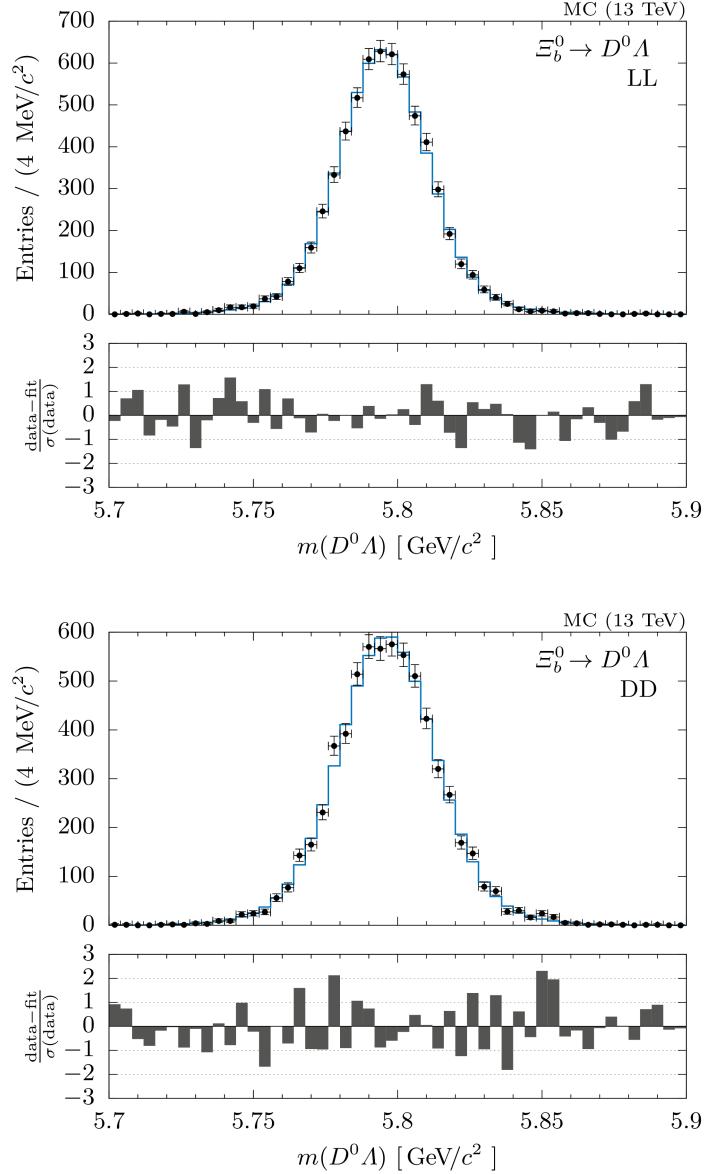


Fig. P.4. Combined invariant mass of D^0 and Λ candidates of track type LL (top) and DD (bottom) from MC simulated $\Xi_b^0 \rightarrow D^0\Lambda$ decays, as well as the corresponding projections of the simultaneous fit in configuration 1.

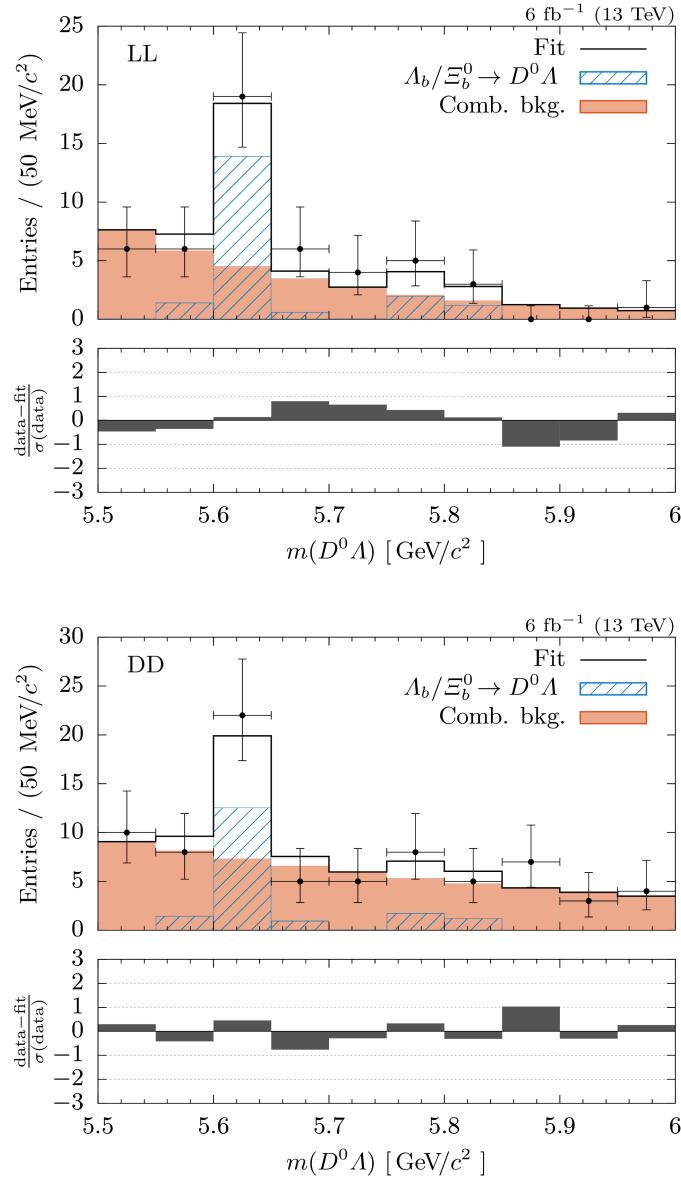


Fig. P.5. Combined invariant mass of D^0 and Λ candidates of track type LL (top) and DD (bottom) from recorded data, as well as the corresponding projections of the simultaneous fit in configuration 1.

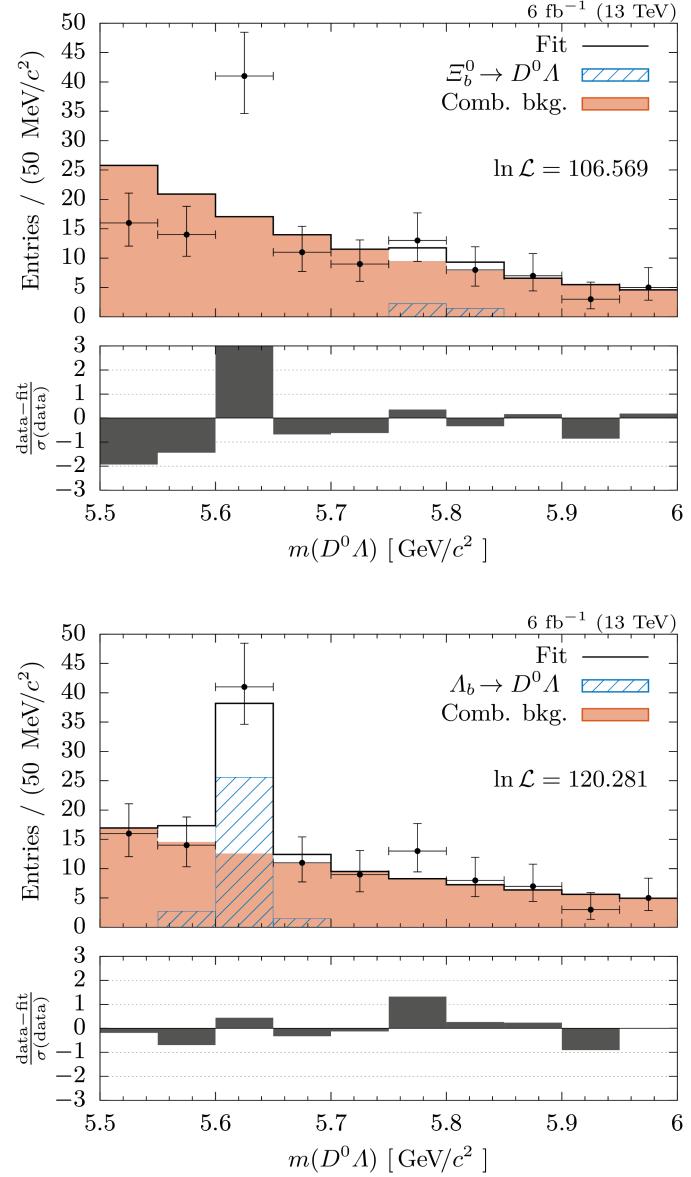


Fig. P.6. Combined invariant mass of D^0 and Λ candidates from recorded data, as well as the corresponding accumulated projections of the simultaneous fit for both track types. These fits are used to extract the yield significance of $\Lambda_b \rightarrow D^0\Lambda$ (top) and $\Xi_b^0 \rightarrow D^0\Lambda$ (bottom). The log-likelihood when both modes are enabled (one DoF more) is $\ln \mathcal{L} = 121.944$.

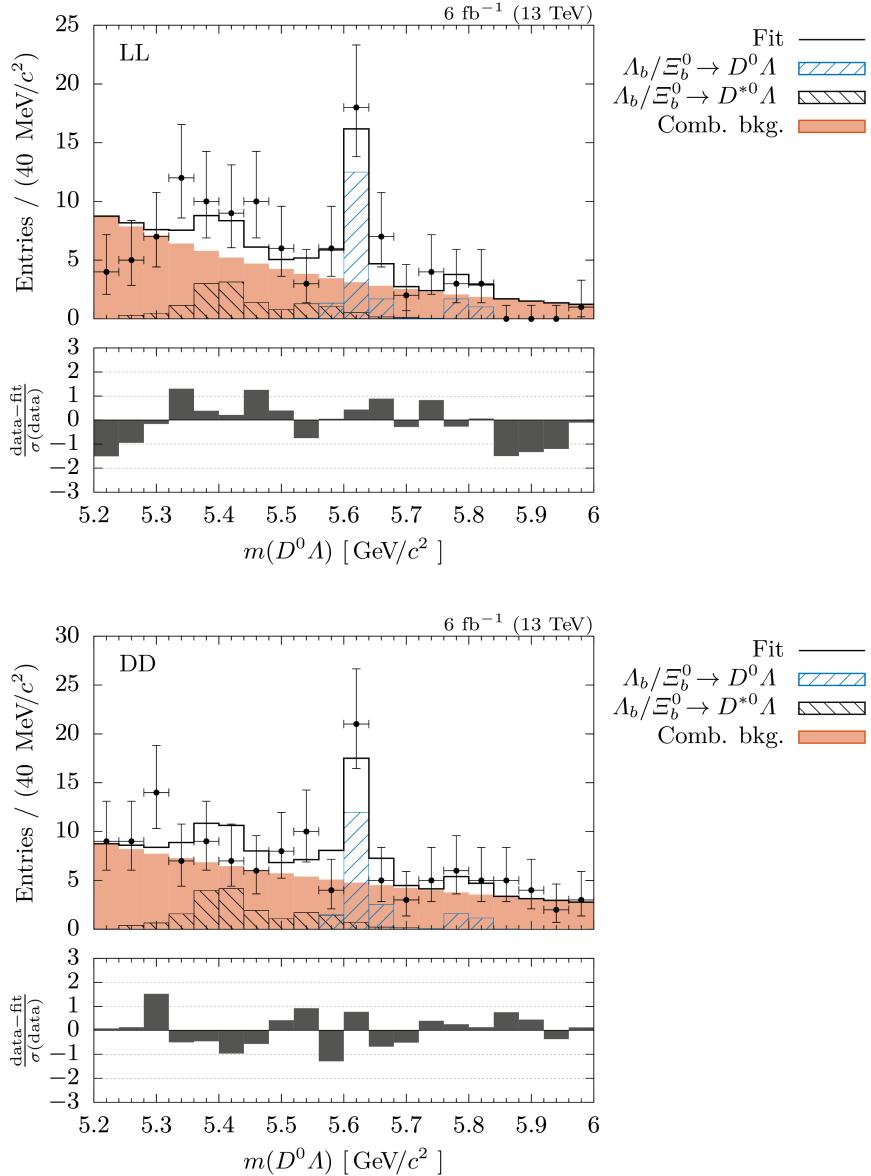


Fig. P.7. Combined invariant mass of D^0 and Λ candidates of track type LL (top) and DD (bottom), as well as the corresponding projections of the fit in configuration 3. The yield of the D^{*0} background is constrained among both track types, presumably causing the overshooting of the combinatorial background in the former mode.

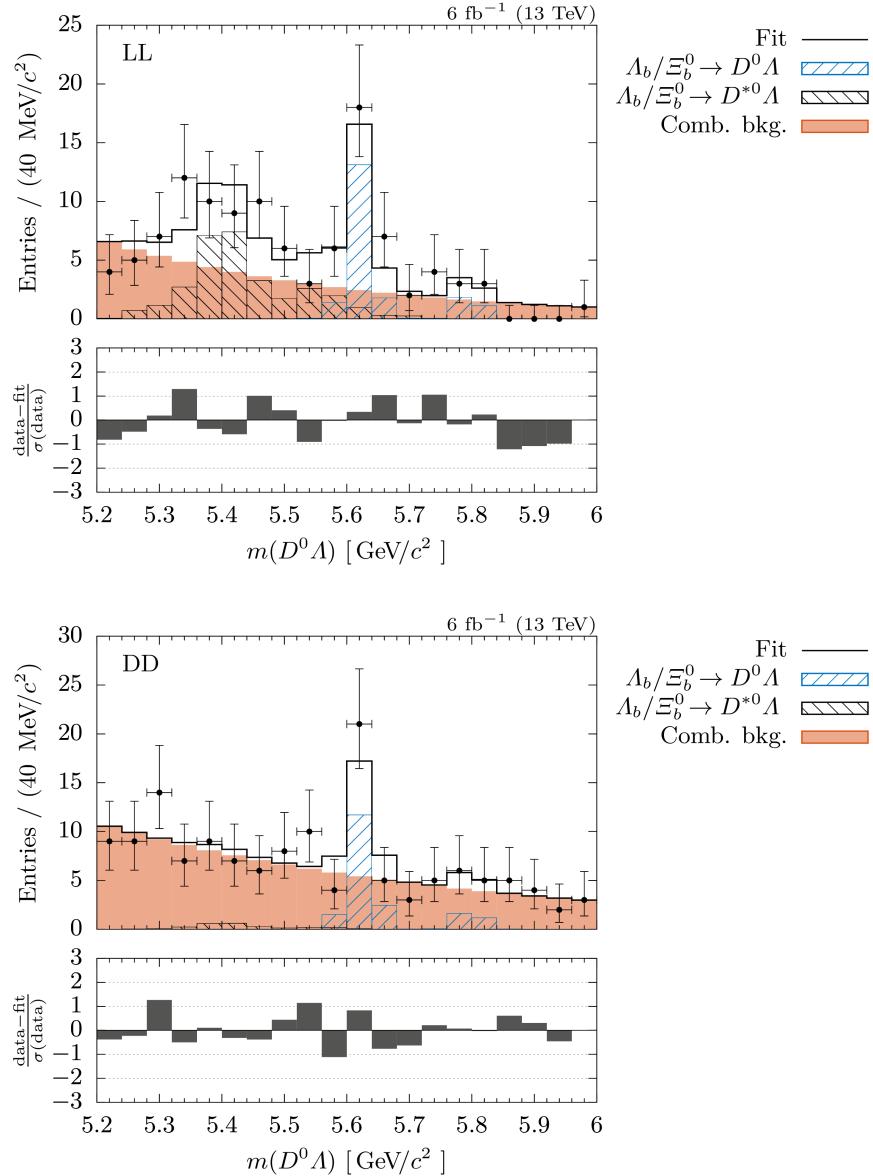


Fig. P.8. Combined invariant mass of D^0 and Λ candidates of track type LL (top) and DD (bottom) from recorded data, as well as the corresponding projections of the simultaneous fit in configuration 4. The fraction f_2 is unconstrained and is allowed to vary among different track types.

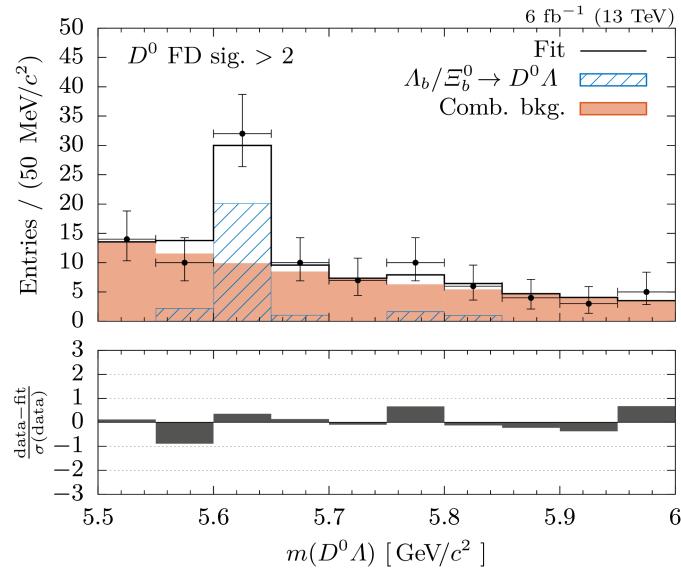


Fig. P.9. Fit to combined invariant mass of D^0 and Λ candidates of both track types from recorded data, when the flight distance significance of D^0 candidates is required to be larger than 2.

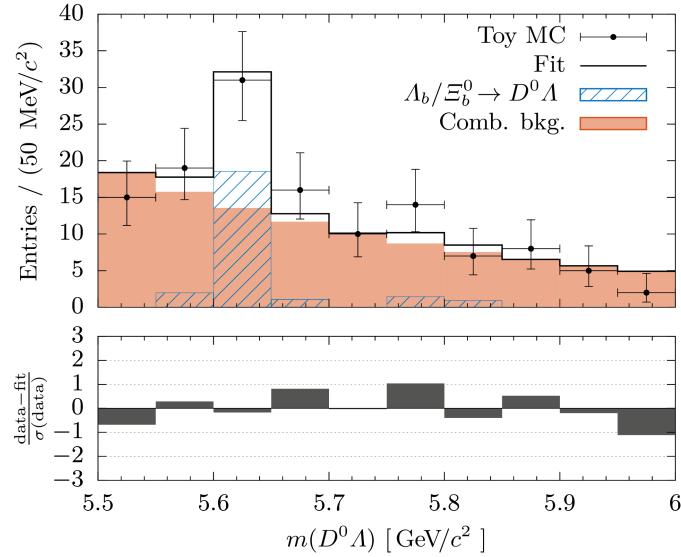


Fig. P.10. Example of a distribution that was generated as part of the pseudo-experiment to test for a bias and the validity of the error estimation of the fit. For the generation and fitting both signals modes were kept enabled.

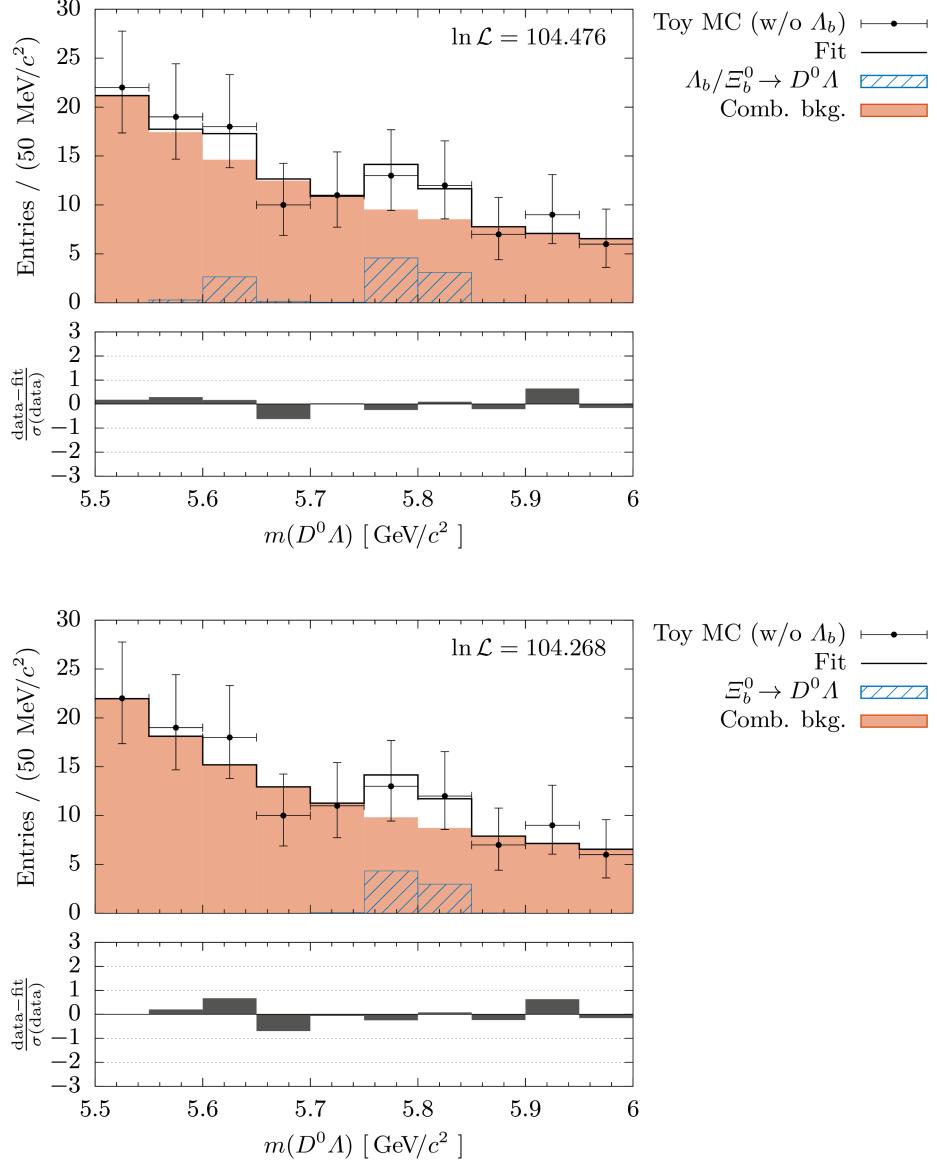


Fig. P.11. Example of a distribution (data points) that was generated as part of the pseudo-experiment to test the validity of the signal yield significance of the $A_b \rightarrow D^0\Lambda$ mode. During generation this mode was kept disabled (null hypothesis). The distribution is fitted twice, first with the A_b signal component enabled (top) and once with disabled A_b signal component (bottom). The difference in the log-likelihood values is then used to benchmark the null-hypothesis.

Appendix Q

The Exponential Function as a PDF

The normalized exponential function $f(x|k)$ has one free parameter k

$$f(x|k) = \mathcal{N} e^{-kx},$$

given a sufficient normalization factor \mathcal{N} . For a given normalization interval $[a_1, a_2]$ parts of the normalization factor \mathcal{N} can be moved into the argument of the exponential function

$$f(x|k) = \frac{k e^{-k(x-a_1)}}{1 - e^{-k(a_2-a_1)}}. \quad (\text{Q.1})$$

In the limit $k \rightarrow 0$ Eq. (Q.1) is well defined

$$\lim_{k \rightarrow 0} f(x|k) = 1,$$

but numerically unstable. Math libraries typically implement the relative error exponential function

$$\text{exprel}(x) := \frac{e^x - 1}{x},$$

to avoid the loss of precision that occurs when x is near zero. Using this definition of `exprel` Eq. (Q.1) can be rewritten

$$f(x|k) = \frac{e^{-k(x-a_1)}}{(a_2 - a_1) \text{exprel}(-k(a_2 - a_1))}, \quad (\text{Q.2})$$

and now allows sign flips of k during the fitting process.

Fitting an exponential function on a single interval can easily be generalized for piecewise fits on two disjoint normalization intervals $[a_1, a_2]$ and $[b_1, b_2]$. Doing so changes the normalization factor and Eq. (Q.2) becomes

$$f(x|k) = \frac{e^{-k(x-a_1)}}{(a_2 - a_1) \text{exprel}(-k(a_2 - a_1)) + (b_2 - b_1) e^{-k(b_1 - a_1)} \text{exprel}(-k(b_2 - b_1))}. \quad (\text{Q.3})$$

In both cases the integral in x can be written as

$$\int_{x_1}^{x_2} dx f(x|k) = (x_2 - x_1) \text{exprel}(-k(x_2 - x_1)) \times f(x_1|k). \quad (\text{Q.4})$$

This relation is useful when a fit for k is used to inter- or extrapolate yields in given regions. An uncertainty approximation by applying ordinary error propagation needs the derivative w.r.t. k of Eq. (Q.4) which is cluttered and again suffers from numerical instabilities. When error propagation is needed, we will therefore revert to a numerical approach. Eq. (Q.4) is then evaluated for Gaussian distributed values of $k \sim \mathcal{G}(\mu, \sigma)$, where μ and σ are the results of the preceding fit, and sorted. The interval that spans a fraction of $\text{erf}(1/\sqrt{2}) \approx 68\%$ of the sorted values, centered at the median is then taken as an approximation of the uncertainty interval.

Appendix R

Clipped Gaussian Distribution

A large part of the free parameters of the fit model that we establish in Chap. 8 are fractions, such as f_1 and f_s , which are physically constrained to the interval $[0, 1]$. These boundaries enhance the numerical stability of the fit and makes it easy to separate or disable components of the likelihood by fixing the fractions to zero or one. By design, $f \in [0, 1]$ guarantees $0 \leq \mathcal{L} \leq 1$ and forbids negative signal yields which are meaningless for the present analysis. As a consequence, the nominal values of fitted yields f which can be assumed to be Gaussian distributed, will accumulate excessively at $f = 0$, *i.e.*, the assumed Gaussian distribution

$$\mathcal{G}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

transforms, since x is not distributed uniformly $\phi(x) = x$, but rather clipped at negative values $\phi(x) = \Theta(x)x$,

$$\mathcal{G}(x|\mu, \sigma) \xrightarrow{\phi(x)=\Theta(x)x} \frac{\delta(x)}{2} \operatorname{erfc} \frac{\mu}{\sqrt{2}\sigma} + \mathcal{G}(x|\mu, \sigma) =: \tilde{\mathcal{G}}(x|\mu, \sigma).$$

This clipping changes the moments non-trivially, in particular,

$$\langle x \rangle_{\tilde{\mathcal{G}}} = \mu \left[1 - \frac{1}{2} \operatorname{erfc} \frac{\mu}{\sqrt{2}\sigma} \right] + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} \neq \mu = \langle x \rangle_{\mathcal{G}}, \quad (\text{R.1})$$

$$\langle x^2 \rangle_{\tilde{\mathcal{G}}} = (\mu^2 + \sigma^2) \left[1 - \frac{1}{2} \operatorname{erfc} \frac{\mu}{\sqrt{2}\sigma} \right] + \frac{\mu\sigma}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} \neq \mu^2 + \sigma^2 = \langle x^2 \rangle_{\mathcal{G}}. \quad (\text{R.2})$$

We note that, although the mean value is shifted, this does not contribute to a potential bias of the fit if the genuine yield (fraction) is non-zero, since it does not affect all values, but only unphysical $f < 0$ fluctuations.

If instead considering $\phi(x) = \Theta(x)x^2$ which would corresponds to a χ_1^2 -distribution with one DoF in the unclipped case with $\mu = 0$ and $\sigma = 1$, the distribution now becomes a clipped χ^2 -distribution:

$$\mathcal{G}(x|0, 1) \xrightarrow{\phi(x)=\Theta(x)x^2} \frac{1 + \delta(x)}{2} \frac{\exp(-x/2)}{\sqrt{2\pi x}} = \frac{1 + \delta(x)}{2} \chi_1^2(x) =: \tilde{\chi}_1^2(x).$$

The mean and variances of $\tilde{\chi}_1^2(x)$ are

$$\langle x \rangle_{\tilde{\chi}_1^2} = \frac{1}{2} \langle x \rangle_{\chi_1^2} = \frac{1}{2}, \quad (\text{R.3})$$

$$\langle x^2 \rangle_{\tilde{\chi}_1^2} - \langle x \rangle_{\tilde{\chi}_1^2}^2 = \frac{1}{2} \langle x^2 \rangle_{\chi_1^2} - \frac{1}{4} \langle x \rangle_{\chi_1^2}^2 = \frac{3}{2} - \frac{1}{4} = \frac{5}{4}. \quad (\text{R.4})$$

Appendix S

Estimations of a CI for the Ξ_b^0 - Λ_b Ratio

The fit parameter f_s of the model that we established in Sec. 8.1 can be used to estimate the branching fraction of the Ξ_b^0 and Λ_b baryon into $D^0\Lambda$ up to corrections due to the different b -fragmentations:

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \rightarrow D^0\Lambda)}{\mathcal{B}(\Lambda_b \rightarrow D^0\Lambda)} = \frac{1 - f_s}{f_s} =: f(\Xi_b^0/\Lambda_b), \quad (\text{S.1})$$

where $f_{\Lambda_b}/f_{\Xi_b^0}$ is the ratio of the fragmentation fractions of b -quarks into Λ_b and Ξ_b^0 baryons.

Two frequentist confidence intervals (CI) according to Ref. [143] are calculated by drawing random events from the fitted PDF where all parameters are fixed, except for f_1 and f_s . While $\hat{f}(\Xi_b^0/\Lambda_b)$ is varied on the interval $[0.0 \dots 0.6]$, the value of f_1 is corrected such that the ratio of the Λ_b signal and the combinatorial background $f_s(1 - f_1)/f_1$ stays constant. For each value of $\hat{f}(\Xi_b^0/\Lambda_b)$, 400 fits are performed. Each of these fits yields a value for f_s which is used to calculate $f_{\text{obs}}(\Xi_b^0/\Lambda_b)$, according to Eq. (S.1). CIs are estimated by finding intervals in $\hat{f}(\Xi_b^0/\Lambda_b)$ such that, according to the pseudo-experiment, for each value of $\hat{f}(\Xi_b^0/\Lambda_b)$ within this interval, intervals with a given coverage in $f_{\text{obs}}(\Xi_b^0/\Lambda_b)$ include the value that was found with the fit to recorded data. In total, we evaluate 25 different $\hat{f}(\Xi_b^0/\Lambda_b)$ values and smooth the estimated boundaries with linear functions.

Two different methods of finding the intervals in $f_{\text{obs}}(\Xi_b^0/\Lambda_b)$ are used: In Fig. S.1 we show the result of the *central* method where the 400 different outcomes are partitioned at the 16% (5%) and 84% (95%) percentiles, *i.e.*, the central interval corresponds to 68% (90%) CL. We note that this method of constructing two-sided intervals implicitly gives the one-sided 84% (95%) CL upper limit, too.

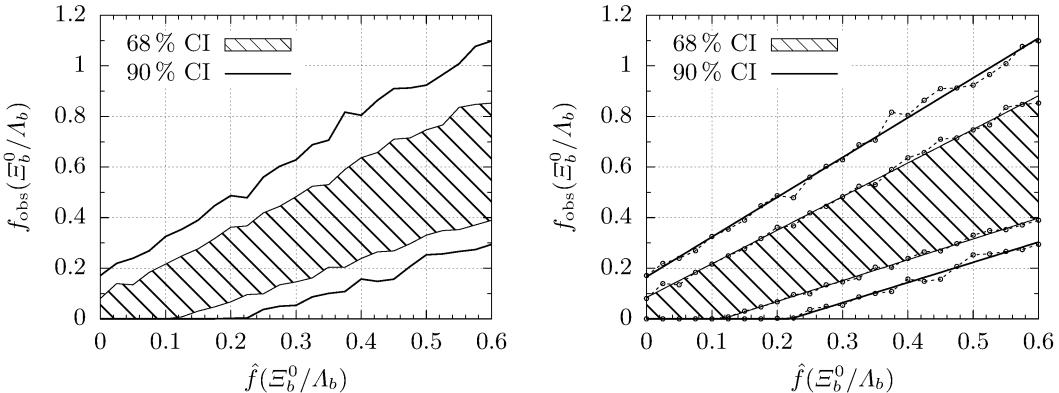


Fig. S.1. Frequentist CIs using the *central* method. Interval boundaries are estimated by percentiles in $f_{\text{obs}}(\Xi_b^0/\Lambda_b)$ at 25 different $\hat{f}(\Xi_b^0/\Lambda_b)$ positions. The boundaries (left) are smoothed with linear functions (right).

In Fig. S.2 we show the result of the *shortest* method where for each value of $\hat{f}(\Xi_b^0/\Lambda_b)$ the

shortest interval in $f_{\text{obs}}(\Xi_b^0/\Lambda_b)$ with a 68% and 90% coverage is estimated. The output of this method is noisier and does not allow an extraction of upper limits.

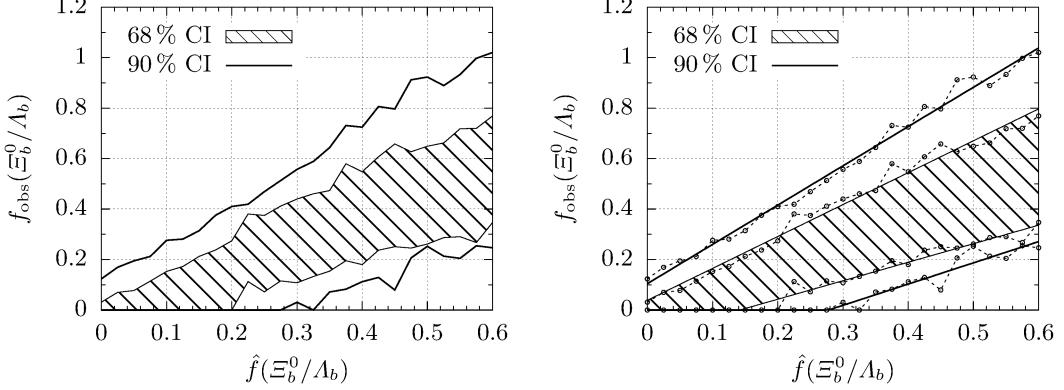


Fig. S.2. Frequentist CIs using the *shortest* method. Interval boundaries are estimated by finding the shortest intervals in $f_{\text{obs}}(\Xi_b^0/\Lambda_b)$ with a coverage of 68% and 90% at 25 different $\hat{f}(\Xi_b^0/\Lambda_b)$ positions. The boundaries (left) are smoothed with linear functions (right).

As an alternative, we also calculate CIs using Bayesian methods by converting the fitted likelihood, as shown in Fig. S.3, into an a posteriori probability density (using Bayes theorem). This transformation involves a normalization and an assumption about the prior probability. The latter is non-obvious and we decide to estimate intervals based on the assumption of a uniform distribution of $f(\Xi_b^0/\Lambda_b)$. In Fig. S.4 we show the results of integrating the PDF when using the *central* and *shortest* method. Again, 84% CL and 95% CL upper limits are implicitly given by the *central* 68% and 90% interval, respectively. A comparison of the presented methods is shown in Fig. 9.2 and Tab. 9.2.

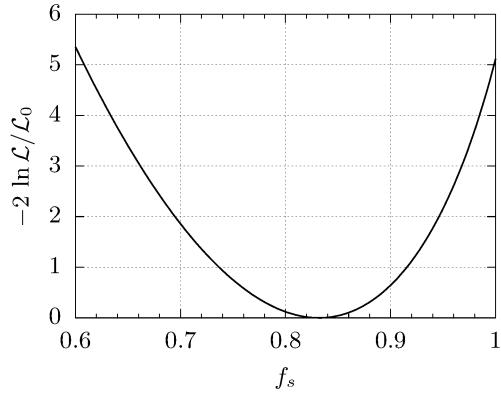


Fig. S.3. Normalized log-likelihood ratio of f_s as estimated by the fitting procedure. The normalization is chose such that intervals using the likelihood ratio method can directly be read off the y -axis.

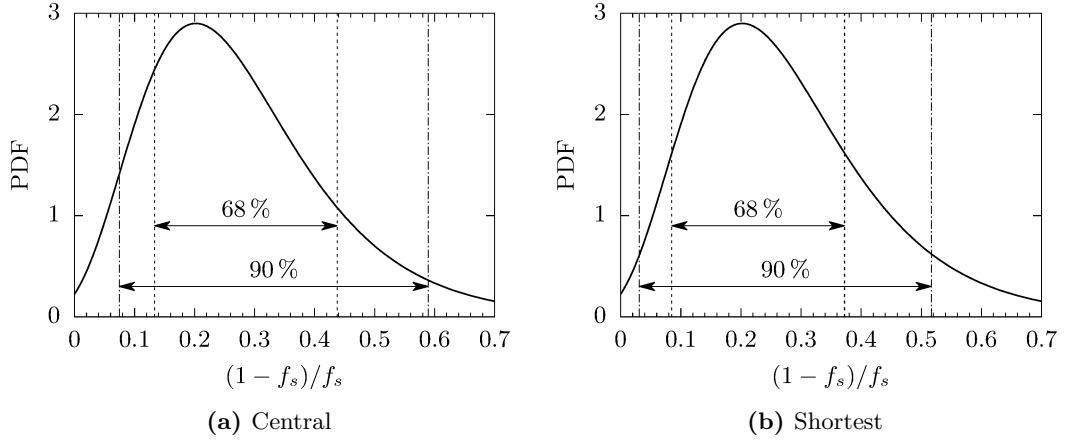


Fig. S.4. Bayesian CIs by integrating the normalized likelihood of $f(\Xi_b^0/\Lambda_b)$ as estimated by the fitting procedure, assuming a uniform distribution of the prior. On the left (right) intervals are chosen according to the *central* (*shortest*) method.

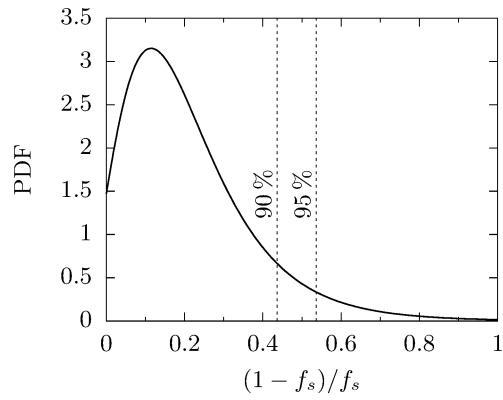


Fig. S.5. Bayesian upper limits by integrating the normalized likelihood of $f(\Xi_b^0/\Lambda_b)$ as estimated by the fitting procedure, assuming a uniform distribution of the prior, when an additional veto against charmless Ξ_b^0 backgrounds is required.

Danksagung

An dieser Stelle möchte ich allen beteiligten Personen danken, die mich während meiner Promotions- und Studienzeit unterstützt haben:

Mein besonderer Dank gilt Priv.-Doz. Dr. Roland Waldi für die ausgezeichnete Betreuung. Sein Wissen und seine Erfahrung sowie die Bereitschaft, sich unabhängig von Arbeitszeiten und Ruhestand tiefgreifend in meine Arbeit hineinzudenken, haben mich bei der Anfertigung dieser Arbeit enorm unterstützt. Seine große Flexibilität gegenüber familiären Situationen sowie das Einräumen von Möglichkeiten mich mit Themen, die über den Kernbereich meiner Arbeit hinausgehen, zu beschäftigen, waren nicht selbstverständlich. Vielen Dank, Roland!

Nicht weniger wichtig ist die großartige Unterstützung meiner Partnerin Sophia. Insbesondere in Zeiten von Corona schulterte sie die den Löwenanteil der Kinderbetreuung und verschaffte mir so die Freiräume, diese Arbeit fertig stellen zu können. Vielen Dank für deine Geduld, Unterstützung und Aufmunterung. Selbiges gilt natürlich auch für unseren Sohn. Ich liebe euch!

Danken möchte ich außerdem den vielen Mitgliedern der LHCb Kollaboration, insbesondere meinen hier gewonnenen Freunden. Ohne euch würde dieser Arbeit die Grundlage fehlen.

Nicht zuletzt danke ich meinen Freunden und Studienkollegen aus Rostock. Die letzten neun Jahre waren fantastisch. Ich werde meine Studienjahre, vor allem wegen euch, vermissen!

Darüber hinaus bedanke ich mich bei meinen Eltern für ihre seelische und finanzielle Unterstützung sowie für ihre hervorragend angenommene Rolle als Großeltern. Auch meinem Bruder möchte ich für seine Hilfe und die stetig aufmunternden Worte danken! Ohne euch würde diese Arbeit nicht existieren.

Vielen Dank euch allen!

“It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of light, it was the season of darkness, it was the spring of hope, it was the winter of despair.”
— Charles Dickens, from his book *A Tale of Two Cities*.