EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH (CERN)

LHCb-THESIS-2015-0XX
December 22, 2014

Measurement of the B_c^+ meson lifetime using $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays with the LHCb detector at CERN

Lucio Anderlini *Supervisor:* Giacomo Graziani

INFN and Università degli Studi di Firenze

Abstract

The precision measurement of the B_c^+ meson lifetime provides an essential test of the models describing the unique open-flavour state composed of two heavy quarks. It is also a necessary input for all measurements of B_c^+ production and decay branching fractions. The first measurement of the B_c^+ lifetime achieved by the LHCb Collaboration is presented in this Thesis. The data sample collected in 2012, in pp collisions at a centre-of-mass energy of 8 TeV, and corresponding to an integrated luminosity of 2 fb^{-1}, is analysed to select $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays. A two-dimensional data-model is developed combining the information on the invariant mass of the $J/\psi\,\mu^+$ combination and the decay time, as measured in the rest-frame of the $J/\psi\,\mu^+$ combination. Data-driven techniques are proposed to model the background sources, including the candidates selected because of the misidentification of a hadron as the muon produced in a B_c^+ decay, for which an original technique has been developed. The template distribution for $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays relies on realistic dynamical models including feed-down decays, and depends on the B_c^+ lifetime through a statistical correction between the B_c^+ and $J/\psi\,\mu^+$ rest frames, known as k-factor. Data-driven cross-checks are used to test the dynamical model and assess the related uncertainties. The measured lifetime is $\tau_{B_c^+} = 509 \pm 8\,(\text{stat}) \pm 12\,(\text{syst})$ fs, where the largest systematic uncertainty is due to the statistical model used to describe the combinatorial background, which is the only one relying on simulation.

UNIVERSITÀ
DEGLI STUDI
FIRENZE

Dipartimento di Fisica e Astronomia
Scuola di Dottorato in Scienze

Tesi di Dottorato di Ricerca in Fisica e Astronomia
Indirizzo Fisica — XXVII ciclo

Measurement of the B_c^+ meson lifetime using $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays with the LHCb detector at CERN

Settore Scientifico Disciplinare: FIS/04

Dottorando:
Lucio Anderlini

Tutore:
Dr. Giacomo Graziani

Coordinatore:
Prof. Roberto Livi

Anni 2012/14

Preface

Since the first time I entered CERN in 2009 I have been fascinated by particle physics, and in particular by its experimental aspects. It is like if the front-edge knowledge and technologies of all the fields of science converged in a single laboratory to be applied to some of the most intriguing questions of our age.

From the physics of the electronic silicon devices to the data-acquisition apparatus; from the internal structure of the proton to the technologies to accelerate bright proton beams to unprecedented energies; from the network technologies used for securing the access to personal and confidential data to the infrastructure of the LHC Computing Grid spreading petabytes of data worldwide; from the computer vision algorithms applied to tracking and particle identification to imaging applied to radio-diagnostic and radio-therapy; from the differential geometry analysis applied to effective models of QCD to the applied statistics to set confidence intervals and limits. The variety of subjects and challenges is astonishing.

During these five years of collaboration with the LHCb experiment at CERN with various assignments and roles, I had the great chance of getting in touch with most of these fields.

This Thesis is devoted to the lifetime measurement of the B_c^+ meson, my main activity during the three years of the Ph.D. programme. The first four chapters are intended to describe the context of the analysis, spanning on the many front-edge technologies and techniques contributing to the final result. The following chapters describe the original work I carried out to achieve this challenging measurement at LHCb.

The B_c^+ meson is the only ground state meson in the Standard Model being composed of two heavy quarks of different flavour. It is therefore a unique state that can be described by the same effective models developed for quarkonium physics, but decays only weakly. QCD-inspired effective theories are used to predict the partial decay widths of the B_c^+ decay channels, and therefore its lifetime. Since different theoretical approaches yield to expectations spanning an order of magnitude, high experimental precision on the lifetime measurement is a benchmark for the theoretical models and their underlying assumptions. The production of the B_c^+ meson is forbidden or suppressed at e^+e^- colliders, so that its experimental history is recent. Discovered at CDF in 1998, it is being studied intensively for the first time by the LHC experiments. A precise lifetime measurement is crucial for a correct assessment of the efficiency of criteria related to the detachment of the B_c^+ decay vertex from the pp collision primary vertex. The uncertainty on the B_c^+ lifetime is therefore an important systematic effect on most of the production and branching fraction

measurements.

The measurement of the B_c^+ lifetime was performed in the past by experiments at TeVatron using both the semileptonic channel $B_c^+ \to J/\psi\,\ell^+\nu_\ell X$ and the fully reconstructed hadronic channel $B_c^+ \to J/\psi\,\pi^+$. The two analyses are very different. The former relies on a larger statistics because of the large branching fraction of the $B_c^+ \to J/\psi\,\ell^+\nu_\ell X$ decay modes, but requires techniques to deal with partially reconstructed decays, since the lifetime is defined in the rest frame of the B_c^+ meson, which can only be approximated by the rest frame of the $J/\psi\,\ell^+$ combination. The hadronic channel is limited by the statistical uncertainty due to the small branching fraction of the decay $B_c^+ \to J/\psi\,\pi^+$, and it requires a signal selection strategy based on variables correlated to the flight distance, making the efficiency a non-trivial function of the decay time. On the other hand, the model of signal and background contributions is much simpler for the hadronic channel, because it relies on the narrow peak in the reconstructed mass of the $J/\psi\,\pi^+$ combination.

The analysis presented here, performed studying the semileptonic $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay mode, relies on the much larger B_c^+ production rate available at LHC and on original techniques to build data-driven models of most of the background sources. The model-dependence introduced by the statistical correction from the $J/\psi\,\mu^+$ rest frame to the B_c^+ rest frame is based on realistic dynamic models including the description of the decay matrix elements, and decays to higher charmonium states decaying to J/ψ. The theoretical description is validated with an original data-driven technique allowing to constrain the model using the whole kinematic information measured from the decays.

The result obtained,

$$\tau_{B_c^+} = 509 \pm 8\,(\text{stat}) \pm 12\,(\text{syst})\,\text{fs}, \tag{1}$$

is the first LHCb measurement of the B_c^+ lifetime, with an uncertainty less than half of that of the previous world average [1].

The first Chapter introduces the theoretical framework of the heavy flavour physics and in particular the theoretical techniques used to predict b-hadron lifetimes and properties of the B_c^+ meson. The second chapter focuses on the experimental results on B_c^+ physics obtained at Tevatron and at the LHC. The description of the LHC and of the LHCb detector is the subject of Chapter III, while the physics software is described in Chapter IV together with an introduction to the LHCb Computing model. The same Chapter includes the description of my contributions to the *Stripping* project, the last centralized step of the data reduction process. Chapters V, VI, and VII are devoted to the original analysis I developed to measure the lifetime of the B_c^+ meson with semileptonic decays, and in particular on the analysis strategy, data model, and systematic uncertainty respectively. Finally in Chapter VIII, I compare the result with other measurements, draw the conclusions and discuss the possible outlook of this study.

Contents

1 *Introduction* **1**

I Elements of theory **3**
I.1 A brief introduction to the quark model . . . 5
I.2 Quantum Chromodynamics . . . 8
I.3 Weak decay of quarks and quark mixing . . . 10
I.3.1 Discrete symmetries and their conservation . . . 11
I.4 Quarkonium states . . . 14
I.4.1 Potential models . . . 15
I.4.2 Quarkonium Decays . . . 15
I.5 The B_c^+ meson . . . 16
I.5.1 Theoretical prediction of the B_c^+ mass . . . 17
I.5.2 Theoretical considerations on the B_c^+ meson production . . . 18
I.5.3 Theoretical description of the B_c^+ meson decays . . . 22
I.5.4 Prospects for studies of CP violation in the $(\bar{b}c)$ system . . . 25
I.5.5 Theoretical framework for the b-hadron lifetimes . . . 28
I.5.6 Predictions of the B_c^+ lifetime . . . 36

II Experimental advances on B_c^+ physics **41**
II.1 Mass measurements . . . 44
II.2 Production measurements . . . 47
II.3 Observation of an excited B_c^+ meson state . . . 48
II.4 B_c^+ decays . . . 50
II.4.1 Observation of $B_c^+ \to B_s^0\pi^+$ decays . . . 50
II.4.2 Decays to J/ψ and hadrons . . . 51
II.4.3 Decays to higher charmonia states: $B_c^+ \to \psi(2S)\pi^+$ . . . 51
II.4.4 Decay $B_c^+ \to J/\psi\, D_s^+$ . . . 52
II.4.5 Semileptonic B_c^+ decays . . . 53
II.5 B_c^+ lifetime measurements, state-of-the-art . . . 53
II.5.1 B_c^+ lifetime measurement using the channel $B_c^+ \to J/\psi\,\pi^+$ . . . 54
II.5.2 B_c^+ lifetime measurements using the channel $B_c^+ \to J/\psi\,\ell^+\nu$ . . . 56
II.5.3 Summary of the B_c^+ lifetime measurements . . . 58

III The Large Hadron Collider and the LHCb experiment **61**
III.1 The Large Hadron Collider . . . 61

III.1.1 The LHC accelerator system . . . 62
III.1.2 The large experiments at the LHC . . . 63
III.1.3 The $b\bar{b}$ production cross section . . . 63
III.1.4 The luminosity and the beam time structure . . . 65
III.2 The LHCb experiment . . . 66
III.2.1 The LHCb detector . . . 68
III.2.2 The LHCb data-taking . . . 89
III.3 Trigger strategy and performance . . . 89
III.3.1 Level-0 trigger . . . 91

IV LHCb physics software and computing model 93
IV.1 Overview and dataflow . . . 93
IV.2 Physics applications . . . 96
IV.2.1 The architecture: Gaudi . . . 97
IV.2.2 High Level Trigger . . . 98
IV.2.3 Reconstruction: vertex fit and decay time measurement . . . 103
IV.2.4 Offline selection (Stripping) . . . 110
IV.3 The LHC Computing Grid . . . 113
IV.3.1 The DIRAC Middleware . . . 115

2 B_c^+ *meson lifetime measurement* 117

V Analysis strategy 121
V.1 Hadronic vs. semileptonic channel . . . 122
V.2 Signal selection . . . 125
V.3 Multiple candidates . . . 128
V.4 Specialized muon identification algorithms . . . 131
V.4.1 Kalman filter . . . 132
V.4.2 Muon Isolation . . . 133
V.4.3 Decays in flight . . . 135
V.4.4 Performance of muon identification algorithms . . . 135

VI Data model 139
VI.1 Signal model . . . 140
VI.1.1 Generator-level distributions . . . 140
VI.1.2 Pseudo-proper decay time . . . 141
VI.1.3 k-factor technique . . . 144
VI.1.4 Feed-down decays . . . 147
VI.2 Misidentification background . . . 150
VI.2.1 Rejection . . . 150
VI.2.2 Model . . . 152
VI.2.3 Construction of the data model . . . 154
VI.2.4 Uncertainties . . . 156

VI.3 Other Background sources and their models . . . 157
VI.3.1 Fake J/ψ background . . . 157
VI.3.2 Combinatorial background . . . 158
VI.4 Fit to data . . . 161
VI.4.1 Fit result . . . 162
VI.4.2 Cross-checks on the statistical uncertainty . . . 163
VI.4.3 Assessment of the goodness of fit . . . 166

VII Systematic uncertainties **169**
VII.1 Systematic uncertainties on the signal model . . . 169
VII.1.1 Data-driven constraints on the theoretical model . . . 169
VII.1.2 Effects of the reconstruction on the signal model . . . 172
VII.2 Uncertainties on the background model . . . 173
VII.2.1 Uncertainties on the combinatorial prompt background . . . 173
VII.2.2 Uncertainties on the combinatorial detached background . . . 175
VII.2.3 Uncertainties on the fake-J/ψ background . . . 176
VII.2.4 Uncertainties on the misidentification background . . . 177
VII.3 Reconstruction and fitting techniques . . . 177
VII.3.1 Validation of fit technique . . . 177
VII.3.2 Momentum scale calibration . . . 177
VII.3.3 Length scale calibration . . . 178
VII.3.4 Dependence of the efficiency on decay time . . . 178
VII.3.5 Uncertainty associated to multiple candidates . . . 179
VII.3.6 Uncertainty due to incorrect primary vertex associations . . . 179
VII.4 Summary of the systematic uncertainty . . . 180
VII.5 Further cross-checks . . . 180

VIII Conclusion and outlook **187**

3 *Supplementary material* **193**

A Density Estimation Trees **195**
A.1 The algorithm . . . 197
A.1.1 Decision Trees . . . 197
A.1.2 Training (or learning) . . . 199
A.1.3 Pruning and cross-validation . . . 201
A.1.4 Smoothing techniques . . . 207
A.2 Implementation . . . 209
A.3 Tests and applications . . . 210
A.3.1 Effect of pruning . . . 212
A.3.2 Smoothing . . . 214
A.4 Timing . . . 216
A.5 Conclusions . . . 218

Acknowledgements **219**

References **221**

Part 1

Introduction

Chapter I Introduction

Elements of theory

The Standard Model of Particle Physics, indicated as SM hereafter, is a set of theoretical models able to explain all the physics phenomena that we observe at the accelerator and cosmic-ray experiments, and most of the physics phenomena in the Universe as we observe them, for example, through telescopes.

The SM is a quantum field theory representing particles as spin-$\frac{1}{2}$ fields and their interaction as spin-1 fields. The recent discovery of the Higgs boson by the ATLAS and CMS Collaborations at the LHC has confirmed the existence of a third kind of field, scalar, defining the inertial properties of matter and interaction fields.

Gravity is the unique known interaction which is not included in the SM since it is not clear whether it can be properly described as a quantum field theory. The other interactions are the Electromagnetic interaction, associated to neutral massless mediators named *photons* (γ), the Nuclear Weak interaction, associated to the charged massive mediator $W^{\pm}$ and the neutral massive mediator Z^0, and the Strong Nuclear interaction associated to neutral massless mediator *gluon* (g).

Two categories of fermions are defined: *leptons* and *quarks*. Leptons are elementary particles which do not interact through the Strong interaction. The electron (e^-), essential constituent of the ordinary matter and responsible for all the chemical properties of elements, is the lightest charged lepton. Two other charged leptons are known: the *muon* (μ^-), and the *tauon* (τ^-). They have all the same unity charge, $-e$. While no electron decay is neither observed nor expected in the SM, μ^- and τ^- leptons can decay to lighter leptons through weak interaction. Such decays involve the emission of neutral "leptons", interacting only through weak interaction, called *neutrinos*. Weak interaction conserves a quantum number, named "lepton flavour", shared by charged and neutral leptons. The three neutrino states corresponding to the three charged lepton are ν_e, ν_μ, and ν_τ, corresponding to the three charged leptons.

Though in the original formulation of the SM, neutrinos were considered massless, it is today clear that at least two of the neutrino eigenstates are massive. Indeed, the process of neutrino oscillations, namely the variation of the neutrino flavour during its time evolution, has been observed and confirmed, and it is only possible if the eigenstates

are a linear combination of the states with defined flavour ν_e, ν_μ, and ν_τ, and if their masses are different. Despite these recent developments, the masses of SM neutrinos, and their differences, are (almost) always negligible in the accelerator experiments, so that neutrinos are considered massless in the computations included in this work.

As opposed to leptons, *quarks* can interact through the strong interaction. The *quantum number* defining this interaction is named *color*, and the theory describing how quarks interact through the strong force is named *Quantum Chromo-Dynamics* (QCD). An important difference between the strong and electromagnetic interactions is the self-interaction of the mediator particles: possible for gluons, impossible for photons. Self-interaction explains why strong-interaction bound states, named *hadrons*, are so different from electromagnetic bound states, for example *atoms*, and why the constituents of the latter (electrons and nuclei) can be observed as free, while quarks can only be observed within hadrons. In the ordinary matter, the most common elementary particles are the *down* quark (d), with charge $-\frac{1}{3}e$, and the *up* quark (u), having charge $+\frac{2}{3}e$. Indeed *up* and *down* quarks are the so-called *valence quarks* constituting the nucleons: protons and neutrons. Actually, because of the nature of the interaction binding quarks, the nucleons contain additional quark–anti-quark pairs named *sea-quarks* which do not change the properties of the nucleon as the charge or the spin, but are very important to describe its interactions.

Protons are bound states *uud*, while neutrons are bound states *udd*. The property of a quark of being *up* or *down* is named hadronic *flavour*. Six different *flavours* exist for quarks: *down* (d), *up* (u), *strange* (s), *charm* (c), *bottom* (b), and *top* (t). For the two latter flavours the alternative names *beauty* and *truth* are often used.

For any elementary matter particle, a respective antiparticle exists. For example the positron (e^+) is the anti-particle of the electron (e^-), while an *up*-quark has a respective anti-quark *anti-up* ($\bar{u}$). Excluding the charge, (almost) all the properties of a particle are (almost) the same as those of the respective anti-particle. So, for example, anti-quarks interact through strong interaction and it is possible to observe strong-interaction bound states of a quark and an anti-quark, named *mesons*. As opposed to bound states of three quarks (named *baryons*), or three anti-quark (*anti-baryons*), no example of infinite-lifetime (*stable*) meson is available: any known meson can decay.

The number of flavours, the complexity of QCD, and the existence of bound states of particles and anti-particles makes the variety of hadrons huge. Because of this complexity, most of the rest of this introductory chapter is devoted to the description of hadrons and effective theories aiming to a simplification of QCD at the cost of a reduction of the applicability range.

Until June 2012, the existence of a scalar field inferring the inertial properties (mass) of the elementary particles, was predicted by the SM, but not proved experimentally.

This scalar field is named *Brout-Englert-Higgs* (BEH) field, after the names of the authors of two independent works in 1964 describing a mechanism attributing mass to the weak-interaction mediators $W^\pm$, and Z^0. During the review of his paper, Peter Higgs were suggested to include a sentence about the possible excitation of this field, manifestation of the mechanism itself. Since in terms of quantum field theories particles are excitations

of fields, the excitation of the BEH field is a particle, and since the field is scalar that particle is named a *boson*: the Higgs boson.

On July 2012, the 4$^{\text{th}}$, the ATLAS and CMS Collaborations collecting and analysing data obtained in proton-proton collisions at the *Large Hadron Collider* (LHC) have reported the first observation of the Higgs boson, corroborating the BEH mechanism after almost 50 years of studies, and confirming the prediction of the SM.

I.1 A brief introduction to the quark model

As mentioned above, quarks are strongly interacting spin-$\frac{1}{2}$ states. A quantum number, named *baryon number* is defined to be $\frac{1}{3}$ for quarks and $-\frac{1}{3}$ for anti-quarks, so that baryons have unity baryon number, meson have zero baryon number, and anti-baryon have -1 baryon number.

The six quark flavour, u, d, c, s, t and b are grouped into three families:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{I.1}$$

Quarks in the first row have charge $+\frac{2}{3}$, while those in the second row have charge $-\frac{1}{3}$.

As mentioned above, quarks can only be observed in bound states, therefore it is hard to define properly their mass, or inertial properties. The definition of mass usually requires prescriptions which may be different for energy ranges where QCD can or can not be treated perturbatively. Still, it is well established that

$$m_u < m_d < m_s < m_c < m_b < m_t \tag{I.2}$$

and that the mass difference $m_d - m_u$ is much smaller than the difference $m_s - m_d$, which is much smaller than $m_c - m_s$. For this reason approximate symmetries assuming $m_u = m_d$ (named *isospin symmetry* or $SU(2)_F$), or $m_u = m_d = m_s$ (named $SU(3)_F$), or even $m_u = m_d = m_s = m_c$ (named $SU(4)_F$) are often helpful when classifying hadrons.

Using the formalism of group theory, one can evaluate the families of hadrons obtained combining a quark and an anti-quark having *up* or *down* flavour as

$$\mathbf{2} \otimes \bar{\mathbf{2}} = \mathbf{3} \oplus \mathbf{1}. \tag{I.3}$$

These four states are divided into two irreducible representation containing 3 elements (named pions, π^+, π^0, and π^-) and 1 element (named η) respectively. The main difference between the two irreducible representations is in the symmetric properties of the states exchanging $u \leftrightarrow d$, and then exchanging every particle with its own antiparticle (G-parity). If the isospin symmetry was perfect, *i.e.* if $m_u = m_d$, then $m(\pi^+) = m(\pi^-) = m(\pi^0)$, while experimentally one measures [1]

$$\begin{aligned} m(\pi^\pm) &= 139.57018 \pm 0.00035 \text{ MeV}/c^2, \\ m(\pi^0) &= 134.9766 \pm 0.0006 \text{ MeV}/c^2. \end{aligned}$$

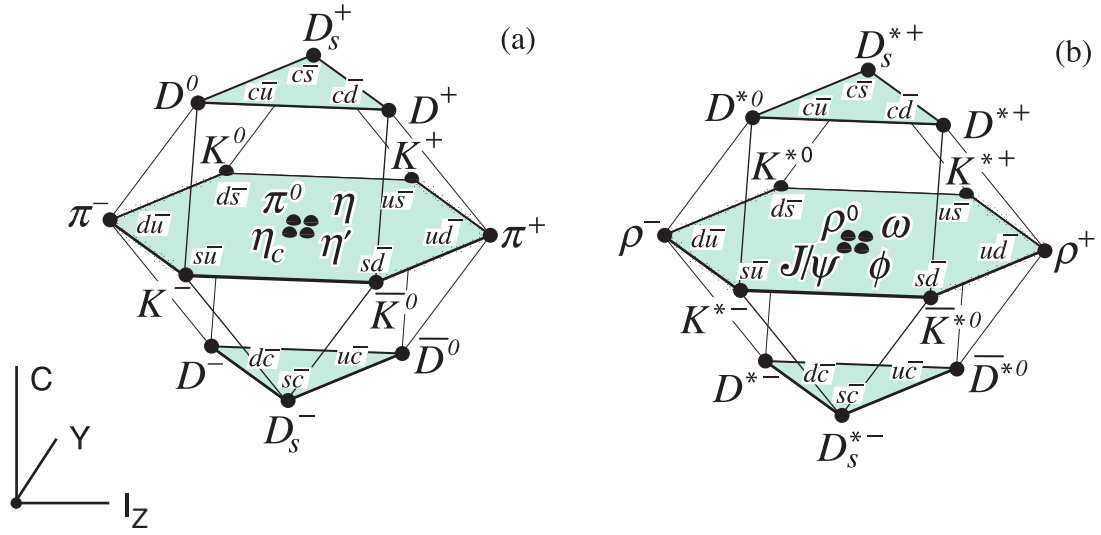

Figure I.1: $SU(4)_F$ weight diagram showing the 16-plets for the pseudoscalar (a) and vector meson (b) made of u, d, s, and c quarks as function of the isospin I, charm, and hypercharge Y. The nonet of light mesons composed of u, d, and s quarks, occupy the central planes to which the $c\bar{c}$ have been added. The figure was published by the Particle Data Group [1].

the small difference shows that the isospin symmetry is broken. Among the causes of isospin symmetry breaking there are the mass difference $m_d - m_u$ and electromagnetic forces.

For applications where a larger symmetry breaking is acceptable, it is common to assume $m_s = m_d = m_u$. Combining a quark and anti-quark with flavours *up*, *down*, or *strange*, one gets

$$\mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1}. \tag{I.4}$$

A fourth quark, the *charm* quark, can be included by extending the $SU(3)_F$ to $SU(4)_F$. Clearly, because of the much heavier c quark, the symmetry breaking is hardly negligible. In the $SU(4)_F$ classification, the sixteen mesons are classified in a 15-plet and a singlet:

$$\mathbf{4} \otimes \bar{\mathbf{4}} = \mathbf{15} \oplus \mathbf{1}. \tag{I.5}$$

In Figure I.1a the $SU(4)_F$ classification of pseudo-scalar meson, composed of quarks with opposite spin vectors, is shown.

An identical classification can be defined for vector mesons, composed of quarks with aligned spin vectors. Such a classification is represented in the diagram of Figure I.1b.

Besides $q\bar{q}$ bound states, QCD allows bound states formed of three quarks (baryons) or three anti-quarks (anti-baryons). As a difference with mesons, the quarks in a baryon can be identical fermionic states, and therefore symmetrical properties of the global wave function are important, since, according to the Fermi-Dirac statistics, states with symmetric global wavefunction cannot exist.

For historical reasons, and for simplicity, when treating the baryon spectroscopy the color wavefunction is always neglected. Since QCD postulates that, for all the strongly-

bound states, the color wavefunction is a totally antisymmetric singlet, the requirement for a global wavefunction to be antisymmetric translates into the requirement for the spin⊗flavour wavefunction to be totally symmetric.

Considering only *up* and *down* quarks the possible combinations in $SU(2)_F$ are

$$\mathbf{2} \otimes \mathbf{2} \otimes \mathbf{2} = \mathbf{2}_{2S} \oplus \mathbf{2}_{2A} \oplus \mathbf{4}_{3S} \tag{I.6}$$

where the first doublet is symmetric when exchanging two quarks of the same flavour, while the second is antisymmetric. The multiplet $\mathbf{4}_{3S}$ is symmetric to any exchange. Protons and neutrons are superpositions of the first two isospin doublets, where the isospin wavefunction is symmetric (antisymmetric) when the spin wavefunction is symmetric (antisymmetric), therefore the spin⊗isospin wavefunction is always symmetric. Similarly, to make quartet states to be totally symmetric, the spin of three quarks has to be aligned, and indeed the baryons in $\mathbf{4}_{3S}$ are Δ resonances, with spin $\frac{3}{2}$.

This interplay between the spin and flavour symmetries becomes more important when considering larger groups, for example including the *strange* quark in an extended $SU(3)_F$ symmetry. The baryon multiplets are:

$$\mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{1}_A \oplus \mathbf{8}_M \oplus \mathbf{8}_M \oplus \mathbf{10}_S \tag{I.7}$$

where the singlet is totally antisymmetric to the exchange of flavours, property which cannot be corrected by an $SU(2)$-spin symmetric wavefunction, because, as seen in Equation I.6, there is no totally antisymmetric singlet when combining $SU(2)$ quantities of three quarks. A totally symmetric flavour function requires a totally symmetric spin wavefunction, which translates into the spin-$\frac{3}{2}$ baryons contained in the decuplet $\mathbf{10}_S$. The eight linear combinations of the states in the two multiplet $\mathbf{8}_M$ are spin-$\frac{1}{2}$ light baryons.

The reasoning can be extended to a $SU(4)_F$ symmetry including the c-quark, the physical multiplets are represented in Figure I.2.

In principle, the symmetry could be further extended to an $SU(5)_F$ symmetry including all the quarks that can originate hadrons (the t-quark has a lifetime shorter than the characteristic time of strong interaction: and thus it cannot combine to other quarks into hadronic bound states). Within the $SU(5)_F$ symmetry, there are four open-flavour b-hadrons: $b\bar{u}$, $b\bar{d}$, $b\bar{s}$, and $b\bar{c}$, named B^+, B^0, B_s^0, and B_c^+ respectively. The latter is the heaviest ground-state meson to be open-flavoured, and is treated in further detail in the next chapter.

Beside baryons and mesons, there are other quark–anti-quark configurations which can be associated to a QCD $SU(3)_C$ singlet. These bound states have two sub-categories that are often considered as the closest alternatives to conventional hadrons. The tetraquark has zero baryon number and is a bound state ($qq\bar{q}\bar{q}$). It is believed that the state $Z_c^+(4430)$, discovered by Belle [2] and recently studied by the LHCb Collaboration [3] is the first unambiguous candidate for a tetraquark.

Pentaquark states, bound states ($qqqq\bar{q}$) are also expected, but no confirmed state is known.

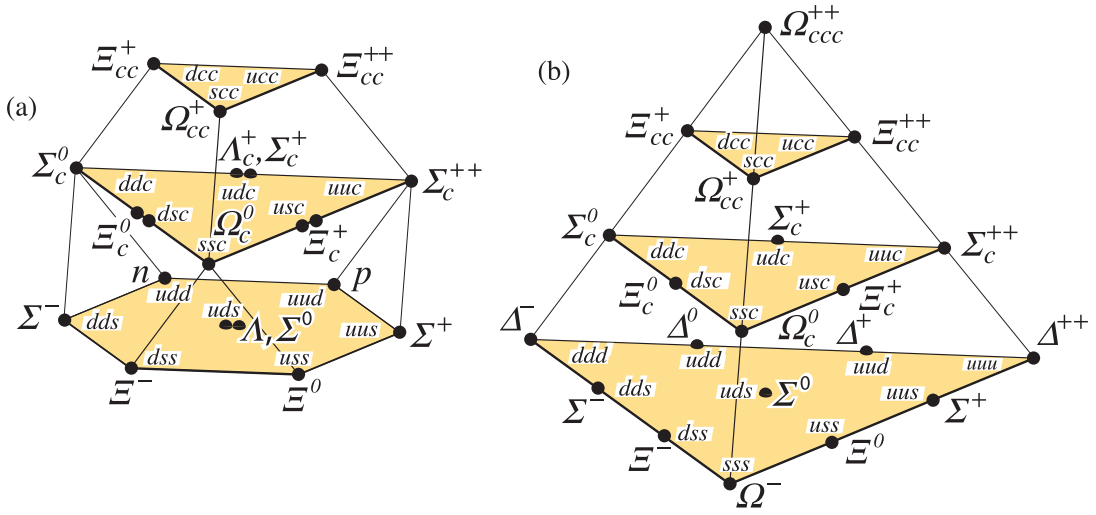

Figure I.2: $SU(4)_F$ weight diagram showing the multiplets of baryons made of u, d, s, and c quarks. On the left (a), the 20-plet with an $SU(3)_F$ octet; on the right (b), the 20-plet with an $SU(3)_F$ decuplet. The figure was published by the Particle Data Group [1].

I.2 Quantum Chromodynamics

The theory that describes strong interaction in the Standard Model is called *quantum chromodynamics* (QCD). As quantum electrodynamics (QED), the theory which describes the electromagnetic interaction, QCD is a *gauge theory* since almost all of its properties can be derived by the assumption of the existence of a symmetry which leaves the theory unvaried. This principle, named *gauge invariance*, has given amazing results describing the electromagnetic and the weak interactions where the symmetry groups involved are $U(1)$ and $SU(2)$ respectively, and is today well established for QCD, where the symmetry group is $SU(3)$. As the mediator particles of the electromagnetic interaction, the photons, interact with the electric charge, the gluons interact with the *color charge*. Leptons, photons and electroweak bosons $W^{\pm}$ and Z^0 have no color charge; quarks can be found in three color states named after the three fundamental colors *red* (r), *green* (g), and *blue* (b); anti-quarks can be found in the respective anti-color states $\bar{r}$, $\bar{g}$, and $\bar{b}$. Gluons can be found in eight states corresponding to color–anti-color combinations that allow to preserve the color charge in the interaction of a quark with a gluon.

The color symmetry follows the same algebra as the flavour symmetry described above for the u, d, and s quarks. While the flavour symmetry $SU(3)_F$ is an *approximate* symmetry obtained neglecting the different masses of the three flavours, the color symmetry $SU(3)_C$ in the SM is a symmetry assumed to be exact and preserved by all the interactions.

The quark confinement, mentioned in the introductory overview, is coded in QCD as a postulate: all hadrons are $SU(3)_C$ singlet states, having a totally anti-symmetric color wave function. Actually, color has been historically introduced to explain the existence of totally symmetric baryon states as the Δ^{++} baryon, composed of three identical, same-

spin *up*-quarks. Without color, the existence of such a baryon would violate the Pauli exclusion principle for quarks as dictated by the Fermi-Dirac statistics. Today, QCD is well established and has important consequences in the theoretical treatment of hadrons. For example, the fact that gluons carry a color charge means that they can interact with each other. The gluon-gluon interaction is a unique feature of QCD which is not present in QED and which leads to an opposite behaviour in terms of quark–gluon coupling as a function of the energy scale.

In QED, the existence of virtual e^+e^- pairs surrounding the interacting fermions have the effect of *screening* its charge. At higher energy the effect of screening is reduced, naively because the length scale of the process is smaller, being inversely proportional to the energy, hence the electromagnetic interaction is stronger at higher energy.

For QCD, the cloud of virtual particles surrounding the quark is composed mainly of gluons, interacting with each-other and with any probe particle in the neighbourhood. The effect is an *antiscreening* which makes the strong nuclear interaction weaker at higher energy. In the limit of infinite energy, the quark is said to be *asymptotically free*. This feature of QCD has important consequences as the existence of the *quark–gluon plasma*, a high-energy high-pressure state whose characterization is the main item in the physics programme of ALICE, one of the four large experiments of the LHC.

The concepts of screening, antiscreening and asymptotic freedom are translated into equations introducing a dependence on the energy scale in the coupling constants α and α_s for QED and QCD respectively. Because of this dependence coupling constants are said to be *running coupling constants*. The strong coupling constant α_s at energy scale Q^2 can be expressed relative to α_s at energy scale Q_0^2 as [1],

$$\alpha_s(Q^2) = \alpha_s(Q_0^2)\left[1 + \frac{(33-2N_f)}{6\pi}\alpha_s(Q_0^2)\ln(Q^2/Q_0^2)\right]^{-1}, \tag{I.8}$$

where N_f is the number of quark flavours, 6 in the Standard Model. Setting $\beta_0 = \frac{33-2N_f}{3}$ one can simplify the relation to

$$\alpha_s(Q^2) = \frac{1}{\beta_0 \ln \frac{Q^2}{\Lambda^2_{\mathrm{QCD}}}}, \tag{I.9}$$

hiding the dependence on the arbitrary scale Q_0^2 in the parameter

$$\Lambda^2_{\mathrm{QCD}} = Q_0^2 \exp\left[-\frac{1}{\beta_0}\frac{1}{\alpha_s(Q_0^2)}\right], \tag{I.10}$$

experimentally measured to be $\Lambda_{\mathrm{QCD}} \sim 200$ MeV.

The constant Λ_{QCD} represents a characteristic energy scale of strong interaction, for energies much larger than Λ_{QCD}, a perturbative treatment of the strong interaction is justified because α_s is much smaller than 1. Instead, for energies much smaller than Λ_{QCD}, perturbation theory cannot be used and QCD is said to be in its non-perturbative regime. The non-perturbative regime requires effective theories which spans from the simple bag

models, describing the non-perturbative states and quark confinement by dividing the space in two regions inside and outside a non-perturbative bag, to the Lattice QCD, involving the usage of large computing clusters to simulate the QCD dynamics in a discrete space-time.

I.3 Weak decay of quarks and quark mixing

While the quark flavour is conserved in the strong interactions described by the QCD Lagrangian, decays of hadrons towards lighter states with different flavour are observed, and interpreted as weak decays of quarks through the emission of a charged W boson.

In 1963, when only the u, d, and s quarks were known, Cabibbo suggested that the negative-charge state interacting with the u quark in this weak interaction process was actually a linear combination of the quarks d and s

$$d' = d\cos\theta_c + s\sin\theta_c \tag{I.11}$$

where θ_c is a mixing angle known as Cabibbo angle. In 1970, Glashow, Iliopoulos, and Maiani predicted the existence of a fourth quark, named *charm*, observing that to explain experimental data on the suppression of flavour-changing neutral currents, a cancellation mechanism involving four quark fields has to be present in the quark loop diagrams. The description of quark mixing can then be expressed by a mixing matrix as

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = V_C \begin{pmatrix} d \\ s \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \tag{I.12}$$

In 1973, Kobayashi and Maskawa to accommodate new experimental results [4] exhibiting the violation of the CP discrete symmetry (see next section) extended the mixing matrix including a third family of quarks

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{I.13}$$

It can be shown that the matrix unitarity allows to express all the matrix elements using only independent physical parameters, three mixing angles and a CP-violating phase.

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}s_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & s_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{I.14}$$

where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$, and δ is the phase responsible for the CP-violating phenomena in flavour-changing processes in the SM. The angles θ_{ij} are chosen to ensure $s_{ij}c_{ij} \geq 0$. It is known experimentally that

$$s_{13} \ll s_{23} \ll s_{12} \ll 1, \tag{I.15}$$

and it is convenient to make this hierarchy explicit using the Wolfenstein parameterization [5],

$$V_{CKM} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\bar{\rho}-i\bar{\eta}) \\ \lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3\big(1-(\bar{\rho}+i\bar{\eta})\big) & A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{I.16}$$

The fit of the Wolfenstein parameters gives [1]

$$\begin{array}{ll} \lambda = 0.22537 \pm 0.00061, & A = 0.814^{+0.023}_{-0.024}, \\ \bar{\rho} = 0.117 \pm 0.021, & \bar{\eta} = 0.353 \pm 0.013. \end{array} \tag{I.17}$$

The magnitudes of the nine CKM elements are

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{pmatrix}. \tag{I.18}$$

I.3.1 Discrete symmetries and their conservation

Modern theoretical physics exploits the concept of symmetry to derive physical and observable effects from very general assumptions about Nature. For example, momentum conservation is considered to be a consequence of the translational invariance of the space-time. Analogous invariances are exploited to state conservation of energy and angular momentum.

In Quantum Field Theory (QFT) it is usual to distinguish between continuous and discrete symmetries. Continuous symmetries involve operators acting on continuous variables, as for example time evolution, position translations, and rotations, whereas discrete symmetries relate a numerable and finite set of states. The Emmy Noëther's theorem states that any continuous symmetry corresponds to a physical quantity conserved in the evolution of the system, justifying the strong interest of modern physics in symmetries.

The fundamental discrete symmetries are C, P and T. C is the charge conjugation symmetry, and relates two states, the second representing the antiparticle of the first. The P symmetry changes the sign of the three space coordinates $\mathbf{x} \to -\mathbf{x}$. Both C and P operators have two eigenvalues: $+1$ and -1. Finally the T symmetry or Time-Reverse parity is a symmetry between states differing for the sign of the temporal coordinate: the two states have opposite time-evolution directions. There are other discrete symmetries, as the isospin-related G-parity and supersymmetric symmetry R, which are not treated further in this Thesis.

Symmetries can be combined to derive other relations between states. For example, the CP symmetry is the combination of charge conjugation and parity symmetry. The interest of this combination arises from the experimental evidence that C and P symmetries are maximally violated by the nuclear weak interaction, *e.g.* the C- or P- transformed neutrino

field, always produced with negative helicity, corresponds to a left-handed anti-neutrino or to a right-handed neutrino, respectively, which are both unobserved in nature. The combined CP operator, relating the two fields seems to recover a reassuring symmetry. However a small violation of the CP symmetry was observed in the kaon system in 1964 and received with surprise by the scientific community, because CP-violation is associated to the violation of the T symmetry. The CPT symmetry is still considered to be an exact symmetry, because it constitutes a fundamental property of any Quantum Field Theory, so that renouncing to CPT invariance means to renounce to QFT, which has given remarkable results in the latest century. No CPT violation has been observed experimentally, so far.

The violation of P symmetry was observed in 1956 by C. S. Wu *et al.* [6] in the β decays of ^{60}Co nuclei. In 1964, CP violation was first observed by J. H. Christenson *et al.* [4] in the neutral kaon system. It is only in 2001 that BaBar and Belle experiments both observed CP violation also in the neutral B meson system [7, 8].

The experimental studies on CP violation are still very active, since testing the CKM paradigm for quark mixing which is a non-trivial prediction of the SM, allows to set stringent constraints on New Physics scenarios.

In the SM, the only source for CP violation is the phase of the CKM matrix which cannot be absorbed in the arbitrary and unphysical phase of the quark fields.

The matrix is unitary by construction: $V_{\rm CKM}^\dagger V_{\rm CKM} = V_{\rm CKM} V_{\rm CKM}^\dagger = \mathbf{1}$. This implies 12 relations constraining the sum of three terms obtained multiplying mixed matrix elements to equal zero. If represented in the complex plane these relations are triangles whose sides are the three contributions concatenated. Among the 12 unitarity triangles, one is particularly relevant because none of the angles is so small to become difficult to measure.

$$(db) \text{ triangle} \qquad V_{ub}^* V_{ud} + V_{cb}^* V_{cd} + V_{tb}^* V_{td} = 0 \tag{I.19}$$

The angles of this triangle have been named

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \qquad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \text{and} \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \tag{I.20}$$

It is because of the non-null value of the complex phase that the area of the unitarity triangles (which is the same for all the triangles) is not zero. For this reason, the area of the unitarity triangles is sometime used as a "measure" of the CP violation in the quark sector.

Three categories of observables related to CP violation are defined: *CP violation in the decay* or *direct CP violation*; *CP violation in the oscillation* or *CP violation in the mixing*; and *CP violation in the interference* between decay and mixing.

CP violation in the decay is observed when the decay rate of a meson to a final state f is different from the decay rate of its CP-conjugate meson towards the CP-conjugate final state $\bar{f}$. Direct CP violation is the best example of difference between the behavior of matter and antimatter. It is the only possible CP violating mechanism in charged mesons, which can not oscillate (see below). Limiting the interest to charged mesons, such a difference between CP-conjugate mesons can be studied as a decay width charge

asymmetry, defining the CP asymmetry:

$$\mathcal{A}_{f^\pm} = \frac{\Gamma(B^- \to f^-) - \Gamma(B^+ \to f^+)}{\Gamma(B^- \to f^-) + \Gamma(B^+ \to f^+)} \tag{I.21}$$

For neutral mesons, where flavour oscillation is possible, CP violation in the mixing and in the interference are possible. Considering a neutral B meson system, the following CP eigenstates are defined

$$|B^0_{CP+}\rangle = \frac{1}{\sqrt{2}}\left(|B^0\rangle + |\bar{B}^0\rangle\right) \quad ; \quad |B^0_{CP-}\rangle = \frac{1}{\sqrt{2}}\left(|B^0\rangle - |\bar{B}^0\rangle\right) \tag{I.22}$$

where $|\bar{B}^0\rangle \equiv CP|B^0\rangle$. Mass eigenstates are given by

$$|B^0_L\rangle = \frac{1}{\sqrt{p^2+q^2}}\left(p|B^0\rangle + q|\bar{B}^0\rangle\right) \quad ; \quad |B^0_H\rangle = \frac{1}{\sqrt{p^2+q^2}}\left(p|B^0\rangle - q|\bar{B}^0\rangle\right) \tag{I.23}$$

where p and q are parameters fixed by Nature and are related to the mass difference between heavier and lighter mass eigenstates and to the difference between their decay widths. An evidence for CP violation in mixing requires

$$\left|\frac{p}{q}\right| \neq 1. \tag{I.24}$$

In such a case, the mass eigenstates $|B^0_L\rangle$ and $|B^0_H\rangle$ differ from CP eigenstates $|B^0_{CP+}\rangle$ and $|B^0_{CP-}\rangle$. Hence it is possible that time evolution modifies the state in the CP eigenstate basis, so that without any external interaction, CP is not conserved.

It is possible to measure CP violation effects in the interference between a decay $B^0 \to f$ ($\bar{B}^0 \to f$) and a decay $\bar{B}^0 \to B^0 \to f$ ($B^0 \to \bar{B}^0 \to f$). This effect only occurs if B^0 and $\bar{B}^0$ share a common decay channel. Decays to CP eigenstates are the cleanest example. CP violation in the interference occurs if

$$Im\left(\frac{q}{p}\frac{\overline{A}_f}{A_f}\right) \neq 0 \tag{I.25}$$

The physical observable is the time-dependent CP asymmetry, defined as

$$\mathcal{A}_{CP}(t) = \frac{\Gamma(B^0(t) \to f) - \Gamma(\bar{B}^0(t) \to f)}{\Gamma(B^0(t) \to f) + \Gamma(\bar{B}^0(t) \to f)} \tag{I.26}$$

where $B^0(t)$ and $\bar{B}^0(t)$ represent neutral B states decaying to f at a time t after they have been tagged as pure B^0 and $\bar{B}^0$ states, respectively. Studying this time dependence it is possible to extract the interference term between the two decays, which only depends on the CP-violating phase introduced in the mixing matrix. This method is the most precise to measure the α and β angles of the unitarity triangle.

$^{2S+1}L_J$	n_r	$c\bar{c}$ state	$b\bar{b}$ state
$^1S_0$	1	η_c	η_b
$^3S_1$	1	J/ψ	$\Upsilon(1S)$
$^3P_0$	1	χ_{c0}	χ_{b0}
$^3P_1$	1	χ_{c1}	χ_{b1}
$^3P_2$	1	χ_{c2}	χ_{b2}
$^1P_1$	1	h_c	
$^3S_1$	2	$\psi(2S)$	$\Upsilon(2S)$
$^3S_1$	2	$\psi(3770)$	$\Upsilon(3S)$
$^3S_1$	4	$\psi(4040)$	$\Upsilon(4S)$

Table I.1: Some of the predicted and observed $c\bar{c}$ and $b\bar{b}$ states classified according the radial-excitation quantum number n_r, and the momentum quantum numbers: S for the spin, L for the orbital momentum, and J for the total angular momentum.

I.4 Quarkonium states

A special family of mesons includes the *quarkonium* states (or simply *quarkonia*), composed of a quark and its anti-quark. Quarkonium is named *charmonium* when formed of a $c\bar{c}$ pair, and bottomonium when formed of a $b\bar{b}$ pair.

Historically, the observation of quarkonium states has been interpreted as the discovery of their constituent quarks. In 1974 the first charmonium state was observed independently at SLAC and at Brookhaven (BNL). The former group proposed the name ψ, while the second J [9, 10]. Nowadays it is still known as J/ψ. The observation was immediately acknowledeged as evidence of the existence of a fourth quark, the *charm* quark, previously predicted by Glashow, Iliopoulos and Maiani (1970) [11].

In 1977 the first bottomonium, named Υ state was observed at Fermilab, by a group led by L. Lederman [12], and interpreted as evidence of the existence of the *bottom* quark predicted by Kobayashi and Maskawa [13].

The study of the many quantum-mechanical states of the quarkonium system is named *spectroscopy*. The J/ψ and Υ resonances are both vector mesons with spin 1. There are lighter states in which the spins of the quark and the anti-quark couple in a singlet state to give spin-0 quarkonium states named η_c and η_b. Even if they are lighter than their vector partners, they are more difficult to produce. Indeed, the production channels having a virtual photon coupling to a $c\bar{c}$ or $b\bar{b}$ pair is one of the dominant for the vector quarkonia, but it is forbidden for the scalar states because the angular momentum of the initial state (the virtual photon) and of the final state (the scalar quarkonium) are different.

Besides spin coupling, also orbital momentum and radial excitations of the quark–anti-quark system introduces higher-mass states in the quarkonium spectroscopy.

In Table I.1 some of the lightest, well established, quarkonium states are classified. A more complete and detailed listing can be found for example in Ref. [1].

I.4.1 Potential models

Being the *charm* and *bottom* quark heavy, their motion within the hadrons is considerably slower than it is for light quarks. In particular, it can be shown that the relativistic corrections to a non-relativistic description of their wavefunctions are small, $\mathcal{O}\left(\frac{m_q}{\Lambda_{\mathrm{QCD}}}\right)$.

In the center-of-mass frame, the Schrödinger equation for the state ψ is

$$-\frac{1}{2\mu}\nabla^2\psi(\boldsymbol{r}) + V(r)\psi(\boldsymbol{r}) = E\psi(\boldsymbol{r}), \tag{I.27}$$

where $\boldsymbol{r}$ is the space vector and r its modulus, $V(r)$ the radial potential, E the energy of the state, and $\mu = m_q/2$ the reduced mass of the two quarks of mass m_q. In the approximation of negligible spin-dependent effects, the mass of the quarkonium states is simply

$$M(q\bar{q}) = 2m_q + \frac{E}{c^2}. \tag{I.28}$$

It is possible to reproduce the observed spectrum through a simple parametrization of $V(r)$ as a weighed sum of Coulomb and harmonic potentials, with weights obtained fitting the measured masses of the quarkonium states.

$$V(r) = -\frac{a}{r} + br, \tag{I.29}$$

with $a = 0.30$, and $b = 0.23\,\mathrm{GeV}^2$.

An alternative potential, which fits experimental data equivalently well, is

$$V'(r) = a\ln(br) \tag{I.30}$$

with $a = 0.75\,\mathrm{GeV}$, and $b = 0.80\mathrm{GeV}$.

The shapes of the two potentials V and V' are in good agreement in the range $0.2\,\mathrm{fm} < r < 1.0\,\mathrm{fm}$, which allows to conclude that the potential is well defined in this region, describing properly the dynamics of the quarkonium states, which is found consistent for charmonium and bottomonium states.

Potential models for quarkonium states constitute an active research field with many different models developed and tuned on experimental data, in a way similar to what is described for the simplistic model above. For example, the use of relativistic quark models, instead of the non relativistic description shown above, allows a coherent description of heavy and light mesons in excellent agreement with experimental data [14].

Many reviews of the potential models for quarkonia state exist, see for example the CERN Yellow Report in Ref. [15] or PDG review in Ref. [16].

I.4.2 Quarkonium Decays

The lifetime of a particle is inversely proportional to its decay width Γ, a weighted sum of the squared-modulus amplitudes of all the possible decay channels of that particle.

Particles whose lifetime is too short to be measured as a flight distance are called *resonances* and the decay width is usually quoted instead of the lifetime. All the quarkonium states are resonances, but their decay widths can be orders-of-magnitude different from state to state. In particular when the mass of the quarkonium state is larger than the kinematical threshold of two heavy mesons, new decay channels become accessible and the decay width increases strongly. For charmonium states the kinematical threshold is $2m(D^0) \approx 3\,739$ MeV/c^2. States heavier than the threshold, like the $\psi(3770)$, can decay to $D^0\bar{D}^0$ or D^+D^- through strong interaction, while below the threshold, many common decay channels require the annihilation of the two heavy quarks, suppressed by the so called Zweig rule [17].

Another notable example comes from the bottomonium system, where the first resonance beyond the threshold is the $\Upsilon(4S)$ decaying to $B^0\bar{B}^0$ or B^+B^-. At the b-factories, the e^+e^- collisions are tuned to happen with a center-of-mass energy equal to the mass of the $\Upsilon(4S)$, 10.58 GeV/c^2, to study the B^0_d and $B^\pm_u$ mesons in clean events where only the $B\bar{B}$ pair is present in the detector. To perform studies on the B^0_s meson, the center-of-mass energy was increased to the mass of the $\Upsilon(5S)$, whose production is much less efficient. For this reason it is impossible for the b-factories to produce beauty baryons, like the Λ^0_b, and heavier mesons like the B^+_c.

Higher quarkonium states can also decay to lighter states with a photon or a pion pair, for example,

$$\begin{aligned} \psi(2S) &\to \pi^+\pi^- J/\psi \qquad Br : 34.45 \pm 0.30\%, \\ \psi(2S) &\to \pi^0\pi^0 J/\psi \qquad Br : 18.13 \pm 0.31\%, \\ \chi_{c0} &\to \gamma J/\psi \qquad Br : 1.27 \pm 0.06\%. \end{aligned}$$

I.5 The B^+_c meson

Between charmonium ($c\bar{c}$) and bottomonium ($b\bar{b}$), the quark model predicts the existence of a mixed state ($c\bar{b}$) named B^+_c meson. Despite this apparent similarity, the fact that the two quarks have different flavours changes significantly the properties of this meson. The B^+_c meson is therefore a unique state in the standard model, a sort of *open-flavour quarkonium* state, where models used to describe quarkonia can be applied to a long-lived state that can only decay through weak interaction.

One of the features shared by the B^+_c meson and by quarkonium states is the theoretical framework used to describe the excited states, unfortunately only the ground state of the ($c\bar{b}$) system, named B^+_c, is well established, though for its radial excitation B^{+**}_c there is an unconfirmed observation [18].

The reason making harder to study B^+_c states than quarkonia is the different production mechanism. To produce a quarkonium state it is sufficient to produce a high-energy virtual photon or gluon producing a $q\bar{q}$ pair which hadronizes to a charmonium state. This can be done easily at lepton colliders, since the annihilation of e^+e^- is likely to produce virtual photons, or at hadronic colliders through the exchange of one or three gluons.

The production of the B^+_c mesons is instead more difficult to achieve because both the c and the $\bar{b}$ quarks and the respective charge-conjugate quark $\bar{c}$ and b have to produced in the

event. The production mechanism of B_c^+ is therefore much different from the production mechanism of quarkonium, and it is also different with respect to the production mechanism of B^+, B^0, and B_s^0 states. According to Ref. [19], the probability of a fragmentation of a b quark towards a B_c^+ state is about 11 order of magnitude lower than a fragmentation towards B^0 or B^+ states.

The dominant production mechanism becomes the gluon-gluon fusion process, requiring fourth-order calculations in α_s, to be described through perturbative QCD models. According to these models, at the energy of the Large Hadron Collider (LHC), the production cross-section of B_c^+ mesons should be only three order of magnitude smaller than for $B_{u,d,s}$ mesons, making its study possible [19].

The description of the production mechanisms of the B_c^+ meson would probably deserve another whole Thesis, hence it is covered here only very partially. Nonetheless, it is relevant to introduce the momentum spectrum of the B_c^+ mesons produced at LHC since its effect on the lifetime measurements is not negligible.

Another aspect that makes the study of B_c^+ meson challenging is the great variety of final states accessible through its decays. The B_c^+ ground state has a mass of approximately 6.28 GeV/c^2, well below the kinematical threshold to strong decays towards B and D mesons: $m(D^0) + m(B^+) \approx 7.144\,\text{GeV}/c^2$. Hence, to all the possible weak decays led by $\bar{b} \to \bar{c}(\bar{u})$ transitions, one should add the decays through $c \to s(d)$ transition and through weak annihilation $\bar{b}c \to W^{*+}$. Each category accounts for dozens or hundreds of possible final states differing by the hadronization of the final state quarks. The theoretical description of the B_c^+ decay is therefore essential to orient the studies towards channels with relatively high branching fraction, with basic understanding of the expected decay kinematics.

I.5.1 Theoretical prediction of the B_c^+ mass

Nowadays, the precision on the B_c^+ mass is very high as discussed in the next chapter, therefore there is no need to use theoretical assumptions on the B_c^+ mass as input of the lifetime measurements. However, it is important to briefly review the theoretical expectations on the mass in order to build confidence in the theoretical understanding of this state.

The most precise theoretical calculation has been obtained by the HPQCD Collaboration using lattice QCD numerical algorithms using the *Highly Improved Staggered Quark* action for u/d, s and c quarks and non-relativistic QCD (NRQCD) for the b quark. The result [20],

$$m_{B_c^+}^{(\text{HPQCD})} = 6280 \pm 10 \ \text{MeV}/c^2,$$

is consistent with the world average [1]

$$m_{B_c^+}^{(\text{PDG})} = 6275.6 \pm 1.1\,\text{MeV}/c^2.$$

I.5.2 Theoretical considerations on the B_c^+ meson production

The production mechanism can be described in three steps:

1. The pp collision creates a $c\bar{c}$ and a $b\bar{b}$ pair;
2. If the produced heavy quarks have small relative momentum, it is possible that they bind to form a $(\bar{b}c)$ state;
3. If the bound state has a mass below the BD kinematical threshold, then it decays with strong or radiative transition to the B_c^+ meson ground-state.

Hence, when calculating the B_c^+ total production cross-section, direct and feed-down productions must be summed.

The production of the c- and $\bar{b}$-quark in a parton collision can be led by several mechanisms. The fragmentation mechanism is very important for the production of single heavy mesons (having one heavy quark and one light quark) such as D^0, $D_{(s)}^+$, B^+, $B_{(s)}^0$. In this mechanism, the heavy quark pair is produced by parton shower; once the heavy quark is produced, light quark–anti-quark pairs are created from vacuum to shield the color charge, and the light anti-quark (quark) binds to the heavy quark (anti-quark) to give a meson. However, the probability to create a pair of heavy quarks from vacuum is much smaller than that of creating a pair of light ones, so that the probability of producing a quark pair of various flavours from the vacuum is expressed relative to the probability of creating an $u\bar{u}$ pair as [19]

$$u : d : s : c = 1.0 : 1.0 : 0.3 - 0.4 : 10^{-10} - 10^{-11}. \tag{I.31}$$

An alternative mechanism is to produce the B_c^+ meson via a excited-beyond-threshold state $\Upsilon(nS)$, where $n \geq 4$ is the radial excitation quantum number. But while this mechanism is reasonable for B^0 and B^+ mesons, no bottomonium state is either observed or expected beyond the $B_c^+ B_c^-$ threshold of $2m(B_c^+) \approx 12.6\,\mathrm{GeV}/c^2$.

A third mechanism one may think of is the direct weak production, with a virtual weak boson W^+ coupling to a pair $c\bar{b}$ which binds to produce a B_c^+ meson. Beside the CKM suppression due to the small V_{cb} matrix element, the mass of the B_c^+ is much smaller than the mass of the W, which should be much off-shell suppressing the propagator.

The model currently accepted for the B_c^+ production requires two heavy quark pairs $c\bar{c}$ and $b\bar{b}$ to be produced in the same parton–parton interaction. There is, then, a sizeable probability that they match to form a B_c^+ meson.

Two categories of parton–parton scattering producing a B_c^+ meson are identified: the gluon-gluon fusion, and the quark–anti-quark fusion. As discussed below, at the LHC energy the latter is suppressed. Figure I.3 shows some of the Feynman diagrams for the two categories.

As it can be noticed, the diagrams of each process count at least four vertices. Translating the diagrams to formulae, this means that the B_c^+ production cross-section is proportional to $\alpha_s(Q^2)^4$. The most ambiguous part of the calculation is the choice of

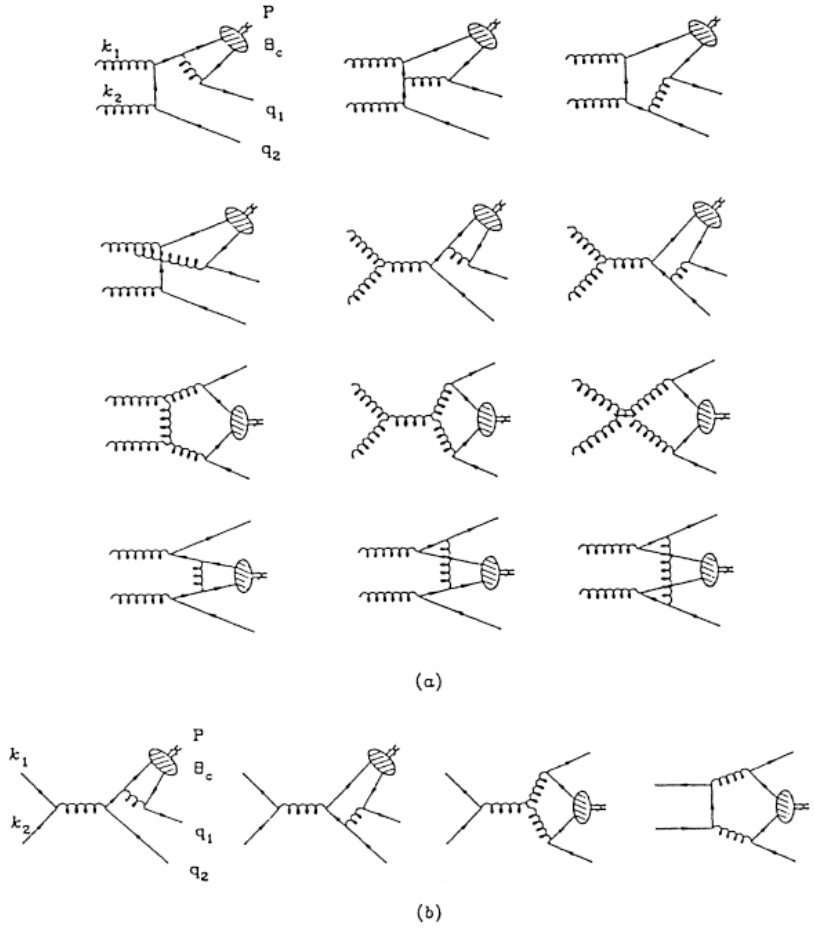

Figure I.3: Feynman diagrams for the parton–parton processes. (a) The typical Feynman diagrams for the gluon–gluon fusion. (b) The Feynman diagrams for the quark–anti-quark fusion. The Figure was published in Ref. [19].

the energy scale Q^2. Several reasonable choices have been proposed in the history of B_c^+ theoretical predictions, spanning from the *charm*-quark mass m_c to the sum of the masses of all the produced heavy quarks $2(m_b + m_c)$. Unfortunately the difference in the predicted cross-section can easily change by an order of magnitude from one extreme to the other, so that in some cases it has been chosen to set α_s to some value, typically 0.2, without discussing the energy scale [21–25]. Modern Monte Carlo generators assume $Q^2 = \hat{s}/4$, where $\hat{s}$ is the squared center-of-mass energy of the parton–parton interaction process [19].

A further complication arises from the existence of excited $(\bar{b}c)$ states which decay through strong or electromagnetic interaction towards a B_c^+ meson. Hence, when studying the production of the latter also the production of all the higher states must be known. All the parton–parton processes that can produce a B_c^+ are therefore reanalyzed requiring each of the possible excited states in the final state. Their contributions are then included in the final result for the production cross-section $\sigma(pp \to B_c^+ + b\bar{c})$.

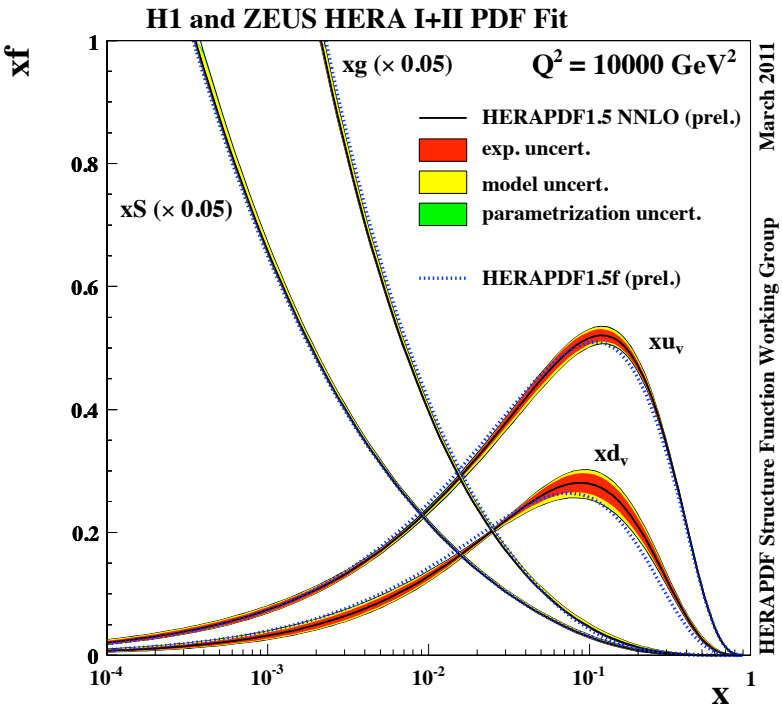

Figure I.4: The parton distribution functions of the proton from *HERAPDF1.5 NNLO* at $Q^2 = 10000$ GeV^2, *i.e.* the region relevant for the hadron colliders as the LHC. The gluon and sea distributions are scaled down by a factor 20. The experimental, model and parametrization uncertainties are shown separately. For comparison, the central value of the *HERAPDF1.0 NNLO* is also shown. The Figure was published in Ref. [27].

Once the parton–parton scattering processes have been defined, they have to be weighted by the probability of finding a given parton i in the proton. Such probability is named proton *parton density function* $F_i(x, Q^2)$, where x is the fraction of the proton momentum carried by the parton i, and the dependence on the energy-scale Q^2 of the hard process is small and known as Bjorken scaling violation. The parton density functions for the gluons, the sea quarks (and anti-quarks), and the valence quarks, are shown in Figure I.4, which summarizes well two important aspects. First, fixed the minimal energy in center-of-mass of the parton–parton system needed to obtain a certain process, the probability of finding a suitable gluon or sea quark increases quickly by raising the proton energy, and therefore decreasing x; this explains the strong dependence on the energy of the accelerator beams of the cross-section for pp collisions towards final states including heavy-quark pairs. Second, at low-x the probability of finding a gluon is much larger than the probability of finding a sea quark, and much larger than the probability of finding a valence quark (or anti-quark) which makes the difference between pp and $p\bar{p}$ collisions negligible in terms of low-x cross-sections. For a more pedagogical introduction to proton form factors see for example Ref. [26].

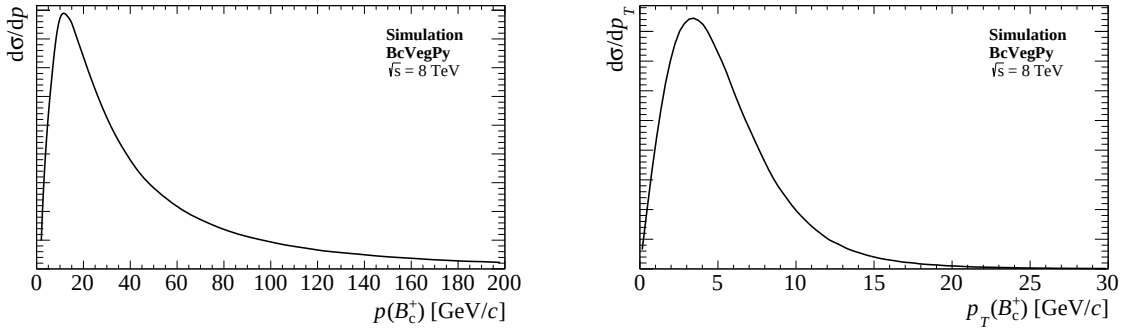

Figure I.5: Dependence of the B_c^+ production cross-section $\sigma(pp \to \bar{b}c + \bar{c} + b)$ on the momentum (left) and transverse momentum (right) of the B_c^+ meson as obtained with a configuration for proton–proton collisions at a center-of-mass energy of 8 TeV by the BCVEGPY generator described in Ref. [28].

Recent calculations have been released as a Monte Carlo (or VEGAS) generator named BCVEGPY [28] which extends PYTHIA 6 [29] by including the parton–parton scattering processes discussed above.

The study of the production of the B_c^+ meson is an active field of research both theoretically and experimentally, but for the analysis described here the relevant results are the p and p_T spectra shown in Figure I.5. As discussed later, the models describing production quantities are subject to some degree of arbitrariness that have to be taken properly into account when defining a robust analysis strategy.

I.5.3 Theoretical description of the B_c^+ meson decays

As mentioned already, the B_c^+ decay processes can be classified in three branches:

- the $\bar{b}$ quark decay with the spectator c-quark,
- the c quark decay with the spectator $\bar{b}$-quark, and
- the annihilation channel $B_c^+ \to W^{*+}$.

In the case of $\bar{b} \to \bar{c}c\bar{s}$ decays, one separates the contribution from the Pauli interferenace (PI) of the c-quark in the final state of the b-quark decay with the c-quark from the initial state. Hence, one can write the total decay width as sum over the partial widths of the categories listed above.

$$\Gamma(B_c^+ \to X) = \Gamma(b \to X) + \Gamma(c \to X) + \Gamma(\text{annihil.}) + \Gamma(PI). \tag{I.32}$$

A reliable estimation for the annihilation decay width can be obtained in the framework of inclusive decays, taking the sum of the quark and leptonic decay modes into account. It can be shown that [30],

$$\Gamma(\text{annihil.}) \propto \sum_{i=c,\tau} m_i^2 \left(1 - \frac{m_i^2}{m_{B_c}^2}\right)^2 \cdot C_i, \tag{I.33}$$

where $C_\tau = 1$ and $C_c = 3|V_{cs}|^2 a_1^2$, and all the other contributions are negligible because suppressed by the inverse of the particle squared mass. In the definition of C_c, V_{cs} is an element of V_{CKM} as defined in Equation I.14, and $a_1 = 1.22 \pm 0.04$ is a correction factor accounting for hard gluon corrections to the effective four-quark interaction of weak corrections.

The other contributions can be studied with the approach of the Operator Product Expansion (OPE), introduced later in Section I.5.5 taking into account the α_s corrections to the free quark decays. Considering the matrix element for the transition operator over the bound meson state allows to take into account the interference between the spectator c-quark and the c quark from $\bar{b} \to c\bar{c}s$ decays, showing that the latter is almost completely suppressed due to Pauli interference. The c-quark decays with spectator $\bar{b}$-quark are expected to dominate with respect to the b-quark decays because $\bar{b} \to \bar{c}$ transitions are CKM-suppressed cross-family decays, but on the other hand they are suppressed in comparison with the free quark decays because of a large bound energy in the initial state, in a similar way to how neutrons have finite lifetime when free, but become stable when bound in stable nuclei.

The conclusion is that, of the whole decay width, about 70% is due to c-quark decays with spectator $\bar{b}$, about 20% is due to $\bar{b}$-quark decays with the c-quark being spectator, and only 10% to weak annihilation.

Since the experimental signature of decays led by $\bar{b} \to \bar{c}$ transitions is clearer because a charmonium state can often be found in the final state, most of the observed and studied decays belong to this category. In particular in this Thesis, the focus is on semileptonic decays $B_c^+ \to J/\psi\, \mu^+ \nu_\mu$.

Form factors of $B_c^+ \to J/\psi\,\mu^+\nu_\mu$

The theoretical description of a decay consists of two main parts: the determination of the decay amplitude, and the determination of the phase-space of the decay. The latter, sometimes referred to as the *kinematics of the decay*, depends only on the number and the masses of the particles involved, and can be easily determined analytically as a function of the momenta of the particles in the final state in the rest frame of the mother particle. Instead, the decay amplitude or *decay dynamics* is evaluated studying the matrix element of the Hamiltonian operator describing the decay transition. In the case of two-body decays, momentum conservation forces the momenta of the two daughter particles to be equal and opposed in the rest frame of the mother particle, and only in case of polarized non-scalar mother particles the decay amplitude has a non-trivial angular dependence. In three-body decays, the relative angular distributions of the daughter particles depend on the decay dynamics which is often unknown. The definition of the angular dependence of the matrix element requires assumptions on several parameters describing the QCD correction to the weak decays, which may be treated with various approaches.

The results obtained by several theorists studying the form factors of the $B_c^+ \to J/\psi\, X$ decays are qualitatively similar but the theoretical uncertainties affect the precision measurement presented in this Thesis.

Different theoretical frameworks are used. Historically, the first calculations were done with potential models, then more complete models, as QCD sum rules [31] or relativistic quark model [32] have been developed showing better agreement with the experimental data on the B mesons, and therefore more trusted in calculations on the B_c^+ meson.

In general, the three-body decay matrix elements are computed through a three-point correlation function. Two-point correlation functions are the common propagators or Green functions, used in perturbative quantum field-theory to describe the virtual particles appearing in Feynman diagrams. Three-point correlation functions define the momenta of the daughter particles given the momentum of the mother particle.

When considering the decay of a pseudoscalar meson (as the B_c^+) to a vector (as the J/ψ), the decay amplitude can be described independently on the chosen dynamic model in terms of the four *form factors* $V(q^2), A_0(q^2), A_1(q^2)$, and $A_2(q^2)$.

The vector and axial currents are defined as [1]

$$\langle J/\psi\,(p',\epsilon)|\bar{c}\gamma^\mu b|B_c^+(p)\rangle = 2i\varepsilon^{\mu\nu\alpha\beta}\frac{\varepsilon_\nu p'_\alpha p_\beta}{m(B_c^+)+m(J/\psi)}V(q^2), \tag{I.34}$$

$$\begin{aligned}\langle J/\psi\,(p',\epsilon)|\bar{c}\gamma^\mu\gamma_5 b|B_c^+(p)\rangle = {} & \big(m(B_c^+)+m(J/\psi)\big)\left[\varepsilon^\mu - \frac{\varepsilon\cdot q q^\mu}{q^2}\right]A_1(q^2) \\ & - \varepsilon\cdot q\left[\frac{(p+p')^\mu}{m(B_c^+)+m(J/\psi)} - \frac{(m(B_c^+)-m(J/\psi))q^\mu}{q^2}\right]A_2(q^2) \\ & + 2m(J/\psi)\frac{\varepsilon\cdot q q^\mu}{q^2}A_0(q^2),\end{aligned} \tag{I.35}$$

[1] The natural system of units, with $c = \hbar = 1$ is used here and in the following sections.

where p is the momentum of the B_c^+ meson, p' the momentum of the J/ψ, and q their difference $q = p - p'$. The polarization of the vector meson (*i.e.* the J/ψ) is indicated with ε. Note that in the limit of massless lepton, $A_0(q^2)$ does not contribute to semileptonic decay rates since $\varepsilon \cdot q = 0$.

By construction

$$m_\ell^2 \geq q^2 \geq (m(B_c^+) - m(J/\psi))^2. \tag{I.36}$$

The vector and axial currents defined in Eq. I.34 and I.35 enter the definition of the differential decay amplitude to give

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}q^2} = \frac{G_F^2 \Delta |V_{qb}|^2}{96\pi^3} \frac{q^2}{m(B_c^+)^2} \left(|H_+(q^2)|^2 + |H_-(q^2)|^2 + |H_0(q^2)|^2 \right), \tag{I.37}$$

where $H_\pm$ and H_0 represent the different helicity amplitudes and are defined as

$$H_\pm(q^2) = \frac{2m(B_c^+)\Delta}{m(B_c^+) + m(J/\psi)} \left[V(q^2) \mp \frac{(m(B_c^+) + m(J/\psi))^2}{2m(B_c^+)\Delta} A_1(q^2) \right], \tag{I.38}$$

$$H_0(q^2) = \frac{1}{2m(J/\psi)\sqrt{q^2}} \Bigg[(m(B_c^+) + m(J/\psi))(m(B_c^+)^2 - m(J/\psi) - q^2) A_1(q^2) - \frac{4m(B_c^+)^2\Delta^2}{m(B_c^+) + m(J/\psi)} A_2(q^2) \Bigg]. \tag{I.39}$$

The quantity Δ represents the recoil momentum of the J/ψ resonance in the B_c^+ rest frame and can be defined as

$$\Delta = \sqrt{\frac{(m(B_c^+)^2 + m(J/\psi) - q^2)^2}{4m(B_c^+) - m(J/\psi)^2}}. \tag{I.40}$$

The difference between different models enters in the definition of form factors $V(q^2)$, $A_0(q^2)$, $A_1(q^2)$ and $A_2(q^2)$.

As an example, the form factors obtained in the relativistic quark framework are explicitly reported in Appendix of Ref. [32].

The dependence of the partial decay width $\Gamma(B_c^+ \to J/\psi\, \mu\nu_\mu)$ on q^2 and on the invariant mass of the combination $J/\psi\, \mu^+$ is shown in Figure I.6 for the phase-space model, obtained neglecting the decay dynamics, and for two form factor models obtained with QCD sum-rules by V. V. Kiselev [33] and in framework of the relativistic quark model by D. Ebert [32].

Another form-factor model which is used in this Thesis is the ISGW2 model [34], an update of the ISGW [35] model named after the authors Isgur, Scora, Grinstein, and Wise based on the non-relativistic quark model and successfully used for many years to describe semileptonic B decays.

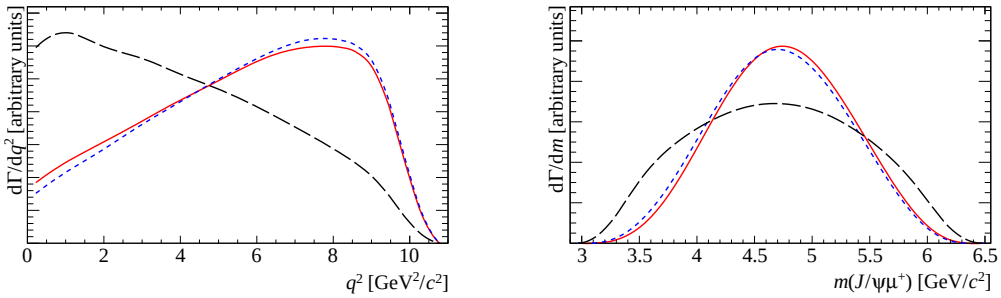

Figure I.6: Dependence of the partial decay width $\Gamma(B_c^+ \to J/\psi\,\mu\nu_\mu)$ with respect to q^2 (left) and to the invariant mass of the $J/\psi\,\mu$ combination (right). The considered models are the pure phase space model (black long-dashed curve), the form factor model obtained using QCD sum-rules [33] (red solid curve), and the form factor model obtained in the framework of relativistic quark model [32] (blue short-dashed curve).

I.5.4 Prospects for studies of CP violation in the $(\bar{b}c)$ system

Since the B_c^+ meson is a charged state it cannot oscillate, hence neither CP-violation in the oscillation, nor in the interference between the oscillation and the decay can arise. The only possible CP violation contribution comes from CP violation in the decay, or direct CP violation.

As a consequence of CPT invariance, the CP operator can only modify the phase of the amplitude described by a single Feynman diagram. Two interfering amplitudes with a common final state f are thus needed to cause CP violating differences of the decay widths $\Gamma(B \to f)$ and $\Gamma(\bar{B} \to \bar{f})$. Measurements with suitable final states provide information on the parameters of the CKM matrix. In Ref. [36], I have summarized the experimental techniques used to measure the γ angle from charmed decays of the charged B mesons.

The same approach can be used in the B_c^+ system, where the statistics is known to be smaller, but the effect of CP-violation is expected to be larger with suitable statistics. A measurement of γ in the $(\bar{b}c)$ system would be probably the easiest-to-achieve independent measurement of this quantity in a loop-pollution-free decay. And would therefore constitute an important test of the CKM picture describing the weak decay of quarks.

The following presentation of the measurement of CP violating observables with the B_c^+ meson summarizes the paper by Kiselev in Ref. [37].

Consider the decays

$$B_c^+ \to D^0 D_s^+, \qquad \text{and} \qquad B_c^+ \to \bar{D}^0 D_s^+,$$

whose diagrams are shown in Figure I.7. The exclusive modes make the penguin terms to be excluded, since the penguins add an even number of charm quarks, *i.e.* two or zero, while the final state contains two charm quarks including one from the initial state, so that only one charm quark has to be added in the $\bar{b}$ quark decay.

Distinguishing between *weak phase* and *strong phase*, where the former changes sign under application of a CP transformation while the latter doesn't, one can write the

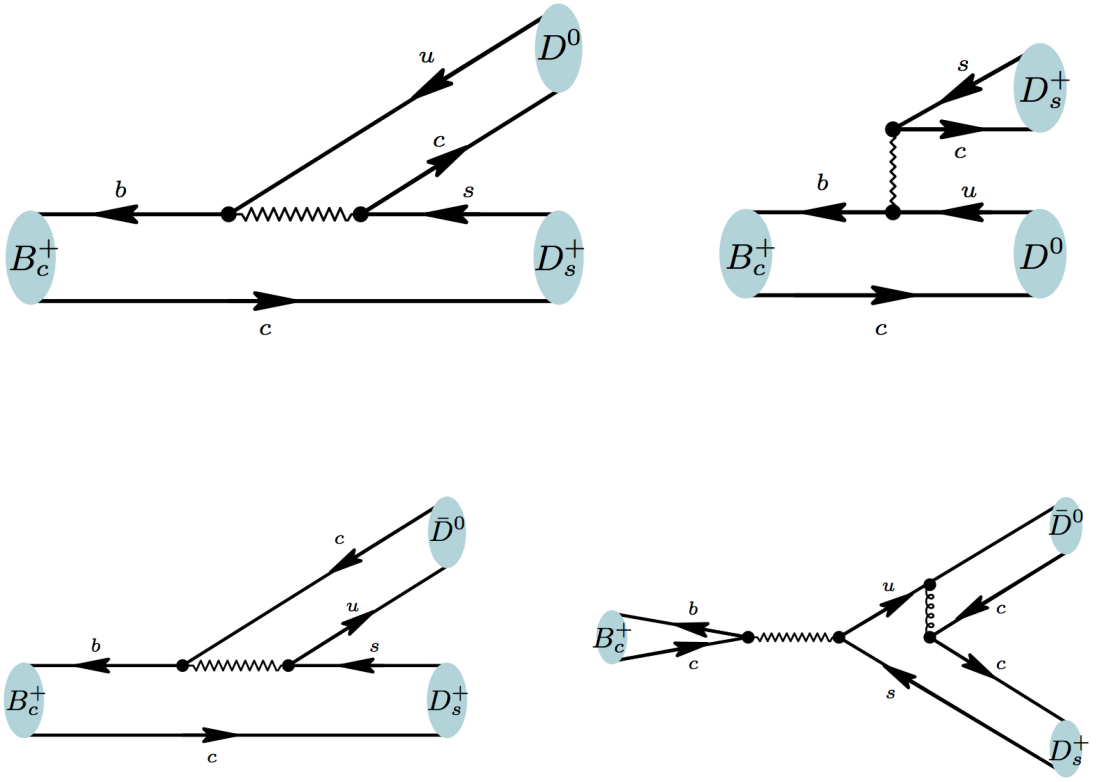

Figure I.7: Feynman diagrams of the decay $B_c^+ \to D^0 D_s^+$ (top) and $B_c^+ \to \bar{D}^0 D_s^+$ (bottom). The two diagrams interfere when D^0 and $\bar{D}^0$ are reconstructed in CP eigenstates as $D^0 \to K^+K^-$.

amplitudes of the decays depicted in Figure I.7 as

$$\mathcal{A}(B_c^+ \to D^0 D_s^+) \equiv \mathcal{A}_D = V_{ub}^* V_{cs} \mathcal{M}_D, \quad \text{and} \quad \mathcal{A}(B_c^+ \to \bar{D}^0 D_s^+) \equiv \mathcal{A}_{\bar{D}} = V_{cb}^* V_{us} \mathcal{M}_{\bar{D}}, \tag{I.41}$$

where $\mathcal{M}_D$ and $\mathcal{M}_{\bar{D}}$ denote the strong-phase factors depending on the dynamics of the strong interaction, and the effects of the weak interaction, including the weak phase, have been factorized as product of CKM-matrix elements.

Recalling the definition of the angle γ of the unitarity triangle expressed in Eq. I.20,

$$\gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right), \tag{I.42}$$

one can write

$$\begin{aligned} \arg\left[\frac{V_{ub}V_{cs}^*}{V_{cb}V_{us}^*}\right] &= \arg\left[\left(-\frac{V_{ub}V_{cs}^*}{V_{cb}V_{us}^*}\right)\cdot\left(-\frac{V_{ud}^*V_{cd}^*}{V_{ud}^*V_{cd}^*}\right)\right] = \arg\left[-\frac{V_{ud}^*V_{ub}}{V_{cd}^*V_{cb}}\right] + \arg\left[-\frac{V_{ud}V_{cd}^*}{V_{cs}V_{us}^*}\right] \\ &= -\gamma + \varphi. \end{aligned} \tag{I.43}$$

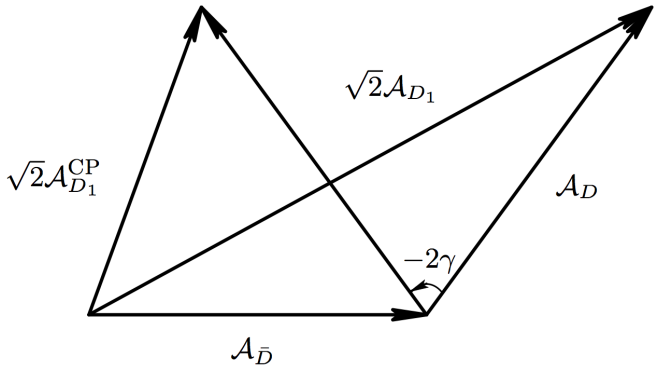

Figure I.8: Triangular relation representing Equations I.47 and I.48 to explain how a weak phase becomes experimentally accessible. The figure was published in Ref. [37].

The phase φ is an angle of the unitarity triangle associated to the relation

$$V_{ud}V_{cd}^* + V_{us}V_{cs}^* + V_{ub}V_{cb}^* = 0. \tag{I.44}$$

Using the Wolfenstein parametrization of the CKM matrix, it is easy to state that two of the sides of such a triangle have order λ, while the $V_{ub}V_{cb}^*$ has order λ^5. Being φ the angle between the long sides, one concludes that it so small to be negligible compared to γ.

Under this excellent approximation, one can therefore write

$$\mathcal{A}(B_c^- \to \bar{D}^0 D_s^-) = e^{-2i\gamma}A_D, \qquad \text{and} \qquad \mathcal{A}(B_c^- \to \bar{D}^0 D_s^-) = A_{\bar{D}}, \tag{I.45}$$

where the choice $V_{cb}V_{us}^* = 0$ sets the global phase convention consistently, for example, with the Wolfenstein parametrization.

Defining a CP-even eigenstate of the D^0 meson

$$D_1 = \frac{1}{\sqrt{2}}\left(D^0 \pm \bar{D}^0\right), \tag{I.46}$$

one gets the relation

$$\sqrt{2}\mathcal{A}(B_c^+ \to D_s^+ D_1) \equiv \sqrt{2}\mathcal{A}_{D_1} = \mathcal{A}_D + \mathcal{A}_{\bar{D}}, \tag{I.47}$$

$$\sqrt{2}\mathcal{A}(B_c^- \to D_s^- D_1) \equiv \sqrt{2}\mathcal{A}^{\rm CP}_{D_1} = e^{-2i\gamma}\mathcal{A}_D + \mathcal{A}_{\bar{D}}. \tag{I.48}$$

$$\tag{I.49}$$

Being the amplitudes complex quantities, Equations I.47 and I.48 translate to the triangular relations shown in Figure I.8.

The length of the sides of the triangle shown in Figure I.8 are experimentally accessible in the form of decay width, or relative branching fractions. The CP asymmetry

$$A_{CP} = \frac{\Gamma(B_c^- \to D_1 D_s^-) - \Gamma(B_c^+ \to D_1 D_s^+)}{\Gamma(B_c^- \to D_1 D_s^-) + \Gamma(B_c^+ \to D_1 D_s^+)}, \tag{I.50}$$

Decay mode	$\mathcal{B}, 10^{-6}$	*Decay mode*	$\mathcal{B}, 10^{-6}$
$B_c^+ \to D^+ \bar{D}^0$	53	$B_c^+ \to D^+ D^0$	0.32
$B_c^+ \to D^+ \bar{D}^{*0}$	75	$B_c^+ \to D^+ D^{*0}$	0.28
$B_c^+ \to D^{*+} \bar{D}^0$	49	$B_c^+ \to D^{*+} D^0$	0.40
$B_c^+ \to D^{*+} \bar{D}^{*0}$	330	$B_c^+ \to D^{*+} D^{*0}$	1.59
$B_c^+ \to D_s^+ \bar{D}^0$	4.8	$B_c^+ \to D_s^+ D^0$	6.6
$B_c^+ \to D_s^+ \bar{D}^{*0}$	7.1	$B_c^+ \to D_s^+ D^{*0}$	6.3
$B_c^+ \to D_s^{*+} \bar{D}^0$	4.5	$B_c^+ \to D_s^{*+} D^0$	8.5
$B_c^+ \to D_s^{*+} \bar{D}^{*0}$	26	$B_c^+ \to D_s^{*+} D^{*0}$	40.4

Table I.2: Branching ratios of exclusive B_c^+ decays according to QCD sum-rules prediction as published in Ref. [37].

and the CP ratio

$$R_{CP} = \frac{\Gamma(B_c^- \to D_1 D_s^-) + \Gamma(B_c^+ \to D_1 D_s^+)}{\frac{1}{2}\left(\Gamma(B_c^- \to D^0 D_s^-) + \Gamma(B_c^+ \to \bar{D}^0 D_s^+)\right)}, \tag{I.51}$$

can thus be used to measure the angle γ of the CKM matrix in a B_c^+ extension of the Gronau-London-Wyler method widely used for the $B^\pm$ meson system.

To enhance effects of CP violation and make the measurement more sensitive to γ, it is important for the length of the sides of the triangle depicted in Figure I.8 to be as similar as possible.

Kiselev analysed different flavours of the decay $B_c^+ \to DD$ including decays to vector D mesons. The results are shown in Table I.2. The branching fractions are proportional to the lengths of the vectors $\mathcal{A}_{\bar{D}}$ and $\mathcal{A}_D$ shown in Figure I.8. The more similar are the branching ratios, the larger is the expected CP asymmetry. None of the decay modes considered have been observed. The detection of $B_c^+ \to DD$ decays is not far from reach in terms of statistics. In particular considering the partially reconstructed decays to excited D mesons. Still, the observation of CP violation in the B_c^+ system would require to increase the integrated luminosity by at least one order of magnitude with respect to the Run I sample, and is certainly not achievable before the LHCb upgrade phase, foreseen to start in 2019 [38].

I.5.5 Theoretical framework for the b-hadron lifetimes

The theoretical tools mentioned above for the study of the semileptonic decay form factors have a wider range of application since they allow to introduce QCD corrections to the calculation of the decay width of inclusive channels as $B_c^+ \to J/\psi X$. Summing up the contributions of all the possible inclusive processes, it is possible to determine the total decay width, inversely proportional to the particle lifetime.

Experimental tests of the lifetime are therefore checks of a digest of the theoretical model of the meson. This includes the form-factor determination, QCD matrix elements

that enter the higher-order corrections (perturbative QCD), and the model chosen to describe the $(\bar{b}c)$ bound state (non-perturbative QCD).

A brief overview of the theoretical frameworks used to compute lifetimes is illustrated in the following paragraphs.

Heavy Quark Effective Theory (HQET)

While not directly related to lifetime predictions, some results of the Heavy Quark Effective Theory (HQET) and notably its Lagrangian are of interest for the next applications, and therefore they deserve a short introduction.

Heavy quark effective theory is an effective theory designed to systematically exploit the simplifications of QCD interactions containing a single heavy quark. Its momentum is decomposed as

$$p = m_b v + k \tag{I.52}$$

where m_b is the mass of the heavy quark, v its *velocity* of the heavy hadron, and k the residual component which is expected to be much smaller than $m_b v$. The heavy quark field Ψ is decomposed as

$$h_v(x) \equiv e^{imv\cdot x}\frac{1+\gamma^\mu v_\mu}{2}\Psi(x), \tag{I.53}$$

$$H_v(x) \equiv e^{imv\cdot x}\frac{1-\gamma^\mu v_\mu}{2}\Psi(x), \tag{I.54}$$

which implies $\Psi(x) = e^{-imv\cdot x}(h_v(x) + H_v(x))$.

The following covariant derivative operators are introduced

$$D_\mu\partial_\mu - igT^a A^a_\mu, \qquad \text{and} \qquad D^\mu_\perp \equiv D^\mu - v^\mu v\cdot D, \tag{I.55}$$

where T^a is the a-th generator of the $SU(3)_C$ symmetry and A_μ is the gluon field.

The Dirac equation $(i\gamma^\mu D_\mu - m_b)\Psi = 0$ can be rewritten as a coupled system of equations for the projections $H_v(x)$ and $h_v(x)$:

$$iv\cdot Dh_v = -i\gamma_\mu D^\mu_\perp H_v \tag{I.56}$$

$$(iv\cdot D + 2m_b)H_v = i\gamma_\mu D^\mu_\perp h_v \tag{I.57}$$

With some algebra, it is easy to obtain the following expression of the fermion Lagrangian

$$\bar{h}_v iv\cdot Dh_v + \bar{h}_v i\gamma_\mu D^\mu_\perp \frac{1}{iv\cdot D + 2m_b} i\gamma_\mu D^\mu_\perp h_v \tag{I.58}$$

which simplifies in the following relation in the limit of $m_b \to +\infty$:

$$\mathcal{L} = \bar{h}_v iv\cdot Dh_v + \frac{1}{2m}\bar{h}_v(iD_\perp)^2 h_v + \frac{g}{4m}\bar{h}_v\sigma^{\mu\nu}G_{\mu\nu}h_v. \tag{I.59}$$

The first term is the lowest-order Lagrangian of HQET describing the residual QCD dynamics of the heavy quark once the kinematic dependence of m_b is separated out. The remaining terms are leading power corrections that can be treated as small perturbations.

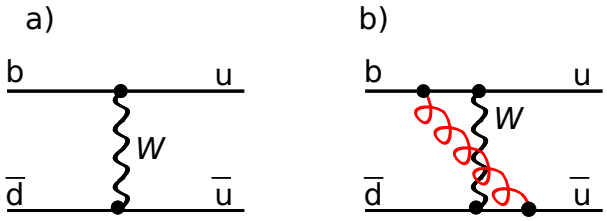

Figure I.9: The two simplest Feynman diagrams representing the transition $b \to ud\bar{u}$. Diagram a) neglects effects of QCD, while diagram b) is the simplest QCD correction diagram.

The second and third terms of Equation I.59 can be treated as perturbations by writing the Hamiltonian as $H = H_0 + H'$, where

$$H' = -\int \mathrm{d}^3x \left(\frac{1}{2m} \bar{h}_v (iD_\perp)^2 h_v + \frac{g}{4m} \bar{h}_v \sigma^{\mu\nu} G_{\mu\nu} h_v \right). \tag{I.60}$$

At the first order, one can write the mass of the heavy hadron as

$$m_B = m_b + \bar{\Lambda} - \frac{\mu_\pi^2 - \mu_G^2}{2m_b}, \tag{I.61}$$

where $\bar{\Lambda}$ is a constant term which accounts for the mass contribution due to light degrees of freedom, and

$$\mu_\pi^2 = \frac{\langle B|\bar{b}(i\vec{D})^2 b|B\rangle}{2m_B} + \mathcal{O}\left(\frac{1}{m_b}\right), \tag{I.62}$$

$$\mu_G^2 = \frac{\langle B|\bar{b}\frac{g_s}{2}\sigma_{\mu\nu}G^{\mu\nu} b|B\rangle}{2m_B} + \mathcal{O}\left(\frac{1}{m_B}\right). \tag{I.63}$$

The quantities μ_π^2 and μ_G^2 are widely used in Heavy Quark Physics and are named kinetic and chromo-magnetic operator, respectively.

Operator Product Expansion

The Operator Product Expansion (OPE) is a basic concept with almost infinite possible variants or specializations. Reading how various author present the topic, it can be noticed how despite its conceptual simplicity almost nobody dares to give a definition preferring a more operative approach using examples. Hereafter, the Operator Product Expansion is presented through an example I found in an enlightening series of lectures about Heavy Quark Physics given by G. Buchalla in 2001 [39].

Buchalla considers the simple process $\bar{B}^0 \to u\bar{u}$ as diagrammatically represented in Figure I.9a. This process triggers the non-leptonic decays such as $\bar{B}^0 \to \pi^+\pi^-$. Clearly, the complexity of the problem increases when one starts considering QCD corrections, indeed all the quarks involved are *"dressed"* of QCD interactions of all kinds.

Instead of trying to give the whole description at once, Buchalla tries to define a suitable expansion parameter, as commonly done in theoretical physics. Since the mass of the W-boson is much larger than the momentum scale of the process, a reasonable expansion parameter could be $\frac{p}{M_W}$. Then he looks for a parametrization of the decay amplitude as

$$A = C_1 \left(\frac{M_W}{Q^2} \right) \langle O_1 \rangle + M_2 \left(\frac{M_W}{Q^2} \right) \langle O_2 \rangle + ... \tag{I.64}$$

where O_1 and O_2 are local four-quark operators and C_1 and C_2 are named Wilson coefficients obtained from further calculations.

Finally, he considers the amplitude of the process neglecting QCD interactions

$$A = \frac{g_W^2}{8} V_{ud}^* V_{ub} \frac{i}{k^2 - M_W^2} (\bar{d}_i u_i)_{V-A} (\bar{u}_j b_j)_{V-A}, \tag{I.65}$$

with the subscripts i and j being colour indices, and applies an Operator Product Expansion to the first order in $\frac{p}{M_W}$:

$$A = \underbrace{\left[-i \frac{G_F}{\sqrt{2}} V_{ud}^* V_{ub} \right]}_{C_1} \cdot \underbrace{\left[(\bar{d}u)_{V-A} (\bar{u}b)_{V-A} \right]}_{\langle O_1 \rangle} + \mathcal{O} \left(\frac{k^2}{M_W^2} \right), \tag{I.66}$$

where $\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8 M_W^2}$.

As stated above, OPE is very simple in principle but difficulties can quickly arise. For example, switching on the strong interaction, QCD generates another operator, represented in Figure I.9b,

$$O_2 = (\bar{d}_i u_j)_{V-A} (\bar{u}_j b_i)_{V-A}, \tag{I.67}$$

which has the same Dirac and flavour structure, but a different colour form. The computation of the Wilson coefficient in this case is out of the scope of this introduction. The choice of the expansion parameter and the technique used to evaluate the Wilson coefficient define the many flavours of the Operator Product Expansion, of which only few are briefly introduced below.

It is important to point out that the most important property of the OPE in Equation I.64 is the *factorization* of long- and short-distance contributions. All effects of QCD above some factorization scale μ (short distances) are contained in the Wilson coefficient C. All the low-energy contributions below μ (long-distances) are collected into the matrix elements of the local operators $\langle O_i \rangle$. In this way the short-distance part of the amplitude can be systematically extracted and calculated in perturbation theory. The problem to evaluate the matrix elements of local operators between hadron states requires in general non-perturbative techniques, as for example lattice QCD or QCD sum-rules, but it is considerably simpler than the original problem of the full standard-model amplitude.

QCD sum rules, or Shifman–Vainshtein–Zakharov (SVZ) sum rules

QCD sum rules relate hadronic parameters, such as masses and magnetic moments to a few characteristics of the vacuum of quantum chromodynamics (QCD): gluon and quark condensate, the expectation value of $\bar{q}q$ in the QCD vacuum: $\langle 0|\bar{q}q|0\rangle$ [31]. The method is based on the Operator Product Expansion, adapted in the mid-1970s to QCD [40].

The great advantage of QCD sum rules over other methods is the analyticity of the described quantities, in other words it is possible to relate quantities in different kinematical regions.

To evaluate lifetimes, the QCD sum rules are applied to a complete set of inclusive decay modes to evaluate the partial decay widths. As already mentioned, calculations on inclusive decay modes suffer from the ambiguity in the definition of the energy scale of the process Q^2, whose choice is the main source of uncertainty in the results obtained with QCD sum rules.

Light-front quantization

The light-front quantization provides an alternative to ordinary equal-time quantization useful to define a relativistic description of bound systems, as for example hadrons, in terms of quantum-mechanical wave functions.

The quantization is based on the choice of light-front coordinates $x^{\pm} \equiv t \pm z$, where x^+ plays the role of time and the corresponding spatial coordinate is x^-. Here, t is the ordinary time, z is one Cartesian coordinate, and c is the speed of light. The other two Cartesian coordinates, x and y, are untouched and often called transverse or perpendicular.

The non-perturbative QCD effects are grouped into the *light-front* wave-function of the hadron. The internal motion of the heavy quark inside the heavy flavour meson H_Q is described by the distribution function $F(x) = \int \mathrm{d}^2 p_\perp |\psi(x, p_\perp^2)|^2$, where $|\psi(x, p_\perp^2)|^2$ represents the probability to find a quark Q carrying a light-front fraction $x = p_Q^+ / P_{H_Q}^+$ of the meson momentum and a transverse relative momentum squared $p_\perp^2$.

Light-front quantization has been adopted successfully to describe B hadron decays and to predict their decay branching fractions.

Heavy Quark Expansion (HQE) or $1/m_Q$ expansion

The most advanced flavour of the Operator Product Expansion is the Heavy Quark Expansion (HQE) which uses the optical theorem and from the Heavy Quark Effective Theory (HQET) to define a relation between the *total* decay width of a particle and few matrix elements that can be evaluated using gauge lattice QCD.

Following the OPE approach, an effective Hamiltonian operator $\mathcal{H}_{\text{eff}}$ is defined as the weighted sum of all the possible operators, as depicted in Figure I.10.

To the operators considered above,

$$O_1^p = (\bar{d}_i p_i)_{V-A} (\bar{p}_j b_j)_{V-A}, \tag{I.68}$$

$$O_2^p = (\bar{d}_i p_j)_{V-A} (\bar{p}_j b_i)_{V-A}, \tag{I.69}$$

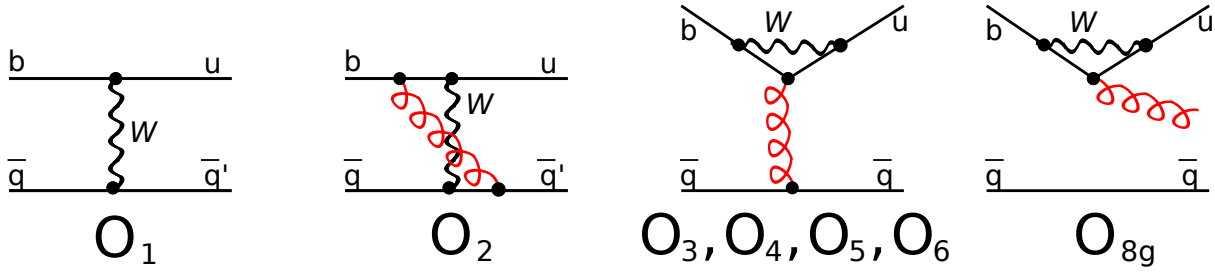

Figure I.10: Feynman diagrams of the seven operators used in the definition of the effective Hamiltonian $\mathcal{H}_{\text{eff}}$.

with $p = u, c$, it is reasonable to add the QCD-penguin-diagram operators

$$O_3 = (\bar{d}_i b_i)_{V-A} \sum_q (\bar{q}_j q_j)_{V-A}, \tag{I.70}$$

$$O_4 = (\bar{d}_i b_j)_{V-A} \sum_q (\bar{q}_j q_i)_{V-A}, \tag{I.71}$$

$$O_5 = (\bar{d}_i b_j)_{V-A} \sum_q (\bar{q}_j q_i)_{V+A}, \tag{I.72}$$

$$O_6 = (\bar{d}_i b_j)_{V-A} \sum_q (\bar{q}_j q_i)_{V+A}. \tag{I.73}$$

In some works, an additional penguin-diagram operator is added to $\mathcal{H}_{\text{eff}}$ to include processes where the gluon does not couple with the spectator quark current. Namely,

$$O_{8g} = -\frac{g}{8\pi^2} m_b \bar{d}_i \sigma^{\mu\nu} (1+\gamma^5) T^a_{ij} b_j G^a_{\mu\nu} \tag{I.74}$$

where T^a is the a-th generator of the $SU(3)_C$ symmetry and $G^a_{\mu\nu}$ represents the field-strength tensor.

The effective Hamiltonian operator can thus be written as

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \sum_{p=u,c} \lambda_p \left[C_1 O_1^p + C_2 O_2^p + \sum_{i=3,4\ldots,6,8g} C_i O_i \right] + h.c. \tag{I.75}$$

where $\lambda_p \equiv V^*_{pd} V_{pb}$ introduces the CKM matrix elements in the expression for the effective Hamiltonian. Note that an additional λ_t term should be included in the sum, but CKM unitarity has been exploited to cancel it.

The effective Hamiltonian defined in I.75 can be used to define the total decay width of a B meson as

$$\Gamma(B \to X) = \frac{1}{m_B} \sum_X \int_{PS} (2\pi)^4 \delta^{(4)}(p_B - p_X) |\langle X | \mathcal{H}_{\text{eff}} | X \rangle|^2, \tag{I.76}$$

using the optical theorem one can write

$$\Gamma(B \to X) = \frac{1}{2m_B}\langle B|\mathcal{T}|B\rangle \equiv \langle \mathcal{T}\rangle, \tag{I.77}$$

where the transition operator $\mathcal{T}$ is defined as

$$\mathcal{T} = \text{Im } i \int \mathrm{d}^4x T\mathcal{H}_{\text{eff}}(x)\mathcal{H}_{\text{eff}}(0), \tag{I.78}$$

with T being the time-ordered product and is commonly used because it allows a direct evaluation in terms of Feynman diagrams.

A second operator-product-expansion, exploiting the large value of the b-quark mass m_b allows to write the transition operator as

$$\mathcal{T} = \frac{G_F^2 m_b^5}{192\pi^3}|V_{cb}|^2 \left[c_{3,b}\bar{b}b + \frac{c_{5,b}}{m_b^2}\frac{\langle B|\bar{b}g_s\sigma_{\mu\nu}G^{\mu\nu}b|B\rangle}{2M_B} + \frac{c_{6,b}}{m_b^3}\frac{\langle B|(b\bar{q})_\Gamma(\bar{q}b)_\Gamma|B\rangle}{M_B} + ...\right]. \tag{I.79}$$

In other words, the transition operator is expressed as a sum of arbitrary operators whose dimension matches the dimension of the m_b^n suppression term in order to preserve the homogeneity of the added terms. Therefore the leading order is provided by the operator with the smallest dimension, which is $\bar{b}b$ having dimension three[2], followed by operators of dimension five representing the one-gluon interaction between the b-quark and the spectator quark ($\bar{b}\sigma_{\mu\nu}G^{\mu\nu}b$), and dimension six representing the b-quark and the spectator quark interacting through a four-fermion interaction ($(b\bar{q})_\Gamma(\bar{q}b)_\Gamma$).

The lack of a dimension-four operator is not a choice, but rather an interesting result of the Heavy Quark Expansion.

The expression for the total decay width

$$\Gamma(B \to X) = \frac{G_F^2 m_b^5}{192\pi^3}|V_{cb}|^2 \left[c_{3,b}\frac{\langle B|\bar{b}b|B\rangle}{2m_B} + \frac{c_{5,b}}{m_b^2}\frac{\langle B|\bar{b}g_s\sigma_{\mu\nu}G^{\mu\nu}b|B\rangle}{2m_B} + \frac{c_{6,b}}{m_b^3}\frac{\langle B|(q\bar{b})_\Gamma(\bar{q}b)_\Gamma|B\rangle}{m_B} + ...\right] \tag{I.80}$$

can be simplified using the relation [41]

$$\frac{\langle B|\bar{b}b|B\rangle}{2m_B} = 1 - \frac{\mu_\pi^2 - \mu_G^2}{2m_b^2} + \mathcal{O}\left(\frac{1}{m_b^3}\right) \tag{I.81}$$

[2] **About dimensional analysis in QFT.** Dimensional analysis is widely used in QFT. The basic idea is that using the natural unit system where $\hbar = c = 1$, the dimension of any quantity can be expressed in units of energy at some power. The Lagrangian action S has to be dimensionless to be evaluated as argument of an exponential function when defining the evolution operator. Given the definition of $S = \int \mathrm{d}^4x\mathcal{L}$ and agreed that the dimension of the coordinates x^μ is the inverse of energy (dimension –1), one concludes that the dimension of the Lagrangian density is four, being the free Lagrangian density of the b quark defined as $\bar{b}(i\gamma^\mu\partial_\mu - m_b)b$, the dimension of $\bar{b}m_b b$ as to be four as well, which makes the fermion field b to have dimension 3/2.

where μ_π^2 and μ_G^2 are defined in Equations I.62 and I.63,

$$\boxed{\Gamma(B \to X) = \frac{G_F^2 m_b^5}{192\pi^3} V_{cb}^2 \left\{ c_{3,b} \left[1 - \frac{\mu_\pi^2 - \mu_G^2}{2m_b^2} + \mathcal{O}\left(\frac{1}{m_b^3}\right) \right] + 2c_{5,b} \left[\frac{\mu_G^2}{m_b^2} + \mathcal{O}\left(\frac{1}{m_b^3}\right) + \frac{c_{6,b}}{m_b^3} \frac{\langle B|(\bar{b}q)_\Gamma(\bar{q}b)_\Gamma|B\rangle}{M_B} \right] + ... \right\}} \quad \text{(I.82)}$$

The importance of Equation I.82 could be hardly overestimated: it expresses the lifetime of a heavy hadron as the lifetime of the quark, the first term, plus corrections weighted by the inverse of the mass of the heavy quark at some power. The fact that the first power does not appear is a non trivial result combination of the fact that dimension-four operators does not enter the decay processes of the heavy mesons, and of the fact that Equation I.81 does not contain a correction term of order $(\frac{1}{m_b})$. This implies that the lifetimes of the hadrons formed of a b quark and light (anti-)quark(s), have to be very close to each other. The coefficients $c_{n,b}$ are obtained through direct comparison of Equation I.82 and Equation I.77, as shown explicitly for example in Ref. [42] for some simple cases.

Because of unfortunate results by the LEP experiments, finding a value of the Λ_b baryon lifetime too small to be consistent with HQET predictions, for many years it has been believed [43] that HQET, and in general most of the OPE techniques, suffer from *local quark-hadron duality violation*.

Hadron-quark duality is a principle which states that the inclusive decay rates as the sum over all inclusive channels and the inclusive rate as predicted by the heavy quark expansion are dual to each other. This means they are both valid representations of the same quantity using different descriptions. The term *local* refers to the fact that the energy scale m_b is a fixed quantity, as opposed to *e.g.* the center-of-mass energy in e^+e^- annihilation which can be averaged to obtain suitably defined *global* quantities. In principle, the hadronic description leads to the true result, as measured in the experiments, but computing all the exclusive processes is technically unfeasible for heavy mesons [39].

A violation of the *quark-hadron duality* is expected but the numerical size is assessed to be small and could not explain the large discrepancy observed for the lifetime of the Λ_b baryon.

The impressive aspect of this story is that HQE is a non trivial effective model which has been under scrutiny for twenty years trying and failing to accommodate experimental values, finally disproved by the LHC experiments [44–46], which found results in good agreement with the HQE predictions. This makes the HQE the most trusted approach to lifetime determinations. Taking as input recently calculated Lattice-QCD matrix elements it has led to lifetime predictions of impressive precision and accuracy, as summarized in Table I.3. A more complete summary of the comparison between HQE predictions and recent experimental values is given in Refs. [43, 47].

	HQE prediction [42]	LHCb experimental results	
$\tau(B^-)/\tau(\bar{B}^0)$	$1.04^{+0.05}_{-0.01} \pm 0.03$	1.079 ± 0.007 ps	[1]
$\tau(\Lambda_b^0)$	1.41 ± 0.08 ps	$1.479 \pm 0.009 \pm 0.010$ ps	[45]
$\tau(\Xi_b^-)$	1.56 ± 0.10 ps	$1.599 \pm 0.041 \pm 0.018 \pm 0.012(\tau_{B^-})$ ps	[48]
$\tau(\bar{B}_s^0)$	1.521 ± 0.008 ps	$1.535 \pm 0.015 \pm 0.012 \pm 0.007(\tau_{B^-})$ ps	[49]
$\tau(\Xi_b^0)$	—	$1.477 \pm 0.026 \pm 0.014 \pm 0.013(\tau_{\Lambda_b})$ ps	[50]

Table I.3: Comparison of the predicted value for the lifetime of some b-hadrons with recent experimental results. Theoretical predictions are taken from the recent review in Ref. [42].

I.5.6 Predictions of the B_c^+ lifetime

The most naive prediction one could make on the B_c^+ lifetime is that since the annihilation decay is negligible, the expression of the total decay width could simplify to

$$\Gamma(B_c^+ \to X) = \Gamma(B_c^+; \bar{b} \to X) + \Gamma(B_c^+; c \to X). \tag{I.83}$$

As discussed in Section I.5.5, the decay width of a heavy quark is well approximated by that of heavy hadrons containing that quark and a light anti-quark. This is true for b-hadrons, it is less appropriate for c-hadrons, but one may still write

$$\Gamma(B_c^+) \sim \Gamma(D^0) + \Gamma(B_d^0). \tag{I.84}$$

To assess the uncertainty on this rough evaluation, one may try using a different c-hadron, for instance the Λ_c baryon: $\Gamma(B_c^+) \sim \Gamma(\Lambda_c^+) + \Gamma(B_d)$. Following these considerations one writes

$$\tau_{B_c^+}^{(\text{HADRONS})} \sim 0.2 \text{ — } 0.6 \text{ ps}.$$

Such a basic estimation predicts also that the $c \to s$ decay dominates the decay in a ratio of roughly 4:1.

However, the phase space available for decays $(\bar{b}c) \to \bar{b}s\bar{d}u$ is more limited than it is in D^0 decays, and it is much more limited than in Λ_c^+ decays. This could make the B_c^+ lifetime longer, and suppress the importance of the $c \to s$ transitions.

Correcting the prediction for the different phase spaces available, but still inheriting the decay width of quarks from other heavy mesons, Lusignoli and Masetti predicted [51]

$$\tau_{B_c^+}^{(\text{HADRONS+PS})} \sim 0.5 \text{ ps}.$$

A different estimation by Eichten and Quigg proposed that the quark decay width as $\Gamma_Q \propto G_F^2 m_Q^5$ should be evaluated using the quark mass *"reduced"* by effect of the binding energy within the hadron. For ordinary b-mesons, this can be seen as a redefinition of the quark mass. However, for the B_c^+ meson, this would translate in subtracting a fixed quantity μ_{BE} (the binding energy) to both the c- and b-quark masses. The relative effect

Quark masses (GeV/c^2)			Partial decay widths (ps^{-1})				$\tau_{B_c^+}$
m_b	m_c	m_s	$\Gamma_{B_c^+}(\bar{b} \to X)$	$\Gamma_{B_c^+}(c \to X)$	$\Gamma_{B_c^+}$(PI)	$\Gamma_{B_c^+}$(annih.)	(ps)
5.0	1.5	0.2	0.694	1.148	-0.115	0.193	0.54
4.8	1.35	0.15	0.576	0.725	-0.132	0.168	0.75
5.1	1.6	0.45	0.635	1.033	-0.101	0.210	0.55
5.1	1.6	0.20	0.626	1.605	-0.101	0.210	0.43
5.05	1.55	0.20	0.623	1.323	-0.107	0.201	0.48
5.0	1.5	0.15	0.620	1.204	-0.114	0.193	0.53

Table I.4: Value of the lifetime of the B_c^+ meson and of the partial contributions to the decay width for different choices of the quark masses. The results have been previously published in Ref. [55].

would suppress more the *charm* decays than the *beauty* decays, and since the mass enters at the fifth power in the decay amplitude, small mass shifts translate in large changes in the predicted lifetime. For a binding energy of $\mu_{BE} = 500$ MeV, the *beauty* quark decay dominates over the charm decay leading to the prediction [52]

$$\tau_{B_c^+}^{(\text{QUARK MASS-SHIFT})} \sim 1.3 \text{ ps}.$$

In the years of the development of the HQE, Bigi applied the expansion to the decay widths $\Gamma(B_c^+; \bar{b} \to X)$ and $\Gamma(B_c^+; c \to X)$ separately, using $\frac{1}{m_b}$ and $\frac{1}{m_c}$ as expansion term. The annihilation term is treated separately and included in the total decay width to obtain the lifetime prediction [53]

$$\tau_{B_c^+}^{(\text{HQE})} \sim 0.4 \text{ ps}.$$

Using a similar approach Beneke and Buchalla found that varying the value assumed for the mass of the charm quark between 1.4 and 1.6 GeV the lifetime prediction spans in the wide range [54]

$$\tau_{B_c^+}^{(\text{HQE2})} \sim 0.4 \text{ — } 0.7 \text{ ps}.$$

Three years later, in 1999, Onishchenko published a study about the dependence of the result obtained using HQE on the values used for the masses of the *beauty*, *charm*, and *strange*. The lifetime obtained spans in a relatively large range [55]

$$\tau_{B_c^+}^{(\text{HQE3})} \sim 0.43 \text{ — } 0.75 \text{ ps},$$

where the detail is given in Table I.4.

An alternative approach is based on potential models as those described in Section I.4.1 for quarkonia. Once defined a potential describing the interaction of the two quarks, the Schrödinger equation is solved to find the wave functions, then the explicit overlap integrals are calculated to find the partial decay widths of several (in principle all) processes that are then summed up to find the total decay width. Obviously, changing the potential,

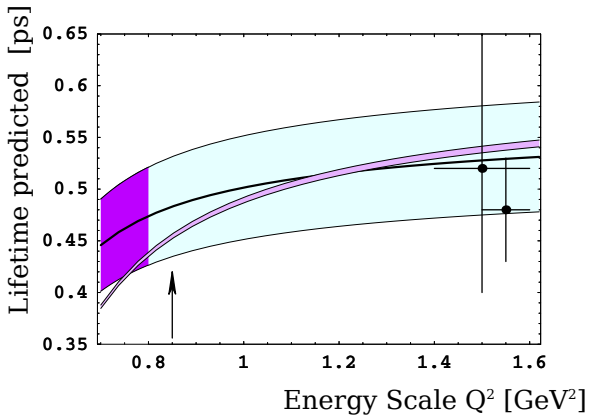

Figure I.11: Dependence of the lifetime prediction as obtained using QCD sum rules on the assumed value for the energy scale Q^2. The markers with error bars represent the predictions by Bigi [53] and Onishchenko [55]. The arrow indicates the assumption on Q^2 used by Kiselev $Q^2 = (0.85\,\mathrm{GeV})^2$ when deriving his prediction using QCD sum rules and is close to the *charm*-quark mass. The Figure appears in many works, for example in Refs [30, 33, 56].

the assumptions of the quantization or the technique used to describe the many decay channels, different results can be found.

Kiselev *et al.* in 2000 published an expected lifetime of [33, 56]

$$\tau_{B_c^+}^{(\mathrm{PM,OPE})} \sim 0.55 \pm 0.15 \text{ ps},$$

obtained using the OPE inclusive method and the potential model exclusive calculation, and averaging the results. In the same paper Kiselev, predicts

$$\tau_{B_c^+}^{(\mathrm{SR})} \sim 0.48 \pm 0.05 \text{ ps},$$

using QCD sum rules. As discussed above, even if the quark mass ambiguity is less relevant for QCD sum rules, still the choice of the energy scale of the process introduces a certain degree of arbitrariness in the lifetime predictions. In Figure I.11 the the lifetime prediction from QCD sum rules is shown as a function of the assumed energy scale.

Using another potential model and the formalism of the light-front quantization which introduces some advantages in the calculation of the non-perturbative part of the QCD corrections, Anisimov *et al.* predicted the value [57]

$$\tau_{B_c^+}^{(\mathrm{LF})} \sim 0.59 \pm 0.06 \text{ ps}$$

obtained by adding the decay widths of 84 exclusive decay modes and 44 inclusive decay modes. In this case the dominant uncertainty is the choice of the light-front wave-functions.

Despite the many new techniques developed to predict lifetime values, the idea of predicting the B_c^+ lifetime using the lifetime of other heavy hadrons as input is still tempting. In 2000, Chang *et al.* applied the HQE to *beauty* and *charmed* mesons and to

Author		Year	$\tau_{B_c^+}$ (ps)	Technique
—			0.2 – 0.6	Sum of the decay widths of b- and c-hadrons
Lusignoli, Masetti	[51]	1990	0.5	Phase-space corrected PM estimation
Eichten, Quigg	[52]	1994	1.3	$\Gamma(c \to s)$ suppressed from binding energy
Bigi	[53]	1995	0.4	Heavy Quark Expansion
Beneke, Buchalla	[54]	1996	0.4 – 0.7	Heavy Quark Expansion (varying m_c)
Anisimov *et al.*	[57]	1998	0.59 ± 0.06	Potential model in light-front quantization
Onishchenko	[55]	1999	0.43 – 0.75	HQE (varying m_b, m_c, and m_s)
Kiselev *et al.*	[56]	2000	0.55 ± 0.15	Potential models, and OPE. Averaged.
Kiselev *et al.*	[56]	2000	0.48 ± 0.05	QCD sum rules
Chang *et al.*	[58]	2000	0.37 – 0.47	HQE with inputs from b- and c-hadrons
Rai *et al.*	[59]	2013	0.29	Potential models

Table I.5: Some theoretical estimations for the lifetime of the B_c^+ meson as obtained with the various techniques described above.

the B_c^+ using the values obtained for the coefficients of the $B_{u,d}$ and D mesons as input for the expansion of the B_c^+ decay width. The result should be more robust against variations of the quark masses, but still suffers from some ambiguity so that the final result lays in the range [58]

$$\tau_{B_c^+}^{(\mathrm{HADRONS+HQE})} \sim 0.37 \text{ — } 0.47 \text{ ps}$$

Theoretical studies trying to predict the B_c^+ lifetime are still ongoing and particularly active in Indian institutes. For example, at the 2013 HADRON Conference (2013), A. K. Rai proposed a new estimation based on potential models [59]

$$\tau_{B_c^+}^{(\mathrm{PM})} \sim 0.29 \text{ ps}$$

The discussed lifetime estimations are summarized in Table I.5

Chapter II

Experimental advances on B_c^+ physics

Introduction

The experimental history of the B_c^+ meson began at CERN at the Large Electron-Positron collider (LEP) operating at an energy in the center-of-mass of the e^+e^- pair of $m(Z^0)c^2 \sim 90$ GeV in order to excite the resonance Z^0, then decaying to a pair of leptons or quarks [60]. The B_c^+ production mechanism discussed in Section I.5.2, does not apply to electron-positron colliders, where the quarks and gluons hard enough to produce a B_c^+ meson in gg- or $q\bar{q}$-fusion are significantly less. The mechanism of a possible B_c^+ production at LEP is therefore the fragmentation as shown in Figure II.1, which is quite unlikely as discussed in the introductory chapter.

Analyzing a dataset of approximately 4.2×10^6 hadronic Z^0 decays collected between 1990 and 1995 the OPAL Collaboration published upper limits to the production cross-section of the B_c^+ meson times the decay branching fraction of the three decay modes they considered: $B_c^+ \to J/\psi\,\pi^+$, $B_c^+ \to J/\psi\,\mu^+\nu_\mu$, and $B_c^+ \to J/\psi\,a_1^+$. The histograms of the invariant mass of the selected candidates is shown in Figure II.2.

As discussed in Section I.5.2, the B_c^+ production is not accessible at e^+e^- colliders operating at the mass of the $\Upsilon(4S)$ or $\Upsilon(5S)$, thus no experimental input from the

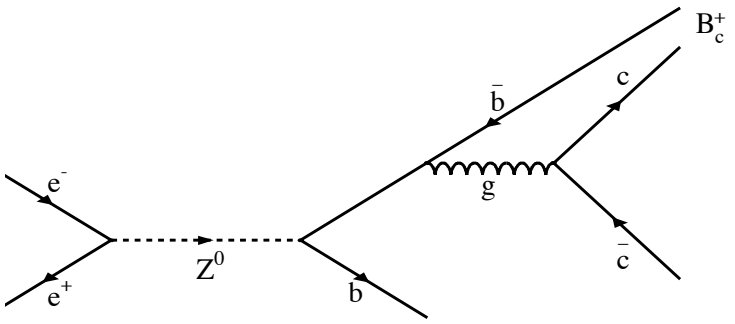

Figure II.1: Dominant production mechanism of the B_c^+ meson at e^+e^- colliders as LEP. The drawing was published by the OPAL Collaboration [60].

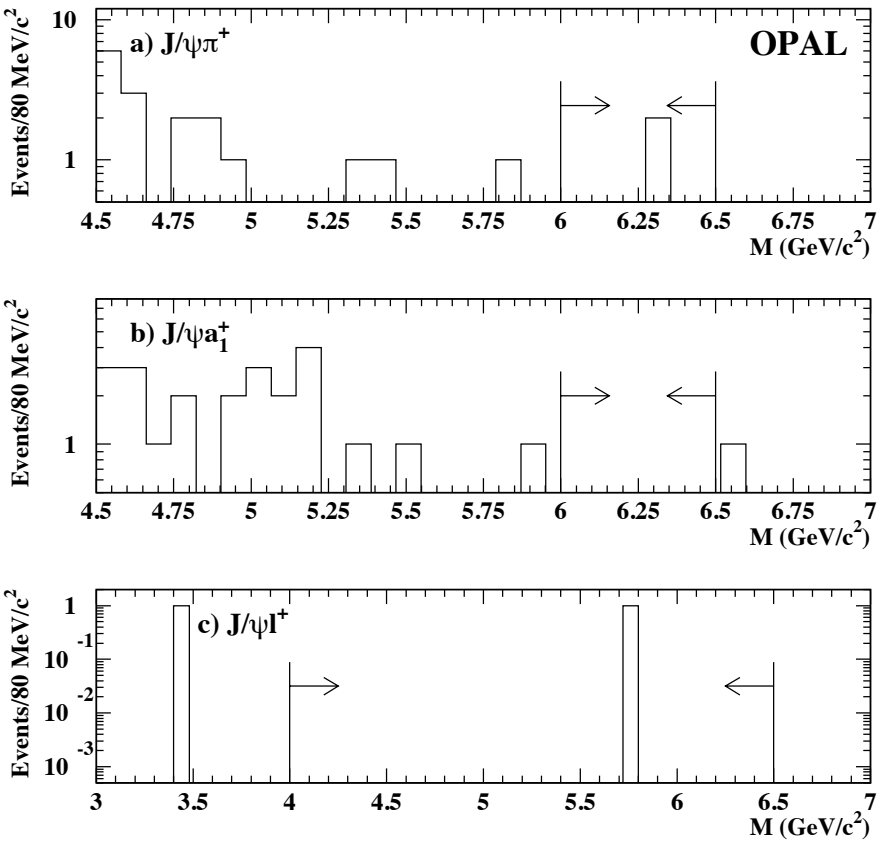

Figure II.2: Invariant mass distribution of (a)$J/\psi\,\pi^+$ combinations, (b) $J/\psi\,{\rm a_1}^+$ combinations, and (c) $J/\psi\,\ell^+$ combinations in the OPAL data-sample collected between 1990 and 1995 corresponding to approximately 4.2×10^6 hadronic Z^0 decays. The signal regions where the B_c^+ candidates are expected to lay are also shown. The figure was published in [60].

b-factories is available.

The first hadronic collider with an energy sufficiently high to produce the B_c^+ meson was the Tevatron at Fermilab, $p\bar{p}$ asymmetric collider operating at an energy in the center of mass $\sqrt{s} = 1.96$ TeV. Two detectors were installed in the interaction points of the accelerator: CDF and D0. While they entered the history of Science with the discovery of the *top* quark in 1995, CDF and D0 also discovered the B_c^+ meson and started the characterization of this new meson performing measurements of its production, mass and lifetime.

The discovery of the B_c^+ meson was performed by the CDF experiment in 1998 studying the final state $J/\psi\,\mu^+\nu$ [61]. As for the case of the LEP analysis, when considering semileptonic decays the signal region in the observed invariant mass spectrum, defined

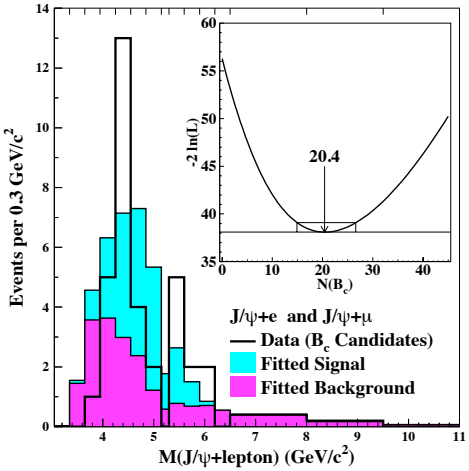

Figure II.3: Histogram of the $J/\psi\,\ell^+$ mass comparing the signal and background contributions determined in the likelihood fit to the combined data for $J/\psi\,e^+$ and $J/\psi\,\mu^+$. The inset shows the log-likelihood function $-2\ln(L)$ vs. the number of B_c^+ mesons. The Figure was published in Ref. [61].

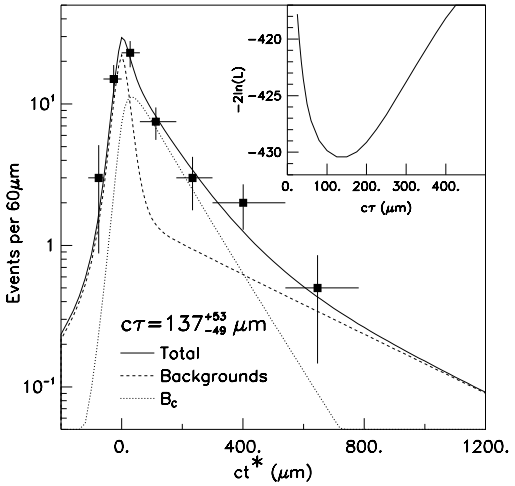

Figure II.4: The distribution in ct for the combined $J/\psi\,\mu^+$ and $J/\psi\,e^+$ data along with the fitted curve and contributions to it from signal and background. The inset shows the log-likelihood function vs. $c\tau$ for B_c^+. The Figure was published in Ref. [61].

here to lay between 4 and 6.5 GeV/c^2, increases significantly because of the additional degree of freedom due to the missing information on the momentum of the neutrino. The wide mass spectrum with the template distributions for the signal and the background whose normalizations are fitted to the observed candidates, is shown in Figure II.3. The likelihood fit used converges to a normalization for the signal template distribution of $20.4^{+6.2}_{-5.5}$ events with a likelihood profile shown in the inset of Figure II.3. A large number of Monte Carlo toy experiments has been generated with candidates following the distribution of the background template. The fraction of such toy experiments that, fitted with the combined template for signal and background, result in a number of signal events of 20.4 or larger is 0.63×10^{-6}; corresponding to a statistical significance of 4.8 standard deviations.

Using the 20.4 events selected, the CDF Collaboration estimated the production cross-section of the B_c^+ meson times the $B_c^+ \to J/\psi\,\ell^+\nu$ branching fraction, normalized to the B^+ production cross-section times the $B^+ \to J/\psi\,K^+$ decay branching fraction, to be

$$\mathcal{R}(J/\psi\,\ell\nu) \equiv \frac{\sigma(B_c^+) \times \mathcal{B}(B_c^+ \to J/\psi\,\ell^+\nu)}{\sigma(B^+) \times \mathcal{B}(B^+ \to J/\psi\,K^+)} = 0.132^{+0.041}_{-0.037}\,(\text{stat}) \pm 0.031\,(\text{syst})^{+0.032}_{-0.020}(\text{lifetime}), \tag{II.1}$$

where the systematic uncertainty is dominated by possible differences between the simulated and collected datasets. The efficiency of selection criteria based on the detachment of the decay vertex from $p\bar{p}$ primary vertex depends on the lifetime of the B_c^+ meson. The uncertainty on the latter reflects into an uncertainty on the selection efficiency and therefore on the production cross-section, quoted separately as *lifetime* uncertainty in the measurement of $\mathcal{R}(J/\psi\,\ell\nu)$.

In the same publication, CDF showed the distribution of the $J/\psi\,\ell$ candidates with respect to the transverse pseudo-proper decay length $ct_{\perp,\mathrm{ps}}$ defined as

$$ct_{\perp,\mathrm{ps}} = \frac{m(J/\psi\,\ell^+)\cdot L_{xy}(J/\psi\,\ell^+)}{|p_T(J/\psi\,\ell^+|)}, \tag{II.2}$$

where $m(J/\psi\,\ell^+)$ is the invariant mass of the combination $J/\psi\,\mu^+$, while $p_T(J/\psi\,\ell^+)$ and $L_{xy}(J/\psi\,\ell^+)$ are the projections on the plane orthogonal to the beam axis of the momentum of the $J/\psi\,\mu^+$ combination and of the flight distance of the B_c^+ candidate. The distribution of the selected candidates, superposed to the template distributions for signal and background is shown in Figure II.4. The template distribution of the decay time has a free parameter defining the slope which is related to the lifetime of the B_c^+ meson. The value of the B_c^+ lifetime which maximizes the agreement of the template with data is

$$\tau_{B_c^+}^{(\mathrm{CDF})} = 0.46^{+0.18}_{-0.16}\,(\mathrm{stat}) \pm 0.03\,(\mathrm{syst})\ \mathrm{ps}, \tag{II.3}$$

where the second uncertainty, systematic, is dominated by the uncertainties on the chosen template distributions.

A new era for the B_c^+ meson begun with the first injection of the Large Hadron Collider (LHC) operating at an energy in the center of mass of $\sqrt{s} = 7$ TeV and designed to ramp up to $\sqrt{s} = 14$ TeV in 2015. The large number of B_c^+ candidates collected by the LHC experiments allows precision measurements and the first spectroscopy studies on electromagnetic decays towards B_c^+ states.

In this chapter, a brief review of the experimental status of the knowledge of the B_c^+ meson is given, which intends to depict the experimental status of the fast paced context in which the work described in this Thesis was developed.

II.1 Mass measurements

Clearly the mass measurement obtained studying the ending point of the distribution of the invariant mass of the $J/\psi\,\ell^+$ combination is not very precise. The discovery paper reported the value

$$m(B_c^+)^{(\mathrm{CDF})} = 6.40 \pm 0.39\,(\mathrm{stat}) \pm 0.13\,(\mathrm{syst})\ \ \mathrm{GeV}/c^2. \tag{II.4}$$

The fully reconstructed channels have smaller branching ratio and therefore the number of events that can be collected in a finite amount of time is smaller. For this reason, the publication about the first observation of a fully reconstructed B_c^+ decay, namely $B_c^+ \to J/\psi\,\pi^+$, is dated 2008. Ten years later than the observation of the meson itself.

The result, obtained from the fit on the selected 108 ± 15 signal candidates, is [62]

$$m(B_c^+)^{(\mathrm{CDF\,II})} = 6.2756 \pm 0.0029\,(\mathrm{stat}) \pm 0.0025\,(\mathrm{syst})\,\mathrm{GeV}/c^2, \tag{II.5}$$

with an uncertainty reduced by more than two orders of magnitude.

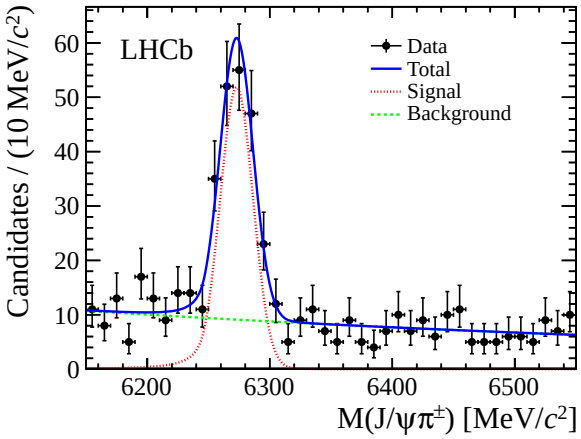

Figure II.5: Invariant mass distribution of $B_c^+ \to J/\psi\,\pi^+$ candidates used in the mass measurement. The fit to data is superimposed. The Figure was published in [64].

With the larger statistics available at the LHC, the mass measurement becomes limited by systematic effects. Using a more suitable decay, a better mass measurement can be obtained with fewer candidates. It has been shown recently that using the $B_c^+ \to D_s^+ J/\psi$ and $B_c^+ \to J/\psi\, p\bar{p}\nu_\mu$ decay modes the B_c^+ mass resolution can be improved significantly.

Mass measurement using the decay $B_c^+ \to J/\psi\,\pi^+$

The latest (and possibly the last) measurement of the B_c^+ meson mass using the $B_c^+ \to J/\psi\,\pi^+$ channel has been achieved by the LHCb Collaboration using the pp data collected in the first half of 2011, with a center-of-mass energy $\sqrt{s} = 7$ TeV.

The data collected are shown in Figure II.5 superimposed to the data model composed of an exponential background summed to a double-sided Crystal-ball function [63].

The mass measurement is [64]

$$m(B_c^+)^{(\mathrm{LHCb,J/\psi\,\pi^+})} = 6.2737 \pm 0.0013\,(\mathrm{stat}) \pm 0.0016\,(\mathrm{syst})\,\mathrm{GeV}/c^2. \qquad \text{(II.6)}$$

The uncertainty on the measurement is dominated by the systematic error on the momentum scale, namely the bias introduced by the LHCb spectrometer in reconstructing the momentum of the charged tracks combined to form the B_c^+ candidates. To reduce the effect of the uncertainty on the momentum scale, the B_c^+-meson mass is measured relatively to the mass of the B^+ meson, as measured reconstructing the $J/\psi\,K^+$ decay mode:

$$m(B_c^+)^{(\mathrm{LHCb,J/\psi\,\pi^+})} - m(B^+)^{(\mathrm{LHCb,J/\psi\,K^+})} = 994 \pm 1.3\,(\mathrm{stat}) \pm 0.6\,(\mathrm{syst})\ \ \mathrm{MeV}/c^2. \qquad \text{(II.7)}$$

To understand how to further reduce the uncertainty it could help to recall that the mass of a particle can be calculated given its four-momentum p simply as $mc = \sqrt{p^2}$. The four-momentum of the B_c^+ meson is reconstructed summing the four-momentum of its

daughter particles. In the rest frame of the B_c^+ meson, and assuming a two-body decay, this reads

$$p_{B_c^+} = \begin{pmatrix} cm_{B_c^+} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \frac{1}{c}\sqrt{m_1^2c^4 + \mathbf{p}_1^2c^2} \\ \mathbf{p}_1 \end{pmatrix} + \begin{pmatrix} \frac{1}{c}\sqrt{m_2^2c^4 + \mathbf{p}_2^2c^2} \\ \mathbf{p}_2 \end{pmatrix}. \tag{II.8}$$

If the B_c^+ mass measurement is performed using a channel where the contribution of m_1 and m_2 to p^2 is larger, and the contribution of $\mathbf{p_1}$ and $\mathbf{p_2}$ is smaller, the impact on the uncertainty on $\mathbf{p}$, notably the momentum scale, but also the statistical error, is reduced.

Defined the Q-value as the difference between the mass of the mother particle and the sum of the masses of the daughters, one concludes that decays with smaller Q-value are more suitable for mass measurements.

Mass measurement using the decay $B_c^+ \to J/\psi\, D_s^+$

The Q-value of the decay $B_c^+ \to J/\psi\, D_s^+$ is much smaller than it is for $B_c^+ \to J/\psi\, \pi^+$, so that the uncertainty due to the momentum scale calibration is reduced to 0.30 MeV/c^2.

The mass measurement obtained [65],

$$m(B_c^+)^{(\mathrm{LHCb,J/\psi\, D_s^+})} = (6.27628 \pm 0.00144 \pm 0.00036)\,\mathrm{GeV}/c^2, \tag{II.9}$$

is the world most precise measurement of the B_c^+ mass.

The mass distribution of the collected events is shown in Figure II.6.

Mass measurement using the decay $B_c^+ \to J/\psi\, p\bar{p}\pi^+$

Recently another low-Q decay has been observed and studied at LHCb. The decay $B_c^+ \to J/\psi\, p\bar{p}\pi^+$ is the first baryonic B_c^+ decay observed. It has been observed studying the combined LHCb datasets collected in 2011 and 2012 with a significance corresponding to 7.3 standard deviations.

Beside the reduced impact of the momentum scale uncertainty, the decay $B_c^+ \to J/\psi\, p\bar{p}\pi^+$ offers a final state composed of particles whose masses are determined with higher precision with respect to the $B_c^+ \to J/\psi\, D_s^+$ decay, reducing the related source of uncertainty on the B_c^+ mass to a negligible level.

The result obtained [66],

$$m(B_c^+)^{(\mathrm{LHCb,J/\psi\, p\bar{p}\pi^+})} = (6.2740 \pm 0.0018\,(\mathrm{stat}) \pm 0.0004\,(\mathrm{syst}))\,\mathrm{GeV}/c^2, \tag{II.10}$$

is almost as precise as the mass measurement obtained studying the final state $J/\psi\, D_s^+$ and constitutes an independent measurement of this important quantity.

The mass distribution of the collected events is shown in Figure II.7.

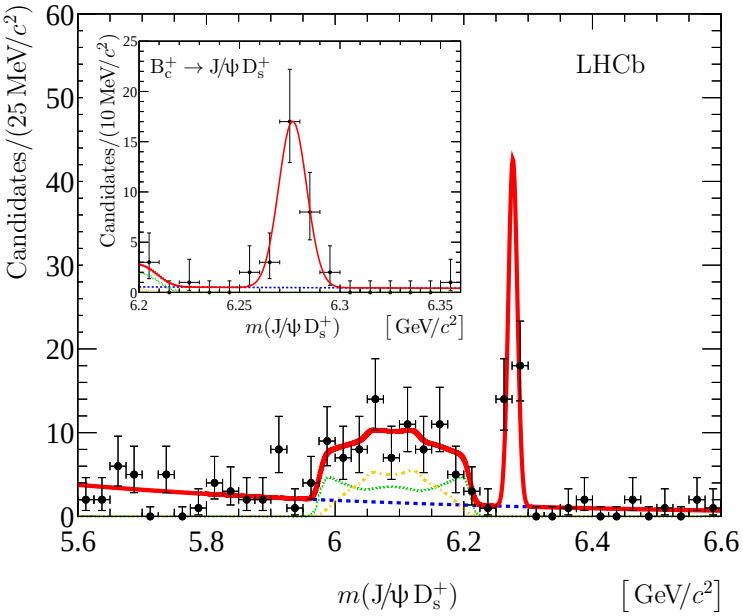

Figure II.6: Invariant mass distribution of the $B_c^+ \to J/\psi D_s^+$ candidates as collected by the LHCb Collaboration. The narrow peak is the signal, while the broad structure at lower mass is due to partially reconstructed $B_c^+ \to J/\psi D_s^{*+}$ decays. The inset shows the narrow peak using a finer binning. The Figure was published in Ref. [65].

II.2 Production measurements

Since no absolute branching fraction has never been measured for B_c^+ states, CMS and LHCb have used $B^+ \to J/\psi K^+$ as normalization channel, measuring the ratio

$$R_{c/u} = \frac{\sigma(B_c^+) \times \mathcal{B}(B_c^+ \to J/\psi \pi^+)}{\sigma(B^+) \times \mathcal{B}(B^+ \to J/\psi K^+)} = \frac{N(B_c^+ \to J/\psi \pi^+)}{\epsilon_{tot}^c} \frac{\epsilon_{tot}^u}{N(B^+ \to J/\psi K^+)} \tag{II.11}$$

where $N(B_c^+ \to J/\psi \pi^+)$ and $N(B^+ \to J/\psi K^+)$ represent the number of reconstructed B_c^+ and B^+ decays, respectively, and are corrected for the total selection efficiencies ϵ_{tot}^c and ϵ_{tot}^u, respectively. The choice of the normalization channel $B^+ \to J/\psi K^+$ makes many of the uncertainties cancel in the ratio.

Recently, experimental values for $R_{c/u}$ were measured by CMS and LHCb in two different kinematical regions using datasets corresponding to an integrated luminosity of 4.7 fb^{-1} and 0.37 fb^{-1} respectively (cf. Equation III.1).

$$R_{c/u}^{\text{CMS}} = \left(0.48 \pm 0.05(\text{stat}) \pm 0.04(\text{syst})^{+0.05}_{-0.03}(\tau_{B_c^+})\right) \times 10^{-2} \tag{II.12}$$

with $p_T(B_c^+) > 15$ GeV/c, $|y| < 1.6$, $\sqrt{s} = 7$ TeV [67].

$$R_{c/u}^{\text{LHCb}} = \left(0.68 \pm 0.10(\text{stat}) \pm 0.03(\text{syst}) \pm 0.05(\tau_{B_c^+})\right) \times 10^{-2} \tag{II.13}$$

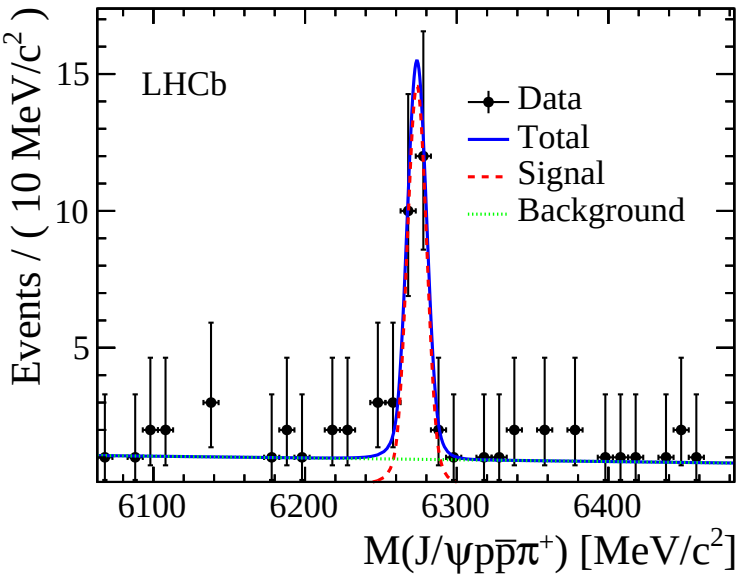

Figure II.7: Invariant mass distribution of the $B_c^+ \to J/\psi p\bar{p}\pi^+$ candidates collected by the LHCb Collaboration. The data model, fitted to the experimental data to measure the B_c^+ mass and its uncertainty, is superposed. The Figure was published in Ref. [66].

with $p_T(B_c^+) > 4$ GeV/c, $2.5 < \eta < 1.6$, $\sqrt{s} = 7$ TeV [64].

The third error is due to the uncertainty on the lifetime of the B_c^+ meson which reflects into an uncertainty on the selection efficiency of criteria correlated to the B_c^+ flight distance. Since the two measurements are performed in two different kinematical regions, they are not expected to be consistent. Indeed, the softer p_T distribution of the B_c^+ meson with respect to the B^+ implies a lower value of $R_{c/u}$ at higher p_T. The measurements are consistent with the expectations.

II.3 Observation of an excited B_c^+ meson state

Few months ago, at the International Conference on High Energy Physics (ICHEP) held in Valencia, the ATLAS Collaboration reported the first observation of a candidate for a B_c^{**+} state decaying to $B_c^+\pi^+\pi^-$. If confirmed, this would be the first observation of an excited state of the $(\bar{b}c)$ system [18].

The analysis was performed using the dataset collected with the ATLAS detector in 2011 and 2012 at $\sqrt{s} = 7$ and 8 TeV, respectively.

The mass distribution is interpreted as an exponential background superposed to a Gaussian *pdf* modelling the signal. The mean of the fitted Gaussian peak in consistent with the world average of the B_c^+ mass measurements [1], and the number of fitted $B_c^+ \to J/\psi\pi^+$ candidates is 100 ± 23 in 2011 data and 227 ± 25 in 2012 data.

The reconstructed B_c^+ candidates are then combined to same-charge pion pairs (B_c^+

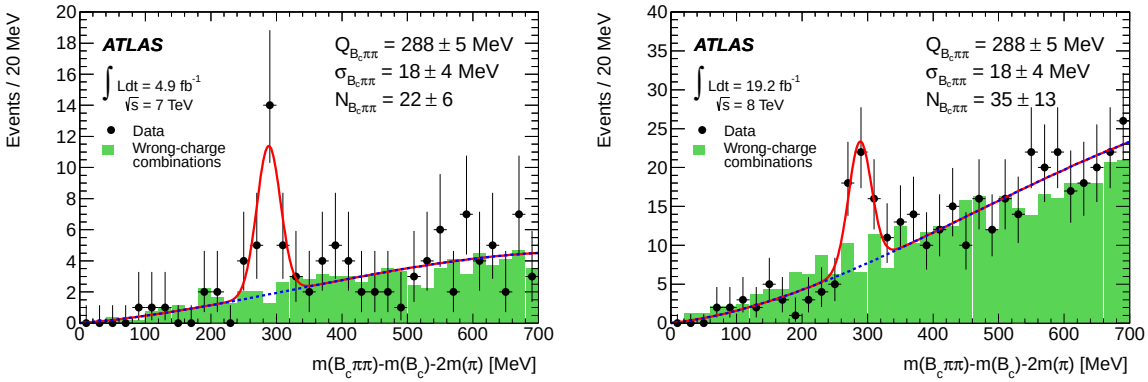

Figure II.8: The Q-value ($Q = c^2\,(m(B_c^+\pi^+\pi^-) - m(B_c^+) - 2m(\pi^\pm))$) distribution for right-charge combinations (shaded histogram) in 7 (left) and 8 (right) TeV data. The wrong-charge combinations are normalized to the same yield as the right-charge background. The solid line is the projection of the results of the unbinned maximum likelihood fit to all candidates in the range 0–700 MeV/c^2.

π^+ π^+, and B_c^+ π^- π^-) to model the expected combinatorial background to $B_c^{**+} \to B_c^+\pi^+\pi^-$ decays. The mass distribution of both $B_c^+\pi^+\pi^-$ combinations, said right-sign combinations since they agree with the expected signal signature, and $B_c^+\pi^\pm\pi^\pm$, said wrong-sign combination, are shown in Figure II.8.

The mass distributions are fitted with the sum of an exponential background and a Gaussian signal, to extract the signal yield and the mass of the new state. The fit yields 22 ± 6 signal events in the 2011 data and 35 ± 13 events in 2012 data, corresponding to a global significance of 5.2 standard deviations.

The Q-value obtained from the fit is

$$Q/c^2 = m(B_c^+\pi^+\pi^-) - m(B_c^+) - 2m(\pi^\pm) = 288.3 \pm 3.5 \pm 4.1\,\text{MeV}/c^2, \qquad \text{(II.14)}$$

corresponding to a mass of the excited B_c^+ state

$$m(B_c^{**+}) = 6.842 \pm 0.004 \pm 0.005\,\text{GeV}/c^2. \qquad \text{(II.15)}$$

The preliminary result for the ratio of the yields observed for B_c^{**+} and B_c^+ candidates,

$$\frac{N(B_c^{**+})}{N(B_c^+)}^{(\text{observed})} = (16 \pm 4)\%, \qquad \text{(II.16)}$$

is found to be consistent with the theoretical expectation [68]

$$\frac{N(B_c^{**+})}{N(B_c^+)}^{(\text{expected})} = \frac{\sigma(B_c^{**+})}{\sigma(B_c^+)} \times \varepsilon(\pi^\pm)^2 \times \mathcal{B}(B_c^{**+} \to B_c^+\pi^+\pi^-) \sim (2 \text{ — } 17)\%, \qquad \text{(II.17)}$$

where $\frac{\sigma(B_c^{**+})}{\sigma(B_c^+)} \sim 0.6$ is the ratio of the production cross-sections of the B_c^{**+} and B_c^+ states, $\varepsilon(\pi^\pm)^2 \sim 0.6$ is the reconstruction efficiency of two charged pions, and $\mathcal{B}(B_c^{**+} \to B_c^+\pi^+\pi^-)$

is the expected decay branching fraction obtained as the branching fraction of hadronic B_c^{**+} decays $\mathcal{B}(B_c^{**+} \to B_c^+\pi\pi) \sim 0.75$ — 0.95 corrected for the isospin factor 0.5 — 0.67.

While consistent to theoretical expectations, the result is very close to the upper bound of a prediction spanning an order of magnitude, enhancing the interest for this recent result.

II.4 B_c^+ decays

Many new decays have been observed by the LHCb experiment, allowing to compare their relative branching fraction to theoretical expectations, therefore constraining the models. However the most important information would come from the relative decay width of transitions $\bar{b} \to \bar{c}(\bar{u})$ to $c \to s(d)$. This section reviews the first step in this direction, describing the observation of a B_c^+ decaying through the transition $c \to s$ achieved by the LHCb Collaboration which reported the observation of the $B_c^+ \to B_s^0\pi^+$ decay, then briefly lists the new decay modes recently observed by the LHCb Collaboration.

II.4.1 Observation of $B_c^+ \to B_s^0\pi^+$ decays

Analysing the pp-collision data collected in 2011 at a center-of-mass energy of 7 TeV, and in 2012 at a center-of-mass energy of 8 TeV, LHCb has performed a search starting from two samples of fully reconstructed B_s^0 mesons decaying to $B_s^0 \to D_s^-\pi^+$ and $B_s^0 \to J/\psi\,\phi$.

$73\,700 \pm 500$ $B_s^0 \to D_s^-\pi^+$ and $103\,760 \pm 380$ $B_s^0 \to J/\psi\,\phi$ candidates are observed and combined to a charged pion to create B_c^+ candidates. The distributions of the invariant mass of such candidates are shown in Figure II.9.

The fitted signal yield for $B_c^+ \to B_s^0(\to D_s^-\pi^+)\pi^+$ decays is 64 ± 10 corresponding to a statistical significance of 7.7σ; for $B_c^+ \to B_s^0(\to J/\psi\,\phi)\pi^+$, 35 ± 8 signal candidates are observed, corresponding to a statistical significance of 6.1σ.

The B_s^0 and B_c^+ yields are corrected for the relative detection efficiencies, to obtain the efficiency-corrected ratios of $B_c^+ \to B_s^0\pi^+$ over B_s^0 yields,

$$\left(2.54 \pm 0.40(\text{stat})\ {}^{+0.23}_{-0.17}(\text{syst})\right) \times 10^{-3}, \quad \text{and} \quad (2.20 \pm 0.49(\text{stat}) \pm 0.23(\text{syst})) \times 10^{-3}. \tag{II.18}$$

for B_s^0 reconstructed as $D_s^-\pi^+$ and $J/\psi\,\phi$ respectively. The systematic uncertainty is dominated by the uncertainty on the lifetime of the B_c^+ meson which results into an uncertainty on the selection efficiency of criteria based on the B_c^+ flight distance. The correlation of such contribution between the two B_s^0 reconstruction channels, is taken into account when combining the results above to give the ratio of production rates multiplied by the branching fraction

$$\frac{\sigma(B_c^+)}{\sigma(B_s^0)} \times \mathcal{B}(B_c^+ \to B_s^0\pi^+) = \left(2.37 \pm 0.31(\text{stat}) \pm 0.11(\text{syst})^{+0.17}_{-0.13}(\tau_{B_c^+})\right). \tag{II.19}$$

Assuming a value for $\sigma(B_c^+)/\sigma(B_s^0)$ of 0.2 [69], one would obtain a branching ratio

$\mathcal{B}(B_c^+ \to B_s^0\pi^+)$ of about 10%, the highest branching fraction ever observed for a b-hadron weak decay.

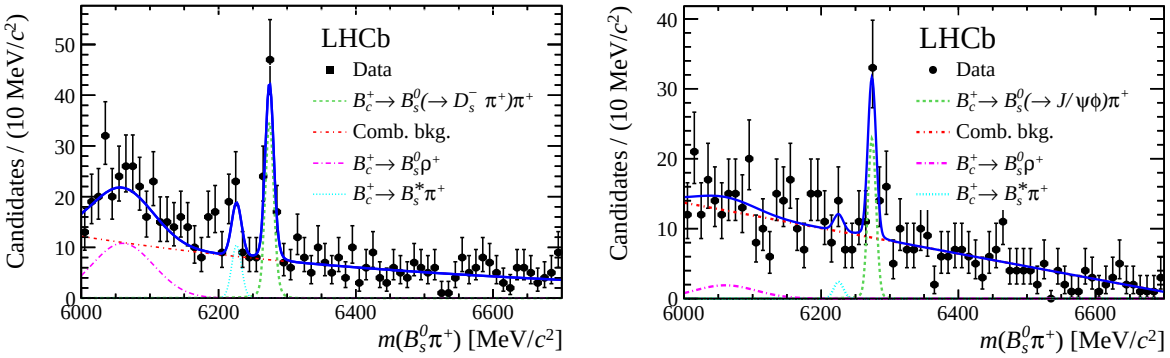

Figure II.9: Invariant mass for the $B_s^0\pi^+$ combinations with the B_s^0 reconstructed as $B_s^0 \to D_s^-\pi^+$ (left) and $B_s^0 \to J/\psi\,\phi$ (right). The plots were published in Ref. [69].

II.4.2 Decays to J/ψ and hadrons

Most of the observed B_c^+ decays have a charmonium resonance in the final state, usually a J/ψ, and a number of light hadrons. Theoretically these decays are well described with a factorization approach treating separately the decay of the B_c^+ meson to a J/ψ and a virtual W^+ boson, followed by the decay of the latter to hadrons. The decay $B_c^+ \to J/\psi W^{*+}$ is described using the semileptonic decay form factors as described in Section I.5.3, while the description of the virtual $W^\pm$ boson benefits of the experimental information from $\tau^- \to \nu_\tau W^{*-}$, where W^{*-} decays to hadrons. Within this theoretical framework, expectations for the branching fraction of the B_c^+ decays to J/ψ +hadrons relative to $B_c^+ \to J/\psi\,\pi^+$ have been formulated and found consistent with experiment.

Recent results and their comparison with the theoretical expectation are reported in Table II.1.

II.4.3 Decays to higher charmonia states: $B_c^+ \to \psi(2S)\pi^+$

Using the dataset collected in 2011, LHCb observed the decay $B_c^+ \to \psi(2S)\pi^+$ with a statistical significance of 5.2 σ. The branching fraction of the decay to $\psi(2S)\pi^+$ relative to the decay $B_c^+ \to J/\psi\,\pi^+$ was measured to be

$$\frac{\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\,\pi)} = 0.250 \pm 0.068\,(\text{stat}) \pm 0.014\,(\text{syst}) \pm 0.006(\mathcal{B}), \qquad \text{(II.20)}$$

where the last uncertainty is inherited by the branching fraction of the decay $\psi(2S) \to \mu^+\mu^-$ relative to the decay $J/\psi \to \mu^+\mu^-$.

Final state	*Ref.*		$\Gamma_i/\Gamma_{B_c^+\to J/\psi\pi^+}$ *Experimental*		*Theoretical*	
$J/\psi\pi^+\pi^-\pi^+$	LHCb 2012	[70]	$2.41 \pm 0.30\,(\text{stat})$	$\pm 0.33\,(\text{syst})$	2.3	[71]
$J/\psi\pi^+\pi^-\pi^+$	CMS 2013	[67]	$2.55 \pm 0.80\,(\text{stat})$	$\pm 0.33\,(\text{syst})\ ^{0.04}_{-0.01}(\tau_{B_c^+})$	2.3	[71]
$J/\psi 3\pi^+2\pi^-$	LHCb 2014	[72]	$1.74 \pm 0.44\,(\text{stat})$	$\pm 0.24\,(\text{syst})$	0.95 — 1.1	[73]
$J/\psi K^+K^-\pi^+$	LHCb 2013	[74]	$0.53 \pm 0.10\,(\text{stat})$	$\pm 0.05\,(\text{syst})$	0.47 — 0.49	[75]
$J/\psi K^+$	LHCb 2013	[76]	$0.069 \pm 0.019\,(\text{stat})$	$\pm 0.05\,(\text{syst})$	0.065 — 0.077	[76]
$J/\psi p\bar{p}\pi^+$	LHCb 2014	[66]	$0.143\ ^{+0.039}_{-0.034}\,(\text{stat})$	$\pm 0.013\,(\text{syst})$	0.17 ± 0.2	[66]

Table II.1: Recently measured branching fractions of B_c^+ decays to J/ψ and hadrons relative to $B_c^+ \to J/\psi\pi^+$ decays.

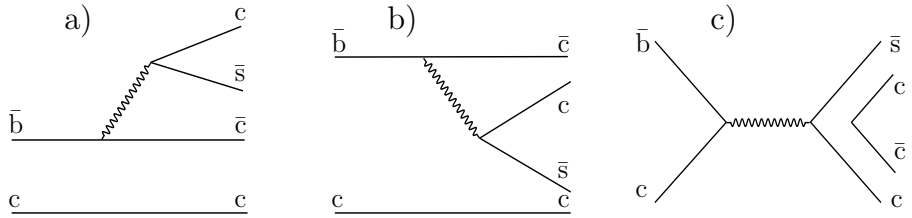

Figure II.10: Feynman diagrams for $B_c^+ \to J/\psi D_s^+$ through the c-spectator, colour suppressed, and weak annihilation processes. The drawing was published in Ref. [65].

The result is consistent with the theoretical expectations obtained in the framework of the relativistic quark model [32],

$$\frac{\mathcal{B}(B_c^+ \to \psi(2S)\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\pi^+)} \sim 0.18. \tag{II.21}$$

II.4.4 Decay $B_c^+ \to J/\psi D_s^+$

The decay mode $B_c^+ \to J/\psi D_s^+$, already mentioned in Section II.1 for the mass measurement, was observed by the LHCb Collaboration analyzing the combined dataset collected in 2011 and 2012 [65]. This decay is particularly interesting because it is led by three concurrent coherent processes as shown in Figure II.10.

The c-spectator diagram, with the $\bar{b}$ quark decaying to a $\bar{c}$ quark with the emission of a virtual W^{*+} boson decaying to a $c\bar{s}$ pair combining into a D_s^+ meson, is shown in Figure II.10a.

If, as expected, the c-spectator diagram dominates the decay, then the factorization principle would motivate the following relation

$$\frac{\Gamma(B_c^+ \to J/\psi D_s^+)}{\Gamma(B_c^+ \to J/\psi\pi^+)} \approx \frac{\Gamma(B \to \overline{D^*} D_s^+)}{\Gamma(B \to \overline{D^*}\pi)}, \tag{II.22}$$

where $\mathcal{O}(0.5\%)$ corrections are expected because of the different phase space.

Experimentally [65],

$$\frac{\mathcal{B}(B_c^+ \to J/\psi\, D_s^+)}{\mathcal{B}(B_c^+ \to J/\psi\, \pi^+)} = 2.37 \pm 0.56\,(\text{stat}) \pm 0.10\,(\text{syst}), \tag{II.23}$$

which is consistent with the factorization expectation when the ratio $\Gamma(B \to \overline{D^*} D_s^+)/\Gamma(B \to \overline{D^*}\pi)$ is evaluated using B^0 decays, and slightly larger when B^+ meson decays are used.

II.4.5 Semileptonic B_c^+ decays

With respect to the hadronic decay $B_c^+ \to J/\psi\,\pi^+$, the decay $B_c^+ \to J/\psi\,\mu^+\nu$ is favoured because of the form-factors of the decay $B_c^+ \to J/\psi\, W^*$ described in Section I.5.3. As already mentioned, the dependence of the decay width on q^2, momentum transferred by the W^+ boson, suppresses the decays with low q^2. In the two-body decay $B_c^+ \to J/\psi\,\pi^+$, the q^2 is naively constrained to the squared mass of the pion, which is low and therefore it is suppressed. Besides, the angular momentum of the virtual W^+ boson is different from the spin of the pion, while it is the same of $\mu^+\nu_\mu$ pair, contributing to suppress the former with respect to the latter.

The first measurement relating semileptonic and hadronic decay rates of the B_c^+ meson is performed using proton-proton collision data collected with the LHCb detector in 2011.

The result [77],

$$\frac{\mathcal{B}(B_c^+ \to J/\psi\,\pi^+)}{\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu)} = 0.0469 \pm 0.0028\,(\text{stat}) \pm 0.0046\,(\text{syst}), \tag{II.24}$$

is dominated by the systematic uncertainty on the B_c^+ decay model. Indeed, in order to remove the background due to b-hadrons decaying to J/ψ and hadrons, misidentified as $J/\psi\,\mu^+$ candidates, the analysis was performed on $J/\psi\,\mu^+$ candidates with invariant mass larger than $5.3\,\text{GeV}/c^2$. The efficiency of this selection requirement depends strongly on the shape of the $m(J/\psi\,\mu^+)$ distribution, described using decay form-factor models. Different models lead to different values for the efficiency, resulting in different estimations for the relative branching fraction.

Figure II.11 shows the simulation for signal and background samples used to determine the threshold on the mass of the $J/\psi\,\mu^+$ combination, and the invariant mass distribution of real-data combinations superposed to the data model.

II.5 B_c^+ lifetime measurements, state-of-the-art

After the lifetime measurement obtained contextually to the observation of the B_c^+ meson and described previously, several other lifetime measurements have been obtained using either the semileptonic $B_c^+ \to J/\psi\,\ell^+\nu$ decay channel or the fully reconstructed $B_c^+ \to J/\psi\,\pi^+$ mode. In this section the lifetime measurements published before the work presented in this Thesis are discussed.

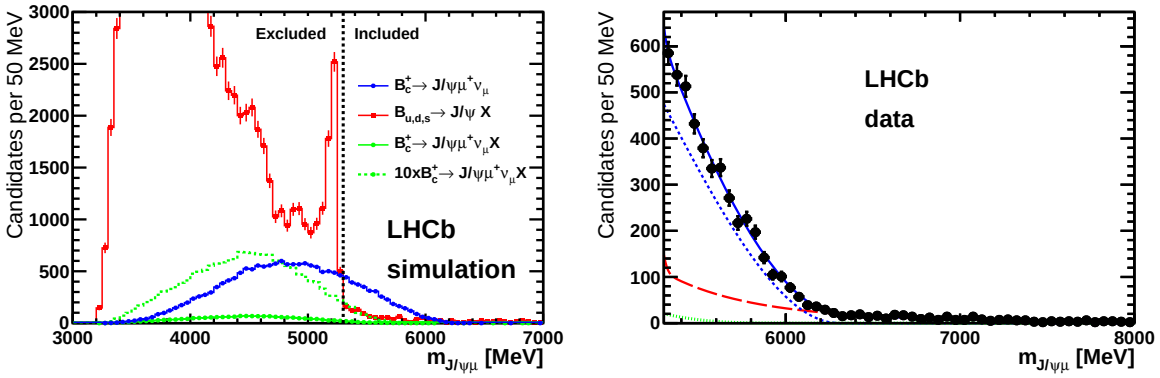

Figure II.11: Distribution of the invariant mass of the combinations $J/\psi\,\mu^+$ obtained through full simulation (left) and in real data (right). The color code is as follows: blue represents the signal components due to $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decays, the contribution due to B_c^+ decays including a J/ψ and a muon in the final state accompanied by other particles (for example $B_c^+ \to (J/\psi\,\gamma)_{\psi(2S)}\mu^+\nu$ decays) is shown in green; finally, background due to misidentification of $H_b \to J/\psi$ + hadrons decays to $J/\psi\,\mu^+$ combinations is represented in red. The large red peak in the left plot is due to $B^+ \to J/\psi\,K^+$ decays and its position is slightly shifted from the B^+ mass due to the incorrect mass assumed for the K^+ particle misidentified as a μ^+. The Figures were published in Ref. [77].

II.5.1 B_c^+ lifetime measurement using the channel $B_c^+ \to J/\psi\,\pi^+$

The only published result of a lifetime measurement using fully reconstructed $B_c^+ \to J/\psi\,\pi^+$ decays has been obtained by the CDF Collaboration in 2013 [78].

The selection of the B_c^+ candidates includes requirements on a projection of the distance between the daughter tracks and the Primary Vertex (Impact Parameter) which are extremely powerful in rejecting combinatorial background, but introduce a dependence of the selection efficiency on the decay time, corrected using simulation. The abundant $B^+ \to J/\psi\,K^+$ decay is used as control channel to ensure that simulation reproduces correctly this dependence. The efficiency for $B_c^+ \to J/\psi\,\pi^+$ and $B_c^+ \to J/\psi\,K^+$ decays is shown in Figure II.12.

The measurement was obtained using the invariant mass distribution of the B_c^+ candidates, shown in Figure II.13, to separate the signal and background components through a simultaneous fit to the mass and decay-length distributions, shown in Figure II.14, where the decay-length is defined as

$$ct \equiv \mathbf{L}_{xy} \cdot \mathbf{p}_T(B_c^+) \frac{cm(B_c^+)}{|p_T(B_c^+)|^2}. \tag{II.25}$$

The background yield is estimated using the mass sidebands, the regions in the mass ranges from 6.16 to 6.21 GeV/c^2, and from 6.33 to 6.60 GeV/c^2 where candidates constitute a pure background sample. The region below 6.16 GeV/c^2 is not used to avoid potential contamination from partially reconstructed decays.

The lifetime obtained from the simultaneous likelihood fit to the selected 272 ± 61 (stat)

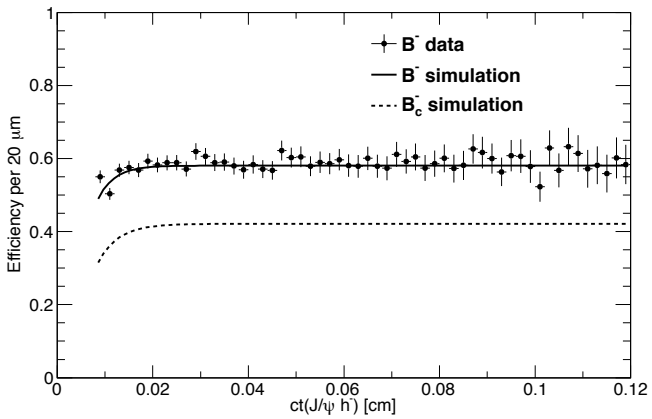

Figure II.12: Dependence of the efficiency on the reconstructed decay time. The quick rise at low-ct is due to the selection criteria based on the distance between the daughter tracks and the beam axis. The Figure was published in Ref. [78].

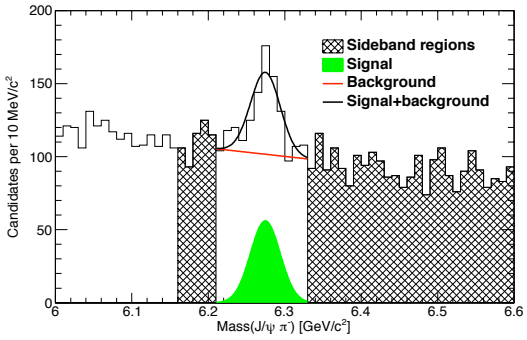

Figure II.13: Invariant-mass distribution of the $J/\psi\ \pi^+$ candidates. The hatched areas are the sideband regions and the signal region lies between them. The fit result is overlaid in the signal region, as well as the signal and the background template distributions. The figure was published in Ref. [78].

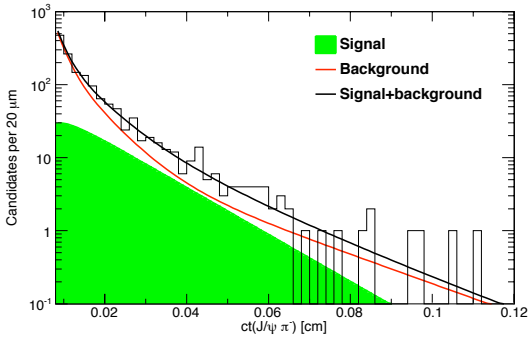

Figure II.14: Decay-length distribution of the $J/\psi\pi^+$ candidates. The projection of the simultaneous fit, together with individual contributions from signal and background is shown. The figure was published in Ref. [78].

signal candidates is

$$\tau_{B_c^+}^{(\mathrm{CDF,J/\psi\,\pi})} = 463^{+73}_{-65}\,(\mathrm{stat}) \pm 36\,(\mathrm{syst})\ \mathrm{fs}, \tag{II.26}$$

where the systematic uncertainty is dominated by the background model, which is constrained using data and therefore limited by the small number of events selected.

Detailed documentation on this analysis is available through the Ph.D. Thesis in Ref. [79].

Another preliminary result, obtained by the CMS Collaboration with a handful of candidates, has been published in a Master Thesis defended at the University of Milano. The result is consistent with the CDF measurement [80],

$$\tau_{B_c^+}^{(\mathrm{CMS,J}/\psi\,\pi)} = 453 \pm 42 \text{ fs} \tag{II.27}$$

This preliminary result is not considered further.

II.5.2 B_c^+ lifetime measurements using the channel $B_c^+ \to J/\psi\,\ell^+\nu$

Both CDF and D0 analyzed the dataset collected in the second run of the Tevatron to measure the lifetime of the B_c^+ meson.

The D0 Collaboration measured the lifetime using the decays $B_c^+ \to J/\psi\,\mu^+\nu$. To measure the lifetime of the B_c^+ meson, the D0 collaboration introduces the *decay length* L related to the decay time t^* of the B_c^+ meson by the equation

$$L = ct^*\beta\gamma = t^*\frac{p(B_c^+)}{m(B_c^+)}, \tag{II.28}$$

its transverse projection is defined as

$$L_{xy} = ct^*(\beta\gamma)_{xy} = t^*\frac{p_T(B_c^+)}{m(B_c^+)}. \tag{II.29}$$

Unfortunately, the proper decay length is not a quantity experimentally accessible for semileptonic decays because since the B_c^+ meson is not fully reconstructed, its (transverse) momentum is not known. The fraction of the transverse momentum of the B_c^+ meson which is reconstructed because taken by the $J/\psi\,\ell^+$ combination is named k-factor. Namely,

$$k^{(\mathrm{D0})} = \frac{p_T(J/\psi\,\mu^+)}{p_T(B_c^+)}. \tag{II.30}$$

The k-factor distribution is studied in six bins of $m(J/\psi\,\mu)$.

The skewed smearing of the signal template distribution for the decay length in the rest frame of the $J/\psi\,\mu^+$ combination (named *pseudo-proper decay time*) is named *k-factor method*, and, in one form or in another, is common to most of the studies on semileptonic decays, and more in general to studies involving partial reconstructions.

Figure II.15 shows the projection of the data model on the invariant mass of the combination $J/\psi\,\mu^+$ where the combinatorial background is included as two separate categories: the background due to three tracks from the primary vertex ($p\bar{p}$ collision); and the background due to real J/ψ produced in a b-hadron decay, attached to a particle misidentified as a muon.

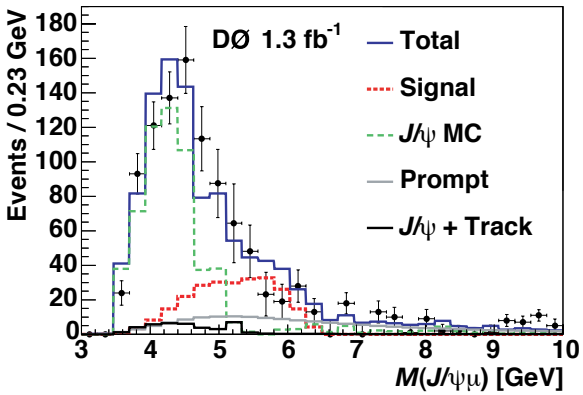

Figure II.15: Fit to the mass of the J/ψ μ^+ vertex as collected by the D0 experiment, with the J/ψ mass sidebands and the B^+ component subtracted and decay-length significance $L_{xy}/\sigma(L_{xy}) > 4$ required. The figure was published in Ref. [81].

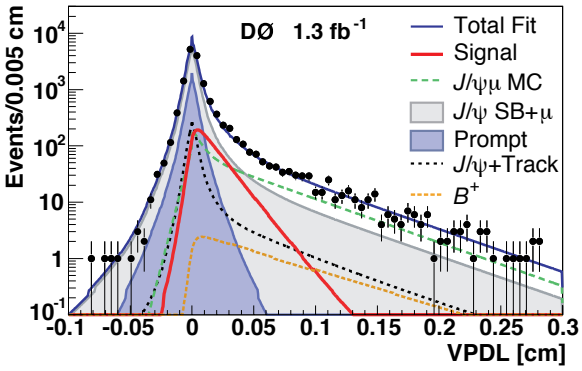

Figure II.16: Distribution of the Visible-Proper Decay Length (VPDL) of the data collected by the D0 experiment. The figure was published in Ref. [81].

Figure II.16 shows the distribution of the so called visible proper decay length (VPDL)

$$\mathrm{VPDL} = L_{xy}\frac{m(B_c^+)}{p_T(J/\psi\,\mu^+)}ck, \tag{II.31}$$

providing the estimation of the proper decay length ct^*. The data model is fitted to the $881 \pm 80\,(\mathrm{stat})$ signal candidates to measure the B_c^+ lifetime [81]

$$\tau_{B_c^+}^{(\mathrm{D0,J}/\psi\,\mu^+)} = 448^{+38}_{-36}\,(\mathrm{stat}) \pm 0.032\,(\mathrm{syst})\ \mathrm{ps}, \tag{II.32}$$

where the systematic uncertainty is dominated by the background model obtained from a combined use of simulated and real data.

Detailed documentation on this analysis is available from the Ph.D. Thesis in Ref. [82].

The latest result by CDF [83] has not been published, but supersedes a previous result [84], extending the statistics, and including the decay channel $J/\psi\,\mu^+\nu_\mu$ beside the channel $J/\psi\,e^+\nu_e$ used in Ref. [84].

The analysis approach of the measurement obtained by CDF is very similar to the one discussed for D0. An interesting difference, worth few lines of description, is the technique used to model the background due to the association of a random track, misidentified as a lepton, to a J/ψ meson, common decay product of a b-hadron. To model this background, any lepton requirement in the selection of the J/ψ + *track* sample is removed. Each of the hadrons in this sample is a candidate to fake a lepton. Estimating the probability that each of these hadrons is a particular particle (among pion, kaon, and proton), F_π, F_K, or F_p, and the probability for a given particle type to fake a lepton, P_π, P_K, and P_p, it is possible to determine a total fake lepton probability for each event and to use it to weight

the events in the J/ψ + *track* sample. The weighting for a given event can be written as

$$W = F_\pi P_\pi + F_K P_K + F_p P_p. \tag{II.33}$$

Once the weight is estimated using control-samples, the background of fake lepton events in the final state can be determined. The particle type composition of the third tracks in the J/ψ + *track* sample is estimated using the $\mathrm{d}E/\mathrm{d}x$ and time of flight (TOF) information. The probability for hadrons to fake a lepton can be estimated using samples in which the hadrons are identified as decay products in relatively pure and fully reconstructed systems. For kaons and pions, the decay $D^{*+} \to D^0\pi^+ \to \pi^+K^-\pi^+$ is used. The sign of the pion from the D^{*+} meson decay identifies the pion and the kaon in the D^0 meson decay. The $\Lambda^0 \to p\pi^-$ decay is used to select a pure sample of protons thanks to the large mass difference between the proton and pion masses.

The template *pdf* obtained weighting each event in the J/ψ + *track* sample is used to subtract this background from the signal sample when determining the B_c^+ meson lifetime.

The results obtained with the $J/\psi\ \mu^+$ and $J/\psi\ e^+$ samples have been used separately and the value obtained for the lifetime is finally combined. CDF determines the lifetime using the measured value $\frac{p_T(J/\psi\,\ell)}{m(J/\psi\,\mu^+)c}$ as an approximation for the boost factor $(\beta\gamma)_{xy} = \frac{p_T(B_c^+)}{m(B_c^+)c}$, resulting in the pseudo-proper decay-length

$$ct_{\mathrm{ps}}^{(\mathrm{CDF})} = L_{xy}\frac{m(J/\psi\,\ell)c}{p_T(J/\psi\,\ell)}. \tag{II.34}$$

The signal template distribution for this quantity is obtained from an exponential distribution convoluted with the appropriate k-factor distribution obtained from simulation.

The result obtained for the combination $J/\psi\ \mu^+$ is

$$\tau_{B_c^+}^{(\mathrm{CDF},J/\psi\,\mu^+)} = 599^{+109}_{-\ 90}\,(\mathrm{stat})\ \mathrm{fs}, \tag{II.35}$$

and the distribution of the pseudo-proper decay length is shown in Figure II.17.

The result obtained for the combination $J/\psi\ e^+$ is

$$\tau_{B_c^+}^{(\mathrm{CDF},J/\psi\,e^+)} = 405^{+52}_{-49}\,(\mathrm{stat})\ \mathrm{fs}, \tag{II.36}$$

and the distribution of the pseudo-proper decay length is shown in Figure II.18.

The combined result is

$$\tau_{B_c^+}^{(\mathrm{CDF})} = 475^{+53}_{-49}\,(\mathrm{stat}) \pm 18\,(\mathrm{syst})\ \mathrm{fs}, \tag{II.37}$$

where the systematic uncertainty is dominated by the background model.

II.5.3 Summary of the B_c^+ lifetime measurements

The summary of the B_c^+ lifetime measurement is shown in Table II.2. Figure II.19 represents the same values with combined systematic and statistical uncertainties.

The world average, using only peer-reviewed results was [1]

$$\tau_{B_c^+}^{(\mathrm{PDG})} = (0.452 \pm 0.033) \times 10^{-12}\ \mathrm{s}. \tag{II.38}$$

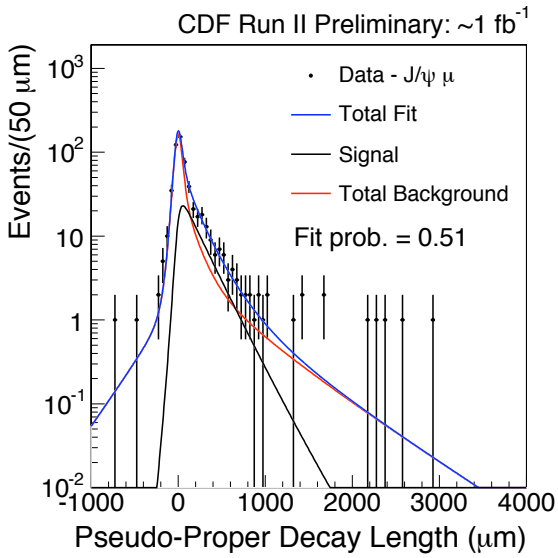

Figure II.17: Fitted *pseudo-proper decay lenght* distribution for the J/ψ μ^+ candidates collected by the CDF experiment. Background distributions are shown combined in a single component. The figure was made public in Ref. [83].

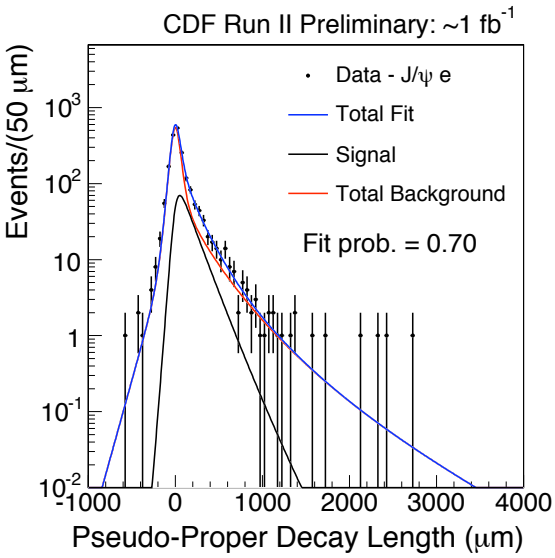

Figure II.18: Fitted *pseudo-proper decay lenght* distribution for the J/ψ e^+ candidates collected by the CDF experiment. Background distributions are shown combined in a single component. The figure was made public in Ref. [83].

Experiment	*Decay channel*	*Ref.*	*Lifetime [ps]*
CDF	$B_c^+ \to J/\psi\, \ell^+\nu$	[61]	$0.46^{+0.18}_{-0.16}$ (stat)$\pm$0.03 (syst)
D0	$B_c^+ \to J/\psi\, \mu^+\nu$	[81]	$0.448^{+0.038}_{-0.036}$ (stat)$\pm$0.032 (syst)
CDF	$B_c^+ \to J/\psi\, e^+\nu$	[84]	$0.463^{+0.073}_{-0.065}$ (stat)$\pm$0.036 (syst)
CDF	$B_c^+ \to J/\psi\, \pi^+$	[78]	0.452$\pm$0.048 (stat)$\pm$0.027 (syst)
CDF	$B_c^+ \to J/\psi\, \ell^+\nu$	[83]	$0.475^{+0.053}_{-0.049}$ (stat)$\pm$0.018 (syst)

Table II.2: Summary of the B_c^+ lifetime measurements prior to the work presented in this Thesis.

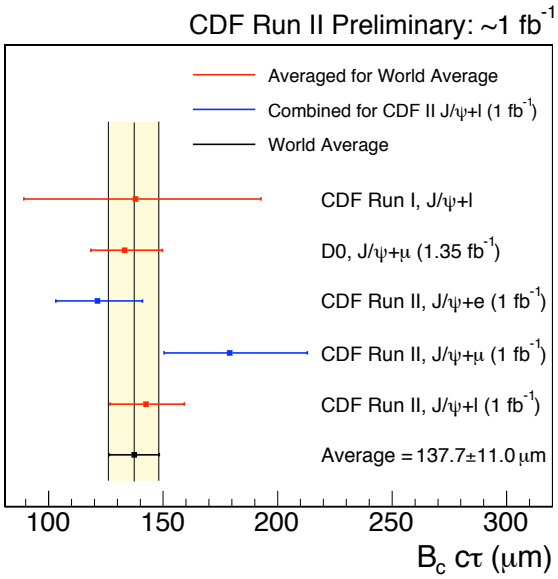

Figure II.19: Summary of the B_c^+ lifetime measurements prior to the work presented in this Thesis. The results are given here for the average proper decay length The Figure was made public in Ref. [83].

Chapter III Introduction

The Large Hadron Collider and the LHCb experiment

The Large Hadron Collider is a long-term project discussed for the first time in a joint CERN–ECFA Workshop meeting in 1984, before starting the works for the tunnel of LEP. The main purpose was to use the same tunnel as LEP to build a Large Hadron Collider (LHC) to *investigate hadron collisions* in the energy range between 10 and 20 TeV in the center of mass.

Approved by the CERN Council in 1994, the LHC delivered the first proton-proton collision in 2009, while four large experiments were recording that unprecedented event: ATLAS, CMS, ALICE, and LHCb.

Three years of data-taking followed that first bunch, alternating pp collisions to p-Pb and Pb-Pb collisions.

On July the 4th, 2012, the ATLAS and CMS Collaborations announced the discovery of the Higgs boson, probably the most glittering result achieved so far at the LHC.

Part of that investigation is performed with the LHCb detector, a forward spectrometer focusing on rare and CP violating processes in the b-quark sector using advanced vertexing, trigger and particle identification techniques.

The excellent performance of the LHC and of the LHCb detector has allowed several important physics achievements to be accomplished.

III.1 The Large Hadron Collider

The *Large Hadron Collider* (LHC) is a proton-proton and heavy ion (lead) collider at CERN. It is located at the French-Swiss border, in a 27 km long tunnel which contains two beam pipes for the two particle beams accelerated in opposite directions. The choice of using proton-proton collisions allows the LHC to reach high luminosity avoiding the production of anti-protons to be used in the collisions, major limit to the luminosity at previous high-energy hadronic accelerators as the Tevatron. During the first run, LHC has been delivering an average instantaneous luminosity to $2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$, with a record of $3.9 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ reached on April, 19th 2012. The instantaneous luminosity is maximal

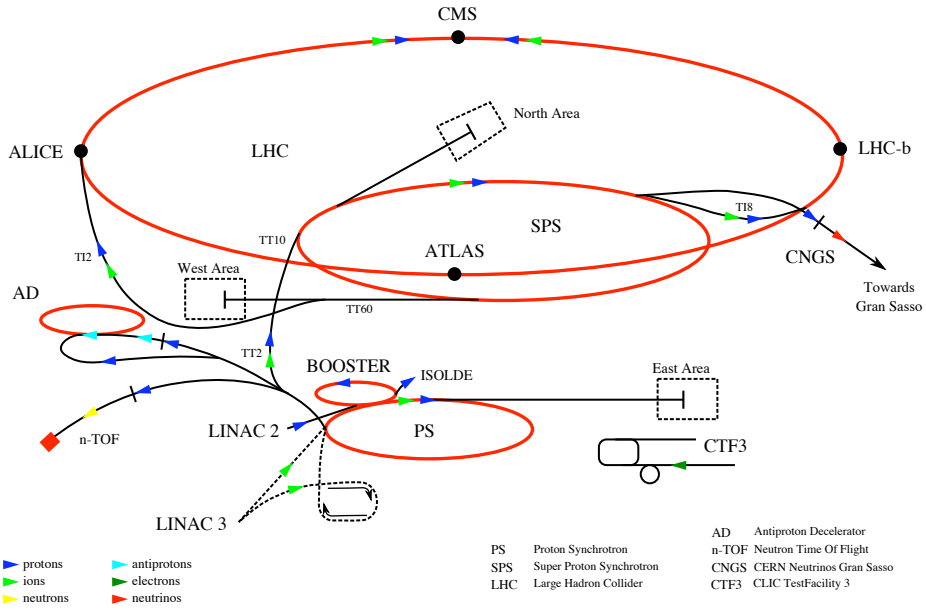

Figure III.1: Schematic representation of the CERN accelerator system. Picture by Forthommel, CC-BY-SA 3.0.

at the beginning of the *fill*, and then decreases because of several factors including beam warming and proton losses. The LHCb experiment uses a technique named *luminosity leveling* to run at lower but constant instantaneous luminosity as described in Section III.2.2. It is the originality of LHC superconducting bending magnets to have the two beams bent and circulating in opposite directions in the same structure. The sizable bending power of the magnets will allow LHC to reach 7 TeV per beam, and a $\sqrt{s}$ value of 14 TeV.

III.1.1 The LHC accelerator system

The LHC is served by other smaller and less powerful particle accelerators which gradually accelerate protons up to 450 GeV before transferring them to the LHC storage rings, where they are further accelerated for about 20 minutes before reaching the operational energy. In Figure III.1, a schematic representation of the accelerators complex of CERN is shown.

For the first acceleration stage a linear accelerator, called LINAC2, is used. The second stage is provided by the Proton Synchrotron Booster which injects protons in the Proton Synchrotron (PS). The PS accelerates protons up to an energy of 25 GeV. The Super Proton Synchrotron, made famous by the discovery of $W^{\pm}$ and Z^0 bosons, brings the energy up to 450 GeV, protons can then be extracted for fixed target experiments or to be injected in the LHC storage rings.

III.1.2 The large experiments at the LHC

There are four large experiments exploiting the LHC proton-proton collisions. CMS (*Compact Muon Solenoid*) is a general purpose experiment whose detector is composed of various concentric sub-detectors around the interaction point. The detector is subdivided in three parts, two end-caps and a barrel-shaped part. It contains a superconducting solenoidal magnet generating a magnetic field of 3.8 T, which allows a very compact structure and motivates the experiment name. The whole detector is 12 500 tons heavy. Some of the main aims of CMS are the Higgs boson studies, searches for particles not expected in the Standard Model, and precision measurements of *top* and *bottom* physics.

Aims shared by the other LHC general purpose detector: ATLAS (*A Toroidal LHC ApparatuS*). It is 46 meters long with a diameter of 25 meters and a weight of 7 000 tons. It is the largest collider experiment ever built and the ATLAS collaboration is the largest scientific collaboration in the world.

ALICE (A Large Ion Collider Experiment) was designed with the main goal of exploiting heavy ion (Pb-Pb) collisions at LHC. ALICE aims to study nuclear matter at high temperature and pressure, and investigate the quark-gluon plasma, a phase where quarks are expected not to be confined inside hadrons and QCD can be described as a perturbative theory.

Finally LHCb, described in some detail in the next section, is the LHC experiment devoted to flavour physics, aiming at precision measurements and indirect searches for physics beyond the Standard Model. The LHCb experiment is a forward spectrometer with excellent vertex and mass resolution, good charged particle identification and a versatile trigger scheme. Its design is oriented to fast reconstruction of b- and c-hadron decay candidates in the earlier stages of data selection.

III.1.3 The $b\bar{b}$ production cross section

The $b\bar{b}$ production cross section depends on the pp-collision energy increasing significantly with it, faster than the total cross-section.

The cross-sections of various processes as functions of the center-of-mass energy $\sqrt{s}$ are presented in Figure III.2 for a comparison between Tevatron and the LHC. As usual, cross-sections are expressed in barn sub-multiples. A barn b equals 10^{-28} m^2, sub-multiples from millibarn (mb) to attobarn (ab $= 10^{-18}$ b) are often used.

The $b\bar{b}$ production cross section at Tevatron ($\sqrt{s} = 1.96$ TeV) in the geometrical acceptance of CDF has been estimated to be $\sim 10\mu$b, the $b\bar{b}$ production cross section at LHC has been estimated to be $\sim 290\mu$b, corresponding to a cross-section of 75μb in the geometrical acceptance of LHCb, for a center-of-mass energy of 7 TeV. At the nominal value of 14 TeV, the cross-section at LHC is expected to exceed 500μb.

The angular distribution of the $b\bar{b}$ production is peaked in a small region, in forward and backward directions. The high correlation between the flight direction of the two hadrons is not surprising if one considers that the $b\bar{b}$ pairs are mainly generated in processes involving two interacting gluons with a high energy and a high momentum. Since the two virtual

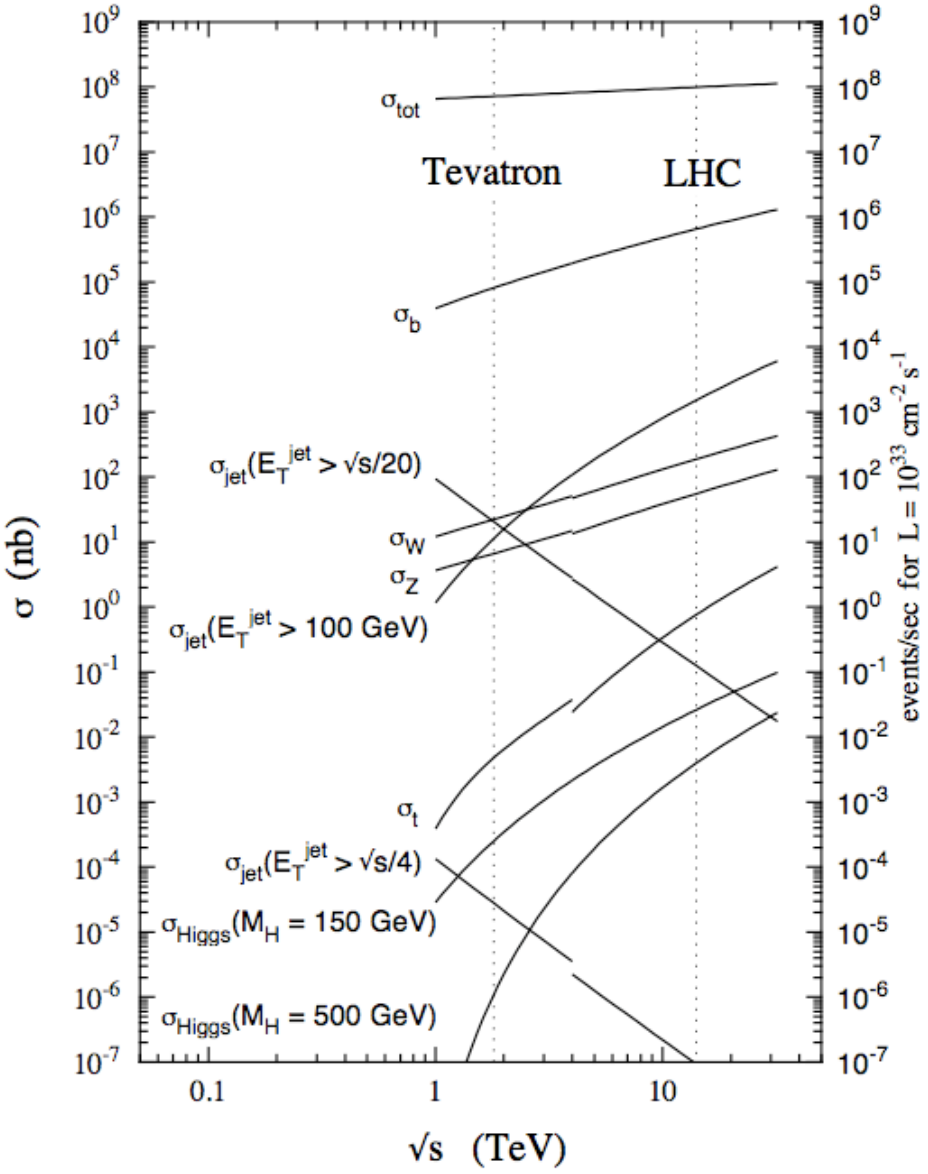

Figure III.2: Cross sections as a function of the energy $\sqrt{s}$ in the center of mass reference system. For lower energies $p\bar{p}$ collision cross section are represented. For higher energies (> 4 TeV) the pp collision cross sections are plotted. The vertical dotted lines indicate the $\sqrt{s}$ energy of Tevatron and the design $\sqrt{s}$ of LHC.

gluons interacting in the $b\bar{b}$ production transport a variable fraction of the colliding protons momentum, they are very unlikely to have exactly opposite momenta. As a consequence, the $b\bar{b}$ pair momentum has a direction close to the effective beam axis as the longitudinal momentum of the $b\bar{b}$ pair is usually much larger than the transverse one. Once the b quarks are generated they collect light quarks, through the fragmentation (*hadronization*) process, generating baryons or more often mesons. Even if in the hadronization process gluons

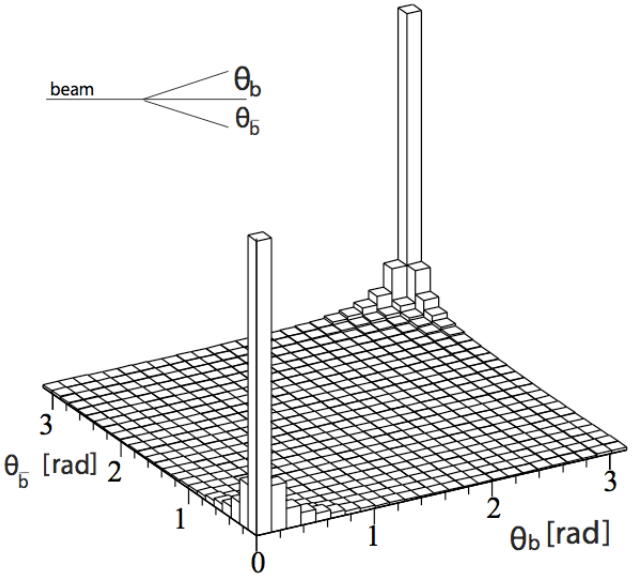

Figure III.3: The two-dimensional histogram representing the correlation in the directions of the B mesons events generated by a PYTHIA simulation of $pp \to B\overline{B}X$ events at LHC. The angles θ_b ($\theta_{\overline{b}}$) is the angle between the B ($\bar{B}$) meson flight direction and the beam axis.

are exchanged, the B meson flight direction does not differ significantly from the original b quark momentum direction. A Monte-Carlo simulation obtained with PYTHIA [29] is shown in Figure III.3. The illustration shows the remarkable correlation between the directions of the B and $\overline{B}$ mesons in the same event, and the favourite flight direction in forward and backward regions.

III.1.4 The luminosity and the beam time structure

The instantaneous luminosity $\mathcal{L}$ is an important parameter for an accelerator allowing to estimate the expected event rate when the cross section is known. It is defined as the ratio between the event rate and the cross section. It depends on various beam parameters:

$$\mathcal{L} = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{III.1}$$

where f is the frequency of colliding bunches, n_1 and n_2 are the number of protons per bunch and σ_x and σ_y represent the effective beam transversal dimensions. Because of the variation of n_1 and n_2 due to beam–beam collisions which eject protons from the beams, and because of the increase in σ_x and σ_y due to beam warming, $\mathcal{L}$ changes during the acquisition period. Hence, in order to estimate the number of events expected in a given data sample, the integrated luminosity $\int \mathcal{L}dt$ is usually preferred. To make easier the

multiplication with a known cross sections to estimate the number of events in a given period, the integrated luminosity is usually expressed in inverse picobarn pb^{-1} (or its multiples: 1 $\text{ab}^{-1} = 10^3$ $\text{fb}^{-1} = 10^6$ pb^{-1}).

At the LHC, the beam is structured in bunches, with a bunch spacing of 25 ns or multiples. In 2011 and 2012 it has been preferred to run with a bunch spacing of 50 ns to optimize the achievable luminosity. Up to 2340 bunches of protons per beam can be accelerated in the storage rings, with 25 ns bunch spacing. In 2011 and 2012 the number of bunches per beam was 1380. An important parameter is μ, the average number of pp collisions per bunch crossing visible in the detector. This parameter depends on the instantaneous luminosity of the accelerator. A high value for μ makes the triggering and the event reconstruction more difficult, worsens background to analyses measuring CP asymmetries, and challenges the radiation hardness and the rate capabilities of the detectors in the regions with higher occupancy. For these reasons the LHCb collaboration chose to operate at a luminosity not larger than $4 \cdot 10^{32}$ $\text{cm}^{-2}\text{s}^{-1}$, corresponding to $\mu \sim 1.7$ at $\sqrt{s} = 8$ TeV. Considering that the $b\bar{b}$ production cross section at 7 TeV, in the LHCb acceptance equals (75.3 ± 14.7) μb [85], 10^{10} $b\bar{b}$ pairs are estimated to be produced in the LHCb acceptance already with the first 37 pb^{-1} of data collected in 2010. BaBar and Belle together have produced $\sim 1.5 \times 10^9$ $b\bar{b}$ pairs in their full life-time with an integrated luminosity of $\sim$1.5 ab^{-1}.

III.2 The LHCb experiment

LHCb is the LHC experiment specialized in studies of b-physics. The experiment has a wide physics programme covering many important aspects of Heavy Flavour, Electroweak and QCD physics. In December 2009, a roadmap document was published to describe six of the key measurements that the LHCb Collaboration was expecting to achieve during the first run of the LHC.

The six measurements included in the document are [86]

- The measurement of the angle γ of the unitarity triangle defined by the CKM matrix, as described in Section I.3 on page 10, using only tree-diagram processes. The angle γ is the worst measured angle of the CKM unitarity triangle. Its measurement using tree-diagram processes is important because it is the only angle of the unitarity triangle which can be measured with negligible contribution from loop diagrams, where potential contributions of New Physics could arise. Besides, it allows to test the closure relation $\alpha + \beta + \gamma = \pi$ which is a non-trivial prediction of the Standard Model, which could be violated in models extending it by, *e.g.*, including additional quark families. The world average of the measurements of the angle γ is now led by the LHCb results, combined in Ref. [87].

- The measurement of the angle γ of the unitarity triangle using processes involving loops. The observation of a discrepancy with the value of γ as determined using tree-diagram processes would mean that New Physics contributions have appeared

in the loops. The LHCb measurement is even more precise using loop processes than tree-diagram decay only, as reported in Ref. [88].

- Observation of the decay $B_s^0 \to \mu^+\mu^-$ and measurement of its branching fraction. This decay is very interesting because part of the family of the flavour-changing neutral currents, and its branching ratio is precisely predicted by the Standard Model. In this case the transition $\bar{b} \to \bar{s}$ can only happen through loop diagrams involving a neutral mediator which in the SM can only be a Z^{*0} or a Higgs boson. A deviations from the predicted branching fraction of the $B_s^0 \to \mu^+\mu^-$ decay would be signature of physics beyond the Standard Model. The decay $B_s^0 \to \mu^+\mu^-$ has been observed by the LHCb Collaboration already with 2011 data [89], and the branching ratio has now been measured to a relative accuracy of 25% by combining the results of the CMS and LHCb Collaborations [90].

- Study of the decay $B^0 \to K^*\mu^-\mu^+$ which is interesting for the same reasons as the $B_s^0 \to \mu^+\mu^-$ but which allows, through angular analysis, to investigate the dynamical model of the decay. Its properties could actually be strongly modified by contributions of new physics arising in the loops. The latest measurement of the CP asymmetries of this decay have been recently published [91], while the branching fraction measurement and the angular analysis were published in Ref. [92]. A small deviation from the SM prediction is rising a considerable interest in the theory community.

- Measurement of the weak phase ϕ_s which is one of the angles of a secondary unitarity triangle of the CKM matrix, associated to the relation $V_{ub}V_{ud}^* + V_{cb}V_{cd}^* + V_{tb}V_{td}^* = 0$. The golden channel to perform this measurement is the time-dependent analysis of the decay modes $B_s^0 \to J/\psi\, K^+K^-$ and $B_s^0 \to J/\psi\, \pi^+\pi^-$, in particular through the resonance $B_s^0 \to J/\psi\, \phi$. The measurement, published in Ref. [93], represents the world best measurement of the phase ϕ_s, which is consistent with the SM expectations.

- Study of radiative B decays, as for example $B^+ \to K^+\pi^+\pi^-\gamma$. Radiative decays are associated, once again, to flavour-changing neutral currents so that they are another interesting probe for physics beyond the Standard Model. The analysis published in Ref. [94] has observed the first direct evidence of photon polarization in a $b \to s\gamma$ transition, which is very sensitive to New Physics, and is found to be consistent with the Standard Model.

The roadmap of the physics analyses accessible at LHCb has then expanded thanks to the excellent performance of the detector and to new ideas. For example,

- LHCb has performed world leading measurements of the lifetime of many b-hadrons as discussed in Section I.5.5. In particular the lifetime measurement of the Λ_b^0 baryon has allowed to correct a historical disagreement between the experimental value and the theoretical predictions. For a recent review see Ref. [47].

- LHCb has given important contributions in the field of *charm physics*. The world leading measurements of the masses of the D mesons [95], the measurement of CP asymmetries challenging for the first time the SM predictions [96, 97], and the observation of the first spin-3 heavy hadron, the $D_{sJ}^{+}(2860)$ meson state [98], are recent examples of important charm physics results achieved at LHCb.

- the production of quarkonia states is also an attractive field where LHCb has given important contributions. The double differential cross-section of most of the $1S$ and $2S$ triplet and singlet states have been measured for both charmonium and bottomonium states [99–106]. The recent result on the decay $\chi_b(3P) \to \Upsilon(3S)\gamma$ shed new light on the production mechanism of the $\Upsilon(3S)$ bottomonium state [107].

- The hadron spectroscopy has seen in the latest years revived interest due to the observation of many unpredicted states, known as exotic hadrons, similar to charmonium states, but not clearly interpreted in the quark model scheme. LHCb has published many studies of these states, for example it achieved the unambiguous observation of an exotic particle which cannot be classified within the quark model, the $Z(4430)$ particle. Previously observed by Belle [108], but not confirmed by Babar [109], this charged state decaying to $\psi(2S)\pi^-$ has been finally confirmed with the publication in Ref. [3] and constitutes the first unambiguous evidence for a tetraquark state.

- LHCb has measured the forward production of the electroweak vector bosons W^+ and Z^0 constraining the parton *pdf* in an unique region $x \sim 10^{-4}$ ($Q^2 = M_{W,Z}^2$), and setting limits to the Higgs boson production [110, 111].

- The Z^0 boson has been observed also in p-Pb collisions, achieving the first observation of Z^0 production in heavy ion collisions [112].

III.2.1 The LHCb detector

The LHCb detector [113], located in the cavern previously occupied by the DELPHI LEP experiment, has been developed as a single-arm detector, in contrast with the other three large LHC detectors (ATLAS, CMS and ALICE) which are called 4π detectors since they cover a solid angle of nearly 4π srad. The choice of having only one side equipped is a trade off between physics expectations and budget constraints.

The geometrical acceptance of LHCb for $b\bar{b}$ pairs is about 18%. This is due to the fact that the detector cannot be too close to the beam to measure particles in the highest pseudorapidity[1] regions, without increasing the background rate due to beams halos. The LHCb detector covers a pseudorapidity range $2 < \eta < 5$, corresponding to the interval 15 $\div$ 250 mrad for the polar angle θ.

The key features of LHCb include

- An excellent vertex and proper time resolution;

[1]The pseudorapidity η is defined as $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$, where θ is the angle between the particle momentum and the beam axis.

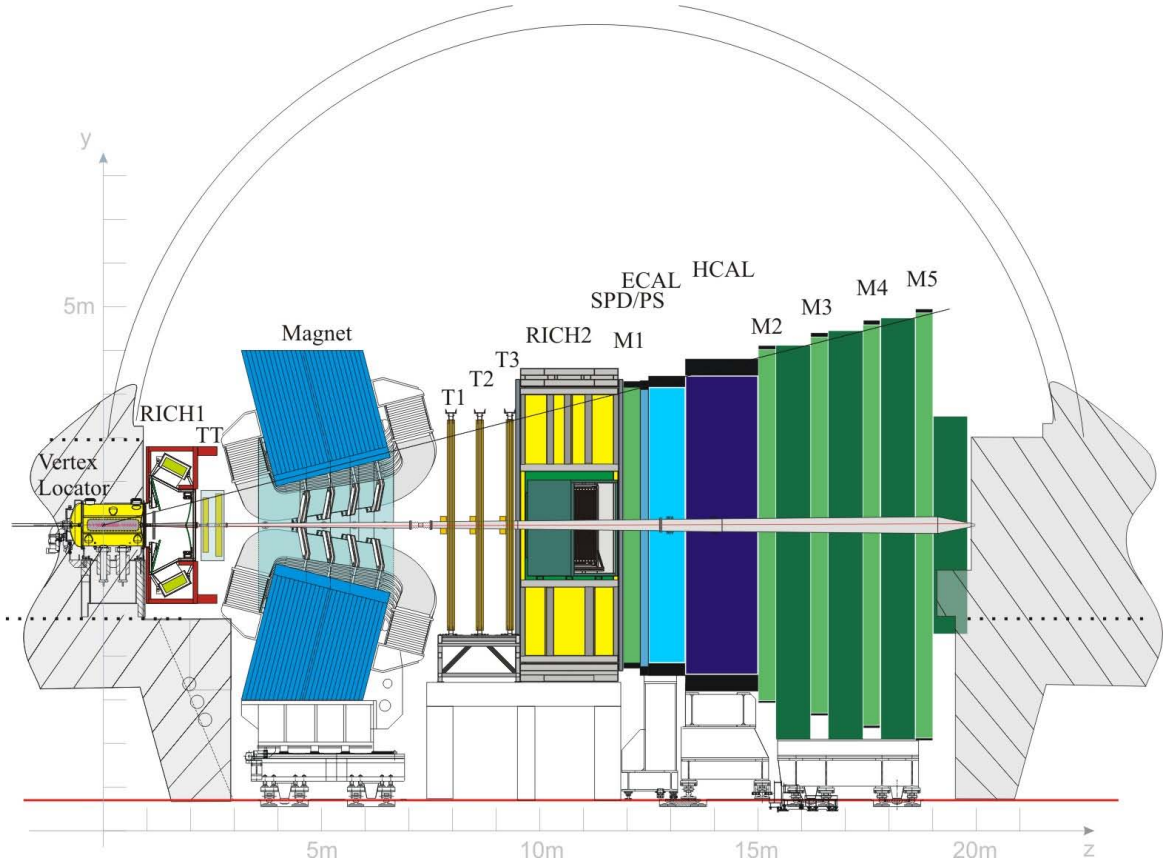

Figure III.4: A schematic representation of the LHCb detector in the non-bending vertical plane. The definition of non-bending is referred to the magnetic field, which bends particles trajectories in a plane orthogonal to the represented one. The origin of the Cartesian reference system is centered on the beam-beam interaction point. The z axis coincides with the beam axis and is directed towards downstream detectors, the y axis is vertical and defined to be parallel to the weight-force direction pointing upwards. The x axis is horizontal and forms a right-hand reference system with the axes defined above.

- Precise particle identification, especially for π-K separation;
- Precise invariant mass reconstruction. This feature is required to efficiently reject background due to random combinations of tracks (combinatorial background) and implies a good momentum resolution.
- A versatile trigger scheme. High efficiency is required in both leptonic and hadronic B decay channels, in order to collect high statistics samples and study the variety of modes with small branching ratios;

The LHCb detector can be conceptually subdivided in two subsystems:

- *The tracking system*, composed of a vertex locator and a set of tracking stations before and after a large dipole magnet,

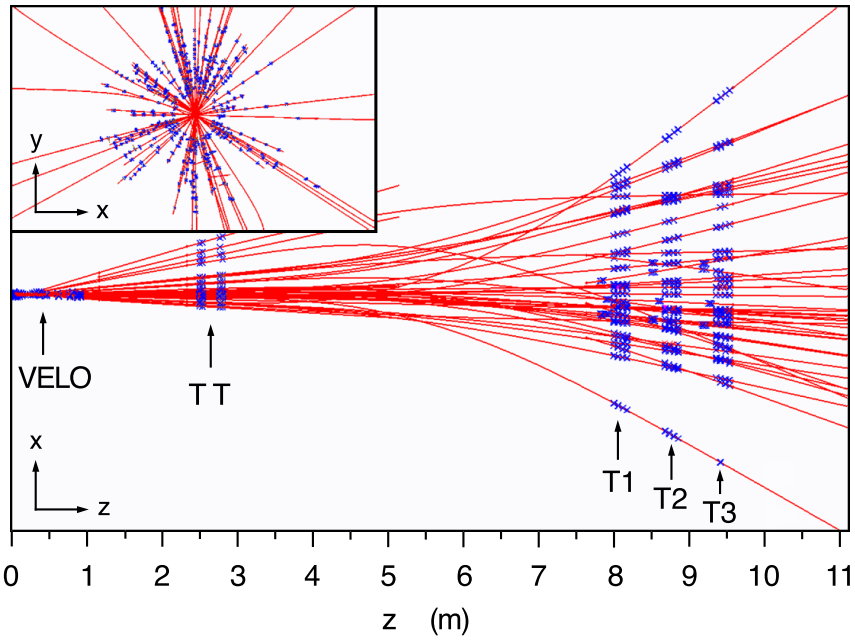

Figure III.5: View of the LHCb tracking system in an average-occupancy event. Blue markers represent the active hits, while red curves are the reconstructed tracks. The main figure shows the *bending plane*, while the inset shows the *transverse plane*. Titled arrows indicates the parts of the tracking system corresponding to the various sets of hits. The Figure was published in Ref. [114].

- *The particle identification system*, composed of a calorimeter system, two Ring Image Cherenkov (RICH) detectors and five muon stations.

The detector has been designed to be projective, in order to reduce dead zones for particle detectors.

The tracking system

The tracking system is based on four classes of detectors: the Vertex Locator (VELO), a set of silicon tracking stations surrounding the pp interaction region. The stations are orthogonal to the beam axis, in extreme proximity to it to enhance the resolution on the impact parameter; Trigger Trackers or *Tracker Turicensis* (TT) placed after the first RICH, before the magnet, and Inner and Outer trackers (IT and OT, named T1, T2 and T3 in Figure III.4) placed after the magnet in the inner and outer region respectively. The positions of the various sensitive parts of the tracking system are shown in Figure III.5 where an event with average occupancy is shown. The main Figure shows a projection on the plane xz, called *bending plane* because orthogonal to the magnetic field, while the inset the *transverse plane* xy. The z coordinate is parallel to the beam direction, the origin of the coordinate system is the nominal interaction point.

The aim of VELO trackers is to detect with the highest precision the position of the

secondary vertices in order to determine the decay time of weakly decaying particles when measuring their lifetimes or processes involving time-dependent CP violation. Its high resolution on vertex positions is also used to improve b-tagging and complex decay chains reconstruction efficiency.

The Trigger Trackers are used to set a reference before the magnetic deflection. With the additional information from trackers after the magnet, it is possible to evaluate the momentum and to trigger on its transversal component in order to select events with heavy-particle daughter tracks. Finally the Inner and Outer trackers are used to measure the particle deflection due to the 4 T·m integrated magnetic field, and therefore the particle momentum. The magnetic field is provided by a large warm dipole magnet represented in Figure III.6.

The profile of the magnetic field alongside the z coordinate is shown in Figure III.9 together with the classification of the reconstructed tracks in LHCb:

- *VELO tracks* are segments of tracks reconstructed in the VELO but with no corresponding hit in the TT. VELO tracks are rarely used for physics because of the poor information on their momentum, but are very useful to reconstruct the position of the primary vertices.

- *Upstream tracks* are segments of tracks reconstructed in the VELO and matched with hits in the TT, but not with hits in the downstream trackers T1, T2 and T3. Upstream tracks are used to reconstruct low momentum tracks that are *kicked* out of acceptance of the downstream trackers by the magnetic field. The information of the momentum of these tracks is given by the bending due to the residual magnetic field between the VELO and the TT.

- *Long tracks* are segments of tracks matching hits in the VELO, in the TT and in the downstream trackers T1, T2, and T3. Long tracks are the most useful for physics and the information on their momentum is the most precise and accurate.

- *Downstream tracks* are segments of tracks reconstructed only by the TT and by the downstream tracker, but not matching hits in the VELO. They are used to study decays of long-lived particles that decay between the vertex locator and the TT, in this case their originating vertex is outside of the VELO and they can only be reconstructed using the other tracking stations. Resolution on momentum is usually good, but resolution on the impact parameter is poor with resepect to long tracks, and the rate of misreconstructed tracks is higher due to the fewer channels constraining the existence of the track.

- *T tracks* are segments of tracks reconstructed only in the downstream tracking stations. They are very rarely used for physics, but are useful for calibration and detector studies.

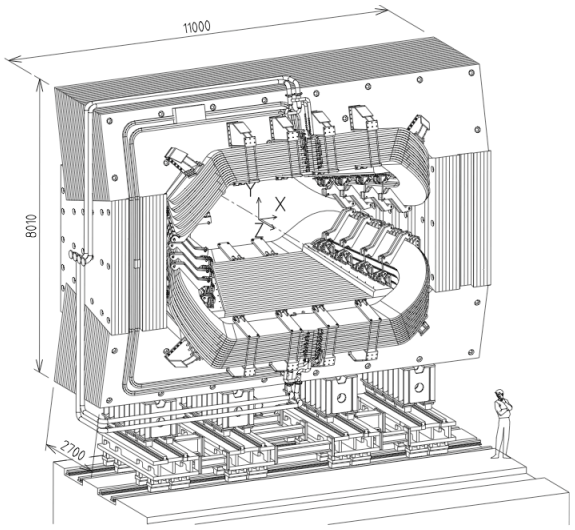

Figure III.6: A schematic representation of the LHCb magnet as published in Ref. [113].

The Vertex Locator

The Vertex Locator (VELO) is a silicon detector developed to provide precise measurement of track coordinates close to the interaction region, which are used to identify the displaced secondary vertices which are a distinctive feature of b and c-hadron decays [113]. The VELO consists of a series of silicon modules, each providing a measure of the radial distance from the origin r and of the azimuthal angle ϕ, arranged along the beam direction (z-axis). Two planes are located upstream of the VELO sensors to improve the determination of the position of the primary vertices.

The interaction region are placed at a radial distance from the beam smaller than the aperture required by the LHC during injection, and therefore it has to be retractable. The detectors are mounted in a vessel that maintains vacuum around the sensors and is separated from the machine vacuum by a thin walled corrugated aluminum sheet, referred hereafter as *RF-foils*, also protecting the sensors from RF background of the machine. The RF-design allows to minimize the material traversed by a charged particle before it crosses the sensors and the geometry is designed to make the two halves of the VELO to overlap when in the closed position.

Figure III.7 represents the VELO detector through the drawing of one of the 23 stations in the xy plane and the arrangement of the station in the xy plane. Each half-station is composed of two layers measuring the r and ϕ coordinates.

The VELO covers the angular acceptance of the downstream detectors, *i.e.* detect particles with pseudorapidity in the range $1.6 < \eta < 5$ emerging from primary vertices in the range $|z| < 10.6$ cm. Such tracks are required to cross at least three VELO stations.

The choice of a cylindrical coordinate system was chosen to enable fast reconstruction in the LHCb trigger. Indeed, the 2D tracking in the plane rz is sufficient to identify quickly events containing tracks with large impact parameter, worth for a more accurate

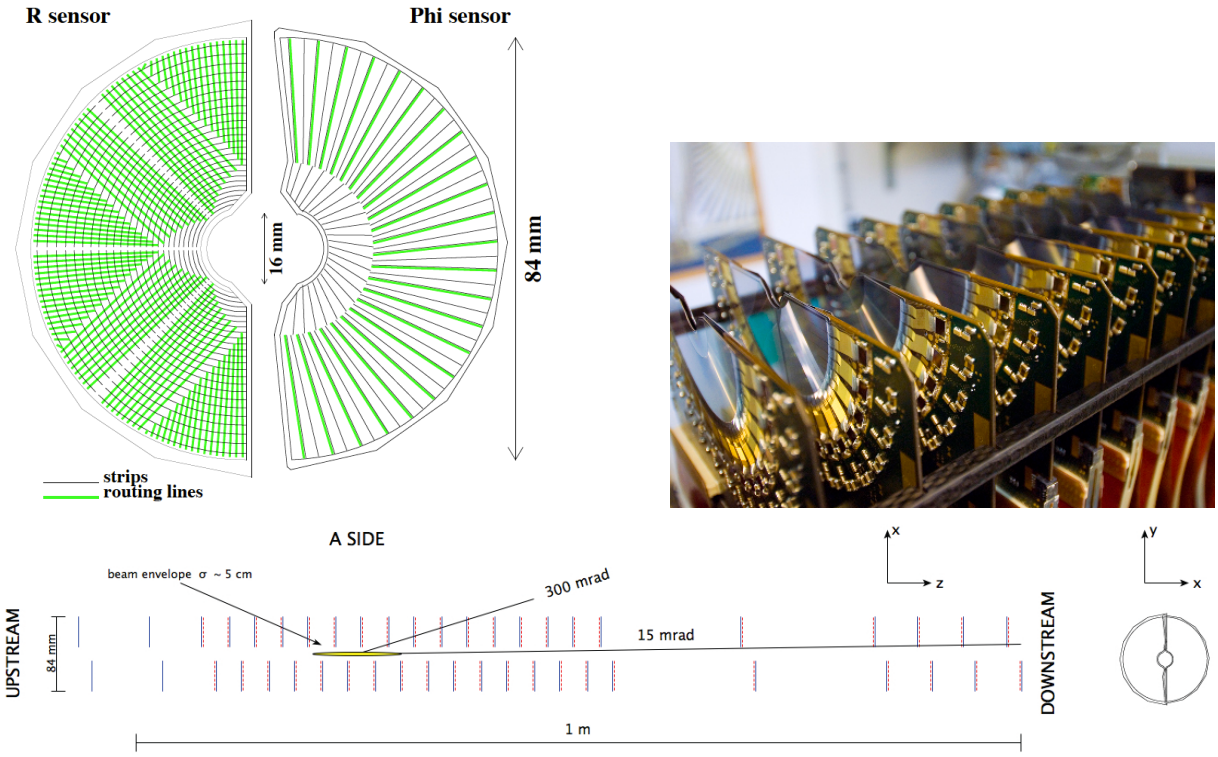

Figure III.7: On top left, sketch illustrating the $r\phi$ geometry of the VELO sensors. For clarity, only a portion of the strips are illustrated. On the top right, a photo of a half of the VELO detector is shown. On the bottom the cross-section in the xy plane is shown with the detector in the fully closed position. The solid blue lines represent the r-strip layers, while ϕ-strip layers are shown as dashed red lines. The Figures were published in Ref. [115].

and time-expensive reconstruction in the 3D space.

The number of physical channels used to readout the VELO detector is 180000, and represents a trade-off between requirements and budgetary limits.

The geometry and the technology of the sensors has been developed to comply with needs in terms of radiation tolerance imposed by the severe radiation environment at 7 mm from the LHC interaction point.

The minimum pitch was imposed by technological limits in the fabrication of the silicon devices and is approximately 32 µm. The pitch increases up to about 100 µm, moving from the inner to the outer region. The choice optimizes the vertex resolution and ensures that measurements along the track contribute to the impact parameter precision with roughly equal weight.

The sensors are allocated on modules which can be retracted during the LHC operational phases of the LHC when the beams are not focused enough to ensure the safety of the detector. Before the LHC ring is filled, the detectors move away from the interaction region by 30 mm in order to allow for beam excursions during injection and ramping. Once the beams reach stable conditions, the detectors are placed into the nominal position

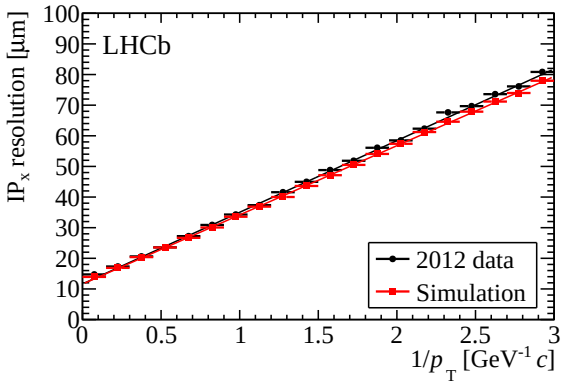

Figure III.8: Schematic view of the tracks in the LHCb spectrometer (bottom, together with the indication of the magnet field (top). The Figure was published in Ref. [114].

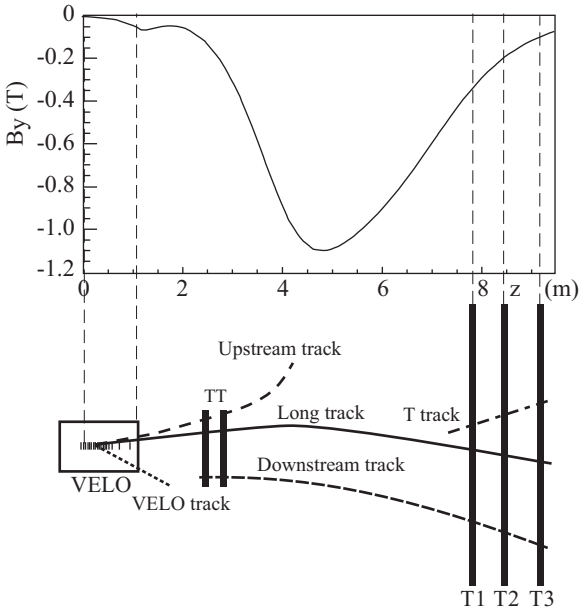

Figure III.9: Impact parameter resolution as a function of the transverse momentum p_T. Good agreement is observed between 2012 data (shown in black) and full simulation (in red). The Figure was published in Ref. [114].

transversally centered around the interaction region. The position can vary from fill to fill over ±5 mm in both x and y. The closing procedure includes therefore a phase of determination of the position of the interaction point in the xy plane, followed by the complete closure of the detector in its stable data-acquisition position. In stable running condition, with the VELO closed, the distance between foil and beam is 7 mm.

The resolution on the position of the primary vertices obtained with the VELO varies between 9 and 35 μm for the x and y coordinates, and between 50 and 280 μm for the z coordinate, depending on the number of tracks used to reconstruct the vertex. Typical events with a 25-track primary vertex have a transversal resolution of approximately 15 μm and a longitudinal resolution of 71 μm [115].

The VELO has the highest resolution on the impact parameters of charged tracks among the large experiments at the LHC varying between 10 and 80 μm depending on the transverse momentum of the considered track. Indeed, the dominant uncertainty on the impact parameter is the multiple scattering of the tracks crossing the VELO stations. Hence, an inverse relation with the transverse momentum of the track is expected and observed, as shown in Figure III.8.

The Trigger Tracker

The Trigger Tracker (TT) is located just upstream the magnet. It consists of two stations separated by 27 cm. Each station has two layers of silicon strip detectors covering the full acceptance. The strips of the four layers are arranged in order to measure the x, u v, x, coordinates where u and v are non-orthogonal axes such that there is a 5 degrees angle between u and x, and between v and x, clockwise and anti-clockwise respectively. This structure is named $xuvx$ and allows a spatial resolution of about 50μm in the x coordinate.

The $xuvx$ geometry has been chosen to enhance resolution in the bending plane, which is relevant for the momentum measurement, and has the great advantage over a orthogonal (or Cartesian) reference system that it allows to solve ambiguities in case of two or more tracks.

Each detection layer is composed of a set of half-modules covering half of the height of the LHCb acceptance. A half-module is a row of seven silicon sensors.

The main advantage of this detector design is that all the front-end electronics and the cooling infrastructure are located outside of the acceptance of the experiment: above or below the active area of the detector.

Inner and outer tracker

The tracking stations located downstream of the magnet (T1, T2 and T3 in Figure III.4) are required to measure the momentum of charged particles deflected by the dipole magnetic field. The inner tracker (IT), closer to the beams, is made of silicon detectors, while the detectors in the outer tracker (OT) are *straw tubes*, gas detectors developed to minimize the material budget before the calorimeters. Inner trackers only cover a region of $\sim 2\%$ of the total area (5 m $\times$ 6 m) of a tracking station, however they measure about 20% of the particle flux, due to the low-angle peak in particle distributions. Each of the three IT stations consists of four individual detector boxes that are arranged around the beam-pipe. Each detector box contains four detection layers and each detection layer consists of seven detector modules. The thickness of the sensors is either 320 or 410 µm for the top–bottom and side detector boxes, respectively. The pitch is 198 µm. The 130 000 strips used are either 11 cm or 22 cm long depending on the installation region and the surface they cover is roughly 4 m^2. The digitization of the electronic signal is performed outside the LHCb acceptance to reduce the material budget.

The four layers of both IT and OT follow the $xvux$ structure described in the previous section. Each OT station consists of 12 double layers of straw tubes. The tubes are 2.4 m long with 4.9 mm inner diameter, and are filled with a gas mixture of $Ar/CO_2/O_2$ (70/28.5/1.5) which guarantees a fast drift-time below 50 ns and good aging properties. The straw tubes are fixed to carbon-fiber panels resulting in gas-tight boxes enclosing stand-alone detector modules. Each module is composed of two layers of 64 drift tubes each. In the longest modules, the straw tubes are split in the middle into two independent readout.

The complete OT detector consists of 168 long and 96 short modules and comprises 53 760 single straw-tube channels.

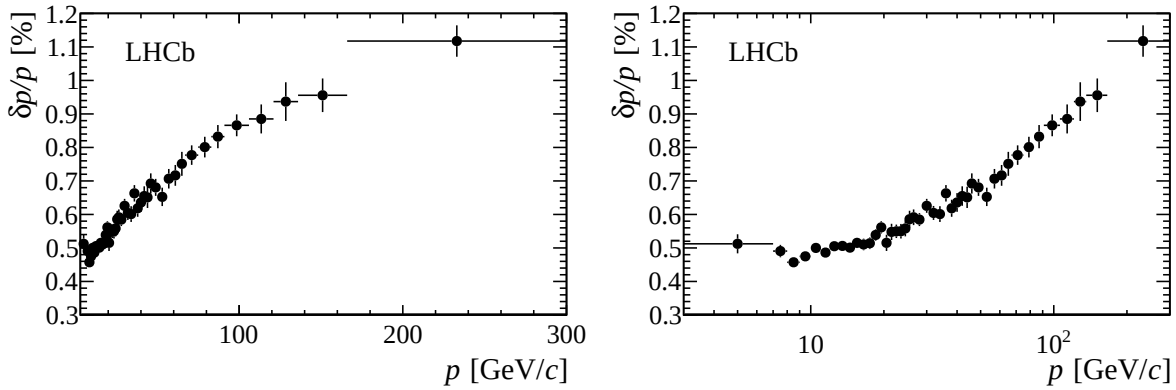

Figure III.10: Resolution on the momentum obtained with the LHCb spectrometer, shown in linear (left) and log (right) scale. The Figure was published in Ref. [114].

The hit resolution along the x-axis is 50 μm and 200 μm for the inner and outer trackers, respectively.

The magnet

The superconducting magnet originally proposed for the spectrometer was too expensive and the construction time was too long. It has been replaced by a warm magnet with saddle-shaped coils matching the detector acceptance. It provides a peak field of 1.1 T and an integrated bending power of 4 T·m (for 10 m long tracks).

The magnetic field was inverted periodically (about all the two weeks) during the data-taking to reduce asymmetries due to dis-uniformities in the detector. This is of particular importance for analyses involving asymmetries between events originated by B mesons of opposite flavour.

Performance of the LHCb tracking system

The combination of the excellent spectrometer, the low material budget before the calorimeters, and the excellent vertex resolution are essential ingredients for a number of high-precision measurements.

The measured resolution on the momentum p is presented in Figure III.10 for long tracks.

As discussed in Section II.1, the accuracy of the momentum measurement is crucial for decays with high Q-value as the typical heavy flavour decays. Decays of well known $c\bar{c}$, $b\bar{b}$ and Z^0 resonances are used to calibrate the momentum measurement. Figure III.11 shows the mass spectrum of the Υ resonances as reconstructed in their decays to dimuon. The mass resolution is obtained with maximum likelihood fits using the data model superimposed. Combining the mean value and resolution of the Υ resonances, with those of J/ψ, $\psi(2S)$ and Z^0 the mass resolution as a function of the mass is interpolated as shown in Figure III.12.

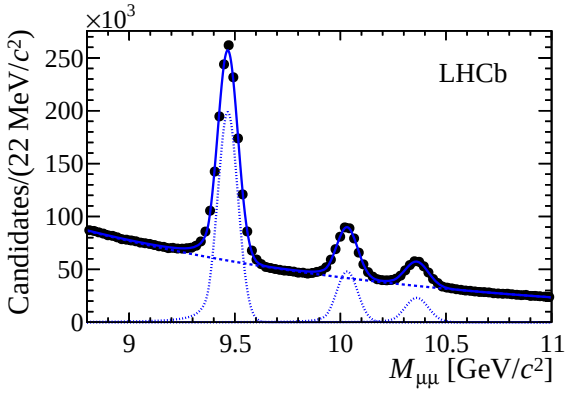

Figure III.11: Mass peaks of the bottomonium resonance $b\bar{b}$: $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$. The Figure was published in Ref. [114].

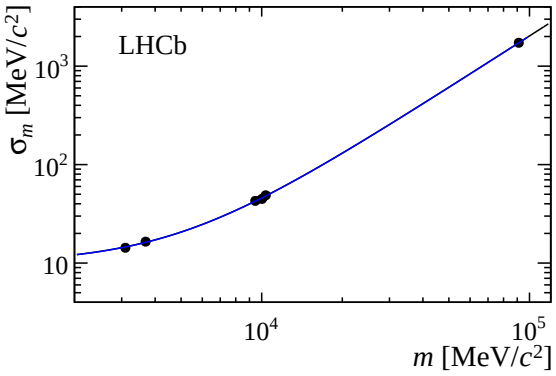

Figure III.12: Dependence of the mass resolution on the mass scale studied using the decays of J/ψ, $\psi(2S)$, $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$, and Z^0 to dimuon. The Figure was published in Ref. [114].

If combined with the information on the decay length, the measurement of the momentum allows to determine the decay time in the rest frame of the mother particle (at least when the latter is fully reconstructed). The proper decay time resolution is essential to perform studies on time-dependent CP asymmetries, which constitute an important part of the physics programme of the LHCb experiment. It also allows for precision lifetime measurements, as the one discussed in this Thesis. The resolution achieved by the LHCb tracking system is enough to resolve the flavour oscillations $B_s^0 \leftrightarrow \overline{B_s^0}$ occurring during time evolution. Figure III.13 shows the proper time distribution of B_s^0 mesons that have been tagged, using the rest of the event (often the other b-hadron) to establish whether the flavour observed in the decay is the same as the flavour at production time (unmixed) or not (mixed). The typical proper decay time resolution for the fully reconstructed B meson decays is 40 to 50 fs.

The RICH system. Charged Particle Identification

If compared to ATLAS and CMS, at the LHC, and CDF and D0, at Tevatron, LHCb is equipped with a more efficient system of charged particle identification to discriminate between protons, electrons, muons, pions and kaons abundantly produced in B and D meson decays. The first RICH detector (RICH 1 hereafter) is placed before the magnet and aims at measuring low momentum particles which can be deflected out of the detector acceptance by the magnetic field. The RICH 1 provides an optical structure made of spherical and plane mirrors which are required to reflect Cherenkov light to the photomultiplier (PMTs) tubes conserving the information on the emission angles. The PMTs are placed outside the geometrical acceptance and a magnetic shield is provided. Two different radiators share the optical system of RICH 1: silicon aerogel with a refractive index tuned to 1.030 and C_4F_{10} with a refractive index slightly dependent on the wavelenght. At $\lambda = 400$ nm,

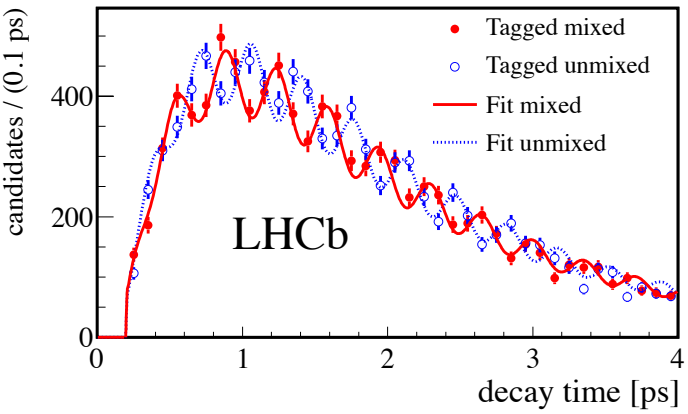

Figure III.13: Decay-time distribution of the decays $B_s^0 \to D_s^+\pi^-$ for the B_s^0 meson tagged as unmixed (same flavour as at the production) or mixed (opposite flavour) studying the rest of the event. The Figure was published in Ref. [114].

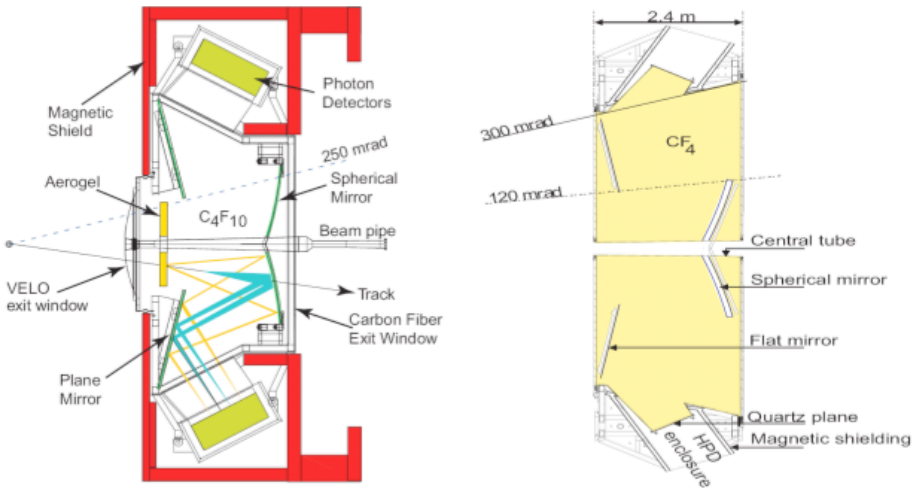

Figure III.14: A schematic representation of RICH 1 and RICH 2 detectors. The Figure was published in Ref. [113].

the gaseous radiator has a refractive index of 1.0014.

Particles with a higher momentum are supposed to remain in the geometrical acceptance after the magnetic deflection. A second RICH (RICH 2), optimised for high momentum particles, is thus placed after the magnet. The optical structure is somehow similar to the RICH 1, but the only radiator is CF_4, having refractive index of 1.0005 at $\lambda = 400$ nm. The RICH1 acceptance requires a momentum between 1 and 70 GeV/c, while for RICH 2 the interval is $12 \div 150$ GeV/c. The RICH detectors are schematically represented in Figure III.14.

RICH detectors are based on the Cherenkov light emission of a particle traversing a material with refractive index n. The angle of light emission, or Cherenkov angle is defined by the relation [116],

$$\cos\theta_c = \frac{1}{n\beta}. \tag{III.2}$$

For a fixed radiator material, the angle θ_c depends only on the mass and momentum of the particle. Figure III.15 represents the dependence of the Cherenkov angle on the momentum of pions, kaons, and protons. The Cherenkov angle of muons can also be distinguished at very small momentum, however the separation from pions is not sufficient to guarantee an effective muon identification, which is achieved by a dedicated muon system.

To identify a charged particle, the information on its direction and the momentum is taken from the tracking system. One ring for each possible mass hypothesis is constructed and the likelihood of each test ring against the positions of the activated PMT channels is evaluated. In Figure III.16 the firing PMT channels are shown with markers and the ring hypotheses associated to the proton, the kaon and the pion masses are superposed. The RICH likelihood is a constructed taking into account the different resolution on the position of the hits available in different regions of the PMT plane. This RICH likelihood value, combined with the information from the calorimeters and the muon system, is associated to each track, allowing to define particle identification criteria in offline analyses. The combined likelihood for the hadron hypothesis h_1 is usually expressed relative to the likelihood of a second hadron h_2 which represents the background to a specific channel. Namely,

$$\mathrm{DLL}_{h_1/h_2} = \log\left(\frac{\text{Combined Likelihood for } h_1}{\text{Combined Likelihood for } h_2}\right). \tag{III.3}$$

Since the most abundant species produced at hadron colliders are pions, it is customary to use the variables $\mathrm{DLL}_{K/\pi}$ and $\mathrm{DLL}_{p/\pi}$ to reject pions when selecting kaons and protons, and the variable $\mathrm{DLL}_{p/K}$ when rejecting kaons in a proton selection. The population of the plane $\mathrm{DLL}_{K/\pi} \perp \mathrm{DLL}_{p/\pi}$ is shown in Figure III.17. Cutting on the variable $\mathrm{DLL}_{K/\pi}$ it is possible to select kaons with an arbitrary efficiency at a cost of a certain amount of misidentification $\pi \to K$. In Figure III.18, the efficiency and the misidentification probability for selecting kaons agains pions is shown as a function of the particle momentum. There is a performance drop around 50 GeV/c, but the $K-\pi$ separation remains acceptable up to 150 GeV/c.

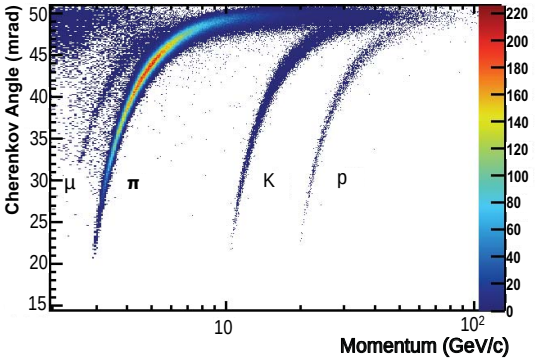

Figure III.15: Depencence of the Cherenkov angle on the momentum of the track for different masses. The picture shows the limits of the RICH 2 radiator which is blind to protons of momentum smaller than 20 GeV/c, and to kaons of momentum smaller than 10 GeV/c. The identification of particles becomes challenging beyond 50 GeV/c where the angle distributions for the mass hypotheses start overlapping. The Figure was published in Ref. [114].

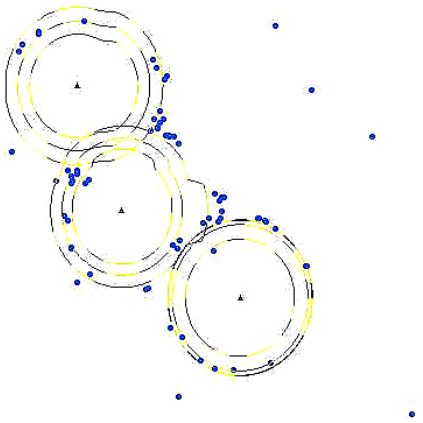

Figure III.16: RICH 2 photo-multiplier panel with pixels and hypothesis rings. The points represent the positions of the photomultiplier channels collecting a photon. Rings are traced taking as input the position and the momentum of tracks from the tracking system, and assuming the mass of the pion, the kaon or the proton. The three hypotheses give the different concentric rings shown in the figure. The picture was published in the outreach LHCb web page.

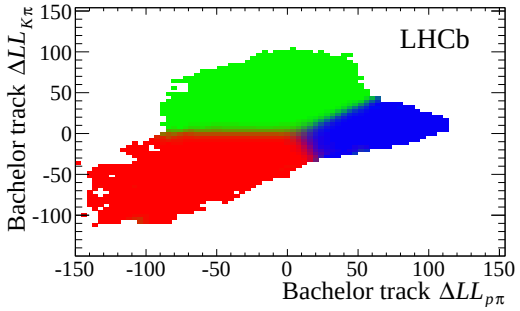

Figure III.17: Graphical representation of the population of the plane $\mathrm{DLL}_{K/\pi}$ $\perp$ $\mathrm{DLL}_{p/\pi}$ by different species of hadrons. The color of each bin is defined by the RGB code having red, green and blue components representing the fraction of pions (red), kaons (green), and protons (blue) found in that bin.

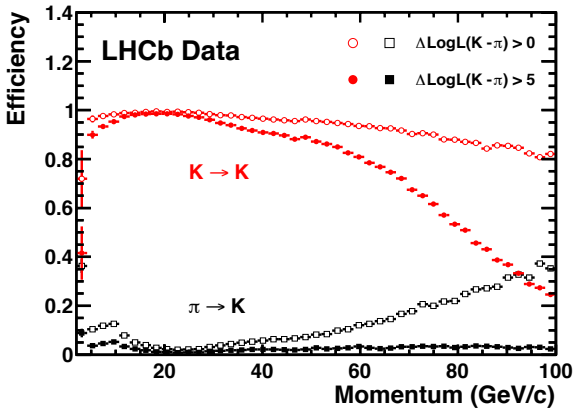

Figure III.18: Kaon efficiency and pion misidentification probability obtained cutting on the $\mathrm{DLL}_{K/\pi}$ variable with threshold 0 and 5, as a function of the particle momentum. The Figure was published in Ref. [114].

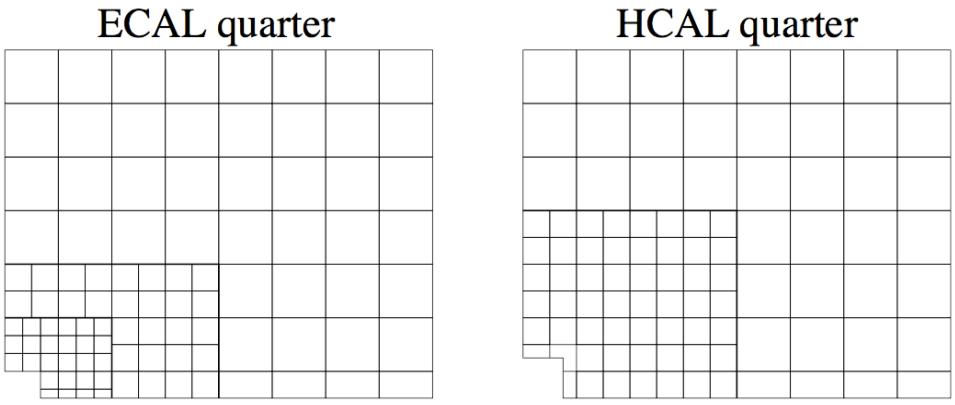

Figure III.19: Calorimeter lateral segmentation: electromagnetic calorimeter (left), hadronic calorimeter (right).

The calorimeters

The LHCb calorimeter system is composed of an electromagnetic calorimeter preceded by a SPD and a PS and followed by and, downstream, a hadronic calorimeter. The former is designed to identify electromagnetic showers generated by $e^{\pm}$ and γ, while the latter records energy mostly deposited by hadronic showers.

The main purposes of the calorimeter system are:

- to provide the transverse energy measurement for charged hadron, electron and photon candidates. This information is used for the lower level trigger decision.
- to provide particle identification to distinguish between electrons, photons and hadrons, including the separations of single photons and π^0 decays. Calorimeters are also intended to measure the position and the energy of neutral particles.
- to achieve high reconstruction accuracy for π^0 and photons.

The electromagnetic calorimeter

The Electromagnetic CALorimeter (ECAL) is a sampling calorimeter of *shashlik* type composed of multiple and alternated layers of lead and scintillator, readout by plastic optical fibers. The thickness of the ECAL is 25 radiation lengths. The layers are perpendicular to the beam direction and segmented transversally into towers. The optical fibers are disposed along the beam direction at regular intervals, and collect the signals from active layers of each tower to the PMT situated on the back of each detector. The lateral segmentation in square-shaped cells provides good shower separation and angular resolution. The cell size varies with the region: in the inner section, closer to the beam, where the ECAL is required to deal with a higher particle rate, it is equipped with 40.4 mm wide cells. The cells in the intermediate region are 60.6 mm wide, while in the outer region a 121.2 mm width is

considered sufficient to deal with the lower particle rate. The ECAL lateral segmentation is shown in Figure III.19, left. The ECAL energy resolution can be parametrized as

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E/1\text{GeV}}} \oplus 1\% \tag{III.4}$$

where the symbol $\oplus$ indicates a squared sum root: $a \oplus b = \sqrt{a^2 + b^2}$.

To distinguish between charged and neutral particles hitting the calorimeter a scintillator pad detector (SPD) is placed in front of the electromagnetic calorimeter. It is a fine laterally segmented scintillator layer which aims to identify charged particles and associate an input position to them before the electromagnetic shower to improve the association of the shower to the corresponding charged track. In order to distinguish electromagnetic and hadronic showers and disentangle overlapping electromagnetic showers casted by different particles a second finely segmented scintillator pad, called pre-shower detector (PS), is placed downstream of a 2.5 X_0 thick lead converter layer, followed by the SPD.

The hadronic calorimeter

The Hadronic CALorimeter (HCAL) is a sampling device made of iron and scintillating tiles, as absorber and active material respectively. In contrast with the ECAL, the scintillator tiles of the HCAL are oriented to be parallel to the beam axis. This feature can be used to obtain a better angular resolution from the HCAL. The HCAL thickness equals 5.6 interaction lengths, which are not sufficient to ensure the whole containment of the hadronic shower, but are enough to achieve a reasonable measurement of the energy.

The HCAL is laterally segmented into square cells of size 131.3 mm in the inner section and 262.6 mm in the outer section, see Figure III.19, at right.

The readout is also made through optical fibers running along the beam direction as in the shashlik design.

The muon system

Muon triggering and offline muon identification are fundamental requirements of the LHCb experiment. Muons are present in the final states of many CP-sensitive B decays, in particular the two "gold-plated" decay modes $B^0 \to J/\psi\,(\mu^+\mu^-)K^0_S$ and $B^0_s \to J/\psi\,(\mu^+\mu^-)K^+K^-$. Moreover, muons from semi-leptonic b decays provide a tag of the initial state flavour of accompanying neutral B mesons. Finally, the flavour changing neutral current decays $B^0_s \to \mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ are rare decays having muons in the final state.

The muon system plays therefore an important role in the physics programme of LHCb, and in particular it is required to provide a high-p_T muon trigger at the earliest trigger level (Level-0). Concerning the requirements for offline muon identification, the muons reconstructed in the high-precision tracking detectors with momenta down to 3 GeV/c must be correctly identified with an efficiency higher than 90%, while keeping the pion misidentification probability below 1.5%.

	M1	M2	M3	M4	M5
R1[mm^2]	10 × 25	6.3 × 31	6.7 × 34	29 × 36	31 × 39
R2[mm^2]	20 × 50	12.5 × 63	13.5 × 68	58 × 73	62 × 77
R3[mm^2]	40 × 100	25 × 125	27 × 135	116 × 145	124 × 155
R4[mm^2]	80 × 200	50 × 250	54 × 270	231 × 290	248 × 309

Table III.1: Logical pad dimensions (horizontal × vertical) for each station and region. Dimensions follows the projective principle of LHCb design.

The detector is structured in five stations named M1 to M5, represented as light green boxes in Figure III.4. The first station (M1) is placed upstream the calorimeter pre-shower, at 12.1 m from the interaction point. M1 is important for the transverse-momentum measurement of the muon track used in the Level-0 muon trigger. The remaining four stations are interleaved with the iron filters at mean positions of 15.2 m (M2), 16.4 m (M3), 17.6 m (M4), and 18.8 (M5). Muons are filtered by the electromagnetic and hadronic calorimeters and three 80-cm thick iron layers for a total absorber-thickness of 20 nuclear interaction lengths. The chambers between the filters are allocated in about 40 cm of space.

The inner and outer angular acceptances of the muon system in the bending (non-bending) plane are 20 (16) mrad and 306 (258) mrad, respectively. The geometrical acceptance of muons from b-decays is about 20% relative to the full solid angle. A side view of the muon system is shown in Figure III.20 while a photo of part of the detector is shown in Figure III.21.

The layout of the muon stations is structured in four quadrants (left-right, and top-bottom) and in four regions R1, R2, R3, and R4. The granularity of the detectors in the four regions scales with the ratio 1:2:4:8, as well as the region width. This allows to obtain a roughly constant occupancy through the detector despite the large variation of the particle flux passing from the central part, close to the beam axis, to the detector border. Regions have different logical-pad dimensions as summarized in Table III.1 and represented in Figure III.22 where the reported numbers refer to the second muon station (M2).

The whole stations M2, to M5, and the three outer regions of M1 are equipped with 1368 *Multi-Wire Proportional Chamber* (MWPC). The gas used is a mixture of $Ar/CO_2/CF_4$ $(40 : 55 : 5)$. Muon chambers in M2...M5 consist of four active layers as represented in Figure III.23, while chambers in M1 have only two active layers to reduce the material budget in front of the calorimeters reducing multiple scattering. This redundant design minimizes the detector failure rate and improves the detector efficiency. Signals from each layer are OR-ed, and an efficiency better than 95% in a 20 ns window at a gas gain $G \sim 10^5$ is achieved.

In the innermost region R1 of the station M1, the the detector is required to stand at a rate up to 500 kHz/cm^2 of charged particles. Due to the large particle flux in this region the chambers must also be especially radiation hard, in order to avoid visible aging

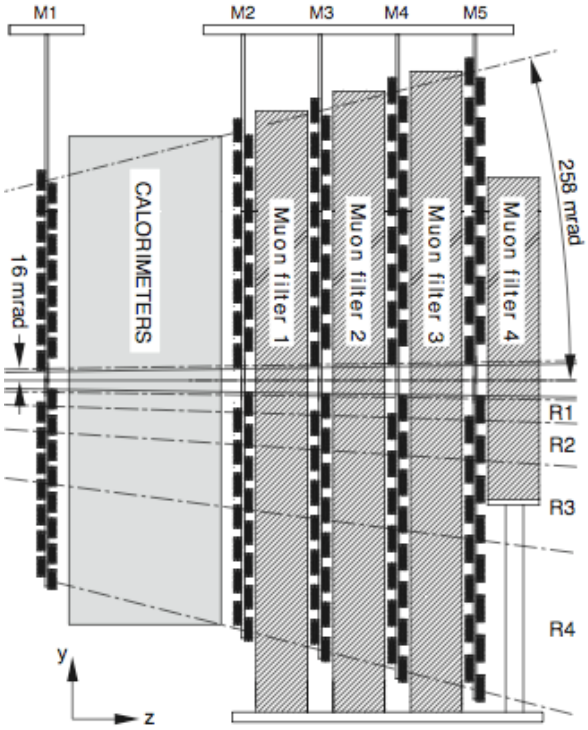

Figure III.20: Side view of the muon system. The figure was published in Ref. [113].

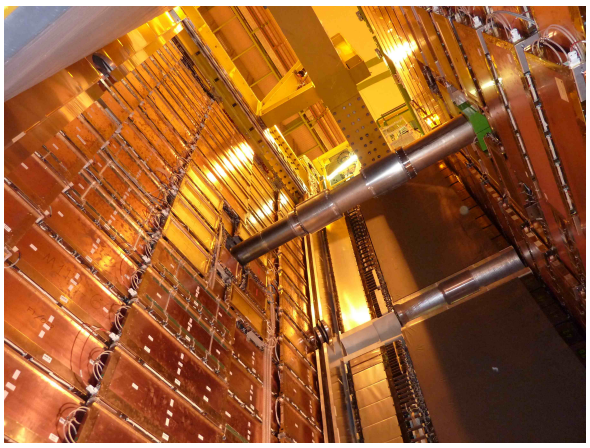

Figure III.21: Photo of the muon system with open calorimeters. M1 and M2 are the copper-colored walls on the sides of the image. The hole in the detector to allocate the beam-pipe, and the beam-pipe itself, are visibile in the center of the figure.

effects in 10 years of LHCb operations. The LHCb Collaboration has therefore chosen an alternative technology, resulting in triple Gas Electron Multiplier (triple-GEM detectors) to cover the 20×24 cm^2 active area in M1R1.

The triple-GEM detector, which consists of three gas electron multiplier (GEM) foils sandwiched between anode and cathode planes, provides comparable time and position resolution as a MWPC, with higher rate capability and radiation hardness. The ionisation electrons, produced in the drift gap between the cathode and the first GEM foil, are attracted by electric fields through the three GEM foils where they are multiplied. Once they cross the last GEM foil, they drift to the anode in the interaction gap, giving raise to an induced current signal on the pads.

A cross-section of the detector is shown in Figure III.24. The best values of the gap fields and of the voltage across the GEM foils were determined experimentally by optimizing time resolution versus discharge probability and are typically $V_1 = 440V$, $V_2 = 430V$, and $V_3 = 410V$ corresponding to electric field varying between 3.5 and 5 kV/cm from gap to gap.

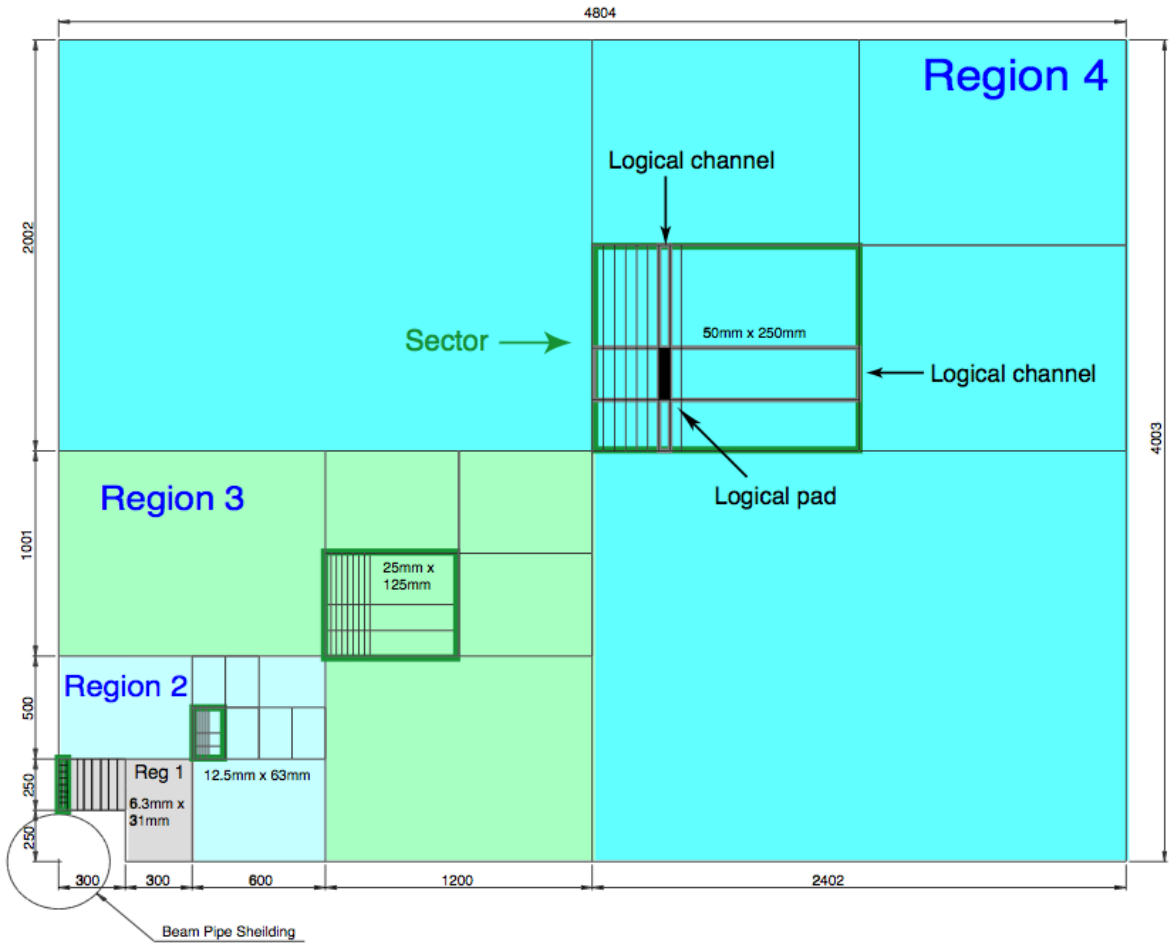

Figure III.22: Front view of a quadrant of muon station M2, showing the dimensions of the regions. Inside each region is shown a sector, defined by the size of the horizontal and vertical strips. The intersection of the horizontal and vertical strip, corresponding to the logical channels, are logical pads. The region and channel dimensions scale by a factor two from one region to the next.

The signal from two triple-GEM layers are OR-ed to raise the efficiency to 96% in a 20 ns window at a gain 6 $\cdot 10^3$.

Standard Muon Identification techniques

The LHCb muon identification strategy has been recently reviewed in Ref. [117], and can be divided in three steps:

- A loose boolean selection of muon candidates based on the penetration of the muon candidate through the calorimeters and ion filters, which provides high efficiency while reducing the misidentification probability of hadrons at the percent level. This criterion, detailed below, is named IsMuon.

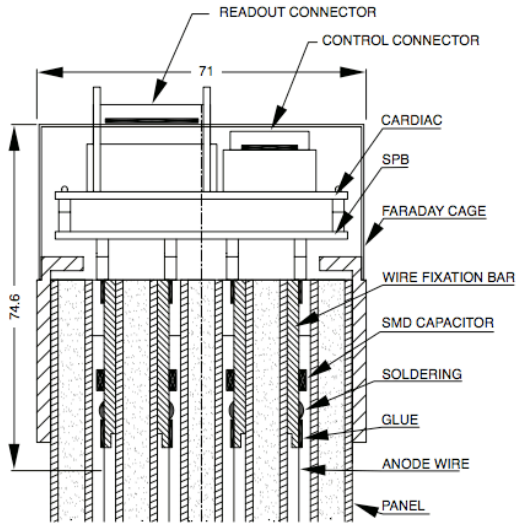

Figure III.23: Cross-section of a muon Multi-Wire Proportional chamber as used in the M2, M3, M4, and M5 stations. Chambers in M1 have two active layers only in order to reduce material budget in front of the calorimeters. The figure was published in Ref. [113].

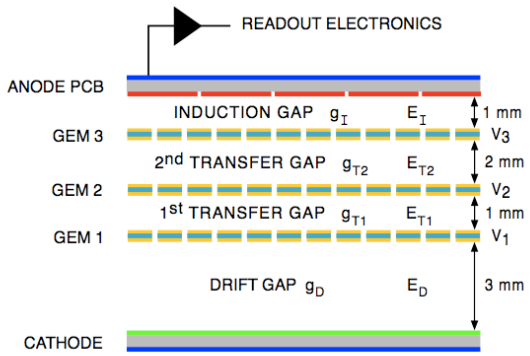

Figure III.24: Schematic representation of the triple-gap Gas Electron Multiplier (GEM) chambers used in the inner region (R1) of the first muon station (M1). The figure was published in Ref. [113].

- Computation of a likelihood for the muon and non-muon hypotheses, based on the pattern of hits around the extrapolation to the different muon stations of the charged particles trajectories reconstructed with high precision in the tracking system. The logarithm of the ratio between the muon and non-muon hypotheses is used as discriminating variable and called muDLL.

- Computation of a combined likelihood for the different particle hypotheses, including information from the calorimeter and RICH systems. The logarithm of the ratio between the muon and pion hypotheses is used as discriminating variable and called $\mathrm{DLL}_{\mu/\pi}$.

The selection based on IsMuon and likelihood criteria are independent and can be applied separately. However, it is customary to refine the IsMuon selection with a subsequent likelihood-based criterion.

The binary selection criterion, IsMuon, is defined according to the number of muon stations where a hit is found within a field of interest (FOI) defined around the track extrapolation. The number of stations required to have a muon signal is a function of the track momentum, as shown in Table III.2. The sizes of the field of interest, which depend on the region and on the particle momentum, are optimized using a data driven method exploiting J/ψ decays to dimuon named *tag and probe*, detailed below. The decay of the J/ψ resonance to dimuon is exploited to select a pure sample of *probe* muons without

Momentum range		*Muon stations*
Minimum [GeV/c]	*Maximum* [GeV/c]	*with hits in the FOI*
3	6	M2 and M3
6	10	M2 and M3 and (M4 or M5)
10	–	M2 and M3 and M4 and M5

Table III.2: Muon stations required by the IsMuon selection as a function of momentum range.

muon identification requirements, by selecting track pairs having invariant mass consistent with the J/ψ, *tagging* one track satisfying tight muon identification requirement, and using the other as *probe*.

The likelihood of the muon hypothesis is evaluated starting from the *average squared distance significance* defined as

$$D^2 = \frac{1}{N} \sum_{i=\mathrm{M2...M5}} \left\{ \left(\frac{x^{(i)}_{\mathrm{closest}} - x^{(i)}_{\mathrm{track}}}{\mathrm{pad}^{(i)}_x} \right)^2 + \left(\frac{y^{(i)}_{\mathrm{closest}} - y^{(i)}_{\mathrm{track}}}{\mathrm{pad}^{(i)}_y} \right)^2 \right\} \qquad \text{(III.5)}$$

where the index i runs over the muon stations containing hits within the FOI, $(x^{(i)}_{\mathrm{closest}}, y^{(i)}_{\mathrm{closest}})$ are the coordinates of the closest hit to the track extrapolation point for each station $(x^{(i)}_{\mathrm{track}}, y^{(i)}_{\mathrm{track}})$, and $\mathrm{pad}^{(i)}_x$ ($\mathrm{pad}^{(i)}_y$) corresponds to half the pad size in the x (y) direction. The total number of stations containing hits within their FOI is denoted by N.

The distribution of D^2 for muons is studied using the tag and probe method with the $J/\psi \to \mu^+\mu^-$ decay, while for the non-muon hypothesis protons from Λ decays are used, since kaons and pions have a muon component due to decays $K^+ \to \mu^+\nu\pi^0$ and $\pi^+ \to \mu^+\nu$ happening before reaching the muon calorimeters. The D^2 distributions obtained from data are shown in Figure III.25.

The likelihood for the muon (or non-muon) hypothesis is defined, for each candidate having average squared distance significance D^2_0 as the integral of the calibrated muon (or proton) D^2 probability density function from 0 to the measured value D^2_0. The likelihood distribution is shown in Figure III.26.

Combining the information on the likelihood with the likelihood from other detectors allows to suppress background due to hadrons properly identified by the RICH detectors. The effect of adding information from other detectors is shown in Figure III.27 where the efficiency and rejection power of the selection are shown for the background hypothesis being represented by pions and kaons.

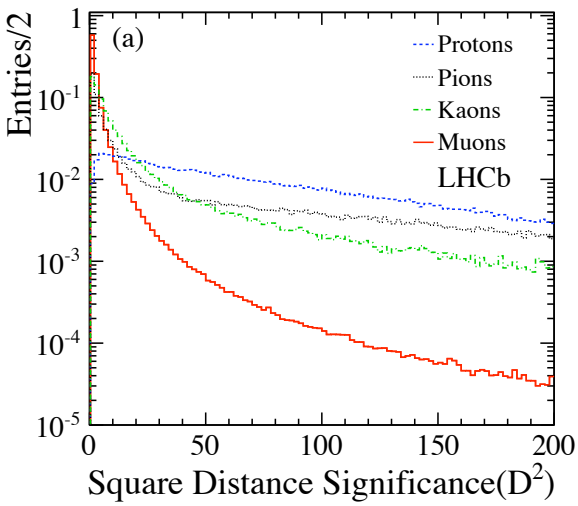

Figure III.25: Average square distance significance distributions for muons, protons, pions and kaons as published in Ref. [117].

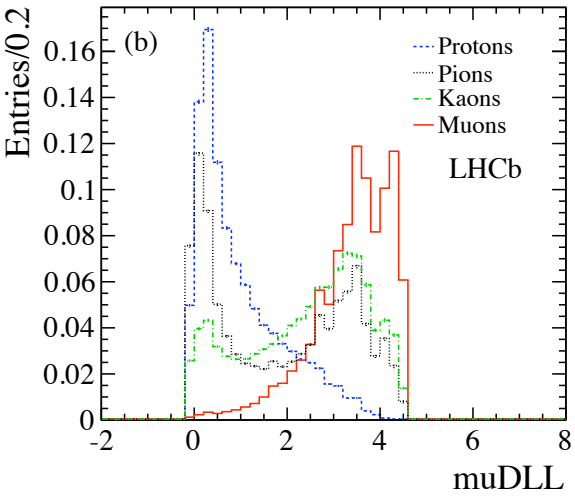

Figure III.26: muDLL likelihood distributions for muons, protons, pions and kaons as published in Ref. [117].

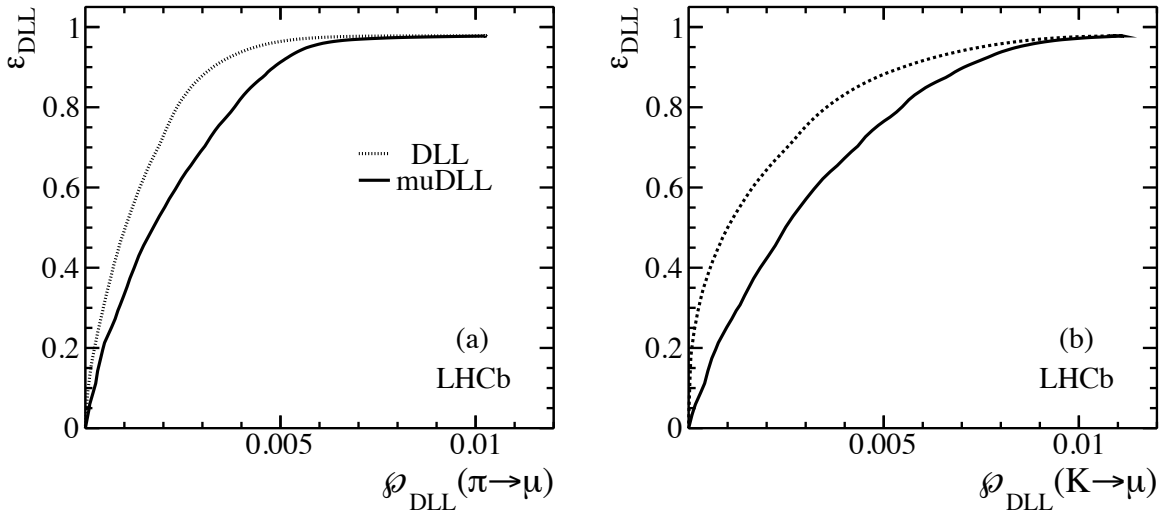

Figure III.27: Average efficiency $\varepsilon_{\mathrm{DLL}}$ as a function of the pion (a) and kaon (b) misidentification probabilities for particles with momentum $p > 3\,\mathrm{GeV}/c$. The dotted lines show the $\mathrm{DLL}_{\mu/\pi}$ performance, while the muDLL performance is shown with a solid line.

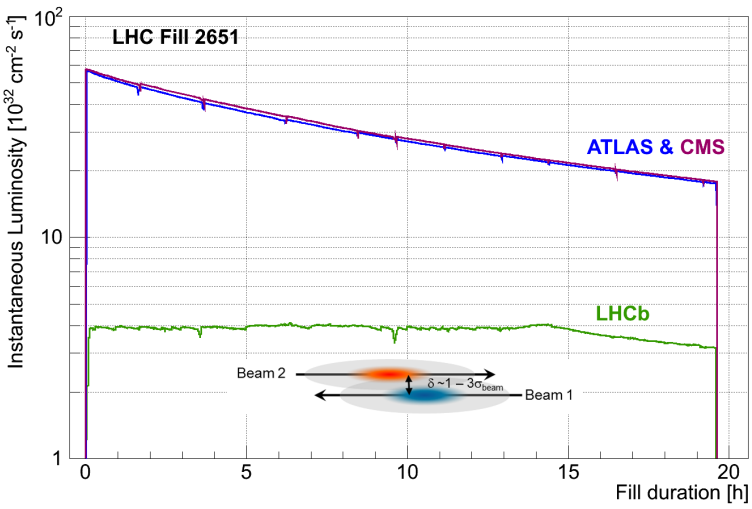

Figure III.28: Instantaneous luminosity delivered at LHCb compared with the instantaneous luminosity delivered at ATLAS and CMS in a typical run (Run 2651 was used). The procedure is named *luminosity leveling*.

III.2.2 The LHCb data-taking

The LHCb experiment runs at a lower luminosity with respect to the other experiments and uses a technique named *luminosity leveling*. An approximately constant constant instantaneous luminosity allows to increase the integrated luminosity and preserving stable trigger conditions during the run. Luminosity leveling is implemented through a misalignment of the beams near to the interaction point followed by a progressive realignment, dynamically driven by the luminosity online monitoring. The process is illustrated in Figure III.28 where the instantaneous luminosity collected by LHCb is compared to the luminosity collected by ATLAS and CMS, in a long fill, when the maximum overlap of the beams is reached, after about 14 hours, the luminosity starts to decrease exponentially with time.

Figure III.29 shows the maximum instantaneous luminosity reached during the data-taking and the average *pile-up* μ, number of visible collisions due to a single bunch-crossing.

The data-taking is divided in three sets according to the year of data-taking, 2010, 2011, and 2012. Table III.3 summarizes the properties of the data samples collected during the *pp* programme of the LHC.

III.3 Trigger strategy and performance

The LHCb trigger is structured in two levels, the Level-0 trigger (L0) and the High Level Trigger (HLT). L0 is said to be a hardware trigger, it is implemented using custom hardware including FPGAs and processes data with a fixed latency of 4 μs. The Level-0

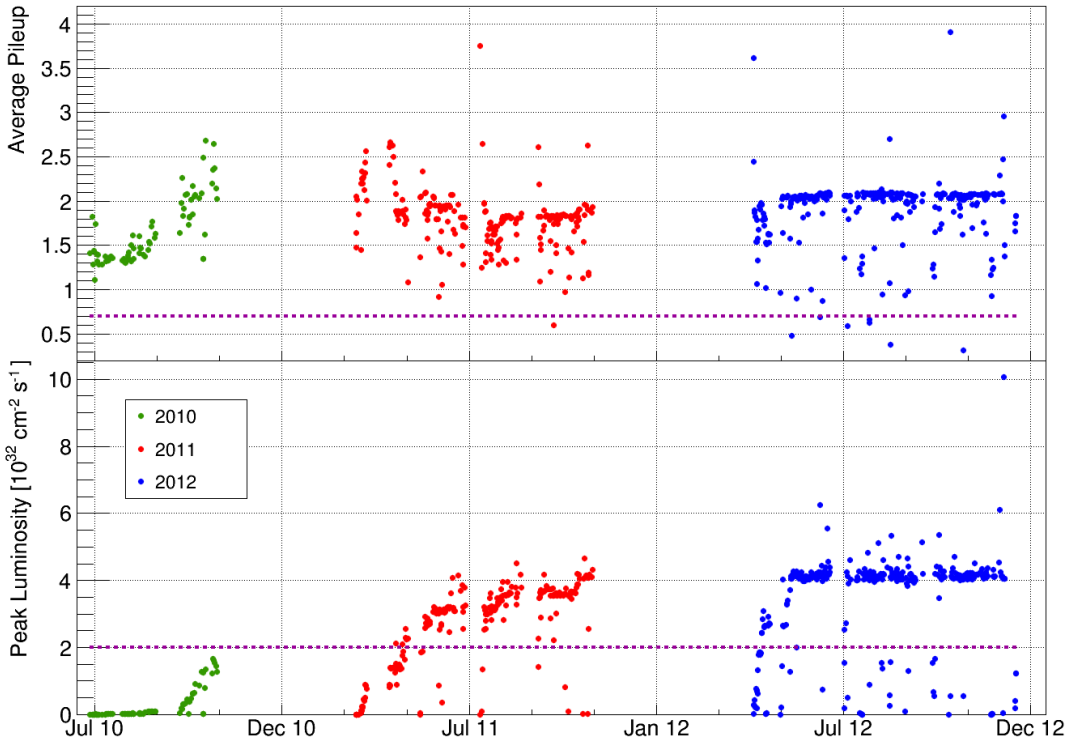

Figure III.29: Operation conditions of the LHCb experiment during the Run-I of LHC.

Year	*Center-of-mass energy*	*Integrated luminosity at LHCb*
2010	7 TeV	0.04 fb^{-1}
2011	7 TeV	1.10 fb^{-1}
2012	8 TeV	2.08 fb^{-1}

Table III.3: Summary of the datasamples collected at LHCb during the main *pp* programme of the LHC.

trigger uses inputs from the calorimeter and muon systems to reduce the rate of bunch crossing with at least one inelastic *pp* collisions down to 1.1 MHz, at which the whole detector can be read out. The maximum rate is imposed by the front-end electronics of the various sub-detectors. The HLT is a software trigger performing a full reconstruction of the event in a massive computer farm. A first partial reconstruction is followed by a first reduction stage, named HLT1, and only events selected are fully reconstructed and selected using the HLT2 algorithms, selecting events to be written on tape. The HLT is implemented taking advantage of the general infrastructure of the LHCb software and therefore it is described in the next chapter.

III.3.1 Level-0 trigger

The L0 trigger is actually composed of three independent triggers. The L0-Calorimeter trigger, the L0-Muon trigger and the L0-PileUp trigger. The last is not used for flavour physics, but only for the determination of the luminosity and will be neglected hereafter.

The *L0-Calorimeter* system uses information from the Scintillator Pad Detector (SPD), PreShower (PS), Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL). Each detector is divided in logical cells which follow the projective geometry of the experiment. The transverse energy E_T is calculated separately for each sub-detector in clusters of 2×2 cells, and is defined as

$$E_T = \sum_{\text{cell}=1}^{4} E_{\text{cell}} \sin\theta_{\text{cell}}, \tag{III.6}$$

where E_{cell} is the energy deposited in the cell and θ_{cell} is the angle between the z-axis and a particle assumed to come from the average interaction point and hitting the center of the cell and the sum runs over the cluster cells.

Using the information on the transverse energy, three independent candidates may be created:

- Hadron candidate (`L0Hadron`): the highest E_T HCAL cluster. In case of alignment between the highest E_T clusters in ECAL and HCAL, the E_T of the hadron candidate is the sum of the HCAL and ECAL clusters.

- Photon candidate (`L0Photon`): the highest E_T ECAL cluster with one or two PS cells hit in front of the ECAL cluster and no hit in the SPD cell corresponding to the PS cells. In the inner zone of the ECAL, an ECAL cluster with 3 or 4 PS cells hits is also accepted as a photon. The E_T of the candidate is the E_T deposited in the ECAL alone.

- Electron candidate (`L0Electron`): same requirements as for a photon candidate, with in addition at least one SPD cell hit in front of the PS cells.

The E_T of the candidates is compared to a fixed threshold and events containing at least one candidate above threshold are retained by L0. Table III.4 reports the thresholds used for the different candidates.

The *L0-Muon* is composed of four processors, one per quadrant, trying to identify the two muon tracks with largest and second largest momentum transverse to the beam axis (p_T) in their quadrant. Track finding is performed by processing units, which combine the data from the five muon stations to form towers pointing towards the interaction region. In the bending plane the search is limited to muons with $p_T > 0.5\,\text{GeV}/c$. The p_T of the muon candidate is evaluated using only the first two muon stations: M1 and M2 assuming that the muon is produced in the interaction region to assess the bending in the magnetic filed and then the transverse momentum. The position of M1, before the calorimeters allows to reduce effect of multiple scattering improving the quality of the

Candidate	*Requirement*	*Threshold* 2011	*Threshold* 2012	*SPD Multiplicity*	*Rate [kHz]*
`L0Hadron`	$E_T >$	3.5 GeV	3.7 GeV	< 600	405
`L0Electron`	$E_T >$	2.5 GeV	3.0 GeV	< 600	165
`L0Photon`	$E_T >$	2.5 GeV	3.0 GeV	< 600	80
`L0Muon`	$p_T >$	1.48 GeV/c	1.76 GeV/c	< 600	340
`L0DiMuon`	$\sqrt{p_T^{\text{largest}} \times p_T^{\text{2nd largest}}} >$	1.296 GeV/c	1.6 GeV/c	< 900	75

Table III.4: Selection of L0 lines and their rates. The table was published in Ref. [118] and updated in Ref. [119].

momentum measurement. The resolution obtained on the muon candidate p_T is around 25%. Two candidates are created, the single muon candidate `L0Muon` and the dimuon candidate `L0DiMuon`. The former requires a minimum value of the largest p_T, while the latter discriminates on the basis of the geometric mean of the largest and second-largest muon transverse momenta. Thresholds are listed in Table III.4.

The trigger is ignored if the event contains too many tracks and is therefore too hard to reconstruct. To avoid selecting too complex events less than SPD 600 hits are required. This requirement is loosen for `L0DiMuon` candidates for which events with up too 900 SPD hits are accepted. While the signal loss is limited, the requirement on the event complexity prevents the software trigger to elaborate events which are too complex, from which it would be hard to extract any useful information, but for which the reconstruction could last hundred times the average, wasting CPU resources and making the HLT latency high and difficult to predict.

Chapter IV Introduction

LHCb physics software and computing model

In the previous chapter, the overview of the LHCb detector and its purpose were described, but the gap between particle tracking and identification, and reliable physics results is large. This chapter is devoted to cover that gap: the LHCb Physics Software and the Computing Model adopted to deliver the service.

IV.1 Overview and dataflow

There are two main ingredients in the formulation of an efficient computing model: data reduction, and data simulation. *Data reduction* is the process of discarding data not containing relevant physics information, using subsequent layers of algorithms with increasing complexity and computation time. Data reduction is unavoidable at the LHC experiments, because the amount of data collected in pp collisions exceeds several hundreds of petabytes making storage impossible with the current technologies, and because applying the full analysis during the data-taking is too unsafe to be acceptable and most of the time technically impossible. *Data simulation* is the process of simulating pp collision events and processing the simulated data through the same work-flow as for real data in order to compare the reconstructed information (mass, decay time, alignments, ...) with the information used in the generation of the event. The latter is named *truth information*.

Data reduction applies also to simulated data, because the samples of real and simulated data have to be as similar as possible. However, events are not really discarded in the simulated samples but *flagged* as kept or discarded. When data reduction processes do not suppress rejected events they are said to run in *flagging mode*. Flagged simulated samples are used to study the selection efficiency during the data reduction process in order to statistically remove (or *unfold*) the effect of the selection from the measured quantities.

In Figure IV.1, the computing logical data flow is sketched. The left arm represents the simulated data flow. An *event generator* simulates the collision of protons, the production and decay of particles. The *generator data* are stored for monitoring and validation purpose, and allow to determine the efficiency of generator-level selections suppressing

particles definitely out of the detector acceptance. Generator data are used as input to the Detector Simulation which simulates the interaction of particles with the geometry of the detector and translates the simulated energy depositions into *hits*, signals on some of the detector electronic channels (digitization). The simulated *detector hits* are used to determine the trigger response to the simulated events. The code used to simulate the response of the software trigger is exactly the same running on real data, but runs in flagging mode and the truth information is saved in the output data, named *Rawmc data*.

The right arm represents the real data processing. The detector hits are read directly from the detector and three subsequent trigger layers are used to implement online data reduction. The first trigger level, L0, is the only hardware level, while the high-level trigger is divided in two software layers named HLT1 and HLT2, running in a large PC farm in the neighborhood of the detector. The data retained by the trigger processing are saved on tape as Raw data which still do not contain the information of tracks and neutral objects, but only the sets of hits and raw information from the trigger and the sub-detectors, named *raw banks*.

Raw data are then spread world wide storing copies in large computing centers connected through the LHC Computing Grid (LCG). The *Reconstruction* runs in parallel on the LCG with different machines running on different events, combining the information from the sub-detectors to produce *track candidates* and *neutral candidates*. Particle Identification information is associated to each candidate in the form of likelihood as discussed in the previous section. The output of the reconstruction is a Full Data Summary Tape, named FULL.DST which is stored in the tape systems of the LCG computing centers.

The pattern recognition algorithms in the reconstruction program make use of calibration and alignment constants stored in two distributed databases storing the detector (geometrical) properties and data-taking conditions, separately. The reconstruction process is repeated about yearly on all the collected data samples using updated reconstruction techniques and improved alignment constants obtained in the continuous improvement of the understanding of the detector.

Finally, reconstructed data are made available to physics analysts through a last data reduction process, named *Stripping*, necessary to filter data relevant for the analyses from tape to disk. During the Stripping physics candidates are created combining tracks and/or neutral objects. For example, two muon candidates are combined to create a J/ψ candidate, which can be combined to a pion candidate to create a $B_c^+ \to J/\psi\,\pi^+$ decay candidate. The output of the Stripping procedure can be either a smaller DST storing the candidates created during the Stripping along with the full reconstructed event, or a so-called microDST which stores only the candidates identified by the Stripping algorithms. Simulated datasets are usually stripped in flagging mode, and the truth information is saved alongside with the reconstructed candidates in either DST and microDST formats. Real data DST output is organized in streams, containing events sharing some experimental signature. This is the chosen trade-off between the need for a fast random access to candidates selected by a given Stripping algorithm, and the reduction of storage wasted by doubling the copies of the same event selected by different Stripping algorithms.

The Stripping output is accessed by the physics analysis program run by the analysts

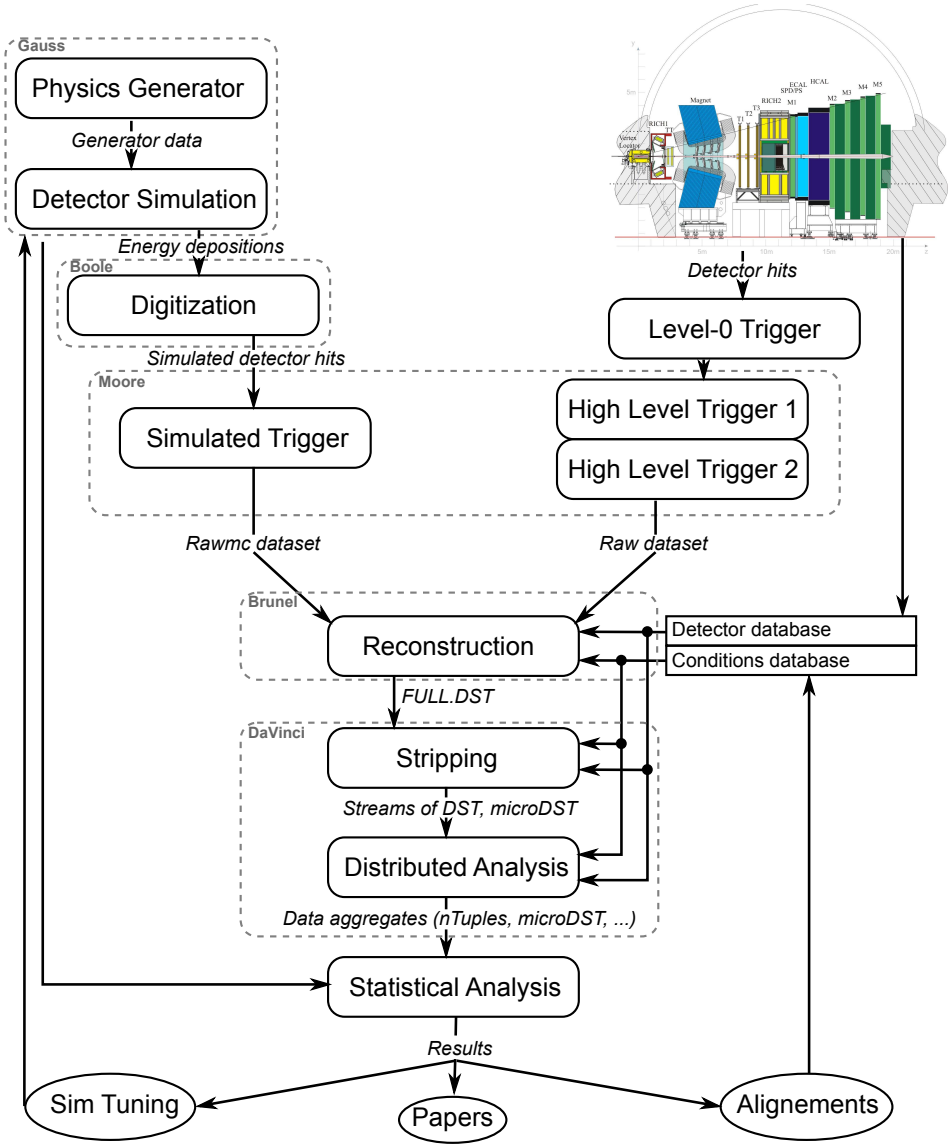

Figure IV.1: The LHCb computing logical dataflow model. Rounded solid boxes represent logical steps in the dataflow, italic labels are data or data formats, and dashed boxes represent different projects in which the LHCb software is organized. See text for detail.

on the LCG nodes creating small files with only the relevant information used to perform statistical analyses as density estimations, fits, and so on. The output of the analysis program is usually few tens of GB and can be easily stored on modern laptops and personal computers.

A more technical description of the LHCb Computing Model, with particular care for the needs in terms of infrastructure (disk, tape, network...), is available in Ref. [120].

IV.2 Physics applications

What is needed to do physics analysis is a set of programs to simulate, reconstruct and visualize events, and a set of tools to ease the writing of analysis programs. The LHCb Collaboration developed an architecture named *Gaudi*, now in use also at ATLAS under the name of Athena, implementing all the parts of an analysis which can be split in single-event blocks, or in other words problems where each *pp* collision is assumed to be uncorrelated to the previous and the next ones (except from effects due to dead-times, treated statistically). In Figure IV.1, all the blocks enclosed in gray dashed boxes are implemented using the Gaudi architecture. As it can be seen, most of the physics applications at LHCb are implemented under this common framework with the notable exceptions of the Level-0 trigger which is implemented using FPGAs which are not compatible with Gaudi, and the Statistical Analysis, including modeling of the whole dataset to which the single-event paradigm does not apply, and for which every analyst has different favorite approach, the most common being ROOT and PyROOT, but Matlab or even Office suites are sometime used.

One of the great advantages of having a common architecture for the physics is that the parallelization of the code on an event basis can be common to all the programs. Having defined an input-output structure which does not mixes events the workload can be easily split on many computing nodes in a Computing Center, or on the LCG, in exactly the same way for the trigger, the simulation, the reconstruction and the signal selection purposes. Technicalities of parallelization through the heterogeneous systems composing the LHCb computing resources are therefore moved to the development and maintenance of the common architecture preserving the energies that physicists spend on physics code.

The applications based on the Gaudi framework are:

- *Gauss* – Simulation program which interfaces Monte Carlo generators as Pythia [29] or BcVegPy [28] with theEvtGen package [121] describing the decays of the produced particles, and the Geant4 toolkit [122, 123], describing the interaction of simulated particles with the detector. The output of Gauss are `sim` files encoding the energy depositions of particles in the sub-detector volumes;

- *Boole* – Digitization program taking as input the `sim` files and translating them to detector hits;

- *Moore* – Trigger program including the code for online selection at High-Level and the fast event reconstruction performed between HLT1 and HLT2. Moore also includes the code used to emulate the Level-0 trigger, running on FPGA for real data, on simulated samples;

- *Brunel* – Reconstruction program creating physics objects (tracks, photon, muon candidates) and performing the refined calculations of the associated quantities (PID likelihoods, track quality, ...);

- *DaVinci* – Physics Analysis or Signal Selection program creating candidates for specific particle decays (like $J/\psi \to \mu^+\mu^-$) and saving the output in a format useful for statistical analyses (histograms, nTuples, likelihoods). DaVinci is also used for the Stripping selection, in which case an output format readable by DaVinci itself is used.

IV.2.1 The architecture: Gaudi

When moving from sequential languages, as Fortran or C, to languages implementing excellent support to objects, like C++ or Java, it is tempting to describe data with complex objects containing the code to calculate derivate quantities. For example, a track object could have a constructor that performs the fit of the track whenever the track object is initialized. This is the classical *Object-Oriented Programming* which has been found to expose the development of large-scale physics program to at least two severe issues. First, objects for which multiple reconstruction algorithms are reasonable should have a different interface for each algorithm, to pass configuration parameters or additional information needed from the algorithm but which have nothing to see with the object itself. Back to the example of the track, one would have a different track object for each algorithm used to reconstruct the track, every object would have a different interface making maintenance and generic coding very hard tasks. Besides the data has to be stored somewhere, and temporary steps in the definition of data might be of interest, while the choice and the configuration of algorithms is usually defined at the launch or compile-time and does not require storage. Mixing algorithms and data makes the management of the memory flow extremely challenging, and nowadays disk and RAM memories are the most expensive computing resources whose waste should be possibly avoided. These and other considerations made the Physics Application communities to stick to the so-called *Data-Oriented Programming* where data objects are containers which are filled by independent algorithm objects. Fast and efficient memory management is then possible for data objects, while optimized scheduling can be provided for algorithms.

Gaudi implements the data flow through the concept of Transient Event Store, a set of containers for all the temporary objects in event processing, including track parameters, vertex quality, PID likelihoods and so on. In Figure IV.2 the difference between the real data flow from and to the TES, and the apparent data flow following the algorithm processing is shown. An important aspect of this technique is the caching of results of algorithms already used in the same event. If two algorithms creating candidates for $B_c^+ \to J/\psi\pi^+$ and $B_c^+ \to J/\psi\mu^+\nu$ decays are used, only the first accessing the TES of reconstructed J/ψ in the event will actually trigger the algorithms for the construction of the J/ψ candidates, while the latter will rely on the candidates already stored in the TES. On the other hand if no algorithm tries to access the TES to read J/ψ candidates, then, obviously, the J/ψ candidates are not produced nor stored. Still, Gaudi enforces a user defined scheduling which is essential, for example, to compare a result before and after the application of a calibration. This mixing of on-demand and user-defined scheduling makes Gaudi extremely flexible preserving high performance, necessary (but not sufficient)

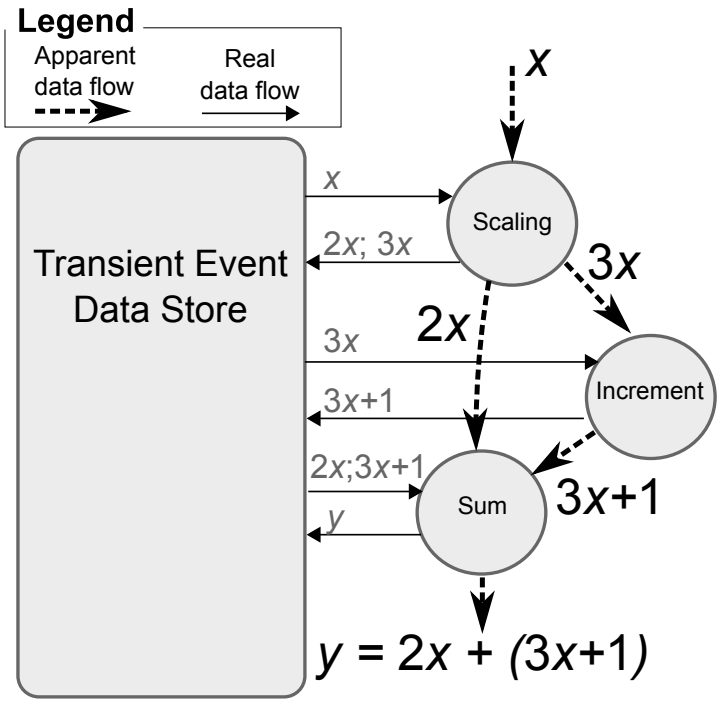

Figure IV.2: Trivial example of usage of the Transient Event Store (TES) to calculate the expression $y = 2x + (3x + 1)$ using the three algorithms *scaling* (returning the double and the triple of the input), *increment* (returning the input incremented by one), and *sum* (returning the sum of the inputs). The difference between the apparent data flow, following the execution of the algorithms, and the real data flow from and to the TES is shown. In the real applications, the mathematical operations are substituted by reconstruction algorithms.

condition to satisfy one of the key requirements in its development.

More details on the Gaudi framework can be found in Ref. [124].

IV.2.2 High Level Trigger

As mentioned in the previous chapter, the LHCb trigger is composed of two levels, the Level-0 trigger (L0) and the High Level Trigger (HLT). L0 is a hardware trigger described in Section III.3.1, while the HLT consists of a software application, *Moore*, that runs on a PC farm close to the detector. The HLT reduces the rate of accepted events to few kHz, and all such events are written to storage as Raw data. An event that is accepted by Level-0 trigger is assembled, merging data from the different sub detectors by an event builder program running on a core of each multi-core node of the trigger PC farm [113].

The event filter process is then divided in two stages. The first stage, HLT1, relies on a partial event reconstruction to perform an inclusive selection of interesting signatures. At the reduced rate of 40/80 kHz (in 2011/2012), the full event reconstruction becomes feasible with minor adjustments to speed up the code with respect to the offline reconstruction sequence. Relying on this reconstruction, a set of inclusive and exclusive selections reduces

the trigger rate to 3/5 kHz, which are stored on tape for further offline analysis. The improvement in the output bandwidth achieved in 2012 run, made possible by the extension of the offline computing capabilities, has allowed to widen the physics program of the LHCb experiment.

The HLT1 reconstruction is a simplified version of the offline tracking algorithm with PID information only coming from the muon system. The candidates created by the HLT1 selection are

- `Hlt1TrackAllL0` selects events with good quality track candidates with high p_T ($p_T > 1.6\,\text{GeV}$) and displaced with respect to the primary vertex;
- `Hlt1TrackPhoton` is the same as `Hlt1TrackAllL0`, but adding the requirement of a `L0Photon` or `L0Electron` candidate at Level-0, and relaxing the p_T and track quality requirements;
- `Hlt1TrackMuon` is similar to `Hlt1TrackAllL0`, but with the additional requirement of matching of the track with hits in the muon chambers;
- `Hlt1SingleMuonHighPT` allows to remove any requirement on the separation of the track from the primary vertex when the muon candidate p_T exceed a tighter threshold;
- `Hlt2DiMuonHighMass` selects dimuon candidates consistent with quarkonia resonances $m_{\mu\mu} > 2.5\,\text{GeV}/c^2$ without requirements on the displacement from the primary vertex;
- `Hlt1DiMuonLowMass` selects dimuon candidates without mass restrictions, but requiring detachment of the dimuon vertex from the primary vertex.

The requirements are summarized in Table IV.1 and Table IV.2 for muon and non-muon lines, respectively.

The second level software trigger (HLT2) relies on a full event reconstruction for all tracks with a p_T of at least 300 MeV/c. Several exclusive and inclusive selections are performed at this level, too many to be discussed here, where the discussion is limited to muon trigger lines which are relevant to decays with a dimuon in the final state. Further detail can be found in Ref. [113, 118, 119].

Both single muon and dimuon selections are implemented using the same muon identification procedure to the one used in offline analyses and described in Section III.2.1. The trigger policy consists again in selecting events with high p_T or well detached muon candidates. Dedicated trigger lines are devoted to the selection of the J/ψ and $\psi(2S)$ resonances, useful to a large amount of LHCb analyses.

The candidates created by the muon HLT2 algorithms are:

- `Hlt2SingleMuon` selects events having a single high p_T muon, for which HLT1 has found a `Hlt1TrackMuon` candidate, and good track quality and detachement from the primary vertex;

HLT1 Line	TrackMuon	SingleMuon HighPT	DiMuon HighMass	DiMuon LowMass
Track IP [mm]	> 0.1	–	–	–
Track IP χ^2	> 16	–	–	> 3
Track p_T [GeV/c]	> 1	> 4.8	> 0.5	> 0.5
Track p [GeV/c]	> 8	> 8	> 6	> 6
Track χ^2/ndf	< 2	< 4	< 4	< 4
DOCA [mm]	–	–	< 0.2	< 0.2
χ^2_{vertex}	–	–	< 25	< 25
Mass [GeV/c^2]	–	–	> 2.6	> 1

Table IV.1: Summary of the requirements applied to the muon candidates as created in the HLT1 selection algorithms. Requirements are on the impact parameter of the track (Track IP), the difference between the primary vertex fit including or excluding the muon track candidate (Track IPχ^2), the momentum and transverse momentum of the muon track candidate (Track p and p_T), the quality of the track measured as a χ^2 of the fitted track with respect to the detector hits (Track χ^2), the Distance Of Closest Approach (DOCA) between the muon track candidates forming a dimuon candidate, the quality of the fit to the common vertex of the muon pair tracks (χ^2_{vertex}), and the invariant mass of the dimuon candidate (Mass).

HLT1 Line	Hlt1TrackAllL0	Hlt1TrackPhoton
Track IP [mm]	> 0.1	> 0.1
Number of VELO hits/track	> 9	> 6
Number of missed VELO hits/track	< 3	< 3
Number OT+IT$\times$2 hits/track	> 16	> 15
Track IPχ^2	> 16	> 16
Track p_T [GeV/c]	> 1.7	> 1.2
Track p [GeV/c]	> 10	> 6
Track χ^2/ndf	< 2.5	< 2.5

Table IV.2: Summary of the requirements applied to the non-muon candidates as created in the HLT1 selection algorithms. Requirements include criteria on the quantities defined in the caption of Table IV.1 plus criteria based on the number of hits in the tracker per track composing the candidate considering hits in the Vertex Locator (VELO hits/track), in the downstream tracker (OT+IT$\times$2), and expected but not found hits in the VELO (missed VELO hits/track).

- `Hlt2SingleMuonHighPT` renounces to all the above criteria except for the muon identification and p_T requirements, and the latter is tightened;
- `Hlt2DiMuonJpsi` and `Hlt2DiMuonPsi2S` select J/ψ and $\psi(2S)$ candidates with minimal requirements, but saving only 20% and 10% of the selected events chosen randomly (pre-scale).

HLT2 line	Hlt2SingleMuon	Hlt2SingleHighPT
Hlt1TrackMuon	Triggered-On-Signal	–
Track IP	> 0.5	–
Track IP χ^2	> 200	–
Track p_T [GeV/c^2]	> 1.3	> 10
Track χ^2/ndf	< 2	–
Pre-scale	0.5	1.

Table IV.3: Summary of the HLT2 algorithms selecting candidates with a single muon.

- Hlt2DiMuonJpsiHighPT and Hlt2DiMuonPsi2SHighPT select J/ψ and $\psi(2S)$ candidates with tighter p_T requirements and without pre-scale;
- Hlt2DiMuonDetached selects dimuon candidates with invariant mass down to 1 GeV/c^2, requiring detachement from the primary vertex both in terms of fitted flight distance and of impact parameter of the daughter muon tracks;
- Hlt2DiMuonDetachedHeavy selects dimuon candidates with higher invariant mass with respect to Hlt2DiMuonDetached, but removing criteria on the impact parameter of the daughter tracks and relaxing the requirement on the fitted flight distance;
- Hlt2DiMuonDetachedJPsi selects J/ψ candidates adding a requirement on the invariant mass of the dimuon to be consistent with the mass of a J/ψ resonance, and relaxing further the flight distance requirements with respect to the Hlt2DiMuonDetachedHeavy candidate.
- Hlt2DiMuonB selects candidates for $B^0_{u,d} \to \mu^+\mu^-$ requiring high invariant mass of the muon pair.

Tables IV.3, IV.4, IV.5 summarize the requirements to create these candidates.

Events containing at least a selected candidate are saved on tape.

Deferred trigger

The LHC delivers stable beams for nearly 30% of the time, using the rest for injection or tuning. In order to avoid the waste of CPU resources during the idle time, the hard disks of the Event Filter Farm nodes are used to save about 20% of the L0-accepted events. The events are then analysed during the inter-fill gaps [125].

This approach, named *Deferred trigger* has allowed to improve the HLT reconstruction by loosening the p_T threshold on the reconstructible tracks from 500 MeV/c in 2011 to 300 MeV/c in 2012 and by implementing a specialized tracking algorithm for the K^0_S decays.

HLT2 line	Hlt2DiMuon JPsi	Hlt2DiMuon Psi2S	Hlt2DiMuon JPsiHighPT	Hlt2DiMuon Psi2SHighPT	Hlt2DiMuon B
Track χ^2/ndf	< 5	< 5	< 5	< 5	< 5
Mass [GeV/c^2]	$M_{J/\psi} \pm 0.12$	$M_{\psi(2S)} \pm 0.12$	$M_{J/\psi} \pm 0.12$	$M_{J/\psi} \pm 0.12$	> 4.7
χ^2_{vertex}	< 25	< 25	< 25	< 25	< 10
$p_T^{\mu\mu}$ [GeV/c]	–	–	> 2	> 3.5	–
Pre-scale	0.2	0.1	1.0	1.0	1.0

Table IV.4: Summary of the HLT2 algorithms selecting dimuon candidates without detachment requirements.

HLT2 line	Hlt2DiMuon Detached	Hlt2DiMuon DetachedHeavy	Hlt2DiMuon DetachedJPsi
Track χ^2/ndf	< 5	< 5	< 5
Track IPχ^2	> 9	–	–
Mass [GeV/c^2]	> 1	> 2.95	$M_{J/\psi} \pm 0.12$
FDχ^2	> 49	> 25	> 9
χ^2_{vetex}	< 25	< 25	< 25
$p_T^{\mu\mu}$	> 1.5	–	–

Table IV.5: Summary of the HLT2 algorithms requiring a detached dimuon candidate.

Performance

The trigger performance with respect to a defined process can be evaluated using a data-driven method comparing the number of events triggered because of that process and the number of events triggered independently of that process.

The two categories are named

- *TOS* (Triggered On Signal): the signal under study is *sufficient* to trigger the event;
- *TIS* (Triggered Independent of Signal): the event would also have been triggered without the signal under study.

Clearly, an event can be TIS and TOS simultaneously (TIS&TOS), or neither TIS nor TOS. The LHCb trigger system records all the information needed for such a classification. The trigger efficiency for selecting an event independently of the signal particles is

$$\varepsilon^{\text{TIS}} = \frac{N^{\text{TIS}}}{N^{\text{SEL}}}, \tag{IV.1}$$

where N^{SEL} is the number of events that would be selected offline if no trigger requirements were applied. This information is clearly available only in simulated samples. If the

efficiency of selecting an event independently on the considered signal is independent of the criteria used to select the signal, then one can assess ε^{TIS} in real data samples,

$$\varepsilon^{\text{TIS}} = \frac{N^{\text{TIS\&TOS}}}{N^{\text{TOS}}}. \tag{IV.2}$$

The total trigger efficiency can therefore be computed as

$$\varepsilon^{\text{TRIG}} = \frac{N^{\text{TRIG}}}{N^{\text{SEL}}} = \varepsilon^{\text{TIS}} \times \frac{N^{\text{TRIG}}}{N^{\text{TIS}}} = \frac{N^{\text{TIS\&TOS}}}{N^{\text{TOS}}} \times \frac{N^{\text{TRIG}}}{N^{\text{TIS}}}. \tag{IV.3}$$

Clearly, if an analysis relies on TOS events only, as it is often the case, then the TOS requirement becomes part of the offline selection and Equation IV.3 simplifies in

$$\varepsilon^{\text{TRIG}} = \frac{N^{\text{TIS\&TOS}}}{N^{\text{TIS}}}. \tag{IV.4}$$

Using this definition it is possible to measure the trigger line efficiency using real data, thus without relying on the accuracy of the simulation. To study the efficiency of the muon and dimuon trigger lines the $B^+ \to J/\psi K^+$ channel is often used. Figure IV.3 shows the efficiencies of few trigger algorithms selecting $B^+ \to J/\psi K^+$ as a function of the p_T of the B^+ meson trigger.

IV.2.3 Reconstruction: vertex fit and decay time measurement

The step following the online or trigger selection in the logical data flow of the LHCb experiment is the event reconstruction, the offline process of finding tracks and neutral objects and associating PID likelihoods to each of them.

Brunel applies a pattern recognition algorithm to reconstruct tracks and to determine the position of the primary vertices. The primary vertices may be refitted at a later stage using the DaVinci application to add further constraints or to test how the fit quality changes when a subset of the tracks associated to a primary vertex are assumed to be produced in the decay of a flying particle, in a *secondary vertex*. Even if refitting may be needed, it is crucial to have a preliminary estimation of the number and position of the primary vertices common to all the analyses. Analyses which uses the primary vertex to reject background rather than for precision measurement of the flight distance, may use the preliminary estimation, which is actually quite good, and save computing resources.

The following pages describe briefly the preliminary primary vertex fit and then move to another part of the event reconstruction, covered by the DaVinci program the reconstruction of decay candidates. The steps to define a decay candidate are: the selection of the decay daughters, the fit of the decay vertex, the assessment of consistency with the decay hypothesis using additional constraints as the mass of the mother particle and the position of primary vertex, and the re-evaluation of properties such as the decay time exploiting the additional constraints.

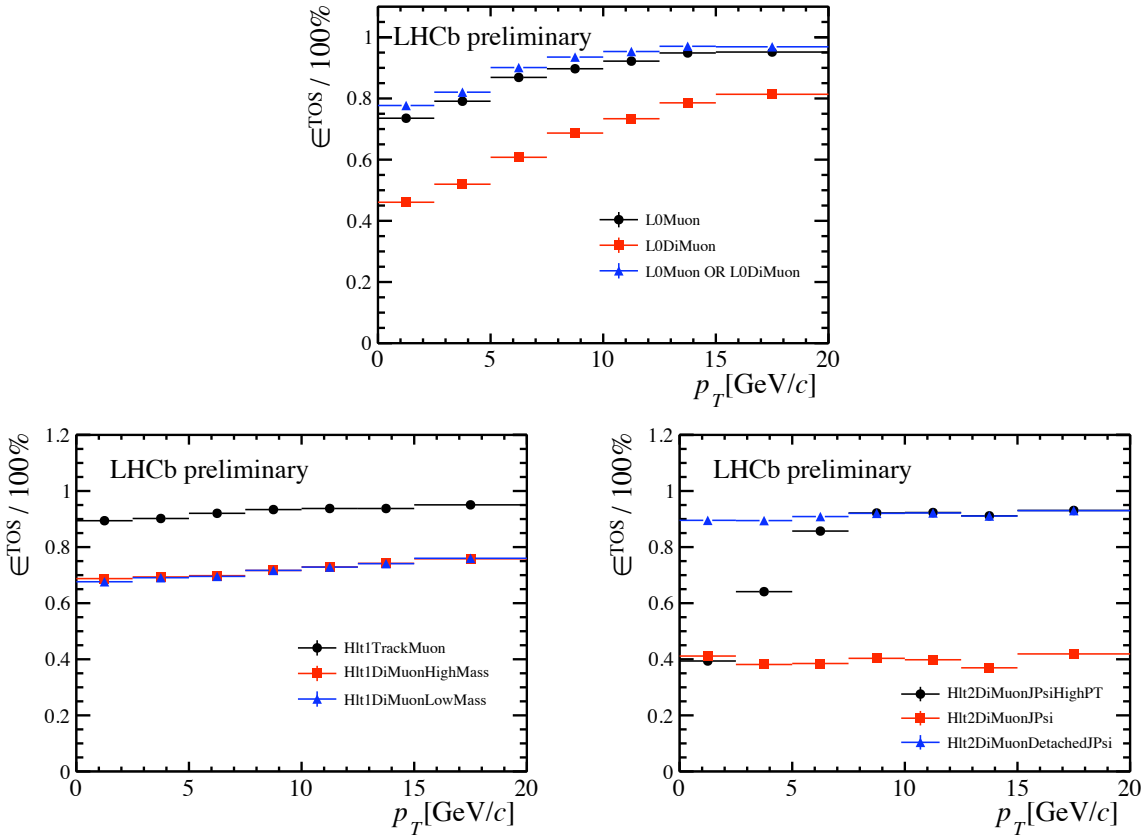

Figure IV.3: Muon Trigger efficiencies considering the $B^+ \to J/\psi K^+$ decays and the algorithms implemented in the Level-0 trigger (top), and High Level Trigger (bottom). The latter is split into two level HLT1 and HLT2. On the horizontal axis the, p_T of the B^+ meson is shown.

Primary vertex reconstruction

The initial estimate of the primary vertex z coordinate is based on the seeding procedure employing the method of analytical clusterization. The cluster is defined by the z^{clu} coordinate and its uncertainty σ_z^{clu}. The procedure starts with initial clusters determined by the closest approaches of the tracks to the z axis. At the first step, clusters are composed of only one track, then they are merged if the condition

$$D^{\text{pair}} = \frac{|z^{\text{clu}_1} + z^{\text{clu}_2}|}{\sqrt{(\sigma_z^{\text{clu}_1})^2 + (\sigma_z^{\text{clu}_2})}} < 5 \tag{IV.5}$$

is satisfied, and the new cluster position defined by the weighted average

$$z^{\text{clu}} = \frac{\frac{z^{\text{clu}_1}}{(\sigma_z^{\text{clu}_1})^2} + \frac{z^{\text{clu}_2}}{(\sigma_z^{\text{clu}_2})^2}}{\frac{1}{(\sigma_z^{\text{clu}_1})^2} + \frac{1}{(\sigma_z^{\text{clu}_2})^2}} \qquad \text{and} \qquad \sigma_z^{\text{clu}} = \sqrt{\frac{1}{\frac{1}{(\sigma_z^{\text{clu}_1})^2} + \frac{1}{(\sigma_z^{\text{clu}_2})^2}}} \tag{IV.6}$$

Finally clusters with a final multiplicity (number of tracks) below 6 are removed. Additional quality conditions are then applied to reduce the number of false clusters. False seeds are usually due to clusters counting few tracks reconstructed with small uncertainty and pointing close to a high-multiplicity primary vertex.

For each primary vertex seed, the fitting procedure is performed in decreasing order of the seed multiplicity. The long, upstream and VELO tracks are used in the reconstruction. All tracks are extrapolated to the position of the seed using the information of their curvature due to the residual magnetic field if available (long and upstream tracks).

The position of the primary vertex is then determined minimizing the function

$$\chi^2_{PV} = \sum_{i=1}^{n_{\text{tracks}}} \frac{d^2_{0i}}{\sigma^2_{d_{0i}}} \tag{IV.7}$$

where d_{0i} denotes the impact parameter of the i-th track and σ_{d0} its uncertainty. The fit is repeated discarding at each iteration the track with the largest impact parameter significance (IPS) defined as d_0/σ_{d_0}, until all tracks satisfy the requirement IPS < 4.

Primary vertices with a final multiplicity smaller than 6 are removed.

The whole procedure is then repeated for the next seed excluding all the tracks already used in the reconstruction of other primary vertices.

Selection of the decay candidates

Only the so-called *basic* particles, having a lifetime long enough to have a flight distance large and comparable to the detector size, are reconstructed during the reconstruction process and stored in the FULL.DST datasets. A notable exception is the π^0 particle which is short lived but its products, two collinear photons, cross the detector and are hardly distinguishable from a single photon. This "diphoton" is therefore considered a basic particle and so is the pion. From the basic particles other common particles are created using the on-demand scheduling described above. Common particles such as J/ψ resonances, K^0_S, and so on, are reconstructed whenever at least one algorithm tries to access the container of candidates for that particles in a given event. The reconstruction usually applies an extremely loose selection which can be refined later through a filter process, extremely less expensive than the reconstruction process itself.

The reconstruction of a given signal decay is performed through a loop on all the containers assigned for the daughter particles. If a daughter particle is a basic particle, like a pion or a kaon, then the container contains the tracks and their PID. The kaon container will contain only tracks whose kaon PID is higher than a loose threshold. If the particle is not a basic particle then the container contains (or is filled with) the candidates constructed in a previous step following the same procedure described here.

Beside the requirement defining whether a particle should appear in a container or not, finely tuned selection based on the momentum, detachment from PV, and on the PID can be applied just before the construction of the new decay candidate.

Once chosen the decay daughters, requirements can be applied on their combination. Usually a minimum total transverse momentum and a reasonable range for the invariant

mass are required at this stage. Then the vertex fit, described in the next paragraph, is performed, where the parameters of the tracks, including their momentum, are recomputed but not overwritten (to allow alternative fits to be performed) under the constraint of a common vertex. A final selection is then applied to the vertex quality and to the new values for the mass and momenta of the mother candidate.

Daughter candidates are classified into the following categories:

- *Flying particles* are particles with a lifetime sufficiently long to travel a measurable distance before their decay. Basic charged particles, neutral and charged B and D mesons are in this category.

- *Vertexed particles* are resonances decaying to at least two flying particles or to at least a vertexed particle. Examples are the $J/\psi \to \mu^+\mu^-$ and $D^{*+} \to D^0\pi^+$ resonances decaying to flying particles, but also $\psi(2S) \to J/\psi\,\gamma$ which decays to another vertexed particle.

- *Non-Vertexed particles* are resonances decaying to one flying particle, for example $D^{*0} \to D^0\gamma$. The determination of the vertex relies on the particles produced together with the mother particle (the D^{*0} in the example).

- *Photons* cannot be used to create vertices because the angular resolution of the calorimeter is not sufficient to resolve the position of the originating vertex. Photons converted to e^+e^- pairs have better resolution, but still not sufficient to constrain significantly the vertex position.

Decay Vertex Fit

There are several techniques used at LHCb to fit the decay vertices. The main difference is between online and offline fitters. The former are designed to be fast and to spend a predictable amount of time to perform the calculation, the offline fitters are focused on precision and assessment of the fit quality and related uncertainties.

The fitter used for the analysis presented here was the default and only stable fitter at the time the analysis was performed and is named *OfflineVertexFitter*.

As the name suggests, it is a fitter used to perform secondary vertex fits in the offline analysis running in DaVinci. The fitter takes as input a list of particles, performs the fit and updates the mother particle with the recalculated values. The input particles are not modified and they can be safely reused to perform other vertex fits with different hypotheses.

The basic idea of the fitter is to factorize the calculation into five types of action:

- *Seeding*: the determination of a vertexed particle or a set of at least two flying particles. This is the first step of any fit. Failed the seeding, the fit fails.

- *Adding a Vertexed particle*: if there are more than one Vertexed particle, then the vertices are merged to the vertex of the seed vertex.

- *Adding a flying particle*: flying particles are added to the seed vertex one by one.
- *Adding a Photon*: after all the vertexed and flying particles are added to the seed vertex, the seed vertex is assumed to be the production vertex of the photon. Its momentum is recalculated using this new constrain and the momentum of the mother particle is updated.
- *Adding a Diphoton*: same as the previous action, but taking into account a possible constraint on the diphoton mass.

Finally, the χ^2 of the vertex is minimized and is stored in the vertex object as additional information, together with the number of degrees of freedom of that particular fit.

Further documentation on the OfflineVertexFitter is available in Ref. [126].

Decay Tree Fitter

Once the decay vertex fit has been performed for every vertex in the decay chain, a global fit of the decay tree can be performed using the Kalman Filter technique [127] ported to particle physics for the Babar experiment and recently integrated into the LHCb software [128].

The Kalman Filter is a recursive or progressive fit technique which is used primarily in technological applications concerning automatic navigation, where each step exploits a new measurement of the position which allows to correct the previous determination of the state of the navigated object, namely its position and its speed. In particle physics the Kalman Filter has been used for many years for the tracking algorithms, where each tracking station constitutes a further measurement and allows to redetermine the track parameters.

Only recently it has been observed that the Kalman Filter can be applied to vertex fitting using the constraints on the vertex, as the mass of the decaying particle, the parameters of the daughter particles, or the position of the production vertex, as subsequent steps in the determination of the vertex parameters (position and covariant matrix). The advantage of this technique with respect to the standard least-square technique is that it allows to split the problem in layers reducing the dimension of the matrices involved in the calculation. This translates into an improved applicability to problems involving a large set of parameters.

Indeed, the least-square fit is limited in this case by the need for inversion of large matrices. If n is the number of parameters in the fit, a $n \times n$ matrix needs to be inverted at each fit iteration with a computational cost which scales as $n^2 \log n$ (or worse).

Hence the Kalman Filter approach is preferred in fits with many parameters as those performed on long decay chains containing several constraints coming from the decay nature (masses) and from measurements (momenta and positions).

The Decay Tree Fitter stores a global χ^2 value which combines the χ^2 of the single vertices taking properly into account the correlation between the many uncertainties. The algorithm also computes the number of degrees of freedom, which is not always a trivial quantity when the complexity of the decay chain increases.

Refit of the Primary Vertex

The primary vertices are reconstructed in *Brunel* once for all the possible candidates, but once the secondary vertices are identified and the tracks are associated, using the procedure described above, the fit of the primary vertex should be repeated using only the tracks not associated to a secondary vertex.

Detailed studies have shown, however, that the bias introduced in the position of the primary vertex because of possible misidentification of secondary tracks as prompt is negligible with respect to the uncertainty on the position of the secondary vertex.

The subsequent step is the association of a decay candidate to the primary vertex which, most likely, has produced it. The association can fail for several reasons: the primary vertex could be reconstructed as a secondary vertex, or not reconstructed at all and hence ignored in the association. It is also possible that a different primary vertex than the one which has generated the candidate decay, is found to have better consistency with it than the one which has generated the event. Luckily, the length of the interaction zone, where the primary vertices distribute is of the order of 10 cm, much larger than the typical decay length (in the experiment frame) of b- and c-hadrons, usually shorter than 1 cm. The effect of wrong association of the primary vertex is thus small and easy to model and control with the low pileup at which LHCb operates.

Determination of the decay time for fully reconstructed decays

Once the position of the primary and secondary vertices is determined, the decay time can be assessed combining their distance with the momentum of the mother particle candidate.

For fully reconstructed decay modes, where the pointing can be correctly assumed, the decay time is obtained with a fit constraining the momentum of the combination to point back towards the primary vertex.

An example of this procedure can be found in Ref. [129]. The decay time in the proper frame of the decaying particle $t_{\rm fit}$ (named in the following *fit decay time*) is obtained minimizing

$$\chi^2(t_{\rm fit}) = \mathcal{R}^T W_O \mathcal{R} \tag{IV.8}$$

where

$$\mathcal{R} = \begin{pmatrix} \mathbf{v} - \tilde{\mathbf{v}} \\ \mathbf{p} - \tilde{\mathbf{p}} \\ \mathbf{x} - (\tilde{\mathbf{v}} - \tilde{t_{\rm fit}}\frac{\tilde{\mathbf{p}}}{m}) \end{pmatrix} \qquad \text{and} \qquad \mathcal{W}_O = \begin{pmatrix} W_v & W_{v,p} & 0 \\ W_{v,p} & W_p & 0 \\ 0 & 0 & W_x \end{pmatrix} \tag{IV.9}$$

with $\mathbf{v}$ and $\mathbf{x}$ measured position of the decay and production vertices, $\mathbf{p}$ the particle momentum, $\tilde{\mathbf{v}}$, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{p}}$ their estimators with the constraint $\mathbf{v} = \mathbf{x} - t_{\rm fit}\frac{\mathbf{p}}{m}$ and m the particle mass. W_O is the covariant matrix. The event-by-event error is estimated from the second derivative of the χ^2 function with respect to $t_{\rm fit}$ at minimum.

Considerations on the partial reconstruction

The partial reconstruction of a decay is the determination of the properties of the decay using only a subset of the daughter particles. There are several reasons that can lead to the need of partial reconstruction, typically it allows to avoid the measurement of particles whose detection efficiency is low. The missing particle is often a neutral particle, a photon, a π^0 or a neutrino for which the tracker gives no information, but it can also be a low-momentum particle which is often out-of-acceptance because of the magnetic field.

There are two techniques used to deal with partial reconstruction. One is based on the k-factor formalism which has been introduced in Chapter II, and is based on the statistical correction of the properties of the combination of the reconstructed tracks. The correction is obtained through a full simulation of the decay including the reconstruction and the selection procedure, with inputs from the theoretical model of the decay (form-factors).

The k-factor formalism is described in some detail for the decay mode $B_c^+ \to J/\psi\,\mu^+\nu$ in Chapter VI. The second method consists in estimating the missing momentum on an event-by-event basis and is briefly discussed here to motivate the choice of the k-factor formalism.

The properties of the partially reconstructed particle with known mass can be inferred from a subset of reconstructed daughter exploiting the information on the flight direction obtained by the relative position of the production and decay vertices of the particle. The angle θ_{DIRA} between the flight direction and the momentum of the combination of the reconstructed particles is a measure of the fraction of the momentum taken from the missing particle(s).

The kinematics of the decay can be exploited to write the quadratic relation

$$\begin{aligned} 4(p_\perp^2 + c^2 m_{\mathrm{comb}}^2)p_\nu^2 + \\ 4(p_{//}(2p_\perp^2 - c^2 m_{\mathrm{comb}}^2 + c^2 M^2))p_\nu + \\ 4(p_\perp^2(2p_{//}^2 + c^2 M^2)) - (c^2 m_{\mathrm{comb}}^2 - c^2 M^2)^2 = 0 \end{aligned} \tag{IV.10}$$

where $p_\perp$ and $p_{//}$ are the components of the momentum of the daughter combination perpendicular and parallel to the flight distance, m_{comb} is the invariant mass of the daughter combination, and M is the mass of the partially-reconstructed mother particle. The equation can be solved event by event with respect to the momentum p_ν of the missing particle. The value is obtained up to a two-fold since the angle between the reconstructed momentum and the flight distance cannot be used to determine whether the neutrino was emitted forward or backward.

The effect of this ambiguity can be mitigated by using a dynamical model for the decay and choosing the solution corresponding to the larger double differential decay width or including in the data modeling both the solutions.

This method has been successfully applied for b-hadron decays but it becomes harder when the lifetime of the mother particle decreases. Shorter lifetime means shorter decay length and larger uncertainty on the angle of the flying particle being more affected by the uncertainties on the positions of the primary and secondary vertices.

This technique has been discarded for the lifetime measurement of the B_c^+ because it has been found that the time-dependent uncertainty introduced in the decay time measurement through the uncertainty on the decay angle exceeds the uncertainties on the theoretical model, and/or forces to suppress statistics taking only well detached events. Still, the technique is used for the assessment of systematic uncertainties on the theoretical model describing the semileptonic B_c^+ decays as discussed in the following chapters.

IV.2.4 Offline selection (Stripping)

The following stage of the LHCb data flow implemented within the Gaudi framework is the signal selection or Stripping.

The Stripping project is a collection of hundreds of algorithms, named Stripping lines, building candidates and applying selections to them. Almost every analysis has its own dedicated Stripping lines, but there are some Stripping lines shared by several analyses. In particular those containing detached or high-momentum J/ψ and $\psi(2S)$ resonances, which are a frequent final state of b-hadron decays.

During the three years of my Ph.D. I have contributed to the development and to the maintenance of the Stripping package as contact person and *stripping liason* for the "quarkonia and b-hadron" working group.

I contributed in particular in two sectors: I developed a pilot project for the monitoring of the Stripping stability and I proposed an object oriented structure for the J/ψ Stripping.

General structure of the Stripping package

Differently from other physics software, the stripping code is expected to be modified by many users. Most of them use only standard tools to define the selections needed by their analysis, while more advanced users may implement dedicated algorithms to implement part of the analysis in the Stripping phase and save computing resources.

Besides stripping contains the selection strategies for all the analyses performed at LHCb, and it is therefore crucial to enforce an advanced version control system which allows not only to roll-back the source code to a given version, but to run the Stripping code for any version possibly used for analysis.

The key requirements are then robustness and flexibility. The Stripping project implements three packages named Stripping Selection, Stripping Settings and Stripping Archive. Only the first of these is accessible by all the analysts and it represents the repository to *develop* the stripping selection strategies.

The analysts provide objects named *Line Builders* which are configured during the central production and register a set of *Stripping lines* sharing some configuration parameters stored in *Configuration dictionaries* which are passed to the Line Builder by the production or test environment.

Facilities for the test of small number of Line Builders are also provided, including tools to write the selection output bandwidth to ROOT files to ease the verification of the selection output. Key requirements to the development of stripping lines are the *bandwidth*

and the *processing time*. The bandwidth takes the name from the trigger jargon and is proportional to the product of the event-size and the number of events stored. Candidates saved using the microDST format have a smaller event size than events stored on DST. Hence, stripping lines writing to microDST are allowed to save more candidates than lines writing to DST, respecting the limit on the bandwidth imposed by the finite amount of storage resources. The processing time has to be shorter than 0.5 ms per event. To achieve this result it is crucial to implement properly the construction of the decay candidates. Filtering pre-constructed candidates as described above and sorting the daughter particles from the rarest to the most common to skip quickly events for which the required particles are not available, are common hints to speed up stripping lines.

The Stripping Settings package contains the definitions of the configuration dictionaries of all the lines that have to run during the production jobs. A dozen of people has write access to the Stripping Settings package: the stripping coordinators and the *stripping liasons*. The latter are deputies of the analysis working group in charge for checking that the stripping lines of their group comply with the requirements, to move the configuration dictionary to the Stripping Settings package, and deploy the line for production. Stripping liasons also check for a reasonable optimization of the stripping lines of their working group and push the line authors to take action in case of need. More aggregate testing tools are available at this level. Stripping liasons rarely have to check the output of the stripping lines and limit their interest to bandwidth, processing time, and number of candidates saved for event. A too large candidate multiplicity is often due to mis-balanced requirements on the daughter particles which make the stripping line to focus on the existence of one of the daughter particles instead of the actual combination. It is also joint responsibility of the conveners and of the stripping liasons to respect the group bandwidth that the Collaboration has decided to allocate to the activities of each group.

Finally, the Stripping Archive package is the repository of all the past configurations of the stripping lines. It backups the Stripping Selection package at the production time in order to preserve a snapshot of the code available in the future to rerun Stripping on simulation or on data collected in a later data-taking. Only Stripping coordinators have write access to the Stripping Archive package.

Monitoring tools for the Stripping development

As it should be evident from the number of pages composing this Chapter, the amount of operations from the detection of particles in the detector sensitive volumes to the output of the event selection is huge. Many different experts cooperate in different parts of the data flow chain, and not always the impact of a modification at some level of the reconstruction has the expected results on the output of the signal selection.

Besides the stripping authors are many physicists having clear the minimal details of their own analysis but often ignoring the details of the reconstruction and almost never following with attention the releases of the reconstruction software.

These considerations have pushed towards a Stripping-based validation of the software releases through a simple tool running on a regular basis on many Stripping lines and

comparing the distribution of some relevant variables to spot changes in the relevant code and to notify promptly the Stripping authors in order to make evident what caused the variation.

This validation tool is composed of three main bodies:

- A simple XML database containing the information on the Stripping lines for which the validation is required and on the variables to test.
- A web interface to fill the database and to check the histograms filled during the validation. The interface includes a simple automatic tool to check all the produced histograms and look for inconsistencies.
- A DaVinci algorithm running on the CERN batch system reading the XML database and analysing with the latest Stripping code a small sample of data from the reconstructed FULL.DST data sample. The algorithm saves the histograms in slots which can be made different because of the version of the software or because of generic conditions under which the user is testing the stability of the Stripping.

The software tool is currently under the scrutiny of the Stripping project coordinators.

Object-oriented structure of the DiMuon Stripping

The current version of the Stripping project for the DiMuon candidates is a pattern of Stripping lines defining the candidates needed for several different analysis with slightly different requirements. For example there are Stripping lines to select J/ψ and $\psi(2S)$ lines which are identical, except for the different mass regions. In a similar way, there are SameSign DiMuon lines which are used to assess the combinatorial background when selecting DiMuons which are supposed to be identical to the OppositeSign partners with the only exception of the charge of the muons.

The current Stripping code makes abundant usage of code repetition so that it often happens that loosening a cut in the OppositeSign line to enhance signal, analysts forget to loosen cuts also for the SameSign version of the line, making the events selected by the new version almost useless because not supported by the control line.

To avoid these severe maintenance issues without loosing generality I proposed to move to a hierarchical structure for the Stripping project, where lines which are supposed to reproduce the behaviour of another line are physically linked to that line, and import all the selection strategy, replacing only requirements which are intended to be different.

The DiMuon Stripping used for the latest Stripping of LHC-RunI has been translated into this hierarchical structure obtaining exact correspondence. The hierarchical structure is depicted in Figure IV.4.

The transition to this improved structure is being considered for future stripping productions.

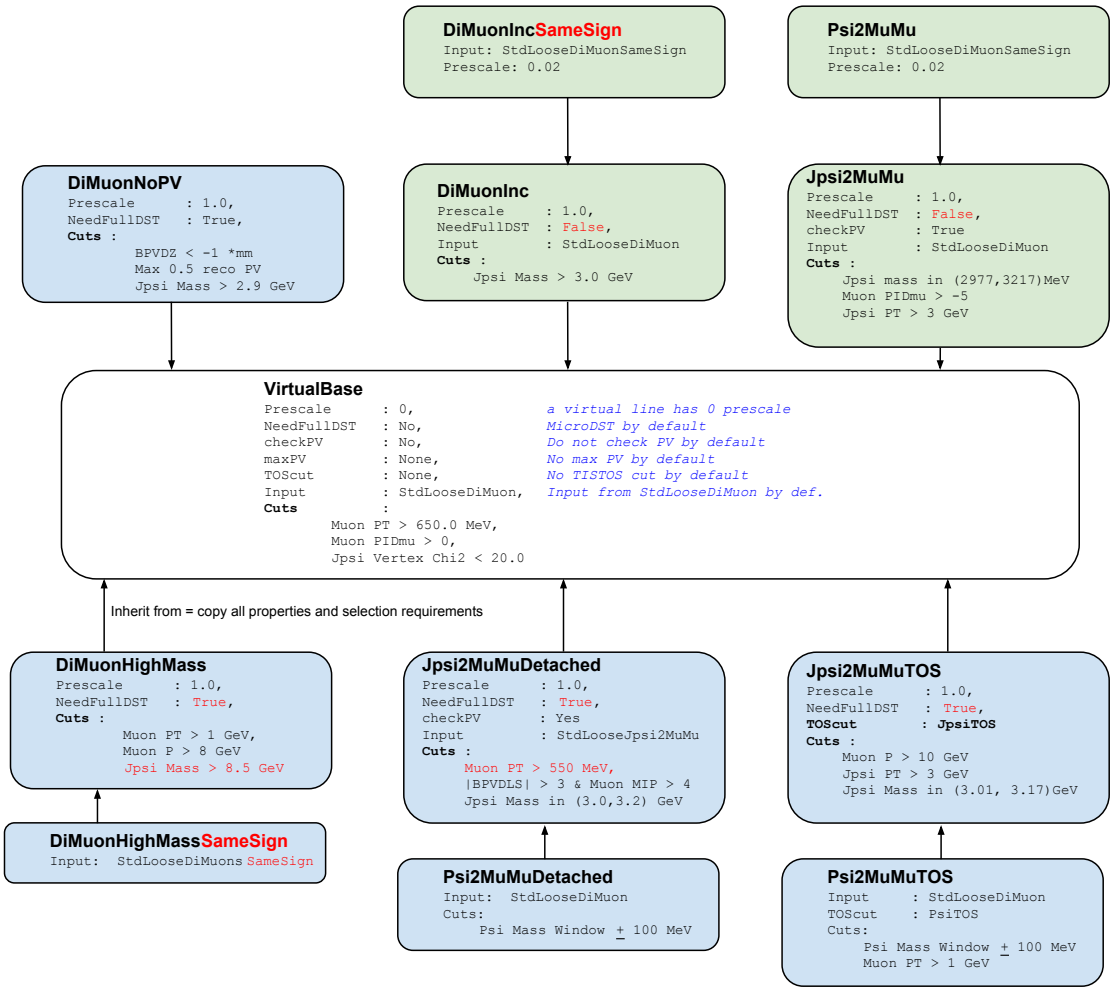

Figure IV.4: Hierarchical structure of the proposed DiMuon Stripping.

IV.3 The LHC Computing Grid

The need to process huge amounts of data is not limited to the LHC community. Data analysis in physics, chemistry, genetics, mathematics, finance and other fields often make use of extensive computing and data storage infrastructure for analysing data. Sometimes, the computing power is too large to be affordable for a single institution or country. There are several reasons to prefer a multi-location strategy:

- availability: if a whole computing center gets off-line for maintenance or network reasons, the data can still be accessed and analysed from the other computing centers;
- reliability: to ease parallel analyses, data are often replicated several times on the grid nodes. If one of the replicas is corrupted or lost, data can be retrieved from the

other computing centers;

- safety: catastrophic events and human errors can challenge the operations of a computing center. But even if all data were lost at the same time, replicas would still be available on other computing centers, maintained by other people in different locations and therefore less subject to correlated disasters.

Metacomputing is a technology intended to integrate multiple computing resources to create a heterogeneous resource making the integration as transparent as possible to the different applications accessing those resources. It is an ancestor of the modern computing grid, recently defined with the following statement [130].

> "A Grid is a hardware and software infrastructure that provides dependable, consistent, and pervasive access to resources to enable sharing of computational resources, utility computing, autonomic computing, collaboration among virtual organizations, and distributed data processing, among others."

The grid architecture is described as a set of layers, each layer connected only to the previous and the next layers, and being dedicated to a specific purpose.

The higher level of the Grid architecture, accessed by *Grid users*, is the *Application and Service* layer, defined by the related software used by the scientific community accessing the grid. This level provides also many management-level functions such as accounting and usage monitoring.

The Application and Service layer is connected to a layer which consists of software systems masking the heterogeneous system by defining and implementing the communication protocols and the application programming interfaces. This layer, named *Middleware*, covers a role similar to the Operating System of a personal computer, hiding the complexity of the access to the hardware.

The lower layer is indeed named *Hardware* layer, and consists of the actual computing resources that are parts of the Grid, including server and storage devices.

Finally the *Network* layer defines the underlaying connectivity for the resources in the Grid.

The implementation of the architecture is specific to the different Grid projects developed in different scientific communities with different scopes. The LHC Computing Grid (LCG) is a distributed computing infrastructure devoted devoted to provide the production and analysis environment for the four major LHC experiments [131]. The aim of the LCG is to use a world-wide Grid infrastructure of computing resources to provide computational, network and storage resources an order of magnitude higher than what was available to previous particle physics experiments.

The amount of data to be stored is too large for a single computing center and it is therefore distributed world-wide to 170 computing centers in 40 countries [131].

The LCG is structured in four tiers. The Tier-0 computing centers, interface between the detectors and the LCG, is located in the Meyrin CERN site, and was extended with a new data centre hosted by the Wigner Research Centre for Physics in Budapest, Hungary. The Tier-0 Data Center stores the raw data of all the LHC experiments and ships replicas

to the Tier-1 Data Centers. Usually there are one or two Tier-1 Data Center per country. Tier-0 and Tier-1 Data Centers reconstruct raw data and ship reconstructed data to Tier-2 for analyses. Tier-3 Data Centers are used for statistical analyses and are usually not included in central production jobs. Recently, some Tier-2 Data Centers have been included in reconstruction jobs to enhance performance. The production of Simulated samples is performed on Tier-2 data centers.

IV.3.1 The DIRAC Middleware

LHCb has developed a dedicated Middleware, named DIRAC, for production and user jobs running on heterogeneous architecture including LCG computing center, the computing power made available by the Yandex Company, and the PC farm used for the Software trigger, available during the technical stops and long shut down periods.

DIRAC provides the implementation of a *Distributed Infrastructure with Remote Agent Control* and introduced the concept of *Pilot Job* in the *Workload Management System* (WMS). The matching of available resources and payloads in the central queue is done remotely by the Pilot Jobs instead of by a central application. The advantage of this technique is an additional layer between the WMS and the heterogeneous architecture of the distributed computing system implementing the pull scheduling paradigm. This additional layer ensures the availability of the resources and checks for hardware faults before matching payloads to that hardware, and provides a uniform interface to the WMS independent of the actual hardware. Finally, Pilot Jobs map the available resources in a much more reliable way, because the Pilot Job asks for payloads only once it has checked the actual availability and integrity of the needed computing resources. Conceptually, the Pilot Job is not too different from a virtual machine with dedicated computing resources.

Every Pilot Job performs an installation of DIRAC including a full download of the latest running configuration. Once the exact working conditions are determined (available disk, memory, and CPU, running platform, available Grid data, ...) a DIRAC Job Agent is executed. The *Job Agent* is responsible for sending the payload request to the central WMS server, and later to execute the received payload. The common execution environment for all payload is set through the instantiation of a *Job Wrapper*, which is the last responsible for the execution of the payload and can execute management commands like killing jobs, uploading the output data of the payload on the Grid storage and so on.

A more technical overview of the Workload Management System implemented through DIRAC was published in Ref. [132], while a complete review of the LCG with particular care to the distributed database containing the information of the available data and replicas can be found in the introductory chapters of the Thesis in Ref. [133].

A further layer, named *Ganga* has been developed to help the users to configure their jobs to be run on the LCG. Ganga is an easy-to-use front-end for job definition and management, implemented in Python. It includes built-in support for configuring and running applications based on the Gaudi framework, described below. Ganga allows trivial switching between testing on a local batch system and large-scale processing on Grid resources [134].

Part 2

B_c^+ *meson lifetime measurement*

The second part of this Thesis is devoted to the description of the analysis I carried out on the LHCb data collected in 2012 aiming at the measurement of the lifetime of the B_c^+ meson using the $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays.

In September 2012, when with my collegues in Firenze we started to consider this analysis, the outcome was quite uncertain. Several difficulties were evident from the beginning, and we realized about further complications only in the following months. The competition with the lifetime measurement using the hadronic $J/\psi\,\pi^+$ decay mode was somewhat hard because the partial reconstruction introduces several uncertainties that challenge the precision and the accuracy of the measurement. Hence, the idea of LHCb publishing a precision measurement obtained with the semileptonic channel before the result from fully reconstructed $B_c^+ \to J/\psi\,\pi^+$ decays caused more than one year of discussions and a tough internal review process, performed blinding the result.

At the end of October 2013, the internal reviewers, Sheldon Stone and Fatima Soomro, accepted to unblind the analysis, and we started preparing the publication, refining the evaluation of systematic uncertainties and finally moving to the collaboration-wide discussion about the analysis and its result.

In January 2014 the result was finally released on the arXiv for the review of the High-Energy Physics community and presented to several international conferences [135–140].

In May 2014 the result appeared on the European Journal of Physics after being received with enthusiasm by the referees [141].

The following chapters are organized as follows. Chapter V introduces the analysis technique, the choice of the datasets, the signal and background sources and motivates the selection strategy.

Chapter VI describes the signal and background models, including the k-factor method used to relate the B_c^+ mean lifetime, defined in the B_c^+ rest frame, to the *pseudo-proper decay time* measured in the rest frame of the combination $J/\psi\,\mu^+$. The model is then applied to data, the result obtained through a maximum likelihood fit, and the statistical uncertainty assessed with several methods.

Chapter VII describes the systematic uncertainties and the techniques used to assess them.

The result is compared to previous measurements in Chapter VIII.

Analysis strategy

Analysing the full dataset collected at LHCb in 2011 and 2012, about three thousand B_c^+ decays to $J/\psi\pi^+$ are observed [65]. As described in Section II.4.5, the branching fraction of the semileptonic decay $B_c^+ \to J/\psi\mu^+\nu$ is about twenty times larger than it is for the hadronic mode $B_c^+ \to J/\psi\pi^+$. With such statistics, the lifetime measurement is likely to be limited by systematic uncertainties. The data samples used for the analysis has been therefore limited to data collected in 2012 in order to avoid complications in the model arising from the different beam energies, trigger configurations, and detector alignments. In particular, a bias on the decay time distribution due to trigger tracking algorithms was found in the precision measurement of the lifetime of b-hadrons. An *a posteriori* correction of the bias is possible, but not effortless, while the effect was removed from the trigger tracking software before the beginning of the 2012 data taking. Comparing the small benefit to the large effort required, it has been decided not to include data collected in 2011 in the analysed dataset.

The online and offline event selection is driven by the need to minimize any bias on the decay time of the signal events, while suppressing the abundant background. Since the analysis is expected to be dominated by systematic uncertainties it is not obvious how to define a proper figure of merit to perform an optimization of the selection criteria. Particular care is taken to avoid the inclusion of requirements related to the decay time of the particle. This includes track pointing variables, less accurate for shorter decay lengths, and explicit detachment requirements on the B_c^+ candidates. However, requirements based on the impact parameters are extremely powerful when rejecting background. Most of the tracks in the event are generated in the pp collision, and therefore they have a small impact parameter. Ignoring tracks with an impact parameter smaller than a threshold when creating the flying particle candidate is a criterion offering high efficiency and excellent background rejection. The drawback is the explicit bias on the decay time distribution. Renouncing to the impact parameter criteria forces to harden particle identification requirements on the three muons in the final state. Muons are relatively rare particles in the typical pp-collision event at the LHC, so that the probability of observing a three-muon vertex is relatively low even without removing tracks consistent with a primary vertex.

The most dangerous source of background is therefore related to the misidentification of hadrons as muons. The invariant mass of the dimuon forming the J/ψ candidate is required to be consistent with the narrow J/ψ mass peak so that standard *Particle IDentification* (PID) requirement are sufficient to reduce *fake* J/ψ background to an acceptable level. On the other hand, the background due to the misidentification of the *bachelor muon* is potentially more dangerous and it is therefore crucial to suppress and model this so-called *misidentification background* source as effectively as possible.

The B_c^+ lifetime is then determined through a two-dimensional fit to the distributions of the invariant mass of the J/ψ μ^+ combination and the *pseudo-proper decay time*, the B_c^+ decay time in the rest-frame of the $J/\psi\,\mu^+$ combination. The statistical model takes into account the correlation between the two variables for most components.

The signal model and the relative template distributions rely on the full simulation of $B_c^+ \to J/\psi\,\mu^+\nu$ decays and *feed-down* decays as $B_c^+ \to \psi(2S)\mu^+\nu$ followed by the decay $\psi(2S) \to J/\psi\,X$. Uncertainties on the form-factor model and on the feed-down decay branching fractions are assessed using data-driven techniques and included in the systematic uncertainty on the data model described in Chapter VII.

A statistical approach is taken in order to relate the *pseudo-proper decay time* $t_{\rm ps}$ to the average lifetime of the B_c^+ meson, defined as the mean value of the *proper decay time* t in the rest frame of the B_c^+ meson.

Background models are obtained from real data whenever possible. In particular, the effect of the multitude of B channels contributing to the *misidentification background* can be measured from the events selected without any requirement on the *bachelor muon*. The resulting distributions are weighted by the probability of the misidentification of the bachelor muon candidate, which is measured on calibration datasets. Using a similar technique to the one employed by CDF as summarized in Section II.5 and in Ref. [83].

Other background sources include *fake* J/ψ candidate, modeled using J/ψ mass sidebands, and random combinations of real J/ψ and muons. The detached part of this background is modeled using a large sample of simulated $B_{u,d,s} \to J/\psi\,X$ decays; while distributions for the prompt component, where both the J/ψ and the muon are consistent with the same primary vertex, are modeled from data, with the yield being a free parameter of the global fit used to extract the lifetime.

V.1 Hadronic vs. semileptonic channel

As mentioned above, two different approaches have been used to measure the B_c^+ lifetime. The hadronic decay mode, $B_c^+ \to J/\psi\,\pi^+$, has the advantage of being a fully reconstructed decay mode. The determination of the rest frame of the B_c^+ meson is easy, there are no theoretical inputs to the measurement, and the signal–background separation benefits of the excellent mass resolution of the LHCb spectrometer since the invariant mass of the signal candidates peaks at the mass of the B_c^+ meson, while the mass distribution of the background events is a smooth distribution, usually parametrized as an exponential or polynomial *probability density function* (*pdf*). Unfortunately, as discussed in Section II.4.5,

the branching fraction of the $B_c^+ \to J/\psi\,\pi^+$ decay is small, and the lifetime measurements are likely to be dominated by the statistical uncertainties. Besides, in order to enhance the signal significance, the selection relies on multivariate techniques often including variables related to the decay time and therefore introducing an explicit, but non-trivial, bias in the decay time distribution. In particular, impact parameter requirements on the pion are unavoidable because of the large number of pion tracks per event consistent with the primary vertex. As a consequence, the challenge of the analysis moves to the measurement of the acceptance as a function of the decay time. There are three methods to measure the acceptance function: it can be obtained from a simulated signal sample, it can be measured studying the decay of a particle whose lifetime is well known to a similar final state, or it can be measured virtually moving the position of the reconstructed primary vertex alongside the momentum of the mother particle and measuring the probability of selection according to the defined criteria. The latter technique is named *swimming*.

None of the three techniques is totally satisfactory. Simulation is a very complex process which is not free from uncertainties, and it is not obvious how to estimate the deviation of the simulated acceptance function from reality. A technique widely used is to simulate the acceptance function for a decay of a particle having well known lifetime and to show that the application of acceptance function to real data allows to retrieve the expected lifetime.

An evolution of this technique is the measurement of the acceptance function assuming an exponential distribution for the decay time of the reference particle. If the decay time of the two particles M (measured) and R (reference) distributes as

$$\Gamma_M(t) \propto \exp\left(-\frac{t}{\tau_M}\right) \quad \text{and} \quad \Gamma_R(t) \propto \exp\left(-\frac{t}{\tau_R}\right), \tag{V.1}$$

and the efficiencies for selecting M and R decays are $\varepsilon_M(t)$ and $\varepsilon_R(t)$, respectively; then the event yields are

$$N_M(t) \propto \varepsilon_M(t) \exp\left(-\frac{t}{\tau_M}\right), \quad \text{and} \quad N_R(t) \propto \varepsilon_R(t) \exp\left(-\frac{t}{\tau_R}\right), \tag{V.2}$$

and their ratio, experimentally accessible, is

$$\frac{N_M(t)}{N_R(t)} \propto \frac{\varepsilon_M(t)}{\varepsilon_R(t)} \exp\left(-t\left(\frac{1}{\tau_M} - \frac{1}{\tau_R}\right)\right). \tag{V.3}$$

Known the value of τ_R, and simulating the efficiency ratio $\frac{\varepsilon_M(t)}{\varepsilon_R(t)}$ which is very close to unity, the measurement of $\frac{N_M(t)}{N_R(t)}$ is sufficient to determine the lifetime. This technique has been successfully used in many precision lifetime measurements of b-hadrons, studying the lifetime of the B^+, B_s^0 and Λ_b^0 hadrons relative to the lifetime of the B_d^0 meson. The technique relies on the similarity of the final states used to reconstruct the target and reference particles. This is appropriate for most b-hadrons, but it is not obviously applicable to the B_c^+ meson, since the kinematics of the natural reference mode $B^+ \to J/\psi\,K^+$ is

different due to the large differences of masses and lifetimes; and the simulated corrections are obtained using a production model which has not been experimentally verified, relying instead on theoretical assumptions for the production mechanism of the B_c^+ meson, which is different from the production mechanism of the B^+ meson, used as reference.

Another alternative is *swimming*, developed at LEP and used at CDF [142, 143]. It consists of an event-by-event determination of the *turning points* of the efficiency at which the single candidate efficiency changes from *not-selected* to *selected* by gradually increasing the distance between the production and decay vertices. Swimming is a good alternative to the previous methods when no reasonable reference decay channel can be defined and when the event yield is large enough to make systematic uncertainties on the simulated acceptance function non negligible compared to the statistical uncertainty. The drawback is that it neglects any potential bias to the decay time distribution due to reconstruction, only considering the selection as possible source of dependence of the acceptance function. The other negative aspect of swimming is that it is technically very complex, and slows down the analyses since it needs large amounts of computing resources to recompute the acceptance function every time that the selection strategy is changed.

The second approach to the B_c^+ lifetime measurement consists in using semileptonic decays. While the acceptance function has to be determined precisely when using hadronic decays, the large branching fraction of the decay $B_c^+ \to J/\psi\,\mu^+\nu$ and the clean experimental signature allows to define a tight selection strategy without variables related to the decay time. The acceptance function is therefore expected to be flat and simulation can be used to confirm this hypothesis and constrain possible residual effects of the reconstruction. Unfortunately the missing energy due to the non-reconstructed neutrino in the final state has unpleasant implications:

- the signal model relies on the theoretical form-factor models to statistically describe the relation between the pseudo-proper and proper decay times. A systematic uncertainty from the decay model is therefore unavoidable;

- feed-down decays as $B_c^+ \to \psi(2S)\mu^+\nu$ with $\psi(2S) \to J/\psi\,X$ are impossible to separate from the signal, but they are described by different decay models polluting the signal sample. These decays are tedious because the shape of the distribution depends also on the lifetime of the B_c^+ meson itself so that they cannot be simply statistically subtracted from the sample, and on the other hand their contribution is not known precisely because most of the feed-down decays have never been observed and are theoretically predicted with large uncertainties;

- the signal lays in the wide interval $3.5 < m(J/\psi\,\mu^+) < 6.3\,\mathrm{GeV}/c^2$ where many background sources, peaking and not peaking, contribute, and for which no trivial parametrization (as an exponential or a polynomial model) can be used.

The combination of the results of the two strategies is highly desirable because the systematic uncertainties are totally uncorrelated and they constitute an excellent mutual cross-check of the necessary analysis techniques.

V.2 Signal selection

There are two major requirements for the selection of decay candidates. The first is to avoid any bias on the decay time distribution of the signal events. The second is that the Trigger and Stripping selection criteria should not use variables related to the PID of the bachelor muon. The second requirement is imposed by the need of modelling the background due to the misidentification of the bachelor muon using a data-driven technique.

In order to comply with the second requirement, the J/ψ candidate is required to be sufficient to trigger the acquisition of the event both in Level-0 and High-Level trigger selections. In other words, the J/ψ candidate is required to be Triggered On Signal (TOS, cf. Section IV.2.2) by the following trigger lines (defined in Section III.3):

- *Level-0* – `L0Muon` or `L0Dimuon`;
- *High-Level* – `Hlt1DimuonHighMass` and `Hlt2DiMuonJpsiHighPT`.

The chosen lines are designed to be lifetime unbiased, since they do not require detachment from the primary vertex of the daughter muons or of the J/ψ candidates.

It is interesting to observe that the complex software trigger of LHCb includes a dedicated line for the $B_c^+ \to J/\psi\,\mu^+\nu$ channel applying looser cuts to the J/ψ resonance when found in combination with a third muon from the same vertex. The dedicated trigger line is very useful for other analyses (see *e.g.* Ref. [77]) but the PID requirement on the bachelor muon discourages its use for the lifetime measurement, where special treatment of the background due to misidentification of the bachelor muon requires a sample of unbiased bachelor tracks. The chosen trigger line relies instead on a harder requirement on the transverse momentum of the J/ψ, introduced to limit the bandwidth of the stripping line to an acceptable level.

A dedicated Stripping algorithm was developed to create and select B_c^+ candidates from $J/\psi\,\mu^+ \to (\mu^+\mu^-)\mu^+$ vertices. No PID requirements are applied by the Stripping algorithm to the bachelor muon. This means that candidates made from $J/\psi\,K^+$, $J/\psi\,\pi^+$, and $J/\psi\,p$ vertices are also selected.

Requirements on the minimal transverse momentum are applied to the two J/ψ decay products (1.4 GeV/c), the J/ψ candidate (2 GeV/c), the bachelor muon (2.5 GeV/c) and the J/ψ μ^+ combination (6 GeV/c). The momentum of the bachelor muon must be between 3.066 and 3.131 GeV/c^2, a range corresponding to about four times the mass resolution. Two sideband mass regions, 3.005 – 3.036 and 3.156 – 3.190 GeV/c^2 are used to evaluate the background from track pairs misidentified as J/ψ candidates. The three muons are required to originate from a common vertex, with a χ^2 per degree of freedom from the fit smaller than 3.0. This restrictive requirement suppresses combinatorial background from random association of real J/ψ and muon candidates not originating from the same vertex. The $J/\psi\,\mu^+$ mass, $m(J/\psi\,\mu^+)$, is reconstructed from a kinematic fit constraining the J/ψ mass to its known value [1], and is required to be between 3.5 and 6.25 GeV/c^2. Particle identification is based on the information from the Cherenkov detectors, calorimeters and

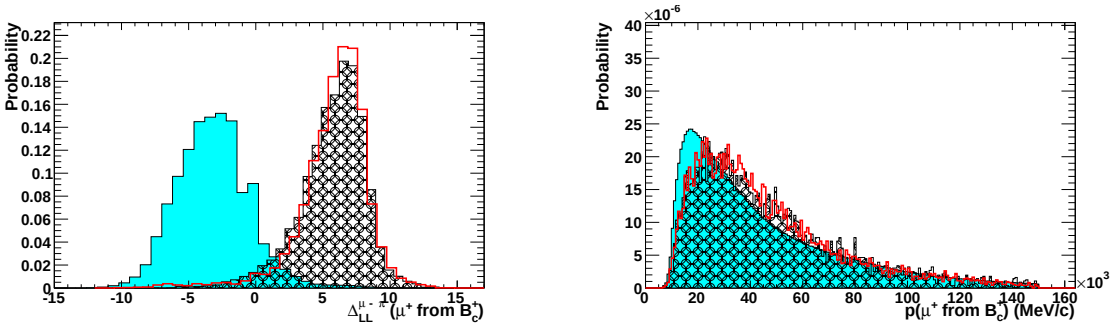

Figure V.1: Normalized distribution of the bachelor muon momentum (right) and muon logarithmic likelihood relative to pion hypothesis (left, cf. Section III.2.1). The distributions of the simulated signal samples are represented as a solid red line, whilst the black patterned distribution is obtained selecting a signal sample (purity $\sim 90\%$) from real data. The cyan shaded histogram represents the background distribution in real data.

muon system, combined in PID likelihood relative to the pion hypothesis as described in Section III.2.1. The requirement $\text{DLL}_{\mu/\pi} > 1$ is applied on the two muon tracks forming the J/ψ candidate. Dedicated, more restrictive identification requirements are applied to the bachelor muon candidate, named BachelorMuonId in the following, developed on purpose for this analysis and described in Section V.4.

In order to reduce the misidentification background a tight requirement on the momentum of the bachelor muon is applied: $p(\mu^+_{B_c^+}) > 13\,\text{GeV}/c$. The benefit from this requirement is twofold. First, it suppresses combinatorial background because the momentum spectrum of the B_c^+ meson is harder than it is for background sources. Second, muon candidates with momentum larger than 10 GeV/c are required to have matched hits in all the four muon stations following the calorimeters in order to pass the binary selection IsMuon (cf. Section III.2.1). The momentum spectrum is shown in Figure V.1 where the distribution of the simulated signal is superposed to the selected signal and to the expected background distributions obtained inverting the requirement on the bachelor muon identification.

The selection criteria are summarized in Table V.1.

The efficiency of the various requirements has been studied using simulation as reported in Table V.2.

The total signal efficiency for signal events, including the detector geometrical acceptance and the trigger, reconstruction and offline selection efficiencies, is predicted from simulation to be 0.25%. Note that much higher signal efficiency and better signal significance can be obtained by renouncing to the decay-time-unbiased requirement.

No dependence on the decay time is observed neither due to the selection nor to the reconstruction. Figure V.2 shows the acceptance function measured on a simulated sample of $B_c^+ \to J/\psi\,\mu^+\nu$ decays as a function of the *true* decay time. A linear fit is superposed. The fitted slope, $3 \pm 6\,(\text{stat})$ ns^{-1}, is consistent with zero. The statistical uncertainty on the slope, rounded up to the confidence interval [–10, 10] ns^{-1} is taken into account as a

Cut variable	*Trigger threshold*	*Stripping threshold*	*Statistical analysis*
Track χ^2/ndf	< 3	< 3	< 3
$p_T(\mu_{J/\psi})$	$> 0.5\,\text{GeV}/c$	$> 1.4\,\text{GeV}/c$	$> 1.4\,\text{GeV}/c$
$\mu_{J/\psi}$ PID	IsMuon	IsMuon	$\text{DLL}_{\mu/\pi} > 1$
$p(\mu_{B_c^+})$	–	–	(13, 150) GeV/c
$p_T(\mu_{B_c^+})$	–	$> 2.5\,\text{GeV}/c$	$> 2.5\,\text{GeV}/c$
$\mu_{B_c^+}$ PID	–	–	BachelorMuonId
$\mu_{B_c^+}$ track ghost prob.	–	< 0.4	< 0.3
J/ψ mass	$\pm 120\,\text{MeV}/c^2$	$\pm 120\,\text{MeV}/c^2$	$(3066, 3131)\,\text{GeV}/c^2$
$p_T(J/\psi)$	$> 2.0\,\text{GeV}/c$	–	$> 2.0\,\text{GeV}/c$
$J/\psi\,\chi^2_{\text{vtx}}$/ndf	< 25	< 9	< 9
$m(J/\psi\,\mu^+)$	–	(3.2, 6.2) GeV/c^2	(3.5, 6.25) GeV/c^2
$J/\psi\,\mu^+$ χ^2_{vtx}/ndf	–	< 9	< 3
$p_T(J/\psi\,\mu^+)$	–	$> 6.0\,\text{GeV}/c$	$> 6.0\,\text{GeV}/c$
J/ψ trigger	–	–	L0, HLT1, HLT2 TOS

Table V.1: Summary of applied cuts in the Trigger and Stripping lines used in the analysis, and in the offline selection. Trigger thresholds include `Hlt1DimuonHighMass` and `Hlt2DiMuonJPsiHighPT` requirements. Different cuts are applied to the muons from J/ψ (labeled $\mu_{J/\psi}$) and to the bachelor muon (labeled $\mu_{B_c^+}$).

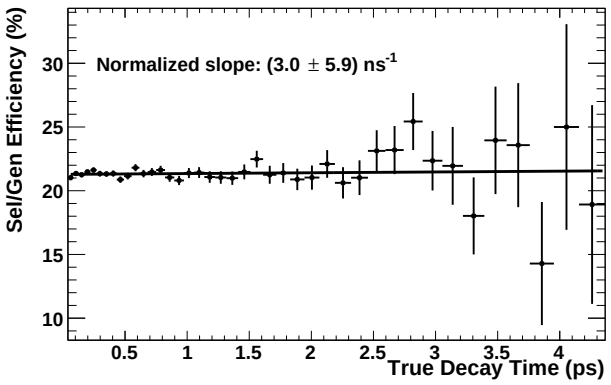

Figure V.2: Dependence of the acceptance function on the decay time of the B_c^+ meson in a simulated sample $B_c^+ \to J/\psi\,\mu^+\nu$.

contribution to the systematic uncertainty on the measured lifetime. The total number of candidates is 29 756.

Level of the selection	*Number of Candidates*	*Efficiency wrt. generated*	*Efficiency wrt. reconstructible*
Generated (4π)	6789935	100%	1144%
Reconstruction			
Reconstructible true candidates	593309	8.74%	100%
Reconstructed true candidates	540399	7.96%	91.1%
Stripping			
$p_T(\mu_{B_c^+}) > 2.5$ GeV/c	142167	2.09%	24.0%
$p_T(\mu_{J/\psi}) > 1.4$ GeV/c	59495	0.88%	10.0%
$p_T(J/\psi) > 2.0$ GeV/c	53618	0.79%	9.04%
$p_T(J/\psi\mu^+) > 6.0$ GeV/c	46349	0.69%	7.81%
All the other Stripping requirements	43808	0.65%	7.38%
Offline selection			
Events selected by the Trigger (TOS)	33646	0.50%	5.67%
PID requirements	26227	0.39%	4.42%
Track and muon isolation	23585	0.35%	3.98%
$m(J/\psi)$ in range (3066, 3131) MeV/c^2	20532	0.30%	3.46%
B_c^+ vertex tight requirement	18454	0.27%	3.11%
All the other offline requirements	17214	0.25%	2.90%

Table V.2: Number of candidates and selection efficiency profiled in the various steps of the analysis. Candidates are *reconstructible* when all the daughter tracks are in acceptance, and *reconstructed* when a reconstructed track is associated to each daughter particle. Offline selection efficiency has been obtained from truth-matched reconstructed Monte Carlo, using reconstructed variables. These numbers are not used in the analysis and are reported for completeness only. The two available Monte Carlo samples, generated with different form factor models, have been combined for this study.

V.3 Multiple candidates

When an event contains more than a candidate for some specific signature it is said to be a *multiple-candidate* event. Large rates of multiple candidates are often a hint of pathologies in the selection strategy due to unbalanced requirements on different parts of the decay candidate, and are a severe issue for analyses based on event counting, as cross-section and branching fraction measurements, because often the probability of an event to contain multiple candidates is different for signal and background events introducing biases in the measurement. The issue is less severe when measuring properties of particles, as mass or lifetime, but it may be not negligible for a high precision measurement.

Three categories of multiple candidates can be individuated: *clone candidates* are candidates where at least an object, usually a track, in the final state is cloned during the reconstruction process. Cloned candidates have special features: their properties (masses, momenta, ...) are almost equal for all the cloned candidates, the angle between the cloned

tracks is very small, and they share detector hits. *Overlap candidates* are candidates where part of the candidate is unique, but multiple *different* objects, usually tracks, can be used to complete the candidate. For example the combination $J/\psi\,\mu^+$ may associate to the same J/ψ several μ^+ candidates reconstructed in the same event. Finally, *genuine multiple candidates* are independent signal candidates which can be found in a single event, for example coming from two different primary vertices, or from the b and $\bar{b}$ decays. Genuine multiple candidates are usually negligible in analyses looking for rare signatures.

For the lifetime measurement of the B_c^+ meson using the semileptonic decays, overlap multiple candidates do not bias the analysis, since the statistical description of random combinations of a J/ψ and random tracks takes them into account, while clone candidates should be rejected since a statistical description of this background is difficult.

In the case of a muon track cloned in the reconstruction procedure, a dimuon from a J/ψ can be misreconstructed as a $J/\psi\,\mu^+$ combination. In this case the invariant mass of the muon triplet peaks at the value

$$m(J/\psi\,\mu^+)^{(\text{clone})} = \sqrt{2m(J/\psi)^2 + m(\mu)^2} = 4379\,\text{MeV}/c^2. \tag{V.4}$$

As it can be seen from Figure V.3 the effect is well visible once all the other selection requirements are applied.

For the study of the $B_c^+ \to J/\psi\,\mu^+\nu$ decay mode, two isolation requirements are combined: *Track isolation* and *Muon isolation*. The purpose of the two criteria is slightly different. The former is intended to suppress clone candidates by removing candidates where the angle between the two same-sign muons is too small, or alternatively rejecting candidates sharing too many hits in the tracking system. The muon isolation is needed to avoid alignment of the bachelor muon track with a J/ψ daughter track which would enhance the misidentification probability of a hadron as a muon, making the background modeling more challenging. Muon isolation is therefore treated in the next section, dealing with the muon identification.

Both the implementations of the track isolation requirement have been studied and comparable performance have been observed as shown in Figure V.3. The threshold on the number of hits in the tracking system shared by the two same-sign muons is set to 20%. Written in formula, if $h(\mu^+_{B_c^+})$ and $h(\mu^+_{J/\psi})$ are the numbers of hits in the tracker used to reconstruct the bachelor muon and the same charge J/ψ daughter; and if the number of shared hits is $h(\mu^+_{B_c^+} \cap \mu^+_{J/\psi})$, the $J/\psi\,\mu^+$ candidate is accepted if

$$\frac{h(\mu^+_{B_c^+} \cap \mu^+_{J/\psi})}{\min(h(\mu^+_{B_c^+}), h(\mu^+_{J/\psi}))} < 20\%. \tag{V.5}$$

The geometrical requirement is applied by requiring that the invariant mass of the $\mu^+\mu^+$ combination is larger than twice the muon mass. The two quantities are exactly the same if the two muon candidates are actually the same muon. The threshold on $m(\mu^+_{B_c^+}\mu^+_{J/\psi})$ was chosen to be 0.3 MeV/c^2 larger than $2m_\mu$ to take into account resolution effects.

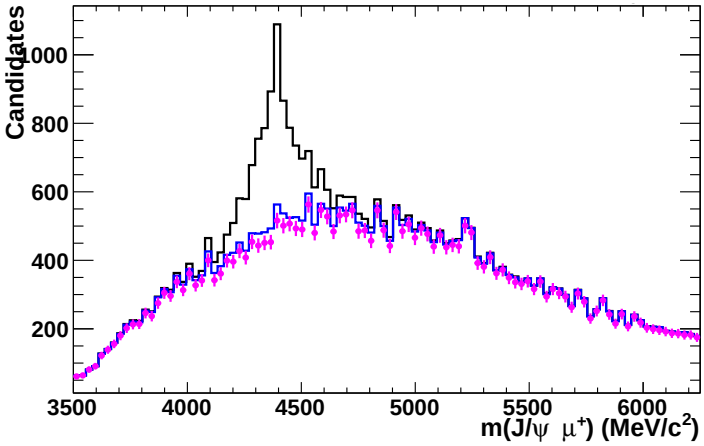

Figure V.3: Invariant mass distribution of the $J/\psi\,\mu^+$ combination before and after the application of the two possible requirements. In blue the hit-based requirements, in magenta the geometrical requirement. The large peak, removed by the application of these requirement is due to J/ψ candidates with one of the two muon cloned and used as bachelor muon in the $J/\psi\ \mu^+$ combination.

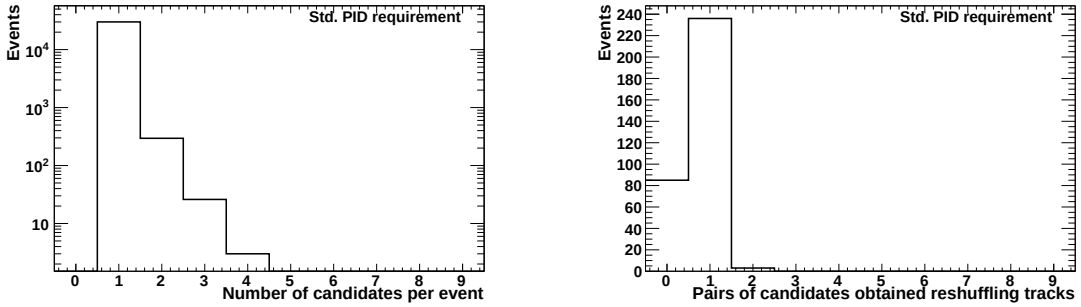

Figure V.4: Number of candidates per event (left) and number of pairs obtained reshuffling the same three tracks by exchanging the two same-sign muon candidates (right). For events with less than two candidates, candidate pairs are not defined, and do not appear in the right plot. This is possible when, by chance, the invariant mass $m(\mu^+_{B_c^+}, \mu^+)$ is consistent with the J/ψ mass. The plots are obtained from the real data sample once the full selection is applied. The right plot shows how in most cases at least one pair of same-event candidates are obtained from reshuffling. In a minority of cases, the existence of same-event candidates is due to other sources.

The technique based on the number of hits shared by the two same-sign muons has been chosen to reject clones because of a wider usage within the Collaboration.

The multiple candidate events after the clone rejection are 0.6%, and they are mainly due to the exchange of the two same-sign muons in the construction of the J/ψ candidate. The effect of multiple candidates on the measurement of the lifetime has been assessed as source of systematic uncertainty and found small, as described in Chapter VII.

V.4 Specialized muon identification algorithms

After discussing the general approach taken to select signal with reasonable purity and particular care in the rejection of some specific background, the focus moves now on the muon identification techniques applied to the bachelor muon candidate. As already mentioned, the random association of a real J/ψ meson with a hadron misidentified as a muon is expected to be the largest source of background to 3–muon secondary vertex candidates. When the J/ψ meson is produced in the decay of a b-hadron it is significantly detached from the primary vertex so that it represents a potential source of bias for the decay time distribution and therefore the lifetime measurement.

Two main sources of muon misidentification background can be distinguished: misidentification due to random association of muon hits to the candidate track, and decays in flight. The strategies to suppress the two backgrounds are different. The accidental association of muon hits to a track can be reduced by improving the association algorithm, for example combining the information from the two views as projected on the x and y axes. Decays in flight can be suppressed by introducing hard vetoes against kaons using the information from the RICH detectors, and by studying the misalignment of the upstream and downstream track segments, to identify a kink in the particle trajectory.

Dealing with the misidentification background requires both *rejection* and *modelling*. The latter is finalized to the statistical subtraction of the background contribution and is described in the next chapter. Rejection of the misidentification background is based on a combination of standard and specialized algorithms to comply with the requirements listed below.

- Excellent rejection of kaons and pions;
- good rejection of protons;
- reduction of the misidentification due to decays in flight $K^+ \to \mu^+\nu X$ and $\pi^+ \to \mu^+\nu$;
- good isolation of the muon track to avoid bias in the muon identification performance with respect to the calibration samples.

A stringent set of requirements, tuned for this analysis, named BachelorMuonId, is satisfied when:

- the condition IsMuon is verified;
- the PID log-likelihoods of the bachelor muon candidate as kaons and protons relative to the pion hypothesis have to be less than 8 and 20,

$$\mathrm{DLL}_{K/\pi} < 8 \qquad \text{and} \qquad \mathrm{DLL}_{p/\pi} < 20; \tag{V.6}$$

- the PID log-likelihood of the bachelor muon as a muon relative to the pion hypothesis has to be larger than 3,

$$\mathrm{DLL}_{\mu/\pi} > 3; \tag{V.7}$$

- after a Kalman filter refit of the track in the muon detector, the muon matches a hit in both detector views in at least four of the five muon stations;
- the χ^2 per degree of freedom (χ^2/ndf) of the Kalman fit is smaller than 1.5;
- the hits associated to the bachelor muon in the muon system must not be associated to the other muons in the three-muon combination;
- the χ^2_{match} describing the quality of the matching between the segments of the long track upstream and downstream the magnet is smaller than 15.

The standard muon identification techniques including IsMuon and DLL requirements are described in Section III.2.1. The others are discussed below.

V.4.1 Kalman filter

The standard technique used to identify muons is very robust and highly efficient, but does not fully exploit the information from the muon system. The Florence LHCb group has been exploring the application of tracking algorithms to the muon identification in various contexts. For example in Ref. [144], Hopfield neural networks are applied to the hits in the muon system to obtain track segments that are then matched to the tracks reconstructed in the tracking system to identify muons. The advantage of neural networks is their speed in the tracking which would allow application of tracking algorithms for muon identification at trigger level and in the online monitoring. However, speed is not a major concern for the bachelor muon identification in the few-million event sample obtained once all of the kinematical requirements on the three-muon combination are applied. Hence, a more standard approach to tracking, based on the Kalman filter, is chosen.

Tracking-based muon identification techniques have at least two advantages:

- in some regions of the muon system, the x and y coordinates of the hits are readout separately reading for example a charge deposited on vertical wires for the x position, and on cathodic horizontal stripes to determine the y position. The standard muon identification techniques were designed to be robust against efficiency losses so that if the information on x (y) is not available, the middle of the pad is associated and to the whole pad width (height) as uncertainty. While enhancing efficiency, this choice is keen to larger misidentification probabilities in case of noisy readout channels affecting only one view. The issue is partially solved by requiring hits in both views. The probability of having noisy channels in both views becomes smaller and therefore the misidentification due to combinatorial association of muon hits to a track is better controlled.
- The standard tool compares the track extrapolation obtained from tracking system with muon hits in each station, without trying to refit the track across the muon detector. The fit procedure based on a Kalman filter is extended with the following

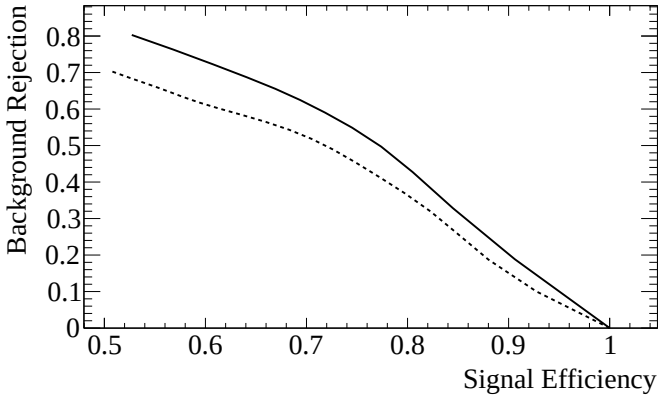

Figure V.5: Signal efficiency and background rejection of the Kalman filter requirement (solid line) compared to the muDLL requirement used as standard technique (dashed line). Signal efficiency is relative to the sample of simulated $B_c^+ \to J/\psi\,\mu^+\nu$ decays selected using only the IsMuon requirement for the PID of the bachelor muon. Analogously, the background rejection is obtained from simulated $B_{u,d,s} \to J/\psi\,X$ decays preselected with the IsMuon requirement for the bachelor track.

iterative process: muon hits within 6 standard deviations from the track extrapolation are attached to the candidate. The Kalman filter is performed and outliers are removed until all hits are within 3 standard deviations from the refitted track. The χ^2 is then calculated using only the best match from each station, called χ^2_K hereafter.

Figure V.5 represents the performance of the Kalman filter technique compared to the standard muDLL criterion described in Section III.2.1.

V.4.2 Muon Isolation

As mentioned above, *muon isolation* is used in this analysis to ensure that the muon identification performance in the sample considered is the same as it is in the calibration samples. As it will be discussed in the next chapter, this is a key requirement of the data-driven technique aiming at a statistical description of the background due to misidentification of the bachelor muon.

The need for isolation requirements arises from the observation that events containing J/ψ+track candidates are expected to have more hits in the muon system with respect to the average event, as used in calibration samples. The two J/ψ daughters are high momentum muons adding hits in the muon detector that can be associated to a random hadron in the event, whose misidentification probability increases. In principle, one would like to remove the hits associated to the J/ψ muons before evaluating the bachelor track muon identification likelihood. However, a more robust approach is chosen, ignoring all the events where there are muon hits associated to both the bachelor muon candidate and one of the J/ψ daughters. The efficiency of this requirement is assessed on simulated

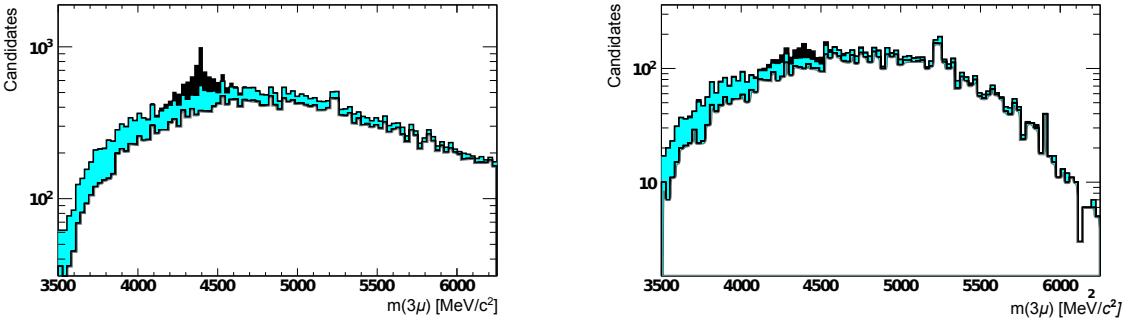

Figure V.6: Deformation of the invariant mass of the three-muon combinations when the isolation requirements are applied in tracking (black histogram is removed), and in the muon system (cyan shaded histogram is removed). On the left plot the whole data sample is shown, while on the right plot only detached events with $t_{\rm ps}$ larger than 250 fs are shown.

$B_c^+ \to J/\psi\,\mu^+$ candidates to be $\sim 99\%$.

The effect of the muon isolation requirement on the distribution of the invariant mass of the combination $J/\psi\,\mu^+$ is shown in Figure V.6.

If it is true that the isolation of the muon candidates affects muon identification performance, it is reasonable (and actually observed) that isolation requirements can improve the background rejection even for analyses dealing with a single muon in the final state.

Several implementations of the muon isolation requirement have been developed within the Collaboration. The most widely used is based on the number of muon hits shared by more than one muon candidate passing the IsMuon criterion. In this case the requirement is not between two specific tracks as in the specialized case described above, but rather between a track and all the others. Using the approach described in the previous section, based on a Kalman filter to determine the muon likelihood based on the muon system only, it is also possible to exploit the resolution on the position of the hits to require a minimal separation between the fitted muon track and the nearest hit associated to another long track. In other words, for each muon hit attached to the candidate with the Kalman fit procedure, the consistency with the extrapolation of all other long tracks with minimal quality requirements and momentum larger than 5 GeV/c is checked. For each station m, the minimum distance $d_m^{\rm other}$ is calculated and express in units of standard deviations. The test statistic $d_{{\rm iso}\mu}$ is then defined as the maximum value of $d_m^{\rm other}$ among muon stations M2 to M5.

The threshold on $d_m^{\rm other}$ has been optimized to 2.75 σ, so that only muon candidates having at least one hit which is not consistent (within 2.75 σ) with any other track are accepted as muons.

The application of the this requirement has been studied on $B_c^+ \to J/\psi\,\mu^+\nu$ candidates and found highly efficient ($\epsilon > 99\%$) and on $B_{u,d,s} \to J/\psi X$ candidates for which a 10% reduction is achieved. Still, the technique was not applied to the final selection because it would be impossible to apply this cut to the calibration samples with the current

reconstruction software.

Only the standard isolation requirement, based on the number of shared hits between the same-sign muon candidates, have been used to select three-muon candidates consistent with the decay $B_c^+ \to J/\psi\,\mu^+\nu$ and used in the analysis.

V.4.3 Decays in flight

Another important handle to reduce the misidentification of hadrons as muons is to study the effect of decays in flight of kaons and pions to final states containing a muon as $K^+ \to \mu^+\nu$.

As described in Chapter III, hadron identification at LHCb takes advantage of two RICH detectors, one upstream the magnet with radiators optimized for low momentum particles that risk to be kicked out of acceptance by the magnetic field, and a second RICH detector for higher-momentum particles. Considering muons with a momentum larger than 13 GeV/c, only the second RICH gives useful information.

In order to remove decays in flight, the information from the RICH detector is crucial because it allows to tag kaons and to reject the muon hypothesis because the hadron hypothesis at RICH level is better.

However if the hadron decays before RICH 2, there is no chance of using the RICH information to tag the hadron. To reduce this source of background, it is reasonable to study the mismatch between the segment of track in the upstream part of the detector, reconstructed using the Vertex Locator and the Tracker Turicensis, and the downstream track segment as reconstructed with the inner and outer tracker system.

The χ^2_{match} is defined as the difference of the track χ^2 and the sum of the χ^2 of the upstream and downstream track segments. For good candidates it is expected to follow a χ^2 distribution with five degrees of freedom, two for the position, two for the slopes, and one for the curvature.

Figure V.7 shows the relation between the decay vertex of a kaon or a pion to a muon and its χ^2_{match} in a simulated sample of $B_{u,d,s} \to J/\psi\,X$ decays.

V.4.4 Performance of muon identification algorithms

The additional criteria provide a 30% reduction of the misidentification background when applied on top of the standard combined PID cuts. This is not surprising since they bring some information not exploited by the standard muon DLL. Table V.3 shows the efficiency and the misidentification background rejection on a simulated sample of $B_c^+ \to J/\psi\,\mu^+\nu$. The misidentification probability has been assessed also on real data studying calibration samples of $D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$ for kaons and pions and $\Lambda^0 \to p\pi^-$ for protons. The study is reported in VI.2.1.

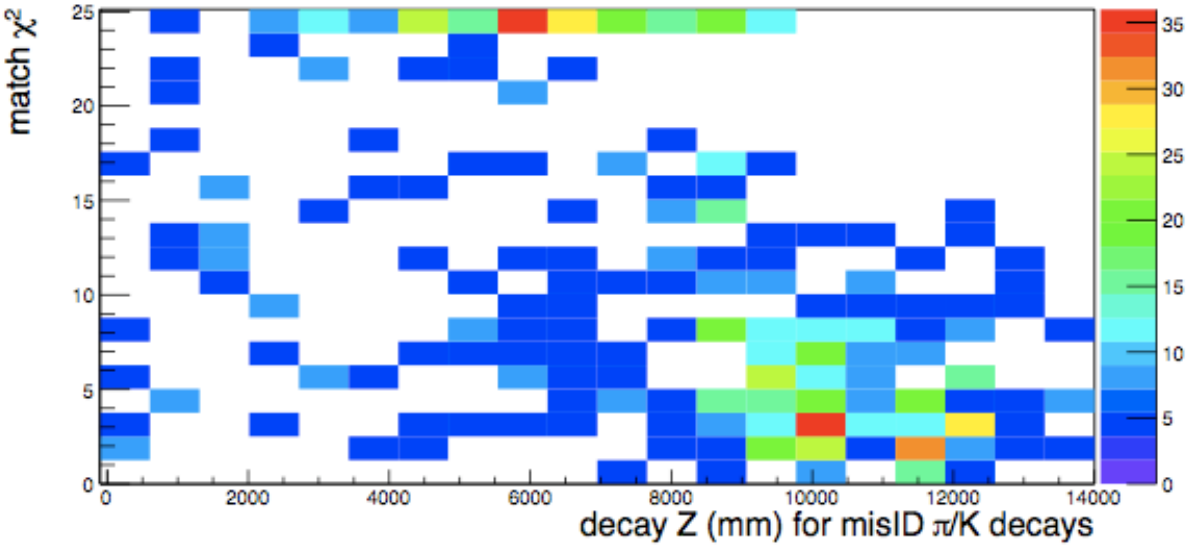

Figure V.7: Distribution of the background to $B_c^+ \to J/\psi\,\mu^+\nu$ due to decays in flight. In both plots, the last bin of the χ^2_{match} variable contains the overflows. The tracker is placed at z between 0 and 9 m. The hotspot between 10 and 12 m corresponds to RICH2.

Requirement	*Signal* [%]	*MisID* [%]	*Fraction of Decays in Flight* [%]
Preselection (hits in >2 muon stations and vertex)	100.00	100.0	0.22
isMuon	98.89	56.5	0.36
$\text{DLL}_{\mu/\pi} > 3$	92.03	16.3	0.82
+ all stations matched in 2 views	92.01	16.2	0.82
+ $\chi^2_{\text{match}} < 15$	89.90	13.6	0.79
+ $\chi^2_K < 1.5$	89.75	13.2	0.81
+ $\text{DLL}_{p/\pi} < 20$	89.72	11.1	0.80
+ $\text{DLL}_{K/\pi} < 8$	89.60	5.6	0.74
+ Isolation	88.97	5.0	0.82

Table V.3: Effect of standard and additional PID requirements, evaluated on a simulated sample of $B_c^+ \to J/\psi\,\mu^+\nu$ decays. Background is studied using a simulated sample of $B_{u,d,s} \to J/\psi\,X$ decays.

	p^+	K^+	π^+
isMuon	0.85%	1.03%	0.56%
$\mathrm{DLL}_{\mu/\pi} > 3$	0.09%	0.50%	0.28%
+ all stations matched in 2 views	0.09%	0.50%	0.23%
+ $\chi^2_{\mathrm{match}} < 15$	0.08%	0.44%	0.20%
+ $\chi^2_K < 1.5$	0.03%	0.42%	0.18%
+ $\mathrm{DLL}_{p/\pi} < 20$	0.02%	0.38%	0.18%
+ $\mathrm{DLL}_{K/\pi} < 8$	0.02%	0.23%	0.18%

Table V.4: Effect of standard and additional PID requirements evaluated on calibration samples for pions, kaons and protons. For the latter the decay $\Lambda^0 \to p^+\pi^-$ is used, for pions and kaons the decay $D^{*+} \to D^0\pi^+$ with $D^0 \to K^-\pi^+$ is used. The muon identification performance varies with the momentum and the rapidity of the track and with the RICH occupancy. The values reported in the Table are an average assuming flat distribution of the probe particles in the ranges $p \in (13, 150)\,\mathrm{GeV}/c$, $\eta \in (2, 5)$, and `nTracks` $\in (0, 800)$. The detailed study is reported in Section VI.2.1.

Data model

The lifetime of the B_c^+ meson is measured from a two-dimensional fit to the joint distribution of the invariant mass $m(J/\psi\,\mu)$ of the combination $J/\psi\,\mu^+$ and of the pseudo-proper decay time $t_{\rm ps}$, defined as the decay time of the B_c^+ meson as measured in the rest frame of the $J/\psi\,\mu^+$ combination.

The template distributions used to compose the data model are classified into two main categories: signal and background. The signal components have a slope of the decay time which depends on the lifetime of the B_c^+ meson. Namely,

- distribution for the main decay $B_c^+ \to J/\psi\,\mu^+\nu$;
- distribution for the feed-down decays $B_c^+ \to J/\psi\,\mu^+\nu X$.

The background category includes

- *Misidentification background*, due to the misidentification of the bachelor track as a muon; this background includes large contributions from $B^+ \to J/\psi\,K^+$, $B^0 \to J/\psi\,K^+\pi^-$ and other B^+, B^0 and B_s^0 decays;
- *Fake* J/ψ, due to the wrong identification of two tracks as a dimuon from a J/ψ decay;
- *Wrong primary vertex association*, due to the association of a primary vertex which is responsible neither for the production of the J/ψ nor of its b-hadron parent;
- *Prompt combinatorial*, due to the combination of a real J/ψ to a real muon, with the J/ψ produced promptly in the pp collision;
- *Detached combinatorial*, due to the combination of a real J/ψ and a real muon, with the J/ψ decay produced in a b-hadron decay.

In this chapter, the models for the various contribution are described.

VI.1 Signal model

VI.1.1 Generator-level distributions

As already mentioned, due to its semileptonic nature and given the presence of spin, the $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decay cannot be simply described with a phase-space distribution. Three theoretical decay models have been interfaced to the Gauss application for physics simulations at LHCb, named for brevity after the authors, Kiselev [33], Ebert [32], and ISGW2 (Isgur, Scora, Grinstein, and Wise) [34]. The models, based on form-factor formalism, are described in Section I.5.3. The form-factor by Kiselev has been chosen as baseline while the others are used for a fast assessment of the dependence of the result on the theoretical model. Systematic uncertainties related to model dependence will be evaluated constraining the possible deviation from these models using the actual distributions observed on data.

The distributions are evaluated assuming a mass of 6276 MeV/c and a lifetime of 0.45 ps for the B_c^+ meson according to the spectra calculated with the Monte Carlo generator BCVEGPY.

Two simulated samples have been generated for the Kiselev model. The first, used to assess efficiencies and optimize the selection contains 10^6 events with only basic acceptance generator-level requirements. After finalizing the event selection strategy, a second $5 \cdot 10^5$-event sample has been produced to increase the statistics needed to model the signal distribution and has been obtained applying harder cuts at generator level in order to save the resources wasted simulating events which are discarded by the final selection. The selected event yields of the second sample is one order of magnitude larger than for the first one.

Figure VI.1 shows the mass distribution at generator level and after the application of the whole selection, including trigger, stripping and offline requirements.

VI.1.2 Pseudo-proper decay time

When reconstructing a decay partially, in this case because of the missing neutrino, it is impossible to define unambiguously the proper frame of the mother particle, and thus to compute the proper decay time from the one measured in the laboratory frame.

As discussed in Section IV.2.3, there are two common approaches to partially reconstructed decays. One exploits the pointing variables to guess the neutrino momentum imposing the alignment of the momentum and flight distance of the B_c^+ meson. The other is a statistical approach based on the theoretical model of the decay which allows to correct statistically the decay time reconstructed in the rest frame of the $J/\psi\,\mu^+$ combination to obtain the average decay time in the rest frame of the B_c^+ meson. The second method, known as k-factor, has been chosen for the lifetime measurement because the usage of pointing variables introduces an important time-dependent bias in the distributions. Pointing variables are used to assess the systematic uncertainty on the theoretical model used to derive the k-factor as described in the next Chapter.

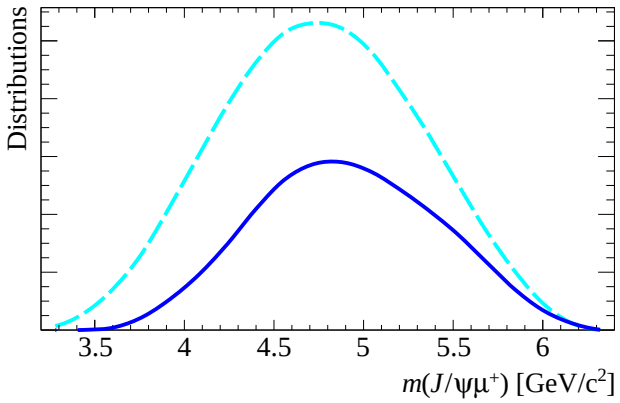

Figure VI.1: Comparison between the mass distribution at generator level as obtained with the Kiselev model (dashed cyan line), and after the application of the whole selection strategy as described in Chapter V. Normalizations are arbitrary.

The decay time in the rest frame of the $J/\psi\,\mu^+$ combination is named *pseudo-proper decay time* because it approximates the proper decay time of the B_c^+ meson, using a different relativistic rest frame. It is defined as

$$t_{\rm ps} = \frac{\mathbf{p}\cdot(\mathbf{v}-\mathbf{x})}{c|\mathbf{p}|}\,\frac{1}{[\beta\gamma]_{J/\psi\,\mu^+}} = \mathbf{p}\cdot(\mathbf{v}-\mathbf{x})\frac{m(J/\psi\,\mu)}{|\mathbf{p}|^2} \tag{VI.1}$$

where $\mathbf{v}$ and $\mathbf{x}$ are the measured positions of the decay and production vertices, and $\mathbf{p}$ is the momentum of the reconstructed $J/\psi\,\mu^+$ combination. The primary vertex associated to each candidate is the one with the minimum $\chi^2_{\rm IP}$, defined as the difference in the vertex fit χ^2 when including or not in the fit the $J/\psi\,\mu^+$ candidate.

It is worth to stress that the pseudo-proper decay time is obtained simply applying Equation VI.1 without additional constraints or fit procedures. This is different from the standard technique used to determine the decay time of a fully reconstructed decay described in Section IV.2.3, based on a refit of the decay kinematics to constrain the mother particle to have flight distance aligned with its momentum. Indeed, since the decay is partially reconstructed, it is not correct to apply this constraint, which would result in a bias to the decay time distribution, due to the fact that the constraint is more important for longer-living particles for which the uncertainty on the flight distance is smaller.

Pseudo-proper decay time distribution for $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decays

We define a formalism for the signal decay time distribution that disentangles explicitly the effect of the unreconstructed final state particles and the reconstruction effects introduced by the finite resolution of the vertex position and momentum measurement. The distribution of the true decay time t^* is a decreasing exponential

$$f_{\rm th}(t^*) = \frac{1}{\tau}\theta(t^*)\exp\left(-\frac{t^*}{\tau}\right) \tag{VI.2}$$

where τ is the lifetime and

$$\theta(t^*) = \begin{cases} 0 & t^* < 0 \\ 1 & t^* \geq 0 \end{cases} \tag{VI.3}$$

is the Heaviside step function.

The true value $t^*_{\rm ps}$ of the pseudo-proper decay time is related to the true proper decay time by a factor k, whose distribution is predicted using simulation

$$t^*_{\rm ps} = kt^*. \tag{VI.4}$$

The distribution of $t^*_{\rm ps}$ is obtained through the Mellin convolution, which states that if $X(x)$, $Y(y)$ and $Z(z)$ are the *pdf*s of two generic random variables x and y, and of their product $z = xy$, then

$$Z(z) = \int_{-\infty}^{+\infty} X\left(\frac{z}{y}\right) Y(y) \frac{{\rm d}y}{|y|}. \tag{VI.5}$$

Therefore, if t^* is distributed according to $f_{\rm th}(t^*)$ as defined in Equation VI.2, and k distributes according to $h(k)$, then

$$f_{\rm ps}(t^*_{\rm ps}) = \int_{-\infty}^{+\infty} \frac{1}{\tau k} \theta\left(\frac{t^*_{\rm ps}}{k}\right) \exp\left(-\frac{t^*_{\rm ps}}{k\tau}\right) h(k) {\rm d}k. \tag{VI.6}$$

Finally, to take into account the effect of the experimental resolution σ_t, and a possible offset t_0, on the measured pseudo decay time, the distribution is convolved with a Gaussian with mean value t_0 and variance σ_t^2.

$$\begin{aligned} f(t_{\rm ps}) &= f_{\rm ps}(t^*_{\rm ps}) * G(t_{\rm ps} - t^*_{\rm ps}; t_0, \sigma_t) \\ &= \int_{-\infty}^{+\infty} {\rm d}t' \int_{-\infty}^{+\infty} {\rm d}k\, h(k) \frac{1}{\tau k} \theta\left(\frac{t'}{k}\right) \exp\left(-\frac{t'}{k\tau}\right) \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(t_{\rm ps} - t_0 - t')^2}{2\sigma_t^2}\right) \\ &= \int_{-\infty}^{+\infty} {\rm d}k\, h(k) \int_{0}^{+\infty} {\rm d}t' \frac{1}{\tau k} \exp\left(-\frac{t'}{k\tau}\right) \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(t_{\rm ps} - t_0 - t')^2}{2\sigma_t^2}\right) \\ &= \int_{-\infty}^{+\infty} {\rm d}k\, h(k) \frac{1}{2k\tau} \exp\left(\frac{\sigma_t^2}{2k^2\tau^2} - \frac{(t_{\rm ps} - t_0)}{k\tau}\right) {\rm erfc}\left(\frac{\sigma_t}{\tau k\sqrt{2}} - \frac{(t_{\rm ps} - t_0)}{\sigma_t\sqrt{2}}\right), \end{aligned} \tag{VI.7}$$

where the symbol $*$ represents the Fourier convolution and erfc is the complementary error function.

Instead of using a single Gaussian-resolution model, a multiple-Gaussian pdf is preferred to describe the experimental resolution function. It is defined as

$$G(t) = \sum_{i=1}^{N_g} \frac{g_i}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{t - t_0^{(i)}}{2\sigma_i^2}\right) \tag{VI.8}$$

where the index i runs over the Gaussian contributions, and $\sum_{i=1}^{N_g} g_i = 1$.

For the linearity of the integral in Equation VI.7, the pseudo-proper decay time pdf becomes

$$f(t_{\rm ps}) = \sum_i g_i f(t_{\rm ps}; \sigma_i t_0^{(i)}). \tag{VI.9}$$

The distribution of the k-factor depends on $m(J/\psi\,\mu)$. Hence, a 2–dimensional distribution is used for the signal once the pdf $h(k)$ is parametrized in bins of $m(J/\psi\,\mu)$.

$$f_{\rm sig}(t_{\rm ps}, m_c) = m_{\rm sig}(m_c) \times f(t_{\rm ps}, m_c) \qquad \text{with} \quad m_c = m(J/\psi\,\mu) \tag{VI.10}$$

where $m_{\rm sig}$ is the normalized distribution of $m(J/\psi\,\mu)$, indicated simply as m_c to lighten the notation.

Resolution model

The pseudo-proper decay time $t_{\rm ps}$ is obtained from the measurement of the primary and secondary vertex positions and muon momenta. The errors, estimated from the vertex fitters, are propagated to an event-by-event error estimate $\delta t_{\rm ps}$. The extension of the data model to a third dimension for $\delta t_{\rm ps}$, in order to exploit the better time resolution on the events with lower $\delta t_{\rm ps}$ has been considered and found irrelevant in terms of statistical uncertainty on the lifetime measurement. On the other hand, the description of the template distributions for $\delta t_{\rm ps}$ and, notably, their correlation with $m(J/\psi\,\mu)$ and $t_{\rm ps}$ introduce a non-negligible source of systematic uncertainty.

The resolution function is obtained empirically by a fit to the $(t_{\rm ps} - t^*_{\rm ps})$ distribution in simulated data. The mean of each Gaussian contribution can be fixed to a unique value ($t_0^{(1)} = t_0^{(2)} = ... = t_0^{(N_g)} = t_0$), or left free to vary for each Gaussian contribution. In the latter scenario the number of parameters increases so that the number of Gaussian contribution that can be added to the final resolution model is limited by the stability of the fit. Table VI.1 summarizes the parameters of the resolution function obtained fitting different multi-Gaussian functions to the simulated dataset for the unique-mean model, the fit parameters obtained using free mean for each Gaussian contribution are reported in Table VI.2.

Figure VI.2 represents the fit to a simulated sample of $B_c^+ \to J/\psi\,\mu^+\nu$ decays used to define the parameters of the resolution function.

While the choice of a multi-Gaussian parametrization could be suboptimal, it leads the great advantage of analyticity in the expression of the decay time distribution as expressed in Equation VI.7.

Following a common practice in the experiment, it has been chosen to adopt a three-Gaussian model with common offset for the resolution function used to quote the central value, using the alternative models for systematic uncertainty assessment. Since the wider Gaussian contribution is most probably due to wrong-PV association, already taken into account with a dedicated pdf in this analysis, the parameters of the three-Gaussian resolution model are fixed to the values obtained in the fit of the four-Gaussian model, but setting $g_4 = 0$ and renormalizing the function.

N_g	1	2	3	4	5
t_0 [fs]	-0.5 ± 1.5	-2.3 ± 0.2	-2.0 ± 0.2	-1.9 ± 0.2	-1.9 ± 0.3
σ_1 [fs]	199.5 ± 1.0	35.3 ± 0.2	30.5 ± 0.3	27.1 ± 0.7	26.3 ± 0.6
g_2 [%]	—	1.86 ± 0.11	8.8 ± 0.9	24 ± 4	29 ± 5
σ_2 [fs]	—	(2.3 ± 0.1) 10^3	94 ± 5	54 ± 3	50 ± 2
g_3 [%]	—	—	0.82 ± 0.08	0.52 ± 0.07	2.0 ± 0.4
σ_3 [ps]	—	—	3.4 ± 0.2	4.4 ± 0.4	0.17 ± 0.02
g_4 [%]	—	—	—	1.6 ± 0.2	$0.05^{+0.10}_{-0.05}$
σ_4 [ps]	—	—	—	0.26 ± 0.04	1.0 ± 0.2
g_5 [%]	—	—	—	—	0.27 ± 0.6
σ_5 [ps]	—	—	—	—	5.9 ± 0.8

Table VI.1: Parametrization of the resolution function as obtained from the simulated sample with a multi-Gaussian data model. The uncertainties are statistical only. Note that while t_0, σ_1 and σ_2 are expressed in femtoseconds, σ_3, σ_4, and σ_5 are expressed in picoseconds.

N_g	1	2	3
$t_0^{(1)}$ [fs]	-0.5 ±1.5	-2.3 ±0.3	-1.6 ±0.3
σ_1 [fs]	199.5 ±1.0	35.3 ±0.2	30.5 ±0.4
$g_2(\%)$	–	1.8 ±0.1	9.0 ±0.9
$t_0^{(2)}$ [ps]	–	0.03 ±0.13	-17 ±3
σ_2 [ps]	–	2.31 ±0.10	93 ±5
$g_3(\%)$	–	–	0.83 ±0.07
$t_0^{(3)}$ [ps]	–	–	0.2 ±0.3
σ_3 [ps]	–	–	3.4 ±2.1

Table VI.2: Parametrization of the resolution function as obtained from the simulated sample with a multi-Gaussian data model, without constraints on the values of means $t_0^{(i)}$. The uncertainties are statistical only. Time units are not homogeneous alongside the table.

A dedicated study was performed to verify that there is no dependence of the parameters of the resolution function on the decay time itself.

VI.1.3 k-factor technique

The k-factor has been defined in Equation VI.4 as

$$k = \frac{t^*_{\rm ps}}{t^*}. \tag{VI.11}$$

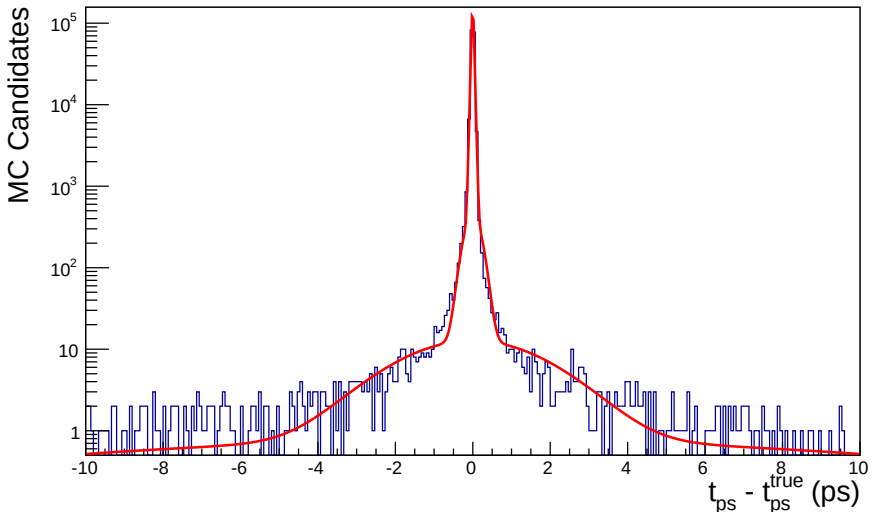

Figure VI.2: Distribution of the difference between the true and reconstructed pseudo-proper decay times as simulated for the decay $B_c^+ \to J/\psi\,\mu^+\nu_\mu$. A five-Gaussian data model, with parameters obtained from an unbinned maximum likelihood fit and means constrained to a unique value, is superposed. Parameters are listed in Table VI.1.

Neglecting the small difference between the measurement of the decay distance obtained as difference between the decay and production vertices, and its projection on the momentum of the $J/\psi\,\mu^+$ combination, this can be approximated by

$$k_{\text{kinematic}} = \frac{[\beta\gamma]_{B_c^+}}{[\beta\gamma]_{J/\psi\,\mu^+}} = \frac{p_{B_c^+}/m_{B_c^+}}{p_{J/\psi\,\mu^+}/m_{J/\psi\,\mu^+}}, \tag{VI.12}$$

which is found to be almost identical to k, with a relative error of $\mathcal{O}(10^{-5})$

A third definition sometime used in the literature is

$$k_{\text{time}} = \frac{t_{\text{ps}}}{t^*} \tag{VI.13}$$

which includes the effects of the experimental resolution in the definition of a k-factor describing the total smearing of the B_c^+ decay time measurement. The comparison between the three definitions is shown in Figure VI.3. It can be seen that the effect from incomplete kinematics dominates over the reconstruction effects.

Possible dependences of the k-factor distribution on the $J/\psi\,\mu^+$ momentum or on the decay time were searched and not found neither in the simulated $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ sample, nor in a real data set of $B^0 \to J/\psi\,K^+\pi^-$ reconstructed both fully and partially to perform a data-driven measurement of the k-factor distribution.

Instead, as already mentioned, the k-factor distribution is expected to depend on the reconstructed mass of the combination $J/\psi\,\mu^+$. A large value of $m(J/\psi\,\mu)$, close to $m(B_c^+)$, is associated to low-momentum neutrino, which introduces therefore a smaller ambiguity in the reconstruction of the B_c^+ rest frame. The k-factor distribution is therefore expected

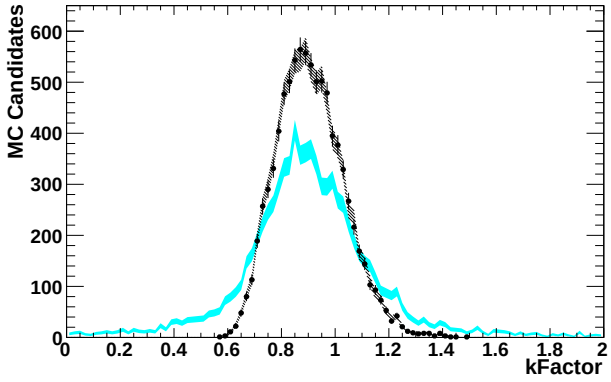

Figure VI.3: Comparison of the k-factor distributions as obtained with three different definitions. When defined as $k = t^*_{\rm ps}/t^*$ as in Equation VI.11, the k factor distribution is shown in black markers with error bars. The distribution is almost identical to that obtained for $k_{\rm kinematic}$ as defined in Equation VI.12, represented by the dashed area. Finally, the shaded area represents the distribution of the $k_{\rm time}$ variable including the smearing due to the experimental resolution.

to be narrower for larger values of $m(J/\psi\,\mu)$ and to broaden moving towards the lower ending point of the $m(J/\psi\,\mu)$ distribution.

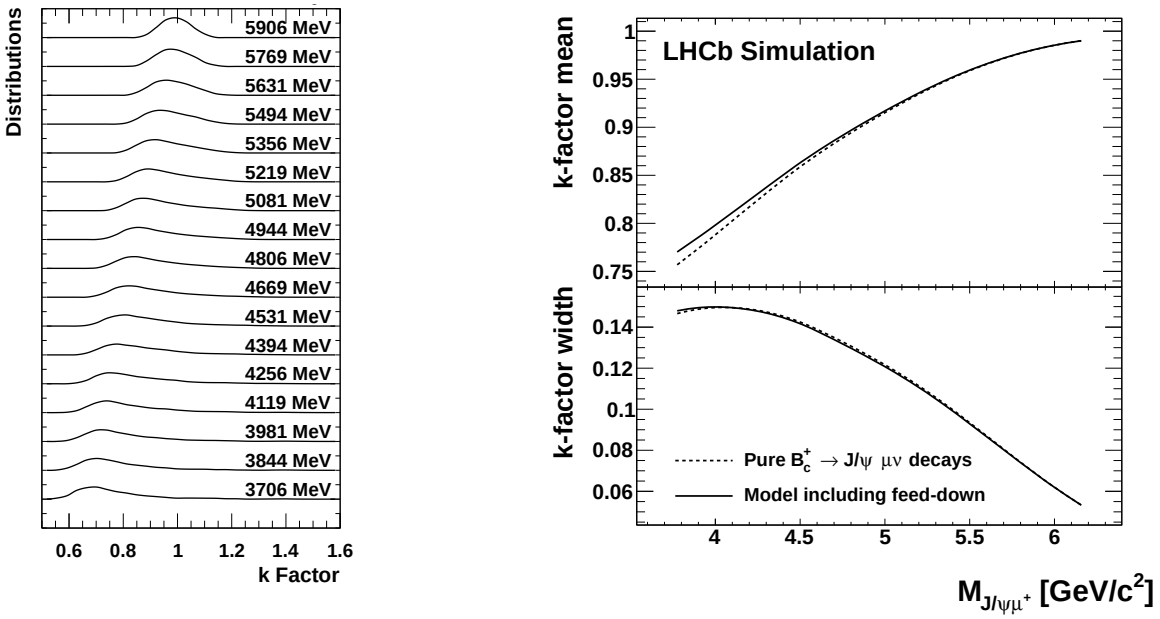

Figure VI.4: Dependence of the k-factor on the invariant mass of the combination $J/\psi\,\mu^+$. At left, the distributions are shown in bins of $m(J/\psi\,\mu)$ centered on the values reported on the right side. The right plot represents the mean and the width of the distributions shown at left, as a function of $m(J/\psi\,\mu)$. The correction due to feed-down decays is also shown. The plots have been obtained using a sample of simulated $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decays.

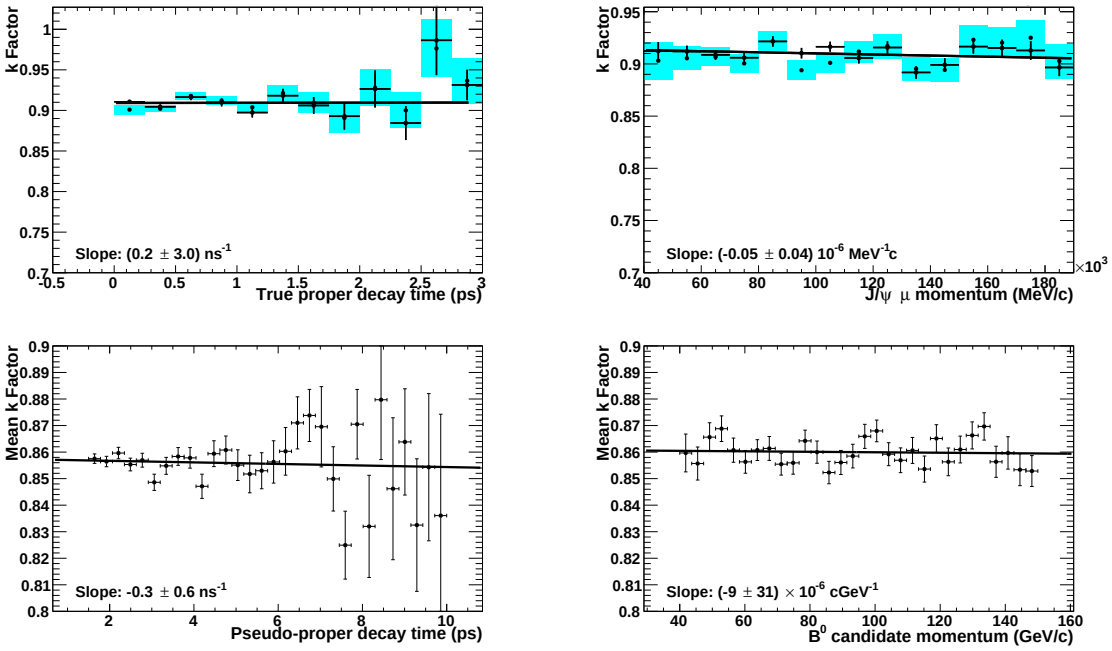

Figure VI.5: Dependence of the mean of the k-factor distribution on the decay time (left column) and on the momentum (right column). On the top a simulated sample of $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decays is used, and the definitions for the k-factor of Equations VI.11 (black marker) and VI.12 (cyan shaded area) are compared. In the bottom row the k-factor is calculated comparing the momentum of fully reconstructed $B^0 \to J/\psi\,K^-\pi^+$ decays, and of the combination $J/\psi\,K^-$.

The k-factor distributions are shown in 17 bins of $m(J/\psi\,\mu)$ in Figure VI.4. The mean and the width values of the distributions are also shown as a function of $m(J/\psi\,\mu)$. The correction due to feed-down decays, treated in the next section, is also shown and found small, but not negligible.

VI.1.4 Feed-down decays

The dominant decay composing the signal is certainly $B_c^+ \to J/\psi\,\mu^+\nu_\mu$, but other B_c^+ decays can provide an experimental signature similar to the studied decay. The following decays are explicitly included in the signal data model:

- $B_c^+ \to \psi(2S)\mu^+\nu$ with $\psi(2S) \to J/\psi\,X$,
- $B_c^+ \to \chi_{cJ}\mu^+\nu$ $(J = 0, 1, 2)$ with $\chi_{cJ} \to J/\psi\,X$,
- $B_c^+ \to h_c\mu^+\nu$ with $h_c \to J/\psi\,X$,
- $B_c^+ \to J/\psi\,\tau^+\nu$ with $\tau^+ \to \mu^+\nu\bar{\nu}$.

	$J/\psi\mu\nu$	$\psi(2S)\mu^+\nu$	$\{\chi_c, h_c\}\mu^+\nu$	$J/\psi\tau^+\nu$
InAcceptance	1 M	200 k	100 k	1 M
ϵ_{Acc}	0.309	0.304	0.140	0.302
Selected	8624	1241	1237	3449
$\epsilon_{Sel}(\times 10^3)$	8.62	6.21	12.37	3.45
R_{BR} [149]	1	0.05	0.33	0.25
$\mathcal{P}$(DecayChain)	1	0.609	0.099	0.177
Contribution to the final sample	0.943	0.020	0.020	0.017

Table VI.3: Summary of the contributions to the signal MC sample of the four considered decays. First the number of events ("InAcceptance") generated with a J/ψ inside the detector acceptance are reported. In order to optimize the MC generation and reconstruction procedure only for the $B_c^+ \to \{\chi_c, h_c\}$ sample a special tool, requiring three muon from the B_c^+ inside the detector acceptance, was developed and applied. The efficiency of the generator–level acceptance cut is also shown: ϵ_{Acc}=InAcceptance/N_{Gen}, where N_{Gen} is the total number of generated events. "Selected" and ϵ_{Sel} represent the number of events which survive to all the applied cuts and the corresponding efficiency: ϵ_{Sel}=Selected/InAccepance, respectively. The various feed-down contributions are normalized to the $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ channel using the relative branching fraction, $R_{BR} = \mathcal{B}(B_c^+ \to i)/\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu)$ and their weight is obtained by multiplying the efficiencies by the probability $\mathcal{P}$(DecayChain), of producing, via a subsequent decay, the same particle configuration of the signal, i.e. a J/ψ and a μ in the final state. The contribution of every channel is finally reported.

These *feed-down* decays produce two effects: first they modify the mass shape of the signal contribution since the ending point is different (more missing mass). Second, they change the k-factor distribution, especially in the lower mass bins.

Table VII.3 lists the numbers of events generated and selected for the four decays considered. To evaluate the expected yield in the final sample, the relative branching fractions of the B_c^+ decays have to be assumed, since none of these decays has been observed so far. For $B_c^+ \to \psi(2S)\mu^+\nu$, the Kiselev prediction of 4.9% is used [56]. The estimation of the contribution of χ_c and h_c relies on the latest theoretical predictions [145] and on the assumption of a conservative branching fraction of 1% [146] for the once seen, but not confirmed $h_c \to J/\psi\pi^0$ decay [147]. The use of an alternative prediction [148] provides a similar contribution to the signal from this source. Since no experimental confirmations of these branching fractions are available, the possible bias introduced by the assumed values is assessed by varying the model within the range given by other available theoretical calculations, and comparing the model with and without each feed–down contribution.

Other potential feed-down sources are considered, as the $B_c^+ \to J/\psi\, D_{u,s}^{(*)}$ decay followed by a semileptonic decay of the charm meson. However, this background is reduced to a negligible level by the tight p_T requirement on the bachelor muon and by the χ^2 vertex criterion. The expected contribution of these decays have been calculated to be one order of magnitude smaller than the other included feed-down contributions.

Another potential source of feed-down decays is due to $B_c^+ \to B_s^0 \mu^+ X$ decays with the B_s^0 decaying to a final state including a J/ψ. These decays are expected to be relatively abundant with respect to B_c^+ decays to open charm final states, and therefore they constitute a more severe issue despite of the longer lifetime of the B_s^0 meson which translates into a statistical worsening of the χ^2 distribution. The branching ratio is expected to be at least one order of magnitude smaller than $\mathcal{B}(B_c^+ \to J/\psi\,\mu^+\nu)$. A further suppression of this mode is obtained with the kinematical selection. A study at generator level, which emulates the stripping phase, shows that the selection efficiency is only 0.05%. Again, a strong rejection is provided by the cut on the transverse momentum of the bachelor muon. Finally, because of the large mass of the B_s^0 meson, the upper ending point of the $m(J/\psi\,\mu)$ distribution for this feed-down contribution is around 4.4 GeV/c^2. In other words, the decays $B_c^+ \to B_s^0 X$ populate a region of the dataset which is dominated by the misidentification background so that any candidate possibly surviving would have negligible statistical weight in the determination of the lifetime. Considering the large uncertainty on the background due to b-hadron decays associated to random muons, which is conceptually very similar to the random association of a B_s^0 meson associated to a real muon eventually originated in the same B_c^+ decay, it has been decided to neglect this source of feed-down at the price of a more conservative estimation of the systematic uncertainty on the model for detached combinatorial background.

From the experience of $B_{u,d}$ semileptonic studies, it is known that non-resonant decay can also lead to a pollution of the sample which is very difficult to quantify theoretically, and for which no measured value is available, yet. According to a generator-level study, generating the non-resonant $B_c^+ \to J/\psi\,\mu^+\nu\pi^0$ decay with a phase-space model, the invariant mass distribution of the $J/\psi\,\mu^+$ combination lays between the $J/\psi\,\tau^+\nu$ and $\psi(2S)\mu\nu$ contributions. Because of the ignorance of the actual shape and yield of this contribution, it has been chosen not to perform a full simulation of the decay mode, which would lack of significance, but rather to take it into account in the systematic uncertainty by allowing large variation of the yield for the other two feed-down components, as discussed in the next Chapter.

Figure VI.6 represents the invariant mass distribution of the $J/\psi\,\mu^+$ combination as simulated using the Kiselev form-factor models [33], and after the correction due to inclusion of the feed-down decay modes considered in this analysis.

To summarize, the following two–dimensional *pdf* is obtained from simulation to describe the ($t_{\rm ps}$, $m(J/\psi\,\mu)$) distribution of the signal in the chosen baseline model.

$$\begin{aligned} f_{\rm sig}(t_{\rm ps}, m(J/\psi\,\mu)) = m_{\rm sig}(m(J/\psi\,\mu))\times \\ \int_{-\infty}^{+\infty} h(k; m(J/\psi\,\mu))\frac{1}{2k\tau}\times \\ \sum_{i=1}^{3} g_i \exp\left(\frac{\sigma_i^2}{2k^2\tau^2} - \frac{(t_{\rm ps}-t_0)}{k\tau}\right)\mathrm{erfc}\left(\frac{\sigma_i}{k\tau\sqrt{2}} - \frac{(t_{\rm ps}-t_0)}{\sigma_i\sqrt{2}}\right). \end{aligned} \tag{VI.14}$$

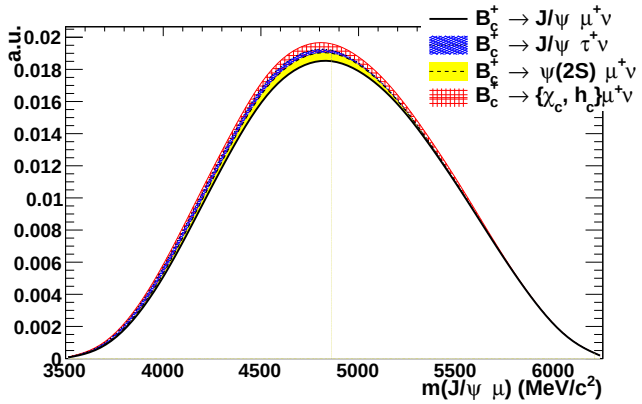

Figure VI.6: Distribution of the invariant mass of the $J/\psi\,\mu^+$ combination as obtained by the pure $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay mode, simulated with the Kiselev form-factor model [33], and corrected by the feed-down decay modes considered in the analysis.

The distribution of the k-factor and of the invariant mass $m(J/\psi\,\mu)$ are obtained from the realistic dynamic models previously described, that includes the feed-down contributions, and takes into account the experimental efficiency, which is assumed to depend only on $m(J/\psi\,\mu)$. In the nominal model, a triple-Gaussian function with a common offset is used to describe the average pseudo-proper time resolution for all events. The alternative resolution models described are used for the assessment of the systematic uncertainty.

VI.2 Misidentification background

The *misidentification background* is defined as the set of candidates selected because of a wrong identification of a hadron as a bachelor muon. Hadrons from $B_{u,s,d} \to J/\psi\,X$ decays or emitted from the primary vertex and combinatorially associated to a J/ψ are included in this background source.

There are two important aspects when dealing with background sources: *rejection* and *modelling*. The latter is essential to perform some kind of statistical background-subtraction from the final sample. However the statistical uncertainty can be only controlled if the contribution of the background with respect to the signal is not too large, and therefore efficient background rejection is needed.

The techniques to enhance misidentification background rejection are described together with the selection strategy in Chapter V, and only briefly summarized here, before focusing on the model for the misidentification background.

VI.2.1 Rejection

The rejection of the misidentification background relies on the following criteria, whose subsequent effect is represented in Figure VI.7, in the order they are presented:

- the binary IsMuon condition has to be verified (black);
- the $\mathrm{DLL}_{\mu/\pi}$ of the bachelor track, obtained combining the information of several the detectors including the muon system and the RICH detectors has to exceed a tight threshold (red);
- at least four muon stations must have hits matched in both x and y views to the candidate muon track (green);
- the χ^2_{match}, test statistics for the good matching of the muon track as reconstructed upstream and downstream of the magnet has to be smaller than a threshold value to suppress hadrons decaying in flight while in the magnet (blue);
- the χ^2 obtained with the application of a Kalman filter to the standalone fitting in the muon system has to be smaller than 1.5 per degree of freedom (yellow);
- the $\mathrm{DLL}_{p/\pi}$ of the muon candidate has to be small in order to reject protons successfully identified by the RICH detectors (purple);
- the $\mathrm{DLL}_{K/\pi}$ of the muon candidate has to be small in order to reject kaons successfully identified by the RICH detectors (cyan).

The misidentification probabilities for pions, kaons and protons are obtained from calibration samples of hadrons whose PID is kinematically identified by the structure of the studied decay. For protons, Λ^0 decays to $p\pi^-$ are selected without PID requirements on the daughter tracks, while for kaons and pions decays $D^{*+} \to D^0\pi^+$ with the D^0 meson decaying to $K^-\pi^+$ are used. The charge of the pion from D^{*+} is used to identify the kaon and the pion.

In order to account for the kinematic differences between hadrons in the calibration and in the signal samples, the misidentification probabilities are measured in bins of momentum, pseudorapidity and number of tracks.

The events selected from calibration samples are required to have the *probe* particle in the muon system acceptance, and to have been triggered independently of it. To subtract the background from the calibration samples a technique, named *Fit&Count* was used, which is simple, robust and widely used in the Collaboration.

The simplest approach to measure misidentification probabilities would be to fit the shape of the background subtracted mother mass distribution (D^{*+} or Λ^0) before and after the application of the muon identification requirements. However, the low statistics available after the application of the requirements makes the fit unreliable. Moreover, the shape parameters would be fixed so that a variation in the resolution or in the mass mean value due to correlation of the PID variables with kinematics, would result in an underestimation of the selection efficiency.

As an alternative, the total muon identification efficiency ϵ_{tot}, for signal and background together, and for background only (ϵ_{bkg}) for far mass sidebands, are measured. A fit of the bin content without muon identification requirements is performed to obtain the total number of signal and background candidates, N_s and N_b, respectively.

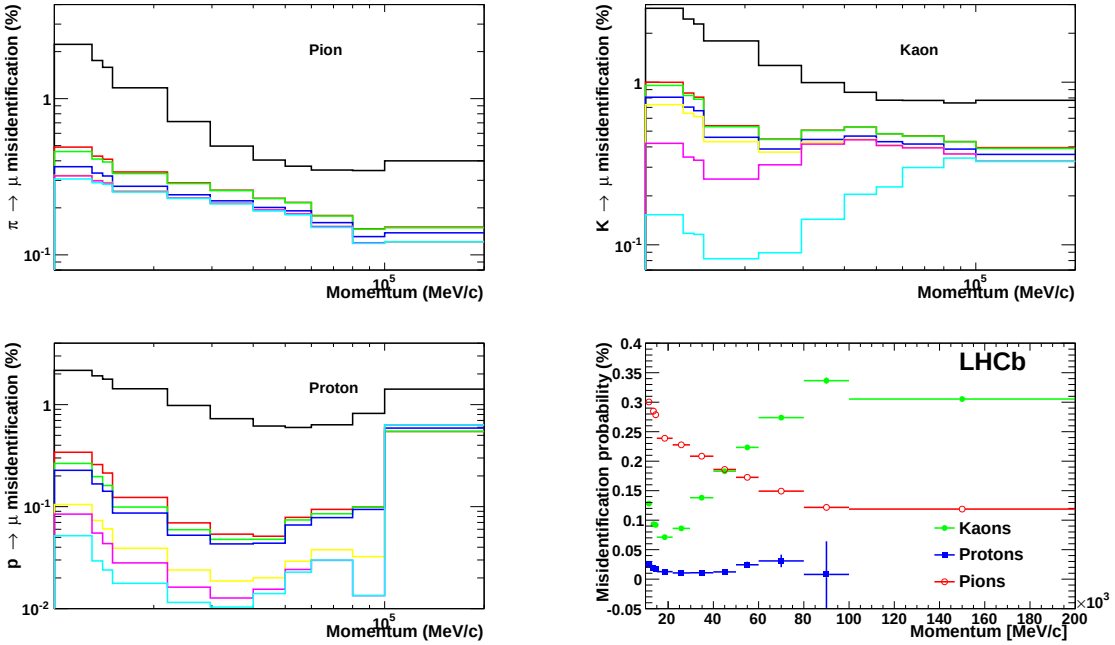

Figure VI.7: Efficiency as a function of the transverse momentum at the subsequent application of the criteria listed in Section VI.2.1 to a pure samples of pions (top-left), kaons (top-right), and protons (bottom-left). The bottom-right plot summarizes the efficiencies (or misidentification probabilities) obtained combining all the requirements and represents it in with linear scales.

The signal efficiency, *i.e.* the misidentification probability, is then obtained as

$$M_h = \epsilon_{\text{sig}} = \frac{N_s + N_b}{N_s}\epsilon_{\text{tot}} - \frac{N_b}{N_s}\epsilon_{\text{bkg}}. \tag{VI.15}$$

The uncertainty is evaluated taking into account the correlation between quantities.

The misidentification probabilities for protons, kaons and pions obtained with this method for the muon identification technique illustrated above, are shown in Figure VI.7, in bins of p.

VI.2.2 Model

In order to model shape and normalization of the $(t_{\text{ps}}, m(J/\psi\,\mu))$ distribution of this source of background, a data sample of decays $B_c^+ \to J/\psi$ `track` is used. It is obtained from the standard data sample, to which all selection criteria are applied with the exception of PID requirements for the bachelor muon. This dataset will be referred as *NoMuonID*. The expected contribution of the misidentification background is then obtained applying to each event a weight W, defined as the weighted average misidentification probability M_h among hadron species h, the weight being the probability P_h for the track to be a hadron

Momentum p [GeV/c]	13	14	15	22	29	40	50	60	80	100	150
Pseudorapidity η	1.5	2.7	3	3.2	4.35	5.0					
Impact Parameter IP [mm]	0.00	0.05	0.10	0.20	0.50	5.0					
Number of tracks (`nTracks`)	0	70	100	200	300	500	800				

Table VI.4: Boundaries used to bin M_h and P_h when building the data model for the misidentification background.

h.

$$W = \sum_{h=\mathrm{h,K},\pi} P_h(\eta, p, \mathrm{IP})\, M_h(\eta, p, \texttt{nTracks}), \tag{VI.16}$$

where h can be a kaon, a pion or a proton. The contribution from leptons to the *NoMuonID* sample is neglected because it is expected to be below 1%. The resulting misidentification background yield will be therefore overestimated by the same relative amount, which is negligible with respect to other uncertainties. For the misidentification probabilities M_h the *Fit&Count* technique of Equation VI.15 is used, while the values for P_h are obtained from the distributions of the PID variables, calibrated from the same control samples used for the determination of M_h.

The distributions of momenta and rapidity of the hadrons obtained from the calibration samples are in general different from those of hadrons included in the *NoMuonID* dataset. Therefore, they are parametrized as a function of the main variables on which they are expected to depend. For P_h, these are the track momentum p, its pseudorapidity η, and the impact parameter (IP) with respect to the primary vertex. The latter is needed to account for the different composition of the decays contributing to the background at different decay time: prompt events will dominate at low t_{ps}, while some exclusive B decays dominate the detached events. For M_h, as described above, the variables are p, η and the number of tracks in the event `nTracks`. The particle identification detectors,in particular those based on the Cherenkov effect, show better performance for clean events with few tracks. Rare events with more than 800 tracks are eliminated, and events with less tracks are divided in 8 bins, as reported in Table VI.4.

The estimation of the probability for a track to be a hadron h ($h = \mathrm{p, K}, \pi$) is performed in bins of p and η. The boundaries used are reported in Table VI.4. Candidates with momentum lower than 13 GeV/c are removed from the sample since background dominates in that region, and the amount of ghosts tracks is not negligible. Rare candidates with momentum p larger than 150 GeV/c are also eliminated. The pseudorapidity boundaries, listed in Table VI.4, cover the whole detector acceptance and therefore no candidates are found outside this range.

In order to determine the fraction of pions, kaons and protons in each bin, the differential log-likelihood $\mathrm{DLL}_{K/\pi}$ and $\mathrm{DLL}_{p/\pi}$ are used. The two–dimensional distribution $\mathrm{DLL}_{K/\pi} \perp \mathrm{DLL}_{p/\pi}$ of the bachelor track is fitted to the sum of the three hadron components obtained from the calibration samples described above. The *sPlot* [150] technique was applied to disentangle the background contributions affecting the calibration samples using

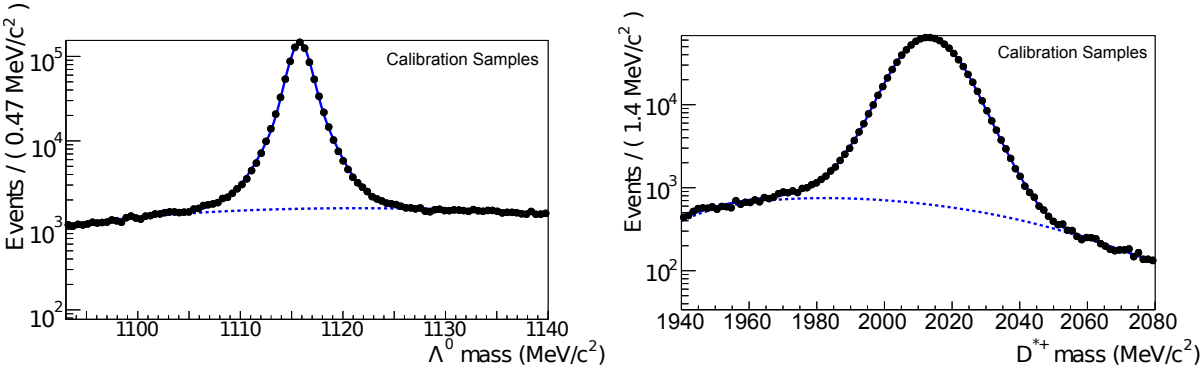

Figure VI.8: Mass fits to the calibration samples $\Lambda^0 \to p\pi^-$ and $D^{*+} \to \pi^+ D^0$ with $D^0 \to K^-\pi^+$ used to determine the template *pdf*s for protons, pions and kaons on the plane $\mathrm{DLL}_{K/\pi} \perp \mathrm{DLL}_{p/\pi}$.

the invariant mass of the D^{*+} and Λ^0 particles as discriminating variables.

The distribution of the invariant mass of the Λ^0 candidates is modeled using a double Crystal ball function [63], the background is modeled with a Chebychev second-order polynomial *pdf*. For the D^{*+} mass, the signal is modeled as sum of a Crystal ball and a Gaussian; the combinatorial background is modeled with a third-order Chebychev polynomial *pdf*. The mass fits are shown in Figure VI.8.

The resulting joint distributions $f_h(\mathrm{DLL}_{K/\pi}, \mathrm{DLL}_{p/\pi})$ are used, as template models, in a bidimensional likelihood fit to the same variables for the bachelor `track` in the *NoMuonID* sample. The fit is performed in bins if p, η, `nTracks` and IP. While the RICH performances (the template *pdf*) are expected to change as a function of `nTracks`, the functions of protons, kaons and pions are not expected to be strongly correlated with the number of tracks per event. The fit is thus performed simultaneously on each `nTracks` bin, using the appropriate reference distribution, but constraining the fractions to be the same for all bins. On the other hand, the detachment from the primary vertex can modify the fraction of kaons, protons and pions in the bin, but it is not supposed to modify the response of the PID detectors. The fit is therefore performed in bins of the track Impact Parameter (IP), using the same PID pdf for all IP bins. Boundaries are listed in Table VI.4.

In order to avoid biased due to correlation between the hadron species and the trigger efficiency, all events in the calibration samples are required to be *Triggered Independently of the Signal* (TIS). The technique is illustrated in Section IV.2.2.

Figure VI.9 illustrates an example of the fit result for a randomly chosen bin.

VI.2.3 Construction of the data model

The P_h and M_h factors are finally merged in a weight table, following the common binning structure in p, η, `nTracks` and IP. The two-dimensional probability density function of (t_{ps}, $M_{J/\psi\,\mu^+}$) representing the misidentification background component is obtained from the distribution of the *NoMuonID* sample after applying the weight W as defined in Equation VI.16. The *pdf* is obtained by smoothing the binned 2D histogram. The expected yield of

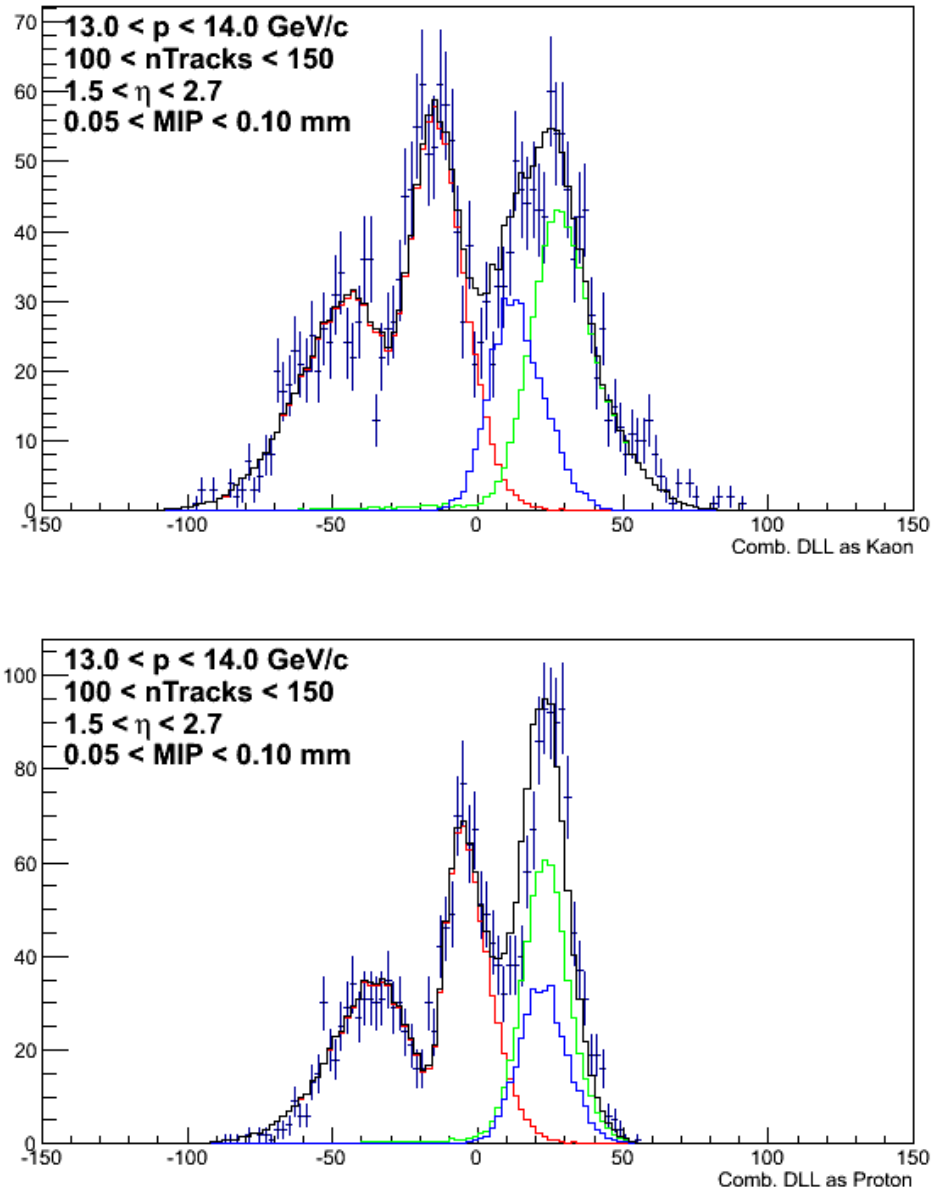

Figure VI.9: An example of $\mathrm{DLL}_{K/\pi} \perp \mathrm{DLL}_{p/\pi}$ fit. The plots show the projections in the two PID variables. Template *pdf*s extracted from the calibration samples are represented in blue for protons, green for kaons, and red for pions. The free parameters of the fit are the fractions of kaons and pions, while the fraction of protons is constrained by the normalization $\sum_h P_h = 1$. The double-peaked structure of the pions is due to the combined usage of the two RICH detectors.

the misidentification background in the final fit is given by the total sum of the weights, corresponding to 10987 ± 110, where the uncertainty is statistical only, corresponding to an average misidentification probability of 0.20%.

The whole technique was validated on a simulated sample of b-hadron decaying to final states including a J/ψ. The shape and normalization of the background are properly reproduced.

VI.2.4 Uncertainties

The uncertainty on the background normalization has been estimated starting from the error on the M_h and P_h factors, as obtained from the *Fit&Count* technique, and from the fit of the hadronic components. The errors are then propagated to assess the uncertainty on the track-by-track weight as:

$$\delta W = M_p\delta P_p + P_p\delta M_p + \sqrt{(P_\pi\delta M_\pi + M_\pi\delta P_\pi)^2 + (P_K\delta M_K + M_K\delta P_K)^2 - 2(M_\pi M_K\delta P_\pi\delta P_K)} \tag{VI.17}$$

which relies on the assumption that the number of protons in much smaller than the number of kaons and pions, and thus that there is full anticorrelation between the kaon and pion fractions. Each bin of the weight table is then associated to its uncertainty. The uncertainty on the global normalization is obtained multiplying the errors of each bin by the number of tracks of the *NoMuonID* sample in that bin, and summing the result in quadrature. In this way the correlation between the weight associated to tracks in the same bin is properly taken into account, while the weights of different bins are considered uncorrelated.

VI.3 Other Background sources and their models

VI.3.1 Fake J/ψ background

J/ψ resonances are reconstructed from the decay $J/\psi \to \mu^+\mu^-$. Fake J/ψ can be reconstructed either because of track misidentification as daughter muons, or because of a combinatorial association of two true muons. In both cases the invariant mass distribution of fake J/ψ daughters is not expected to peak at the J/ψ nominal mass. In this analysis the dimuon mass is required to lay in the region with boundaries -30 and +35 MeV/c^2 with respect to the J/ψ nominal mass, this corresponds to about twice the experimental resolution. The yield of fake J/ψ candidates within such signal window was estimated extrapolating into this region the background yield determined using the sidebands. This is achieved through an exponential fit to events with dimuon mass values more than 60 MeV/c^2 away from the nominal J/ψ mass. The determined fake-J/ψ contamination amounts to 4102 ± 64 events. This value is corrected for the bachelor muon misidentification yield as described below, and constrained in the final fit. Figure VI.10 shows the dimuon invariant mass distribution of real data. The events in the J/ψ mass sidebands are assumed to model correctly the background due to fake J/ψ in the signal region. The distributions of the J/ψ μ^+ mass and B_c^+ decay time for the two sidebands are found consistent within the statistical uncertainties. Possible small differences, due to different average momentum in the higher and lower sidebands are expected to be negligible since both sidebands are used and contribute with similar statistics. The systematic effect on the lifetime measurement due to this approximation was assessed and found small.

A two-dimensional non-parametric Gaussian-kernel *pdf* [151] is used to describe the contribution without neglecting possible correlation between the invariant mass of the combination $J/\psi\,\mu^+$, and the pseudo-proper decay time $t_{\rm ps}$. In order to dilute the contribution of very detached events, which would be modeled by the Kernel *pdf* as isolated Gaussian bumps without a real physical meaning, the Kernel *pdf* is sampled in bins of different size. Preserving resolution where large statistics is available, and forcing an unimodal distribution in the tails. The distribution obtained for $t_{\rm ps}$ is shown in Figure VI.11.

Remove candidates with fake J/ψ and fake bachelor muon.

Events with a fake J/ψ candidate with mass consistent with the nominal J/ψ mass, and associated to a misidentified hadron are included in both the fake J/ψ and misidentification background sources. In order to avoid double counting, the latter component is subtracted from the fake J/ψ background model.

The same technique used to evaluate the misidentification background contribution in the $J/\psi\,\mu^+$ sample is applied to the J/ψ candidates in the mass sidebands. Namely to weight each event in the J/ψ mass sidebands removing the requirements on muon identification by the weight W, defined in Equation VI.16. The *pdf* obtained represents the distribution of events having a fake J/ψ and a misidentified muon, and it is subtracted from the fake-J/ψ model *pdf*.

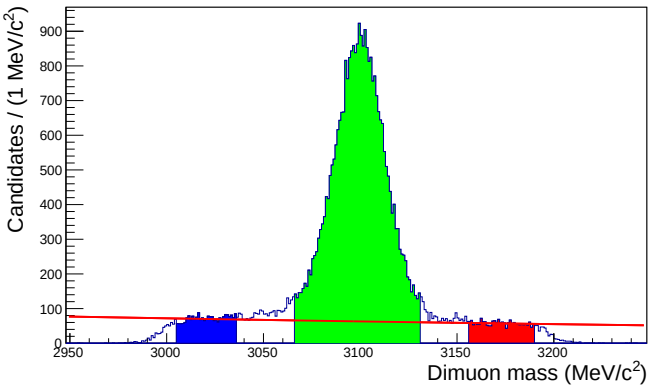

Figure VI.10: Dimuon invariant mass on real data. Sidebands shaded in blue and red, are used to determine the parameters of an exponential function (superimposed) through a fit procedure. The estimation of the number of background events is obtained by integrating the fit function over the signal region. The number of candidates in the colored sidebands is 4178, the number of candidates in the signal box is 30700, while the whole sample contains 40576 candidates. The borders of the distribution are not sharp because the requirement on the invariant mass of the $\mu^+\mu^-$ combination is applied before performing the vertex fit. The latter modifies slightly the value of the dimuon mass and introduces these resolution effects.

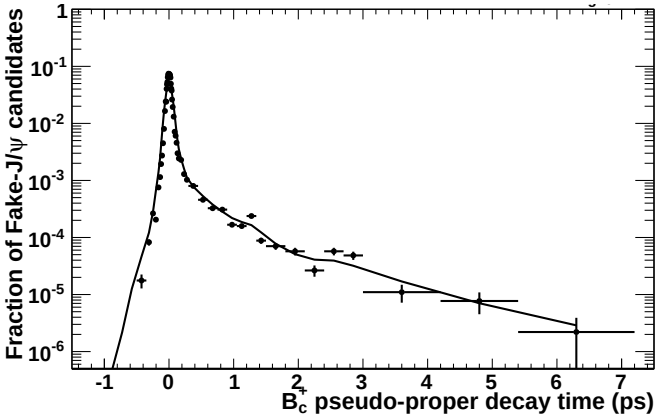

Figure VI.11: Distribution of the pseudo-proper decay time $t_{\rm ps}$ for the events in the J/ψ mass sidebands. The solid line represents the *pdf* obtained with a sampled kernel density estimation.

VI.3.2 Combinatorial background

Combinatorial background is defined as the set of candidates selected because of a random association of a true J/ψ to a real muon. If the J/ψ is produced in the pp collision then it will fit the description of a *prompt combinatorial background*, while if it is produced in b-hadron decay, it will fit the description of *detached combinatorial background*.

Prompt combinatorial background

There are two possible strategies for dealing with prompt combinatorial background. Rejection or modelling.

Rejecting prompt background is easy because it is sufficient to consider only events whose reconstructed $t_{\rm ps}$ exceeds some threshold. On the other hand the choice of the threshold introduces arbitrariness in the measurement, and because of the short lifetime of the B_c^+ meson, hardening the threshold is not for free in terms of signal statistics. Therefore it has been preferred to use this method for the assessment of systematic uncertainties, quoting the best value for the B_c^+ lifetime using the whole $t_{\rm ps}$ distribution and including in the fit a model for the combinatorial prompt contribution.

Modelling is based on a *pdf* extracted from data neglecting the correlation between $t_{\rm ps}$ and $m(J/\psi\,\mu)$. The decay time projection is modeled with a Gaussian *pdf* whose parameters are free in the global fit. Alternative models will be considered to evaluate systematic uncertainties. The mass projection is obtained from the prompt peak, with a tight cut $-500 < t_{\rm ps} < 10$ fs to remove the signal component and enforcing the muon identification requirements for the bachelor muon to $\mathrm{DLL}_{\mu/\pi} > 5$ to suppress the contamination from misidentification background.

The candidates selected with the standard analysis, plus the two requirements listed above, are fitted with a polynomial *pdf* to obtain a smooth model. The two-dimensional *pdf* is obtained as direct product of the two projections.

Detached combinatorial background

Detached combinatorial background is the only background source for which the data model must rely on simulation. An 18 million event sample of simulated $H_b \to J/\psi\, X$ decays with realistic branching fractions was generated. The J/ψ and bachelor muon candidates are required to be "truth-matched", *i.e.* it required that each reconstructed track is close enough to a simulated muon track, and that the two muon tracks associated to the J/ψ daughter, were actually produced in a J/ψ decay it the simulation. Most of the candidates surviving the selection are of combinatorial nature: none of them has the three muons coming from the same hadron decay, with the exception of $B_s^0 \to J/\psi\,\phi$ decays, with both the J/ψ and the ϕ decaying to dimuon, giving a contribution of around 2% to the combinatorial background model. The projected time and mass distributions from the modest background sample obtained from the simulation are shown in Figure VI.12. To model the pseudo-proper decay time distribution, a double exponential was used, while the mass was modeled with a Gaussian kernel *pdf*.

The yield of the expected contribution to the final sample was evaluated from the number $N_{\rm sel}$ of selected events (including trigger), over the number of generated events $N_{\rm gen}$, scaled to the number of produced mesons for the integrated luminosity of our data

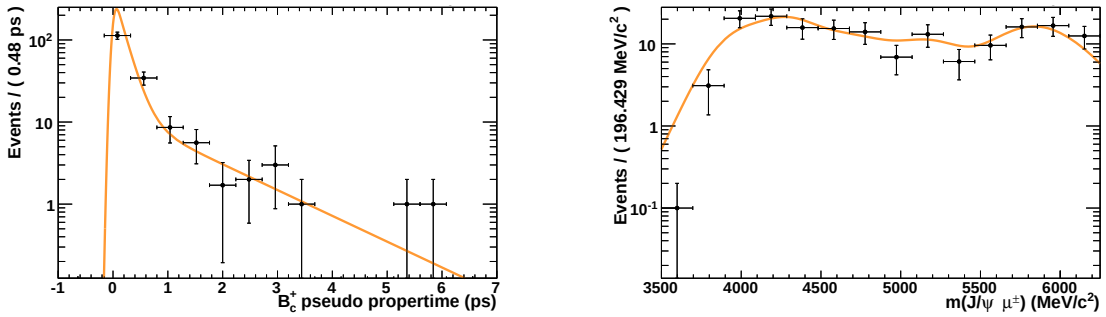

Figure VI.12: Combinatorial background from true J/ψ combined with a true muon, estimated using 18 millions of simulated $H_b \to J/\psi X$ decays. Time projection on the left, mass projection on the right plot.

sample:

$$N_{\text{comb}} = \frac{N_{\text{sel}}}{N_{\text{gen}}} \times \int_{2012} \mathrm{d}t\, \mathcal{L} \times \sigma(J/\psi \text{ from } b) \times \epsilon_g \times \mathcal{B}(J/\psi \to \mu^+\mu^-) = 974 \pm 168\,(\text{stat}), \tag{VI.18}$$

where the production cross-section $\sigma(J/\psi \text{ from} b)$ at $\sqrt{8}$ TeV measured by LHCb was used [105]. The correction factor $\epsilon_g = 0.82$ takes into account the different requirements on the pseudorapidity and p_T of the J/ψ in [105] with respect to the generator level requirements applied in to the H_h sample. The error is dominated by the statistical uncertainty on the ratio $\frac{N_{\text{sel}}}{N_{\text{gen}}}$.

In order to provide a data-driven cross-check to the simulated data model, events with a reconstructed $m(J/\psi\,\mu)$ beyond the nominal mass of the B_c^+ meson were used. Candidates with $m(J/\psi\,\mu) > m(B_c^+)$ are expected to be dominated by combinatorial background because all other background contribution are kinematically forbidden.

The results are shown in Figure VI.13. The selection of simulated events yields 37 events, corresponding to an expected yield of (201 ± 73) in data. Applying the same selection in data yields 221 ± 14 candidates, where the uncertainty is statistical only. The two values are consistent and also the shapes agree within the large statistical uncertainties of the samples.

Despite the excellent result of this cross-check, the simulation of particle identification variables is known to be not totally reliable. A conservative uncertainty of $\pm 50\%$ on the predicted yield of this background is therefore considered when assessing the systematic uncertainties, as it will be discussed in Chapter VII.

Incorrect primary vertex association

A source of background found relevant in other similar studies is the association of the B_c^+ candidate to the wrong primary vertex. In this scenario, the pseudo-proper decay time projection of the *pdf* is almost flat and extends symmetrically at positive and negative times.

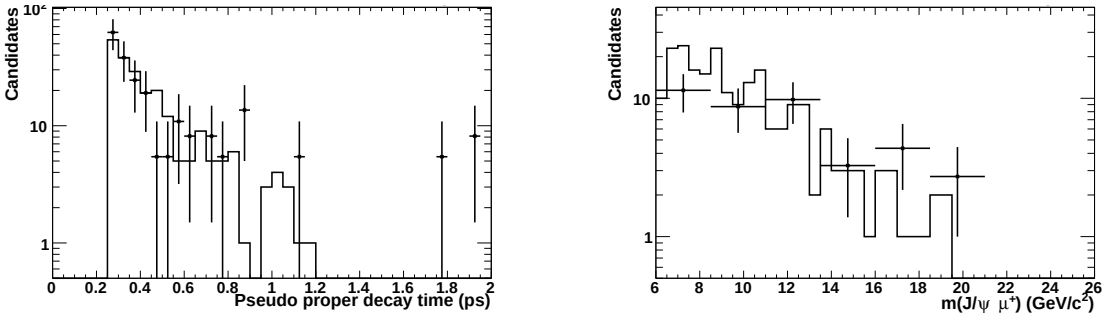

Figure VI.13: Data-driven cross-check of the detached combinatorial background, obtained comparing the yield and distributions of predicted and observed beyond the mass of the B_c^+ meson.

For $B_{u,d} \to J/\psi X$ analyses, the effect was treated by applying a veto to the next-to-best primary vertex, requiring the impact parameter of the B meson with respect to the second best primary vertex to be larger than 50 [152]. In this way, ambiguities are removed, but events for which the correct primary vertex is not reconstructed are not. An alternative technique is to take this background into account in the data model. The background distributions are obtained from data associating the selected B_c^+ candidate in one event to the primary vertex of the previous selected event. This is equivalent to choosing randomly the position of the primary vertex according to a realistic spatial distribution.

The yield of the contribution is free in the final fit and is well constrained by the far negative tail of the $t_{\rm ps}$ distribution, where other contributions are negligible.

VI.4 Fit to data

The B_c^+ lifetime is determined from a maximum likelihood unbinned fit to the joint $f(t_{\rm ps}, m(J/\psi\,\mu))$ distribution of the selected sample, in the range $-1.5 < t_{\rm ps} < 8$ ps and $3.5 < m(J/\psi\,\mu) < 6.25\,\text{GeV}/c^2$.

The fit technique adopted is a maximization of the log-likelihood with respect to the free parameters. The likelihood is not extended with a Poissonian term for the total normalization, which instead is constrained to the total number of candidates.

The free parameters of the fit are:

- the B_c^+ proper lifetime;
- the Signal yield;
- the mean and the width(s) of the pseudo-proper decay time resolution function;
- the number of candidates with wrong PV association.

The B_c^+ proper lifetime has been blinded during the development of the analysis, by adding a constant and unknown randomly generated offset to the value displayed during the

fit procedure. Once the internal reviewers agreed on the quality of the analysis including the assessment of the systematic uncertainties, the B_c^+ lifetime has been unblinded and the result published without further adjustments.

The global fit is performed using the Minuit-MIGRAD package [153]. The probability distribution functions for all the background components, except the prompt one, are sampled once over a grid of 10000 ($t_{\rm ps}$) $\times$ 1000 ($m(J/\psi\,\mu)$) points. Linear interpolation is used to evaluate the probability associated to each candidate. The normalization of the *pdf*s for these contributions is evaluated by numerical integration before the fit, and since there are no free parameters, the normalization is not recalculated at further iterations.

The signal and prompt background contributions do have free parameters, notably the B_c^+ lifetime, as well as the mean and width of the prompt peak which are let free. For the prompt background component, the correlation between $t_{\rm ps}$ and $m(J/\psi\,\mu)$ is neglected, so that the normalization can be computed as product of separate 1D integrals on the two variables. The mass projection is integrated numerically at the beginning of the fit procedure since it is constant. The decay time *pdf* is described by an analytical expression with a known normalization, recomputed through analytical integration at each step.

The same applies to the signal component: the $m(J/\psi\,\mu)$ distribution is fixed, while the $t_{\rm ps}$ distribution, which in this case depends on the $m(J/\psi\,\mu)$ value is recomputed at each step.

For the yields of the misidentification, combinatorial and fake-J/ψ backgrounds, we apply a Gaussian constraint to allow for Poissonian fluctuations around their expected values, which are fixed. The prompt peak yield is constrained by the closure relation

$$\sum \text{yields} = \mathrm{N}_{\rm tot}, \tag{VI.19}$$

with $N_{\rm tot}$ being the number of candidates in the sample.

VI.4.1 Fit result

The central value of the fit for the B_c^+ lifetime is

$$\tau_{B_c^+} = (508.7 \pm 7.7\,(\text{stat}))\ \text{fs} \tag{VI.20}$$

from a number of signal candidates of 8995 ± 103. The total number of background candidates is $20\,760 \pm 120$, of which 2585 have $t_{\rm ps} > 150$ fs. In the detached region, signal decays dominate the sample as it is shown in Figure VI.14.

The fitted width of the prompt peak is (32.8 ± 0.7) fs which is in good agreement with what expected from resolution model studies on simulation, spanning from 32 to 35 fs depending on the resolution model.

The mean value of the prompt peak converges at (-2.1 ± 0.9) fs, again in perfect agreement with the expectations from the signal Monte Carlo studies spanning from -2.6 to -1.8 fs for different resolution models. The correlations between the lifetime and the other free parameters are all under 20%. The most important is the correlation with the

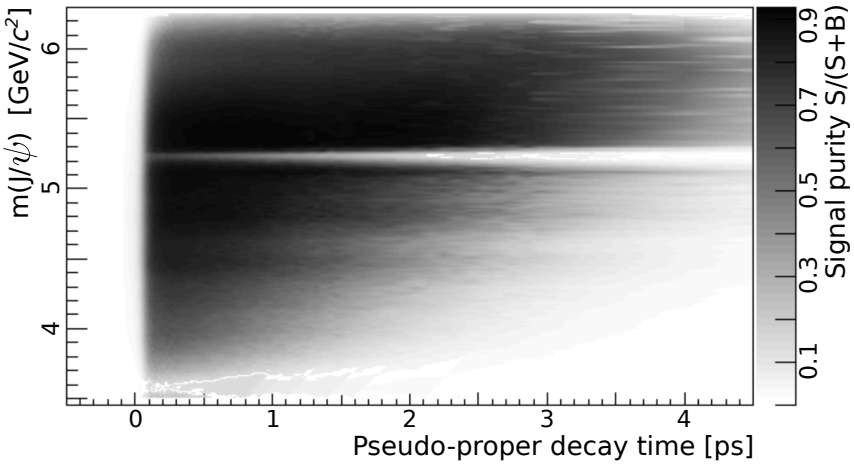

Figure VI.14: Signal Purity across the plane $m(J/\psi\,\mu) \perp t_{\rm ps}$. The horizontal white stripe is dominated by $B^+ \to J/\psi\,K^+$ decays.

signal yield (19.9%), followed by the correlation with the prompt peak parameters: mean (14%) and width (15%).

The projections of the fit, showing the contribution of the different components, superposed to the histograms of the selected dataset, are shown in Figures VI.15 and VI.16.

VI.4.2 Cross-checks on the statistical uncertainty

The statistical uncertainty is estimated using the MIGRAD algorithm with parabolic approximation of the negative log-likelihood. The estimated value is 7.7 fs. As a cross-check to the MIGRAD assessment, the MINOS technique, based on asymmetric likelihood profiling, was used. The result,

$$-7.4 < t_{\rm ps} < 7.5 \text{ fs} \qquad \text{(at 68\% of C.L.)}, \tag{VI.21}$$

is consistent with the MIGRAD estimation.

A further check is performed by fitting 300 pseudo-experiment samples, generated according to the distribution of the nominal fit model. The fitted lifetime values exhibit a distribution consistent with a Gaussian *pdf* with a standard deviation of 7.5 ± 0.3 fs, perfectly consistent with the quoted error.

Finally the bootstrap technique has been used, by defining 260 datasets of the same size of the nominal sample, with candidates selected randomly allowing repetitions [154]. The fits performed on the bootstrap datasets have results for the B_c^+ lifetime normally distributed, with a mean of 509.9 ± 0.5 fs, and a standard deviation of 7.7 ± 0.3 fs, consistent with the other assessments.

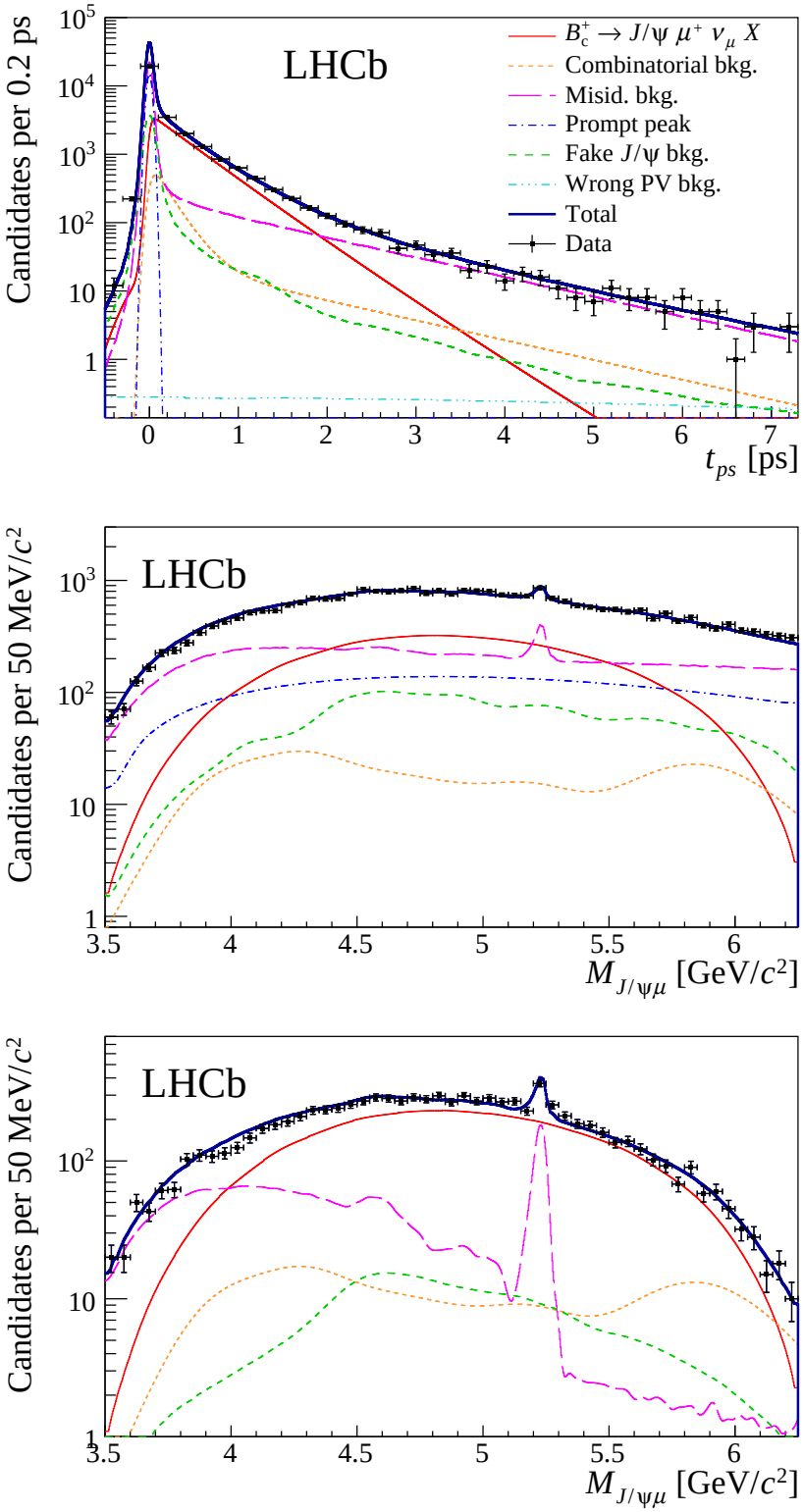

Figure VI.15: Projections of the two-dimensional fit of the $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ data model of 2012 data. The $t_{\rm ps}$ and $m(J/\psi\,\mu)$ projections are shown on the top and middle plot, respectively. The bottom figure displays the mass projection of the detached events, having $t_{\rm ps} > 150$ fs.

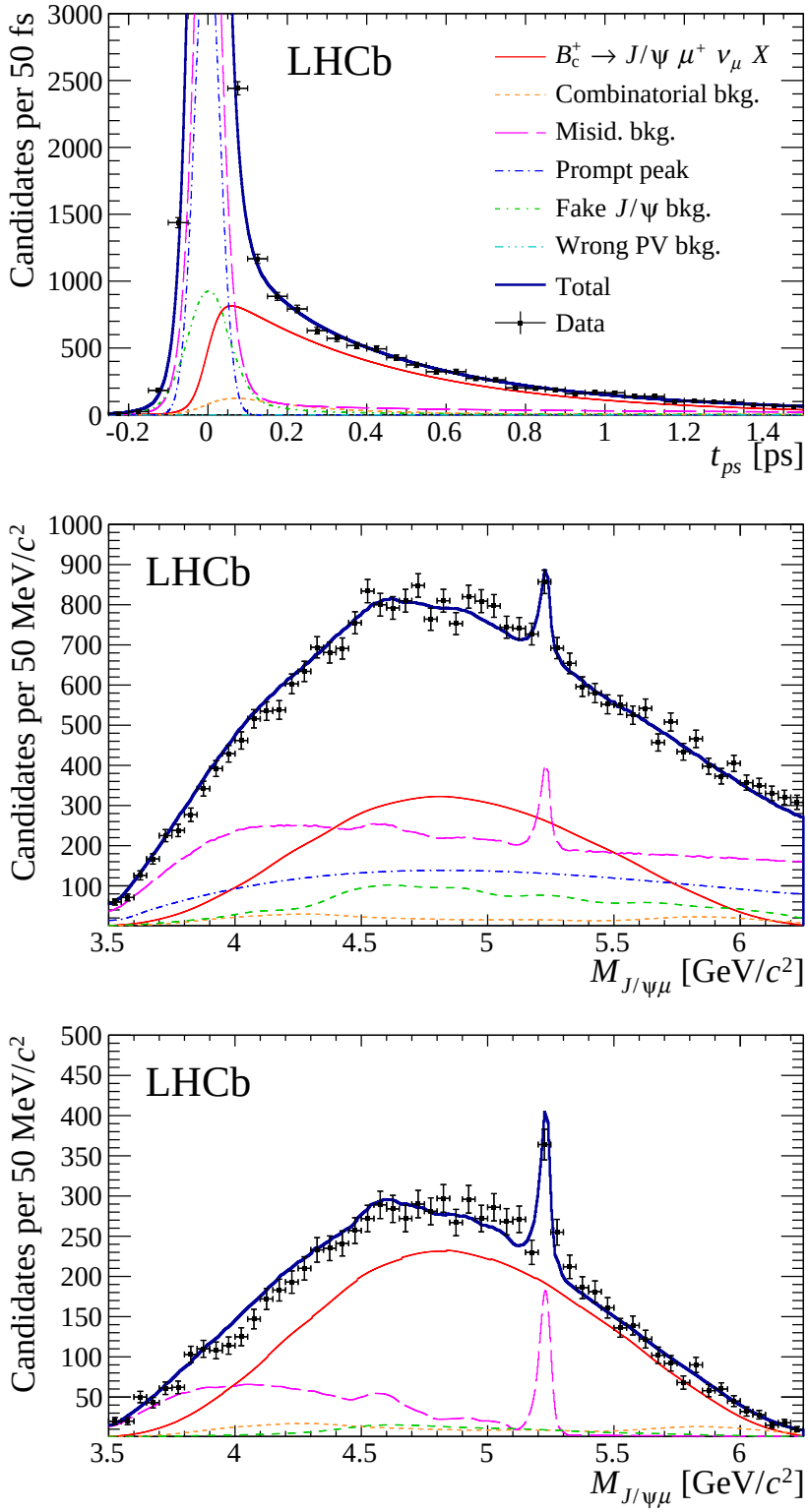

Figure VI.16: Projections of the two-dimensional fit of the $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ data model of 2012 data in linear scale. The $t_{\rm ps}$ and $m(J/\psi\,\mu)$ projections are shown on the top and middle plot, respectively. The bottom figure displays the mass projection of the detached events, having $t_{\rm ps} > 150$ fs.

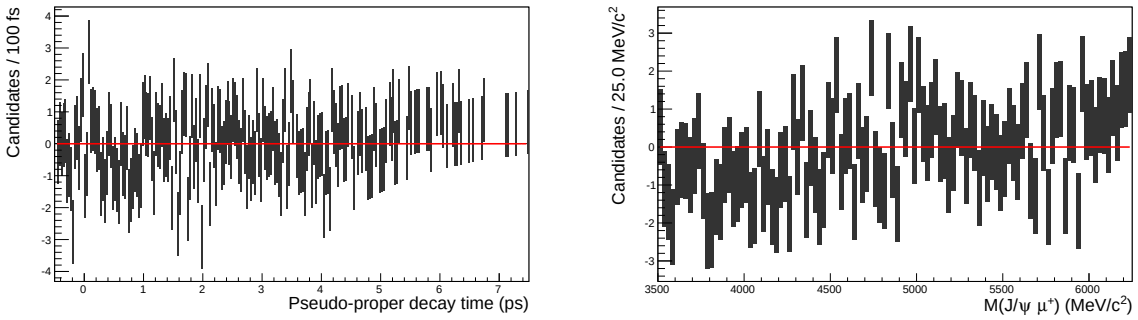

Figure VI.17: Pull distributions of the projected data model fitted to 2012 data. At left the $t_{\rm ps}$ pull distribution, at right the $m(J/\psi\,\mu)$ pull distribution.

VI.4.3 Assessment of the goodness of fit

The fit quality can be qualitatively assessed by studying the pull distributions of the data-model projections as shown in Figure VI.17.

The goodness of the fit is assessed bu using a binned method in order to be totally disentangled from the minimization of the likelihood.

A 100 ($t_{\rm ps}$) $\times$ 100 ($m(J/\psi\,\mu)$) binning grid is defined to perform this test. To avoid contribution from almost-empty bins, we evaluate the χ^2 only for those bins for which the expected event yield is larger than 0.5. The p-value is obtained assuming as number of degrees of freedom the difference between the number of bins with expected yield larger than 0.5, and the number of free parameters in the fit.

The obtained p-value has a distribution consistent with the flat hypothesis, as verified using 100 toy Monte Carlo samples, generated according to the distribution fitted on data. The χ^2 per degree of freedom of the distribution obtained against the flat hypothesis in ten p-value bins is $\chi^2/ndf = 9.1/9$.

The p-value obtained for the nominal fit is 0.20.

As a further cross-check the goodness of fit is assessed with the so-called method of the nearest-neighbour. The idea behind this method is that candidates for which the nearest candidate is closer, should occur in regions where the *pdf* is larger. The method defines, for the i-th candidate, the random variable

$$U_i = \exp\left(-n_d f_0(\mathbf{x}_i) V_D(R_i^{nn})\right), \tag{VI.22}$$

where n_d is the total number of candidates, $f_0(\mathbf{x}_i)$ is the value of the *pdf* for the set of parameters $\mathbf{x}_i$, $V_D(R_i^{nn})$ is the volume of the hyper-sphere of radius R_i^{nn}, and R_i^{nn} is the distance to the nearest neighbour.

If the fit model is correct, the U_i distribution is expected to be approximately uniform. The obtained U_i distribution was tested against the flat hypothesis obtaining a p-value of 0.4. The distribution of the U_i statistic is shown in Figure VI.18.

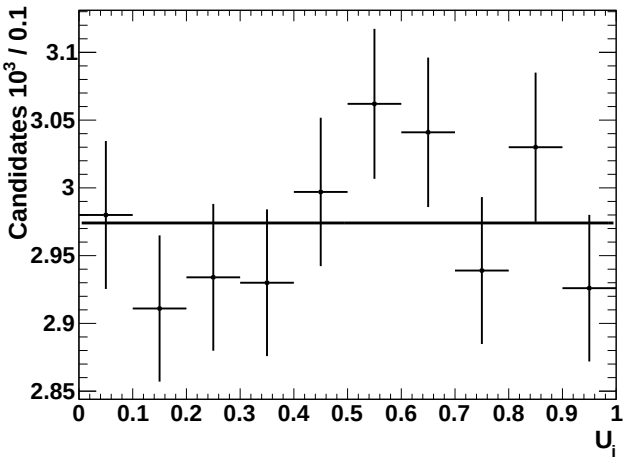

Figure VI.18: Distribution of the U_i random variable obtained with the distance-to-nearest-neighbour method. The histogram contains an entry per candidate in the final fit sample. For each candidate the nearest-neighbour is found and U_i is evaluated. The distribution is consistent with being uniform, as shown by the fit superposed to the histogram, as expected if the model pdf correctly describes the data.

VII Systematic uncertainties

The assessment of the systematic uncertainty on the lifetime measurement is structured in two main categories: uncertainties associated to the data model, assessed by changing the components within their uncertainties, both in shape and normalization; uncertainties associated to the reconstruction and fit techniques, assessed either by using alternative algorithms or quoting results from previous more specific studies.

The summary of the estimated systematic effects is reported in Table VII.2.

VII.1 Systematic uncertainties on the signal model

VII.1.1 Data-driven constraints on the theoretical model

The theoretical models described in Chapter I enter the data model in two ways:

- describing the production spectra of the B_c^+ meson, which define the shape of the k-factor and $m(J/\psi\,\mu)$ distributions *a posteriori* of the kinematic selection;
- describing the decay model of the B_c^+ meson to final states including a J/ψ and a muon, including form-factor models and feed-down decays.

A wrong production spectrum for the B_c^+ meson can, through acceptance effects, distort the $m(J/\psi\,\mu)$ distribution and, to a lower extent, the $t_{\rm ps}$ distribution. The spectra obtained by the BCVEGPY generator are tested using fully reconstructed $B_c^+ \to J/\psi\,\pi^+$ in the 2012 data sample. The model is modified by applying a linear distortion to the simulated spectra. Events are re-weighted in turn according to the p_T and η values with the following weights,

$$w_\alpha = 1 + \alpha(\eta - \overline{\eta}) \tag{VII.1}$$
$$w_\beta = 1 + \beta(p_T - \overline{p_T}) \tag{VII.2}$$

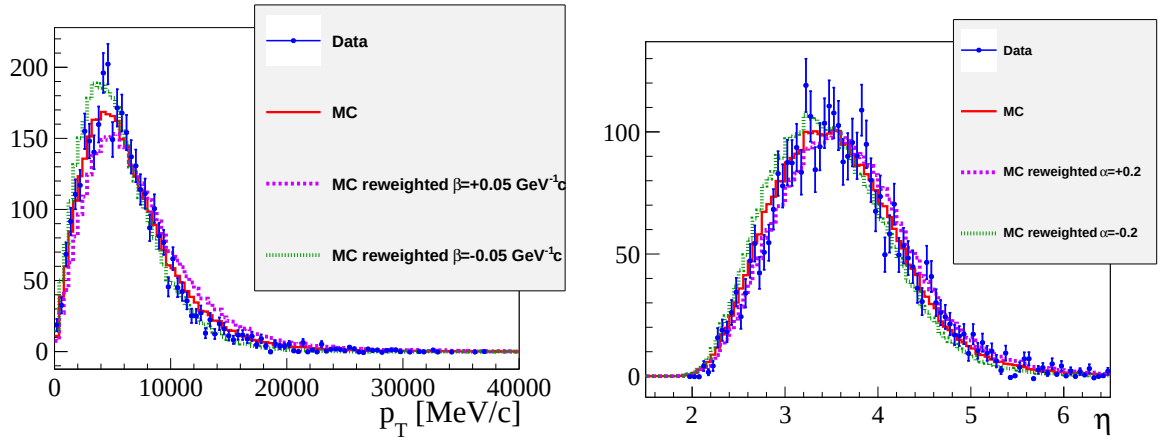

Figure VII.1: Comparison between the distribution of the transverse momentum (left) and pseudorapidity (right) of the selected $B_c^+ \to J/\psi\pi^+$ decays in a simulated sample (solid red histogram), and in the full 2012 dataset. Background subtraction is used for real data. The histograms obtained for the simulated sample reweighted described in the text are shown as dashed histograms.

Linear deformation	*Lifetime deviation* [fs]
Increasing p_T	+0.6
Decreasing p_T	–0.2
Increasing η	–0.8
Decreasing η	+0.6

Table VII.1: Deviation from the nominal lifetime obtained by fitting the data sample applying a linear deformation to the B_c^+ production spectra.

where $\overline{\eta}$ and $\overline{p_T}$ are the average values of the pseudorapidity and transverse momentum, respectively; and the deformation parameters are set to

$$\alpha = \pm 0.2 \qquad \text{and} \qquad \beta = \pm 0.05\,\text{GeV}^{-1}c, \tag{VII.3}$$

corresponding to a worsening of about 1σ of the agreement of the simulated with observed spectra for the decay $B_c^+ \to J/\psi\pi^+$, as illustrated in Figure VII.2. The fit is repeated with the re-weighted models. The effect on the lifetime is within the interval ± 1.0 fs centered in the nominal value, assigned as systematic uncertainty. The lifetime obtained are listed in Table VII.1.

The decay model depends on many physical parameters (form-factors and relative amplitudes of feed-down decays) for which only theoretical calculations and no experimental data are available. In order not to rely on the theoretical assumptions, the large size of the selected sample and its high purity are used to constrain any generic variation of the decay dynamical model and feed-down contributions using the kinematic distributions observed

in data. The approach to assess the related uncertainty is very similar to that used for the assessment on the systematic uncertainty on the production spectra, but in this case the deformation is applied to the plane $m(J/\psi\,\mu)^2 \perp m(\mu\nu)^2$, where $m(\mu\nu)^2$ is the squared invariant mass of the $\mu^+\nu$ combination, often referred to as q^2. The decay model D' is obtained as a transformation of the nominal decay model D, predicted by Kiselev [33],

$$D'(m(J/\psi\,\mu)^2, m(\mu\nu)^2) = D(m(J/\psi\,\mu)^2, m(\mu\nu)^2) \times \exp\left(\alpha_\psi m(J/\psi\,\mu) + \alpha_\nu m(\mu\nu)\right), \tag{VII.4}$$

where α_ψ and α_ν are named *deformation parameters* hereafter. The choice of an exponential deformation is driven by the need of avoiding negative weights for extreme values of the mass range, keeping an approximately linear deformation for small values of the deformation parameters.

The q^2 is reconstructed using the partial reconstruction technique described in Section IV.2.3, which exploits the alignment of the flight distance and B_c^+ velocity to guess the neutrino momentum up to a two-fold ambiguity. The q^2 solution is named q^2_H (q^2_L) when determined using the higher (lower) solution for $p(B_c^+)$. In order to improve the reconstruction quality of the flight distance, and to obtain a sample dominated by signal decays, only events with $t_{\rm ps}$ larger than 150 fs are retained for this study.

The distributions of $m(J/\psi\,\mu)$, q^2_H, and q^2_L in the data sample are shown in Figure VII.2 superposed to the stack plot of the signal an background contributions. For the former, the nominal model was used, background was obtained by applying the same techniques described in Chapter VI.

The distributions of the data sample are in good agreement with the data, though the shape of the mass distribution is clearly not perfect. The agreement is assessed by evaluating the χ^2 of the three binned distributions and combining them with the Fisher's method [1], which provides the following recipe to combine n uncorrelated p-values into a χ^2 statistic with $2n$ degrees of freedoms,

$$\chi^2_{2n} = -2\sum_i \log(p_i). \tag{VII.5}$$

The p-value associated to χ^2_{2n}, is shown in Figure VII.3 as a function of the deformation parameters. The lifetime offset obtained from fits with deformed models are also shown. When using the combination of deformation parameters maximizing the agreement with data, one obtains the data model depicted in Figure VII.4. It can be noted comparing Figure VI.15c and VII.4 that the fit quality is significantly improved after applying the deformation that maximises the combined p-value.

Among the models consistent with the data at 90% of confidence level (combined p-value > 0.1), the maximum offset of the B_c^+ lifetime is 4.1 fs. Conservatively, a ± 5 fs uncertainty is assigned as a data-driven model-independent systematic effect of the signal decay model.

Two other fully simulated form-factor models have been used to cross-check effects of the decay model: the Ebert [32] and ISGW2 [34] models described in Section I.5.3. Building the signal template distribution with the Ebert form-factor model, leads to a

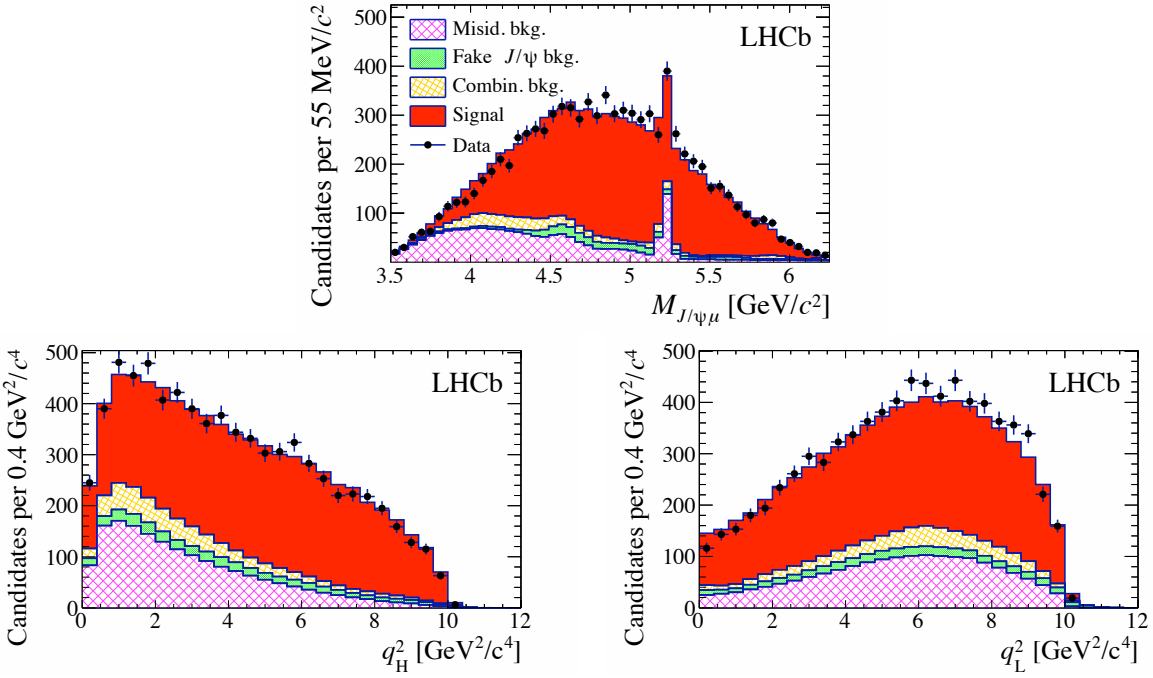

Figure VII.2: Binned distributions of the invariant mass of J/ψ and μ^+, and of the two solutions of the q^2 for detached signal ($t_{\rm ps} > 150$ fs). Signal distribution is obtained with the non-deformed Kiselev model, and background shapes and yields are obtained with the techniques described in Chapter VI. Signal yield is forced to be complementary with the background yield, which is fixed, to the total number of data candidates.

lifetime measurement shifted by +2.0 fs with respect to the nominal measurement. The shift is –1.5 fs for the ISGW2 model. The generator level distributions of the Ebert and ISGW2 models are fitted by deforming the Kiselev model as described by Equation VII.4, letting α_μ and α_ψ being free parameters of the deformation. Both cross-checks lead to a variation of the measured lifetime within the model-independent data-driven constraint.

VII.1.2 Effects of the reconstruction on the signal model

In the definition of the pseudo-proper decay time discussed in Equation VI.1, the momentum of the $J/\psi\,\mu^+$ combination is projected along the flight direction. It is therefore possible that a bias is introduced in the decay time value if the model does not describe correctly the distribution of the angle between the two directions. However, this angle is very small on average, and the effect is small. By substituting the k-factor definition with the kinematic k-factor defined in Equation VI.12, the lifetime is only 0.1 fs larger than the nominal result. The effect is therefore negligible.

The number of $m(J/\psi\,\mu)$ bins used for the discrete description of the dependence of the k-factor on the mass of the $J/\psi\,\mu^+$ combination has been arbitrarily chosen to be 15.

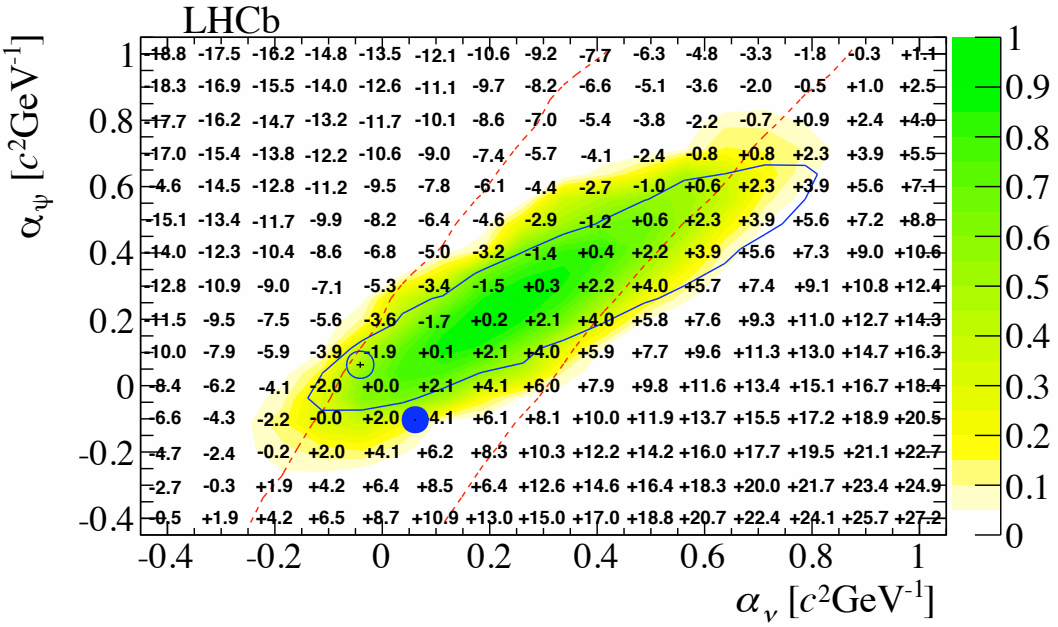

Figure VII.3: Lifetime offset (expressed in fs) obtained with different deformation parameters of the signal model. The z colored axis represents the combined p value of the goodness-of-fit test for the three variables. The red dashed lines show the region having p-value larger than 0.32 for the q^2 test only, the blue solid curve is the same for the $m(J/\psi\mu)$ test only. The filled (empty) blue marker indicates the deformation parameters that best fit the Ebert [32] (ISGW2 [34]) model.

Figure VII.5 shows the dependence of the result obtained as a function of the number of bins used. Results obtained with more than 10 bins are stable within ± 0.1 fs. No systematic uncertainty is thus associated to the k-factor binning.

The time resolution model of the signal is varied from the triple-Gaussian model used to quote the central value to two-Gaussian and four-Gaussian model; the offset on the resulting lifetime are $+1.2$ and -1.3 fs, respectively. A ± 1.3 fs uncertainty is assigned to resolution effects.

VII.2 Uncertainties on the background model

VII.2.1 Uncertainties on the combinatorial prompt background

Though signal and prompt background have very different time and mass distributions, it is the most abundant background source and its yield can not be predicted precisely. Imperfections in its model can lead to a bias of the signal parameters (both yield and lifetime).

Double- and triple-Gaussian shapes, instead of the single-Gaussian used in the nominal

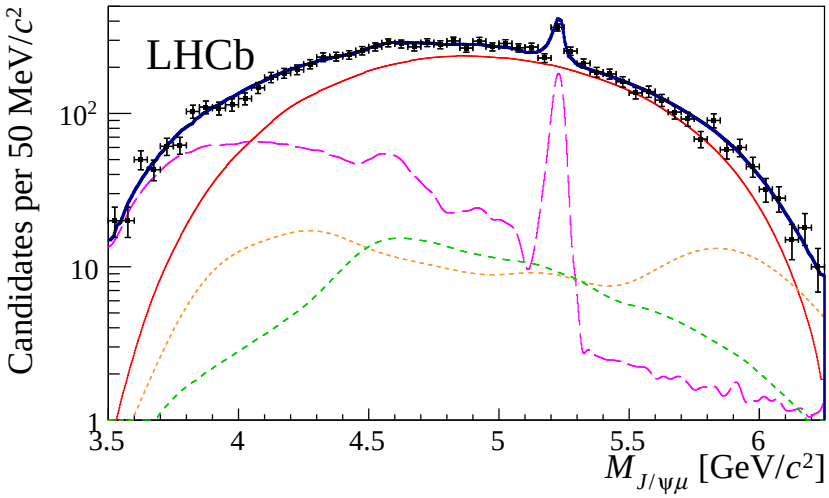

Figure VII.4: Data model for the $m(J/\psi\,\mu)$ projection compared to data when using the signal model with optimal α_ψ and α_ν as obtained in the scan of Figure VII.3. The fit components follow the legend of Figure VI.15.

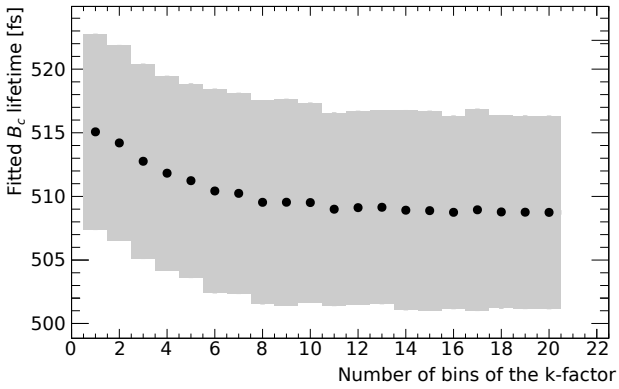

Figure VII.5: Dependence of the fitted B_c^+ lifetime values obtained with different uniform binning schemes for the k-factor distribution. The shaded region represents the statistical uncertainty associated to each fit result and is reported for completeness only.

fit, were used to assess the systematic uncertainty on the time *pdf*. The maximal variation, obtained with the triple-Gaussian model, is +1.1 fs, which is assigned as systematic uncertainty.

The $m(J/\psi\,\mu)$ distribution in the nominal model was obtained from events in the prompt region with tighter muon identification criteria, to remove the misidentification background component, and assuming null correlation with the time distribution. In order to check the robustness of the lifetime result against imperfections of the model, the lifetime is measured assuming alternative distributions for the prompt contribution.

An alternative distribution for the mass of the combination $J/\psi\,\mu^+$ is obtained from simulation of prompt $J/\psi \to \mu^+\mu^-$ decays. The simulated J/ψ is then combined with a random muon candidate from the same event. The lifetime obtained using this alternative model differs by -6.1 fs. The small number of simulated candidates surviving the selection makes the shape obtained for the *pdf* not very precise, and the fit quality is significantly reduced when using the simulated shape.

Instead of trying varying the model, the lifetime fit is performed only on the events with $t_{\rm ps}$ exceeding a threshold $t_{\rm ps}^{\rm th}$. The fit result is shown in Figure VII.6 as a function of the threshold $t_{\rm ps}$. Superposed, the expected statistical dilution due to the loss of signal candidates is shown. The lifetime values resulting from the cropped data sample are consistent with the expected fluctuation because of the reduced size of the sample. To estimate the variance of such fluctuation, the statistically independent results $\tau_<$ and $\tau_>$ are considered. The former is obtained from events below the threshold $t_{\rm ps}^{\rm th}$, the latter from events beyond the threshold. The two results are affected by a statistical error $\sigma_<(T)$ and $\sigma_>(T)$. The nominal result τ_0, obtained from all events and having error σ_0 is expected to be the weighted average of the two results,

$$\tau_0 = \frac{\tau_</\sigma_<^2 + \tau_>/\sigma_>^2}{1/\sigma_<^2 + 1/\sigma_>^2} \qquad \text{and} \qquad \sigma_0^2 = \frac{1}{1/\sigma_<^2 + 1/\sigma_>^2}. \tag{VII.6}$$

The uncertainty shown in the Figure VII.6 is the statistical error on $\delta\tau = t_> - t_0$, which can be written as

$$\sigma(\delta\tau) = \frac{\sigma_>}{\sqrt{1 + \frac{\sigma_0^2}{\sigma_>^2 - \sigma_0^2}}}. \tag{VII.7}$$

Assuming the statistical error scaling with the inverse of the squared root of the signal event statistics, one would have $\sigma_>^2 = \sigma_0^2/(1 - f_<)$, where $f_<$ is the fraction of signal events below the threshold T. In such case the above formula would give $\sigma(\delta\tau) = \sigma_0\sqrt{f_</(1 - f_<)}$. However this assumption does not take into account the different background levels for different values of $t_{\rm ps}$, and therefore it overestimates the statistical loss of information in the prompt peak region. The uncertainty assessed through the likelihood second derivative when fitting the candidates beyond threshold is used instead.

The lifetime variation obtained with the $t_{\rm ps} > 150$ fs requirement, which removes most of the prompt candidates, is $+6.4$ fs, corresponding to 1.5 times the expected statistical error. Conservatively, a ± 6.4 fs uncertainty is assigned to systematic effects associated to the prompt peak model. This is consistent with the other assessments discussed above.

The largest deviation from the nominal value observed varying the threshold $t_{\rm ps}^{\rm th}$ is found at $t_{\rm ps}^{\rm th} = 0.25$ ps, corresponding to 13 fs, or 2.2σ. This is still consistent with a statistical fluctuation, and becomes perfectly reasonable when including the conservative uncertainty on possible effects described above.

VII.2.2 Uncertainties on the combinatorial detached background

Since the combinatorial detached contribution is the only background source whose model relies on simulation, data-driven checks are performed to evaluate the uncertainty on its

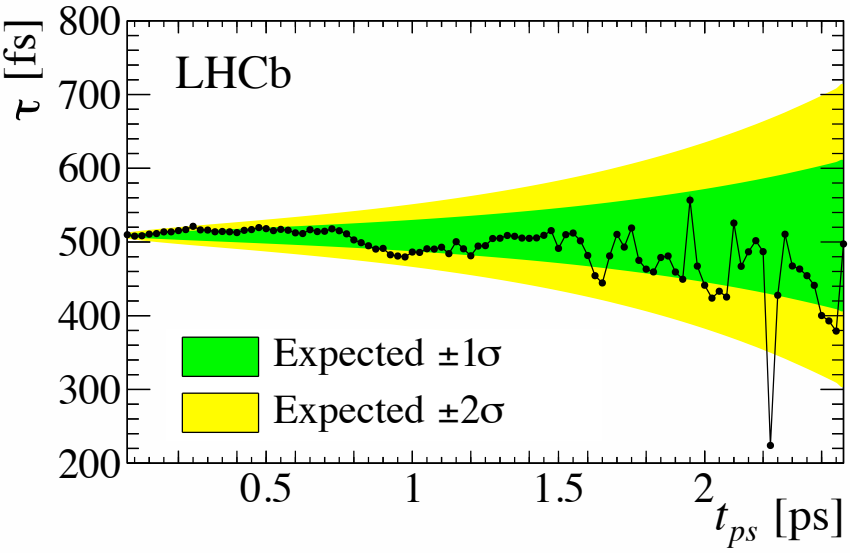

Figure VII.6: Fitted lifetime obtained removing the events with pseudo-proper decay time $t_{\rm ps}$ shorter than the threshold reported on the x-axis. The green (yellow) band shows the expected $\pm 1(2)\,\sigma$ statistical fluctuation of the $\delta\tau$ variation due to the loss of the removed signal candidates.

predicted yield. The yield of detached candidates before the bachelor muon identification requirements, which is expected to be dominated by b-hadron decays, is measured and found to differ by 35% from the value predicted by the simulation. To account for a further uncertainty related to the efficiency of the muon identification criteria, a systematic uncertainty of $\pm 50\%$ is assigned on the combinatorial background yield. The quoted uncertainty on the yield corresponds to ± 3.4 fs on τ.

The pseudo-proper time and mass distributions are also found to agree. The uncertainty on the *pdf* is dominated by the shape of the $t_{\rm ps}$ distribution. A single exponential rather than a double exponential function is used, and the parameters of the nominal function are varied within their statistical uncertainty. The maximum variation is -7.3 fs.

VII.2.3 Uncertainties on the fake-J/ψ background

For the fake J/ψ background, the expectation value of its yield is varied within its statistical uncertainty, assessed by combining the uncertainty on the expected number of candidates due to combinations including a fake J/ψ, and on the number of candidates obtained combining a fake J/ψ and a fake muon,

$$\sigma(\text{fake } J/\psi\,;\text{real } \mu^+) = \sqrt{\sigma^2(\text{fake } J/\psi\,) + \sigma^2(\text{fake } J/\psi\,;\text{fake } \mu^+)}. \qquad \text{(VII.8)}$$

The systematic uncertainty associated to the yield of fake J/ψ candidates is ± 0.4 fs.

The uncertainty on the shape of the *pdf* is studied using two alternative models obtained using only one of the sidebands. Using the left (right) sideband only, the fitted

lifetime changes of +2.3 fs (−0.8 fs). A symmetric uncertainty of ±2.3 fs is assigned to effects due to imperfections in the model of the fake J/ψ decay.

VII.2.4 Uncertainties on the misidentification background

The number of candidates built with a hadron misidentified as bachelor muon is estimated with the technique described in Section VI.2. The expected yield is 10987 ± 110, where the uncertainty is statistical, corresponding to 1687 candidates with $t_{\rm ps}$ larger than 150 fs.

In order to assess the systematic uncertainty on this model, the lifetime fit is performed letting the normalization of the misidentification background to float without constraint. The yield converges to 11418 ± 136, corresponding to a shift of the lifetime obtained in the fit procedure of +0.8 fs. The shift is larger than the one obtained by artificially increasing (decreasing) the fixed normalization of the misidentification background by 1σ, corresponding to a shift in the lifetime of −0.5 fs (+0.6 fs).

The accuracy of the *pdf* is limited by the size of the calibration and *NoMuonID*, since the misidentification probability W, defined in Equation VI.16, is parametrized in bins of several variables. The effect of the uncertainty in each bin is estimated by simulating 1000 alternative *pdf*s after applying random offsets to the W values, according to their statistical uncertainty. The maximal variation of the lifetime is −1.2 fs.

The systematic uncertainty assigned to the misidentification background is ±0.8 fs for effects of the normalization, and ±1.2 for effects of the shape of the *pdf*.

VII.3 Reconstruction and fitting techniques

VII.3.1 Validation of fit technique

A possible bias from the fit procedure is explored using simulated pseudo-experiments generated according to our nominal fit model, with a fitted lifetime of 508.7 fs.

The mean value the results obtained from 300 pseudo-experiment samples, generated according to the nominal fit model is 508.9 ± 0.5 fs, in good agreement with the generated value. A conservative 0.5 fs systematic uncertainty is accounted for possible bias of the maximum likelihood estimator.

VII.3.2 Momentum scale calibration

Momentum scale calibration can affect the lifetime measurement if the measured momentum is different from the real one in two ways: first, the pseudo-proper decay time is based on the kinematical factor in its definition, and second, a different scale in simulated and real data introducing a bias through the k-factor parameterization in bins of $m(J/\psi\,\mu)$. Both the effects are expected to be very small, errors in the momentum calibration cancel at first order when evaluating the ratio $p(J/\psi\,\mu^+)/m(J/\psi\,\mu)$, and the k-factor binning is very coarse with respect to the error introduced by possible momentum scale mis-calibration.

To assess the systematic error associated to momentum scale, we refit the data sample after the application of an offline momentum scale calibration procedure, based on the fit of the experimental bias when reconstructing masses of quarkonium resonances and beauty mesons with respect to the relative world averages.

The effect of the calibration on the fitted lifetime is negligible.

VII.3.3 Length scale calibration

The uncertainty on the length calibration, found in other analyses to be dominated by the z-scale calibration, is extrapolated from the assessment performed in the analysis on the B_s^0 effective lifetime measured studying the decay $B_s^0 \to K^+K^-$ [155]. The value assigned as systematic uncertainty on the B_s^0 lifetime measurement was ± 1.0 fs. This value is corrected for the different transverse momentum spectra of the B_s^0 and B_c^+ mesons through the factor $\frac{\langle p_{B_c^+} \rangle}{\langle p_{B_s^0} \rangle} \approx 1.3$. The systematic uncertainty on the B_c^+ lifetime measurement due to the length scale calibration is therefore ± 1.3 fs.

VII.3.4 Dependence of the efficiency on decay time

The efficiency function has been discussed in Chapter V and found consistent with being independent of the decay time. According to the standard parameterization of efficiency losses at large decay time, the efficiency is modeled as $\epsilon \propto (1 - \beta t_{\rm ps})$. For an efficiency independent of the decay time, β is expected to be consistent with zero. The effect of the β effect on the measured lifetime is a bias $\Delta\tau$ which can be assessed by considering that

$$\begin{aligned}
\exp\left(-\frac{t}{\tau+\Delta\tau}\right) &= \exp\left(-\frac{t}{\tau}\right)(1+\beta t) \\
\exp\left(\frac{t}{\tau} - \frac{t}{\tau+\Delta\tau}\right) &= 1+\beta t \\
\exp\left(\frac{\Delta\tau}{\tau^2\left(1+\frac{\Delta\tau}{\tau}\right)}\right) &= 1+\beta t \\
1 + \frac{\Delta\tau}{\tau^2\left(1+\frac{\Delta\tau}{\tau}\right)} t + \mathcal{O}\left(\frac{\delta\tau^2}{\tau^2}\right) &= 1+\beta t \\
\Delta\tau = \frac{\beta\tau^2}{1-\beta\tau} + \mathcal{O}\left(\frac{\Delta\tau^2}{\tau^2}\right) &\underset{\beta\tau\ll 1}{\longrightarrow} \beta\tau^2
\end{aligned} \tag{VII.9}$$

The β effect has been studied in detail for the measurement of b-hadron lifetimes using final states containing a J/ψ [152]. The major contribution to efficiency losses at large decay time is associated to the reconstruction algorithms of the Vertex Locator at trigger level, and is correctly reproduced in the detector simulation. The effect is more important in 2011 data than it is in 2012, because of a change in the reconstruction algorithms, but it is not totally solved. To assign a systematic uncertainty, the statistical uncertainty on the value of β obtained from simulation is considered. The value of β is expected to lie in

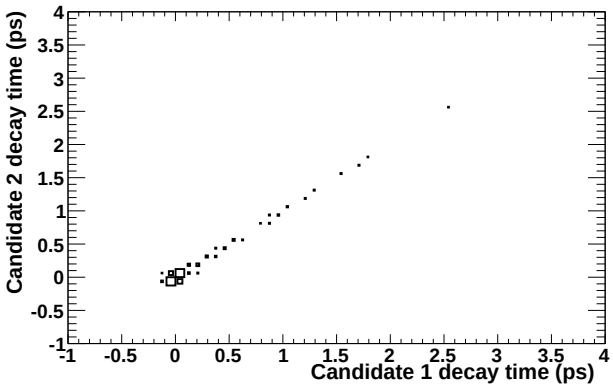

Figure VII.7: Correlation between the decay time reconstructed for pairs of same-event candidates.

the interval $[-3, 9]$ ns^{-1} (at 68% of C. L.). Conservatively, a total uncertainty of ± 10 ns^{-1} is assigned to β, translating into an uncertainty $\beta\tau^2 = 2.6$ fs on the measured lifetime.

VII.3.5 Uncertainty associated to multiple candidates

In order to assess the systematic uncertainty due to the correlation of multiple candidates in the same event, the maximal possible bias is introduced by discarding all multiple candidates in an event, but the one with the largest (or smaller) pseudo-proper decay time. The fits obtained on the samples differ from the nominal value by less than 1 fs, conservatively accounted as systematic uncertainty.

It is interesting to observe that the deviations in the fitted lifetime obtained for the two modified samples have the same sign, this happens because the effect of reducing the yield is larger than the explicit bias introduced by selecting only candidates with the largest (or shortest) decay time. This is confirmed by the scatter plot of the decay time of pairs of candidates found in the same events, shown in Figure VII.7, displaying large correlation between the decay times reconstructed for same-event candidates.

VII.3.6 Uncertainty due to incorrect primary vertex associations

To estimate the uncertainty on the modelling of events with an incorrectly associated primary vertex, the fit is repeated removing events where more than on primary vertex is consistent with the candidate decay. The lifetime changes by $+1.8$ fs. A symmetric uncertainty of ± 1.8 fs is accounted accordingly.

Source	*Assigned systematic [fs]*
B_c^+ production model	1.0
B_c^+ decay model	5.0
Signal resoution model	1.3
Fake J/ψ background yield	0.4
Fake J/ψ background shape	2.3
Prompt combinatorial background model	6.4
Detached combinatorial background yield	3.4
Detached combinatorial background shape	7.3
Misidentification background yield	0.8
Misidentification background shape	1.2
Length-scale calibration	1.3
Momentum scale calibration	0.2
Efficiency function	2.6
Incorrect association to primary vertex	1.8
Multiple candidates	1.0
Fit validation	0.5
Quadratic sum	12.4

Table VII.2: Summary of the contributions to the systematic uncertainty on the B_c^+ lifetime.

VII.4 Summary of the systematic uncertainty

The sum in quadrature of the mentioned contributions is 12.4 fs, and is dominated by effects on the decay model. In particular, the combinatorial background, both prompt and detached, introduces the largest part of the systematic uncertainty.

The theoretical uncertainty on the signal decay model is under control, and conservatively estimated with an original data-driven technique.

The single contributions are listed in Table VII.2.

VII.5 Further cross-checks

Besides the checks described above, aiming at the determination of the systematic uncertainty associated to some specific step of the analysis, several further consistency checks have been performed to probe residual biases not accounted for by the assigned systematic uncertainty.

Reducing misidentification background removing long-lived events

In order to check for possible mismodelling of the misidentification background, the fit is repeated by varying the maximum $t_{\rm ps}$ requirement between 2 and 8 ps. The long-lived tail

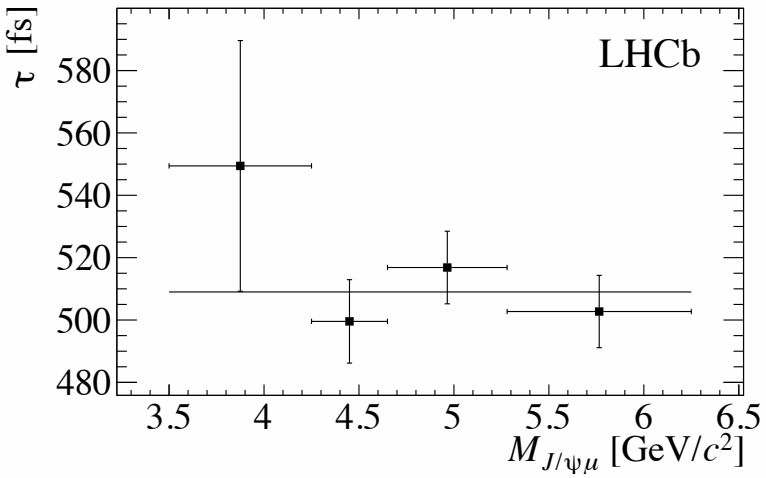

Figure VII.8: Results of the fits performed in mass bins. The horizontal line represents their average, consistent with the result of the nominal fit.

of the $t_{\rm ps}$ distribution is dominated by b-hadron decays with a hadron misidentified as a muon.

Removing events with a reconstructed pseudo-proper decay time down to 2.0 fs, variations of the fitted lifetime within ± 1.5 fs are observed. This is consistent with the systematic uncertainties assessed above.

Fit in $m(J/\psi\,\mu)$ bins

The fit is also performed in four $m(J/\psi\,\mu)$ bins, defined by the boundaries: 3500, 4250, 4650, and 6250 MeV/c^2.

Unfortunately, only 20% of the candidates falls in this region, thus the statistical uncertainty on the result is expected to be of the order of $8/\sqrt{20\%} = 19$ fs. Since the nominal value has been obtained with a sample which includes these candidates, the difference between the two fit results has a statistical uncertainty of 16 fs.

The fit results shown in Figure VII.8 are in excellent agreement. The weighted average of the four fitted values is 508.2 fs, perfectly consistent with the result of the nominal fit, 508.7 fs, within the systematic uncertainties quoted for the model.

Extended likelihood fit

The likelihood fit used to quote the central value does not rely on the likelihood extension to allow fluctuations on the total number of events. Extended likelihood fits are usually recommended when fitting event yields for cross-section measurements, while it is not necessary when fitting shape parameters, as masses or lifetimes.

The extension of the likelihood is obtained as

$$-2\log\mathcal{L}_{\rm extended} = -2\log\mathcal{L} - 2\,(-M + N\log(M) - N\log(N) - N), \qquad \text{(VII.10)}$$

where N is the number of entries in the fit sample and M is the sum of candidates associated to each component (signal, misidentification background, ...).

The comparison between the result obtained with and without the likelihood extension is a useful cross-check of the stability of the fit technique. The lifetime result changes by less than 0.1 fs.

Separate fits on upwards and downwards magnet polarity data samples

As a further cross-check of the possible reconstruction biases, the global fit is performed on the samples acquired with upwards and downwards polarity separately. The *pdf*s of signal and combinatorial background, obtained from simulation are not regenerated and the same *pdf*s as for the global fit on the whole sample are used, the other *pdf*s, obtained with data-driven methods, are rebuilt using the split data sample, with the same techniques as described in Chapter VI.

The lifetime obtained from the fit on the upwards polarity sample is (515.0 ± 11.0) fs, while for the downward polarity (500.5 ± 11.0) fs is obtained. The two results agree within 1σ: the difference is consistent with being statistical only. The combination of the two results is (507.5 ± 7.7) fs, 0.5 fs far from the nominal value. Such a small discrepancy is considered as already taken into account in the systematic uncertainty on the shape of the *pdfs* and on the fit technique, so that no further uncertainty is accounted for this effect.

Vertex $J/\psi\,\mu^+$ and rejection of the combinatorial background

In order to cross-check the value expected for the detached combinatorial background, due to the random association of a real muon with a real J/ψ produced in a b decay, the fit is repeated without constraint on the normalization of this contribution, for different values of the threshold on the vertex χ^2 criterion.

Since the distribution of combinatorial candidates with respect to the $J/\psi\,\mu^+$ vertex χ^2 is almost flat, it is reasonable to predict that the fitted yield $N_{\rm comb}^{\rm (obs)}$ scales linearly with the threshold.

The fitted combinatorial yield is reported in Figure VII.9. The straight line forced to intercept the y-axis at $y = 0$, fitting the experimental points is also superposed.

The agreement is excellent, and the effect on the fitted lifetime lays in the range ± 2.5 fs centered on the nominal value.

Composition of the misidentification background

The composition of the misidentification background was studied using the simulated sample of $B_{uds} \to J/\psi\,X$ decays. Candidates passing the selection are associated to the true decay simulated and classified according to the final state.

Figure VII.10.a shows a stack plot of the simulated contributions to the misidentification background, for two different muon identification criteria applied to the bachelor muon. As it can be seen, the exclusive modes with the largest relative contribution to the background are $B^0 \to J/\psi\,K^-\pi^+$ and $B^+ \to J/\psi\,K^+$ The contribution of $B^0 \to J/\psi\,K^-\pi^+$ decays in

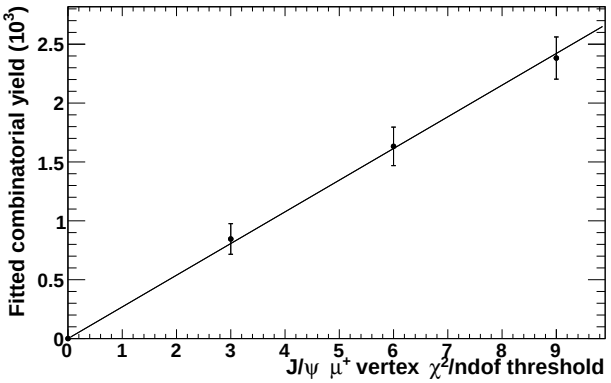

Figure VII.9: Fitted yields for the detached combinatorial background obtained with different thresholds on the maximum $J/\psi\,\mu^+$ vertex χ^2 per degree of freedom. The prediction of linear scaling for combinatorial background is superposed as a solid line.

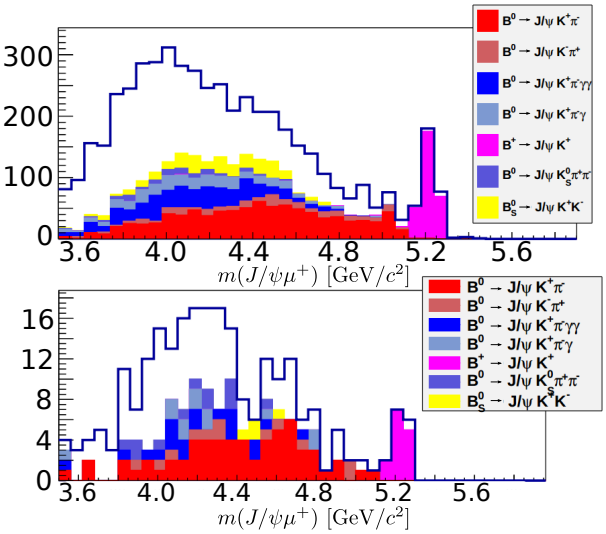

Figure VII.10: Classification of the misidentification background based on a simulated sample of $B_{uds} \to J/\psi X$. Reconstructed and selected candidates are associated *a posteriori* to the simulated decay. The selection includes loose muon identification criteria in the top figure, and the tight requirement used in the analysis in the bottom plot.

real data was studied by combining the $J/\psi\,\mu^+$ candidate to the long track giving the best vertex χ^2. If the difference between the four-track vertex χ^2 and the three-track vertex χ^2 is larger than 10, the combination is discarded. The invariant mass of the combination $J/\psi\,\mu^+$ + track is calculated assuming the pion mass for the additional track. Figure VII.11 shows the distribution of the invariant mass of the four-track combination in the data sample obtained without muon identification requirements (*NoMuonID*). Two peaks corresponding to the decays $B^0 \to J/\psi\,K^-\pi^+$ and $B_s^0 \to J/\psi\,K^+K^-$ are well visible, but

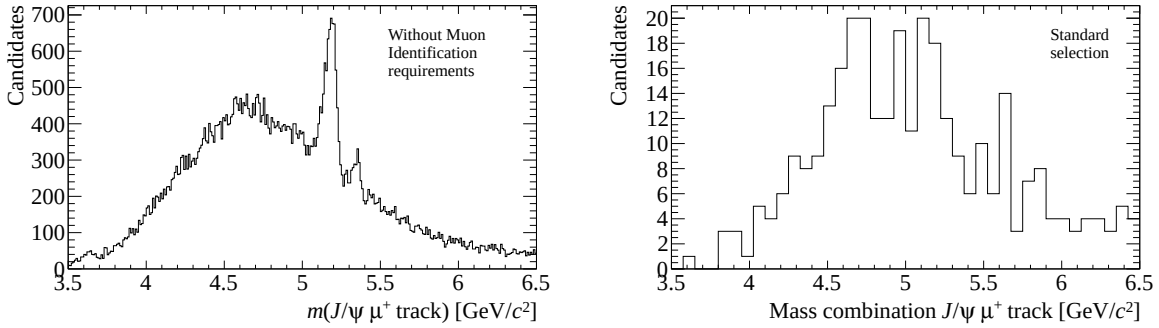

Figure VII.11: Distribution of the invariant mass of the combination $J/\psi\,\mu^+\mathsf{track}$, in the data sample without muon identification requirements on the bachelor muon, at left, and to the standard-selection sample at right. The invariant mass is evaluated assuming the pion mass for the additional track. The two peaks visible in the left plot are due to $B^0 \to J/\psi\,K^+\pi^-$ and $B_s^0 \to J/\psi\,K^+K^-$ decays, the mass value are shifted because of the wrong mass assumptions for the kaons.

the masses are shifted towards lower values because of the wrong mass assumption for the kaons.

Requirements on the four-track invariant mass have been considered as possible improvements of the misidentification background rejection. Unfortunately, once the muon identification criteria are applied, only 316 candidates survive the requirement on the four-track vertex χ^2, of which only 86 fall in the region between 5.0 and 5.5 GeV/c^2. Their distribution is shown in Figure VII.11.b. Removing the reconstructed candidates to the ten thousand candidates composing the misidentification background would be worthless even if the all the 86 candidates were due to misidentification. Considering the distortion to the signal model that such a kinematical requirement would introduce, and the relative additional complications, it has been chosen not to include these isolation requirements in the selection strategy.

Test of the k-factor with $B^0 \to J/\psi\,K^{*0}$ with $K^* \to K^+\pi^-$

In order to check the reliability of the simulation used to predict the k-factor distribution, including the resolution effects, the decay $B^0 \to J/\psi\,K^{*0}$ with $K^* \to K^+\pi^-$ is reconstructed in real data with and without the pion. Using the information from the fully reconstructed decay, one can calculate the k-factor from real data (defined in this case as the ratio of the two reconstructed decay time $t_{\rm ps}/t$) and compare it to the same quantity as predicted by simulation. The selected candidates are weighted by a quadratic function of $m(J/\psi\,K^-)$ to force the agreement of simulated and real data mass distributions. The k-factor distribution of the reweighted simulated candidates agrees with the real data distribution without further reweighting. The mass and k-factor distributions of the real and simulated samples are shown in Figure VII.12. This is an important result since it shows that the systematic uncertainty on the k-factor distribution is well controlled by acting on the distribution of

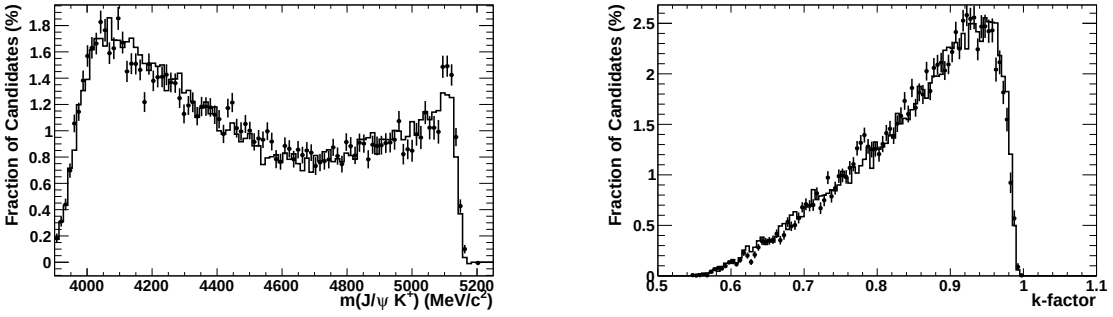

Figure VII.12: Comparison of the distribution and invariant mass of the $J/\psi\,\mu^+$ combination (left) and of k-factor (right) from $B^0 \to J/\psi\,K^{*+}$ decays in real data (markers) and simulation (solid histogram).

the reconstructed invariant mass, either by applying a weight function as done here, or analysing the events in mass bins as done in the lifetime measurement.

After reweighting, the average k-factor is predicted to better than 0.1%, corresponding to a bias on the lifetime below 0.5 fs.

Further details on the $B^0 \to J/\psi\,K^-\pi^+$ control channel can be found in Ref. [156].

Feed-down contributions

The contribution of feed-down decays with different topology modifies the $m(J/\psi\,\mu)$ and k-factor distributions of the $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays with respect to the pure $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay mode. The effect on the mass and k-factor distributions is discussed in Chapter VI for the nominal value, while the assessment of the systematic uncertainty is described in the following.

The relative decay widths with respect to the $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay mode are varied according to the range of values predicted in Refs. [32, 33, 57, 145, 148, 157–160]. More conservatively, each modelled component is varied in turn by $\pm 100\%$ in order to take into account possible smaller contributions, such as the non-resonant $B_c^+ \to J/\psi\,\mu^+\nu\pi^0$ decays, which have not been modelled explicitly and whose *pdf* shapes are intermediate between the considered ones. The maximum variation with respect to the nominal fit is 0.3 fs. Table VII.3 summarizes the predicted intervals, the values used to quote the nominal lifetime, the variation range used for the systematic check, and the maximal lifetime offset. The latter is found to be in any case much smaller than the accounted systematic uncertainty on the theoretical model.

Final state	$\psi(2S)\mu^+\nu$	$J/\psi\,\tau^+\nu$	$\chi_c(h_c)\mu^+\nu$
Theoretical prediction	0.1 – 7.8%	(25.0 ± 0.5)%	10 – 50%
Nominal value	5%	25%	33%
Range of systematic check	0 – 10%	0 – 50 %	0 – 66%
Maximal lifetime variation [fs]	0.1	0.2	0.1

Table VII.3: Explored ranges of relative feed-down contribution for the main modes. The theoretical prediction and the maximal lifetime offset obtained varying each branching fraction within its variation range are also listed. Theoretical predictions are published in Refs. [32, 33, 57, 145, 148, 157–160].

VIII Conclusion and outlook

The LHCb experiment is the smallest of the four large experiments of the *Large Hadron Collider*, with a physics programme originally focused on b-physics, CP violating processes and rare decays, intended for indirect searches for physics beyond the standard model as unexpected contribution arising in loop diagrams. The excellent performance of the LHC and of the LHCb experiment allowed to significantly extend the physics programme, so that during the three years of my Doctorate I saw the one-sentence presentation of the LHCb experiment at conferences evolving from the "the b-physics experiment" to "the general purpose experiment in the forward region". Key features of the LHCb detector are the high precision on the impact parameter of tracks and vertex position; low material budget before the calorimeters and an excellent tracking system to precisely measure track momenta; a particle identification apparatus composed of two RICH detectors with three radiators, two calorimeters, and a 4+1 station muon system. All the detectors are read with an average frequency of 1 MHz when the Level-0 hardware trigger, operating at 40 MHz with fixed latency, detects high-p_T muons, large impact parameter tracks, or important energy depositions in the calorimeters. Three subsequent layers of reconstruction and data-reduction follow: a first High-Level Trigger stage based on partial event reconstruction, a second High-Level Trigger stage based on full but fast reconstruction, and a Stripping stage relying on full offline reconstruction. Data selected by the High-Level Trigger 2 are archived on tape, data selected by the Stripping are available on disk for analysis. LHCb has published many analyses of the dataset collected in the first two years of data-taking whose results lead the world averages, in both the b- and c-quark sectors. World leading results on many parameters of the CKM matrix, the exclusion of a wide range of parameters of supersymmetric models, and first observation of several extremely rare Flavour Changing Neutral Current decays are examples of results achieved in the mainstream programme of LHCb, which is extended and complemented by the new results on proton-ion collisions, exotic hadron spectroscopy, and quarkonium production.

For the first time, the B_c^+ meson has become subject of an extensive research programme. The LHCb Collaboration has published precision measurements of the mass using several channels with independent uncertainties, and observed many decay modes; among the

others, the first observations of decays including in the final state a $\psi(2S)$, or a $p\bar{p}$ pair, or a B_s^0 were achieved. The latter also represents the first observation of a B_c^+ decay led by a $c \to s$ transition. In the meanwhile the ATLAS Collaboration has reported the first observation of an excited $(\bar{b}c)$ state, interpreted as a $B_c^{**+}(2S)$ resonance decaying to a B_c^+ meson and a $\pi^+\pi^-$ pair.

LHCb has performed high-precision measurements of B hadron lifetimes, with excellent results in particular in the relative measurements assessing the lifetime difference between two states. The lifetime of most b-hadrons has been measured for the first time, or with a precision much higher than at previous experiments. Among the results obtained for b-hadrons, the lifetime of the Λ_b^0 baryon has been measured by the LHCb Collaboration and found consistent with theoretical predictions based on the *Heavy Quark Expansion*, disproving previous experimental results. Heavy Quark Expansion is today considered the most powerful tool to predict lifetimes. The key feature of HQE, and in general of the Operator Product Expansions, is the separation of the perturbative and non-perturbative contributions to the decay width predictions. The non-perturbative calculations require input from effective models for the strong nuclear interaction, and the choice of the effective model differentiates the several flavours of HQE predictions. Recent predictions based on Lattice-QCD calculations have been found in excellent agreement with the experimental results.

In this Thesis, I discussed the first lifetime measurement of the B_c^+ meson performed by the LHCb Collaboration, of which I am the main contributor and corresponding author for the relative publication [141].

The decay mode chosen to perform the study is the inclusive semileptonic decay $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$. The large branching fraction of the B_c^+ decay towards final states containing a J/ψ and a muon allows to avoid selection criteria based on the impact parameter of the final state tracks, tightening the selection on the muon identification to exploit the rarity of the three-muon signature to reject background. Dedicated muon algorithms have been developed and discussed, with particular care devoted to possible bias introduced in the misidentification probability of the bachelor track as a muon because of the two additional muons from the J/ψ decay. The rejection is complemented by a sophisticated data-driven modelling technique used to statistically subtract the misidentification background when measuring the B_c^+ lifetime in a two-dimensional fit to the joint distribution of the invariant mass $m(J/\psi\,\mu)$ of the $J/\psi\,\mu^+$ combination, and the decay time reconstructed in the $J/\psi\,\mu^+$ rest frame. The effort is finally rewarded by a systematic uncertainty associated to the misidentification background reduced to a small level.

Other background sources include candidates with a fake J/ψ meson associated to a random real muon, and candidates obtained combining a real muon and a real J/ψ but not originated in a B_c^+ decay. The latter, named combinatorial background, is split in two components: prompt and detached. All the data-model of the background sources are obtained through data-driven techniques with the exception of the combinatorial detached background for which simulation was used. Data-driven cross-checks confirm that the model is reasonable, but a large systematic uncertainty is associated for a possible

mismatch between the simulated and real dataset. This reflects into a large uncertainty on the lifetime due to the combinatorial background, both prompt and detached, which dominates the systematic uncertainty on the final result.

The semileptonic nature of the chosen decay requires techniques to estimate the relation between them momenta of the $J/\psi\,\mu^+$ combination and of the B_c^+ meson. A statistical approach based on theoretical decay models, known as k-factor model, has been chosen for the lifetime measurement. Our original approach to the k-factor technique allowed to disentangle analytically the uncertainties introduced in the signal model by the missing kinematic information and the reconstruction effects. The resulting signal model, based on theoretical assumptions was cross-checked using the actual kinematic distributions obtained with a partial-reconstruction technique exploiting the alignment of the momentum of the B_c^+ and its flight direction (vector connecting the primary and decay vertices). This original data-driven cross-check of the theoretical decay model has allowed to assess the systematic uncertainty on the B_c^+ lifetime due to possible mismatch between the theoretical and the real decay model. The uncertainty includes the effect of the form-factor description of the dominant $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decay, and contribution from the feed-down $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ decays, such as $B_c^+ \to J/\psi\,\tau^+\nu$, $B_c^+ \to \psi(2S)\mu^+\nu$, and $B_c^+ \to \chi_{cJ}\mu^+\nu$ decay modes.

The result obtained from the blind analysis of the dataset collected by the LHCb experiment in 2012 in pp collisions, and corresponding to an integrated luminosity of 2 fb^{-1}, is

$$\tau_{B_c^+}^{(\mathrm{LHCb},J/\psi\,\mu^+)} = 509 \pm 8\,(\mathrm{stat}) \pm 12\,(\mathrm{syst})\ \mathrm{fs}. \tag{VIII.1}$$

Published on the European Physics Journal C (see Ref. [141]), it dominates the current world average as calculated by the *Heavy Flavour Averaging* group [161],

$$\tau_{B_c^+}^{(\mathrm{HFAG})} = 500 \pm 13\ \mathrm{fs}. \tag{VIII.2}$$

This measurement represents an important milestone in the B_c^+ physics as it reduces the uncertainty on many B_c^+ physics analyses with selection based on the detachment of the B_c^+ meson from the primary vertices. While extremely powerful, these criteria introduce an efficiency depending on the lifetime of the B_c^+ meson. The uncertainties on the latter reflects into uncertainty on the efficiency corrected number of events, affecting branching fraction and production measurements. Among the recent analyses benefiting of the improved precision in the lifetime measurement, two notable examples are the measurement of the branching ratio of $B_c^+ \to J/\psi\,\mu^+\nu_\mu$ decays relative to $B_c^+ \to J/\psi\,\pi^+$ [77] and the first observation of a baryonic B_c^+ decay [66].

The importance of the result has been acknowledged with the publication of an outreach article on the public web page of the LHCb Collaboration, and on the CERN Courier [162].

A further input to the average is expected to be published soon by the LHCb Collaboration analysing the fully reconstructed hadronic decay $J/\psi\,\pi^+$. The analysis has totally independent statistical and systematic uncertainty and is competitive with the measurement obtained studying semileptonic decays. The B_c^+ lifetime measured with

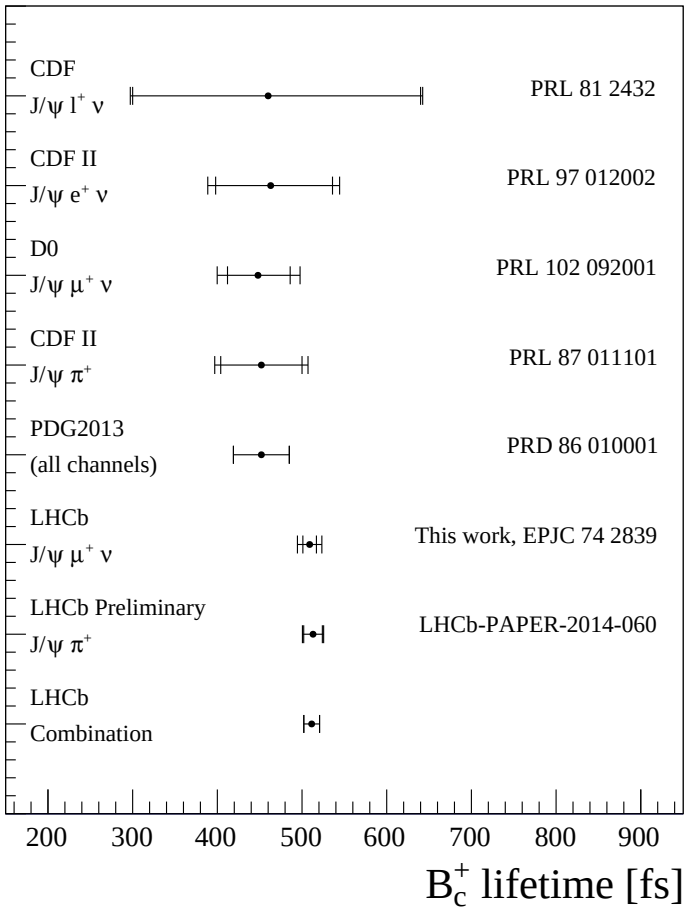

Figure VIII.1: Comparison of some experimental results on the B_c^+ lifetime.

studying fully reconstructed $B_c^+ \to J/\psi\,\pi^+$ decays is [163]

$$\tau_{B_c^+}^{(\mathrm{LHCb},J/\psi\,\pi^+)} = 513 \pm 11 \pm 6 \text{ fs}, \tag{VIII.3}$$

where the systematic uncertainty is dominated by the determination of the acceptance function based on simulation.

The combination of the LHCb results yields to the B_c^+ lifetime

$$\tau_{B_c^+}^{(\mathrm{LHCb,\ combination})} = 511 \pm 9 \text{ fs}. \tag{VIII.4}$$

The comparison of the experimental results obtained for the B_c^+ lifetime is shown in Figure VIII.1.

The dominant systematic uncertainty of the measurement achieved with semileptonic decays is associated to the template distributions used to model background components,

and since these templates are obtained with data-driven methods, the large datasets that will be collected in the future runs of LHCb may improve the precision of the measurement. Indeed, background could be better rejected and modeled thanks to the increased statistics. Still, it is more reasonable to imagine that the next lifetime measurement will be performed with high statistics fully reconstructed decays such as $B_c^+ \to J/\psi\,\pi^+$, already competitive with the measurement presented here, but still dominated by the statistical uncertainty.

During the development of the analysis, I had the opportunity to develop a few original general-purpose techniques now made available to the LHCb collaboration. The most important examples are: the k-formalism disentangling analytically the kinematics and resolution effects on the data model; the misidentification background modelling technique based on the fit of a PID unbiased sample using template distributions obtained from calibration samples, and the relative treatment of the statistical uncertainties; and the data-driven technique used to constrain the systematic uncertainty due to model dependence.

Measurements of the B_c^+ lifetime are also possible at the other LHC experiments, so that it is reasonable to expect contributions from ATLAS and CMS especially when the large statistics of new runs will be available, if their trigger strategy will include B_c^+ candidates. The production of B_c^+ mesons is kinematically forbidden at Super-Belle. Contributions to B_c^+ physics by experiments at future e^+e^- colliders seem also unlikely because of the small B_c^+ production cross-section.

Part 3

Supplementary material

Appendix

Supplementary material

A Density Estimation Trees

The usage of nonparametric density estimation techniques has seen a quick growth in the latest years both in High Energy Physics and in other fields of science dealing with large data samples. Indeed, the improved computing resources and data visualization techniques have made large dimensionality problems approachable, but multi-dimensional parametrization are often non trivial. Consider the probability density function $f(x)$ of a continuous-valued random variable X, and assume that $x_1, x_2, ..., x_{N_{\text{tot}}}$ is a sample of observations, independent realizations of X. The idea of the nonparametric approach is to avoid restrictive assumptions about the form $f(x)$ and to estimate it directly from the data. Histograms are well-known, widely used nonparametric density estimators. The model of the misidentification background discussed in this Thesis, is obtained with a smoothed histogram, because the joint two-dimensional distribution cannot be easily described by some analytical model with a reasonable number of parameters.

Histograms have the advantage of simplicity, they can be trained quickly over large data samples, and read even faster. Among disadvantages there is the lack of continuity, which can be only partially solved by smoothing, and the arbitrariness in the definition of the binning scheme. Besides, severe issues arise using histograms to model distributions with low-statistics long tails, because empty bins often cause problems when trying to consider the normalized histogram as a *pdf*. Figure A.1a and A.1b represent a density estimation based on histograms with different starting points of the binning scheme. The outcome of the same analysis using the former or the latter *pdf* as template function of some data model may be very different.

A possible alternative is to "build the bins" centered on each data entry and obtain the value of the *pdf* in x_0 by counting the number of bins including x_0. This technique, whose application is illustrated in Figure A.1c, is named box-kernel density estimation, and reduces the degrees of arbitrariness in the definition of a *pdf* to the choice of the kernel width. As a drawback, the distribution obtained has border effects including fast ripples which are unphysical because they can lay well below the experimental resolution, especially for large data samples. The generalization of the box-kernel technique is the

kernel density estimation. Instead of the box kernel,

$$k_{\text{box}}(x) = \begin{cases} 1 & \text{if } |x| < \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases} , \tag{A.1}$$

a generic kernel function $k(x)$ is used, to define the *pdf* estimator as

$$\hat{f}(x) = \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} \frac{1}{h} k\left(\frac{x - x_i}{h}\right). \tag{A.2}$$

The width of the kernel is regulated through the parameter h, named bandwidth, and the normalization of the *pdf* is guaranteed if the condition

$$\int \mathrm{d}z \; k(z) = 1 \tag{A.3}$$

is satisfied. A list of kernel functions used in the literature is presented in Table A.1, in High Energy Physics the Gaussian kernel is the most used [151]. In the analysis presented in this Thesis, the Gaussian kernel estimation is used for both the fake-J/ψ background model and the mass distribution of the combinatorial background. These are two examples of low-statistics data samples for which kernel density estimation requires a reasonable amount of computing resources. For larger datasets, with $N_{\text{tot}} \sim \mathcal{O}(10^6)$ entries and in a d-dimensional space, the computational cost of storing the $d \times N_{\text{tot}}$ parameters of kernels and summing a million of kernel functions per evaluation of the *pdf* estimator may become prohibitive. Some implementations cache the estimated *pdf* on a fine grid and use interpolation to improve the speed of the evaluation of the *pdf*, notably during fit operations [164]. However, the memory management of such a sampled *pdf* may be non trivial as the number of dimensions increases. Finally, the choice of the optimal bandwidth has been treated in some depth for one-dimensional kernels [165], but it is not mature for multi-dimensional problems. The optimal bandwidth is usually determined by minimizing the *Mean Integrated Squared Error* (MISE)

$$\text{MISE}(h) = \text{E}\left(\int \left(\hat{f}(x) - f(x)\right)^2 \mathrm{d}x\right), \tag{A.4}$$

varying the bandwidth h. Here, E denotes the expected value with respect to a generic sample of N_{tot} entries. It can be shown that [165]

$$\text{MISE}(h) \approx \int \left(\frac{1}{4} h^4 k_2^2 \left(f''(x)\right)^2 + \frac{1}{N_{\text{tot}} h} f(x) j_2\right) \mathrm{d}x \tag{A.5}$$

where N_{tot} is the number of entries in the sample,

$$k_2 = \int z^2 k(z) \mathrm{d}z, \qquad \text{and} \qquad j_2 = \int k(z)^2 \mathrm{d}z. \tag{A.6}$$

Assuming no dependence of the bandwidth on the variable x, *i.e.* renouncing to adaptive bandwidth, one can solve the equation

$$\frac{\mathrm{dMISE}(h)}{\mathrm{d}h} = 0 \tag{A.7}$$

to find the optimal bandwidth

$$h_{\mathrm{opt}} = \left(\frac{1}{N_{\mathrm{tot}}} \frac{j_2}{k_2^2} \frac{1}{\beta(f)} \right)^{\frac{1}{5}}. \tag{A.8}$$

The second derivative of the generating *pdf* $f(x)$ is enclosed in the parameter

$$\beta(f) = \int f''(x)^2 \mathrm{d}x, \tag{A.9}$$

and is unknown as it is unknown the *pdf* $f(x)$. It is customary to adopt the optimal value of the normal distribution

$$\beta(\mathrm{Gaussian}) = \frac{3}{8} \frac{\sigma^{-5}}{\sqrt{\pi}}, \tag{A.10}$$

because of the good properties of the estimator of its standard deviation which allows to write

$$h_{\mathrm{opt}} = \left(\frac{1}{N_{\mathrm{tot}}} \frac{j_2}{k_2^2} \frac{3}{8\sqrt{\pi}} \right)^{\frac{1}{5}} \hat{\sigma}. \tag{A.11}$$

with

$$\hat{\sigma} = \sqrt{\frac{1}{N_{\mathrm{tot}} - 1} \sum_{i=1}^{N_{\mathrm{tot}}} (x_i - \overline{x})^2}. \tag{A.12}$$

Considering the tent or triangular kernel function, the optimal bandwidth further simplifies into

$$h_{\mathrm{opt}}^{(\mathrm{tent})} = \frac{2^{\frac{6}{5}} \, \pi^{\frac{1}{10}}}{N_{\mathrm{tot}}^{\frac{1}{5}}} \hat{\sigma} \sim 2.58 \frac{\hat{\sigma}}{N_{\mathrm{tot}}^{1/5}} \tag{A.13}$$

This important result will be useful in the treatment of another nonparametric density estimator based on decision trees which is the subject of this Chapter. The *pdf* estimated with decision trees is less accurate than what can be achieved using kernel density estimation, but both training and evaluation are extremely faster. The speed and the high scalability to problems in many dimensions make the decision tree estimation a useful tool for exploratory data analysis.

A.1 The algorithm

A.1.1 Decision Trees

A decision tree is an algorithm or a flowchart composed of internal *nodes* representing tests of a variable or of a property. Nodes are connected to form *branches*, which terminate into a *leaf*, associated to a *decision*.

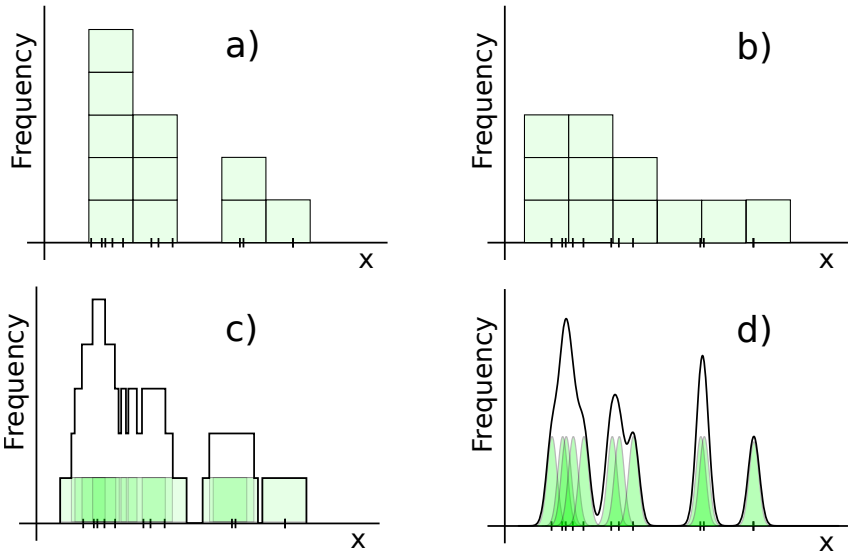

Figure A.1: Four examples of nonparametric density estimation techniques applied to the same sample. Figure a) and b) illustrate how different can result the pdf obtained from an histogram varying only the offset of the binning scheme. Figure c) represents a box-kernel density estimation, which reduces the degrees of arbitrariness to the choice of the width of the kernel, at the price of worsening of border effects including fast ripples beyond the actual resolution of the measurement. Figure d) shows the Gaussian kernel density estimation with a non-optimal choice of the width of the Gaussian functions.

Epancechnikv	$\begin{cases} \frac{3}{4}(1-\frac{1}{5}t^2)/\sqrt{5} & \text{if } \lvert t\rvert < \sqrt{5} \\ 0 & \text{elsewhere} \end{cases}$
Biweight	$\begin{cases} \frac{15}{16}(1-t^2)^2 & \text{if } \lvert t\rvert < 1 \\ 0 & \text{elsewhere} \end{cases}$
Triangular	$\begin{cases} 1-\lvert t\rvert & \text{for } \lvert t\rvert < 1 \\ 0 & \text{elsewhere} \end{cases}$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-t^2/2}$
Box	$\begin{cases} 1 & \text{if } \lvert x\rvert < \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$

Table A.1: Six kernel functions discussed for example in Ref. [165].

Decision Trees are widely used in *classification* problems. In High Energy Physics the *decisions* are usually *Signal* and *Background*, and the variables tested in the *nodes* are quantities obtained through the candidate reconstruction [166].

More rarely, Decision Trees are used in *regression* problems where the *decision* is the

best estimation of some parameter given the variables tested in the decision nodes.

Both *classification* and *regression* problems are so-called supervised machine learning algorithms because they need a training sample modelling the relation between the input variables and the decision. Such a relation is then "learned" by the decision tree in a *training* phase, and later applied to real data (*application*). The complementary class of machine learning algorithms is named *unsupervised* machine learning and is based on clustering or multi-modal density estimation. Density estimation is an unsupervised machine learning algorithm not based on clustering (but clusters can be identified studying the local maxima of the estimated *pdf*).

Density Estimation Trees (DET) are algorithms used to estimate the joint probability density function of a d-dimensional data sample by defining a piecewise constant estimator structured as a decision tree.

A.1.2 Training (or learning)

Consider the generic data sample $\mathcal{S}$ of N_{tot} observations in $\mathbb{R}^d$. The i-th observation is associated to d coordinates $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, ..., x_i^{(d)})$. The piecewise constant density estimate of the sample $\mathcal{S}$ is defined as

$$\hat{f}(\mathbf{x}) = \begin{cases} \frac{N(\text{bin}_1)}{V(\text{bin}_1)N_{\text{tot}}} & \text{if } \mathbf{x} \in \text{bin}_1 \\ \frac{N(\text{bin}_2)}{V(\text{bin}_2)N_{\text{tot}}} & \text{if } \mathbf{x} \in \text{bin}_2 \\ ... & \\ \frac{N(\text{bin}_{N_{\text{bin}}})}{V(\text{bin}_{N_{\text{bin}}})N_{\text{tot}}} & \text{if } \mathbf{x} \in \text{bin}_{N_{\text{bin}}} \end{cases}, \tag{A.14}$$

where $N(\text{bin}_i)$ is the number of entries included in the i-th bin, and $V(\text{bin}_i)$ is its volume. N_{bin} is the number of d-dimensional bins. The normalization of the estimated *pdf* is guaranteed by the closure relations

$$\sum_{i=1}^{N_{\text{bin}}} N(\text{bin}_i) = N_{\text{tot}} \qquad \text{and} \qquad \sum_{i=1}^{N_{\text{bin}}} V(\text{bin}_i) = V\left(\bigcup_{i=1}^{N_{\text{bin}}} \text{bin}_i\right). \tag{A.15}$$

Introducing the characteristic function

$$\mathbb{I}(\mathbf{x} \in \text{bin}_i) = \begin{cases} 1 & \text{if } x \in \text{bin}_i \\ 0 & \text{if } x \notin \text{bin}_i \end{cases}, \tag{A.16}$$

Equation A.14 can be shortened to

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{N_{\text{bin}}} \frac{1}{N_{\text{tot}}} \frac{N(\text{bin}_i)}{V(\text{bin}_i)} \mathbb{I}(\mathbf{x} \in \text{bin}_i). \tag{A.17}$$

The optimal approximation of the probability density function $f(\mathbf{x})$, minimizes the Integrated Squared Error (ISE), which has a similar definition to the MISE defined in

Equation A.4, but it is specific to the data sample considered instead of taking the expected value for a generic realization of the *pdf* $f(x)$. Namely,

$$\mathcal{R} = \mathrm{ISE}(\hat{f}, f) = \int_{\mathbb{R}^d} (\hat{f}(\mathbf{x})) - f(\mathbf{x}))^2 \mathrm{d}\mathbf{x} \tag{A.18}$$

which can be written as

$$\mathcal{R} = \int_{\mathbb{R}^d} \left[f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right]^2 \mathrm{d}x = \int_{\mathbb{R}^d} f(\mathbf{x})^2 \mathrm{d}\mathbf{x} + \int_{\mathbb{R}^d} \hat{f}(\mathbf{x})^2 \mathrm{d}\mathbf{x} - 2 \int_{\mathbb{R}^d} \hat{f}(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{A.19}$$

The first integral does not contribute to the optimization of the piecewise *pdf* because is independent of the binning choice, depending only on the real *pdf* $f(\mathbf{x})$. The second integral can be rewritten replacing $f(\mathbf{x})$ with its definition as shown in Equation A.17,

$$\int_{\mathbb{R}^d} \hat{f}(\mathbf{x})^2 \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^d} \left[\sum_{i=1}^{N_{\mathrm{bin}}} \frac{1}{N_{\mathrm{tot}}} \frac{N(\mathrm{bin}_i)}{V(\mathrm{bin}_i)} \mathbb{I}(\mathbf{x} \in \mathrm{bin}_i) \right]^2 \mathrm{d}\mathbf{x}. \tag{A.20}$$

The cross terms in the expansion of the squared sum vanish because of the following property of the characteristic function,

$$\mathbb{I}(\mathbf{x} \in \mathrm{bin}_i) \mathbb{I}(\mathbf{x} \in \mathrm{bin}_j) = \delta_{ij}\ \mathbb{I}(\mathbf{x} \in \mathrm{bin}_i), \tag{A.21}$$

therefore,

$$\int_{\mathbb{R}^d} \hat{f}(\mathbf{x})^2 \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^d} \sum_{i=1}^{N_{\mathrm{bin}}} \frac{1}{N_{\mathrm{tot}}^2} \frac{N(\mathrm{bin}_i)^2}{V(\mathrm{bin}_i)^2}\ \mathbb{I}(x \in \mathrm{bin}_i) \mathrm{d}\mathbf{x}. \tag{A.22}$$

The only term depending on $\mathbf{x}$ is the characteristic function, all of the other terms can be factorized to give

$$\int_{\mathbb{R}^d} \hat{f}(\mathbf{x})^2 \mathrm{d}\mathbf{x} = \sum_{i=1}^{N_{\mathrm{bin}}} \frac{1}{N_{\mathrm{tot}}^2} \frac{N(\mathrm{bin}_i)^2}{V(\mathrm{bin}_i)^2}\ \int_{\mathbb{R}^d} \mathbb{I}(x \in \mathrm{bin}_i) \mathrm{d}\mathbf{x} = \sum_{i=1}^{N_{\mathrm{bin}}} \frac{1}{N_{\mathrm{tot}}^2} \frac{N(\mathrm{bin}_i)^2}{V(\mathrm{bin}_i)^2}\ V(\mathrm{bin}_i). \tag{A.23}$$

The third integral of Equation A.19 can be approximated by a sum considering the following property of Monte Carlo integration,

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} g(x_i) = \int_{\mathbb{R}} g(x) h(x) \mathrm{d}x \qquad \text{where } x_1, x_2, ..., x_N \text{ distribute as } h(x). \tag{A.24}$$

Hence, in the approximation of a large sample $\mathcal{S}$,

$$-2 \int_{\mathbb{R}^d} \hat{f}(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x} = -\frac{2}{N_{\mathrm{tot}}} \sum_{i=1}^{N_{\mathrm{tot}}} \sum_{j=1}^{N_{\mathrm{bin}}} \frac{1}{N_{\mathrm{tot}}} \frac{N(\mathrm{bin}_j)}{V(\mathrm{bin}_j)} \mathbb{I}(\mathbf{x}_i \in \mathrm{bin}_j), \tag{A.25}$$

through a simple factorization,

$$-2\int_{\mathbb{R}^d}\hat{f}(\mathbf{x})f(\mathbf{x})\mathrm{d}\mathbf{x} = -2\sum_{j=1}^{N_{\mathrm{bin}}}\frac{1}{N_{\mathrm{tot}}^2}\frac{N(\mathrm{bin}_j)}{V(\mathrm{bin}_j)}\sum_{i=1}^{N_{\mathrm{tot}}}\mathbb{I}(\mathbf{x}_i\in\mathrm{bin}_j) = -2\sum_{j=1}^{N_{\mathrm{bin}}}\frac{1}{N_{\mathrm{tot}}^2}\frac{N(\mathrm{bin}_j)^2}{V(\mathrm{bin}_j)}. \tag{A.26}$$

The optimization problem reduces to the minimization of the quantity

$$\begin{aligned}\mathcal{R} &= \int_{\mathbb{R}^d}[f(x)^2 - \hat{f}(x)^2]\mathrm{d}\mathbf{x}\\ &= \sum_{i=1}^{N_{\mathrm{bin}}}\left(\frac{N^2(\mathrm{bin}_i)}{N_{\mathrm{tot}}^2V(\mathrm{bin}_i)}\right) - 2\sum_{i=1}^{N_{\mathrm{bin}}}\left(\frac{N^2(\mathrm{bin}_i)}{N_{\mathrm{tot}}^2V(\mathrm{bin}_i)}\right) + \mathrm{const}\\ &= \sum_{i=1}^{N_{\mathrm{bin}}}\left(-\frac{N^2(\mathrm{bin}_i)}{N_{\mathrm{tot}}^2V(\mathrm{bin}_i)}\right) + \mathrm{const}.\end{aligned} \tag{A.27}$$

In the construction of the density estimation tree, the single contributions to the sum are minimized separately, by defining the single-bin replacement error

$$R(\mathrm{bin}_i) = -\frac{N^2(\mathrm{bin}_i)}{N_{\mathrm{tot}}^2V(\mathrm{bin}_i)}, \tag{A.28}$$

and iteratively splitting each bin b to two sub-bins b_L and b_R according to the criteria maximising the residual gain

$$G(b) = R(b) - R(b_L) - R(b_R), \tag{A.29}$$

chosen among the $d\times(N(b)-1)$ possible criteria combining one of the d variables and one of the thresholds set at the middle point between two adjacent candidates in bin b. The iterative process terminates when a stop condition is matched usually a minimal number of entries per bin. Namely, the loop stops when the condition

$$N(b_L) < N_{\mathrm{min}} \qquad \text{or} \qquad N(b_R) < N_{\mathrm{min}} \tag{A.30}$$

is matched. The procedure is sketched in Figure A.2.

In order to contrast effects of overtraining I found useful to add an alternative stop condition based on the widths of the bin in each dimension. Setting this minimal width to a value well below the resolution of the points does not reduce the adaptability of the algorithm, but avoid spikes when a few points are found to be aligned in at least one projection. Technically, this condition denies the splitting of a bin when one of the two sub-bins has a width below the threshold, but allows the splitting in any other dimension, or even in the same dimension with a different threshold.

A.1.3 Pruning and cross-validation

As it happens for supervised machine learning techniques, Density Estimation Trees can be overtrained. Overtraining (or overfitting) occurs when the statistical model obtained

through the Density Estimation Tree describes random errors or noise instead of the underlying *pdf*. The effect is an exaggerate care payed to minor fluctuations in the data, resulting in isolated bins with small volume corresponding to high *pdf* values.

Overtraining can be reduced by stopping the growth of the decision tree through a criterion defined *a posteriori*, comparing the statistical agreement of the estimated *pdf* with some alternative estimator. The procedure of merging contiguous bins to retrace the decision tree towards the top, is named *pruning*. Clearly, pruning tree branches makes the Integrated Squared Error defined in Equation A.18 larger. To compensate for this effect, a complexity penalty is assigned to poorly populated bins. The idea is to sort the constructed bins in terms of the actual improvement they introduce in the statistical description of the data model and to remove those whose effect is a worsening of the agreement with the alternative, usually more expensive, estimation.

Following the common procedure for classification and regression trees, I define the regularized error

$$R_\alpha(\text{bin}_i) = \sum_{j \in \text{leaves of bin}_i} R(\text{bin}_j) + \alpha C(\text{bin}_i), \tag{A.31}$$

where α is named *regularization parameter*, and the index j runs over the sub-bins of bin_i with no further sub-bins, or leaves. $C(\text{bin}_i)$ is the complexity function of bin_i, as depicted in Figure A.3. In the literature $C(\text{bin}_i)$ usually is the number of terminal nodes, or leaves attached to a subtree, or branch. This complexity function is very useful for classification trees because it provides a top-down simplification technique complementary to the stopping rule. For density estimation trees, I found that the choice of the stopping rule is so important that providing an optimized pruning of a tree trained with a suboptimal stopping rule is totally useless. Therefore I prefer choosing the bin depth as complexity function, providing a down-top pruning which can be seen as an *a posteriori* optimization of the stopping rule. The bin depth of the root node is defined to be 1, its two sub-bins have depth 2, their sub-bins 3, and so on. The accuracy achieved with the complexity function based on the number of leaves can be slightly better if the stop condition is optimal. On the other hand, the pruned tree obtained with depth-based complexity function is independent on the stop criterion (provided it is loose enough not to be competitive with the pruning itself), and therefore is a worth choice for large samples and exploratory data analysis.

If $R_\alpha(\text{bin}_i) > R(\text{bin}_j)$ the splitting of the i-th bin is considered useless and all its sub-bins are merged, or pruned. A threshold value α_i of the regularization parameter α is associated to each i-th bin, defining the minimum value of α which would cause the bin merging:

$$\alpha_i = \frac{1}{C(\text{bin}_i)} \left(R(\text{bin}_i) - \sum_{j \in \text{leaves of bin}_i} R(\text{bin}_j) \right). \tag{A.32}$$

All the bins having sub-bins are then sorted according to the regularization parameter α_i. Defined some quality function $Q(\alpha)$, discussed later, the optimal binning is obtained for

$$\alpha = \alpha_{\text{best}} \quad : \quad Q(\alpha_{\text{best}}) = \max_{\alpha \in \{\alpha_i\}_i} Q(\alpha). \tag{A.33}$$

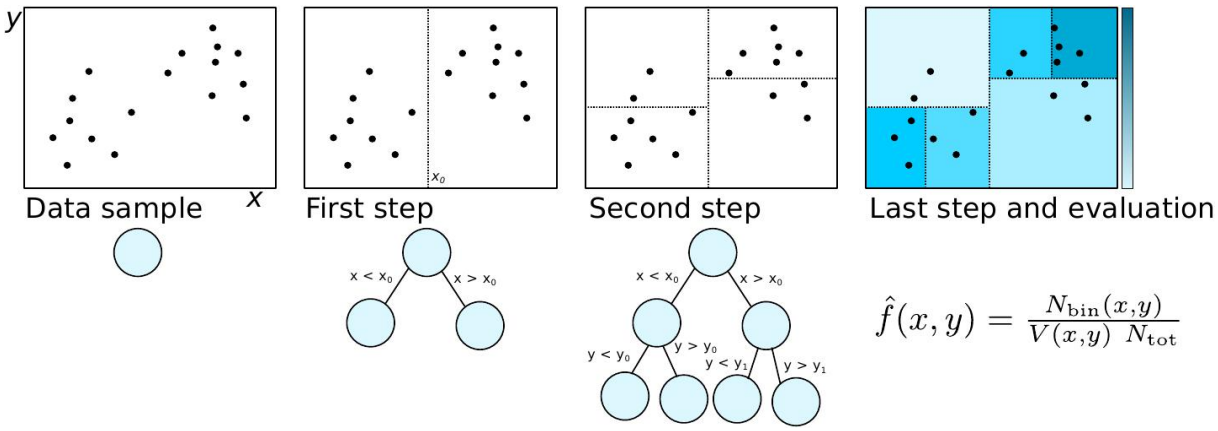

Figure A.2: Simple example of training of a density estimation tree over a two-dimensional sample.

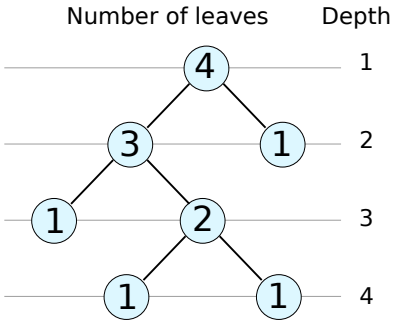

Figure A.3: Two examples of complexity function based on the number of leaves or subtrees, or on the node depth.

The determination of the optimal regularization parameter is named *cross-validation* and many different techniques are possible, depending on the choice of the quality function $Q(\alpha)$.

Leave-One-Out Cross Validation (LOO-CV)

A commonly adopted cross validation technique consists in the estimation of the *pdf* through the resamping of the original dataset. When training supervised machine learning algorithms it is common to split the known sample in two statistically identical subsamples used for training and validation. In order not to loose the statistical power of the full dataset, it is customary to split the training in n steps. At each step a fraction $\frac{1}{n}$ of the sample is used to validate the algorithm, and the rest for the training. If the number of steps equals the number of entries in the sample, then the cross validation technique is named *Leave-One-Out cross validation* (LOO-CV). The principle of LOO-CV is extended to unsupervised machine learning algorithms, and in particular to density estimation, by

re-defining the Integrated Squared Error

$$\mathcal{R}_{\mathrm{LOO}}(\alpha) = \int_{\mathbb{R}^d} \left(\hat{f}^{\alpha}(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x} - \frac{2}{N}\sum_{i=1}^{N_{\mathrm{tot}}} \hat{f}^{\alpha}_{\mathrm{not}\,i}(\mathbf{x}_i) \tag{A.34}$$

where $\hat{f}^{\alpha}(\mathbf{x})$ is the density estimator pruned with complexity parameter α, and $\hat{f}^{\alpha}_{\mathrm{not}\,i}$ is the density estimator built from a data sample without the i-th entry $\mathbf{x}_i$. The quality function is simply

$$Q(\alpha) = -R_{\mathrm{LOO}}(\alpha). \tag{A.35}$$

The application of Leave-one-out cross validation is extremely slow, besides it requires to build one decision tree per entry. When considering the application of Density Estimation Trees to very large samples, with more than a million of entries, the construction of one million of decision trees per one million of possible pruning constant in unreasonable. Even accepting large CPU time expenses, the method would require a large amount of fast memory to run efficiently.

Nearest-Neighbour distance

As a difference with supervised machine learning techniques, the density estimation can benefit of validation techniques developed for the assessment of the goodness of unbinned fits in multiple dimensions. In Ref. [167], a review of the techniques used in high-energy physics is given. Most of the techniques discussed involve random numbers either for resampling or for Monte Carlo integrations. In order to preserve some determinism in the behaviour of the density estimation tree, essential during its development, I preferred avoiding these methods and considering instead the comparison between the *nearest-neighbour distance* and the value of the *pdf* to assess the quality of the agreement as a function of the regularization parameter.

The method has been discussed in Section VI.4.3 to test the goodness of the two-dimensional fit used to measure the B_c^+ lifetime. The random variable U_i is defined as

$$U_i = \exp\left(-N_{\mathrm{tot}}\int_{|\mathbf{x}-\mathbf{x}_i|<R_i^{nn}} \hat{f}_{\alpha}(\mathbf{x})\mathrm{d}\mathbf{x}\right) \approx \exp(-N_{\mathrm{tot}}\hat{f}_{\alpha}(\mathbf{x})V_d(R_i^{nn})), \tag{A.36}$$

where R_i^{nn} is the distance from the i^{th} event to its nearest neighbour and $V_d(R)$ is the d-dimensional hyper-spherical volume of radius R. The approximation is valid for a smooth *pdf* which can be assumed constant within the integration domain. Under the hypothesis $f(\mathbf{x}) = f_{\alpha}(\mathbf{x})$, the distribution of U_i is expected to be flat.

There are many possible methods to test for flatness. I perform a binned χ^2 test and use as quality function

$$Q^{nn}(\alpha) = -\chi^2_{\alpha} \tag{A.37}$$

The technique has the great advantage of being independent of any additional parameter, but for each value of α, all the U_i random variables have to be recalculated, even caching the distance to the closest neighbour. This requires to evaluate the pruned tree $N_{\mathrm{tot}} \times N_{\mathrm{bin}}$, leading to high computational cost for large samples and dimensions.

Kernel-based cross-validation

The computational cost of the nearest neighbour method can be partially reduced comparing the *pdf* $f^{\alpha}(\mathbf{x})$ assessed using a pruned decision tree, to a kernel density estimation $f_k(\mathbf{x})$ and defining the quality function as

$$Q^{ker}(\alpha) = -\int_{\mathbb{R}^d} \left(f_\alpha(\mathbf{x}) - f_k(\mathbf{x})\right)^2 \mathrm{d}\mathbf{x}. \tag{A.38}$$

The squared term is expanded to give

$$\begin{aligned} Q^{ker}(\alpha) = & -\int_{\mathbb{R}^d} \left(f_\alpha(x)^2 - 2f_\alpha(x)f_k(x) + f_k(x)^2\right) \mathrm{d}\mathbf{x} \\ = & -\int_{\mathbb{R}^d} f_\alpha(x)^2 \mathrm{d}\mathbf{x} + 2\int_{\mathbb{R}^d} f_\alpha(x)f_k(x)\mathrm{d}\mathbf{x} - \int_{\mathbb{R}^d} f_k(x)^2 \mathrm{d}\mathbf{x} \end{aligned}$$

The third integral term is dropped because it does not depend on α and it is therefore useless in the process of maximization of the quality function. The first two integrals are treated separately below, substituting to $f_\alpha(x)$ and $f_k(x)$ their expressions as in Equations A.17 and A.2, respectively. The first integral was already solved in Equation A.23 and reads

$$\int_{\mathbb{R}^d} \hat{f}_\alpha(\mathbf{x})^2 \mathrm{d}\mathbf{x} = \frac{1}{N_{\text{tot}}^2} \sum_{i=1}^{N_{\text{bin}}} \frac{N(\text{bin}_i^\alpha)^2}{V(\text{bin}_i^\alpha)} \tag{A.39}$$

The second integral can be written as

$$\begin{aligned} -2\int_{\mathbb{R}^d} f_0(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x} = & \\ -2\int_{\mathbb{R}^d} \left\{ \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} k_d\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_d}\right) \right\} & \left\{ \sum_{j=1}^{N_{\text{bin}}} \frac{1}{N_{\text{tot}}} \frac{N(\text{bin}_j)}{V(\text{bin}_j)} \mathbb{I}(\mathbf{x} \in \text{bin}_j) \right\} \mathrm{d}\mathbf{x}. \end{aligned} \tag{A.40}$$

With this notation, k_d is a normalized multidimensional kernel function. To explicit the normalization and the different bandwidth associated to each dimension of the problem, one writes

$$k_d\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_d}\right) = \prod_{k=1}^{d} \frac{1}{h_k} k\left(\frac{x^{(k)} - x_i^{(k)}}{h_k}\right). \tag{A.41}$$

Using a triangular kernel simplifies the problem in terms of computational cost because integrals may be computed analytically and efficiently, therefore I choose

$$k_d\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_d}\right) = \prod_{k=1}^{N_k} \frac{1}{h_k} \left(1 - \frac{|x^{(k)} - x_i^{(k)}|}{h_k}\right) \theta\left(1 - \frac{|x^{(k)} - x_i^{(k)}|}{h_k}\right), \tag{A.42}$$

where $\theta(x)$ is the Heaviside step function.

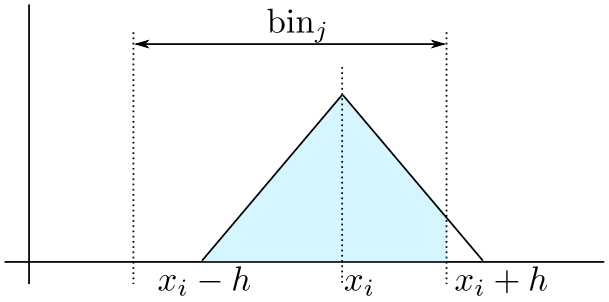

Figure A.4: Representation of the integral of the kernel function as in Equation A.43.

With a simple factorization Equation A.40 can be rewritten as

$$
\begin{aligned}
-2\int_{\mathbb{R}^d} f_0(\mathbf{x}) f(\mathbf{x}) \mathrm{d}\mathbf{x} = \\
- \frac{2}{N_{\mathrm{tot}}^2} \sum_{i=1}^{N_{\mathrm{tot}}} \sum_{j=1}^{N_{\mathrm{bin}}} \frac{N(\mathrm{bin}_j)}{V(\mathrm{bin}_j)} \times \\
\times \prod_{k=1}^{d} \frac{1}{h_k} \int_{-\infty}^{+\infty} \left(1 - \frac{|x^{(k)} - x_i^{(k)}|}{h_k}\right) \theta\left(1 - \frac{|x^{(k)} - x_i^{(k)}|}{h_k}\right) \mathbb{I}\left(x^{(k)} \in \mathrm{bin}_j^{(k)}\right) \mathrm{d}x^{(k)}.
\end{aligned}
\tag{A.43}
$$

The integral is illustrated in Figure A.4 and can be rewritten as

$$
\mathcal{I}_{jk}(\mathbf{x}_i; h_k) = \int_{\ell_{ij}^{(k)}}^{u_{ij}^{(k)}} \left(1 - \frac{|x^{(k)} - x_i^{(k)}|}{h_k}\right) \mathrm{d}x^{(k)} \quad \text{with} \quad \begin{cases} u_{ij}^{(k)} = \min(x_{\max}^{(k)}(\mathrm{bin}_j), x_i^{(k)} + h_k) \\ \ell_{ij}^{(k)} = \max(x_{\min}^{(k)}(\mathrm{bin}_j), x_i^{(k)} - h_k) \end{cases},
\tag{A.44}
$$

where $x_{\min}^{(k)}(\mathrm{bin}_j)$ and $x_{\max}^{(k)}(\mathrm{bin}_j)$ represent the boundaries of the j-th bin. Splitting the absolute value into a case function, one gets

$$
\begin{aligned}
\mathcal{I}_{jk}(\mathbf{x}_i; h_k) = \int_{\ell_i^{(k)}}^{u_i^{(k)}} \left(1 - \frac{|x^{(k)} - x_i^{(k)}|}{h_k}\right) \mathrm{d}x^{(k)} = \\
= \int_{\ell_{ij}^{(k)}}^{x_i^{(k)}} \left(1 + \frac{x^{(k)} - x_i^{(k)}}{h_k} \operatorname{sign}(x_i^{(k)} - \ell_{ij}^{(k)})\right) \mathrm{d}x^{(k)} + \\
+ \int_{x_j^{(k)}}^{u_{ij}^{(k)}} \left(1 - \frac{x^{(k)} - x_i^{(k)}}{h_k} \operatorname{sign}(u_{ij}^{(k)} - x_i^{(k)})\right) \mathrm{d}x^{(k)} \\
= u_{ij}^{(k)} - \ell_{ij}^{(k)} - \frac{(u_{ij}^{(k)} - x_i^{(k)})^2}{2h_k} \operatorname{sign}\left(u_{ij}^{(k)} - x_i^{(k)}\right) + \\
+ \frac{(\ell_{ij}^{(k)} - x_i^{(k)})^2}{2h_k} \operatorname{sign}\left(x_i^{(k)} - \ell_{ij}^{(k)}\right)
\end{aligned}
\tag{A.45}
$$

where $\text{sign}(x) = 2\theta(x) - 1$. One concludes that

$$-2\int_{\mathbb{R}^d} f_0(\mathbf{x})f(\mathbf{x})\mathrm{d}\mathbf{x} = -\frac{2}{N_{\text{tot}}^2}\sum_{i=1}^{N_{\text{tot}}}\sum_{j=1}^{N_{\text{bin}}}\prod_{k=1}^{d}\frac{N(\text{bin}_j^\alpha)}{V(\text{bin}_j^\alpha)}\mathcal{I}_{jk}(\mathbf{x}_i; h_k), \tag{A.46}$$

or equivalently

$$-2\int_{\mathbb{R}^d} f_0(\mathbf{x})f(\mathbf{x})\mathrm{d}\mathbf{x} = -\frac{2}{N_{\text{tot}}^2}\sum_{j=1}^{N_{\text{bin}}}\frac{N(\text{bin}_j^\alpha)}{V(\text{bin}_j^\alpha)}\mathcal{N}_j, \qquad \text{with} \qquad \mathcal{N}_j = \sum_{i=1}^{N_{\text{tot}}}\prod_{k=1}^{d}\mathcal{I}_{jk}(\mathbf{x}_i; h_k). \tag{A.47}$$

Note that $\mathcal{N}_j$ can be computed only once for each tree node and cached during the maximization of the quality function

$$Q^{ker}(\alpha) = -\int_{\mathbb{R}^d}(f_\alpha(\mathbf{x}) - f_k(\mathbf{x}))^2\mathrm{d}\mathbf{x} = \frac{1}{N_{\text{tot}}^2}\sum_{j=1}^{N_{\text{bin}}}\frac{N(\text{bin}_j^\alpha)}{V(\text{bin}_j^\alpha)}\left(2\mathcal{N}_j - N(\text{bin}_j)\right) + \text{const.} \tag{A.48}$$

This is an important result because it allows to reduce the complexity of the computation of the quality function down to N_{bin}, reducing by a factor N_{tot} the CPU time with respect to the nearest neighbour method. The disadvantage is in the dependence on the additional parameter h. As a baseline, the optimal value of Equation A.13,

$$h = 2.58\frac{\hat{\sigma}}{N_{\text{tot}}^{1/5}}, \tag{A.49}$$

is used, however this value is optimal for $f(x)$ being a Gaussian distribution, which may be a very wrong assumption. The bandwidth h is arbitrary and constant alongside the data model, its optimal definition and techniques to let it vary in different regions of the data model are possible outlook for the development of these algorithms.

A.1.4 Smoothing techniques

One of the major limitations of the density estimation trees are the sharp boundaries of the bins. When using an estimated density to model some contribution in a fit, this means that an infinitesimal displacement of a data point may result in a finite difference of the fit result, and this may be annoying in particular for toy Monte Carlo studies.

Simple smoothing techniques are discussed bearing in mind that the problem is substantially different from the smoothing of an n-dimensional histogram, because the nearest bins are not trivially defined. The techniques I considered belong to two families, linear interpolation and smearing. The former is computationally very expensive and it is not trivially extendible to more than two dimensions, while the latter has the disadvantage of relying on an additional parameter which is the weight-function width, or smearing resolution, and it is much less accurate than the linear interpolation.

Further work is needed in this sector. For example considering trigonometric interpolation based on Discrete Fourier Transforms.

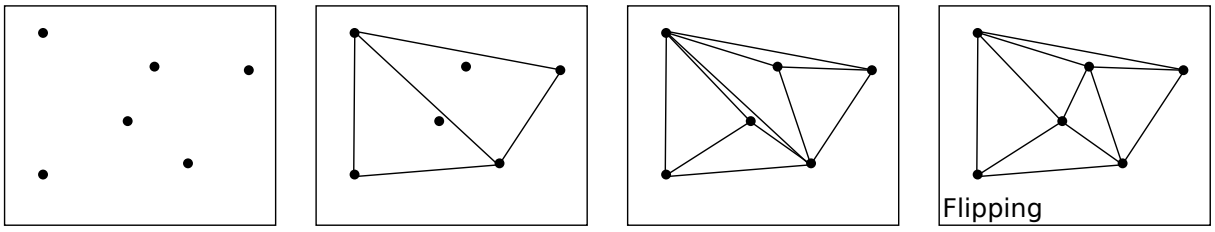

Figure A.5: Example of recursive Delaunay triangulation.

Linear interpolation

Given the complexity of the task, I will restrain the discussion of this technique to the two-dimensional case. Extensions to more dimensions are subject of active research and are definitely beyond the purpose of this discussion. Consider a set T of leaves of a trained decision tree. Each leaf is defined by the coordinates of its center (x_i, y_i), and by the estimated *pdf* f_i. Given a generic point $P \equiv (x, y)$, the method consists in defining the three points (x_a, y_a), (x_b, y_b), and (x_c, y_c), centers of three different leaves, forming the triangle inscribing P with the smallest area. The unique plane passing through the points (x_a, y_a, f_a), (x_b, y_b, f_b), and (x_c, y_c, f_c) can then be defined and its equation used to determine an approximation of $f(x, y)$.

The determination of the three points (x_a, y_a), (x_b, y_b), and (x_c, y_c) is usually based on the Delaunay triangulation of Dirichlet tesselation. For a set S of points in the Euclidean plane, the unique triangulation $DT(S)$ of S such that no point in S is inside the circumcircle of any triangle in $DT(S)$. $DT(S)$ is the dual of the Voronoi diagram of S. If n is the number of points in S, the Voronoi diagram of S is the partitioning of the plane containing S points into n convex polygons such that each polygon contains exactly one point and every point in a given polygon is closer to its central point than to any other.

It can be shown that considering two triangles with a common side, the sum of the opposite angles is less than π only if both are Delaunay triangles. This property can be exploited to build a trivial algorithm to perform a Delaunay triangulation recursively. Given a large triangle, a random point inscribed in the triangle is used to split the triangle in three subtriangles. The procedure is repeated until all the triangles are empty. Then considering all the pairs of triangles with a common side, those which do not correspond to Delaunay triangles are *flipped*, as shown in Figure A.5.

Linear interpolation through Delaunay triangulation is implemented in the ROOT libraries as support for the graphical representation of `TGraph2D` objects. Extensions to further dimensions are not trivial and are not treated in this document.

Triangular smearing

An alternative technique to remove the unphysical discontinuities at bin boundaries is to average the estimated *pdf* $\hat{f}$ in a region centered on $\mathbf{x}$ instead of taking $\hat{f}(\mathbf{x})$ itself. The

smoothed *pdf* will be defined as

$$\hat{f}_s(\mathbf{x}) = \int_{\mathbb{R}^d} \hat{f}(\mathbf{z}) w(\mathbf{x} - \mathbf{z}) \mathrm{d}\mathbf{z} \tag{A.50}$$

where $w(\mathbf{x})$ is a normalized weight function. Choosing a triangular kernel as weight function allows to simplify the integral as discussed in Section A.1.3,

$$\hat{f}_s(\mathbf{x}) = \sum_{j=1}^{N_{\mathrm{bin}}} \prod_{k=1}^{d} \mathcal{I}_{jk}(\mathbf{x}; h_k), \tag{A.51}$$

where $I_{jk}(\mathbf{x}; h_k)$ was defined in Equation A.45. The result of the smearing depends strongly on the choice of the weight function, and of its bandwidth h_k.

Note that the evaluation of $f_s(\mathbf{x})$ does not require to loop on the entries, and that the required integration, for a triangular weight function, is obtained with arithmetical operations only. Hence, the smoothing does not worsen significantly the CPU time needed to read the density estimation tree.

A.2 Implementation

The implementation used to study the various algorithms described above and to test their performance has been realized in C++, using ROOT libraries to ease the import of a dataset stored in a `TTree` class.

The main object is devoted to the training, the pruning and the evaluation of the density estimator, while a subclass `Bin` is introduced to represent the nodes of the decision tree, allowing dynamical allocation of the resources needed to store the tree structure.

`Bin` objects are constructed during the training, each node is responsible for the construction and the destruction of its sub-nodes. At construction, nodes are registered by reference in the main class so that they can be accessed either through a fall-through approach from the root to the leaf, or in a sequential way. The former access method is used at the evaluation time, following the branching down to the terminal node, the latter is used for pruning.

Each `Bin` object contains a copy of the data points it includes, but to save memory, once the bin is split into two sub-bins, the data are transfered to the child bins. In this way, only one copy of each dataset is stored in memory.

Data are organized in columns, each column representing a coordinate. The first step of the bin-split procedure to divide a node is to sort the data of the parent bin. Each column is copied and sorted separately, temporarily loosing any information on the correlation of variables, as only the variable projections are actually considered. The sorted projections are used to define the thresholds, typically the mid point between two adjacent data entries, to be tested in the minimization of the replacement error R. All the thresholds of all the projections are tested, and the one minimizing R is used to create two sub bins.

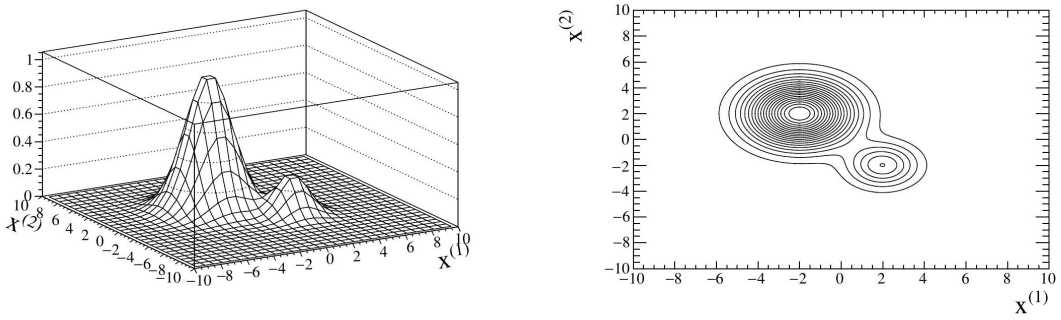

Figure A.6: Theoretical *pdf* used to test the performance of the Density Tree Estimator.

A.3 Tests and applications

Defined the algorithms and briefly discussed its implementation, here I show some application and performance tests. Chosen a generating *pdf* defined analytically, a sample of Monte Carlo generated events is produced and used to train the Density Estimation Tree. The comparison of the Density Estimation Tree, of the histogram and of the original *pdf* allows to discuss some strength and some weakness of the method.

The test function considered is the sum of two two-dimensional Gaussian distributions as shown in Figure A.6.

Random samples of 10, 100, 1000, and 10000 entries are generated and used to train Density Estimation Trees. For this test self-optimization is not used. The growth of the tree is stopped when the bin contains only one element, or when the width of one of its sub-bins would be smaller than 0.3.

The plots of the obtained decision tree are shown in Figure A.7. Overtraining is evident.

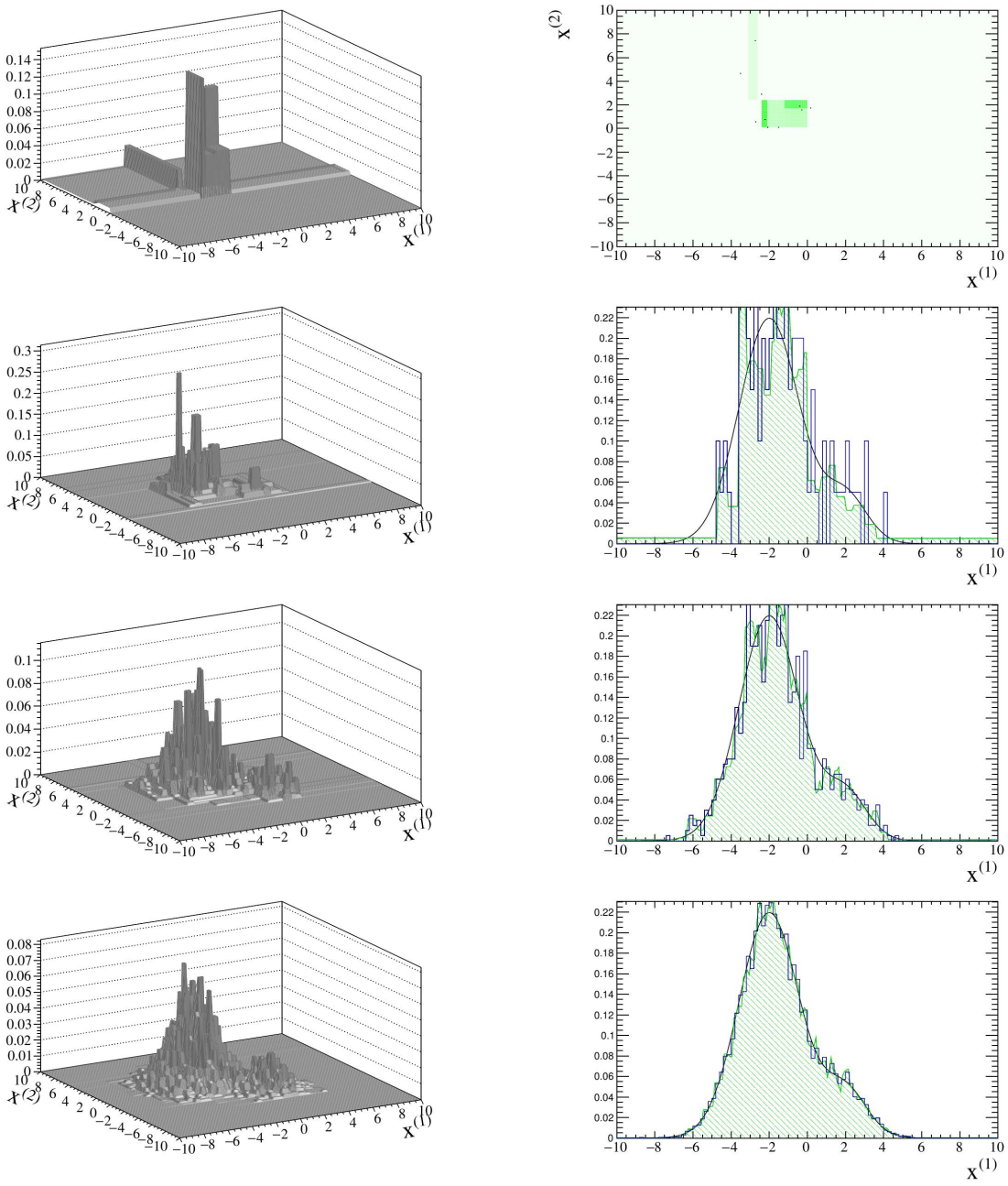

Figure A.7: Density estimations obtained with Density Estimation Trees trained over samples of 10, 100, 1000 and 10000 entries (each row correspond to a different sample). On the left the 3D view of the Density Estimation, on the right the Density Estimation is projected on the coordinate $x^{(1)}$. The black smooth curve represents the generating *pdf*, the histogram of the events is represented as a blue curve while the density estimation based on a decision tree is shown as a dash-filled green line. For the 10-entry sample, the position of the entries is shown superposed to a representation of the *pdf*.

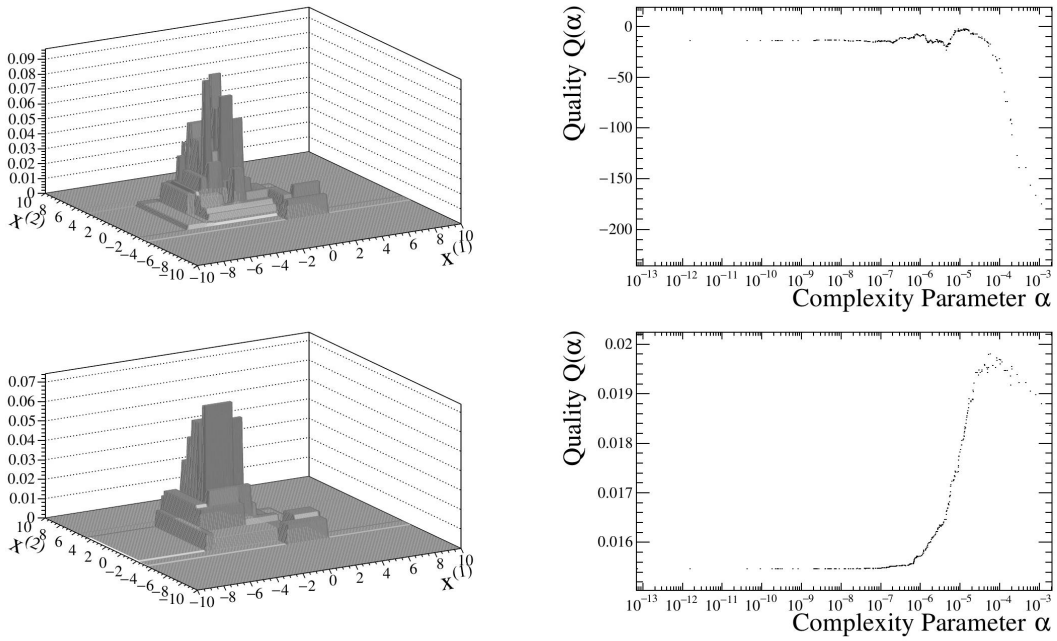

Figure A.8: Density estimation obtained for a sample of 1000 entries and the self-optimization (or pruning) based on the depth-complexity function and using the nearest-neighbour method (top) or the comparison with a triangular kernel density estimation (bottom). On the right, the quality function is shown as a function of the regularization parameter α. The discrete values assumed by α are the threshold value calculated for the each node of the main tree.

A.3.1 Effect of pruning

The effect of pruning is studied over the same small sample of 1000 entries used to train the Density Estimation Tree shown in Figure A.7, using the same stop condition.

Figure A.8 shows the density estimations obtained using the self-optimization based on the depth complexity function and the cross validation techniques based on the nearest neighbour method and the comparison with the triangular kernel density estimation.

Overtraining is less evident than without self-optimization, but still present.

A clever choice of the stop condition and the usage of the complexity function based on the number of leaves can lead to much better results, but may require a few trials to choose a satisfactory stop condition. Figure A.9 shows the density estimation obtained using the number-of-leaves complexity function on a tree built with the following stop condition: no node with less than 3 entries, no bin-width below 0.8. The quality function is also reported.

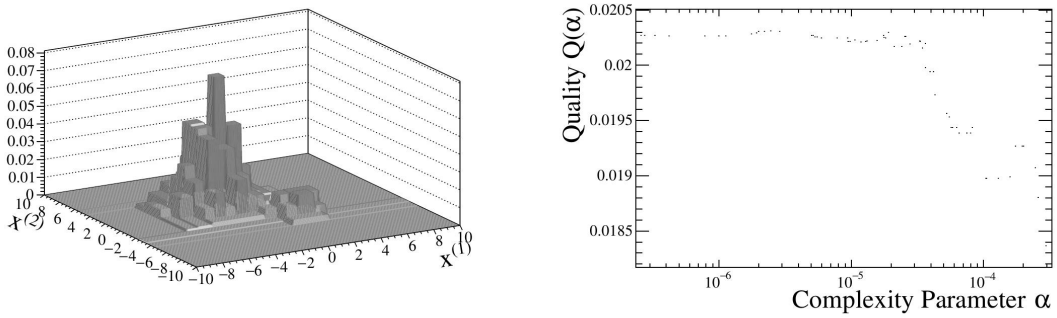

Figure A.9: Pruning obtained using a tuned stop criterion (no node with less than 3 entries, no bin-width below 0.8) and a self-optimization using the number of leaves as complexity function.

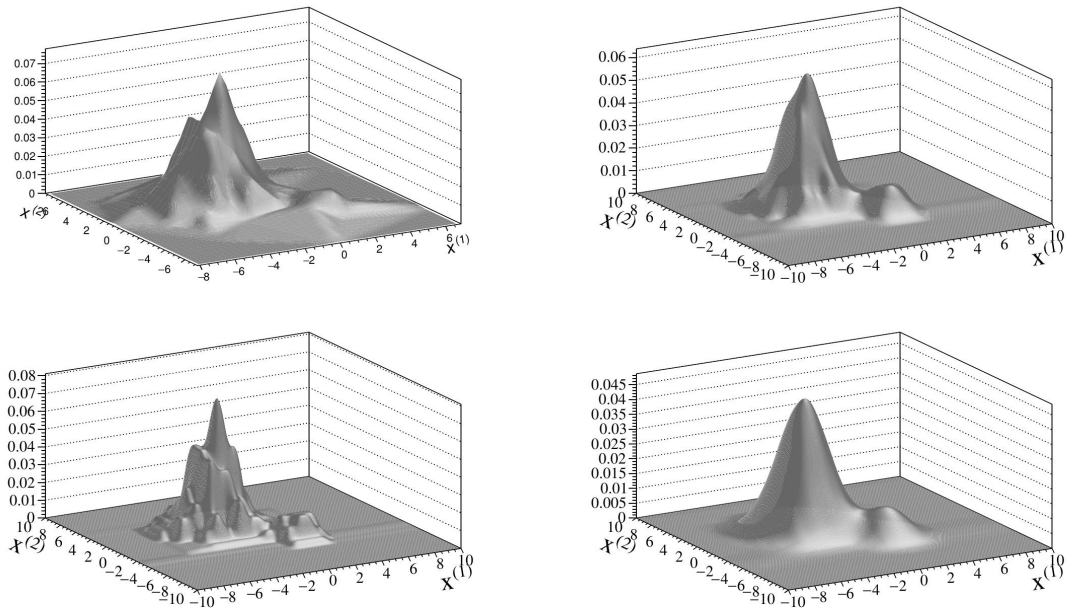

Figure A.10: Smoothed version of the density estimation discussed in Figure A.9. At top-left the Delaunay triangulation is used, in the other figures the smoothing is performed with a triangular-weighted average. The bandwidth of the weight function is $h = 1.0$ (top-right), $h = 0.5$ (bottom-left), and $h = 2.0$ (bottom-right).

Figure A.11: Effect of the choice of the bandwidth when smoothing through a triangular-weight average. The density estimation tree is the same as in Figures A.10 and A.9, but at left bandwidth $h_{\text{left}} = 0.5$ is used, while at right, $h_{\text{right}} = 2.0$.

A.3.2 Smoothing

As discussed in Section A.1.4, smoothing is useful to remove the border effects and bin boundaries. Two techniques have been considered: Delaunay triangulation and triangular smearing. The former can only be applied in two dimensions and is very sensitive to overtraining because the value of the *pdf* assumed in the middle of each bin is reproduced as is in the smoothed *pdf*. This limits the effects of resolution loss where many bins are present, but does not contribute to cure overtrained trees. Oppositely, triangular smearing, based on a weighted average of the *pdf*, can reduce effects of the overtraining by averaging fast oscillations. The density estimations of Figure A.9 is smoothed with the Delaunay triangulation and with triangular smearing with different bandwidths and shown in Figure A.10. Note how the choice of the bandwidth can significantly change the shape of the smoothed density estimation.

Triangular smearing can be used to cure overtraining, at the expenses of the loss of dynamic adjustment of the resolution, which is one of the most interesting features of the density estimation trees. Figure A.12 shows an overtrained density estimation and its

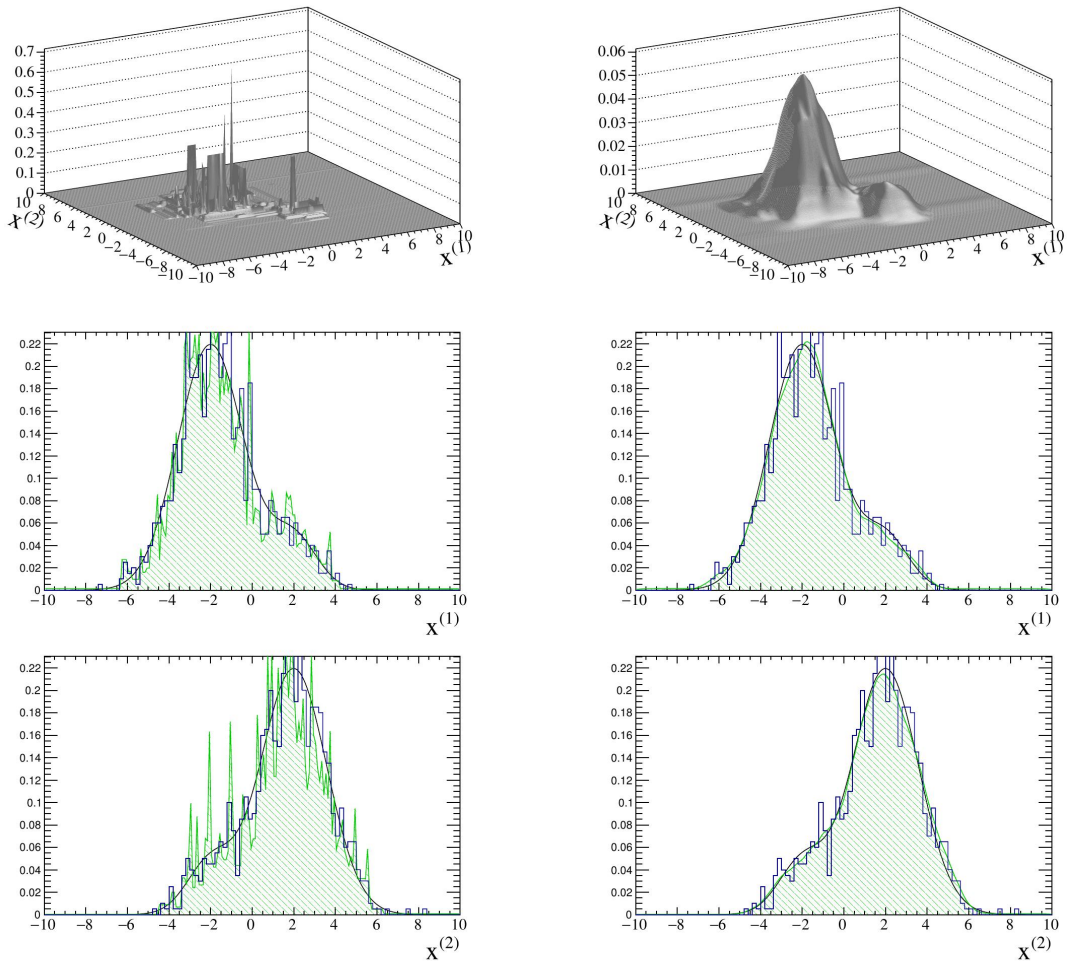

Figure A.12: Effect of the triangular smearing on an overtrained density estimation tree. The left column shows the density estimation obtained reading the overtrained density estimation tree, the right column shows the smoothed version of the *pdf*. The first row represent the three-dimensional views, while the projections are reported in the second and in the third rows. The black smooth curve represents the generating *pdf*, the histogram of the events is represented as a blue curve while the density estimation based on a decision tree is shown as a dash-filled green line.

smoothed version. The projections of the two distributions onto the $x^{(1)}$ variable are also shown.

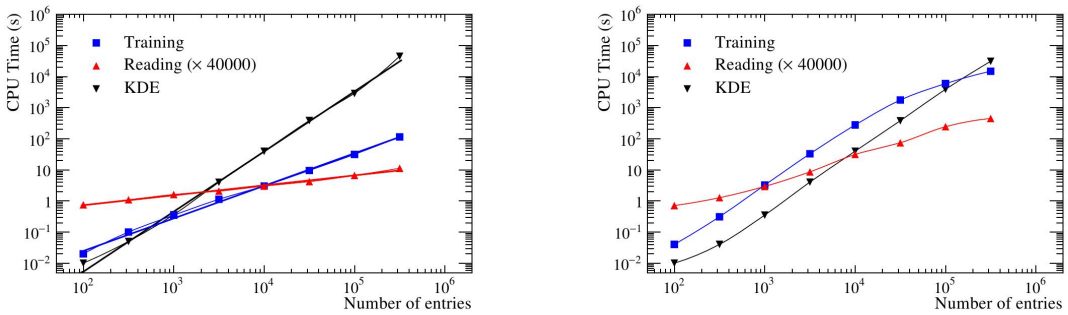

Figure A.13: CPU time to train and evaluate a self-optimized decision tree as a function of the number of entries N_{tot}. At left a stop criterion including a reasonable bin-width threshold (0.3) is used, at right it is replaced with a very loose threshold (0.03). The time required to train a RooFit Kernel Density Estimator (KDE) is also reported for comparison. The points in the left plot are fitted with power laws and the fit results are reported in Table A.2.

A.4 Timing

As already mentioned, one of the advantages of density estimation trees is the speed of their training and evaluation. The complexity of the algorithm is $N_{\text{bin}} \times N_{\text{tot}}$. For a well behaved tree, trained on a large number of events, N_{bin} is roughly independent of N_{tot} and therefore the complexity of the algorithm is linear in N_{tot}. Though, if the stop condition is very loose it may happen that the decision tree tries to produce a bin per event, or per few events. If no bin-width condition is used, and the growth of the tree is stopped when the bin contains n_{th} entries, or less, then

$$N_{\text{bin}} \sim \frac{N_{\text{tot}}}{n_{\text{th}}}, \tag{A.52}$$

and the complexity of the algorithm becomes quadratic in N_{tot}, becoming comparable to kernel density estimation.

Figure A.13 compare the CPU time required to train, optimize, and sample on a 200×200 grid, a smoothed density estimation with the time required to build a using the widely used Roofit package [164]. The comparison is repeated twice, for a stop condition including a reasonable (0.3) and a very loose (0.03) bin-width threshold. For the reasonable criterion the dependence of the CPU time on N_{tot} is consistent with a linear dependence, while for the loose criterion it is consistent with a quadratic dependence for small N_{tot} and reaches a linear behaviour for very large samples. The CPU time t_{CPU} needed to train and read the density estimation tree with reasonable bin-width threshold, and to train the Kernel Density Estimation, are fitted with a power law

$$t_{\text{CPU}} = a(N_{\text{tot}})^b, \tag{A.53}$$

and the parameters a and b reported in Table A.2.

$t_{\mathrm{CPU}} = a(N_{\mathrm{tot}})^b$	a	b
Training KDE	$(0.73 \pm 0.11)\mu$s	1.931 ± 0.017
Training DET	(0.20 ± 0.2) ms	1.046 ± 0.012
Reading DET ($\times$40000)	(0.17 ± 0.2) s	0.320 ± 0.014

Table A.2: Parameters of the power laws describing the dependence of the CPU time on the sample size N_{tot}. For the kernel density estimation the relation is nearly quadratic, for the training of the density estimation tree it is nearly linear, reading the density estimation tree is expected to be proportional to N_{bin}, which with the chosen stop condition is found to be proportional to $(N_{\mathrm{tot}})^{0.3}$. Uncertainties are statistical only.

A.5 Conclusions

I discussed Density Estimation Trees, a class of algorithms to perform nonparametric density estimation using binary decision trees. Density Estimation Trees present interesting features for exploratory data analysis being fast and robust algorithms. An original cross-validation technique based on triangular kernel estimation has been discussed and it has been shown that it can lead to self-optimization of the stopping rule of the tree growth with computational complexity much reduced with respect to the most common methods, opening the way to the application of self-optimization techniques to very large samples with $\mathcal{O}(10^6)$ or more entries. The algorithm has been implemented and tested on a simple but realistic distribution. Smoothing techniques are important to reduce the issues introduced from unphysical discontinuities in the density estimation. Two techniques have been discussed: linear interpolation based on Delaunay triangulation, and triangular smearing. Both the techniques have a reasonable computational cost which depends linearly on the number of bins, but the former can be applied only in two dimensions, while the second may introduce a loss of resolution in applications where the adaptability of the density estimation tree is important. Variable correlation negatively affects the performance of the algorithm. A stage of linear decorrelation could precede the training of the decision tree to improve the accuracy of the result.

Acknowledgements

During these three years of Ph.D. I have learned a lot, about physics, about coding, and about the scientific community and its moods, but the most important lesson is certainly the importance of people walking down that pathway with you. Listening to collaborators and colleagues is seeding ideas which will grow and benefit your work, while sharing your work and your point to an attentive eye is testing the robustness of the basements you are building on. At the same time, family and friends are the team it is worth to play for, for which you can spend that one more night on your laptop, or keep pushing for your results to be considered.

My supervisor, Giacomo Graziani, has been at the same time colleague, collaborator, friend and in some cases I would say even family, when collecting my frustration or sharing some success. I am very grateful to him for his patience, constant availability and care in the analysis of my own work, and for the practical contributions he offered in the development of the analysis presented in this Thesis, and of the other activities still private to the LHCb Collaboration.

A big thank is also due to Michele Veltri, active member of the analysis committee working on the B_c^+ lifetime measurement. We developed together the automated infrastructure to produce and test the large simulated samples used in the analysis, with the several dynamical models he retrieved in the literature with a remarkable bibliographic effort.

Giovanni Passaleva, our team leader, has always made of his best to ease my scientific activities. I would like to thank him, and the whole Istituto Nazionale di Fisica Nucleare, for the year I have been allowed to spend at CERN as INFN-Associate, an invaluable benefit to the development of my professional skills and scientific achievements, and a great personal experience.

Andrea Bizzeti, professor in Modena, has offered me the opportunity of sitting for a while on the other side of the desk, co-tutoring the Master Thesis of Antonio Spagnolo, an enterprising student with enough fantasy and technical culture to propose, implement and develop innovative data analysis techniques.

Maddalena Frosini, young post-doc in Florence, has been my office mate for about two years, including six months of overlap in Geneva. She is today among my best friends, but I cannot avoid to acknowledge the great technical support she offered in the first year of my Ph.D. when I was facing for the first time the huge complexity of the LHCb Collaboration software. In that period my office-mate was Antonio Cassese. That is interesting because he is today my flat-mate and host. I do not mean to spend the hundred pages I would need to list the reasons for which I am grateful to Antonio, he welcomed me the first day in Florence, shared with me a few nights in Geneva (could I forget how he got me to discover the *penne alla siciliana à la Meyrinose*?) and he is still of major support in these last few weeks of Ph.D. for both my academic and, more importantly, personal life.

To the conveners of my working group, Giulia Manca, Vanya Belyaev and Marco Pappagallo, and to internal referees of the analysis presented in this Thesis, Fatima Soomro and Sheldon Stone, I acknowledge a careful and attentive reading of the internal

documentation, whose quality — I ought to confess — was sometime suboptimal. They have all pushed to improve the reliability of the final result, proposing clever cross-checks and suggesting to explore new techniques. I am glad of having worked with them and proud of having finally managed to get their expert and wise consensus to the publication of the result.

Tim Gershon has been physics coordinator of the LHCb Collaboration for the first two years of my Doctorate. I consider Tim as one of the most brilliant experimental physicists in the field of High Energies and I am honored of the support he has never denied to my candidatures to cover both LHCb and non-LHCb roles.

Mario Calvetti and Sandra Malvezzi have been assigned as referees to this Thesis by the Doctoral School. Their comments to this document have been essential to its development and finalization.

I would like to thank also a few friends of the LHCb Collaboration I met in Geneva or at the LHCb Weeks around Europe. In particular, Stefano Perazzini, Vincenzo Vagnoni and Maria Zangoli, from Bologna, Flavio Archilli, Barbara Sciascia and Ricardo Vazquez Gomez from LNF, Andrea Contu and Alessandro Cardini from Cagliari, and Francesca Dordei working in Heidelberg, are to my memory the best of the LHCb Collaboration. A real team of people always ready to answer other's questions and keeping offering their help to overcome the many technical issues.

Under a more personal perspective, I thank my family and I apologize for the rarity of my phone or Skype calls in the rush periods (ahem...). For a while they have been trying to call, but the third time I answered from my office at 9 pm, on Sunday, they capitulated, and still today my mother watches my Skype status because "as long as it is green, she knows I am fine".

Livio Bianchi, Stefania Bufalino, Cristina Ferro, Sandro Gonzi, Andrea Beraudo, Roberta Volpe, Elisa Prandini and Andrea Rossi, have filled the tough year we spent together in Geneva with funny and pleasant memories. Caterina, daughter of Andrea and Elisa was born when I started the B_c^+ lifetime measurement. When my analysis was approved, Caterina was starting walking. When it got published, she was starting talking. In practice, in the time I produced a paper, Elisa and Andrea *produced* a beautiful speaking baby. That resizes Science.

Back in Florence, for the third year of Ph.D., I found an old network of people, among others, Eduardo Grossi, Davide Nuzzi, Barbara Olmi, Federico Tommasi and Lucia Pettinato. We shared the day-by-day life in the Physics department as the uncertainties about the future, both ours and for public research.

Valentina Gori, colleague and friend both in Florence and in Geneva, deserves a special thank for having introduced me to the *Laboratorio Teatrale Ludico*, an acting company which welcomed me and offered the opportunity to restart practicing a passion I dropped when moving to Florence three years ago. Maria Carla Lallo, Federico Galeotti, Lorenzo Gheri, Valeria Zoni, Erica Nieri, Matteo Tusa and Stefania Bartoli, members of the group, are the few non-physics-speaking people I met this year and I can today call friends.

Another member of the group, Martina Gasperi, is actually more than a friend. The reader won't mind if I prefer thanking her privately, but she should be acknowledged at

least for the patience to tolerate, smiling, my being late and in a bad mood, when dealing with HEP software, which, as I told her, never gives the right answer before 8 pm.

By design.

I finally thank Niccheri, even if he will hardly ever read these words.

Bibliography

[1] Particle Data Group, K. A. Olive *et al.*, *Review of Particle Physics (RPP)*, Chin. Phys. **C38** (2014) 090001.

[2] BELLE Collaboration, S. Choi *et al.*, *Observation of a resonance-like structure in the $\pi^{\pm}$ psi-prime mass distribution in exclusive $B \to K\,\pi^{\pm}$ psi-prime decays*, Phys. Rev. Lett. **100** (2008) 142001, arXiv:0708.1790.

[3] LHCb collaboration, R. Aaij *et al.*, *Observation of the resonant character of the $Z(4430)^{-}$ state*, Phys. Rev. Lett. **112** (2014) 222002, arXiv:1404.1903.

[4] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, *Evidence for the 2 pi Decay of the k(2)0 Meson*, Phys. Rev. Lett. **13** (1964) 138.

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

[6] C. Wu *et al.*, *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (1957) 1413.

[7] BELLE Collaboration, A. Abashian *et al.*, *Measurement of the CP violation parameter $\sin 2\phi_1$ in B^0_d meson decays*, Phys. Rev. Lett. **86** (2001) 2509, arXiv:hep-ex/0102018.

[8] BaBar Collaboration, B. Aubert *et al.*, *Measurement of CP violating asymmetries in B^0 decays to CP eigenstates*, Phys. Rev. Lett. **86** (2001) 2515, arXiv:hep-ex/0102030.

[9] SLAC-SP-017 Collaboration, J. Augustin *et al.*, *Discovery of a Narrow Resonance in e+ e- Annihilation*, Phys. Rev. Lett. **33** (1974) 1406.

[10] E598 Collaboration, J. Aubert *et al.*, *Experimental Observation of a Heavy Particle J*, Phys. Rev. Lett. **33** (1974) 1404.

[11] S. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285.

[12] S. Herb *et al.*, *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39** (1977) 252.

[13] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[14] R. Giachetti and E. Sorace, *Unified treatment of hyperfine splittings in $b\bar{b}$, $c\bar{c}$ and $s\bar{s}$*, Phys. Rev. **D87** (2013), no. 3 034021, arXiv:1207.3301.

[15] Quarkonium Working Group, N. Brambilla *et al.*, *Heavy quarkonium physics*, arXiv:hep-ph/0412158.

[16] G. T. Bodwin *et al.*, *Quarkonium at the Frontiers of High Energy Physics: A Snowmass White Paper*, arXiv:1307.7425.

[17] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, CERN-TH-412, NP-14146, PRINT-64-170, 1964.

[18] ATLAS Collaboration, G. Aad *et al.*, *Observation of an Excited $B_c^{\pm}$ Meson State with the ATLAS Detector*, arXiv:1407.1032.

[19] C.-H. Chang and Y.-Q. Chen, *The hadronic production of the B(c) meson at Tevatron, CERN LHC and SSC*, Phys. Rev. **D48** (1993) 4086.

[20] E. Gregory *et al.*, *A Prediction of the B*(c) mass in full lattice QCD*, Phys. Rev. Lett. **104** (2010) 022001, arXiv:0909.4462.

[21] A. Berezhnoy, A. Likhoded, and M. Shevlyagin, *Hadronic production of B(c) mesons*, Phys. Atom. Nucl. **58** (1995) 672, arXiv:hep-ph/9408284.

[22] A. Berezhnoy, A. Likhoded, and O. Yushenko, *Hadronic production of B(c) mesons*, Phys. Atom. Nucl. **59** (1996) 742.

[23] C.-H. Chang, Y.-Q. Chen, G.-P. Han, and H.-T. Jiang, *On hadronic production of the B(c) meson*, Phys. Lett. **B364** (1995) 78, arXiv:hep-ph/9408242.

[24] K. Kolodziej, A. Leike, and R. Ruckl, *Production of B(c) mesons in hadronic collisions*, Phys. Lett. **B355** (1995) 337, arXiv:hep-ph/9505298.

[25] A. Berezhnoy, V. Kiselev, and A. Likhoded, *Hadronic production of S and P wave states of anti-b c quarkonium*, Z. Phys. **A356** (1996) 79, arXiv:hep-ph/9602347.

[26] B. R. Martin and G. Shaw, *Particle physics*, ISBN-9780470032947, 2008.

[27] R. Placakyte, *Parton Distribution Functions*, arXiv:1111.5452.

[28] C.-H. Chang, C. Driouichi, P. Eerola, and X. G. Wu, *BCVEGPY: an event generator for hadronic production of the B_c meson*, Comput. Phys. Commun. **159** (2004) 192, arXiv:hep-ph/0309120.

[29] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[30] I. P. Gouz *et al.*, *Prospects for the B_c studies at LHCb*, Phys. Atom. Nucl. **67** (2004) 1559, arXiv:hep-ph/0211432.

[31] M. Shifman, *Shifman-Vainshtein-Zakharov sum rules*, Scholarpedia **8** (2013), no. 11 8790, revision 137127.

[32] D. Ebert, R. N. Faustov, and V. O. Galkin, *Weak decays of the B_c meson to charmonium and D mesons in the relativistic quark model*, Phys. Rev. **D68** (2003) 094020, arXiv:hep-ph/0306306.

[33] V. V. Kiselev, *Exclusive decays and lifetime of B_c meson in QCD sum rules*, arXiv:hep-ph/0211021.

[34] D. Scora and N. Isgur, *Semileptonic meson decays in the quark model: An update*, Phys. Rev. **D52** (1995) 2783, arXiv:hep-ph/9503486.

[35] N. Isgur, D. Scora, B. Grinstein, and M. B. Wise, *Semileptonic B and D Decays in the Quark Model*, Phys. Rev. **D39** (1989) 799.

[36] L. Anderlini, *Preliminary study for a measurement of the CKM angle γ using the tree $B^{\pm} \to D^{(*)0}K^{\pm}$ decays with the LHCb detector at CERN*, Presented 23 Sep 2011, Modena University, Advisor: A. Bizzeti.

[37] V. Kiselev, *Gold plated mode of CP violation in decays of B_c meson from QCD sum rules*, J. Phys. **G30** (2004) 1445, arXiv:hep-ph/0302241.

[38] LHCb collaboration, I. Bediaga *et al.*, *Framework TDR for the LHCb Upgrade: Technical Design Report*, Tech. Rep. CERN-LHCC-2012-007. LHCb-TDR-12, CERN, Geneva, Apr, 2012.

[39] G. Buchalla, *Heavy quark theory*, arXiv:hep-ph/0202092.

[40] M. A. Shifman, A. Vainshtein, and V. I. Zakharov, *QCD and Resonance Physics. Sum Rules*, Nucl. Phys. **B147** (1979) 385.

[41] N. Uraltsev, *Topics in the heavy quark expansion*, arXiv:hep-ph/0010328.

[42] A. Lenz, *Lifetimes and HQE*, arXiv:1405.3601.

[43] S. Stone, *Lifetimes of some b-flavored hadrons*, arXiv:1406.6497.

[44] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the Λ_b^0 baryon lifetime*, Phys. Rev. Lett. **111** (2013) 102003, arXiv:1307.2476.

[45] 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.

[46] ATLAS Collaboration, G. Aad *et al.*, *Measurement of the Λ_b lifetime and mass in the ATLAS experiment*, Phys. Rev. **D87** (2013), no. 3 032002, arXiv:1207.2284.

[47] L. Anderlini, *Masses and lifetimes at LHCb*, Journal of Physics: Conference Series (2014), Proceedings of the XI International Conference on hyperons, charm and beauty hadrons.

[48] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the mass and lifetime of the Ξ_b^- baryon*, arXiv:1409.8568.

[49] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the $\bar{B}_s^0$ meson lifetime in $D_s^+\pi^-$ decays*, arXiv:1407.5873.

[50] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the mass and lifetime of the Ξ_b^0 baryon*, Phys. Rev. Lett. **113** (2014) 032001, arXiv:1405.7223.

[51] M. Lusignoli and M. Masetti, *B_c decays*, Z. Phys. **C51** (1991) 549.

[52] E. J. Eichten and C. Quigg, *Mesons with beauty and charm: Spectroscopy*, Phys. Rev. **D49** (1994) 5845, arXiv:hep-ph/9402210.

[53] I. I. Bigi, *Inclusive B_c decays as a QCD lab*, Phys. Lett. **B371** (1996) 105, arXiv:hep-ph/9510325.

[54] M. Beneke and G. Buchalla, *B_c meson lifetime*, Phys. Rev. **D53** (1996) 4991, arXiv:hep-ph/9601249.

[55] A. I. Onishchenko, *Doubly heavy systems: Decays and OPE*, arXiv:hep-ph/9912424.

[56] V. V. Kiselev, A. E. Kovalsky, and A. K. Likhoded, *B_c decays and lifetime in QCD sum rules*, Nucl. Phys. **B585** (2000) 353, arXiv:hep-ph/0002127.

[57] A. Y. Anisimov, I. M. Narodetsky, C. Semay, and B. Silvestre-Brac, *The B_c meson lifetime in the light-front constituent quark model*, Phys. Lett. **B452** (1999) 129, arXiv:hep-ph/9812514.

[58] C.-H. Chang, S.-L. Chen, T.-F. Feng, and X.-Q. Li, *Lifetime of the B_c meson and some relevant problems*, Phys. Rev. **D64** (2001) 014003, arXiv:hep-ph/0007162.

[59] A. K. Rai and N. Devlani, *Mass Spectrum and decay properties of the B_c meson*, PoS **Hadron2013** (2013) 045.

[60] OPAL Collaboration, K. Ackerstaff *et al.*, *Search for the B_c meson in hadronic Z^0 decays*, Phys. Lett. **B420** (1998) 157, arXiv:hep-ex/9801026.

[61] CDF collaboration, F. Abe *et al.*, *Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. Lett. **81** (1998) 2432, arXiv:hep-ex/9805034.

[62] CDF collaboration, T. Aaltonen *et al.*, *Observation of the Decay B^+ -(c) $\to J/\psi\pi^\pm$ and Measurement of the B^+ -(c) Mass*, Phys. Rev. Lett. **100** (2008) 182002, arXiv:0712.1506.

[63] 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.

[64] LHCb Collaboration, R. Aaij *et al.*, *Measurements of B_c^+ production and mass with the $B_c^+ \to J/\psi\pi^+$ decay*, Phys. Rev. Lett. **109** (2012) 232001, arXiv:1209.5634.

[65] LHCb collaboration, R. Aaij *et al.*, *Observation of $B_c^+ \to J/\psi D_s^+$ and $B_c^+ \to J/\psi D_s^{*+}$ decays*, Phys. Rev. **D87** (2013), no. 11 112012, arXiv:1304.4530.

[66] LHCb collaboration, R. Aaij *et al.*, *First observation of a baryonic B_c^+ decay*, arXiv:1408.0971.

[67] CMS Collaboration, V. Khachatryan *et al.*, *Measurement of the ratio $\mathcal{B}(\mathrm{B_c^+} \to \mathrm{J}/\psi\,\pi^+\pi^+\pi^-)/\mathcal{B}(\mathrm{B_c^+} \to \mathrm{J}/\psi\,\pi^+)$ and the production cross sections times branching fractions of $\mathrm{B_c^+} \to \mathrm{J}/\psi\,\pi^+$ and $\mathrm{B^+} \to \mathrm{J}/\psi\,\mathrm{K^+}$ in pp collisions at $\sqrt{s} = 7$ TeV*, arXiv:1410.5729.

[68] K. Toms, *Invited talk at the b-hadron and quarkonia lhcb meeting: Observation of the $B_c(2S)$ state at atlas*, indico.cern.ch/event/270004, July, 2014.

[69] LHCb collaboration, R. Aaij *et al.*, *Observation of the Decay $B_c^+ \to B^0\pi^+$*, Phys. Rev. Lett. **111** (2013), no. 18 181801, arXiv:1308.4544.

[70] LHCb Collaboration, R. Aaij *et al.*, *First observation of the decay $B_c^+ \to J/\psi\pi^+\pi^-\pi^+$*, Phys. Rev. Lett. **108** (2012) 251802, arXiv:1204.0079.

[71] A. Likhoded and A. Luchinsky, *Light hadron production in $B_c^+ \to J/\psi + X$ decays*, Phys. Rev. **D81** (2010) 014015, arXiv:0910.3089.

[72] LHCb collaboration, R. Aaij *et al.*, *Evidence for the decay $B_c^+ \to J/\psi 3\pi^+ 2\pi^-$*, JHEP **1405** (2014) 148, arXiv:1404.0287.

[73] A. Luchinsky, *Production of charged π-mesons in exclusive $B_c \to V(P) + n\pi$ decays*, Phys. Rev. **D86** (2012) 074024, arXiv:1208.1398.

[74] LHCb, R. Aaij *et al.*, *Observation of the decay $B_c \to J/\psi K^+K^-\pi^+$*, JHEP **1311** (2013) 094, arXiv:1309.0587.

[75] A. Luchinsky, *Production of K mesons in exclusive B_c decays*, arXiv:1307.0953.

[76] LHCb, R. Aaij *et al.*, *First observation of the decay $B_c^+ \to J/\psi K^+$*, JHEP **1309** (2013) 075, arXiv:1306.6723.

[77] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the ratio of B_c^+ branching fractions to $J/\psi\pi^+$ and $J/\psi\mu^+\nu_\mu$*, Phys. Rev. **D90** (2014) 032009, arXiv:1407.2126.

[78] CDF collaboration, T. Aaltonen *et al.*, *Measurement of the B_c^- meson lifetime in the decay $B_c^- \to J/\psi\ \pi^-$*, Phys. Rev. **D87** (2013) 011101, arXiv:1210.2366.

[79] H. Song, *Measurement of the $B_c^\pm$ meson lifetime using $B_c^\pm \to J/\psi\ \pi^\pm$ decays*, FERMILAB-THESIS-2013-05.

[80] Colombo, Simone, *Study of the $B_c^+ \to J/\psi\pi^+$ and $B_c^+ \to J/\psi 3\pi$ decay channels in the CMS experiment at LHC*, INFN THESIS 6987, 2011.

[81] D0 collaboration, V. M. Abazov *et al.*, *Measurement of the lifetime of the $B_c^\pm$ meson in the semileptonic decay channel*, Phys. Rev. Lett. **102** (2009) 092001, arXiv:0805.2614.

[82] L. C. Welty-Rieger, *The Lifetime of a beautiful and charming meson: B_c lifetime measured using the D0 detector*, FERMILAB-THESIS-2008-66, 2008.

[83] CDF Collaboration, A. Abulencia *et al.*, *Measurement of the B_c^+ Lifetime in $B_c^+ \to J/\psi + l + X$ Decays*, http://goo.gl/9Kkt5M.

[84] CDF collaboration, A. Abulencia *et al.*, *Measurement of the B_c^+ meson lifetime using the decay mode $B_c^+ \to J/\psi\, e^+\nu_e$*, Phys. Rev. Lett. **97** (2006) 012002, arXiv:hep-ex/0603027.

[85] 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.

[86] LHCb Collaboration, B. Adeva *et al.*, *Roadmap for selected key measurements of LHCb*, arXiv:0912.4179.

[87] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CKM angle γ from a combination of $B^\pm \to Dh^\pm$ analyses*, Phys. Lett. **B726** (2013) 151, arXiv:1305.2050.

[88] LHCb collaboration, R. Aaij *et al.*, *Determination of γ and $-2\beta_s$ from charmless two-body decays of beauty mesons*, arXiv:1408.4368.

[89] LHCb Collaboration, R. Aaij *et al.*, *First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. **110** (2013) 021801, arXiv:1211.2674.

[90] LHCb Collaboration, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Sep, 2014. Submitted to Nature.

[91] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetries in the decays $B^0 \to K^{*0}\mu^+\mu^-$ and $B^+ \to K^+\mu^+\mu^-$*, arXiv:1408.0978.

[92] LHCb Collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay* $B^0 \to K^{*0}\mu^+\mu^-$, JHEP **1308** (2013) 131, arXiv:1304.6325.

[93] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation and the* B^0_s *meson decay width difference with* $B^0_s \to J/\psi K^+K^-$ *and* $B^0_s \to J/\psi pi^+pi^-$ *decays*, Phys. Rev. **D87** (2013), no. 11 112010, arXiv:1304.2600.

[94] LHCb Collaboration, R. Aaij *et al.*, *Observation of photon polarization in the* $b \to s\gamma$ *transition*, Phys. Rev. Lett. **112** (2014) 161801, arXiv:1402.6852.

[95] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of D meson mass differences*, JHEP **1306** (2013) 065, arXiv:1304.6865.

[96] LHCb collaboration, R. Aaij *et al.*, *Measurement of D0-D0bar mixing parameters and search for CP violation using* $D0 \to K^+\pi^-$ *decays*, Phys. Rev. Lett. **111** (2013) 251801, arXiv:1309.6534.

[97] LHCb Collaboration, R. Aaij *et al.*, *Observation of* $D^0 - \overline{D}^0$ *oscillations*, Phys. Rev. Lett. **110** (2013), no. 10 101802, arXiv:1211.1230.

[98] LHCb Collaboration, R. Aaij *et al.*, *Study of* D_{sJ} *decays to* $D^+K^0_S$ *and* D^0K^+ *final states in pp collisions*, JHEP **1210** (2012) 151, arXiv:1207.6016.

[99] LHCb Collaboration, R. Aaij *et al.*, *Observation of* J/ψ *pair production in pp collisions at* $\sqrt{s} = 7TeV$, Phys. Lett. **B707** (2012) 52, arXiv:1109.0963.

[100] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the* $B^\pm$ *production cross-section in pp collisions at* $\sqrt{s} = 7$ *TeV*, JHEP **1204** (2012) 093, arXiv:1202.4812.

[101] LHCb Collaboration, R. Aaij *et al.*, *Measurement of Upsilon production in pp collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. **C72** (2012) 2025, arXiv:1202.6579.

[102] LHCb Collaboration, R. Aaij *et al.*, *Measurement of* $\psi(2S)$ *meson production in pp collisions at sqrt(s)=7 TeV*, Eur. Phys. J. **C72** (2012) 2100, arXiv:1204.1258.

[103] LHCb collaboration, R. Aaij *et al.*, *Measurement of* Υ *production in pp collisions at* $\sqrt{s} = 2.76$ *TeV*, Eur. Phys. J. **C74** (2014) 2835, arXiv:1402.2539.

[104] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the relative rate of prompt* χ_{c0}, χ_{c1} *and* χ_{c2} *production at* $\sqrt{s} = 7TeV$, JHEP **1310** (2013) 115, arXiv:1307.4285.

[105] LHCb collaboration, R. Aaij *et al.*, *Production of* J/ψ *and* Υ *mesons in pp collisions at* $\sqrt{s} = 8$ *TeV*, JHEP **1306** (2013) 064, arXiv:1304.6977.

[106] LHCb collaboration, R. Aaij *et al.*, *Measurement of* J/ψ *production in pp collisions at* $\sqrt{s} = 2.76$ *TeV*, JHEP **1302** (2013) 041, arXiv:1212.1045.

[107] LHCb collaboration, R. Aaij *et al.*, *Study of χ_b meson production in pp collisions at $\sqrt{s} = 7$ and 8TeV and observation of the decay $\chi_b(3P) \to \Upsilon(3S)\gamma$*, arXiv:1407.7734.

[108] BELLE Collaboration, R. Mizuk *et al.*, *Dalitz analysis of $B \to K$ pi+ psi-prime decays and the Z(4430)+*, Phys. Rev. **D80** (2009) 031104, arXiv:0905.2869.

[109] BaBar Collaboration, B. Aubert *et al.*, *Search for the Z(4430)- at BABAR*, Phys. Rev. **D79** (2009) 112001, arXiv:0811.0564.

[110] LHCb Collaboration, R. Aaij *et al.*, *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, JHEP **1206** (2012) 058, arXiv:1204.1620.

[111] LHCb collaboration, R. Aaij *et al.*, *Limits on neutral Higgs boson production in the forward region in pp collisions at $\sqrt{s} = 7$ TeV*, JHEP **1305** (2013) 132, arXiv:1304.2591.

[112] LHCb Collaboration, R. Aaij *et al.*, *Observation of Z production in proton-lead collisions at LHCb*, JHEP **1409** (2014) 030, arXiv:1406.2885.

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

[114] LHCb Collaboration, *LHCb detector performance*, To be submitted to IJMPA (2014), In preparation.

[115] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007, arXiv:1405.7808.

[116] H. Bichel, D. E. Groom, and S. R. Klein, *Passage of particles through matter*, Particle Data Group Review **32** (2013).

[117] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[118] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, J. Instrum. **8** (2012) P04022.

[119] LHCb HLT project, J. Albrecht, V. Gligorov, G. Raven, and S. Tolk, *Performance of the LHCb High Level Trigger in 2012*, J. Phys. Conf. Ser. **513** (2014) 012001, arXiv:1310.8544.

[120] N. Brook, *LHCb Computing Model*, Tech. Rep. LHCb-2004-119. CERN-LHCb-2004-119, CERN, Geneva, Dec, 2004.

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

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

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

[124] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[125] M. Frank *et al.*, *Deferred High Level Trigger in LHCb: A Boost to CPU Resource Utilization*, J. Phys. Conf. Ser. **513** (2014) 012006.

[126] Y. Xie, I. Belyaev, and Y. Amhis, *Vertex Fitters*, LHCb tWiki pages, 2010.

[127] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME–Journal of Basic Engineering **82** (1960), no. Series D 35.

[128] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nuclear Instruments and Methods in Physics Research A **552** (2005) 566, `arXiv:physics/0503191`.

[129] G. Raven, *Selection of $B_s \to J/\psi\phi$ and $B^+ \to J/\psi K^+$*, LHCb-2003-118 (2003).

[130] M. L. Bote-Lorenzo, Y. A. Dimitriadis, and E. Gómez-Sánchez, *Grid characteristics and uses: a grid definition*, in *Grid Computing*, pp. 291–298. Springer Berlin Heidelberg, 2004.

[131] *LHC Computing Grid Web Site*, http://wlcg.web.cern.ch.

[132] LHCb DIRAC Collaboration, A. Casajus, R. Graciani, S. Paterson, and A. Tsaregorodtsev, *DIRAC pilot framework and the DIRAC workload management system*, J. Phys. Conf. Ser. **219** (2010) 062049.

[133] Z. Mathe, *Feicim: A browser and analysis tool for distributed data in particle physics*, CERN-THESIS-2012-156.

[134] F. Brochu *et al.*, *Ganga: a tool for computational-task management and easy access to grid resources*, CoRR **abs/0902.2685** (2009).

[135] T. Gershon and M. Needham, *Heavy Flavour Physics at the LHC*, `arXiv:1408.0403`.

[136] Y. Li, *Heavy flavour spectroscopy at LHC*, `arXiv:1409.4020`.

[137] LHCb Collaboration, P. d. Simone, *b and c spectroscopy at LHCb*, EPJ Web Conf. **73** (2014) 03007.

[138] LHCb Collaboration, I. Polyakov, *b and c hadron spectroscopy at LHCb*, `arXiv:1404.7613`.

[139] L. Anderlini, *Properties and Decays of the B_c^+ meson*, `arXiv:1407.8066`.

[140] L. Anderlini, *Properties and decays of the B_c^+ meson*, PoS(DIS2014)177 (2014).

[141] LHCb collaboration, R. Aaij *et al.*, *Measurement of the B_c^+ meson lifetime using $B_c^+ \to J/\psi\mu^+\nu_\mu X$ decays*, Eur. Phys. J. **C74** (2014) 2839, arXiv:1401.6932.

[142] CDF Collaboration, T. Aaltonen *et al.*, *Measurement of the B^- lifetime using a simulation free approach for trigger bias correction*, Phys. Rev. **D83** (2011) 032008, arXiv:1004.4855.

[143] R. Bailey *et al.*, *Measurement of the lifetime of charged and neutral D mesons with high resolution silicon strip detectors*, Z. Phys. **C28** (1985) 357.

[144] T. Mazzoni, *Identificazione di muoni nell'esperimento LHCb mediante ricostruzione di tracce con reti neurali ricorsive*, Tesi INFN 7477, 2012.

[145] Z.-h. Wang, G.-L. Wang, and C.-H. Chang, *The B_c decays to a P-wave charmonium by the improved Bethe-Salpeter approach*, J. Phys. **G39** (2012) 015009, arXiv:1107.0474.

[146] Y.-P. Kuang, S. F. Tuan, and T.-M. Yan, *Hadronic transitions and $^1P_1$ states of heavy quarkonia*, Phys. Rev. **D37** (1988) 1210.

[147] Fermilab E835 collaboration, M. Andreotti *et al.*, *Results of a search for the $h_c(^1P_1)$ state of charmonium in the $\eta_c\gamma$ and $J/\psi\pi^0$ decay modes*, Phys. Rev. **D72** (2005) 032001.

[148] D. Ebert, R. N. Faustov, and V. O. Galkin, *Semileptonic and nonleptonic decays of B_c mesons to orbitally excited heavy mesons in the relativistic quark model*, Phys. Rev. **D82** (2010) 034019, arXiv:1007.1369.

[149] V. V. Kiselev and A. V. Tkabladze, *Semileptonic B_c decays from QCD sum rules*, Phys. Rev. **D48** (1993) 5208.

[150] M. Pivk and F. R. Le Diberder, *Plots: A statistical tool to unfold data distributions*, Nuclear Instruments and Methods in Physics Research A **555** (2005) 356, arXiv:physics/0402083.

[151] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198, arXiv:hep-ex/0011057.

[152] LHCb collaboration, R. Aaij *et al.*, *Measurements of the B^+, B^0, B_s^0 meson and Λ_b^0 baryon lifetimes*, JHEP **1404** (2014) 114, arXiv:1402.2554.

[153] F. James and M. Roos, *Minuit – a system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun. **10** (1975) 343.

[154] P. Hall, *The bootstrap and edgeworth expansion*, Springer Series in Statistics, Springer, New York, 1992.

[155] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the effective $B^0_s \to K^+K^-$ lifetime*, Phys. Lett. **B716** (2012) 393, arXiv:1207.5993.

[156] L. Sestini, *Studio del decadimento $B^0_d \to D^{*-}\mu^+\nu$ per l'ottimizzazione di algoritmi di identificazione del sapore nell'esperimento LHCb*, INFN Thesis 8680, 2013.

[157] M. A. Ivanov, J. G. Korner, and P. Santorelli, *Semileptonic decays of the B_c meson*, Phys. Rev. **D63** (2001) 074010, arXiv:hep-ph/0007169.

[158] P. Colangelo and F. De Fazio, *Using heavy quark spin symmetry in semileptonic B_c decays*, Phys. Rev. **D61** (2000) 034012, arXiv:hep-ph/9909423.

[159] J.-F. Liu and K.-T. Chao, *B_c meson weak decays and* CP *violation*, Phys. Rev. **D56** (1997) 4133.

[160] C.-H. Chang, Y.-Q. Chen, G.-L. Wang, and H.-S. Zong, *Decays of the meson B_c to a P-wave charmonium state χ_c or h_c*, Phys. Rev. **D65** (2001) 014017, arXiv:hep-ph/0103036.

[161] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of B-Hadron, C-Hadron, and tau-lepton properties as of early 2012*, arXiv:1207.1158.

[162] LHCb Collaboration, *How long can beauty and charm live together?*, CERN Courier **January/February** (2014).

[163] LHCb collaboration, R. Aaij *et al.*, *Measurement of the lifetime of the B^+_c meson using the $B^+_c \to J/\psi\pi^+$ decay mode*, arXiv:1411.6899.

[164] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, arXiv:physics/0306116.

[165] W. Zucchini, A. Berzel, and O. Nenadic, *Applied smoothing techniques*, Part I: Kernel Density Estimation (2003) 15.

[166] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*, CRC press, 1984.

[167] M. Williams, *How good are your fits? Unbinned multivariate goodness-of-fit tests in high energy physics*, JINST **5** (2010) P09004, arXiv:1006.3019.