

Doctoral Dissertation

博士論文

**Measurement of the Flavor Changing
Neutral Current Decays $B \rightarrow K\ell^+\ell^-$
at the Belle II Experiment**

(Belle II実験におけるフレーバー変換中性カレント崩壊
 $B \rightarrow K\ell^+\ell^-$ の測定)

A Dissertation Submitted for the Degree of

Doctor of Philosophy

令和2年7月博士（理学）申請

July 2020

Department of Physics, Graduate School of Science

The University of Tokyo

東京大学大学院理学系研究科物理学専攻

Johnny Alejandro Mora Grimaldo

ジョニー アレハンドロ モラ グリマルド

To my parents, of course

Acknowledgements

There is a long list of names of people whose support was by different means crucial for the completion of this work, and I believe any attempt to write it will surely leave half of them behind. Nevertheless, I would like to start with the most obvious ones, Yoni Mora and Geny Grimaldo, my parents, for their unconditional support—even across half the world—and my brother, Daniel Mora, for representing everything I like from my homeland.

I would like to show my gratitude to professor Hiroaki Aihara, for granting me the opportunity to grow as a scientist during these five years; to professor Yoshiyuki Onuki, for his kind—and almost paternal—support; and to professor Masashi Yokoyama, for showing me what can be achieved when we love what we do.

I surely must thank professor Akimasa Ishikawa, for his generous and selfless guidance, which took me out of the shadows more than once, and without which this study would not have been started.

I must also mention Kuniko Kono, whose diligence and support was always relieving.

The members of my research laboratory too aided me in different ways, from my daily life in a new country, to the problems that arise in every research project, and also by just providing a space to rest and relax within the academic environment. They are, without any order of preference: Naruhiro Chikuma, Konosuke Iwamoto, Taichiro Koga, Nobuhiro Shimizu, Yusuke Suda, Hiroko Niikura, Fuminao Hosomi, Wan Kun, Tianmeng Lou, Tatsuki Ose, Kohei Matsushita, Riku Tamura, Yifan Jin, Tang Shenli, Zhang Tingyu, Natsumi Ogawa, Haruto Kikutani, Aoi Eguchi, Yoshimi Yoshimoto and Ian Watson.

Many thanks to the members of the Tetsumon Soccer Club of The University of Tokyo, which welcomed me as a peer from day one, and offered me the chance to enjoy one of my passions.

My appreciation also goes to all the members of the Belle II collaboration; to the Electro Weak Penguin Group members, for their valuable insights and comments through this work, with a special mention to Simon Whele and Saurabh Sandilya,

who were always willing to help as group conveners; to Sam Cunliffe, for his outstanding role as leader of the software group; to Patrick Ecker for his previous work on the new Bremsstrahlung Module used in this analysis, and to Thomas Keck for the fastBDT implementation of Boosted Decision Trees; to Ilya Komarov, for his dedication into making the new commers familiar with the software analysis tools; to Miho Fujii, Kentaro Kuze and Toshiki Yoshinobu, for making me feel part of the japanese students at Belle II; to Thomas Czank, Michell Villanueva and Marxil Sánchez, for their friendship and to all the people who joined the soccer matches after the Belle II General Meetings.

Of course, I have an enormous debt with Yo Sato, whom with I first worked with in the Monte Carlo matching code for Bremsstrahlung particles; his comments, advises and support were fundamental not only for the code development, but also for almost all aspects of this study.

I would like to show my gratitude as well to the members of the committee, for their valuable insights. To professor Yasuyuki Okumura specially, for his unlimited patience and guidance.

Last, but not least, I would like to thank all the people who made my life in Japan richer and enjoyable, and that supported me in the daily struggles of living in a foreign country: Yukiko Yoshida, Miyuki Sagae, Hitomi Yanashima, Sakurai Yumeko, Haruka Higuchi, Haruka Kajiwara, Anne Lee, David Jiménez, Álex Cruz, Naruhiro Chikuma (again), Marco Bertolio, Marc-Olivier Jaekel, Neil Waters, Alexander Eloi, Kei and Yoshie Arashima, and many, many others. May the winds of fortune be at your back.

Abstract

Flavor Changing Neutral Currents (FCNC) constitute excellent probes for physics beyond the Standard Model, since their branching fractions can be affected by the presence of new, heavy particles as mediators. Of special interest are the $B \rightarrow K\ell^+\ell^-$ decays—where B is either a B^0 or a B^+ meson, K is either a K^+ or a K^0 , and ℓ is either a μ or an e —given their relatively high branching fraction and smaller theoretical uncertainties (in comparison with other FCNC), and the multiple new couplings they can test. Great efforts have been put into measuring the value of R_K , defined as

$$R_K = \frac{\mathcal{B}(B \rightarrow K\mu^+\mu^-)}{\mathcal{B}(B \rightarrow Ke^+e^-)},$$

since it is theoretically very clean, and recent measurements show tension with the Standard Model predicted value of 1.

These decays are extremely rare, and as such, their uncertainties are dominated by the size of the data sample used to study them. New particle accelerators, focused on extreme luminosities, have the potential to narrow these uncertainties and to provide more precise measurements. The Belle II Experiment, which started recording physics events from the summer of 2019, aims for an integrated luminosity of 50 ab^{-1} , by achieving a record of instantaneous luminosity ($\sim 8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$). Since it is operating at the SuperKEK-B, a B meson factory, its reconstruction efficiency for B decays is higher than more general experiments such as the LHC; thus, it represents an ideal setup for a more precise study of the $B \rightarrow K\ell^+\ell^-$.

The current work is the first analysis of the $B \rightarrow K\ell^+\ell^-$ decays using the Belle II experiment data accumulated in its first physics runs up to March of 2020. It consists of 11.5 fb^{-1} of collisions recorded at a Center of Mass Energy equal to the $Y(4S)$ mass; besides the novelty of being the first study of these processes at the Belle II Experiment, this analysis also introduces a new, improved algorithm for Bremsstrahlung recovery, uses Boosted Decision Trees as multivariate classifiers in order to reduce the different background components in the data samples, and

applies the technique of *boosting to flatness* in order to avoid introducing bias through the multivariate classifiers' outputs. In the study, the ratio R_K is measured for the dilepton invariant mass regions corresponding to the J/ψ and $\psi(2S)$ resonances to be

$$\frac{\mathcal{B}(B \rightarrow KJ/\psi [e^+e^-])}{\mathcal{B}(B \rightarrow KJ/\psi [\mu^+\mu^-])} = 0.99 \pm 0.09 \text{ (stat)} \pm 0.01 \text{ (sys)},$$
$$\frac{\mathcal{B}(B \rightarrow K\psi(2S) [e^+e^-])}{\mathcal{B}(B \rightarrow K\psi(2S) [\mu^+\mu^-])} = 1.03 \pm 0.41 \text{ (stat)} \pm 0.01 \text{ (sys)}.$$

The decay mode $B^+ \rightarrow K^+e^+e^-$ is observed with a significance of 1.35 and a 90% C.L. upper bound on its branching fraction is measured to be

$$\mathcal{B}(B^+ \rightarrow K^+e^+e^-) < 3.5 \times 10^{-6}.$$

The other three decays are not observed; the 90% C.L. upper bound on their branching fractions are:

$$\mathcal{B}(B^0 \rightarrow K^0\mu^+\mu^-) < 3.2 \times 10^{-6},$$
$$\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-) < 9.9 \times 10^{-7},$$
$$\mathcal{B}(B^0 \rightarrow K^0e^+e^-) < 2.3 \times 10^{-6}.$$

All of these are consistent with measurements by previous experiments, and with the predictions made by the Standard Model.

We also perform a sensitivity projection for both the branching fractions and the R_K ratio at higher luminosities, and show that the improvements in the detector performance are required to meet previous expectations for the R_K ratio at Belle II—derived from extrapolations of the Belle measurements—; based on these projections, and pending on the improvements mentioned, we predict that the Belle II experiment will be able to settle the R_K anomaly with a 5σ significance at 35 ab^{-1} .

Table of contents

1	Introduction	1
1.1	Contributions and Novelties Introduced	1
1.2	The Standard Model	2
1.2.1	NP Searches in $b \rightarrow s\ell^+\ell^-$ Transitions	7
2	Flavor Changing Neutral Currents in B decays and Analysis Strategy	17
2.1	Phenomenology of the Standard Model and Flavor Changing Neutral Currents in B Decays	17
2.1.1	The Operators Product Expansion and the Effective Hamiltonian	17
2.1.2	Effective Hamiltonian for the $B \rightarrow K\ell^+\ell^-$ Transitions	19
2.1.3	The Decay Rate	21
2.2	Strategy for this Study	25
3	Experimental Apparatus	29
3.1	The Accelerator: SuperKEKB	29
3.2	The Detector: Belle II	31
3.2.1	VXD	33
3.2.2	CDC	35
3.2.3	PID Detectors	36
3.2.4	ECL	39
3.2.5	KLM	40
3.2.6	Triggering and DAQ	41
3.3	The Analysis Framework: basf2	44
4	Event Reconstruction and Background Analysis	47
4.1	MC Datasets	48
4.2	Event Selection	48
4.2.1	Charged Particle Identification	49

4.2.2	Skimming	50
4.2.3	Selection of Final State Charged Particles	51
4.2.4	K_S^0 Reconstruction	56
4.2.5	B Reconstruction	58
4.2.6	Sum-up of the event selection cuts and ROE	61
4.3	Background analysis	63
4.3.1	Peaking backgrounds	64
4.3.2	Continuum Suppression	70
4.3.3	$B\bar{B}$ Suppression	81
4.3.4	Selection of the Classifiers' Cuts	89
5	Data Analysis Preliminaries	95
5.1	Dataset	95
5.2	Efficiency Corrections	96
5.2.1	Track Finding	96
5.2.2	Lepton ID	99
5.2.3	Kaon ID	100
5.2.4	K_S^0 Selection	102
5.2.5	MVA Classifiers	106
5.3	Final Efficiencies	108
6	Signal Extraction	115
6.1	Signal Extraction	116
6.1.1	PDF Components	116
6.1.2	Fitting	127
6.2	Error Analysis	130
6.2.1	Systematic Errors from the Fitting Parameters	130
7	Discussion and Conclusions	135
7.1	Discussion	135
7.1.1	Branching Fractions	135
7.1.2	R_K	148
7.2	Conclusions	152
	References	155

Appendix A A Speed Course in the Standard Model	163
A.1 Particles	163
A.2 Interactions	166
A.3 Observables, Scattering Matrix and Feynman Diagrams	172
A.4 The Hamiltonian Approach	173
Appendix B The Cabibbo-Kobayashi-Maskawa Matrix	175
Appendix C How is the Bremsstrahlung Relation Set	179
Appendix D Figure of Merit Plots for the Charmonium Veto Regions	183
Appendix E Boosted Decision Trees	187
E.1 Boosted Decision Trees	188
E.2 Boosting to Flatness	190
Appendix F Distribution of the MVA Training Variables	193
Appendix G Eff. and Fake Rate Correction Tables from PID Studies	205
Appendix H Fits to MC and Data Inside the Charmonium Sidebands	221
Appendix I Fit Results for the K_S^0 Systematics	225
Appendix J The Extended Maximum Likelihood Technique	239
J.1 Likelihood fit	239
J.2 Extended likelihood	240
J.3 Significance	241
Appendix K Extra Fits	243
K.1 Non-peaking backgrounds	243
K.2 Leaked charmonium backgrounds	245
K.3 $K\pi^+\pi^-$ backgrounds	246
K.4 Signal	248
K.5 Combined	249
Appendix L Systematic Errors for the Peaking Background Yields	251
Appendix M Fits to MC Data Using Belle Parameterization for the NP Bkg.	253

Appendix N Statistical Error Calculation for R_K at Higher Selection Effs.	257
--	------------

Chapter 1

Introduction

1.1 Contributions and Novelty Introduced

This work represents the first study of the $B \rightarrow K\ell^+\ell^-$ processes in the Belle II experiment; as such, it is intended to properly systematize the analysis procedure, and to serve as the main guide for future studies in the topic. We have put special emphasis in the derivation of the reconstruction and background rejection strategies, shown in chapter 4, which is sometimes overlooked in this kind of work. The methodology followed here, far from being limited to the current analysis, can be easily extended to studies where a multidimensional fit is used to extract the signal yields, or where the quantity to be optimized is other than the pseudo-significance (which is the target variable each of the selection cuts in this analysis tries to maximize).

This work also introduced a new Bremsstrahlung recovery method, which is now widely used by the Belle II collaboration. It is an optimized version of the algorithm used at Belle, and has shown promising results in the rejection of the charmonium background that plagues the $B \rightarrow K\ell^+\ell^-$ decays, without incurring in any major degradation of the reconstruction efficiency for the electron modes. This novel method is described in chapter 4.

We also perform the rejection of certain background components by making use of multivariate classifiers. In contrast with previous studies at B factories, which mostly employ Neural Networks as the classifier architecture, we opt for *Boosted Decision Trees*, given their robustness to correlations among the input variables, the short time required for their tuning/training, and their interpretability. Moreover, this analysis includes a novel additional term in the loss function of these classifiers, in order to assure that their output is not correlated with the m_{bc} , which is the

fit variable from which the signal yields are obtained. This with the intention of avoiding introducing any bias —through artificial peaks— in the m_{bc} distribution. The details of the training process and this new extra term are given in chapter 4 and in appendix E, respectively.

Finally, in chapter 7, the present study ends with a projection on the sensitivity for the branching fraction and the R_K measurements at higher luminosities, in optimistic and conservative scenarios. This provides future studies with a threshold against to which compare their own performance.

Without further introduction, let us begin.

1.2 The Standard Model

From the first mention of the word *atom* by the Greek philosophers Leucippus and his pupil Democritus, more than 2400 years ago, humanity has traveled a long and enlightening path towards the understanding of the structure of matter. The joint efforts of scientists all over the world (and centuries) allowed for the establishment and corroboration of what is now known as the Standard Model of Physics: a theory that describes the interaction of entities known as *elementary particles*, the basic constituents of our observable universe. It does so by treating these entities as quantized excitations of fields that extend throughout space and time; as such, the Standard Model is a *Quantum Field Theory* (QFT). In fact, it is the most complete QFT up to date, and it correctly predicts a large number of the experimental observations in Particle Physics. The discovery of the Higgs boson on July of 2012 by the CMS [1] and the ATLAS [2] collaborations constituted the last piece of evidence of the veracity of the model.

As a quantum field theory, the Standard Model is based on both Quantum Mechanics and Special Relativity, the greatest revolutions in physics during the 20th century. The first postulates that particles can be considered as excited states, or quanta, of their underlying fields, which are functions of space and time. The second, that the laws of physics are the same in any inertial frame¹, and that speed of light c

¹The motion of any object can be described only relative to an abstract coordinate system; such system is known as a frame of reference. A frame of reference in which space is homogeneous and isotropic (this is, in which laws of physics do not vary with the spatial position) and time is homogenous (laws of physics do not change between different instants of time) is known as an *inertial* frame of reference. In such a frame, an object subject to no forces —a free body— which is at rest at one instant remains always at rest [3].

is the maximum attainable speed in the universe—and hence, that there cannot be instantaneous interactions between particles at different points in space.

As first proven by Wigner [4], these assumptions are enough to develop the framework for Quantum Field Theories. Complemented by years of experimental observations², such theories were continuously depurated, and finally the current picture of our observable universe came to be: matter is composed of quanta of fermionic³ fields; they interact with each other by exchanging quanta of bosonic fields. There are three types of bosonic fields, corresponding to the three elementary forces⁴: the electromagnetic, the weak and the strong forces. The mediator of the electromagnetic field is the photon (γ) and the ones of the strong field are eight particles known as gluons (g); the weak force has three mediators: the electrically charged W^\pm , and the Z .

Fermions can be further separated in two types, depending on whether they participate in the strong interaction or not. Those that do are known as quarks, and those that don't, as leptons, from which the *electron* is the most common one. Both quarks and leptons are divided in *generations*, which correspond to pairs that interact with each other through the exchange of a W^\pm boson. Besides their mass, the different quark (lepton) generations have very similar properties between each other.

The Higgs boson is the only *scalar* boson in the theory⁵; it gives mass to the fermionic fields, and to the W and Z bosons⁶. The mechanism by which this is achieved is known as the *Higgs* mechanism, and it is briefly discussed in Appendix A.

Table 1.1 resumes the Standard Model particles and interactions. Each column in the fermion group correspond to a generation (they are usually termed first, second and third generation, due to the chronological order of their discovery). Leptons on the first row have an electric charge of zero and are known as *neutrinos*, whereas

²The interested reader can consult any book on the history of particle physics or, for that matter, any book on particle physics. See, for example, the first chapter of [5].

³The fermionic and bosonic nature of elementary particles dictates how many of them can share the same physical state, and how do their underlying fields transform when described in different inertial frames.

⁴Strictly speaking, gravity *is* an elementary force as well; however, there is no consensus on how to quantize it. This remains an open question to this day, and it represents one of the big problems with the Standard Model.

⁵The other bosons are *vector* bosons. This terminology has to do with the magnitude of their intrinsic angular momentum, or spin.

⁶the large mass of its mediators effectively limits the range of the weak force; this is the reason why it is *weak*.

those in the second one carry a charge of -1 . For quarks, elements in the first row have an electric charge of $2/3$, and those in the second one of $-1/3$. For leptons, mass increases while moving down and to the right in the table; for quarks, mass increases while moving up and to the right, except for the first column, where the mass of the u quark is smaller than the mass of the d quark. As for bosons, mass increases while going down.

Table 1.1. The Standard Model of Physics

Fermions				Bosons	
				Name	Interaction
Leptons	ν_e	ν_μ	ν_τ	Photon	Electromagnetic
	e	μ	τ	Gluon	Strong
Quarks	u	c	t	W^\pm, Z^0	Weak
	d	s	b	Higgs	Higgs interaction

A few more comments are on point (most of these are treated somehow more rigorously in Appendix A):

- Each field carries an intrinsic property called *spin*, which is the quantum analogous of angular momentum: fermions have spin $1/2$, scalar bosons have spin zero, and vector bosons have spin one. This property dictates how the fields transform when described in different inertial frames, and if many particles can occupy the same physical state⁷. Moreover, when particles are put under a magnetic field, the spin dictates how strong is their coupling with it. The naming is somewhat misleading, however; elementary particles are not truly spinning, at least not in the classical sense.
- Besides spin, there is another intrinsic property of particles known as *chirality*, which further determines how the fields' descriptions transform between inertial frames, and whose possible values are -1 and 1 .
- For each particle there is a corresponding antiparticle, which has the same mass, spin and parity, but opposite charges⁸. The Higgs, the Z , the γ and the g are their own antiparticle (and thus all of their charges are zero).

⁷Particles described by fields with a half-integer spin cannot, and are said to obey the Fermi-Dirac statistics; those with integer spin are allowed to, and are described by the Bose-Einstein statistics. Hence, the names *fermion* and *boson*.

⁸The concept of charge here is not limited to the electric charge; in the strong interaction, for example, the equivalent of electric charge is color charge, for which there are three possible values: red, green, and blue. An up quark carrying electric charge $2/3$ and a red color charge, will have as antiparticle an anti-up quark, with electric charge $-2/3$ and an anti-red color charge.

- The charged weak force violates parity symmetry maximally, *id est*, it only couples to particles (and antiparticles) with negative parity. The neutral weak interaction also violates parity symmetry, but particles with either parity can still take part on it, with different coupling strengths. Moreover, this is the only interaction that can change the flavor of a particle (by mixing the first and second component of each generation of quarks and leptons).
- The interaction of a fermion with the Higgs field *flips* the chirality of said fermion. A left-handed particle becomes a right-handed one and vice versa.
- The classical potential of the strong force can be described by a term proportional to the distance: $V_{\text{strong}}(r) \propto kr$. This implies that a system of two or more distant quarks carries a large amount of energy, and, in order to get to a stable configuration, it will create quark-antiquark ($q\bar{q}$) pairs which will then form bounded states. Hence, all free particles are color neutral. This neutrality can be achieved by combinations of three quarks of different color, which are known as baryons⁹, or by the combination of a quark and an antiquark with opposite color charges, known as mesons.
- Due to the form of its classical potential, theoretical calculations using perturbation techniques in the low energy region —where it is not possible to probe short distances, and hence $V(r)$ is large— are not as accurate for the strong interaction as they are for the other forces.

Challenges to the Standard Model and Searches for New Physics

As remarkable as the theoretical framework of the Standard Model is for the description of many the natural world, it is evident that our understanding of all its processes is far from complete. Indeed, the SM gets in difficulties when trying to interpret a variety of phenomena, among which we can mention:

- The matter-antimatter asymmetry in the universe: if in the early stages of the universe particles and antiparticles were created on equal amounts, why is it that our observable universe seems mostly constituted by particles?
- The hierarchy problem, or why is the mass of the Higgs boson so small, despite the expected quantum corrections from loops of virtual particles?

⁹The most famous ones being the proton, which is composed of two u quarks and one d quark, and the neutron, which is composed of two d quarks and one u quark.

- The nature of Dark Matter and Dark Energy.
- The strong CP problem, or, how is it that strong interactions do not violate CP symmetry when, in principle, they can?
- The origin of neutrino masses and the neutrino mass scale.
- The total number of Higgs bosons.
- The stability of the proton.
- The quantization of the gravitational field or, more precisely, a quantization of the gravitational field which is compatible with General Relativity.

This inability to offer an explanation to any of the previously mentioned items (as well as many others) suggests that the SM is an incomplete, or effective, theory, an approximation to a more fundamental model which is still unknown. Multiple hypothesis about the nature of this model have been proposed and, as with the Standard Model, they need to be corroborated (or refuted) by precise measurements performed in dedicated experiments; these include cosmological observations (searches for dark matter, cosmic rays studies, cosmic neutrino observatories, general relativity tests with celestial bodies, and studies on Big Bang remnants), neutrino experiments (from both reactor and accelerator sources) and particle accelerators.

Accelerator searches can be further divided into two types: the Energy Frontier and the Intensity Frontier programs. The first one probes for New Physics by reaching higher and higher energy scales where new, heavy particles can be directly produced. The LHC accelerator is a good example of a machine build under this scheme, holding the world record for the center of mass energy of an accelerator at 14 TeV, with some plans of rising it to 27 TeV [6].

On the other side, the Intensity Frontier strives for maximizing the number of observed "interesting" events, by increasing the intensity of the colliding beams (thus raising the number of collisions per unit of time) and by the implementation of special detection and reconstruction techniques. Examples of experiments built under the intensity frontier program include BaBar, Belle, and its successor, the Belle II experiment, which aims to reach a new world record in instant luminosity: $8 \times 10^{35} \text{cm}^{-2}\text{s}^{-1}$.

The intensity frontier can provide signs of New Physics in two ways: one is by looking for processes which are forbidden in the Standard Model, for which a measurement is an automatic sign of New Physics. This could either probe new

couplings, or new particles with mass scales well below the Energy Frontier, but with a very weak coupling to the Standard Model particles.

The second strategy is to look for discrepancies between measurements and the Standard Model predictions for processes whose theoretical uncertainties are well under control. This is often performed with decays which occur, at leading order, through loop diagrams. In this case new particles can appear as virtual particles in the loop, and change considerably the rates and the properties of the process with respect to its Standard Model values, since both the New Physics and the Standard Model contribution would be of the same order.

Flavor Changing Neutral Currents (FCNCs) present an ideal environment for studies of this last nature. They correspond to processes in which the flavor of a quark changes, but its charge remains unaltered. From table 1.1, they can be understood as transitions from one quark column to another, within the same row. FCNCs can only occur through loop diagrams in the Standard Model (more on Feynman diagrams can be read in appendix A and the literature referenced there). Moreover, theoretical uncertainties within these processes can be palliated by performing measurements on inclusive decays (where the hadronic component of the final state is not unique), or by choosing a proper set of clean observables.

1.2.1 New Physics Searches in $b \rightarrow s\ell^+\ell^-$ Transitions

Multiple extensions of the Standard Model produce contributions to Flavor Changing Neutral Current processes by the introduction of new, massive particles, which can alter the value of the observables measured by experimentalists. Of special interest are the processes in which the decaying quark is a b quark¹⁰. Moreover, radiative (where the $b \rightarrow s$ transition is accompanied by the emission of a photon) and semileptonic (where the $b \rightarrow s$ transition occurs together with the production of a lepton pair) decays provide cleaner theoretical environments in comparison to purely hadronic decays, where the calculation of the strong interaction terms between all the quarks involved leads to higher uncertainties.

The $b \rightarrow s\gamma$ Decay

The radiative $b \rightarrow s\gamma$ transition occurs through the effective diagram shown in figure 1.1. The decay rate for this process is governed by the parameter C_7 , one of the *Wilson coefficients* used in the treatment of rare decays of B mesons. New Physics scenarios

¹⁰Because of reasons explained in the appendix B.

can manifest by modifying this coefficient, and by the introduction of the chirality-flipped coefficient \mathcal{C}'_7 , which, in the SM, is suppressed by a factor of m_s/m_b , where m_s and m_b are the masses of the s and b quarks, respectively. The inclusive decay $\bar{B} \rightarrow X_s \gamma$, where X_s corresponds to any hadron containing an s quark and without a c quark, and B stands for both \bar{B}^0 ($b\bar{d}$ quarks) and B^- ($b\bar{u}$ quarks) mesons, is one of the best probes for this process, as inclusive branching ratios are theoretically clean. On the contrary, the exclusive decays $\bar{B} \rightarrow \bar{K}^{(*)} \gamma$ —where \bar{K} is either \bar{K}^0 ($s\bar{d}$) or K^- ($s\bar{u}$)—are theoretically challenging, but offer a way of measuring the ratio $\mathcal{C}'_7/\mathcal{C}_7$, either by computing the difference in the decay rates for CP -conjugated processes, or by looking at the chirality of the outgoing photon in decays with three-body hadronic final states. The Standard Model prediction for the branching fraction for photons with energies $E_\gamma > 1.6$ GeV is given by [7]

$$\text{BR}(\bar{B} \rightarrow X_s \gamma) = \frac{G_F^2 \alpha_e m_b^5}{32\pi^4} |\lambda_t \mathcal{C}_7|^2 = (3.36 \pm 0.23) \times 10^{-4} \quad (1.1)$$

where G_F is the Fermi constant for weak decays, α_e is the fine structure constant for electromagnetism, and $\lambda_t = V_{tb} V_{ts}^*$ a product of components of the CKM matrix (see appendix B).

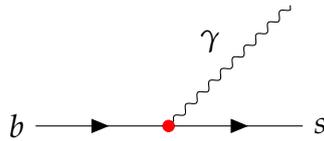


Fig. 1.1 Effective diagram for the $b \rightarrow s \gamma$ decay. The effective vertex (shown in red) encapsulates the short distance physics from the weak interaction responsible for $b \rightarrow s$ FCNC. It is also the source of the \mathcal{C}_7 term.

The latest experimental measurements carried on at dedicated B factories are shown in figure 1.2. They are in good agreement with the theoretical value.

The Belle II experiment should be able to perform a measurement of $\text{BR}(B \rightarrow X_s \gamma)$ within a statistical precision of around 6%, and of the CP -asymmetry in $B \rightarrow K^* \gamma$ and $B \rightarrow \rho \gamma$ of 0.05 and 0.15, respectively [9].

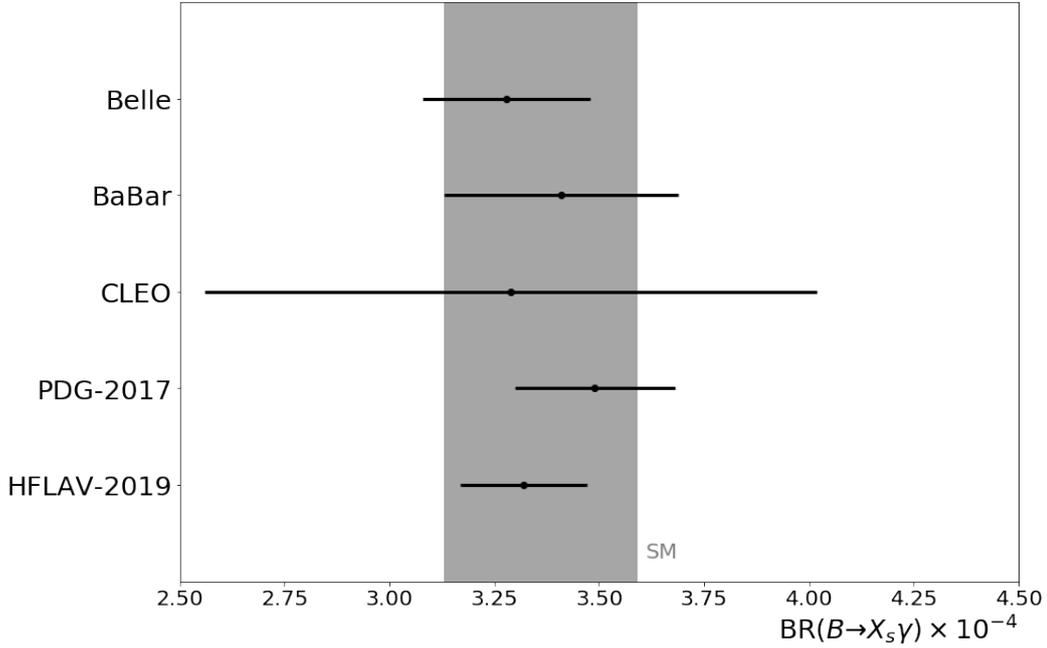


Fig. 1.2 Experimental measurements for the $b \rightarrow s\gamma$ branching ratio at 90% C.L., as reported by the Heavy Flavor Averaging Group (HFLAV) [8]. The values for the Belle and BaBar experiments are the averages of several results, obtained by different methods. Though theoretical calculations are done for $E_\gamma > 1.6$ GeV, measurements are performed for different energies, all above $E_\gamma > 1.7$ GeV; the values shown here have been extrapolated down to 1.6 GeV from such measurements. The horizontal bars correspond to the total error (adding the statistical and systematic errors in quadrature). The shadowed band corresponds to the prediction region of the Standard Model.

The $b \rightarrow s\ell^+\ell^-$ Decay

Compared to radiative and leptonic¹¹ decays, the semileptonic $b \rightarrow s\ell^+\ell^-$ process, where ℓ can be either an electron or a muon¹², is mediated, at leading order, by two

¹¹These have not been (and will not be) discussed here. They correspond to processes of the form $B \rightarrow \ell^+\ell^-$, where ℓ can be either an electron or a muon. In the SM, their production rate is dictated by the Wilson coefficient \mathcal{C}_{10} . For a brief introduction to the experimental physics analyses on this sector, the reader can refer to [9]. In a joint effort, the CMS and the LHCb collaboration were the ones to report the first observation of the $B_s^0 \rightarrow \mu^+\mu^-$ process; the LHCb later reported the most precise measurement to date of the branching fraction for this process [10]. The ATLAS collaboration provided a 95% C.L. upper bound to the branching fraction of $B^0 \rightarrow \mu^+\mu^-$ decays, and reported results within 2.6 standard deviations from the Standard Model predictions [11].

¹²In principle, ℓ^- can be a τ^- lepton too. However, the τ is much more massive than the electron and the muon, and hence its production rate is suppressed. Moreover, it promptly decays into states with two or more neutrinos, which makes its detection very challenging. The BaBar collaboration performed a measurement on the branching fraction of the $B^+ \rightarrow K^+\tau^+\tau^-$ decay, and reported an upper limit of 2.25×10^{-3} at 90% C.L. [12], which is about four orders of magnitude above the SM prediction of $\text{BR}(B^+ \rightarrow K^+\tau^+\tau^-) = 1.44(15) \times 10^{-7}$ [13].

effective diagrams (see figure fig:effective-bsll), which makes it sensitive to a broader spectrum of new physics contributions.



Fig. 1.3 Effective leading diagrams for the $b \rightarrow s \ell^+ \ell^-$ decay. The FCNC vertices are shown in red. While the first diagram represent the contribution from C_7 , the second one includes both the effects of C_9 and C_{10} .

The decay rate in the absence of QED corrections can be expressed in terms of the squared dilepton invariant mass q^2 and θ_ℓ , the angle between ℓ^- and the B meson in the $\ell^+ \ell^-$ center of mass frame, as [14]

$$\frac{d^2\Gamma}{dq^2 d\cos\theta_\ell} = \frac{3}{8} \left[\left(1 + \cos^2\theta_\ell\right) H_T(q^2) + 2H_A(q^2) \cos\theta_\ell + 2 \left(1 \cos^2\theta_\ell\right)^2 H_L(q^2) \right] \quad (1.2)$$

where the $H_i(q^2)$ functions contain the dependence on the Wilson coefficients, and are independent of $\cos\theta_\ell$. The total branching fraction will be independent of H_A , but the forward-backward asymmetry, defined as

$$A_{FB} = \frac{\int_0^1 d\cos\theta_\ell \frac{d\Gamma}{d\cos\theta_\ell} - \int_{-1}^0 d\cos\theta_\ell \frac{d\Gamma}{d\cos\theta_\ell}}{\int_{-1}^1 d\cos\theta_\ell \frac{d\Gamma}{d\cos\theta_\ell}} \quad (1.3)$$

does, and provides another observable to test for NP.

As with the $b \rightarrow s \gamma$ transition, the semileptonic FCNC can be studied both in inclusive and exclusive decays. The $B \rightarrow X_s \ell^+ \ell^-$ decays, although theoretically clean, are experimentally challenging, due to the high multiplicity of final states and the large backgrounds expected in the high m_{X_s} region; usually a cut on this hadronic invariant mass is imposed, which both suppresses the decay rate in the low q^2 region and introduces additional uncertainties. The latest measurements on the inclusive branching fractions are shown in table 1.2.

The only measurement of A_{FB} in inclusive decays was performed by the Belle collaboration by using the sum of 10 exclusive decays in the range $m_{X_s} > 1.1$ GeV [20], and the results are in agreement with the Standard Model within the theoretical and statistical uncertainties.

Table 1.2. Latest measurements and averages at 90% C.L. of the branching fractions ($\times 10^{-6}$) for the $B \rightarrow X_s \ell^+ \ell^-$ modes [8, 15]. The Belle measurements include the region $q^2 > 0.2$ GeV, whereas the BaBar ones take $q^2 > 0.1$ GeV, and are in agreement (within 2σ) with the SM predictions, which are: $(4.6 \pm 0.8) \times 10^{-6}$ for the $B \rightarrow X_s \ell^+ \ell^-$ mode [16], $(4.15 \pm 0.71) \times 10^{-6}$ for $B \rightarrow X_s \mu^+ \mu^-$ and $(6.89 \pm 1.01) \times 10^{-6}$ for $B \rightarrow X_s e^+ e^-$ [17].

Mode	BaBar [18]	Belle [19]	PDG-2020	HFLAV-2019
$B \rightarrow X_s e^+ e^-$	$7.69^{+0.82+0.71}_{-0.77-0.6}$	$4.05 \pm 1.30^{+0.87}_{-0.83}$	6.7 ± 1.7	6.67 ± 0.82
$B \rightarrow X_s \mu^+ \mu^-$	$4.41^{+1.31+0.63}_{-1.17-0.50}$	$4.13 \pm 1.05^{+0.85}_{-0.81}$	4.3 ± 1.0	$4.26^{+0.98}_{-0.91}$
$B \rightarrow X_s \ell^+ \ell^-$	$6.73^{+0.70+0.60}_{-0.64-0.56}$	$4.11 \pm 0.83^{+0.85}_{-0.81}$	5.8 ± 1.3	5.84 ± 0.69

On the other side, exclusive $B \rightarrow M \ell^+ \ell^-$ decays are experimentally easier and deliver a larger number of observables, but with the cost of higher theoretical uncertainties, since the calculation of the branching fractions requires the knowledge of the $B \rightarrow M$ form factors; moreover, while in the $b \rightarrow s \gamma$ case this form factor needs only to be calculated for the $q^2 = 0$ case (where q^2 is the invariant mass squared of the outgoing photon), semileptonic currents demand the knowledge of this form factor in the full kinematic range $4m_\ell^2 < q^2 < (m_B - m_M)^2$. Such uncertainties are smaller in decays where the meson is either a K or a π , since there are only three independent form factors, of which only two contribute to the SM prediction in the massless lepton limit (a good approximation for electrons and muons); moreover, these mesons are stable under strong interactions, which facilitates their reconstruction¹³.

After the form factors, the next most significant uncertainties are hadronic ones associated to non-factorizable corrections. These are smaller for decays in which $M = K^{(*)}$ [9]; hence, the $B \rightarrow K \ell^+ \ell^-$ processes represent the *best compromise between experimental simplicity, prediction power and theoretical cleanliness*, and they constitute the focus point of the present work. Their theoretical uncertainties can be further alleviated by the measurement of observables made of ratios —such as A_{FB} in equation 1.3—, where the form factor terms partially cancel. The golden quantity is

¹³The analysis of decays with vector mesons is, however, richer in observables, due to the dependence of the decay rate on three different measurable angles [21], in comparison to the pseudoscalar decays, where it only depends on θ_ℓ . This gives rise to a group of twelve angular observables, which can be further combined in new observables, canceling as many theoretical uncertainties as possible [22]. Deviations of 2.1σ , 3.4σ and 2.7σ from the SM predictions are observed in the observable P'_5 by the Belle [23], LHCb [24] and ATLAS [25] collaborations, respectively.

given by

$$R_K = \frac{\text{BR}(B \rightarrow K\mu^+\mu^-)}{\text{BR}(B \rightarrow Ke^+e^-)}. \quad (1.4)$$

In the Standard Model, 1.4 is, to a very high precision, equal to one [26]. Thus, any significant deviation would be a clear signal of a new physics mechanism which favors one lepton flavor over the other. The latest reported values of the branching fractions for the $B \rightarrow K\ell^+\ell^-$ processes are listed in table 1.3.

Table 1.3. Values of the branching fractions for the $B \rightarrow K\ell^+\ell^-$ modes from the PDG averages [15] and the Standard Model predictions. For the predictions, the form factors are the ones derived in [27], and the values of the Wilson coefficients are obtained from [28] at a scale $\mu_b = 4.8$ GeV.

Mode	PDG-2020	Standard Model
$B^+ \rightarrow K^+e^+e^-$	$(5.5 \pm 0.7) \times 10^{-7}$	$(6.020 \pm 0.862) \times 10^{-7}$
$B^+ \rightarrow K^+\mu^+\mu^-$	$(4.41 \pm 0.22) \times 10^{-7}$	$(6.027 \pm 0.864) \times 10^{-7}$
$B^+ \rightarrow K^+\ell^+\ell^-$	$(4.51 \pm 0.23) \times 10^{-7}$	$(6.02 \pm 0.86) \times 10^{-7}$
$B^0 \rightarrow K^0e^+e^-$	$(1.6_{-0.8}^{+1.0}) \times 10^{-7}$	$(5.575 \pm 0.735) \times 10^{-7}$
$B^0 \rightarrow K^0\mu^+\mu^-$	$(3.39 \pm 0.34) \times 10^{-7}$	$(5.581 \pm 0.756) \times 10^{-7}$
$B^0 \rightarrow K^0\ell^+\ell^-$	$(3.1_{-0.7}^{+0.8}) \times 10^{-7}$	$(5.58 \pm 0.76) \times 10^{-7}$

The most precise measurement to date was performed by the LHCb collaboration, using 5.0 fb^{-1} of data taken at center of mass energies of 7, 8 and 14 TeV. The reported value for the ratio in the $q^2 \in [1, 6] \text{ GeV}^2$ is [29]

$$R_K = 0.846_{-0.054-0.014}^{+0.060+0.016} \quad (1.5)$$

where the first error is statistical and the second is systematic. The result is consistent with the SM predictions at the level of 2.5 standard deviations. The Belle collaboration also reported the value for the R_K ratio in the same q^2 range using their full 711 fb^{-1} [30]

$$R_K = 0.98_{-0.23}^{+0.27} \pm 0.06 \quad (1.6)$$

(the first uncertainty is statistical, and the second systematic) well in accordance with the theoretical predictions.

The results from these measurements (and some others not included here, such as those from the $\Lambda_b \rightarrow \Lambda \ell^+ \ell^-$ processes) can be combined in order to constrain possible new physics scenarios. Figure 1.4 depicts the most-likely values for the NP corrections to the semi-leptonic Wilson Coefficients C_9 , C'_9 and C_{10} in the case where the final lepton pair corresponds to a $\mu^+ \mu^-$ pair. Though a modification to these coefficients from NP is a tantalizing solution to the discrepancies between SM predictions and the experimental measurements performed, the authors openly recognize that, at present, statistical and underestimated systematic uncertainties cannot be ruled out as possible explanations of such differences [31].

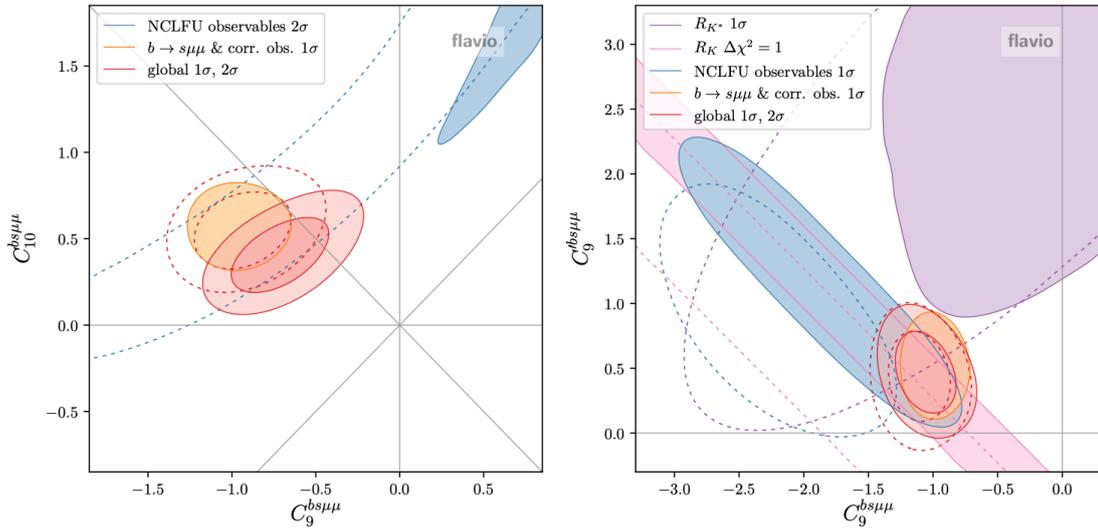


Fig. 1.4 Likelihood contours of the global fit and several fits to a subset observables in the plane of new physics contributions to the pair of operators C_9, C_{10} (left) and C_9, C_{10} (right) for muonic channels, assuming that the contribution to any other operator is zero. The solid (dashed) contours include (exclude) the latest measurements on R_K (from LHCb) and R_{K^*} (from Belle). The contours from NCLFU observables are obtained from measurements in neutral current lepton flavor universality observables: R_K, R_{K^*}, D'_{P_4} and D'_{P_5} . The orange regions correspond to 1σ constrains from $b \rightarrow s\mu\mu$ observables and other observables whose uncertainties are correlated with those of the $b \rightarrow s\mu\mu$ observables. For more details, the reader is referred to the original paper [31], which is also the source of this image.

Indeed, despite the multiple efforts, statistical error is still the dominating source of uncertainty for the measurements of R_K and R_{K^*} . Though the number of B mesons produced at the LHC accelerator is relatively high, analyses tend to suffer from low reconstruction efficiencies in the electron channels, due to the background levels expected from such high Center of Mass energies. This also limits the LHC studies

to modes with charged kaons, since the reconstruction of $K^0 \rightarrow \pi\pi(\pi)$ events is quite challenging.

In comparison, B factories background sources are better understood, and dedicated reconstruction procedures lead to similar efficiencies for both electron and muon channels, and open the possibility to the study of decays involving neutral kaons. In this sense, the Belle II Detector¹⁴, offers a great opportunity for the analysis of the semileptonic $b \rightarrow s\ell^+\ell^-$ transitions at a new level of statistical precision. The SuperKEKB aims to break the world record for instantaneous luminosity (average number of collisions per second) for an accelerator experiment, currently held by the LHC accelerator at $2.14 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$, by reaching a value of $\sim 8 \times 10^{35} \text{cm}^{-2}\text{s}^{-1}$ [32], in order to accumulate up to 50ab^{-1} of data in 10 years of operation. Figure 1.5 shows the expected uncertainty in the measurement of R_K at Belle II, as a function of the integrated luminosity. The error is expected to be comparable to the LHCb measurement at around 11ab^{-1} . However, this estimation is done by extrapolations of the uncertainties obtained by Belle analyses, and thus a projection based on the Belle II performance at early stages is bound to be more realistic, and to provide new insights on what needs to be improved in order to get closer to (or surpass) the extrapolated values.

Following these motivations, this thesis presents the first analysis on the $B \rightarrow K\ell\ell$ decay modes, aiming to measure the R_K ratio, in the Belle II Detector. This study was performed using the data from the first physics runs carried on with the whole detector setup —excluding one layer of the innermost pixel vertex detector—, from March of 2019 to March of 2020; the integrated offline luminosity is $11.5329 \pm 0.0016 \text{fb}^{-1}$, which corresponds to approximately 12 million of $B\bar{B}$ pairs. We make a detailed calculation of the systematic and statistical uncertainties related to the measurements, and calculate their expected values at the Belle II target luminosity of 50ab^{-1} ; we then obtain a revised luminosity threshold at which the Belle II measurements can compete with the Belle and the LHCb ones in terms of precision.

The outline of this document is as follows: in chapter 2, a brief account of the theoretical framework used in the study of $B \rightarrow K\ell^+\ell^-$ decays and the strategy followed in this work are presented; chapter 3 gives a description of the SuperKEKB accelerator, the Belle II Detector and the analysis tools employed in this study; chap-

¹⁴A second-generation multipurpose spectrometer which operates at the SuperKEKB accelerator facility in Tsukuba, Japan. See chapter 3. The detector is designed to measure the properties of particles produced in electron-positron collisions at a center of mass energy of 10.58 GeV. This energy corresponds to the invariant mass of the $Y(4S)$ resonance, which decays 96% of the time to a pair of B mesons. This allows for the observation of a large number of B decays in a low background environment.

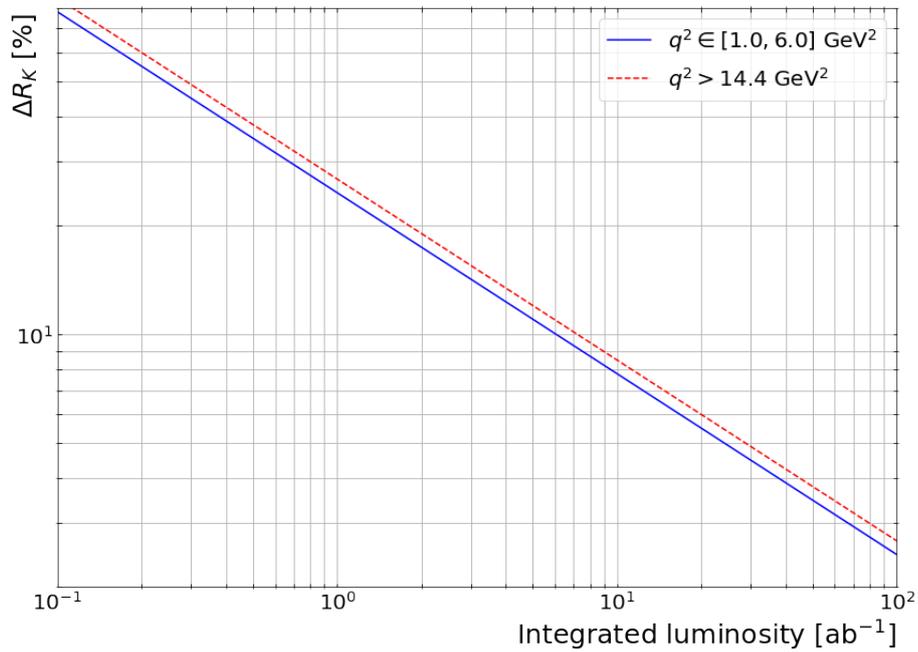


Fig. 1.5 Expected sensitivity on R_K measurements at Belle II as a function of the integrated luminosity, as appears in [33, 34]; based on extrapolations of Belle data performed in [35].

ter 4 deals with the signal background reduction strategies and the reconstruction process, where a novel Bremsstrahlung recovery module is introduced, and Boosted Decision Trees are used to suppress certain types of background events; chapter 5 presents the efficiency correction factors and the systematic errors introduced into the final efficiencies; chapter 6 then outlines the signal extraction procedure following the reconstruction process, and applies it to both the experimental and the simulated data; finally, chapter 7 presents the discussion on the results, the error projections, and the conclusions of this work.

Chapter 2

Flavor Changing Neutral Currents in B decays and Analysis Strategy

This chapter is divided in two parts: the first one presents a small review of the theoretical treatment done for Flavor Changing Neutral Currents, putting an emphasis on the set of observables which are sensitive to new physics contributions; the second one is a description of the analysis strategy followed in this study in order to measure some of these observables.

2.1 Phenomenology of the Standard Model and Flavor Changing Neutral Currents in B Decays

2.1.1 The Operators Product Expansion and the Effective Hamiltonian

Flavor changing neutral processes occur through the interplay of strong and weak interactions, and thus are dependent of two very different length scales: the weak scale, of the order¹ of $\mathcal{O}(1/m_W)$ (here m_W is the mass of the W boson) and the QCD scale, which depends on the mass of the decaying quark $\mathcal{O}(1/m_q)$. To include the effects of both forces the Operator Product Expansion (OPE), first proposed by K. G. Wilson [36] and later formalized by him and Zimmerman [37], is used. In the OPE, short-ranged forces are treated as effective point-like interactions between fermion currents; the nature of weak and highly energetic strong processes are then encapsulated in what are known as the *Wilson coefficients*, C_i , whereas the long

¹This document uses natural units, in which $c = \hbar = 1$.

distance strong interactions, which cannot be treated by perturbative methods, are represented by local operators O_i . The C_i can be viewed as coupling constants for the effective vertices O_i , and one can construct an effective Hamiltonian for the $b \rightarrow s$ FCNC given by²:

$$H_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} \sum_{q=u,c,t} V_{qb} V_{qs}^* \left(\sum_i C_i(m_q, \mu) O_i(\mu) \right) \quad (2.1)$$

where μ is the boundary between the heavy mass terms—the ones that get integrated—and the light ones,

$$G_F = \frac{\sqrt{2}}{8} \frac{g_W^2}{m_W^2}$$

is the Fermi coupling constant (g_W is the coupling constant for the weak interaction), V_{qb} and V_{qs} are components of the CKM matrix (see appendix B) and m_q is the mass of quark q . The complete set of operators is [38]:

$$\begin{aligned} O_1 &= (\bar{s}_\alpha \gamma_\mu P_L c_\beta) (\bar{c}_\beta \gamma^\mu P_L b_\alpha), & O_2 &= (\bar{s}_\alpha \gamma_\mu P_L c_\alpha) (\bar{c}_\beta \gamma^\mu P_L b_\beta), \\ O_3 &= (\bar{s}_\alpha \gamma_\mu P_L b_\alpha) \left(\sum_{q=u,d,c,s,b} \bar{q}_\beta \gamma^\mu P_L q_\beta \right), & O_4 &= (\bar{s}_\alpha \gamma_\mu P_L c_\beta) \left(\sum_{q=u,d,c,s,b} \bar{q}_\beta \gamma^\mu P_L q_\alpha \right), \\ O_5 &= (\bar{s}_\alpha \gamma_\mu P_L b_\alpha) \left(\sum_{q=u,d,c,s,b} \bar{q}_\beta \gamma^\mu P_R q_\beta \right), & O_6 &= (\bar{s}_\alpha \gamma_\mu P_L c_\beta) \left(\sum_{q=u,d,c,s,b} \bar{q}_\beta \gamma^\mu P_R q_\alpha \right), \\ O_7 &= \frac{g_e}{16\pi^2} m_b \bar{s}_\alpha \sigma_{\mu\nu} P_R b_\alpha F^{\mu\nu}, & O_8 &= \frac{g_s}{16\pi^2} m_b \bar{s}_\alpha \sigma_{\mu\nu} P_R T_{\alpha\beta}^a b_\beta G_a^{\mu\nu}, \\ O_9 &= \frac{g_e^2}{16\pi^2} \bar{s}_\alpha \gamma_\mu P_L b_\alpha \bar{\ell} \gamma^\mu \ell, & O_{10} &= \frac{g_e^2}{16\pi^2} \bar{s}_\alpha \gamma_\mu P_L b_\alpha \bar{\ell} \gamma^\mu \gamma_5 \ell, \end{aligned} \quad (2.2)$$

with α and β color indices (which are summed if repeated), $P_L = (1 - \gamma^5)/2$ and $P_R = (1 + \gamma^5)/2$ the left and right-handed spinor projectors, g_e and g_s the electromagnetic and strong coupling constants, respectively, $F^{\mu\nu}$ the electromagnetic tensor field and $G_a^{\mu\nu}$ the color tensor fields. The b quark mass m_b introduced here corresponds to the mass calculated in the $\overline{\text{MS}}$ regularization scheme at the scale μ .

²Notice that both the coefficients and the operators depend on the scale μ chosen for the integration. Due to *asymptotic freedom*, short distance QCD effects (such as gluon exchanges with the virtual quarks) can be calculated in perturbation theory, if μ is not too small, and included in C_i ; these effects are the ones that govern the μ dependency of the coefficients. It is customary to choose μ of the order of the mass of the decaying quark, which is $\mathcal{O}(m_b)$ for B decays; since the effective Hamiltonian should be independent of this scale, the operators O_i must also carry some dependence, which cancels the one from the coefficients.

2.1.2 Effective Hamiltonian for the $B \rightarrow K\ell^+\ell^-$ Transitions

In order to perform calculations on observables such as branching fractions and decay rates, the matrix element $\langle H_{\text{eff}} \rangle$ must be calculated. This is done in three steps [39]:

1. The matching of the full theory to the effective theory at a scale μ_W of the order of the mass of the W boson. This allows for the extraction of the Wilson coefficients $C_i(m_q, \mu_W)$ to the desired order in α_s .
2. Making use of the renormalization group equations, evolving the Wilson coefficients down to the appropriate low energy scale μ of the order of the mass of the decaying quark (m_b for the case we are interested at), thus obtaining $C_i(m_q, \mu)$.
3. Calculation of the matrix elements $\langle \mathcal{O}_i(\mu) \rangle$ by means of some non-perturbative QCD method.

For the $B \rightarrow K\ell^+\ell^-$ decays, the leading Feynman diagrams of the full theory are shown in figure . Though we have included here the u and c quarks as mediators, the C_i , rather than depending on the quark masses m_i , are functions of the ratios $x_i = m_i/m_W$, so it is safe to ignore the contribution of the lighter quarks.

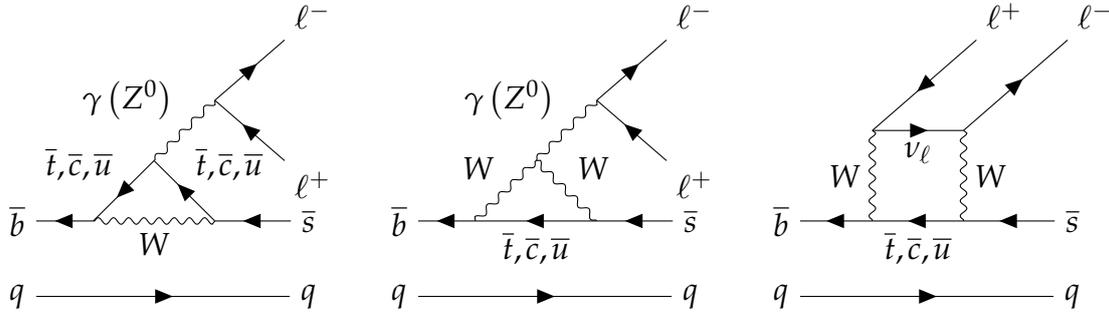


Fig. 2.1 Leading order diagrams for the $B \rightarrow K\ell^+\ell^-$ decay. The first two are known as penguin diagrams (by exercising the imagination it is possible to see the figure of this bird drawn by the diagram lines), whereas the last one is a box diagram. Notice that both a photon and a Z^0 can couple to the lepton pair in the penguin diagrams; moreover, albeit in the first diagram the coupling is depicted with the \bar{t} quark, the $\gamma(Z^0)$ can also be emitted from the other quarks. Here, q stands for either an up or a down quark and ℓ is either an electron or a muon.

At the scale μ_W , asymptotic freedom occurs and strong interactions can be treated perturbatively. The amplitudes of the full diagrams can be then computed to the

desired order in g_s , and made equal to the amplitudes calculated at the same order for the effective diagrams shown in figure 2.2.

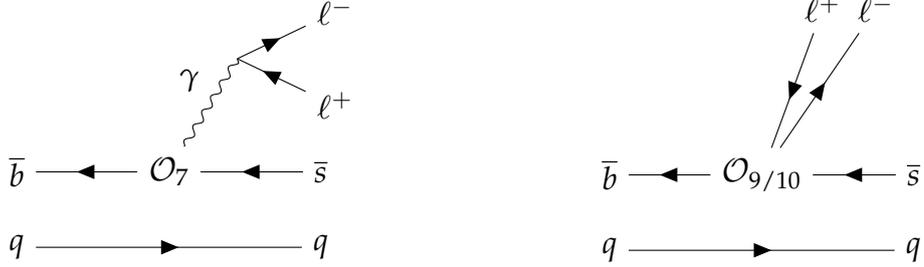


Fig. 2.2 Effective leading Feynman diagrams for the $B \rightarrow K\ell^+\ell^-$ decay.

By equating the effective and full amplitudes, it is possible to obtain an expression for the C_i at the μ_W scale. However, the decay occurs at a lower CM energy, so the operators are evolved accordingly, down to the correct energy scale, which is of the order of the mass of the decaying b quark. This is done by using the renormalization group equations (RGE), which correspond to differential equations that describe the change of renormalized quantities with respect to μ . The dependence of the Wilson coefficients $\vec{C} = [C_1, \dots, C_{10}]$ on μ is given by

$$\frac{d}{d \ln \mu} \vec{C}(\mu) = \gamma^T(\alpha_s) \vec{C}(\mu) \quad (2.3)$$

where $\gamma(\alpha_s)$ is known as the anomalous-dimension matrix, and $\alpha_s = g_s^2/4\pi$.

Thus, though in principle, from the group of ten operators given in equation 2.1 only \mathcal{O}_7 , \mathcal{O}_9 and \mathcal{O}_{10} contribute at tree level to the decay amplitude of the $B \rightarrow K\ell^+\ell^-$ processes (see figure 2.2), operators \mathcal{O}_{1-6} and \mathcal{O}_8 appear as one and two loop corrections [40] of the tree level effective diagrams.

The effective Hamiltonian is then written as

$$H_{\text{eff}} = \frac{4G_F}{\sqrt{2}} \lambda_t \left[C_7^{\text{eff}}(m_t, \mu_b) \mathcal{O}_7(\mu_b) + C_9^{\text{eff}}(m_t, \mu_b) \mathcal{O}_9(\mu_b) + C_{10}(m_t, \mu_b) \mathcal{O}_{10}(\mu_b) \right] \quad (2.4)$$

C_7^{eff} contains contributions from C_2 and C_8 , whereas C_9^{eff} includes those of C_{1-6} .

The most recent calculations of these coefficients include the next-to-next-to-leading order (NNLO) QCD corrections, and next-to-leading order QED corrections [14].

2.1.3 The Decay Rate

Making use of the expression for H_{eff} given in equation 2.4, the matrix element for the decay amplitude of the $B \rightarrow K\ell^+\ell^-$ process can be written as

$$\mathcal{M} = -\langle H_{\text{eff}} \rangle = \frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \left(\sum_{i=7,9,10} C_i \langle K\ell^+\ell^- | \mathcal{O}_i | B \rangle \right), \quad (2.5)$$

where, for brevity, $C_{7,9} = C_{7,9}^{\text{eff}}$ and we have omitted the dependence of C_i and \mathcal{O}_i on the scale μ . Replacing the operators by their explicit forms³ in equation 2.2,

$$\begin{aligned} \mathcal{M} = & \frac{G_F}{\sqrt{2}} \frac{\alpha_e}{2\pi} \lambda_t \langle K\ell^+\ell^- | \frac{-2im_b}{q^2} C_7 \left[\bar{s}\sigma^{\mu\nu} q_\nu (1 + \gamma_5) b\bar{\ell}\gamma_\mu \ell \right] + \\ & C_9 \left[\bar{s}\gamma_\mu (1 - \gamma_5) b\bar{\ell}\gamma^\mu \ell \right] + C_{10} \left[\bar{s}\gamma_\mu (1 - \gamma_5) b\bar{\ell}\gamma^\mu \gamma_5 \ell \right] | B \rangle. \end{aligned} \quad (2.6)$$

Due to the shape of the six-dimensional operators \mathcal{O}_i , the components $\langle K\ell^+\ell^- | \mathcal{O}_i | B \rangle$ can be factorized as $\langle \ell^+\ell^- | \mathcal{O}_i^{(\ell)} | 0 \rangle \langle K | \mathcal{O}_i^{(b \rightarrow s)} | B \rangle$, where $\mathcal{O}_i^{(\ell)}$ is the part of the operator \mathcal{O}_i corresponding to the leptonic currents, and $\mathcal{O}_i^{(b \rightarrow s)}$ the one corresponding to the hadronic ones. Moreover, these hadronic components can be written in terms of the $f_0(q^2)$, $f_+(q^2)$ and $f_T(q^2)$ form factors [27]:

$$\langle K | \bar{s}\gamma^\mu b | B \rangle = \frac{m_B^2 - m_K^2}{q^2} q^\mu (f_0 - f_+) + (p_B + p_K)^\mu f_+, \quad (2.7)$$

$$\langle K | \bar{s}\gamma^\mu \gamma_5 b | B \rangle = 0, \quad (2.8)$$

$$\langle K | \bar{s}\sigma^{\mu\nu} b | B \rangle = \frac{if_T}{m_B + m_K} \left[(p_B + p_K)^\mu q^\nu - q^\mu (p_B + p_K)^\nu \right]. \quad (2.9)$$

These identities, together with the fact that $\sigma^{\mu\nu} \gamma_5 = \frac{i}{2} \varepsilon^{\mu\nu\alpha\beta} \sigma_{\alpha\beta}$ (where $\varepsilon^{\mu\nu\alpha\beta}$ is the Levi-Civita symbol), completely determine the hadronic form factors involved in

³Notice that, in equation 2.2, \mathcal{O}_7 represents the emission of a *real* photon (hence the presence of the $F^{\mu\nu}$ tensor). However, due to spin conservation, the photon in the Feynman diagram of figure 2.2 is always a *virtual* one. Thus, $F^{\mu\nu}$ gets replaced by the term

$$-2i \frac{q_\nu}{q^2} \bar{\ell}\gamma_\mu \ell$$

where q is the momentum of the dilepton system.

the decay⁴. Hence:

$$\begin{aligned} \mathcal{M} = & \frac{G_F \alpha_e}{\sqrt{2} 2\pi} \lambda_t \left[\frac{2m_b}{q^2} \frac{f_T}{m_B + m_K} \left(q^2 (p_B + p_K)^\mu - (m_B^2 - m_K^2) q^\mu \right) C_7 [\bar{\ell} \gamma_\mu \ell] + \right. \\ & \left. + \left[\frac{m_B^2 - m_K^2}{q^2} q^\mu (f_0 - f_+) + (p_B + p_K)^\mu f_+ \right] \left(C_9 [\bar{\ell} \gamma_\mu \ell] + C_{10} [\bar{\ell} \gamma_\mu \gamma_5 \ell] \right) \right]. \end{aligned} \quad (2.10)$$

Defining

$$f_-(q^2) = \frac{m_B^2 - m_K^2}{q^2} [f_0(q^2) - f_+(q^2)] \quad (2.11)$$

the matrix element can be concisely written as

$$\mathcal{M} = \frac{G_F \alpha_e}{\sqrt{2} 2\pi} \lambda_t f_+ \left(\mathcal{A}_V^\mu [\bar{\ell} \gamma_\mu \ell] + \mathcal{A}_A^\mu [\bar{\ell} \gamma_\mu \gamma_5 \ell] \right) \quad (2.12)$$

with

$$\mathcal{A}_V^\mu = A (p_B + p_K)^\mu + B q^\mu, \quad \mathcal{A}_A^\mu = C (p_B + p_K)^\mu + D q^\mu \quad (2.13)$$

and

$$A = C_9 f_+ + \frac{2m_b}{m_B + m_K} C_7 f_T, \quad (2.14)$$

$$B = C_9 f_- - \frac{2m_b (m_B - m_K)}{q^2} C_7 f_-, \quad (2.15)$$

$$C = C_{10} f_+, \quad (2.16)$$

$$D = C_{10} f_-. \quad (2.17)$$

Due to the equation of motion for the lepton fields, the term in B does not contribute to the matrix element, and the term in D gets suppressed by one power of the lepton

⁴Indeed, one then has:

$$\langle K | \bar{s} \sigma^{\mu\nu} \gamma_5 b | B \rangle = -\frac{f_T}{m_B + m_K} \varepsilon^{\mu\nu\alpha\beta} (p_B + p_K)_\alpha q_\beta.$$

Notice, however, that when contracted with q_ν the previous term vanishes.

mass [41]. With this in mind, the decay rate is given by

$$\begin{aligned} \frac{d^2\Gamma_\ell}{dq^2 d\cos\theta_\ell} &= \frac{1}{4} \sum_{\text{spin}} |\mathcal{M}|^2 \\ &= \frac{G_F^2 \alpha_e^2}{2^{11} \pi^5} m_B \lambda_t^2 \left[a_\ell(q^2) + c_\ell(q^2) \cos^2\theta_\ell \right] \end{aligned} \quad (2.18)$$

where θ_ℓ is the angle between the direction of motion of the B meson and the negatively charged lepton ℓ^- in the dilepton center of mass frame (see figure 2.3), and

$$\beta_\ell(q^2) = \sqrt{1 - \frac{4m_\ell^2}{q^2}}, \quad (2.19)$$

$$\lambda(q^2) = m_B^4 + m_K^4 + q^4 - 2(m_B^2 m_K^2 + q^2 m_B^2 + q^2 m_K^2), \quad (2.20)$$

$$\begin{aligned} \frac{a_\ell(q^2)}{\beta_\ell(q^2) \sqrt{\lambda(q^2)}} &= \lambda(q^2) (|A|^2 + |C|^2) + |C|^2 4m_\ell^2 (2m_B^2 + 2m_K^2 - q^2) \\ &\quad + 8m_\ell^2 (m_B^2 - m_K^2 + q^2) \Re e(CD^*) + 4m_\ell^2 q^2 |D|^2, \end{aligned} \quad (2.21)$$

$$\frac{c_\ell(q^2)}{\beta_\ell(q^2) \sqrt{\lambda(q^2)}} = -\lambda(q^2) \left(1 - \frac{4m_\ell^2}{q^2}\right) (|A|^2 + |C|^2). \quad (2.22)$$

For the region of large hadronic recoil (where the K energy is high, and hence, $q^2 \ll m_B^2$), the Light Cone Sum Rules (LCSRs) technique [27] has been used to obtain expressions for the form factors f_+ , f_- and f_T . The results get then extrapolated to higher values of q^2 where they cannot be directly calculated. Such values are the ones used for the Monte Carlo simulations in this study.

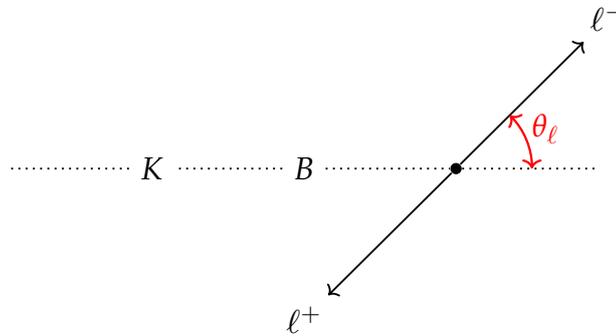


Fig. 2.3 The angle θ_ℓ .

Due to the limited statistics, sometimes an analysis over the double decay rate is not possible. In this case, partial decay rates over q^2 and $\cos\theta_\ell$ are useful. The first is given by

$$\frac{d\Gamma_\ell}{dq^2} = 2\Gamma_0 \left[a_\ell(q^2) + \frac{1}{3}c_\ell(q^2) \right] \quad (2.23)$$

where we have written

$$\Gamma_0 = \frac{G_F^2 \alpha_e^2}{2^{11} \pi^5} m_B \lambda_t^2, \quad (2.24)$$

and the second by

$$\frac{d\Gamma_\ell}{d\cos\theta_\ell} = \Gamma_0 \left[A_\ell + C_\ell \cos^2\theta_\ell \right] \quad (2.25)$$

with

$$A_\ell = \int_{q_{\min}^2}^{q_{\max}^2} dq^2 a_\ell(q)^2, \quad C_\ell = \int_{q_{\min}^2}^{q_{\max}^2} dq^2 c_\ell(q)^2. \quad (2.26)$$

Finally, the integrated decay rate is

$$\Gamma_\ell = 2\Gamma_0 \left(A_\ell + \frac{1}{3}C_\ell \right). \quad (2.27)$$

Clean observables are obtained by taking the ratio of calculated quantities, so uncertainties can cancel out. As shown in [38], for the multiple decay rates the lepton mass effects in the low q^2 region are of the order of $\mathcal{O}(m_\ell^4/q^4)$, which are about $m_\mu^4/q^4 \sim 10^{-4}$ for muons and $m_e^4/q^4 \sim 10^{-5}$ for electrons; hence, in the SM, the decay rates are independent of the lepton flavor. See, for example, figure 2.4.

The main sources of systematic uncertainties in the decay rates' calculations are the form factors, the value of V_{ts} and the renormalization scale μ ; however, at large hadronic recoil, symmetry relations constrain the ratios f_T/f_+ and f_-/f_+ , and corrections for these expressions of the order of $\mathcal{O}(\alpha_s, q^2/m_B^2)$ have been obtained [44], allowing to isolate the form factor uncertainties in f_+ . This indeed makes the ratio R_K , given by

$$R_K = \frac{\Gamma_\mu}{\Gamma_e} = \frac{A_\mu + C_\mu/3}{A_e + C_e/3} \quad (2.28)$$

extremely clean. As Γ_ℓ is practically independent of the lepton flavor, R_K is expected to be, to a very good precision, equal to one. The errors associated with f_+ and μ_b are similar in magnitude, and of the order of 10^{-4} , and the deviation of R_K from 1 is mainly due to the lepton mass effect in the decay rates [38]. This indeed shows that R_K is an excellent probe of physics beyond the SM, since a value far from unity would sign the presence of new interactions (operators) which couple

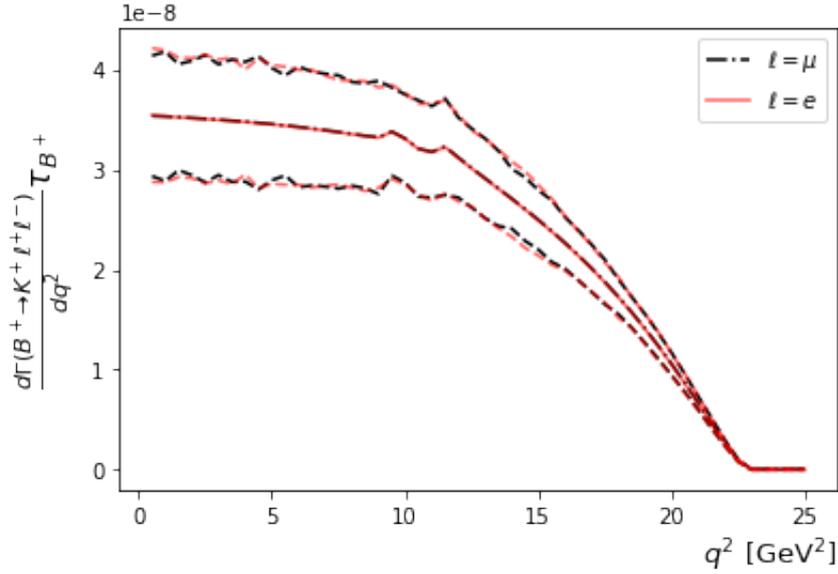


Fig. 2.4 Branching fraction —written as the decay ratio times the mean lifetime of the B^+ meson, τ_{B^+} , using the form factors provided in [42]. The mean value is given by the red, continuous (black, dash-dotted) line for the electron (muon) channel, and the theoretical errors by the dashed lines of the same color. Notice that the mean values are practically the same for both modes, and that the discrepancies in the theoretical uncertainties decrease towards the right side of the plot, showing that the lepton mass corrections get smaller with increasing values of q^2 . Image generated using *flavio* [43].

with differently (which is to say, with different Wilson coefficients) to muons and electrons⁵.

2.2 Strategy for this Study

This study aims to measure the total branching fraction of the $B \rightarrow K\ell^+\ell^-$ decays by reconstructing them from experimental data. Specifically, the decay channels of interest in this study are four⁶:

- $B^0 \rightarrow K^0 e^+ e^-$,
- $B^0 \rightarrow K^0 \mu^+ \mu^-$,

⁵On a different note, notice that the forward backward asymmetry is zero in the Standard Model; this is easily seen from equation 2.25. Hence, A_{FB} constitutes another clean probe for NP scenarios; it requires, however, a larger number of events than R_K , since the differential decay rate is needed.

⁶For the neutral modes, only the K_s^0 are reconstructed; moreover, K_s^0 is only reconstructed through its decay in a pair of oppositely charged pions; though this is its main decay channel (branching fraction close to 70%), this indeed reduces the reconstruction efficiency for these channels.

- $B^+ \rightarrow K^+ e^+ e^-$,
- $B^+ \rightarrow K^+ \mu^+ \mu^-$.

Unless otherwise stated, charge conjugation is implied throughout this study.

These decays are reconstructed from the data recorded by the Belle II detector, and we perform an Extended Unbinned Maximum Likelihood (EUML) fit to the beam constrained mass of the reconstructed B meson (see appendix J for more on the EUML fit method, and chapter 4 for an explanation on what the beam constrained mass is); from this fit we obtain the number of signal events in the sample $N_{\ell\ell}^i$ (where i indicates the B meson flavor and ℓ the lepton flavor), from which the branching fraction is calculated as

$$\mathcal{B}(B^i \rightarrow K^i \ell^+ \ell^-) = \frac{N_{\ell\ell}^i}{2\varepsilon_{\ell\ell}^i N_{B\bar{B}}^i}, \quad (2.29)$$

with $\varepsilon_{\ell\ell}^i$ the reconstruction efficiency of the decay mode, and $N_{B\bar{B}}^i$ the number of $B^i \bar{B}^i$ pairs produced.

On the other side, in order to reduce the total error reported, the experimental R_K is taken as the weighted average of the R_K in the charged ($B = B^+$) and neutral ($B = B^0$) modes, where the weights are inversely proportional to the square of the error in the measurement. For a single mode, the ratio R_K^i is

$$R_K^i = \frac{\varepsilon_{ee}^i N_{\mu\mu}^i}{\varepsilon_{\mu\mu}^i N_{ee}^i}, \quad (2.30)$$

and

$$R_K = \frac{(\Delta R_K^+ \Delta R_K^0)^2}{(\Delta R_K^+)^2 + (\Delta R_K^0)^2} \left(\frac{R_K^+}{(\Delta R_K^+)^2} + \frac{R_K^0}{(\Delta R_K^0)^2} \right), \quad (2.31)$$

where we have written the total error for the R_K^i measurement as ΔR_K^i .

In order to keep the systematic uncertainties arising from the fitting under control, we aim to reduce the background component as much as possible while maximizing the pseudo-significance through every step of the reconstruction process. We define the pseudo-significance of a data sample as

$$s = \frac{N_S}{\sqrt{N_S + N_B}} \quad (2.32)$$

where N_S is the number of signal events (this is, of $B^i \rightarrow K^i \ell^+ \ell^-$ decays) present in the sample, and N_B is the number of background events. In this way, we assure that the signal component can be extracted by a relatively simple fit, reducing the uncertainties related to it.

As a way to avoid experimenter bias, this analysis was performed in a blinded way: the reconstruction and fitting procedures were derived from simulations and from data which is not used in the final signal extraction, and only when this procedure was proven to be adequate on the simulated data, was it applied to the experimental one.

Chapter 3

Experimental Apparatus

This chapter deals with the experimental setup used for this study. The first section describes the SuperKEKB accelerator, where the B mesons are copiously produced in electron-positron collisions; the second one deals with the Belle II detector, a multipurpose mass spectrometer that allows for the reconstruction of the B meson decays with unprecedented precision; finally, the third section shortly describes the software framework used for the analysis of the data recorded by the detector.

3.1 The Accelerator: SuperKEKB

The SuperKEKB is located at the High Energy Research Organization Center close to Tsukuba-shi, one of the largest cities of Ibaraki prefecture, Japan. As an extended upgrade of its predecessor, the KEKB accelerator¹, the SuperKEKB is a *B-factory*, a facility where a large number of B mesons are produced in electron-positron collisions. These collisions occur at a center of mass energy equal to the invariant mass of the $\Upsilon(4S)$ resonance, $\sqrt{s} = 10.579$ GeV. This resonance decays 96% of the time into a pair of B mesons, and its mass is just above their production threshold; this implies that the B mesons are produced almost at rest in the center of mass frame. As studies on CP violation require the separation of the decay products of the two mesons produced, the SuperKEKB is an asymmetric collider, so that, in the laboratory frame, the two B 's would carry a momentum component along the

¹Which operated for more than a decade, between 1998 and 2010, and achieved the world's highest instantaneous luminosity ($2.11 \times 10^{34} \text{cm}^{-2}\text{s}^{-1}$), with a total integrated luminosity of 1.04ab^{-1} ; the collected data allowed for the corroboration of the Kobayashi-Maskawa theory for explaining the presence of CP violation in the Standard Model (see [B](#)).

beam axis which would allow for the separated reconstruction of both of them². It aims for an instantaneous luminosity record of $\sim 8 \times 10^{35} \text{cm}^{-2}\text{s}^{-1}$. The accelerator was commissioned for Phase 1 in early 2016, during which the behavior of single beams was studied; Phase 2 started in early 2018, during which most of the Belle II detector was assembled (the vertex detector was still missing) and the first collisions took place. Currently, the Phase 3, which began in March of 2019, contains the first collision events with all the components of the Belle II detector installed.

Figure 3.1 depicts an overview of the accelerator complex. The electron bunches are produced at the start of the linear accelerator (linac) and taken up to an energy of 4 GeV. One in every two bunches is used to produce positrons by collisions with a tungsten target [45]. After their generation, the positrons go through a 1 GeV damping ring used to reduce their emittance (their spatial and energy spread); finally, both electrons and positrons are accelerated to their final energies (7 GeV and 4 GeV, respectively) in the second half of the LINAC, and are then injected into their respective storage rings.

Once in the storage rings, the electron and positron bunches will meet at the interaction point (IP) inside the Belle II detector at frequency of 250 millions of collisions per second. In order to achieve such a high luminosity, multiple modifications took place: the beam currents were increased, and the beam vertical sizes at the IP are squeezed down to a value of 59 nm, which in turn requires the reduction of the effective longitudinal beam size (this is, the length of the beam portion which is crossing with the other beam at a given instant at the IP), due to the Hourglass effect³. One option to alleviate this was to reduce the bunch longitudinal size, but this would have increased the energy spread of the beams⁴. The implemented solution is known as the nano-beam scheme, first proposed for the Super B factory

²If the collider was symmetric, both mesons would decay at almost the same position in the detector region. Since the B lifetime is around 1.5 to 1.7 ps, the determination of Δt , the difference in the decay times of both mesons—which is of primary importance in CP violation studies—, would demand an extremely precise time resolution. On the other side, in an asymmetric collider, the knowledge of Δz , the difference in the decay vertex position of each meson in the laboratory frame, allows for the determination of $\Delta t = \Delta z / \beta\gamma$. This requires an accurate measurement of the decay vertexes, which is still challenging, but more feasible in comparison with the symmetric collider case; this is also the main reason why the Belle II detector is equipped with an excellent vertex detector.

³The Hourglass effect is the parabolic increase of the β function, and hence the linear increase of the beam size, for points away from the IP (assuming the IP is the point of minimum beam size). If the effective interaction length is too large, contributions to the luminosity from such points gets reduced, since collisions happen outside the detector acceptance.

⁴The energy spread measures how much the energy of single particles in the bunch deviate from the nominal beam energy. Due to Liouville's theorem, the phase space density must remain constant for a system which obeys conservation of energy. If the spatial portion of the phase space is reduced, the momentum portion must increase. Thus decreasing the beam size increases the energy spread.

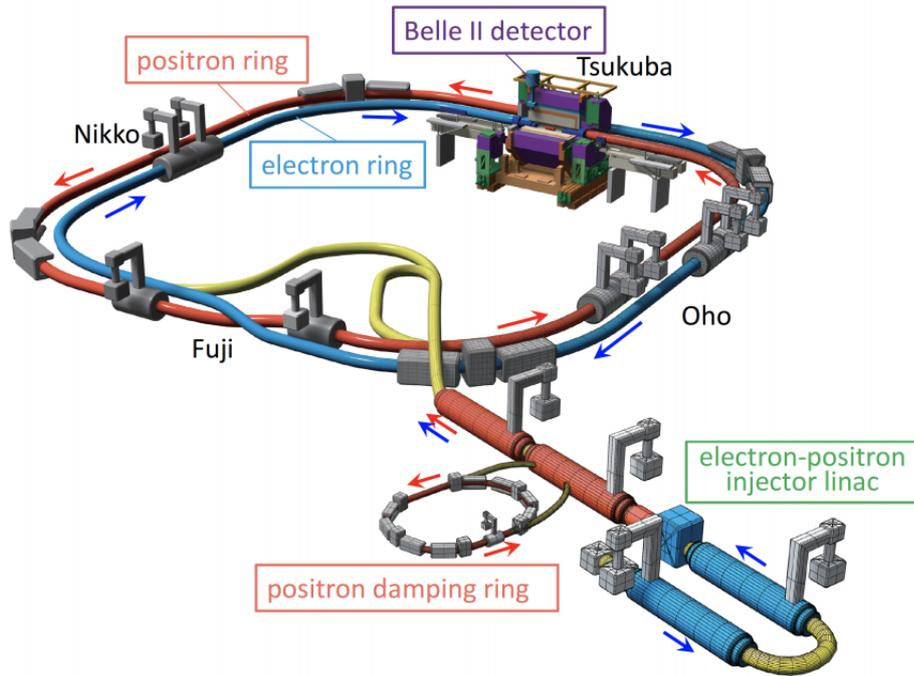


Fig. 3.1 Schematic view of the SuperKEKB collider. The interaction point, where the Belle II detector is located, is in the Tsukuba section. Original image taken from [32].

project in Italy [46], where the horizontal beam size at the IP gets squeezed down to $7.75 \mu\text{m}$ and the effective bunch length is reduced by making the beams meet at a relatively large crossing angle of 41.5 mrad .

Table 3.1 shows a comparison of the main parameters of the KEKB and SuperKEKB.

3.2 The Detector: Belle II

As the accelerator facility it belongs, the Belle II detector is also an upgrade of its predecessor, the successful Belle detector. Both are large solid-angle magnetic spectrometers, installed at the accelerators IP, where they are supposed to measure the properties of the final state particles produced by the electron-positron collisions⁵;

⁵with the exception of neutrinos, which escape the detector volume without leaving any trace; these are usually reconstructed by calculating the difference between the initial energy of the beams and the total energy of the detected particles. This is one of the advantages of B factories in comparison with high energy accelerators like the LHC, where the multiplicity of the physics events makes it impossible to account for the whole detectable final state particles.

Table 3.1. KEKB vs. SuperKEKB parameters [32]

Parameter	KEKB		SuperKEKB	
	LER	HER	LER	HER
Beam Energy [GeV]	3.5	8.0	4.0	7.007
Beam pipe radius at IP [mm]	20.5		10	
Circumference [m]	3016.262		3016.262	
Half crossing angle [mrad]	0		41.5	
Beta function at IP (β_x^*/β_y^*) [mm]	1200/5.9	1200/5.9	32/0.7	25/0.30
Horizontal beam size [μm]	147	170	10.1	10.7
Vertical beam size [nm]	940	940	48	62
Energy spread [$\times 10^{-4}$]	7.3	6.7	7.92	6.37
Beam current [A]	1.64	1.19	4.60	2.60
Number of bunches	1584		2500	
Bunch length [mm]	7	7	6.0	5.0
Instant luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	2.108×10^{34}		8×10^{35}	
Integrated luminosity [ab^{-1}]	1.041		50	

they both consist of multiple layers of subdetectors, each with different tasks, surrounding the interaction region. A schematic of the Belle II detector is shown in figure 3.2. It can be roughly divided in the next components:

- A tracking system, composed of a vertex detector (VXD) and a central drift chamber (CDC). The first is divided in the innermost pixel detector (PXD), and the four layered strip vertex detector (SVD).
- A Time of Propagation (TOP) counter in the barrel region, and a proximity-focusing aerogel ring imaging Čerenkov (ARICH) detector in the forward endcap region, constitute the particle identification (PID) system.
- An electromagnetic calorimeter (ECL) used for measuring the energy of electrons and photons.
- A muon and K_L detection system (KLM). It consists of layers of RPCs (resistive plate chambers) in the barrel region and scintillator strips in the endcaps. It

also works as a return yoke for the magnetic field used to bend the trajectories of charged particles (for momentum measurement and PID purposes).

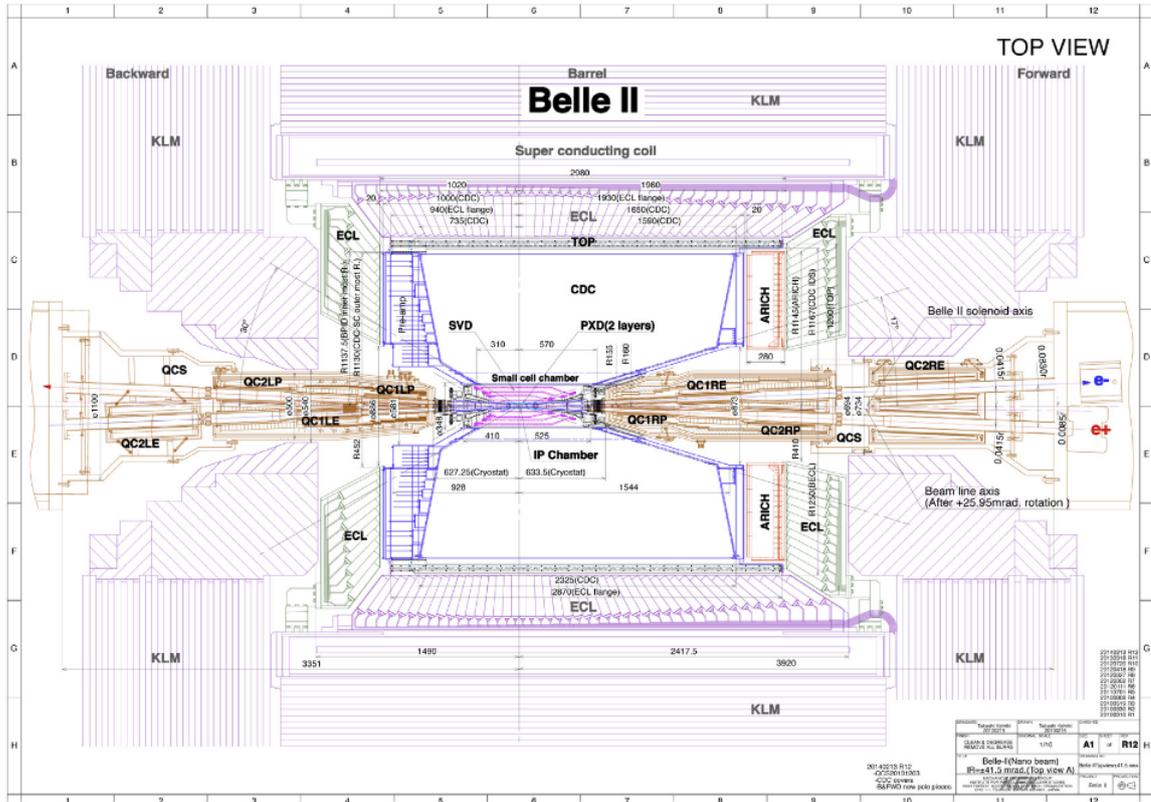


Fig. 3.2 Top view of the Belle II detector. The asymmetry in its design is evident from the picture; most of the detector active region is placed towards the forward end of the geometry (which corresponds to the direction of motion of the high energy electron beam), since the decay products are boosted in its direction. The focusing devices are also visible surrounding the beam pipes outside the detector, as well as the superconducting solenoid, in charge of producing the magnetic field used to bend the path of charged particles. Original image taken from [47].

3.2.1 VXD

The change in the beam energies reported in table 3.1 leads to a smaller Lorentz boost factor, which in turns translates into a shorter decay length in the laboratory frame for the B mesons. Since the correct location of the decaying vertices is of utmost importance for CP violation studies, the vertex detector resolution was consequently improved. It now consists of two devices, the Silicon Pixel Detector (PXD) and the

Silicon Vertex Detector (SVD), which in total comprise six layers around the beam pipe. The two innermost layers, at radii of 14 mm and 22 mm, make up the PXD; the remaining four, which are double-sided silicon strip ladders, at radii of 39 mm, 80 mm, 104 mm and 135 mm, constitute the SVD. In comparison, the innermost layer of the Belle SVD (it did not have a PXD detector) was positioned at 30 mm from the interaction point, and its outermost at 60.5 mm. A schematic of the VXD components is shown in figure 3.3.

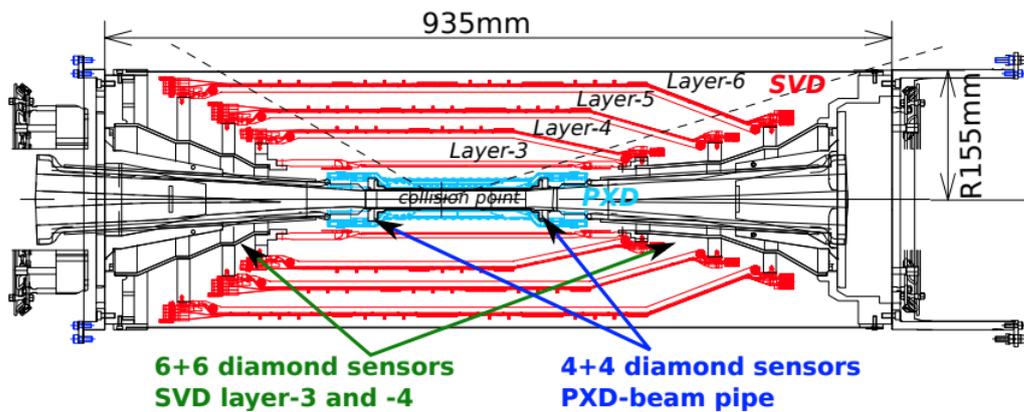


Fig. 3.3 Cross-sectional view of the VXD detector. The PXD layers are in blue, whereas the SVD ones are in red. The locations of some of the diamond sensors, used for beam background monitoring, are depicted too. Notice also that layers 4, 5 and 6 are slightly slanted towards the forward region, which increases the spatial precision in the boost direction. Original image taken from [48].

The PXD layers are matrices of DEPFET pixels, field effect transistors (FET) implanted on fully depleted n-type silicon; interactions between the substrate and incoming particles would excite the electrons on the first, which are —by adding a deep n-dopping implantation— then collected just under the FET gate, and used to modulate the current on the transistor. In order not to disturb the tracking detectors outside the VXD, multiple Coulomb scattering is avoided by keeping the material budget at very low levels: the thickness of each pixel is $75 \mu\text{m}$, which complies with the original design requirement of a maximum thickness of $0.2\%X_0$ ⁶ [47]. The first layer consists of 8 planar sensors, with a width of 15 mm and a sensitive length of 90 mm; the second one contains 12 modules with a width of 15mm and a sensitive length of 123 mm [49]. A total of 7.7 million channels are employed. Given the large

⁶ X_0 corresponds to the radiation length of the material: it is the mean distance over which a high-energy electron loses all but $1/e$ of its energy through Bremsstrahlung radiation, and $7/9$ of the mean free path for pair production for a high-energy photon [15].

number of channels, it is not possible to read all of them within the trigger latency; instead, only pixels within a region of interest (ROI), determined by extrapolation of the SVD and CDC track to the PXD volume, are readout, and used in the final event reconstruction.

The SVD, on the other side, consists of four layers of double-sided silicon strip detectors (DSSD) ladders in a windmill-like arrangement. The average material budget of each layer is about 0.7% of the radiation length, cooling pipes included [48]. The p-strips in the sensor run along the longitudinal (beam-axis) direction, and the n-strips in the transverse one. Similar to the PXD operation scheme, each ladder consists of n-type silicon which, by interactions with the impinging particles, produce electron-hole pairs. The electrons are collected in the p-strips, whereas the holes move towards the n-strips. The signal on both strips is used to locate the hit in the subdetector volume. In order to minimize the capacitive noise, the readout electronics are situated as close to the sensors as possible: for the DSSD on both ends of the ladders, the readout chips are located on the far end of the detector geometry, outside its acceptance. The chips for sensors in the middle of the arrangement are, on the other hand, placed upon them; since these chips are within the detector acceptance, they are thinned down to $100\ \mu\text{m}$ in order to reduce the material budget. The total number of channels is 224×10^3 .

The use of read DSSD instead of pixels is justified due to the financial cost and the readout complexity the later would carry. However, use of DSSD can lead to increments in the detector occupancy and to a detriment in its signal-to-noise ratio due to the emergence of ghost hits, which correspond to wrong combinations of n-strip and p-strip signals when more than one hit are registered in the same readout window⁷.

3.2.2 CDC

The 1.5 T magnetic field produced by the superconducting solenoid, which flows along the beam axis towards the forward endcap, defines the z axis of the detector geometry. The curvature of the tracks left by charged particles in the tracking system in the xy plane gives information on their transverse momentum. Using the measured angle between the track and the z axis, the full momentum can be calculated. This requires a precise measurement of the particle path, which is done with the aid of the central drift chamber.

⁷For n hits, at most n p-strips and n n-strips are activated. Hence, there are n^2 possible combinations for the hits coordinates, $n(n-1)$ of which are not real —this is, ghost— hits.

The CDC is a multi-wire proportional chamber, with a mixture of Helium and Ethane in equal amounts as active component. An electric field between the field wires and the sense wires drives the electrons (produced from the ionization of the active component by the incoming particle) towards the later ones⁸. By measuring the drift time of these electrons it is possible to determine the distance from the wire at which the particle passed. This provides the azimuthal angle of the particle position at the moment of ionization. Since the wires are located at constant values of r , this coordinate is also known. To obtain the z coordinate of the particle position, stereo wires are introduced: these correspond to wires with both end points rotated by an stereo angle α with respect to the z axis. The azimuthal wire position becomes a linear function and its radial coordinate a quadratic function of z . A comparison between the CDC of the Belle and Belle II detectors is shown in figure 3.4. Owing to the new, smaller PID detector in the barrel region, and in order to cope with the higher luminosity, the Belle II CDC covers a larger radius, and the number of wires is roughly twice as much as for the Belle case.

Most of the layers, except for the ones in the first super-layer, consist of normal cells, also shown in figure 3.4; the 8 wire layers' group closest to the IP consists of what are known as small cells: the tightly spaced wires help to mitigate the increase in the occupancy levels due to the higher background rates. It covers the polar angle region between 17° and 150° .

3.2.3 PID Detectors

The two main sub-detectors used for particle identification, the TOP and ARICH, are based on the fact that particles moving faster than light in a highly refractive medium emit photons in a direction determined by the particle velocity; this is known as Čerenkov radiation. This radiation consists of a cone around the momentum vector of the particle, with an angle θ given by the expression

$$\cos\theta = \frac{1}{\beta n} \quad (3.1)$$

where $\beta = |v|$ and n is the refractive index of the medium.

The TOP detector, which takes care of particle identification on the barrel region (covers the polar angle region between 32° and 120°), consists of 16 radiators, each

⁸The ions produced together with the electron have much smaller drift velocities, and they recombine before they reach any measuring device.

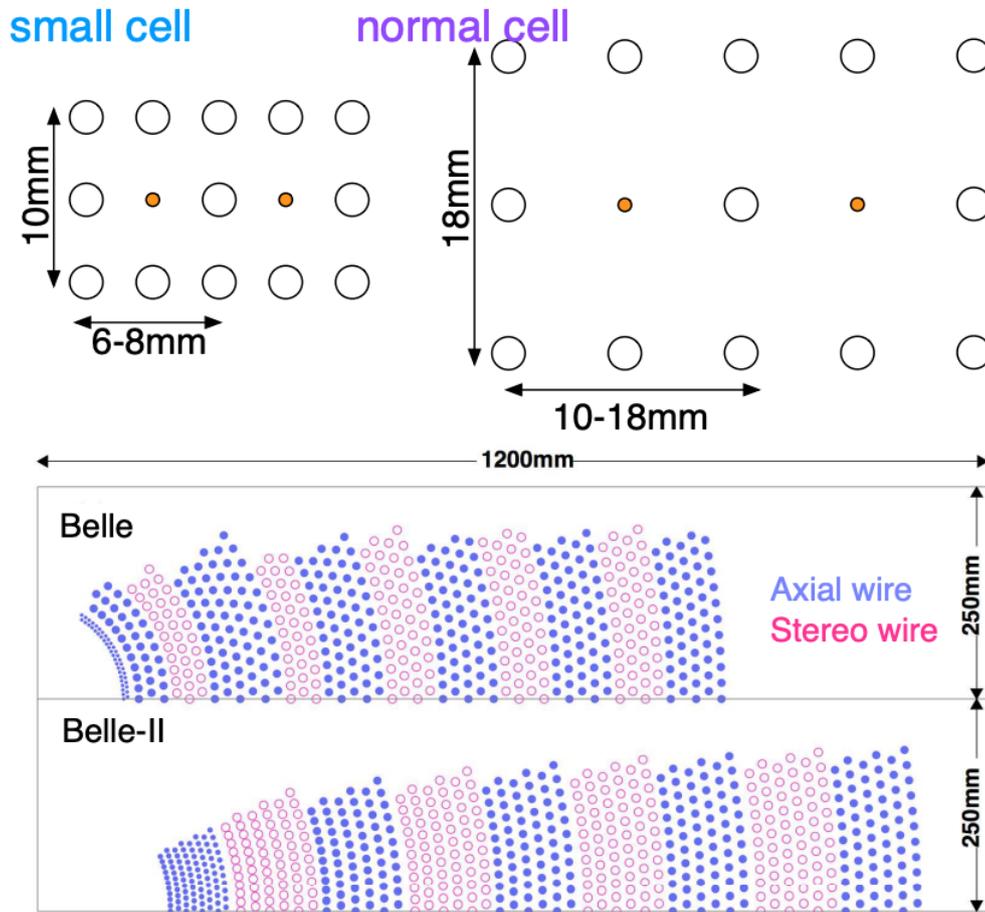


Fig. 3.4 Front view of the CDC for the Belle and Belle II detectors. Layers of wires are collected in groups of six and denoted as super-layers (the first super-layer, however, consists of 8 wire layers). Axial and stereo super-layers intercalate across the radial axis. Subsequent stereo super-layers have stereo angles of opposite sign. Original image taken from [50].

of one made of two quartz bars (acting as the highly refractive medium), a mirror (at the forward end of the bar, to drive the photons back to the readout system) and a small expansion prism (see figure 3.5). The photons produced in the quartz go through total internal reflections all the way to the back end of the bars, where they get registered by an array of micro-channel plate photomultipliers (MCP-PMTs). The position of the photons in the detection array and their arrival time is used to discriminate between different particle hypotheses (see the next chapter for more on particle ID).

Particle identification is performed by measuring the time of flight of the incident radiation —by using the Čerenkov angle expression from equation 3.1— and by

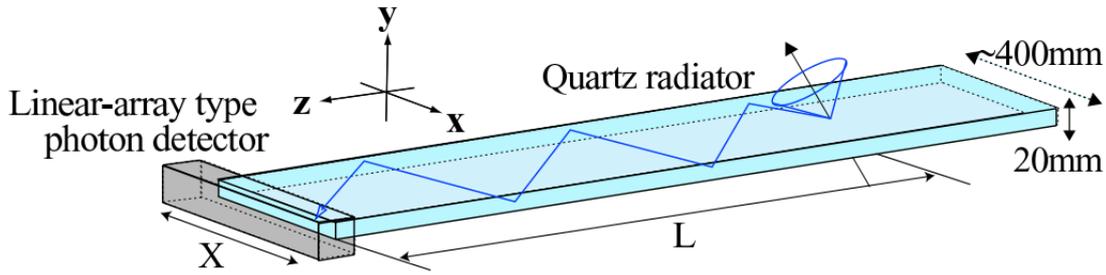


Fig. 3.5 Conceptual design for one of the TOP detector modules. Original image taken from [51].

using the momentum p and flight length L information from the CDC, through the relation

$$T = \frac{L}{\beta} = nL \cos \theta = L \sqrt{1 + \left(\frac{m}{p}\right)^2}. \quad (3.2)$$

On the forward cap (polar angle between 17° and 35°) the ARICH detector is located. It is a proximity ring imaging Čerenkov detector; it consists of two-layered tiles of hydrophobic silica aerogel with different refractive indices (1.045 for the first layer and 1.055 for the second) which are used as radiators, and the photons emitted are collected in the hybrid avalanche photo-detectors (HAPD), located 20 cm apart from the tiles (see figure 3.6).

The Čerenkov angle resolution per track improves with the number of photons detected. In order to increase the emission of Čerenkov photons (and thus the number of photons detected), without increasing the thickness of the aerogel (which would degrade the single photon Čerenkov angle resolution due to the increased uncertainty in the photon emission point), two aerogel radiators with different refractive indices are used. The one with the lower refractive index is located upstream, since it would deliver photons with a smaller angle. If the index of the second radiator is well adjusted, the corresponding two rings overlap completely, increasing the number of photons without coarsening the thickness of the final ring.

The Belle II ARICH detector consists of 248 wedge-shaped $17 \times 17 \text{ cm}^2$ tiles, consisting of two aerogel layers, each one of which is 2 cm thick. The photon detection is performed by 420 HAPDs, with 144 channels each.

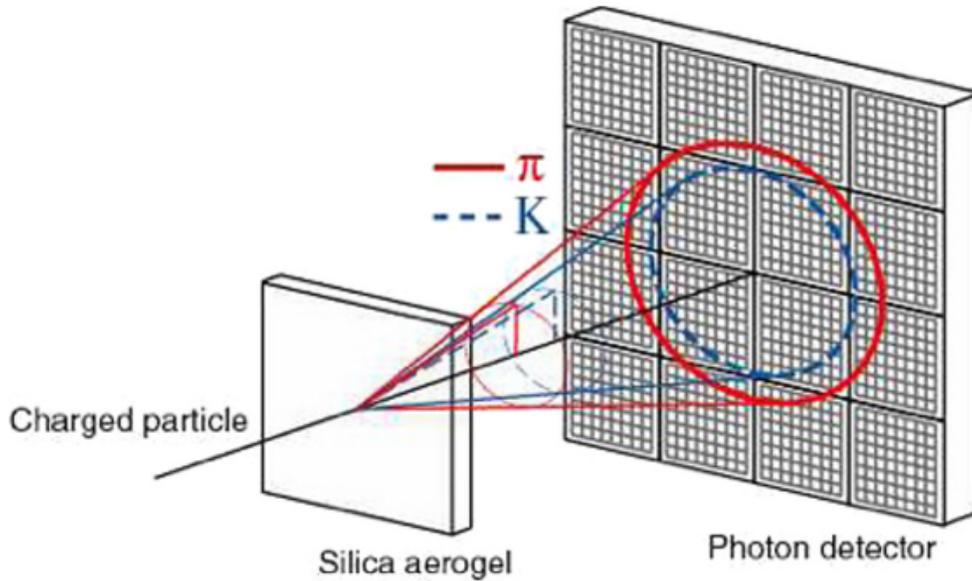


Fig. 3.6 Basic principle of the ARICH detector. The 2cm thick aerogel emits Čerenkov radiation at an angle given by equation 3.1. The photons travel across the 20 cm gap between the gel and the HAPD, where they describe a ring whose radius depends on the original emission angle and the gap distance. Particles of different flavor with the same momentum will produce rings of different sizes. Original image taken from [52].

3.2.4 ECL

The Belle II electromagnetic calorimeter is the same as the one used for the Belle experiment, with an improved readout system, in order to cope with the higher event rate. It is composed of Thallium activated Cesium Iodide —CsI(Tl)— scintillation crystals, with a truncated pyramidal shape; their average cross section is $60 \text{ mm} \times 60 \text{ mm}$, and have a thickness of 300 mm (in the barrel region), which corresponds to $16.2X_0$. The total length of the detector in the barrel region is 3 m, with an inner radius of 1.25 m; its polar angle coverage ranges from $\theta = 32.2^\circ$ to $\theta = 128.7^\circ$. The annular endcaps (in green in figure 3.2) are located at $z = 2 \text{ m}$ (forward endcap, with angular coverage between 12.4° and 31.4°) and $z = -1 \text{ m}$ (backward endcap, angular coverage between 130.7° and 155.1°) from the IP. Notice there are two gaps between the barrel and the endcaps, of approximately 1° , which are not covered. In total, the ECL covers about 90% of the solid angle in the center of mass system.

Two photodiodes, each one with a sensitive area of $10 \times 20 \text{ mm}^2$, are attached at the back of the crystals. The signal yield from these photodiodes is then fed to a charge-sensitive amplifier, which is mounted on the crystal as well.

Due to the increased background rates, the relatively long scintillation decay time of the CsI(Tl) (around $1 \mu\text{s}$) increases the overlapping of pulses from neighboring events. In order to cope with this, new readout electronics were implemented, which reduced the shaping time from 1 to $0.5 \mu\text{s}$, and use a pipe-line readout with waveform analysis [53].

3.2.5 KLM

The outermost subdetector is the K-long and muon calorimeter. It consists of layers of active detector elements interleaved with 4.7 cm iron plates, used as return yokes for the 1.5 T magnetic field produced by the solenoid. It covers the region between $\theta = 20^\circ$ and $\theta = 155^\circ$. In the barrel region, the active layers are the same used for the Belle experiment, consisting of two glass-electrode resistive plate chambers (RPCs) stacked on top of each other, with a thickness of around 32 mm (see figure 3.7a). The area of each module ranges from $2.2 \times 1.5 \text{ m}^2$ to $2.2 \times 2.7 \text{ m}^2$, except near the solenoid's helium chimney, where they are shorter by 0.63 m. An incoming particle ionizes the gas molecules along its path, and the electron (ion pairs) are accelerated by the electric field towards the anode (cathode) of the RPC. Due to the electric field strength, the primary electrons further ionize the gas molecules, leading to a streamer between the electrodes. This streamer is imaged on a plane of external pickup strips, each with a width of about 5cm. These pickup strips are arranged in two orthogonal planes located on each side of the RPCs stack, and give information on the (ϕ, z) coordinates of the particle [54].

Notice that the RPC carries an intrinsic dead time associated with the recovery of its electric field near a discharge; hence, the particle detection efficiency for the KLM depends on the background rate (the higher the background, the higher the number of inoperative modules, thus reducing the sensitivity of the detector). Due to the expected increase in the neutron rate for the Belle II detector, the efficiency in the endcaps is expected to drop to zero [54]. Hence, the two innermost active layers of the barrel region, as well as the totality of the active layers in the endcaps (14 layers each), consist of scintillators strips with wavelength shifting (WLS) fibers (see figure 3.7b). The produced photons are then readout by silicon photomultipliers, which are used instead of common PMT due to the spatial limitations and the presence of a strong magnetic field [55].

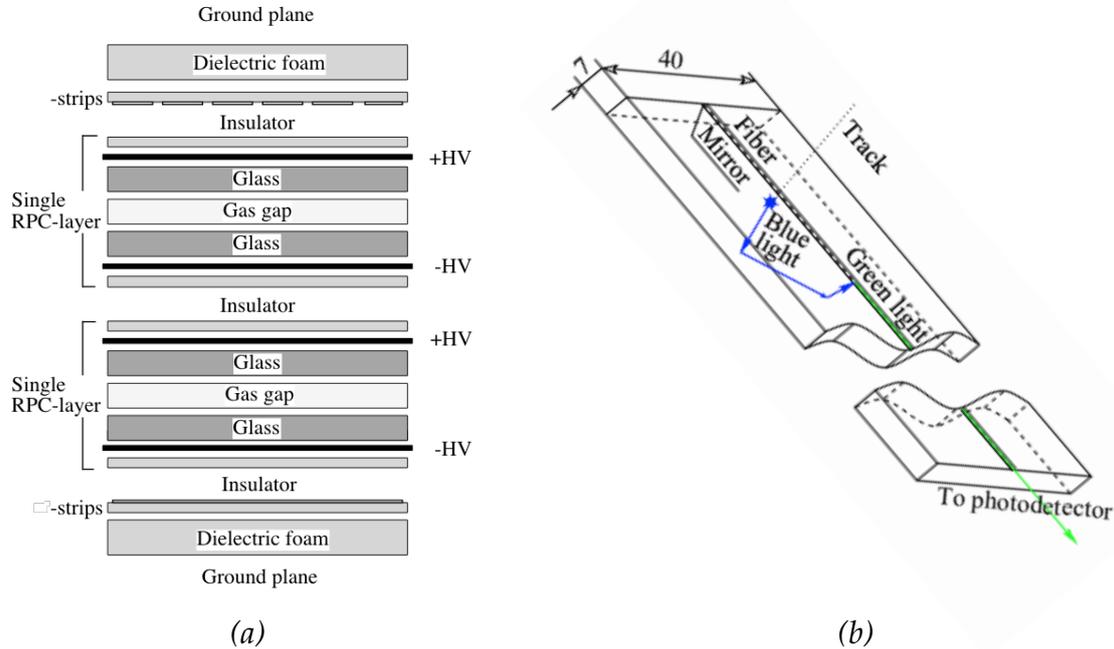


Fig. 3.7 (a) Active detector for the outermost layers of the KLM in the barrel region. The gas is a mixture of 62% HFC-134a, 30% argon, and 8% butane-silver (the later being a mixture of 70% *n*-butane and 30% isobutane). The glass is made of silicon dioxide (73%), sodium oxide (13%), calcium oxide (9%) and uncategorized compounds (4%), and has a thickness of 2.4 mm. (b) Light detection in the scintillator strip for the endcaps and innermost layers of the barrel components of the KLM. Original figures taken from [54].

3.2.6 Triggering and DAQ

Most of the events resulting from the e^+e^- collisions at the SuperKEKB will not correspond to the production of the $Y(4S)$ resonance. Indeed, table 3.2 shows the cross section at a center of mass energy of 10.58 GeV for several physics processes, as well as their occurrence rate, for the goal luminosity of $8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$.

As such, a trigger logic is required, able to select hadronic events with high efficiencies, with an average rate of no more than 30 kHz, good timing precision (maximum 10 ns) and event separation (minimum 200 ns) and with a fixed latency of about $5 \mu\text{s}$ [45]. Moreover, it should trigger on neutral exotic physics signatures such as $e^+e^- \rightarrow \gamma$ or $e^+e^- \rightarrow 3\gamma$, in the presence of large QED backgrounds. This is achieved by an improvement over the Belle triggering scheme, which consists of two trigger logics: the L1 trigger and the High Level Trigger.

The L1 trigger is a hardware based trigger, which consists of sub-trigger systems associated to each of the subdetectors (excluding the VXD), whose outputs get

Table 3.2. Total cross section and production rates of various physics processes at the SuperKEKB for a luminosity of $8 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$ [56]

Physics process	Cross section [nb]	Rate [Hz]
$\Upsilon(4S) \rightarrow B\bar{B}$	1.05	840
$e^+e^- \rightarrow q\bar{q}$	3.69	2952
$e^+e^- \rightarrow \mu^+\mu^-$	1.148	918
$e^+e^- \rightarrow \tau^+\tau^-$	0.919	735
Bhabha	74.4 (after cuts)	59.52×10^3
$e^+e^- \rightarrow \gamma\gamma$	3.3 (after cuts)	2640

combined in the Global Decision Logic, from which the final trigger decision is made (see figure 3.8). Besides the new information provided by each system, for Belle II all components and connections were replaced with new technologies: each sub-trigger performs its trigger logic by means of field programmable gate arrays (FPGAs), and all data flows through high speed serial links, which enables to deliver high amounts of data to one FPGA. The L1 trigger latency is about $5 \mu\text{s}$.

Once the L1 trigger signal is obtained (which also contains the type of event that activated the trigger), the data from each subdetector —except for the PXD— is transferred from their front-end electronics to the COPPER (Common Pipeline Platform for Electronics Readout) CPU boards, using the Belle2Link connection [57]. Data transfer from the SVD is the most critical one, since it takes 26.4 (13.2) μs to send the data consisting of 6 (3) signal samples downstream, which is more than five times the required latency. In order to alleviate this, each front-end electronics board is equipped with a buffer, which stores both the detector data and the trigger decision for the event. The PXD requires 20 μs to read out the entire event in one frame, where a frame contains all the hits in the past 20 μs , and it is expected that these hits correspond to two or more triggers with a high probability. As long as the occupancy of the PXD is kept below certain threshold, the readout from this subdetector does not introduce any dead time.

The data in the COPPER system is then transferred to event builder switches, where a first reconstruction of the event is performed, and from it to the High Level Trigger (HLT) logic through an ethernet connection. The HLT consists of more than 5000 CPUs, dedicated to software triggering using the information from the partially reconstructed events. The data flow is reduced to 2/3 of its size, owing to

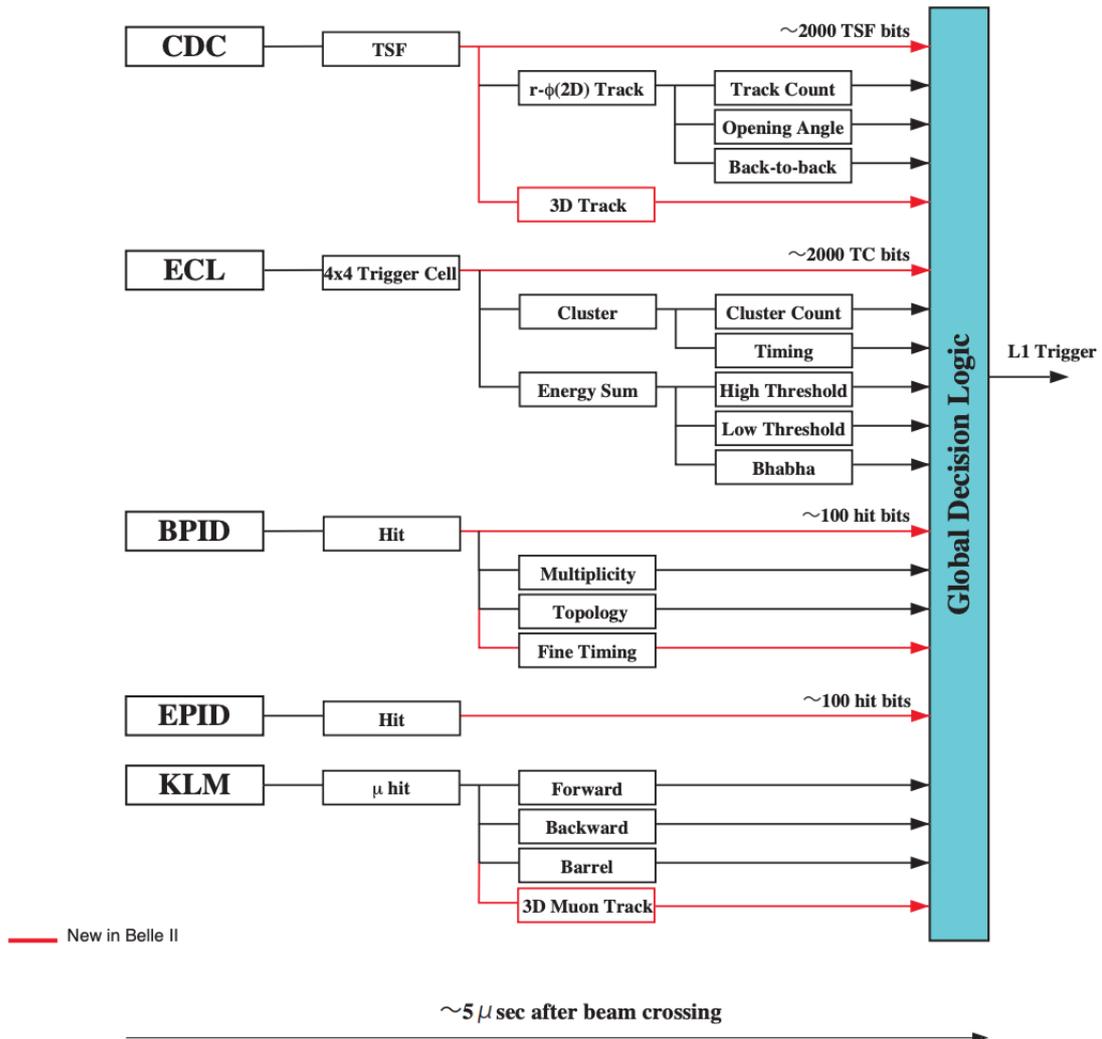


Fig. 3.8 Schematic of the L1 trigger system for the Belle II detector. Each subdetector system sends its own output to the Global Decision Logic (GDL). Here, the final trigger decision is made. Lines in red are information that was not used during the Belle experiment, and that was added to the system of the new detector. BPID corresponds to the PID subdetector in the barrel region (TOP), and EPID is that in the endcap (ARICH). Original image taken from [45].

a reduction of 1/3 in the number of events, and the duplication of the information available for each event, since it now includes the reconstruction results as well. The HLT also makes use of the tracking information from the CDC and the SVD to derive the Region Of Interest in the PXD, and provides it to the Online Selector Node (ONSEN). The ROI allows to reduce the data rate from the PXD to 1/10 of its initial value; together with the HLT event selection, the total reduction is of 1/30.

The ONSEN and HLT data are finally sent to a storage unit which is physically next to the HLT unit, copied later to the KEK computing center (KEKCC) and then distributed to the grid⁹. The whole process is depicted in the box diagram of figure 3.9.

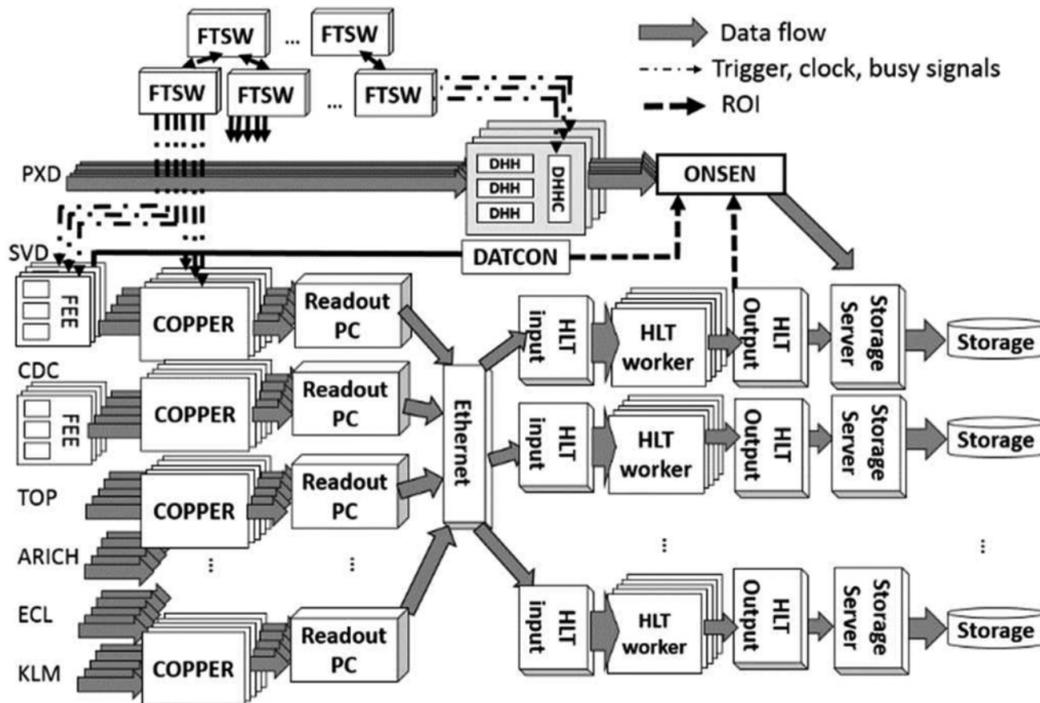


Fig. 3.9 DAQ system for the Belle II experiment. Original image taken from [58].

3.3 The Analysis Framework: basf2

After being stored at the KEKCC, the data is sent for offline reconstruction, using more detailed algorithms. This reconstruction takes into account the detector and accelerator conditions during the data acquisition, which are determined by using recorded physics processes with a distinctive signature and a well-known topology, that allow for the calculation of the multiple calibration and alignment variables.

⁹The grid corresponds to a computer network infrastructure employing the resources of the 25 Belle II member nations. Originally developed in the LHC collaboration, it is designed to handle large amounts of data by distributing it between the multiple computing centers available.

These calculations are done regularly, and their results are stored into a database, which is later accessed whenever a physics analysis is performed.

Once the offline reconstruction is completed, its output, in the form of an mDST (mini data summary table, a Belle II file format based on ROOT) file, is made available for the collaboration members. These files are organized into a group of runs with different duration, each containing a sequence of independent events. An event corresponds to the measurements of the by-products of the initial e^+e^- collisions, or of cosmic rays. A group of runs with similar detector conditions is known as an experiment. Experiments, runs and events are all labeled by integers [59].

The study of this data is carried out using the Belle II analysis software framework (basf2); this same framework is used for the data acquisition, reconstruction and Monte Carlo simulations. It consists of a series of C++ *modules* loaded dynamically which are executed sequentially over the mDST files. The configuration of such modules is done with steering files written in Python, and its parallel processing (over the events) feature exploits the potential of multi-core CPUs [47]. Multiple external libraries are able in the framework; examples are EvtGen [60] and Geant4 [61] for the Monte Carlo data generation, and ROOT [62] for the data handling.

A large number of simulated events is produced in what are known as Monte Carlo campaigns, in order to reduce the discrepancies that arise between analyses using different tools for generating their own data, and to allow for the reproducibility of the analysis results by any member of the collaboration. Whenever a detector configuration changes, the data production system evolves, the analysis framework improves, or the understanding of the background conditions gets better, new campaigns are conducted in order to have simulated data as close as possible to the real one. So far there have been 13 Monte Carlo campaigns, from which the latest 3 correspond to the configuration of Phase III. In the following chapters, analysis performed on simulated data will explicitly specify the campaign to which the data belongs to.

Chapter 4

Event Reconstruction and Background Analysis

Once Belle II has stored the relevant information for a physics event, it is up to the analyst to interpret it in a reliable and systematic way in order to determine the nature of the process involved. This is known as the *event reconstruction* step. To make sure that the reconstructed events correspond to those of interest in this study, a group of selection criteria is applied to the data of each event recorded by the detector, and only events passing these cuts are considered signal candidates. However, it is possible for background events to make it through these cuts, and for signal events to fail them. To assess these sources of error, we perform Monte Carlo studies in which the retention rate of signal events or *efficiency*, and the rejection power on background events or *purity* of the reconstruction are measured. The efficiency and purity are later used as scale factors when extracting the number of signal and background events from the true data, and the events that make it through the selection criteria in the Monte Carlo study are used to model the signal and background shapes.

This chapter describes the reconstruction process and the selection cuts employed. It is divided in three parts: in the first one, the simulated datasets used for the study are presented; the second part deals with the event selection for the decays of interest, which comprises four steps: skimming, selection of final state particles (FSP), reconstruction of intermediate composite particles, and reconstruction of B mesons; here we also present the new Bremsstrahlung recovery method employed in this study. The final part then describes the main background sources for the study and the strategies applied to reduce these background yields.

4.1 MC Datasets

The Monte Carlo data for this analysis was generated during the latest (to date) MC Campaign, MC13, using the release-04-00-03 version of the basf2 software. It consists of four distinct types of simulated events, using the whole configuration of the Belle II detector:

1. $B \rightarrow K\ell^+\ell^-$ signal events: generated using the the decay kinematic model described by Ali et al., implemented in the EvtGen package; the branching fractions for their yield estimation in table 4.1 were obtained from the most recent PDG tables.
2. Generic continuum $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) events: the largest source of background for the analysis at hand, their yield (table 4.1) is obtained from the cross sections used in the Belle II Physics Generators (see chapter 3). We use the simulated data to train multi-variate classifiers that separate these from signal events.
3. Generic $B\bar{B}$ events: semileptonic events (events with one or more neutrinos in their final states), higher charmonium resonances and hadronic decays with misidentified particles can mimic signal candidates. We use the data provided to train multi-variate classifiers in order to suppress as much of this background source as possible while still keeping a large fraction of the true reconstructed events. Notice that this data *contains signal events too*. Their yield (table 4.1) is also calculated from the branching fractions mentioned in the previous chapter.
4. $B \rightarrow KJ/\psi [\rightarrow \ell^+\ell^-]$ and $B \rightarrow K\psi(2S) [\rightarrow \ell^+\ell^-]$ events (1 million each); all K^0 mesons in these events are K_S^0 which decay through a pair of charged pions.

4.2 Event Selection

In the Belle II experiment, final state particles (particles with lifetimes long enough to decay outside the detector region) are associated to tracks and clusters found during the pre-processing of the detector data; in the current study, electrons, muons, charged kaons and charged pions are reconstructed from tracks. Clusters in the ECL without any track related to them are assigned to photons, which are used in this

Table 4.1. MC datasets

Type of event	MC ($\times 10^6$)	data	data/MC
Signal	$B^0 \rightarrow K^0 \mu^+ \mu^-$	5	3.06×10^{-7}
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	5	8.44×10^{-7}
	$B^0 \rightarrow K^0 e^+ e^-$	10	7.23×10^{-8}
	$B^+ \rightarrow K^+ e^+ e^-$	10	5.26×10^{-7}
$B^+ B^-$	1080	4.79×10^6	4.43×10^{-3}
$B^0 \bar{B}^0$	1020	4.52×10^6	4.43×10^{-3}
$q\bar{q}$	7380	32.70×10^6	4.43×10^{-3}

study to account for the energy loss of electrons due to Bremsstrahlung radiation while traversing the detector material.

4.2.1 Charged Particle Identification

The Belle II experiment uses the concept of likelihoods to determine the type of particle corresponding to a given track. Although in the previous chapter the TOP and the ARICH subdetectors were presented as the particle identification system, in reality all subdetectors provide complementary information on this aspect. For each one, a likelihood profile on a measurable quantity is constructed for each particle hypothesis; these are then used to define the global likelihood for a particle hypothesis α as

$$\mathcal{L}_\alpha(\mathbf{x}) = \prod_i \mathcal{L}_{i,\alpha}(\mathbf{x}, p) = \prod_i \mathcal{P}_i(\mathbf{x}|\alpha, p) \quad (4.1)$$

where \mathbf{x} represents the measurements associated to the given track, and $\mathcal{P}_i(\mathbf{x}|\alpha, p)$ can be interpreted as the probability of detector i registering the values \mathbf{x} , given the fact that the track corresponds to a particle α and carries momentum p ¹. The actual algorithm uses the logarithm of the likelihoods in order to avoid numerical inaccuracies.

The SVD and CDC detectors use the truncated mean of the energy loss dE/dx of each particle in their respective volumes as the measurable quantity. This is specially useful for low momentum tracks.

¹The precision in the momentum measurement done by the CDC is sufficient for using it as a parameter for the probability distributions.

The TOP and ARICH detectors, as mentioned in chapter 3, use Čerenkov radiation emitted by particles in a highly-refractive medium to measure their speed (or time of propagation) and, together with the momentum information from the CDC, their mass, hence determining the type of particle. The likelihood for the TOP detector takes as measured quantity the number of photons arriving at time intervals Δt in each channel of the photon detector array, over the whole time-of-arrival interval t . ARICH likelihood calculations use a similar approach, using the probability of each pixel of the detector to register a hit given the track momentum and direction and certain particle hypothesis. Likelihoods for charged hadrons —protons, pions, kaons and deuterons— use the information of these two detectors, together with the SVD and CDC probabilities.

The ECL uses the clusters related to a track to measure the shower energy deposited by the particle; it then calculates the ratio of this energy over the particle momentum measured by the CDC (E_{ECL}/p) and uses it to obtain the related probabilities $\mathcal{P}_{\text{ECL}}(E_{\text{ECL}}/p|\alpha, p)$. The discrimination power of the electron likelihood comes primary from the ECL measurements, and is complemented by the probabilities obtained from the SVD, CDC, TOP and ARICH detectors. Notice, however, that the separation power of the ECL likelihood is drastically reduced for low momentum electrons, since the radius of curvature of the particle path is very small, and hence the track never reaches the ECL.

Finally, the KLM uses the difference in longitudinal penetration depth and transverse scattering of the extrapolated track from the CDC. The information from this detector, is mainly used for K_L and muon identification, together with information from the SVD, CDC, TOP and ARICH.

Once the likelihoods from all subdetectors are combined, the global particle ID is obtained as the ratio

$$\text{PID}_\alpha(\mathbf{x}) = \frac{\mathcal{L}_\alpha(\mathbf{x})}{\sum_\beta \mathcal{L}_\beta(\mathbf{x})}. \quad (4.2)$$

This is the quantity used at the analysis level to determine the particle nature of the reconstructed track; the interested reader is referred to [35] for further details.

4.2.2 Skimming

Since the amount of data the Belle II Experiment will record is extremely large, and most of the processes taking place do not produce a $B\bar{B}$ pair (see table 3.2), loose preselection cuts are applied to the whole data set, which are intended to preserve most of the interesting events for a certain physics analysis, while removing a large

portion of the undesired background. This reduces the number of data that analysts in the collaboration need to handle. This process is known as skimming. Usually, skims are designed by the physics working groups of the collaboration, and handed to the data production group to be applied on both the real and simulated datasets. Each working group focuses on physics processes of similar nature, and hence that require the same type of skim.

The skim used for this analysis was designed to select $B \rightarrow X\ell^+\ell^-$ decays, where X is any hadron. The cuts applied are summed up in table 4.2.

Table 4.2. $B \rightarrow X\ell^+\ell^-$ skim

Applied to	Cut
	Normalized Fox-Wolfram moment $R_2 < 0.5$
Event	Number of tracks ≥ 3 $E_{\ell^+}^* + E_{\ell^-}^* > 1.5 \text{ GeV}$
All lepton candidates	$ dz < 5 \text{ cm}$ and $dr < 2 \text{ cm}$ $p > 0.395 \text{ GeV}$
Electrons candidates	electronID > 0.1
Muon candidates	muonID > 0.5 electronID < 0.1

Values with an asterisk are calculated in the Center of Mass frame of the experiment. The variable dz (dr) is the distance between the IP and the PCA on the z axis ($r - \phi$ plane). For a brief explanation of the R_2 variable, see the section on continuum suppression on this chapter.

4.2.3 Selection of Final State Charged Particles

In order to further reduce the presence of undesired background, mostly coming from the large number of pions generated and misidentified as leptons (or kaons), tighter cuts in the PID values for electrons, muons and charged Kaons were applied. For muons, we also require a momentum above 0.9 GeV in order to assure that the particle track reaches the KLM, which is the main subdetector used for muon identification. Pion tracks, used in the reconstruction of K_s^0 , do not get any cut. Table 4.3 shows the cut values.

Table 4.3. Cuts for final state particles

Particle	Cut
Electron	electronID > 0.9
Muon	muonID > 0.9 and $p > 0.9$ GeV
Charged kaons	kaonID > 0.6 and $ dz < 5$ cm and $dr < 2$ cm

Bremsstrahlung Recovery for Electrons

Electrons may lose an important portion of their energy through Bremsstrahlung radiation² while moving inside the detector volume. The energy lost is carried by a photon emitted in a direction tangential to the electron trajectory, which is eventually absorbed in the ECL (if its energy is high enough —larger than 10 MeV).

At the Belle experiment, Bremsstrahlung recovery was done during the offline reconstruction step. The strategy, depicted in figure 4.1, consisted of adding the energy of all ECL clusters which were not related to a track, and which were located within a cone around the electron momentum, to the electron energy. The electron momentum was then updated to

$$\mathbf{p}_e = \mathbf{p}_e^{\text{track}} + \sum_{\gamma} \mathbf{p}_{\gamma} \quad (4.3)$$

where \mathbf{p}_{γ} is the momentum of the photon associated with the ECL cluster, and the sum runs over all ECL clusters whose energy was added to the electron's one. For electrons with small energies and for photons radiated far from the electron track origin, this approach is expected to yield a low finding efficiency, since the photon emission direction can greatly differ from the electron initial direction.

At Belle II, this issue is tackled by a novel procedure introduced by Ecker, Hauth and Bernlocher [63] (we briefly describe their work here; for a complete picture of it, the reader is referred to the bibliography). Through MC simulations of $Y(4S)$ decays, the places in the Belle II detector where an electron is more likely to emit

²Bremsstrahlung radiation corresponds to the electromagnetic radiation that takes place when the momentum of a charged particle changes. The radiated power is given by the expression

$$P = \frac{q^2 \gamma^4}{6\pi} \left(\dot{\boldsymbol{\beta}}^2 - (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})^2 \right)$$

where q is the charge of the particle, $\boldsymbol{\beta}$ is its velocity, and $\gamma = E/m$ (E and m are the energy and mass of the particle). This implies that the energy lost due to Bremsstrahlung is much larger in light particles (like electrons).

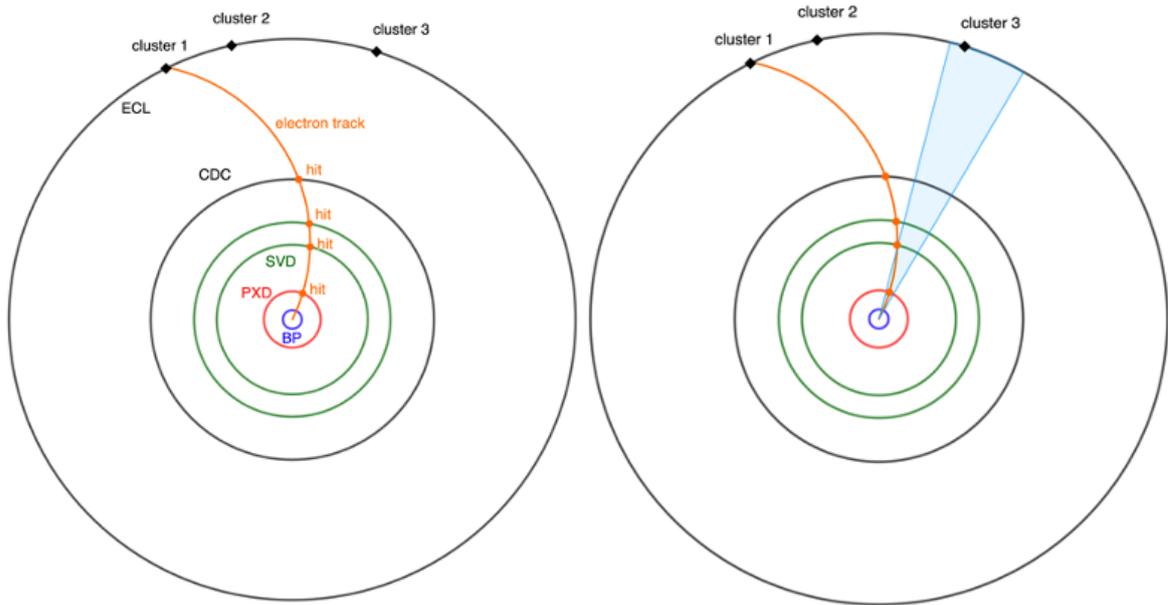


Fig. 4.1 Belle-like Bremsstrahlung recovery method. *Left:* the electron depicted emits three Bremsstrahlung photons, associated with the clusters 1, 2 and 3. Cluster 1 energy is added to the electron one, since it is in the particle path. *Right:* ECL clusters located inside a cone of 5 mrad around the initial direction of the electron momentum get added to the electron energy. In this case, the energy carried out by cluster 3 gets recovered. However, cluster 2 is not.

Bremsstrahlung radiation were detected; they are located within the innermost 20 cm of the detector volume, and correspond to the VXD layers, the beam pipe, and at the inner edge of the CDC (see figure 4.2). From the emitted photons, around 65% have energies above 10 MeV, and from these, 75% are detected by the ECL. The new method iterates over each of the track hits in the VXD region and the virtual hits at the beam pipe and the inner CDC wall, performing an extrapolation onto the ECL, and searches for secondary ECL clusters (ECL clusters with no track associated to them) that match these extrapolations. This matching is done by assigning acceptance ranges to both the extrapolated hit and the cluster, and checking if both ranges overlap. The acceptance ranges depend on the position uncertainty of the cluster and the hit, and on an acceptance factor define by the user; in their work, Ecker, Hauth and Bernlocher determined through an MC study that a value of 3.0 for the acceptance factor represents an optimal balance between photon detection efficiency and fake rates. Once the matching is performed, the ECL cluster is labeled as a potential Bremsstrahlung photon associated with the track, and a relation weight is assigned to them, which measures the quality of the match (the larger the weight,

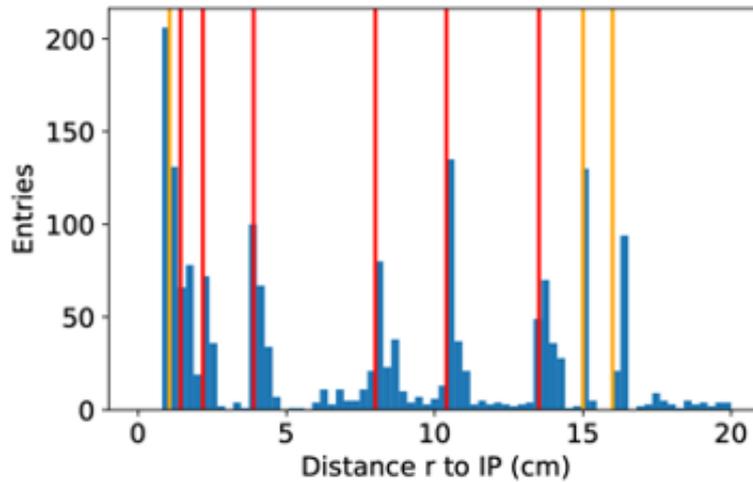


Fig. 4.2 Distribution of the distance between the Bremsstrahlung photon vertices and the interaction point in 10^4 MC simulations of $Y(4S)$ decays without beam background. The vertical lines, from left to right, correspond to: beam pipe wall (leftmost yellow line), VXD layers (2 PXD layers and 4 SVD layers, red lines), inner and outer face of the inner CDC wall (rightmost yellow lines). Taken from [63].

the worse the match). The algorithm is outlined in appendix C, and is performed during the online reconstruction.

Figure 4.4 shows a comparison of the Bremsstrahlung recovery efficiency for the Belle method and the Belle II method, as evaluated in [63]; here, the efficiency is defined as the number of correctly matched Bremsstrahlung photons over the total number of Bremsstrahlung photons detected by the ECL. As expected, in the new method the efficiency is mostly independent of the electron energy and of the distance between the photon emission vertex and the IP.

In order to apply this method, we start by reconstructing photons with the selection cuts given in table 4.4. E_1/E_9 corresponds to the ratio of the highest energy deposited in a single ECL crystal, divided by the energy deposited in an array of 3×3 crystals with the one with the highest energy deposition in the center. Next, we check

Table 4.4. Bremsstrahlung photons selection cuts

Variable	Cut
ECL cluster polar angle	in the range $(17^\circ, 150^\circ)$
ECL cluster error timing	less than 1 ms
Energy	higher than 75 MeV or $E_1/E_9 > 0.4$

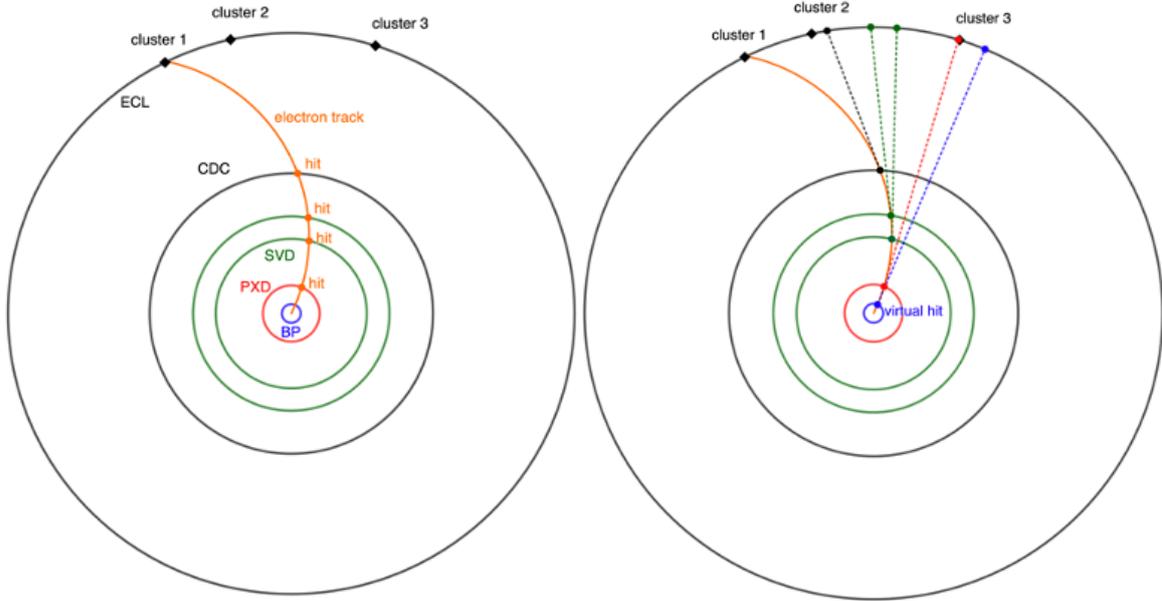


Fig. 4.3 Belle II-like Bremsstrahlung recovery method. *Left*: the electron depicted emits three Bremsstrahlung photons, associated with the clusters 1, 2 and 3. Cluster 1 energy is added to the electron one, since it is in the particle path. *Right*: ECL clusters located within the acceptance region of the projected hits (real and virtual) are related to the electron track, with a weight proportional to the distance between the projected hit and the cluster. The user can later determine the maximum weight allowed for a photon to be considered a Bremsstrahlung photon of a given electron. If more than two tracks get the same cluster associated, the cluster is assigned to the track with the smallest weight.

if any of the reconstructed photon has been associated as a Bremsstrahlung photon of any of the electron (or positron) tracks we have previously selected. If it has, we add its momentum to the momentum of the electron. Since one cluster (photon) can be related to multiple tracks (electrons), the photon momentum is added only to the electron with the lowest relation weight. Electrons are allowed to have as many Bremsstrahlung photons as possible. Figure 4.5 shows the invariant mass distribution of the electron-positron pair for $B \rightarrow K\psi(nS) [\rightarrow e^+e^-]$ reconstructed events from the dedicated MC samples, before and after applying the novel recovery method.

As a way of measuring the improvement achieved with the algorithm, we calculate the ranges of q which contain a certain percentage of the reconstructed events before and after the Bremsstrahlung recovery is performed. The results are shown in table 4.5. We see that after the algorithm is applied, the width of these ranges is reduced by 82% for the J/ψ decays, and by 85% for the $\psi(2S)$ decays. Reducing the width of these ranges is of uttermost importance for the analysis, since these

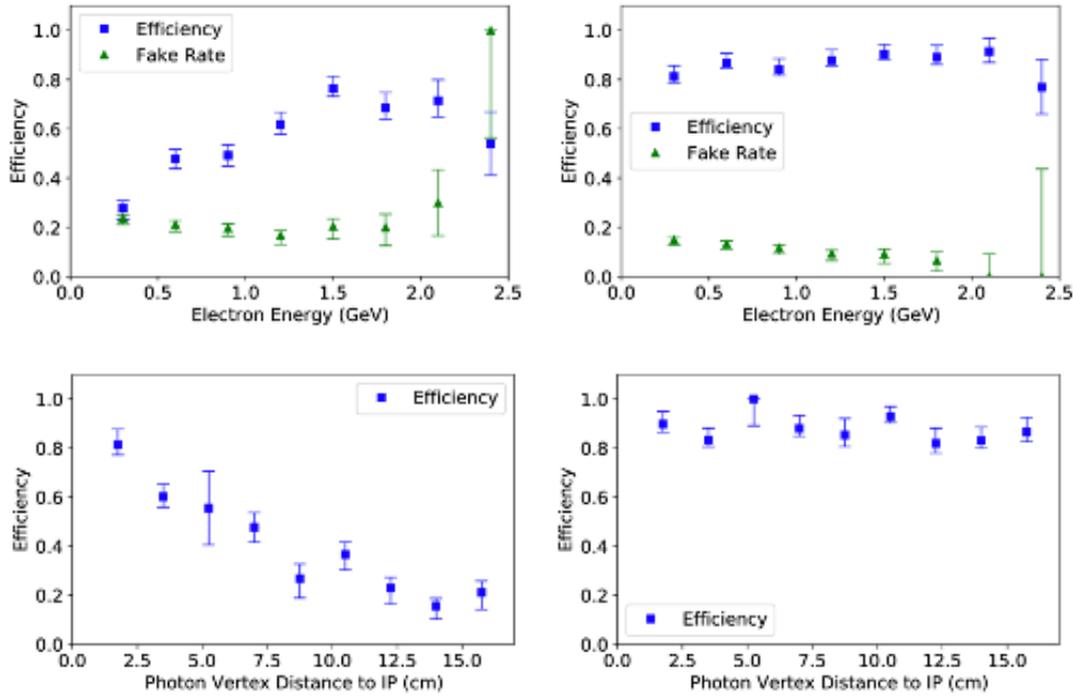


Fig. 4.4 Belle (left) and Belle II (right) Bremsstrahlung recovery efficiencies, evaluated in 10^4 MC simulations of $Y(4S)$ decays without any beam background. The upper two plots depict the efficiency in function of the electron energy, whereas the bottom two present the efficiency in function of the distance between the photon emission vertex and the IP. All the images are taken from [63].

resonant decays constitute one of the largest source of background for the analysis, and they must be removed by applying a veto in q around the nominal mass of the resonances. A better mass resolution implies a narrower veto window, which benefits the signal efficiency.

4.2.4 K_S^0 Reconstruction

The K_S^0 meson has a mean lifetime of 8.954×10^{-11} [15] seconds; as a first approximation, assuming it moves at the speed of light, it will travel a few millimeters before decaying inside the Belle II detector volume. Its main decays modes are into a pair of charged pions (69.2% of the time), and into a pair of neutral pions (30.69%). Neutral pions decay even faster than the K_S^0 to a pair of photons. The reconstruction of the decay $K_S^0 \rightarrow \pi^0(\gamma\gamma)\pi^0(\gamma\gamma)$ requires then the detection of 4 low energetic photons, which can be very difficult considering the background levels due to the increased

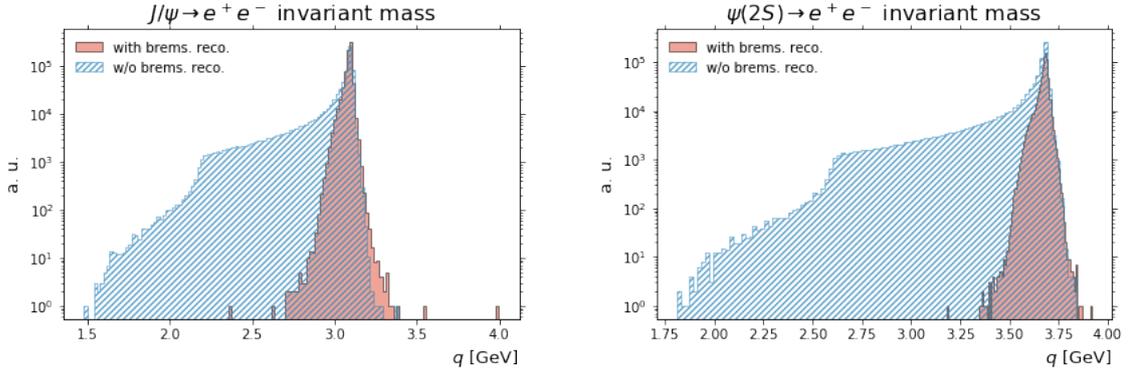


Fig. 4.5 Dilepton invariant mass [GeV] distribution for the $B \rightarrow K\psi(nS) [\rightarrow e^+e^-]$ decays. The tail towards lower values gets reduced and the number of events peaking at the nominal masses increases after the new Bremsstrahlung recovery method is applied.

Table 4.5. q ranges containing a percentage of the correctly reconstructed charmonium events from MC

resonance		q range [GeV]		
		68%	95%	99.7%
$J/\psi \rightarrow e^+e^-$	w/o brems. corr.	(2.88,3.10)	(2.37,3.11)	(2.06,3.14)
	with brems. corr.	(3.07,3.10)	(3.02,3.12)	(2.95,3.16)
	width ratio	0.16	0.14	0.19
$\psi(2S) \rightarrow e^+e^-$	w/o brems. corr.	(3.43,3.69)	(2.82,3.70)	(2.47,3.74)
	with brems. corr.	(3.66,3.69)	(3.60,3.71)	(3.54,3.75)
	width ratio	0.15	0.13	0.16

luminosity in the machine. This is the reason why only the decay mode $K_S^0 \rightarrow \pi^+\pi^-$ is used throughout this analysis.

The K_S^0 candidates are obtained from vertex objects found during the preprocessing, and from a combination of oppositely charged tracks under the pion hypothesis. The invariant mass of the candidate is loosely constrained to lie within the K_S^0 nominal mass. Due to the fact that the kinematic variables of the pions are calculated assuming that they were generated at the interaction point, their total four momentum will differ from the true momentum of the K_S^0 . In order to correct this discrepancy a vertex fit is performed, where the four momenta of the pions are recalculated at the vertex position of the K_S^0 using a Kalman fitter, a technique first

developed at BaBar [64], and adapted to the Belle II experiment under the name *TreeFitter* [65].

During its more than ten years of operation, the Belle collaboration came up with numerous and creative ways of reducing the amount of undesired data. A selection of cuts on the parameters of the pion tracks, as well as a tighter cut in the invariant mass of the K_S^0 , define what is known as a *goodBelleKshort*. These cuts depend on the K_S^0 momentum in the laboratory frame, and are resumed in table 4.6. The track parameters are

- dr : minimum between the distances of closest approach of both pion tracks to the interaction point in the $r - \phi$ plane.
- $d\phi$: angle between the K_S^0 momentum vector and the directional vector from the interaction point to the K_S^0 decay vertex.
- dz : distance in the beam direction between the two pion tracks.
- ρ : flight length of the K_S^0 in the $r - \phi$ plane; this is, the distance between the interaction point and the K_S^0 decay vertex perpendicular to the beam direction.

Table 4.6. Track parameters' cuts for *goodBelleKshort*. The mass window is $0.468 \text{ GeV} < m(K_S^0) < 0.528 \text{ GeV}$

p_{lab} [GeV]	dr [cm]	$d\phi$ [mrad]	dz [cm]	ρ [cm]
< 0.5	> 0.05	< 0.3	< 0.8	≥ 0.00
$0.5 - 1.5$	> 0.03	< 0.1	< 1.8	≥ 0.08
≥ 1.5	> 0.02	< 0.03	< 2.4	≥ 0.22

An MC analysis of the decay $D^{*+} \rightarrow D^0 [\rightarrow K_S^0 \pi^+ \pi^-] \pi^+$ showed that the implementation of these cuts reduced the background component in 97.4% while conserving 83% of the signal events (see the section on K_S^0 systematics on the next chapter). Under the light of these results, the cuts are incorporated in the reconstruction process.

4.2.5 B Reconstruction

For the case of B mesons, as neatly exposed in [66], the fact that the $Y(4S)$ resonance energy is just above the needed to produce a pair of B mesons means that, in the center of mass system of the experiment, both of these mesons will be almost at rest.

Each one will have, to a very good approximation, a Center of Mass (CM) energy of $\sqrt{s}/2$, where \sqrt{s} is the CM Energy of the whole system, which for the Super KEKB accelerator is 10.579 GeV (the mass of the $Y(4S)$ resonance). Correctly reconstructed B mesons should have energies close to this nominal value. Hence, the variable

$$\Delta E = E_{\text{reco}}^* - \frac{\sqrt{s}}{2} \quad (4.4)$$

(where E_{reco}^* is the reconstructed energy of the B meson in the CM system), should be close to zero for true candidates.

The invariant mass of the reconstructed particle is also a good variable to discriminate between true events and combinatorial background. However, in the case of B mesons, it is strongly correlated to ΔE ³ and hence, a new variable, known as the *beam constrained mass*, is used:

$$m_{bc}^2 = \frac{s}{4} - \left(\sum_i \mathbf{p}_i^* \right)^2. \quad (4.5)$$

The energy difference ΔE is quite sensitive to particle missidentification: particles with different masses will contribute in different ways to the total value of E_{reco}^* . This implies that tight cuts on this variable are more useful to reject background coming from missidentification than cuts in the beam constrained mass. We use m_{bc} as the fitting variable.

In order to collect enough statistics on the generic and continuum backgrounds, needed to determine their shape and yield, a first set of very loose cuts are applied to both m_{bc} and ΔE : $5.2 \text{ GeV} < m_{bc} < 5.3 \text{ GeV}$ and $-0.5 \text{ GeV} < \Delta E < 0.5 \text{ GeV}$. The cuts on ΔE are tightened after looking at the distribution of the main background sources, as we will soon explain. Figure 4.6 shows the ΔE distribution for correctly reconstructed signal events in the electron and muon channels.

Due to Bremsstrahlung radiation, and despite the Bremsstrahlung recovery procedure described previously, a tail towards negative values in the ΔE distribution is observed in decay modes containing electrons. Hence, it is necessary to apply separate ΔE cuts for each decay mode according to its lepton flavor.

³Indeed, in the CM system, the sum of momenta of the decay products of the B meson is close to zero (since it is almost at rest); hence, the major contribution to the invariant mass is obtained from the sum of energies of the decay products; but this is exactly E_{reco}^* in equation (4.4). The beam constrained mass partially solves this, since it replaces E_{reco}^* by $\sqrt{s}/2$.

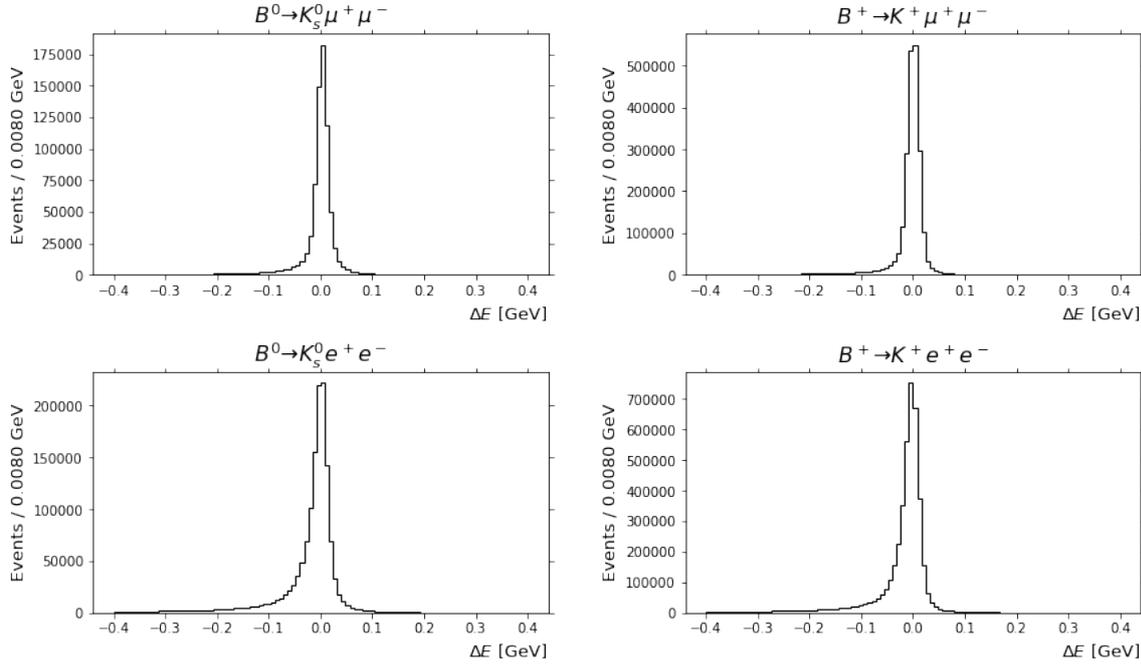


Fig. 4.6 ΔE distribution for the reconstructed signal events. Notice that the decay modes containing electrons have a longer radiative tail, due to Bremsstrahlung losses in the detector volume.

In order to obtain these cuts, we first define a signal window from the m_{bc} distributions shown in figure 4.7. More than 90% of the reconstructed signal events are in the region $m_{bc} > 5.27$, and thus the signal region is defined as $m_{bc} \in [5.27, 5.3]$ GeV for all modes.

Next, the new ΔE cuts are obtained from a MC study using a 2 ab^{-1} mixture of generic $B\bar{B}$ (containing a proper amount of signal events) and $q\bar{q}$ samples. After performing the reconstruction procedure on this data, we select upper and lower cuts close to the values that maximize the pseudo-significance inside the signal region:

$$\text{FOM} = \frac{N_S}{\sqrt{N_B + N_S}}. \quad (4.6)$$

Here N_S is the number of signal events inside the signal region and the ΔE window, and N_B is the number of background events that fall inside the signal region and within the window for ΔE . Figure 4.8 shows the value of the FOM for different combinations of the lower and upper limits for ΔE . We decide to use a common range for decays with the same lepton flavor, since the ΔE distributions are independent of the K meson flavor. The final cuts are shown in table 4.7.

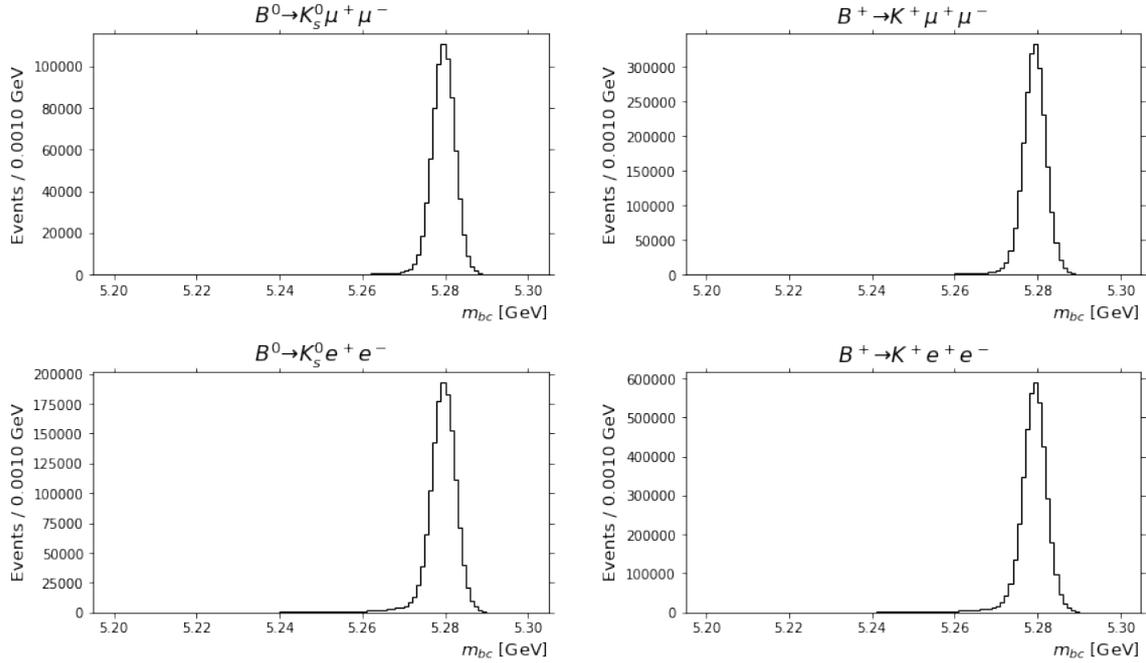


Fig. 4.7 m_{bc} distribution for the reconstructed signal events, after the loose cuts in m_{bc} and ΔE . The percentage of events in the signal region (this is, with $m_{bc} > 5.27$ GeV) are: 91.7% ($B^0 \rightarrow K_s^0 \mu^+ \mu^-$), 90.4% ($B^+ \rightarrow K^+ \mu^+ \mu^-$), 91.0% ($B^0 \rightarrow K_s^0 e^+ e^-$) and 90.0% ($B^+ \rightarrow K^+ e^+ e^-$).

Table 4.7. ΔE cuts per decay mode

Decay mode	cut
$B \rightarrow K \mu^+ \mu^-$	$-0.1 \text{ GeV} < \Delta E < 0.05 \text{ GeV}$
$B \rightarrow K e^+ e^-$	$-0.12 \text{ GeV} < \Delta E < 0.07 \text{ GeV}$

4.2.6 Sum-up of the event selection cuts and ROE

We end this section by summing up all the selection cuts mentioned previously.

We require for our events to contain at least two oppositely electron tracks, or two oppositely muon tracks. A lepton track is defined as a track with a momentum in the laboratory frame larger than 0.395 GeV and with an electronID larger than 0.9; a muon track has a momentum larger than 0.9 GeV and a muonID larger than 0.9. Energy lost by the electrons due to Bremsstrahlung radiation is recovered by collecting the momenta of all the photons in the vicinity of the projected hits on the beam pipe, VXD layers, and CDC inner walls onto the ECL. The events must also contain at least one charged Kaon track, which is selected by a cut in the kaonID larger than 0.6, or one neutral Kaon, obtained either from a vertex detected during

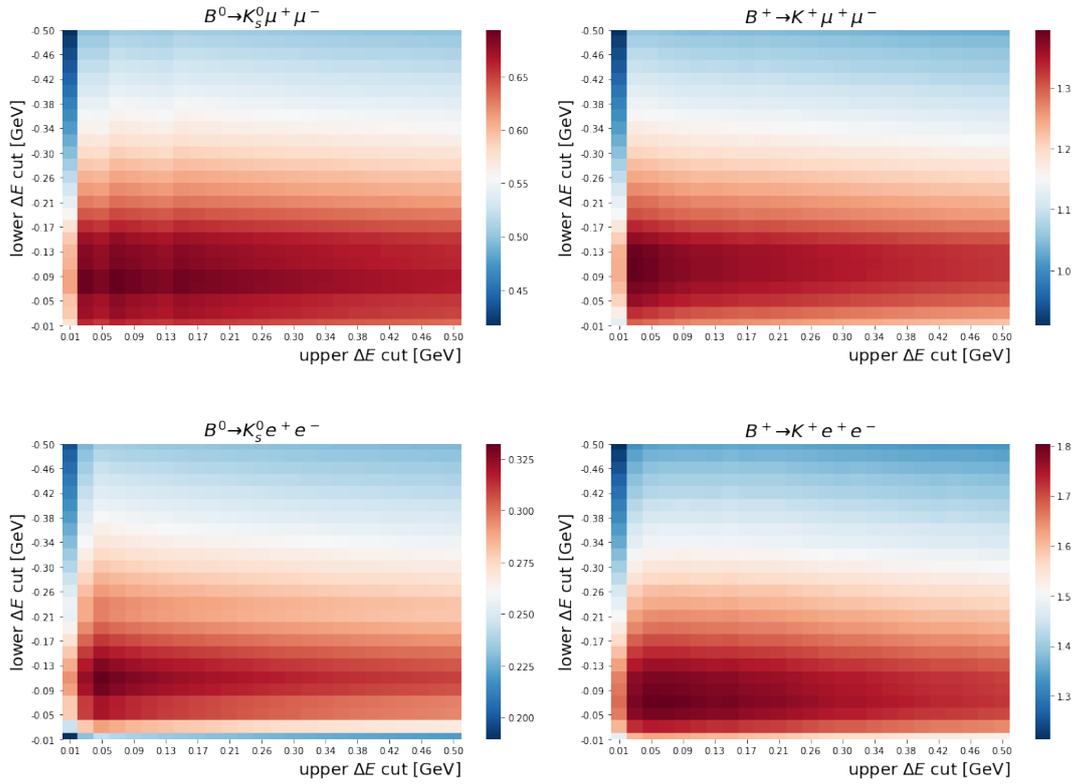


Fig. 4.8 Figure of merit distribution for different values of ΔE_{low} and ΔE_{high} . The signal and background yields in equation 4.6 are those inside the signal region $m_{bc} > 5.27$ GeV.

the online reconstruction, or by the combination of two oppositely charged tracks. The neutral kaon must also pass a vertex fit procedure, have a mass within 30 MeV of the nominal K_S^0 mass, and additional cuts according to its momentum, as detailed in table 4.6.

The B meson is finally reconstructed by adding up the two lepton tracks and the kaon track/vertex, and we require its beam constrained mass to be within $[5.27, 5.3]$ GeV; we define the sideband region from events where the beam constrained mass is between $[5.2, 5.27]$ GeV. We also require the energy difference of the reconstructed B meson to be within $[-0.12, 0.07]$ GeV if the meson is made up from electron tracks, or within $[-0.1, 0.05]$ GeV if the B meson is made up from muon tracks. No vertex fit is applied on the B meson decay.

Once the B meson candidate is reconstructed, the rest of the tracks, ECL and KLM clusters in the event are grouped in what is known as the Rest Of the Event (ROE). Tracks are assumed to belong to charged pions, isolated ECL clusters to photons, and isolated KLM clusters to K_L^0 . In order to reduce the contamination due to the beam background, a series of cuts are applied to the reconstructed objects:

tracks are required to have a momentum in the CM system less than 3.2 GeV, and a transverse momentum larger than 0.5 GeV, and ECL cluster must have an energy above 0.5 GeV.

4.3 Background analysis

Figure 4.9 presents the m_{bc} distribution for each decay mode after applying the reconstruction steps and cuts described in the previous section. Notice that the signal contribution is hardly discernible. In order to alleviate this, this section presents different strategies employed to reduce the background presence in the final data sample.

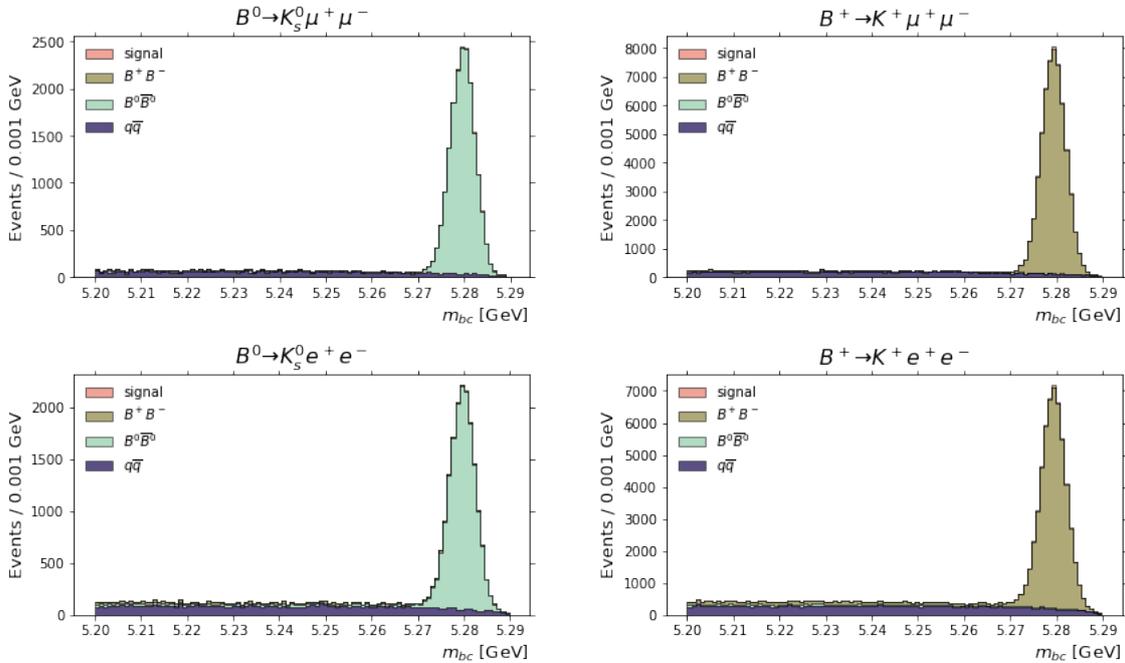


Fig. 4.9 m_{bc} distribution for the initial reconstruction performed on the four channels of interest.

The main sources of background can be categorized in three groups:

- Peaking backgrounds, which correspond to decays that contain some resonances, and hence appear as a clear peak in the invariant mass distribution of a group of decay products. In the case at hand, three sources have been identified: first, the processes $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$, which include the charmonium resonances $\psi(nS)$; for these, the invariant mass of the lepton pair peaks around

the mass of the resonance. Second, the processes in which a π^+ is combined with a kaon and a π^- coming from the decay of a \bar{D} meson; here, a double missidentification of the pair of pions as leptons, or a triple missidentification⁴ of a pion as a kaon, and of the other two hadrons as leptons, present peaks in the invariant mass distribution of the $K\ell^-$ pair (in the double miss-ID case) or in the invariant mass of the lepton pair (in the case of triple miss-ID), around the D meson mass. Finally, for decay modes containing electrons, continuum processes in which a photon undergoes pair conversion, or a π^0 undergoes a Dalitz decay, peak at small values of the invariant mass of the electron pair.

- Continuum background, which is a combinatorial background that originates from decays of the form $e^+e^- \rightarrow q\bar{q}$, where q is any quark lighter than the bottom quark, and mimics the kinematical signature of true events. Luckily, the topologies of $e^+e^- \rightarrow q\bar{q}$ and $e^+e^- \rightarrow B\bar{B}$ decays are very different in B factories, allowing for an effective way to suppress this source [66].
- Combinatorial background from $B\bar{B}$ events. Most of this background comes from missidentification of the final state particles, and from wrong combinatorial errors.

4.3.1 Peaking backgrounds

π^0 Dalitz decay and photon conversion

Figure 4.10 shows the $q = M(\ell^+\ell^-)$ distribution for 2 ab^{-1} of MC data. The peaks from the $q\bar{q}$ events in the low q region for the electron channels are associated with Dalitz π^0 decays and photon conversions. This first source of background is suppressed by a cut on the minimum value of q for electron channels. From the histograms, a common cut of $q > 0.16 \text{ GeV}$ is decided for both modes.

Charmonium resonances

The prominent peaks in the charged (neutral) channels in figure 4.10 originating from B^+B^- ($B^0\bar{B}^0$) decays are related to the processes $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$. As these events are completely indistinguishable from our signal events, we veto the q region around the nominal mass of the $\psi(nS)$ resonances. Given the branching

⁴Only in decays involving charged B mesons.

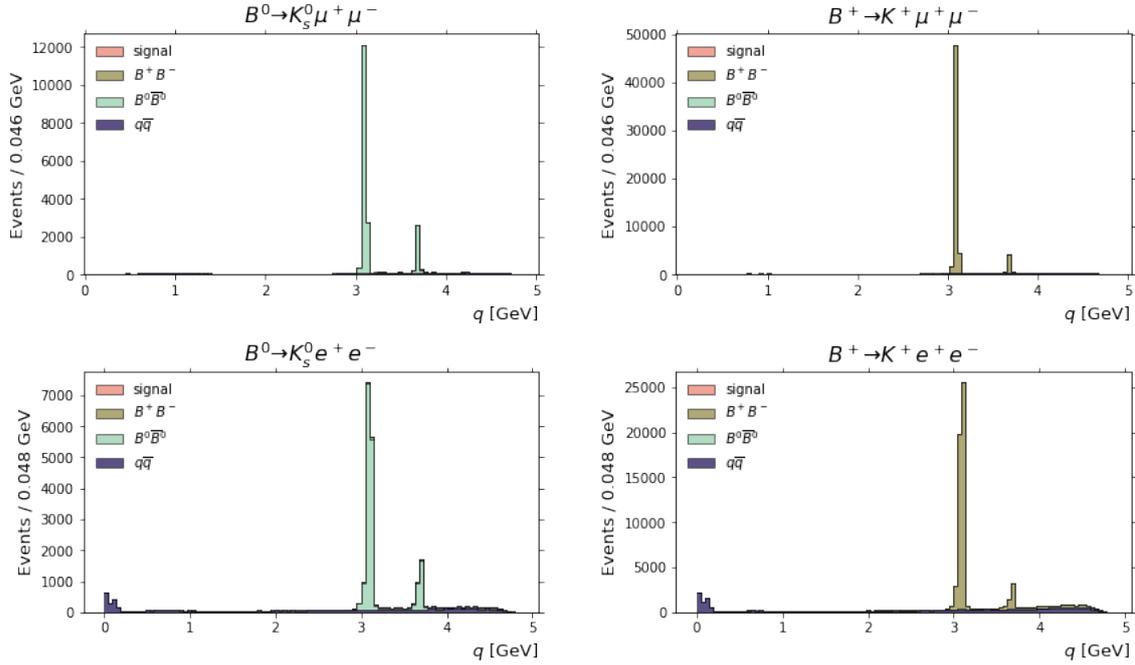


Fig. 4.10 Dilepton invariant mass distribution for the reconstruction performed on the four channels of interest.

ration of the multiple resonances $\psi(nS) \rightarrow \ell^+ \ell^-$, it is enough to consider only the first tow, J/ψ and $\psi(2s)$ ⁵.

We follow the same strategy than with the ΔE cuts: select a veto region in q such that the expression

$$\text{FOM} = \frac{N_S}{\sqrt{N_S + N_B}} \quad (4.7)$$

is maximized. The data sample corresponds to our previously reconstructed events, with the tighter ΔE cuts and Dalitz vetoes (in the electron channels) applied. For details, see Appendix D.

For the electron modes, some resonances may escape the veto region in two ways: energy losses can lead to an invariant mass value which is smaller than the lower cut, or the Bremsstrahlung recovery can increase the invariant mass to a value above the veto region. In order to account for this, the veto is applied not only to the invariant mass of the corrected electron pair, but also to the invariant mass of the electrons before the recovery, and to the invariant masses obtained by recovering the momentum of only one of the two electrons.

⁵Indeed, the branching fraction for the leptonic decay of the next resonance, $\psi(3770) \rightarrow e^+ e^-$, is already three orders of magnitude lower than the $\psi(2s) \rightarrow e^+ e^-$ one. This is also clear from figure 4.10.

The final veto regions, and the percentage of resonances present after the vetoes, are listed in tables 4.8 and 4.9 and depicted in figures 4.11 and 4.12.

Table 4.8. Veto regions for the J/ψ resonance

Decay mode	Veto region
$B \rightarrow K\mu^+\mu^-$	$-0.12 \text{ GeV} < q - m(J/\psi) < 0.06 \text{ GeV}$
$B \rightarrow Ke^+e^-$	$-0.17 \text{ GeV} < q - m(J/\psi) < 0.07 \text{ GeV}$

Table 4.9. Veto regions for the $\psi(2S)$ resonance

Decay mode	Veto region
$B \rightarrow K\mu^+\mu^-$	$-0.11 \text{ GeV} < q - m(\psi(2S)) < 0.05 \text{ GeV}$
$B \rightarrow Ke^+e^-$	$-0.13 \text{ GeV} < q - m(\psi(2S)) < 0.07 \text{ GeV}$

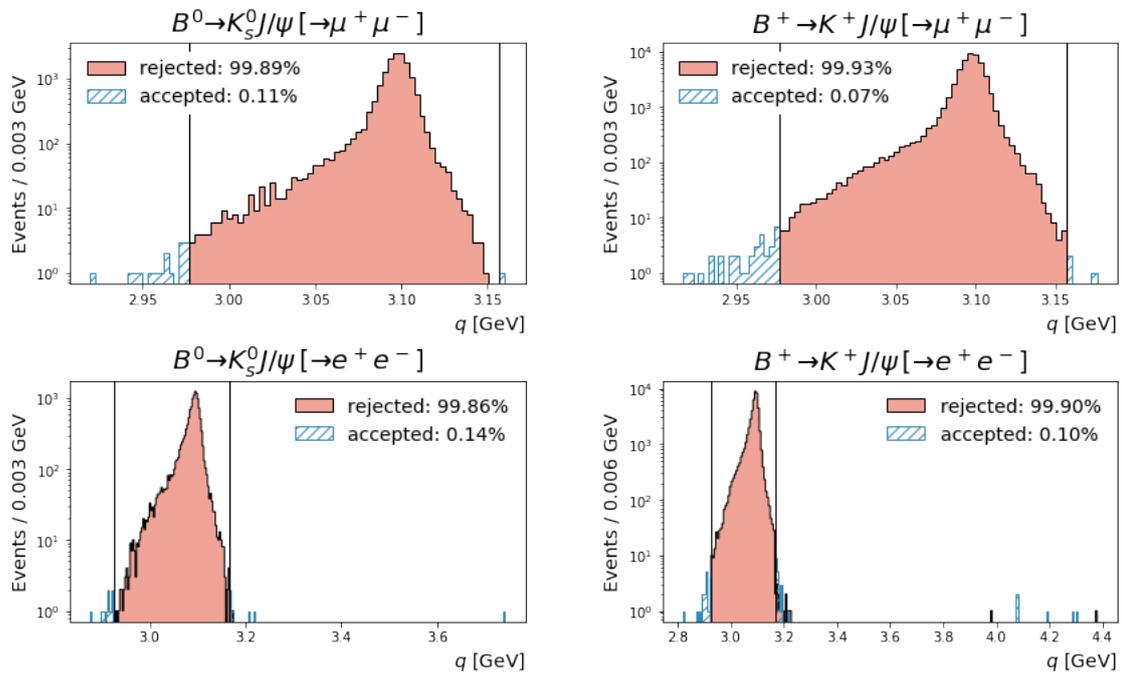


Fig. 4.11 q distribution for the resonant $B \rightarrow KJ/\psi [\rightarrow \ell^+ \ell^-]$ decays. The vertical black lines denote the veto region. Events inside this region get rejected. Notice, however, that for the decays containing electrons, events *outside* the region also get rejected; this is because at least one of the dilepton invariant masses calculated without Bremsstrahlung recoveries falls inside the veto.

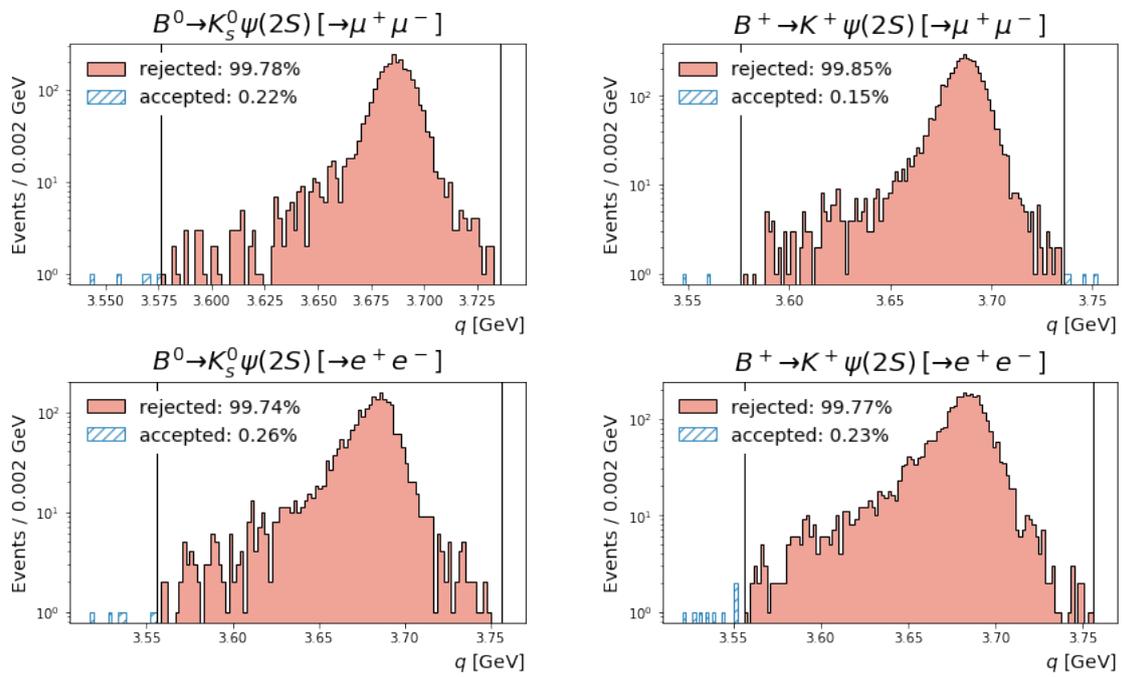


Fig. 4.12 q distribution for the resonant $B \rightarrow K\psi(2S) [\rightarrow \ell^+\ell^-]$ decays. The vertical black lines denote the veto region.

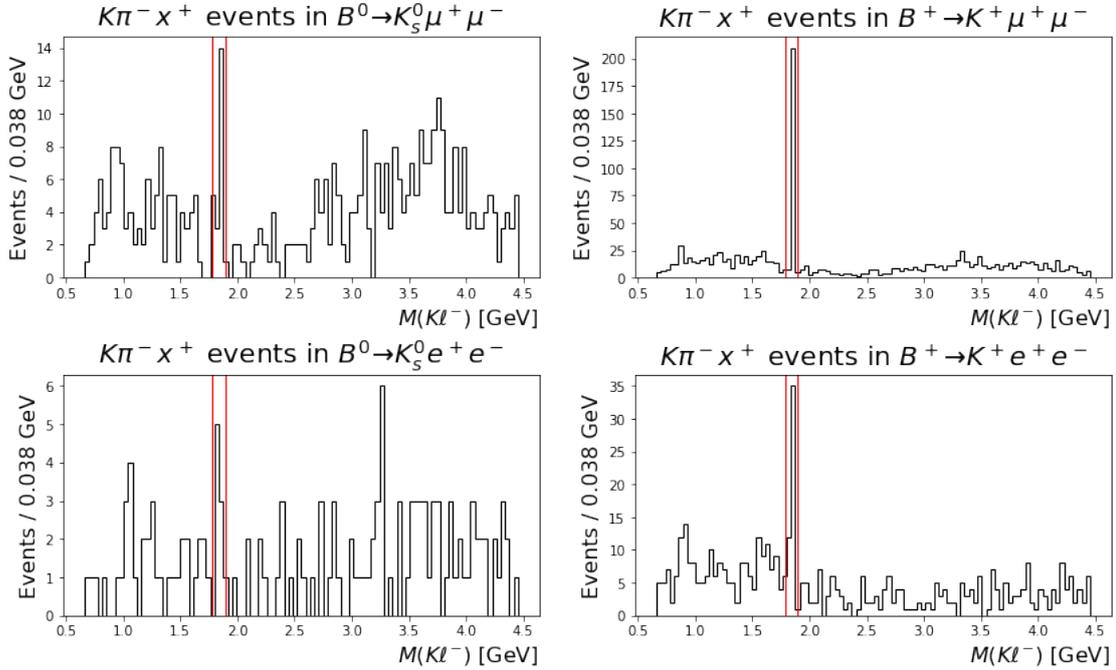
\bar{D} events

Fig. 4.13 $M(K\ell^-)$ distribution for $K\pi^-x^+$ events in the 2 ab^{-1} MC sample. The vertical red lines denote the veto region for the D meson.

Figure 4.13 shows the invariant mass distribution for the $K\ell^-$ pair for the events consisting of a correctly identified kaon, a pion misidentified as lepton, and an oppositely charged particle. A clear peak around the D^0 meson mass is present in the distribution for $B^+ \rightarrow K^+\ell^+\ell^-$ events, and a less prominent one in the $B^0 \rightarrow K_S^0\ell^+\ell^-$ around the D^+ mass. We suppress these peaks with a veto around the nominal mass of the respective D meson, as described in table 4.10.

Table 4.10. Veto regions for the D resonance

Decay mode	Veto region
$B^0 \rightarrow K_S^0\ell^+\ell^-$	$-0.09\text{ GeV} < M(K_S^0\ell^-) - m(D^+) < 0.03\text{ GeV}$
$B^+ \rightarrow K^+\ell^+\ell^-$	$-0.08\text{ GeV} < M(K^+\ell^-) - m(D^0) < 0.03\text{ GeV}$

4.3.2 Continuum Suppression

One of the advantages of B factories above energy-frontier facilities is the particular topology of the B meson decays, which can be used to discriminate between signal and combinatorial background events. This topology affects the phase-space distribution of the decay products, and multiple variables have been devised in order to quantify these effects. Reference [66] presents a comprehensive summary of the most common ones. We mention here only those used for the current analysis:

1. $\cos\theta_B$: The cosine of the angle between the reconstructed momentum of the B candidate and the z (beam) axis. As B mesons are scalar particles, whereas the $Y(4S)$ is a vector meson, the conservation of angular momentum requires for the B to follow a $\sin^2\theta_B$ distribution, whereas in $e^+e^- \rightarrow q\bar{q}$ processes, all particles involved have spin $1/2$, which implies that the two resulting jets follow a $1 + \cos^2\theta_B$ distribution.
2. *Thrust*: for a B meson, the thrust axis \mathbf{T} is defined as the unit vector that maximizes the expression

$$T = \frac{\sum_i |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_i |\mathbf{p}_i|} \quad (4.8)$$

where \mathbf{p}_i is the momentum of particle i , and the sum is done over the momenta of all the B decay particles. In the same way, it is possible to define a thrust axis for all the particles which do not belong to the decay of interest. These are referred to as the *Rest Of the Event* (ROE). The values of T (equation 4.8) for the B candidate and the ROE, the angle between the thrust axis for the B candidate and the beam axis (θ_{TBz}), and the angle between the thrust axes for the B candidate and the ROE (θ_{TBO}) are used as inputs. For a correctly reconstructed B meson, the decay products are produced isotropically, and thus both $|\cos\theta_{TBz}|$ and $|\cos\theta_{TBO}|$ should be uniformly distributed between $[0,1]$. For events containing jets, both the B candidate and the ROE are strongly collimated, and the distribution of the cosines peak at high values.

3. *Fox-Wolfram moments*: initially introduced by Fox and Wolfram [67], these quantities present another way of understanding the phase-space composition of the B products. They are defined as

$$H_k = \sum_{i,j} \frac{|\mathbf{p}_i| |\mathbf{p}_j|}{s} P_k(\cos\Omega_{ij}) \quad (4.9)$$

where the sum is over all possible pairs of particles in the event, Ω_{ij} is the angle between the momenta of particles i and j , and P_k refers to the k -th Legendre polynomial. The normalized Fox-Wolfram moments, R_k , are defined as $R_k = H_k/H_0$. For jet like structures, the angle between momenta is close to either zero (if the particles belong to the same jet) or π (for particles in different jets); R_k will then take values close to zero (one) for odd (even) values of k .

4. *Kakuna Super Fox-Wolfram moments*: a combination of multiple FW moments calculated only over a part (charged, neutral or missing) of the B primary daughters, or the ROE. They correspond to the R_j^{oo} and H_{xl}^{so} presented in Chapter 9.5.2 of [66].
5. *Cleo Cones*: developed by the CLEO collaboration, a set of nine variables CC_i measuring the momentum flow through nine cones around the thrust axis of the B candidate, with half angles evenly spaced between 10° and 90° .
6. E_T : the component of the event energy associated with the total transverse momentum of the event. For collimated jet-like events, this value should be far less than for isotropic B decays.
7. $\Delta z(\ell^+, \ell^-)$: the distance along the beam axis between the two oppositely charged leptons. For random combinations of tracks, this distance is in average larger than for true signal events.
8. M_{miss}^2 : the missing squared mass of the event. It is calculated as

$$M_{\text{miss}}^2 = E_{\text{miss}}^2 - \mathbf{p}_{\text{miss}}^2 \quad (4.10)$$

where E_{miss} and \mathbf{p}_{miss} are the missing energy and momentum of the event. In the CM system, the sum of the energies of the decay products should be equal to the initial energy of the system, \sqrt{s} , and the sum of their momenta should be equal to zero. E_{miss} and \mathbf{p}_{miss} measure the difference between these theoretical values and the actual total energy and momentum of the event.

A multivariate classifier is trained to provide a discriminant using these inputs. We choose boosted decision trees (BDTs) as our multivariate algorithm, since they offer multiple advantages over Neural Networks, while achieving the same separation power:

- By construction, BDTs are robust to correlations in the input parameters; on the other side, the training time for a Neural Network can be negatively impacted by correlations [68].
- BDTs are faster to train, and require the tuning of a fewer number of hyperparameters; this facilitates the search for an optimal architecture.
- Interpretability. BDTs apply cuts to the input variables, and have a defined metric for the importance of each input variable; Neural Networks, on the other side, work as black boxes, and as the network deepens, the correlation between variables increases, making it harder to disentangle the effects of each input on the output.

Among the multiple BDTs implementations, the fastBDT [69] one has been selected, due to its speed and the option of *boosting to flatness*, which penalizes the classifier if its outputs are correlated with variables that should remind unbiased for the analysis. In this case, we use this functionality to avoid introducing an artificial peak in the signal m_{bc} region (see appendix E for more on BDTs).

We trained a classifier for each lepton flavor, removing variables strongly correlated to the m_{bc} for each channel (we use the approach described in [70] to measure the correlation between two variables), and taking care of overfitting by keeping a validation sample which is not used during the training phase. Moreover, we compared the distributions of all the remaining training variables in MC and real data samples in the charmonium sidebands —after applying the cuts from the previous sections—, and performed a Kolmogorov-Smirnoff test to evaluate their similitude. The real data corresponds to the full experimental dataset⁶. The results are shown in table 4.11, and the distributions can be consulted in appendix F.

All variables, except for T_{ROE} and H_{n2}^{s0} in the $K^+e^+e^-$ decay, and $|\cos\theta_{TBTz}|$ in the $K_S^0e^+e^-$ decay, have a p-value greater than 0.05. We exclude these three variables from the inputs for the continuum BDT for electrons.

The training dataset consists of signal and background events in a 1:1 ratio. The total number of available samples is given in table 4.12. For each lepton flavor, the samples of both charged and neutral modes are combined. From these, 80% were used as training data, and the remaining 20% for the validation. The final set of variables used for each BDT is given in table 4.13, and their distribution in the signal and background samples for the training set are shown in figures 4.14 and 4.15. The

⁶For a more detailed description of this dataset, please refer to the next chapter.

Table 4.11. p -value of the KS test for the Continuum Suppression MVA training variables between MC and experimental data in the charmonium sidebands

Variable	KS test for			
	$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	$B^+ \rightarrow K^+ \mu^+ \mu^-$	$B^0 \rightarrow K_s^0 e^+ e^-$	$B^+ \rightarrow K^+ e^+ e^-$
R_0^{00}	0.30	0.77	0.10	0.42
R_2^{00}	0.69	0.61	0.03	0.57
$ \cos \theta_{TBTz} $	0.03	0.41	0.90	0.02
CC_3	0.77	0.65	1.00	0.69
T_{ROE}	0.93	0.01	0.99	0.01
M_{miss}^2	0.26	0.63	0.12	0.42
T	0.08	0.83	0.00	0.03
CC_1	0.08	0.13	0.39	0.68
E_T	0.73	0.82	0.96	0.87
H_{m2}^{s0}	0.98	0.69	0.43	0.64
H_{n2}^{s0}	0.17	0.98	0.63	0.84
R_2^{00}	0.59	0.07	0.38	0.02
R_2	0.07	0.29	0.00	0.34
CC_2	0.25	0.92	0.97	1.00
H_{c2}^{s0}	0.38	0.58	0.65	0.43
$\Delta z(\ell^+, \ell^-)$	0.23	0.36	0.30	0.42
$ \cos \theta_{TBT0} $	0.58	0.52	0.12	0.80
R_4^{00}	0.91	0.96	0.66	0.60

hyperparameters used for each BDT are given in table 4.14, and the outputs of the classifier for the validation data are shown in figure 4.16.

Table 4.12. Total number of available samples per decay channel for the $q\bar{q}$ classifier. Half of them correspond to signal events, and half of them to continuum ones.

$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	$B^+ \rightarrow K^+ \mu^+ \mu^-$	$B^0 \rightarrow K_S^0 e^+ e^-$	$B^+ \rightarrow K^+ e^+ e^-$
2×5439	2×16950	2×5217	2×16556

Table 4.13. Variables used in the continuum suppression BDTs

Variable	in $B \rightarrow K\mu^+\mu^-$ fBDT	in $B \rightarrow Ke^+e^-$ fBDT
CC_3	yes	yes
$ \cos\theta_{TBTO} $	yes	yes
M_{miss}^2	yes	yes
R_2	yes	yes
$ \cos\theta_{TBTz} $	yes	no
CC_2	yes	yes
T_{ROE}	yes	no
H_{n2}^{s0}	yes	yes
E_T	no	yes
T	yes	yes
$\Delta z(\ell^+, \ell^-)$	yes	yes
H_{c2}^{s0}	no	yes
R_0^{00}	yes	yes
R_2^{00}	yes	yes
CC_1	yes	yes
R_4^{00}	yes	no
R_2^{00}	yes	yes

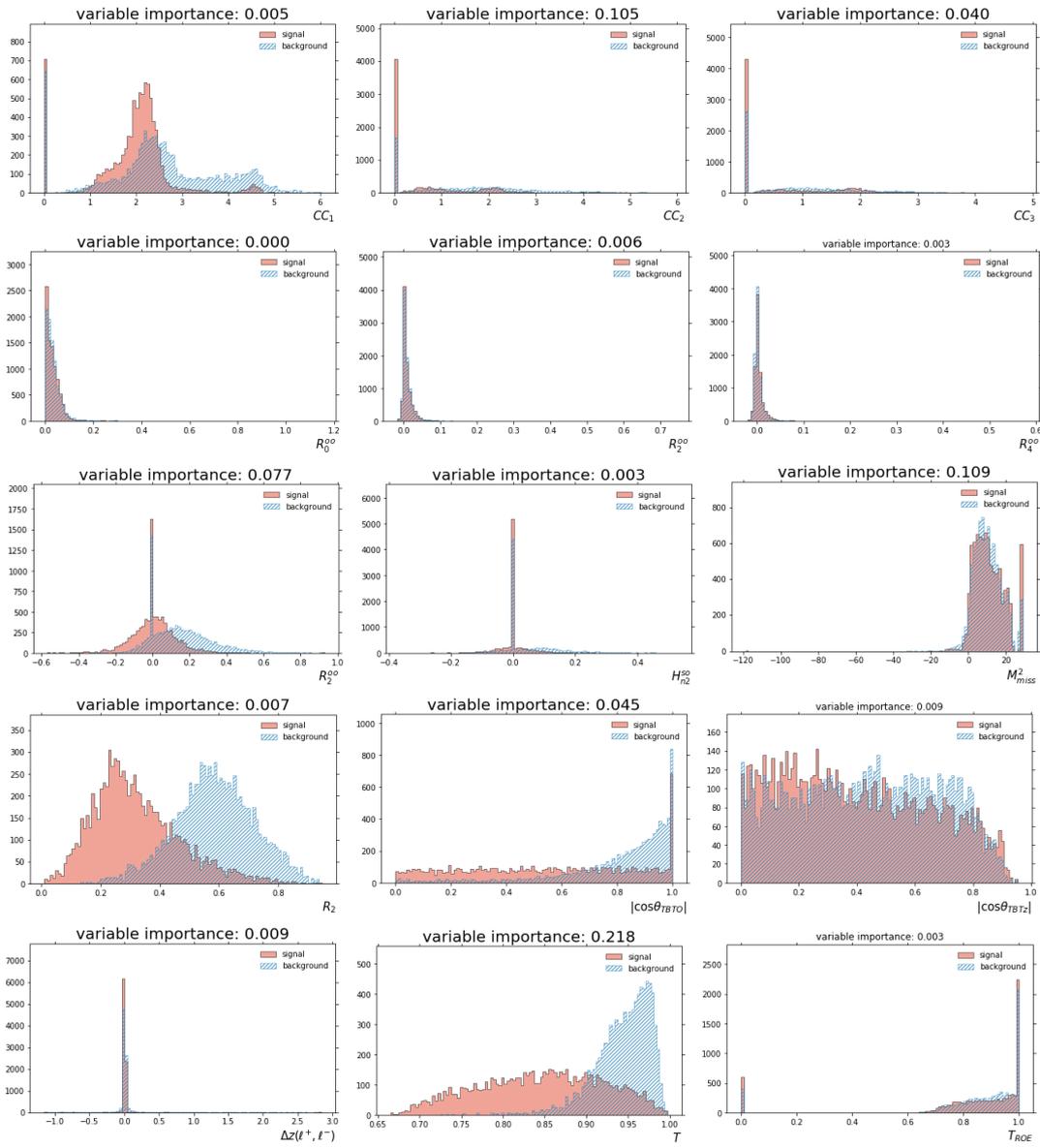


Fig. 4.14 Distribution of the continuum suppression classifier input variables, for the muon decay modes. The number above each plot is the relative importance of the variable for the classifier, measured as the average reduction of the Gini impurity achieved after a cut in the given variable is applied.

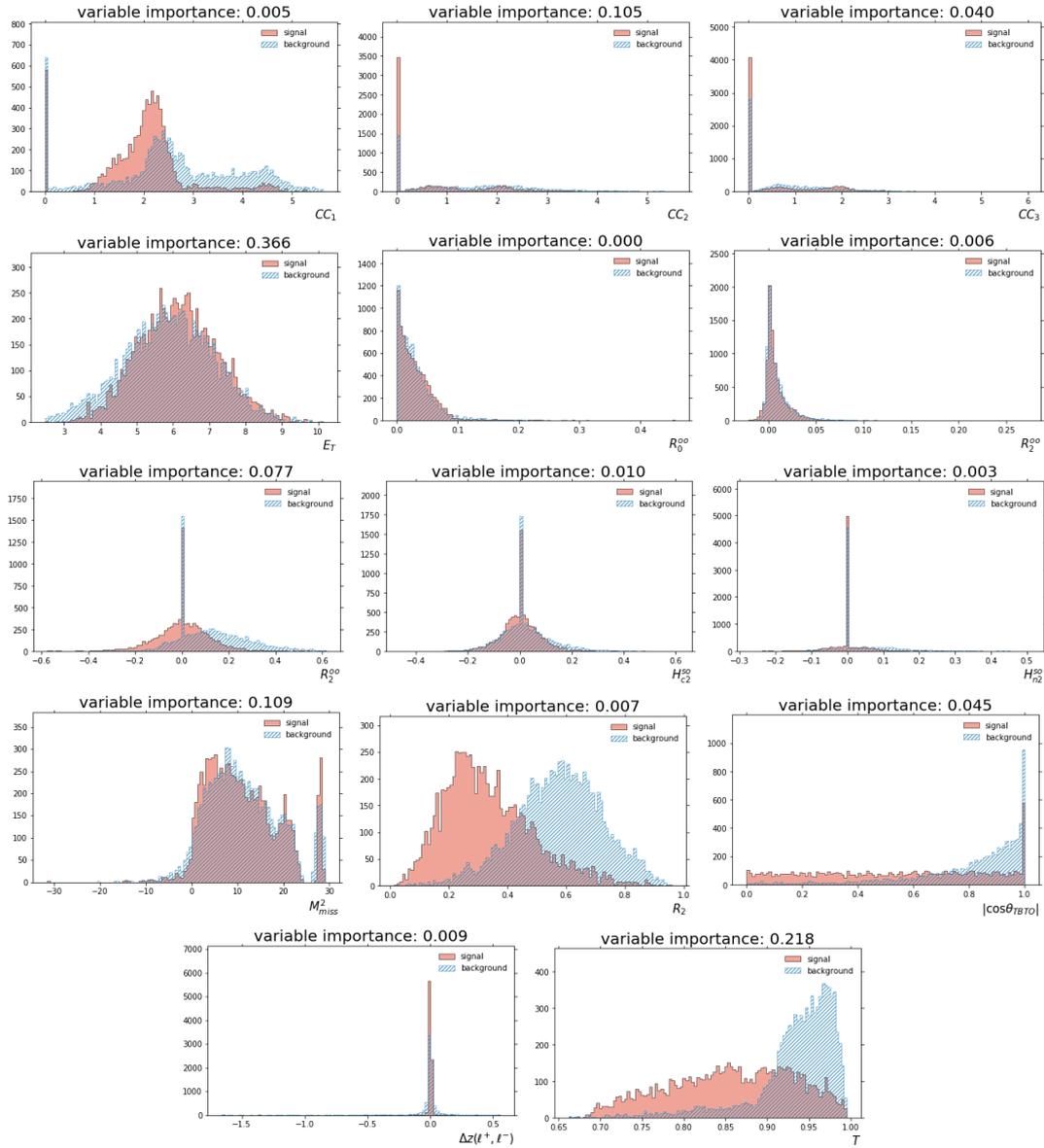


Fig. 4.15 Distribution of the continuum suppression classifier input variables, for the electron decay modes. The number above each plot is the relative importance of the variable for the classifier, measured as the average reduction of the Gini impurity achieved after a cut in the given variable is applied.

Figure 4.17 shows the background efficiency for the validation dataset in three different regions: the sideband ($m_{bc} < 5.27$ GeV); the skirts of the signal region, or signal region 1 ($5.27 \text{ GeV} \leq m_{bc} \leq 5.275 \text{ GeV}$ and $m_{bc} \geq 5.285 \text{ GeV}$); and the central part of the signal region, or signal region 2 ($5.275 < m_{bc} < 5.285 \text{ GeV}$). The three regions show similar efficiency values for most of the possible cuts; this proves that

the flatness loss is effective in avoiding boosting the background yield in the signal region.

As a check for possible overfitting, figure 4.18 shows the Receiver-Over-Operator (ROC) curve of the classifiers' outputs for the different training samples. Since the curves for the training and validation data are very similar, it can be concluded that the BDTs generalize well to unobserved data, and hence there is no overfitting.

In order to increase the number of available data for a second set multivariate classifiers, we do not apply a cut on the outputs of the continuum BDTs yet. Instead, we optimize the cuts for both the continuum and the $B\bar{B}$ classifiers at the same time. We thus proceed to present the training procedure for $B\bar{B}$ background suppression.

Table 4.14. Continuum suppression BDT hyperparameters

Hyperparameter	$B \rightarrow Ke^+e^-$ BDT	$B \rightarrow K\mu^+\mu^-$ BDT
Number of trees	600	600
Learning rate	0.05	0.05
Tree depth	3	3
Flatness penalization weight	7	7

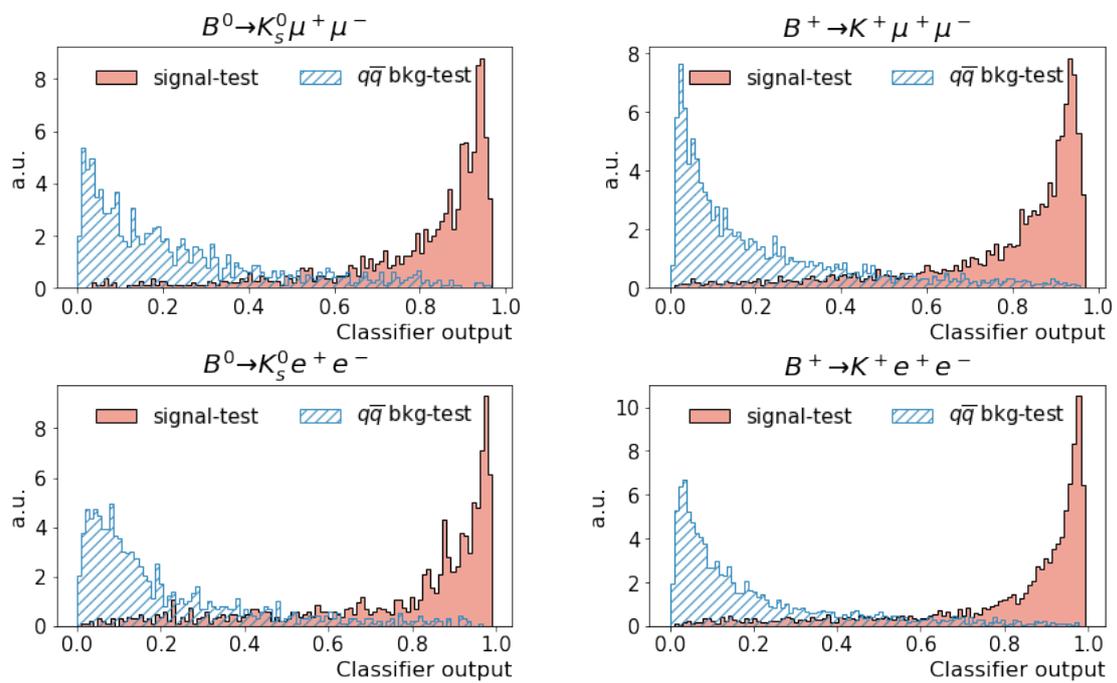


Fig. 4.16 Distribution of the continuum background classifier output for the validation samples.

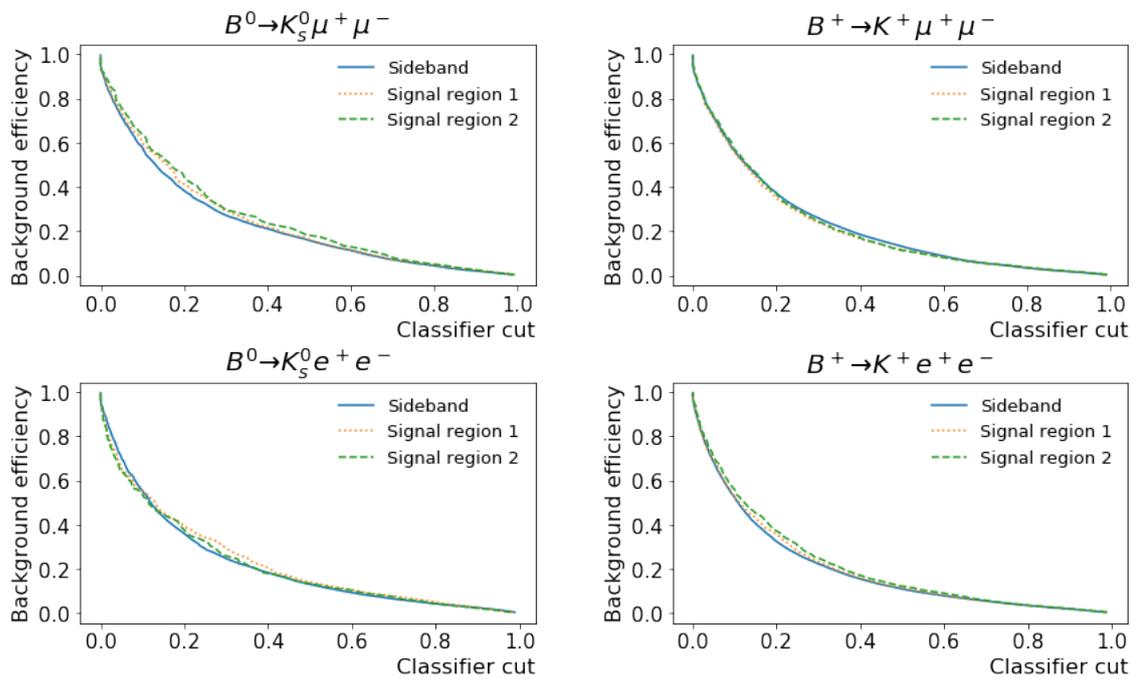


Fig. 4.17 Background efficiency in the validation dataset as a function of the cut value on the continuum classifier output. The three regions correspond to events for which $m_{bc} < 5.27$ GeV (blue, solid line), $5.27 \text{ GeV} \leq m_{bc} \leq 5.275 \text{ GeV}$ and $m_{bc} > 5.285 \text{ GeV}$ (yellow, dotted line) and $5.275 \text{ GeV} < m_{bc} < 5.285 \text{ GeV}$ (green, dashed line). A cut points where the lines are close implies a uniform selection over m_{bc} .

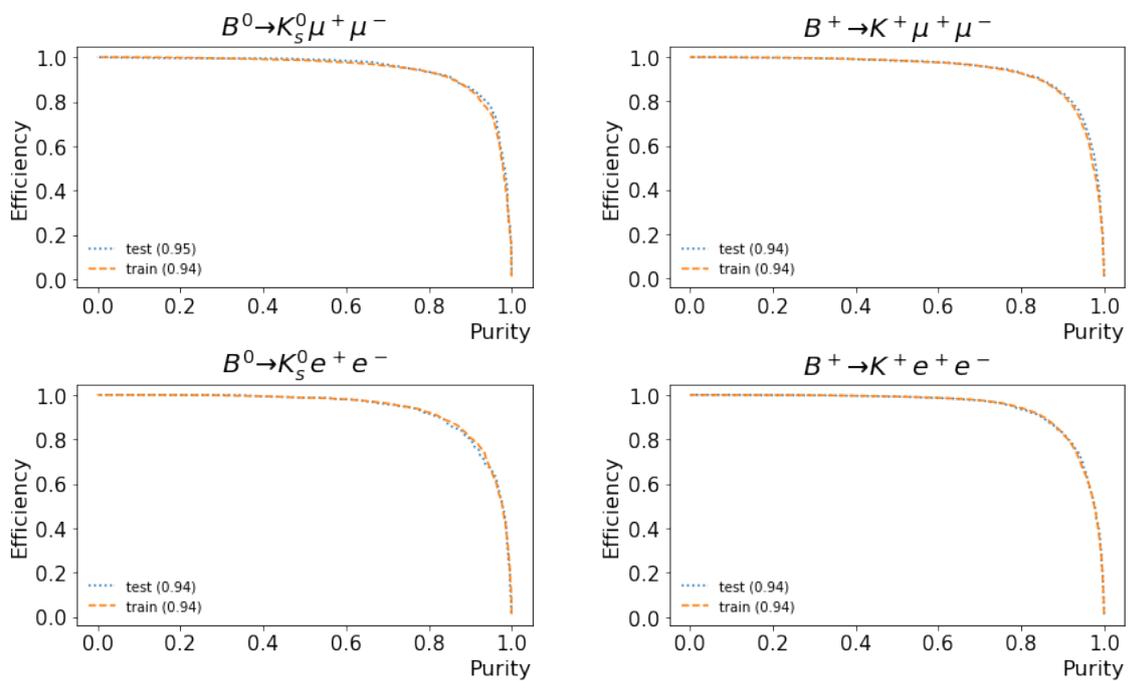


Fig. 4.18 ROC curves of the continuum suppression classifier for the validation sample of each decay mode. The number in parentheses corresponds to the area under the curve (AUC). The closer it is to one, the better the performance of the classifier.

4.3.3 $B\bar{B}$ Suppression

In order to suppress the background originating in other B meson decays, a classifier is trained for each decay mode. The sizes of the datasets are shown in table 4.15. We

Table 4.15. Total number of available samples per decay channel for the $B\bar{B}$ classifier. Half of them correspond to signal events, and half of them to continuum ones.

$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	$B^+ \rightarrow K^+ \mu^+ \mu^-$	$B^0 \rightarrow K_S^0 e^+ e^-$	$B^+ \rightarrow K^+ e^+ e^-$
2×1428	2×4414	2×2276	2×8836

employ some of the variables described before. Also, in order to reject semileptonic events, which constitute the main source of generic $B\bar{B}$ background, variables that are sensitive to the presence of neutrinos are fed as inputs to the classifiers. These include the missing mass of the event and the ROE, the missing energy of the event and the ROE, and the value of ΔE for the B candidate. In order to reject random combinations of final state particles, we include $\Delta z (\ell^+ \ell^-)$ and some of the thrust-related variables presented previously. Table 4.16 depicts all the variables considered, together with the p -value of the Kolmogorov-Smirnoff test between the MC and the real data distributions.

Table 4.16. p -value of the KS test for the $B\bar{B}$ Suppression MVA training variables between MC and experimental data in the charmonium sidebands

Variable	KS test for			
	$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	$B^+ \rightarrow K^+ \mu^+ \mu^-$	$B^0 \rightarrow K_S^0 e^+ e^-$	$B^+ \rightarrow K^+ e^+ e^-$
$ \cos \theta_{TBTz} $	0.03	0.41	0.90	0.02
T_{ROE}	0.93	0.01	0.99	0.01
M_{miss}^2	0.26	0.63	0.12	0.42
E_T	0.73	0.82	0.96	0.87
$\Delta z (\ell^+, \ell^-)$	0.23	0.36	0.30	0.42
$ \cos \theta_{TBT0} $	0.58	0.52	0.12	0.80
ΔE	0.00	0.00	0.01	0.01
$E_{miss} (ROE)$	0.78	0.83	0.45	0.57
$M_{miss}^2 (ROE)$	0.51	0.40	0.35	0.20

As in the previous section, we exclude those variables with a p -value less than 0.05. The final sets of variables used for each classifier are listed in table 4.17.

In comparison with the continuum classifier, in the $B\bar{B}$ the presence of peaking background is expected, mostly due to charmonium decays leaking through the veto regions and $B \rightarrow K\pi + \pi^-$ events, where both pions are misidentified as leptons. We relax the flatness requirement by assigning a lower weight in comparison with the classifiers from the previous section. The hyperparameter values are presented in table 4.18.

Table 4.17. Variables used as inputs for the $B\bar{B}$ BDTs. The ROE variables are calculated used only information from the rest of the event.

Variable	$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	$B^+ \rightarrow K^+ \mu^+ \mu^-$	$B^0 \rightarrow K_S^0 e^+ e^-$	$B^+ \rightarrow K^+ e^+ e^-$
T_{ROE}	yes	yes	yes	no
$ \cos\theta_{TBTz} $	yes	yes	no	yes
$E_{miss}(ROE)$	yes	yes	yes	yes
M_{miss}^2	yes	yes	yes	yes
$ \cos\theta_{TBTO} $	yes	yes	yes	yes
E_T	yes	yes	yes	yes
ΔE	yes	yes	yes	yes
$\Delta z(\ell^+, \ell^-)$	yes	yes	yes	yes
$M_{miss}^2(ROE)$	yes	yes	yes	yes

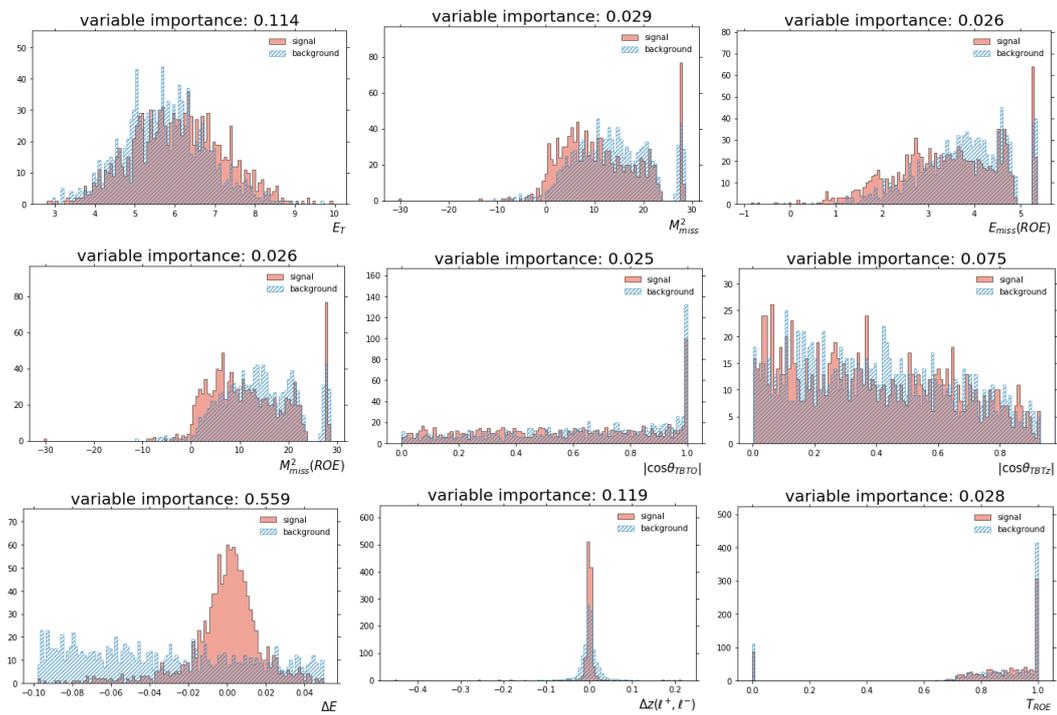


Fig. 4.19 Distribution of the $B\bar{B}$ suppression classifier input variables, for the $B^0 \rightarrow K_S^0 \mu^+ \mu^-$ decay mode. The number above each plot is the relative importance of the variable for the classifier, measured as the average reduction of the Gini impurity achieved after a cut in the given variable is applied.

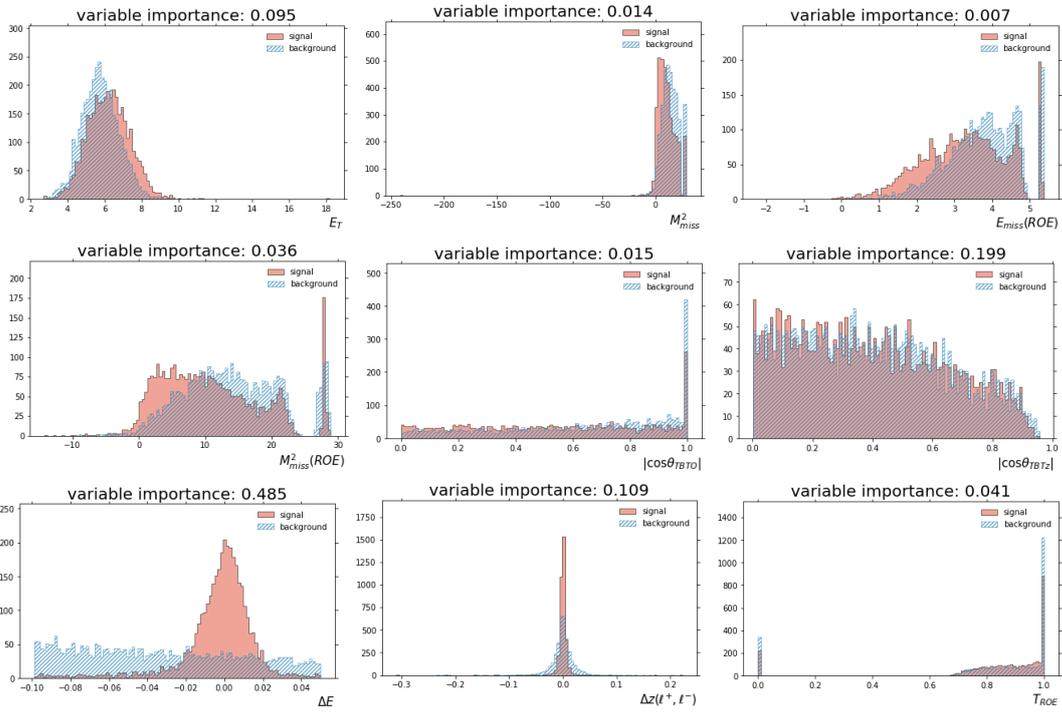


Fig. 4.20 Distribution of the $B\bar{B}$ suppression classifier input variables, for the $B^+ \rightarrow K^+ \mu^+ \mu^-$ decay mode. The number above each plot is the relative importance of the variable for the classifier, measured as the average reduction of the Gini impurity achieved after a cut in the given variable is applied.

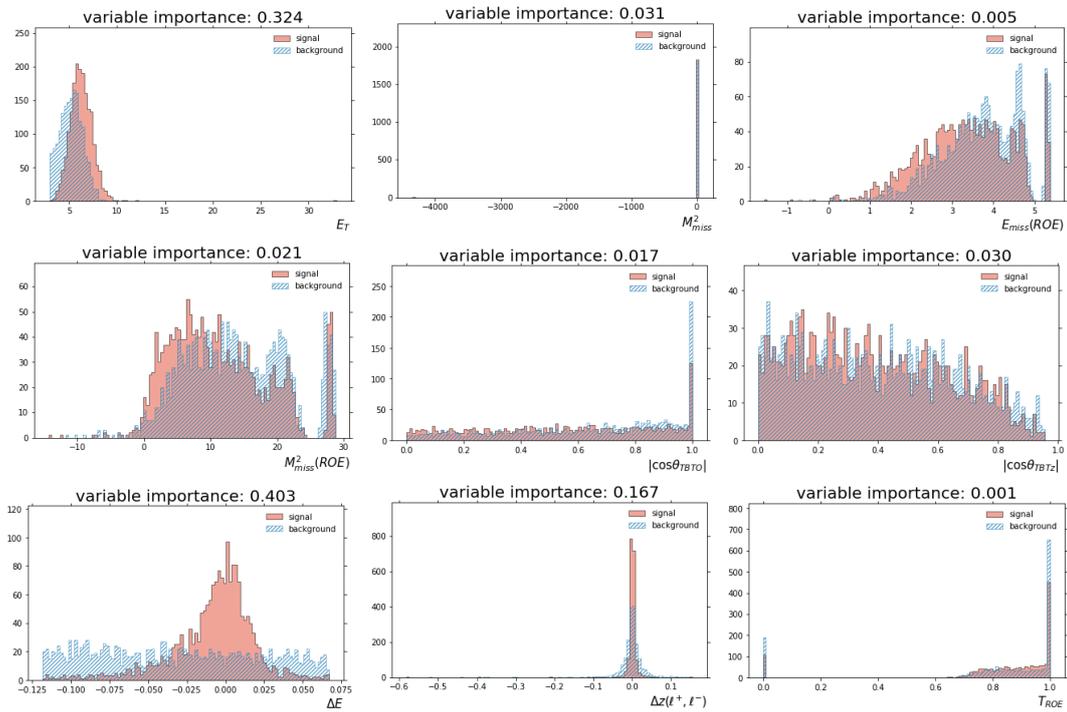


Fig. 4.21 Distribution of the $B\bar{B}$ suppression classifier input variables, for the $B^0 \rightarrow K_S^0 e^+ e^-$ decay mode. The number above each plot is the relative importance of the variable for the classifier, measured as the average reduction of the Gini impurity achieved after a cut in the given variable is applied.

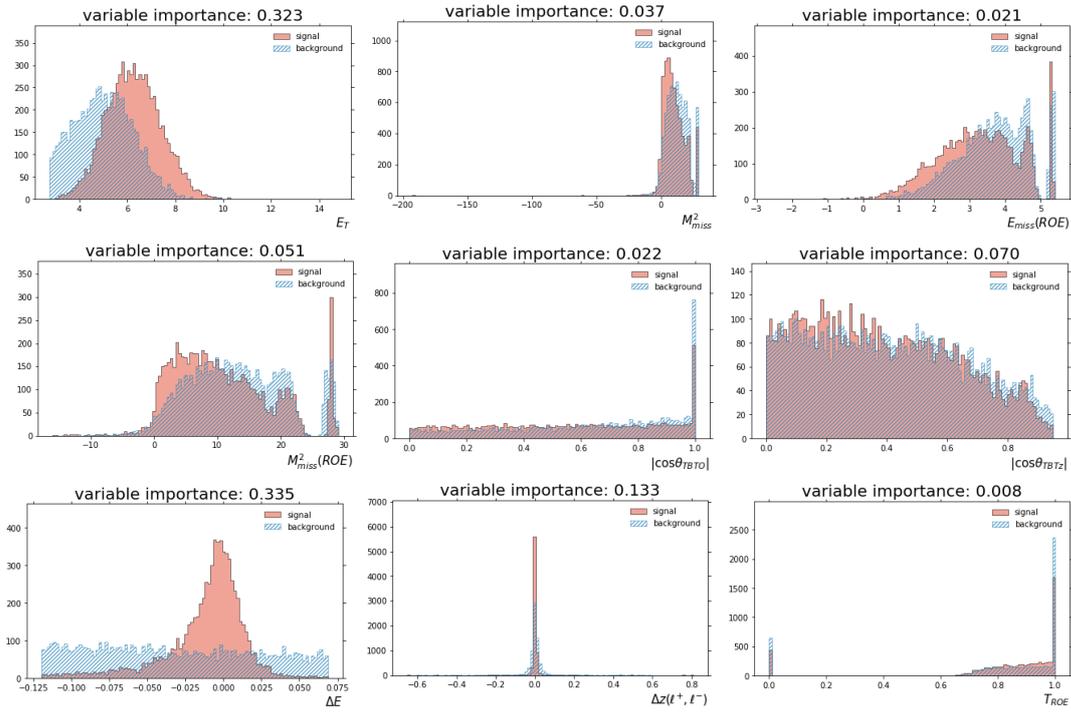
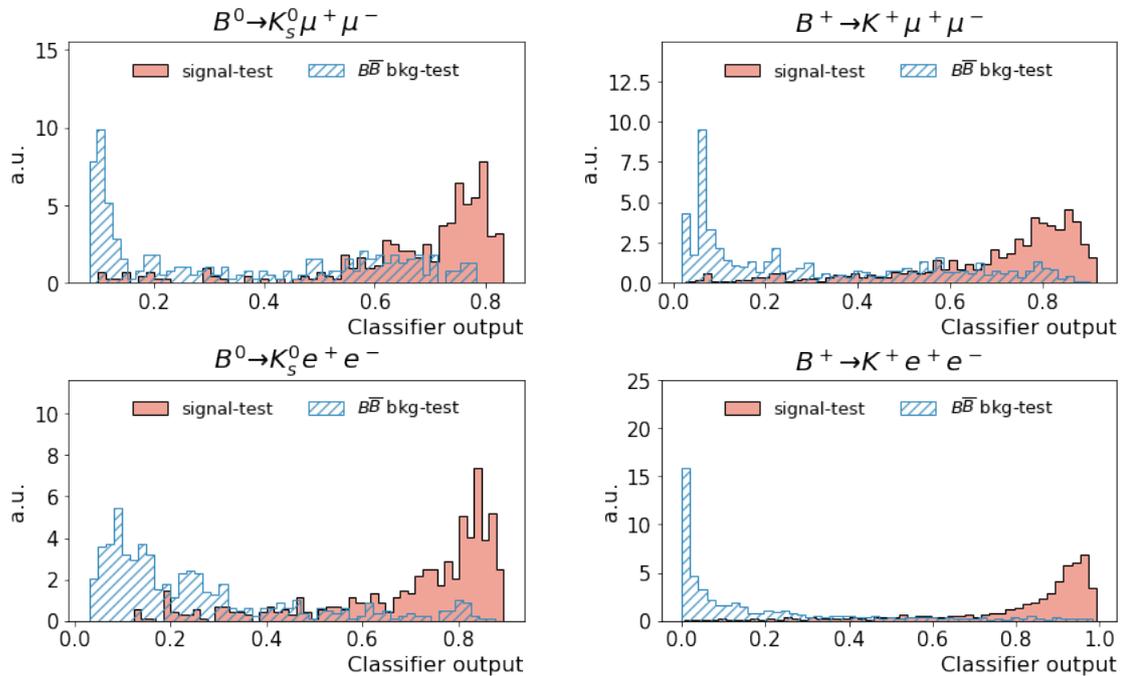


Fig. 4.22 Distribution of the $B\bar{B}$ suppression classifier input variables, for the $B^+ \rightarrow K^+ e^+ e^-$ decay mode. The number above each plot is the relative importance of the variable for the classifier, measured as the average reduction of the Gini impurity achieved after a cut in the given variable is applied.

Table 4.18. $B\bar{B}$ suppression BDT hyperparameters

Hyperparameter	μ modes		e modes	
	neutral	charged	neutral	charged
Number of trees	250	300	300	400
Learning rate	0.005	0.01	0.01	0.05
Tree depth	3	3	3	3
Flatness weight	4	11.5	0.1	0.1

Figure 4.23 shows the classifier output distribution for the four decays of interest. The background efficiency, depicted in figure 4.24, presents variations for events inside and outside the signal region in all channels, with bigger differences in the muon channels, where the peaking background is expected to be larger. Finally, figure 4.25 presents the ROC curves for the four classifiers.

Fig. 4.23 Distribution of the $B\bar{B}$ background classifier output for the validation samples.

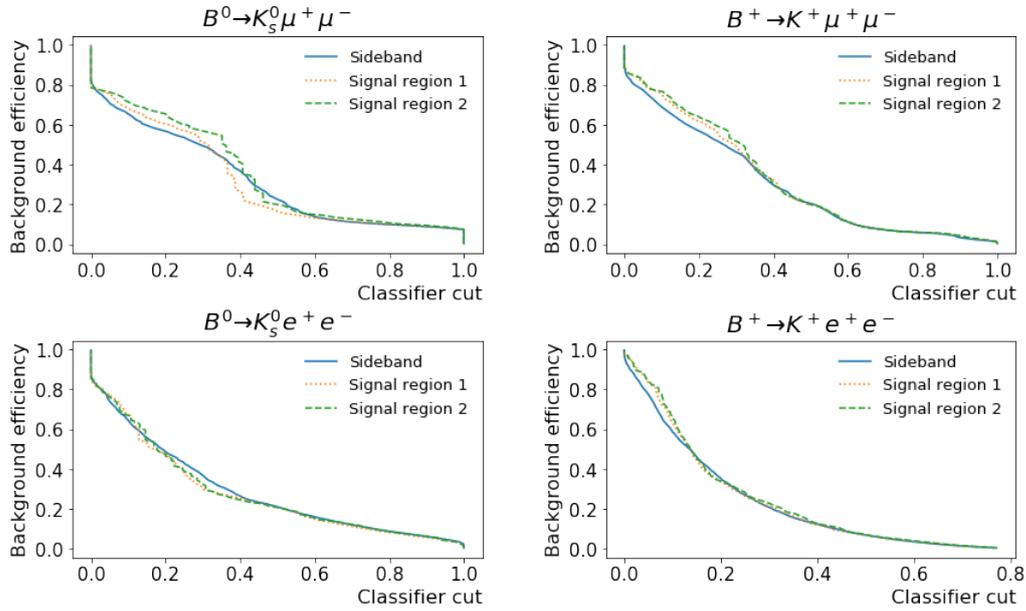


Fig. 4.24 Background efficiency as a function of the cut value in the validation dataset on the $B\bar{B}$ classifier output. The three regions correspond to events for which $m_{bc} < 5.27$ GeV (blue, solid line), $5.27 \text{ GeV} \leq m_{bc} \leq 5.275$ GeV (yellow, dotted line) and $5.275 \text{ GeV} < m_{bc} < 5.285$ GeV (green, dashed line). A cut points where the lines are close implies a uniform selection over m_{bc} .

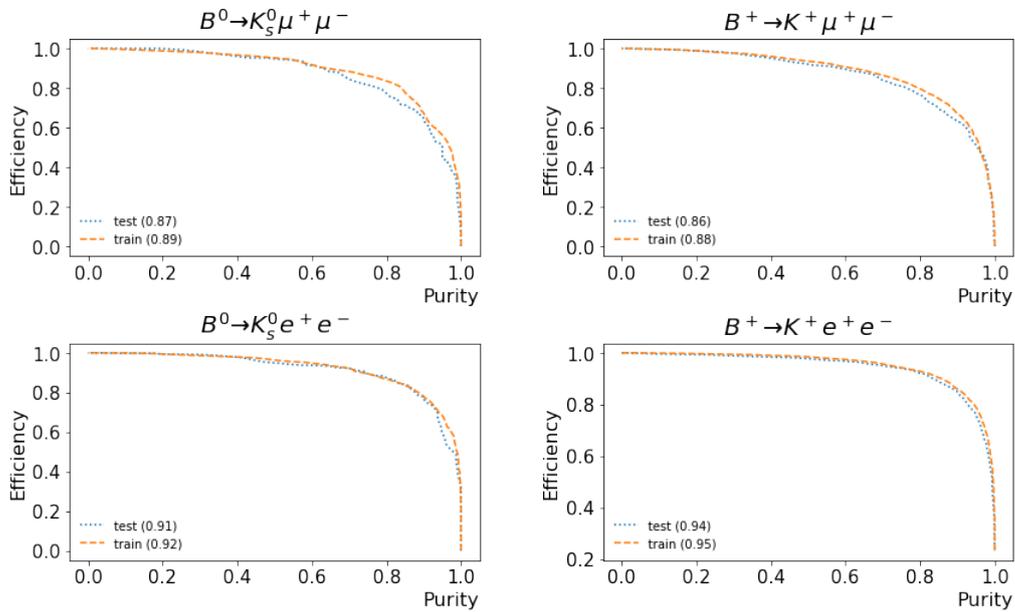


Fig. 4.25 ROC curves of the $B\bar{B}$ suppression classifier for the validation samples of each decay mode. The number in parentheses corresponds to the area under the curve (AUC).

4.3.4 Selection of the Classifiers' Cuts

In order to select the best cuts for each pair of classifiers in each decay mode, we maximize once again the figure of merit

$$FOM = \frac{N_S}{\sqrt{N_S + N_B}} \quad (4.11)$$

where N_S and N_B are the number of signal and background events, respectively, that survive the cuts, estimated from the original 2 ab^{-1} MC data after the reconstruction and veto cuts have been applied. The two-dimensional FOM distributions are depicted in figure 4.26. The selected cuts are summarized in table 4.19.

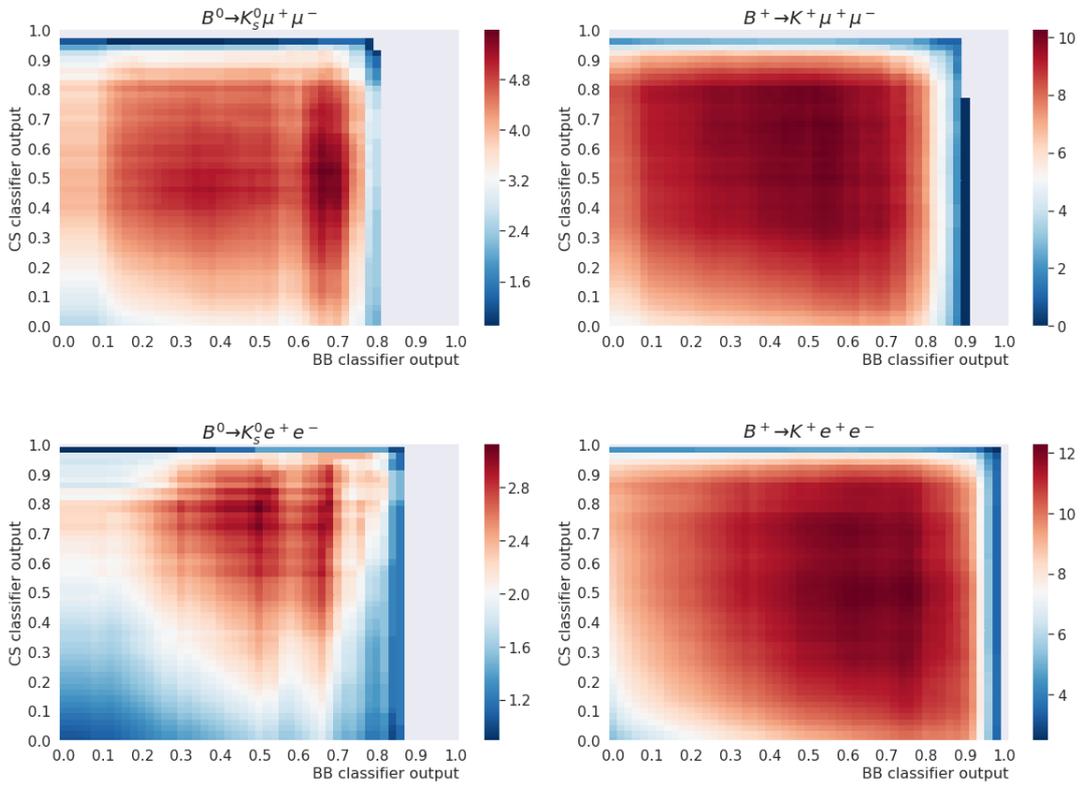


Fig. 4.26 FOM distribution for the multivariate classifier cuts in each decay mode.

Table 4.19. Selected cuts on the MVA classifier outputs

Classifier	$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	$B^+ \rightarrow K^+ \mu^+ \mu^-$	$B^0 \rightarrow K_S^0 e^+ e^-$	$B^+ \rightarrow K^+ e^+ e^-$
Continuum	0.52	0.68	0.78	0.52
$B\bar{B}$	0.66	0.52	0.5	0.76

In the case of multiple candidates per event, we select the one with the lowest value of $|\Delta E|$. The final m_{bc} distributions are shown in figure 4.27, for the luminosity of the MC sample. The cutflows for each decay mode are shown in tables 4.20 to 4.23. In the tables, ε corresponds to the overall efficiency, and ε_{SR} to the efficiency inside the signal region; the signal to noise ratio, N_S/N_B , and the pseudo-significances, are also calculated using only events in the signal region. The pseudo-significances are first obtained for the MC luminosity, and extrapolated to a luminosity of 11.53 fb^{-1} (corresponding to the luminosity of the dataset in this study) and 50 ab^{-1} (the target luminosity of the Belle II collaboration). These values are summed up in table 4.24.

As a sanity check, the MVA classifier distributions for data and MC inside the charmonium sidebands are shown in figure 4.28.

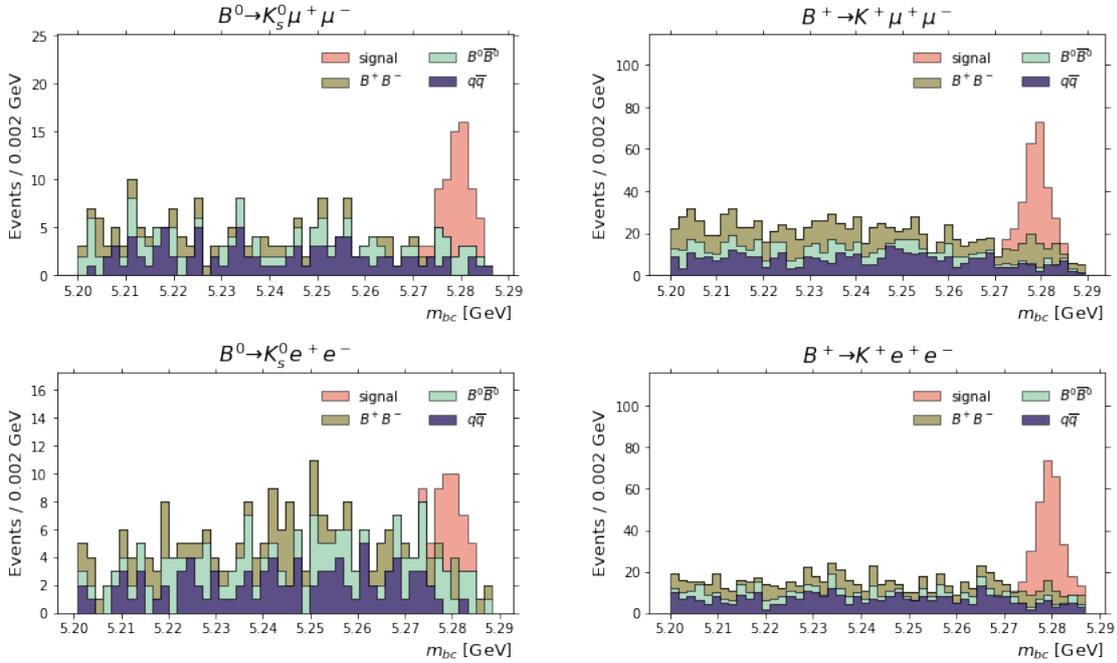


Fig. 4.27 Final m_{bc} distribution for each decay mode for 2 ab^{-1} MC data.

Table 4.20. $B^0 \rightarrow K^0 \mu^+ \mu^-$ cutflow table for 2 ab^{-1} MC data

Cut	Signal	$B^+ B^-$	$B^0 \bar{B}^0$	$q\bar{q}$	N_S/N_B	$\frac{N_S}{\sqrt{N_S+N_B}}$	ϵ [%]	ϵ_{SR} [%]
reco.	101	20764	38534	42680	0.002	0.499	13.55	13.38
ΔE	95	849	19058	6247	0.005	0.685	12.72	12.70
Dalitz	95	849	19058	6247	0.005	0.685	12.72	12.70
J/ψ	87	719	3738	5898	0.023	1.385	11.85	11.84
$\psi(2S)$	84	632	845	5553	0.084	2.557	11.01	11.00
D veto	83	611	817	5439	0.086	2.570	10.70	10.69
CS	72	373	588	709	0.293	4.038	9.42	9.42
$B\bar{B}S$	48	40	68	96	1.846	5.580	6.12	6.12
best cand.	48	40	68	95	1.920	5.618	6.10	6.10

Table 4.21. $B^+ \rightarrow K^+ \mu^+ \mu^-$ cutflow table for 2 ab^{-1} MC data

Cut	Signal	$B^+ B^-$	$B^0 \bar{B}^0$	$q\bar{q}$	N_S/N_B	$\frac{N_S}{\sqrt{N_S+N_B}}$	ϵ [%]	ϵ_{SR} [%]
reco.	357	125746	46982	138316	0.003	1.032	40.94	40.38
ΔE	344	61641	2047	19769	0.006	1.388	38.65	38.61
J/ψ	319	7767	1725	18645	0.043	3.607	36.05	36.01
$\psi(2S)$	293	3323	1489	17530	0.091	4.936	33.57	33.53
D veto	287	3003	1411	16950	0.097	5.041	32.63	32.59
CS	208	1514	711	1047	0.481	8.222	25.34	25.31
$B\bar{B}S$	181	433	199	371	1.382	10.247	21.82	21.81
best cand.	181	432	199	369	1.382	10.247	21.80	21.79

Table 4.22. $B^0 \rightarrow K^0 e^+ e^-$ cutflow table for 2 ab^{-1} MC data

Cut	Signal	$B^+ B^-$	$B^0 \bar{B}^0$	$q\bar{q}$	N_S/N_B	$\frac{N_S}{\sqrt{N_S+N_B}}$	ϵ [%]	ϵ_{SR} [%]
reco.	46	21095	38762	44385	0.001	0.228	14.04	13.67
ΔE	44	1520	18051	7663	0.002	0.329	12.69	12.65
Dalitz	44	1519	18043	6242	0.002	0.331	12.67	12.64
J/ψ	40	1237	3878	5746	0.011	0.667	11.33	11.31
$\psi(2S)$	36	1050	1307	5313	0.032	1.063	10.23	10.21
D veto	35	1012	1264	5217	0.032	1.047	9.94	9.91
CS	27	182	272	170	0.231	2.250	6.62	6.61
$B\bar{B}S$	25	67	86	92	0.641	3.125	5.83	5.83
best cand.	25	67	85	92	0.658	3.150	5.81	5.81

Table 4.23. $B^+ \rightarrow K^+ e^+ e^-$ cutflow table for 2 ab^{-1} MC data

Cut	Signal	$B^+ B^-$	$B^0 \bar{B}^0$	$q\bar{q}$	N_S/N_B	$\frac{N_S}{\sqrt{N_S+N_B}}$	ϵ [%]	ϵ_{SR} [%]
reco.	465	129686	52531	142388	0.004	1.334	42.27	41.16
ΔE	427	58860	4566	25006	0.008	1.786	38.54	38.44
Dalitz	426	58830	4565	19853	0.008	1.795	38.50	38.40
J/ψ	380	10051	3852	18357	0.050	4.251	34.47	34.39
$\psi(2S)$	341	5802	3443	16954	0.091	5.318	31.15	31.08
D veto	333	5543	3293	16556	0.092	5.277	30.24	30.17
CS	279	2339	1217	1699	0.353	8.529	25.78	25.72
$B\bar{B}S$	220	246	146	336	2.178	12.279	19.51	19.50
best cand.	219	246	145	336	2.168	12.242	19.49	19.47

Table 4.24. Pseudo-significance for the decays of interest at the current luminosity

Decay	at 2 ab ⁻¹ (MC luminosity)	at 11.53 fb ⁻¹ (this study)	at 50 ab ⁻¹ (target)
$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	5.62	0.42	28.1
$B^+ \rightarrow K^+ \mu^+ \mu^-$	10.25	0.77	51.25
$B \rightarrow K \mu^+ \mu^-$	11.41	0.86	57.05
$B^0 \rightarrow K_S^0 e^+ e^-$	3.15	0.24	15.75
$B^+ \rightarrow K^+ e^+ e^-$	12.24	0.93	61.2
$B \rightarrow K e^+ e^-$	12.98	0.98	64.9
$B \rightarrow K \ell^+ \ell^-$	17.12	1.29	85.6

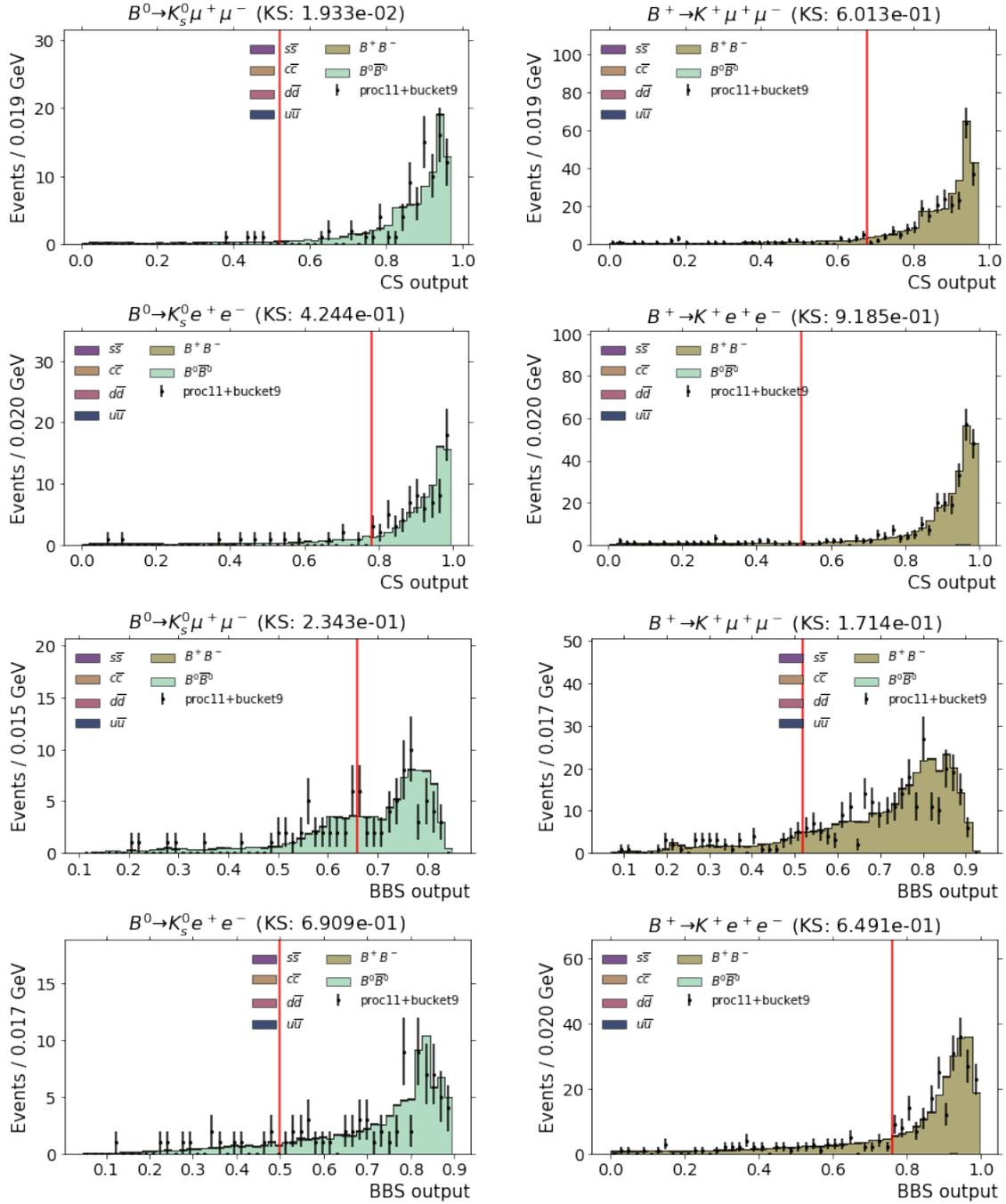


Fig. 4.28 Continuum (top) and $B\bar{B}$ (bottom) classifiers' distributions for data and MC inside the charmonium sidebands. The red vertical lines correspond to the cut values, and the number in parenthesis is the p -value of the Kolmogorov-Smirnoff test. The MC sample has been scaled to the luminosity of the dataset, which is 11.53 fb^{-1} .

Chapter 5

Data Analysis Preliminaries

This chapter deals with the data analysis previous to the signal yield extraction, and it is divided in three parts: in the first one, the dataset used for this analysis is explained. In the second one, the systematic uncertainties and efficiency corrections related to tracking, lepton ID, Kaon ID, K_S^0 reconstruction and the MVA cuts are obtained. In the final section, the corrected efficiencies, together with their systematic and statistical errors, are reported.

5.1 Dataset

The real data sample used in this analysis consists of events recorded at a CM Energy close to the $Y(4s)$ resonance, and stored in `proc11` and `bucket9`, which correspond to a total luminosity of 11.53 fb^{-1} ; table 5.1 describes in detail the experiments and runs included in the dataset¹, their corresponding integrated luminosities, and the version of the `basf2` software used for their processing. The errors in the luminosities are statistical; a 1% systematic error is also expected, and included in the final computations. Contrary to the MC data, the real experiment is missing a large portion of the second layer of the PXD; this is expected to have a minor impact the K_S^0 reconstruction efficiency, as there are no strict cuts on the pion tracks, and we expect to take this difference into account in the efficiency correction factor due to the K_S^0 reconstruction. The efficiency detection for the other particles are more likely to remain unaffected, since the PID and momentum cuts assure that their tracks are measured by other subdetectors.

¹within an experiment, only runs performed with the proper detector configuration are considered

Table 5.1. Real dataset used for this analysis. All luminosities were calculated using Bhabha events

	experiment	runs	luminosity (pb^{-1})	basf2 version
proc11	exp7	909-4120	425.5 ± 0.3	release-04-02-02
		43-1022		
	exp8	1036-1554	4597.4 ± 0.9	release-04-02-02
		1835-3123		
	exp10	3130-5902	3741.3 ± 1.1	release-04-02-02
bucket9	exp12	0-1969	2741.3 ± 1.1	release-04-02-02

An online skim is applied to the events in this data sample, known as the High Level Trigger hadron skim (hlt_hadron). It consists of a Bhabha event veto and a cut of at least three good tracks in the event. The good tracks must satisfy $p_T > 0.2$ GeV, $|d_0| < 2$ cm, and $z_0 < 4$ cm, where p_T is the transverse momentum of the track and d_0 and z_0 its helix parameters.

5.2 Efficiency Corrections

Since the efficiencies calculated in the previous chapter were obtained from MC samples, they must be corrected in order to account for any discrepancy between the simulation model and the real experiment.

5.2.1 Track Finding

The tracking efficiency (how many of the generated tracks are properly reconstructed by the detector) was evaluated in [71] using $e^+e^- \rightarrow \tau^+\tau^-$ events. One of the *tau* leptons decays leptonically ($\tau \rightarrow \ell\nu_\ell\nu_\tau$, with $\ell = e, \mu$) while the other decays hadronically ($\tau \rightarrow 3\pi^\pm\nu_\tau + n\pi^0$). In this type of decays, 4 charged particles are produced, and hence 4 tracks are expected. Three (tagging) tracks, with a total charge of ± 1 , are used to tag these decays. The existence of the fourth (probing) track can be inferred by charge conservation. The track associated with the τ decaying leptonically is referred to as the 1-prong track, whereas the tracks associated with the hadronically decaying τ are termed 3-prong tracks. The tracking efficiency is

measured as

$$\varepsilon_{\text{track}} \times A = \frac{N_4}{N_4 + N_3} \quad (5.1)$$

where N_4 (N_3) is the number of events where 4 (3) tracks were reconstructed, and A is a factor that takes into account the acceptance of the Belle II detector for the fourth track. The data-MC discrepancy is defined as

$$\delta_{\text{track}} = 1 - \frac{\varepsilon_{\text{track}}^{\text{data}}}{\varepsilon_{\text{track}}^{\text{MC}}}. \quad (5.2)$$

In order to properly reconstruct these decays, the event must fire the ECL trigger *hie*, which has an energy threshold of 1 GeV and a Bhabha veto requirement. Different classes of tagging tracks are defined, according to the particle they associate to: electrons, muons or pions. The probing track is assumed to always correspond to a pion. Depending on the channel (electronic or muonic) and the number of tracks detected, events are required to satisfy different track multiplicity constraints for each class. The definitions for each track class are summarized in table 5.2, and the multiplicity criteria in table 5.3.

Table 5.2. Track classes

	Probe (π^\pm)	Tag pion	Tag electron	Tag muon
p_T [MeV]	-	> 200	> 200	> 200
$ z_0 $ [cm]	< 5	< 5	< 5	< 5
$ d_0 $ [cm]	< 2	< 2	< 2	< 2
$E_{\text{cluster}}^{\text{ECL}}/p$	< 0.8	< 0.6	(0.8, 1.2)	< 0.4
$E_{\text{cluster}}^{\text{ECL}}$ [MeV]	-	> 0	-	(0, 400)

Table 5.3. Track multiplicity criteria

	N_π^{probe}	N_π^{tag}	N_μ^{probe}	N_e^{probe}
Electron channel, 4-tracks	3	≥ 2	1	-
Electron channel, 3-tracks	2	2	1	< 5
Muon channel, 4-tracks	4	≥ 3	-	≥ 1
Muon channel, 3-tracks	3	3	-	≥ 1

In order to suppress background arising from continuum events and radiative dilepton processes, the next additional cuts are applied:

- $\cos\theta_{12} < -0.5$, where θ_{12} is the opening angle in the CMS between the 1-prong track and each of the two tag-tracks in the 3-prong τ decay. Since in the CMS both τ are expected to be back-to-back, their jets will also have a large angular separation.
- When the two tagging tracks in the 3-prong τ decay have opposite charge (OS), a requirement of $|m_{\pi\pi} - m_{\rho}| < 100$ MeV is applied, which selects hadronic τ decays with an intermediate ρ meson ($\rho \rightarrow \pi^+ \pi^-$); when they have the same charge (SS), the requirement is loosened to $300 \text{ MeV} < m_{\pi\pi} < m_{\tau}$.
- The 1-prong track momentum in the CMS is required to be in the range $(0.2E_{\text{beam}}, 0.8E_{\text{beam}})$. The upper bound reduces the radiative dilepton background contribution, and the lower one diminishes the number continuum events.
- For the electron-OS channels, $M_{\text{miss}}^2 < 20 \text{ GeV}^2$, and the polar angle in the CMS of the missing momentum must be below 40° or above 135° .

Finally, the number of MC events is scaled down to match the data luminosity, and the number of 3-track and 4-track events is obtained for all channels. The values of $\varepsilon_{\text{track}}$ are calculated for data and MC, and δ_{track} is obtained from them.

A calibration factor for the efficiency estimator (equation 5.1) is included, in order to take into account the overlooked dependence on the number and the electric charge of the reconstructed tracks. This calibration factor is calculated as follows: a new set of MC events is generated, with an *a priori* track missing probability q which is intended to account for the overlooked effects; the tracking efficiency ε'_{MC} is calculated for this new sample, by following the same procedure as for the original MC. This efficiency, however, can also be calculated as $\varepsilon'_{\text{MC}} = \varepsilon_{\text{MC}}^* (1 - q)$, where $\varepsilon_{\text{MC}}^*$ is the complete MC tracking efficiency, which also includes the dependency on the number of tracks and their electric charge; the calibration factor k can be therefore defined as the ratio between the original measured MC tracking efficiency (which does not include such dependencies) and the complete one (which does):

$$k = \frac{\varepsilon_{\text{MC}}^*}{\varepsilon_{\text{MC}}} = \frac{\varepsilon'_{\text{MC}}}{\varepsilon_{\text{MC}}(1 - q)}. \quad (5.3)$$

In the analysis, k is obtained by performing a linear regression after measuring ϵ'_{MC} for $q = 0.025, 0.05, 0.075, 0.1$.

Systematic errors related to the luminosity of the data sample, the ECL trigger efficiency, the calibration factor k and the number of background events, are taken into account by varying the respective nominal values by $\pm 1\sigma$ and measuring the new yields; the uncertainty introduced by the trigger is estimated by switching off this requirement and measuring the new yields; and the uncertainty introduced by the charge of the probe track is measured by measuring the yields for each charge separately. The overall final value for δ_{track} is found to be

$$\delta_{\text{track}} = 0.19 \pm 0.14 \text{ (stat)} \pm 0.78 \text{ (sys)} \% \quad (5.4)$$

We assign a correction factor of 99.81%, and a total error equal to the sum in quadrature of the systematic and statistical errors:

$$\text{sys}_{\text{track}} = \sqrt{0.14^2 + 0.78^2} = 0.79\% \quad (5.5)$$

5.2.2 Lepton ID

The analysis on lepton ID is performed in [72] by looking at $J/\psi \rightarrow \ell^+\ell^-$ decays. Data events were further filtered using the `h1t_hadron` skim, which is an online skim consisting of a Bhabha veto and a minimum cut of at least three good tracks in the event. Electron and muon candidates are selected through the next cuts:

- $|dr| < 2.0 \text{ cm}$,
- $|dz| < 5.0 \text{ cm}$,
- $p > 0.1 \text{ GeV}$,

Muons are also required to deposit less than 0.8 GeV in the ECL, and electrons are required to have a E/p ratio larger than 0.3. The `correctBrems` module is used to recover Bremsstrahlung energy losses for the electron candidates. Next, the following cuts are applied to the reconstructed J/ψ candidate:

- $2.8 < m(\ell^+\ell^-) < 3.3 \text{ GeV}$.
- A vertex fit, using the `KFit` package. Candidates for which the fitting does not converge are discarded.

Finally, in order to reduce background contamination, the event is subject to the next cuts:

- $R_2 < 0.4$,
- minimum of 5 tracks per event,
- each lepton track must have primary ECL cluster associated.

The number of signal and background events is estimated from a fit the the invariant mass of the J/ψ candidate $-m(\ell^+\ell^-)$. For electrons, the signal PDF is modeled as a sum of a Crystal-Ball function, a bifurcated Gaussian function and a Gaussian function, while a second order Chebyshev polynomial is used to model the background. For muons, the signal PDF is the sum of a Gaussian and a bifurcated Gaussian distributions, and the background as a second order Chebyshev polynomial. The fit is performed in bins of momentum and polar angle. The bins definitions are given in tables 5.4 and 5.5; the tables with the correction factors for each bin of momentum, polar angle and charge are presented in appendix G.

Table 5.4. Electron momentum and $\cos\theta_e$ bins

p bins [GeV]	$\cos\theta_e$ bins
(0.4,0.5)	(0.847,0.976)
(0.5,1.0)	(0.427,0.847)
(1.0,1.5)	(0.0,0.427)
(1.5,2.0)	(-0.304,0.0)
(2.0,2.5)	(0.612,0.304)
(2.5,3.0)	(0.908,0.612)

We re-weight each signal MC event according to the angular and momentum distribution of its leptons, and propagate the error in the weights to the final efficiency calculation.

5.2.3 Kaon ID

Efficiency corrections due to kaon ID discrepancies between data and MC were calculated in [73] using $D^{*+} \rightarrow D^0 [\rightarrow K^- \pi^+] \pi^+$ decays. Since the mass of the D^{*+} is only about 6 MeV above the $D^0 \pi^+$ threshold, the pion has a low momentum, and

Table 5.5. Muon momentum and $\cos\theta_\mu$ bins

p bins [GeV]	$\cos\theta_\mu$ bins
(0.4,0.5)	(0.802,0.921)
(0.5,0.7)	(0.682,0.802)
(0.7,1.0)	(0.399,0.682)
(1.0,1.5)	(0.111,0.399)
(1.5,2.0)	(0.208,0.111)
(2.0,2.5)	(0.531,0.208)
(2.5,3.0)	(0.605,0.531)
-	(0.857,0.605)

can be used to determine the flavor of the D meson, as well as the charge of the pion (same as the slow pion) and the kaon (opposite to the slow pion) in the D meson decay. The kaon ID efficiency is then calculated as

$$\varepsilon_K = \frac{\text{number of kaon tracks passing the kaon ID cut}}{\text{number of kaon tracks}}, \quad (5.6)$$

and the correction factor between data and MC as

$$r_K = \frac{\varepsilon_K^{\text{data}}}{\varepsilon_K^{\text{MC}}}. \quad (5.7)$$

In order to select the desired events, all tracks are required to be in the vicinity of the IP by applying the cuts $dr < 2$ cm and $|dz| < 4$ cm, and to have at least 20 CDC hits. Two tracks with opposite charges are assigned a kaon and pion mass hypothesis and then combined to form the D^0 meson candidate. Another charged track with a pion mass hypothesis is then combined with the D^0 candidate to yield a D^{*+} candidate. In order to select D^{*+} originating from $e^+e^- \rightarrow c\bar{c}$ continuum processes, a cut in the momentum of the D^{*+} in the CM frame is applied, $P_{D^{*+}} > 2.5$ GeV.

In order to reduce the number of background events, the invariant mass of the D^0 meson is required to be in the range (1.8,1.95) GeV, and the difference between the invariant masses of the D^0 and the D^{*+} meson are required to be in the range $0.138 \text{ GeV} < M_{D^{*+}} - M_{D^0} < 0.155 \text{ GeV}$.

The number of events for data and MC are obtained from an extended unbinned maximum likelihood fit over the D^0 invariant mass distribution, using the sum of two Gaussians with a common mean centered at the nominal value of m_{D^0} to model the signal events, and a second order polynomial to model the background component. The yields, efficiencies and correction factors are calculated for different bins of p_K and $\cos\theta_K$, the momentum and the cosine of the polar angle of the kaon track. These bins are presented in table 5.6, and the correction factors for each bin in table G.1, in appendix G

Table 5.6. Kaon momentum and $\cos\theta_K$ bins

p [GeV] bins	$\cos\theta_K$ bins
(0.5, 1)	(0.883, 0.956)
(1, 1.5)	(0.766, 0.883)
(1.5, 2)	(0.5, 0.766)
(2, 2.5)	(0.225, 0.5)
(2.5, 3)	(-0.104, 0.225)
(3, 3.5)	(-0.423, -0.104)
(3.5, 4.5)	(-0.682, -0.423)
-	(-0.866, -0.682)

We re-weight each signal MC event according to the angular and momentum distribution of its charged kaon, and propagate the error in the weights to the final efficiency calculation.

5.2.4 K_S^0 Selection

The efficiency correction due to the K_S^0 selections cuts applied during the reconstruction is obtained from the $D^{*+} \rightarrow D^0 [\rightarrow K_S^0 \pi^+ \pi^-] \pi^+$ decay. We reconstruct this decay from a subsample of 200 fb^{-1} MC events, and from the `proc10+bucket8` dataset, after the `h1t_hadron` skim has been applied.

We require all tracks not associated with the K_S^0 to originate in the vicinity of the IP, by applying the cuts $dr < 0.2 \text{ cm}$ and $|dz| < 2 \text{ cm}$, and by requiring for the tracks to have at least one VXD hit. Pion candidates (except the pair that combine into the K_S^0) are required to have a global pion ID above 0.6. In order to suppress

D^{*+} mesons coming from B decays, a momentum cut in the CMS of $P_{D^{*+}}^* > 2.5$ GeV is applied.

The K_S^0 candidates are obtained from vertex objects found during the preprocessing, and from a combination of oppositely charged tracks under the pion hypothesis. The invariant mass of the candidates is required to be in the range (0.3 GeV, 0.7 GeV). Next, A vertex fit, using the *TreeFitter* function, is performed (without applying any cut on the fit result). Finally, a cut on the recalculated invariant mass of the K_S^0 is performed, and it is required to lie within $0.45 \text{ GeV} < m_{K_S^0} < 0.55 \text{ GeV}$.

The K_S^0 is combined with two pion tracks to form a D^0 candidate, whose invariant mass is required to be in the range (1.82 GeV, 1.91 GeV) (see figure 5.2). The D^0 meson is finally combined with a slow pion to form a D^{*+} ; the energy release difference of the decay, $dQ = Q - 5.85 \text{ MeV}$, where $Q = m_D^* - m_{D^0} - m_{\pi_s}$ is constrained to lie between $(-1.5 \text{ MeV}, 1.5 \text{ MeV})$ (figure 5.2). If multiple candidates for the same event are obtained, the one with the value of dQ closest to zero is selected.

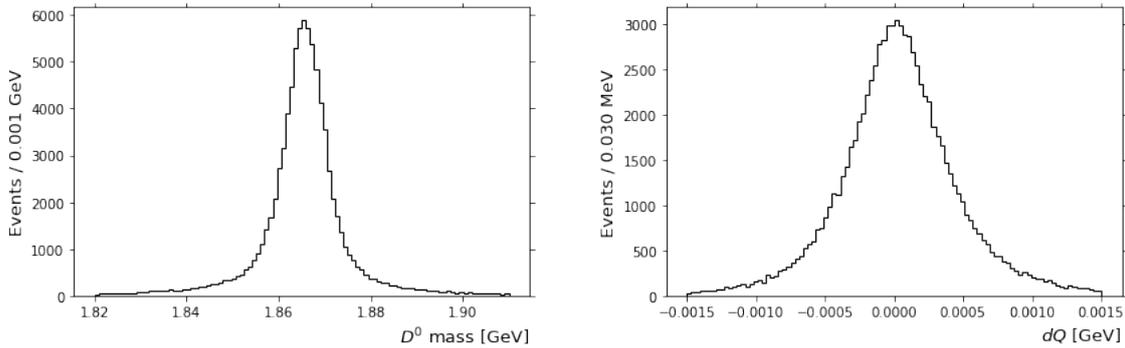


Fig. 5.1 Invariant mass distribution for the true D^0 (left) and dQ distribution of the true D^{*+} mesons (right) in MC.

Figure shows the signal and background components for the K_S^0 invariant mass distribution (after the vertex fit), before and after applying the goodBelleKShort cuts; the fraction of background events surviving the cut is 3.05%, while the number of selected signal events is 84.55%.

The signal component is fitted using the sum of three Gaussians $\mathcal{G}(x : \mu, \sigma)$ with a common mean:

$$\text{PDF}_{\text{signal}}(m_{K_S^0}) = f_1 \mathcal{G}(m_{K_S^0} : \mu, \sigma_1) + (1 - f_1) \left[f_2 \mathcal{G}(m_{K_S^0} : \mu, \sigma_2) + (1 - f_2) f_1 \mathcal{G}(m_{K_S^0} : \mu, \sigma_3) \right]. \quad (5.8)$$

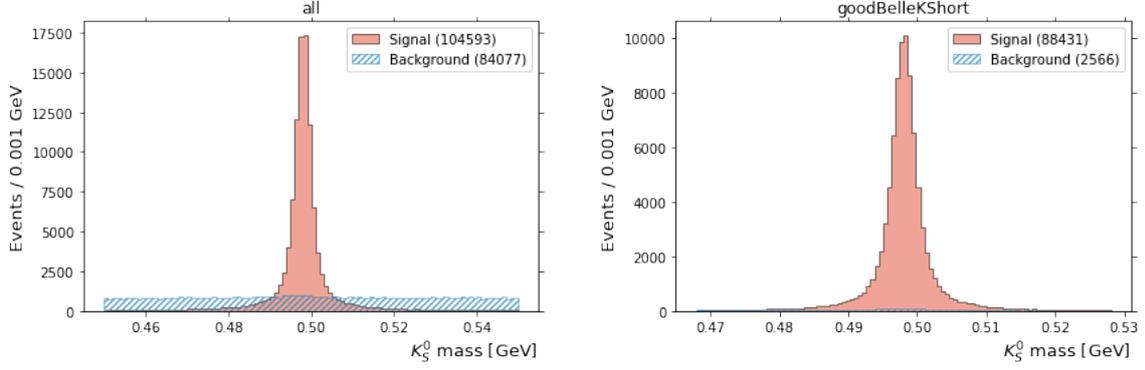


Fig. 5.2 Invariant mass distribution for the true (pink, solid) and fake (blue, hatched) K_S^0 in the MC data sample. The plot on the left depicts the distribution of all K_S^0 reconstructed, while the one in the right shows only those K_S^0 which pass the goodBelleKShort requirements.

The background component is modeled as a linear function²:

$$\text{PDF}_{\text{bkg}}(m_{K_S^0}) = a_1 m_{K_S^0} + 1. \quad (5.9)$$

Since the goodBelleKShort cuts depend on the momentum of the K_S^0 meson, the efficiency correction is calculated independently for the three different momentum regions $p_K < 0.5$ GeV, $0.5 \text{ GeV} \leq p_K < 1.5$ and $p_K > 1.5$ GeV. For each momentum bin we perform the fit for four different bins in $\cos\theta_K$, where θ_K is the polar angle of the K_S^0 candidate: $\cos\theta_K < -0.5$, $0.5 \leq \cos\theta_K < 0$, $0 \leq \cos\theta_K < 0.5$ and $\cos\theta_K > 0.5$. The fit results, together with the yields, are shown in appendix I. Table 5.7 presents the goodBelleKShort cut efficiency in data and MC for each combination of the momentum and angle regions, together with the efficiency correction factors $r_{K_S^0}$.

Figure 5.3 shows the fit results for the complete MC and data samples before and after applying the K_S^0 selection cuts.

The systematics of this study are obtained by varying the dQ and tightening the m_{D^0} windows, and recalculating the signal yields for data and MC, the efficiencies and the $r_{K_S^0}$. The systematic error associated with the selection cuts is calculated as the maximum difference between the original data/MC ratio and the ratios obtained after tightening the cuts. The cuts on $|dQ|$ are varied by $|dQ| < 1.7, 1.6,$ and 1.4 MeV. The largest difference in ratios is of 0.27%, which is assigned as the systematic error for the dQ cuts.

²The constant term is set to 1 without loss of generality, since the normalization condition of the PDF effectively removes one degree of freedom.

Table 5.7. Efficiency ratios and correction for the K_S^0 reconstruction

$p_{K_S^0}[\text{GeV}]$ bins	$\cos\theta_K$ bins	$\varepsilon_{\text{data}}$	ε_{MC}	$r_{K_S^0}$
< 0.5 GeV	< -0.5	0.64 ± 0.21	0.71 ± 0.08	0.89 ± 0.32
	0.5 - 0	0.64 ± 0.13	0.81 ± 0.05	0.78 ± 0.17
	0 - 0.5	0.85 ± 0.19	0.75 ± 0.05	1.13 ± 0.26
	> 0.5	0.58 ± 0.20	0.67 ± 0.18	0.87 ± 0.37
	all	0.71 ± 0.13	0.77 ± 0.03	0.92 ± 0.17
0.5 GeV - 1.5 GeV	< -0.5	0.83 ± 0.12	0.88 ± 0.03	0.94 ± 0.13
	0.5 - 0	0.94 ± 0.07	0.88 ± 0.01	1.07 ± 0.08
	0 - 0.5	0.87 ± 0.05	0.88 ± 0.01	0.98 ± 0.06
	> 0.5	0.89 ± 0.05	0.84 ± 0.01	1.06 ± 0.07
	all	0.88 ± 0.03	0.86 ± 0.01	1.02 ± 0.04
> 1.5 GeV	< -0.5	0.95 ± 1.18	0.86 ± 0.04	1.10 ± 1.38
	0.5 - 0	0.79 ± 0.10	0.91 ± 0.02	0.88 ± 0.11
	0 - 0.5	0.89 ± 0.07	0.87 ± 0.02	1.02 ± 0.09
	> 0.5	0.87 ± 0.05	0.86 ± 0.01	1.02 ± 0.06
	all	0.81 ± 0.05	0.88 ± 0.01	0.92 ± 0.06
all	< -0.5	0.81 ± 0.02	0.86 ± 0.02	0.94 ± 0.04
	0.5 - 0	0.87 ± 0.06	0.88 ± 0.01	0.99 ± 0.07
	0 - 0.5	0.87 ± 0.04	0.87 ± 0.01	1.00 ± 0.05
	> 0.5	0.80 ± 0.04	0.84 ± 0.01	0.95 ± 0.05
	all	0.87 ± 0.02	0.86 ± 0.01	1.01 ± 0.03

The mass window for the D^0 meson are tightened to (1.83,1.90), (1.84,1.89) and (1.85,1.88) GeV. The largest ratio difference is 2.23%.

The final global ratio value is then

$$r_{K_S^0} = 100.56 \pm 2.80(\text{stat}) \pm 2.25(\text{sys})\%. \quad (5.10)$$

We use these results to weight the MC events of the neutral channels according to the momentum and direction of the K_S^0 meson. We assign a common systematic uncertainty to the efficiency correction weights equal to the systematic error in the

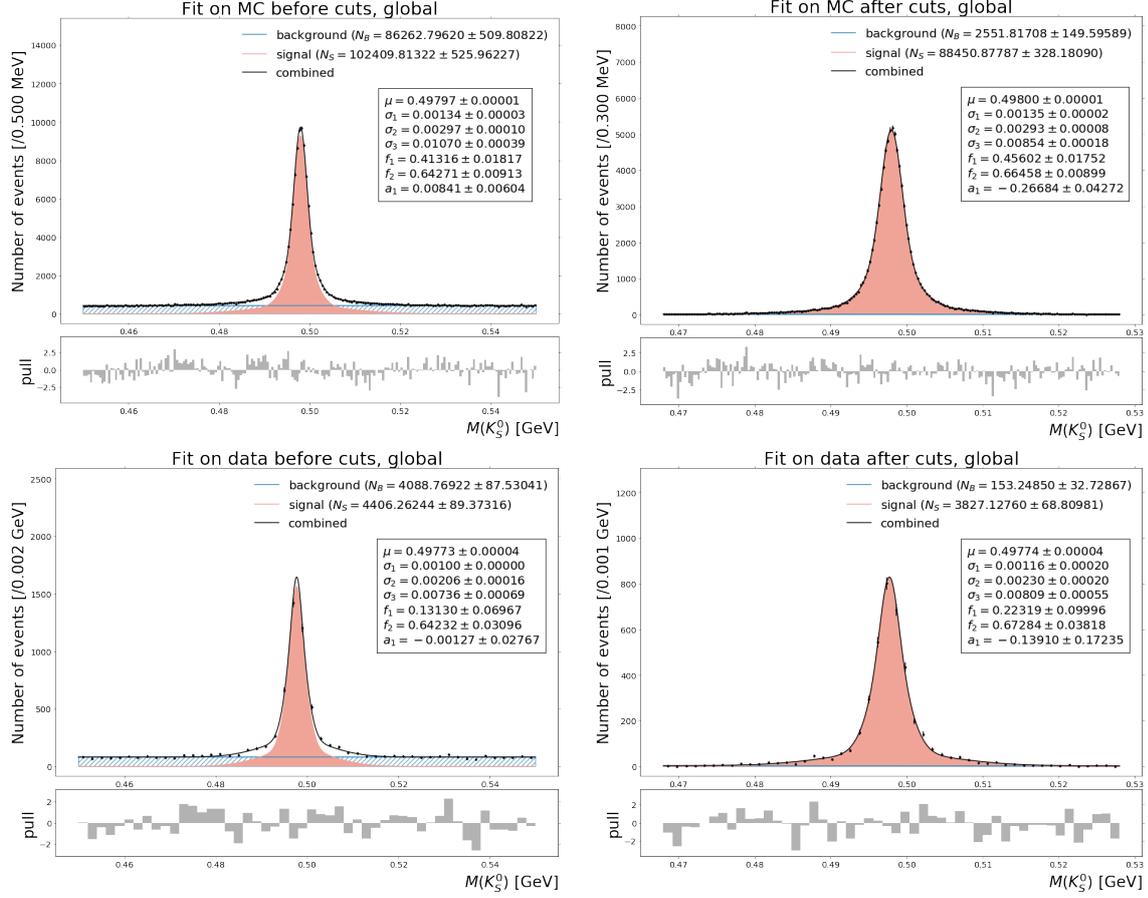


Fig. 5.3 Fit results on the K_S^0 mass from $D^{*+} \rightarrow D^0 [\rightarrow K_S^0 \pi^+ \pi^-] \pi^+$ decays on data and MC samples.

global ratio, and propagate this and the statistical errors for the weights to the final efficiency calculation.

5.2.5 MVA Classifiers

In order to evaluate the efficiency discrepancies between data and MC due to the MVA classifiers, we make use of the $B \rightarrow K\psi(nS) [\rightarrow \ell^+ \ell^-]$ resonances; we apply the same selection cuts as for the $B \rightarrow K\ell^+ \ell^-$ decays, and invert the charmonium vetoes.

The MVA efficiency is defined as

$$\varepsilon_{\text{MVA}} = \frac{\text{Number of events after MVA cuts}}{\text{Number of events before MVA cuts}} \quad (5.11)$$

and the correction factor as

$$r_{\text{MVA}} = \frac{\varepsilon_{\text{MVA}}^{\text{data}}}{\varepsilon_{\text{MVA}}^{\text{MC}}} \quad (5.12)$$

We obtain the number of events by performing an extended unbinned maximum likelihood fit to the m_{bc} distribution of the samples. The signal component is modeled as a Gaussian, where both μ and σ are allowed to vary. Incidentally, the Gaussian parameters obtained from the fit to the experimental data will also be used in the next chapter to model the signal component of the final m_{bc} distribution.

The small background portion of the distributions is fitted using an ARGUS function [74]:

$$f(m_{bc}) = \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} \frac{m_{bc}}{c^2} \sqrt{1 - \frac{m_{bc}^2}{c^2}} \exp\left\{-\frac{1}{2}\chi^2\left(1 - \frac{m_{bc}^2}{c^2}\right)\right\}, \quad (5.13)$$

for $0 \leq m_{bc} < c$ and

$$\Psi(\chi) = \Phi(\chi) - \phi(\chi) - \frac{1}{2} \quad (5.14)$$

where $\Phi(\chi)$ and $\phi(\chi)$ correspond to the cumulative distribution and the probability density functions of the standard normal distribution. We allow χ to vary freely, while fixing c to 5.3 GeV. The parameters obtained after the fit for the Gaussian function, for each lepton and kaon flavor, and for both data and MC, are shown in table 5.8. The fitted distributions' plots are included in appendix H. The obtained yields, efficiencies and correction factors are presented in table 5.9.

Table 5.8. Gaussian parameters for the m_{bc} distribution of $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$ decays

			neutral (B^0)		charged (B^+)	
			μ	e	μ	e
μ [GeV]	w/o MVA	data	5.28008(28)	5.27975(36)	5.27951(16)	5.27908(15)
		MC	5.27952(2)	5.27953(2)	5.27924(1)	5.27920(1)
	with MVA	data	5.28024(35)	5.27975(38)	5.27936(17)	5.27899(16)
		MC	5.27963(2)	5.27954(2)	5.27922(1)	5.27923(1)
σ [GeV]	w/o MVA	data	0.00251(22)	0.00276(30)	0.00264(12)	0.00246(11)
		MC	0.00260(1)	0.00264(2)	0.00256(1)	0.00265(1)
	with MVA	data	0.00248(26)	0.00287(33)	0.00245(12)	0.00235(12)
		MC	0.00253(2)	0.00261(2)	0.00254(1)	0.00260(1)

Table 5.9. Yields from the $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$ decays

		w/o MVA	with MVA	efficiency (ϵ_{MVA})	data/MC (r_{MVA})
$B = B^0, \ell = \mu$	data	84_{-8}^{+9}	51_{-6}^{+7}	0.62 ± 0.11	1.00 ± 0.18
	MC	17012_{-130}^{+131}	10502_{-102}^{+102}	0.62 ± 0.01	
$B = B^+, \ell = \mu$	data	284_{-16}^{+17}	217_{-14}^{+15}	0.76 ± 0.07	0.98 ± 0.09
	MC	55026_{-232}^{+238}	42865_{-207}^{+207}	0.78 ± 0.01	
$B = B^0, \ell = e$	data	70_{-8}^{+9}	63_{-7}^{+8}	0.90 ± 0.17	1.17 ± 0.22
	MC	14964_{-122}^{+124}	11446_{-106}^{+107}	0.77 ± 0.01	
$B = B^+, \ell = e$	data	266_{-16}^{+16}	204_{-13}^{+14}	0.77 ± 0.07	1.05 ± 0.10
	MC	47888_{-219}^{+220}	35066_{-188}^{+188}	0.73 ± 0.01	

We propagate these ratios and their errors into the final efficiency as explained in the next section.

5.3 Final Efficiencies

The final reconstruction efficiencies for the decay modes of interest, after taking into account all the corrections discussed in this chapter, are summarized in table 5.10.

Table 5.10. Final efficiencies

decay	ϵ [%]	$\Delta\epsilon_{\text{sys}}$ [%]	$\Delta\epsilon_{\text{sys}}$ at 50 ab^{-1} [%]	$\Delta\epsilon_{\text{stat}}$ [%]
$B^0 \rightarrow K^0\mu^+\mu^-$	6.13	1.15	0.10	0.01
$B^+ \rightarrow K^+\mu^+\mu^-$	18.14	1.69	0.42	0.02
$B^0 \rightarrow K^0e^+e^-$	7.19	1.35	0.11	0.01
$B^+ \rightarrow K^+e^+e^-$	18.29	1.81	0.44	0.01

The systematic errors were calculated by the TOY MC method. Here, we make use of the vector \mathbf{r} , containing all the nominal correction factors

$$\mathbf{r} = \{r_{\text{track}}, r_{\text{MVA}}, r_{\ell}(p_1, \theta_1), r_{\ell}(p_1, \theta_2), \dots, r_K(p_1, \theta_1), r_K(p_1, \theta_2), \dots\}$$

as the mean of a multidimensional normal distribution, with a covariance given by

$$\text{cov} = \begin{bmatrix} \Delta r_{\text{track}}^2 & 0 & \cdots & 0 \\ 0 & \Delta r_{\text{MVA}}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}, \quad (5.15)$$

where Δr_i corresponds to the total error associated with the correction factor r_i ³. We then perform $M = 1000$ simulations, in which the effective correction factors are sampled from this distribution. This is,

$$r_{\text{eff}}^{(m)} \sim \mathcal{N}(\mathbf{r}, \text{cov}), \quad m = 1, \dots, M. \quad (5.16)$$

We use these effective factors to re-weight each MC signal event which survived the selection cuts⁴. If any of the events is missing at least one of these correction factors⁵, we discard it. We then calculate the efficiency for the m -th simulation as:

$$\varepsilon^{(m)} = \frac{1}{N_0} \sum_i w_i^{(k)} \quad (5.17)$$

where i ranges over all the surviving signal events, N_0 is the initial number of generated events and $w_i^{(m)}$ is the product of all the effective correction factors for the event i . Finally, we report the efficiency value as the mean of the efficiencies for each simulation

$$\varepsilon = \langle \varepsilon^{(m)} \rangle = \frac{1}{M} \sum_{m=1}^M \varepsilon^{(m)} \quad (5.18)$$

and its systematic error as the standard deviation of these efficiencies

$$\Delta \varepsilon_{\text{sys}} = \frac{1}{M} \sqrt{\sum_{m=1}^M (\varepsilon^{(m)} - \varepsilon)^2}. \quad (5.19)$$

Finally, the statistical error is obtained from the statistical error of a binomial variable in the next fashion: we can model the selection of an event as a Bernoulli trial with a

³These errors are calculated as the sum of squares of the statistical and systematic errors. In case only the statistical error is provided, the systematic error is taken to be zero.

⁴Notice that some components of the vector \mathbf{r}_{eff} enter the calculation of the weight more than once. For example, for charged B meson decays, there are three tracks in the final state, so the factor $r_{\text{track,eff}}$ appears three times in the calculation of each event weight.

⁵certain combinations of bins in the efficiency studies do not contain enough statistics for the fit to converge. In these cases there is no correction factor reported.

probability $p = \varepsilon$, and hence the number of selected events n —or, in other words, the number of events passing the cuts—in a total of N_0 trials is distributed as a binomial variable with mean pN_0 and variance $p(1-p)N_0$. Performing the usual error propagation

$$\Delta\varepsilon_{\text{stat}} = \frac{d\varepsilon}{dn}\Delta n = \frac{1}{N_0}\sqrt{p(1-p)N_0} = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_0}}. \quad (5.20)$$

We have also reported the expected systematic error at the target luminosity of 50 ab^{-1} , which was calculated by scaling the statistical errors for each correction factor by a value of

$$\sqrt{\frac{\text{luminosity for data used in the study}}{\text{target luminosity}}}$$

and performing the simulations again.

Table 5.11 shows the contribution of each correction factor to the overall systematic error of the efficiency. These values were obtained by the same TOY MC method described before, but making $\Delta r_i = 0$ for all correction factors i different from the one being studied. At the current luminosity, the largest source of uncertainty is associated with the continuum and $B\bar{B}$ classifiers. This is expected to improve as the data size increases, and at the target luminosity of 50 ab^{-1} the largest source of systematic uncertainties for all decay modes is found in the tracking efficiency. However, the statistical error from the MVA classifiers is still the second largest source of systematic uncertainties; possible solutions can be to train a smaller number of BDTs, thus combining the statistics of multiple decay modes; nevertheless, this approach may result in lower efficiencies. Another option would be to choose a set of looser selection cuts applied before the continuum and $B\bar{B}$ suppression, but this can increase the number of background events in the final samples. Notice that in these calculations the systematic error for all the K^+ ID correction factors is still zero, which is overly optimistic. As the data size increases, further studies of the systematic factors affecting the kaon ID efficiency must be evaluated.

As a way of validating the analysis done so far, we calculate the ratio of the branching fractions

$$R_K^{\text{reso}} = \frac{\mathcal{B}(B \rightarrow K\psi(nS) [\rightarrow e^+e^-])}{\mathcal{B}(B \rightarrow K\psi(nS) [\rightarrow \mu^+\mu^-])} \quad (5.21)$$

Table 5.11. Contribution to $\Delta\epsilon_{\text{sys}}$ from the different correction factors

decay	source	$\Delta\epsilon_{\text{sys}}$ [%]	$\Delta\epsilon_{\text{sys}}$ at 50 ab^{-1} [%]
$B^0 \rightarrow K^0 \mu^+ \mu^-$	tracking	0.10	0.10
	lepton ID	0.01	0.00
	K_S^0 reco.	0.03	0.01
	MVA	1.13	0.02
$B^+ \rightarrow K^+ \mu^+ \mu^-$	tracking	0.44	0.42
	lepton ID	0.02	0.00
	K^+ ID	0.01	0.00
	MVA	1.63	0.03
$B^0 \rightarrow K^0 e^+ e^-$	tracking	0.12	0.11
	lepton ID	0.12	0.01
	K_S^0 reco.	0.07	0.01
	MVA	1.33	0.02
$B^+ \rightarrow K^+ e^+ e^-$	tracking	0.44	0.43
	lepton ID	0.23	0.01
	K^+ ID	0.01	0.00
	MVA	1.74	0.03

for $n = 1, 2$, using the whole dataset. The determination of this ratio is of crucial importance in the study of FCNC decays, since it can be used to reduce the systematic uncertainties for R_K in the non-resonant processes. Indeed, the recent LHCb result [29] uses the double ratio

$$R_K = \frac{\frac{\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi[\mu^+ \mu^-])}}{\frac{\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)}{\mathcal{B}(B^+ \rightarrow K^+ J/\psi[e^+ e^-])}}. \quad (5.22)$$

This introduces additional uncertainties related to the reconstruction efficiency of the resonant decays and to the ratio of the $J/\psi \rightarrow \ell^+ \ell^-$ branching fractions, but has the advantage of cancelling the systematic errors associated with the reconstruction process. As the data size increases and the systematic uncertainties overcome the statistical ones, this approach may help in delivering cleaner measurements.

We start by applying all the selection cuts described in the previous chapter, except for the charmonium vetoes, which are reversed, and the MVA cuts. We perform a fit on the m_{bc} distribution for each decay mode, and in each veto region separately. We model the signal component as a Gaussian, and the background component as an ARGUS function, and float all parameters except for the cut value of the ARGUS, which is fixed at 5.3 GeV. We calculate the reconstruction efficiencies from the dedicated MC samples for each resonant decay, and incorporate the correction factors through a toy MC simulation.

The final ratios are shown in table 5.12, where the systematic error is calculated as the sum of squares of the systematic errors in the efficiencies⁶ for each decay mode, and the statistical error is the sum of squares of the errors in the yields obtained from the fit.

Table 5.12. Branching ratio of $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$ decays, $\ell = e, \mu$

resonance	decay mode	ratio ($e^+e^-/\mu^+\mu^-$)	error (stat)	error (sys)	PDG value
J/ψ	neutral	0.89	0.16	0.01	
	charged	1.04	0.12	0.01	1.002
	average	0.99	0.09	0.01	
$\psi(2S)$	neutral	1.30	0.81	0.01	
	charged	0.94	0.45	0.01	1.003
	average	1.03	0.41	0.01	

All measurements are consistent with the PDG values within 1σ . Notice that in all measurements the statistical error is the dominant source of uncertainties, and it is expected that larger data samples will improve this precision.

We make further use of the $B \rightarrow Ke^+e^-$ events inside the J/ψ veto region to assess the effectiveness of the Bremsstrahlung recovery method; these samples are virtually free from background contamination, and thus almost all the events correspond to $B \rightarrow KJ/\psi [\rightarrow e^+e^-]$ decays. Figure 5.4 shows the dilepton invariant mass distributions before and after performing the Bremsstrahlung recovery. As observed, once the recovery is done, the distribution narrows towards the nominal J/ψ mass, which is 3.1 GeV. The radiative tail gets shorter, and the range of q

⁶Here, we only include the systematics due to the lepton ID correction factors, since the other ones cancel out.

containing a given percentage of the reconstructed events, shown in table 5.13 shrinks by a scale similar as in the MC samples studied in chapter 4. This proves that the performance of the method is similar between data and MC, and confirms the improvement achieved with it.

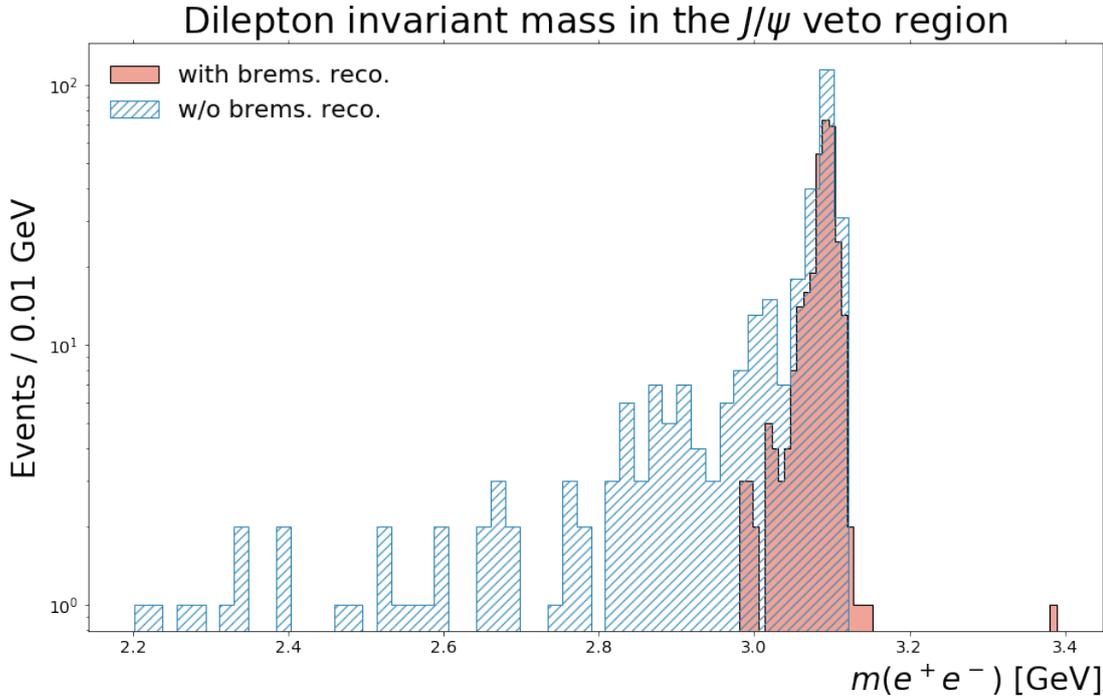


Fig. 5.4 Dilepton invariant mass distribution for the $B \rightarrow Ke^+e^-$ decays inside the J/ψ sidebands, with and without Bremsstrahlung recovery.

Table 5.13. q ranges containing a percentage of the reconstructed $B \rightarrow Ke^+e^-$ events inside the J/ψ veto region

	range containing		
	68% [GeV]	95% [GeV]	99.7% [GeV]
w/o brems. corr.	(2.89,3.10)	(2.40,3.11)	(2.22,3.12)
with brems. corr.	(3.07,3.10)	(3.02,3.12)	(2.98,3.15)
width ratio	0.16	0.14	0.18

Chapter 6

Signal Extraction

This chapter presents the m_{bc} distributions from the experimental data after performing the reconstruction of the four channels of interest. The expected number of events for the current luminosity are shown in table 6.1, while figure 6.1 shows the m_{bc} distribution for MC data scaled to the current luminosity.

Table 6.1. Expected number of events at $\mathcal{L} = 11.53 \text{ fb}^{-1}$, based on the latest PDG values for the branching fractions

event	expected number
$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	0.24 ± 0.05
$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.99 ± 0.11
$B^0 \rightarrow K_s^0 e^+ e^-$	0.13 ± 0.09
$B^+ \rightarrow K^+ e^+ e^-$	1.25 ± 0.20

The extraction procedure is based on an extended unbinned maximum likelihood fit to the m_{bc} distribution of the selected events. This chapter is divided in two parts: in the first one, the different components of the PDF are described, and the fit is performed. In the second one, the systematic errors associated with the fitting parameters are calculated, and a parameterization of the statistical and systematic errors of the R_K ratio as a function of the luminosity is derived, from which a luminosity threshold for competitive results is obtained.

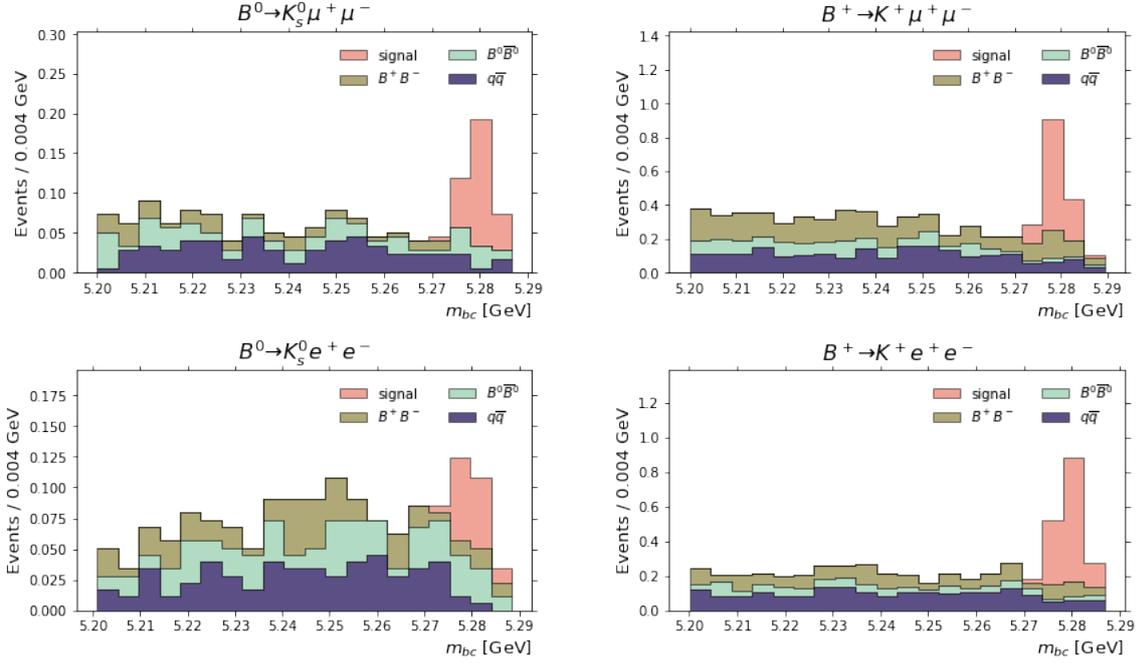


Fig. 6.1 Final m_{bc} distribution for MC data scaled to a luminosity of 11.53 fb^{-1} .

6.1 Signal Extraction

6.1.1 PDF Components

The signal component is modeled as a Gaussian, and its mean and variance are fixed to the values obtained from the fits performed in the previous chapter on the m_{bc} distribution inside the charmonium sidebands; we reproduce these parameters one more time in table 6.2.

Table 6.2. Gaussian parameters for the signal components in the m_{bc} distribution of $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$ decays

Parameter	neutral (B^0)		charged (B^+)	
	$\ell = \mu$	$\ell = e$	$\ell = \mu$	$\ell = e$
μ [GeV]	5.28024(35)	5.27975(38)	5.27936(17)	5.27899(16)
σ [GeV]	0.00248(26)	0.00287(33)	0.00245(12)	0.00235(12)

The non-peaking background is modeled as an ARGUS function, and we fix both of its parameters from fits to the generic MC background samples, where we require at least one of the leptons to be properly identified. The distribution of

these backgrounds and their fitting parameters are shown in figure 6.2. This sample contains $B\bar{B}$ and $q\bar{q}$ events, and we assume that the composition ratio of each type of event is well modeled (and thus no uncertainty related to this composition is taken into account), by noticing that the distributions of the multiple MVA scores show good agreement between data and MC (figure 4.28).

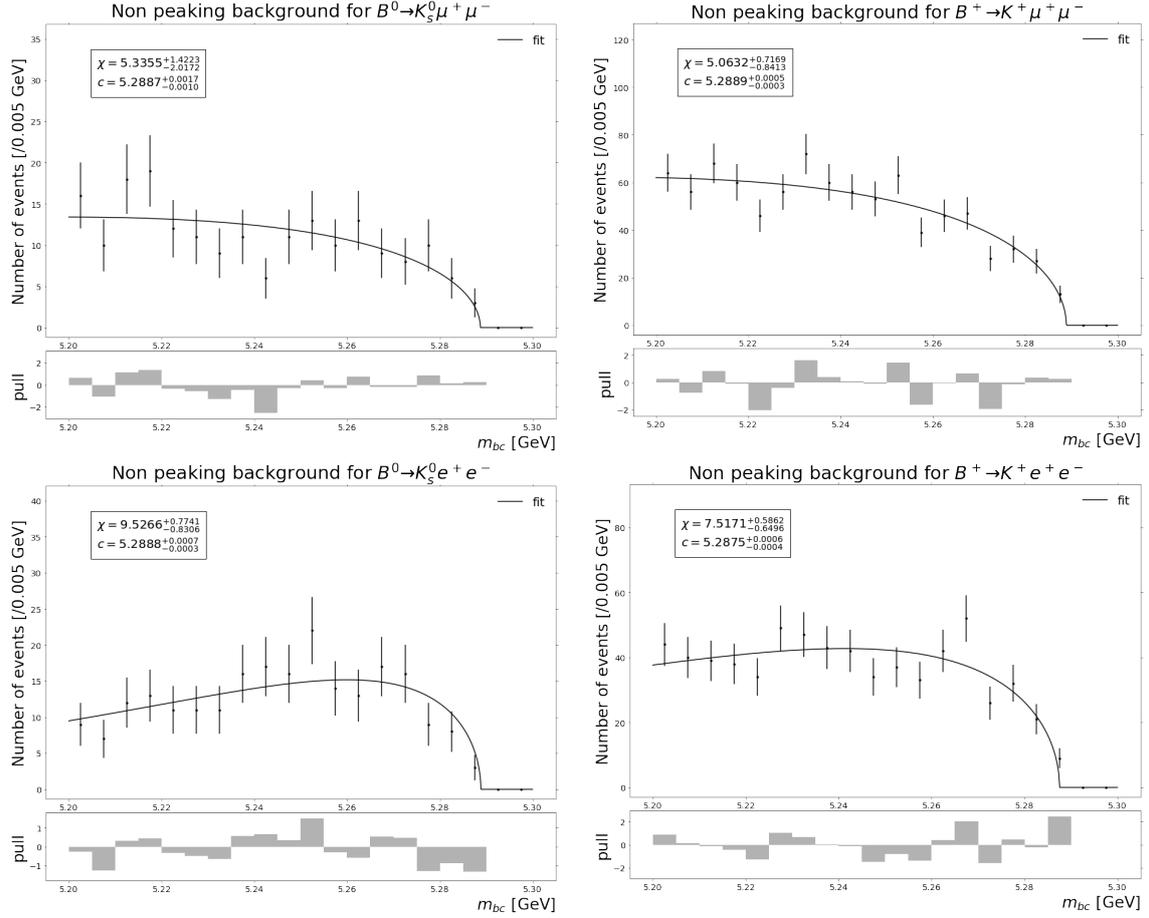


Fig. 6.2 m_{bc} distributions and fits for the non-peaking backgrounds from MC.

We also consider two sources of peaking backgrounds: J/ψ and $\psi(2S)$ events that make it through the charmonium vetoes, and $B \rightarrow K\pi^+\pi^-$ where both pions are misidentified as leptons. We estimate the contribution of the first one by performing the signal reconstruction on the dedicated MC $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$ samples mentioned in chapter 4. Before the fit, each reconstructed event is weighted by the factor

$$w = \frac{2 \times N^{+/0} \times \mathcal{B}(B^{+/0} \rightarrow K^{+/0}\psi(nS)) \times a_K^{+/0} \times \mathcal{B}(\psi(nS) \rightarrow \ell^+\ell^-)}{N_0}, \quad (6.1)$$

where $N^{+/0}$ is the expected number of B^+B^- ($B^0\bar{B}^0$) pairs produced in 1 fb^{-1} of e^+e^- collisions at the Belle II detector, $\mathcal{B}(B^{+/0} \rightarrow K^{+/0}\psi(nS))$ and $\mathcal{B}(\psi(nS) \rightarrow \ell^+\ell^-)$ are the corresponding branching fractions for the processes as reported by the PDG, N_0 is the number of generated MC events, and

$$a_K^{+/0} = \begin{cases} 1 & \text{if } B = B^+ \\ 0.5 \times \mathcal{B}(K_s^0 \rightarrow \pi^+\pi^-) & \text{if } B = B^0 \end{cases} \quad (6.2)$$

is a correction factor which takes into account that, for neutral decays, during the production of these MC samples, $K^0 = K_s^0$ and $\mathcal{B}(K_s^0 \rightarrow \pi^+\pi^-) = 1$. This reweighting procedure allows us to introduce the same PDF at different luminosities by simply scaling the yield. The values used for the different parameters in equation 6.1 are listed in table 6.3. Their error is propagated to the weight error, and the weights errors are finally propagated to the yield by a toy MC approach, in a procedure similar to the one used in the previous chapter (see appendix L).

Table 6.3. Parameters used in the weighting procedure of the charmonium background events

parameter	value	error
N^0	0.51×10^6	0.01×10^6
N^+	0.54×10^6	0.01×10^6
$\mathcal{B}(B^+ \rightarrow K^+J/\psi)$	1.01×10^{-3}	0.03×10^{-3}
$\mathcal{B}(B^+ \rightarrow K^+\psi(2S))$	6.19×10^{-4}	0.22×10^{-4}
$\mathcal{B}(B^0 \rightarrow K^0J/\psi)$	8.68×10^{-4}	0.30×10^{-4}
$\mathcal{B}(B^0 \rightarrow K^0\psi(2S))$	5.8×10^{-4}	0.5×10^{-4}
$\mathcal{B}(J/\psi \rightarrow \mu^+\mu^-)$	5.96%	0.03%
$\mathcal{B}(J/\psi \rightarrow e^+e^-)$	5.97%	0.03%
$\mathcal{B}(\psi(2S) \rightarrow \mu^+\mu^-)$	8.0%	0.6%
$\mathcal{B}(\psi(2S) \rightarrow e^+e^-)$	7.93%	0.17%
a_K^+	1	0
a_K^0	0.35	0

The MC analysis reveals that, at the current luminosity, no events from resonant decays are expected (see figure 6.3). However, in order to account for this background at larger luminosities, we model the m_{bc} distribution for resonant events

inside the signal region as a Gaussian, and allow the mean and the standard deviation to vary freely in the fit. The results are shown in figure 6.4; the yields reported correspond to the expected number of events for a luminosity of 1 fb^{-1} .

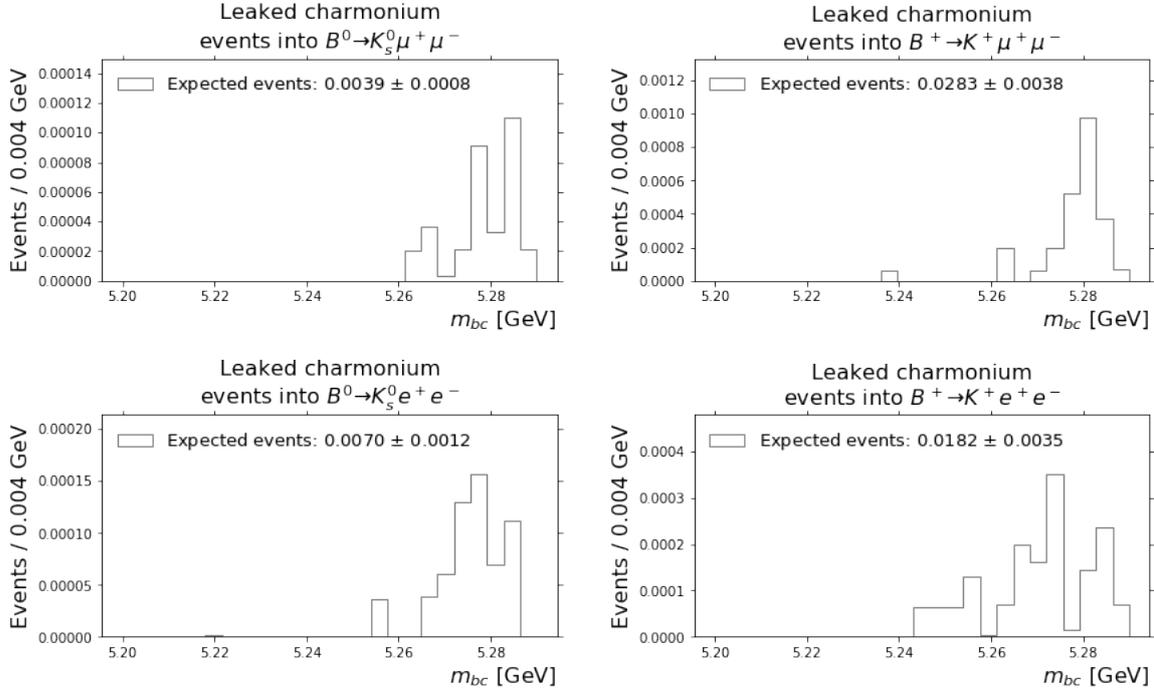


Fig. 6.3 m_{bc} distributions of true charmonium decays in MC data that pass the selection criteria.

The contribution from the second type of peaking background is calculated directly from data, by applying the same selection cuts, except for the lepton ID requirements, which are inverted. We then weight the selected events by their appropriate pion to lepton fake rate, and fit the resulting m_{bc} distribution as the fractional sum of a Gaussian and an ARGUS function:

$$\text{PDF}_{\text{miss-ID}}(m_{bc}) = f\mathcal{G}(m_{bc} : \mu, \sigma) + (1 - f)\text{ARGUS}(m_{bc} : \chi, c). \quad (6.3)$$

The fake rates are obtained from $K_S^0 \rightarrow \pi^+ \pi^-$ events, and the complete results are fully detailed in [72]; here we give just a small description of the study. The pions are selected by the next cuts:

- $|dr| < 2.0 \text{ cm}$,
- $|dz| < 5.0 \text{ cm}$,
- $p > 0.1 \text{ GeV}$.

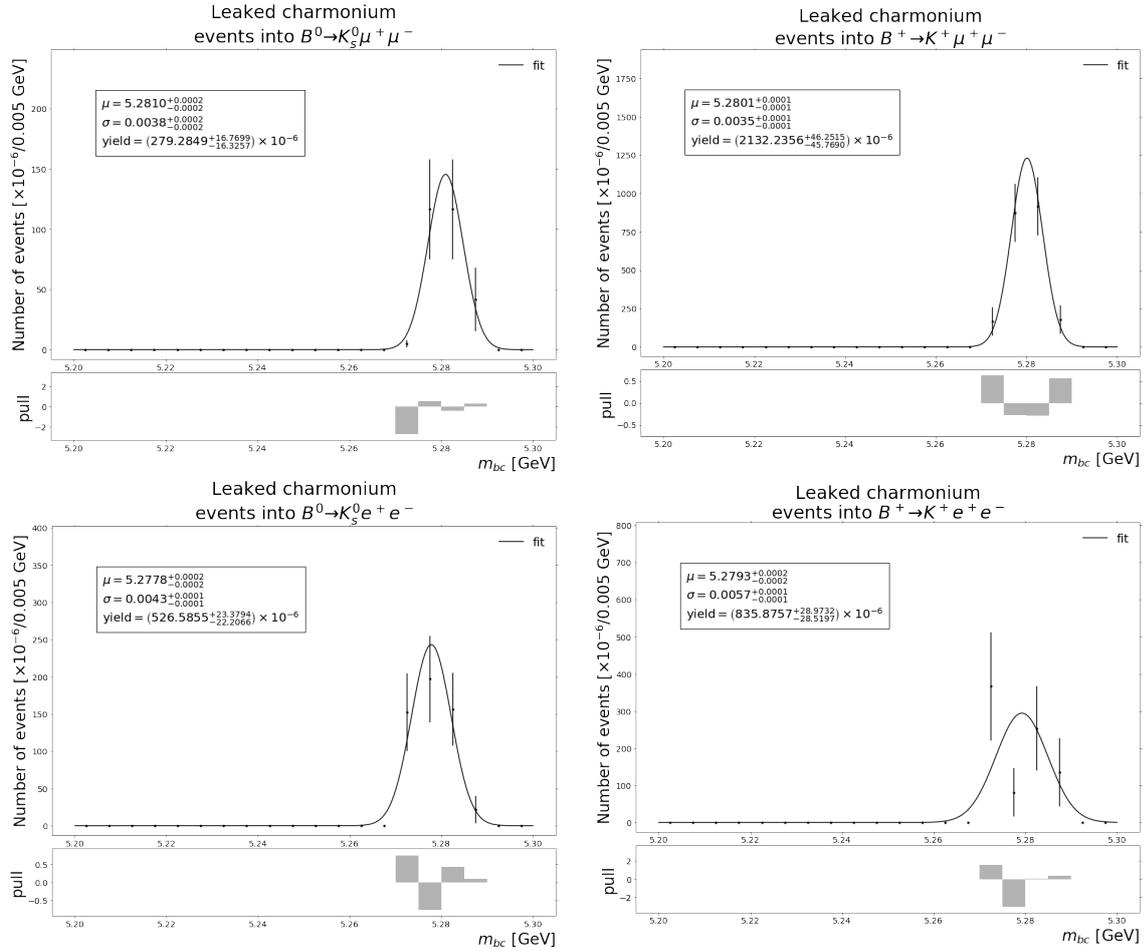


Fig. 6.4 Fits to the m_{bc} distributions inside the signal region for the MC resonant events that make it through the selection cuts (scaled to a luminosity of 1 fb^{-1}).

And the K_S^0 candidates are then reconstructed by requiring $0.45 < m(\pi^+ \pi^-) < 0.55$ GeV. In order to achieve an environment similar to that of B -enriched events, only events with $R_2 < 0.4$ are kept. Next, a vertex fit, using the `KFit` package, is performed, and only those events for which the fit converged are retained. Finally, after the fit, the cosine of the angle between the K_S^0 momentum and the decay vertex position vector is required to be larger than 0.998. The fake rates are obtained by counting the number of pion tracks kept before and after applying a lepton ID cut, and are reported for both muon and electron IDs, and for tracks of positive and negative charges separately. They are calculated for the same bins of momentum and polar angle as the ones used for the lepton ID and muon ID efficiency corrections in the previous chapter. From this study, it is revealed that the pion fake rates for electrons in data are more than 6 times larger than in MC, specially in the low momentum

region; on the contrary, the pion fake rates for muons are *smaller* in data than in MC (see figure 6.5). The explanation for this was found in a poor modeling of the input variables used in the PID process by the CDC. Indeed, figure 6.6 shows the distribution of the mean energy loss for pions in data and MC. Moreover, figure 6.7 (6.8) presents the electron (muon) PID distributions for pions in data and MC in each subdetector. The mismatch for the PID obtained by the CDC is evident. We thus obtain different fitting parameters for MC and data, according to the weights applied.

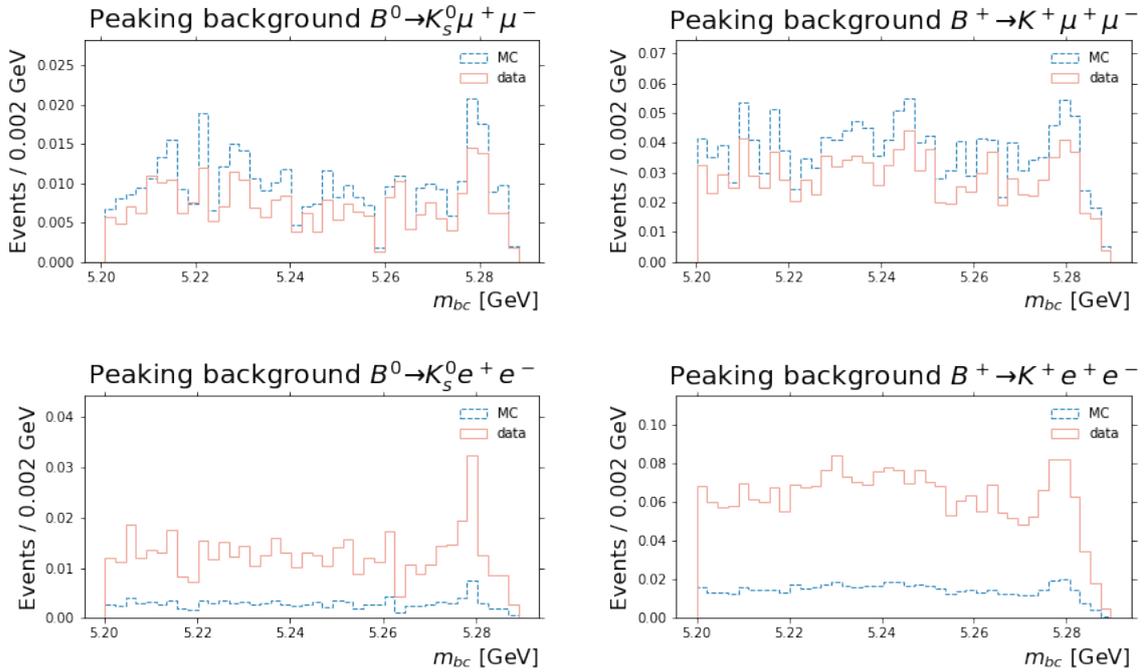


Fig. 6.5 Scaled m_{bc} distributions for double lepton miss-ID events from data, for data and Monte Carlo.

The fit results for data are shown in figure 6.9. The corresponding results for MC are shown in appendix K.

Once the shape parameters and the yield for the weighted $B \rightarrow K\pi^+\pi^+$ events are obtained, they are fixed and included in the final extended PDF. Table 6.4 resumes the components of the final PDF considered for the fitting.

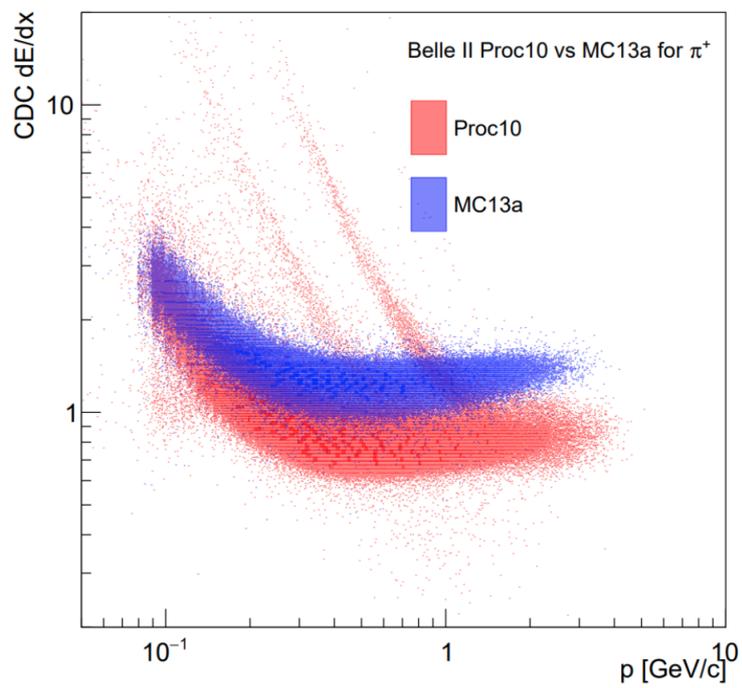


Fig. 6.6 dE/dx as a function of momentum over the whole angular range for π^+ for data and MC, with some background components present in data. The discrepancy is clearly visible. Image taken from [72].

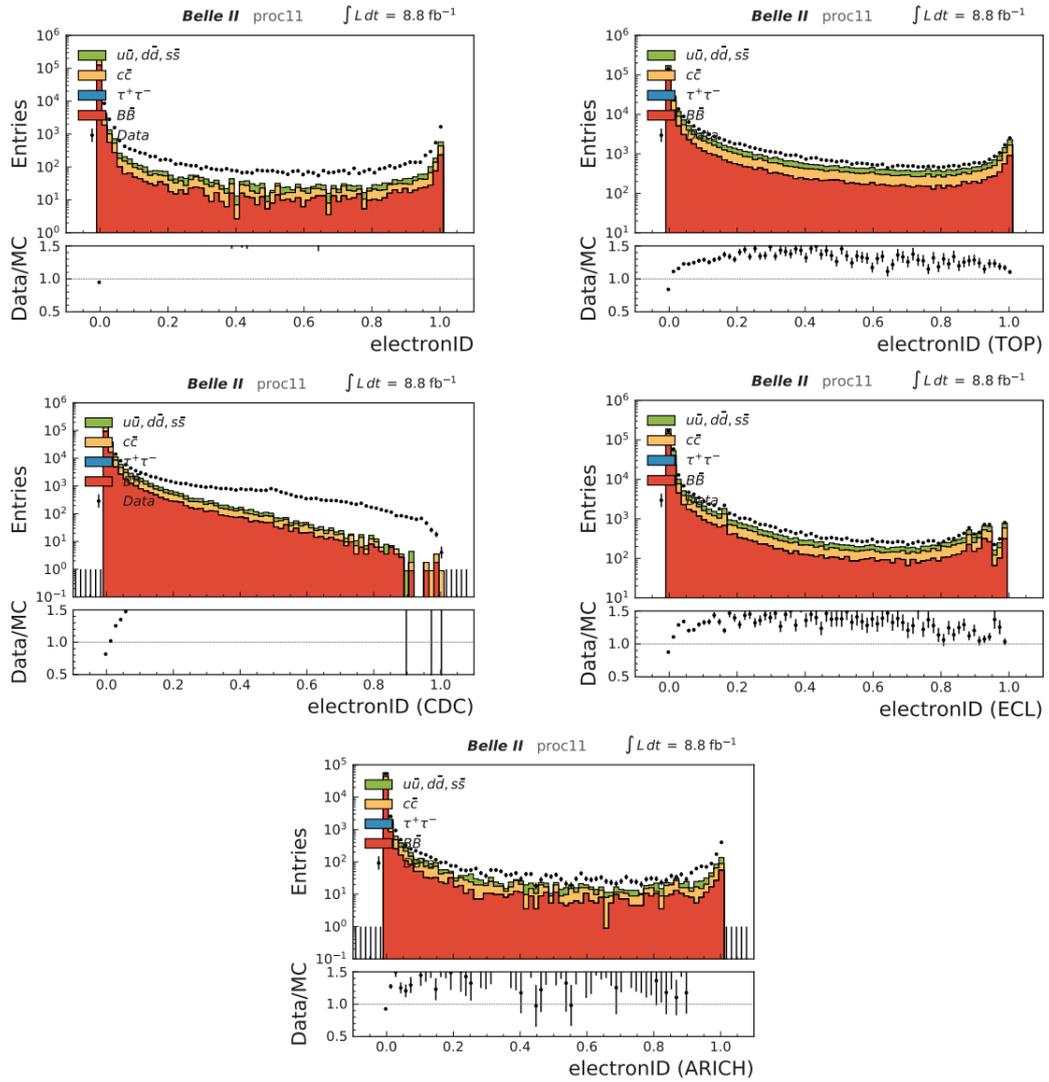


Fig. 6.7 Electron PID distributions for data and MC for each of the subdetectors. Image taken from [72].

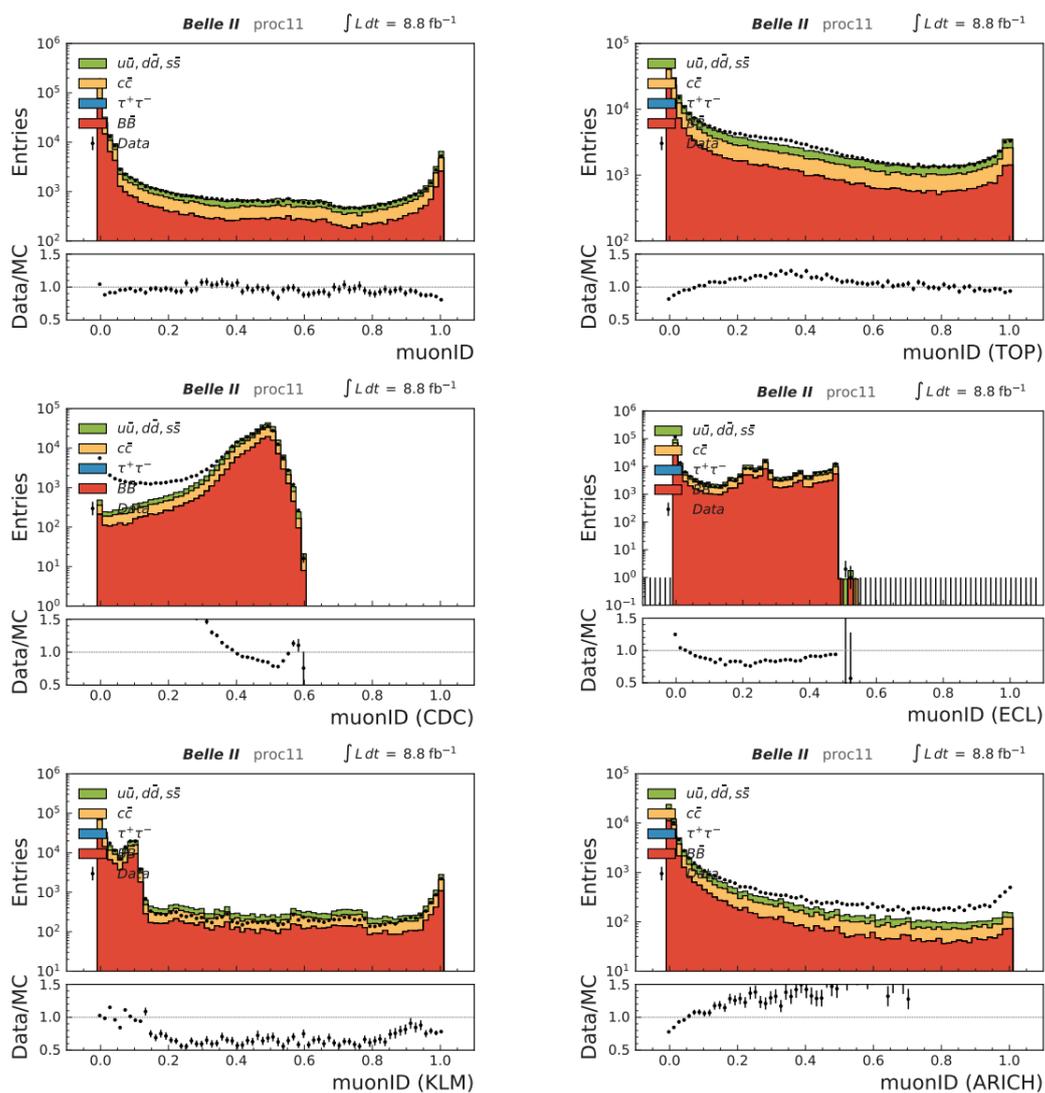


Fig. 6.8 Muon PID distributions for data and MC for each of the subdetectors. Image taken from [72].

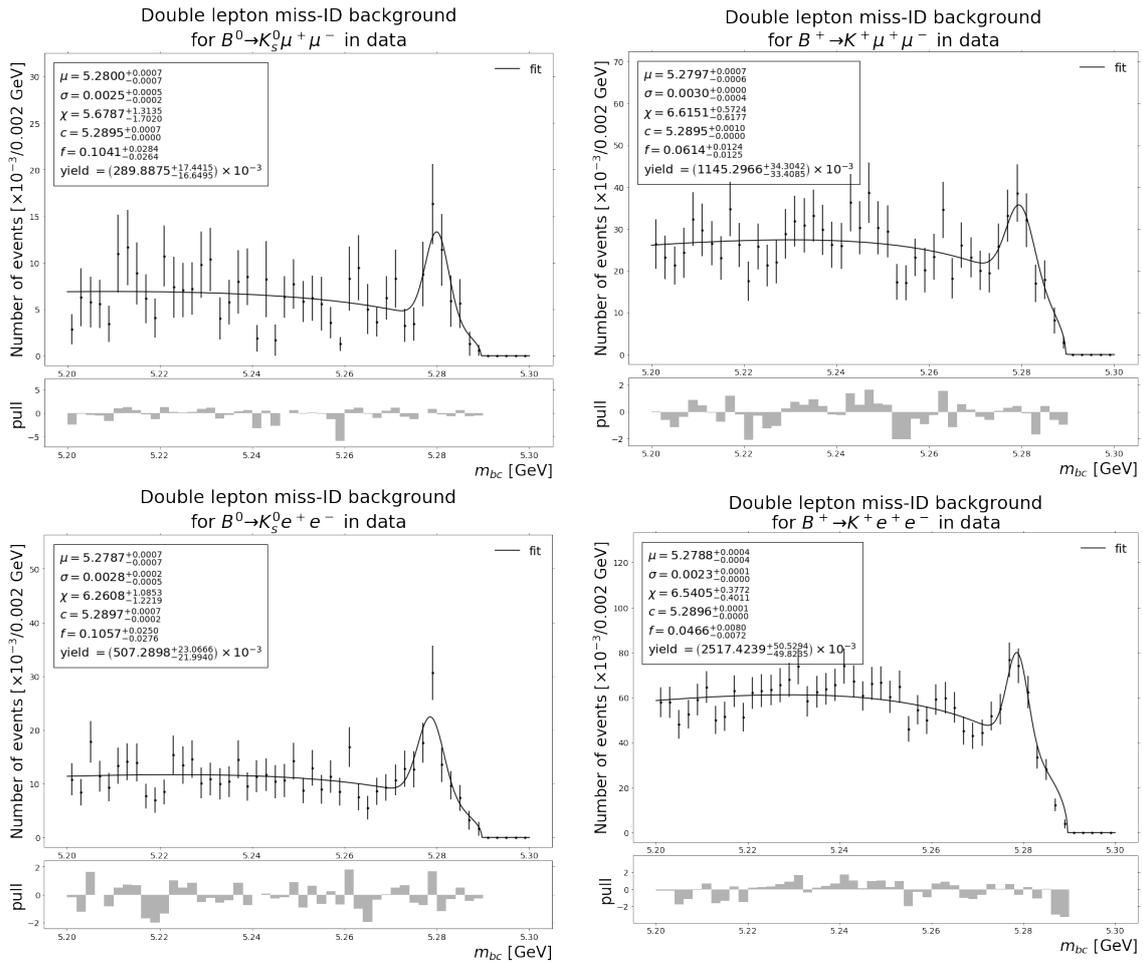


Fig. 6.9 Fit to the weighted m_{bc} distributions of $K\pi^+\pi^+$ background events from 11.53 fb^{-1} of experimental data.

Table 6.4. Final components of the m_{bc} extended PDF

component	PDF	parameter	fixed
Signal	Gaussian	μ	yes
		σ	yes
		yield	no
Non-peaking bkg.	ARGUS	χ	yes
		c	yes
		yield	no
Leaked charmonium bkg.	Gaussian	μ	yes
		σ	yes
		yield	yes
Double lepton miss-ID bkg.	Gaussian+ARGUS	μ	yes
		σ	yes
		χ	yes
		c	yes
		f	yes
		yield	yes

6.1.2 Fitting

Monte Carlo

We start by revisiting the m_{bc} distributions for the MC data shown in figure 4.27. In chapter 4, we gave an estimate of the pseudo-significance at different luminosities, based only on the number of signal and background events passing the selection criteria. Here, we use the PDF information for the different components to perform an extended unbinned maximum likelihood fit, in order to get a more precise value of the significance. We also use the fit results to get an estimate of the statistical error expected at different luminosities.

As the luminosity for the MC data is 2 ab^{-1} , we rescale the yields for the peaking background components accordingly. The fit results are shown in figure 6.10. Notice that the yields for signal events are in close agreement with the true number of signal events in the sample.

We define the statistical significance of the fit as

$$S = \sqrt{-2 \log \left(\frac{\lambda_0}{\lambda_{\max}} \right)} \quad (6.4)$$

where λ_0 is the likelihood obtained after fitting the distribution following a null hypothesis —this is, under the assumption that the signal yield is zero— and λ_{\max} is the likelihood obtained after performing the fit under the alternate hypothesis —this is, letting the signal yield vary freely— (we have changed the naming conventions and referred to the likelihood with the letter λ in order to distinguish it from the luminosity). We performed the fits again, this time constraining the signal yield to be zero, and calculated the likelihood. The significances for each decay mode are listed in table 6.5. As expected, the additional information regarding the shape of the components increases the statistical power of the measurement, except for the $B^0 \rightarrow K_S^0 e^+ e^-$ decay; this is due to the fact that the percentage of events in the signal region which are background is higher in comparison with the other decays. This makes it difficult to clearly distinguish the signal peak. This, in turn, is caused by the fact that the separation power of the MVA classifiers is not optimal yet, as the continuum suppression classifier is trained with a dataset mostly composed of $B^+ \rightarrow K^+ e^+ e^-$ events, and the number of training samples for the $B\bar{B}$ classifier is still not large enough to generalize properly. This is expected to improve as larger MC datasets are produced.

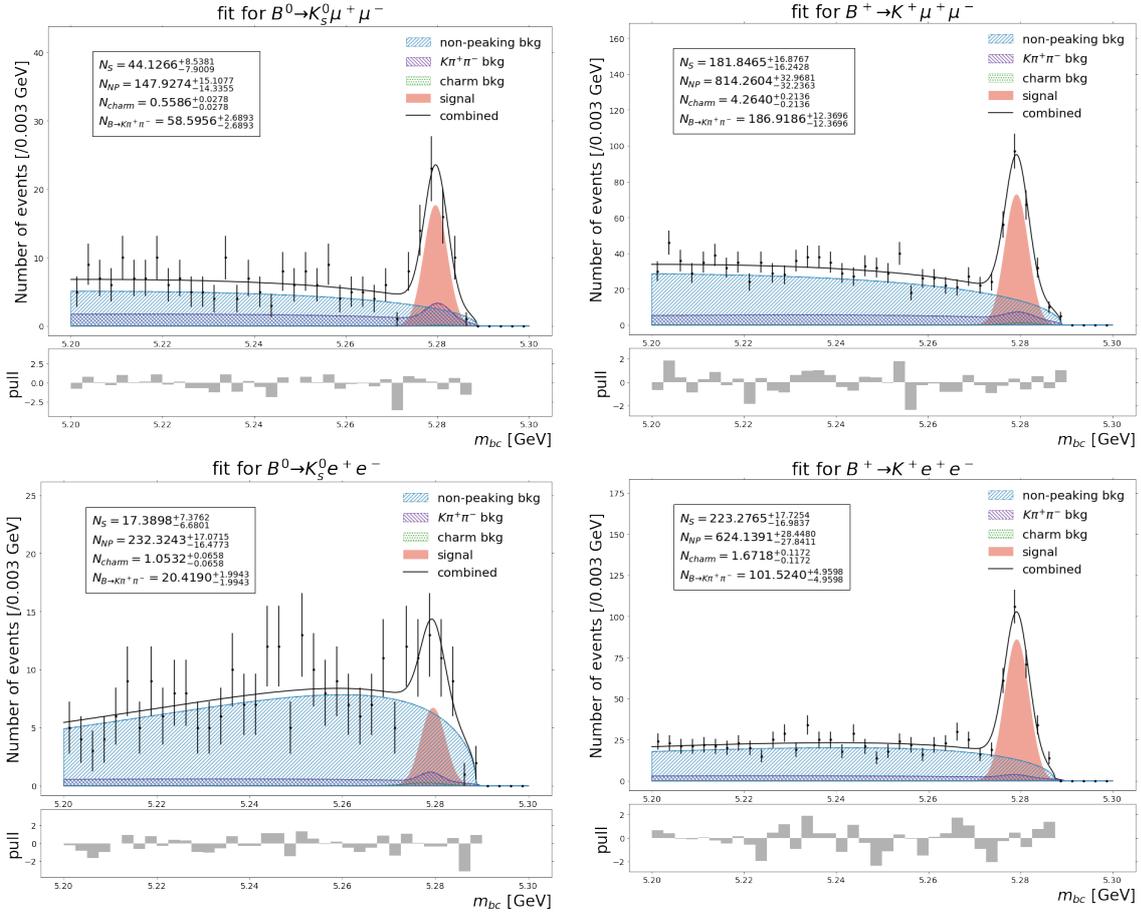


Fig. 6.10 Final fits to the MC generic sample, corresponding to 2 ab^{-1} of data taken at the $Y(4S)$ resonance. The errors in the yields for the peaking components are *not* obtained from the fit, and *do not* correspond to the statistical error. Rather, they are obtained through error propagation from the original yields, the experimental luminosity, and their weight factors. See appendix L for a detailed explanation.

We also list the expected luminosity at which an observation of the decay can be done with a significance of 5σ .

Data

The number of observed events obtained from performing the reconstruction procedure on data are shown in table 6.6. Only the $B^+ \rightarrow K^+ e^+ e^-$ mode presents events in the signal region. We show the m_{bc} distribution for all decay modes together with their fits in figure 6.11.

The significance of the fit for the $B^+ \rightarrow K^+ e^+ e^-$ decay is 1.35. Following the Feldman and Cousins approach, we calculate the number of expected background

Table 6.5. Expected significance at different luminosities (from MC). The fits required for the combination of various decay modes are shown in appendix K

decay	at 2 ab ⁻¹ (MC luminosity)	at 11.53 fb ⁻¹ (this study)	at 50 ab ⁻¹ (target)	luminosity for $S = 5\sigma$ [fb ⁻¹]
$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	7.24	0.55	36.18	894
$B^+ \rightarrow K^+ \mu^+ \mu^-$	15.46	1.17	77.30	209
$B \rightarrow K \mu^+ \mu^-$	17.38	1.32	86.91	165
$B^0 \rightarrow K_S^0 e^+ e^-$	2.81	0.21	14.05	6096
$B^+ \rightarrow K^+ e^+ e^-$	19.08	1.45	95.41	135
$B \rightarrow K e^+ e^-$	18.01	1.37	90.05	154
$B \rightarrow K \ell^+ \ell^-$	25.01	1.90	125.04	80

Table 6.6. Final number of events in the experimental dataset

decay	observed events	background yield	bkg. events in the signal box	signal yield
$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	1	$1.0^{+1.26}_{-0.69}$	0.16	-
$B^+ \rightarrow K^+ \mu^+ \mu^-$	4	$4.1^{+2.4}_{-1.7}$	0.60	-
$B^0 \rightarrow K_S^0 e^+ e^-$	1	$1.0^{+1.3}_{-0.5}$	0.22	-
$B^+ \rightarrow K^+ e^+ e^-$	7	$4.5^{+2.7}_{-2.0}$	0.74	$1.5^{+2.0}_{-1.3}$

events in the signal region by integration of the background PDF and, for a 90% C.L., report the upper limit of the Branching fractions as

$$\mathcal{B}(B^0 \rightarrow K^0 \mu^+ \mu^-) < 3.2 \times 10^{-6},$$

$$\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-) < 9.8 \times 10^{-7},$$

$$\mathcal{B}(B^0 \rightarrow K^0 e^+ e^-) < 2.2 \times 10^{-6},$$

$$\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-) < 3.5 \times 10^{-6}.$$

We proceed now with the systematic error analysis.

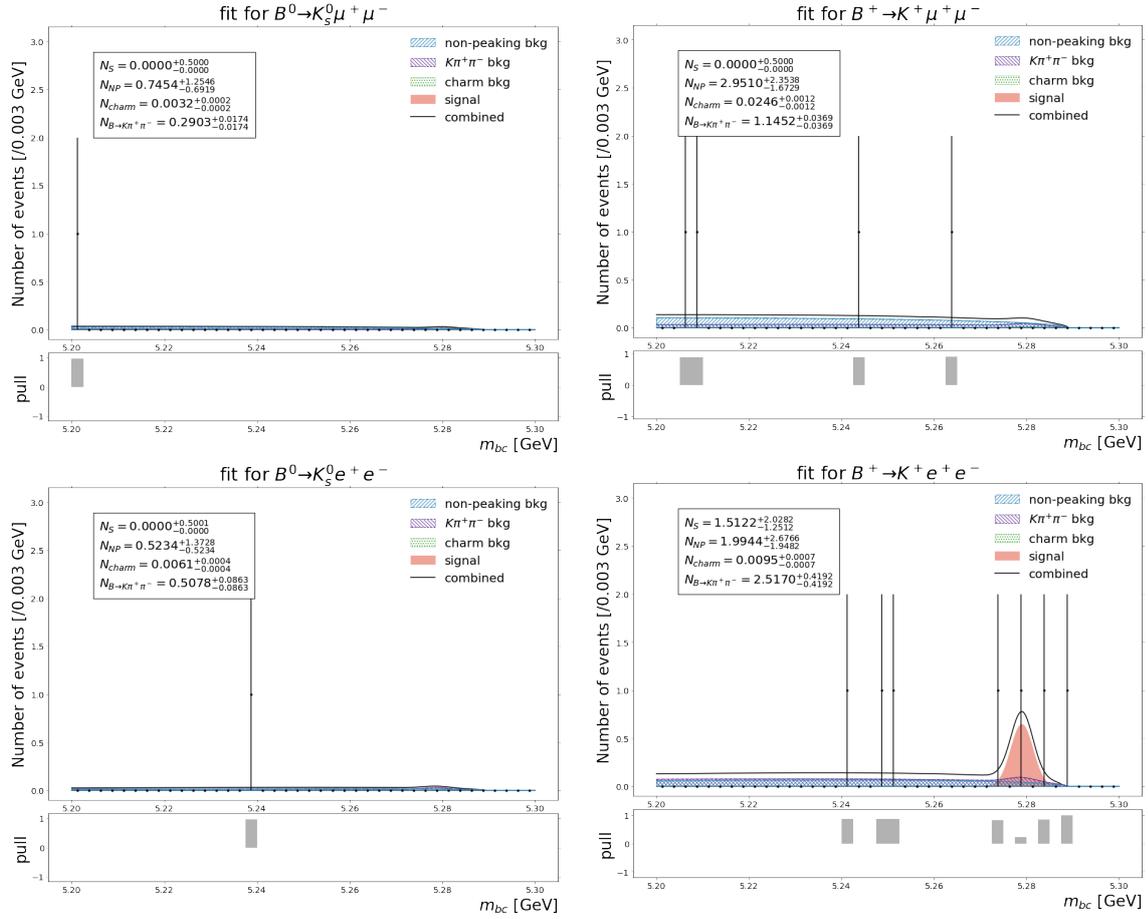


Fig. 6.11 Fit to the m_{bc} distribution for the reconstructed $B \rightarrow K\ell^+\ell^-$ decays in data.

6.2 Error Analysis

6.2.1 Systematic Errors from the Fitting Parameters

We calculate the systematic error in the floating yields due to the PDF parameterization by varying the next parameters by $\pm 1\sigma$ and performing the fit again:

1. signal mean,
2. signal standard deviation,
3. non-peaking χ parameter,
4. charmonium yields, and
5. $B \rightarrow K\pi^+\pi^-$ yields.

Since most of the decay modes do not have events in the signal region, we calculate the error in the background yield inside the signal region, after varying each of the parameters listed. The results are given in table 6.7.

Table 6.7. Relative error of the background yield inside the signal region due to the fit parameterization in experimental data. The error related to the charmonium normalization is negligible.

parameter	decay	relative error in background yield [%]
μ_{signal}	$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	0.0
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.0
	$B^0 \rightarrow K_s^0 e^+ e^-$	0.0
	$B^+ \rightarrow K^+ e^+ e^-$	1.7
σ_{signal}	$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	0.0
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.0
	$B^0 \rightarrow K_s^0 e^+ e^-$	0.0
	$B^+ \rightarrow K^+ e^+ e^-$	3.2
$\chi_{\text{no-peak}}$	$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	11.4
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	4.1
	$B^0 \rightarrow K_s^0 e^+ e^-$	4.6
	$B^+ \rightarrow K^+ e^+ e^-$	4.7
$\text{yield}_{B \rightarrow K \pi^+ \pi^-}$	$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	2.4
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	1.2
	$B^0 \rightarrow K_s^0 e^+ e^-$	3.7
	$B^+ \rightarrow K^+ e^+ e^-$	3.1
total	$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	11.7
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	4.3
	$B^0 \rightarrow K_s^0 e^+ e^-$	5.9
	$B^+ \rightarrow K^+ e^+ e^-$	6.7

We recalculate the 90% C.L. upper bound in the average number of signal events, using as the expected number of background events

$$N'_B = N_B (1 - \delta N_B) \quad (6.5)$$

where N_B is the original number of background events in the signal box (table 6.6), and δN_B is the total relative error (table 6.7).

We incorporate the systematic uncertainties due to the efficiency and the number of B^+B^- pairs produced by adding their relative errors in quadrature and scaling the upper bound accordingly:

$$\text{UB}' = \text{UB} \left(1 + \left(\delta \varepsilon_{\ell\ell}^i \right)^2 + \left(\delta N_{BB}^i \right)^2 \right), \quad (6.6)$$

following [75] (page 145). The number of B meson pairs produced is calculated as

$$N_{BB}^i = \mathcal{L} \times \sigma(e^+e^- \rightarrow B\bar{B}) \times f^i \quad (6.7)$$

where $\sigma(e^+e^- \rightarrow B\bar{B})$ is the cross section for $B\bar{B}$ production at the $Y(4S)$ CM Energy, and f^i is the fraction representing the relative production rates for B^+B^- (f^+) and $B^0\bar{B}^0$ (f^0) pairs [15]. The values used for this calculation, together with their errors, are shown in table 6.8.

Table 6.8. Parameter values and errors for the calculation of the number of $B\bar{B}$ pairs produced

parameter	value	error
\mathcal{L}	11.53 fb ⁻¹	0.12 fb ⁻¹
$\sigma(e^+e^- \rightarrow B\bar{B})$	1.05 nb	0.01 nb
f^+	0.51	0.01
f^0	0.49	0.01

The final reported upper bounds are shown in table 6.9. We have also included the central value obtained from the EUML fit for the $B^+ \rightarrow K^+e^+e^-$ decay (the systematic error reported here is calculated in the next chapter), and the central value obtained from the fits to the MC samples scaled to the current luminosity (shown in figure 6.1). These upper bounds are in agreement with the predictions from the Standard Model and the latest values reported by the PDG (see table 1.3).

Table 6.9. Final results on the branching fractions for the $B \rightarrow K\ell^+\ell^-$ processes. For the central values, the first error is statistical and the second is systematic

decay	upper bound (90% C.L.)	central values [$\times 10^{-7}$]	
		data	scaled MC
$\mathcal{B}(B^0 \rightarrow K^0\mu^+\mu^-)$	3.2×10^{-6}	-	$3.6_{-3.5}^{+13.8} \pm 8.8$
$\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-)$	9.9×10^{-7}	-	$4.6_{-4.2}^{+7.2} \pm 1.2$
$\mathcal{B}(B^0 \rightarrow K^0e^+e^-)$	2.3×10^{-6}	-	$1.3_{-1.3}^{+10.7} \pm 21$
$\mathcal{B}(B^+ \rightarrow K^+e^+e^-)$	3.5×10^{-6}	$6.7_{-5.6}^{+9.1} \pm 1.2$	$5.8_{-4.5}^{+7.2} \pm 22.9$

Chapter 7

Discussion and Conclusions

7.1 Discussion

7.1.1 Branching Fractions

We have performed the first analysis on the $B \rightarrow K\ell^+\ell^-$ processes in the Belle II experiment. In doing so, we calculated the 90% C.L. upper bounds for the branching fractions of these decays, using 11.53 fb^{-1} of data; the obtained values agree with previous measurements.

We also estimated the luminosity required for a 5σ rediscovery of these decays from Monte Carlo simulations; a bold comparison with a similar Belle analysis [76], may indicate that these "rediscovery" luminosities are higher for the muon modes, and very similar for the charged electron mode (see table 7.1). Notice, however, that each study uses different values for the branching fractions: this work uses the most recent PDG values, which are different from the results obtained by Belle. These differences must be taken into account when evaluating the performance of each analysis; we do this by calculating the gain in the pseudo-significance attained with an increment in the number of observed events, and assume that this is the same improvement achieved in the significance. The gain in pseudo-significance is given by

$$g_s = g_S \sqrt{\frac{N_S + N_B}{g_S N_S + N_B}} \quad (7.1)$$

where g_S is the gain in the number of signal events, N_S is the number of signal events, and N_B is the number of background events. For the $B^0 \rightarrow K^0\mu^+\mu^-$ decay, $g_S = 1.65$, $N_S = 48$ and $N_B = 25$ (see table 4.20). This gives a gain in significance of

1.38, and thus the “rediscovery” luminosity scales down by a factor of 1.9, to a value of 469 fb^{-1} , still larger, but much closer, to the Belle one.

Table 7.1. Luminosities required for a 5σ observation, reconstruction efficiencies, measured branching fractions, and background rates of the $B \rightarrow K\ell^+\ell^-$ decay modes in Belle (extrapolated from [76]) and in the MC data of this study

decay		5σ lumi. [fb^{-1}]	reco. eff. [%]	BF [$\times 10^{-7}$]	bkg. rate [$\times 10^{-3}/\text{fb}$]
$B^0 \rightarrow K^0\mu^+\mu^-$	Belle	364	8.5	5.6	14.7
	this study (MC)	894	6.1	3.4	14.7
$B^+ \rightarrow K^+\mu^+\mu^-$	Belle	165	23.6	4.5	75.7
	this study (MC)	209	21.8	4.4	58.4
$B^0 \rightarrow K^0e^+e^-$	Belle	-	5.0	-	21.5
	this study (MC)	6096	5.8	1.6	21.4
$B^+ \rightarrow K^+e^+e^-$	Belle	135	16.6	6.3	52.5
	this study (MC)	135	19.5	5.5	51.9

The larger values for the “rediscovery” luminosity in the muon channels are due to the smaller reconstruction efficiencies. Indeed, following the same methodology as before, we calculate the gain in the number of signal events due to an increase in efficiency for the Belle II analysis. In this case, for the $B^+ \rightarrow K^+\mu^+\mu^-$ decay, $g_S = 1.08$, $N_S = 181$ and $N_B = 130$ (see table 4.21), and the “rediscovery” luminosity becomes 188 fb^{-1} . A similar procedure for the $B^0 \rightarrow K^0\mu^+\mu^-$ decay mode (taking into account the different Branching Fractions as well) gives a “rediscovery” luminosity of 309 fb^{-1} .

We noticed that the Belle study uses a slightly different parameterization for the non-peaking background; we performed the signal extraction procedure on MC using this model, and found that the results are consistent with the ones obtained through the parameterization presented in chapter 6. The details, for the interested reader, are shown in appendix M.

Despite the better rejection power expected from the MVA classifiers (in comparison with the Fisher discriminant used by Belle to reject generic $B\bar{B}$ and continuum events), the background rates are fairly similar among the two studies. The explanation for this is found in the higher pion-to-lepton fake rates, which affect the number of $K\pi^+\pi^-$ background events.

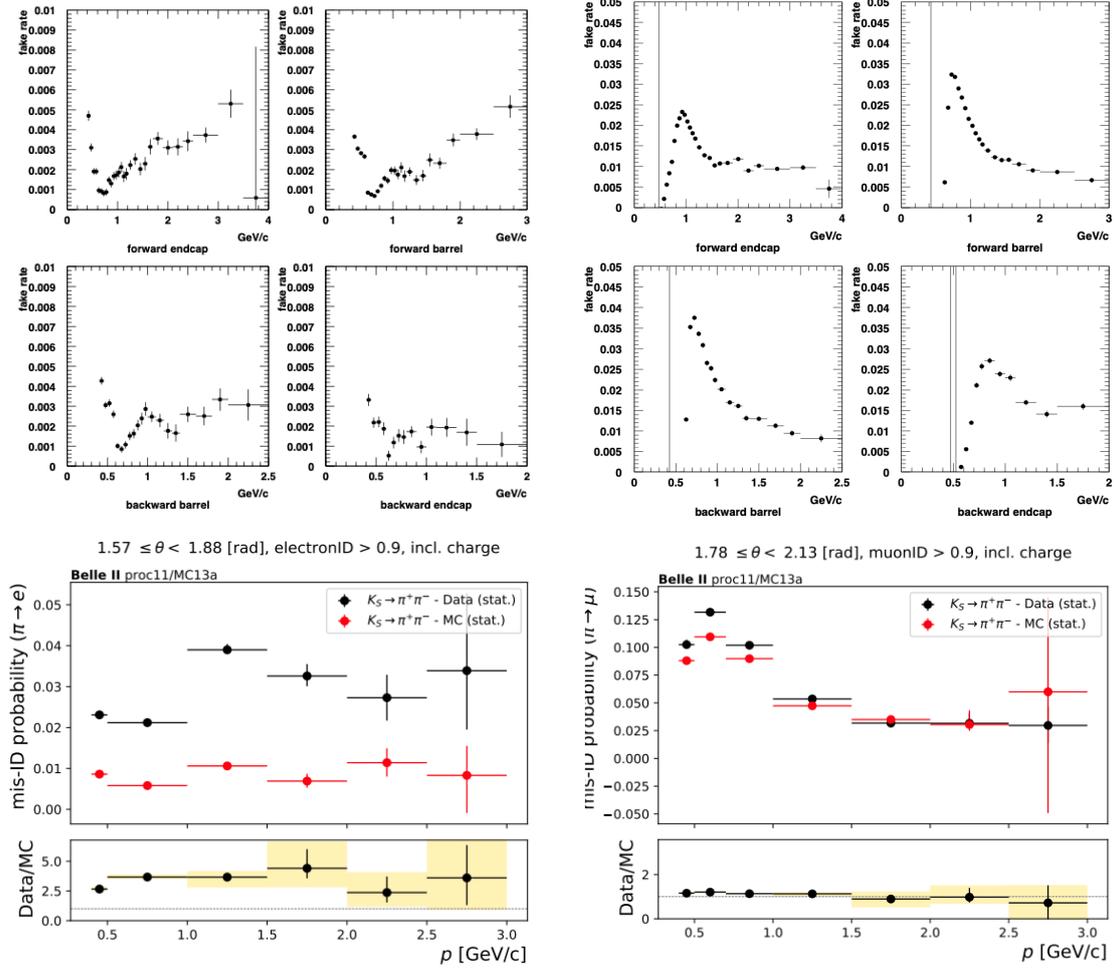


Fig. 7.1 Pion-to-electron (left) and pion-to-muon (right) fake-rates for Belle (top row) and Belle II (bottom). The PID requirements for the Belle plots are $ID_{e:\pi} > 0.2$, and $ID_{\mu:\pi} > 0.9$, where $ID_{\alpha:\beta}$ is the binary ID, defined as the ratio of the likelihoods between hypothesis α and β . For the Belle plots, the fake rates are reported for the next regions: forward endcap ($\theta < 0.9$ rad), forward barrel ($0.9 \leq \theta < 1.57$ rad), backward barrel ($1.57 \leq \theta < 2.04$ rad) and backward endcap ($\theta \geq 2.04$). For the Belle II plots, the figures correspond to the barrel region of the ECL (for the electron fake rates), and to the backward endcap of the KLM, close to the solenoid chimney (for the muon fake rates). The Belle plots were taken from [77], and the Belle II ones from [72].

Figure 7.1 shows the pion to lepton fake rates for certain regions of the detectors. We observe that the pion-to-electron (pion-to-muon) fake rates at Belle II are between 2 to 7 (3 to 8) times larger than Belle's. The reason behind this performance gap is unknown, and additional studies will be required in order to identify and correct it. As so, these larger rates affect the $K\pi^+\pi^-$ contributions: figure 7.2 shows the expected m_{bc} distributions of the $K\pi^+\pi^-$ background in the Belle study; the

procedure followed to obtain it was analogous to the one presented in chapter 6 of this work: all the selections cuts were applied, except for the lepton ID ones, which were reversed. The distributions are described by the sum of an ARGUS function (as in equation M.1) and a Gaussian, for a dataset of 140 fb^{-1} .

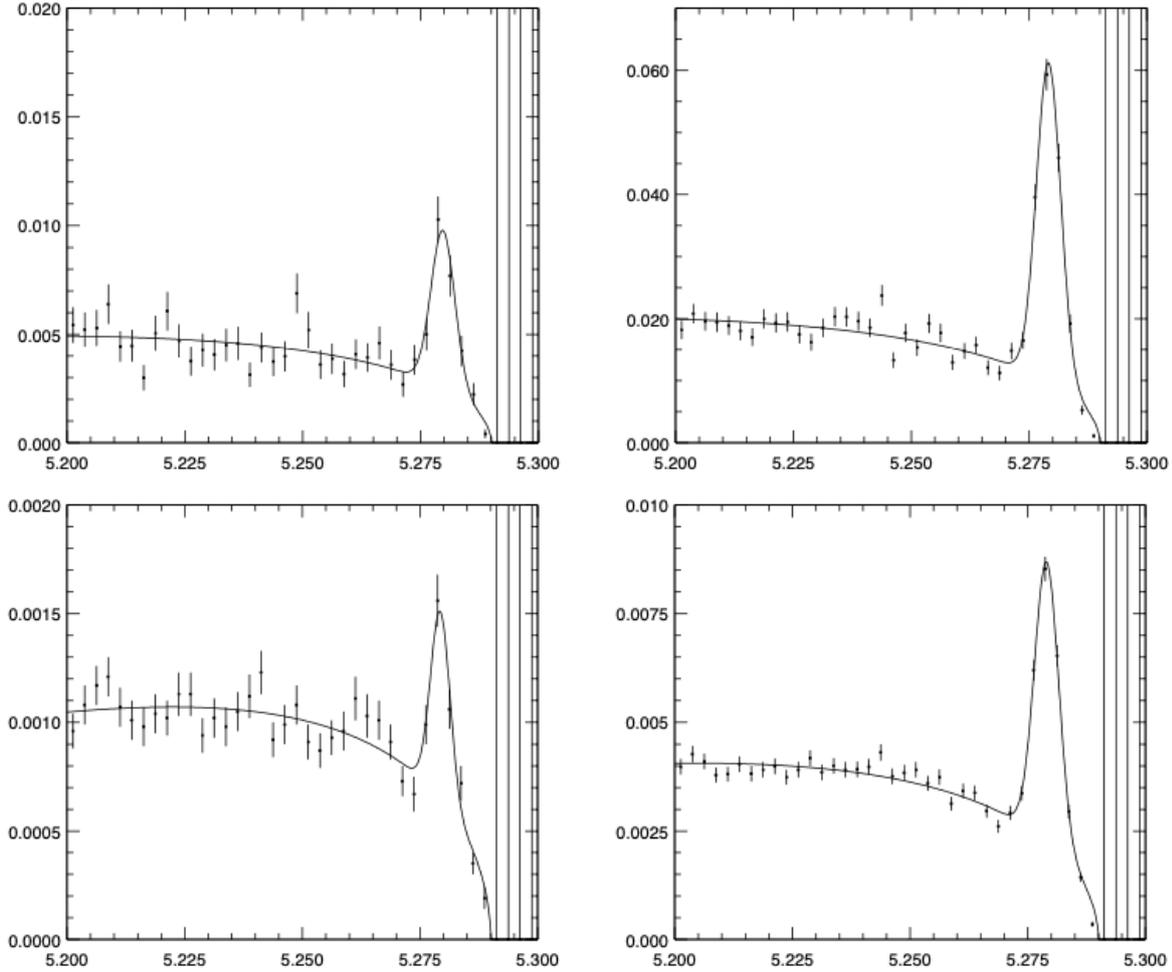


Fig. 7.2 Models of the $K\pi^+\pi^-$ background in the Belle analysis for $B^0 \rightarrow K_S^0\mu^+\mu^-$ (top left), $B^+ \rightarrow K^+\mu^+\mu^-$ (top right), $B^0 \rightarrow K_S^0e^+e^-$ (bottom left), and $B^+ \rightarrow K^+e^+e^-$ (bottom right) from 140 fb^{-1} of data. Images obtained from [77].

A comparison between these and the distributions obtained in this study (shown in section K.3 of appendix K for the MC samples) reveals that the $K\pi^+\pi^-$ component is most than 10 times larger than at Belle¹, even with a correctly calibrated set of subdetectors (since the subdetectors in MC are assumed to be properly calibrated).

¹Recall that these distributions are obtained for the luminosities of each study. Hence, the distributions of appendix K correspond to the expected number of $K\pi^+\pi^-$ events in 11.53 fb^{-1} of data, whereas the ones in figure 7.2 correspond to an integrated luminosity of 140 fb^{-1} .

In order to measure the effect of these differences between the two studies, we re-calculate the background rate at Belle II while conservatively scaling down by a factor of 5 the background component from $K\pi^+\pi^-$ events. The results, depicted in table 7.2, effectively show that once these conditions are taken into account, the background rates for the Belle II analysis become smaller than the ones at Belle.

Table 7.2. Background rates for the Belle II analysis with a reduced $K\pi^+\pi^-$ component

decay	background rate [$10^{-3}/\text{fb}$]
$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	9.5
$B^+ \rightarrow K^+ \mu^+ \mu^-$	43.9
$B^0 \rightarrow K_s^0 e^+ e^-$	19.6
$B^+ \rightarrow K^+ e^+ e^-$	44.4

The significance is also expected to improve as these background rates go down. Taking the charged muon mode as an example, we calculate the gain in significance due to a reduction by a factor of $g_{\text{bkg}} = 0.75$ in the background levels as:

$$g_s = \sqrt{\frac{N_S + N_B}{N_S + g_B N_B'}} \quad (7.2)$$

which gives $g_s = 1.06$. This, together with the increment in efficiency up to the same level as Belle, would bring the “rediscovery” luminosity down to 166 fb^{-1} . This shows that the gap in the pion-to-lepton fake rates is a central aspect to work on in order to maximize the sensitivity in these decay channels.

We continue now with the projection of the branching fractions’ sensitivities for higher luminosities. As shown in the SBelle study [78], the reconstruction efficiency for the $B^0 \rightarrow K_s^0 J/\psi$ channel is smaller by 8.8% at the highest expected background rate—which is assumed to be 20 times larger than at Belle²—in comparison with the base (Belle) case due to a higher occupancy in the VXD, while the performance of the PID systems is expected to remain unaltered. In order to incorporate this into the projections, and following a conservative approach, we multiply the efficiencies of each decay mode by a luminosity dependent scaling factor. This dependence is assumed to be of the form

$$\exp\left(-k \left[\frac{\mathcal{L} - \mathcal{L}_0}{\mathcal{L}_0}\right]\right),$$

²This is, the number of hits recorded by each subdetector is 20 times larger.

where \mathcal{L}_0 is the average instantaneous luminosity for the accelerator during this study (which we take as $2 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$) and k is chosen so as to make this factor equal to 0.91 when $\mathcal{L} = 8 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$; we pick this functional form since it is the one that adjusts the best to the simulation points obtained in the SBelle study (in contrast with a power law, or a logarithmic functional). We make the conservative simplification of assuming that this efficiency drop is the same for all decay modes³.

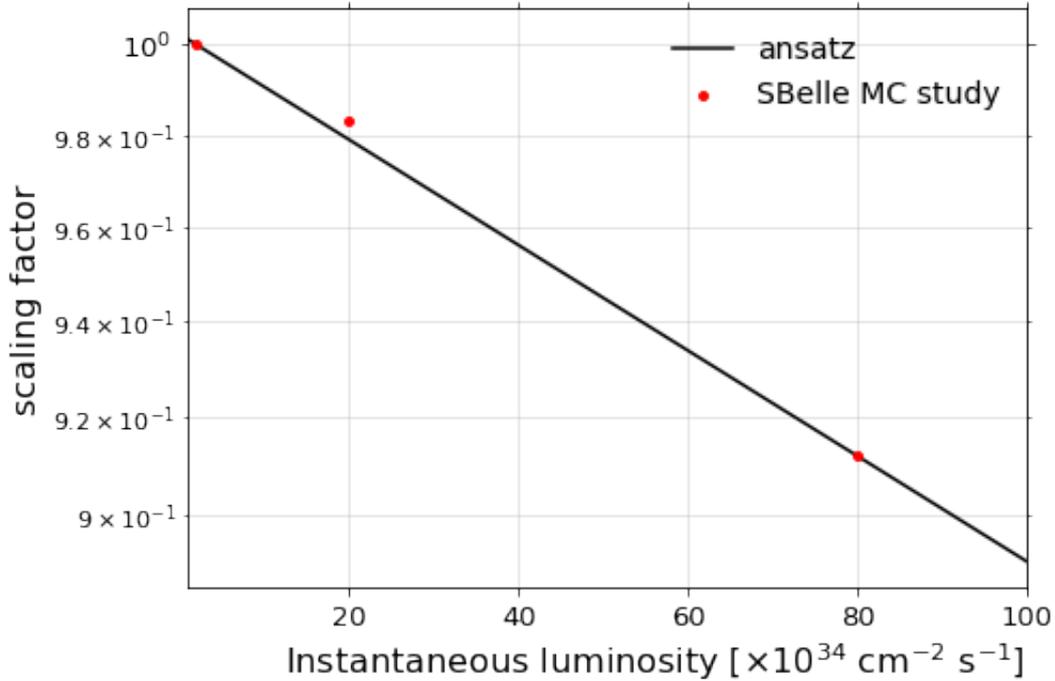


Fig. 7.3 Scaling factor for the reconstruction efficiencies in function of the instantaneous luminosity. The SBelle study can be found at [78].

We also assume that the signal to background ratio remains constant as the luminosity goes up; this is justified provided that, inside the beam-induced background components, the ones that present a larger increase with the luminosity are the Touschek scattering⁴—which scales with the square of the beam current, and with the inverse of the beam vertical size—, and the physics backgrounds related to e^+e^-

³This is not really the case, as the drop in efficiency in the $B^0 \rightarrow K_S^0 J/\psi$ decay mode has mostly to do with the presence of two (or three, if a vertex fit is also applied to the B meson) vertices that have to be correctly reconstructed in a more contaminated environment. The decays of interest in this study involve at most two vertices, for which we only fit one, so the efficiency drop is most likely less severe. This effectively results in a conservative estimation of the efficiency drop, and the real effect will be, more likely, less drastic.

⁴Touschek scattering refers to the energy transfer between particles inside the same beam bunch; once the energy of both particles fall outside of the dynamic aperture of the ring, their trajectory starts to deviate from the design orbit. After a while, these particles end up hitting the beampipe

collisions, namely radiative Bhabha scattering and two photon processes —which scale linearly with the luminosity. The other components of the beam background: synchrotron radiation (photon emission by the particles being bent in a magnetic field) and Beam-gas scattering (Coulomb scattering of the bunch particles by residual gas molecules), scale with the beam energy and the magnetic field strength, and with the beam current and the pressure inside the beam pipe, respectively, and thus their relative contributions are expected to become more and more negligible as the luminosity increases.

As the beam background arising from physics processes presents a similar dependence on luminosity as the signal components, its ratio to the signal component is effectively constant throughout the experiment. On the other side, the Touschek behaviour is hard to model, but we expect it to be non-critical, as there are plans to mitigate its effect by tuning the beam optics and the collimator settings and improving the vacuum condition. We are further assuming that the background rejection performance will not depend on the instantaneous luminosity; we justify this assumption by noting that the degradation in pion fake rates is expected to be less than 1% in the TOP counter [45], and that the ECL and KLM, which are central in the lepton PID calculation, are far enough from the IR to be significantly affected. The ARICH detector, on the other side, has been shielded with polyethylene in order to cope with the neutron background produced by photons impinging on the beam pipe.

Figure 7.4 shows the planned luminosity profile for the SuperKEKB accelerator. Based on this, we assume that the first 20 ab^{-1} of data are taken in conditions that do not present a significant decrement to the reconstruction efficiencies. The last 30 ab^{-1} are collected under a more demanding environment; we assume an instantaneous luminosity of $6.5 \times 10^{-35} \text{ cm}^{-2} \text{ s}^{-1}$ (which corresponds to a new, more conservative, target, owing to the technical difficulties found during the first set of physics runs), and correct the efficiency with the scaling factor obtained previously.

We perform projections for two possible scenarios. The conservative one, in which the efficiencies are the ones derived in this study; and the optimistic one, in which new efficiencies are calculated assuming the next improvements are introduced:

walls, producing electromagnetic showers. If these showers happen near the interaction region, they can produce hits in the detector. The rate of this background is proportional to the beam current, and inversely proportional to the beamsizes; given these dependencies, it is expected to be one of the major sources of beam background at SuperKEKB at the target luminosity.

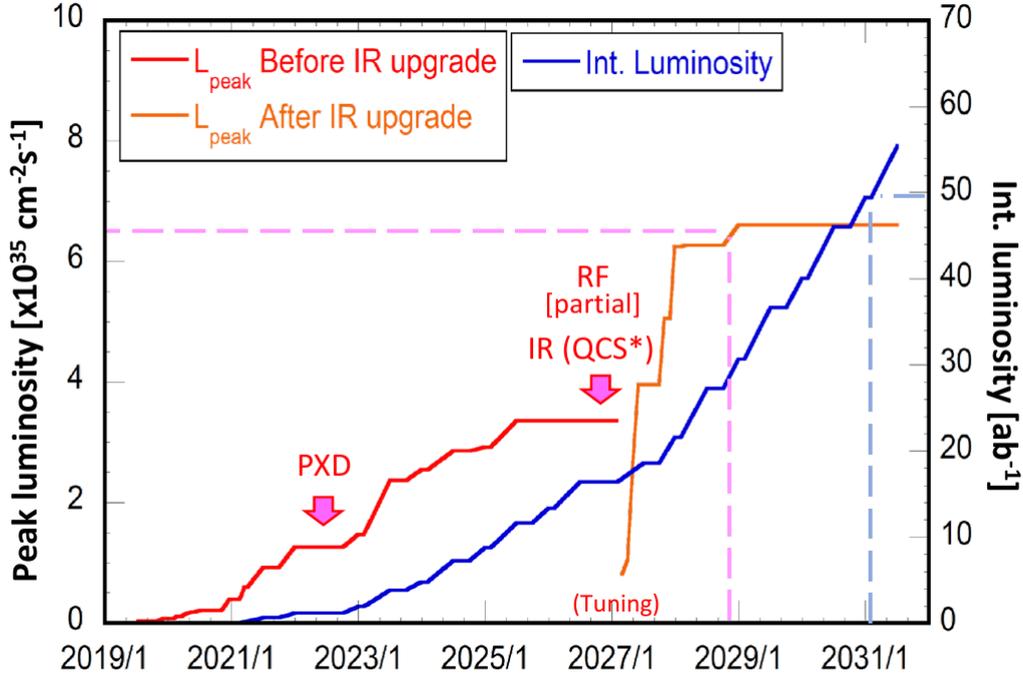


Fig. 7.4 Plan for the instantaneous and integrated luminosities at the SuperKEKB accelerator. PXD refers to the exchange of the subdetector, and RF/IR refers to improvements to the RF gun (power upgrade) and the Interaction Region optics, required to achieve the design instantaneous luminosity.

1. A K_S^0 reconstruction procedure based on MVAs, as Belle had during its later years of operation. Assuming the same efficiencies as then (see, for example, [30]), this constitutes an improvement of 30% in the efficiency of the $B^0 \rightarrow K_S^0 e^+ e^-$ decay, and of 40% in the efficiency of the $B^0 \rightarrow K_S^0 \mu^+ \mu^-$ decay.
2. An improved PID system. The efficiency gain is estimated by the next means: taking as proxy the MC PID efficiencies, we average the data to MC efficiency ratios of the events in the signal MC samples. This revealed that there is a difference of 6% (10%) in the reconstruction efficiency of the decays between data and MC due to electron (muon) PID; hence, we conservatively expect an improvement of this level. The same analysis yields an expected increase of 2% in the reconstruction efficiency due to improvements in the kaon ID⁵.

⁵We are, again, being moderate in these estimations. The performance—both real and simulated—of the detectors is expected to improve in the future, due to the introduction of new PID techniques based on Machine Learning, and on the upgrade of the detector. The effects of these changes, however, is hard to quantize at the current stage, so we offer here a very simple optimistic scenario.

3. We also take into account the reduction of background events in the signal region due to PID improvements and the introduction of stronger classifiers (which may contain, for example, information of a vertex fit performed on the B meson decay), by removing the degradation factor in the background-to-signal ratios introduced previously to account for the different performance in the pion-to-lepton fake rate between data and MC.

We model the relative statistical error as

$$\delta\mathcal{B}_{\ell\ell}^i(\text{stat}) = \sqrt{\frac{1}{\alpha_{\ell\ell}^i \mathcal{L}}} \quad (7.3)$$

where $\alpha_{\ell\ell}^i$ is a proportionality factor that gives the number of events per unit of luminosity in the signal region. It is calculated as

$$\alpha_{\ell\ell}^i = \frac{2 \times \sigma(e^+e^- \rightarrow B\bar{B}) \times f^i \times \mathcal{B}(B^i \rightarrow K^i \ell^+ \ell^-) \times \epsilon_{\ell\ell}^i}{1 + r_{\ell\ell}^i} \quad (7.4)$$

where the branching fraction values are taken from the latest PDG data [15]; $r_{\ell\ell}^i$ represents the background to signal ratio for each decay inside the signal region; these are obtained from the MC analysis performed in chapter 4, and their values are shown in table 7.3. We scale these ratios by a factor of 4 (3) for the charged (neutral) electron mode, in order to take into account the much larger peaking component in data, in comparison to MC (see figure 6.5). We include the effects of the higher

Table 7.3. Background to signal ratio inside the signal region for MC

decay	$r_{\ell\ell}^i$
$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	0.52
$B^+ \rightarrow K_+ \mu^+ \mu^-$	0.72
$B^0 \rightarrow K_S^0 e^+ e^-$	1.52
$B^+ \rightarrow K^+ e^+ e^-$	0.46

instantaneous luminosity and the reconstruction improvements only in the relative statistical error.

The systematic uncertainty associated with the efficiency is modeled as

$$\delta\epsilon_{\ell\ell}^i = \sqrt{\frac{a_{\ell\ell}^i}{\mathcal{L}} + b_{\ell\ell}^i} \quad (7.5)$$

and the parameters $a_{\ell\ell}^i$ and $b_{\ell\ell}^i$ are obtained from a regression on the efficiency errors calculated at 11.53 fb^{-1} and 50 ab^{-1} in chapter 5. Their values are given in table 7.5.

We keep the uncertainty due to the counting of $B\bar{B}$ pairs constant at a conservative value of 2%.

The uncertainties due to the fitting procedure are expected to decrease as the data sample grows, since the values of the parameters get better constrained; however, the exact form of the relation is not as straightforward as for the previous components. We use the ansatz:

$$\delta \text{fit}_{\ell\ell}^i = \left(\frac{\beta_{\ell\ell}^i}{\mathcal{L}} \right)^n. \quad (7.6)$$

In order to estimate $\beta_{\ell\ell}^i$ and n , we calculate the systematics of the fit in the MC data by varying the values of the fixed parameters by $\pm 1\sigma$ and taking the largest difference between the original signal yields and the ones obtained after each variation. The results are shown in table 7.4. We repeat the same procedure for the fit result of the $B^+ \rightarrow K^+ e^+ e^-$ decay in the experimental data (table 7.6), and derive n as

$$n = \frac{\log \delta \text{fit}_{ee,\text{exp}}^+ - \log \delta \text{fit}_{ee,\text{MC}}^+}{\log \mathcal{L}_{\text{MC}} - \log \mathcal{L}_{\text{exp}}}. \quad (7.7)$$

We obtain a value of $n = 0.42$. The values of the β parameters are listed in table 7.5

The projected sensitivities for the multiple scenarios are shown in figure 7.5, together with the error in the latest PDG report; from these plots, it is clear that the statistical error will dominate the uncertainty in the branching fraction measurements all the way up to the target luminosity for the Belle II experiment. The non-reducible systematic error will start playing a major role for the charged modes above 10 ab^{-1} , because of the larger statistics. Notice that the expected sensitivity will overcome the PDG reports at around 1.5 ab^{-1} (1 ab^{-1}) for the charged (neutral) electron mode, in the conservative approach. On the other side, in the optimistic scenario, the precision of the electron modes receives a significant boost from the background reduction expected due to the PID and MVA improvements; with these, the integrated luminosities required to deliver world leading results in terms of precision are 450 fb^{-1} and 750 fb^{-1} for the neutral and charged modes, respectively. This is because the most precise measurements for these decays are performed in B factories, so overcoming the integrated luminosity of Belle with a similar reconstruction performance (or a smaller luminosity with an improved reconstruction performance) will allow Belle II to deliver more precise results; on the other side, the branching fractions for the muon modes are dominated by the LHCb studies

Table 7.4. Relative error of the signal yield due to the fit parameterization in Monte Carlo. The error related to the charmonium normalization is negligible.

parameter	decay	relative error in signal yield [%]
μ_{signal}	$B^0 \rightarrow K^0 \mu^+ \mu^-$	3.9
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.4
	$B^0 \rightarrow K^0 e^+ e^-$	4.3
	$B^+ \rightarrow K^+ e^+ e^-$	0.3
σ_{signal}	$B^0 \rightarrow K^0 \mu^+ \mu^-$	3.8
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	1.6
	$B^0 \rightarrow K^0 e^+ e^-$	4.1
	$B^+ \rightarrow K^+ e^+ e^-$	1.6
$\chi_{\text{no-peak}}$	$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.3
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.2
	$B^0 \rightarrow K^0 e^+ e^-$	0.4
	$B^+ \rightarrow K^+ e^+ e^-$	0.0
$\text{yield}_{B \rightarrow K \pi^+ \pi^-}$	$B^0 \rightarrow K^0 \mu^+ \mu^-$	1.0
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.6
	$B^0 \rightarrow K^0 e^+ e^-$	1.0
	$B^+ \rightarrow K^+ e^+ e^-$	0.1
total	$B^0 \rightarrow K^0 \mu^+ \mu^-$	5.5
	$B^+ \rightarrow K^+ \mu^+ \mu^-$	1.8
	$B^0 \rightarrow K^0 e^+ e^-$	6.1
	$B^+ \rightarrow K^+ e^+ e^-$	1.6

of these decays; thus, a larger dataset is required in order to surpass the current sensitivity.

As a cross-check of these predictions, we compare their values at 11.53 fb^{-1} with the errors in the central values obtained from data and MC in the previous chapter (see table 6.9). They are displayed in table 7.7; we see excellent agreement between data and the projection for the $B^+ \rightarrow K^+ e^+ e^-$ decay; the agreement between the projection and MC is good for the charged channels, where the statistics are larger and thus the relative error is better described by the model. For the neutral modes,

Table 7.5. Values for the parameters a , b and β modeling the systematic uncertainty in the branching fractions.

decay	a [fb ⁻¹]	b [$\times 10^{-3}$]	β [fb ⁻¹]
$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.40	0.24	2.11
$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.09	0.54	0.15
$B^0 \rightarrow K^0 e^+ e^-$	0.40	0.24	2.66
$B^+ \rightarrow K^+ e^+ e^-$	0.11	0.56	0.12

Table 7.6. Relative error in the central value of the signal yield due to the fit parameterization for the $B^+ \rightarrow K^+ e^+ e^-$ decay in the experimental data. The error related to the charmonium normalization is negligible.

parameter	relative error in signal yield [%]
μ_{signal}	0.00
σ_{signal}	13.86
$\chi_{\text{no-peak}}$	3.55
$\text{yield}_{B \rightarrow K \pi^+ \pi^-}$	2.53
total	14.53

the discrepancy between the projections and the MC errors is larger; this is again explained by their smaller data sample, for which the approximation made when modeling the statistical error does not hold.

Table 7.7. Relative errors [%] in the branching fractions of the $B \rightarrow K \ell^+ \ell^-$ decays at 11.53 fb⁻¹

decay	from data	from MC	from projection (conservative)	from projection (optimistic)
$\mathcal{B}(B^0 \rightarrow K^0 \mu^+ \mu^-)$	-	387	255	195
$\mathcal{B}(B^+ \rightarrow K^+ \mu^+ \mu^-)$	-	155	133	111
$\mathcal{B}(B^0 \rightarrow K^0 e^+ e^-)$	-	834	644	367
$\mathcal{B}(B^+ \rightarrow K^+ e^+ e^-)$	137	126	139	100

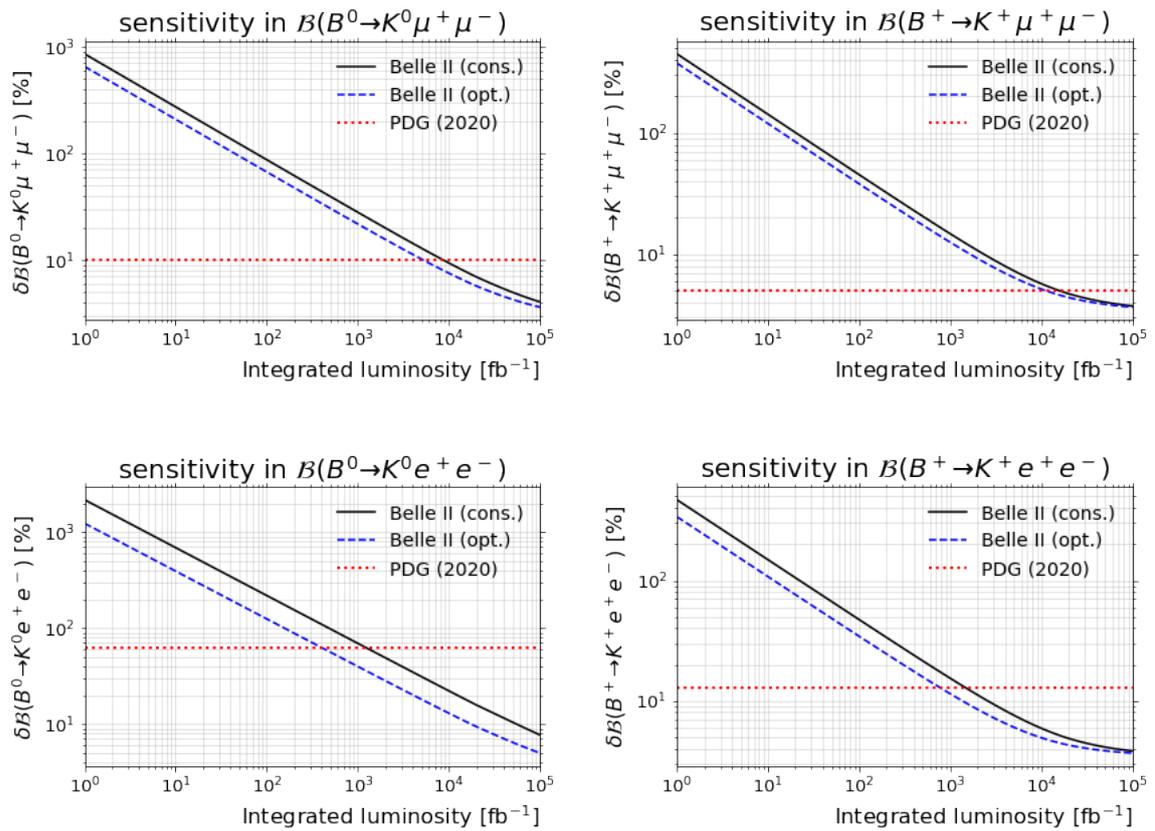


Fig. 7.5 Relative error in the branching fraction measurement of the $B \rightarrow K \ell^+ \ell^-$ processes, as a function of the integrated luminosity. The blue, dashed line corresponds to the optimistic scenario, whereas the black, solid line is the conservative one.

7.1.2 R_K

Due to the limited statistics it was not possible to obtain a first measurement of this ratio in the current study; however, we have gathered almost all the pieces needed for a more realistic estimation of the Belle II sensitivity on this observable. In order to perform a comparison with the latest results from LHCb and Belle, we will limit the following calculations to the low q^2 region⁶, corresponding to the range $q^2 \in [1, 6]$ GeV².

Recall that, in this study, R_K was intended to be obtained as the weighted average of the ratio from the charged (R_K^+) and neutral (R_K^0) channels:

$$R_K = \frac{(\Delta R_K^+ \Delta R_K^0)^2}{(\Delta R_K^+)^2 + (\Delta R_K^0)^2} \left(\frac{R_K^+}{(\Delta R_K^+)^2} + \frac{R_K^0}{(\Delta R_K^0)^2} \right). \quad (7.8)$$

Here, ΔR_K^i is the square sum of the statistical and systematic error for the ratio in the i channel. The formula for R_K^i is:

$$R_K^i = \frac{\varepsilon_{ee}^i N_{\mu\mu}^i}{\varepsilon_{\mu\mu}^i N_{ee}^i}, \quad (7.9)$$

The statistical error in R_K^i is the sum of squares of the statistical error of the yields for the electron and the muon modes:

$$\delta R_K^i (\text{stat}) = \frac{1}{\sqrt{\mathcal{L}}} \sqrt{\frac{1}{\alpha_{ee}^i} + \frac{1}{\alpha_{\mu\mu}^i}}. \quad (7.10)$$

This time, in order to take into account the fact that we are limiting the measurement to the low q^2 region, we multiply $\alpha_{\ell\ell}^i$ in equation 7.10 by a factor $C_{\ell\ell}^i$, which corresponds to the ratio between the averaged branching fraction of the decay for $q^2 \in [1, 6]$ GeV² and the total branching fraction. These ratios are obtained from the Standard Model predictions using flavio [43], with the Wilson coefficients derived from [28], and the form factors calculated in [27]. Their values are found to be independent of the lepton and B meson flavors up to a precision of 1%, and equal to 0.34.

On the other side, the systematic error for R_K^i is a combination of four terms: the errors in the reconstruction efficiencies $\varepsilon_{\ell\ell}^i$, and the systematic errors in the signal yields due to the fitting parameters. Notice, however, that most of the correction

⁶Once again, q corresponds to the dilepton invariant mass of the decay, $q = m(\ell^+ \ell^-)$.

factors for the efficiency cancel when taking the ratio: only those related to the MVA classifiers and the lepton ID remain. We thus recalculate the systematic error in the efficiency arising only from these two components. The results are shown in table 7.8.

Table 7.8. Systematic error in the efficiency arising only from the lepton ID and MVA correction factors

decay	$\Delta\varepsilon_{\text{sys}}$ [%]	$\Delta\varepsilon_{\text{sys}}$ at 50 ab^{-1} [%]
$B^0 \rightarrow K_s^0 \mu^+ \mu^-$	1.09	0.02
$B^+ \rightarrow K^+ \mu^+ \mu^-$	1.68	0.02
$B^0 \rightarrow K_s^0 e^+ e^-$	1.33	0.02
$B^+ \rightarrow K^+ e^+ e^-$	1.73	0.03

The new values for the parameters a and b are shown in table 7.9.

Table 7.9. Values for the parameters a , b modeling the systematic uncertainty due to the efficiency in R_K .

decay	a [fb^{-1}]	b [$\times 10^{-6}$]
$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.37	2.86
$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.10	0.73
$B^0 \rightarrow K^0 e^+ e^-$	0.40	1.71
$B^+ \rightarrow K^+ e^+ e^-$	0.10	0.29

Figure 7.6 shows the expected relative statistical and systematic errors in R_K as a function of the luminosity, from the parameterization derived, in the low q^2 region. As R_K is the addition of two terms, its relative error is not computable without knowing the values of R_K^+ and R_K^0 . In the figures shown, we assume that lepton universality holds, and thus the measured values for R_K^+ and R_K^0 are equal to 1.

The statistical error dominates the uncertainty up to the target luminosity, and becomes smaller than the LHCb (Belle) one at around 24.7 ab^{-1} (1.5 ab^{-1}), and around 15.6 ab^{-1} (1 ab^{-1}) in the optimistic scenario; the overall error is thus competitive at these luminosities, where the biggest gain in precision is obtained by the reduction of the background components in the signal region. This is larger than the

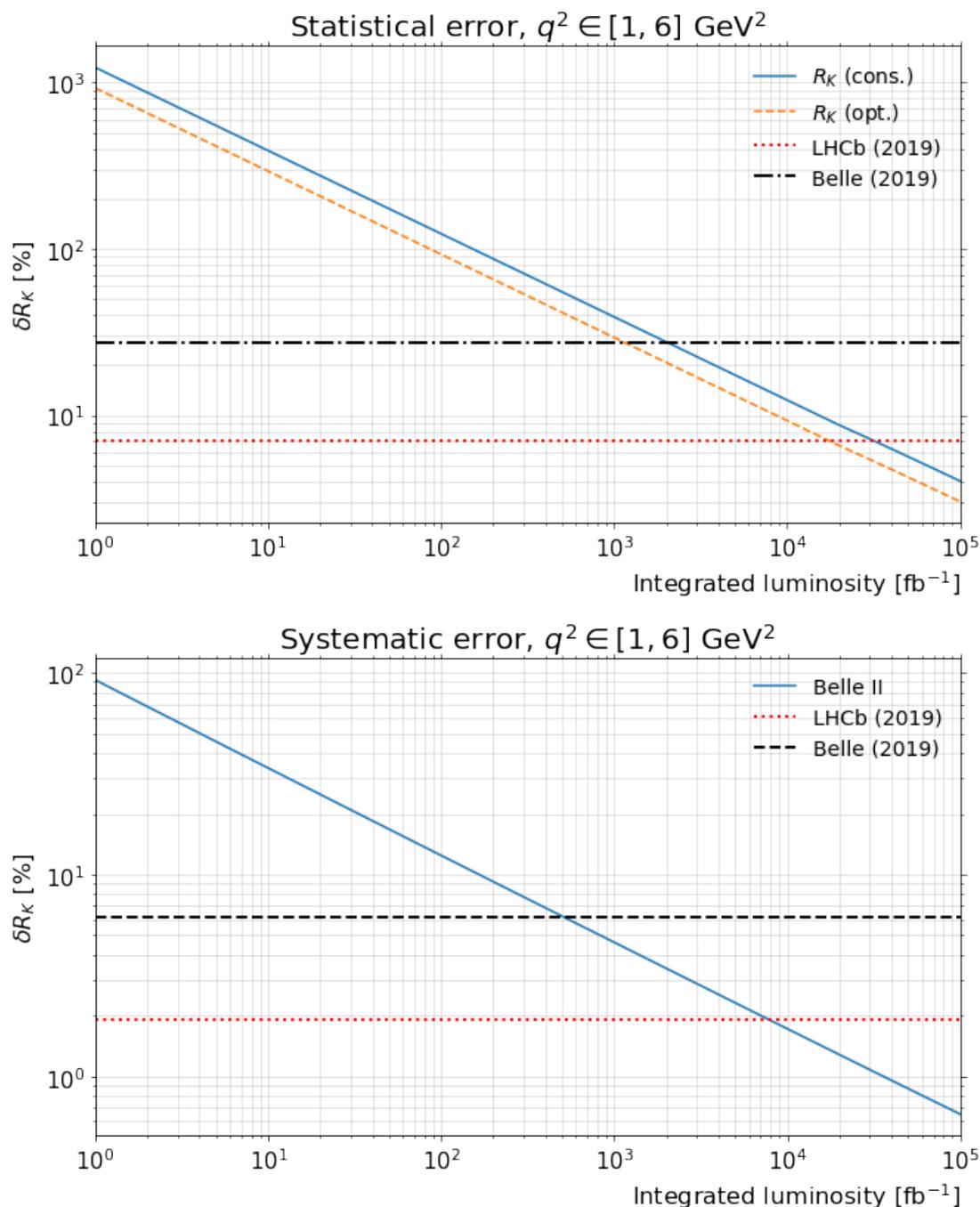


Fig. 7.6 Statistical (top) and systematic (bottom) errors for the R_K observable as a function of the luminosity, over the low q^2 region. Notice that, in our model, the systematic error is independent of the possible upgrades introduced in the analysis, so it has the same values for either the conservative or the optimistic scenarios.

original expectations obtained from the extrapolation of Belle results⁷. We find the explanation for this in:

⁷Bear in mind, however, that these expectations were meant to provide only the order of magnitude of the required luminosity, and not to be taken as exact values.

- the way most of the cuts are derived in this study. At each step, we aimed to maximize the pseudo-significance, which implies loses in efficiency in exchange for a cleaner signal peak. On the contrary, the Belle study emphasizes signal efficiency: it uses looser ΔE and MVA cuts, and it also trains a single classifier for all decay modes to suppress both continuum and $B\bar{B}$ background.
- The particle reconstruction efficiency; as the projection in figure 7.6 shows, a better PID (which would imply a cleaner, larger, data sample) translates into smaller thresholds for the integrated luminosity.

As a way of testing our first hypothesis, we recalculated the cut values for both the continuum and the $B\bar{B}$ classifiers trained for this study to obtain an efficiency drop of 5% after *each one of them is applied*, which is the loss reported by Belle [30]. We then estimate the new reconstruction efficiencies for each decay mode, and the relative statistical error for these new efficiencies; the procedures are detailed in appendix N. It is observed that, with these new cuts, the procedure presented in this work achieves the same relative statistical error than the Belle study at a luminosity of 1 ab^{-1} (673 fb^{-1} in the optimistic scenario) —provided that the signal yields can be extracted precisely—, while obtaining higher values for the background purity gain in three of the four decay modes. Also, with these new cuts, the measurements on R_K are competitive against the LHCb results at 15.6 ab^{-1} in the conservative scenario, and at 11 ab^{-1} in the optimistic one (comparing the statistical error only).

However, and despite this purity gain, the background present in the Belle II data is still larger than in Belle. This new approach would then require deeper control of the PDF components, in order to correctly extract the yields from a more contaminated data sample; this has been proven to be already difficult for the neutral electron mode in the current study. A multidimensional fit can be used to alleviate this; the latest Belle study uses ΔE and the MVA classifier output distributions to this end. This strategy, however, would increase the background levels more, since ΔE , which is deterministic in removing generic $B\bar{B}$ background events, cannot longer be used as an input for the classifier; it also increases the sources of systematic uncertainties in the measurement, possibly reducing the impact of the improvement achieved by the higher statistics: indeed, the Belle study reports a systematic error of 6%, which is almost the same as the total error in the LHCb result. The exchange between the size of the statistical and systematic uncertainties may be evaluated at different luminosities in order to select the procedure which delivers the most precise results. We leave the choice to future studies.

Finally, from the projections in 7.6, we observe that the Belle II experiment will only be able to settle down the anomaly in R_K with a 5σ precision in the optimistic scenario; effectively, for the conservative case, an integrated luminosity of 56 ab^{-1} would be required to reach this level of precision, whereas in the optimistic case, the same sensitivity is achieved at 35 ab^{-1} .

7.2 Conclusions

We have performed the first study on the $B \rightarrow K\ell^+\ell^-$ processes, which constitute very sensitive probes for New Physics, in the Belle II experiment. We analyzed 11.53 fb^{-1} of data taken at a Center of Mass Energy equal to the $Y(4S)$ resonance mass, which constitutes roughly 12.1 million of $B\bar{B}$ pairs, employing a novel Bremsstrahlung recovery method, and using Boosted Decision Trees in order to reduce the background from $q\bar{q}$ and $B\bar{B}$ events. In the process, we also measured the ratios

$$\frac{\mathcal{B}(B \rightarrow KJ/\psi [e^+e^-])}{\mathcal{B}(B \rightarrow KJ/\psi [\mu^+\mu^-])} = 0.99 \pm 0.09 \text{ (stat)} \pm 0.01 \text{ (sys)},$$

$$\frac{\mathcal{B}(B \rightarrow K\psi(2S) [e^+e^-])}{\mathcal{B}(B \rightarrow K\psi(2S) [\mu^+\mu^-])} = 1.03 \pm 0.41 \text{ (stat)} \pm 0.01 \text{ (sys)}$$

which are in agreement with the latest PDG reported values. In the calculation for R_K , these ratios provide a way of replacing the systematic uncertainties related to the reconstruction efficiency with the uncertainty in their values. Thus, future measurements are of utmost relevance for the analysis of the Flavor Changing Neutral Currents, specially for experiments in which the reconstruction efficiencies are highly dependent on the lepton flavor.

Given the luminosity available and the values of the branching fractions for the decays of interest, only the decay $B^+ \rightarrow K^+e^+e^-$ had events populating the signal region. We calculated the 90% C.L. upper bound for the branching fractions of the four decay modes as

$$\mathcal{B}(B^0 \rightarrow K^0\mu^+\mu^-) < 3.2 \times 10^{-6},$$

$$\mathcal{B}(B^+ \rightarrow K^+\mu^+\mu^-) < 9.9 \times 10^{-7},$$

$$\mathcal{B}(B^0 \rightarrow K^0e^+e^-) < 2.3 \times 10^{-6},$$

$$\mathcal{B}(B^+ \rightarrow K^+e^+e^-) < 3.5 \times 10^{-6},$$

which are consistent with the PDG reported measurements and with theoretical predictions. The largest contribution to the systematics of the branching fractions comes from the fitting parameters and the correction factor in efficiency due to the MVA classifier cuts. Both are expected to decrease as the data size increases and, at higher luminosities, the error in the tracking efficiency is expected to be dominant. Most of this error has its origin in the trigger efficiency uncertainty; hence, future studies must strive for a better understanding of the triggering system at Belle II. We also noted that the statistical significance for the muon modes is lower than in the Belle experiment, and found that this is explained by both a smaller reconstruction efficiencies, and the larger backgrounds. The pion to electron fake rates were observed to be significantly larger than at Belle, and further studies to elucidate the cause for this are of utmost importance, since a reduction of these rates has the potential to increase the sensitivity of the measurements by more than 20%.

We carried out an error analysis on the R_K ratio, in which the statistical and systematic components were modeled as functions of the luminosity, and the parameters for said models were deduced from experimental data (when available) and from Monte Carlo (when experimental data was missing). Following the reconstruction procedure presented in this work, the statistical error is expected to dominate the uncertainty of R_K all the way up to the Belle II target luminosity of 50 ab^{-1} ; the projections, although in gross agreement with the values obtained from the extrapolation of Belle results, are more realistic, and give a threshold of 24.7 ab^{-1} (1.5 ab^{-1}) in the luminosity in order to overcome the precision achieved by the LHCb (Belle) measurements. These values are lowered to 15.6 ab^{-1} (against LHCb) and 1 ab^{-1} (against Belle) if improvements in the PID, K_S^0 reconstruction, and MVA classifier efficiencies are made. The only non-reducible systematic uncertainty for R_K is the one related to lepton identification, which at the target luminosity is still smaller than the one introduced by the MVA classifiers. Focusing in increasing the significance of the signal components facilitated the fitting procedure and reduced the systematic uncertainties related to it; these are already smaller than at Belle for a luminosity of 600 fb^{-1} .

If existent, new physics in the $b \rightarrow sl^+\ell^-$ systems related to Lepton Flavor Non-Universality will only be observed with a 5σ precision at a luminosity of 35 ab^{-1} , provided that improvements in the reconstruction efficiency and background rejection take place. It is thus of utmost importance that future analysis strive for

a more sophisticated K_S^0 reconstruction, for a more powerful PID process, and for more robust MVA classifiers.

References

- [1] The CMS Collaboration, “Observation of a new boson at a mass of 125 gev with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012.
- [2] The ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.
- [3] L. Landau and E. Lifshitz, *Mechanics*. No. v. 1, Elsevier Science, 1982.
- [4] E. Wigner, “On Unitary Representations of the Inhomogeneous Lorentz Group,” *Annals of Mathematics*, vol. 40, no. 1, pp. 149–204, 1939.
- [5] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*. Physics textbook, New York, NY: Wiley, 2008.
- [6] J. Abelleira *et al.*, “High-Energy LHC design,” *Journal of Physics: Conference Series*, vol. 1067, p. 022009, sep 2018.
- [7] M. Misiak *et al.*, “Updated Next-to-Next-to-Leading-Order QCD predictions for the weak radiative B -meson decays,” *Physical Review Letters*, vol. 114, Jun 2015.
- [8] Y. S. Amhis *et al.*, “Averages of b -hadron, c -hadron, and τ -lepton properties as of 2018,” 2019. Updated results and plots available at <https://hflav.web.cern.ch/>.
- [9] T. Blake, G. Lanfranchi, and D. M. Straub, “Rare B decays as tests of the Standard Model,” *Progress in Particle and Nuclear Physics*, vol. 92, p. 50–91, Jan 2017.
- [10] R. Aaij *et al.*, “Measurement of the $B_s^0 \rightarrow \mu^+$ Branching Fraction and Effective Lifetime and Search for $B^0 \rightarrow \mu^+$ Decays,” *Physical Review Letters*, vol. 118, May 2017.
- [11] M. Aaboud *et al.*, “Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector,” *Journal of High Energy Physics*, vol. 2019, Apr 2019.
- [12] The BaBar Collaboration, “Search for $B^+ \rightarrow K^+ \tau^+ \tau^-$ at the BaBar experiment,” 2016.

- [13] C. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, "Standard Model predictions for $B \rightarrow K\ell\ell$ with form factors from lattice QCD," 2013.
- [14] T. Huber, T. Hurth, and E. Lunghi, "Inclusive $B \rightarrow X_s\ell^+\ell^-$: Complete angular analysis and a thorough study of collinear photons," 2015.
- [15] The Particle Data Group, "Review of Particle Physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018.
- [16] A. Ghinculov, T. Hurth, G. Isidori, and Y.-P. Yao, "The rare decay $B \rightarrow X_s\ell^+\ell^-$ to NNLL precision for arbitrary dilepton invariant mass," *Nuclear Physics B*, vol. 685, p. 351–392, May 2004.
- [17] A. Ali, E. Lunghi, C. Greub, and G. Hiller, "Improved model-independent analysis of semileptonic and radiative rare B decays," *Physical Review D*, vol. 66, Aug 2002.
- [18] J. Lees *et al.*, "Measurement of the $B \rightarrow X_s\ell^+\ell^-$ Branching Fraction and Search for Direct CP-Violation from a Sum of Exclusive Final States," *Physical Review Letters*, vol. 112, May 2014.
- [19] M. Iwasaki *et al.*, "Improved measurement of the electroweak penguin process $B \rightarrow X_s\ell^+\ell^-$," *Physical Review D*, vol. 72, Nov 2005.
- [20] Y. Sato and others, "Measurement of the Lepton Forward-Backward Asymmetry in $B \rightarrow X_s\ell^+\ell^-$ Decays with a Sum of Exclusive Modes," 2014.
- [21] W. Altmannshofer, P. Ball, A. Bharucha, A. J. Buras, D. M. Straub, and M. Wick, "Symmetries and asymmetries of $B \rightarrow K^*\mu^+\mu^-$ decays in the Standard Model and beyond," *Journal of High Energy Physics*, vol. 2009, pp. 019–019, Jan 2009.
- [22] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, "Optimizing the basis of $B \rightarrow K^*\ell^+\ell^-$ observables in the full kinematic range," *Journal of High Energy Physics*, vol. 2013, May 2013.
- [23] The Belle Collaboration, "Angular analysis of $B^0 \rightarrow K^*(892)^0\ell^+\ell^-$," 2016.
- [24] R. Aaij *et al.*, "Angular analysis of the $B^0 \rightarrow K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity," *Journal of High Energy Physics*, vol. 2016, Feb 2016.
- [25] The ATLAS collaboration, "Angular analysis of $B_d^0 \rightarrow K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8\text{ TeV}$ with the ATLAS detector," 2017.
- [26] C. Bobeth, G. Hiller, and G. Piranishvili, "Angular distributions of $\bar{B} \rightarrow \bar{K}\ell^+\ell^-$ decays," *JHEP*, vol. 12, p. 040, 2007.
- [27] N. Gubernari, A. Kokulu, and D. van Dyk, " $B \rightarrow P$ and $B \rightarrow V$ form factors from B-meson light-cone sum rules beyond leading twist," *Journal of High Energy Physics*, vol. 2019, Jan 2019.

- [28] T. Huber, E. Lunghi, M. Misiak, and D. Wyler, “Electromagnetic logarithms in $B \rightarrow X_s \ell^+ \ell^-$,” *Nuclear Physics B*, vol. 740, p. 105–137, Apr 2006.
- [29] R. Aaij *et al.*, “Search for lepton-universality violation in $B^+ \rightarrow K^+ \ell^+ \ell^-$ decays,” *Phys. Rev. Lett.*, vol. 122, no. 19, p. 191801, 2019.
- [30] “Test of lepton flavor universality in $B \rightarrow K \ell^+ \ell^-$ decays.”
- [31] J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl, and D. M. Straub, “B-decay discrepancies after Moriond 2019,” 2019.
- [32] K. Akai, K. Furukawa, and H. Koiso, “SuperKEKB Collider,” *Nucl. Instrum. Meth.*, vol. A907, pp. 188–199, 2018.
- [33] S. Sandilya, “Prospects for rare decays at belle ii,” Dec 2016.
- [34] C. L. L. (conference presentation on behalf the Belle II Collaboration), “Belle ii status and physics prospects,” Jul 2019.
- [35] W. Altmannshofer *et al.*, “The Belle II Physics Book,” 2018.
- [36] K. G. Wilson, “Non-Lagrangian Models of Current Algebra,” *Phys. Rev.*, vol. 179, pp. 1499–1512, Mar 1969.
- [37] K. G. Wilson and W. Zimmermann, “Operator product expansions and composite field operators in the general framework of quantum field theory,” *Communications in Mathematical Physics*, vol. 24, pp. 87–106, Jun 1972.
- [38] C. Bobeth, G. Hiller, and G. Piranishvili, “Angular distributions of $\bar{B} \rightarrow K \bar{\ell} \ell$ decays,” *Journal of High Energy Physics*, vol. 2007, p. 040–040, Dec 2007.
- [39] A. J. Buras, “Weak Hamiltonian, CP violation and rare decays,” in *Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2*, pp. 281–539, 1998.
- [40] A. J. Buras and M. Münz, “Effective Hamiltonian for $B_s e^+ e^-$ beyond leading logarithms in the naive dimensional regularization and ’t Hooft–Veltman schemes,” *Physical Review D*, vol. 52, p. 186–195, Jul 1995.
- [41] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, “A comparative study of the decays $B \rightarrow (K, K^*) \ell^+ \ell^-$ in the standard model and supersymmetric theories,” *Physical Review D*, vol. 61, Mar 2000.
- [42] J. A. Bailey *et al.*, “ $B \rightarrow K \ell^+ \ell^-$ decay form factors from three-flavor lattice QCD,” *Physical Review D*, vol. 93, Jan 2016.
- [43] D. M. Straub, “flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond,” 2018.

- [44] M. Beneke and T. Feldmann, "Symmetry-breaking corrections to heavy-to-light B meson form factors at large recoil," *Nuclear Physics B*, vol. 592, p. 3–34, Jan 2001.
- [45] T. Abe *et al.*, "Belle II Technical Design Report," 2010.
- [46] M. Baszczyk *et al.*, "SuperB Technical Design Report," 2013.
- [47] I. Adachi, T. E. Browder, P. Križan, S. Tanaka, and Y. Ushiroda, "Detectors for extreme luminosity: Belle II," *Nucl. Instrum. Meth.*, vol. A907, pp. 46–59, 2018.
- [48] K. R. Nakamura *et al.*, "The Belle II SVD detector," *PoS*, vol. Vertex2016, p. 012, 2017.
- [49] S. Furlotov, "The Belle II pixel vertex detector," *Journal of Instrumentation*, vol. 7, p. C01014, 01 2012.
- [50] N. Taniguchi, "Central Drift Chamber for Belle-II," *Journal of Instrumentation*, vol. 12, pp. C06014–C06014, jun 2017.
- [51] U. Tamponi, "The TOP counter of Belle II: status and first results," in *10th Workshop on Ring Imaging Cherenkov Detectors (RICH2018) Moscow, Russia, July 29-August 4, 2018*, vol. A, 2019.
- [52] H. Kindo *et al.*, "Initial performance of the Aerogel RICH detector of the Belle II experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2019.
- [53] V. Aulchenko *et al.*, "Electromagnetic calorimeter for Belle II," *Journal of Physics: Conference Series*, vol. 587, p. 012045, feb 2015.
- [54] T. Abe *et al.*, "Belle II Technical Design Report," 2010.
- [55] T. Aushev *et al.*, "A scintillator based endcap KL and muon detector for the Belle II experiment," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 789, p. 134–142, Jul 2015.
- [56] T. Ferber and P. Urquijo, "Overview of the Belle II Physics Generators (Belle II internal note)." URL: <https://docs.belle2.org/record/282/files/BELLE2-NOTE-PH-2015-006-v2.pdf>, 01 2016.
- [57] D. Sun, Z. Liua, J. Zhao, and H. Xu, "Belle2Link: A Global Data Readout and Transmission for Belle II Experiment at KEK," *Physics Procedia*, vol. 37, pp. 1933–1939, 2012. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).
- [58] S. Yamada *et al.*, "Data Acquisition System for the Belle II Experiment," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 3, pp. 1175–1180, 2015.
- [59] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun, "'the belle ii core software'," *Computing and Software for Big Science*, vol. 3, Nov 2018.

- [60] D. J. Lange, "The EvtGen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1, pp. 152 – 155, 2001. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.
- [61] S. Agostinelli *et al.*, "Geant4– a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.
- [62] R. Brun and F. Rademakers, "Root– an object oriented data analysis framework," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 389, no. 1, pp. 81 – 86, 1997. New Computing Techniques in Physics Research V.
- [63] P. Ecker, "Study of Bremsstrahlung Finding at the Belle II Experiment," 2018.
- [64] W. D. Hulsbergen, "Decay chain fitting with a Kalman filter," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 552, p. 566–575, Nov 2005.
- [65] Tenchini, Francesco and Krohn, Jo-Frederik, "Decay Chain Reconstruction at Belle II," *EPJ Web Conf.*, vol. 214, p. 06023, 2019.
- [66] A. J. Bevan *et al.*, "The Physics of the B Factories," *Eur. Phys. J.*, vol. C74, p. 3026, 2014.
- [67] "Event shapes in e^+e^- annihilation,"
- [68] Søren Halkjær and Winther, Ole, "The Effect of Correlated Input Data on the Dynamics of Learning," in *Advances in Neural Information Processing Systems 9* (M. C. Mozer and M. I. Jordan and T. Petsche, ed.), pp. 169–175, MIT Press, 1997.
- [69] T. Keck, "FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification," *CoRR*, vol. abs/1609.06119, 2016.
- [70] M. Feindt and M. Prim, "An algorithm for quantifying dependence in multivariate data sets," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 698, pp. 84 – 89, 2013.
- [71] A. Glazov, P. Rados, A. Rostomyan, E. Paoloni, and L. Zani, "Measurement of the track reconstruction efficiency in Phase 3 data using $e^+e^- \rightarrow \tau^+\tau^-$ events," 2020.
- [72] Daniel Ferlewicz and Marco Milesi and Phillip Urquijo, "Lepton identification in Proc10 and MC13a using $J/\psi \rightarrow \ell^+\ell^-$ and $K_S^0 \rightarrow \pi^+\pi^-$," Apr 2020.
- [73] S. Sandilya and A. Schwartz, "Study of Kaon and Pion Identification Performance in Phase III data using the decay $D^{*+} \rightarrow D^0 [K^- \pi^+]$," 05 2020.

- [74] H. Albrecht *et al.*, "Search for hadronic $b \rightarrow u$ decays," *Physics Letters B*, vol. 241, no. 2, pp. 278 – 282, 1990.
- [75] G. D'Agostini, *Bayesian reasoning in high-energy physics: principles and applications*. CERN Yellow Reports: Monographs, Geneva: CERN, 1999.
- [76] A. Ishikawa *et al.*, "Observation of $B \rightarrow K^* \ell^+ \ell^-$," *Phys. Rev. Lett.*, vol. 91, p. 261601, 2003.
- [77] Ishikawa, Akimasa, "Observation of the Electroweak Penguin Decay $B \rightarrow K^* \ell^+ \ell^-$," Feb 2001.
- [78] I. Adachi *et al.*, "sbelle design study report," 2008.
- [79] M. Srednicki, *Quantum Field Theory*. Cambridge: Cambridge Univ. Press, 2007.
- [80] S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 2005.
- [81] W.-K. Tung, *Group Theory in Physics*. 57 Shelton Street, Covent Garden, London: World Scientific Publishing Co., 1999.
- [82] J. Schwichtenberg, *Physics from Symmetry*. Springer: Springer, 2015.
- [83] E. Noether, "Invariante Variationsprobleme," *Transport Theory and Statistical Physics*, vol. 1, no. 3, pp. 186–207, 1971.
- [84] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental Test of Parity Conservation in Beta Decay," *Phys. Rev.*, vol. 105, pp. 1413–1415, Feb 1957.
- [85] R. P. Feynman, "Space-Time Approach to Non-Relativistic Quantum Mechanics," *Rev. Mod. Phys.*, vol. 20, pp. 367–387, Apr 1948.
- [86] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley, 1995.
- [87] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak Interactions with Lepton-Hadron Symmetry," *Phys. Rev. D*, vol. 2, pp. 1285–1292, Oct 1970.
- [88] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction," *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973.
- [89] A. D. Sakharov, "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe," *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967. [*Usp. Fiz. Nauk*161,no.5,61(1991)].
- [90] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa Matrix," *Phys. Rev. Lett.*, vol. 51, pp. 1945–1947, Nov 1983.
- [91] J. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, 11 2000.

-
- [92] A. Rogozhnikov, A. Bukva, V. Gligorov, A. Ustyuzhanin, and M. Williams, "New approaches for boosting to uniformity," *Journal of Instrumentation*, vol. 10, p. T03002–T03002, Mar 2015.

Appendix A

A Speed Course in the Standard Model

A.1 Particles

The Standard Model is composed of elementary particle *fields* and the interaction terms between them. The starting point for its derivation is the merging between the basic axioms of quantum mechanics: that particles can be considered as excited states, or quanta, of their underlying fields, which are functions of the space and time; and those of special relativity: **1.** That the speed of light has the same value c in all inertial frames of reference, and **2.** That physics should be the same in any of these frames. In addition, it is possible to introduce two more empirical conditions: that laws of physics are homogeneous and isotropic, this is, they do not depend on the orientation or the position of the observer.

From this, it is possible to derive a mathematical representation of the fields existing in a universe governed by these axioms, as first proven by Wigner [4]. The first starts by recalling that inertial frames are related to one another by Poincarè transformations, which are defined as the group of transformations that leave the line element

$$(dx)^2 = g_{\mu\nu} dx^\mu dx^\nu = (cdt)^2 - (d\mathbf{x})^2 \quad (\text{A.1})$$

invariant. Here $g_{\mu\nu} = \text{diag}(1, -1, -1, -1)$ is the Minkowski metric, and we have written an infinitesimal displacement vector in space-time as $dx = (cdt, dx^1, dx^2, dx^3)$. Notice that, from equation A.1, an object moving at speed c will have a line element $(dx)^2 = 0$ in any frame related by a Poincarè transformation, and hence the same

speed c in any of these frames: the first assumption of special relativity is then fulfilled.

The second assumption can be translated as requiring for the equations of motion of the theory to be covariant under Poincarè transformations. In quantum mechanics, under the Lagrangian formalism, the equations of motion of a system described by a collection of *fields* $\phi_i(x)$ can be obtained by finding the extrema of a functional \mathcal{S} , called the action, defined as

$$\mathcal{S} = \int d^4x \mathcal{L} [\phi_1(x), \dots, \phi_N(x), \partial_\mu \phi_1(x), \dots, \partial_\mu \phi_N(x)] \quad (\text{A.2})$$

where \mathcal{L} , the Lagrangian, is a functional of the fields. The equations of motion are

$$\frac{\partial \mathcal{L}}{\partial \phi_i} - \partial_\mu \left[\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi_i)} \right] = 0. \quad (\text{A.3})$$

These should be covariant under Poincarè transformations, which implies that the fields $\phi_i(x)$ must transform under representations of the Poincarè group. This group is composed of Lorentz transformations (rotations in space and velocity boosts) together with translations by a constant vector $a = (a^0, \mathbf{a})$. As the Lorentz group has four disconnected components —this is, one cannot go from an element in one component to another in a different component by a continuous curve lying within the group. See, for example, [79], chapter 2—, so does the Poincarè group. The component —or subgroup— which contains the identity element (a transformation which does not change anything), known as the inhomogeneous proper orthochronous Lorentz group $ISO(3,1)^\uparrow$, corresponds to the fundamental symmetry group of the Standard Model, and from now on any mention of the Poincarè group will refer to this subgroup unless stated otherwise¹. It contains four generators for translations P^μ and six² generators for the Lorentz transformations $M^{\mu\nu}$: 3 for boosts in any of the three spatial dimensions ($K_i = M^{i0}$), and the remaining 3 for rotations ($J_i = \varepsilon_{ijk} M^{jk}/2$, with ε_{ijk} the totally antisymmetric tensor).

¹The full Poincarè group $IO(1,3)$ can be seen as the set

$$IO(3) = \{ISO(1,3), P \times ISO(1,3), T \times ISO(1,3), PT \times ISO(1,3)\},$$

where P and T correspond to the space —or parity— and time reversal operators.

²Though, in principle, the index notation may suggest there are 16 generators (μ and ν run from 0 to 3), the M operators are antisymmetric (see [79], chapter 1); this reduces the number from 16 to 6.

Of special interest are the irreducible representations of the Poincarè group, since they can be labeled by the eigenvalues of their Cassimir operators³, and are then perfect candidates for describing physical states with clear observables, which correspond to one-particle states [80]. In other words, they correspond to the elementary particles of the theory.

The Poincarè group has two Cassimir operators:

$$C_1 = P^\mu P_\mu, \quad (\text{A.4})$$

$$C_2 = W^\mu W_\mu, \text{ with } W_\mu = \frac{1}{2} \varepsilon_{\mu\nu\rho\sigma} P^\nu M^{\rho\sigma} \quad (\text{A.5})$$

and $\varepsilon_{\mu\nu\rho\sigma}$ the totally antisymmetric tensor in four dimensions. W_μ is known as the Pauli-Lubansky vector. Notice that the scalar values of the first Cassimir operator correspond to $p^2 = m^2$, which is to be interpreted as a squared mass; the second Cassimir operator requires a special treatment, for whether $p^2 > 0$, $p^2 = 0$ or $p^2 < 0$ (the last case is, however, disregarded as unphysical, and hence not contemplated by the Standard Model theory). For a detailed treatment the reader can refer to chapter 10 of [81], or chapter 2 of [80]. The result, in any case, is that, for massive states, the scalar value W^2 is related to the internal spin of the state, whereas for massless ones, it is related to their helicity, which is the projection of the spin along the direction of motion.

The concept of particles with defined parity emerges from the fact that the cover of the Lorentz Group is $SU(2) \times SU(2)$. This simply means that the Lorentz group contains two copies of the $SU(2)$ algebra, given by the two set of operators (see, for example, [79], chapter 33):

$$SU(2)_L = \{A_1, A_2, A_3\} \text{ with } A_i = \frac{1}{2} (J_i + iK_i), \quad (\text{A.6})$$

$$SU(2)_R = \{B_1, B_2, B_3\} \text{ with } B_i = \frac{1}{2} (J_i - iK_i). \quad (\text{A.7})$$

Notice that $J_i = A_i + B_i$, so the values for the spin j are determined by the eigenvalues of the two operators A^2 and B^2 ⁴. Indeed, $j = |a - b|, |a - b| + 1, \dots, a + b$. Each representation can be labeled as (a, b) , and carries spin $j \in (|a - b|, |a - b| + 1, \dots, a + b)$

³The Cassimir operators are defined as operators which commute with all the generators of their group. Hence, they commute with all elements in the group and their eigenvalues remain invariant under the group transformations.

⁴Since the A_i and B_i operators conform an $SU(2)$ algebra, it is possible to find eigenstates of their Cassimir operators, A^2 and B^2 . In analogy with the spin treatment usually done in quantum mechanics for the J_i operators, one finds that the eigenvalues of A^2 and B^2 operators are of the form

and mass m . Particles with the same values of j and m can then belong to different representations. The Standard Model is composed by the representations shown in table A.1. Fermions correspond to particles with spin $1/2$, which can belong to

Table A.1. Representations of the Lorentz group included in the Standard Model.

Particle	Representation	Spin	In the Standard Model
Spinor Fermions	$\left(\frac{1}{2}, 0\right) \oplus \left(0, \frac{1}{2}\right)$	$\frac{1}{2}$	Quarks and Leptons
Vector Bosons	$\left(\frac{1}{2}, \frac{1}{2}\right)$	1	Photon, Gluon, Z^0 , W^\pm
Scalar Bosons	$(0, 0)$	0	Higgs

either the $\left(\frac{1}{2}, 0\right)$ or the $\left(0, \frac{1}{2}\right)$ representations of the group. The first are termed left handed fermions, while the second correspond to right handed ones.

Applying the parity operator P on A_i transforms it into B_i and vice versa⁵. Because of this, any theory that respects parity should include both representations at the same time; this is the reason why the Spinor Fermion representation is the direct sum of $\left(\frac{1}{2}, 0\right)$ and $\left(0, \frac{1}{2}\right)$, instead of a copy of one of these. Electromagnetic, strong and gravitational forces are invariant under P , so states with the same mass and spin belonging to the left or right representation are indistinguishable, and hence assigned to the same particle. This is not the case, however, for weak interactions, as shall be shown next.

A.2 Interactions

So far, the only difference between particles belonging to the same representation is their mass. The second ingredient for the Standard Model is the introduction of interactions between the different fields. The derivation of the free (non-interacting) Lagrangians for different representations in the Standard Model is done, for example, in chapter 6 of [82]. They are listed in table A.2.

$a(a+1)$ and $b(b+1)$, where a and b are integers or semi-integers. At the same time, the maximum allowed value for $a_3 = a$ and for $b_3 = b$.

⁵Intuitively, \mathbf{J} oughts to remain unchanged under parity transformations, as it is an axial vector. On the other side, a boost in a certain direction would point in the opposite direction after an inversion of the spatial axes. It is also possible to convince one self by performing the operations on the fundamental representation of the group algebra, as done in the chapter 3 of [82]

⁶Recall that γ^μ correspond to the Gamma matrices and $\bar{\psi} = \psi^\dagger \gamma^0$

Table A.2. Free Field Lagrangians in the Standard Model

Particle	Representation	Lagrangian name	General Form
Fermions	$\left(\frac{1}{2}, 0\right) \oplus \left(0, \frac{1}{2}\right)$	Dirac ⁶	$\mathcal{L} = \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi$
Vector Bosons	$\left(\frac{1}{2}, \frac{1}{2}\right)$	Proca ⁷	$\mathcal{L} = -\frac{1}{2} G^{\mu\nu} ^2 + m^2 A^\mu ^2$
Scalar Bosons	$(0, 0)$	Klein-Gordon	$\mathcal{L} = \frac{1}{2} \left(\partial_\mu \phi ^2 - m^2 \phi ^2 \right)$

Notice that the Dirac Lagrangian is invariant under $U(1)$ transformations (this is, under multiplication by a complex scalar $e^{i\theta}$, where θ is real). As first stated by Noether [83], this symmetry implies the existence of a conserved quantity: the electric charge. The construction of the interactive theory⁸ begins with the postulate that the Dirac Lagrangian should also be invariant under local $U(1)$ transformations⁹, leading to the quantization of the electromagnetic field¹⁰. Indeed, this assertion requires the introduction of a massless vector field, or *gauge* field, B_μ , which transforms like $B_\mu \rightarrow B_\mu - \frac{1}{q} \partial_\mu \theta(x)$ and that couples with the spinor field ψ with a term proportional to g —a coupling constant, dictating the strength of the interaction—, and has an associated tensor field $G^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu$; the coupling with the field has opposite signs for the spinors ψ and $\psi^C = C\psi^*$, where C is an orthogonal matrix such that $C\gamma^\mu C^T = -\gamma^{\mu*}$, and the asterisk denotes complex conjugation. Hence, the spinor ψ^C is identified as the antiparticle of ψ .

The question arises for whether there are other internal symmetries in the Lagrangian. As it turns out, a Lagrangian of two spinor fields, written in terms of the doublet $\Psi = (\psi_1, \psi_2)^T$

$$\mathcal{L}_\Psi = i\bar{\Psi} \gamma_\mu \partial^\mu \Psi \quad (\text{A.8})$$

is invariant under $SU(2)$ global transformations as long as the two fields are massless. As with the $U(1)$ case, extending this symmetry to local $SU(2)$ transformations — which correspond to a multiplication by a 2×2 matrix of the form $e^{i\theta_i(x)\sigma^i}$, with σ_i

⁷With $G^{\mu\nu}$ a tensor field which depends on the nature of the interaction mediated by the boson. See the discussion in the text.

⁸The interested reader can find in the chapter 7 of [82] the detailed procedure of what is about to be sketched in what follows.

⁹This is, $U(1)$ transformations where the transformation parameter θ is now a real function of the time-space coordinates

¹⁰To be more rigorous, as will be shown next, the electromagnetic field is a superposition of this original $U(1)$ gauge field with one of $SU(2)$ gauge fields

being the Pauli matrices, the generators of the group— requires the introduction of three (one for each generator) massless vector fields, W_μ^0 , W_μ^1 and W_μ^2 , which couple to this doublet¹¹. These vector fields transform as

$$W_\mu^i \rightarrow W_\mu^i - \frac{1}{g} \partial_\mu \theta_i(x) - \varepsilon_{ijk} \theta_j(x) W_\mu^k$$

(where we have assumed summation over repeated latin indexes), and each generates a tensor field $W_i^{\mu\nu}$ of the form

$$W_i^{\mu\nu} = \partial^\mu W_i^\nu - \partial^\nu W_i^\mu + \varepsilon_{ijk} W_j^\mu W_k^\nu. \quad (\text{A.9})$$

This symmetry is used as the base for the model of weak interactions and the components of the doublets are suggested by experimental results: left handed electrons, muons and taons are always transformed by weak interactions into the same particle with different momentum, or into their respective neutrinos¹². On the other side, positively charged mass eigenstates of quarks (up, charm, top) transform into a superposition of negatively charged mass eigenstates (down, strange, bottom). The matrix V_{CKM} describing this superposition is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix, discussed in appendix B. The six $SU(2)_L$ doublets are:

$$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}, \begin{pmatrix} u_L \\ d'_L \end{pmatrix}, \begin{pmatrix} c_L \\ s'_L \end{pmatrix}, \begin{pmatrix} t_L \\ b'_L \end{pmatrix}, \quad (\text{A.10})$$

with

$$\begin{pmatrix} d'_L \\ s'_L \\ b'_L \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}. \quad (\text{A.11})$$

¹¹Though electromagnetic and strong interactions respect parity, weak interactions do not, as first observed by Wu [84]: only particles belonging to the left representation of the Lorentz group participate in them. This implies that, under $SU(2)$ transformations, particles belonging to the $\left(\frac{1}{2}, 0\right)$

transform as doublets—as expressed before—but particles belonging to the $\left(0, \frac{1}{2}\right)$ representation behave as singlets. Since a spinor ψ contains both a left handed and a right handed component, the interaction term between the $SU(2)$ gauge bosons W_i and the doublet Ψ contains a term which selects, or *projects*, the left-handed component of the two spinors in Ψ .

¹²With the discovery of neutrino oscillations, this may not be the case anymore. Instead, as with quarks, the neutrinos of the weak interaction become a superposition of the three mass-eigenstates ν_i , and hence an electron-neutrino can transform into a muon-neutrino, since these states are not longer orthogonal.

From Noether's Theorem, $SU(2)$ symmetry implies the conservation of three quantities (one for each generator):

$$Q_i = \bar{\Psi} \gamma^0 \sigma_i \Psi. \quad (\text{A.12})$$

However, only one of the generators can be diagonalized at the same time—or, in other words, $SU(2)$ has only one Cartan generator—which implies that there is only one charge that can be assigned to each component of the doublet. This charge receives the name of weak isospin, and its eigenvalues are $\frac{1}{2}$ for the first component, and $-\frac{1}{2}$ for the second one.

The bosons of these interactions have zero mass, whereas experiments showed that the mediators of the weak force, the W^\pm and the Z^0 , are massive. Moreover, a fermion mass term also spoils the invariance, so this model predicted a theory in which all particles are massless. The solution to this dilemma was given by the introduction of a scalar field doublet $\Phi = (\phi_1, \phi_2)^T$, with a Lagrangian of the form

$$\mathcal{L}_\Phi = D_\mu \Phi^\dagger D^\mu \Phi - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (\text{A.13})$$

where D_μ is the covariant derivative of the fields ϕ_i , given by $D_\mu = \partial_\mu - i\frac{1}{g}\sigma_i W_\mu^i - i\frac{1}{q}B_\mu$. The last two terms in A.13 correspond to what is known as the Higgs potential $V(\Phi) = \langle \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \rangle$, with $\lambda > 0$ ¹³; the beauty of this equation resides in the assumption that the value of μ^2 depends on the temperature, so the minimum of the potential shifts as the temperature decreases. In the early universe, $\mu^2 > 0$ and the minimum energy configuration was achieved by having $\langle \Phi \rangle = 0$. However, after the temperature decreased down to a critical value, a phase transition took place, in which μ^2 flipped sign. The shape of the potential $V(\Phi)$ changed, and a continuum of minima appeared¹⁴:

$$\langle \Phi_{\min} \rangle = e^{i\phi} \begin{pmatrix} 0 \\ \frac{|\mu|}{\sqrt{2\lambda}} \end{pmatrix}. \quad (\text{A.14})$$

¹³As the λ term describes quartic self interactions among these scalar fields, vacuum stability demands for λ to be greater than zero

¹⁴Recall that the components of the $SU(2)$ doublets differ by a unit in their electrical charge. This is also the case for Φ , $q_{\phi_1} - q_{\phi_2} = 1$. Since so far experimental observations agree with the conservation of electrical charge, ϕ_2 is assigned a charge of zero, and is allowed to acquire a non-zero vacuum expectation value. ϕ_1 has then charge +1 and so its vacuum expectation value is made zero. Because of this, $U(1)$ symmetry is preserved, electrical charge conserved, and the photon is still a massless boson.

Expanding the doublet around one of the multiple minima gives rise to a physical field h with a zero vacuum-expectation-value, known as the Higgs field:

$$\Phi = \langle \Phi_{\min} \rangle + \begin{pmatrix} 0 \\ h \end{pmatrix}. \quad (\text{A.15})$$

Interactions between this scalar doublet and the massless boson fields — B^μ and W_i^μ — give rise to the photon field A^μ , the W^\pm boson fields, and the Z^0 boson field as superpositions of the original ones:

$$W^\pm = \frac{1}{\sqrt{2}} (W_1 \mp iW_2), \quad A = \frac{1}{\sqrt{g_e^2 + g_w^2}} (g_e W_3 + g_w B), \quad Z^0 = \frac{1}{\sqrt{g_e^2 + g_w^2}} (g_w W_3 - g_e B) \quad (\text{A.16})$$

where g_e is the coupling constant for the B field, and g_w the coupling constant for the W_i fields.

Moreover, this expansion also generates a mass term for the W^\pm and the Z^0 . The *electromagnetic* and *weak* interactions are then obtained from the symmetry breaking of the $U(1) \times SU(2)$ *electroweak* interaction.

It is possible to follow this line of thought and add an interaction between a left spinor doublet $\Psi^L = \Psi = (\psi_1^L \psi_2^L)^T$, two right spinor singlets ψ_1^R and ψ_2^R , and the scalar doublet Φ , known as Yukawa coupling, of the form

$$\mathcal{L}_{\text{Yukawa}} = -\mu_1 (\bar{\Psi}^L \Phi^C \psi_1^R + \bar{\psi}_1^R \Phi^{C,\dagger} \Psi^L) - \mu_2 (\bar{\Psi}^L \Phi \psi_2^R + \bar{\psi}_2^R \Phi^\dagger \Psi^L) \quad (\text{A.17})$$

where Φ^C is the charge-conjugated field $\Phi^C = (\phi_2^*, -\phi_1^*)^T$. Once symmetry is broken, this expression gives fermions their mass¹⁵.

In the same fashion, a triplet of spinors is invariant under $SU(3)$ local transformations (multiplications by a 3×3 matrix of the form $e^{i\theta_i(x)\mu^i}$ where μ^i are the Gell-Mann matrices, the generators of the group), provided, again, that eight massless vector fields g_i^μ are introduced in the theory. For the $SU(3)$ case, there are eight conserved charges, but each component in the triplet can be labeled by only two of

¹⁵This is, up to the current date, the accepted mechanism for generating the charged fermions masses. Neutrinos, which do not have electrical charge, are able to acquire mass—which current experiments on neutrino oscillations have shown is different from zero—through a different mechanism, reserved for what are known as Majorana particles. These interactions have the form:

$$\mathcal{L}_{\text{Majorana}} = \frac{(\Phi^{C,\dagger} \Psi^L)^2}{\mu}.$$

This, however, presents the problem of being a non-renormalizable term.

them, since $SU(3)$ has only two Cartan generators. Moreover, the two generators share the same eigenstates, so the eigenvalues of both generators are grouped as a single charge, known as *color*. The two Cartan operators are μ^3 and μ^8 :

$$\mu^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mu^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (\text{A.18})$$

A triplet with eigenvalues $(1, \frac{1}{\sqrt{3}})$ has a *red* color charge, one with $(-1, \frac{1}{\sqrt{3}})$ has blue charge, and one with $(0, -\frac{2}{\sqrt{3}})$ has green charge. The components of a triplet correspond then to the same particle with different colors. This interaction is known as the *strong* interaction, and only quarks (and anti-quarks) are known to participate in it.

Table A.3, also shown in chapter 1, resumes the Standard Model particles and interactions. Each column in the fermion group gives the $SU(2)$ doublets for both leptons and quarks. The first doublet component for leptons has an electric charge of zero, whereas the second one has a charge of -1 . For quarks doublets, the first component has an electric charge of $2/3$, and the second one of $-1/3$. Quarks carry color charge, and interact through the strong force, whereas leptons do not. Except for the Higgs boson, which is a scalar boson—this is, transforms under the $(0,0)$ representation of the Poincaré group—, all interactions are mediated by vector bosons. For the fermion sector, mass increases while moving down and to the right in the table; for bosons, mass increases as going down.

Table A.3. The Standard Model of Physics

Fermions				Bosons	
				Name	Interaction
Leptons	ν_e	ν_μ	ν_τ	Photon	Electromagnetic
	e	μ	τ	Gluon	Strong
Quarks	u	c	t	Weak	W^\pm, Z^0
	d	s	b	Higgs	Higgs interaction

A.3 Observables, Scattering Matrix and Feynman Diagrams

In order to test the validity of the Standard Model as a mathematical framework able to explain physical events, it is necessary for it to make predictions on quantities that can be measured. In the quantum formalism, these dynamical variables are represented by hermitian operators \mathcal{A} —in order for them to have real eigenvalues—that act on physical states $|s\rangle$. The expectation values of these observables on the given states are represented by the inner product

$$\langle \mathcal{A} \rangle = \langle s | \mathcal{A} | s \rangle. \quad (\text{A.19})$$

In most experiments, the researcher aims to determine how does the state of a system evolves with time under the action of physical interactions. Particularly, in High Energy Physics, one is interested in the probability amplitudes for scattering processes and decays, this is, the probability amplitudes for transitions between states that at $t \rightarrow -\infty$ or $t \rightarrow +\infty$ contain definite numbers of particles of various types. These are known as *in* ($|\alpha\rangle$) and *out* ($|\beta\rangle$) states, where α and β represent the multiple labels used to identify them: momenta of the particles contained, their spin—or helicity, if they are massless—, and species. The collections of amplitudes, given by

$$S_{\beta\alpha} = \langle \beta | \alpha \rangle, \quad (\text{A.20})$$

is known as the S -matrix, and its components are the probability amplitudes for transitions between the states $\alpha \rightarrow \beta$. If there were no interactions S would be proportional to the identity matrix; thus, of primary interest is the matrix T , defined as

$$iT = S - 1. \quad (\text{A.21})$$

The calculation of the components $T_{\alpha\beta}$ is done with the aid of *Feynman Diagrams*, which are graphical representations of the processes to analyze¹⁶: each particle

¹⁶The treatment given here obscures the connection between this and the previous section; the S matrix, however, can be seen as a non-finite representation of the operator $\hat{S} = \hat{\Omega}_+^\dagger \hat{\Omega}_-$, where $\hat{\Omega}_+$ and $\hat{\Omega}_-$ are known as the Möller operators, each of which take a state at time t and projects it into a state at time $t \rightarrow \pm\infty$. As first shown by Feynman in his celebrated paper [85] for the non-relativistic case, under certain conditions (fulfilled by the Standard Model Lagrangian) this product is a functional of the action integral

$$\mathcal{S} = \int d^4x \mathcal{L};$$

is associated with a line, and each interaction with a vertex joining these lines; particles in the *in* and *out* states correspond to external lines, they fulfill the *on-shell* condition $p^2 = m^2$, and contribute to the matrix element calculation with their Poincarè representation; particles associated with internal lines are known as virtual particles, and due to the Heisenberg Uncertainty Principle, are able to be off-shell (this is, they do not obey the relation $p^2 = m^2$); their contribution is in the form *propagators*.

Notice, however, that it is possible to construct diagrams as complicated as desired, by introducing more and more vertexes and internal lines: as long as the external lines correspond to the same particles, all diagrams are contributions to the same transition amplitude. This is remediated by the fact that each interaction is governed by its own coupling constant g ; for electromagnetic and weak interactions, these constants are smaller than one, and hence diagrams with extra vertexes (which imply more interactions) are suppressed by a factor of g^n , where n is the number of vertexes. It is then possible to achieve very good agreements between experiments and theory with a fairly low number of relatively simple diagrams. However, the coupling constant for the strong interaction does not allow for the same treatment, so alternative methods for their analysis have been developed, and are grouped under the name of quantum chromodynamics (QCD).

A.4 The Hamiltonian Approach

In the previous section the S matrix was introduced as a representation of the operator \hat{S} which, under certain conditions, can be written in terms of a functional of the action \mathcal{S} :

$$\hat{S} \propto e^{i \int d^4x \mathcal{L}}.$$

Though convenient for theoretical remarks, this expression is rarely used for calculations. Instead, one relies on what is known as *perturbation theory*, where the main quantity is not the Lagrangian anymore, but the Hamiltonian H , the operator whose spectrum is the possible energies of the system. It is related to the Lagrangian by the transformation

$$H = \sum_i \left[\frac{\partial \mathcal{L}}{\partial (\partial_0 \phi_i)} \partial_0 \phi_i \right] - \mathcal{L} \quad (\text{A.22})$$

for more on this, see chapter 9 of [80], where the Feynman rules are also derived from this formulation.

where $\partial_0 = \partial/\partial t$, and ϕ_i represents a generic field. This Hamiltonian can be separated in free (H_0) and interacting (H_I) parts, which correspond to the free and interacting elements in the Lagrangian, respectively. As shown in [80], it is possible to expand the operator \hat{S} as a *Dyson's series*, of the form

$$\hat{S} = 1 + \sum_{n=1}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} d^4x_1 d^4x_2 \dots d^4x_n T \{H_I(x_1) \dots H_I(x_n)\} \quad (\text{A.23})$$

where T denotes a time ordered product: operators happening at an earlier time should act first, so they go to the right. It can be mathematically written as:

$$T \{H_I(x_1)H_I(x_2)\} = \theta(x_1^0 - x_2^0) H_I(x_1)H_I(x_2) + \theta(x_2^0 - x_1^0) H_I(x_2)H_I(x_1) \quad (\text{A.24})$$

with $\theta(y)$ the Heaviside step function.

As a first approximation, then, it is possible to write

$$\hat{S} = 1 - i \int_{-\infty}^{\infty} d^4x H_I(x) \quad (\text{A.25})$$

The matrix element $T_{\alpha\beta}$ containing the actual interaction term between states α and β is then given, at leading order, by

$$T_{\alpha\beta} = -i \langle \hat{S} - 1 \rangle = - \left\langle \int_{-\infty}^{\infty} d^4x H_I(t) \right\rangle. \quad (\text{A.26})$$

Though not proven here (the reader can consult chapter 4 of [86]), $T_{\alpha\beta}$ can be further divided into a term containing the dynamics of the interaction, and one describing the kinematics. This is,

$$T_{\alpha\beta} = (2\pi)^4 \delta^{(4)}(p_\alpha - p_\beta) \mathcal{M}(\alpha \rightarrow \beta). \quad (\text{A.27})$$

The delta function takes care of energy and momentum conservation, whereas the dynamics of the interaction is described by the *matrix element* \mathcal{M} .

The calculation of the $B \rightarrow K\ell^+\ell^-$ decay rate in chapter 2 is performed by effectively calculating the expectation value of the interaction Hamiltonian between the initial and final states.

Appendix B

The Cabibbo-Kobayashi-Maskawa Matrix

First proposed by Nicola Cabibbo in 1963 when the quark theory was just an hypothesis, the CKM matrix began with a single term, the Cabibbo angle θ_c , which described the weak interaction of down and strange quarks with up quarks. It was postulated that the object coupling with the up quark through the charged weak interaction was a linear combination of the down and strange quarks

$$d' = \cos\theta_c d + \sin\theta_c s. \quad (\text{B.1})$$

However, this led to predictions for the rates of $\Delta S = 1$ processes (this is, processes in which the number of s quarks changed by one unit) which were in clear conflict with experimental observations. The explanation came with the introduction of a new quark, c , which coupled to the linear superposition

$$s' = -\sin\theta_c d + \cos\theta_c s. \quad (\text{B.2})$$

This postulate, which solves the puzzle of why are FCNC highly suppressed in nature, is known as the GIM mechanism [87]; it predicted the charm quark four years before its discovery in 1974.

Notice that it is possible to write

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \quad (\text{B.3})$$

and that the mixing matrix in B.3, which will be denoted by U , corresponds to a rotation in the $d - s$ plane, and is hence unitary. Therefore, the interaction term of the Z^0 boson with quarks, which is a linear combination of the diagonal $SU(2)$ operator σ_3 and the identity matrix, is proportional to a linear combination of

$$\begin{pmatrix} UU^T & 0 \\ 0 & -UU^T \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (\text{B.4})$$

(which corresponds to the $SU(2)$ generator W_3 part of the Z^0) and

$$\begin{pmatrix} UU^T & 0 \\ 0 & UU^T \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (\text{B.5})$$

(which corresponds to the $U(1)$ generator B part of the Z^0), which is flavor diagonal. Because of this, FCNC are not possible at tree level in the Standard Model. Moreover, looking at figure B.1 it is easy to see that processes mediated by loop diagrams with W boson exchanges are also suppressed, in agreement with experimental observations. In 1973, Kobayashi and Maskawa proved that there is

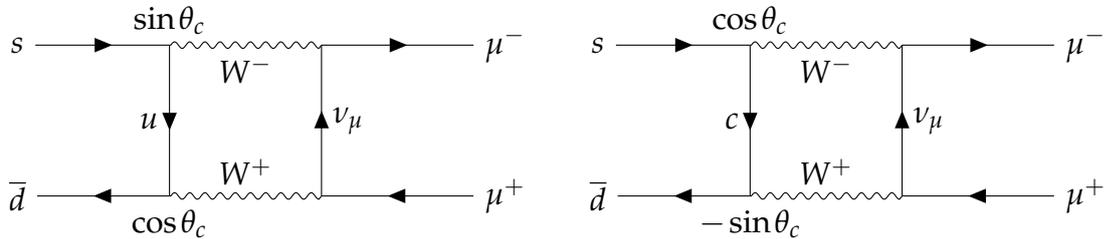


Fig. B.1 Some of the leading order diagrams for the $K \rightarrow \mu^+ \mu^-$ decay. The quark-quark vertices are labeled with the components of the Cabibbo matrix. Due to the sign difference interference of both diagrams is destructive, which leads to a suppression of this process in nature.

no room for CP violating terms in a Lagrangian constructed from only four quarks [88]. CP violation is required, among others, for the explanation of the asymmetry between matter and antimatter in the universe, as explained by Sakharov in 1967 [89]. In order to introduce CP violating terms in the Standard Model, Kobayashi and Maskawa postulated the existence of a new doublet, for then one of the phases in the weak mixing matrix is not longer superfluous (it cannot be absorbed in a redefinition of the quark fields), and this induces CP violation. This mixing matrix

is known as the Cabibbo-Kobayashi-Maskawa matrix:

$$\begin{aligned}
V_{\text{CKM}} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\
&= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}
\end{aligned}$$

(B.6)

where $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$, θ_{12} , θ_{13} and θ_{23} are the 3 Euler angles needed to describe a rotation in 3 dimensions and δ is a complex phase. Notice that the unitarity of the CKM matrix implies

$$\sum_i V_{ij}V_{ik}^* = \delta_{jk} \text{ and } \sum_j V_{ij}V_{kj}^* = \delta_{ik}. \quad (\text{B.7})$$

Experimentally $s_{13} \ll s_{23} \ll s_{12} \ll 1$, and it is convenient to express this hierarchy using the Wolfenstein parameterization [90]. With:

$$s_{12} = \lambda, \quad s_{23} = A\lambda^2, \quad s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta)$$

it is possible to write V_{CKM} up to $\mathcal{O}(\lambda^4)$ terms as

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (\text{B.8})$$

In the Effective Hamiltonian treatment of FCNC, the Wilson coefficients can be written as [39]

$$\mathcal{C}_k \propto \sum_{i=u,c,t} \lambda_i F(x_i) \text{ or } \sum_{i=u,c,t} \lambda_i \lambda_j \tilde{F}(x_i, x_j) \quad (\text{B.9})$$

where F and \tilde{F} are functions of $x_i = m_i^2/m_W^2$, and

$$\lambda_i = \begin{cases} V_{is}^* V_{id} & \text{for } s \rightarrow d \text{ transitions,} \\ V_{ib}^* V_{id} & \text{for } b \rightarrow d \text{ transitions,} \\ V_{ib}^* V_{is} & \text{for } b \rightarrow s \text{ transitions;} \end{cases} \quad (\text{B.10})$$

thus, equation B.7 can be written as

$$\lambda_t + \lambda_c + \lambda_u = 0 \quad (\text{B.11})$$

which implies that $\mathcal{C}_k = 0$ if $x_u = x_c = x_t$. This is an effect, again, of the GIM mechanism. What forbids a complete suppression of these processes and breaks the mechanism is the disparity of the quark masses. Notice, however, that $m_c, m_u \ll m_W$, which implies $x_c \approx x_u \approx 0$. This highlights the importance of B processes, in which the appearance of the top quark in the internal loop effectively removes the GIM suppression, making the B decays particularly useful for studying FCNC transitions, in comparison of D decays, where only d, s and b quarks appear in the internal loops.

Appendix C

How is the Bremsstrahlung Relation Set

The procedure for creating the Bremsstrahlung named relation between a track and a ECL cluster occurs during the offline reconstruction, after clusters, tracks and vertices, have been identified.

It starts with the selection of all tracks in the event for which the electron ID is higher than any other particle ID and with a momentum of less than 5 GeV. For these tracks, hits at the beampipe (at a radius of 1 cm) and at the outer SVD wall (at a radius of 16 cm) are created by extrapolation. Then, the polar (θ_h) and azimuthal (ϕ_h) angles for the extrapolated hits and all the track hits in the VXD region are retrieved, and their errors, $\Delta\theta_h$ and $\Delta\phi_h$, calculated.

Next, ECL clusters which were *not* associated with a track during the reconstruction, and with an energy which is a fraction of the track momentum higher than 2%, are fetched; for each of these clusters, the polar (θ_c) an azimuthal (ϕ_c) angles and their respective errors ($\Delta\theta_c$ and $\Delta\phi_c$), are obtained. Then, the algorithm loops over all the hits obtained previously and checks if $\theta_c - \Delta\theta_c \leq \theta_h \leq \theta_c + \Delta\theta_c$ and $\phi_c - \Delta\phi_c \leq \phi_h \leq \phi_c + \Delta\phi_c$. If this is the case, it then calculates the difference between the directional vectors of the hit and the cluster:

$$s = \left| \hat{\mathbf{d}}_h - \hat{\mathbf{d}}_c \right|, \text{ where } \hat{\mathbf{d}}_i = (\sin\theta_i \cos\phi_i, \sin\theta_i \sin\phi_i, \cos\phi_i) \quad (\text{C.1})$$

and then selects the hit with the smallest s . From this hit, the effective acceptance factor is obtained as

$$\min \left(\frac{|\cos\theta_c - \cos\theta_h|}{\Delta\phi_c + \Delta\phi_h}, \frac{|\sin\theta_c \sin\phi_c - \sin\theta_h \sin\phi_h|}{\Delta\theta_c + \Delta\theta_h} \right) \quad (\text{C.2})$$

Finally, if the hit selected has a transverse momentum which is smaller than 4 GeV, and s is less than 0.05¹, a *Bremsstrahlung* named relation is established between the track and the cluster. The weight of the relation is then given by the effective acceptance factor. The pseudocode for this procedure is shown in algorithm 1.

¹Notice that, for such small value of s , the distance between the two vectors can be interpreted as an arc length. Moreover, since the two vectors are unitary, this arc length is nothing but the angle between the two. Hence, this restriction on s can be interpreted as a veto in the angle between the photon and the electron directions above 50 mrad.

Algorithm 1: Bremsstrahlung recovery algorithm

Data: Tracks and ECL Clusters for the event
Result: Bremsstrahlung relation between some tracks and clusters
best hit \leftarrow None;
smallest distance \leftarrow None;
effective acceptance factor \leftarrow None;
forall *Tracks* **do**
 if *track energy < 5 GeV and track most probable particle = electron* **then**
 extrapolate hits to the beam pipe and outer SVD walls;
 forall *ECL clusters* **do**
 if $0.02 \times \text{track energy} < \text{cluster energy} < \text{track energy}$ and cluster has
 not relation **from** any track in *Tracks* **then**
 get $\theta_c, \phi_c, \Delta\theta_c$ and $\Delta\phi_c$;
 forall *VXD hits and extrapolated hits* **do**
 get $\theta_h, \phi_h, \Delta\theta_h$ and $\Delta\phi_h$;
 if $\theta_c - \Delta\theta_c \leq \theta_h \leq \theta_c + \Delta\theta_c$ and $\phi_c - \Delta\phi_c \leq \phi_h \leq \phi_c + \Delta\phi_c$
 then
 get acceptance factor for the hit (equation C.2);
 get s (equation C.1);
 if *best hit = None* **then**
 best hit \leftarrow hit;
 smallest distance \leftarrow s ;
 effective acceptance factor \leftarrow hit acceptance factor;
 end
 else if $s < \text{smallest distance}$ **then**
 best hit \leftarrow hit;
 smallest distance \leftarrow s ;
 effective acceptance factor \leftarrow hit acceptance factor;
 end
 end
 end
 if *best hit \neq None and best hit transverse momentum < 4 GeV and
 smallest distance < 0.05* **then**
 set Bremsstrahlung cluster relation **to** track;
 relation weight \leftarrow effective acceptance factor;
 end
 end
 end
 end
end

Appendix D

Figure of Merit Plots for the Charmonium Veto Regions

This appendix presents the resulting values for the figure of merit

$$\text{FOM} = \frac{N_S}{\sqrt{N_S + N_B}} \quad (\text{D.1})$$

for multiple candidates of the charmonium veto regions. N_S is the number of signal events which survive the veto, and N_B the number of background events. These are obtained from a 2 ab^{-1} MC sample of $B\bar{B}+q\bar{q}$ events after the initial reconstruction procedure and the Dalitz/conversion veto. The calculation of the FOM for the $\psi(2S)$ resonances is done after the veto for the J/ψ resonance is applied.

From the plot, the regions that maximize the FOM value are:

$$\begin{array}{l|l} B^0 \rightarrow K_s^0 \mu^+ \mu^- & -0.058 \text{ GeV} < q - m(J/\psi) < 0.034 \text{ GeV} \\ B^+ \rightarrow K^+ \mu^+ \mu^- & -0.120 \text{ GeV} < q - m(J/\psi) < 0.044 \text{ GeV} \\ B^0 \rightarrow K_s^0 e^+ e^- & -0.165 \text{ GeV} < q - m(J/\psi) < 0.055 \text{ GeV} \\ B^+ \rightarrow K^+ e^+ e^- & -0.120 \text{ GeV} < q - m(J/\psi) < 0.055 \text{ GeV} \end{array}$$

We choose to apply a single veto on a group decays with the same lepton flavor, and select it by joining the best regions for each decay in the group. The final veto regions are then

$$\begin{array}{l|l} B \rightarrow K \mu^+ \mu^- & -0.12 \text{ GeV} < q - m(J/\psi) < 0.06 \text{ GeV} \\ B^+ \rightarrow K^+ e^+ e^- & -0.17 \text{ GeV} < q - m(J/\psi) < 0.07 \text{ GeV} \end{array}$$

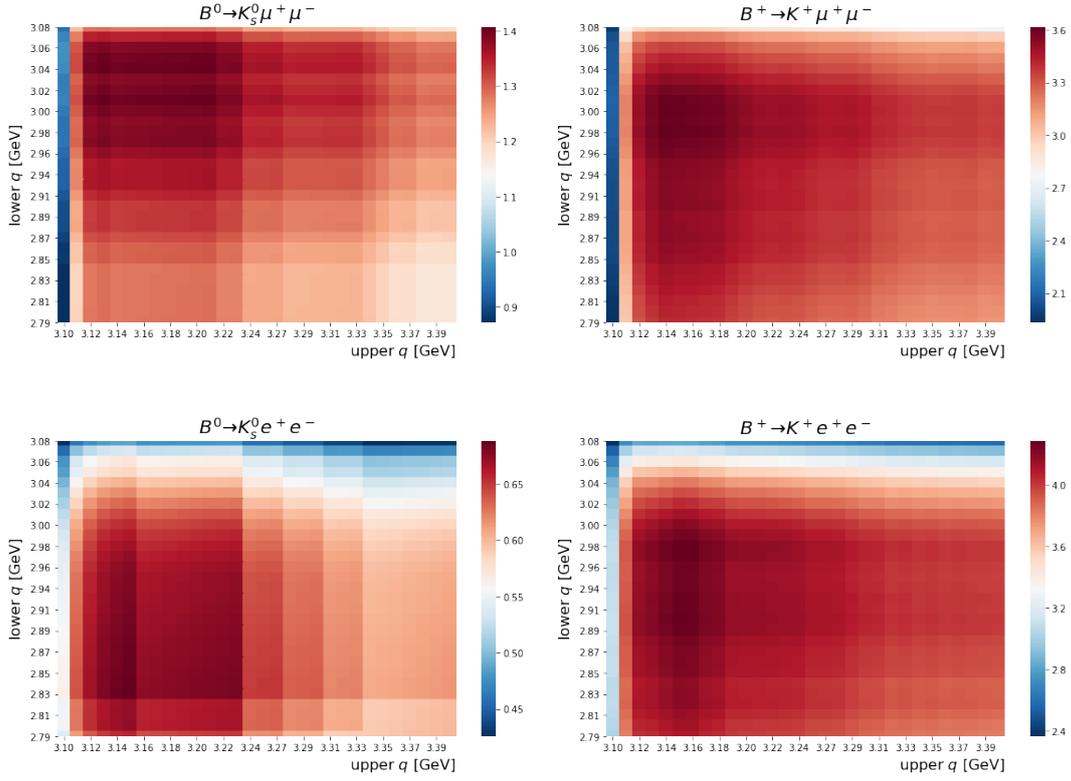


Fig. D.1 Figure of merit distribution for the different J/ψ veto windows $[q_{\text{low}}, q_{\text{high}}]$.

The procedure for $\psi(2S)$ is analogous. From figure D.2, the best regions are:

$$\begin{array}{l|l}
 B^0 \rightarrow K_s^0 \mu^+ \mu^- & -0.179 \text{ GeV} < q - m(\psi(2S)) < 0.041 \text{ GeV} \\
 B^+ \rightarrow K^+ \mu^+ \mu^- & -0.084 \text{ GeV} < q - m(\psi(2S)) < 0.031 \text{ GeV} \\
 B^0 \rightarrow K_s^0 e^+ e^- & -0.094 \text{ GeV} < q - m(\psi(2S)) < 0.031 \text{ GeV} \\
 B^+ \rightarrow K^+ e^+ e^- & -0.073 \text{ GeV} < q - m(\psi(2S)) < 0.063 \text{ GeV}
 \end{array}$$

In this case the best region for the neutral muon channel is specially broad; this is due to low statistics of the sample, since increasing the veto window would still get rid of non-peaking background as well. If the reduction of non-peaking background achieved by this is greater than the reduction in the number of signal events selected, the FOM gets maximized. Hence, this time, we rely on the percentage of rejected charmonium events to select the veto region. In order to reject at least 99.7% of the peaking background, the cuts applied are:

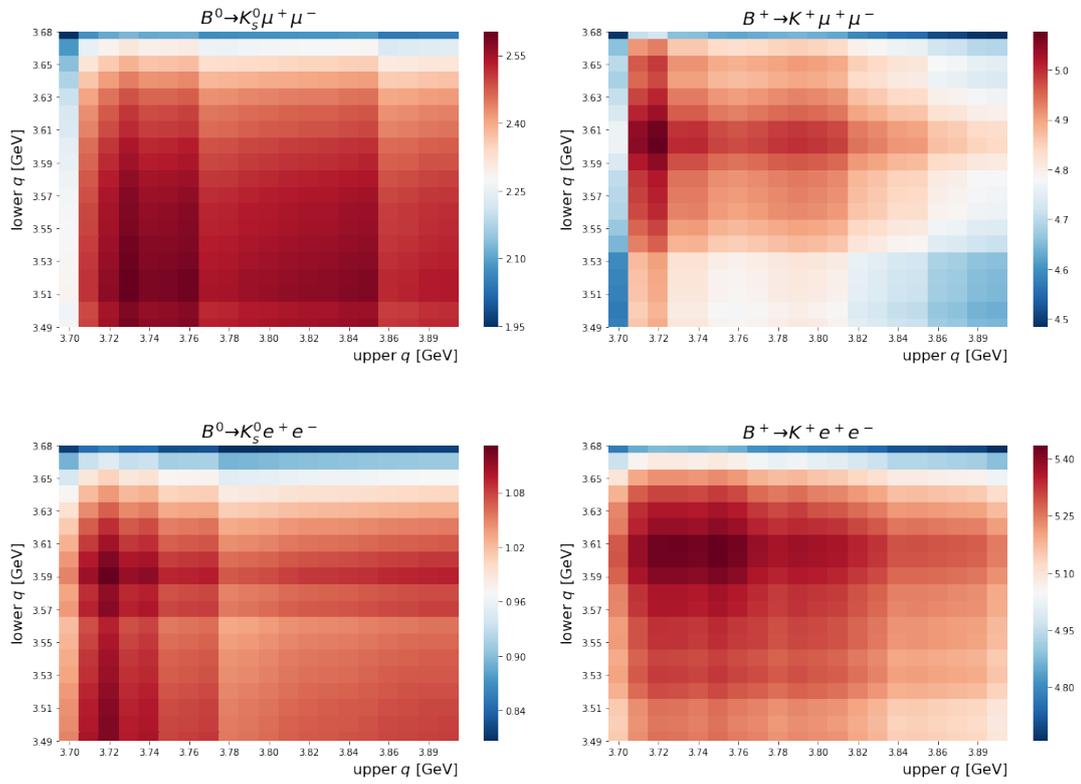


Fig. D.2 Figure of merit distribution for the different $\psi(2S)$ veto windows $[q_{\text{low}}, q_{\text{high}}]$.

$$\begin{array}{l}
 B \rightarrow K \mu^+ \mu^- \\
 B^+ \rightarrow K^+ e^+ e^-
 \end{array}
 \left| \begin{array}{l}
 -0.11 \text{ GeV} < q - m(J/\psi) < 0.05 \text{ GeV} \\
 -0.13 \text{ GeV} < q - m(J/\psi) < 0.07 \text{ GeV}
 \end{array} \right.$$

Appendix E

Boosted Decision Trees

In this appendix we briefly discuss the use of a collection of weak classifiers, known as decision trees, to deliver a robust and powerful classification output. A thorough overview of boosting can be found in [91].

Decision trees are a supervised learning method for classification and regression. They are composed of nodes, which represent tests on a variable (e.g., is the variable higher or lower than certain threshold value), and from which two or more branches lead to different nodes, according to the result of the test performed. Once a node with no more branches is reached (a leaf node), the output of the classification is calculated by different means: in classification, it is usual to classify the incoming event according to the largest class in the leaf the event is assigned to. The maximum number of nodes between the first one (or root) and each one of the leaves is known as the height (or depth) of the tree.

In the case at hand, we aim to establish the probability $y^{(i)}$ that an event, described by the parameters $\mathbf{x}^{(i)} = \{x_1^{(i)}, \dots, x_N^{(i)}\}$, corresponds to a signal event. A perfect classifier would have $y^{(i)} \in \{0, 1\}$ for every i ; this would imply a perfect knowledge of all possible values of \mathbf{x} for signal events. However, given the finite nature of the training data on which these classifiers are built, their output is instead a continuous variable in the range $[0, 1]$, following a distribution different to $y^{(i)}$; let this approximation be $\hat{y}^{(i)}$. The closer $\hat{y}^{(i)}$ is to one (zero), the more likely it is that the event corresponds to a signal (background) one.

Decision trees work by selecting as the variable for the next node the one that maximizes a certain criterion, usually the information gain, or minimize some other, such as the Gini impurity. Deep decision trees can reduce the impurity of each of their leaf nodes to zero when performing over the training dataset, at the cost of generalizing poorly to unseen cases; this is, the trees overfit the training distribution;

on the contrary, a shallow tree can generalize well to new data, but has a low separation power.

E.1 Boosted Decision Trees

BDTs follow a different approach: they approximate the probability distribution $y(x_1, \dots, x_N)$ as the sum of shallow trees:

$$\hat{y}(x_1, \dots, x_N) = \sum_{m=0}^M f_m(x_1, \dots, x_N). \quad (\text{E.1})$$

The f_m functions are chosen in order to minimize the expected value of a loss function, $L(y, \hat{y})$, which measures the goodness of the fit. For signal/background classification, each event can be seen as an independent Bernoulli trial, for which the proper success probability $\hat{y}^{(i)}$ needs to be determined. Assume that, for N events, the first n where signal events, and the remaining $N - n$ background¹. The estimated probability of observing this distribution is given by

$$\hat{y}^{(1)} \cdot \dots \cdot \hat{y}^{(n)} \left(1 - \hat{y}^{(n+1)}\right) \cdot \dots \cdot \left(1 - \hat{y}^{(N)}\right) \quad (\text{E.2})$$

since this is precisely the distribution we observe, it would be natural to expect this probability to be as high as possible. This is, the values of each $\hat{y}^{(i)}$ are such that the product is a maximum. In order to simplify the calculations, we use the logarithm of the previous function, as the logarithm is a monotonically increasing function; hence, we define the loss function as the negative of the logarithm, and the maximization problem becomes a minimization one:

$$L(y, \hat{y}) = - \left[\sum_i \ln \left(\hat{y}^{(i)}\right) + \sum_j \ln \left(1 - \hat{y}^{(j)}\right) \right] \quad (\text{E.3})$$

where i runs over all the signal events, and j over all the background ones. It is possible to make an explicit use of the target variable y by keeping in mind that, for

¹Notice that the order of signal and background events does matter, as the corresponding parameters $\mathbf{x}^{(i)}$ will be different.

signal (background) events, $y^{(i)} = 1(0)$. Hence,

$$L(y, \hat{y}) = - \left[\sum_{i=1}^N y^{(i)} \ln(\hat{y}^{(i)}) + (1 - y^{(i)}) \ln(1 - \hat{y}^{(i)}) \right]. \quad (\text{E.4})$$

Boosting then constructs the function \hat{y} by adding multiple weak learners in a weighted fashion

$$\hat{y}(x_1, \dots, x_N) = \sum_{m=0}^M \beta_m h(x_1, \dots, x_N | \mathbf{a}_m). \quad (\text{E.5})$$

Here, h is the mapping function of the weak classifiers, \mathbf{a}_m are the specific parameters for the m -th classifier, and β_m is the weight associated with it. The values for β_m and \mathbf{a}_m are obtained from a fit to the training data in a *stage-wise* approach. An initial proxy for \hat{y} ,

$$\hat{y}(x_1, \dots, x_N) = F_0(x_1, \dots, x_N)$$

is proposed, and then, for $m = 1, \dots, M$, β_m and \mathbf{a}_m are obtained by minimizing the loss function for the new proxy $F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta h(\mathbf{x} | \mathbf{a})$:

$$(\beta_m, \mathbf{a}_m) = \underset{\beta, \mathbf{a}}{\operatorname{argmin}} L(y, F_{m-1}(\mathbf{x}) + \beta h(\mathbf{x} | \mathbf{a})). \quad (\text{E.6})$$

Gradient boosting is an approximate solution to equation E.6, that works for any differentiable loss function. It is divided in two steps; in the first one, the gradient of the loss function with respect to the previous proxy approximation is obtained for each training example. These gradients are known as the pseudo-residuals $r^{(i)}$

$$r_m^{(i)} = \frac{\partial L(y^{(i)}, F_{m-1}(x^{(i)}))}{\partial F_{m-1}(x^{(i)})}, \quad (\text{E.7})$$

where, in the case of signal/background classification

$$L(y^{(i)}, F(x^{(i)})) = \begin{cases} -\ln(F(x^{(i)})), & \text{if } y^{(i)} = 1 \\ -\ln(1 - F(x^{(i)})), & \text{if } y^{(i)} = 0. \end{cases} \quad (\text{E.8})$$

These pseudo-residuals are then used to obtain the the parameters \mathbf{a}_m by a least-squares approach:

$$\mathbf{a}_m = \underset{\mathbf{a}, \rho}{\operatorname{argmin}} \sum_i \left[r_m^{(i)} - \rho h(\mathbf{x}^{(i)} | \mathbf{a}) \right]. \quad (\text{E.9})$$

Next, the value of the parameter β_m is obtained by a simpler version of equation E.6, in which the \mathbf{a}_m is fixed:

$$\beta_m = \underset{\beta}{\operatorname{argmin}} L(y, F_{m-1}(\mathbf{x}) + \beta h(\mathbf{x}|\mathbf{a}_m)). \quad (\text{E.10})$$

BDTs are hence a particular type of boosted gradient classifiers, in which the weak learners correspond to shallow trees, and the \mathbf{a}_m parameters represent the variable/cut-value pairs for the nodes in the tree. At each m iteration, a classification tree partitions the \mathbf{x} -space into K disjoint regions $\{R_m^{(k)}\}_{k=1,\dots,K}$, where K corresponds to the number of leaf-nodes in the tree, and predicts a constant value for each one. As mentioned before, this constant value is usually the mean over the \mathbf{x} -subspace mapped to that leaf. With trees, equation E.10 can be solved separately within each region. Empirically, it was found that the introduction of a learning rate (or *shrinkage*) parameter $0 < \nu \leq 1$ in the update step for the solution leads to a smaller generalization error. Hence, for $m = 1, \dots, M$:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu \cdot \beta_m h(\mathbf{x}|\mathbf{a}_m). \quad (\text{E.11})$$

E.2 Boosting to Flatness

Rogozhnikov et al [92] introduced some novel methods in order to assure that the BDT output is not correlated with variables of interest, which would introduce undesirable biases in the analysis. The one used in this analysis is what they call the uGBFL(bin) algorithm; it introduces an extra term into the loss function to be minimized by the BDT, which grows larger the more non-uniform the BDT output is over the range of the variables of interest; this term is based in the fact that, if the data pertaining to the variables of interest is binned, any cut in the output of a uniform classifier will result in roughly the same selection efficiency (within statistical fluctuations) for each of the bins. The exact form of the flatness penalty term is:

$$L_{\text{flat}} = \sum_b w_b \int |F_b(\hat{y}) - F(\hat{y})|^2 d\hat{y}, \quad (\text{E.12})$$

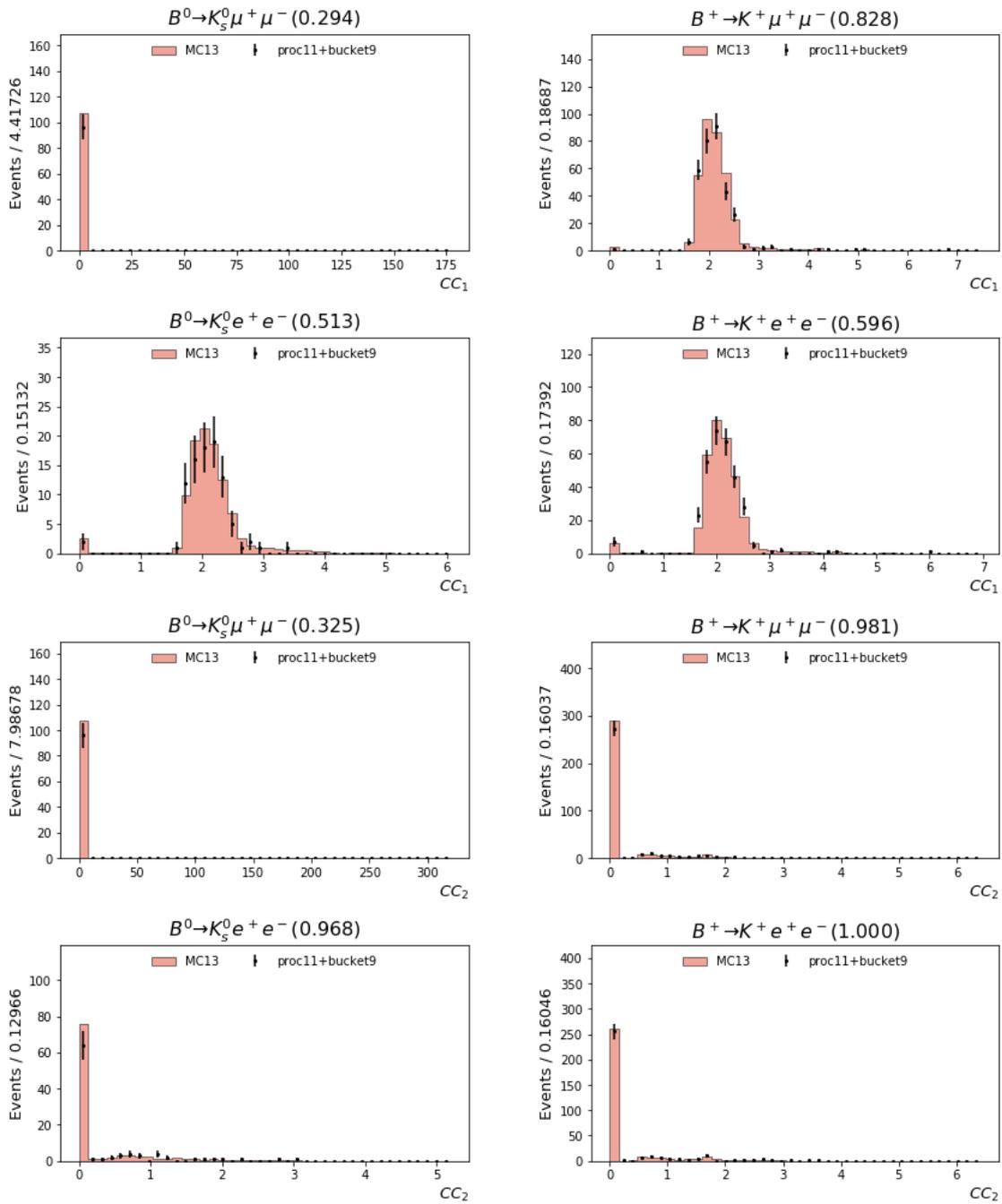
where b ranges over all the bins and w_b is the (weighted) fraction of events in bin b ; $F_b(\hat{y})$ and $F(\hat{y})$ are similar, and correspond to the cumulative distributions of $f_b(\hat{y})$ and $f(\hat{y})$, respectively, where $f_b(\hat{y})$ is the distribution of the classifier output in the bin b and $f(\hat{y})$ is the global distribution of the classifier output. This loss

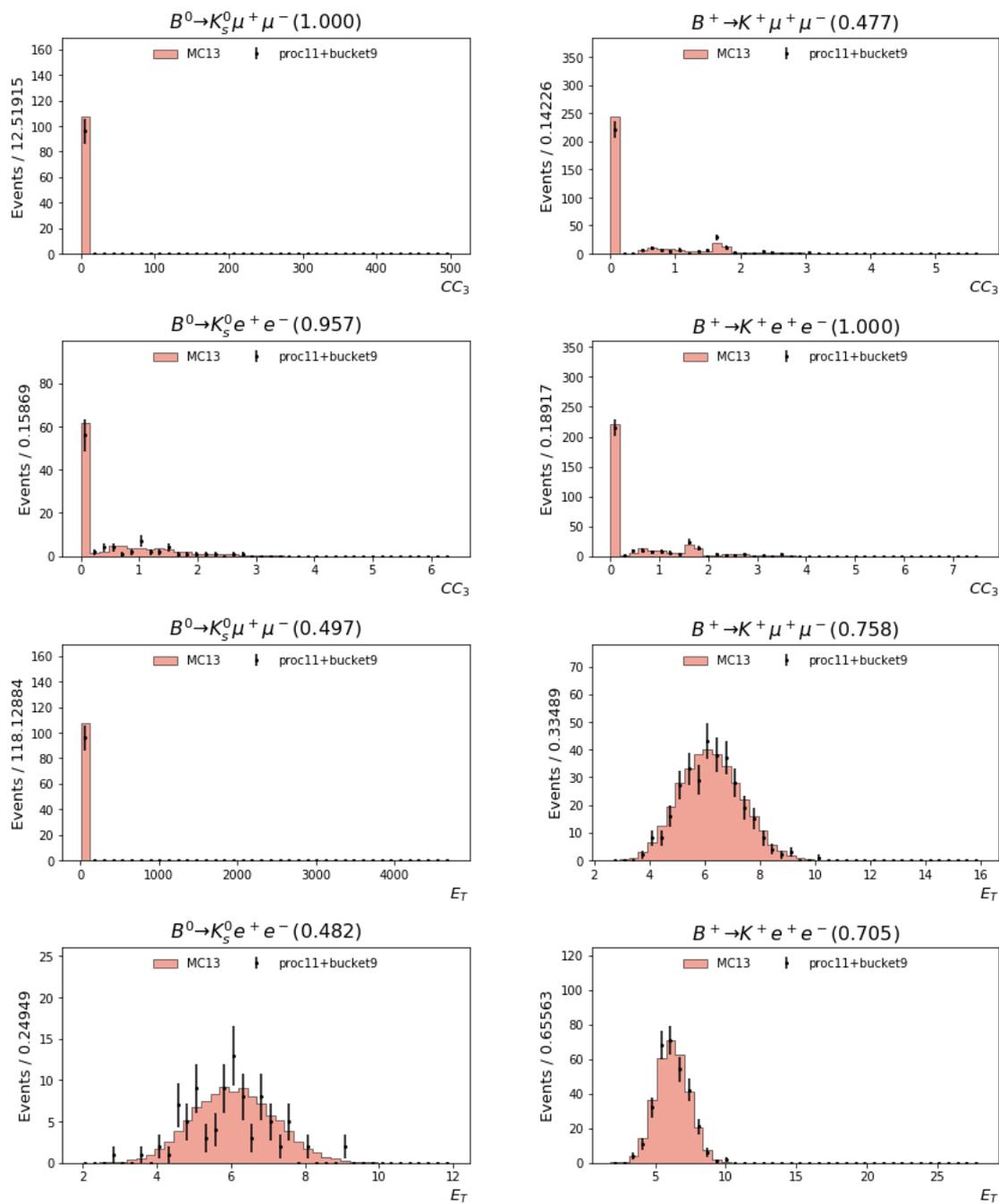
function is added to the one in equation [E.4](#) after being weighted by a factor α , which determines the relative importance of having a flat distribution against having a strong separation power.

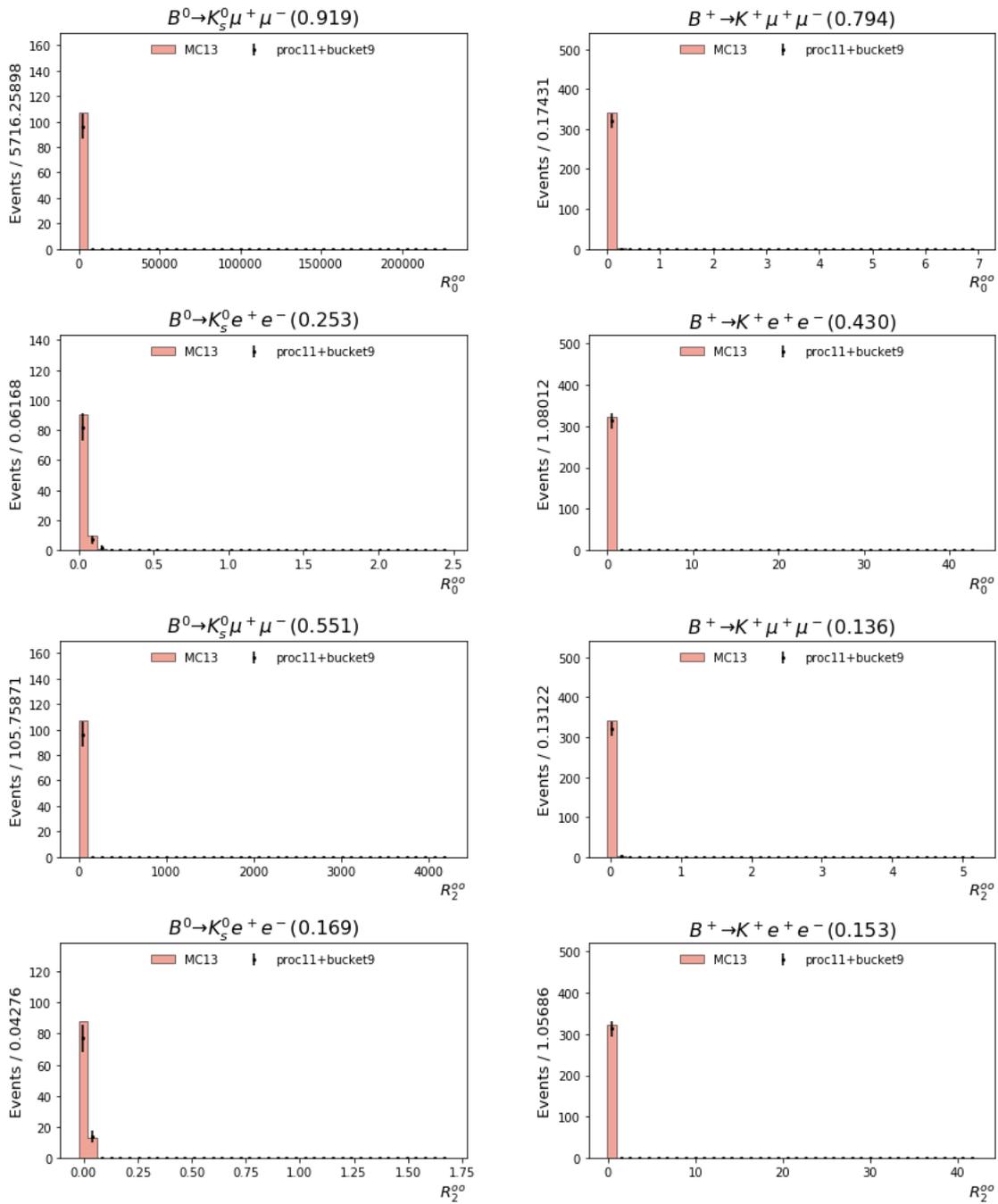
Appendix F

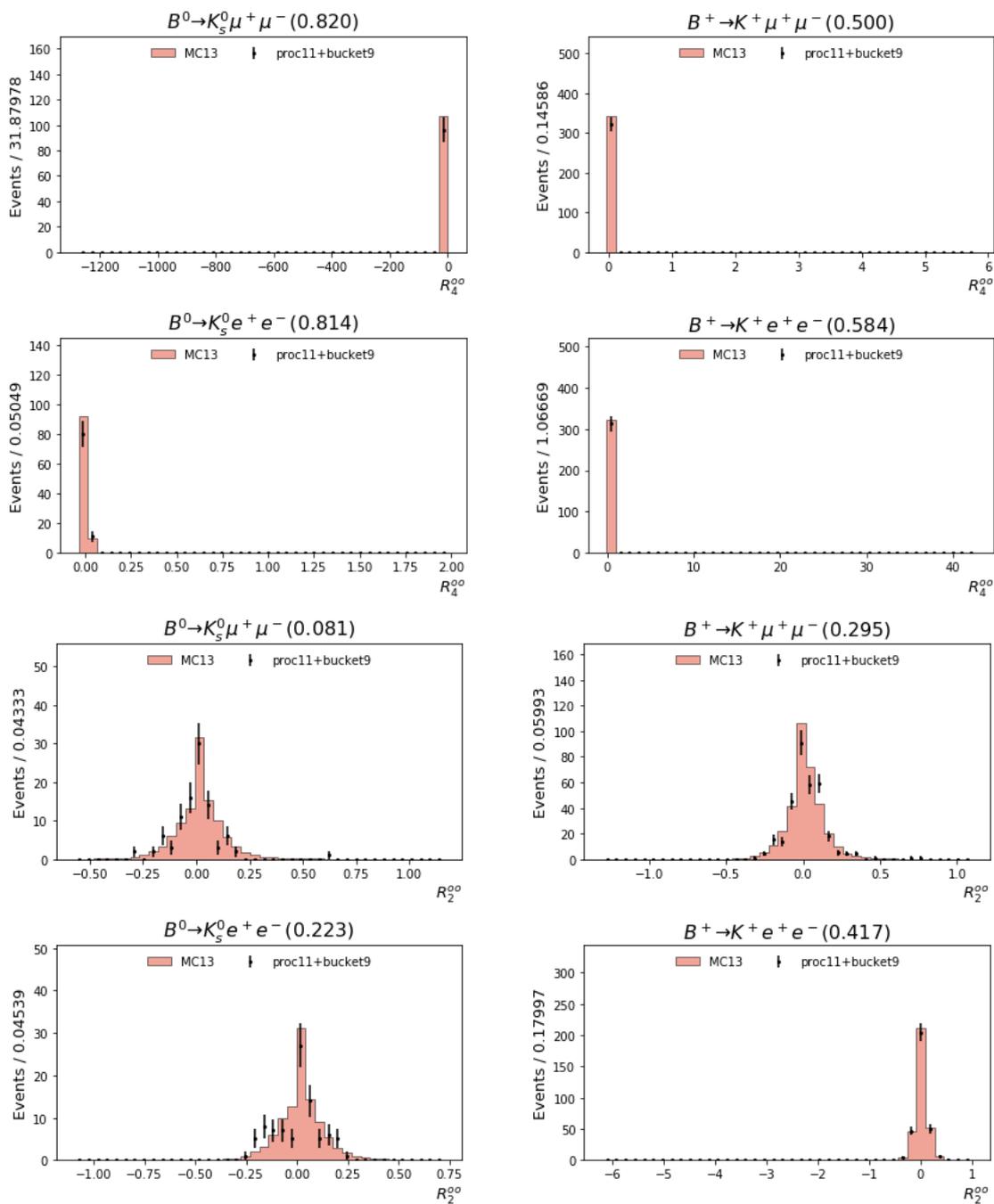
Distribution of the MVA Training Variables

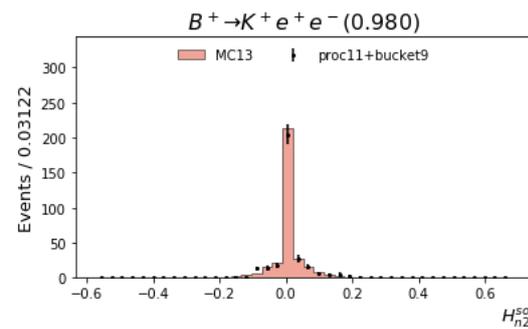
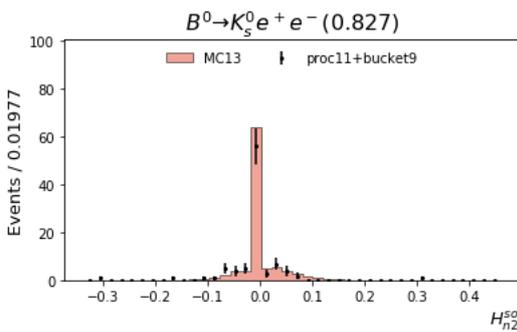
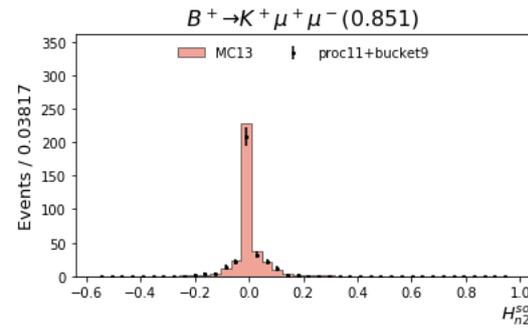
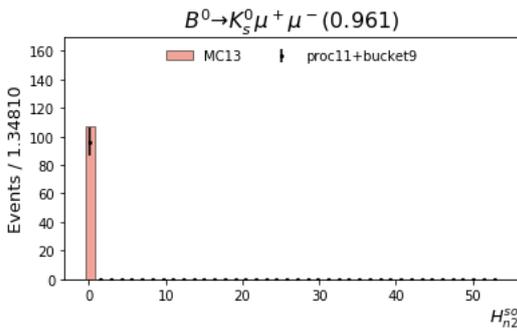
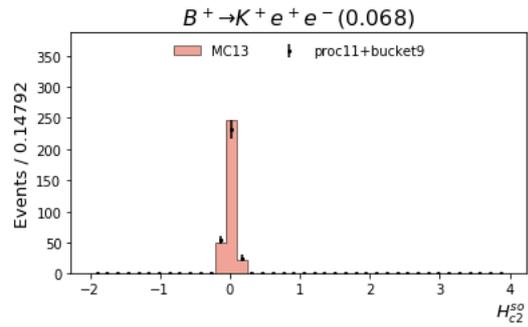
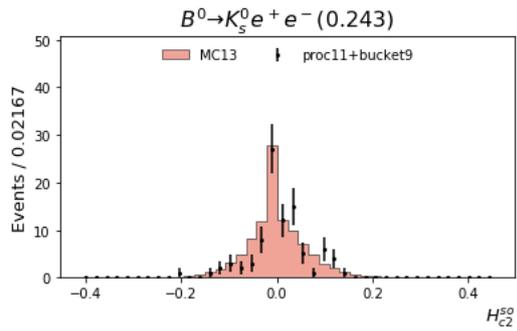
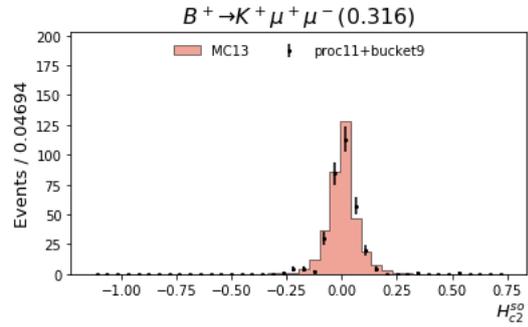
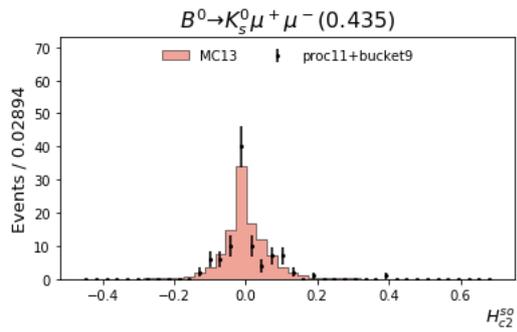
The following plots depict the distribution of the all the available training variables for the MVA classifiers, for events in the charmonium veto regions, after applying the initial reconstruction cuts and the D meson and Dalitz vetoes. The filled histogram represents the distribution for MC13 data, composed of 2 ab^{-1} of $B\bar{B}$ and $q\bar{q}$ events, using as cross sections the values presented in table 3.2; the black dots correspond to 11.53 fb^{-1} of experimental data. The numbers after each decay mode represent the p -value of the Kolmogorov-Smirnoff test, which measures how similar two distributions are.

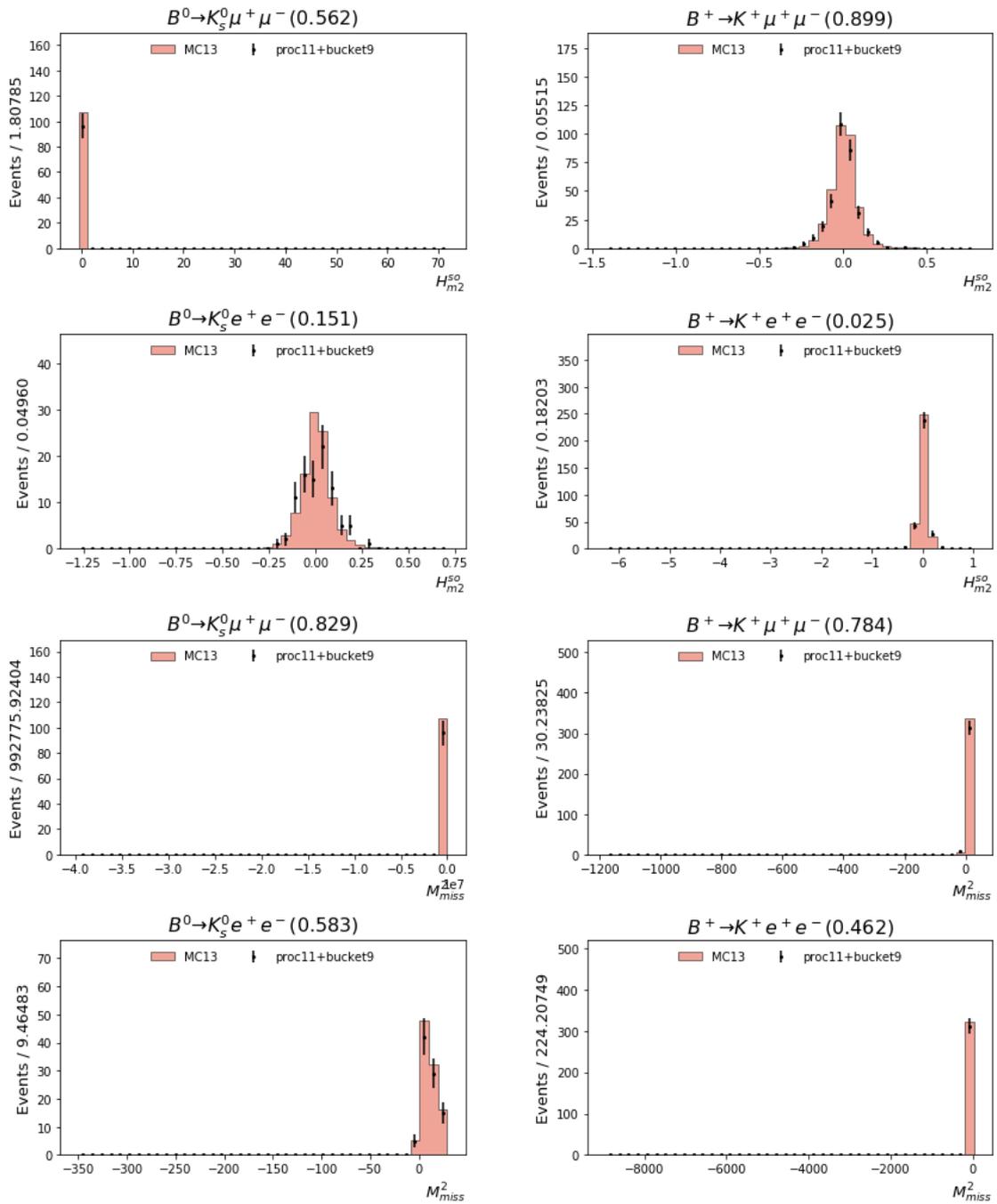


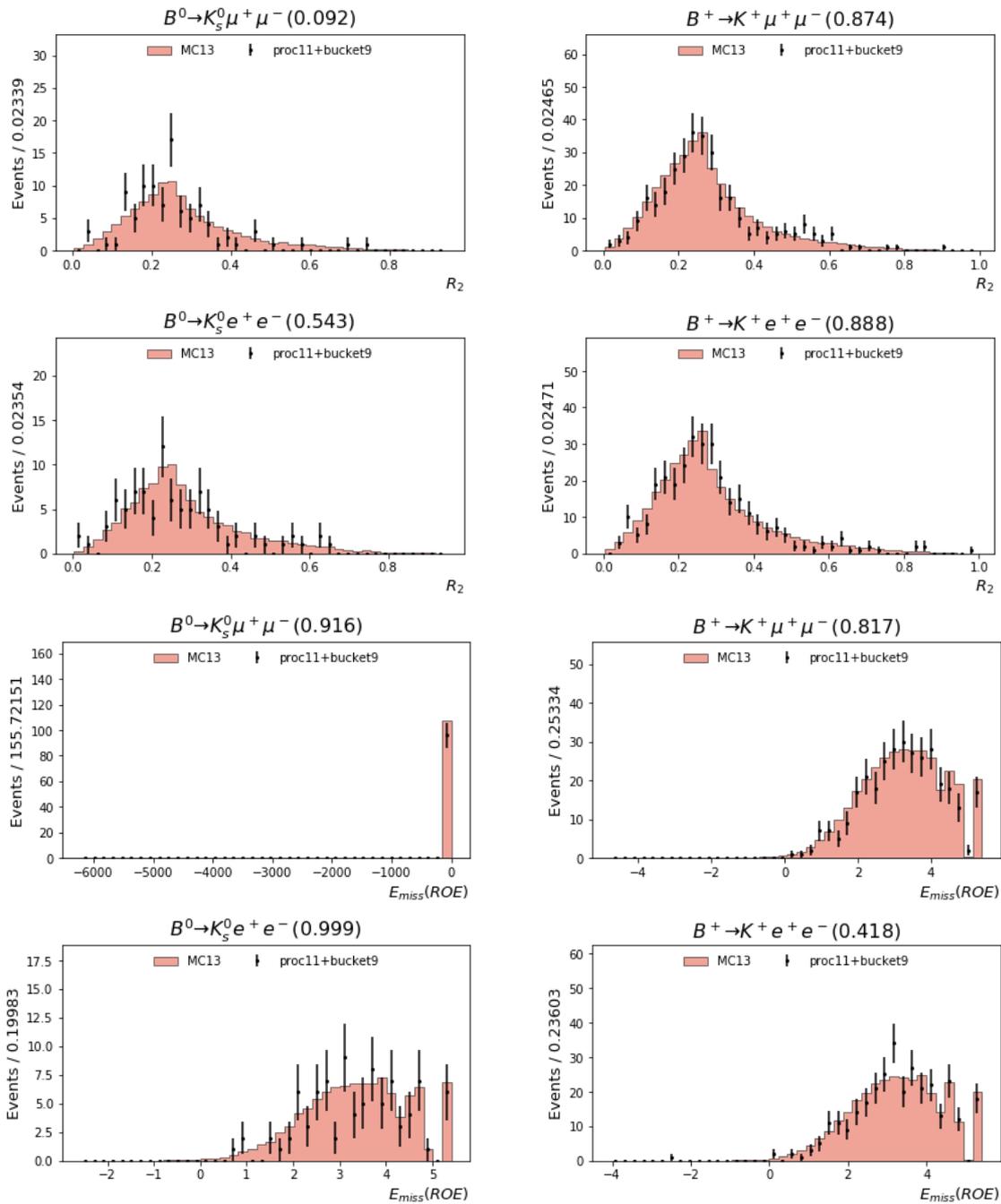


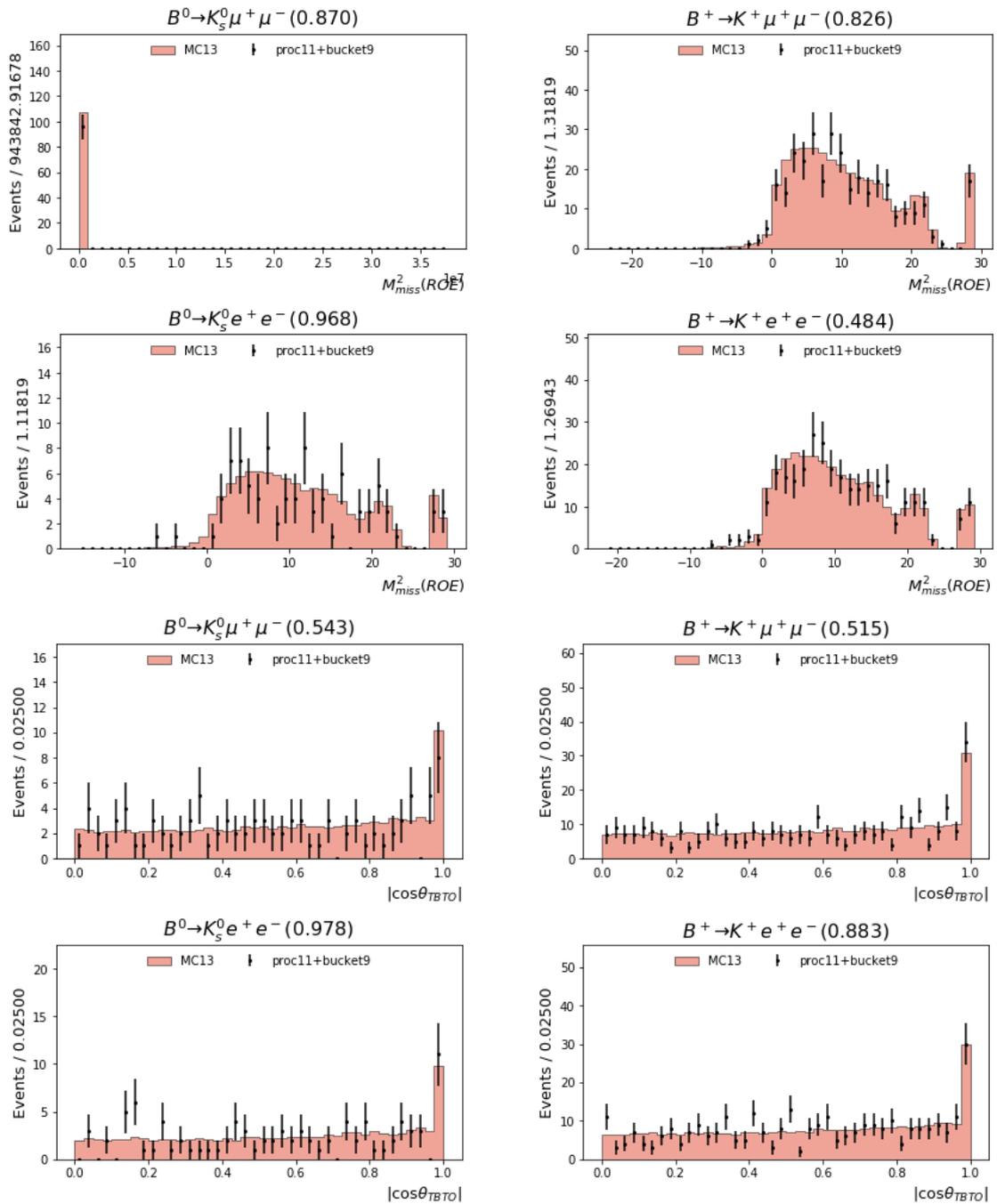


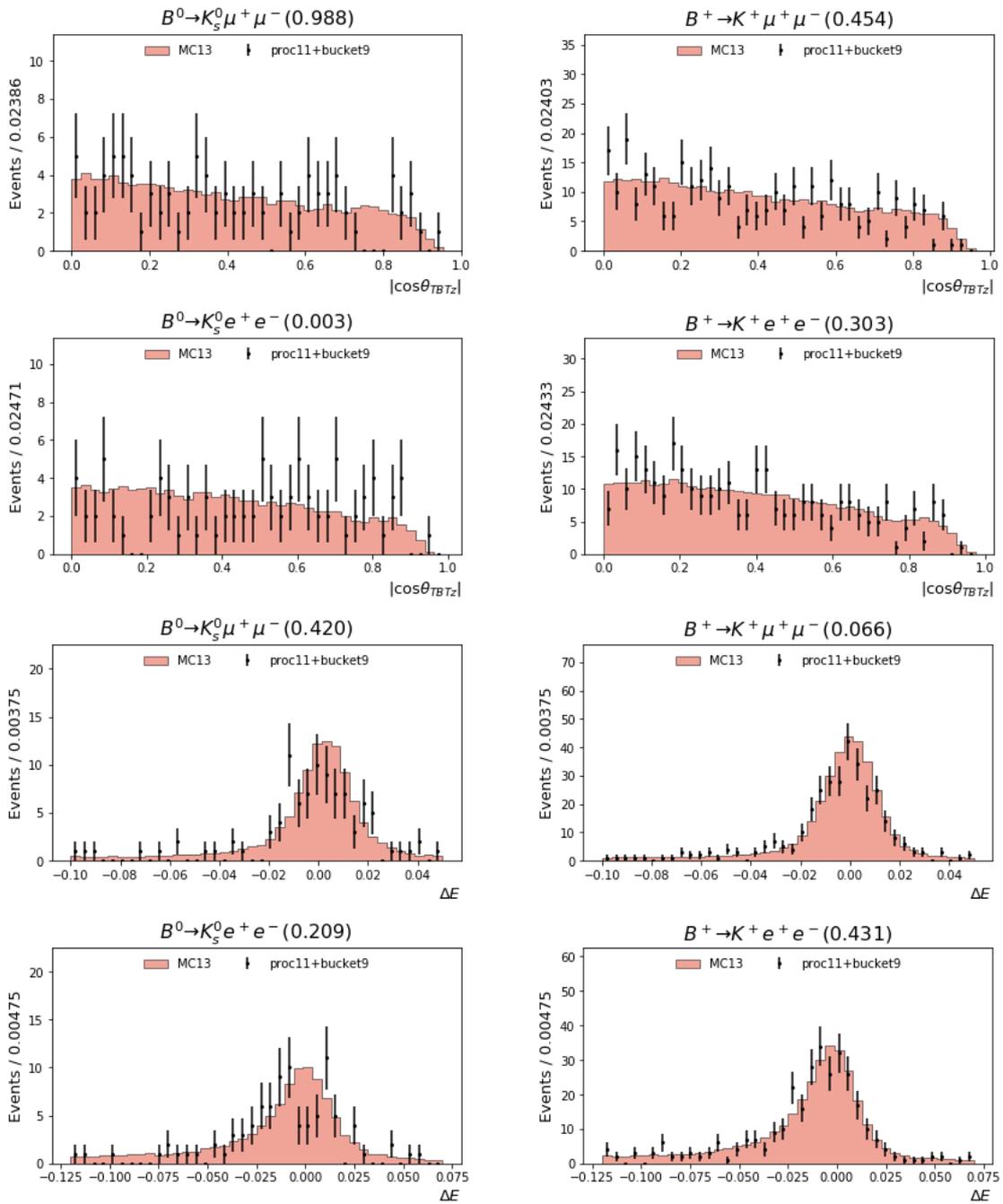


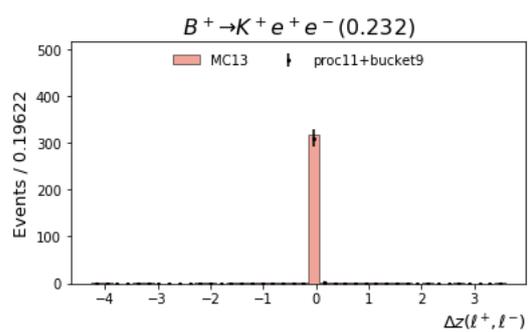
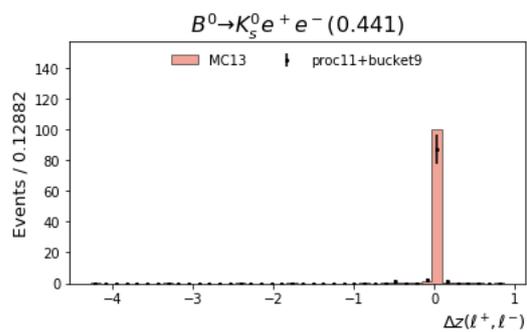
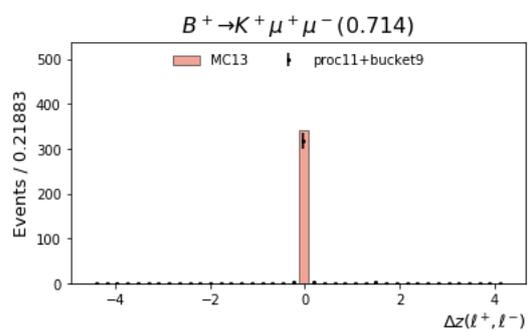
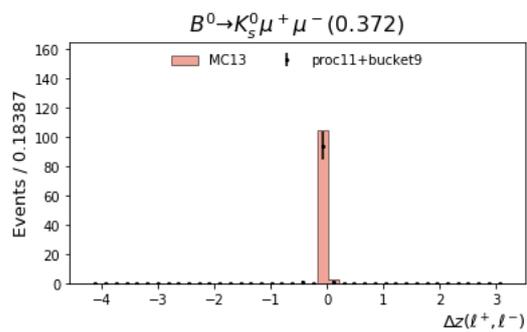


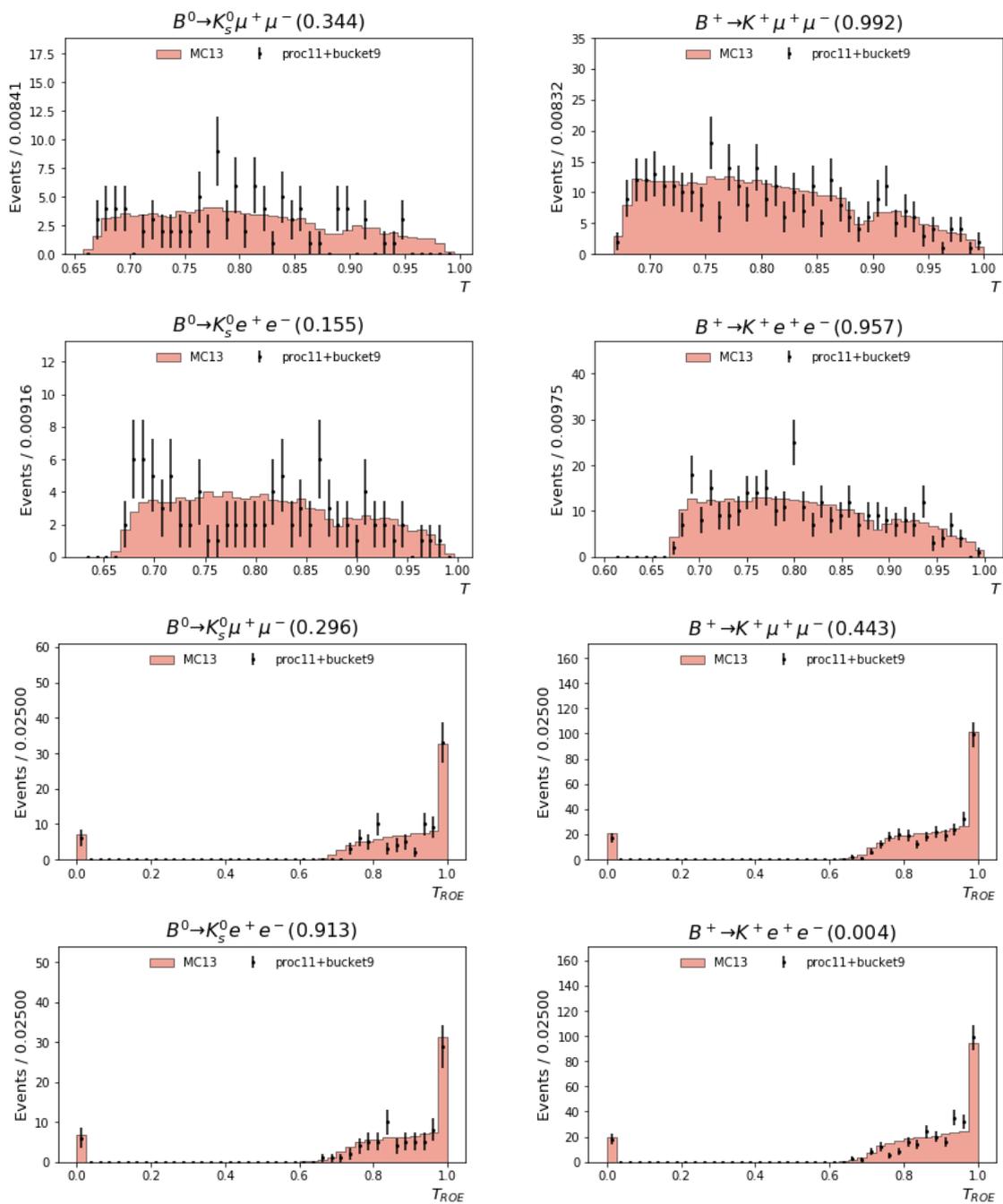












Appendix G

Efficiency and Fake Rate Correction Tables from Particle ID Studies

This appendix presents the longtables with the reconstruction efficiencies for data and MC used to obtain the correction factors for the different particle IDs used in this study; it also includes the longtables for pion to lepton fakes rates used in chapter 6. Some bins lack the sufficient number of events required to obtain a reliable correction factor; consequently, reconstructed events that fall inside these regions are properly discarded in the different step of the data analysis.

Table G.1. Selection efficiencies and efficiency correction terms for the kaon ID

$p_K[\text{GeV}]$	$\cos\theta_K$	$\varepsilon_{\text{data}}$	ε_{MC}	r_{kaonID}
(0.50,1.00)	(-0.866, -0.682)	0.34 ± 0.02	0.33 ± 0.01	1.04 ± 0.07
	(-0.682, -0.423)	0.52 ± 0.02	0.62 ± 0.01	0.84 ± 0.03
	(-0.423, -0.104)	0.87 ± 0.01	0.88 ± 0.00	1.00 ± 0.01
	(-0.104, 0.225)	0.84 ± 0.01	0.85 ± 0.00	0.98 ± 0.01
	(0.225, 0.500)	0.83 ± 0.01	0.82 ± 0.01	1.01 ± 0.02
	(0.500, 0.766)	0.82 ± 0.01	0.80 ± 0.01	1.02 ± 0.02
	(0.766, 0.883)	0.81 ± 0.02	0.83 ± 0.01	0.97 ± 0.02
	(0.883, 0.956)	0.81 ± 0.03	0.88 ± 0.01	0.93 ± 0.03
(1.00,1.50)	(-0.866, -0.682)	0.06 ± 0.01	0.12 ± 0.01	0.46 ± 0.09
	(-0.682, -0.423)	0.41 ± 0.02	0.45 ± 0.01	0.93 ± 0.04

Table G.1. Selection efficiencies and efficiency correction terms for the kaon ID

$p_K [GeV]$	$\cos\theta_K$	ϵ_{data}	ϵ_{MC}	r_{kaonID}
	(-0.423, -0.104)	0.87 ± 0.01	0.87 ± 0.00	1.00 ± 0.01
	(-0.104, 0.225)	0.79 ± 0.01	0.83 ± 0.00	0.95 ± 0.01
	(0.225, 0.500)	0.80 ± 0.01	0.81 ± 0.00	0.99 ± 0.02
	(0.500, 0.766)	0.82 ± 0.01	0.83 ± 0.00	0.99 ± 0.01
	(0.766, 0.883)	0.79 ± 0.02	0.80 ± 0.01	0.99 ± 0.02
	(0.883, 0.956)	0.78 ± 0.02	0.78 ± 0.01	1.01 ± 0.03
(1.50, 2.00)	(-0.866, -0.682)	0.26 ± 0.02	0.22 ± 0.01	1.16 ± 0.11
	(-0.682, -0.423)	0.58 ± 0.01	0.46 ± 0.01	1.24 ± 0.04
	(-0.423, -0.104)	0.88 ± 0.01	0.89 ± 0.00	0.99 ± 0.01
	(-0.104, 0.225)	0.84 ± 0.01	0.87 ± 0.00	0.97 ± 0.01
	(0.225, 0.500)	0.80 ± 0.01	0.83 ± 0.00	0.96 ± 0.01
	(0.500, 0.766)	0.81 ± 0.01	0.85 ± 0.00	0.96 ± 0.01
	(0.766, 0.883)	0.79 ± 0.02	0.77 ± 0.01	1.04 ± 0.02
	(0.883, 0.956)	0.73 ± 0.03	0.68 ± 0.01	1.08 ± 0.04
(2.00, 2.50)	(-0.866, -0.682)	0.12 ± 0.02	0.10 ± 0.01	1.17 ± 0.24
	(-0.682, -0.423)	0.40 ± 0.02	0.30 ± 0.01	1.33 ± 0.06
	(-0.423, -0.104)	0.88 ± 0.01	0.86 ± 0.00	1.02 ± 0.01
	(-0.104, 0.225)	0.78 ± 0.01	0.83 ± 0.00	0.93 ± 0.02
	(0.225, 0.500)	0.74 ± 0.01	0.79 ± 0.00	0.94 ± 0.02
	(0.500, 0.766)	0.77 ± 0.01	0.80 ± 0.00	0.96 ± 0.01
	(0.766, 0.883)	0.78 ± 0.02	0.78 ± 0.01	1.00 ± 0.02
	(0.883, 0.956)	0.88 ± 0.02	0.85 ± 0.01	1.04 ± 0.03
(2.50, 3.00)	(-0.866, -0.682)	0.04 ± 0.03	0.06 ± 0.01	0.69 ± 0.52
	(-0.682, -0.423)	0.26 ± 0.02	0.23 ± 0.01	1.13 ± 0.09
	(-0.423, -0.104)	0.79 ± 0.02	0.84 ± 0.01	0.95 ± 0.02

Table G.1. Selection efficiencies and efficiency correction terms for the kaon ID

$p_K[\text{GeV}]$	$\cos\theta_K$	$\varepsilon_{\text{data}}$	ε_{MC}	r_{kaonID}
	(-0.104,0.225)	0.76 ± 0.02	0.80 ± 0.01	0.94 ± 0.02
	(0.225,0.500)	0.67 ± 0.01	0.72 ± 0.01	0.94 ± 0.02
	(0.500,0.766)	0.70 ± 0.01	0.73 ± 0.00	0.95 ± 0.02
	(0.766,0.883)	0.76 ± 0.02	0.81 ± 0.01	0.94 ± 0.02
	(0.883,0.956)	0.89 ± 0.02	0.87 ± 0.01	1.01 ± 0.03
(3.00,3.50)	(-0.866,-0.682)	0.06 ± 0.05	0.01 ± 0.01	8.73 ± 13.50
	(-0.682,-0.423)	0.20 ± 0.03	0.22 ± 0.01	0.93 ± 0.16
	(-0.423,-0.104)	0.74 ± 0.03	0.79 ± 0.01	0.93 ± 0.03
	(-0.104,0.225)	0.66 ± 0.02	0.73 ± 0.01	0.90 ± 0.03
	(0.225,0.500)	0.61 ± 0.02	0.64 ± 0.01	0.95 ± 0.03
	(0.500,0.766)	0.63 ± 0.01	0.68 ± 0.01	0.92 ± 0.02
	(0.766,0.883)	0.73 ± 0.02	0.79 ± 0.01	0.93 ± 0.02
	(0.883,0.956)	0.89 ± 0.02	0.91 ± 0.01	0.98 ± 0.03
(3.50,4.50)	(-0.866,-0.682)	0.00 ± 70607.40	0.00 ± 0.06	$0.00 \pm \text{nan}$
	(-0.682,-0.423)	0.27 ± 0.10	0.24 ± 0.04	1.11 ± 0.44
	(-0.423,-0.104)	0.70 ± 0.05	0.77 ± 0.02	0.91 ± 0.07
	(-0.104,0.225)	0.61 ± 0.03	0.67 ± 0.01	0.90 ± 0.05
	(0.225,0.500)	0.58 ± 0.02	0.61 ± 0.01	0.95 ± 0.04
	(0.500,0.766)	0.57 ± 0.02	0.68 ± 0.01	0.85 ± 0.02
	(0.766,0.883)	0.63 ± 0.02	0.76 ± 0.01	0.82 ± 0.03
	(0.883,0.956)	0.86 ± 0.03	0.90 ± 0.01	0.96 ± 0.03

Table G.2. Selection efficiencies and efficiency correction terms for the electron ID

$p_e[\text{GeV}]$	$\cos\theta_e$	charge	$\varepsilon_{\text{data}}$	ε_{MC}	$r_{\text{electronID}}$	
(0.50,1.00)	(-0.908, -0.612)	+	0.48 ± 0.25	0.86 ± 0.07	0.55 ± 0.29	
		-	0.29 ± 0.21	0.46 ± 0.07	0.69 ± 0.57	
	(-0.612, -0.304)	+	0.88 ± 0.15	1.03 ± 0.05	0.86 ± 0.15	
		-	0.79 ± 0.14	0.96 ± 0.03	0.82 ± 0.14	
	(-0.304, 0.001)	+	0.93 ± 0.16	0.93 ± 0.04	1.01 ± 0.18	
	(0.001, 0.427)	-	0.77 ± 0.18	0.98 ± 0.04	0.79 ± 0.19	
	(1.00,1.50)	(-0.908, -0.612)	+	0.80 ± 0.05	0.90 ± 0.02	0.89 ± 0.06
			-	0.28 ± 0.03	0.48 ± 0.02	0.58 ± 0.07
(-0.612, -0.304)		+	0.95 ± 0.02	0.98 ± 0.01	0.97 ± 0.02	
		-	0.77 ± 0.02	0.86 ± 0.01	0.90 ± 0.02	
(-0.304, 0.001)		+	1.00 ± 0.02	0.99 ± 0.01	1.02 ± 0.02	
		-	0.96 ± 0.02	0.98 ± 0.01	0.98 ± 0.02	
(0.001, 0.427)		+	0.97 ± 0.01	0.99 ± 0.01	0.99 ± 0.02	
		-	0.94 ± 0.02	0.98 ± 0.01	0.97 ± 0.02	
(0.427, 0.847)		+	1.01 ± 0.03	1.01 ± 0.01	1.00 ± 0.03	
		-	0.94 ± 0.03	1.00 ± 0.01	0.95 ± 0.03	
(0.847, 0.976)		+	0.94 ± 0.16	1.03 ± 0.06	0.93 ± 0.16	
		-	0.37 ± 0.22	1.14 ± 0.09	0.34 ± 0.21	
(1.50,2.00)	(-0.908, -0.612)	+	0.77 ± 0.09	0.95 ± 0.02	0.81 ± 0.10	
		-	0.49 ± 0.07	0.76 ± 0.03	0.64 ± 0.09	
	(-0.612, -0.304)	+	0.98 ± 0.02	1.00 ± 0.01	0.99 ± 0.02	
		-	0.92 ± 0.02	0.95 ± 0.01	0.97 ± 0.02	
	(-0.304, 0.001)	+	0.99 ± 0.01	1.00 ± 0.00	0.99 ± 0.01	
		-	0.98 ± 0.01	0.99 ± 0.00	0.99 ± 0.01	
	(0.001, 0.427)	+	0.99 ± 0.01	1.01 ± 0.00	0.99 ± 0.01	

Table G.2. Selection efficiencies and efficiency correction terms for the electron ID

$p_e[\text{GeV}]$	$\cos\theta_e$	charge	$\varepsilon_{\text{data}}$	ε_{MC}	$r_{\text{electronID}}$	
		–	0.96 ± 0.01	0.99 ± 0.00	0.98 ± 0.01	
		+	1.02 ± 0.01	0.99 ± 0.00	1.03 ± 0.01	
	(0.427,0.847)	–	0.97 ± 0.01	0.97 ± 0.00	0.99 ± 0.01	
		+	0.97 ± 0.04	1.01 ± 0.01	0.96 ± 0.05	
	(0.847,0.976)	–	0.96 ± 0.05	0.97 ± 0.01	0.99 ± 0.05	
(2.00,2.50)	(–0.908,–0.612)	–	0.36 ± 0.49	0.78 ± 0.36	0.51 ± 0.86	
		+	0.89 ± 0.12	1.00 ± 0.03	0.89 ± 0.12	
	(–0.612,–0.304)	+	0.96 ± 0.05	1.00 ± 0.01	0.96 ± 0.05	
		–	0.92 ± 0.05	0.98 ± 0.01	0.94 ± 0.05	
	(0.001,0.427)	+	1.00 ± 0.02	1.00 ± 0.01	1.00 ± 0.02	
		–	0.94 ± 0.02	0.99 ± 0.01	0.95 ± 0.02	
	(0.427,0.847)	+	0.99 ± 0.01	1.00 ± 0.00	1.00 ± 0.01	
		–	0.97 ± 0.01	0.98 ± 0.00	0.99 ± 0.01	
	(0.847,0.976)	+	1.00 ± 0.03	1.00 ± 0.01	1.00 ± 0.04	
		–	0.94 ± 0.04	0.97 ± 0.01	0.96 ± 0.04	
	(2.50,3.00)	(–0.304,0.001)	+	0.89 ± 0.22	1.03 ± 0.16	0.81 ± 0.31
			+	1.03 ± 0.12	1.00 ± 0.02	1.06 ± 0.10
(0.001,0.427)		–	0.94 ± 0.09	0.99 ± 0.03	0.93 ± 0.09	
		+	0.96 ± 0.03	1.00 ± 0.01	0.95 ± 0.03	
(0.427,0.847)		–	0.98 ± 0.03	0.99 ± 0.01	0.99 ± 0.04	
		+	1.03 ± 0.10	1.01 ± 0.02	1.03 ± 0.09	
(0.847,0.976)	–	1.02 ± 0.07	0.97 ± 0.02	1.05 ± 0.08		

Table G.3. Selection efficiencies and efficiency correction terms for the muon ID

$p_\mu[\text{GeV}]$	$\cos\theta_\mu$	charge	$\varepsilon_{\text{data}}$	ε_{MC}	r_{muonID}	
(0.50,1.00)	(-0.908, -0.612)	+	0.48 ± 0.25	0.86 ± 0.07	0.55 ± 0.29	
		-	0.29 ± 0.21	0.46 ± 0.07	0.69 ± 0.57	
	(-0.612, -0.304)	+	0.88 ± 0.15	1.03 ± 0.05	0.86 ± 0.15	
		-	0.79 ± 0.14	0.96 ± 0.03	0.82 ± 0.14	
	(-0.304, 0.001)	+	0.93 ± 0.16	0.93 ± 0.04	1.01 ± 0.18	
	(0.001, 0.427)	-	0.77 ± 0.18	0.98 ± 0.04	0.79 ± 0.19	
	(1.00,1.50)	(-0.908, -0.612)	+	0.80 ± 0.05	0.90 ± 0.02	0.89 ± 0.06
			-	0.28 ± 0.03	0.48 ± 0.02	0.58 ± 0.07
(-0.612, -0.304)		+	0.95 ± 0.02	0.98 ± 0.01	0.97 ± 0.02	
		-	0.77 ± 0.02	0.86 ± 0.01	0.90 ± 0.02	
(-0.304, 0.001)		+	1.00 ± 0.02	0.99 ± 0.01	1.02 ± 0.02	
		-	0.96 ± 0.02	0.98 ± 0.01	0.98 ± 0.02	
(0.001, 0.427)		+	0.97 ± 0.01	0.99 ± 0.01	0.99 ± 0.02	
		-	0.94 ± 0.02	0.98 ± 0.01	0.97 ± 0.02	
(0.427, 0.847)		+	1.01 ± 0.03	1.01 ± 0.01	1.00 ± 0.03	
		-	0.94 ± 0.03	1.00 ± 0.01	0.95 ± 0.03	
(0.847, 0.976)		+	0.94 ± 0.16	1.03 ± 0.06	0.93 ± 0.16	
		-	0.37 ± 0.22	1.14 ± 0.09	0.34 ± 0.21	
(1.50,2.00)	(-0.908, -0.612)	+	0.77 ± 0.09	0.95 ± 0.02	0.81 ± 0.10	
		-	0.49 ± 0.07	0.76 ± 0.03	0.64 ± 0.09	
	(-0.612, -0.304)	+	0.98 ± 0.02	1.00 ± 0.01	0.99 ± 0.02	
		-	0.92 ± 0.02	0.95 ± 0.01	0.97 ± 0.02	
	(-0.304, 0.001)	+	0.99 ± 0.01	1.00 ± 0.00	0.99 ± 0.01	
		-	0.98 ± 0.01	0.99 ± 0.00	0.99 ± 0.01	
	(0.001, 0.427)	+	0.99 ± 0.01	1.01 ± 0.00	0.99 ± 0.01	

Table G.3. Selection efficiencies and efficiency correction terms for the muon ID

$p_\mu[\text{GeV}]$	$\cos\theta_\mu$	charge	$\varepsilon_{\text{data}}$	ε_{MC}	r_{muonID}	
	(0.427,0.847)	–	0.96 ± 0.01	0.99 ± 0.00	0.98 ± 0.01	
		+	1.02 ± 0.01	0.99 ± 0.00	1.03 ± 0.01	
	(0.847,0.976)	–	0.97 ± 0.01	0.97 ± 0.00	0.99 ± 0.01	
		+	0.97 ± 0.04	1.01 ± 0.01	0.96 ± 0.05	
	(2.00,2.50)	(–0.908,–0.612)	–	0.36 ± 0.49	0.78 ± 0.36	0.51 ± 0.86
			+	0.89 ± 0.12	1.00 ± 0.03	0.89 ± 0.12
(–0.304,0.001)		+	0.96 ± 0.05	1.00 ± 0.01	0.96 ± 0.05	
		–	0.92 ± 0.05	0.98 ± 0.01	0.94 ± 0.05	
(0.001,0.427)		+	1.00 ± 0.02	1.00 ± 0.01	1.00 ± 0.02	
		–	0.94 ± 0.02	0.99 ± 0.01	0.95 ± 0.02	
(0.427,0.847)	+	0.99 ± 0.01	1.00 ± 0.00	1.00 ± 0.01		
	–	0.97 ± 0.01	0.98 ± 0.00	0.99 ± 0.01		
(2.50,3.00)	(0.847,0.976)	+	1.00 ± 0.03	1.00 ± 0.01	1.00 ± 0.04	
		–	0.94 ± 0.04	0.97 ± 0.01	0.96 ± 0.04	
	(–0.304,0.001)	+	0.89 ± 0.22	1.03 ± 0.16	0.81 ± 0.31	
		–	0.94 ± 0.09	0.99 ± 0.03	0.93 ± 0.09	
	(0.001,0.427)	+	1.03 ± 0.12	1.00 ± 0.02	1.06 ± 0.10	
		–	0.94 ± 0.09	0.99 ± 0.03	0.93 ± 0.09	
(0.427,0.847)	+	0.96 ± 0.03	1.00 ± 0.01	0.95 ± 0.03		
	–	0.98 ± 0.03	0.99 ± 0.01	0.99 ± 0.04		
(0.847,0.976)	+	1.03 ± 0.10	1.01 ± 0.02	1.03 ± 0.09		
	–	1.02 ± 0.07	0.97 ± 0.02	1.05 ± 0.08		

Table G.4. Pion to electron fake rates

$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
(0.40,0.50)	(-0.908,-0.612)	+	0.022 ± 0.001	0.003 ± 0.000	6.78 ± 0.74
		-	0.014 ± 0.001	0.003 ± 0.000	5.18 ± 0.72
	(-0.612,-0.304)	+	0.032 ± 0.001	0.006 ± 0.000	5.17 ± 0.43
		-	0.026 ± 0.001	0.007 ± 0.000	3.62 ± 0.26
	(-0.304,0.001)	+	0.026 ± 0.001	0.008 ± 0.000	3.11 ± 0.24
		-	0.021 ± 0.001	0.009 ± 0.000	2.25 ± 0.16
	(0.001,0.427)	+	0.031 ± 0.001	0.010 ± 0.000	3.22 ± 0.16
		-	0.025 ± 0.001	0.011 ± 0.001	2.20 ± 0.14
	(0.427,0.847)	+	0.040 ± 0.001	0.019 ± 0.000	2.14 ± 0.07
		-	0.038 ± 0.001	0.020 ± 0.000	1.86 ± 0.05
	(0.847,0.976)	+	0.034 ± 0.001	0.005 ± 0.000	6.68 ± 0.68
		-	0.032 ± 0.001	0.006 ± 0.000	5.22 ± 0.49
(0.50,1.00)	(-0.908,-0.612)	+	0.005 ± 0.000	0.001 ± 0.000	3.69 ± 0.71
		-	0.001 ± 0.000	0.000 ± 0.000	2.35 ± 1.00
	(-0.612,-0.304)	+	0.027 ± 0.001	0.014 ± 0.000	2.00 ± 0.09
		-	0.022 ± 0.001	0.014 ± 0.000	1.53 ± 0.07
	(-0.304,0.001)	+	0.025 ± 0.001	0.006 ± 0.000	4.30 ± 0.43
		-	0.017 ± 0.001	0.006 ± 0.000	2.90 ± 0.20
	(0.001,0.427)	+	0.020 ± 0.000	0.004 ± 0.000	5.18 ± 0.22
		-	0.012 ± 0.000	0.004 ± 0.000	2.90 ± 0.20
	(0.427,0.847)	+	0.020 ± 0.000	0.008 ± 0.000	2.52 ± 0.06
		-	0.014 ± 0.000	0.008 ± 0.000	1.69 ± 0.05
	(0.847,0.976)	+	0.018 ± 0.004	0.003 ± 0.001	5.42 ± 1.92
		-	0.019 ± 0.003	0.004 ± 0.001	5.23 ± 1.08
(1.00,1.50)	(-0.908,-0.612)	+	0.007 ± 0.003	0.002 ± 0.001	3.01 ± 2.26

Table G.4. Pion to electron fake rates

$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
		–	0.001 ± 0.001	0.000 ± 0.001	2.04 ± 4.45
	$(-0.612, -0.304)$	+	0.038 ± 0.009	0.021 ± 0.007	1.76 ± 0.65
		–	0.030 ± 0.004	0.024 ± 0.007	1.22 ± 0.35
	$(-0.304, 0.001)$	+	0.044 ± 0.002	0.010 ± 0.001	4.28 ± 0.58
		–	0.033 ± 0.003	0.011 ± 0.002	2.95 ± 0.48
	$(0.001, 0.427)$	+	0.022 ± 0.002	0.003 ± 0.001	6.02 ± 1.31
		–	0.013 ± 0.001	0.003 ± 0.001	3.77 ± 0.96
	$(0.427, 0.847)$	+	0.019 ± 0.001	0.012 ± 0.000	1.65 ± 0.09
		–	0.013 ± 0.001	0.012 ± 0.001	1.05 ± 0.08
	$(0.847, 0.976)$	+	0.015 ± 0.003	0.003 ± 0.001	4.87 ± 1.85
		–	0.019 ± 0.003	0.003 ± 0.001	5.85 ± 1.90
$(1.50, 2.00)$	$(-0.908, -0.612)$	+	0.021 ± 0.007	0.007 ± 0.005	2.92 ± 2.78
		–	0.002 ± 0.002	0.002 ± 0.007	1.30 ± 2.51
	$(-0.612, -0.304)$	+	0.027 ± 0.012	0.019 ± 0.009	1.33 ± 0.87
		–	0.022 ± 0.008	0.017 ± 0.007	1.18 ± 0.67
	$(-0.304, 0.001)$	+	0.036 ± 0.010	0.004 ± 0.003	6.10 ± 3.53
		–	0.032 ± 0.010	0.013 ± 0.005	2.30 ± 1.22
	$(0.001, 0.427)$	+	0.022 ± 0.005	0.006 ± 0.002	3.28 ± 1.50
		–	0.015 ± 0.003	0.006 ± 0.002	2.58 ± 1.47
	$(0.427, 0.847)$	+	0.022 ± 0.007	0.017 ± 0.003	1.26 ± 0.43
		–	0.014 ± 0.005	0.013 ± 0.003	1.03 ± 0.43
	$(0.847, 0.976)$	+	0.019 ± 0.002	0.005 ± 0.001	3.36 ± 1.35
		–	0.020 ± 0.002	0.004 ± 0.002	4.12 ± 1.91
$(2.00, 2.50)$	$(-0.908, -0.612)$	+	0.015 ± 0.054	0.017 ± 0.013	1.00 ± 1.98
		–	0.022 ± 0.076	0.007 ± 0.012	1.67 ± 2.97

Table G.4. Pion to electron fake rates

$p_\pi[GeV]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
	(-0.612, -0.304)	+	0.023 ± 0.013	0.015 ± 0.007	1.70 ± 1.70
		-	0.015 ± 0.017	0.018 ± 0.013	0.92 ± 1.44
	(-0.304, 0.001)	+	0.031 ± 0.015	0.009 ± 0.009	3.40 ± 2.46
		-	0.020 ± 0.011	0.010 ± 0.009	1.98 ± 1.88
	(0.001, 0.427)	+	0.018 ± 0.008	0.007 ± 0.003	2.79 ± 2.12
		-	0.014 ± 0.005	0.008 ± 0.004	1.85 ± 1.68
	(0.427, 0.847)	+	0.028 ± 0.006	0.023 ± 0.004	1.15 ± 0.30
		-	0.016 ± 0.002	0.022 ± 0.003	0.67 ± 0.16
	(0.847, 0.976)	+	0.017 ± 0.004	0.011 ± 0.003	1.63 ± 1.15
		-	0.015 ± 0.004	0.007 ± 0.003	2.39 ± 1.82
(2.50, 3.00)	(-0.908, -0.612)	-	1.015 ± 0.488	0.854 ± 0.278	1.27 ± 0.89
		-	0.065 ± 0.124	0.038 ± 0.105	2.15 ± 3.54
	(-0.304, 0.001)	+	0.030 ± 0.172	0.012 ± 0.047	2.79 ± 5.27
		-	0.035 ± 0.033	0.016 ± 0.075	2.50 ± 3.98
	(0.001, 0.427)	+	0.020 ± 0.013	0.009 ± 0.009	2.50 ± 2.46
		-	0.016 ± 0.013	0.012 ± 0.007	1.75 ± 2.41
	(0.427, 0.847)	+	0.035 ± 0.008	0.016 ± 0.005	2.11 ± 1.15
		-	0.023 ± 0.008	0.024 ± 0.007	0.91 ± 0.45
	(0.847, 0.976)	+	0.022 ± 0.007	0.008 ± 0.006	2.70 ± 2.40
		-	0.023 ± 0.008	0.006 ± 0.004	3.32 ± 2.54

Table G.5. Pion to muon fake rates

$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
(0.40,0.50)	(-0.857, -0.605)	+	0.001 ± 0.000	0.002 ± 0.000	0.25 ± 0.17
		-	0.001 ± 0.002	0.002 ± 0.000	0.59 ± 0.88
	(-0.531, -0.208)	+	0.096 ± 0.002	0.088 ± 0.002	1.09 ± 0.03
		-	0.107 ± 0.002	0.085 ± 0.002	1.25 ± 0.04
	(-0.208, 0.111)	+	0.060 ± 0.001	0.056 ± 0.001	1.08 ± 0.03
		-	0.067 ± 0.002	0.055 ± 0.001	1.23 ± 0.05
	(0.111, 0.399)	+	0.063 ± 0.001	0.066 ± 0.001	0.95 ± 0.03
		-	0.071 ± 0.002	0.064 ± 0.001	1.12 ± 0.04
	(0.399, 0.682)	+	0.115 ± 0.002	0.132 ± 0.002	0.88 ± 0.02
		-	0.125 ± 0.003	0.126 ± 0.002	1.00 ± 0.02
	(0.682, 0.802)	+	0.137 ± 0.002	0.118 ± 0.003	1.16 ± 0.03
		-	0.145 ± 0.003	0.112 ± 0.003	1.28 ± 0.03
	(0.802, 0.921)	+	0.083 ± 0.002	0.029 ± 0.001	2.88 ± 0.09
		-	0.088 ± 0.002	0.028 ± 0.001	3.20 ± 0.11
(0.50,0.70)	(-0.857, -0.605)	+	0.014 ± 0.002	0.037 ± 0.002	0.39 ± 0.05
		-	0.013 ± 0.001	0.035 ± 0.002	0.37 ± 0.03
	(-0.605, -0.531)	+	0.001 ± 0.002	0.003 ± 0.001	0.16 ± 0.55
		-	0.001 ± 0.001	0.005 ± 0.001	0.16 ± 0.14
	(-0.531, -0.208)	+	0.119 ± 0.003	0.111 ± 0.002	1.07 ± 0.03
		-	0.140 ± 0.004	0.107 ± 0.002	1.31 ± 0.04
	(-0.208, 0.111)	+	0.073 ± 0.002	0.068 ± 0.002	1.07 ± 0.04
		-	0.088 ± 0.003	0.066 ± 0.001	1.35 ± 0.05
	(0.111, 0.399)	+	0.061 ± 0.001	0.066 ± 0.001	0.93 ± 0.03
		-	0.072 ± 0.002	0.065 ± 0.001	1.11 ± 0.03
	(0.399, 0.682)	+	0.085 ± 0.001	0.102 ± 0.001	0.83 ± 0.01

Table G.5. Pion to muon fake rates

$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
		–	0.099 ± 0.001	0.101 ± 0.001	0.98 ± 0.02
	(0.682,0.802)	+	0.132 ± 0.002	0.153 ± 0.001	0.86 ± 0.01
		–	0.154 ± 0.003	0.149 ± 0.001	1.03 ± 0.03
	(0.802,0.921)	+	0.067 ± 0.001	0.037 ± 0.001	1.80 ± 0.05
		–	0.075 ± 0.001	0.035 ± 0.001	2.11 ± 0.06
(0.70,1.00)	(–0.857,–0.605)	+	0.028 ± 0.001	0.056 ± 0.002	0.50 ± 0.03
		–	0.027 ± 0.001	0.054 ± 0.002	0.49 ± 0.03
	(–0.605,–0.531)	+	0.021 ± 0.002	0.040 ± 0.003	0.52 ± 0.07
		–	0.022 ± 0.003	0.046 ± 0.003	0.48 ± 0.07
	(–0.531,–0.208)	+	0.098 ± 0.002	0.091 ± 0.001	1.07 ± 0.03
		–	0.104 ± 0.002	0.088 ± 0.002	1.19 ± 0.03
	(–0.208,0.111)	+	0.080 ± 0.002	0.076 ± 0.001	1.05 ± 0.03
		–	0.080 ± 0.001	0.069 ± 0.002	1.16 ± 0.03
	(0.111,0.399)	+	0.060 ± 0.001	0.071 ± 0.002	0.84 ± 0.03
		–	0.060 ± 0.001	0.068 ± 0.001	0.89 ± 0.02
	(0.399,0.682)	+	0.066 ± 0.001	0.078 ± 0.001	0.85 ± 0.02
		–	0.064 ± 0.001	0.076 ± 0.001	0.84 ± 0.02
	(0.682,0.802)	+	0.090 ± 0.002	0.117 ± 0.002	0.77 ± 0.02
		–	0.099 ± 0.002	0.119 ± 0.002	0.83 ± 0.02
	(0.802,0.921)	+	0.063 ± 0.001	0.056 ± 0.001	1.11 ± 0.03
		–	0.067 ± 0.001	0.056 ± 0.001	1.20 ± 0.03
(1.00,1.50)	(–0.857,–0.605)	+	0.027 ± 0.009	0.044 ± 0.010	0.60 ± 0.23
		–	0.024 ± 0.008	0.041 ± 0.010	0.57 ± 0.20
	(–0.605,–0.531)	+	0.043 ± 0.020	0.051 ± 0.021	0.82 ± 0.46
		–	0.035 ± 0.015	0.045 ± 0.019	0.76 ± 0.43

Table G.5. Pion to muon fake rates

$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
	$(-0.531, -0.208)$	+	0.056 ± 0.003	0.049 ± 0.009	1.14 ± 0.19
		-	0.053 ± 0.003	0.048 ± 0.004	1.11 ± 0.11
	$(-0.208, 0.111)$	+	0.045 ± 0.002	0.043 ± 0.002	1.03 ± 0.06
		-	0.045 ± 0.002	0.043 ± 0.001	1.05 ± 0.05
	$(0.111, 0.399)$	+	0.040 ± 0.002	0.043 ± 0.002	0.94 ± 0.06
		-	0.038 ± 0.002	0.040 ± 0.002	0.93 ± 0.05
	$(0.399, 0.682)$	+	0.039 ± 0.001	0.043 ± 0.001	0.90 ± 0.03
		-	0.037 ± 0.001	0.040 ± 0.001	0.92 ± 0.04
	$(0.682, 0.802)$	+	0.044 ± 0.002	0.054 ± 0.002	0.81 ± 0.04
		-	0.040 ± 0.001	0.054 ± 0.002	0.74 ± 0.03
	$(0.802, 0.921)$	+	0.042 ± 0.002	0.044 ± 0.002	0.96 ± 0.05
		-	0.039 ± 0.002	0.042 ± 0.001	0.92 ± 0.05
(1.50, 2.00)	$(-0.857, -0.605)$	+	0.020 ± 0.006	0.021 ± 0.007	0.93 ± 0.51
		-	0.014 ± 0.006	0.031 ± 0.006	0.47 ± 0.26
	$(-0.605, -0.531)$	+	0.019 ± 0.013	0.021 ± 0.010	1.15 ± 1.23
		-	0.024 ± 0.009	0.020 ± 0.011	1.10 ± 1.00
	$(-0.531, -0.208)$	+	0.039 ± 0.014	0.039 ± 0.014	0.95 ± 0.47
		-	0.028 ± 0.010	0.032 ± 0.010	0.79 ± 0.33
	$(-0.208, 0.111)$	+	0.027 ± 0.007	0.025 ± 0.008	1.02 ± 0.40
		-	0.025 ± 0.007	0.025 ± 0.008	0.94 ± 0.38
	$(0.111, 0.399)$	+	0.023 ± 0.007	0.025 ± 0.007	0.88 ± 0.35
		-	0.024 ± 0.008	0.024 ± 0.006	0.96 ± 0.35
	$(0.399, 0.682)$	+	0.022 ± 0.004	0.025 ± 0.004	0.87 ± 0.19
		-	0.022 ± 0.004	0.022 ± 0.005	0.98 ± 0.27
$(0.682, 0.802)$	+	0.030 ± 0.004	0.028 ± 0.003	1.03 ± 0.18	

Table G.5. Pion to muon fake rates

$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
	(0.802,0.921)	-	0.025 ± 0.005	0.025 ± 0.003	0.94 ± 0.25
		+	0.024 ± 0.002	0.026 ± 0.002	0.91 ± 0.13
		-	0.022 ± 0.002	0.024 ± 0.002	0.91 ± 0.14
(2.00,2.50)	(-0.857,-0.605)	+	0.024 ± 0.012	0.016 ± 0.010	1.74 ± 2.40
		-	0.056 ± 0.156	0.070 ± 0.093	0.84 ± 2.24
	(-0.531,-0.208)	+	0.051 ± 0.020	0.033 ± 0.021	1.51 ± 1.01
		-	0.022 ± 0.019	0.038 ± 0.014	0.61 ± 0.53
	(-0.208,0.111)	+	0.036 ± 0.014	0.026 ± 0.013	1.29 ± 0.83
		-	0.038 ± 0.007	0.024 ± 0.007	1.48 ± 0.68
	(0.111,0.399)	+	0.022 ± 0.014	0.037 ± 0.012	0.56 ± 0.40
		-	0.021 ± 0.009	0.029 ± 0.009	0.68 ± 0.37
	(0.399,0.682)	+	0.032 ± 0.009	0.042 ± 0.010	0.71 ± 0.28
		-	0.021 ± 0.007	0.031 ± 0.004	0.68 ± 0.29
	(0.682,0.802)	+	0.023 ± 0.004	0.024 ± 0.004	0.90 ± 0.33
		-	0.024 ± 0.004	0.019 ± 0.005	1.17 ± 0.49
(0.802,0.921)	+	0.019 ± 0.003	0.031 ± 0.007	0.53 ± 0.15	
	-	0.022 ± 0.004	0.029 ± 0.005	0.70 ± 0.23	
(2.50,3.50)	(-0.208,0.111)	+	0.015 ± 0.040	0.021 ± 0.116	0.78 ± 1.84
		-	0.018 ± 0.034	0.041 ± 0.030	0.57 ± 1.24
	(0.111,0.399)	+	0.024 ± 0.016	0.045 ± 0.063	0.54 ± 0.56
		-	0.015 ± 0.011	0.028 ± 0.014	0.64 ± 0.80
	(0.399,0.682)	+	0.040 ± 0.011	0.047 ± 0.012	0.77 ± 0.32
		-	0.037 ± 0.008	0.043 ± 0.009	0.79 ± 0.30
	(0.682,0.802)	+	0.026 ± 0.010	0.032 ± 0.010	0.72 ± 0.39
		-	0.020 ± 0.011	0.025 ± 0.008	0.87 ± 0.62

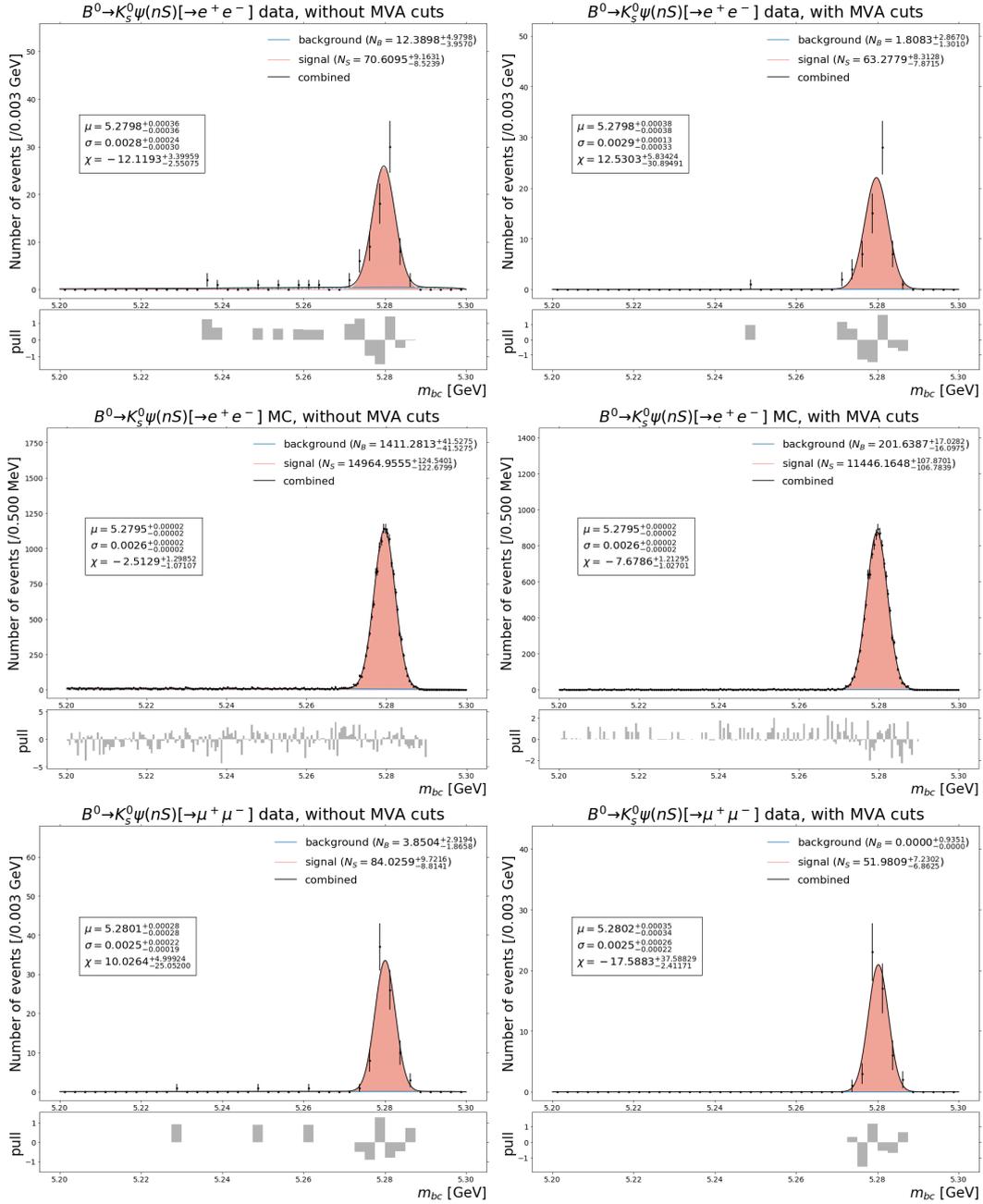
Table G.5. Pion to muon fake rates

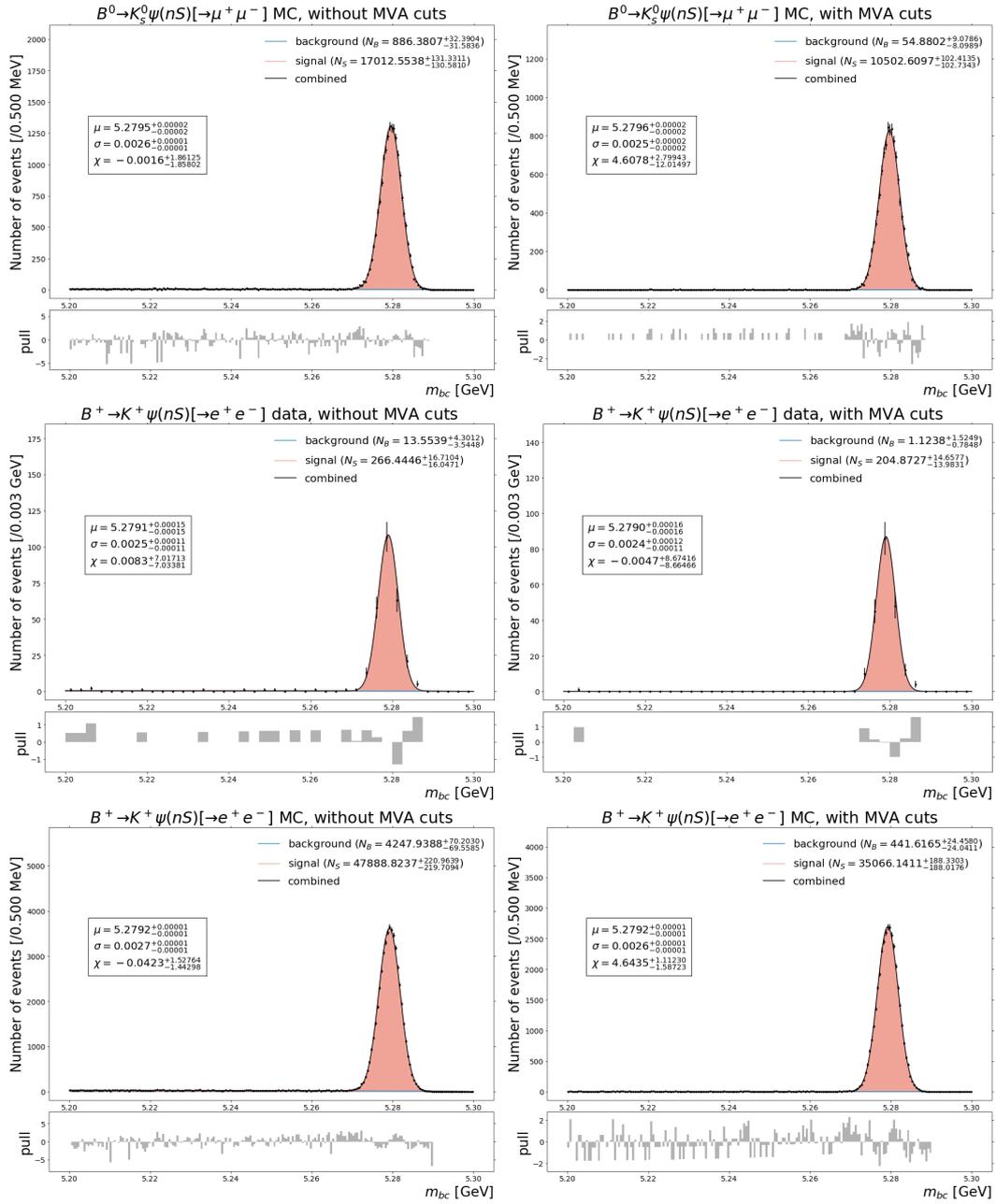
$p_\pi[\text{GeV}]$	$\cos\theta_\pi$	charge	fake rate (data)	fake rate (MC)	fake rate ratio
	(0.802,0.921)	+	0.017 ± 0.005	0.036 ± 0.007	0.44 ± 0.19
		-	0.022 ± 0.005	0.033 ± 0.007	0.66 ± 0.28

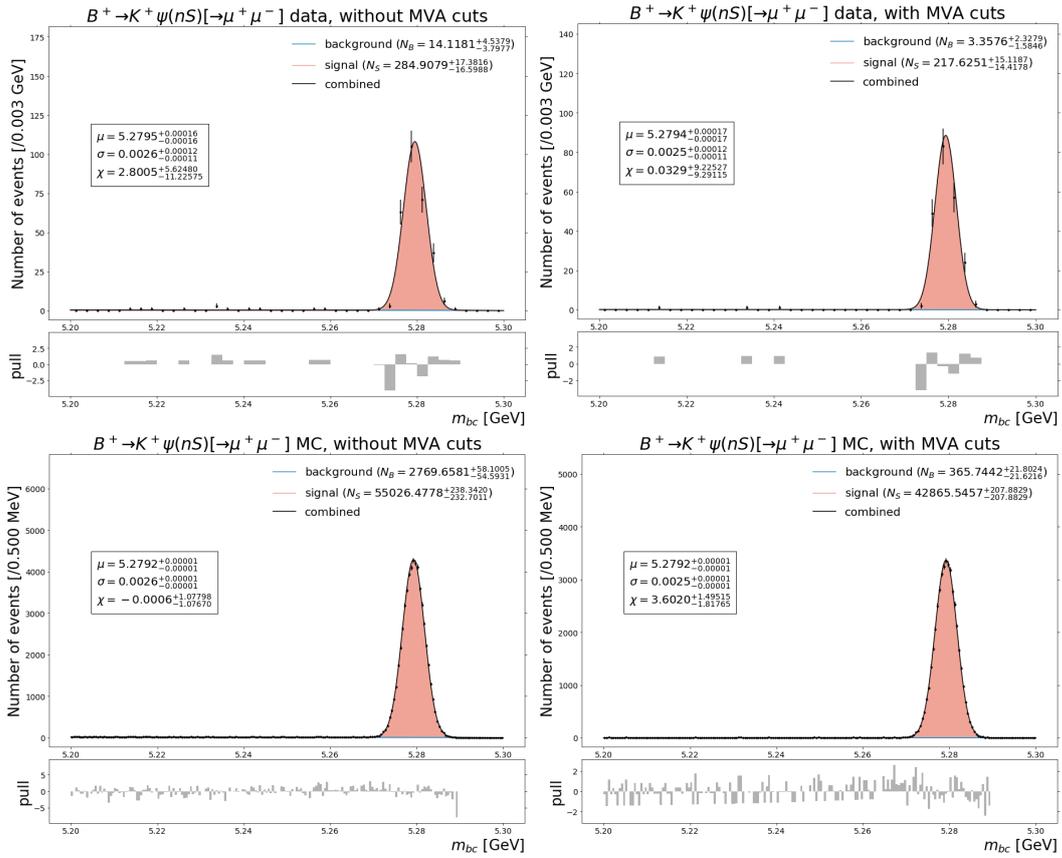
Appendix H

Fits to MC and to Experimental Data Inside the Charmonium Sidebands

This appendix shows the fits to the m_{bc} distributions of the $B \rightarrow K\psi(nS) [\rightarrow \ell^+\ell^-]$ decays, from which the signal yields are obtained, and the efficiency correction and systematic uncertainties calculated.



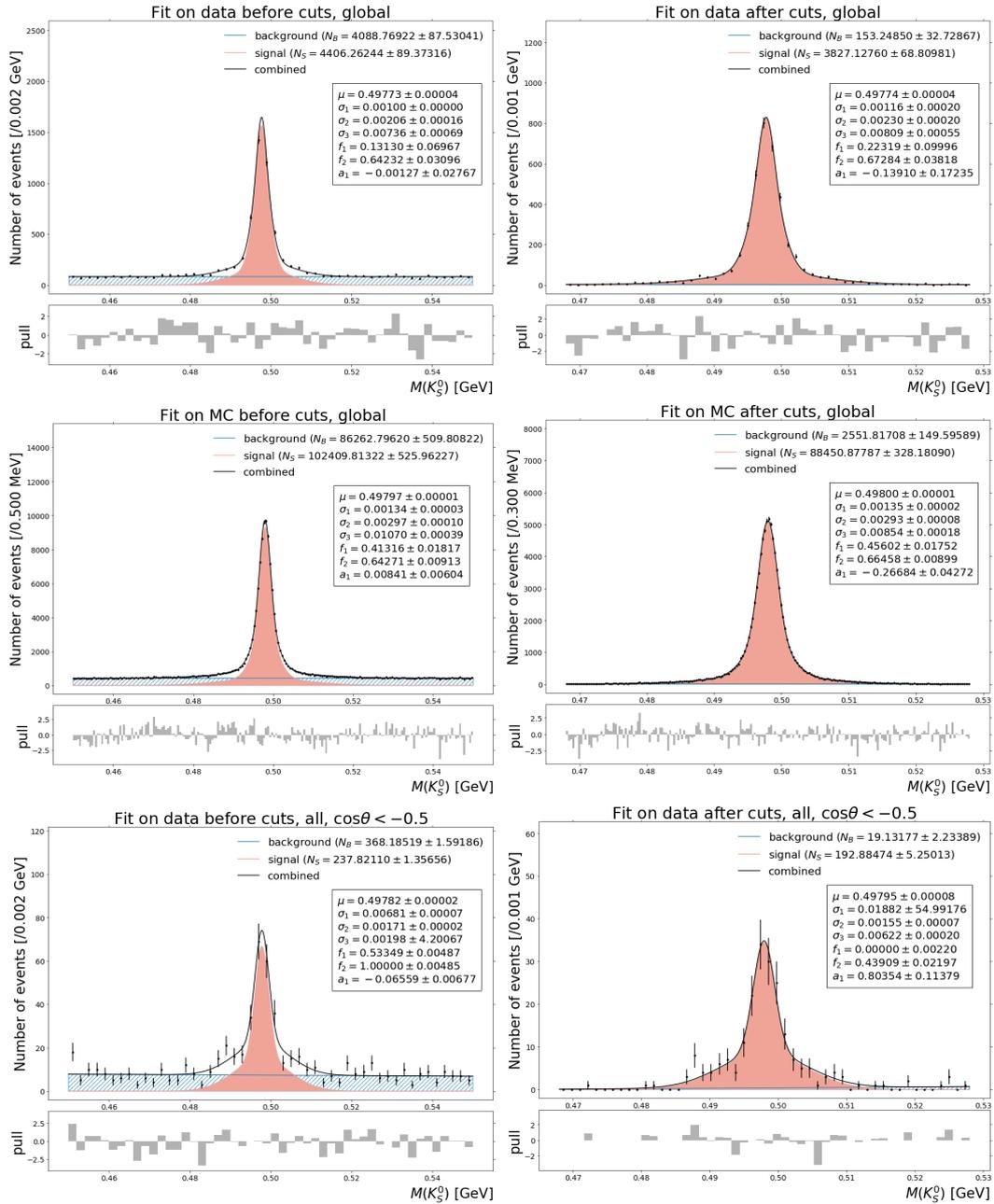


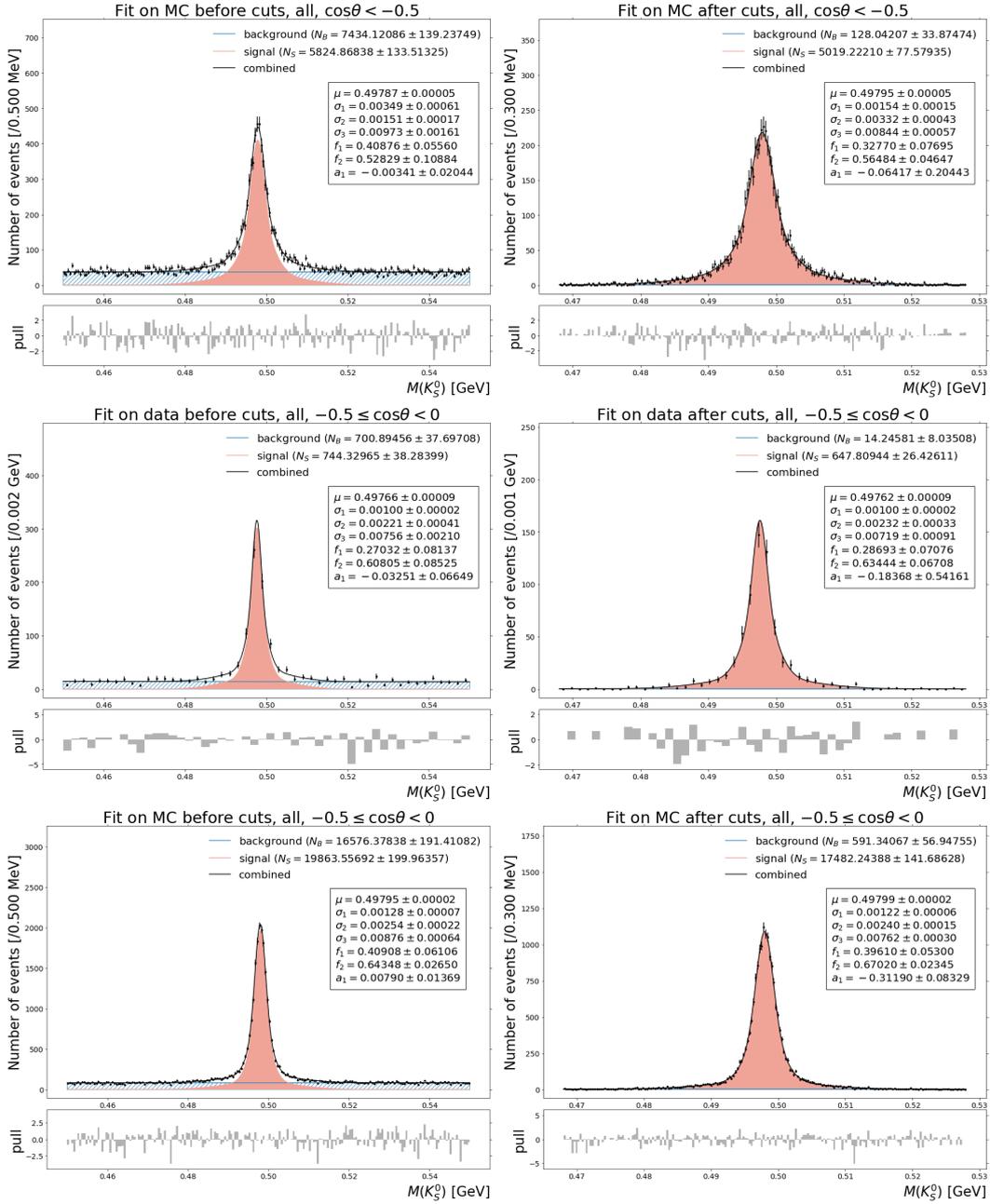


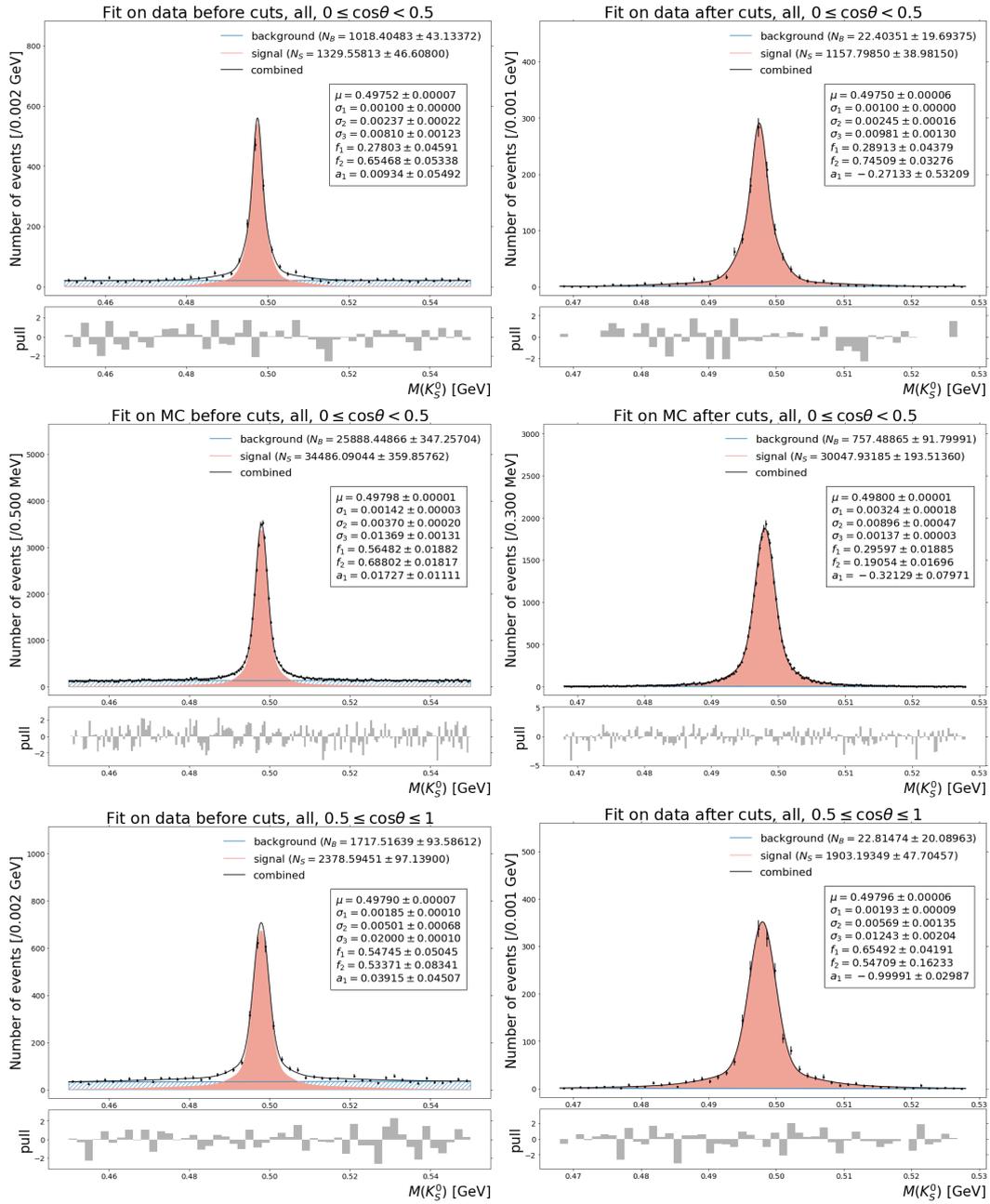
Appendix I

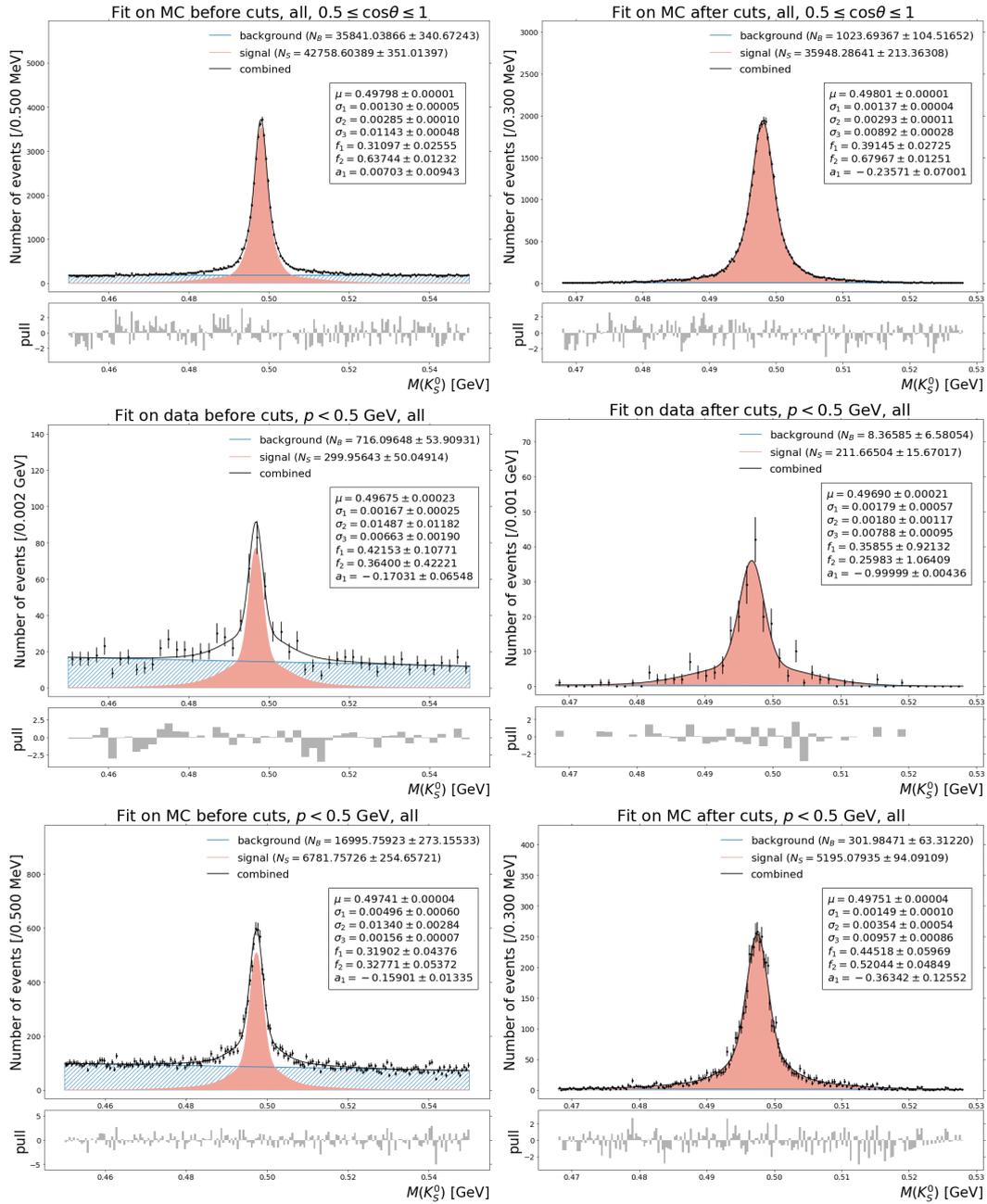
Fit Results for the K_S^0 Systematics

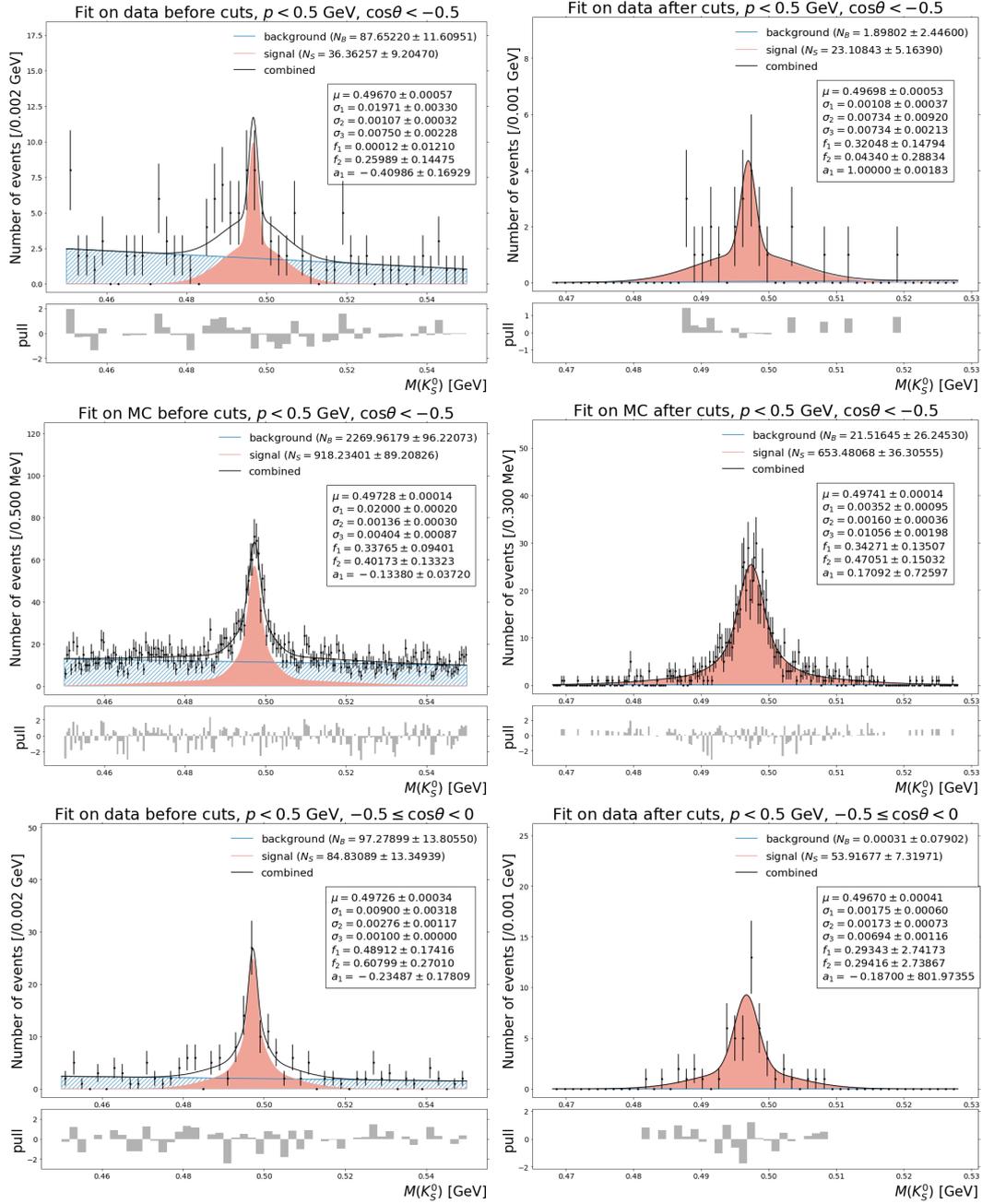
Next the fit results for the K_S^0 systematics study are shown. The MC data corresponds to 200 fb^{-1} of $B\bar{B}$ and $q\bar{q}$ events, using as cross sections the values presented in table 3.2; the experimental data corresponds to the whole 11.53 fb^{-1} .

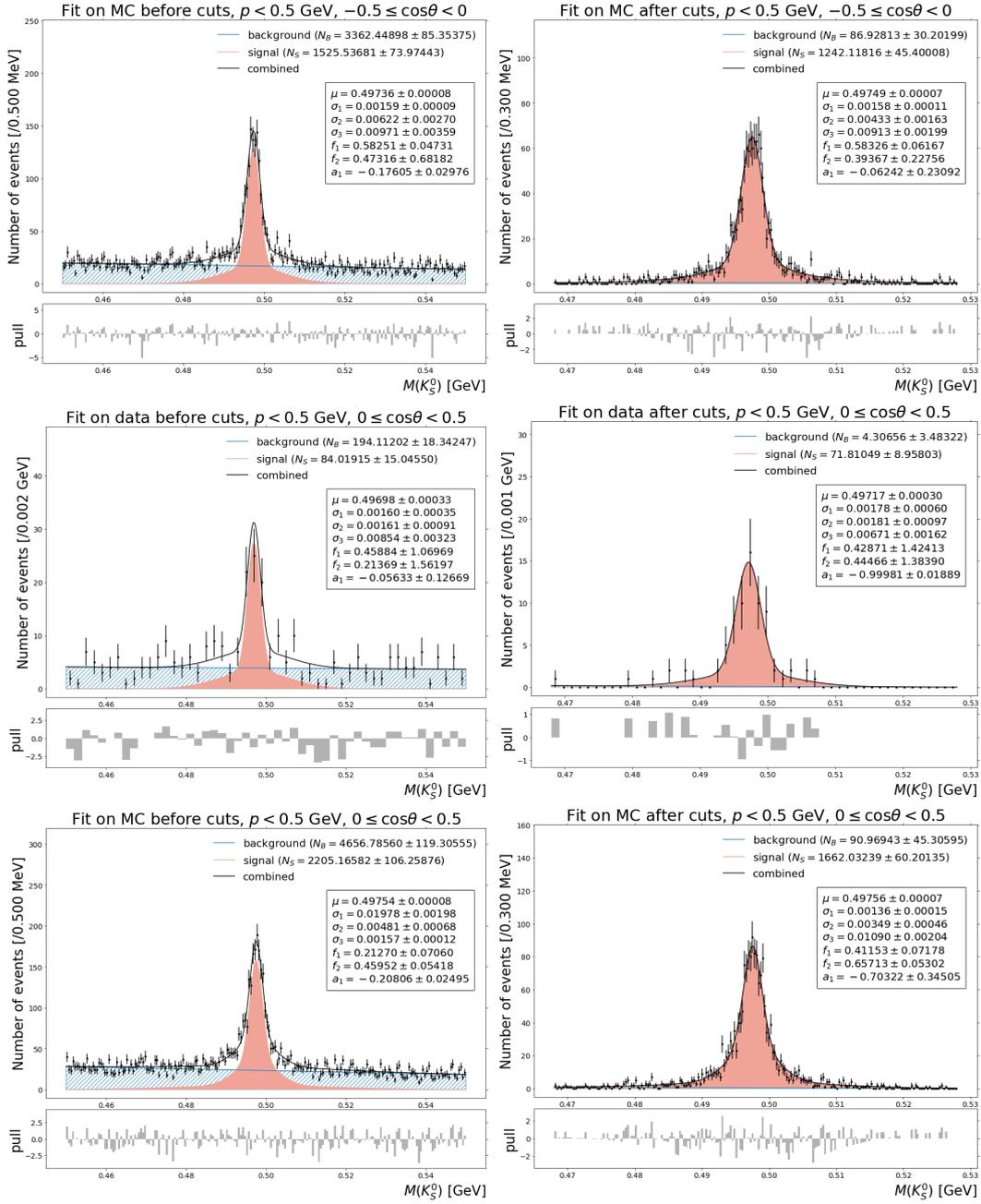


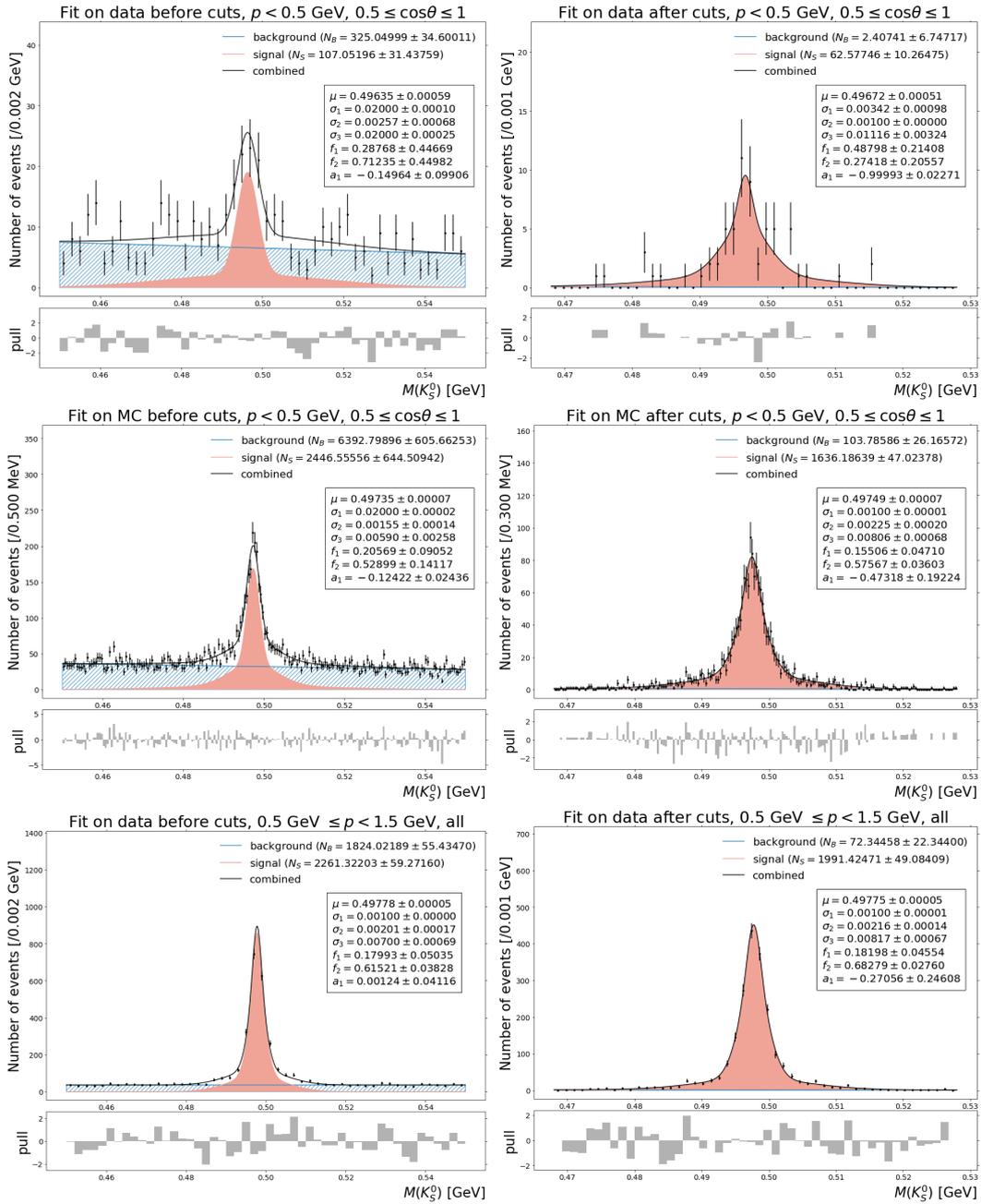


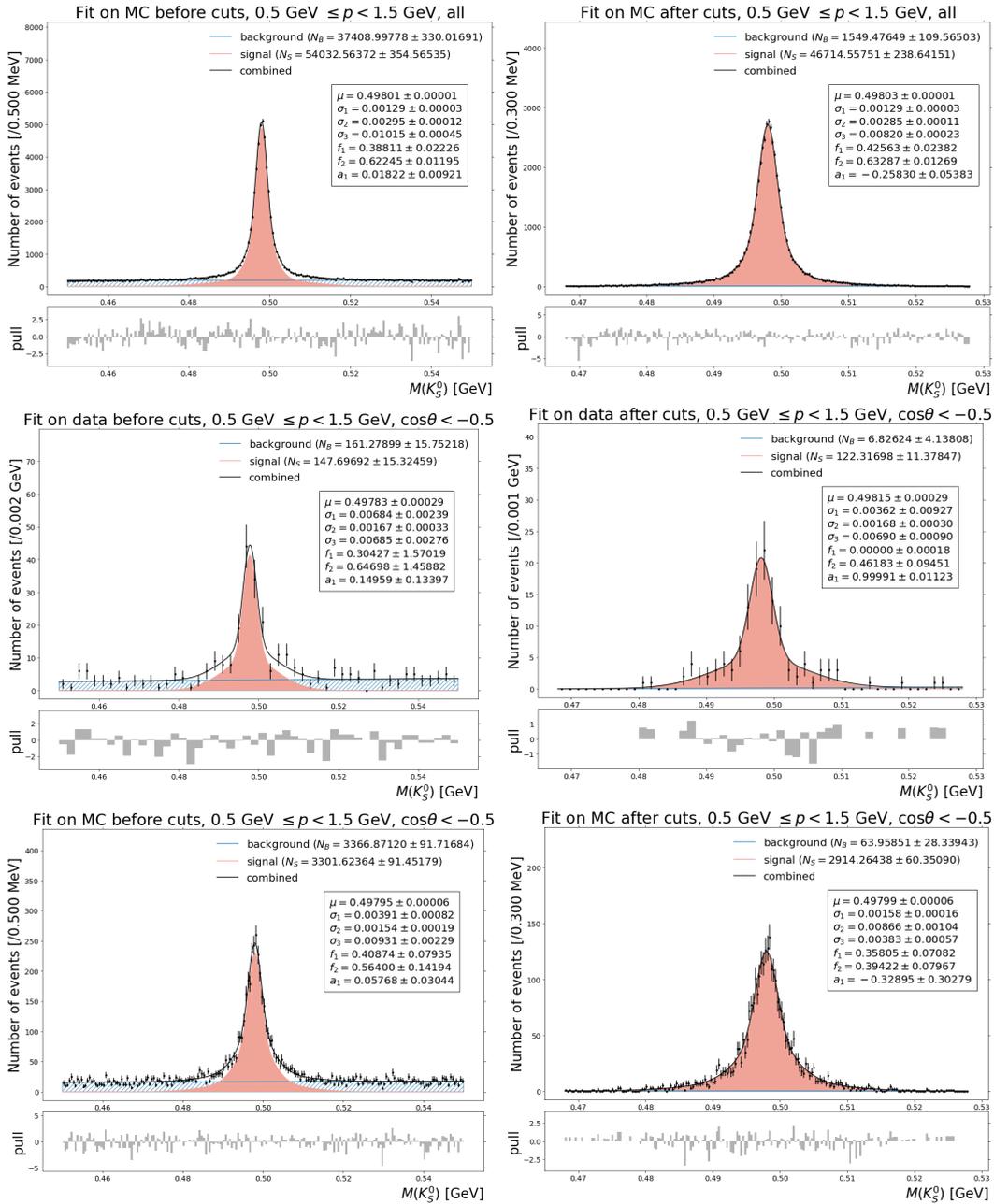


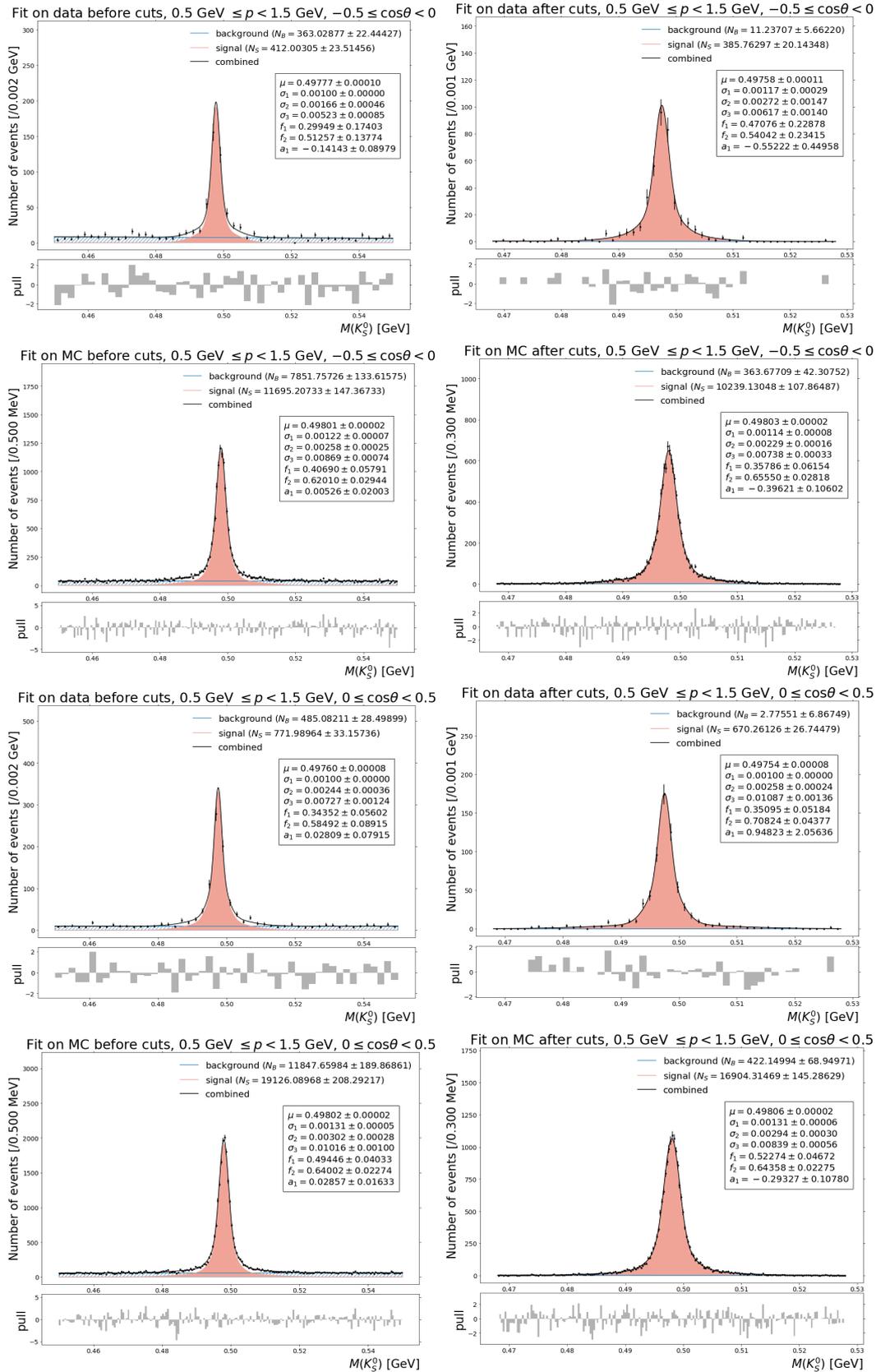


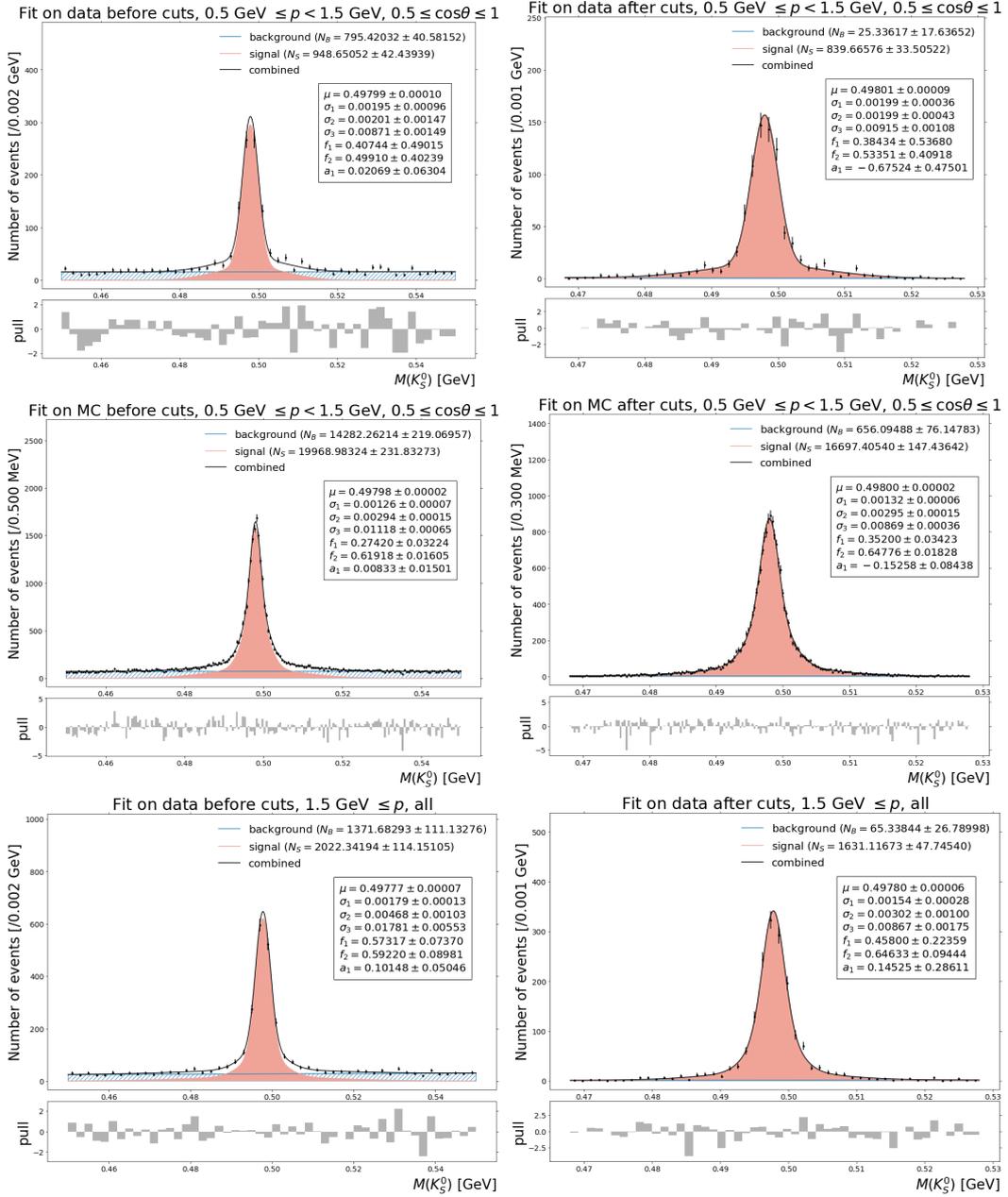


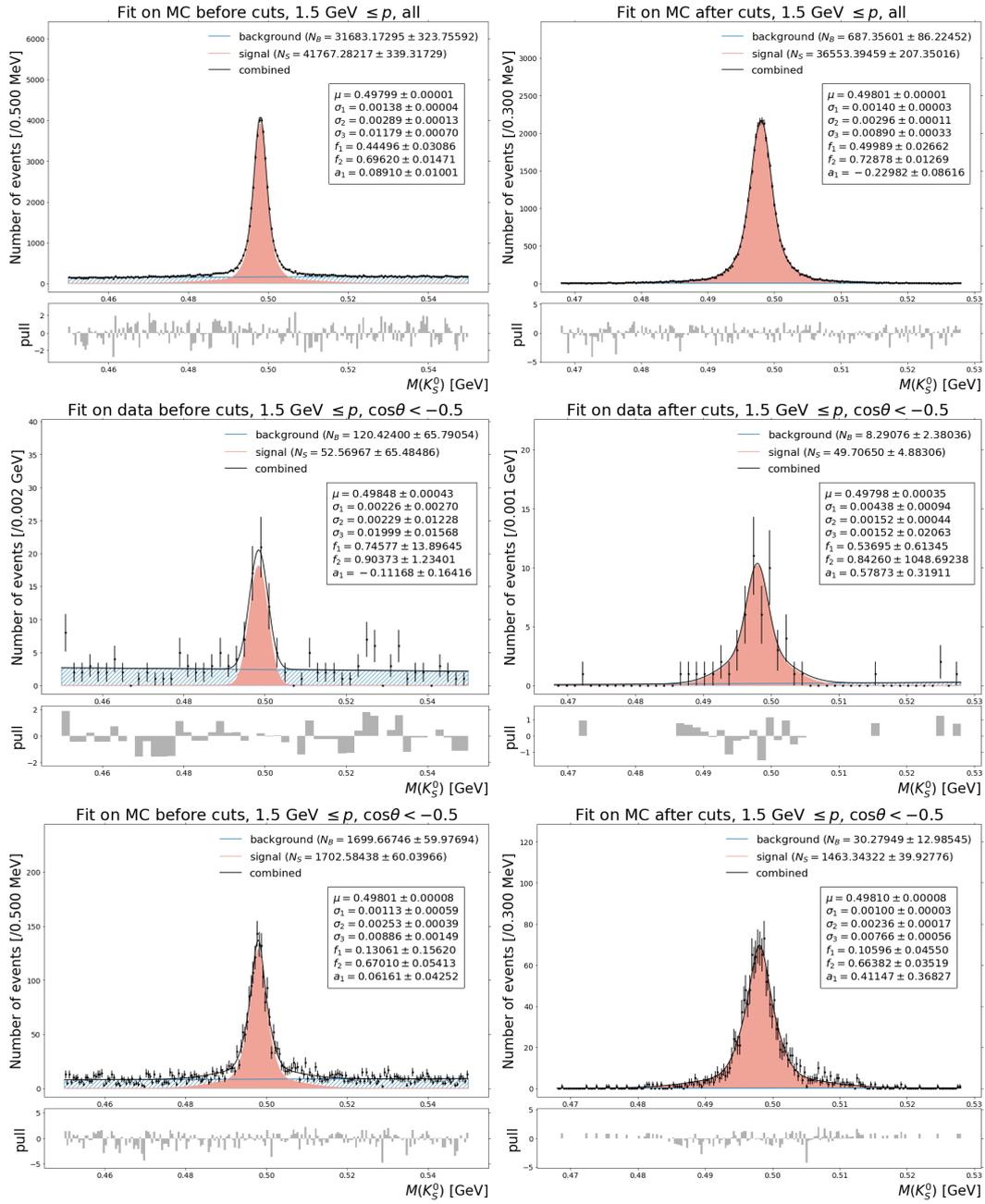


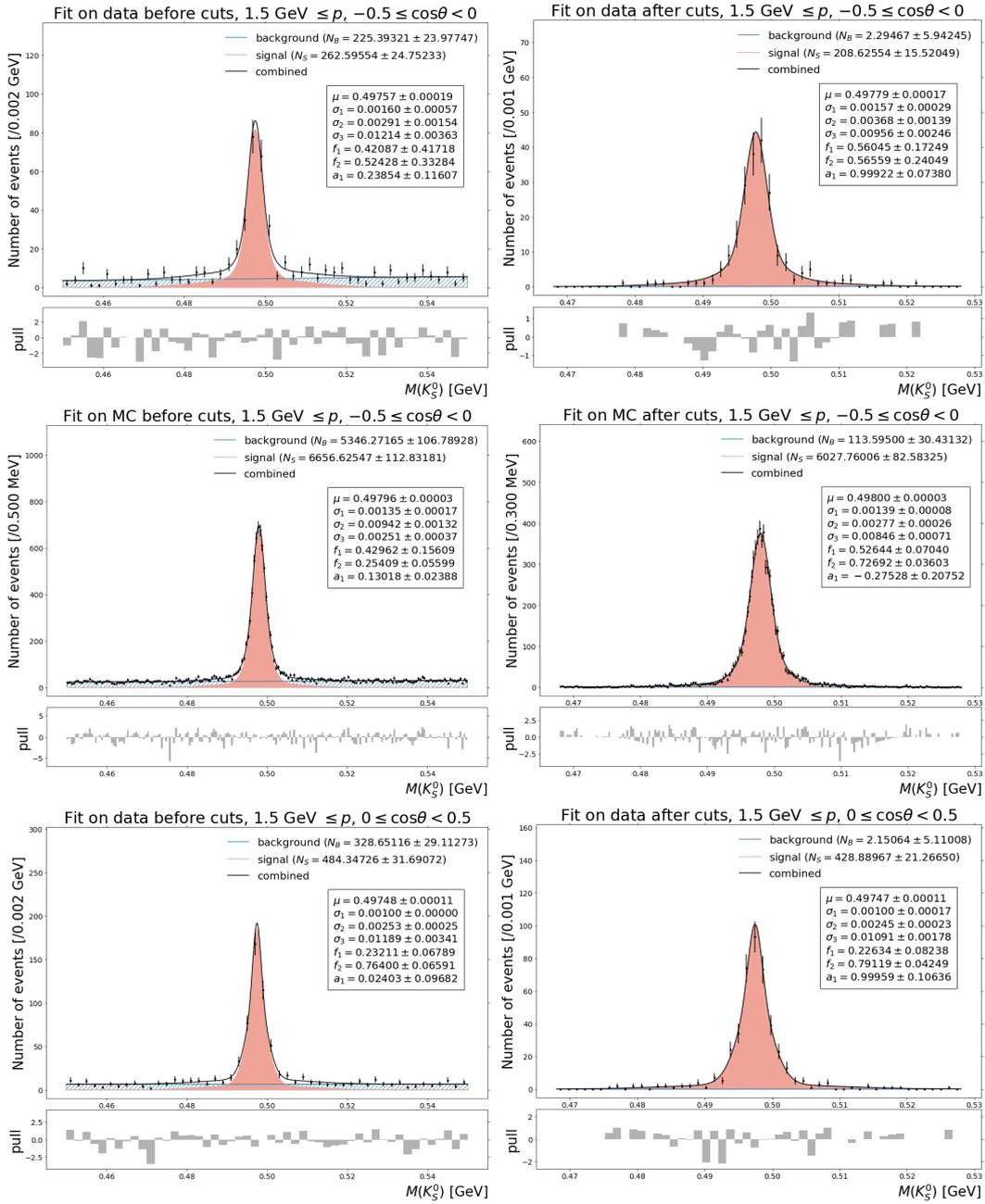


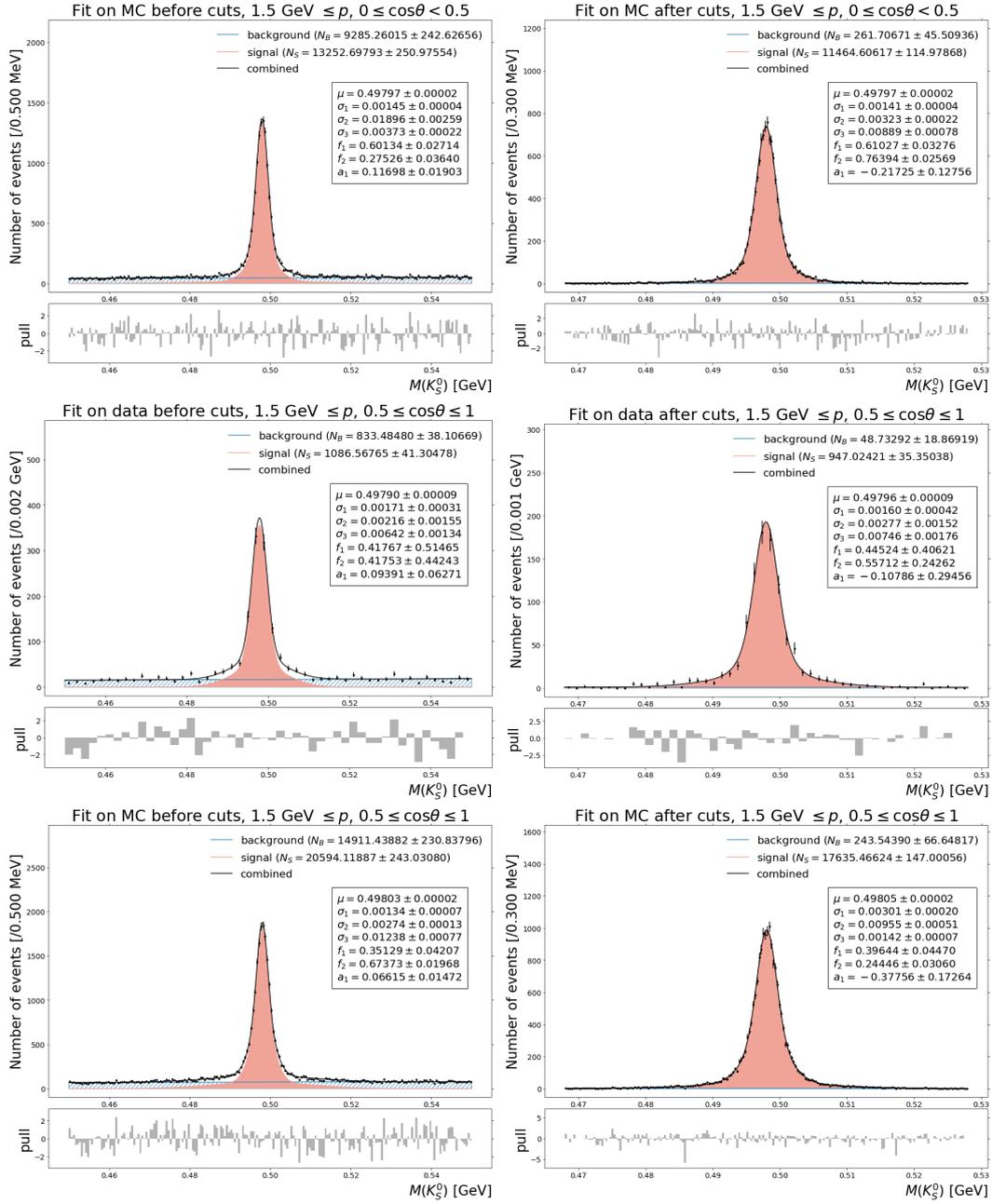












Appendix J

The Extended Maximum Likelihood Technique

In this appendix we briefly discuss the technique used throughout the study to obtain the parameters of a fit, known as the extended maximum likelihood technique.

J.1 Likelihood fit

The concept of likelihood was mentioned in appendix E, in the context of function approximations through weak learners. As then, we start by assuming that we perform a set of identical and statistically independent observations $\{x_1, \dots, x_n\}$ of a variable x . From these observations, we want to derive the probability density function for the observable x . Whereas with Boosted Decision Trees (and other MVA methods) no strong assumption over the shape of the PDF is made, in the case of a likelihood an ansatz for the type of function is provided. This is, we assume that the observable x is well described by a function $f(x|\theta_1, \dots, \theta_m)$ that depends on a series of parameters θ_i whose values we intend to determine. A logical approach is to look for those values $\hat{\theta}_i$ that maximize the quantity

$$\mathcal{L}(\theta_1, \dots, \theta_m) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_m). \quad (\text{J.1})$$

\mathcal{L} is known as the likelihood. We might as well define the problem as a minimization one, and simplify the product by introducing a logarithm. Hence, the final objective is the minimization of the expression

$$-\log \mathcal{L}(\theta_1, \dots, \theta_m) = -\sum_{i=1}^n \log f(x_i|\theta_1, \dots, \theta_m). \quad (\text{J.2})$$

Once the minimization is done (through a numerical or an exact procedure), one can expand the log-likelihood around its minimum:

$$-\log \mathcal{L}(\theta_1, \dots, \theta_m) \approx -\log \mathcal{L}(\hat{\theta}_1, \dots, \hat{\theta}_m) + \frac{1}{2} \sum_{i,j=1}^m \left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i \partial \theta_j} \right] (\theta_i - \hat{\theta}_i) (\theta_j - \hat{\theta}_j) \quad (\text{J.3})$$

and notice that, the likelihood follows a multidimensional Gaussian profile¹:

$$\mathcal{L}(\theta_1, \dots, \theta_m) \sim \exp \left(-\Delta \theta^T \mathbf{M} \Delta \theta \right) \quad (\text{J.4})$$

where $\Delta \theta^T = (\theta_1 - \hat{\theta}_1, \dots, \theta_m - \hat{\theta}_m)$ and \mathbf{M} is the $m \times m$ matrix:

$$\mathbf{M} = \begin{pmatrix} \frac{\partial^2 \log \mathcal{L}}{\partial^2 \theta_1} & \frac{\partial^2 \log \mathcal{L}}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \log \mathcal{L}}{\partial \theta_1 \partial \theta_m} \\ \frac{\partial^2 \log \mathcal{L}}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \log \mathcal{L}}{\partial^2 \theta_2} & \dots & \frac{\partial^2 \log \mathcal{L}}{\partial \theta_2 \partial \theta_m} \\ \vdots & \ddots & \dots & \vdots \\ \frac{\partial^2 \log \mathcal{L}}{\partial \theta_m \partial \theta_1} & \frac{\partial^2 \log \mathcal{L}}{\partial \theta_m \partial \theta_2} & \dots & \frac{\partial^2 \log \mathcal{L}}{\partial^2 \theta_m} \end{pmatrix}. \quad (\text{J.5})$$

It is then possible to numerically derive the error in the fit for the parameter θ_i , σ_i , by noting that

$$-\log \mathcal{L}(\theta_1, \dots, \hat{\theta}_i + \sigma_i, \dots, \hat{\theta}_m) + \log \mathcal{L}(\hat{\theta}_1, \dots, \hat{\theta}_i, \dots, \hat{\theta}_m) \approx \frac{1}{2} \left[\frac{\partial^2 \ln \mathcal{L}}{\partial \theta_i^2} \right] \sigma_i^2 = \frac{1}{2}. \quad (\text{J.6})$$

The strategy is then straightforward: keep varying the value of the parameter until the log-likelihood value increases by 0.5 above its minimum.

J.2 Extended likelihood

It is often the case that, in a study, one wants to determine the absolute rate of a process, and not only its probability. The concept of extended likelihood incorporates the information pertaining these rates, by introducing the probability of observing n events given an expected rate of ν :

$$\mathcal{L}(\theta_1, \dots, \theta_m, \nu) = \frac{\nu^n e^{-\nu}}{n!} \prod_{i=1}^n f(x_i | \theta_1, \dots, \theta_m). \quad (\text{J.7})$$

¹This is no longer true if one includes higher terms in the expansion; this results in an asymmetric profile, which is the cause of the asymmetrical errors in the parameters.

J.3 Significance

In chapter 6 the statistical significance of a fit under two hypotheses (h_0 and h_1) is defined as

$$S = \sqrt{-2 \log \left(\frac{\mathcal{L}_0}{\mathcal{L}_1} \right)} \quad (\text{J.8})$$

where \mathcal{L}_0 and \mathcal{L}_1 are the minimized likelihoods for the respective hypothesis. Since the alternative hypothesis h_1 has more free parameters than the null one h_0 , it will always fit at least as well as h_0 . Thus, the ratio inside the logarithm has an upper bound of one, and S is larger than or equal to zero. Moreover, by Wilks' theorem, as the sample size $n \rightarrow \infty$, under the null hypothesis, the significance follows a χ^2 distribution in D dimensions, where D is the difference in the number of free parameters between h_0 and h_1 ; in other words, if h_0 is true, then \mathcal{L}_1 cannot differ from \mathcal{L}_0 by a fraction larger than the statistical uncertainties. The meaning of the statistical significance is then clear: it is a statistic that, in the limit of large sample sizes, behaves as a χ^2 variable, and can be used to measure the probability of the alternate hypothesis to be wrongly accepted when the null one is true.

From the expression of the extended likelihood, the significance can be written as

$$S = \sqrt{2 \left[(v_0 - v_1) + \sum_{i=1}^n \log \left(\frac{v_1 p_1(x_i)}{v_0 p_0(x_i)} \right) \right]} \quad (\text{J.9})$$

where v_k and p_k are the expected rate and the PDF for hypothesis k . Notice that both terms inside the squared brackets are extensive, meaning they scale with the number of events. It is then expected for the significance to scale with the square root of the number of events. Since in an accelerator the number of experiments is proportional to the luminosity, we expect S to scale with the square root of the luminosity.

Appendix K

Extra Fits

This chapter presents the plots for the fits mentioned in the main corpus of this work, which do not appear there.

K.1 Non-peaking backgrounds

These fits were performed to obtain the shape parameters of the non-peaking background component of the final PDF for their respective decays from 2 ab^{-1} generic MC background events where at least one lepton is properly identified.

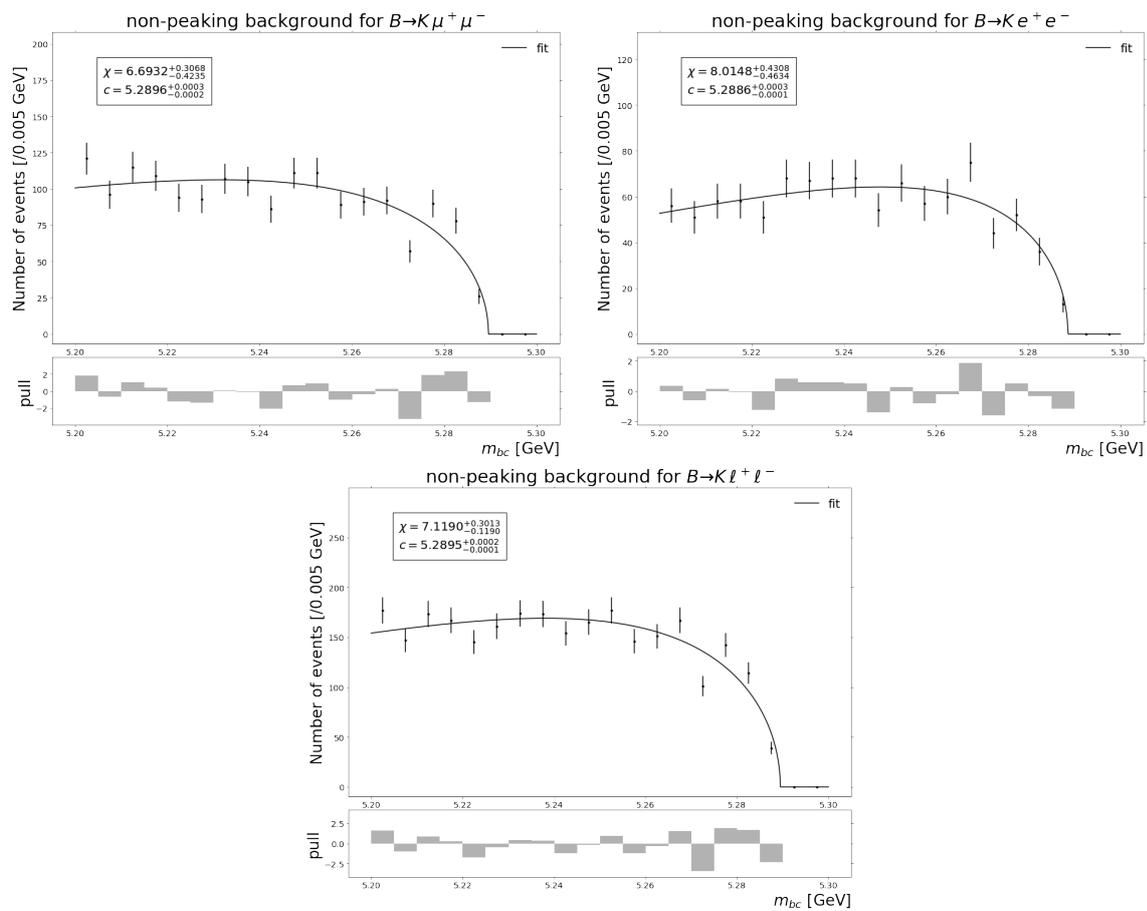


Fig. K.1 Extra m_{bc} distributions and fits for the non-peaking backgrounds from MC.

K.2 Leaked charmonium backgrounds

These fits were performed to obtain the shape parameters and yields of the peaking background originating from charmonium resonances making it through the veto from dedicated Monte Carlo simulations of the decays $B \rightarrow K\psi(nS) [\rightarrow \ell^+ \ell^-]$.

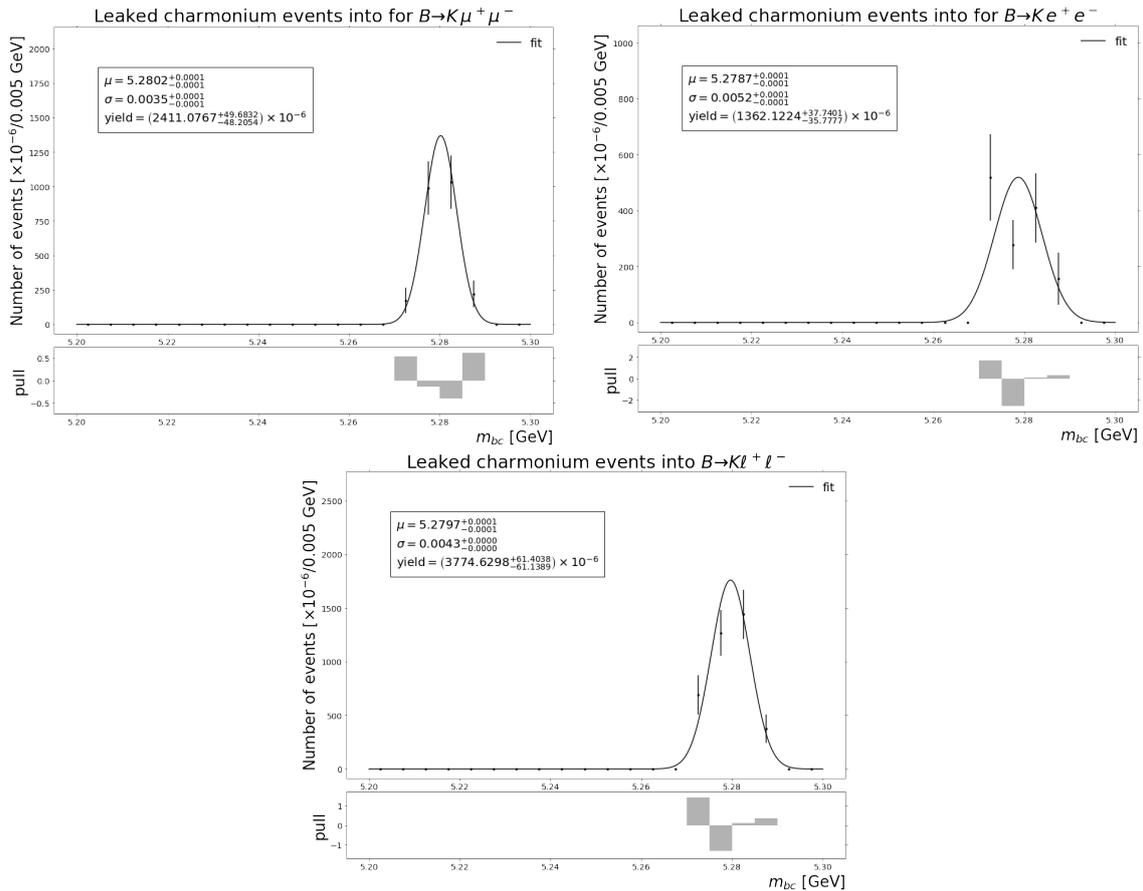


Fig. K.2 Extra m_{bc} distributions and fits for the leaking charmonium component of the background, from dedicated MC samples.

K.3 $K\pi^+\pi^-$ backgrounds

These fits were performed to obtain the shape parameters and yields of the peaking background originating from $B \rightarrow K\pi^+\pi^-$ where both pions are misidentified as leptons from experimental data where all the selection cuts are applied, except for the lepton ID ones, which are reversed. These events were then weighted by the fake rate for pions in Monte Carlo.

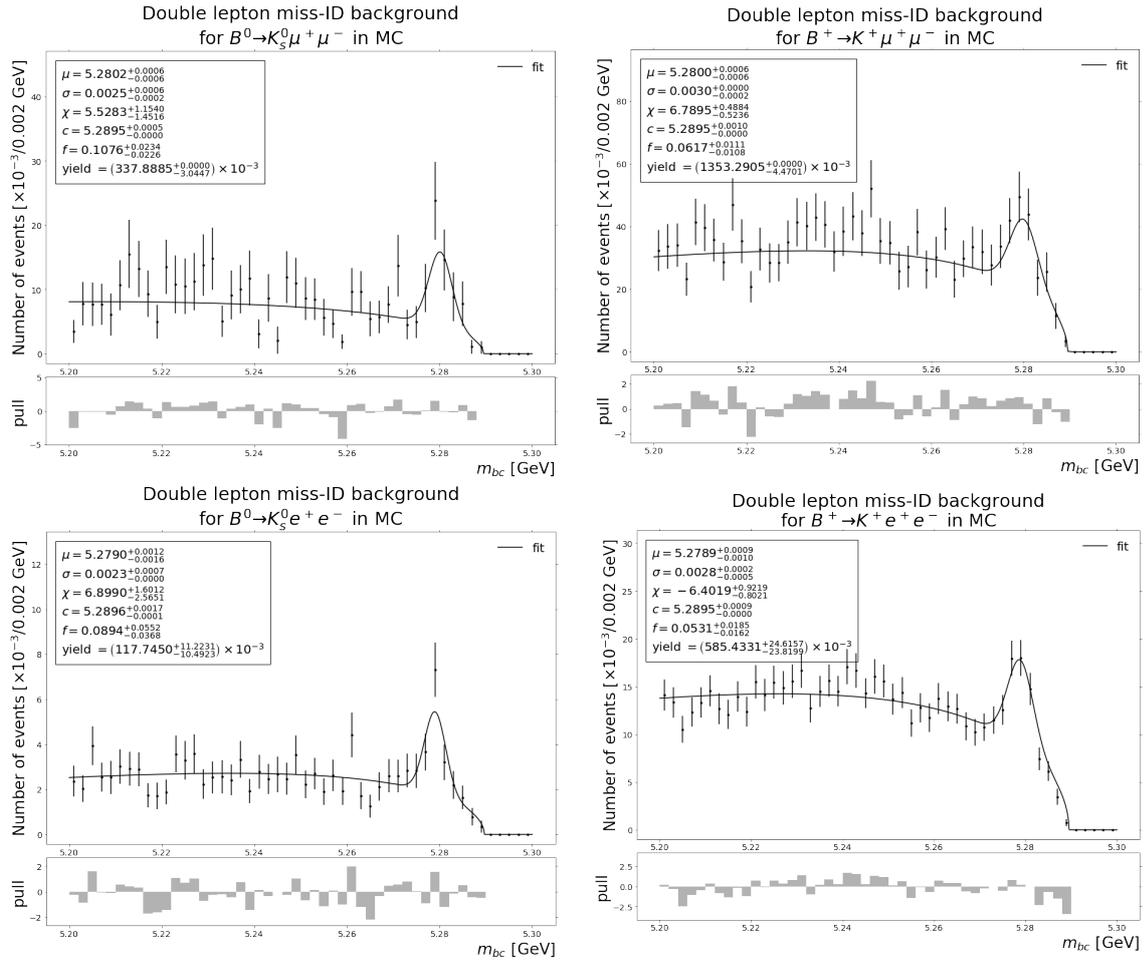


Fig. K.3 m_{bc} distributions and fits for the $K\pi^+\pi^-$ component of the background, using the pion-to-lepton fake rates in MC.

K.4 Signal

These fits were performed to obtain the shape parameters of the signal component of the final PDF from 11.53 fb^{-1} of experimental data where all the selection cuts are applied, except for the charmonium vetoes, which are reversed.

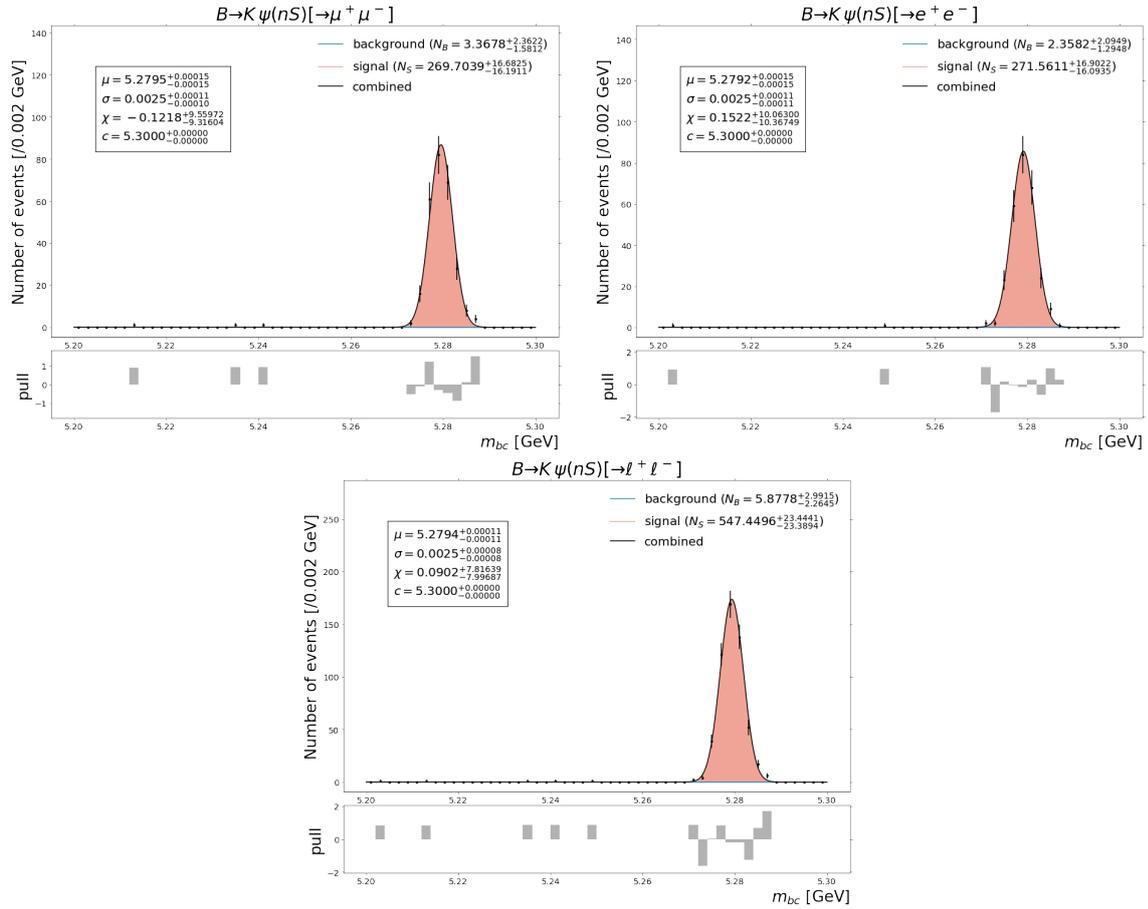


Fig. K.5 m_{bc} distributions and fits for the signal component, using data from the charmonium sidebands.

K.5 Combined

These fits were performed on the generic MC data after applying the reconstruction cuts, in order to determine the significance of the fit at 2 ab^{-1} .

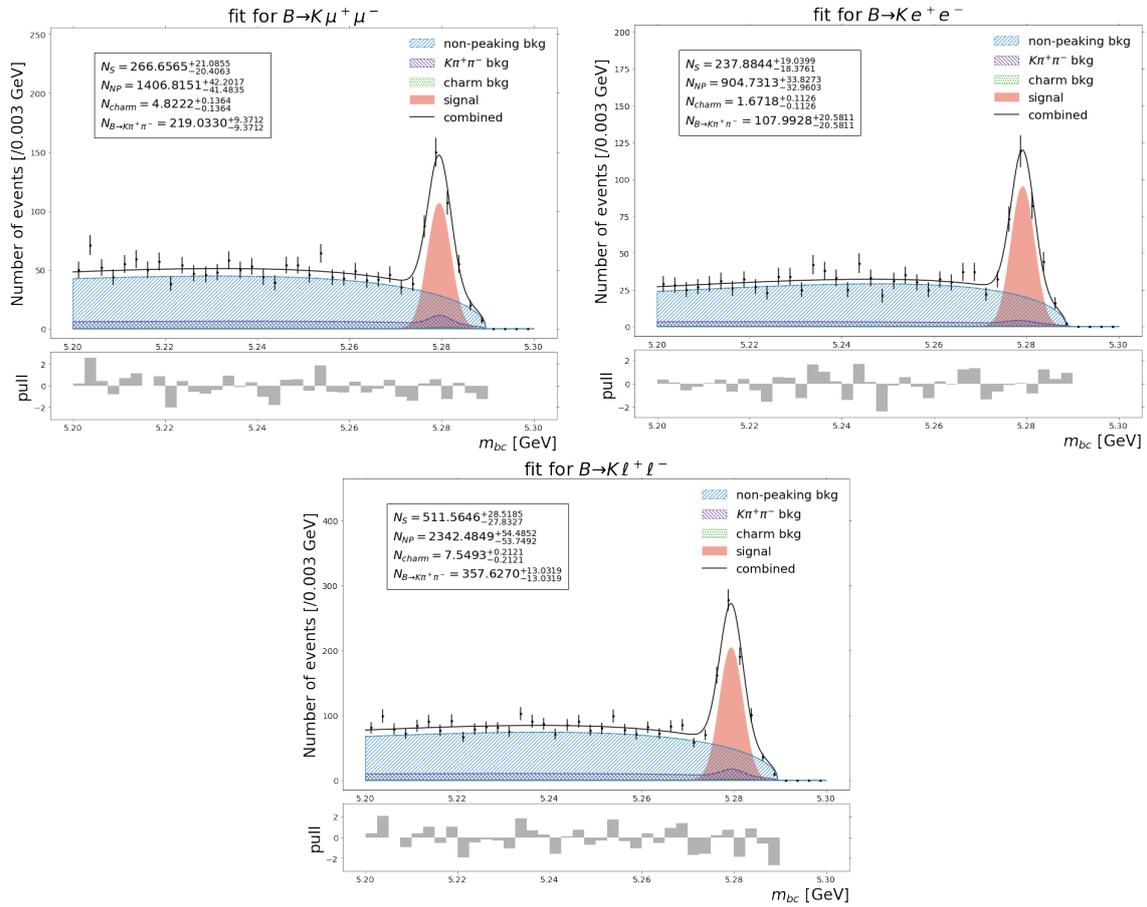


Fig. K.6 Extra m_{bc} distributions and fits for multiple combinations of the decays of interests, in MC.

Appendix L

Systematic Errors for the Peaking Background Yields

This appendix resumes the procedure for the calculation of the errors in the expected yields for the peaking background components (charmonium resonances and $K\pi^+\pi^-$) events.

Each event in the final sample is weighted by a factor w : in the charmonium background, this weight is basically the expected number of decays at a luminosity of 1 fb^{-1} , which depends on the type of resonance, the flavor of the decaying B meson and (in a much weaker way) in the flavor of the lepton pair; for the $K\pi^+\pi^-$, the weight is the pion to lepton fake rate, which depends on the charge, flavor, momentum and angle of the lepton. All weights carry an error: for the charmonium resonances, this error is obtained by propagating the error in the $ee \rightarrow B\bar{B}$ cross section, in the fraction of B^+B^- to $B^0\bar{B}^0$ pairs produced, and in all the branching fractions. For the double lepton miss-id background, it is the sum of squares of the statistical and systematic error of the weights, obtained in [72].

In order to propagate these errors, we perform $M = 1000$ simulations, in which we obtain new effective weights associated from a multivariate normal distribution. This distribution has a mean vector containing the corresponding weight values, and a diagonal covariance matrix where each non-zero entry corresponds to the inverse of the squared error of each weight. For each one of these simulations, the yield is calculated as the sum of these effective weights, and we assign a systematic error equal to the standard deviation of these 1000 yields.

Since the statistical error in the weights for $K\pi^+\pi^-$ decays is large, there is room for improvement from larger data samples. Hence, we also calculate the systematic error in the yield for $K\pi^+\pi^-$ decays at a luminosity of 50 ab^{-1} by rescaling the statistical error in the weights and repeating the previous procedure.

The total error in the yield is then taken as the sum of squares of this systematic error, and the statistical error obtained from the fits. For the $B \rightarrow K\pi^+\pi^-$, since the luminosity

used for normalization is the one from the experimental dataset, we also propagate its error into the yield error.

Appendix M

Fits to the MC Data Using the Belle Parameterization for the Non-peaking Background

In this work, the PDF modeling the non-peaking background followed the form described by equation 5.13; on the other side, in the Belle analysis mentioned in chapter 7 [76], the parameterization of this component is given by

$$f(m_{bc}) = m_{bc} \left(1 - \left[\frac{m_{bc}}{c}\right]^2\right)^p \exp\left(\chi \left[1 - \left(\frac{m_{bc}}{c}\right)^2\right]\right). \quad (\text{M.1})$$

with $p = 0.5$ (notice that the sign of the χ parameter matters, contrary to equation 5.13). In order to check the effect of this in the “rediscovery” luminosity, we parameterized the non-peaking background following equation M.1. The results are shown in figure M.1.

Comparing the two parameterizations, it is seen that the shape parameter χ_{Belle} in equation M.1 and the $\chi_{\text{Belle II}}$ one in equation 5.13, should be related by the expression

$$\chi_{\text{Belle}} = -\frac{\chi_{\text{Belle II}}^2}{2}. \quad (\text{M.2})$$

This is the case for the electron modes; for the muon ones, though the central values do not follow this relation, the error in the χ_{Belle} values are large enough to dim these results consistent; the shapes of the parameterizations are pretty much the same (compare figures 6.2 and M.1).

As a further check, we perform the fitting on the MC final m_{bc} distributions (see figure M.2), and extract the signal yields using the new parameterization. Notice that these yields are in very agreement with the ones obtained in chapter 6, thus showing once again that both models are consistent.

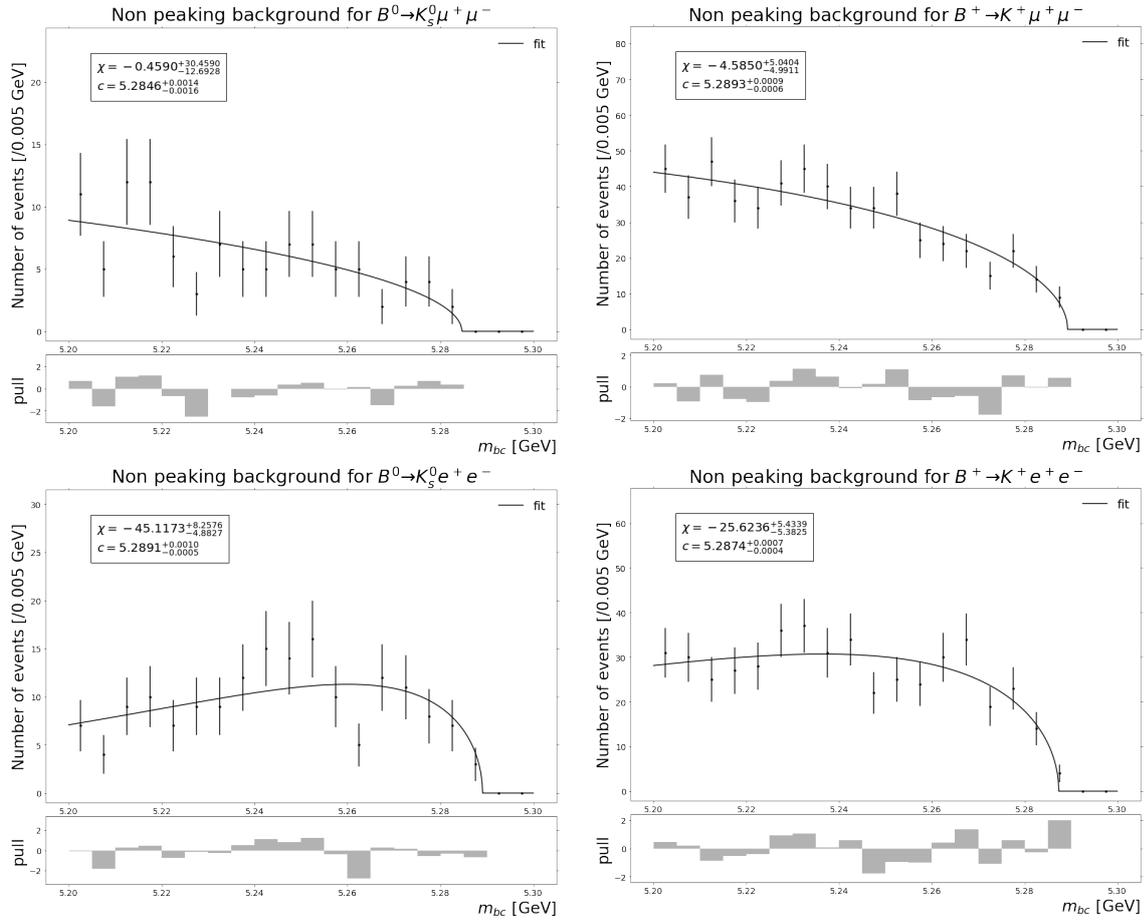


Fig. M.1 m_{bc} distributions and fits for the non-peaking backgrounds from MC, following the parameterization presented in equation M.1.

Notice, also, that the true number of signal events in MC for the charged muon mode is, although consistent with the ground MC truth, farther away from the true value (181), in comparison with the yield obtained in 6; on the other side, the new yield for the neutral muon mode is closer to the true number of signal events than the ones obtained in chapter 6. Although this maybe just an effect of the statistical fluctuations in the data sample, it may also hint that future analyses could benefit from using different parameterizations for the background components according to the flavor of the B meson and the leptons involved.

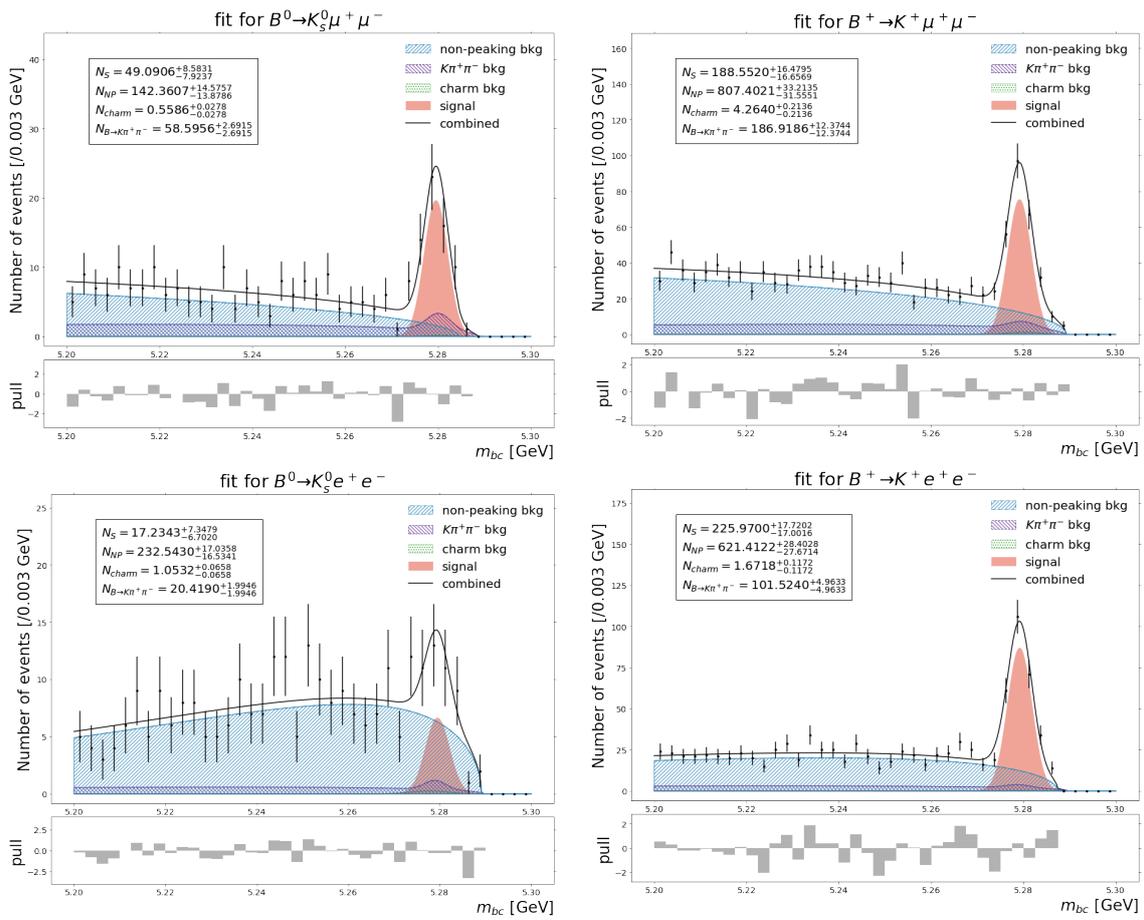


Fig. M.2 Final fits to the MC generic sample, corresponding to 2 ab^{-1} of data taken at the $Y(4S)$ resonance; the non-peaking background follows the shapes shown in figure M.1.

Appendix N

Statistical Error Calculation for R_K at Higher Selection Efficiencies

In this appendix we recalculate the relative statistical error in R_K , obtained after modifying the cuts in the MVA classifiers in order to preserve a larger portion of the signal events. We first estimate the new cut values for the MVA outputs from the testing datasets. Figures N.1 and N.2 present the cumulative distributions for the outputs' values; we select the cuts for which 5% of the signal events are lost (black, vertical line), and calculate the efficiency and the purity gains after applying these new cuts to the generic MC samples. The results are listed in table N.1. Notice that the purity gain is higher than 75% —which is the purity gain obtained by the MVA classifier in the Belle study— for all decays except $B^0 \rightarrow K_S^0 e^+ e^-$.

Table N.1. Cut values for the MVA outputs for a 5% signal efficiency loss (per classifier)

decay	CS cut	$B\bar{B}S$ cut	eff. loss [%]	purity gain [%]	final eff. [%]
$B^0 \rightarrow K_S^0 \mu^+ \mu^-$	0.37	0.28	12.05	83.02	9.40
$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.27	0.22	11.5	79.80	28.84
$B^0 \rightarrow K_S^0 e^+ e^-$	0.22	0.24	5.7	69.22	9.35
$B^+ \rightarrow K^+ e^+ e^-$	0.22	0.028	8.13	82.65	27.72

We use these efficiencies to estimate new values for the parameters $\alpha_{\ell\ell}^i$ in equation 7.10 as

$$\alpha_{\ell\ell}^i = \frac{2 \times C_{\ell\ell}^i \times \sigma(e^+e^- \rightarrow B\bar{B}) \times f^i \times \mathcal{B}(B^i \rightarrow K^i \ell^+ \ell^-)}{1 + r_{\ell\ell}^i} \quad (\text{N.1})$$

where the branching fraction values are taken from the latest PDG data [15].

From these, the expected statistical error in R_K is obtained in the same fashion as in chapter 6. The results are shown in figure N.3. From figure N.4 notice, however, that the

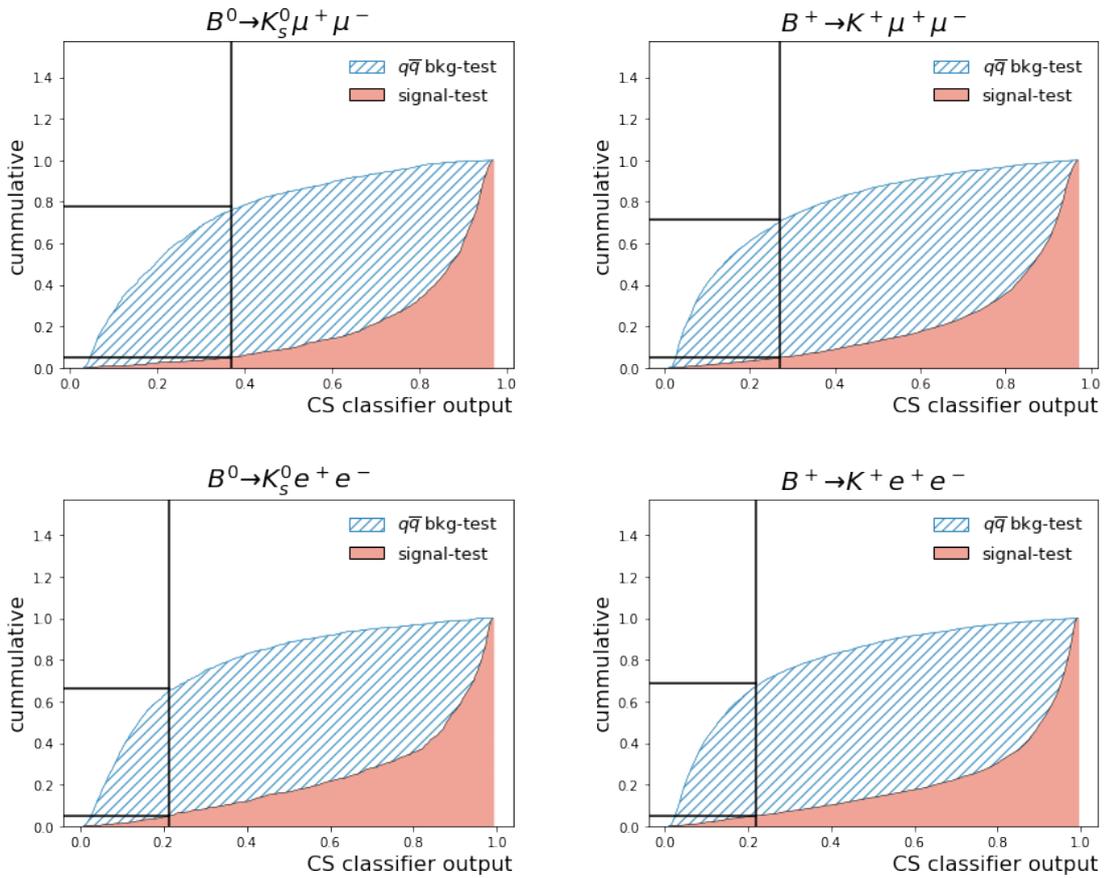


Fig. N.1 Cummulative distributions for the continuum classifiers' outputs on the validation datasets. The vertical line indicates the cut value at which 5% of the signal component is rejected.

signal peak is less distinguishable (specially for neutral modes), and that there is a larger background component peaking in the signal region. Thus, the signal extraction procedure would require additional care of these conditions.

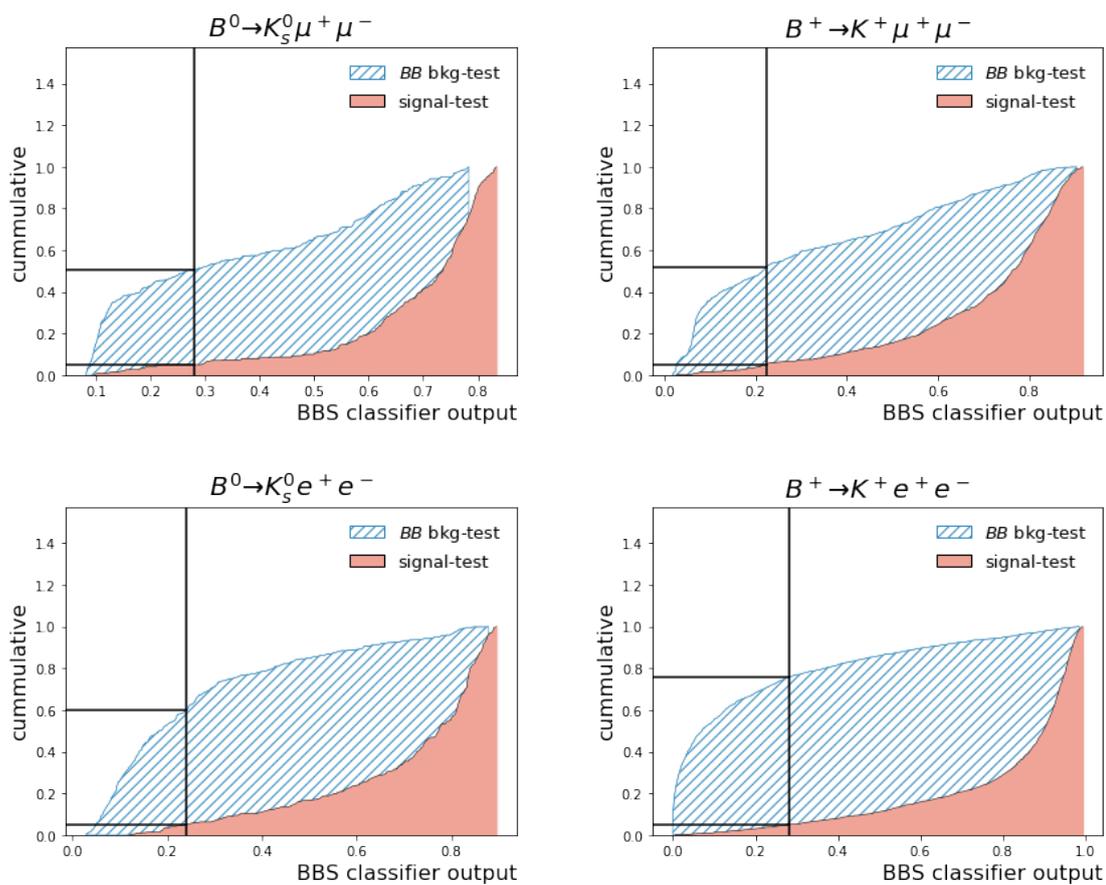


Fig. N.2 Cummulative distributions for the $B\bar{B}$ classifiers' outputs on the validation datasets. The vertical line indicates the cut value at which 5% of the signal component is rejected.

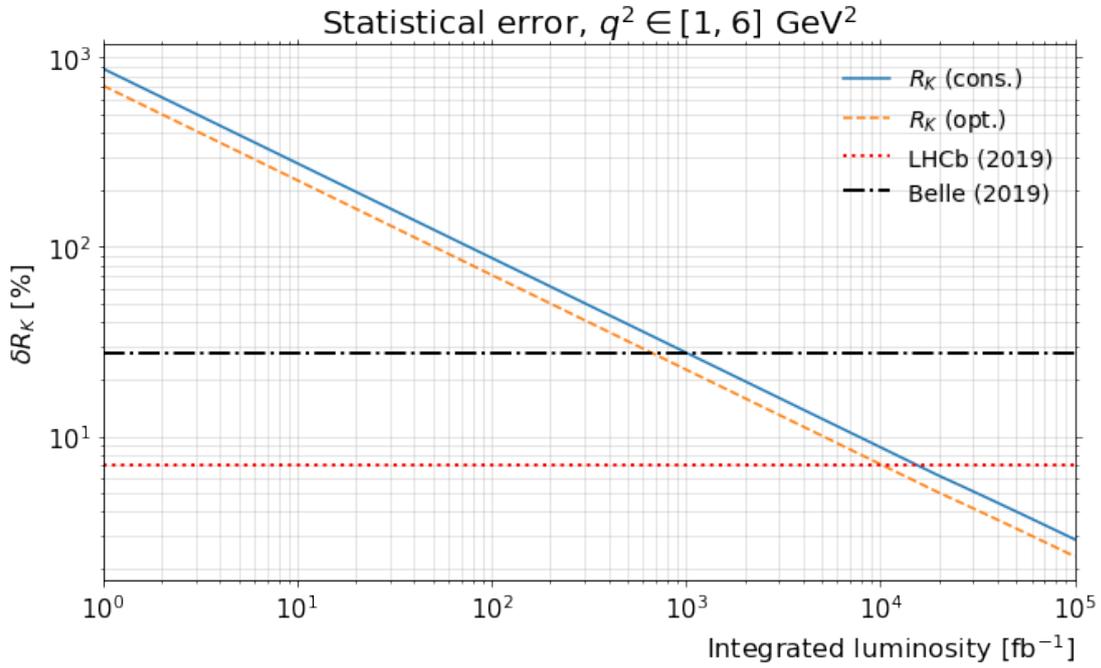


Fig. N.3 Statistical error projection in R_K for the looser MVA cuts.

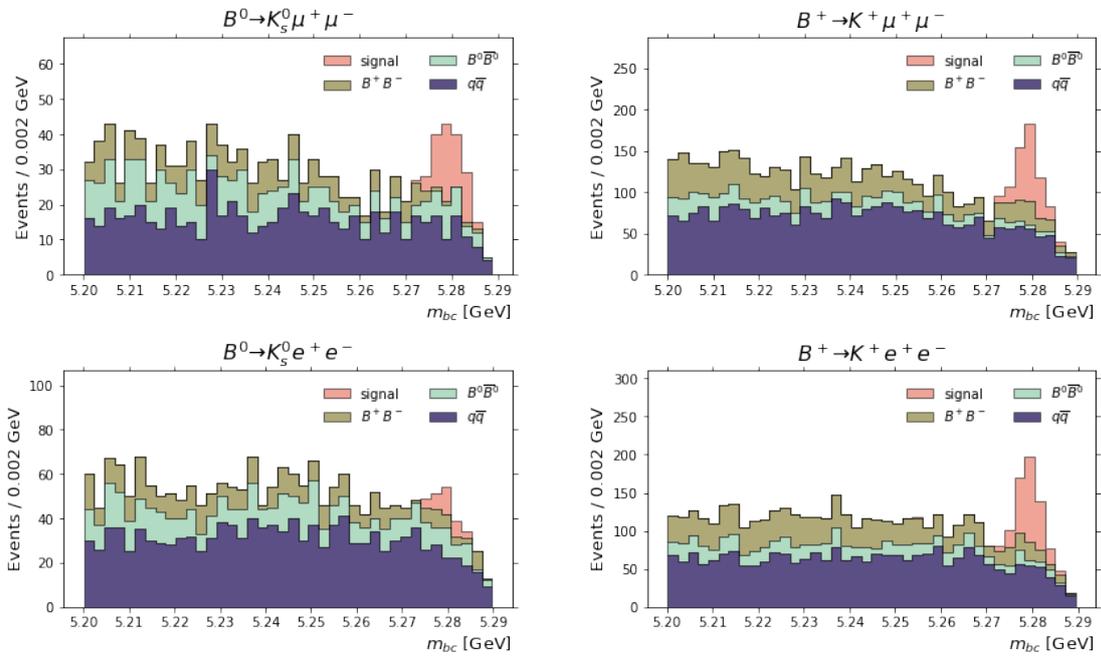


Fig. N.4 m_{bc} distribution for the selected MC events.