UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Scienze e Tecnologie

Corso di Laurea Magistrale in Fisica

Tesi di Laurea Magistrale

Search for CP violation in $\Lambda_b \to pK^-\mu^+\mu^-$ decays

Candidato:

Daniele Marangotto

N$^{\text{o}}$ matr. 842629

Codice PACS: 13.30.Ce

Relatore Interno:

Prof. Fernando Palombo

Relatore Esterno:

Dott. Nicola Neri

Anno Accademico 2014-2015

Abstract

In this thesis the first measurement of CP-violation effects ever performed on rare heavy beauty baryon $\Lambda_b \to pK^-\mu^+\mu^-$ decay is described. This analysis is carried out on the whole dataset recorded by the LHCb experiment during 2011 and 2012.

The CP symmetry violation study is one of the most promising method for searching physics beyond the standard model, the most accurate theory describing interactions of elementary particles to date. Indeed, the measured amount of CP-violation in high-energy physics experiments, even though compatible with standard model expectations, is not sufficient for explaining the observed matter-antimatter asymmetry of our universe. Hence, new sources of CP-violation could come from physics beyond standard model.

The $\Lambda_b \to pK^-\mu^+\mu^-$ decays occur via electroweak loop diagrams which, on the one hand make this process very rare in the standard model, on the other hand allow possible new physics fields to give this process a sensible contribution. For the $\Lambda_b \to pK^-\mu^+\mu^-$ decay a limited quantity of CP-violation is expected in the standard model, therefore a significatively large CP-violation observation in this analysis would be a sign of beyond standard model physics. In this thesis, CP-violation is searched exploiting direct CP asymmetries and triple-products based on T-odd correlations, two complementary measurements sensible to different new physics contributions.

Experimentally, the study of $\Lambda_b \to pK^-\mu^+\mu^-$ transition is very clean thanks to the high performance of the muon detection and particle identification systems of LHCb experiment, this enabling to discard most of the background events. The $\Lambda_b \to pK^-J/\psi$ decays are used as control channel, since they have a far higher yield as well as the same final particles.

The measured asymmetries

$$\Delta\mathcal{A}_{CP} = (-4.8 \pm 5.0\ (\text{stat}) \pm 0.3\ (\text{syst})) \times 10^{-2}$$

$$A_T = (-2.8 \pm 7.2\ (\text{stat}) \pm 1.0\ (\text{syst})) \times 10^{-2}$$

$$\overline{A}_T = (-4.0 \pm 6.9\ (\text{stat}) \pm 1.0\ (\text{syst})) \times 10^{-2}$$

$$a_{CP}^{T_{\text{odd}}} = (+0.6 \pm 5.0\ (\text{stat}) \pm 0.7\ (\text{syst})) \times 10^{-2}$$

are found compatible with the CP-conservation hypothesis.

This thesis is organized in the following sections:

- In section 1 there is the introduction to the analysis. First, the role of symmetries in the standard model and the study of discrete symmetry violation are presented, then the section details the theoretical motivations of the study of *CP*-violating asymmetries in $\Lambda_b \to pK^-\mu^+\mu^-$ decay and the performed *CP*-violation measurements. Finally the current experimental status regarding *CP*-violation searches and loop electroweak processes studies in the Λ_b sector is summarized.

- In section 2 the LHCb experiment is described. It is given an overview of the LHC ring, of LHCb subdetectors and of the LHCb trigger system. Moreover, the LHCb analysis and simulation software are depicted.

- In section 3 the muon identification procedure implemented at LHCb is detailed. Thanks to the very high performance of the muon identification, processes with two final-state muons like $\Lambda_b \to pK^-\mu^+\mu^-$ decay are among the cleanest ones in terms of background rejection.

- In section 4 it is explained how data relative to $\Lambda_b \to pK^-\mu^+\mu^-$ decay are extracted from the whole LHCb Run I dataset. Details regarding the stripping procedure, the applied preselection cuts and trigger requirements as well as the vetoes needed for removing physical background components are reported.

- In section 5 signal and background probability density functions (PDF) composing the fit model are chosen. Signal PDFs are derived from simulated datasets, differently for signal and control decays.

- Section 6 is devoted to the selection of signal events over the combinatorial background surviving preselection cuts. A multivariate (machine-learning) technique exploits differences in signal and background distributions of several variables for building a classifier. The cut on such a classifier is chosen in order to maximize a figure-of-merit minimizing the *CP*-violating observable statistical uncertainties. The selection efficiency is crucial for this analysis, so every detail of this selection is studied and optimized.

- In section 7 *CP*-observables are extracted using unbinned extended simultaneous maximum likelihood fits to selected data. The fit procedure is validated using several simulated toy MC experiments.

- Section 8 is devoted to the study of the possible systematic uncertainties affecting the *CP*-violating asymmetries measurements. Moreover, the performed cross-checks are described.

- In section 9 the final results and the analysis conclusions are reported.

- In the appendices some technical aspects of the analysis are detailed. It is derived the expected branching fraction between signal $\Lambda_b \to pK^-\mu^+\mu^-$ and control $\Lambda_b \to pK^-J/\psi$ decays. It is explained the resampling of particle identification variables and the kinematic reweighting of the simulated signal sample. Finally, a table of the employed variables is given for a better clarity.

Contents

1 Introduction

1.1 Theoretical introduction

1.1.1 Symmetries and invariances

Symmetries are the transformations of a physical system such that the transformed system is undistinguishable from the original one. The identification of the symmetries associated to a physical system is of great relevance for understanding its basic properties and for its description. Global symmetries, not depending on system coordinates, are connected to conserved quantities via Noether's theorem [1]. It states that for every continuous symmetry a corresponding conservation law exist, where the generators of the symmetry group are the conserved quantities, *i.e.* they are constant of the motion. As an example, it will be shown how translational symmetry is associated to the conservation of spatial momentum in the quantum mechanism formalism:
From basic quantum mechanics it is known that translations on a quantum system are represented by a three-parameter group of unitary operators $\mathcal{U}(\vec{a})$. The action of the translation of a spatial vector $\vec{a}$ on a quantum state is defined as

$$\mathcal{U}(\vec{a})\,|\psi(\vec{x})\rangle = |\psi(\vec{x}+\vec{a})\rangle \tag{1}$$

The translation properties imply these operators satisfy

$$\mathcal{U}(\vec{a})\mathcal{U}(\vec{b}) = \mathcal{U}(\vec{a}+\vec{b}) \qquad \mathcal{U}(\vec{a})\mathcal{U}(-\vec{a}) = 1 \tag{2}$$

while the symmetry requirement for every scalar product

$$\langle\psi(\vec{x})\,|\,\psi(\vec{y})\rangle = \langle\psi(\vec{x}+\vec{a})\,|\,\psi(\vec{y}+\vec{a})\rangle = \left\langle\psi(\vec{x})\,\middle|\,\mathcal{U}^\dagger(\vec{a})\mathcal{U}(\vec{a})\,\middle|\,\psi(\vec{y})\right\rangle \tag{3}$$

implies their unitarity

$$\mathcal{U}^\dagger(\vec{a}) = \mathcal{U}^{-1}(\vec{a}) = \mathcal{U}(-\vec{a}) \tag{4}$$

Translation operators can be expressed in terms of their generators $\vec{P}$, the spatial momentum operators, via the exponential map

$$\mathcal{U}(\vec{a}) = \exp\left(-\frac{i}{\hbar}\vec{P}\cdot\vec{a}\right) \tag{5}$$

For infinitesimal translations of parameter $\vec{\epsilon}$ this reduces to

$$\mathcal{U}(\vec{\epsilon}) = 1 - \frac{i}{\hbar}\vec{P}\cdot\vec{\epsilon} \tag{6}$$

Symmetry	Conserved quantity
Translations	Spatial momentum ($\vec{P}$)
Rotations	Angular momentum ($\vec{J}$)
Time translation	Energy (E)
Phase invariance	Electric charge (q)

Table 1: Symmetry-conserved quantity relations following from Noether's theorem.

The unitary time-evolution operator describing the evolution of quantum states is defined as

$$|\psi(t)\rangle = \mathcal{T}(t-t_0)\,|\psi(t_0)\rangle \tag{7}$$

for the evolution of a quantum state from a reference time t_0 to time t.
As a consequence of translational symmetry, the translation operators commute with the time-evolution operator

$$\begin{aligned}
&\langle \psi(\vec{x}-\vec{\epsilon},t_0)\,|\,\mathcal{T}(t-t_0)\,|\,\psi(\vec{y}-\vec{\epsilon},t_0)\rangle = \langle \psi(\vec{x}-\vec{\epsilon},t_0)\,|\,\psi(\vec{y}-\vec{\epsilon},t)\rangle \\
&= \langle \psi(\vec{x},t_0)\,|\,\psi(\vec{y},t)\rangle \\
&= \langle \psi(\vec{x},t_0)\,|\,\mathcal{T}(t-t_0)\,|\,\psi(\vec{y},t_0)\rangle \\
&= \left\langle \psi(\vec{x}-\vec{\epsilon},t_0)\,\middle|\,\mathcal{U}^\dagger(\vec{\epsilon})\mathcal{T}(t-t_0)\mathcal{U}(\vec{\epsilon})\,\middle|\,\psi(\vec{y}-\vec{\epsilon},t_0)\right\rangle
\end{aligned} \tag{8}$$

for every couple of states $|\psi(\vec{x})\rangle$,$|\psi(\vec{y})\rangle$. Therefore, $\mathcal{U}^\dagger(\vec{\epsilon})\mathcal{T}(t-t_0)\mathcal{U}(\vec{\epsilon}) = \mathcal{T}(t-t_0)$, implying $[\mathcal{U}(\vec{\epsilon}),\mathcal{T}(t-t_0)] = 0$. Substituting the expression for the infinitesimal translation Eq. 6 it is straightforward to obtain

$$[P_i,\mathcal{T}(t-t_0)] = 0 \tag{9}$$

for each momentum component. Commutation of the momentum operators with time-evolution implies conservation of momentum expectation values on every quantum state

$$\begin{aligned}
\langle P_i(t)\rangle &= \langle \psi(t)\,|\,P_i\,|\,\psi(t)\rangle \\
&= \left\langle \psi(t_0)\,\middle|\,\mathcal{T}^\dagger(t-t_0)P_i\mathcal{T}(t-t_0)\,\middle|\,\psi(t_0)\right\rangle \\
&= \left\langle \psi(t_0)\,\middle|\,\mathcal{T}^\dagger(t-t_0)\mathcal{T}(t-t_0)P_i\,\middle|\,\psi(t_0)\right\rangle \\
&= \langle \psi(t_0)\,|\,P_i\,|\,\psi(t_0)\rangle
\end{aligned} \tag{10}$$

therefore translational symmetry entails spatial momentum conservation.
Table 1 reports the relation between symmetry and conserved quantities following from Noether's theorem for some important symmetries.

A very important role in physics is played by local symmetries (depending on system coordinates). In fact, the invariance of a system under a local symmetry can determine the dynamics of a system. The first application of this principle is general relativity, where the invariance under arbitrary spacetime coordinate transformations completely constrains the gravitational interaction as curvature of the spacetime. This later inspired Yang-Mills theories [2], where the dynamics of the system is determined by the request of local invariance under a given symmetry group (called gauge group). Each generator of the symmetry group introduces a massless vector field mediating the interaction among those fields having a non-trivial transformation law under the gauge group. The interaction form results to be fixed by the local symmetry requirement except for the coupling constant value.

1.1.2 The standard model

The Standard Model (SM) of particle physics, currently the most reliable theory describing strong, weak and electromagnetic interactions, is a Lorentz-invariant Yang-Mills theory based on $SU(3) \otimes SU(2)_L \otimes U(1)_Y$ gauge group [3–6]. The $SU(3)$ group acts in the three-dimensional space of color charges giving rise to the strong interaction, mediated by 8 massless gluons, involving only quark fields. The $SU(2)_L \otimes U(1)_Y$ group generates the unified electroweak interaction. $SU(2)_L$ group acts in the two-dimensional space of weak isospin, where the L subscript indicates that only left-handed chirality fields transforms under this group. They transform as isospin doublets where the three neutrinos are paired with their associated leptons and the three up-type quarks with down-type ones,

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \qquad \begin{pmatrix} u \\ d_W \end{pmatrix}, \begin{pmatrix} c \\ s_W \end{pmatrix}, \begin{pmatrix} t \\ b_W \end{pmatrix} \tag{11}$$

where d_W, s_W, b_W are combinations of the three down-type quark mass eigenstates, a phenomenon known as quark mixing (Sec. 1.1.4). It is important to note that also ν_l are not neutrino mass eigenstate, this being a completely analogous lepton mixing phenomenon. The $U(1)_Y$ group acts as a phase transformation on fields, the Y subscript indicates that this is not the $U(1)$ gauge group of electromagnetism, but a different one based on hypercharge conservation. The electroweak interaction is mediated by 4 vector fields, corresponding to the 4 generators of the $SU(2)_L \otimes U(1)_Y$ group. The spontaneous symmetry breaking of the electroweak gauge group (Brout-Englert-Higgs mechanism [7, 8]) accounts for the huge mass of $W^\pm$ and Z^0 vector fields by means of an interaction with a complex scalar field[1]. The residual

[1]whose associated particle is called Higgs boson.

symmetry group $U(1)_{EM}$ is the gauge group of electromagnetism, associated to the massless photon.

1.1.3 Discrete symmetries

Both the mentioned global and local symmetries in Sec. 1.1.1 are continuous symmetries, as they are represented by a group of operators labelled by one or more continuous parameters. For example, rotations around a fixed axis are parametrized by the continuous angle θ of rotation. However, a very important role in physics is played also by discrete symmetries, described by a finite (or at least countable) number of operators, not continuously connected with the identity. In this case Noether's theorem does not hold. The most important discrete symmetries in particle physics are parity (P), charge-conjugation (C) and time-reversal (T) transformations as well as their sequential combinations CP and CPT [9].
Parity transformation is the inversion of spatial coordinates, equivalent to the transformation acted by a mirror[2]. Table 2 shows the effect of parity transformation on some physical quantities.
On quantum states it is represented by a unitary operator P acting as

$$P\,|\psi(\vec{x},t)\rangle = \eta_P\,|\psi(-\vec{x},t)\rangle \tag{12}$$

The normalization of quantum states is not changed by parity

$$\begin{aligned} 1 &= \langle\psi(-\vec{x},t)\,|\,\psi(-\vec{x},t)\rangle = \langle\psi(\vec{x},t)\,|\,\psi(\vec{x},t)\rangle \\ &= \left\langle\psi(\vec{x},t)\,\middle|\,P^\dagger P\,\middle|\,\psi(\vec{x},t)\right\rangle = |\eta_P|^2\,\langle\psi(-\vec{x},t)\,|\,\psi(-\vec{x},t)\rangle = |\eta_P|^2 \end{aligned} \tag{13}$$

implying η_P being a phase $e^{i\theta}$, with real θ. The consideration that a second application of parity restores the initial system, $P^2 = 1$, along with unitarity $PP^\dagger = 1$ imply that parity is also an hermitian operator $P = P^{-1} = P^\dagger$. Its eigenstates are said to be parity-defined states and their real eigenvalues ± 1 represent a multiplicative quantum number called P-parity.
Charge conjugation transforms each particle into its own antiparticle, changing sign to all charges associated to the quantum state. Table 2 reports the effect of charge-conjugation transformation on some physical quantities.
As parity, it is represented by a unitary and hermitian operator C acting as

$$C\,|\psi\rangle = \eta_C\,\left|\overline{\psi}\right\rangle \tag{14}$$

with η_C being a phase and a C-parity quantum number can be defined on charge-conjugation eigenstates.

[2]Actually the mirror transformation is an inversion of spatial coordinates followed by a rotation of angle π.

Observable		P-transformed	C-transformed	T-transformed
Time	t	t	t	$-t$
Position	$\vec{x}$	$-\vec{x}$	$\vec{x}$	$\vec{x}$
Momentum	$\vec{p}$	$-\vec{p}$	$\vec{p}$	$-\vec{p}$
Energy	E	E	E	E
Angular momentum	$\vec{J}$	$\vec{J}$	$\vec{J}$	$-\vec{J}$
Electric charge	q	q	$-q$	q
Electric field	$\vec{E}$	$-\vec{E}$	$-\vec{E}$	$\vec{E}$
Magnetic field	$\vec{B}$	$\vec{B}$	$-\vec{B}$	$-\vec{B}$

Table 2: Transformation of some physical quantities under parity, charge-conjugation and time-reversal transformations.

Existence of P- and C-parity allows to test directly the conservation of these symmetries on physical processes. If parity is a symmetry, transitions between states with different parities are forbidden and any observable A odd under parity (such that $PAP^\dagger = -A$) must have zero expectation value on every state of the system.
Time-reversal transformation reverses the time direction: hence it leaves unchanged the spatial coordinates but it reverses the direction of all momenta and swaps initial and final states. From conservation of the canonical commutation relation $[x, p] = i\hbar$ under time reversal operator T follows $TiT^\dagger = -i$, meaning that time-reversal is represented by an antiunitary (unitary and antilinear) operator. For this reason, investigation of T-conservation must be quite different from what possible for P and C symmetries. For instance, the observation of a non-zero expectation value of a T-odd observable does not necessarily imply T-violation [9]. In practice, conservation of time-reversal symmetry is usually tested indirectly, because of the impossibility of building a time-reversed version of a quantum state, apart few (too) simple cases. Obviously, it is not possible, for example, to experimentally implement the time-reversed state of a decaying particle starting from the incoherent superposition of its asymptotic daughter particles states.
Gravity, electromagnetic and strong interaction conserve separately P, C and T symmetries, as far as it is currently known. This is not true for electroweak interaction. The first suggestion that parity could not be a symmetry for weak interaction was put forward in 1956 by Lee and Yang [10] as an explanation of the so-called $\sigma - \tau$ puzzle, while the first experimental evidence of parity violation came one year later from study of cobalt-60 β-decay [11]. Due to the tight link between P and C symmetries in weak interaction, many of the

early experiments showing parity violation implied also charge-conjugation violation. It was soon realized that weak interaction maximally violate P and C symmetries, this because weak interaction involves only left-handed fields. In fact, given a left-handed neutrino, from weak interaction perspective does not exist neither the right-handed neutrino (its P-transformed) nor the left-handed antineutrino (its C-transformed). On the contrary, the fact that right-handed antineutrino existed (its CP-transformed) suggested that CP symmetry could be conserved by weak interaction. The discovery of CP-violation in kaon decays [12] was as unexpected as inexplicably small, raising a great interest in CPV studies, still lively now.
On the contrary, no evidence for CPT-violation has been observed to date. This is a very important fact because it is demonstrated that CPT-violation it is not compatible with a local, Lorentz-invariant field theory like the SM (CPT theorem [13]).

1.1.4 Quark mixing and CP-violation in the standard model

The idea of quark mixing aroused when in the '60 it was tried to classify the known u, d and s quarks into weak isospin representations of $SU(2)_L$ group. In fact, the simple classification as a doublet plus a singlet

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} s \end{pmatrix} \tag{15}$$

was not compatible with the observed $s \to u$ transitions. This problem was solved by the seminal idea of Cabibbo [14] who proposed that the isospin eigenstates could be a combination of d and s quark mass eigenstates characterized by a mixing angle θ

$$\begin{pmatrix} u \\ \cos\theta d + \sin\theta s \end{pmatrix}, \begin{pmatrix} -\sin\theta d + \cos\theta s \end{pmatrix} \tag{16}$$

However, it was later noticed that this classification allowed flavour-changing neutral current processes (FCNC, mediated by Z^0 field) to proceed with a much higher probability than observed. This conflict was settled by the introduction of the charm (c) quark (GIM mechanism [15]) and the isospin classification involving two doublets

$$\begin{pmatrix} u \\ \cos\theta d + \sin\theta s \end{pmatrix}, \begin{pmatrix} c \\ -\sin\theta d + \cos\theta s \end{pmatrix} \tag{17}$$

which correctly predicts the FCNC process suppression.
However, this scheme was still not complete because it did not include the

observed *CP*-violation in weak interaction. Indeed, a *CP*-violating term in the weak interaction Lagrangian must come from the presence of complex coefficients, nevertheless, with two quark generations (Eq. 17) the only degree of freedom is the real mixing angle θ. In 1973, Kobayashi and Maskawa [16] theorized the existence of a third quark generation introducing bottom (b) and top (t) quarks. This way, mixing of the three down-type quark isospin eigenstates (d_W, s_W, b_W) in terms of the mass eigenstates (d, s, b) is expressed by a unitary matrix called Cabibbo-Kobayashi-Maskawa (CKM) matrix

$$\begin{pmatrix} d_W \\ s_W \\ b_W \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{18}$$

With three quark generations mixing is parametrized by three real mixing angles and a complex phase allowing for *CP*-violation.
CKM matrix features an interesting hierarchical structure[3], captured by the Wolfenstein CKM matrix parametrization [17]

$$\begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} \tag{19}$$

as an approximation to third order of the hierarchical parameter $\lambda \sim 0.22$, ruling favoured and suppressed weak interaction processes. The amount of CPV provided by CKM matrix is very small, as needed, since the only matrix elements containing a complex phase V_{ub} and V_{td} are suppressed by a factor $\lambda^3 \sim 0.01$.
At present, the CKM mechanism successfully accounts for the observed amount of CPV in hadron processes. However, the measured CPV quantity is still insufficient for explaining the observed matter-antimatter asymmetry of the universe [18], for this reason new sources of CPV are expected to show up in beyond standard model (BSM) physics.
Another source of CPV within SM should come from the three generations lepton mixing, where CPV is given by the complex phase into the lepton mixing matrix. Leptogenesis models, aimed at explaining the observed matter-antimatter asymmetry via lepton mixing CPV, have been proposed [19]. Still in SM domain, the strong interaction Lagrangian admits a *CP*-violating term, but its presence has never been detected. The presence of such a CPV term can produce a non-zero neutron electric dipole momentum, which has been experimentally constrained to a very tiny amount. The discussion on

[3] Interestingly not observed in the analogous lepton mixing matrix.

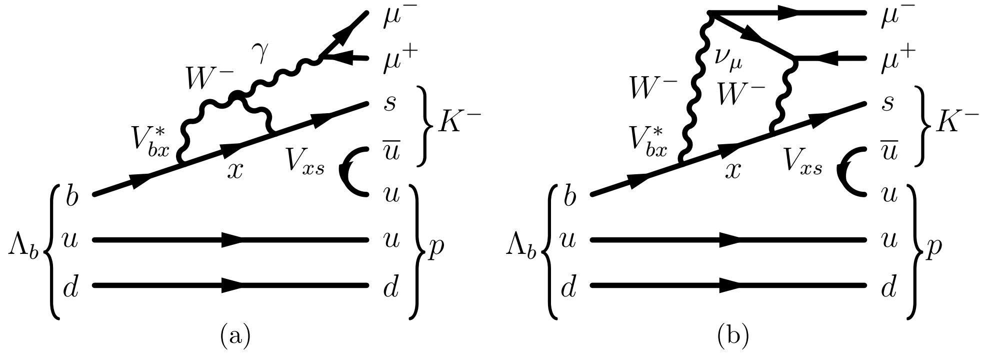


Figure 1: Electroweak penguin (a) and box (b) Feynman diagrams for $\Lambda_b \to pK^-\mu^+\mu^-$ decay: x represents one of the three up-type quarks u,c and t. The $u\overline{u}$ pair originates from the hadronization process.

why this term is so suppressed is known as strong *CP* problem, to address this issue different solutions have been proposed. Some solutions introduce new particles, like the axion predicted by Peccei-Quinn theory [20], while others try to explain the absence of the strong *CPV* term within the SM (*e.g.* [21]).

1.2 Analysis motivation

The $\Lambda_b \to pK^-\mu^+\mu^-$ decay is a flavour changing neutral current electroweak process, occurring only via electroweak loop (penguin) and $W^\pm$ box Feynman diagrams (Fig. 1), being prohibited at tree-level in SM. As this is a very rare process, it is a sensitive channel to look for BSM contributions. In fact, new particles can contribute to the process by circulating into loops or by providing new tree-level contributions, hopefully showing up in differences between SM-predicted and measured physics observables.

For $\Lambda_b \to pK^-\mu^+\mu^-$ decay the expected amount of *CPV* is expected to be very small in SM, because, looking at the CKM matrix Wolfenstein parametrization Eq. 19, amplitudes containing a *CPV* term are proportional to $V_{bu}V_{us} \sim \lambda^4$, while the others are proportional to $V_{bt}V_{ts} \sim \lambda^2$ and $V_{bc}V_{cs} \sim \lambda^2$ (Fig. 1). Hence, the *CP*-violating amplitudes are suppressed by a factor $\lambda^2 \sim 4.8\%$ relative to the leading ones. Contribution of possible BSM physics providing additional sources of *CPV* might therefore be detectable in a *CPV* measurement on this decay channel. In any case, evidence for *CP*-violation in $\Lambda_b \to pK^-\mu^+\mu^-$ decays, also if compatible with SM expectation, would be a remarkable result since no evidence of *CPV* has been found in any baryon decay. The only evidence for *CPV* in baryonic final-state pro-

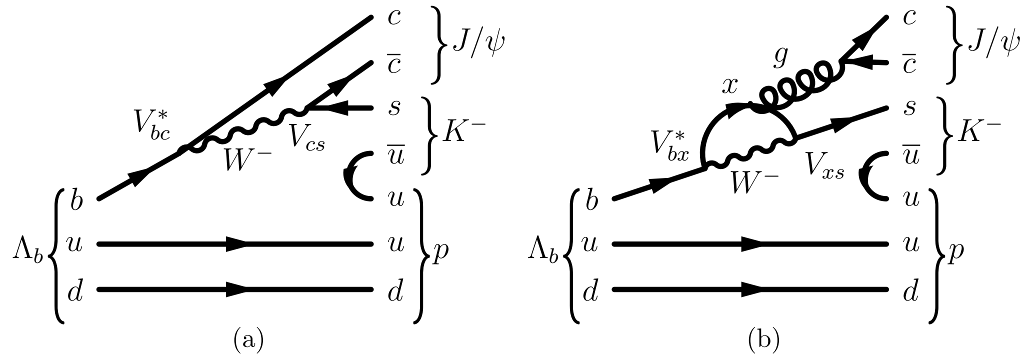


Figure 2: Tree (a) and electroweak penguin (b) Feynman diagrams for control $\Lambda_b \to pK^- J/\psi$ decay. Here x is one of the three up-type quarks, but only the t quark loop diagram has a significant effect being enhanced by the large t quark mass.

cesses was found for the decay $B^+ \to p\overline{p}K^+$ [22].
The $\Lambda_b \to pK^- J/\psi$ decay is used as a control channel, where J/ψ resonance decays to two muons (Fig. 2). It is mediated by a $b \to c$ transition not particularly suppressed by the CKM mechanism with far higher yield than $\Lambda_b \to pK^-\mu^+\mu^-$. $\Lambda_b \to pK^- J/\psi$ decay is especially suited as control channel because it has the same final particles as $\Lambda_b \to pK^-\mu^+\mu^-$, the main difference between the two processes is their kinematics, as $\Lambda_b \to pK^- J/\psi$ is a 3-body rather than a 4-body decay.

1.3 *CP*-violation measurements

CP-violation effects become manifest when a given process is compared with its *CP*-conjugate one, this can be done by building observables sensitive to differences of the two process differential rates. In principle such a comparison can be performed phase-space point by phase-space point. Following [23], $\mathcal{M}(\{\lambda_i, p_i\})$ represents the process amplitude in terms of final particle helicities λ_i and momenta p_i, while $\overline{\mathcal{M}}(\{\lambda_i, p_i\})$ denotes its *CP*-conjugate amplitude. CPV at any phase-space point takes the form of a difference between the squared moduli of $\mathcal{M}(\{\lambda_i, p_i\})$ and $\overline{\mathcal{M}}(\{\lambda_i, -p_i\})$, the latter having CP-transformed helicities (*CP*-even) and momenta (*CP*-odd). Therefore it is the comparison of differential rates of two processes involving *CP*-conjugate particles with identical helicities and opposite momenta.
It is useful to introduce the motion-reversal transformation $\hat{T}$ reversing both momentum and angular momentum, constituting the unitary component of the time-reversal transformation presented in Sec. 1.1.3, because motion-

reversal does not swap initial and final states.
The two considered amplitudes can be decomposed into $\hat{T}$-even and $\hat{T}$-odd parts:

$$\begin{aligned}\mathcal{M}(\{\lambda_i, p_i\}) &= \mathcal{M}_e(\{\lambda_i, p_i\}) + \mathcal{M}_o(\{\lambda_i, p_i\}) \\ \overline{\mathcal{M}}(\{\lambda_i, p_i\}) &= \overline{\mathcal{M}}_e(\{\lambda_i, -p_i\}) + \overline{\mathcal{M}}_o(\{\lambda_i, -p_i\}) \\ &= \overline{\mathcal{M}}_e(\{\lambda_i, p_i\}) - \overline{\mathcal{M}}_o(\{\lambda_i, p_i\})\end{aligned} \tag{20}$$

These four pieces can be expressed as modulus-phase decomposition

$$\begin{aligned}\mathcal{M}_e(\{\lambda_i, p_i\}) &= \sum_j a_e^j \exp\{i\left(\delta_e^j + \varphi_e^j\right)\} \\ \overline{\mathcal{M}}_e(\{\lambda_i, p_i\}) &= \sum_j a_e^j \exp\{i\left(\delta_e^j - \varphi_e^j\right)\} \\ \mathcal{M}_o(\{\lambda_i, p_i\}) &= \sum_k a_o^k \exp\{i\left(\delta_o^k + \left[\varphi_o^k + \pi/2\right]\right)\} \\ \overline{\mathcal{M}}_o(\{\lambda_i, p_i\}) &= \sum_k a_o^k \exp\{i\left(\delta_o^k - \left[\varphi_o^k + \pi/2\right]\right)\}\end{aligned} \tag{21}$$

where $a_{e,o}^{j,k}$, $\delta_{e,o}^{j,k}$, $\varphi_{e,o}^{j,k}$ are real functions of the helicities and momenta. All the *CPV* effects are included in the *CP*-odd weak phases $\varphi_{e,o}^{j,k}$, while $\delta_{e,o}^{j,k}$ are *CP*-even strong phases. These *CP*-even phases come from the contribution of all orders in QCD perturbation theory (known as effective Hamiltonian operator rescattering) representing final-state interactions (FSI) not described by the QCD factorization approach [24].
The differential rate of any pair of *CP*-conjugate processes can be decomposed into four parts with definite *CP* and $\hat{T}$ transformation properties:

$$\left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-}^{\mathrm{even}}_{\mathrm{odd}}}^{\hat{\mathrm{T}}-^{\mathrm{even}}_{\mathrm{odd}}} \equiv \frac{1\pm\mathrm{T}}{2}\frac{1\pm\mathrm{CP}}{2}\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi} \tag{22}$$

where $\Phi \equiv \{\lambda_i, p_i\}$. Assuming, for simplicity, that in the process under study there are respectively two $\hat{T}$-even and one $\hat{T}$-odd contributions, the amplitudes becomes

$$\begin{aligned}\mathcal{M}(\{\lambda_i, p_i\}) &= a_e^1 \exp\{i\left(\delta_e^1 + \varphi_e^1\right)\} + a_e^2 \exp\{i\left(\delta_e^2 + \varphi_e^2\right)\} \\ &\quad + i a_o^1 \exp\{i\left(\delta_o^1 + \varphi_o^1\right)\} \\ \overline{\mathcal{M}}(\{\lambda_i, p_i\}) &= a_e^1 \exp\{i\left(\delta_e^1 - \varphi_e^1\right)\} + a_e^2 \exp\{i\left(\delta_e^2 - \varphi_e^2\right)\} \\ &\quad + i a_o^1 \exp\{i\left(\delta_o^1 - \varphi_o^1\right)\}\end{aligned} \tag{23}$$

where all functions are evaluated at the same phase-space point Φ. The squared modulus of the above expressions provide the differential rates which can be decomposed as in Eq. 22. The *CP*-even terms are

$$\left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-even}}^{\hat{\mathrm{T}}-\mathrm{even}} = a_e^1 a_e^1 + a_e^2 a_e^2 + a_o^1 a_o^1 + 2a_e^1 a_e^2 \cos(\delta_e^1 - \delta_e^2)\cos(\varphi_e^1 - \varphi_e^2) \tag{24}$$

$$\left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-even}}^{\hat{\mathrm{T}}-\mathrm{odd}} = 2a_e^1 a_o^1 \sin(\delta_e^1 - \delta_o^1)\cos(\varphi_e^1 - \varphi_o^1) + 2a_e^2 a_o^1 \sin(\delta_e^2 - \delta_o^1)\cos(\varphi_e^2 - \varphi_o^1) \tag{25}$$

For *CP*-violation studies the relevant parts are the *CP*-odd terms, vanishing in the *CP*-conservation hypothesis, which take the form

$$\left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-odd}}^{\hat{\mathrm{T}}-\mathrm{even}} = -2a_e^1 a_e^2 \sin(\delta_e^1 - \delta_e^2)\sin(\varphi_e^1 - \varphi_e^2) \tag{26}$$

$$\left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-odd}}^{\hat{\mathrm{T}}-\mathrm{odd}} = 2a_e^1 a_o^1 \cos(\delta_e^1 - \delta_o^1)\sin(\varphi_e^1 - \varphi_o^1) + 2a_e^2 a_o^1 \cos(\delta_e^2 - \delta_o^1)\sin(\varphi_e^2 - \varphi_o^1) \tag{27}$$

These constitute two different types of *CP*-violating differential rates: a non-zero $\hat{T}$-even rate needs strong phase differences in at least two $\hat{T}$-even interfering amplitudes, while the $\hat{T}$-odd does not. In fact, in absence of strong phase contribution to the amplitude $\hat{T}$ is equivalent to T, so that *CPT* conservation implies any *CP*-odd quantity to be also $\hat{T}$-odd [25].
Although no phase-space integration of *CP*-odd rates is in principle required for testing *CP* conservation, practical constraints like the finite available statistics impose the introduction of phase-space integrated observables. In this analysis two classes of integrated observables are measured: direct *CP* asymmetries derived from $\hat{T}$-even differential rate and triple-product asymmetries resulting from $\hat{T}$-odd rate integration.

1.3.1 Direct *CP* asymmetry

Direct *CP* asymmetry is defined as the phase-space integration of the $\hat{T}$-even *CP*-odd differential rate over the $\hat{T}$-even *CP*-even one

$$\mathcal{A}_{CP} \equiv \frac{\int \mathrm{d}\Phi \left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-odd}}^{\hat{\mathrm{T}}-\mathrm{even}}}{\int \mathrm{d}\Phi \left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\mathrm{CP-even}}^{\hat{\mathrm{T}}-\mathrm{even}}} = \frac{\Gamma(i \to f) - \overline{\Gamma}(\bar{i} \to \bar{f})}{\Gamma(i \to f) + \overline{\Gamma}(\bar{i} \to \bar{f})} \tag{28}$$

reducing to the total rate asymmetry between the process under study and its *CP*-conjugate. Different observables may be defined multiplying the $\hat{T}$-even *CP*-odd differential rate by any $\hat{T}$-even function $f(\Phi)$.

An upper limit to direct *CP* asymmetry expectation in the SM for $\Lambda_b \to pK^-\mu^+\mu^-$ decay can be computed as follows: substituting the differential rates Eqs. 24,26 into direct *CP* asymmetry definition Eq. 28 one obtains

$$\mathcal{A}_{CP} = \frac{2A_e^1 A_e^2 \sin(\Delta_e^1 - \Delta_e^2)\sin(\phi_e^1 - \phi_e^2)}{A_e^1 A_e^1 + A_e^2 A_e^2 + 2A_e^1 A_e^2 \cos(\Delta_e^1 - \Delta_e^2)\cos(\phi_e^1 - \phi_e^2)} \tag{29}$$

where A_e^j, Δ_e^j and ϕ_e^j are the phase-space integrated versions of a_e^j, δ_e^j and φ_e^j, while $\hat{T}$-odd terms a_o^j are neglected. Referring to Fig. 1, let us suppose that amplitude $j = 1$ refers to a *CP*-conserving one and amplitude $j = 2$ to a *CP*-violating one. As seen in Sec. 1.2, the CKM hierarchy implies

$$A_e^2/A_e^1 \sim \lambda^2 = 4.8\%$$

therefore the denominator can be replaced by

$$(A_e^1)^2 \leq A_e^1 A_e^1 + A_e^2 A_e^2 + 2A_e^1 A_e^2 \cos(\Delta_e^1 - \Delta_e^2)\cos(\phi_e^1 - \phi_e^2)$$

The weak phase ϕ_e^1 is zero as no CPV term is present in amplitude 1, while amplitude 2 contains $V_{ub} = |V_{ub}|e^{i\phi}$ with a weak phase of $\phi = 1.2$ rad, hence $\sin(\phi_e^1 - \phi_e^2) = \sin(1.2) = 0.93$. Strong phases are hard to calculate, since they involve non-perturbative strong interaction effects, therefore the conservative upper limit $\sin(\Delta_e^1 - \Delta_e^2) \leq 1$ is placed. Following these considerations, the upper limit on $\mathcal{A}_{CP}$ is

$$\mathcal{A}_{CP} \leq \frac{2A_e^1 A_e^2 \sin(\Delta_e^1 - \Delta_e^2)\sin(\phi_e^1 - \phi_e^2)}{(A_e^1)^2} = \frac{2 \cdot 1 \cdot 0.93 A_e^2}{A_e^1} \leq 8.9\% \tag{30}$$

Evidence for a direct *CP* asymmetry well above this limit would be a clear sign of BSM physics.
Experimentally, direct *CP* asymmetries are measured starting from raw asymmetries

$$\begin{aligned}\mathcal{A}_{\text{raw}}(\Lambda_b \to pK^-\mu^+\mu^-) &\equiv \frac{N(\Lambda_b \to pK^-\mu^+\mu^-) - N(\overline{\Lambda_b} \to \overline{p}K^+\mu^+\mu^-)}{N(\Lambda_b \to pK^-\mu^+\mu^-) + N(\overline{\Lambda_b} \to \overline{p}K^+\mu^+\mu^-)} \\ &= \mathcal{A}_{CP}(\Lambda_b \to pK^-\mu^+\mu^-) + \mathcal{A}_{\text{prod}}(\Lambda_b) - \mathcal{A}_{\text{det}}(K) + \mathcal{A}_{\text{det}}(p) \end{aligned} \tag{31}$$

where terms in Eq. 31 are the *CP*, Λ_b production and detection asymmetries

$$\mathcal{A}_{\text{det}}(X) = \frac{\epsilon(X) - \epsilon(\overline{X})}{\epsilon(X) + \epsilon(\overline{X})} \tag{32}$$

with $\epsilon(X)$ being the detection efficiency on particle X. Inverting Eq. 31, the direct *CP* asymmetry (corresponding to that of Eq. 28) is expressed as

$$\begin{aligned}\mathcal{A}_{CP}(\Lambda_b \to pK^-\mu^+\mu^-) = \mathcal{A}_{\text{raw}}&(\Lambda_b \to pK^-\mu^+\mu^-) \\ &- \mathcal{A}_{\text{prod}}(\Lambda_b) + \mathcal{A}_{\text{det}}(K) - \mathcal{A}_{\text{det}}(p)\end{aligned} \tag{33}$$

Therefore $\mathcal{A}_{\text{raw}}$ is not an experimentally clean observable, because production and detection asymmetries are poorly known. But a similar expression holds also for the control decay $\Lambda_b \to pK^- J/\psi$

$$\begin{aligned}\mathcal{A}_{CP}(\Lambda_b \to pK^- J/\psi) = \mathcal{A}_{\text{raw}}(\Lambda_b \to pK^- J/\psi) \\ - \mathcal{A}_{\text{prod}}(\Lambda_b) + \mathcal{A}_{\text{det}}(K) - \mathcal{A}_{\text{det}}(p)\end{aligned} \tag{34}$$

The difference between Eqs. 33 and 34 is called $\Delta\mathcal{A}_{CP}$

$$\begin{aligned}\Delta\mathcal{A}_{CP} &\equiv \mathcal{A}_{CP}(\Lambda_b \to pK^-\mu^+\mu^-) - \mathcal{A}_{CP}(\Lambda_b \to pK^- J/\psi) \\ &= \mathcal{A}_{\text{raw}}(\Lambda_b \to pK^-\mu^+\mu^-) - \mathcal{A}_{\text{raw}}(\Lambda_b \to pK^- J/\psi)\end{aligned} \tag{35}$$

This observable is simply the difference between the two real *CP* asymmetries, independent of production and detection asymmetries, hence an experimentally clean observable. Since detection asymmetries depend on the decay kinematics, Eq. 35 is actually true only when the different kinematics between signal and control decays is taken into account. This is done by assigning a systematic uncertainty to $\Delta\mathcal{A}_{CP}$ measurement (Sec. 8.1.1).
Sensitivity of $\Delta\mathcal{A}_{CP}$ to *CP*-violation should not be decreased with respect to $\mathcal{A}_{CP}(\Lambda_b \to pK^-\mu^+\mu^-)$, because the $\Lambda_b \to pK^- J/\psi$ decay is expected to have a basically zero $\mathcal{A}_{CP}$ in the SM. In fact, the only diagram providing a CPV term is the penguin with a u quark loop one (Fig. 2(b)), suppressed with respect to the tree diagram (2(a)) and further suppressed by the tiny u quark mass and the CKM hierarchy factor λ^2.
As seen, a non-negligible direct *CP* asymmetry requires a large strong phase difference between two interfering amplitudes. For this reason, direct *CP* asymmetries are sensitive to SM-BSM interference contributions, because of the sizeable SM strong phase contributions, but it can be shown that only vector-axial BSM operators can produce a considerable CPV [24]. Such contributions can not arise from BSM-BSM interference since negligible BSM strong phase contributions are expected [26].

1.3.2 *CP*-violating asymmetries based on T-odd correlations

CPV effects can also be studied by using the so-called T-odd correlations [9, 23, 27, 28], that is by exploiting the $\hat{T}$-odd part of the differential rates Eqs. 26,27. Phase-space integrated $\hat{T}$-odd observables are constructed from differential rates as

$$\int \mathrm{d}\Phi f(\Phi) \left.\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\right|_{\text{CP}-\text{odd}}^{\hat{\text{T}}-\text{odd}} \tag{36}$$

with some $\hat{T}$-odd function $f(\Phi)$ without which the phase-space integral would vanish. One of the simpler way to define such a function is to take the asymmetry with respect a $\hat{T}$-odd observable $C_T(\{\lambda_i, p_i\})$, that is

$$f(\Phi) = \text{sign}\{C_T(\Phi)\}$$

One class of such observables are triple products $\vec{v}_1 \cdot (\vec{v}_2 \times \vec{v}_3)$, where $\vec{v}_i$ can be spins or momenta of final particles in the mother particle rest frame.
In the literature are usually defined the following asymmetries

$$\begin{aligned} A_T &\equiv \frac{\int \mathrm{d}\Phi f(\Phi) \left[\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{odd}}_{\mathrm{CP-even}} + \frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{odd}}_{\mathrm{CP-odd}} \right]}{\int \mathrm{d}\Phi \left[\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{even}}_{\mathrm{CP-even}} + \frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{even}}_{\mathrm{CP-odd}} \right]} \\ &= \frac{\Gamma(i \to f, C_T > 0) - \Gamma(i \to f, C_T < 0)}{\Gamma(i \to f, C_T > 0) + \Gamma(i \to f, C_T < 0)} \end{aligned} \tag{37}$$

$$\begin{aligned} \overline{A}_T &\equiv \frac{\int \mathrm{d}\Phi f(\Phi) \left[\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{odd}}_{\mathrm{CP-even}} - \frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{odd}}_{\mathrm{CP-odd}} \right]}{\int \mathrm{d}\Phi \left[\frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{even}}_{\mathrm{CP-even}} - \frac{\mathrm{d}\Gamma}{\mathrm{d}\Phi}\Big|^{\hat{\mathrm{T}}-\mathrm{even}}_{\mathrm{CP-odd}} \right]} \\ &= \frac{\overline{\Gamma}(\bar{i} \to \overline{f}, -\overline{C}_T > 0) - \overline{\Gamma}(\bar{i} \to \overline{f}, -\overline{C}_T < 0)}{\overline{\Gamma}(\bar{i} \to \overline{f}, -\overline{C}_T > 0) + \overline{\Gamma}(\bar{i} \to \overline{f}, -\overline{C}_T < 0)} \end{aligned} \tag{38}$$

where $\overline{C}_T$ is the C_T observable evaluated for the *CP*-conjugate process, that is using helicities and momenta of the *CP*-conjugate process final particles. Since A_T and $\overline{A}_T$ refer separately to the process under study and to its *CP*-conjugate[4], they have no definite *CP* transformation properties, therefore a non-vanishing value of these observables is not a sign of *CP*-violation.
On the contrary, their value is influenced by the presence of final-state interactions as A_T turns out to be proportional to

$$A_T \propto \sum_{j,k} 2A_e^j A_o^k \sin\left[(\Delta_e^j - \Delta_e^k) + (\phi_e^j - \phi_o^k)\right] \tag{39}$$

where a non-vanishing value can be produced by differences among strong phases Δ_e^j of interfering *CP*-even amplitudes [23]. Moreover, if the chosen C_T observable is *P*-odd, A_T becomes a *P*-odd observable, therefore its value being sensible also to parity-violation effects. This is the case when using triple-products like $\vec{p}_1 \cdot (\vec{p}_2 \times \vec{p}_3)$ or $\vec{p}_1 \cdot (\vec{\lambda}_1 \times \vec{\lambda}_2)$ as C_T observable.
A *CP*-odd observable allowing to unambiguously witness *CP*-violation is the difference between the two *CP*-conjugated process asymmetries, defined as

$$a_{CP}^{T_{\mathrm{odd}}} \equiv \frac{1}{2}\left(A_T - \overline{A}_T\right) \tag{40}$$

[4]Or equivalently since they are a combination of *CP*-even and *CP*-odd differential rates.

where the CP-even final-state contributions cancel out.
For $\Lambda_b \to pK^-\mu^+\mu^-$ decay, triple products (TP) are defined consistently with [29]:

$$C_T \equiv \vec{p}_{\mu^+} \cdot (\vec{p}_p \times \vec{p}_{K^-}) \qquad \overline{C}_T \equiv \vec{p}_{\mu^-} \cdot (\vec{p}_{\bar{p}} \times \vec{p}_{K^+}) \tag{41}$$

with momenta Lorentz-transformed to the Λ_b rest frame. TP asymmetries, corresponding to Eqs. 37,38, are defined as

$$A_T \equiv \frac{N_{\Lambda_b}(C_T > 0) - N_{\Lambda_b}(C_T < 0)}{N_{\Lambda_b}(C_T > 0) + N_{\Lambda_b}(C_T < 0)} \quad \overline{A}_T \equiv \frac{N_{\overline{\Lambda}_b}(-\overline{C}_T > 0) - N_{\overline{\Lambda}_b}(-\overline{C}_T < 0)}{N_{\overline{\Lambda}_b}(-\overline{C}_T > 0) + N_{\overline{\Lambda}_b}(-\overline{C}_T < 0)} \tag{42}$$

The $a_{CP}^{T_{\text{odd}}}$ observable defined in Eq. 40 is independent of Λ_b production asymmetry as well as reconstruction efficiency, because this efficiency is identical for both $C_T > 0$ and $C_T < 0$ events, separately for Λ_b and $\overline{\Lambda}_b$ decays. This fact is explicitly checked on simulated $\Lambda_b \to pK^-\mu^+\mu^-$ events (Sec. 8.3.2)
Contrary to direct CP asymmetry, a non-zero $a_{CP}^{T_{\text{odd}}}$ value needs only a weak phase difference to be produced, no strong phases between interfering amplitudes are required. For this reason CPV effects on this observable may stem from both SM-BSM and BSM-BSM contributions: in the first case important BSM effects can arise only from vector-axial operators, while in the latter one BSM contributions can mainly arise from scalar-pseudoscalar or tensor operators [24].

1.4 Experimental status of Λ_b particle study

Λ_b particle is the lightest beauty baryon, first observed by UA1 experiment at CERN [30]. The first extensive studies on this particle were performed by DØ and CDF experiments at Fermilab and by LHCb collaboration which precisely measured its mass [31] and lifetime [32].
The first CPV studies in Λ_b sector were performed by CDF collaboration who measured the direct CP asymmetry in $\Lambda_b \to pK^-$ and $\Lambda_b \to p\pi^-$ decays [33]

$$\begin{aligned} \mathcal{A}_{CP}(\Lambda_b \to pK^-) &= +0.37 \pm 0.17(\text{stat}) \pm 0.03(\text{syst}) \\ \mathcal{A}_{CP}(\Lambda_b \to p\pi^-) &= +0.03 \pm 0.17(\text{stat}) \pm 0.05(\text{syst}) \end{aligned} \tag{43}$$

LHCb experiment has investigated direct CP asymmetries in $\Lambda_b \to \overline{K}^0 p\pi^-$ [34] and $\Lambda_b \to p\pi^- J/\psi$ [35] decays

$$\mathcal{A}_{CP}(\Lambda_b \to \overline{K}^0 p\pi^-) = +0.22 \pm 0.13(\text{stat}) \pm 0.03(\text{syst}) \tag{44}$$

$$\mathcal{A}_{CP}(\Lambda_b \to p\pi^- J/\psi) - \mathcal{A}_{CP}(\Lambda_b \to pK^- J/\psi) = (+5.7 \pm 2.4(\text{stat}) \pm 1.2(\text{syst}))\%$$

All these CPV measurements are compatible with the CP conservation hypothesis. Study of CP-violation using T-odd correlations in $\Lambda_b \rightarrow ph^-h^+h^-$ decays in LHCb experiment is ongoing.
Λ_b decays involving $b \rightarrow s$ quark transitions occurring via electroweak penguin diagrams, like $\Lambda_b \rightarrow pK^-\mu^+\mu^-$, have been studied by LHCb experiment in $\Lambda_b \rightarrow \Lambda\mu^+\mu^-$ decays, first observed by CDF experiment [36], measuring its differential branching fraction and forward-backward asymmetries [37]. Measurement of $\Lambda_b \rightarrow pK^-\mu^+\mu^-$ differential branching fraction in LHCb experiment is ongoing.

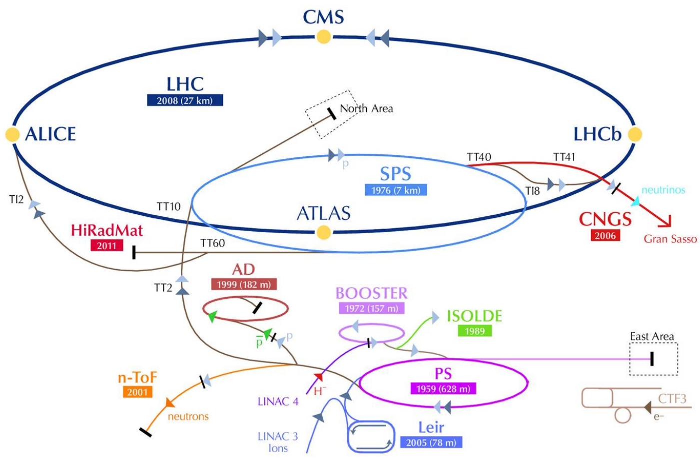


Figure 3: The CERN accelerator complex and the different experiments.

2 The LHCb experiment

LHCb is one of the main experiments at CERN (Conseil Européen pour la Recherche Nucléaire) in Meyrin, Geneva canton, Switzerland. The purpose of the experiment is to search for indirect evidence of new physics making very precise measurements on heavy hadron (composed by charm and beauty quarks) decays, especially by looking for rare processes and *CP*-violating effects. LHCb experiment analyses data recorded from proton-proton collisions taking place inside the LHC (Large Hadron Collider) accelerator. Throughout this section an overview of the LHCb detector during Run I is given, since this analysis is performed on the Run I data. The major improvements implemented for Run II are mainly related to the trigger system.
Figures reported in this section are taken from CERN or LHCb related websites.

2.1 Large hadron collider

The particle accelerator LHC [38] is built inside a 26.7 km circular tunnel at a depth of about 100 m underground, between Switzerland and France. Up to now, it is the most powerful particle accelerator ever built, operating at a centre of mass energy $\sqrt{s}$ of 7 TeV from 2010 to 2011, 8 TeV during 2012 and 13 TeV from 2015, after the two years planned stop called Long Shutdown 1. The accelerator complex at CERN, consisting of four smaller accelerators (see Fig. 3), injects the proton flux into the LHC pipes after a stepwise preliminary acceleration up to a proton momentum of 450 GeV/c.

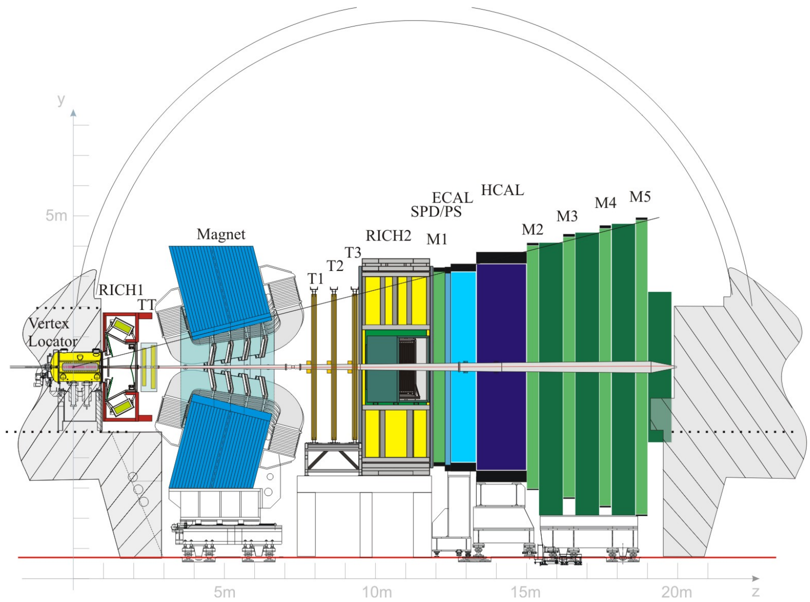


Figure 4: Lateral view of the LHCb detector.

LHC is composed of two beam pipes where the proton bunches travel in opposite directions in ultra-high vacuum. The necessary acceleration is given by a system of helium-cooled superconducting magnets operating at $1.9K$ temperature placed alongside the ring. The protons are then forced to collide in four points (corresponding to the four experiments on LHC ring: IP1 - ATLAS, IP2 - ALICE, IP5 - CMS, IP8 - LHCb) with a crossing rate of 40 MHz, namely 25 ns between two consecutive collisions. This corresponds to a nominal peak luminosity of $L = 10^{34} cm^{-2} s^{-1}$.

2.2 LHCb detector

The LHCb detector has been specifically designed for flavour physics studies at LHC accelerator. Its peculiar structure is much more similar to a fixed-target experiment design than to the "barrel plus endcaps" design which characterizes the other large detectors at CERN. In fact, the dominant production mechanism of $b\bar{b}$ quark pairs is the fusion of two (or more) gluons radiated by the proton quark constituents. This results in a peaked angular distributions at low polar angles (Fig. 5), so beauty particles are very likely to be produced in the same forward (or backward) direction at very small angles with respect to the beam axis. The LHCb detector is thus designed as a single-arm forward spectrometer with an angular acceptance from ≈10 mrad to 300 mrad in the magnet bending plane and to 250 mrad in the non-bending

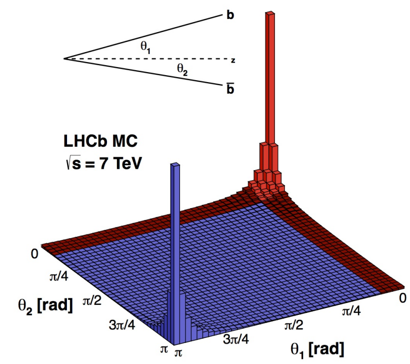


Figure 5: $b\bar{b}$ production angle in LHCb experiment at $\sqrt{s} = 7\,\text{TeV}$.

plane, which corresponds to a pseudorapidity range of $2 < \eta < 5$ [39]. The LHCb detector is composed by several subdetectors, each one has a specific design and is optimized to measure a different physical quantity (see Fig. 4). The LHCb experiment in Run I operated at a nominal average luminosity of $2 \times 10^{32}\ cm^{-2}s^{-1}$ which is about 1/50 of the nominal peak luminosity of LHC. This choice was made in order to have a smaller number of interaction (usually one p-p collision per bunch crossing) for a better vertex reconstruction, which is a fundamental requirement for flavour physics. Indeed, multiple p-p interactions in the same event may be mistaken for secondary (detached from the interaction point) vertices.
At LHCb experiment a large quantity of beauty particles (compared to other B-factories experiment) are produced thanks to the rather high $b\bar{b}$ cross-section. At $\sqrt{s} = 7\,\text{TeV}$ p-p center of mass energy the cross-section to produce beauty hadrons is measured to be [40]

$$\sigma(pp \to H_b X) = (75.3 \pm 5.4_{stat} \pm 13.0_{syst})\ \mu\text{b} \tag{45}$$

Extrapolating this result to the full pseudorapidity region the total $b\bar{b}$ cross-section is found to be

$$\sigma(pp \to b\bar{b}X) = (284 \pm 20_{stat} \pm 49_{syst})\ \mu\text{b} \tag{46}$$

2.2.1 Vertex locator (VELO)

Starting from the interaction point, the first subdetector is the Vertex Locator (VELO) [41]. Displaced secondary vertices are a distinctive feature

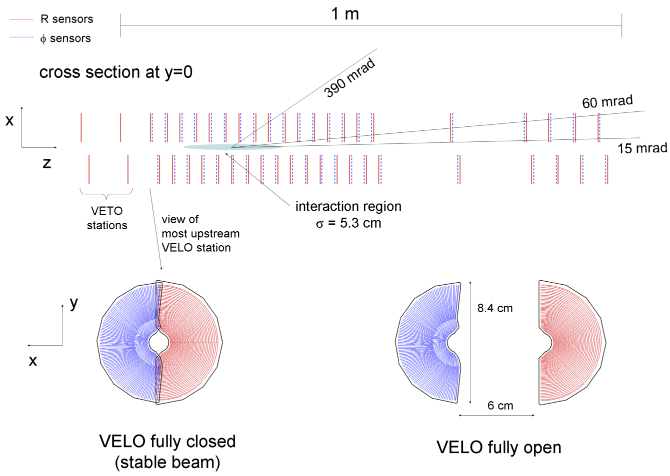


Figure 6: VELO configuration and VELO sensors.

of charm and beauty hadron decays, so a fundamental requirement for the LHCb experiment is a precise measurement of track coordinates close to the interaction point. The VELO system consists in 25 silicon stations placed along the beam direction, each one composed by one left module and one right module. Each module is composed by two measuring sensors: a radial one (R sensor) and an angular one (Φ sensor). The azimuthal coverage is about 182° for each sensor, giving a small overlap between the right and left modules in order to simplify the relative alignment and to guarantee a full azimuthal acceptance. The VELO system configuration and a schematic view of the sensors are shown in Fig. 6.
The VELO system is able to measure tracks in the full LHCb angular acceptance, moreover, the backward hemisphere is also partly covered (for improving the primary vertex resolution) and the two most upstream modules are used as a pile-up veto counter for the Level-0 trigger. The error on the primary vertex position is mainly related with the number of tracks produced in a p-p collision. For a typical event, the resolution in the beam direction is 42 μm and 10 μm in the perpendicular direction.

2.2.2 Ring-imaging Cherenkov (RICH) detectors

Hadron identification in LHCb is performed by using a high performance Ring-Imaging Cherenkov system composed by two detectors aiming at different momentum ranges [42]. RICH1 is located upstream of the magnet and it identifies low momentum particles (from 1 GeV/c up to about 60 GeV/c).

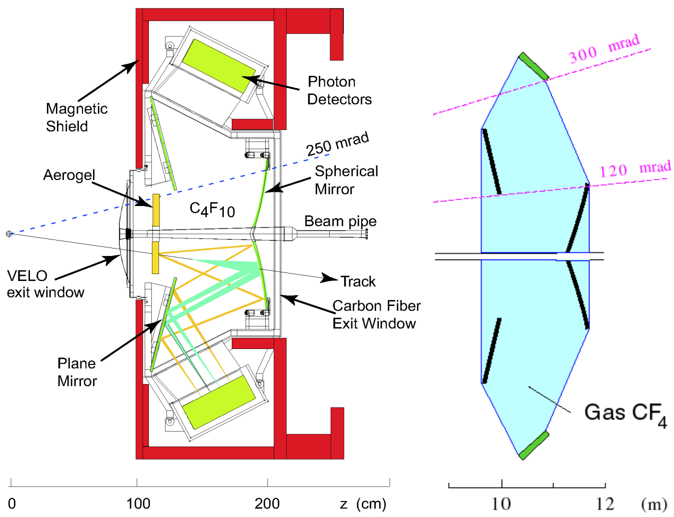


Figure 7: (left) RICH 1 and (right) RICH 2 schemes.

It is made by silica aerogel and C_4F_{10} gas radiators with a polar angle acceptance from 25 to 300 mrad. RICH2, located downstream of the magnet and the tracking stations, has a more limited angular acceptance (from 15 to 120 mrad in the horizontal plane and from 15 to 100 mrad in the vertical plane) and it covers the high momentum range, from about 15 GeV/c up to about 100 GeV/c by means of a CF_4 radiator. Both RICH1 and RICH2 layouts are shown in Fig. 7.
Cherenkov light is focused onto the photon detector planes using tilted spherical mirrors and secondary plane mirrors, in order to reflect the image out of the spectrometer acceptance. The baseline photon detectors are multianode photomultiplier tubes (MaPMT).

2.2.3 Tracking system

The tracking system consists of 4 stations (TT, T1, T2, T3) between the vertex detector and the calorimeters. It provides efficient reconstruction and precise momentum measurement of charged tracks, track directions for ring reconstruction in the RICH and information for the higher level trigger [43, 44]. Each tracking station (except for TT) is composed by an Inner Tracker (IT) located in an elliptical shaped region around the beam pipe and an Outer Tracker (OT) which covers most of the acceptance. In the region covered by the IT the particle flux is about 20 times bigger than the flux in the OT region, so IT requires a finer granularity and a different technology to deal with a greater flux.

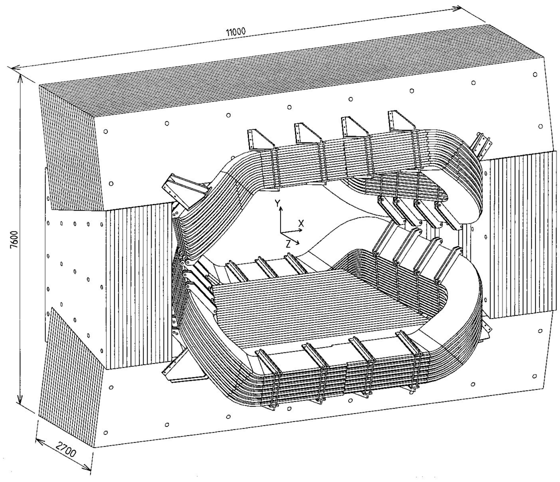


Figure 8: LHCb magnet.

The Trigger Tracker (TT) is located between RICH1 and the magnet. Due to its reduced dimensions it contains only IT modules. The other stations are located between the magnet and RICH2. The silicon tracker system (composed by TT and IT) uses approximately 270000 silicon microstrips detectors with a strip pitch of 198 μm for the IT and 183 μm for the TT. For improving track reconstruction, these detectors are composed by four layers in which strips are vertical in the first and in the last layer whereas the other two layers are rotated by stereo angles of $\pm 5°$, providing sensitivity in the vertical direction. The Outer Tracker is composed by an array of straw tubes modules each one consisting of two panels and two sidewalls forming a stable, gas-tight box containing up to 256 straw tubes. Each tube has an inner diameter of 4.9 mm, filled with a mixture of argon (70%) and carbon dioxide (30%) guaranteeing fast drift time and a sufficient drift-coordinate resolution (200 μm). Like the silicon tracker, also OT modules are composed by four layers arranged in a similar geometry.

2.2.4 Magnet

For achieving a good momentum resolution the LHCb experiment employs a dipole warm magnet for bending charged particles tracks [45]. The magnet is composed by two trapezoidal coils bent at 45° on the two transverse sides and placed mirror-symmetrically (see Fig. 8). It has a bending power (given by the integrated magnetic field) of 4 Tm that allows to measure momenta of charged particles up to 200 GeV/c within 0.5% uncertainty.

2.2.5 Calorimeters

The LHCb calorimeter system [46] has been devised for the identification of high transverse energy hadrons, electrons and photons candidates. It measures their energy and position providing essential information for the Level-0 trigger candidate selection. Its structure consists of a single-layer pre-shower detector (PSD) made by 14 mm thick lead plates and 10 mm square scintillators, followed by an electromagnetic calorimeter (ECAL) and an hadronic calorimeter (HCAL). A scintillator pad detector (SPD) is located before the PSD.

The ECAL is made up of 70 layers consisting of 2 mm thick lead plates and 4 mm thick polystyrene-based scintillator plates. It has the important function of detecting photons with adequate granularity and energy resolution for prompt photons and π^0 reconstruction. For $p_{\mathrm{T}} < 2\,\mathrm{GeV}/c$ neutral pions are mainly reconstructed combining two separate clusters in the ECAL (resolved π^0), while for greater transverse momenta the two photons are measured as a single cluster (merged π^0).

The HCAL consists of 16 mm thin iron plates spaced with 4 mm thick scintillating tiles arranged parallel to the beam pipe.

2.2.6 Muon system

The muon system for the LHCb experiment [47] consists of five tracking stations placed along the beam axis, the first one (M1) lies in front of the calorimeters and the other four, interleaved with three iron filters, lie after the calorimeters (see Fig. 9). The muon stations are equipped with Multi Wire Proportional Chambers (MWPCs) except for the inner region of M1. This region is equipped with Gas Electron Multiplier (GEM) chambers which have better ageing properties as they have to stand a greater particle flux. The efficiency for each chamber is required to be enough to achieve a 95% trigger efficiency on the trigger algorithm, which requires a five-fold coincidence in a time lapse smaller than the p-p collision period (25 ns). This gives a lower momentum threshold for efficient muon triggering of about $6\,\mathrm{GeV}/c$, since the total absorber thickness is about 20 interaction lengths.

The inner and outer angular acceptances of the muon system are 20 (16) mrad and 306 (258) mrad in the bending (non-bending) plane, similar to the tracking system acceptances. This corresponds to a geometrical acceptance of about 20% of the full solid angle for muons coming from b quark decays. Each station is subdivided in four regions R1-R4 whose linear dimensions and logical pad size scale in the ratio 1:2:4:8 with the beam distance.

The detector granularity varies such that its contribution to the muon p_{T}

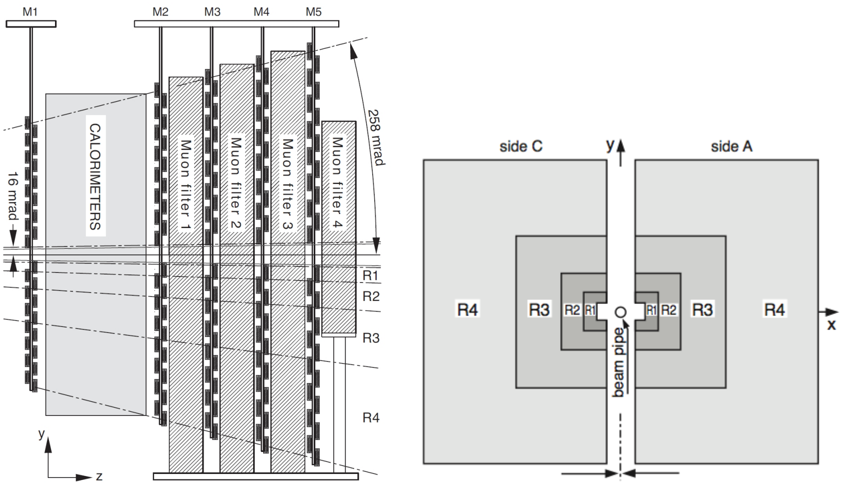


Figure 9: (left) Side view of the Muon System and (right) Station layout with the four regions R1-R4, from [48].

resolution compensates the multiple-scattering contribution. This latter contribution increases as the muons energy decrease, therefore the detector resolution must increase with the distance from the beam axis because high-momentum tracks tend to be closer to the beam.

2.2.7 Trigger

In LHCb detector useful proton-proton interactions occur at a rate of 10 MHz but this rate needs to be reduced at about 2-5 kHz in order to be able to store all the interesting data for the offline analysis. This is achieved by means of two trigger selections: the Level-0 trigger (L0) and the High Level Trigger (HLT) [39, 49, 50].
L0 is hardware-implemented and reduces the event rate at 1 MHz, at which the detector can be read out. L0 selects candidates combining information from calorimeter and muon system. HLT is software-based and reduces the event rate at 2 kHz using a full event reconstruction based on particle identification, track measurement, vertex reconstruction and impact parameter measurement. It consists of two different levels HLT1 and HLT2 acting sequentially, where HLT1 is designed to be faster than HLT2 to cope with the L0 output event rate while HLT2 performs a careful event reconstruction very close to the offline one.

2.3 LHCb analysis software

All LHCb analysis software is developed under `Gaudi` [51, 52] framework, an experiment independent open project providing interfaces and services for building HEP experiment frameworks for event data processing applications. Dedicated packages take care of the different steps necessary to obtain completely reconstructed candidates from raw data as well as accurate event simulations describing both particle physics and detector response.

2.3.1 Offline data reconstruction

Data selected by both L0 and HLT trigger levels, consisting of digitized data coming from the different subdetectors, are stored for the offline analysis. Candidates reconstruction is performed by `Brunel` [53] package: it takes care of track identification starting from measured hits in each detector, reconstructs particles momentum and energy as well as primary and secondary vertexes (derived from the intersection of two or more tracks) and performs the particle identification process combining data from RICH, ECAL, HCAL and muon system and computing particles PID variables (see for example Sec. 3.5 for muon PID variables).
Next, data go through the *stripping* process, where candidates are divided into categories called *streams* grouping decays with common features. For example, events with at least a couple of high p_{T} muons with opposite charges like $\Lambda_b \to pK^-\mu^+\mu^-$ decays belong to the `Dimuon` stream. For each stream several loose selection criteria called *stripping lines* are defined. Each stripping line is a set of standard requirements for restricting the search for a particular decay channel to a subset of the whole data. $\Lambda_b \to pK^-\mu^+\mu^-$ decays are searched by applying the `B2XMuMu` stripping line, selecting beauty particle candidates decaying to a dimuon final state. Note that this line selects also other channels like $\Lambda_b \to pK^-J/\psi$, our control channel, or like $B_s^0 \to \phi\mu^+\mu^-$ that will show up as physical background. The stripping procedure as well as the complete decay chain reconstruction is performed by `DaVinci` [54] package.

2.3.2 Monte Carlo simulation

Simulated events generated with the Monte Carlo (MC) method are extensively used throughout this analysis for studying decay properties, in particular for the rare $\Lambda_b \to pK^-\mu^+\mu^-$ decay which is kept blind. MC dataset production process is carried out by `Gauss` [55] package, interfaced with specific packages for handling generation and simulation phases. Generation of the physical process including the proton-proton interaction described by the known theoretical models and the subsequent hadronization process is performed by `PYTHIA` [56] package, while simulation of the mother B particle decay of interest is done by `EVTGEN` [57] package. The simulation phase consisting in describing the interaction among particles and detectors, taking into account all the geometric and material details of each subdetector, is made by `Geant4` [58, 59] package. Output produced by `Gauss` package is further elaborated by `Boole` [60] package, the LHCb digitization application, simulating the detector response, the readout electronics behaviour and also the Level-0 trigger level. Its output, digitized data reproducing real data coming from the real detector, are then reconstructed exactly in the same way as real data.

3 Muon identification

Muons are the particles that can be best identified in HEP experiments thanks to their very little interaction with matter. For this reason, dimuon final state decays like $\Lambda_b \to pK^-\mu^+\mu^-$ provide channels with excellent background suppression and hence are "golden modes" for seeking new physics. Throughout this section it is explained how muon identification is carried out in LHCb.

3.1 At detector level

The identification of a reconstructed track in the tracking system as a muon is based on the association of the measured hits inside the muon system around its extrapolated trajectory. A search for hits around the extrapolation points is performed within rectangular windows whose dimensions change as a function of the track momentum at each station and separately for each muon system region.

A muon needs a minimum momentum of 3 GeV/c to pass through the calorimeters and reach the M2 and M3 stations, while a minimum momentum of 6 GeV/c is needed for traversing all the five stations.

The possible misidentification of a proton as a muon may be due to either a combination of wrong hits in the muon stations that are aligned with a proton track or the existence of a true muon whose track inside the muon system points exactly in the direction of a proton track. This muon can be produced either close to the interaction point or in the calorimeter shower. Since the window dimension decreases with momentum and an increasing number of hits is required for tracks above 10 GeV/c, a strong reduction of the proton misidentification rate is seen in the interval $3 - 30$ GeV/c and for $p_{\mathrm{T}} < 1.7$ GeV/c. Decays in flight are the main cause of misidentification of pions and kaons as muons. To a good approximation, the misidentification rate is the sum of the contribution of decays in flight and the proton misidentification probability.

3.2 At L0 trigger level

There are two main physics trigger in L0 devoted to muon selection: the `L0Muon` and `L0DiMuon` lines [61]. These L0 muon processors look for the two highest p_{T} muon tracks reconstructed in the muon system. The position of a track in the muon chambers alone allows the determination of its p_{T} with a precision of roughly 20%. The `L0Muon` line sets a threshold on the largest

transverse momentum $p_{\mathrm{T}}^{largest}$ (1.48 GeV/c in 2011, 1.76 GeV/c in 2012) of the muon candidates. The `L0DiMuon` line sets a threshold on the product of the two largest transverse momentum, $p_{\mathrm{T}}^{largest} \times p_{\mathrm{T}}^{2^{nd}largest}$((1.30 GeV/$c$)^2 in 2011, (1.60 GeV/$c$)^2 in 2012) of the muon candidates [62]. The tightening of L0 thresholds in the 2012 configuration is a consequence of the increased luminosity and beam energy.
Moreover, all events are required to be clean enough. This is reached by limiting the hit multiplicity in the SPD detector at 900 SPD hits for L0 dimuon triggers and at 600 SPD for all other L0 triggers.
The trigger efficiencies are assessed by using events reconstructed with the full offline software and are computed with respect to candidates selected by the full offline analysis of the respective channel excluding trigger. This way is therefore possible to measure the inefficiencies due to simplified reconstruction, possible misalignments and reduced resolution relative to the offline reconstruction. They also account for the selection requirements made to satisfy rate and processing time limitations.

3.3 At HLT1 level

The muon triggers in HLT1 are executed only on events which have been triggered by `L0Muon` or `L0DiMuon` lines. The HLT1 reconstruction and selection sequence is a faster version of the offline one:

- A fast version of the offine Velo tracking is executed, followed by a vertex reconstruction stage. Vertices found in a 300μm radius of the mean *p*-*p* interaction region are considered primary vertices (PV).
- Next, the track candidates are filtered using a fast muon identification algorithm. Only Velo track candidates which have sufficient muon hits associated are then used as seeds for the forward track pattern recognition step.
- The produced long tracks (tracks having hits in both VELO and tracking system) are then fitted with a fast version of the offline Kalman track fitter and the offline muon ID based on the fitted long tracks is applied to increase the purity of the muon sample.

There are three main purposes for single muon triggers in HLT1:

- **Inclusive signal trigger:** for B mesons, D mesons and τ lepton decays with one or more muons in the final state, where the inclusive selection is based on only one of the muons. The displaced single muon trigger `Hlt1TrackMuon` discriminates signal from background exploiting the muon p_{T} and impact parameter (respect to the primary vertex) χ^2 (χ^2_{IP}).

- **Electroweak signal trigger:** for muons coming from $W^{\pm}$ and Z^0 decays. It exploits the very high p_{T} of prompt muons for efficiently triggering these electroweak events. It is implemented in the prompt high p_{T} single muon trigger `Hlt1SingleMuonHighPt`.

- **Calibration triggers:** there is a variety of single muon triggers for calibration and the determination of tracking and trigger efficiencies.

The HLT1 dimuon triggers act as signal trigger for many analyses with two or more muons in the final state. The trigger decisions are based on the mass of the dimuon candidate and the separation of the dimuon vertex in terms of the muon χ^2_{IP}. There are two different selections:

- **High mass dimuons:** Dimuon vertices with invariant mass at or above the J/ψ one can be selected by the `Hlt1DiMuonHighMass` trigger line without requiring any vertex separation. Both muons are required to pass the muon ID steps and they are then combined if their distance of closest approach (DOCA) is below 0.2 mm. The dimuon vertex is required to have a reasonable fit quality ($\chi^2 < 25$). Dimuon candidate vertexes with an invariant mass above 2.7 GeV/c^2 are then accepted by the HLT1 trigger. No vertex separation related quantity is cut on to allow a precise analysis of trigger acceptance.

- **Detached dimuons:** Combinations of muons with lower masses are selected in the trigger line `Hlt1DiMuonLowMass`. This trigger line accepts dimuon vertexes with a mass above 1 GeV/c^2 where both tracks must have a significant χ^2_{IP} with respect to their best primary vertex (the one which minimizes the IP ($\chi^2_{\mathrm{IP}} < 3$).

3.4 At HLT2 level

HLT2 performs a full reconstruction of the event at the output rate of HLT1 (40 kHz). The track reconstruction sequence of HLT2 is designed to be as similar as possible to the offline sequence provided the limited time available. With respect to HLT1 sequence there are only two differences, here the non-pointing VELO tracks are also reconstructed (as in the offline sequence) and the tracks are globally reconstructed.
The single muon triggers in HLT2 are devised for the following purposes:

- **Semileptonic signal trigger:** Semileptonic decays of B and D mesons can be triggered with a detached single muon trigger without biasing the hadronic part of the event. This is one of the purposes of the trigger line `Hlt2SingleMuon`.

- **Calibration trigger:** Many calibration and efficiency methods use $J/\psi \to \mu^+\mu^-$ decays in tag & probe methods. Reasonable efficiency for this task is provided by the `Hlt2SingleMuon` detached single muon trigger. For lowering the huge rate of single muon events this line firstly requires the candidate to have been triggered by the `Hlt1TrackMuon` line (TOS required, see definition in Sec. 4). In addition, a very good track quality (track $\chi^2 < 2$), $p_{\mathrm{T}} > 1.3$ GeV/c and very large impact parameter and χ^2_{IP} are requested (IP > 0.5 mm and $\chi^2_{\mathrm{IP}} > 200$). The p_{T} requirement is kept not too high for not biasing the hadronic part of semileptonic decays.

- **Electroweak signal trigger:** $W^{\pm}$ and Z^0 decays to one or two muons can be triggered effectively by exploiting the high p_{T} of the tracks by the prompt single muon trigger line `Hlt2SingleMuonHighPt` which is similar to its HLT1 counterpart but with tighter cuts ($p_{\mathrm{T}} > 10$ GeV/c) for suitably reducing the event rate.

Inclusive dimuon triggers consists of prompt trigger selections, where the dominant discriminants used to distinguish between signal and background are the dimuon mass and muon identification algorithm response, and detached dimuon trigger selections, based on the decay length significance (DLS) of the dimuon vertex. Prompt dimuon triggers are organized in three lines: two with mass windows of 120 MeV/c^2 around the J/ψ (`Hlt2DiMuonJPsi`) and $\psi(2S)$ (`Hlt2DiMuonPsi2S`) masses whereas the third line accepts all dimuons with invariant mass above 4.7 GeV/c^2 (`Hlt2DiMuonB`). There is no prompt trigger selection for dimuon vertices with mass under the J/ψ mass, for reducing the events rate. For these lines no p_{T} cut is applied for not biasing the muon's final state angular distribution, while a good track and vertex quality are required (track χ^2/ndf[5] < 5, vertex $\chi^2 < 25$, $\chi^2 < 10$ for `Hlt2DiMuonB`). There exist also two lines which apply cuts on the dimuon resonance p_{T}: `Hlt2DiMuonJPsiHighPT` ($p_{\mathrm{T}}(\mu\mu) > 2$ GeV/c for J/ψ events) and `Hlt2DiMuonPsi2SHighPT` ($p_{\mathrm{T}}(\mu\mu) > 3.5$ GeV/c for $\psi(2S)$ events).
Detached dimuons triggers in HLT2 are organized in three inclusive lines:

- `Hlt2DiMuonDetached` is primarily devoted to select dimuons with mass below the J/ψ one. The candidate must have $m(\mu\mu)$ above 1 GeV/c^2,

[5]Here ndf is the number of degrees of freedom available for the track fitting

a significant χ^2_{IP} of the muons and a vertex separation with respect to its closest PV ($\chi^2_{\text{IP}} > 9$ and $DLS > 7$) and a minimum $p_{\text{T}}(\mu\mu)$ of 1.5 GeV/c.

- `Hlt2DiMuonDetachedHeavy` is a similar line for dimuons resonating at the J/ψ mass or at heavier mass resonances with the non-mass cuts loosened with respect to the previous line ($m(\mu\mu) > 2.95$ GeV/c^2, $DLS > 5$ and no requirements on χ^2_{IP} and $p_{\text{T}}(\mu\mu)$).
- `Hlt2DiMuonDetachedJPsi` enhances the efficiency within the relatively small J/ψ mass window ($\Delta m < 120$ MeV/c^2) by requiring a reduced vertex separation compared to the `Hlt2DiMuonDetachedJPsi` line ($DLS > 3$).

3.5 Muon PID variables

As part of the LHCb muon identification procedure [63] different particle identification variables (PID) are defined and used for background rejection

- **IsMuon**: it is a boolean (yes/no) decision that states if a given track is compatible with hits in the muon system. Given a track reconstructed in the tracking system, hits in the muon stations are searched around the track extrapolation in a suitably defined Field of Interest (FOI). The decision is positive if a track has at least one hit within FOI in a certain number of stations depending on track momentum (see Table 3).
- **Delta Log Likelihood (DLL)**: using tracks passing the IsMuon selection, likelihoods for the muon and non-muon hypotheses P_μ and $P_{\not\mu}$ are computed for each muon candidate. These quantities are estimated by basing on the average squared distance D of the N hits within the FOI from track extrapolation

$$D = \sum_{i=0}^{N} \frac{1}{N} \left\{ \left(\frac{x_i - x_{track}}{pad_x} \right)^2 + \left(\frac{y_i - y_{track}}{pad_y} \right)^2 \right\} \qquad (47)$$

where x_i and x_{track} are respectively the hits and extrapolated x coordinates and pad_x is the logical pad size in direction x, for normalizing this distance to the granularity in that region of the muon detector. Then, the logarithm of the ratio between the two hypotheses

$$\Delta \log \frac{P_\mu}{P_{\not\mu}} \qquad (48)$$

can be used as discriminating variable for background rejection.

Momentum Range (GeV/c)	Muon Stations
$3 < p < 6$	M2,M3
$6 < p < 10$	M2,M3,M4 or M5
$p > 10$	M2,M3,M4,M5

Table 3: Muon stations required for a positive `IsMuon` decision as a function of momentum range.

- **ProbNN variables**: these are more powerful PID variables, because they combine information obtained separately from tracking, muon, RICH, and calorimeter systems. The best method found to combine this information consist of taking into account correlations among the different detector systems and by including additional information. This is carried out using multivariate techniques, combining PID information from each sub-system into a single probability value for each particle hypothesis.

Samples of decays which can be selected without using PID information for their relatively high statistics and low background (e.g. $J/\psi \to \mu^+\mu^-$) are essential for variable calibration, *ProbNN* multivariate analysis training and performance evaluation. In fact, the variables definition presented earlier are to be tuned properly using a test MC sample for taking into account differences arising when dealing with real detectors, like efficiency losses.

Variable	Requirement
Global Event Cut (GEC)	
SPD hits	< 600
Muon tracks	
track χ^2/ndf	< 5.0
ghost probability	< 0.4
min IPχ^2(PV)	> 9
Muon PID	
DLL$_{\mu\pi}$	> -3.0
`isMuon`	`true`
$\mu^+\mu^-$	
$m(\mu^+\mu^-)$	$< 7100.0\,\mathrm{MeV}/c$
Vertex χ^2/ndf	< 12.0
DIRA	> -0.9
PV separation χ^2	> 9.0
Λ_b	
$\lvert Q \rvert$	< 3
Mass	$4600 < \mathrm{M} < 7000\,\mathrm{MeV}/c$
Vertex χ^2/ndf	< 8
IPχ^2(related PV)	< 16
DIRA	> 0.9999
related PV separation χ^2	> 121.0

Table 4: `B2XMuMu` stripping line cuts. Here $|Q|$ is the particle charge equal to the sum of daughters charge. See Table 21 for variable definitions.

4 Stripping and pre-selection

This analysis is carried out using the full LHCb Run I data sample took in 2011 and 2012 and corresponding to an integrated luminosity of $3\,\mathrm{fb}^{-1}$. Candidates are reconstructed using the `B2XMuMu` stripping line (see Sec. 2.3.1 for stripping process description), whose requirements are detailed in Table 4. Table 5 lists the pre-selection cuts applied to stripped data for further reducing background events. These are quite standard cuts, apart from the one on $m(pK^-)$, which is placed because it excludes a lot of background events. Few signal events are expected above this cut since the $m(pK^-)$ spectrum observed in $\Lambda_b \to pK^- J/\psi$ channel [64] consists mainly of a main $\Lambda^*(1520)$ and many other Λ^* intermediate resonances, all with $m(pK^-)$ below the cho-

Quantity	Requirement
any track $\arcsin(p_{\mathrm{T}}/p)$	$\in [0, 0.4]$ rad
$prob$(DTF)	$> 10^{-5}$
$DTFm(pK^-\mu^+\mu^-)$	$\in [5350, 6000]\,\mathrm{MeV}/c^2$
$\mathrm{DLL}_{p\pi}(p)$	> -3
$\mathrm{DLL}_{pK}(p)$	> -9
$\mathrm{DLL}_{K\pi}(K)$	> -3
$\mathrm{DLL}_{Kp}(K)$	> -8
p,K^- tracks	`hasRich`
$m(pK^-)$	$> 2350\,\mathrm{MeV}/c^2$

Table 5: Preselection requirements applied to stripped data. See Table 21 for variable definitions.

Stage	Trigger Lines
L0	`L0Muon_TOS`
HLT1	`Hlt1TrackAllL0_TOS` `Hlt1TrackMuon_TOS`

HLT2	`Hlt2TopoMu[2,3,4]BodyBBDT_TOS`
	`Hlt2Topo[2,3,4]BodyBBDT_TOS`
	`Hlt2SingleMuon_TOS`
	`Hlt2DiMuonDetached_TOS`

Table 6: Trigger requirements for candidate events.

sen cut.
In addition standard trigger requirements are applied, listed in Table 6. Two categories of triggered events are usually defined according to the information required to trigger the event: TOS (Trigger On Signal) where signal particles are sufficient to trigger the event and TIS (Trigger Independent of Signal) where non-signal particles are sufficient to trigger the event. Note that a given event can belong to both categories (TIS&TOS) or to neither of the two. For this analysis only TOS trigger requirements are put. Most of these requirements are muon trigger lines (or combinations of more than one line) described in Sec. 3, `TrackAllL0` refers to any track selected by L0 and `Hlt2Topo` refers to HLT2 topological trigger lines, usually employed to trigger inclusively on n-body B decays [65].
$\Lambda_b \to pK^-J/\psi$ and $\Lambda_b \to pK^-\psi(2S)$ candidates are removed from $\Lambda_b \to pK^-\mu^+\mu^-$ sample by applying the standard LHCb veto for $c\bar{c}$ resonances on

the dimuon invariant mass $q^2 \equiv m(\mu^+\mu^-)^2$.

$$8 < q^2 < 11\,\mathrm{GeV}^2/c^4 \quad \texttt{and} \quad 12.5 < q^2 < 15\,\mathrm{GeV}^2/c^4$$

Candidate reconstruction is improved by using the `Decay Tree Fitter` (DTF) tool [66] which performs a kinematic simultaneous fit of the whole decay topology imposing suitable constraints. Hence, a better resolution on the Λ_b mass distribution is attained using DTF than the usual sequential decay reconstruction. DTF is set to constrain the Λ_b particle origin to the interaction point, while for $\Lambda_b \to pK^-J/\psi$ decays the J/ψ meson is not mass-constrained, for reconstructing these events in exactly the same way as $\Lambda_b \to pK^-\mu^+\mu^-$ decays.
MC events are selected exactly in the same way as data. For $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^-J/\psi$ MC samples are considered only truth-matched[6] events.

4.1 Peaking component vetoes

The dominant peaking components affecting this analysis comprise $B_s^0 \to K^+K^-\mu^+\mu^-$ and $B^0 \to K\pi\mu^+\mu^-$ decays with hadron mis-identifications, $\Lambda_b \to pK^-\mu^+\mu^-$ signal decays with a $p \leftrightarrow K$ swap and muons mis-identified as kaons, where the muon is from a J/ψ or $\psi(2S)$ resonance. For the B^0/B_s^0 peaking components, the entire event is refitted using DTF to the specific peaking component topology. The variables employed for setting B^0/B_s^0 vetoes are then calculated from the reconstruction corresponding to the best refitted B^0/B_s^0 topology. For MC studies the PID variables are corrected using `PIDCalib` package (see Appendix B for PID variable resampling).
No evidence of muons mis-identified as protons and of non-muonic peaking components (like $\Lambda_b \to pK^-\pi^+\pi^-$ with pions misidentified as muons) is seen. In the next subsections the vetoes set for excluding these peaking components from the signal dataset are detailed. All the invariant mass requirements are given in MeV/c^2 units.

4.1.1 $B_s^0 \to K^+K^-\mu^+\mu^-$ removal

The $B_s^0 \to K^+K^-\mu^+\mu^-$ background is dominated by the $\phi(1020) \to K^+K^-$ resonance, although some $f_2(1525) \to K^+K^-$ contribution is expected. Unfortunately, the proton PID variables are not discriminating enough to render an effective separation between signal and B_s^0 events. This is depicted

[6] A $\Lambda_b \to pK^-\mu^+\mu^-$ MC truth-matched event is a generated $\Lambda_b \to pK^-\mu^+\mu^-$ event. Truth-matching excludes generated background events reconstructed as $\Lambda_b \to pK^-\mu^+\mu^-$ events.

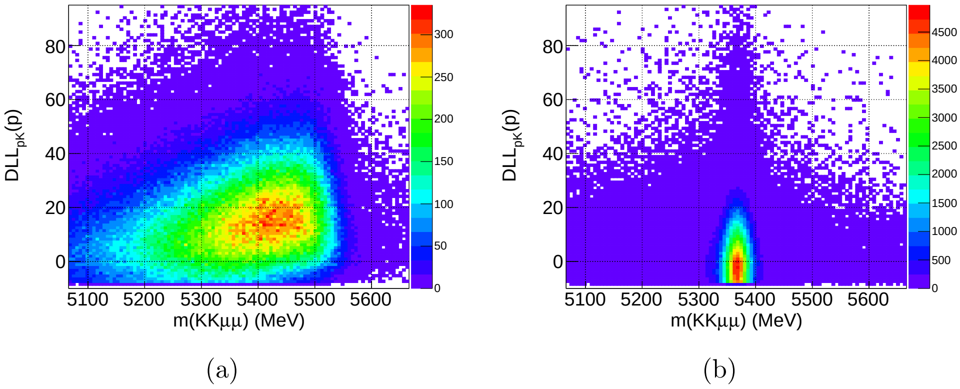


Figure 10: $B_s^0 \to K^+K^-\mu^+\mu^-$ removal: 2-D scatter plots of $\text{DLL}_{pK}(p)$ versus the reconstructed B_s^0 mass for (a) $\Lambda_b \to pK^-\mu^+\mu^-$ and (b) $B_s^0 \to \phi\mu^+\mu^-$ MC.

in Fig. 10, where the B_s^0 invariant mass is taken from the best reconstructed $B_s^0 \to K^+K^-\mu^+\mu^-$ candidate in each event. Relying on the fact that the narrow $\phi(1020)$ resonance is the main background component, candidates satisfying the following requirements are rejected

$$\begin{aligned}
&\{m(KK\mu\mu) \in [5300, 5420], m(KK) \in [1010, 1035]\} \texttt{ or}\\
&\{|m(KK\mu\mu) - 5366| < 40, m(KK) < 1080,\ \text{DLL}_{pK}(p) < 30\} \texttt{ or}\\
&\{|m(KK\mu\mu) - 5366| < 10, \text{DLL}_{pK}(p) < 0\}
\end{aligned} \tag{49}$$

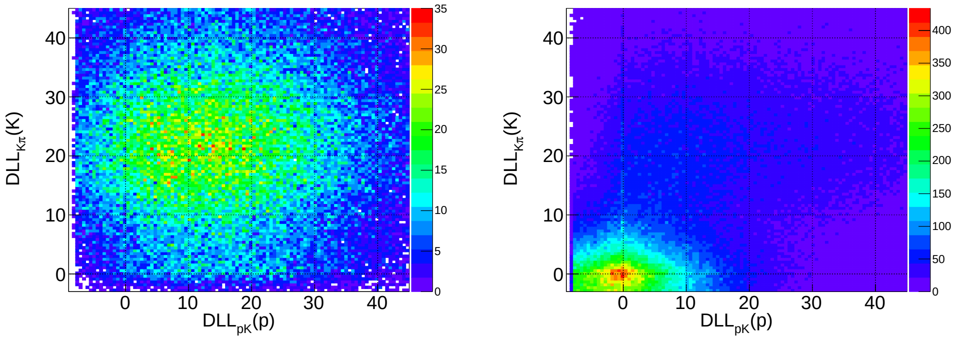
After these cuts, in the $\Lambda_b \to pK^-J/\psi$ mass spectrum some B_s^0 events not coming from $\phi(1020)$ resonance are still observed. No evidence of such events is seen in the $\Lambda_b \to pK^-\mu^+\mu^-$ mass spectrum (Sec. 8.3.4, Fig. 40). For this reason the following additional cuts are applied, only to $\Lambda_b \to pK^-J/\psi$ control channel

$$\begin{aligned}
&\{m(KK\mu\mu) \in [5300, 5420], m(KK) < 1300,\ \text{DLL}_{pK}(p) < 40\} \texttt{ or}\\
&\{|m(KK\mu\mu) - 5366| < 40, \text{DLL}_{pK}(p) < 20\}
\end{aligned} \tag{50}$$

4.1.2 $B^0 \to K\pi\mu^+\mu^-$ removal

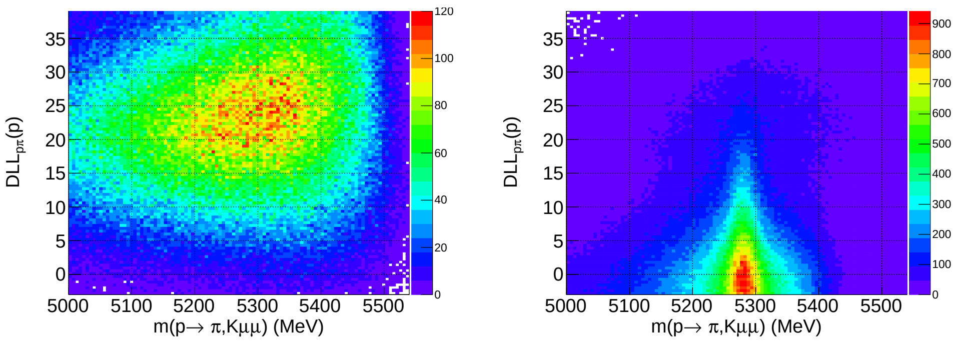
The $B^0 \to K\pi\mu^+\mu^-$ background is dominated by the broad $K^*(892) \to K\pi$ and $f_2(1430) \to K\pi$ resonances. Other possible partial wave contribution are even wider in terms of $m(K\pi)$ invariant mass. PID swaps can occur in two possible ways:

- Double swap $\{p \leftrightarrow K, K \leftrightarrow \pi\}$ candidates. Figure 11 shows the plot of $\text{DLL}_{K\pi}(K)$ versus $\text{DLL}_{pK}(p)$ for signal and B^0 simulated events, where



(a) (b)

Figure 11: $B^0 \to K\pi\mu^+\mu^-$ "double swap" removal: 2-D scatter plots of $\mathrm{DLL}_{K\pi}(K)$ versus $\mathrm{DLL}_{pK}(p)$ for (a) $\Lambda_b \to pK^-\mu^+\mu^-$ and (b) $B^0 \to K\pi\mu^+\mu^-$ MC. A cut has been placed on the reconstructed B^0 mass, $|m(p \to K, K \to \pi, \mu\mu) - 5279| < 45$.



(a) (b)

Figure 12: $B^0 \to K\pi\mu^+\mu^-$ "$p \leftrightarrow \pi$" swap removal: 2-D scatter plots of $\mathrm{DLL}_{p\pi}(p)$ versus $m(p \to K, K\mu\mu)$ for (a) $\Lambda_b \to pK^-\mu^+\mu^-$ and (b) $B^0 \to K\pi\mu^+\mu^-$ MC.

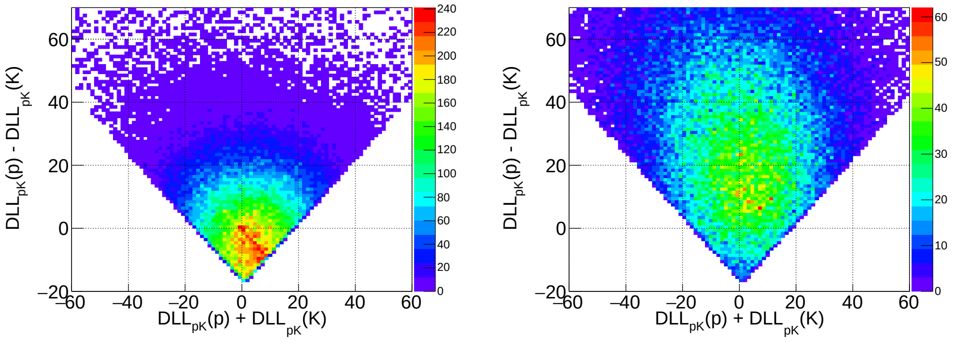


Figure 13: $p \leftrightarrow K$ swap removal: 2-D scatter plots of the PID variables for (a) non-truth-matched and (b) truth-matched MC $\Lambda_b \to pK^-\mu^+\mu^-$. A $|m(p \leftrightarrow K, \mu\mu) - 5620| < 45$ MeV selection has been placed.

the double-swapped B^0 mass $m_B \equiv m(p \to K, K \to \pi, \mu\mu)$[7] is required to satisfy $|m_B - 5279| < 45$. Defining $m_X \equiv m(p \to K, K \to \pi)$, candidates are rejected if

$$\begin{aligned}
&\{|m_X - 895| < 100, |m_B - 5279| < 45, \mathrm{DLL}_{K\pi}(K) < 5, \mathrm{DLL}_{pK}(p) < 15\} \texttt{ or} \\
&\{|m_X - 1430| < 100, |m_B - 5279| < 45, \mathrm{DLL}_{K\pi}(K) < 5, \mathrm{DLL}_{pK}(p) < 15\} \texttt{ or} \\
&\{|m_B - 5279| < 30, \mathrm{DLL}_{K\pi}(K) < 3, \mathrm{DLL}_{pK}(p) < 10\} \qquad (51)
\end{aligned}$$

- $p \leftrightarrow \pi$ swap candidates. Figure 12 shows the plot of $\mathrm{DLL}_{p\pi}(p)$ versus $m(p \to K, \mu\mu)$ for signal and B^0 MC events. Defining $m_B^1 \equiv m(p \to \pi, K\mu\mu)$ and $m_X^1 \equiv m(p \to \pi, K)$, candidates are rejected if

$$\begin{aligned}
\{|m_B^1 - 5279| < 50,\ \mathrm{DLL}_{p\pi}(p) < 4\} \texttt{ or} \\
\{|m_X^1 - 895| < 50,\ |m_B^1 - 5279| < 30,\ \mathrm{DLL}_{p\pi}(p) < 10\} \qquad (52)
\end{aligned}$$

4.1.3 $p \leftrightarrow K$ swap removal

The convenient variables to study $p \leftrightarrow K$ swaps are the PID variables $x \equiv \mathrm{DLL}_{pK}(p) + \mathrm{DLL}_{pK}(K)$, $y \equiv \mathrm{DLL}_{pK}(p) - \mathrm{DLL}_{pK}(K)$ and the Λ_b invariant mass calculated swapping proton and kaon mass hypothesis, $m(p \leftrightarrow K, \mu\mu)$. Figure 13 shows the 2-D scatter plots of these PID variables around

[7]This expression means that the invariant mass is computed taking the K mass hypothesis for the particle identified as a proton and the π mass for the particle identified as a K.

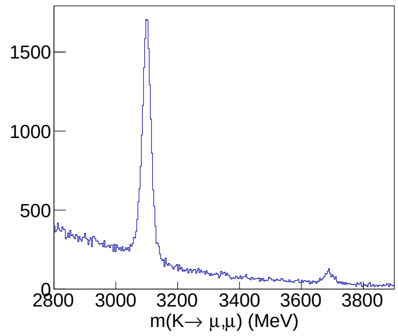


Figure 14: $\mu \leftrightarrow K$ removal: $m(K \to \mu, \mu)$ spectrum from data with "K`isMuon or` $\mathrm{DLL}_{\mu\pi}(K) > -2$" requirements on the kaon. The J/ψ and $\psi(2S)$ peaks are clearly visible.

the Λ_b region $|m(p \leftrightarrow K, \mu\mu) - 5620| < 45$. The $p \leftrightarrow K$ swaps are subsequently removed by rejecting candidates satisfying

$$\begin{aligned}\{\sqrt{x^2+y^2} < 15, |m(p \leftrightarrow K, \mu\mu) - 5620| < 20\} \texttt{ or} \\ \{\sqrt{x^2+y^2} < 7, |m(p \leftrightarrow K, \mu\mu) - 5620| < 40\} \texttt{ or} \\ \{y < 0, |m(p \leftrightarrow K, \mu\mu) - 5620| < 45\}\end{aligned} \tag{53}$$

4.1.4 $\mu \leftrightarrow K$ swap removal

Since charmonium resonances are very copiously produced in LHCb, it is not uncommon for a true muon from a J/ψ or $\psi(2S)$ resonance to be misidentified as a kaon. Figure 14 shows the invariant mass $m(K \to \mu, \mu)$ directly from data, where charmonium resonances are clearly visible. Swaps are removed by rejecting candidates satisfying the further criteria

$$|m(K \to \mu, \mu) - 3096| < 100 \texttt{ or } |m(K \to \mu, \mu) - 3686| < 60 \tag{54}$$

4.1.5 Signal loss

The signal loss due to the applied veto cuts is checked on MC $\Lambda_b \to pK^-\mu^+\mu^-$ events by counting how many signal events are discarded when applying a specific veto (Table 7). As done before, only resampled PID variables are considered.

Veto	Λ_b Signal loss [%]
$B_s^0 \to K^+K^-\mu^+\mu^-$	1.6
$B^0 \to K\pi\mu^+\mu^-$ $p \leftrightarrow \pi$ swap	0.6
$B^0 \to K\pi\mu^+\mu^-$ double swap	0.3
$K \leftrightarrow p$ mis-ID	1.7
$K \leftrightarrow \mu$ mis-ID	0.9

Table 7: The percentage signal loss after application of each peaking component veto cut.

5 Fit model

For fitting in a suitable way the Λ_b mass distribution it is mandatory to understand how signal and background shapes can be parametrized. In similar analyses the signal shape is usually parametrized using one or more Crystal Ball (CB) probability density function (PDF) [67,68]. The CB PDF is made of a Gaussian PDF connected with a one power-law tail, so as to be a $\mathcal{C}^1$ function[8]:

$$CB(x;\overline{x},\sigma,\alpha,n) = N \begin{cases} \exp\left(\frac{(x-\overline{x})^2}{2\sigma^2}\right) & \text{for } \alpha > 0, \frac{(x-\overline{x})}{\sigma} > -\alpha \\ & \text{for } \alpha < 0, \frac{(x-\overline{x})}{\sigma} < -\alpha \\ A\left(B - \frac{(x-\overline{x})}{\sigma}\right)^{-n} & \text{for } \alpha > 0, \frac{(x-\overline{x})}{\sigma} < -\alpha \\ & \text{for } \alpha < 0, \frac{(x-\overline{x})}{\sigma} < -\alpha \end{cases} \tag{55}$$

where $A = \left(\frac{n}{|\alpha|}\right)^n \exp\left(-\frac{|\alpha^2|}{2}\right)$, $B = \frac{n}{|\alpha|} - |\alpha|$ and N is the normalization factor. Here, $\overline{x}$ and σ are the usual Gaussian PDF[9] parameters, α sets the power-law tail threshold (for $\alpha > 0$ the tail is on the low-mass side and vice versa) and n controls the power exponent.
The CB PDF is particularly well suited to parametrize radiative decays like the Λ_b decay under study, characterized by power-law tails in the mother particle mass distribution.
The combinatorial background, consisting of candidates where one or more tracks are paired randomly, is fitted with an exponential PDF, as done in similar analyses (*e.g.* [35, 64, 69]).
After a careful study of $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^-J/\psi$ MC samples, the signal shape has been parametrized by means of a PDF composed by 2 CB and one Gaussian functions. For these shapes, all the PDF components share

[8]Derivable function with continuous derivative.
[9]For CB function $\overline{x}$ is the most probable value, not the mean value.

the same mean and one CB has the radiative tail on the high mass side ($\alpha < 0$), allowing to take into account events whose mass is not properly measured. Tail parameters α and n of the two CB, relative fractions and relative widths are then fixed to the ones determined by fitting the Λ_b mass distribution of the MC data, differently for $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^- J/\psi$ decays (Fig. 15, Table 8). The composition of m PDFs $F(x)$ is normalized to the total fitted number of events N_{tot} as

$$F(x) = N_1 \mathrm{PDF}_1(x) + N_2 \mathrm{PDF}_2(x) + ... + (N_{tot} - \sum_{i=1}^{m-1} N_i)\mathrm{PDF}_m \tag{56}$$

where N_i is the number of events associated to the component PDF_i. The relative fraction for the component PDF_i is then defined as N_i/N_{tot}, while the relative width is computed with respect to the first CB component as $\sigma(\mathrm{PDF}_i)/\sigma(\mathrm{CB1})$. Therefore, when fitting data, the only floating parameters will be the normalization, the common mean and one overall width[10], this for considering possible resolution differences between real and simulated events. Fits are performed using the unbinned extended maximum likelihood method, consisting in maximizing the likelihood defined as

$$\mathcal{L}(\nu, \vec{a}) = \exp(-\nu)\frac{\nu^n}{n!}\prod_{i=1}^{n} p(x_i, \vec{a}) \tag{57}$$

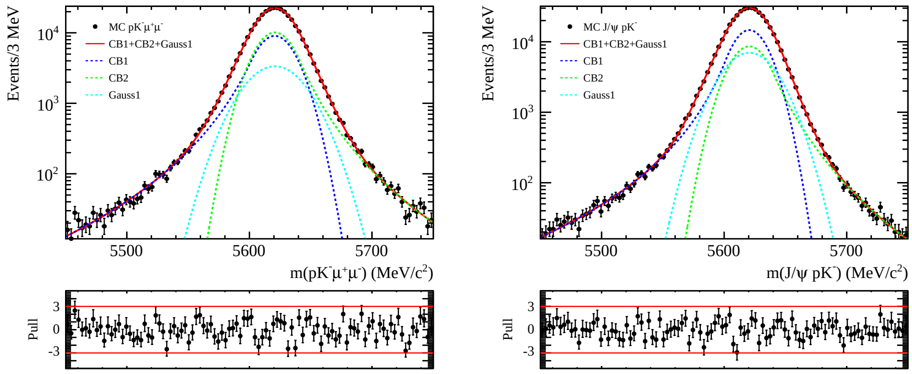
over the PDF free parameters $\vec{a}$ and the fitted number of events ν. Here x_i are the n events to be fitted and $p(x_i, \vec{a})$ is the fitting PDF evaluated at given $(x_i, \vec{a})$. Usually the likelihood maximization is done by minimizing $-\log \mathcal{L}$, which is a more manageable expression since the multiplication turns into a sum

$$-\log \mathcal{L}(\nu, \vec{a}) = -\sum_{i=1}^{n} \log p(x_i, \vec{a}) - n \log \nu + \nu + \mathrm{const.} \tag{58}$$

The extended feature consists in determining the number of fitted events ν by adding a Poissonian distribution, with n as mean number of events, to the likelihood. In fact, for repeated experiments the number of events will fluctuate according to Poisson statistics. The employed method is unbinned as no histogram is used during the maximum likelihood fitting. The histograms reported in the following sections are used for graphical reasons, where the plotted fitted PDFs are suitably normalized.

The fit quality is checked by reporting the distribution of the pulls, defined

[10] In practice the first CB width is left floating, while the other widths are linked to the floating one using the fixed ratio derived from MC.



(a) (b)

Figure 15: Fit to Montecarlo $\Lambda_b \to pK^-\mu^+\mu^-$ (a) and $\Lambda_b \to pK^-J/\psi$ (b) signal samples. Below the fit plots, the pulls between data points and fitted function are reported.

for each histogram bin as the difference between the bin value and the fitted function value, difference then divided by the associated bin error.

6 Selection

A selection for rejecting combinatorial background events is made by using a multivariate classifier (MVA) based on a *Boosted Decision Tree* (BDT) [70,71] implemented as one method of the *Toolkit for Multivariate Data Analysis* (TMVA) [72], a package developed towards high energy physics (HEP) tasks. Such a classifier learns how to distinguish signal events from background events by exploiting differences between signal and background PDF distributions of a set of chosen variables. During this stage, called training, a discriminating variable (BDT variable), interpretable as a measure of signal-likeness, is built. The more signal and background variable distributions are different, the more the classifier will be able to separate signal and background events.

BDT is one of the two commonly used classifiers in HEP analyses, together with artificial neural networks (NN). For this analysis BDT is chosen because it is a simpler method than neural networks and its performance is stable with respect to training variable number, separation and correlations among them. For example, addition of a poorly discriminating variable to the training set do not decrease BDT performance, rather may further improve it. In

Variable	MC $\Lambda_b \to pK^-\mu^+\mu^-$ Fit Value	Error	MC $\Lambda_b \to pK^-J/\psi$ Fit Value	Error
CB1 Events (10^3)	208.4	$\pm$ 6.2	138	$\pm$ 18
CB2 Events (10^3)	185.6	$\pm$ 7.9	103	$\pm$ 20
Gauss1 Events (10^3)	99.2	$\pm$ 7.0	99.4	$\pm$ 8.8
Mean (MeV/c^2)	5620.70	$\pm$ 0.04	5620.70	$\pm$ 0.04
CB1 σ (MeV/c^2)	14.87	$\pm$ 0.17	13.58	$\pm$ 0.19
CB2 σ (MeV/c^2)	15.04	$\pm$ 0.17	14.96	$\pm$ 0.51
Gauss1 σ (MeV/c^2)	21.56	$\pm$ 0.43	20.11	$\pm$ 0.48
CB1 α	1.322	$\pm$ 0.026	1.469	$\pm$ 0.086
CB2 α	-1.330	$\pm$ 0.045	-1.47	$\pm$ 0.15
CB1 n	3.67	$\pm$ 0.13	2.88	$\pm$ 0.18
CB2 n	4.23	$\pm$ 0.28	4.03	$\pm$ 0.70

Table 8: Fit results for MC $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^-J/\psi$ signal samples.

general, this is not true for neural networks. Moreover, neural networks are far more resource consuming and difficult to tune because of their several free parameters.
Once BDT variable is produced, the cut value is chosen by optimizing a proper figure of merit (FoM). The best sensitivity to CP asymmetry (minimization of statistical uncertainties) is usually achieved by optimizing the significance FoM defined as

$$\mathrm{FoM} = \frac{S}{\sqrt{S+B}} \tag{59}$$

Here S stands for the number of signal events after BDT cut, B is the number of background events after BDT cut in the chosen signal region $m(\Lambda_b) \in [5500, 5700]\,\mathrm{MeV}/c^2$. They are computed rescaling the number of signal and background events before BDT selection with signal and background efficiencies obtained applying the BDT cut to the training samples. This way, the optimization process is not biased by the statistical fluctuations of the data sample. The expected number of $\Lambda_b \to pK^-\mu^+\mu^-$ events before selection is taken as the fitted number of $\Lambda_b \to pK^-J/\psi$ events after preselection, rescaled by the theoretical branching fraction (derived in Appendix A, Eq. 77)

$$\frac{\mathcal{B}(\Lambda_b \to pK^-\mu^+\mu^-)}{\mathcal{B}(\Lambda_b \to pK^-J/\psi(\to \mu^+\mu^-))} = 1.57\%$$

The number of background events is derived by counting the number of $\Lambda_b \to pK^-\mu^+\mu^-$ events in the signal region before selection minus the expected $\Lambda_b \to pK^-\mu^+\mu^-$ signal events.
As a cross check, an alternative BDT cut is chosen by optimizing the Punzi FoM [73]

$$\text{FoM} = \frac{\epsilon_s}{\sqrt{B} + \frac{a}{2}} \tag{60}$$

devised for gaining the best evidence for rare decays. Here ϵ_s is the BDT signal efficiency, B is the number of background events, as before, and a is the significance w.r.t the null hypothesis in (Gaussian equivalent) sigma units that one would like to achieve: to obtain an evidence for a certain decay a significance of 3 σ is required (chosen), while a significance of 5 σ is required for a decay observation. This FoM needs no estimate of the signal yield before BDT cut.
The selection has been prepared first for the control sample $\Lambda_b \to pK^-J/\psi$, because for this decay it is possible to better assess possible issues by comparing the MC $\Lambda_b \to pK^-J/\psi$ variable distributions to the data signal distributions obtained using the *sPlot* statistical tool [74]. This method allow to extract the signal distributions by reweighting each candidate with a weight, called *s-Weight*, derived starting from a data fit.
J/ψ particle is not mass-constrained for keeping the same reconstruction procedure as for $\Lambda_b \to pK^-\mu^+\mu^-$ decay.
For the training stage a Montecarlo data sample of $\Lambda_b \to pK^-J/\psi$ events is used as signal sample and the $\Lambda_b \to pK^-J/\psi$ data right sideband $m(\Lambda_b) > 5800\,\text{MeV}$ as background sample. All preselection and peaking component veto cuts are applied before developing the BDT selection.
Several variable types have been considered as possible training variables: kinematic, topological and different classes of isolation variables (related to extra tracks that can be added to the decay chain). Use of particle identification variables is investigated later. The discriminating power of each variable is measured as the statistical separation $\langle S^2 \rangle$ between signal and background PDFs, defined as

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\text{PDF}_S(y) - \text{PDF}_B(y))^2}{\text{PDF}_S(y) + \text{PDF}_B(y)} dy \tag{61}$$

where $\text{PDF}_{S,B}$ are the signal and background PDF of the variable y under study.
Variables with higher signal/background statistical separation have been chosen as training variables, taking into account the correlations among them[11].

[11] Obviously, given a set of variables with fixed statistical separation, the resulting BDT discriminating power will be maximum when variables are uncorrelated, hence the effective

These are (see Appendix D for variable definition, X refers to the pK^- system):

- Kinematic variables: $p_{\mathrm{T}}(X)$, $p(X)$, $p_{\mathrm{T}}(\Lambda_b)$
- Topological variables: $\log(\chi^2_{\mathrm{IP}}(\Lambda_b, K))$, $\log\arccos(\mathrm{DIRA}(\Lambda_b))$, $\tau_S(\Lambda_b)$, $prob(\mathrm{DTF})$
- Isolation Variables (see Sec. 6.1)

Next, delta log likelihood PID variables (defined in Table 21) are considered for the training stage. Only the hadronic $\mathrm{DLL}_{p\pi}(p)$, $\mathrm{DLL}_{K\pi}(K)$, $\mathrm{DLL}_{K\pi}(p)$, $\mathrm{DLL}_{p\pi}(K)$ are added to the training variable set. DLL regarding muons are discarded because of their very low separating power and this is a sign of the effectiveness of muon identification in LHCb experiment. All PID variables of MC samples are resampled using the `PIDCalib` packages for correcting discrepancies between data and MC PID distributions (Appendix B). DLLs have been exploited as BDT training variables also by [75, 76] analyses.
Differences observed between MC and *s-Weighted* variable distributions are corrected by reweighting the MC $\Lambda_b \to pK^- J/\psi$ sample (Appendix C). This correction is very important because small data/MC discrepancies in training variable distributions can result in a sizeable difference between data/MC BDT distributions, that would lead to a wrong estimation of the signal efficiency and in turn a wrong optimization process.
Finally, also ProbNN PID variables (defined in Table 21) are considered as training variables, but even though they showed a good data/MC agreement after PID resampling (Fig. 42) the resulting BDT distributions for MC and *s-Weighted* $\Lambda_b \to pK^- J/\psi$ samples are still quite different (Fig. 16). This problem is observed for both $v2$ and $v3$ ProbNN versions and it is not understood yet.
Once the training variable set is selected, different BDT training algorithms are tested. The *Adaptive boosting* algorithm [77] is chosen as this showed the best discriminating performance.
A testing stage is done after the training one, where the obtained BDT is applied to another set of signal and background events, different from the training one. Compatibility of the BDT distributions obtained for training and testing datasets is checked using the Kolmogorov-Smirnov test, which returns the probability that the two distributions are compatible within statistical fluctuations. Failure of this test (probability $< 5\%$) would imply the BDT classifier to be too specific for the samples used for building it (overtraining), therefore not applicable to other datasets. BDT booking options

separating power of a variable is decreased by its correlations with other variables.

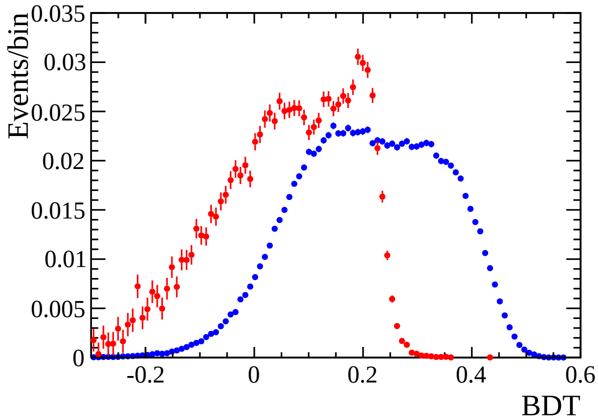


Figure 16: Comparison between MC (blue) and *s-Weighted* (red) $\Lambda_b \to pK^- J/\psi$ BDT normalized distributions for a test BDT employing ProbNN $v3$ in place of DLL PID variables.

are carefully chosen as a compromise between discrimination increase and overtraining control.
BDT selection for $\Lambda_b \to pK^-\mu^+\mu^-$ candidates is set as for $\Lambda_b \to pK^- J/\psi$ channel, using 160000 events of MC $\Lambda_b \to pK^-\mu^+\mu^-$ reweighted using the weights derived for MC $\Lambda_b \to pK^- J/\psi$ sample as signal sample and 160000 events of $\Lambda_b \to pK^-\mu^+\mu^-$ data right sideband $m(\Lambda_b) > 5800\,\text{MeV}$ as background sample. This way, it is assumed that MC reweighting corrects the Λ_b production effects for $\Lambda_b \to pK^-\mu^+\mu^-$ events and that it is independent of the particular Λ_b decay process. No data right sideband event is later used for *CP*-violating observables extraction, whose relative fits are done for $m(\Lambda_b) \in [5400, 5800]\,\text{MeV}/c^2$.
The testing stage is performed on different signal and background samples of the same number of events used for the training stage. Overtraining check is reasonably good, reported in Fig. 17. Fig. 18 shows that the data/MC comparison for the BDT variable is fine.
Training variable distributions for signal and background samples are compared in Fig. 19, their statistical separation is given in Table 9. In Fig. 20 is shown the signal/background comparison for ProbNN $v3$ variables, their statistical separation, reported in Table 10, is larger than the corresponding DLL variables.
Correlations of BDT variable with reconstructed Λ_b mass and C_T triple-product have been checked to be low enough to exclude a biasing effect on the latter variables (Fig. 21). BDT linear correlations are found to be -0.3% with $m(\Lambda_b)$ and 1.7% with C_T.

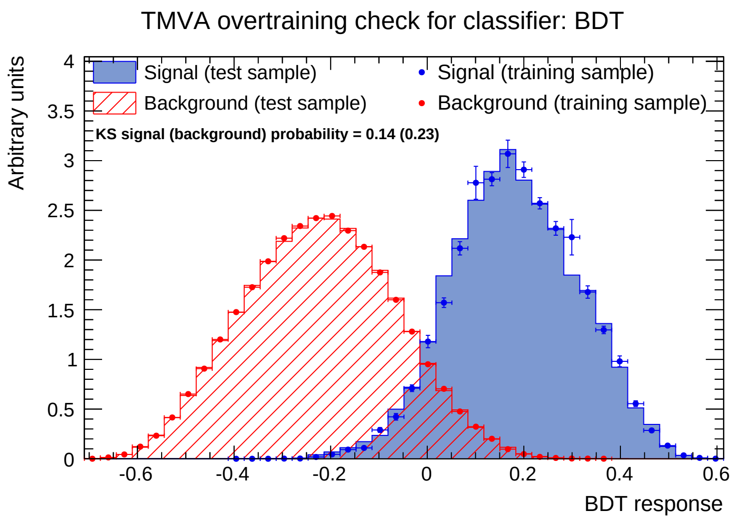


Figure 17: Overtraining check for BDT classifier.

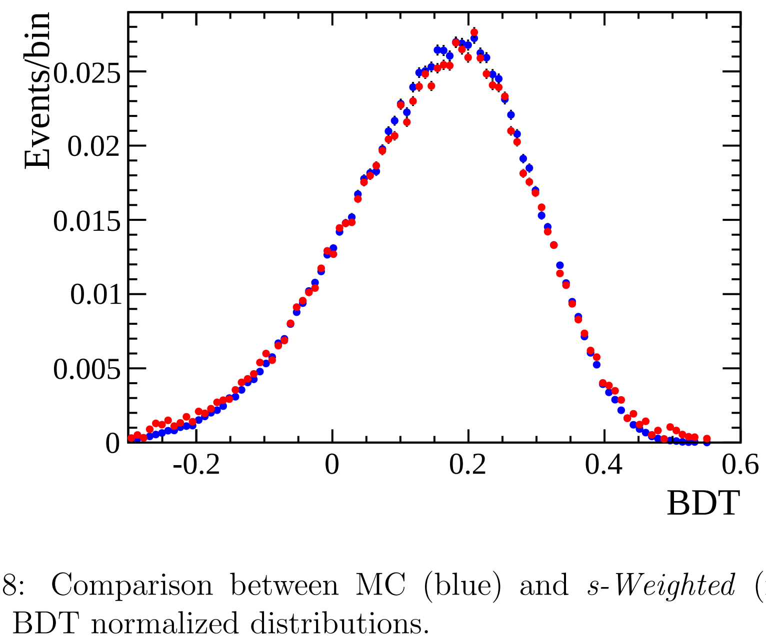


Figure 18: Comparison between MC (blue) and *s-Weighted* (red) $\Lambda_b \to pK^- J/\psi$ BDT normalized distributions.

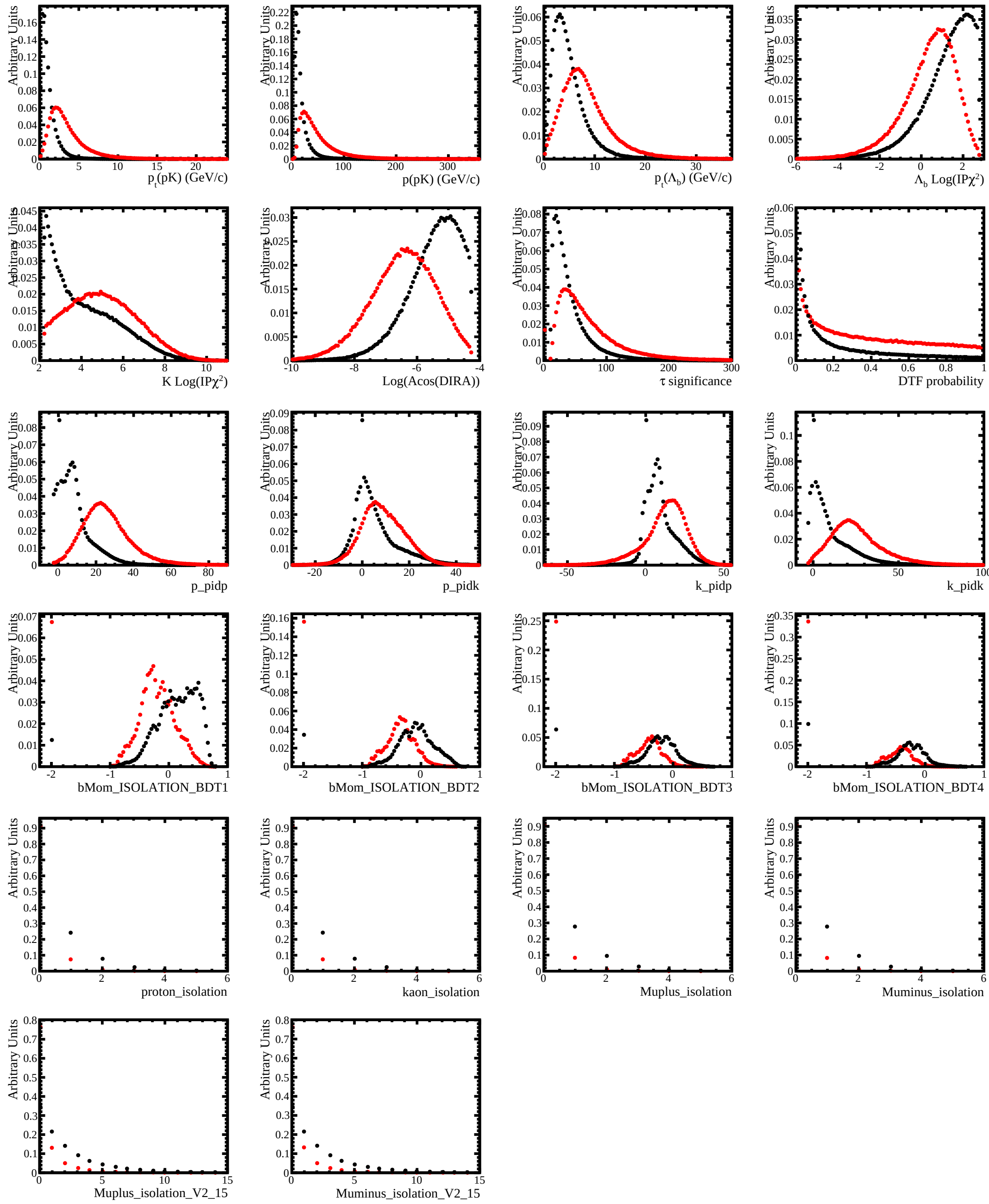


Figure 19: Comparison between reweighted MC $\Lambda_b \to pK^-J/\psi$ (red) and data right sideband (black) normalized distributions for MVA training variables.

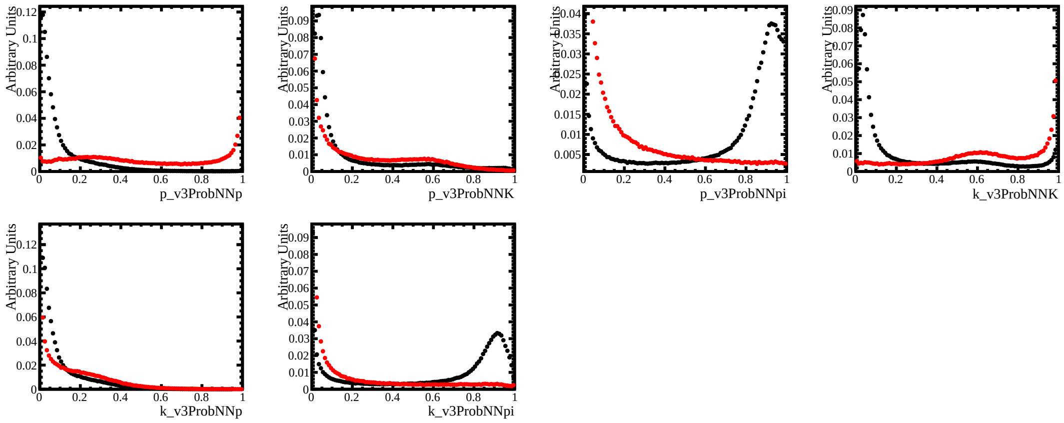


Figure 20: Comparison between reweighted MC $\Lambda_b \to pK^- J/\psi$ (red) and data right sideband (black) normalized distributions for ProbNN $v3$ variables.

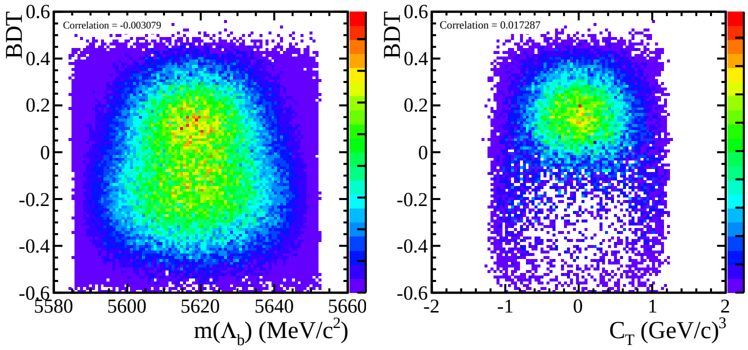


Figure 21: BDT variable scatter plots with reconstructed Λ_b mass and C_T.

Rank	Variable	$\langle S^2 \rangle$ (%)
1	$\mathrm{DLL}_{p\pi}(p)$	43.56
2	$p_{\mathrm{T}}(X)$	37.28
3	$p(X)$	36.70
4	$\mathrm{DLL}_{K\pi}(K)$	26.70
5	$prob(\mathrm{DTF})$	21.52
6	Isolation_BDT1	20.45
7	$\log\arccos(\mathrm{DIRA}(\Lambda_b))$	17.78
8	$\mathrm{DLL}_{p\pi}(K)$	17.53
9	Isolation_BDT2	17.20
10	$\log(\chi^2_{\mathrm{IP}}(K))$	14.41
11	Isolation_BDT3	14.13
12	$\tau_S(\Lambda_b)$	13.91
13	$\log(\chi^2_{\mathrm{IP}}(\Lambda_b))$	13.21
14	Isolation_BDT4	12.99
15	$\mathrm{DLL}_{K\pi}(p)$	12.87
16	Isolation_V2_15(μ^+)	12.85
17	Isolation_V2_15(μ^-)	12.79
18	Isolation(μ^-)	8.351
19	Isolation(μ^+)	7.918
20	$p_{\mathrm{T}}(\Lambda_b)$	7.597
21	Isolation(p)	6.644
22	Isolation(K^-)	6.539

Table 9: MVA training variables ranked according to their statistical separation.

Rank	Variable	$\langle S^2 \rangle$ (%)
1	$\mathrm{ProbNN}_p(p)$	46.38
2	$\mathrm{ProbNN}_\pi(p)$	35.24
3	$\mathrm{ProbNN}_K(K)$	28.73
4	$\mathrm{ProbNN}_\pi(K)$	27.98
5	$\mathrm{ProbNN}_K(p)$	5.294
6	$\mathrm{ProbNN}_p(K)$	4.317

Table 10: ProbNN variables ranked according to their statistical separation.

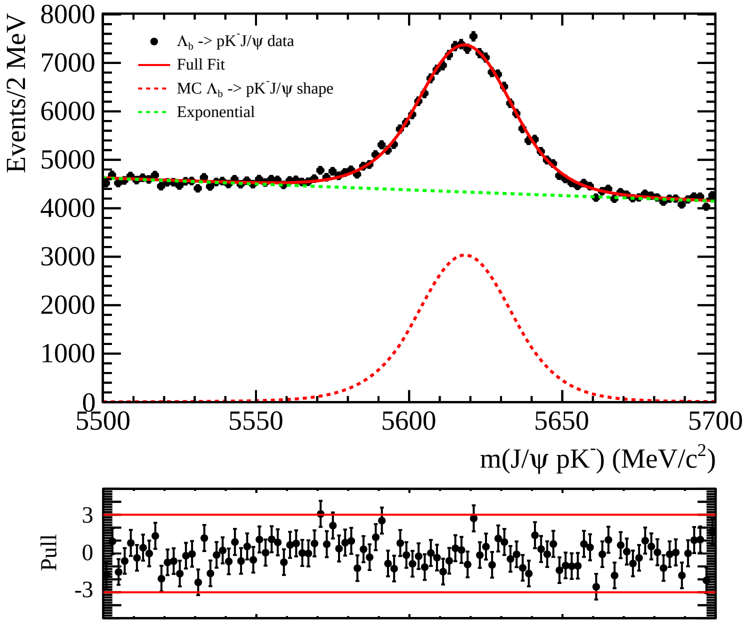
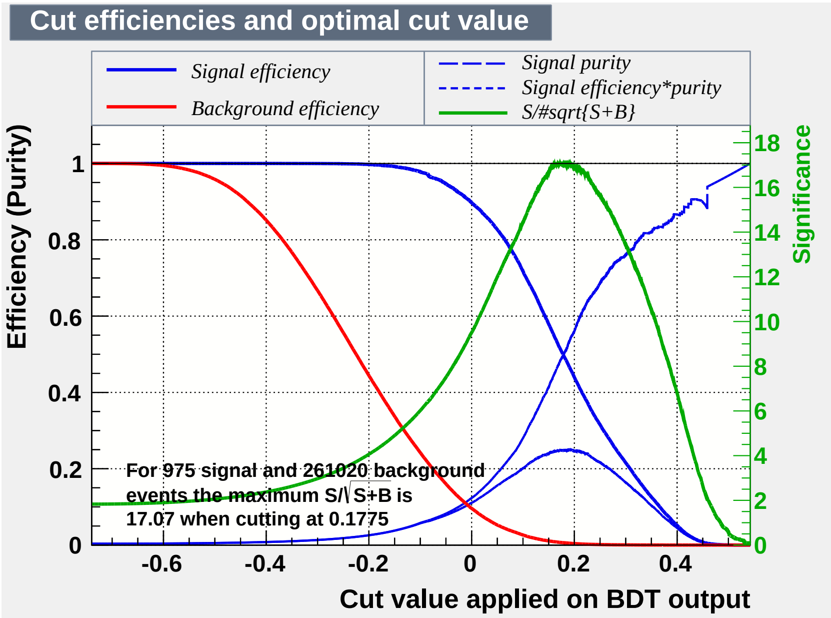


Figure 22: Fit to $\Lambda_b \to pK^- J/\psi$ data before BDT selection. Pulls defined as in Sec. 5.

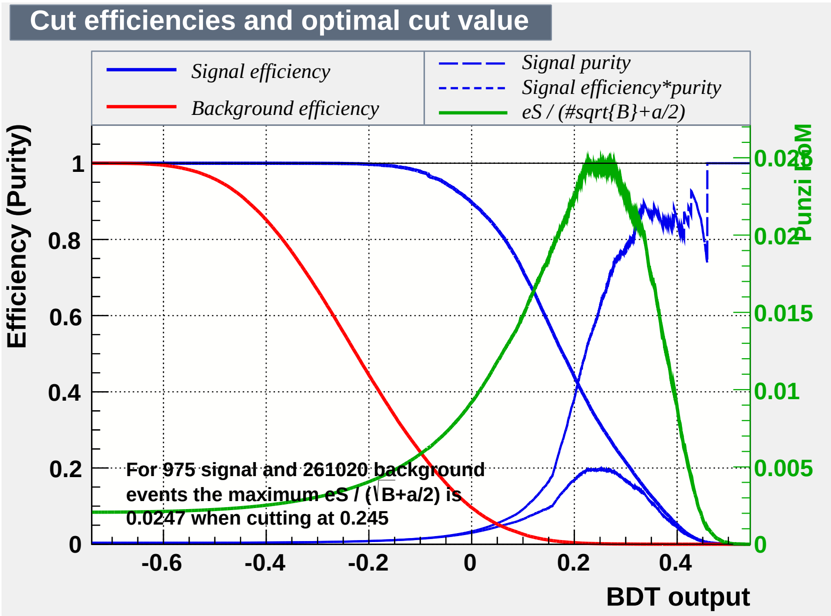
For what concerns the optimization process, the fit to $\Lambda_b \to pK^- J/\psi$ data before BDT selection and relative fit results are showed respectively in Fig. 22 and Table 11. Starting from $\Lambda_b \to pK^- J/\psi$ yield, the expected number of $\Lambda_b \to pK^- \mu^+ \mu^-$ events is 975, while the expected background events in the signal region is 261020.
The significance optimization is reported in Fig. 23(a), suggesting a cut on the BDT variable at 0.1775 with a maximum expected significance of 17.07. The alternative Punzi FoM optimization is showed in Fig. 23(b) suggesting a tighter cut at 0.245.

Variable	Fitted value	Error
Signal Events (10^3)	62.13	$\pm$ 0.73
BKG Events (10^3)	438.16	$\pm$ 0.95
Exponential c (10^{-4})	-5.46	$\pm$ 0.28
Mean	5618.20	$\pm$ 0.16
CB1 σ	13.96	$\pm$ 0.17

Table 11: Fit results for $\Lambda_b \to pK^- J/\psi$ data fit before BDT selection.



(a)



(b)

Figure 23: Significance (a) and Punzi (b) optimizations. The oscillations visible for high values of BDT cut are a consequence of the finite size of the training samples, which makes the computed background efficiency to be discontinuous. This effect does not compromise the optimization procedure since the chosen BDT cuts lie outside that BDT region.

6.1 Isolation variables

Many background events come from decays where additional charged tracks are not reconstructed. These tracks will form good vertices with such $\Lambda_b \to pK^-\mu^+\mu^-$ candidates, while for a true signal candidate it should not be possible to add extra tracks to the decay topology. Isolation variables are built for distinguishing this kind of background and therefore they are good signal/background discriminators. Use of isolation variables for BDT training is useful because, being not correlated with kinematic, topological or PID variables, they provide a different S/B discriminating information. Moreover they are not correlated with $m(\Lambda_b)$ or C_T triple-product. The following isolation variables, exploited by [69, 76] analyses, are added to the BDT training variable set.

- **Hadron isolation** [Isolation(p, K^-)]: These variables, built for [76] selection, consider the extra tracks (not belonging to the decay looked for) forming a vertex with the hadron track. For each extra track the following quantities are taken into account:
  - Minimum distance between the extra track and the PV (pvdist).
  - Minimum distance between the extra track and the Λ_b vertex (svdist).
  - The *distance of closest approach* (DOCA) between the extra track and the hadron one.
  - Extra track IP significance ($\mathrm{IP}/\mathrm{IP}_{\mathrm{error}}$).
  - Angle between the extra track and the hadron one (β).
  - The quantity defined as:

$$f_c = \frac{|\vec{p}_{\mathrm{h}} + \vec{p}_{\mathrm{trk}}|\alpha^{\mathrm{h+trk,PV}}}{|\vec{p}_{\mathrm{h}} + \vec{p}_{\mathrm{trk}}|\alpha^{\mathrm{h+trk,PV}} + p_{\mathrm{T,h}} + p_{\mathrm{T,trk}}} \tag{62}$$

    where $\alpha^{\mathrm{h+trk,PV}}$ is the angle between the hadron and the track candidate, $p_{\mathrm{T,h}}$ and $p_{\mathrm{T,trk}}$ are the hadron and extra track transverse momenta with respect to the beam line.

  Hadron isolation variables are built counting how many extra tracks satisfy the following conditions:

  - pvdist $\in [0.5, 40]$ mm
  - svdist $\in [-0.15, 30]$ mm
  - DOCA < 0.13 mm

- Track IP significance > 3
- $\beta < 0.27$
- $f_c < 0.6$

- **Muon isolation** [Isolation(μ^+, μ^-)]: these variables are built like the hadron isolation ones (still for [76] selection) but employ a multivariate approach: the six variables considered before for hadron tracks plus extra track IP, extra track p_{T} and extra track χ^2/ndf are fed into a BDT which is trained using a phase-space $B^0 \to K^*\mu^+\mu^-$ MC sample and an inclusive dimuon background sample. For this analysis two different variables for each muon track corresponding to different BDT training versions are added to the MVA selection.

- **BDT isolation** [Isolation_BDT_$(1-4)$]: these are four different BDT trainings considering the whole decay topology for selecting additional extra tracks, developed for [69] selection. These BDTs include extra track p_{T}, the angle between $\vec{p}_{\mathrm{track}}$ and $\vec{p}_{p\mu}$, extra track ghost probability, extra track χ^2 and extra track χ^2_{IP} with respect to the $p\mu$ vertex. Behaviour of these BDT trainings on our signal and background samples changes gradually from Isolation_BDT_1 which rejects more background events to Isolation_BDT_4 which has an higher signal efficiency (Fig. 19), for this reason they are all included in our BDT selection.

6.2 Comparison of different selections

The goal of this section is to show that our MVA selection strategy, that is including both isolation and PID as BDT training variables, is better in terms of attained significance FoM (Eq. 59) with respect to other strategies, used in other analyses, which do not include isolation variables or they apply simple PID cuts. To this end four selection strategies are tested (variable classes refer to Sec. 6):

- **BDT1**: Trained employing kinematic and topological variables.
- **BDT2**: Trained employing kinematic, topological and isolation variables.
- **BDT1+C**: Same as BDT1, but followed by cuts on hadronic ProbNN PID variables.
- **BDT3**: Our selection (Sec. 6), namely trained employing kinematic, topological, isolation and hadronic DLL PID variables.

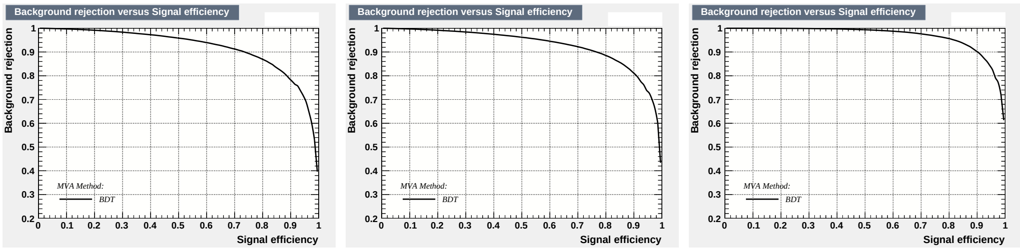


Figure 24: ROC curves for (left) BDT1, (center) BDT2 and (right) BDT3. The squared grid highlights how the BDT rejection power increases from BDT1 to BDT3.

BDT	1	2	1+C	3
ϵ_{sig} (%)	36	45	27	41
ϵ_{bkg} (10^{-4})	35	52	4.7	8.7
N_{sig}	350	440	263	400
Significance FoM	9.79	10.35	13.41	15.96
Punzi FoM	0.0110	0.0117	0.0215	0.0247
Purity (%)	27	25	68	64

Table 12: Performance comparison of the four considered selection strategies.

All the BDT configurations are trained on the same signal and background samples with same BDT booking options and are subsequently optimized using the Punzi FoM[12] (Eq. 60 with $a = 3$). The ROC curves (signal efficiency vs background rejection plots) for the three BDTs are reported in Fig. 24. Performance comparison of the four selections is shown in Table 12. According to significance FoM (minimizing uncertainties on CPV observables) our selection clearly outperforms the other considered strategies.

7 Extraction of *CP*-violating observables

7.1 Direct *CP* asymmetry

Data samples after selection are split into Λ_b and $\overline{\Lambda}_b$ subsamples according to the kaon charge. An unbinned extended simultaneous maximum likelihood fit to the Λ_b mass distribution of both subsamples is performed for extracting the raw *CP* asymmetry $\mathcal{A}_{\text{raw}}$. The simultaneous fit is done using the fit model depicted in Sec. 5, where the signal and background shape free parameters[13] are constrained to be equal for each of the two subsamples. The raw asymmetry is extracted directly from the fit model by expressing the $\overline{\Lambda}_b$ yield as (Eq. 31)

$$N_{\overline{\Lambda}_b} = N_{\Lambda_b} \frac{1 - \mathcal{A}_{\text{raw}}}{1 + \mathcal{A}_{\text{raw}}} \tag{63}$$

while background yields are estimated separately for each subsample.
The minimized log-likelihood corresponding to this fit procedure is

$$\begin{aligned}
-\log \mathcal{L}(\overline{x}, \sigma, c, N_{\Lambda_b}, \mathcal{A}_{\text{raw}}, N_{\text{exp},\Lambda_b}, N_{\text{exp},\overline{\Lambda}_b}) = \\
& -\sum_{i=1}^{n_{\Lambda_b}} \log\left(\frac{N_{\Lambda_b}}{N_{\Lambda_b} + N_{\text{exp},\Lambda_b}} S(x_i, \overline{x}, \sigma) + \frac{N_{\text{exp},\Lambda_b}}{N_{\Lambda_b} + N_{\text{exp},\Lambda_b}} B(x_i, c)\right) \\
& - n_{\Lambda_b} \log(N_{\Lambda_b} + N_{\text{exp},\Lambda_b}) + N_{\Lambda_b} + N_{\text{exp},\Lambda_b} \\
& -\sum_{i=1}^{n_{\overline{\Lambda}_b}} \log\left(\frac{N_{\overline{\Lambda}_b}}{N_{\overline{\Lambda}_b} + N_{\text{exp},\overline{\Lambda}_b}} S(y_i, \overline{x}, \sigma) + \frac{N_{\text{exp},\overline{\Lambda}_b}}{N_{\overline{\Lambda}_b} + N_{\text{exp},\overline{\Lambda}_b}} B(y_i, c)\right) \\
& - n_{\overline{\Lambda}_b} \log(N_{\overline{\Lambda}_b} + N_{\text{exp},\overline{\Lambda}_b}) + N_{\overline{\Lambda}_b} + N_{\text{exp},\overline{\Lambda}_b}
\end{aligned} \tag{64}$$

where $\overline{x}$ is the signal shape mean, σ is the first CB PDF width, c is the exponential parameter, N_{Λ_b} is the number of fitted Λ_b signal events, N_{exp,Λ_b} ($N_{\text{exp},\overline{\Lambda}_b}$)

[12] In BDT1+C strategy ProbNN cuts are optimized on Punzi FoM after application of BDT1 selection.

[13] These are the signal shape mean, the first CB component width and the exponential c parameter.

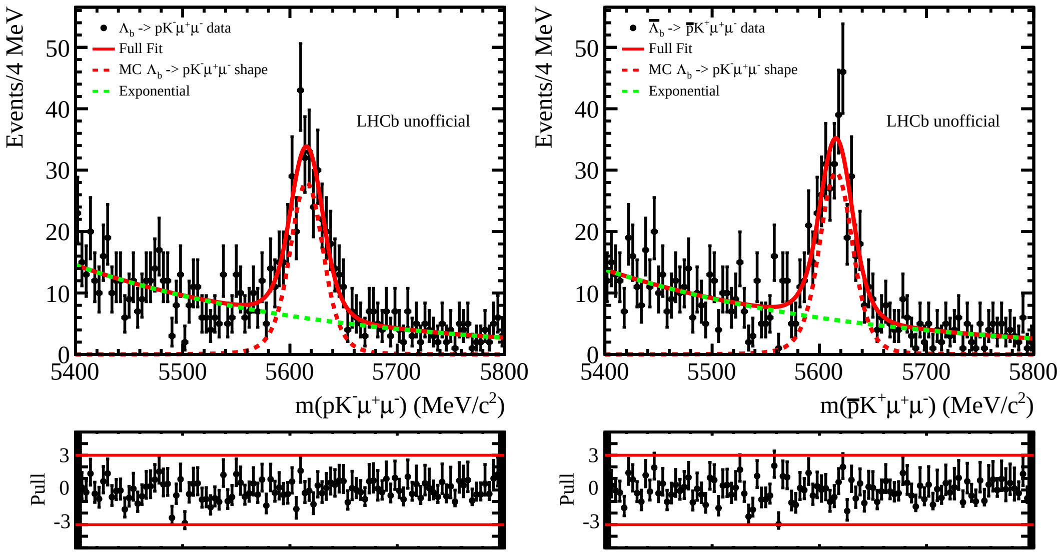


Figure 25: Simultaneous fit for $\mathcal{A}_{\text{raw}}$ extraction to $\Lambda_b \to pK^-\mu^+\mu^-$ data.

is the number of fitted background Λ_b ($\overline{\Lambda}_b$) events and $N_{\overline{\Lambda}_b}$ is a function of N_{Λ_b} and $\mathcal{A}_{\text{raw}}$ via Eq. 63. n_{Λ_b} and $n_{\overline{\Lambda}_b}$ are the number of Λ_b and $\overline{\Lambda}_b$ candidates to be fitted, x_i and y_i refers to the i-th Λ_b and $\overline{\Lambda}_b$ candidate, $S(x_i, \overline{x}, \sigma)$ and $B(x_i, c)$ are the signal and background shapes defined in Sec. 5.
Simultaneous fits to $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^-J/\psi$ data are shown respectively in Fig. 25 and Fig. 26, the relative fit results are reported in Table 13.
$\Delta\mathcal{A}_{CP}$ is then derived from fitted raw asymmetries, obtaining

$$\Delta\mathcal{A}_{CP} = (-4.8 \pm 5.0) \times 10^{-2}$$

where the reported uncertainty is statistical only.
The goodness of fit is first checked looking at the pulls, defined for each histogram bin as the difference between the bin value and the fitted function value, difference then divided by the associated bin error. The pull distribution should be compatible with statistical fluctuations if the resulting fit parametrization is good. No significant structure in the pull distributions is visible. A more careful study of the fit behaviour on toy MC experiments is carried out in Sec. 7.3.

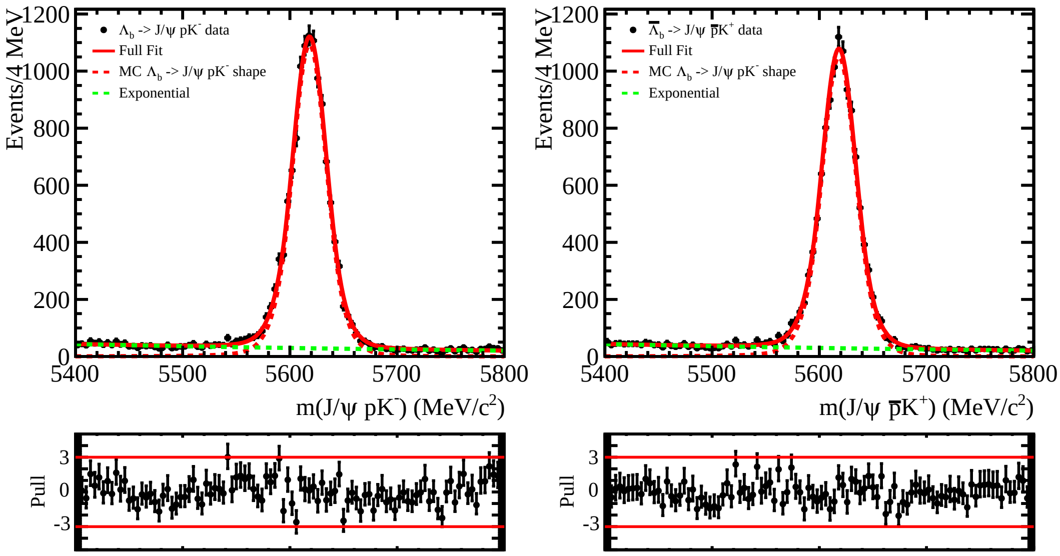
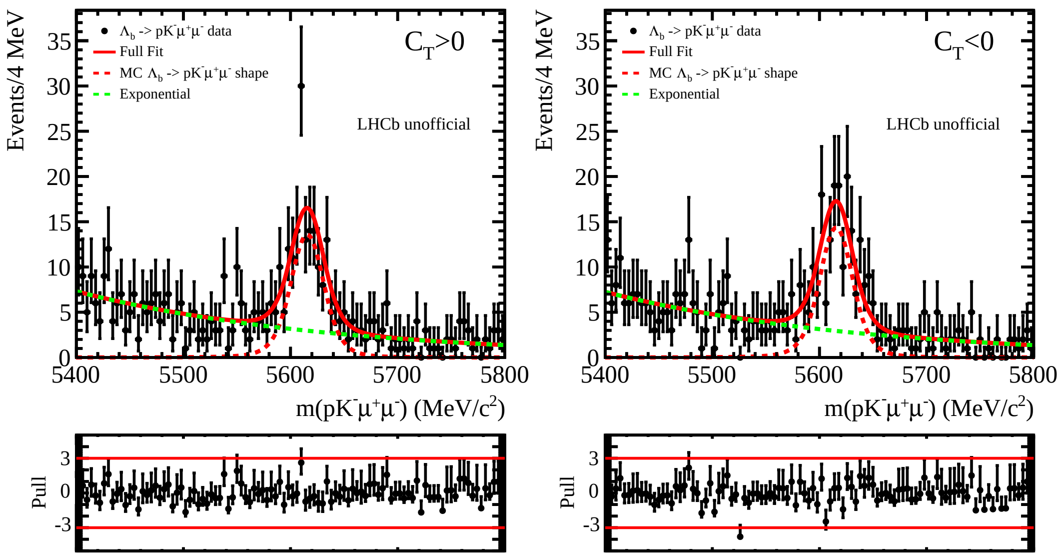


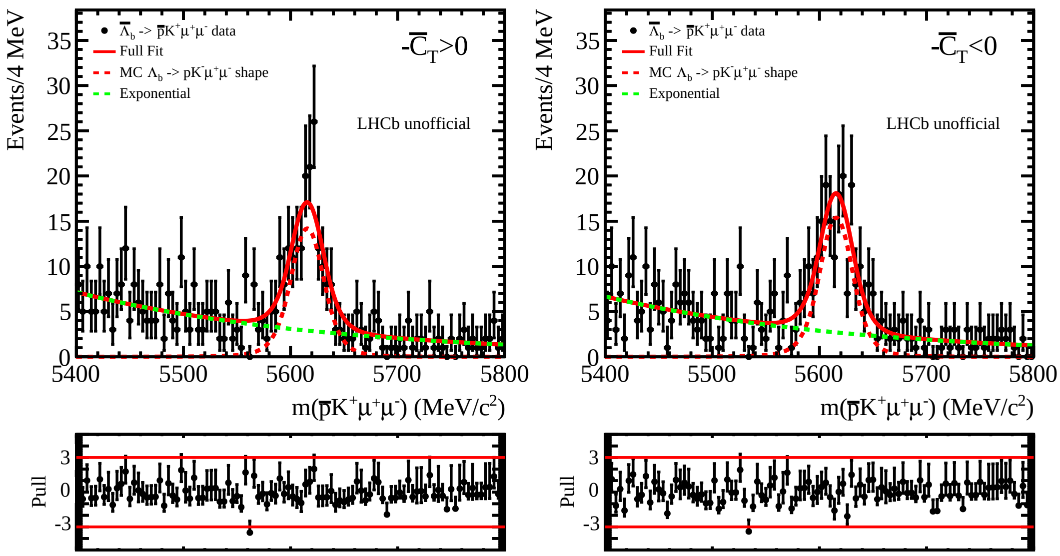
Figure 26: Simultaneous fit for $\mathcal{A}_{\mathrm{raw}}$ extraction to $\Lambda_b \to pK^- J/\psi$ data.

	$\Lambda_b \to pK^-\mu^+\mu^-$		$\Lambda_b \to pK^- J/\psi$	
Variable	Fit Value	Error	Fit Value	Error
$\mathcal{A}_{\mathrm{raw}}(10^{-2})$	-2.8	$\pm$ 5.0	1.97	$\pm$ 0.70
Λ_b Signal N	291	$\pm$ 22	11674	$\pm$ 115
Λ_b BKG N	706	$\pm$ 30	3036	$\pm$ 67
$\overline{\Lambda}_b$ BKG N	671	$\pm$ 29	3049	$\pm$ 67
Mean (MeV/c^2)	5615.60	$\pm$ 0.94	5618.30	$\pm$ 0.13
CB1 σ (MeV/c^2)	14.58	$\pm$ 0.88	14.29	$\pm$ 0.10
Exponential c (10^{-3})	-4.18	$\pm$ 0.27	-1.89	$\pm$ 0.12

Table 13: Fit results for direct CP asymmetry simultaneous fits to $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^- J/\psi$ samples.



(a)



(b)

Figure 27: Simultaneous fit for A_T, $\overline{A}_T$ extraction to $\Lambda_b \to pK^-\mu^+\mu^-$ data divided in (a) Λ_b and (b) $\overline{\Lambda}_b$ datasets. Left (right) plots are $C_T > 0$ ($C_T < 0$) subsamples.

Variable	Fit Value	Error
$A_T(10^{-2})$	-2.8	$\pm$ 7.2
$\overline{A}_T(10^{-2})$	-4.0	$\pm$ 6.9
Λ_b Signal Events	291	$\pm$ 22
$\overline{\Lambda}_b$ Signal Events	309	$\pm$ 22
Λ_b $C_T > 0$ BKG Events	354	$\pm$ 21
Λ_b $C_T < 0$ BKG Events	351	$\pm$ 21
$\overline{\Lambda}_b$ $\overline{C}_T > 0$ BKG Events	347	$\pm$ 21
$\overline{\Lambda}_b$ $\overline{C}_T < 0$ BKG Events	323	$\pm$ 20
Mean (MeV/c^2)	5615.60	$\pm$ 0.95
CB1 σ (MeV/c^2)	14.63	$\pm$ 0.88
Exponential c (10^{-3})	-4.19	$\pm$ 0.27

Table 14: Fit results for A_T, $\overline{A}_T$ extraction simultaneous fit to $\Lambda_b \to pK^-\mu^+\mu^-$ sample.

7.2 Triple-product asymmetries

The $\Lambda_b \to pK^-\mu^+\mu^-$ data sample after selection is split into 4 subsamples according to Λ_b flavour and $C_T/\overline{C}_T$ sign. The Λ_b mass distributions of all subsamples are simultaneously fit in analogy to what done for direct *CP* asymmetry in the previous section, where now triple-product asymmetries A_T, $\overline{A}_T$ are extracted from the fit model by writing the following yields as (Eq. 42)

$$\begin{aligned}
N_{\Lambda_b, C_T>0} &= \frac{1}{2} N_{\Lambda_b}(1 + A_T) \\
N_{\Lambda_b, C_T<0} &= \frac{1}{2} N_{\Lambda_b}(1 - A_T) \\
N_{\overline{\Lambda}_b, -\overline{C}_T>0} &= \frac{1}{2} N_{\overline{\Lambda}_b}(1 + \overline{A}_T) \\
N_{\overline{\Lambda}_b, -\overline{C}_T<0} &= \frac{1}{2} N_{\overline{\Lambda}_b}(1 - \overline{A}_T)
\end{aligned} \tag{65}$$

The minimized log-likelihood is analogous to Eq. 64, with four terms corresponding to the four considered datasets.
Simultaneous fit to $\Lambda_b \to pK^-\mu^+\mu^-$ data is shown in Fig. 27, results are reported in Table 14. $a_{CP}^{T_{\rm odd}}$ is then computed from fitted A_T, $\overline{A}_T$, obtaining

$$a_{CP}^{T_{\rm odd}} = (0.6 \pm 5.0) \times 10^{-2}$$

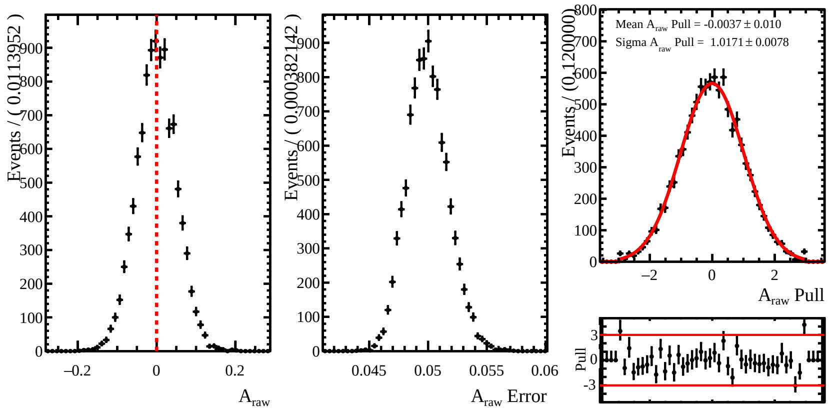
where the reported uncertainty is statistical only. The goodness of fit is first checked looking at the pulls of the four histograms, no significant structure in the pull distributions is visible. A more careful study of fit behaviour on toy MC experiments is carried out in Sec. 7.3.

7.3 Fit validation

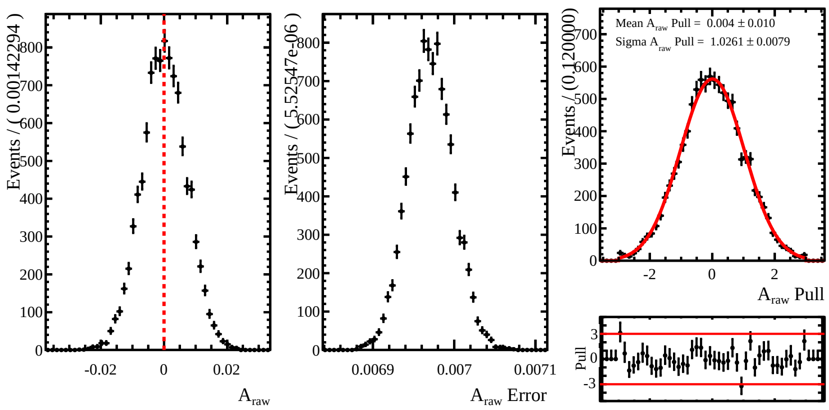
Inaccuracies in data parametrization due to the chosen fit model may produce a biasing effect on CPV observables. This effect is studied by generating 10000 simulated experiments (toy MC). For each toy, the data subsamples are generated according to the $m(\Lambda_b)$ data distribution after selection, the number of generated events per subsample is one half of the real data events after selection, plus a Poissonian fluctuation different for each subsample. This ensures that every toy MC is generated starting from a zero asymmetry condition, this way a possible resulting bias can not be due to a toy generation asymmetry, but must come from a fit model effect. Toys are then fitted using the unbinned simultaneous maximum likelihood fits set for data fit, described in Secs. 7.1,7.2. The pull distribution for each asymmetry A is defined as

$$\frac{A_{\text{fit}} - A_{\text{gen}}}{\sigma(A_{\text{fit}})} \tag{66}$$

where $A_{\text{gen}} = 0$ from toy generation and it is fitted with a Gaussian function, which should have mean compatible with zero if the fit model cause no bias. The fit validation for $\mathcal{A}_{\text{raw}}$ observable is carried out separately for signal and control channels (Fig. 28). The fit validation for triple-product asymmetries is done only for signal events (Fig. 29). No bias is observed as the Gaussian function means are all compatible with zero.

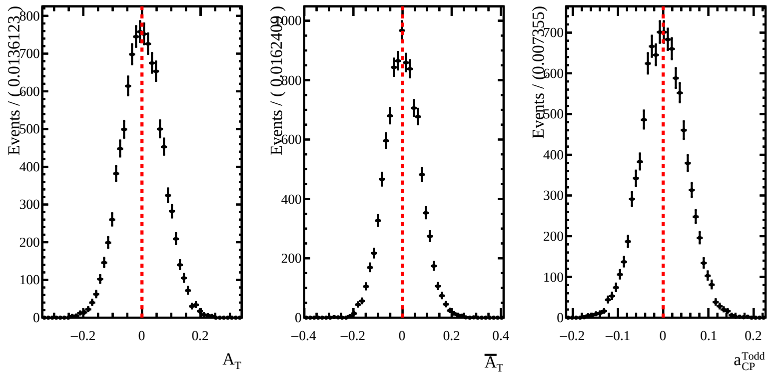


(a)

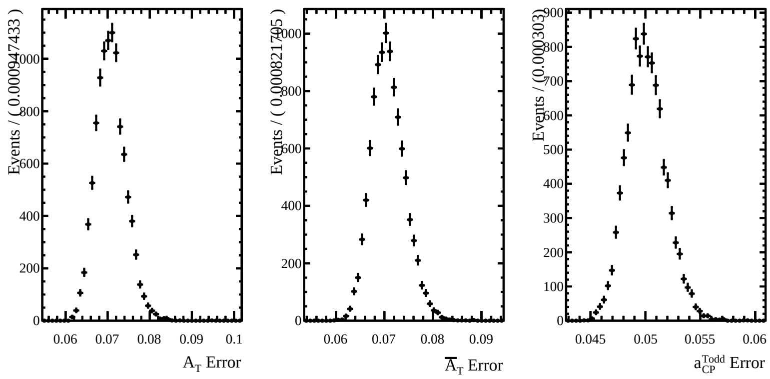


(b)

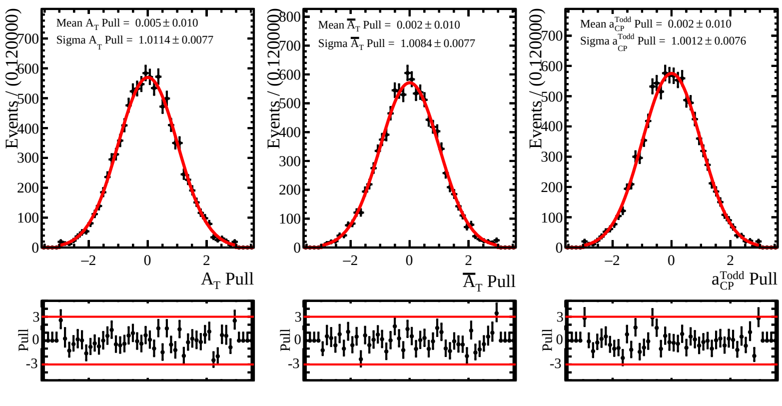
Figure 28: Toy MC study on $\mathcal{A}_{\rm raw}$ observable for (a) $\Lambda_b \to pK^-\mu^+\mu^-$ and (b) $\Lambda_b \to pK^-J/\psi$ events. Left plots are the fitted $\mathcal{A}_{\rm raw}$ values, the dashed red line represents the generated zero asymmetry. Central plots are the fitted $\mathcal{A}_{\rm raw}$ errors and right plots are the pull distributions, fitted with a Gaussian function.



(a)



(b)



(c)

Figure 29: Toy MC study on triple-product asymmetries for $\Lambda_b \to pK^-\mu^+\mu^-$ data. (a) are the fitted values, where dashed red lines correspond to the generated zero asymmetries. (b) are the fitted statistical errors and (c) are the pull distributions, fitted with a Gaussian function.

Variable	Fit Value	Error
$\mathcal{A}_{\rm raw}(10^{-2})$	1.69	$\pm$ 0.68
Λ_b Signal N	11925	$\pm$ 114
Λ_b BKG N	2767	$\pm$ 65
$\overline{\Lambda}_b$ BKG N	2762	$\pm$ 65
Mean (MeV/c^2)	5618.50	$\pm$ 0.13
CB1 σ (MeV/c^2)	14.25	$\pm$ 0.10
Exponential c (10^{-3})	-2.03	$\pm$ 0.13

Table 15: Fit results for $\mathcal{A}_{\rm raw}$ extraction simultaneous fit to reweighted $\Lambda_b \to pK^-J/\psi$ data.

8 Systematic uncertainties

8.1 Direct *CP* asymmetry measurement

The study of the possible systematic uncertainties sources affecting the direct *CP* asymmetry measurement is carried out following [35].

8.1.1 Experimental bias

As mentioned in Sec. 1.3.1 proton and kaon detection asymmetries do not completely cancel out in Eq. 35 because they depend on the different kinematics of control and signal decays, possibly introducing fake $\Delta\mathcal{A}_{CP}$ asymmetry. This effect is estimated by reweighting $\Lambda_b \to pK^-J/\psi$ candidates using weights derived from proton and kaon momentum distribution comparison between MC $\Lambda_b \to pK^-\mu^+\mu^-$ and MC $\Lambda_b \to pK^-J/\psi$ samples. The observed difference in $\mathcal{A}_{\rm raw}$ ($\Lambda_b \to pK^-J/\psi$) between simultaneous fits made using reweighted or not $\Lambda_b \to pK^-J/\psi$ data is taken as systematic uncertainty due to the experimental setup bias.
Simultaneous fit to reweighted $\Lambda_b \to pK^-J/\psi$ data is shown in Fig. 30, relative fit results are reported in Table 15, to be compared to the unweighted $\Lambda_b \to pK^-J/\psi$ data simultaneous fit results (Table 13). The difference between the two fitted $\mathcal{A}_{\rm raw}$ ($\Lambda_b \to pK^-J/\psi$) gives an experimental bias systematic uncertainty on $\Delta\mathcal{A}_{CP}$ of $\pm 0.3 \times 10^{-2}$.

8.1.2 Fit model

The choice of the shapes parametrizing signal and background events (Sec. 5) is in practice not unique nor perfect. The measured *CPV* observables may therefore depend on the particular fit model chosen for the simultaneous fit.

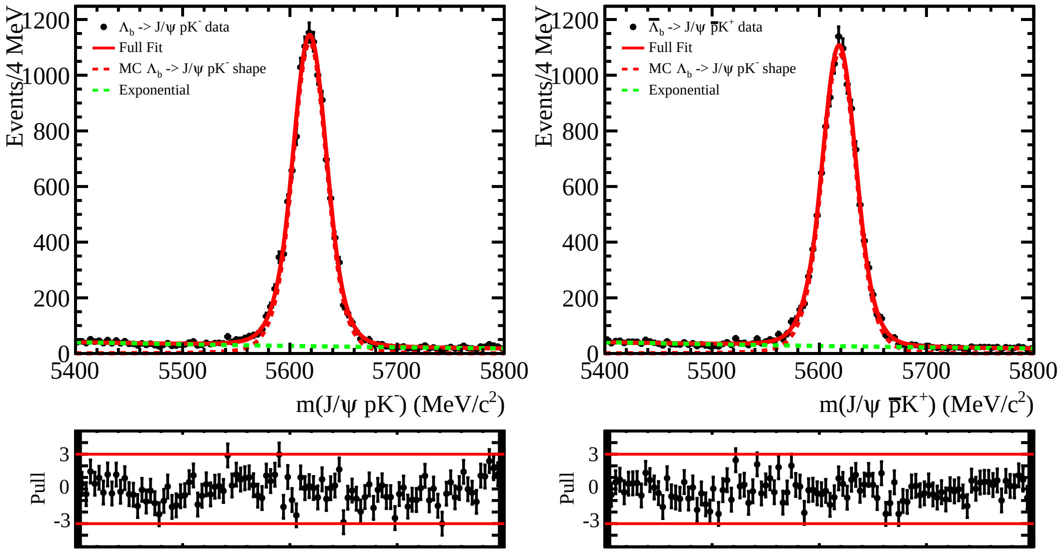


Figure 30: Simultaneous fit for $\mathcal{A}_{\rm raw}$ extraction to reweighted $\Lambda_b \to pK^- J/\psi$ data.

To assess possible fit model biasing effects a different fit model, alternative to the nominal one of Sec. 5, is set as follows.
The signal shapes for signal and control decays are made of one Gaussian and two bifurcated Gaussian functions[14], sharing the same mean. Signal shape parameters are fixed from MC fit as done for the nominal ones, leaving the overall normalization, the common mean and the Gaussian width as free parameters. Different signal shapes like one bifurcated plus one standard Gaussian, double Gaussian or Johnson SU[15] PDFs have been tested but they are not able to parametrize the observed signal tails.
As background shape a linear function is chosen in place of the exponential one.
The possible fit model biasing effect on CPV observables is studied by gen-

[14]Bifurcated means that the Gaussian function has different widths for left and right sides.

[15]The Johnson SU function is a four-parameter PDF defined as

$$J_{SU}(x) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+(\frac{x-\epsilon}{\lambda})^2}} \exp\left\{-\frac{1}{2}\left[\gamma + \delta \sinh^{-1}\left(\frac{x-\epsilon}{\lambda}\right)\right]^2\right\}.$$

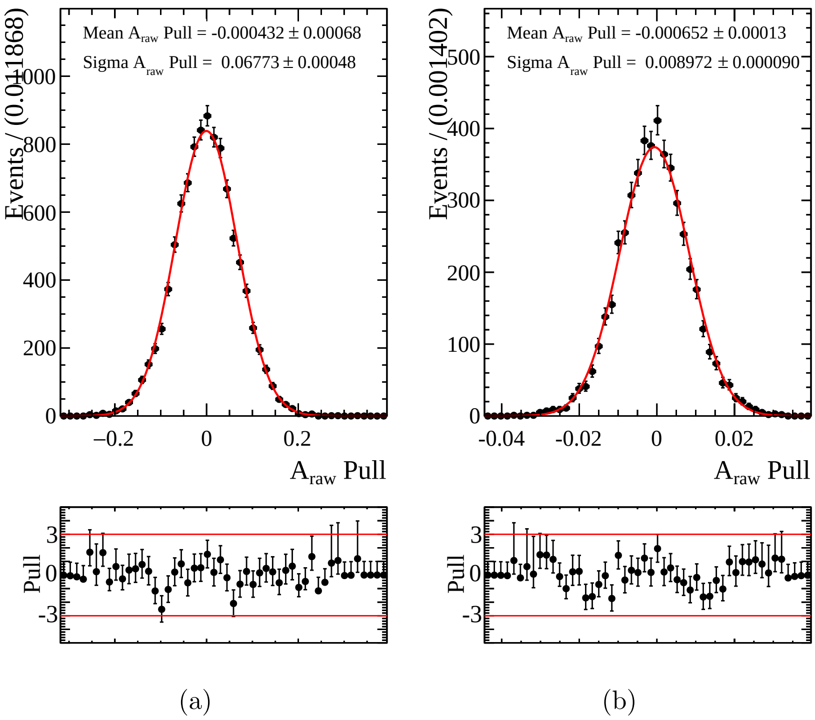


Figure 31: Pull distributions (Eq. 67) fitted with a Gaussian function, obtained from the toy MC study for fit model bias on $\mathcal{A}_{\rm raw}$ observable for (a) $\Lambda_b \to pK^-\mu^+\mu^-$ and (b) $\Lambda_b \to pK^-J/\psi$ events.

erating 10000 toy MC for signal decays and 5000 for control decays[16]. As done for fit validation (Sec. 7.3), Λ_b and $\overline{\Lambda}_b$ subsamples are generated according to the $m(\Lambda_b)$ data distribution after selection, the number of generated events is one half of the data events after selection per subsample, plus a Poissonian fluctuation different for each sample. This ensures that each MC toy is generated starting from a zero asymmetry condition. MC toys are then fitted using the unbinned simultaneous maximum likelihood fit set in Sec. 7.1 employing both the nominal and the alternative fit models. The pull distribution, for an asymmetry A, is defined as

$$\frac{A_{\text{nom}}}{\sigma(A_{\text{nom}})} - \frac{A_{\text{alt}}}{\sigma(A_{\text{alt}})} \tag{67}$$

where A_{nom} and A_{alt} are the fitted asymmetries using the two models. This pull distribution is then fitted with a Gaussian function, whose mean should be compatible with zero if differences in fitted asymmetries between the two models are compatible with statistical fluctuations.
The fit model study for $\mathcal{A}_{\text{raw}}$ observable is carried out separately for signal and control channel (Fig. 31). No bias is observed in the pull distribution for signal decays. The uncertainty on the fitted Gaussian mean for signal decays $\pm 0.068 \times 10^{-2}$, rescaled by the $\mathcal{A}_{\text{raw}}(\Lambda_b \to pK^-\mu^+\mu^-)$ statistical uncertainty (Table 13) gives a very small systematic uncertainty on $\mathcal{A}_{\text{raw}}$ $(\Lambda_b \to pK^-\mu^+\mu^-)$ of $\pm 0.0034 \times 10^{-2}$. A fit model bias is observed in control decays, however, the fitted Gaussian mean -0.0652×10^{-2}, rescaled by the $\mathcal{A}_{\text{raw}}(\Lambda_b \to pK^-J/\psi)$ statistical uncertainty (Table 13) gives an even smaller systematic uncertainty on $\mathcal{A}_{\text{raw}}(\Lambda_b \to pK^-J/\psi)$ of $\pm 0.00045 \times 10^{-2}$. Therefore the systematic uncertainty due to a possible fit model bias on $\Delta\mathcal{A}_{CP}$ observable is considered negligible.

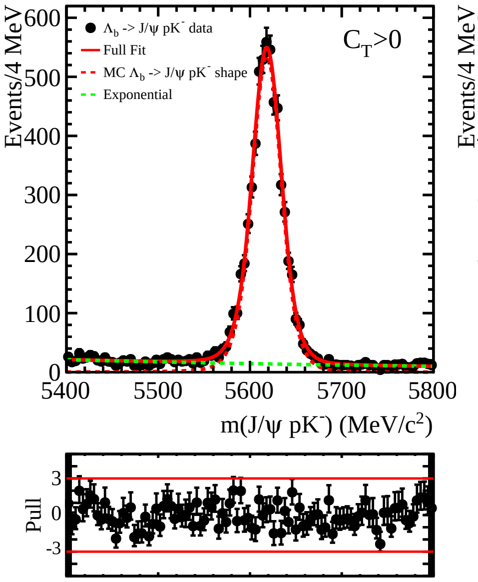
8.2 Triple-product asymmetries measurement

The study of the possible sources of systematic uncertainties affecting the triple-product asymmetries measurement is carried out following [29].

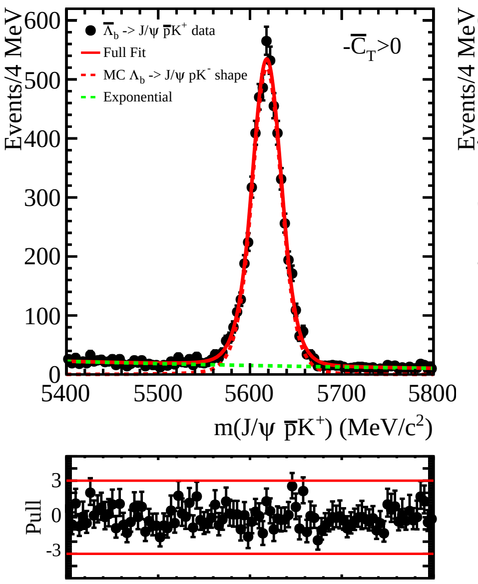
8.2.1 Experimental bias

An experimental bias on triple-product asymmetries may be introduced by both the selection made on $\Lambda_b \to pK^-\mu^+\mu^-$ candidates and the detector acceptance. This is assessed in the first place by checking the correlation

[16] Less MC toys are considered for control decays because of their higher computing resource consumption.



(a)



(b)

Figure 32: Simultaneous fit for A_T, $\overline{A}_T$ extraction to $\Lambda_b \to pK^- J/\psi$ data divided in (a) Λ_b and (b) $\overline{\Lambda}_b$ datasets. Left (right) plots are $C_T > 0$ ($C_T < 0$) subsamples.

Variable	Fit Value	Error
$A_T(10^{-2})$	-1.97	$\pm$ 0.97
$\overline{A}_T(10^{-2})$	-1.0	$\pm$ 1.0
Λ_b Signal Events (10^3)	11.67	$\pm$ 0.11
$\overline{\Lambda}_b$ Signal Events (10^3)	11.22	$\pm$ 0.11
Λ_b $C_T > 0$ BKG Events	1481	$\pm$ 46
Λ_b $C_T < 0$ BKG Events	1555	$\pm$ 48
$\overline{\Lambda}_b$ $\overline{C}_T > 0$ BKG Events	1563	$\pm$ 47
$\overline{\Lambda}_b$ $\overline{C}_T < 0$ BKG Events	1486	$\pm$ 46
Mean (MeV/c^2)	5618.30	$\pm$ 0.13
CB1 σ (MeV/c^2)	14.29	$\pm$ 0.10
Exponential c (10^{-3})	-1.89	$\pm$ 0.11

Table 16: Fit results for A_T, $\overline{A}_T$ extraction simultaneous fit to $\Lambda_b \to pK^-J/\psi$ sample.

between C_T and BDT variable, which is found negligible (Fig. 21). In the second place, the $a_{CP}^{T_{\rm odd}}$ observable is measured for our control channel $\Lambda_b \to pK^-J/\psi$. Any CP-violating effect is expected to be zero for this process, since the only diagram providing a CPV term is the penguin with a u quark loop one (Fig. 2(b)), suppressed with respect to the tree diagram (2(a)) as well as by the tiny u quark mass and the CKM hierarchy factor λ^2. Hence, a non-zero $a_{CP}^{T_{\rm odd}}$ value can be regarded as an experimental bias effect. Simultaneous fit to $\Lambda_b \to pK^-J/\psi$ data is shown in Fig. 32, relative results are reported in Table 16.
For our control channel $a_{CP}^{T_{\rm odd}} = (0.48 \pm 0.70) \times 10^{-2}$ is found compatible with zero as expected, where the reported uncertainty is only statistical. This statistical uncertainty is assigned as systematic error on $a_{CP}^{T_{\rm odd}}$ for $\Lambda_b \to pK^-\mu^+\mu^-$ decay. Even though final-state interactions and parity-violation effects may produce a non-zero expectation value of A_T and $\overline{A}_T$ asymmetries, also for our control channel, these are found to be quite small. Since the statistical error on the two asymmetries is basically the same and no correlation between them is observed, the systematic uncertainties on A_T and $\overline{A}_T$ observables are conservatively estimated as $\sigma(A_T) = \sigma(\overline{A}_T) = \sqrt{2}\sigma(a_{CP}^{T_{\rm odd}})$. The experimental bias systematic uncertainties on triple-product observables are summarized in Table 17.

	ΔA_T (10^{-2})	$\Delta \overline{A}_T$ (10^{-2})	$\Delta a_{CP}^{T_{\rm odd}}$ (10^{-2})
Syst. error	$\pm$ 1.0	$\pm$ 1.0	$\pm$ 0.7

Table 17: Systematic uncertainties assigned for a possible experimental bias.

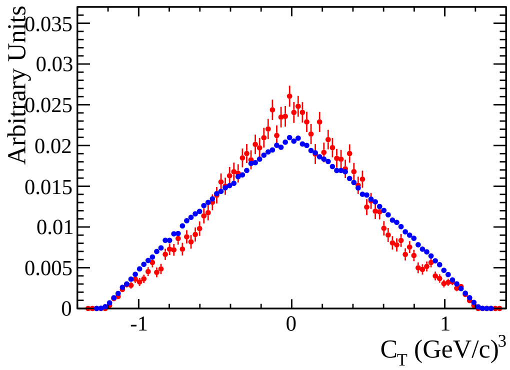


Figure 33: Comparison between C_T normalized distributions for MC $\Lambda_b \to pK^-J/\psi$ (blue) and *s-Weighted* $\Lambda_b \to pK^-J/\psi$ (red) samples.

8.2.2 Detector resolution bias on C_T

The possible biasing effect due to detector resolution is estimated by studying the residual distribution defined as $(C_{T,\rm reco} - C_{T,\rm gen})$ on MC $\Lambda_b \to pK^-\mu^+\mu^-$ events, where $C_{T,\rm reco}$ is the reconstructed C_T value and $C_{T,\rm gen}$ is the true generated value. A small data/MC discrepancy for C_T distributions, due to differences between real and simulated decay dynamics between the two samples, is observed (Fig. 33), but it is considered irrelevant for the purpose of this study.
The residual distribution $(C_{T,\rm reco} - C_{T,\rm gen})$ is found to be compatible with no bias (symmetric with zero mean) and its dependence on $C_{T,\rm gen}$ is almost flat (Fig. 34) as desirable.
Then, triple-product asymmetries are measured by counting the number of Λ_b and $\overline{\Lambda}_b$ events belonging to each $C_T > 0$ and $C_T < 0$ subsample for both reconstructed and generated C_T variables. Differences between reconstructed and generated values are taken as detector resolution systematic uncertainties (Table 18).

8.2.3 Fit model

The possible biasing effect of the chosen fit model on triple-product asymmetries is carried out in analogy to what done for direct CP asymmetries

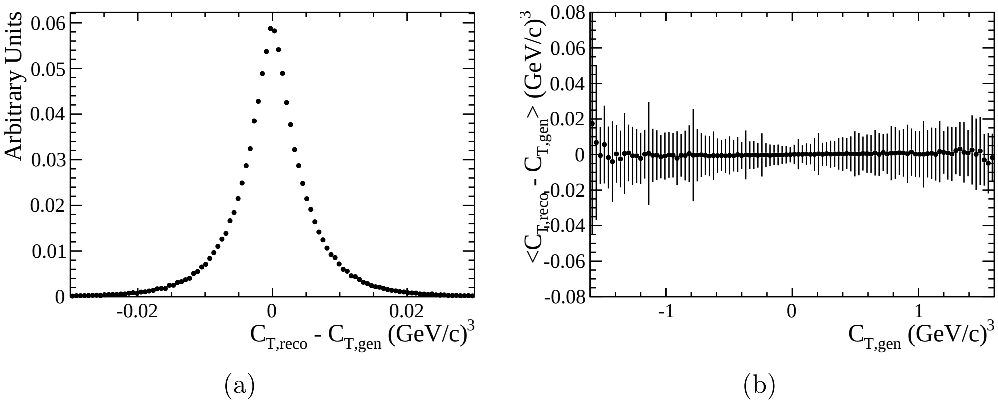


Figure 34: (a) Residual distribution for MC $\Lambda_b \to pK^-\mu^+\mu^-$ and (b) the same as a function of $C_{T,\text{gen}}$. Black points and error bars correspond respectively to mean and standard deviation of the residual distribution in each $C_{T,\text{gen}}$ bin.

	A_T (10^{-2})	$\overline{A}_T$ (10^{-2})	$a_{CP}^{T_\text{odd}}$ (10^{-2})
Reco	0.09 ± 0.58	0.75 ± 0.60	-0.33 ± 0.42
Gen	0.09 ± 0.58	0.71 ± 0.60	-0.31 ± 0.42
	ΔA_T (10^{-2})	$\Delta \overline{A}_T$ (10^{-2})	$\Delta a_{CP}^{T_\text{odd}}$ (10^{-2})
Syst. error	± 0.00	± 0.04	± 0.02

Table 18: Triple-product asymmetries measured on MC $\Lambda_b \to pK^-\mu^+\mu^-$ events using reconstructed and generated C_T observables. Differences between the two values are taken as systematic uncertainties due to resolution bias.

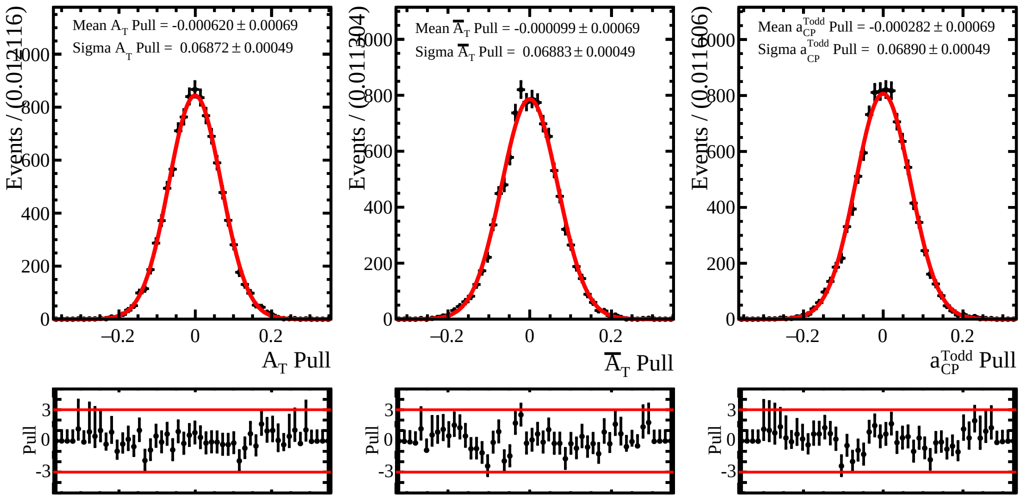


Figure 35: Pull distributions (Eq. 67) fitted with a Gaussian function, obtained from the toy MC study for fit model bias on triple-product asymmetries.

(Sec. 8.1.2). 10000 MC toys are generated, where for each of the four $\Lambda_b/\overline{\Lambda}_b$ and $C_T < 0/C_T > 0$ subsamples a number of events equal to one quarter of the data events after selection plus different Poissonian fluctuations are generated according to the $m(\Lambda_b)$ data distribution. Toys are then fitted using the unbinned simultaneous maximum likelihood fit described in Sec. 7.2 using both nominal and alternative models. For each observable, the pull distribution defined in Eq. 67 is fitted with a Gaussian function.
The fit model study for triple-product asymmetries is reported in Fig. 35. No bias is observed in the pull distributions. As the resulting systematic uncertainties resulting from the Gaussian means statistical uncertainties are tiny (0.0034×10^{-2} for $a_{CP}^{T_{\text{odd}}}$) the systematic uncertainties due to a possible fit model bias on triple-product asymmetries are considered negligible.

8.2.4 Λ_b polarization effects

In contrast to spinless B meson decays, Λ_b baryon is a spin 1/2 particle which may be produced with a non-zero polarization. While Λ_b longitudinal polarization is expected to vanish for parity conservation in strong interactions, heavy-quark effective theory predicts Λ_b baryon to retain the initial b quark transverse polarization after the hadronization process. Previous measurements found Λ_b polarization in LHCb to be compatible with zero, although statistical uncertainties do not exclude a polarization of a few %: [78] mea-

sured Λ_b polarization to be $P = 0.06 \pm 0.07_{\text{stat}} \pm 0.02_{\text{syst}}$; [64] found a polarization value of $P = -0.020 \pm 0.023$.
The C_T triple-product (Eq. 41) is built using only particles momenta, without spin vectors, but nonetheless they may be sensible to the daughter particles momentum angular distribution.
The $a_{CP}^{T_{\text{odd}}}$ observable is insensible to polarization effects if Λ_b and $\overline{\Lambda}_b$ polarizations are equal, this because in absence of CPV Λ_b and $\overline{\Lambda}_b$ decay amplitudes are identical, thus any polarization effect would be equal for A_T and $\overline{A}_T$ observables and would cancel out in $a_{CP}^{T_{\text{odd}}}$ expression. Actually, the different production processes of Λ_b and $\overline{\Lambda}_b$ particles may cause differences in their polarizations, but [78] found them equal within statistical uncertainties.
Recently the ongoing $\Lambda_b \to ph^-h^+h^-$ analysis measured triple-product asymmetries on a Λ_b polarized generated sample showing that A_T, $\overline{A}_T$ and $a_{CP}^{T_{\text{odd}}}$ observables are all insensitive to any Λ_b polarization effects.
For these reasons polarizations effects are not further investigated and no systematic uncertainty is given.

8.3 Cross-checks

8.3.1 Stability of results

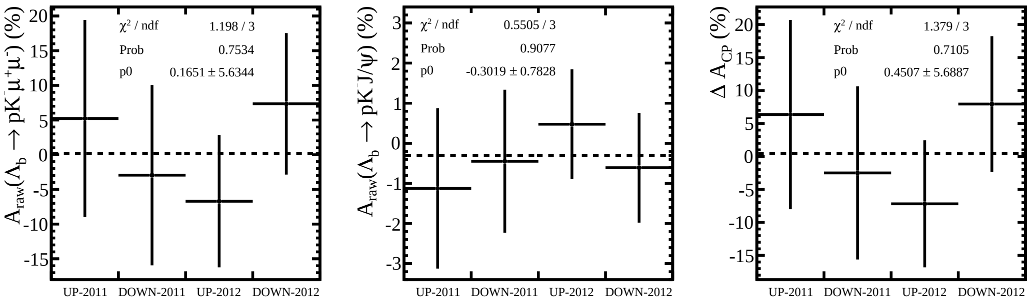
Stability of measured CPV observables with respect to magnet polarity and data taking year has been checked by measuring the CPV observables into the four up/down polarity and 2011/2012 year data subsamples (Fig. 36). No significant deviation from the measured CPV observables is seen. Stability of measured CPV observables with respect to BDT cut has been checked by applying different BDT cuts (Fig. 37). No significant dependence of measured CPV observables on the BDT cut is observed. It is important to note that BDT is trained including PID variables, so this check implies that dependence of CPV observables on the PID variables used for the selection process is under control.

8.3.2 Reconstruction efficiency vs C_T

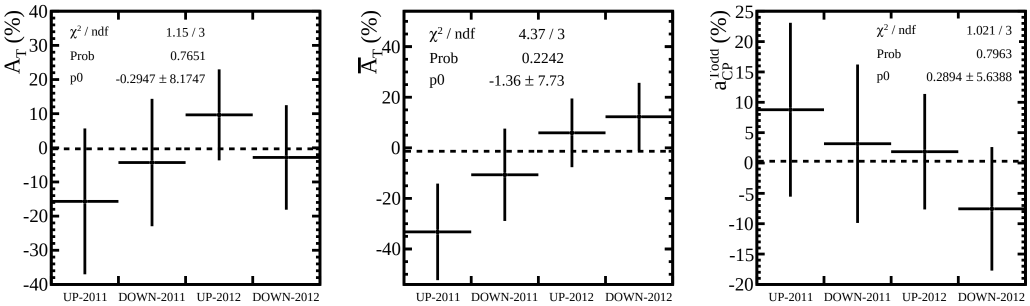
Reconstruction efficiency of the employed selection for positive and negative C_T events, defined as

$$\epsilon(+) = \frac{N(C_T > 0, \text{BDT} > 0.1775)}{N(C_T > 0)} \quad \epsilon(-) = \frac{N(C_T < 0, \text{BDT} > 0.1775)}{N(C_T < 0)} \tag{68}$$

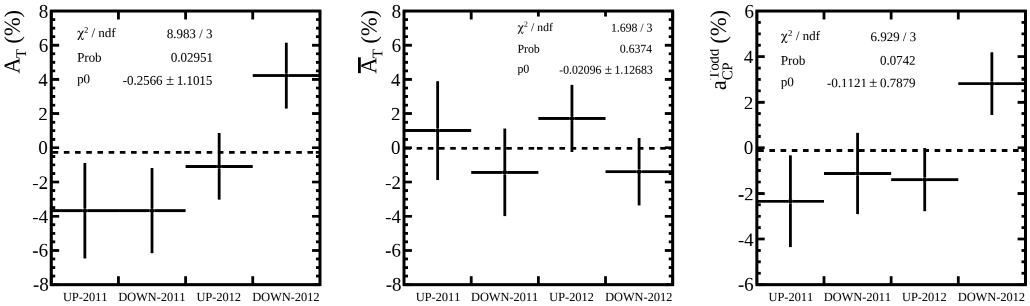
are studied on MC $\Lambda_b \to pK^-\mu^+\mu^-$ events, separately for Λ_b and $\overline{\Lambda}_b$ events. A slight dependence of the reconstruction efficiency on $|C_T|$ is observed



(a)

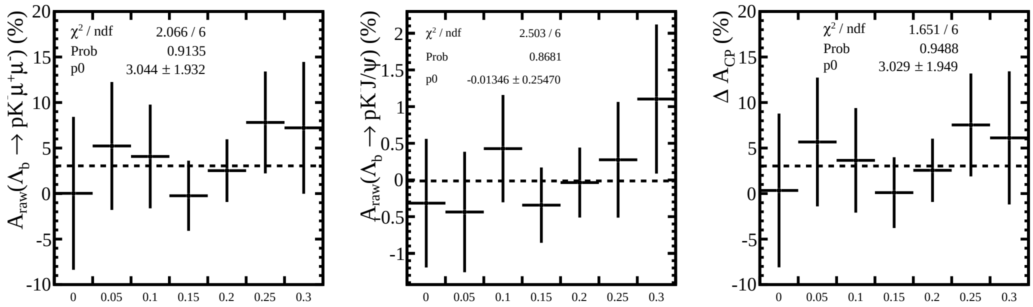


(b)

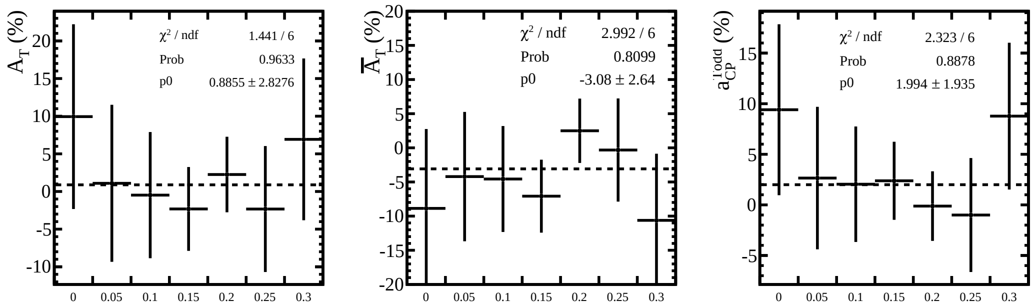


(c)

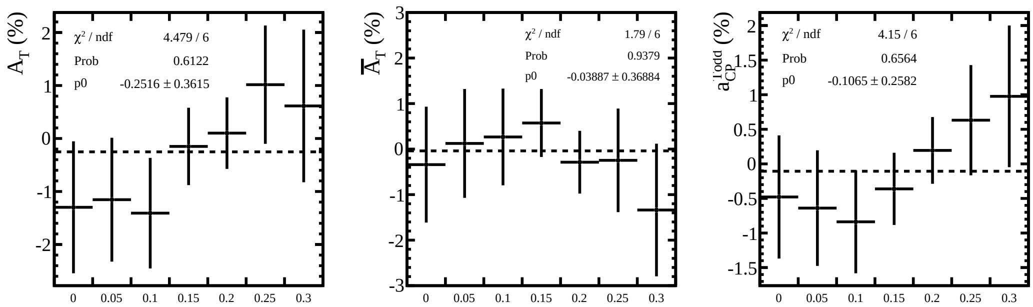
Figure 36: Stability of CPV observables with respect to magnet polarity/year for (a) direct CP asymmetries and triple-product asymmetries measured on (b) signal and (c) control decays. The plotted results are the shifts with respect the measured values (Tables 13,14), the dashed line is the fit with a constant function. Shift statistical uncertainties are computed taking into account the correlation among datasets.



(a)

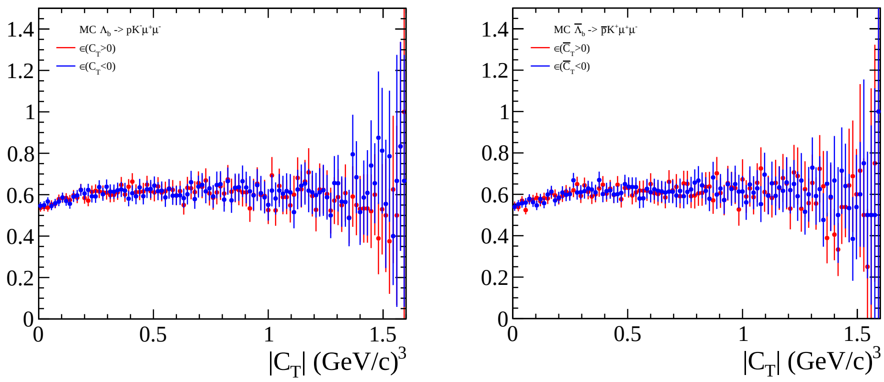


(b)



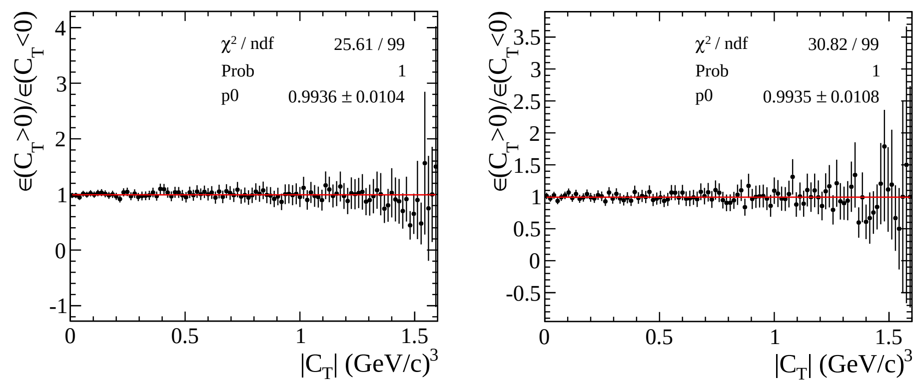
(c)

Figure 37: Stability of CPV observables with respect to BDT cut for (a) direct CP asymmetries and triple-product asymmetries measured on (b) signal and (c) control decays. The plotted results are the shifts with respect the measured values (Tables 13,14), the dashed line is the fit with a constant function. Shift statistical uncertainties are computed taking into account the correlation among datasets.



(a) (b)

Figure 38: Reconstruction efficiency of the employed selection for MC $\Lambda_b \to pK^-\mu^+\mu^-$ $C_T > 0$ (red) and $C_T < 0$ (blue) events as a function of $|C_T|$, separately for (a) Λ_b and (b) $\overline{\Lambda}_b$ events.



(a) (b)

Figure 39: Reconstruction efficiency ratios $\epsilon(+)/\epsilon(-)$ for MC $\Lambda_b \to pK^-\mu^+\mu^-$ events as a function of $|C_T|$, separately for (a) Λ_b and (b) $\overline{\Lambda}_b$ events. Red line is the fit with a constant function.

	$\Delta(\Delta\mathcal{A}_{CP})$ (10^{-2})	ΔA_T (10^{-2})	$\Delta\overline{A}_T$ (10^{-2})	$\Delta a_{CP}^{T_{\rm odd}}$ (10^{-2})
Shift	0.26 ± 0.66	0.05 ± 0.97	0.10 ± 0.93	0.08 ± 0.67

Table 19: Difference between fitted observables with or without multiple candidates. Correlation between the two fit samples is taken into account.

(Fig. 38), but their ratio is found compatible with one for both Λ_b and $\overline{\Lambda}_b$ events (Fig. 39). This ensures that the $a_{CP}^{T_{\rm odd}}$ observable is independent of reconstruction efficiencies, as mentioned in Sec. 1.3.2.

8.3.3 Multiple candidates

The fraction of multiple candidates after the full selection procedure is observed to be 0.9% in the signal sample and 0.8% in the control sample. The effect of multiple candidates is checked by measuring the CPV observables retaining only the best $\Lambda_b \to pK^-\mu^+\mu^-$ candidate per event according to the higher DTF probability. Differences with respect to the nominal fit results are found compatible with their statistical uncertainties (Table 19), therefore no systematic uncertainty is assigned.

8.3.4 Effectiveness of the veto cuts

The yields of each peaking component are checked directly on data after preselection, $c\overline{c}$ veto and BDT selection, but without any peaking component veto (Fig. 40). Apparently, the BDT selection is sufficient to exclude the major part of peaking background events, as the only significant component observed is the $\mu \leftrightarrow K$ swap one. After the application of its relative veto cut Eq. 54 its residual yield is negligible (Fig. 41). For this reason and given that the signal loss due to the applied vetoes is very small (Table 7) no biasing effect of the chosen vetoes on CPV observable is investigated and no systematic uncertainty is given.

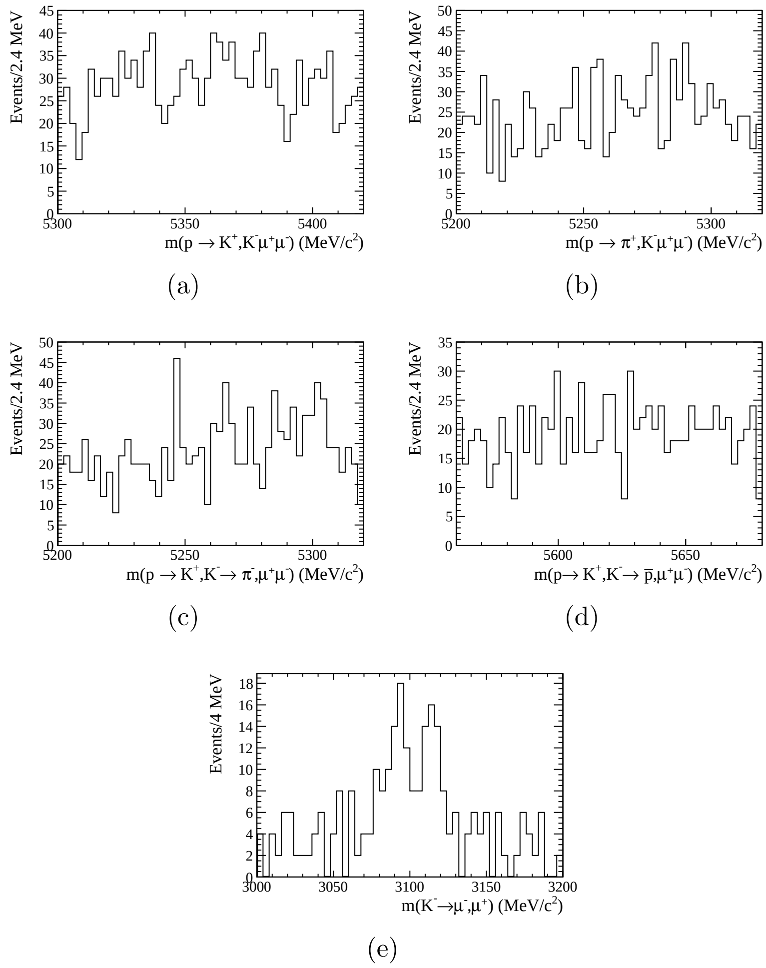


Figure 40: Peaking component yields after full data selection apart from the peaking component vetoes: (a) $B_s^0 \to K^+K^-\mu^+\mu^-$, (b) $B^0 \to K\pi\mu^+\mu^-$ single swap, (c) $B^0 \to K\pi\mu^+\mu^-$ double swap, (d) $p \leftrightarrow K$ swap, (e) $\mu \leftrightarrow K$ swap.

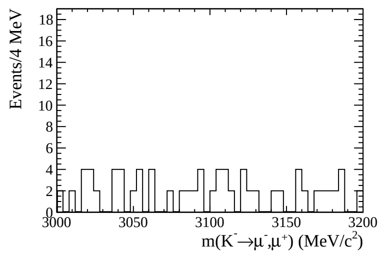


Figure 41: $\mu \leftrightarrow K$ swap component residual yield after full data selection and relative veto cut Eq. 54.

Source	$\Delta\mathcal{A}_{CP}$ (10^{-2})	A_T (10^{-2})	$\overline{A}_T$ (10^{-2})	$a_{CP}^{T_{\rm odd}}$ (10^{-2})
Experimental bias	$\pm$ 0.3	$\pm$ 1.0	$\pm$ 1.0	$\pm$ 0.7
C_T resolution bias	unaffected	$\pm$ 0.00	$\pm$ 0.04	$\pm$ 0.02
Fit model	negligible		negligible	
Λ_b polarization	unaffected		negligible	
Multiple candidates	negligible		negligible	
Veto cuts	negligible		negligible	

Table 20: Summary of the systematic uncertainty studies performed in Sec. 8.

9 Conclusions

The first measurement of $\Delta\mathcal{A}_{CP}$ observable (defined in Sec. 1.3.1) and triple-product asymmetries A_T, $\overline{A}_T$ and $a_{CP}^{T_{\rm odd}}$ (defined in Sec. 1.3.2) is performed on the full Run I dataset recorded by the LHCb experiment. After an accurate selection procedure (Secs. 4,6), these observables are extracted from data by means of unbinned extended simultaneous maximum likelihood fits (Secs. 5,7). Several studies have been performed for identifying the systematic uncertainties affecting the measured observables (Sec. 8), summarized in Table 20. Only the experimental bias source has a significant contribution to the overall uncertainty. The measured observables are

$$\begin{aligned}
\Delta\mathcal{A}_{CP} &= (-4.8 \pm 5.0\ (\text{stat}) \pm 0.3\ (\text{syst})) \times 10^{-2} \\
A_T &= (-2.8 \pm 7.2\ (\text{stat}) \pm 1.0\ (\text{syst})) \times 10^{-2} \\
\overline{A}_T &= (-4.0 \pm 6.9\ (\text{stat}) \pm 1.0\ (\text{syst})) \times 10^{-2} \\
a_{CP}^{T_{\rm odd}} &= (+0.6 \pm 5.0\ (\text{stat}) \pm 0.7\ (\text{syst})) \times 10^{-2}
\end{aligned} \tag{69}$$

All these observables are compatible with the CP-conservation hypothesis.

A Signal-to-control decay branching fraction estimate

An estimate for $\mathcal{B}(\Lambda_b \to pK^-\mu^+\mu^-)/\mathcal{B}(\Lambda_b \to pK^-J/\psi(\to \mu^+\mu^-))$ branching fraction is necessary for optimizing the selection using the significance figure-of-merit (FoM) while keeping the $\Lambda_b \to pK^-\mu^+\mu^-$ signal blinded.
The estimation is based on the $\Lambda_b \to \Lambda^*\mu^+\mu^-$ BR expectations for SM derived in [79] and the amplitude fit to the $m(pK^-)$ spectrum for $\Lambda_b \to pK^-J/\psi$ channel obtained in [64].
We start from the SM predictions for

$$\mathcal{B}(\Lambda_b \to \Lambda^*(1520)\mu^+\mu^-) = 1.3\ (2.1) \times 10^{-7} \text{ for SCA (MCN) FF} \tag{70}$$

where the two values refers to two different sets of Form Factors (FF) used for the computation. These form factors are derived within the framework of Heavy Quark Effective Theory (HQET), an effective description for the phenomenology of hadrons containing a single heavy quark. SCA form factors are obtained by using a non-relativistic reduction of the quark spinors and computed using single component wave functions. On the contrary, MCN form factors are obtained using the full relativistic form for the quark spinors and the full (multi-component) quark model wave function. The latter method should provide the more reliable results, since it makes use of fewer approximations, for this reason the BF estimate will be derived using the MCN predictions. Then, $\mathcal{B}(\Lambda_b \to pK^-\mu^+\mu^-)$ is found rescaling the above prediction by

$$\mathcal{B}(\Lambda^*(1520) \to pK^-) = 22.5\% \tag{71}$$

from [80] and

$$\frac{\mathcal{B}(\Lambda_b \to pK^-\mu^+\mu^-)}{\mathcal{B}(\Lambda_b \to \Lambda^*(1520)(\to pK^-)\mu^+\mu^-)} = 6.05 \tag{72}$$

derived from the ratio between the sum of all the $\Lambda_b \to \Lambda^* J/\psi$ fractions found in the $\Lambda_b \to pK^-J/\psi$ amplitude fit [64] (excluding the two pentaquark states) and the $\Lambda^*(1520)$ fraction. The resulting BR estimate is

$$\mathcal{B}(\Lambda_b \to pK^-\mu^+\mu^-) = 2.86 \times 10^{-7} \tag{73}$$

$\mathcal{B}(\Lambda_b \to pK^-J/\psi(\to \mu^+\mu^-))$ is taken from the first LHCb measurement [81]

$$\mathcal{B}(\Lambda_b \to pK^-J/\psi(\to \text{any final state})) = 3.04 \times 10^{-4} \tag{74}$$

rescaled by

$$\mathcal{B}(J/\psi \to \mu^+\mu^-) = 5.961\% \tag{75}$$

from [80], obtaining

$$\mathcal{B}(\Lambda_b \to pK^- J/\psi(\to \mu^+\mu^-)) = 1.81 \times 10^{-5} \tag{76}$$

Hence, the estimated BF turns out to be

$$\frac{\mathcal{B}(\Lambda_b \to pK^-\mu^+\mu^-)}{\mathcal{B}(\Lambda_b \to pK^- J/\psi(\to \mu^+\mu^-))} = 1.57\% \tag{77}$$

B PID resampling

PID is an important aspect of the LHCb detector capabilities very useful to obtain a clean sample of signal Λ_b events. It was used in Sec. 4.1 to veto peaking background components. PID is also a powerful input to the multivariate classifier described in Sec. 6. However, it is important to ensure a good agreement between data and MC PID distributions, so as to not introduce biases. In order to improve the agreement, the distributions for each PID variable is resampled using the `MCResampling` tool of the `PIDCalib` packages.
The PID resampling is done in two steps. First, histograms of the PID variables are produced in bins of `nTracks`, pseudorapidity and p_{T} using standard `PIDCalib` data samples. These include $D^{*+} \to D^0(\to K^-\pi^+)\pi^-$, prompt $\Lambda \to p\pi^-$ and $J/\psi \to \mu^+\mu^-$ decays. Protons from hadronic and semi-leptonic decays of Λ_c were also included. In the second step, the PID variables in the MC samples are updated using the corresponding histograms, such that the resampled MC distributions look like the data. This resampled value replaces the PID variable for the simulated event and it is used in subsequent operations. PID variables after resampling show a rather good data/MC comparison (Figs. 42,43), apart some discrepancies in DLL variables involving K^- particle.

C MC reweighting

Comparison between MC $\Lambda_b \to pK^- J/\psi$ and *s-Weighted* $\Lambda_b \to pK^- J/\psi$ data highlight sizeable differences in some of the variable distributions employed for the MVA selection (Fig. 43), which is a known issue due to the incorrect simulation of Λ_b production (observed also in [64]).
MC variable distributions are corrected by reweighting each event in order to recover the *s-Weighted* data distributions. Reweighting is done in three sequential steps:

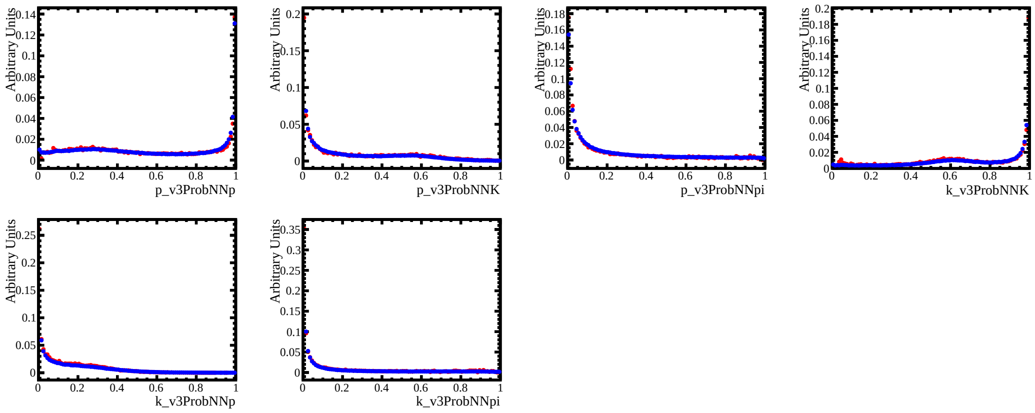


Figure 42: Comparison between MC $\Lambda_b \to pK^- J/\psi$ (blue) and *s-Weighted* $\Lambda_b \to pK^- J/\psi$ data (red) normalized distributions for ProbNN *v*3 PID variables after PID resampling.

- A first reweighting is done using `nTracks`, `nSPDHits`, $(p_{\mathrm{T}}(\Lambda_b),p(\Lambda_b))$ and $\log(\chi^2_{\mathrm{IP}}(K))$ variables, where brackets stand for a multi-dimensional reweighting.
- This preliminary weight is employed to reweight $(p(p),\mathrm{DLL}_{p\pi}(p),\mathrm{DLL}_{K\pi}(p))$ variable distributions.
- In turn, the resulting weight is used for reweighting $(p(K),\mathrm{DLL}_{K\pi}(K),\mathrm{DLL}_{p\pi}(K))$ variables and the final MC weight is produced.

The comparison between variable distributions after reweighting shows that discrepancies are satisfactorily removed (Fig. 44).

C.1 MC decay model

Throughout this analysis both the employed MC $\Lambda_b \to pK^-\mu^+\mu^-$ and $\Lambda_b \to pK^- J/\psi$ samples are generated using a phase-space model, where no underlying decay structure is simulated. Actually a rich structure is observed in the $m(pK^-)$ spectrum [64], consisting of a main $\Lambda^*(1520)$ and many other Λ^* intermediate resonances.
As a test, MC $\Lambda_b \to pK^- J/\psi$ dataset is reweighted to the $m(pK^-)$ spectrum derived from the amplitude fit in [64], but no sensible differences in MVA variable distributions is seen.

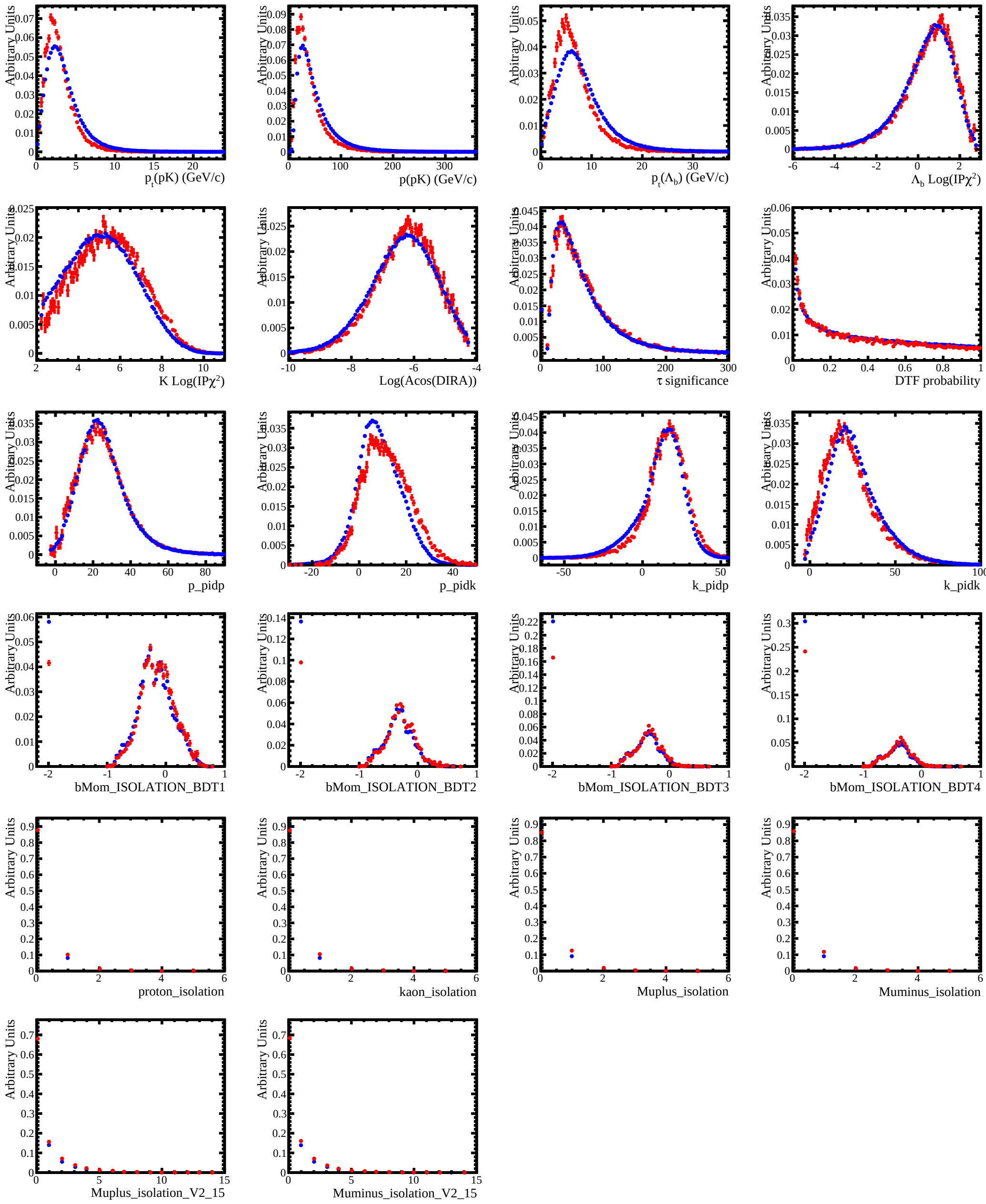


Figure 43: Comparison between MC $\Lambda_b \to pK^- J/\psi$ (blue) and *s-Weighted* $\Lambda_b \to pK^- J/\psi$ data (red) normalized distributions for MVA variables before MC reweighting, after PID resampling.

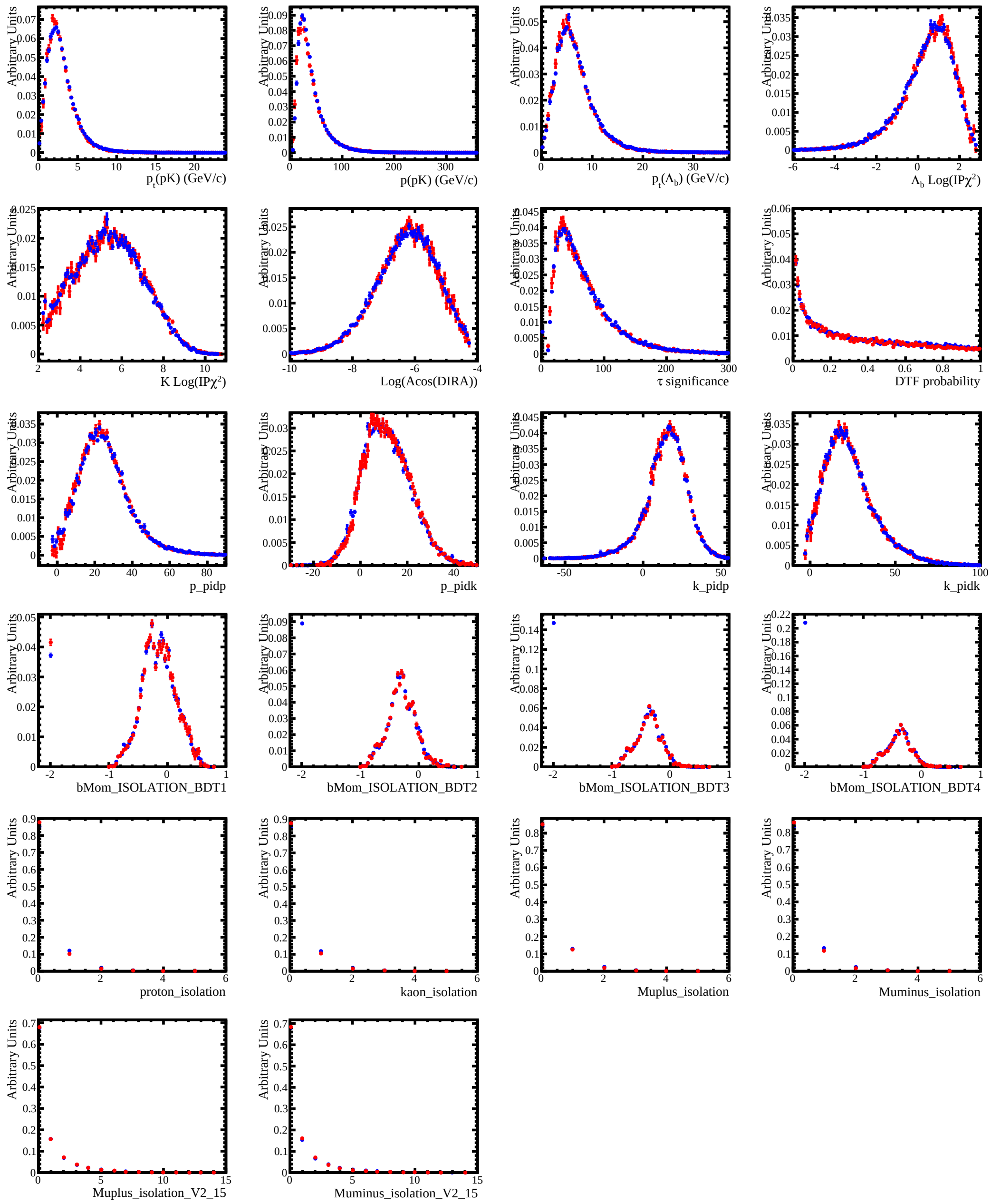


Figure 44: Comparison between MC $\Lambda_b \to pK^- J/\psi$ (blue) and *s-Weighted* $\Lambda_b \to pK^- J/\psi$ data (red) normalized distributions for MVA variables after MC reweighting.

D Variables

In Table 21 all the variables employed throughout this thesis are defined.

Variable	Notation	Description
Kinematic Variables		
Momentum	p	Modulus of the particle momentum vector $\lvert p\rvert = \sqrt{p_x^2 + p_y^2 + p_z^2}$.
Transverse momentum	p_{T}	Modulus of the particle momentum vector projection on the xy plane normal to the z-axis defined as the proton beam direction $\lvert p_{\mathrm{T}}\rvert = \sqrt{p_x^2 + p_y^2}$.
Pseudorapidity	η	Quantity describing the angle α between the particle and the proton beam direction as $\eta = -\log(\tan(\alpha/2))$. It can expressed in terms of the particle momentum as $\eta = \frac{1}{2}\log\left(\frac{p+p_z}{p-p_z}\right)$.
Isolation Variables		Defined in Sec. 6.1.
PID Variables		
Delta log likelihood	$\mathrm{DLL}_{YZ}(X)$	Combined likelihood for particle X to be a Y particle with respect to Z hypothesis, obtained by adding the likelihood information produced by each LHCb subdetector linearly.
ProbNN	$\mathrm{ProbNN}_{Y}(X)$	Probability value for particle X to be a Y particle. These variables combine the PID information produced by each LHCb subdetector taking into account correlations among them as well as additional information employing a multivariate approach (see also Sec. 3.5).
Topological Variables		
DIRA angle	DIRA	Cosine of the angle between the momentum of the particle and the flight direction from primary vertex to Λ_b decay vertex.
Distance of closest approach	DOCA	Minimum distance between two given tracks.

DTF probability	$prob$(DTF)	Probability that the fitted decay chain using `Decay Tree Fitter` tool is compatible with $\Lambda_b \to pK^-\mu^+\mu^-$ hypothesis.
Impact parameter	IP	Minimum distance between a given track and a reference vertex (in this analysis the primary vertex).
IP χ^2	χ^2_{IP}	IP given as a χ^2 with respect to the zero IP hypothesis.
τ significance	τ_S	τ/τ_{err} ratio, where τ is the particle proper lifetime and τ_{err} is its measurement error.
Track Variables		
Ghost probability		Probability that the reconstructed track is a fake (ghost) track.
`hasRich`		Boolean decision true when a given track has hits in RICH detectors (see Sec. 2.2.2).
`isMuon`		Boolean decision true if a given track is compatible with hits in the muon system (details in Sec. 3.5).
`nTracks`		Number of total identified tracks in one event as output of the whole tracking algorithm.
`nSPDHits`		Number of hits in the scintillator pad detector (SPD, see Sec. 2.2.5).
PV separation χ^2		Separation between the given vertex and the primary vertex expressed as χ^2.
Track χ^2/ndf		χ^2 per degree of freedom of the fitted track.
Vertex χ^2/ndf		χ^2 per degree of freedom of the fitted vertex.

References

[1] E. Noether, *Invariante variationsprobleme*, Gott. Nachr. **1918** (1918) 235, arXiv:physics/0503066.

[2] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191.

[3] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961), no. 4 579 .

[4] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[5] A. Salam, *Weak and electromagnetic interactions*, Originally printed in N. Svartholm: Elementary particle theory, Proceedings of the Nobel symposium 1968, Lerum, Sweden (1968) 367 .

[6] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, Nuclear Physics B **44** (1972), no. 1 189 .

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321.

[8] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508.

[9] M. S. Sozzi, *Discrete symmetries and CP violation: from experiment to theory*, Oxford Graduate Texts, Oxford Univ. Press, New York, NY, 2008.

[10] T. D. Lee and C. N. Yang, *Parity nonconservation and a two-component theory of the neutrino*, Phys. Rev. **105** (1957) 1671.

[11] C. S. Wu *et al.*, *Experimental test of parity conservation in beta decay*, Phys. Rev. **105** (1957) 1413.

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

[13] O. W. Greenberg, *CPT violation implies violation of Lorentz invariance*, Phys. Rev. Lett. **89** (2002) 231602, arXiv:hep-ph/0201258.

[14] N. Cabibbo, *Unitary symmetry and leptonic decays*, Physical Review Letters **10** (1963) 531.

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

[16] M. Kobayashi and T. Maskawa, *CP-Violation in the renormalizable theory of weak interaction*, Progress of Theoretical Physics **49** (1973) 652.

[17] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[18] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991) 392.

[19] H. Nunokawa, S. J. Parke, and J. W. F. Valle, *CP violation and neutrino oscillations*, Prog. Part. Nucl. Phys. **60** (2008) 338, `arXiv:0710.0554`.

[20] R. D. Peccei and H. R. Quinn, *CP conservation in the presence of pseudoparticles*, Phys. Rev. Lett. **38** (1977) 1440.

[21] H. Banerjee, D. Chatterjee, and P. Mitra, *Is there still a strong CP problem?*, Physics Letters B **573** (2003) 109 .

[22] LHCb collaboration, R. Aaij *et al.*, *Evidence for CP violation in $B^+ \to p\overline{p}K^+$ decays*, Phys. Rev. Lett. **113** (2014) 141801, `arXiv:1407.5907`.

[23] G. Durieux and Y. Grossman, *Probing CP violation systematically in differential distributions*, Phys. Rev. **D92** (2015), no. 7 076013, `arXiv:1508.03054`.

[24] A. K. Alok *et al.*, *New physics in $b \to s\mu^+\mu^-$: CP-violating observables*, JHEP **11** (2011) 122, `arXiv:1103.5344`.

[25] S. D. Rindani, *CP violation at colliders*, Pramana **45** (1995) S263, `arXiv:hep-ph/9411398`.

[26] A. Datta and D. London, *Measuring new physics parameters in B penguin decays*, Phys. Lett. **B595** (2004) 453, `arXiv:hep-ph/0404130`.

[27] A. Datta and D. London, *Triple-product correlations in $B \to V_1V_2$ decays and new physics*, Int. J. Mod. Phys. **A19** (2004) 2505, `arXiv:hep-ph/0303159`.

[28] G. Valencia, *Angular correlations in the decay $B \to VV$ and CP Violation*, Phys. Rev. **D39** (1989) 3339.

[29] LHCb collaboration, R. Aaij *et al.*, *Search for CP violation using T-odd correlations in $D^0 \to K^+K^-\pi^+\pi^-$ decays*, JHEP **10** (2014) 005, arXiv:1408.1299.

[30] C. Albajar *et al.*, *First observation of the beauty baryon Λ_b in the decay channel $\Lambda_b \to \Lambda J/\psi$ at the CERN proton-antiproton collider*, Physics Letters B **273** (1991), no. 4 540 .

[31] LHCb collaboration, R. Aaij *et al.*, *Study of beauty hadron decays into pairs of charm hadrons*, Phys. Rev. Lett. **112** (2014) 202001, arXiv:1403.3606.

[32] 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, arXiv:1402.6242.

[33] CDF Collaboration, T. Aaltonen *et al.*, *Measurements of direct CP violating asymmetries in charmless decays of strange bottom mesons and bottom baryons*, Phys. Rev. Lett. **106** (2011) 181802.

[34] LHCb collaboration, R. Aaij *et al.*, *Searches for Λ_b^0 and Ξ_b^0 decays to $K_S^0 p\pi^-$ and $K_S^0 pK^-$ final states with first observation of the $\Lambda_b^0 \to K_S^0 p\pi^-$ decay*, JHEP **04** (2014) 087, arXiv:1402.0770.

[35] 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.

[36] CDF, T. Aaltonen *et al.*, *Observation of the baryonic flavor-changing neutral current decay $\Lambda_b \to \Lambda\mu^+\mu^-$*, Phys. Rev. Lett. **107** (2011) 201802, arXiv:1107.3753.

[37] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular anaysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, JHEP **06** (2015) 115, arXiv:1503.07138.

[38] O. S. Bruning *et al.*, *LHC design report Vol.1: the LHC main ring*, CERN-2004-003-V-1.

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

[40] 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, arXiv:1009.2731.

[41] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011. LHCb-TDR-005.

[42] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037. LHCb-TDR-003.

[43] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024. LHCb-TDR-006.

[44] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029. LHCb-TDR-008.

[45] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007. LHCb-TDR-001.

[46] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036. LHCb-TDR-002.

[47] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010. LHCb-TDR-004.

[48] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, `arXiv:1306.0249`.

[49] LHCb collaboration, *LHCb online system, data acquisition and experiment control: Technical Design Report*, CERN-LHCC-2001-040. LHCb-TDR-007.

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

[51] G. Barrand *et al.*, *GAUDI: A software architecture and framework for building HEP data processing applications*, Computer Physics Communications **140** (2001), no. 12 45 , CHEP2000.

[52] LHCb collaboration, *The Gaudi Project*, `http://proj-gaudi.web.cern.ch/proj-gaudi/`.

[53] LHCb collaboration, *The Brunel Project*, `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/`.

[54] LHCb collaboration, *The DaVinci Project*, `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/`.

[55] LHCb collaboration, *The Gauss Project*, `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/gauss/`.

[56] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.

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

[58] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[59] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[60] LHCb collaboration, *The Boole Project*, `http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/boole/`.

[61] R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, LHCb-PUB-2011-017.

[62] A. Puig, *The LHCb trigger in 2011 and 2012*, LHCb-PUB-2014-046.

[63] G. Lanfranchi *et al.*, *The muon identification procedure of the LHCb experiment for the first data*, LHCb-PUB-2009-013.

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

[65] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, `arXiv:1210.6861`.

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

[67] M. Oreglia, *A Study of the Reactions $\psi' \to \gamma\gamma\psi$*, PhD thesis, SLAC, 1980.

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

[69] LHCb collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Physics **11** (2015) 743, arXiv:1504.01568.

[70] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[71] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Meth. **A543** (2005) 577, arXiv:physics/0408124.

[72] P. Speckmayer, A. Hocker, J. Stelzer, and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, J. Phys. Conf. Ser. **219** (2010) 032057.

[73] G. Punzi, *Sensitivity of searches for new signals and its optimization*, in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (L. Lyons, R. Mount, and R. Reitmeyer, eds.), p. 79, 2003. arXiv:physics/0308063.

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

[75] LHCb collaboration, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **09** (2015) 179, arXiv:1506.08777.

[76] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104, arXiv:1512.04442.

[77] R. E. Schapire and Y. Freund, *A decision-theoretic generalization of on-line learning and an application to boosting*, Jour. Comp. and Syst. Sc. **55** (1997) 119.

[78] LHCb collaboration, R. Aaij *et al.*, *Measurements of the $\Lambda_b^0 \to J/\psi\Lambda$ decay amplitudes and the Λ_b^0 polarisation in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Lett. **B724** (2013) 27, arXiv:1302.5578.

[79] L. Mott and W. Roberts, *Rare dileptonic decays of Λ_b in a quark model*, Int. J. Mod. Phys. A **27** (2012) 1250016, arXiv:1108.6129.

[80] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38** (2014) 090001.

[81] LHCb collaboration, R. Aaij *et al.*, *Study of the productions of Λ_b^0 and $\bar{B}^0$ hadrons in pp collisions and first measurement of the $\Lambda_b^0 \to J/\psi p K^-$ branching fraction*, Chin. Phys. C **40** (2015) 011001, `arXiv:1509.00292`.

Acknowledgements

Having reached the end of the long ascent of my master degree, I want to look back for a while to thank everyone helped me in reaching this important goal.
I am grateful to Nicola, who guided me through this long and lively analysis and to Biplab for his essential collaboration. Thanks also to Fernando for the thesis revision.
Many thanks to the LHCb Milano group, in particular to Andrea and Jinlin for their fruitful suggestions. I am also indebted to Angelo, my supervisor as CERN summer student, for having introduced me to LHCb data analysis.
Moreover, I would like to acknowledge both INFN and CERN institutions for the opportunity of working in a cutting-edge experiment like LHCb and for their practical support.
Special thanks to all my friends, who made me understand that everything is very serious as nothing is really serious and that often foolishness is more fruitful than wisdom.
Finally, very special thanks to my parents, whose special support made all this possible.