Search for $\Lambda_b \to K^-\mu^+$ at LHCb

Dissertation

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

submitted by

Dipl.-Phys. Oliver Grünberg

to the Faculty of Mathematics and Natural Sciences
of the University of Rostock, Germany

Rostock, 2016

urn:nbn:de:gbv:28-diss2017-0057-9

Referees:

Priv.-Doz. Dr. Roland Waldi

Institute of Physics, University of Rostock
Albert-Einstein-Str. 23
D-18059 Rostock, Germany

Prof. Dr. Stephanie Hansmann-Menzemer

Physikalisches Institut, Heidelberg University
Im Neuenheimer Feld 226
D-69120 Heidelberg, Germany

Opening of the doctoral degree procedure: 09.01.2017

Date of scientific colloquium: 07.04.2017

"Was wir alleine nicht schaffen
Das schaffen wir dann zusammen
Dazu brauchen wir keinerlei Waffen
Unsere Waffe nennt sich unser Verstand"

— Xavier Naidoo. *"Was wir alleine nicht schaffen"*.

Abstract

In this thesis the search for the decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ is presented using the Run I data of the LHCb experiment. Both decays offer an ansatz to the description of the matter-antimatter asymmetry in our universe and is based on the Sakharov conditions that predict the existence of processes violating the baryon and lepton number conservation. Especially the measured baryon-to-photon ratio $\eta \approx 6 \cdot 10^{-10}$ allows to estimate the probability of such processes in this order of magnitude. In the underlying data set including about 50 billion recorded Λ_b decays no signal events are observed and an upper limit on the combined branching fraction is determined at a confidence level of 90 %:

$$\left[\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.95 \cdot 10^{-9} \ .$$

The result is equivalent to the value of $[\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)]$ but excludes the large uncertainty of $\mathcal{B}(\Lambda_b \to pK^-) \approx 3.1 \cdot 10^{-6}$ that origins from another measurement. In the $\Xi_b^0 \to K^-\mu^+$ case the determined upper limit at a confidence level of 90 % is

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \left[\mathcal{B}(\Xi_b^0 \to K^-\mu^+) + \mathcal{B}(\Xi_b^0 \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.25 \cdot 10^{-9} \ ,$$

and includes the unknown production ratio of Ξ_b^0 to Λ_b. In addition an indication ($\chi = 2.7$) for the yet unobserved decay $\Xi_b^0 \to pK^-$ is found with 25 ± 10 signal events.

Kurzfassung

In dieser Dissertation wird die Suche nach den Zerfällen $\Lambda_b \to K^-\mu^+$ und $\Xi_b^0 \to K^-\mu^+$ mit den Run I Daten des LHCb Experiments präsentiert. Beide Zerfälle bieten einen Ansatz zur Beschreibung der Materie-Antimaterie Asymmetrie in unserem Universum auf Grundlage der Sacharowkriterien, welche die Existenz von Prozessen vorhersagen, die die Erhaltung der Baryon- und Leptonzahl verletzen. Insbesondere lässt sich aus dem gemessenen Baryon-zu-Photon Verhältnis $\eta \approx 6 \cdot 10^{-10}$ eben diese Größenordnung für die Wahrscheinlichkeit solcher Prozesse schätzen. In dem zugrundeliegenden Datensatz mit etwa 50 Milliarden aufgezeichneten Λ_b Zerfällen wurden keine Signalereignisse beobachtet und eine obere Grenze auf das kombinierte Verzweigungsverhältnis ermittelt mit einem Konfidenzniveau von 90 %:

$$\left[\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.95 \cdot 10^{-9} \ .$$

Das Resultat entspricht numerisch dem Wert von $[\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)]$, klammert jedoch die große Unsicherheit von $\mathcal{B}(\Lambda_b \to pK^-) \approx 3.1 \cdot 10^{-6}$ aus, welches aus einer anderen Messung stammt. Im Falle von $\Xi_b^0 \to K^-\mu^+$ wurde eine obere Grenze mit einem Konfidenzniveau von 90 % bestimmt:

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \left[\mathcal{B}(\Xi_b^0 \to K^-\mu^+) + \mathcal{B}(\Xi_b^0 \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.25 \cdot 10^{-9} \ ,$$

welches das unbekannte Produktionsverhältnis von Ξ_b^0 zu Λ_b beinhaltet. Des weiteren wurde ein Hinweis ($\chi = 2.7$) auf den bisher unbeobachteten Zerfall $\Xi_b^0 \to pK^-$ mit 25 ± 10 Signalereignissen gefunden.

Contents

"I bet our entire universe is in a tiny glass jar placed neatly on a shelf in some alien child's room as a science project he got a C on"

— Author unknown

1
Introduction

1.1 A universe made of matter

The genesis of a universe that contains only matter is vital for the existence of mankind and a central topic of cosmology. According to the Big-Bang theory the universe started to exist from an initial singularity of super high density and temperature. In the course of time the universe cooled down and expanded rapidly while running through eras dominated by different and partly unknown forces that finally left only matter. Figure 1.1 illustrates the evolution of the universe starting from the Big Bang at $t = 0$ until the creation of neutral atoms. Little is known about the times $t \lesssim 10^{-10}s$ that cannot be described by present theories.

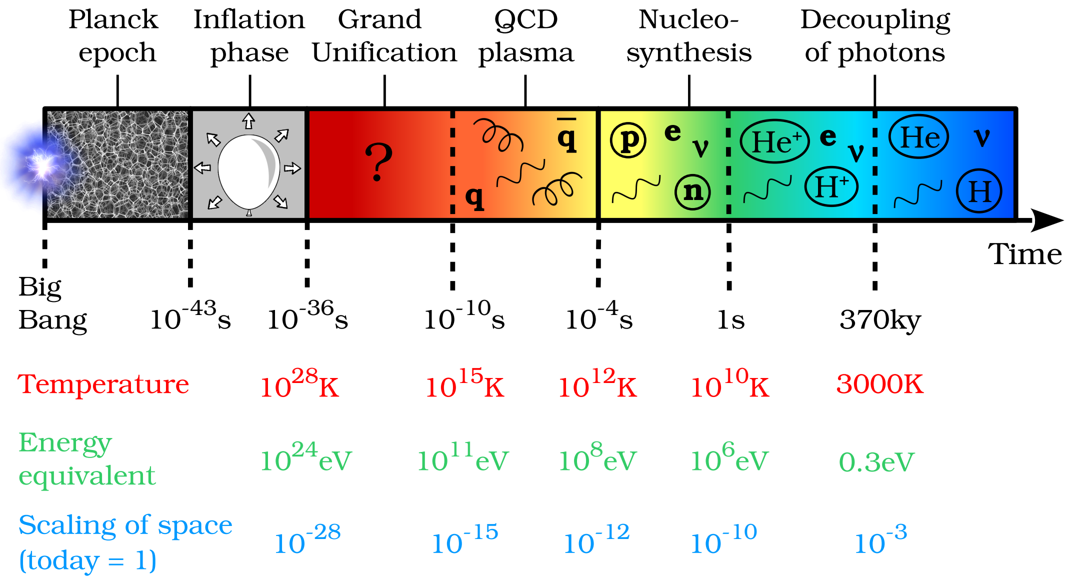


Figure 1.1: Evolution of the universe until the formation of the lighter elements hydrogen and helium. The shown temperatures, energies and scaling factors are derived from the Big-Bang cosmology model that is described below.

Already after the first second all anti-matter had disappeared and the universe was filled with a large number of photons and ordinary matter in form of a hot plasma of protons, neutrons and electrons. Today, the ordinary matter exists mainly in form of intergalactic gas of hydrogen and helium and more than one trillion galaxies [1] while the photons shine through the universe as an isotropic blackbody radiation with a temperature of about 2.73 K. An important hint on the processes that lead to the matter excess is the baryon-to-photon ratio $\eta \approx 6 \cdot 10^{-10}$, *i.e.* there is about one nucleon per 2 billion photons. In present cosmology this excess is attributed to a mutual extinction of matter and anti-matter and can be understood by the following simplified model: in a very early phase of the universe where $k_{\mathrm{B}}\,T \gg 2\,m_x\,c^2$, equal amounts of photons (γ), and particles of matter x and anti-matter $\bar{x}$ existed in an equilibrium state, $n_\gamma = n_x = n_{\bar{x}}$, that was dominated by the reactions $\gamma\gamma \leftrightarrow x\bar{x}$. During the cooling of the universe the creation of matter and anti-matter stopped when $k_{\mathrm{B}}\,T \approx 2\,m_x\,c^2$ and most of the matter and antimatter annihilated into photons. A tiny fraction in the order of η evaded extinction by undergoing dynamic processes that finally left only matter. In cosmology this process is called “baryogenesis” which actually means the creation of baryons and leptons, the constituents of matter. In 1966 Andrei Sakharov derived three conditions that have to be satisfied by “baryogenesis” [2].

1. **Baryon and lepton number violation (BLNV)**
   This condition is obvious assuming that the initial baryon and lepton number was zero.

2. **Violation of C and CP invariance**
   In order to prefer matter over antimatter, the rates of C or CP conjugated processes must be different. This concerns in particular decays of x and $\bar{x}$ or oscillations $x \leftrightarrow \bar{x}$.

3. **Deviation from thermodynamical equilibrium**
   During the cooling of the universe, phase transitions and times of non-equilibrium are likely to have existed to favour the creation of matter w.r.t. the reverse process.

A possible candidate for x is the heavy and short-living b quark that has been studied in many ways since its discovery in 1977. Although the Standard Model of particle physics (SM) forbids b decays under violation of the baryon and lepton number ($\mathbb{B}$ and $\mathbb{L}$), new interactions or particles may allow such decays with a probability in the order of η. Due to the proton-electron symmetry in the electrically neutral universe it is reasonable to additionally require $\mathbb{B}-\mathbb{L}$ to be conserved to create equal amounts of leptons and baryons at the same time. Another constraint comes from the long proton lifetime of $\tau(p) \gtrsim 10^{40}s$ that excludes the violation of $\mathbb{B}$ and $\mathbb{L}$ in the first family of fermions. Figure 1.2 sketches a process that violates $\mathbb{B}$ and $\mathbb{L}$ separately but conserves $\mathbb{B}-\mathbb{L}$; a heavy antiquark ($\overline{Q}$) decays to two lighter quarks (q) and a lepton (ℓ^-), *i.e.* matter is created from antimatter. This process might be allowed due to virtual contributions by unknown particles. In this example the C-conjugated process $Q \to \bar{q}\bar{q}\ell^+$ should occur at a lower rate (C violation) to finally produce more matter than antimatter starting from $n_Q = n_{\overline{Q}}$.

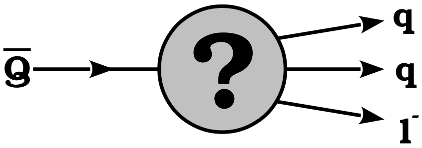


Figure 1.2: Illustration of a hypothetical decay $\overline{Q} \to qq\ell^-$ violating $\mathbb{B}$ and $\mathbb{L}$ but conserving $\mathbb{B}-\mathbb{L}$.

A direct way to search for BLNV in b quark decays is the exclusive reconstruction of b hadrons like Λ_b in such particular final states. LHCb is the first experiment that provides the necessary number of Λ_b to possibly observe these decays at a rate around η. A discovery would mark a milestone in high-energy physics and set a foundation for new theories about cosmological baryogenesis and extensions to the SM.

1.2 Big-Bang cosmology

The following section gives a brief introduction into the standard model of cosmology and the role of the parameter η.

In the 20th century the understanding of cosmology has made great progress by finding a set of equations by A. Friedmann and G. Lemaître that describe the expansion of a homogeneous and isotropic universe:

$$\text{Friedmann equation:} \quad H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\,\pi\,G}{3}\,\rho - \frac{k}{a^2} + \frac{\Lambda}{3}\ , \tag{1.1}$$

$$\text{Fluid equation:} \quad \dot{\rho} + 3\,(1+w)\,\rho\,\left(\frac{\dot{a}}{a}\right) = 0\ . \tag{1.2}$$

Both equations can be derived from general relativity and model the motion of the universe on large scales as a perfect fluid that can be characterised by its overall mass density ρ and pressure $p = w\,\rho$, where w is the equation of state parameter. The dimension of these scales is in the order of megaparsec at which the structure of the universe becomes regular. $a = r/r_0$ is the dimensionless scale factor of space expansion that is normalised to be one today. G is the gravitational constant, $k = (+1, 0, -1)$ is the spatial curvature constant and Λ is the cosmological constant. The Hubble parameter H describes the homogeneous expansion of space per length and time interval $H = \frac{dr}{dt\,r}$. This also leads to the useful relation $v = H \cdot r$ between the seeming escape velocity v of galaxies at large distances r.

The density of the universe ρ can be separated into three different components: the relativistically moving photons (γ) and neutrinos (ν), matter in form of ordinary matter (b) and cold dark matter (c), and vacuum energy (v) as a consequence of Λ. Thus the total density is $\rho_{\text{tot}} = (\rho_\gamma + \rho_\nu)_{\text{r}} + (\rho_{\text{b}} + \rho_{\text{c}})_{\text{m}} + \rho_{\text{v}}$ with $\rho_{\text{v}} = \Lambda/(8\,\pi\,G)$.

1.2.1 A simple model of cosmological expansion

On the basis of the Friedmann-Lemaître equations the evolution of the universe can be described on a simple yet adequate level assuming a flat universe with $k = 0$ in agreement with cosmological observations [3]. The solution of the fluid equation with $\dot{\rho} \neq 0$ leads to $\rho \sim a^{-3(1+w)}$. For relativistic particles w_{r} is 1/3 and for pressureless matter w_{m} is zero. Thus the dependency of densities on a is $\rho_{\text{r}} \sim 1/a^4$ and $\rho_{\text{m}} \sim 1/a^3$. The different development of ρ_{r} and ρ_{m} can be understood by the redshift of photons due to space expansion. Since $\rho_{\text{v}} = \mathit{const.}$ the fluid equation reduces to $w_{\text{v}} = -1$ and hence ρ_{v} causes a negative pressure. In the Friedmann equation this leads to the dependencies $(\dot{a}/a)^2 \sim (1/a^4)_{\text{r}} + (1/a^3)_{\text{m}} + (1)_{\text{v}}$. As a result this means that in an expanding universe the evolution is dominated first by radiation, later by matter, and much later by the vacuum energy. The time dependency of a can be derived from the Friedmann equation and is $a_{\text{r}} \sim t^{1/2}$ during the radiation phase, $a_{\text{m}} \sim t^{2/3}$ during matter domination and $a_{\text{v}} \sim \exp(t)$ for large times. The corresponding space expansion rates are $H_{\text{r}} = 1/(2t)$, $H_{\text{m}} = 2/(3t)$, and $H_{\text{v}} = \mathit{const.}$. In conclusion the expansion of the universe is relatively small in times when the positive radiation pressure was dominating and grows exponentially once the radiation and matter density becomes negligible and the negative pressure of the vacuum energy gains the upper hand. The development of the temperature along with space expansion can be obtained from Planck's law $\rho_\gamma \sim T^4$ which leads to the relation $T \cdot a = \mathit{const.}$ and hence $T \cdot a = T_0 \cdot a_0 = 2.73\,\text{K} \approx \mathcal{O}(1)$. Table 1.1 summarises the time dependencies of a, H, T and ρ during the different epochs.

Table 1.1: Dynamic of a, T, H, ρ_r and ρ_m during the three epochs of expansion.

Epoch	a	$H = \dot{a}/a$	$T \sim 1/a$	$\rho_\mathrm{r} \sim T^4$	$\rho_\mathrm{m} \sim T^3$
Radiation	$\sim t^{1/2}$	$1/(2t)$	$\sim t^{-1/2}$	$\sim t^{-2}$	$\sim t^{-3/2}$
Matter	$\sim t^{2/3}$	$2/(3t)$	$\sim t^{-2/3}$	$\sim t^{-8/3}$	$\sim t^{-2}$
Vacuum energy	$\sim e^{t}$	*const.*	$\sim e^{-t}$	$\sim e^{-4t}$	$\sim e^{-3t}$

1.2.2 Introduction of dimensionless parameters and η_{10}

In cosmology it is often made use of dimensionless parameters Ω and h instead of ρ and H. The cosmological density parameter $\Omega = \rho/\rho_\mathrm{crit}$ is defined with the help of the critical density $\rho_\mathrm{crit} \equiv 3\,H^2/(8\,\pi\,G)$. Therefore $\Omega_\mathrm{tot} = (\Omega_\gamma + \Omega_\nu)_\mathrm{r} + (\Omega_\mathrm{b} + \Omega_\mathrm{c})_\mathrm{m} + \Omega_\mathrm{v}$ and the Friedmann equation can be rewritten as $k/a^2 = H^2\,(\Omega_\mathrm{tot} - 1)$. Instead of H the scaled Hubble parameter h is defined by $H \equiv h \cdot 100\,\mathrm{km\,s^{-1}\ Mpc^{-1}} \approx \mathrm{h} \cdot 100\,\%/(10^{10}\,\mathrm{y})$. The development of $\Omega_i \sim \rho_i/H^2$ over time depends on the particular epoch. In the radiation era it can be stated $\Omega_\mathrm{r} = const.$, $\Omega_\mathrm{m} \sim \sqrt{t}$, and in the matter era $\Omega_\mathrm{r} = t^{-2/3}$ and $\Omega_\mathrm{m} = const.$. The normalised vacuum density develops as $\Omega_\mathrm{v} \sim t^2$ in both eras. Once the vacuum energy is dominating Ω_r and Ω_m decrease exponentially while Ω_v remains constant.

The baryon-to-photon ratio $\eta = \eta_{10} \cdot 10^{10} = n_\mathrm{b}/n_\gamma$ plays a central role in the description of the matter-antimatter asymmetry and the dynamic creation of nucleons (nucleosynthesis). n_b is approximately $\rho_\mathrm{b}/m_p = \Omega_\mathrm{b} \cdot \rho_\mathrm{crit}/m_p$ assuming that most of the baryons are protons. The number of photons per volume is $n_\gamma \sim T^3$ (see App. A.1). Hence the scaled baryon-to-photon ratio η_{10} can be determined from Ω_b, h and T (see App. A.2), and is time independent:

$$\eta_{10} = \frac{\Omega_\mathrm{b}\,\rho_\mathrm{crit}}{\langle m_\mathrm{b}\rangle\,n_\gamma} \cdot 10^{-10} \approx 273.5 \cdot \Omega_\mathrm{b}\,h^2\ . \tag{1.3}$$

Table 1.2 summarises the relevant cosmological parameters according to the Review of Particle Physics (RPP) [3]. These data suggest that our universe is flat ($k = 0$) and the total density is close to the critical one. Although ordinary matter and photons are everything that is known to mankind, it contributes only about 5 % to the total density budget. The content of the universe is dominated by the vacuum energy ($\Omega_\mathrm{v} \approx 70\,\%$) and cold dark matter ($\Omega_\mathrm{c} \approx 25\,\%$). It seems that the fate of the universe depends strongly on the exact properties of the vacuum energy that drives the expansion and is currently the most puzzling element in cosmology.

1.2.3 Time scale of the expansion

A time scale for the development of the universe can be obtained by finding a unique relation $T = f(t)$. If one assumes that the past expansion of the universe has been dominated by matter it can be stated $T = T_0 \cdot (t_0/t)^{2/3}$ with $T_0 \approx 2.73\,\mathrm{K}$ and $t_0 \approx 13.8\,\mathrm{Gyr}$. The relation is valid since the equilibrium time t_eq and temperature T_eq when the matter density started to overbalance the radiation density. T_eq can be found from the condition $\rho_\mathrm{m} = \rho_\mathrm{r}$ and hence $\Omega_\mathrm{m}\,\rho_\mathrm{crit}\,(T_\mathrm{eq}/T_0)^3 = \Omega_\mathrm{r}\,\rho_\mathrm{crit}\,(T_\mathrm{eq}/T_0)^4$. The solution is $T_\mathrm{eq} = \Omega_\mathrm{m}\,T_0/\Omega_\mathrm{r} \approx 9300\,\mathrm{K}$. This takes into account that the total relativistic energy density is $\Omega_\mathrm{r} = \Omega_\nu + \Omega_\gamma \approx 1.68\,\Omega_\gamma$ [3, 4] assuming that neutrinos are massless and have not heated up during the e^+e^- annihilation when $k_\mathrm{B}\,T$ became smaller than $2\,m_e$. The corresponding time is $t_\mathrm{eq} = t_0\,(T_0/T_\mathrm{eq})^{3/2} \approx 70\,\mathrm{kyr}$. Thus the coordinates $(T_\mathrm{eq}, t_\mathrm{eq})$ set the fix point for $T \sim t^{-1/2}$ in the radiation dominated era: $T = T_\mathrm{eq} \cdot (t_\mathrm{eq}/t)^{1/2}$. The resulting time scale of the Big Bang is presented in Fig. 1.1.

Table 1.2: Present cosmological parameters according to the RPP [3].

Property	Value	Property	Value
η_{10}	$5.8\ldots6.6$ (95 % CL)	Ω_{tot}	1.005 ± 0.017
h	0.678 ± 0.009	Ω_{v}	0.692 ± 0.012
T_0	$(2.7255 \pm 0.0006)\,\mathrm{K}$	Ω_{c}	0.258 ± 0.011
t_0	$(13.80 \pm 0.04)\,\mathrm{Gyr}$	Ω_{b}	0.0484 ± 0.010
c/H_0	$(1.374 \pm 0.018) \cdot 10^{26}\,\mathrm{m}$	Ω_{γ}	$(5.38 \pm 0.15) \cdot 10^{-5}$
ρ_{crit}	$\approx 5.1\, m_p/\,\mathrm{m}^3$	Ω_{ν}	$0.0012\ldots0.016$

1.3 Experimental results on η_{10}

In the past years the precise determination of η_{10} has made great progress and it became possible to determine its value independently following two different strategies. The first one is related to the theory of Big-Bang nucleosynthesis (BBN) that describes the primordial creation of the light isotopes deuterium (D), ^{3}He, ^{4}He, and ^{7}Li. The second one is based on the measurement of anisotropies in the cosmic microwave background (CMB). A comprehensive overview about the present status of BBN and the measurement of η can be found in Ref. [5].

The production of light elements took place in the radiation dominated universe around one second after the Big Bang. When the temperature fell below $k_{\mathrm{B}}\, T_f \approx 0.8\,\mathrm{MeV}$ the neutrons froze out from the $n \leftrightarrow p$ equilibrium and started to decay. Due to the low binding energy of deuterium, $E_{\mathrm{B}} = 2.2\,\mathrm{MeV}$, the initiation of nucleosynthesis was delayed until $k_{\mathrm{B}}\, T \approx E_{\mathrm{B}}/\ln(\eta^{-1}) \approx 0.1\,\mathrm{MeV}$ when the deuterium production rate overbalanced the destruction rate due to photo dissociation ("Deuterium bottleneck"). About 25 % of the neutrons decayed before the rest entered a chain of fusion processes that mainly left ^{4}He, some D and ^{3}He, and a tiny fraction of ^{7}Li. The final neutron-to-proton ratio was about 1/7. The theory of BBN is able to predict the primordial ($|_{\mathrm{p}}$) abundances of isotopes relative to hydrogen in dependence of η_{10}, the gravitational constant G, the mean neutron lifetime τ_n, the number of light neutrino families N_ν, and nuclear reaction cross sections. As a result the primordial abundance of D is particularly sensitive to η_{10} [5],

$$\begin{aligned}\mathrm{D/H}|_{\mathrm{p}} \sim (\eta_{10})^{-1.6} \cdot (\mathrm{G})^{0.95} \cdot (\tau_{\mathrm{n}})^{0.41} \cdot (\mathrm{N}_\nu)^{0.395} \\ \times\, \sigma(\mathrm{D\,D} \to {}^3\mathrm{He\,n})^{-0.53} \cdot \sigma(\mathrm{D\,D} \to \mathrm{T\,p})^{-0.47} \cdot \sigma(\mathrm{D\,p} \to {}^3\mathrm{He}\,\gamma)^{-0.31} \\ \times\, \sigma(\mathrm{p\,n} \to \mathrm{D}\,\gamma)^{-0.19} \cdot \sigma({}^3\mathrm{He\,n} \to \mathrm{T\,p})^{0.023} \cdot \sigma({}^3\mathrm{He\,D} \to {}^4\mathrm{He\,p})^{-0.012}\ .\end{aligned} \tag{1.4}$$

The rates of ^{4}He/H, ^{3}He/H, and ^{7}Li/H are proportional to $(\eta_{10})^{0.039}$, $(\eta_{10})^{-0.59}$, and $(\eta_{10})^{2.11}$, respectively. Though ^{7}Li/H has a stronger dependence on η_{10}, the sensitivity is deteriorated by theoretical and experimental uncertainties [5]. Recently it was possible to measure $\mathrm{D/H}|_{\mathrm{p}}$ precisely by the analysis of isotope-shifted absorption lines in the spectrum of distant and bright light sources behind intergalactic gas [6]. Since no substantial stellar chemical processing has taken place in the gas, the composition of elements has been mostly conserved since its creation. The determined abundance is $\mathrm{D/H}|_{\mathrm{p}} = (2.53 \pm 0.04) \cdot 10^{-5}$ which corresponds to $\eta_{10} \approx 6.03$.

An even more precise measurement of η_{10} was achieved by the analysis of the CMB by the *PLANCK* collaboration [7]. The CMB is a snapshot of the universe from the era of the last photon scattering with matter about 370 kyr after the Big Bang. All information about the structure of the universe at that time are saved in it. *PLANCK* has analysed anisotropies in the angular distribution of the CMB temperature that can partly be attributed to baryonic-acoustic-oscillations (BAO). In short, these are periodic density fluctuations of the baryon-electron plasma in the early universe caused by the interplay of attractive gravitational forces of dark matter and repulsive forces from the radiation pressure. These anisotropies can be connected to properties of the matter content in the early universe [8]. *PLANCK* published $\Omega_\mathrm{b}\, h^2 = (2.226 \pm 0.023) \cdot 10^{-2}$ [9] which corresponds to $\eta_{10} = 6.09 \pm 0.06$ according to Eq. 1.3.

Figure 1.3 shows the calculated abundances of the light elements according to BBN theory in dependence of η_{10} and $\Omega_\mathrm{b}\, h^2$ as well as different measurements according to the current RPP [3]. There is a notable disagreement of the observed ^{7}Li abundance with $\eta_{10} \approx 6$ that has not been understood yet ("Lithium problem").

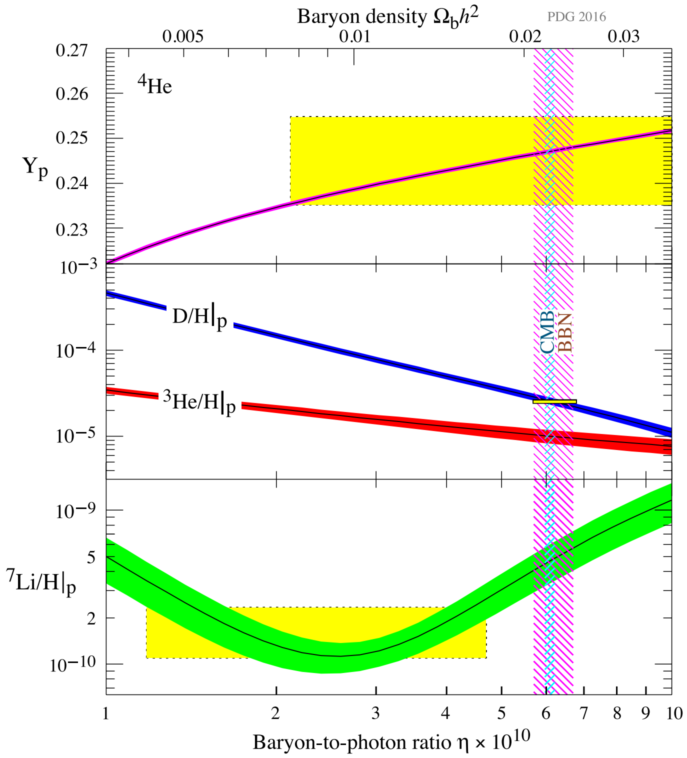


Figure 1.3: Results from BBN theory and experiment on η_{10} according to the RPP [3].

1.4 The Standard Model of particle physics

The present knowledge about elementary particles and their interactions is summarised in the Standard Model[1] of particle physics. In this theory the elementary particles are separated by their intrinsic spin S into the class of bosons that have integer spin $S/\hbar = 0, 1, 2$, etc. and fermions that have half-integer spin $S/\hbar = 1/2, 3/2$, etc.. There are three fundamental forces described by the SM: the strong force, the electromagnetic force and the weak force. In order to undergo one of these interactions a fermion must carry a corresponding charge: colour charge for the strong interaction, electric charge (Q) for the electromagnetic (EM) interaction, and weak isospin for the weak interaction. The known elementary fermions all have $|S| = 1/2$ and are divided into quarks, q, that participate in all interactions, charged leptons ℓ^- that interact weakly and electromagnetically, and neutral leptons ν (neutrinos) that only interact weakly. In summary

---

[1] It is actually more of a theory.

there are six flavour types of quarks and three lepton flavours that can be distinguished, see Tab. 1.3. Among the fermions there is a mass hierarchy that allows to group these particles into three families. In addition to each presented fermion there is an antifermion that has the same mass and spin but carries opposite-sign charges, *e.g.* $\bar{u}$, μ^+, $\overline{\nu}_\tau$. A special feature of the strong interaction is the color confinement that forces quarks to be bound in color singlet states called hadrons. The most common hadrons are mesons $= |q\bar{q}\rangle$ and baryons $= |qqq\rangle$.

Table 1.3: Elementary fermions of the Standard Model. The stated masses are rounded and the quark masses are calculated using the $\overline{\text{MS}}$ scheme [3].

	Quarks			Leptons		
Family	Flavour	Mass (MeV/c^2)	Q (e)	Flavour	Mass (MeV/c^2)	Q (e)
1.	u	2	+2/3	ν_e	< 0.002	0
	d	5	−1/3	e^-	0.51	−1
2.	c	1270	+2/3	ν_μ	< 0.2	0
	s	100	−1/3	μ^-	105.7	−1
3.	t	$170 \cdot 10^3$	+2/3	ν_τ	< 20	0
	b	4180	−1/3	τ^-	1777	−1

Within the SM three of the four known fundamental forces are described in the framework of quantum field theories that allow to make quantitative predictions such as scattering cross sections and decay rates. Interactions between particles are mediated by gauge bosons that transfer momentum and energy and hence make dynamic processes possible in the first place. Table 1.4 summarises the three forces and their mediators. Gauge bosons may also interact with each other which is in particular true for gluons. Another essential component of the SM is the Higgs mechanism that describes how bosons and fermions gain masses in a quantum field theory. The discovery of a new particle at the LHC in 2012-2013 that is compatible with the predicted Higgs boson (H^0) is widely considered as confirmation of the Higgs mechanism and thus completes the SM to be a consistent theory up to energies around 1 TeV. However, at the time being it is unclear which extensions have to be made to describe processes at even higher energies.

Beside the SM there is gravitation that is too weak to play a role in present particle physics and that cannot be described by quantum theory yet. Effects of quantum gravitation will contribute at particle energies around 10^{19} GeV (Planck scale).

Table 1.4: Forces and gauge bosons of the Standard Model with masses according to the RPP [3]. The relative strengths consider the force between two neighbouring protons in a nuclei.

Force	Relative strength	Typical time scale	Mediator	Mass (GeV/c^2)
Strong	100	$10^{-23}s$	gluon (G)	0
EM	1	$10^{-18}s$	photon (γ)	0
Weak	10^{-5}	$10^{-12}s$	$W^\pm$	80.4
		(b decays)	Z^0	91.2

1.5 Phenomenological considerations

In the following section the decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ are classified within the SM. Here and in the rest of the document charge conjugated (C) processes are always implied unless stated otherwise. This means the formulation $\Lambda_b \to K^-\mu^+$ also includes the decay $\overline{\Lambda}_b \to K^+\mu^-$ and observables such as the invariant mass $m(K^-\mu^+)$ also consider $K^+\mu^-$ combinations.

1.5.1 b quarks in the early universe

During the radiation era the b quarks have frozen out from the $\gamma\gamma \leftrightarrow b\bar{b}$ equilibrium around $k_\mathrm{B}\,T = 2\,m_b\,c^2 \approx 10\,\mathrm{GeV}$ at $t \approx 10^{-8}s$. At these high temperatures the b quarks can be expected to exist freely in a quark-gluon plasma. The critical temperature for the phase transition of quarks and gluons into hadrons is expected to be around $k_\mathrm{B}\,T_\mathrm{c} \approx 150\,\mathrm{MeV}$ [10] that was achieved at about $t = 10^{-4}s$ and marks the beginning of the hadron era. At this time all b quarks have already decayed since the b quark lifetime is about $10^{-12}s$.

1.5.2 Motivation for $\mathbb{B} - \mathbb{L} = const.$

Beside the observed matter asymmetry in the visible universe there is also good evidence for a charge neutral universe [2], *i.e.* $n_p = n_e$. Therefore it is reasonable to assume that $\mathbb{B}$ and $\mathbb{L}$ violation occurred in a correlated process to finally obtain an equal amount of protons and electrons. A simple way to assure a symmetric production of baryons and leptons is to claim $\mathbb{B} - \mathbb{L} = const.$ which is equivalent to $\Delta\mathbb{B} = \Delta\mathbb{L}$.

Another motivation for $\mathbb{B} - \mathbb{L} = const.$ comes from spin conservation, $\vec{J} = \vec{L} + \vec{S} = const.$. Since fermions have $|S| = 1/2$ but the angular momentum L is quantized in steps of one, fermions have to be created pairwise so that the overall spin can be conserved.

The searched decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ violate $\mathbb{B}$ and $\mathbb{L}$ separately but conserve $\mathbb{B} - \mathbb{L}$. Since the flavour of the initial state is unknown, *i.e.* a signal in $m(K^-\mu^+)$ around the Λ_b mass could be a Λ_b or a $\overline{\Lambda}_b$, this analysis also explores the possibility of $\mathbb{B} + \mathbb{L} = const.$.

1.5.3 Properties of the participating particles

The four relevant particles of the searched decays are the baryons Λ_b and Ξ_b^0, the charged K meson, and the muon lepton. Some of their most relevant properties are summarised in Tab. 1.5.

Table 1.5: Properties of the participating particles according to the RPP [3].

Particle	Mass (MeV/c^2)	Lifetime τ	$I(J^P)$	$(\mathbb{B}, \mathbb{L})$
μ^+	105.7	2.197 µs	$0(1/2)$	$(0, -1)$
K^-	493.677 ± 0.016	(12.380 ± 0.021) ns	$1/2(0^-)$	$(0, 0)$
Λ_b	5619.51 ± 0.23	(1.466 ± 0.010) ps	$0(1/2^+)^*$	$(1, 0)$
Ξ_b^0	5791.9 ± 0.5	(1.464 ± 0.031) ps	$1/2(1/2^+)^*$	$(1, 0)$

* Quark model prediction

The initial particles Λ_b and the Ξ_b^0 are baryons with the quark content $\Lambda_b = |bdu\rangle$ and $\Xi_b^0 = |bsu\rangle$. The Λ_b is the lightest baryon that contains a b quark. In terms of the strong isospin it is the singlet state $(I = 0, I_3 = 0)$ of the $3 \oplus 1$ multiplet $(\Sigma_b^+, \Sigma_b^0, \Sigma_b^-) \oplus (\Lambda_b)$. It has the same quark content as the triplet member Σ_b^0 $(I = 1, I_3 = 0)$ but a different isospin wave function:

---

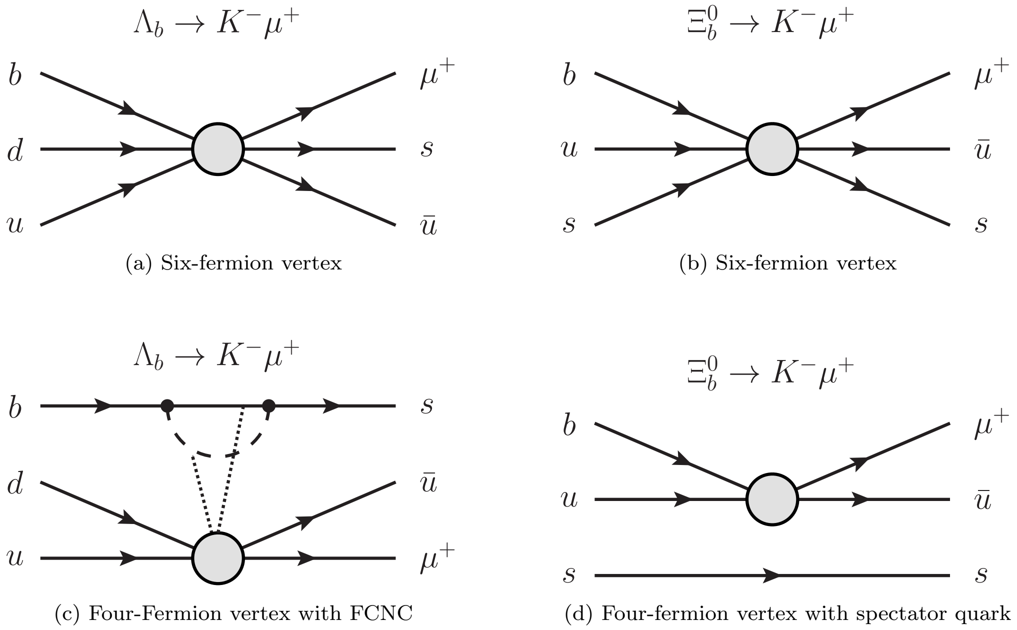
[2] Ref. [11] reports an excess charge per baryon $q_{e-p} < 10^{-26}e$

$$\Lambda_b = \frac{1}{\sqrt{2}}\left(|bud\rangle - |bdu\rangle\right) \ , \ \Sigma_b^0 = \frac{1}{\sqrt{2}}\left(|bud\rangle + |bdu\rangle\right)$$

Due to the large mass difference $\Delta m(\Sigma_b, \Lambda_b) \approx 190\,\mathrm{MeV}/c^2$ all Σ_b decay almost instantly into $\Lambda_b \pi$ via the strong force. The Ξ_b^0 is a member of the isospin doublet (Ξ_b^0, Ξ_b^-). Besides the mass and mean lifetime of Λ_b and Ξ_b^0 only little is known about exclusive branching fractions and possible CP violation in their decays. The $K^- = |s\bar{u}\rangle$ is the lightest meson with a s quark and has been studied in detail since its discovery in 1947. It has a relatively long average lifetime and decays only weakly in the SM. The most frequent final states are $\mu\bar{\nu}$ (64 %) and $\pi^+\pi^0$ (21 %). The muon is the charged lepton of the second fermion family with a very long average lifetime of about 2.2 µs. Like the kaons it decays weakly and is well described by the SM.

1.5.4 Decay diagrams

Although the searched decays are forbidden in the SM, it is possible to picture them in quark-flow diagrams and discuss them on a qualitative level. Figure 1.4 shows four conceivable diagrams of the decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$. The topology of the processes can be divided into three types. Figure 1.4(a,b) depicts a six-fermion vertex where all initial quarks vanish and a μ^+, a s quark, and a $\bar{u}$ quark are created. The actual interaction is hidden in the big vertex circle and requires some sort of new interaction due to the BLNV. Additional SM processes may be also contained. In Figure 1.4(c) the b quark is converted into a s quark due to flavour-changing neutral currents (FCNC) $b \to tW^- \to s$ where the W^- and/or the t quark possibly emit a virtual particle that interacts with the four-fermion vertex. The diagram 1.4(d) of the decay $\Xi_b^0 \to K^-\mu^+$ is an alternative to Fig. 1.4(b) in which the initial s quark does not participate in the interaction.



(a) Six-fermion vertex (b) Six-fermion vertex

(c) Four-Fermion vertex with FCNC (d) Four-fermion vertex with spectator quark

Figure 1.4: Possible quark-flow diagrams of the decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$.

1.5.5 Definition of the branching fraction

The partial decay rate of a two-body decay $i \to f = f_1 + f_2$ can be stated using "Fermi's rule"

$$\lambda_{if} = \lambda(i \to f) = \frac{2\pi}{\hbar} \int \left| \sum \mathcal{M}_{if}(p_1, p_2) \right|^2 \, d\varrho_f \; . \tag{1.5}$$

$\left| \sum_f \mathcal{M}_{if}(p_1, p_2) \right|^2$ is the coherent sum over all matrix elements $\mathcal{M}_{if} \sim \langle i | \hat{H}_{int} | f \rangle$ that contribute to the decay and generally depend on the four-momenta p_1 and p_2 of the final state particles. $\hat{H}_{int}$ is the Hamiltonian of the underlying interaction. $d\varrho_f = |\vec{p}|/(16\,\pi^2\, m_i) d\Omega$ is the differential phase space density of a two-body decay that only depends on the angular distribution relative to the spin of the initial state particle. $|\vec{p}| = |\vec{p}_1| = |\vec{p}_2|$ is the absolute momentum in the center-of-mass frame of the intial particle. The branching ratio is defined as

$$\mathcal{B}(i \to f) = \frac{\lambda_{if}}{\sum_{f'} \lambda_{if'}} = \lambda_{if} \cdot \tau(i) = 0 \ldots 1 \; , \tag{1.6}$$

where in the denominator the sum is made over all possible final states. The last result makes clear that the branching fraction $\mathcal{B}(\Lambda_b \to K^- \mu^+) = \lambda(\Lambda_b \to K^- \mu^+) \cdot \tau(\Lambda_b)$ can be expected to be much larger than $\mathcal{B}(\Sigma_b^0 \to K^- \mu^+)$ because $\tau(\Sigma_b^0) \approx 10^{-10}\, \tau(\Lambda_b)$ and assuming that $\lambda(\Sigma_b^0 \to K^- \mu^+) \approx \lambda(\Lambda_b \to K^- \mu^+)$ due to the same quark content of Σ_b^0 and Λ_b.

1.5.6 Connection between η and $\mathcal{B}(\Lambda_b \to K^- \mu^+)$

It is possible to describe an excess of matter and the value of η due to Λ_b decays with baryon and lepton number violation. In App. A.4 a toy model is described that starts with an equal amount of Λ_b and $\overline{\Lambda}_b$ and assumes two kinds of decays. The first one are SM-type decays that require $\Delta\mathbb{B} = \Delta\mathbb{L} = 0$, *e.g.* $\Lambda_b \to pK^-$, and are denoted as $\Lambda_b \to \mathsf{B} + X$ where B is a baryon. The other one are exotic decays with $\Delta\mathbb{B} = \Delta\mathbb{L} = -1$, *e.g.* $\overline{\Lambda}_b \to K^+ \mu^-$, and denoted as $\overline{\Lambda}_b \to \mathsf{L} + X$ where L is a lepton. In both cases X stands for a generic set of particles with $\mathbb{B} = \mathbb{L} = 0$. An equal excess of baryons and leptons can be created if there is also CP violation in these decays, for example if $\mathcal{B}(\overline{\Lambda}_b \to K^+ \mu^-) > \mathcal{B}(\Lambda_b \to K^- \mu^+)$. The major result of this model is the relation

$$\eta \approx a_{\mathsf{L}}\, \mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X) = a_{\mathsf{B}}\, \mathcal{B}(\Lambda_b \to \mathsf{B} + X) \; . \tag{1.7}$$

$\mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X)$ and $\mathcal{B}(\Lambda_b \to \mathsf{B} + X)$ are the inclusive branching fractions of both decay types. $a_{\mathsf{B,L}} = 0 \ldots 1$ are the overall decay asymmetries of both decay types to create more fermions (B and L) than antifermions ($\overline{\mathsf{B}}$ and $\overline{\mathsf{L}}$). The rest of the Λ_b and $\overline{\Lambda}_b$ and their decay products annihilate into photon pairs. The result shows that only a combination of CP violation and BLNV leads to a matter excess and that, in principle, a branching fraction $\mathcal{B}(\Lambda_b \to K^- \mu^+) \lesssim \eta / a_{\mathsf{L}}$ in the order of 10^{-10} is possible. However, the model has the shortcoming that only b quarks existed in the early phase of the universe and no Λ_b baryons. As a result it can be concluded that if primordial b quarks are at least partly responsible for the matter-antimatter asymmetry in the universe and their decay properties are measurable in Λ_b decays, the decay $\Lambda_b \to K^- \mu^+$ may exist with a branching fraction in the order of η.

1.5.7 Implication for the proton lifetime

Due to general symmetry arguments (crossing symmetry) the existence of $\Lambda_b \to K^- \mu^+$ would have an influence on the proton lifetime due to the virtual b quark content of the proton. One possible way for the proton to decay would be $p \to (\Lambda_b W^+)^* \to K^- \mu^+ \pi^+$ where the proton decays to a virtual $(^*)$ $\Lambda_b W^+$ pair via $u \to bW^+$ which is allowed within short times due to the uncertainty principle. In the end a final state is reached that is lighter than the proton mass via the decays $\Lambda_b \to K^- \mu^+$ and, for example, $W^+ \to u\bar{d} = \pi^+$. Reference [12] presents calculations on the basis of the long proton lifetime that lead to upper limits on branching fractions of baryon and lepton number violating decays in the order of 10^{-30}, *e.g.* $\mathcal{B}(B^0 \to \Lambda_c \ell^-) < 4 \cdot 10^{-30}$. However, a solution could be the interference between the Λ_b and its isospin partner Σ_b^0. Since both baryons have the same quark content, there are always two processes that have to be considered: $p \to (\Lambda_b W^+)^* \to K^-\mu^+ + X$ and $p \to (\Sigma_b^0 W^+)^* \to K^-\mu^+ + X$. Both amplitudes may cancel due to destructive interference. As a result $\Lambda_b \to K^-\mu^+$ could be allowed without affecting the proton lifetime.

1.5.8 Prior measurements

There have been a number of searches for BLNV in the past whose results are tabulated in App. A.5. A very stringent measurement comes from the Super-Kamiokande experiment (Super-K) that searched for proton decays into $\ell^+\pi^0$ leading to the result $\tau_p \gtrsim 10^{33}$ yr which practically rules out BLNV in the first family of fermions. Two other notable measurements that involve fermions of the second and third family are $\mathcal{B}(B^- \to \Lambda\mu^-) < 6.2 \cdot 10^{-8}$ by BaBar and $\mathcal{B}(\tau^- \to \overline{p}\mu^+\mu^-) < 3.3 \cdot 10^{-7}$ by LHCb. However, none of the measurements involving fermions of the two higher families reached a sensitivity in the order of $\eta \approx 10^{-10}$.

1.6 Beyond the Standard Model

Despite its big success the present SM implies a number of problems which lead to the conclusion that there must be some sort of superior theory (Grand Unified Theory - GUT) that sets a common foundation for the three known forces and can describe different aspects of the SM at once. These are, for example, the mass hierarchy and flavour structure of quarks and leptons, and CP violation and mixing in the quark and neutrino sector. A prototype for a GUT is the SU(5) theory [13] by H. Georgi and S. Glashow that considers the three forces of the SM as different manifestations of one fundamental interaction with one supercharge and one coupling constant. The unification breaks into the known interactions below a certain energy $\sqrt{s} = m_X$ where m_X corresponds to the mass scale of the SU(5) gauge bosons. Although the theory has failed due to a too small prediction of the proton lifetime it still acts as a role model and demonstrates the challenges and pitfalls of a GUT. Today there are many ideas for a GUT that all have their advantages and shortcomings. One common aspect of GUTs is the description of quarks and leptons as members of a fermion (f) multiplet that may interact with each other via a gauge boson X as depicted in Fig. 1.5(a). Such interplay between quarks and leptons automatically results in a proton decay as illustrated in Fig. 1.5(b). The proton lifetime is calculable and given by $\tau_p \sim m_X^4/m_p^5$. A proton lifetime consistent with the measured lower limit of $\tau(p) \gtrsim 10^{33}$ yr requires masses m_X in the order of 10^{15} GeV/c^2 or higher.

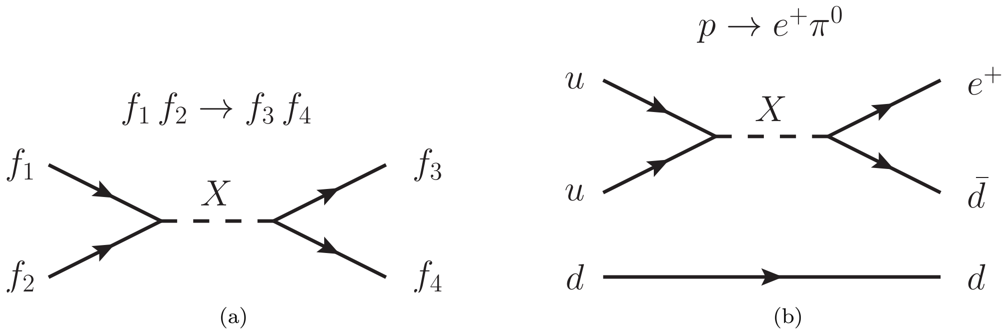
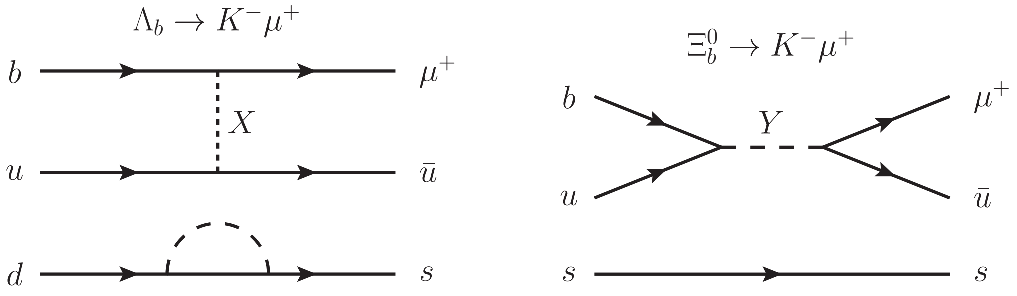


Figure 1.5: Fundamental fermion interaction of a GUT via the gauge boson X and, as a consequence, the allowed proton decay.

1.6.1 Leptoquarks

Leptoquarks (LQs) are hypothetical bosons that carry both baryonic and leptonic quantum numbers and allow interactions between leptons and quarks as expected in any GUT. The first ideas about LQs came up by J.Pati and A.Salam that proposed the lepton number to be a fourth "color" [14]. A possible representation of LQs is X^C and Y^C and their respective antiparticles where $C = \{r, g, b\}$ is the color charge and $Q(X) = +4/3\,e$ and $Q(Y) = +1/3\,e$. However, LQs exist in many extensions of the SM with model-dependent couplings and masses, and may induce flavor-changing neutral currents and lepton-family violation. Direct and indirect searches at collider experiments set model-dependent lower limits on their masses in the order of $500 \ldots 1000\,\mathrm{GeV}/c^2$, see also the review about Leptoquarks in Ref. [3]. Figure 1.6 shows two Feynman diagrams of $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ that are made possible by LQs. If LQs have a coupling constant that is proportional to the family number minus one, the direct interaction in the first family of fermions is forbidden and the proton decay is suppressed while BLNV processes are allowed in the second and third family.



(a) (b)

Figure 1.6: Possible Feynman diagrams of the decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ including LQs.

“Measure what is measurable,
and make measurable what is not so.”
— Galileo Galilei (1564 - 1642)

2

The LHCb experiment

The data for this analysis were recorded at the LHCb detector that is situated at the Large Hadron Collider (LHC) complex at the European Organisation for Nuclear Research (CERN) near Geneva. The following chapter gives a brief description of the LHCb experiment at the LHC where heavy b and c hadrons are created and measured.

2.1 The Large Hadron Collider

The LHC is a ring accelerator for protons and heavy ions that was built between 1998 and 2008. It is installed 175 m below the surface in a 27 km long tunnel around the CERN area. In the main operation mode there are two adjacent beams of oppositely flying protons that are boosted to a maximum energy of 6.5 TeV per beam. At four dedicated interaction points the two beams are crossing to induce head-on collisions of protons to create new heavy particles like b quarks. Each interaction point is home to a particle physics experiment, namely ALICE, ATLAS, CMS, and LHCb, where data of the collisions are recorded. Figure 2.1 gives a schematic overview about the LHC ring and the position of the four experiments. The main focus of the LHC experiments is the study of the fundamental forces in nature, the constituents of matter, and the discovery of new particles and interactions. The LHC has completed its first period of data taking (Run I) between April 2010 and February 2013 at center-of-mass (cms) energies of $\sqrt{s} = 7\,\text{TeV}$ until 2011 and 8 TeV afterwards. Some major outcomes of that time are the discovery of the Higgs boson by ATLAS [15] and CMS [16], the observation of the very rare decay $B_s^0 \to \mu^+\mu^-$ by CMS [17] and LHCb [18], as well as the first-time reconstruction of pentaquarks and other exotic hadrons by LHCb [19,20]. Summing over all participating scientists and engineers, the LHC involves more than 10 000 people from over 100 countries. At the time being it is the largest and most complex physical experiment ever to be developed and operated by men.

Before protons are injected into the LHC they run through a succession of smaller accelerators that speed them up to an energy of 450 GeV and at the same time pack them into bunches. In the LHC ring there are a couple of radio-frequency cavities for the final acceleration. The major challenge is the stable running of the beams while the connected experiments are using them for proton-proton (pp) collisions. Therefore the LHC has installed more than 1600 superconducting magnets along the ring to bend, deflect, and focus the beams. The cooling is achieved by 120 tons of superfluid Helium-4 at a temperature of 1.9 K. To minimise the interaction with beam gas remnants an ultra vacuum of less than 10^{-10} mbar is created within the pipes. During the first years of running both LHC beams consisted of up to 1380 bunches with a population in the order of 10^{11} protons. Each bunch had a nominal length of 9.4 cm along the beam axis and a radius of roughly 19 µm. In the standard configuration the bunches were collided with a rate of 20 MHz and at a full crossing angle of 0.29 mrad in the vertical plane leading to a maximum luminosity of $7.7 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$.

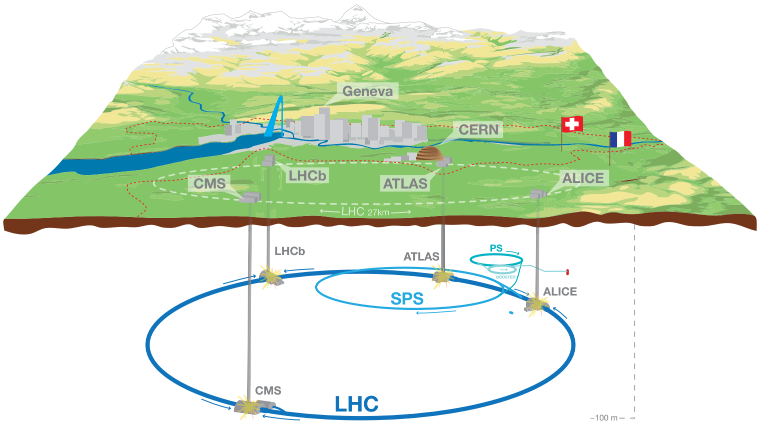


Figure 2.1: Overview of the LHC complex and the affiliated experiments at CERN [21].

2.2 The LHCb detector

LHCb is a dedicated heavy flavour experiment aiming for the reconstruction of decays of heavy hadrons containing b and c quarks. Its main physics programme is the search for new sources of *CP* violation as well as rare and forbidden decays in the SM to explain the matter-antimatter asymmetry in the universe. Therefore LHCb conducts precision measurements of physics parameters such as branching ratios and CP asymmetries to seek for deviations between experimental results and SM predictions that could be explained by new particles and interactions. Such indirect searches can be sensitive to new phenomena manifesting themselves as quantum corrections to physical processes at scales beyond the actual cms energy of the protons and complement direct searches by ATLAS and CMS. In addition, LHCb is able to address a variety of other physics questions like hadron spectroscopy and studies of quark-gluon plasma in proton-lead collisions. The overall cost for the development and commissioning of the LHCb detector is about 65 million Euro (BMBF calculations).

In the following sections the set up and basic functionality of the LHCb detector is explained. More detailed information can be found in Refs. [22, 23].

2.2.1 Detector setup and Run I data taking

The LHCb detector is a single-arm spectrometer for charged particles and photons in the momentum range of $5-200\,\text{GeV}/c$. It has a forward angular coverage from about 15 mrad to 300(250) mrad in the bending (non-bending) plane which corresponds to a pseudorapidity of $\eta = 1.8 - 4.9$. Figure 2.2 shows the general layout of the detector during Run I. A description of the components is given in the next sections.

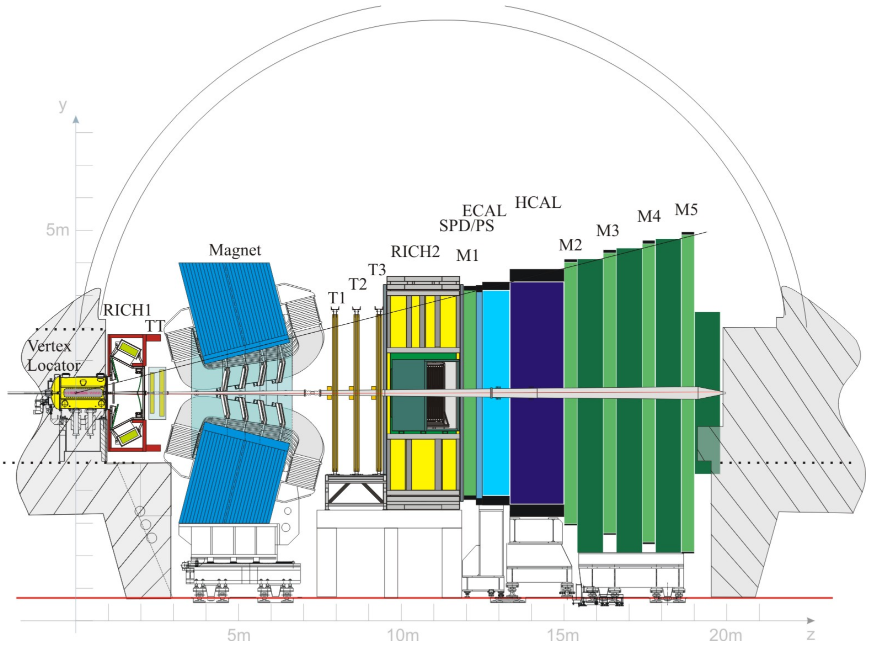


Figure 2.2: Layout of the LHCb detector during Run I [24].

The design is driven by the fact that the $b\bar{b}$ quark pair production mainly happens in a narrow cone in the forward or backward direction around the beam axis. Figure 2.3 shows the simulated $b\bar{b}$ production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ and compares the acceptance regions of LHCb to the general purpose detectors (GPD) ATLAS and CMS. According to the simulation LHCb detects either one or both of the produced b quarks with a probability of 27 % and 25 %.

The first pp collisions in LHCb took place at the end of 2009 at $\sqrt{s} = 0.9\,\text{TeV}$ and provided a small amount of data to finalise the commissioning of the detector and the reconstruction software. In 2010 the cms energy increased to $\sqrt{s} = 7\,\text{TeV}$ along with the luminosity and allowed first studies of the $b\bar{b}$ cross section [25]. In the years 2011 and 2012 the LHC delivered stable beams for long periods and enabled LHCb to record more than $3\,\text{fb}^{-1}$ of high-quality data for analyses. A key role played the technique of luminosity levelling; by adjusting the beam offset in the transversal plane the luminosity and the number of visible pp interactions (pile-up) can be kept at a constant and small value for the detector to cope with. This avoids premature ageing of the components and ensures steady data taking conditions. In 2011 the majority of the data was taken at a luminosity of $3.5 \cdot 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$ and a pile-up of 1.4, and in 2012 the average luminosity was $4.0 \cdot 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$ with a pile-up of 1.8. A multistage trigger system selects events of interest from the original collision rate of 20 MHz to a retention rate of several kilohertz. The polarity of the dipole magnet was regularly inverted to cancel out left-right asymmetries in charge detection in the combined data. In summary the average operational efficiency, defined as the ratio of recorded luminosity over delivered luminosity, during Run I was about 93 %. The missing 7 % are mainly due to dead times in the read-out system and short technical problems. About 99 % of the recorded data is available for physics analyses.

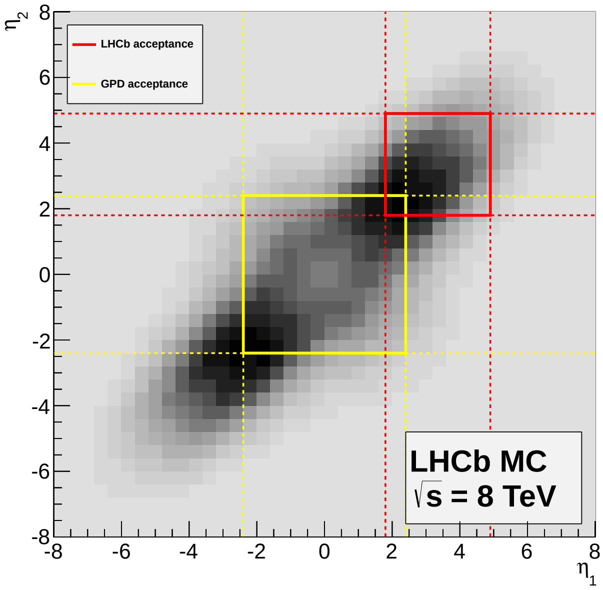


Figure 2.3: Correlation of the $b\bar{b}$ production in dependence of their pseudorapidities according to simulations by Pythia8 and CTEQ6 NLO [24].

2.2.2 Recorded luminosities

Table 2.1 summarises the available integrated luminosities per year and magnet polarity for this analysis. The measurement of the integrated luminosity is based on two independent methods that are carried out by LHCb: “van der Meer scan” and “beam-gas imaging” [26]. The combined relative uncertainty is about 1.2 %. Throughout this work the formulation “2011 data” and “2012 data” is synonymous for data recorded at $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$, respectively.

Table 2.1: Integrated luminosity per magnet polarity that is available for data analysis.

$\int \mathcal{L}\,(\text{pb}^{-1})$	2011 ($\sqrt{s} = 7\,\text{TeV}$)	2012 ($\sqrt{s} = 8\,\text{TeV}$)
MagDown	573(1)	1027(2)
MagUp	434(1)	1035(2)
Sum	1007(1)	2062(3)
Ratio	1.32 : 1	≈ 1 : 1

2.2.3 Charged particle reconstruction

The reconstruction of long-lived charged particles, namely $e^\pm$, $\mu^\pm$, $\pi^\pm$, $K^\pm$ and $p/\overline{p}$, and the measurement of their momenta is the major task of the LHCb detector. Charged particles traversing the detector are reconstructed from hits in the tracking system that consists of the Vertex Locator (VELO), the Tracker Turicensis (TT), and the T stations (T1-3). The curvature of the tracks for the momentum measurement is caused by a warm dipole magnet between the TT and the T stations providing a bending power of 4 Tm. Depending on which tracking stations have been passed the tracks are classified into Long tracks, Upstream tracks, Downstream tracks, VELO tracks and T tracks. Figure 2.4 illustrates the tracking system of LHCb and the track types. The magnetic field is aligned along the vertical y axis. Its magnitude depends on the z position and is small within the tracking stations which results in straight trajectories here. This analysis uses only Long tracks which have the best momentum and spatial resolution.

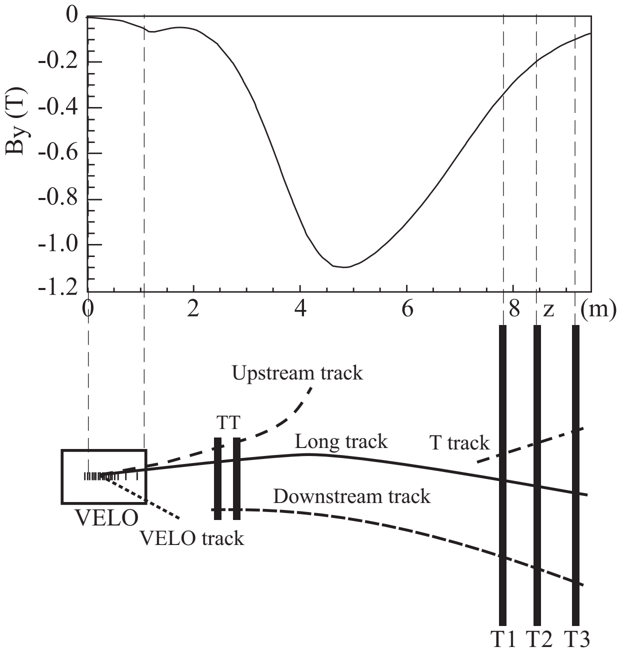
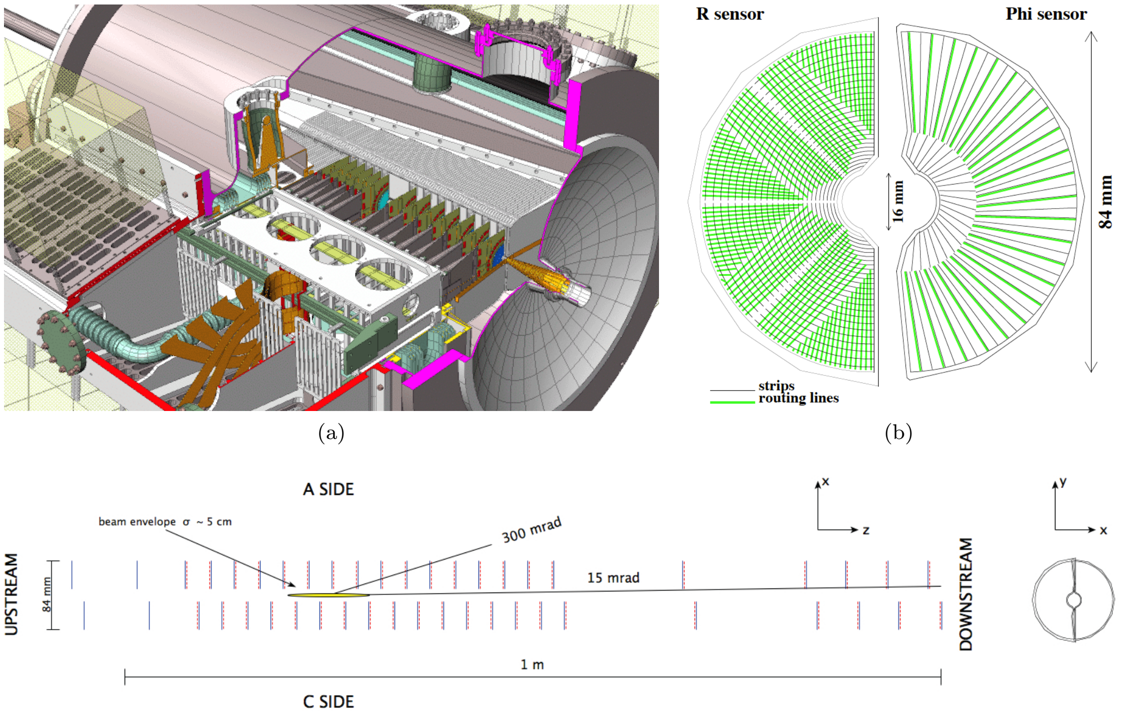


Figure 2.4: Illustration of the LHCb tracking system including a map of the magnetic field strength and definitions of the track types [23].

VELO

The task of the VELO is the precise detection of primary vertices (PVs) from pp collisions and displaced secondary vertices (SVs) of decaying b and c hadrons by reconstructing tracks coming from these points. It is located around the interaction point in a 1.4 m long evacuated vessel with a diameter of about 1.1 m. In the center of the VELO is a series of silicon strip modules to measure the r and ϕ coordinates of hits induced by tracks penetrating and hence ionizing the sensitive material. The basic structure of the VELO is illustrated in Fig. 2.5(a). There are altogether 42 pairs of R-ϕ sensors along the beam line. Their set up is driven by the constraint that charged tracks within the foreseen angular acceptance of $15 - 300\,\mathrm{mrad}$ must pass at least three pairs of R-ϕ modules for a decent track reconstruction. The design of the two sensor types is shown in Fig. 2.5(b). Each sensor is 300 µm thick and consists of a layer of strips for detection and a layer of routing lines for the read out. The VELO is divided into two movable halves that are retracted during proton injection to protect the sensors from the increased beam aperture. Once there are stable beams the VELO is closed and the sensitive area starts 8.2 mm away from the nominal beam axis to achieve an optimal vertex resolution and angular coverage. Figure 2.5(c) sketches the closed VELO around the beam axis in the (x, z) plane at $y = 0$. In addition to the 42 R-ϕ modules there are four R-sensors upstream of the VELO that work as pile-up veto system for the trigger. The average amount of material seen by tracks passing the VELO corresponds to $0.22\,X_0$. See also Ref. [27] for more information about the design of the VELO.

The overall performance of the VELO during Run I is described in detail in Ref. [28]. Relevant to physics analyses are the resolution of the decay time, the PV position and the impact parameter (IP) of a track, measuring the minimal distance of the track's trajectory to the set of PVs in an event. The PV resolution depends on the number of tracks that are used for reconstruction. For the required minimum of five tracks the resolution in the x and y coordinate is 35 µm, respectively, going down to 13 µm for a PV reconstructed from 25 tracks. The IP resolution of a track depends linearly on $1/p_{\mathrm{T}}$ and ranges from 80 µm for low-p_{T} tracks to a resolution better than 35 µm for tracks with a transversal momentum larger than 1 GeV/c. A decay time resolution of about $50 f_s$ is achieved for the four-pronged decay $B_s^0 \to J/\psi\,(\mu^+\mu^-)\phi(K^+K^-)$. The VELO has shown to withstand the massive radiation level during Run I without reducing the hit efficiency of typically 98 %.



(a) (b)

(c)

Figure 2.5: Illustrations of the VELO vessel (a), R-ϕ sensors (b), and (x, z) plane of the cross section at $y = 0$ [28].

TT

The Tracker Turicensis (Turicensis is derived from the earliest written roman record of the city of Zurich where the detector has been developed: “Statio Turicensis Quadragesima Galliarum”) is a silicon microstrip detector that is located behind RICH1 and before the magnet. It consists of four planar stations. Each station has four sensitive layers in an $(x - u - v - x)$ arrangement, where x are vertical layers and the u and v layers are tilted by $-5°$ and $+5°$ for stereo-angle detection. The total active area of the TT is $8.4\,\mathrm{m}^2$ covering the complete LHCb acceptance region. The detector is vital for the reconstruction of Downstream tracks originating from the long-lived states K_{S}^0 and Λ that decay mostly outside of the VELO. In addition the TT can add information to the reconstruction of Long tracks to improve the momentum measurement. In practice the TT fulfills the anticipated spatial resolution of 50 µm and has an overall hit efficiency of 99.7 %.

Magnet

LHCb uses a warm dipole magnet that is cooled by water and provides a bending power $\int B\,dl$ of 4 Tm for tracks of 10 m length. The coils are made of aluminium and achieve a maximum field strength of one Tesla in the center and dropping down to less than 50 mT at the RICH detectors. The field map was scanned with a spatial precisions of 1 mm and a relative uncertainty on the magnitude of about 10^{-4}. More details about the LHCb magnet can also be found in Ref. [29].

Tracking stations

There are three identical T stations that each consist of an Inner Tracker (IT) around the beam pipe and an Outer Tracker (OT) covering the rest of the acceptance region (see Fig. 2.6(a)). This design choice was made due to the fact that the inner region has much larger track occupancies demanding for radiation hard silicon technology. In particular the IT uses the same silicon microstrip technology and stereo-layer geometry as the TT. Each IT station has four individual boxes that partially overlap and cover an active area of $4\,\mathrm{m}^2$.

The OT is a drift-time detector made of large arrays of gas-tight tubes with a diameter of 4.9 mm. Each tube has a single anode wire (see Fig. 2.6(b)) and is filled with a mixture of Argon (70 %) and CO_2 (30 %) resulting in a maximum drift time of 36 ns. During operation the wires are supplied with +1550V, corresponding to a gas gain of about $5 \cdot 10^4$. The tubes are packed in two staggered layers making up one module (see Fig. 2.6(c)). Each station consists of four modules in an $(x - u - v - x)$ arrangement. Altogether there are about 55 000 single straw-tube channels. The performance of the OT is described in detail in Ref. [30]. During Run I the average detection efficiency in the inner part until 1.5 mm of the tube radius is 99.2 % and falls down to 95 % at $r = 2.0$ mm, and to 50 % at the edges. No degradation of the hit efficiency due to radiation damage was observed.

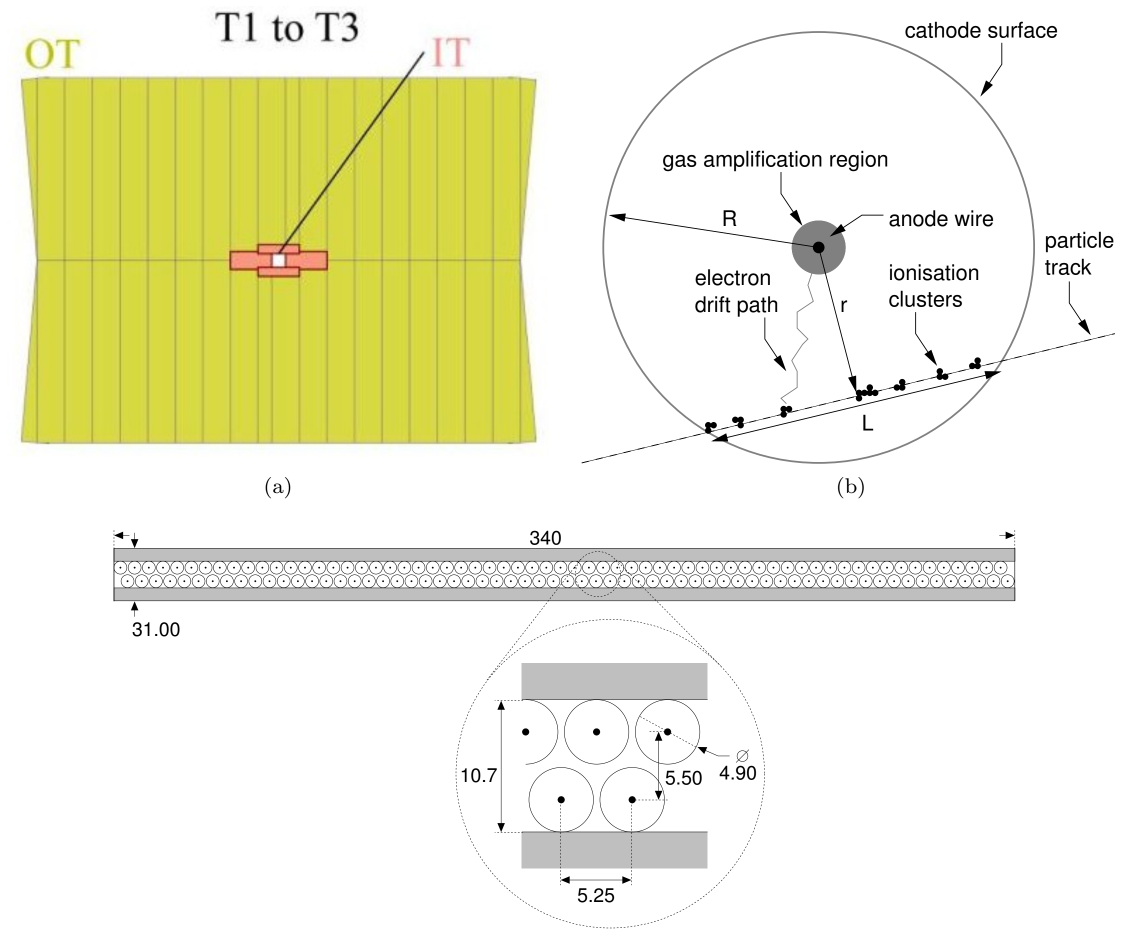
Reconstruction of Long tracks

The reconstructed decays in this analysis use exclusively Long tracks. At LHCb there are two complementary algorithms to find such tracks: Forward tracking and Track matching. Forward tracking combines VELO tracks with a single hit in one T station and searches for additional hits in the other T stations along the extrapolated trajectory. The best possible combination of hits defines the Long track. Track matching combines VELO and T tracks. Hereby the T tracks are reconstructed via an independent algorithm requiring them to have at least one hit in the x layers and one in the stereo layers of each T station. If there are two or more tracks found that have many hits in common (clones), only the one with the most hits is retained.

Once a Long track has been detected it is searched for additional hits in the TT along the trajectory to improve the reconstruction quality. Finally the tracks are refitted with a Kalman filter to account for multiple scattering and $\mathrm{d}E/\mathrm{d}x$ energy loss correction. See also Refs. [32–35] for more information about track reconstruction at LHCb. Figure 2.7(a) illustrates a typical LHCb event seen by the tracking system. In summary, the momentum resolution of Long track ranges from $\delta p/p = 0.5\,\%$ for particles below $p = 20\,\mathrm{GeV}/c$ to 1.0 % at $p = 200\,\mathrm{GeV}/c$ (see Fig. 2.7(b)). The average track finding efficiency is about 96 %.

2.2.4 Particle identification

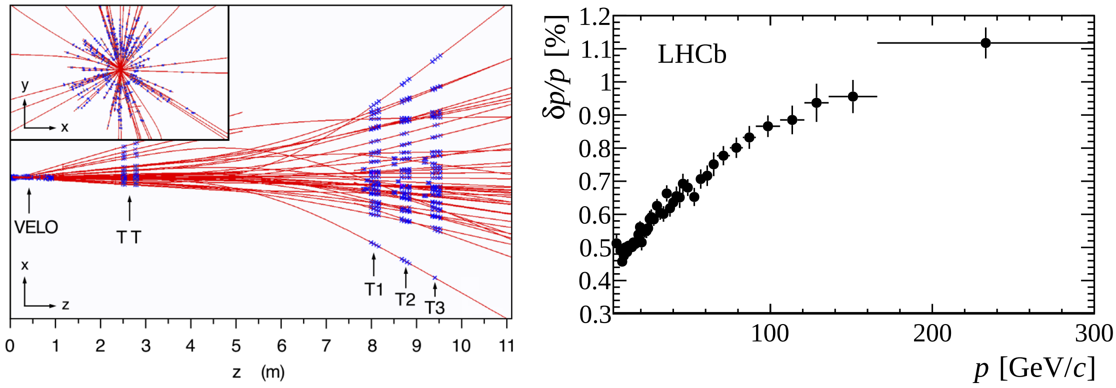
A multilayered particle identification (PID) system enables LHCb to separate different final states of b-hadron decays and to reduce combinatorial background due to the large number of pion tracks. As an example, a typical LHCb event at $\sqrt{s} = 7\,\mathrm{TeV}$ contains more than 100 charged tracks whereby pions play the dominating role — the hadronic track composition $N(\pi^{\pm}) : N(K^{\pm}) : N(p/\overline{p})$ is about 77 : 16 : 7 in the region $p_{\mathrm{T}} < 0.8\,\mathrm{GeV}/c$ [36]. The PID system comprises two Ring Imaging Cerenkov (RICH) detectors, an electronic and a hadronic calorimeter (ECAL, HCAL) and a muon system.



(a) (b)

(c)

Figure 2.6: Design of a T station (a) [31], working principle of a drift tube (b), and cross section of one monolayer (c) [30].

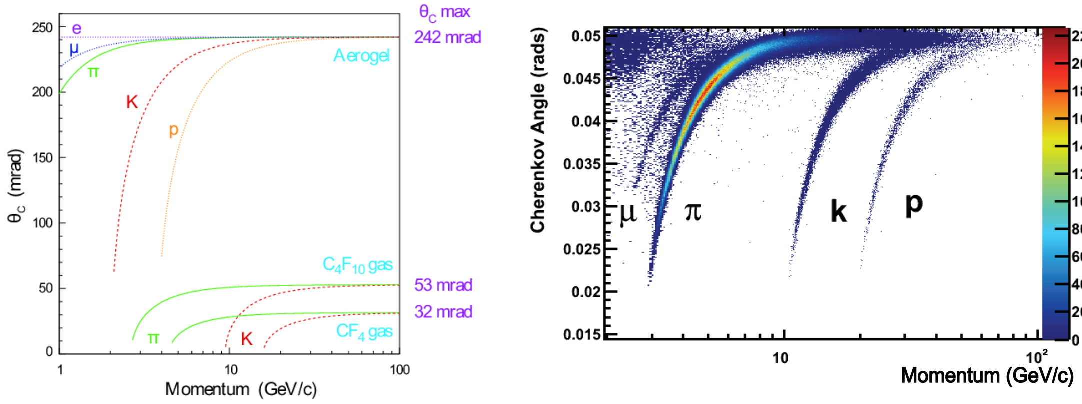


(a) (b)

Figure 2.7: Illustration of a typical LHCb event seen by the tracking stations (a) and the relative momentum resolution of Long tracks from $J/\psi \to \mu^+\mu^-$ decays in data (b) [23].

RICH detectors

Both RICH detectors (RICH 1,2) are mainly dedicated to the distinction of charged hadron tracks and their design is adapted to the characteristics of such tracks. The first one is located behind the VELO and measures tracks at large polar angles, *i.e.* ±25 to ±300(250) mrad in the bending (non-bending) plane, that have a soft momentum spectrum in the range of $1-60$ GeV/c. RICH 2 is placed behind the T stations and identifies tracks at small polar angles, *i.e.* ±15 to ±120(100) mrad in the bending (non-bending) plane, with a momentum of $15-100$ GeV/c. Both detectors utilize planar and spherical mirrors to redirect and focus Cerenkov light in the spectral range of $200-600$ nm to an array of pixel hybrid photo detectors. RICH 1 uses two separate radiators made of Aerogel and C_4F_{10} (flourobutane) gas while RICH 2 employs CF_4 gas. The refraction index of these materials for light with $\lambda = 400$ nm is 1.03, 1.0014 and 1.0005, respectively. Figure 2.8(a) compares the anticipated Cerenkov angle (θ_c) versus particle momentum for the different radiators and (b) for real data from the C_4F_{10} radiator. Details about the performance of the RICH system during Run I can also be found in Ref. [37]. In summary both detectors have met the anticipated demands in terms of the Cerenkov angle resolution and the number of produced photo electrons.



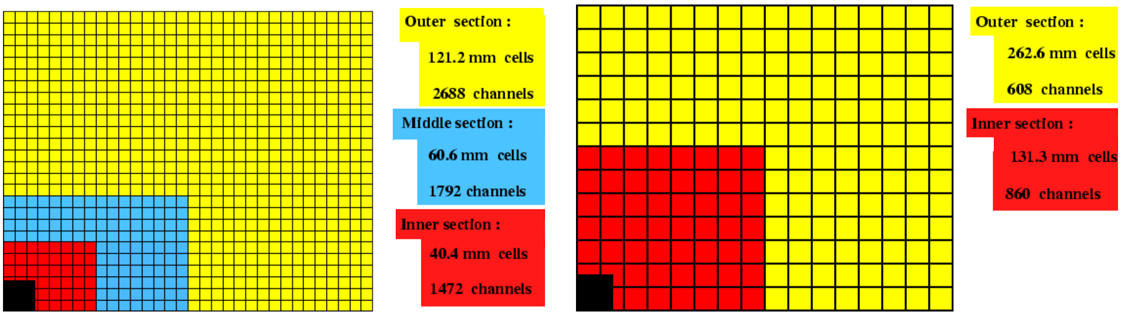
(a) (b)

Figure 2.8: Cerenkov angle versus particle momentum for different radiators in theory (a) [22] and for real data from the C_4F_{10} radiator (b) [37].

Calorimeter system

The main tasks of the calorimeter system are the identification of electrons and the measurement of the energy and position of hadrons, electrons and photons. These data are also relevant for the hardware trigger that needs to decide after 4 µs if an event will be recorded or not. In front of the aforementioned ECAL and HCAL there are two small subdetectors complementing the calorimeter system: the scintillating pad detector (SPD) and the preshower detector (PS). The SPD differs between electrons and photons while the PS lies behind a lead absorber and enables the separation of electrons and pions before entering the ECAL. Both calorimeters are of the shashlik type and cover the full angular acceptance region except for the inner part of $15-25$ mrad due to the high radiation level here. The ECAL uses altogether 66 layers of 2 mm lead absorber and 4 mm scintillator material corresponding to a total length of 25 X_0 and 1.1 λ_{int}. The HCAL uses altogether about 500 tons of iron as absorber whose length is equivalent to 5.6 λ_{int}. Figure 2.9 shows the lateral segmentation of ECAL and HCAL into cells and their combination into sections. The geometric design of the sections is projective, *i.e.* each section of each calorimeter subsystem covers the same angular acceptance region. For the detection of the scintillation light the calorimeters utilize wavelength-shifting fibres and photo multiplier tubes. The relative energy resolution $\sigma(E)/E$ of the ECAL was investigated with electron test beams and is smaller than 3 %

in the measuring range of $E > 10\,\text{GeV}$. In case of the HCAL the relative energy resolution was measured to be about 23 % for pions with $E = 10\,\text{GeV}$ and improving to 12 % at $E = 80\,\text{GeV}$. Lateral scans show that the detection of electrons and hadrons is sufficiently uniform across the planes. More information about the calorimeter performance can also be found in Ref. [38].



(a) ECAL (b) HCAL

Figure 2.9: Lateral segmentation of the ECAL (a) and HCAL (b). One quarter of the detector front is shown. The cells dimensions are given for ECAL and reduce by about 1.5 % for SPD and PS [22].

Muon system

A decent muon identification is vital for the operation of the detector in terms of calibration, triggering, flavour tagging and a high sensitivity on rare decays including muons in the final state. LHCb has a muon system made of five stations (M1-5) where each station consists of four concentric regions (R1-4) (see Fig. 2.10). The area of the regions increases with growing distance from the beam pipe so that an approximately uniform muon hit rate per region is achieved. Starting behind M2 there are muon filters made of iron with a total length equivalent to $20\,\lambda_{int}$ that filter hadron tracks penetrating the calorimeter. To pass all stations a muon requires a minimal momentum of $6\,\text{GeV}/c$. The muon system applies two technologies for hit detection: the inner region of M1 (M1R1) uses triple GEM (gas election multiplier) detectors to cope with the high hit density while the rest of the stations utilise multi-wire proportional chambers (MWPC). All regions are segmented along the x and y axis to obtain a spatial hit resolution.

The detection of muons requires a five-fold coincidence in all stations in a time window of 25 ns. Hereby stations M1-3 provide precise track finding and the measurement of the transversal momentum with a relative precision of 20 % using a fast standalone track reconstruction. The stations M4-5 are mainly for the validation of the muon tracks behind the iron absorbers. During Run I the overall detection efficiency was larger than 95 %. More details about the performance of the muon system during Run I can also be found in Ref. [39].

Performance of the particle identification

For the offline data analysis information from all PID subdetectors are summarised in a set of combined likelihood ($\mathcal{L}$) ratios $\text{DLL}(\text{x}-\text{y}) = \sum_{\text{PID}} \ln(\mathcal{L}_\text{x}/\mathcal{L}_\text{y})$ with $x, y \in \{e, \mu, \pi, K, p\}$ testing the mass hypothesis of x against y. Figure 2.11 illustrates the performance of the DLL variables for hadron tracks in different samples from real data. Pure samples of muons, pions, kaons and protons are obtained by the reconstruction of $J/\psi \to \mu^+\mu^-$, $D^{*+} \to D^0(K^-\pi^+)\pi^+$, and $\Lambda \to p\pi$, respectively, including kinematic selections to suppress background. In summary an adequate track identification is possible in the momentum range of $5 \ldots 100\,\text{GeV}/c$.

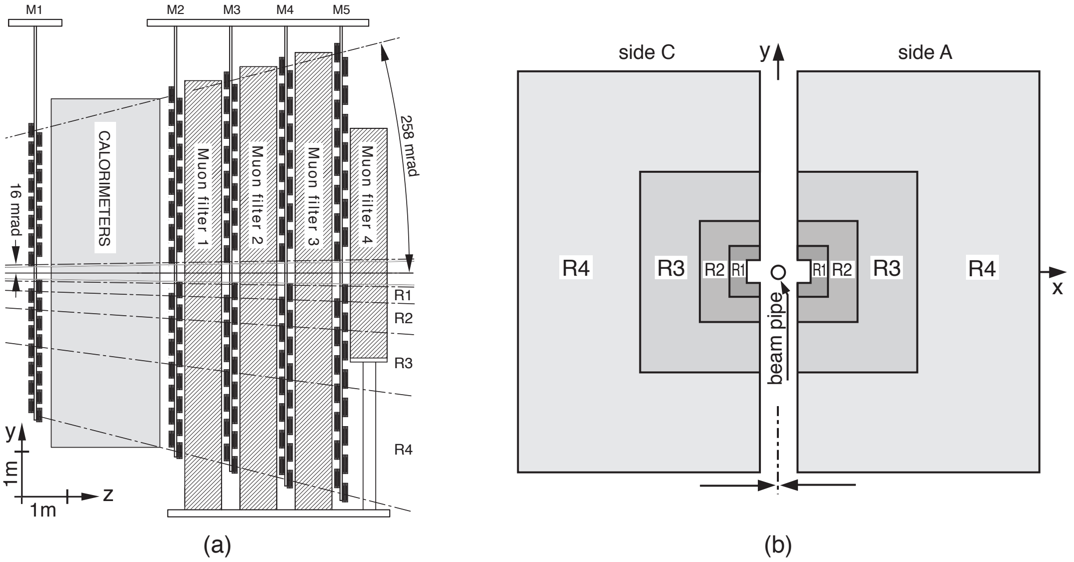
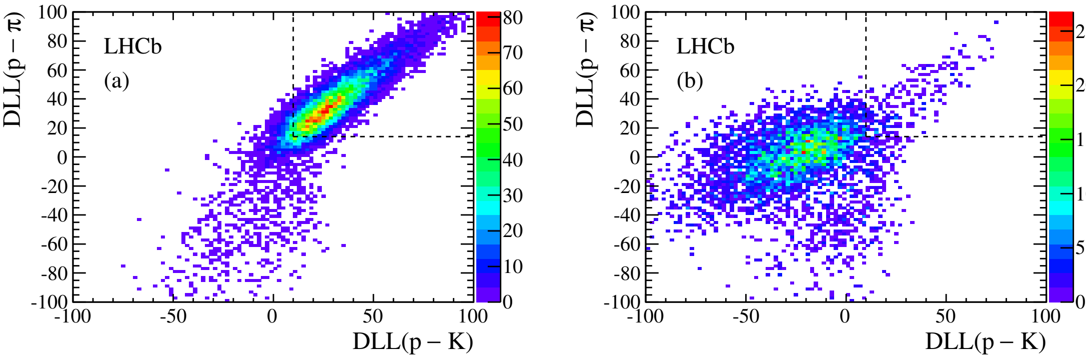
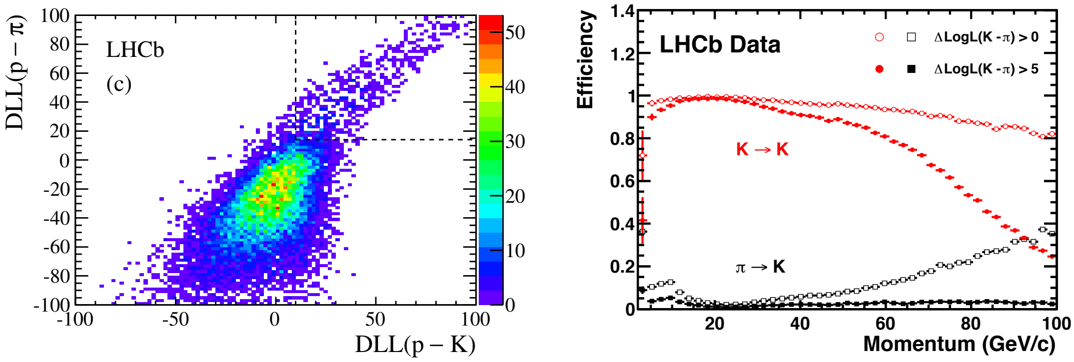


Figure 2.10: Design of the Muon stations in the (x, z) plane (a) and the (x, y) plane (b) [39].



(a) Proton sample

(b) Kaon sample

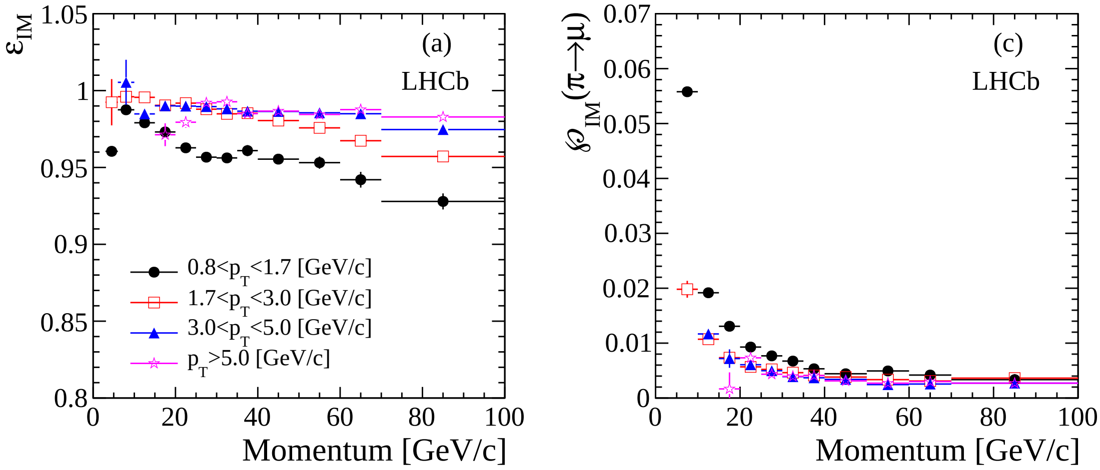


(c) Pion sample

(d) $K \to K$ and $\pi \to K$ efficiencies

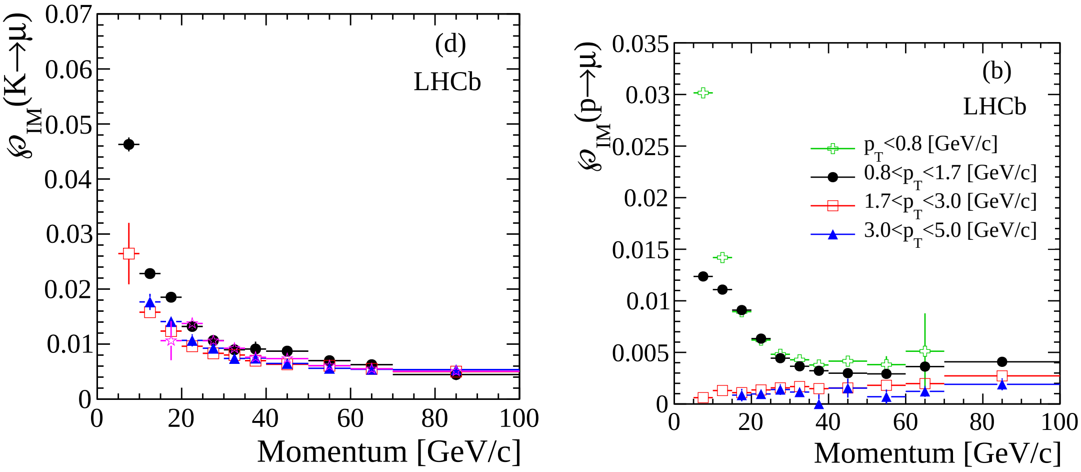
Figure 2.11: Performance of the DLL variables [23, 36].

The binary selector IsMuon={true,false} is commonly used at LHCb for the identification of muons. A positive decision requires hits in the muon stations within fields of interest (FOI) along the extrapolated track. The value of FOI is defined in dependence of the momentum p as $\mathrm{FOI} = a + b \cdot \exp(-c \cdot p)$ whereby the parameters a, b, c were determined from simulation studies and consider the momentum-dependent multiple scattering in the material. Figure 2.12 shows the performance of the IsMuon selector for different data samples. In conclusion a good separation of hadrons and muons is achieved for tracks with a momentum of $p > 10\,\mathrm{GeV}/c$. More information about the performance of the muon identification at LHCb can also be found in Ref. [40].



(a) Muon sample

(b) Pion sample

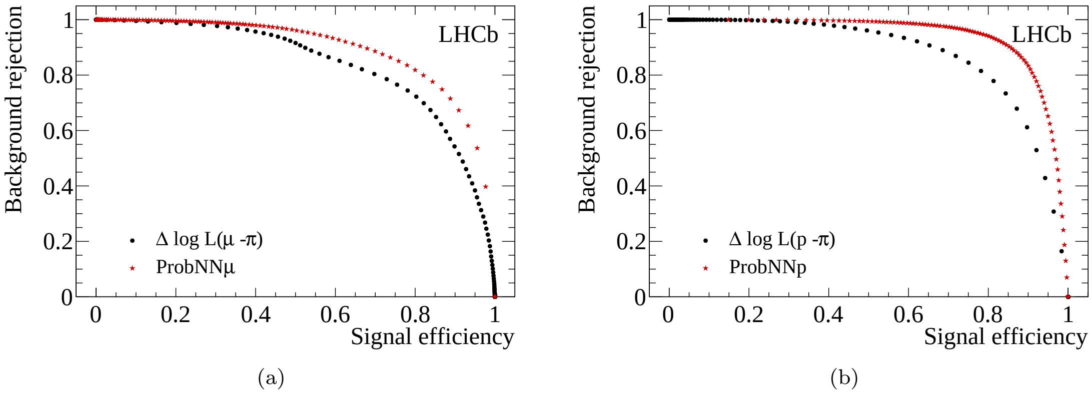


(c) Kaon sample

(d) Proton sample

Figure 2.12: Performance of the binary IsMuon (IM) selector [40].

In addition to the DLL variables there is a second class of observables that is based on a neural network classifier involving information from the PID system and track reconstruction. These ProbNN variables show a better performance than the likelihoods and can be used to optimise the final event selection. For illustration Fig. 2.13 compares the performance of ProbNN to DLL variables. The underlying data sets are $\Sigma^+ \to p\mu^+\mu^-$ events from signal MC and data sidebands including the selections $p(\mu) = 5\ldots10\,\mathrm{GeV}/c$ and $p(p) = 5\ldots50\,\mathrm{GeV}/c$.



(a) (b)

Figure 2.13: Comparison of the variables ProbNN to DLL for muons and protons [23].

In summary, the LHCb PID system is able to identify tracks well over a large momentum range. However, comparisons of real data to MC events have shown discrepancies in the simulation of the particle identification demanding for data-driven methods to measure selection efficiencies.

2.2.5 Online event selection (trigger)

At LHCb the visible interaction rate is about 10 MHz. Hereby a visible interaction is defined as an event where at least two charged particles exist that leave a sufficient number of hits in the tracking system to be reconstructible. A three-level trigger system at LHCb reduces this rate to several kilohertz (2011: about 3.5 kHz, 2012: about 5 kHz) to save events containing most likely b and c hadrons to disks for offline data analysis. The three triggers are the hardware based L0 trigger and two software based high-level triggers (HLT1 and HLT2) that are briefly explained below. To be able to adapt on changing beam conditions and to allow special data taking runs the behaviour of all triggers is defined in unique trigger configuration keys (TCKs). Only one TCK can be run at the same time.

L0 trigger

The L0 trigger reduces the event rate to 1 MHz given by the maximum read-out frequency of the detector. It uses custom made electronics that run synchronised with the LHC beam clock and at a fixed latency of 4 µs. During this time information from the pile-up stations, the calorimeter and the muon stations are processed and a trigger decision is made that is transferred to the Readout Supervisor (RS). The RS also considers random and unbiased trigger decisions collecting data dedicated to calibration measurements (*e.g.* integrated luminosity). Finally the RS decides if the event is accepted or not and calls for detector read out. The bandwidth of 1 MHz during Run1 is divided into 400 kHz for events with high-p_{T} muons, 500 kHz for events with high-E_{T} hadrons and 150 kHz for events with high-E electrons or photons, including overlaps of the three independent systems. The pile-up trigger may veto events with multiple interactions based on measurements from the two R-modules in the upstream region of the VELO. In addition the calorimeter system may also reject overstuffed events in which the track multiplicity measured by the SPD is too high or the total energy deposited in the ECAL and HCAL is too large. Technically, the L0 trigger consists of several hardware-implemented channels searching for a particular signature in one of the subdetectors. For this analysis the two channels `L0Hadron` and `L0Muon` are relevant that are explained in the section about the event selection.

HLT

The HLT is a software application that runs in 29 500 independent instances on a computer farm (Event Filter Farm - EFF). In the first stage (HLT1) a fast event reconstruction is done that searches for primary vertices and tracks using the Forward tracking algorithm and a simplified detector model. By imposing kinematic (p, p_{T}) and geometric (IP) selections on the tracks, events involving hadron and muon tracks originating from displaced vertices are selected. In addition Global Event Cuts (GEC) are made to reject events with too many hits in the tracking system that would otherwise spoil the processing speed. The HLT1 reduces the input event rate of 1 MHz to 80 kHz that is passed to the HLT2.

During the HLT2 stage a full event reconstruction is made that is close to the one of offline analyses and that is used by HLT2 trigger lines. A major step is the application of the Forward tracking algorithm on all VELO tracks including cuts on p and p_{T} to reduce the processing time. Additionally the HLT2 can make use of the IsMuon selector to treat muons separately. In comparison to the offline event reconstruction the overall track finding efficiency in the HLT2 is only $1-2\,\%$ lower.

In more detail each HLT stage consists of a large number of independent algorithms called "trigger lines" that search for particular event and decay signatures such as two-pronged and displaced vertices from b hadron decays. A special feature of 2012 data taking was a deferred trigger that saved 20 % of the L0-accepted events to local disks. These events were processed later by the EFF during periods without stable beams which make up about 70 % of the year. As a consequence the available processing time per event for the HLT has increased allowing lower thresholds and more data to be recorded.

In summary of the Run I data taking the LHCb trigger system has shown a steady and high performance for a variety of decay channels. For example the trigger efficiency on the two-body decay $B^0 \to K^-\pi^+$ in 2012 is 53 % in the L0 trigger, 97 % in the HLT1, and 80 % in the HLT2. More information about the LHCb trigger performance and configuration can also be found in Refs. [41,42]

2.3 Λ_b production at LHCb

Heavy quarks are produced in pp collisions by the scattering of the proton constituents (partons) via the strong interaction. These partons are gluons g and mostly light quarks $q = \{u, d\}$ and anti quarks $\bar{q} = \{\bar{u}, \bar{d}\}$. The dominant processes that contribute to the $b\bar{b}$ pair production at LHCb can be described by pair production $gg \to b\bar{b}$ and $q\bar{q} \to b\bar{b}$ in the order of magnitude $\mathcal{O}(\alpha_S^2)$ (leading order), flavour excitation $gg \to b\bar{b}g$ and $gq \to b\bar{b}q$, and by gluon splitting $gg \to gg^*, g^* \to b\bar{b}$ which are all in the order of magnitude $\mathcal{O}(\alpha_S^3)$ [43]. Beauty quarks form b hadrons (H_b) due to the strong interaction and finally decay after $1-2$ ps due to the weak interaction. These b hadrons are mainly B^0, B^+, B_s^0 and Λ_b. One of the first LHCb results was the measurement of the inclusive production cross section of such hadrons within detector acceptance to be $\sigma(pp \to H_b\,X) \approx 75\,\mu\mathrm{b}$ at $\sqrt{s} = 7\,\mathrm{TeV}$ [25]. As a result there are in the order of 10^{11} b hadrons created per $1\,\mathrm{fb}^{-1}$ at LHCb.

There are three LHCb results using Run I data that allow to estimate the number of produced Λ_b baryons:

- the measurement of the relative fragmentation rate of b quarks into Λ_b and B^0 [44]

$$f_{\Lambda_b}/f_d \approx 0.40$$

- the inclusive measurement of the B^0 production cross section [45]

$$\sigma\left[pp \to B^0X\ (p_{\mathrm{T}} = 0\ldots40\,\mathrm{GeV}/c, y = 2.0\ldots4.5)\right] = (38.1 \pm 6.0)\,\mu\mathrm{b}\ ,$$

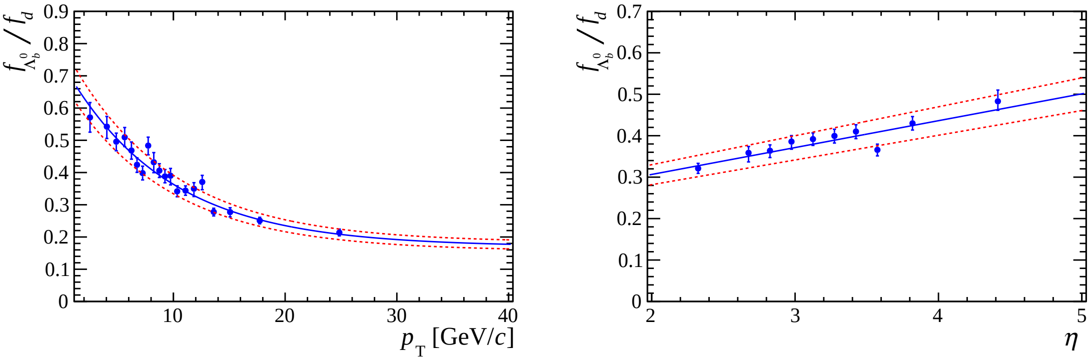
- the determination of the relative Λ_b production cross section at 7 and 8 TeV [46]

$$\sigma(8\,\text{TeV})/\sigma(7\,\text{TeV}) \approx 1.2\ .$$

Therefore the estimated number of created Λ_b baryons within detector acceptance during Run I is $N(\Lambda_b) \approx (38\,\mu\text{b} \times 1\,\text{fb}^{-1}) \times 0.4 \times (1 \cdot 1.0 + 2 \cdot 1.2) = 50 \cdot 10^9$. As a consequence the anticipated sensitivity on $\mathcal{B}(\Lambda_b \to K^-\mu^+)$ is in the order of 10^{-9} assuming an overall reconstruction efficiency of typically a few per cent.

2.3.1 Properties of Λ_b baryons produced at LHCb

An unresolved feature of the Λ_b production w.r.t. B^0 mesons is the strong dependence from the transversal momentum p_T while the variation in dependence of the pseudorapidity η is small, see Fig. 2.14 (a) and (b).

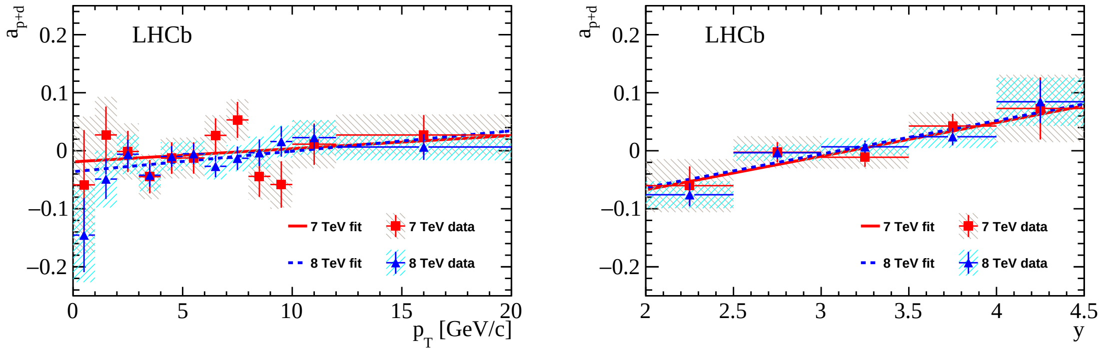


(a) $f_{\Lambda_b}/f_d\,(p_\text{T})$.

(b) $f_{\Lambda_b}/f_d\,(\eta)$.

Figure 2.14: Λ_b production w.r.t. B^0 in $\sqrt{s} = 7$ TeV data [44].

The asymmetry of the $\Lambda_b/\overline{\Lambda}_b$ production was investigated at LHCb using Run I data [46]. Figure 2.15 shows the combined asymmetry of production and decay ($a_\text{p+d}$) in dependence of p_T (a) and the rapidity (b), whereby $a_\text{p+d} > 0$ corresponds to an increased Λ_b production w.r.t. $\overline{\Lambda}_b$. The fitted slope of the p_T distribution is consistent with zero while the slope of the y distribution is 0.058 ± 0.014 in the combined data of 2011 and 2012, respectively. This could be due to a increased Λ_b production w.r.t. $\overline{\Lambda}_b$ at higher rapidities. There are theoretical ideas to explain an asymmetry, see for example Refs. [47, 48]. No transversal polarisation of the Λ_b is observed [49].



(a) $f_{\Lambda_b}/f_d(p_\text{T})$ for $\sqrt{s} = 7$ TeV data [44].

(b) Asymmetry of the $\Lambda_b/\overline{\Lambda}_b$ production [46].

Figure 2.15: Properties of the Λ_b production at LHCb.

"The good ones go into the pot,
the bad ones go into your crop."
— Cinderella (Grimms' Fairy Tales)

3

Event selection

3.1 Analysis strategy

The primary goal of this analysis is the search for the decay $\Lambda_b \to K^- \mu^+$. At the same time the selected events also allow the search for the decay of the about 170 MeV/c^2 heavier Ξ_b^0 baryon into the same final state. The used data consist of 1 fb^{-1} from 2011 and 2 fb^{-1} from 2012. Due to the fact that the 2012 data make up the major share all studies are made here in the first place. To be unbiased a blind analysis is conducted that uses simulated signal events, $K^-\mu^+$ combinations from the mass sidebands, and wrong-charge combinations of $K^+\mu^+$ to investigate the mainly combinatorial background and optimise the signal selection. Another important step is the validation of the LHCb event- and detector simulation by comparing real and simulated data from similar decay channels such as $B^0 \to K^+\pi^-$ and $\Lambda_b \to \Lambda_c \pi^-$. After the determination of the signal yield the branching fraction is obtained with the help of the normalisation channel $\Lambda_b \to pK^-$. The following outline summarises the milestones of the analysis.

1. Event selection
    (a) Description of the selection of $K\mu$ combinations from LHCb data
    (b) Identification of efficient trigger lines
    (c) Study of background contributions from other b hadron decays
2. Control channel studies
    (a) Study of differences between simulation and real data
3. Signal selection
    (a) Development of a multivariate event classifier
    (b) Optimisation of the signal selection
    (c) Unblinding and determination of the signal yield
4. Determination of the number of $\Lambda_b \to pK^-$ events and the selection efficiencies
5. Calculation of the branching ratio of $\Lambda_b \to K^- \mu^+$

3.2 Characteristics of b hadron decays at LHCb

The first step in the analysis is the preselection of $K\mu$ pairs from reconstructed events in the detector. To suppress the large amount of random combinations arising from the high track multiplicity in a pp collision, a number of selection criteria are applied. These criteria make explicit use of the characteristics of b hadron decays at LHCb as well as properties of the track reconstruction and identification and are described below.

At LHCb b hadrons are produced with a great boost of $\beta\gamma \approx 10\ldots40$. Due to their large mean lifetime of around 1.5 ps these particles travel on average a few mm before they decay within the VELO. Thanks to the high spatial resolution of $< 100\,\mu\text{m}$ the VELO allows a good separation of secondary tracks coming from a displaced b hadron decay vertex and prompt tracks originating from primary vertices. Figure 3.1 sketches the topology of a reconstructed $\Lambda_b \to K^-\mu^+$ event at LHCb and illustrates a number of observables that are used for selection. The most important quantity is the impact parameter (IP). It measures the minimal distance between the projected trajectory of a particle and any primary vertex in an event and is expected to be zero for particles coming from a PV and large for secondary tracks. For a reconstructed Λ_b the best PV (BPV) is chosen to be the one with the smallest IP. Other important observables are the flight distance ($FD = |\vec{F}|$), the direction angle (ϕ_{Dira}) and the proper lifetime (τ). The flight distance, as indicated in Fig. 3.1, is the vector between the BPV = PV(2) and the secondary vertex (SV) of the decay. For random combinations of tracks, the length of the flight distance is typically close to zero. ϕ_{Dira} is the angle between $\vec{F}$ and the momentum vector $\vec{p}$ of the reconstructed particle and zero for correctly reconstructed b hadrons. The proper lifetime can be calculated from FD, the momentum p and the reconstructed invariant mass m of the b hadron according to $\tau = FD \cdot m/p$.

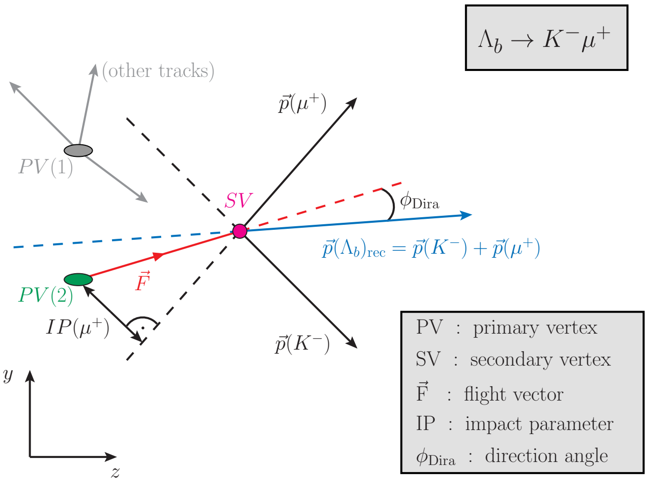


Figure 3.1: Illustration of some LHCb observables for a $\Lambda_b \to K^-\mu^+$ event.

3.3 Selection of $\Lambda_b \to K\mu$ events

For the reconstruction and filtering of $K\mu$ combinations a selection algorithm has been developed that is similar to the one of the very rare $B_s^0 \to \mu^+\mu^-$ decay that was found at LHCb with $\mathcal{B}(B_s^0 \to \mu^+\mu^-) \approx 3 \times 10^{-9}$ [18]. The preselection begins with the combination of tracks that are required to satisfy a number of criteria. One track uses kaon mass hypothesis and the other one muon mass hypothesis. All charge combination are considered, *i.e.* $K^-\mu^+ + c.c.$ and $K^+\mu^+ + c.c.$, whereby the first ones are right-sign (RS) combinations for the signal search and the second ones are wrong-sign (WS) combinations for background studies. The chosen tracks are then fitted to a common vertex and their four-momenta are added to form a Λ_b candidate on which additional selections are imposed to be compatible with originating from a Λ_b decay.

3.3.1 Track selection criteria

Both tracks are required to satisfy the following selection criteria.

Impact parameter significance: $\chi^2(\mathrm{IP}) > 25$

For a given track and its BPV the quality of the primary vertex fit χ^2_{Fit} is measured with and without the track. The difference $\chi^2(\mathrm{IP}) = \Delta\chi^2_{\mathrm{Fit}}$ is large for tracks coming from displaced vertices and small for prompt tracks.

Track fit quality: $\chi^2_{/\mathrm{DOF}}(\mathrm{Tr}) < 3$

The reduced $\chi^2_{/\mathrm{DOF}}(\mathrm{Tr}) = \chi^2/DOF$ describes the goodness of the track fit to the hits in the tracking system.

Ghost rate: $\mathrm{P(Gh)} < 0.3$

Ghost tracks are fake combinations of hits. In simulated events a track is defined as a ghost when it uses less than 70 % of the hits that belong to one generated track. On the basis of a likelihood method the classifier $\mathrm{P(Gh)} = 0 \dots 1$ is able to discriminate ghost tracks in addition to $\chi^2_{/\mathrm{DOF}}(\mathrm{Tr})$ by combining information of the track fit with the energy deposit of track hits in the silicon detectors. This allows to reject fake tracks which use spillover hits from previous pp collisions that typically have lower energy deposits [50].

Transversal momentum: $p_{\mathrm{T}} > 250\,\mathrm{MeV}/c^2$

The transversal momentum of a track w.r.t. the beam axis is typically larger for particles from b decays compared to tracks coming from the underlying event.

To suppress background from the large number of pion tracks in an event, information from the PID systems are used as described in Sec. 2.2.4.

$K - \pi$ **discrimination:** $\mathrm{DLL(K} - \pi) > 0$

The kaons must be measured to be more kaon- than pion-like.

Hits in the Muon detector: IsMuon(μ)=true

All muons are required to satisfy the binary selector IsMuon.

3.3.2 Candidate selection criteria

After passing the track selection criteria, the kaon and muon combinations must satisfy $m(K,\mu) = (5620 \pm 500)\,\mathrm{MeV}/c^2$ and a distance of closest approach of their trajectories (DOCA) less than 0.3 mm. The pairs passing these criteria are fitted to a common vertex and form a Λ_b candidate if the fit is successful. To reject unphysical candidates a number of selections are applied: $\chi^2(\mathrm{IP}) < 25$, $\chi^2(\mathrm{FD}) = [\mathrm{FD}/\delta(\mathrm{FD})]^2 > 100$, $\tau > 0$ and $\chi^2_{/\mathrm{DOF}}(\mathrm{Vtx}) < 9$. The described event selections have been tested in a sample of simulated signal events and in a background test sample of LHCb events without preselection (minimum bias events). As a result the signal efficiency is about 30 % and the background acceptance is approximately $5 \cdot 10^{-5}$.

3.3.3 Selection of $B^0 \to K^+\pi^-$ and $\Lambda_b \to pK^-$ events

For the selection of the decays $\Lambda_b \to pK^-$ (normalisation channel) and $B^0 \to K^+\pi^-$ (control channel) the data of a generic $B \to \mu^+\mu^-$ reconstruction are used that was developed for the $B^0_s \to \mu\mu$ analysis in 2010 [18]. The algorithm combines oppositely charged tracks with muon mass hypothesis and without PID requirements in the invariant mass range $m(\mu^+\mu^-) = (5366 \pm 500)\,\mathrm{MeV}/c^2$. The selections applied on the mother particles are very similar to the $\Lambda_b \to K\mu$ case. The reconstruction efficiency in simulated signal events is about 20 %. Table 3.1 summarises and compares the two event selections.

Table 3.1: Selections for $\Lambda_b \to K\mu$ and $B \to \mu^+\mu^-$ events.

Property	$\Lambda_b \to K\mu$	$B \to \mu^+\mu^-$
PID selections		
Track 1	$\mathrm{DLL(K-\pi)} > 0$	—
Track 2	IsMuon=true	—
Track selections		
p_T	$> 250\,\mathrm{MeV}/c$	$> 250\,\mathrm{MeV}/c$
$\chi^2(\mathrm{IP})$	> 25	> 25
$\chi^2_{/\mathrm{DOF}}(\mathrm{Tr})$	< 3	< 3
P(Gh)	< 0.3	< 0.3
Combination selections		
Invariant mass	$(5620 \pm 500)\,\mathrm{MeV}/c^2$	$(5366 \pm 500)\,\mathrm{MeV}/c^2$
DOCA	$< 0.3\,\mathrm{mm}$	$< 0.3\,\mathrm{mm}$
Mother candidate selections		
$\chi^2_{/\mathrm{DOF}}(\mathrm{Vtx})$	< 9	< 9
$\chi^2(\mathrm{IP})$	< 25	< 25
$\chi^2(\mathrm{FD})$	> 100	> 225
τ	$> 0\,\mathrm{ps}$	$0 - 13.248\,\mathrm{ps}$
p_T	—	$> 350\,\mathrm{MeV}/c$

3.4 Studies on simulated events

The following studies are made using a sample of generated and reconstructed $\Lambda_b \to K^-\mu^+$ decays from a full Monte-Carlo (MC) simulation of pp collisions at LHCb. In the simulation, pp collisions are generated using PYTHIA [51, 52] with a specific LHCb configuration [53]. Decays of hadronic particles are described by EVTGEN [54], in which final-state radiation is generated using PHOTOS [55]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit [56, 57] as described in Ref. [58].

The events are reconstructed using the same preselections as in the data except for the PID selections due to the fact that these variables are not correctly simulated. The used sample comprises about 82 k correctly reconstructed events from 2012 MC.

3.4.1 Trigger selections

The use of trigger selections allows to significantly reduce the amount of background events while keeping most of the signal events. Moreover, the trigger efficiency can only be determined with sufficient precision from simulated events if well defined conditions on the kaons, muons and their combinations are applied. On the basis of simulated $\Lambda_b \to K^-\mu^+$ events, that include an emulation of the complete online event selection, Tab. 3.2 summarises the most efficient triggers that are found from correctly reconstructed decays after event selection and within detector acceptance in 2012 signal MC. Within each trigger level the lines are combined by a logical `OR` while the different levels are combined by a logical `AND` (&). Due to imperfections of the trigger emulation the stated efficiencies are only a preliminary estimate. The chosen set of triggers is sensitive to $\Lambda_b \to K\mu$ as well as the normalisation channel $\Lambda_b \to pK^-$ and explained below. In summary the total trigger efficiency is 71 % in the sample of $\Lambda_b \to K^-\mu^+$ events after the event selection and 84 % of reconstructed and selected $\Lambda_b \to K^-\mu^+$ events that are triggered anyhow at LHCb.

Table 3.2: Trigger efficiencies for 2012 signal MC of $\Lambda_b \to K^-\mu^+$.

Level	Precondition	Trigger line(s)	Efficiency
All	reconstructed	All physical triggers	84 %
L0	reconstructed	L0Muon-TOS	86 %
	and triggered	L0Hadron-TOS	30 %
	by LHCb = obs	L0-TIS	35 %
		L0_{sel} = combination	100 %
HLT1	obs & L0_{sel}	AllTrack-TOS	84 %
		MuonTrack-TOS	77 %
		Hlt1_{sel} = combination	92 %
HLT2	obs & L0_{sel} & Hlt1_{sel}	Muon2Body-TOS	91 %
		All2Body-TOS	81 %
		Hlt2_{sel} = combination	92 %
All	reconstructed	L0_{sel} & Hlt1_{sel} & Hlt2_{sel}	71 %

Trigger categories

Each positive trigger decision can be attributed to objects measured in the detector. These objects can be, for example, energy deposits in particular cells of the calorimeters or hits in certain straw tubes of the T stations. The trigger software records a unique identifier of all triggering objects and is able to associate them later on to signal tracks in the offline reconstruction. Depending on the association the signal candidate can be classified as "Triggered On Signal" (TOS), "Triggered Independent of Signal" (TIS), or "Triggered on Both" (TOB). If an event is TIS and TOS at the same time it is declared as TISTOS. The definitions of TOS, TIS and TOB according to [59] are:

TOS : These are events for which the presence of the signal is sufficient to generate a positive decision.

TIS : The "rest" of the event is sufficient to generate a positive trigger decision, where the rest of the event is defined through an operation procedure consisting in removing the signal and all detector hits belonging to it.

TOB : These are events that are neither TIS nor TOS, neither the presence of the signal nor the rest of the event alone are sufficient to generate a positive trigger decision, but rather both are necessary.

The category TOB is not relevant in this analysis. For a positive association 70 % of the online reconstructed trigger candidate hits need to be contained within the set of hits forming the offline reconstructed track.

L0Hadron

As shown in Fig. 2.9(b) the HCAL is laterally segmented into cells. The front-end electronics considers clusters of 2 × 2 cells and calculates the transversal energy as $E_\mathrm{T} = \sum E_i \cdot \sin(\theta_i)$. Hereby E_i is the deposited energy in each cell and θ_i is the angle between the beam axis and the direction between the center of the pp-interaction envelope and the center of the cell. The L0Hadron algorithm searches for the cluster with the highest E_T and makes a positive decision if $E_\mathrm{T} > 3.5(3.7)\,\mathrm{GeV}$ during 2011 (2012) data taking. In addition the number of SPD hits are required to be smaller than 600. This trigger category is mainly caused by the kaon from $\Lambda_b \to K^-\mu^+$.

L0Muon

In each quadrant of the muon system a dedicated L0 Muon processor determines the two muons with the largest p_T value that hit all stations. The L0Muon algorithm sets a threshold of $1.48\,\mathrm{GeV}/c$ in 2011 and $1.76\,\mathrm{GeV}/c$ in 2012 that has to be satisfied by at least one of the four highest-p_T muons for a positive trigger decision. L0Muon is especially sensitive to the muon from $\Lambda_b \to K^-\mu^+$.

HLT1 AllTrack and MuonTrack

The AllTrack trigger line, technically named `HLT1TrackAllL0`, is the "work horse" of the HLT1 exploiting 58 kHz of the total 80 kHz bandwidth. It uses Long tracks from the fast HLT1 track reconstruction as input that require the preselections $p > \{3, 6\}\,\mathrm{GeV}/c$ and $p_\mathrm{T} > 0.5\,\mathrm{GeV}/c$, where the lower momentum threshold was introduced in 2012. The AllTrack line selects good quality tracks from displaced vertices by requiring a minimum number of hits in the VELO and T stations, an adequate track fit and kinematic and geometric properties summarised in Table 3.3. MuonTrack, technically named `HLT1TrackMuon`, is a similar line that is dedicated to tracks that are matched to muon hits. Lower thresholds on the number of hits and kinematic observables are possible due to their smaller rate.

Table 3.3: HLT1 line selections in 2011 (2012) data.

Variable	AllTrack	MuonTrack
#(VELO hits)	> 9	—
#(missing VELO hits)	< 3	—
#(T stations hits)	> 16	—
p_T (GeV/c)	$> 1.7(1.6)$	> 1.0
p (GeV/c)	> 10	> 8.0
IP	$> 0.1\,\mathrm{mm}$	$> 0.1\,\mathrm{mm}$
$\chi^2_{/\mathrm{DOF}}(\mathrm{Tr})$	(< 2)	(< 2.5)
IsMuon	—	true

HLT2 topological two-body lines

The HLT2 trigger lines have access to fully reconstructed events. For this analysis the HLT2 triggering on signal events is based on two topological lines that both apply a specially developed boosted decision tree called “bonsai boosted decision tree” (BBDT) [60]. The working principle of a BDT is described in more detail in Sec. 5. Generally speaking it is an algorithm that uses a number of observables to classify an event as more signal or more background like. Before the BBDT is used, two-body decay vertices are created from reconstructed tracks that satisfy $\chi^2(\mathrm{IP}) > 4$, $\chi^2(\mathrm{Tr}) < 4$ for muons and $\chi^2(\mathrm{Tr}) < 3$ for other tracks. In addition the track combinations require $\mathrm{DOCA} < 0.2\,\mathrm{mm}$, $\chi^2(\mathrm{FD}) > 100$, $\tau > 0$ and $\sum |p_\mathrm{T}| > 3\,\mathrm{GeV}$ and are classified by the BBDT using seven variables: $\sum |p_\mathrm{T}|$, the minimum p_T of both tracks, DOCA, $\chi^2(\mathrm{FD})$, $\chi^2(\mathrm{IP})$ of the combination, the invariant mass and the corrected invariant mass $m_\mathrm{cor} = \sqrt{m^2 + |{p_\mathrm{T}}^\mathrm{miss}|}+|{p_\mathrm{T}}^\mathrm{miss}|$. $|{p_\mathrm{T}}^\mathrm{miss}|$ is defined as the missing momentum transversal to the flight direction defined by the best primary vertex and the decay vertex. A specific feature of the BBDT is the discretisation of the variables into a number of intervals with an optimised size. These intervals are larger than the respective resolution of the observable and ensure robustness against minor changes of the data taking conditions as well as universality about the characteristics of b hadron decays. On the other side they are still small enough to distinguish signal and background events. For the training of the BBDT a set of simulated B^+, B^0, B^0_s and Λ_b decays is used as signal sample and minimum bias events from data as background sample. As a consequence the BBDT is a fast and powerful classifier. It rejects more than 90 % of the events selected by the HLT1 but retains $70 - 90\,\%$ of b hadrons depending on the topology of the decay. For example $B^+ \to J/\psi\, K^+$ is more often accepted than $B^+ \to D^0\pi^+$. The technical names of the lines are `Hlt2TopoMu2BodyBBDTDecision` and `Hlt2Topo2BodyBBDTDecision` whereby the second one requires at least one of the two tracks to be identified as a muon. This allows lower thresholds due to the smaller number of candidates.

3.4.2 Mass resolution

In the sample of reconstructed and triggered $\Lambda_b \to K^-\mu^+$ events in 2012 MC the mass resolution is studied and the blinded mass window is defined. Figure 3.2 shows the $m(K^-\mu^+)$ spectrum and the result of a binned likelihood fit to the distribution. The signal component is modeled by the sum of two Crystal Ball functions, see App. A.6 and A.7, with a common mean value to describe the asymmetric shape due to missing energy from final-state radiation. In addition a first-order polynomial is used for the combinatorial background. As a result about 65 % of the signal events are within a mass window of $\pm 20\,\mathrm{MeV}/c^2$ around the fitted mean value. This corresponds to an effective resolution of $\sigma_\mathrm{eff} = 20\,\mathrm{MeV}/c^2$. The fitted mass of the Λ_b is $(5620.1 \pm 0.1)\,\mathrm{MeV}/c^2$ and slightly larger than the generated mass of $5619.4\,\mathrm{MeV}/c^2$. The reason for this is unclear.

Fits to the $m(K^-\mu^+)$ distribution of simulated $\Xi_b^0 \to K^-\mu^+$ events provide the same results in terms of the effective resolution. For the definition of the blinded region a mass window of $\pm 50\,\mathrm{MeV}/c^2$ around mean mass is chosen. This selection contains $\geq 93\,\%$ of reconstructed and triggered signal events and is consistent with other rare h_b decay searches at the LHCb experiment such as $B^0 \to p\overline{p}$. The center of the blinded mass region is determined from data to consider the shift of the absolute mass scale w.r.t. simulated events. According to the studies using the control channel $B^0 \to K^+\pi^-$, which is described in Ch. 4, the shift is approximately $+5\,\mathrm{MeV}/c^2$.

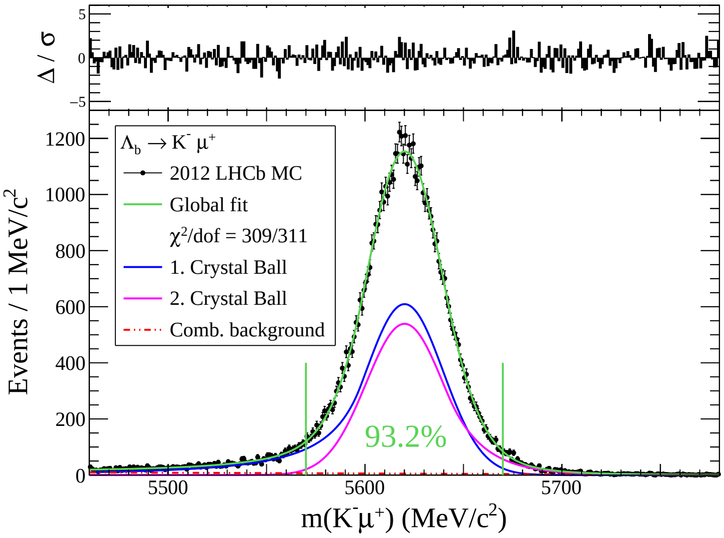


Figure 3.2: Fit to the $m(K^-\mu^+)$ distribution after trigger selections in 2012 MC.

3.5 Background studies

Apart from random track combinations, decays of b hadrons can make contributions to the invariant mass spectrum depending on their abundance in the data. Due to similar event properties w.r.t. $\Lambda_b \to K^-\mu^+$ many of these physical background sources cannot be sufficiently rejected during event selection and may possibly contaminate the signal region in the $m(K^-\mu^+)$ distribution. In the following the b hadron decays are classified into two groups and investigated using kinematic considerations as well as simulated events.

(1) Reflections: $h_b^0 \to h^-h^+$
Reflections are fully reconstructed b hadron decays into two charged hadrons, *e.g.* $B^0 \to K^+\pi^-$, that may manifest as narrow signal enhancement in the invariant mass spectrum. Due to the wrong mass hypothesis for at least one track the reconstructed mass is more or less shifted w.r.t. the true mass of the mother particle.

(2) Partially reconstructed decays: $h_b \to h^-\ell^+ + X$
This background class consists of semileptonic decays such as $B_s^0 \to K^-\mu^+\nu_\mu$ or $B_c^- \to J/\psi\,(\mu^+\mu^-)K^-$ where at least one daughter particle is not used for reconstruction. The invariant mass spectrum of such events is usually broad and falls down towards a distinct upper limit.

Partially reconstructed decays with more than one missing track are not considered due to the large amount of missing energy that prevents $K\mu$ combinations to fall into the selected mass window $m(K\mu) = m(\Lambda_b) \pm 500\,\mathrm{MeV}/c^2$.

3.5.1 Background from reflections

Kinematic considerations

The impact of reflections is estimated by determining the shifted mass M' in the cms system of the mother particle using correct momenta and wrong mass hypotheses (cms approximation of M'). The calculation neglects the boost $\beta\gamma$ of the particles in the detector frame and the helicity angle θ of the daughters. However, the approximation is precise to the level of a few ten MeV/c^2 depending mainly on the differences between the used mass hypotheses.

For a given two-body decay $M \to m_1 \, m_2$, where M is the mass of the mother and m_1 and m_2 of the daughters, respectively, the momenta and energies of the decay products in the cms of the mother are

$$p_1 = p_2 = p = \frac{\sqrt{\left(M^2 - (m_1 + m_2)^2\right)\left(M^2 - (m_1 - m_2)^2\right)}}{2M} \, ,$$

The invariant mass is $M = E_1 + E_2 = \sqrt{p^2 + m_1^2} + \sqrt{p^2 + m_2^2}$. For the reconstruction $\Lambda_b \to K\mu$ the masses $m(K)$ and $m(\mu)$ are used. Therefore the energy of the daughters change to $E(\to K) = \sqrt{p^2 + m_K^2}$ and $E(\to \mu) = \sqrt{p^2 + m_\mu^2}$. The shifted mass M' of the reconstructed candidate is $M' = E(\to K) + E(\to \mu)$. Table 3.4 shows M' for a number of two-body decays of b hadrons. All of these reflections are expected to have an invariant mass distribution that is centered below the Λ_b mass.

Table 3.4: Shifted mass for a number of reflections using the cms approximation.

Decay	$\approx \mathcal{B}$	p (MeV/c)	M' (MeV/c^2)
$B^0 \to \pi^+\pi^-$	5×10^{-6}	2636	5320
$B_s^0 \to \pi^+\pi^-$	8×10^{-7}	2680	5407
$B^0 \to K^+\pi^-$	2×10^{-5}	2615	5278
$B_s^0 \to K^-\pi^+$	6×10^{-6}	2659	5365
$B_s^0 \to K^+K^-$	3×10^{-5}	2637	5322
$\Lambda_b \to p\pi^-$	5×10^{-6}	2730	5506
$\Lambda_b \to pK^-$	5×10^{-6}	2709	5465

Reconstruction of reflections in simulated events

In the following the two decay modes $B^0 \to K^+\pi^-$ and $\Lambda_b \to p\pi^-$ are reconstructed in simulated events using the $\Lambda_b \to K^-\mu^+$ event selection. Both modes are kinematically and topologically similar to $\Lambda_b \to K^-\mu^+$ and allow to study the invariant mass spectra and estimate their rate in the data. However, the quantitative results that are presented can only give a ballpark estimate due to the imperfect simulation of particle identification and trigger efficiencies. Additional requirements from the final event selection are also not included.

$B^0 \to K^+\pi^-$

This decay has a relatively large branching fraction of $\mathcal{B}(B^0 \to K^+\pi^-) \approx 2 \cdot 10^{-5}$ [3] and is expected to provide the biggest contribution among the reflections. From a sample of 1 530 k generated events within LHCb acceptance there are about 1 000 candidates left after event and trigger selections. This corresponds to a selection efficiency of $\varepsilon_{\text{sel}} \approx 7 \cdot 10^{-4}$. According to a fit to the mass distribution of reconstructed events there are on average about $2 \cdot 10^{-3}$ events in the signal region $m(K^-\mu^+) = m(\Lambda_b) \pm 50\,\text{MeV}/c^2$. Hence the mean selection efficiency here is $\varepsilon_{\text{sel}} \approx 2 \cdot 10^{-3}/1530\,k \approx 2 \cdot 10^{-9}$. The expected average number of events in the data is

$N(B^0 \to K^+\pi^-) = N(B^0) \times \mathcal{B}(B^0 \to K^+\pi^-) \times \varepsilon_{\text{sel}}$ which is about $10^{11} \times 2 \cdot 10^{-5} \times 7 \cdot 10^{-4} = 1400$ in total and $< 1/100$ in the $\Lambda_b \to K^-\mu^+$ signal region.

Figure 3.3 shows the invariant mass distribution of the $B^0 \to K^+\pi^-$ reflection after event and trigger selections. The $B^0 \to K^-\mu^+$ reflection is separated into the two possible misidentification modes (1) $K \to K$, $\pi \to \mu$ and (2) $K \to \mu$, $\pi \to K$. For comparison the mass spectrum of simulated and reconstructed $\Lambda_b \to K^-\mu^+$ events is also included.

In conclusion this decay mode is well separated from the signal channel and the mass shift is in good agreement with the cms approximation. The impact from other two-body B^0 decays can be expected to be smaller due to the lower branching fractions. In particular the decay modes $B_s^0 \to K^+K^-$ and $B_s^0 \to K^+\pi^-$ should be less frequent due to the four times smaller production rate of B_s^0 mesons.

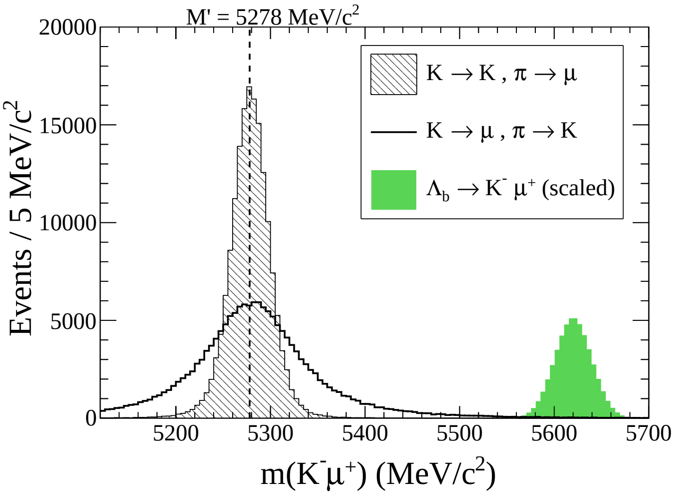


Figure 3.3: Mass distribution of the $B^0 \to K^+\pi^-$ reflection after selections.

$\Lambda_b \to p\pi^-$

The second mode that is investigated has a branching ratio of $\mathcal{B}(\Lambda_b \to p\pi^-) \approx 5 \cdot 10^{-6}$ [3]. From a sample of 1 510 k generated events within LHCb acceptance there are about 700 candidates left after event and trigger selections. This corresponds to an efficiency of $5 \cdot 10^{-4}$. In the Λ_b signal window there are about 50 events which leads to an efficiency of $4 \cdot 10^{-5}$. These numbers translate to $N(\Lambda_b \to p\pi^-) = N(\Lambda_b) \cdot \mathcal{B}(\Lambda_b \to p\pi^-) \cdot \varepsilon_{\text{sel}} \approx 120$ events in total and $N(\Lambda_b \to p\pi^-) \approx 10$ events that are expected in the Λ_b signal window in the data. Here, the number of Λ_b baryons within detector acceptance is estimated to be around 50×10^9, see Sec. 2.3.

Fig. 3.4 shows the invariant mass distribution for the two different misidentification modes along with the expected signal of $\Lambda_b \to K^-\mu^+$. In conclusion the mass distribution is broad but overlaps clearly with the signal window. The influence of $\Lambda_b \to pK^-$ can be expected to be similar due to a comparable branching ratio [3].

3.5.2 Background from partially reconstructed decays

Another relevant source of physical background are decays where one daughter particle is not used for reconstruction. Due to the fact that the energy of the missing particle is a spectrum, the invariant mass distribution of the remaining particles is broad. The most relevant cases to consider are on one side the Cabbibo-suppressed, semileptonic decays $\Lambda_b \to p\mu^-\overline{\nu}_\mu$ and $B_s^0 \to K^+\mu^-\overline{\nu}_\mu$, and on the other side $B_c^+ \to J/\psi\,(\mu^+\mu^-)K^+$ and $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$.

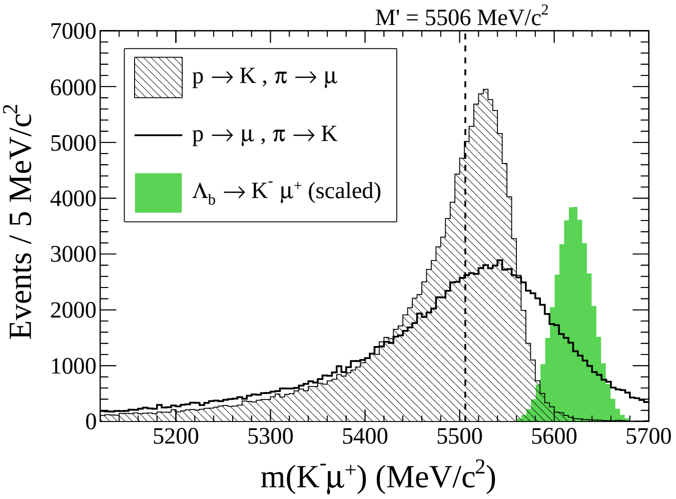


Figure 3.4: Mass distribution of the $\Lambda_b \to p\pi^-$ reflection after selections.

Semileptonic decays of the sort $h_b \to h^+\mu^-\overline{\nu}_\mu$

The relatively large branching ratio of $\mathcal{B}(B^0 \to \pi^+\ell^-\nu_\ell) \approx 1.5 \times 10^{-4}$ leads to the expectation that the decay rates of $\Lambda_b \to p\mu^-\overline{\nu}_\mu$ and $B_s^0 \to K^+\mu^-\overline{\nu}_\mu$, that have not been measured until now, are in the same order of magnitude. Such events can be selected due to the displaced vertex of the b hadron and the muon in the final state.

From the kinematic point of view the distribution of $m(h,\mu)$ will end around the invariant mass of the original mother particle in the case that the neutrino is produced at rest in the cms of the mother particle. This leads to the expectation that $\Lambda_b \to p\mu^-\overline{\nu}_\mu$ will make the largest portion of partially reconstructed decays in the Λ_b mass region.

Reconstruction of $\Lambda_b \to p\mu^-\nu_\mu$ in simulated events

The used sample consist of 2 576 k events within LHCb acceptance that were generated using light-cone sum rules to model the $\Lambda_b \to p$ form factor [61]. After event and trigger selections there are about 770 events left leading to a reconstruction efficiency of $\varepsilon_{\text{sel}} \approx 3 \times 10^{-4}$. Assuming a branching fraction of $\mathcal{B}(\Lambda_b \to p\mu^-\overline{\nu}_\mu) = 10^{-4}$, similar to $\mathcal{B}(B^0 \to \pi^-\ell^+\nu)$, and $N(\Lambda_b) \approx 5 \times 10^{10}$ within LHCb acceptance, there are about 1500 events left in the data after event and trigger selections. Figure 3.5 shows the invariant mass spectrum for the two misidentification modes after selections. The majority of the events are reconstructed below the Λ_b signal region. Due to the smaller mass of the B_s^0 meson and its lower rate in the data, the contribution from $B_s^0 \to K^+\mu^-\nu_\mu$ is expected to be less.

The resonant two-body decay $B_c^- \to J/\psi\,(\mu^+\mu^-)K^-$

The decay $B_c^- \to J/\psi\,(\mu^+\mu^-)K^-$ is a particular case of physical background because combinations of real kaons and muons can be reconstructed to a common and displaced vertex. In addition it is not clear in the first place if the spectrum of $m(K^-\mu^+)$ ends below $m(\Lambda_b)$ due to the large mass $m(B_c^+) \approx 6275\,\text{MeV}/c^2$. However, starting from kinematic considerations, the range of $m(K^-\mu^+)$ can be determined to be $607\ldots5450\,\text{MeV}/c^2$, *i.e.* the end point is significantly below $m(\Lambda_b) = 5620\,\text{MeV}/c^2$. A detailed calculation is presented in App. A.8.

Reconstruction of $B_c^- \to J/\psi\,(\mu^+\mu^-)K^-$ in simulated events

Using a sample of 2 048 k generated events within LHCb acceptance, there are about 5 900 candidates left after event and trigger selections. The efficiency is $\varepsilon_{\text{sel}} \approx 3 \times 10^{-3}$. The ratio of

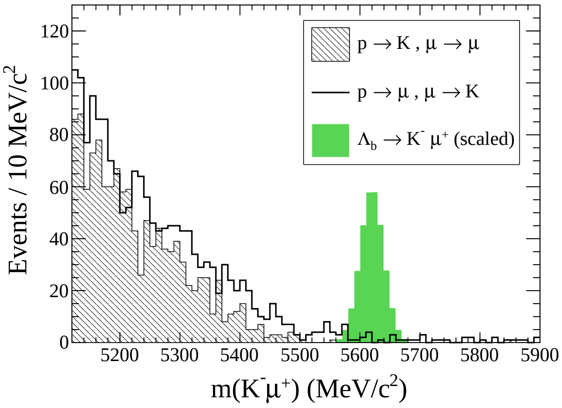
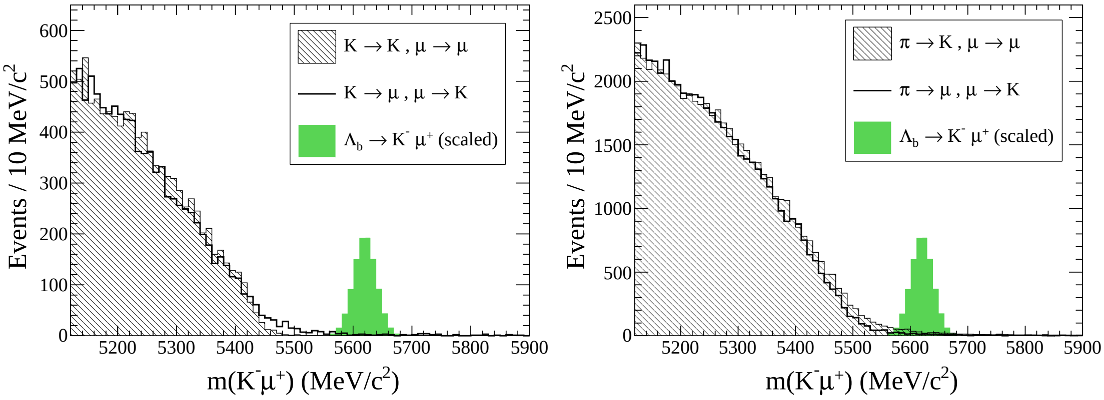


Figure 3.5: Mass distribution of partially reconstructed $\Lambda_b \to p\mu^- \overline{\nu}_\mu$ events after selections.

the production cross sections of B_c^+ over B^+ can be estimated to be about 1/415, see App. A.9. Therefore the anticipated number of background events is roughly $N(B^+) \times f_c/f_u \times \mathcal{B}(B_c^- \to J/\psi\, K^-) \times \mathcal{B}(J/\psi \to \mu^+\mu^-) \times \varepsilon_{\text{sel}} \approx 10^{11} \times 1/400 \times 3\cdot 10^{-3} \times 6\cdot 10^{-2} \times 3\cdot 10^{-3} = 130$. Figure 3.6(a) shows the invariant mass spectrum after selections for the two possible assignments of particles. It can be seen that the calculated end point of $m(K^-\mu^+) = 5450\,\text{MeV}/c^2$ is correct for the right particle assignment apart from some smearing effect due to the finite momentum resolution.



(a) $B_c^+ \to J/\psi\,(\mu^+\mu^-)K^+$ events
(b) $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ events

Figure 3.6: $m(K^-\mu^+)$ distribution after selections in $B_c^+ \to J/\psi\,(\mu^+\mu^-)K^+$ events (a) and $B_c^+ \to J/\psi\,(\mu^+\mu^-)\pi^+$ events (b).

The same studies have been repeated on a sample of roughly 8 000 k simulated events of $B_c^- \to J/\psi\,(\mu^+\mu^-)\pi^-$. In contrast to $B_c^- \to J/\psi\,(\mu^+\mu^-)K^-$ this decay is not Cabbibo suppressed. However, the selection efficiency is reduced by a factor of about 10 resulting from to the PID requirement on the pion track. Figure 3.6(b) shows the invariant mass spectrum after selections.

In summary none of the two B_c^- decays are expected to be effectively harmful due to the small number of events that are widely spread below $m(\Lambda_b)$.

"Control is an illusion, you infantile egomaniac."

— Dr. Claire Lewicki (from the movie "Days of Thunder")

4

Control channel studies

In the following a number of known decays are investigated in real data and MC events to validate the different aspects of the LHCb event simulation. This is in particular necessary to check to what extend signal MC of $\Lambda_b \to K^- \mu^+$ can be used for the optimisation of the signal selection. These decay channels are $B^0 \to K^+\pi^-$ and $\Lambda_b \to \Lambda_c \pi^-$ to check the simulation of event variables such as $p_\mathrm{T}(\Lambda_b)$ and DOCA. The simulation of trigger lines are examined using $B^+ \to J/\psi K^+$ events. In addition, reconstructed $\Lambda_b \to J/\psi p K^-$ events allow to assess the influence of muon and hadron triggers on the distribution of observables in Λ_b events. The main properties of the relevant particles are summarised in App. A.10

4.1 $B^0 \to K^+\pi^-$ for studies of event variables

The following studies are made using the sample of inclusive $B \to \mu^+\mu^-$ events that were refitted by applying kaon and pion mass hypotheses to the two tracks t_1 and t_2. To assign the correct masses to the tracks t_1 and t_2, on which no PID selections have been made in the original sample, it is chosen $K = t_1$ if $\mathrm{DLL(K-\pi)(t_1)} > \mathrm{DLL(K-\pi)(t_2)}$, *i.e.* track 1 is more kaon like than track 2, and $K = t_2$ in the other case. Afterwards a loose PID selection of $\mathrm{DLL(K-\pi)} > -5$ is required for the kaon which reduces the background level by a factor of three while keeping almost 100 % of the signal.

4.1.1 Trigger selection

Figure 4.1 shows the $m(K^+\pi^-)$ distribution of events using all triggers, after applying the signal triggers and the distribution of rejected events, respectively. The trigger requirement removes about 2/3 of the background events while only a small amount of signal events is lost. In conclusion the use of triggers is justified due to the significantly improved signal-to-background ratio.

4.1.2 Determination of the mass shift

The absolute scale of the invariant mass of two particles is sensitive to a correct momentum measurement. Due to imperfections in the mapping of the magnetic field the measured momenta are biased and reconstructed masses are shifted which has to be considered in the analysis of the data. A fit is made to the $m(K^+\pi^-)$ distribution to validate the mass shapes provided by simulated events and measure the mass shift. The fit model is described in detail in App. A.11 and involves

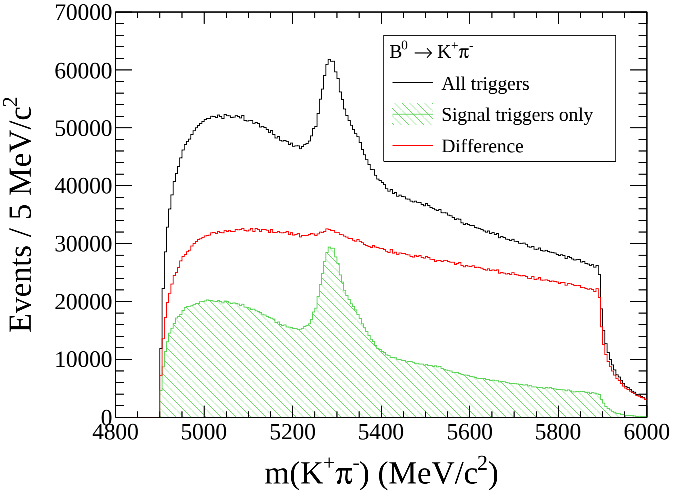


Figure 4.1: Impact of the trigger on the $m(K^+\pi^-)$ distribution in 2012 data.

the decay modes: $B^0 \to K^+\pi^-$, $B_s^0 \to K^+\pi^-$, $B_s^0 \to K^+K^-$, $B^0 \to \pi^+\pi^-$, $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$. In the fit to the data all parameters are fixed to the values determined from MC and a floating parameter D is introduced to allow all mean masses to be shifted commonly, *i.e.* $\hat{m}_{\text{Fit}}^{\text{data}} = \hat{m}_{\text{Fit}}^{\text{MC}} + D$. The background is described by a second order polynomial. Contributions from partially reconstructed decays like $B^0 \to \pi^+\mu^-\nu_\mu$ are assumed to be small in the fit range $m(K^+\pi^-) = 5150\,\text{MeV}/c^2 \ldots 5890\,\text{MeV}/c^2$ and to be absorbed by the background model. Figure 4.2 shows the result of the binned likelihood fit to the mass distribution. In summary the simulated shapes of the $m(K^+\pi^-)$ spectrum is compatible with real data and the fitted mass shift is $D = (4.13 \pm 0.17)\,\text{MeV}/c^2$. Due to the fact that the $\Lambda_b \to K^-\mu^+$ reconstruction uses similar mass hypotheses and high-momentum tracks it is expected that the mass shift of the Λ_b will be comparable.

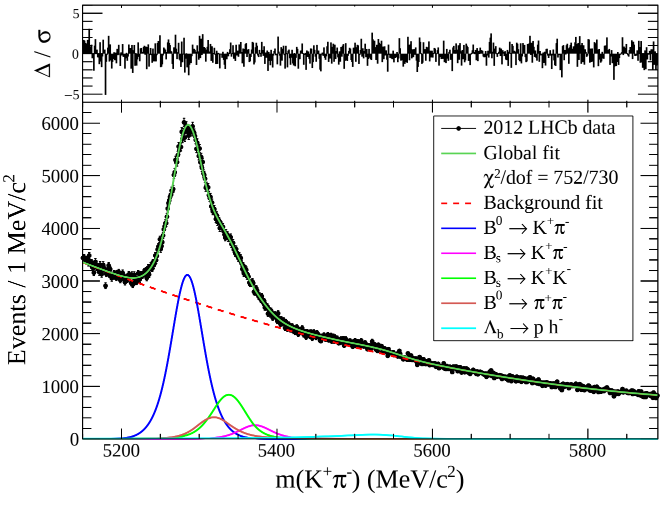


Figure 4.2: Fit to the $m(K^+\pi^-)$ distribution from selected 2012 data.

Along with the mass shift the normalisation $N(B^0 \to K^+\pi^-)$ is fitted as well as the suppression of the other modes relative to $B^0 \to K^+\pi^-$: $\mathcal{S}(x) = N_{\text{Fit}}(B^0 \to K^+\pi^-)/N_{\text{Fit}}(x)$. The fit results are $N(B^0 \to K^+\pi^-) = 168\,115 \pm 934$, $\mathcal{S}(B_s^0 \to K^+\pi^-) = 12.0 \pm 0.6$, $\mathcal{S}(B_s^0 \to K^+K^-) = 3.51 \pm 0.08$, $\mathcal{S}(B^0 \to \pi^+\pi^-) = 6.6 \pm 0.4$, $\mathcal{S}(\Lambda_b \to p\pi^-) = 21.4 \pm 2.2$, $\mathcal{S}(\Lambda_b \to pK^-) = 39 \pm 20$. It is visible that both Λ_b decay modes are strongly suppressed. Due to the fact that the $\Lambda_b \to K^-\mu^+$ event selection is similar to this one, *i.e.* the other track than the kaon requires in addition IsMuon=true, reflections from $\Lambda_b \to p\pi^-$ and $\Lambda_b \to pK^-$ are also expected to provide only small contributions relative to the $B_{(s)}$ decay modes.

4.1.3 Discrepancies between data and signal MC

The large number of $B^0 \to K^+\pi^-$ events and the good signal-to-background ratio is exploited to make comparisons between real data and about 165 k simulated events in a number of variables that can be used to separate signal and background. Due to the similarity of $B^0 \to K^+\pi^-$ and $\Lambda_b \to K^-\mu^+$ events, *i.e.* two-body b-hadron decays with light final state particles and a large energy release, it is assumed that well simulated variables can be used for the optimisation of the signal selection using $\Lambda_b \to K^-\mu^+$ MC. These studies are complemented by comparisons of reconstructed $\Lambda_b \to \Lambda_c\pi^-$ events in real data to MC below.

Figure 4.3 shows the $m(K^+\pi^-)$ distribution from the sample of $B^0 \to K^+\pi^-$ events requiring the signal triggers. To obtain background-suppressed distributions of signal variables, a sideband subtraction is made using the sidebands I and II and the signal range of $m(K^+\pi^-) = 5235\ldots5335\,\mathrm{MeV}/c^2$ as indicated in Fig. 4.3. The signal band is centered $5\,\mathrm{MeV}/c^2$ above the nominal B^0 mass to consider the mass shift that is rounded up within the precision given by the binning of $2\,\mathrm{MeV}/c^2$. Comparisons in signal MC show that contributions from other decays in the signal range, namely $B^0_s \to K^+K^-$ and $B^0 \to \pi^+\pi^-$, do not differ in the considered variables. A third mass sideband (III) is considered to compare the signal variable distributions in dependence of the invariant mass. A binned likelihood fit is performed to the mass sidebands $m(K^+\pi^-) = 5150\,\mathrm{MeV}/c^2\ldots5200\,\mathrm{MeV}/c^2$ and $m(K^+\pi^-) = 5600\,\mathrm{MeV}/c^2\ldots5890\,\mathrm{MeV}/c^2$ using a second order polynomial to interpolate the number of background events in the signal range. The signal-to-background ratio here is 1/1.3.

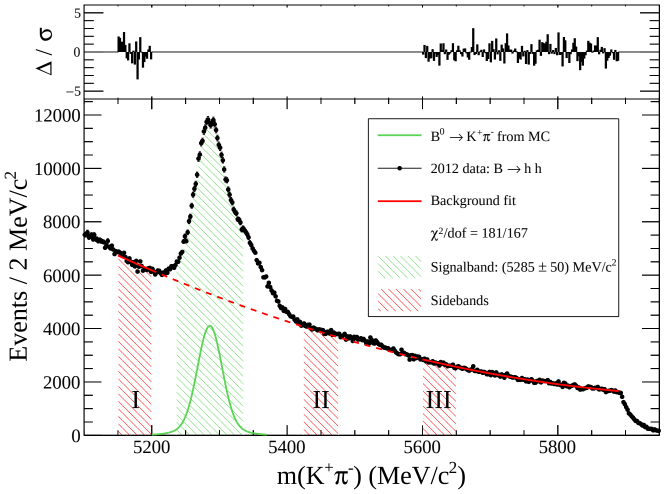


Figure 4.3: $m(K^+\pi^-)$ distribution from 2012 data requiring signal triggers.

In addition to the variables that are used for event selection, a number of other variables are considered in the data-MC comparison which may be used to discriminate background events.

Pseudorapidity: η

The pseudorapidity of a particle is defined as

$$\eta = \frac{1}{2} \cdot \ln\left(\frac{p + p_\mathrm{L}}{p - p_\mathrm{L}}\right) = -\ln\left[\tan(\theta/2)\right] ,$$

where θ is the polar angle between the momentum and the beam axis and $p_\mathrm{L} = p \cdot \cos(\theta)$.

Track multiplicity: nTracks

The number of unique reconstructed tracks per event as a measure for the event multiplicity.

CDF isolation: $\mathrm{Iso}_{\mathrm{CDF}}$

The CDF isolation is defined as follows:

$$\mathrm{Iso}_{\mathrm{CDF}} = \frac{p_{\mathrm{T}}(h_b)}{p_{\mathrm{T}}(h_b) + \sum_{\mathrm{tracks}} |p_{\mathrm{T}}|} \, .$$

Here, h_b stands for a reconstructed b hadron. The summation is made over all tracks that are not used for the reconstruction of h_b in the cone $\sqrt{\delta\eta^2 + \delta\theta^2} < 1.0$, where $\delta\eta$ and $\delta\theta$ are the difference in the pseudorapidity and the polar angle between the track and h_b.

Track isolation: Iso5

For a given track Iso5 is the number of tracks in the cone $\sqrt{\delta\eta^2 + \delta\theta^2} < 0.5$. Tracks from other daughter particles of the same decay are not counted.

Combined daughter variables: DauMax[X] and DauMin[X]

Signal variables X of the daughter tracks are combined by taking their maximum or minimum. The maximum is chosen if the variable is typically larger for background events than for signal events, *e.g.* DauMax[Iso5], and the minimum in the other case, *e.g.* DauMin[p_{T}]. In both cases the idea is to choose the less signal-like value.

The SP value to quantify discrepancies

In order to quantify the difference between two histograms $\mathrm{H_X}$ and $\mathrm{H_Y}$, an empirical quantity named separating power (SP) is introduced. The value of SP is determined from the normalised histograms $\mathrm{h_X}$ and $\mathrm{h_Y}$ by using their bin contents $\mathrm{h(i) = H(i)/N,\ N = \sum_i H(i)}$ according to

$$\mathrm{SP(X,Y)} = \frac{1}{2}\sum_{\mathrm{i}}^{\mathrm{bins}} |\mathrm{h_X(i) - h_Y(i)}| = 0 \ldots 100\,\% \, . \tag{4.1}$$

SP can have values in the range $(0, 1)$ where 0 means total agreement of both distributions and no separating power and 1 stands for no overlap and maximum separating power. The definition of the SP variable is illustrated in Fig. 4.4. The statistical uncertainty of the SP value is calculated from error propagation of the bin contents, *i.e.* $\delta^2[\mathrm{h(i)}] = \delta^2[\mathrm{H(i)}]/\mathrm{N}^2$. Studies with random samples of the same underlying distribution show that the SP value scales with $1/\sqrt{n}$, where $n = N/B$ is the average number of events per bin. The statistical uncertainty scales with $1/\sqrt{N}$.

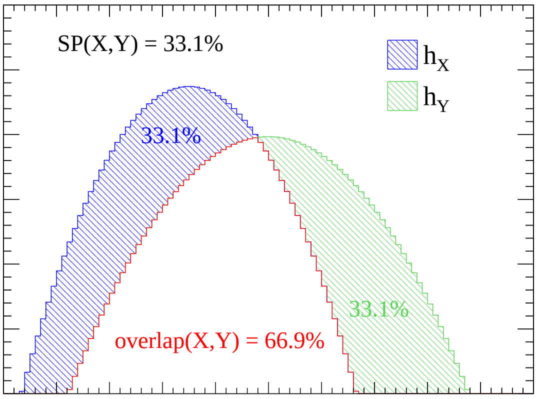


Figure 4.4: Illustration of the SP variable.

Table 4.1 lists in the first column the SP value comparing sideband subtracted signal distributions to signal MC and in brackets the expected SP value by comparing one half of the signal MC sample with the other half. The value is scaled by $1/\sqrt{2}$ to extrapolate to the full sample size. It allows to assess systematic biases due to the particular shape of the distribution and the number of entries. The statistical uncertainty of all SP values is in order of 0.2 %. A good agreement with less than 5 % difference is found for most of the variables besides the track quality variables and the track multiplicity.

Table 4.1: SP from a sideband subtraction in $B^0 \to K^+\pi^-$ events from 2012 data and signal MC. The numbers in brackets are the expected values as described in the text.

Signal variables	**SP(Sub,MC)**		**SP(Sub,Sb)**
ϕ_{Dira}	0.86 %	[0.41 %]	44.5 %
DOCA	1.27 %	[0.72 %]	27.0 %
$\tau(B^0)$	1.63 %	[0.93 %]	25.7 %
IP	1.70 %	[0.86 %]	45.4 %
$p/M(B^0)$	1.89 %	[1.31 %]	11.1 %
$\chi^2_{/\text{DOF}}(\text{Vtx})$	1.93 %	[0.88 %]	15.1 %
$\eta(B^0)$	2.04 %	[0.69 %]	39.3 %
DauMin[χ^2(IP)]	2.05 %	[1.27 %]	45.7 %
DauMin[IP]	2.10 %	[1.55 %]	37.1 %
χ^2(FD)	2.25 %	[1.40 %]	31.6 %
$p(B^0)$	2.29 %	[1.66 %]	11.8 %
FD	2.61 %	[1.59 %]	24.3 %
χ^2(IP)	2.97 %	[0.80 %]	40.9 %
DauMin[p_{T}]	3.17 %	[0.95 %]	31.8 %
$p_{\text{T}}(B^0)$	4.11 %	[1.24 %]	51.3 %
DauMax[Iso5]	4.83 %	[0.21 %]	51.1 %
Iso$_{\text{CDF}}$	5.98 %	[1.30 %]	37.5 %
DauMax[$\chi^2_{/\text{DOF}}$(Tr)]	10.0 %	[0.83 %]	14.0 %
nTracks	12.0 %	[1.15 %]	16.6 %
DauMax[P(Gh)]	30.5 %	[0.72 %]	21.8 %
$p_{\text{T}}(K^+)$	2.13 %	[1.18 %]	32.5 %
$p_{\text{T}}(\pi^-)$	2.41 %	[1.33 %]	29.7 %
$p(K^+)$	2.48 %	[1.65 %]	11.6 %
$p(\pi^-)$	2.85 %	[1.81 %]	8.51 %

The second column show the SP value comparing sideband-subtracted signal distributions with the combined distribution of the sideband regions I and II. A number of variables are well simulated in MC and at the same time provide a good signal-to-background separation, for example ϕ_{Dira}, IP, DauMin[χ^2(IP)], DOCA and DauMax[Iso5]. Plots of all variables can be found in App. A.12. It can be concluded that geometric variables that are sensitive to a correct decay vertex reconstruction, *e.g.* ϕ_{Dira}, IP or $\chi^2_{/\mathrm{DOF}}(\mathrm{Vtx})$, are well described in simulated events. Due to the underestimation of the track multiplicity in signal MC, on average about 25 tracks less w.r.t. data events, the isolation variables are slightly biased. Variables that are especially related to the Λ_b production, *e.g.* $\eta(\Lambda_b)$, are studied separately using $\Lambda_b \to \Lambda_c \pi^-$ events.

4.2 $B^+ \to J/\psi K^+$ for trigger studies

For this studies $B^+ \to J/\psi K^+$ events are used where the J/ψ is reconstructed as $J/\psi \to \mu^+\mu^-$. The decay has a trigger signature similar to $\Lambda_b \to K^-\mu^+$ and a large product branching fraction of approximately $6 \cdot 10^{-5}$ providing a large and almost background-free sample of signal events. This allows comparisons between recorded and simulated events to study the precision of the trigger simulation.

The event selection combines two identified muon tracks with an invariant mass of $100\,\mathrm{MeV}/c^2$ around the nominal J/ψ mass to a detached decay vertex. Loose selections on properties of the vertex and track reconstruction are implied. In the next step the J/ψ candidates are combined with a charged track that has kaon-mass hypothesis and no PID selections. The reconstructed B^+ requires $m(J/\psi K^+) = m(B^+) \pm 500\,\mathrm{MeV}/c^2$, $\chi^2(\mathrm{Vtx}) < 45$ and $\chi^2(\mathrm{IP}) < 25$. The event selection has an efficiency of approximately 25 % in simulated events and is summarised in App. A.13. Figure 4.5 shows the spectrum of $\Delta m(B^+, J/\psi) = m(\mu^+\mu^-K^+) - m(\mu^+\mu^-)$ after event selections and using all LHCb triggers. To suppress combinatorial background and remnants of $B^+ \to J/\psi \pi^-$ events the invariant $\mu^+\mu^-$ mass is required to be in a mass window of $\pm 50\,\mathrm{MeV}/c^2$ around the nominal J/ψ mass and $\mathrm{DLL}(\mathrm{K}-\pi) > 0$ for the kaon track. The distribution is fitted using the sum of two Crystal Ball functions with a common mean value for the signal component and a first-order polynomial for combinatorial background. There are approximately $845 \cdot 10^3$ signal events in total.

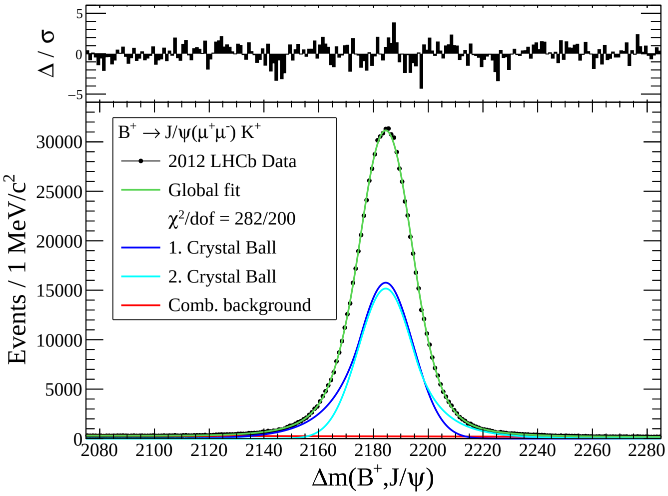


Figure 4.5: $\Delta m(B^+, J/\psi)$ distribution of $B^+ \to J/\psi K^+$ events from 2012 data.

4.2.1 Trigger studies

The large sample of $B^+ \to J/\psi K^+$ events allows studies on the the accuracy of the trigger simulation by comparing the trigger efficiency between data and MC. In signal MC the trigger efficiencies are measured as the ratio between the number of signal events after requiring one of the TOS lines and the total number of events that are selected by all physical trigger lines in all stages (T_0), *i.e.* $\varepsilon_{\mathrm{MC}}(\mathrm{TOS}) = N(\mathrm{TOS}\,\&\,T_0)/N(T_0)$. In real data each trigger efficiency is determined with the help of TIS triggered events which provide a sample of $B^+ \to J/\psi K^+$ decays that is selected due to the rest of the event and hence independently from the B^+ decay products. Due to the fact that the decisions of the three trigger stages have only small dependencies, the efficiency of, *e.g.*, a particular L0 trigger can be determined in the sample of events that only satisfy L0-TIS, but not HLT1-TIS and HLT2-TIS. This means that 100 % corresponds to all $B^+ \to J/\psi K^+$ events that are TIS in the related trigger stage. In 2012 data there are approximately 225 k L0-TIS events, 120 k HLT1-TIS events and 26 k HLT2-TIS events. The efficiency of each TOS trigger is measured as $\varepsilon_{\mathrm{Data}}(\mathrm{TOS}) = N(\mathrm{TOS}\,\&\,\mathrm{TIS}\,\&\,T_0)/N(\mathrm{TIS}\,\&\,T_0)$, whereby the T_0 requirement is a natural constraint of the experiment. This method is abbreviated as TISTOS method. In all studies the number of signal events is determined from a binned likelihood fit to the invariant mass spectrum as it was done before.

Table 4.2 lists the resulting efficiencies. The binomial uncertainties on all efficiencies are $< 0.1\,\%$ except for the HLT2 trigger in data where $\delta(\varepsilon) \approx 0.2\,\%$ due to the limited number of signal events. Good agreement is found for all HLT lines. However, the efficiencies of the L0 trigger is overestimated in MC by more than 5 % w.r.t. the data. This may be caused by the fact that (1) in MC only one trigger configuration is simulated while different configurations were used during data taking and (2) the L0 trigger is not correctly simulated. In order to obtain correct L0 efficiencies for normalisation, samples of $B^+ \to J/\psi K^+$ and $\Lambda_b \to \Lambda_c\pi^-$ events from real data are used for calibration which automatically considers the right mix of different trigger configurations.

Table 4.2: Trigger efficiencies in 2012 data and signal MC.

Trigger	$\varepsilon_{\mathrm{Data}}$	$\varepsilon_{\mathrm{MC}}$	$\varepsilon_{\mathrm{MC}}/\varepsilon_{\mathrm{Data}} - 1$
L0Hadron-TOS	14.5 %	15.3 %	5.2 %
L0Muon-TOS	86.6 %	94.1 %	8.7 %
Combination	87.6 %	95.5 %	9.1 %
AllTrack-TOS	86.3 %	86.3 %	0.0 %
MuonTrack-TOS	92.7 %	91.7 %	−1.0 %
Combination	97.1 %	96.5 %	−0.6 %
Topo2Body-TOS	83.6 %	82.5 %	−1.3 %
TopoMu2Body-TOS	89.3 %	89.5 %	0.2 %
Combination	89.3 %	89.5 %	0.2 %

4.3 $\Lambda_b \to \Lambda_c\pi^-$ for Λ_b properties

$\Lambda_b \to \Lambda_c\pi^-$ events are reconstructed using the subsequent decay $\Lambda_c \to pK^-\pi^+$. It has a product branching fraction of around $3 \cdot 10^{-4}$ and provides a large number of events to validate the simulation of Λ_b properties such as boost (p/m), proper lifetime and pseudorapidity.

For the event selection Λ_c candidates are formed from three charged tracks with proton, kaon, and pion mass hypothesis, respectively, and without PID requirements. Each track has to satisfy $p > 1\,\mathrm{GeV}/c$, $p_\mathrm{T} > 0.1\,\mathrm{GeV}/c$, $\mathrm{P(Gh)} < 0.3$, $\chi^2(\mathrm{IP}) > 4$, and $\chi^2_{/\mathrm{DOF}}(\mathrm{Tr}) < 3$. To suppress combinatorial background, soft selections on the invariant mass, the momentum and the DOCA of all possible track pairs are made. A good vertex fit is required as well as a significant flight distance w.r.t. all PV's. Λ_b decays are reconstructed by combining the Λ_c candidates with a pion track that requires the same criteria as the other tracks. Wrong Λ_b candidates are rejected by imposing selections on their momentum, lifetime, impact parameter and vertex quality. A summary of the event selection can be found in App. A.14. The efficiency in simulated events is around 3 %. To filter the large data sample after the soft event selection, the $pK^-\pi^+$ combinations must satisfy $m(pK^-\pi^+) = 2265\ldots2310\,\mathrm{MeV}/c^2$ which selects more than 99 % of the reconstructed Λ_c baryons, see App. A.14.1, and suppresses the reflections from $\Lambda_b \to D^0 p\pi^-$ and $\Lambda_b \to D^0 pK^-$ to a negligible amount. Additional selections are the signal trigger conditions, $\mathrm{ProbNN}(p) > 5\,\%$, $\mathrm{ProbNN}(K) > 5\,\%$, and, to suppress contributions from $\Lambda_b \to \Lambda_c K^-$, $\mathrm{ProbNN}(\pi) > 5\,\%$ for the pion from the Λ_b decay. Figure 4.6 show the resulting mass spectrum with a large and clean signal peak.

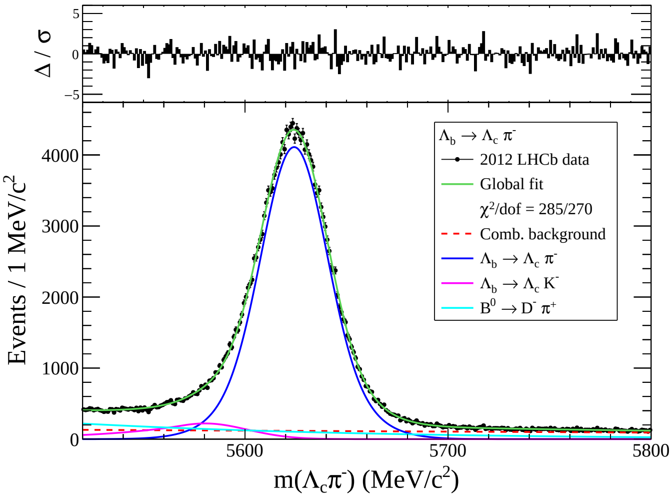
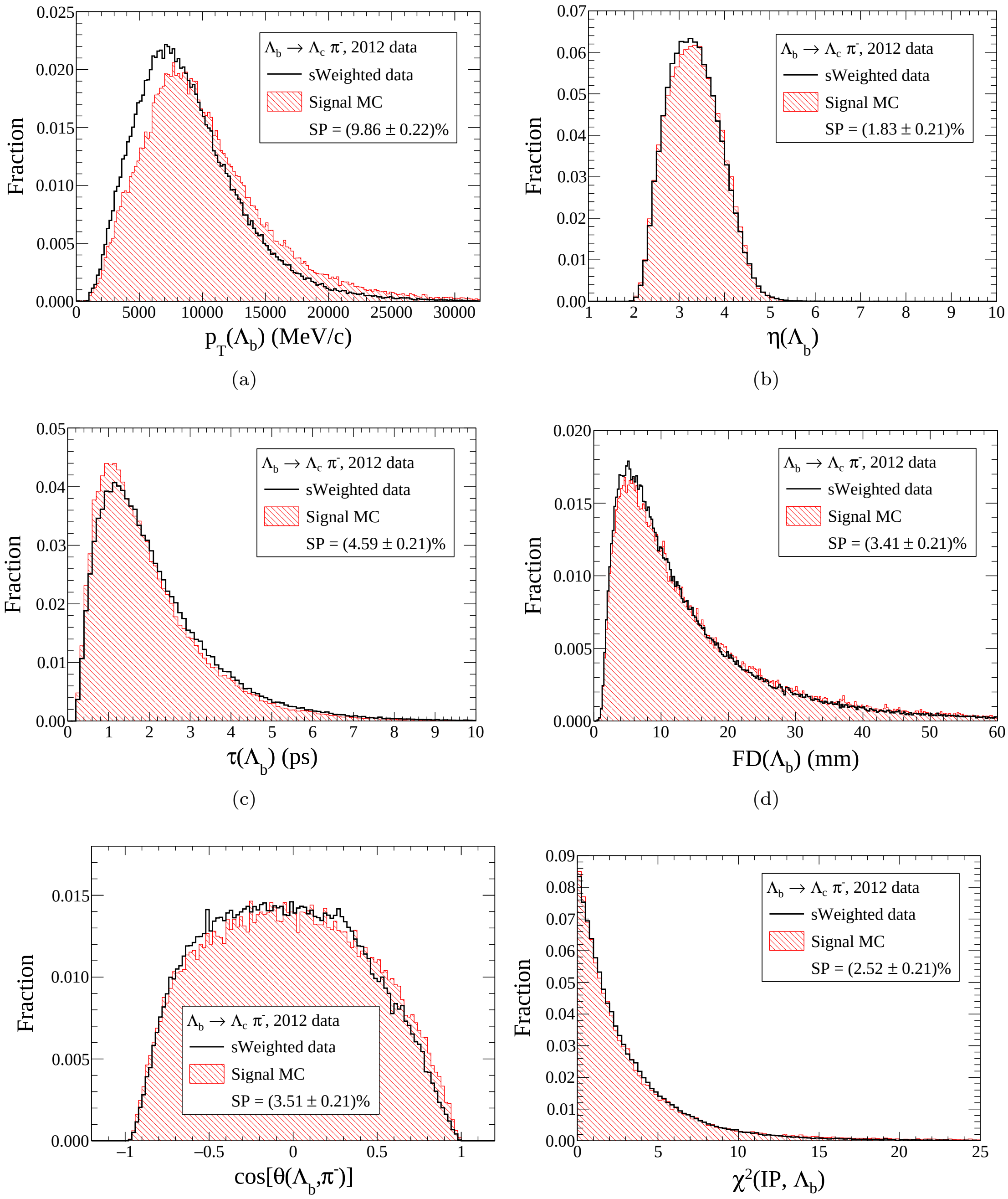


Figure 4.6: $m(\Lambda_c\pi^-)$ distribution of selected events from 2012 data.

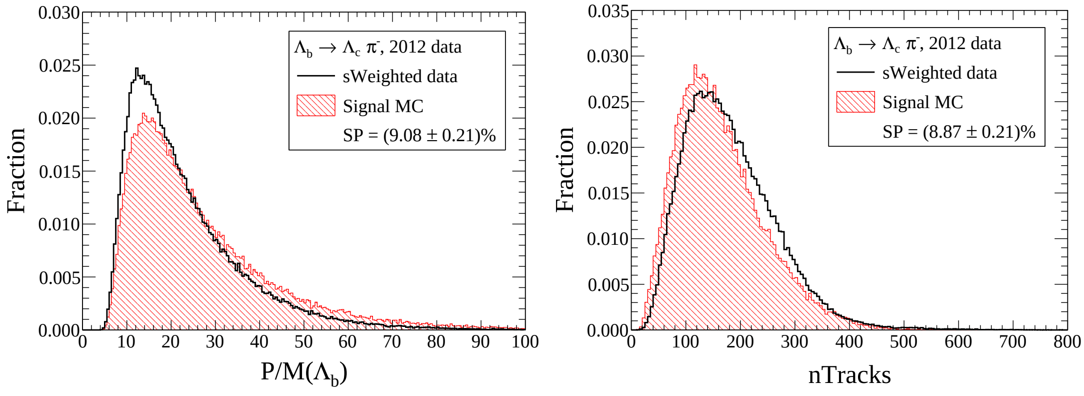
A fit is made to the mass distribution in the range $m(\Lambda_c\pi^-) = 5520\ldots5800\,\mathrm{MeV}/c^2$ to calculate *sWeights* which allow to obtain sideband-subtracted distributions of variables (*sPlots*) [62]. The model consists of a first-order polynomial for the background and three components to describe the decays $\Lambda_b \to \Lambda_c\pi^-$, $\Lambda_b \to \Lambda_c K^-$, and $B^0 \to D^+\pi^-$, respectively. All models are tested in fits to simulated events that are reconstructed using the $\Lambda_b \to \Lambda_c\pi^-$ event selection. The parameters of the modes $\Lambda_b \to \Lambda_c K^-$ and $B^0 \to D^+\pi^-$ are fixed except for the normalisation. Due to the large number of $\Lambda_b \to \Lambda_c\pi^-$ signal events all parameters are free except for the mean mass that is fixed to the value obtained from signal MC. A free fit parameter is used to commonly shift the mean mass of all modes. Details about the different fit models can be found in App. A.14.2. According to the fit there are $N(\Lambda_b \to \Lambda_c\pi^-) = 182\,814 \pm 794$ events.

The sample of signal events is used to assess differences between real data and simulated events. All events are required to satisfy the selection described above including the signal triggers. Figures 4.7 and 4.8 compare the distributions from reconstructed signal MC and background subtracted data using *sWeights*. There is a good agreement for the variables $\eta(\Lambda_b)$, $\mathrm{FD}(\Lambda_b)$, and $\chi^2(\mathrm{IP})$. On average the Λ_b are generated with a too large boost (MC: 26.4, Data: 22.8) which is also visible in the $p_\mathrm{T}(\Lambda_b)$ distribution. The proper lifetime is a bit too small in the simulation due to the fact that the Λ_b are generated with an average lifetime of $\tau_\mathrm{gen} = 1.38\,\mathrm{ps}$ while the present world average is measured to be $\tau = (1.466 \pm 0.010)$ ps [3]. These two effects mostly cancel out in the simulation of the flight distance. The distribution of the observed helicity angle $\cos[\theta(\Lambda_b, \pi^-)]$ indicates a slight forward-backward asymmetry in data w.r.t. reconstructed signal MC. A comparison of the track multiplicity shows that there are too little tracks per event in the simulation (MC: 163, Data: 180). To obtain correct values for the Λ_b boost and the track multiplicity a calibration is made that is explained in the next sections.



(a) (b) (c) (d) (e) (f)

Figure 4.7: *sPlot*'ted distributions from $\Lambda_b \to \Lambda_c\pi^-$ events in 2012 data and signal MC.



(a) (b)

Figure 4.8: *sPlot*'ted distributions from $\Lambda_b \to \Lambda_c\pi^-$ events in 2012 data and signal MC.

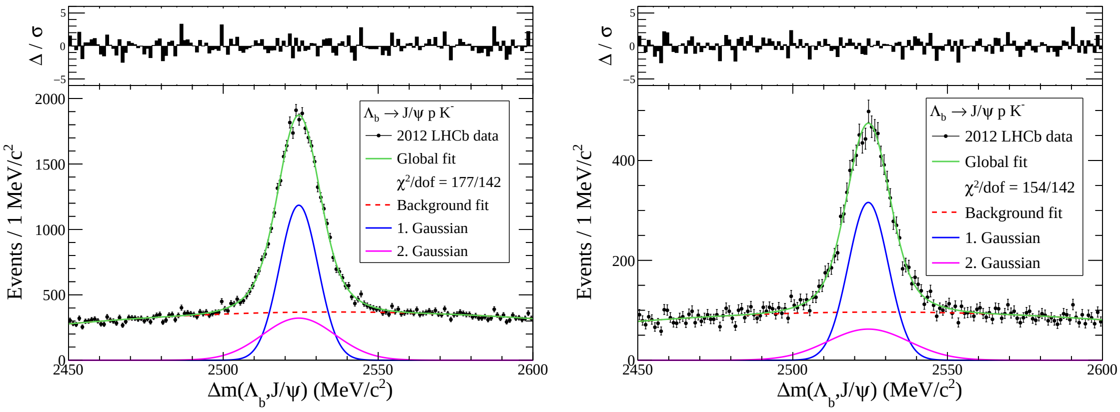
4.4 $\Lambda_b \to J/\psi pK^-$ for trigger bias studies

The decay $\Lambda_b \to J/\psi pK^-$ with $J/\psi \to \mu^+\mu^-$ was observed for the first time at LHCb in 2014 [63]. These events allow studies on differences between muon and hadron triggered events and only hadron triggered events.

For the event selection J/ψ candidates are reconstructed using two identified muon tracks with an invariant mass of $80\,\mathrm{MeV}/c^2$ around the nominal J/ψ mass. Loose requirements on the vertex and track quality are imposed as well as $p_\mathrm{T}(\mu) > 500\,\mathrm{MeV}/c^2$. The Λ_b decay is reconstructed by combining the J/ψ candidates with two well fitted tracks with proton and kaon mass hypothesis, $\chi^2(\mathrm{IP})(p) > 9$, $\chi^2(\mathrm{IP})(K) > 9$, $\mathrm{DLL}(\mathrm{p}-\pi)(\mathrm{p}) > -5$, and soft selections on the invariant mass and vertex-fit quality. The event selection is summarised in App. A.15. To suppress the large background resulting from the soft event selections additional demands are imposed: $\chi^2(\mathrm{IP}) < 25$ for the Λ_b candidate, $\chi^2(\mathrm{IP}) > 25$ for the J/ψ and the muon tracks, and $\mathrm{ProbNNk}(K) > 5\,\%$ and $\mathrm{ProbNNp}(p) > 5\,\%$. The overall selection efficiency is $\approx 8\,\%$ in simulated events.

Due to the fact that $\Lambda_b \to J/\psi pK^-$ has muons and hadrons in the final state, two different trigger categories can be studied: (1) events that require the complete set of signal triggers and (2) the subset of events in which the muon-TOS triggers are not explicitly required. Figure 4.9(a) shows the distribution of $\Delta m(\Lambda_b, J/\psi) = m(\mu^+\mu^-pK^-) - m(\mu^+\mu^-)$ and a binned likelihood fit to the sample of events requiring the complete set of signal triggers. The fit model uses a second order polynomial for the background and the sum of two Gaussian distributions with a common mean value for the signal. All parameters may float. There are $N_\mathrm{Sig} \approx 27.9\,\mathrm{k}$ signal events. Figure 4.9(b) shows the same fit to the sample of events without the muon-TOS triggers and $N_\mathrm{Sig} \approx 7.2\,\mathrm{k}$ events are found. This means that about three out of four events require the muon-TOS triggers. The mass difference $\Delta m(\Lambda_b, J/\psi)$ is used instead of $m(J/\psi pK^-)$ due to the improved signal resolution of around a factor of two.

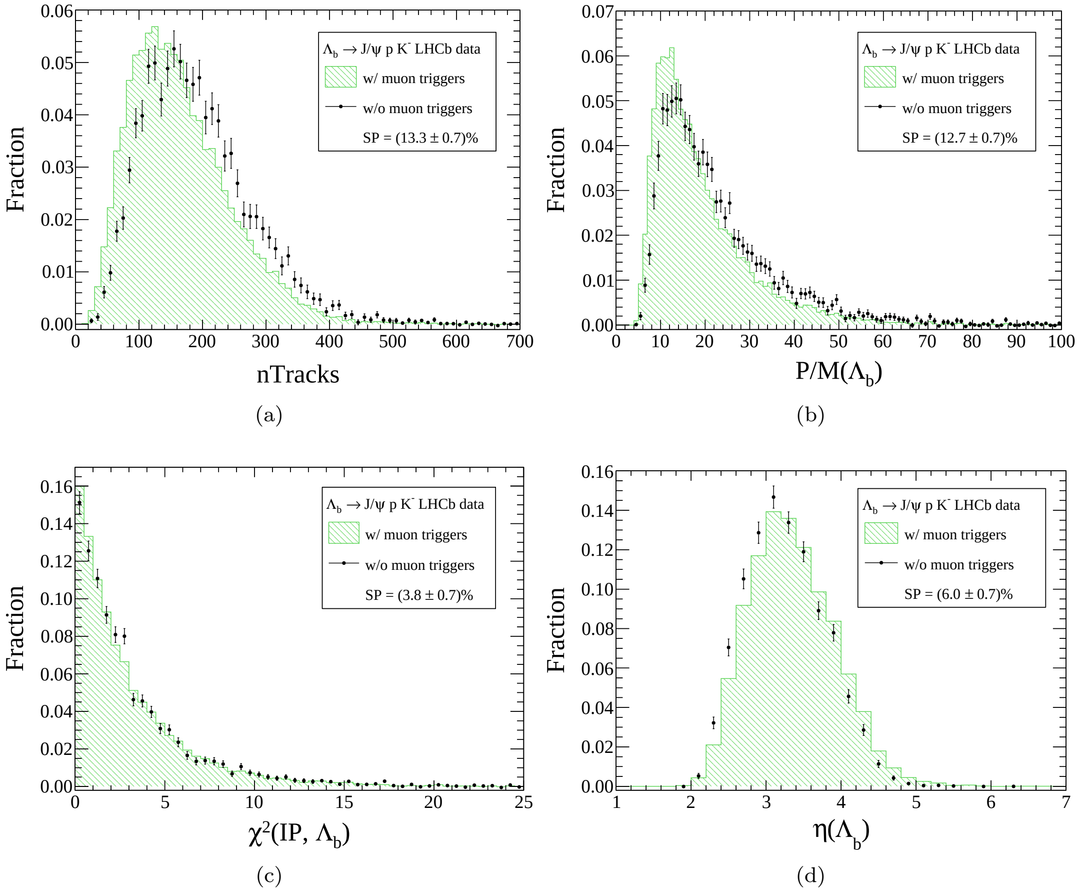
Studies in $\Lambda_b \to J/\psi pK^-$ signal MC indicate that the distributions of the Λ_b boost and the track multiplicity, that have to be calibrated using data, depend on the question if muon triggers are used or not. This behaviour can be explained by the fact that muon triggers require softer selection thresholds which results in smaller biases on event- and kinematic variables. The sample of $\Lambda_b \to J/\psi pK^-$ events from real data is used to validate these differences. Background subtracted distributions from data are obtained using the *sPlot* technique exploiting the prior fit results. Figure 4.10 shows the distributions of different variables requiring muon triggers, *i.e.* all signal triggers are used, and without explicitly demanding muon triggers. The distributions are visibly deviating for the Λ_b boost and the track multiplicity. Other variables are less affected by the different trigger conditions.



(a) all signal triggers (b) w/o requiring muon-TOS triggers

Figure 4.9: Fit to mass distribution from $\Lambda_b \to J/\psi p K^-$ events in 2012 data.

In conclusion the calibration of the Λ_b boost and the track multiplicity will be done separately for simulated events of $\Lambda_b \to K^-\mu^+$ and $\Lambda_b \to pK^-$ to consider the fact that $\Lambda_b \to pK^-$ events are only hadron triggered. Additionally, it is also necessary to optimise the signal selections for both channels separately to minimise the effects of trigger-induced biases.



(a) (b)

(c) (d)

Figure 4.10: *sPlot*'ted distributions from $\Lambda_b \to J/\psi p K^-$ events using different trigger conditions.

4.5 Track multiplicity calibration

The track multiplicity is not correctly simulated. A data-driven calibration is made using $\Lambda_b \to J/\psi pK^-$ and $\Lambda_b \to \Lambda_c\pi^-$ events from real data and signal MC to calculate a corrected track multiplicity in simulated $\Lambda_b \to K^-\mu^+$ and $\Lambda_b \to pK^-$ events. This is required for the calculation of PID selection efficiencies that are tabulated in bins of p, η, and nTracks. For the determination of a corrected track multiplicity it is distinguished between events that are mainly muon triggered: $\Lambda_b \to K^-\mu^+$ and $\Lambda_b \to J/\psi pK^-$, and only hadron triggered: $\Lambda_b \to pK^-$ and $\Lambda_b \to \Lambda_c\pi^-$. This differentiation is necessary due to the trigger-induced bias that was shown before.

To correct the track multiplicity from simulated events, a linear transformation is made that uses the mean μ and the variation σ of the nTracks distribution from reconstructed signal MC and sWeigted data, respectively. Equation 4.2 describes how nTracks is transformed into nTracks′. The parameters μ' and σ' are determined from real events and μ and σ from MC. The formula can be rewritten as linear equation with the slope A and the shift B.

$$\text{nTracks}' = \left(\frac{\text{nTracks}-\mu}{\sigma}\right)\cdot\sigma'+\mu' \quad\rightarrow\quad \text{nTracks}' = \text{nTracks}\cdot\left(\frac{\sigma'}{\sigma}\right)_{\text{A}} + \left(\mu'-\mu\cdot\frac{\sigma'}{\sigma}\right)_{\text{B}} \tag{4.2}$$

Figure 4.11 shows the calibration curves for the two Λ_b modes and for comparison the curves for $B^0 \to K^+\pi^-$ events and if MC was correct ($\text{nTracks}' = \text{nTracks}$). The selectively shown uncertainties are statistical and calculated as $\delta(\text{nTracks}') = \sqrt{\text{nTracks}^2\cdot\delta^2(\text{A})+\delta^2(\text{B})}$. In addition the track multiplicity distribution from simulated $\Lambda_b \to K^-\mu^+$ events is drawn to see where the correction is mainly applied to.

It can be seen that the track multiplicity of muon triggered events $\Lambda_b \to J/\psi pK^-$ is closer to the ideal curve and hence less biased. In general the discrepancy between MC and data is smaller for simulated Λ_b events than for $B^0 \to K^+\pi^-$ events. Therefore the small bias in DauMax[Iso5] and Iso_{CDF} that was observed in the $B^0 \to K^+\pi^-$ studies can be expected to be smaller in Λ_b events due to the more realistic simulation of the track multiplicity.

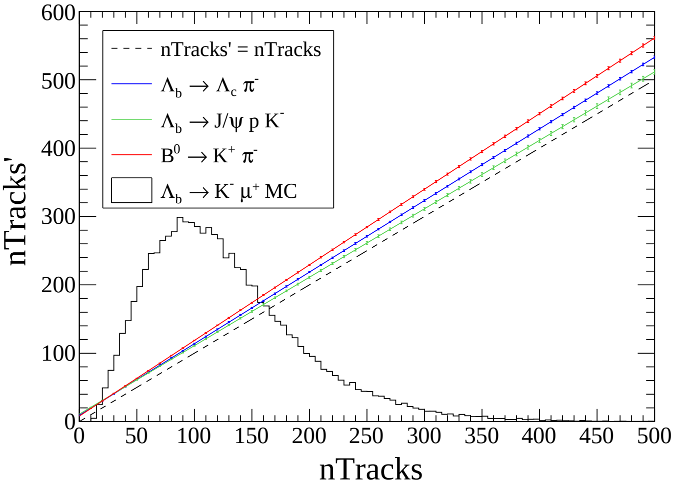
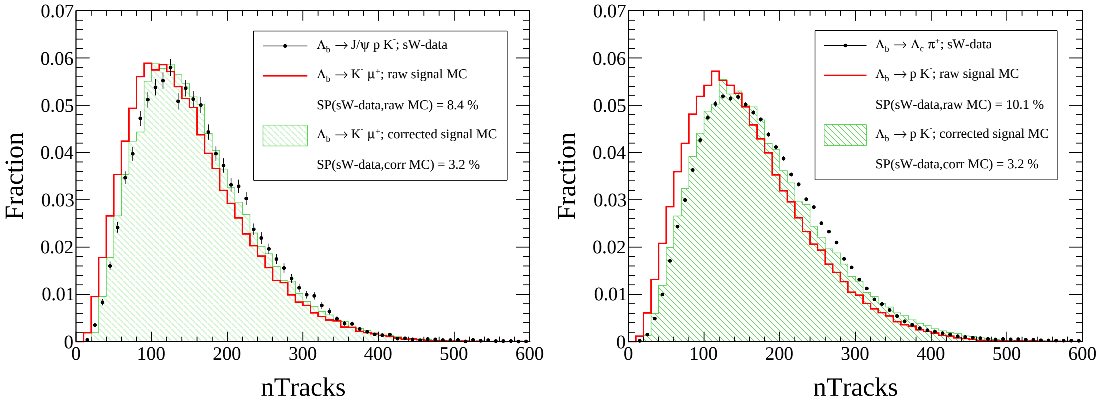


Figure 4.11: Calibration curves for the track multiplicity. The histogram shows the nTracks distribution from simulated $\Lambda_b \to K^-\mu^+$ events in arbitrary units.

The obtained parameters are for $\Lambda_b \to J/\psi pK^-$: $A = 1.001 \pm 0.007$, $B = 10.9 \pm 1.4$, for $\Lambda_b \to \Lambda_c\pi^-$: $A = 1.048\pm0.004$, $B = 9.13\pm0.70$, and for $B^0 \to K^+\pi^-$: $A = 1.106\pm0.005$, $B = 8.0\pm0.9$. The uncertainties are statistical and arise from the finite size of the used samples. Figure 4.12 compares the agreement of the uncorrected and corrected track multiplicity distribution to the sWeighted distribution of nTracks-2 from the calibration channels. Hereby it is assumed that the track multiplicity is on average the same for Λ_b events. As a result the corrected distribution are now compatible with the data.



(a) (b)

Figure 4.12: Distribution of nTracks before and after the correction.

4.6 Λ_b boost calibration

Due to the fact that the boost P/M of the Λ_b are too large in the simulation, the momenta of the daughter particles are incorrect which biases the determination of their PID efficiencies. Therefore a calibration is made using real and simulated events of $\Lambda_b \to J/\psi pK^-$ and $\Lambda_b \to \Lambda_c\pi$ that allows to calculate corrected boosts in signal MC. To conduct a correction of the momenta of a daughter particle x its original four momentum $P(x)$ is boosted into the Λ_b rest frame using the original boost, and afterwards boosted again applying the corrected value $(P/M)'$. The boost direction of the Λ_b remains unchanged which is motivated by the fact that the pseudorapidity η is correctly described in MC. In order to consider the impact of the different triggers selections, decays of $\Lambda_b \to J/\psi pK^-$ are used for the calibration of simulated $\Lambda_b \to K^-\mu^+$ events, *i.e.* in case of having muonic and hadronic triggers, and, correspondingly, events of $\Lambda_b \to \Lambda_c\pi$ are used to calibrate $\Lambda_b \to pK^-$ events, *i.e.* for events to which only hadron triggers are sensitive to. The method is tested by applying the $\Lambda_b \to \Lambda_c\pi$ calibration curve on simulated and only hadron triggered $\Lambda_b \to J/\psi pK^-$ events and make comparisons to sWeighted data.

Figure 4.13 shows the calibration curves obtained from $\Lambda_b \to J/\psi pK^-$ and $\Lambda_b \to \Lambda_c\pi$ events, and for comparison the curve from $B^0 \to K^+\pi^-$ events.

The parameters of the curve are described in Eq. 4.3. The slope A and the shift B are determined from the standard deviation σ and mean μ value of the P/M distribution from reconstructed signal MC, and, correspondingly, σ' and μ' from sWeighted data.

$$(P/M)' = P/M \cdot \left(\frac{\sigma'}{\sigma}\right)_A + \left(\mu' - \mu \cdot \frac{\sigma'}{\sigma}\right)_B \tag{4.3}$$

The determined parameters are for $\Lambda_b \to \Lambda_c\pi^-$: $A = 0.8249 \pm 0.0024$ and $B = 1.04 \pm 0.09$, for $\Lambda_b \to J/\psi pK^-$: $A = 0.813 \pm 0.005$ and $B = 0.84 \pm 0.14$, and for $B^0 \to K^+\pi^-$: $A = 1.010 \pm 0.004$ and $B = -0.08 \pm 0.10$. In conclusion there are only small differences between the two Λ_b calibration curves resulting from the different trigger conditions. The calibration curve for B^0 events is compatible with the ideal curve $(P/M)' = P/M$ due to the good description of the B momentum in simulated events.

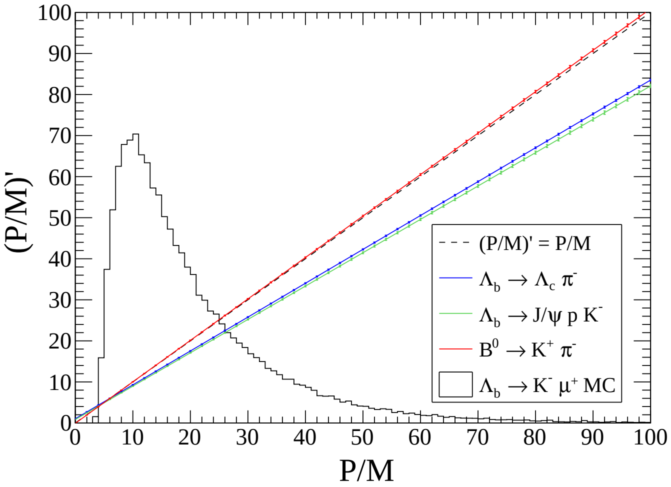


Figure 4.13: Calibration curves for the Λ_b boost. The histogram shows the P/M distribution from simulated $\Lambda_b \to K^- \mu^+$ events in arbitrary units.

Figure 4.14 shows how the $\Lambda_b \to \Lambda_c \pi$ calibration acts on simulated and only hadron triggered $\Lambda_b \to J/\psi p K^-$ events. In conclusion the corrected distribution of the Λ_b boost shows a good agreement with real data.

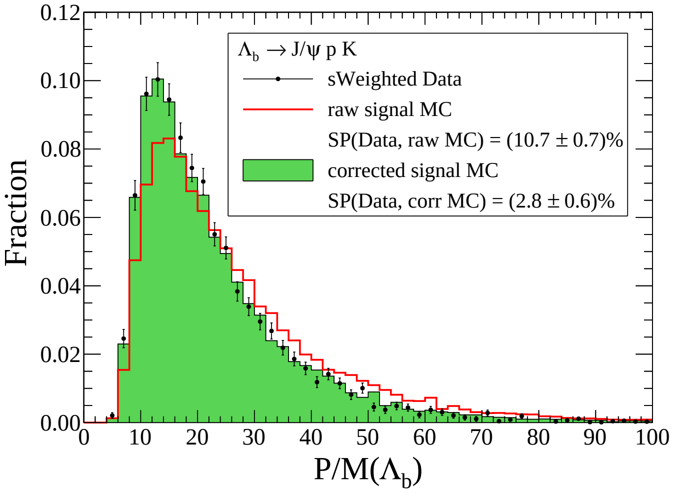


Figure 4.14: Comparison of the boost distribution from MC to data before and after the correction.

Figure 4.15 compares the simulated momentum of the kaons before and after the correction to sWeighted data. A clear improvement is visible. However, a small differences remain which can be explained by the limited number of events and the incomplete simulation of $\Lambda_b \to J/\psi p K^-$ decays that does not include observed intermediate states such as pentaquarks $P_c^+ \to J/\psi p$ and $\Lambda^* \to p K^-$ resonances [20], which results in a systematic bias of the kaon momentum distribution in MC.

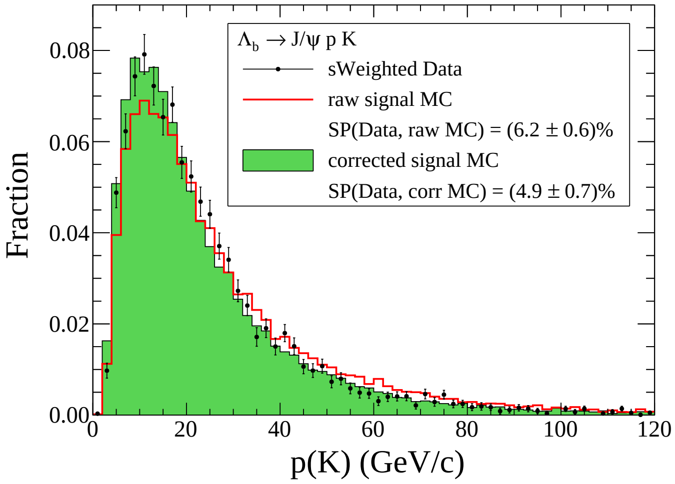


Figure 4.15: Comparison of the kaon momentum distribution from MC to data before and after the correction.

4.7 Properties of 2011 data

In this section a number of control channel studies are repeated in 2011 data. As a conclusion the same behaviour is observed as in the 2012 data in terms of the trigger simulation and data-MC agreement.

4.7.1 Mass shapes and relative b-hadron production

The sample of reconstructed and triggered $B^0 \to K^+\pi^-$ events from 2011 is used to study the compatibility of the mass spectrum w.r.t. 2012 data. Figure 4.16 shows the invariant mass distribution of the selected $K^+\pi^-$ combinations. The distribution is fitted using the same model as for the 2012 data. All parameters are fixed to the values determined there except for the normalisation $N(B^0 \to K^+\pi^-)$, the mass shift D, and the number of background events.

As a result there are $N(B^0 \to K^+\pi^-) = 66\,444 \pm 339$ events which correspond to about 40 % of the signal events found in 2012 data. The fitted mass shift is $D = (4.37 \pm 0.18)\,\mathrm{MeV}/c^2$ and compatible with the 2012 result $D = (4.13 \pm 0.17)\,\mathrm{MeV}/c^2$. The number of fitted background events corresponds to 33 % of 2012 data.

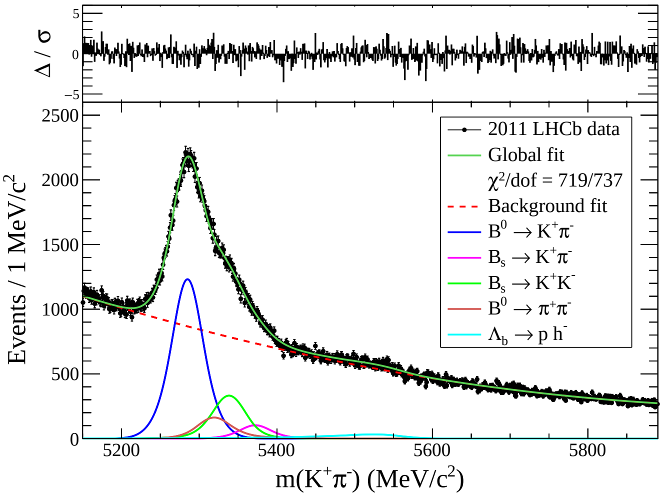


Figure 4.16: Fit to the 2011 $m(K^+\pi^-)$ distribution using the 2012 fit model with fixed parameters.

Other control channels

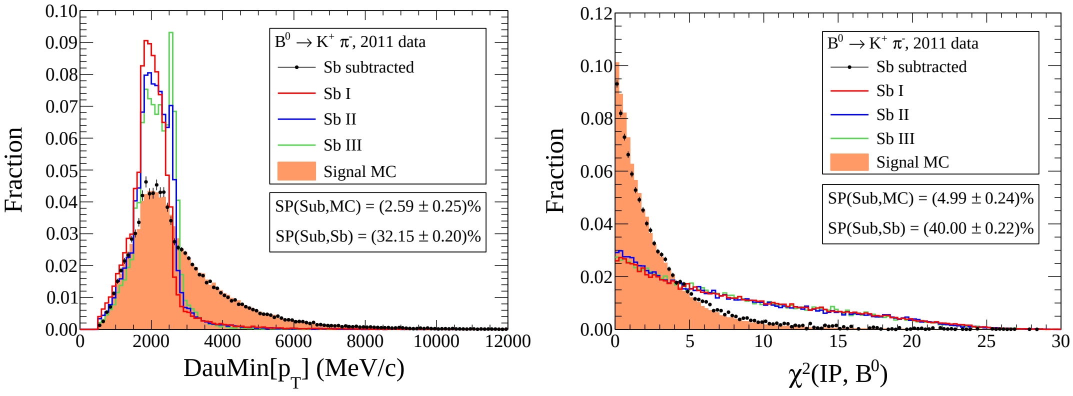
Complementary to the $B^0 \to K^+\pi^-$ studies, the mass spectra of the control channels $B^+ \to J/\psi\,(\mu^+\mu^-)K^+$, $\Lambda_b \to \Lambda_c(pK^-\pi^+)\pi^-$, and $\Lambda_b \to J/\psi\,(\mu^+\mu^-)pK^-$ are fitted to validate the agreement with 2012 data. Apart from the normalisations all parameters are fixed from the fit to the 2012 data. In summary the mass spectra of 2011 and 2012 data are compatible and the same fit model can be used. The plots of the fits can be found in App. A.16. The fraction of fitted signal (background) events relative to the 2012 data set are: 40 %(29 %) for $\Lambda_b \to \Lambda_c\pi^-$, 42 %(37 %) for $B^+ \to J/\psi\,K^+$, 42 %(38 %) for $\Lambda_b \to J/\psi\,pK^-$ both using all signal triggers and without requiring the muon-TOS triggers. Effectively, the number of observed b hadrons has increased by about 20 % per fb^{-1} in $\sqrt{s} = 8\,\mathrm{TeV}$ data which is compatible with results from another LHCb measurement [46]. In comparison the background level per fb^{-1} has grown by about 30 % in 2012 data. This can be explained by larger cross sections and a higher trigger rate and pile-up.

4.7.2 MC-Data comparisons

In analogy to the 2012 studies in Sec. 4.1.3 a sideband subtraction in the $m(K^+\pi^-)$ distribution is made to study discrepancies between $B^0 \to K^+\pi^-$ signal events in 2011 data and MC. The events from real data and signal MC, with a sample size of about 85 k entries, imply the trigger and event selections. Table 4.3 shows the SP values for different variables and in brackets the scaled expectation value of SP from signal MC that is split in two halves and compared to each other. A good agreement with a difference of less than five per cent is found for all variables except for nTracks which will be corrected as described below for the measurement of the PID selection efficiency. Figure 4.17 compares the distributions of DauMin[p_T] and χ^2(IP) in simulated events and real data. Plots of the other variables can be found in App. A.17.

Table 4.3: SP from a sideband subtraction in $B^0 \to K^+\pi^-$ data of 2011. The statistical uncertainties are around 0.3 %. The numbers is brackets are the scaled expectation values of SP.

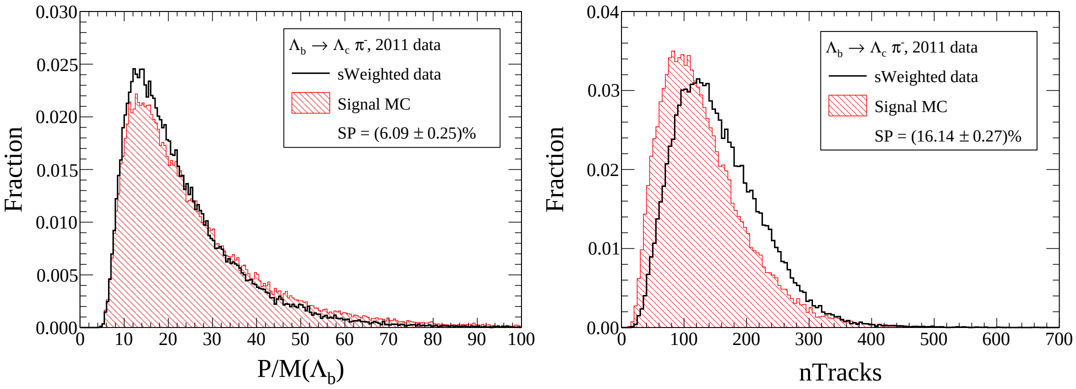
Signal variables	**SP(Sig,MC)**	**Signal variables**	**SP(Sig,MC)**
$\chi^2_{/\mathrm{DOF}}$(Vtx)	2.02 % [1.46 %]	$\eta(B^0)$	3.71 % [1.14 %]
ϕ_Dira	2.26 % [0.74 %]	FD	3.98 % [2.49 %]
DOCA	2.26 % [1.13 %]	DauMax[Iso5]	4.75 % [0.03 %]
DauMin[p_T]	2.59 % [1.54 %]	χ^2(IP)	4.99 % [1.27 %]
DauMin[IP]	2.76 % [1.99 %]	Iso$_\mathrm{CDF}$	6.14 % [1.69 %]
$p/M(B^0)$	3.61 % [1.73 %]	nTracks	12.2 % [1.43 %]



(a) (b)

Figure 4.17: Comparison of signal variables between data and signal MC.

The sample of background subtracted $\Lambda_b \to \Lambda_c\pi$ events is compared to signal MC in projections of the Λ_b IP significance, flight distance, pseudorapidity, proper lifetime, boost, and track multiplicity per event. The first three distributions, that are used in the multivariate analysis, show a good agreement and the SP values are below 5 %, respectively. As in the studies of the 2012 data the proper lifetime is slightly biased in signal MC due to the smaller average Λ_b lifetime in the simulation. All corresponding plots can be found in App. A.18. Figures 4.18(a) and (b) compare the distributions of the Λ_b boost and nTracks. As for the 2012 data sample both distributions are biased in simulated events and require a calibration.



(a) Λ_b boost (b) nTracks

Figure 4.18: Comparison of 2011 data and signal MC for $\Lambda_b \to \Lambda_c\pi^-$ events.

4.7.3 Trigger studies

The TISTOS method is applied on the sample of $B^+ \to J/\psi K^+$ events to validate the accuracy of the used trigger lines with simulated events. Table 4.4 shows the results. In the data there are around 100 k L0-TIS events, 43 k HLT1-TIS events, and 12 k Hlt2-TIS events. The binomial uncertainties on the efficiencies are $< 0.1\,\%$ for simulated events and $0.1 - 0.4\,\%$ for real data. In comparison to the results from 2012, see Sec. 4.2.1, there is a consistent overestimation of the L0-TOS efficiencies in signal MC. The efficiencies of the HLT lines are generally in better agreement. However, the two Muon-TOS lines are slightly overestimated in simulated events.

Table 4.4: Trigger line efficiencies in 2011 data and signal MC.

Trigger	$\varepsilon_{\text{Data}}$	ε_{MC}	$\varepsilon_{\text{MC}}/\varepsilon_{\text{Data}} - 1$
L0Hadron-TOS	14.6 %	16.1 %	10.6 %
L0Muon-TOS	89.6 %	95.8 %	6.9 %
Combination	90.5 %	97.0 %	7.2 %
AllTrack-TOS	85.0 %	86.5 %	1.8 %
MuonTrack-TOS	86.8 %	90.4 %	4.1 %
Combination	93.2 %	95.5 %	2.5 %
Topo2Body-TOS	77.4 %	77.6 %	0.2 %
TopoMu2Body-TOS	83.1 %	85.2 %	2.6 %
Combination	83.1 %	85.2 %	2.6 %

4.7.4 MC Calibration

The calibration curves for the track multiplicity and the Λ_b boost are obtained using 66.4 k $B^0 \to K^+\pi^-$ events, 72.6 k $\Lambda_b \to \Lambda_c\pi^-$ events, and 11.8 k $\Lambda_b \to J/\psi pK^-$ from 2011 data, respectively. All results are obtained using the same method as for the 2012 data, see Sec. 4.5 and 4.6, and are slightly different while the qualitative findings are the same.

Figure 4.19 shows the calibration curves for the track multiplicity. There is a visible bias for the purely hadron-triggered decays $B^0 \to K^+\pi^-$ and $\Lambda_b \to \Lambda_c\pi^-$ while the muon-triggered event class $\Lambda_b \to J/\psi pK^-$ fits better to the ideal curve. The determined parameters are $A = 1.012 \pm 0.004$ and $B = 22.2 \pm 0.7$ for $\Lambda_b \to \Lambda_c\pi^-$, $A = 0.951 \pm 0.009$ and $B = 14.6 \pm 1.5$ for $\Lambda_b \to J/\psi pK^-$, and $A = 1.057 \pm 0.006$ and $B = 12.3 \pm 0.9$ for $B^0 \to K^+\pi^-$.

In Fig. 4.20 the calibration curves for the Λ_b boost can be seen. The parameters are $A = 0.841 \pm 0.004$ and $B = 1.68 \pm 0.11$ for $\Lambda_b \to \Lambda_c\pi^-$, $A = 0.822 \pm 0.007$ and $B = 1.2 \pm 0.2$ for $\Lambda_b \to J/\psi pK^-$, and $A = 1.012 \pm 0.005$ and $B = 0.35 \pm 0.13$ for $B^0 \to K^+\pi^-$. The uncertainties are purely statistical and shown for selected points in the figure.

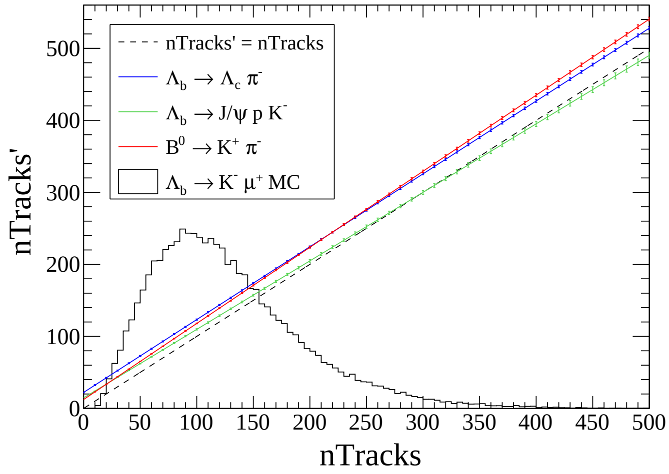


Figure 4.19: Calibration curves for the track multiplicity from 2011 data. The histogram shows the nTracks distribution from simulated $\Lambda_b \to K^-\mu^+$ events in arbitrary units.

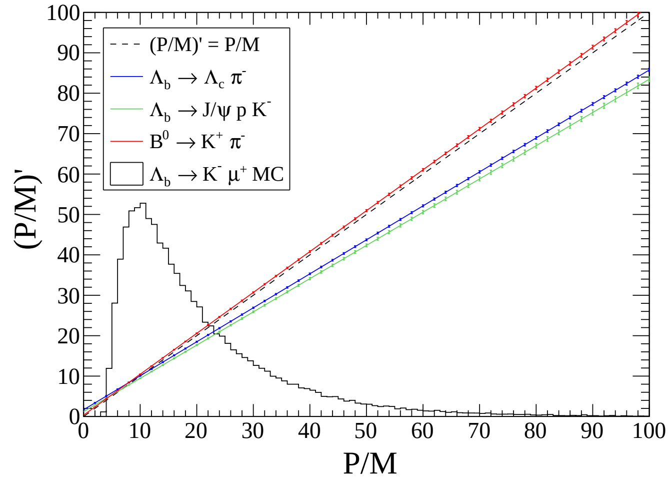


Figure 4.20: Calibration curves for the Λ_b boost from 2011 data. The histogram shows the P/M distribution from simulated $\Lambda_b \to K^-\mu^+$ events in arbitrary units.

"Seek and you will find"

— The Bible, Matthew 7:7

5

Signal selection

This chapter describes the development of a final event selection to maximise the sensitivity in the search for the decay $\Lambda_b \to K^-\mu^+$. The major role in this process plays a machine-learning algorithm that is trained to classify $K^-\mu^+$ combinations as more signal- or more background-like. Hereby it is made use of the statistically independent WS data sample to mimic the combinatorial background from random $K^-\mu^+$ pairs.

5.1 Comparison of RS and WS events

The properties of $K^+\mu^+$ and $K^-\mu^+$ pairs are compared to justify the use of WS data as a model for random $K^-\mu^+$ pairs. The underlying data set are $K^-\mu^+$ and $K^+\mu^+$ combinations from 2012 satisfying the same event and trigger selections. This sample is blinded in the Λ_b and Ξ_b^0 mass region of $(5625 \pm 50)\,\mathrm{MeV}/c^2$ and $(5795 \pm 50)\,\mathrm{MeV}/c^2$ including a shift of $+5\,\mathrm{MeV}/c^2$. The shift is rounded up by about $0.8\,\mathrm{MeV}/c^2$ to an integer value. Since the $5\,\mathrm{MeV}/c^2$ binning of the mass histograms is larger, it is assumed that the rounding has no relevant influence on fits to the distributions.

The available data are divided into three sidebands.

- Lower sideband: $m(K^-\mu^+) = 5120 \ldots 5575\,\mathrm{MeV}/c^2$
- Middle sideband: $m(K^-\mu^+) = 5675 \ldots 5745\,\mathrm{MeV}/c^2$
- Upper sideband: $m(K^-\mu^+) = 5845 \ldots 6120\,\mathrm{MeV}/c^2$

Figure 5.1 shows the invariant mass spectrum of the blinded RS data and the entire WS data that is normalised to the middle and upper sidebands which are expected to be completely dominated by combinatorial background. In these ranges there are 14.7 % more RS events. As a result there is a good agreement between both shapes except for the lower sideband where hints for contributions from $B_{(s)} \to h^+h^-$ events can be seen around $m(K^-\mu^+) = 5300\,\mathrm{MeV}/c^2$.

From the sample of WS and RS events that lie in the middle or upper sideband, the distributions of a number of event variables are compared. Table 5.1 summarises the corresponding SP values. In summary a good agreement is found for all variables. Comparisons of the distributions are presented in App. A.19. The correlation of the event variables with the invariant mass in both samples is in agreement and small, see Fig. A.22 in App. A.19. Similar studies in 2011 data confirm the agreement between RS and WS data, see Tab. A.10 in App. A.19, though the consistence in general is slightly worse.

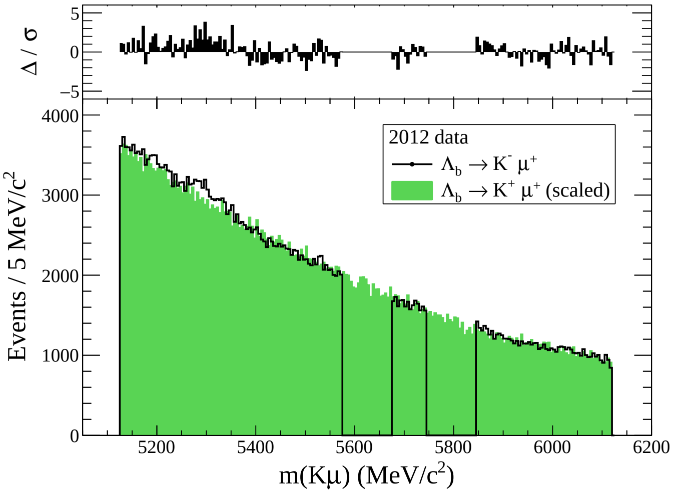


Figure 5.1: Comparison of the invariant mass spectrum of $K^-\mu^+$ and $K^+\mu^+$ pairs.

Table 5.1: SP values comparing RS and WS data from 2012. The statistical uncertainties are smaller than 0.25 %, respectively.

Signal variables	**SP(RS,WS)**	**Signal variables**	**SP(RS,WS)**
DauMin[p_T]	1.19 %	χ^2(IP)	1.98 %
DauMax[Iso5]	1.29 %	χ^2(FD)	2.33 %
$p_\mathrm{T}(\Lambda_b)$	1.49 %	$\mathrm{Iso}_\mathrm{CDF}$	2.34 %
$\chi^2_{/\mathrm{DOF}}$(Vtx)	1.69 %	P/M	2.35 %
η	1.72 %	DauMin[χ^2(IP)]	2.57 %
IP	1.73 %	FD	2.67 %
DOCA	1.78 %	DauMin[IP]	2.96 %
ϕ_Dira	1.80 %	τ	3.05 %
nTracks	1.95 %		

5.2 Multivariate event classification

For the multivariate analysis the software package TMVA (version 4.2.0 with Root 5.34/24) and its method BDT (Boosted Decision Tree with *adaptive* boost algorithm) is applied, which was also used in other searches for rare decays, *e.g.* $B^0 \to p\overline{p}$ and $B^0_s \to \mu^+\mu^-$, at LHCb.

5.2.1 Working principle of a BDT

A BDT is a machine-learning algorithm that uses two input samples; one consisting of signal events and the other one of background events. The BDT compares the distributions of a number of certain signal variables and their correlations. To achieve a precise separation of both event classes the distributions should differ as much as possible. During the training phase a tree-like structure of one-dimensional splits (nodes) is created for each variable. At each node a threshold value is defined that decides if the event is more background or signal like. The boosting is a method that mixes and reweights events differently to train many trees and finally develops a classifier as an average of these trees. After the training events are classified by traversing the tree of nodes in dependence of the values of the used variables. Finally, the quality of being more signal- or more background like is quantified by an output variable $\mathrm{x_{BDT}} \in (-1, 1)$ that is used for selections. More information about BDTs can be found, for example, in Refs. [64,65].

The behaviour of a BDT classifier is defined by the input samples and the event variables that are used for the training, and tuning parameters that can be varied to improve the performance. These are namely: number of trees (NTrees), number of cuts (NCuts), the maximum depth (Depth) and the minimal fraction of used events per node (NodeSize). After the training, the BDT is applied on a statistically independent test sample to detect biases (overtraining) and determine the signal and background efficiencies. The used input samples for the training are the complete sample of triggered WS events and one half of the sample of correctly reconstructed and triggered signal MC events. For the testing the middle and upper sideband of triggered RS data is used and the second half of the signal MC sample. The WS data have a sample size of ≈ 350 000 events, the RS data consist of about 85 000 events, and the used sample of simulated events comprises about 60 000 events.

To assess the performance of the BDT three efficiencies are measured that focus on the rejection power of RS events in the middle and upper sideband: $\varepsilon_{1\%}$: the signal efficiency when the background level is reduced to 1 %, ε_{100}: the signal efficiency when the background level is reduced to 100 events, and ε_{50}: the signal efficiency when the background level is reduced to 50 events. The primary performance number is $\varepsilon_{1\%}$ and the other two are taken as a cross check.

5.2.2 BDT optimisation

After preliminary studies with different BDT tuning parameters and variables, the following choice is made for a reference BDT: (NTrees/NCuts/Depth/NodeSize) = (1000/30/2/5 %), and ten variables: χ^2(IP), FD, η, $\chi^2_{/\mathrm{DOF}}$(Vtx), $\mathrm{Iso_{CDF}}$, ϕ_{Dira}, DOCA, DauMin[IP], DauMin[p_{T}], and DauMax[Iso5]. The resulting performance is $(\varepsilon_{1\%}/\varepsilon_{100}/\varepsilon_{50}) = (73.4\,\%/51.4\,\%/45.5\,\%)$.

Figure 5.2 shows the distribution of the BDT output variable $\mathrm{x_{BDT}}$ for the training and test samples. For a consistent comparison of the background samples only events from the mass range of the middle and upper sideband are considered in the training and test sample. As a result there is a good agreement between the training and test samples which is quantified by an unbinned Kolmogorov-Smirnoff (KS) test of the $\mathrm{x_{BDT}}$ variable.

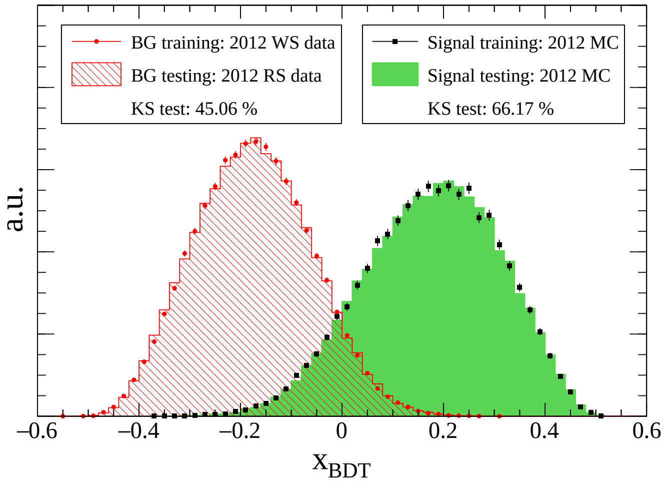


Figure 5.2: $\mathrm{x_{BDT}}$ distribution for the reference BDT.

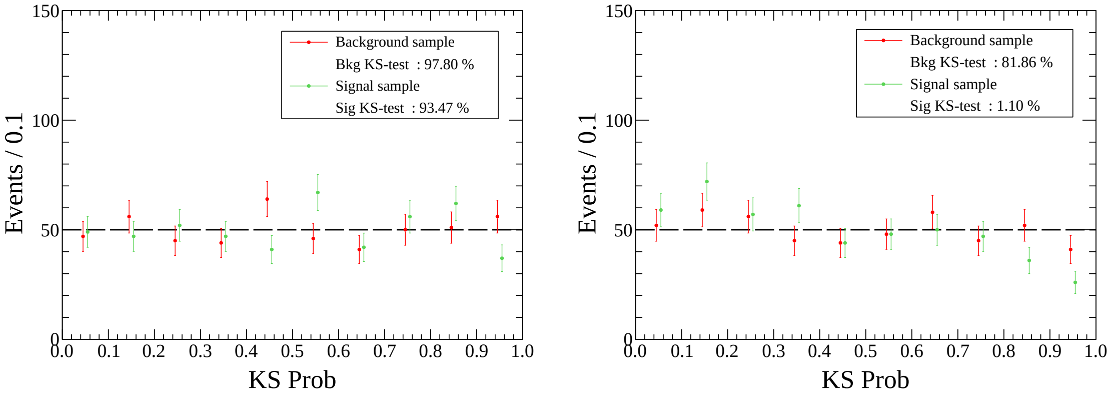
The BDT performance was systematically checked by varying the set of used event variables, tuning parameters and by comparing to other multivariate analysis methods. These studies are summarised in App. A.20. It is found that the reference BDT has an optimal performance that cannot be improved by choosing different signal variables or tuning parameters. It has also the best performance among other BDT methods.

5.2.3 Overtraining

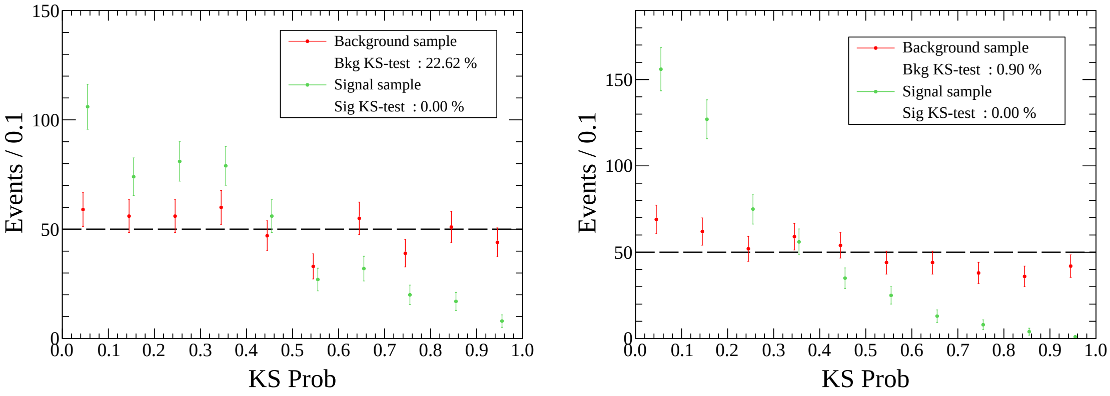
The overtraining of the used BDT is checked according to a specially developed method described in App. B. The basic idea of the method is to train and test the BDT on many different partitions of the input samples. Generally speaking a partition of a set is a grouping of the set's elements into non-empty and disjoint subsets. In the following only partitions of two equally sized subsets are of interest. One subset represents the training sample and the other one is for testing. A set consisting of $2N$ events can be divided into $\binom{2N}{N} \approx 2^{2N}/\sqrt{\pi \cdot N}$ possible partitions with equal size and using Stirling's approximation for the factorials. Hence there is an almost infinite number of partitions for the given input samples. The behaviour of two BDTs that are trained on randomly chosen partitions of the same input sample can only differ statistically. To check for biases a KS test is made comparing the $\mathrm{x}_{\mathrm{BDT}}$ distribution of the training and the test samples for 500 partitions. For each partition the probability of a positive KS test is determined. An unbiased BDT must have a uniform probability distribution while a biased BDT has not. The most relevant tuning parameter is the maximum depth of the tree which is illustrated in Figure 5.3. The other BDT tuning parameters are the same as for the reference BDT. If the Depth is two the KS distribution for both samples is compatible with being uniform. For a Depth of three, four, and five the KS distribution of the signal sample becomes more and more biased while the distribution of the background sample is robust and deforms only slightly. The different behaviour can be explained by the different sizes of the input samples; the signal sample is about four times smaller than the background sample and hence the BDT starts earlier to train on statistical fluctuations which results in a bias. In conclusion the reference BDT is not overtrained. Similar checks with other parameters shows that they have no measurable influence on the overtraining behaviour for values around the reference one.

Full training samples

Since the reference BDT shows no signs of overtraining, *i.e.* the performance is the same on the test and training sample, from now on a BDT is applied that uses the complete signal MC sample for training. This is motivated by the fact that performance will improve and the impact of statistical fluctuations is reduced due to the two times larger amount of training events. Figure 5.4 shows a comparison of the Receiver Operating Characteristic (ROC) curves where the signal efficiency $\varepsilon_{\mathrm{sig}}$ in signal MC is plotted against the background rejection efficiency $1 - \varepsilon_{\mathrm{bg}}$ in the sample of RS events from the middle and upper sideband. In the region $\varepsilon_{\mathrm{sig}} > 50\,\%$ the performance of the BDT using the full test sample is slightly better. In the region with a low background level, *i.e.* $\varepsilon_{\mathrm{sig}} < 50\,\%$, the ROC curves suffer from statistical fluctuations due to the small number (< 100) of background events. The integral over the ROC curve is approximately $98\,\%$.



(a) Depth = 2 (b) Depth = 3



(c) Depth = 4 (d) Depth = 5

Figure 5.3: BDT overtraining in dependence of the Depth parameter.

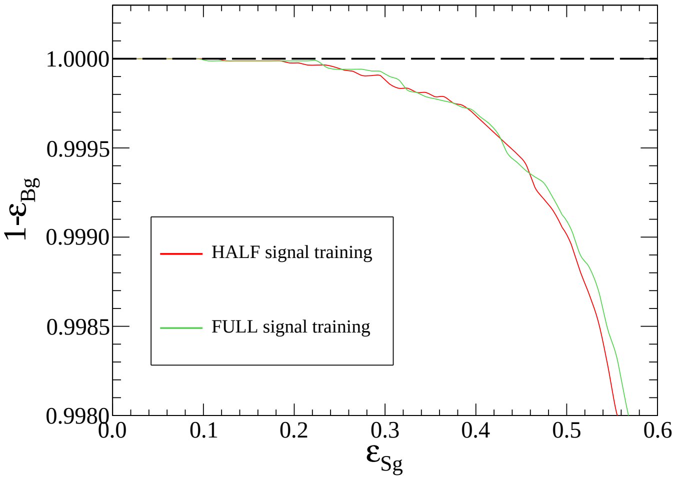


Figure 5.4: Variation of the efficiencies in the WS test samples.

5.2.4 Influence of the BDT classifier on the mass spectra

The influence of the BDT classifier is assessed on signal and background events to check for biases in the invariant mass distribution due to BDT selections.

Signal MC events

Figures 5.5 and 5.6 compare the invariant mass spectrum of truth-matched and triggered MC events of $\Lambda_b \to K^-\mu^+$ and $B^0 \to K^+\pi^-$, respectively, before and after a tight selection of $\mathrm{x_{BDT}} > 0.3$ ($\varepsilon_{\mathrm{sig}} \approx 30\,\%$). No significant change of the signal shape is observed. In conclusion the shape of the invariant mass distribution of signal events is not affected by a BDT selection.

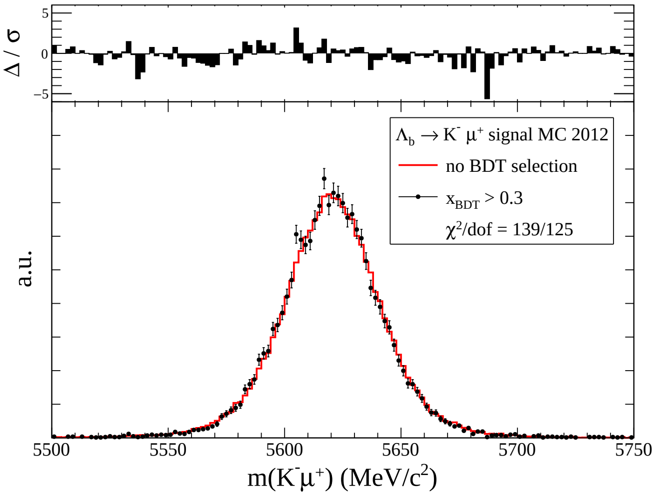


Figure 5.5: Influence of the BDT classifier on the $\Lambda_b \to K^-\mu^+$ mass spectrum.

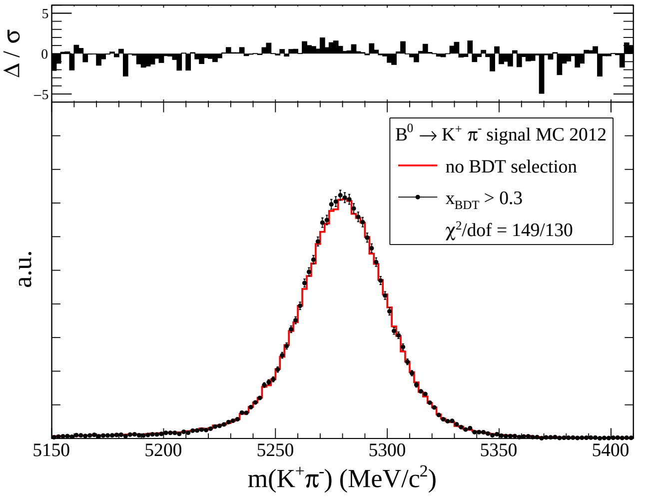


Figure 5.6: Influence of the BDT classifier on the $B^0 \to K^+\pi^-$ mass spectrum.

$K^+\mu^+$ background events

Figure 5.7 compares the invariant mass spectrum of $K^+\mu^+$ events from 2012 data after trigger selections and requiring different $\mathrm{x_{BDT}}$ selections starting from $\mathrm{x_{BDT}} > -0.6$ ($\varepsilon_{\mathrm{bg}} \approx 100\,\%$) to

$x_{BDT} > 0.1$ ($\varepsilon_{bg} \approx 1\,\%$). Each distribution is fitted using a second order polynomial where the fit parameter p_2 belongs to the quadratic term. The insert histogram shows the significance $p_2/\delta(p_2)$ from the fit in dependence of the BDT selection. It can be concluded that the curvature of the background shape is negligible for selections $x_{BDT} > 0.0$ and the background becomes linear. This behaviour can be explained by the fact that the BDT classifier uses the variable DauMin[p_T] that is slightly correlated with the invariant mass ($\rho \approx 19\,\%$, see Fig. A.22).

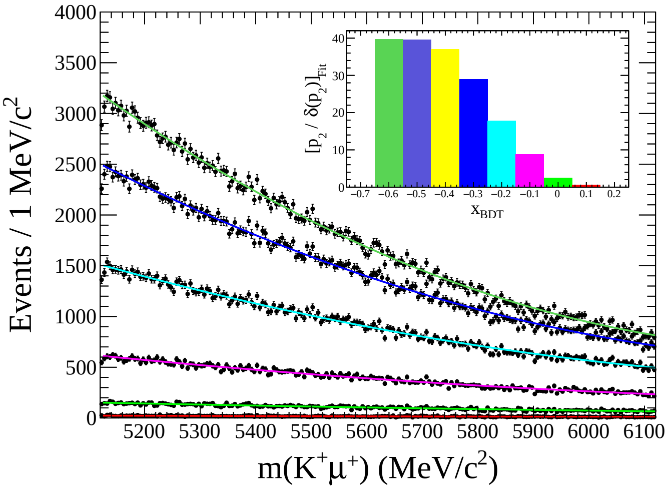


Figure 5.7: Influence of the BDT classifier on the $K^+\mu^+$ mass spectrum.

5.2.5 BDT classifier for 2011 data

In the 2011 data there are available about 67 k (110 % of 2012) signal MC events, 27 k (32 %) $K^-\mu^+$ events in the middle and upper sidebands, and 114 k (33 %) $K^+\mu^+$ events for training and testing. As a first cross check the developed BDT using 2012 data is applied on the WS data and signal MC from 2011. Figure 5.8 shows the comparison of the respective BDT distributions. In conclusion there are differences between both samples that lead to a negative KS test. As a result a separate BDT is trained to avoid systematic biases and possibly improve the sensitivity.

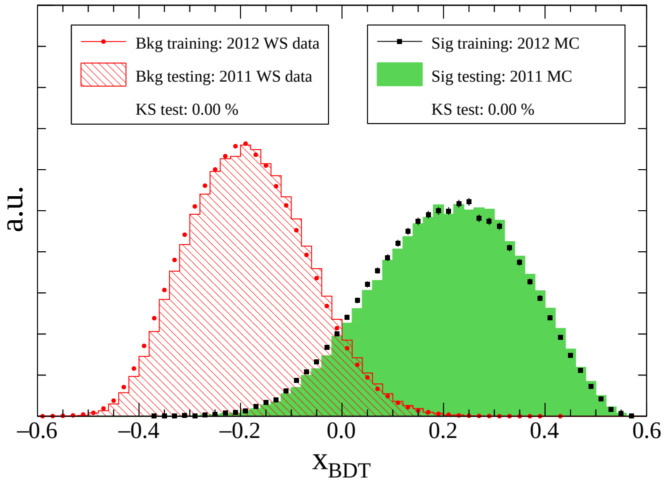
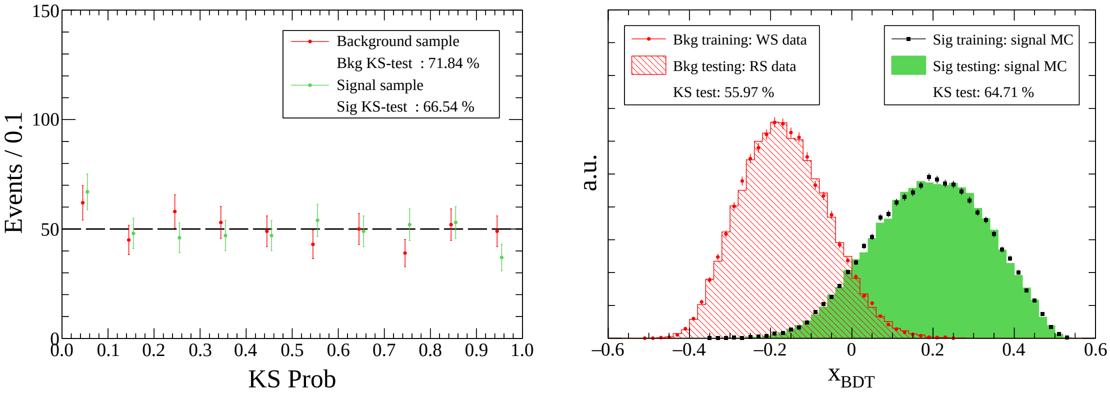


Figure 5.8: BDT response comparison using 2012 data for training and 2011 data for testing.

The BDT for the 2011 data is developed using the same tuning parameters and signal variables as the 2012 version. Figure 5.9(a) shows the result of the overtraining test from 500 random partitions of WS data and signal MC. No indication for overtraining is observed. Figure 5.9(b) compares the BDT responses for an arbitrarily chosen partition. In summary a good agreement is found and hence the finally applied BDT also exploits the full 2011 data for training.



(a) (b)

Figure 5.9: Overtraining tests for the 2011 BDT.

5.3 Optimisation of the $\Lambda_b \to K^-\mu^+$ signal selection

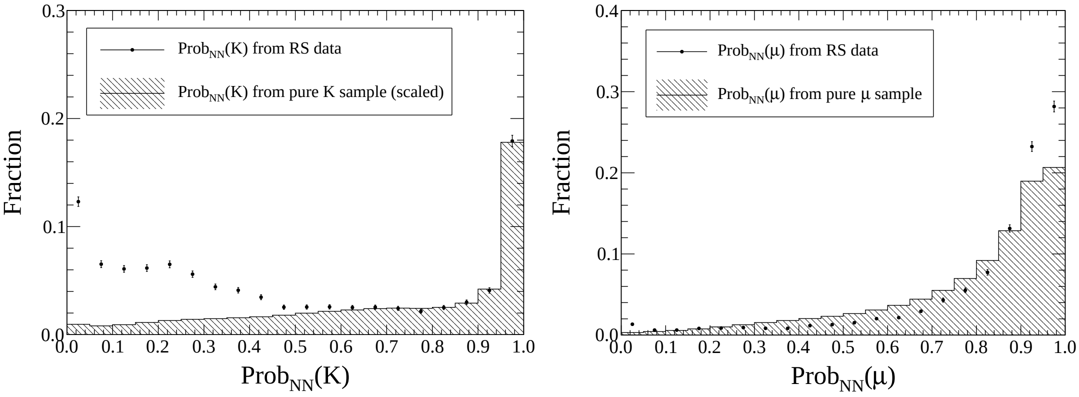
For the optimisation of the signal selection the statistical significance for an upper limit $S/\sqrt{B}$ is considered as figure of merit (FoM), where S is the number of signal events and B the expected number of background events in the signal range resulting from a fit to the $m(K^-\mu^+)$ spectrum. Since S is unknown, it is replaced by the signal efficiency $\varepsilon_{\mathrm{sel}}$ from the sample of simulated events. It is assumed that S and $\varepsilon_{\mathrm{sel}}$ only differ by the unknown but constant factor S_0, *i.e.* $\varepsilon_{\mathrm{sel}} = S/S_0$, denoting the number of signal events in real data after event and trigger selections. Hence the optimum of $S/\sqrt{B}$ would be the same as of $\varepsilon_{\mathrm{sel}}/\sqrt{B}$. Thus the FoM that is optimised for the final signal selection is $\varepsilon_{\mathrm{sig}}/\sqrt{B}$. As a cross check for selections with large uncertainties $\delta(B)$ the value $\mathrm{FoM}' = \varepsilon_{\mathrm{sig}}/\sqrt{B + \delta(B)}$ is considered to assess the impact of statistical fluctuations.

In the following the FoM is optimised in dependence of the BDT variable $\mathrm{x_{BDT}}$ and the PID variables ProbNN(K) and ProbNN(μ) separately in 2011 and 2012 data. In all cases the signal triggers are required.

5.3.1 Preliminary PID studies

To assess the purity of the kaons and muons in the middle and upper sideband, the distributions of the ProbNN variables are compared to the pure sample of K and μ from $B^+ \to J/\psi K^+$ events of 2012. For these kaons and muons the same PID selections are required as for the $K^-\mu^+$ combinations, namely IsMuon(μ)=true and DLL(K − π)(K) $>$ 0. The $B^+ \to J/\psi K^+$ events are selected in the range $\Delta m(B^+, J/\psi) = (2155\ldots2215)\,\mathrm{MeV}/c^2$ and $m(\mu^+\mu^-) = m(J/\psi) \pm 50\,\mathrm{MeV}/c^2$, and have a purity $S/(S+B)$ of about 98 %. The selected kaons and muons from the sidebands require $\mathrm{x_{BDT}} > 0$ and comprise about 6400 events. Figure 5.10 shows the distributions of the ProbNN variables. As a result the muon sample is clean while there is a considerable background from misidentified tracks as kaons in the range ProbNN(K)$\lesssim$ 0.5. Both variables are largely uncorrelated with a correlation coefficient of $\rho = 0.7\,\%$. The purity of the muons can be explained the low misidentification rates $\varepsilon(\pi \to \mu) \approx 0.9\,\%$ and $\varepsilon(K \to \mu) \approx 1.7\,\%$ measured in the kinematically similar decay $B^0 \to K^+\pi^-$, see App. A.21.

In conclusion the FoM is mainly sensitive to the variables $\mathrm{x_{BDT}}$ and ProbNN(K) to suppress background from random $K^-\mu^+$ combinations and misidentified kaon tracks. The variable ProbNN(μ) can be expected to play only a minor role.



(a) (b)

Figure 5.10: Comparison of ProbNN variables from combinatorial background in the middle and upper sideband of RS data and $B^+ \to J/\psi K^+$ events of 2012.

5.3.2 Figure of merit optimisation

The optimisation of the FoM $= \varepsilon_{\mathrm{sig}}/\sqrt{B}$ is done in dependence of $\mathrm{x_{BDT}}$ and ProbNN(K) and without additional selections on the muons due to their great purity. The signal efficiency is determined as

$$\varepsilon_{\mathrm{sig}} = \varepsilon_{\mathrm{BDT}} \cdot \varepsilon_{\mathrm{PID}}(K) = \frac{N_{\mathrm{BDT}}}{N_{\mathrm{Presel}}} \cdot \varepsilon_{\mathrm{PID}}(K) \;, \tag{5.1}$$

where N_{Presel} is the number of truth-matched $\Lambda_b \to K^-\mu^+$ events from signal MC after event and trigger selections, N_{BDT} is the number of remaining events after the selection in $\mathrm{x_{BDT}}$, and $\varepsilon_{\mathrm{PID(K)}}$ the efficiency of the respective ProbNN(K) requirement.

The value of $\varepsilon_{\mathrm{PID(K)}}$ is determined using the efficiencies from the PID tables and the number of truth-matched events from signal MC in bins of $\eta(K)$, $p(K)$ and nTracks. Hereby, the momentum of the kaons and the track multiplicity per event are corrected as described in Sec. 4.5 and 4.6. The uncertainty of $\varepsilon_{\mathrm{BDT}}$ is determined as binomial uncertainty $\delta\varepsilon_{\mathrm{BDT}} = \sqrt{\varepsilon_{\mathrm{BDT}}\,(1-\varepsilon_{\mathrm{BDT}})/N_{\mathrm{Presel}}}$ and $\delta\varepsilon_{\mathrm{PID(K)}}$ is determined together with $\varepsilon_{\mathrm{PID(K)}}$. The total uncertainty $\delta\varepsilon_{\mathrm{sig}}$ results from quadratic error combination.

The number of background events in the Λ_b signal region is determined as

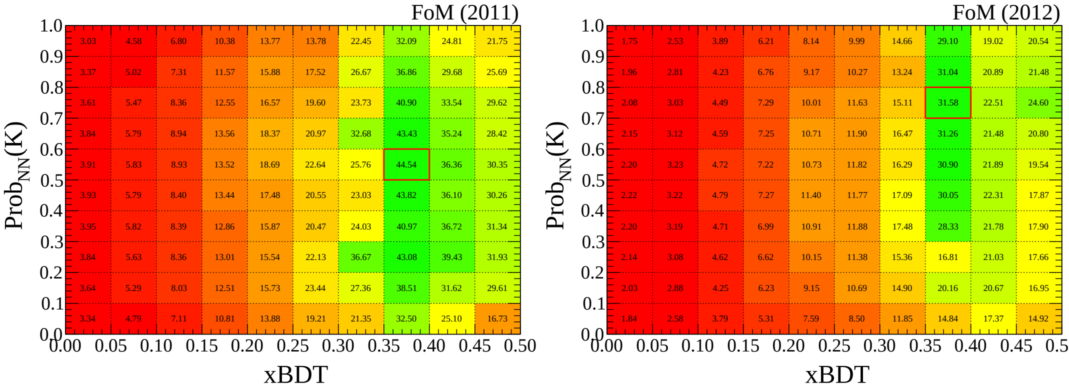
$$B = \sum_i f_i \cdot N_i + N_{\mathrm{bg}} \cdot \int\limits_{5575\,\mathrm{MeV}/c^2}^{5675\,\mathrm{MeV}/c^2} g_{\mathrm{bg}}(m)dm \;, \tag{5.2}$$

from a fit to the invariant mass spectrum of $K^-\mu^+$ using the modes $i \in \{\Lambda_b \to p\mu^-\overline{\nu}_\mu, \Lambda_b \to pK^-, \Lambda_b \to p\pi^-\}$. The variable f_i is the fraction of the respective fit model in the Λ_b signal region, N_i is the normalisation, and $g_{\mathrm{bg}}(m)$ the normalised fit function describing the combinatorial background. The complete fit model is explained in detail in App. A.23. It includes a linear function for the combinatorial background and several signal components to describe reflections from other b hadron decays whose individual shapes have been determined using simulated events.

The uncertainty $\delta(B)$ is determined from error propagation (see App. A.22) of Eq. 5.2 taking into account the covariance matrix of the fit parameters. The determination of B and $\varepsilon_{\mathrm{PID}}$ is done in steps of $\Delta \mathrm{x_{BDT}} = 0.05$ and $\Delta\mathrm{ProbNN}(K) = 0.10$. This binning is a compromise between precision and to avoid optimisation on statistical fluctuations.

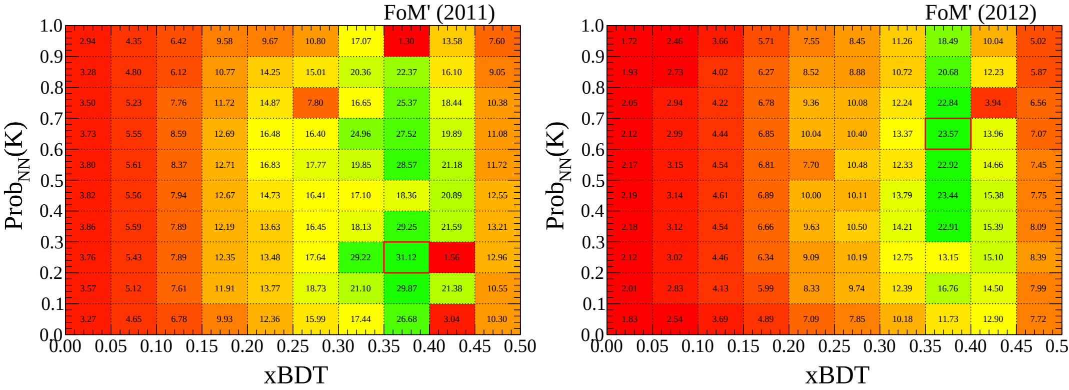
Results

Figures 5.11 and 5.12 show the FoM and FoM′ distribution of 2011 and 2012 data. In both years the optimal selection is achieved for $\mathrm{x_{BDT}} > 0.35$ while the optimal ProbNN(K) selection varies. The significance in 2011 data is generally higher due to the better signal-to-background ratio which can be explained by the looser trigger selections in 2012 data. For 2011 data the optimal selection is in the range $\mathrm{ProbNN}(K) > 0.2 \ldots 0.6$ and for 2012 data in the range $\mathrm{ProbNN}(K) > 0.4 \ldots 0.8$. However, the precision of the FoM suffers from the large uncertainty on B which is in the order $\delta(B)/B \approx 50 \ldots 170\,\%$ in the optimal range. Hence a unique optimum cannot be found and for the sake of simplicity a common selection of $\mathrm{x_{BDT}} > 0.35$ and $\mathrm{ProbNN}(K) > 0.4$ is chosen that is compatible with the results from both years. The anticipated value of B from a fit to the combined $m(K^-\mu^+)$ spectrum is 0.36 ± 0.17, *i.e.* there are most likely no background events in the signal region. The signal efficiency according to simulated $\Lambda_b \to K^-\mu^+$ events is around $15\,\%$.



(a) (b)

Figure 5.11: Distribution of FoM in dependence of ProbNN(K) and $\mathrm{x_{BDT}}$.



(a) (b)

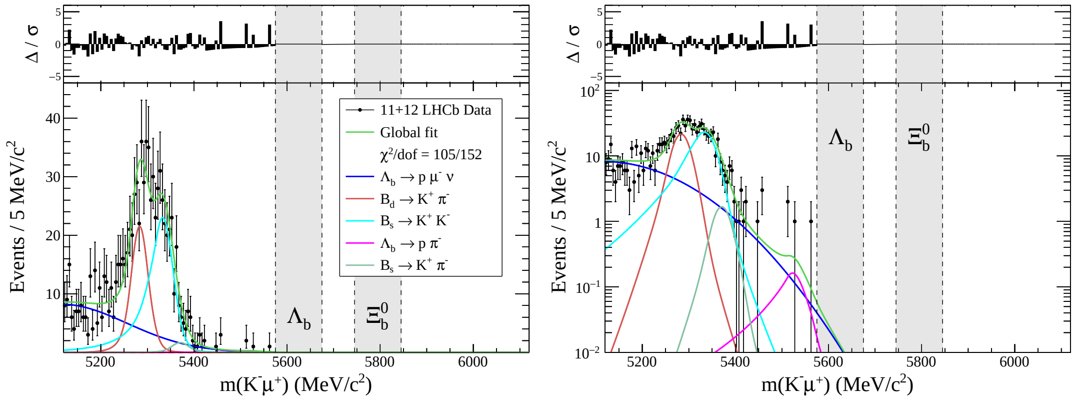
Figure 5.12: Distribution of FoM′ in dependence of ProbNN(K) and $\mathrm{x_{BDT}}$.

5.3.3 Cross check with Punzi-FoM

For comparison the optimal selection according to the Punzi-FoM $= \frac{\varepsilon_{\rm sig}}{\frac{c}{2}+\sqrt{B}}$ is determined. The parameter c is a constant that can be interpreted as the anticipated significance of the signal. A typical value is $c = 3$ to find the optimal selection for an evidence. The used values of $\varepsilon_{\rm sig}$ and B for both years are the same as before. In 2011 data the optimum is at $x_{\rm BDT} > 0.3$ and ProbNN$(K) > 0.2$ with a background yield of $B = 0.61 \pm 0.35$ and in 2012 data the optimum is at $x_{\rm BDT} > 0.3$ and ProbNN$(K) > 0.3$ with a background yield of $B = 1.9 \pm 1.0$. The signal efficiency according to signal MC is around 26 % in both years. In comparison to the optimum using the upper-limit FoM the average background rate has increased by a factor of at least six while the signal efficiency increases by less than a factor of two. This leads to the conclusion that the Punzi-FoM results in a less optimal selection due to the much larger background.

5.3.4 Summary

The optimisation of the final selection leads to similar results for 2011 and 2012 data. For the sake of simplicity a common set of cuts, namely $x_{\rm BDT} > 0.35$ and ProbNN$(K) > 0.4$, is chosen. It is abstained from an additional requirement on the muon PID due to an already vanishingly small background rate with a large statistical uncertainty and due to the great purity of the muons. Figure 5.13 shows the $m(K^-\mu^+)$ distribution of the combined data after the final selection and a fit to the spectrum. The average background rate in the Λ_b signal region is $\langle B\rangle_\Lambda = 0.36 \pm 0.17$ and $\langle B\rangle_\Xi = 0.06 \pm 0.15$ in the Ξ_b^0 signal region. According to signal MC the selection efficiencies are about 17 % and 14 % for $\Lambda_b \to K^-\mu^+$ in 2011 and 2012, respectively, and about 19 % and 16 % for $\Xi_b^0 \to K^-\mu^+$ in 2011 and 2012, respectively. The fitted number of events from the other decay channels are $N(\Lambda_b \to p\mu^-\nu_\mu) = 282 \pm 23$, $N(B^0 \to K^+\pi^-) = 212 \pm 24$, $N(B_s^0 \to K^+K^-) = 345 \pm 31$, $N(B_s^0 \to K^+\pi^-) = 16 \pm 11$ and $N(\Lambda_b \to p\pi^-) = 3 \pm 3$. There are no measurable contributions from the combinatorial background and $\Lambda_b \to pK^-$.



(a) (b)

Figure 5.13: $m(K^-\mu^+)$ distribution of 2011 and 2012 requiring the optimal signal selections.

5.3.5 Validation of the BDT selection efficiency

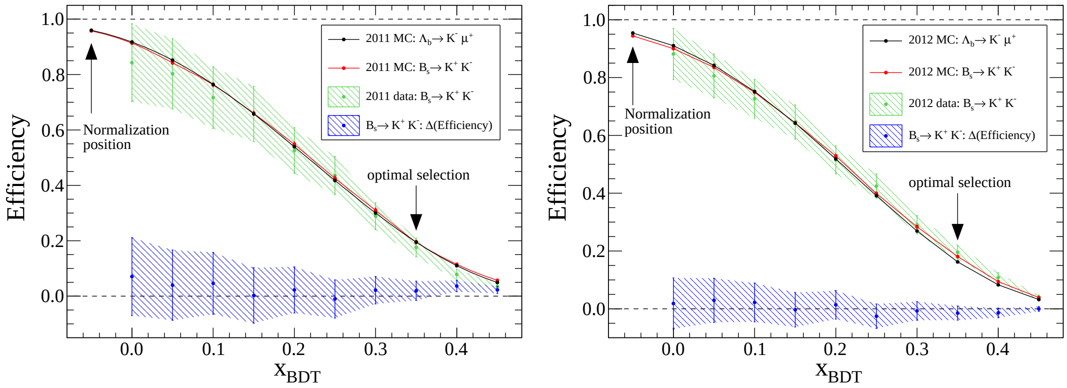
The efficiency of the BDT selection, that was determined from simulated $\Lambda_b \to K^-\mu^+$ events, is validated with the help of the control channel $B_s^0 \to K^+K^-$ to check for differences between data and MC. To do that the respective BDT classifier for 2011 and 2012 data is applied on simulated events of $B_s^0 \to K^+K^-$ and $\Lambda_b \to K^-\mu^+$, both after $\Lambda_b \to K^-\mu^+$ reconstruction and trigger selection, and $B_s^0 \to K^+K^-$ events from real data that are visible as a reflection in the $m(K^-\mu^+)$ mass spectrum. From the global fit to the $m(K^-\mu^+)$ spectrum the number of

$B_s^0 \to K^+K^-$ events (N_i) are determined in dependence of the BDT cut. For all mass spectra the events have to satisfy the optimal selection ProbNN(K) > 0.4. The efficiency of the BDT selection in data is determined using the number of fitted $B_s^0 \to K^+K^-$ events requiring $\mathrm{x_{BDT}} > -0.05$ for normalisation. This cut is a compromise between a soft selection to find the total number of events and a small combinatorial background to determine the number of signal events with sufficient precision. Hereby it is assumed that the selection efficiency at $\mathrm{x_{BDT}} > -0.05$, which is around 96 % in signal MC, deviates to a negligible amount between data and MC. This is motivated by the fact that the selection efficiency for soft BDT cuts is only slightly sensitive to Data/MC differences and the statistical uncertainty on N_i, arising mainly from the combinatorial background in the fit to the data, is dominating. The efficiency in simulated events is $\varepsilon'_i = N'_i/N'_0$ where $N'_{(0)}$ is the (total) number of counted events after BDT selection. The efficiency in data is $\varepsilon_i = \varepsilon'_0 \cdot N_i/N_0$ where ε'_0 is the MC efficiency at the normalisation position $\mathrm{x_{BDT}} > -0.05$.

In signal MC the uncertainty of $\varepsilon'_i = N'_i/N'_0$ is the binomial uncertainty $\delta(\varepsilon'_i) = \varepsilon'_i \cdot (1-\varepsilon'_i)/N'_0$. The uncertainty of $\varepsilon_i = \varepsilon'_0 \cdot N_i/N_0$ in data is determined from error propagation of N_i and N_0. The respective uncertainties of N_i and N_0 are estimated on the basis of the uncertainty from the fit $\delta_{\mathrm{Fit}}(N_i)$ which is modified (1) by subtracting the Poisson uncertainty $\delta^2_{\mathrm{Poisson}}(N_i) = N_i$ of the average value of N and (2) by adding the binomial error $\delta^2_{\mathrm{Binom}}(N_i) = N_i \cdot (1-N_i/N_0)$. This is done to obtain only the binomial uncertainty of ε_i in the used sample in addition to the one from the combinatorial background. As a result $\delta^2(N_i)$ is summarised as $\delta^2_{\mathrm{Fit}}(N_i) - \delta^2_{\mathrm{Poisson}}(N_i) + \delta^2_{\mathrm{Binom}}(N_i)$. The uncertainty of ε'_0 from signal MC is negligible. Finally the uncertainty of ε_i is

$$\delta(\varepsilon_i) = \varepsilon_i \cdot \sqrt{\left(\frac{\delta(N_i)}{N_i}\right)^2 + \left(\frac{\delta(N_0)}{N_0}\right)^2}\,. \tag{5.3}$$

Figure 5.14 illustrates the result of the studies for both years. The estimated selection efficiency of simulated $\Lambda_b \to K^-\mu^+$ events and simulated $B_s^0 \to K^+K^-$ events are very similar and compatible with the efficiency from data within its uncertainty. This justifies the use of signal MC to optimise the FoM optimisation. At the optimal $\mathrm{x_{BDT}}$ selection the signal efficiency for $B_s^0 \to K^+K^-$ events from 2011 are (17.6 ± 3.4) % in data and (19.5 ± 0.7) % in signal MC. In 2012 data the efficiency is (19.6 ± 2.4) % and (18.1 ± 0.5) % in signal MC.



(a) 2011 data (b) 2012 data

Figure 5.14: Validation of the BDT selection efficiency using $B_s^0 \to K^+K^-$ events.

In summary the efficiencies show a good agreement which can be explained mainly by the fact that the used signal variables of the BDT classifier have only little data/MC discrepancies. In combination with the efficiency of the PID selection ProbNN(K) > 0.4, which is 85.2 % in 2011 data and 85.4 % in 2012 data, and a negligible uncertainty, the overall efficiency of the $\Lambda_b \to K^-\mu^+$

signal selection is $(15.0 \pm 2.9)\,\%$ in 2011 data and $(16.7 \pm 2.1)\,\%$ in 2012 data. For the overall selection efficiency of $\Xi_b^0 \to K^-\mu^+$ events it is assumed that $\varepsilon_{\mathrm{BDT}}(\Xi_b^0 \to K^-\mu^+)/\varepsilon_{\mathrm{BDT}}(\Lambda_b \to K^-\mu^+) = 1.14 = \text{const.}$, which is determined from signal MC in both years. Hence the resulting efficiencies for $\Xi_b^0 \to K^-\mu^+$ is $(17.1 \pm 3.3)\,\%$ in 2011 data and $(19.0 \pm 2.4)\,\%$ in 2012 data. These numbers are used for the normalisation.

5.4 $\Lambda_b \to pK^-$ signal selection

For the measurement of the signal events in the normalisation channel the same strategy is applied as for the signal channel. However, due to the large number of $\Lambda_b \to pK^-$ events of approximately 10 000 in the LHCb data only a soft BDT selection is necessary. In addition no extra PID selections are made due to the fact that the BDT is more powerful in rejecting combinatorial background and contributions from reflections can be well described by a global fit to the $m(pK^-)$ spectrum.

5.4.1 2012 data

For the selection of the $\Lambda_b \to pK^-$ events from the data a BDT is trained to discriminate the combinatorial background. The background sample contains around 156 000 pK^- combinations from the upper sideband $m(pK^-) = (5720\ldots6020)\,\mathrm{MeV}/c^2$ that is indicated in Fig. 5.15. The signal sample consists of about 135 000 correctly reconstructed $\Lambda_b \to pK^-$ events from signal MC. Equivalent to the BDT classifier for the $\Lambda_b \to K^-\mu^+$ selection, both input samples imply the signal trigger selections and the same tuning parameters (NTrees/NCuts/Depth/NodeSize) are used.

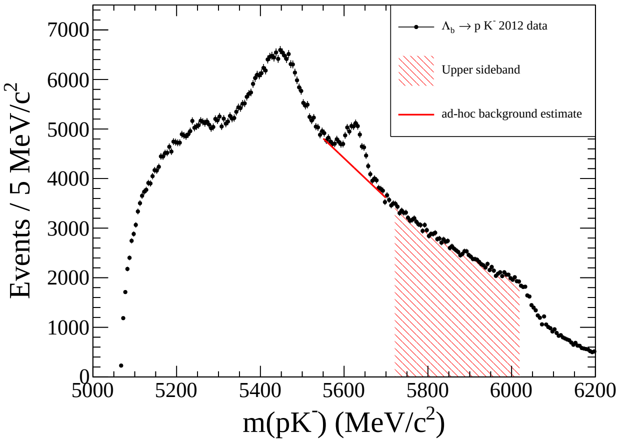
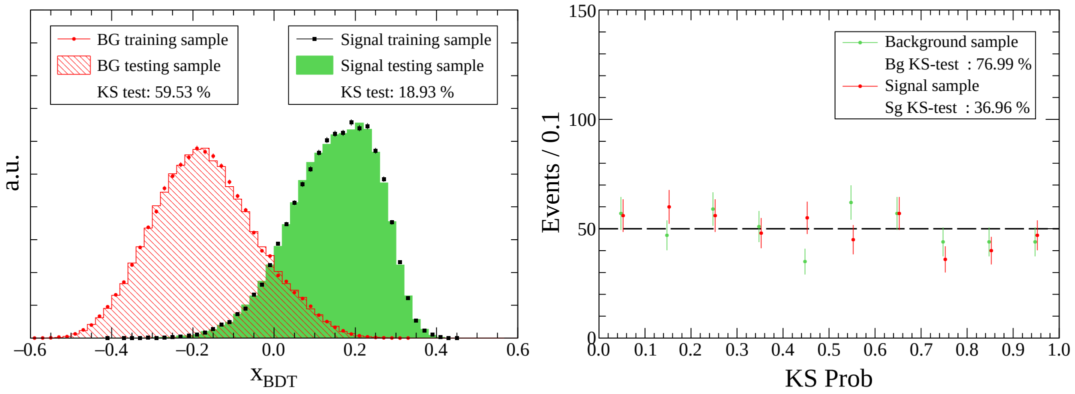


Figure 5.15: Mass spectrum of pK^- combinations after trigger selection.

Figure 5.16(a) shows the performance of the BDT for a random partition of the training and test samples. Figure 5.16(b) shows the overtraining behaviour from 500 different partitions. No overtraining is observed and hence the final BDT is trained on the whole background and signal sample to improve the performance and further reduce the impact of overtraining.

For the optimisation of the BDT selection the number of signal (S) and background events (B) is roughly estimated from an ad-hoc background estimation with a linear function in the Λ_b signal region as shown in Fig. 5.15. There are around 125 000 background events under the curve and $S \approx$ 10 000 signal events. Figure 5.17 shows the significance $S/\sqrt{S+B}$ and the background and signal efficiency in dependence of the BDT output variable. The optimal cut value is around $\mathrm{x_{BDT}} > 0.10$. For the determination of $N(\Lambda_b \to pK^-)$ from a fit to the mass spectrum a slightly softer selection of $\mathrm{x_{BDT}} > 0.05$ is chosen. According to signal MC the efficiency here is approximately 83 % and the background rejection rate is about 94 %.



(a) BDT output distribution. (b) Overtraining test.

Figure 5.16: Performance plots for the BDT.

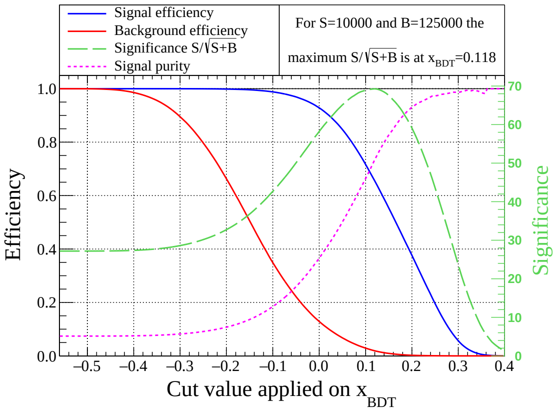


Figure 5.17: Comparison of the BDT performance.

5.4.2 2011 data

$\Lambda_b \to pK^-$ events are reconstructed using the same selections and triggers as for the 2012 data. Figure 5.18 shows the invariant mass distribution of the resulting sample of proton and kaon combinations. For the training and testing of the BDT classifier there are 72 k correctly reconstructed MC events of $\Lambda_b \to pK^-$ available and 52 k background events from the upper sideband $m(pK^-) = 5720 \ldots 6020\,\mathrm{MeV}/c^2$ from real data.

Figure 5.19(a) shows the response of the BDT classifier in the training and test samples which have a good agreement, respectively. No overtraining is observed from a sample of 500 different partitions. Therefore the final BDT for signal selection uses the full signal and background samples as it was done in the case of 2012 data. In order to find an optimal x_{BDT} selection an ad-hoc background estimate is made using the same linear function as for 2012 data but with a different scaling. As a result there are about 4 500 signal events and 42 500 background events in the $\Lambda_b \to pK^-$ signal range. The maximum significance $S/\sqrt{S+B}$ is achieved for $x_{\mathrm{BDT}} > 0.087$ which is shown in Fig. 5.19(b). A slightly softer cut at $x_{\mathrm{BDT}} > 0.0$ is chosen that has a signal efficiency of about 91 % in simulated events. The background rejection is about 89 % for events from the upper sideband.

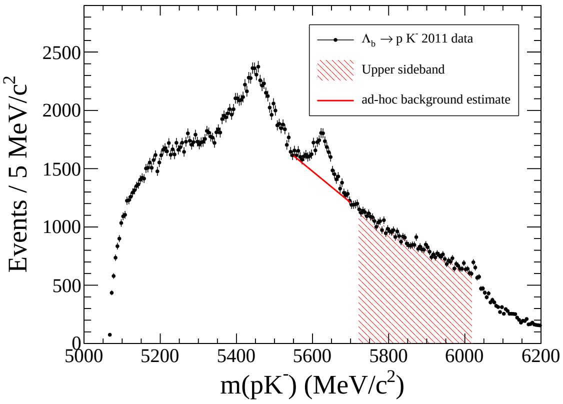
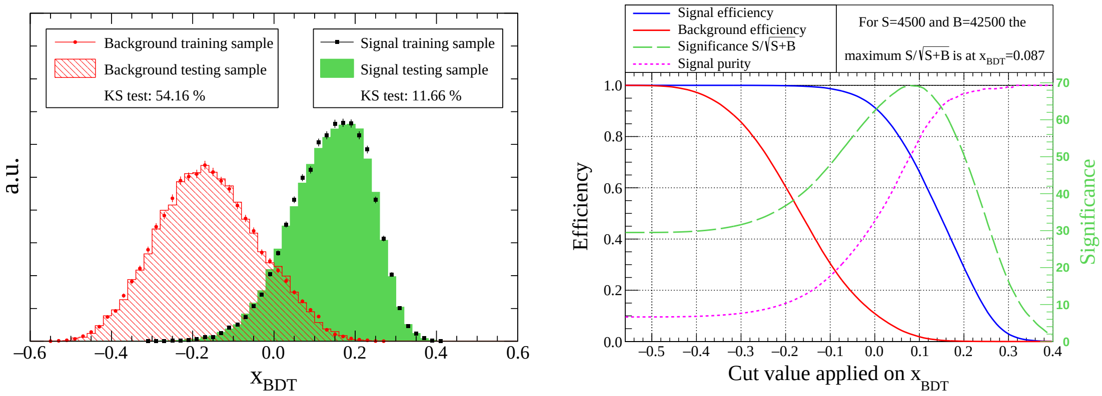


Figure 5.18: $m(pK^-)$ distribution from 2011 data after trigger selection.



(a) (b)

Figure 5.19: Performance of the BDT classifier in 2011 data.

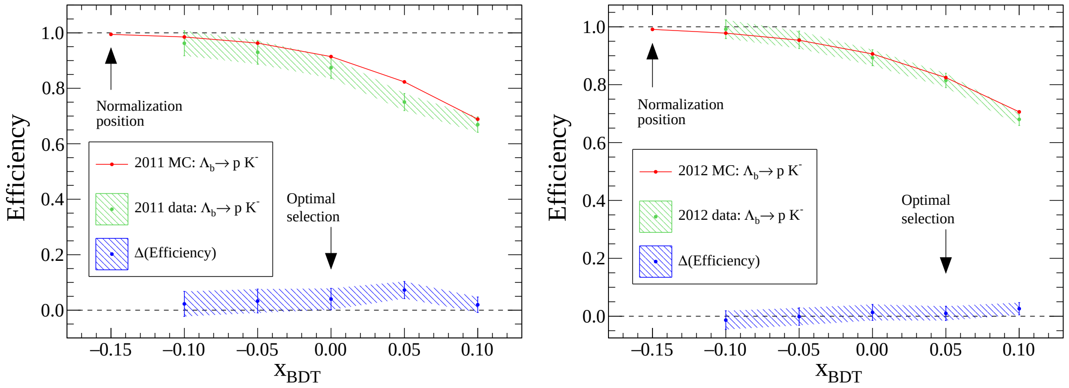
5.4.3 Validation of the BDT selection efficiency

The BDT selection efficiency is validated with the help $\Lambda_b \rightarrow pK^-$ events from data and the same method as described in Sec. 5.3.5. From fits to the $m(pK^-)$ spectrum the number of $\Lambda_b \rightarrow pK^-$ events is determined in dependence of the BDT selection. The fit model and the determination of $N(\Lambda_b \rightarrow pK^-)$ is described in more detail in Sec. 6.1. Figure 5.20 compares the resulting efficiency in 2011 and 2012 data. Requiring the optimal BDT selection $\mathrm{x_{BDT}} > 0.0$ for 2011 the efficiency in signal MC is $(91.4 \pm 0.1)\,\%$ and $(87.4 \pm 3.9)\,\%$ in real data. For the year 2012 and requiring $\mathrm{x_{BDT}} > 0.05$ the efficiency in signal MC is $(82.5 \pm 0.1)\,\%$ and $(81.5 \pm 2.5)\,\%$ for real data. The efficiency from signal MC at the normalisation position $\mathrm{x_{BDT}} > -0.15$ is larger than $99\,\%$.

5.5 Determination of the signal yield

5.5.1 Expected significance

On the basis of the expected average background rate in the respective signal regions the necessary number of observed events (N_{obs}) is determined to claim an evidence or observation of the sought decays. Therefore the probability is calculated that $N \geq N_{\mathrm{obs}}$ assuming a Poisson distributed background with a mean value $\langle B \rangle$, *i.e.* $p(N_{\mathrm{obs}}) = \sum_{N=N_{\mathrm{obs}}}^{\infty} P(N|\langle B \rangle)$ (p-value). The p-value is translated into a significance interval with the help of a Gauss distribution. Figure 5.21 shows



(a) (b)

Figure 5.20: Validation of the BDT selection efficiency using $B_s^0 \to K^+K^-$ events.

the results. The 1-σ confidence interval arising from the uncertainty on the expected average background rate is illustrated as well. Due to the lower background the observation of $\Xi_b^0 \to K^-\mu^+$ requires less observed events in comparison to $\Lambda_b \to K^-\mu^+$. To claim an observation about seven events are required in the $\Lambda_b \to K^-\mu^+$ signal region and about five events in the $\Xi_b^0 \to K^-\mu^+$ signal region.

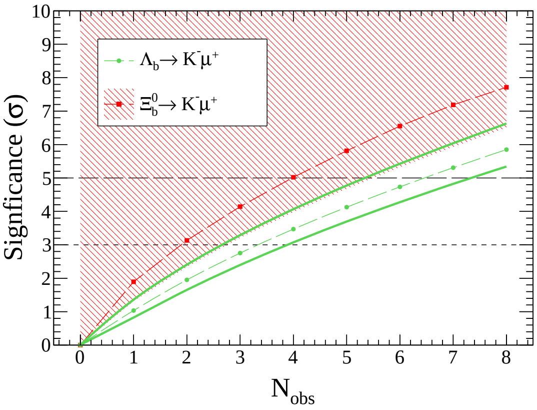
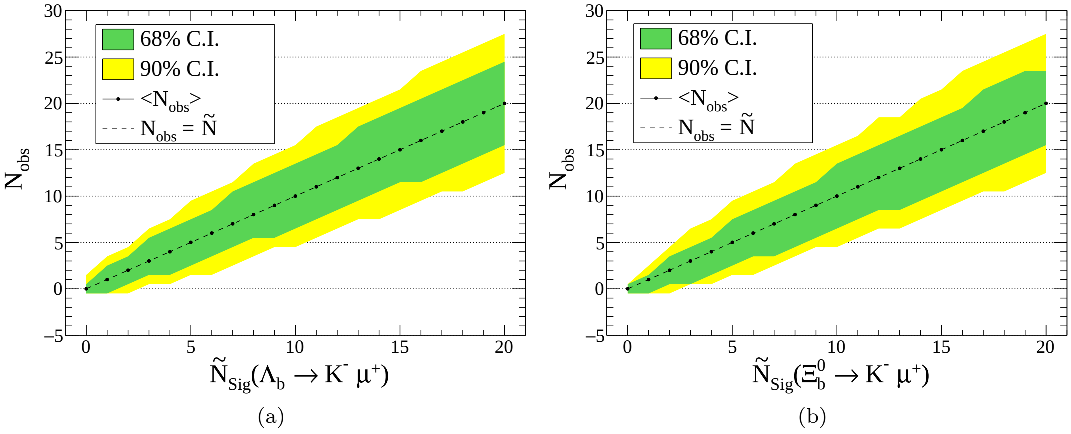


Figure 5.21: Expected significance for an observation in dependence of N_{obs}.

5.5.2 Frequentist confidence intervals

The confidence belt for the signal yield in dependence of the observed number of events is constructed following the Neyman method. Random experiments are made where the number of events in the two signal regions are simulated as the sum of background events B and signal events S. Both variables follow a Poisson distribution with the mean values $\langle B\rangle_{\Lambda} = 0.36$ and $\langle B\rangle_{\Xi} = 0.06$, and $\langle S\rangle = \tilde{N}_{\mathrm{Sig}}$, the true average number of signal events, as tunable parameter. For each simulated value of $\langle S\rangle$ 200 k random experiments are made. The confidence intervals are created such that they represent the narrowest range that contains at least 68 % and 90 % of all events, respectively. The estimator for $\tilde{N}_{\mathrm{Sig}}$ is N_{obs} due to the fact that the anticipated number of background events in both signal regions is most likely zero.

Figure 5.22 illustrates the resulting confidence belts for $\tilde{N}_{\mathrm{Sig}}(\Lambda_b \to K^-\mu^+)$ and $\tilde{N}_{\mathrm{Sig}}(\Xi_b^0 \to K^-\mu^+)$. The random experiments show that the estimator $N_{\mathrm{obs}} = \tilde{N}_{\mathrm{Sig}}$ is unbiased. On the basis of the belts the confidence interval of the number of signal events can be read once N_{obs} is determined from data.



(a) (b)

Figure 5.22: Confidence belts of $\tilde{N}_{\mathrm{Sig}}$ in dependence of N_{obs}.

5.5.3 Unblinding

Figure 5.23 shows the unblinded $m(K^-\mu^+)$ distribution from 2011 and 2012 data requiring the signal triggers and the final event selection. As a result there is one event in the Λ_b signal region and no event in the Ξ_b^0 signal region. A binned likelihood fit is made to the complete mass distribution. In comparison to the prior fit to the blinded mass spectrum, see Fig. 5.13, the fit model here includes two additional signal components for $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ events. Both modes use the sum of two Crystal Ball functions with a common mean value and all parameters except for the normalisation are fixed from a fit to reconstructed signal MC, respectively. The fit can be seen in Fig. 3.2. Due to the fact that the combinatorial background and the decay $\Lambda_b \to pK^-$ were found to be negligible in the fit to the blinded mass spectrum these components are not considered. There are altogether seven normalisations fitted to a histogram that consists of 200 bins. As a result both signal yields of the searched decay modes are zero. Figure 5.24 shows the same fit and data using a logarithmic scale for the number of events per bin.

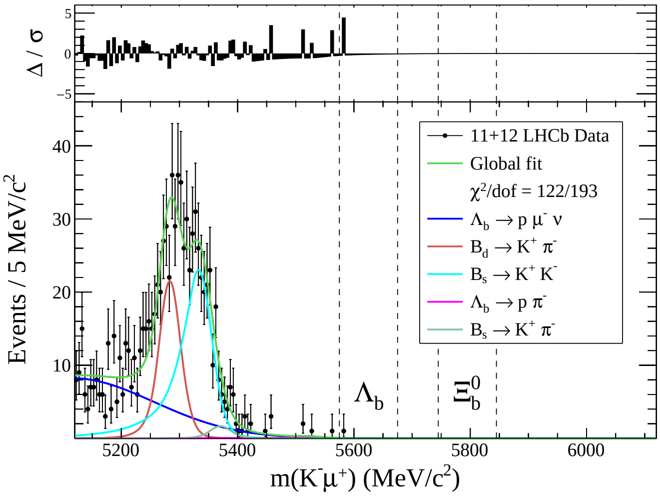


Figure 5.23: Fit to the unblinded $m(K^-\mu^+)$ distribution from 2011 and 2012 data.

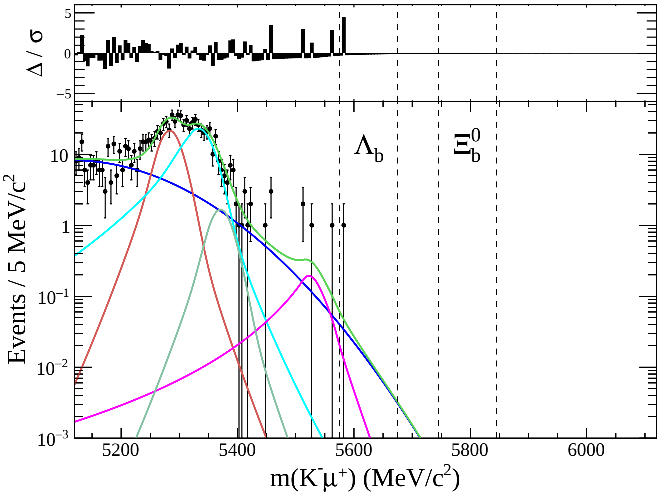


Figure 5.24: Fit to the unblinded $m(K^-\mu^+)$ distribution from 2011 and 2012 data.

Frequentist upper limit

Using the results of $N_{\text{obs}} = 1$ in the Λ_b signal range and $N_{\text{obs}} = 0$ in the Ξ_b^0 signal range the 90 % confidence intervals can be obtained from Fig. 5.22; $N(\Lambda_b \to K^-\mu^+) < 4.5$ and $N(\Xi_b^0 \to K^-\mu^+) < 2.5$. The first result does not consider the position of the event in the Λ_b signal range and is therefore a conservative upper limit.

Bayesian upper limit

For the Bayesian upper limit the likelihood of the fit is considered in dependence of the fit parameter $N(\Lambda_b \to K^-\mu^+)$ and $N(\Xi_b^0 \to K^-\mu^+)$, respectively. The likelihood $l(\tilde{N}|\vec{n})$ is converted into an *a posteriori* probability density $f(\tilde{N}|\vec{n})$ using the Bayes Theorem

$$f(\tilde{N}|\vec{n}) = \frac{l(\tilde{N}|\vec{n})}{\int l(N'|\vec{n})\, f(N')\, dN'} \cdot f(\tilde{N}) \qquad [66]\ , \tag{5.4}$$

where N is the fitted number of signal events, $\tilde{N} = \langle N \rangle$ is the true number of signal events that is estimated by N, and $\vec{n}$ the measured number of events per bin in the $m(K^-\mu^+)$ spectrum. The prior probability $f(\tilde{N})$ is assumed to be constant and non vanishing in the range $\tilde{N} \geq 0$ and zero everywhere else. This leads to the normalised *subjective* probability density

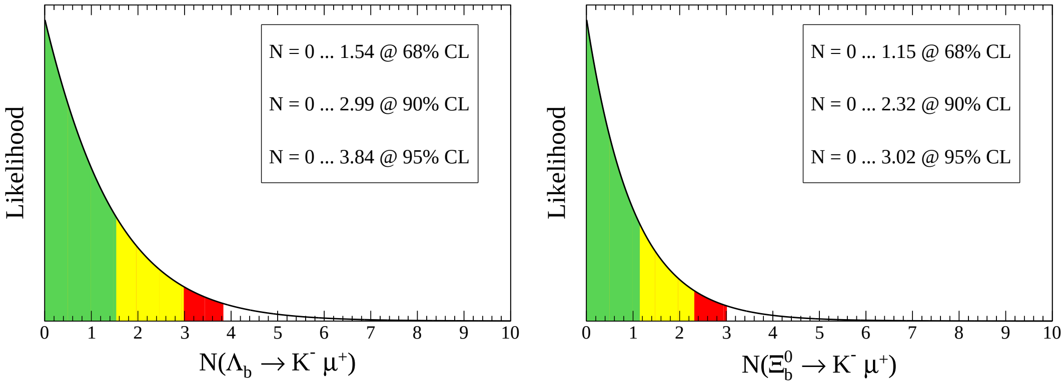
$$c(\tilde{N}|\vec{n}) = \frac{l(\tilde{N}|\vec{n})}{\int l(N'|\vec{n})\, dN'} \qquad [66] \tag{5.5}$$

that is used to calculate a Bayesian upper limit on the true number signal events N_{up}

$$\int_0^{N_{\text{up}}} c(N|\vec{n})\, dN = \text{CL}\ . \tag{5.6}$$

Figure 5.25 shows the likelihood distribution of the fit. The 90 % confidence intervals are $N(\Lambda_b \to K^-\mu^+) < 2.99$ and $N(\Xi_b^0 \to K^-\mu^+) < 2.32$. The upper limit of $N(\Xi_b^0 \to K^-\mu^+)$ is slightly larger than 2.3 one would obtain applying a Poisson distribution with zero observed events. This is due to the fact that the empty Ξ_b^0 signal window covers only about 93 % of the signal distribution.

Therefore there is a non-vanishing probability that one or more of the events in the complete mass spectrum result from $\Xi_b^0 \to K^-\mu^+$. The assumption was tested by determining the Bayesian upper limit using a $\Xi_b^0 \to K^-\mu^+$ fit model where the signal width is reduced by a factor of ten and hence the Ξ_b^0 signal window contains 100 % of the signal distribution. As a result the upper limits are in agreement with the ones resulting from a Poisson distribution with zero observed events.



(a) (b)

Figure 5.25: Likelihood distribution in dependence of $N(\Lambda_b \to K^-\mu^+)$ (a) and $N(\Xi_b^0 \to K^-\mu^+)$ (b). The green, yellow and red ranges represent the 68 %, 90 %, and 95 % confidence interval, respectively.

> *"If enough data is collected, anything may be proven by statistical methods. "*
>
> — Arthur Bloch (1948- , from his books "Murphy's Law")

6

Normalisation

The measurement of the branching fraction is done relative to the normalisation channel $\Lambda_b \to pK^-$. This improves the precision of the result since systematic uncertainties are partly canceling in ratios of efficiencies and signal yields. These biases can be due to the imperfect simulation of LHCb events and the interaction of particles with the detector, and biases in the determination of the selection efficiencies that is done for both decays in a similar way. Another advantage of a relative measurement is the fact that no knowledge about the absolute number of Λ_b produced at LHCb is required. Equation 6.1 shows how $\mathcal{B}_{\text{sig}}$ is determined. The factor $f_{\text{norm}}/f_{\text{sig}}$ describes the relative production rate of the decaying particles that is unity in the $\Lambda_b \to K^-\mu^+$ case.

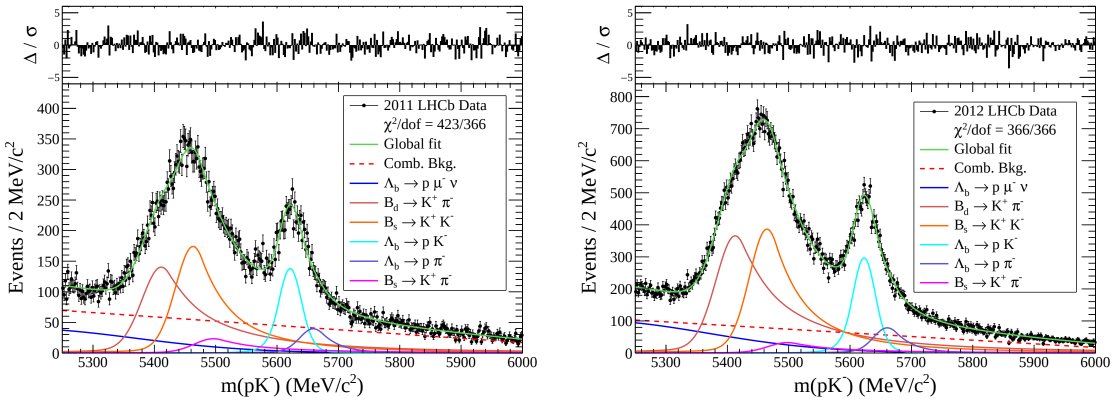
$$\mathcal{B}_{\text{sig}} = \mathcal{B}_{\text{norm}} \times \frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}} \times \frac{N_{\text{sig}}}{N_{\text{norm}}} \times \frac{f_{\text{norm}}}{f_{\text{sig}}} \tag{6.1}$$

In the following $N_{\text{norm}} = N(\Lambda_b \to pK^-)$ and the efficiencies are measured.

6.1 Determination of $N(\Lambda_b \to pK^-)$

The number of $\Lambda_b \to pK^-$ signal events is determined from a fit to to the pK^- mass spectrum in 2011 and 2012 data, respectively. Apart from the event and trigger selections, the pK^- candidates are required to satisfy $\mathrm{DLL}(\mathrm{p}-\pi)(\mathrm{p}) > -5$, $\mathrm{DLL}(\mathrm{K}-\pi)(\mathrm{K}^-) > -5$, $\mathrm{DLL}(\mathrm{p}-\mathrm{K}^-)(\mathrm{p}) > -5$, and the MVA selections $\mathrm{x}_{\mathrm{BDT}} > 0.00$ for 2011 data and $\mathrm{x}_{\mathrm{BDT}} > 0.05$ for 2012 data to suppress the combinatorial background as described in Sec. 5.4.2 and 5.4.1. The fit model includes two-body decays of B^0, B_s^0 and Λ_b whose shapes are fixed to the ones found in reconstructed signal MC. To describe the combinatorial background a first-order polynomial is used. The mass shift is considered by a floating parameter D that is added commonly to all mean masses. A detailed description of the fit model is presented in App. A.24.

Figure 6.1 shows the fit results and Tab. 6.1 summarises the signal yields. In total there are about 10 800 $\Lambda_b \to pK^-$ events in the combined data. There is an insignificant excess of Λ_b to $\overline{\Lambda}_b$ events visible in both years. According to MC studies a ratio $N(\Lambda_b) : N(\overline{\Lambda}_b)$ of 1.083(2011) and 1.070(2012) is expected after trigger and signal selections and starting with equal amounts of Λ_b and $\overline{\Lambda}_b$. Therefore an equal production of Λ_b and $\overline{\Lambda}_b$ in data is assumed for the calculation of efficiencies.



(a) 2011 data (b) 2012 data

Figure 6.1: Fits to the $m(pK^-)$ distribution in 2011 and 2012 data.

Table 6.1: Number of measured $\Lambda_b \to pK^-$ and $\overline{\Lambda}_b \to \overline{p}K^+$ events in both years.

Year	$N(\Lambda_b \to pK^-)$	$N(\overline{\Lambda}_b \to \overline{p}K^+)$	$\sum$	$N(\Lambda_b) : N(\overline{\Lambda}_b)$
2011	1727 ± 72	1715 ± 78	3442 ± 106	1.01(6)
2012	3785 ± 110	3559 ± 114	7344 ± 158	1.06(5)

6.2 Determination of the total efficiency

The total efficiency is defined as the number of signal events N_{Sig} that are found after all selections divided by the number of decays that have happened N_0. To simplify the procedure the total efficiency is factorised into four terms that are determined separately. Equation 6.2 shows how the efficiency is factorised.

$$\varepsilon_{\mathrm{tot}} = \frac{N_{\mathrm{Sig}}}{N_0} = \frac{N_{\mathrm{Acc}}}{N_0} \times \frac{N_{\mathrm{Sel}}}{N_{\mathrm{Acc}}} \times \frac{N_{\mathrm{Trig}}}{N_{\mathrm{Sel}}} \times \frac{N_{\mathrm{Sig}}}{N_{\mathrm{Trig}}} \,. \tag{6.2}$$

These four factors are: the detector acceptance $\varepsilon_{\mathrm{Acc}} = N_{\mathrm{Acc}}/N_0$, the event selection efficiency within detector acceptance $\varepsilon_{\mathrm{sel}} = N_{\mathrm{Sel}}/N_{\mathrm{Acc}}$, the trigger efficiency of selected events $\varepsilon_{\mathrm{Trig}} = N_{\mathrm{Trig}}/N_{\mathrm{Sel}}$, and the efficiency of the signal selection $\varepsilon_{\mathrm{sig}} = N_{\mathrm{Sig}}/N_{\mathrm{Trig}}$. Hence the total efficiency is $\varepsilon_{\mathrm{tot}} = \varepsilon_{\mathrm{Acc}} \times \varepsilon_{\mathrm{sel}} \times \varepsilon_{\mathrm{Trig}} \times \varepsilon_{\mathrm{sig}}$.

6.3 Detector acceptance

The detector acceptance is measured in the course of generation and simulation by summing up the number of signal events N_{Acc} in which all daughter particles traverse the detector and allow a correct decay reconstruction. N_{Acc} is divided by the total number of generated events N_0. Table 6.2 shows the detector acceptance that is found for the different years and magnet polarities. The mean value considers the different fractions of the integrated luminosities from data according to Tab. 2.1. All uncertainties are statistical.

Table 6.2: Detector acceptance ε_{Acc} for signal and control channels

$\varepsilon_{\text{Acc}}(\%)$		$\Lambda_b \to K^-\mu^+$	$\Xi_b^0 \to K^-\mu^+$	$\Lambda_b \to pK^-$
2011	MagDown	18.70 ± 0.06	18.55 ± 0.06	19.64 ± 0.07
	MagUp	18.65 ± 0.06	18.48 ± 0.06	19.43 ± 0.07
	mean	18.68 ± 0.05	18.52 ± 0.05	19.55 ± 0.05
2012	MagDown	18.95 ± 0.08	18.80 ± 0.06	19.99 ± 0.10
	MagUp	18.89 ± 0.08	18.86 ± 0.06	19.80 ± 0.10
	mean	18.92 ± 0.06	18.83 ± 0.05	19.90 ± 0.07

6.4 Event selection efficiency

The event selection efficiency ε_{sel} is separated according to Eq. 6.3 into the efficiency of the PID requirements ε_{PID} and the efficiency of the remaining selections of the decay reconstruction ε_{Rec}. This is done due to the fact that the PID variables are not correctly simulated and ε_{PID} has to be determined separately with the help of calibration samples from data.

$$\varepsilon_{\text{sel}} = \frac{N_{\text{Sel}}}{N_{\text{Acc}}} = \frac{N_{\text{Rec}}}{N_{\text{Acc}}} \cdot \frac{N_{\text{Sel}}}{N_{\text{Rec}}} = \varepsilon_{\text{Rec}} \cdot \varepsilon_{\text{PID}} \ . \tag{6.3}$$

6.4.1 Determination of the reconstruction efficiency

All reconstruction efficiencies are determined from binned likelihood fits to the invariant mass spectra of simulated events that pass the set of event selection criteria except for the PID requirements. In addition, the muons are required to be in the acceptance region of the muon chambers to allow the IsMuon decision to be made. Events with a double misidentification, *i.e.* the cases $K = \mu$ and $\mu = K$ for $\Lambda_b \to K^-\mu^+$ and $p = K$ and $K = p$ for $\Lambda_b \to pK^-$, that are reconstructed in almost every simulated signal event, are not considered. It is made use of the same fit model as before in the studies of simulated events, see Sec. 3.4.2. The efficiency is determined as the ratio of the fitted number of signal events divided by the total number of generated decays within detector acceptance. Table 6.3 summarises the results, the uncertainties are statistical.

Table 6.3: Reconstruction efficiencies for signal and normalisation channels.

$\varepsilon_{\text{Rec}}(\%)$	$\Lambda_b \to K^-\mu^+$	$\Xi_b^0 \to K^-\mu^+$	$\Lambda_b \to pK^-$
2011	34.10 ± 0.09	35.29 ± 0.09	23.10 ± 0.04
2012	31.47 ± 0.09	32.89 ± 0.09	21.49 ± 0.04

6.4.2 Determination of the PID efficiency

The PID efficiency of the event selection is measured with the help of reconstructed MC events and PID efficiency tables. These tables are provided by a dedicated group of LHCb scientists and are the recommended method to determine the performance of PID selections. They were created using real data of reconstructed $J/\psi \to \mu^+\mu^-$ events for muons, $D^{*+} \to D^0(K^+\pi^-)\pi^+$

events for kaons and $\Lambda \rightarrow p\pi^-$ and $\Lambda_b \rightarrow \Lambda_c(pK^-\mu^+)\mu^-\nu_\mu$ for protons and without any PID requirement on the track of interest, respectively. After applying a certain PID selection, the fraction of remaining signal events is measured in the invariant mass spectrum and in bins of the track multiplicity, magnet polarity, the track's momentum, and the track's pseudorapidity. Beside the efficiency ε_i its binomial uncertainty $\delta\varepsilon_i$ is provided in the tables as well. Thanks to the large production rate of J/ψ and D^{*+} at LHCb the efficiencies have a high precision ($\delta\varepsilon/\varepsilon < 0.1\,\%$) in most of the bins. The precision of the proton PID tables is low in certain regions of the phase space due to a limited number of protons in the edge regions of η.

The PID efficiency $\varepsilon_{\mathrm{PID}}(x)$ for the track x from a given decay mode is determined as a weighted sum over the number of signal events N_i and the corresponding efficiency ε_i, where i indicates a specific bin in the four-dimensional space. Equation 6.4 summarises how $\varepsilon_{\mathrm{PID}}$ and its uncertainty $\delta\varepsilon_{\mathrm{PID}}$ is determined.

$$\varepsilon_{\mathrm{PID}} = \frac{\sum N_i \cdot \varepsilon_i}{\sum N_i} \quad , \quad \delta\varepsilon_{\mathrm{PID}} = \sqrt{\frac{\sum N_i \cdot \delta^2\varepsilon_i}{\sum N_i}} \ . \tag{6.4}$$

$\Lambda_b \rightarrow K^-\mu^+$ PID efficiency

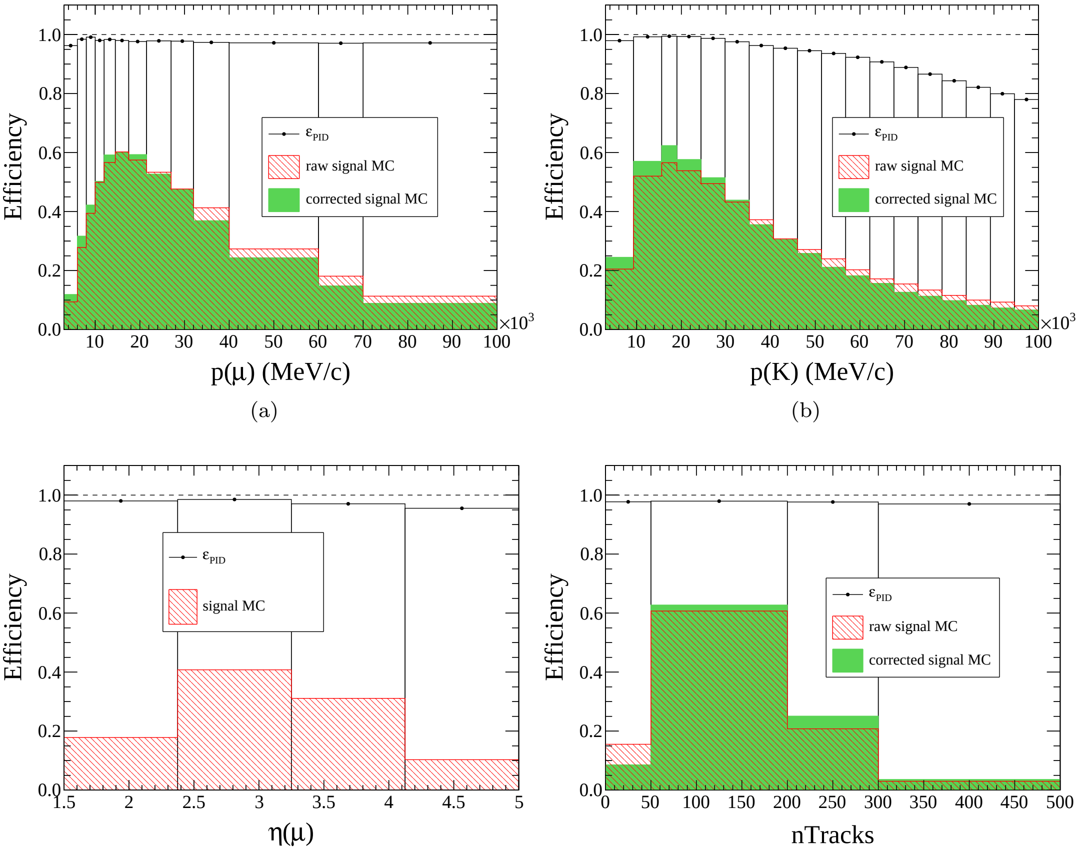
The PID efficiency of the $\Lambda_b \rightarrow K^-\mu^+$ event selection is measured independently for both tracks according to $\varepsilon_{\mathrm{PID}} = \varepsilon_{\mathrm{PID}}(K) \cdot \varepsilon_{\mathrm{PID}}(\mu)$. The calculation uses corrected momenta and track multiplicities as described in Sec. 4.5 and 4.6. Table 6.4 shows the results for $\Lambda_b \rightarrow K^-\mu^+$ events in dependence of the two magnet polarities. For comparison the PID efficiency using uncorrected signal MC is shown as well. The calculation of the mean value considers the different luminosities per magnet polarity in both years. In conclusion there is no significant deviation between them. For the muons the correction has only a small influence while $\varepsilon_{\mathrm{PID}}(K)$ differs relatively by about 1.6 %. Figure 6.2 shows one-dimensional projections of the PID efficiencies and the histograms of $\Lambda_b \rightarrow K^-\mu^+$ signal MC events with and without correction. The binning is given by the PID tables and both MC histograms are normalised to the bin width and the scale is in arbitrary units. In summary the largest dependence of the PID efficiency is given by the momentum distribution of the tracks. Table 6.5 summarises the PID efficiencies for $\Lambda_b \rightarrow K^-\mu^+$ and $\Xi_b^0 \rightarrow K^-\mu^+$ in both years.

Table 6.4: Calculated PID efficiencies for $\Lambda_b \rightarrow K^-\mu^+$ events in 2012 data.

$\varepsilon_{\mathrm{PID}}$ (%)	MagDown	MagUp	Mean
IsMuon(μ)=true, uncorrected	97.85 ± 0.43	97.74 ± 0.43	97.75 ± 0.31
IsMuon(μ)=true, corrected	97.65 ± 0.44	97.56 ± 0.44	97.60 ± 0.31
DLL(K − π)(K), uncorrected	91.88 ± 0.18	91.57 ± 0.17	91.73 ± 0.12
DLL(K − π)(K), corrected	93.34 ± 0.17	93.07 ± 0.16	93.21 ± 0.12

Table 6.5: PID efficiencies for $\Lambda_b \rightarrow K^-\mu^+$ and $\Xi_b^0 \rightarrow K^-\mu^+$ events.

$\varepsilon_{\mathrm{PID}}$ (%)	$\Lambda_b \rightarrow K^-\mu^+$	$\Xi_b^0 \rightarrow K^-\mu^+$
2011	91.3 ± 0.5	91.2 ± 0.4
2012	91.0 ± 0.3	90.8 ± 0.3



(a) (b)

(c) (d)

Figure 6.2: Projections in p and η of the PID efficiency and the distribution of simulated $\Lambda_b \to K^-\mu^+$ events with and without correction in arbitrary units and normalised to the bin width.

$\Lambda_b \to pK^-$ PID efficiency

The PID selection of the $\Lambda_b \to pK^-$ events are $\mathrm{DLL(K-\pi)(K)} > -5.0$, $\mathrm{DLL(p-\pi)(p)} > -5.0$, and $\mathrm{DLL(p-K)(p)} > -5.0$. For the kaon selection the same strategy is applied as before, including the correction of the momenta and track multiplicity, and it is determined $\varepsilon_{\mathrm{PID}}(K) = (99.37 \pm 0.13)\,\%$ for 2011 data and $\varepsilon_{\mathrm{PID}}(K) = (99.34 \pm 0.07)\,\%$ for 2012 data.

The precision of the proton PID efficiency suffers from the limited statistics of the PID tables. For example 35 % of the simulated protons are located in bins with an uncertainty $\delta(\varepsilon) > 1\,\%$ and still 10 % are in bins with $\delta(\varepsilon) > 50\,\%$. As an alternative approach $\varepsilon_{\mathrm{PID}}(p)$ is estimated using the kaon PID tables instead. This is approach is motivated by the fact that the $K - \pi$ separation of a given track with a certain momentum is always worse than the $p - K$ and $p - \pi$ separation if the particle identification is based on the measurement of the Cerenkov angle, see App. A.25. The combined proton PID efficiency is therefore estimated as $\varepsilon_{\mathrm{PID}}(p) = \varepsilon\,[\mathrm{DLL(p-K)(p)} > -5.0] \cdot \varepsilon\,[\mathrm{DLL(p-\pi)(p)} > -5.0] \approx \varepsilon\,[\mathrm{DLL(K-\pi)(p)} > -5.0]$ and is $(99.32 \pm 0.09)\,\%$ in 2011 data and $(99.38 \pm 0.08)\,\%$ in 2012 data. In conclusion the overall efficiency is $\varepsilon_{\mathrm{PID}} \approx (98.69 \pm 0.15)\,\%$ for 2011 data and $\varepsilon_{\mathrm{PID}} \approx (98.66 \pm 0.17)\,\%$ for 2012 data.

6.4.3 Summary of the event selection efficiencies

The results of the reconstruction and PID selection efficiencies are combined according to $\varepsilon_{\text{sel}} = \varepsilon_{\text{Rec}} \cdot \varepsilon_{\text{PID}}$ and summarised in Tab. 6.6. The stated uncertainties result from quadratic combination of the individual errors.

Table 6.6: Summary of the event selection efficiencies $\varepsilon_{\text{sel}} = \varepsilon_{\text{Rec}} \cdot \varepsilon_{\text{PID}}$.

$\varepsilon_{\text{sel}}(\%)$	$\Lambda_b \to K^-\mu^+$	$\Xi_b^0 \to K^-\mu^+$	$\Lambda_b \to pK^-$
2011	31.13 ± 0.19	32.18 ± 0.16	22.80 ± 0.05
2012	28.64 ± 0.12	29.86 ± 0.13	21.20 ± 0.05

6.5 Determination of the trigger efficiency

The determination of the trigger efficiency is done in two steps: (1) the determination of the L0 trigger efficiency (ε^{L0}) using data driven method and (2) the HLT efficiency (ε^{HLT}) using signal MC. The distinction is necessary due to the fact that the L0 trigger is not correctly simulated as it was shown in Sec. 4.2.1. In addition, the use of a data-driven method has the advantage that the mix of different trigger conditions during data taking is automatically included in the efficiencies. In contrast to the L0 trigger the discrepancies of the HLT efficiencies between events from real data and simulation are small. The total trigger efficiency is factorised according to equation 6.5 which is justified by the fact that the hardware and software trigger are deciding independently from each other.

$$\varepsilon_{\text{Trig}} = \frac{N_{\text{trig}}}{N_{\text{sel}}} = \frac{N_{L0}}{N_{\text{sel}}} \cdot \frac{N_{\text{trig}}}{N_{L0}} = \varepsilon^{L0} \cdot \varepsilon^{HLT} \tag{6.5}$$

6.5.1 $\Lambda_b \to pK^-$ trigger efficiency

The total L0 trigger efficiency is calculated as

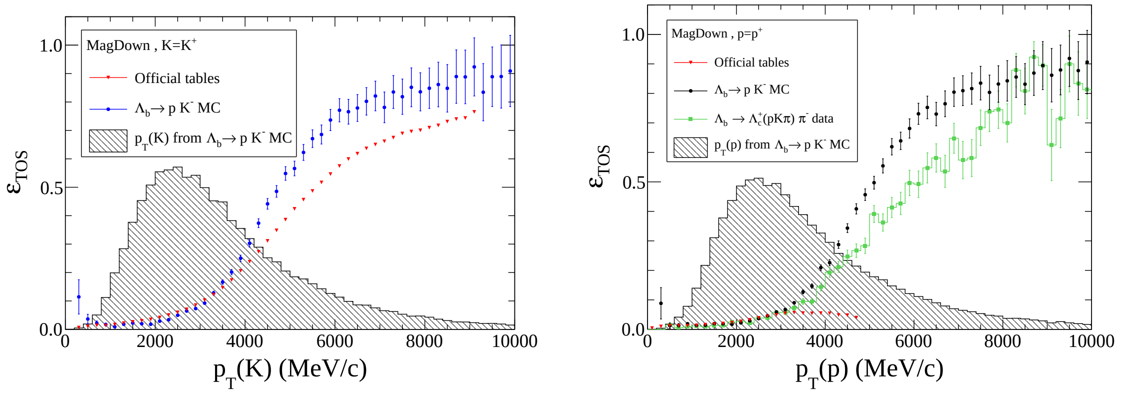
$$\varepsilon^{L0} = \frac{N_{TOS} + N_{TIS} - N_{TISTOS}}{N_{\text{sel}}} = \frac{N_{TOS}}{N_{\text{sel}}} \cdot \frac{N_{TOS} + N_{TIS} - N_{TISTOS}}{N_{TOS}} \tag{6.6}$$

$$\varepsilon^{L0} = \underbrace{\frac{N_{TOS}}{N_{\text{sel}}}}_{\varepsilon_{TOS}} \cdot \left(1 + \underbrace{\frac{N_{TIS} - N_{TISTOS}}{N_{TOS}}}_{\varepsilon_{TIS}}\right) . \tag{6.7}$$

For the mode $\Lambda_b \to pK^-$ the relevant L0 triggers are L0-Hadron-TOS (abbrev. as TOS) and L0-Global-TIS (abbrev. as TIS). The TOS efficiency of kaons is determined with the help of official LHCb tables that were created by applying the TISTOS method on $D^{*+} \to D^0(K^+\pi^-)\pi^+$ events. This means that from a sample of TIS-triggered signal events the TOS efficiency is measured as $\varepsilon^{TOS}(p_{\text{T}}) = N_{TIS\&TOS}(p_{\text{T}})/N_{TIS}(p_{\text{T}})$ where $N(p_{\text{T}})$ results from a fit to the mass spectrum of reconstructed D^{*+} events in a certain p_{T} interval of the investigated track. The tables distinguish between both magnet polarities and both charges.

Figure 6.3(a) shows the efficiency distribution of kaons from simulated $\Lambda_b \to pK^-$ events and from the official tables along with the distribution of the corrected transversal momentum of the kaon in arbitrary units. The magnet polarity is MagDown and the kaon charge is positive. There is a clear overestimation visible in simulated events starting around $p_{\text{T}}(K) = 3000\,\text{MeV}/c$. This affects more than 50 % of the $\Lambda_b \to pK^-$ events. About 2.5 % of all kaons have a p_T that list outside of the range of the official tables and are weighted using the efficiency from the highest p_T bin.

For the TOS efficiency of the protons the official tables, derived from $\Lambda \to p\pi^-$ events, hold no information in high-p_T regions. Therefore it is made use of tables that were created by applying the TISTOS method on a sample of reconstructed $\Lambda_b \to \Lambda_c(\to pK^-\pi^+)\pi^-$ events [67]. These tables include information up to $p_\mathrm{T}(p) = 10000\,\mathrm{MeV}/c$. Less than 2 % of the protons are above this threshold and are weighted using the efficiency from the highest p_T bin. Figure 6.3(b) compares the proton TOS efficiency of simulated $\Lambda_b \to pK^-$ events to the $\Lambda_b \to \Lambda_c(\to pK^-\pi^+)\pi^-$ calibration samples. The efficiency is overestimated for simulated protons with $p_\mathrm{T} \gtrsim 4000\,\mathrm{MeV}/c$.



(a) (b)

Figure 6.3: Comparison of L0-Hadron-TOS efficiency for p and K from 2012 data and MC. The hatched histograms show the p_T distribution from simulated events with arbitrary scaling.

For the determination of the trigger efficiency of the track type $x = \{p, K\}$, events from reconstructed signal MC are separated in bins of the corrected transversal momentum p_T. A weighted efficiency and its uncertainty is calculated according to Eq. 6.8.

$$\varepsilon_{TOS}(x) = \frac{\sum N_i(p_\mathrm{T}) \cdot \varepsilon_i(p_\mathrm{T})}{\sum N_i(p_\mathrm{T})} \;, \quad \delta\varepsilon_{TOS}(x) = \sqrt{\frac{\sum N_i(p_\mathrm{T}) \cdot \delta^2\varepsilon_i(p_\mathrm{T})}{\sum N_i(p_\mathrm{T})}} \tag{6.8}$$

Here N_i is the number of MC events and $\varepsilon_i \pm \delta\varepsilon_i$ is the TOS efficiency from the tables in the p_T bin i, respectively. MC events with a $p_\mathrm{T}(x)$ that is not mapped in the tables, which concerns about 3 % of all events, are weighted with the efficiency from the highest p_T bin that is available. Table 6.7 shows the resulting efficiencies in dependence of charge and magnet polarity. The calculation of the average value considers the different shares of integrated luminosity per magnet polarity. $\Lambda_b \to pK^-$ and its charge conjugated decay are weighted $1:1$ assuming that the $\Lambda_b/\overline{\Lambda}_b$ production is symmetric and any deviation that is seen, see Tab. 6.1, is due to detector effects.
The total TOS efficiency can be written as

$$\varepsilon(p = TOS \texttt{ OR } K = TOS) = \varepsilon_{TOS}(p \texttt{ OR } K) = \varepsilon_{TOS}(p) + \varepsilon_{TOS}(K) - \varepsilon_{TOS}(p \texttt{ AND } K)\;. \tag{6.9}$$

Due to momentum conservation the transversal momentum of the proton and the kaon are not independent from each other, *i.e.* $\vec{p_\mathrm{T}}(p) + \vec{p_\mathrm{T}}(K) = \vec{p_\mathrm{T}}(\Lambda_b)$. Figure 6.4 shows the distribution of $p_\mathrm{T}(p)$ *vs.* $p_\mathrm{T}(K)$. The correlation is $-19\,\%$. Therefore the combined TOS efficiency does not factorise according to $\varepsilon_{TOS}(p \texttt{ AND } K) = \varepsilon_{TOS}(p) \cdot \varepsilon_{TOS}(K)$.

Table 6.7: Weighted L0-TOS efficiencies of p and K from $\Lambda_b \to pK^-$ events.

Year	$\varepsilon_{TOS}(\%)$	p	$\overline{p}$	K^+	K^-
2011	MagDown	19 ± 5	22 ± 5	24.4 ± 0.3	24.3 ± 0.3
	MagUp	18 ± 5	20 ± 6	24.1 ± 0.4	24.2 ± 0.3
2011	Average	20.1 ± 2.5		24.3 ± 1.6	
2012	MagDown	18.4 ± 2.9	20.2 ± 2.8	20.77 ± 0.14	20.47 ± 0.14
	MagUp	17.9 ± 3.2	19.4 ± 3.0	20.42 ± 0.15	20.47 ± 0.14
2012	Average	19.0 ± 1.5		20.53 ± 0.07	

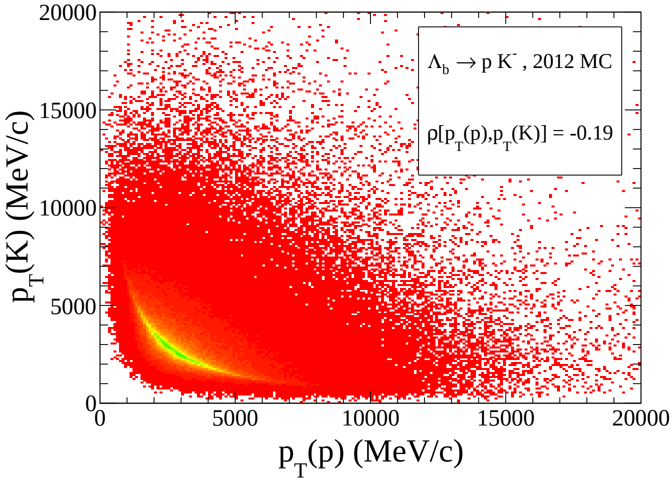


Figure 6.4: Correlation of the track's p_{T} in reconstructed $\Lambda_b \to pK^-$ events in 2012 MC.

To consider the dependence, the combined TOS efficiency of proton and kaon is factorised by introducing a factor $R \geq 0$ according to

$$\varepsilon_{TOS}(p \texttt{ OR } K) = \varepsilon_{TOS}(p) + \varepsilon_{TOS}(K) - \varepsilon_{TOS}(p) \cdot \varepsilon_{TOS}(K) \cdot R \ . \tag{6.10}$$

R is determined using signal MC events and according to

$$R = \frac{\varepsilon_{TOS}(p) + \varepsilon_{TOS}(K) - \varepsilon_{TOS}(p \texttt{ OR } K)}{\varepsilon_{TOS}(p) \cdot \varepsilon_{TOS}(K)} \ . \tag{6.11}$$

The values for both years are $R(2011) = (0.814 \pm 0.005)$ and $R(2012) = (0.849 \pm 0.004)$ where the uncertainty on R results from quadratic summation of the respective binomial uncertainties on the efficiencies. The L0-Hadron-TOS efficiencies of $\Lambda_b \to pK^-$ are $\varepsilon_{TOS}(2011) = (40.4 \pm 2.1)\,\%$ and $\varepsilon_{TOS}(2012) = (36.2 \pm 1.2)\,\%$. To determine the complete L0 trigger efficiency according to Eq. 6.7 the values N_{TOS} and $N_{TIS} - N_{TISTOS}$ are determined from fits to the $m(pK^-)$ spectrum after requiring the corresponding L0 trigger conditions. The results are for 2011 data is $\varepsilon_{TIS} = (1369 \pm 73)/(2864 \pm 110) = (47.8 \pm 3.1)\,\%$ and for 2012 data $\varepsilon_{TIS} = (2723 \pm 89)/(5315 \pm 122) =$

(51.2±2.0) %. For comparison, in signal MC the TIS efficiencies are (32.2±0.3) % and (35.3±0.2) % in the 2011 and 2012 samples which differs significantly from real data. This could be due to the lower event multiplicity and an incorrect simulation of the underlying event that causes L0-TIS. In conclusion the L0 trigger efficiencies in both years are $\varepsilon^{L0}(2011) = (59.7 \pm 3.3)\,\%$ and $\varepsilon^{L0}(2012) = (54.7 \pm 2.0)\,\%$

HLT efficiency and total trigger efficiency

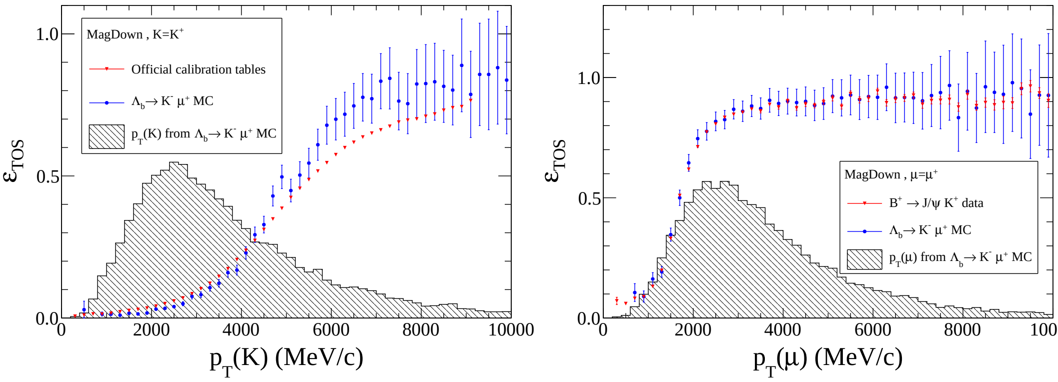
The HLT efficiency is determined in signal MC according to $\varepsilon^{HLT} = N_{\mathrm{trig}}/N_{L0}$. N_{L0} is the number of correctly reconstructed signal events after the L0 trigger and N_{trig} the number of correctly reconstructed signal events requiring all signal triggers. The results are $\varepsilon^{HLT}(2011) = (64.61 \pm 0.14)\,\%$ and $\varepsilon^{HLT}(2012) = (70.62 \pm 0.10)\,\%$ with binomial uncertainties. The overall trigger efficiency $\varepsilon_{\mathrm{Trig}} = \varepsilon^{L0} \cdot \varepsilon^{HLT}$ of $\Lambda_b \to pK^-$ events is $\varepsilon_{\mathrm{Trig}}(2011) = (38.6 \pm 2.1)\,\%$ and $\varepsilon_{\mathrm{Trig}}(2012) = (38.6 \pm 1.4)\,\%$. For comparison, the overall trigger efficiency $\varepsilon_{\mathrm{Trig}} = N_{\mathrm{trig}}/N_{\mathrm{sel}}$ from simulated events is about 41.1 % in 2011 MC and 42.6 % in 2012 MC. This corresponds to a relative deviation of about 7 % for 2011 data and 10 % for 2012 data.

6.5.2 $\Lambda_b \to K^-\mu^+$ trigger efficiency

For the determination of $\varepsilon_{\mathrm{Trig}}(\Lambda_b \to K^-\mu^+)$ the same strategy as for $\Lambda_b \to pK^-$ events is used.

L0-Trigger efficiency

The L0-TOS efficiency of the kaons is determined using the the official calibration data as in case of $\Lambda_b \to pK^-$. For the muons calibration tables of $\varepsilon_{TOS}(p_{\mathrm{T}})$ are created applying the *TISTOS* method on $B^+ \to J/\psi K^+$ events, see Appendix A.26. Figure 6.5 shows the differences between data and MC. There is no significant deviation in case of the muons, while the kaon efficiency shows a visible discrepancy.



(a) (b)

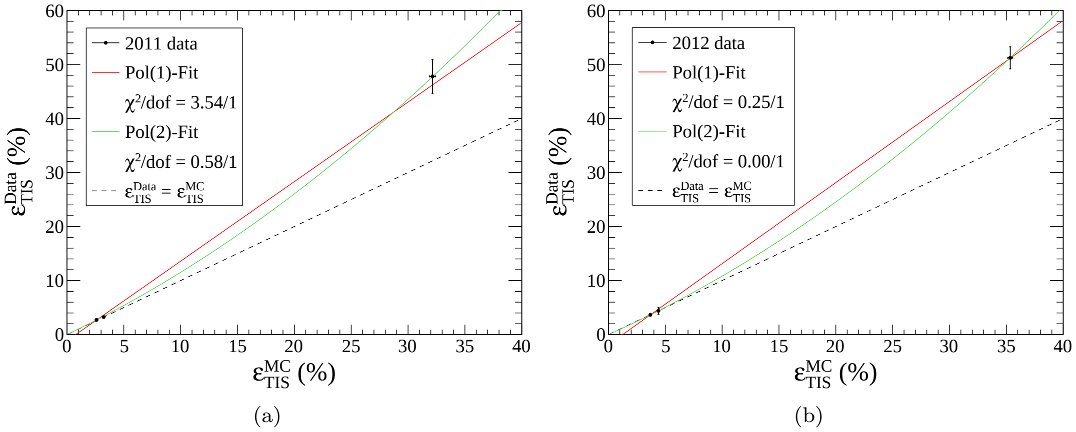
Figure 6.5: Comparison of L0-Muon-TOS efficiency in 2012 MC and 2012 data. The hatched histograms show the p_{T} distribution from simulated events with arbitrary scaling.

Table 6.8 shows the resulting L0-TOS efficiencies for kaons and muons in dependence of charge and magnet polarity. The average values take into account the different sizes of the data samples per magnet polarity and assumes an equal production of Λ_b and $\overline{\Lambda}_b$. The calculation of the L0-TOS efficiency includes the correlation of $p_{\mathrm{T}}(K)$ and $p_{\mathrm{T}}(\mu)$. These are $R(2011) = (0.889 \pm 0.006)$ and $R(2012) = (0.864 \pm 0.007)$ in the signal MC samples of the respective years. In conclusion the total L0-TOS efficiencies for both years are $\varepsilon_{TOS}(2011) = (86.6 \pm 0.6)\,\%$ and $\varepsilon_{TOS}(2012) = (83.8 \pm 0.4)\,\%$.

Table 6.8: Weighted L0-TOS efficiencies of K and μ for $\Lambda_b \to K^-\mu^+$ events.

Year	$\varepsilon_{TOS}(\%)$	K^+	K^-	μ^+	μ^-
2011	MagDown	22.81 ± 0.28	23.40 ± 0.29	80.0 ± 1.4	80.1 ± 1.4
	MagUp	22.92 ± 0.35	23.28 ± 0.34	79.7 ± 1.7	79.6 ± 1.7
2011	Average	23.10 ± 0.16		79.9 ± 0.8	
2012	MagDown	20.15 ± 0.14	20.27 ± 0.14	77.5 ± 1.0	77.1 ± 1.0
	MagUp	19.85 ± 0.14	20.71 ± 0.15	76.7 ± 1.0	77.0 ± 1.0
2012	Average	20.25 ± 0.07		77.1 ± 0.5	

The determination of the TIS efficiency $\varepsilon_{TIS} = (N_{TIS} - N_{TISTOS})/N_{TOS}$ for $\Lambda_b \to K^-\mu^+$ is done with the help of the control channels $B^+ \to J/\psi K^+$, $\Lambda_b \to J/\psi pK^+$, and $\Lambda_b \to pK^-$ due to the fact that this efficiency is in general not correctly simulated as it was seen in the $\Lambda_b \to pK^-$ case above. For each of these channels and for each year the TIS efficiency is measured in data and signal MC. As a result there are three data points $(\varepsilon_{TIS}^{MC}, \varepsilon_{TIS}^{Data})$ for each year. A calibration curve $\varepsilon_{TIS}^{Data} = f(\varepsilon_{TIS}^{MC})$ is fitted to the data points that allows to assess $\varepsilon_{TIS}^{Data}(\Lambda_b \to K^-\mu^+)$ from the TIS efficiency in MC. Two functional forms show a good agreement with the data: the first one is a linear function $f_1(x) = m\,x + n$ and the other one is a second-order polynomial of the form $f_2(x) = a\,x^2 + b\,x$, where m, n, a, b are fit parameters. Figure 6.6 shows the data and the fits for both years. As a result the second-order polynomial has better conformity with the data. In the simulation of $\Lambda_b \to K^-\mu^+$ the efficiencies are $\varepsilon_{TIS} = (4.05 \pm 0.07)\,\%$ in 2011 and $\varepsilon_{TIS} = (4.86 \pm 0.10)\,\%$ in 2012. The calculated efficiencies using f_2 are $\varepsilon_{TIS} = (4.31 \pm 0.11(\text{stat}) \pm 0.52(\text{syst}))\,\%$ in 2011 and $\varepsilon_{TIS} = (4.90 \pm 0.11(\text{stat}) \pm 0.52(\text{syst}))\,\%$ in 2012. The statistical uncertainty includes the uncertainty and covariance of the fit parameters and the uncertainty of ε_{TIS} due to the limited size of the MC sample, and the systematic error comes from the difference $|f_2(\varepsilon) - f_1(\varepsilon)|$. In App. A.27 the underlying data can be found. The resulting TIS efficiencies are $\varepsilon_{TIS}^{Data} = (4.3 \pm 0.5)\,\%$ in 2011 and $\varepsilon_{TIS}^{Data} = (4.9 \pm 0.5)\,\%$ in 2012. The total L0 trigger efficiencies are $\varepsilon^{L0}(2011) = (90.3 \pm 0.8)\,\%$ and $\varepsilon^{L0}(2012) = (87.9 \pm 0.6)\,\%$.



(a) (b)

Figure 6.6: Comparison of L0-Muon-TOS efficiency in 2012 MC and 2012 data.

HLT efficiency and total trigger efficiency

The HLT efficiency is determined in signal MC by dividing the number of correctly reconstructed events after the L0 trigger by the number of correctly reconstructed events after all signal triggers. These efficiencies are $\varepsilon^{HLT}(2011) = (70.66 \pm 0.15)\,\%$ and $\varepsilon^{HLT}(2012) = (79.04 \pm 0.16)\,\%$. The total trigger efficiency $\varepsilon_{\text{Trig}} = \varepsilon^{L0} \cdot \varepsilon^{HLT}$ for both years is $\varepsilon_{\text{Trig}}(2011) = (63.8 \pm 0.6)\,\%$ and $\varepsilon_{\text{Trig}}(2012) = (69.5 \pm 0.5)\,\%$.

6.6 Additional systematic uncertainties

6.6.1 $N(\Lambda_b \to pK^-)$

The systematic uncertainty on $N(\Lambda_b \to pK^-)$ due to the fixed parameters determined from MC is estimated by repeating the fit to the $m(pK^-)$ distribution and allow one of these parameters to float as well as the normalisation $N(\Lambda_b \to pK^-)$. These parameters are all parameters of the $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ fit model to assess uncertainties due the signal form. All other parameters are fixed to the result from the original fit shown in Fig. 6.1(a,b). In addition the original fit is repeated using a second order polynomial for the combinatorial background to estimate the impact of the more or less flat background. After the fits all deviations from the original value of $N(\Lambda_b \to pK^-)$ are added in quadrature. The final results of the determination of $N(\Lambda_b \to pK^-)$ in 2011 and 2012 data are:

$$N(\Lambda_b \to pK^-)_{2011} = 3442 \pm 106(\text{stat}) \pm 153(\text{syst}) = 3442 \pm 186 = 3442 \times (1 \pm 5.4\,\%)\ , \quad (6.12)$$

$$N(\Lambda_b \to pK^-)_{2012} = 7344 \pm 158(\text{stat}) \pm 113(\text{syst}) = 7344 \pm 194 = 7344 \times (1 \pm 2.6\,\%)\ . \quad (6.13)$$

6.6.2 Event selection efficiency

Systematic uncertainties on the event selection efficiency are due to differences between MC and real data in the selection variables and in the track reconstruction efficiency. The former uncertainty is superseded by the much tighter final signal selection whose systematic uncertainty has been determined in Sec. 5.3.5 and 5.4.3. Differences in the track reconstruction efficiency at LHCb are investigated in [68] and lead to the conclusion that the overall efficiency ratio $R = \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{MC}}}$ is compatible with unity for 2011 and 2012 MC. These studies have been made with the help of long tracks from $J/\psi \to \mu^+\mu^-$ decays using a tag-and-probe approach where the tag track is reconstructed using the complete LHCb tracking system and the reconstruction of the probe track does not involve information from at least one of the stations whose tracking efficiency is investigated. As a result of the studies there are tables provided that allow to measure R and its uncertainty in dependence of p and η with the help of simulated events. A correlation of R with the charge of the track and the magnet polarity is not observed.

Table 6.9 shows the combined ratios $R = R(K^-) \times R(\mu^+)$ and $R = R(p) \times R(K^-)$ that have been determined with the corresponding signal MC from 2011 and 2012. The value $R(t)$ of a particular track t is determined as the weighted mean $R(t) = \frac{\sum_i R_i \cdot N_i}{\sum_i N_i}$ summing over the individual ratios R_i and the number of MC events N_i in bins i of p and η. For muons the uncertainty of R is calculated as weighted mean of $\delta(R_i)$ taken from the tables. For hadrons the value of $\delta(R_i)$ is scaled with a factor of three to consider uncertainties on the tracking efficiency that arise from the imperfect simulation of hadronic interaction and large-angle scattering with the detector material [68].

Table 6.9: Ratio of tracking efficiencies from real data and signal MC.

$R(\%)$	$\Lambda_b \to K^-\mu^+$	$\Xi_b^0 \to K^-\mu^+$	$\Lambda_b \to pK^+$
2011	99.5 ± 0.7	99.5 ± 0.7	99.5 ± 1.0
2012	100.4 ± 0.8	100.3 ± 0.8	100.3 ± 1.1

The value and the uncertainty of R is combined with the original event selection efficiencies $\varepsilon'_{\text{sel}}$ that are summarised in Sec. 6.4.3 according to $\varepsilon_{\text{sel}} = \varepsilon'_{\text{sel}} \cdot R$ and quadratic summation of the uncertainties. The results are shown in Tab. 6.10.

Table 6.10: Event selection efficiencies including the tracking efficiency correction.

$\varepsilon_{\text{sel}}(\%)$	$\Lambda_b \to K^-\mu^+$	$\Xi_b^0 \to K^-\mu^+$	$\Lambda_b \to pK^-$
2011	30.79 ± 0.29	32.02 ± 0.28	22.69 ± 0.23
2012	28.75 ± 0.26	29.95 ± 0.27	21.26 ± 0.24

6.7 Summary

Table 6.11 summarises the individual efficiencies of $\Lambda_b \to K^-\mu^+$, $\Xi_b^0 \to K^-\mu^+$, and $\Lambda_b \to pK^+$ for both years. The values for $\Xi_b^0 \to K^-\mu^+$ are obtained in the same way as for $\Lambda_b \to K^-\mu^+$ but using simulated $\Xi_b^0 \to K^-\mu^+$ events. The uncertainties on the total efficiency results from quadratic combination of the individual errors. In comparison the efficiency of $\Lambda_b \to pK^+$ is larger due to the good signal-to-background ratio in the data making only soft selections necessary. The $K^-\mu^+$ final state benefits from a higher trigger and event selection efficiency but suffers a loss from the final signal selection.

6.7.1 Overall selection efficiency

The selection efficiencies of both years are combined using the efficiency (ε') and the number of events (N') from the control channel according to

$$\varepsilon_{\text{tot}} = \frac{\left(\frac{N'_{11}}{\varepsilon'_{11}} \cdot \varepsilon_{11} + \frac{N'_{12}}{\varepsilon'_{12}} \cdot \varepsilon_{12}\right)}{\left(\frac{N'_{11}}{\varepsilon'_{11}} + \frac{N'_{12}}{\varepsilon'_{12}}\right)} . \tag{6.14}$$

The resulting efficiencies are $\varepsilon_{\text{tot}}(\Lambda_b \to K^-\mu^+) = (0.61 \pm 0.07)\,\%$, $\varepsilon_{\text{tot}}(\Xi_b^0 \to K^-\mu^+) = (0.74 \pm 0.08)\,\%$, and $\varepsilon_{\text{tot}}(\Lambda_b \to pK^-) = (1.38 \pm 0.07)\,\%$

Table 6.11: Summary of the efficiencies, all numbers are in per cent.

$\varepsilon(\%)$	$\Lambda_b \to K^-\mu^+$	$\Xi_b^0 \to K^-\mu^+$	$\Lambda_b \to pK^+$
2011			
Acceptance	18.68 ± 0.05	18.52 ± 0.05	19.55 ± 0.05
Event selection	30.79 ± 0.29	32.02 ± 0.28	22.69 ± 0.23
Trigger	63.8 ± 0.6	64.8 ± 0.5	38.6 ± 2.1
Signal selection	15.0 ± 2.9	17.1 ± 3.3	87.4 ± 3.9
Total efficiency	0.55 ± 0.11	0.66 ± 0.13	1.50 ± 0.11
2012			
Acceptance	18.92 ± 0.06	18.83 ± 0.05	19.90 ± 0.01
Event selection	28.75 ± 0.26	29.95 ± 0.27	21.26 ± 0.24
Trigger	69.5 ± 0.5	71.6 ± 0.4	38.6 ± 1.4
Signal selection	16.7 ± 2.1	19.0 ± 2.4	81.5 ± 2.5
Total efficiency	0.63 ± 0.08	0.77 ± 0.10	1.33 ± 0.07

6.7.2 Normalisation factor

Starting from Eq. 6.1 a normalisation factor α is introduced combining the results and uncertainties from the normalisation according to

$$\frac{f_{\text{sig}}}{f_{\text{norm}}} \times \frac{\mathcal{B}_{\text{sig}}}{\mathcal{B}_{\text{norm}}} = N_{\text{sig}} \times \underbrace{\frac{1}{\varepsilon_{\text{sig}}} \times \frac{\varepsilon_{\text{norm}}}{N_{\text{norm}}}}_{\alpha} \, . \tag{6.15}$$

Using the abbreviations $N_{\text{norm}} = N'$, $\varepsilon_{\text{norm}} = \varepsilon'$, and $\varepsilon_{\text{sig}} = \varepsilon$ the normalisation factor of the combined data can be calculated as

$$\alpha = \frac{1}{\varepsilon} / \frac{N'}{\varepsilon'} = \frac{\left(\frac{N'_{11}}{\varepsilon'_{11}} + \frac{N'_{12}}{\varepsilon'_{12}}\right)}{\left(\frac{N'_{11}}{\varepsilon'_{11}} \cdot \varepsilon_{11} + \frac{N'_{12}}{\varepsilon'_{12}} \cdot \varepsilon_{12}\right)} / \left(\frac{N'_{11}}{\varepsilon'_{11}} + \frac{N'_{12}}{\varepsilon'_{12}}\right) \, , \tag{6.16}$$

$$\alpha = \left(\frac{N'_{11}}{\varepsilon'_{11}} \cdot \varepsilon_{11} + \frac{N'_{12}}{\varepsilon'_{12}} \cdot \varepsilon_{12}\right)^{-1} \, . \tag{6.17}$$

The normalisation factor of the two signal channels are $\alpha(\Lambda_b \to K^-\mu^+) = (2.10 \pm 0.25) \cdot 10^{-4}$ and $\alpha(\Xi_b^0 \to K^-\mu^+) = (1.73 \pm 0.21) \cdot 10^{-4}$ including a quadratic summation of the individual uncertainties.

6.7.3 Single-event sensitivity

The single-event sensitivity, abbreviated here as β, is an experimental factor that can be interpreted as the branching ratio that is needed to see one signal event. Its value depends on the total number of initial state particles in the data (N_0) and the total selection efficiency ($\varepsilon_{\mathrm{sig}}$)

$$\beta = \frac{\mathcal{B}_{\mathrm{sig}}}{N_{\mathrm{sig}}} = \frac{1}{\varepsilon_{\mathrm{sig}} \cdot N_0} \ . \tag{6.18}$$

β can be written in dependence of α and the external parameters according to

$$\beta = \alpha \times \mathcal{B}_{\mathrm{norm}} \times \frac{f_{\mathrm{norm}}}{f_{\mathrm{sig}}} \ . \tag{6.19}$$

The single-event sensitivity for $\mathit{\Lambda}_b \to K^- \mu^+$ is determined using the present world-average value $\mathcal{B}(\mathit{\Lambda}_b \to pK^-) \approx 5.1 \cdot 10^{-6}$ [3] and $f_{\mathrm{sig}}/f_{\mathrm{norm}} = f_{\mathit{\Lambda}_b}/f_{\mathit{\Lambda}_b} = 1$ to be $\beta \approx 1.1 \cdot 10^{-9}$. The single-event sensitivity for $\mathit{\Xi}_b^0 \to K^- \mu^+$ can be expected to be smaller due to the suppressed $\mathit{\Xi}_b^0$ production w.r.t. $\mathit{\Lambda}_b$ baryons.

“Oh, listen, just one more thing.”
— Inspector Columbo (from the American TV series “Columbo”)

7

Results and conclusions

7.1 Upper limit on the ratio of branching fractions

An upper limit on the ratio of branching fractions

$$\tilde{R} = \frac{\mathcal{B}(\Lambda_b \to K^-\mu^+)}{\mathcal{B}(\Lambda_b \to pK^-)} = \tilde{N} \cdot \tilde{\alpha} \tag{7.1}$$

is calculated using a Bayesian approach. This ratio is the most precise result that can be derived from this analysis since the numerical value of $\mathcal{B}(\Lambda_b \to pK^-)$ and its uncertainty has to be taken from another measurement. $\tilde{R}, \tilde{N}$ and $\tilde{\alpha}$ denote the true and unknown value of the ratio, the number of signal events and the normalisation factor, respectively. The two-dimensional likelihood of $\tilde{N}$ and $\tilde{\alpha}$ results from a factorisation ansatz assuming that both values are independent from each other, *i.e.* $l(\tilde{N}, \tilde{\alpha}|\vec{n}, \alpha, \delta_\alpha) = l_1(\tilde{N}|\vec{n}) \cdot l_2(\tilde{\alpha}|\alpha, \delta_\alpha)$, where $\vec{n}$ describes the observed number of events per bin in the $m(K^-\mu^+)$ distribution and α and δ_α the measured normalisation factor and its uncertainty. $l_1(\tilde{N}|\vec{n})$ is described by the measured likelihood in dependence of $N(\Lambda_b/\Xi_b^0 \to K^-\mu^+)$ from the fit to the $m(K^-\mu^+)$ distribution. $l_2(\tilde{\alpha}|\alpha, \delta_\alpha)$ is described by a Gaussian that uses α as mean value and δ_α as standard deviation. This is motivated by the fact that the normalisation factor α absorbs a number of measured values and hence can be approximated by a Gauss distribution according to the central limit theorem. Figure 7.1(a) shows the distribution of $l(\tilde{N}, \tilde{\alpha}|\vec{n}, \alpha, \delta_\alpha)$. The likelihood $l(\tilde{R} = \tilde{N} \cdot \tilde{\alpha}|\vec{n}, \alpha, \delta_\alpha)$ is determined by MC integration along the hyperbola $\tilde{R} = \tilde{N} \cdot \tilde{\alpha}$ and shown in Fig. 7.1(b). To obtain a Bayesian confidence level the likelihood is converted into an *a posteriori* probability density

$$f(\tilde{R}|\vec{n}, \alpha, \delta_\alpha) = \frac{l(\tilde{R}|\vec{n}, \alpha, \delta_\alpha)}{\int l(R|\vec{n}, \alpha, \delta_\alpha)\, f(R)\, dR} \cdot f(\tilde{R}) \; . \tag{7.2}$$

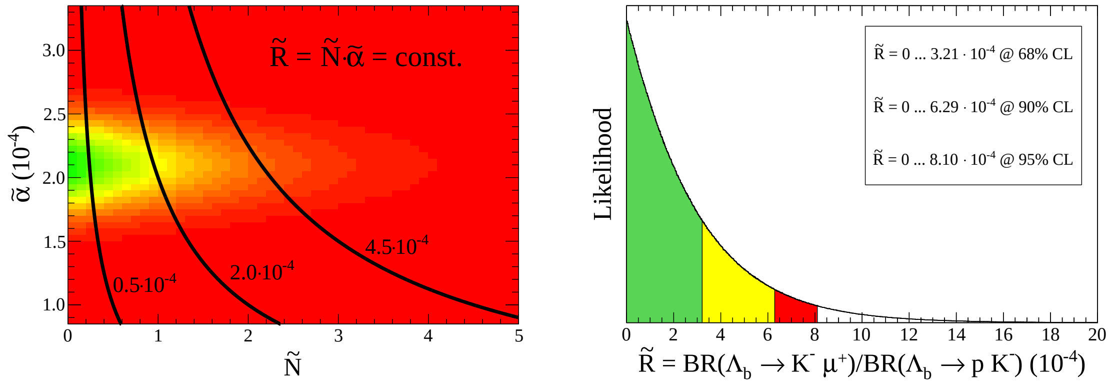
A *subjective* probability density $c(\tilde{R}|\vec{n}, \alpha, \delta_\alpha)$ [66] is defined by assuming that the prior probability $f(\tilde{R})$ is constant and non vanishing for $\tilde{R} \geq 0$ and $f(\tilde{R}) = 0$ everywhere else. This corresponds to the assumption that nothing is known about the value of $\tilde{R}$.

$$c(\tilde{R}|\vec{n}, \alpha, \delta_\alpha) = \frac{l(\tilde{R}|\vec{n}, \alpha, \delta_\alpha)}{\int l(R|\vec{n}, \alpha, \delta_\alpha)\, dR} \tag{7.3}$$

The confidence level (CL) of an upper limit $\tilde{R}_{\mathrm{up}}$ is calculated according to

$$\mathrm{CL}(\tilde{R}_{\mathrm{up}}) = \int\limits_0^{\tilde{R}_{\mathrm{up}}} c(\tilde{R}|\vec{n},\alpha,\delta_\alpha)\, d\tilde{R}\ . \tag{7.4}$$

The 90 % confidence interval of $\tilde{R}$ is $0 \ldots 6.29 \cdot 10^{-4}$. The upper limit rises to $6.30 \cdot 10^{-4}$ if the uncertainty δ_α is increased by 25 %. If one neglects the uncertainty δ_α, $\tilde{R}_{\mathrm{up}}$ is not smeared by the uncertainty on α and is simply $\tilde{R}_{\mathrm{up}} = 2.99 \times 2.10 \cdot 10^{-4} = 6.28 \cdot 10^{-4}$ @ CL = 90 %. This shows that the uncertainty on α has little influence and the upper limit is dominated by the distribution of $l_1(\tilde{N}|\vec{n})$.



(a) (b)

Figure 7.1: Likelihood distribution of $l(\tilde{N},\tilde{\alpha}|\vec{n},\alpha,\delta_\alpha)$ (a) and $l(\tilde{R}|\vec{n},\alpha,\delta_\alpha)$ (b).

The same method has been repeated for the upper limit on

$$\tilde{R}' = \frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \to K^-\mu^+)}{\mathcal{B}(\Lambda_b \to pK^-)} = \tilde{N}' \cdot \tilde{\alpha}'\ , \tag{7.5}$$

using the likelihood distribution for $N(\Xi_b^0 \to K^-\mu^+)$ from the fit to the data and $\alpha' = \alpha(\Xi_b^0 \to K^-\mu^+) = (1.73 \pm 0.21) \cdot 10^{-4}$. The resulting upper limits are $1.99 \cdot 10^{-4}$, $4.02 \cdot 10^{-4}$, and $5.26 \cdot 10^{-4}$ at the confidence levels of 68 %, 90 %, and 95 %, respectively. In summary the 90 % confidence intervals are

$$\frac{\mathcal{B}(\Lambda_b \to K^-\mu^+)}{\mathcal{B}(\Lambda_b \to pK^-)} < 6.29 \cdot 10^{-4} \quad , \quad \frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \to K^-\mu^+)}{\mathcal{B}(\Lambda_b \to pK^-)} < 4.02 \cdot 10^{-4}\ . \tag{7.6}$$

7.2 $\mathcal{B}(\Lambda_b \to K^-\mu^+)$ and $\mathcal{B}(\Xi_b^0 \to K^-\mu^+)$

A statement about the absolute branching fraction necessarily requires knowledge about $\mathcal{B}(\Lambda_b \to pK^-)$. Currently there is only one published value $\mathcal{B}(\Lambda_b \to pK^-) = (5.6 \pm 0.8(\mathrm{stat}) \pm 1.5(\mathrm{syst})) \cdot 10^{-6}$ [69] by CDF using $1\,\mathrm{fb}^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$. The result includes a large uncertainty from the used fragmentation ratio $f_{\Lambda_b}/f_d = 0.230 \pm 0.052$ of b quarks into Λ_b and B^0 that was published by the Particle Data Group in 2008. This ratio is based on measurements in $Z \to b\bar{b}$ events at the Z resonance. However, subsequent results by CDF and LHCb indicate different fragmentation rates in hadron colliders, *e.g.* $f_{\Lambda_b}/f_d \approx 40\,\%$ at $\sqrt{s}(p\bar{p}) = 1.96\,\mathrm{TeV}$, (see also Sec. "Production and decay of b-flavored hadrons" in [3]) which compromises the stated $\mathcal{B}(\Lambda_b \to pK^-)$ result. Therefore the value of $\mathcal{B}(\Lambda_b \to pK^-)$ is recalculated using the unbiased ratio

$$\frac{f_{\Lambda_b}}{f_d} \times \frac{\mathcal{B}(\Lambda_b \to pK^-)}{\mathcal{B}(B^0 \to K^+\pi^-)} = 0.066 \pm 0.009(\text{stat}) \pm 0.008(\text{syst}) \quad [69]\,, \tag{7.7}$$

$\mathcal{B}(B^0 \to K^+\pi^-) = (1.96 \pm 0.05) \cdot 10^{-5}$ [3], and $\Gamma(\bar{b} \to b \text{ baryon})/\left[\Gamma(\bar{b} \to B^+) + \Gamma(\bar{b} \to B^0)\right] = 0.21^{+0.08}_{-0.06} \pm 0.01$ [3]. The latter results is based on a measurement by CDF at $\sqrt{s}(p\bar{p}) = 1.96\,\text{TeV}$ from 2008 and is approximately the same as $f_{\Lambda_b}/(2f_d)$ assuming isospin symmetry and a negligible fraction of other b baryons than Λ_b. Combining these numbers, the resulting branching fraction is

$$\mathcal{B}(\Lambda_b \to pK^-) = [0.066 \pm 0.012] \times \left[(1.96 \pm 0.05) \cdot 10^{-5}\right] \div \left[0.42^{+0.16}_{-0.12}\right] \;, \tag{7.8}$$

$$\mathcal{B}(\Lambda_b \to pK^-) = (3.1^{+1.3}_{-1.0}) \cdot 10^{-6} \;. \tag{7.9}$$

The result is checked by an independent approach calculating the branching ratio directly from the LHCb results on $\Lambda_b \to pK^-$ taking into account $N(\Lambda_b) \approx 50 \cdot 10^9$ (see Sec 2.3), $N_{\text{obs}}(\Lambda_b \to pK^-) \approx 10\,800$, and the total signal efficiency $\varepsilon \approx 1.38\,\%/20\,\% = 6.9\,\%$ within detector acceptance,

$$\mathcal{B}(\Lambda_b \to pK^-) = \frac{N(\Lambda_b \to pK^-)}{\varepsilon \cdot N(\Lambda_b)} \approx 3.1 \cdot 10^{-6} \;.$$

This estimate is consistent with Eq. 7.9. However, the result has a relative uncertainty of not less than 31 % due to $N(\Lambda_b) \sim f_{\Lambda_b}/f_d \times \sigma(pp \to B^0 + X)$ with relative errors of 27 % and 16 %, respectively. Figure 7.2 shows the different upper limits (ULs) on $\mathcal{B}(\Lambda_b \to K^-\mu^+)$ in dependence of $\mathcal{B}(\Lambda_b \to pK^-)$.

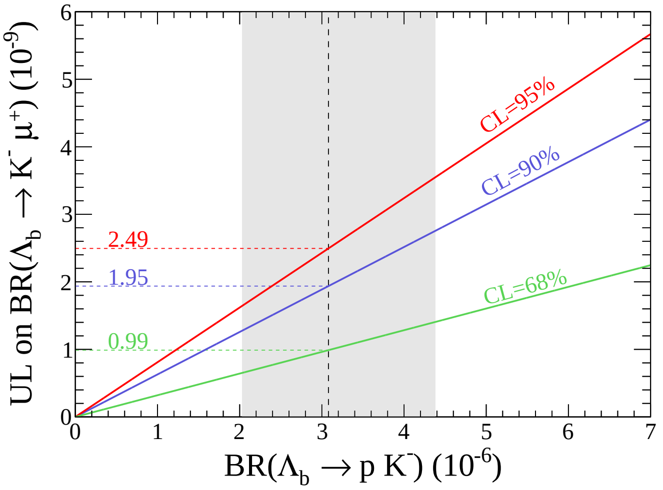


Figure 7.2: Upper limits on $\mathcal{B}(\Lambda_b \to K^-\mu^+)$ in dependence of $\mathcal{B}(\Lambda_b \to pK^-)$.

Since it cannot be distinguished between $\Lambda_b \to K^-\mu^+$ and $\Lambda_b \to K^+\mu^-$ the final result is

$$\left[\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.95 \cdot 10^{-9} \quad \text{at CL} = 90\,\% \;, \tag{7.10}$$

where the large uncertainty on $\mathcal{B}(\Lambda_b \to pK^-)$ is excluded. Concerning the Ξ_b^0 decay it can be stated

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \left[\mathcal{B}(\Xi_b^0 \to K^-\mu^+) + \mathcal{B}(\Xi_b^0 \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.25{\cdot}10^{-9} \text{ at CL} = 90\,\% \,. \quad (7.11)$$

Here, the large uncertainty on $\mathcal{B}(\Lambda_b \to pK^-)$ is excluded.

7.3 Search for $\Xi_b^0 \to pK^-$

The high number of about eleven thousand $\Lambda_b \to pK^-$ events in the data sample offers the chance to search for the decay $\Xi_b^0 \to pK^-$ that has not been observed yet. Due to the large background level in the Ξ_b^0 mass region around 5790 MeV/c^2 in the fit to the $m(pK^-)$ distribution, see Fig. 6.1, the cuts on $\mathrm{x_{BDT}}$, ProbNN(p), and ProbNN(K^-) are tightened to be sensitive on a presumably small number of $\Xi_b^0 \to pK^-$ events. The binned likelihood fit to the mass spectrum includes a decay model for $\Xi_b^0 \to pK^-$ events which is the same as for $\Lambda_b \to pK^-$ but with parameters fixed from a fit to simulated $\Xi_b^0 \to pK^-$ events. Figure 7.3 shows a fit to the $m(pK^-)$ spectrum of combined 2011 and 2012 data after a loose optimisation of $S/\sqrt{B}$ of the $\Lambda_b \to pK^-$ signal leading to the selections $\mathrm{x_{BDT}}(2011) > 0.05$, $\mathrm{x_{BDT}}(2012) > 0.10$, ProbNN($p$)$> 0.70$, and ProbNN($K^-$)$> 0.70$. The significance of the $\Xi_b^0 \to pK^-$ signal is determined by a likelihood-ratio test [66] that compares the log-likelihoods of the fit with (L_1) and without (L_0) signal hypothesis. The fit without signal hypothesis includes the constraint $N(\Xi_b^0 \to pK^-) = 0$ and has one degree of freedom less. The results are $N(\Lambda_b \to pK^-) = 3\,796 \pm 67$, $N(\Xi_b^0 \to pK^-) = 25 \pm 10$, and $\chi = \sqrt{2(L_1 - L_0)} = 2.7$. In summary the significance is too low to claim an evidence but can be interpreted as an indication for $\Xi_b^0 \to pK^-$.

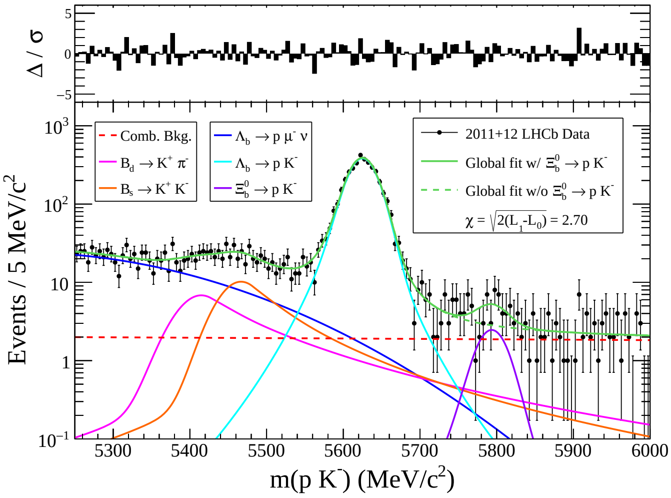


Figure 7.3: Fit to the combined $m(pK^-)$ distribution of 2011 and 2012 data after selections.

7.4 Conclusions

A search for the decays $\Lambda_b \to K^-\mu^+$ and $\Xi_b^0 \to K^-\mu^+$ has been presented using LHCb data from 2011 and 2012 corresponding to an integrated luminosity of $3\,\text{fb}^{-1}$ and containing about $50 \cdot 10^9$ Λ_b decays. The analysis is motivated by the matter-antimatter asymmetry in our universe that might be explained by baryon and lepton number violation in b decays with a probability in the order of 10^{-10}. In summary no signal events are observed and the following upper limits are determined at a confidence level of 90 %

$$\frac{\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)}{\mathcal{B}(\Lambda_b \to pK^-)} < 6.29 \cdot 10^{-4} \tag{7.12}$$

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \frac{\mathcal{B}(\Xi_b^0 \to K^-\mu^+) + \mathcal{B}(\Xi_b^0 \to K^+\mu^-)}{\mathcal{B}(\Lambda_b \to pK^-)} < 4.02 \cdot 10^{-4} \tag{7.13}$$

$$\left[\mathcal{B}(\Lambda_b \to K^-\mu^+) + \mathcal{B}(\Lambda_b \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.95 \cdot 10^{-9} \tag{7.14}$$

$$\frac{f_{\Xi_b^0}}{f_{\Lambda_b}} \times \left[\mathcal{B}(\Xi_b^0 \to K^-\mu^+) + \mathcal{B}(\Xi_b^0 \to K^+\mu^-)\right] \times \frac{3.1 \cdot 10^{-6}}{\mathcal{B}(\Lambda_b \to pK^-)} < 1.25 \cdot 10^{-9} \tag{7.15}$$

In addition an indication for the yet unobserved decay $\Xi_b^0 \to pK^-$ is found with 25 ± 10 signal events and a significance of $\chi = 2.7$.

The results concerning $\Lambda_b \to K^-\mu^+$ are the most sensitive upper limits on b hadron decays including the violation of baryon and lepton number conservation so far. The reason for this is on one side the high Λ_b production rate and the large trigger and reconstruction efficiency at LHCb. On the other side the signal region is mostly populated by combinatorial background that can be removed by applying an optimised multivariate selection. A blind analysis was conducted whereby it was made use of simulated $\Lambda_b \to K^-\mu^+$ decays and wrong-sign $K^+\mu^+$ combinations from real data to maximise the sensitivity for a signal and without being biased by the data in the signal region. In addition a number of control channels were studied in MC and real data to validate the LHCb event and detector simulation. In particular the known and similar decay $\Lambda_b \to pK^-$ has been investigated to act as normalisation mode and check the search strategy. As a result the achieved single-event sensitivity on $\Lambda_b \to K^-\mu^+$ events is about $6 \cdot 10^{-10}$, *i.e.* on average one event is visible if $\mathcal{B}(\Lambda_b \to K^-\mu^+) = 6 \cdot 10^{-10}$.

An improvement of the upper limit can be expected using the Run II data of LHCb at $\sqrt{s}(pp) = 13\,\text{TeV}$. Preliminary studies show that the recorded number of b hadrons has approximately doubled per $1\,\text{fb}^{-1}$. Thus, taking into account the anticipated integrated luminosity of $5\,\text{fb}^{-1}$ during Run II, the use of the combined data increases the sensitivity by a factor of about four.

Due to the high production rate of b and c quarks at LHCb there are a number of other conceivable decay modes for BLNV that are interesting to search for, for example $\Lambda_b \to D^-\mu^+$, $\Lambda_b \to D_s^-\mu^+$, $\overline{B}^0 \to \Lambda_c\mu^-$, $B^- \to \Lambda\mu^-$ and $\Lambda_c \to \mu^+\mu^-\mu^+$. A widening of the spectrum of investigated channels is reasonable since the mechanism, if any, behind BLNV is unclear and might actually not allow $\Lambda_b \to K^-\mu^+$ but one of these decays. However, due to the large Λ_b production rate at LHCb the search for BLNV in B decays is problematic since semileptonic decays of the sorts $\Lambda_b \to \mathsf{B}\,\mathsf{L}\,\nu$ and $\Lambda_b \to \mathsf{B}\,\mathsf{L}\,\overline{\mathsf{L}}$ contaminate the signal region of $B \to \mathsf{B}\,\mathsf{L}$ decays in the $m(\mathsf{B}\,\mathsf{L})$ spectrum. Therefore a dedicated high-luminosity B experiment operating at the $\Upsilon(4S) \to B\overline{B}$ production threshold like Belle II might be more suitable for these searches.

A

Appendix

A.1 Planck's law

In 1900 Max Planck derived a formula for the energy spectrum of an ideal black body absorbing all light and emitting photons due to his finite temperature. The number of photons per volume and in dependence of the frequency is

$$n(f) = \frac{8\pi}{c^3} \cdot \frac{f^2}{\exp[hf/(k_\mathrm{B}T)] - 1} \ ,$$

and the energy per volume is

$$\varepsilon(f) = n(f) \cdot hf = \frac{8\pi}{c^3} \cdot \frac{hf^3}{\exp[hf/(k_\mathrm{B}T)] - 1} \ .$$

By introducing the dimensionless ratio $R = hf/(k_\mathrm{B}T)$ both distributions can be transformed into

$$n(R) = n(f) \cdot \left|\frac{df}{dR}\right| = \frac{8\pi}{h^3c^3} \cdot (k_\mathrm{B}\,T)^3 \cdot \frac{R^2}{\exp(R) - 1} \ ,$$

$$\varepsilon(R) = \varepsilon(f) \cdot \left|\frac{df}{dR}\right| = \frac{8\pi}{h^3c^3} \cdot (k_\mathrm{B}\,T)^4 \cdot \frac{R^3}{\exp(R) - 1} \ .$$

The overall densities can be calculated by using the identity

$$\int\limits_0^\infty \frac{R^{z-1}}{\exp(R) - 1} dR = \Gamma(z) \cdot \zeta(z) \quad \text{for } \mathrm{Re}(z) > 1 \ ,$$

where $\Gamma(z)$ is the gamma function and $\zeta(z)$ is the Riemann zeta function. For integer numbers $\Gamma(n) = (n-1)!$, $\zeta(4) = \pi^4/90$, and $\zeta(3) = 1.2020569\ldots$. Substituting h by $2\pi\hbar$ the total energy per volume is

$$\varepsilon_\gamma = \int\limits_0^\infty \varepsilon(R)\,dR = \frac{\pi^2}{15\,\hbar^3c^3} \cdot (k_\mathrm{B}\,T)^4 \approx 0.262 \cdot \left(\frac{T}{2.73\,\mathrm{K}}\right)^4 \ \mathrm{eV\ cm^{-3}} \ ,$$

and the total number of photons per volume is

$$n_\gamma = \int_0^\infty n(R)\,dR = \frac{2}{\pi^2\hbar^3c^3}\cdot\zeta(3)\cdot(k_\mathrm{B}\,T)^3 \approx 412\cdot\left(\frac{T}{2.73\,\mathrm{K}}\right)^3\ \mathrm{cm}^{-3}\ ,$$

The average photon energy is

$$\bar{E}_\gamma = \frac{\varepsilon_\gamma}{n_\gamma} = \frac{\pi^4}{30\cdot\zeta(3)}\cdot k_\mathrm{B}\,T = 2.7012\ldots\ k_\mathrm{B}\,T \approx 0.64\cdot\left(\frac{T}{2.73\,\mathrm{K}}\right)\ \mathrm{meV}\ .$$

A.2 Formula for η_{10}

The formula of $\eta_{10} = \eta\cdot 10^{10}$ can be found from the expected number of photons per volume n_γ and the cosmological baryon density $n_\mathrm{b} = \Omega_\mathrm{b}\,\rho_\mathrm{crit}/\langle m_\mathrm{b}\rangle$ that actually includes electrons, neutrons and protons. The average baryon mass is assumed to be $\langle m_\mathrm{b}\rangle \approx m_p$ neglecting the lowered mass of isotopes due to the binding energy. However, this effect is very small as shown in App. A.3.

$$\eta = \eta_{10}\cdot 10^{-10} = \frac{n_\mathrm{b}}{n_\gamma} = \frac{\Omega_\mathrm{b}\,\rho_\mathrm{crit}}{\langle m_\mathrm{b}\rangle\,n_\gamma} \approx \frac{\Omega_\mathrm{b}}{m_p}\cdot\frac{3\,H^2}{8\,\pi\,G}\cdot\frac{\pi^2\hbar^3\,c^3}{2\,\zeta(3)\,(k_\mathrm{B}\,T)^3}$$

$$\eta_{10} \approx \frac{\Omega_\mathrm{b}\,h^2}{T^3}\cdot\frac{3\pi\cdot 10^{10}}{16\,\zeta(3)\,G\,m_p}\cdot\left(\frac{\hbar c}{k_\mathrm{B}}\right)^3\cdot\left(100\,\frac{\mathrm{km}}{s\ \mathrm{Mpc}}\right)^2 \approx 273.5\,\Omega_\mathrm{b}\,h^2 \quad \text{for } T = 2.7255\,\mathrm{K}$$

A.3 Next-order approximation of $\langle m_\mathrm{b}\rangle$

The average baryon mass in the universe $\langle m_\mathrm{b}\rangle$ is determined considering two corrections. The first one accounts the lowered mass of isotopes due their binding energy E_b and the second one also includes the contributions of the two most abundant nuclei beside hydrogen.

$$\langle m_\mathrm{b}\rangle = \sum_i f_i\left(\frac{M_i - E_{\mathrm{b},i}}{A_i}\right)\quad ,\quad i = \mathrm{H}, \mathrm{He}, \mathrm{D}\ .$$

The fractions f_i can be calculated from the known mass fractions He/H = 24.7 % and D/H = $2.5\cdot 10^{-5}$ and the normalisation $\sum_i f_i \approx f_\mathrm{H} + f_\mathrm{He} + f_\mathrm{D} = f_\mathrm{H}\,(1 + \mathrm{He/H} + \mathrm{D/H}) = 1$. As a result the fractions are $f_\mathrm{H} = 80.2\,\%$, $f_\mathrm{He} = 19.8\,\%$, and $f_\mathrm{D} = 2.0\cdot 10^{-5}$. In summary the next-order approximation of $\langle m_\mathrm{b}\rangle$ is

$$\langle m_\mathrm{b}\rangle \approx 80.2\,\%\,[m_p]_\mathrm{H} + 19.8\,\%\left[\frac{2(m_p+m_n) - 28.3\,\mathrm{MeV}/c^2}{4}\right]_\mathrm{He} + 2.0\cdot 10^{-5}\left[\frac{(m_p+m_n) - 2.22\,\mathrm{MeV}/c^2}{2}\right]_\mathrm{D}\ ,$$

$$\langle m_\mathrm{b}\rangle \approx 80.2\,\%\left[938.3\,\mathrm{MeV}/c^2\right]_\mathrm{H} + 19.8\,\%\left[931.8\,\mathrm{MeV}/c^2\right]_\mathrm{He} + 2.0\cdot 10^{-5}\ \left[937.8\,\mathrm{MeV}/c^2\right]_\mathrm{D}\ ,$$

$$\langle m_\mathrm{b}\rangle \approx 937.0\,\mathrm{MeV}/c^2 = (1 - 1.3\cdot 10^{-3})\cdot m_p\ . \tag{A.1}$$

A.4 A toy model for Λ_b to achieve an excess of matter

The basic idea of the following model is to create an equivalent excess of baryons (B) and leptons (L) out of an initially equal amount of Λ_b and $\overline{\Lambda}_b$ by introducing a new class of decays like $\Lambda_b \to K^-\mu^+$ that violate $\mathbb{B}$ and $\mathbb{L}$ separately, but conserve $\mathbb{B}-\mathbb{L}$. The two considered classes are:

1. SM-type decays with $\Delta\mathbb{B} = \Delta\mathbb{L} = 0$ to produce B and $\overline{\mathsf{B}}$

$$\Lambda_b \to \mathsf{B} + X \quad , \quad x = \sum \lambda_f \, \tau(\Lambda_b)$$

$$\overline{\Lambda}_b \to \overline{\mathsf{B}} + X \quad , \quad \bar{x} = \sum \lambda_{\bar{f}} \, \tau(\overline{\Lambda}_b)$$

2. Exotic decays with $\Delta\mathbb{B} = \Delta\mathbb{L} = \pm 1$ to produce $\overline{\mathsf{L}}$ and L

$$\Lambda_b \to \overline{\mathsf{L}} + X \quad , \quad y = \sum \lambda_f \, \tau(\Lambda_b)$$

$$\overline{\Lambda}_b \to \mathsf{L} + X \quad , \quad \bar{y} = \sum \lambda_{\bar{f}} \, \tau(\overline{\Lambda}_b)$$

where:

- X stands for a generic amount of particles with $\mathbb{B} = 0$ and $\mathbb{L} = 0$
- $\sum \lambda_f$ is the sum over all partial decay rates of $\Lambda_b \to f$ etc.
- $\tau = 1/\sum_\infty \lambda_f$ the sum over all partial decay rates
- $x, y, \bar{x}, \bar{y}$ are inclusive branching ratios

From CPT invariance it follows that $\tau(\Lambda_b) = \tau(\overline{\Lambda}_b)$ and hence $x + y = \bar{x} + \bar{y} = 1$. Thus it can be stated $x - \bar{x} = \bar{y} - y$. An excess of matter requires $x > \bar{x}$ and $\bar{y} > y$ starting from equal amounts of Λ_b and $\overline{\Lambda}_b$. CP violation in the two decay classes can be expressed by the asymmetry parameters

$$a_{\mathsf{B}} = \frac{x - \bar{x}}{x + \bar{x}} = 0 \ldots 1 \quad \text{and} \quad a_{\mathsf{L}} = \frac{\bar{y} - y}{\bar{y} + y} = 0 \ldots 1 \; . \tag{A.2}$$

By introducing the abbreviations $\mathcal{B}(\Lambda_b \to \mathsf{B} + X) = x + \bar{x}$ and $\mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X) = y + \bar{y}$ for the inclusive branching fractions the equation $x - \bar{x} = \bar{y} - y$ can be rewritten as

$$\boxed{a_{\mathsf{B}} \cdot \mathcal{B}(\Lambda_b \to \mathsf{B} + X) = a_{\mathsf{L}} \cdot \mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X)} \; . \tag{A.3}$$

Equation A.3 can be depicted by a balanced see-saw as shown in Fig. A.1. The anticipated large value of $\mathcal{B}(\Lambda_b \to \mathsf{B} + X)$ in comparison to $\mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X)$ is compensated by a much larger overall CP violation in non-SM-type decays, *i.e.* $a_{\mathsf{L}} \gg a_{\mathsf{B}}$.

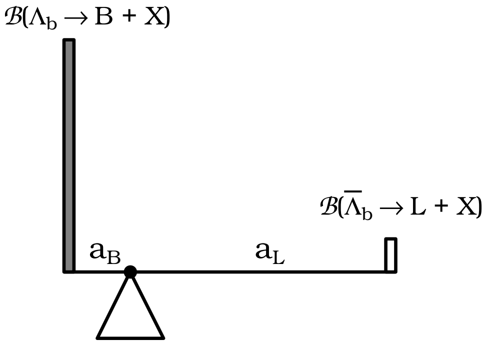


Figure A.1: Illustration of $a_\mathsf{B} \cdot \mathcal{B}(\Lambda_b \to \mathsf{B} + X) = a_\mathsf{L} \cdot \mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X)$ by a balanced see-saw.

A.4.1 Connection between η and $\mathcal{B}(\Lambda_b \to K^-\mu^+)$

A relation to $\eta = n_\mathsf{B}/n_\gamma$ can be found by the idea that Λ_b and $\overline{\Lambda}_b$ create photons by their mutual extinction $\Lambda_b\overline{\Lambda}_b \to \gamma\gamma$ and baryons by the asymmetry a_B in SM-type decays. This can be stated by the equation

$$\eta = \frac{n_\mathsf{B}}{n_\gamma} = \frac{n(\Lambda_b)\, a_\mathsf{B}\, \mathcal{B}(\Lambda_b \to \mathsf{B} + X)}{n(\Lambda_b)\,(1 - a_\mathsf{B})} \;. \tag{A.4}$$

By the use of Eq. A.3 and assuming $a_\mathsf{B} \ll 1$ it follows

$$\mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X) \approx \frac{\eta}{a_\mathsf{L}} \;. \tag{A.5}$$

This would mean that an observed baryon-photon-ratio η can be explained by a combination of non-SM-type decays of Λ_b and an overall CP violation that effectively prefers the creation of leptons. Since $a_\mathsf{L} \leq 1$ the sum of all these branching ratios is bigger as or equal to η. The particular decay $\mathcal{B}(\Lambda_b \to K^-\mu^+)$ is one of many conceivable decays of this kind, *i.e.* $\mathcal{B}(\Lambda_b \to K^-\mu^+) = f \cdot \mathcal{B}(\overline{\Lambda}_b \to \mathsf{L} + X)$ with $f = 0 \ldots 1$, which leads to the final approximation

$$\mathcal{B}(\Lambda_b \to K^-\mu^+) \approx \eta\,\frac{f}{a_\mathsf{L}} \approx 6 \cdot 10^{-10}\,\frac{f}{a_\mathsf{L}} \;. \tag{A.6}$$

A.4.2 Additional creation of baryons and leptons

In addition to the decays $\overline{\Lambda}_b \to \mathsf{L} + X$ the large mass of the Λ_b also allows decays of the sort $\overline{\Lambda}_b \to \mathsf{L} + n\,(\mathsf{B} + \mathsf{L}) + \bar{n}\,(\overline{\mathsf{B}} + \overline{\mathsf{L}}) + X$, *e.g.* $\overline{\Lambda}_b \to \mu^- + 2\,(\Lambda\mu^-) + \overline{\Lambda}_c\mu^+$, where additional baryon-lepton pairs are created. An overall excess of matter $\langle a_m \rangle \sim \langle n \rangle - \langle \bar{n} \rangle$ would be created if the respective average multiplicities of pair production are not the same, *i.e.* $\langle n \rangle \neq \langle \bar{n} \rangle$. On the basis of the general assumption that matter and antimatter are always created in a symmetric process and any asymmetry results from decays, it is assumed $\langle n \rangle = \langle \bar{n} \rangle$ and hence no additional baryon and lepton production will occur.

A.5 Prior measurements related to BLNV

Table A.1: All upper limits are at the 90 % CL if not states otherwise.

Year	Experiment	Results	Reference
1998	OPAL	$\mathcal{B}(Z^0 \to pe^-) < 1.84 \cdot 10^{-6}$ $\mathcal{B}(Z^0 \to p\mu^-) < 1.76 \cdot 10^{-6}$	[70]
1999	CLEO	$\mathcal{B}(\tau^- \to \overline{p}\gamma) < 3.5 \cdot 10^{-6}$ $\mathcal{B}(\tau^- \to \overline{p}\pi^0) < 15 \cdot 10^{-6}$ $\mathcal{B}(\tau^- \to \overline{p}\eta) < 8.9 \cdot 10^{-6}$ $\mathcal{B}(\tau^- \to \overline{p}\,2\pi^0) < 33 \cdot 10^{-6}$ $\mathcal{B}(\tau^- \to \overline{p}\pi^0\eta) < 27 \cdot 10^{-6}$	[71]
2005	Belle	$\mathcal{B}(\tau^- \to \Lambda\pi^-) < 0.72 \cdot 10^{-7}$ $\mathcal{B}(\tau^- \to \overline{\Lambda}\pi^-) < 1.4 \cdot 10^{-7}$	[72]
2009	Super-K	$\tau_p/\mathcal{B}(p \to e^+\pi^0) > 8.2 \cdot 10^{33}\,\mathrm{yr}$ $\tau_p/\mathcal{B}(p \to \mu^+\pi^0) > 6.6 \cdot 10^{33}\,\mathrm{yr}$	[73]
2009	CLEO	$\mathcal{B}(D^0 \to pe^-) < 1.0 \cdot 10^{-5}$ $\mathcal{B}(D^0 \to \overline{p}e^+) < 1.1 \cdot 10^{-5}$	[74]
2011	BaBar	$\mathcal{B}(B^0 \to \Lambda_c e^-) < 5.2 \cdot 10^{-6}$ $\mathcal{B}(B^0 \to \Lambda_c\mu^-) < 1.8 \cdot 10^{-6}$ $\mathcal{B}(B^- \to \Lambda e^-) < 8.1 \cdot 10^{-8}$ $\mathcal{B}(B^- \to \Lambda\mu^-) < 6.2 \cdot 10^{-8}$ $\mathcal{B}(B^- \to \overline{\Lambda}e^-) < 3.2 \cdot 10^{-8}$ $\mathcal{B}(B^- \to \overline{\Lambda}\mu^-) < 6.1 \cdot 10^{-8}$	[75]
2013	LHCb	$\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 8.0 \cdot 10^{-8}$ $\mathcal{B}(\tau^- \to \overline{p}\mu^+\mu^-) < 3.3 \cdot 10^{-7}$ $\mathcal{B}(\tau^- \to p\mu^-\mu^-) < 4.4 \cdot 10^{-7}$	[76]
2014	CMS	$\mathcal{B}(t \to \ell^+\bar{q}\bar{q}') < 1.5 \cdot 10^{-3}$ at CL $= 95\,\%$ $\ell^+ = \{e^+, \mu^+\}, \bar{q} = \{\bar{d}, \bar{s}, \bar{b}\}, \bar{q}' = \{\bar{u}, \bar{c}\}$	[77]

A.6 Crystal Ball fit function

The Crystal Ball function (CB) [78] is a phenomenological fit model to describe the invariant mass spectra of charged track combinations including radiative energy losses of the tracks due to interactions with the detector material. The function is separated into two parts at the threshold $-|\alpha|$: a Gaussian distribution on one side and a power-law tail $\sim x^{-n}$ on the other side. For the fits the following definition is used

$$CB(x|\mu,\sigma,\alpha,n) = \frac{1}{N_0} \cdot \begin{cases} \exp\left[-\frac{1}{2}\,X^2\right] \ , & \text{if } X \geq -|\alpha| \\ \dfrac{\exp\left[-\frac{1}{2}\,\alpha^2\right] \left(\frac{n}{|\alpha|}\right)^n}{\left(\frac{n}{|\alpha|} - |\alpha| - X\right)^n} \ , & \text{if } X < -|\alpha| \end{cases} \tag{A.7}$$

where: $\{x,\mu,\sigma,\alpha,n\} \in \mathbb{R}, \quad \alpha \neq 0 \ , \ \sigma > 0 \ , \ n = 0 \ \texttt{OR} \ n > 1 \ ,$

$$X = \left(\frac{x-\mu}{\sigma}\right) \ , \ \text{ if } \ \alpha > 0 \rightarrow \text{radiative tail is below the threshold } -|\alpha|$$

$$X = -\left(\frac{x-\mu}{\sigma}\right) \ , \ \text{ if } \ \alpha < 0 \rightarrow \text{radiative tail is above the threshold } -|\alpha|$$

The finite integral including the transformation $dx = \sigma dX$ and $\vec{p} = (\mu,\sigma,\alpha,n)$ is

$$N(X_1,X_2|\vec{p}) = \sigma \int\limits_{X_1}^{X_2} CB(X|\vec{p})dX = \sigma \int\limits_{X_1}^{-|\alpha|} CB(X|\vec{p})dX + \sigma \int\limits_{-|\alpha|}^{X_2} CB(X|\vec{p})dX \ , \tag{A.8}$$

$$N(X_1,X_2|\vec{p}) = \sigma \left[\frac{\frac{n}{|\alpha|}\,\exp\left(-\frac{1}{2}\,\alpha^2\right)}{n-1} \left(1 - \frac{1}{\left[1 - \frac{|\alpha|^2}{n} - \frac{|\alpha|}{n}\,X_1\right]^{n-1}} \right) \right] + \sigma \left[\sqrt{\frac{\pi}{2}} \left(\text{Erf}\left[\frac{X_2}{\sqrt{2}}\right] + \text{Erf}\left[\frac{|\alpha|}{\sqrt{2}}\right] \right) \right] \ . \tag{A.9}$$

The normalisation factor N_0 is

$$N_0 = \lim_{-X_1,X_2\rightarrow\infty} N(X_1,X_2|\vec{p}) = \sigma \left[\frac{\frac{n}{|\alpha|}\,\exp\left(-\frac{1}{2}\,\alpha^2\right)}{n-1} + \sqrt{\frac{\pi}{2}} \left(1 + \text{Erf}\left[\frac{|\alpha|}{\sqrt{2}}\right] \right) \right] \ . \tag{A.10}$$

A.7 Combination of fit functions

If $F_i(x|\vec{p}_i)$ are a normalised fit functions with certain parameters $\vec{p}_i$, then the combination of fit functions is defined as

$$F(x|\vec{p},\vec{f}\,) = \sum_i f_i \cdot F_i(x|\vec{p}_i) \ , \ \sum_i f_i = 1 \ \rightarrow \ f_1 = 1 - \sum_{i\neq 1} f_i \ . \tag{A.11}$$

Hereby the parameter f_i denotes the relative fraction of the component F_i.

A.8 Kinematic considerations on $B_c^- \to J/\psi\,(\mu^+\mu^-)K^-$

The allowed mass range of $m(K^-\mu^+)$ depends only on the invariant masses of the mother and daughter particles and the helicity angle θ. Figure A.2 shows the general setup of a resonant two-body decay.

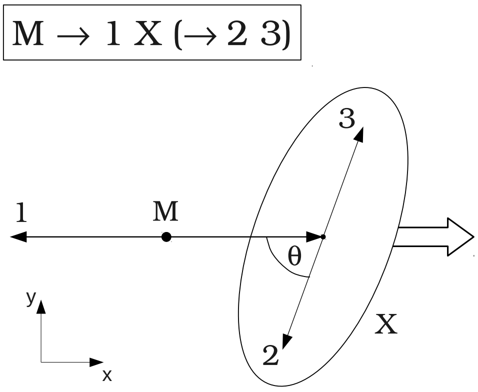


Figure A.2: Illustration of a resonant two-body decay $M \to 1\,X(\to 1\,2)$.

The general formula for the invariant mass of two particles is

$$m_{12}^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2 = m_1^2 + m_2^2 + 2(E_1E_2 - p_1p_2\cos\theta)\ .$$

The values m_1 and m_2 are known, $\theta = 0 \ldots \pi$ and (E_2, p_2) can be calculated in the cms of X

$$E_2 = \frac{m_X^2 - m_3^2 + m_2^2}{2m_X}\ , \quad p_2 = \frac{\sqrt{[m_X^2 - (m_2+m_3)^2]\,[m_X^2 - (m_2-m_3)^2]}}{2m_X}\ .$$

(E_1, p_1) can be calculated in the cms of X by doing a Lorentz boost of (E_1^*, p_1^*) from the cms of M. The boost vector is $\vec{\beta}_{\text{boost}} = \vec{\beta}_X$ in the cms of M. Therefore the boost is calculated from

$$\begin{pmatrix} E_1 \\ p_{1\,\parallel} \end{pmatrix} = \begin{pmatrix} \gamma_X & -\gamma_X\vec{\beta}_X \\ -\gamma_X\vec{\beta}_X & \gamma_X \end{pmatrix} \cdot \begin{pmatrix} E_1^* \\ p^*_{1\,\parallel} \end{pmatrix}\ , \quad p_{1\,\perp} = p^*_{1\,\perp} = 0\ .$$

The latter relation is due to the fact that $\vec{\beta}_X$ and $\vec{p}_1^{\,*}$ are antiparallel and hence $\vec{p}_1^{\,*}$ has no component orthogonal to $\vec{\beta}_X$. From that it follows that $\vec{p}_1^{\,*} = \vec{p}^{\,*}_{1\,\parallel}$ as well as $\vec{p}_1^{\,*} = -\vec{p}_X^{\,*}$ and $p_1^* = p_X^* = p^*$ in the cms of M.

The boost parameters γ_X and $\gamma_X\vec{\beta}_X$ are

$$\gamma_X = \frac{E_X^*}{m_X}\ , \quad \gamma_X\vec{\beta}_X = \frac{\vec{p}_X^{\,*}}{m_X}\ .$$

Solving the the first row of the matrix equation yields

$$E_1 = \gamma_X E_1^* - \gamma_X\vec{\beta}_X\vec{p}^{\,*}_{1\,\parallel} = \frac{E_X^*E_1^*}{m_X} - \frac{\vec{p}_X^{\,*}}{m_X}\cdot(-\vec{p}_X^{\,*}) = \frac{E_X^*E_1^* + p^{*2}}{m_X}\ .$$

Using the relation $E_X^* = M - E_1^*$ and $E_1^* = \dfrac{M^2 - m_X^2 + m_1^2}{2M}$ it is

$$E_1 = \frac{M E_1^* - m_1^2}{m_X} = \frac{M \cdot \frac{M^2 - m_X^2 + m_1^2}{2M} - m_1^2}{m_X} \ ,$$

$$\boxed{E_1 = \frac{M^2 - m_X^2 - m_1^2}{2\,m_X}} \ .$$

The momentum p_1 can be obtained from the relation $E^2 = p^2 + m^2$

$$\boxed{p_1 = \sqrt{E_1^2 - m_1^2}} \ .$$

Figure A.3 shows the invariant mass spectrum of $m(K^+\mu^-)$ in dependence of θ. For the partially reconstructed decay $B_c^+ \to J/\psi\,(\mu^+\mu^-)K^+$ the $m(K^-\mu^+)$ spectrum ends at $5450\,\text{MeV}/c^2$.

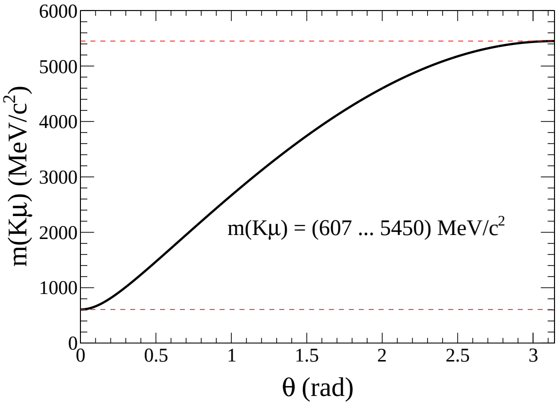


Figure A.3: Spectrum of $m(K^+\mu^-)$ in dependence of θ for the decay $B_c^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$.

A.9 Estimation of the B_c^+ production at LHCb

There are three published results that allow estimate the B_c^+ production at LHCb. The first one is the measurement of the ratio of production cross section times branching fraction between the decays $B_c^- \to J/\psi\,\pi^-$ and $B^+ \to J/\psi\,K^+$ [79]. The result is

$$R = \frac{f_c}{f_u} \times \frac{\mathcal{B}(B_c^- \to J/\psi\,\pi^-)}{\mathcal{B}(B^+ \to J/\psi\,K^+)} = (0.683 \pm 0.018 \pm 0.009)\,\% \ .$$

The second one is $\mathcal{B}(B^+ \to J/\psi\,K^+) = (1.026 \pm 0.031) \cdot 10^{-3}$ [3]. Finally the branching ratio $\mathcal{B}(B_c^+ \to J/\psi\,\pi^-) = (2.91^{+0.15}_{-0.42}{}^{+0.40}_{-0.27}) \cdot 10^{-3}$ was calculated by theorists including next-to-leading order corrections [80]. As a result the estimated ratio of the production cross sections is

$$\frac{f_c}{f_u} = R \cdot \frac{\mathcal{B}(B^+ \to J/\psi\,K^+)}{\mathcal{B}(B_c^- \to J/\psi\,\pi^-)} \approx \frac{1}{415} = 0.24\,\% \ .$$

A.10 Properties of relevant particles

Table A.2: Properties of the used particles according to the RPP [3]. Values without uncertainty are rounded before the first significant figure.

Particle	Mass (MeV/c^2)	Lifetime $\tau = \hbar/\Gamma$	$c \cdot \tau$
$\mu^\pm$	105.7	2.197 µs	659 m
$\pi^\pm$	139.6	26.03 ns	7.80 m
$K^\pm$	493.7	(12.380 ± 0.021) ns	3.71 m
$B^\pm$	5279.31 ± 0.15	(1.638 ± 0.004) ps	0.491 mm
B^0	5279.62 ± 0.15	(1.520 ± 0.004) ps	0.456 mm
B_s^0	5366.82 ± 0.22	(1.510 ± 0.005) ps	0.453 mm
B_c^+	6275.1 ± 1.0	(0.507 ± 0.009) ps	0.152 mm
J/ψ	3096.9	$\approx \hbar/(93\,\mathrm{keV})$	$\mathcal{O}(1\,\mathrm{pm})$
p	938.3	$\gtrsim 10^{29}$ yr	—
Λ_c	2286.46 ± 0.14	(0.200 ± 0.006) ps	0.060 mm
Λ_b	5619.51 ± 0.23	(1.466 ± 0.010) ps	0.439 mm
Ξ_b^0	5791.9 ± 0.5	(1.464 ± 0.031) ps	0.439 mm

Additional information concerning Λ_b and Ξ_b^0 according to RPP [3]

- $m(\Lambda_b) - m(B^0) = (339.2 \pm 1.4)\,\mathrm{MeV}/c^2$
- $m(\Lambda_b) - m(B^+) = (339.72 \pm 0.28)\,\mathrm{MeV}/c^2$
- $m(\Xi_b^0) - m(\Lambda_b) = (172.5 \pm 0.4)\,\mathrm{MeV}/c^2$
- $\dfrac{\tau(\Lambda_b)}{\tau(B^0)} = 0.964 \pm 0.007$

A.10.1 Published LHCb measurements related to $\mathit{\Lambda}_b$ or $\mathit{\Xi}_b^0$

The following table summarises published LHCb measurements related to properties of $\mathit{\Lambda}_b$ and/or $\mathit{\Xi}_b^0$ and the used amount of integrated luminosity.

Table A.3: LHCb measurements concerning $\mathit{\Lambda}_b$ and $\mathit{\Xi}_b^0$

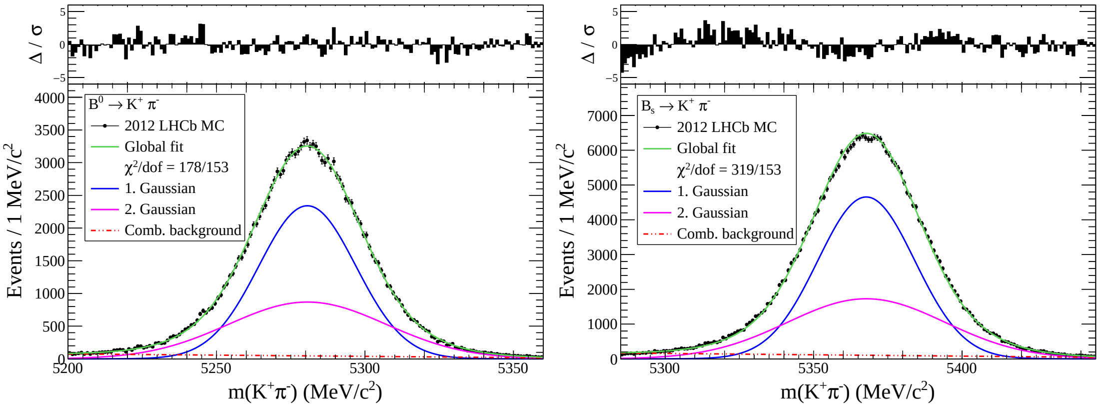
Phys.Rev.Lett. 113 (2014) 032001 (July 2014) [81]	$3\,\mathrm{fb}^{-1}$
$m(\mathit{\Xi}_b^0) - m(\mathit{\Lambda}_b) = (172.44 \pm 0.43)\,\mathrm{MeV}/c^2$ $\frac{\tau(\mathit{\Xi}_b^0)}{\tau(\mathit{\Lambda}_b)} = 1.006 \pm 0.021$ $\frac{f_{\mathit{\Xi}_b^0}}{f_{\mathit{\Lambda}_b}} \cdot \frac{\mathcal{B}(\mathit{\Xi}_b^0 \to \mathit{\Xi}_c^+ \pi^-)}{\mathcal{B}(\mathit{\Lambda}_b \to \mathit{\Lambda}_c \pi^-)} \cdot \frac{\mathcal{B}(\mathit{\Xi}_c^+ \to pK^-\pi^+)}{\mathcal{B}(\mathit{\Lambda}_c \to pK^-\pi^+)} = (1.88 \pm 0.05) \times 10^{-2}$	
Phys.Lett. B734 (2014) 122-130 (June 2014) [82]	$3\,\mathrm{fb}^{-1}$
$\frac{\tau(\mathit{\Lambda}_b)}{\tau(\overline{B}^0)} = 0.974 \pm 0.007$	
JHEP 1404 (2014) 114 (April 2014) [83]	$1\,\mathrm{fb}^{-1}$
$\tau_{\mathit{\Lambda}_b \to J/\psi\, \mathit{\Lambda}} = (1.415 \pm 0.028)\,\mathrm{ps}$	
Phys.Rev. D89 (2014) 032001 (Feb. 2014) [84]	$1\,\mathrm{fb}^{-1}$
$m(\mathit{\Xi}_b^0) - m(\mathit{\Lambda}_b) = (174.8 \pm 2.5)\,\mathrm{MeV}/c^2$	
Phys.Rev.Lett. 111 (2013) 102003 (Sep. 2013) [85]	$1\,\mathrm{fb}^{-1}$
$\frac{\tau(\mathit{\Lambda}_b)}{\tau(\overline{B}^0)} = 0.976 \pm 0.014$	
Phys.Rev.Lett. 110 (2013) 182001 (May 2013) [86]	$1\,\mathrm{fb}^{-1}$
$m(\mathit{\Lambda}_b) = (5619.44 \pm 0.40)\,\mathrm{MeV}/c^2$	
JHEP 1210 (2012) 037 (Oct. 2012) [87]	$0.37\,\mathrm{fb}^{-1}$
$\frac{\mathcal{B}(\mathit{\Lambda}_b \to p\pi^-)}{\mathcal{B}(\mathit{\Lambda}_b \to pK^-)} = 0.86 \pm 0.10$	

A.11 Fit model for the $m(K^+\pi^-)$ distribution

The fit model for the $B^0 \to K^+\pi^-$ distribution is developed using simulated events of $B^0 \to K^+\pi^-$, $B_s^0 \to K^+\pi^-$, $B^0 \to \pi^+\pi^-$, $B_s^0 \to K^+K^-$, $\Lambda_b \to p\pi^-$, and $\Lambda_b \to pK^-$. Each decay is reconstructed using the $B^0 \to K^+\pi^-$ event selection and the signal triggers are applied.

A.11.1 Modelling of $B_{(s)}^0 \to K^+\pi^-$ decays

These two decays are reconstructed using the correct mass hypothesis for the daughter particles. The two-fold misidentification $K \to \pi$ and $\pi \to K$ can be neglected due the fact that the event selection requires $\mathrm{DLL(K-\pi)(K)} > \mathrm{DLL(K-\pi)(\pi)}$. In simulated events this condition suppresses double misidentification cases. The invariant mass spectrum of the $B^0 \to K^+\pi^-$ mass spectrum is symmetric and described by the sum of two Gaussian distributions with a common mean for the signal. A first-order polynomial is applied for the combinatorial background. For the fit of the $B_s^0 \to K^+\pi^-$ mass spectrum the obtained parameters for the two widths and the relative ratio of the two Gaussians are fixed. Figure A.4 shows the results of the two fits and Tab. A.4 summarises the fitted parameters of the signal components. The shape of the two invariant mass spectra are compatible.



(a) $B^0 \to K^+\pi^-$ (b) $B_s^0 \to K^+\pi^-$

Figure A.4: Fits to the $m(K^+\pi^-)$ distributions from simulated $B_{(s)}^0 \to K^+\pi^-$ events.

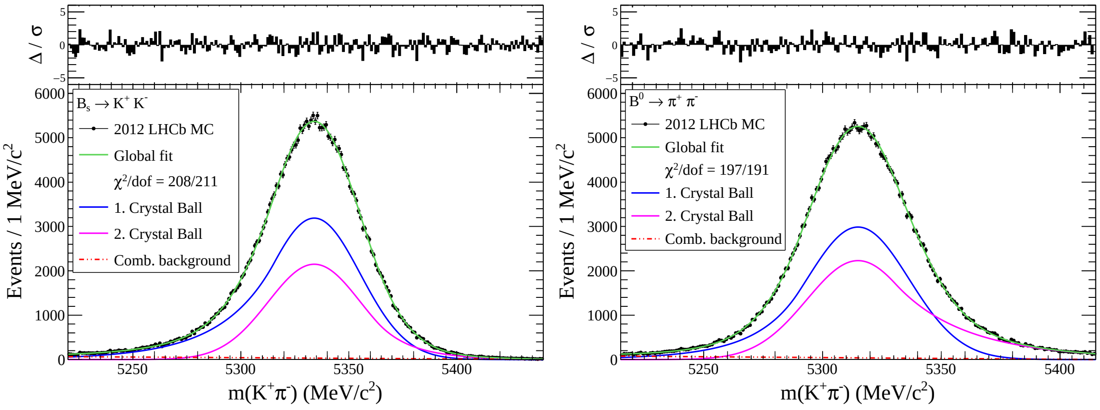
Table A.4: Values of the fitted parameters of the signal components.

Parameter	$B^0 \to K^+\pi^-$	$B_s^0 \to K^+\pi^-$
μ	$(5280.59 \pm 0.06)\,\mathrm{MeV}/c^2$	$(5367.75 \pm 0.04)\,\mathrm{MeV}/c^2$
f_2	$(38 \pm 6)\,\%$	fixed
σ_1	$(16.4 \pm 0.4)\,\mathrm{MeV}/c^2$	fixed
σ_2	$(26.6 \pm 1.4)\,\mathrm{MeV}/c^2$	fixed

A.11.2 Modelling of the decays $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$

The reflections $B_s^0 \to K^+K^-$ and $B^0 \to \pi^+\pi^-$ have an asymmetric mass shape due to the wrong mass hypothesis for one of the daughters. Both mass spectra are modeled by the sum of two

Crystal Ball profiles with a common mean value for signal component and a first-order polynomial for the combinatorial background. Figure A.5 shows the results of the fits and Tab. A.5 the fitted parameters.



(a) $B_s^0 \to K^+K^-$ (b) $B^0 \to \pi^+\pi^-$

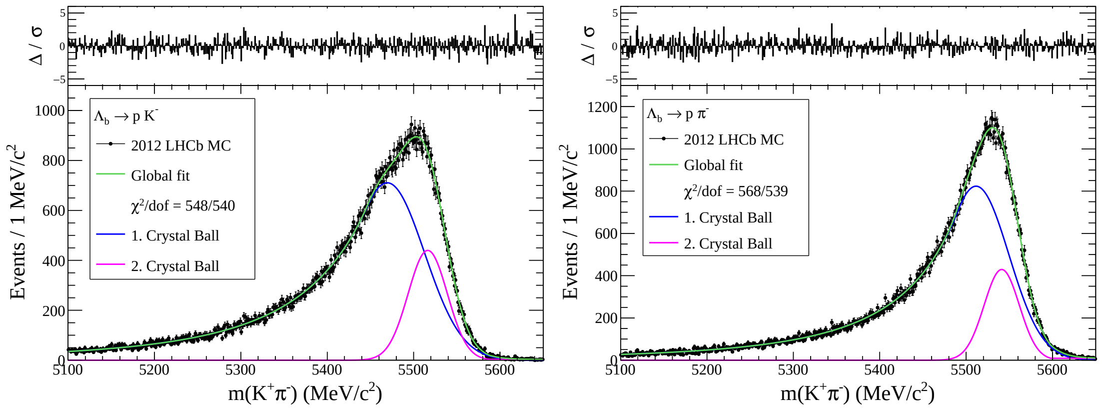
Figure A.5: Fits to the $m(K^+\pi^-)$ distributions from simulated $B_s^0 \to K^+K^-$ and $B^0 \to \pi^+\pi^-$ events.

Table A.5: Values of the fitted parameters of the signal components.

Parameter	$B_s^0 \to K^+K^-$	$B^0 \to \pi^+\pi^-$
μ	$(5334.00 \pm 0.06)\,\mathrm{MeV}/c^2$	$(5314.95 \pm 0.07)\,\mathrm{MeV}/c^2$
f_2	$(39.0 \pm 2.7)\,\%$	$(45.4 \pm 2.1)\,\%$
σ_1	$(21.34 \pm 0.07)\,\mathrm{MeV}/c^2$	$(21.53 \pm 0.11)\,\mathrm{MeV}/c^2$
σ_2	$\equiv \sigma_1$	$\equiv \sigma_1$
α_1	0.878 ± 0.026	1.16 ± 0.04
α_2	-1.40 ± 0.04	-0.733 ± 0.033
n_1	16.4 ± 2.6	12.5 ± 3.6
n_2	$\equiv n_1$	$\equiv n_1$

A.11.3 Modelling of the decays $\Lambda_b \to p\pi^-$ and $\Lambda_b \to pK^-$

The invariant mass distribution has a long tail to the left due to the large amount of missing energy from the $p \to \pi$ or $p \to K$ misidentification. The spectrum is described by the sum of two Crystal Ball functions without common parameters. There are no relevant contributions from random combinations in simulated events. Figure A.6 shows the results of the fits and Tab. A.6 the fitted parameters.



(a) $\Lambda_b \to pK^-$ (b) $\Lambda_b \to p\pi^-$

Figure A.6: Fits to the $m(K^+\pi^-)$ distributions from simulated $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ events.

Table A.6: Values of the fitted parameters of the signal components.

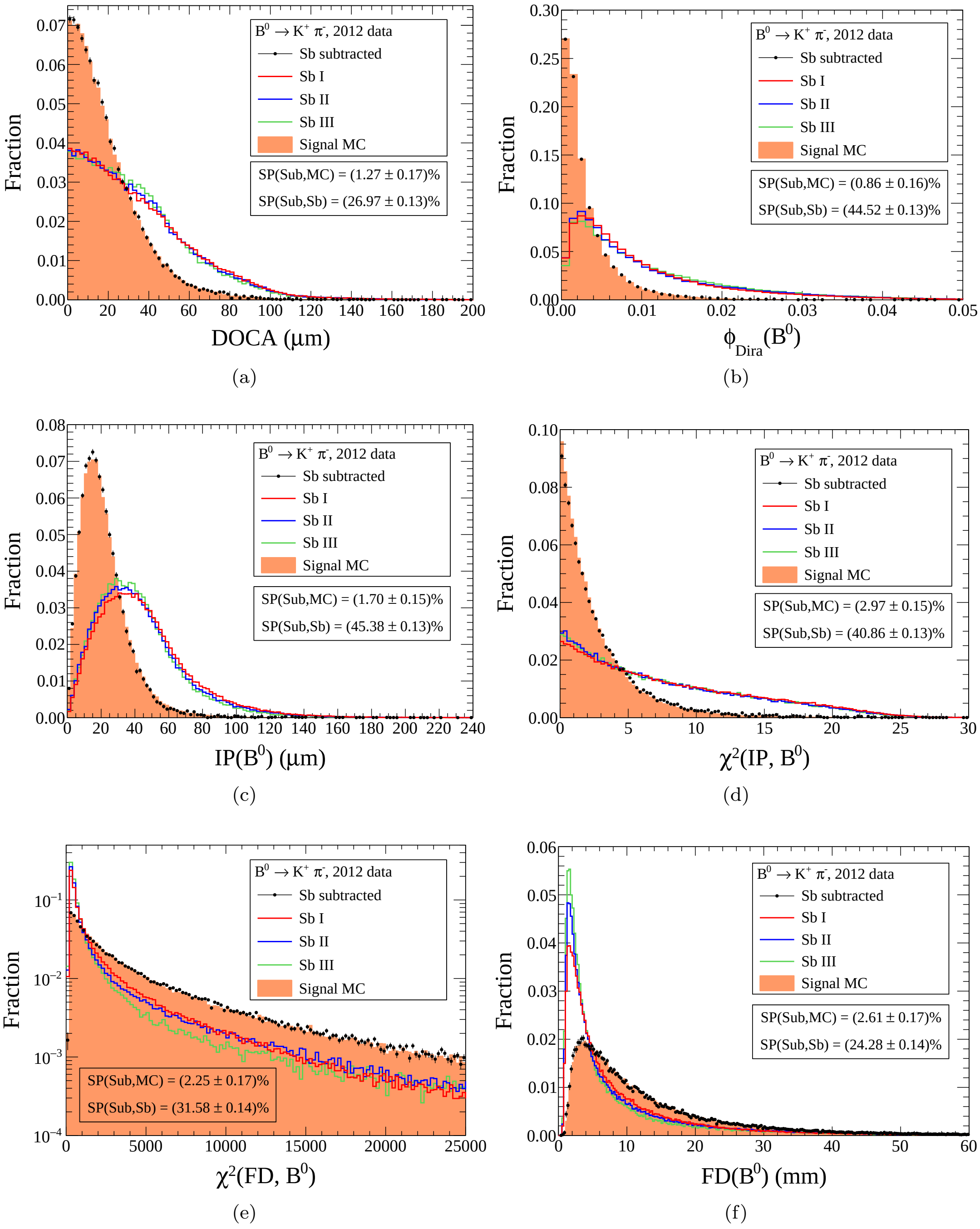
Parameter	$\Lambda_b \to pK^-$	$\Lambda_b \to p\pi^-$
f_2	$(18.19 \pm 0.8)\,\%$	$(15.8 \pm 1.0)\,\%$
μ_1	$(5469.5 \pm 1.2)\,\mathrm{MeV}/c^2$	$(5511.5 \pm 1.0)\,\mathrm{MeV}/c^2$
μ_2	$(5516.4 \pm 0.6)\,\mathrm{MeV}/c^2$	$(5541.2 \pm 0.6)\,\mathrm{MeV}/c^2$
σ_1	$(43.1 \pm 0.6)\,\mathrm{MeV}/c^2$	$(38.1 \pm 0.6)\,\mathrm{MeV}/c^2$
σ_2	$(22.7 \pm 0.3)\,\mathrm{MeV}/c^2$	$(20.4 \pm 0.4)\,\mathrm{MeV}/c^2$
α_1	0.536 ± 0.012	0.553 ± 0.010
α_2	-3.14 ± 0.04	-2.78 ± 0.04
n_1	3.76 ± 0.20	3.17 ± 0.11
n_2	$= 0$ (fixed)	$= 0$ (fixed)

A.11.4 Fit strategy for data

For the fit to the data all parameters are fixed to the values obtained from MC. To determine the mass shift in real data all mean masses are commonly shifted by a free fit parameter.

A.12 Comparison of MC and data in 2012 $B^0 \to K^+\pi^-$ events

The following plots compare signal and background events from reconstructed $K^+\pi^-$ events from 2012 data and simulated $B^0 \to K^+\pi^-$ events. All events require the signal triggers. More information about the used samples can be found in Sec. 4.1.3.

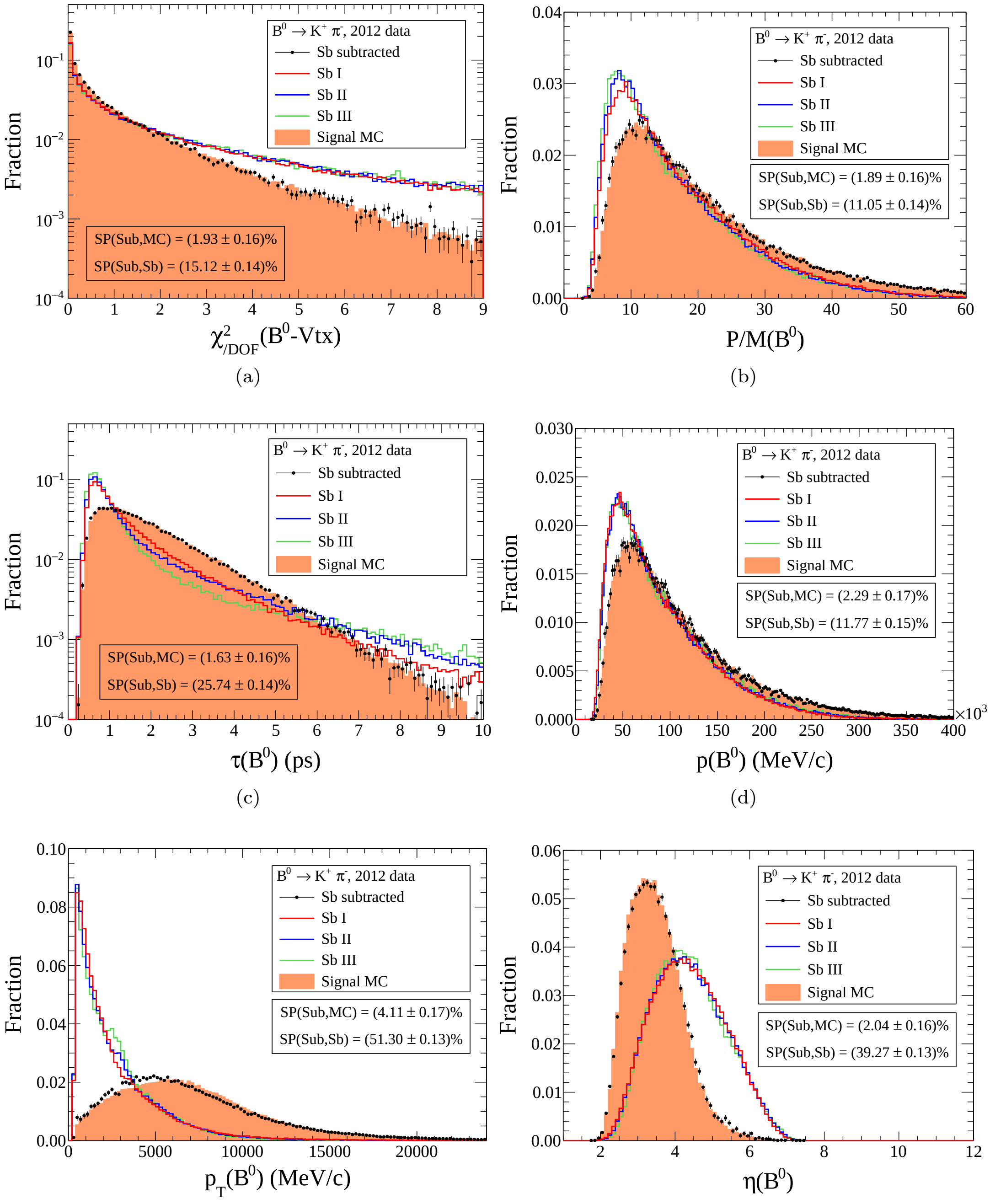


(a) (b)

(c) (d)

(e) (f)

Figure A.7: Comparison of signal variables in different data samples - Part 1.

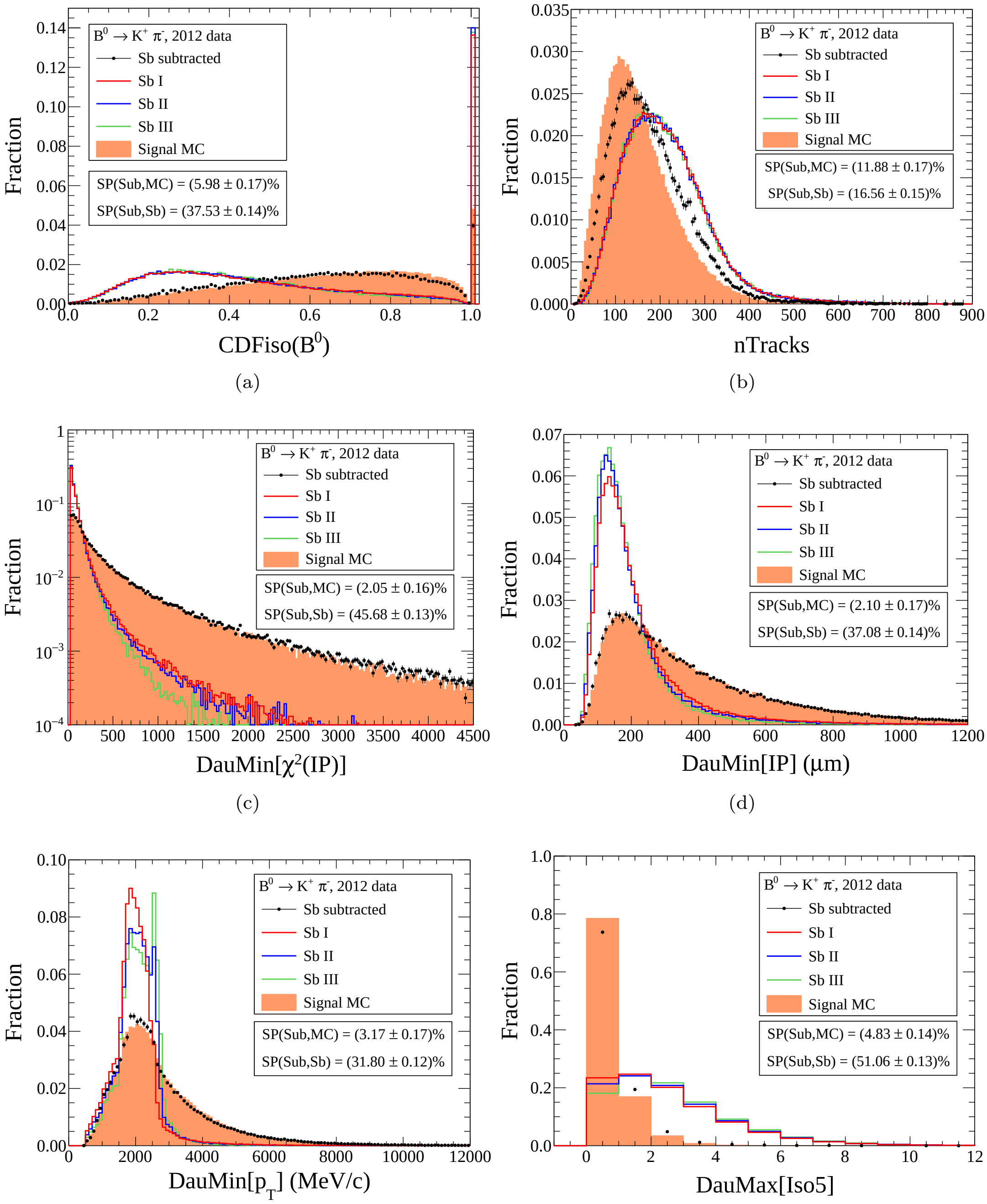


(a) (b)

(c) (d)

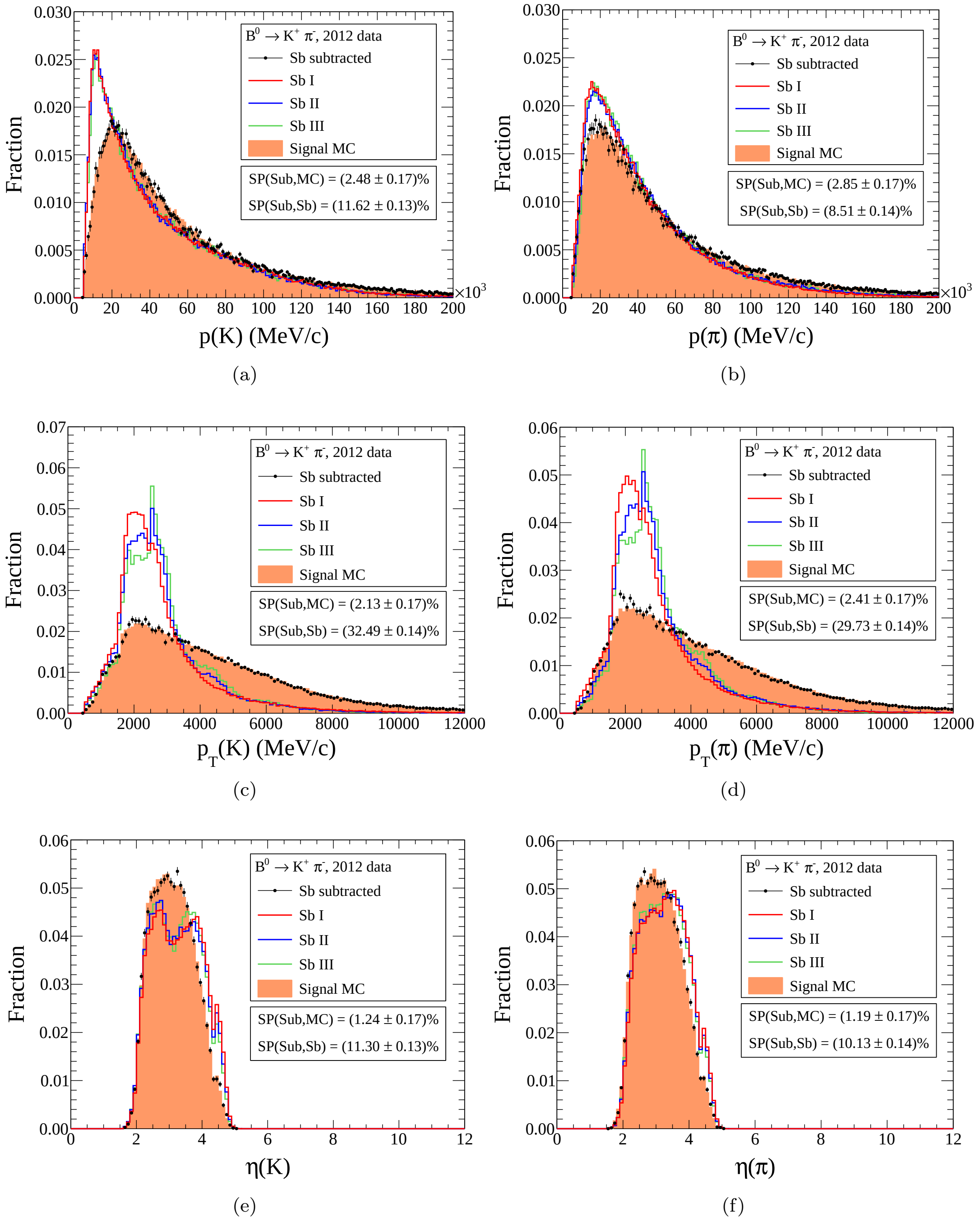
(e) (f)

Figure A.8: Comparison of signal variables in different data samples - Part 2.



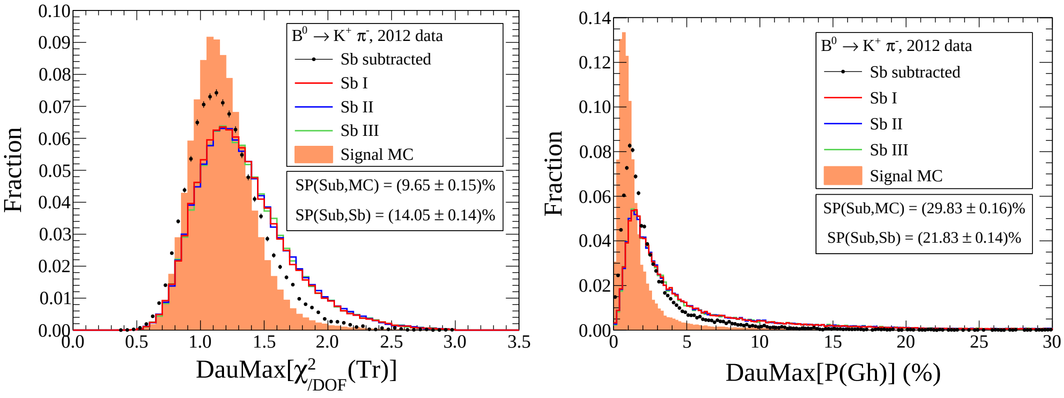
(a) (b) (c) (d) (e) (f)

Figure A.9: Comparison of signal variables in different data samples - Part 3.



(a) (b) (c) (d) (e) (f)

Figure A.10: Comparison of signal variables in different data samples - Part 4.



(a) (b)

Figure A.11: Comparison of signal variables in different data samples - Part 5.

A.13 $B^+ \to J/\psi\, K^+$ event selection

Table A.7: $B^+ \to J/\psi\,(\mu^+\mu^-)K^+$ event selections.

Property \ Decay	$J/\psi \to \mu^+\mu^-$	$B^+ \to J/\psi\, K^+$
PID selections		
Track 1	IsMuon=true	—
Track 2	IsMuon=true	—
Track selections		
p_{T}	$> 250\,\mathrm{MeV}/c^2$	$> 250\,\mathrm{MeV}/c^2$
$\chi^2(\mathrm{IP})$	> 25	> 25
$\chi^2_{/\mathrm{DOF}}(\mathrm{Tr})$	< 3	< 3
Selections on combinations		
Invariant mass	$(3097 \pm 100)\,\mathrm{MeV}/c^2$	$(5280 \pm 500)\,\mathrm{MeV}/c^2$
DOCA	$< 0.3\,\mathrm{mm}$	—
Selections after vertex fit		
$\chi^2(\mathrm{Vtx})$	< 9	< 45
ϕ_{Dira}	$< 90°$	—
$\chi^2(\mathrm{IP})$	—	< 25
$\chi^2(\mathrm{FD})$	> 169	—

A.14 $\Lambda_b \to \Lambda_c\pi^-$ event selection

Table A.8: $\Lambda_b \to \Lambda_c(pK^-\pi^+)\pi^-$ event selections.

Property \ Decay	$\Lambda_c \to pK^-\pi^+$	$\Lambda_b \to \Lambda_c\pi^-$
Track selections		
(p, p_{T})	$> (1, 0.1)$ GeV/c	$> (1, 0.1)$ GeV/c
P(Gh)	< 0.3	< 0.3
χ^2(IP)	> 4	> 4
$\chi^2_{/\mathrm{DOF}}$(Tr)	< 3	< 3
Selections on combinations		
Invariant Mass	(2286 ± 110) MeV/c^2	$(5200 \ldots 7000)$ MeV/c^2
$\chi^2_{/\mathrm{DOF}}$(Tr)	< 2.5	< 2.5
$\sum \lvert p_{\mathrm{T\,Track}} \rvert$	> 1800 MeV/c	> 5000 MeV/c
(p, p_{T})	$> (5, 0.5)$ GeV/c	—
DOCA	< 0.5 mm	—
Selections after vertex fit		
χ^2(IP)(Tracks)	—	> 16
IP(Tracks)	—	> 0.1 mm
$\chi^2_{/\mathrm{DOF}}$(Vtx)	< 10	< 10
χ^2(FD)	> 36	< 5
$\cos(\phi_{\mathrm{Dira}})$	> 0	> 0.999
(p, p_{T})	—	$> (10, 1.7)$ GeV/c
τ	—	> 0.2 ps
χ^2(IP)	—	< 25

A.14.1 $\varLambda_c$ selection in $\varLambda_b \to \varLambda_c\pi^-$ events

To reject the large amount of random $pK^-\pi^+$ combinations, a $\varLambda_c$ selection in $m(pK^-\pi^+)$ is developed. Figure A.12 shows the invariant mass spectrum of $pK^-\pi^+$ combinations from reconstructed $\varLambda_b \to \varLambda_c(pK^-\pi^+)\pi^-$ decays. All events require the signal triggers and the soft selections $m(\varLambda_c\pi) = 5550\ldots5700\,\text{MeV}/c^2$, ProbNN$(p) > 1\,\%$ and ProbNN$(K) > 1\,\%$. The distribution is fitted using the sum of two Gaussians with a common mean value to model the signal component and a linear function to describe the combinatorial background. All parameters are allowed to float during the fit. About 66 % of the signal events are found in the range of $\pm 6\,\text{MeV}/c^2$ around the fitted $\varLambda_c$ mass of $2287.5\,\text{MeV}/c^2$. This corresponds to an effective resolution of $\sigma_{\text{eff}}(\varLambda_c) = 6\,\text{MeV}/c^2$. For the $\varLambda_c$ selection it is required $m(pK^-\pi^+) = 2265\ldots2310\,\text{MeV}/c^2$ which is satisfied by more than 99 % of the $\varLambda_c$ candidates according to the fit. The fitted parameters of the two Gaussians are: $\mu = (2287.46 \pm 0.02)\,\text{MeV}/c^2$, $f_2 = (36.8 \pm 2.5)\,\%$, $\sigma_1 = (4.91 \pm 0.07)\,\text{MeV}/c^2$, and $\sigma_2 = (9.36 \pm 0.26)\,\text{MeV}/c^2$.

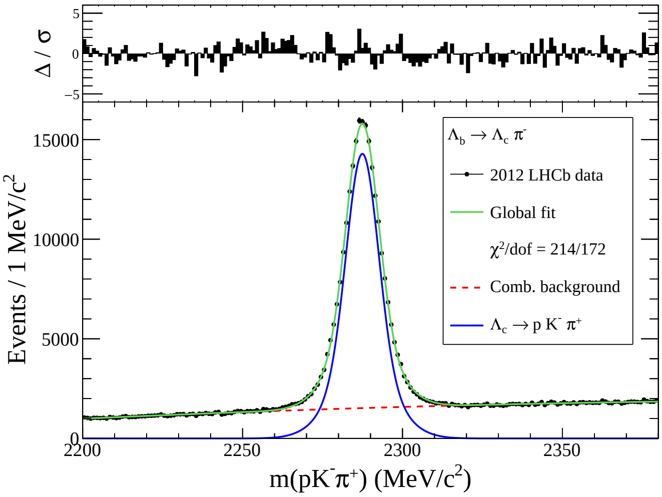


Figure A.12: Fit to the $m(pK^-\pi^+)$ distribution of selected events from 2012 data.

A.14.2 Fit model for the $m(\varLambda_c\pi^-)$ distribution

For the fit of the $m(\varLambda_c\pi^-)$ distribution three signal components are considered; $\varLambda_b \to \varLambda_c\pi^-$, $\varLambda_b \to \varLambda_c K^-$, and $B^0 \to D^+\pi^-$. All modes are reconstructed in simulated events using the $\varLambda_b \to \varLambda_c\pi^-$ event selection and requiring the signal triggers. All fits are binned likelihood fits to the shown distributions.

Modelling of $\varLambda_b \to \varLambda_c\pi^-$

This mode is modeled by the sum of two Gaussian distributions with a common mean value for the signal and a constant for the combinatorial background. Figure A.13 shows the result of the fit to the $m(\varLambda_c\pi^-)$ spectrum. The fitted parameters of the two Gaussians are: $\mu = (5620.81 \pm 0.07)\,\text{MeV}/c^2$, $f_2 = (28.7 \pm 1.1)\,\%$, $\sigma_1 = (14.21 \pm 0.14)\,\text{MeV}/c^2$, and $\sigma_2 = (32.91 \pm 0.7)\,\text{MeV}/c^2$.

Modelling of $\varLambda_b \to \varLambda_c K^-$

The asymmetric mass distribution of $\varLambda_b \to \varLambda_c K^-$ events is described by a single Crystal Ball profile for the signal and a constant for the combinatorial background. Figure A.14(a) shows the result of the fit to the $m(\varLambda_c\pi^-)$ spectrum. The fitted parameters of the Crystal Ball function are: $\mu = (5576.9 \pm 0.5)\,\text{MeV}/c^2$, $\sigma = (20.8 \pm 0.4)\,\text{MeV}/c^2$, $\alpha = 0.509 \pm 0.013$, and $n = 100 \pm 6$.

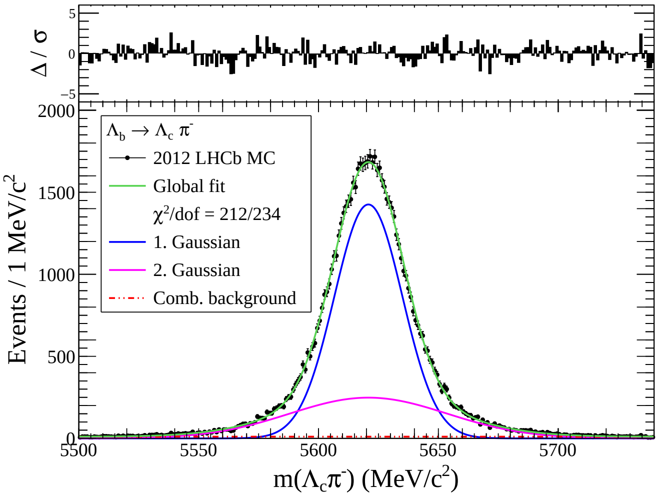
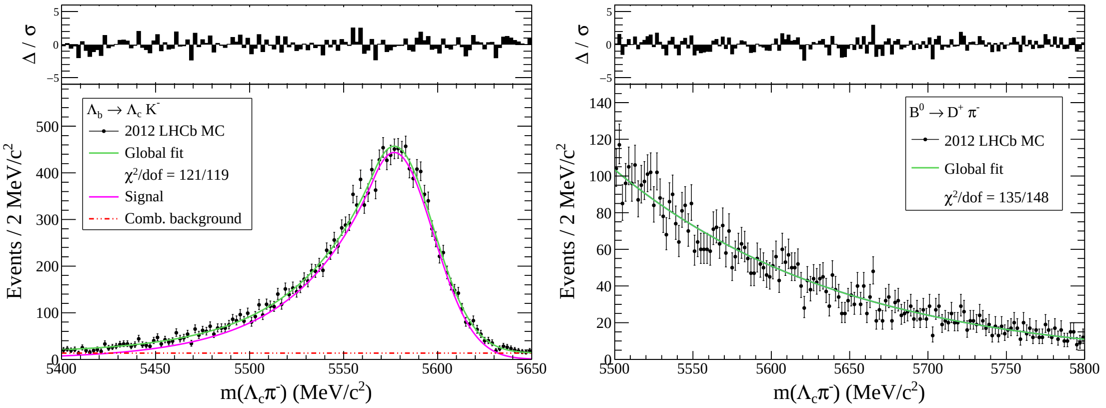


Figure A.13: Fit to the $m(\Lambda_c\pi^-)$ spectrum of simulated $\Lambda_b \to \Lambda_c\pi^-$ events.

Modelling of $B^0 \to D^+\pi^-$

The smoothly falling mass spectrum is described by a Gaussian in the range $m(\Lambda_c\pi^-) = 5500\ldots5800\,\mathrm{MeV}/c^2$. Due to the large correlation between the two parameters the width is fixed to the arbitrarily chosen value $500\,\mathrm{MeV}/c^2$ and the mean is fitted. Figure A.14(b) shows the result of the fit. The fitted value is $\mu = (3779 \pm 41)\,\mathrm{MeV}/c^2$.



(a) $\Lambda_b \to \Lambda_c K^-$

(b) $B^0 \to D^+\pi^-$

Figure A.14: Fits to the $m(\Lambda_c\pi^-)$ spectrum of simulated $\Lambda_b \to \Lambda_c K^-$ and $B^0 \to D^+\pi^-$ events.

Fit to real data

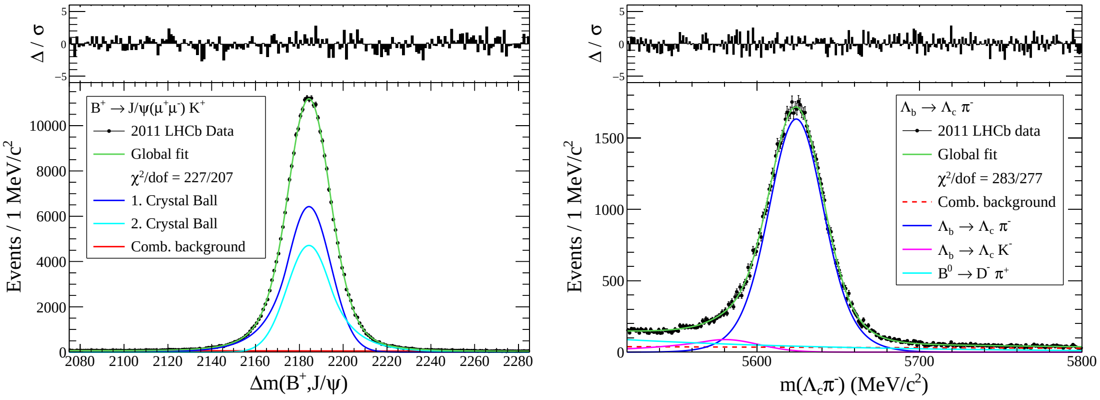
In the fit to real data all parameters of the modes $\Lambda_b \to \Lambda_c K^-$ and $B^0 \to D^+\pi^-$ are fixed as well as the obtained mean value of $\Lambda_b \to \Lambda_c\pi^-$. The remaining parameters of the $\Lambda_b \to \Lambda_c\pi^-$ fit model are kept free due to the large number of signal events in data. The mass shift is considered by a free fit parameter that is added equally to the mean of all signal components.

A.15 $\Lambda_b \to J/\psi\, pK^-$ event selection

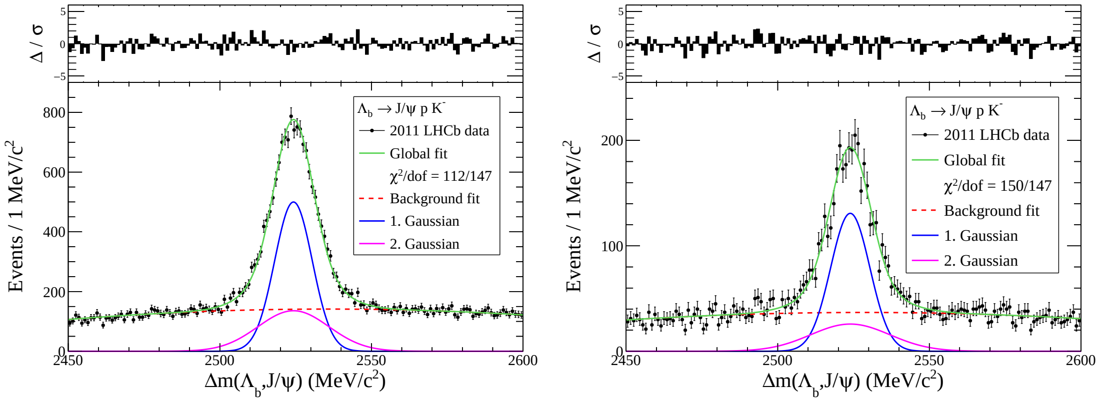
Table A.9: $\Lambda_b \to J/\psi\,(\mu^+\mu^-)pK^-$ event selections.

Property \ Decay	$J/\psi \to \mu^+\mu^-$	$\Lambda_b \to J/\psi\, pK^-$
PID selections		
Track 1	IsMuon=true	$\mathrm{DLL}(\mathrm{p}-\pi)(\mathrm{p}) > -5$
Track 2	IsMuon=true	—
Track 1&2	$\mathrm{DLL}(\mu-\pi)(\mu) > 0$	—
Track selections		
p_T	$> 500\,\mathrm{MeV}/c^2$	$> 500\,\mathrm{MeV}/c^2$
$\chi^2(\mathrm{IP})$	–	> 9
$\chi^2_{/\mathrm{DOF}}(\mathrm{Tr})$	< 3	< 3
Selections on combinations		
Invariant mass	$(3097 \pm 80)\,\mathrm{MeV}/c^2$	$(4800\ldots6200)\,\mathrm{MeV}/c^2$
$\chi^2(\mathrm{DOCA})$	< 20	—
Selections after vertex fit		
Invariant mass	—	$(4900\ldots6100)\,\mathrm{MeV}/c^2$
$\chi^2(\mathrm{Vtx})$	< 16	—
$\chi^2_{/\mathrm{DOF}}(\mathrm{Vtx})$	—	< 5

A.16 Fits to 2011 mass spectra of control channels



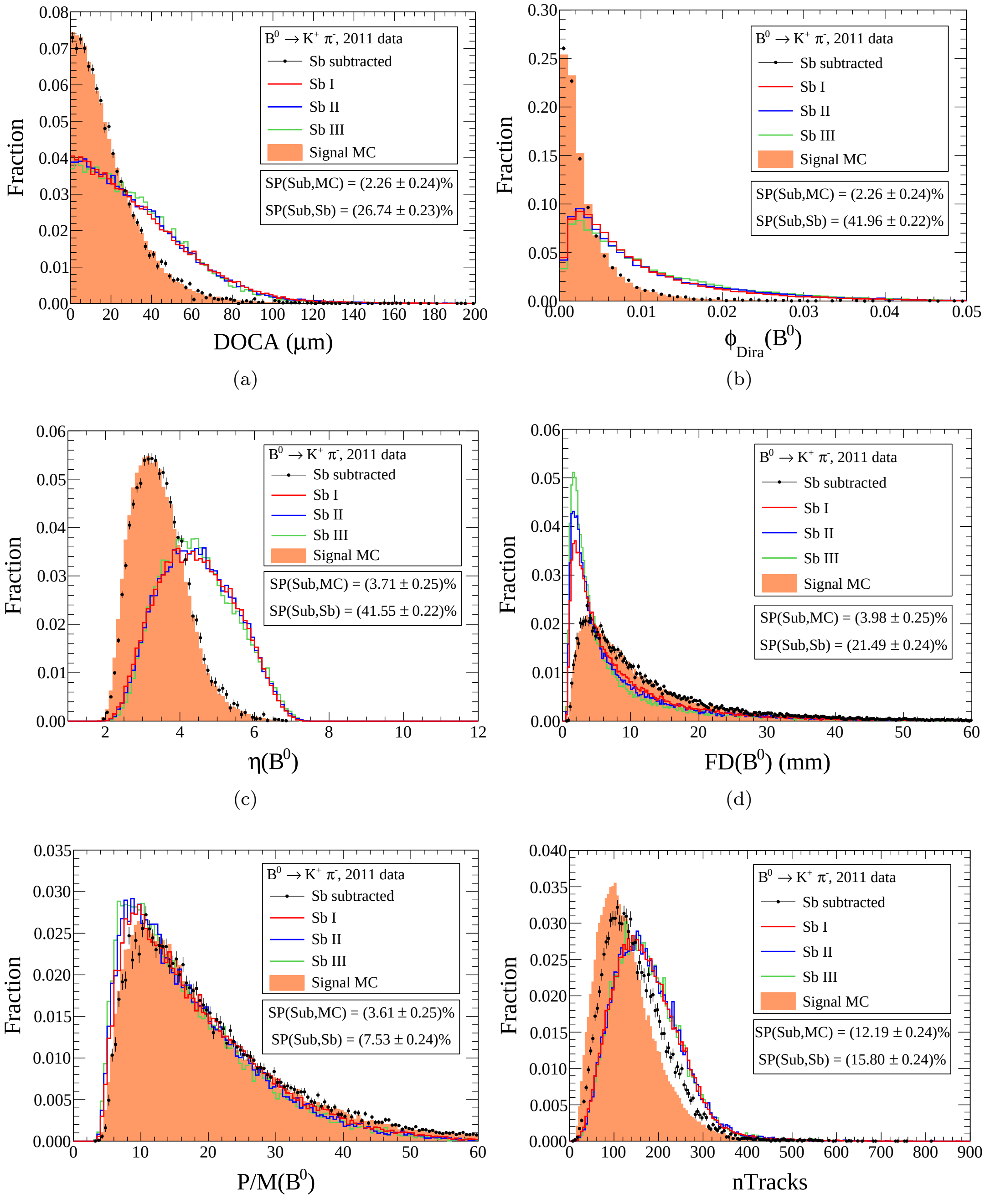
(a) $B^+ \to J/\psi K^+$ (b) $\Lambda_b \to \Lambda_c \pi^-$



(c) $\Lambda_b \to J/\psi pK^-$ w/ muon triggers (d) $\Lambda_b \to J/\psi pK^-$ w/o muon-TOS triggers

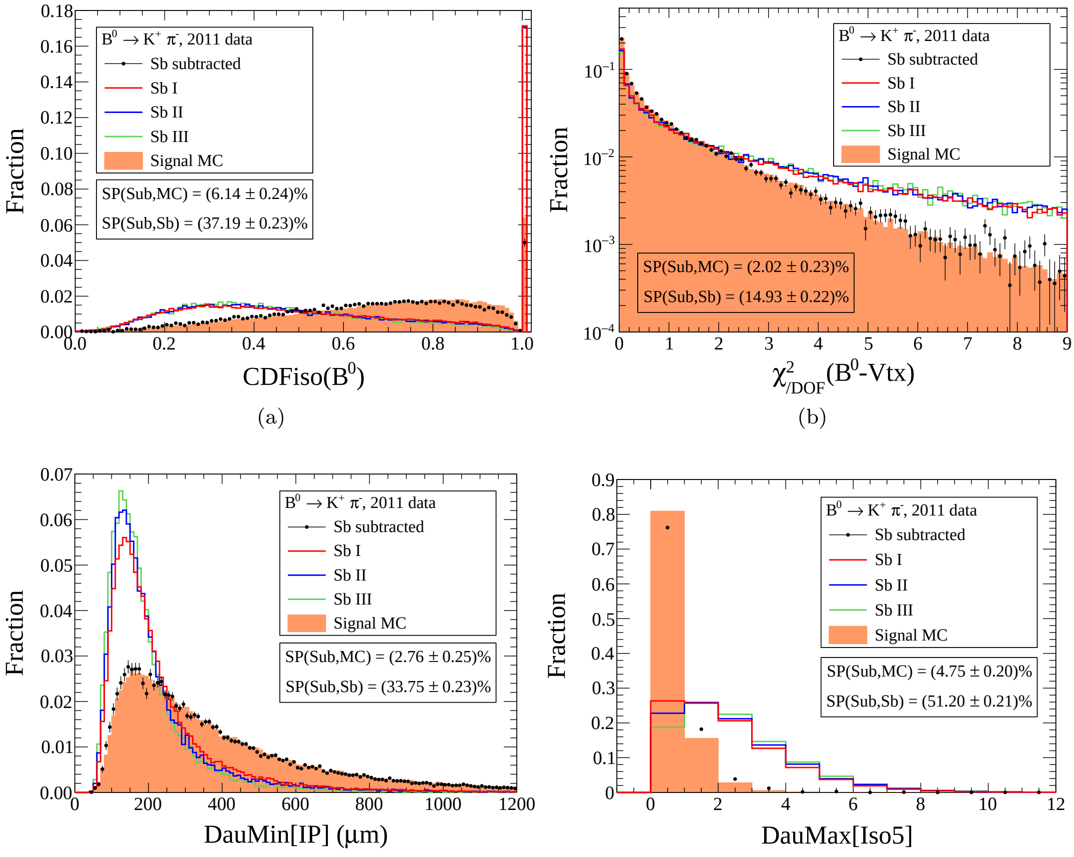
Figure A.15: Fits to 2011 mass distributions using the 2012 fit models with fixed parameters. There are about 303 k $B^+ \to J/\psi K^+$ events, 72.6 k $\Lambda_b \to \Lambda_c\pi^-$ events, 11.8 k $\Lambda_b \to J/\psi pK^-$ events with muon triggers, and about 3.0 k $\Lambda_b \to J/\psi pK^-$ events without muon triggers.

A.17 Comparison of MC and data in 2011 $B^0 \to K^+\pi^-$ events



(a) (b) (c) (d) (e) (f)

Figure A.16: Comparison of signal variables in different $B^0 \to K^+\pi^-$ data samples - Part 1.

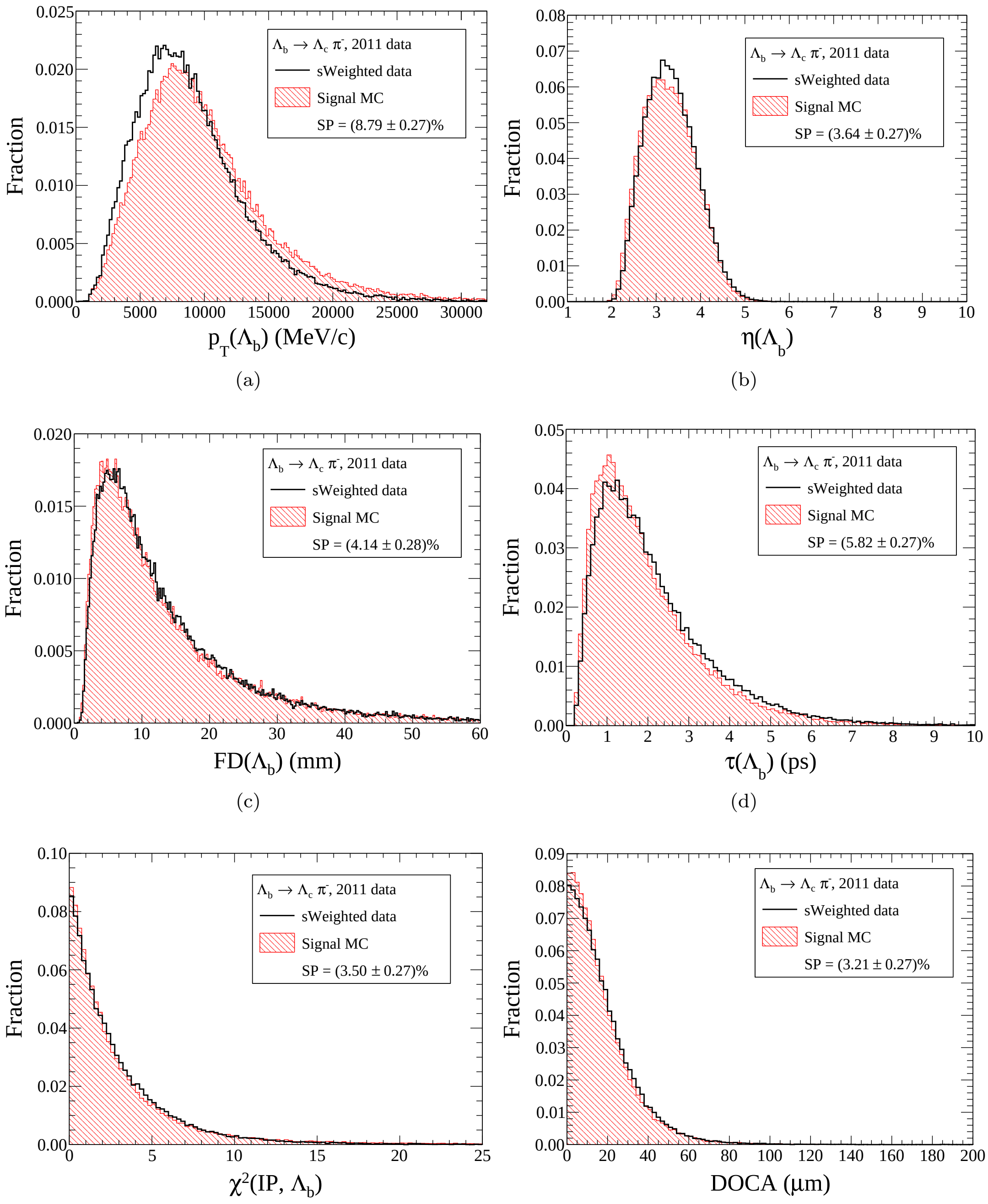


(a) (b)

(c) (d)

Figure A.17: Comparison of signal variables in different $B^0 \to K^+\pi^-$ data samples - Part 2.

A.18 Comparison of MC and data in 2011 $\Lambda_b \to \Lambda_c \pi^-$ events



(a) (b)

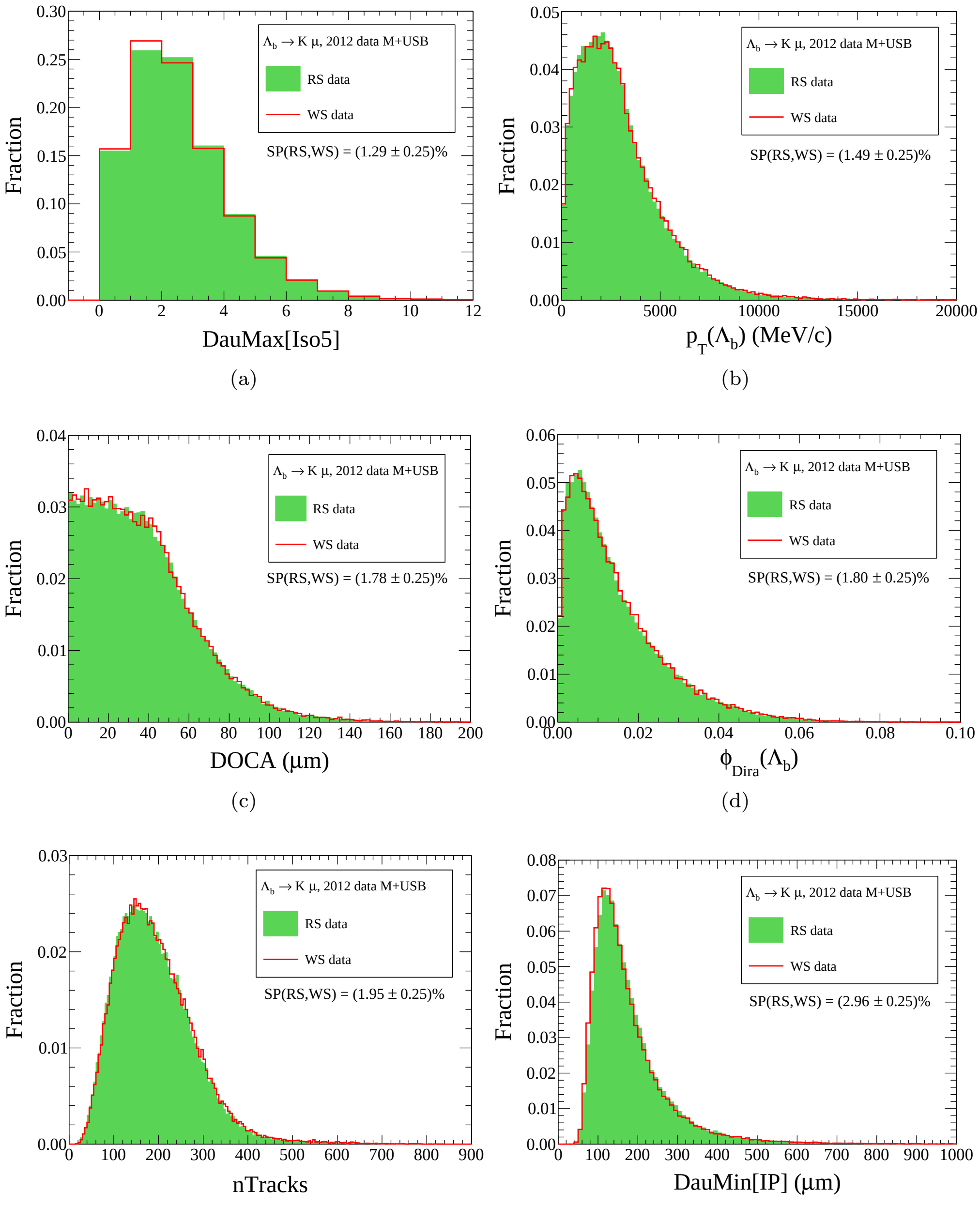
(c) (d)

(e) (f)

Figure A.18: Comparison of signal variables in different $\Lambda_b \to \Lambda_c \pi^-$ data samples.

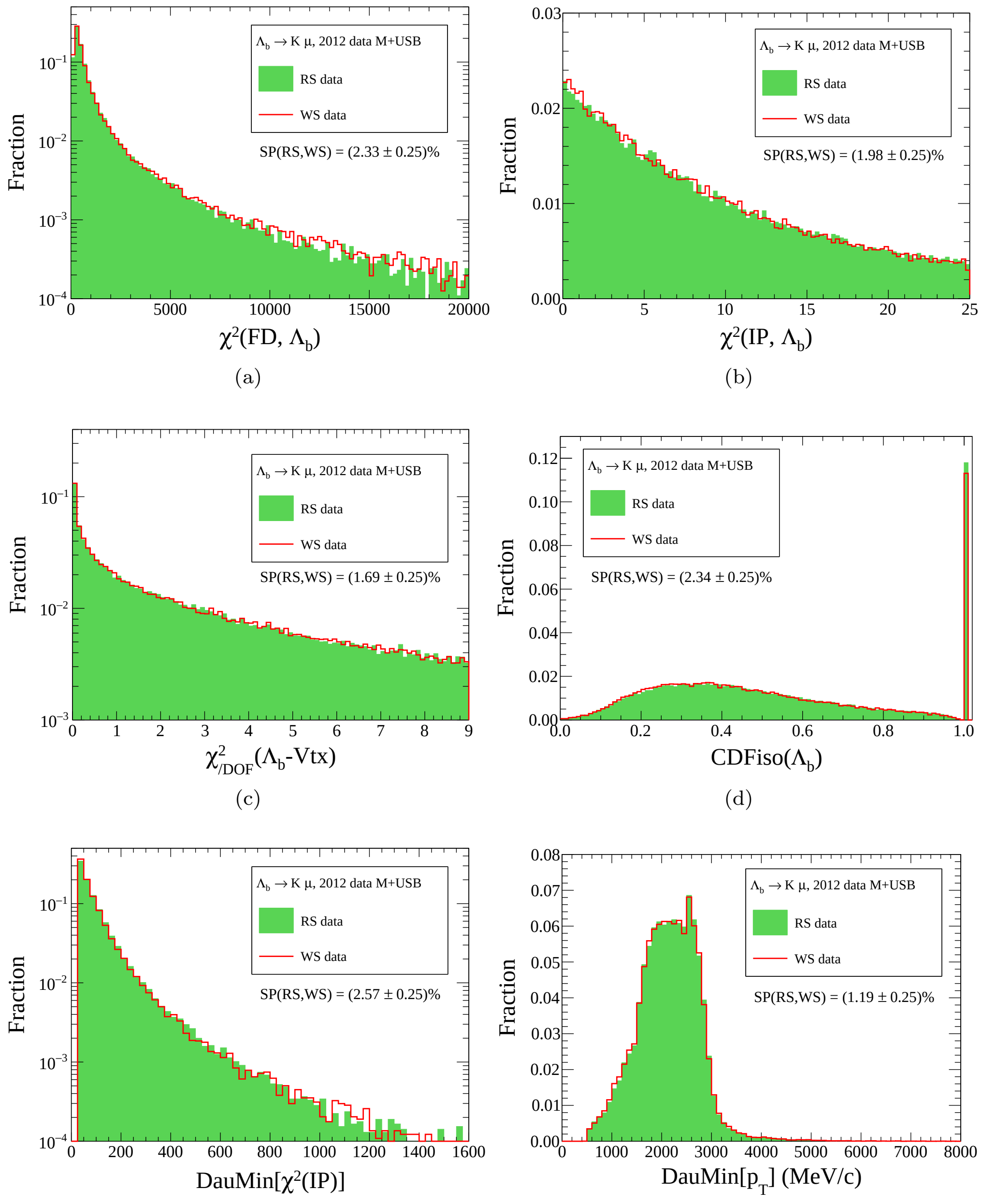
A.19 Comparison of RS and WS data

Figures A.19, A.20, and A.21 show comparisons of several variable distributions between RS and WS data from 2012 in the middle and upper sideband (M+USB).



(a) (b) (c) (d) (e) (f)

Figure A.19: Comparison of different variables in RS and WS data - Part 1.

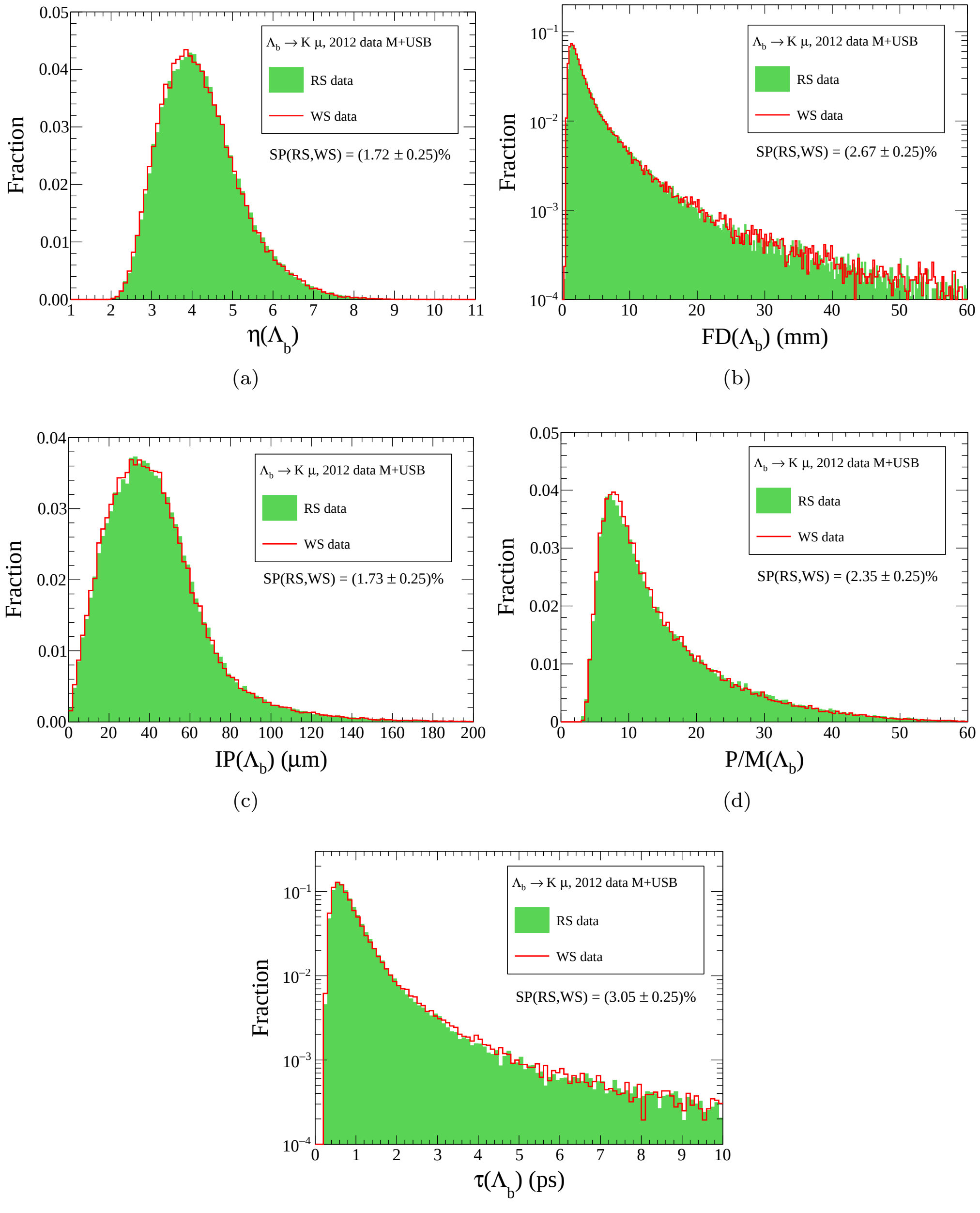


(a) (b)

(c) (d)

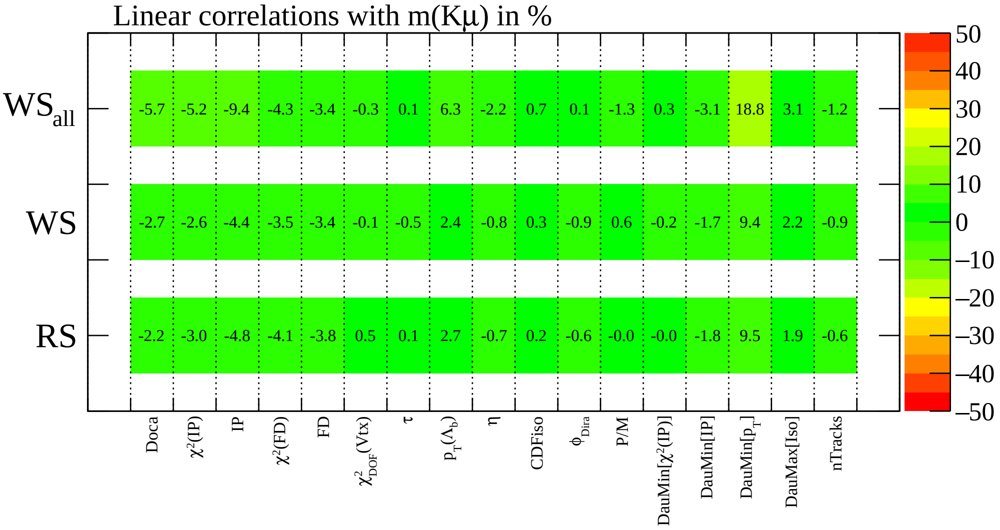
(e) (f)

Figure A.20: Comparison of variables in RS and WS data - Part 2.



(a) (b) (c) (d) (e)

Figure A.21: Comparison of variables in RS and WS data - Part 3.



	Doca	χ^2(IP)	IP	χ^2(FD)	FD	χ^2_{DOF}(Vtx)	τ	$p_T(\Lambda_b)$	η	CDFiso	ϕ_{Dira}	P/M	DauMin[χ^2(IP)]	DauMin[IP]	DauMin[p_T]	DauMax[Iso]	nTracks
WS_{all}	-5.7	-5.2	-9.4	-4.3	-3.4	-0.3	0.1	6.3	-2.2	0.7	0.1	-1.3	0.3	-3.1	18.8	3.1	-1.2
WS	-2.7	-2.6	-4.4	-3.5	-3.4	-0.1	-0.5	2.4	-0.8	0.3	-0.9	0.6	-0.2	-1.7	9.4	2.2	-0.9
RS	-2.2	-3.0	-4.8	-4.1	-3.8	0.5	0.1	2.7	-0.7	0.2	-0.6	-0.0	-0.0	-1.8	9.5	1.9	-0.6

Figure A.22: Linear correlation between $m(K\mu)$ and event variables. The row WS_{all} shows the correlation factors in the complete mass range of $K^+\mu^+$ combinations.

Table A.10: SP values comparing RS and WS data from 2011. The statistical uncertainties are smaller than 0.5 %, respectively.

Signal variables	**SP(RS,WS)**	**Signal variables**	**SP(RS,WS)**
DauMin[p_T]	2.11 %	χ^2(IP)	3.48 %
DauMax[Iso5]	2.47 %	χ^2(FD)	3.78 %
$p_\text{T}(\Lambda_b)$	2.76 %	Iso_CDF	3.10 %
$\chi^2_{/\text{DOF}}$(Vtx)	3.11 %	P/M(Λ_b)	3.15 %
$\eta(\Lambda_b)$	2.69 %	DauMin[χ^2(IP)]	2.98 %
IP(Λ_b)	2.93 %	FD	4.37 %
DOCA	2.67 %	DauMin[IP]	4.14 %
ϕ_Dira	2.88 %	τ	4.14 %
nTracks	3.11 %		

A.20 Systematic BDT studies

A.20.1 Importance of the used variables

To assess the importance of each variable, the performance of the BDT was evaluated when one of the event variables is removed. Table A.11 shows the loss of BDT performance for each variable that is omitted. According to the studies the most important variable is η and the least important variable is FD.

Table A.11: BDT efficiency loss after omitting a variable.

omitted Variable	$\Delta\varepsilon_{1\%}$	$\Delta\varepsilon_{100}$	$\Delta\varepsilon_{50}$
$\eta(\Lambda_b)$	−8.8 %	−13.8 %	−13.4 %
DauMax[Iso5]	−8.1 %	−8.2 %	−8.1 %
DauMin[IP]	−5.0 %	−6.2 %	−6.9 %
χ^2(IP)	−2.6 %	−3.0 %	−3.6 %
DOCA	−1.9 %	−3.6 %	−2.3 %
$\mathrm{Iso}_{\mathrm{CDF}}$	−1.7 %	−2.5 %	−3.3 %
DauMin[p_{T}]	−1.5 %	−2.4 %	−3.6 %
ϕ_{Dira}	−1.4 %	−1.2 %	−1.3 %
$\chi^2_{/\mathrm{DOF}}$(Vtx)	−0.8 %	−1.7 %	−1.8 %
FD(Λ_b)	−0.8 %	−0.0 %	−0.1 %

A.20.2 Addition of variables

None of the variables IP, χ^2(FD), p_{T}, τ, and DauMin[χ^2(IP)] provided a significant benefit $\Delta\varepsilon_{1\,\%} > 0.5\,\%$ to the BDT performance.

A.20.3 Substitution of variables

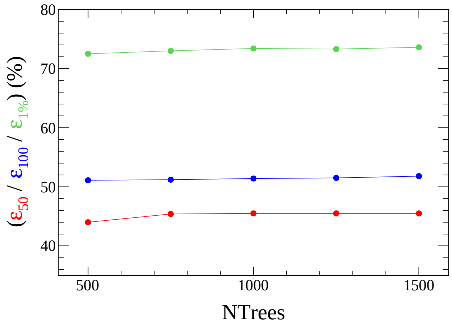
None of the following substitutions of physically similar variables provided a benefit to the BDT performance: χ^2(IP) → IP, FD → χ^2(FD), $\eta \to p_{\mathrm{T}}$, DauMin[IP] → DauMin[χ^2(IP)].

A.20.4 Variation of BDT tuning parameters

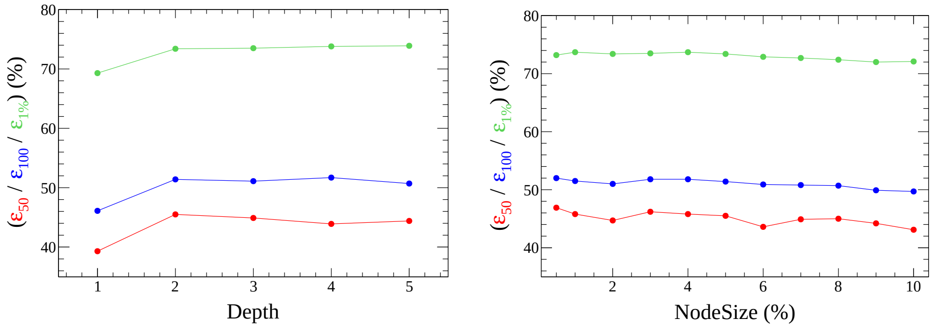
For this studies all four BDT parameters are varied independently from each other and the variation of the performance numbers ($\varepsilon_{1\%}/\varepsilon_{100}/\varepsilon_{50}$) is determined. Figure A.23 illustrates how these efficiencies change. In conclusion there is no big deviation of the efficiencies around the reference parameters (NTrees/NCuts/Depth/NodeSize) = (1000/30/2/5 %).

A.20.5 Comparison to other TMVA methods

Figure A.24 compares different TMVA methods including all BDT-like methods and a linear Fisher discriminant. All BDTs use the same parameters (Trees/NCuts/Depth/NodeSize) = (1000/30/2/5%). In conclusion the initial BDT method shows the best performance along with BDTG.

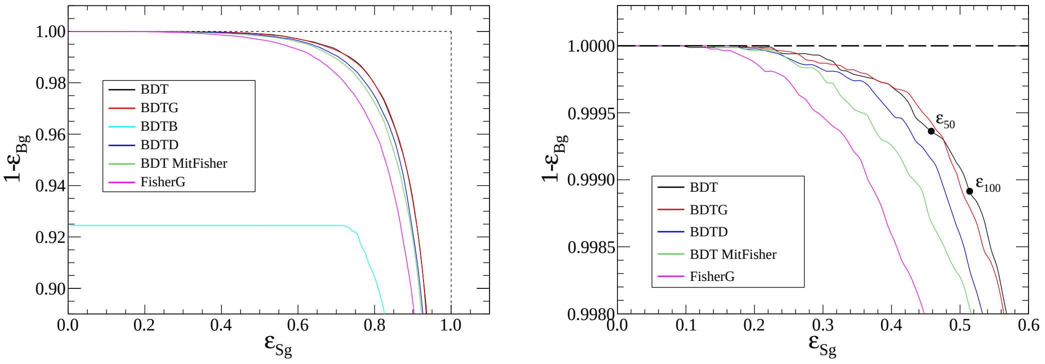


(a) NTrees (b) NCuts



(c) Depth (d) NodeSize

Figure A.23: Performance numbers in dependence of the BDT tuning parameters.

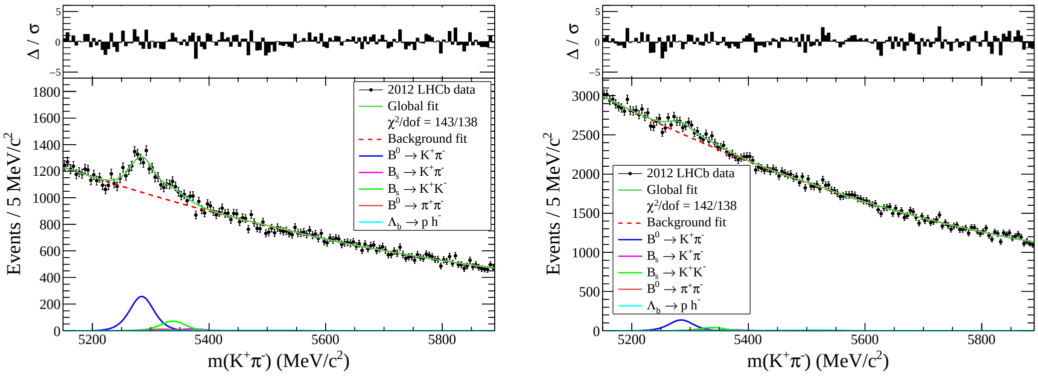


(a) Global performance (b) Performance in the low background region

Figure A.24: Comparison of different TMVA methods.

A.21 Muon misidentification in $B^0 \to K^+\pi^-$ events

The muon misidentification rate of the most common particles π and K is measured in the decay $B^0 \to K^+\pi^-$ whose tracks are kinematically similar to the ones of $\Lambda_b \to K^-\mu^+$. The misidentification may happen due to tracks that penetrate the calorimeter ("punch-through events") or in-flight decays $\{K, \pi\} \to \mu\nu_\mu$ before the calorimeter. Figure A.25 shows the $m(K^+\pi^-)$ distribution from 2012 data requiring IsMuon(K)=true (a) and IsMuon(π)=true (b). From the fits to the mass spectrum without muon requirement (see Fig. 4.2) the total number of events is around 168 k events. About 2800 kaons and 1500 pions satisfy IsMuon=true. This corresponds to $\varepsilon(K \to \mu) \approx 1.7\,\%$ and $\varepsilon(\pi \to \mu) \approx 0.9\,\%$ averaged over all momenta which is consistent with the PID studies concerning the IsMuon selector from Ref. [40] and shown in Fig. 2.12(b,c). The higher rate of kaons over pions can be understood on a qualitative level by the higher decay rate $\Gamma(K \to \mu\nu_\mu) \approx 1.34\,\Gamma(\pi \to \mu\nu_\mu)$ and the smaller nucleon cross section in the momentum range $5 - 100\,\mathrm{GeV}/c$ that makes kaons more likely to pass the calorimeters. In conclusion the muon misidentification rate is negligible for small numbers of $\Lambda_b \to K^-\mu^+$ events.



(a) IsMuon(K)=true

(b) IsMuon(π)=true

Figure A.25: Muon fake rate of kaons and pions in $B^0 \to K^+\pi^-$ events from 2012.

A.22 Error propagation

If $y = f(\boldsymbol{x})$ is a function that depends on several observables $\boldsymbol{x} = (x_i)$, the function can be developed as a Taylor series around the measured parameters $\tilde{\boldsymbol{x}} = (\tilde{x}_i)$ that have errors ε_i:

$$f(\boldsymbol{x})|_{\boldsymbol{x}=\tilde{\boldsymbol{x}}} = f(\tilde{\boldsymbol{x}}) + \sum_i \frac{\partial f}{\partial x_i}\,\varepsilon_i + \mathcal{O}(\varepsilon^2) \;.$$

In the first-order error approximation the second-order term is neglected. The standard deviation of ε_i is δ_i and the total standard deviation δy of $\varepsilon_y = f(\boldsymbol{x}) - f(\tilde{\boldsymbol{x}})$ can be estimated as

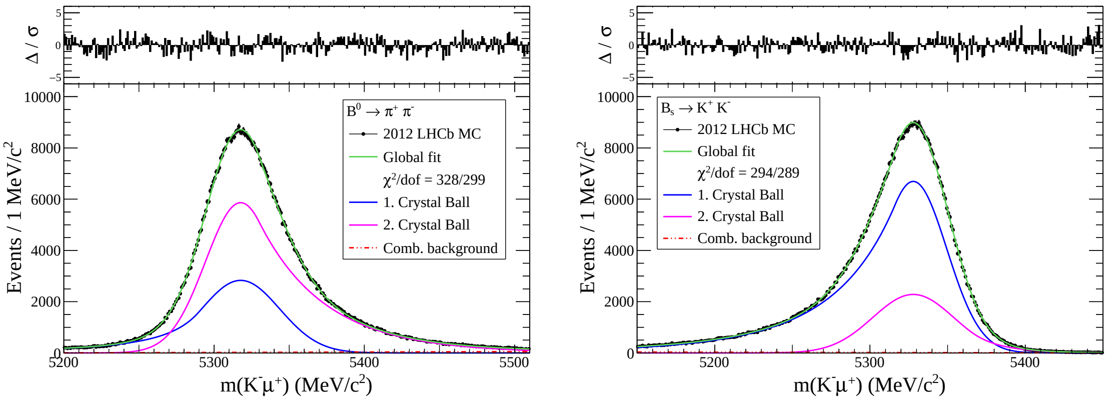
$$\delta y \approx \sqrt{\sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \delta_i^2 + 2\sum_{i\neq j} \mathrm{Cov}(x_i, x_j) \cdot \frac{\partial f}{\partial x_i} \cdot \frac{\partial f}{\partial x_j}} \;. \tag{A.12}$$

A.23 Fit model for the $m(K^-\mu^+)$ distribution

In the following the fit models for the different background components of the $m(K^-\mu^+)$ distribution is described. All fits are binned likelihood fits to the shown histograms.

Modelling of $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$

From fits to signal MC events, reconstructed as $\Lambda_b \to K^-\mu^+$, the shapes of the respective mass distributions can be described well by the sum of two Crystal Ball functions with a common mean value. Figure A.26 shows the result of the fits. For the fit to the $m(K^-\mu^+)$ distribution in data all parameters are fixed to the values determined from signal MC that are shown in Tab. A.12.



(a) $B^0 \to \pi^+\pi^-$ events (b) $B_s^0 \to K^+K^-$ events

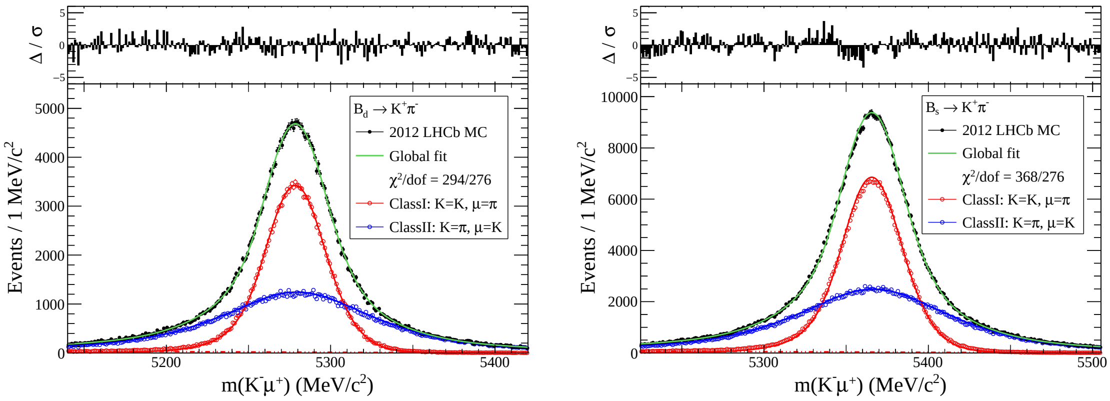
Figure A.26: Fits to the $m(K^-\mu^+)$ distribution from $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$ MC events.

Table A.12: Values of the fitted parameters of the signal components.

Parameter	$B^0 \to \pi^+\pi^-$	$B_s^0 \to K^+K^-$
μ	$(5317.69 \pm 0.09)\,\mathrm{MeV}/c^2$	$(5327.87 \pm 0.09)\,\mathrm{MeV}/c^2$
f_2	$(69.3 \pm 0.9)\,\%$	$(23.13 \pm 1.5)\,\%$
σ_1	$(26.00 \pm 0.22)\,\mathrm{MeV}/c^2$	$(21.62 \pm 0.20)\,\mathrm{MeV}/c^2$
σ_2	$(23.48 \pm 0.17)\,\mathrm{MeV}/c^2$	$(26.1 \pm 0.5)\,\mathrm{MeV}/c^2$
α_1	1.09 ± 0.025	0.535 ± 0.012
α_2	-0.534 ± 0.008	-1.45 ± 0.06
n_1	2.14 ± 0.16	7.8 ± 1.1
n_2	44 ± 8	17 ± 4

Modelling of $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+\pi^-$

The shape of $B^0 \to K^+\pi^-$ events, reconstructed as $\Lambda_b \to K^-\mu^+$, is determined from signal MC studies. For a correct parametrisation two event classes are considered: (I) events where $(K = K, \mu = \pi)$ and (II) $(K = \pi, \mu = K)$. Both event classes can be fitted well using the sum of two Crystal Ball functions with a common mean value. Figure A.27(a) shows the results of the binned likelihood fits to the two event classes of simulated $B^0 \to K^+\pi^-$ events. For the fit to the combined distribution all parameters are fixed except for the fraction of the second mode f_2 and the normalisation N. For $B_s^0 \to K^+\pi^-$ events the same fit model with fixed parameters from $B^0 \to K^+\pi^-$ MC are used except for the mean value and the normalisation. The fitted mean values of the two event classes are $\mu_I = (5365.61 \pm 0.05)\,\mathrm{MeV}/c^2$ and $\mu_{II} = (5366.47 \pm 0.15)\,\mathrm{MeV}/c^2$. In conclusion the result is slightly worse in terms of χ^2/dof, but, however, the parametrisation matches the data still well.



(a) $B^0 \to K^+\pi^-$ events (b) $B_s^0 \to K^+\pi^-$ events

Figure A.27: Fits to the $m(K^-\mu^+)$ distribution from $B^0/B_s^0 \to K^+\pi^-$ MC events.

Table A.13: Values of the fitted parameters of the two event classes of $B^0 \to K^+\pi^-$.

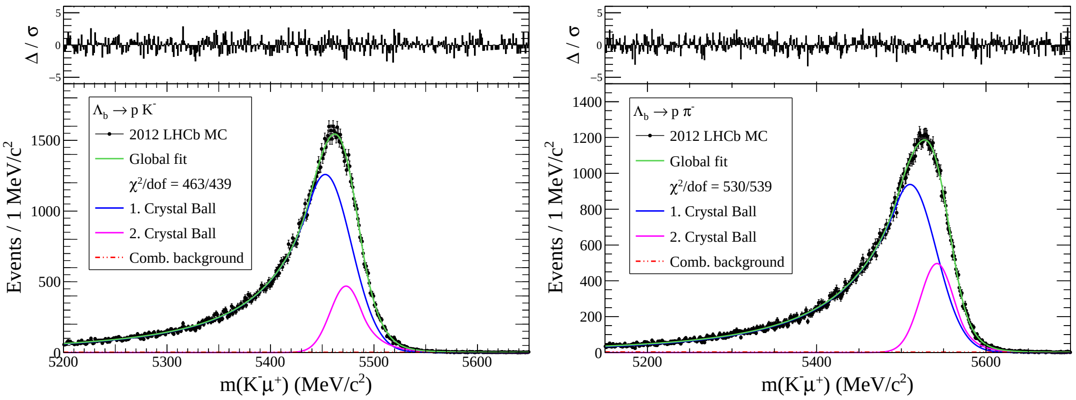
Parameter	$I\ (K = K, \mu = \pi)$	$II\ (K = \pi, \mu = K)$
μ	$(5278.46 \pm 0.07)\,\mathrm{MeV}/c^2$	$(5279.38 \pm 0.22)\,\mathrm{MeV}/c^2$
f_2	$(34.7 \pm 1.1)\,\%$	$(45.1 \pm 0.8)\,\%$
σ_1	$(19.56 \pm 0.18)\,\mathrm{MeV}/c^2$	$(43.1 \pm 0.4)\,\mathrm{MeV}/c^2$
σ_2	$(16.82 \pm 0.21)\,\mathrm{MeV}/c^2$	$\equiv \sigma_1$
α_1	1.142 ± 0.018	0.567 ± 0.012
α_2	$\equiv -\alpha_1$	$\equiv -\alpha_1$
n_1	24 ± 6	10 (fixed)
n_2	$\equiv n_1$	$\equiv n_1$
χ^2/dof	302/271	299/273

Modelling of $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$

The mass shape of the decay modes $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$, both reconstructed as $\Lambda_b \to K^-\mu^+$, have been determined in simulated events. Both event classes are fitted separately using the sum of two Crystal Ball functions. The fits to the complete sample of reconstructed Λ_b decays in simulated data is shown in Fig. A.28. Here, all parameters are fixed except for the normalisation and the ratio of the two event classes. After requiring the PID criteria from the event selection only the event class $K = p, \mu = K$ remains for $\Lambda_b \to pK^-$ events and $K = p, \mu = \pi$ for $\Lambda_b \to p\pi^-$ events. This can be understood by the fact that kaons and pions may decay in the detector and have a muon in the final state being only slightly deflected from the kaon or pion flight direction. In the fit to the data only this event class is considered. Table A.14 shows the fitted parameters.

Modelling of $B_s^0 \to K^+\mu^-\nu_\mu$ and $\Lambda_b \to p\mu^-\nu_\mu$

These two partially reconstructed decay modes can be described well by the right tail of a Gaussian distribution. Studies in simulated events shows that only the event class with $\mu = \mu$ and $K =$



(a) $B^0 \to K^+\pi^-$ events (b) $B_s^0 \to K^+\pi^-$ events

Figure A.28: Fits to the $m(K^-\mu^+)$ distribution from $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ MC events.

Table A.14: Values of the fitted parameters for $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ MC events.

Parameter	$\Lambda_b \to pK^-$ $(K = p, \mu = K)$	$\Lambda_b \to p\pi^-$ $(K = p, \mu = \pi)$
f_2	$(14.0 \pm 1.7)\,\%$	$(17.8 \pm 1.7)\,\%$
μ_1	$(5452.91 \pm 0.9)\,\mathrm{MeV}/c^2$	$(5510.3 \pm 1.3)\,\mathrm{MeV}/c^2$
μ_2	$(5473.03 \pm 0.9)\,\mathrm{MeV}/c^2$	$(5542.1 \pm 1.0)\,\mathrm{MeV}/c^2$
σ_1	$(25.3 \pm 0.5)\,\mathrm{MeV}/c^2$	$(31.2 \pm 0.8)\,\mathrm{MeV}/c^2$
σ_2	$(15.4 \pm 0.6)\,\mathrm{MeV}/c^2$	$(19.9 \pm 0.5)\,\mathrm{MeV}/c^2$
α_1	0.572 ± 0.014	0.480 ± 0.014
α_2	-1.03 ± 0.08	-1.39 ± 0.06
n_1	2.46 ± 0.17	3.5 ± 0.4
n_2	10 (fixed)	10 (fixed)

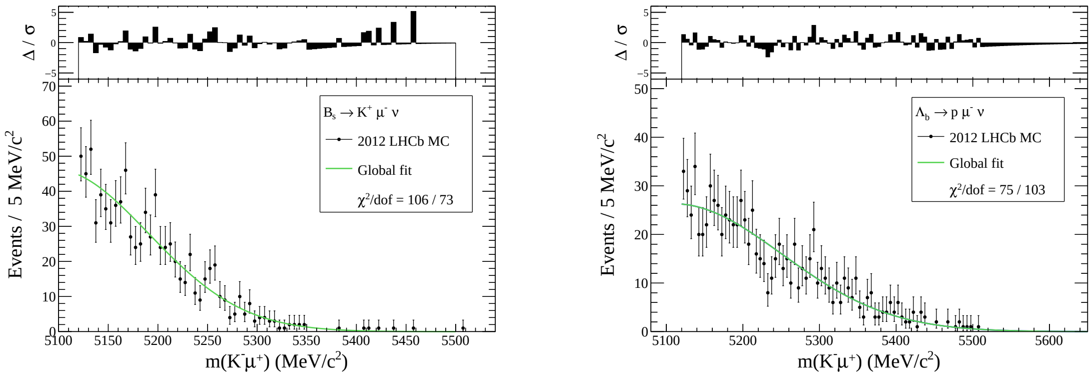
$\{K, p\}$ relevant due to the IsMuon=true requirement on the muon track. Figure A.29 shows the result of a binned likelihood fit. For the fit to the $m(K^-\mu^+)$ distribution in data the parameters are fixed to the results from MC. The fitted parameters are $\mu = (5094 \pm 20)\,\mathrm{MeV}/c^2$ and $\sigma = (96 \pm 7)\,\mathrm{MeV}/c^2$ in the $B_s^0 \to K^+\mu^-\nu_\mu$ case and $\mu = (5110 \pm 24)\,\mathrm{MeV}/c^2$ and $\sigma = (141 \pm 10)\,\mathrm{MeV}/c^2$ for simulated $\Lambda_b \to p\mu^-\nu_\mu$ events.

Combinatorial background

The shape of the combinatorial is determined from fits to the $m(K^+\mu^+)$ distribution after stripping and BDT selections, and is studied in Sec. 5.2.4. As a result the background can be described well by a first-order polynomial for selections $\mathrm{x_{BDT}} > 0$. In the fit to the $m(K^-\mu^+)$ distribution in data the slope is required to be negative to allow only a monotonically decreasing background.

Test of the fit model in real data

The complete fit model is tested in data using all models as described above. A mass shift of $+5\,\mathrm{MeV}/c^2$ is added to the mean values of all b-hadron shapes as determined from the fit to



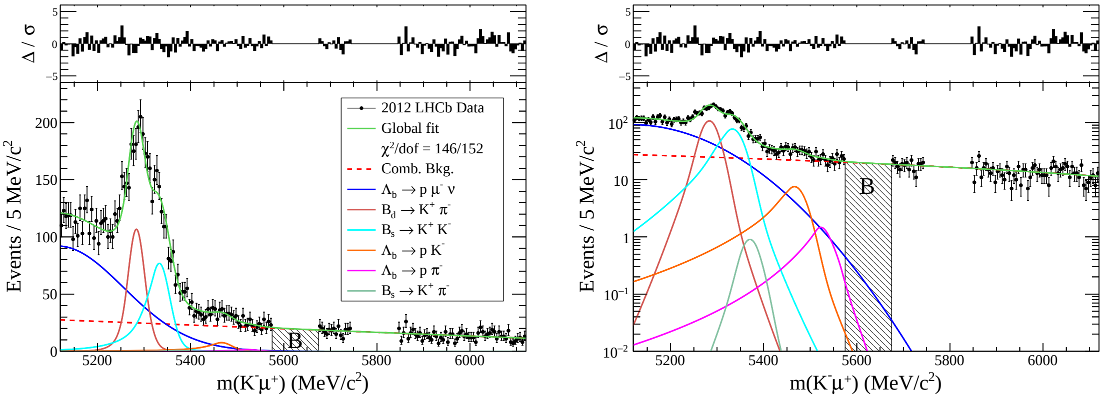
(a) $B_s^0 \to K^+\mu^-\nu_\mu$ events
(b) $\Lambda_b \to p\mu^-\nu_\mu$ events

Figure A.29: Fits to the $m(K^-\mu^+)$ distribution from $B_s^0 \to K^+\mu^-\nu_\mu$ and $\Lambda_b \to p\mu^-\nu_\mu$ MC events.

the $B \to K^+\pi^-$ spectrum in Sec. 4.1.2. To improve the sensitivity on all modes, the amount of combinatorial background is reduced by requiring $\mathrm{x_{BDT}} > 0.1$. As a result of the fit the significances of the decay modes $B^0 \to \pi^+\pi^-$, $B_s^0 \to K^+\pi^-$ and $B_s^0 \to K^-\mu^+\nu_\mu$, measured as $N/\delta(N)$, are close to zero and thus these modes are neglected. Figure A.30 shows the result of the fit without the these modes using (a) linear and (b) logarithmic scales. According to the fit there are 3152 ± 123 $\Lambda_b \to p\mu^-\nu_\mu$ events, 1049 ± 59 $B^0 \to K^+\pi^-$ events, 1159 ± 67 $B_s^0 \to K^+K^-$ events, 141 ± 45 $\Lambda_b \to pK^-$ events, 23 ± 40 $\Lambda_b \to p\pi^-$ events, and 3904 ± 203 background events.

The absence of $B_s^0 \to K^-\mu^+\nu_\mu$ events in comparison to $\Lambda_b \to p\mu^-\nu_\mu$ can be explained by the lower mass and production rate of B_s^0 mesons. The lack of $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+\pi^-$ events can be explained by their smaller branching ratio w.r.t. $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+K^-$, and as a consequence the few signal events are absorbed in the fitted normalisation of the dominant modes.

In conclusion the fit model describes the mass spectrum well and allows the interpolation of the expected number of background events B for the optimisation of the signal selection. According to the fit the correlations factors of B with the normalisations are: $-70\,\%$ with $N(\Lambda_b \to p\mu^-\nu_\mu)$, $-3\,\%$ with $N(B^0 \to K^+\pi^-)$, $-26\,\%$ with $B_s^0 \to K^+K^-$, $+51\,\%$ with $\Lambda_b \to pK^-$, and $-62\,\%$ with $\Lambda_b \to p\pi^-$.



(a)
(b)

Figure A.30: Fit to the $m(K^-\mu^+)$ distribution from data requiring $\mathrm{x_{BDT}} > 0.1$.

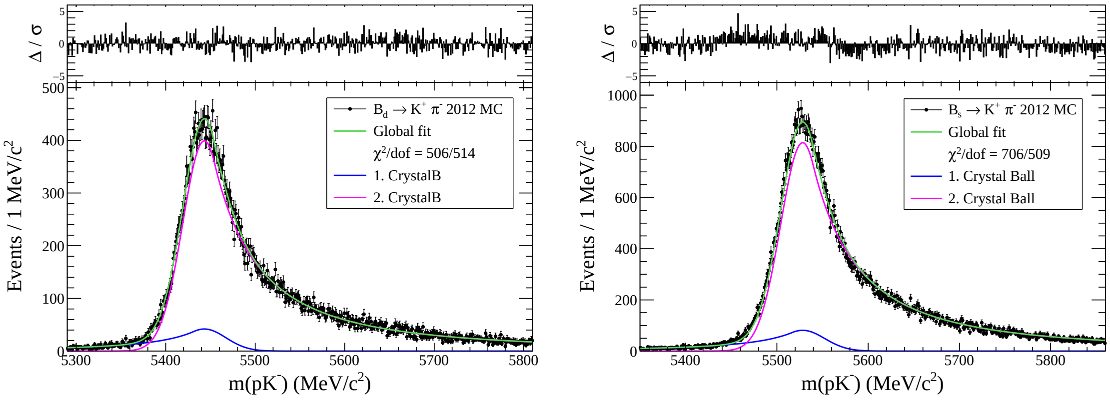
A.24 Fit model for the $m(pK^-)$ distribution

To develop a fit model for the $m(pK^-)$ distribution, simulated events of the most common two-body decays of B^0, B_s^0 and Λ_b are reconstructed as $\Lambda_b \to pK^-$ and their mass spectrum is fitted. These modes are $B^0 \to \pi^+\pi^-$, $B^0 \to K^+\pi^-$, $B_s^0 \to K^+\pi^-$, $B_s^0 \to K^+K^-$, $B_s^0 \to K^-\mu^+\nu_\mu$, $\Lambda_b \to pK^-$, $\Lambda_b \to p\pi^-$ and $\Lambda_b \to p\mu^-\nu_\mu$. For all reconstructed candidates the signal triggers are required and no PID selections are made. All fits are binned likelihood fits.

A.24.1 $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+\pi^-$

A dominant reflection origins from the decay $B^0 \to K^+\pi^-$ with two kinds of misID (I) $p = K$ and $K = \pi$, and (II) $p = \pi$ and $K = K$. For $B^0 \to K^+\pi^-$ events the mass distributions of both classes are fitted separately. In both cases the sum of two Crystal Ball functions with a common mean value is used. For $B_s^0 \to K^+\pi^-$ events the same fit model and parameters are used except for the invariant mass that is determined from an unconstrained fit. In test fits to the $m(pK^-)$ distribution in data, all parameters are fixed from MC except for the normalisation and the ratio of the two modes. It turns out that the first event class is dominating to about 100 % and hence the second events class can be neglected to simplify the fit.

Figure A.31 shows the fit to the $m(pK^-)$ spectrum of the first events class from simulated $B^0 \to K^+\pi^-$ events (a) and $B_s^0 \to K^+\pi^-$ events (b). The mass spectrum of $B_s^0 \to K^+\pi^-$ events can be described by the $B^0 \to K^+\pi^-$ fit model. The separately fitted mass is $(5528.24 \pm 0.25)\,\mathrm{MeV}/c^2$. Table A.31 shows the fit parameters.



(a) $B^0 \to K^+\pi^-$. (b) $B_s^0 \to K^+\pi^-$.

Figure A.31: Fits to the mass distributions of $B_{(s)}^0 \to K^+\pi^-$ events in 2012 MC.

A.24.2 $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$

The spectrum of both modes is fitted using the sum of two Crystal Ball functions that have a common mean value. Figure A.32 shows the results of the fits to both distributions. The addition of a linear function to describe combinatorial background events is not necessary.

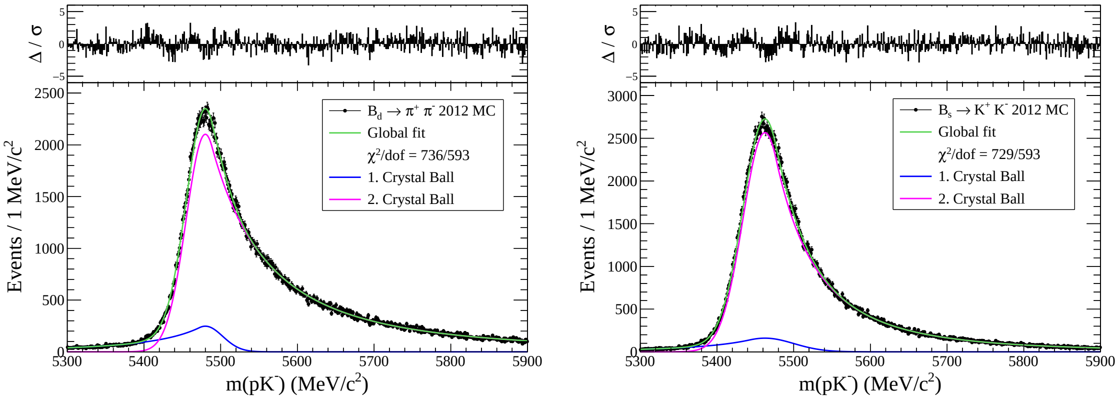
From the fit to the data it can be concluded that the decay mode $B^0 \to \pi^+\pi^-$ is irrelevant which can be explained by a double suppression pion tracks due to the PID selections.

A.24.3 $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$

The signal mode $\Lambda_b \to pK^-$ is fitted using the sum of two Crystal Ball functions with a common mean value. In MC the event class $p = K$ and $K = p$ is largely suppressed by the event selection and not considered. Figure A.33(a) shows the result of the fit.

Table A.15: Values of the fitted parameters to describe $B^0 \to K^+\pi^-$ and $B_s^0 \to K^+\pi^-$ events.

Parameter	$I(p = K, K = \pi)$
f_2	$(91.9 \pm 0.3)\,\%$
μ	$(5443.2 \pm 0.4)\,\mathrm{MeV}/c^2$
σ_1	$(24.135 \pm 0.3)\,\mathrm{MeV}/c^2$
σ_2	$\equiv \sigma_1$
α_1	0.572 ± 0.014
α_2	$\equiv -\alpha_1$
n_1	1.98 ± 0.07
n_2	$\equiv n_2$



(a) $B^0 \to \pi^+\pi^-$.

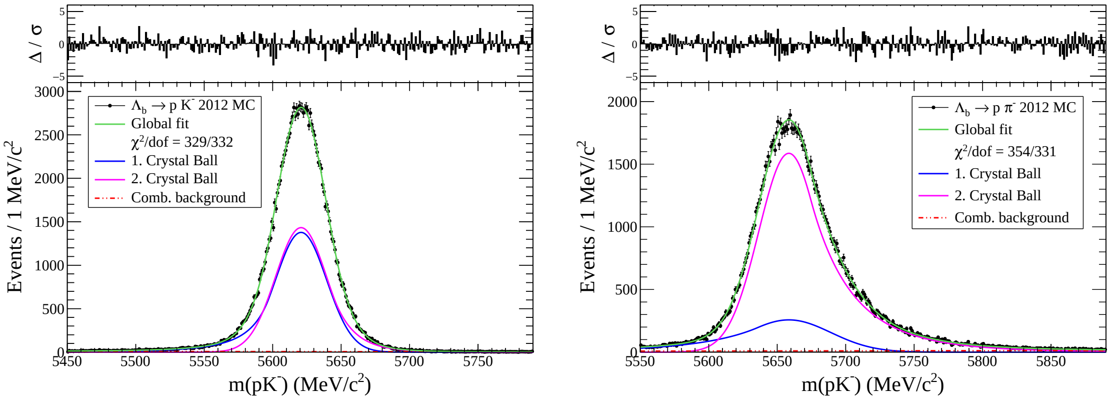
(b) $B_s^0 \to K^+K^-$.

Figure A.32: Fits to the mass distributions of $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$ events in 2012 MC.

Table A.16: Values of the fitted parameters for $B^0 \to \pi^+\pi^-$ and $B_s^0 \to K^+K^-$ events.

Parameter	$B^0 \to \pi^+\pi^-$	$B_s^0 \to K^+K^-$
f_2	$(91.71 \pm 0.14)\,\%$	$(93.62 \pm 0.16)\,\%$
μ	$(5480.0 \pm 0.2)\,\mathrm{MeV}/c^2$	$(5462.52 \pm 0.16)\,\mathrm{MeV}/c^2$
σ_1	$(22.6 \pm 0.9)\,\mathrm{MeV}/c^2$	$(37.6 \pm 2.0)\,\mathrm{MeV}/c^2$
σ_2	$(25.1 \pm 0.2)\,\mathrm{MeV}/c^2$	$(28.63 \pm 0.13)\,\mathrm{MeV}/c^2$
α_1	0.344 ± 0.004	0.528 ± 0.006
α_2	$\equiv -\alpha_1$	$\equiv -\alpha_1$
n_1	2.45 ± 0.05	3.10 ± 0.06
n_2	$\equiv n_1$	$\equiv n_1$

For the mode $\Lambda_b \to p\pi^-$ the situation the same fit model is used. The event class $p = \pi$ and $K = p$ is suppressed by the event selection and not considered. Figure A.33(b) shows the result of the fit and Tab. A.17 the determined parameters.



(a) $\Lambda_b \to pK^-$.

(b) $\Lambda_b \to p\pi^-$.

Figure A.33: Fits to the mass distributions of $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ events in 2012 MC.

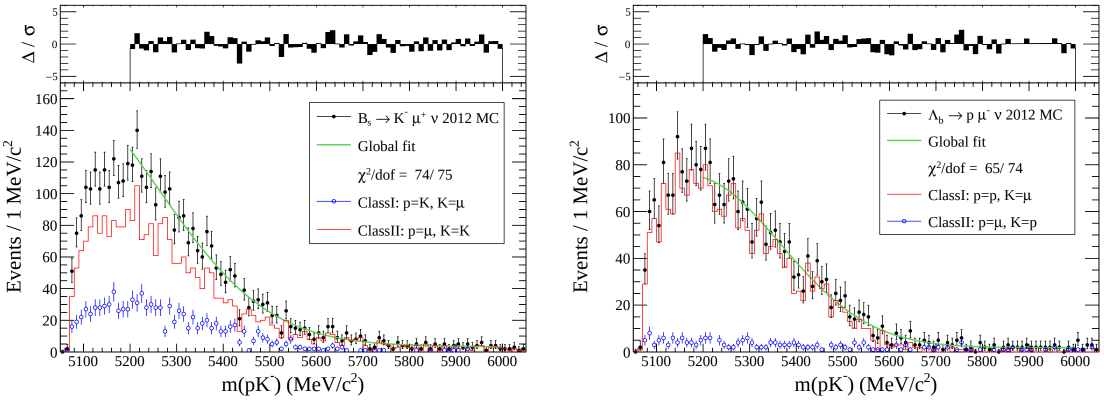
Table A.17: Values of the fitted parameters for $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ events.

Parameter	$\Lambda_b \to pK^-$	$\Lambda_b \to p\pi^-$
f_2	50 % (fixed)	(82.6 ± 1.4) %
μ	(5620.71 ± 0.06) MeV/c^2	(5648.5 ± 0.14) MeV/c^2
σ_1	(18.60 ± 0.08) MeV/c^2	(30.8 ± 1.8) MeV/c^2
σ_2	$\equiv \sigma_1$	(21.98 ± 0.19) MeV/c^2
α_1	1.260 ± 0.019	0.730 ± 0.023
α_2	$\equiv -\alpha_1$	$\equiv -\alpha_1$
n_1	3.6 ± 0.4	18.10 ± 0.06
n_2	30 ± 3	$\equiv n_1$

A.24.4 Partially reconstructed decays $B_s^0 \to K^-\mu^+\nu_\mu$ and $\Lambda_b \to p\mu^-\overline{\nu}_\mu$

For both decay modes a separation into the two possible misID classes is not made due to the similar shape of both distributions and the small number of events. Both mass spectra a fitted using the sum of a first-order polynomial and a Gaussian distribution to describe the right shoulder of the spectrum. Figure A.34(a) and (b) show the results of the fits. The fitted parameters are $\mu = (5033 \pm 111)$ MeV/c^2 and $\sigma = (230 \pm 41)$ MeV/c^2 for $B_s^0 \to K^-\mu^+\nu_\mu$ and $\mu = (5173 \pm 39)$ MeV/c^2 and $\sigma = (193 \pm 22)$ MeV/c^2 for $\Lambda_b \to p\mu^-\overline{\nu}_\mu$.

In data the contribution from the $B_s^0 \to K^-\mu^+\nu_\mu$ mode is negligible and hence is not included to simplify the fit.



(a) $B_s^0 \to K^+\mu^-\nu_\mu$.

(b) $\Lambda_b \to p\mu^-\nu_\mu$.

Figure A.34: Fits to the mass distributions of $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ events in 2012 MC.

A.24.5 Fit to the data

For the fit to the data the models as described before are used with all parameters fixed from MC. All mean values are increased by a floating parameter D to consider the mass shift. For the combinatorial background a first-order polynomial is used.

A.25 Particle identification using the Cerenkov Effect

The Cerenkov angle is $\cos(\theta_c) = 1/(\beta \cdot n)$ and the separation $x - y$ of two different mass hypotheses $m = m_x$ and $m = m_y$ is

$$\Delta \cos(xy) = \cos(\theta_c)_x - \cos(\theta_c)_y = \frac{1}{\beta_x \cdot n} - \frac{1}{\beta_y \cdot n} \tag{A.13}$$

for tracks with a given momentum p, velocity $\beta_x = p/E_x$, and the radiator with a refractive index n. The condition for the radiation of Cerenkov light is $\beta\, n > 1$ or equivalent to this $p > m/\sqrt{n^2 - 1}$. Equation A.13 can be rewritten as

$$\Delta \cos(xy) = \frac{E_x - E_y}{p \cdot n} \tag{A.14}$$

The energy difference can be written as $E_x - E_y = \sqrt{p^2 + m_x^2} - \sqrt{p^2 + m_y^2}$. In case $p \gg m$ the root can be approximated as $\sqrt{p^2 + m^2} \approx p + m^2/2/p$ and $E_x - E_y \approx (m_x^2 - m_y^2)/2/p$. The ratio of two different mass separations is therefore

$$\frac{\Delta \cos(xy)}{\Delta \cos(ab)} = \frac{E_x - E_y}{E_a - E_b} \approx \frac{m_x^2 - m_y^2}{m_a^2 - m_b^2} \tag{A.15}$$

The numerical values of the approximation are

$$\frac{\Delta \cos(p\pi)}{\Delta \cos(K\pi)} \approx \frac{3.83}{1}\ , \qquad \frac{\Delta \cos(pK)}{\Delta \cos(K\pi)} \approx \frac{2.84}{1}\ . \tag{A.16}$$

Figure A.35 shows the exact formula for both ratios. The limit is achieved rapidly. In summary the $p - K$ and $p - \pi$ separation is always better than the $K - \pi$ separation.

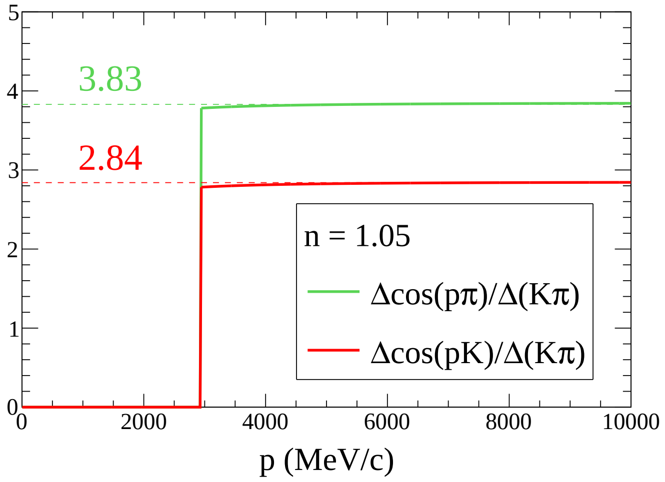


Figure A.35: Ratio of the PID separation pK *vs.* $K\pi$ and $p\pi$ *vs.* $K\pi$.

A.26 L0-Muon-TOS efficiency determination

For the determination of the L0-Muon-TOS efficiency the data from the $B^+ \to J/\psi K^+$ stripping line are used. The efficiencies are calculated in bins of $p_\mathrm{T}(\mu)$ using the the TISTOS method assuming that the TOS efficiency in a sample of TIS triggered events is identical with the true TOS efficiency in a sample of untriggered events:

$$\varepsilon_{TOS}(p_\mathrm{T}) := \frac{N_{TOS}(p_\mathrm{T})}{N_\mathrm{sel}(p_\mathrm{T})} \overset{!}{=} \frac{N_{TISTOS}(p_\mathrm{T})}{N_{TIS}(p_\mathrm{T})}$$

The reasoning for this is assumption is the fact that the decay of a b hadron is independent from the properties of the underlying event. Hence the trigger signatures of a certain b-hadron decay (TOS) and the underlying event (TIS) are largely uncorrelated.

Figure A.36 shows the total sample of L0-TIS triggered $B^+ \to J/\psi K^+$ events and a linear fit to the left and right sideband. The number of signal events (S) is determined from a sideband subtraction; the number of background events in the left (L) and right (R) sideband is determined and the number of background events in the signal region (B) is estimated. S is then determined as $S = A - B$ where A is the total number of events in the signal region. In total there are about 225 000 TIS triggered events.

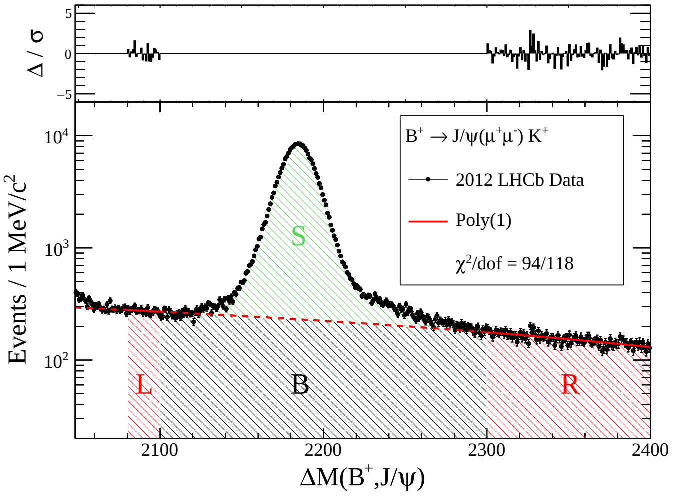
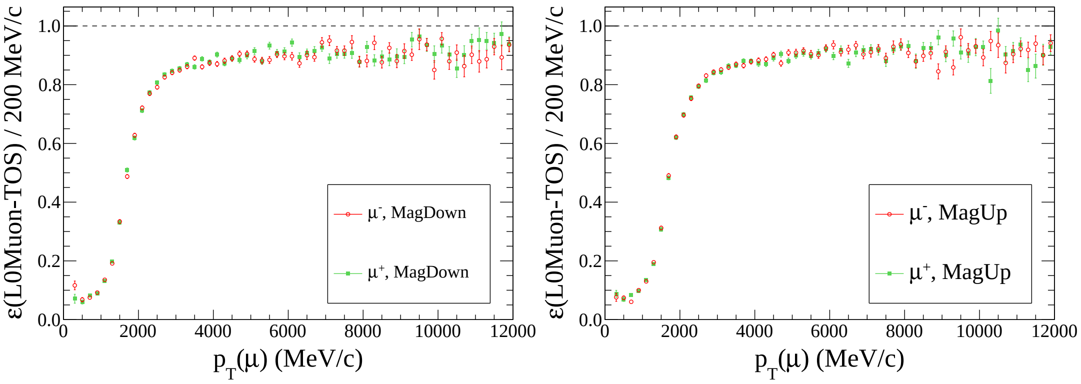


Figure A.36: $\Delta m(B^+, J/\psi)$ spectrum of TIS triggered events and fit to the background.

The values of $\varepsilon_{TOS}(p_\mathrm{T}) = N_{TISTOS}/N_{TIS}$ are determined for both charges and magnet polarities separately and using a bin width of 200 MeV/c. For each p_T bin the number of "L0-Global-TIS" events N_{TIS} and "L0-TIS `AND` L0-Muon-TOS" events N_{TISTOS} is determined from a sideband subtraction as demonstrated in Fig. A.36. Figure A.37 shows the calibration histograms for both magnet polarities.

The statistical uncertainty on the efficiency is calculated as follows: For a particular p_T interval the following number of the signal range are given: the total number of TIS events N_0, the estimated number of TIS background events in the signal band B_0, the total number of TISTOS events N_1, and the estimated number of TISTOS background events in the signal band B_1. It is assumed that N_1/N_0 and B_1/B_0 follow a binomial distribution, *i.e.* $\delta(N_1) = \sqrt{N_1 \cdot (1 - N_1/N_0)}$ and $\delta(B_1) = \sqrt{B_1 \cdot (1 - B_1/B_0)}$. The uncertainty on the number of TISTOS signal events is $\delta(S_1) = \sqrt{\delta^2(N_1) + \delta^2(B_1)}$ and the efficiency including the uncertainty is:

$$\varepsilon_{TOS} = \frac{N_1 - B_1}{N_0 - B_0} \pm \frac{\sqrt{\delta^2(N_1) + \delta^2(B_1)}}{N_0 - B_0} = \frac{S_1}{S_0} \pm \frac{\delta(S_1)}{S_0}$$



(a) (b)

Figure A.37: $\varepsilon_{TOS}(p_{\mathrm{T}})$ determined from $B^+ \to J/\psi K^+$ events.

A.27 L0-TIS data points and fit parameters

Table A.18: Weighted L0-TOS efficiencies for K and μ.

	2011	2012
$\varepsilon_{TIS}^{MC}(B^+ \to J/\psi K^+)$	$(2.60 \pm 0.03)\,\%$	$(3.69 \pm 0.02)\,\%$
$\varepsilon_{TIS}^{Data}(B^+ \to J/\psi K^+)$	$(2.72 \pm 0.03)\,\%$	$(3.65 \pm 0.02)\,\%$
$\varepsilon_{TIS}^{MC}(\Lambda_b \to J/\psi p K^+)$	$(3.23 \pm 0.06)\,\%$	$(4.40 \pm 0.08)\,\%$
$\varepsilon_{TIS}^{Data}(\Lambda_b \to J/\psi p K^+)$	$(3.24 \pm 0.19)\,\%$	$(4.4 \pm 0.6)\,\%$
$\varepsilon_{TIS}^{MC}(\Lambda_b \to p K^-)$	$(32.15 \pm 0.26)\,\%$	$(35.34 \pm 0.22)\,\%$
$\varepsilon_{TIS}^{Data}(\Lambda_b \to p K^-)$	$(47.8 \pm 3.1)\,\%$	$(51.2 \pm 2.0)\,\%$
$f_1 = m\,x + n$	$m = 1.47 \pm 0.10$	$m = 1.50 \pm 0.07$
	$n = (-1.13 \pm 0.28)\,\%$	$n = (-1.88 \pm 0.25)\,\%$
	$\mathrm{Cov}(m,n) = -2.9 \cdot 10^{-2}$	$\mathrm{Cov}(m,n) = -1.6 \cdot 10^{-2}$
$f_2 = a\,x^2 + b\,x$	$a = (1.50 \pm 0.33) \cdot 10^{-2}$	$a = (1.45 \pm 0.19) \cdot 10^{-2}$
	$b = 1.003 \pm 0.002$	$b = 0.938 \pm 0.012$
	$\mathrm{Cov}(a,b) = -3.9 \cdot 10^{-5}$	$\mathrm{Cov}(a,b) = -1.6 \cdot 10^{-2}$
$\varepsilon_{TIS}^{MC}(\Lambda_b \to K^- \mu^+)$	$(4.05 \pm 0.07)\,\%$	$(4.86 \pm 0.10)\,\%$
$\varepsilon_{TIS}^{MC}(\Xi_b^0 \to K^- \mu^+)$	$(3.27 \pm 0.06)\,\%$	$(3.75 \pm 0.07)\,\%$

B

A strategy to measure overtraining of BDTs

B.1 Abstract

Overtraining of BDTs can spoil their performance and may lead to unforeseeable behaviour. It is shown that the usual TMVA method is problematic and not conclusive. A more meaningful strategy is presented to measure overtraining of BDTs using the TMVA package.

B.2 Preface

For the all studies that are presented the following software is used: ROOT 5.34.24 including TMVA `4.2.0`.

B.3 The TMVA method

A common check for overtraining is the comparison of the BDT response of the training sample to the test sample. After the BDT training, TMVA creates histograms of the BDT response for the training and test sample and saves the output, along with other data, in the file `TMVA.root`. The TMVA script `TMVAGui.C` takes access to `TMVA.root` and provides a function, namely `‘‘(4b) Classifier Output distributions (test and training samples superimposed)’’`, to draw the BDT histograms for the signal and background function (see Fig. B.1).
In this plot there is also the result of a Kolmogorov-Smirnov (KS) test, that compares the training and test samples for signal and background, respectively. This number is determined by the `ROOT` method `TH1::KolmogorovTest(const TH1* h2, Option_t* option = "X")`. The used histograms can be found in `TMVA.root`, e.g. for the BDT classifier it is compared

- `Method_BDT/BDT/MVA_BDT_B` to `Method_BDT/BDT/MVA_BDT_train_B`
- `Method_BDT/BDT/MVA_BDT_S` to `Method_BDT/BDT/MVA_BDT_train_S`

The corresponding source code for the KS test can be found in `$ROOTSYS/tmva/test/mvas.C`.

```
// perform K-S test
cout << "--- Perform Kolmogorov-Smirnov tests" << endl;
Double_t kolS = sig->KolmogorovTest( sigOv, "X" );
```

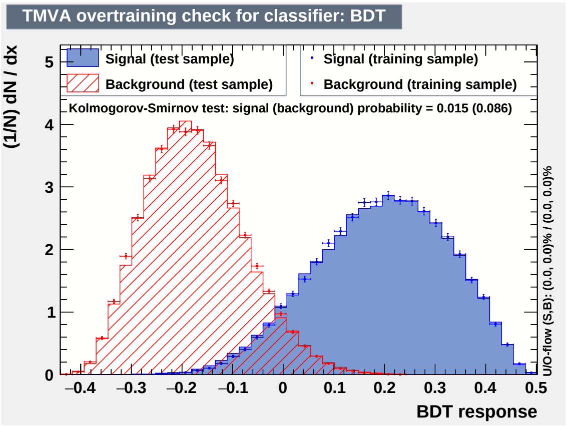


Figure B.1: BDT response on training and test samples and results from KS tests.

```
Double_t kolB = bgd->KolmogorovTest( bgdOv, "X" );
cout << "--- Goodness of signal (background) consistency: \
" << kolS << " (" << kolB << ")" << endl;

TString probatext = Form( "Kolmogorov-Smirnov test: \
signal (background) probability = %5.3g (%5.3g)", kolS, kolB );
TText* tt = new TText( 0.12, 0.74, probatext );
tt->SetNDC(); tt->SetTextSize( 0.032 ); tt->AppendPad();
```

It is necessary to mention that the option "X" is used, which means according to the TH1 class description of `ROOT` (https://root.cern.ch/root/html/TH1.html#TH1:KolmogorovTest (25.02.2015));

```
"X":
Run the pseudo experiments post-processor with the following procedure:
make pseudo experiments based on random values from the parent
distribution and compare the KS distance of the pseudo experiment
to the parent distribution. Bin the KS distances in a histogram,
and then take the integral of all the KS values above the value
obtained from the original data to Monte Carlo distribution.
The number of pseudo-experiments nEXPT is currently fixed at 1000.
The function returns the integral.
(thanks to Ben Kilminster to submit this procedure). Note that
this option "X" is much slower.
```

However, this method, that is named pseudo-KS test from now on, has some systematic problems:

1. The result depends on the seed of the random number generator of `ROOT`
2. It uses histograms, *i.e.* the result depends on the binwidth
3. htrain->KolmogorovTest(htest,"X") $\neq$ htest->KolmogorovTest(htrain,"X"), *i.e.* the method is not symmetric, even with the same seed
4. The pseudo-KS test is biased
5. Only one test is not enough to show if two distributions come from the same underlying p.d.f

The statements 1-2 are obvious and 3-5 is proved in the next sections.

B.3.1 Kolmogorov-Smirnov test in a nutshell

The KS test is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare two samples (”two-sample Kolmogorov–Smirnov test”), *i.e.* two sets $(x_i)_{i=1}^{n}$ and $(y_i)_{i=1}^{m}$. The test statistic $D_{n,m}$ is the maximum distance between the cumulative probability distributions $F_n(x)$ and $F_m(y)$ as described in Eq. B.1 and illustrated in Fig. B.2.

$$D_{n,m} = \sup |F_n(x) - F_m(y)| \ , \qquad \text{sup is the supremum function} \tag{B.1}$$

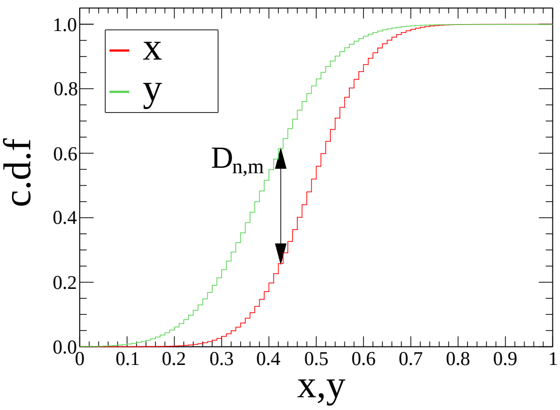


Figure B.2: Illustration of the two-sample Kolmogorov–Smirnov test statistic.

The null hypothesis that both samples have the same p.d.f. is accepted at the confidence level $1 - \alpha$ if

$$D_{n,m} < c(\alpha) \cdot \sqrt{\frac{n+m}{n \cdot m}}.$$

$1-\alpha$	90 %	95 %	99 %	99.9 %
$c(\alpha)$	1.22	1.36	1.63	1.95

B.3.2 Kolmogorov-Smirnov tests in ROOT

`ROOT` has three methods to make a KS test. As a result the confidence level for the null hypothesis is returned.

- unbinned:
  `TMath::KolmogorovTest(Int_t nx, Double_t *x, Int_t ny, Double_t *y)`
- binned:
  `TH1::KolmogorovTest(const TH1* h2, Option_t* option = "")`
- pseudo-KS:
  `TH1::KolmogorovTest(const TH1* h2, Option_t* option = "X")`

The first test uses two arrays x and y of real numbers that have to be sorted in increasing order. This test is considered to be the ”true” KS test, that calculates the correct confidence level (see `NOTE3` below). The second test uses two histograms to calculate the KS probability. Without setting an option, it takes the bin entries as samples x_i and y_i. If one sets the option `"X"` pseudo experiments are made as described above. The following quotations from the ROOT homepage explicitly recommends using the unbinned test, since the binned test in general has no expected distribution.

```
NOTE3 (Jan Conrad, Fred James)
  "The returned value PROB is calculated such that it will be
  uniformly distributed between zero and one for compatible histograms,
  provided the data are not binned (or the number of bins is very large
  compared with the number of events). Users who have access to unbinned
  data and wish exact confidence levels should therefore not put their data
  into histograms, but should call directly TMath::KolmogorovTest. On
  the other hand, since TH1 is a convenient way of collecting data and
  saving space, this function has been provided. However, the values of
  PROB for binned data will be shifted slightly higher than expected,
  depending on the effects of the binning.
```

https://root.cern.ch/root/html/TH1.html#TH1:KolmogorovTest (25.02.2015)

B.3.3 Behaviour of the KS tests from ROOT

In the following 500 random experiments are made to check the behaviour of the different KS tests. In each experiment 10000 random events of variables x and y are created, that both are uniformly distributed between 0 and 1. Figure B.3 shows one example of the distributions.

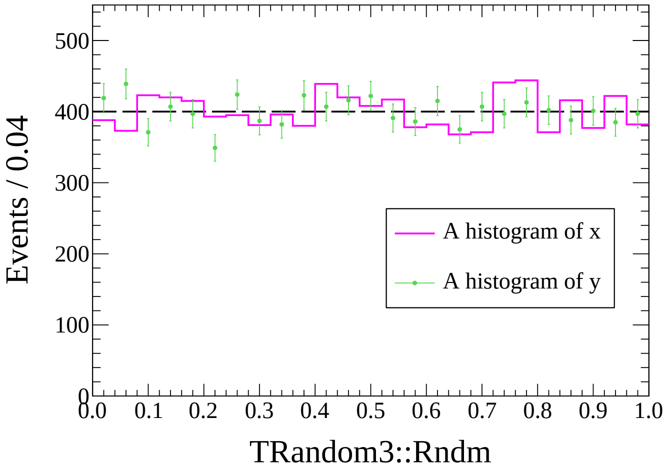


Figure B.3: A histogram of x and y.

For each of the 500 experiments all three KS tests that `ROOT` offers are made, and since the same starting seed is used for the random experiments, the tested samples are always identical. It is expected that the result of the 500 KS tests are distributed uniformly if the method is unbiased. This can be checked (1) graphically by creating a histogram of the KS tests which is then compared to a uniform distribution, and (2) by making the unbinned KS test of ROOT using the array of 500 KS tests as one sample and 500 uniformly distributed numbers that are created according to $n_i = \frac{2 \cdot i+1}{2 \cdot 500}\Big|_0^{499} = 1/1000, 3/1000, \ldots 999/1000$ as the other one.

In addition, a check is made if swapping the input samples in the method call has an effect on the result, *i.e.* it is calculated `TMath::KolmogorovTest(nx,x,ny,y)` and `TMath::KolmogorovTest(ny,y,nx,x)`. In case of the pseudo-KS tests, both methods use the same starting seed, to have equal conditions. For the binned tests, the influence of the binwidth is investigated as well.

Unbinned result

Figure B.4 shows how the 500 unbinned KS test are distributed. Swapping the input samples in the method call has no influence. In conclusion the unbinned KS test provides a unique result that is in agreement with the expected uniform distribution. The result of 93.5%, comparing the expected uniform distribution to the unbinned KS tests, is considered to be the true value for the used sample of 500 × 10 000 events.

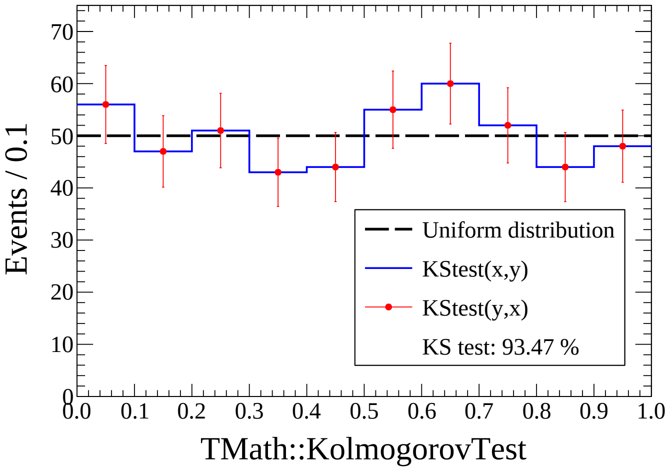
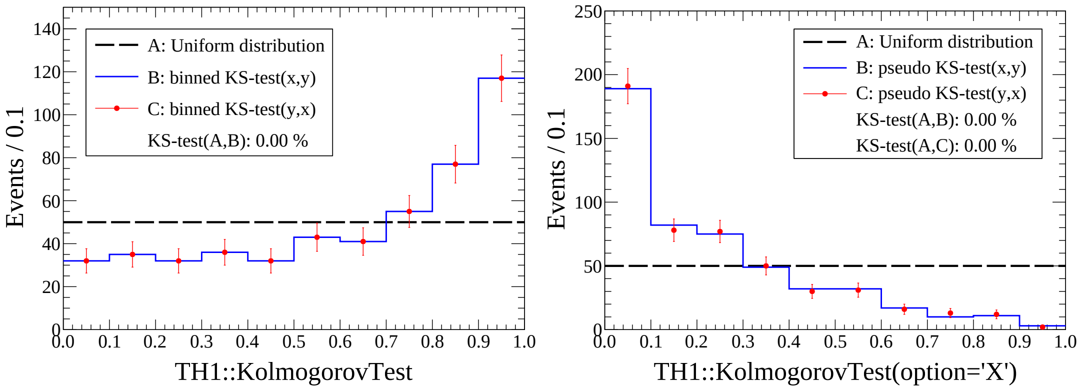


Figure B.4: Distribution of the unbinned KS test.

Large bins for x and y

The 10000 events of x and y, respectively, are put in a histograms with 25 bins, *i.e.* ≈ 400 events per bin. One of the 300 examples can be seen in Fig. B.5(a). Figure B.5(b) and (c) show how the binned KS tests are distributed. Both distributions deviate from the expected uniform distribution. The binned KS test is shifted toward large values of KS probability and the pseudo-KS test is shifted towards small values. In addition the swapping of the input samples in the method call of the pseudo-KS test leads to different results.



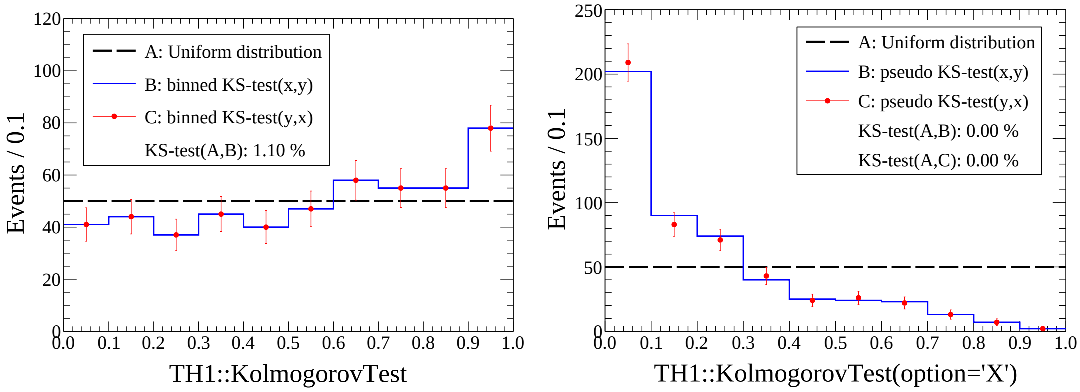
(a) Distribution of binned KS tests

(b) Distribution of pseudo-KS tests

Figure B.5: Results for binned KS tests with large bins.

Medium sized bins for x and y

Here the number of bins is 100, *i.e.* there are ≈ 100 events per bin. The resulting histograms can be seen in Fig. B.6. In conclusion, the binned KS test is improving, the distribution is in better agreement with the expected uniform distribution. The pseudo-KS test is changing only slightly.

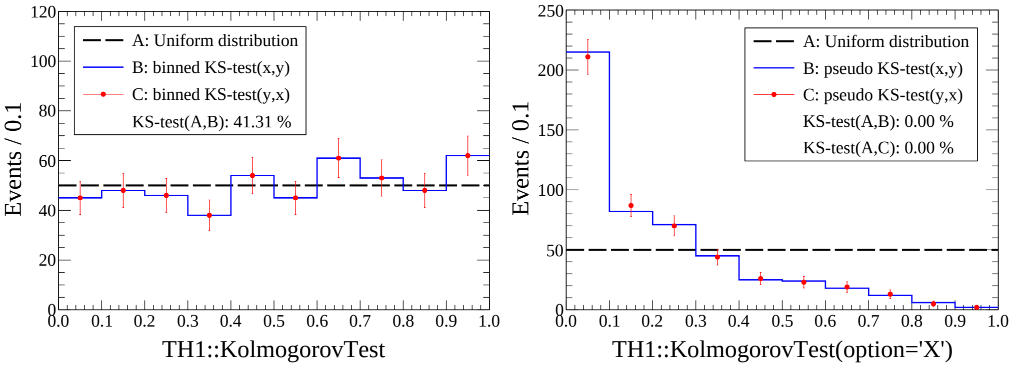


(a) Distribution of binned KS test (b) Distribution of pseudo-KS test

Figure B.6: Results for binned KS tests with medium size bins.

Small bins for x and y

Here the number of bins is 500, *i.e.* there are ≈ 20 events per bin. Figure B.7 shows the distribution of the KS tests. The distribution of the binned KS tests in, shown in Fig. B.7(a), is improving and benefits from the smaller binning, while the pseudo-KS test distribution changes only slightly and is still far away from a uniform distribution.



(a) Distribution of binned KS test (b) Distribution of pseudo-KS test

Figure B.7: Results for binned KS tests with small bins.

Conclusions

In conclusion the result of the pseudo-KS test is biased and inappropriate. The binned KS test provides only reasonable results for a very fine binning. Only the unbinned and recommended KS test is able to provide correct results.

B.4 A new strategy to check for overtraining

Based on the experience from Sec. B.3 it is advisable to not use the pseudo-KS test to check a BDT for overtraining effects. Hereby, another method is proposed that has two distinct advantages:

1. the unbinned KS test is made, providing correct confidence levels
2. the BDT is trained many times and the distribution of the KS tests is investigated

The basic idea behind point 2 is to train a specific BDT, that is defined by its parameters (Minimal NodeSize, Number of Trees, Number of Cuts, Maximum Depth) and the used input samples (signal and background), on various partitions of the input samples to achieve a statistical variation.

A partition P of a set M is a group of non-empty subsets in such a way that every element of M is included in one and only subset. In the following, a partition P always means the splitting of M into two subsamples with equal size, *i.e.* $P = \{A, B\}$ with the properties $M = A \cup B$, $|A| = |B|$ and $A \cap B = \varnothing$. M can be considered as the complete input sample and A and B are the training and test sample, respectively. Due to the fact that A and B are disjoint, they are statistically independent. For two randomly chosen partitions $P = \{A, B\}$ and $P' = \{A', B'\}$ the behaviour of a BDT trained and tested on P, BDT(P), and a BDT trained and tested on P', BDT(P'), can vary only statistically. The behaviour of each BDT is measured as the probability of a KS test (KS prob) comparing the distribution of the BDT output variable on the training sample and the test sample. If the BDT is not overtrained, the KS prob must be uniformly distributed like before where random subsamples of two identically distributed variables x and y have been compared.

The proposed overtraining check works as follows:

- Create a random partition of signal and background sample
- Train and test on this partition, obtain `TMVA.root` as output
- calculate the unbinned KS test from the `TMVA.root` file
- repeat the points above until a few hundred different partitions have been trained and tested and plot the distribution of the unbinned KS test
- if the unbinned KS test is distributed uniformly, the BDT is not overtrained

Training and testing in `TMVA` is configured in the file `TMVAClassification.C`. To have different partitions, the method `PrepareTrainingAndTestTree` can be used. Random splitting is called by `SplitMode=Random` and different partitions can be created by using different integer values of n in `SplitSeed=n`. `SplitSeed=0` means a random split seed is chosen. To get comparable results values of $1 \ldots n$ are used as splitseed. The complete command is:

```
factory->PrepareTrainingAndTestTree( mycuts, mycutb,
"nTrain_Signal=0:nTest_Signal=0:nTrain_Background=0:nTest_Background=0:
SplitMode=Random:SplitSeed=n:NormMode=NumEvents:!V" );
```

B.4.1 Statistical properties of the partitions

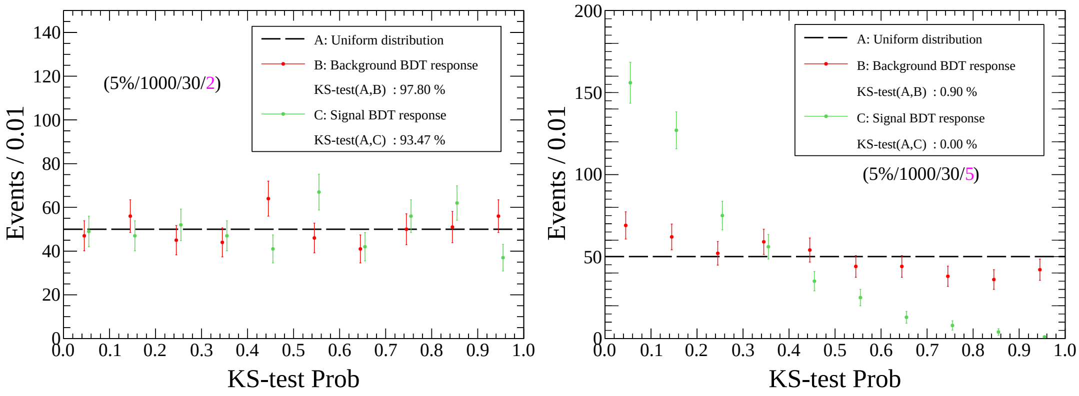
If there are $2N$ events in a sample, there are $\binom{2N}{N}$ possible partitions. For example if $2N = 1000$, there are $\approx 10^{300}$ different partitions. On average, two randomly chosen partitions differ by $N/2$ elements. The corresponding calculations are shown in Appendix B.6.

B.4.2 An example from TMVA

The different behaviour of the KS test methods of `ROOT` is demonstrated using a typical user case. In particular it is checked how the unbinned and the pseudo-KS test differ for an overtrained and a not overtrained BDT. The signal sample is $\Lambda_b \to K^- \mu^+$ signal MC and the background sample are WS data. Training and testing is done on 500 different partitions. The sample of simulated events has a size of ca. 60 000 events and the background sample has a size of ca. 350 000 events. The overtrained and the not overtrained BDT use the same set of partitions, *i.e.* the same seeds are used. An overtraining can be achieved by increasing the parameter `MaxDepth` from 2 to 5. For each of the 500 partitions the unbinned and the pseudo-KS test is calculated.

Figures B.8(a) and (b) compare the result of the unbinned KS test for both parameters. Here and in all other histograms the number in brackets denote (minimal node size / no. of trees / no of cuts / maximum depth). The distribution of the KS test is uniform for `MaxDept=2` in both samples. Hence, this BDT is considered as to be not overtrained. In contrast it shows that the distributions of the KS test of the BDT that uses `MaxDept=5` is clearly deviating from the expected uniform distribution. This BDT is considered as overtrained. In addition it is visible that the overtraining effect is more significant in the smaller sample of signal events. This is comprehensible if one considers overtraining as an effect that occurs when the number of degrees of freedom is small and the BDT trains on statistical fluctuations.

Figure B.9 shows the linear correlation between the unbinned KS test of signal and background. There is no significant correlation which can be explained by the fact the sampling of the signal and background partitions happens independently from each other.

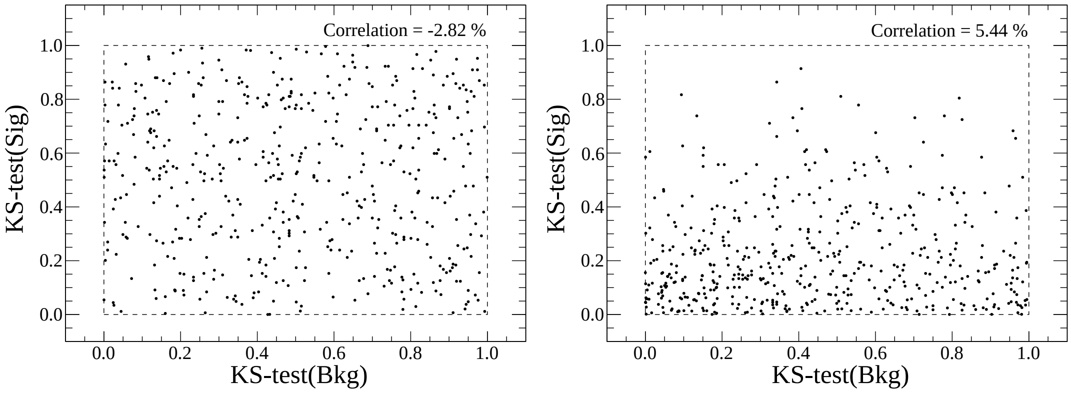


(a) Not overtrained BDT (b) Overtrained BDT

Figure B.8: KS test for two different BDTs.

Figure B.10 show the result of the pseudo-KS test from the same 500 partitions that were used before. In every case the KS probability is calculated like `TMVAGui.C` does it. From the distributions it can be seen that both of the pseudo-KS tests are not able to distinguish the overtrained from the not-overtrained BDT.

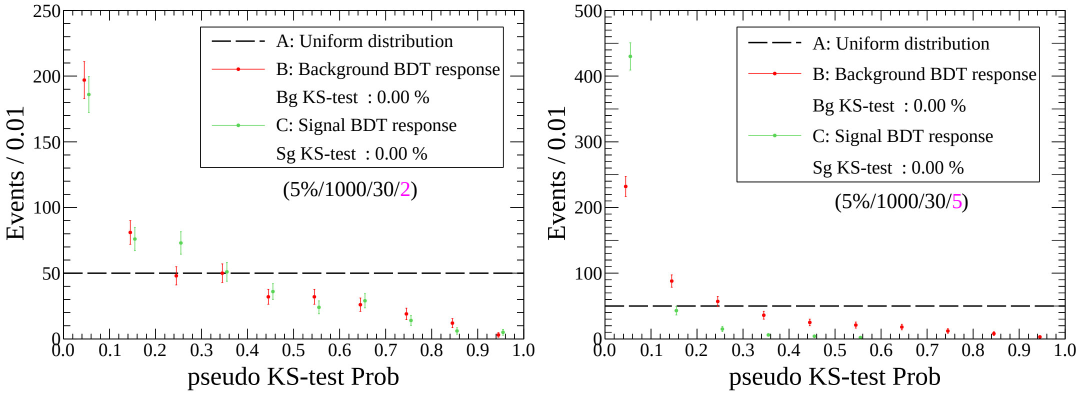
Usually one does only a single test on one partition and accepts a BDT as not overtrained when both KS tests (signal and background) are in the range of 10% . . . 90%. In this case the not overtrained BDT is accepted with a probability of about $\left(\frac{500-200}{500}\right)_{Bg} \times \left(\frac{500-190}{500}\right)_{Sg} = 37.2\,\%$, where 200 and 190 are the number of cases in which the KS test of signal and background is outside of the acceptance range, respectively. The overtrained BDT is accepted with a probability of about $\left(\frac{500-430}{500}\right)_{Bg} \times \left(\frac{500-240}{500}\right)_{Sg} = 7.3\,\%$. This can be interpreted such that the pseudo-KS test does not prefer an overtrained BDT over a not overtrained one, but is too conservative since it is more likely ($> 50\%$) to reject a good BDT. For comparison, using the unbinned KS test from Fig. B.10 one would accept the not-overtrained BDT with a probability of ca. $0.8 \cdot 0.8 = 64\%$ and the overtrained one with a probability of about $\left(\frac{500-110}{500}\right)_{Bg} \times \left(\frac{500-160}{500}\right)_{Sg} = 53.0\,\%$. In summary,



(a) not-overtrained BDT (b) overtrained BDT

Figure B.9: Correlation of the signal vs. background KS tests for the two BDTs.

whatever method one uses to calculate the KS test, the result of only one partition is not conclusive and the pseudo-KS test is not able to distinguish an overtrained from a not overtrained BDT.



(a) not overtrained BDT (b) overtrained BDT

Figure B.10: Pseudo-KS test for two different BDTs.

B.4.3 Choosing a BDT

In case the unbinned KS test shows that a BDT is not overtrained, a strategy to choose a particular BDT classification can be to require that the KS tests for signal and background are both around 50 % (as illustrated in Fig B.11). This can be interpreted such that this is an "average" BDT classifier, *i.e.* about 50 % have a higher KS probability and about 50 % have a lower KS probability.

B.4.4 An alternative check

An alternative check for overtraining is the comparison of the BDT performance on the training and test sample, quantified by the integral over the ROC curve. If the BDT is not overtrained this number should be similarly distributed due to the fact that the BDT is not adapted to the features of the training sample. Figure B.12 shows the distribution of the ROC integrals. Both distributions are compatible in case of the not overtrained BDT while there is a visible deviation for the overtrained BDT.

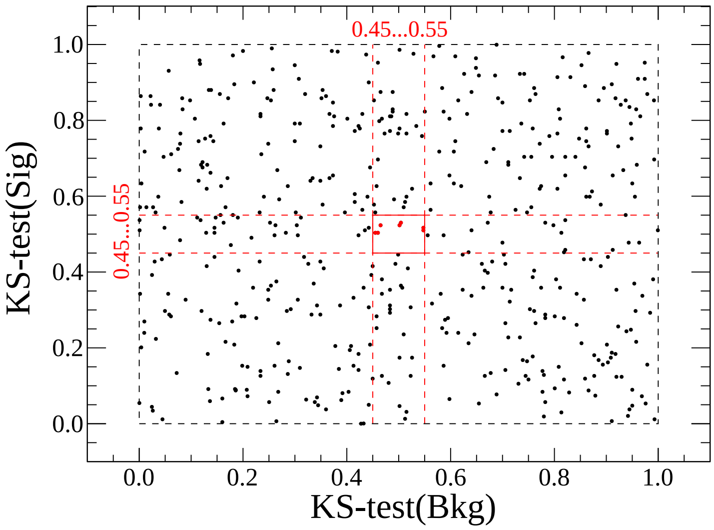
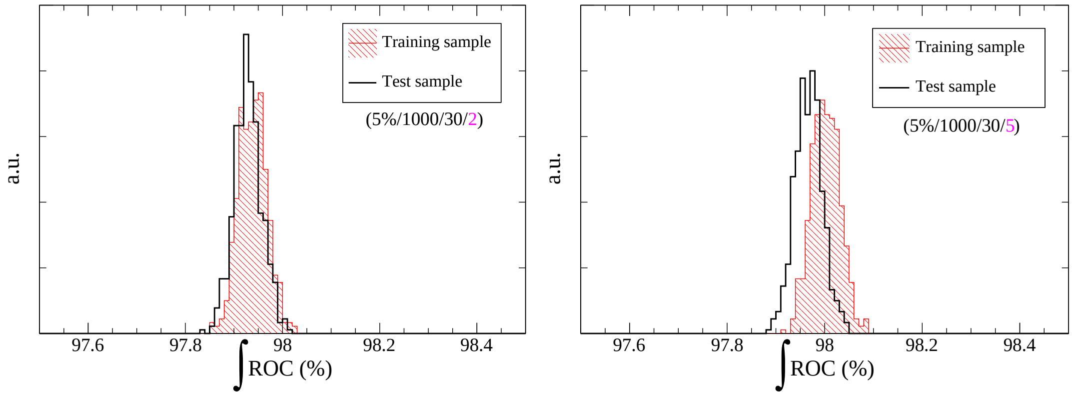


Figure B.11: Selection range for a particular BDT.



(a) not overtrained BDT

(b) overtrained BDT

Figure B.12: Comparison of ROC integrals for two different BDTs.

B.5 Conclusions

It has been shown that the present KS test of TMVA (pseudo-KS test) is biased and that a single result is not able to judge if a BDT is overtrained or not. A solution has been presented, that firstly uses unbinned data to provide exact and unbiased confidence levels and secondly requires to test the BDT on several random partitions of the input samples to check the distribution of the unbinned KS test. In addition it has been shown that the overtraining can result from too small input samples. The fact that an unbiased BDT is not able to distinguish between training and test samples leads to the useful practical consequence that one can use the entire input sample for training. This helps to reduce the influence of statistical fluctuations and possibly improves the BDT performance due to the higher number of events.

B.6 Appendix: statistical properties of partitions

If (A,B) is a random partition of $2N$ elements, a new partition (A',B') can be created by swapping $k = 0 \dots N$ pairs of elements between A and B. There are $\binom{N}{k}$ ways to select k elements from partition A and B, respectively. Hence there are $\binom{N}{k} \cdot \binom{N}{k}$ partitions $(A', B')_{(k)}$ where k swaps have been made. The sum of all possible k-partitions is $\sum_k (A', B')_{(k)} = \sum_k \binom{N}{k} \cdot \binom{N}{k} = \binom{2N}{N}$ according to the Che-Vandermont identity. This is identical with the number of possible partitions from simple combinatorics: there are $\binom{2N}{N}$ ways to select N events out of $2N$.

The probability that a randomly chosen partition (A',B') differs by k elements to a given partition (A,B) is the number of possible partitions divided by the number of all partitions assuming that all possible realisations of (A',B') have the same probability to be chosen (Laplace probability model).

$$p(k) = \frac{\binom{N}{k} \cdot \binom{N}{k}}{\binom{2N}{N}}$$

The expectation value of k is:

$$\langle k \rangle = \sum_{k=0}^{N} k \cdot p(k) = \frac{1}{\binom{2N}{N}} \cdot \sum_{k=0}^{N} k \cdot \binom{N}{k}^2 = \frac{1}{\binom{2N}{N}} \cdot \frac{N}{2} \cdot \binom{2N}{N} = \frac{N}{2}$$

where it is used $\binom{N}{k} = \binom{N}{N-k}$ and

$$\sum_{k=0}^{N} k \cdot \binom{N}{k}^2 = \left[(0+N) \cdot \binom{N}{0}^2 \right] + \left[(1+N-1) \cdot \binom{N}{1}^2 \right] + \dots \tag{B.2}$$

$$= N \cdot \sum_{k=0}^{N/2} \binom{N}{k}^2 = N \cdot \frac{1}{2} \sum_{k=0}^{N} \binom{N}{k}^2 = N \cdot \frac{1}{2} \binom{2N}{N} \tag{B.3}$$

An alternative approach is the following; starting from a given partition (A,B), a new partition (A',B') can be achieved by randomly filling the elements of A into the sets A' and B' and after that filling the elements from B into A' and B'. Since A' and B' shall have equal sizes, the probability that a certain element from A falls into A' or B' is $1/2$, respectively. The probability that k of N elements from A fall into A' is described by the binomial distribution,

$$p(k|N, p = 1/2) = \binom{N}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{N-k} = \left(\frac{1}{2}\right)^N \cdot \binom{N}{k}$$

This means the expected number of events from A falling into A' is $\langle k \rangle = N \cdot p = N/2$, following from the properties of the binomial distribution and in agreement with the result above. The variance is $\sigma(k)^2 = N \cdot p \cdot (1-p) = N/4$ and hence the standard deviation of k is $\sqrt{N}/2$.

C

List of used acronyms

Collaborations and experiments

- ALICE: A Large Ion Collider Experiment
- ATLAS: A Toroidal LHC ApparatuS
- CDF: Collider Detector at Fermilab
- CERN: Conseil Européen pour la Recherche Nucléaire
- CMS: Compact Muon Solenoid
- LHC: Large Hadron Collider
- LHCb: Large Hadron Collider beauty

LHCb experiment

- BDT: Boosted decision tree
- BPV: Best primary vertex
- ECAL: Electromagnetic calorimeter
- EFF: Event filter farm
- GEC: Global event cut
- GEM: Gas electron multiplier
- GPD: General purpose detector
- DAQ: Data aquisition
- DIRA: Direction angle
- DOCA: Distance of closest approach
- DOF/dof: Degrees of freedom (fit)
- FD: Flight distance

- FoM: Figure of merit
- HCAL: Hadronic calorimeter
- HLT: High-level trigger
- HPD: Hybrid photon detector
- IP: Impact parameter
- IT: Inner Tracker
- L0: Level-0 trigger
- MC: Monte-Carlo (simulated events)
- MWPC: Multi-wire plate chamber
- OT: Outer Tracker
- PID: Particle identification
- PMT: Photo multiplier tubes
- PS: Preshower detector
- PV/SV: Primary/Secondary vertex
- RICH: Ring imaging Cherenkov detector
- RS: Right-sign (combinations of $K^-\mu^+$ + c.c.)
- ROC: Receiver operating characteristic
- SP: Seperating power
- SPD: Scintillating pad detector
- ST: Silicon Tracker
- TCK: Trigger configuration key
- TIS: Trigger independent of signal
- TOS: Trigger on signal
- TT: Tracker Turicencis
- VELO: Vertex Locator
- WS: Wrong-sign (combinations of $K^+\mu^+$ + c.c.)

Physics

- BAO: Baryonic accoustic oscillations
- BBN: Big-Bang nucleosynthesis
- BLNV: Baryon and lepton number violation
- CMB: Cosmic microwave background
- EM: Electromagnetic
- FCNC: Flavour-changing neutral currents

- GUT: Grand Unified Theory
- LQ: Leptoquark
- SM: Standard Model

Miscellaneous

- BMBF: Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research)
- RPP: The Review of Particle Physics

Bibliography

[1] C. J. Conselice, A. Wilkinson, K. Duncan, and A. Mortlock, "The Evolution of Galaxy Number Density at $z < 8$ and Its Implications," *Astrophys. J.* **830** (2016) no. 2, 83, `arXiv:1607.03909 [astro-ph.GA]`.

[2] A. Sakharov, "Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe," *Pisma Zh.Eksp.Teor.Fiz.* **5** (1967) 32–35.

[3] Patrignani, C. *et al.* (Particle Data Group), "Review of Particle Physics," *Chin. Phys.* **C40** (2016) no. 10, 100001.

[4] A. Liddle, *An introduction to modern cosmology; 2nd ed.* Wiley, Chichester, 2003. `https://cds.cern.ch/record/1010476`.

[5] R. H. Cyburt, B. D. Fields, K. A. Olive, and T.-H. Yeh, "Big Bang Nucleosynthesis: 2015," *Rev. Mod. Phys.* **88** (2016) 015004, `arXiv:1505.01076 [astro-ph.CO]`.

[6] R. Cooke, M. Pettini, R. A. Jorgenson, M. T. Murphy, and C. C. Steidel, "Precision measures of the primordial abundance of deuterium," *Astrophys. J.* **781** (2014) no. 1, 31, `arXiv:1308.3240 [astro-ph.CO]`.

[7] **Planck** Collaboration, "Planck science team home." `http://www.cosmos.esa.int/web/planck`. [Online; accessed 17-November-2016].

[8] W. Hu and S. Dodelson, "Cosmic microwave background anisotropies," *Ann. Rev. Astron. Astrophys.* **40** (2002) 171–216, `arXiv:astro-ph/0110414 [astro-ph]`.

[9] **Planck** Collaboration, P. A. R. Ade *et al.*, "Planck 2015 results. XIII. Cosmological parameters," *Astron. Astrophys.* **594** (2016) A13, `arXiv:1502.01589 [astro-ph.CO]`.

[10] F. Karsch, E. Laermann, and A. Peikert, "Quark mass and flavor dependence of the QCD phase transition," *Nucl. Phys.* **B605** (2001) 579–599, `arXiv:hep-lat/0012023 [hep-lat]`.

[11] C. Caprini, S. Biller, and P. G. Ferreira, "Constraints on the electrical charge asymmetry of the universe," *JCAP* **0502** (2005) 006, `arXiv:hep-ph/0310066 [hep-ph]`.

[12] W.-S. Hou, M. Nagashima, and A. Soddu, "Baryon number violation involving higher generations," *Phys. Rev.* **D72** (2005) 095001, `arXiv:hep-ph/0509006 [hep-ph]`.

[13] H. Georgi and S. L. Glashow, "Unity of All Elementary Particle Forces," *Phys. Rev. Lett.* **32** (1974) 438–441.

[14] J. C. Pati and A. Salam, "Lepton Number as the Fourth Color," *Phys. Rev.* **D10** (1974) 275–289. [Erratum: Phys. Rev.D11,703(1975)].

[15] **ATLAS** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett.* **B716** (2012) 1–29, `arXiv:1207.7214 [hep-ex]`.

[16] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett.* **B716** (2012) 30–61, `arXiv:1207.7235 [hep-ex]`.

[17] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ with the CMS Experiment," *Phys. Rev. Lett.* **111** (2013) 101804, `arXiv:1307.5025 [hep-ex]`.

[18] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment," *Phys. Rev. Lett.* **111** (2013) 101805, `arXiv:1307.5024 [hep-ex]`.

[19] **LHCb** Collaboration, R. Aaij *et al.*, "Observation of two new Ξ_b^- baryon resonances," *Phys. Rev. Lett.* **114** (2015) 062004, `arXiv:1411.4849 [hep-ex]`.

[20] **LHCb** Collaboration, R. Aaij *et al.*, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays," *Phys. Rev. Lett.* **115** (2015) 072001, `arXiv:1507.03414 [hep-ex]`.

[21] **Welt der Physik**, "Reparatur und Umbau des LHC." `http://www.weltderphysik.de/gebiet/teilchen/experimente/teilchenbeschleuniger/lhc/reparatur-und-umbau-des-lhc/`. [Online; accessed 24-August-2016].

[22] **LHCb** Collaboration, A. A. Alves, Jr. *et al.*, "The LHCb Detector at the LHC," *JINST* **3** (2008) S08005.

[23] **LHCb** Collaboration, R. Aaij *et al.*, "LHCb Detector Performance," *Int. J. Mod. Phys.* **A30** (2015) no. 07, 1530022, `arXiv:1412.6352 [hep-ex]`.

[24] **LHCb** Collaboration, "Speaker's bureau, material for presentations." `http://lhcb.web.cern.ch/lhcb/speakersbureau/html/Material_for_Presentations.html`. [Online; accessed 23-August-2016].

[25] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region," *Phys. Lett.* **B694** (2010) 209–216, `arXiv:1009.2731 [hep-ex]`.

[26] **LHCb** Collaboration, R. Aaij *et al.*, "Precision luminosity measurements at LHCb," *JINST* **9** (2014) no. 12, P12005, `arXiv:1410.0149 [hep-ex]`.

[27] **LHCb** Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb VELO (VErtex LOcator): Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2001. `https://cds.cern.ch/record/504321`.

[28] **LHCb** Collaboration, R. Aaij *et al.*, "Performance of the LHCb Vertex Locator," *JINST* **9** (2014) 09007, `arXiv:1405.7808 [physics.ins-det]`.

[29] **LHCb** Collaboration, S. Amato *et al.*, *LHCb magnet: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2000. `http://cds.cern.ch/record/424338`.

[30] **LHCb** Collaboration, R. Arink *et al.*, "Performance of the LHCb Outer Tracker," *JINST* **9** (2014) no. 01, P01002, `arXiv:1311.3893 [physics.ins-det]`.

[31] **LHCb** Collaboration, "LHCb public page, Tracking system." `https://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Trackers2-en.html`. [Online; accessed 24-August-2016].

[32] M. Needham and J. Van Tilburg, "Performance of the track matching," Tech. Rep. LHCb-2007-020. CERN-LHCb-2007-020, CERN, Geneva, Mar, 2007. `https://cds.cern.ch/record/1020304`.

[33] M. Needham, "Performance of the Track Matching," Tech. Rep. LHCb-2007-129. CERN-LHCb-2007-129, CERN, Geneva, Oct, 2007. `https://cds.cern.ch/record/1060807`.

[34] O. Callot and M. Schiller, "PatSeeding: A Standalone Track Reconstruction Algorithm," Tech. Rep. LHCb-2008-042. CERN-LHCb-2008-042, CERN, Geneva, Aug, 2008. `https://cds.cern.ch/record/1119095`.

[35] J. A. N. van Tilburg, *Track Simulation and Reconstruction in LHCb*. PhD thesis, Vrije U., Amsterdam, 2005. `http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.3.T67T35::2005`.

[36] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of prompt hadron production ratios in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV," *Eur. Phys. J.* **C72** (2012) 2168, `arXiv:1206.5160 [hep-ex]`.

[37] **LHCb** Collaboration, M. Adinolfi *et al.*, "Performance of the LHCb RICH detector at the LHC," *Eur. Phys. J.* **C73** (2013) 2431, `arXiv:1211.6759 [physics.ins-det]`.

[38] **LHCb** Collaboration, P. Perret and X. Vilasís-Cardona, "Performance of the LHCb calorimeters during the period 2010-2012," *J. Phys. Conf. Ser.* **587** (2015) no. 1, 012012.

[39] **LHCb** Collaboration, A. A. Alves, Jr. *et al.*, "Performance of the LHCb muon system," *JINST* **8** (2013) P02022, `arXiv:1211.1346 [physics.ins-det]`.

[40] **LHCb** Collaboration, F. Archilli *et al.*, "Performance of the Muon Identification at LHCb," *JINST* **8** (2013) P10020, `arXiv:1306.0249 [physics.ins-det]`.

[41] A. Puig, "The LHCb trigger in 2011 and 2012," Tech. Rep. LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov, 2014. `https://cds.cern.ch/record/1970930`.

[42] **LHCb** Collaboration, R. Aaij *et al.*, "The LHCb Trigger and its Performance in 2011," *JINST* **8** (2013) P04022, `arXiv:1211.3055 [hep-ex]`.

[43] E. Norrbin and T. Sjostrand, "Production and hadronization of heavy quarks," *Eur. Phys. J.* **C17** (2000) 137–161, `arXiv:hep-ph/0005110 [hep-ph]`.

[44] **LHCb** Collaboration, R. Aaij *et al.*, "Study of the kinematic dependences of Λ_b^0 production in pp collisions and a measurement of the $\Lambda_b^0 \to \Lambda_c^+ \pi^-$ branching fraction," *JHEP* **08** (2014) 143, `arXiv:1405.6842 [hep-ex]`.

[45] **LHCb** Collaboration, R. Aaij *et al.*, "Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s} = 7$ TeV," *JHEP* **08** (2013) 117, `arXiv:1306.3663`.

[46] **LHCb** Collaboration, R. Aaij *et al.*, "Study of the production of Λ_b^0 and $\overline{B}^0$ hadrons in pp collisions and first measurement of the $\Lambda_b^0 \to J/\psi pK^-$ branching fraction," *Chin. Phys.* **C40** (2016) no. 1, 011001, `arXiv:1509.00292 [hep-ex]`.

[47] J. L. Rosner, "Asymmetry in Λ_b and $\bar{\Lambda}_b$ production," *Phys. Rev.* **D90** (2014) no. 1, 014023, `arXiv:1405.2885 [hep-ph]`.

[48] E. Norrbin and R. Vogt, "Bottom production asymmetries at the LHC," in *Fifth Workshop on electronics for LHC experiments, Snowmass, CO, USA, 20-24 Sep 1999: Proceedings*. 2000. `arXiv:hep-ph/0003056 [hep-ph]`.

[49] **LHCb** Collaboration, R. Aaij *et al.*, "Measurements of the $\Lambda_b \to J/\psi \Lambda$ decay amplitudes and the Λ_b polarisation in pp collisions at $\sqrt{s} = 7\,\text{TeV}$," *Phys. Lett.* **B724** (2013) 27–35, `arXiv:1302.5578 [hep-ex]`.

[50] M. Needham, "Identification of Ghost Tracks using a Likelihood Method," Tech. Rep. LHCb-2008-026. CERN-LHCb-2008-026. LPHE-2008-004, CERN, Geneva, May, 2008. `http://cds.cern.ch/record/1107564`.

[51] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP* **05** (2006) 026, `arXiv:hep-ph/0603175 [hep-ph]`.

[52] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.* **178** (2008) 852–867, `arXiv:0710.3820 [hep-ph]`.

[53] I. Belyaev *et al.*, “Handling of the generation of primary events in gauss, the lhcb simulation framework,” *Journal of Physics: Conference Series* **331** (2011) no. 3, 032047. `http://stacks.iop.org/1742-6596/331/i=3/a=032047`.

[54] D. J. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[55] P. Golonka and Z. Was, “PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays,” *Eur. Phys. J.* **C45** (2006) 97–107, `arXiv:hep-ph/0506026 [hep-ph]`.

[56] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[57] **GEANT4** Collaboration, S. Agostinelli *et al.*, “GEANT4: A Simulation toolkit,” *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[58] **LHCb** Collaboration, M. Clemencic *et al.*, “The LHCb Simulation Application, Gauss: Design, Evolution and Experience,” *Journal of Physics: Conference Series* **331** (2011) no. 3, 032023. `http://stacks.iop.org/1742-6596/331/i=3/a=032023`.

[59] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, “Data driven trigger efficiency determination at LHCb,” Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014. `https://cds.cern.ch/record/1701134`.

[60] V. V. Gligorov, C. Thomas, and M. Williams, “The HLT inclusive B triggers,” Tech. Rep. LHCb-PUB-2011-016. CERN-LHCb-PUB-2011-016. LHCb-INT-2011-030, CERN, Geneva, Sep, 2011. `http://cds.cern.ch/record/1384380`.

[61] A. Khodjamirian, C. Klein, T. Mannel, and Y. M. Wang, “Form Factors and Strong Couplings of Heavy Baryons from QCD Light-Cone Sum Rules,” *JHEP* **09** (2011) 106, `arXiv:1108.2971 [hep-ph]`.

[62] M. Pivk and F. R. Le Diberder, “SPlot: A Statistical tool to unfold data distributions,” *Nucl. Instrum. Meth.* **A555** (2005) 356–369, `arXiv:physics/0402083 [physics.data-an]`.

[63] **LHCb** Collaboration, R. Aaij *et al.*, “Observation of the $\Lambda_b^0 \to J/\psi p \pi^-$ decay,” *JHEP* **07** (2014) 103, `arXiv:1406.0755 [hep-ex]`.

[64] A. Hocker *et al.*, “TMVA - Toolkit for Multivariate Data Analysis,” *PoS* **ACAT** (2007) 040, `arXiv:physics/0703039 [PHYSICS]`.

[65] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[66] R. Waldi, *Statistische Datenanalyse - Grundlagen und Methoden für Physiker*. Springer-Verlag, Berlin Heidelberg New York, 2015.

[67] **LHCb** Collaboration, A. Poluektov, “L0hadron efficiency correction for protons.” `https://twiki.cern.ch/twiki/bin/view/LHCb/LbLcPiL0HadronCorrection`. [Online; accessed 25-January-2016].

[68] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of the track reconstruction efficiency at LHCb,” *JINST* **10** (2015) no. 02, P02007, `arXiv:1408.1251 [hep-ex]`.

[69] **CDF** Collaboration, T. Aaltonen *et al.*, “Observation of New Charmless Decays of Bottom Hadrons,” *Phys. Rev. Lett.* **103** (2009) 031801, `arXiv:0812.4271 [hep-ex]`.

[70] **OPAL** Collaboration, G. Abbiendi *et al.*, “Search for baryon and lepton number violating Z0 decays,” *Phys. Lett.* **B447** (1999) 157–166, `arXiv:hep-ex/9901011 [hep-ex]`.

[71] **CLEO** Collaboration, R. Godang *et al.*, “Search for baryon and lepton number violating decays of the tau lepton,” *Phys. Rev.* **D59** (1999) 091303, `arXiv:hep-ex/9902005 [hep-ex]`.

[72] **Belle** Collaboration, Y. Miyazaki *et al.*, “Search for lepton and baryon number violating τ^- decays into $\overline{\Lambda}\pi^-$ and $\Lambda\pi^-$,” *Phys. Lett.* **B632** (2006) 51–57, `arXiv:hep-ex/0508044 [hep-ex]`.

[73] **Super-Kamiokande** Collaboration, H. Nishino *et al.*, “Search for Proton Decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ in a Large Water Cherenkov Detector,” *Phys. Rev. Lett.* **102** (2009) 141801, `arXiv:0903.0676 [hep-ex]`.

[74] **CLEO** Collaboration, P. Rubin *et al.*, “Search for $D^0 \to \overline{p}e^+$ and $D^0 \to pe^-$,” *Phys. Rev.* **D79** (2009) 097101, `arXiv:0904.1619 [hep-ex]`.

[75] **BaBar** Collaboration, P. del Amo Sanchez *et al.*, “Searches for the baryon- and lepton-number violating decays $B^0 \to \Lambda_c^+\ell^-$, $B^- \to \Lambda\ell^-$, and $B^- \to \overline{\Lambda}\ell^-$,” *Phys. Rev.* **D83** (2011) 091101, `arXiv:1101.3830 [hep-ex]`.

[76] **LHCb** Collaboration, R. Aaij *et al.*, “Searches for violation of lepton flavour and baryon number in tau lepton decays at LHCb,” *Phys. Lett.* **B724** (2013) 36–45, `arXiv:1304.4518 [hep-ex]`.

[77] **CMS** Collaboration, S. Chatrchyan *et al.*, “Search for baryon number violation in top-quark decays,” *Phys. Lett.* **B731** (2014) 173–196, `arXiv:1310.1618 [hep-ex]`.

[78] J. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'*. PhD thesis, SLAC, 1982. `http://inspirehep.net/record/183554`.

[79] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of B_c^+ production in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *Phys. Rev. Lett.* **114** (2015) 132001, `arXiv:1411.2943 [hep-ex]`.

[80] C.-F. Qiao, P. Sun, D. Yang, and R.-L. Zhu, “B_c^+ exclusive decays to charmonium and a light meson at next-to-leading order accuracy,” *Phys. Rev.* **D89** (2014) no. 3, 034008, `arXiv:1209.5859 [hep-ph]`.

[81] **LHCb** Collaboration, R. Aaij *et al.*, “Precision measurement of the mass and lifetime of the Ξ_b^0 baryon,” *Phys. Rev. Lett.* **113** (2014) 032001, `arXiv:1405.7223 [hep-ex]`.

[82] **LHCb** Collaboration, R. Aaij *et al.*, “Precision measurement of the ratio of the Λ_b^0 to $\overline{B}^0$ lifetimes,” *Phys. Lett.* **B734** (2014) 122–130, `arXiv:1402.6242 [hep-ex]`.

[83] **LHCb** Collaboration, R. Aaij *et al.*, “Measurements of the B^+, B^0, B_s^0 meson and Λ_b^0 baryon lifetimes,” *JHEP* **04** (2014) 114, `arXiv:1402.2554 [hep-ex]`.

[84] **LHCb** Collaboration, R. Aaij *et al.*, “Studies of beauty baryon decays to D^0ph^- and $\Lambda_c h^-$ final states,” *Phys. Rev.* **D89** (2014) no. 3, 032001, `arXiv:1311.4823 [hep-ex]`.

[85] **LHCb** Collaboration, R. Aaij *et al.*, “Precision measurement of the Λ_b baryon lifetime,” *Phys. Rev. Lett.* **111** (2013) 102003, `arXiv:1307.2476 [hep-ex]`.

[86] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of the Λ_b, Ξ_b^- and Ω_b^- baryon masses,” *Phys. Rev. Lett.* **110** (2013) no. 18, 182001, `arXiv:1302.1072 [hep-ex]`.

[87] **LHCb** Collaboration, R. Aaij *et al.*, “Measurement of b-hadron branching fractions for two-body decays into charmless charged hadrons,” *JHEP* **10** (2012) 037, `arXiv:1206.2794 [hep-ex]`.

Danksagung

Tja, wer hätte gedacht, dass ein kleiner Junge aus dem wilden Rostocker Stadtteil Dierkow mal an einer deutschen Universität eine Dissertation über Teilchenphysik einreicht? Für das Gelingen dieser Arbeit möchte ich mich vor allem bei meinem Chef Roland Waldi bedanken, der mich seit meiner Diplomzeit immer hervorragend betreut hat und dessen Scharfsinn und Wissensschatz ich immer sehr bewundert habe. Desweiteren möchte ich mich bedanken bei dem leider viel zu früh verstorbenen Prof. Henning Schröder, der mich und uns in Rostock für die Suche nach Baryon- und Leptonzahlverletzung inspiriert hat. Großen Dank gebührt ebenso allen Mitarbeitern der LHCb Kollaboration, die die experimentelle Grundlage für die erfolgreiche Datenanalyse gelegt haben. Insbesondere sei hier Johannes Albrecht erwähnt, der mir bei den ersten Schritten geholfen hat, Vanya Belyaev, der immer schnell auf meine Fragen antwortet, sowie meine liebe Bürokollegin Miriam Heß, die regelmäßig ein paar gute Tipps für mich hatte um im Dschungel der LHCb Software klar zu kommen.

In dem Paralleluniversum außerhalb der Teilchenphysik möchte ich mich bei meinen Eltern und Freunden bedanken, die mich während der letzten Jahre unterstützt und begleitet haben. Insbesondere erwähnt seien Christian (Chrischi), Matthias (Matze), Peter (Peti), der rote Richard und meine geliebte Flatmate Marieke, die sich hiermit alle ganz doll gedrückt fühlen dürfen. Und so schließe ich mit einem Liedzitat von Klaus Lage das mir gerade in den Sinn kommt:

“Egal wohin die Zeit uns treibt
die Liebe bleibt.”