
**Development of a B^0 flavor tagger and
performance study of a novel
time-dependent CP analysis of the decay
 $B^0 \rightarrow \pi^0 \pi^0$ at Belle II**

Fernando Abudinén



München 2018

**Development of a B^0 flavor tagger and
performance study of a novel
time-dependent CP analysis of the decay
 $B^0 \rightarrow \pi^0 \pi^0$ at Belle II**

Fernando Abudinén

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Fernando Abudinén
aus Barranquilla, Kolumbien

München, den 28. September 2018



Dissertation

an der Fakultät für Physik
der Ludwig-Maximilians-Universität München
vorgelegt von Fernando Jesús Abudinén Gallego
am 28. September 2018.

Erstgutachter: Prof. Dr. Christian Kiesling
Zweitgutachter: Prof. Dr. Thomas Kuhr

Tag der mündlichen Prüfung: 19. Oktober 2018
Vorsitzender der Prüfungskommission: Prof. Dr. Gerhard Buchalla
Schriftführender Prüfer: Prof. Dr. Joachim Rädler

Max-Planck-Institut für Physik,
München, den 22. Oktober 2018.

„Und läg’ er nur noch immer in dem Grase!
In jede[s] Quark begräbt er seine Nase.“
Johann Wolfgang von Goethe, Faust I.

Contents

Zusammenfassung	xi
Abstract	xiii
1 Introduction	1
2 Physics at Belle II	3
2.1 The Standard Model of particle physics	3
2.2 Discrete symmetries	7
2.2.1 Parity inversion	8
2.2.2 Charge conjugation	9
2.2.3 Time reversal	10
2.2.4 <i>CPT</i> symmetry	11
2.2.5 <i>CP</i> symmetry	11
2.3 The Cabibbo angle and the GIM mechanism	13
2.4 The CKM matrix	14
2.4.1 Parameters in a mixing matrix	15
2.4.2 Parametrization of the CKM matrix	16
2.4.3 The unitarity triangle	18
2.5 Flavor mixing of flavored neutral mesons	20
2.6 Types of <i>CP</i> Violation	25
2.7 The <i>B</i> -meson sector	26
2.7.1 $B^0 - \bar{B}^0$ mixing	27
2.7.2 Time-dependent <i>CP</i> violation	28
2.7.3 Measurements of <i>CP</i> violation	30
2.7.4 Measurements of $B^0 - \bar{B}^0$ mixing	35
2.8 LHCb and Belle II	36
2.9 The unitarity angles ϕ_1 and ϕ_3	38
2.10 The unitarity angle ϕ_2	44
2.10.1 Properties of π and ρ mesons	48
2.10.2 Isospin analysis for $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$	48
2.10.3 Determination from $B \rightarrow \rho\pi$	54
2.11 Prospects for the unitarity angles	55

3	From Belle to Belle II	57
3.1	From KEKB to SuperKEKB	57
3.2	From the Belle to the Belle II detector	61
3.3	Track reconstruction	66
3.4	Vertex Reconstruction	68
3.5	Particle identification	69
4	The Belle II flavor tagger	71
4.1	Introduction	71
4.2	Definitions	73
4.3	Tagging categories	75
4.4	The category-based algorithm	91
4.5	Optimization and performance on MC	98
4.5.1	The MC samples	98
4.5.2	Total effective efficiency	100
4.5.3	Optimization of the target identification	102
4.5.4	Performance of individual categories	106
4.5.5	Performance of combiners on Belle II MC and on Belle MC	112
4.6	Understanding the asymmetries	117
4.7	Correlations with the generated Δt	131
4.8	Performance on Belle data	136
4.9	Test of robustness	142
4.10	The tag-side interference effect	151
4.11	Performance using only primary-lepton categories	154
4.12	Comparison with the alternative deep-learning approach	156
4.13	Comparison with Belle and BaBar	158
4.14	Calibrating the flavor tagger	159
4.15	Conclusion	161
5	Results from Belle II MC with background	163
5.1	Performance on Belle II MC with background	163
5.2	Conclusion	168
6	Performance study of the decay $B^0 \rightarrow \pi^0\pi^0$	169
6.1	Introduction	169
6.2	Generation of events	170
6.3	Analysis strategy	173
6.4	Reconstruction of the signal side	174
6.5	Distinguishing between Dalitz π^0 s and converted photons	181
6.6	Measurement of Δt	183
6.6.1	Signal-side vertex reconstruction	184

6.6.2	Tag-side vertex reconstruction	187
6.6.3	Δt resolution and Δt selection	188
6.7	Continuum Suppression	191
6.7.1	Input Variables	192
6.7.2	Performance and selection criterion	197
6.8	Flavor tagging	199
6.9	Final selection and efficiency	200
6.10	Estimation of precision with pseudo experiments	205
6.10.1	The extended unbinned maximum-likelihood fit	206
6.10.2	The four-photon model	207
6.10.3	The Δt resolution function	211
6.10.4	The Dalitz-Conversion model	213
6.10.5	Fit projections	221
6.10.6	Fit validation and results	225
6.11	Comparison with Belle	228
6.12	Conclusion	229
7	The unitarity angle ϕ_2 at Belle II	231
7.1	Introduction	231
7.2	The isospin analysis fit	231
7.3	Extrapolation of uncertainties	233
7.4	The $B \rightarrow \pi\pi$ system	235
7.4.1	Expected precision at Belle II	236
7.4.2	Results of the isospin analysis	240
7.5	The $B \rightarrow \rho\rho$ system	244
7.5.1	Expected precision at Belle II	244
7.5.2	Results of the isospin analysis	249
7.6	Expected precision combining the $B \rightarrow \pi\pi$ and the $B \rightarrow \rho\rho$ systems	250
7.7	Concluding remarks and prospects	254
8	Conclusion and outlook	257
A	Physics	261
A.1	The Δt probability density for time-dependent CP analysis	261
A.2	Probability density for time-integrated CP analysis	262
A.3	Probability densities for t_{sig} and t_{tag}	263
A.4	Derivation of the triangular isospin relations for the decays $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$	266
B	The Belle II flavor tagger	273
B.1	Probability density for CP violation analyses considering flavor tagging	273

B.2	Calculation of statistical uncertainties	274
B.3	Input variables for each category	278
B.4	Studies without impact parameters	299
B.5	Test of Robustness	307
C	Probability Density Functions	323
C.1	Gaussian distribution	323
C.2	ARGUS distribution	323
C.3	Crystal Ball function	324
C.4	Kernel density estimation	324
D	Performance study of the decay $B^0 \rightarrow \pi^0\pi^0$	327
D.1	Distributions of the pull and the residuals	327
	Acknowledgments	345

Zusammenfassung

Das Belle II Experiment befindet sich am neuen Elektron-Positron Beschleuniger SuperKEKB, eine B -Mesonenfabrik der zweiten Generation am Forschungszentrum KEK in Tsukuba, Japan. Mit einer Design-Luminosität von $8 \cdot 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ wird SuperKEKB um ca. zwei Größenordnungen den Weltrekord seines Vorgängers KEKB übertreffen. Dies wird schließlich zu einem Datensatz führen, der einer integrierten Luminosität von ca. 50 ab^{-1} entsprechen wird, einem Wert, fünfzig mal größer als der des Vorgängerexperimentes Belle. Der neue Belle II Detektor ist ausgestattet mit modernster Technik, um den Herausforderungen aufgrund der hohen Luminosität und des erhöhten Untergrunds standhalten zu können. Das Hauptziel des Belle II Experimentes ist die Entdeckung möglicher Beiträge von neuer Physik jenseits des Standard Modells, in dem die Eigenschaften von B - und D -Mesonen als auch von τ -Leptonen, die am SuperKEKB Beschleuniger produziert werden, präzise untersucht werden.

Ein Großteil des Belle II Physikprogramms ist fokussiert auf Messungen der CP -Verletzung in B -Mesonzerfällen. Für diese Analysen, muss das Flavor von einem der zwei ausschließlich produzierten B^0 -Mesonen bestimmt werden, eine Aufgabe, die mit Hilfe von speziellen Algorithmen durchgeführt wird. Im Rahmen dieser Dissertation wurde für das Belle II Experiment ein neuer Algorithmus zur Flavor-Bestimmung entwickelt. Dieser wurde sorgfältig optimiert, um die Fähigkeiten des neuen Detektors auszunutzen. Der Algorithmus basiert auf neuen, multivariaten Methoden, welche von den spezifischen Flavor-Signaturen von B^0 -Zerfällen optimalen Gebrauch machen. Da Belle II Daten noch nicht zur Verfügung stehen, wurde der neue Algorithmus mit Belle Daten ausführlich validiert. Die effektive Effizienz, die der neue Algorithmus mit Belle Daten erreichte, ist $\varepsilon_{\text{eff}} = (33.5 \pm 0.5 \text{ (stat)})\%$, ein Wert, der um ca. 10% die Leistung des Vorgängeralgorithmus von Belle übertrifft. Mit Belle II Monte Carlo Ereignissen, die mit Untergrund simuliert wurden, erreicht der neue Algorithmus eine effektive Effizienz von $\varepsilon_{\text{eff}} = (33.89 \pm 0.04 \text{ (stat)})\%$.

Aufgrund der Größe des erwarteten Datensatzes und des Präzisionsvermögens für die Vertexrekonstruktion wird Belle II im Stande sein, Messungen durchzuführen, die bei Vorgängerexperimenten nicht möglich waren. Ein weiterer, wesentlicher Teil dieser Dissertation ist der Vorbereitung der Datenanalyse für die Vermessung des CP -Verletzungsparameters \mathcal{S}_{CP} im Zerfall $B^0 \rightarrow \pi^0 \pi^0$ gewidmet. Diese Messung ist essenziell für die Reduktion der 8-fachen Ambiguität in der Bestimmung des Unitaritätswinkels ϕ_2/α , wenn man $B \rightarrow \pi\pi$ Zerfälle betrachtet. Die Vermessung von $\mathcal{S}^{\pi^0 \pi^0}$ benötigt eine präzise Bestimmung des B^0 -Zerfallvertex, welche im dominanten Endzustand mit vier Photonen nicht erreicht werden kann. Aus diesem Grund wurde eine Analysemethode entwickelt, die seltene Zerfälle mit Dalitz-Zerfällen oder mit in der Strahlröhre konvertierten Photonen nutzt. Die Studien zeigten, dass mit dem gesamten Belle II Datensatz von 50 ab^{-1} die Vermessung von $\mathcal{S}^{\pi^0 \pi^0}$ machbar sein wird. Die erwartete Unsicherheit von $\mathcal{S}^{\pi^0 \pi^0}$ wurde zu $\mathcal{S}^{\pi^0 \pi^0} = \pm 0.28 \text{ (stat)} \pm 0.03 \text{ (syst)}$ bestimmt. Mit dieser Präzision führt $\mathcal{S}^{\pi^0 \pi^0}$ zu einer Reduktion der 8-fachen ϕ_2 -Ambiguität zu einer 2-fachen. Bei einer gleichzeitigen Betrachtung der Isospin-Analysen von den Zerfällen $B \rightarrow \pi\pi$ und $B \rightarrow \rho\rho$ wurde die erwartete Belle II Präzision für ϕ_2 auf insgesamt $\delta\phi_2 \approx 0.6^\circ$ abgeschätzt, eine Genauigkeit, die mehr als fünfmal kleiner ist, als die des jetzigen Weltmittelwertes.

Abstract

The Belle II experiment is located at the energy-asymmetric electron-positron collider SuperKEKB, the next-generation B factory at KEK in Tsukuba, Japan. With a design luminosity of $8 \cdot 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, SuperKEKB will overtake by nearly two orders of magnitude the record reached by its predecessor KEKB, leading ultimately to a data sample corresponding to a recorded integrated luminosity of about 50 ab^{-1} , which is fifty times larger than the one collected by the predecessor experiment Belle. The Belle II detector is a new state-of-the-art detector designed to cope with the challenging high-luminosity conditions and with the larger beam-backgrounds. The primary goal of the Belle II experiment is to reveal, or to impose stringent constraints on, possible new physics contributions beyond the Standard Model (SM) by studying accurately the properties of B and D mesons as well as of τ leptons produced at SuperKEKB.

A major part of the Belle II physics program is focused on measuring the violation of CP symmetry in B -meson decays. For a CP analysis of neutral B -meson decays at B factories, the flavor of one of the two exclusively produced neutral B mesons has to be determined, a task that is performed by flavor tagging algorithms, or shortly flavor taggers. In this thesis, a new flavor tagging algorithm was developed for Belle II with a specific and dedicated optimization to exploit the capabilities of the new detector and to cope with the harsher experimental conditions at SuperKEKB. The new Belle II flavor tagger developed in this thesis is based on novel multivariate methods exploiting the flavor specific signatures of B^0 decays. In the absence of Belle II data, the new Belle II flavor tagger was comprehensively validated using Belle data, on which the algorithm reached an effective tagging efficiency of $\varepsilon_{\text{eff}} = (33.5 \pm 0.5 \text{ (stat)})\%$, outperforming the previous algorithm at Belle by about 10%. On Belle II Monte Carlo events simulated with background, the algorithm reaches an effective tagging efficiency of $\varepsilon_{\text{eff}} = (33.89 \pm 0.04 \text{ (stat)})\%$.

Thanks to the expected large size of its data sample and to its vertex reconstruction capabilities, Belle II will be able to make measurements that were not possible at its predecessors. A substantial part of this thesis is dedicated to prepare the data analysis for the measurement of the time-dependent CP -violation parameter \mathcal{S}_{CP} of the decay $B^0 \rightarrow \pi^0 \pi^0$, which is essential to reduce the 8-fold ambiguity in the determination of the unitarity angle ϕ_2/α considering $B \rightarrow \pi\pi$ decays. The time-dependent CP analysis of $B^0 \rightarrow \pi^0 \pi^0$ requires a precise determination of the B^0 -decay position, which cannot be achieved in the dominant four-photons final state. Thus, a novel analysis method was developed to make use of rare events with photons converting into e^+e^- pairs or neutral pions undergoing Dalitz $\pi^0 \rightarrow e^+e^-\gamma$ decays. The studies showed that, with the full Belle II data sample of 50 ab^{-1} , the measurement of $\mathcal{S}^{\pi^0 \pi^0}$ will be feasible. The expected uncertainty of $\mathcal{S}^{\pi^0 \pi^0}$ was estimated to be $\delta \mathcal{S}^{\pi^0 \pi^0} = \pm 0.28 \text{ (stat)} \pm 0.03 \text{ (syst)}$, resulting in a reduction of the current 8-fold ϕ_2 -ambiguity to a 2-fold one. Considering the isospin analyses of both $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ decays, the overall expected Belle II precision for ϕ_2 at 50 ab^{-1} was estimated to be around $\delta \phi_2 \approx 0.6^\circ$, which is more than five times smaller than the current world average.

1 Introduction

The standard model of particle physics (SM) succeeds in describing all known particles and the interactions relevant for their dynamics, passing all the experimental tests to date. However, a wide range of observations and several open questions of theoretical and phenomenological nature indicate that the SM is an effective theory that is accurate at the energy scales probed so far, but is incomplete at higher scales. The SM does not, for example, include gravity, nor offers an explanation for the existence of Dark Matter or Dark Energy, and cannot, based on the well-founded Big Bang scenario, explain the dominance of matter over anti-matter in the observable universe.

The key goal of particle physics today is to find experimental indications of non-SM phenomena. The two largest experiments at the Large Hadron Collider (LHC) at CERN, ATLAS and CMS, aim at discovering new particles produced directly in proton-proton collisions at center-of-mass energies up to 14 TeV. The direct searches at LHC, however, have yielded null results so far. Since more powerful colliders are not foreseen in the near future, indirect searches at the precision frontier offer promising opportunities to find physics beyond the SM in the next decade.

Indirect searches rely on comparing precise measurements with similarly precise SM predictions to reveal significant deviations, indicating New Physics (NP). Located at the next-generation B factory SuperKEKB, the Belle II experiment is one of the leading facilities for precision indirect searches. A major part of the Belle II physics program is focused on the study of quark-flavor dynamics with emphasis in the B -meson sector. Because of its rich phenomenology, quark-flavor physics offers unique opportunities to test a broad class of hypothetical SM extensions. Typically, SM extensions introduce new heavy particles that, for example, may enhance the violation of CP symmetry in B -meson decays via quantum-loop effects, a condition necessary to explain the matter-antimatter asymmetry in the universe. The high-precision measurements at Belle II have the potential to be sensitive to additional very high mass contributions in the quantum loops far beyond the range that is directly accessible at LHC.

The ultimate precision and probing power in many key measurements associated with B -meson dynamics requires the identification of the B -meson flavor, i.e. to discriminate between neutral B mesons and their antiparticles. This task is referred to as flavor tagging and is accomplished using dedicated flavor tagging algorithms. Due to the challenging high-luminosity conditions and the expected large beam backgrounds at SuperKEKB, a novel flavor tagging algorithm had to be developed for Belle II with a specific and dedicated optimization

in order to cope with the new harsh experimental conditions and to exploit the capabilities of the new Belle II detector. A major part of this thesis is devoted to the development, optimization, performance characterization, and validation of the new flavor tagging algorithm which has now become the official Belle II flavor tagger. Because of unexpected delays of the SuperKEKB machine, the algorithm could not be tested with data from Belle II, but had to be validated using simulated Monte Carlo events as well as data from the predecessor experiment Belle. Due to the novel small size of the interaction region at SuperKEKB, the performance of the flavor tagger at Belle II was found to be influenced by the amount of CP violation. This small but interesting new effect was carefully studied for the first time in this thesis.

One of the currently limiting factors for testing the SM description of quark-flavor dynamics is the unitarity angle ϕ_2/α . Important channels to measure this angle are the $B \rightarrow \pi\pi$ decays. Here, however, a crucial measurement is missing, namely the observable $\mathcal{S}^{\pi^0\pi^0}$, for which a time-dependent CP -violation analysis of the decay $B^0 \rightarrow \pi^0\pi^0$ is required. A measurement of $\mathcal{S}^{\pi^0\pi^0}$, which has not been feasible yet, would offer the opportunity to make substantial advancements in understanding the consistency of the model.

This thesis presents a comprehensive study launched to prepare the data analysis for the measurement of the physical observables associated with the $B^0 \rightarrow \pi^0\pi^0$ decay. The aim of the study was to prove that Belle II will be able to measure $\mathcal{S}^{\pi^0\pi^0}$ and to assess the future Belle II precision for this observable. The analysis served also to apply for the first time the developed flavor tagging algorithm. The final part of this thesis is dedicated to study the impact of the novel measurement of $\mathcal{S}^{\pi^0\pi^0}$ on the determination of ϕ_2 and to assess the overall Belle II precision for ϕ_2 .

The outline of this thesis is as follows:

- Chapter 2 gives a brief introduction into the Standard Model of particle physics focusing on the study of discrete symmetries in quark-flavor dynamics, and especially on the study of CP violation in the B -meson sector.
- Chapter 3 will introduce the KEKB accelerator and its upgrade to SuperKEKB, together with the Belle and the Belle II experiments.
- Chapter 4 is devoted to the Belle II flavor tagger, covering in detail the concept behind the algorithm, as well as its optimization, performance characterization, and validation.
- Chapter 5 presents the performance characterization of the flavor tagger on Belle II Monte Carlo events simulated with background.
- Chapter 6 is dedicated to the analysis of the decay $B^0 \rightarrow \pi^0\pi^0$, and to assess the Belle II precision for the measurement of the physical observables associated with this decay.
- Chapter 7 is devoted to study the impact of the measurement of $\mathcal{S}^{\pi^0\pi^0}$ on the determination of the unitarity angle ϕ_2 , and to assess the overall Belle II precision for ϕ_2 .
- Chapter 8 summarizes the studies and the results presented in this thesis, giving an outlook to future measurements at Belle II.

2 Physics at Belle II

The Belle experiment, together with its competitor experiment BaBar, succeeded in establishing the violation of CP symmetry in the B -meson sector. The results of both experiments, along with those from kaon and hadron-collider experiments, agree with the SM predictions. However, further measurements are required to increase the sensitivity on possible non-SM contributions. The primary motivation of the upgrade of the Belle experiment to Belle II is the search for deviations providing a hint at such contributions.

This chapter introduces the SM of particle physics and focuses on the violation of CP symmetry. The chapter sets the theoretical basis for the studies presented in this thesis covering the CP violation in the B -meson system, its measurement at B factories, and the determination of the so-called unitarity angles. An emphasis is placed on the unitarity angle ϕ_2 , which plays a central role in this thesis.

The Belle II experiment has a rich physics program which goes beyond the CP violation in the B -meson sector. A comprehensive overview of the full Belle II program can be found in [1].

2.1 The Standard Model of particle physics

The SM is a quantum field theory (QFT) based on a special class of continuous symmetries such as local gauge invariance and Lorentz invariance. It describes the three fundamental forces relevant for the description of dynamics at the microscopic level: the electromagnetic, the weak, and the strong force. The forces are mediated by the exchange of virtual spin-1 particles called gauge bosons. The gauge bosons result from the invariance under local gauge transformations which generate mathematical symmetry groups. The strength of the interaction is determined by a constant of the gauge transformation and corresponds to the charge on which the force acts.

Tab. 2.1 presents a list with the four known fundamental interactions in nature and their mediators. They are listed in ascending order according to their strength relative to the strong force. The range of the forces is inversely proportional to the mediator mass [2]. The only massive mediators are the W^+ , the W^- and the Z^0 bosons; their masses give a natural scale for the range of the weak interaction. Since the strong-force mediator, the gluon, has zero mass, the strong force has an infinite range. However, due to the dynamics of the strong force, the effective range at which its effect manifests is around 2 fm (about a proton diameter), corresponding to the mass of the pion which can be considered as effective mediator [2].

Table 2.1: The four fundamental forces in nature. They are ordered according to their strength values, which are determined by their respective coupling constants, and normalized to the strength of the strong force at a distance of ca. 1 fm (about the size of a proton). The strongest force is in the bottom. The question mark (?) means that the graviton is a hypothetical particle that has not been observed yet.

Interaction	Acts on	Relative strength	Range [m]	Mediator
Gravity	Mass	10^{-38}	∞	Graviton (?)
Weak	Weak isospin	10^{-5}	10^{-18}	W^+ , W^- , Z^0
Electromagnetic	Electric charge	$1/137$	∞	Photon (γ)
Strong	Color charge	1	∞	Gluon (g)

The constituents of matter are elementary spin- $\frac{1}{2}$ particles, which can be sorted into two types: *quarks* and *leptons*. The interactions of quarks and leptons with the photon are described by the theory of quantum electrodynamics (QED) which is based on the Abelian, i.e. commutative, symmetry group of unitary 1×1 operators called $U(1)_{EM}$. Here, the subscript EM stays for electromagnetic. The strength of the interaction corresponds to the electric charge of the particle.

Quarks have never been observed as free particles, they exist only within *hadrons*. The force responsible for holding hadrons (and atomic nuclei) together is the strong force. Strong interactions are described by quantum chromodynamics (QCD) which is based on the non-Abelian symmetry group $SU(3)_c$. The subscript c stands for color charge, or just *color*, which is the quantum number corresponding to the charge of the interaction. The three degrees of freedom of the symmetry group are the degrees of freedom of the color charge, which are labeled in analogy to the three kinds of colors humans perceive: red, green, blue. Apart from electric charge, every quark carries one of the three different color charges.

In contrast to the photon, the gluon carries the interaction charge. This property allows a gluon to interact with other gluons. In the $SU(3)_c$ group, there are $3^2 - 1 = 8$ independent generators (color octets), where the term -1 refers to an excluded not-observed color-neutral gluon (color singlet). Each generator gives rise to a gluon with a specific color-anticolor combination.

Both quarks and gluons form hadrons, which are color-neutral, or “white”, composite particles. There are two known types of hadrons: *mesons* and *baryons*. Mesons are bound states of a quark and an antiquark, where the first carries one of the three possible colors and the second the respective anti-color. Baryons, on the other hand, are composed of three quarks, each one with a different color; antibaryons are composed of three antiquarks, each one with a different anti-color.

A striking phenomenon in QCD is the *asymptotic freedom* associated with the non-Abelian (non-commutative) nature of the $SU(3)_c$ symmetry group [3, 4]. Asymptotic freedom means

that the strength of the interaction tends to vanish at small distances (large energy scales). If the distance between two colored objects increases, the strength of the strong interaction increases indefinitely. When two colored objects try to move apart, a flux tube of gluons is created. If the distance increases, the flux tube becomes “stretched” and the gluon force increases until the flux tube breaks by the generation of a quark-antiquark pair. In such wise, a color-neutral hadron breaks up into one or more color-neutral hadrons. The fact that colored objects do not exist as free particles is called *color confinement*.

Considering now weak interactions, they differ drastically from the electromagnetic and the strong interactions. The symmetry group of the weak interaction is the non-Abelian $SU(2)_L$ group, where the subscript L stays for left-handed, indicating that the fermions participating in weak interactions are left-handed (see Sect. 2.4). Left-handed fermions form so-called weak isospin doublets ($I = \frac{1}{2}$) with third isospin component $I_3 = \pm\frac{1}{2}$, such as

$$\begin{array}{l} I_3 = +\frac{1}{2} \\ I_3 = -\frac{1}{2} \end{array} \quad \left(\begin{array}{c} u \\ d \end{array} \right)_L, \quad \left(\begin{array}{c} \nu_e \\ e^- \end{array} \right)_L,$$

while right-handed fermions, such as u_R , d_R and e^-_R , are weak isospin singlets ($I = 0$) with $I_3 = 0$. The charged mediators of the weak interaction, the W^\pm bosons, couple to all quark and lepton doublets; they are the only gauge bosons in the SM that can change the flavor of particles and are therefore responsible for most of the particle decays.

The $SU(2)_L$ symmetry group requires three massless spin-1 bosons as carriers of the interaction; they form the W -boson triplet ($I = 1$) W^+ , W^- and W^0 (W_μ^i , $i = 1, 2, 3$) which couples only to left-handed fermions. Experiments have shown, however, that the three bosons mediating the weak interactions are W^+ , W^- and Z^0 , where the Z^0 couples partially to right-handed fermions, indicating that it is not a pure W^0 . The Z^0 boson results from the unification of electromagnetic and weak interactions. The unified interaction is called electroweak interaction and is described by the electroweak theory which is based on the symmetry group $SU(2)_L \otimes U(1)_Y$, where the subscript Y stays for hypercharge [5–7]. In addition to the W -boson triplet, this symmetry group requires a neutral gauge boson singlet associated with the group $U(1)_Y$: the gauge field B^0 (B_μ^0). In the electroweak theory, the observed photon γ and the observed Z^0 boson are linear combinations of the two electrically neutral components W^0 and B^0 . In terms of the neutral fields W_μ^3 and B_μ^0 , the photon field A_μ and the Z -boson field Z_μ are equivalent to

$$\begin{aligned} A_\mu &= B_\mu^0 \cos \theta_W + W_\mu^3 \sin \theta_W, \\ Z_\mu &= -B_\mu^0 \sin \theta_W + W_\mu^3 \cos \theta_W, \end{aligned}$$

where the mixing angle θ_W is called Weinberg angle. The triplet of W bosons couples with the strength g of the $SU(2)_L$ group to the doublets of left-handed fermions, while the neutral gauge field B_μ^0 couples with the strength g' of the group $U(1)_Y$ to left- and to right-handed fermions in

proportion to their respective hypercharges $Y = 2(Q_{\text{EM}} - I_3)$, where Q_{EM} denotes the electric charge. The strengths g and g' fulfill $g \sin \theta_W = g' \cos \theta_W = e$, where e is the elementary electric charge.

In the electroweak theory, the mechanism explaining how the $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken is the Higgs mechanism [8–10]. The Higgs mechanism introduces a complex scalar field giving rise to a scalar spin-0 particle, the *Higgs boson*. The scalar field is symmetric under global gauge transformations, but has a non-zero vacuum expectation value v , breaking the gauge invariance locally. Through the spontaneous symmetry breaking, the Higgs mechanism provides masses for the W^+ , the W^- and the Z^0 bosons. Additionally, the fermions acquire mass by coupling to the Higgs field.

By grouping together QCD and the electroweak theory, one obtains the Standard Model of particle physics, which is based on the combined symmetry group

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y.$$

The SM describes a number of quantum fields whose excitations are the observed elementary particles. Figure 2.1 shows a diagram with all the elementary particles of the SM: three generations of quarks and leptons (spin- $\frac{1}{2}$ fermions), four spin-1 gauge bosons and the scalar spin-0 Higgs boson. Each of the different fermions is associated with a *flavor quantum number*. For each of the fermions, there is a corresponding antiparticle with opposite flavor. The

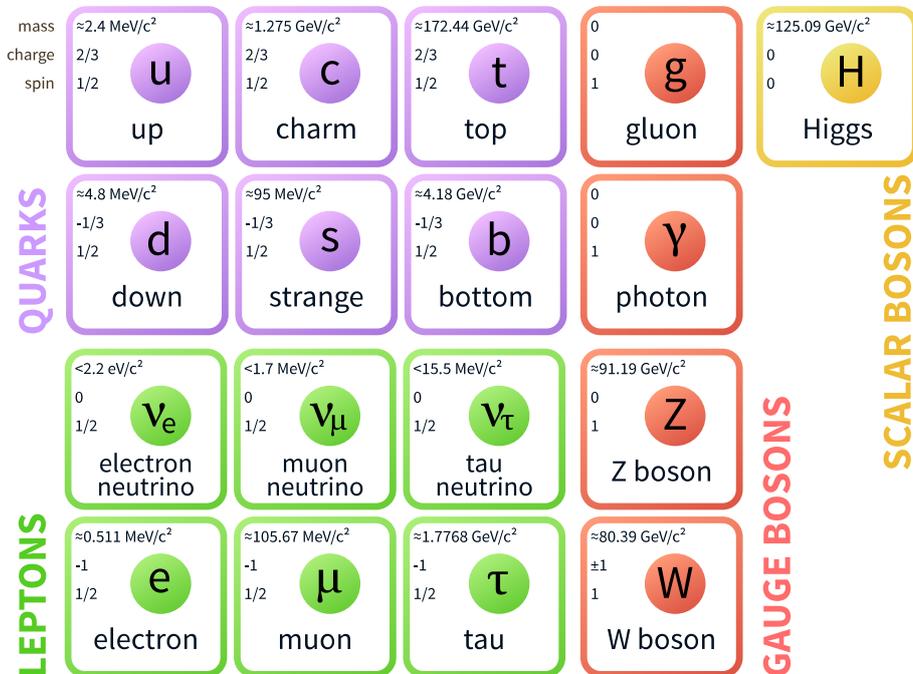


Figure 2.1: The elementary particles of the Standard Model [11]. They are divided into three generations of quarks and leptons, gauge bosons as force carriers, and the Higgs boson. The fermions are shown in the same order as in the weak isospin doublets.

three generations of fermions represent a mass hierarchy. The first generation corresponds to the lightest fermions: the two quarks of this family, together with the electron, build all the visible matter. The heavier particles of the second and the third generations are unstable; they decay into the lighter particles of the first generation.

The ultimate piece of the standard model was the Higgs boson, which was discovered in 2012 by the ATLAS and the CMS collaborations [12, 13]. However, there are several open questions suggesting that the SM cannot be the ultimate theory. Extensions of the SM most likely consist in new massive particles not contained in it.

Finding the first indications for New Physics (NP) beyond the SM is the chief goal of particle physics today. To search for such indications, there are two complementary approaches. The first one are direct searches for new heavy particles at high-energy colliders (*energy frontier*), an approach that have been successful since the invention of colliders. The second one are indirect searches for non-SM particles by comparing precise measurements with precise SM predictions (*intensity frontier*). Indirect searches have been also successful in the past; these searches are focused on processes where new particles can contribute as virtual particles via quantum-loop corrections. Since the masses of new particles can be arbitrarily high, indirect searches offer a way to explore energy scales that are higher than those directly accessible at colliders. At present, direct searches at LHC have yielded null results. Thus, the indirect approach is a very promising one in the next decade.

The quark-flavor dynamics in the standard model is described by the CKM mechanism. Quark-flavor dynamics offers a large variety and abundance of processes that are both, sensitive to non-SM dynamics up to much higher energies than directly reachable at LHC and stringently constrained into a simple structure with only four free parameters. Thus, quark-flavor dynamics offers among the most promising opportunities for indirect searches exploring the next energy scale in the next decade and beyond. This thesis targets at making an essential contribution to the indirect searches at Belle II and is focused on the flavor dynamics of B mesons, in particular, on the phenomena related to the violation of the discrete charge-parity symmetry in this system.

2.2 Discrete symmetries

Symmetries are fundamental in physics. They give rise to conservation laws [14], and provide an insight into the dynamics of a given system. Symmetries help simplify models and calculations, making them solvable in most of the cases. A symmetry exists when a physical law is invariant under certain transformations. There are two types of transformations: continuous transformations, such as rotations or translations, and discrete transformations, such as reflections or charge conjugations. The parity transformation P , the charge conjugation C , the time inversion T as well as the product transformations CP and CPT belong to the discrete transformations and will be introduced in the following.

2.2.1 Parity inversion

The parity transformation P is defined as an inversion of the spatial coordinates at the origin, i.e. $(x, y, z) \rightarrow (-x, -y, -z)$. This is equivalent to a mirror reflection followed by a 180° rotation, as illustrated in Fig. 2.2. Physical observables can be classified into four types according to their transformation properties with respect to P : vectors, axial vectors, scalars and pseudoscalars. Polar vectors (called usually just vectors), like the momentum \mathbf{p} pointing in the direction of motion of a particle, are reversed under a P transformation. Axial vectors like the angular momentum \mathbf{L} and the spin \mathbf{s} remain constant. Scalar observables like the squared momentum $\mathbf{p} \cdot \mathbf{p}$ are invariant under parity transformations, while pseudoscalar observables like the helicity $\hat{\mathbf{p}} \cdot \hat{\mathbf{s}} = \frac{\mathbf{p} \cdot \mathbf{s}}{|\mathbf{p}| |\mathbf{s}|}$ flip sign under P . If a parity transformation is applied for a second time, the spatial coordinates are transformed back to their original directions. In quantum mechanics a parity transformation corresponds thus to an operator P , which is unitary and Hermitian¹, fulfilling $P^2 = 1$. Consequently, the possible eigenvalues of an eigenstate of P are $\eta_P = \pm 1$.

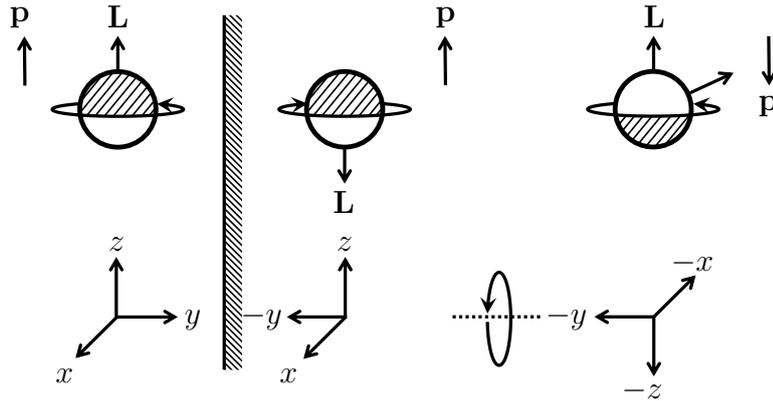


Figure 2.2: A mirror reflection at an arbitrary plane followed by a rotation of 180° around the axis orthogonal to the plane corresponds to the parity transformation P . The graphic illustrates the effect of P on linear and angular momenta \mathbf{p} and \mathbf{L} .

The angular momentum part of a particle wave function $\psi(t, \mathbf{x})$ can be expressed in terms of spherical harmonic functions, which are either symmetric or antisymmetric under parity transformation, with respective eigenvalues $+1$ or -1 . Additionally, an *intrinsic parity* of ± 1 can be assigned to particles. For a system of several particles, the total parity eigenvalue is the product of the intrinsic parities times $(-1)^L$, where L is the quantum number corresponding to the angular momentum of the system. By convention, the intrinsic parity of the proton was defined to be positive as its electric charge.

¹In QFT, the parity symmetry corresponds in general to some discrete internal symmetry, and in very special cases, which are not relevant for the SM particles considered in this thesis, the parity operator fulfills $P^2 = -1$ (see e.g. [15, chapter 3]).

Parity symmetry is conserved in strong and in electromagnetic interactions. The conservation of parity was considered as a dogma until the so-called $\theta - \tau$ puzzle motivated a rethinking in the mid 1950s. By 1956, two strange particles, called at that time θ^+ and τ^+ (not to be confused with the τ lepton), had been observed and they appeared to be identical in every respect (mass, lifetime, production cross section, spin and charge) except in their decay: θ^+ decayed into two pions (positive parity) and τ^+ into three pions (negative parity). In the same year, Tsung-Dao Lee and Chen-Ning Yang suggested that θ^+ and τ^+ could be just different decay modes of the same particle, which today is called K^+ . They noted that parity conservation in weak interactions had not been accurately tested before and proposed several experiments to perform this [16]. In particular, they proposed to test parity conservation in β decays.

By the end of 1956, Chien-Shiung Wu and collaborators carried out the proposed experiment testing parity conservation in β decays; the experiment was based on the β^- decay ${}^{60}_{27}\text{Co} \rightarrow {}^{60}_{28}\text{Ni}^* + e^- + \bar{\nu}_e$ [17]. ${}^{60}_{27}\text{Co}$ has a total nuclear spin $J = 5$, and ${}^{60}_{28}\text{Ni}^*$ has $J = 4$. The nuclear spins were aligned in a magnetic field. Due to conservation of angular momentum, the spins of the anti-neutrino and the electron had to be parallel to the magnetic field. The observation was that electrons were emitted predominantly opposite to the magnetic field, i.e. opposite to the spin direction. This means that the emission of electrons with negative helicity $\hat{p} \cdot \hat{s} = -1$ is favored, while the emission of electrons with positive helicity is suppressed. Since parity transformation flips the sign of the helicity, the experiment proved that parity symmetry is maximally violated in β decays. In turn, Lee and Yang received the Nobel Prize in 1957.

2.2.2 Charge conjugation

The charge conjugation C transforms a particle into its anti-particle, and vice-versa. It reverts all additive quantum numbers such as electric charge, flavor, baryon number and lepton number, leaving unchanged the mass, the momentum, the angular momentum and the spin. In quantum mechanics this corresponds to the charge conjugation operator C , which is unitary and Hermitian, and thus idempotent ($C^2 = 1$). Consequently, the possible eigenvalues (charge parity) are $\eta_C = \pm 1$. Only particles for which all additive quantum numbers are zero, i.e. particles that are their own antiparticles, are eigenstates of C . Examples are the photon γ and the neutral pion π^0 .

The charge parity of the photon is $\eta_C(\gamma) = -1$ as it is described by a vector potential A produced by charges and currents which change sign under charge conjugation². A neutral pion decays electromagnetically into two photons and thus $\eta_C(\pi^0) = +1$, since $C|\pi^0\rangle = (-1)(-1)|\gamma\gamma\rangle$. In the case of charged pions, they are transformed by C into each other, i.e. $C|\pi^\pm\rangle = |\pi^\mp\rangle$.

In general, eigenstates of C can be formed from linear combinations of particle and antiparticle states, e.g. a fermion-antifermion bound state $|f\bar{f}\rangle$. Application of charge conjugation to this system exchanges effectively the two particles. The corresponding eigenvalue is given

²This argument leads to the right charge parity of the photon. However, a sufficient argument requires to consider the effect of C on a massless spin-1 particle in QFT (see e.g. [15, chapter 5]).

by $(-1) \cdot (-1)^L \cdot (-1)^{S+1} = (-1)^{L+S}$, where L and S are the quantum numbers for the total orbital angular momentum and the total spin. The first factor (-1) results from the swap of the fermion charges due to Fermi-Dirac symmetry, the second factor $(-1)^L$ from the swap of the positions, and the third factor $(-1)^{S+1}$ from the swap of the spin states. For a two-boson system, the eigenvalue is $(-1)^L(-1)^S = (-1)^{L+S}$, where the $(-1)^S$ factor results from the spin exchange for bosons. In the case of two identical neutral bosons like photons or neutral pions, L must be even and thus the charge parity is always $+1$. Since $\eta_C(\pi^0) = +1$, the charge parity of any system composed only of neutral pions is $+1$.

Charge conjugation symmetry requires that particles and antiparticles behave identically within fundamental interactions. This means that any process should occur with the same rate as its respective charge conjugated process. The only fundamental force that has been found to violate C symmetry, is the weak force. Just after the discovery of P symmetry violation in weak interactions, Lee and Yang together with Reinhard Oehme [18] indicated that charge conjugation symmetry is also violated in nature. Their argument was based on the fact that the combined CPT symmetry, which will be introduced in Sect. 2.2.4, has to be conserved.

2.2.3 Time reversal

The time reversal T transforms the time t as $t \rightarrow -t$. This means that it transforms a process into one which is occurring backwards in time (time-reverted). Position and charge of a particle remain unchanged, while the direction of motion and the angular momentum change sign under T .

While T symmetry is apparently violated at the macroscopic scale due to the second law of thermodynamics (entropy increases with time), at the microscopic scale, T symmetry is found to be conserved in strong and in electromagnetic interactions. The first hint of T violation at the microscopic scale was the discovery of CP violation in weak decays: in order to conserve the combined symmetry CPT , T symmetry has to be violated if CP symmetry is violated. To directly proof T violation one has for example to measure that a given reaction and its respective inverse reaction occur at different rates. As it will be discussed in Sect. 2.2.5, CP violation was discovered in decays of neutral kaons to final states containing only pions. These final states do not allow to unambiguously determine the final flavor of the kaon, i.e. if the final-state kaon is a K^0 or a \bar{K}^0 . A way to determine the final flavor is to reconstruct flavor-specific semileptonic decays $K^0 \rightarrow e^+\nu_e\pi^-$ ($\bar{K}^0 \rightarrow e^-\bar{\nu}_e\pi^+$) which have a small branching fraction of order 10^{-4} .

In 1998, the CPLEAR experiment at CERN succeed in measuring for the first time an asymmetry between the rates of the transitions $K^0 \rightarrow \bar{K}^0$ and $\bar{K}^0 \rightarrow K^0$ using semileptonic decays [19]. In this experiment, the initial neutral kaon was produced strongly together with an accompanying charged kaon (associated $s\bar{s}$ production); the flavor of the initial neutral kaon could be unambiguously determined by the charge of the charged kaon.

The observation of T symmetry violation at CPLEAR is the only one considered as a direct evidence [20]. Other observations of T asymmetries in $K_L \rightarrow \pi^+\pi^-e^+e^-$ decays at the KTEV

experiment [21], and in the B -meson sector at BaBar [22] are considered to be doubtful as direct evidence [23, 24]. The violation of T symmetry measured by the CPLEAR experiment, however, is equivalent to the violation of CP symmetry in mixing (see Sect. 2.5 and 2.6). At present, there is no evidence of T symmetry violation which is uncorrelated with the CP violation in weak interactions: a violation of T symmetry involving a conservation of CP symmetry would necessarily violate CPT symmetry.

2.2.4 CPT symmetry

The CPT transformation is the product of the three discrete transformations introduced in the previous sections: charge conjugation C , parity transformation P and time reversal T . The CPT transformation is connected with the CPT theorem. This theorem states that any QFT that is invariant under Lorentz transformations automatically conserves symmetry under CPT transformations (the order of the C , P and T operators is irrelevant). In simplified words, if Lorentz invariance is given, every conceivable Hermitian Hamiltonian commutes with the CPT operator, implying that particles and their respective antiparticles have exactly the same mass and the same lifetime [15, 25]. If CPT symmetry is conserved, every process involving particles should occur with the same rate as the parity-reflected time-reversed process involving antiparticles.

The first proof of the CPT theorem was provided in 1954 by Gerhart Lüders [26], and then generalized one year later by Pauli [27]. When the violation of P symmetry was discovered in weak interactions (see Sect. 2.2.1), the CPT theorem gained relevance, since, based on it, Lee, Yang and Oehme [18] showed that C symmetry is also violated in weak interactions. After the discovery of CP violation, it became clear that T symmetry is violated if CPT symmetry is conserved. Meanwhile, the violation of T symmetry has been also observed (see Sect. 2.2.3).

CPT invariance can be tested for example by measuring the mass and the lifetime of a particle and of its respective antiparticle. Any difference would be an evidence of CPT violation. Several experiments have searched for such differences between K^0 and \bar{K}^0 mesons [28], and between B^0 and \bar{B}^0 mesons [29]. At present, there is no evidence of CPT violation.

2.2.5 CP symmetry

The product transformation of charge conjugation C and parity inversion P corresponds to a CP transformation. Experiments testing P symmetry in the mid 1950's indicated that C and P symmetries are not conserved in weak interactions. However, they left open the possibility that the combined CP symmetry is conserved. And for some years, it seemed that CP symmetry was generally conserved until 1964, when James Harlowe Christenson, James Watson Cronin, Val Logsdon Fitch and René Turlay observed for the first time the violation of CP symmetry in the neutral kaon system [30].

A neutral kaon $K^0(d\bar{s})$ and its antiparticle $\bar{K}^0(\bar{d}s)$ are flavor eigenstates. This means that they can be distinguished by their flavor quantum number, corresponding in this case to their strangeness. Linear combinations of the flavor eigenstates K^0 and \bar{K}^0 form the eigenstates of the weak interaction, K_1^0 and K_2^0 , which are states of definite mass and lifetime (mass eigen-

states). After the discovery of P and C violation, it was assumed that the weak eigenstates K_1^0 and K_2^0 were CP eigenstates, since the first one had been observed to decay to two pions in s -wave ($L = 0$), i.e. to a CP even state, while the second had been observed to decay to three pions, i.e. to a CP odd state. Under this assumption, the weak (mass) eigenstates are

$$\begin{aligned} |K_1^0\rangle &= \frac{1}{\sqrt{2}} (|K^0\rangle + |\bar{K}^0\rangle) & CP|K_1^0\rangle &= +|K_1^0\rangle, \\ |K_2^0\rangle &= \frac{1}{\sqrt{2}} (|K^0\rangle - |\bar{K}^0\rangle) & CP|K_2^0\rangle &= -|K_2^0\rangle. \end{aligned}$$

The mass of three pions of about $420 \text{ MeV}/c^2$ is slightly below the mass of the kaon of about $500 \text{ MeV}/c^2$. Therefore, the kinematic phase space available for the decay $K_2^0 \rightarrow \pi\pi\pi$ is considerably smaller than for the decay $K_1^0 \rightarrow \pi\pi$, resulting in a mean lifetime of K_2^0 (51.16 ns) of about three orders of magnitude larger than that of K_1^0 (89.56 ps) [28]. The mass eigenstates are called therefore short (S) and long (L): $K_1^0 \equiv K_S^0$ and $K_2^0 \equiv K_L^0$.

Around 1964, two groups, one from Princeton [30] and one from Illinois [31], performed experiments to set a lower limit on the decay of a K_2^0 into two pions. Both experiments were run at the Brookhaven Alternating Gradient Synchrotron (AGS). They prepared neutral kaon beams by hitting fixed targets with protons. Neutral kaons were produced via strong interactions in an admixture of K_1^0 and K_2^0 . Both experiments made use of the large lifetime difference between the two mass eigenstates by placing their detectors at distances from the target corresponding to more than 500 lifetimes of the K_1^0 . The detectors were used to measure the three momentum of the charged products, assuming that they had the mass of the charged pion. Both groups reconstructed kaon candidates using pairs of oppositely charged particles with an invariant mass $m_{\pi^+\pi^-}$ in the range of the kaon mass. Both found surprisingly that the decay $K_2^0 \rightarrow \pi^+\pi^-$ in fact occur. The Princeton group of Cronin and Fitch observed the decay two weeks before the Illinois group and gave a more precise estimation of the branching fraction $\mathcal{B}(K_2^0 \rightarrow \pi^+\pi^-/K_2^0 \rightarrow \text{all charged modes}) = (2.0 \pm 0.4) \cdot 10^{-3}$. For their discovery of CP violation Cronin and Fitch received the Nobel Prize in 1980.

The conclusion after the CP experiments was that CP symmetry is an approximative symmetry, i.e. it is only slightly violated in weak interactions (the effect is of order $\mathcal{O}(10^{-3})$), while the individual C and P symmetries are maximally violated. The weak eigenstates are thus not exactly CP eigenstates. However, they can be parametrized in terms of the CP eigenstates introducing a small complex mixing parameter ε :

$$\begin{aligned} |K_S^0\rangle &= \frac{1}{\sqrt{1+|\varepsilon|^2}} (|K_1^0\rangle + \varepsilon|K_2^0\rangle) = \frac{1}{\sqrt{2(1+|\varepsilon|^2)}} ((1+\varepsilon)|K^0\rangle + (1-\varepsilon)|\bar{K}^0\rangle), \\ |K_L^0\rangle &= \frac{1}{\sqrt{1+|\varepsilon|^2}} (\varepsilon|K_1^0\rangle + |K_2^0\rangle) = \frac{1}{\sqrt{2(1+|\varepsilon|^2)}} ((1+\varepsilon)|K^0\rangle - (1-\varepsilon)|\bar{K}^0\rangle). \end{aligned}$$

The parameter ε is related to the so-called CP violation in mixing (see Sect. 2.6) and can be measured by comparing the decay rate of the K_L^0 with the flavor-specific semileptonic decays $K^0 \rightarrow e^+\nu_e\pi^-$ ($\bar{K}^0 \rightarrow e^-\bar{\nu}_e\pi^+$). The current value is $|\varepsilon| = (2.228 \pm 0.011) \cdot 10^{-3}$ [28].

2.3 The Cabibbo angle and the GIM mechanism

About a decade before the discovery of CP violation, the strangeness quantum number had been introduced to describe those particles whose weak decay rates were much smaller than predicted by the weak theory of Enrico Fermi [32] (Nobel Prize, 1938). Fermi's theory can be regarded as an effective theory in which the interaction mediated by a W boson is described as a point interaction of four fermions with a universal coupling constant $G_F \equiv \sqrt{2}g^2/(8m_W^2)$. Fermi's theory provides a good description of low energy phenomena, such as β decays, if one takes into account the violation of discrete symmetries. For high energies, Fermi's theory leads to inconsistent predictions.

Experiments in the 1950's and 1960's were still far below the electroweak scale. The striking observation at that time was that the decay rates of strange particles to leptonic final states, such as $K^+(u\bar{s}) \rightarrow \mu^+\nu_\mu$, were considerably smaller than the rates for similar decays of non-strange particles, such as $\pi^+(u\bar{d}) \rightarrow \mu^+\nu_\mu$. To explain this, Nicola Cabibbo suggested in 1963 that the weak current of hadronic particles J_μ is a linear combination of a strangeness conserving current $J_\mu^{(0)}$ and a strangeness changing current $J_\mu^{(1)}$ [33]

$$J_\mu = \cos \theta \cdot J_\mu^{(0)} + \sin \theta \cdot J_\mu^{(1)},$$

where he introduced a mixing angle θ , which became to be called Cabibbo angle θ_C , to conserve the norm of the total current. The coupling constant g had to be corrected to $g \cdot \sin \theta_C$ for decays of strange hadrons, and to $g \cdot \cos \theta_C$ for decays of non-strange hadrons. By comparing the decay rates of several channels measured at different experiments, Cabibbo found a consistent value of θ_C around 0.26 rad. Additionally, his hypothesis could explain the difference between the decay rates of β decays, and the decays rates of muon decays. In terms of the quark model, the Cabibbo angle corresponds to a mixing angle between quark flavors. Thus, the weak interaction couples to the linear combination

$$|d'\rangle = \cos \theta_C \cdot |d\rangle + \sin \theta_C \cdot |s\rangle.$$

Another striking effect in the strange sector was the suppression of flavor changing neutral currents, such as the decay $K^0 \rightarrow \mu^+\mu^-$. Around 1970, the experimental limit for the branching fraction of this decay was 10^{-4} [34]. Today, the measured value is $6.84 \cdot 10^{-9}$ [28]. This indicates that there are no first order contributions to this process. However, this decay is possible as a second-order loop processes, as Fig. 2.3 shows.

Considering only the diagram with the up quark in the loop, the expected branching fraction is several orders of magnitude larger than the measured branching fraction. In 1970, Glashow together with John Iliopoulos and Luciano Maiani noticed that such processes could be sup-

pressed if there was an additional quark with the same charge as the up quark, but with a new flavor quantum number, which they called charm [35]. They proposed a four-quark model, which became to be called Glashow-Iliopoulos-Maiani (GIM) model, with the three quarks known at that time and the new charm quark. In their new model, the mixing angle introduced by Cabibbo was extended to a rotation matrix \mathbf{V}_C (Cabibbo matrix) mixing the down and the strange quarks:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \mathbf{V}_C \begin{pmatrix} d \\ s \end{pmatrix} = \begin{pmatrix} \cos \theta_C & \sin \theta_C \\ -\sin \theta_C & \cos \theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \quad (2.1)$$

Consequently, in weak interactions, the u quark couples to d' and the charm quark couples to s' . This model gives rise to two important effects. First, since the Cabibbo matrix \mathbf{V}_C is unitary, the interacting states d' and s' are orthogonal to each other, explaining why first-order flavor-changing neutral currents do not exist. Second, in the higher-order processes, additional diagrams appear with a charm quark instead of an up quark inside the loop, such as in Fig. 2.3 (right). Up and charm diagrams interfere destructively due to the minus sign in the Cabibbo matrix; just a small contribution remains due to the mass difference between the up and the charm quarks, explaining why processes such as $K^0 \rightarrow \mu^+ \mu^-$ are highly suppressed. This way of suppressing flavor-changing neutral currents is referred to as GIM mechanism.

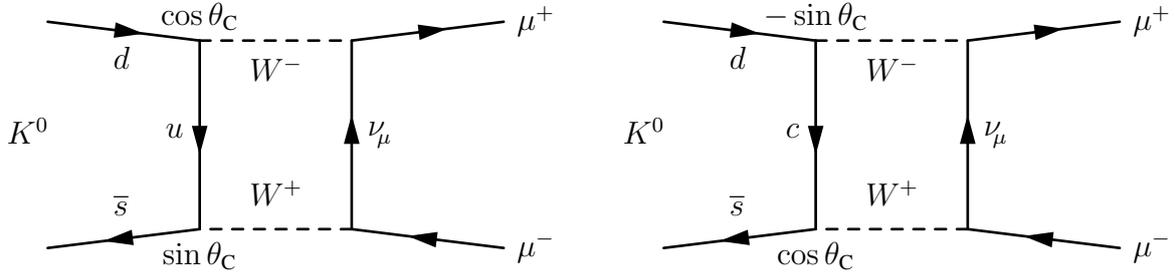


Figure 2.3: Feynman diagrams contributing to the process $K^0 \rightarrow \mu^+ \mu^-$: the diagram on the left (with an up quark in the loop) and the diagram on the right (with a charm quark in the loop) cancel each other almost exactly, resulting in a large suppression of the decay.

2.4 The CKM matrix

As observed in nature, CP symmetry is violated in the SM by the charged weak interactions of quarks. The charged-current Lagrangian describing the coupling of charged W bosons to quarks is given by [36]

$$\begin{aligned} \mathcal{L}_{cc} &= -\frac{g}{\sqrt{2}} \left(\bar{u}_i \gamma^\mu \frac{1 - \gamma^5}{2} \mathbf{V}_{ij} d_j W_\mu^+ + \bar{d}_i \gamma^\mu \frac{1 - \gamma^5}{2} \mathbf{V}_{ij}^\dagger u_j W_\mu^- \right) \\ \xrightarrow{CP} \mathcal{L}_{cc}^{CP} &= -\frac{g}{\sqrt{2}} \left(\bar{d}_j \gamma^\mu \frac{1 - \gamma^5}{2} \mathbf{V}_{ij} u_i W_\mu^- + \bar{u}_j \gamma^\mu \frac{1 - \gamma^5}{2} \mathbf{V}_{ij}^\dagger d_i W_\mu^+ \right), \end{aligned}$$

where the triplet of up-type quarks u is coupled through the Dirac matrices γ^μ to the triplet of down-type quarks d . Here, the chirality operator P_L acts on a Dirac field ψ as

$$P_L \psi = \frac{1 - \gamma_5}{2} \psi = \psi_L \quad \Rightarrow \quad \bar{\psi} \gamma^\mu \frac{1 - \gamma_5}{2} \psi = \bar{\psi}_L \gamma^\mu \psi_L,$$

projecting the left-handed components of the quark triplets. Since γ^μ is a vector (V) operator and $\gamma^\mu \gamma^5$ is an axial vector (A) operator, the resulting weak operator $\gamma^\mu P_L$ is a V–A superposition. The V–A structure of the weak interactions violates maximally P and C symmetries, since it allows only the left-handed chirality components of fermions (right-handed components of antifermions) to participate in weak interactions.

The matrix V in the charged-current Lagrangian \mathcal{L}_{cc} is the generalized version of the Cabibbo matrix V_C in eq. (2.1). It is a unitary matrix which mixes the down-type quarks of the different generations. The Lagrangian \mathcal{L}_{cc} is invariant under CP transformations if $V_{ij}^\dagger = V_{ji}$, i.e. if it is real. Thus, in order to include the observed violation of CP symmetry, the mixing matrix has to contain at least one irreducible complex phase. A complex phase is irreducible if it cannot be removed by adjusting the phase of the quark fields (see next section).

In 1973, Makoto Kobayashi and Toshihide Maskawa noticed that the mixing matrix could bear an irreducible complex phase if there are at least three generations of quarks. They proposed to incorporate a mixing matrix for three generations into the electroweak theory, predicting at that time three new quarks. This matrix became to be called the Cabibbo-Kobayashi-Maskawa matrix V_{CKM} :

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

In the expression above, the physical quarks (d, s, b) are called mass eigenstates, while the rotated quark states (d', s', b') are called flavor eigenstates. Fig. 2.4 shows an example illustrating the effect of a CP transformation on a weak interaction vertex: the mixing matrix coefficient in the coupling becomes complex conjugated.

2.4.1 Parameters in a mixing matrix

In general, a unitary $n \times n$ matrix V contains n^2 free parameters. An easy way to see this is to write V as e^{iH} , where H is a Hermitian matrix. Both V and H have equal number of free parameters. The parameters of H can be counted easily: n real values in the diagonal, $\frac{1}{2}n(n-1)$ real values above the diagonal and $\frac{1}{2}n(n-1)$ imaginary values above the diagonal, yielding in total n^2 free parameters.

Considering now the matrix V , unitarity requires that the column vectors are normalized and orthogonal to each other. Thus, in the first column, there are $(n-1)$ degrees of freedom for the magnitude of the entries; in the second column, there are $(n-2)$, and so forth, giving

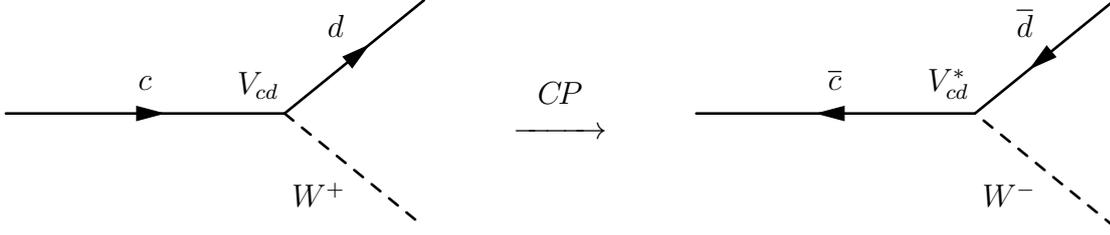


Figure 2.4: For the decay of the charm quark, the coupling to a down quark is proportional to the mixing matrix coefficient V_{cd} . In the charge-parity conjugated decay (right), the complex conjugated coefficient V_{cd}^* appears.

in total $\frac{1}{2}n(n-1)$ real parameters, which are expressed commonly as mixing angles, i.e. generalized Cabibbo angles. The remaining $\frac{1}{2}n(n+1)$ degrees of freedom of \mathbf{V} correspond to complex phases.

From the physical point of view, not all of these phases are observable. The Lagrangian \mathcal{L}_{cc} is invariant under simultaneous phase transformations of the quark-mass-eigenstates fields such as

$$u_i \rightarrow e^{i\phi_i} u_i, \quad d_j \rightarrow e^{i\phi_j} d_j, \quad V_{ij} \rightarrow e^{i(\phi_i - \phi_j)} V_{ij}.$$

The phase differences $(\phi_i - \phi_j)$ can be chosen such that the transformation above eliminates $(2n-1)$ phases of the total number of phases in \mathbf{V} [36]. This can be seen in the following way: there are n generations of quarks with $2n$ quarks in total. Since the effect of one single global phase for all quarks does not change the properties of \mathbf{V} , there are $2n-1$ degrees of freedom that can be eliminated by adjusting the phase factors of the quark fields. The net number of irreducible complex phases with potential to cause CP violation is then

$$n_{CP} = \frac{n(n+1)}{2} - (2n-1) = \frac{1}{2}(n-2)(n-1).$$

If $n=1$, or $n=2$, this number vanishes. If $n=3$, i.e. in the six-quark model proposed by Kobayashi and Maskawa, there is exactly one irreducible complex phase in the mixing matrix. Thus, the CKM matrix entails only four free parameters, three real parameters (rotation angles) and one complex phase.

2.4.2 Parametrization of the CKM matrix

Depending on the definition of the quark-field phases and the order of rotations, the CKM matrix can be parametrized in different ways leading to the same physical results. A widely

used parametrization first proposed by [37], but following the notation in [36], is given by

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}c_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \quad (2.2)$$

where $c_{ij} \equiv \cos \theta_{ij}$ and $s_{ij} \equiv \sin \theta_{ij}$ for $i, j = 1, 2, 3$. The phase δ is the irreducible complex phase causing CP violation unless $\delta \in \{0, \pi\}$. One advantage of this parametrization is that the three angles $\theta_{ij} \in [0, \frac{\pi}{2}]$ correspond to mixing angles between the quark generations i and j . If one angle vanishes, i.e. $\theta_{ij} = 0$, the generations i and j are decoupled. In the limit $\theta_{23} = \theta_{13} = 0$, the angle θ_{12} is equal to the Cabibbo angle θ_C .

For phenomenological analyses, especially for B physics, a useful and widely spread parametrization is the one proposed by Lincoln Wolfenstein [38]. He expanded the CKM matrix in powers of a parameter $\lambda = \sin(\theta_{12}) \approx 0.22$, where the angle θ_{12} corresponds to the Cabibbo angle θ_C , the largest mixing angle in the CKM matrix. Wolfenstein noted that $|V_{cb}| \approx \mathcal{O}(\lambda^2)$ and introduced four parameters λ , A , ρ and η , such that

$$\begin{aligned} \sin \theta_{12} = \lambda, & \quad \Rightarrow \quad \cos \theta_{12} = \sqrt{1 - \lambda^2}, \\ \sin \theta_{23} = A\lambda^2, & \quad \Rightarrow \quad \cos \theta_{23} = \sqrt{1 - A^2\lambda^4}, \\ \sin \theta_{13} e^{-i\delta} = A\lambda^3(\rho - i\eta), & \quad \Rightarrow \quad \cos \theta_{13} = \sqrt{1 - A^2\lambda^3(\rho^2 + \eta^2)}, \end{aligned}$$

where A , ρ and η are real parameters in the range $(0, 1)$. If $\eta \neq 0$, CP symmetry is violated. Up to $\mathcal{O}(\lambda^3)$, the CKM matrix is given by

$$\mathbf{V}_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3[1 - \rho - i\eta] & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (2.3)$$

The main advantage of this so-called Wolfenstein's parametrization is that it makes directly visible the nearly diagonal structure of the matrix and the hierarchy between generations: the largest coupling ($\mathcal{O}(1)$) is the one between quarks of the same generation; the second largest ($\mathcal{O}(\lambda)$) between the first and the second generation; the third largest ($\mathcal{O}(\lambda^2)$) between between the second and the third generation; and the smallest ($\mathcal{O}(\lambda^3)$) between the first and the third generation. The CP violation parameter η appears only in the smallest parameter as well as in other higher order terms. Especially for K -meson physics, one has to be aware that η enters in V_{cd} and V_{cs} at higher order. This reflects that the CP violation in the SM is a very small effect.

2.4.3 The unitarity triangle

The model with three quark generations proposed by Kobayashi and Maskawa is the minimal possible model that includes CP violation. It can be tested by measuring the individual coefficients of the CKM matrix and proofing if the resulting matrix fulfills unitarity.

The unitarity condition gives rise to twelve distinct complex relations among the matrix coefficients:

$$\mathbf{V}_{\text{CKM}}^\dagger \mathbf{V}_{\text{CKM}} = \mathbb{1} \Rightarrow \sum_{j \in \{u,c,t\}} V_{ji}^* V_{jk} = \delta_{ik} \text{ with } i, k \in \{d, s, b\}, \quad (2.4)$$

$$\mathbf{V}_{\text{CKM}} \mathbf{V}_{\text{CKM}}^\dagger = \mathbb{1} \Rightarrow \sum_{j \in \{d,s,b\}} V_{ij} V_{kj}^* = \delta_{ik} \text{ with } i, k \in \{u, c, t\}. \quad (2.5)$$

Six of the relations above, those with $i \neq k$, can be represented geometrically as triangles in the complex plane. They are called unitarity triangles. Their interior angles represent observable quantities that measure CP violation. The geometrical consequence of CP violation is that these triangles are not degenerate into lines, i.e. that the three vectors describing the sides cannot be made real. All triangles have the same area, commonly written as $J_{CP}/2$, where J_{CP} is the so-called Jarlskog parameter, which is a parametrization independent and universal measure of CP violation in the SM [39]. Using the parametrization in eq. (2.2), the Jarlskog parameter depends on the CP violating phase as $J_{CP} \propto \sin \delta$. In the SM, all CP violation phenomena are correlated since the CP violation is generated by the same irreducible complex phase.

Using Wolfenstein's parametrization (see previous subsection), the sides of the six triangles can be expressed in powers of λ . From the unitarity condition in eq. (2.4), one obtains three triangles which are relevant for the physics of K , B and B_s mesons. The sides of these triangles contain CKM coefficients related to processes involving the aforementioned mesons. The K triangle is obtained by multiplying columns 1 and 2 of the CKM matrix; the B triangle by multiplying columns 1 and 3; and the B_s triangle by multiplying columns 2 and 3. The triangles are

$$K \text{ triangle: } V_{ud}V_{us}^* + V_{cd}V_{cs}^* + V_{td}V_{ts}^* = 0 \Rightarrow \mathcal{O}(\lambda) + \mathcal{O}(\lambda) + \mathcal{O}(\lambda^5) = 0, \quad (2.6)$$

$$B \text{ triangle: } V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \Rightarrow \mathcal{O}(\lambda^3) + \mathcal{O}(\lambda^3) + \mathcal{O}(\lambda^3) = 0, \quad (2.7)$$

$$B_s \text{ triangle: } V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0 \Rightarrow \mathcal{O}(\lambda^4) + \mathcal{O}(\lambda^2) + \mathcal{O}(\lambda^2) = 0. \quad (2.8)$$

By comparing the orders of magnitude of the triangle sides, it is evident that the B triangle has sides of about the same size (angles of similar size), while the K and the B_s triangles are almost degenerate to lines (one angle is close to zero). For the K triangle one side is four orders of magnitude smaller than the other sides, and for the B_s triangle one side is two orders of magnitude smaller than the others. Thus, the CP violation effects in the B -meson system are expected to be larger than in the other systems. This has been one of the main motivations

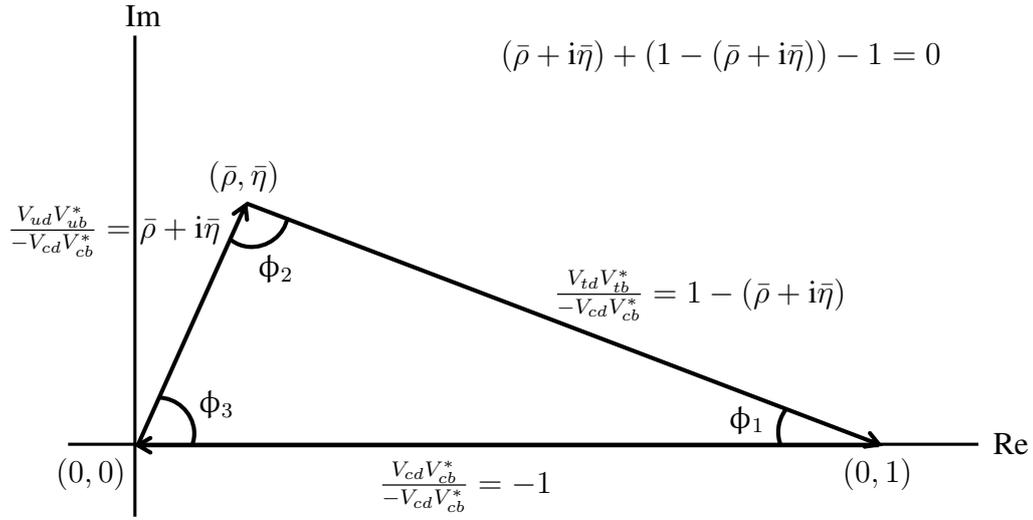


Figure 2.5: Visualization of the unitarity triangle for the B -meson system (eq. (2.7)) in the complex plane. The sides are normalized by $-V_{cd}V_{cb}^*$, fixing two corners of the triangle at $(0, 0)$ and $(1, 0)$.

for the construction of experiments focused on exploring the physics of B -meson decays.

Fig. 2.5 shows a visualization of the B triangle. By convention, all sides are divided by $-V_{cd}V_{cb}^*$, such that one side is real with unit length. This fixes two of the corners of the triangle. The position of the third corner is determined by the parameters

$$\bar{\rho} = \rho \left(1 - \frac{\lambda^2}{2} + \dots \right), \quad \text{and} \quad \bar{\eta} = \eta \left(1 - \frac{\lambda^2}{2} + \dots \right),$$

which can be expanded to different orders in powers of λ . In the same way, the area below the triangle can be written as

$$\frac{J_{CP}}{2} = \frac{\lambda^6 A^2 \bar{\eta}}{2} + \dots$$

The inner angles of the triangle, the so-called unitarity angles, are defined as

$$\phi_1 \equiv \beta \equiv \arg \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right) = \arg \left(\frac{1}{1 - \bar{\rho} - i\bar{\eta}} \right), \quad (2.9)$$

$$\phi_2 \equiv \alpha \equiv \arg \left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right) = \arg \left(-\frac{1 - \bar{\rho} - i\bar{\eta}}{\bar{\rho} + i\bar{\eta}} \right), \quad (2.10)$$

$$\phi_3 \equiv \gamma \equiv \arg \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right) = \arg (\bar{\rho} + i\bar{\eta}), \quad (2.11)$$

where the angle ϕ_3 coincides to a good approximation with the phase δ of the parametrization in eq. (2.2). In the literature, the unitarity angles are usually called β , α , and γ . However, in this thesis, the angles are called ϕ_1 , ϕ_2 and ϕ_3 since this is the convention at Belle and at Belle II.

The position of the corner $(\bar{\rho}, \bar{\eta})$ can be determined by measuring the three angles and the length of the two non-normalized sides. Since the normalized triangle is unambiguously defined by two of these parameters, the system is evidently over-constrained, offering an excellent test of the SM: if the model is correct and complete, the unitarity condition of the CKM matrix must be fulfilled, implying that the triangle closes.

2.5 Flavor mixing of flavored neutral mesons

Flavored neutral mesons have a characteristic flavor quantum number referring to their quark compositions: “strangeness”, related to s quarks; “charm”, related to c quarks; and “beauty”, related to b quarks. By convention, the sign of the flavor quantum number agrees with the sign of the electric charge of the quark determining the flavor. For example, $K^0(d\bar{s})$, $D^0(\bar{u}c)$ and $B^0(d\bar{b})$ have positive flavor, while their antiparticles $\bar{K}^0(\bar{d}s)$, $\bar{D}^0(u\bar{c})$ and $\bar{B}^0(\bar{d}b)$ have negative flavor. In the following, the flavor eigenstate of a flavored neutral meson with flavor positive will be denoted by $|M\rangle$. Since $|M\rangle$ is not its own antiparticle, it is not an eigenstate of the charge conjugation operator:

$$C|M\rangle = |\bar{M}\rangle.$$

Flavor quantum numbers are conserved in strong interactions and thus the flavor eigenstates are simultaneously eigenstates of strong interactions. In contrast, weak interactions do not conserve flavor quantum numbers allowing $|M\rangle$ to undergo a transition into $|\bar{M}\rangle$ (or vice versa), changing the flavor quantum number by two units. Since flavored neutral mesons are subjected to oscillations in the flavor space, the mass eigenstates are superpositions of the pure flavor eigenstates $|M\rangle$ and $|\bar{M}\rangle$. This phenomenon is the so-called flavor mixing. As an example, Figure 2.6 shows the leading order Feynman diagrams contributing to neutral $B^0 - \bar{B}^0$ mixing.

A flavored neutral meson can be considered as a two-state quantum system with the generic wave function:

$$|\psi(t)\rangle = \begin{pmatrix} a(t) \\ b(t) \end{pmatrix} = a(t)|M\rangle + b(t)|\bar{M}\rangle.$$

The corresponding time evolution is approximated by the time-dependent Schrödinger equation (Weisskopf-Wigner approximation)

$$i\hbar \frac{d}{dt} |\psi\rangle = \mathcal{H}_{\text{eff}} |\psi\rangle,$$

where the effective Hamiltonian \mathcal{H}_{eff} is a 2×2 matrix that can be decomposed into a mass and a decay matrix

$$\mathcal{H}_{\text{eff}} = M - \frac{i}{2}\Gamma,$$

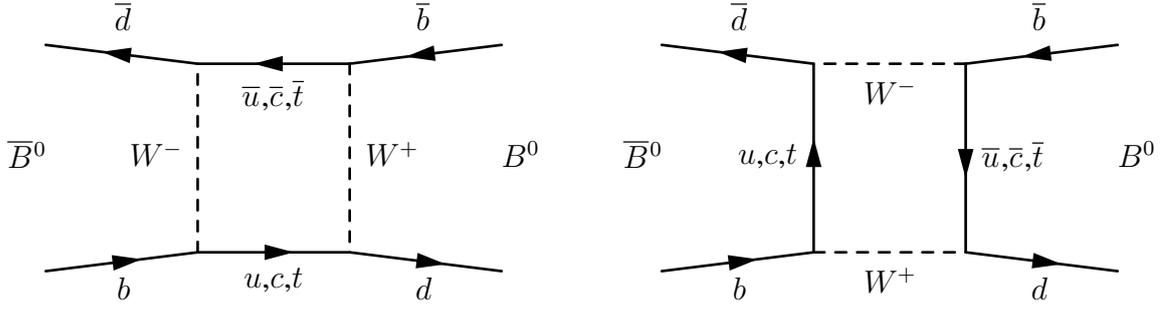


Figure 2.6: Leading order Feynman diagrams contributing to neutral $B^0 - \bar{B}^0$ mixing. A B^0 meson can transform into a \bar{B}^0 and viceversa by the exchange of two W -bosons and a virtual $u, c, \text{ or } t$ quark.

where M and Γ are hermitian matrices. The effective Hamiltonian \mathcal{H}_{eff} is not hermitian and thus probability is not conserved. This corresponds to the decrease in the total number of M and \bar{M} mesons since the mesons decay as time evolves. The diagonal elements of M and Γ are related respectively to the mass and the lifetime of the pure flavor eigenstates. Conservation of CPT symmetry (see Sect. 2.2.4) requires that the M and the \bar{M} mesons have the same mass and the same lifetime implying that $M_{11} = M_{22} = M$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$. These diagonal elements are determined by the quark masses and by strong and electromagnetic (EM) interactions. Thus, the diagonal elements of \mathcal{H}_{eff} can be identified as the strong and electromagnetic (EM) Hamiltonian \mathcal{H}_0 . On the other hand, the off-diagonal elements have to fulfill $M_{12} = M_{21}^*$ and $\Gamma_{12} = \Gamma_{21}^*$ due to hermiticity. These elements involve weak interactions and can be merged into a weak perturbation Hamiltonian \mathcal{H}_w . One obtains

$$\mathcal{H}_{\text{eff}} = \begin{pmatrix} H_0 & 0 \\ 0 & H_0 \end{pmatrix} + \mathcal{H}_w = \begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{pmatrix}. \quad (2.12)$$

Since the phase factors of $|M\rangle$ and $|\bar{M}\rangle$ can be arbitrarily adjusted without changing the physics results, only the difference between them is relevant. The phase difference can be quantified by a complex weak phase ϕ_{12} . In total, the flavor mixing can be parametrized by five real parameters, by convention, these are [40]

$$M_{11}, \quad \Gamma_{11}, \quad |M_{12}|, \quad |\Gamma_{12}| \quad \text{and} \quad \phi_{12} = \arg\left(-\frac{M_{12}}{\Gamma_{12}}\right).$$

By diagonalizing the Hamilton operator in eq. (2.12), the time evolution of the meson-antimeson system can be described in terms of its eigenvalues and eigenstates. The eigenvalues ω_{\pm} , are given by

$$\omega_{\pm} = M - \frac{i}{2}\Gamma \pm \sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)} = M - \frac{i}{2}\Gamma \pm pq. \quad (2.13)$$

The two associated eigenstates $v_{\pm} = (1, \pm \frac{q}{p})^T$ represent the physical mass eigenstates

$$\begin{aligned} |M_+(t)\rangle &\equiv p|M(t)\rangle + q|\bar{M}(t)\rangle, \\ |M_-(t)\rangle &\equiv p|M(t)\rangle - q|\bar{M}(t)\rangle. \end{aligned} \quad (2.14)$$

The complex numbers p and q satisfy

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}} = e^{-i\phi_M} \sqrt{\frac{|M_{12}| + \frac{i}{2}|\Gamma_{12}|e^{i\phi_{12}}}{|M_{12}| + \frac{i}{2}|\Gamma_{12}|e^{-i\phi_{12}}}} \quad \text{with } |p|^2 + |q|^2 = 1, \quad (2.15)$$

where the absolute mixing weak phase $\phi_M \equiv \arg(M_{12})$ is introduced. The mass eigenstates $|M_{\pm}(t)\rangle$ are pure CP eigenstates if they are orthogonal to each other (even and odd states). This is the case when the phase ϕ_{12} is zero or an odd multiple of π , i.e. if the absolute value of $|q/p|$ is equal to 1.

The masses m_{\pm} and the lifetimes Γ_{\pm} for the mass eigenstates states can be derived from the eigenvalues in eq. (2.13),

$$\begin{aligned} m_+ &= \text{Re}(\omega_+), & \Gamma_+ &= -2\text{Im}(\omega_+), \\ m_- &= \text{Re}(\omega_-), & \Gamma_- &= -2\text{Im}(\omega_-), \end{aligned}$$

which can be recasted in terms of the following observables

$$\begin{aligned} m &\equiv \frac{1}{2}(m_+ + m_-) = M, & \Gamma &\equiv \frac{1}{2}(\Gamma_+ + \Gamma_-) = \Gamma = \frac{1}{\tau}, \\ \Delta m &\equiv m_- - m_+ = -2\text{Re}(pq), & \Delta \Gamma &\equiv \Gamma_- - \Gamma_+ = 4\text{Im}(pq). \end{aligned}$$

Subsequently, the time evolution of $|M_{\pm}\rangle$ is given by

$$\begin{aligned} |M_+(t)\rangle &= e^{-i\omega_+ t} (p|M\rangle + q|\bar{M}\rangle), \\ |M_-(t)\rangle &= e^{-i\omega_- t} (p|M\rangle - q|\bar{M}\rangle), \end{aligned} \quad (2.16)$$

where both mass eigenstates decay as $e^{-\Gamma_{\pm} t}$ with a modulation generated by the complex phase $e^{im_{\pm}}$. Considering now eq. (2.14), the time evolution of the pure flavor states can be written as

$$\begin{aligned} |M(t)\rangle &\equiv \frac{1}{2p} (|M_+(t)\rangle + |M_-(t)\rangle), \\ |\bar{M}(t)\rangle &\equiv \frac{1}{2q} (|M_+(t)\rangle - |M_-(t)\rangle). \end{aligned}$$

Insertion of the time evolution in 2.16 yields

$$\begin{aligned} |M(t)\rangle &= \frac{1}{2} \left((e^{-i\omega_+t} + e^{-i\omega_-t}) |M\rangle + \frac{q}{p} (e^{-i\omega_+t} - e^{-i\omega_-t}) |\bar{M}\rangle \right) \\ &= \frac{1}{2} e^{-i\omega_+t} \left(\left(1 + e^{-i(\Delta m - \frac{1}{2}\Delta\Gamma)t} \right) |M\rangle + \frac{q}{p} \left(1 - e^{-i(\Delta m - \frac{1}{2}\Delta\Gamma)t} \right) |\bar{M}\rangle \right), \end{aligned} \quad (2.17)$$

$$\begin{aligned} |\bar{M}(t)\rangle &= \frac{1}{2} \left(\frac{p}{q} (e^{-i\omega_+t} - e^{-i\omega_-t}) |M\rangle + (e^{-i\omega_+t} + e^{-i\omega_-t}) |\bar{M}\rangle \right) \\ &= \frac{1}{2} e^{-i\omega_+t} \left(\frac{p}{q} \left(1 - e^{-i(\Delta m - \frac{1}{2}\Delta\Gamma)t} \right) |M\rangle + \left(1 + e^{-i(\Delta m - \frac{1}{2}\Delta\Gamma)t} \right) |\bar{M}\rangle \right). \end{aligned} \quad (2.18)$$

With non vanishing Δm and $\Delta\Gamma$, which are related to the weak off-diagonal components of the Hamiltonian M_{12} and Γ_{12} , an initial pure flavor state will mix with its opposite flavor state before it decays. Thus, decay rates can be only properly defined, if the initial state is specified either as $|M\rangle$ or as $|\bar{M}\rangle$ [41]. Starting with a specific flavor eigenstate $|M\rangle$ at $t = 0$, and taking into account the orthogonality between flavor eigenstates, the probability to observe the state $|\bar{M}\rangle$ at time t is

$$|\langle \bar{M} | M(t) \rangle|^2 = \frac{1}{2} \left| \frac{q}{p} \right|^2 e^{-\Gamma t} \left(\cosh\left(\frac{\Delta\Gamma t}{2}\right) - \cos(\Delta m t) \right), \quad (2.19)$$

and to observe $|M\rangle$ is

$$|\langle M | M(t) \rangle|^2 = \frac{1}{2} e^{-\Gamma t} \left(\cosh\left(\frac{\Delta\Gamma t}{2}\right) + \cos(\Delta m t) \right). \quad (2.20)$$

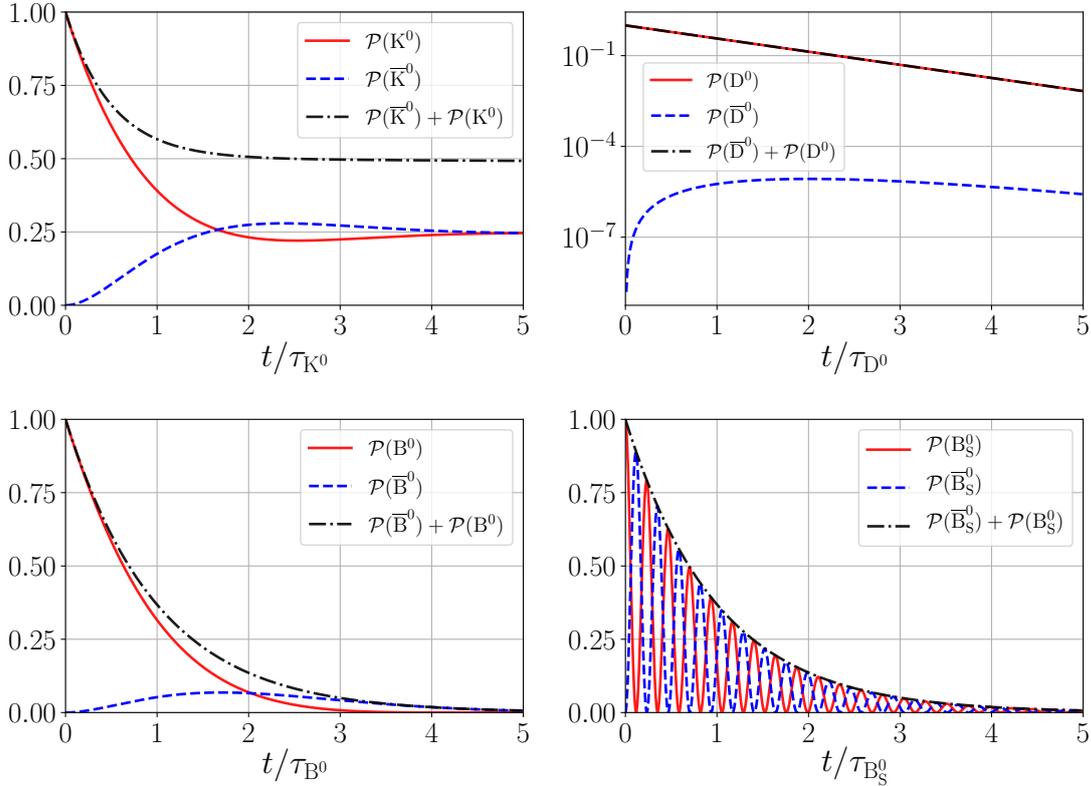
Fig. 2.7 shows flavor oscillations for K^0 , D^0 , B^0 and B_s^0 mesons basing on the measured oscillation parameters in Tab. 2.2. In the B -meson system, the mass eigenstates $|M_\pm\rangle$ are commonly called $|B_{H,L}^0\rangle$, where H stand for heavy and L for light. In the kaon system the convention light and heavy is interchanged with short and long, i.e. $|K_{S,L}^0\rangle$. In the case of K^0 and D^0 mesons, the M_{12} and Γ_{12} are dominated by long-distance contributions making the theoretical estimates for Δm and $\Delta\Gamma$ difficult to compute. For B^0 mesons, the long-distance effects are CKM suppressed simplifying the estimation of Δm and $\Delta\Gamma$ [29].

Historically, flavor mixing was predicted in 1954 by Gell-Mann together with Abraham Pais [43]. They assumed that the mass eigenstates of flavored neutral mesons were C eigenstates. In 1956, a Columbia-Brookhaven group could observe this effect for the first time in the $K^0 - \bar{K}^0$ system [44]. Few months later, the P and the C violation in weak interactions was discovered. And it was thought then that the mass eigenstates were CP eigenstates. Only after the discovery of the CP violation in 1964, it became clear that the neutral kaon mass eigenstates are neither C nor CP eigenstates. In the $B^0 - \bar{B}^0$ system, flavor mixing was discovered in 1987 by the ARGUS collaboration [45]. In the B_s^0 -meson system, flavor mixing

Table 2.2: Oscillation parameters for K^0 , B^0 , B_s^0 and D^0 mesons [28].

Parameter	K^0	D^0
$\Delta m [\hbar \cdot \text{s}^{-1}]$	$(0.5293 \pm 0.0009) \cdot 10^{10}$	$0.95_{-0.44}^{+0.41} \cdot 10^{10}$
$\tau_S = \frac{1}{\Gamma_S} [\text{s}]$	$(0.8954 \pm 0.0004) \cdot 10^{-10}$	$(410.1 \pm 1.5) \cdot 10^{-15}$
$\tau_L = \frac{1}{\Gamma_L} [\text{s}]$	$(5.1116 \pm 0.021) \cdot 10^{-8}$	
$\Delta \Gamma [\text{s}^{-1}]$	$(-1.115 \pm 0.001) \cdot 10^{10}$	$1.29_{-0.18}^{+0.14} \cdot \Gamma$
$x = \frac{\Delta m}{\Gamma}$	0.946 ± 0.002	$0.0046_{-0.0015}^{+0.0014}$ [42]
$y = \frac{\Delta \Gamma}{2\Gamma}$	-0.997 ± 0.001	0.0062 ± 0.0008 [42]

Parameter	B^0	B_s^0
$\Delta m [\hbar \cdot \text{s}^{-1}]$	$(0.5096 \pm 0.0034) \cdot 10^{12}$	$(17.757 \pm 0.021) \cdot 10^{12}$
$\tau = \frac{1}{\Gamma} [\text{s}]$	$(1.520 \pm 0.004) \cdot 10^{-12}$	$(1.510 \pm 0.005) \cdot 10^{-12}$
$\Delta \Gamma [\text{s}^{-1}]$	$(-0.002 \pm 0.010) \cdot \Gamma$ [42]	$(0.082 \pm 0.007) \cdot 10^{12}$
$x = \frac{\Delta m}{\Gamma}$	0.775 ± 0.006	26.81 ± 0.10
$y = \frac{\Delta \Gamma}{2\Gamma}$	< 0.001 [42]	0.062 ± 0.006

Figure 2.7: Mixing probabilities (eqs. (2.19) and (2.20)) for initial pure K^0 (top left), D^0 (top right), B^0 (bottom left) and B_s^0 (bottom right) mesons as functions of their lifetimes.

was identified as an important effect already after the first measurements [46]. However, its mixing frequency could be resolved first in 2006 by the CDF collaboration [47]. Finally, the $D^0 - \bar{D}^0$ mixing was first observed in 2007 at the B factories [29].

2.6 Types of CP Violation

Depending on the way how the CP violation complex phase appears, the different CP violation mechanisms can be classified into three types. For this purpose, one considers the amplitudes of the transitions of a flavored meson $|M\rangle$ and of its antiparticle $|\bar{M}\rangle$, into a final state $|f\rangle$ and into the CP conjugate $|\bar{f}\rangle$,

$$\begin{aligned} A_f &= \langle f | \mathcal{H}_{\text{eff}} | M \rangle = A(|M\rangle \rightarrow |f\rangle), & \bar{A}_f &= \langle f | \mathcal{H}_{\text{eff}} | \bar{M} \rangle = A(|\bar{M}\rangle \rightarrow |f\rangle), \\ A_{\bar{f}} &= \langle \bar{f} | \mathcal{H}_{\text{eff}} | M \rangle = A(|M\rangle \rightarrow |\bar{f}\rangle), & \bar{A}_{\bar{f}} &= \langle \bar{f} | \mathcal{H}_{\text{eff}} | \bar{M} \rangle = A(|\bar{M}\rangle \rightarrow |\bar{f}\rangle), \end{aligned}$$

with the effective Hamiltonian operator \mathcal{H}_{eff} . The different types of CP violation are then:

- a) **CP violation in decay** also called direct CP violation, occurs if

$$\left| \frac{A_f}{\bar{A}_{\bar{f}}} \right| \neq 1 \quad \text{or} \quad \left| \frac{\bar{A}_f}{A_{\bar{f}}} \right| \neq 1. \quad (2.21)$$

This is the only kind of CP violation that is possible for both, neutral mesons and charged mesons. Other types of CP violation exclude charged mesons, since they do not mix with their antiparticles. The phase difference between the decay amplitudes A_f and $\bar{A}_{\bar{f}}$ can be quantified by a complex phase ϕ_D , which is the superposition of the complex weak phases and the strong phases involved in the process. Considering that, in general, different intermediate states contribute to a decay, the total decay amplitudes A_f and $\bar{A}_{\bar{f}}$ can be written as

$$A_f = \sum_i |A_i| e^{i(\delta_i + \phi_i)}, \quad \bar{A}_{\bar{f}} = \sum_i |A_i| e^{i(\delta_i - \phi_i)},$$

where the phases δ_i are strong phases, and the phases ϕ_i are weak phases. Since strong processes do not violate CP symmetry, the strong phases remain invariant under CP transformation. On the other hand, the weak phases are associated with the coefficients of the CKM matrix that appear in the amplitudes of a given decay. Since a CP transformation turns CKM coefficients into their complex conjugate, the weak phases flip sign under CP transformations. The direct CP violation condition in eq. (2.21) can be also formulated as

$$|A_f|^2 - |\bar{A}_{\bar{f}}|^2 = -2 \sum_{i,j} |A_i| |A_j| \sin(\delta_i - \delta_j) \sin(\phi_i - \phi_j) \neq 0.$$

Thus, CP violation in decay can occur only if at least two intermediate states with different strong and weak phases contribute to a decay process. The squared absolute value of the total decay amplitudes $|A_f|^2$ and $|\bar{A}_f|^2$ can be accessed experimentally since they are proportional to the total decay rates, while the individual amplitudes $|A_i|$ and the strong phases δ_i are very difficult to compute theoretically leading to large uncertainties. Thus, observables that depend only on the weak phases, such as the CKM angles, allow to test the SM predictions in a much cleaner way.

b) **CP violation in mixing** or indirect CP violation means,

$$\left| \frac{q}{p} \right| = \left| \frac{1 - \varepsilon}{1 + \varepsilon} \right| \neq 1 \Rightarrow |\varepsilon| \neq 0. \quad (2.22)$$

In this case, CP violation arises as a consequence of an asymmetry in the flavor oscillation $|M\rangle \rightarrow |\bar{M}\rangle \rightarrow |M\rangle$, meaning that physical mass eigenstates are not CP eigenstates, as explained in Sect. 2.5. In other words, there is a complex weak phase ϕ_{12} between M_{12} and Γ_{12} . The ε parameter corresponds to the parametrization used in Sect. 2.2.5.

c) **CP violation by interference of mixing and decay** is possible to be observed when the neutral mesons $|M\rangle$ and $|\bar{M}\rangle$ have both a common final state $|f\rangle$, preferentially a pure CP eigenstate

$$CP|f_{CP}\rangle = \pm|f_{CP}\rangle.$$

Even if there is no CP violation in mixing and decay separately, i.e. if $|\bar{A}_{f_{CP}}/A_{f_{CP}}| = |q/p| = 1$, the sum of the decay and the mixing phases ϕ_D and ϕ_M can give rise to a total phase difference and thus to an interference between these two processes, generating consequently a violation of CP symmetry. Introducing a new complex quantity λ_{CP} , the condition for this kind of CP violation can be written as

$$\text{Im}(\lambda_{CP}) \neq 0, \quad \text{where} \quad \lambda_{CP} \equiv \frac{q}{p} \cdot \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} = \left| \frac{q}{p} \right| \cdot \left| \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} \right| e^{-i(\phi_M + \phi_D)}.$$

2.7 The B -meson sector

B mesons consist of a heavy b or \bar{b} quark and a lighter quark. Depending on the lighter quark, they can be classified into two families, a neutral and a charged one. The neutral B mesons are³: B^0 , represented by $B^0(d\bar{b})$ and $\bar{B}^0(\bar{d}b)$, with a mass $m_{B^0} = 5.28 \text{ GeV}/c^2$; and B_s^0 , represented by $B_s^0(s\bar{b})$ and $\bar{B}_s^0(\bar{s}b)$, with a mass $m_{B_s^0} = 5.37 \text{ GeV}/c^2$. Analogously, the charged B mesons are: B^\pm , represented by $B^+(u\bar{b})$ and $B^-(\bar{u}b)$, with a mass $m_{B^\pm} = 5.28 \text{ GeV}/c^2$; and B_c^\pm , represented by $B_c^+(c\bar{b})$ and $B_c^-(\bar{c}b)$, with a mass $m_{B_c^\pm} = 6.28 \text{ GeV}/c^2$.

³The particle data in this section is taken from the Particle Data Group (PDG) [28].

As they are the major topic of study in this thesis, the discussions in the following will focus on the B mesons B^0 and B^\pm , which form $SU(2)$ isospin doublets ($I = \frac{1}{2}$):

- $B^0(d\bar{b})$, $B^+(u\bar{b})$ with $I_3 = -\frac{1}{2}, +\frac{1}{2}$, and
- $B^-(\bar{u}b)$, $\bar{B}^0(\bar{d}b)$ with $I_3 = -\frac{1}{2}, +\frac{1}{2}$,

with a very small mass difference $\Delta m_{B^0 B^\pm} = 0.31 \pm 0.06 \text{ MeV}/c^2$ between charged and neutral B mesons.

As explained in Sect. 2.4.3, the B -meson system offers the possibility to accurately test the CKM mechanism which predicts a large CP violation in it. For the study of CP violation and for the search for NP, some key properties of B mesons are advantageous from the experimental and from the theoretical point of view.

From the experimental point of view, B mesons have relatively long lifetimes ($\tau_{B^0} = 1.52 \text{ ps}$ and $\tau_{B^\pm} = 1.64 \text{ ps}$), facilitating the study of $B^0 - \bar{B}^0$ oscillations as well as of B decays. The long lifetimes of B mesons are a consequence of the small couplings between the third quark family (t, b) and the other two, (c, s) and (u, d) ($V_{cb} \sim \mathcal{O}(10^{-2})$, $V_{ub} \sim \mathcal{O}(10^{-3})$). Due to this ‘‘hierarchical’’ structure of the CKM matrix, decays via $b \rightarrow c$ transitions have larger branching fractions than those via $b \rightarrow u$ transitions. The hierarchy of the CKM matrix ($V_{td} \sim \mathcal{O}(10^{-3})$, $V_{ts} \sim \mathcal{O}(10^{-2})$) is also responsible for the fact that the oscillation frequency of B^0 mesons is lower than the oscillation frequency of B_s^0 mesons (see 2.7 in Sect. 2.7.1). Consequently, it is easier to measure the oscillation of neutral B^0 mesons.

From the theoretical point of view, the high mass of the b quark makes the calculations of hadronic corrections more reliable. The mass $m_b(\overline{MS}) \approx 4.2 \text{ GeV}$ is relatively far from the hadronic scale $\Lambda_{\text{QCD}} \approx 0.34 \text{ GeV}/c^2$ [48], at which the effects of strong interactions cannot be treated perturbatively anymore. This, together with the property of asymptotic freedom of QCD (see Sect. 2.1), and the fact that $\alpha_s(m_b) \ll 1$, opens the possibility of systematic approximations that can be exploited in various theoretical applications [49].

The following subsections are dedicated to the $B^0 - \bar{B}^0$ flavor mixing as well as to the CP violation in $B^0(\bar{B}^0)$ and in B^\pm decays. These phenomena play a central role at Belle II, and in the studies performed in this thesis.

2.7.1 $B^0 - \bar{B}^0$ mixing

For B^0 mesons and their antiparticles \bar{B}^0 , the time evolution description expressed in eq. (2.17) and eq. (2.18) can be simplified because the lifetime difference between the two physical eigenstates, B_H^0 and B_L^0 , is practically negligible [42]:

$$\left. \frac{\Delta\Gamma}{\Gamma} \right|_{B^0} = -0.002 \pm 0.010.$$

Thus, one can set $\Delta\Gamma = 0$ and redefine the lifetime as $\Gamma \equiv \Gamma_H = \Gamma_L$. The time evolution for neutral B mesons, starting as a pure flavor state, can be then written as

$$|B^0(t)\rangle = \frac{1}{2} e^{-\frac{\Gamma}{2}t} \left((1 + e^{-i\Delta mt}) |B^0\rangle + \frac{q}{p} (1 - e^{-i\Delta mt}) |\bar{B}^0\rangle \right), \quad (2.23)$$

$$|\bar{B}^0(t)\rangle = \frac{1}{2} e^{-\frac{\Gamma}{2}t} \left(\frac{p}{q} (1 - e^{-i\Delta mt}) |B^0\rangle + (1 + e^{-i\Delta mt}) |\bar{B}^0\rangle \right), \quad (2.24)$$

where $e^{-im_H t}$ was removed by phase convention. Considering the leading order amplitudes contributing to $B^0 - \bar{B}^0$ mixing (Fig. 2.6), there are virtual u , c and t quarks inside loops. However, on the scale of the B -meson mass, the light u and c quarks have nearly the same mass and their contributions cancel out due to the GIM-mechanism [50]. Consequently, the t -quark contributions dominate. It follows from eq. (2.15) that

$$\left. \frac{q}{p} \right|_{B^0} = \sqrt{\frac{M_{12}^*}{M_{12}}} + \mathcal{O}\left(\frac{\Gamma_{12}}{M_{12}}\right) \Big|_{B^0} \simeq \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*}, \quad \Rightarrow \left| \frac{q}{p} \right|_{B^0} \approx 1,$$

since $\Gamma_{12}/M_{12} \simeq \mathcal{O}(m_b^2/m_t^2) \sim \mathcal{O}(10^{-3})$ [51]. This is experimentally confirmed [42]:

$$\left| \frac{q}{p} \right|_{B^0} = 1.0010 \pm 0.0008.$$

Thus, CP violation in $B^0 - \bar{B}^0$ mixing is very small and the eigenstates B_H^0 and B_L^0 can be considered to a good approximation as pure CP eigenstates. Moreover, in the neutral B meson system, the complex parameter λ_{CP} can be written as

$$\begin{aligned} \lambda_{CP} &= e^{-i\phi_M(B)} \cdot \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \cdot \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}}, \quad (2.25) \\ \Rightarrow |\lambda_{CP}| &= \left| \frac{\bar{A}_{f_{CP}}}{A_{f_{CP}}} \right|. \end{aligned}$$

2.7.2 Time-dependent CP violation

Considering now the decay of a $B^0(\bar{B}^0)$ meson into a CP eigenstate f_{CP} , the corresponding decay amplitudes are given by

$$A_{CP} = \langle f_{CP} | H | B^0 \rangle, \quad \bar{A}_{CP} = \langle f_{CP} | H | \bar{B}^0 \rangle.$$

Application of the time evolution in eqs. (2.23) and (2.24) yields

$$A_{CP}(t) = \langle f_{CP} | H | B^0(t) \rangle = \frac{1}{2} e^{-\frac{\Gamma}{2}t} \left((1 + e^{-i\Delta mt}) A_{CP} + \frac{q}{p} (1 - e^{-i\Delta mt}) \bar{A}_{CP} \right)$$

$$\begin{aligned}
&= \frac{1}{2} e^{-\frac{\Gamma}{2}t} A_{CP} [(1 + e^{-i\Delta mt}) + \lambda_{CP} (1 - e^{-i\Delta mt})], \\
\bar{A}_{CP}(t) = \langle f_{CP} | H | \bar{B}^0(t) \rangle &= \frac{1}{2} e^{-\frac{\Gamma}{2}t} \left(\frac{p}{q} (1 - e^{-i\Delta mt}) A_{CP} + (1 + e^{-i\Delta mt}) \bar{A}_{CP} \right) \\
&= \frac{1}{2} e^{-\frac{\Gamma}{2}t} A_{CP} \frac{p}{q} [(1 - e^{-i\Delta mt}) + \lambda_{CP} (1 + e^{-i\Delta mt})].
\end{aligned}$$

This leads to the time-dependent decay rates

$$\begin{aligned}
\Gamma(B^0(t) \rightarrow f_{CP}) &= |\langle f_{CP} | H | B^0(t) \rangle|^2 \\
&= \frac{1}{2} e^{-\Gamma t} |A_{CP}|^2 \\
&\quad \cdot [1 + \cos(\Delta mt) - 2\text{Im}(\lambda_{CP}) \sin(\Delta mt) + (1 - \cos(\Delta mt)) |\lambda_{CP}|^2], \\
\Gamma(\bar{B}^0(t) \rightarrow f_{CP}) &= |\langle f_{CP} | H | \bar{B}^0(t) \rangle|^2 \\
&= \frac{1}{2} e^{-\Gamma t} |A_{CP}|^2 \\
&\quad \cdot [1 - \cos(\Delta mt) + 2\text{Im}(\lambda_{CP}) \sin(\Delta mt) + (1 + \cos(\Delta mt)) |\lambda_{CP}|^2].
\end{aligned}$$

CP violation gives rise to a difference between the decay rates above. Since the effect can be very small, the asymmetry between the decay rates is considered as observable, with the advantage that in this way many systematic effects cancel out reciprocally. The time-dependent CP asymmetry between decay rates is defined as

$$\begin{aligned}
a_{CP}(t) &\equiv \frac{\Gamma(\bar{B}^0(t) \rightarrow f_{CP}) - \Gamma(B^0(t) \rightarrow f_{CP})}{\Gamma(\bar{B}^0(t) \rightarrow f_{CP}) + \Gamma(B^0(t) \rightarrow f_{CP})} \\
&= \frac{(|\lambda_{CP}|^2 - 1) \cos(\Delta mt) + 2\text{Im}(\lambda_{CP}) \sin(\Delta mt)}{1 + |\lambda_{CP}|^2} \\
&= \mathcal{A}_{CP} \cos(\Delta mt) + \mathcal{S}_{CP} \sin(\Delta mt), \tag{2.26}
\end{aligned}$$

where the CP -violation parameters

$$\mathcal{A}_{CP} \equiv \frac{|\lambda_{CP}|^2 - 1}{|\lambda_{CP}|^2 + 1} \quad \text{and} \quad \mathcal{S}_{CP} \equiv \frac{2 \cdot \text{Im}(\lambda_{CP})}{|\lambda_{CP}|^2 + 1}, \tag{2.27}$$

have been introduced. Since in this case $|\lambda_{CP}| = |\bar{A}_{CP}/A_{CP}|$, the \mathcal{A}_{CP} parameter corresponds to direct violation of CP symmetry, i.e. $\mathcal{A}_{CP} \neq 0$ if $|\lambda_{CP}| \neq 1$. Respectively, \mathcal{S}_{CP} is related to the mixing-induced CP violation, i.e. to CP violation effects by interference between mixing and decay. Due to the approximation $\Delta\Gamma_B \approx 0$ in eqs. (2.23) and (2.24), a third CP violation

parameter is neglected in $a_{CP}(t)$ (eq. (2.26)), that is

$$\mathcal{A}_{\Delta\Gamma} = \frac{2 \cdot \text{Re}(\lambda_{CP})}{|\lambda_{CP}|^2 + 1} \stackrel{\Delta\Gamma=0}{\approx} 0.$$

Since $a_{CP}(t)$ is normalized, the CP violation parameters have to fulfill

$$\mathcal{A}_{CP}^2 + \mathcal{S}_{CP}^2 + \mathcal{A}_{\Delta\Gamma}^2 = 1 \quad \Rightarrow \quad \mathcal{A}_{CP}^2 + \mathcal{S}_{CP}^2 \leq 1.$$

The experimental challenge is then to measure precisely the time-dependent rates of B^0 and \bar{B}^0 decaying into CP eigenstates.

2.7.3 Measurements of CP violation

Since B mesons have a relatively large mass, they have myriad possible decay modes. Those that are useful for CP violation analyses are very rare with branching fractions around $10^{-4} - 10^{-6}$. Significant measurements require thus a huge amount of B meson events of the order of 10^8 .

B mesons can be produced in hadron colliders together with many other hadrons; or exclusively in a very clean environment at B factories (see detailed comparison in Sect. 2.8). B factories are asymmetric-energy e^+e^- colliders optimized for time-dependent studies of B meson decays. The first B factories were the KEKB at KEK in Tsukuba, Japan and the Positron-Electron Project IIs (PEP-II) at SLAC, California⁴. In the following, the discussion will focus on the measurement of CP violation at B factories.

At B factories, electrons and positrons collide with a center of mass of 10.58 GeV corresponding to the rest energy of the $\Upsilon(4S)$ resonance. The $\Upsilon(4S)$ is the third radially excited 3S_1 $b\bar{b}$ bound state, and the first hadronic $b\bar{b}$ resonance above the $B\bar{B}$ production threshold (see Fig. 2.8). Thus, B mesons are essentially at rest in the $\Upsilon(4S)$ center-of-mass frame. The $\Upsilon(4S)$ decays to more than 96% of the cases into $B\bar{B}$ meson pairs, and roughly into equal amounts of charged and neutral pairs, namely $\Gamma(B^+B^-)/\Gamma(B^0\bar{B}^0) = 1.058 \pm 0.024$ [28]. Charged B^+B^- meson pairs allow to measure the direct CP violation, whereas $B^0\bar{B}^0$ allow to measure the direct, and the mixing-induced CP violation.

The quantum numbers of the $\Upsilon(4S)$ resonance are $J^{PC} = 1^{--}$, and the quantum numbers of a B meson are $J^P = 0^-$. Conservation of angular momentum dictates the $B\bar{B}$ system to be in a p -wave configuration, i.e. the orbital angular momentum is $L = 1$. Since P and C are conserved in strong interactions, the quantum numbers of the $B\bar{B}$ pair are the same of the $\Upsilon(4S)$ resonance. The wave function of the $\Upsilon(4S)$ is antisymmetric, forbidding the final states B^0B^0 and $\bar{B}^0\bar{B}^0$, which are symmetric by means of Bose-Einstein statistics.

Considering a neutral pair $B^0\bar{B}^0$, if the two mesons in the system are allowed to oscillate independently, there is a non vanishing probability to find a B^0B^0 or a $\bar{B}^0\bar{B}^0$ meson pair at

⁴KEK stays for High Energy Accelerator Research Organisation in Japanese: Kō Enerugī Kasokuki kenkyū kikou. SLAC stays for Stanford Linear Accelerator Center.

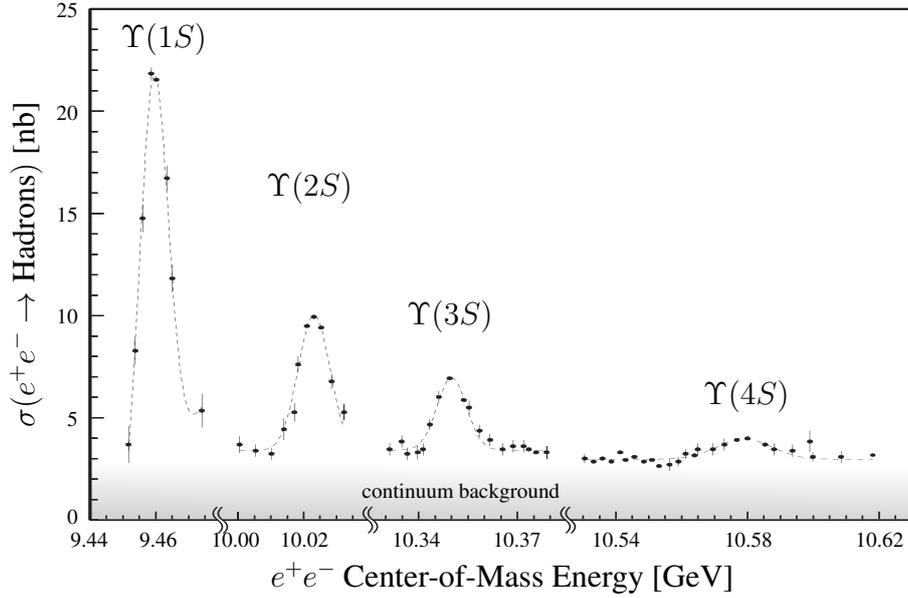


Figure 2.8: Hadronic cross section for e^+e^- collisions as a function of the e^+e^- center-of-mass energy in the region at and above the $\Upsilon(1S)$ resonance [52].

a given time. Since its wave function has to be anti-symmetric as well, when a neutral $B^0\bar{B}^0$ pair is produced, the two-meson system has to be in a quantum entangled state. The entangled state can be written as

$$|B_1^0(t_1), B_2^0(t_2)\rangle = \frac{1}{\sqrt{2}} (|B_1^0(t_1)\rangle|\bar{B}_2^0(t_2)\rangle - |\bar{B}_1^0(t_1)\rangle|B_2^0(t_2)\rangle). \quad (2.28)$$

The entanglement prevails until one of the two mesons decays at a time t_1 . At t_1 , the B mesons have opposite flavors, for example if one is a B^0 , then the other is a \bar{B}^0 . The second meson is free to oscillate up to its decay at a time t_2 . Therefore, events with two B^0 or two \bar{B}^0 decays are possible. The probability that this occurs depends on the time difference Δt between the two decays. Using eqs. (2.23) and (2.24), and defining $\Delta t \equiv t_1 - t_2$, eq. (2.28) becomes

$$|B_1^0(t_1), B_2^0(t_2)\rangle = \frac{1}{2\sqrt{2}} e^{-\frac{\Gamma}{2}(t_1+t_2)} \left[(1 + e^{-i\Delta m\Delta t}) (|B_1^0\rangle|\bar{B}_2^0\rangle - |\bar{B}_1^0\rangle|B_2^0\rangle) + (1 - e^{-i\Delta m\Delta t}) \left(\frac{q}{p} |\bar{B}_1^0\rangle|\bar{B}_2^0\rangle - \frac{p}{q} |B_1^0\rangle|B_2^0\rangle \right) \right]. \quad (2.29)$$

If the flavor of one of the two B mesons can be determined by a flavor-specific decay, the flavor and the time evolution of the other B is known. Considering a $B^0\bar{B}^0$ pair, if one meson (B_{CP}^0) decays into a CP eigenstate, and the other (B_{tag}^0) is tagged by a flavor-specific decay, the

time-dependent decay rates can be written as

$$\begin{aligned}
\Gamma(f_{CP}, f_{\text{tag}}) &= \left| \langle f_{CP}, f_{\text{tag}} | H | B_{CP}^0(t_{CP}), B_{\text{tag}}^0(t_{\text{tag}}) \rangle \right|^2 \\
&= \frac{1}{4} e^{-\Gamma(t_{CP}+t_{\text{tag}})} |A_{CP}|^2 |A_{\text{tag}}|^2 \left| \frac{p}{q} \right|^2 \\
&\quad \cdot \left[1 - \cos(\Delta m \Delta t) + 2\text{Im}(\lambda_{CP}) \sin(\Delta m \Delta t) + (1 + \cos(\Delta m \Delta t)) |\lambda_{CP}|^2 \right], \tag{2.30}
\end{aligned}$$

$$\begin{aligned}
\Gamma(f_{CP}, \bar{f}_{\text{tag}}) &= \left| \langle f_{CP}, \bar{f}_{\text{tag}} | H | B_{CP}^0(t_{CP}), B_{\text{tag}}^0(t_{\text{tag}}) \rangle \right|^2 \\
&= \frac{1}{4} e^{-\Gamma(t_{CP}+t_{\text{tag}})} |A_{CP}|^2 |A_{\text{tag}}|^2 \\
&\quad \cdot \left[1 + \cos(\Delta m \Delta t) - 2\text{Im}(\lambda_{CP}) \sin(\Delta m \Delta t) + (1 - \cos(\Delta m \Delta t)) |\lambda_{CP}|^2 \right], \tag{2.31}
\end{aligned}$$

where the decay amplitude $A_{\text{tag}} = \langle f_{\text{tag}} | H | B^0 \rangle = \langle \bar{f}_{\text{tag}} | H | \bar{B}^0 \rangle$ was introduced. Considering that CP violation in B meson mixing is negligible, i.e. $|p/q| \approx 1$ (eq. (2.25)), for the time-dependent CP asymmetry one obtains

$$\begin{aligned}
a_{CP}(\Delta t) &\equiv \frac{\Gamma(f_{CP}, f_{\text{tag}}) - \Gamma(f_{CP}, \bar{f}_{\text{tag}})}{\Gamma(f_{CP}, f_{\text{tag}}) + \Gamma(f_{CP}, \bar{f}_{\text{tag}})} \\
&= \mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t). \tag{2.32}
\end{aligned}$$

Normalizing eqs. (2.30) and (2.31) to unity in the region $-\infty < \Delta t < \infty$ (see derivation in App. A.1), the Δt probability for a flavor $q = +1(-1)$ corresponding to $B_{\text{tag}}^0 = B^0(\bar{B}^0)$ is given by

$$\mathcal{P}(\Delta t, q) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} [1 + q(\mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t))], \tag{2.33}$$

where τ_{B^0} is the B^0 lifetime. The CP violation parameters \mathcal{A}_{CP} and \mathcal{S}_{CP} can be measured by comparing the Δt distributions for $q = +1$ and $q = -1$. Fig. 2.9 shows the probability $\mathcal{P}(\Delta t, q)$ together with the resulting CP asymmetry a_{CP} for two different values of the CP -violation parameters \mathcal{A}_{CP} and \mathcal{S}_{CP} . A time-dependent CP violation analysis requires then to accomplish three tasks: the reconstruction of the B_{CP}^0 decay to a specific f_{CP} final state, the determination of the flavor q of B_{tag}^0 , and the measurement of the decay time difference Δt .

The reconstruction of the signal requires special techniques which depend on the considered CP -decay mode. The determination of the flavor of B_{tag}^0 is performed extracting flavor signatures from flavor-specific B^0 decays. A prominent example is the semileptonic decay $B_{\text{tag}}^0 \rightarrow X \ell^+ (\bar{B}_{\text{tag}}^0 \rightarrow X \ell^-)$, where the charge of the lepton ‘‘tags’’ unambiguously the flavor $q = -1(+1)$. One of the goals of this work was the development and the validation of a *flavor tagging algorithm* for the Belle II experiment. The algorithm will be introduced in chapter 4.

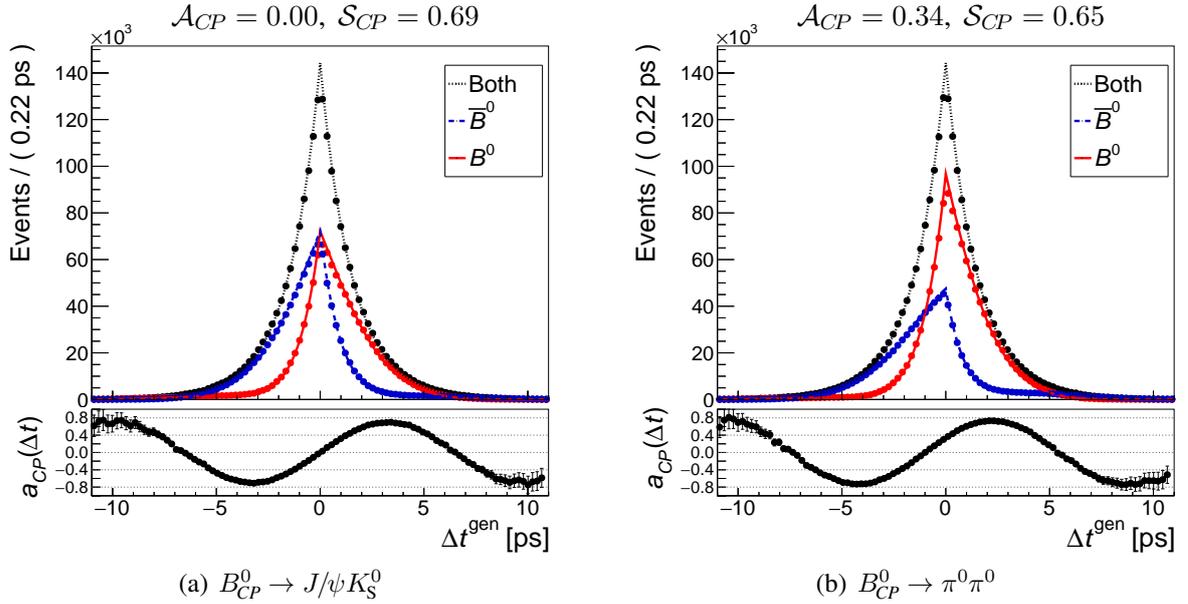


Figure 2.9: Two different examples of how CP violation can manifest in the probability $\mathcal{P}(\Delta t, q)$ and the related CP asymmetry a_{CP} . The distributions are shown for two different B_{CP}^0 decay channels. They were obtained from simulated events used for the studies in Chapter 4 and 6. The generated Δt distributions are shown for events where B_{tag}^0 is a B^0 ($q = +1$), and where B_{tag}^0 is a \bar{B}^0 ($q = -1$) at the time of its decay. No CP violation means no asymmetry, i.e. perfect matching of the distributions for B^0 and for \bar{B}^0 .

The measurement of the decay time difference Δt requires to achieve a precision at the order of magnitude of the B -meson lifetime, $\Delta t \sim \mathcal{O}(\text{ps})$. The reason why B factories have asymmetric beam energies is to translate this tiny time difference into a measurable decay length: the $\Upsilon(4S)$ system is produced with a Lorentz boost $\beta\gamma$ which amplifies the separation between decay vertices. Since the B mesons are produced almost at rest in the $\Upsilon(4S)$ frame, they travel to a good approximation along the boost direction. The time difference can be then measured as

$$\Delta t = \frac{\Delta l}{\beta\gamma c}, \quad (2.34)$$

where Δl is the difference between the decay vertex positions projected on the boost axis, and c is the speed of light. Fig. 2.10 shows a simple schematic of the production of two neutral B mesons at the $\Upsilon(4S)$.

In cases where tracks are absent in the final state, like in the decay $B_{CP}^0 \rightarrow \pi^0 \pi^0 \rightarrow 4\gamma$ (if no photon converts), the vertex location of B_{CP}^0 cannot be reconstructed and the time-dependent CP analysis is not possible. However, the direct CP -violation parameter \mathcal{A}_{CP} can be measured performing a time-integrated CP -violation analysis. The probability function for this analysis

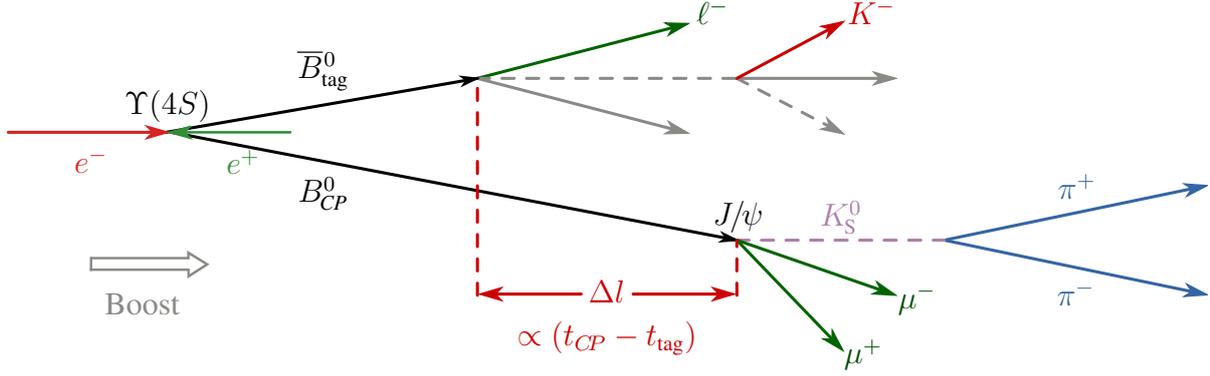


Figure 2.10: Simple schematic of production of a $B^0 \bar{B}^0$ pair at the $\Upsilon(4S)$. One of the mesons decays semileptonically and the other into the CP eigenstate $J/\psi K_S^0$.

is obtained integrating eq. (2.33) over $-\infty < \Delta t < \infty$ (see details in App. A.2),

$$\mathcal{P}(q) = \frac{1}{2} \left[1 + q \mathcal{A}_{CP} \frac{1}{(\tau_{B^0} \Delta m)^2 + 1} \right] = \frac{1}{2} [1 + q \mathcal{A}_{CP} (1 - 2 \cdot \chi_d)], \quad (2.35)$$

and the corresponding CP asymmetry is

$$a_{CP} = \frac{\mathcal{P}(q) - \mathcal{P}(-q)}{\mathcal{P}(q) + \mathcal{P}(-q)} = \mathcal{A}_{CP} (1 - 2 \cdot \chi_d), \quad (2.36)$$

where

$$1 - 2 \cdot \chi_d = \frac{1}{1 + (\tau_{B^0} \Delta m)^2}.$$

The parameter $\chi_d = 0.1860 \pm 0.0011$ [42] is defined as

$$\chi_d = \frac{x^2 + y^2}{2(x^2 + 1)},$$

where (see Tab. 2.2)

$$x = \frac{\Delta m}{\Gamma} = \tau_{B^0} \Delta m, \quad \text{and} \quad y = \left. \frac{\Delta \Gamma}{2\Gamma} \right|_{B^0} \approx 0.$$

Considering decays of charged B mesons, the direct CP violation parameter \mathcal{A}_{CP} is given by,

$$\mathcal{A}_{CP} = \frac{\Gamma(B^- \rightarrow f^-) - \Gamma(B^+ \rightarrow f^+)}{\Gamma(B^- \rightarrow f^-) + \Gamma(B^+ \rightarrow f^+)},$$

where $\Gamma(B^\pm \rightarrow f^\pm)$ are the decay rates of B^\pm to a specific final state f^\pm . Since the electric charge of the B meson is the same of the final state, measurements of CP violation in charged B meson decays require only the full reconstruction of the signal B decay.

2.7.4 Measurements of $B^0 - \bar{B}^0$ mixing

The measurement of $B^0 - \bar{B}^0$ oscillations allows the determination of the lifetime τ_{B^0} and the oscillation frequency Δm , which corresponds to the mass difference between the two mass eigenstates of a neutral B -meson. These measurements are essential to validate and to calibrate the algorithms used for the reconstruction of B -decay vertices and for the determination of the B^0 -flavor.

Focusing again on the experimental environment at B factories, for the measurement of $B^0 - \bar{B}^0$ mixing, one considers events where the flavor of the two exclusively produced B mesons can be determined. Usually, one reconstructs one of the B mesons to a flavor specific decay (B_{sig}^0) and tags the flavor of the other (B_{tag}^0) using a flavor tagging algorithm. Denoting the full-reconstructed final state as f_{sig} and the final state of the accompanying B meson as f_{tag} , one obtains the following decay rates using eq. (2.28)

$$\begin{aligned} \Gamma(f_{\text{sig}}, f_{\text{tag}}) &= |\langle f_{\text{sig}}, f_{\text{tag}} | H | B_{\text{sig}}^0(t_{\text{sig}}), B_{\text{tag}}^0(t_{\text{tag}}) \rangle|^2 \\ &= \frac{1}{4} e^{-\Gamma(t_{\text{sig}}+t_{\text{tag}})} |A_{\text{sig}}|^2 |A_{\text{tag}}|^2 \left| \frac{p}{q} \right|^2 [1 - \cos(\Delta m \Delta t)], \end{aligned} \quad (2.37)$$

$$\begin{aligned} \Gamma(\bar{f}_{\text{sig}}, \bar{f}_{\text{tag}}) &= |\langle \bar{f}_{\text{sig}}, \bar{f}_{\text{tag}} | H | B_{\text{sig}}^0(t_{\text{sig}}), B_{\text{tag}}^0(t_{\text{tag}}) \rangle|^2 \\ &= \frac{1}{4} e^{-\Gamma(t_{\text{sig}}+t_{\text{tag}})} |A_{\text{sig}}|^2 |A_{\text{tag}}|^2 \left| \frac{q}{p} \right|^2 [1 - \cos(\Delta m \Delta t)], \end{aligned} \quad (2.38)$$

$$\begin{aligned} \Gamma(f_{\text{sig}}, \bar{f}_{\text{tag}}) &= |\langle f_{\text{sig}}, \bar{f}_{\text{tag}} | H | B_{\text{sig}}^0(t_{\text{sig}}), B_{\text{tag}}^0(t_{\text{tag}}) \rangle|^2 \\ &= \frac{1}{4} e^{-\Gamma(t_{\text{sig}}+t_{\text{tag}})} |A_{\text{sig}}|^2 |A_{\text{tag}}|^2 [1 + \cos(\Delta m \Delta t)] = \Gamma(\bar{f}_{\text{sig}}, f_{\text{tag}}), \end{aligned} \quad (2.39)$$

where $A_{\text{sig}} = \langle f_{\text{sig}} | H | B^0 \rangle = \langle \bar{f}_{\text{sig}} | H | \bar{B}^0 \rangle$. Considering again that $|p/q| \approx 1$, one defines the mixing asymmetry as

$$a_{\text{mix}}(\Delta t) \equiv \frac{\Gamma(f_{\text{sig}}, \bar{f}_{\text{tag}}) + \Gamma(\bar{f}_{\text{sig}}, f_{\text{tag}}) - (\Gamma(f_{\text{sig}}, f_{\text{tag}}) + \Gamma(\bar{f}_{\text{sig}}, \bar{f}_{\text{tag}}))}{\Gamma(f_{\text{sig}}, \bar{f}_{\text{tag}}) + \Gamma(\bar{f}_{\text{sig}}, f_{\text{tag}}) + (\Gamma(f_{\text{sig}}, f_{\text{tag}}) + \Gamma(\bar{f}_{\text{sig}}, \bar{f}_{\text{tag}}))} = \cos(\Delta m \Delta t).$$

The eqs. (2.37) to (2.39) can be normalized to unity in the region $-\infty < \Delta t < \infty$, in the same way as eqs. (2.30) and (2.31) in Sect. 2.7.3. This leads to a Δt probability which is given by

$$\mathcal{P}(\Delta t, q_{\text{sig}}, q_{\text{tag}}) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} [1 - q_{\text{sig}} \cdot q_{\text{tag}} \cdot \cos(\Delta m \Delta t)], \quad (2.40)$$

where the flavor $q_{\text{sig}} = +1(-1)$ corresponds to $B_{\text{sig}}^0 = B^0(\bar{B}^0)$, and a flavor $q_{\text{tag}} = +1(-1)$ corresponds to $B_{\text{tag}}^0 = B^0(\bar{B}^0)$. For time-integrated validation and calibration studies, the equation above can be integrated in the same way as eq. (2.33) in Sect. 2.7.3 (the calculation

is performed in App. A.2). One obtains,

$$\mathcal{P}(q_{\text{sig}}, q_{\text{tag}}) = \frac{1}{2} [1 - q_{\text{sig}} \cdot q_{\text{tag}} \cdot (1 - 2 \cdot \chi_d)]. \quad (2.41)$$

2.8 LHCb and Belle II

For the next ten years the only experiments specialized to explore the physics of b hadrons, and especially the CP -violation in the B system, will be the Belle II and the LHCb experiments. As its name indicates, the large hadron collider beauty (LHCb) experiment [53] is located at the LHC at CERN.

While at B factories the B mesons are produced resonantly in e^+e^- collisions at the $\Upsilon(4S)$ (exactly at the $B\bar{B}$ threshold), at LHCb they are produced in proton-proton collisions at a high center-of-mass energy (currently 13 TeV). At an energy of 13 TeV, the hadronic $b\bar{b}$ cross section is about $560 \mu\text{b}$ [54] which is about six orders of magnitudes larger than the $B\bar{B}$ cross section at B factories ($\sim 1 \text{ nb}$). This allows LHCb to collect large data samples while operating at a relatively low luminosity of about $2 \cdot 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ (the world record reached at KEKB is $2.11 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$). The low luminosity at LHCb is important to avoid events with multiple proton-proton interactions per bunch crossing. Such events are very difficult to analyze; events with single interactions are thus preferred. The relatively low luminosity helps reduce the detector occupancy and the radiation damage. By changing the beam focus at the interaction point, LHCb can tune the luminosity independently from the interaction points of the other experiments at the LHC.

In proton-proton collisions, $b\bar{b}$ pairs are produced together with several other particles which form a so-called underlying event. The large number of particles per event complicates the reconstruction of b -hadron decays and has in particular a negative impact on the B -flavor tagging: the effective B -tagging efficiency obtained at LHCb is about ten times lower than at B factories, where $B\bar{B}$ pairs are produced without additional particles.

The $b\bar{b}$ production at LHCb occurs mainly through gluon-gluon fusions in which the momenta of the incoming partons is highly asymmetric in the lab frame [55]. This results in a large boost of the $b\bar{b}$ pair along the direction of the gluon with the higher momentum, which can be either in forward or in backward direction. The LHCb detector is a single arm spectrometer, covering only the forward region, such that about 35% of the produced $b\bar{b}$ pairs are in the detector acceptance [56].

A positive consequence of the large boost is the excellent time resolution: LHCb can reach resolutions of about 50 fs and resolve the frequency of B_s^0 flavor oscillations [56]. This is not possible at B factories where the time resolution is around 1 ps. The large boost at LHCb, however, is counterproductive for the reconstruction of K_S^0 particles since a considerable fraction of them decay after the vertex detector. Several decay channels with K_S^0 particles in the final state are not covered by LHCb. For example the $b \rightarrow s\bar{s}s$ channels, which are highly sensitive to NP effects (see Sect. 2.9).

Table 2.3: Past, current and future LHCb run periods and their characteristics: center-of-mass energy \sqrt{s} , collected/projected integrated luminosity, estimated cumulative yields of hadronic B decays with respect to Run 1, and the year of completion. The cumulative yields take into account the integrated luminosity, the $b\bar{b}$ cross section in dependence of the energy, and the expected higher trigger efficiency after the phase-1 upgrade. The high-luminosity upgrade phase-2 is currently at the conceptual level [57].

Run period	\sqrt{s} [TeV]	Collected / Projected $\int \mathcal{L} dt$ [fb^{-1}]	Cum. Yield w.r.t. Run 1	Year attained
Run 1	7, 8	3	1	2012
Run 2	13	5	4	2018
Phase-1 upgrade	14	50	60	2030
Phase-2 upgrade (?)	14	300	~ 400	2035

Due to the harsh hadronic environment at LHCb, the reconstruction of final states with neutral particles like neutrinos, photons or K_L^0 particles is very difficult, in particular in the case of channels with several photons in the final state like those including neutral pions and η mesons. Due to the large boost, a considerable fraction of neutral pions cannot be resolved as pairs of photons in the electromagnetic calorimeter. An additional difficulty is to determine the direction of the momentum of neutral particles since the detector geometry is squeezed toward the interaction point: small angular uncertainties result in large spatial uncertainties when the particles' direction is pointed back to the interaction point. Therefore, LHCb does not cover most of the channels with one neutral particle and none of the channels with more than one.

An advantage at LHCb is the possibility to study b hadrons that are heavier than the B_s^0 , like the Λ_b and the B_c^+ . Due to the huge number of produced $b\bar{b}$ pairs, LHCb offers also unique opportunities for the search of rare decays, such as $B^0 \rightarrow \mu^+\mu^-$. However, rare decays to invisible final states, such as $B \rightarrow \nu\bar{\nu}$, need a full event reconstruction: a task that only Belle II can fulfill due to the tight kinematical constraints on the initial state at a B factory.

For the measurement of absolute branching fractions, the total number of produced B mesons is required. At B factories, this can be performed to a good level of precision (the measurement is explained in Sect. 7.3). At LHCb, the complexity of the hadronic interactions in the proton-proton collision impedes the estimation of the total number of produced B mesons. LHCb circumvents this obstacle by measuring ratios between two branching fractions.

LHCb started to operate in March 2010. The data collected until 2012 (Run 1) corresponds to 3 fb^{-1} and to about $3 \cdot 10^{11}$ $b\bar{b}$ pairs in the detector acceptance [56]. Just for a comparison, the Belle experiment recorded about $0.8 \cdot 10^9$ $B\bar{B}$ pairs with a dataset of about 0.8 ab^{-1} at the $\Upsilon(4S)$; and Belle II aims to collect about $5 \cdot 10^{10}$ $B\bar{B}$ pairs with full integrated luminosity (by 2025). Tab. 2.3 [57] details the past, current and future LHCb run periods with an estimate of

the increase in the cumulative yields of hadronic B decays in comparison with those attained during the first run period. These increases take into account the integrated luminosity, the $b\bar{b}$ cross section in dependence of the energy, and the expected higher trigger efficiency after the phase-1 upgrade [57].

The previous discussion summarizes only few key points which are important for further discussions in this thesis. Belle II and LHCb have comprehensive physics programs which cannot be compared here in full extent. In addition to b -hadron physics, both experiments include charm physics, τ -physics and hadron spectroscopy. Depending on the particular decay channel or physics observable, the experiments will compete against or complement each other.

2.9 The unitarity angles ϕ_1 and ϕ_3

At the current level of experimental precision, the CKM mechanism gives a remarkably good description of quark-flavor dynamics and in particular of the CP violation in the B meson system, which was established by the Belle and the BaBar experiments, an achievement that, along with the discovery of the six quarks, merited the Nobel Prize for Kobayashi and Maskawa in 2008.

Fig. 2.11 shows the unitarity triangle for the B -meson system calculated using the available data [58]. Among the three unitarity angles, ϕ_1 is currently the one determined with the highest precision. The least determined one is ϕ_3 . Within the uncertainties, the unitarity condition is fulfilled. However, many constraints on the triangle, especially those associated with loop-processes, are still limited by the experimental precision. Therefore, Belle II aims at improving the precision on the determination of the sides and the angles of the unitary triangle, especially of ϕ_2 , ϕ_3 and V_{ub} , which, along with concurrent lattice and phenomenological advances, will give a sharper insight into possible NP contributions.

In the following, the determination of the unitary triangles ϕ_1 and ϕ_3 is briefly discussed putting emphasis on ϕ_1 since the flavor tagger developed in this thesis plays an essential role in its determination. The discussion is based on the general ones in [1, 29, 36, 42]. Since the unitarity angle ϕ_2 plays a central role in this thesis, its determination will be discussed in detail in Sect. 2.10.

The unitarity angle ϕ_1

For the measurement of the unitarity angle ϕ_1 , the useful B_{CP}^0 decay modes can be classified into four main groups according to the underlying quark transitions. These transitions are $b \rightarrow c\bar{c}s$, $b \rightarrow c\bar{c}d$, $b \rightarrow c\bar{u}d$, and $b \rightarrow q\bar{q}s$. Especially interesting are the $b \rightarrow c\bar{c}s$ and the $b \rightarrow q\bar{q}s$ transitions, which will be discussed in the following. The $b \rightarrow c\bar{c}s$ transitions are the sources of the most precise measurement of ϕ_1 ; and $b \rightarrow q\bar{q}s$ transitions offer unique opportunities for the search of NP effects.

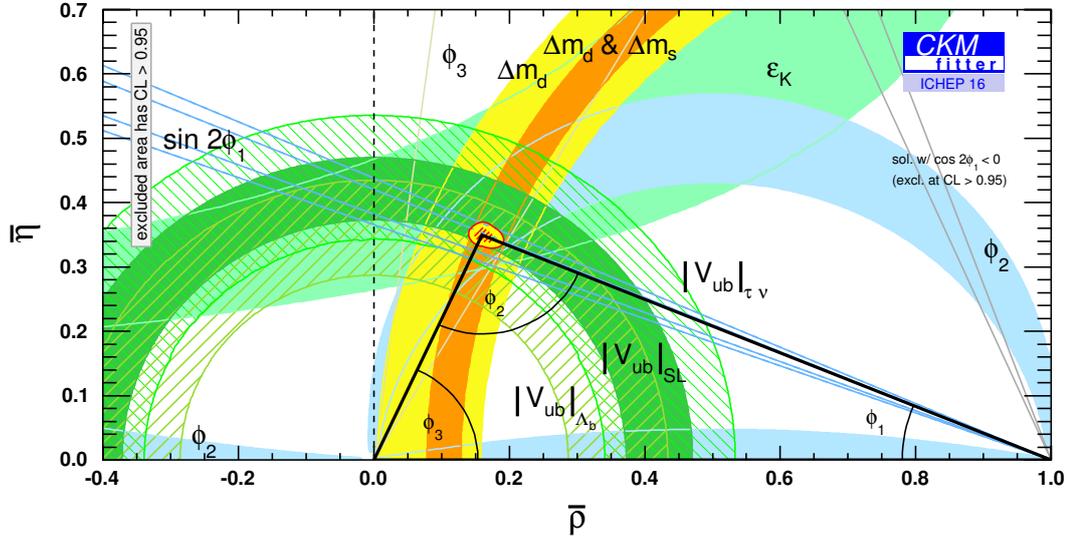


Figure 2.11: Current constraints on the unitarity triangle for the B -meson system using measurements up to 2016 [58].

Color-suppressed $b \rightarrow c\bar{c}s$ transitions: these processes include B_{CP}^0 decays into a charmonium resonance ($c\bar{c}$) and a K^0 (K_S^0 or K_L^0), or a K^{*0} that decays into a CP eigenstate ($K_S^0\pi^0$ or $K_L^0\pi^0$). Fig. 2.12 shows the leading order Feynman diagrams contributing to these processes. In the SM, the contributions of the penguin diagrams are doubly CKM-suppressed and can be neglected to an approximation of the order of 1% [1]. Considering the $K^0 - \bar{K}^0$ mixing in the case of $(cc)K^0$ decays, the CP violation complex quantity $\lambda_{(c\bar{c})K^0}$ is a product of three phases:

$$\begin{aligned} \lambda_{(c\bar{c})K^0} &= \left(\frac{q}{p}\right)_{B^0} \cdot \left(\frac{\bar{A}_{(c\bar{c})K^0}}{A_{(c\bar{c})K^0}}\right) \cdot \left(\frac{q}{p}\right)_{K^0} \simeq \frac{V_{tb}^*V_{td}}{V_{tb}V_{td}^*} \cdot \frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}} \cdot \frac{V_{cs}V_{cd}^*}{V_{cs}^*V_{cd}} = \frac{V_{tb}^*V_{td}V_{cb}V_{cd}^*}{V_{tb}V_{td}^*V_{cb}^*V_{cd}} \\ &= \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right)^{-1} \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right)^* = (e^{i\phi_1})^{-1} \cdot (e^{i\phi_1})^* = e^{-2i\phi_1}, \end{aligned} \quad (2.42)$$

where ϕ_1 is the unitarity angle defined in eq. (2.9) of Sect. 2.4.3. For the CP violation parameters (eq. (2.27)), one obtains

$$\lambda_{CP} = \eta_{CP}\lambda_{(c\bar{c})K^0}, \quad \Rightarrow \quad \mathcal{A}_{CP} = 0, \quad \mathcal{S}_{CP} = -\eta_{CP}\sin(2\phi_1),$$

where η_{CP} is the CP eigenvalue of the final state. The η_{CP} eigenvalue is given by the product of the individual CP eigenvalues of the final state mesons times $(-1)^L$ (L is the orbital angular momentum). Since a B^0 is a spin-0 particle, the total angular momentum of the final state must be zero. For example, in the case of vector-pseudoscalar final states (spin-1-spin-0),

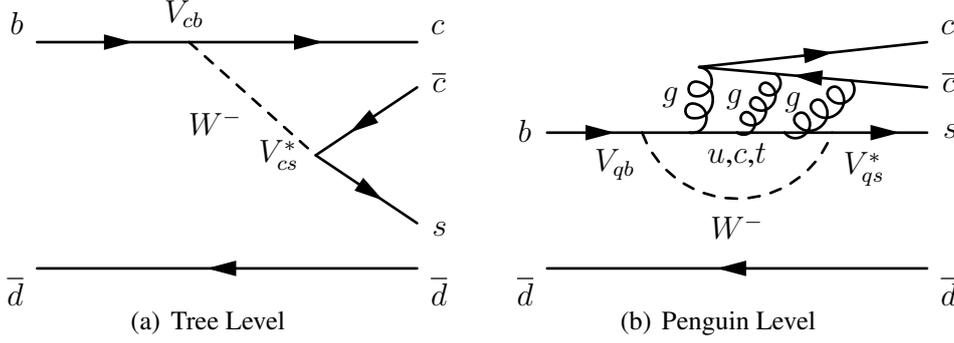


Figure 2.12: Leading order Feynman diagrams contributing to decays via $b \rightarrow c\bar{c}s$ transitions. The contribution of the penguin diagram can be neglected to an approximation that is better than one percent.

the mesons are emitted in a p -wave configuration ($L = 1$). Considering the K_S^0 and the K_L^0 as nearly CP -even and CP -odd states, the decay channels analyzed by Belle and BaBar can be classified into CP -odd and CP -even channels. The channels $J/\psi K_S^0$, $\psi(2S)K_S^0$, $\chi_{c1}K_S^0$, $\eta_c K_S^0$ are CP -odd channels, while the $J/\psi K_L^0$ channel is CP -even. The vector-vector channels, such as the $J/\psi K^{*0}$, are admixtures of CP -odd and CP -even channels and require an angular analysis to separate the amplitudes of definite CP . Among all these channels, the channel $B_{CP}^0 \rightarrow J/\psi K_S^0$ is regarded as the *golden channel* because of the small theoretical uncertainties, the relatively large branching fraction (around $8 \cdot 10^{-4}$), the high reconstruction efficiency, the large signal-to-background ratio, and the good resolution achievable for the vertex reconstruction of B_{CP}^0 . A precise determination of $\sin(2\phi_1)$, however, is achieved by considering simultaneously CP -odd and CP -even channels.

Exploiting their full integrated luminosities, Belle and BaBar measured the following CP -violation parameters combining the $b \rightarrow c\bar{c}s$ channels:

$$\text{Belle [59]:} \quad \sin(2\phi_1) = 0.667 \pm 0.023 \pm 0.12, \quad \mathcal{A}_{CP} = 0.006 \pm 0.016 \pm 0.012,$$

$$\text{BaBar [60]:} \quad \sin(2\phi_1) = 0.687 \pm 0.028 \pm 0.012, \quad \mathcal{A}_{CP} = -0.024 \pm 0.020 \pm 0.016.$$

On the other hand, the LHCb collaboration has covered only the $B^0 \rightarrow J/\psi K_S^0$ channel. Using the full Run 1 data set, LHCb measures

$$\text{LHCb [61]:} \quad \sin(2\phi_1) = 0.731 \pm 0.035 \pm 0.020, \quad \mathcal{A}_{CP} = 0.038 \pm 0.032 \pm 0.005.$$

At B factories, the precision for \mathcal{A}_{CP} , is close to be dominated by systematic uncertainties: the dominating source in the case of \mathcal{A}_{CP} is the tag-side interference effect, which will be discussed in Sect. 4.10.

Extrapolating the precision of Belle to the full Belle II data sample, the Belle II collaboration estimates the expected uncertainties $\delta \sin(2\phi_1) = 0.0052$ and $\delta \mathcal{A}_{CP} = 0.0090$ (statistical and

systematic uncertainties are added in quadrature) [1]. To extract the value of ϕ_1 at this level of precision, penguin contributions have to be taken into account. This can be performed by adding a correction term, such as in

$$\mathcal{S}_{CP}^{J/\psi K_S^0} \equiv \sin(2\phi_1 + \delta\phi^{J/\psi K_S^0}).$$

However, the value of $\delta\phi^{J/\psi K_S^0}$ is very difficult to determine. Different ways to perform this are currently studied. Very promising are sophisticated data-driven techniques which rely on approximative flavor symmetries to extract the value of $\delta\phi^{J/\psi K_S^0}$ (an example is [62]). Most of the flavor-symmetric channels, like $B^0 \rightarrow J/\psi\pi^0$, are not covered by LHCb. Thus, further studies will have to wait for high-precision measurements at Belle II.

Considering now the extraction of ϕ_1 from $\sin(2\phi_1)$, there are four possible solutions:

$$(\phi_1) \leftrightarrow \left(\frac{\pi}{2} - \phi_1\right), (\phi_1 + \pi), \left(\frac{3\pi}{2} - \phi_1\right).$$

To partially resolve the ambiguities $(\phi_1) \leftrightarrow (\phi_1 + \pi)$ and $(\frac{3\pi}{2} - \phi_1)$, one performs time-dependent analyses of decays where the probability function describing the Δt distribution contains $\sin(2\phi_1)$ as well as $\cos(2\phi_1)$ terms. For example one can perform full angular analyses of vector-vector decays such as $B_{CP}^0 \rightarrow J/\psi K^{*0}$, time-dependent Dalitz-plot analyses of three-body decays such as $B_{CP}^0 \rightarrow D^{(*)0}h^0$ (a $b \rightarrow c\bar{u}d$ transition with an unflavored meson h^0), or time-dependent analysis in two Dalitz regions such as $B_{CP}^0 \rightarrow D^{*+}D^{*-}K^0$ (a three-body $b \rightarrow c\bar{c}s$ transition). On the other hand, the $(\frac{\pi}{2} - \phi_1)$ ambiguity can be removed only by combining ϕ_1 with other CKM measurements.

The ϕ_1 measurements performed by Belle and BaBar dominate the precision of the current world average; combining the $b \rightarrow c\bar{c}s$ measurements, the SM compatible value of ϕ_1 is $\phi_1 = (21.9 \pm 0.7)^\circ$ [42].

Penguin-only or penguin-dominated $b \rightarrow q\bar{q}s$ transitions: the penguin amplitude dominating these processes is shown in Fig. 2.13. For some of these processes, like for most of those occurring via $b \rightarrow s\bar{s}s$ transitions, for example the $B^0 \rightarrow \phi K_S^0$ decay, tree level contributions are completely absent. For others, like in the case of $b \rightarrow u\bar{u}s$ transitions, for example the decay $B \rightarrow \eta' K_S^0$, color-suppressed and Cabibbo-suppressed tree diagrams may contribute too, however, their contributions are expected to be negligible with respect to the penguin contributions. The dominant penguin contribution has exactly the same weak phase as the $b \rightarrow c\bar{c}s$ tree diagram and leads to the same complex quantity λ_{CP} (eq. (2.42)). Thus, any tiny deviation of $\sin(2\phi_1)$ from the value obtained using $b \rightarrow c\bar{c}s$ processes is a clear indication of a NP effect, as soon as it is beyond the theoretical uncertainties for the predicted deviation

$$\Delta\mathcal{S}_f = -\eta_{CP}\mathcal{S}_{CP} - \sin(2\phi_1)^{c\bar{c}s}.$$

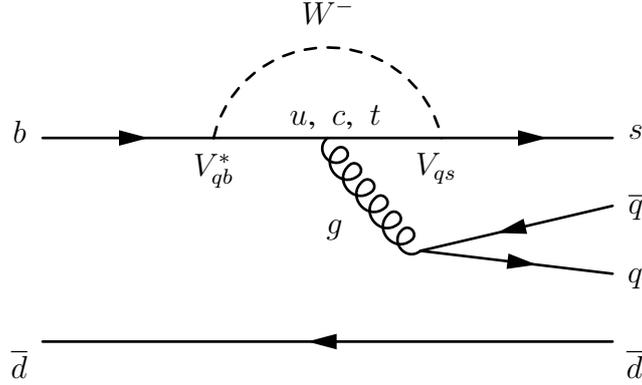


Figure 2.13: Feynman diagram for penguin-only or penguin-dominated $b \rightarrow q\bar{q}s$ processes.

The value of $\Delta\mathcal{S}_f$ is predicted to be very small, especially for the final states $\eta'K_S^0$ and ϕK_S^0 . At present, the measurements of $\Delta\mathcal{S}_f$ are still dominated by statistical uncertainties: the total experimental uncertainties are about an order of magnitude above the theoretical ones [42]. At this level of precision, there are no significant tensions between the data and the predictions. Since the theoretical uncertainties on $\Delta\mathcal{S}_f$ are very small, these channels are optimal for the search of NP effects.

The unitarity angle ϕ_3

Several approaches have been proposed to measure the unitarity angle ϕ_3 (see discussions in [29, 36]). The theoretically clean ones base on the interference between the tree amplitudes $b \rightarrow c\bar{u}s$ and $b \rightarrow u\bar{c}s$ in charged B -meson decays $B^\pm \rightarrow D^{(*)}K^{(*)\pm}$, where $D^{(*)}$ represents a superposition of $D^{(*)0}$ and $\bar{D}^{(*)0}$. Fig. 2.14 shows the contributing Feynman diagrams. With respect to the $b \rightarrow c\bar{u}s$ amplitude, the $b \rightarrow u\bar{c}s$ amplitude is color-suppressed by a factor $c_f \approx 0.1$. The $D^0 - \bar{D}^0$ mixing can be neglected to a good approximation. There are also no penguin contributions: the channels arise solely from the interference of the two tree level diagrams, which differ in their weak and their strong phases. Considering for example the final state DK^- , the ratio between color-suppressed and color-favored amplitudes can be written as

$$\frac{A(B^- \rightarrow \bar{D}^0 K^-)}{A(B^- \rightarrow D^0 K^-)} = r_B e^{i(\delta_B - \phi_3)},$$

where $r_B = c_f |V_{cs}V_{ub}^*/V_{us}V_{cb}^*|$ is the ratio of the magnitudes and δ_B is the strong phase difference. The same expression applies for the CP conjugated decay with a B^+ ; the only difference is that the relative weak phase becomes complex conjugate: $\phi_3 \rightarrow -\phi_3$. The parameters r_B , δ_B and ϕ_3 can be determined experimentally and, due to the absence of penguin contributions, they are free of theoretical uncertainties.

There are three established methods to extract ϕ_3 basing on different D decay channels. They are known by the initials of their proponents. The Gronau-London-Wyler method (GLW)

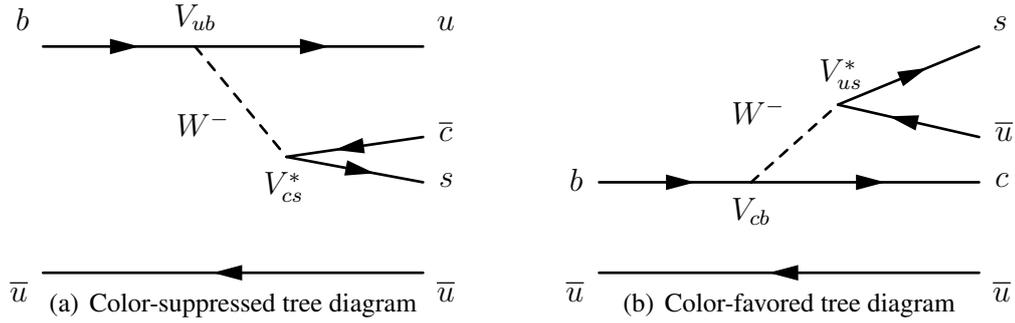


Figure 2.14: Dominant Feynman diagrams contributing to $B^\pm \rightarrow D^{(*)}K^{*\pm}$ decays. The color-suppressed diagram ($b \rightarrow u\bar{c}s$) is suppressed additionally by the small value of $|V_{ub}|$. The color-favored diagram on the left ($b \rightarrow \bar{c}us$ transition) is only singly Cabibbo suppressed. The total ratio $|V_{cs}V_{ub}^*/V_{us}V_{cb}^*|$ is of order $\mathcal{O}(1)$.

bases on Cabibbo-suppressed D decays to CP eigenstates, such as K^+K^- (CP -even) and $K_S^0\pi^0$ (CP -odd) [63]. The Atwood-Dunietz-Soni method considers the case when D^0 and \bar{D}^0 decay to a common final state which is doubly Cabibbo-suppressed for one of the two, for example when the color-favored decay $B^- \rightarrow D^0K^-$ is followed by the doubly-Cabibbo suppressed $D^0 \rightarrow K^+\pi^-$ decay, while the color-suppressed decay $B^- \rightarrow \bar{D}^0K^-$ is followed by the Cabibbo favored $\bar{D}^0 \rightarrow K^+\pi^-$ decay (ADS) [64]. In this way, the double Cabibbo suppression in one diagram compensates the color-suppression in the other, such that the ratio r_B can get close to unity, potentially giving rise to large CP -violation effects. The third method is the Giri-Grossman-Soffer-Zupan method (GGSZ) [65] which bases on the Dalitz-plot distribution of the products of D decays to self-conjugate multi-bodi final states, such as $D^0(\bar{D}^0) \rightarrow K_S^0\pi^+\pi^-$. The GGSZ method exploits the entire resonant structure of the three-body D decay, with interference of doubly Cabibbo-suppressed, Cabibbo-allowed, and CP eigenstate amplitudes. This results in a large sensitivity to ϕ_3 . Additionally, the branching fraction of such decays is larger with respect to the decays considered by the other methods. However, the common problem for all the three methods are the small overall branching fractions within $5 \cdot 10^{-6}$ to $5 \cdot 10^{-9}$. A precise determination of ϕ_3 requires thus large data samples. Combining all the measurements of Belle and BaBar, the average value is $\phi_3 = (67 \pm 11)^\circ$ [29]. The achieved precision is still limited by the sample sizes and is about one order of magnitude worse than for ϕ_1 .

At B factories, the Golden Channel for the determination of ϕ_3 is the channel $B^\pm \rightarrow D(\rightarrow K_S^0\pi^+\pi^-)K^\pm$, which is analyzed following the GGSZ method. This analysis requires as input the strong D -decay phases, which can be determined in different ways. The Belle II collaboration plans to employ auxiliary measurements of the strong D -decay phases that will be performed at the Beijing Spectrometer III (BES III), a charm factory running at the $\psi(3770)$ resonance ($D\bar{D}$ production threshold).

At LHCb, the ADS/GLW and the GGSZ decay channels are almost equally accessible, because of the larger production cross section with respect to B factories (compensating the

small branching fractions), and because of the lower selection efficiency and the worse vertex and mass resolutions for K_S^0 particles decaying beyond the vertex detector (thus lower performance in GGSZ analyses). Combining all LHCb measurements, the average ϕ_3 value is the most precise value from a single experiment $\phi_3 = (72.2_{-7.3}^{+6.8})^\circ$ [66]. LHCb has performed extrapolations and predicts a precision around 4° with the Run 2 data set (attained by 2018), and aims at reaching a precision of 1° with the full dataset collected by 2030 [57]. On the other hand, the Belle II collaboration predicts a precision around 1.6° with its full data set collected by 2025 [1].

2.10 The unitarity angle ϕ_2

Any B decay mode induced by a $b \rightarrow u\bar{u}d$ transition is sensitive to ϕ_2 . Thus, decays of charged and neutral B mesons to two, three or four pions in the final state can be used to measure ϕ_2 . Useful decays with three and four pions in the final state occur through intermediate resonances, for example $B \rightarrow \rho\pi \rightarrow 3\pi$, $B \rightarrow \rho\rho \rightarrow 4\pi$ and $B \rightarrow a_1(1260)\pi \rightarrow 4\pi$. Fig. 2.15 shows the leading order tree and QCD-penguin Feynman diagrams, together with the color-suppressed EW penguin diagrams, contributing to the decays $B^{i+j} \rightarrow h_1^i h_2^j$ of charged and neutral B mesons. The final state $h_1^i h_2^j$ stays for a pair of unflavoured light mesons such that $h \in \{\pi, \rho\}$ and $i, j, i+j \in \{-, 0, +\}$. An overview with more higher order Feynman diagrams can be found in [67, 68].

Considering for an instant only the decays of neutral B^0 mesons, one can assume for a first approximation that only tree level diagrams contribute (Figs. 2.15(a) and 2.15(b)). Taking into account the B^0 meson mixing, one obtains for the CP quantity λ_{hh} (eq. (2.25)),

$$\lambda_{hh} = \left(\frac{q}{p}\right)_{B^0} \cdot \left(\frac{\bar{A}^{hh}}{A^{hh}}\right) = \frac{V_{tb}^* V_{td} V_{ud}^* V_{ub}}{V_{tb} V_{td}^* V_{ud} V_{ub}^*} = e^{2i\phi_2}, \quad (2.43)$$

where the definition of the ϕ_2 angle in eq. (2.10) was used. For the corresponding CP violation parameters one obtains

$$\lambda_{CP} = \eta_{CP} \lambda_{hh} \Rightarrow \mathcal{A}_{CP} = 0, \quad \mathcal{S}_{CP} = \eta_{CP} \sin(2\phi_2).$$

The measured direct CP -violation parameter \mathcal{A}_{CP} for the decay $B^0 \rightarrow \rho^+ \rho^-$ is $\mathcal{A}^{\rho^+ \rho^-} = 0.05 \pm 0.13$. This shows that the penguin contributions are very small for this decay. However, for the decay $B^0 \rightarrow \pi^+ \pi^-$ the measured value of $\mathcal{A}^{\pi^+ \pi^-} = 0.31 \pm 0.05$ differs considerably from zero, revealing that penguin contributions are not negligible in general⁵. For the decays $B^0 \rightarrow h^0 h^0$, penguin contributions are expected to be even larger, as the leading order tree contribution is color suppressed. Despite the relatively large experimental uncertainties, the measured value of $\mathcal{A}^{\pi^0 \pi^0} = 0.43 \pm 0.24$ confirms this expectation.

Coming back to the general case of $B^{i+j} \rightarrow h_1^i h_2^j$ decays, penguin contributions have to be

⁵The values for the measured CP -violation parameters given in this section are the current world averages by the Particle Data Group (PDG) [28].

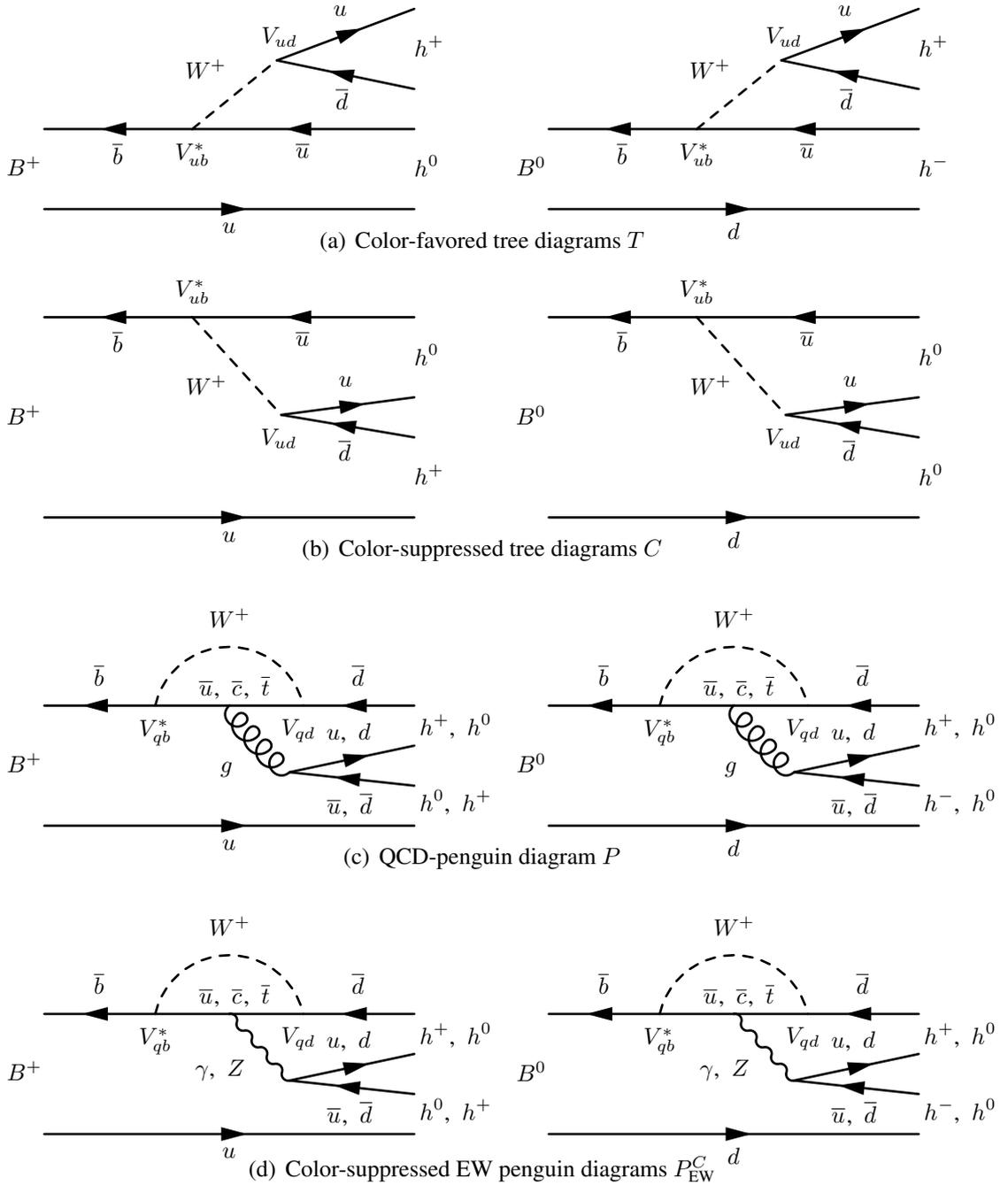


Figure 2.15: The dominant tree level and QCD-penguin Feynman diagrams together with the color-suppressed EW penguin diagrams contributing to $B^+ \rightarrow h^- h^0$ (left) and to $B^0 \rightarrow h^+ h^-, h^0 h^0$ (right). The generic h stays for a pion or for a ρ meson.

considered. Since the EM coupling coefficient α is smaller than the strong coupling coefficient α_s , EW penguin amplitudes are expected to be of $\mathcal{O}(10\%)$ with respect to the QCD penguin amplitudes [69, 70] and, therefore, they can be neglected to a good approximation. Accounting for the CKM coefficients in the tree diagrams and in the QCD-penguin diagrams, the decay amplitude for decays $B \rightarrow h_1^i h_2^j$ can be written as [71]

$$A^{ij} = \langle h_1^i h_2^j | \mathcal{H}_{\text{eff}} | B \rangle = V_{ud} V_{ub}^* (T_u^{ij} + P_u^{ij}) + V_{cd} V_{cb}^* P_c^{ij} + V_{td} V_{tb}^* P_t^{ij},$$

where \mathcal{H}_{eff} is the effective Hamiltonian describing the transition, T_u^{ij} is the hadronic tree amplitude, and P_u^{ij} , P_c^{ij} and P_t^{ij} are the hadronic QCD-penguin amplitudes with quarks u , c , and t in the W loop. Using the unitarity relation of the CKM matrix given in eq. (2.7), the decay amplitude A^{ij} can be rewritten for three different conventions:

$$\begin{aligned} \mathcal{U} \text{ convention: } & A^{ij} = V_{cd} V_{cb}^* (P_c^{ij} - T_u^{ij} - P_u^{ij}) + V_{td} V_{tb}^* (P_t^{ij} - T_u^{ij} - P_u^{ij}), \\ \mathcal{C} \text{ convention: } & A^{ij} = V_{ud} V_{ub}^* (T_u^{ij} + P_u^{ij} - P_c^{ij}) + V_{td} V_{tb}^* (P_t^{ij} - P_c^{ij}), \\ \mathcal{T} \text{ convention: } & A^{ij} = V_{td} V_{tb}^* (T_u^{ij} + P_u^{ij} - P_t^{ij}) + V_{cd} V_{cb}^* (P_c^{ij} - P_t^{ij}). \end{aligned}$$

Being all three conventions equivalent, the choice of convention can be done arbitrary without any physical implication. This is referred to as CKM ambiguity [72]. In the following, the \mathcal{C} convention is adopted. Defining the tree and the penguin amplitudes as $T_{uc}^{ij} = T_u^{ij} + P_u^{ij} - P_c^{ij}$ and $P_{tc}^{ij} = P_t^{ij} - P_c^{ij}$, one obtains

$$A^{ij} = V_{ud} V_{ub}^* T_{uc}^{ij} + V_{td} V_{tb}^* P_{tc}^{ij}.$$

Including the magnitudes squared of the CKM products in the amplitudes, i.e. using $T_{uc}^{ij} = T^{ij} / |V_{ud} V_{ub}^*|^2$ and $P_{tc}^{ij} = -P^{ij} / |V_{td} V_{tb}^*|^2$, the CP -violation quantity λ_{hh} can be written as

$$\begin{aligned} \lambda_{hh} &= \frac{q}{p} \cdot \frac{\bar{A}^{ij}}{A^{ij}} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \frac{1}{V_{ud} V_{ub}^*} \bar{T}^{ij} - \frac{1}{V_{td} V_{tb}^*} \bar{P}^{ij} = \frac{V_{tb}^* V_{td}}{V_{tb} V_{td}^*} \frac{V_{cd} V_{cb}^*}{V_{ud} V_{ub}^*} \bar{T}^{ij} - \frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*} \bar{P}^{ij} \\ &= e^{-2i\phi_1} \frac{e^{-i\phi_3} \bar{T}^{ij} - e^{i\phi_1} \bar{P}^{ij}}{e^{i\phi_3} T^{ij} - e^{-i\phi_1} P^{ij}}, \end{aligned}$$

where the definitions of the unitarity triangles ϕ_1 and ϕ_3 (eqs. (2.9) and (2.11)) were used. The amplitudes T^{ij} and P^{ij} can be written in terms of magnitude and hadronic phase, leading to

$$A^{ij} = e^{i\phi_3} e^{i\delta_T} |T^{ij}| - e^{-i\phi_1} e^{i\delta_P} |P^{ij}|.$$

The CP invariance of strong interaction requires the hadronic amplitudes T^{ij} and P^{ij} to be invariant under CP transformations. Thus, a CP transformation on A^{ij} applies a complex conjugation on weak phases only. Rotating consistently all CP transformed amplitudes \bar{A}^{ij} in

order to absorb the mixing phase, i.e. $\tilde{A}^{ij} = e^{-2i\phi_1} \bar{A}^{ij}$, the ratio between amplitudes \tilde{A}^{ij} and A^{ij} yields

$$\frac{\tilde{A}^{ij}}{A^{ij}} = \lambda_{hh} = \frac{e^{-i(\phi_1+\phi_3)} e^{i(\delta_T-\delta_P)} |T^{ij}| - |P^{ij}|}{e^{i(\phi_1+\phi_3)} e^{i(\delta_T-\delta_P)} |T^{ij}| - |P^{ij}|} = \frac{e^{i(\delta+\phi_2)} |T^{ij}| + |P^{ij}|}{e^{i(\delta-\phi_2)} |T^{ij}| + |P^{ij}|}, \quad (2.44)$$

where $\delta = \delta_T - \delta_P$ and $\phi_2 - \pi = -\phi_1 - \phi_3$ (see Sect. 2.4.3). In case of negligible penguin contributions, i.e. $|P^{ij}| \simeq 0$, one obtains the same result for λ_{hh} as in eq. (2.43). In case of non-negligible penguin contributions, one introduces an effective angle $\phi_{2,\text{eff}}$, which is also referred to as penguin *polluted* angle, obtaining

$$\frac{\tilde{A}^{ij}}{A^{ij}} = \lambda_{hh} = \left| \frac{\bar{A}^{ij}}{A^{ij}} \right| e^{2i\phi_{2,\text{eff}}}. \quad (2.45)$$

The non-negligible penguin contributions prevent a direct determination of ϕ_2 from measured CP -violation parameters. In 1990, two days before the birth of this thesis' author, Michael Gronau and David London proposed to determine the amount of penguin pollution from data making use of relations among the hadronic amplitudes [73]. These relations were obtained through an $SU(2)$ isospin analysis of the decay system $B \rightarrow \pi\pi$. A later review of the Gronau-London isospin analysis, clarified that the method could be used as well to extract amplitude relations for the decay systems $B \rightarrow \rho\rho$ and $B \rightarrow \rho\pi$ [74].

A further approach to determine the penguin contribution is to make use of the approximate flavor symmetry $SU(3)_f$ between the decays $B \rightarrow K\pi$, $B \rightarrow KK$ and $B \rightarrow \pi\pi$ [75, 76]. This approach works similarly for the decays $B \rightarrow K^*\rho$ and $B \rightarrow \rho\rho$ [77]. For the system $B \rightarrow \rho\pi$, isospin symmetry in a Dalitz plot analysis, and flavour symmetry $SU(3)_f$ have been applied to determine ϕ_2 [78, 79]. A treatment based on $SU(3)_f$ was proposed also for the $B \rightarrow a_0(1260)_1^{+-}\pi$ system [80]. Other approaches consider the extraction of ϕ_2 from $B \rightarrow a_0(980)\pi \rightarrow \eta\pi\pi$, $B \rightarrow a_0(980)\rho \rightarrow \eta\pi\pi\pi$, and other B decays to resonances with isospin $I = 1$ [81].

With the available data samples, the determination of ϕ_2 is dominated by the $B \rightarrow \rho\rho$ system, and to a lesser extent, by the $B \rightarrow \pi\pi$ system [71]. Considering the data of Belle, BaBar and LHCb for the systems $B \rightarrow \pi\pi$, $B \rightarrow \rho\rho$ and $B \rightarrow \rho\pi$, the current SM value of ϕ_2 obtained by the CKMfitter group is $(86.2_{-4.0}^{+4.4})^\circ$ [71]. One of the main goals of this thesis' work was to estimate the precision for ϕ_2 by projecting the measurements associated with the $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ systems. The estimated Belle II precision is $\Delta\phi_2 \approx 0.6^\circ$ (see chapter 7).

In the following, the determination of ϕ_2 performing isospin analyses of the systems $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ will be discussed in depth. Subsequently, the determination of ϕ_2 using $B \rightarrow \rho\pi$ decays will be discussed briefly.

2.10.1 Properties of π and ρ mesons

To address the isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$, one needs first to consider the properties of the final-state particles in these decays. The π and the ρ mesons are respectively the lightest pseudoscalar (spin-0) and vector (spin-1) mesons. Both particles build $SU(2)$ isospin triplets ($I = 1$):

- $\pi^+(u\bar{d})$, $\pi^0(\frac{u\bar{u}-d\bar{d}}{\sqrt{2}})$, $\pi^-(\bar{u}d)$ with $I_3 = -1, 0, +1$, and
- $\rho^+(u\bar{d})$, $\rho^0(\frac{u\bar{u}-d\bar{d}}{\sqrt{2}})$, $\rho^-(\bar{u}d)$ with $I_3 = -1, 0, +1$.

Charged pions π^\pm have a mass $m_{\pi^\pm} = 139.6 \text{ MeV}/c^2$ and decay weakly to about 99.99% into $\mu^+\nu_\mu(\mu^-\bar{\nu}_\mu)$. The mean lifetime is $\tau_{\pi^\pm} = 2.6 \cdot 10^{-8} \text{ s}$, corresponding to $c\tau_{\pi^\pm} \approx 7.8 \text{ m}$. The quantum numbers are $J^P = 0^-$.

Neutral pions π^0 have a mass $m_{\pi^0} = 135 \text{ MeV}/c^2$ which differs by about $4.6 \text{ MeV}/c^2$ from the charged pion mass. Neutral pions decay electromagnetically to $(98.823 \pm 0.034)\%$ into two photons and to $1.174 \pm 0.035\%$ into $e^+e^-\gamma$. The latter decay is referred to as Dalitz decay, after R. H. Dalitz, who suggested in 1951 that a π^0 could decay into a real and a virtual photon; The latter then produces directly an e^+e^- pair (Dalitz pair) [82]. The mean lifetime is $\tau_{\pi^0} = 8.4 \cdot 10^{-17} \text{ s}$, corresponding to $c\tau_{\pi^0} \approx 25 \text{ nm}$. Being its own antiparticle, the quantum numbers of a π^0 are $J^{PC} = 0^{-+}$.

The ρ mesons decay via the strong interaction and have therefore a very short lifetime corresponding to a broad width of $\Gamma_\rho = 149.1 \text{ MeV}$, which is approximately a fifth of its mass $m_\rho = 775.3 \text{ MeV}/c^2$. The quantum numbers of ρ^\pm are $J^P = 1^-$, and those of ρ^0 , which is its own antiparticle, are $J^{PC} = 1^{--}$. The ρ mesons decay almost exclusively into two pions. Since charge parity is conserved in strong interactions, ρ^0 decays only into $\pi^+\pi^-$.

The particle properties quoted above are world averages by the Particle Data Group [28].

2.10.2 Isospin analysis for $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$

The validity of the Gronau-London isospin analysis relies on the fact that isospin symmetry is almost fully conserved during the hadronization process in $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ decays. Thus, the relations between the decay amplitudes derived from the isospin symmetry can be used to disentangle the effects of the tree and the strong penguin contributions to obtain the unpolluted value of ϕ_2 [73].

Given a final state $h_1^i h_2^j$ with two identical mesons $h_1 = h_2 = h \in \pi, \rho$, the decay amplitudes A^{+0} , A^{+-} and A^{00} can be written as

$$A^{ij} \equiv \langle h^i h^j | \mathcal{H}_{\text{eff}} | B^{i+j} \rangle, \quad (2.46)$$

where \mathcal{H}_{eff} is the effective Hamiltonian describing the transition. Since the final-state pions and ρ mesons are bosons, the total wave function of the final state $h_1^i h_2^j$ must be symmetric

under particle exchange. For $i \neq j$, the symmetrized final states can be written as

$$|h^i h^j\rangle = \sqrt{\frac{1}{2}} (|h_1^i h_2^j\rangle + |h_1^j h_2^i\rangle).$$

Since B mesons are spin-0 particles, the total angular momentum of the final state is $J = 0$. Regarding now the isospin, for a single π or ρ the value is 1. Spin sum rules dictate that the total final-state isospin I_f for a pair $\pi\pi$ or $\rho\rho$ can be 0, 1 or 2. However, due to Bose statistics only final states with $I_f = 0$ or 2 are allowed. A final state $I_f = 1$ would be antisymmetric as the symmetry under particle exchange considering isospin is $(-1)^{J+I}$ [74]. This statement is exact if the particles in the final state have equal masses. In the case of ρ mesons, which have a significant width, the possible mass difference between the particles in the final state could give rise to a final state with $I_f = 1$ [83]. Nevertheless, it was shown that in absence of a particular enhancement of the $I_f = 1$ amplitude, the results for ϕ_2 are completely insensitive to the ρ width [84]: the possible $I_f = 1$ contribution can be precisely estimated only with high statistics data at SuperKEKB and will be therefore neglected in this work. Considering now the possible values for I_f , the three relevant final states are,

$$|h^+ h^0\rangle = \sqrt{\frac{1}{2}} (|h_1^+ h_2^0\rangle + |h_1^0 h_2^+\rangle) = |2, 1\rangle, \quad (2.47)$$

$$|h^+ h^-\rangle = \sqrt{\frac{1}{2}} (|h_1^+ h_2^-\rangle + |h_1^- h_2^+\rangle) = \sqrt{\frac{1}{3}} |2, 0\rangle + \sqrt{\frac{2}{3}} |0, 0\rangle, \quad (2.48)$$

$$|h^0 h^0\rangle = \sqrt{\frac{2}{3}} |2, 0\rangle - \sqrt{\frac{1}{3}} |0, 0\rangle. \quad (2.49)$$

For the isospin analysis, the decay amplitudes in eq. (2.46) can be factorized in two parts: the weak decay $b \rightarrow u\bar{u}d$ corresponding to an isospin transition ΔI , and the hadronization into two light mesons [71]. Following Harry Lipkin, Helen Quinn et al. [74], the Hamiltonian for the quark transition $\bar{b} \rightarrow \bar{u}u\bar{d}$, in terms of $A_{\Delta I}$ amplitudes, is of the form

$$\mathcal{H}_{\text{eff}} = A_{\frac{3}{2}} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle + A_{\frac{1}{2}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle, \quad (2.50)$$

Multiplication with the initial states B^+ and B^0 gives,

$$\mathcal{H}_{\text{eff}}|B^+\rangle = \mathcal{H}_{\text{eff}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \sqrt{\frac{3}{4}} A_{\frac{3}{2}} |2, 1\rangle + (A_{\frac{1}{2}} - \frac{1}{2} A_{\frac{3}{2}}) |1, 1\rangle, \quad (2.51)$$

$$\mathcal{H}_{\text{eff}}|B^0\rangle = \mathcal{H}_{\text{eff}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} A_{\frac{3}{2}} |2, 0\rangle + \sqrt{\frac{1}{2}} (A_{\frac{1}{2}} + A_{\frac{3}{2}}) |1, 0\rangle + \sqrt{\frac{1}{2}} A_{\frac{1}{2}} |0, 0\rangle. \quad (2.52)$$

The equations above correspond to four-quark states $|\bar{u}u\bar{d}, q\rangle$ with $q = u$ or d , while the states in eqs. (2.47) to (2.49) are two-meson states. The transition between the two necessarily

involves hadronization and other rescattering effects. Applying the Wigner-Eckhart theorem, these amplitudes can be expressed in terms of reduced matrix elements $A_{\Delta I, I_f}$, where ΔI is the isospin shift and I_f is the final-state isospin. Hadronization and strong rescattering effects absent in the amplitudes $A_{\Delta I}$ are included into the amplitudes $A_{\Delta I, I_f}$. A detailed derivation of eqs. (2.47) to (2.52) can be found in App. A.4. The projection of the states $\mathcal{H}_{\text{eff}}|B^{i+j}\rangle$ onto the final states gives,

$$A^{+0} = \langle h^+ h^0 | \mathcal{H}_{\text{eff}} | B^+ \rangle = \sqrt{\frac{3}{4}} A_{\frac{3}{2}, 2}, \quad (2.53)$$

$$A^{+-} = \langle h^+ h^- | \mathcal{H}_{\text{eff}} | B^0 \rangle = \sqrt{\frac{1}{6}} A_{\frac{3}{2}, 2} + \sqrt{\frac{1}{3}} A_{\frac{1}{2}, 0}, \quad (2.54)$$

$$A^{00} = \langle h^0 h^0 | \mathcal{H}_{\text{eff}} | B^0 \rangle = \sqrt{\frac{1}{3}} A_{\frac{3}{2}, 2} - \sqrt{\frac{1}{6}} A_{\frac{1}{2}, 0}. \quad (2.55)$$

Similarly, one obtains for the CP conjugated decay amplitudes \bar{A}^{ij} (see App. A.4)

$$\bar{A}^{+0} = \langle h^- h^0 | \mathcal{H}_{\text{eff}} | B^- \rangle = \sqrt{\frac{3}{4}} \bar{A}_{\frac{3}{2}, 2}, \quad (2.56)$$

$$\bar{A}^{+-} = \langle h^+ h^- | \mathcal{H}_{\text{eff}} | \bar{B}^0 \rangle = \sqrt{\frac{1}{6}} \bar{A}_{\frac{3}{2}, 2} - \sqrt{\frac{1}{3}} \bar{A}_{\frac{1}{2}, 0}, \quad (2.57)$$

$$\bar{A}^{00} = \langle h^0 h^0 | \mathcal{H}_{\text{eff}} | \bar{B}^0 \rangle = \sqrt{\frac{1}{3}} \bar{A}_{\frac{3}{2}, 2} + \sqrt{\frac{1}{6}} \bar{A}_{\frac{1}{2}, 0}. \quad (2.58)$$

The CP conjugated amplitudes $\bar{A}_{\Delta I, I_f}$ carry strong phases identical to those of $A_{\Delta I, I_f}$, but opposite weak phases. From eqs. (2.53) to (2.55) and eqs. (2.56) to (2.58), one obtains finally two relations among the hadronic amplitudes, which allow to extract the penguin pollution,

$$\begin{aligned} A^{+0} - A^{00} &= \sqrt{\frac{1}{2}} A^{+-}, \\ \bar{A}^{+0} - \bar{A}^{00} &= \sqrt{\frac{1}{2}} \bar{A}^{+-}. \end{aligned} \quad (2.59)$$

These equations are referred to as isospin triangles since they can be represented as triangles in the complex space. The triangle relations hold also for consistently rotated amplitudes. In this work, the convention $\tilde{A}^{ij} = e^{-2i\phi_1} \bar{A}^{ij}$ is used.

Strong penguin diagrams can lead only to $\Delta I = \frac{1}{2}$ transitions. Since the amplitudes in eqs. (2.53) and (2.56) lack of $\Delta I = \frac{1}{2}$ components, the decay $B^\pm \rightarrow h^\pm h^0$ occurs purely as a tree diagram. Using equation eq. (2.44), one obtains

$$\frac{\tilde{A}^{+0}}{A^{+0}} = \frac{e^{i(\delta+\phi_2)}}{e^{i(\delta-\phi_2)}} = e^{2i\phi_2} \Rightarrow \mathcal{A}^{h^+ h^-} = 0. \quad (2.60)$$

In this case, the imaginary component of the ratio cannot be measured since there is no mixing-induced CP Parameter. However, the measured CP parameters $\mathcal{A}^{\pi^+\pi^0} = 0.03 \pm 0.04$ and $\mathcal{A}^{\rho^+\rho^0} = -0.05 \pm 0.05$ show good agreement with zero (PDG world averages [28]). Therefore, $\mathcal{A}^{h^+h^-}$ is not taken into account for the determination of ϕ_2 . Using now eq. (2.45) for the penguin polluted decay $B^0 \rightarrow h^+h^-$, the triangle relations in eq. (2.59), and the relation in eq. (2.60), one obtains all ingredients to determine ϕ_2 considering isospin symmetry. By convention, one amplitude can be chosen to be real; in this case A^{+-} is chosen. One obtains [85],

$$\begin{aligned}
A^{+0} &= |A^{+0}| e^{i(\delta-\phi_2)}, & \tilde{A}^{+0} &= |A^{+0}| e^{i(\delta+\phi_2)}, \\
A^{+-} &= |A^{+-}|, & \tilde{A}^{+-} &= |\bar{A}^{+-}| e^{2i\phi_{2,\text{eff}}}, \\
A^{00} &= A^{+0} - \frac{A^{+-}}{\sqrt{2}} = e^{i(\delta-\phi_2)} \left(|A^{+0}| - \frac{|A^{+-}|}{\sqrt{2}} e^{-i(\delta-\phi_2)} \right), \\
\tilde{A}^{00} &= \tilde{A}^{+0} - \frac{\tilde{A}^{+-}}{\sqrt{2}} = e^{i(\delta+\phi_2)} \left(|A^{+0}| - \frac{|\bar{A}^{+-}|}{\sqrt{2}} e^{-i(\delta+\phi_2-2\phi_{2,\text{eff}})} \right).
\end{aligned} \tag{2.61}$$

For the branching fractions \mathcal{B}^{ij} and the CP parameters \mathcal{A}^{ij} and \mathcal{S}^{ij} it follows

$$\begin{aligned}
\mathcal{B}^{+0} &= \tau_{B^+} |A^{+0}|^2, & \mathcal{B}^{+-} &= \tau_{B^0} \frac{|A^{+-}|^2 + |\bar{A}^{+-}|^2}{2}, \\
\mathcal{B}^{00} &= \tau_{B^0} \left(|A^{+0}|^2 + \frac{|A^{+-}|^2 + |\bar{A}^{+-}|^2}{4} - \frac{|A^{+0}|}{\sqrt{2}} (|A^{+-}|c + |\bar{A}^{+-}|\bar{c}) \right), \\
\mathcal{A}^{+-} &= \frac{|\bar{A}^{+-}|^2 - |A^{+-}|^2}{|\bar{A}^{+-}|^2 + |A^{+-}|^2}, & \mathcal{S}^{+-} &= \frac{2|\bar{A}^{+-}||A^{+-}|\sin(2\phi_{2,\text{eff}})}{|\bar{A}^{+-}|^2 + |A^{+-}|^2}, \\
\mathcal{A}^{00} &= \frac{|\bar{A}^{+-}|^2 - |A^{+-}|^2 - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{c} - |A^{+-}|c)}{4|A^{+0}|^2 + |\bar{A}^{+-}|^2 + |A^{+-}|^2 - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{c} + |A^{+-}|c)},
\end{aligned} \tag{2.62}$$

where $c = \cos(\phi_2 - \delta)$ and $\bar{c} = \cos(\phi_2 + \delta - 2\phi_{2,\text{eff}})$. The different lifetimes for charged and neutral B mesons τ_{B^+} and τ_{B^0} are included in the branching fractions. In case that all branching fractions and CP -violation parameters in eq. (2.62) are measured, one obtains a system of 6 linear independent equations with 6 real positive variables: $|A^{+0}|$, $|A^{+-}|$, $|\bar{A}^{+-}|$, δ , $\phi_{2,\text{eff}}$ and ϕ_2 . From these equations, the value of ϕ_2 can be determined up to an 8-fold ambiguity in the range $[0, \pi]$ because of two reasons. First, each isospin triangle can have two possible orientations, leading to a 4-fold trigonometric ambiguity

$$(\phi_2, \delta) \leftrightarrow (\delta, \phi_2), (2\phi_{2,\text{eff}} - \phi_2, 2\phi_{2,\text{eff}} - \delta), (2\phi_{2,\text{eff}} - \delta, 2\phi_{2,\text{eff}} - \phi_2). \tag{2.63}$$

Second, there is an additional symmetry involving also $\phi_{2,\text{eff}}$: the functions c , \bar{c} and \mathcal{S}^{+-} are invariant under the reflection:

$$(\phi_{2,\text{eff}}, \phi_2, \delta) \leftrightarrow \left(\frac{\pi}{2} - \phi_{2,\text{eff}}, \frac{\pi}{2} - \phi_2, \frac{\pi}{2} - \delta \right). \tag{2.64}$$

An illustration of the 8-fold ϕ_2 ambiguity is shown in Fig. 2.16.

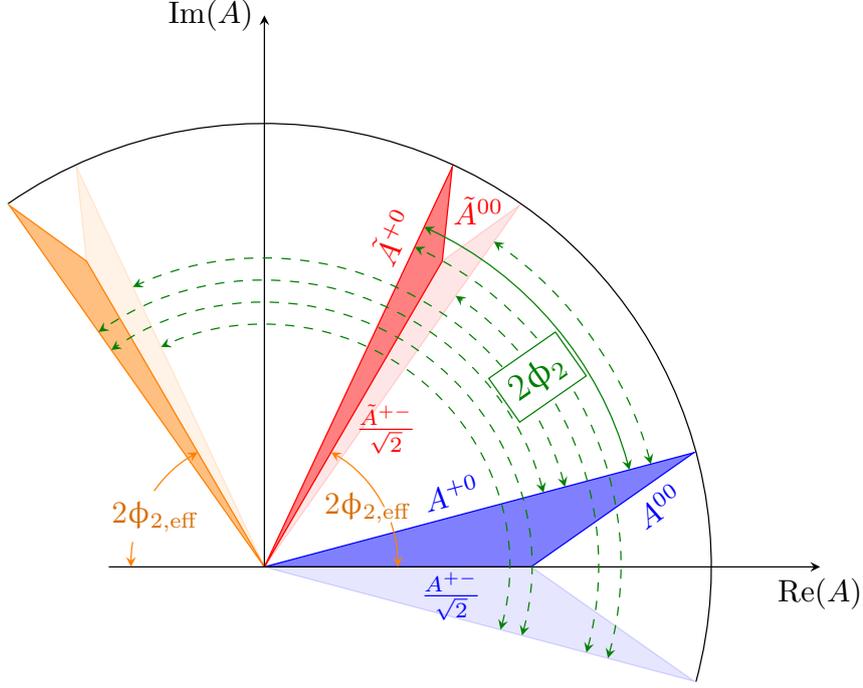


Figure 2.16: Geometrical representation of the isospin triangular relations (eq. (2.59)) in the complex plane of $B^{i+j} \rightarrow h^i h^j$ amplitudes. The blue and the red shaded areas correspond to the isospin triangles. The angle between the CP conjugate charged amplitudes A^{+-} and \bar{A}^{+-} corresponds to twice the weak phase $\phi_{2,\text{eff}}$ (orange solid lines). The angle between the CP conjugate charged amplitudes A^{+0} and \bar{A}^{+0} corresponds to twice the CKM angle ϕ_2 (green solid line). The other triangles with lighter shade represent the mirror solutions allowed by the discrete ambiguities in eq. (2.62), with the corresponding values for ϕ_2 represented by the green dashed lines.

The \mathcal{S}^{00} constraint

An additional constraint can partially lift the 8-fold ambiguity in the determination of ϕ_2 : the mixing-induced CP -violation parameter of the decay $B^0 \rightarrow h^0 h^0$,

$$\mathcal{S}^{00} = \frac{4|A^{+0}|^2 \sin(2\phi_2) + 2|A^{+-}||\bar{A}^{+-}| \sin(2\phi_{2,\text{eff}}) - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{s} + |A^{+-}|s)}{4|A^{+0}|^2 + |\bar{A}^{+-}|^2 + |A^{+-}|^2 - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{c} + |A^{+-}|c)}, \quad (2.65)$$

where $s = \sin(\phi_2 + \delta)$ and $\bar{s} = \sin(\phi_2 - \delta + 2\phi_{2,\text{eff}})$ are not invariant under the transformations of (ϕ_2, δ) in eq. (2.63), thus fixing the orientation of each isospin triangle.

Since the postulation of the isospin analysis in 1990, the measurement of \mathcal{S}^{00} for the decay $B^0 \rightarrow \pi^0 \pi^0$ has not been feasible. Its measurement is a real experimental challenge because the time-dependent analysis of $B^0 \rightarrow \pi^0 \pi^0$ requires a precise reconstruction of the B^0 -decay vertex, which cannot be achieved in the dominant four-photons final state. To make use of rare

events with photon conversions $\gamma_c \rightarrow e^+e^-$ or with neutral pions undergoing Dalitz decays $\pi^0 \rightarrow e^+e^-\gamma$, large samples of $B^0 \rightarrow \pi^0\pi^0$ decays are needed. At present, the size of the Belle and the BaBar data samples is insufficient and, at LHCb, the reconstruction of these kind of events is impossible due to the harsh hadronic environment (see Sect. 2.8). In this thesis, a novel analysis method was developed to make use of e^+e^- pairs to reconstruct the B^0 -decay vertex. The goal was to figure out if Belle II will be able to measure $\mathcal{S}^{\pi^0\pi^0}$. The study will be presented in chapter 6.

For the decay $B^0 \rightarrow \rho^0\rho^0$, a single measurement of $\mathcal{S}^{\rho^0\rho^0}$ by BaBar is available with large uncertainties $\mathcal{S}^{\rho^0\rho^0} = 0.3 \pm 0.7 \pm 0.2$ [86]. However, for $B \rightarrow \rho\rho$, the sides A^{00} and \tilde{A}^{00} of the isospin triangles are much smaller than the other sides, and the triangles are shrunk into lines. The 8-fold ambiguity is therefore already reduced by a factor four (see Fig. 7.5 in Sect. 7.5.2).

Differences between $\pi\pi$ and $\rho\rho$

A major difference between the $\pi\pi$ and the $\rho\rho$ final states are the possible angular momentum configurations. This is important because the CP eigenvalue of the final state $h^i h^j$ is given by $\eta_{CP} = (-1)^L$. In the case of $\pi\pi$, the CP eigenvalue is even: the total spin S is zero, and the orbital angular momentum L has to be zero to conserve the initial $J = L + S = 0$. On the contrary, the pair of spin-1 ρ mesons can have spin configurations $S = 0, 1$ and 2 . The conservation of $J = 0$ imposes the orbital angular momentum L to be equal and oppositely aligned to S , leading to $L = 0, 1$ or 2 ; where $L = 0$ and 2 correspond to CP -even, and $L = 1$ to CP -odd final states. Thus, independent isospin analyses have to be performed for the CP -even and the CP -odd modes, requiring the branching fractions and the CP asymmetries in eqs. (2.62) and (2.65) to be measured in both cases.

Performing an angular analysis of the decays $B^0 \rightarrow \rho^+\rho^-$ and $B^+ \rightarrow \rho^+\rho^0$, it was measured that the fraction f_L of decays leading to longitudinally polarized mesons $\rho_L\rho_L$, i.e. with $L = 0$, dominates over a negligible fraction of events leading to final states with $L \neq 0$ [87–90]. The current world averages are $f_{L,\rho^+\rho^0} = 0.95 \pm 0.016$ and $f_{L,\rho^+\rho^-} = 0.990_{-0.019}^{+0.021}$ [28]. Consequently, the isospin analysis of $B \rightarrow \rho\rho$ is performed only for decays leading to longitudinally polarized ρ_L mesons.

Isospin breaking effects

In chapter 7, the Gronau-London isospin analysis is applied to estimate the Belle II precision for the determination of ϕ_2 . However, for a precise determination of ϕ_2 , one needs to consider isospin-breaking effects coming mainly from electroweak contributions and from the small mass difference between the u and the d quarks.

EW penguins contribute for example to $\Delta I = \frac{1}{2}$ and $\Delta I = \frac{3}{2}$ transitions, leaving the isospin relations (eq. (2.59)) invariant, but affecting the additional relation $A^{+0} = e^{-2i\phi_2} \tilde{A}^{+0}$ in eq. (2.60), as these amplitudes are no longer pure tree level processes [91, 92]. Other contributions from EW penguins introduce $A_{\frac{3}{2},2}$ amplitudes in eqs. (2.53) to (2.58) invalidating the isospin relations [93, 94].

Considering the mass difference between u and d quarks, this difference indicates that there is a mismatch between isospin eigenstates and mass eigenstates. This mismatch causes mixing between mesons: $\pi^0 - \eta, \eta'$ [95] and $\rho - \omega$ mixing [96]. Furthermore, in the $B \rightarrow \pi\pi$ system, the $\pi^0 - \eta, \eta'$ mixing introduces an additional $I_f = 1$ amplitude [92].

At present, the CKMfitter group considers only part of the isospin-breaking effects affecting the isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$; the group estimates a shift $\Delta\phi_2$ around -2° near the SM-compatible solution and an increase in the uncertainty at 68% confidence level of around 1.5° [71]. At Belle II, the precision for ϕ_2 performing isospin analyses is expected to be $\lesssim 1^\circ$ (see chapter 7) and, therefore, isospin-breaking effects will have to be covered accurately in the determination of ϕ_2 . For this, detailed studies analyzing the inclusion of isospin-breaking effects are still needed [97].

2.10.3 Determination from $B \rightarrow \rho\pi$

The determination of ϕ_2 from $B \rightarrow \rho\pi$ is more complicated than the determination from $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$. The reason is that the final states $\rho^\pm\pi^\mp$ are not CP eigenstates. Following Jure Zupan [98], there are essentially two approaches: to perform a full time-dependent Dalitz plot analysis of $B \rightarrow \pi^+\pi^-\pi^0$ considering isospin symmetry [74, 78], or to consider only the $\rho^\pm\pi^\mp$ region together with $SU(3)_f$ related modes of the type $B \rightarrow K^*\pi$ and $B \rightarrow \rho K$ [79]. The first method requires a particular analysis which has been pioneered by Belle [99] and by BaBar [100], and will be briefly discussed hereafter. The second requires only the individual measurement of the branching fractions and the CP parameters of the respective modes, but suffers from larger theoretical uncertainties due to $SU(3)_f$ -breaking effects which have been only partially studied [79, 98].

For the isospin analysis of $B \rightarrow \rho\pi$, one has again the possible transitions $\Delta I = \frac{1}{2}, \frac{3}{2}$ and eqs. (2.50) to (2.52) hold. However, for the final states one has $I_f \in \{0, 1, 2\}$. The analysis leads to two independent relations, one among the hadronic A^{ij} amplitudes and another among the \tilde{A}^{ij} amplitudes, which can be represented as pentagons in the complex space and are therefore called, in analogy to the isospin triangles, isospin pentagons. The phases between the amplitudes A^{ij} , and the CP conjugated amplitudes \tilde{A}^{ij} can be extracted through a time-dependent Dalitz plot analysis. This analysis is sensitive to the interference between the strong and the weak amplitudes in the regions where the ρ^+, ρ^- and ρ^0 resonances overlap.

Unlike for $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$, the isospin relations for $B \rightarrow \rho\pi$ relate only the penguin amplitudes. Since the penguin amplitudes are smaller than the tree amplitudes, the uncertainty on ϕ_2 due to isospin-breaking effects is expected to very small around 0.1° [92]. Furthermore, ϕ_2 can be unambiguously determined from $B \rightarrow \rho\pi$. There are eleven independent observables, which are obtained from a set of 27 mutually dependent measurable coefficients [101]. The observables are described by 12 variables: ϕ_2 , 6 magnitudes of tree and penguin amplitudes, and 5 strong phases between these amplitudes. Due to the pentagon relations, the number of free variables are only 6 including ϕ_2 and, thus, the system is over-constrained providing ϕ_2 without ambiguities.

Both collaborations, Belle and BaBar, succeeded in extracting meaningful information about ϕ_2 and discovered the existence of two “secondary solutions” on either side of the expected primary solution [99, 100]. The secondary solutions do not arise from ambiguities, but are rather artifacts resulting from the small amount of signal events and are expected to vanish with more data. This strongly motivates the repetition of the analysis at Belle II, with a data sample of at least few ab^{-1} . Experimentally, the major challenge will consist in modeling correctly the tails of the ρ bands and in distinguishing the $\rho\pi$ signal among the interfering resonant and non-resonant contributions.

It is very difficult to simulate the interfering resonant and non-resonant contributions to the $B \rightarrow \pi^+\pi^-\pi^0$ decay, and to realize a performance study of the full time-dependent Dalitz-plot analysis. Furthermore, the uncertainties of the 27 measurable coefficients, which are needed as input to extract ϕ_2 , cannot be projected in a simple way. Therefore, the decay $B \rightarrow \rho\pi$ is not considered in this work for the projection of the ϕ_2 sensitivity at Belle II.

2.11 Prospects for the unitarity angles

Since the projection of the measurement of the unitarity angle ϕ_2 at Belle II plays a central role in this thesis, it is interesting to compare briefly which are the prospects for the measurements of all the three unitarity angles. Considering the unitarity angles ϕ_1 and ϕ_2 , the precision on the value of these two angles will be certainly dominated by Belle II. Regarding the angle ϕ_1 , LHCb potentially will be competitive in the case of the golden channel $B^0 \rightarrow J/\psi K_S^0$. However, a precise determination of ϕ_1 will require other measurements that are not covered by LHCb, such as $B^0 \rightarrow J/\psi K_L^0$, which are fundamental part of the Belle II program.

In the case of the angle ϕ_2 , most of the useful channels contain one or two neutral pions. Thus LHCb will not be able to compete for ϕ_2 with Belle II. However, at the level of precision expected at Belle II, which is $\lesssim 0.1^\circ$ for ϕ_1 and $\lesssim 1^\circ$ for ϕ_2 , it is possible that the values of these angles will be dominated by theoretical uncertainties (see Sect. 2.9 and Sect. 2.10.2).

This is not the case for ϕ_3 . Its determination is free of theoretical uncertainties, but, at the same time, it requires huge sample sizes, and (depending on the method) also auxiliary charm factory measurements to reach a precision around few degrees. To be able to reach a high precision in ϕ_3 , Belle II and LHCb plan to make use of auxiliary measurements provided by BES III, and aim to reach a precision around 1° on similar time scales [1, 57]. Thus, in the area of flavor physics, the race for ϕ_3 will be a very exciting one. Keep your seat belts fasten!

3 From Belle to Belle II

The Belle experiment was designed to precisely measure the CP -asymmetries in the decays of the B mesons that were produced at the KEKB collider. Thanks to the high instantaneous luminosity at KEKB, whose peak luminosity around $2.11 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ is the current world record, Belle was able to collect a data sample corresponding to about 1 ab^{-1} at five $\Upsilon(nS)$ resonances. Of this sample, about 0.8 ab^{-1} were collected at the $\Upsilon(4S)$.

Despite the large integrated luminosity, the precision at Belle is still limited by the size of the data sample (see Sect. 2.9). To increase the luminosity, the KEKB collider was shut down in June of 2010, after more than ten years of running, to be upgraded to the new SuperKEKB collider. SuperKEKB is scheduled to start the physics run in spring of 2019 and is designed to reach an instantaneous luminosity of $80 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, leading ultimately to an integrated luminosity of 50 ab^{-1} by 2025.

To cope with the challenging new experimental conditions, the Belle detector was substantially upgraded to become the new Belle II detector. The new Belle II detector is equipped with novel sub-detectors for particle identification and particle track reconstruction, especially a novel pixel detector, which will provide position measurements for the precise reconstruction of the decay vertex of short-lived particles. To maximally exploit the capabilities of the new detector, the Belle II collaboration started the development of completely new reconstruction and analysis algorithms.

Thanks to the expected large integrated luminosity at SuperKEKB, to the improvements in the detector, in the reconstruction algorithms, and in the analysis algorithms, and thanks to the novel triggers, Belle II will be able to reach a much higher precision than Belle, and to search for New Physics in sectors that remained uncovered at Belle.

In this thesis, most of the studies were performed using Belle II MC events. Belle MC and Belle data were used to validate the developed flavor tagging algorithm (chapter 4). Additionally, Belle measurements served as reference to find reconstruction criteria for the decay $B^0 \rightarrow \pi^0\pi^0$ (chapter 6), and to estimate the Belle II precision for the unitarity angle ϕ_2 (chapter 7). Therefore, this chapter aims at briefly introducing the Belle and the Belle II experiments together with the KEKB and the SuperKEKB colliders, basing on the comprehensive descriptions in [102–105].

3.1 From KEKB to SuperKEKB

KEKB was an asymmetric e^+e^- collider, where the term “asymmetric” refers to the energy difference between the colliding electrons and positrons. Fig. 3.1 shows a schematic overview

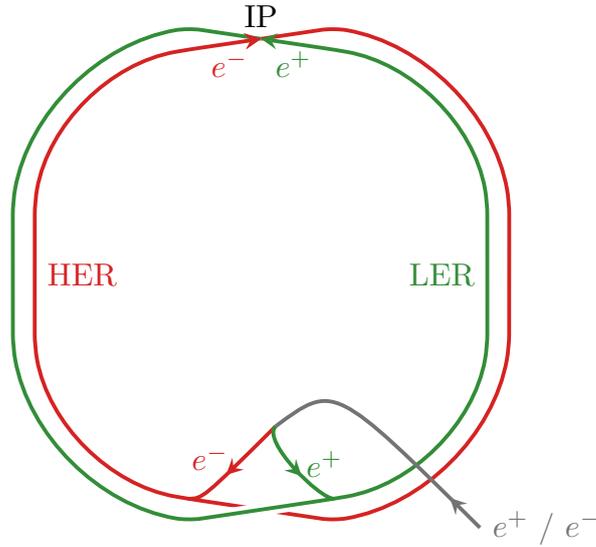


Figure 3.1: The KEKB double-ring structure. Shown are the low energy ring (LER), the high energy ring (HER), and the injection tunnel leading to the linear accelerator. The electrons and positrons collide at the interaction point (IP), which is surrounded by the Belle II detector.

of KEKB. It consisted of a linear accelerator and a double storage ring of about 3 km circumference [104]. The linear accelerator brought the electrons and the positrons to their respective target energies before injecting them in bunches into the independent storage rings: the *high-energy ring* (HER) for the electrons, and the *low-energy ring* (LER) for the positrons. The electrons and positrons were brought to collision at the interaction point (IP), which was surrounded by the Belle detector. The electron and the positron beams crossed with an angle of 22 mrad.

The SuperKEKB collider adopted the linear accelerator and the double storage ring of KEKB. However, several modifications were needed in order to reach the design instantaneous luminosity. Tab. 3.1 presents a summary of the most significant machine parameters at KEKB and at SuperKEKB. These parameters and the respective modifications will be discussed briefly in the following.

The instantaneous luminosity \mathcal{L} is a measure for the collider's performance. Given \mathcal{L} , the event rate for a process with cross section σ can be written as

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma.$$

To increase the event rate, SuperKEKB aims at increasing the instantaneous luminosity by a factor 40 compared to the peak luminosity at KEKB. For two beams with a Gaussian profile of horizontal and vertical size σ_x and σ_y , the instantaneous luminosity is given by

$$\mathcal{L} = \frac{N_{e^+} N_{e^-} f_c}{4\pi\sigma_x\sigma_y} \cdot R_L,$$

Table 3.1: Fundamental parameters at KEKB and at SuperKEKB [104, 105].

		KEKB	SuperKEKB
Energy (LER/HER) [GeV]	E	3.5/8.0	4.0/7.0
Instantaneous luminosity [$10^{34} \text{ cm}^{-2} \text{ s}^{-1}$]	\mathcal{L}	2.11	80
Beam current (LER/HER) [A]	I	1.64/1.19	3.60/2.62
Vertical Beta Function at IP [mm]	β_y^*	5.9/5.9	0.27/0.41
Beam crossing angle [mrad]	ϕ_c	22 ($\approx 1.3^\circ$)	83 ($\approx 4.8^\circ$)
Horizontal beam size at IP [μm]	σ_x^*	90	9
Vertical beam size at IP [nm]	σ_y^*	1900	60
Boost	$\beta\gamma$	0.425	0.287

where N_{e^+} and N_{e^-} are the number of particles in a positron and in an electron bunch, f_c is the bunch crossing frequency, and R_L is a reduction factor accounting for geometrical effects associated with the finite crossing angle and the bunch length. To increase the luminosity, SuperKEKB aims at reducing the beam sizes, and at increasing the beam currents $I_{e^+/e^-} \propto N_{e^-/e^+} f_c$ by a factor of two compared to KEKB.

At any given point s along the accelerator ring, the beam sizes can be written as

$$\sigma_{x/y}(s) = \sqrt{\epsilon_{x/y} \beta_{x/y}(s)},$$

where ϵ is the *emittance* and β is the so-called *beta function*. The beta function is a measure of the transverse beam size along the ideal trajectory, describing the effect of focusing and defocusing magnets, and it is minimized at the collision point. The emittance is associated with the initial beam divergence at the injection into the storage ring and remains unaffected by the focusing system. To achieve a low emittance, a new *damping ring* was constructed for the positrons in order to cool their transverse momentum. Since it is easier to achieve a low emittance with a higher energy, the beam energies at SuperKEKB were modified to further reduce the emittance of the positron beam.

The major new feature in the luminosity upgrade of SuperKEKB is the so-called *nano beam scheme*, where the vertical beta function at the IP, β_y^* , is squeezed down to about 0.3 mm (around 20 times smaller than at KEKB) resulting in a vertical beam size of only $\sigma_y^* \approx 60$ nm. To achieve the nano beam scheme, the beam crossing angle ϕ_c was increased by about a factor of four compared to KEKB. Figure 3.2 shows a schematic view of the beam crossing at KEKB and at SuperKEKB. At SuperKEKB, the size of the interaction region in z -direction σ_z is about 20 times smaller than at KEKB [106]. This novel experimental condition has an impact on the performance of the flavor tagger, as it will be discussed in Sect. 4.6 and chapter 5.

The most important running mode of SuperKEKB is at the $\Upsilon(4S)$, corresponding to a center-of-mass energy of $\sqrt{s} = 2\sqrt{E_{\text{LER}} E_{\text{HER}}} = 10.58$ GeV. Due to the asymmetric beam energies

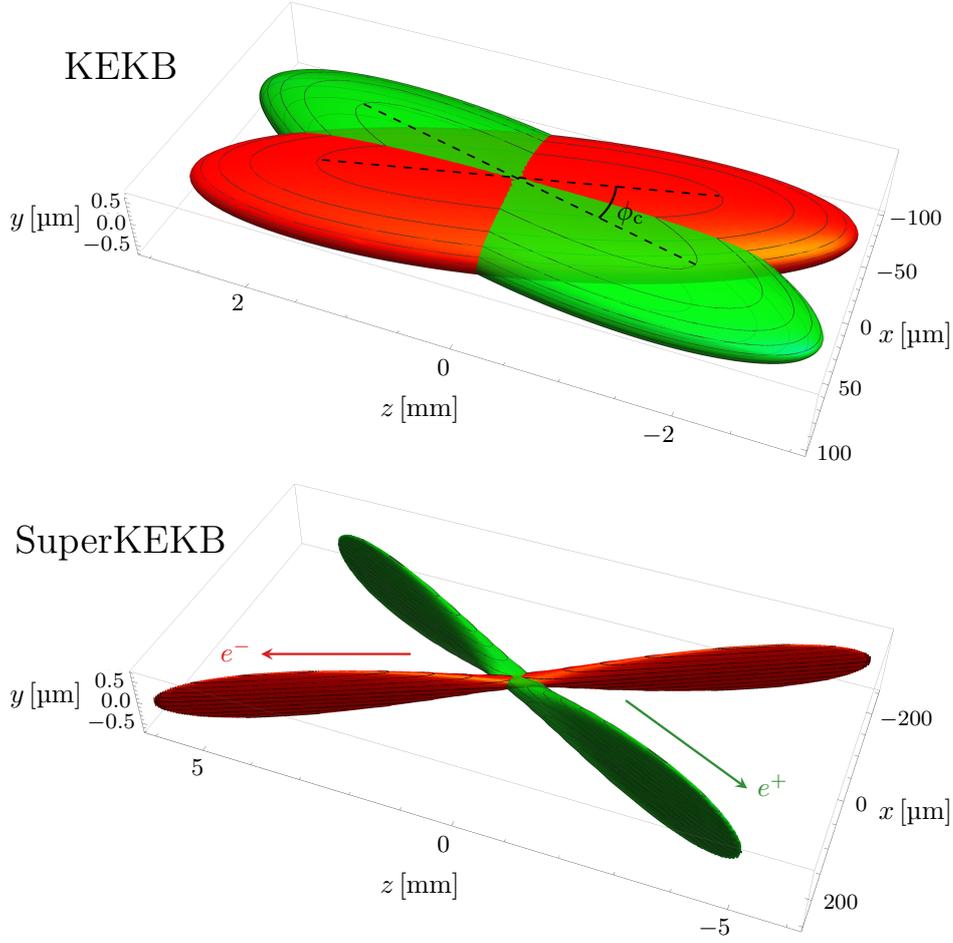


Figure 3.2: Schematic view of the beam crossing at KEKB (top) and at SuperKEKB (bottom). The definition of the crossing angle is shown in the top. The beam directions are shown in the bottom. The size of the interaction region in z -direction at KEKB was around 10 mm, while at SuperKEKB it is around 0.5 mm [106].

of the electron and the positron beams, the $\Upsilon(4S)$ is produced with a boost $\beta\gamma$. The boost depends on the beam energies and on the crossing angle between them. For a head-on collision, the boost magnitude is given by

$$\beta = \frac{p_{\Upsilon(4S)}c}{E_{\Upsilon(4S)}} = \frac{E_{\text{HER}} - E_{\text{LER}}}{E_{\text{HER}} + E_{\text{LER}}}, \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}, \quad \beta\gamma = \frac{E_{\text{HER}} - E_{\text{LER}}}{\sqrt{s}}.$$

As discussed in Sect. 2.7.3, the boost is needed to measure the time difference between the decays of the two B mesons produced at the $\Upsilon(4S)$ (eq. (2.34)). Due to the reduction of the beam-energy asymmetry at SuperKEKB, the boost is reduced by factor $\approx 2/3$ compared to KEKB. Thus, the average distance Δl (in boost direction) between the two decay vertices, of about 190 μm at KEKB, is reduced to about 130 μm at SuperKEKB. To avoid a deterioration of

the Δt resolution, the new Belle II experiment has a novel vertex detector, as well as improved track and vertex reconstruction algorithms.

3.2 From the Belle to the Belle II detector

The Belle detector was a large solid-angle magnetic spectrometer installed at the interaction point of KEKB [102]. Besides being able to precisely measure the time-dependent CP -asymmetries in B -meson decays, Belle was able to cover many other topics thanks to its detector capabilities [29]. It contained various charged particle tracking and particle identification (PID) detectors as well as an electromagnetic calorimeter. These components were enclosed in a magnetic solenoid with an instrumented iron return yoke.

Before reaching the Belle detector, all particles had to traverse a double-walled beryllium beam pipe, which had a small radius of 20 mm (15 mm after 2003) to optimize the vertexing performance. The gap between the beryllium layers served as cooling channel. To shield the detector against synchrotron radiation (< 5 keV), the outer surface of the beam pipe was covered with a 20 μm gold foil. The total material thickness of the beryllium was about 0.3% of a radiation length, and the one of the gold foil was about 0.6%.

The innermost part of the detector was a silicon vertex detector (SVD). The task of the SVD was to provide position measurements for the precise reconstruction of the vertex location of short lived particles. It had a barrel design and consisted of three layers of double-sided silicon strip detectors (SVD1) until 2003, when it was replaced by one with four layers (SVD2). The layers of the SVD were reinforced with support ribs.

The major charged particle tracking detector was the central drift chamber (CDC). It served to precisely measure the momentum of charged particles from their track curvature. The CDC provided also particle identification by measuring the energy loss per unit length of the particles, dE/dx , which depends on the particle velocity. This information was specially important for low-momentum particles that did not reach the outer PID detectors. The CDC had 50 cylindrical layers of sense wires. Approximately half of them were parallel to the z -axis to measure the transverse momentum p_t (axial wires), while the other half was slanted by a small angle of about ± 50 mrad to measure the polar angle θ of the tracks (stereo wires).

For charged particle identification, Belle had two dedicated sub-detectors: an aerogel Cherenkov counter (ACC) and a time of flight counter (TOF). These detectors exploited the Cherenkov effect: a particle emits a cone of Cherenkov photons when it passes through a dielectric medium at a speed greater than the speed of light in the medium. Since the emission angle depends only on the particle's velocity, the velocity can be measured independently from the momentum to obtain the particle's mass. A part of the ACC was located in the barrel region surrounding the CDC, and another part in the front-end cap. The TOF was placed in the barrel region surrounding the ACC.

The electromagnetic calorimeter (ECL) enclosed all the previously mentioned sub-detectors. The task of the ECL was to detect photons, which are invisible for the tracking detectors, and to measure their energy and position. The ECL was also an important source of particle iden-

tification for electrons, which can be identified by comparing the momentum and the energy deposition of particles. Additionally, the ECL served to detect K_L^0 particles. Its structure was divided into three parts: a barrel, a front-end cap, and a back-end cap part. It consisted of 8736 thallium-doped cesium iodide (CsI(Tl)) crystals with a cross section of about $6 \times 6 \text{ cm}^2$, and a length of 30 cm, which corresponds to 16.2 radiation lengths. The crystals pointed towards a narrow region around the IP.

Surrounding the ECL, the superconducting solenoid was placed. It produced an homogeneous magnetic field of 1.5 T parallel to the z -direction. The magnetic field is needed to measure the momentum of charged particles. Charged particles experience a Lorentz force in the plane transverse to the field. The trajectory of the particles follows a helix, whose radius is proportional to the transverse momentum p_t . By measuring p_t and the polar angle θ , the total momentum of a particle can be obtained.

The magnetic flux returned through the iron return yoke, which served as absorber for the outermost detector, the detector to identify K_L^0 and muons (KLM). The KLM consisted of resistive plate counters interleaved in the plates of the iron return yoke. The KLM provided identification for muons, and for long-lived kaons. Dense material is needed to detect these particles, as their interaction probability with material is small. Charged kaons and muons are characterized by tracks in the CDC. Kaons produce hadronic showers in the ECL and in the iron plates of the KLM. On the contrary, muons do not produce showers, but they are expected to traverse all KLM layers. Neutral KLM clusters, i.e. those that are not associated with a track, identify the long-lived K_L^0 .

The Belle experiment had a trigger system and a global decision logic (GDL) to discriminate between the events that should be recorded and the background events caused for example by cosmic rays, synchrotron radiation, beam gas, and “non-interesting” physics processes such as Bhabha scattering. For hadronic events, the trigger system obtained an efficiency of more than 99.5%.

To be able to operate at much higher occupancies due to the increased luminosity at SuperKEKB, the Belle detector had to be substantially upgraded. Despite the harsh experimental conditions, the new Belle II detector is designed to reach a high overall physics performance by making use of novel sub-detectors.

Figure 3.3 shows an overview of the Belle II detector, as well as the Belle II coordinate system. The z -axis points roughly in the direction of the electron beam, the y -axis points upward, and the x -axis points radially away from the center of the storage ring. The origin is at the IP. The xy -plane is referred to as transverse plane. In spherical coordinates, the azimuthal angle ϕ is measured in the transverse plane from the x -axis; and the polar angle θ is measured from the z -axis. As its predecessor Belle, the Belle II detector is symmetric around the z -axis and, to account for the boost, it has an asymmetric acceptance covering the polar angles 17° to 150° .

To compensate the reduction in the boost, Belle II was designed to reach a more precise spatial resolution than Belle. For this, a new beam pipe with an outer radius of just 12 mm was

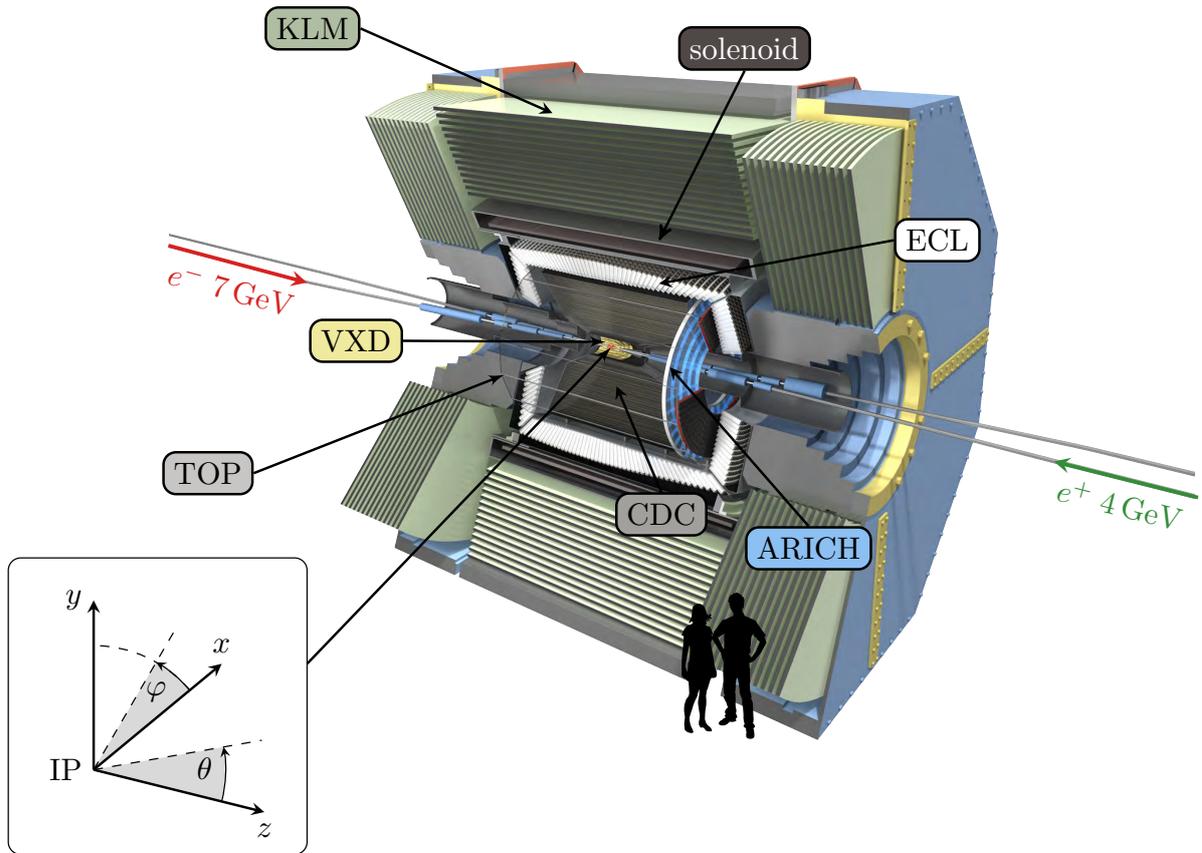


Figure 3.3: Overview of the Belle II detector with the vertex detector (VXD), the central drift chamber (CDC), the time of propagation counter (TOP), the Aerogel RICH detector (ARICH), the electromagnetic calorimeter (ECL), the solenoid, the K_L^0 and μ detector (KLM) and a schematic of the Belle II coordinate system. The origin corresponds to the interaction point (IP) [107].

installed; and a new vertex detector (VXD) was mounted directly on the beam pipe at a radius of only 14 mm from the IP. The total material thickness of beryllium in the beam pipe is the same as at Belle, but the shielding gold foil covers the inner wall, instead of the outer wall, and has a thickness of only $10\ \mu\text{m}$. Thinning the gold foil helps reduce multiple scattering of charged particles; and placing it in the inner surface helps improve the impact parameter resolution (the impact parameters will be introduced in Sect. 3.3).

Figure 3.4 shows a schematic overview of the VXD at Belle and at Belle II in the transverse plane. In the region where the first two VXD layers are mounted at Belle II, the background is very large and is dominated by an irreducible luminosity dependent component. Silicon strip detectors in this region would have unacceptable occupancies. Therefore, pixel detectors (PXD) were required for this first two layers. The four outer layers of the VXD employ silicon strip detectors (SVD) since they are mounted at a radial distance where the occupancy is tolerable.

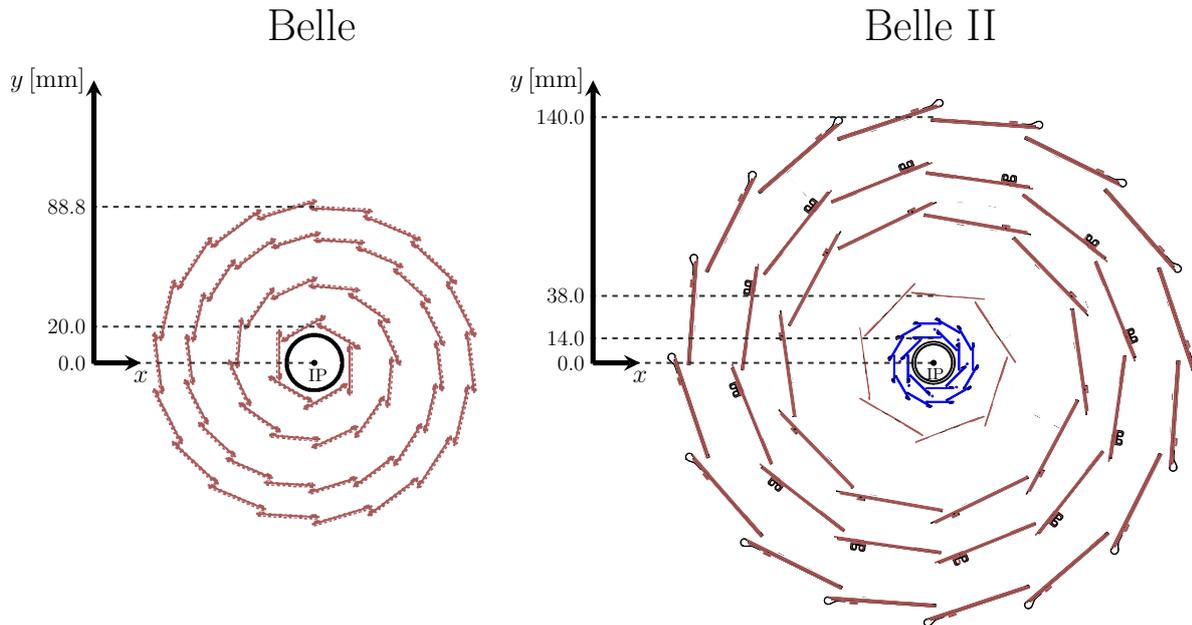


Figure 3.4: Front view of the Belle SVD2 (left) and of the Belle II VXD (right). The beam pipes are colored black, the novel pixel detectors are colored blue, and the strip detectors are colored guava. The Belle II illustration is courtesy of Karlheinz Ackermann.

The PXD is based on the novel DEpleted p -channel Field Effect Transistor (DEPFET) technology [108, 109]. It is designed to fulfill the physics requirements set by time-dependent CP-violation analyses and to cope with the anticipated large background. An advantage of this system is its small material budget which reduces the effect of multiple scattering by charged particles. Due to the internal signal amplification and the high signal-to-noise ratio, DEPFET sensors can be built very thin; the sensitive part has a thickness of $75\ \mu\text{m}$. Mechanical stability is provided by a narrow unthinned rim surrounding the sensor ladders. Furthermore, the read out electronics (except for a few control switchers) is mounted outside of the detector acceptance. A detailed hardware description of the Belle II PXD can be found in [110].

The Belle II SVD has about the same module thickness around $300\ \mu\text{m}$ as the Belle SVD. However, the Belle II SVD covers a larger radius and, to improve the spatial precision in boost direction, the forward sensors in the outer three layers are slanted towards the z -axis, as Figure 3.5 shows.

The CDC at Belle II is also completely new. It covers a larger radius and has 56 layers of sense wires, 6 more than the Belle CDC. Figure 3.6 shows a comparison of the layer configurations at Belle and at Belle II. The layers are arranged in so-called superlayers: at Belle II there are 9 superlayers, and at Belle there were 11. Within a superlayer, all layers have the same orientation. In contrast to Belle, at Belle II each superlayer contains the same number of layers, namely six. Only the innermost superlayer has two additional layers with smaller drift cells to cope with the higher background level.

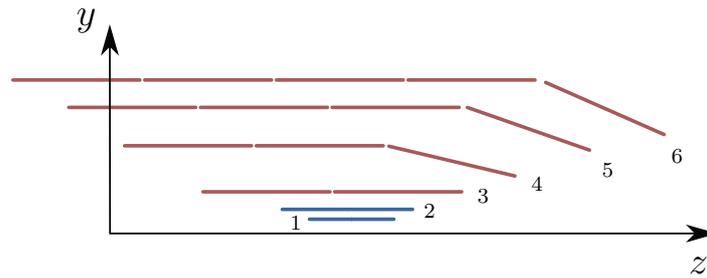


Figure 3.5: Alignment of the Belle II VXD layers around the beam pipe together with their corresponding numbering. The PXD layers are shown in blue and the SVD layers in guava [111].

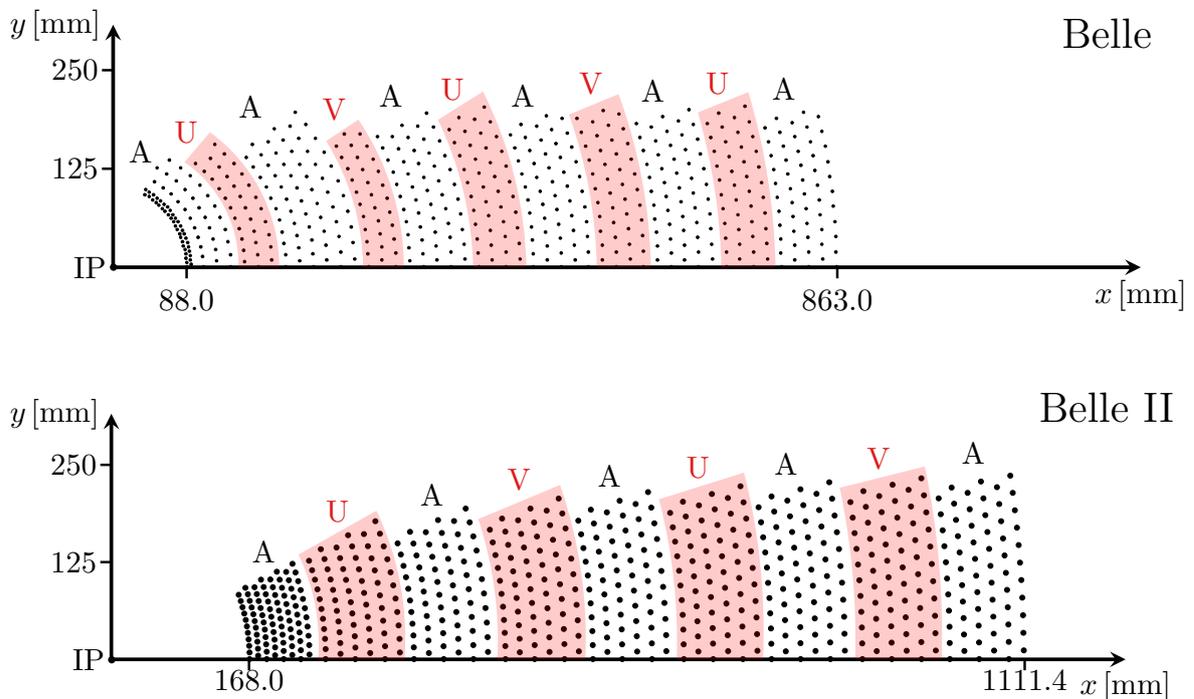


Figure 3.6: Layer configuration of the CDC at Belle (top) and at Belle II (bottom). The letter A stays for axial superlayers; the letter U for stereo superlayers with positive stereo angle; and the letter V for stereo superlayers with negative stereo angle.

To improve the discrimination power between charged particles, especially between kaons and pions, and to cope with the higher background level, the Cherenkov detectors of Belle were replaced by the new ARICH and TOP detectors. These new detectors have also a reduced material budget to improve the ECL response. The aerogel ring-imaging Cherenkov detector (ARICH) is installed in the front-end cap covering the polar range between 17° and 35° ; and the time of propagation counter (TOP) is installed in the barrel region on the outer wall of the CDC. It covers the polar angle range between 32° and 120° .

The Belle II experiment adopted the electromagnetic calorimeter of Belle. However, the readout system was upgraded to cope with the higher event rate. The ECL has a slightly larger acceptance than the tracking detectors covering the polar range between 12° and 155° .

The magnetic solenoid and the iron return yoke at Belle II were taken also from Belle. The resistive plate counters of the KLM in the barrel region were adopted from the Belle detector as well. To cope with the higher background rate in the front-end and in the back-end parts of the KLM, the resistive plate counters in these regions were replaced with scintillators.

The trigger system at Belle II is adopted from Belle. However, the trigger system has been upgraded and made more flexible to accommodate novel subtrigger systems such as the z -vertex and the single-photon triggers [1, 112] in order to cope with the increased event rate, and to search for new physics in low-multiplicity sectors that remained uncovered at Belle.

3.3 Track reconstruction

Track reconstruction is essential to reconstruct charged particles originating from primary and secondary vertices and to distinguish them against background. A precise track reconstruction is essential also to reconstruct the decay vertices and to align the detector. Precise vertex reconstruction and precise detector alignment are crucial to perform time-dependent CP violation analyses and to reduce their systematic uncertainties.

At Belle II, the track reconstruction consists of two steps, *track finding* and *track fitting*. Track finding consists in recognizing the patterns of VXD and CDC hits that belong to the track of a given charged particle. The trajectory of the charged particle is obtained then in the track fitting step from a fit to the hit positions. The track reconstruction in Belle II Analysis Software Framework (BASF2) is continuously under development. It is based on the GENFIT tracking framework [113], an optimized Kalman filter [114] and novel multivariate methods [115]. A detailed description can be found in [1].

The trajectory of a charged particle propagating in a constant homogenous magnetic field follows a helix described by five parameters at a point P of the trajectory. At Belle II, the point P corresponds to the point of closest approach to the IP. The point of closest approach is determined by extrapolating the track to the global z -axis. The five track parameters in the Belle II tracking software are the following:

- d_0 : the signed distance from the IP to the point of closest approach in the transverse plane. The sign convention is defined as follows: moving along the track into the direction of the particle's momentum, the sign is positive (negative) if the origin is to the right (left) of the track at the point of closest approach.
- z_0 : the longitudinal signed distance from the IP to the point of closest approach.
- ϕ_0 : the ϕ angle at the point of closest approach.
- $\tan \lambda$: the tangent of the angle between the momentum at the point of closest approach and the transverse plane.

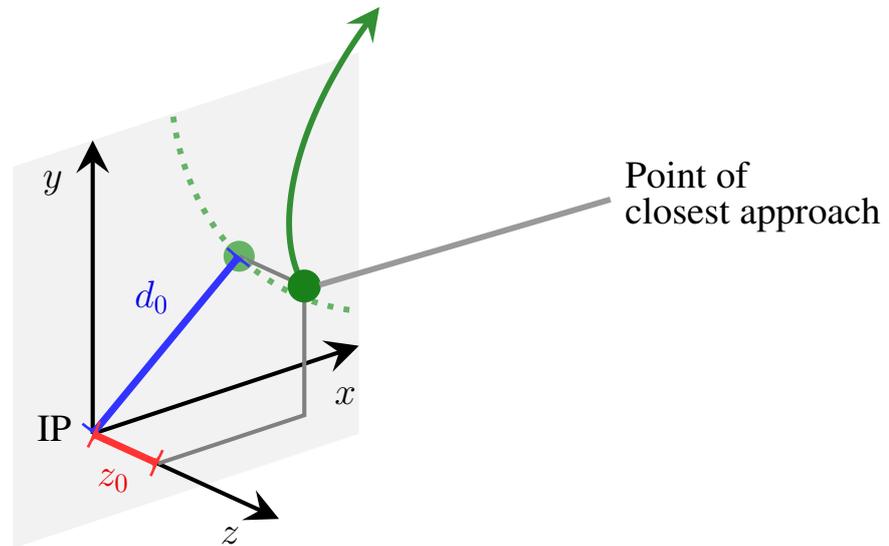


Figure 3.7: The solid green line represents the track of a particle, and the dashed green line its projection onto the xy plane. The impact parameters are defined as follows: d_0 is the signed distance between the origin and the point of closest approach on the xy plane; z_0 is the z coordinate of the track at the point of closest approach.

- ω : the curvature, whose sign corresponds to the charge of the track.

Figure 3.7 illustrates the definition of the impact parameters d_0 and z_0 at Belle II. To characterize the precision of the track reconstruction, the impact parameter resolutions σ^{d_0} and σ^{z_0} are some of the most important quantities. They are given by [116]:

$$\sigma^{d_0} = \sqrt{a^2 + \frac{b^2}{p\beta \sin(\theta)^{3/2}}}, \quad \sigma^{z_0} = \sqrt{a^2 + \frac{b^2}{p\beta \sin(\theta)^{5/2}}},$$

where a is the intrinsic detector resolution, b is the multiple scattering coefficient, and $p\beta \sin(\theta)^{3/2}$ and $p\beta \sin(\theta)^{5/2}$ are so-called pseudo-momenta chosen to take into account the effect of multiple scattering. Figure 3.8 shows the impact parameter resolutions expected at Belle II. The results for Belle II MC events with a single muon track are compared with the results for Belle cosmic events [117]. In comparison with Belle, Belle II reaches an improvement in the resolution on both impact parameters by almost a factor two.

In general, the trajectories of charged particles are not ideal helices because the particles interact with the detector material losing a fraction of their energy and undergoing multiple scattering. Additionally, the solenoid field is not perfectly homogenous. The Belle II track reconstruction algorithms take these effects into account. To correctly treat the interaction of particles with matter, the track candidates are fitted with different mass hypotheses depending

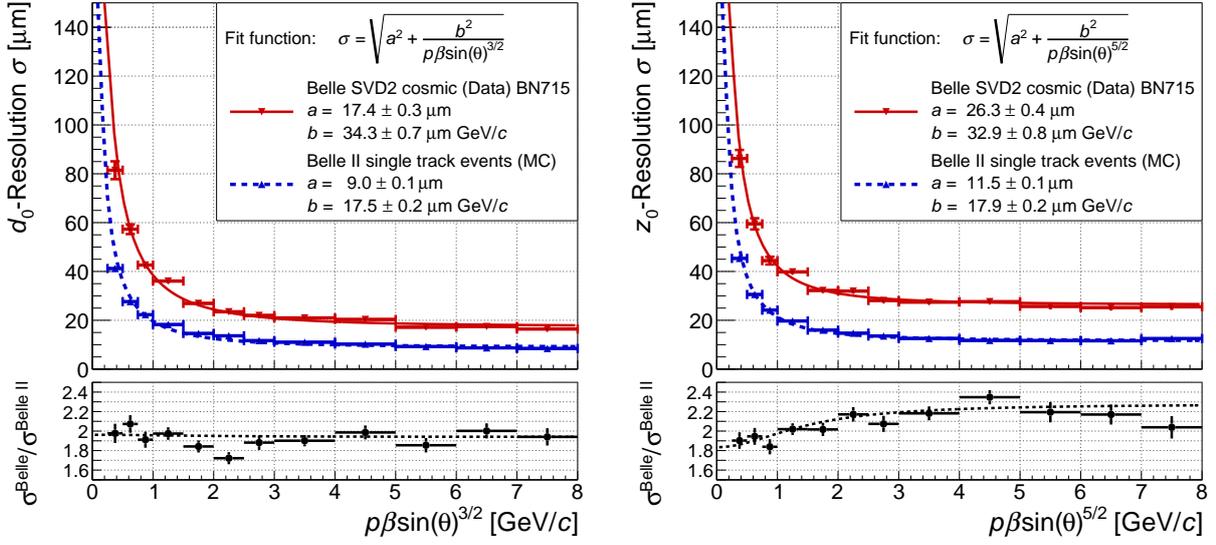


Figure 3.8: Resolution of the transverse d_0 (left) and longitudinal z_0 (right) impact parameters. The results for MC events with a single muon track using the Belle II tracking algorithm are compared with the results for Belle cosmic events [117]. The resolution in each bin is estimated using the σ value of a single Gaussian function fitted in a region containing 90% of the data around the mean value of the distributions.

on their momenta. For example, a pion and a kaon with the same momentum have similar interactions with matter and therefore a single hypothesis is sufficient. While an electron loses more energy due to bremsstrahlung than a pion with the same momentum.

By default, every track is fitted with the pion mass hypothesis. The five supported hypotheses are electron, muon, kaon, pion and proton. When particle candidates are built at the analysis level, if the track fit result for the corresponding mass hypothesis is available, then this result is used to form the particle candidate. Otherwise the result for the closest available mass hypothesis is taken.

3.4 Vertex Reconstruction

The Belle II experiment employs two different implementations of the vertex reconstruction: the Kinematic Fit (KFit) library developed for the Belle experiment [118], and the Reconstruction in an Abstract Vertices Environment (RAVE) [119], a standalone package of the CMS vertexing libraries.

At Belle II, the benchmark B -decay mode to validate the vertex reconstruction algorithms is the golden channel $B_{\text{sig}}^0 \rightarrow J/\psi K_S^0$, where the B_{sig}^0 vertex can be reconstructed with high precision using the muon tracks produced in the J/ψ decay. Figure 3.9 (left) shows the fit residuals for the J/ψ vertex. The residuals are the values obtained by subtracting the generated MC vertex from the reconstructed vertex. Here, $l_{\text{sig}} = \mathbf{x}_{\text{sig}} \cdot \hat{\beta}$, where \mathbf{x}_{sig} is the three-dimensional J/ψ vertex in the lab frame, and $\hat{\beta}$ is a unit vector in boost direction. Each muon track is

required to have at least one associated PXD hit. The resolution of $24\ \mu\text{m}$ corresponds to an improvement in the resolution by about a factor 2 with respect to the Belle MC [120], and is consistent with the improvement in the resolution of the impact parameters shown in Figure 3.8.

In measurements of time-dependent CP asymmetries, the precision is strongly correlated with the Δt resolution. The time difference Δt is calculated from the difference Δl in boost direction between the reconstructed vertices of B_{sig}^0 and B_{tag}^0 (eq. (2.34)). To exploit the expected large data sample, the major challenge at Belle II consists in maintaining a Δt resolution which should be comparable with that of Belle or better. For the reconstruction of the B_{tag}^0 -vertex, Belle II has a novel algorithm, which will be introduced in Sect. 6.6.2. This algorithm obtains also a considerable improvement with respect to Belle.

For fully reconstructed $B_{\text{sig}}^0 \rightarrow J/\psi(\rightarrow \mu^+\mu^-) K_S^0(\rightarrow \pi^+\pi^-)$ decays, Fig. 3.9 (right) shows the Δt residuals. The resolution (81 ps) and the bias (0 ps) represent an improvement with respect to Belle [120], despite the reduction in the boost. The MC events used to produce Figure 3.9 correspond to the latest official MC validation sample used for the studies in chapter 5.

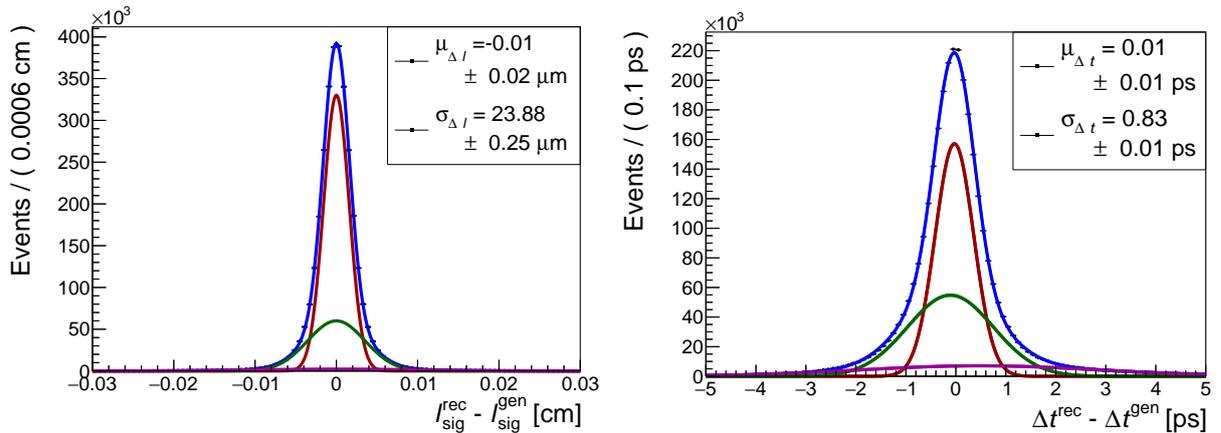


Figure 3.9: Left: $J/\psi \rightarrow \mu^+\mu^-$ vertex fit residuals. Right: Δt resolution for $B_{\text{sig}}^0 \rightarrow J/\psi K_S^0$. Both fits are performed using the sum of three Gaussian functions. The shift μ and the resolution σ given in the figures correspond to the weighted averages of the mean values and the standard deviations of the three Gaussian functions.

3.5 Particle identification

A good particle identification is crucial for distinguishing between final state hadrons and leptons, and especially for flavor tagging, as it will be shown in chapter 4. Belle II has an upgraded PID system consisting of the TOP and the ARICH sub-detectors. The information from these sub-detectors is combined with measurements of dE/dx obtained from the SVD and the CDC to provide the primary sources for charged hadron identification. For electron

identification, the ECL provides the primary source of identification. And for muons, the KLM provides the primary source.

At Belle II, the charged particle identification relies on likelihood based selectors. The information from each PID system is analyzed independently to determine a likelihood for each charged particle hypothesis. These likelihoods can be used individually, or to construct a combined likelihood ratio.

The likelihood selectors rely on likelihood ratios constructed in the following way. First, the PID log likelihoods from each sub-detector are summed to create a combined PID likelihood for each of six long-lived charged particle hypotheses: electron, muon, pion, kaon, proton and deuteron. Next, the difference in log likelihood between two particle hypotheses is used to construct a PID value \mathcal{L}_α according to

$$\mathcal{L}_\alpha = \mathcal{L}_{\alpha/\beta} = \mathcal{L}(\alpha : \beta) = \frac{\prod_i \mathcal{L}(\alpha)}{\prod_i \mathcal{L}(\alpha) + \prod_i \mathcal{L}(\beta)}, \quad (3.1)$$

where α and β represent two different particle types, and i extends over the active sub-detectors for the PID type of interest. The value $\mathcal{L}(\alpha : \beta)$ is greater than 0.5 for a charged track that more closely resembles a particle of type α than one of type β , and is less than 0.5 otherwise. Sect. 3.5 presents the standard combinations of α and β . For a comprehensive description of the PID algorithms at Belle II see [1].

	\mathcal{L}_e	\mathcal{L}_μ	\mathcal{L}_K	\mathcal{L}_π	\mathcal{L}_p
α	e	μ	K	π	p
β	π	π	π	K	π

Table 3.2: Standard combinations of α and β for the PID likelihoods defined in eq. (3.1).

4 The Belle II flavor tagger

4.1 Introduction

At B factories, B mesons pairs are produced at the $\Upsilon(4S)$ resonance exclusively, i.e. without additional particles, and the fraction of events with multiple $e^+ e^-$ interactions (“pile up”) is negligible. When the decay of one of the B mesons is fully reconstructed (signal side), the remaining reconstructed tracks and neutral ECL and KLM clusters in the event necessarily belong to the decay of the accompanying B meson (tag side). Most measurements of CP -violation and of B -meson mixing require the full reconstruction of a neutral B meson (B_{sig}^0), and the determination of the flavor of the accompanying B^0 meson (B_{tag}^0) at the time of its decay. The determination of the B_{tag}^0 flavor is referred to as flavor tagging and is accomplished using so-called flavor tagging algorithms, or shortly flavor taggers.

B mesons have a myriad of possible decay channels. Many of these decays provide flavor signatures through flavor-specific final states. Flavor signatures correspond to characteristics of the decay products that are correlated with the electric charge sign of the b -quark in the B^0 meson. Because of the wide range of possible decay channels, it is unfeasible to fully reconstruct a large fraction of flavor-specific B_{tag}^0 decays. Instead of a full reconstruction, flavor tagging algorithms apply inclusive techniques to maximally exploit the information provided by the different flavor signatures in flavor-specific decays.

The predecessors of Belle II, the Belle and the BaBar experiments, developed two different flavor taggers [29]. Due to the novel high-luminosity conditions and the increased beam backgrounds at Belle II, a new flavor tagger had to be developed with a specific and dedicated optimization in order to cope with the harsher experimental conditions. Within this thesis’ work, a new flavor tagger was developed for Belle II. Since the previous algorithms at Belle and at BaBar relied on the physics of B meson decays, several useful concepts of these algorithms were adopted for the new Belle II algorithm. The major adopted concept was the sorting of different flavor signatures into different tagging categories.

The algorithm developed at BaBar was based on multivariate methods and reached a better performance than the one at Belle. Therefore, the new category-based Belle II algorithm was developed using multivariate methods as well, exploiting also the capabilities of the new Fast Boosted Decision Tree (FBDT) developed especially for the Belle II collaboration [115].

The development of the new flavor tagger started almost four years ago. After the first year, a first version of the algorithm was released in collaboration with Moritz Gelb [121]. Afterwards, the algorithm was further developed and improved to consider more flavor signa-

tures than the predecessor algorithms and to exploit the improving reconstruction capabilities at Belle II. About two years ago, a second flavor tagging algorithm started to be developed at Belle II [122]. This second algorithm takes advantage of deep-learning multivariate methods [123].

In the category-based approach, the algorithm identifies the decay products that stem from flavor-specific decays and are correlated with the B_{tag}^0 flavor. For this, the algorithm employs the PID information and the kinematic properties of the tag-side tracks, as well as global information provided by all the tag-side tracks and clusters. The characteristics of the identified decay products are then used to determine the flavor of the B_{tag}^0 meson. In the deep-learning algorithm, the B_{tag}^0 flavor is determined using the decay products without pre-identification, employing only the PID information and the kinematic properties of the tag-side tracks. A deep neural network is assumed to learn the flavor signatures present in flavor-specific decays.

This chapter is organized as follows. First, the quantities characterizing the performance of a flavor tagger are defined. The sources of flavor information and the corresponding discriminating input variables are explained in the next sections. Subsequently, the category-based flavor tagging algorithm created in this thesis is introduced. The key steps for its optimization and the evaluation of its performance are covered afterwards. The performance is evaluated using Belle II MC without simulated background, and using Belle MC with simulated background. The evaluation included studies on new possible effects on the performance of the flavor tagger caused by the amount of CP violation.

The category-based flavor tagger was comprehensively validated using Belle data. The validation included a test of robustness in which the amount of input information was systematically varied using Belle data. The sections presenting these results are followed by a discussion about the systematic effect caused by the tag-side interference, and the possible ways to mitigate it. Towards the end of this chapter, the category-based flavor tagger is compared with the alternative deep-learning flavor tagger [122], and with the previous algorithms at Belle and at BaBar. Finally, the plans for calibration of the flavor taggers at Belle II are discussed.

In the last stage of this thesis, new official MC samples with simulated background became available. The results of the category-based algorithm using these new samples are presented in chapter 5.

4.2 Definitions

Given a total number of events N , the efficiency ε is defined as the fraction of events to which the flavor tagging algorithm can assign a flavor tag, i.e.

$$\varepsilon = \frac{N^{\text{tag}}}{N}$$

where N^{tag} is the number of tagged events. The fraction of wrong identifications over the number of tagged events is denoted by w . Thus, the number of tagged B and \bar{B} events is given by

$$\begin{aligned} N_{B^0}^{\text{tag}} &= \varepsilon(1-w)N_{B^0} + \varepsilon w N_{\bar{B}^0} \\ N_{\bar{B}^0}^{\text{tag}} &= \varepsilon(1-w)N_{\bar{B}^0} + \varepsilon w N_{B^0}, \end{aligned}$$

where N_{B^0} and $N_{\bar{B}^0}$ are the true number of B^0 and \bar{B}^0 mesons on the tag side. The asymmetry observed in CP -violation analysis is then

$$a_{CP}^{\text{obs}} = \frac{N_{B^0}^{\text{tag}} - N_{\bar{B}^0}^{\text{tag}}}{N_{B^0}^{\text{tag}} + N_{\bar{B}^0}^{\text{tag}}} = (1-2w) \cdot \frac{N_{B^0} - N_{\bar{B}^0}}{N_{B^0} + N_{\bar{B}^0}} = (1-2w) \cdot a_{CP},$$

where a_{CP} corresponds to the CP asymmetry in CP analyses, i.e. to eq. (2.32) for time-dependent measurements or to eq. (2.36) for time-integrated (see Sect. 2.7.3). Thus, in order to minimize systematic uncertainties, the value of w has to be precisely measured. The strength of the observed CP asymmetry is proportional to $|1-2w|$, i.e. the CP asymmetry becomes “diluted” because of the wrong-tag fraction. The so-called dilution factor is defined as

$$r \equiv |1-2w|, \quad (4.1)$$

where $r = 0$ means no flavor information ($w = 0.5$) and $r = 1$ corresponds to an unambiguous tag ($w = 0, 1$). The statistical uncertainty of a_{CP} is

$$\delta a_{CP} = \frac{\delta a_{CP}^{\text{obs}}}{1-2w}.$$

Assuming that a_{CP}^{obs} is small, i.e. $N_{B^0}^{\text{tag}} \approx N_{\bar{B}^0}^{\text{tag}}$, one obtains for the statistical uncertainty of a_{CP}^{obs}

$$\delta a_{CP}^{\text{obs}} \stackrel{N_{B^0}^{\text{tag}} \approx N_{\bar{B}^0}^{\text{tag}}}{=} \frac{1}{\sqrt{N^{\text{tag}}}}.$$

Thus, one finds that

$$\delta a_{CP} = \frac{1}{\sqrt{N^{\text{tag}}(1-2w)}}. \quad (4.2)$$

The effective tagging efficiency ε_{eff} of a flavor tagging algorithm is defined such that the statistical uncertainty on the measured asymmetry a_{CP} is related to the effective number of tagged events N^{eff} by

$$\delta a_{CP} = \frac{1}{\sqrt{N^{\text{eff}}}} = \frac{1}{\sqrt{\varepsilon_{\text{eff}} \cdot N}}. \quad (4.3)$$

So, the statistical uncertainty on a_{CP} would be the same if one would have N^{eff} perfectly tagged events instead of N events tagged with an effective efficiency ε_{eff} . Comparing eq. (4.2) with eq. (4.3), one obtains

$$\varepsilon_{\text{eff}} = \frac{N^{\text{tag}}}{N} \cdot (1 - 2w)^2 = \varepsilon \cdot r^2. \quad (4.4)$$

A maximization of the effective efficiency results in a minimization of the statistical uncertainty. Therefore, CP -violation analyses are performed usually in bins of the dilution factor r : depending on the choice of the binning, the effective efficiency can increase or decrease since the flavor discrimination power is minimal in the range where r is close to zero, and maximal in the range where r is close to 1. When r binning is used, the efficiency in eq. (4.4) takes a binned form which will be introduced in Sect. 4.5.2. The r binning is required for the calibration of the flavor tagger, as it will be explained in Sect. 4.14.

In general, the scaling of δa_{CP} with ε_{eff} (eq. (4.3)) is only approximate. For a likelihood-based analysis, the expected statistical uncertainty of an estimated CP or mixing asymmetry can be obtained from the maximum-likelihood estimator. An example is presented in Sect. 6.10.6.

Up to now, w and ε have been considered to be equal for $q = +1(-1)$. However, a slight difference can arise as a result of a charge-asymmetric detector performance. To take this effect into account, one redefines

$$\varepsilon = \frac{\varepsilon_{B^0} + \varepsilon_{\bar{B}^0}}{2}, \quad w = \frac{w_{B^0} + w_{\bar{B}^0}}{2}, \quad (4.5)$$

and introduces the differences

$$\Delta\varepsilon = \varepsilon_{B^0} - \varepsilon_{\bar{B}^0}, \quad \Delta w = w_{B^0} - w_{\bar{B}^0}, \quad (4.6)$$

where the index corresponds to the true flavor, e.g. w_{B^0} is the fraction of true B^0 mesons that were wrongly classified as \bar{B}^0 . In terms of these quantities, the Δt probability $\mathcal{P}(\Delta t, q)$ for time-dependent CP -violation analyses given by eq. (2.33) in Sect. 2.7.3 becomes

$$\begin{aligned} \mathcal{P}^{\text{obs}}(\Delta t, q) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \varepsilon \left[1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) \right. \\ \left. + [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot a_{CP}(\Delta t) \right], \quad (4.7) \end{aligned}$$

where $\mu = \Delta\varepsilon/(2\varepsilon)$ and $a_{CP}(\Delta t)$ corresponds to eq. (2.32). For time-integrated CP -violation analyses, the probability $\mathcal{P}(q)$ in eq. (2.35) becomes

$$\mathcal{P}^{\text{obs}}(q) = \frac{1}{2}\varepsilon \left[1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) + [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot \mathcal{A}_{CP}(1 - 2 \cdot \chi_d) \right]. \quad (4.8)$$

A detailed derivation of eqs. (4.7) and (4.8) is presented in App. B.1. In the same way, the Δt probability for $B^0 - \bar{B}^0$ mixing in eq. (2.40) of Sect. 2.7.4 becomes:

$$\mathcal{P}(\Delta t, q_{\text{sig}}, q) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \varepsilon \left[1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) - [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot q_{\text{sl}} \cdot \cos(\Delta m \Delta t) \right], \quad (4.9)$$

where q_{sig} is the flavor of the signal B^0 decaying into a flavor-specific mode, and q is the flavor of the accompanying B_{tag}^0 determined by the flavor tagger. For time integrated measurements one obtains

$$\mathcal{P}(q_{\text{sig}}, q) = \frac{1}{2}\varepsilon \left[1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) - [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot q_{\text{sl}} \cdot (1 - 2 \cdot \chi_d) \right]. \quad (4.10)$$

In eqs. (4.9) and (4.10) it is assumed that the flavor q_{sl} of B_{sl}^0 is always determined correctly.

4.3 Tagging categories

The category-based flavor tagger relies on flavor-specific decay modes with relatively high branching fractions ($\gtrsim 2\%$). Each decay mode has a particular decay topology and provides a flavor-specific signature. Some additional signatures are obtained by combining similar or complementary decay modes. Within a category, a particular flavor signature is considered separately. The final flavor tagger is based on 13 categories which are presented in Tab. 4.1.

The decay modes are characterized by flavor-specific final state particles. These particles are treated as targets since their charges are correlated with the flavor of B_{tag}^0 . To extract these flavor-specific signatures, the targets have to be identified among all available particle candidates. To accomplish this task, discriminating input variables are calculated for each particle candidate. An overview of the input variables for each category is presented in Tab. 4.2. Except for the Maximum- p^* category, particle identification (PID) variables are used for all the categories. A description of the particle identification at Belle II and the calculation of the PID variables is provided in Sect. 3.5.

Table 4.1: Tagging categories and their targets (left) with some characteristic examples of the considered decay modes (right). Here, p^* stays for momentum in the centre-of-mass frame and ℓ^\pm for charged leptons (μ^- or e^-).

Categories	Targets for \overline{B}^0	Underlying decay modes
Electron	e^-	
Intermediate Electron	e^+	$\overline{B}^0 \rightarrow D^{*+} \bar{\nu}_\ell \ell^-$
Muon	μ^-	$\hookrightarrow D^0 \pi^+$
Intermediate Muon	μ^+	$\hookrightarrow X K^-$
Kinetic Lepton	ℓ^-	
Intermediate Kinetic Lepton	ℓ^+	$\overline{B}^0 \rightarrow D^+ \pi^- (K^-)$
Kaon	K^-	$\hookrightarrow K^0 \nu_\ell \ell^+$
Kaon-Pion	K^-, π^+	
Slow Pion	π^+	
Maximum p^*	ℓ^-, π^-	$\overline{B}^0 \rightarrow \Lambda_c^+ X^-$
Fast-Slow-Correlated (FSC)	ℓ^-, π^+	$\hookrightarrow \Lambda \pi^+$
Fast Hadron	π^+, K^-	$\hookrightarrow p \pi^-$
Lambda	Λ	

In the following, the flavor signatures and the input variables for each category are explained in detail. The distributions of the input variables and their correlations on MC are shown for some characteristic examples. For each category, the particle candidates corresponding to the targets will be referred to as *signal*, and all other candidates as *background*. For a better visibility, the distributions of the input variables are normalized such that the areas under the signal and under the background distributions equal unity. The correlations R_{ij} between two variables x_i and x_j correspond to

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}}, \quad (4.11)$$

where C is the covariance matrix

$$C_{ij} = \langle (x_i - \mu_i)(x_j - \mu_j) \rangle = \langle x_i x_j \rangle - \mu_i \mu_j,$$

and μ_i , and μ_j are the mean values $\langle x_i \rangle$, and $\langle x_j \rangle$.

The distributions and the correlations presented in the following subsections are obtained from the official Belle II MC used during the development of the flavor tagger. The distributions obtained from Belle MC will be shown in some cases where there are considerable differences. When Belle MC is used, this is indicated explicitly. The MC samples correspond to $B^0 \overline{B}^0$ events where B_{sig}^0 is fully reconstructed and correctly matched with the MC decay

chain. In these events, B_{tag}^0 decays to any possible final state according to the known branching fractions [28]. Further details about the MC samples will be given in Sect. 4.5. For all the input variables, the respective distributions are presented in App. B.3 together with detailed illustrations of their ranking and their correlations.

Table 4.2: Discriminating input variables for each category. For some of the categories the p -value of the track fit is taken into account. For the Lambda category, the p -value of the reconstructed Λ decay vertex is used. All variables are calculated for each considered particle candidate.

Categories	Discriminating input variables
Electron	$\mathcal{L}_e, \mathcal{L}_e^{\text{dE/dx}}, \mathcal{L}_e^{\text{TOP}}, \mathcal{L}_e^{\text{ARICH}}, \mathcal{L}_e^{\text{ECL}}, p^*, p_t^*, p, p_t, \cos \theta, d_0, \xi_0,$
Int. Electron	$M_{\text{rec}}^2, E_{90}^W, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*, \cos \theta_{\text{T}}^* , p\text{-val.}$
Muon	$\mathcal{L}_\mu, \mathcal{L}_\mu^{\text{dE/dx}}, \mathcal{L}_\mu^{\text{TOP}}, \mathcal{L}_\mu^{\text{ARICH}}, \mathcal{L}_\mu^{\text{KLM}}, p^*, p_t^*, p, p_t, \cos \theta, d_0, \xi_0,$
Int. Muon	$M_{\text{rec}}^2, E_{90}^W, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*, \cos \theta_{\text{T}}^* , p\text{-val.}$
Kin. Lepton	$\mathcal{L}_e, \mathcal{L}_e^{\text{dE/dx}}, \mathcal{L}_e^{\text{TOP}}, \mathcal{L}_e^{\text{ARICH}}, \mathcal{L}_e^{\text{ECL}}, \mathcal{L}_\mu, \mathcal{L}_\mu^{\text{dE/dx}}, \mathcal{L}_\mu^{\text{TOP}}, \mathcal{L}_\mu^{\text{ARICH}}, \mathcal{L}_\mu^{\text{KLM}},$
Int. Kin. Lep.	$p^*, p_t^*, p, p_t, \cos \theta, d_0, \xi_0, M_{\text{rec}}^2, E_{90}^W, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*, \cos \theta_{\text{T}}^* , p\text{-val.}$
Kaon	$\mathcal{L}_K, \mathcal{L}_K^{\text{dE/dx}}, \mathcal{L}_K^{\text{TOP}}, \mathcal{L}_K^{\text{ARICH}}, p^*, p_t^*, p_t, \cos \theta, d_0, \xi_0,$ $n_{K_S^0}, \sum p_t^2, M_{\text{rec}}^2, E_{90}^W, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*, \cos \theta_{\text{T}}^* , p\text{-val.}$
Slow Pion	$\mathcal{L}_\pi, \mathcal{L}_\pi^{\text{dE/dx}}, \mathcal{L}_\pi^{\text{TOP}}, \mathcal{L}_\pi^{\text{ARICH}}, \mathcal{L}_{\pi/e}^{\text{dE/dx}}, \mathcal{L}_e, \mathcal{L}_K, \mathcal{L}_K^{\text{dE/dx}}, \mathcal{L}_K^{\text{TOP}}, \mathcal{L}_K^{\text{ARICH}},$
Fast Hadron	$p^*, p_t^*, p, p_t, \cos \theta, d_0, \xi_0, n_{K_S^0}, \sum p_t^2, M_{\text{rec}}^2, E_{90}^W, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*, \cos \theta_{\text{T}}^* , p\text{-val.}$
Kaon-Pion	$\mathcal{L}_K, y_{\text{Kaon}}, y_{\text{SlowPion}}, \cos \theta_{K\pi}^*, q_K \cdot q_\pi$
Maximum p^*	$p^*, p_t^*, p, p_t, d_0, \xi_0, \cos \theta_{\text{T}}^* $
FSC	$\mathcal{L}_{K\text{Slow}}, p_{\text{Slow}}^*, p_{\text{Fast}}^*, \cos \theta_{\text{T, Slow}}^* , \cos \theta_{\text{T, Fast}}^* , \cos \theta_{\text{SlowFast}}^*, q_{\text{Slow}} \cdot q_{\text{Fast}}$
Lambda	$\mathcal{L}_p, \mathcal{L}_\pi, p_\Lambda^*, p_\Lambda, p_\pi^*, p_\pi, p_\pi^*, p_\pi, q_\Lambda, M_\Lambda, n_{K_S^0}, \cos \theta_{\mathbf{x}_\Lambda, \mathbf{p}_\Lambda}, \mathbf{x}_\Lambda , \sigma_\Lambda^{zz}, p\text{-val.}$

Leptons

Primary and secondary leptons from B^0 decays are used as target particles for different categories. In the first case, the leptons stem from $\bar{B}^0 \rightarrow X^+ \ell^- \bar{\nu}_\ell$ decays occurring via transitions $b \rightarrow c \ell^- \bar{\nu}_\ell$, or $b \rightarrow u \ell^- \bar{\nu}_\ell$. A negatively charged primary lepton unambiguously tags a \bar{B}^0 meson, and a positively charged one a B^0 meson. Primary electrons and muons are the targets of the *Electron* and the *Muon* categories. Both are considered as targets in the *Kinetic Lepton* category.

Secondary leptons that are produced through the decay of charmed mesons and baryons of the kind $\bar{B}^0 \rightarrow X_c (\rightarrow \ell^+ \nu_\ell X_{s(d)}) X$, occurring via transitions $b \rightarrow c \rightarrow s (d) \ell^+ \nu_\ell$, tag as well the flavor of the B^0 meson. In this case the charge-flavor correspondence is reversed,

i.e. a positively charged secondary lepton tags a \bar{B}^0 meson, and a negatively charged one a B^0 meson. Since their momentum spectrum is much softer in comparison with the primary leptons, secondary leptons are referred to as intermediate leptons. Intermediate electrons and intermediate muons are the targets of the *Intermediate Electron* and the *Intermediate Muon* categories. Both are considered as targets in the *Intermediate Kinetic Lepton* category.

In order to distinguish primary and secondary leptons from all other candidates, kinematic and PID variables, are used as input variables. The PID variables have high discrimination power. For electrons, the PID variables are the individual likelihoods $\mathcal{L}_e^{dE/dx}$, $\mathcal{L}_e^{\text{TOP}}$, $\mathcal{L}_e^{\text{ARICH}}$ and $\mathcal{L}_e^{\text{ECL}}$, and the combined \mathcal{L}_e . For muons, the PID variables are the individual likelihoods $\mathcal{L}_\mu^{dE/dx}$, $\mathcal{L}_\mu^{\text{TOP}}$, $\mathcal{L}_\mu^{\text{ARICH}}$ and $\mathcal{L}_\mu^{\text{KLM}}$, and the combined \mathcal{L}_μ . The combined likelihoods have in general the highest discrimination power within the PID variables. However, the combined and the individual likelihoods together have a larger discrimination power. Figure 4.1 shows the individual likelihoods with the lowest and the highest discrimination power as well as the combined ratios for the Electron and the Muon categories.

Within the kinematic variables, the momentum variables, such as the absolute momentum p^* and the transverse momentum p_t^* in the $\Upsilon(4S)$ centre-of-mass frame, as well as the absolute momentum p and the transverse momentum p_t in the laboratory frame, have the highest dis-

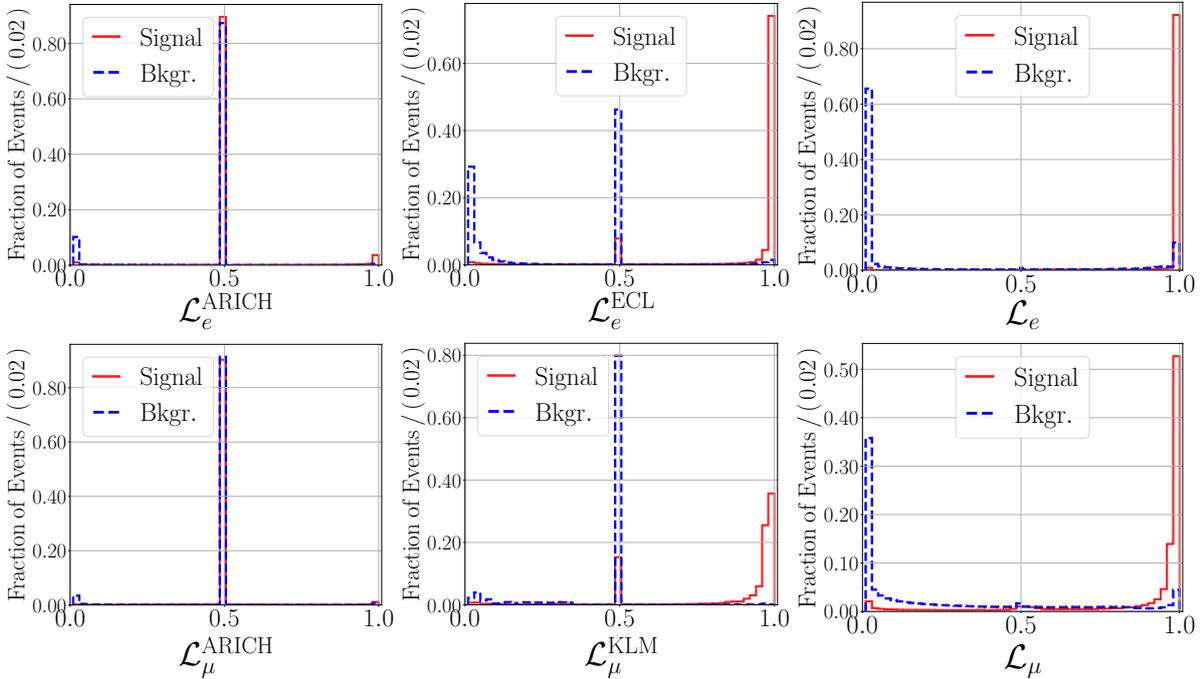


Figure 4.1: PID likelihoods for the Electron (top) and the Muon categories (bottom). Individual likelihoods with the lowest (left) and the highest (center) discrimination power are shown together with the combined ratios (right). A value of 0.5 is assigned to tracks without available PID information.

crimination power, especially for primary leptons. Intermediate leptons are more difficult to distinguish from other candidates because of their softer momentum spectrum. Additionally, the kinematic variables include $\cos\theta$, the cosine of the polar angle of the momentum in the laboratory frame.

Primary leptons are produced at the B_{tag}^0 decay vertex and have thus small impact parameters. In comparison, intermediate leptons have larger ones. The impact parameter corresponds to the distance between the track's point of closest approach and the interaction point (IP) (see Sect. 3.3): the impact parameter in the xy -plane is d_0 , and in three dimensions is $\xi_0 = \sqrt{d_0^2 + z_0^2}$. The use of impact parameters enhances the discrimination power of the flavor tagger. At Belle II, however, the use of impact parameters causes asymmetries in the performance of the flavor tagger giving rise to small systematic effects, which will be discussed in Sect. 4.6 and 4.7.

Figure 4.2 shows the distributions of p^* , $\cos\theta$ and ξ_0 for the Kinetic Lepton and the Intermediate Kinetic Lepton categories. In the $\cos\theta$ distributions, one can recognize small dips around $\cos\theta = 0$, i.e. around $\theta \approx 90^\circ$. Around this angle, soft momentum tracks curl multiple times through the tracking subdetectors¹, potentially crossing the same detector cells or pixels

¹A charged particle needs at least a transverse momentum of about $0.25 \text{ GeV}/c$ to traverse all tracking subdetectors at Belle II (VXD and CDC).

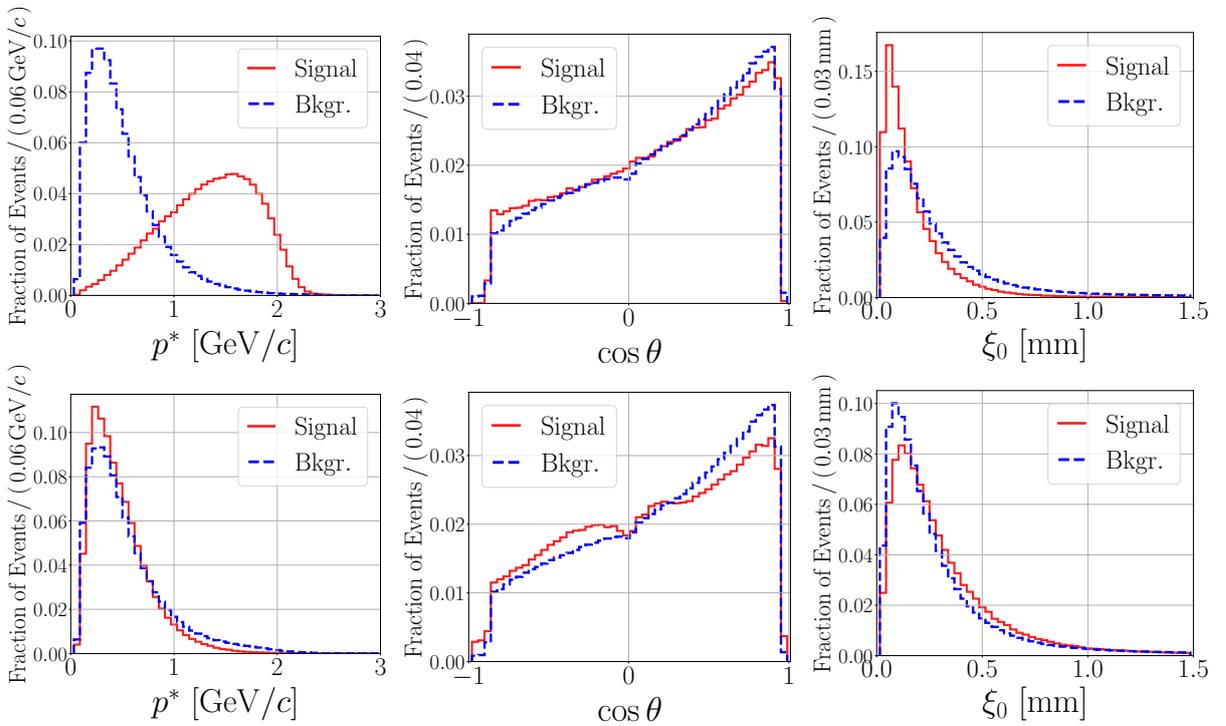


Figure 4.2: Distributions of p^* , $\cos\theta$ and ξ_0 for the Kinetic Lepton (top) and the Intermediate Kinetic Lepton categories (bottom).

(or those close by), difficulting the track identification, and causing thus a loss of track finding efficiency.

Further discrimination power is obtained from the following kinematic variables calculated in the $\Upsilon(4S)$ centre-of-mass frame:

- M_{rec}^2 , the squared invariant mass of the recoiling system X whose four-momentum is defined by:

$$p_X^\mu = \sum_{i \neq \ell} p_i^\mu,$$

where the index i goes over all charged and neutral candidates in the tag side and ℓ corresponds to the index of the lepton candidate. The squared recoil invariant mass is then

$$M_{\text{rec}}^2 = m_X^2 = g_{\mu,\nu} p_X^\mu p_X^\nu.$$

- p_{miss}^* , the absolute value of the missing momentum $\mathbf{p}_{\text{miss}}^*$ which is defined by

$$\mathbf{p}_{\text{miss}}^* = \mathbf{p}_{B_{\text{tag}}^0}^* - \mathbf{p}_X^* - \mathbf{p}_\ell^*.$$

Taking into account that the B_{tag}^0 meson is produced at rest in the $\Upsilon(4S)$ frame, i.e. $\mathbf{p}_{B_{\text{tag}}^0}^* \approx \mathbf{0}$, one obtains

$$\mathbf{p}_{\text{miss}}^* \approx -(\mathbf{p}_X^* + \mathbf{p}_\ell^*). \quad (4.12)$$

- $\cos \theta_{\text{miss}}^*$, the cosine of the angle between the momentum \mathbf{p}_ℓ^* of the lepton candidate and the missing momentum $\mathbf{p}_{\text{miss}}^*$.
- E_{90}^W , the energy in the hemisphere defined by the direction of the virtual W^\pm in the B_{tag}^0 meson decay. The momentum of the virtual W^\pm is given by

$$\mathbf{p}_W^* = \mathbf{p}_\ell^* + \mathbf{p}_\nu^* \approx \mathbf{p}_\ell^* + \mathbf{p}_{\text{miss}}^* = -\mathbf{p}_X^*,$$

where the momentum \mathbf{p}_ν of the neutrino is estimated using the missing momentum $\mathbf{p}_{\text{miss}}^*$. The sum of energies for E_{90}^W extends over all charged and neutral candidates in the recoiling system X that are in the same hemisphere with respect to the W^\pm :

$$E_{90}^W = \sum_{i \in X, \mathbf{p}_i^* \cdot \mathbf{p}_W^* > 0} E_i.$$

- $|\cos \theta_T^*|$, the cosine of the angle between the lepton candidate's momentum \mathbf{p}_ℓ^* and the thrust axis of the B_{tag}^0 in the $\Upsilon(4S)$ center-of-mass frame (see Sect. 6.7.1). The calculation of T_{tag} involves all charged and neutral candidates in the tag side.

Figure 4.3 shows the distributions of M_{rec}^2 and $M_{\text{rec}} = |M_{\text{rec}}^2|^{\frac{1}{2}}$ for the Kinetic Lepton category. The distributions show tails toward the lower side due to missing particles. The M_{rec} distribution for the signal (target leptons) peaks around the mass of a D meson ($m_D \approx 1.87 \text{ GeV}/c^2$), and the background (all other candidates) peaks around half of the mass of a B meson. The M_{rec}^2 distributions are much broader and peak at slightly lower values, but preserve the same discrimination power. In order to save computing resources, only M_{rec}^2 is used.

Figure 4.4 shows the distributions of p_{miss}^* , $\cos \theta_{\text{miss}}^*$ and $|\cos \theta_{\text{T}}^*|$ for the Kinetic Lepton category. The p_{miss}^* and the $\cos \theta_{\text{miss}}^*$ signal distributions correspond to the high momentum neutrino emitted back-to-back with the target lepton in the frame of the virtual W^\pm . Therefore, it propagates almost in opposite direction to the target lepton in the $\Upsilon(4S)$ frame. On the other hand, the $|\cos \theta_{\text{T}}^*|$ signal distribution corresponds to the target lepton, which is emitted almost in opposite direction to the remaining B_{tag}^0 decay products in the $\Upsilon(4S)$ frame. Therefore, its momentum is almost parallel (or antiparallel) to the thrust axis, which corresponds to a good approximation to the direction of the remaining B_{tag}^0 decay products.

For target electrons and muons, the distributions of M_{rec}^2 , p_{miss}^* , $\cos \theta_{\text{miss}}^*$ and $|\cos \theta_{\text{T}}^*|$ are similar, and the distributions of E_{90}^W are different. Figure 4.5 shows E_{90}^W distributions for the Electron and the Muon category obtained from Belle II and Belle MC. Primary electrons deposit a large amount of energy in the ECL and, therefore, the energy in the direction of the virtual W^\pm peaks at a larger value for the signal. A second peak in the signal distribution at lower energies is caused by inefficiencies in the ECL clustering, which is currently under development. As Fig. 4.5 (top right) shows, the peak at lower energies is less pronounced on Belle MC. Considering now the muons, they leave small energy depositions in the ECL and, therefore, the distribution peaks for them at a lower value than for the background.

Figure 4.6 shows the correlations between the input variables used for the Kinetic Lepton category. High correlations are present among the momentum variables p^* , p_t^* , p and p_t . The missing momentum p_{miss}^* is highly negatively correlated with M_{rec}^2 as expected from the definition in eq. (4.12). Other relatively high correlations are present among E_{90}^W and some PID variables, as well as among PID variables themselves.

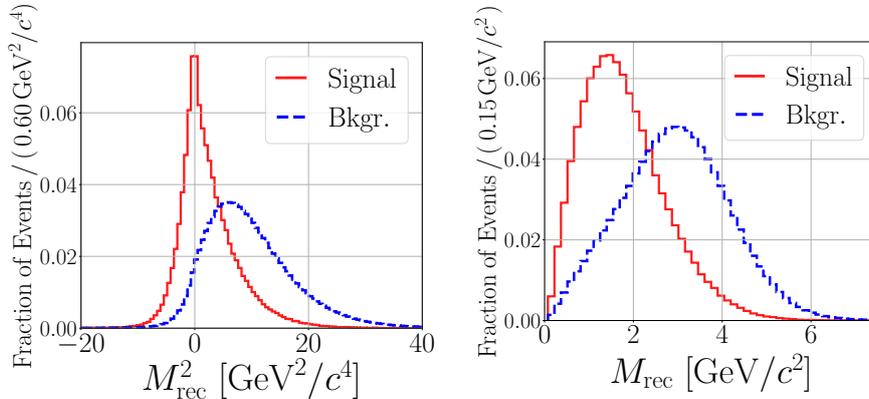


Figure 4.3: Distributions of M_{rec}^2 and $M_{\text{rec}} = |M_{\text{rec}}^2|^{\frac{1}{2}}$ for the Kinetic Lepton category.

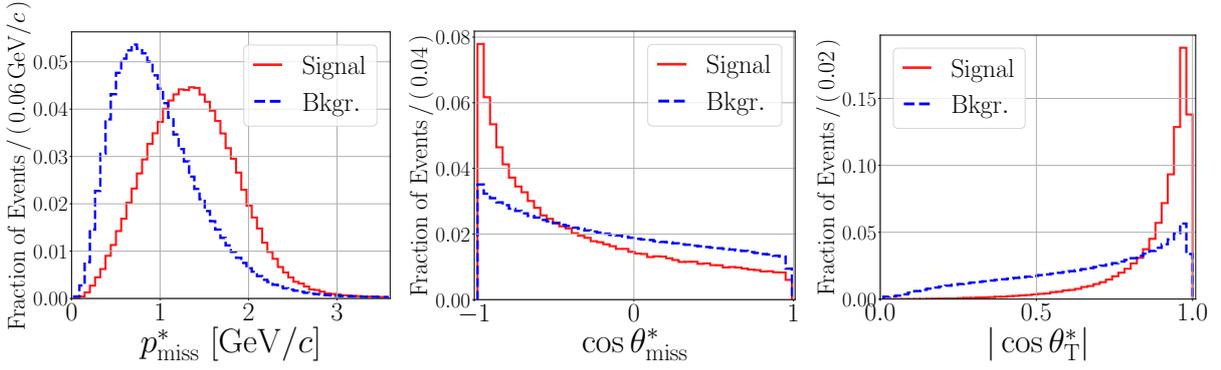


Figure 4.4: Distributions of p_{miss}^* , $\cos \theta_{\text{miss}}^*$ and $|\cos \theta_{\text{T}}^*|$ for the Kinetic Lepton category.

For all categories considering leptons as targets, Fig. 4.9 shows the ranking of the input variables provided by the FBDT algorithm. This ranking is estimated as explained in [115, 124] and is a relative ranking. This means that the variable with the highest discrimination power has a ranking equal to 100 and the lowest a ranking equal to 0. To be noted is that in cases where several variables having large discrimination power are highly correlated, one of the variables is ranked high, while the others are ranked low. In general, the kinematic variables have relatively high discrimination power for direct leptons. Intermediate leptons are kinematically more similar to the background and thus the kinematic variables have rather

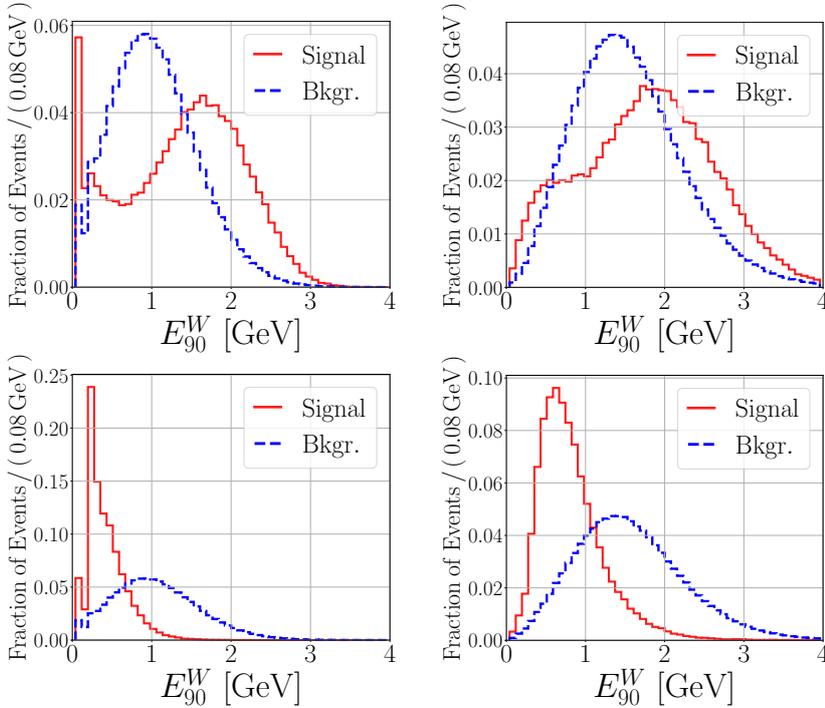


Figure 4.5: Distributions of E_{90}^W for the Electron (top) and the Muon categories (bottom) obtained from Belle II (left) and Belle MC (right).

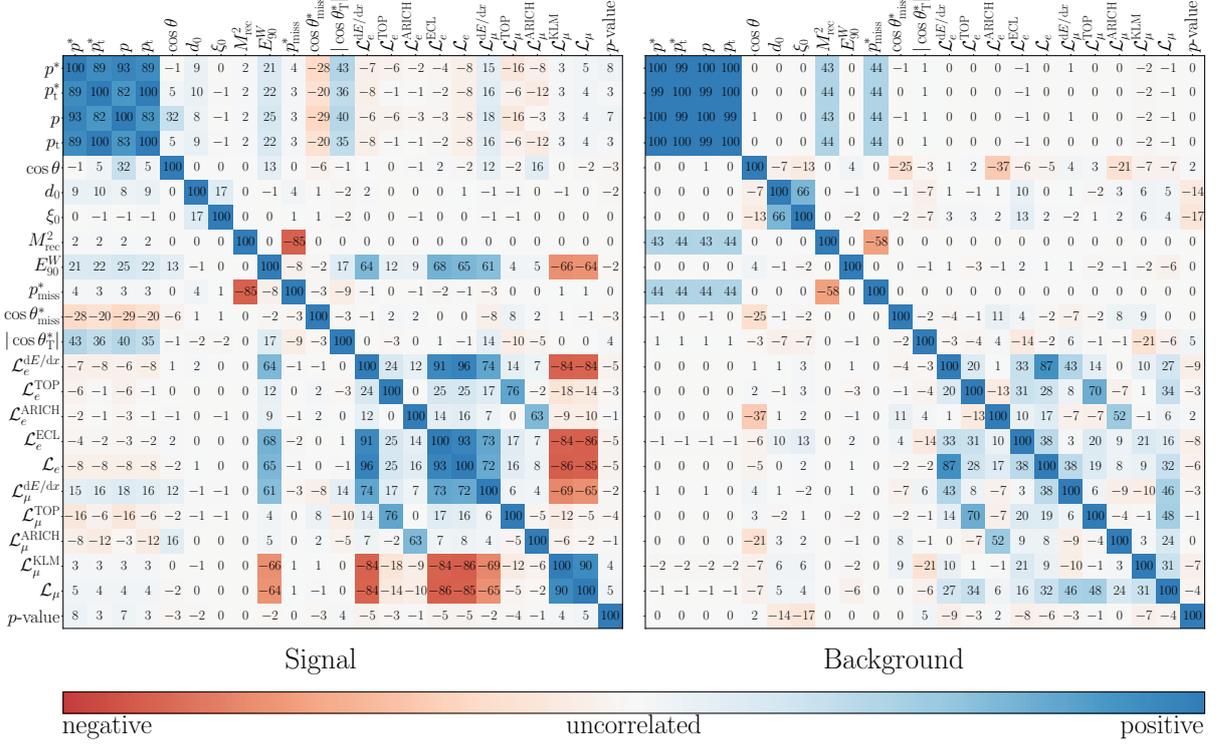


Figure 4.6: Correlations between the input variables used for the Kinetic Lepton category in percent.

low discrimination power for them. On the other hand, the PID variables have in both cases relatively high discrimination power, and dominate in the case of intermediate leptons.

Kaons

Kaons are produced predominantly through decays of charmed mesons via $b \rightarrow c \rightarrow s$ transitions, e.g. $\bar{B}^0 \rightarrow D^0 (\rightarrow K^- \pi^+ \pi^0) X$. Kaons stemming from such decays, and from decays of charmed baryons via $b \rightarrow c \rightarrow s$ transitions, tag a \bar{B}^0 if they are negatively charged, and a B^0 if they are positively charged. These kaons are referred to as *right sign* kaons.

The kaon category has the highest power to identify the B_{tag}^0 flavor because of the high inclusive branching fraction $\mathcal{B}(B^\pm/B^0 \rightarrow K^\pm) = (78.9 \pm 2.5)\%$, and because the fraction of right sign kaons ($\sim 66\%$) is much higher than the fraction of wrong sign kaons ($\sim 13\%$) [28]. Wrong sign kaons are produced through CKM-suppressed processes of the kind

$$\bar{b} \rightarrow W^+ (\rightarrow c\bar{s}/c\bar{d}) X, \quad \text{with } c \rightarrow s \rightarrow K^-,$$

and they have the opposite charge-flavor correspondence, i.e. they stem from a \bar{B}^0 decay if they are positively charged, and from a B^0 decay if they are negatively charged.

To identify target kaons, the input variables include the PID variables (\mathcal{L}_K , $\mathcal{L}_K^{\text{dE/dx}}$, $\mathcal{L}_K^{\text{TOP}}$ and $\mathcal{L}_K^{\text{ARICH}}$), the momentum variables (p^* , p_t^* , p , p_t , and $\cos \theta$) and the impact parameters (d_0 and ξ_0). The following variables provide additional discrimination power:

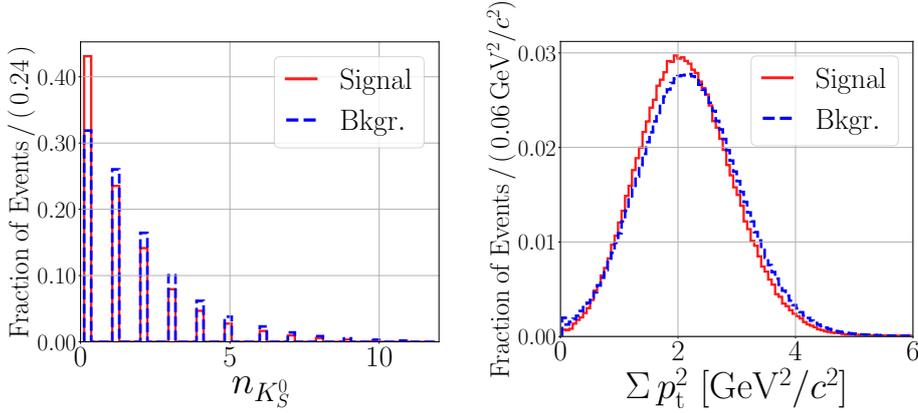


Figure 4.7: Distributions of $n_{K_S^0}$ and $\sum p_t^2$ for the Kaon category.

- $n_{K_S^0}$, the number of reconstructed K_S^0 on the tag side.
- $\sum p_t^2$, the sum of the squared transverse momentum of all tracks on the tag side in the laboratory frame.
- M_{rec}^2 , E_{90}^W , p_{miss}^* , $\cos \theta_{\text{miss}}^*$ and $|\cos \theta_T^*|$, the variables in the $\Upsilon(4S)$ frame which distinguish the lepton background.

Figure 4.7 shows the distributions of $n_{K_S^0}$ and $\sum p_t^2$ for the Kaon category. As the signal distribution of $n_{K_S^0}$ shows, a charged kaon without K_S^0 has a higher probability to be the target right sign kaon. Charged kaons produced through $b \rightarrow c\bar{c}s/c\bar{c}d$ transitions or through hadronisation of $s\bar{s}$ out of the vacuum are usually accompanied by one or more K_S^0 . An example of a $b \rightarrow c\bar{c}s$ process with hadronisation of $s\bar{s}$ out of the vacuum is

$$\bar{B}^0 \rightarrow (c\bar{c}) \left(\rightarrow K^+K^-K_S^0 \right) (s\bar{s}) \left(\rightarrow K^+K^- \right) K_S^0.$$

Considering now the variable $\sum p_t^2$, its signal distribution is less broad and peaks at a slightly lower value in comparison with the background distribution.

Figure 4.8 shows the correlations between the input variables used for the Kaon category. High correlations are present between the momentum variables, as well as between M_{rec}^2 , p_{miss}^* and $\sum p_t^2$. These last correlations are expected from the definition in eq. (4.12). The ranking of the input variables for the Kaon category is presented in Fig. 4.9. In this case, the PID variables and the impact parameters have the highest discrimination power.

Slow Pions

Pions are the most common final state particles. Primary and secondary pions are considered as target particles for several categories. The charge of secondary pions from decays $\bar{B}^0 \rightarrow XD^{*+} (\rightarrow D^0\pi^+)$ provides tagging information. Due to the small mass difference between D^{*+} and D^0 ($\sim 145 \text{ MeV}/c^2$), the secondary pions have a soft momentum spectrum and are therefore referred to as slow pions.

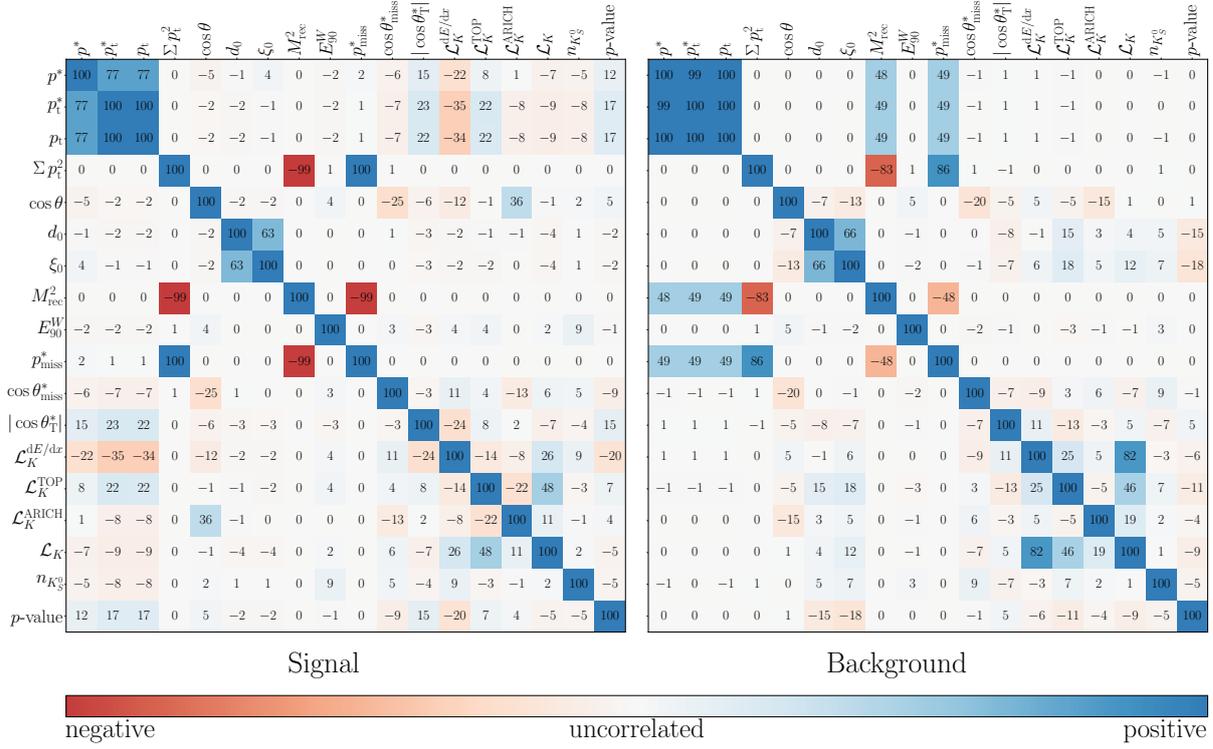


Figure 4.8: Correlations between the input variables used for the Kaon category in percent.

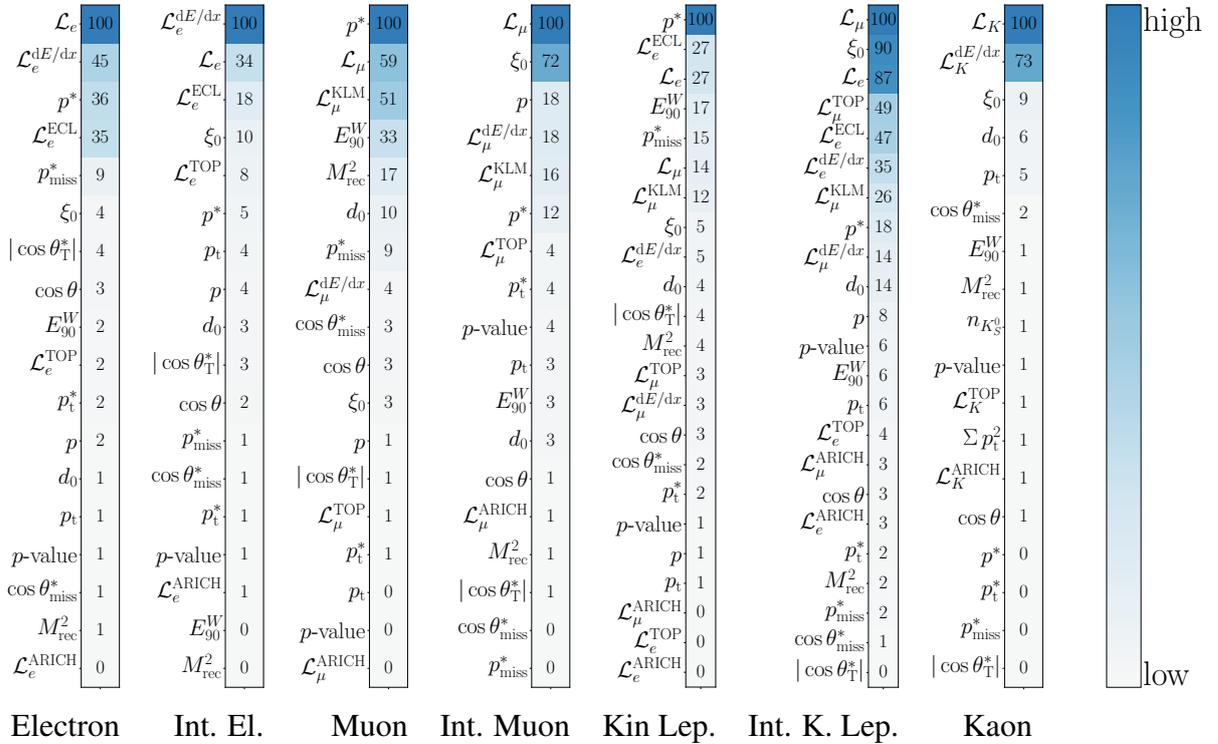


Figure 4.9: Ranking of input variables provided by the FBBDT algorithm for different categories.

The Slow Pion category uses all the variables of the Kaon category, which include some variables of the Kinetic Lepton category. Additionally, the Slow Pion category uses the pion and the electron PID variables \mathcal{L}_π , $\mathcal{L}_\pi^{\text{dE/dx}}$, $\mathcal{L}_{\pi/e}^{\text{dE/dx}}$, \mathcal{L}_e , $\mathcal{L}_\pi^{\text{TOP}}$ and $\mathcal{L}_\pi^{\text{ARICH}}$. The electron PID variables help to discriminate the target slow pions from the electrons created either through photon conversions, or through $\pi^0 \rightarrow e^+ e^- \gamma$ Dalitz decays.

Figure 4.10 shows the distributions of p^* , ξ_0 , $\cos \theta$ and $|\cos \theta_T^*|$ for the Slow Pion category. The soft momentum spectrum of the slow pions can be recognized in the signal p^* distribution. The signal distribution peaks at around 100 MeV/c. For soft momentum particles, the effect of multiple scattering increases as it is proportional to $\sim (\beta cp)^{-1}$. Thus, the reconstruction of the slow pion tracks is more difficult than e.g. for primary leptons, resulting in relatively large impact parameters as the signal distribution of ξ_0 shows.

A considerable discrimination power is provided by the variables $\cos \theta$ and $|\cos \theta_T^*|$. Slow pions are produced nearly at rest in the D^{*+} frame together with the D^0 meson. Therefore, in the $\Upsilon(4S)$ frame the slow pions fly close to the D^0 decay products and opposite to the other B_{tag}^0 decay products, while in the lab frame they travel close to the beam boost direction. Consequently, slow pions can be distinguished considering the θ angle in the lab frame, and the angle to the thrust axis θ_T^* in the $\Upsilon(4S)$ frame, as the signal distributions of $\cos \theta$ and $|\cos \theta_T^*|$ show.

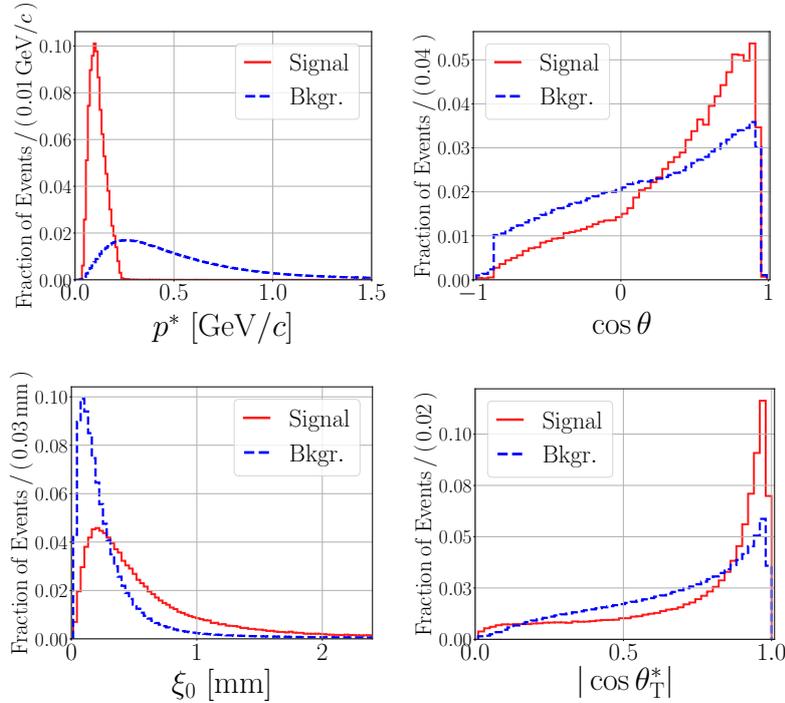


Figure 4.10: Distributions of p^* , ξ_0 , $\cos \theta$ and $|\cos \theta_T^*|$ for the Slow Pion category.

Fast Hadrons

The targets of the Fast Hadron category are kaons and pions from the W boson in $b \rightarrow c W^-$ decays, or in $b \rightarrow u W^-$ decays, and from *one-prong* decays of primary tauons, i.e.

$$\bar{B}^0 \rightarrow \tau^- (\rightarrow h^- \nu_\tau) \bar{\nu}_\tau X,$$

where h^- stays for a π^- or a K^- meson. The category considers as targets also those kaons and pions that are produced through intermediate resonances, which decay via strong processes conserving the flavor information, for example

$$\bar{B}^0 \rightarrow K^{*-} (\rightarrow K^- \pi^0) X^+,$$

and

$$\bar{B}^0 \rightarrow \tau^- (\rightarrow \rho^- (\rightarrow \pi^- \pi^0) \nu_\tau) \bar{\nu}_\tau X^+.$$

The target kaons and pions are referred to as fast hadrons because of their hard momentum spectrum. A negatively (positively) charged fast hadron indicates a \bar{B}^0 (B^0) meson.

The Fast Hadron category uses the same set of variables applied within the Slow Pion category since these variables distinguish also fast kaons and fast pions among the background of slow pions, kaons from decays of charmed hadrons and leptons.

Correlation between kaons and slow pions (Kaon-Pion)

If an event contains both a target kaon and a target slow pion, as for example in the decay

$$\bar{B}^0 \rightarrow D^{*+} (\rightarrow D^0 (\rightarrow K^- X') \pi^+) X^-,$$

additional flavor information can be gained by exploiting the angular and the charge correlations between both targets. For this, the targets have to be identified. For this, the simplest way is to use the output information of the Kaon and the Slow Pion categories.

The Kaon-Pion category uses the following variables:

- \mathcal{L}_K , the combined kaon PID likelihood.
- y_{Kaon} , the probability of being a target kaon obtained from the individual kaon category.
- y_{SlowPion} , the probability of being a target slow pion obtained from the individual slow pion category.
- $\cos \theta_{K\pi}$, the cosine of the angle between the kaon and the slow pion momentum in the $\Upsilon(4S)$ frame.
- $q_K \cdot q_\pi$, the charge product of the kaon and the slow pion candidates.

Figure 4.11 shows the distributions of y_{Kaon} , y_{SlowPion} , $\cos \theta_{K\pi}$ and $q_K \cdot q_\pi$. The discrimination power of the individual outputs from the Kaon and the Slow Pion categories can be recognized

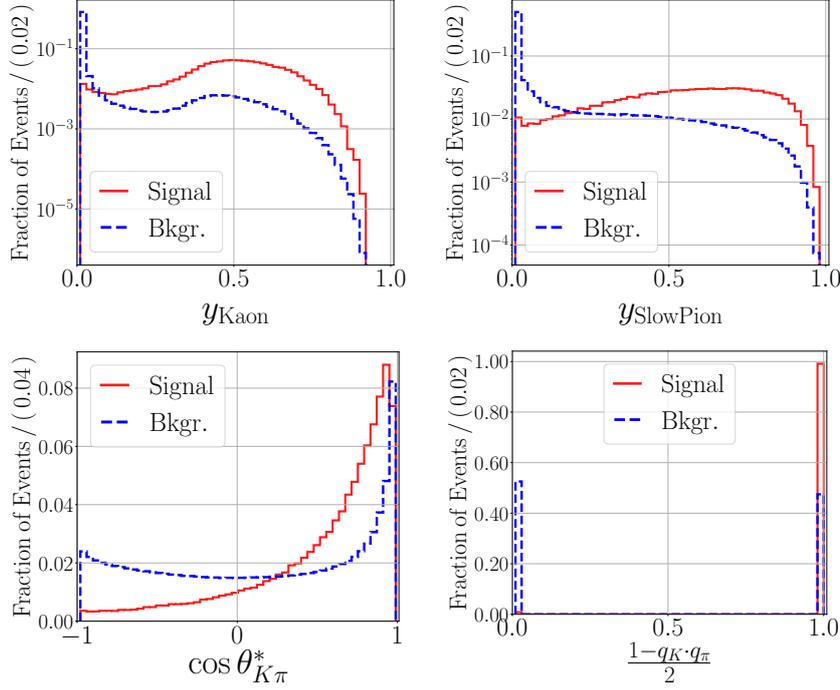


Figure 4.11: Distributions of y_{Kaon} , y_{SlowPion} , $\cos \theta_{K\pi}$ and $q_K \cdot q_\pi$ for the Kaon-Pion category. The last variable is transformed to be in the region between 0 and 1.

in the distributions of y_{Kaon} and y_{SlowPion} . As the signal $\cos \theta_{K\pi}$ distribution shows, a right sign kaon and the corresponding slow pion are emitted in approximately the same direction in the $\Upsilon(4S)$ frame, and they are produced with opposite charges in agreement with their individual charge-flavor correspondence, as the signal distribution in Fig. 4.11 (bottom right) shows.

High momentum particles (Maximum p^*)

Primary hadrons and leptons from the W^\pm boson in $b \rightarrow c W^-$, or in $b \rightarrow u W^-$, are characterised by a very hard momentum spectrum. A very inclusive tag can be performed by selecting the track with the highest momentum in the $\Upsilon(4S)$ frame and using its charge as a flavor tag. A negatively charged fast particle indicates a \bar{B}^0 , and a positively charged a B^0 meson.

The purpose is to recover flavor tagging information from primary particles that may have not been selected either as a primary lepton or as a fast hadron. The discriminating input variables are the momentum variables (p^* , p_t^* , p , p_t , and $\cos \theta$), the impact parameters (d_0 and ξ_0) and $|\cos \theta_T^*|$.

Correlation between fast and slow particles (FSC)

Events of the kind $\bar{B}^0 \rightarrow D^{*+} W^-$ contain both a target slow pion and a high momentum primary particle originating from the W^\pm boson. In that case, additional flavor tagging information can be gained by using the correlations between the slow pion and the high momentum particle.

The W^- and the D^{*+} are produced back-to-back in the B_{tag}^0 center-of-mass frame. Therefore, the angle between the track of the target fast particle and the target slow pion is expected to be very large. The discriminating input variables are:

- p_{Slow}^* , the momentum of the slow pion candidate in the $\Upsilon(4S)$ frame.
- p_{Fast}^* , the momentum of the high momentum candidate in the $\Upsilon(4S)$ frame.
- \mathcal{L}_K , the PID kaon likelihood.
- $|\cos \theta_{\text{T, Slow}}^*|$, the cosine of the angle between the thrust axis and the slow pion candidate.
- $|\cos \theta_{\text{T, Fast}}^*|$, the cosine of the angle between the thrust axis and the high momentum candidate.
- $\cos \theta_{\text{SlowFast}}^*$, the cosine of the angle between the slow and the high momentum candidates.
- $q_{\text{Slow}} \cdot q_{\text{Fast}}$, the charge product of the slow pion and the high momentum candidates.

Figure 4.12 shows the distributions of $|\cos \theta_{\text{T, Fast}}^*|$, $\cos \theta_{\text{SlowFast}}^*$ and $q_{\text{Slow}} \cdot q_{\text{Fast}}$. Since the virtual W^\pm boson is emitted back-to-back to the D^{*+} meson in the B_{tag}^0 frame, primary particles from the W^\pm move close to the thrust axis of B_{tag}^0 , and in opposite direction to the slow pion in the $\Upsilon(4S)$ frame, as the signal distributions of $|\cos \theta_{\text{T, Fast}}^*|$ and $\cos \theta_{\text{SlowFast}}^*$ show. In agreement with their individual charge-flavor correspondence, primary particles and slow pions have to be produced with opposite charges, as Fig. 4.12 (right) shows.

Lambda baryons

Additional flavor tagging information can be obtained by considering the flavor of Λ baryons produced via the cascade transition $b \rightarrow c \rightarrow s$ as a flavor signature. The presence of such a Λ baryon indicates a \bar{B}^0 , and the presence of such a $\bar{\Lambda}$ a B^0 . Although the fraction of events containing a target Λ is rather small, they provide relatively clean and independent flavor tagging information which complements the other categories.

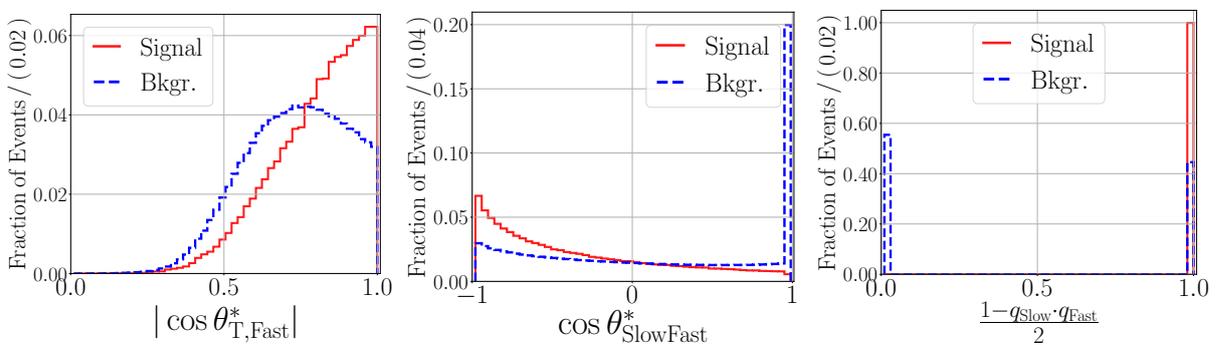


Figure 4.12: Distributions of $|\cos \theta_{\text{T, Fast}}^*|$, $\cos \theta_{\text{SlowFast}}^*$ and $q_{\text{Slow}} \cdot q_{\text{Fast}}$ for the FSC category. The last variable is transformed to be in the region between 0 and 1.

The Λ candidates are obtained by reconstructing $\Lambda \rightarrow p\pi^-$ decays, or $\bar{\Lambda} \rightarrow \bar{p}\pi^+$ decays. This is performed by combining proton and pion candidates on the tag side. In addition to the momentum variables of the reconstructed Λ , and the momentum variables of the proton and the pion used for the reconstruction, the following input variables are used:

- $\mathcal{L}_p, \mathcal{L}_\pi$ the PID likelihoods of the proton and the pion.
- q_Λ , the flavor of the Λ baryon.
- M_Λ , the reconstructed mass of the Λ .
- $n_{K_S^0}$, the number of reconstructed K_S^0 on the tag side.
- $\cos \theta_{\mathbf{x}_\Lambda \mathbf{p}_\Lambda}$, the cosine of the angle between the Λ momentum \mathbf{p}_Λ and the direction from the IP to the reconstructed Λ vertex \mathbf{x}_Λ in the laboratory frame.
- $|\mathbf{x}_\Lambda|$, the absolute distance between the Λ vertex and the IP.
- σ_Λ^{zz} , the error of the Λ vertex fit in z -direction.
- χ_Λ^2 , the χ^2 probability of the reconstructed Λ decay vertex.

Figure 4.13 shows the distributions of M_Λ , $n_{K_S^0}$, $\cos \theta_{\mathbf{x}_\Lambda \mathbf{p}_\Lambda}$ and $|\mathbf{x}_\Lambda|$. As discussed for the Kaon category, decays with hadronisation of a $s\bar{s}$ pair out of the vacuum are usually accompanied by one or more K_S^0 . An example of such a decay is

$$\bar{B}^0 \rightarrow \Lambda \bar{\Lambda} K_S^0.$$

Thus, a lambda candidate produced without accompanying K_S^0 particles has a higher probability to be the target particle, as the signal distribution of $n_{K_S^0}$ shows. Since the lambda trajectory is not influenced by the solenoid field, the reconstructed momentum \mathbf{p}_Λ should be parallel to the direction from the IP to the reconstructed vertex \mathbf{x}_Λ in the lab frame, as the signal $\cos \theta_{\mathbf{x}_\Lambda \mathbf{p}_\Lambda}$ distribution shows. The signal $|\mathbf{x}_\Lambda|$ distribution with a long tail towards large values is caused by the relatively long lifetime $\tau_\Lambda \approx 0.26$ ns, resulting in $c \cdot \tau_\Lambda \approx 8$ cm.

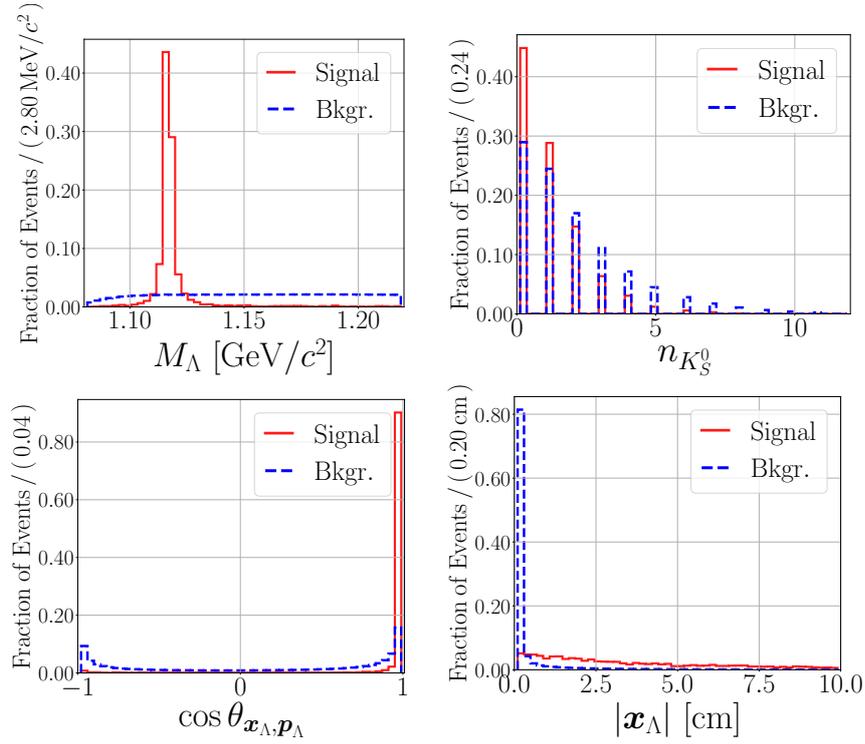


Figure 4.13: Distributions of M_Λ , $n_{K_S^0}$, $\cos \theta_{\mathbf{x}_\Lambda, \mathbf{p}_\Lambda}$ and $|\mathbf{x}_\Lambda|$ for the Lambda category.

4.4 The category-based algorithm

The Belle II category-based flavor tagger created in this thesis is a modular algorithm based on multivariate methods. It provides an output value $y \in [-1, 1]$ equivalent to $y = q \cdot r$, where $q = \text{sgn}(y)$ is the tagged B_{tag}^0 flavor, and $r = |y|$ is the flavor dilution factor. When $y = -1$, B_{tag}^0 is perfectly tagged as \bar{B}^0 , and when $y = +1$, B_{tag}^0 is perfectly tagged as B^0 .

The flavor of B_{tag}^0 results from a combination of the thirteen flavor signatures discussed in the previous section. Each of these signatures corresponds to the output of a single category which can be understood as an individual sub-tagger. A schematic overview of the information flow in the algorithm is presented in Fig. 4.14.

The algorithm of the flavor tagger is a 2-level-process: event and combiner level. The event level process is performed within each individual category. On this level, a multivariate method assigns to each particle candidate an output value $y_{\text{cat}} \in [0, 1]$. This output value can be interpreted as the probability of being the target of the corresponding category providing the right flavor tag.

The particle candidates correspond to the tracks in the tag side. To form a particle candidate from a track, an invariant mass is assigned to it. For the five types of considered particles, e , μ , K , π and p , an independent list of candidates is built by assigning the corresponding PDG mass to the tracks in the tag side. Each category considers the list of candidates belonging to its own

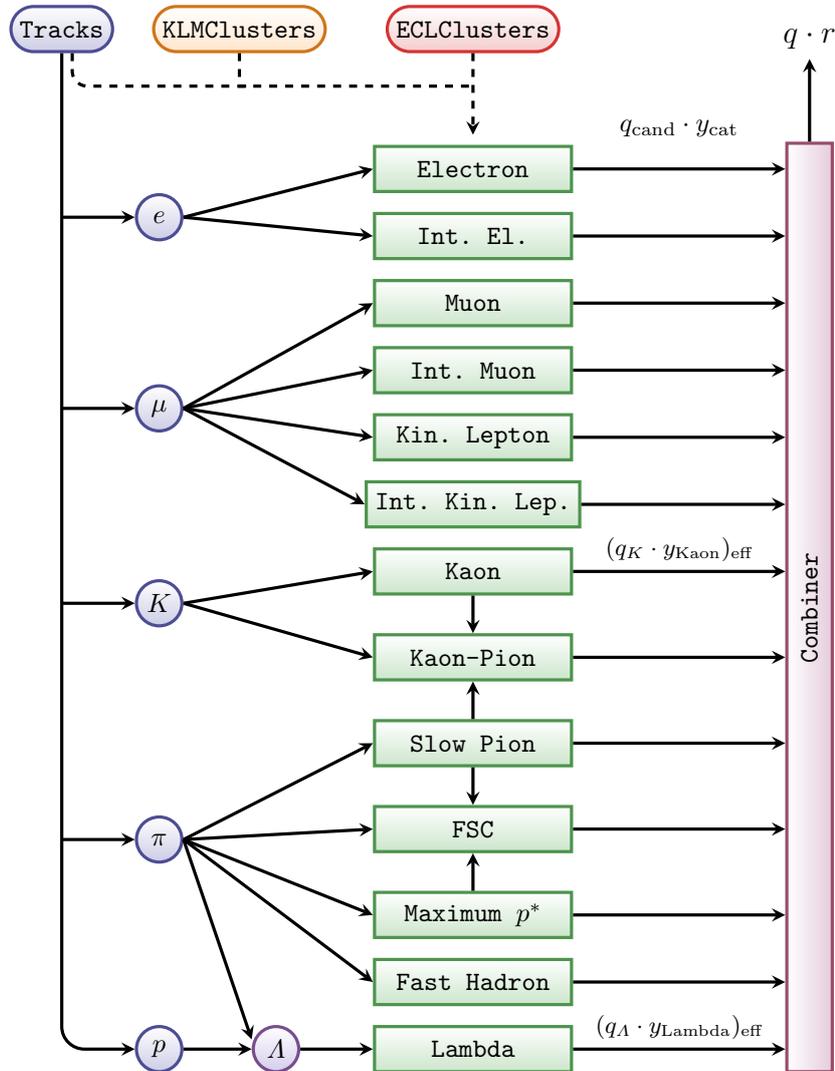


Figure 4.14: Schematic overview of the category-based flavor tagger. The tracks in the tag side are used to build five different lists of candidates (e , μ , K , π and p). Each category considers the list of candidates belonging to its own targets. The different categories are represented by green boxes, and the combiner by a magenta box.

targets. For the special case of the Lambda category, the target candidates are Λ particles reconstructed from pairs of proton and pion candidates.

In order to determine y_{cat} , the event-level multivariate methods get the input variables of the corresponding category (see Tab. 4.2). Some input variables require information from all reconstructed tracks and all neutral ECL and KLM clusters in the tag side. There are two special categories which get information from other categories: the Kaon-Pion category, which gets input information from the Kaon and the Slow Pion categories, and the FSC category which gets input information from the Slow Pion and the Maximum p^* categories.

Summing the input variables for all categories yields a total number $n_{\text{input}} = 220$. Some variables are used several times for the same candidates in different categories. To save computing time, each variable is calculated only once for each candidate. For example, E_{90}^W is used in the six leptonic categories, and in the Kaon, the Slow Pion, and the Fast Hadron categories. Thus, it counts nine times in the calculation of n_{input} . The Electron and the Intermediate Lepton categories consider the list of electron candidates; the other four leptonic categories consider the list of muon candidates; the Kaon category considers the list of kaon candidates; and the Slow Pion and the Fast Hadron categories consider the list of pion candidates. For each track, the variable E_{90}^W is calculated only four times: one time for the electron, for the muon, for the kaon, and for the pion candidate. Thus, E_{90}^W counts as four different variables. Considering now all input variables, the total number of different variables is $n_{\text{unique}} = 108$.

Within each category, the particle candidates are ranked according to the values of y_{cat} . The candidate with the highest y_{cat} is selected as target. Only for the Maximum p^* category, the target is the candidate with the largest momentum in the $\Upsilon(4S)$ frame.

The procedure within each single category is illustrated in Fig. 4.15. All categories contribute to the final tag. This fact helps to improve the performance of the flavor tagger since the B_{tag}^0 can decay in such a way that there is more than one flavor-specific signature.

The combiner level is the last step in the process. It corresponds to a multivariate method that receives thirteen input values, i.e. one input value from each category, and gives $y = q \cdot r$ as output. Each input value is the product $q_{\text{cand}} \cdot y_{\text{cat}}$ of each category, where the charge q_{cand} and the probability y_{cat} correspond to the particle candidate selected as target. For two special cases, the Kaon and the Lambda categories, the input value is the effective product

$$(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}} = \frac{\prod_i (1 + (q_{\text{cand}} \cdot y_{\text{cat}})_i) - \prod_i (1 - (q_{\text{cand}} \cdot y_{\text{cat}})_i)}{\prod_i (1 + (q_{\text{cand}} \cdot y_{\text{cat}})_i) + \prod_i (1 - (q_{\text{cand}} \cdot y_{\text{cat}})_i)}, \quad (4.13)$$

where the products extend over the three particles with the highest y_{cat} value. For the Lambda category, q_{cand} corresponds to the B^0 flavor tagged by the Λ candidate, i.e. $q_{\Lambda} = -1(+1)$ for $\Lambda(\bar{\Lambda})$. The use of $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ for the Kaon category improves considerably the performance of the flavor tagger. This will be discussed in Sect. 4.5.4.

The multivariate method chosen for the event and the combiner level is the Fast Boosted Decision Tree (FBDT) [115]. The FBDT is a stochastic gradient-boosted decision tree developed especially for Belle II by Thomas Keck and is the default multivariate method of the experiment. It incorporates several mechanisms for regularization and is optimized to save

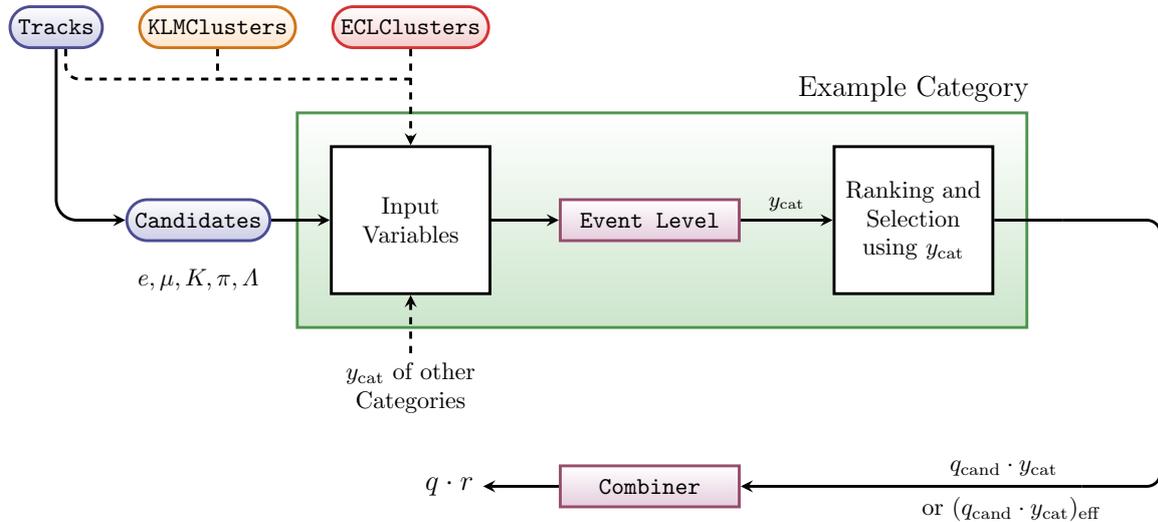


Figure 4.15: Procedure for each single category (green box): the candidates correspond to the reconstructed tracks for a specific mass hypothesis. The input variables are presented in Tab. 4.2; some of them consider all reconstructed tracks and all neutral ECL and KLM clusters on the tag side. The magenta boxes represent multivariate methods: y_{cat} is the output of the event level. The output of the combiner is equivalent to the product $q \cdot r$.

computing resources during its application and during its training procedure. To cross-check the result of the FBDT, an independent multivariate method, a multi-layer perceptron (MLP), is employed for the combiner level. The implementation of the MLP is based on the open-source library FANN [125], where the acronym stays for Fast Artificial Neural Network.

For both combiner level methods, the input values are identical. The flavor tagger provides the output of both the FBDT and the MLP combiners. The cross-check with a MLP is not performed for the event level because the MLP implementation does not include the preprocessing of input variables performed by the FBDT algorithm. The FBDT maps the shape of the input variables to a uniform integer distribution. This preprocessing is called equal-frequency binning and is illustrated in Fig. 4.16. To avoid a computing overhead during the application phase, the FBDT applies the equal-frequency binning only during the training process. After the training, the FBDT parameters are transformed to match the original scale of the input variables.

The equal-frequency binning allows the flavor tagger to exploit the full range of the input variables considering large tails and outliers in the input variable distributions, making the algorithm more robust and avoiding cuts and expensive scaling computations. Avoiding cuts is crucial because the performance of the track reconstruction and PID algorithms changes periodically. Under these conditions, the amount of time and work needed to continuously optimize the cuts by hand would increase indefinitely.

For the MLP, the input variables are usually scaled to $[-1, 1]$ in order to make full use of the region where the \tanh activation function behaves linearly with respect to the input. At the

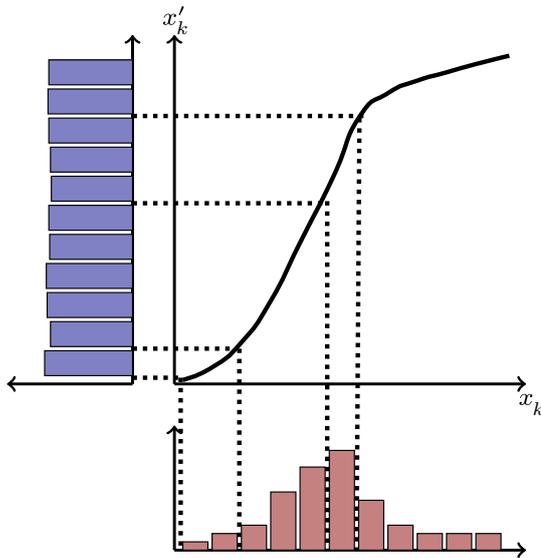


Figure 4.16: Equal-frequency binning. The variable x_k is binned so that each bin contains roughly the same amount of events. The dotted lines indicate the bin boundaries [115].

combiner level, this scaling is not necessary because the inputs from each category are within the range $[-1, 1]$ by definition.

To avoid any bias from a possible statistical correlation, the flavor tagger is trained using two statistically independent MC samples: one sample for the event level, and one sample for the combiner level. At each training step, one half of the sample is used as training sample and the other half as a test (validation) sample for an unbiased evaluation of the performance against over-fitting. The event level is trained first and each category is trained independently. The FBDT and the MLP combiners are trained afterwards.

The evaluation against over-fitting for each of the multi-variate methods is performed by comparing the distribution of the output on the training sample and on the test sample. The output on the training sample has to overlap with the output on the testing sample within the statistical uncertainties. At the event level, the evaluation is performed separately for signal and for background components and, at the combiner level, for the B^0 and for the \bar{B}^0 components. Figures 4.17 and 4.18 show examples of evaluations against over-fitting for the Kinetic Lepton and for the Kaon categories, and for the FBDT and the MLP combiners. Several different mechanisms inherent to the FBDT serve to prevent over-fitting [115]. In the case of the MLP, the validation sample is used actively for the stopping criterion during the training procedure as explained in [112, 126].

To control the structure of the multivariate-method model, the learning procedure, and the preprocessing of the input data, the FBDT and the MLP implementations provide different so-called *hyper-parameters*. For the FBDTs, the flavor tagger uses the default hyper-parameter values which were optimized by the FBDT developer for the identification of B mesons and other intermediate particles in the Full Event Interpretation (FEI) algorithm [115, 127]. For the flavor tagger, only the *number of trees* was optimized to 500. For the MLP combiner, the hyper-parameter values were adopted from previous studies for the Belle II z -vertex trigger [112, 126]. Only the *cost function* was modified to be linear.

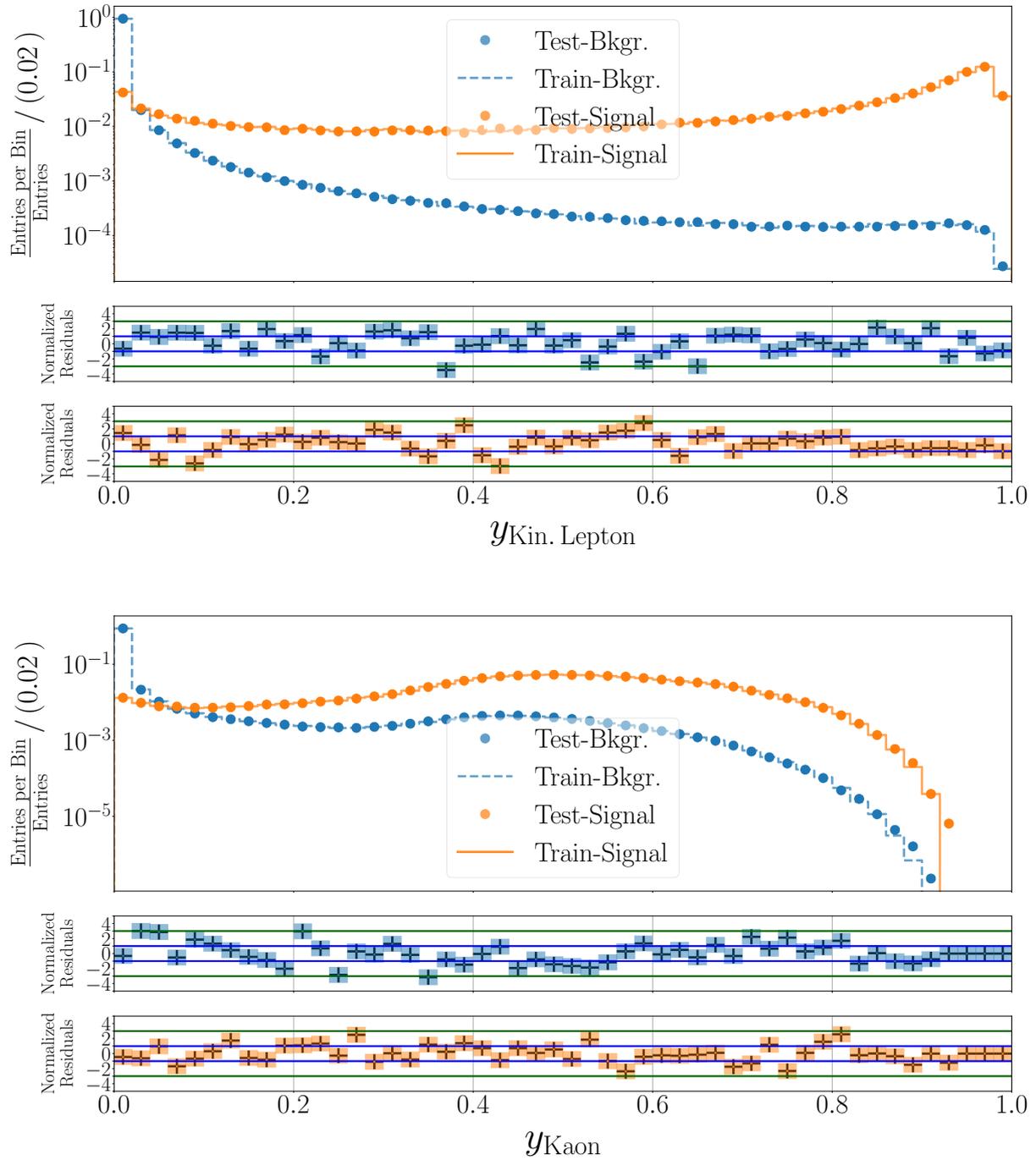


Figure 4.17: Evaluations against over-fitting for the Kinetic Lepton and for the Kaon categories. The test is performed separately for signal and for background components.

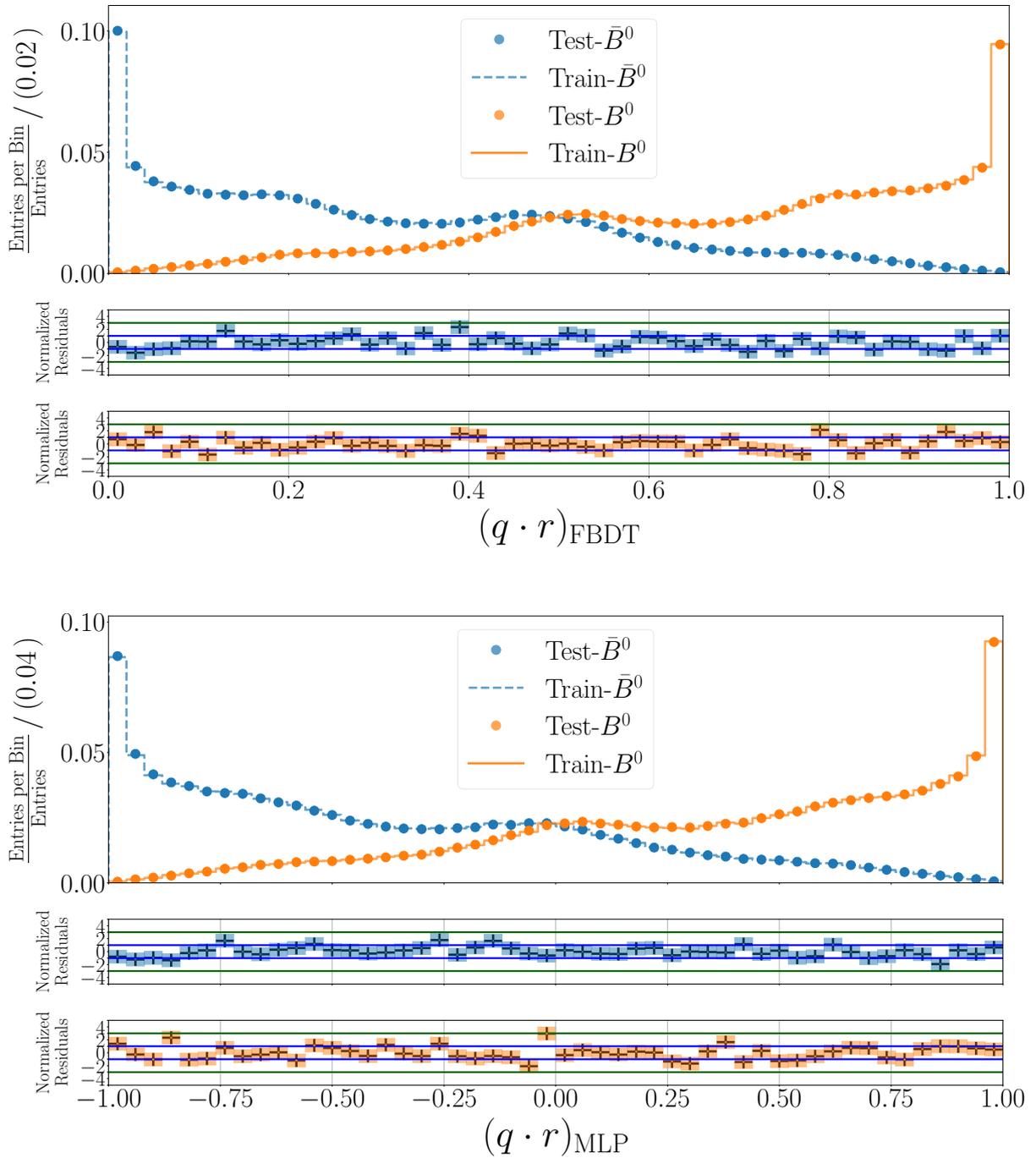


Figure 4.18: Evaluations against over-fitting for the FBDT and the MLP combiners. The output of the FBDT combiner is mapped to the region between 0 and 1. The test is performed separately for the B^0 and for the \bar{B}^0 components.

4.5 Optimization and performance on MC

This section covers the key steps in the optimization and the evaluations of the performance of the flavor tagger on MC simulations for Belle and Belle II. The key goal of the optimization was to maximize the effective efficiency since a high effective efficiency is essential to reduce the uncertainties on CP -violation parameters and thus to enhance the sensitivity to NP effects.

The purpose of using Belle II MC as well as Belle MC was to evaluate the improvements of the algorithm on two independent MC samples which simulate the same physical processes, but have to cope with different experimental conditions and have realized different tracking and PID algorithms. Additionally, a validation on data becomes possible by comparing the results obtained with Belle MC and with Belle data. The results on both MC samples are presented always together. But this is not to be understood as a comparison between the capabilities of Belle II and Belle: the used Belle II MC was simulated without background and cannot yet be validated on data, while the Belle MC was simulated with background and is the product of several years of development and validation using Belle data. The evaluations using Belle data will be presented in Sect. 4.8 and 4.9.

In the following, the generation and the reconstruction of the MC samples used to optimize and to characterize the performance of the algorithm will be described. The calculation of the effective efficiency and the concept of the linearity check, an important step in the validation of the flavor tagger, will be explained afterward. The next subsection introduces the identification of target particles, a crucial task for ensuring an optimal use of the flavor information in B^0 meson decays. Then, the evaluation of the performance of the individual categories is presented together with the optimization of the input preprocessing for the combiners. The last subsection presents the overall evaluation of the performance of the FBDT and the MLP combiners.

4.5.1 The MC samples

The MC events used for training and for testing correspond to $B^0\bar{B}^0$ pairs in which one meson, namely B_{sig}^0 , decays to $J/\psi K_S^0$, while B_{tag}^0 decays to any possible final state according to the known branching fractions [28]. In these events, the signal J/ψ decays into two muons and the signal K_S^0 into two charged pions.

The Belle II MC events used for the development and the validation of the flavor tagger were generated within an official MC campaign using the Belle II Analysis Software Framework (BASF2). These events serve also to validate the vertex reconstruction tools (see Sect. 3.4). When the flavor tagger was developed, two official MC samples were available: one generated without background, and another with simulated background. However, the track reconstruction algorithm for events with background was very poor and, therefore, only the sample without background was used during the development and the optimization of the flavor tagger. In this chapter, all the results are obtained using Belle II MC samples without background.

The official Belle II MC samples are simulated with CP violation such that $\mathcal{A}_{CP} = 0$ and $\mathcal{S}_{CP} = 0.69$. These values of \mathcal{A}_{CP} and \mathcal{S}_{CP} agree within one standard deviation with the

current world averages for $b \rightarrow c\bar{c}s$ transitions [42]. To study the impact of different CP -violation parameters on the performance of the flavor tagger, private Belle II MC samples were generated with the same BASF2 release used to produce the official samples. The private samples were generated with $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$, and with $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, -0.69)$. These samples have the same sizes of the official ones.

In the case of Belle MC events, the simulation as well as the reconstruction of clusters, tracks and K_S^0 mesons is performed using the Belle Analysis Software Framework (BASF). These events were simulated taking into account the effect of the background, as well as the size and the position of the interaction region at KEKB for the production vertex of the $\Upsilon(4S)$. The CP -violation parameters are the same as for the official Belle II MC samples. To be analyzed within BASF2, the Belle MC events have to be converted. The conversion routine is documented in [127].

In the Belle MC events, the K_S^0 candidates are reconstructed from two oppositely charged tracks using BASF as explained in [128, 129]. In the case of Belle II MC, K_S^0 candidates are reconstructed from two oppositely charged tracks as $\pi^+\pi^-$ candidates, requiring only that the reconstructed invariant mass $M_{\pi^+\pi^-}$ of the $\pi^+\pi^-$ candidates is within

$$-0.25 \text{ GeV}/c^2 < M_{\pi^+\pi^-} - m_{K_S^0} < +0.25 \text{ GeV}/c^2,$$

where $m_{K_S^0}$ denotes the nominal K_S^0 mass $m_{K_S^0} = 0.498 \text{ GeV}/c^2$ [28].

Except for the reconstruction of K_S^0 mesons, the reconstruction of the decay channel $B_{\text{sig}}^0 \rightarrow J/\psi (\rightarrow \mu^+\mu^-) K_S^0 (\rightarrow \pi^+\pi^-)$ is performed in the same way for both Belle and Belle II MC events using BASF2. To include almost 100% of the B_{sig}^0 candidates, the kinematic requirements on the K_S^0 , on the J/ψ , and on the B_{sig}^0 candidates were chosen to be broad enough.

The J/ψ mesons are reconstructed using pairs of oppositely charged tracks as $\mu^+\mu^-$ candidates, requiring the invariant mass $M_{\mu^+\mu^-}$ of the $\mu^+\mu^-$ candidates to be within

$$-0.11 \text{ GeV}/c^2 < M_{\mu^+\mu^-} - m_{J/\psi} < +0.11 \text{ GeV}/c^2,$$

where $m_{J/\psi}$ denotes the nominal J/ψ mass $m_{J/\psi} = 3.097 \text{ GeV}/c^2$ [28].

Reconstructed K_S^0 and J/ψ candidates are combined to form B_{sig}^0 candidates by applying requirements on two kinematic variables, the energy difference $\Delta E = E_B^* - E_{\text{beam}}^*$ and the beam-constrained mass $m_{\text{bc}} = \sqrt{(E_{\text{beam}}^*)^2 - (p_B^*)^2}$. E_{beam}^* is the beam energy in the $\Upsilon(4S)$ frame, and E_B^* and p_B^* are the reconstructed energy and momentum of the B -meson candidate in the $\Upsilon(4S)$ frame. These variables are discussed in more detail in Sect. 6.4. The requirements are $|\Delta E| < 0.15 \text{ GeV}$ and $m_{\text{bc}} > 5.2 \text{ GeV}/c^2$.

To be used for training and testing, the reconstructed MC events have to fulfill two additional requirements. First, the signal decay channel $B_{\text{sig}}^0 \rightarrow J/\psi (\rightarrow \mu^+\mu^-) K_S^0 (\rightarrow \pi^+\pi^-)$ has to be fully reconstructed and correctly matched with the MC decay chain. Second, none of the reconstructed tracks remaining on the tag side is associated with the signal B_{sig}^0 meson.

After the selection, the Belle II and the Belle MC events are split into two training samples (one for the event level and one for the combiner level) and one testing sample. The size of each training sample is about 1.3 million events for Belle II MC, and about 1 million events for Belle MC. And the size of the training sample is about 2.6 million events for Belle II MC, and about 2 million events for Belle MC.

4.5.2 Total effective efficiency

During the development of the flavor tagger, the training procedure, the choice of the categories and the choice of the input variables have been arranged in order to maximize the total effective efficiency ε_{eff} . The effective efficiency ε_{eff} introduced in eq. (4.4) is redefined by sorting the tagged events into bins of the dilution factor $r = |y| = |q \cdot r|$, where y is the output of the flavor tagger. The binning is required because the calibration of the flavor tagger can be performed only in bins of r , as it will be explained in Sect. 4.14. The binning used in this work corresponds to the same r binning used by the Belle experiment [130]. Using this binning, the performance of the Belle II flavor tagger can be compared with the performance of the Belle flavor tagger. Just for visualization, Fig. 4.19 shows the Belle binning and the distribution of the dilution factor r obtained from the output of the FBDT combiner using the official Belle II MC.

In bins of r , the total effective efficiency on MC is

$$\varepsilon_{\text{eff}} = \varepsilon \sum_i \frac{n_i}{N_{\text{tag}}} \cdot r_{\text{MC},i}^2 = \sum_i \frac{n_i}{N} \cdot r_{\text{MC},i}^2 = \sum_i \varepsilon_i \cdot r_{\text{MC},i}^2, \quad \text{using } \varepsilon_i = \frac{n_i}{N}, \quad (4.14)$$

where the sum extends over all r bins. The quantities n_i and $r_{\text{MC},i}$ are the number of tagged events and the true dilution factor in the i -th bin. The tagging efficiency corresponds to $\varepsilon = \frac{N_{\text{tag}}}{N}$, with N_{tag} being the number of taggable events and N the total number of events. Since the flavor signatures are provided by tracks, all events containing at least one reconstructed track on the tag side are taggable. The value of ε determined on Belle MC and on Belle II MC is 99.9%.

For each bin i , the true dilution factor $r_{\text{MC},i}$ (eq. (4.1)) is determined from the wrong-tag fraction $w_{\text{MC},i}$ in the bin. An event is wrongly tagged if the sign of the combiner output is opposite to the true flavor q_{MC} of B_{tag}^0 , i.e. if $q_{\text{MC}} \neq q = \text{sgn}(q \cdot r)$. When a flavor tagger is correctly calibrated on MC, the true dilution factor $r_{\text{MC},i}$ corresponds to the average $\langle r \rangle_i$ of the output dilution $r = |y| = |q \cdot r|$ in each r bin. This correspondence can be checked by plotting $r_{\text{MC},i}$ against $\langle r \rangle_i$. Such *linearity checks* are presented from Sect. 4.5.5 on.

The effective efficiency ε_{eff} can be calculated using the mean values $\langle r \rangle_i$, or using the true values $r_{\text{MC},i}$. For a flavor tagger correctly calibrated on MC, ε_{eff} has the same value in both cases. Since the MC can be biased with respect to the data, to perform CP violation analyses it is necessary to measure the true wrong-tag fraction $w_{\text{data},i}$ on data, and from it the true dilution $r_{\text{data},i}$, as it will be discussed in Sect. 4.14. If the MC and the flavor tagger are correctly calibrated, $r_{\text{MC},i}$ agrees with $r_{\text{data},i}$.

To quantify potential asymmetries in the performance of the flavor tagger, the effective efficiency on MC events is calculated separately for events where the B_{tag}^0 meson is a B^0 ($q_{\text{MC}} = +1$), and for events where the B_{tag}^0 meson is a \bar{B}^0 ($q_{\text{MC}} = -1$). The flavor q_{MC} corresponds to the true flavor of B_{tag}^0 determined using MC information. In bins of r , the total effective efficiency for events where $B_{\text{tag}}^0 = B^0$ is

$$\varepsilon_{\text{eff}}(B^0) = \sum_i \frac{n_{B^0}^i}{N_{B^0}} (r_{\text{MC},B^0})_i^2 = \sum_i \varepsilon_{B^0}^i (r_{\text{MC},B^0})_i^2, \quad \text{using} \quad \varepsilon_{B^0}^i = \frac{n_{B^0}^i}{N_{B^0}}. \quad (4.15)$$

The efficiency $\varepsilon_{\text{eff}}(\bar{B}^0)$ can be written analogously by replacing consistently B^0 with \bar{B}^0 . For the studies on MC events, the difference between effective efficiencies is then given by

$$\Delta\varepsilon_{\text{eff}} = \varepsilon_{\text{eff}}(B^0) - \varepsilon_{\text{eff}}(\bar{B}^0). \quad (4.16)$$

And the total average effective efficiency (eq. (4.14)) is calculated from the dilution factor (eq. (4.1)) obtained from

$$w_{\text{MC},i} = \frac{(w_{\text{MC},B^0})_i + (w_{\text{MC},\bar{B}^0})_i}{2}. \quad (4.17)$$

The statistical uncertainties of the effective efficiencies and of the different quantities characterizing the performance of the flavor tagger are calculated via Gaussian error propagation. The corresponding formulas can be found in App. B.2.

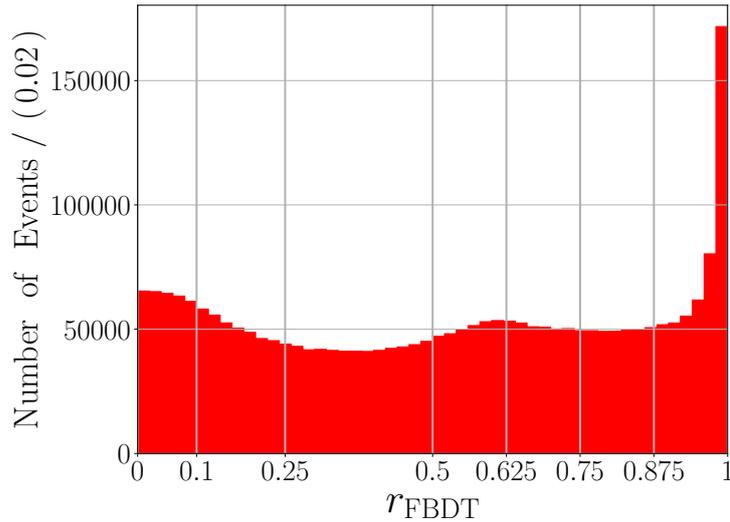


Figure 4.19: Flavor dilution factor $r_{\text{FBDT}} = |y_{\text{FBDT}}|$ obtained from the output $y_{\text{FBDT}} = (q \cdot r)_{\text{FBDT}}$ of the FBDT combiner using Belle II MC. The figure shows the r binning which was introduced by the Belle collaboration [29] and that is used throughout this thesis.

4.5.3 Optimization of the target identification

Each individual tagging category has an associated event-level multivariate method which identifies the tracks, or the target Λ candidates, corresponding to the respective target particles. When the method is trained, the input variables have to be provided together with a target variable, which has different values for signal (target) and for background candidates, for example +1 for signal candidates and -1 for background candidates. For this, one has to identify the signal candidates by considering the MC particles associated with the candidates.

For the Electron, the Muon and the Kinetic Lepton categories, it is very easy to identify the target tracks: if the associated MC particle is the considered lepton with the correct charge-flavor correspondence, and if its mother particle is B_{tag}^0 , then the track is classified as target track. Two other categories for which the identification of the target track is very easy are the Slow Pion and the Maximum p^* categories. For the Slow Pion category, a candidate is classified as target if the associated MC particle is a pion with the right charge-flavor correspondence and if its mother particle is a charged D^* . The charged D^* has to be the daughter of B_{tag}^0 . For the Maximum p^* category, the target track is required only to have the charge-flavor correspondence of a primary lepton, or a primary meson, i.e. to be positively charged if B_{tag}^0 is a B^0 , or negatively charged if B_{tag}^0 is a \bar{B}^0 .

For the other eight categories, the identification of the targets is more complicated. In the following, the developed target identification is explained for these cases.

The target identification

Resulting from a series of optimization steps, the developed target identification enabled the flavor tagger to maximally exploit the available flavor information. The terminology which will be used to describe the procedure refers to the MC numbering scheme of the PDG [28]. For example, when charmed baryons or charmed mesons are mentioned, this refers to the complete lists of charmed baryons and charmed mesons which can be found in the review of the PDG group. A $q\bar{q}$ hadronic resonance denotes any of the neutral particles in the list of light $I = 1$ mesons, the list of light $I = 0$ mesons, and the list of $c\bar{c}$ mesons. When the decay chain producing a MC particle is examined, this means that its mother, its grandmother, its grand-grandmother and so forth are examined up to the B_{tag}^0 meson.

- For the **Kaon** category, a track is classified as signal if the associated MC particle is a kaon with the right charge-flavor correspondence, and if a charmed meson, or a charmed baryon, is found in the decay chain producing it. In this way, kaons produced via $b \rightarrow c \rightarrow s$ transitions with one or several intermediate hadronic resonances are correctly identified as signal. An example are kaons from decays such as

$$\bar{B}^0 \rightarrow D^+ (\rightarrow K^{*0} (\rightarrow K^- \pi^+) X'^+) X^-.$$

The procedure identifies as well the kaons that stem from baryonic decays occurring via $b \rightarrow c \rightarrow s$ transitions, for example kaons from decays such as

$$\bar{B}^0 \rightarrow \Lambda_c (\rightarrow K^- p \pi^+) X^-.$$

Kaons produced in decays with intermediate resonances and in baryonic decays are similar in terms of kinematics and particle identification to those produced in decays of primary charmed mesons, and they provide flavor information. Considering them as signal during the training procedure improves considerably the tagging performance of the whole algorithm. Additionally, it is required that no $q\bar{q}$ hadronic resonance is present in the decay chain. This avoids classifying as signal the kaons produced in decays such as

$$\bar{B}^0 \rightarrow D^0 (\rightarrow \phi (\rightarrow K^+ K^-) X') X.$$

The only kaons that can be wrongly selected as signal are non-resonant kaons from decays such as $D^0 \rightarrow K^+ K^- (\pi^0)$. To filter out these kaons, a full scan of the whole B_{tag}^0 decay chain would be needed for each event. Besides being very expensive in terms of computing resources, such a scan brings no considerable improvement: these decays are very rare, e.g. $\mathcal{B}(D^0 \rightarrow K^+ K^-) \approx 4 \cdot 10^{-3}$, and they are CP eigenstates causing no bias of the tagged flavor. Their effect can be thus safely neglected.

- For the **Kaon-Pion** category, an event is selected as signal if the candidates selected by the Slow Pion and by the Kaon categories are associated with a MC pion and a MC kaon having the right charge-flavor correspondences, and if the mother of the pion is a primary $D^{*\pm}$, which is found in the decay chain producing the kaon.
- For the **Lambda** category, a Λ candidate is selected as signal if the associated MC particle is a Λ with the right flavor correspondence, and if a charmed baryon is found in the decay chain producing it. This identification implicitly rejects Λ particles produced through hadronisation of $s\bar{s}$ out of the vacuum having by chance the right flavor as, for example, those from the decay $\bar{B}^0 \rightarrow \bar{\Lambda} \Lambda D^0$.
- For the three **Intermediate-Lepton** categories, a track is selected as signal if the associated MC particle is the considered lepton (or one of the two for the Int. Kin. Lepton category) with the right charge-flavor correspondence, and if there is no $q\bar{q}$ resonance in the decay chain producing it. In this way, intermediate leptons from decays with hadronic resonances are considered as signal. An example of such a process is

$$\bar{B}^0 \rightarrow D^{*+} (\rightarrow D^0 (\rightarrow K^- e^+ \nu_e) \pi^+) X^-,$$

where the intermediate electron is produced via a $b \rightarrow c \rightarrow s e^+ \nu_e$ transition.

- For the **Fast Hadron** category, the signal candidates are required to be associated with a MC kaon or a MC pion with the right charge-flavor correspondence. Additionally, the

MC particle has to fulfill one of the following three requirements. Requirement one: it is a primary particle. Requirement two: the decay chain producing it contains one charged resonance belonging to the list of light $I = 1$ mesons, or to the list of strange mesons, and no $q\bar{q}$ resonance. This approach considers as target fast pions providing flavor information that are produced in decays with intermediate resonances. An example of such a decay is $\bar{B}^0 \rightarrow \rho^-(\rightarrow \pi^-\pi^0) X^+$. And requirement three: the MC particle is produced in a one-prong decay of a primary tau lepton such as $\bar{B}^0 \rightarrow \tau^-(\rightarrow \pi^-\nu_\tau) \bar{\nu}_\tau X^+$. Examples with fast kaons can be obtained by exchanging the π^- for a K^- , and the ρ^- for a K^{*-} in the previously given examples. Processes producing fast kaons have smaller branching fractions than those producing fast pions because they are Cabibbo suppressed, but the fast kaons provide flavor information as well.

Particles that can be wrongly classified as fast hadron targets are pions and kaons from non-resonant decays such as $\bar{B}^0 \rightarrow \pi^+\pi^-$, $\bar{B}^0 \rightarrow K^+\pi^-$, and $\bar{B}^0 \rightarrow K^0K^+K^-$. However, these decays are very rare with branching fractions around 10^{-5} or smaller. Their effect is thus negligible for flavor tagging.

- Finally, for the **Fast-Slow-Correlation** category, an event is considered as signal if the particle selected by the Maximum p^* category is a daughter of B_{tag}^0 , and if $q\bar{q}$ resonances are absent in the decay chain producing the slow pion. The MC particle associated with the slow pion candidate is required to not be a primary particle. Both fast and slow candidates are required to have the right charge-flavor correspondences. Since the slow pion candidate that the FSC category receives is selected by the trained Slow Pion category (the trained methods do not receive MC information), there is a probability that the slow pion candidate is a background particle. Requiring $q\bar{q}$ resonances to be absent in the decay chain producing it rejects background candidates having by chance the right charge-flavor correspondence.

Figure 4.20 presents the output qr distributions of the FBDT and the MLP combiners on Belle II MC and on Belle MC. To produce these figures, the MC information is examined for each event. If the target particle of a certain category is found in the tag side, this category is labeled as true category in this event. The event contains then one true category. If the targets for two categories are found, then the event contains two true categories, and so forth. In the case of the Kaon-Pion and the FSC categories, they are labeled as true if the two respective target particles are found in the tag side. After the examination, the events are sorted according to the number of true categories that are found in them.

As Fig. 4.20 shows, the distributions for B^0 and \bar{B}^0 overlap almost perfectly in the cases where no category can be attributed to the MC events (True Categories = 0). These distributions overlap so well because they belong to residual events lacking in discrimination power to distinguish between B^0 and \bar{B}^0 . The fraction of events for which no category can be attributed is 7.17%. Within these events, only a tiny fraction of events such as $B^0 \rightarrow J/\psi K^{*0}(\rightarrow K^+\pi^-)$ still contain flavor information. However, such events have branching fractions $\lesssim 10^{-3}$ and,

therefore, they cannot bring a considerable improvement. All particles providing clean flavor signatures are correctly considered as signal by the individual categories. In this sense, the flavor tagger uses all the available flavor information.

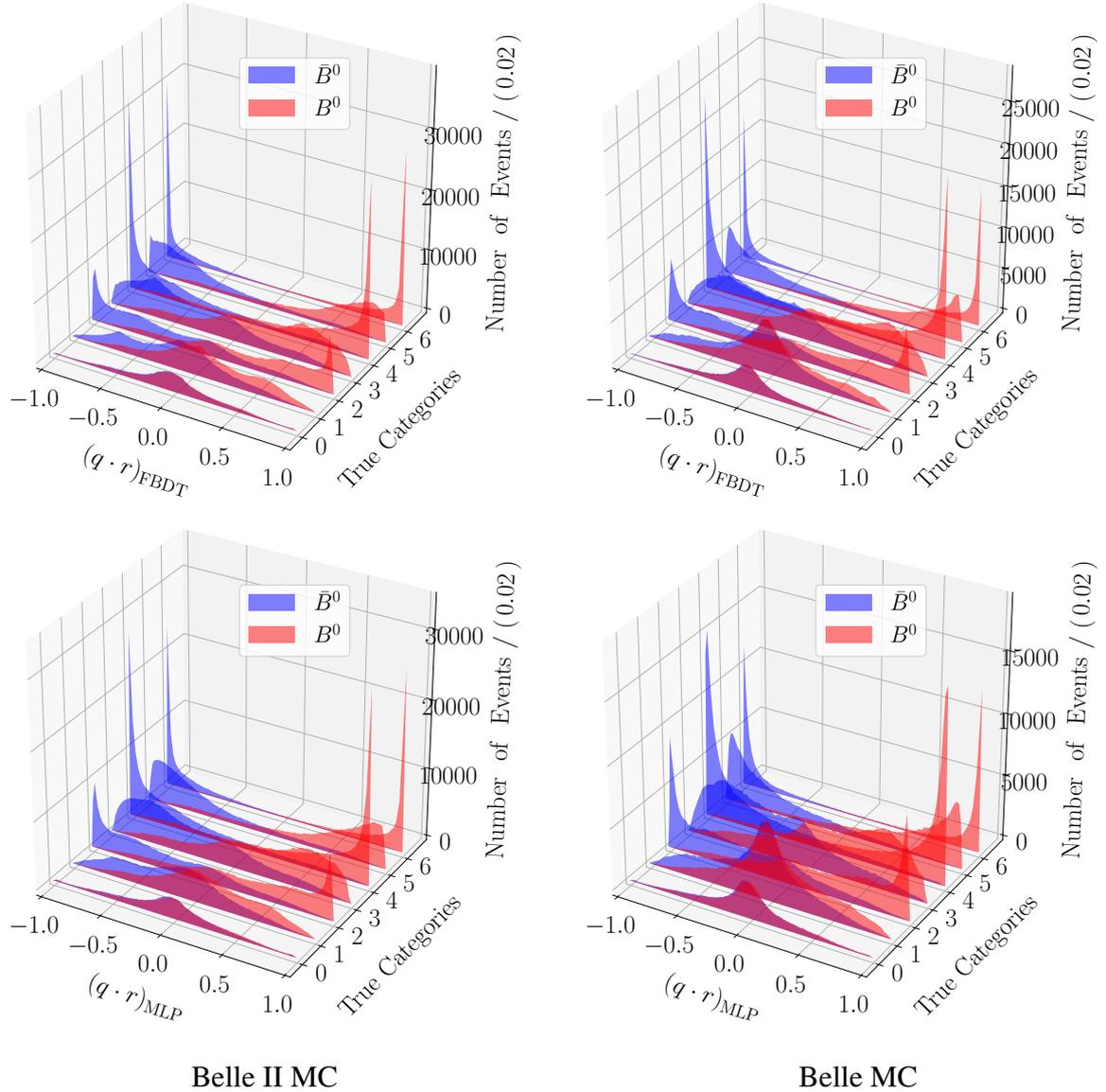


Figure 4.20: Output $q \cdot r$ distributions of the FBDT (top) and the MLP (bottom) combiners on Belle II MC (left) and on Belle MC (right). The distributions are shown for events where B_{tag}^0 is a B^0 and for events where it is \bar{B}^0 . The events are sorted according to the number of true categories. A category is considered as true for a certain event if its target particles are found in the tag side of the events by examining the MC information.

4.5.4 Performance of individual categories

Within each category, the respective event-level multivariate method assigns a value y_{cat} to each particle candidate. The candidates are ranked according to y_{cat} , or, in the case of the Maximum p^* category, according to the momentum p^* in the $\Upsilon(4S)$ frame. The candidate with the highest y_{cat} , or p^* , is selected and the product of its charge q_{cand} times y_{cat} is used as input for the combiner methods. Only for the Kaon and the Lambda categories, the combiner input value is the effective product $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ (eq. (4.13)). The effective product combines the flavor information provided by the three candidates with the highest y_{cat} value. The number three was chosen since it led to the best results.

The distributions of the thirteen combiner input values are presented in Figs. 4.21 to 4.23 for events where the B_{tag}^0 meson is a B^0 , and for events where the B_{tag}^0 meson is a \bar{B}^0 . These distributions were obtained using Belle II MC. The large peaks at zero, which are present in all the distributions apart from that belonging to Maximum p^* (Fig. 4.23 (left)), correspond to events where the selected candidate is very unlikely to provide flavor information. In general, a value close to zero indicates that the probability of finding a certain flavor-specific signature within the B_{tag}^0 final state is very low. A value closer to ± 1 indicates a more reliable flavor tag. In the case of Maximum p^* , there is no peak at zero since this category is inclusive for all tracks, i.e. any event with tracks on the tag side provides flavor information.

Figure 4.23 (right) shows the relative ranking provided by the FBDT combiner. The ranking provided by the FBDT is explained in detail in [115, 124]. The correlations between the combiner input values are presented in Fig. 4.24.

For each individual category, an effective efficiency is calculated by taking the corresponding combiner input value $q_{\text{cand}} \cdot y_{\text{cat}}$, or $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$, as single flavor tag, i.e. considering each category as a ‘‘sub-tagger’’. The effective efficiency is obtained from the distribution of the combiner input values (see for example Figs. 4.21 to 4.23) in the following way. The combiner input values are sorted into 100 bins with $j = 1, 2, \dots, 100$. The dilution factor for each bin j is calculated separately for $B_{\text{tag}}^0 = B^0$ ($q_{\text{MC}} = +1$) and for $B_{\text{tag}}^0 = \bar{B}^0$ ($q_{\text{MC}} = -1$) events as

$$r_{B^0}^j = |1 - 2w_{B^0}^j| = \left| 1 - 2 \cdot \frac{n_{B^0}^j}{n_{B^0}^j + n_{\bar{B}^0}^j} \right|, \quad (4.18)$$

where $n_{B^0}^j$ and $n_{\bar{B}^0}^j$ are the number of tagged B^0 ($q_{\text{MC}} = +1$) and \bar{B}^0 ($q_{\text{MC}} = -1$) events in this bin. The dilution factor $r_{\bar{B}^0}^j$ is calculated by swapping B^0 with \bar{B}^0 in eq. (4.18). Subsequently, the $r_{B^0}^j$ values are grouped depending on their values into the seven r bins shown in Fig. 4.19 (Sect. 4.5.2). For each r bin, the average dilution factor $\langle r_{B^0} \rangle_i$ and the number of

tagged B^0 events $n_{B^0}^i$ are calculated as

$$\langle r_{B^0} \rangle_i = \frac{\sum_k n_{B^0}^k \cdot r_{B^0}^k}{n_{B^0}^i}, \quad \text{and} \quad n_{B^0}^i = \sum_k n_{B^0}^k, \quad (4.19)$$

where k extends over all entries in the i -th r bin. Similarly, the average dilution factor $\langle r_{\bar{B}^0} \rangle_i$ and the number of tagged \bar{B}^0 events $n_{\bar{B}^0}^i$ are obtained by swapping B^0 with \bar{B}^0 in eq. (4.19). Using these quantities, the effective efficiency ε_{eff} and the difference $\Delta\varepsilon_{\text{eff}}$ are calculated as in eqs. (4.15) and (4.16).

Table 4.3 presents a comparison of the effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ for individual categories calculated using $q_{\text{cand}} \cdot y_{\text{cat}}$ and using $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$. The results are shown for Belle MC and for Belle II MC. The only category for which the use of $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ leads to an increase in effective efficiency of more than one percent point on Belle MC and on Belle II MC is the Kaon category. For the Electron, the Muon, the Kinetic Lepton, the Kaon-Pion, the Slow Pion and the Fast Hadron categories, the use of $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ leads only to slight increases. For the other categories, the effective efficiency either remains constant or decreases slightly.

Table 4.4 shows the effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ of the FBDT and the MLP combiners for three different cases: I) the product $q_{\text{cand}} \cdot y_{\text{cat}}$ is used for all categories; II) the effective product $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ is used only for the Kaon and the Lambda categories; III) the effective product is used for the categories whose individual performance was improved by its use (all except the Int. Muon, the Int. Kinetic Lepton, the Maximum p^* , and the Lambda categories). Table 4.4 presents the results on Belle MC and on the official Belle II MC. In comparison with case I, in case II the FBDT and the MLP combiners reach an increase in effective efficiency by about two percent points on Belle MC and on Belle II MC. This increase agrees with the improvement of the performance of the Kaon category due to the use of the effective product. In comparison with case II, in case III the total effective efficiency of both combiners decreases slightly for Belle II MC and stays almost constant for Belle MC.

As a conclusion, the final version of the category-based flavor tagger uses the effective product as combiner input variable only for the Kaon and the Lambda categories (case II in Tab. 4.4). For other categories, the effective product is not used because it does not improve the total effective efficiency of the FBDT and the MLP combiners, and because its use only increases the number of computations per B_{tag}^0 meson candidate.

Independent of the use of the effective product, the categories considering primary leptons, kaons, and pions as targets provide clean flavor signatures and reach sufficiently high effective efficiencies. These categories, luckily, are found frequently in flavor-specific decays with high branching fractions. The intermediate lepton categories and the lambda category provide marginal tagging power.

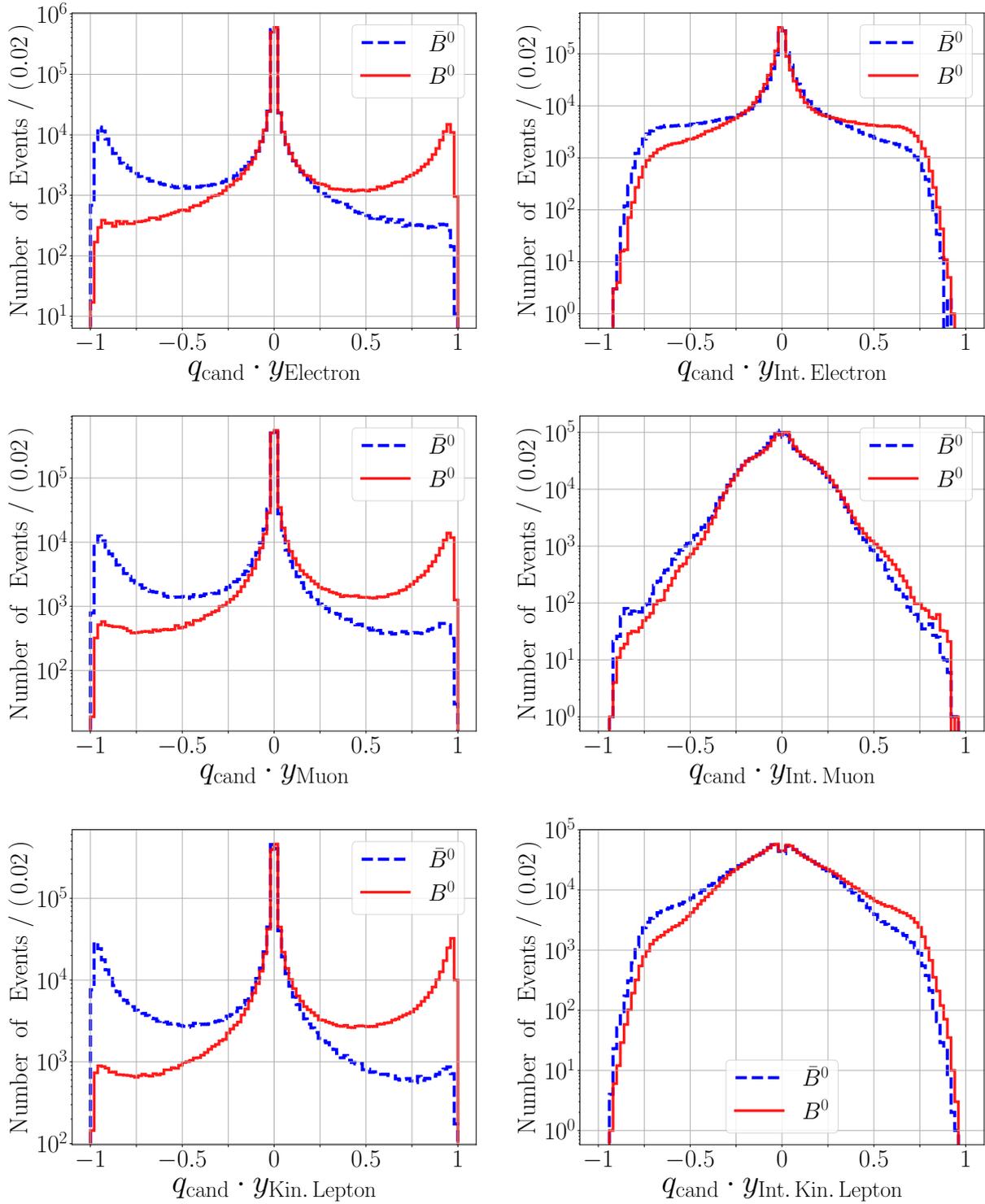


Figure 4.21: Combiner input distributions for the categories: Electron, Intermediate Electron, Muon, Intermediate Muon, Kinetic Lepton and Intermediate Kinetic Lepton. The distributions are shown for events with a true B^0 and with a true \bar{B}^0 in the tag side.

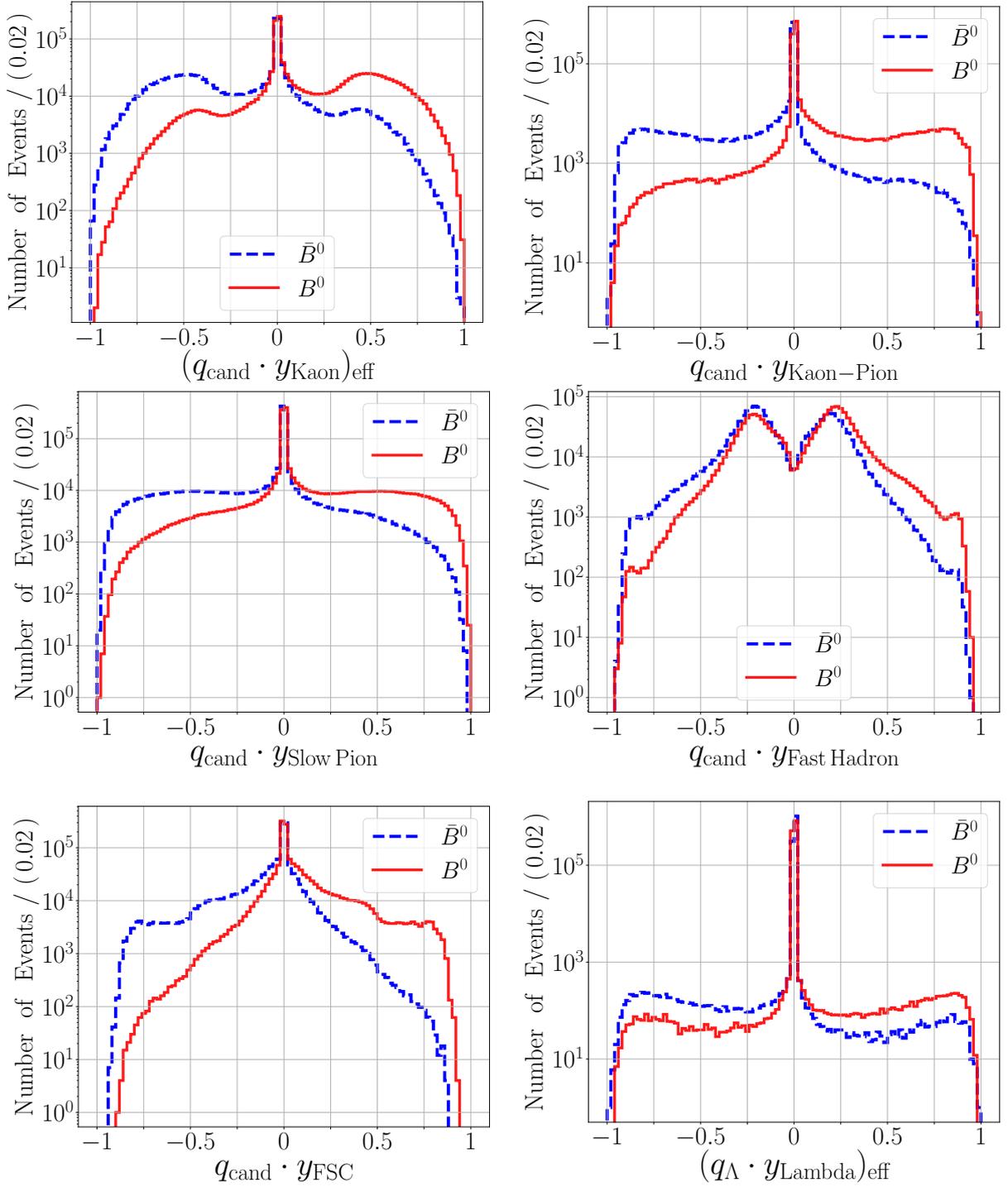


Figure 4.22: Combiner input distributions for the categories: Kaon, Kaon-Pion, Slow Pion, Fast Hadron, FSC and Lambda. The distributions are shown for events with a true B^0 and with a true \bar{B}^0 in the tag side.

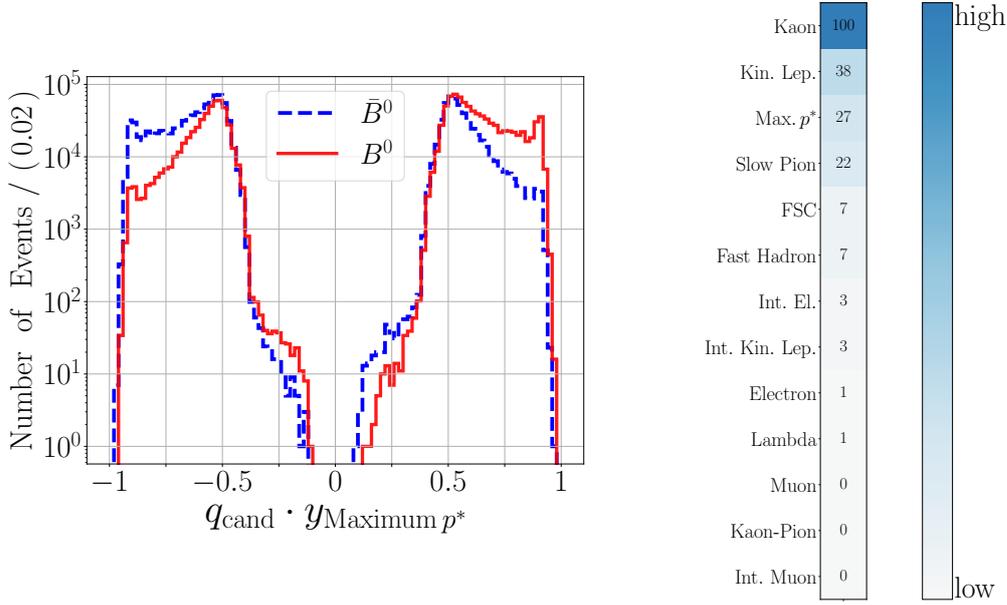


Figure 4.23: Combiner input distribution for the Maximum p^* category (left), and ranking of the combiner inputs obtained from the different tagging categories (right). The ranking is provided by the FBDT algorithm (right).

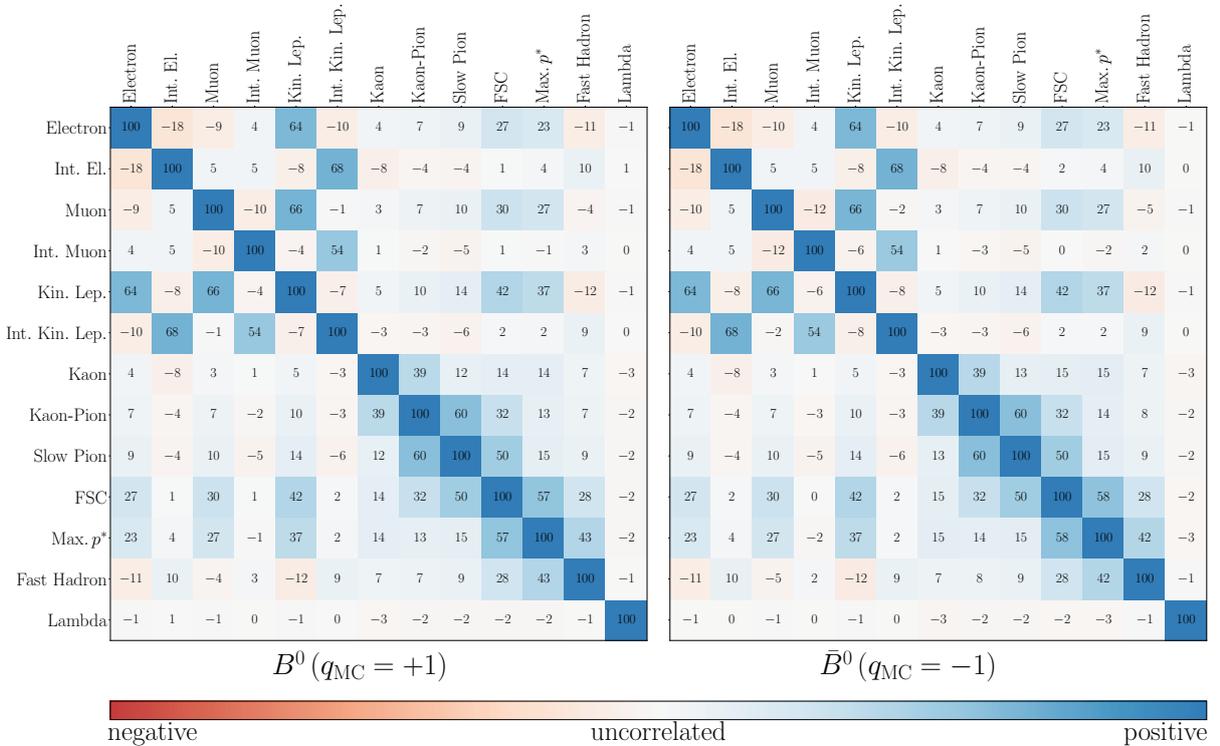


Figure 4.24: Correlations between the combiner inputs obtained from the different tagging categories.

Table 4.3: Performance of individual categories calculated using the product $q_{\text{cand}} \cdot y_{\text{cat}}$ and using the effective product $(q_{\text{cand}} \cdot y_{\text{cat}})_{\text{eff}}$ as combiner input variable. All values are given in percent.

Category	Variable	Belle II MC		Belle MC	
		$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$
Electron	$q \cdot y$	5.58 ± 0.01	0.25 ± 0.02	5.80 ± 0.01	-0.04 ± 0.03
	$(q \cdot y)_{\text{eff}}$	5.68 ± 0.01	0.30 ± 0.02	5.91 ± 0.03	0.10 ± 0.06
Int. Electron	$q \cdot y$	1.40 ± 0.01	-0.22 ± 0.01	0.74 ± 0.01	0.00 ± 0.01
	$(q \cdot y)_{\text{eff}}$	1.43 ± 0.01	-0.18 ± 0.01	0.74 ± 0.01	0.03 ± 0.01
Muon	$q \cdot y$	5.64 ± 0.01	0.27 ± 0.02	5.74 ± 0.01	0.08 ± 0.03
	$(q \cdot y)_{\text{eff}}$	6.04 ± 0.01	0.30 ± 0.03	5.98 ± 0.03	0.10 ± 0.06
Int. Muon	$q \cdot y$	0.30 ± 0.01	-0.02 ± 0.01	0.33 ± 0.01	0.00 ± 0.01
	$(q \cdot y)_{\text{eff}}$	0.28 ± 0.01	0.00 ± 0.01	0.37 ± 0.01	0.01 ± 0.01
Kin. Lepton	$q \cdot y$	11.44 ± 0.02	0.43 ± 0.03	11.70 ± 0.02	0.08 ± 0.04
	$(q \cdot y)_{\text{eff}}$	12.05 ± 0.02	0.51 ± 0.04	12.04 ± 0.04	0.16 ± 0.09
Int. Kin. Lep.	$q \cdot y$	1.31 ± 0.01	-0.11 ± 0.01	0.56 ± 0.01	0.00 ± 0.01
	$(q \cdot y)_{\text{eff}}$	1.03 ± 0.01	-0.03 ± 0.01	0.60 ± 0.01	0.02 ± 0.01
Kaon	$q \cdot y$	17.57 ± 0.01	-0.56 ± 0.03	16.15 ± 0.03	-0.17 ± 0.07
	$(q \cdot y)_{\text{eff}}$	21.19 ± 0.02	-0.76 ± 0.04	19.28 ± 0.02	-0.29 ± 0.04
Kaon-Pion	$q \cdot y$	14.52 ± 0.01	-0.25 ± 0.03	15.15 ± 0.02	-0.26 ± 0.04
	$(q \cdot y)_{\text{eff}}$	15.43 ± 0.01	-0.35 ± 0.03	15.78 ± 0.04	-0.01 ± 0.08
Slow Pion	$q \cdot y$	9.89 ± 0.01	0.06 ± 0.02	9.27 ± 0.01	-0.05 ± 0.03
	$(q \cdot y)_{\text{eff}}$	10.82 ± 0.01	0.10 ± 0.03	9.49 ± 0.03	0.03 ± 0.06
FSC	$q \cdot y$	13.74 ± 0.02	-0.17 ± 0.03	11.54 ± 0.01	-0.11 ± 0.03
	$(q \cdot y)_{\text{eff}}$	14.64 ± 0.02	-0.11 ± 0.03	11.30 ± 0.03	0.04 ± 0.06
Maximum p^*	$q \cdot y$	12.99 ± 0.01	1.40 ± 0.03	11.96 ± 0.02	0.05 ± 0.03
	$(q \cdot y)_{\text{eff}}$	12.54 ± 0.01	0.93 ± 0.03	11.82 ± 0.03	0.10 ± 0.06
Fast Hadron	$q \cdot y$	4.70 ± 0.01	1.14 ± 0.01	1.54 ± 0.01	-0.04 ± 0.01
	$(q \cdot y)_{\text{eff}}$	6.22 ± 0.01	1.46 ± 0.01	1.55 ± 0.01	-0.02 ± 0.02
Lambda	$q \cdot y$	3.05 ± 0.01	0.79 ± 0.01	1.56 ± 0.01	0.26 ± 0.01
	$(q \cdot y)_{\text{eff}}$	2.41 ± 0.01	0.62 ± 0.01	1.53 ± 0.01	0.24 ± 0.01

Table 4.4: Performance of FBDT and of MLP combiners on MC for three different cases (see text). Case II corresponds to the final version of the algorithm. All values are given in percent.

Combiner	Case	Belle II MC		Belle MC	
		$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$
FBDT	I	34.33 ± 0.05	-0.98 ± 0.10	32.03 ± 0.12	-0.31 ± 0.24
	II	37.03 ± 0.05	-1.21 ± 0.10	34.18 ± 0.06	-0.18 ± 0.11
	III	36.66 ± 0.05	-0.80 ± 0.10	34.19 ± 0.06	-0.10 ± 0.11
MLP	I	33.97 ± 0.05	-0.53 ± 0.10	31.90 ± 0.12	-0.31 ± 0.24
	II	36.90 ± 0.05	-1.87 ± 0.10	34.07 ± 0.06	-0.03 ± 0.11
	III	36.50 ± 0.05	-0.32 ± 0.10	34.11 ± 0.06	0.05 ± 0.11

4.5.5 Performance of combiners on Belle II MC and on Belle MC

The combiner level is the last step in the process of the category-based flavor tagger. At this level, a FBDT method combines the flavor information provided by the individual categories. An independent multivariate method, the MLP combiner, is used to cross-check the result of the FBDT. Both combiners receive the same input information and deliver independent output values $y = q \cdot r$ corresponding to the flavor of B_{tag}^0 times the dilution factor r .

The official Belle II MC and the Belle MC samples are generated with the CP violation parameters $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0.69)$. After training and testing the flavor tagger with both MC samples, the output $y = q \cdot r$ is distributed as in Fig. 4.25 (left). Figure 4.25 (right) shows the distributions of the product $q_{\text{MC}} \cdot q \cdot r$, where q_{MC} is the true flavor of B_{tag}^0 at the time of its decay. For the FBDT and the MLP combiners, the distributions on Belle II MC show a considerable asymmetry in the central region, and very small ones at -1 and at $+1$. On Belle MC, only very small asymmetries are present at -1 and at $+1$ for both combiners. The asymmetries in the performance of the flavor tagger will be discussed in Sect. 4.6.

The performance of the flavor tagger has an important impact on the measurement of the CP -violation parameters and of the $B^0 - \bar{B}^0$ mixing. As explained in Sect. 4.2, the time-dependent and the time-integrated probability functions $\mathcal{P}(\Delta t, q)$ and $\mathcal{P}(q)$ require as input the values of the efficiency ε , the parameter $\mu = \Delta\varepsilon/(2\varepsilon)$, the wrong-tag fraction w , and the difference Δw for each r bin. Figures 4.26(a) and 4.26(b) show the respective values for the FBDT and the MLP combiners on the official Belle II MC. The combiners have slightly different performances depending on the r bin. The largest bin-to-bin differences are observed for the parameter μ . On the other hand, the values of ε and w follow similar patterns for both combiners. The values of Δw are relatively large for the two lowest bins. This is related with the asymmetries in the central $q \cdot r$ regions. For the other r bins, the values of Δw are close to zero.

Figure 4.26(c) shows a linearity check between the true dilution r_{MC} determined using MC information and the mean $\langle r \rangle$ of the dilution provided by the combiners, separately for B^0 and \bar{B}^0 tags. The concept of the linearity check was explained in Sect. 4.5.2.

The true dilution corresponds to $r_{\text{MC}} = |1 - 2w_{\text{MC}}|$, where the wrong-tag fraction w_{MC} is calculated as in eq. (4.17). And the mean value $\langle r \rangle$ corresponds to $(\langle r_{B^0} \rangle + \langle r_{\bar{B}^0} \rangle)/2$. The linearity check verifies the equivalence on average between the output $q \cdot r$ and the product $q_{\text{MC}} \cdot r_{\text{MC}}$. In general, if the dilution $r = |y|$ provided by the combiners is linear with respect to the true dilution, the wrong-tag fraction w can be extracted from the combiner output $y = q \cdot r$ through

$$w = \frac{1 - |y|}{2}. \quad (4.20)$$

As Fig. 4.26(c) shows, for the official Belle II MC the linearity is fulfilled on average for both combiners. To reach the linear calibration for the MLP combiner on average, one of its hyperparameters, namely the cost function for the training procedure, had to be chosen to be linear. However, considering independently the results for B^0 and for \bar{B}^0 events, the linearity is not fulfilled. Thus, the output wrong-tag fraction w is equal to w_{MC} on average, but the individual fractions w_{B^0} and $w_{\bar{B}^0}$ do not correspond to w_{MC,B^0} and w_{MC,\bar{B}^0} in the two lowest r bins. This effect is also caused by the asymmetry in the central $q \cdot r$ region, which will be discussed in Sect. 4.6.

This configuration was crucial to reach a linear calibration of the MLP combiner. The concept of linear calibration will be introduced in Sect. 4.5.2.

Figures 4.27(a) and 4.27(b) show the values of ε , $\Delta\varepsilon$, w , and Δw for the FBDT and the MLP combiners on Belle MC. Also for Belle MC, the combiners have slightly different performances depending on the r bin. The values of the parameter μ have the largest differences when comparing the FBDT and the MLP combiners. On the other hand, the values of ε and w are similar, and the values of Δw are all relatively close to zero. However, the pattern of Δw value differs slightly between combiners.

For Belle MC, the linearity is fulfilled for both combiners on average, and independently for B^0 and for \bar{B}^0 events, as Fig. 4.27(c) shows. In this case eq. (4.20) is valid for w , for w_{B^0} and for $w_{\bar{B}^0}$.

Table 4.5 contains the total effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ for the FBDT and the MLP combiners, on the official Belle II and on Belle MC. The differences between the results for the FBDT and for the MLP combiner are very small. On Belle II MC the performance of both combiners is about 3% higher than on Belle MC. This is expected because the Belle II MC is simulated without background, and because Belle II has a more accurate track reconstruction and particle identification (see Sect. 3.2 and 3.3).

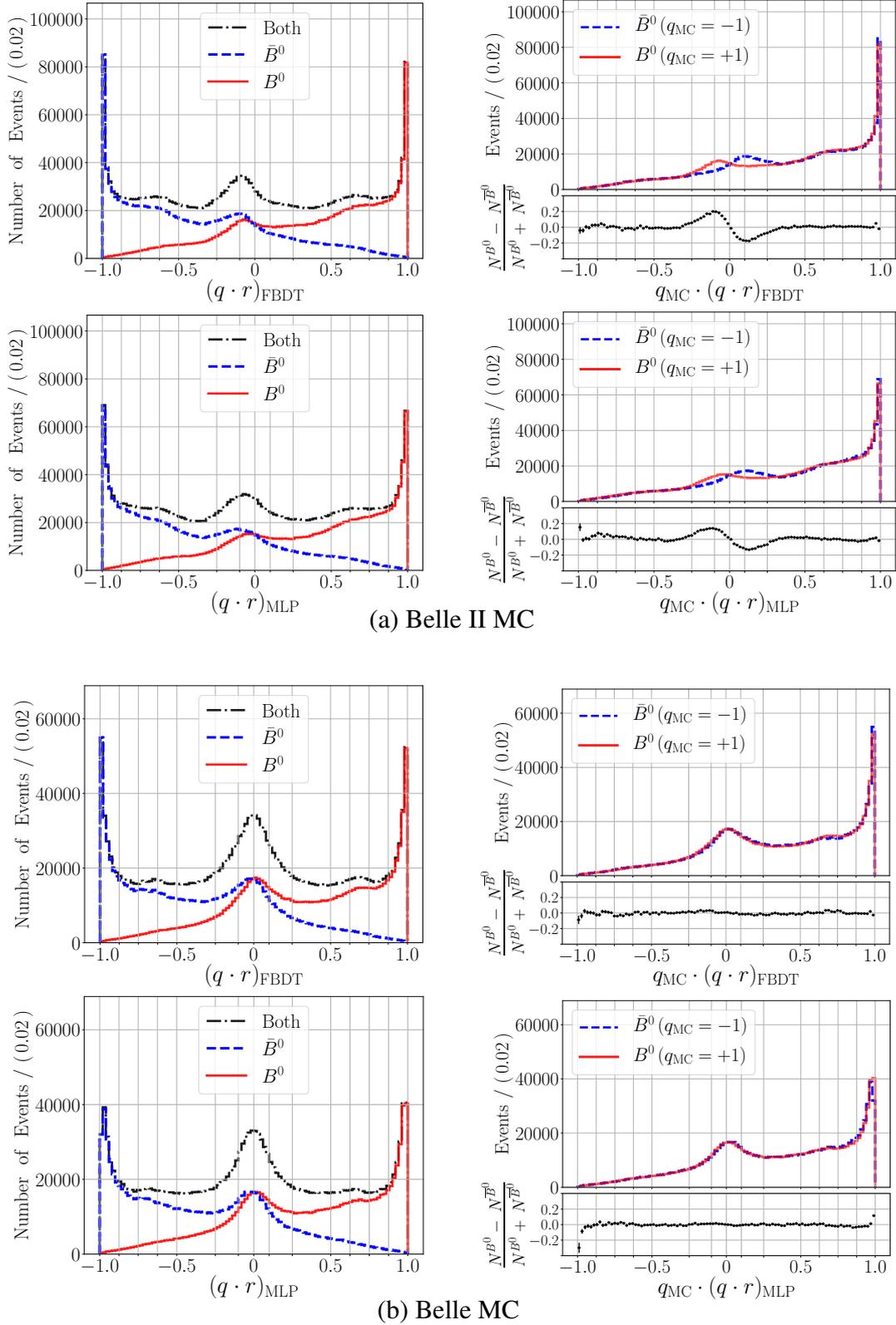
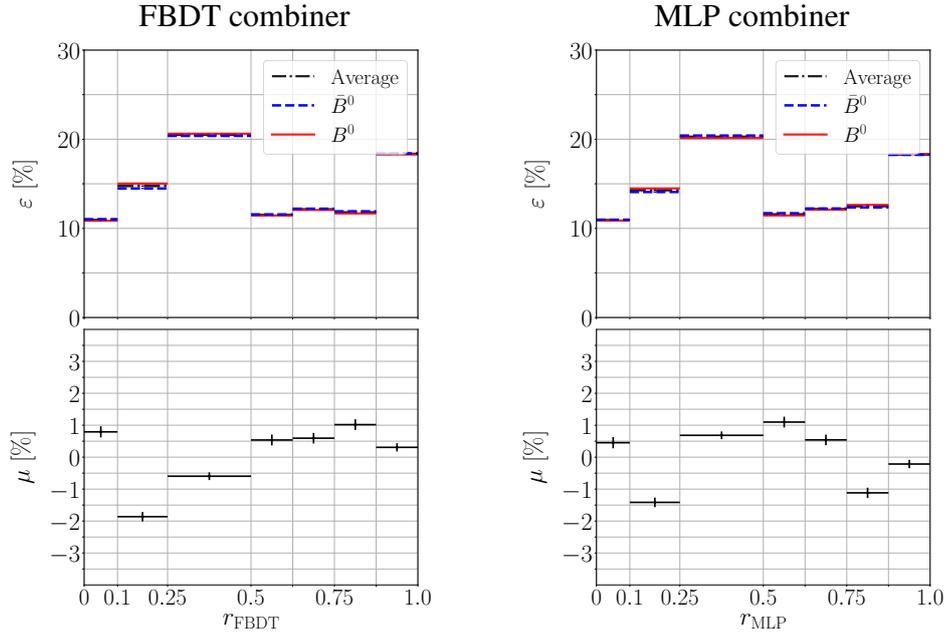
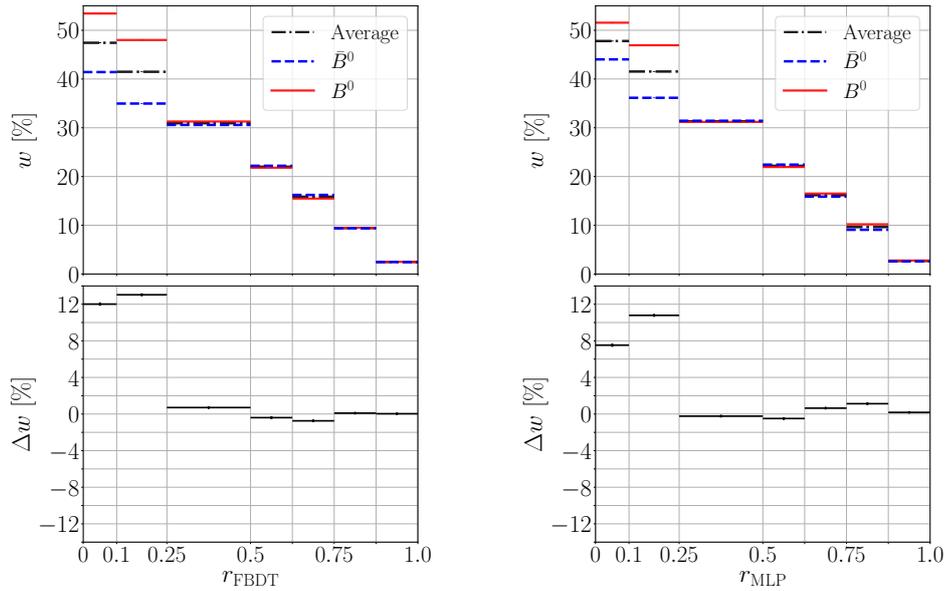


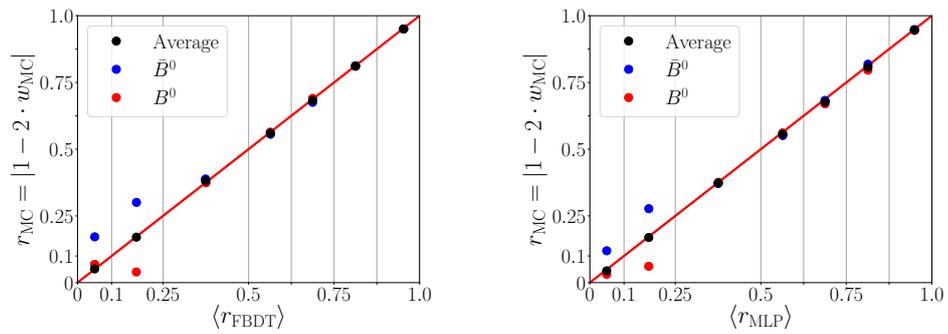
Figure 4.25: Performance of the FBDT and of the MLP combiners on (a) Belle II MC and on (b) Belle MC: (left) combiner output; (right) combiner output multiplied by the true flavor of $B_{\text{tag}}^0 q_{\text{MC}}$.



(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0}) / (\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.



(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r bin.



(c) Correlations between $r_{MC} = |1 - 2 \cdot w_{MC}|$ and $\langle |q \cdot r| \rangle$ for each r bin.

Figure 4.26: Belle II MC: performance of the FBDT (left) and of the MLP (right) combiners.

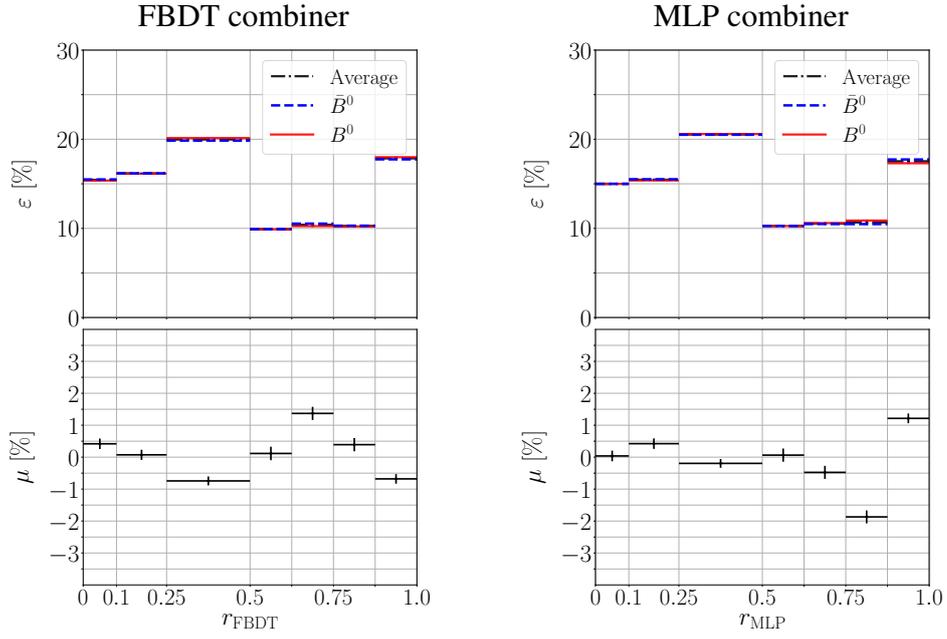
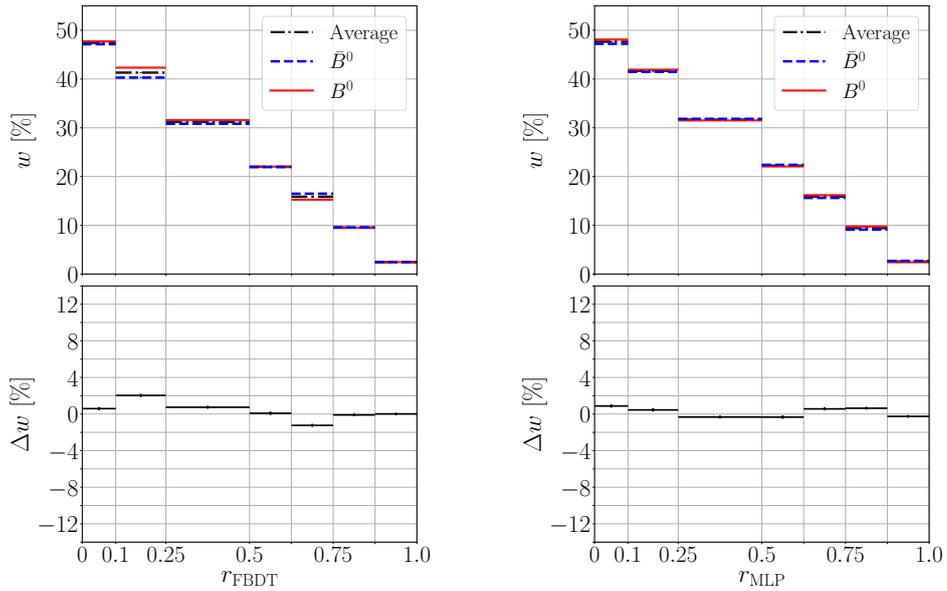
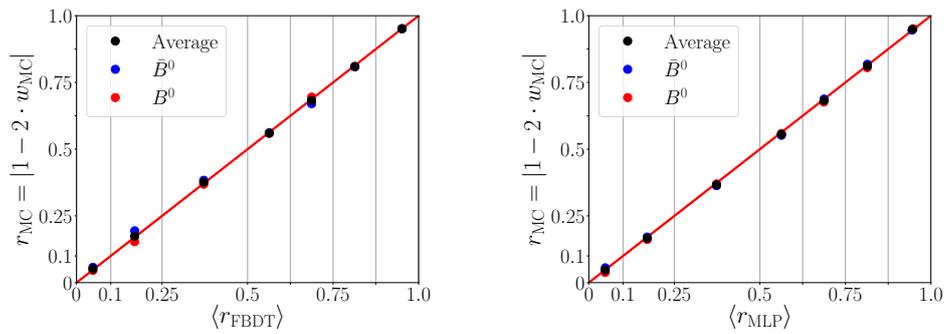
(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0}) / (\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r bin.(c) Correlations between $r_{MC} = |1 - 2 \cdot w_{MC}|$ and $\langle |q \cdot r| \rangle$ for each r bin.

Figure 4.27: Belle MC: performance of the FBDT (left) and of the MLP (right) combiners.

Table 4.5: Performance of FBDT and of MLP combiners on Belle II MC and on Belle MC. All values are given in percent. The Belle II results were obtained using the official Belle II MC.

MC type	FBDT combiner		MLP combiner	
	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$
Belle II	37.03 ± 0.05	-1.21 ± 0.10	36.90 ± 0.05	-1.87 ± 0.10
Belle	34.18 ± 0.06	-0.18 ± 0.11	34.07 ± 0.06	-0.03 ± 0.11

4.6 Understanding the asymmetries

Asymmetries in the performance of the flavor tagger might be expected due to particle-charge dependent detector performance. Also the correlation between the impact parameters and the CP -violation parameters may lead to such asymmetries. Impact parameters are correlated with the decay vertex position of B_{tag}^0 which depends on the decay time t_{tag} of B_{tag}^0 and the boost $\beta\gamma$. Figure 4.28 shows the MC distributions of $\Delta t = t_{\text{sig}} - t_{\text{tag}}$ and t_{tag} , as well as the MC distributions of $z_{\text{tag}}^{\text{dec}} - z_{\text{tag}}^{\text{prod}}$ for Belle II and for Belle. The variables $z_{\text{tag}}^{\text{dec}}$ and $z_{\text{tag}}^{\text{prod}}$ are the positions of the decay and the production vertices of B_{tag}^0 along the z axis. The CP -violation parameters correspond to $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0.69)$. The distributions are shown as functions of the flavor of B_{tag}^0 at the time of its decay. Due to the CP violation, the Δt and the t_{tag} distributions are different for events where B_{tag}^0 is a B^0 , and for events where B_{tag}^0 is a \bar{B}^0 . Therefore, the $z_{\text{tag}}^{\text{dec}} - z_{\text{tag}}^{\text{prod}}$ distribution is also different for B^0 and \bar{B}^0 events. At Belle II, the average boost $\langle\beta\gamma\rangle$ is smaller than at Belle II and, therefore, the $z_{\text{tag}}^{\text{dec}} - z_{\text{tag}}^{\text{prod}}$ distribution at Belle II decreases more steeply compared to Belle.

Figure 4.29 shows the MC distributions of Δt , t_{sig} , and t_{tag} for the three different values of \mathcal{S}_{CP} considered for the flavor tagging studies in this thesis. The Δt distributions are described by eq. (2.33) (see Sect. 2.7.3). The probability density functions describing the t_{sig} and the t_{tag} distributions can be found in App. A.3. The official Belle II MC and the Belle MC samples were generated with $\mathcal{S}_{CP} = 0.69$. The other values of \mathcal{S}_{CP} correspond to private MC samples. As Fig. 4.29 shows, depending on the CP -violation parameters, the distributions of Δt , t_{sig} , and t_{tag} are different for events where B_{tag}^0 is a B^0 , and for events where B_{tag}^0 is a \bar{B}^0 .

Since the t_{tag} distributions for B^0 and for \bar{B}^0 are different, the respective decay vertex distributions of B_{tag}^0 also depend on the value of \mathcal{S}_{CP} . Considering now the daughters of B_{tag}^0 , their production vertex corresponds to the B_{tag}^0 decay vertex. For a charged particle, the production vertex is correlated with the impact parameters of the associated reconstructed track. Therefore, the impact parameters of tracks providing flavor information is different for positively and for negatively charged particles, depending on the CP -violation parameters.

Figures 4.30 and 4.31 show the impact parameter distributions of the target particles of the Electron, the Muon, the Kaon and the Slow Pion categories for the three considered values of \mathcal{S}_{CP} . The distributions for positively and for negatively charged particles are normalized such that the area below the curves equals unity. Each plot shows in the lower part the asymmetry

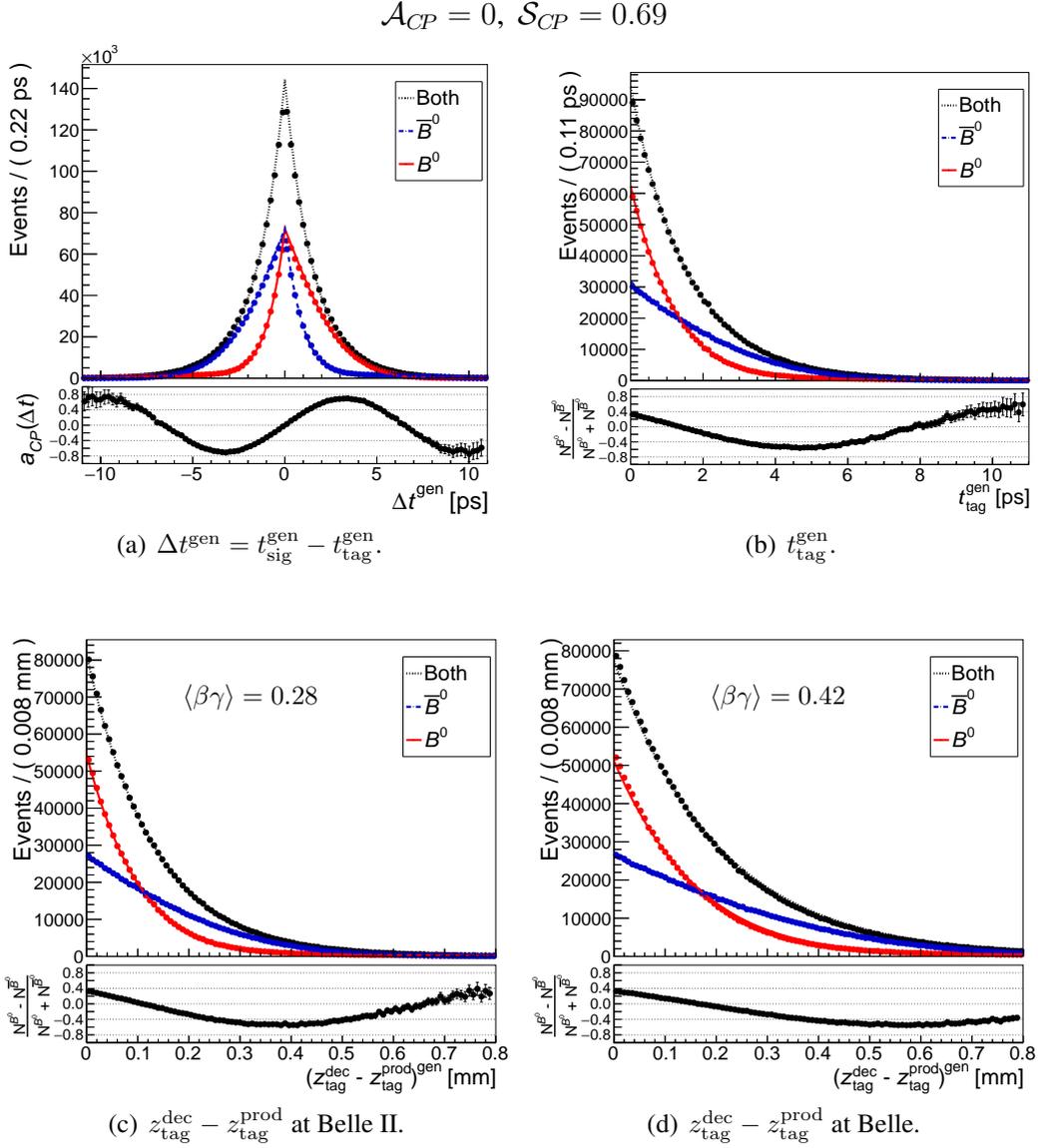


Figure 4.28: MC distributions of (a) $\Delta t^{\text{gen}} = t_{\text{sig}}^{\text{gen}} - t_{\text{tag}}^{\text{gen}}$ and (b) $t_{\text{tag}}^{\text{gen}}$, as well as the MC distributions of $z_{\text{tag}}^{\text{dec}} - z_{\text{tag}}^{\text{prod}}$ for (c) Belle II and for (d) Belle. The distributions are shown for all events, for events where B_{tag}^0 is a B^0 , and for events where B_{tag}^0 is a \bar{B}^0 .

between the fractions of positively and negatively charged particles, f^+ and f^- . The effect of the CP -violation parameters is larger in the z -axis (beam axis), and therefore it is large for ξ_0 , which is equal to $\sqrt{d_0^2 + z_0^2}$. Due to the crossing angle between electron and positron beams (see Sect. 3.1), the boost direction is not perfectly parallel to the z -axis. Therefore, the effect of the CP -violation parameters is also visible for d_0 which is the impact parameter on the plane orthogonal to the z -axis.

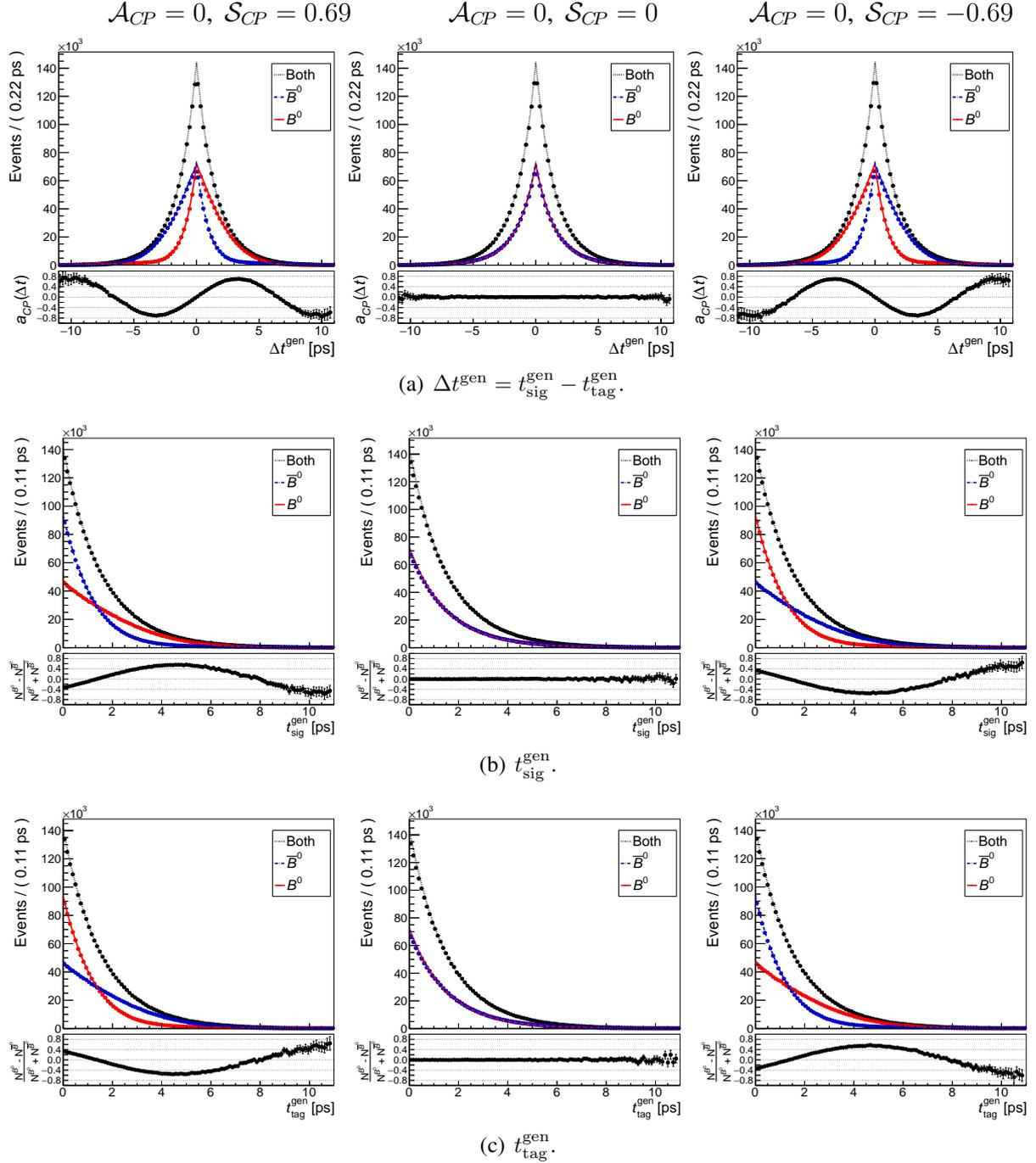


Figure 4.29: Distributions of Δt^{gen} , $t_{\text{sig}}^{\text{gen}}$ and $t_{\text{tag}}^{\text{gen}}$ for the three considered values of \mathcal{S}_{CP} .

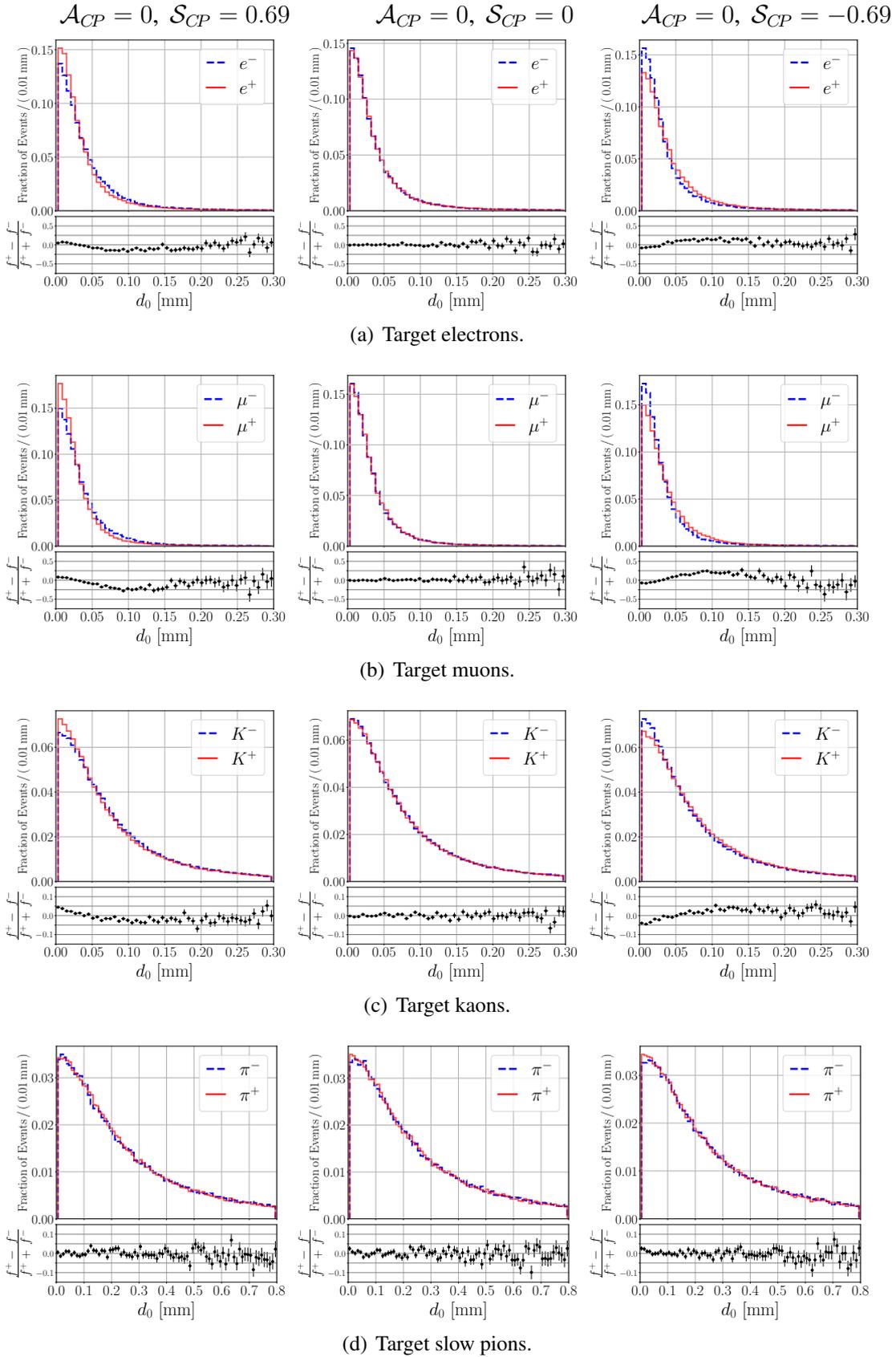


Figure 4.30: Belle II MC distributions of the impact parameter d_0 for the target particles of the Electron, the Muon, the Kaon and the Slow Pion categories.

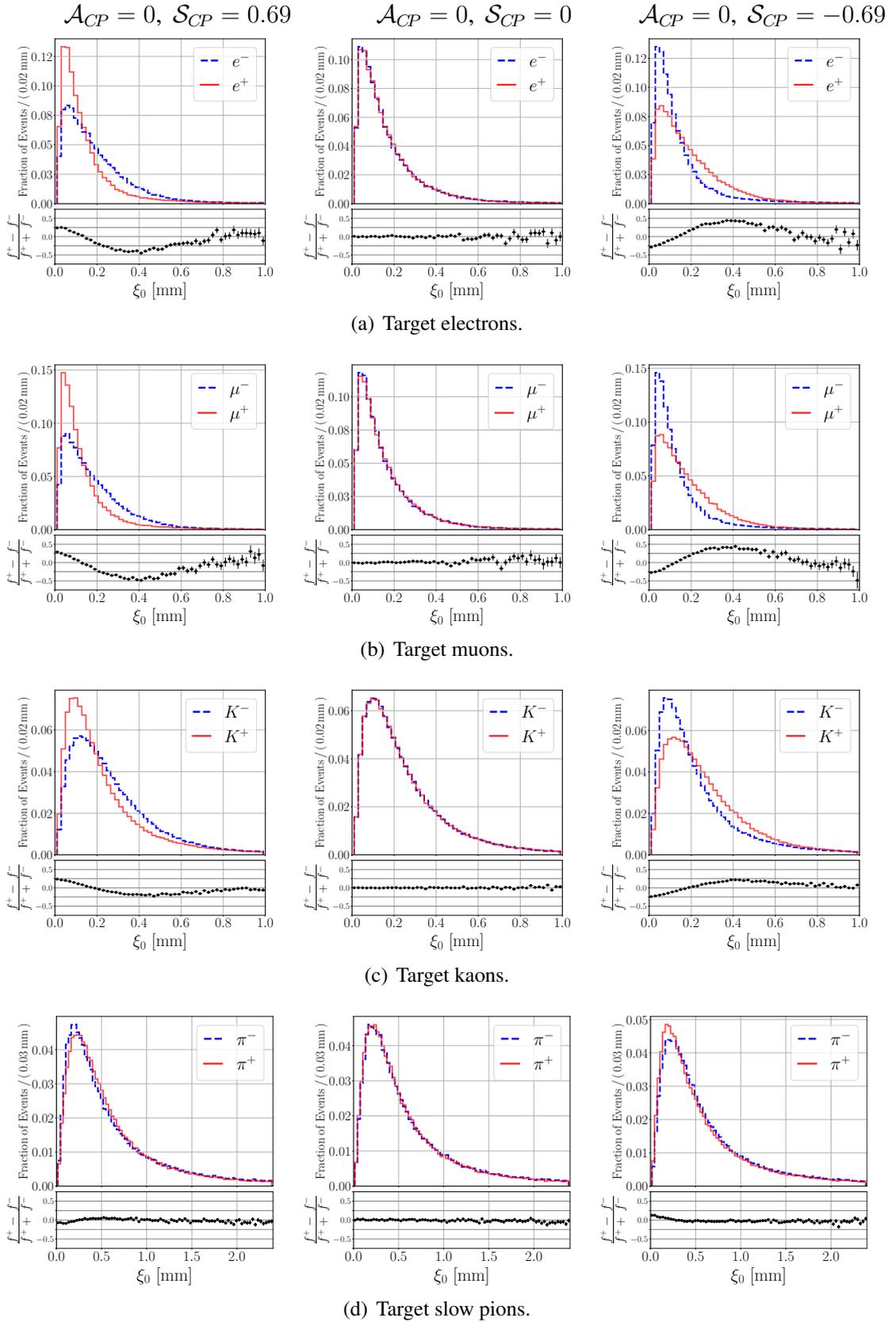


Figure 4.31: Belle II MC distributions of the impact parameter $\xi_0 = \sqrt{d_0^2 + z_0^2}$ for the target particles of the Electron, the Muon, the Kaon and the Slow Pion categories.

In the Belle II MC samples used to develop and to optimize the flavor tagger, the production vertex of the B mesons is always at the origin. This means that the decay vertex of B_{tag}^0 is distributed as in Fig. 4.28(c). In the Belle MC, the production vertex of the B mesons is distributed according to the size and the position of the interaction region at KEKB. The spread of the B_{tag}^0 production vertex blurs the asymmetry of the B_{tag}^0 decay vertex for B^0 and \bar{B}^0 events, as Fig. 4.32 shows. The respective impact parameter distributions are presented in Figs. 4.33(a) to 4.33(d). As these figures show, the asymmetries are negligible at Belle.

At SuperKEKB, the interaction region has in reality a certain size, and the asymmetry of the B_{tag}^0 decay vertex becomes blurred too. However, the size of the interaction region at SuperKEKB (about 0.5 mm in z -direction) is about twenty times smaller than at KEKB. Recent Belle II MC takes this into account. The results using these new samples show that the asymmetry of the B_{tag}^0 decay vertex will play indeed a role at Belle II. These results will be shown in chapter 5.

Impact parameters have a high power to discriminate primary from secondary particles and help thus identify particles providing clean flavor signatures such as primary leptons. However, if impact parameters are used as input variables for flavor tagging, their correlation with the CP -violation parameters may cause systematic effects in CP -violation analyses. To investigate such effects, the flavor tagger was trained for three different values of \mathcal{S}_{CP} using Belle II MC. Next, the trained flavor tagger was tested in each case for the three different values of \mathcal{S}_{CP} .

Figure 4.34 (left) shows the output of the flavor tagger after being trained using the official Belle II MC sample ($\mathcal{S}_{CP} = 0.69$). The trained flavor tagger was tested for the three different values of \mathcal{S}_{CP} . In all the cases, the central bump in the $q \cdot r$ distribution is shifted towards negative values. The total effective efficiency and the difference $\Delta\varepsilon_{\text{eff}}$ have different values for the different values of \mathcal{S}_{CP} (the values are presented in the figure). Figure 4.34 (right) shows the output of the flavor tagger multiplied by the true flavor q_{MC} of B_{tag}^0 . For all values of \mathcal{S}_{CP} , the asymmetry between \bar{B}^0 ($q_{MC} = -1$) and B^0 ($q_{MC} = 1$) is similar.

Figure 4.35 presents the corresponding values of ε , μ , w and Δw for each r bin, as well as the linearity checks. As Fig. 4.35(a) and Fig. 4.35(b) show, the values of ε , μ , w and Δw change slightly depending on the values of \mathcal{S}_{CP} . The changes are largest for the parameter μ . Figure 4.35(c) shows the linearity checks for the three different values of \mathcal{S}_{CP} . On average, the output dilution r_{FBDT} corresponds to the true dilution r_{MC} . However, this does not hold if one considers separately the events where B_{tag}^0 is a B^0 or a \bar{B}^0 : for the two lowest r bins, the value of r_{FBDT} differs considerably from r_{MC} . However, for all the three values of \mathcal{S}_{CP} , the linearity plots have a similar pattern.

The results show that the multivariate methods in the flavor tagger learn the asymmetry of the training sample. To check if this can be avoided by removing the impact parameters, the flavor tagger was trained using the official Belle II MC without d_0 and ξ_0 . In this case, the central bump in the $q \cdot r$ distribution is placed at zero for the different values of \mathcal{S}_{CP} . However, other asymmetries at different r bins become prominent. Additionally, the effective

efficiency drops by about 3 percent points from about 37.0% to about 33.5%. The results obtained without d_0 and ξ_0 can be seen in App. B.4.

Figure 4.36 shows the distributions of the $q \cdot r$ output and the product $q_{MC} \cdot q \cdot r$ for the three different \mathcal{S}_{CP} cases. In each case, the flavor tagger was trained and tested for the same value of \mathcal{S}_{CP} using d_0 and ξ_0 . The training and testing MC samples were generated with the same value of \mathcal{S}_{CP} . As Fig. 4.36 (left) shows, for $\mathcal{S}_{CP} = 0$, the central bump in the $q \cdot r$ distribution is placed at zero, while for $\mathcal{S}_{CP} = +0.69$ it is shifted towards negative values, and for $\mathcal{S}_{CP} = -0.69$ it is shifted towards positive values. Figure 4.36 (right) shows the output of the flavor tagger multiplied by q_{MC} . In the case of $\mathcal{S}_{CP} = 0$, the distributions for \overline{B}^0 and for B^0 overlap in the whole range except near -1 and near $+1$. In the case of $\mathcal{S}_{CP} = -0.69$, the distributions for \overline{B}^0 and for B^0 are asymmetric relative to each other. And in the case of $\mathcal{S}_{CP} = +0.69$, the asymmetry is flipped.

Figure 4.37 presents the corresponding values of ε , μ , w and Δw for each r bin, as well as the linearity checks. Considering first the results for $\mathcal{S}_{CP} = 0$, the distribution of μ has values within 0% and 1% except for the highest r bin, where it reaches a value of about -1.5% . The values of Δw are less than 2% for each r bin. On the other hand, the respective linearity plot shows that the average dilution as well as the dilutions for B^0 and for \overline{B}^0 are correctly calibrated.

Considering now the results for $\mathcal{S}_{CP} = 0.69$ and for $\mathcal{S}_{CP} = -0.69$, the μ distributions in Fig. 4.37(a) have different patterns and they reach values close to -2% and close to $+2\%$ in different r bins. The respective Δw distributions in Fig. 4.37(b) have values close to zero, except for the two lowest r bins, where Δw reaches about 12% for $\mathcal{S}_{CP} = 0.69$, and about -12% for $\mathcal{S}_{CP} = -0.69$. Similarly, the linearity plots in Fig. 4.37(c) show that the dilutions for B^0 and for \overline{B}^0 are not correctly calibrated in the two lowest r bins. The pattern in the linearity plot for $\mathcal{S}_{CP} = 0.69$ is flipped relative to the case $\mathcal{S}_{CP} = -0.69$.

To prevent the multivariate methods from learning the asymmetries caused by the input CP violation, the flavor tagger has to be trained for $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. Figure 4.38 shows the $q \cdot r$ output and the product $q_{MC} \cdot q \cdot r$ after training the flavor tagger with a Belle II MC sample for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. The flavor tagger is tested for the three considered values of \mathcal{S}_{CP} . In all the cases, the $q \cdot r$ distributions are symmetric and the $q_{MC} \cdot q \cdot r$ distributions for B^0 and for \overline{B}^0 overlap. Only near -1 and near $+1$, small asymmetries are present.

As Fig. 4.39 shows, if the flavor tagger is trained using a sample for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$, the values of μ still change slightly depending on the value of \mathcal{S}_{CP} , but the values of Δw remain constant close to zero. Considering the linearity plots, the dilution factor is correctly calibrated on average, and separately for B^0 and for \overline{B}^0 in all the cases.

To check if the correlation between μ and \mathcal{S}_{CP} disappears, the flavor tagger was trained and tested without d_0 and ξ_0 for $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. Even in this case, a small correlation between μ and \mathcal{S}_{CP} remains. Furthermore, other asymmetries in central r bins become prominent, and the effective efficiency reduces by about 3 percent points from 36.7% to 33.5%. The results can be seen in App. B.4.

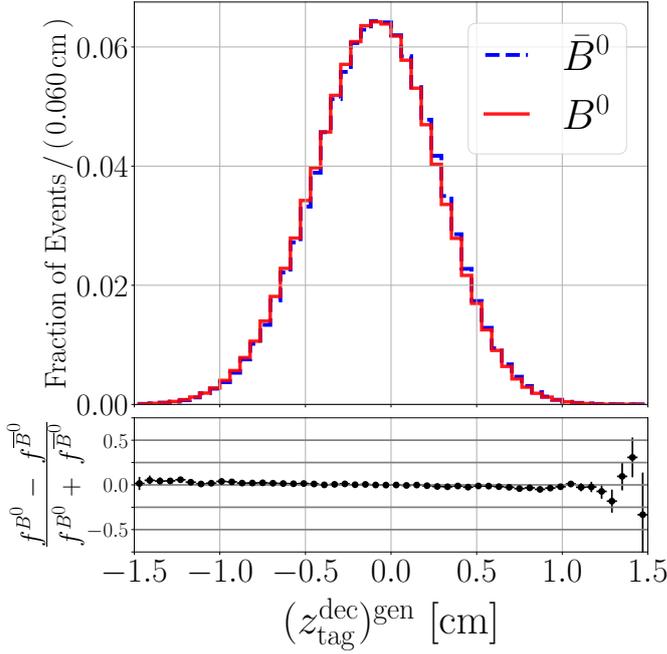
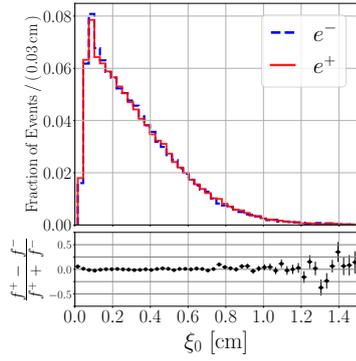
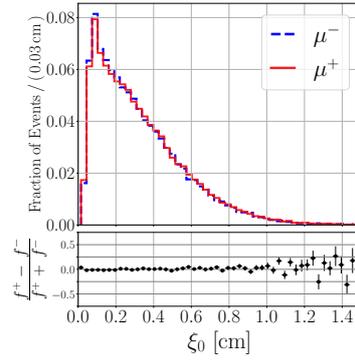


Figure 4.32: Belle MC: distribution of the decay vertex of B_{tag}^0 depending on its flavor at the time of its decay. The CP -parameters used to generate the MC are $\mathcal{A}_{CP} = 0$ and $\mathcal{S}_{CP} = 0.69$.

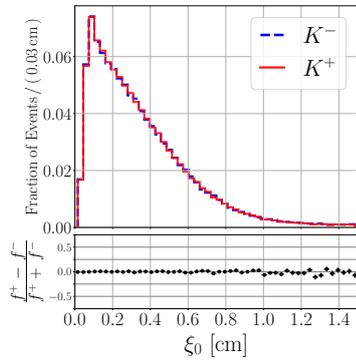
$$\mathcal{A}_{CP} = 0, \mathcal{S}_{CP} = 0.69$$



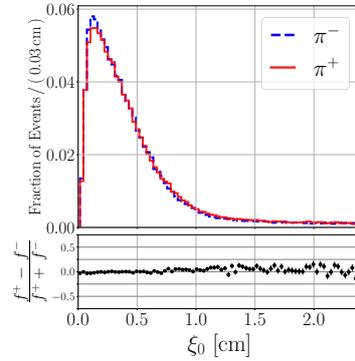
(a) ξ_0 for target electrons.



(b) ξ_0 for target muons.



(c) ξ_0 for target kaons.



(d) ξ_0 for target slow pions.

Figure 4.33: Belle MC: distributions of the impact parameter $\xi_0 = \sqrt{d_0^2 + z_0^2}$ for the target particles of (a) the Electron, (b) the Muon, (c) the Kaon and (d) the Slow Pion categories.

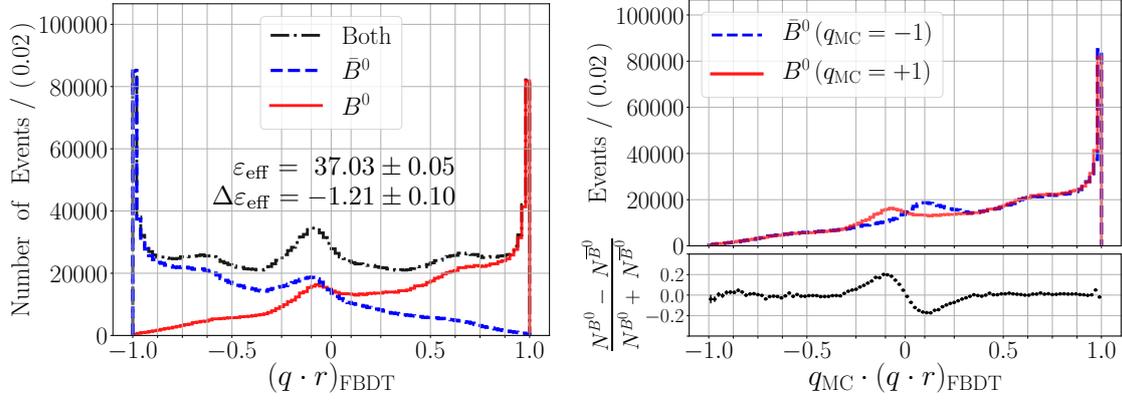
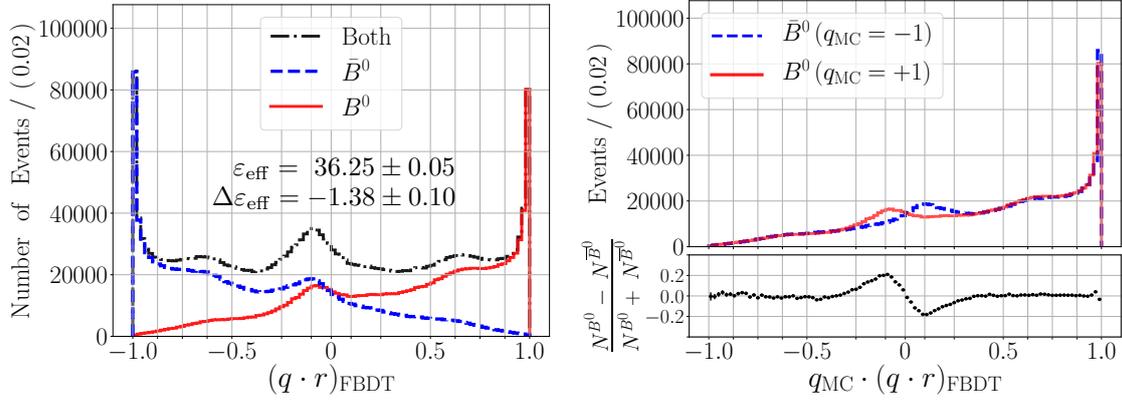
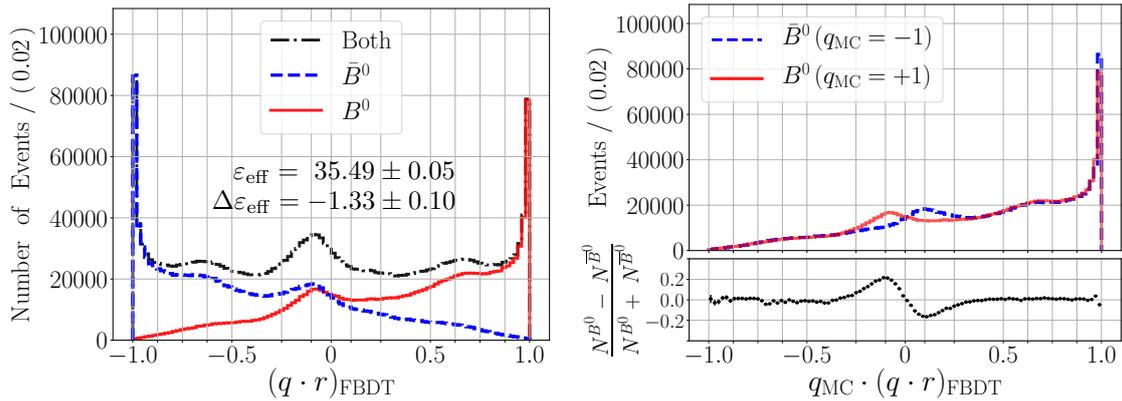
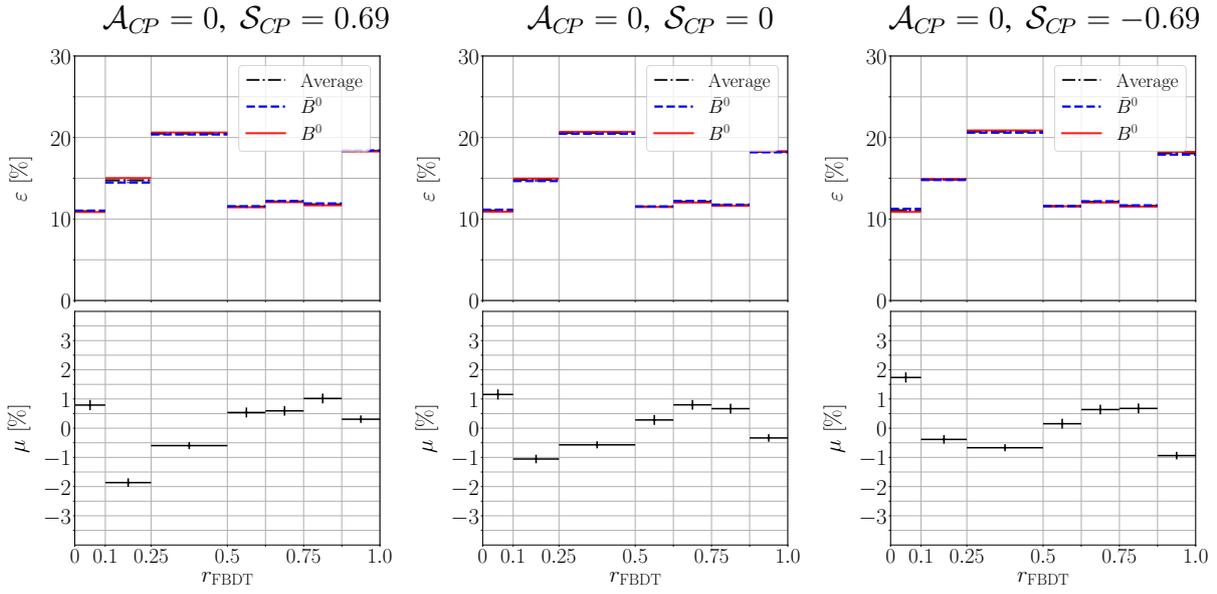
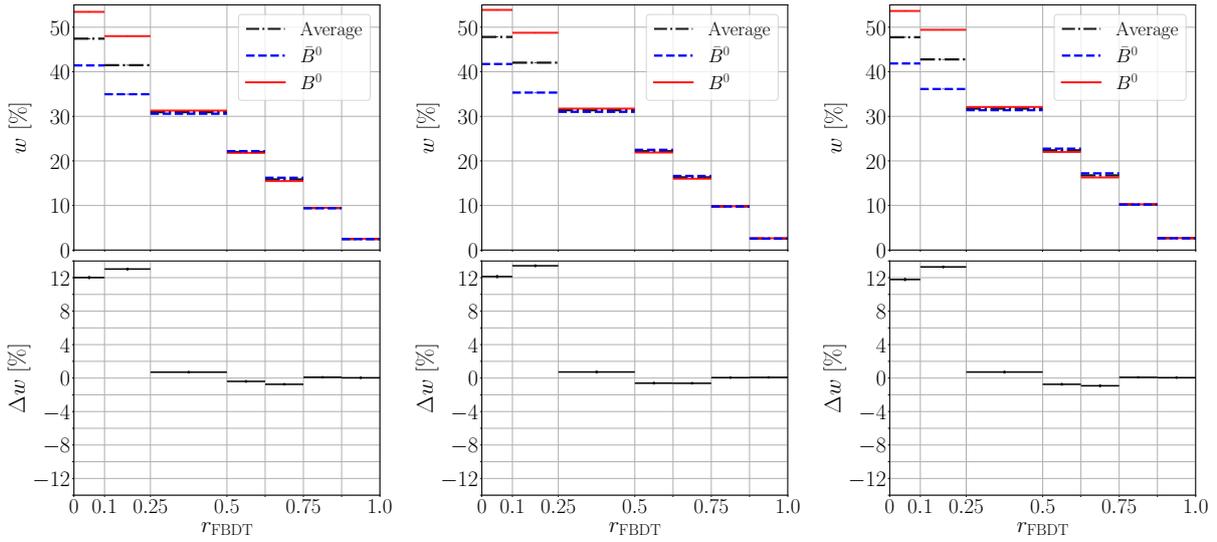
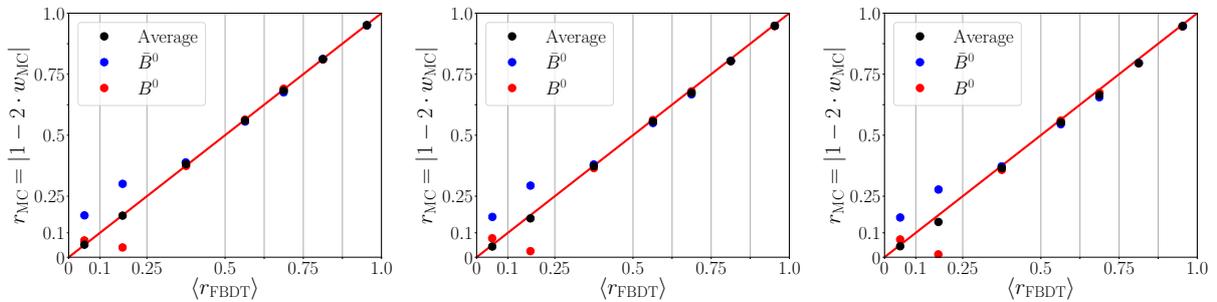
(a) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0.69$.(b) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0$.(c) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = -0.69$.

Figure 4.34: Performance of the FBBDT combiner for three values of \mathcal{S}_{CP} : (left) combiner output; (right) combiner output multiplied by the true flavor of B_{tag}^0 q_{MC} . The effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained using the official Belle II MC sample for which $\mathcal{S}_{CP} = 0.69$.

(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0})/(\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r -bin.(c) Correlations between $|r_{MC} = 1 - 2 \cdot w_{MC}|$ and $\langle |q \cdot r| \rangle$ for each r bin.Figure 4.35: Performance of the FBDT combiner for three different values of S_{CP} . The flavor tagger was trained using the official Belle II MC sample for which $S_{CP} = 0.69$.

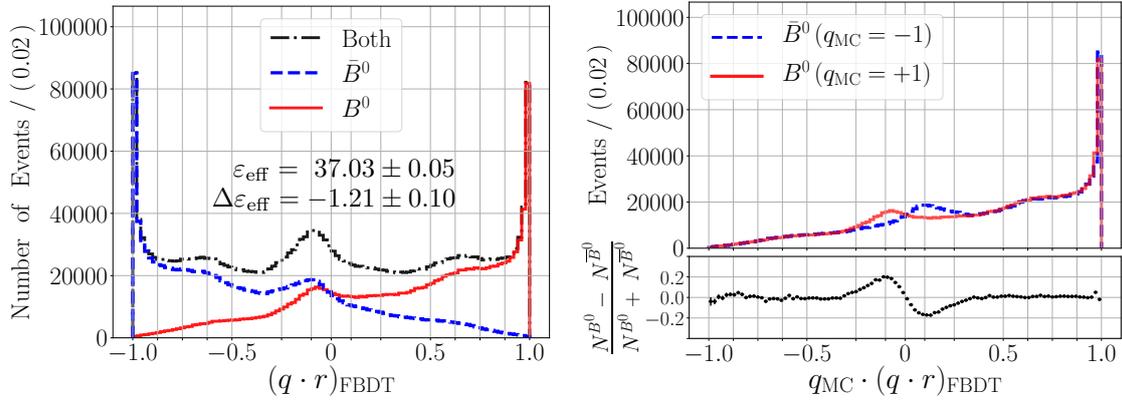
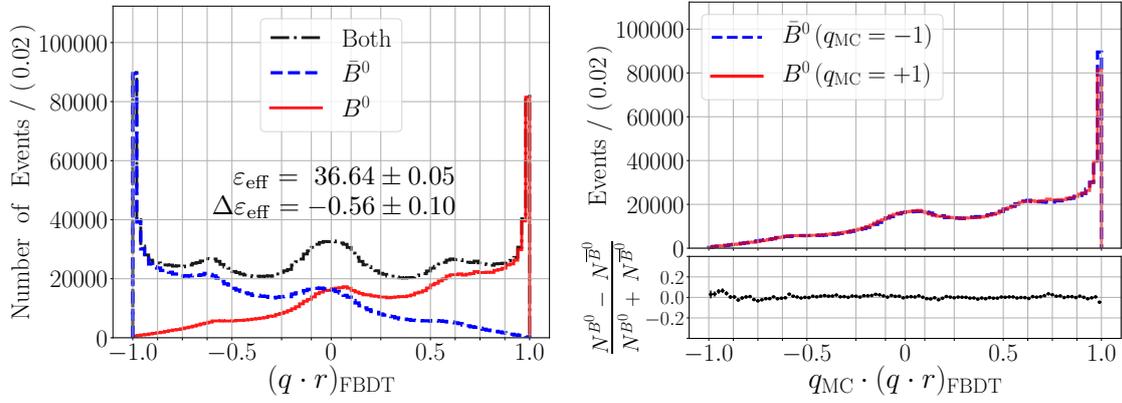
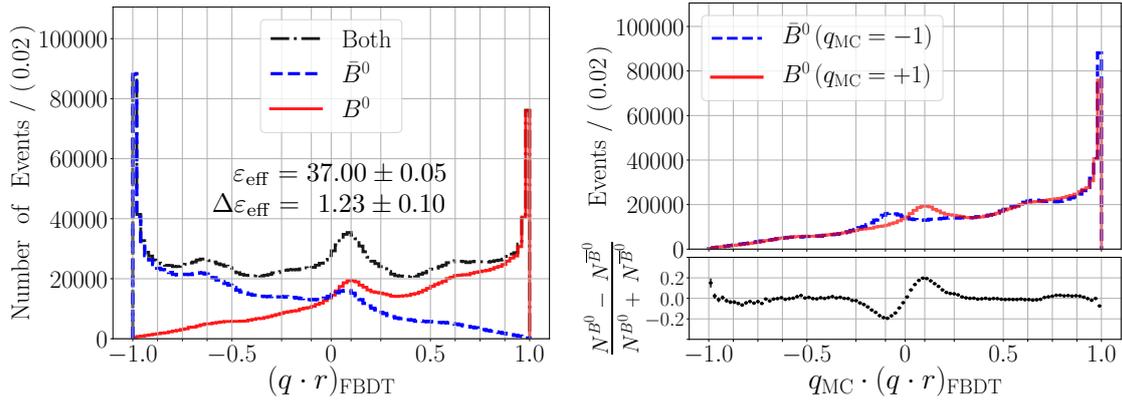
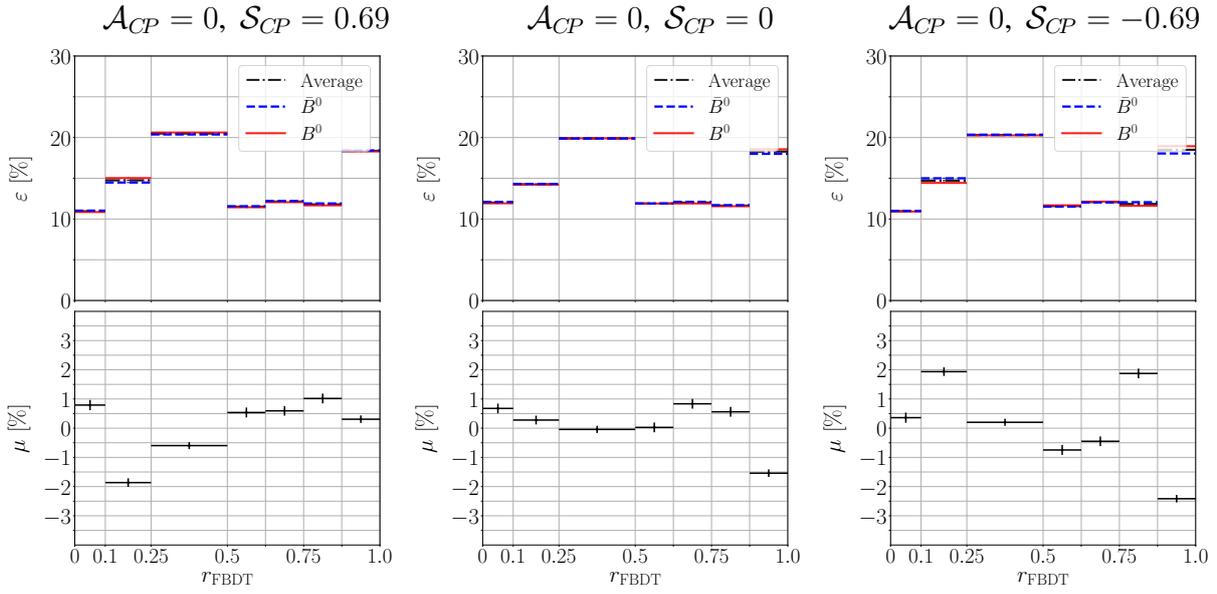
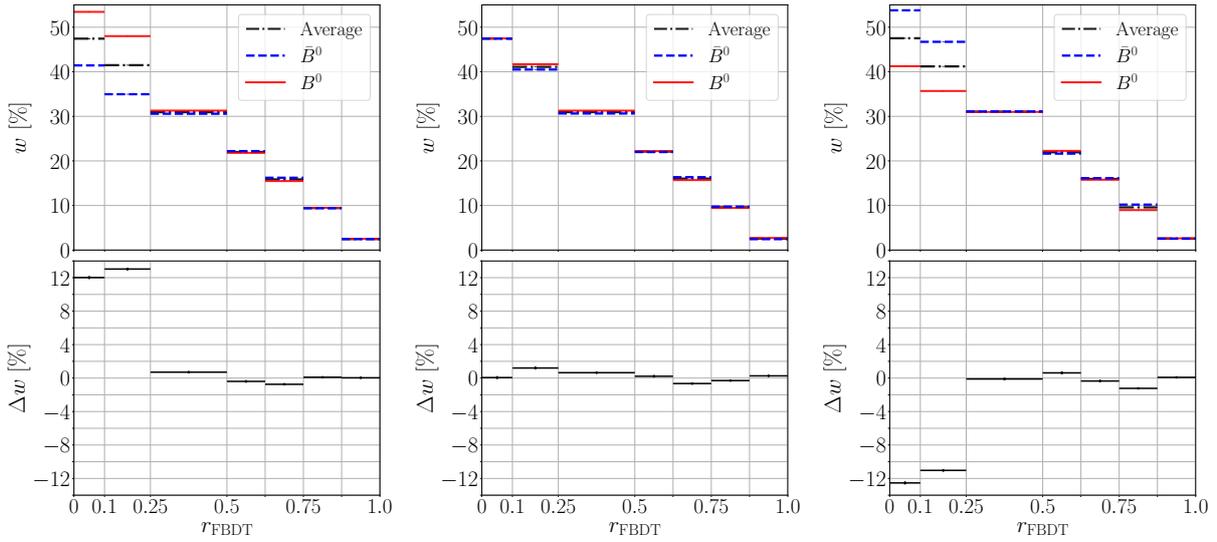
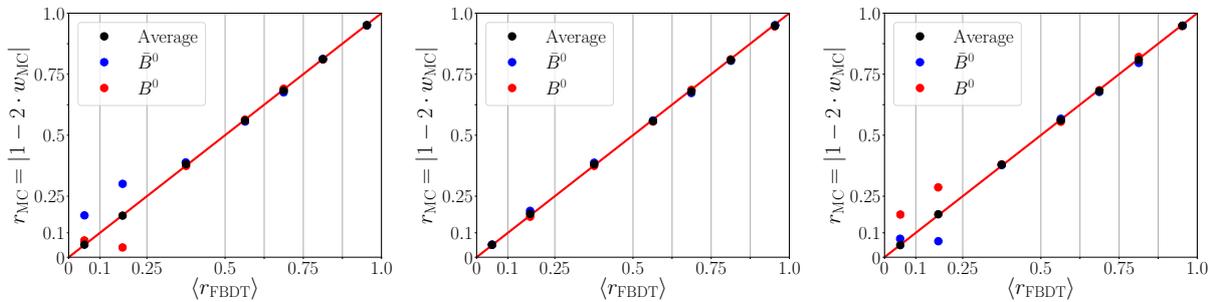
(a) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0.69$.(b) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0$.(c) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = -0.69$.

Figure 4.36: Performance of the FBDDT combiner for three values of \mathcal{S}_{CP} : (left) combiner output; (right) combiner output multiplied by the true flavor of B_{tag}^0 q_{MC} . The effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained in each case for the respective value of \mathcal{S}_{CP} .

(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0})/(\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r bin.(c) Correlations between $|r_{MC} = 1 - 2 \cdot w_{MC}|$ and $\langle |q \cdot r| \rangle$ for each r bin.Figure 4.37: Performance of the FBDT combiner for three different values of \mathcal{S}_{CP} . The flavor tagger was trained in each case for the respective value of \mathcal{S}_{CP} .

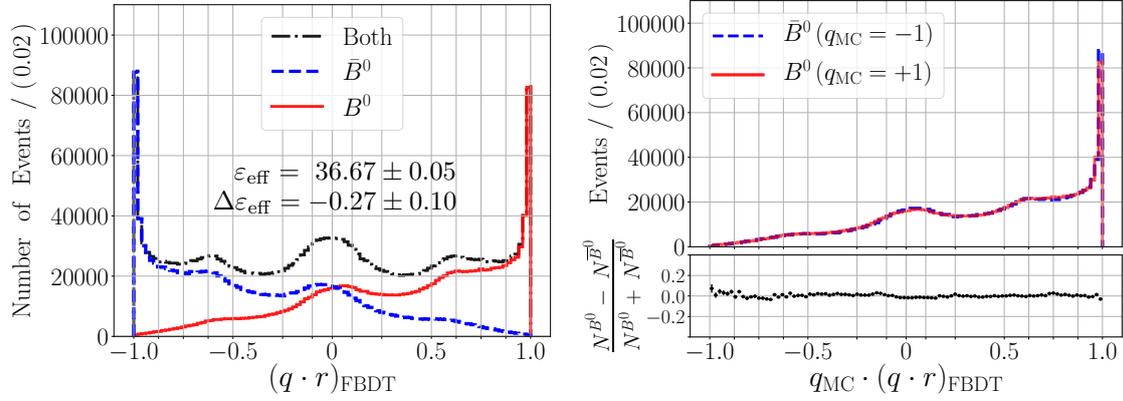
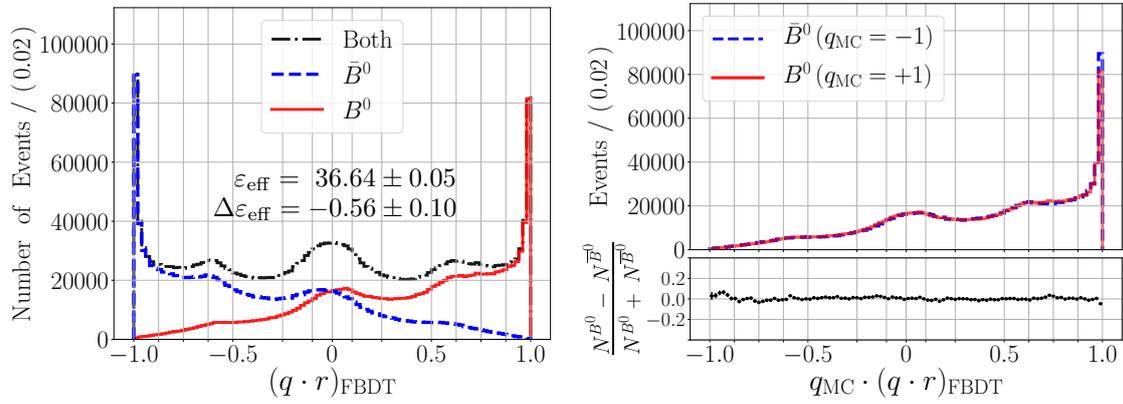
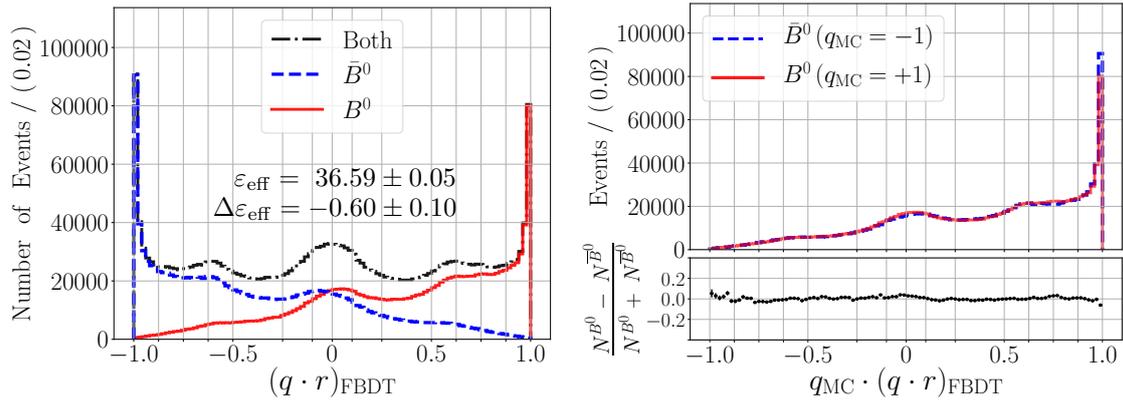
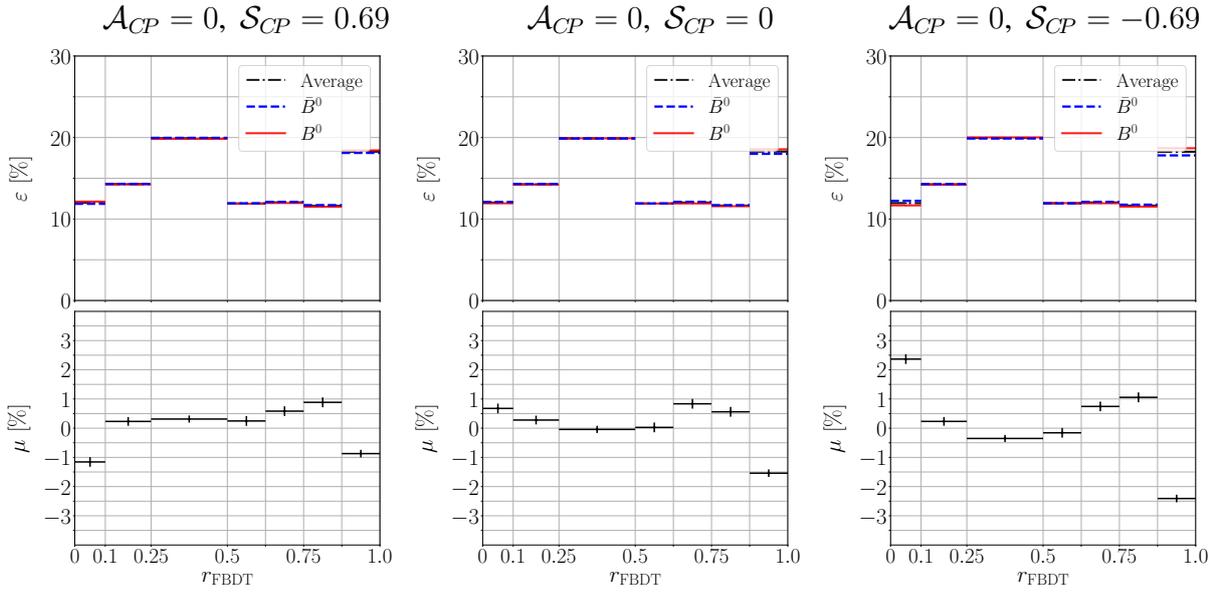
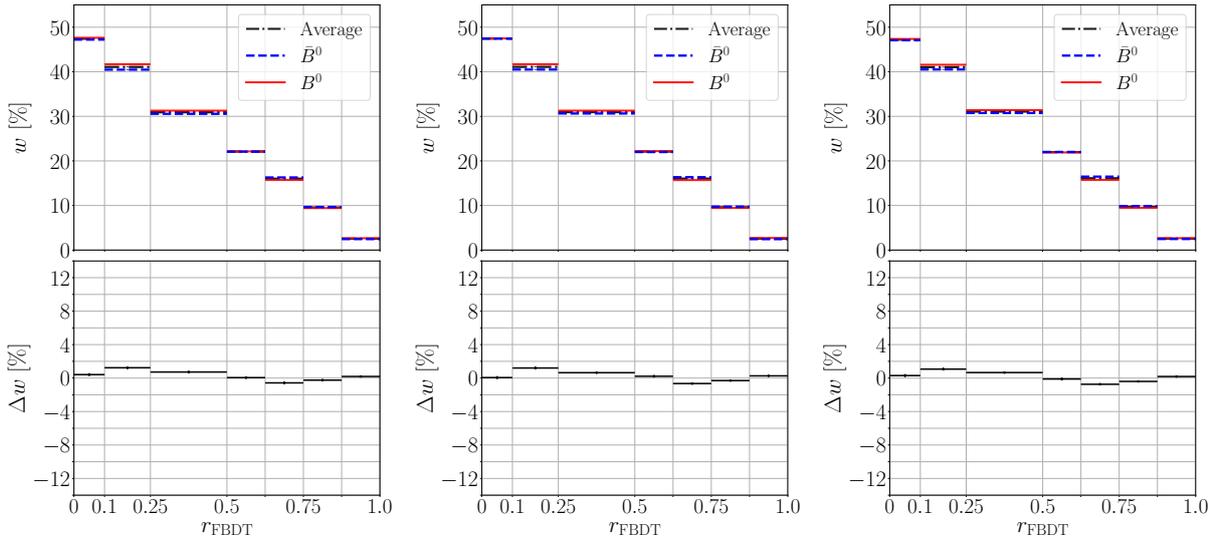
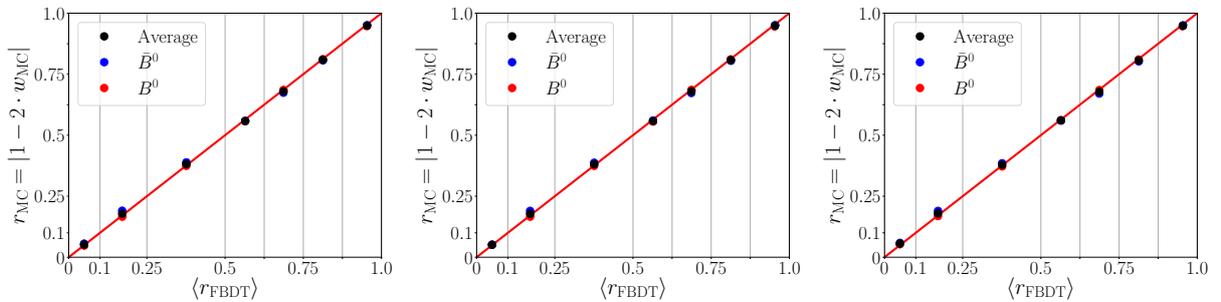
(a) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0.69$.(b) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0$.(c) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = -0.69$.

Figure 4.38: Performance of the FBDT combiner for three values of \mathcal{S}_{CP} : (left) combiner output; (right) combiner output multiplied by the true flavor of $B_{\text{tag}}^0 q_{MC}$. The effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained using a private Belle II MC sample for which $\mathcal{S}_{CP} = 0$.

(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0})/(\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r bin.(c) Correlations between $|r_{MC} = 1 - 2 \cdot w_{MC}|$ and $\langle |q \cdot r| \rangle$ for each r bin.Figure 4.39: Performance of the FBDT combiner for three different values of S_{CP} . The flavor tagger was trained using a private Belle II MC sample for which $S_{CP} = 0$.

To summarize, asymmetries in the performance of the flavor tagger are observed if the algorithm is trained with $\mathcal{S}_{CP} \neq 0$. Without impact parameters, the effect becomes smaller, but the effective efficiency decreases by about 3 percent points. Thus the solution is to keep the impact parameters, but training the flavor tagger with MC samples generated without CP -violation. However, a small dependency between μ and \mathcal{S}_{CP} remains. This correlation potentially causes systematic effects, which will have to be studied in the future. For the performance study presented in this thesis, this effect is negligible as it will be shown in Sect. 6.8.

4.7 Correlations with the generated Δt

The time difference Δt is defined as $t_{\text{sig}} - t_{\text{tag}}$. When B_{tag}^0 decays first, the value of Δt is positive; and when it decays after B_{sig}^0 , Δt is negative. Figure 4.40 shows the MC distributions of t_{tag} and of t_{sig} for positive and for negative values of Δt . The distributions were obtained from the official Belle II MC sample for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0.69)$. The distributions are shown for all events, for events where B_{tag}^0 is a B^0 and for events where B_{tag}^0 is a \bar{B}^0 . Considering first the distributions for all events (black dotted lines), the decay time of the B meson decaying first follows an exponential distribution, while the decay time of the one decaying second follows a distribution that starts at zero, reaches a maximum at $-\tau_{B^0} \cdot \log \frac{1}{2} \approx 1.06$ ps, and then decreases exponentially towards zero.

Considering now t_{tag} when B_{tag}^0 decays first (Fig. 4.40(a)), the distributions for B^0 and for \bar{B}^0 follow the same exponential decay, but they have different starting heights. This difference occurs because the CP violation on the signal side changes the fractions of B^0 and of \bar{B}^0 on the tag side. The difference in height depends on the CP -violation parameters \mathcal{A}_{CP} and \mathcal{S}_{CP} . When B_{tag}^0 decays second (Fig. 4.40(b)), the distributions for B^0 and for \bar{B}^0 have different shapes, but they start both at zero. The difference in shape depends on \mathcal{A}_{CP} and \mathcal{S}_{CP} .

Considering t_{sig} when B_{sig}^0 decays first (Fig. 4.40(d)), the distributions for B^0 and for \bar{B}^0 are exchanged in comparison with the distributions for t_{tag} when B_{tag}^0 decays first (Fig. 4.40(a)). Similarly, when B_{sig}^0 decays second, the distributions for t_{sig} (Fig. 4.40(c)) are exchanged in comparison with the distributions for t_{tag} when B_{tag}^0 decays second (Fig. 4.40(a)).

Figure 4.41 shows the MC distributions of t_{tag} and of t_{sig} obtained from the private Belle II MC sample for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. In this case, all the distributions for B^0 overlap with the respective distributions for \bar{B}^0 . However, the distributions for all the events have the same shape as when \mathcal{A}_{CP} and \mathcal{S}_{CP} are not zero. These shapes depend on the order the B mesons decayed. The probability density functions describing the MC distributions of t_{tag} and of t_{sig} can be found in App. A.3.

As shown in the previous section, impact parameters are correlated with the decay vertex of B_{tag}^0 and, in turn, also correlated with the decay time t_{tag} . Using impact parameters as input variables correlates the performance of the flavor tagger with the decay time t_{tag} and, consequently, with the time difference Δt .

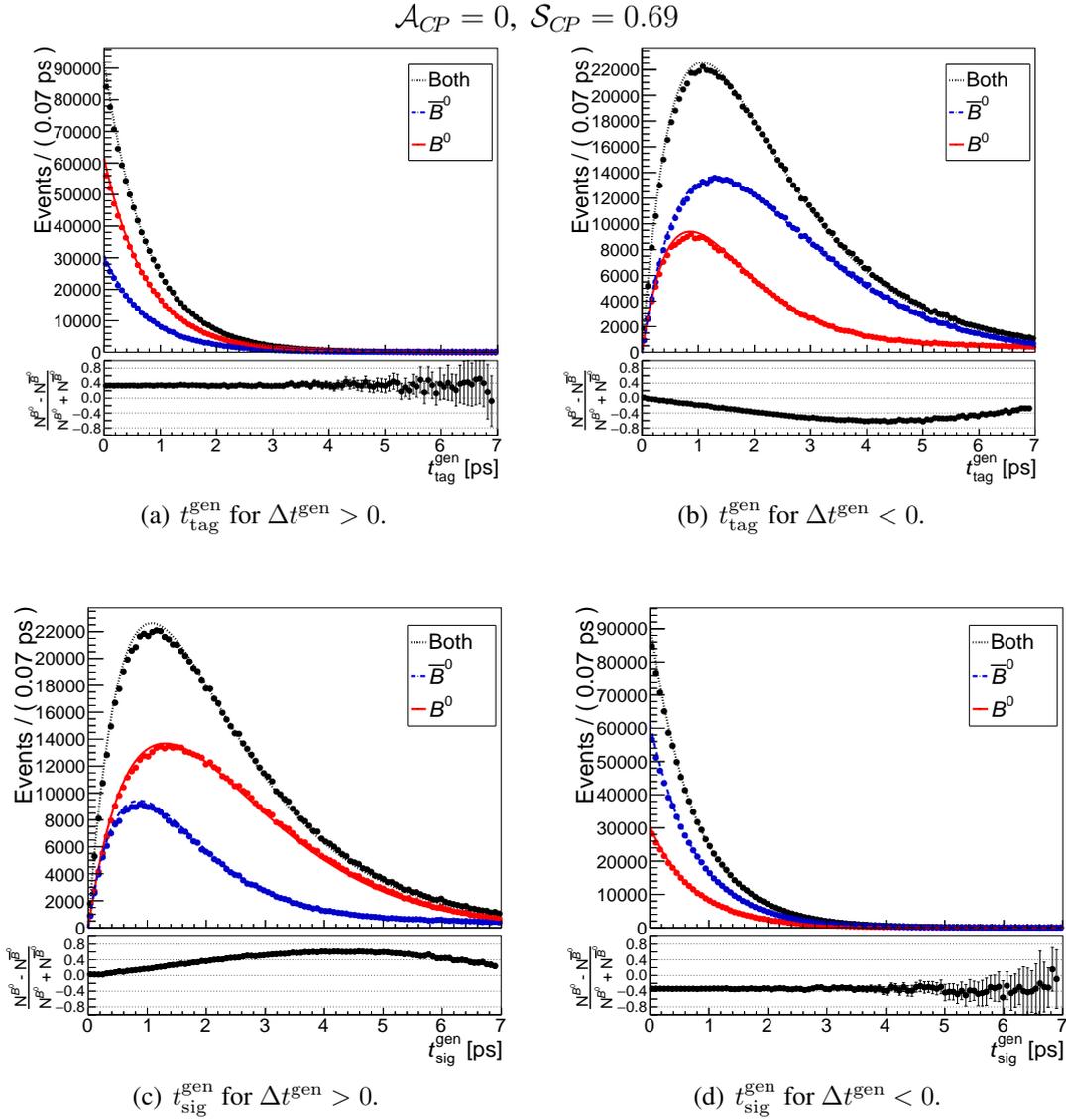


Figure 4.40: MC distributions of t_{tag} and t_{sig} for positive and for negative values of the generated Δt . The distributions are shown for all events, for events where B_{tag}^0 is a B^0 , and for events where B_{tag}^0 is a \bar{B}^0 .

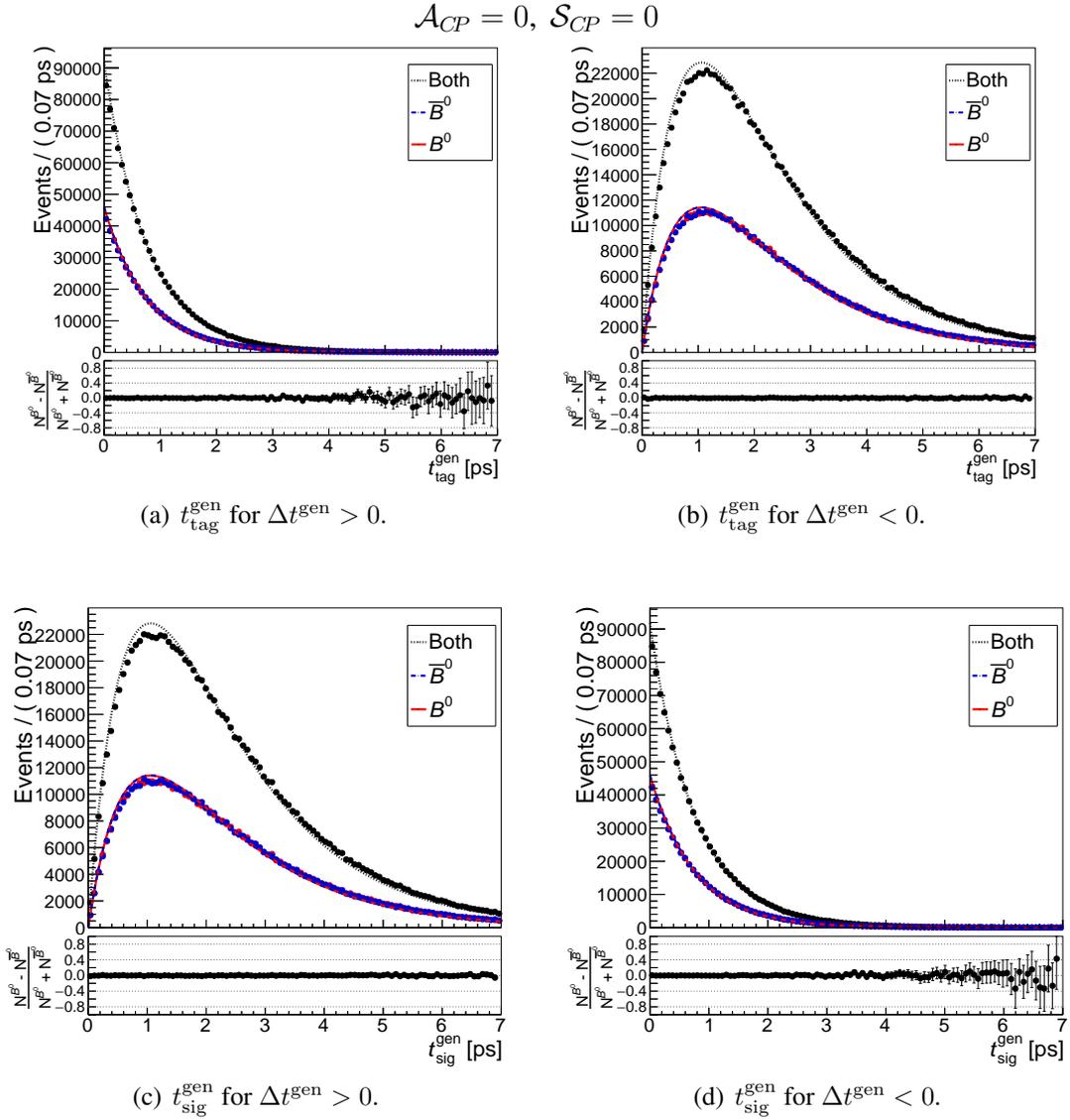


Figure 4.41: MC distributions of t_{tag} and t_{sig} for positive and for negative values of the generated Δt . The distributions are shown for all events, for events where B_{tag}^0 is a B^0 , and for events where B_{tag}^0 is a \bar{B}^0 .

Figure 4.42 shows the $q \cdot r$ output of the flavor tagger and the product $q_{\text{MC}} \cdot q \cdot r$ for positive and for negative values of Δt^{gen} . The flavor tagger was trained and tested with the Belle II MC samples for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. These samples allow the study of the correlations with the generated Δt^{gen} avoiding the correlations with the CP -violation parameters. As Fig. 4.42 shows, the distributions for positive and for negative values of Δt^{gen} are slightly different. The effective efficiency for positive values of Δt^{gen} is about 0.5% higher than for negative values of Δt^{gen} . When Δt^{gen} is negative, the $t_{\text{tag}}^{\text{gen}}$ distribution has a large population at late times. Thus, the performance of the flavor tagger is expected to be worse. Considering $\Delta \varepsilon_{\text{eff}}$, the difference between positive and negative Δt^{gen} regions is only about 0.2 percent points.

Figure 4.43 shows the respective values of ε , μ , w and Δw for each r bin, and for the positive and the negative Δt^{gen} regions. The differences between the positive and the negative Δt^{gen} regions are very small, and they are largest for the parameter μ .

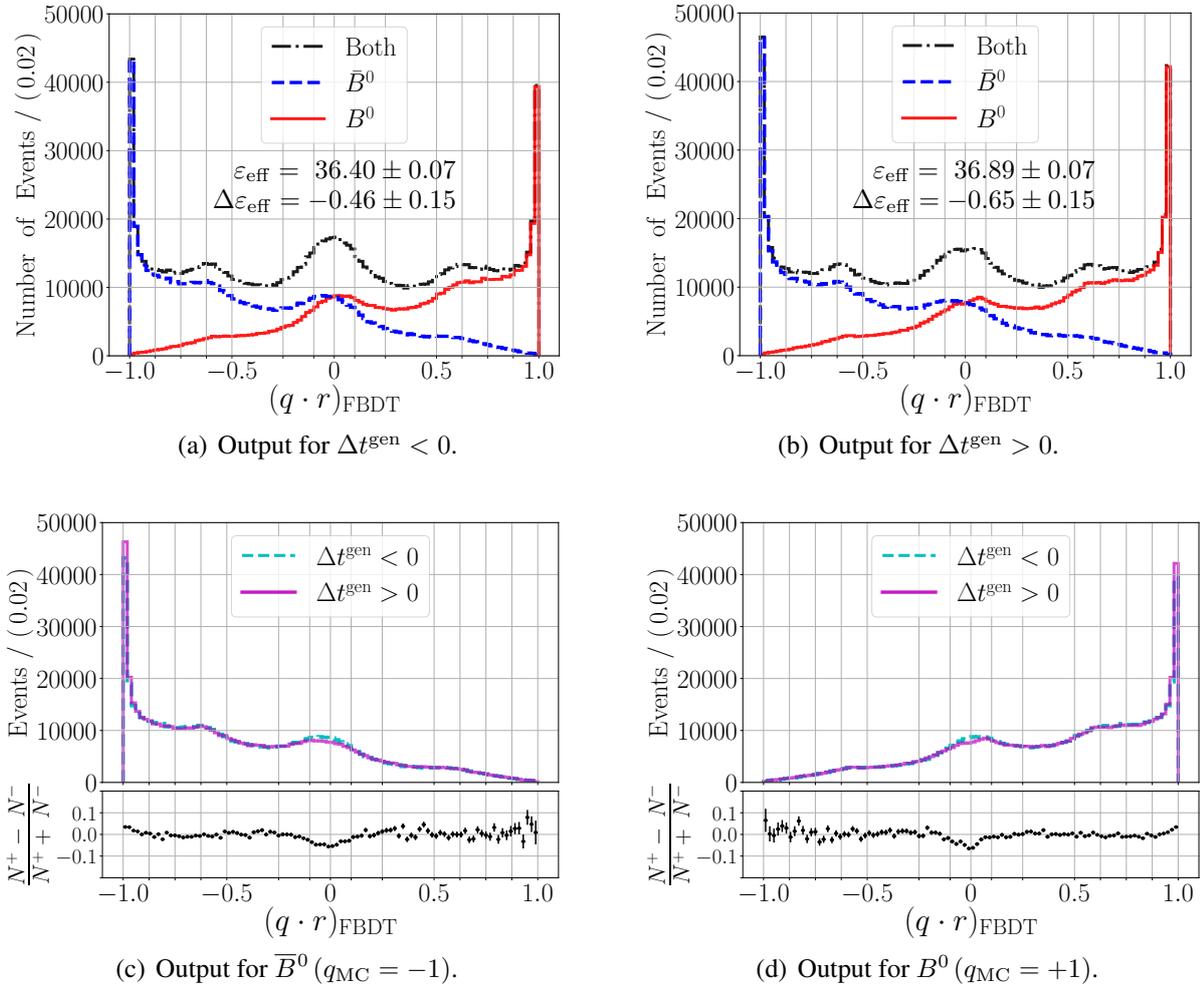
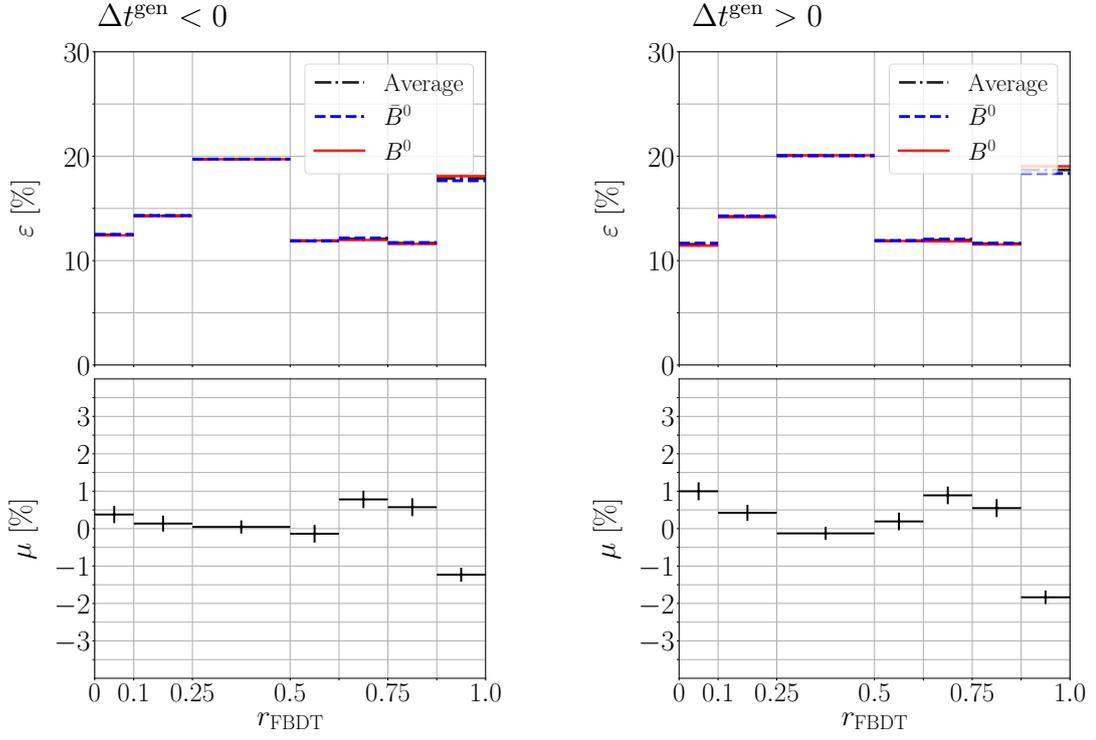
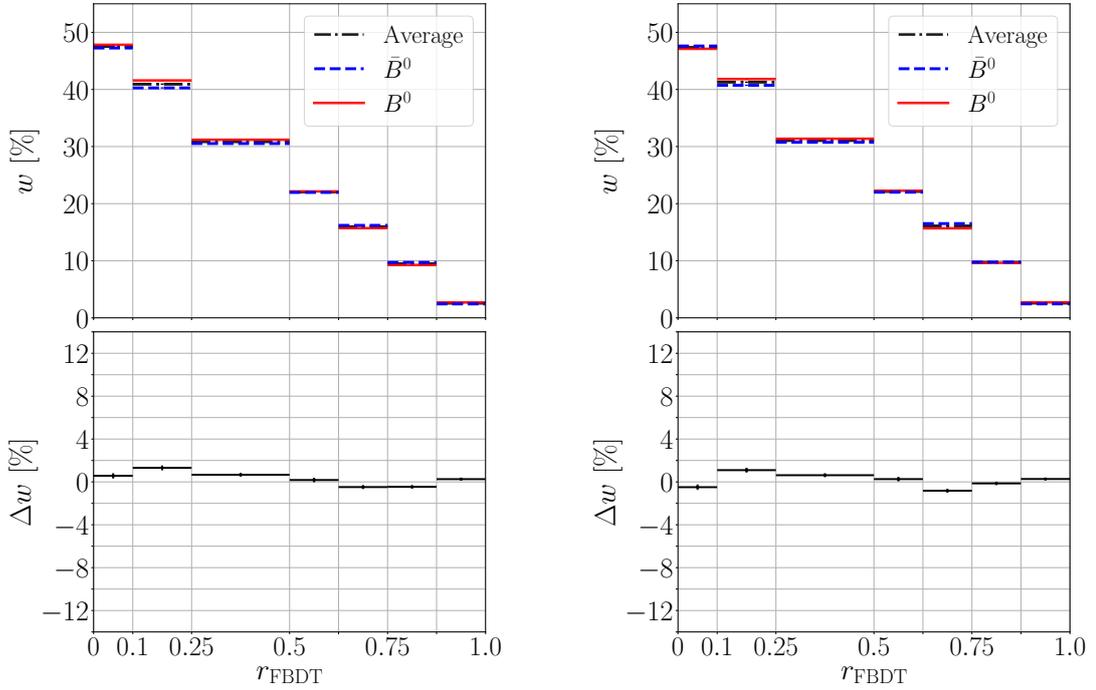


Figure 4.42: Performance of the FBDT combiner for different Δt^{gen} ranges. The effective efficiencies ε_{eff} and the differences $\Delta \varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained and tested with Belle II MC generated with $\mathcal{S}_{CP} = 0$.

(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0}) / (\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r bin.Figure 4.43: Performance of the FBDT combiner for different Δt^{gen} ranges. The flavor tagger was trained and tested with Belle II MC generated with $\mathcal{S}_{CP} = 0$.

To test if the small correlation with Δt^{gen} can be removed by avoiding the impact parameters, the flavor tagger was trained and tested without using impact parameters as input variables. In this case, the differences between the positive and the negative Δt^{gen} ranges become smaller, but they still remain, especially for the parameter μ . The results can be seen in App. B.4.

In the Belle II MC samples used for the studies in this chapter, the production vertex of the $\Upsilon(4S)$ is fixed at the origin. However, at SuperKEKB the interaction region has a size of about 0.5 mm (in z -direction). Under this condition, the correlation of the flavor tagger performance with Δt^{gen} becomes smaller as the production vertex of B_{tag}^0 is not fixed at the origin. Furthermore, the reconstructed Δt is not exactly equal to Δt^{gen} . Although the correlation between the performance of the flavor tagger and the reconstructed Δt is found to be negligible at Belle, this correlation will have to be studied for time-dependent analyses at Belle II. For the performance study of $B^0 \rightarrow \pi^0\pi^0$ presented in chapter 6, this effect is negligible because of the relatively coarse Δt resolution for events with Dalitz π^0 s and with converted photons.

4.8 Performance on Belle data

At present, the only possibility to evaluate the performance of the analysis algorithms developed for Belle II is to use Belle data. The evaluation of the category-based flavor tagger employs a set of $B^0 \bar{B}^0$ pairs, where the same decay channel used for training and for testing on MC is reconstructed in the signal side, i.e. $B_{\text{sig}}^0 \rightarrow J/\psi (\rightarrow \mu^+\mu^-) K_S^0 (\rightarrow \pi^+\pi^-)$. The data sample corresponds to the events selected for the latest Belle analysis of the decay $B^0 \rightarrow J/\psi K_S^0$ [59]. This analysis employed the full Belle data sample collected at the $\Upsilon(4S)$ resonance corresponding to 711 fb^{-1} . For the evaluations performed in this work, only the events with $J/\psi (\rightarrow \mu^+\mu^-)$ candidates were employed. The reconstruction of the J/ψ candidates at Belle is described in [131, 132].

Since the reconstruction of the decay chain was performed using BASF, the events associated with the B_{sig}^0 candidates were converted to be analyzed using BASF2 (see documentation in [127]). The converted events contain only the reconstructed high-level objects, i.e. the tracks, the clusters and the K_S^0 candidates. Therefore, the signal decay channel must be reconstructed again using BASF2 as explained in Sect. 4.5.1.

To distinguish between the signal and the continuum background, an $sPlot$ analysis is performed. The $sPlot$ is a statistical tool developed at B factories to unfold the distributions of the different components in a data sample [29, 133]. It was developed to verify that each component describes the data well in all control variables. For this, the analysis needs at least one discriminating variable for which the distribution of each component is known. The essential requirement is that the discriminating variables are uncorrelated with the studied control variables. The advantage of the $sPlot$ is that all events in the sample can contribute to the distributions of the control variables because they do not have to fulfill any requirements.

In this thesis, the beam constrained mass m_{bc} is used as discriminating variable to unfold the signal and the continuum components in the $q \cdot r$ distribution. Since m_{bc} is not correlated with the output $q \cdot r$ of the flavor tagger, the essential requirement is fulfilled.

The *sPlot* analysis requires the probability density functions (PDF) describing the distribution of the discriminating variable for the signal and the background components. The signal m_{bc} distribution is modeled as the sum of two Gaussian functions (see App. C.1):

$$\mathcal{P}_{\text{sig}}(m_{bc}) \equiv f_1 \cdot G(m_{bc}; \mu_1, \sigma_1) + (1 - f_1) \cdot G(m_{bc}; \mu_1 - \mu_2, \frac{\sigma_1}{\sigma_2}).$$

This PDF is fitted to the m_{bc} distribution of the Belle MC events used for testing (about 2 million events) to determine the parameters f_1 , μ_2 and σ_2 .

The continuum m_{bc} PDF is modeled using an ARGUS function (see App. C.2):

$$\mathcal{P}_{\text{cont}}(m_{bc}) \equiv \text{ARGUS}(m_{bc}; a, c).$$

The parameters μ_1 , σ_1 , a , and c are determined by performing a maximum likelihood fit to the Belle data. The total likelihood for the fit is given by

$$\mathcal{L} \equiv \prod_{i=1} (f_{\text{sig}} \cdot \mathcal{P}_{\text{sig}}(m_{bc}) + (1 - f_{\text{sig}}) \cdot \mathcal{P}_{\text{cont}}(m_{bc})),$$

where i extends over all events in the data sample, and f_{sig} is the fraction of signal events. After the fit, all parameters of the signal and the continuum PDFs are fixed. Figure 4.44 shows the projection of the fit result onto the m_{bc} distribution of B_{sig}^0 candidates. The fitted signal and continuum PDFs are used for the *sPlot* analysis. The analysis performs an extended maximum likelihood fit (see Sect. 6.10) to extract from the data the number of signal and continuum events N_{sig} and N_{cont} , as well as the correlation coefficient between them. Finally, it provides so-called sWeights for each event. The signal and continuum sWeights w_{sig} and w_{cont} are defined as

$$\begin{pmatrix} w_{\text{sig}} \\ w_{\text{cont}} \end{pmatrix} = \frac{\mathbf{C}^{-1}}{N_{\text{sig}} \cdot \mathcal{P}_{\text{sig}} + N_{\text{cont}} \cdot \mathcal{P}_{\text{cont}}} \cdot \begin{pmatrix} \mathcal{P}_{\text{sig}} \\ \mathcal{P}_{\text{cont}} \end{pmatrix},$$

where \mathbf{C}^{-1} is the inverse of the covariance matrix between N_{sig} and N_{cont} .

On Belle data, the measured value of the efficiency $\varepsilon = \frac{N_{\text{tag}}}{N}$, i.e. the percentage of events with at least one track on the tag side (taggable events), is 99.8%, which is equal to the previous value measured by Belle using the Belle flavor tagger [130] and is consistent with the value of 99.9% obtained using the Belle II flavor tagger on Belle MC and on Belle II MC.

To evaluate the performance of the flavor tagger on data, two histograms of $q \cdot r$, one for the signal and one for the continuum component, are filled by weighting the events with the respective sWeights. Each histogram is normalized such that the integral below its curve is equal to one. After this, the $q \cdot r$ distribution obtained on the testing Belle MC sample is

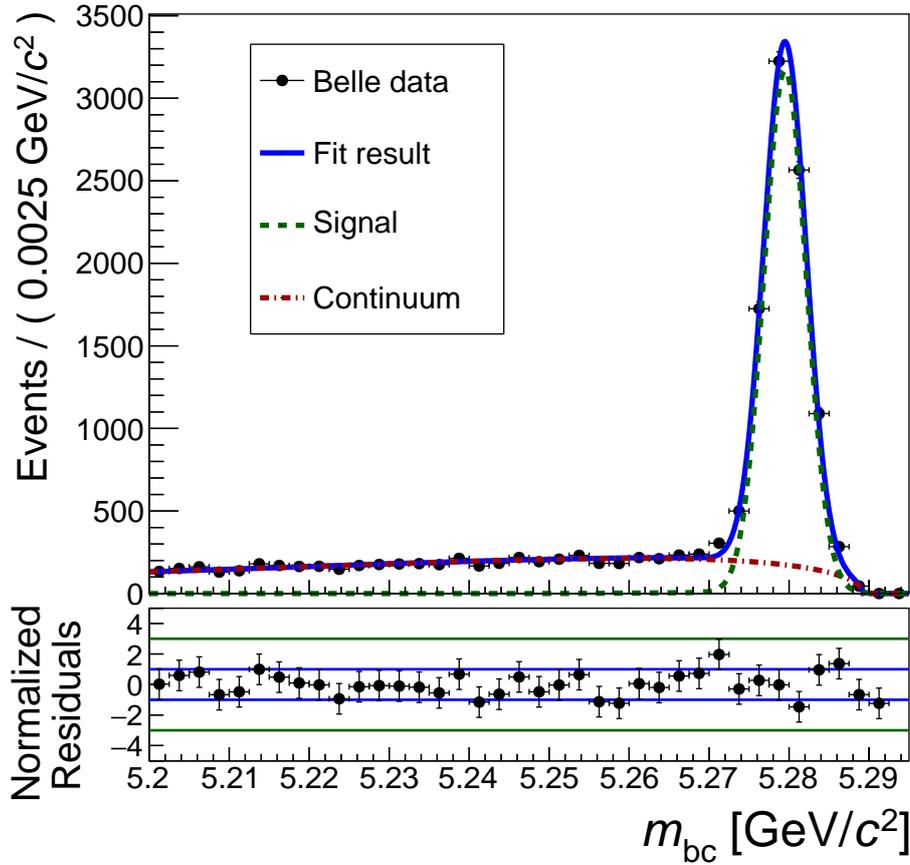


Figure 4.44: Fit projection of the total m_{bc} PDF for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow J/\psi K_S^0$ using Belle data. The number of signal events determined with the $sPlot$ analysis is 8508 ± 98 , and of continuum events is 6332 ± 86 . The given uncertainties are only statistical.

normalized and compared with the signal histogram on data. Figure 4.45 shows the normalized $q \cdot r$ distributions of the signal component on Belle data and on Belle MC for the FBDT and for the MLP combiners. Within the uncertainties, the shapes of the distributions for Belle data and Belle MC show good agreement. Figure 4.46 shows the normalized $q \cdot r$ distributions for the continuum component on Belle data for FBDT and for MLP combiners. Both distributions have a large peak at zero and decrease in population towards the boundaries ($r = |y| = 1$), where the signal distributions peak (Fig. 4.45).

The discrimination power of $q \cdot r$ between signal and continuum components strongly motivates the use of the flavor tagger output as additional variable for continuum suppression in analyses where flavor tagging is not required, e.g. in analysis of charged B^\pm decays. An advantage is also that $q \cdot r$ is not correlated with the kinematic variables m_{bc} and ΔE . The correlation coefficient (eq. (4.11)) between $q \cdot r$ and m_{bc} is -0.2% for both the FBDT and the MLP combiners. The correlation coefficient between $q \cdot r$ and ΔE is 0.2% for the FBDT combiner, and less than 0.1% for the MLP combiner. These coefficients were calculated using Belle data. On Belle MC, all aforementioned coefficients are smaller than 0.1% .

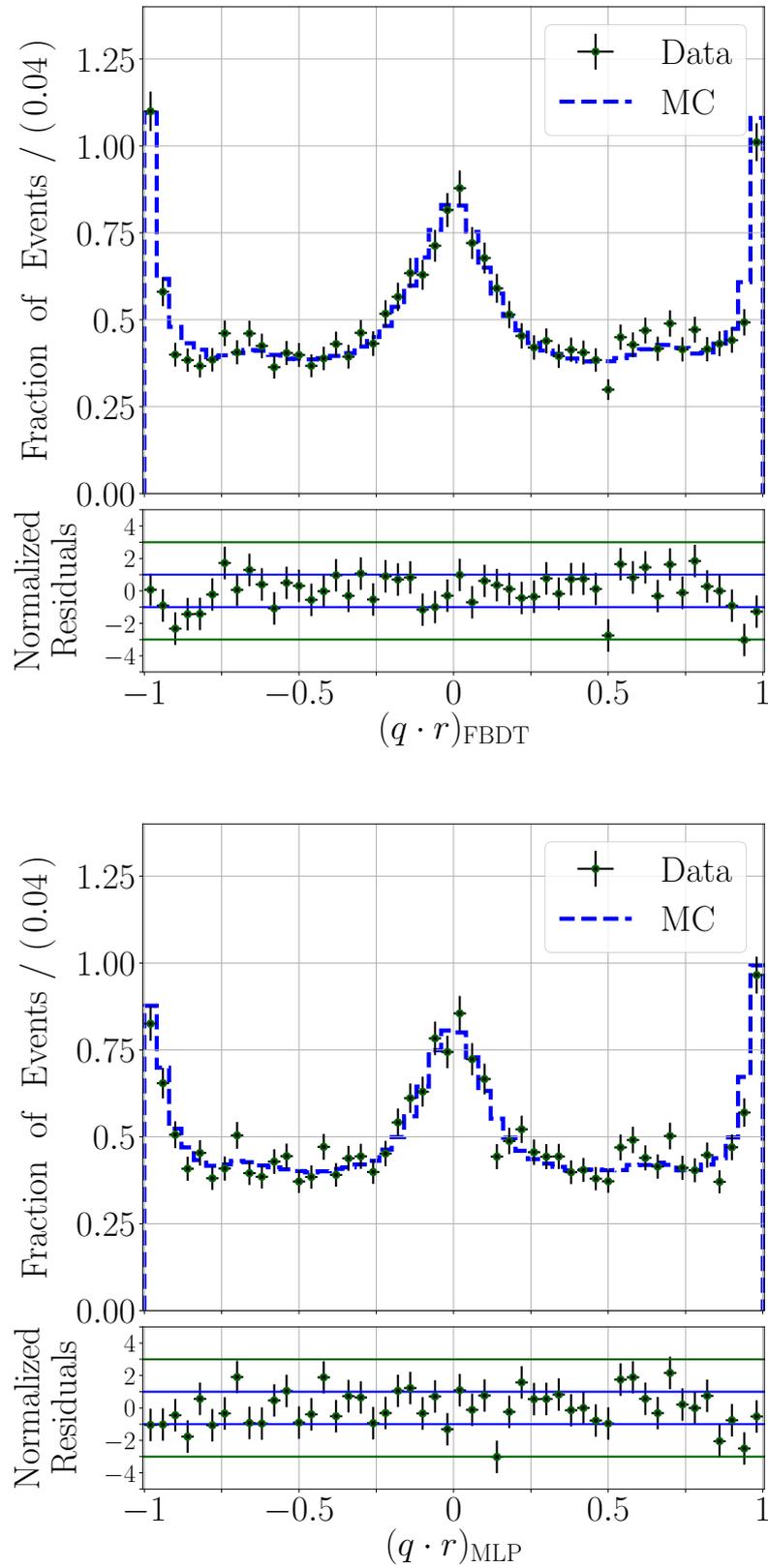


Figure 4.45: Normalized $q \cdot r$ distributions on Belle data and on Belle MC for FBDT and for MLP combiners. Only the contribution from the signal component of the Belle data is shown.

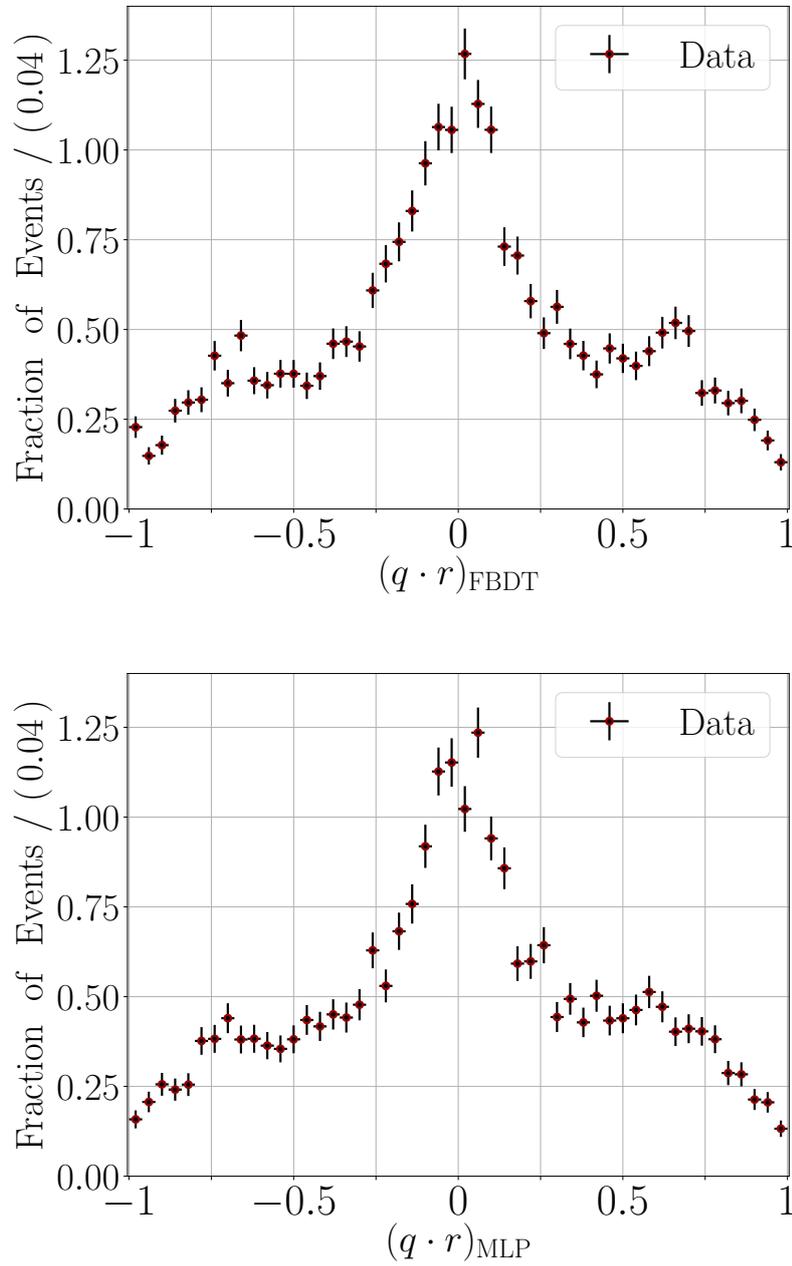


Figure 4.46: Normalized $q \cdot r$ distributions on Belle data for FBDT and MLP combiners. Only the contribution from the continuum component is shown.

Since the $q \cdot r$ distributions for Belle data and Belle MC show very good agreement, and since the flavor tagger is correctly calibrated on Belle MC (see Fig. 4.27(c)), the mean values $\langle r \rangle_i$ obtained from the combiner output are assumed to be equal to the true dilution factors $r_{\text{data},i}$ to estimate the effective efficiency ε_{eff} (eq. (4.14)) on Belle data. In other words, the mean wrong-tag fractions w_i calculated from the flavor tagger output are assumed to be correct. As will be discussed in Sect. 4.14, the wrong-tag fractions w_i and the differences Δw_i have to be determined on data in order to employ the Belle II flavor tagger in future data analyses.

The effective efficiency estimated for each r bin on Belle data is presented in Tab. 4.6 for the FBDT and the MLP combiners. Only statistical uncertainties are considered. The FBDT combiner achieves a total effective efficiency ε_{eff} of $(33.5 \pm 0.5)\%$, identical to the MLP combiner $(33.4 \pm 0.5)\%$.

Table 4.6: Performance of the FBDT and the MLP combiners on Belle data. All values are given in percent considering only statistical uncertainties.

FBDT Combiner			
r - Interval	$\varepsilon_i \pm \delta\varepsilon_i$	$w_i \pm \delta w_i$	$\varepsilon_{\text{eff},i} \pm \delta\varepsilon_{\text{eff},i}$
0.000 – 0.100	15.2 ± 0.4	47.64 ± 0.04	0.034 ± 0.001
0.100 – 0.250	16.5 ± 0.4	41.50 ± 0.06	0.477 ± 0.013
0.250 – 0.500	20.3 ± 0.4	31.39 ± 0.09	2.803 ± 0.066
0.500 – 0.625	10.0 ± 0.3	21.74 ± 0.06	3.204 ± 0.105
0.625 – 0.750	11.1 ± 0.4	15.63 ± 0.06	5.222 ± 0.162
0.750 – 0.875	10.3 ± 0.3	9.40 ± 0.06	6.807 ± 0.218
0.875 – 1.000	16.4 ± 0.4	2.33 ± 0.05	14.863 ± 0.366
Total	$\varepsilon_{\text{eff}} = \sum_i \varepsilon_i \cdot \langle 1 - 2w_i \rangle^2 = 33.5 \pm 0.5$		

MLP Combiner			
r - Interval	$\varepsilon_i \pm \delta\varepsilon_i$	$w_i \pm \delta w_i$	$\varepsilon_{\text{eff},i} \pm \delta\varepsilon_{\text{eff},i}$
0.000 – 0.100	15.13 ± 0.4	47.58 ± 0.04	0.035 ± 0.001
0.100 – 0.250	15.48 ± 0.4	41.47 ± 0.06	0.450 ± 0.013
0.250 – 0.500	21.05 ± 0.4	31.34 ± 0.09	2.927 ± 0.067
0.500 – 0.625	10.63 ± 0.3	21.77 ± 0.06	3.382 ± 0.107
0.625 – 0.750	10.77 ± 0.3	15.64 ± 0.06	5.076 ± 0.159
0.750 – 0.875	10.43 ± 0.3	9.43 ± 0.06	6.857 ± 0.219
0.875 – 1.000	16.34 ± 0.4	2.69 ± 0.05	14.602 ± 0.360
Total	$\varepsilon_{\text{eff}} = \sum_i \varepsilon_i \cdot \langle 1 - 2w_i \rangle^2 = 33.4 \pm 0.5$		

4.9 Test of robustness

As presented in the previous section, the shapes of the $q \cdot r$ output distributions of the FBDT and the MLP combiners on Belle MC show very good agreement with the shapes of the distributions on Belle data. The effective efficiencies achieved on Belle data and on Belle MC are also compatible within the statistical uncertainties.

To reach a high effective efficiency, the Belle II flavor tagger is optimized to maximally exploit the available information on the B_{tag}^0 side. For this reason, a large number of input variables is used for the different flavor tagging categories. The number of input variables is 108. Some of them are used several times for different categories, yielding in total 220 input variables. This leads to the question if the compatibility observed using Belle MC and Belle data is not inherent to the choice of the input variables. In order to answer this, the flavor tagger was trained with Belle MC using different subsets of input variables, and tested with both, Belle MC and Belle data. Obviously, the effective efficiency is expected to change by changing the number of input variables, but the compatibility between the results obtained with data and with MC should remain constant, if the simulation of input variables is correct. In this way, potential differences between the results on MC and on data are evaluated by altering the input information.

For this study, the input variables used at the event level are divided into four subsets according to their type, the ranking provided by the FBDT method and the correlations among them. As explained in Sect. 4.3, input variables can be classified into two types, PID and kinematic variables. Each of these types can be split again into two types, “high” and “low”, according to the FBDT ranking and according to the correlations between the variables. E.g., p^* is usually ranked high within several categories, but p_t^* is ranked low as it is correlated to about 99% with p^* . Thus, p^* and p_t^* are classified into the same group. The same criterion is followed for the other variables.

Table 4.7 shows the four considered subsets of input variables for each category: PID-high, PID-low, kinematic-high and kinematic-low. The number of possible combinations of two subsets is six; and the number of possible combinations of three subsets is four; together with the four single subsets, the sum yields 14 different combinations of variables. For each combination, the flavor tagger was trained using Belle MC and tested on Belle MC and on Belle data.

For some combinations of input variables, there are certain categories that have only one single variable. For example, the Kaon-Pion category has only one kinematic-high variable. When only one input variable is available, the event-level multivariate method cannot be trained. Therefore, the tests using only kinematic-low variables and using only kinematic-high variables do not include the Kaon-Pion category; the test using only PID-low variables does not include the Kaon-Pion, the Maximum p^* , the FSC and the Lambda category; and the tests using PID-high and using all PID variables do not include the Maximum p^* , the FSC, and the Lambda categories. Here, there is a small detail: the PID information for protons is not

available in the Belle MC or in the Belle data samples converted to be analyzed with BASF2. Therefore, for Belle MC and Belle data, the Lambda category does not use PID information and thus n_{unique} and n_{input} (see definition in Sect. 4.4) have two variable less compared to the case for Belle II MC.

Table 4.8 presents the effective efficiencies reached by the FBDT and the MLP combiners on Belle data and Belle MC for each combination of input variables. The results are ordered according to the increase in effective efficiency obtained by the FBDT combiner on data. For a better visualization of the results, Fig. 4.47 shows the effective efficiencies of the FBDT and the MLP combiners for each combination of input variables. The results are shown in the same order as in Tab. 4.8. For both combiners, the highest efficiency on data and on MC is reached using all input variables as in the final algorithm. For each combination of input variables, the efficiencies of both combiners on Belle MC agree with those on Belle data within the uncertainties.

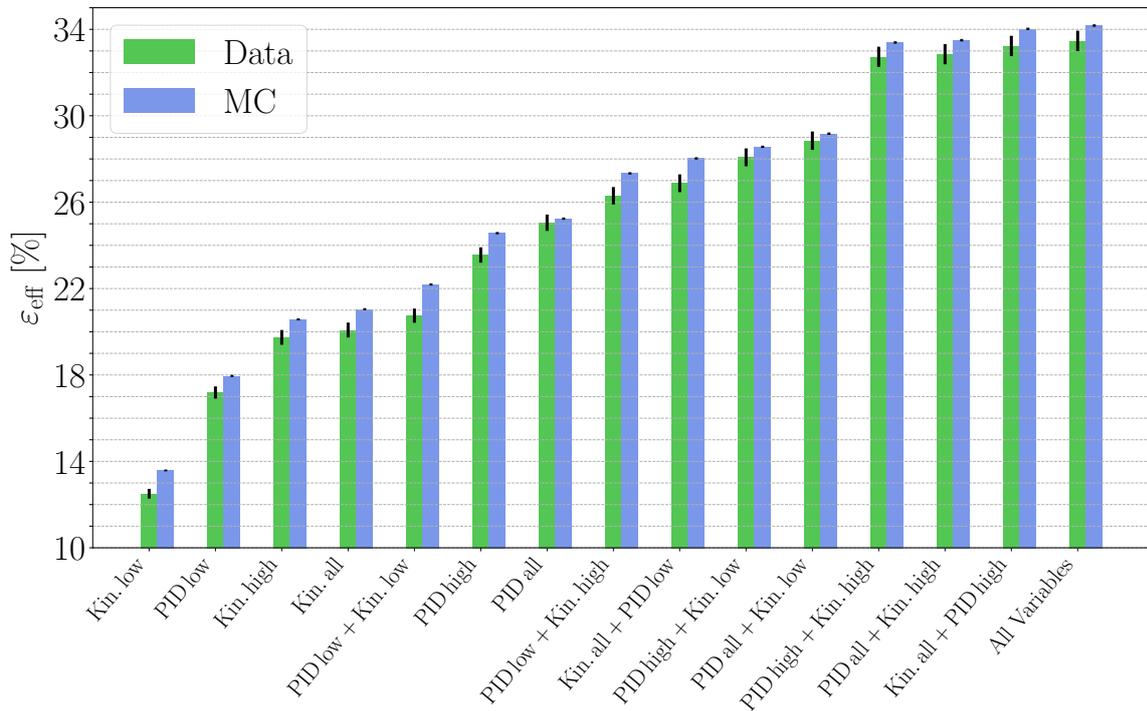
For each combination of input variables, the $q \cdot r$ output distributions of both combiners on Belle MC and on Belle data can be found in App. B.5. Since the efficiencies on data are calculated assuming that the linearity between $\langle r \rangle$ and r_{MC} is valid for data, this linearity is checked for each combiner and for each combination of input variables. The results of all the linearity checks are also presented in App. B.5. Figures 4.48 to 4.51 show the results for the cases with the lowest and with the highest effective efficiencies, and for two intermediate cases. Ordered in increasing value of the effective efficiency, these cases correspond to the following combinations of input variables: kinematic-low only; PID-low together with kinematic-low; PID-high together with kinematic-low; and all variables together. For each combination, the average dilution $\langle r \rangle$ determined from the output of the combiners is linear to the dilution r_{MC} determined using MC information. Within the statistical uncertainties, the $q \cdot r$ output distributions of the FBDT and the MLP combiners on Belle MC show good agreement with those on Belle data for each combination of input variables. This supports the confidence in the accurate description of Belle MC and justifies the multivariate approach with a large number of input variables.

Table 4.7: The four subsets of discriminating input variables used for the test of robustness.

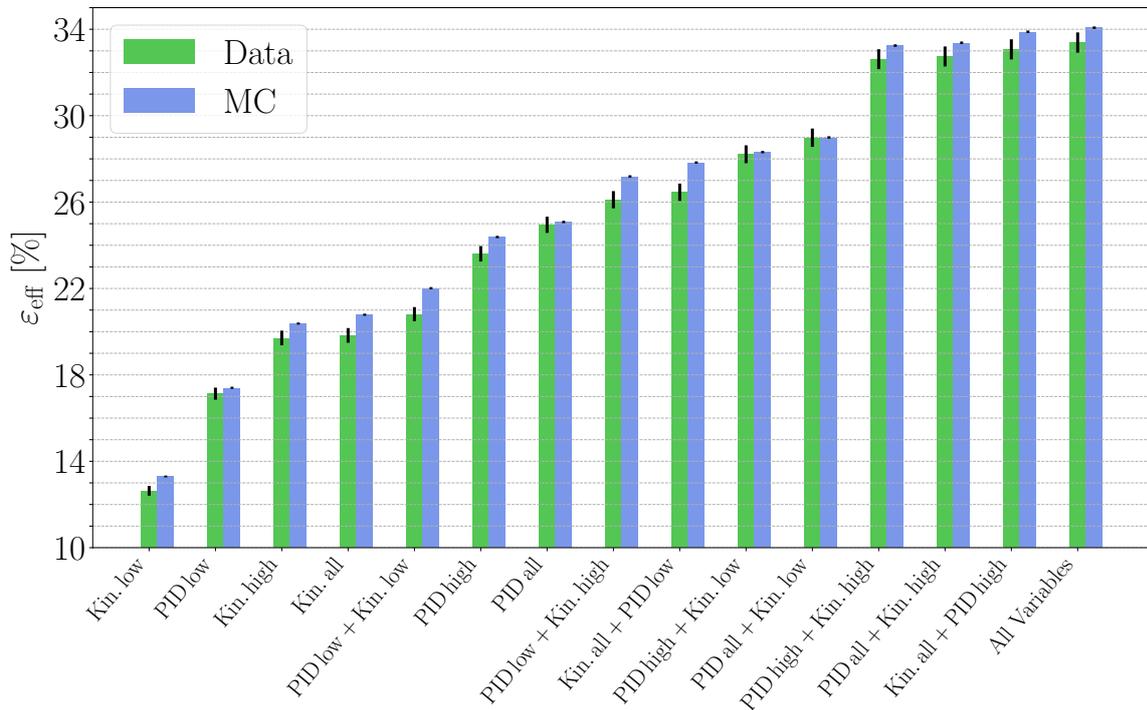
Categories	PID-high variables	Categories	PID-low variables
Electron	$\mathcal{L}_e, \mathcal{L}_e^{dE/dx}, \mathcal{L}_e^{\text{ECL}}$	Electron	$\mathcal{L}_e^{\text{TOP}}, \mathcal{L}_e^{\text{ARICH}}$
Int. Electron		Int. Electron	
Muon	$\mathcal{L}_\mu, \mathcal{L}_\mu^{dE/dx}, \mathcal{L}_\mu^{\text{KLM}}$	Muon	$\mathcal{L}_\mu^{\text{TOP}}, \mathcal{L}_\mu^{\text{ARICH}}$
Int. Muon		Int. Muon	
Kin. Lepton	$\mathcal{L}_e, \mathcal{L}_e^{dE/dx}, \mathcal{L}_e^{\text{ECL}},$ $\mathcal{L}_\mu, \mathcal{L}_\mu^{dE/dx}, \mathcal{L}_\mu^{\text{KLM}}$	Kin. Lepton	$\mathcal{L}_e^{\text{TOP}}, \mathcal{L}_e^{\text{ARICH}},$ $\mathcal{L}_\mu^{\text{TOP}}, \mathcal{L}_\mu^{\text{ARICH}}$
Int. Kin. Lep.		Int. Kin. Lep.	
Kaon	$\mathcal{L}_K, \mathcal{L}_K^{dE/dx}$	Kaon	$\mathcal{L}_K^{\text{TOP}}, \mathcal{L}_K^{\text{ARICH}}$
Slow Pion	$\mathcal{L}_\pi^{\text{ARICH}}, \mathcal{L}_{\pi/e}^{dE/dx}, \mathcal{L}_e,$ $\mathcal{L}_K, \mathcal{L}_K^{\text{ARICH}}$	Slow Pion	$\mathcal{L}_\pi, \mathcal{L}_\pi^{dE/dx}, \mathcal{L}_\pi^{\text{TOP}},$ $\mathcal{L}_K^{dE/dx}, \mathcal{L}_K^{\text{TOP}}$
Fast Hadron		Fast Hadron	
Kaon-Pion	$\mathcal{L}_K, y_{\text{Kaon}}$	Kaon-Pion	y_{SlowPion}
Maximum p^*	-	Maximum	-
FSC	-	FSC	$\mathcal{L}_{K\text{Slow}}$
Lambda	\mathcal{L}_p	Lambda	\mathcal{L}_π
Categories	Kinematic-high variables	Categories	Kinematic-low variables
Electron	$p^*, p_t^*, p, p_t,$	Electron	$M_{\text{rec}}^2, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*,$
Int. Electron	d_0, ξ_0, E_{90}^W	Int. Electron	$\cos \theta, \cos \theta_{\text{T}}^* , p\text{-val.}$
Muon	$p^*, p_t^*, p, p_t,$	Muon	$M_{\text{rec}}^2, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*,$
Int. Muon	d_0, ξ_0, E_{90}^W	Int. Muon	$\cos \theta, \cos \theta_{\text{T}}^* , p\text{-val.}$
Kin. Lepton	$p^*, p_t^*, p, p_t,$	Kin. Lepton	$M_{\text{rec}}^2, p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*,$
Int. Kin. Lep.	d_0, ξ_0, E_{90}^W	Int. Kin. Lep.	$\cos \theta, \cos \theta_{\text{T}}^* , p\text{-val.}$
Kaon	$p^*, p_t^*, p_t,$ $d_0, \xi_0, n_{K_S^0}$	Kaon	$\sum p_t^2, M_{\text{rec}}^2, E_{90}^W, p_{\text{miss}}^*,$ $\cos \theta, \cos \theta_{\text{miss}}^*,$ $ \cos \theta_{\text{T}}^* , p\text{-val.}$
Slow Pion	$p^*, p_t^*, p, p_t, d_0,$	Slow Pion	$n_{K_S^0}, \sum p_t^2, M_{\text{rec}}^2, E_{90}^W,$
Fast Hadron	$\xi_0, \cos \theta, \cos \theta_{\text{T}}^* $	Fast Hadron	$p_{\text{miss}}^*, \cos \theta_{\text{miss}}^*, p\text{-val.}$
Kaon-Pion	$q_K \cdot q_\pi$	Kaon-Pion	$\cos \theta_{K\pi}^*$
Maximum p^*	p^*, p_t^*, p, p_t	Maximum p^*	$d_0, \xi_0, \cos \theta_{\text{T}}^* $
FSC	$p_{\text{Fast}}^*, \cos \theta_{\text{T, Fast}}^* ,$ $\cos \theta_{\text{SlowFast}}^*, q_{\text{Slow}} \cdot q_{\text{Fast}}$	FSC	$p_{\text{Slow}}^*, \cos \theta_{\text{T, Slow}}^* ,$
Lambda	$p_\Lambda^*, p_\Lambda, p_p^*, p_p,$ $M_\Lambda, \cos \theta_{x_\Lambda, p_\Lambda}, x_\Lambda $	Lambda	$p_\pi^*, p_\pi, q_\Lambda, n_{K_S^0},$ $\sigma_\Lambda^{zz}, p\text{-val.}$

Table 4.8: Effective efficiencies of FBDT and of MLP combiners on Belle Data and on Belle MC using different combinations of the subsets of input variables given in Tab. 4.7. The total number of input variables is n_{input} , and the total number of different input variables is n_{unique} (see Sect. 4.4). The results are sorted according to the effective efficiency obtained by the FBDT Combiner on data. All numbers are given in percent with statistical uncertainties only.

Used Subsets	n_{unique}	n_{input}	FBDT Combiner		MLP Combiner	
			Belle data	Belle MC	Belle data	Belle MC
			$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$
Kin. low	37	70	12.5 ± 0.2	13.58 ± 0.04	12.6 ± 0.2	13.30 ± 0.04
PID low	13	30	17.2 ± 0.3	17.95 ± 0.05	17.1 ± 0.3	17.40 ± 0.05
Kin. high	39	80	19.7 ± 0.3	20.57 ± 0.05	19.7 ± 0.3	20.38 ± 0.05
Kin. all	74	150	20.1 ± 0.3	21.04 ± 0.05	19.8 ± 0.3	20.79 ± 0.05
PID low and Kin. low	53	100	20.7 ± 0.3	22.19 ± 0.05	20.8 ± 0.3	22.01 ± 0.05
PID high	18	38	23.6 ± 0.4	24.57 ± 0.05	23.6 ± 0.4	24.38 ± 0.05
PID all	32	68	25.0 ± 0.4	25.24 ± 0.05	24.9 ± 0.4	25.08 ± 0.05
PID low and Kin. high	55	110	26.3 ± 0.4	27.33 ± 0.05	26.1 ± 0.4	27.19 ± 0.05
Kin. all and PID low	89	180	26.9 ± 0.4	28.03 ± 0.05	26.5 ± 0.4	27.83 ± 0.05
PID high and Kin. low	56	108	28.1 ± 0.4	28.56 ± 0.05	28.2 ± 0.4	28.31 ± 0.05
PID all and Kin. low	70	138	28.8 ± 0.4	29.17 ± 0.06	29.0 ± 0.4	28.99 ± 0.06
PID high and Kin. high	58	118	32.7 ± 0.5	33.39 ± 0.06	32.6 ± 0.5	33.25 ± 0.06
PID all and Kin. high	72	148	32.8 ± 0.5	33.50 ± 0.06	32.7 ± 0.5	33.38 ± 0.06
Kin. all and PID high	92	188	33.2 ± 0.5	34.02 ± 0.06	33.1 ± 0.5	33.89 ± 0.06
All Variables	106	218	33.5 ± 0.5	34.18 ± 0.06	33.4 ± 0.5	34.07 ± 0.06

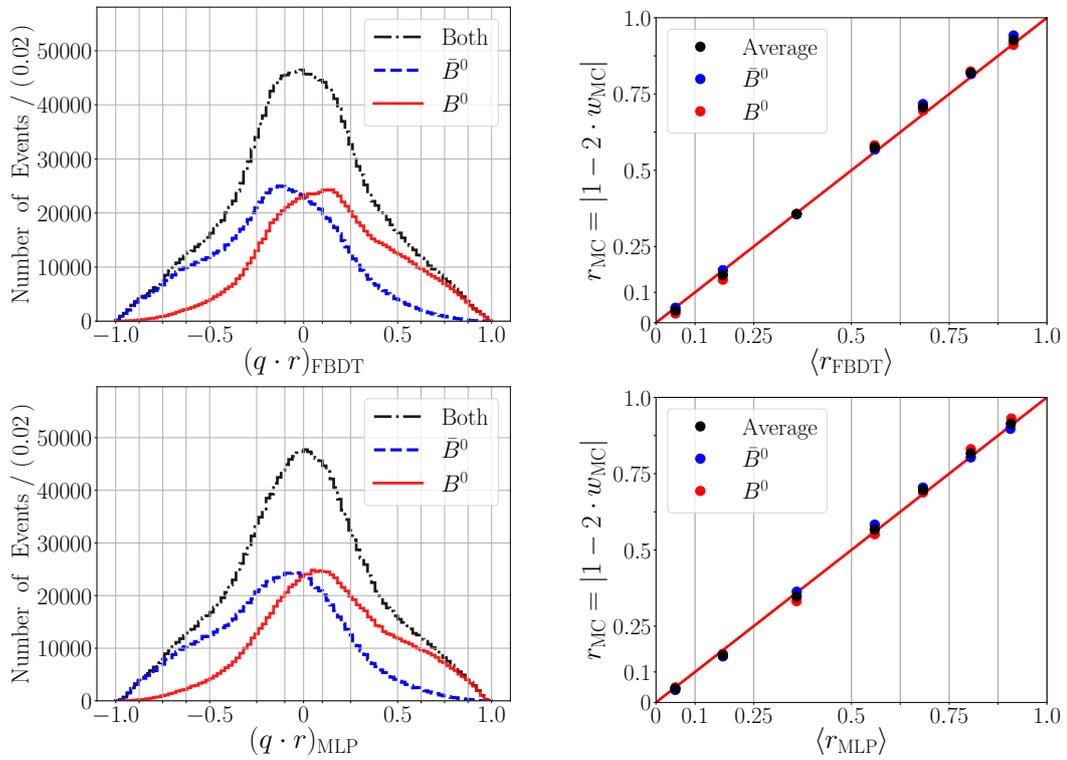


(a) FBDT Combiner

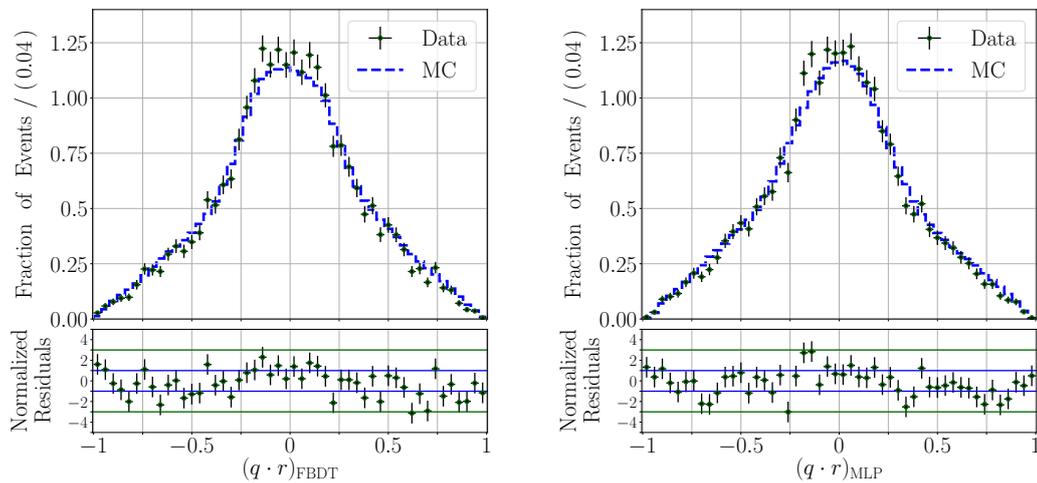


(b) MLP Combiner

Figure 4.47: Effective efficiencies of FBDT and of MLP combiners on Belle data and on Belle MC using different combinations of the subsets of input variables defined in Tab. 4.7. The results correspond to the values in Tab. 4.8 and are sorted according to the increase in effective efficiency obtained by the FBDT combiner on data.

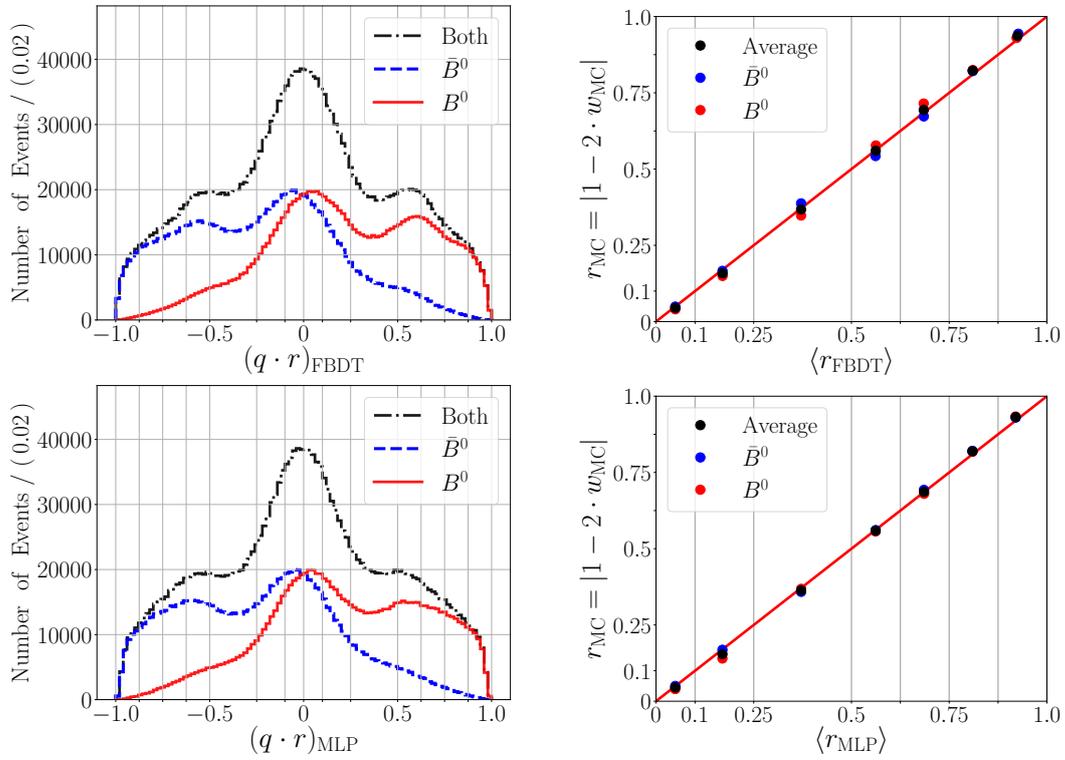


(a) Belle MC.

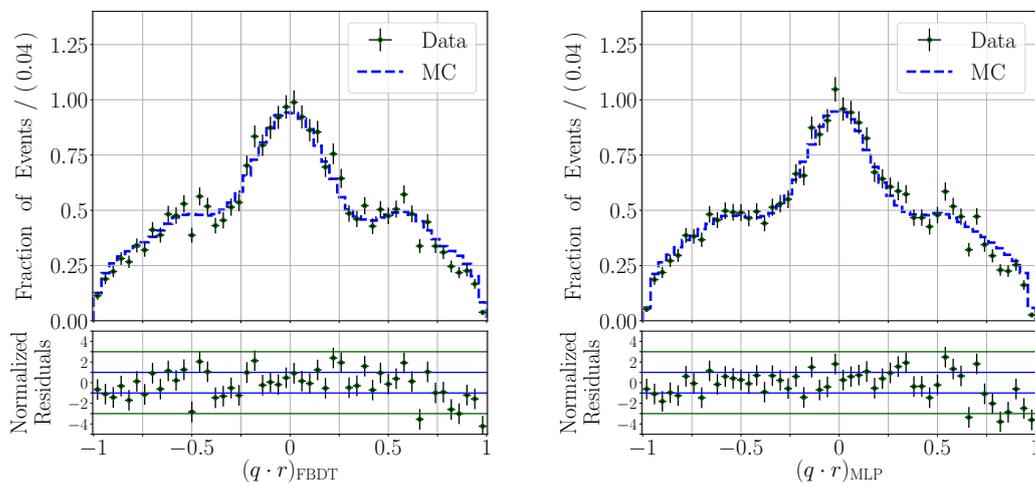


(b) Belle data and Belle MC.

Figure 4.48: Performance of FBDT and of MLP combiners using the kinematic-low variables: (a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); (b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

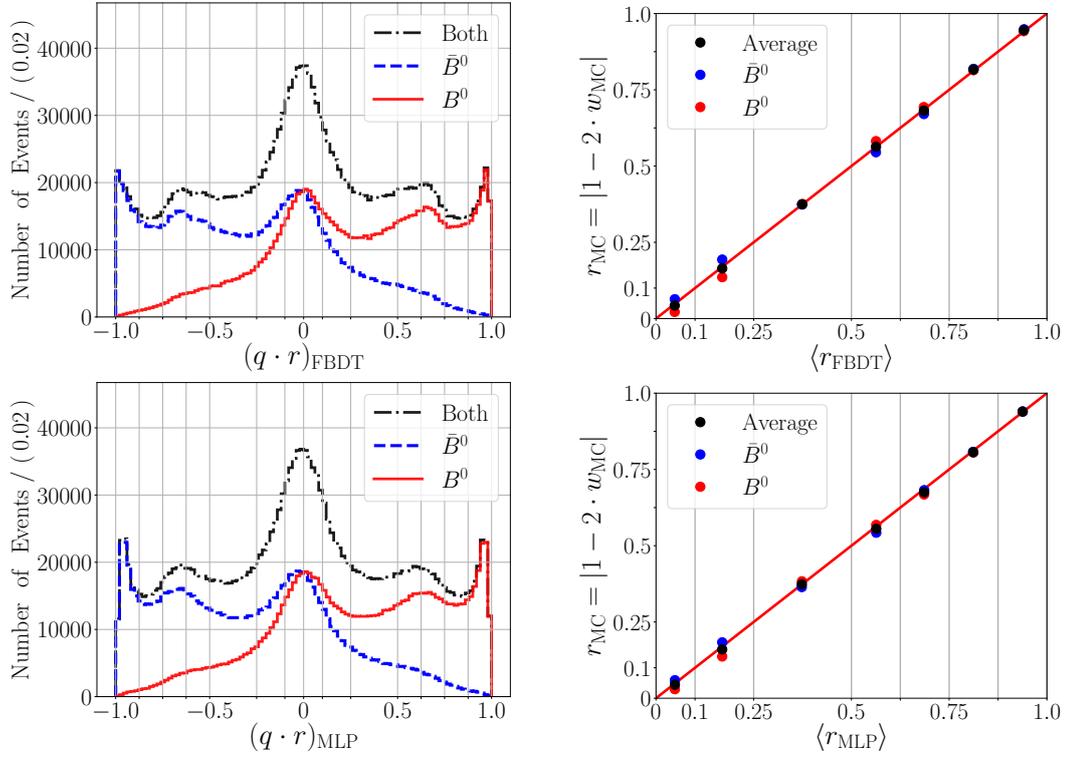


(a) Belle MC.

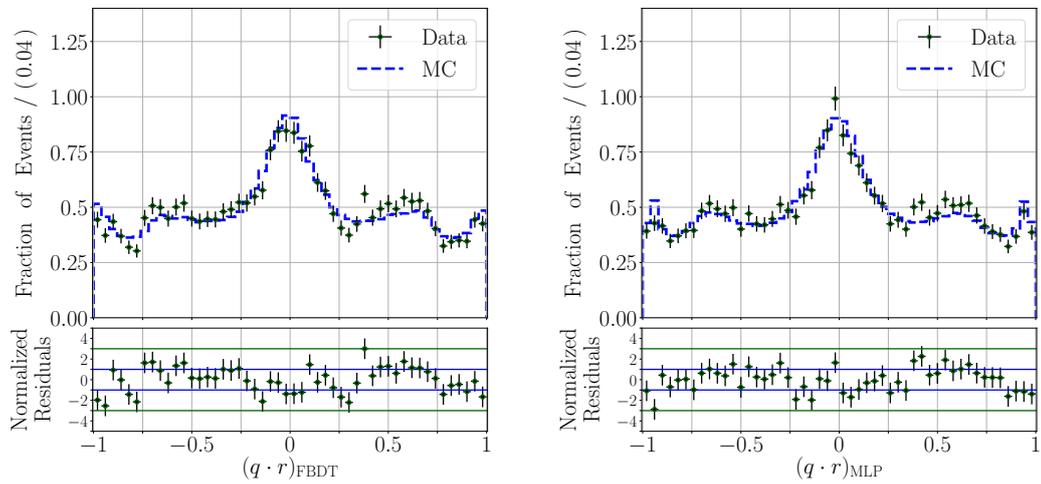


(b) Belle data and Belle MC.

Figure 4.49: Performance of FBDT and of MLP combiners using the PID-low variables together with the kinematic-low variables: (a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); (b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.



(a) Belle MC.



(b) Belle data and Belle MC.

Figure 4.50: Performance of FBDT and of MLP combiners using the PID-high variables together with the kinematic-low variables: (a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = |1 - 2 \cdot w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); (b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

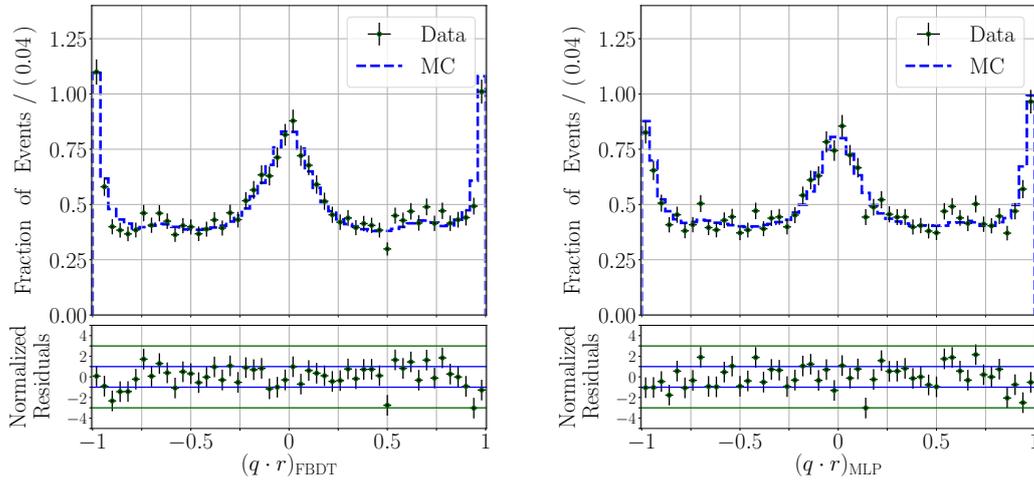
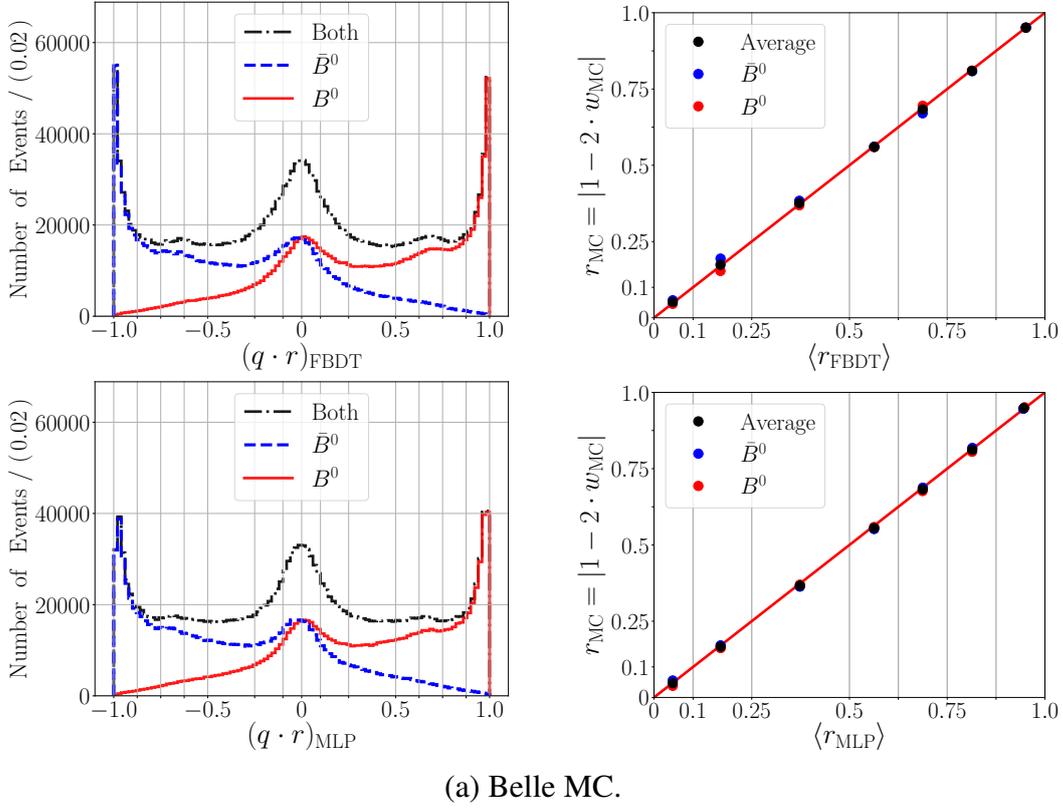


Figure 4.51: Performance of FBDT and of MLP combiners using all variables: (a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); (b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

4.10 The tag-side interference effect

Two systematic effects associated with flavor tagging affect the measurements of CP -violation parameters. The first one is caused by the uncertainties of the flavor-tagging parameters. These are the average wrong-tag fraction w , the difference Δw , the efficiency ε , and the parameter $\mu = \Delta\varepsilon/(2\varepsilon)$ defined in Sect. 4.2. The respective uncertainties are measured using control samples and scale with the integrated luminosity. The second effect is the tag-side interference effect which was pointed out by Owen Long, Max Baak, Robert N. Cahn and David Kirkby [134].

Some hadronic final states used for flavor tagging are possible final states of a B^0 decaying via a CKM-favored transition, and of a \bar{B}^0 decaying via a doubly-CKM-suppressed transition. Since at B factories $B^0\bar{B}^0$ pairs are created in an entangled state, the interference between CKM-favored and doubly CKM-suppressed decay amplitudes in the tag side have an impact on the measurement of the CP -violation parameters in the signal decay channel.

Prominent examples of such hadronic final states are $D^+\pi^-$ and $D^{*+}\pi^-$, which can be the final states of a B^0 decaying via the CKM-favored transition $b \rightarrow c\bar{u}d$, or of a \bar{B}^0 decaying via the doubly-CKM-suppressed transition $\bar{b} \rightarrow \bar{u}c\bar{d}$. Figure 4.52 shows the respective Feynman diagrams. The doubly-CKM-suppressed decay amplitude is suppressed by a about factor $|V_{ub}^*V_{cd}/(V_{cb}V_{ud}^*)| \approx 0.02$ with respect to the CKM-favored one. The relative weak phase between the two amplitudes is ϕ_3 .

Considering the dominant $\bar{b} \rightarrow \bar{u}c\bar{d}$ and the doubly-CKM-suppressed transition $\bar{b} \rightarrow \bar{u}c\bar{d}$, for a specific final state f , the strength of the tag-side interference effect can be parametrized in terms of the real ratio r_f and the strong phase δ_f between the doubly-CKM-suppressed and the CKM-favored amplitudes, and a weak phase $\Phi \equiv 2\phi_1 + \phi_3$ resulting from the difference in the weak phases between the two amplitudes, and from the $B^0 - \bar{B}^0$ mixing.

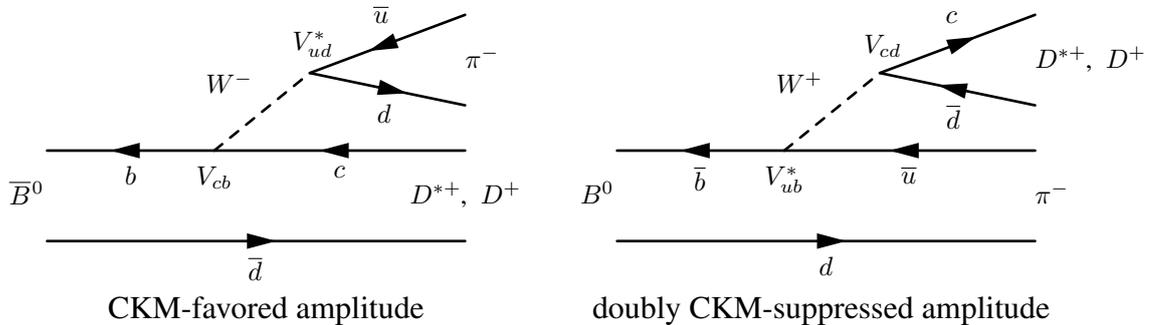


Figure 4.52: The CKM-favored amplitude (left) and the doubly-CKM-suppressed amplitude (right) for final states $D^{*+}\pi^+$ ($D^+\pi^+$). With respect to the dominant contribution, the latter is suppressed by the approximate ratio $|V_{ub}^*V_{cd}/(V_{cb}V_{ud}^*)| \approx 0.02$, and has a relative weak phase difference of ϕ_3 [134].

Since a tagging category combines several final states f of a certain group of decays, one introduces the effective values r' and δ' . The value of r' and δ' cannot be determined analytically. However, r' is expected to be very small (below 0.02), since the contributions from different tagging states tend to cancel unless all contributions have nearly the same strong phase. Using this parametrization, one can calculate corrections for the measured mixing parameters and for the measured CP -violation parameters of the signal side B_{sig}^0 .

Considering the Δt probability for events where the signal B_{sig}^0 is reconstructed to an apparent flavor eigenstate, such as $D^+\pi^-$, it turns out that the dependencies proportional to 1 and to $\cos \Delta m \Delta t$ remain unaffected, while a small $\sin \Delta m \Delta t$ term is induced. Here, proportional to 1 refers to constant terms that are multiplied with the exponential decay $e^{-|\Delta t|/\tau_{B^0}}$ in the Δt probability.

In measurements of $B^0 - \bar{B}^0$ mixing, the induced sinus term is assumed to be zero. Since the estimation of the mixing parameters depends only on the terms proportional to 1 and to $\cos \Delta m \Delta t$ (see eq. (4.9)), these measurements are unaffected by doubly-CKM-suppressed amplitude contributions, either on the tag side or on the reconstructed side. Furthermore, the time-integrated mixing probability (eq. (4.10)) does not contain any corrections due to the $\sin \Delta m \Delta t$ term since it integrates to zero. Since the measurement of the wrong-tag fraction stays unaffected, corrections for the tag-side interference effect cannot be simply absorbed into it [134].

In measurements of CP -violation parameters, the corrections due to the tag-side interference effect are not negligible, but they can be treated as perturbations on the usual measurements. The perturbations introduce a systematic shift and an associated uncertainty; it is possible to relate the true values of the CP -violation parameters \mathcal{A}_{CP} and \mathcal{S}_{CP} to the values \mathcal{A}_{fit} and \mathcal{S}_{fit} extracted from maximum-likelihood fit to the data [134].

At Belle and at BaBar, the effect on \mathcal{A}_{fit} and \mathcal{S}_{fit} was estimated by comparing the fit results of pseudo-experiments which were simulated without and with the tag-side interference terms, this means with $r' = 0$ and with $r' \neq 0$. In the second case, the value of the amplitude ratio $r_{D\pi} \approx 0.021$ [134] between $B^0 \rightarrow D^+\pi^-$ and $B^0 \rightarrow D^-\pi^+$ was taken to determine r' . The pseudo-experiments with interference terms were simulated assuming this value of r' and a uniformly distributed phase difference δ' , because of the lack of knowledge about strong phase differences [29, 134]. With this approach, Belle and BaBar found no indication of a significant shift in the measured values of \mathcal{A}_{CP} and \mathcal{S}_{CP} [29]. For decays occurring via $b \rightarrow c\bar{c}s$, $b \rightarrow c\bar{c}d$, and $b \rightarrow q\bar{q}s$ transitions, the estimated systematic uncertainty is typically around 0.014 for \mathcal{A}_{CP} , and around 0.001 for \mathcal{S}_{CP} [29, 59, 60].

For the extraction of ϕ_1 from the \mathcal{S}_{CP} measurements of the $b \rightarrow c\bar{c}s$ channels, the tag-side interference effect is negligible. However, for precision measurements of \mathcal{A}_{CP} , like for example in the case of the final state $J/\psi K_S^0$, the tag-side interference effect is the dominant source of systematic uncertainties. Belle noted, however, that there is a partial cancellation when combining measurements of CP even and of CP odd decay channels, like when combining measurements for the final states $J/\psi K_S^0$ and $J/\psi K_L^0$ [59].

For decays of B^0 occurring via $b \rightarrow u\bar{u}d$ transitions, like for $B^0 \rightarrow \pi^+\pi^-$, the evaluation of the tag-side interference effect is more complicated because the decay amplitudes receive a significant penguin contribution [29, 134]. In the case of $B^0 \rightarrow \rho^+\rho^-$ the penguin contributions are small and it is therefore easier. Belle estimated for the decay $B^0 \rightarrow \pi^+\pi^-$ a tag-side interference uncertainty of 0.03 for \mathcal{A}_{CP} , and of 0.002 for \mathcal{S}_{CP} ; and for $B^0 \rightarrow \rho\rho$, an uncertainty of 0.01 for \mathcal{A}_{CP} , and of 0.001 for \mathcal{S}_{CP} . As it will be shown in chapter 7, this effect has no impact in the determination of ϕ_2 since the systematic uncertainties of the CP -violation parameters are not the dominant uncertainties entering the isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$.

In principle, one could mitigate the tag-side interference effect by measuring r' and δ' using control samples. However, this measurement is very challenging. As mentioned before, in time-dependent measurements of $B^0 - \bar{B}^0$ mixing the tag side interference effect leaves unchanged the terms proportional to 1 and to $\cos \Delta m \Delta t$ in the Δt probability, but it introduces a $\sin \Delta m \Delta t$ term. This sinus term depends on r' , δ' , and Φ , which depend on the channels considered by the flavor tagger in the tag-side, and on the values of r_f and δ_f for the channel on the signal side. Furthermore, this sinus term changes depending on the flavors of B_{sig}^0 and of B_{tag}^0 [134].

If the reconstructed signal channel is a semileptonic one, for example the high-purity decay channel $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$, then r_f is zero and the sinus term is simplified. With a large sample size—larger than the ones at Belle and at BaBar since they did not succeed in measuring the correction terms—Belle II could pioneer a measurement of r' and δ' . Since the flavor-tagging parameters w , Δw and μ enter this measurement also, one could determine the values of w , Δw and μ from a time-integrated calibration, such that only r' and δ' remain as fit variables. Using the values of r' and δ' with their uncertainties, the tag-side interference terms could be included into the statistical model for measurements of CP -violation parameters. In case that an analysis is performed in bins of the dilution factor, then r' and δ' have to be measured for each bin.

At Belle II, the tag-side interference effect will be especially relevant for precision measurements of \mathcal{A}_{CP} . To fully get rid of the systematic uncertainties associated with the tag-side interference effect, the only way is to use only the tagging categories that consider primary leptons as targets. For these decays, there is no doubly CKM-suppressed decay amplitude and thus $r' = 0$ for these categories. Therefore, a dedicated study was launched to characterize and to validate the performance of the flavor tagger using only the primary-lepton categories.

4.11 Performance using only primary-lepton categories

Since the primary-lepton categories will play a crucial role in precision measurements at Belle II (see previous section), it is necessary to evaluate and to validate the performance of the flavor tagger when using only the Electron, the Muon, and the Kinetic Lepton category. This configuration of the Belle II flavor tagger can be regarded as the semileptonic Belle II flavor tagger. To characterize the performance of this flavor tagger, a dedicated study was launched in which this flavor tagger was trained and tested using Belle II MC and Belle MC. The MC samples are those described in Sect. 4.5.1. The performance was validated again using Belle data as described in Sect. 4.8.

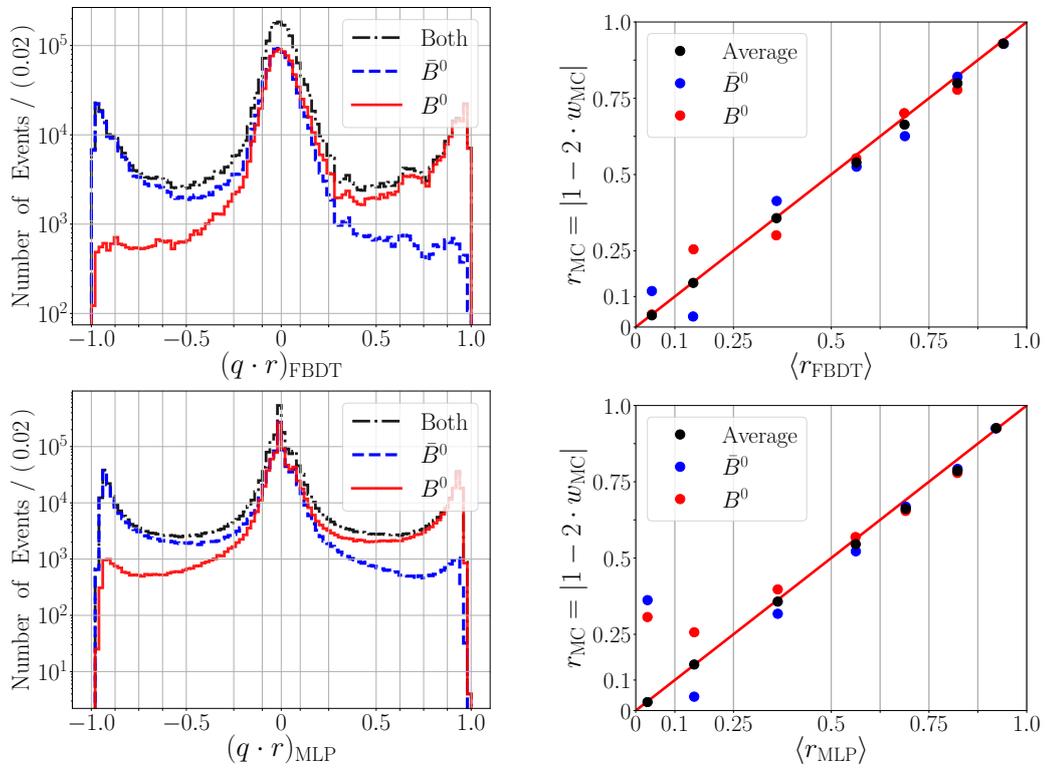
Figure 4.53 shows the $q \cdot r$ distributions of the FBDT and the MLP combiners on Belle MC and on Belle data. As the figure shows, the shapes of the $q \cdot r$ distributions on Belle MC agree with the shapes of the distributions on Belle data within the uncertainties. To calculate the respective effective efficiencies, the linearity between r_{MC} and $\langle r \rangle$ is checked and assumed to be valid for data. Figure 4.53 shows also the results of the linearity checks.

Table 4.9 presents the effective efficiencies of both combiners on Belle II and on Belle MC, as well as on Belle data. The obtained efficiencies are around 12% and about one third of the effective efficiency achieved when using all the tagging categories. This translates into an increase of the statistical uncertainties on CP -violation parameters by about a factor $\sqrt{3} \approx 1.7$: the cost for eliminating the tag-side interference effect.

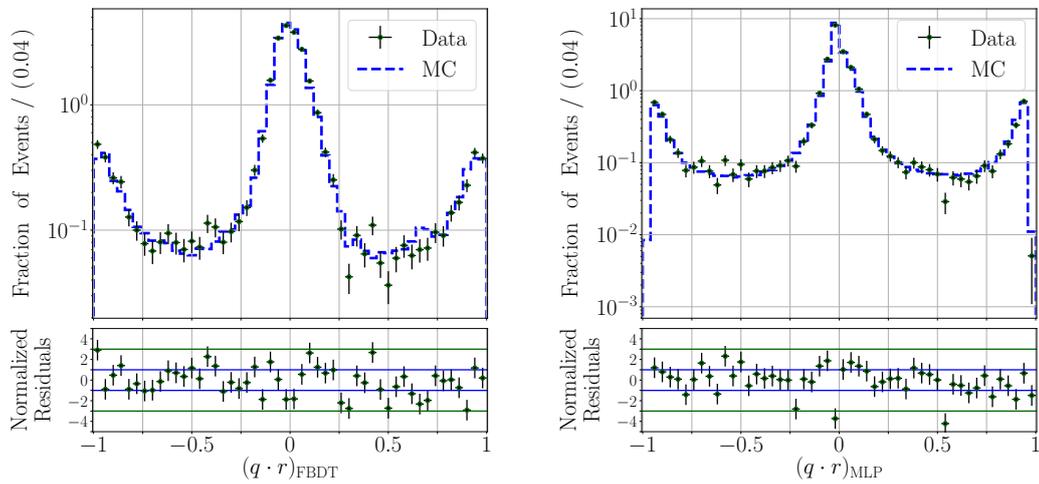
The use of the semileptonic flavor tagger to measure the final values of the CP -violation parameters will depend on the particular case. The main criterion for this will be that the reduction in the systemic uncertainty compensates the increase in the statistical one. In any case, the semileptonic flavor tagger will be required to investigate the impact of the tag-side interference effect. And for this, the values of the flavor-tagging parameters w , Δw and μ for this particular flavor tagger will have to be measured on data.

Table 4.9: Performance of FBDT and MLP combiners on Belle Data, on Belle MC and on Belle II MC using only the three semileptonic categories: Electron, Muon and Kinetic Lepton.

Sample	FBDT Combiner		MLP Combiner	
	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$
Belle Data	12.6 ± 0.3	-	12.2 ± 0.3	-
Belle MC	11.99 ± 0.03	-0.09 ± 0.07	11.81 ± 0.03	0.07 ± 0.07
Belle II MC	12.35 ± 0.03	0.23 ± 0.06	11.85 ± 0.03	0.34 ± 0.06



(a) Belle MC.



(b) Belle data and Belle MC.

Figure 4.53: Performance of FBDT and of MLP combiners using only the semileptonic categories: Electron, Muon and Kinetic Lepton. (a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); (b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

4.12 Comparison with the alternative deep-learning approach

The deep-learning flavor tagger is an independent algorithm at Belle II developed by Jochen Gemmler [122]. The algorithm is based on a multi-layer perceptron with eight hidden layers. The deep-learning MLP receives as input variables the characteristics of the reconstructed tracks on the tag side, and gives as output the product $q \cdot r$.

The deep-learning approach avoids the pre-selection of candidates performed by the individual categories in the category-based approach. Because of its complex architecture, the deep-learning MLP is assumed to learn the correlations between the characteristics of the tag-side tracks and the flavor of B_{tag}^0 . The deep-learning MLP implementation is based on the machine learning library Pylearn2 [135] which builds on the numerical computation library Theano [136]. For more details see [122].

The deep-learning flavor tagger sorts the tracks on the tag side into two groups, a positive and a negative one, according to their electric charge. The algorithm ranks the tracks in each group according to their momenta in the $\Upsilon(4S)$ frame, and selects the top 5 tracks in each group. If an event contains less than 5 positive or less than 5 negative tracks, the algorithm sets the input variables for the missing candidates to zero.

For each candidate, the deep-learning MLP receives 14 input variables corresponding to the momentum, the cosine of the polar angle and the azimuthal angles in the $\Upsilon(4S)$, the combined PID likelihoods for the five different particle hypotheses (see Sect. 3.5), the number of hits in the PXD, in the SVD, and in the CDC, the impact parameters d_0 and z_0 (see Sect. 3.3), and the p -value of the track fit. Multiplying the number of input variables by the number of candidates yields 140, corresponding to the number of input nodes.

When the category-based and the deep-learning flavor taggers were developed, the track reconstruction algorithm at Belle II reached a poor performance in events with simulated background. Therefore, only Belle II MC samples without background were employed to develop and to optimize the algorithms. Both algorithms have been trained and tested also on Belle MC samples with background.

Table 4.10 presents the effective efficiencies of the category-based and the deep-learning flavor taggers on Belle II MC and on Belle MC. They are calculated using the r binning of Belle. The Belle II MC samples used to train and to test the flavor taggers were generated without CP violation. The Belle MC samples used for the deep-learning flavor tagger were generated without CP violation, while for the category-based flavor tagger the Belle MC samples were generated with $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0.69)$. As shown in Sect. 4.6, CP violation has a negligible effect on flavor tagging at Belle.

The difference between the performances of the flavor taggers is relatively large on Belle II MC events, but is negligible on Belle MC events. Further studies are needed to understand why the deep-learning approach outperforms the category-based approach on Belle II MC, but not

Table 4.10: Effective efficiencies of the category-based flavor tagger and the deep-learning flavor tagger on Belle II MC and on Belle MC. All values are given in percent.

Approach	Belle II MC	Belle MC
	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$
Category-based	36.64 ± 0.05	34.18 ± 0.06
Deep-learning [122]	40.69 ± 0.03	34.42 ± 0.09

on Belle MC. At present, the deep learning-based flavor tagger can be trained and tested only by its developer; a direct comparison of both taggers for each r bin is still to be performed.

As explained in Sect. 4.5.3, the category-based flavor tagger exploits the known flavor signatures of B -meson decays. To investigate why the performance of the flavor taggers differ, the output of the individual categories can be compared with the output of the deep-learning flavor tagger. In this way, one can test if the MLP of the deep-learning flavor tagger learns physical information or MC features.

An advantage of the category-based approach is that it can be trained using only primary-lepton categories. As explained in Sect. 4.10, using only primary-lepton categories is the only way to fully eliminate the systematic uncertainties caused by the tag-side interference effect.

Because of the complex MLP architecture, the deep-learning flavor tagger requires much larger training samples than the category-based one. Tests with different training sample sizes showed that the best results could be obtained with the largest available sample of 55 million events. The tendency shows that the performance could still improve with larger training samples. In comparison, the category-based approach reaches good results with samples of about 2 million events (one million for the event level and one million for the combiner level). Tests with larger training samples showed no considerable improvement.

Because of the MLP complexity and the large training sample sizes, the deep-learning flavor tagger requires more computing resources than the category-based one to train the algorithm and to prepare the training samples. To exploit their parallel computation capabilities, Graphics Processing Units (GPU) are used to train the deep-learning MLP. However, the deep-learning MLP can be tested on machines using only Central Processing Units (CPU). On a GTX970 GPU, the training procedure for the 8-layers MLP takes about 48 hours. For the category-based flavor tagger, the training procedure takes about 5 hours running on a single CPU core.

At present, only the category-based flavor tagger has been validated on Belle data. Additionally, it has been trained and tested using Belle II MC with simulated background. The results are shown in chapter 5. For the deep-learning flavor tagger, the validation on Belle data is currently ongoing, and tests using Belle II MC with simulated background still have to be performed. At Belle II, the goal is to validate and to calibrate both flavor taggers. Comparing the performance of both flavor taggers offers a unique possibility to crosscheck their results and to investigate differences between data and MC.

4.13 Comparison with Belle and BaBar

At Belle II, the category-based flavor tagger was developed building on useful concepts of the earlier algorithms used by Belle and by BaBar. In the following, the Belle II category-based flavor tagger will be compared with these previous algorithms, and it will be referred to as Belle II flavor tagger.

With respect to the Belle flavor tagger, the major improvements consist in: the inclusion of three complementary flavor signatures corresponding to the Kaon-Pion, the FSC and the Maximum p^* categories; the consideration of fast kaons as targets in the Fast Hadron category (the Belle algorithm used only fast pions); the use of more input variables within each category; and the employment of robust FBDT and MLP multivariate methods. The Belle flavor tagger was based on multi-dimensional lookup tables and considered 10 flavor signatures, which correspond to the same used by the Belle II flavor tagger apart from the three complementary signatures mentioned above. The flavor signatures used by the Belle flavor tagger were sorted into four categories (Lepton, Kaon, Slow Pion and Lambda). Fast pions were included as targets in the Lepton and in the Kaon category. In comparison, the Belle II flavor tagger considers in total 13 flavor signatures.

A significant difference between the Belle and the Belle II algorithms is that the Belle algorithm sorted out the candidates into the different categories according to specific values of the input variables. The drawback of this approach is that optimal values have to be found each time the performance in detection and reconstruction changes. On the contrary, the Belle II algorithm selects the best candidate (or the best candidates) within each category according to the output of the respective event level multivariate method. If the performance in detection and reconstruction changes, the multivariate methods just need to be retrained.

In the Belle flavor tagger, each particle candidate was used only once within a certain category. If a target particle was classified into the wrong category, it was missing as candidate for its true category. In the Belle II algorithm, each particle candidate is used for all categories (disregarding the Lambda category). Thus, the algorithm is robust against this kind of misclassification.

The Belle flavor tagger reached an effective efficiency of $(30.1 \pm 0.4)\%$ on Belle data [29]. In comparison, the Belle II flavor tagger reaches an effective efficiency of $(33.5 \pm 0.4)\%$ on Belle data. The increase in effective efficiency corresponds to the improvements in the Belle II tagging algorithm.

Considering now the BaBar flavor tagger, its structure is similar to the one of the Belle II flavor tagger. The BaBar flavor tagger was based on 9 categories on the event level, and combined the outputs of the single categories into a final $y = q \cdot r$ output. The multivariate methods were all MLPs. In comparison, the major improvements in the Belle II flavor tagger are: the inclusion of four additional categories, the three categories considering intermediate leptons and the Fast Hadron category; the use of more input variables within each category; and the employment of the robust FBDT multivariate method. The MLP used in the Belle II flavor tagger serves only to perform a cross check.

In the BaBar flavor tagger, the events were sorted into event categories depending on the output of the individual tagging categories and depending on the absolute value of the combiner output $|y|$. Thus, the different event categories were related to different ranges of the dilution factor r . At Belle II, the events are sorted depending only on $r = |y|$, using the same r binning of Belle.

Regarding the effective efficiency, BaBar reached $(33.1 \pm 0.3)\%$ on the full BaBar data [29]. The performance of the Belle II flavor tagger on Belle data is slightly better, corresponding to the improvements in the algorithm.

As explained in Sect. 4.2, the values of μ , w and Δw are required for the measurements of CP -violation parameters and of $B^0 - \bar{B}^0$ mixing. Literature suggests that Belle ignored the parameter μ . BaBar measured it (the values are around -2.4% to $+1.5\%$ depending on the r -bin), but it is not clear how it was used in data analysis [29]. For precision measurements at Belle II, the parameter μ will have to be taken into account. For this, further studies are required. As it was shown in Sect. 4.6, on Belle II MC without background and with the IP fixed at zero, the value of μ changes slightly (the values are between -2% and $+2\%$) depending on the CP violation parameters.

4.14 Calibrating the flavor tagger

The measurements of CP violation and of $B^0 - \bar{B}^0$ mixing that require flavor tagging, require also the values of the flavor tagging parameters ε , μ , w and Δw (see Sect. 4.2). The calibration of a flavor tagger consists in determining these parameters from data to avoid systematic uncertainties caused by the possible differences between data and MC simulations. From the value of the wrong-tag fraction on data, one derives the true dilution factor $r_{\text{data}} = |1 - 2 \cdot w_{\text{data}}|$, and checks if it corresponds to the average dilution $\langle r \rangle = \langle |y| \rangle$, where y corresponds to the output of the flavor tagger.

Besides the flavor tagging parameters, time-dependent analyses require to model the Δt distribution taking into account the systematic effects caused by the detector alignment, the detector resolution and the vertex reconstruction. This is performed by convoluting the measured Δt distribution with an experimental resolution function. At Belle and BaBar, the resolution function $R(\delta t, \sigma_{\Delta t})$ depended on the difference $\delta t = \Delta t - \Delta t_{\text{true}}$ between the measured time difference and the true time difference, and the uncertainty $\sigma_{\Delta t}$ on the measured Δt . The final Δt probability density function was [29]

$$\begin{aligned} \mathcal{P}^{\text{final}}(\Delta t, q) &= \int_{-\infty}^{\infty} \mathcal{P}^{\text{obs}}(\Delta t_{\text{true}}, q_{\text{sig}}, q, \varepsilon, \mu, w, \Delta w) \cdot \mathcal{R}(\delta t, \sigma_{\Delta t}) \cdot d\Delta t_{\text{true}} \\ &= \mathcal{P}^{\text{obs}}(\Delta t, q_{\text{sig}}, q, \varepsilon, \mu, w, \Delta w) \otimes \mathcal{R}(\delta t, \sigma_{\Delta t}), \end{aligned} \quad (4.21)$$

where the function \mathcal{P} corresponds to eq. (4.9).

The parameters of the resolution function can be determined together with the flavor tagging parameters. For this, Belle and BaBar measured the $B^0 - \bar{B}^0$ mixing using control samples

of self-tagging decay modes. The parameters of the functions are extracted performing an unbinned maximum likelihood fit to the data. The signal and the background events are distinguished by the shapes of discriminating variables. Clearly, the flavor tagging parameters cannot be determined for each event, but for groups of events. Belle sorted the events into bins of the dilution factor r , and BaBar into event categories (see Sect. 4.13). Belle and BaBar did not consider the possible asymmetry between the efficiencies for B^0 and for \bar{B}^0 [29] and, thus, the parameter $\mu = \Delta\varepsilon/(2\varepsilon)$ was set to zero. For precision measurements at Belle II, the effect of this parameter will have to be taken into account.

Belle used as control samples the channels $B \rightarrow D^{(*)-}\pi^+, D^{*-}\rho^+, D^{*-}\ell^+\nu_\ell$, as well as the charmonium channels $J/\psi K_S^0$ and $J/\psi K^{*0}$. The last two channels were used as a cross-check and not to extract the flavor tagging parameters. BaBar used the channels $B \rightarrow D^{(*)-}(\pi^+, \rho^+, a_1^+)$, and $D^{*-}\ell^+\nu_\ell$. The latter sample was used only as a cross-check since it had a larger background than the other samples [29]. To perform an independent validation, Belle and BaBar measured the CP -violation parameter S_{CP} for charged B -decay modes. Since S_{CP} is related to $B^0 - \bar{B}^0$ mixing, it has to be consistent with zero for charged B -decay modes. The golden channel for such validations is $B^+ \rightarrow J/\psi K^+$.

At Belle, the dominant sources of systematic uncertainties on the measurement of flavor tagging parameters were the vertex reconstruction and the resolution parameters [130]. The other systematic uncertainties were caused by the uncertainty in the signal fraction, the background shape, the semileptonic D^{**} composition, the background mixing, and the physical parameters τ_{B^0} and Δm (see Sect. 2.5). At Belle II, the uncertainties due to the vertex reconstruction and the resolution parameters are expected to reduce by about a factor 2, as will be discussed in Sect. 7.3. All other uncertainties affecting the measurement of flavor tagging parameters will decrease with larger sample sizes.

In principle, it is possible to calibrate the flavor tagger performing time-integrated analyses. This is possible due to the relatively small time-integrated mixing probability $\chi_d = 0.1860 \pm 0.0011$ [42] (see Sect. 2.7.3). In this case, the signal probability function \mathcal{P} is given by eq. (4.10). The advantage of this approach is the absence of systematic uncertainties caused by the vertex reconstruction and by the resolution parameters. However, it comes at the cost of a decreased sensitivity due to the factor $(1 - 2\chi_d) = 0.628$.

Belle and BaBar, did not consider the tag-side interference effect for measurements of $B^0 - \bar{B}^0$ mixing. As explained in Sect. 4.10, Belle II could pioneer the measurement of correction terms needed for CP -violation analyses.

At present, the Belle II category-based flavor tagger is ready to be calibrated and to be used in analyses. The calibration on Belle II data will be performed when Belle II will have collected enough integrated luminosity to be competitive with its predecessor Belle (at least about 1 ab^{-1}).

4.15 Conclusion

This chapter was devoted to the development and the validation of a new flavor tagging algorithm at Belle II. The algorithm is based on the concept of tagging categories which identify the flavor signatures provided by flavor-specific B -meson decays. This concept, as well as the use of certain input variables, was adopted from previous algorithms developed by the Belle and by the BaBar collaborations. In comparison with its predecessors, the new algorithm at Belle II considers more flavor signatures and exploits the reconstruction capabilities at Belle II in order to cope with the expected challenging experimental conditions.

Relying on the physics of B -meson decays, the new algorithm employs a new multivariate method, the Fast Boosted Decision Tree, developed especially for Belle II [115]. For each tagging category, an FBDT is trained to identify the particles providing the corresponding flavor signature. The FBDTs receive input variables describing particle identification information and kinematic properties of the tracks on the tag-side, as well as global information provided by all tag-side tracks and clusters in the ECL and KLM subdetectors of Belle II.

The use of multivariate methods was inspired by the BaBar algorithm. This algorithm reached a better performance than the Belle's one, which was based on cuts on the input variables. The Belle II algorithm selects the best particle candidates based on the output of the event-level FBDTs without any cuts on the input variables. In this way, the Belle II algorithm does not need to be optimized again if the detection and the reconstruction performance at Belle II change. The FBDTs just need to be retrained.

The flavor information provided by the selected particle candidates is combined in a second step by another FBDT. This combiner provides an output corresponding to the product $q \cdot r$, where q is the flavor of the tag-side B meson and r is the flavor dilution factor. A second multivariate method, a multi-layer perceptron, receiving the same input, is used as a cross-check. Using the output of the FBDT combiner, the Belle II flavor tagger reaches on Belle data an effective efficiency of

$$\varepsilon_{\text{eff}} = (33.5 \pm 0.5)\%,$$

while the MLP combiner reaches $(33.4 \pm 0.5)\%$. The uncertainties are only statistical. The performance of the flavor tagger was validated with the full Belle data sample corresponding to 711 fb^{-1} collected at the $\Upsilon(4S)$ resonance using the decay $B \rightarrow J/\psi K_S^0$ on the signal side.

For comparison, the Belle algorithm reached an effective efficiency of $(30.1 \pm 0.4)\%$ [29]. The increase of 10% in effective efficiency is due to the improvements in the new algorithm at Belle II.

The Belle II algorithm was developed and optimized using official Belle II MC events which were generated, however, without background and with the production vertex of the $\Upsilon(4S)$ fixed at the origin. In the last stage of this thesis, a new official sample became available. The new events were generated with background and with the production vertex of the $\Upsilon(4S)$ distributed according to the expected position and size of the interaction region at SuperKEKB. The results with this new sample are presented in the next chapter.

The Belle II collaboration has developed a second flavor tagging algorithm based on a deep learning MLP. This algorithm determines the flavor of the tag-side B meson using the PID information and the kinematic properties of the tag-side tracks, without pre-identifying flavor signatures. The deep-learning flavor tagger reaches a higher effective efficiency compared to the category-based algorithm on Belle II MC events. But on Belle MC events, the effective efficiencies are almost identical for both flavor taggers. Further studies are required to understand why the deep-learning flavor tagger outperforms the category-based one on Belle II MC, but not on Belle MC. By comparing the output of the individual categories with the output of the deep-learning algorithm one can check if the deep MLP is learning physical information or MC features. The deep-learning algorithm has not yet been tested on Belle II MC with background. Currently, a validation on Belle MC is ongoing. A systematic comparison of the performance of both flavor taggers would offer a unique possibility to crosscheck their results and to investigate differences between data and MC.

At Belle and at BaBar, the effect of possible wrong reconstructions of B_{sig}^0 on the performance of the flavor tagging algorithms was negligible [29]. At Belle II, the background levels will be much higher and this effect will have to be studied. Moreover, due to the smaller interaction region and the better track reconstruction algorithms, the flavor tagging performance at Belle II can be influenced by the decay time of B_{tag}^0 . This can cause the flavor tagging performance to be correlated with the CP -violation parameters and with the time difference Δt . The correlations will depend on the final experimental conditions at Belle II, and thus further studies will be required.

A major conclusion of this chapter is that the flavor tagger at Belle II has to be trained using events simulated without CP violation. This is needed to counteract the correlation between the CP -violation parameters and the performance of the flavor tagger.

The flavor tagger developed in this thesis is ready to be calibrated and to be used in physics analyses. Depending on the sample sizes, Belle II could pioneer the measurement of corrections to counteract the uncertainties caused by the tag-side interference effect. In this chapter, a possible way to measure these corrections during the calibration was proposed. However, further studies are needed to evaluate the feasibility of such a measurement.

In the Belle II physics program, the flavor tagger plays a crucial role. It is required for most key measurements of CP -violation parameters and of $B^0 - \bar{B}^0$ mixing. Additionally, the output of the flavor tagger can be used as a further continuum suppression variable in measurements of charged and neutral B mesons that do not require flavor tagging. The flavor tagger developed in this thesis has been adopted by the Belle II collaboration as its official flavor tagger. It has been already employed in several performance studies at Belle II, for example in performance studies of the decay channels $B^0 \rightarrow \phi K_S^0$ and $B^0 \rightarrow \eta' K_S^0$ [1], and of the decay channel $B^0 \rightarrow \pi^0 \pi^0$ presented in this thesis. As the results of these studies show, the flavor tagger developed in this thesis will contribute to reach world-leading precision in the determination of the CKM angles ϕ_1 and ϕ_2 , and ultimately to increase Belle II's sensitivity to possible New Physics contributions beyond the Standard Model.

5 Results from Belle II MC with background

In the last stage of this thesis, a new official Belle II MC sample with machine background became available. The track reconstruction algorithms were improved and validated on first Belle II commissioning data samples. In addition, their performance on MC events with simulated background improved considerably. Furthermore, the PID algorithms were also improved. In this chapter, the performance of the category-based flavor tagger is evaluated using the new official MC samples with simulated background assuming the instantaneous design luminosity at Belle II of $8 \cdot 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$. The results of the background simulation have not been validated on Belle II data yet.

5.1 Performance on Belle II MC with background

The new official Belle II MC sample corresponds to events where the B -meson decays were simulated as in the previous official Belle II MC sample described in Sect. 4.5.1: B_{sig}^0 decays to $J/\psi K_S^0$ and B_{tag}^0 decays generically according to the known branching fractions. The CP -violation parameters for the decay of B_{sig}^0 correspond to $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0.69)$ as in the previous sample.

In contrast to the previous official Belle II MC events, the new ones were generated taking into account the expected position and size of the interaction region at SuperKEKB (see Sect. 3.1). Figure 5.1(a) shows the MC distributions of the $\Upsilon(4S)$ production vertex in z -direction at Belle II and Belle. Due to the nano-beam scheme, the width of the expected distribution at Belle II is about twenty times narrower compared to Belle. Figure 5.1(b) shows the corresponding distributions of the decay vertex of B_{tag}^0 in dependence of the B_{tag}^0 flavor at the time of the B_{tag}^0 decay. The asymmetry between B^0 and \bar{B}^0 is caused by the simulated CP violation. Each of the distributions in Fig. 5.1 is normalized such that the area below the curve is equal to unity. For a better visibility, the normalized Belle MC distributions were multiplied by 10.

The background simulation takes into account the beam-induced backgrounds caused by Touschek effect (scattering and loss of beam particles), and by beam-gas scattering, as well as the luminosity-dependent backgrounds caused by Bhabha scattering and by two-photon QED processes. More information about the expected backgrounds at Belle II is given in [105, 137].

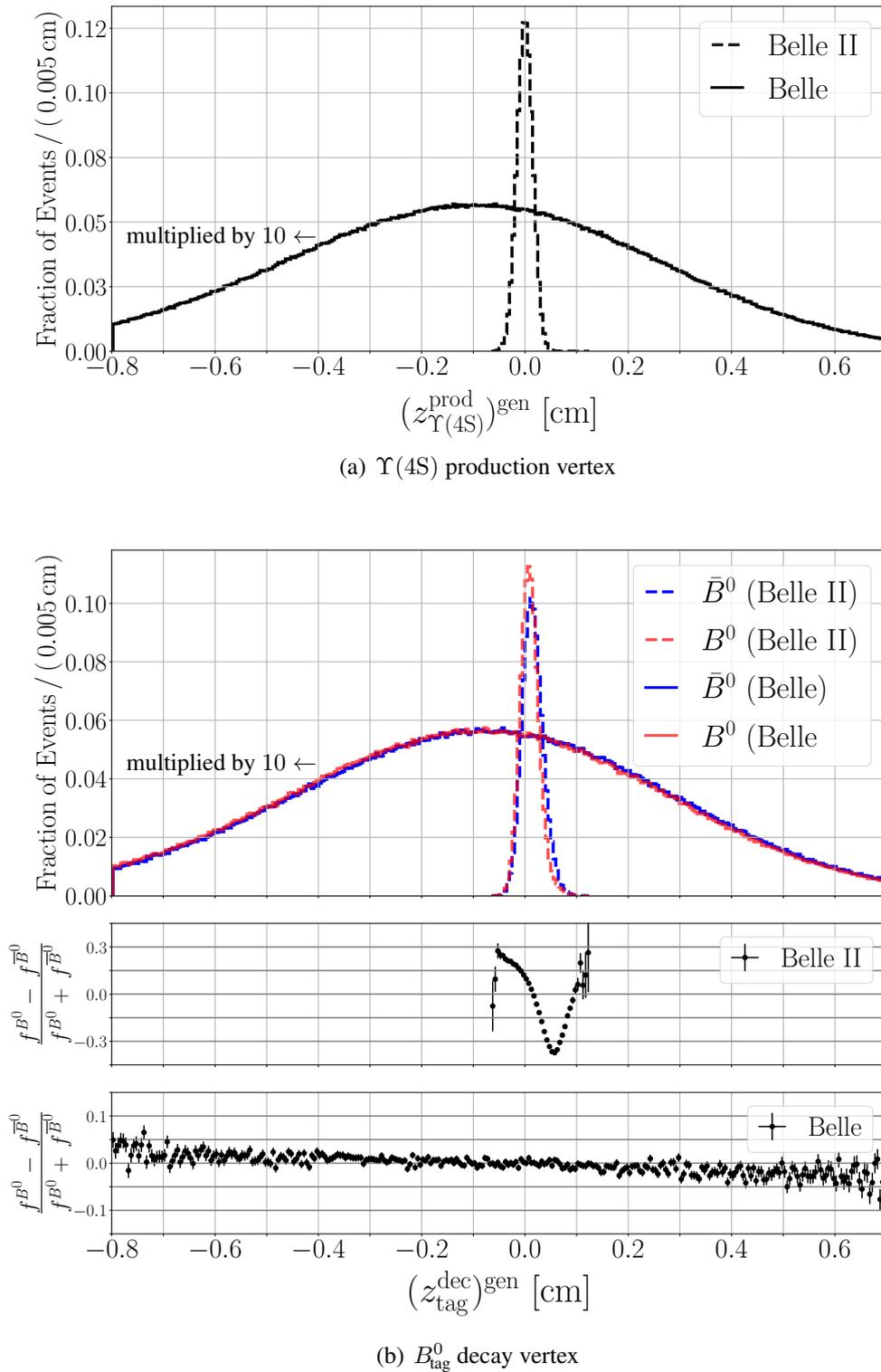


Figure 5.1: Belle II and Belle MC distributions: (a) Production vertex of the $\Upsilon(4S)$ resonance, (b) decay vertex of B_{tag}^0 depending on the B_{tag}^0 flavor at the time of the B_{tag}^0 decay.

The reconstruction of the signal decay channel and the selection of the reconstructed events is performed as described in Sect. 4.5.1. The selected events are again split into three samples: two training samples (one for the event level and one for the combiner level), and one testing sample. The size of each training sample is about 2.5 million events, and the size of the testing sample is about 5 million events.

After training and testing the flavor tagger using the new MC samples, the output $y = q \cdot r$ of the FBDT and the MLP combiners is distributed as in Fig. 5.2 (left). On the right side, Fig. 5.2 shows the distributions of the product $q_{\text{MC}} \cdot q \cdot r$, where q_{MC} is the true B_{tag}^0 flavor at the time B_{tag}^0 decayed. In the region between -0.1 and 0.1 , the distributions show an asymmetry which is caused by the correlation between the CP asymmetry of the B_{tag}^0 -decay vertex (Fig. 5.1(b)) and the distributions of the impact parameters, as it was discussed in Sect. 4.6. The CP asymmetry has an impact also within other r ranges, but to a lesser extent. For the output of the MLP combiner, the central asymmetry is less prominent compared to the FBDT combiner.

Figure 5.3 shows the corresponding values of ε , μ , w , and Δw for each r bin, as well as the linearity checks. The combiners have slightly different performances depending on the r bin. The differences again being largest for the parameter μ . For the values of ε and w , the FBDT

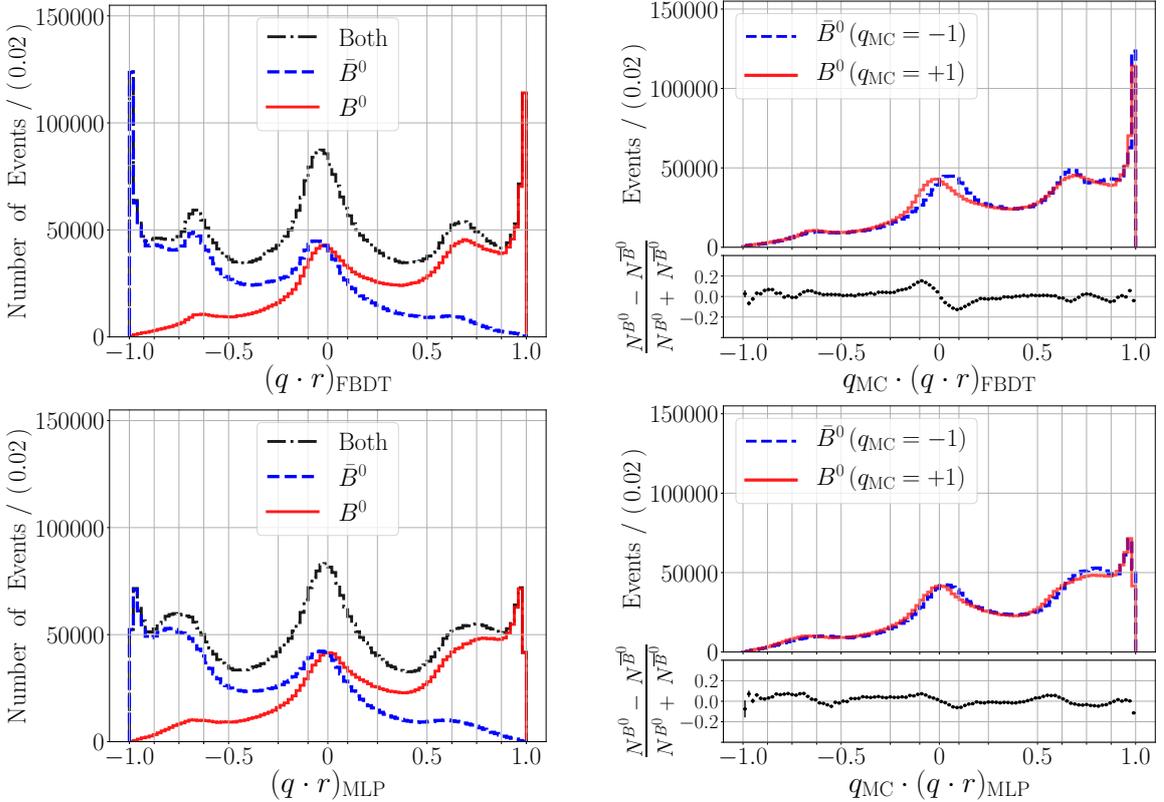


Figure 5.2: Performance of the FBDT and of the MLP combiners on Belle II MC with background (left) combiner output; (right) combiner output multiplied by the true flavor of B_{tag}^0 q_{MC} .

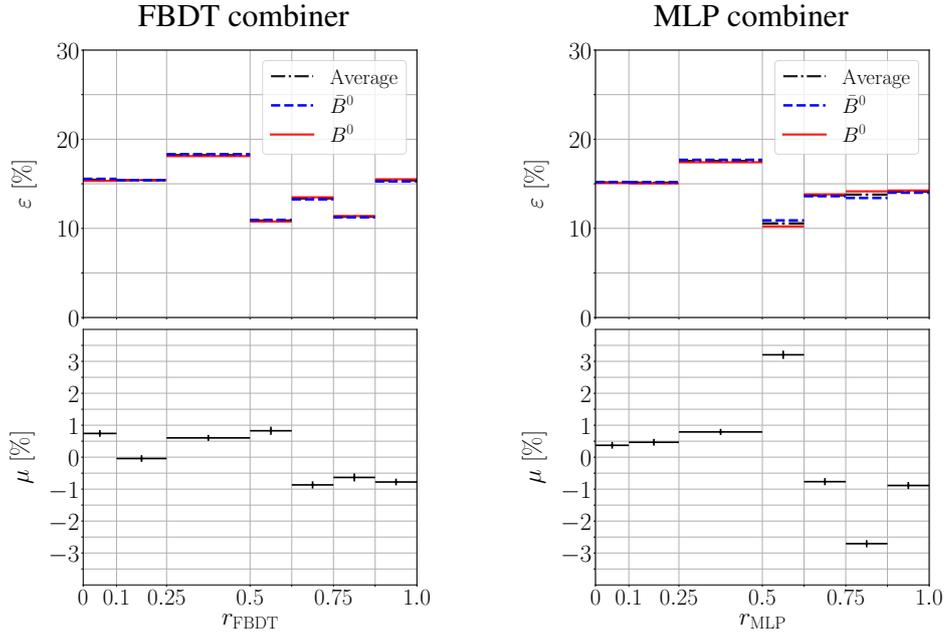
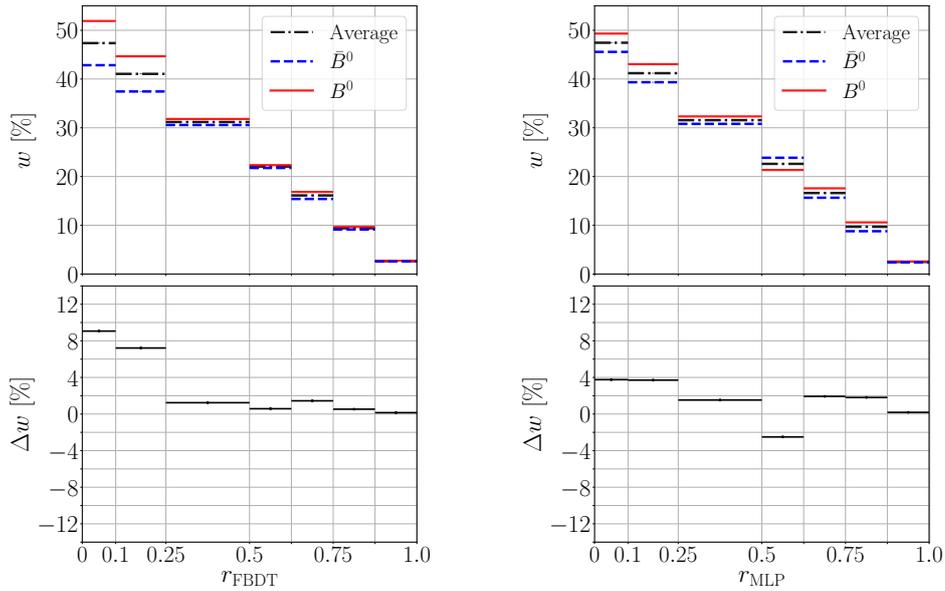
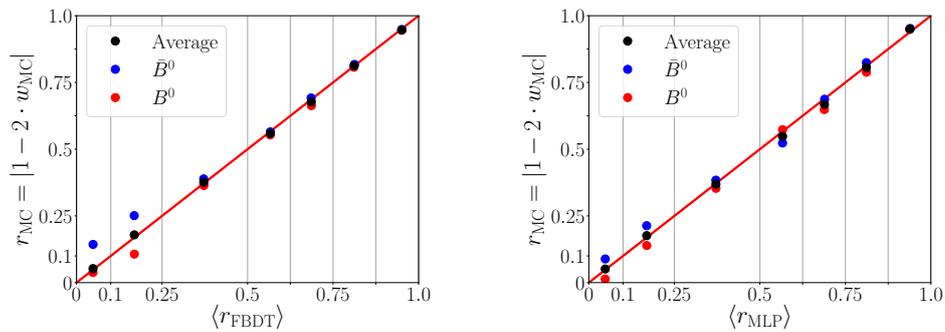
(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0}) / (\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r bin.(c) Correlations between $|r_{MC} = 1 - 2 \cdot w_{MC}|$ and $\langle |q \cdot r| \rangle$ for each r bin.

Figure 5.3: Belle II MC: performance of the (left) FBDT and of the (right) MLP combiners.

and the MLP follow similar patterns. For both combiners, the values of Δw are relatively large for the two lowest r bins. As in the case without background for $\mathcal{S}_{CP} = 0.69$, the linearity check shows that the calibration is correct on average, but the individual average dilutions for B^0 and for \bar{B}^0 events are not correctly calibrated. Both, the large Δw values and the individual miscalibration in the lowest two r bins are associated with the asymmetries in the central $q \cdot r$ regions, i.e. with the CP asymmetry of the B_{tag}^0 -decay vertex.

Considering now $\Delta\varepsilon_{\text{eff}}$, the larger asymmetry at Belle II occurs also because of the effect of the CP violation. As shown in Sect. 4.6, in order to reduce the impact of the CP -violation asymmetry on the performance of the flavor tagger, the algorithm has to be trained using MC events generated without CP violation. Such an official MC sample is not available yet, but will have to be produced in future.

Table 5.1 presents the values of the total effective efficiency ε_{eff} and the difference $\Delta\varepsilon_{\text{eff}}$ for the FBDT and the MLP combiners trained with the new official Belle II MC sample, as well as on Belle MC and on Belle data.

The total effective efficiency on the new Belle II MC is only about 0.2 percent points lower than on Belle MC, despite the fact that at Belle II the background levels are much higher than at Belle. The improvements in the detector, in the reconstruction algorithms, and in the flavor tagger compensate for the higher background, avoiding a considerable deterioration of the flavor tagging performance.

Table 5.1: Performance of FBDT and of MLP combiners on Belle II MC events, on Belle MC events, and on Belle data. The MC events were simulated with background. All values are given in percent considering only statistical uncertainties. The Belle II results were obtained using the latest official Belle II MC.

Sample type	FBDT combiner		MLP combiner	
	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$	$\varepsilon_{\text{eff}} \pm \delta\varepsilon_{\text{eff}}$	$\Delta\varepsilon_{\text{eff}} \pm \delta\Delta\varepsilon_{\text{eff}}$
Belle II MC	33.89 ± 0.04	-2.70 ± 0.10	33.88 ± 0.05	-2.50 ± 0.07
Belle MC	34.18 ± 0.06	-0.18 ± 0.11	34.07 ± 0.06	-0.03 ± 0.11
Belle data	33.5 ± 0.5	—	33.4 ± 0.4	—

5.2 Conclusion

This chapter presented the performance of the Belle II category-based flavor tagger on the newest official Belle II MC sample including background.

After training and testing the flavor tagger with the new MC samples, the performance of the algorithm is roughly as good as for Belle MC. Using the output of the FBDT combiner, the effective efficiency is

$$\varepsilon_{\text{eff}} = (33.89 \pm 0.04)\%.$$

As a cross check, the MLP combiner reaches $(33.88 \pm 0.05)\%$. At Belle II, the expected and simulated background levels are much higher compared to Belle. Nevertheless, the effective efficiency of the flavor tagger is measured at around 34% because of the improved flavor tagger at Belle II, the improved reconstruction algorithms, and the improved detector. In comparison, the Belle flavor tagger reached an effective efficiency around 30% on Belle data. Even under the expected harsh background conditions, Belle II will outperform its predecessor by about 10%.

As in the case of the Belle II MC without background, the results from the new MC show that the built-in CP -asymmetry will certainly influence the flavor tagging performance at Belle II. Undoubtedly, an important conclusion is again that it is mandatory to train the flavor tagger using MC events without built-in CP violation.

At Belle, the effect of the built-in CP asymmetry was negligible because of the wide B_{tag}^0 decay vertex distribution. In contrast, the effect plays a role at Belle II because of the about twenty times narrower (in z -direction) B_{tag}^0 -decay vertex distribution, which is a consequence of the novel nano-beam scheme at SuperKEKB (see Fig. 3.2 in Sect. 3.1).

6 Performance study of the decay

$$B^0 \rightarrow \pi^0 \pi^0$$

6.1 Introduction

With the available measurements, an isospin analysis of $B \rightarrow \pi\pi$ decays leads to eight different solutions for the unitarity angle ϕ_2 (see Sect. 2.10.2). The only way to reduce this ambiguity is to include the missing observable $\mathcal{S}^{\pi^0\pi^0}$, for which a time-dependent CP -violation analysis of the decay $B^0 \rightarrow \pi^0\pi^0$ is required. This analysis needs a precise determination of the B^0 -decay vertex position, which cannot be achieved in the dominant four-photon final state $B^0 \rightarrow \pi^0\pi^0 \rightarrow 4\gamma$. Events with photons converting in the innermost detector material, or with Dalitz $\pi^0 \rightarrow e^+e^-\gamma$ decays provide electron-positron pairs to reconstruct the decay vertex of the signal B^0 meson. Such events, however, are very rare. And large sample sizes are required to make use of them. At present, the size of the Belle and the BaBar data samples is insufficient. And at LHCb, $B^0 \rightarrow \pi^0\pi^0$ decays are not reconstructed.

The Belle II experiment aims at collecting a data sample around fifty times larger than the sample of Belle. The Belle collaboration showed that, with such a large sample, a measurement of $\mathcal{S}^{\pi^0\pi^0}$ is in principle feasible [138]. Given the novel experimental conditions, and the completely new reconstruction and analysis algorithms, Belle II still had to show that its performance is good enough to measure it.

This chapter presents a study of the Belle II performance for the measurement of the CP -violation parameters $\mathcal{A}^{\pi^0\pi^0}$ and, for the first time, $\mathcal{S}^{\pi^0\pi^0}$, together with the branching fraction $\mathcal{B}^{\pi^0\pi^0}$. The proposed method of measuring $\mathcal{S}^{\pi^0\pi^0}$ uses photons converted in the beam pipe and in the pixel detector (PXD), as well as Dalitz π^0 decays. The method of measuring $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{B}^{\pi^0\pi^0}$ uses only photons reconstructed as clusters in the electromagnetic calorimeter to perform a time-integrated CP -violation analysis.

In the following, the generation of MC events and the analysis strategy are explained. Next, the reconstruction of the $B^0 \rightarrow \pi^0\pi^0$ decay chain is described. The subsequent section shows how the decay-vertex positions of the signal and the tag-side B^0 mesons are reconstructed to measure the time difference Δt . The suppression of non- $B\bar{B}$ events and the final selection are presented afterward. The flavor of the tag-side B^0 meson was determined using the flavor tagger developed in this thesis. The last section is dedicated to the time-integrated and the time-dependent analyses performed on simulated events.

6.2 Generation of events

In addition to the $\Upsilon(4S)$ resonance, at B factories, there are contributions from non-resonant processes, where lighter quark and antiquark pairs ($u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$) as well as lepton antilepton pairs (e^+e^- , $\mu^+\mu^-$, $\tau^+\tau^-$) are produced. For the analysis of B -meson decays, events with e^+e^- and with $\mu^+\mu^-$ pairs can be easily filtered out because of their low track multiplicity and because of their topology: the two leptons have high momenta around half the total accelerator energy, and they are emitted back-to-back in the center-of-mass frame. Thus, for the analysis of B -meson decays the relevant backgrounds are events with lighter $q\bar{q}$ pairs (referred to as continuum events), $\tau^+\tau^-$ events, and events with $B\bar{B}$ pairs that do not contain the signal decay. For $B^0 \rightarrow \pi^0\pi^0$, as for most charmless B decays, continuum events are the dominant background source.

The analyses presented in this chapter base on Belle II MC samples. The samples of continuum, $\tau^+\tau^-$, B^+B^- and $B^0\bar{B}^0$ events were produced in an official MC campaign. The size of the continuum sample corresponds to an integrated luminosity of 1.8 ab^{-1} , and the sizes of the B^+B^- and $B^0\bar{B}^0$ samples to 4 ab^{-1} . Signal MC events were generated privately with the same software release used for the official campaign. At the time when the studies presented in this chapter were performed, the reconstruction algorithms at Belle II were not optimized for events with background. Furthermore, realistic background simulations were not available. Therefore, the official and the private MC events used for this study were generated without beam-induced and without luminosity-dependent backgrounds.

The signal MC events correspond to $B^0\bar{B}^0$ pairs in which one meson decays into the signal channel $\pi^0\pi^0$, while the other decays into any possible final state according to the known branching fractions [28]. The π^0 mesons decay also according to the known branching fractions [28]; they decay to about 98.823% into two photons, and to about 1.174% into Dalitz decays $\pi^0 \rightarrow e^+e^-\gamma$.

The EvtGen generator is used to simulate B -meson decays [139]. For the signal events, the SSS_CP model of EvtGen is used to simulate the decay of the scalar B^0 -meson into the two scalar π^0 mesons. In this model, the decay amplitudes for B^0 and for \bar{B}^0 mesons, A and \bar{A} , and the CP -violation quantity λ_{CP} are parametrized as

$$A = |A| e^{i \arg A}, \quad \bar{A} = |\bar{A}| e^{i \arg \bar{A}}, \quad \lambda_{CP} = e^{-2i\alpha} \frac{\bar{A}}{A},$$

where $|A|$, $\arg A$, $|\bar{A}|$, $\arg \bar{A}$, and α are free input parameters. The simulated values of the CP -violation parameters correspond to

$$\mathcal{A}_{CP} = \frac{|\lambda_{CP}|^2 - 1}{|\lambda_{CP}|^2 + 1} = \frac{|\bar{A}|^2 - |A|^2}{|\bar{A}|^2 + |A|^2},$$

$$\mathcal{S}_{CP} = \frac{2 \cdot \text{Im}(\lambda_{CP})}{|\lambda_{CP}|^2 + 1} = \frac{2 \cdot |\bar{A}| \cdot |A| \cdot \sin(\arg \bar{A} - \arg A - 2\alpha)}{|\bar{A}|^2 + |A|^2}.$$

To match specific values of \mathcal{A}_{CP} and \mathcal{S}_{CP} , the free input parameters of the model can be modified. For this performance study, six signal MC samples were generated with different combinations of \mathcal{A}_{CP} and \mathcal{S}_{CP} . Figure 6.1 shows the Δt^{gen} distributions corresponding to the different samples. The values of \mathcal{A}_{CP} and \mathcal{S}_{CP} of the first sample (Fig. 6.1(a)) were chosen to be compatible with the world average of \mathcal{A}_{CP} and the predicted value of \mathcal{S}_{CP} determined by the CKMfitter group [71]. In the second sample (Fig. 6.1(b)), the value of \mathcal{A}_{CP} is compatible with the world average of the Particle Data Group [28]. In the other samples (Figs. 6.1(c) to 6.1(f)), the value of \mathcal{A}_{CP} corresponds to the latest Belle measurement [140], and the four different values of \mathcal{S}_{CP} are compatible with the ϕ_2 solutions of the $B \rightarrow \pi\pi$ isospin analysis at Belle (see Sect. 7.4.2).

At Belle II, the interaction of particles with the detector material is simulated using GEANT4 [141]. After the simulation, ca. 2% of the signal events have at least one conversion inside the beam pipe, and ca. 1% has at least one conversion in the PXD. Additionally, about 6% of the signal events contain at least one converted photon in the SVD volume outside of the PXD. These numbers correspond to events in which the e^+e^- pairs and all the photons stemming from the signal B -meson traverse the detector within the acceptance. Figure 6.2 shows the MC vertices of the converted photons inside of the PXD in the xy -plane.

Considering π^0 s undergoing Dalitz decay, about 2% of the signal events contain at least one of them. This number corresponds to events in which the Dalitz e^+e^- pairs and all the photons from the signal B -meson traverse the detector within the acceptance.

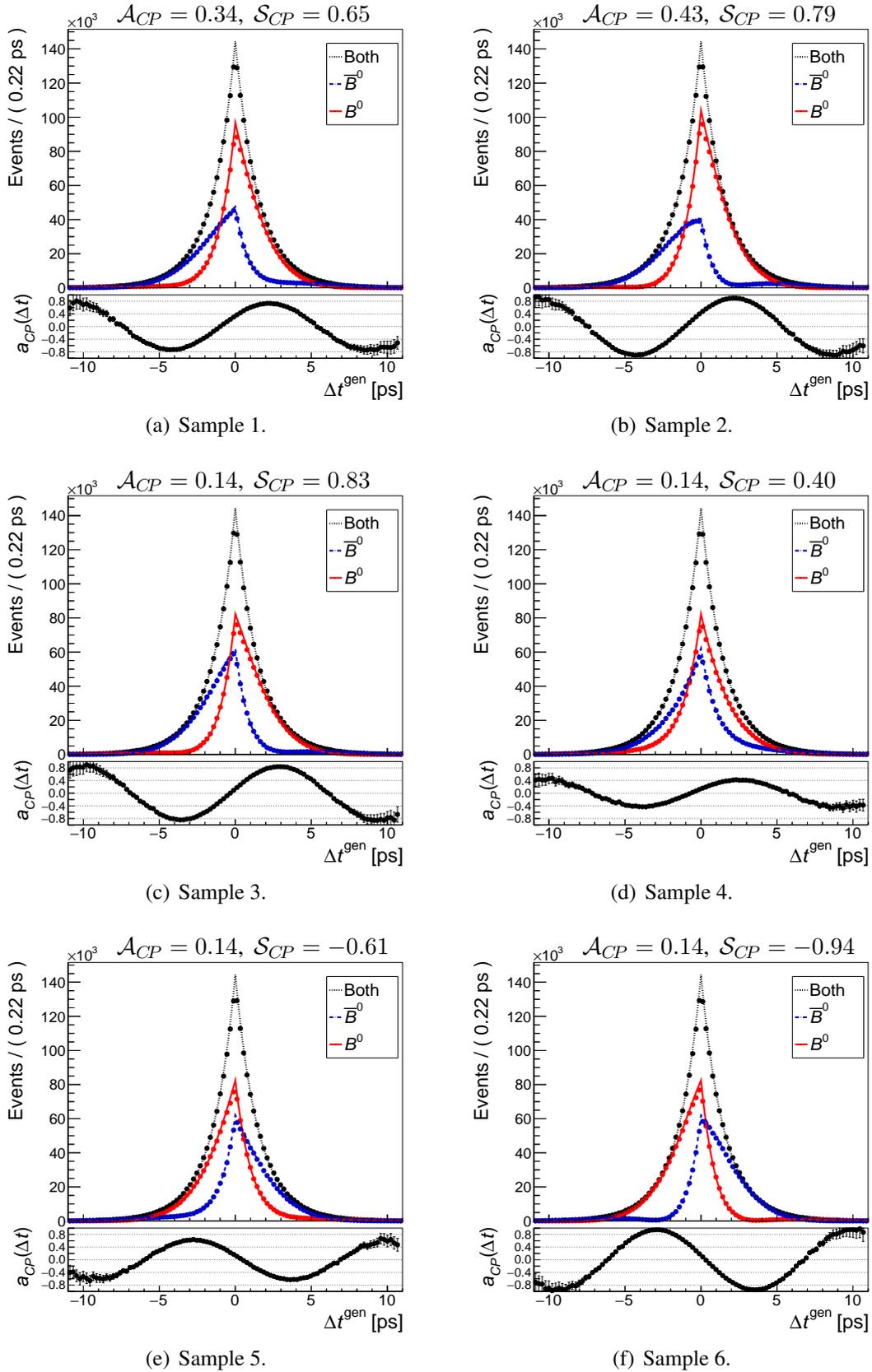


Figure 6.1: MC distributions of $\Delta t^{\text{gen}} = t_{\text{sig}}^{\text{gen}} - t_{\text{tag}}^{\text{gen}}$ for the six generated MC samples. The distributions are shown for all events, for events where B_{tag}^0 is a B^0 , and for events where B_{tag}^0 is a \bar{B}^0 .

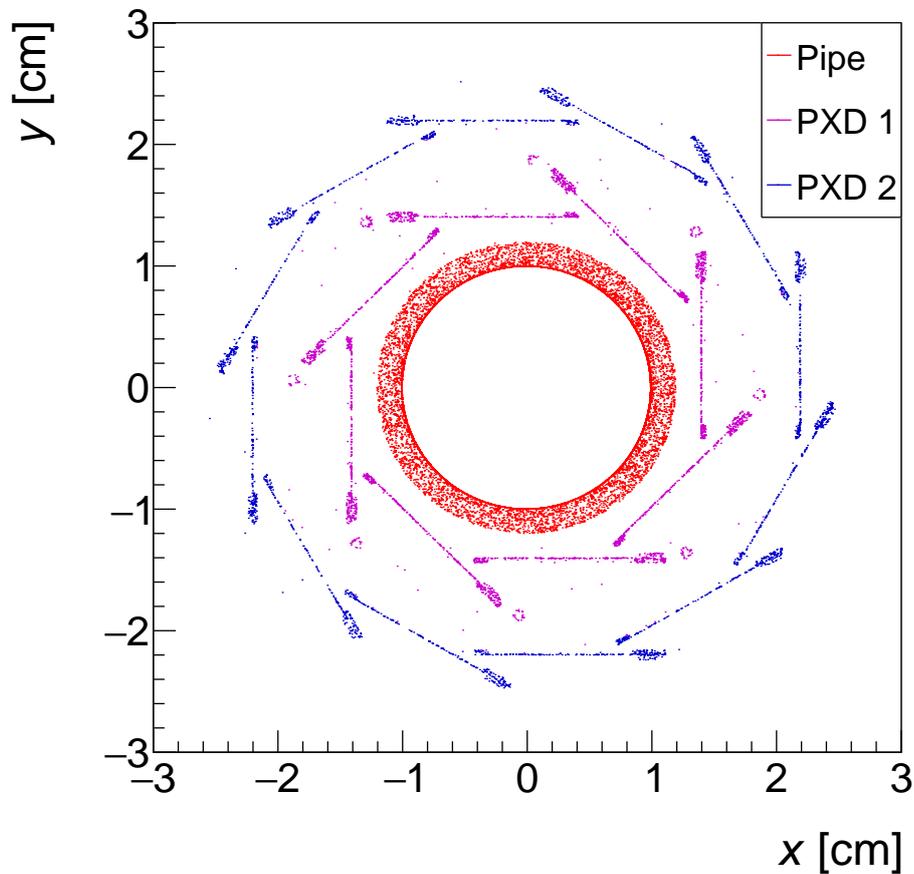


Figure 6.2: Conversion vertices in the xy -plane inside the PXD. Conversions in the beam pipe as well as in the first and second layers of the PXD are shown in red, magenta and blue, respectively. Most of the conversions in the PXD occur in the supporting silicon rim, and in the parts where the control switchers are mounted.

6.3 Analysis strategy

The analysis strategy in this chapter follows the concept of a “blind analysis”. In a blind analysis, the whole analysis procedure is optimized on MC events and eventually also on data outside of the signal region and on control samples extracted from the data. The first step in the procedure is to reconstruct and to select the events. The second step is to extract the physical parameters of interest from the distributions of observables. The extraction of the physical parameters requires the formulation of a statistical model describing the distributions of the observables.

The analysis procedure and the statistical model are validated by performing so-called “pseudo-experiments”. In a pseudo-experiment, the distributions of the observables are obtained from fully simulated MC events that were reconstructed and selected, or from probabi-

lity density functions modeled on such events. The MC events for the pseudo-experiment must be statistically independent from those used to optimize the analysis. The statistical model is then fitted to the distributions of the observables and the physical parameters are extracted. After performing a sufficiently large number of pseudo-experiments, the analysis is validated by checking that the results for the physical parameters are not statistically biased, and that the statistical errors are neither over- nor under estimated.

Once the analysis is validated, it is applied to real data and the physical parameters are extracted. This action is referred to as “unblinding” or “opening the box”. The purpose of the blind analysis technique is to prevent from biased results due to the influence of the preferences and the expectations of the analyzer. After opening the box, systematic effects affecting the measurement are studied using control samples extracted from real data to estimate the systematic uncertainties¹.

In the absence of real Belle II data, the analysis in this chapter fulfills all the steps that have to be performed on MC events. As a final step, the expected statistical uncertainties on the physical parameters of interest are estimated from the results of the pseudo-experiments.

6.4 Reconstruction of the signal side

To perform the analysis of the generated MC events, the following three decay modes are considered as signal:

1. $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0(\rightarrow \gamma\gamma) \pi_{\gamma\gamma}^0(\rightarrow \gamma\gamma)$,
2. $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0(\rightarrow e^+e^-\gamma) \pi_{\gamma\gamma}^0(\rightarrow \gamma\gamma)$,
3. $B_{\text{sig}}^0 \rightarrow \pi_{\gamma_C\gamma}^0(\rightarrow \gamma_C(\rightarrow e^+e^-\gamma) \pi_{\gamma\gamma}^0(\rightarrow \gamma\gamma)$.

The fraction of signal events with more than one conversion or more than one π_{Dal}^0 decay is less than 0.5%. Thus, such events are not reconstructed.

For each event, signal B -meson candidates are found by reconstructing the whole decay chain for the three considered signal decay modes. The reconstruction is performed by applying some selection criteria. These criteria aim at enhancing the number of true candidates, i.e. those who match the generated MC particles, with respect to the number of fake candidates resulting from random combinations of particles. The goal is to remove fake candidates while conserving as many signal candidates as possible. Because of the similarities between Belle and Belle II, some selection criteria were adopted from previous analyses at Belle. In the following, the reconstruction and the selection criteria are explained in the order they are applied. Percentages in this section are given with a precision about ± 0.5 percent points.

γ and $\pi_{\gamma\gamma}^0$ candidates

Photons (γ) and neutral pions ($\pi_{\gamma\gamma}^0$) are reconstructed in a similar way as in previous analyses at Belle [140, 142, 143]. Photons correspond to neutral clusters in the electromagnetic

¹Some systematic uncertainties are not evaluated on control samples (see discussion in Sect. 7.3).

calorimeter. At Belle II, the ECL crystals are the same as at Belle. Obviously, due to the novel background levels and due to the upgrades of the ECL readout system at Belle II, a dedicated optimization of the photon selection has to be performed. In the absence of an optimization for the Belle II ECL at the time of this analysis, the photon-selection criteria at Belle were taken as a reference: an ECL cluster that is not associated with a track, i.e. a neutral cluster, is selected as photon candidate if its energy is greater than 50 MeV, 100 MeV and 150 MeV in the barrel, front-end cap, and back-end cap of the ECL, respectively.

The reconstruction of $\pi_{\gamma\gamma}^0$ is performed using pairs of photons with invariant masses in the range

$$105 \text{ MeV}/c^2 < M_{\gamma\gamma} < 165 \text{ MeV}/c^2 \quad (6.1)$$

corresponding to about $\pm 2.5\sigma$ around the nominal π^0 mass. The standard deviation σ is obtained by fitting a Gaussian function to the mass distribution of $\pi_{\gamma\gamma}^0$ candidates corresponding to true pions decaying into two photons. Figure 6.3 shows the mass distributions of true and of fake $\pi_{\gamma\gamma}^0$ candidates as well as the fit to the mass distribution of true candidates. The selection criterion on $M_{\gamma\gamma}$, shown with red lines, keeps 97% of the signal candidates.

To reduce the number of fake pions, $\pi_{\gamma\gamma}^0$ candidates with small helicity angles such that $|\cos(\theta_H)| > 0.95$ are rejected. The helicity angle θ_H is defined as the angle between the π^0 boost direction in the laboratory frame and the momentum of one of the two γ daughters in the π^0 rest frame. The γ used for the calculation is chosen randomly. Figure 6.4 (left) illustrates the definition of θ_H . And Fig. 6.4 (right) shows the distribution of $\cos(\theta_H)$ for true and for fake candidates. Since neutral pions are pseudo-scalar spin-0 particles, the momenta of their decay products have no directional preference. Thus, the $\cos(\theta_H)$ distribution for true neutral pions is uniform. Several fake pions result from the combination of a soft photon moving parallel or antiparallel to a hard one in the π^0 frame. They cause the peaks at -1 and at 1 in the $\cos(\theta_H)$ distribution. The requirement at $|\cos(\theta_H)| > 0.95$ was adopted from previous analyses at Belle [140, 142, 143] since it was optimized based on the kinematics of the decay. The cut removes 20% of the fake candidates, and only 3% of the signal candidates. The final fraction of true $\pi_{\gamma\gamma}^0$ candidates passing the selection is 94%.

γ_C , $\pi_{\gamma_C\gamma}^0$ and π_{Dal}^0 candidates

The reconstruction and the selection of converted photons (γ_C), π^0 s with one converted photon ($\pi_{\gamma_C\gamma}^0$), and Dalitz decays (π_{Dal}^0) is independent of the $\pi_{\gamma\gamma}^0$ selection. The first step is the reconstruction of electron-positron pairs. They are reconstructed using pairs of oppositely charged tracks, where both tracks have a combined electron PID likelihood of $\mathcal{L}_e > 0.8$ (see Sect. 3.5) and an impact parameter $d_0 < 5$ cm. Figure 6.5 (left) shows the \mathcal{L}_e distribution for all tracks, for tracks that do not correspond to electrons and positrons, and for true electron and positron tracks. Figure 6.5 (right) shows the d_0 distributions for tracks originating from a conversion or a Dalitz decay inside the PXD volume, and for all other tracks. The PXD volume (inside the outer PXD radius) encloses the PXD and the beam pipe. The red lines in the plots show the PID and the d_0 requirements. By applying the PID requirement, 86% of the tracks that are

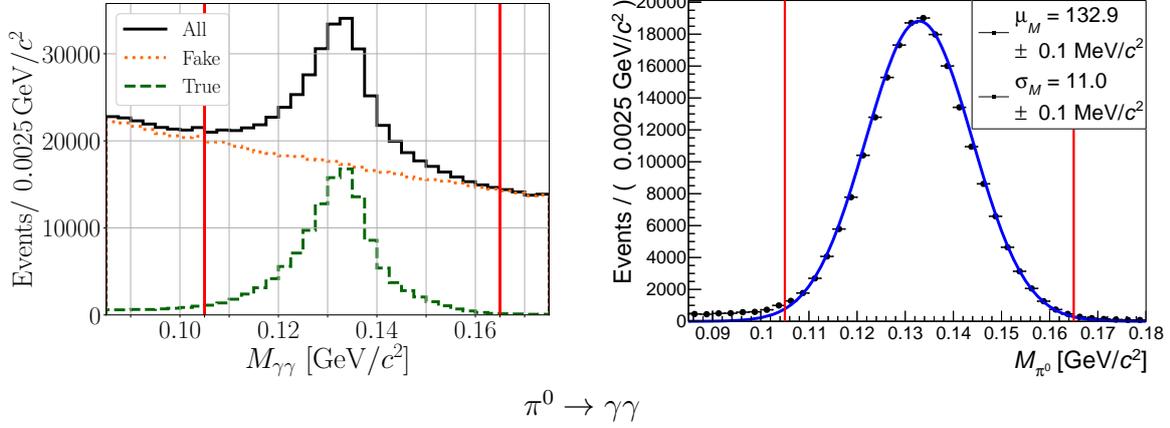


Figure 6.3: Left: mass distributions of π^0 candidates reconstructed from two photons. The distributions are shown for all candidates, for true candidates and for fake candidates. Right: fit to the mass distribution of true candidates. The red vertical lines correspond to the selection criteria. The fit function on the right is a single Gaussian function. The parameters μ_M and σ_M are the mean and the standard deviation of the Gaussian function. For later comparison, the 95% trimmed mean and standard deviation are $\mu_M^{(95)} = 132.9 \pm 0.1$ and $\sigma_M^{(95)} = 10.5 \pm 0.1$.

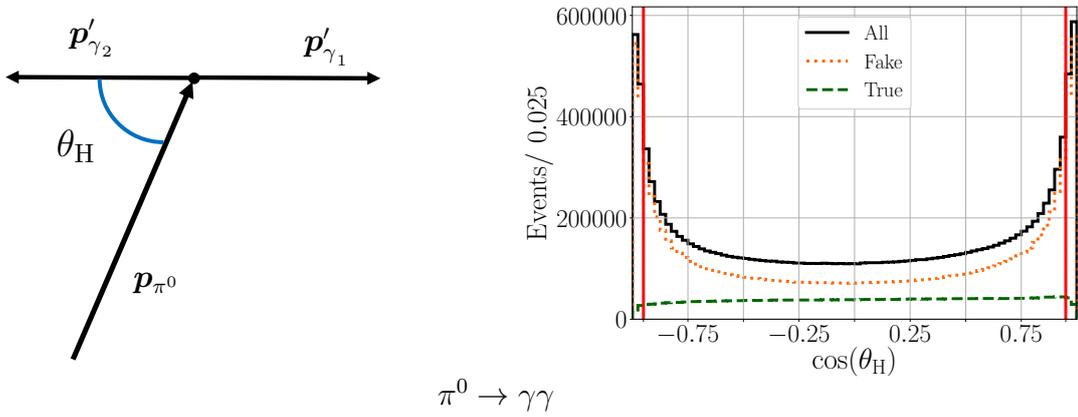


Figure 6.4: Left: the definition of the helicity angle θ_H . The vector p_{π^0} stays for the three-momentum of the pion in the lab frame, and the vectors p'_{γ_1} , and p'_{γ_2} for the momenta of the photons in the rest frame of the pion. The photons are randomly labeled γ_1 or γ_2 . Right: Distribution of $\cos(\theta_H)$ for π^0 candidates reconstructed from two photons. The distribution is shown for all candidates, for true and for fake candidates. The red vertical lines correspond to the selection criteria.

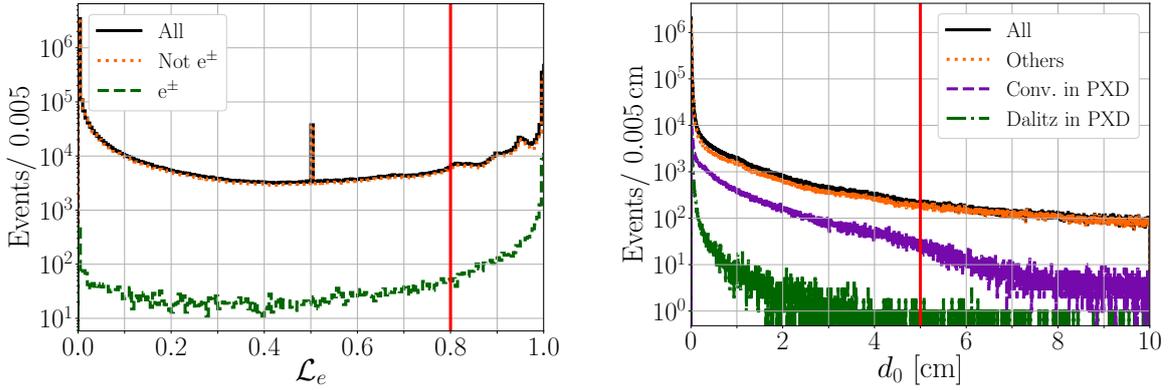


Figure 6.5: Left: electron PID likelihood for tracks that do not correspond to electrons or positrons, and for electron and positron tracks. Right: the d_0 impact parameter for all tracks, for tracks originating from a conversion or a Dalitz decay inside the PXD volume and for all other tracks. The red lines correspond to the selection criteria: a track is selected as electron candidate if its PID likelihood is larger than 0.8, and if its d_0 value is smaller than 5 cm.

not associated with an electron or a positron are removed, and about 83% of the true electron tracks are kept. Additionally, the d_0 requirement removes background tracks from conversions or from Dalitz π^0 s far outside of the PXD volume. Most of the π^0 s undergoing Dalitz decay outside of the PXD come from $K_S^0 \rightarrow \pi^0\pi^0$ decays. To a lesser extent they come also from baryonic decays like $\Lambda \rightarrow n\pi^0$ or $\Sigma^+ \rightarrow p\pi^0$. Of the true electrons and positrons from conversions or from Dalitz- π^0 decays inside the PXD, more than 99% pass the d_0 requirement, and about 85% pass both the d_0 and the PID requirements. The fractions are the same for both kinds of candidates.

Electron-positron pairs are selected if one or both tracks have at least one PXD hit. About 62% of the pairs from a conversion inside the PXD fulfill this requirement. And for the pairs from a Dalitz decay inside the PXD, about 96% fulfill it. The difference between the fractions of pairs from conversions and from Dalitz π^0 s is because some of the conversions occur in the PXD material outside the sensors, without producing PXD hits. Events with electron-positron pairs without PXD hits, but with SVD hits, are not suitable for a time-dependent CP analysis since their time resolution was found to be at least a factor 3 worse than the time resolution of the events passing the one-hit PXD requirement.

Converted photons γ_C are reconstructed using e^+e^- pairs. The γ_C three-momentum is obtained by adding the three-momenta of the e^+ and the e^- tracks. The reconstruction of neutral pions with one converted photon $\pi_{\gamma_C\gamma}^0$ is performed using a reconstructed converted photon γ_C and a photon γ corresponding to a neutral ECL cluster. Dalitz candidates π_{Dal}^0 are reconstructed using three particles for each candidate: an e^+e^- pair and a photon corresponding to a neutral cluster. For this, the four-momenta of the three particles are added.

All three kinds of pions are reconstructed within the same mass range chosen for $\pi_{\gamma\gamma}^0$ candidates (eq. (6.1)). The reason is that the mass distributions of true $\pi_{\gamma_C\gamma}^0$ and of true π_{Dal}^0

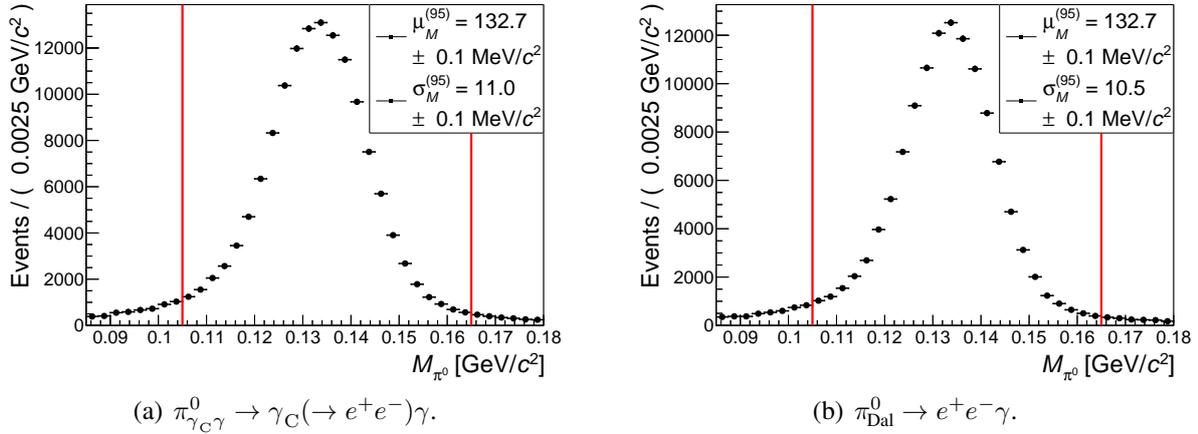


Figure 6.6: Mass distributions of reconstructed true π^0 candidates: (a) neutral pions with one converted photon, (b) neutral pions undergoing Dalitz decays. The parameters $\mu_M^{(95)}$ and $\sigma_M^{(95)}$ are the 95% trimmed mean and standard deviation of the sample.

candidates have almost the same mean and the same standard deviation as the mass distribution of true $\pi_{\gamma\gamma}^0$ candidates. Figure 6.6 shows the mass distributions of true $\pi_{\gamma\gamma}^0$ and of true π_{Dal}^0 candidates. The distributions have pronounced tails due to the radiation losses of the electrons and the positrons, and, in the case of $\pi_{\gamma_C\gamma}^0$ candidates, due to the momentum transfer to the nucleus (or electron) involved in the conversion.

For a neutral pion undergoing a Dalitz decay and for a neutral pion with one converted photon, the number of final state particles is the same. The final state particles are almost in the same kinematic phase space; and the topology of the decays are very similar if at least one of the e^+e^- tracks has a PXD hit. When π_{Dal}^0 candidates are reconstructed, also π^0 s with conversions pass the selection criteria. Thus, the number of true π_{Dal}^0 candidates is almost the number of π_{Dal}^0 candidates corresponding to neutral pions with one converted photon. These candidates are considered as signal candidates as well. The same applies for reconstructed $\pi_{\gamma_C\gamma}^0$ candidates: many of them are in reality Dalitz π^0 s. And they are also considered as signal as well.

For $\pi_{\gamma_C\gamma}^0$ and π_{Dal}^0 candidates, the helicity angle θ_H is defined as the angle between the π^0 boost direction in the laboratory frame and the momentum of the neutral ECL cluster γ in the π^0 rest frame. Candidates with helicity angles such that $|\cos(\theta_H)| > 0.95$ are rejected. The requirement is the same as for $\pi_{\gamma\gamma}^0$ candidates because the momenta of the π^0 -decay products are always isotropically distributed and thus the shape of the signal distribution is in principle the same.

Figure 6.7 shows the $\cos(\theta_H)$ distributions for true and for fake $\pi_{\gamma_C\gamma}^0$ and π_{Dal}^0 candidates. The asymmetry in the case of $\pi_{\gamma_C\gamma}^0$ and π_{Dal}^0 candidates occurs because tracks are less affected by low momentum background in comparison with neutral clusters (the momentum threshold for tracks is higher than for neutral clusters). The fraction of fake candidates that gets removed by

the helicity requirement is about 30% for $\pi_{\gamma C\gamma}^0$ candidates, and about 20% for π_{Dal}^0 candidates. The fraction of true candidates (including π^0 s with conversions and Dalitz π^0 s) passing the helicity and the invariant mass selection is about 90%. The number is the same if the candidates are reconstructed as $\pi_{\gamma C\gamma}^0$ or as π_{Dal}^0 candidates.

B_{sig}^0 candidates

In the final step, B_{sig}^0 candidates are reconstructed requiring that the beam-constrained mass

$$m_{\text{bc}} = \sqrt{(E_{\text{beam}}^*)^2 - (p_B^*)^2},$$

is larger than $5.26 \text{ GeV}/c^2$. And the beam-constrained energy

$$\Delta E = E_B^* - E_{\text{beam}}^*, \quad (6.2)$$

is required to be larger than -0.3 GeV and smaller than 0.2 GeV . In the equations above, the beam energy $E_{\text{beam}}^* = \sqrt{s}/2$ is half the total energy in the $\Upsilon(4S)$ frame, and E_B^* and p_B^* are the reconstructed energy and momentum of the B_{sig}^0 candidate in the $\Upsilon(4S)$ frame.

The requirements on m_{bc} and on ΔE are applied for the three considered signal decay modes. The use of m_{bc} and ΔE exploits the precise knowledge on the beam energy at B factories. Because of energy conservation, the energy of each B meson in the $\Upsilon(4S)$ frame is half the total energy.

The resolution of m_{bc} is determined by the resolution of the beam energy (several MeV) and is much better than the resolution of the B mass determined from the four momenta of the reconstructed decay products. Naturally, m_{bc} has an upper limit of $5.3 \text{ GeV}/c^2$ corresponding to the maximum beam energy.

The resolution of ΔE is dominated by the uncertainty on the reconstructed energy of the B meson. Its distribution has a tail towards negative values because of systematic losses in the photon reconstruction. Therefore, the requirement on ΔE is asymmetric. Both observables, m_{bc} and ΔE , are used in the maximum-likelihood fits which will be described in Sect. 6.10. The respective distributions will be shown there and in Sect. 6.7.

The requirements on m_{bc} and on ΔE used in this chapter were applied previously by Belle in the latest time-integrated analysis of $B^0 \rightarrow \pi^0\pi^0$ [140]. For each of the three considered decay modes, about 94% of the true B_{sig}^0 candidates fulfill them.

Coming back to the discussion about the partial overlap between π_{Dal}^0 and $\pi_{\gamma C\gamma}^0$ candidates, this effect happens also for B_{sig}^0 candidates: many B_{sig}^0 candidates with a reconstructed Dalitz π^0 are in reality B^0 s with a converted photon. And the same applies for B_{sig}^0 candidates with a reconstructed converted photon. For B_{sig}^0 candidates with two $\pi_{\gamma\gamma}^0$ candidates (four-photon final state), there is no candidate corresponding to a decay with one converted photon or with one Dalitz π^0 . The same applies for B_{sig}^0 candidates with one Dalitz π^0 or with one converted photon: none of them matches a true B^0 decaying into a four-photon final state.

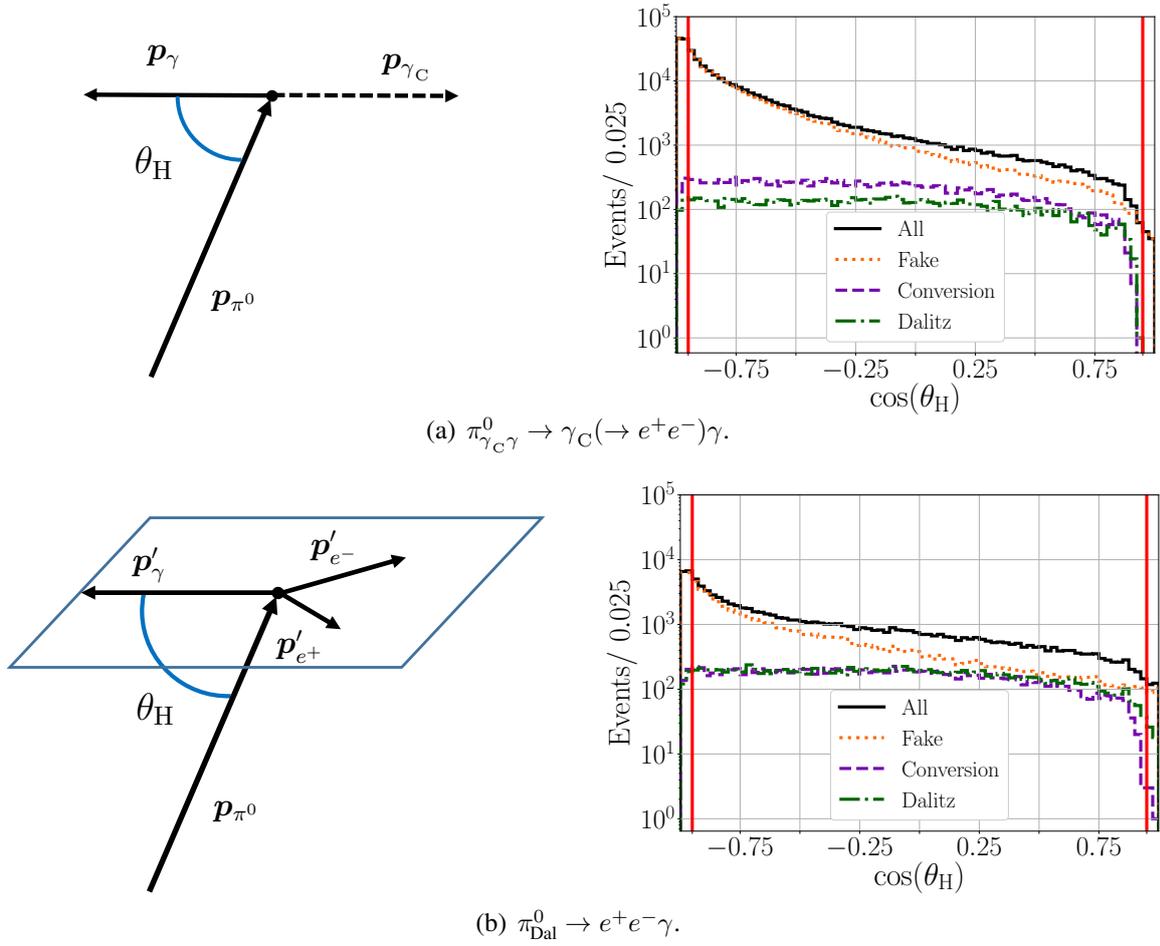


Figure 6.7: The helicity angle θ_H for (a) neutral pion candidates reconstructed with one converted photon and (b) neutral pion candidates reconstructed as Dalitz π^0 s. Left: the definition of the helicity angle. The vector \mathbf{p}_{π^0} stays for the three-momentum of the pion in the lab frame, and the vectors \mathbf{p}'_{γ} , \mathbf{p}'_{γ_c} , \mathbf{p}'_{e^-} , \mathbf{p}'_{e^+} for the momenta of the ECL-cluster photon, the converted photon, the electron and the positron in the rest frame of the pion. Right: Distributions of $\cos(\theta_H)$ for all candidates, for the two signal candidates and for fake candidates. The red vertical lines correspond to the selection criteria.

6.5 Distinguishing between Dalitz π^0 s and converted photons

Candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ and as $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$ contain candidates with a true Dalitz π_{Dal}^0 decay and with a true converted photon in about the same proportion. This was determined by checking the MC information. To perform a CP -violation analysis, one needs to distinguish the two kinds of signal events because they have different vertex resolutions, as it will be shown in the next section, and because the relative fraction of conversions to Dalitz decays can considerably differ between real data and MC events. Considering all candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ (or as $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$) as a single kind of signal candidates, as is done in this thesis, could potentially introduce sizable systematic effects.

To distinguish the two types of signal candidates, an FBDT method [115] was trained with four input variables that provide discrimination power. These input variables make use of the information related to the geometrical properties of the reconstructed e^+e^- pair, for which at least one of the tracks is required to have a PXD hit. The first two variables, r_L and r_U , correspond to the possible two solutions for the e^+e^- vertex in the xy -plane ($r\phi$ -plane): r_L is the solution closest to the IP and r_U the farthest one. For the calculation, the e^+e^- tracks are considered as two intersecting circles with radii r_1 and r_2 . An illustration of the approach is shown in Figure 6.8. The line connecting the two intersection points crosses the line connecting the centers \mathbf{c}_1 and \mathbf{c}_2 of the two circles at the point

$$\mathbf{r}_C = \mathbf{c}_1 + \frac{\mathbf{c}_2 - \mathbf{c}_1}{2} \left(1 + \frac{r_1^2 - r_2^2}{|\mathbf{c}_2 - \mathbf{c}_1|^2} \right).$$

One can then write

$$\mathbf{r}_{U,L} = \mathbf{r}_C \pm y \cdot \mathbf{n},$$

where \mathbf{n} is a unit vector perpendicular to the line connecting the centers of the two circles. If there are two intersections, y is a real number larger than zero. If there is only one intersection $y = 0$. In case that there is no intersection, the values of r_L and r_U are set to be equal r_C .

The fourth and the fifth variables correspond to the angular differences $\Delta\theta_{e^+e^-} = \theta_{e^+} - \theta_{e^-}$ and $\Delta\phi_{e^+e^-} = \phi_{e^+} - \phi_{e^-}$, where the angles θ and ϕ for each track are calculated using the momentum at the respective point of closest approach to the IP; at Belle II, the tracks are represented by the helix parameters calculated at the point of closest approach to the IP (see Sect. 3.3).

Given the momentum $\mathbf{p} = (p_x, p_y, p_z)^T$ for the e^+ and the e^- tracks, the angular difference $\Delta\theta_{e^+e^-}$ is given by

$$\Delta\theta_{e^+e^-} = \arccos\left(\frac{p_z^{e^+}}{|\mathbf{p}^{e^+}|}\right) - \arccos\left(\frac{p_z^{e^-}}{|\mathbf{p}^{e^-}|}\right),$$

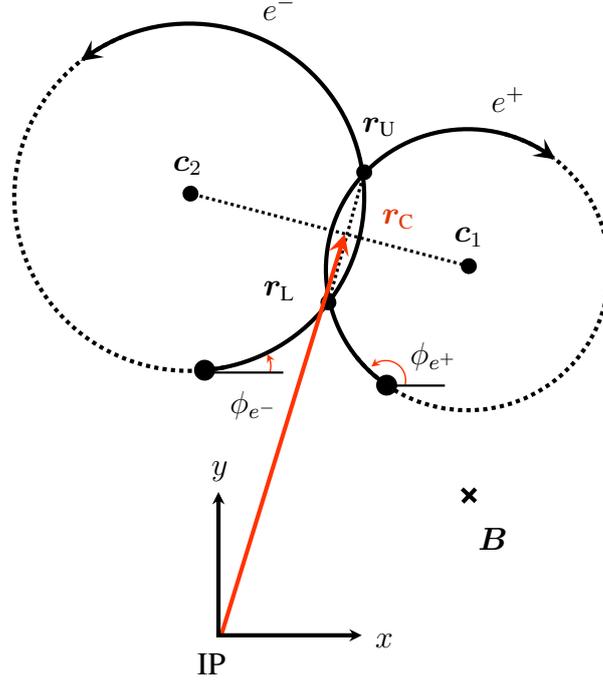


Figure 6.8: Illustration of c_1 , c_2 , r_C , r_L , r_U , ϕ_{e^+} and ϕ_{e^-} on the xy -plane ($r\phi$ -plane). The vector B corresponds to the magnetic field.

and the difference $\Delta\phi_{e^+e^-}$ by

$$\Delta\phi_{e^+e^-} = \text{atan2}(p_y^{e^+}, p_x^{e^+}) - \text{atan2}(p_y^{e^-}, p_x^{e^-}).$$

Because of the definition of atan2 , the values of $\Delta\phi_{e^+e^-}$ can peak around -2π , 0 , and 2π . The distribution has the same shape around each peak. To train a multivariate method, it is useful to have only one peak around 0 . This is achieved by mapping the $\Delta\phi_{e^+e^-}$ distribution in the following way

$$\Delta\phi_{e^+e^-}^{\text{mapped}} = \Delta\phi_{e^+e^-} - \text{truncate}\left(\frac{\Delta\phi_{e^+e^-} + 3\pi}{2\pi}\right) \cdot 2\pi + 2\pi,$$

where the function truncate is a rounding function towards zero². In the following, $\Delta\phi_{e^+e^-}^{\text{mapped}}$ will be referred to just as $\Delta\phi_{e^+e^-}$.

Figure 6.9 shows the distributions of the four input variables for B_{sig}^0 candidates with true Dalitz π_{Dal}^0 decays and for B_{sig}^0 candidates with true converted photons. In particular, the input variable $\Delta\phi_{e^+e^-}$ has a large separation power. For events with Dalitz π_{Dal}^0 decays, the $\Delta\phi_{e^+e^-}$ distribution is symmetric around 0° because the B^0 decay vertex is only a few microns away from the IP in the $r\phi$ -plane and because the π_{Dal}^0 decay follows immediately. Therefore, the angles at the point of closest approach correspond to a good approximation to the true opening

²When using C++ the mapping can be realized with the `fmod` function.

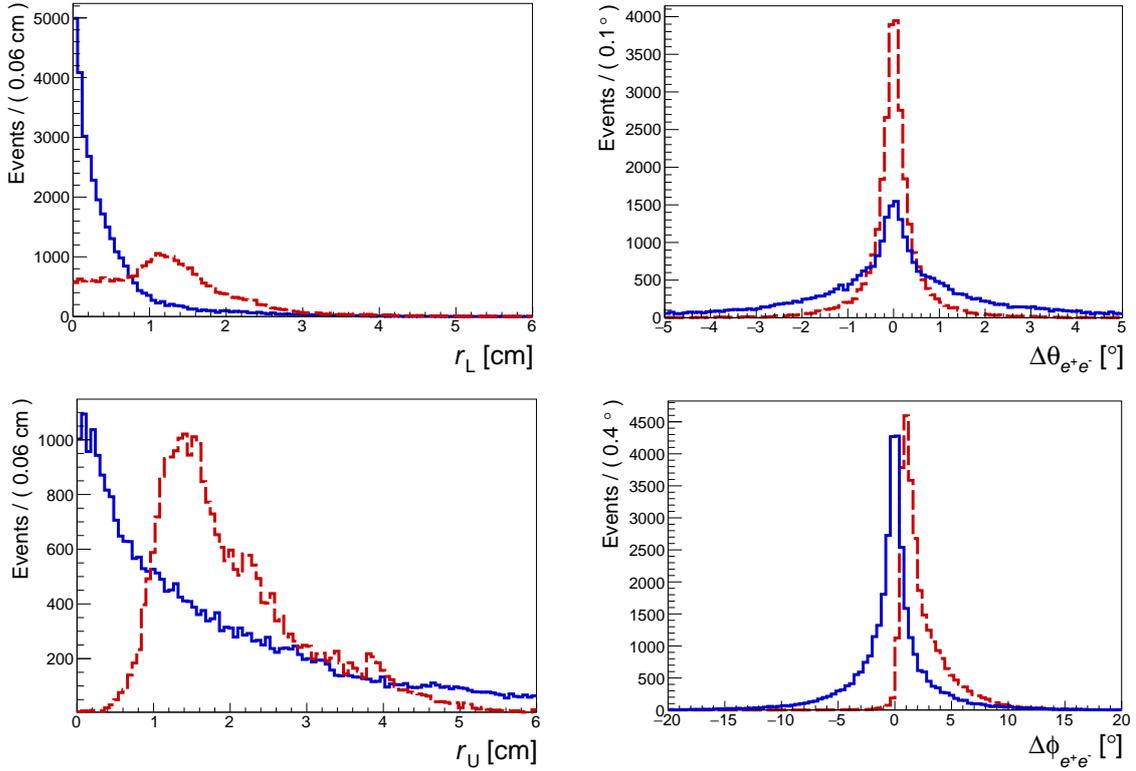


Figure 6.9: The four input variables, r_L (left top), r_U (left bottom), $\Delta\theta_{e^+e^-}$ (right top) and $\Delta\phi_{e^+e^-}$ (right bottom). B_{sig}^0 candidates with true Dalitz π_{Dal}^0 decays are shown by the solid blue curves, and with true converted photons by the dashed red curves.

angles at the decay vertex. In contrast, the distribution of $\Delta\phi_{e^+e^-}$ for converted photons is asymmetric. For these events, the vertex is far from the interaction point (at least 1 cm corresponding to the inner beam pipe radius). Thus, the ϕ angles at the point of closest approach to the IP are biased with respect to the true values at the conversion vertex. Because the curvature of reconstructed electrons and positrons have opposite signs, the angular difference $\Delta\phi_{e^+e^-}$ is always biased in the same direction.

The FBDT classifier was trained with a sample of about 200,000 events, with roughly equal fractions of Dalitz and conversion events. The test against over-fitting was performed in the same way as for the FBDT combiner of the flavor tagger (see Sect. 4.4). Figure 6.10 shows the output y_{DC} of the Dalitz-Conversion FBDT classifier. The excellent discrimination comes from the high power of the input variables.

6.6 Measurement of Δt

To measure the mixing-induced CP -violation parameter S_{CP} , a time-dependent analysis has to be performed. The analysis requires the reconstruction of the decay vertices of the signal-side and the tag-side mesons B_{sig}^0 and B_{tag}^0 . In the case of the decay $B_{\text{sig}}^0 \rightarrow \pi^0\pi^0$, only for B_{sig}^0 candidates with at least one converted photon or with one Dalitz π^0 a vertex reconstruction is

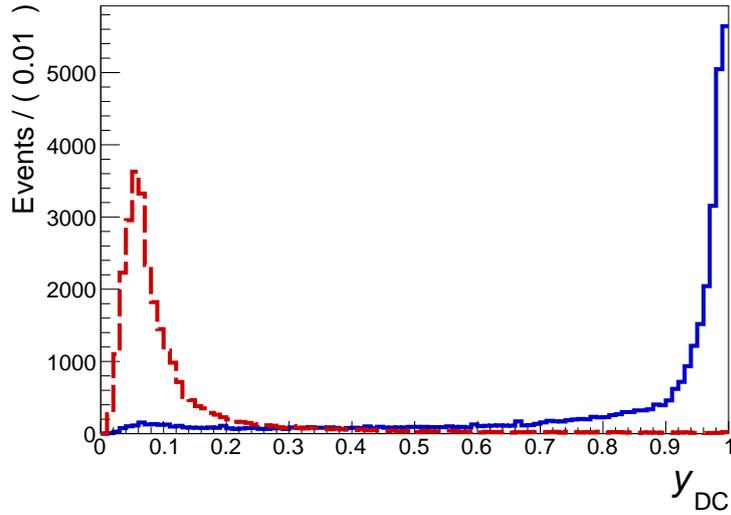


Figure 6.10: The output y_{DC} of the Dalitz-Conversion FBDT classifier. B_{sig}^0 candidates with true Dalitz π_{Dal}^0 decays are shown by the solid blue curve, and with true converted photons by the dashed red curve.

possible. In the following, the techniques employed for the vertex reconstruction of B_{sig}^0 and of B_{tag}^0 are explained. Next, the obtained Δt resolutions will be presented. The available vertex reconstruction algorithms at Belle II are referenced in Sect. 3.4. The results presented in this thesis were obtained using RAVE.

6.6.1 Signal-side vertex reconstruction

Since neutral pions decay promptly, the decay vertices of the neutral pions coincide with the vertex of B_{sig}^0 . A vertex reconstruction requires at least two tracks, or one track together with a decay-position constraint. For B_{sig}^0 candidates with a π_{Dal}^0 candidate, the e^+e^- pair is used to reconstruct B_{sig}^0 -decay vertex. To improve the vertex resolution, the e^+e^- -vertex fit is performed with a decay-position constraint. At Belle II, a commonly used decay-position constraint for B_{sig}^0 candidates is an ellipsoid oriented along the boost direction [1]. The constraint exploits the precise knowledge about the size and the position of the interaction region at Belle II, and the fact that B mesons are produced almost at rest in the $\Upsilon(4S)$ frame. The transverse size of the ellipsoid corresponds to the beam size at the IP. Its size in boost direction is 2 cm, about two orders of magnitude larger than the decay length of B_{sig}^0 .

Figure 6.11 shows the vertex distributions of $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0\pi_{\gamma\gamma}^0$ candidates. The π_{Dal}^0 candidate is required to be a true pion undergoing a Dalitz decay. Figure 6.11(a) shows the distribution of the generated MC decay vertex. Figure 6.11(b) shows the distribution of the B_{sig}^0 -decay vertex when it is reconstructed using the e^+e^- pair without decay-position constraint. Figure 6.11(c) shows the distribution when the B_{sig}^0 vertex is reconstructed by fitting the e^+e^- -vertex together with the ellipsoid constraint. The shape of the ellipsoid is visible from the distribution.

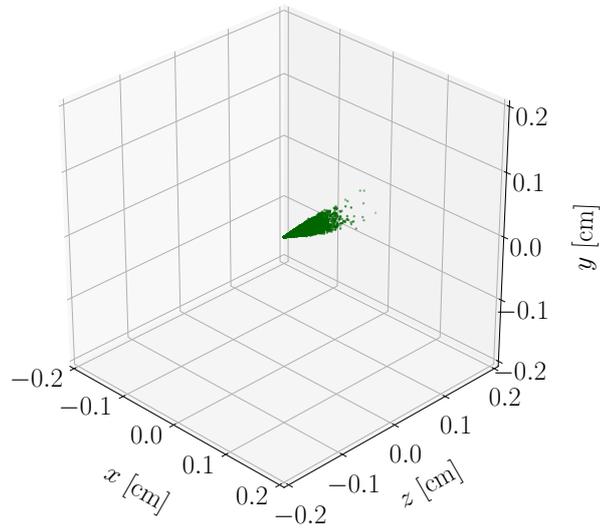
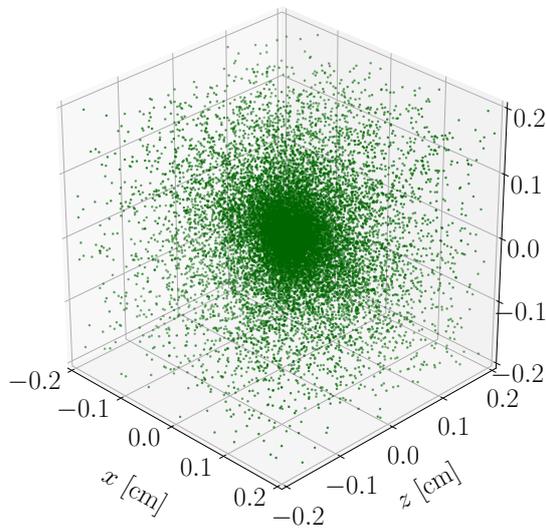
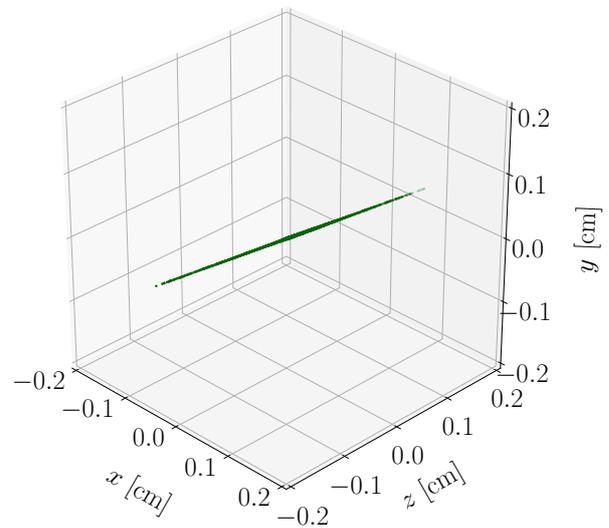
(a) Generated B_{sig}^0 -decay vertex.(b) Non-constrained e^+e^- -vertex fit.(c) e^+e^- -vertex fit with ellipsoid constraint.

Figure 6.11: B_{sig}^0 -vertex position for $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates. The π_{Dal}^0 candidate is required to be a true Dalitz π^0 . Top: MC decay vertex. Bottom: Reconstructed B_{sig}^0 vertex.

Figure 6.12 shows the corresponding residuals of the B_{sig}^0 vertex in boost direction. The residuals are defined as

$$l_{\text{sig}}^{\text{rec}} - l_{\text{sig}}^{\text{gen}} = (\mathbf{x}_{\text{sig}}^{\text{rec}} - \mathbf{x}_{\text{sig}}^{\text{gen}}) \cdot \hat{\beta},$$

where $\hat{\beta}$ is a unit vector pointing in boost direction, and $\mathbf{x}_{\text{sig}}^{\text{rec}}$ and $\mathbf{x}_{\text{sig}}^{\text{gen}}$ are the reconstructed and the generated 3D vertex of B_{sig}^0 . Figure 6.12(a) shows the residuals when the e^+e^- -vertex is reconstructed without decay-position constraint and taken as B_{sig}^0 vertex. And Fig. 6.12(b) shows the residuals when the B_{sig}^0 vertex is reconstructed by fitting the e^+e^- -vertex together with the ellipsoid constraint. With the ellipsoid constraint, the vertex resolution improves by about 85%.

For B_{sig}^0 candidates with one converted photon, the vertex of the e^+e^- pair is far from the B_{sig}^0 decay vertex. Thus, the e^+e^- tracks are used to reconstruct the converted photon as a pseudo track. The converted photon pseudo track is used together with the ellipsoid constraint to reconstruct the decay vertex of B_{sig}^0 .

As explained in the previous section, the B_{sig}^0 candidates with one reconstructed conversion and the B_{sig}^0 candidates with one reconstructed Dalitz π^0 have a certain overlap. Thus, four kinds of candidates with vertex can be distinguished. First, B_{sig}^0 candidates reconstructed with one true π_{Dal}^0 . Second, B_{sig}^0 candidates reconstructed with one π_{Dal}^0 that is in reality a π^0 with one converted photon. Third, B_{sig}^0 candidates reconstructed with one true γ_C . Fourth, B_{sig}^0 candidates reconstructed with one γ_C where the $\pi_{\gamma_C\gamma}$ candidate is in reality a Dalitz π^0 .

Figure 6.13 shows the residuals of the B_{sig}^0 vertex for the four kinds of candidates. Since Dalitz decays occur closer to the IP than conversions, the vertex resolution of B_{sig}^0 candidates with true Dalitz π^0 s is better than for candidates with one true conversion, even if the Dalitz π^0 s are wrongly reconstructed as candidates with one conversion (see Figs. 6.13(a) and 6.13(d)).

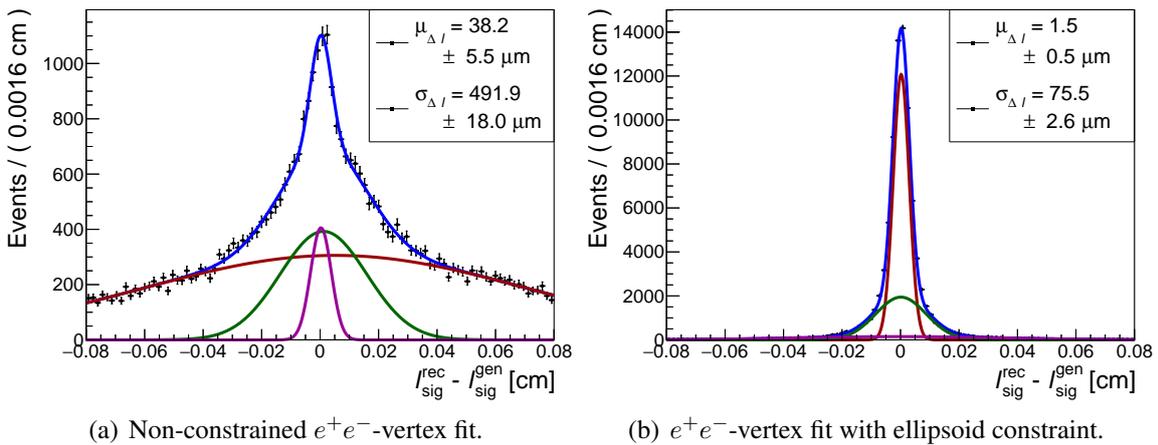


Figure 6.12: B_{sig}^0 -vertex residuals in boost direction for $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0\pi_{\gamma\gamma}^0$ candidates. The π_{Dal}^0 candidate is required to be a true Dalitz π^0 . The fits to the residuals are performed with three Gaussian functions. The shift $\mu_{\Delta l}$ and the resolution $\sigma_{\Delta l}$ are the weighted averages of the mean values and the standard deviations of the three Gaussian functions.

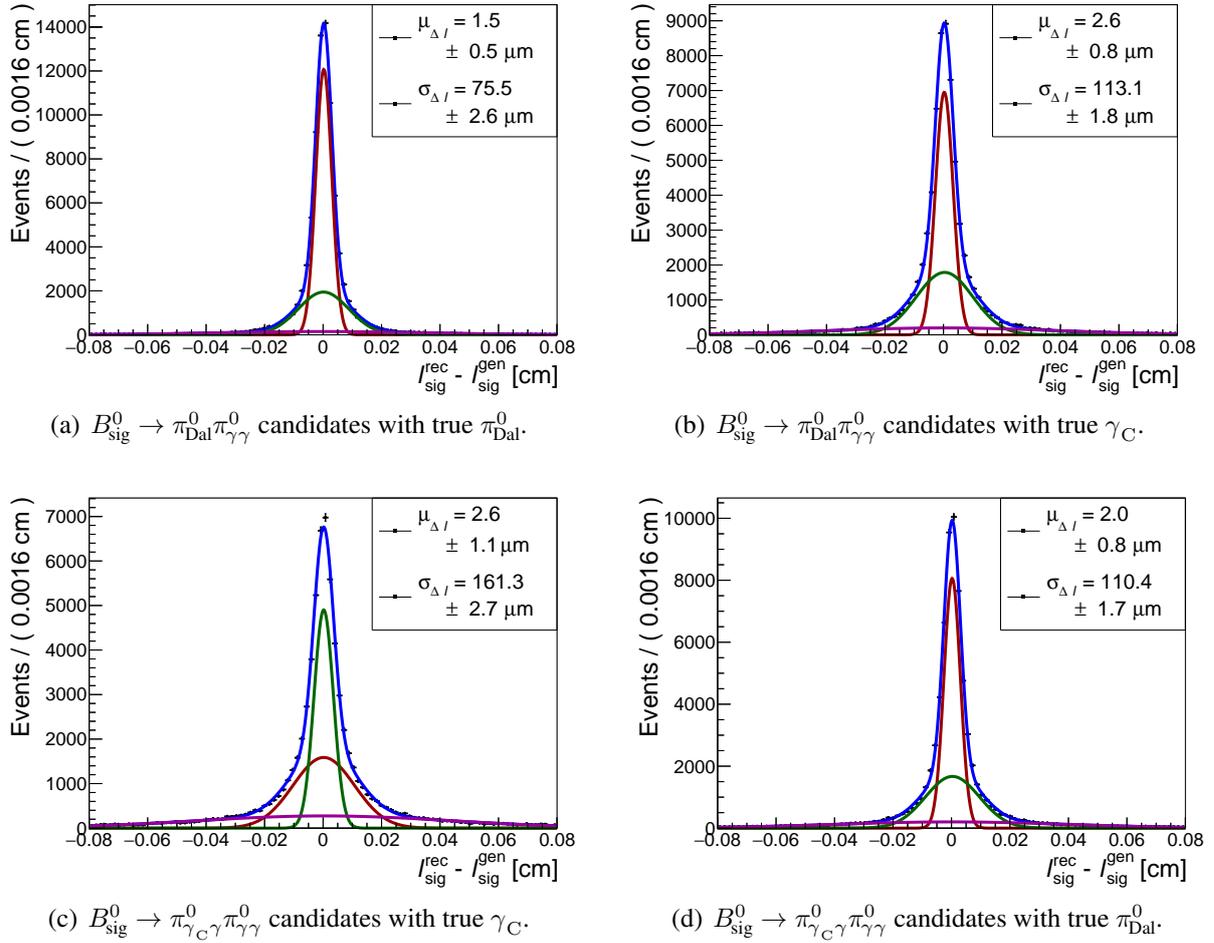


Figure 6.13: B_{sig}^0 -vertex residuals in boost direction for the four kinds of candidates: (a) B_{sig}^0 candidates reconstructed with one true π_{Dal}^0 ; (b) B_{sig}^0 candidates reconstructed with one π_{Dal}^0 that is in reality a π^0 with one converted photon; (c) B_{sig}^0 candidates reconstructed with one true γ_C ; and (d) B_{sig}^0 candidates reconstructed with one γ_C where the $\pi_{\gamma_C\gamma}^0$ candidate is in reality a Dalitz π^0 . The fits to the residuals are performed with three Gaussian functions. The shift $\mu_{\Delta l}$ and the resolution $\sigma_{\Delta l}$ are the weighted averages of the mean values and the standard deviations of the three Gaussian functions.

6.6.2 Tag-side vertex reconstruction

At B factories, the algorithms reconstructing the tag-side vertex profit from the isolated production of the $B\bar{B}$ pairs, in the same way as the flavor tagging algorithms. Since the $B\bar{B}$ pairs are produced at the $\Upsilon(4S)$, without other accompanying particles, the tracks on the tag-side must belong to the decay of B_{tag} and are therefore used for the vertex reconstruction.

The decay tracks stemming from a B^0 -meson decay can be classified into three groups: primary tracks from the B decay, including also daughters from intermediate hadronic resonances, e.g. muons from $B \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) X$ decays; tracks from D mesons, whose decay vertex is in average about $100 \mu\text{m}$ away from the B -decay vertex; and tracks from K_S^0 decays,

whose decay vertex is in average few centimeters away from the B -decay vertex. For the B_{tag}^0 -vertex fit, the tracks from K_S^0 are always excluded.

At Belle II, the RAVE Adaptive Vertex Fit (AVF) is the employed algorithm [144]. The AVF assigns outlying tracks a weight which depends on the compatibility of the track with the vertex; the weight can be therefore understood as a probability of a single track to belong to a given vertex. In order to reduce the weights of the tracks from D mesons, the fit is constrained in a first instance with an ellipsoid around the boost axis. This ellipsoid constraint covers a region where a B meson has a higher probability to decay in comparison with a possible D meson: its length in boost direction corresponds to 1.6 times the B^0 lifetime and its transverse size corresponds to the beam size at the IP. If the fit does not converge, the constraint is redefined enlarging its size in boost direction to about hundred times the decay length of a B^0 . Then, the fit is performed again. In first instance, the tracks with at least one PXD hit on the tag side are used as input. If the fit does not converge, the tracks with and without PXD hits are used as input in a second instance.

When the signal decay channel is fully reconstructed and correctly matched with the MC decay chain, the vertex of B_{tag}^0 does not depend on the reconstructed signal decay channel. To check if there are problems in the reconstruction and the MC association, the distribution of the residuals for each candidate kind was evaluated. Figure 6.14 shows the residuals of the tag-side vertex fit for the four kinds of B_{sig}^0 candidates with vertex. As Fig. 6.14 shows, the mean values and the standard deviations of the four distributions agree within the uncertainties.

6.6.3 Δt resolution and Δt selection

In time-dependent CP analyses, the CP -violation parameters are extracted from the Δt distribution, which can be calculated from the difference between the decay vertices of B_{sig}^0 and B_{tag}^0 (see Sect. 2.7.3). Since the B mesons are produced almost at rest in the $\Upsilon(4S)$, they travel to a good approximation only along the boost direction. Therefore, the time difference Δt can be measured as

$$\Delta t = \frac{(\mathbf{x}_{\text{sig}} - \mathbf{x}_{\text{tag}}) \cdot \hat{\boldsymbol{\beta}}}{\beta\gamma \cdot c},$$

where $\hat{\boldsymbol{\beta}}$ is a unit vector pointing in boost direction, \mathbf{x}_{sig} and \mathbf{x}_{tag} are the 3D vertices of B_{sig}^0 and B_{tag}^0 , and $\beta\gamma$ is the Lorentz boost (the average is $\langle\beta\gamma\rangle = 0.28$). The boost is determined from the beam parameters for each event. After the measurement of Δt , the B_{sig}^0 candidates with one π_{Dal}^0 or with one $\pi_{\gamma C\gamma}^0$ are required to fulfill $|\Delta t| \leq 25$ ps.

Figure 6.15 shows the Δt residuals for the four kinds of candidates. Note that B_{sig}^0 candidates with one true Dalitz π^0 have a better Δt resolution than those with one true conversion. For all kinds of candidates, the Δt resolution is limited by the signal-side vertex resolution. In contrast, for the benchmark decay channel $B_{\text{sig}}^0 \rightarrow J/\psi K_S^0$, the Δt resolution is limited by the tag-side vertex resolution. While for the decay channel $B_{\text{sig}}^0 \rightarrow \pi^0\pi^0$, the B_{sig}^0 vertex is reconstructed using the e^+e^- tracks from a Dalitz decay or from a converted photon (together

with a decay-position constraint), for $B_{\text{sig}}^0 \rightarrow J/\psi K_S^0$, the signal-side vertex is reconstructed with high precision using the muons from the J/ψ decay (see Fig. 3.9).

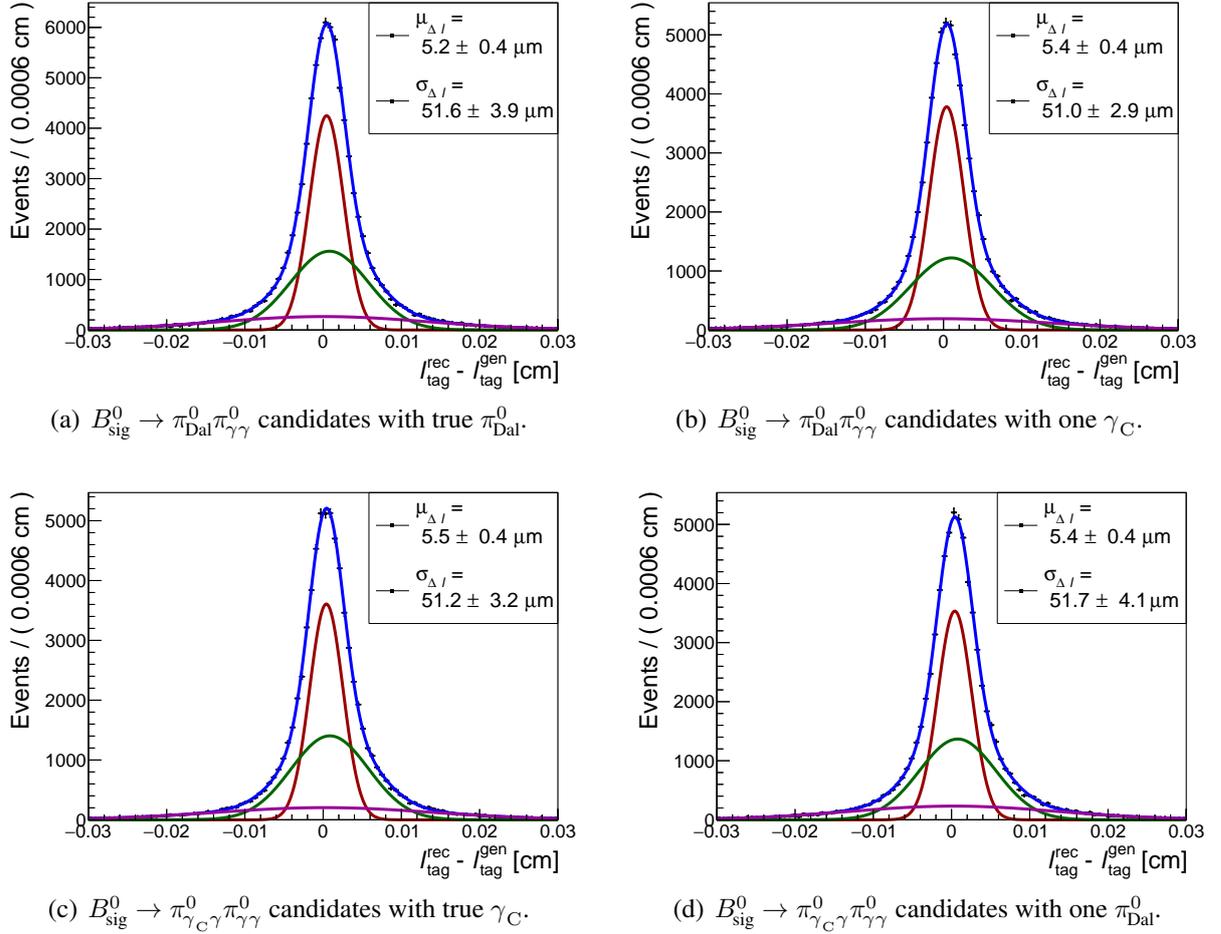


Figure 6.14: B_{tag}^0 -vertex residuals in boost direction for the four kinds of candidates. The fits to the residuals are performed with three Gaussian functions. The shift $\mu_{\Delta l}$ and the resolution $\sigma_{\Delta l}$ are the weighted averages of the mean values and the standard deviations of the three Gaussian functions.

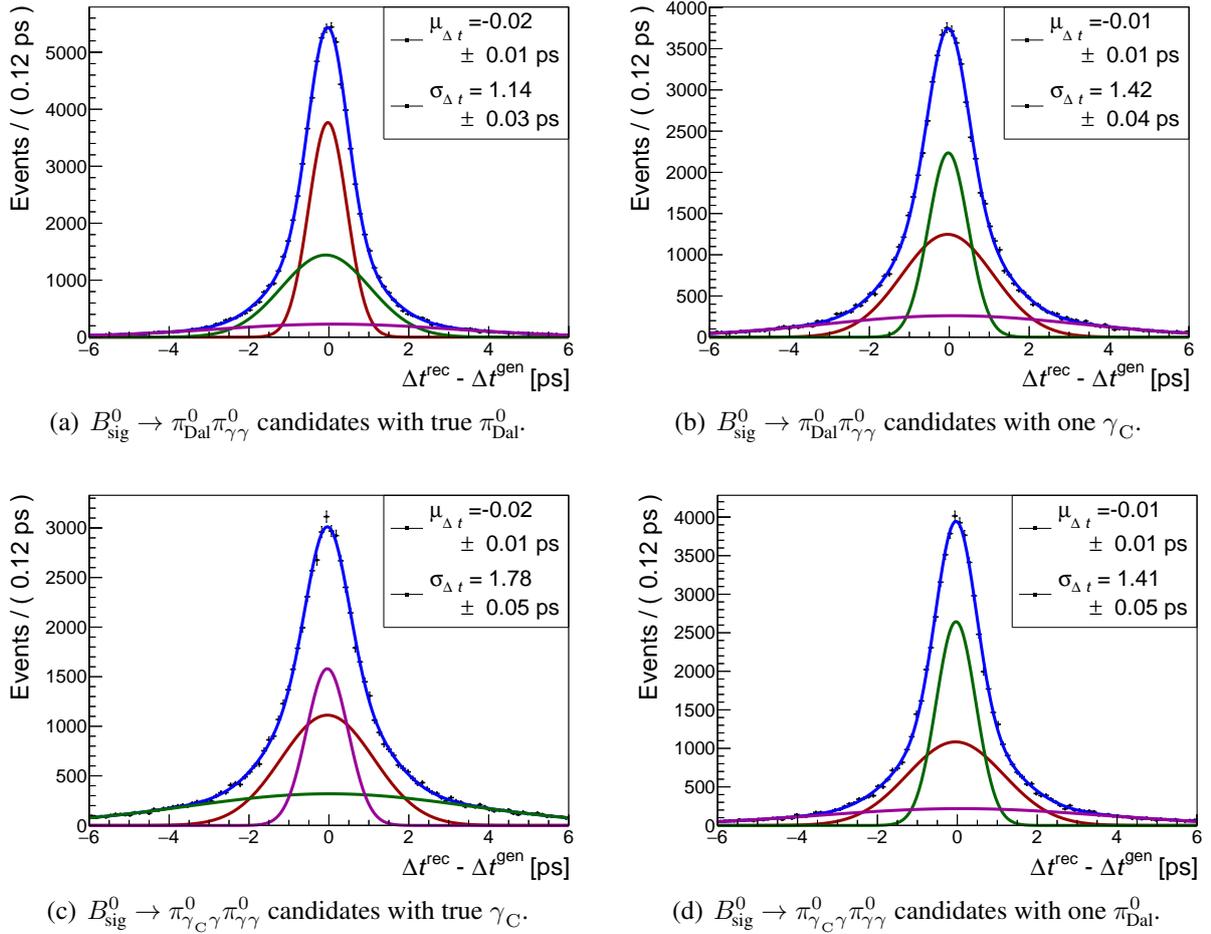


Figure 6.15: Δt residuals for the four kinds of candidates. The fits to the residuals are performed with three Gaussian functions. The shift $\mu_{\Delta t}$ and the resolution $\sigma_{\Delta t}$ are the weighted averages of the mean values and the standard deviations of the three Gaussian functions.

6.7 Continuum Suppression

At the $\Upsilon(4S)$, the fraction of non- $B\bar{B}$ events is larger than the fraction of events with $B\bar{B}$ pairs. Combinations of particles produced in non- $B\bar{B}$ events give rise to large amounts of background events usually outnumbering the signal events by orders of magnitude. For the signal decay channel $B^0 \rightarrow \pi^0\pi^0$, the relevant background from non- $B\bar{B}$ events corresponds to hadronic events (continuum background), where the beam electron and positron collide producing a $q\bar{q}$ pair where q is a quark lighter than a b quark. Such events occur about three times more often than $B\bar{B}$ events. Random combinations of particles from $q\bar{q}$ events can easily fulfill the selection criteria for B_{sig}^0 candidates. Thus, to perform an analysis of the signal decay channel, the continuum events have to be suppressed.

Belle and BaBar developed several techniques to suppress continuum events building on the experience of previous collaborations like CLEO and ARGUS. The techniques vary depending on how well they are suited for the analyzed signal channel. A review of the different techniques can be found in the Physics of the B factories book [29]. They have in common that they use variables exploiting the shape differences between $q\bar{q}$ and $B\bar{B}$ events.

Figure 6.16 shows an illustration of the event shapes for $q\bar{q}$ and for $B\bar{B}$ events. Pairs of light quarks are produced back-to-back in the center-of-mass frame. They have high momenta around the beam energy. When they fragment, they usually produce hadrons with low transverse momenta around $0.3 \text{ GeV}/c$. The produced hadrons form thus jet-like events. On the other hand, $B\bar{B}$ pairs are produced almost at rest in the center-of-mass frame. Because B mesons have 0 spin, their daughters spread isotropically in space forming spherically-shaped events.

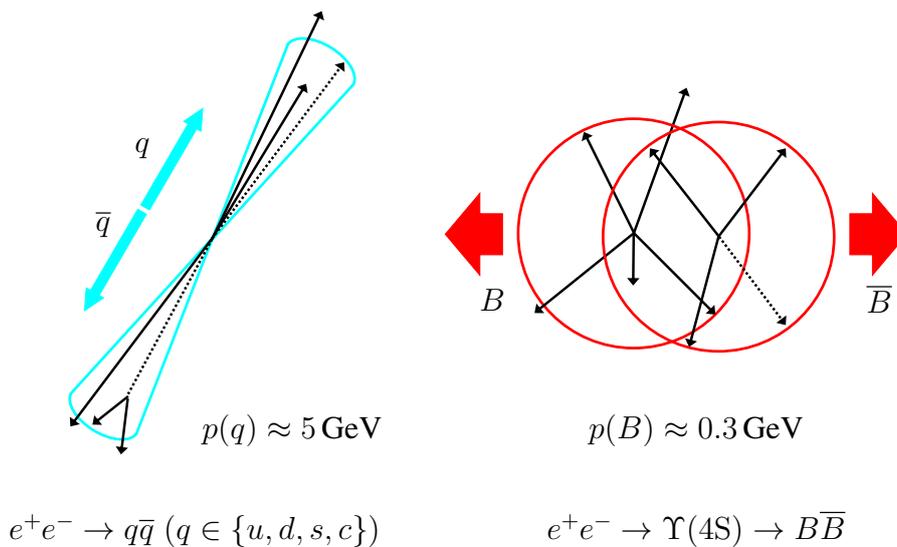


Figure 6.16: Illustration of the event shapes for jet-like continuum events (left) and for spherical $B\bar{B}$ events (right).

Event Type	Fraction [%]
$u\bar{u}$	50
$d\bar{d}$	13
$s\bar{s}$	5
$c\bar{c}$	15
$\tau^+\tau^-$	17

Table 6.1: Fractions of different non- $B\bar{B}$ events passing the reconstruction and the selection described in Sect. 6.4. The uncertainties are about $\pm 1\%$.

In the case of $B^0 \rightarrow \pi^0\pi^0$, a considerable fraction of $\tau^+\tau^-$ events pass the selection for B_{sig}^0 candidates, in addition to the light $q\bar{q}$ events. Events with $\tau^+\tau^-$ pairs occur almost as often as $B\bar{B}$ events. And the tauons can decay into final states with hadrons including one or several neutral pions. Random combinations of such final-state particles can thus easily fulfill the B_{sig}^0 -selection criteria. Table 6.1 presents the fractions of non- $B\bar{B}$ events passing the B_{sig}^0 reconstruction and selection described in Sect. 6.4.

Being still considerably lighter than B mesons, τ leptons are produced with relatively high momenta. Therefore, they also form jet-like events. Events with $\tau^+\tau^-$ pairs differ from light $q\bar{q}$ events, for example due to the presence of neutrinos. However, the continuum suppression variables have a very good discrimination power against them. In this thesis, the continuum suppression served therefore to suppress $\tau^+\tau^-$ events as well.

Belle II has adopted the variables used for continuum suppression at Belle. They serve as input for an FBDT method [115] that discriminates the signal channel from the continuum events. Depending on the analysis and on the signal channel, a subset of these variables can be selected.

This section introduces the variables used in the standard continuum suppression at Belle II following [1, 29]. In this thesis, the continuum suppression FBDT classifier was trained to reject continuum and $\tau^+\tau^-$ events. These two kinds of events will be referred to in the following as non- $B\bar{B}$ events when the distributions of the continuum suppression variables will be shown. The choice of the variables for the analysis of $B \rightarrow \pi^0\pi^0$ will be justified next. Afterward, the performance of the FBDT classifier and the selection criterion on its output will be explained.

6.7.1 Input Variables

- **Thrust:** For a collection of N momenta \mathbf{p}_i ($i = 1, \dots, N$), the thrust scalar T (or thrust) is defined as

$$T = \frac{\sum_{i=1}^N |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_{i=1}^N |\mathbf{p}_i|},$$

where the thrust axis \mathbf{T} is the unit vector that maximizes T . The direction of \mathbf{T} is such that the sum of the absolute values of the momenta, projected onto \mathbf{T} , is maximal. The

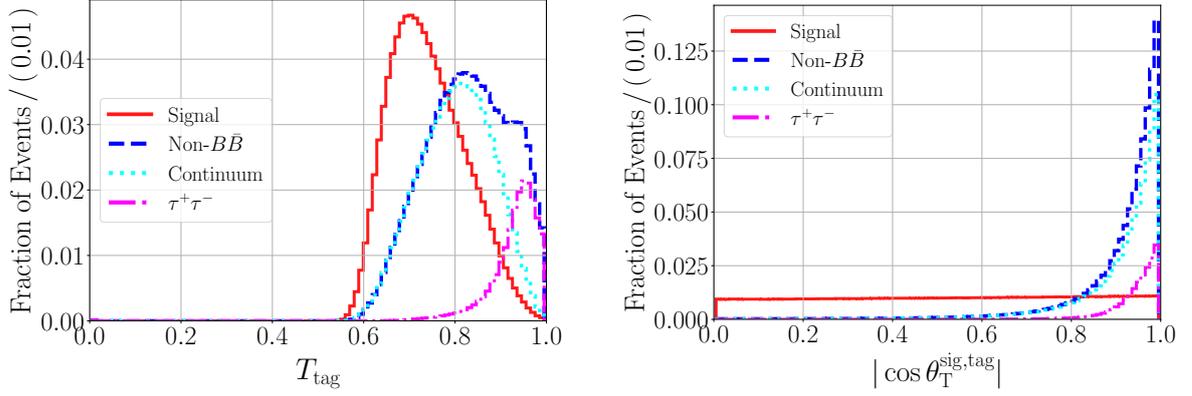


Figure 6.17: The distributions of the thrust T_{tag} (left) and $|\cos \theta_T^{\text{sig,tag}}|$ (right) for signal, for non- $B\bar{B}$, for continuum, and for $\tau^+\tau^-$ MC events. The non- $B\bar{B}$ events correspond to continuum and to $\tau^+\tau^-$ events together. The distributions for signal events and for non- $B\bar{B}$ MC events are normalized such that the area below each curve equals unity, leaving the proportions of continuum and $\tau^+\tau^-$ events as given in Tab. 6.1.

thrust T and the thrust axis \mathbf{T} are calculated in the $\Upsilon(4S)$ frame for the momenta of the B_{sig}^0 candidate decay particles (T_{sig} and \mathbf{T}_{sig}), and for the momenta of all the particles in the tag side (T_{tag} and \mathbf{T}_{tag})³. The values of T_{sig} and T_{tag} have a discrimination power between signal and non- $B\bar{B}$ events. Figure 6.17 (left) shows the distribution of T_{tag} for signal and for non- $B\bar{B}$ MC events.

- **Thrust angles:** The cosines of two thrust angles are used as input variables. The first one is $|\cos \theta_T^{\text{sig,tag}}|$, where $\theta_T^{\text{sig,tag}}$ is the angle between \mathbf{T}_{sig} and \mathbf{T}_{tag} . Among all continuum suppression variables, this variable is the most powerful one. Figure 6.17 (right) shows its distribution for signal and for non- $B\bar{B}$ MC events. Since the momenta of the decay particles of B_{sig}^0 and of B_{tag}^0 are isotropically distributed, \mathbf{T}_{sig} and \mathbf{T}_{tag} are randomly distributed, and thus $|\cos \theta_T^{\text{sig,tag}}|$ follows a uniform distribution. For light $q\bar{q}$ and for $\tau^+\tau^-$ events, the momenta of the particles follow the direction of the jets in the event, and as a consequence \mathbf{T}_{sig} and \mathbf{T}_{tag} are strongly directional and collimated, yielding a $|\cos \theta_T^{\text{sig,tag}}|$ distribution strongly peaked at 1.

The second variable is $|\cos \theta_T^{\text{sig,beam}}|$, where $\theta_T^{\text{sig,beam}}$ is the angle between \mathbf{T}_{sig} and the beam axis (in z -direction).

- **CLEO Cones:** The CLEO collaboration introduced a set of 9 variables for measurements of charmless B_{sig}^0 decays which typically suffer from large continuum backgrounds as $B^0 \rightarrow \pi^0\pi^0$ [145]. They are in principle a refinement of the concept of thrust: they are based on the sum of the absolute values of the momenta of all par-

³The thrust axis of the tag side, T_{tag} , is required also for the calculation of $|\cos \theta_T^*|$ within the flavor tagger (see Sect. 4.3).

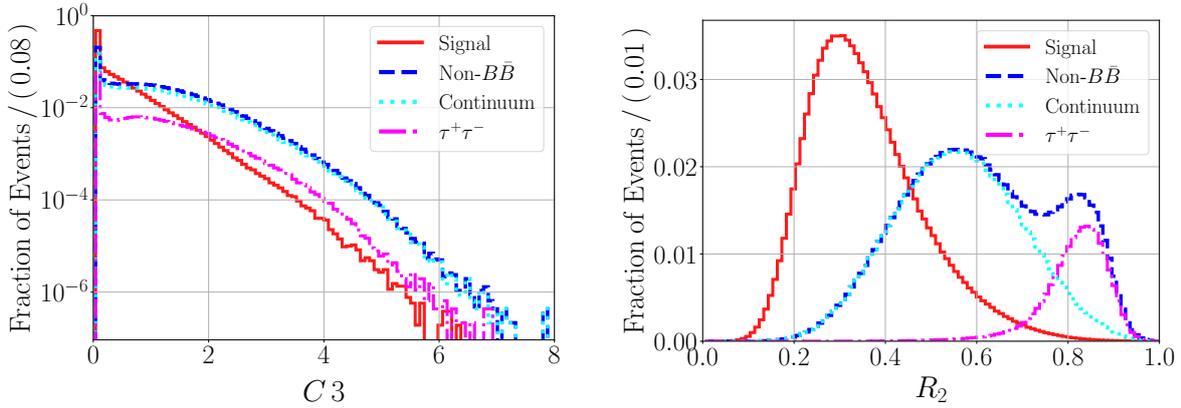


Figure 6.18: The distributions of (left) the third CLEO cone and (right) the ratio R_2 of second to zeroth Fox-Wolfram moment.

ticles within angular sectors around the thrust axis, in intervals of 10° , resulting in 9 concentric cones. For this thesis, the CLEO cones are calculated from all final state particles in the event. The CLEO cone variables have slightly similar distributions. As an example, Fig. 6.18 (left) shows the distribution of the third CLEO cone C_3 , the one with the highest discrimination power.

- **Fox-Wolfram moments:** Given a total number of N particles in an event, with momenta \mathbf{p}_i^* in the $\Upsilon(4S)$ frame, the l -th order Fox-Wolfram moment H_l [146] is defined as

$$H_l = \sum_{i,j}^N \frac{|\mathbf{p}_i^*| \cdot |\mathbf{p}_j^*|}{s} \cdot P_l(\cos \theta_{i,j}^*),$$

where $\theta_{i,j}^*$ is the angle between \mathbf{p}_i^* and \mathbf{p}_j^* , \sqrt{s} is the total energy in the $\Upsilon(4S)$ frame, and P_l is the l -th order Legendre polynomial. In the limit of negligible invariant masses of the particles, the conservation of energy and momentum requires that H_0 is equal to one, which is why the normalized ratios $R_l = H_l/H_0$ are often used. At Belle II, only the normalized ratio R_2 is used. It is a strongly discriminating variable. Figure 6.18 (right) shows its distribution for signal and for non- $B\bar{B}$ MC events.

- **Kakuno-Super-Fox-Wolfram moments:** The discrimination power provided by the Fox-Wolfram moments deteriorates when particles are missing. To account for missing particles, the Belle collaboration developed the Kakuno-Super-Fox-Wolfram moments (KSFW) H_{xl}^{so} and H_l^{oo} ($l \in [0, 4]$) [147]. All reconstructed particles in an event are divided into two types: B_{sig}^0 candidate daughters (denoted as s), and tag-side particles (denoted as o). The H_{xl}^{so} moments are sorted into three categories (denoted as x) depending on whether the particle is charged ($x = c$), neutral ($x = n$), or missing ($x = m$). For the H_{xl}^{so} moments, the missing momentum of an event is treated as an additional

particle. For even l ,

$$H_{xl}^{\text{so}} = \sum_i \sum_{j_x} |\mathbf{p}_{j_x}^*| \cdot P_l(\cos \theta_{i,j_x}^*),$$

where i extends over the primary B_{sig}^0 daughters, and j_x over all the tag-side particles belonging to category x ; $\mathbf{p}_{j_x}^*$ is the momentum of the particle j_x in the $\Upsilon(4S)$ frame; and $P_l(\cos \theta_{i,j_x}^*)$ is the l -th order Legendre polynomial of the cosine of the angle between the particles i and j_x in the $\Upsilon(4S)$ frame. For odd l , one obtains $H_{nl}^{\text{so}} = H_{ml}^{\text{so}} = 0$ and

$$H_{cl}^{\text{so}} = \sum_i \sum_{j_x} q_i \cdot q_{j_x} \cdot |\mathbf{p}_{j_x}^*| \cdot P_l(\cos \theta_{i,j_x}^*), \quad (6.3)$$

where q_i and q_{j_x} are the charges of the particles i and j_x . There are a total of eleven H_{xl}^{so} moments: two for $l = 1, 3$; and nine (3×3) for $l = 0, 2, 4$. The moments for $l = 1, 3$ are known to be correlated with ΔE and m_{bc} . They are therefore not used by default at Belle II.

The definition of the five H_l^{oo} moments is

$$H_l^{\text{oo}} = \begin{cases} \sum_j \sum_k |\mathbf{p}_j^*| \cdot |\mathbf{p}_k^*| \cdot P_l(\cos \theta_{j,k}^*) & (l = \text{even}) \\ \sum_j \sum_k q_j \cdot q_k \cdot |\mathbf{p}_j^*| \cdot |\mathbf{p}_k^*| \cdot P_l(\cos \theta_{j,k}^*) & (l = \text{odd}), \end{cases}$$

where j and k extend over all the tag side particles. The H_{xl}^{so} moments are normalized to H_0^{max} , and the H_l^{oo} moments to $(H_0^{\text{max}})^2$, where $H_0^{\text{max}} = 2(\sqrt{s} - \Delta E)$. The normalization is performed to avoid dependencies on ΔE (eq. (6.2)).

Figure 6.19 shows the distributions of the KSFW moments for $l = 2$, the ones with the highest discrimination power.

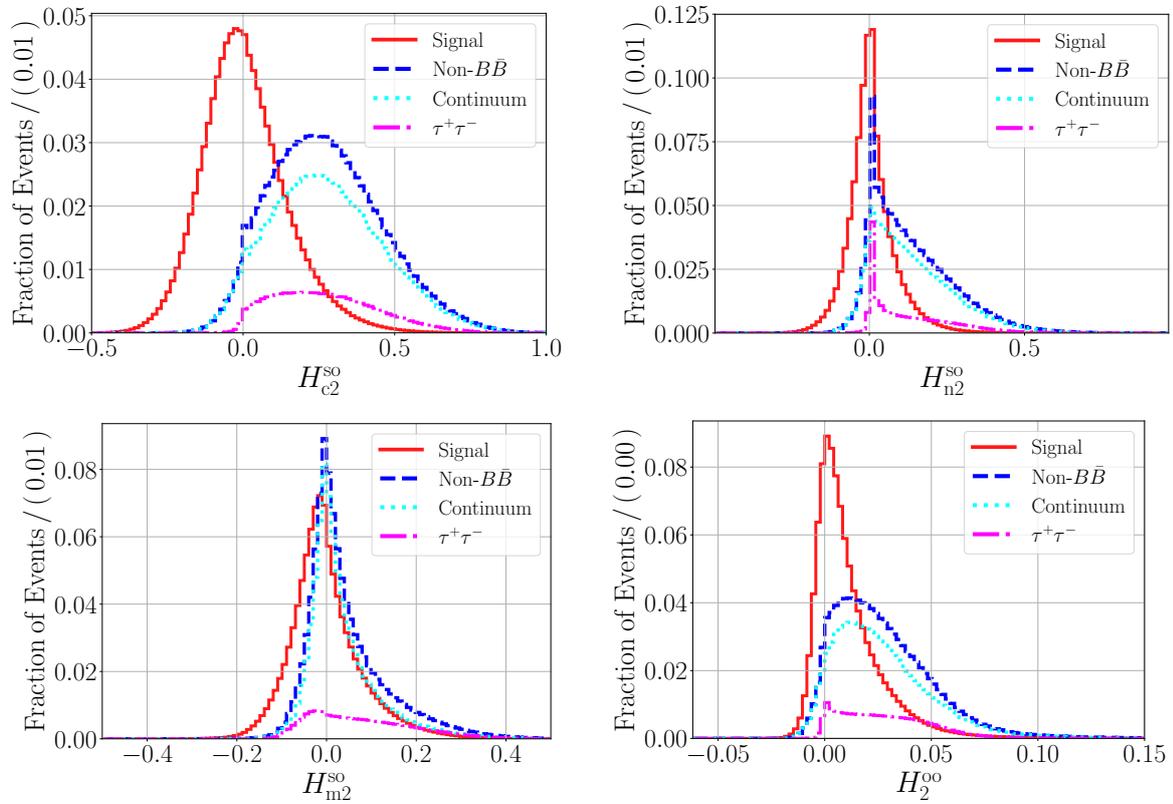
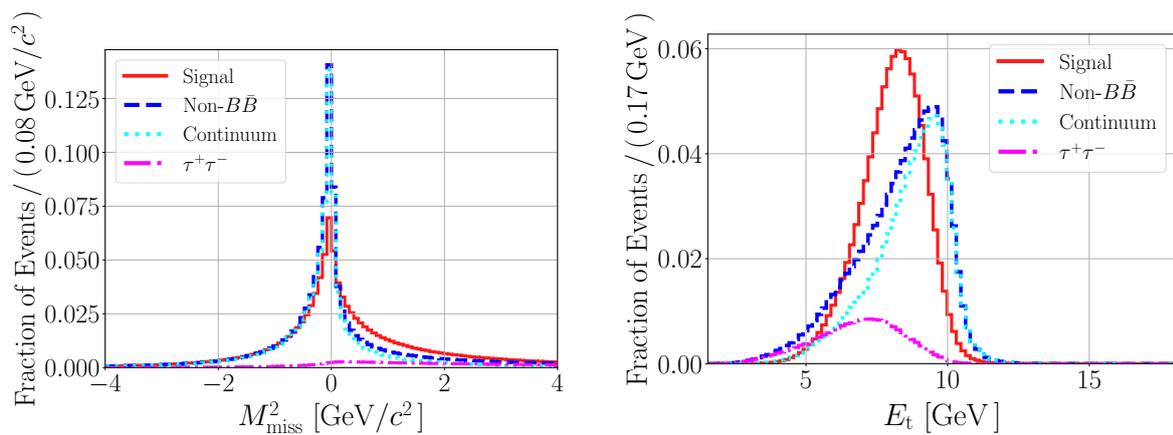
- **Missing mass and transverse energy:** The missing mass squared is defined as

$$M_{\text{miss}}^2 = \left(\sqrt{s} - \sum_{i=1}^N E_i^* \right)^2 - \sum_{i=1}^N |\mathbf{p}_i^*|^2,$$

where E_i^* and \mathbf{p}_i^* are the energy and the momentum of the i -th particle in the $\Upsilon(4S)$ frame, and N is the total number of particles in the event. The transverse energy

$$E_t = \sum_{i=1}^N |(\mathbf{p}_i^*)_t|, \quad (6.4)$$

is the scalar sum of the transverse momenta of all particles in the event. The distributions of M_{miss}^2 and E_t are shown in Fig. 6.20.

Figure 6.19: The distributions of the KSWF moments for $l = 2$.Figure 6.20: The distributions of (left) the missing mass squared M_{miss}^2 and (right) the transverse energy E_t .

Selection of variables

From the 30 variables used in the standard continuum suppression, two variables, T_{sig} and C_1 , were discarded because they are correlated with the fit observables m_{bc} and ΔE . By using them as input, the output of the FBDT classifier becomes correlated with the fit observables, a situation that has to be avoided because otherwise requirements on the FBDT output modify the shape of the continuum background such that it mimics the signal.

Figure 6.21 presents the correlation coefficients between m_{bc} , ΔE and all the variables used in the standard continuum suppression. The thrust value T_{sig} is strongly correlated with m_{bc} , especially for continuum events. And the first CLEO cone is considerably correlated with ΔE . Fig. 6.22 shows the distributions of T_{sig} and m_{bc} , and of C_3 and ΔE .

	m_{bc}	ΔE	T_{sig}	T_{tag}	$c_{s,t}$	$c_{s,b}$	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	R_2	$H_{\text{so } c0}$	$H_{\text{so } c2}$	$H_{\text{so } c4}$	$H_{\text{so } n0}$	$H_{\text{so } n2}$	$H_{\text{so } n4}$	$H_{\text{so } m0}$	$H_{\text{so } m2}$	$H_{\text{so } m4}$	$H_{\text{oo } 0}$	$H_{\text{oo } 1}$	$H_{\text{oo } 2}$	$H_{\text{oo } 3}$	$H_{\text{oo } 4}$	M_{miss}^2	E_t	
m_{bc}	100	10	38	1	-1	-2	3	-2	0	0	0	0	0	0	0	0	1	-1	0	-1	-2	-2	0	-2	0	0	0	1	0	0	0	0	1
ΔE	10	100	2	0	0	-11	33	0	1	1	1	0	-1	-1	-1	4	5	2	0	2	-1	-3	1	-2	0	6	0	2	0	1	-6	14	

Signal

	m_{bc}	ΔE	T_{sig}	T_{tag}	$c_{s,t}$	$c_{s,b}$	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	R_2	$H_{\text{so } c0}$	$H_{\text{so } c2}$	$H_{\text{so } c4}$	$H_{\text{so } n0}$	$H_{\text{so } n2}$	$H_{\text{so } n4}$	$H_{\text{so } m0}$	$H_{\text{so } m2}$	$H_{\text{so } m4}$	$H_{\text{oo } 0}$	$H_{\text{oo } 1}$	$H_{\text{oo } 2}$	$H_{\text{oo } 3}$	$H_{\text{oo } 4}$	M_{miss}^2	E_t
m_{bc}	100	0	63	0	0	0	5	-3	1	0	0	0	0	0	0	1	0	1	1	1	1	2	-2	0	1	0	0	1	0	2	0	0
ΔE	0	100	12	0	-2	0	17	-5	-3	-3	-2	-2	-2	-1	-1	5	0	-1	-1	-2	-2	-2	1	0	-1	-1	1	-2	1	-1	0	0

Non- $B\bar{B}$

Figure 6.21: Correlation coefficients between m_{bc} , ΔE and all the variables used in the standard continuum suppression. These variables are the thrust values T_{sig} and T_{tag} , the absolute values of the cosines of the thrust angles $\theta_{\text{T}}^{\text{sig,tag}}$ and $\theta_{\text{T}}^{\text{sig,beam}}$ (denoted as $c_{\text{T } s, t}$ and $c_{\text{T } s, b}$), the nine cleo cones $C_1 - 9$, the normalized Fox-Wolfram moment R_2 , the nine even KSFW moments H_{xl}^{so} and H_l^{oo} , the missing mass squared M_{miss}^2 , and the transverse energy E_t . The correlation coefficients are calculated as in Sect. 4.3 and are given in percent.

6.7.2 Performance and selection criterion

The FBDT classifier for the optimized continuum suppression was trained using B_{sig}^0 candidates reconstructed with four ECL photons in the final state. The size of the training sample was about one million events. The events were sampled after the reconstruction and the selection described in Sect. 6.4. The fractions of signal and of non- $B\bar{B}$ events were roughly the same. The non- $B\bar{B}$ sample, corresponding to about 0.4 ab^{-1} , contained events with light $q\bar{q}$ and $\tau^+\tau^-$ pairs in the proportions presented in Tab. 6.1. Including $\tau^+\tau^-$ events in the training sample had no impact on the discrimination power of the FBDT classifier on continuum

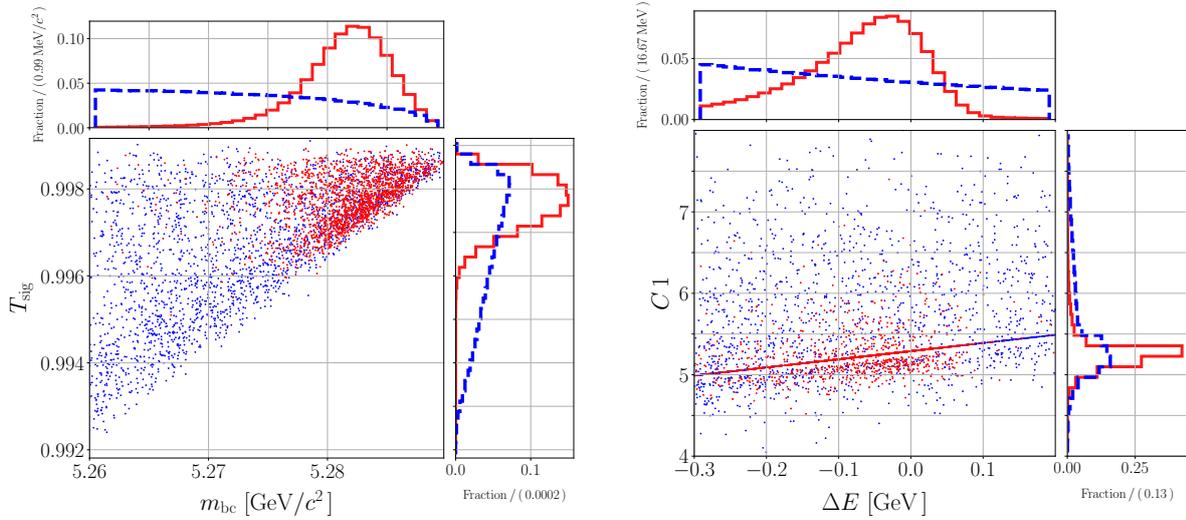


Figure 6.22: The 2D distributions of (left) the thrust value T_{sig} against m_{bc} , and (right) the first CLEO cone $C1$ against ΔE for signal (solid red) and for non- $B\bar{B}$ (dashed blue) MC events.

events. After the training procedure, the test against over-fitting was performed in the same way as for the FBDT combiner of the flavor tagger (see Sect. 4.4).

The performance of the final continuum-suppression FBDT classifier was similar for B_{sig}^0 candidates reconstructed with four ECL photons, and for B_{sig}^0 candidates reconstructed either with one converted photon or with one Dalitz π^0 . This is expected because the difference in shapes between continuum and signal events does not depend on the reconstructed B_{sig}^0 decay mode.

To enhance the fraction of signal events, a selection criterion was applied on the output of the FBDT for all B_{sig}^0 candidates. This selection criterion was chosen such that the Figure of Merit (FoM)

$$\text{FoM} = \frac{N_{\text{sig}}}{\sqrt{N_{\text{sig}} + N_{\text{cont}} + N_{\tau^+\tau^-}}},$$

reaches its maximum value. Here, N_{sig} , N_{cont} , and $N_{\tau^+\tau^-}$ are the numbers of signal, of continuum, and of $\tau^+\tau^-$ events passing the selection. The FoM reaches its maximum value at the same place if the number of $\tau^+\tau^-$ events $N_{\tau^+\tau^-}$ is removed from the denominator. Figure 6.23 (left) presents the distributions of the FBDT output y_{CS} for signal, for non- $B\bar{B}$, for continuum, and for $\tau^+\tau^-$ MC events. The distributions for signal and for non- $B\bar{B}$ MC events are normalized such that the area below each curve equals unity, leaving the proportions of continuum and $\tau^+\tau^-$ events as given in Tab. 6.1. Figure 6.23 (right) shows the values of the FoM depending on the requirement on y_{CS} . The maximum is reached at 0.976 and its position is denoted by a vertical black line. Thus, all B_{sig}^0 candidates are required to have a value greater than or equal to it.

Since some of the employed continuum-suppression variables are slightly correlated with m_{bc} and ΔE (see Fig. 6.21), the output of the final continuum-suppression FBDT classifier

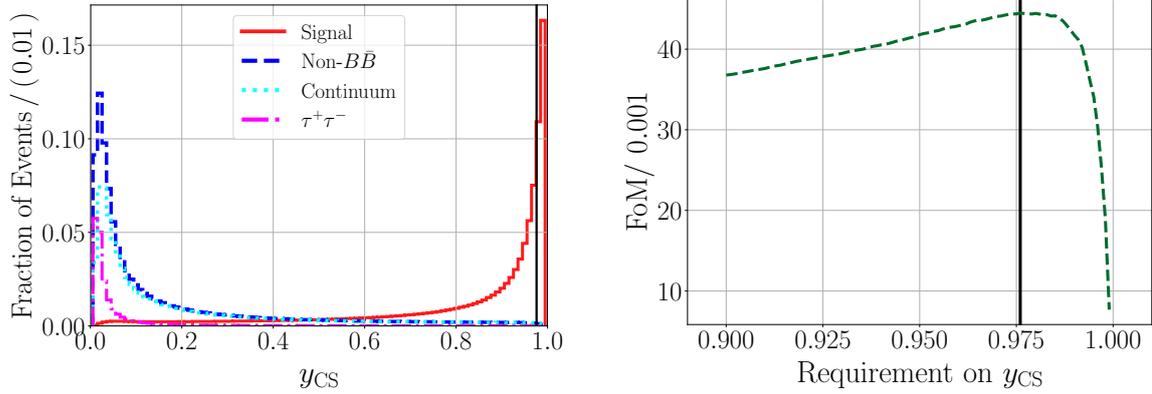


Figure 6.23: Left: the output of the FBDT method used for continuum suppression y_{CS} . The distribution of y_{CS} is shown for signal and for non- $B\bar{B}$ MC events. Right: value of the FoM depending on the requirement on y_{CS} . The maximum is reached at $y_{CS} = 0.976$, corresponding to the vertical black lines.

is also slightly correlated with m_{bc} and ΔE . The correlation coefficient between y_{CS} and m_{bc} is about 2% for signal events and less than 1% for non- $B\bar{B}$ events. And the correlation coefficient between y_{CS} and ΔE is less than 1% for signal events and about 3% for non- $B\bar{B}$ events. Thus, after the requirement on y_{CS} the shapes of the distributions and the correlation coefficients between m_{bc} and ΔE change slightly. Figure 6.24 shows the 2D distributions for m_{bc} and ΔE before and after the requirement on y_{CS} . The shapes of the distributions change slightly. However, the discrimination power between signal and non- $B\bar{B}$ events is not considerably affected.

6.8 Flavor tagging

When the studies presented in this chapter were carried out, the flavor tagger was not fully optimized. The total effective efficiency of the FBDT combiner, calculated using the r binning of Belle, was $\varepsilon_{\text{eff}} = 33.42 \pm 0.10$. To tag the flavor of B_{tag}^0 , the output of the FBDT combiner was used; and the performance of this combiner was taken into account to model the signal Δt distribution, and the Δt distribution of wrongly reconstructed signal events.

For the set of pseudo-experiments which will be presented in Sect. 6.10, the continuum and the $B\bar{B}$ background probability density functions were modeled using the official MC samples corresponding to 1.8 and to 4.0 ab^{-1} . The sizes of these samples are too small to model the background PDFs for each of the r bins of the Belle binning. Therefore, the pseudo-experiments were performed in a single r bin corresponding to $[0.1, 1.0]$. The range $[0, 0.1]$ was discarded because the continuum background peaks at $r = 0$, and because the signal events in this range provide no flavor information. The $q \cdot r$ distribution of the continuum background is shown for example in Sect. 4.8.

The Δt probability function $\mathcal{P}(\Delta t, q)$, and the time-integrated probability $\mathcal{P}(q)$ for the pseudo-experiments correspond to eqs. (4.7) and (4.8) in Sect. 4.2. These functions

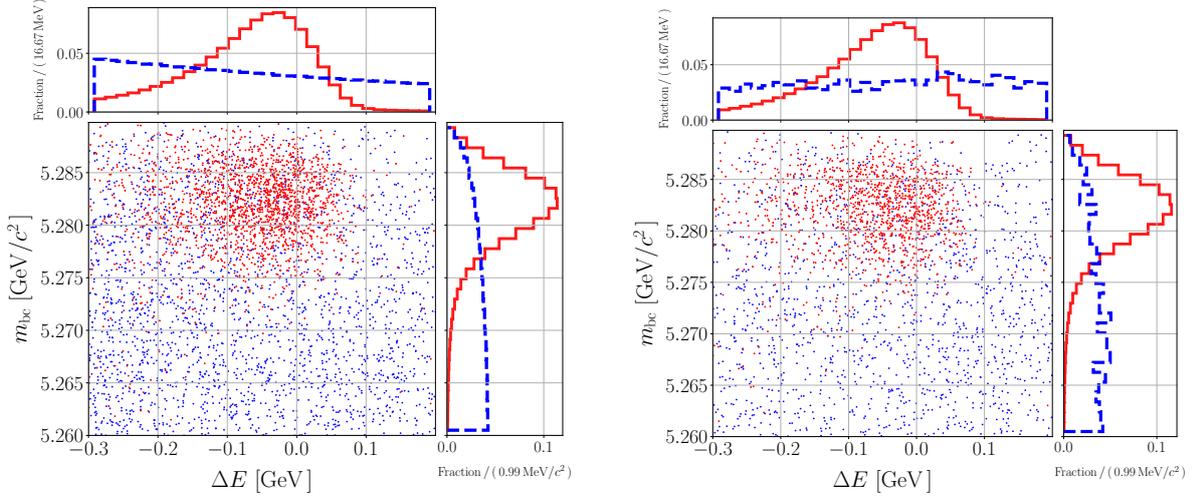


Figure 6.24: The 2D distributions of m_{bc} against ΔE before (left) and after (right) the suppression of non- $B\bar{B}$ events by requiring $y_{CS} \geq 0.976$. The correlation coefficient between m_{bc} and ΔE before the non- $B\bar{B}$ suppression was about 10% for the signal component and 0% for the non- $B\bar{B}$ component; After the non- $B\bar{B}$ suppression, it was about 9% for the signal component and 1% for the non- $B\bar{B}$ component.

include the value of the wrong-tag fraction w , the difference Δw , and the parameter $\mu = \varepsilon_{B^0} - \varepsilon_{\bar{B}^0} / (\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r bin. These values were determined using the official Belle II MC sample used to test the flavor tagger. Table 6.2 presents the numbers characterizing the performance of the FBDT combiner in the intervals $r \in [0.0, 0.1]$ and $r \in [0.1, 1.0]$. In the latter interval, the value of μ_i equals $(-0.06 \pm 0.05) \cdot 10^{-2}$. This value is very small and can be neglected to a good approximation. For the pseudo-experiments, μ was set to zero. The values in Tab. 6.2 correspond to the performance of the flavor tagger used for the study.

Table 6.2: Performance of the FBDT combiner on Belle II MC using the binning employed in the performance study of $B^0 \rightarrow \pi^0\pi^0$. All values are given in percent.

FBDT Combiner on Belle II MC							
r -Interval	ε_i	$\Delta\varepsilon_i$	$w_i \pm \delta w_i$	$\Delta w_i \pm \delta\Delta w_i$	$\varepsilon_{\text{eff},i} \pm \delta\varepsilon_{\text{eff},i}$	$\Delta\varepsilon_{\text{eff},i} \pm \delta\Delta\varepsilon_{\text{eff},i}$	
0.000 – 0.100	12.5	0.10	47.68 ± 0.17	1.12 ± 0.34	0.027 ± 0.004	-0.026 ± 0.008	
0.100 – 1.000	87.4	-0.10	22.10 ± 0.05	0.67 ± 0.11	27.213 ± 0.104	-1.339 ± 0.207	

6.9 Final selection and efficiency

After the reconstruction and the application of the requirements on ΔE , m_{bc} , Δt , continuum-suppression FBDT output, and flavor dilution factor, an event can in principle contain several B_{sig}^0 candidates. For the three reconstructed B_{sig}^0 -decay modes, Fig. 6.25 shows the number of events containing one or several B_{sig}^0 candidates. The average of such a distribution, the

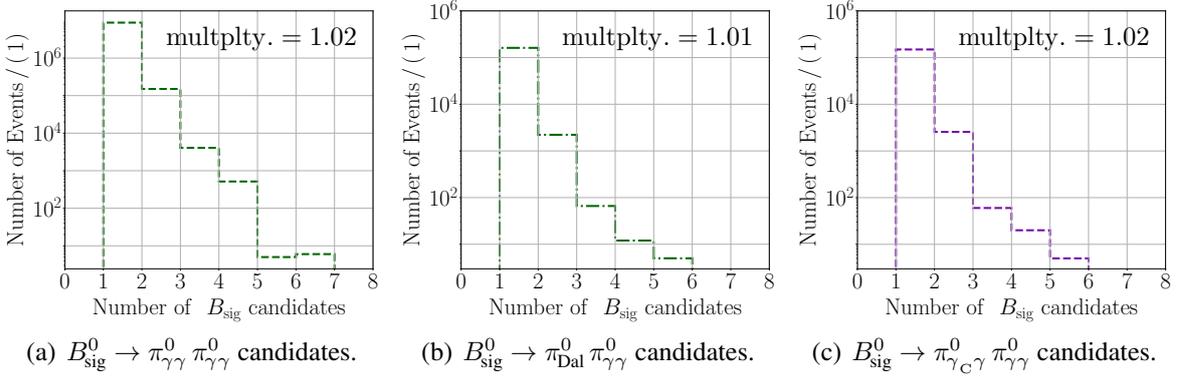


Figure 6.25: Number of reconstructed B_{sig}^0 candidates per event and average multiplicities.

so-called event multiplicity, is 1.02 for the decay modes with four ECL photons and with one converted photon, and 1.01 for the decay mode with one Dalitz π^0 .

Only one candidate per event can be correct: the probability that the reconstructed decay is present on the tag and on the signal side is less than 10^{-12} (the branching fraction for $B^0 \rightarrow \pi^0 \pi^0$ is around 10^{-6}). Thus, for each B_{sig}^0 -decay mode, only one candidate is selected. To perform this, the candidates are ranked according to the flavor dilution factor r , and the candidate with the highest r value is taken as best candidate. A larger dilution factor means better flavor discrimination. However, the fraction of events requiring such a selection is relatively rare and thus the way of selecting the best candidate for each decay mode has no considerable effect. Since the event multiplicity is so small, the best candidate could be selected even arbitrarily.

The best candidate selection is applied separately for each decay mode. An event still can contain one best candidate for more than one decay mode, e.g. one best candidate for the $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ and one best candidate for the $B_{\text{sig}}^0 \rightarrow \pi_{\gamma C \gamma}^0 \pi_{\gamma\gamma}^0$ decay mode. However, this is not the case for events containing B_{sig}^0 candidates with four ECL photons. They have no overlap with other decay modes and are ready to be analyzed after the best candidate selection.

Table 6.3 presents the number of events containing one best B_{sig}^0 candidate with four ECL photons. The numbers are given for correctly reconstructed signal events (denoted as sig), wrongly reconstructed signal events (denoted as WRS), and for $B^+ B^-$, $B^0 \bar{B}^0$, continuum, and $\tau^+ \tau^-$ events. In wrongly reconstructed signal events, the B_{sig}^0 candidate is a random combination of particles which do not match the generated signal B^0 meson. The numbers of correctly and of wrongly reconstructed signal events were determined on a sample corresponding to $73 \times 50 \text{ ab}^{-1}$, and scaled down to the expected Belle II integrated luminosity of 50 ab^{-1} . These numbers, as well as all the numbers for signal events in the following discussion, were obtained assuming the PDG branching fraction $\mathcal{B}^{\pi^0 \pi^0} = 1.91 \cdot 10^{-6}$ [28]. For an integrated luminosity of 50 ab^{-1} , the total number of expected $B^0 \rightarrow \pi^0 \pi^0$ decays at Belle II is about 103,000. For non-signal events, the numbers are given for the integrated luminosities of the available MC samples, and for 50 ab^{-1} . Table 6.3 presents also the fraction of wrongly

		50 ab ⁻¹
	N_{sig}	15068
	N_{WRS}	153
		4 ab ⁻¹
$N_{B^+B^-}$	827	10337
$N_{B^0\bar{B}^0}$	136	1700
		1.8 ab ⁻¹
N_{cont}	2442	67805
$N_{\tau^+\tau^-}$	17	472
$N_{\text{WRS}}/(N_{\text{sig}} + N_{\text{WRS}})$		1.0%
Purity		15.8%

Table 6.3: Number of events passing the final selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$ decay mode: correctly reconstructed signal events N_{sig} , wrongly reconstructed signal events N_{WRS} , continuum events N_{cont} , $\tau^+\tau^-$ events, as well as $B^0\bar{B}^0$ and B^+B^- events. The numbers of correctly and of wrongly reconstructed signal events were determined on a sample corresponding to $73 \times 50 \text{ ab}^{-1}$ and scaled down to 50 ab^{-1} . The other numbers are given for the integrated luminosities of the available MC samples. In the right column, these numbers are scaled up to a total integrated luminosity of 50 ab^{-1} . The fraction of wrongly reconstructed signal events (WRS), and the purity value are given for the scaled numbers.

reconstructed signal events and the purity. The purity is defined as

$$\text{Purity} = \frac{N_{\text{sig}}}{N_{\text{sig}} + N_{\text{WRS}} + N_{B^+B^-} + N_{B^0\bar{B}^0} + N_{\text{cont}} + N_{\tau^+\tau^-}}.$$

The background $B\bar{B}$ events were obtained from the available 4 ab^{-1} MC sample. For B^+B^- events, the largest contribution comes from resonant $B^+ \rightarrow \rho^+(\rightarrow \pi^+\pi^0)\pi^0$ decays, and to a lesser extent, from non-resonant $B^+ \rightarrow \pi^+\pi^0\pi^0$ decays, where in each case the π^+ gets lost. For $B^0\bar{B}^0$ events, the largest contribution comes from $B^0 \rightarrow \pi^0\pi^0\pi^0$ decays and other similar decays with three π^0 s in the final state, such as $B^0 \rightarrow K_S^0(\rightarrow \pi^0\pi^0)\pi^0$ and $B^0 \rightarrow f_0(\rightarrow \pi^0\pi^0)\pi^0$. In these events, one of the π^0 s gets lost.

Considering now the events containing B_{sig}^0 candidates with one converted photon or with one Dalitz π^0 , for an integrated luminosity of 50 ab^{-1} , only 348 events pass the best candidate selections. Around half of them have two B_{sig}^0 candidates: one passing the best candidate selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$ mode, and one passing the best candidate selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ mode. This occurs because of the similarities between the e^+e^- pairs from conversions and from Dalitz decays explained in Sect. 6.4. Figure 6.26 shows a pie chart illustrating the situation. The events can be sorted into three groups. The events in Group I contain only one B_{sig}^0 candidate passing the best candidate selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ decay mode. The events in Group II contain only one B_{sig}^0 candidate passing the best candidate selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$ decay mode. And the events in Group III are those containing two best candidates, one for each decay mode.

One candidate has to be selected per event, and there are several possible ways to perform this. The most simple way is to prefer one of the two reconstructed decay modes. In this case,

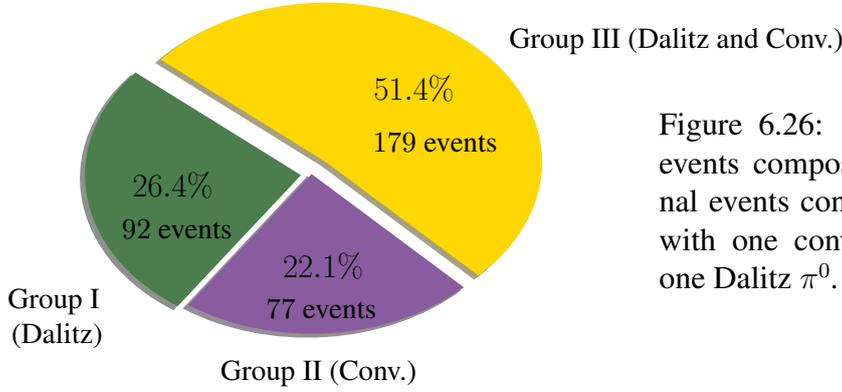


Figure 6.26: The three groups of events composing the total 348 signal events containing B_{sig}^0 candidates with one converted photon or with one Dalitz π^0 .

Table 6.4: Number of signal events where B_{sig}^0 matches the generated signal B^0 (denoted as sig), fractions of signal events with true Dalitz π^0 s and with true converted photons, fraction of signal events where B_{sig}^0 does not match the generated signal B^0 (denoted as WRS), and purity value for two selection options.

Option 1: $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates have priority.

Candidate type	N_{sig}	$N_{\text{Dal}}/N_{\text{sig}}$ [%]	$N_{\text{Conv}}/N_{\text{sig}}$ [%]	$N_{\text{WRS}}/(N_{\text{sig}} + N_{\text{WRS}})$ [%]	Purity
$B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$	271	54.2	45.8	1.1	17.7
$B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$	77	28.8	71.2	3.9	14.8

Option 2: $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$ candidates have priority.

Candidate type	N_{sig}	$N_{\text{Dal}}/N_{\text{sig}}$ [%]	$N_{\text{Conv}}/N_{\text{sig}}$ [%]	$N_{\text{WRS}}/(N_{\text{sig}} + N_{\text{WRS}})$ [%]	Purity
$B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$	92	46.5	53.5	1.3	14.0
$B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$	256	49.5	50.5	1.9	18.0

one has two options, either to select the B_{sig}^0 candidate with the reconstructed Dalitz π^0 when an event belongs to Group III, or to select the B_{sig}^0 candidate with the reconstructed converted photon. For the two selection options, Tab. 6.4 presents the number of signal events where B_{sig}^0 matches the generated signal B^0 , the fractions of signal events with true Dalitz π^0 s and with true converted photons, the fraction of signal events where B_{sig}^0 does not match the generated B^0 , and the purity. The numbers are for 50 ab^{-1} . Note that for each candidate type there is always a large fraction (in three of the cases around a half) of events which corresponds in reality to the other type. And candidates with reconstructed converted photons have slightly larger fractions of wrongly reconstructed candidates. The shapes of the distributions for these

		50 ab ⁻¹	
	N_{Dal}	147	
	N_{Conv}	124	
	N_{WRS}	3	
		4 ab ⁻¹	
$N_{B^+B^-}$	12	150	
$N_{B^0\bar{B}^0}$	0	-	
		1.8 ab ⁻¹	
N_{cont}	40	1111	
$N_{\tau^+\tau^-}$	0	-	

Table 6.5: Number of events passing the final selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ decay mode: Dalitz events N_{Dal} , conversion events N_{C} , wrongly reconstructed signal events N_{WRS} , as well as $B^0\bar{B}^0$, B^+B^- , continuum, and $\tau^+\tau^-$ events. The numbers of correctly and of wrongly reconstructed signal events were determined on a sample corresponding to $73 \times 50 \text{ ab}^{-1}$ and scaled down to 50 ab^{-1} . The other numbers are given for the integrated luminosities of the available MC samples. All numbers are scaled to 50 ab^{-1} .

candidates are very difficult to parametrize when performing a maximum-likelihood fit and thus, this is one first argument in favor of option 1.

As shown in Sect. 6.6, the B_{sig}^0 -vertex resolution is much better for candidates with one Dalitz π^0 than for candidates with one conversion. In the case of option two, almost half of the candidates with one conversion corresponds in reality to decays with one Dalitz π^0 . Thus, reconstructing the vertex of the Dalitz π^0 as a conversion would cause a considerable deterioration of the B_{sig}^0 -vertex resolution. Furthermore, as also shown in Sect. 6.6, the Δt resolution for decays with conversions that are wrongly reconstructed as decays with Dalitz π^0 s is better than for those reconstructed correctly. Therefore, for this thesis the priority was given to the Dalitz candidates as in option 1.

For the time-dependent CP -violation analysis, only the $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates (271 events) were taken. The remaining $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{C}\gamma}^0 \pi_{\gamma\gamma}^0$ candidates (77 events) were excluded because of their worse B_{sig}^0 -vertex resolution and their worse purity. These events could be recovered in future when the reconstruction of electrons and converted photons will improve.

Table 6.5 presents the number of events of each kind passing the final selection for the $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ decay mode. The total number of 271 events matching the generated signal B^0 is composed of 147 events with true Dalitz π^0 s, and 124 events with true converted photons. These numbers and the number of signal events where B_{sig}^0 does not match the generated signal B^0 (WRS) were determined on a sample with a large number of events corresponding to $73 \times 50 \text{ ab}^{-1}$, and scaled down to 50 ab^{-1} . The other numbers are given for the luminosities of the available MC samples. When the number of events for one of the samples differed from zero, it was projected to 50 ab^{-1} .

Considering the background $B\bar{B}$ events, only twelve B^+B^- events passed the final selection. Of these events, eleven contain a resonant $B^+ \rightarrow \rho^+ (\rightarrow \pi^+ \pi^0) \pi^0$ decay (or the charge conjugate) and one contains a non-resonant $B^- \rightarrow \pi^- \pi^0 \pi^0$ decay. In all events, the charged pion is lost, and one of the π^0 s either has a converted photon or is a Dalitz π^0 .

For the three considered decay modes, Tab. 6.6 presents the absolute reconstruction efficiency and the efficiency after the final selection. For events with converted photons and with Dalitz π^0 s, the efficiencies correspond only to those reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. To perform the set of pseudo-experiments, it was verified that the final selection efficiency is constant over the whole Δt fit range.

Table 6.6: Fraction of generated events in the acceptance $N_{\text{gen}}^{\text{acc}}/N_{\text{gen}}$, reconstruction efficiency $N_{\text{rec}}/N_{\text{gen}}^{\text{acc}}$ and efficiency after final selection $N_{\text{rec}}^{\text{FS}}/N_{\text{gen}}^{\text{acc}}$ (the efficiencies are normalised to the number of generated events in the acceptance $N_{\text{gen}}^{\text{acc}}$). Events with converted photons and Dalitz π^0 s (first and second rows) were reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. The highlighted row corresponds to the whole set used for time-dependent CP -analysis.

Decay. Channel	$N_{\text{gen}}^{\text{acc}}/N_{\text{gen}}$ [%]	$N_{\text{rec}}/N_{\text{gen}}^{\text{acc}}$ [%]	$N_{\text{rec}}^{\text{FS}}/N_{\text{gen}}^{\text{acc}}$ [%]
$B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$	2.0	52.0	7.2
$B_{\text{sig}}^0 \rightarrow \pi_{\gamma\text{c}\gamma}^0 \pi_{\gamma\gamma}^0$	3.0	48.8	4.2
Dal + Conv	5.0	50.1	5.4
$B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$	76.2	86.0	19.2

6.10 Estimation of precision with pseudo experiments

In this chapter, the goal is to estimate whether Belle II will be able to measure the mixing induced CP -violation parameter $\mathcal{S}^{\pi^0\pi^0}$ after collecting its full integrated luminosity of 50 ab^{-1} . The other physical parameters related to the decay $B^0 \rightarrow \pi^0\pi^0$, namely the branching fraction $\mathcal{B}^{\pi^0\pi^0}$ and the direct CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$, have been measured already, but the precision is still limited by the sizes of the samples at Belle and BaBar.

The estimation of the precision for $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{B}^{\pi^0\pi^0}$ requires to formulate a model describing the distributions of the discriminating variables ΔE and m_{bc} . For $\mathcal{S}^{\pi^0\pi^0}$, a model describing these variables as well as the Δt distribution is required. Therefore, two independent models were developed. The first one describes the ΔE and the m_{bc} distributions of B_{sig}^0 candidates with four ECL photons. This model is used to estimate the precision for $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{B}^{\pi^0\pi^0}$. The second model describes the ΔE and the m_{bc} distributions for candidates with a reconstructed Dalitz π^0 . It describes as well the respective Δt distributions and the output distributions of the Dalitz-Conversion classifier, and is used to estimate the expected precision for $\mathcal{S}^{\pi^0\pi^0}$.

In an eventual analysis of the full data sample, the physical parameters describing the two models would be used to extract the values of $\mathcal{A}^{\pi^0\pi^0}$, $\mathcal{B}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ from fits to the data. Since the data are not yet available, the procedure is simulated in pseudo-experiments by using MC samples with sizes corresponding to the expected ones.

In the following, the fit technique is briefly introduced. Next, the two developed models and the performed pseudo-experiments will be explained. The validation of the procedure and the results of the pseudo-experiments will be presented afterward.

6.10.1 The extended unbinned maximum-likelihood fit

A standard technique to extract the values of physical parameters from a fit of a model to the data is the maximum-likelihood fit. A detailed description of this technique focusing on its use at B factories can be found in [29].

The first step in the procedure is to find appropriate probability density functions (PDF) describing the distribution of each observable used in the fit. This is commonly referred to as “model building”. Each PDF depends on a certain number of parameters. The parameters can be either physical parameters of interest, or parameters representing quantities that affect the relation between the physical parameters and the observables used in the fit. For example, in time-dependent analyses, the Δt PDF depends on the CP -violation parameters, which are the targets of the whole effort, and the parameters ε , μ , w , and Δw characterizing the performance of the flavor tagger (see Sect. 4.2). When the observables used in the fit are uncorrelated, their PDFs can be multiplied to form a multidimensional PDF. Otherwise the situation is more complicated and, for example, conditional PDFs are required.

In general, several signal and background components contribute and the final PDF is formed by adding the products of the PDF times the event fraction for each component. The event fractions, as well as the parameters of the individual PDFs, build the total set of parameters on which the final PDF depends.

Given a final PDF, its value at a certain point in its space is defined as a *likelihood*. For measurements consisting of several statistically independent points following the same PDF, the total likelihood is the product of the likelihood at each point. Such a total likelihood is called *unbinned* because it is evaluated at each data point, and no binning of the data is needed.

When the probability to observe a certain number of events, i.e. data points, depends on the parameters of the final PDF, this has to be taken into account in the total likelihood. Usually, the number of expected events follows a Poisson distribution. In this case, the total likelihood is a product of the Poisson distribution times the product of the likelihood at each data point. Such a likelihood is called *extended*. And the number of events for a certain component is often called *yield*.

The parameters of the final PDF are determined by maximizing the total likelihood on data points. In other words, the measured parameters are those for which the total likelihood is maximal on a given set of data points. In practice, the logarithm of the total likelihood multiplied by -2 is minimized. This so-called negative log-likelihood has the same extremum and is easier to calculate.

The parameters of the final PDF that are not physical parameters are determined on MC samples or on control samples. For example, considering again the Δt PDF of time-dependent CP -violation analyses, the parameters ε , μ , w , and Δw are determined from data on independent control samples of self-tagging modes, while the CP -violation parameters are determined from the data on the signal channel under investigation.

For the pseudo-experiments presented in this chapter, the parameters characterizing the performance of the flavor tagger were measured on an official MC validation sample as described in Sect. 6.8. The parameters of the background components were determined on the official MC samples described in Sect. 6.2; and the parameters of the signal components were determined on a signal MC sample, generated as explained in Sect. 6.2, with a large number of signal events corresponding to 73 times the full Belle II integrated luminosity.

For the extended unbinned maximum-likelihood fit, the `RoobarFit` package was employed [148]. It is a user friendly framework developed at BaBar basing on `ROOT` and `Roofit`, which use the fitting software `MINUIT` for numerical minimization [149, 150].

6.10.2 The four-photon model

For events with a $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$ candidate, the yield of correctly reconstructed signal events N_{sig} and the direct CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ are extracted from an extended unbinned maximum-likelihood fit in three dimensions. The three observables used in the fit are

- $m_{\text{bc}} \in [5.26, 5.29] \text{ GeV}/c^2$,
- $\Delta E \in [-0.3, 0.2] \text{ GeV}$,
- $q \in \{-1, 1\}$,

where the flavor q is the sign of the flavor tagger output y_{FBDT} . The event model includes four components:

- correctly reconstructed signal events, which will be called just signal events hereafter,
- wrongly reconstructed signal events (WRS),
- $B\bar{B}$ events excluding the signal channel,
- continuum events (hadronic events from light quark pairs and $\tau^+\tau^-$ events).

Each component has its own three-dimensional PDF which is the product of the PDFs describing the distributions of the used observables individually:

$$\mathcal{P}_j(m_{\text{bc}}, \Delta E, q) \equiv \mathcal{P}_j(m_{\text{bc}}) \times \mathcal{P}_j(\Delta E) \times \mathcal{P}_j^{\text{obs}}(q),$$

where j extends over the four components given above. Forming the three-dimensional PDF as a product of individual PDFs neglects the correlations between the observables used in the fit. The correlation coefficient between m_{bc} and ΔE is largest for signal events, determined by MC to around 9%, and is lowest for continuum events (around 1%). The output of the flavor tagger is found to be uncorrelated with m_{bc} and ΔE . Correlations of more than 1% have to be carefully taken into account to keep systematic effects under control. Since control samples were not available for the study of systematic effects, and since the model neglecting the correlations led to almost fully unbiased results, the correlations between m_{bc} and ΔE could be safely ignored.

The total extended likelihood is given by

$$\mathcal{L} \equiv \frac{e^{-\sum_j N_j}}{N!} \prod_{i=1}^N \sum_j N_j \cdot \mathcal{P}_j(m_{bc}^i, \Delta E^i, q^i),$$

where i extends over the events in the MC sample used for a pseudo-experiment, and j over the four components: signal, WRS, $B\bar{B}$, and continuum. The free parameters of the fit are the four yields N_{sig} , N_{WRS} , $N_{B\bar{B}}$ and N_{cont} , as well as the CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$. All other parameters are fixed. The PDFs for the different components have no common parameters. Each one is independent. In the following, the PDFs are described explicitly for each signal and background component. The results of the fits will be presented in Sect. 6.10.5 and 6.10.6.

Signal events

Signal events contain a B_{sig}^0 candidate matching the generated signal B^0 meson. For these events, the PDFs describing the m_{bc} and the ΔE distributions are determined using a large MC sample. The PDF describing the m_{bc} distribution is the sum of an Argus function (see App. C.2) and four Gaussian functions,

$$\begin{aligned} \mathcal{P}_{\text{sig}}(m_{bc}) \equiv & f_A \cdot \text{ARGUS}(m_{bc}; a, c) + f_{3G} \cdot \left(f_1 \cdot G(m_{bc}; \mu_1, \sigma_1) \right. \\ & \left. + f_2 \cdot G(m_{bc}; \mu_2, \sigma_2) + (1 - f_1 - f_2) \cdot G(m_{bc}; \mu_3, \sigma_3) \right) \\ & + (1 - f_A - f_{3G}) \cdot G(m_{bc}; \mu_4, \sigma_4). \end{aligned} \quad (6.5)$$

And the ΔE PDF is the sum of two Gaussian functions and a Crystal Ball function (see App. C.3),

$$\begin{aligned} \mathcal{P}_{\text{sig}}(\Delta E) \equiv & f_1 \cdot G(\Delta E; \mu_1, \sigma_1) + f_2 \cdot G(\Delta E; \mu_2, \sigma_2) \\ & + (1 - f_1 - f_2) \cdot \text{CB}(\Delta E; \alpha, n, \mu_3, \sigma_3). \end{aligned}$$

The parameters of the m_{bc} and the ΔE PDFs are independent from each other. They are fixed after fitting the PDFs to the MC sample. Figure 6.27 shows the fit projections for m_{bc} and for ΔE on signal MC events.

For signal events, the time-integrated PDF describing the CP -Asymmetry is given by

$$\mathcal{P}_{\text{sig}}^{\text{obs}}(q) \equiv \frac{1}{2} \left[1 - q \cdot \Delta w + q \cdot (1 - 2w) \cdot \mathcal{A}^{\pi^0\pi^0} (1 - 2 \cdot \chi_d) \right],$$

where the parameters w , and Δw were determined as explained in Sect. 6.8. The CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ is a free parameter of the fit.

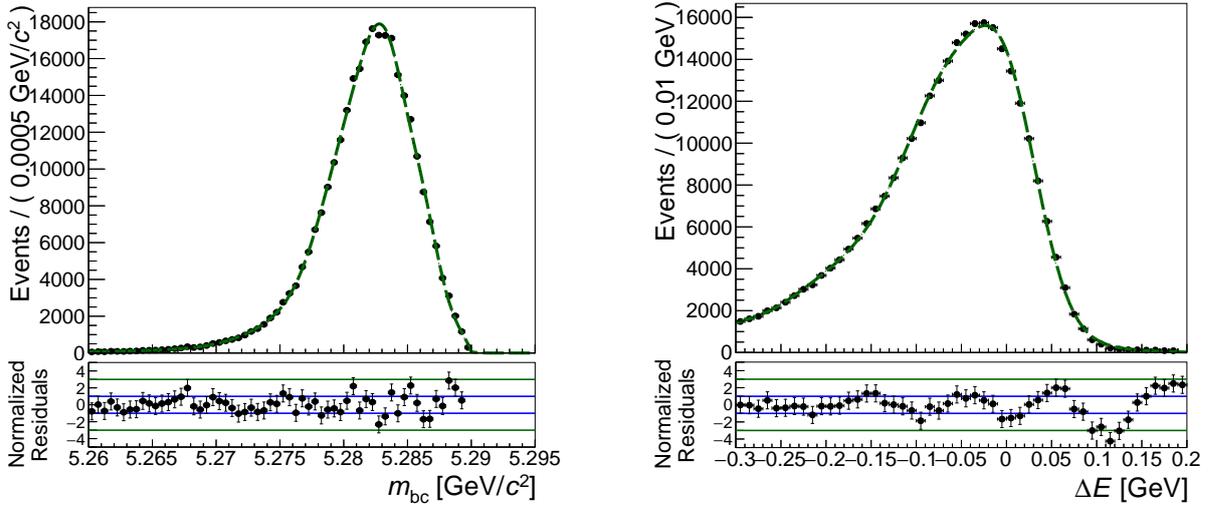


Figure 6.27: Fit projections for true signal $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$ candidates. The fit projections are shown onto m_{bc} (left) and ΔE (right). The points (with error bars) represent the MC sample used to find the PDFs.

Wrongly reconstructed signal events

Wrongly reconstructed signal events (WRS) contain a B_{sig}^0 candidate that does not match the generated signal B^0 meson. The PDFs for these events are determined on the same MC sample used for signal events. The m_{bc} PDF for wrongly reconstructed signal events is the sum of an Argus function and three Gaussian functions (a fourth one turned out not to be necessary),

$$\begin{aligned} \mathcal{P}_{\text{WRS}}(m_{bc}) \equiv & f_A \cdot \text{ARGUS}(m_{bc}; a, c) + (1 - f_A) \cdot \left(f_1 \cdot G(m_{bc}; \mu_1, \sigma_1) \right. \\ & \left. + f_2 \cdot G(m_{bc}; \mu_2, \sigma_2) + (1 - f_1 - f_2) \cdot G(m_{bc}; \mu_3, \sigma_3) \right). \end{aligned} \quad (6.6)$$

And the ΔE PDF is the sum of three Gaussian functions,

$$\begin{aligned} \mathcal{P}_{\text{WRS}}(\Delta E) \equiv & f_1 \cdot G(\Delta E; \mu_1, \sigma_1) + f_2 \cdot G(\Delta E; \mu_2, \sigma_2) \\ & + (1 - f_1 - f_2) \cdot G(\Delta E; \mu_3, \sigma_3). \end{aligned} \quad (6.7)$$

The parameters of the m_{bc} and the ΔE PDFs are also found to be independent from each other, and are fixed after fitting the PDFs to the MC sample. Figure 6.28 shows the fit projections for m_{bc} and for ΔE on wrongly reconstructed signal MC events.

The time-integrated PDF for wrongly reconstructed signal events is

$$\mathcal{P}_{\text{WRS}}^{\text{obs}}(q) \equiv \frac{1}{2} \left[1 - q \cdot \Delta w^{\text{WRS}} \right], \quad (6.8)$$

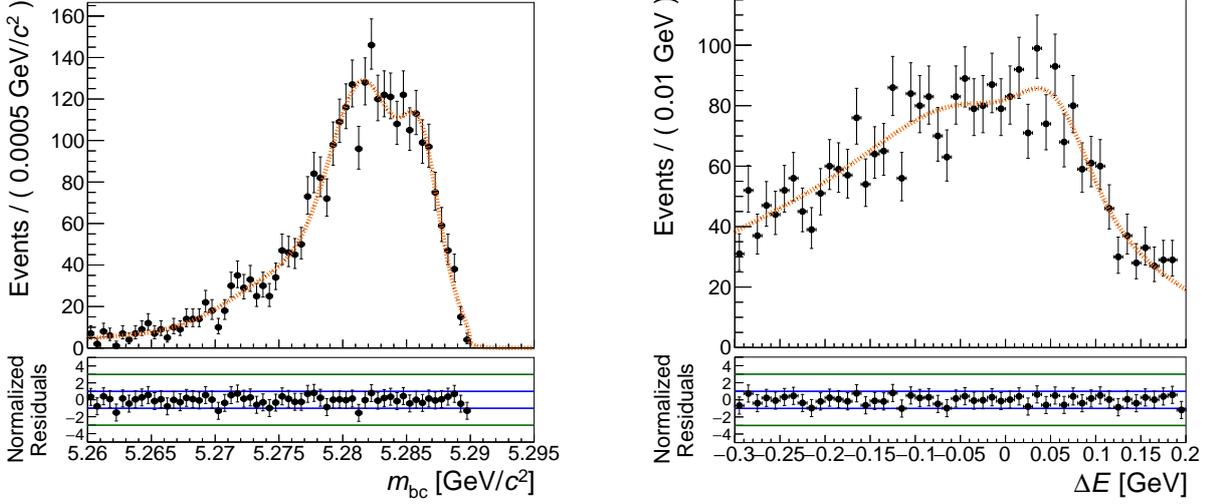


Figure 6.28: Fit projections for wrongly reconstructed signal $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$ candidates (WRS). The fit projections are shown onto m_{bc} (left) and ΔE (right). The points (with error bars) represent the MC sample used to find the PDFs.

where Δw^{WRS} is modified to match the asymmetry in the fractions of wrongly reconstructed signal events tagged as B^0 or as \bar{B}^0 . The number found on MC is $\Delta w^{\text{WRS}} = (-19 \pm 6)\%$.

$B\bar{B}$ events

Most of the background $B\bar{B}$ events come from events where one of the two B mesons decays into a final state with three pions and one of the pions is lost. The largest contributions come from $B^+ \rightarrow \rho^+(\rightarrow \pi^+\pi^0)\pi^0$ decays where the π^+ is lost, and from $B^0 \rightarrow \pi^0\pi^0\pi^0$ decays where one π^0 is lost. The m_{bc} distribution for these events peaks almost at the same value as for signal events. But the ΔE distribution is shifted towards negative values due to the missing pion. Since the distributions are similar for charged and for neutral $B\bar{B}$ events, they are grouped together to form one single background class.

Figure 6.29 shows the fit projections for m_{bc} and for ΔE on background $B\bar{B}$ events. The m_{bc} PDF is the sum of an Argus function and three Gaussian functions as for wrongly reconstructed events (eq. (6.6)). The ΔE PDF can be described by a single Gaussian function. The time-integrated PDF for $B\bar{B}$ events is defined in the same way as for wrongly reconstructed signal events (eq. (6.8)). The parameters of the PDFs are fixed after fitting them to the official $B\bar{B}$ MC sample corresponding to 4 ab^{-1} .

Continuum events

The largest background contribution comes from continuum events. Continuum events and $\tau^+\tau^-$ events are grouped together to form a single background class because they have similar m_{bc} and ΔE distributions. This background class will be called just continuum hereafter. The m_{bc} distribution is described by the PDF $\mathcal{P}_{\text{cont}}(m_{bc})$ with an ARGUS function. The ΔE PDF

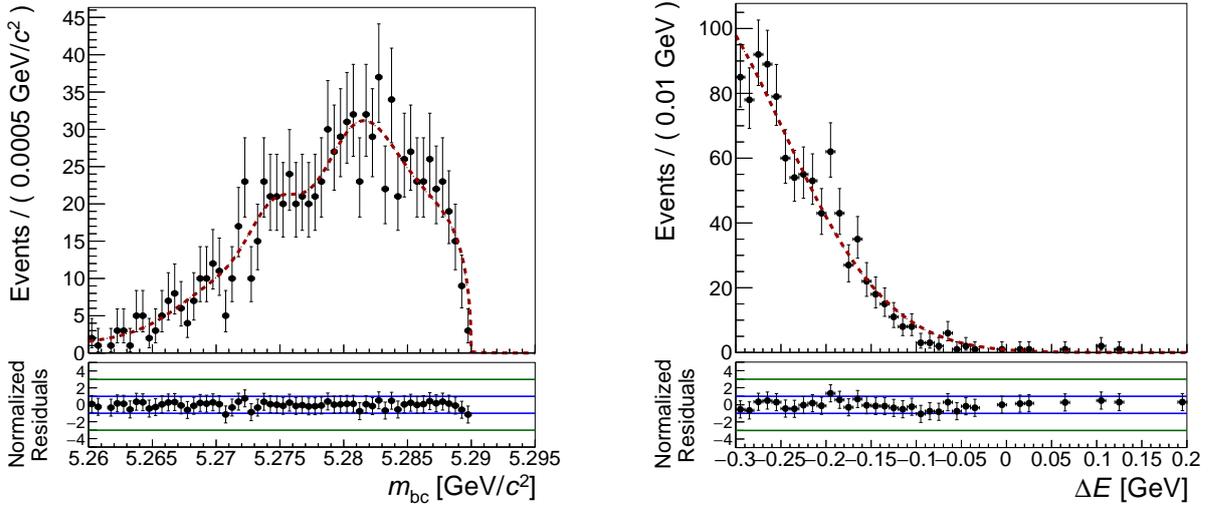


Figure 6.29: Fit projections for background $B\bar{B}$ candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$. The fit projections are shown onto m_{bc} (left) and ΔE (right). The points (with error bars) represent the MC sample used to find the PDFs.

is a first order polynomial,

$$\mathcal{P}_{\text{cont}}(\Delta E) \equiv a_0 + a_1 \cdot \Delta E, \quad (6.9)$$

where a_0 and a_1 are constants. Figure 6.30 shows the fit projections for m_{bc} and for ΔE on continuum. The parameters of the functions are fixed after fitting the PDFs to the official MC sample corresponding to 1.8 ab^{-1} .

For continuum events there is no CP asymmetry and thus $\mathcal{P}_{\text{cont}}^{\text{obs}}(q) = 1$.

6.10.3 The Δt resolution function

To measure the value of $\mathcal{S}^{\pi^0\pi^0}$, a time-dependent CP violation-analysis has to be performed. Such an analysis requires the description of the measured Δt distribution. The theoretical Δt PDF for time dependent CP -violation analyses $\mathcal{P}^{\text{obs}}(q, \Delta t, \varepsilon, \mu, w, \Delta w)$ (eq. (4.7)) is derived from the quantum mechanical PDF $\mathcal{P}(\Delta t, q)$ (eq. (2.33)) and includes the effects caused by the non-perfect performance of the flavor tagger. It excludes, however, the effects caused by the detector alignment, the detector resolution and the vertex reconstruction algorithms on the measured value of Δt . At Belle and at BaBar, these effects were taken into account by convoluting the theoretical Δt PDF with a resolution function $\mathcal{R}(\delta t, \sigma_{\Delta t})$ (see eq. (4.21) in Sect. 4.14).

For the pseudo-experiments presented in this section, the sum of three Gaussian functions was used as Δt resolution function. This approach was followed by BaBar. At BaBar, the mean values and the widths of the two central Gaussian functions scaled according to $\sigma_{\Delta t}$

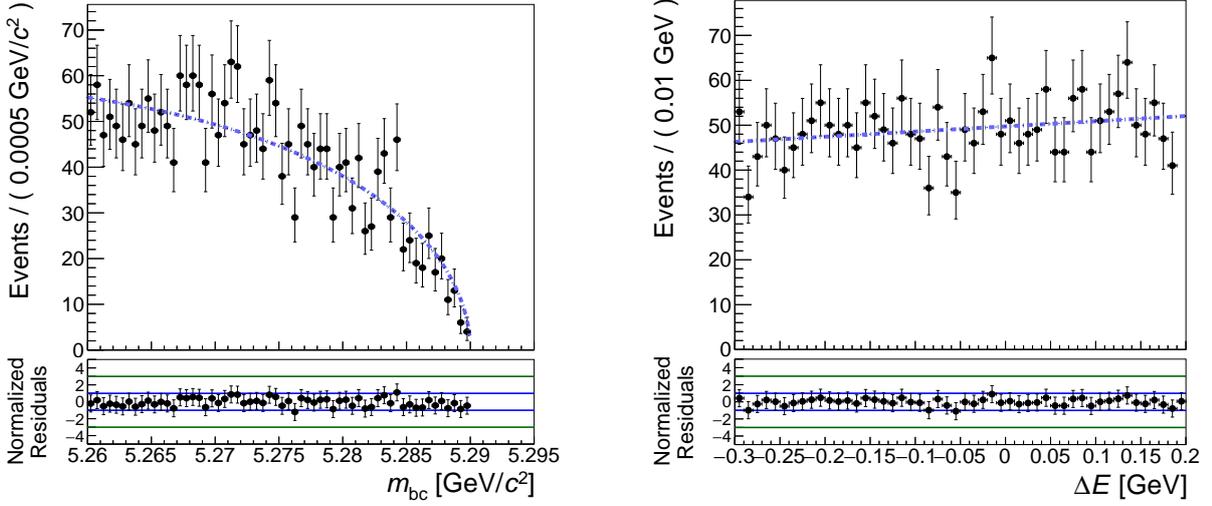


Figure 6.30: Fit projections for continuum $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$ candidates. The fit projections are shown onto m_{bc} (left) and ΔE (right). The points (with error bars) represent the MC sample used to find the PDFs.

on an event-by-event basis. The event-by-event corrections accounted for systematic effects arriving from the changes in the experimental conditions.

In the absence of control samples from data, event-dependent systematic effects affecting the measurement of Δt could not be studied. The simulated experimental conditions were also constant. Therefore, the parameters of the function were set to constant values for all events. Using the three Gaussian resolution function, the final Δt PDF described very well the distribution of the reconstructed Δt . And the results for the CP -violation parameters were unbiased. Additional tests where the B^0 lifetime was measured showed also no bias, and neither over- nor underestimation of the error.

Denoting the mean values and the widths of the three Gaussian functions by μ_i and σ_i , where $i = \text{core, tail, and outlier}$ (in order of increasing width), the Δt resolution function used in this thesis is given by

$$\begin{aligned} \mathcal{R}(\delta t) \equiv & f_{\text{core}} \cdot G(\delta t; \mu_{\text{core}}, \sigma_{\text{core}}) + (1 - f_{\text{core}} - f_{\text{outlier}}) \cdot G(\delta t; \mu_{\text{tail}}, \sigma_{\text{tail}}) \\ & + f_{\text{outlier}} \cdot G(\delta t; \mu_{\text{outlier}}, \sigma_{\text{outlier}}), \end{aligned}$$

where the parameters f_i stay for the respective fractions of the Gaussian functions.

The pseudo-experiments were performed for six different combinations of $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ (see Sect. 6.2). Since the B_{tag}^0 -decay vertex distribution changes depending on the CP -violation parameters, the parameters of the resolution function were determined for each sample. In general, the parameters of the resolution function were adapted to match the Δt distribution for each type of events.

Considering now $e^+e^- \rightarrow q\bar{q}$ events, the proper-time distribution of the final state hadrons has a prompt component, and a lifetime component. To model the Δt PDF for $q\bar{q}$ events, the BaBar approach [29] was also followed: the three Gaussian Δt resolution function was convoluted with a delta function describing the prompt component; and the widths of the three Gaussian functions were adapted to fit the lifetime component.

6.10.4 The Dalitz-Conversion model

The Dalitz-Conversion model is more complex than the four-photon model. It was developed for events with $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates. In addition to m_{bc} and ΔE , it includes two more observables: the time difference Δt and the output of the Dalitz-Conversion classifier y_{DC} . The output y_{DC} is needed to distinguish between the two signal components, Dalitz and conversion events, which differ in their Δt resolutions. The background components are the same as for the four-photon model. However, to model the background PDFs only very few events were available because of the small conversion and Dalitz probabilities.

The yield of Dalitz events N_{Dal} , conversion events N_{Conv} , as well as the CP -violation parameters $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ are extracted from an extended unbinned maximum-likelihood fit in five dimensions. The five observables to be described by the fit are

- $m_{\text{bc}} \in [5.26, 5.29] \text{ GeV}/c^2$,
- $\Delta E \in [-0.3, 0.2] \text{ GeV}$,
- $y_{\text{CD}} \in [0, 1]$,
- $\Delta t \in [-25, 25] \text{ ps}$,
- $q \in \{-1, 1\}$.

The event model includes five components:

- Dalitz events,
- conversion events,
- wrongly reconstructed signal events (WRS),
- $B\bar{B}$ events,
- continuum events.

Each component has its own five-dimensional PDF which is the product of the PDFs describing the distributions of the observables:

$$\mathcal{P}_j(m_{\text{bc}}, \Delta E, y_{\text{CD}}, \Delta t, q) \equiv \mathcal{P}_j(m_{\text{bc}}) \times \mathcal{P}_j(\Delta E) \times \mathcal{P}_j(y_{\text{DC}}) \times \mathcal{P}_j^{\text{final}}(\Delta t, q),$$

where j extends over the five components. As for the four-photon model, the correlations between the variables were disregarded in the fit.

The total extended likelihood for the Dalitz-Conversion model is given by

$$\mathcal{L} \equiv \frac{e^{-\sum_j N_j}}{N!} \prod_{i=1}^N \sum_j N_j \cdot \mathcal{P}_j(m_{bc}^i, \Delta E^i, y_{CD}^i, \Delta t^i, q^i),$$

where i extends over the events in the MC sample used for a pseudo-experiment, and j over the five components: Dalitz, conversion, WRS, $B\bar{B}$, and continuum. The free parameters of the fit are the five yields N_{Dal} , N_{Conv} , N_{WRS} , $N_{B\bar{B}}$ and N_{cont} , as well as the CP -violation parameters $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$. All other parameters are fixed.

Only the Δt PDFs for Dalitz and for Conversion events share common parameters, namely the B^0 lifetime τ_{B^0} , the mass difference of the mass eigenstates Δm (see Sect. 2.5), the CP -violation parameters, and the parameters characterizing the performance of the flavor tagger. The mass difference Δm enters also the Δt PDF for wrongly reconstructed signal events. The other PDFs have no common parameters and are independent from each other. In the following, the single PDFs are described for each component.

Dalitz events

Dalitz events contain a $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidate matching the generated signal B^0 meson. And the reconstructed π_{Dal}^0 matches a generated Dalitz π^0 coming from the signal B^0 decay. The PDFs for these candidates are determined on the same MC sample used for the signal events of the four-photon model. The PDF describing the m_{bc} distribution is the sum of three Gaussian functions,

$$\begin{aligned} \mathcal{P}_{\text{Dal}}(m_{bc}) &\equiv f_1 \cdot G(m_{bc}; \mu_1, \sigma_1) + f_2 \cdot G(m_{bc}; \mu_2, \sigma_2) \\ &\quad + (1 - f_1 - f_2) \cdot G(m_{bc}; \mu_3, \sigma_3). \end{aligned} \quad (6.10)$$

The ΔE PDF is the sum of two Gaussian functions and a Crystal Ball function,

$$\begin{aligned} \mathcal{P}_{\text{Dal}}(\Delta E) &\equiv f_1 \cdot G(\Delta E; \mu_1, \sigma_1) + f_2 \cdot G(\Delta E; \mu_2, \sigma_2) \\ &\quad + (1 - f_1 - f_2) \cdot \text{CB}(\Delta E; \alpha, n, \mu_3, \sigma_3), \end{aligned} \quad (6.11)$$

The distribution of the output of the Dalitz-Conversion classifier y_{DC} is modeled performing an adaptive kernel density estimation (see App. C.4),

$$\mathcal{P}_{\text{Dal}}(y_{\text{DC}}) \equiv \text{Key}(y_{\text{DC}}). \quad (6.12)$$

And the final Δt PDF (see Sect. 6.10.3) is given by

$$\mathcal{P}_{\text{Dal}}^{\text{final}}(\Delta t, q) \equiv \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \left[1 - q \cdot \Delta w + q \cdot (1 - 2w) \right]$$

$$\cdot \left(\mathcal{A}^{\pi^0\pi^0} \cos(\Delta m \Delta t) + \mathcal{S}^{\pi^0\pi^0} \sin(\Delta m \Delta t) \right) \Big] \otimes \mathcal{R}_{\text{Dal}}(\delta t), \quad (6.13)$$

where the parameters w , and Δw were determined as explained in Sect. 6.8. The value of τ_{B^0} was fixed to the true MC B^0 lifetime of 1.525 ps. And the value of Δm was fixed to the true MC mass difference of 0.507 s^{-1} . Figure 6.31 shows the fit projections for m_{bc} , ΔE , y_{DC} and Δt on the Dalitz MC events.

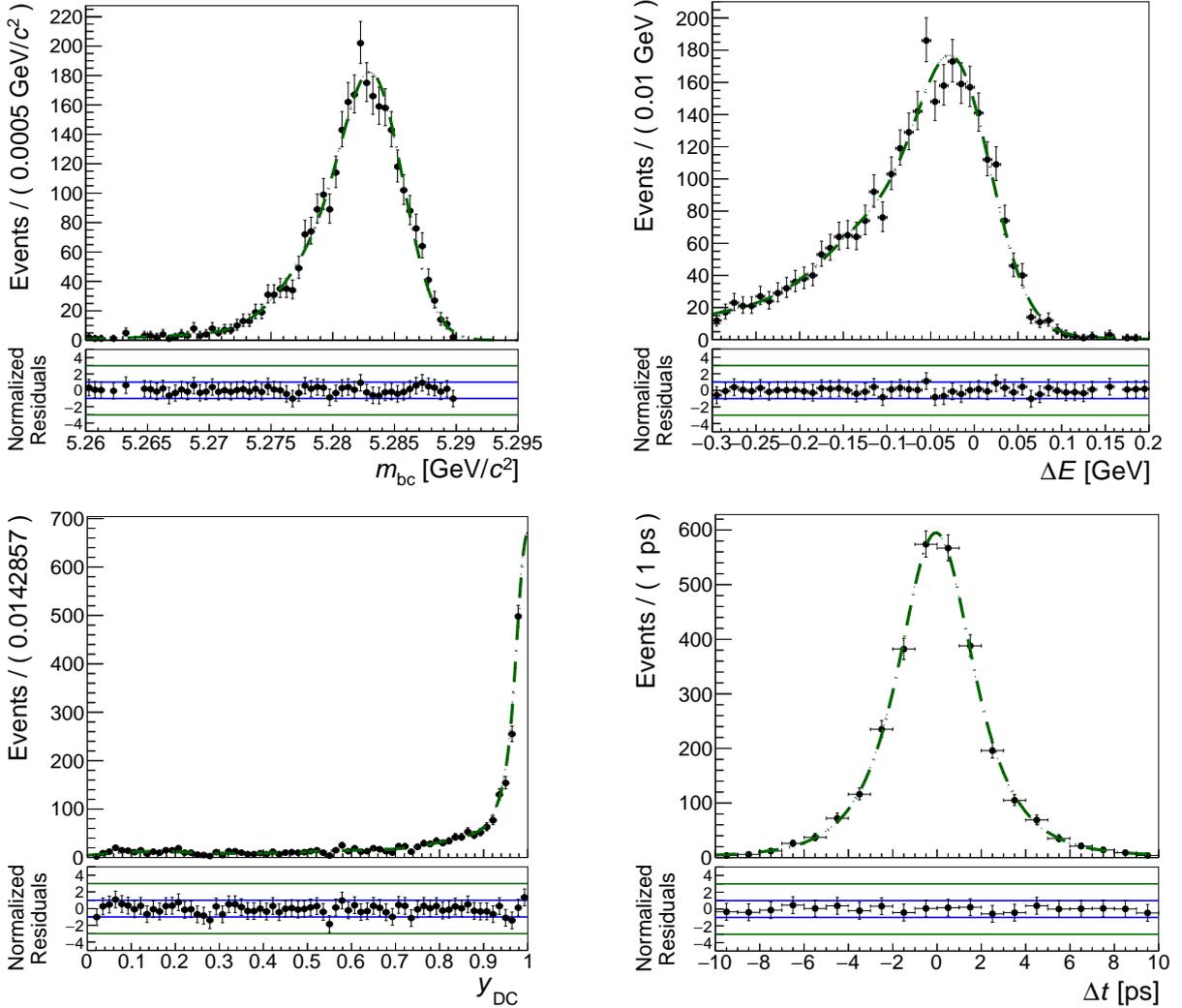


Figure 6.31: Fit projections for true Dalitz B_{sig}^0 candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. The fit projections are shown onto m_{bc} (top left), ΔE (top right), y_{DC} (bottom left), and Δt (bottom right). The points (with error bars) represent the MC sample used to find the PDFs.

Conversion events

Conversion events contain a $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidate matching the generated signal B^0 meson. But the reconstructed π_{Dal}^0 matches a generated π^0 with one converted photon coming from

the signal B^0 decay. The PDFs for these candidates are determined as well on the MC sample used for the signal events of the four-photon model.

The m_{bc} , the ΔE , and the y_{DC} PDFs for conversion events are defined in the same way as for Dalitz events: $\mathcal{P}_{\text{Conv}}(m_{bc})$ is the sum of three Gaussian functions as in eq. (6.10); $\mathcal{P}_{\text{Conv}}(\Delta E)$ is the sum of two Gaussian functions and a Crystal Ball function as in eq. (6.11); and $\mathcal{P}_{\text{Conv}}(y_{DC})$ is modeled performing an adaptive kernel density estimation as in eq. (6.12).

The final Δt PDF is given by

$$\mathcal{P}_{\text{Conv}}^{\text{final}}(\Delta t, q) \equiv \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \left[1 - q \cdot \Delta w + q \cdot (1 - 2w) \cdot \left(\mathcal{A}^{\pi^0\pi^0} \cos(\Delta m \Delta t) + \mathcal{S}^{\pi^0\pi^0} \sin(\Delta m \Delta t) \right) \right] \otimes \mathcal{R}_{\text{Conv}}(\delta t),$$

where the mass difference Δm , the B^0 lifetime τ_{B^0} , and the parameters w , Δw , $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ are the same parameters entering the final Δt PDF for Dalitz events (eq. (6.13)). Figure 6.32 shows the fit projections for m_{bc} , ΔE , y_{DC} and Δt on the Conversion MC events.

Wrongly reconstructed signal events

The PDFs for wrongly reconstructed signal events are determined also on the MC sample used for the signal events of the four-photon model.

The m_{bc} PDF, $\mathcal{P}_{\text{WRS}}(m_{bc})$, is an ARGUS function. The ΔE PDF, $\mathcal{P}_{\text{WRS}}(\Delta E)$, is a single Gaussian function. The distribution of y_{DC} is described by the sum of a two Gaussian function, a bifurcated Gaussian function (see App. C.1), and a first order polynomial,

$$\begin{aligned} \mathcal{P}_{\text{WRS}}(y_{DC}) \equiv & f_{2G} \cdot (f_1 \cdot G(y_{DC}; \mu_1, \sigma_1) + (1 - f_1) \cdot G(y_{DC}; \mu_2, \sigma_2)) \\ & + f_{Gb} \cdot G_b(y_{DC}; \mu_3, \sigma_L, \sigma_R) \\ & + (1 - f_{2G} - f_{Gb}) \cdot (1 + a \cdot y_{DC}). \end{aligned}$$

And the final Δt PDF is given by

$$\mathcal{P}_{\text{WRS}}^{\text{final}}(\Delta t, q) \equiv \frac{e^{-\frac{|\Delta t|}{\tau_{\text{WRS}}}}}{4\tau_{\text{WRS}}} \left[1 - q \cdot \Delta w^{\text{WRS}} + q \cdot (1 - 2w^{\text{WRS}}) \cdot \left(\mathcal{A}^{\text{WRS}} \cos(\Delta m \Delta t) + \mathcal{S}^{\text{WRS}} \sin(\Delta m \Delta t) \right) \right] \otimes \mathcal{R}_{\text{WRS}}(\delta t),$$

where the parameters w^{WRS} , Δw^{WRS} , \mathcal{A}^{WRS} and \mathcal{S}^{WRS} were determined by fitting the Δt distribution. The lifetime τ_{WRS} is fixed to the same value of the B^0 lifetime τ_{B^0} . However, τ_{WRS} and τ_{B^0} are independent fit parameters. Figure 6.33 shows the fit projections for m_{bc} , ΔE , y_{DC} and Δt on wrongly reconstructed signal MC events.

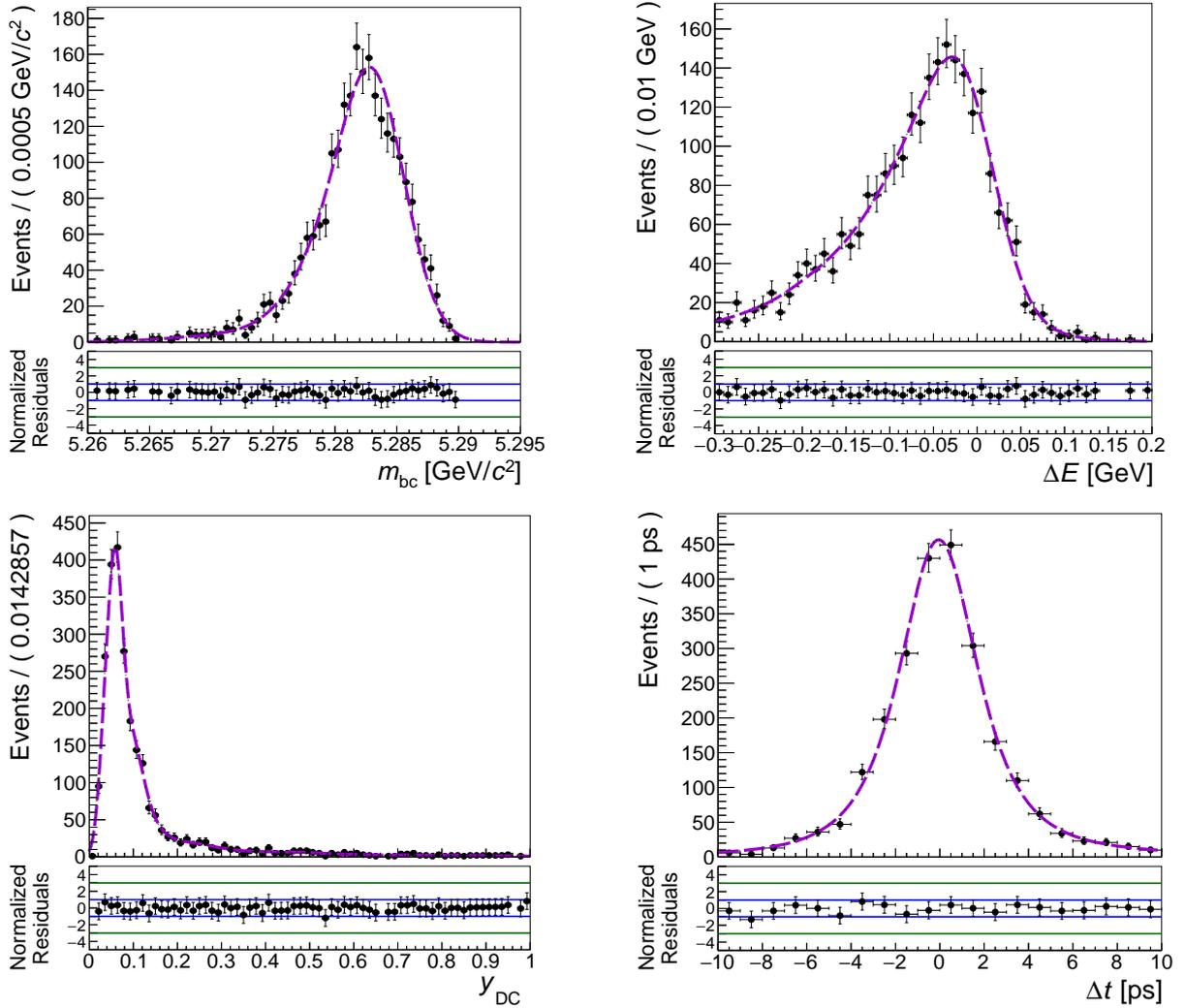


Figure 6.32: Fit projections for true conversion B_{sig}^0 candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. The fit projections are shown onto m_{bc} (top left), ΔE (top right), y_{DC} (bottom left), and Δt (bottom right). The points (with error bars) represent the MC sample used to find the PDFs.

$B\bar{B}$ events

The few background $B\bar{B}$ events in 4 ab^{-1} come all from B^+B^- events where one of the B mesons decayed into a final state with one charged and two neutral pions. The charged pion is always lost, and one of the neutral pions either has a converted photon or is a Dalitz π^0 . The m_{bc} and the ΔE distributions resemble those of four-photon $B\bar{B}$ events: m_{bc} peaks almost at the same value as the signal events, and ΔE is negatively shifted due to the missing pion.

The m_{bc} distribution is described by a PDF, $\mathcal{P}_{B\bar{B}}(m_{bc})$, which is an ARGUS function. The ΔE PDF, $\mathcal{P}_{B\bar{B}}(\Delta E)$, is a single Gaussian function. The y_{DC} PDF, $\mathcal{P}_{B\bar{B}}(y_{\text{DC}})$, is modeled performing a kernel density estimation as for Dalitz and for Conversion events. And the final

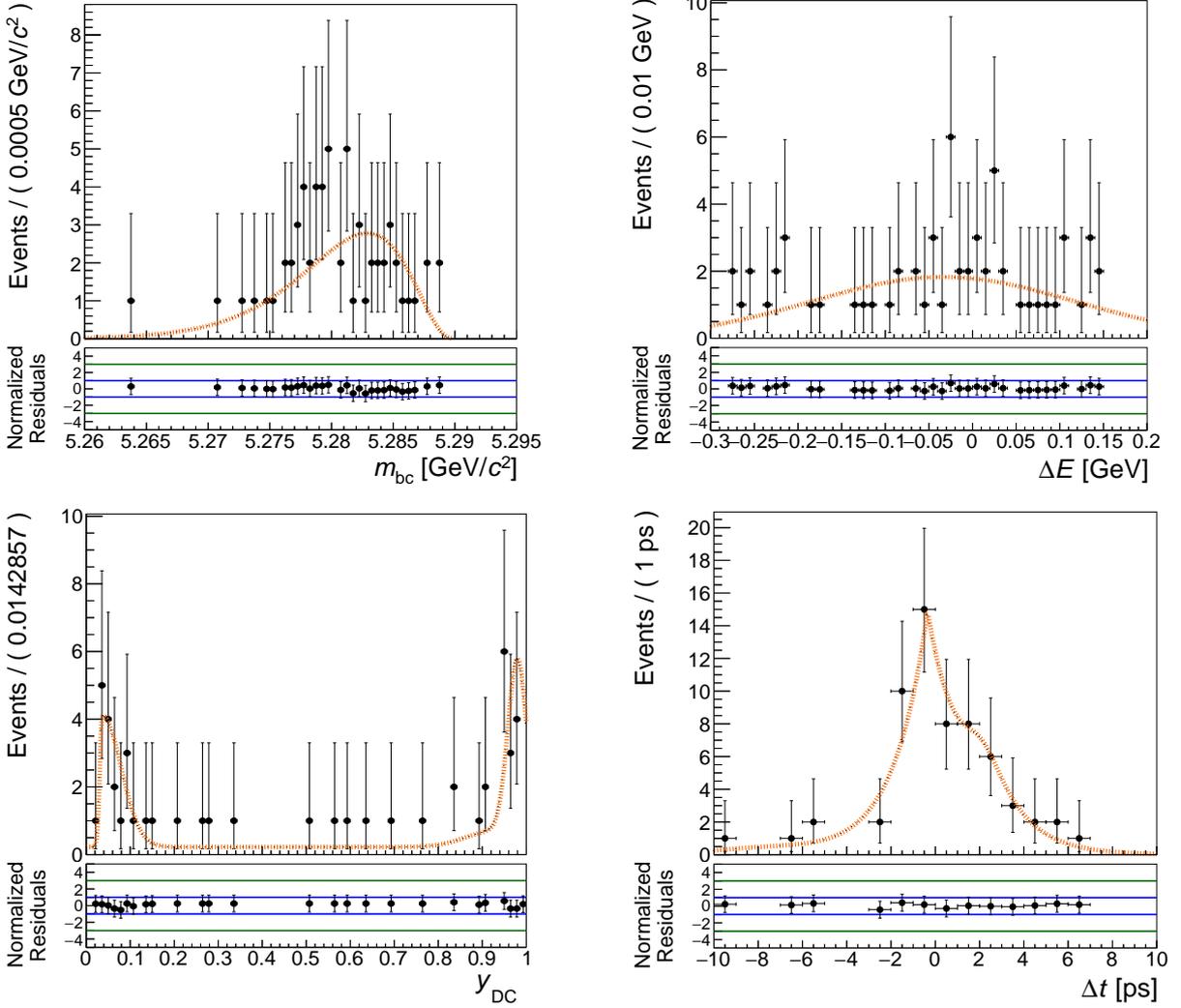


Figure 6.33: Fit projections for wrongly reconstructed signal $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0\pi_{\gamma\gamma}^0$ candidates (WRS). The fit projections are shown onto m_{bc} (top left), ΔE (top right), y_{DC} (bottom left), and Δt (bottom right). The points (with error bars) represent the MC sample used to find the PDFs.

Δt PDF is given by

$$\mathcal{P}_{B\bar{B}}^{\text{final}}(\Delta t, q) \equiv \frac{e^{-\frac{|\Delta t|}{\tau_{\text{eff}}}}}{4\tau_{\text{eff}}} \left[1 - q \cdot \Delta w^{B\bar{B}} \right] \otimes \mathcal{R}_{B\bar{B}}(\delta t),$$

where the effective lifetime τ_{eff} and the parameter $\Delta w^{B\bar{B}}$ were determined by fitting the Δt distribution on the MC events. Figure 6.34 shows the respective fit projections for m_{bc} , ΔE , y_{DC} and Δt .

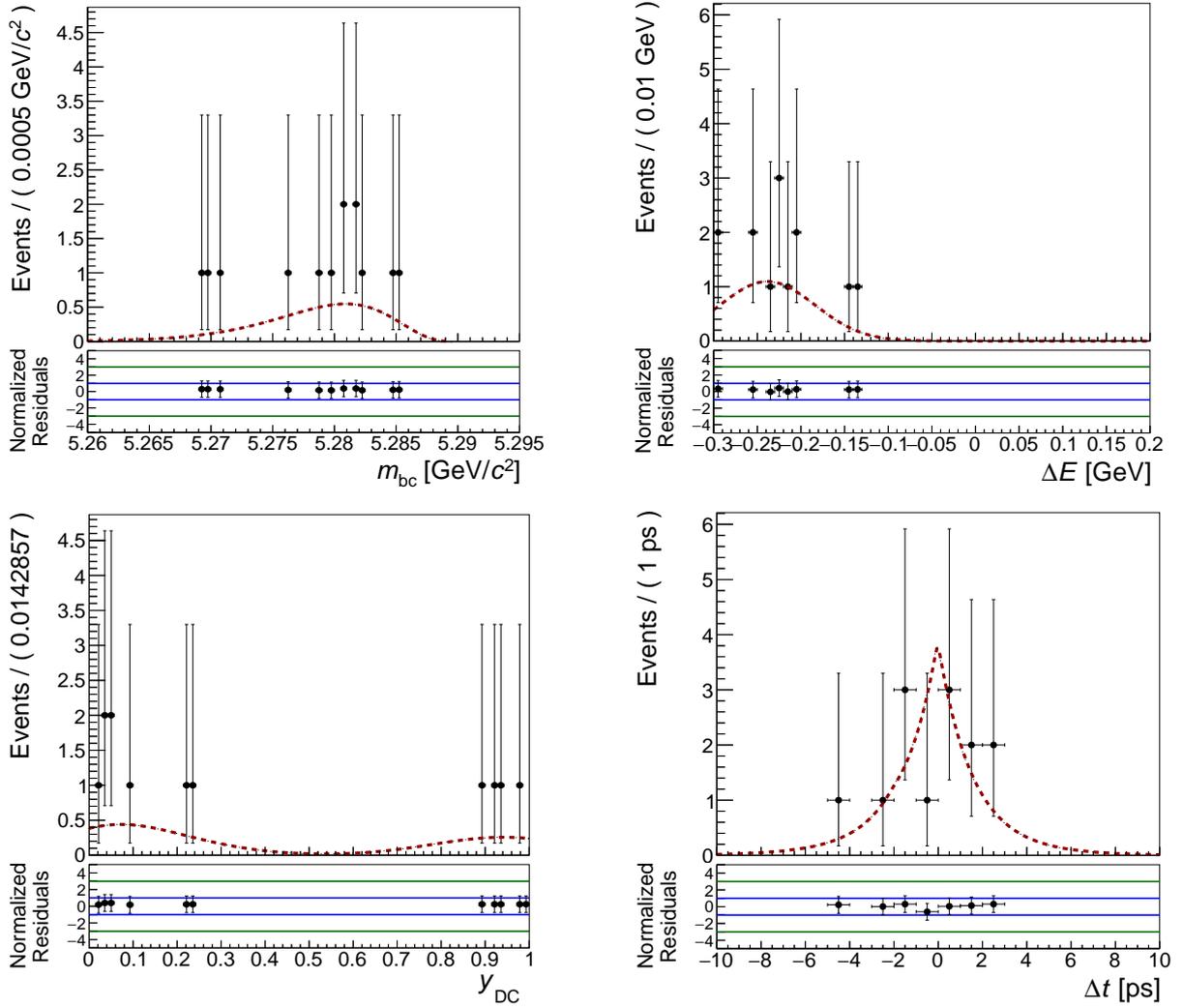


Figure 6.34: Fit projections for background $B\bar{B}$ candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. The fit projections are shown onto m_{bc} (top left), ΔE (top right), y_{DC} (bottom left), and Δt (bottom right). The points (with error bars) represent the MC sample used to find the PDFs.

Continuum events

As in the four-photon model, continuum events are the major background source. In the Dalitz-Conversion model, continuum events are modeled without $\tau^+\tau^-$ events, because none of them passed the final selection in the available 1.8 ab^{-1} MC sample.

The m_{bc} PDF for continuum events, $\mathcal{P}_{\text{cont}}(m_{bc})$, is an ARGUS function. The ΔE PDF, $\mathcal{P}_{\text{cont}}(\Delta E)$, is a first order polynomial defined as in eq. (6.9). The y_{DC} PDF for continuum events is the sum of two Gaussian functions and a first order polynomial,

$$\mathcal{P}_{\text{cont}}(y_{\text{DC}}) \equiv f_1 \cdot G(y_{\text{DC}}; \mu_1, \sigma_1) + f_2 \cdot G(y_{\text{DC}}; \mu_2, \sigma_2)$$

$$+ (1 - f_1 - f_2) \cdot (1 + a \cdot y_{\text{DC}}).$$

And the final Δt PDF is given by

$$\mathcal{P}_{\text{cont}}^{\text{final}}(\Delta t) \equiv \delta(\Delta t) \otimes \mathcal{R}_{\text{cont}}(\delta t).$$

Since continuum events have no CP asymmetry, their Δt PDF does not depend on q . The fit projections for m_{bc} , ΔE , y_{DC} and Δt on continuum events are shown in Fig. 6.35.

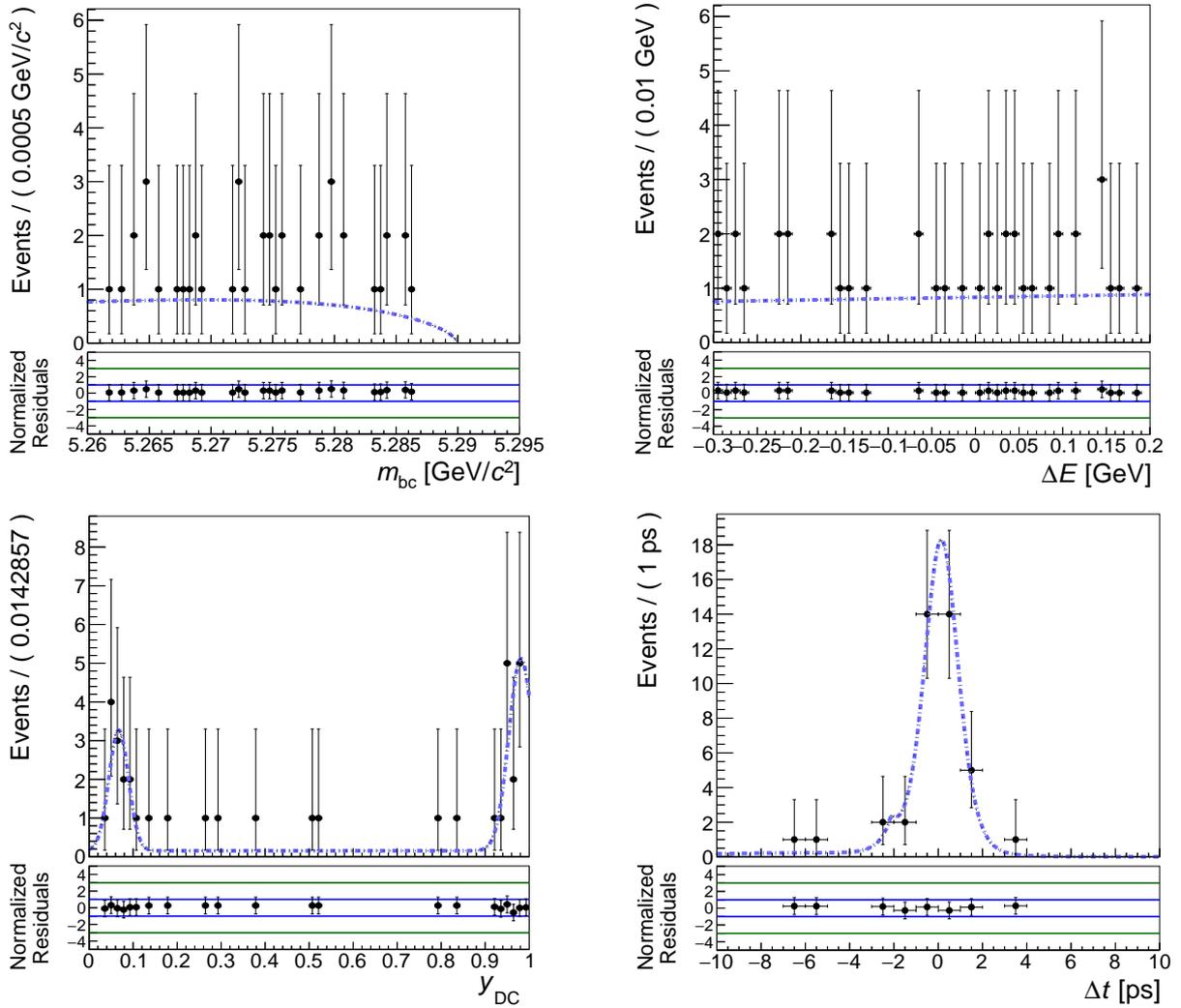


Figure 6.35: Fit projections for continuum $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates. The fit projections are shown onto m_{bc} (top left), ΔE (top right), y_{DC} (bottom left), and Δt (bottom right). The points (with error bars) represent the MC sample used to find the PDFs.

6.10.5 Fit projections

The expected statistical uncertainties on $\mathcal{B}^{\pi^0\pi^0}$, $\mathcal{A}^{\pi^0\pi^0}$, and $\mathcal{S}^{\pi^0\pi^0}$ at Belle II are estimated by performing pseudo-experiments on simulated MC events. For each pseudo-experiment, a sample corresponding to the Belle II integrated luminosity of 50 ab^{-1} is generated. Each sample contains events from all considered components in the same proportion as expected in real data. The signal and the background from wrongly reconstructed signal events are obtained by extracting random sub-samples from fully simulated signal MC events. Due to the small size of the respective MC samples, the $B\bar{B}$ and the continuum events are generated from the PDFs modeled using the official MC samples.

For events with $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$ candidates, a time-integrated CP -analysis is performed using the four-photon model. The fit projections for one of the pseudo-experiments are shown in Fig. 6.36. The fit projections are shown onto m_{bc} and ΔE for the four components (signal, WRS, $B\bar{B}$, and continuum events) as well as for the full event model.

For events with $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates, a time-dependent CP -analysis is performed using the Dalitz-Conversion model (both Dalitz and conversion events treated as Dalitz decays, see Sect. 6.9). Figure 6.38 shows the fit projections for one of the pseudo-experiments. The fit projections onto m_{bc} , ΔE , y_{CD} , and Δt are shown for the five fit components and for the full event model. To show the projections onto y_{CD} and Δt , the sample was enhanced in the signal region by requiring the likelihood ratio $\mathcal{P}_{\text{sig}}/(\mathcal{P}_{\text{sig}} + \mathcal{P}_{\text{bkg}})$ to be larger than 0.65 for each event. For events where the tag-side was tagged as B^0 , and for events where it was tagged as \bar{B}^0 , Figure 6.38 shows the fit projections onto Δt separately. The projections are

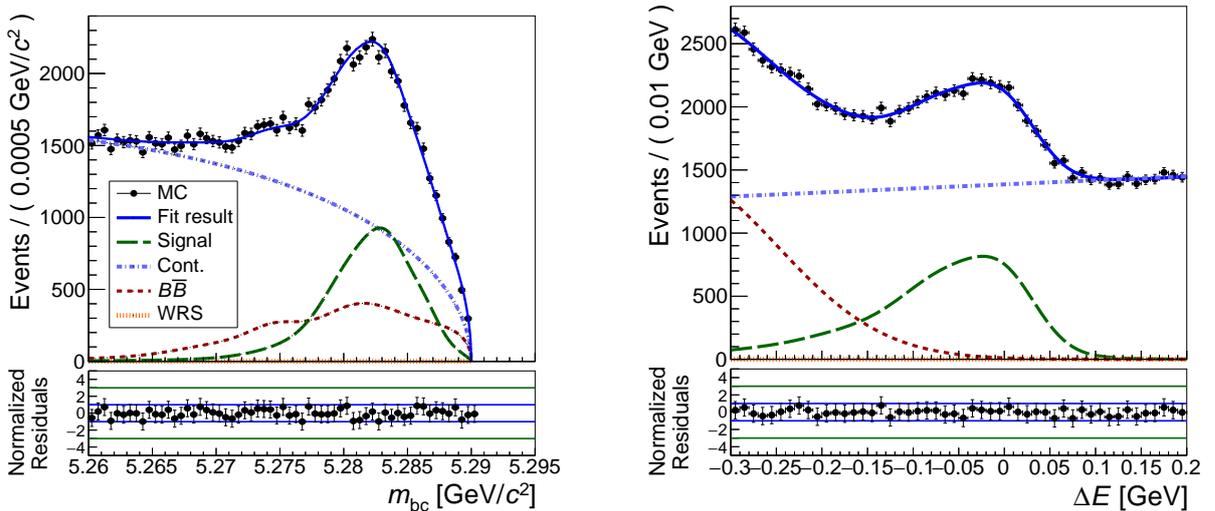


Figure 6.36: Projections of the fit results for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0 \pi_{\gamma\gamma}^0$. The fit projections onto m_{bc} (left) and ΔE (right) are shown for one pseudo-experiment. The points (with error bars) represent the MC sample used for the pseudo-experiment. The full fit results are shown by the solid blue curves.

shown likewise for all the components and for the full event model by enhancing the sample in the signal region.

For the same example pseudo-experiment, Figure 6.39 shows the time-dependent CP asymmetry. The fit projections onto Δt correspond to the results of the full event model for B^0 and for \bar{B}^0 shown in Fig. 6.38. The asymmetry below corresponds to $(N^{B^0} - N^{\bar{B}^0}) / (N^{B^0} + N^{\bar{B}^0})$.

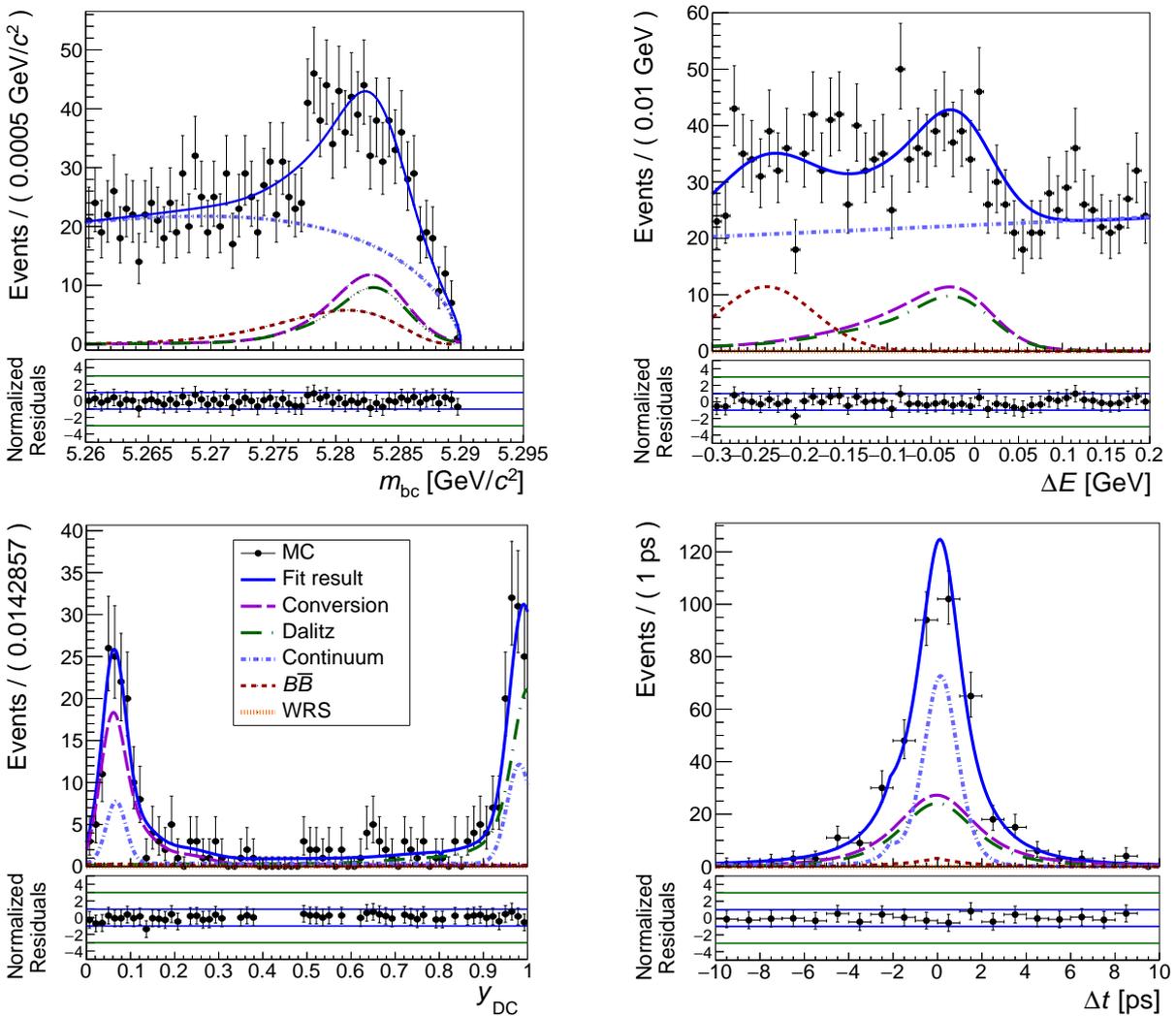


Figure 6.37: Projections of the fit results for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. The fit projections onto m_{bc} (top left), ΔE (top right), y_{DC} (bottom left), and Δt (bottom right) are shown for one pseudo-experiment. The points (with error bars) represent the MC sample used for the pseudo-experiment. For the projections in the bottom, the MC sample was enhanced in the signal region.

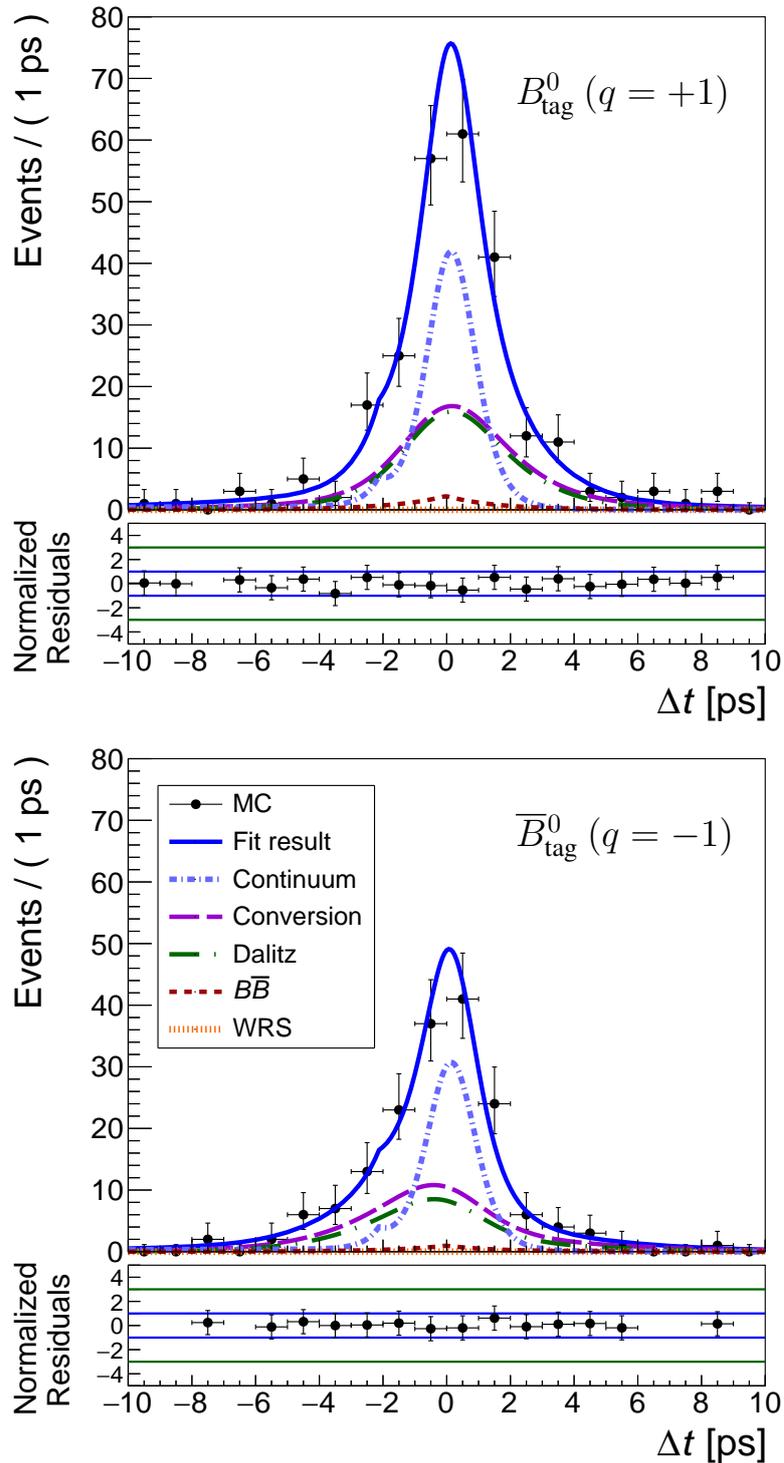


Figure 6.38: Projections of the fit results for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$. The fit projections for one example pseudo-experiment are shown onto Δt . The MC sample was enhanced in the signal region. The projection is shown for events where B_{tag}^0 was tagged as B^0 (top) and as \bar{B}^0 (bottom). The points (with error bars) represent the MC sample used for the pseudo-experiment. The sample (Sample 1) was generated with the parameters $\mathcal{A}_{CP} = 0.34$ and $\mathcal{S}_{CP} = 0.65$. For $q = +1$, the signal distributions (Dalitz and Conv.) have large tails to the right, and for $q = -1$ to the left. The first one is the current world average, and the second one the predicted value by the CKMfitter group.

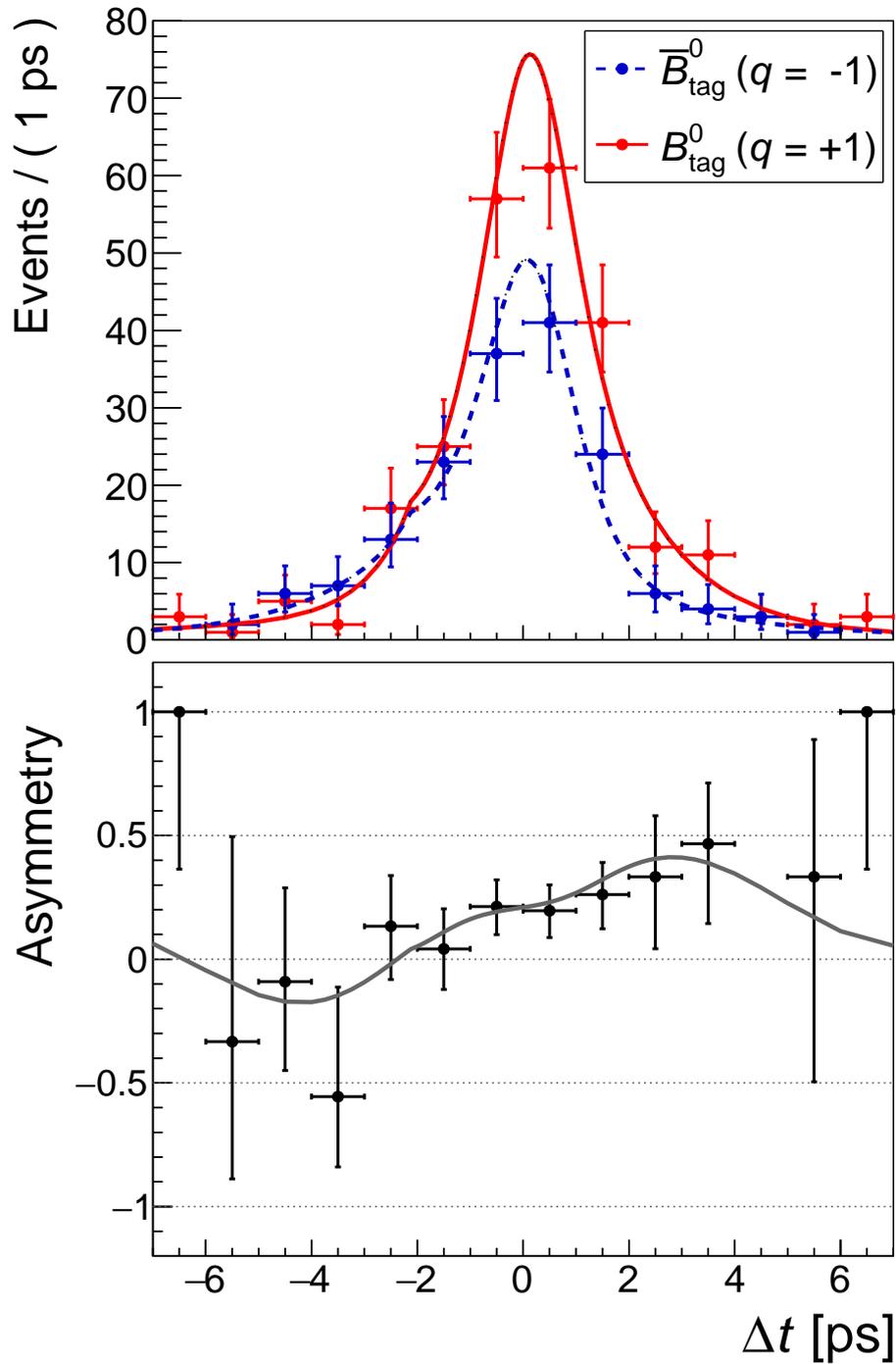


Figure 6.39: Projection of the fit results for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0\pi_{\gamma}^0$ onto Δt . The fit projection is shown for one example pseudo-experiment. The MC sample was enhanced in the signal region. The projection is shown for events where B_{tag}^0 was tagged as B^0 and as \bar{B}^0 . The asymmetry $(N^{B^0} - N^{\bar{B}^0})/(N^{B^0} + N^{\bar{B}^0})$ is shown below. The points (with error bars) represent the MC sample used for the pseudo-experiment. The sample (Sample 1) was generated with the parameters $\mathcal{A}_{CP} = 0.34$ and $\mathcal{S}_{CP} = 0.65$.

6.10.6 Fit validation and results

To verify that the signal yields and the CP -violation parameters are determined without bias and without over- or underestimation of the error, a sufficiently large number of pseudo-experiments (527 in this case) was carried out for each combination of input values $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ (see Sect. 6.2). From the fit results, the pulls and the residuals are calculated. The pull is defined as the error weighted residual,

$$\text{Pull}(x_{\text{output}}) = \frac{x_{\text{output}} - x_{\text{input}}}{\delta x_{\text{output}}},$$

where x_{output} is the fit output, x_{input} is the input value, and δx_{output} is the fit uncertainty. In the case of correctly determined uncertainties, the pull distribution follows a normal distribution with mean equal to 0 and standard deviation equal to 1. A mean differing from 0 corresponds to a bias expressed in units of the statistical uncertainty, while a standard deviation smaller (larger) than 1 corresponds to an over-(under) estimation of the statistical uncertainty.

Figure 6.40 shows the pull and the distribution of the residuals for the signal yield N_{sig} , and for the CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ extracted from time-integrated analyses. The input values for these pseudo-experiments are $N_{\text{sig}} = 15068$, and $\mathcal{A}^{\pi^0\pi^0} = 0.34$. The fits to the pull distributions are performed with a single Gaussian function. For the signal yield, the mean value shows a bias around 13% arriving from the imperfect description of the signal events. Considering the pull standard deviations, there is no large over- or underestimation of the error.

To estimate the statistical errors of the measured parameters, a single Gaussian function is fitted to the distribution of the residuals $x_{\text{output}} - x_{\text{input}}$ and its standard deviation σ is taken as statistical uncertainty. For the branching fraction $\mathcal{B}^{\pi^0\pi^0}$, the statistical uncertainty is estimated using the relation

$$\frac{\delta \mathcal{B}^{\pi^0\pi^0}}{\mathcal{B}^{\pi^0\pi^0}} = \frac{\sigma_{N_{\text{sig}}}}{N_{\text{sig}}}.$$

The measurement of branching fractions will be explained in Sect. 7.3.

Considering now time-dependent analyses, Figure 6.41 shows the pull distributions for the yields of Dalitz and of conversion events, as well as the pull and the distributions of the residuals for $\mathcal{A}^{\pi^0\pi^0}$ and for $\mathcal{S}^{\pi^0\pi^0}$. The input values used for these pseudo-experiments are $N_{\text{Dalitz}} = 147$, $N_{\text{Conversion}} = 124$, $\mathcal{A}_{CP} = 0.34$ and $\mathcal{S}_{CP} = 0.65$. Within the uncertainties, the mean values for all pull distributions agree well with 0, and the respective standard deviations agree well with 1.

Due to the small number of events, the statistical uncertainty of $\mathcal{A}^{\pi^0\pi^0}$ is larger than the one obtained from the four-photon sample. Thus, a time-dependent analysis of the Dalitz-conversion sample is essential only to measure $\mathcal{S}^{\pi^0\pi^0}$.

Table 6.7 presents the final results of this performance study. It contains the estimated values of $\delta \mathcal{B}^{\pi^0\pi^0} / \mathcal{B}^{\pi^0\pi^0}$, $\delta \mathcal{A}^{\pi^0\pi^0}$ and $\delta \mathcal{S}^{\pi^0\pi^0}$ for the different input MC values of $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$. The pull distributions and the residuals for the different MC input values can be found in App. D.1.

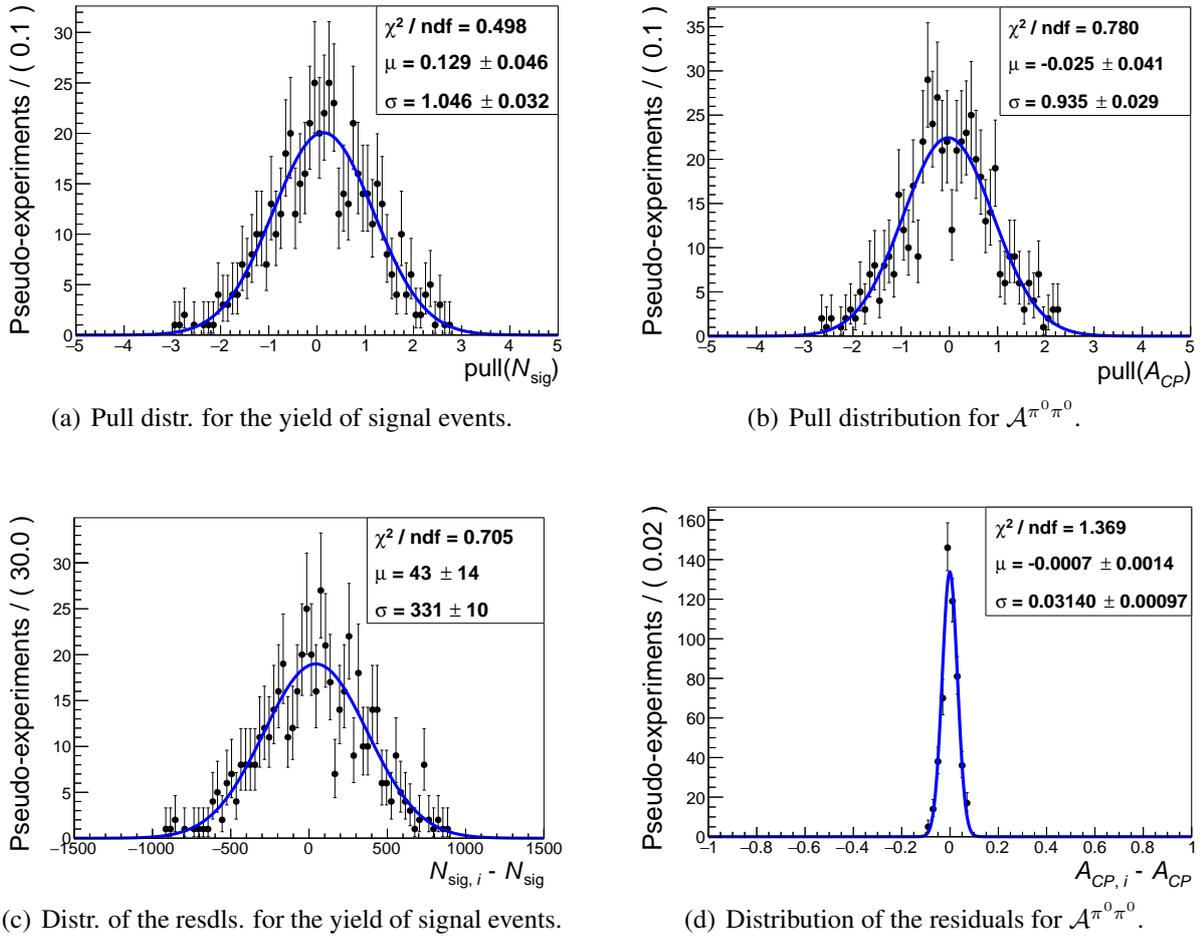
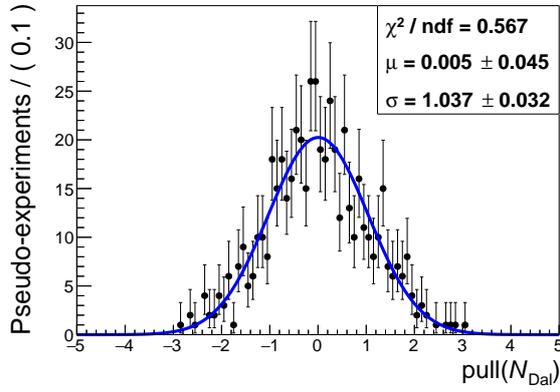
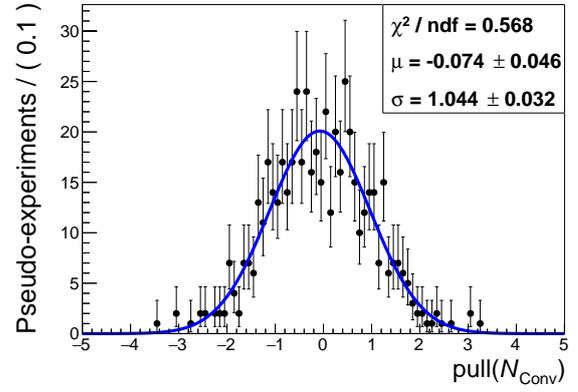


Figure 6.40: Pull distributions and distributions of the residuals of the yield of signal events, and of the CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ for the fit of $B_{\text{sig}}^0 \rightarrow \pi_{\gamma\gamma}^0\pi_{\gamma\gamma}^0$ candidates. The input values used for these pseudo-experiments are $N_{\text{sig}} = 15068$, $\mathcal{A}_{CP} = 0.34$ and $\mathcal{S}_{CP} = 0.65$.

To test the validity of the Δt resolution function and the reconstruction procedure, the B^0 lifetime was extracted from the simulated events performing separate time-dependent analyses. This time, the CP -violation parameters $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ were fixed to the MC input values and the B^0 lifetime τ_{B^0} was released as free fit parameter. The yields for each component were left free as well. Figure 6.42 shows the pull and the distributions of the residuals for the B^0 lifetime. The MC input value corresponds to 1.525 ps. Within the uncertainties, the mean value of the pull distribution is consistent with 0, and the respective standard deviation is consistent with 1.



(a) Pull distr. for the yield of Dalitz events.



(b) Pull distr. for the yield of conversion events.

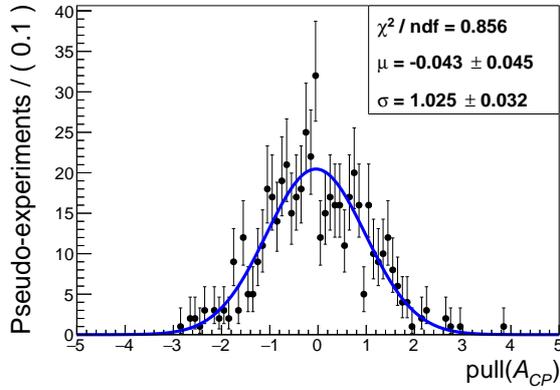
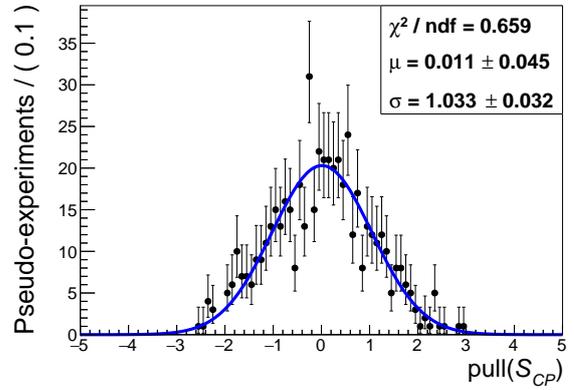
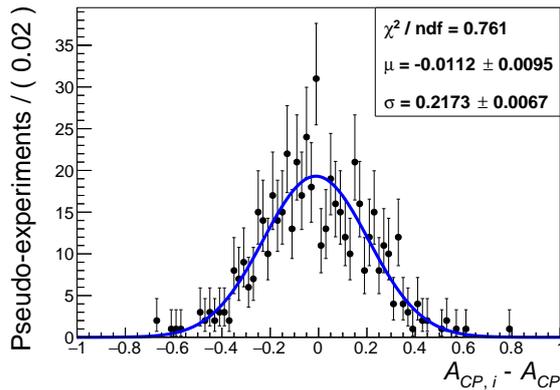
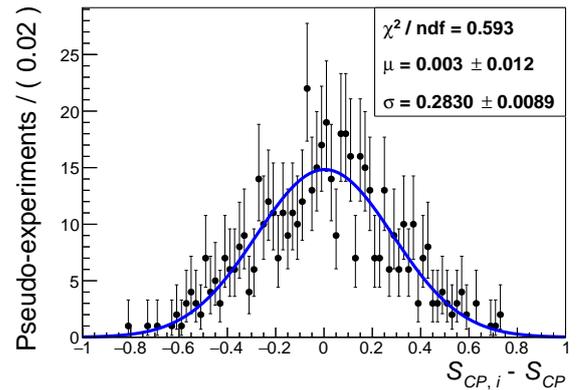
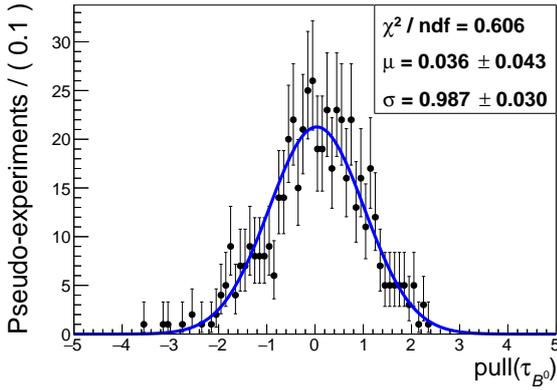
(c) Pull distribution for $\mathcal{A}^{\pi^0\pi^0}$.(d) Pull distribution for $\mathcal{S}^{\pi^0\pi^0}$.(e) Distribution of the residuals for $\mathcal{A}^{\pi^0\pi^0}$.(f) Distribution of the residuals for $\mathcal{S}^{\pi^0\pi^0}$.

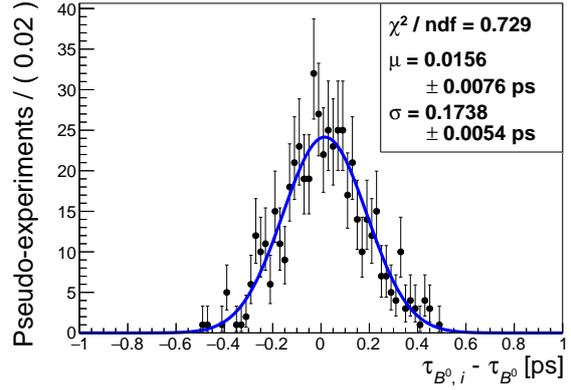
Figure 6.41: Pull distributions, and distribution of the residuals of the CP -violation parameter $\mathcal{S}^{\pi^0\pi^0}$ for the fit of $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ candidates. The input values used for these pseudo-experiments are $N_{\text{Dal}} = 147$, $N_{\text{Conv}} = 124$, $\mathcal{A}_{CP} = 0.34$ and $\mathcal{S}_{CP} = 0.65$.

Table 6.7: Statistical uncertainties $\delta\mathcal{A}^{\pi^0\pi^0}$, $\delta\mathcal{S}^{\pi^0\pi^0}$ and $\delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0}$ expected for the full Belle II data sample using the different input values of $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ employed to generate the signal MC events (see Sect. 6.2).

Input values		Time-integrated	Time-dependent		
$\mathcal{A}^{\pi^0\pi^0}$	$\mathcal{S}^{\pi^0\pi^0}$	$\delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0}$ [%]	$\delta\mathcal{A}^{\pi^0\pi^0}$	$\delta\mathcal{A}^{\pi^0\pi^0}$	$\delta\mathcal{S}^{\pi^0\pi^0}$
0.34 [71]	0.65 [71]	2.2	0.03	0.28	0.22
0.43 [28]	0.79	2.2	0.03	0.29	0.23
0.14 [140]	0.83	2.4	0.03	0.26	0.21
0.14 [140]	0.40	2.3	0.03	0.29	0.20
0.14 [140]	-0.61	2.3	0.03	0.27	0.22
0.14 [140]	-0.94	2.4	0.03	0.28	0.22



(a) Pull distribution for τ_{B^0} .



(b) Distribution of the residuals for τ_{B^0} .

Figure 6.42: Pull and distributions of the residuals of the B^0 lifetime for the fit of $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0\pi_{\gamma\gamma}^0$ candidates. The MC input value is $\tau_{B^0} = 1.525$ ps.

6.11 Comparison with Belle

The latest time-integrated measurement of $\mathcal{B}^{\pi^0\pi^0}$ and $\mathcal{A}^{\pi^0\pi^0}$ at Belle will be discussed in depth in Sect. 7.4.1 and 7.4.1. Considering only statistical uncertainties, the ratio $\delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0}$ was about 15%, and $\delta\mathcal{A}^{\pi^0\pi^0}$ was 0.36. These values are expected to decrease inversely proportional to the square root of the sample size and can be thus scaled using the relation

$$\sqrt{\frac{\mathcal{L}_{\text{Belle}}}{\mathcal{L}_{\text{Belle II}}}} \simeq \sqrt{\frac{0.8 \text{ ab}^{-1}}{50 \text{ ab}^{-1}}} \approx 0.13. \quad (6.14)$$

For the full Belle II data sample, this leads to an expected uncertainty $\delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0} = 2.0\%$, which agrees with the value of 2.2% estimated in this thesis (see Tab. 6.7), and to an expected uncertainty $\delta\mathcal{A}^{\pi^0\pi^0} = 0.05$, which is close to the estimated uncertainty of 0.03.

6.12 Conclusion

The goal of this chapter was to show that Belle II will be able to measure the mixing-induced CP -violation parameter $\mathcal{S}^{\pi^0\pi^0}$ after having collected its full data sample. The results show that, in fact, the measurement is feasible for the expected sample size corresponding to 50 ab^{-1} . The estimated statistical uncertainty is

$$\delta\mathcal{S}^{\pi^0\pi^0} = 0.28.$$

A better reconstruction of electrons and of converted photons will potentially improve the Δt resolution, which is already good enough for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\text{Dal}}^0 \pi_{\gamma\gamma}^0$ decays, but very poor for candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\gamma C}^0 \gamma$ decays. Candidates reconstructed as $B_{\text{sig}}^0 \rightarrow \pi_{\gamma C}^0 \gamma$ decays were therefore excluded from the analysis. Apart from that, when this performance study was carried out, the flavor tagger was not fully optimized. And due to the small amount of MC events available to model the background distributions, the analysis had to be performed in a single bin of the dilution factor r , without exploiting the full flavor tagging power. All these arguments show that an improvement in the $\mathcal{S}^{\pi^0\pi^0}$ precision is possible. However, the Belle II performance was studied using MC events which were generated without beam-induced and without luminosity-dependent backgrounds. The reason is that the track reconstruction algorithms were not optimized for events with additional background when this study was carried out. At Belle II, the background levels are much higher than at Belle. Therefore, the improvements sketched above will be needed to keep the precision estimated in this chapter.

To validate the procedure for $\mathcal{S}^{\pi^0\pi^0}$, a time-integrated analysis was performed to estimate the Belle II precision for the direct CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ and the branching fraction $\mathcal{B}^{\pi^0\pi^0}$. The estimated statistical uncertainties for the full Belle II integrated luminosity yield

$$\delta\mathcal{A}^{\pi^0\pi^0} = 0.03, \quad \text{and} \quad \delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0} = 2.2\%,$$

proving the consistency with the previous Belle measurements for which the extrapolated values are $\delta\mathcal{A}^{\pi^0\pi^0} = 0.05$, and $\delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0} = 2.0\%$.

The analysis presented in this chapter completed all the steps that have to be done on MC events to perform a blind analysis. In the absence of control samples from Belle II data, differences between MC events and real data could not be studied to estimate the systematic uncertainties. For the measurement of $\mathcal{S}^{\pi^0\pi^0}$, however, the total systematic uncertainty is expected to be much smaller than the statistical one due to the expected large sizes of control samples at Belle II. The systematic uncertainties affecting the measurements of $\mathcal{S}^{\pi^0\pi^0}$, $\mathcal{A}^{\pi^0\pi^0}$, and $\mathcal{B}^{\pi^0\pi^0}$ will be estimated in the next chapter by projecting the systematic uncertainties at Belle.

7 The unitarity angle ϕ_2 at Belle II

7.1 Introduction

One of the key quantities in the examination of the CKM mechanism is the unitary angle ϕ_2 . The major goal of this chapter is to estimate how precise Belle II will be able to determine it.

One of the limitations in the determination of ϕ_2 has been the absence of the input observable $\mathcal{S}^{\pi^0\pi^0}$. As shown in the previous chapter, a measurement of $\mathcal{S}^{\pi^0\pi^0}$ will be feasible when Belle II will collect its expected integrated luminosity of about 50 ab^{-1} . One of the objectives of this chapter, is to study the impact of this novel measurement.

The studies and the estimations are based on the isospin analyses of the decay channels $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ introduced in Sect. 2.10.2. The $B \rightarrow \pi\pi$ and the $B \rightarrow \rho\rho$ channels are currently the ones dominating the precision for ϕ_2 [71]. After a brief introduction of the analysis algorithm, the latest measurements of the input observables are reviewed. Next, the systematic uncertainties expected at Belle II are discussed in detail. The Belle II precision is estimated by projecting the statistical and the systematic uncertainties affecting the measurements of the input observables. As major reference, the available Belle measurements are taken. Using these measurements and the projections for Belle II, the isospin analyses are performed, resulting in the expected uncertainty for the unitarity angle ϕ_2 .

7.2 The isospin analysis fit

The unitarity angle ϕ_2 can be determined from the CP -violation parameters associated with the decay channels $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$. Here, non-negligible penguin contributions (penguin pollution) to these decays prevent a direct determination of ϕ_2 . The goal of the isospin analysis is to determine the pollution-corrected value of ϕ_2 using isospin relations among the involved decay amplitudes. For $B \rightarrow \pi\pi$ and for $B \rightarrow \rho\rho$ decays, the isospin relations form so-called isospin triangles (see Sect. 2.10.2). To perform the analysis, the decay amplitudes are parametrized in such a way that they fulfill these triangle conditions.

The algorithm employed for the fit procedure involved in the isospin analysis is based on the approach followed by Belle [87, 140], by BaBar [89, 151] and by the CKMfitter group [71, 152]. The algorithm assumes the parametrization given by eq. (2.61) in Sect. 2.10.2, which is similar to the one proposed by Muriel Pivk and François Le Diberder [85]. This parametrization was slightly modified for the use in this work in order to simplify the calculations. In the

algorithm, the absolute values of the decay amplitudes are given by

$$\begin{aligned}
|A^{+0}| &= |\tilde{A}^{+0}| = \sqrt{a^{+0}}, \\
|A^{+-}| &= \sqrt{a^{+-} (1 - \mathcal{A}_{CP})}, \\
|\tilde{A}^{+-}| &= \sqrt{a^{+-} (1 + \mathcal{A}_{CP})}, \\
|A^{00}| &= \left(\frac{|A^{+0}|^2}{2} + |A^{+0}|^2 - \sqrt{2}|A^{+0}||A^{+0}| \cos(\phi_2 - \delta) \right)^{\frac{1}{2}}, \\
|\tilde{A}^{00}| &= \left(\frac{|\tilde{A}^{+0}|^2}{2} + |A^{+0}|^2 - \sqrt{2}|\tilde{A}^{+0}||A^{+0}| \cos(\phi_2 + \delta - 2\phi_{2,\text{eff}}) \right)^{\frac{1}{2}},
\end{aligned}$$

where a^{+0} , a^{+-} , \mathcal{A}_{CP} , δ , $\phi_{2,\text{eff}}$ and ϕ_2 are real positive variables, and $\tilde{A}^{ij} = e^{-2i\phi_1} \bar{A}^{ij}$. Here, $\phi_{2,\text{eff}}$ is not corrected for penguin pollution bias. The amplitudes are used to calculate the theoretical predictions for the physical observables,

$$\begin{aligned}
\mathcal{B}_{\text{theo}}^{+0} &= \frac{\tau^{B^+}}{\tau^{B^0}} |A^{+0}|^2, & \mathcal{B}_{\text{theo}}^{+-} &= \frac{|A^{+-}|^2 + |\tilde{A}^{+-}|^2}{2}, \\
\mathcal{B}_{\text{theo}}^{00} &= \frac{|A^{00}|^2 + |\tilde{A}^{00}|^2}{2}, & \mathcal{A}_{\text{theo}}^{+-} &= \mathcal{A}_{CP}, \\
\mathcal{S}_{\text{theo}}^{+-} &= \sqrt{1 - \mathcal{A}_{CP}^2} \sin(2\phi_{2,\text{eff}}), & \mathcal{A}_{\text{theo}}^{00} &= \frac{|\tilde{A}^{00}| - |A^{00}|}{|\tilde{A}^{00}| + |A^{00}|},
\end{aligned}$$

and

$$\mathcal{S}_{\text{theo}}^{00} = \frac{2|A^{+0}|^2 \sin(2\phi_2) + |A^{+-}| |\tilde{A}^{+-}| \sin(2\phi_{2,\text{eff}}) - \sqrt{2}|A^{+0}| (|\tilde{A}^{+-}| \bar{s} + |A^{+-}| s)}{|\tilde{A}^{00}| + |A^{00}|}, \quad (7.1)$$

where $s = \sin(\phi_2 + \delta)$ and $\bar{s} = \sin(\phi_2 - \delta + 2\phi_{2,\text{eff}})$. For the determination of ϕ_2 , a scan of the confidence is performed by translating a χ^2 distribution into a confidence level (CL). The χ^2 -distribution is obtained by minimizing $\chi^2 \equiv -2 \log(\mathcal{L})$ at each value of ϕ_2 . The likelihood \mathcal{L} has the form of a multivariate normal distribution

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{\exp\left(\frac{1}{2} (\mathbf{x}_{\text{data}} - \mathbf{x}_{\text{theo}}(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{\text{data}} - \mathbf{x}_{\text{theo}}(\boldsymbol{\theta}))\right)}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}},$$

which quantifies the agreement between the measured values \mathbf{x}_{data} and the theoretical predictions \mathbf{x}_{theo} of the input observables,

$$\mathbf{x} = (\mathcal{B}^{+-}, \mathcal{B}^{00}, \mathcal{B}^{+0}, A^{+-}, \mathcal{S}^{+-}, A^{00}, \mathcal{S}^{00})^T,$$

where T stays for transposed. If no measurement is available for \mathcal{S}^{00} , a χ^2 minimization is also possible. The parameters

$$\boldsymbol{\theta} = \{a^{+0}, a^{+-}, \mathcal{A}_{CP}, \phi_2, \delta, \phi_{2,\text{eff}}\}, \quad (7.2)$$

correspond to the six unknown parameters of the analysis. The covariance matrix Σ contains the uncertainties $\delta x_i = \sqrt{\delta x_{i,\text{stat}}^2 + \delta x_{i,\text{sys}}^2}$ and the measured correlations c_{ij} between the observables in the non-diagonal part. The respective coefficients are

$$\Sigma_{ij} = c_{ij} \cdot \delta x_i \delta x_j, \quad c_{ij} = 1 \text{ for } i = j. \quad (7.3)$$

The scan of the confidence for ϕ_2 is performed in steps of 0.5° in the range $[0, 180^\circ]$. At each step, the value of ϕ_2 is fixed and $\chi^2(\boldsymbol{\theta}) = -2 \log(\mathcal{L}(\boldsymbol{\theta}))$ is minimized. The CL is obtained from the χ^2 value using incomplete Γ functions [152]

$$\text{CL} = \text{Prob}(\chi^2(\boldsymbol{\theta}), N_{\text{dof}}) = \frac{1}{\sqrt{2^{N_{\text{dof}}} \Gamma(N_{\text{dof}}/2)}} \int_{\chi^2(\boldsymbol{\theta})}^{\infty} e^{-\frac{t}{2}} \cdot t^{\frac{N_{\text{dof}}}{2}-1} dt,$$

where $\chi^2(\boldsymbol{\theta})$ is the minimized χ^2 value, and N_{dof} is the number of degrees of freedom, which in this case is equal to 1, because ϕ_2 remains unmodified during the χ^2 minimization.

7.3 Extrapolation of uncertainties

To estimate the future Belle II precision for the measurement of ϕ_2 , the uncertainties of the input observables entering the isospin analyses have to correspond to the full integrated luminosity of 50 ab^{-1} . A way to perform this is to extrapolate the uncertainties of previous Belle measurements. This approach is motivated by the similarities between Belle and Belle II.

Statistical uncertainties are expected to decrease inversely proportional to the square root of the sample size, and thus to scale with the integrated luminosity as in eq. (6.14). In the following sections, the statistical uncertainties of all input observables are scaled in this way, except for $\mathcal{B}^{\pi^0\pi^0}$, $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$, where the statistical uncertainties were estimated performing pseudo-experiments corresponding to the full Belle II integrated luminosity (see Sect. 6.10.6).

Systematic uncertainties can be classified into two types: *reducible* and *irreducible*. Reducible systematic uncertainties are evaluated by comparing control samples from real data with MC events from the corresponding channels. Therefore, they are expected to scale with the integrated luminosity, as the statistical ones. Examples of systematic effects causing reducible uncertainties are: detection and tracking efficiencies, detector alignment, PID calibration, fractions of mis-reconstructed events, wrong-tag fractions (flavor tagging), histogram shapes, model shapes, etc. Irreducible uncertainties depend on underlying physical processes, on the detector resolution, on the algorithms used for the data analysis, or on the algorithms used for the MC generation. They usually do not scale with the integrated luminosity.

Two systematic uncertainties are considered to be irreducible for the measurement of branching fractions: the imperfect simulation of final state radiation, which plays a role for the measurement of $\mathcal{B}^{\pi^+\pi^-}$ (see Sect. 7.4.1), and the number $N_{B\bar{B}}^{\text{prod}}$ of produced $B\bar{B}$ pairs¹. The branching fraction $\mathcal{B}(B \rightarrow f_{\text{sig}})$ for a certain final state f_{sig} is measured as

$$\mathcal{B}(B \rightarrow f_{\text{sig}}) = \frac{N_{\text{sig}}}{\epsilon_{\text{sig}} N_{B\bar{B}}^{\text{prod}}},$$

where N_{sig} is the signal yield measured on data and ϵ_{sig} is the efficiency for f_{sig} determined on signal MC events. The uncertainty on the simulation of final state radiation affects the determination of ϵ_{sig} for certain decay channels, and is caused by the uncertainties inherent to the simulation algorithm, which at Belle and at Belle II is the widely used algorithm PHOTOS [153].

Considering now $N_{B\bar{B}}^{\text{prod}}$, at Belle, the number $N_{B\bar{B}}^{\text{prod}}$ in the hadronic skim² was given by

$$N_{B\bar{B}}^{\text{prod}} = N_{\text{on-resonance}} - a \cdot c \cdot N_{\text{off-resonance}},$$

where $N_{\text{on-resonance}}$ is the number of recorded events at the $\Upsilon(4S)$ ($\sqrt{s} = 10.58 \text{ GeV}$) and $N_{\text{off-resonance}}$ the number at 80 MeV below the resonance. The luminosity scale factor a is measured using the number of recorded Bhabha $N^{e^+e^-}$ and di-muon events $N^{\mu^+\mu^-}$ on and off-resonance before the hadronic skim

$$a \equiv \frac{1}{2} \left(\frac{N_{\text{on-res}}^{e^+e^-}}{N_{\text{off-res}}^{e^+e^-}} + \frac{N_{\text{on-res}}^{\mu^+\mu^-}}{N_{\text{off-res}}^{\mu^+\mu^-}} \right).$$

The parameter c corresponds to the ratio

$$c \equiv \frac{\epsilon_{\text{on-res}}^{q\bar{q}}}{\epsilon_{\text{off-res}}^{q\bar{q}}},$$

where $\epsilon_{\text{on-res}}^{q\bar{q}}$ and $\epsilon_{\text{off-res}}^{q\bar{q}}$ are the skim efficiencies on and off-resonance determined on continuum MC. The continuum MC has imperfections and, at present, it is not clear how to reduce them. Since the value of c relies fully on continuum MC, its uncertainty does not scale with the size of the samples and limits the precision for $N_{B\bar{B}}^{\text{prod}}$. Therefore, the uncertainty of $N_{B\bar{B}}^{\text{prod}}$ is assumed to be irreducible. At Belle II, the optimal method to measure $N_{B\bar{B}}^{\text{prod}}$ is not established yet and further studies are required.

For the measurement of CP -violation parameters, two systematic uncertainties do not scale with the integrated luminosity: the one associated with the Δt -resolution function, and the one associated with the tag-side interference effect (see Sect. 4.10). The first one is related

¹This effect is also referred to as ‘‘Luminosity’’ in the literature.

²A detailed description of the measurement of $N_{B\bar{B}}^{\text{prod}}$ can be found in [154].

to the uncertainties on the reconstructed B -vertex position and on the B -vertex resolution. These uncertainties are related to the alignment of the vertex detector, the vertex reconstruction algorithms and the vertex resolution. The alignment of the vertex detector will be performed using control samples from the data, such as di-muon or two-body decays, and is thus expected to improve at high luminosity. Due to the new nano-beam scheme (see Sect. 3.1), Belle II will have also a more precise knowledge of the position and the size of the interaction region. Additionally, the new Belle II vertex algorithm for the tag side reaches an improvement in resolution by almost a factor two [155]. The signal-side B^0 -vertex resolution is expected to improve also by a factor two as the impact parameter resolution (see Sect. 3.3 and 3.4), thanks to the new pixel detector (PXD). Therefore, the systematic uncertainties associated with the B -vertex are assumed to decrease by a factor two [1], and, in a similar way, the uncertainty due to the Δt -resolution function is assumed to be partially reducible and to decrease by a factor two. The second effect, the one associated with the tag-side interference, is assumed to be fully irreducible since a way to mitigate it is not established yet (see discussion in Sect. 4.10).

7.4 The $B \rightarrow \pi\pi$ system

In this section, the Belle II precision for ϕ_2 is estimated considering the isospin analysis of the $B \rightarrow \pi\pi$ system. After a brief discussion of the current experimental status of the input observables, the projections of the input observables are explained. Using these projections, the isospin analysis is performed without and with the $\mathcal{S}^{\pi^0\pi^0}$ constraint.

Table 7.1 presents the latest measurements of the observables entering the isospin analysis of the $B \rightarrow \pi\pi$ system. The measurements were performed at Belle, at BaBar and at LHCb. Belle and BaBar used data samples collected at the $\Upsilon(4S)$ corresponding respectively to 0.8 ab^{-1} and to 0.5 ab^{-1} . They succeeded in measuring all the observables involved in the isospin analysis, except $\mathcal{S}^{\pi^0\pi^0}$.

LHCb has covered only those decays that have only charged pions in the final states. The value of $\mathcal{B}^{\pi^+\pi^-}$ was measured using a subsample from the Run 1 data set and the values of $\mathcal{A}^{\pi^+\pi^-}$ and $\mathcal{S}^{\pi^+\pi^-}$ using the full Run 1 data set (see Sect. 2.8). Since LHCb can measure only ratios between two branching fractions, the absolute branching fraction $\mathcal{B}^{\pi^+\pi^-}$ was obtained by measuring the ratio

$$\mathcal{B}^{\pi^+\pi^-} / \mathcal{B}^{K^+\pi^-} = 0.262 \pm 0.009 \pm 0.017 [156],$$

and multiplying it with the current world average for $\mathcal{B}^{K^+\pi^-}$. The measurements of $\mathcal{A}^{\pi^+\pi^-}$ and $\mathcal{S}^{\pi^+\pi^-}$ are very precise and they are still limited by the size of the analyzed samples. Updating these last two measurements by using the full Run 1 and Run 2 data samples, LHCb could reach a precision level comparable with that expected at Belle II. However, LHCb is not competitive for decays involving neutral pions because of the harsh hadronic environment. Therefore, a considerable improvement in the precision for ϕ_2 requires new measurements at Belle II.

Table 7.1: Latest measurements of the branching fractions and CP -violation parameters entering in the isospin analysis of the $B \rightarrow \pi\pi$ system, performed by the Belle, the BaBar and the LHCb experiments. The first uncertainty is the statistical one and the second is the systematic one.

	Belle	BaBar	LHCb
$\mathcal{B}^{\pi^+\pi^-}_{[10^{-6}]}$	$5.04 \pm 0.21 \pm 0.18$ [157]	$5.5 \pm 0.4 \pm 0.3$ [158]	$5.08 \pm 0.17 \pm 0.37$ [156]
$\mathcal{B}^{\pi^+\pi^0}_{[10^{-6}]}$	$5.86 \pm 0.26 \pm 0.38$ [157]	$5.02 \pm 0.46 \pm 0.29$ [159]	
$\mathcal{B}^{\pi^0\pi^0}_{[10^{-6}]}$	$1.31 \pm 0.19 \pm 0.19$ [140]	$1.83 \pm 0.21 \pm 0.13$ [151]	
$\mathcal{A}^{\pi^0\pi^0}$	$0.14 \pm 0.36 \pm 0.10$ [140]	$0.43 \pm 0.26 \pm 0.05$ [151]	
$\mathcal{A}^{\pi^+\pi^-}$	$0.33 \pm 0.06 \pm 0.03$ [160]	$0.25 \pm 0.08 \pm 0.02$ [151]	$0.24 \pm 0.07 \pm 0.01$ [161]
$\mathcal{S}^{\pi^+\pi^-}$	$-0.64 \pm 0.08 \pm 0.03$ [160]	$-0.68 \pm 0.10 \pm 0.03$ [151]	$-0.68 \pm 0.06 \pm 0.01$ [161]

7.4.1 Expected precision at Belle II

The Belle measurements presented in Tab. 7.1 are extrapolated to the full Belle II luminosity following the guidelines in Sect. 7.3. The statistical uncertainties of $\mathcal{B}^{\pi^+\pi^-}$, $\mathcal{B}^{\pi^+\pi^0}$, $\mathcal{A}^{\pi^+\pi^-}$ and $\mathcal{S}^{\pi^+\pi^-}$ are extrapolated according to eq. (6.14). For $\mathcal{B}^{\pi^0\pi^0}$, $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$, the statistical uncertainties are taken from the performance study presented in Sect. 6.10.6; the values of these uncertainties correspond to the results of the pseudo-experiments where the current world averages for $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ were taken as input (first row in Tab. 6.7).

For each input observable, the individual systematic uncertainties are summed in quadrature. Reducible systematic uncertainties are extrapolated in the same way as the statistical uncertainties. In the following, the extrapolations of the systematic uncertainties affecting the Belle II measurements are discussed.

Table 7.3 [157] shows the effects for $\mathcal{B}^{\pi^+\pi^-}$ and $\mathcal{B}^{\pi^+\pi^0}$. At high luminosity, the large data samples allow one to improve the modeling of fit components and, therefore, the fitting uncertainties are expected to be reducible. In Table 7.3, the only irreducible sources are the number of B mesons, $N_{B\bar{B}}^{\text{prod}}$, and the contribution from the PHOTOS simulation algorithm. Improvements in the determination of $N_{B\bar{B}}^{\text{prod}}$ and in the PHOTOS algorithm cannot be predicted. The associated uncertainties are therefore not scaled.

Table 7.4 [140] shows the systematic effects for $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{B}^{\pi^0\pi^0}$. Considering first $\mathcal{A}^{\pi^0\pi^0}$, the fitting uncertainties due to the parametrization (modeling) of $\rho\pi$ and other rare backgrounds ($B\bar{B}$ background), as well as the continuum background are expected to decrease with larger data samples and be thus reducible. Similarly, the fit bias is also considered to be reducible. The Belle measurement of $\mathcal{A}^{\pi^0\pi^0}$ [140] does not include uncertainties associated with flavor tagging. Therefore, a reducible flavor-tagging uncertainty is added to this list; the

³This source was included from Table VI of [151].

Table 7.2: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{B}^{\pi^+\pi^0}$ and of $\mathcal{B}^{\pi^+\pi^-}$. The values at 0.8 ab^{-1} correspond to Belle measurements and are taken from Table II in [157].

Source	$\delta\mathcal{B}^{\pi^+\pi^0}/\mathcal{B}^{\pi^+\pi^0} [\%]$		$\delta\mathcal{B}^{\pi^+\pi^-}/\mathcal{B}^{\pi^+\pi^-} [\%]$	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
Tracking	0.35	0.04	0.70	0.09
K - π PID ratio	0.86	0.11	1.72	0.22
PID Ratio cut	0.92	0.12	0.24	0.03
MC statistics	0.17	0.02	0.15	0.02
$N_{B\bar{B}}^{\text{prod}}$	1.37	1.37	1.37	1.37
π^0 det. efficiency	4.00	0.51	-	-
Signal PDF	0.73	0.09	0.50	0.06
Feed-across	1.19	0.15	1.50	0.19
PHOTOS	-	-	0.80	0.80
Charmless B bkg.	4.53	0.57	1.77	0.22
Total	6.49	1.59	3.42	1.63

value was taken from the BaBar $\mathcal{A}^{\pi^0\pi^0}$ measurement [151]. Considering now $\mathcal{B}^{\pi^0\pi^0}$, all the effects in Tab. 7.4 are reducible apart from the luminosity effect (number of neutral B mesons). The reducible effects include fitting effects, which were already discussed, and effects that depend on control samples like the π^0 detection efficiency.

There are other two particular effects affecting only the measurement of $\mathcal{B}^{\pi^0\pi^0}$; they are caused by the recovery of events with converted photons and by a special timing cut. At Belle II, the estimated number of signal events with e^+e^- pairs (with converted photons or Dalitz decays) after the final selection is about 2% of the estimated number of events with four ECL photons (see Sect. 6.9). Using events with e^+e^- pairs in addition to those with four ECL photons would improve the statistical error only by about 1%. Since this improvement is negligible, it is assumed here that e^+e^- pairs will be discarded for the measurement of $\mathcal{B}^{\pi^0\pi^0}$. Thus, the systematic effect caused by the recovery of converted photons is not considered.

The timing-cut effect originates from a special requirement on the trigger time of the ECL crystals. The requirement was employed to suppress out-of-time background events originating from e^+e^- interactions, such as Bhabha scattering, which leave large energy depositions in the ECL. These so-called ‘‘pileup’’ events are similar to genuine hadronic events with high energy back-to-back photons in the $\Upsilon(4S)$ frame and they pass the first-level trigger. At Belle II, the control samples to study the effect of this timing cut will increase according to the integrated luminosity. Hence, the respective uncertainty is expected to decrease.

Table 7.3: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{B}^{\pi^+\pi^0}$ and of $\mathcal{B}^{\pi^+\pi^-}$. The values at 0.8 ab^{-1} correspond to Belle measurements and are taken from Table II in [157].

Source	$\delta\mathcal{B}^{\pi^+\pi^0}/\mathcal{B}^{\pi^+\pi^0} [\%]$		$\delta\mathcal{B}^{\pi^+\pi^-}/\mathcal{B}^{\pi^+\pi^-} [\%]$	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
Tracking	0.35	0.04	0.70	0.09
K - π PID ratio	0.86	0.11	1.72	0.22
PID Ratio cut	0.92	0.12	0.24	0.03
MC statistics	0.17	0.02	0.15	0.02
$N_{B\bar{B}}^{\text{prod}}$	1.37	1.37	1.37	1.37
π^0 det. efficiency	4.00	0.51	-	-
Signal PDF	0.73	0.09	0.50	0.06
Feed-across	1.19	0.15	1.50	0.19
PHOTOS	-	-	0.80	0.80
Charmless B bkg.	4.53	0.57	1.77	0.22
Total	6.49	1.59	3.42	1.63

Table 7.4: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{A}^{\pi^0\pi^0}$ and of $\mathcal{B}^{\pi^0\pi^0}$. The values at 0.8 ab^{-1} are taken from the latest Belle measurement [140].

Source	$\delta\mathcal{A}^{\pi^0\pi^0}$		$\delta\mathcal{B}^{\pi^0\pi^0}/\mathcal{B}^{\pi^0\pi^0} [\%]$	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
$B\bar{B}$ bkg. param.	0.06	0.008	Cont. bkg. param	11.0
Cont. bkg. param	0.08	0.010	π^0 det. efficiency	4.4
Fit bias	0.02	0.003	Cont. ΔE and m_{bc}	4.0
Flavor tagging ³	0.034	0.0034	Assumed $\mathcal{B}^{\rho\pi}$	4.0
Total	0.12	0.01	Off-res cont. bkg.	3.0
			Assumed $\mathcal{B}^{\text{rare}}$	3.0
			Cont. fraction	1.8
			Fitted region	1.5
			$B\bar{B}$ bkg. fractions	1.5
			Luminosity	1.4
			Fit bias	1.0
			Rec. conv. photon	1.0
			Timing cut	0.5
			Total	14.2

For $\mathcal{A}^{\pi^+\pi^-}$ and $\mathcal{S}^{\pi^+\pi^-}$, the systematic effects given in Tab. 7.5 [160] are taken into account for the extrapolation. Apart from the effects caused by the tag side interference and by the Δt -resolution function, all effects in this list are assumed to scale with the luminosity. The effect denoted as flavor tagging, corresponds to the uncertainty on the wrong-tag fraction, which will improve with larger data samples. Most of the uncertainties in Tab. 7.5 are fitting uncertainties and depend on the accuracy of the shape description of the fit components. These uncertainties are considered to be reducible. The only irreducible uncertainties are the ones associated with the Δt resolution function and with the tag-side interference effect.

As Table 7.5 shows, the tag-side interference is the dominant systematic effect for $\mathcal{A}^{\pi^+\pi^-}$. As discussed in Sect. 4.10 and 4.11, the only way to completely eliminate this uncertainty is to use only semileptonic categories for flavor tagging, at the cost of an increase in the statistical uncertainty by about a factor $\sqrt{3} \approx 1.78$. However, the impact of this possible improvement on the measurement of ϕ_2 is minimal since the extrapolated systematic uncertainty of $\mathcal{A}^{\pi^+\pi^-}$ is not the largest uncertainty entering the isospin analysis.

Table 7.5: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{A}^{\pi^+\pi^-}$ and of $\mathcal{S}^{\pi^+\pi^-}$. The values at 0.8 ab^{-1} correspond to Belle measurements and are taken from Table II in [160].

Source	$\delta\mathcal{A}^{\pi^+\pi^-} [10^{-2}]$		$\delta\mathcal{S}^{\pi^+\pi^-} [10^{-2}]$	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
IP profile	0.13	0.016	1.19	0.151
B_{tag} track selection	0.30	0.038	0.33	0.042
Track helix	0.00	0.000	0.01	0.001
Δt selection	0.01	0.001	0.03	0.004
Vertex selection	0.37	0.047	0.23	0.029
Δz bias	0.50	0.063	0.40	0.051
Misalignment	0.40	0.051	0.20	0.025
τ_{B^0} and m_d	0.12	0.015	0.09	0.011
Data/MC shape	0.15	0.019	0.19	0.024
Δt resolution function	0.83	0.415	2.02	1.010
Flavor tagging	0.40	0.051	0.31	0.039
Bkg. param. shape	0.15	0.019	0.28	0.035
Bkg. nonparam. shape	0.37	0.047	0.57	0.072
Fit bias	0.54	0.068	0.86	0.109
Tag-side interference	3.18	3.180	0.17	0.170
Total	3.48	3.210	2.68	1.048

Table 7.6: Branching fractions and CP asymmetry parameters entering in the isospin analysis of the $B \rightarrow \pi\pi$ system: Belle measurements at 0.8 ab^{-1} together with the expected Belle II sensitivity at 50 ab^{-1} .

	Value	0.8 ab^{-1}	50 ab^{-1}
$\mathcal{B}^{\pi^+\pi^-}_{[10^{-6}]}$	5.04	$\pm 0.21 \pm 0.18$ [157]	$\pm 0.03 \pm 0.08$
$\mathcal{B}^{\pi^+\pi^0}_{[10^{-6}]}$	5.86	$\pm 0.26 \pm 0.38$ [157]	$\pm 0.03 \pm 0.09$
$\mathcal{B}^{\pi^0\pi^0}_{[10^{-6}]}$	1.31	$\pm 0.19 \pm 0.19$ [140]	$\pm 0.03 \pm 0.03$
$\mathcal{A}^{\pi^+\pi^0}$	0.14	$\pm 0.36 \pm 0.10$ [140]	$\pm 0.03 \pm 0.01$
$\mathcal{A}^{\pi^+\pi^-}$	0.33	$\pm 0.06 \pm 0.03$ [160]	$\pm 0.01 \pm 0.03$
$\mathcal{S}^{\pi^+\pi^-}$	-0.64	$\pm 0.08 \pm 0.03$ [160]	$\pm 0.01 \pm 0.01$

Table 7.6 presents a summary of the Belle measurements and the extrapolated uncertainties. For $\mathcal{S}^{\pi^0\pi^0}$, the systematic effects are expected to be similar to those affecting the measurement of $\mathcal{A}^{\pi^+\pi^-}$ and $\mathcal{S}^{\pi^+\pi^-}$. Thus, a systematic error of about $\delta\mathcal{S}^{\pi^0\pi^0} \approx 0.03$ is assumed. This gives in total

$$\delta\mathcal{S}^{\pi^0\pi^0} = \pm 0.28 \pm 0.03.$$

7.4.2 Results of the isospin analysis

The isospin analysis of $B \rightarrow \pi\pi$ was performed using the Belle measurements and the projections for Belle II summarized in Tab. 7.6. The analysis was performed without and with the $\mathcal{S}^{\pi^0\pi^0}$ constraint. The total uncertainties of the input variables are obtained by adding in quadrature the statistical and the systematic uncertainties. At Belle, only the correlations c_{ij} (see eq. (7.3)) related to the decay $B^0 \rightarrow \pi^+\pi^-$ were measured. The values of these correlation coefficients can be found in [162]. All the other coefficients were assumed to be null.

Figure 7.1 shows a geometrical representation of the isospin triangles in the complex plane of $B \rightarrow \pi\pi$ amplitudes. In Fig. 7.1 top, the weak phases and the proportions between the magnitudes of the amplitudes correspond to the results of the isospin analysis for $B \rightarrow \pi\pi$ using Belle measurements. Figure 7.1 bottom shows the results using the same data together with the constraint from $\mathcal{S}^{\pi^0\pi^0} = 0.83$. There are two isospin triangles, one for the amplitudes A^{ij} and another for the CP conjugate amplitudes \tilde{A}^{ij} . The latter can be mirrored due to the $\phi_{2,\text{eff}}$ symmetry in the isospin equations (see eq. (2.64)). Additionally, each isospin triangle can have two possible orientations leading in total to an 8-fold ambiguity in the determination of ϕ_2 . Applying the $\mathcal{S}^{\pi^0\pi^0}$ constraint, the orientation of each triangle is fixed, reducing the ambiguity by a factor 4.

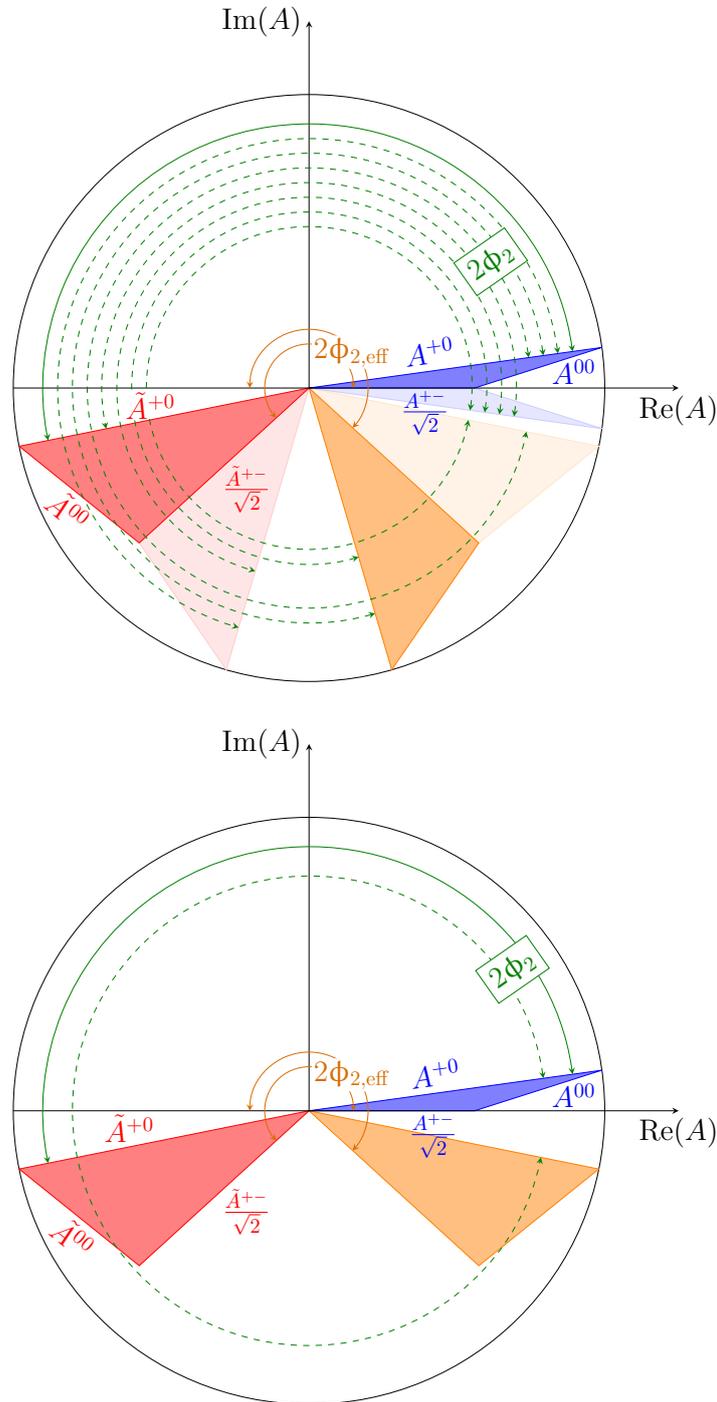


Figure 7.1: Geometrical representation of the isospin triangles. The weak phases and the proportions between the amplitudes correspond to the results of the isospin analysis for $B \rightarrow \pi\pi$ using Belle measurements (top) and using the same data together with the constraint $S^{\pi^0\pi^0} = 0.83$ (bottom). The SM compatible solution for $2\Phi_2$ is represented by the solid green line. The Mirror solutions correspond to the green dashed lines.

Figure 7.2 shows the results of the scan of the confidence for ϕ_2 . The results of the scan using the current Belle measurements show eight possible solutions. At the level of one σ (68% CL), Belle excludes the range $\phi_2 \in [9.5^\circ, 81.6^\circ]$ [140]. Figure 7.2 shows also the improvement of exclusion power at Belle II, still without a future measurement of $\mathcal{S}^{\pi^0\pi^0}$.

The scan including the $\mathcal{S}^{\pi^0\pi^0}$ constraint was performed for several possible values of $\mathcal{S}^{\pi^0\pi^0}$. For some of them, the ϕ_2 solutions are compatible with those obtained without the constraint. The compatible solutions for $\mathcal{S}^{\pi^0\pi^0}$ can be estimated by calculating the theoretical predictions with eq. (7.1). This calculation takes as input the fit parameters θ (see eq. (7.2)) obtained at the solutions of the scan without the constraint. There are four compatible $\mathcal{S}^{\pi^0\pi^0}$ solutions. For the generation of representative samples of signal MC events in chapter 6, these values were taken as input.

Because of the expected experimental uncertainty, it is possible that the value of $\mathcal{S}^{\pi^0\pi^0}$ will differ from the compatible solutions. Figure 7.3 shows different possible scenarios. The results of the scan of the confidence applying the $\mathcal{S}^{\pi^0\pi^0}$ constraint are presented for the four compatible solutions, and for two non-compatible values. From the eight ϕ_2 solutions, only two remain after including the $\mathcal{S}^{\pi^0\pi^0}$ constraint, reducing the number of possible solutions by a factor 4. Even in the cases where the value of $\mathcal{S}^{\pi^0\pi^0}$ is non-compatible, a large range can be excluded at one σ .

For the compatible value $\mathcal{S}^{\pi^0\pi^0} = 0.83$, the result of the scan is consistent with the current measurements of the unitarity triangle. The width of the solution around 92° is about 4° at the confidence level of one σ , thus $\delta\phi_2 \approx 2^\circ$.

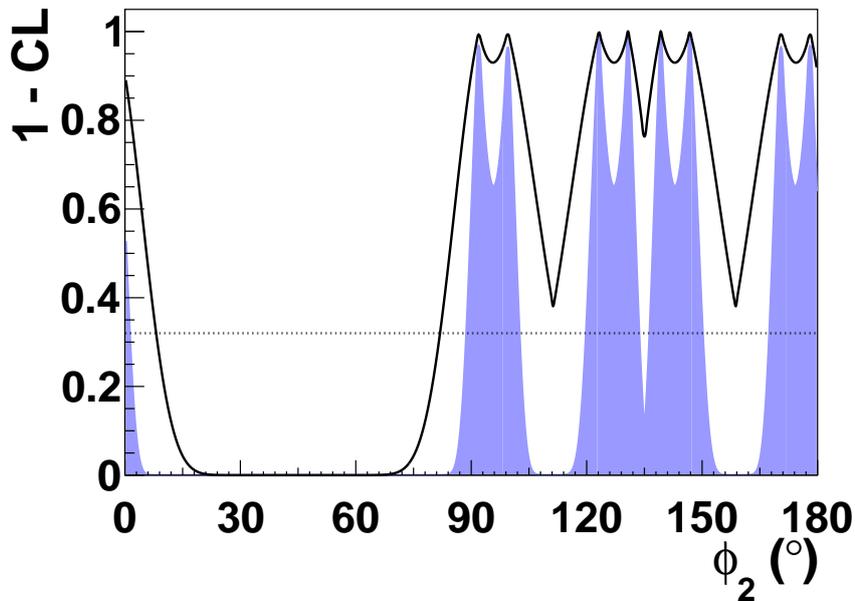


Figure 7.2: Scan of the confidence for ϕ_2 performing isospin analysis of the $B \rightarrow \pi\pi$ system. The black solid line (left) shows the result of the scan using data from Belle measurements (see Tab. 7.6). The blue shaded area shows the projection for Belle II without $\mathcal{S}^{\pi^0\pi^0}$ constraint. The dotted horizontal lines correspond to a CL of 68%.

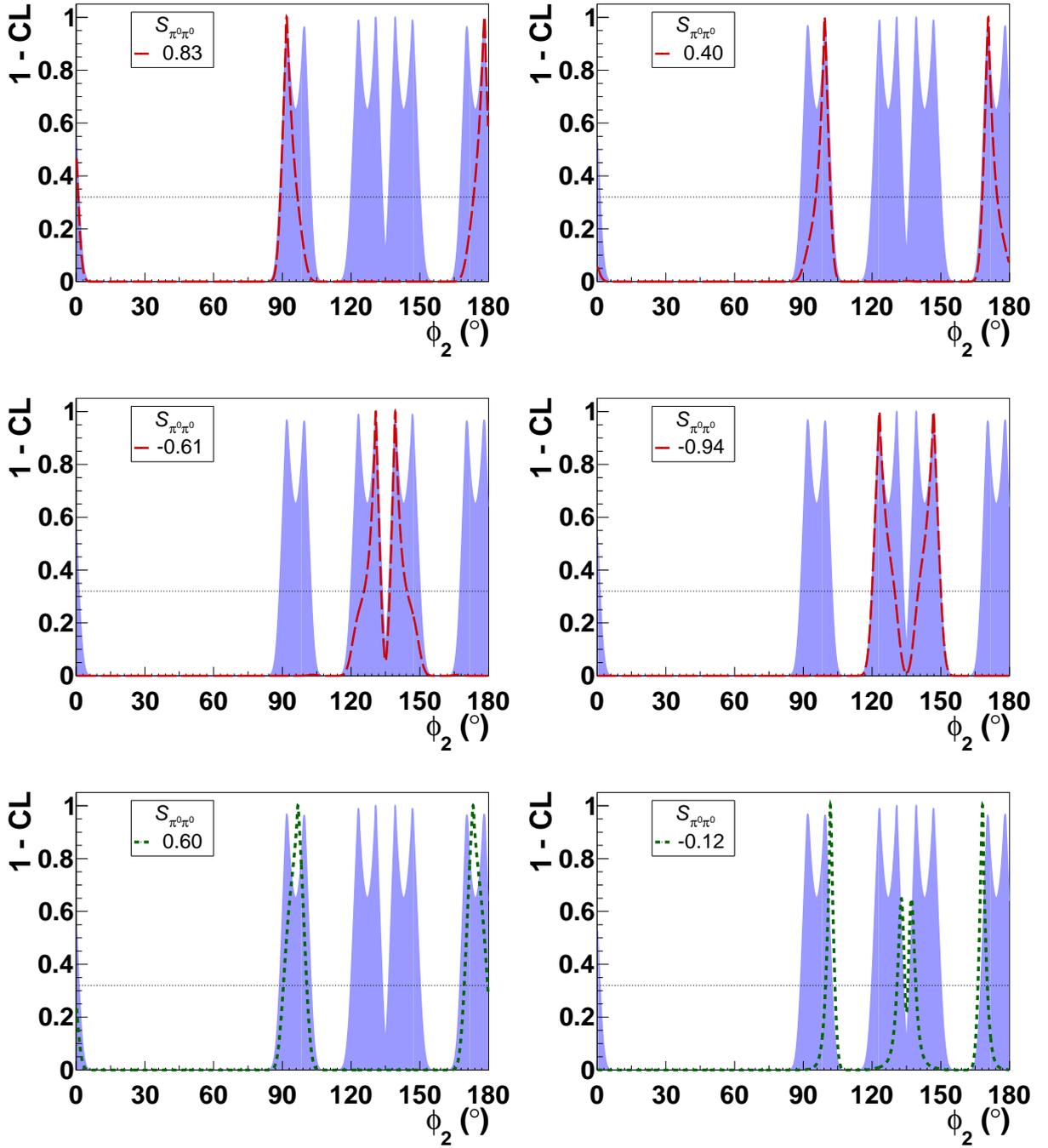


Figure 7.3: Scans of the confidence for ϕ_2 performing isospin analysis of the $B \rightarrow \pi\pi$ system. The blue shaded area in all plots shows the projection for Belle II without $\mathcal{S}^{\pi^0\pi^0}$ constraint. Results of the scan adding the $\mathcal{S}^{\pi^0\pi^0}$ constraint are shown for different $\mathcal{S}^{\pi^0\pi^0}$ values. The red dashed lines show the results for the four compatible values of $\mathcal{S}^{\pi^0\pi^0}$, and the green dotted lines for two non-compatible values. The dotted horizontal lines correspond to a CL of 68%.

7.5 The $B \rightarrow \rho\rho$ system

The measurement of ϕ_2 can be further improved by considering also the decay channel $B \rightarrow \rho\rho$. In the following, the Belle II precision for ϕ_2 is estimated considering, as for $B \rightarrow \pi\pi$, the isospin analysis of the $B \rightarrow \rho\rho$ system.

Table 7.7 presents the latest measurements of the observables entering the isospin analysis of the $B \rightarrow \rho\rho$ system. The measurements were performed at Belle, at BaBar and at LHCb. They are statistically compatible within one standard deviation, except for $f_L^{\rho^0\rho^0}$, where the measurement at Belle is compatible within two standard deviations with those at BaBar and at LHCb. As in the case of $B \rightarrow \pi\pi$ decays, LHCb has not covered those $B \rightarrow \rho\rho$ decays where neutral pions are present in the final state.

In contrast to the $B \rightarrow \pi\pi$ system, the final states of the $B \rightarrow \rho\rho$ system can be CP even (with orbital angular momentum $L = 0, 2$), or CP odd (with $L = 1$), each one requiring an independent isospin analysis. The isospin analysis is performed only for the dominant fraction of decays leading to longitudinally polarized final states, which are pure CP even states.

7.5.1 Expected precision at Belle II

In the same way as for the $B \rightarrow \pi\pi$ system in Sect. 7.4.1, the expected uncertainties for all the input parameters were estimated by extrapolating previous Belle measurements. Only for $\mathcal{A}^{\rho^0\rho^0}$ and $\mathcal{S}^{\rho^0\rho^0}$, BaBar measurements served as reference since there is no Belle measurement

Table 7.7: Latest measurements of the branching fractions, fractions of longitudinally polarised events and CP -violation parameters entering the isospin analysis of the $B \rightarrow \rho\rho$ system, performed by the Belle, the BaBar and the LHCb experiments.

	Belle	BaBar	LHCb
$f_L^{\rho^0\rho^0}$	$0.21 \pm 0.20 \pm 0.15$ [163]	$0.75^{+0.11}_{-0.14} \pm 0.05$ [86]	$0.745^{+0.048}_{-0.058} \pm 0.034$ [164]
$\mathcal{B}^{\rho^0\rho^0}_{[10^{-6}]}$	$1.02 \pm 0.30 \pm 0.15$ [163]	$0.92 \pm 0.32 \pm 0.14$ [86]	$0.94 \pm 0.17 \pm 0.09$ [164]
$f_L^{\rho^+\rho^-}$	$0.988 \pm 0.012 \pm 0.023$ [87]	$0.992 \pm 0.024^{+0.026}_{-0.013}$ [89]	
$\mathcal{B}^{\rho^+\rho^-}_{[10^{-6}]}$	$28.3 \pm 1.5 \pm 1.5$ [87]	$25.5 \pm 2.1^{+3.6}_{-3.9}$ [89]	
$\mathcal{A}^{\rho^+\rho^-}$	$0.00 \pm 0.10 \pm 0.06$ [87]	$-0.01 \pm 0.15 \pm 0.06$ [89]	
$\mathcal{S}^{\rho^+\rho^-}$	$-0.13 \pm 0.15 \pm 0.05$ [87]	$-0.17 \pm 0.20^{+0.05}_{-0.06}$ [89]	
$f_L^{\rho^+\rho^0}$	$0.95 \pm 0.11 \pm 0.02$ [88]	$0.950 \pm 0.015 \pm 0.006$ [90]	
$\mathcal{B}^{\rho^+\rho^0}_{[10^{-6}]}$	$31.7 \pm 7.1 \pm 5.3$ [88]	$23.7 \pm 1.4 \pm 1.4$ [90]	
$\mathcal{A}^{\rho^0\rho^0}$		$-0.2 \pm 0.8 \pm 0.3$ [86]	
$\mathcal{S}^{\rho^0\rho^0}$		$0.3 \pm 0.7 \pm 0.2$ [86]	

for these observables. The statistical uncertainties and the reducible systematic uncertainties were extrapolated as in eq. (6.14). The total systematic uncertainties were calculated by adding in quadrature the irreducible and the extrapolated reducible systematic uncertainties. Just as for $B \rightarrow \pi\pi$, the tables contained in the cited publications were used for the extrapolation of systematic errors.

Table 7.8 [163] shows the systematic effects considered for $\mathcal{B}^{\rho^0\rho^0}$ and for $f_L^{\rho^0\rho^0}$. Apart from the number of B mesons, all effects in this list are reducible. The uncertainties due to the track-reconstruction and due to the particle identification efficiencies were determined using independent control samples, and are thus expected to decrease with larger sample sizes. The error due to a difference in the fraction of misreconstructed events between data and MC is determined by varying this fraction and repeating the fit. The error on this fraction of events is expected to reduce with larger sample sizes. All other uncertainties in this list are associated with imperfections in the modeling of the fit components and are limited by the size of the data sample. The effect “ $\rho^0\pi\pi$ helicity” denotes the uncertainty due to the modeling of the $B^0 \rightarrow \rho^0\pi^+\pi^-$ helicity PDF.

At Belle, the interference between the four-pion final states was neglected in the measurements of $\mathcal{B}^{\rho^0\rho^0}$, $f_L^{\rho^0\rho^0}$, $\mathcal{B}^{\rho^+\rho^-}$, $f_L^{\rho^+\rho^-}$, $\mathcal{A}^{\rho^+\rho^-}$ and $\mathcal{S}^{\rho^+\rho^-}$. The impact of this effect was deter-

Table 7.8: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{B}^{\rho^0\rho^0}$ and of $f_L^{\rho^0\rho^0}$. The values at 0.8 ab^{-1} correspond to Belle measurements and are taken from Table VIII in [163].

Source	$\delta\mathcal{B}^{\rho^0\rho^0}/\mathcal{B}^{\rho^0\rho^0}[\%]$		$\delta f_L^{\rho^0\rho^0}$	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
$N_{B\bar{B}}^{\text{prod}}$	1.40	1.40	-	-
Tracking	1.40	0.18	-	-
PID	2.50	0.32	-	-
Misrec. fraction	2.40	0.30	0.030	0.004
Resonance shape	0.20	0.03	< 0.001	0.000
Model shape	5.10	0.65	0.110	0.014
Histogram shape	8.50	1.08	0.080	0.010
$\mathcal{B}(B \rightarrow a_1\pi)$	0.40	0.05	0.030	0.004
$\mathcal{B}(B \rightarrow b_1\pi)$	< 0.10	< 0.01	< 0.001	0.000
$\mathcal{B}(B \rightarrow a_2\pi)$	< 0.10	< 0.01	< 0.001	0.000
Fit bias	1.90	0.24	0.030	0.004
$\rho^0\pi\pi$ helicity	6.30	0.80	0.050	0.006
Interference	8.40	1.06	0.030	0.004
Total	15.11	2.36	0.157	0.020

Table 7.9: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{B}^{\rho^+\rho^-}$ and of $f_L^{\rho^+\rho^-}$. The values at 0.8 ab^{-1} correspond to Belle measurements and are taken from Table VIII in [87].

Source	$\delta\mathcal{B}^{\rho^+\rho^-}/\mathcal{B}^{\rho^+\rho^-}$ [%]		$\delta f_L^{\rho^+\rho^-}$ [10^{-2}]	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
$N_{B\bar{B}}^{\text{prod}}$	1.38	1.380	-	-
Tracking	0.70	0.089	-	-
PID	2.50	0.316	-	-
π^0 det. efficiency	2.98	0.377	-	-
IP profile	0.01	0.001	0.10	0.01
Δt selection	0.00	0.000	0.10	0.01
Vertex quality	0.16	0.020	0.00	0.00
B_{tag} track selection	0.01	0.001	0.10	0.01
Flavor tagging	0.07	0.009	0.20	0.03
Model shape	3.47	0.439	0.30	0.04
Histogram shape	0.17	0.022	0.20	0.03
MC composition	0.04	0.005	0.70	0.09
Misrec. fraction	0.01	0.001	0.10	0.01
Fixed bkg. yields	0.00	0.000	0.10	0.01
$B \rightarrow a_1^\pm \pi^\pm$ model	0.01	0.001	0.20	0.03
Fit bias	0.53	0.067	0.20	0.03
$B \rightarrow 4\pi$ CP violation	0.03	0.004	0.60	0.08
$(\rho^\pm \pi^\pm \pi^0)_{\text{NR}}$ helicity	0.04	0.005	2.00	0.25
Interference	0.01	0.001	0.20	0.03
Total	5.47	1.534	2.24	0.28

mined using MC sets which were generated such that the obtained yield ratio was equivalent to the one found in the data. Each MC set was fitted with an incoherent model; and the standard deviation of the fit results with respect to a calibration MC set was taken as systematic uncertainty for each observable. Large data samples allow to determine yield ratios and to model components more precisely. Thus this uncertainty can be considered to be reducible.

Table 7.9 [87] shows the list of systematic effects considered for $\mathcal{B}^{\rho^+\rho^-}$ and $f_L^{\rho^+\rho^-}$. Most of the effects in this list were already discussed for the measurement of $\mathcal{B}^{\rho^0\rho^0}$ and $f_L^{\rho^0\rho^0}$; The only irreducible effect is the uncertainty of the number of produced B mesons. The uncertainty due to the π^0 -detection efficiency is determined using control samples and is thus reducible. The “ $(\rho^\pm \pi^\pm \pi^0)_{\text{NR}}$ helicity” effect denotes the uncertainty due to the modeling of the helicity PDF for non-resonant $B^0 \rightarrow \rho^\pm \pi^\pm \pi^0$ decays. As previously discussed, effects related to modeling

of fit components are expected to be reducible. A particular effect in Tab. 7.9 is the one associated with flavor tagging. The measurement of $\mathcal{B}^{\rho^+\rho^-}$ and $f_L^{\rho^+\rho^-}$ was performed in bins of the flavor dilution factor r (see Sect. 4.2). This binning plays a role in the measurement of CP -violation parameters since the fraction of wrongly tagged events is different in each r bin. For the measurement of branching fractions this binning is not necessary. It is useful, for example, if one wants to disregard r bins where the signal-to-continuum ratio is very small, as discussed in Sect. 4.8. At Belle, the four observables $\mathcal{B}^{\rho^+\rho^-}$, $f_L^{\rho^+\rho^-}$, $\mathcal{A}^{\rho^+\rho^-}$ and $\mathcal{S}^{\rho^+\rho^-}$ were measured in a single analysis. Since the yield fraction is different in each r bin, this effect had to be taken into account for the determination of $\mathcal{B}^{\rho^+\rho^-}$ and $f_L^{\rho^+\rho^-}$. With larger data samples, the uncertainties on the wrong-tag fractions and on the yield fractions will decrease and therefore this effect is reducible.

For $\mathcal{A}^{\rho^+\rho^-}$ and $\mathcal{S}^{\rho^+\rho^-}$, the considered systematic effects are shown in Tab. 7.10 [87]. Apart from the tag side interference, all sources in this list are reducible. The uncertainties due to the IP profile, the vertex quality, the Δz bias, and the detector misalignment are expected to decrease due to the better alignment and the better knowledge of the size and the position of the interaction region at Belle II. The uncertainties due to selection criteria, such as the Δt selection and the B_{tag} track selection, are expected to decrease with larger data samples. Also the uncertainty of the MC composition is expected to decrease. The precision of physics parameters is assumed to improve with high-precision measurements at Belle and at LHCb. The uncertainty due to the possible CP violation in the $B\bar{B}$ backgrounds is expected to decrease since, with larger sample sizes, the accuracy in the modeling of the Δt distributions will improve. Among all systematic effects, the dominant source for $\mathcal{A}^{\rho^+\rho^-}$ is the tag-side interference. However, its effect in the determination of ϕ_2 is negligible since the uncertainty of $\mathcal{A}^{\rho^+\rho^-}$ is not the largest uncertainty entering the isospin analysis.

Table 7.11 [88] shows the systematic effects considered for $\mathcal{B}^{\rho^+\rho^0}$ and $f_L^{\rho^+\rho^0}$. Here again, all uncertainties are assumed to be reducible, except the one associated with the number of B mesons. The uncertainty due to continuum suppression was estimated using control samples, which will increase at high luminosity. Most of the other uncertainties are fitting uncertainties which also scale according to the integrated luminosity. In this measurement, there was an additional systematic effect due to a dependence of the detection efficiency on the ρ polarization. The uncertainty on this efficiency is expected to decrease with a larger dataset.

Considering now $\mathcal{A}^{\rho^0\rho^0}$ and $\mathcal{S}^{\rho^0\rho^0}$, although the reference measurement was performed by BaBar, the systematic effects affecting this measurement at Belle II are expected to be similar to those affecting the Belle measurement of $\mathcal{A}^{\rho^+\rho^-}$ and $\mathcal{S}^{\rho^+\rho^-}$. Therefore, the projected systematic uncertainties of $\mathcal{A}^{\rho^+\rho^-}$ and $\mathcal{S}^{\rho^+\rho^-}$ are taken also for $\mathcal{A}^{\rho^0\rho^0}$ and $\mathcal{S}^{\rho^0\rho^0}$. A summary of the Belle and the BaBar measurements together with the extrapolated uncertainties is presented in Tab. 7.12.

Table 7.10: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{A}^{\rho^+\rho^-}$ and of $\mathcal{S}^{\rho^+\rho^-}$. The values at 0.8 ab^{-1} correspond to Belle measurements and are taken from Table VIII in [87].

Source	$\delta\mathcal{A}^{\rho^+\rho^-} [10^{-2}]$		$\delta\mathcal{S}^{\rho^+\rho^-} [10^{-2}]$	
	0.8 ab^{-1}	50 ab^{-1}	0.8 ab^{-1}	50 ab^{-1}
IP profile	0.68	0.086	0.94	0.119
Δt selection	0.04	0.005	0.06	0.008
Track helix error	0.02	0.003	0.01	0.001
Vertex quality	1.20	0.152	0.60	0.076
B_{tag} track selection	0.84	0.106	0.95	0.120
Δz bias	0.50	0.063	0.40	0.200
Misalignment	0.40	0.051	0.20	0.025
Flavor tagging	0.71	0.090	0.51	0.065
Tag-side interference	1.02	1.020	0.08	0.080
Model shape	0.30	0.038	0.60	0.076
Histogram shape	0.19	0.024	0.31	0.039
Physics params.	0.02	0.003	0.02	0.003
MC composition	0.64	0.081	1.34	0.169
Misrec. fraction	0.60	0.076	0.50	0.063
Fixed bkg. yields	0.04	0.005	0.08	0.010
$B \rightarrow a_1^\pm \pi^\pm$ model	0.09	0.011	0.20	0.025
Fit bias	0.50	0.063	0.74	0.094
Bkg. CP violation	4.92	0.622	2.75	0.348
$B \rightarrow 4\pi$ CP violation	3.03	0.383	3.65	0.462
$(\rho^\pm \pi^\pm \pi^0)_{\text{NR}}$ helicity	0.12	0.015	0.77	0.097
Interference	0.12	0.015	0.15	0.019
Total	6.26	1.284	5.22	0.692

Table 7.11: Extrapolation of the systematic uncertainties affecting the measurement of $\mathcal{B}^{\rho^+\rho^0}$ and of $f_L^{\rho^+\rho^0}$. The values at 0.08 ab^{-1} are obtained from an early Belle measurement [88].

Source	$\delta\mathcal{B}^{\rho^+\rho^0}/\mathcal{B}^{\rho^+\rho^0}[\%]$		$\delta f_L^{\rho^+\rho^0}$	
	0.08 ab^{-1}	50 ab^{-1}	0.08 ab^{-1}	50 ab^{-1}
Tracking	3.40	0.14	-	-
PID	3.60	0.14	-	-
π^0 det. efficiency	4.00	0.16	-	-
Continuum suppression	5.40	0.22	-	-
ΔE fit	7.00	0.28	-	-
$N_{B\bar{B}}^{\text{prod}}$	1.38	1.38	-	-
Non-resonant decays	8.35	0.33	-	-
Rare B decays	3.30	0.13	-	-
Long. Polarization eff.	4.90	0.20	0.021	0.001
Total	15.01	1.50	0.021	0.001

7.5.2 Results of the isospin analysis

The isospin analysis of $B \rightarrow \rho\rho$ decays was performed using the Belle measurements and the projections for Belle II summarized in Tab. 7.12. For Belle II, the analysis was performed without and with $\mathcal{S}^{\rho^0\rho^0}$ constraint. For $B \rightarrow \rho\rho$ decays, the branching fractions have to be multiplied by the fractions of events with longitudinally polarized final states. For the decays $B^0 \rightarrow \rho^+\rho^-$ and $B^+ \rightarrow \rho^+\rho^0$, the fraction of events with longitudinally polarized final states is larger than 90%. The correlations between the observables are taken from the Belle measurements. Only the coefficients for the decay $B^0 \rightarrow \rho^+\rho^-$ are available. They can be found in [87]. All other coefficients are assumed to be null.

Figure 7.4 shows the results of the scan of the confidence for ϕ_2 . The results of the scan using the current Belle measurements are consistent with the published Belle results in [163]: there are two possible solutions, one that is compatible with the SM, $\phi_2 = (93.7 \pm 10.6)^\circ$, and a mirror solution $\phi_2 = (176.4 \pm 10.6)^\circ$.

Without $\mathcal{S}^{\rho^0\rho^0}$ constraint, the Belle II precision is about $\delta\phi_2 \sim 1^\circ$ at the one- σ level (68% CL). To estimate the precision with $\mathcal{S}^{\rho^0\rho^0}$ constraint, the value of $\mathcal{S}^{\rho^0\rho^0}$ is estimated such that the solutions of the scan applying the constraint are compatible with the solutions of the scan without it. The compatible value is $\mathcal{S}^{\rho^0\rho^0} = -0.14$. Figure 7.4 (bottom) shows the projections for Belle II without and with $\mathcal{S}^{\rho^0\rho^0}$. For a better visibility, the range that is consistent with the current measurements of the unitarity triangle was zoomed. Including the $\mathcal{S}^{\rho^0\rho^0}$ constraint, the expected uncertainty of ϕ_2 at the one- σ level reduces by about one third: $\delta\phi_2 \sim 0.7^\circ$.

Table 7.12: Branching fractions, fractions of longitudinally polarised events and CP violation parameters entering the isospin analysis of the $B \rightarrow \rho\rho$ system: Belle measurements at 0.8 ab^{-1} and 0.08 ab^{-1} , BaBar measurements at 0.5 ab^{-1} and expected Belle II sensitivity at 50 ab^{-1} .

	Value	0.8 ab^{-1}	50 ab^{-1}
$f_L^{\rho^0\rho^0}$	0.21	$\pm 0.20 \pm 0.15$ [163]	$\pm 0.03 \pm 0.02$
$\mathcal{B}^{\rho^0\rho^0}_{[10^{-6}]}$	1.02	$\pm 0.30 \pm 0.15$ [163]	$\pm 0.04 \pm 0.02$
$f_L^{\rho^+\rho^-}$	0.988	$\pm 0.012 \pm 0.023$ [87]	$\pm 0.002 \pm 0.003$
$\mathcal{B}^{\rho^+\rho^-}_{[10^{-6}]}$	28.3	$\pm 1.5 \pm 1.5$ [87]	$\pm 0.19 \pm 0.4$
$\mathcal{A}^{\rho^+\rho^-}$	0.00	$\pm 0.10 \pm 0.06$ [87]	$\pm 0.01 \pm 0.01$
$\mathcal{S}^{\rho^+\rho^-}$	-0.13	$\pm 0.15 \pm 0.05$ [87]	$\pm 0.02 \pm 0.01$
	Value	0.08 ab^{-1}	50 ab^{-1}
$f_L^{\rho^+\rho^0}$	0.95	$\pm 0.11 \pm 0.02$ [88]	$\pm 0.004 \pm 0.001$
$\mathcal{B}^{\rho^+\rho^0}_{[10^{-6}]}$	31.7	$\pm 7.1 \pm 5.3$ [88]	$\pm 0.3 \pm 0.5$
	Value	0.5 ab^{-1}	50 ab^{-1}
$\mathcal{A}^{\rho^0\rho^0}$	-0.2	$\pm 0.8 \pm 0.3$ [86]	$\pm 0.08 \pm 0.01$
$\mathcal{S}^{\rho^0\rho^0}$	0.3	$\pm 0.7 \pm 0.2$ [86]	$\pm 0.07 \pm 0.01$

For $B \rightarrow \rho\rho$ decays, the isospin analysis results in only two ϕ_2 solutions event without the $\mathcal{S}^{\rho^0\rho^0}$ constraint. This occurs because, for the decay $B^0 \rightarrow \rho^0\rho^0$, the product of branching fraction and fraction of longitudinally polarized events is roughly hundred times smaller than for the other $B \rightarrow \rho\rho$ decays. Thus, the A^{00} amplitude is very small. Figure 7.5 shows a geometrical representation of the respective isospin triangles. The weak phases and the proportions between the magnitude of the amplitudes are obtained from the isospin analysis. As Figure 7.5 shows, the triangles shrink nearly into lines, leading to a collapse of the eight possible solutions into two distinct solutions.

7.6 Expected precision combining the $B \rightarrow \pi\pi$ and the $B \rightarrow \rho\rho$ systems

At the current level of precision, the world average for ϕ_2 is dominated by the $B \rightarrow \rho\rho$ system and to a lesser extent by the $B \rightarrow \pi\pi$ system [71]. The future ϕ_2 precision at Belle II can be thus estimated by combining the isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$. For this, the projections for Belle II in Tabs. 7.6 and 7.12 are taken as input. The combination is performed

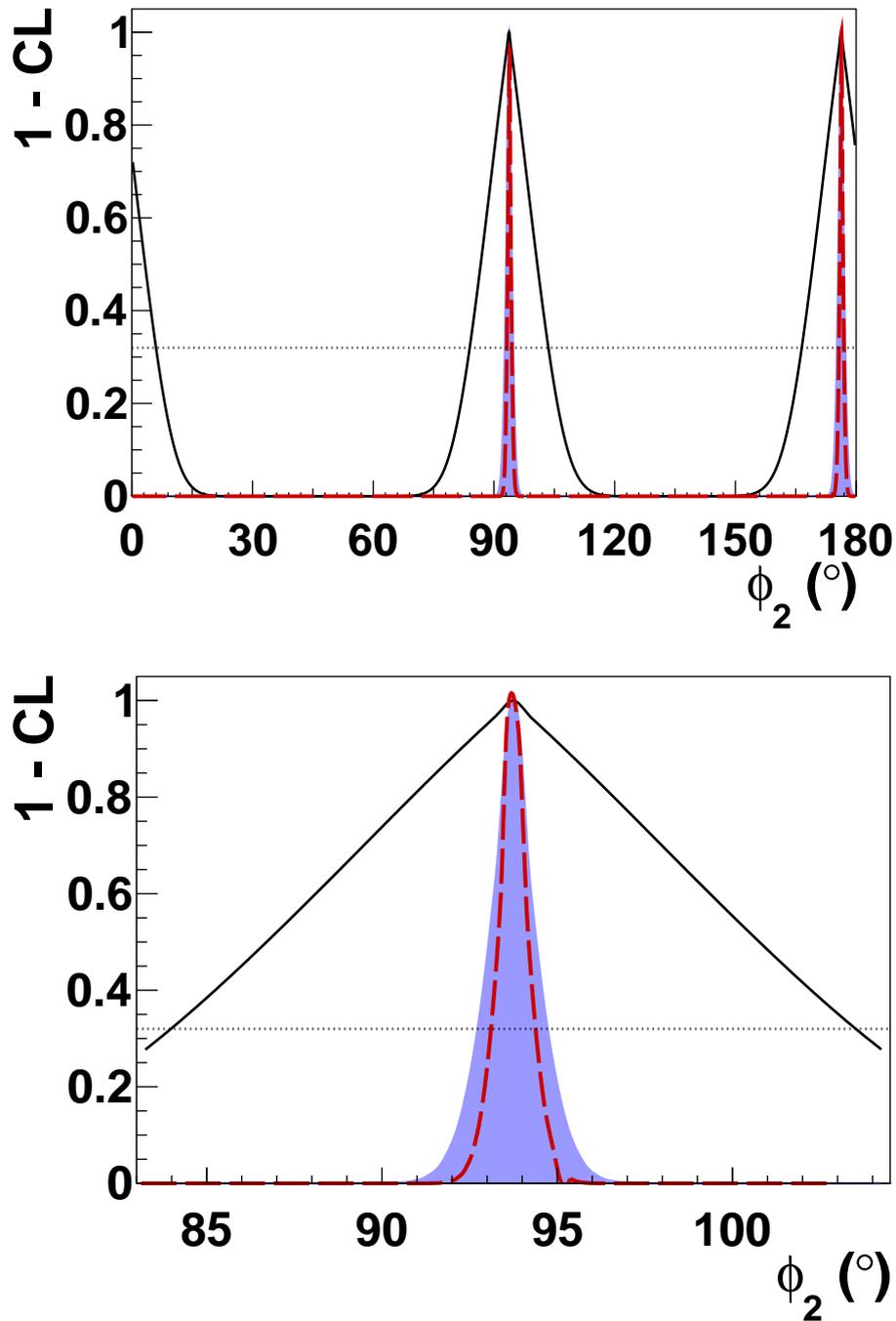


Figure 7.4: Scan of the confidence level of ϕ_2 performing isospin analysis of the $B \rightarrow \rho\rho$ system. The black solid line shows the result of the scan using data from Belle measurements (see Tab. 7.12). The blue shaded area shows the projection for Belle II without the $\mathcal{S}^{\rho^0\rho^0}$ constraint. The red dashed line shows the result adding the constraint $\mathcal{S}^{\rho^0\rho^0} = -0.14 \pm 0.07 \pm 0.01$. The full range $\phi_2 \in [0^\circ, 104^\circ]$ is shown on the top. For a better visibility, the range $\phi_2 \in [83^\circ, 104^\circ]$ is zoomed (bottom). The dotted horizontal lines correspond to a CL of 68%.

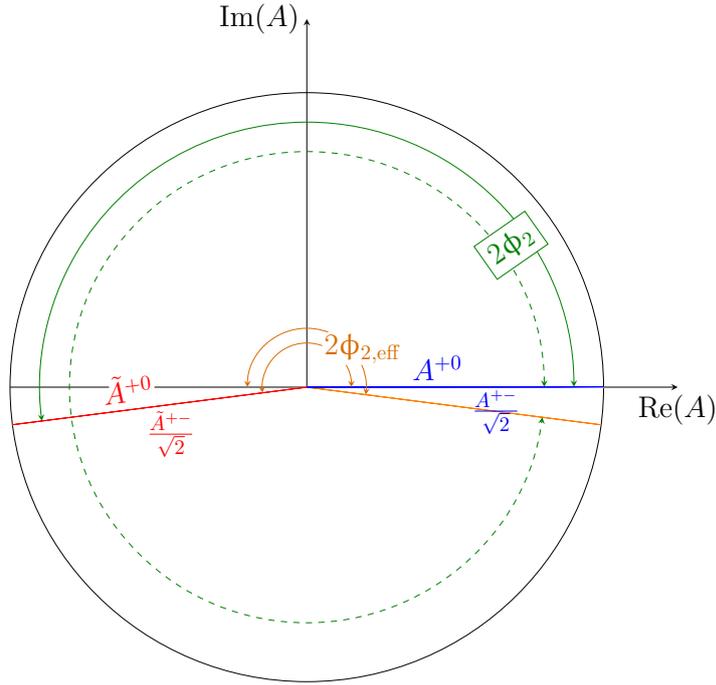


Figure 7.5: Geometrical representation of the isospin triangular relations in the complex plane of $B \rightarrow \rho\rho$ amplitudes. The weak phases and the proportions between the magnitudes of the amplitudes correspond to the results of the isospin analysis for $B \rightarrow \rho\rho$ using Belle measurements. The SM compatible solution for $2\phi_2$ is represented by the solid green line. Due to the small size of the amplitude \mathcal{A}^{00} , the isospin triangles contract and become practically lines.

by adding the individual $\chi^2(\boldsymbol{\theta})$ values, and using the total $\chi^2(\boldsymbol{\theta})$ to calculate the confidence level.

At the level of precision at Belle, the solutions of ϕ_2 from the individual isospin analyses are compatible within one σ . Using the central values of the physical observables from Belle, this is not the case at the expected level of precision at Belle II. In order to have central values of the input observables that are consistent with one value of ϕ_2 , the value of $\mathcal{B}^{\pi^0\pi^0}$ is adjusted by 3% downwards (the adjusted value is $1.27 \cdot 10^{-6}$). This adjustment, which is within one σ of the measured value, ensures that the solutions of the isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ correspond to the same value of ϕ_2 .

Figure 7.6 shows the results of the scan of the confidence for ϕ_2 combining the isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$. As for the individual analyses, the correlations between the observables are taken from Belle measurements [87, 162]. The ϕ_2 scan using current Belle measurements is performed without the \mathcal{S}^{00} constraints. The projections for Belle II are performed for both cases, without and with the \mathcal{S}^{00} constraints.

Without the constraints, the estimated sensitivity is found to be about $\delta\phi_2 \sim 1^\circ$. For the case with the \mathcal{S}^{00} constraints, the analysis is performed with central values of $\mathcal{S}^{\pi^0\pi^0}$ and $\mathcal{S}^{\rho^0\rho^0}$ that are compatible in terms of ϕ_2 . These values are $\mathcal{S}^{\pi^0\pi^0} = 0.75 \pm 0.28 \pm 0.03$ and

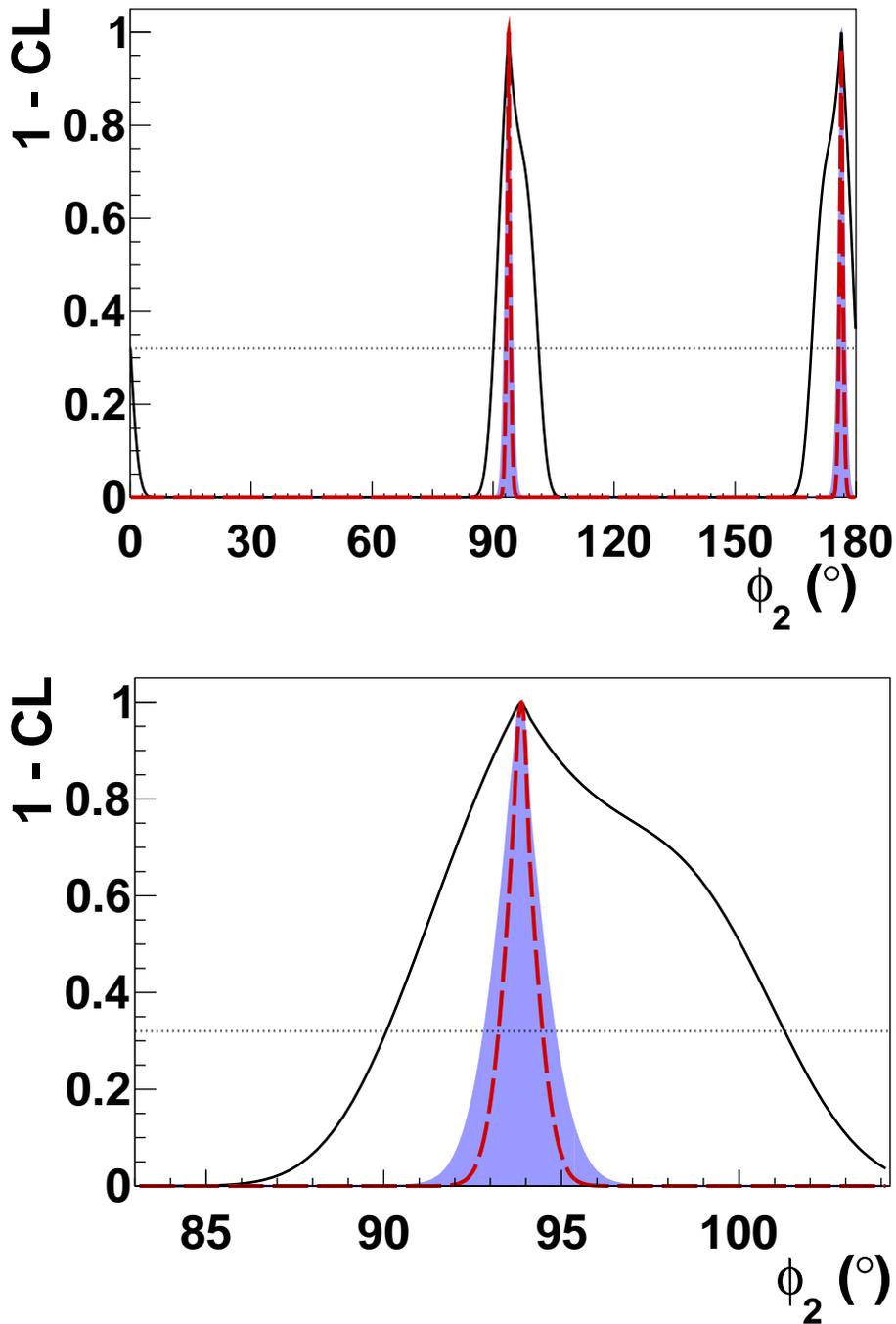


Figure 7.6: Scan of the confidence for ϕ_2 combining the isospin analyses of the $B \rightarrow \pi\pi$ and the $B \rightarrow \rho\rho$ systems. The black solid line shows the result of the scan using data from Belle measurements. The blue shaded area shows the projection for Belle II without S^{00} constraints. The red dashed line shows the result adding the constraints $S^{\pi^0\pi^0} = 0.75$ and $S^{\rho^0\rho^0} = -0.14$. The full range $\phi_2 \in [0^\circ, 104^\circ]$ is shown on the top. For a better visibility, the range $\phi_2 \in [83^\circ, 104^\circ]$ is zoomed (bottom). The dotted horizontal lines correspond to a CL of 68%.

$\mathcal{S}^{\rho^0\rho^0} = -0.14 \pm 0.07 \pm 0.01$. With the constraints, the improvement in the precision at the one- σ level is about factor 2: from $\delta\phi_2 \sim 1^\circ$ to $\delta\phi_2 \sim 0.6^\circ$. In both cases, however, a 2-fold ambiguity remains. As Figure 7.6 shows, the results are dominated by the $B \rightarrow \rho\rho$ system.

7.7 Concluding remarks and prospects

This chapter was devoted to estimate the expected precision for the unitarity angle ϕ_2 at Belle II. For this, isospin analyses of the systems $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ were performed. At present, the uncertainties on ϕ_2 are dominated by the measurements at Belle and at BaBar. There is an 8-fold ambiguity in the determination of ϕ_2 using $B \rightarrow \pi\pi$, which can be reduced to a 2-fold ambiguity using $B \rightarrow \rho\rho$.

The Belle II precision for the input observables was estimated by projecting the available Belle measurements. For the observables $\mathcal{A}^{\rho^0\rho^0}$ and $\mathcal{S}^{\rho^0\rho^0}$, the BaBar measurement was considered. Statistical uncertainties were extrapolated assuming a data sample corresponding to an integrated luminosity of 50 ab^{-1} . Systematic uncertainties were estimated depending on whether they are reducible or irreducible.

Using the projected measurements as input, the precision for ϕ_2 was estimated via individual and via combined isospin analysis of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$. The analyses were performed with and without the \mathcal{S}^{00} constraints. In the case of $B \rightarrow \pi\pi$, applying the \mathcal{S}^{00} constraint with the estimated uncertainty of $\delta\mathcal{S}^{\pi^0\pi^0} = 0.28 \pm 0.03$ reduces the 8-fold ambiguity by a factor 4. Since the value of $\mathcal{S}^{\pi^0\pi^0}$ can differ from the expected value within the SM, its measurement offers the opportunity to find further tensions giving a hint at NP.

Assuming values of \mathcal{S}^{00} that are consistent with the current ϕ_2 solutions, the \mathcal{S}^{00} constraints reduce the uncertainty of ϕ_2 . However, a 2-fold ambiguity remains: one solution is compatible with the SM, and the other is a mirror solution. The expected uncertainties of ϕ_2 are summarized in Tab. 7.13.

Table 7.13: Current world average error [71] and expected uncertainties of the determination of ϕ_2 performing isospin analyses of the decay systems $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ together with a combined isospin analysis of these two systems. For the current world average error, also the decay system $B \rightarrow \rho\pi$ was considered.

Channel	$\delta\phi_2$ [$^\circ$]
Current world average	+4.4 -4.0
$B \rightarrow \pi\pi$	2.0
$B \rightarrow \rho\rho$	0.7
$B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ Combined	0.6

Isospin-breaking effects were ignored. At present, isospin-breaking effects can be only partially included, and are of the order of magnitude of $1^\circ - 2^\circ$ [71]. At the level of precision expected at Belle II, the inclusion of these effects still has to be examined [97].

In the projections performed in this chapter, the system $B \rightarrow \rho\pi$ was not considered. The determination of ϕ_2 from this system is not framed in a way that allows a simple estimation of the future precision at Belle II: the interference between the contributing resonances is very difficult to simulate in pseudo-experiments; and the uncertainties of the 27 input coefficients, which have to be extracted via time-dependent Dalitz-plot analysis, cannot be easily extrapolated. However, since there is no inherent ambiguity in the determination of ϕ_2 using $B \rightarrow \rho\pi$, this system will play an important role at Belle II: it will give additional input to rule out the mirror solutions which are not compatible with the SM (see Sect. 2.10.3). Theoretically, this system is the less affected by isospin-breaking effects. But experimentally it is the most challenging. The precision for ϕ_2 is thus expected to be dominated by $B \rightarrow \rho\rho$ and $B \rightarrow \pi\pi$.

One should note that a major improvement in the precision for ϕ_2 requires new precise measurements at Belle II because LHCb is not competitive for decays involving neutral pions.

8 Conclusion and outlook

Within this thesis, three major goals were achieved. First, a new flavor tagging algorithm for the Belle II experiment was developed, optimized, characterized and validated. This algorithm now serves as the official Belle II flavor tagger. Second, the Belle II performance for the time-dependent analysis of the decay $B^0 \rightarrow \pi^0\pi^0$ was studied, leading ultimately to an estimation of the precision for the two CP -violation parameters $\mathcal{S}^{\pi^0\pi^0}$ and $\mathcal{A}^{\pi^0\pi^0}$, and for the branching fraction $\mathcal{B}^{\pi^0\pi^0}$. Third, the precision on the future Belle II determination of the unitarity angle ϕ_2 was estimated. The studies presented in this thesis were published in dedicated sections of the Belle II physics book [1].

A novel flavor tagger

The new Belle II flavor tagging algorithm developed in this thesis is designed to cope with the unprecedented high-luminosity conditions and the increased beam backgrounds at SuperKEKB. It exploits the capabilities of the new Belle II detector and the new Belle II reconstruction algorithms. It is based on the concept of tagging categories inspired by the signatures present in flavor-specific B^0 decays. The algorithm employs a novel multivariate method developed especially for Belle II, the Fast Boosted Decision Tree (FBDT). The development of the algorithm is now complete; further information contained in the events beyond the ones used in the algorithm was shown to have no additional discrimination power.

In the absence of Belle II data, the new Belle II flavor tagger was validated and characterized on simulated Monte Carlo events and on Belle data. The validation sample was obtained from the full Belle data sample, corresponding to 711 fb^{-1} collected at the $\Upsilon(4S)$ resonance, and corresponds to events where the golden decay mode $B \rightarrow J/\psi K_S^0$ was fully reconstructed on the signal side. The validation included a comprehensive test of robustness in which the input information was systematically varied and the results on Belle MC were compared with the results on Belle data. In all the cases, the results on data and on MC events were consistent within the uncertainties.

In the algorithm, the information provided by the individual categories is combined by an FBDT method. To cross-check the results, a multi-layer perceptron (MLP) is used to combine independently the flavor information. Using the output of the FBDT combiner, the Belle II flavor tagger reaches, on Belle data, an effective efficiency of

$$\varepsilon_{\text{eff}} = (33.5 \pm 0.5)\%.$$

As a cross-check, the MLP combiner reaches $(33.4 \pm 0.5)\%$. The uncertainties are only statistical; true systematic uncertainties can be determined once data from Belle II are available. For comparison, the Belle algorithm reached an effective efficiency of $(30.1 \pm 0.4)\%$ [29]. The increase of 10% in effective efficiency corresponds to the improvements in the new Belle II algorithm.

Using the output of the FBDT combiner, the effective efficiency on the latest official Belle II MC sample (July 2018), which was generated with simulated background, is

$$\varepsilon_{\text{eff}} = (33.89 \pm 0.04)\%.$$

As a cross check, the MLP combiner reaches $(33.88 \pm 0.05)\%$. The uncertainties are again only statistical. Even under the expected challenging background conditions, Belle II will outperform its predecessor by about 10%. The effective efficiency is measured at around 34% at Belle II because of the improved flavor tagger, the improved reconstruction algorithms, and the improved detector.

Due to the novel nano-beam scheme at SuperKEKB, the distribution of the B_{tag}^0 -decay vertex at Belle II is about twenty times narrower (in z -direction) than at Belle. Therefore, the performance of the flavor tagger at Belle II will be slightly influenced by the effect of the CP -violation on the B_{tag}^0 -decay vertex. At Belle, this effect did not play a role due to the wide B_{tag}^0 -decay vertex distribution. One of the major conclusions is that, to counteract this effect, it is mandatory to train the flavor tagger using Monte Carlo events generated without CP -violation.

Time-dependent CP analysis of $B^0 \rightarrow \pi^0\pi^0$

The second part of this thesis was dedicated to the detailed study of the Belle II performance for the time-dependent analysis of the decay $B^0 \rightarrow \pi^0\pi^0$, an analysis that has never been performed, and for which experiments at hadron colliders cannot be competitive. The results of the study show that the analysis is feasible for the expected Belle II sample size of 50 ab^{-1} . The proposed method to measure $\mathcal{S}^{\pi^0\pi^0}$ uses e^+e^- tracks from converted photons $\gamma_C \rightarrow e^+e^-$ and from Dalitz $\pi_{\text{Dal}}^0 \rightarrow e^+e^-\gamma$ decays to reconstruct the B^0 -decay vertex. Since events with Dalitz decays have a better vertex resolution compared to events with converted photons, a simple, but rather effective method was developed to distinguish between the two kinds of events exploiting the geometrical properties of the e^+e^- track pairs.

The statistical uncertainty on $\mathcal{S}^{\pi^0\pi^0}$ was estimated performing sets of pseudo-experiments. In the absence of control samples from real Belle II data, the systematic uncertainties were assumed to be similar to the ones from the extrapolated measurement of $\mathcal{S}^{\pi^+\pi^-}$. Assuming a data sample of 50 ab^{-1} , the result is

$$\delta\mathcal{S}^{\pi^0\pi^0} = \pm 0.28 \text{ (stat)} \pm 0.03 \text{ (syst)}.$$

To validate the procedure for $\mathcal{S}^{\pi^0\pi^0}$, a time-integrated analysis was performed to estimate the Belle II precision for the direct CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ and the branching fraction $\mathcal{B}^{\pi^0\pi^0}$. This analysis used events where the four final-state photons were reconstructed as clusters in the electromagnetic calorimeter. The statistical uncertainties were estimated by performing pseudo-experiments, and the systematic uncertainties by extrapolating the latest Belle measurement of $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{B}^{\pi^0\pi^0}$. Assuming a data sample of 50 ab^{-1} , the result is

$$\delta\mathcal{A}^{\pi^0\pi^0} = \pm 0.03 \text{ (stat)} \pm 0.01 \text{ (syst)}, \quad \delta\mathcal{B}_{\pi^0\pi^0} = \pm 0.03 \text{ (stat)} \pm 0.03 \text{ (syst)}.$$

Expected precision for the unitarity angle ϕ_2

In the last part of this thesis, the impact of the $\mathcal{S}^{\pi^0\pi^0}$ measurement on the determination of the unitarity angle ϕ_2 was studied based on the isospin analysis of the decays $B \rightarrow \pi\pi$. The studies showed that, applying the $\mathcal{S}^{\pi^0\pi^0}$ constraint, with the estimated uncertainty for $\mathcal{S}^{\pi^0\pi^0}$, reduces the 8-fold ϕ_2 ambiguity to a 2-fold one.

As a result of the combined isospin analyses of $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$, the overall expected Belle II precision for ϕ_2 at 50 ab^{-1} is $\delta\phi_2 \approx 0.6^\circ$, which is more than five times smaller than the current world average. An important remark is that, since LHCb is not covering B decays involving neutral pions, new precise Belle II measurements are required to improve the precision for ϕ_2 .

Outlook

The new Belle II flavor tagger developed in this thesis will be an essential instrument to ensure the success of the Belle II physics program. Even under the new challenging experimental conditions at SuperKEKB, the new flavor tagger outperforms its predecessors reaching a higher effective efficiency. The high-performance of the Belle II flavor tagger will help reducing the uncertainties on CP -violation measurements, and will boost Belle II's sensitivity to possible New Physics effects. What still has to be performed is to calibrate the algorithm on Belle II data using control samples of self-tagging modes, and to study the impact of the new systematic effect associated with the novel nano-beam scheme.

The performance study for the decay $B^0 \rightarrow \pi^0\pi^0$ showed that the measurement of the missing observable $\mathcal{S}^{\pi^0\pi^0}$ is feasible at Belle II. However, since the measurement is very challenging, the full expected data sample of 50 ab^{-1} will be required. With the expected precision for $\mathcal{S}^{\pi^0\pi^0}$, Belle II will be able to reduce the current ambiguities on the unitarity angle ϕ_2 by a factor 4, but may, depending on the measured $\mathcal{S}^{\pi^0\pi^0}$ value, also contribute together with other measurements to further tensions in the SM, pointing ultimately to New Physics.

A Physics

A.1 The Δt probability density for time-dependent CP analysis

Starting from the decay rates eqs. (2.30) and (2.31) and taking into account the the GIM cancellations, i.e. setting $|\frac{p}{q}| = 1$ (eq. (2.25)), one can write

$$\begin{aligned}
 \Gamma(t_1, t_2, q) &= |\langle f_1, f_2 | H | B_1(t_1), B_2(t_2) \rangle|^2 \\
 &= \frac{1}{4} e^{-\Gamma(t_1+t_2)} |A_1|^2 |A_2|^2 \\
 &\quad \cdot [(|\lambda_{CP}|^2 + 1) + q [(|\lambda_{CP}|^2 - 1) \cos(\Delta m \Delta t) + 2\text{Im}(\lambda_{CP}) \sin(\Delta m \Delta t)]].
 \end{aligned} \tag{A.1}$$

The goal is to calculate the Δt probability density function $P(\Delta t, q)$. The sum of both flavor contributions, i.e. the sum over the two possible flavors on the tag side have to be normalized. Since only the time difference $\Delta t = t_1 - t_2$ is experimentally accessible and $t_1, t_2 \in [0, \infty)$, one normalizes to unity in the region $-\infty < \Delta t < \infty$, i.e.

$$\int_{-\infty}^{\infty} \sum_q P(\Delta t, q) \cdot d\Delta t = \frac{1}{N} \int_{-\infty}^{\infty} \sum_q \Gamma(t_1, t_2, q) \cdot d\Delta t = 1.$$

Hence, the norm factor can be calculated calculated by:

$$\begin{aligned}
 N &= \int_{-\infty}^{\infty} \sum_q \Gamma(t_1, t_2, q) \cdot d\Delta t \\
 &= \int_{-\infty}^{\infty} \frac{1}{2} e^{-\Gamma(t_1+t_2)} A_1^2 A_2^2 [1 + |\lambda_{CP}|^2] \cdot d\Delta t \\
 &\quad \text{with } c := \frac{1}{2} A_1^2 A_2^2 [|\lambda_{CP}|^2 + 1] \\
 &= \int_{-\infty}^{\infty} c e^{-\Gamma(t_1+t_2)} \cdot d\Delta t.
 \end{aligned}$$

For $t_1 > t_2$, i.e. $\Delta t > 0$,

$$N = c e^{-2\Gamma t_2} \int_0^\infty e^{-\Gamma \Delta t} \cdot d\Delta t,$$

and for $t_1 < t_2$, i.e. $\Delta t < 0$,

$$N = c e^{-2\Gamma t_1} \int_{-\infty}^0 e^{\Gamma \Delta t} \cdot d\Delta t.$$

Now one defines $t' = \min(t_1, t_2)$:

$$\begin{aligned} N &= c e^{-2\Gamma t'} \int_{-\infty}^\infty e^{-\Gamma |\Delta t|} \cdot d\Delta t \\ &= \frac{2}{\Gamma} c e^{-2\Gamma t'} \\ &= \tau e^{-2\Gamma t'} A_1^2 A_2^2 [|\lambda_{CP}|^2 + 1], \end{aligned}$$

where $\Gamma := \frac{1}{\tau}$. Rewriting now eq. (A.1) one obtains:

$$\begin{aligned} \Gamma(t_1, t_2, q) &= \Gamma(t', |\Delta t|, q) \\ &= \frac{1}{4} e^{-2\Gamma t'} \cdot e^{-\Gamma |\Delta t|} |A_1|^2 |A_2|^2 \\ &\quad \cdot [(|\lambda_{CP}|^2 + 1) + q [(|\lambda_{CP}|^2 - 1) \cos(\Delta m \Delta t) + 2\text{Im}(\lambda_{CP}) \sin(\Delta m \Delta t)]]. \end{aligned}$$

This leads finally to

$$\begin{aligned} P(\Delta t, q) &= \frac{1}{N} \Gamma(t', |\Delta t|, q) \\ &= \frac{1}{4} \frac{e^{-2\Gamma t'} \cdot e^{-\Gamma |\Delta t|} A_1^2 A_2^2}{\tau e^{-2\Gamma t'} A_1^2 A_2^2} \\ &\quad \cdot \left[\frac{|\lambda_{CP}|^2 + 1}{|\lambda_{CP}|^2 + 1} + q \left[\frac{|\lambda_{CP}|^2 - 1}{|\lambda_{CP}|^2 + 1} \cos(\Delta m \Delta t) + \frac{2\text{Im}(\lambda_{CP})}{|\lambda_{CP}|^2 + 1} \sin(\Delta m \Delta t) \right] \right] \\ &= \frac{e^{-\Gamma |\Delta t|}}{4\tau} [1 + q (\mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t))]. \end{aligned}$$

A.2 Probability density for time-integrated CP analysis

To obtain a probability function that is independent of Δt , but that takes the effect of B^0 mixing into account, one needs to integrate eq. (2.33) over the whole range $-\infty < \Delta t < \infty$,

$$\int_{-\infty}^{\infty} \mathcal{P}(\Delta t, q) \cdot d\Delta t = \int_{-\infty}^{\infty} \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} [1 + q (\mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t))] \cdot d\Delta t.$$

First, one should consider that $e^{-|x|}$ and $\cos(x)$ are even functions and that $\sin(x)$ is an odd function. The product of two even functions yields again an even function, and the product of an even function with an odd function yields an odd function. The integral of an odd function in a range $x \in [-a, a]$ is zero for any a and thus $\int_{-\infty}^{\infty} e^{-|x|} \sin x = 0$. Consequently, one needs to integrate only

$$\int_{-\infty}^{\infty} \mathcal{P}(\Delta t, q) \cdot d\Delta t = \frac{1}{4\tau_{B^0}} \left[\int_{-\infty}^{\infty} e^{-\frac{|\Delta t|}{\tau_{B^0}}} \cdot d\Delta t + q\mathcal{A}_{CP} \int_{-\infty}^{\infty} e^{-\frac{|\Delta t|}{\tau_{B^0}}} \cos(\Delta m \Delta t) \cdot d\Delta t \right].$$

The integral of an even function in a range $x \in [-a, a]$ is twice the integral of the same function in the range $x \in [0, a]$. Taking this into account, one obtains

$$\int_{-\infty}^{\infty} \mathcal{P}(\Delta t, q) \cdot d\Delta t = \frac{2}{4\tau_{B^0}} \left[\int_0^{\infty} e^{-\frac{\Delta t}{\tau_{B^0}}} \cdot d\Delta t + q\mathcal{A}_{CP} \int_0^{\infty} e^{-\frac{\Delta t}{\tau_{B^0}}} \cos(\Delta m \Delta t) \cdot d\Delta t \right].$$

Using partial integration (or looking into a mathematical compendium like [165]), the integration gives

$$\begin{aligned} \int_{-\infty}^{\infty} \mathcal{P}(\Delta t, q) \cdot d\Delta t &= \frac{1}{2\tau_{B^0}} \left[-\tau_{B^0} e^{-\frac{\Delta t}{\tau_{B^0}}} \right]_0^{\infty} \\ &\quad + q\mathcal{A}_{CP} \frac{e^{-\frac{\Delta t}{\tau_{B^0}}}}{1 + (\tau_{B^0} \Delta m)^2} \left[-\tau_{B^0} \cos(\Delta m \Delta t) + \tau_{B^0}^2 \Delta m \sin(\Delta m \Delta t) \right] \Big|_0^{\infty} \\ &= \frac{1}{2} \left[1 + q\mathcal{A}_{CP} \frac{1}{1 + (\tau_{B^0} \Delta m)^2} \right] = \mathcal{P}(q), \end{aligned}$$

where

$$\frac{1}{1 + (\tau_{B^0} \Delta m)^2} = 1 - 2 \cdot \chi_d.$$

A.3 Probability densities for t_{sig} and t_{tag}

Starting from eq. (A.1) and using the definitions $t_1 = t_{\text{sig}}$, $t_2 = t_{\text{tag}}$, one can obtain the time-dependent decay rate

$$\begin{aligned} \Gamma(t_{\text{sig}}, t_{\text{tag}}, q) &= |\langle f_{\text{sig}}, f_{\text{tag}} | H | B_{\text{sig}}(t_{\text{sig}}), B_{\text{tag}}(t_{\text{tag}}) \rangle|^2 \\ &= \frac{e^{-\Gamma(t_{\text{sig}}+t_{\text{tag}})}}{4 \cdot (1 + |\lambda_{CP}|^2)^{-1}} \cdot |A_{\text{sig}}|^2 \cdot |A_{\text{tag}}|^2 \\ &\quad \cdot \left[1 + q \left[\mathcal{A}_{CP} \cos(\Delta m \cdot (t_{\text{sig}} - t_{\text{tag}})) + \mathcal{S}_{CP} \sin(\Delta m \cdot (t_{\text{sig}} - t_{\text{tag}})) \right] \right] \end{aligned} \tag{A.2}$$

$$= \Gamma \cdot \frac{e^{-\Gamma t_{\text{tag}}}}{2} \left[1 + \frac{q \mathcal{A}_{CP}}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left[\frac{\Delta m}{\Gamma} \sin(\Delta m t_{\text{tag}}) + \cos(\Delta m t_{\text{tag}}) \right] + \frac{q \mathcal{S}_{CP}}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left[\frac{\Delta m}{\Gamma} \cos(\Delta m t_{\text{tag}}) - \sin(\Delta m t_{\text{tag}}) \right] \right].$$

In the same way, one can obtain $\mathcal{P}(t_{\text{sig}}, q)$ by integrating eq. (A.2) over the full t_{tag} range and normalizing the result. This leads to

$$\mathcal{P}(t_{\text{sig}}, q) = \Gamma \cdot \frac{e^{-\Gamma t_{\text{sig}}}}{2} \left[1 + \frac{q \mathcal{A}_{CP}}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left[\frac{\Delta m}{\Gamma} \sin(\Delta m t_{\text{sig}}) + \cos(\Delta m t_{\text{sig}}) \right] + \frac{q \mathcal{S}_{CP}}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left[\sin(\Delta m t_{\text{sig}}) - \frac{\Delta m}{\Gamma} \cos(\Delta m t_{\text{sig}}) \right] \right].$$

In Sect. 4.7, the t_{sig} and the t_{tag} distributions were shown for $\Delta t > 0$ and for $\Delta t < 0$, where $\Delta t = t_{\text{sig}} - t_{\text{tag}}$. To obtain for example the PDF describing the t_{tag} distribution for $\Delta t \leq 0$, one needs to integrate eq. (A.2) over the range $t_{\text{sig}} \leq t_{\text{tag}}$, and then normalize the result. This means

$$F(t_{\text{tag}}, q | \Delta t \leq 0) = \int_0^{t_{\text{tag}}} \Gamma(t_{\text{sig}}, t_{\text{tag}}, q) \cdot dt_{\text{sig}}$$

$$\Rightarrow \mathcal{P}(t_{\text{tag}}, q | \Delta t \leq 0) = \frac{F(t_{\text{tag}}, q | \Delta t \leq 0)}{\int_0^\infty \left[\sum_q F(t_{\text{tag}}, q | \Delta t \leq 0) \right] dt_{\text{tag}}}.$$

This yields

$$\mathcal{P}(t_{\text{tag}}, q | \Delta t \leq 0) = \Gamma \left[e^{-\Gamma t_{\text{tag}}} \left[1 + \frac{q}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left\{ \mathcal{A}_{CP} \left(\frac{\Delta m}{\Gamma} \sin \Delta m t_{\text{tag}} + \cos \Delta m t_{\text{tag}} \right) + \mathcal{S}_{CP} \left(\frac{\Delta m}{\Gamma} \cos \Delta m t_{\text{tag}} - \sin \Delta m t_{\text{tag}} \right) \right\} \right] - e^{-2\Gamma t_{\text{tag}}} \left[1 + \frac{q}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left\{ \mathcal{A}_{CP} + \mathcal{S}_{CP} \frac{\Delta m}{\Gamma} \right\} \right] \right].$$

For the PDF describing the t_{tag} distribution for $\Delta t \geq 0$, one repeats the procedure, but integrating over the range $t_{\text{sig}} \geq t_{\text{tag}}$. The result is

$$\mathcal{P}(t_{\text{tag}}, q | \Delta t \geq 0) = \Gamma e^{-2\Gamma t_{\text{tag}}} \left[1 + \frac{q}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left\{ \mathcal{A}_{CP} + \mathcal{S}_{CP} \frac{\Delta m}{\Gamma} \right\} \right].$$

For the PDFs describing the t_{sig} distributions for $\Delta t > 0$ and for $\Delta t < 0$, the procedure is similar. One obtains,

$$\begin{aligned} \mathcal{P}(t_{\text{sig}}, q | \Delta t \leq 0) &= \Gamma e^{-2\Gamma t_{\text{sig}}} \left[1 + \frac{q}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left\{ \mathcal{A}_{CP} - \mathcal{S}_{CP} \frac{\Delta m}{\Gamma} \right\} \right], \\ \mathcal{P}(t_{\text{sig}}, q | \Delta t \geq 0) &= \Gamma \left[e^{-\Gamma t_{\text{sig}}} \left[1 + \frac{q}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left\{ \mathcal{A}_{CP} \left(\frac{\Delta m}{\Gamma} \sin \Delta m t_{\text{sig}} + \cos \Delta m t_{\text{sig}} \right) \right. \right. \right. \\ &\quad \left. \left. \left. + \mathcal{S}_{CP} \left(\sin \Delta m t_{\text{sig}} - \frac{\Delta m}{\Gamma} \cos \Delta m t_{\text{sig}} \right) \right\} \right] \right. \\ &\quad \left. - e^{-2\Gamma t_{\text{sig}}} \left[1 + \frac{q}{1 + \left(\frac{\Delta m}{\Gamma}\right)^2} \left\{ \mathcal{A}_{CP} - \mathcal{S}_{CP} \frac{\Delta m}{\Gamma} \right\} \right] \right]. \end{aligned}$$

A.4 Derivation of the triangular isospin relations for the decays $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$

For the derivation of the possible isospin configurations of an isovector pair $h^i h^j$, where $h \in \{\pi, \rho\}$ and $i, j, i + j \in \{-, 0, +\}$, one should keep in mind a few basic concepts. First, Isospin symmetry is approximate: the quarks u and d are considered as a spin- $\frac{1}{2}$ system ($SU(2)$ doublet) disregarding the mass and the electric charge differences between them. Using the notation $|I, m_I\rangle$, where I and $m_I = -I, -I + 1, \dots, I$ are respectively the isospin and the secondary isospin quantum numbers (the latter corresponding to the spin z -projection), the isospin quark eigenstates can be written as

$$\begin{aligned} |u\rangle &= \left| \frac{1}{2}, +\frac{1}{2} \right\rangle & |\bar{u}\rangle &= \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \\ |d\rangle &= \left| \frac{1}{2}, -\frac{1}{2} \right\rangle & |\bar{d}\rangle &= -\left| \frac{1}{2}, +\frac{1}{2} \right\rangle. \end{aligned}$$

One should briefly remember the spin addition rules. For a system of two particles with spin I_1 and spin I_2 , the total spin operator is $\mathbf{I} = \mathbf{I}_1 + \mathbf{I}_2$. The possible total spin quantum numbers are

$$I = |I_1 - I_2|, |I_1 - I_2 + 1|, \dots, I_1 + I_2$$

The main task in the addition of angular momentum is a change of basis, from one which corresponds to the eigenstates of $\mathbf{I}_1^2, \mathbf{I}_2^2, \mathbf{I}_{z,1}, \mathbf{I}_{z,2}$ to a basis which corresponds to the eigenstates of $\mathbf{I}^2, \mathbf{I}_z, \mathbf{I}_1^2, \mathbf{I}_2^2$. Symbolically one writes,

$$I_1 \otimes I_2 = (I_1 + I_2) \oplus (I_1 + I_2 - 1) \oplus \dots \oplus |I_1 - I_2|.$$

To be noted is that the symbolic equation above becomes a correct equation if one replaces the quantum numbers for each state with the multiplicities $2I + 1$ and removes the circles around the addition and multiplication symbols.

Considering now the eigenstates with $I = I_1 + I_2$, the top and the bottom eigenstates of I are

$$\begin{aligned} |I, m_I^{\max}\rangle &= |I_1 + I_2, I_1 + I_2\rangle = |I_1, I_1\rangle |I_2, I_2\rangle, \\ |I, m_I^{\min}\rangle &= |I_1 + I_2, -(I_1 + I_2)\rangle = |I_1, -I_1\rangle |I_2, -I_2\rangle. \end{aligned}$$

The eigenstates in-between can be obtained by applying the lowering (raising) operator

$$\mathbf{I}_{\pm} = \mathbf{I}_{1,\pm} + \mathbf{I}_{2,\pm}, \quad \mathbf{I}_{\pm} |I, m_I\rangle = \hbar \sqrt{(I \mp m_I)(I \pm m_I + 1)} |I, m_I \pm 1\rangle.$$

onto the top (bottom) eigenstates. The top eigenstate with $I = I_1 + I_2 - 1$ can be obtained by combining all possible kets with $m_I = I_1 + I_2 - 1$ while requiring normalization and orthogonality to the state $|I_1 + I_2, I_1 + I_2 - 1\rangle$. The lower eigenstates can be obtained again by lowering. The same result is obtained starting from the bottom eigenstate and then raising. This procedure is repeated until one arrives at $I = |I_1 - I_2|$, whose top eigenstate has to be orthogonal to all the previous states with $I_z = |I_1 - I_2|$. In general, one can write

$$|I, m_I\rangle = \sum_{m_{I,1}=-I_1}^{I_1} \sum_{m_{I,2}=-I_2}^{I_2} |I_1, m_{I,1}\rangle |I_2, m_{I,2}\rangle \langle I_1, m_{I,1} | \langle I_2, m_{I,2} | I, m_I\rangle,$$

where the so-called Clebsch-Gordan coefficients $\langle I_1, m_{I,1} | \langle I_2, m_{I,2} | I, m_I\rangle$ can be found in tables for the well-known common cases (e.g. in [28]) or calculated with a general formula derived by G. Racah (See e.g. in [166]).

For a system composed of two isospin- $\frac{1}{2}$ particles, one obtains the following $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$ eigenstates:

$$|1, +1\rangle = \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = -|u\bar{d}\rangle, \quad (\text{A.3})$$

$$|1, 0\rangle = \frac{1}{\sqrt{2}} \left(\left| \frac{1}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \right) = \frac{1}{\sqrt{2}} (|u\bar{u}\rangle - |d\bar{d}\rangle), \quad (\text{A.4})$$

$$|1, -1\rangle = \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = |\bar{u}d\rangle, \quad (\text{A.5})$$

which form an $SU(2)$ triplet corresponding to isovector (isospin-1) mesons like pions ($J = 0$) and ρ mesons ($J = 1$); and

$$|0, 0\rangle = \frac{1}{\sqrt{2}} \left(\left| \frac{1}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle - \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \right) = \frac{1}{\sqrt{2}} (|u\bar{u}\rangle + |d\bar{d}\rangle), \quad (\text{A.6})$$

corresponding to an isosinglet state, which, together with a $|s\bar{s}\rangle$ component, forms the isospin-0 mesons like the η and the η' mesons ($J = 0$), and the ω and the ϕ mesons ($J = 1$). From eqs. (A.4) and (A.6) one obtains also the following useful equations

$$\left| \frac{1}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} (|1, 0\rangle + |0, 0\rangle), \quad (\text{A.7})$$

$$\left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} (|1, 0\rangle - |0, 0\rangle). \quad (\text{A.8})$$

For a system composed of two isospin-1 particles, the possible $1 \otimes 1 = 2 \oplus 1 \oplus 0$ eigenstates are

$$|2, +2\rangle = |1, +1\rangle|1, +1\rangle, \quad (\text{A.9})$$

$$|2, +1\rangle = \frac{1}{\sqrt{2}} (|1, +1\rangle|1, 0\rangle + |1, 0\rangle|1, +1\rangle), \quad (\text{A.10})$$

$$|2, 0\rangle = \frac{1}{\sqrt{6}} (|1, +1\rangle|1, -1\rangle + |1, -1\rangle|1, +1\rangle + 2|1, 0\rangle|1, 0\rangle), \quad (\text{A.11})$$

$$|2, -1\rangle = \frac{1}{\sqrt{2}} (|1, -1\rangle|1, 0\rangle + |1, 0\rangle|1, -1\rangle), \quad (\text{A.12})$$

$$|2, -2\rangle = |1, -1\rangle|1, -1\rangle, \quad (\text{A.13})$$

$$|1, +1\rangle = \frac{1}{\sqrt{2}} (|1, +1\rangle|1, 0\rangle - |1, 0\rangle|1, +1\rangle), \quad (\text{A.14})$$

$$|1, 0\rangle = \frac{1}{\sqrt{2}} (|1, +1\rangle|1, -1\rangle - |1, -1\rangle|1, +1\rangle), \quad (\text{A.15})$$

$$|1, -1\rangle = \frac{1}{\sqrt{2}} (|1, 0\rangle|1, -1\rangle - |1, -1\rangle|1, 0\rangle), \quad (\text{A.16})$$

$$|0, 0\rangle = \frac{1}{\sqrt{3}} (|1, +1\rangle|1, -1\rangle + |1, -1\rangle|1, +1\rangle - |1, 0\rangle|1, 0\rangle). \quad (\text{A.17})$$

From eqs. (A.11) and (A.17), one obtains the useful equations

$$\frac{1}{\sqrt{2}} (|1, +1\rangle|1, -1\rangle + |1, -1\rangle|1, +1\rangle) = \frac{1}{\sqrt{3}}|2, 0\rangle + \sqrt{\frac{2}{3}}|0, 0\rangle, \quad (\text{A.18})$$

$$|1, 0\rangle|1, 0\rangle = \sqrt{\frac{2}{3}}|2, 0\rangle - \frac{1}{\sqrt{3}}|0, 0\rangle. \quad (\text{A.19})$$

For the final states of the decays $B^{i+j} \rightarrow h^i h^j$, which have to be symmetric under particle exchange, it follows

$$|h^+ h^0\rangle = \frac{1}{\sqrt{2}} (|h_1^+ h_2^0\rangle + |h_1^0 h_2^+\rangle) = \frac{1}{\sqrt{2}} (|1, +1\rangle|1, 0\rangle + |1, 0\rangle|1, +1\rangle) = |2, +1\rangle, \quad (\text{A.20})$$

$$\begin{aligned} |h^+h^- \rangle &= \frac{1}{\sqrt{2}} (|h_1^+h_2^- \rangle + |h_1^-h_2^+ \rangle) = \frac{1}{\sqrt{2}} (|1, +1 \rangle|1, -1 \rangle + |1, -1 \rangle|1, +1 \rangle) \\ &= \frac{1}{\sqrt{3}} |2, 0 \rangle + \sqrt{\frac{2}{3}} |0, 0 \rangle, \end{aligned} \quad (\text{A.21})$$

$$|h^-h^0 \rangle = \frac{1}{\sqrt{2}} (|h_1^-h_2^0 \rangle + |h_1^0h_2^- \rangle) = \frac{1}{\sqrt{2}} (|1, -1 \rangle|1, 0 \rangle + |1, 0 \rangle|1, -1 \rangle) = |2, -1 \rangle, \quad (\text{A.22})$$

$$|h^0h^0 \rangle = |h^0h^0 \rangle = |1, 0 \rangle|1, 0 \rangle = \sqrt{\frac{2}{3}} |2, 0 \rangle - \frac{1}{\sqrt{3}} |0, 0 \rangle. \quad (\text{A.23})$$

Considering now the effective Hamiltonian \mathcal{H} for the transitions $\bar{b} \rightarrow \bar{u}u\bar{d}$ and $b \rightarrow u\bar{u}d$, the final states can have $I = |\frac{1}{2} + \frac{1}{2} - \frac{1}{2}|, (\frac{1}{2} + \frac{1}{2} + \frac{1}{2}) = \frac{1}{2}, \frac{3}{2}$. The z -components of the final states are $m = -\frac{1}{2} + \frac{1}{2} + \frac{1}{2} = +\frac{1}{2}$ for $\bar{u}u\bar{d}$, and $m = +\frac{1}{2} - \frac{1}{2} - \frac{1}{2} = -\frac{1}{2}$ for $u\bar{u}d$. In terms of the amplitudes $A_{\Delta I, I_f}$, where ΔI denotes the isospin shift and I_f the final-state isospin, the Hamiltonian operators can be written as

$$\bar{b} \rightarrow \bar{u}u\bar{d} \quad \rightarrow \quad \mathcal{H} = A_{\frac{3}{2}} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle + A_{\frac{1}{2}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle, \quad (\text{A.24})$$

$$b \rightarrow u\bar{u}d \quad \rightarrow \quad \mathcal{H} = \bar{A}_{\frac{3}{2}} \left| \frac{3}{2}, -\frac{1}{2} \right\rangle + \bar{A}_{\frac{1}{2}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle. \quad (\text{A.25})$$

To calculate the decay amplitudes $A^{ij} \equiv \langle h^i h^j | \mathcal{H} | B^{i+j} \rangle$ (eq. (2.46)), the basis of the terms $\mathcal{H} | B^{i+j} \rangle$ needs to be changed,

$$\mathcal{H} | B^+ \rangle = \mathcal{H} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = A_{\frac{3}{2}} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle + A_{\frac{1}{2}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle, \quad (\text{A.26})$$

$$\mathcal{H} | B^0 \rangle = \mathcal{H} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = A_{\frac{3}{2}} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + A_{\frac{1}{2}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle, \quad (\text{A.27})$$

$$\mathcal{H} | B^- \rangle = \mathcal{H} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \bar{A}_{\frac{3}{2}} \left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \bar{A}_{\frac{1}{2}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle, \quad (\text{A.28})$$

$$\mathcal{H} | \bar{B}^0 \rangle = \mathcal{H} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \bar{A}_{\frac{3}{2}} \left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle + \bar{A}_{\frac{1}{2}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle. \quad (\text{A.29})$$

The possible $\frac{3}{2} \otimes \frac{1}{2} = 2 \oplus 1$ eigenstates are

$$|2, +2 \rangle = \left| \frac{3}{2}, +\frac{3}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \quad (\text{A.30})$$

$$|2, +1 \rangle = \sqrt{\frac{3}{4}} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle + \frac{1}{2} \left| \frac{3}{2}, +\frac{3}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \quad (\text{A.31})$$

$$|2, 0 \rangle = \frac{1}{\sqrt{2}} \left(\left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle + \left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \right) \quad (\text{A.32})$$

$$|2, -1\rangle = \sqrt{\frac{3}{4}} \left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle + \frac{1}{2} \left| \frac{3}{2}, -\frac{3}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \quad (\text{A.33})$$

$$|2, -2\rangle = \left| \frac{3}{2}, -\frac{3}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle \quad (\text{A.34})$$

$$|1, +1\rangle = \sqrt{\frac{3}{4}} \left| \frac{3}{2}, +\frac{3}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle - \frac{1}{2} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \quad (\text{A.35})$$

$$|1, 0\rangle = \frac{1}{\sqrt{2}} \left(\left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle - \left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \right) \quad (\text{A.36})$$

$$|1, -1\rangle = \frac{1}{2} \left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle - \sqrt{\frac{3}{4}} \left| \frac{3}{2}, -\frac{3}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle \quad (\text{A.37})$$

$$(\text{A.38})$$

From the equations above, one obtains the useful equations

$$\left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \frac{\sqrt{3}}{2} |2, +1\rangle - \frac{1}{2} |1, +1\rangle, \quad (\text{A.39})$$

$$\left| \frac{3}{2}, +\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} (|2, 0\rangle + |1, 0\rangle), \quad (\text{A.40})$$

$$\left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \frac{\sqrt{3}}{2} |2, -1\rangle + \frac{1}{2} |1, -1\rangle, \quad (\text{A.41})$$

$$\left| \frac{3}{2}, -\frac{1}{2} \right\rangle \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \frac{1}{\sqrt{2}} (|2, 0\rangle - |1, 0\rangle). \quad (\text{A.42})$$

Using eqs. (A.39) to (A.42) together with eqs. (A.3), (A.5), (A.7) and (A.8), the terms in eqs. (A.26) to (A.29) give

$$\mathcal{H}|B^+\rangle = \mathcal{H} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \sqrt{\frac{3}{4}} A_{\frac{3}{2}} |2, 1\rangle + (A_{\frac{1}{2}} - \frac{1}{2} A_{\frac{3}{2}}) |1, 1\rangle, \quad (\text{A.43})$$

$$\mathcal{H}|B^0\rangle = \mathcal{H} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} A_{\frac{3}{2}} |2, 0\rangle + \sqrt{\frac{1}{2}} (A_{\frac{1}{2}} + A_{\frac{3}{2}}) |1, 0\rangle + \sqrt{\frac{1}{2}} A_{\frac{1}{2}} |0, 0\rangle, \quad (\text{A.44})$$

$$\mathcal{H}|B^-\rangle = \mathcal{H} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{3}{4}} \bar{A}_{\frac{3}{2}} |2, -1\rangle + (\bar{A}_{\frac{1}{2}} + \frac{1}{2} \bar{A}_{\frac{3}{2}}) |1, -1\rangle, \quad (\text{A.45})$$

$$\mathcal{H}|\bar{B}^0\rangle = \mathcal{H} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} \bar{A}_{\frac{3}{2}} |2, 0\rangle + \sqrt{\frac{1}{2}} (\bar{A}_{\frac{1}{2}} - \bar{A}_{\frac{3}{2}}) |1, 0\rangle - \sqrt{\frac{1}{2}} \bar{A}_{\frac{1}{2}} |0, 0\rangle. \quad (\text{A.46})$$

Projection of the states $\mathcal{H}|B^{i+j}\rangle$ onto the final states in eqs. (A.20) to (A.23) gives,

$$A^{+0} = \langle h^+ h^0 | \mathcal{H} | B^+ \rangle = \sqrt{\frac{3}{4}} A_{\frac{3}{2}}, \quad (\text{A.47})$$

$$A^{+-} = \langle h^+ h^- | \mathcal{H} | B^0 \rangle = \sqrt{\frac{1}{6}} A_{\frac{3}{2}} + \sqrt{\frac{1}{3}} A_{\frac{1}{2}}, \quad (\text{A.48})$$

$$A^{00} = \langle h^0 h^0 | \mathcal{H} | B^0 \rangle = \sqrt{\frac{1}{3}} A_{\frac{3}{2}} - \sqrt{\frac{1}{6}} A_{\frac{1}{2}}, \quad (\text{A.49})$$

$$\bar{A}^{+0} = A^{-0} = \langle h^- h^0 | \mathcal{H} | B^- \rangle = \sqrt{\frac{3}{4}} \bar{A}_{\frac{3}{2}}, \quad (\text{A.50})$$

$$\bar{A}^{+-} = \langle h^+ h^- | \mathcal{H} | \bar{B}^0 \rangle = \sqrt{\frac{1}{6}} \bar{A}_{\frac{3}{2}} - \sqrt{\frac{1}{3}} \bar{A}_{\frac{1}{2}}, \quad (\text{A.51})$$

$$\bar{A}^{00} = \langle h^0 h^0 | \mathcal{H} | \bar{B}^0 \rangle = \sqrt{\frac{1}{3}} \bar{A}_{\frac{3}{2}} + \sqrt{\frac{1}{6}} \bar{A}_{\frac{1}{2}}. \quad (\text{A.52})$$

From eqs. (A.47) to (A.49) and eqs. (A.50) to (A.52), one obtains finally the two isospin relations called isospin triangles,

$$\begin{aligned} A^{+0} - A^{00} &= \sqrt{\frac{1}{2}} A^{+-}, \\ \bar{A}^{+0} - \bar{A}^{00} &= \sqrt{\frac{1}{2}} \bar{A}^{+-}. \end{aligned} \quad (\text{A.53})$$

B The Belle II flavor tagger

B.1 Probability density for CP violation analyses considering flavor tagging

To obtain the probabilities $\mathcal{P}(\Delta t, q)^{\text{obs}}$ and $\mathcal{P}(q)^{\text{obs}}$ for CP violation analysis considering the performance of a flavor tagging algorithm, one expresses first the efficiencies ε_{B^0} and $\varepsilon_{\bar{B}^0}$, and the wrong tag fractions w_{B^0} and $w_{\bar{B}^0}$ in terms of ε , $\Delta\varepsilon$, w and Δw . From eqs. (4.5) and (4.6), one obtains

$$\begin{aligned}\varepsilon_{B^0} &= \varepsilon + \frac{\Delta\varepsilon}{2}, & \varepsilon_{\bar{B}^0} &= \varepsilon - \frac{\Delta\varepsilon}{2}, \\ w_{B^0} &= w + \frac{\Delta w}{2}, & w_{\bar{B}^0} &= w - \frac{\Delta w}{2}.\end{aligned}\tag{B.1}$$

In terms of these quantities, the probability $\mathcal{P}^{\text{obs}}(\Delta t, q)$ for $q = \pm 1$ is given by

$$\begin{aligned}\mathcal{P}^{\text{obs}}(\Delta t, q = +1) &= \varepsilon_{B^0}(1 - w_{B^0}) \cdot \mathcal{P}(\Delta t, q = +1) + \varepsilon_{\bar{B}^0}w_{\bar{B}^0} \cdot \mathcal{P}(\Delta t, q = -1), \\ \mathcal{P}^{\text{obs}}(\Delta t, q = -1) &= \varepsilon_{\bar{B}^0}(1 - w_{\bar{B}^0}) \cdot \mathcal{P}(\Delta t, q = -1) + \varepsilon_{B^0}w_{B^0} \cdot \mathcal{P}(\Delta t, q = +1).\end{aligned}$$

Insertion of $\mathcal{P}(\Delta t, q)$ (eq. (2.33)) for $q = \pm 1$ yields,

$$\begin{aligned}\mathcal{P}^{\text{obs}}(\Delta t, q = +1) &= \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \left[\varepsilon_{B^0}(1 - w_{B^0}) + \varepsilon_{\bar{B}^0}w_{\bar{B}^0} + \left[\varepsilon_{B^0}(1 - w_{B^0}) - \varepsilon_{\bar{B}^0}w_{\bar{B}^0} \right] a_{CP}(\Delta t) \right] \\ \mathcal{P}^{\text{obs}}(\Delta t, q = -1) &= \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \left[\varepsilon_{\bar{B}^0}(1 - w_{\bar{B}^0}) + \varepsilon_{B^0}w_{B^0} - \left[\varepsilon_{\bar{B}^0}(1 - w_{\bar{B}^0}) - \varepsilon_{B^0}w_{B^0} \right] a_{CP}(\Delta t) \right]\end{aligned}\tag{B.2}$$

where $a_{CP}(\Delta t) = \mathcal{A}_{CP} \cos(\Delta m \Delta t) + \mathcal{S}_{CP} \sin(\Delta m \Delta t)$ (eq. (2.32)). Using eq. (B.1), the terms in the equations above give

$$\varepsilon_{B^0}(1 - w_{B^0}) + \varepsilon_{\bar{B}^0}w_{\bar{B}^0} = \varepsilon \left[1 - \Delta w + \frac{\Delta\varepsilon}{2\varepsilon}(1 - 2w) \right],\tag{B.3}$$

$$\varepsilon_{B^0}(1 - w_{B^0}) - \varepsilon_{\bar{B}^0}w_{\bar{B}^0} = \varepsilon \left[1 - 2w + \frac{\Delta\varepsilon}{2\varepsilon}(1 - \Delta w) \right],\tag{B.4}$$

$$\varepsilon_{\overline{B}^0}(1 - w_{\overline{B}^0}) + \varepsilon_{B^0}w_{B^0} = \varepsilon \left[1 + \Delta w - \frac{\Delta\varepsilon}{2\varepsilon}(1 - 2w) \right], \quad (\text{B.5})$$

$$\varepsilon_{\overline{B}^0}(1 - w_{\overline{B}^0}) - \varepsilon_{B^0}w_{B^0} = \varepsilon \left[1 - 2w - \frac{\Delta\varepsilon}{2\varepsilon}(1 - \Delta w) \right]. \quad (\text{B.6})$$

The probabilities in eq. (B.2) can be rewritten using eqs. (B.3) to (B.6) and defining $\mu = \Delta\varepsilon/(2\varepsilon)$. One obtains,

$$\mathcal{P}^{\text{obs}}(\Delta t, q = +1) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \varepsilon \left[1 - \Delta w + \mu(1 - 2w) + [(1 - 2w) + \mu(1 - \Delta w)] a_{CP}(\Delta t) \right]$$

$$\mathcal{P}^{\text{obs}}(\Delta t, q = -1) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \varepsilon \left[1 + \Delta w - \mu(1 - 2w) + [-(1 - 2w) + \mu(1 + \Delta w)] a_{CP}(\Delta t) \right],$$

leading ultimately to

$$\mathcal{P}^{\text{obs}}(\Delta t, q) = \frac{e^{-\frac{|\Delta t|}{\tau_{B^0}}}}{4\tau_{B^0}} \varepsilon \left[1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) \right. \\ \left. + [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot a_{CP}(\Delta t) \right].$$

For time-integrated CP violation analyses, one integrates $\int_{-\infty}^{\infty} \mathcal{P}^{\text{obs}}(\Delta t, q) \cdot d\Delta t$ as in App. A.2. The integration gives,

$$\mathcal{P}^{\text{obs}}(q) = \frac{1}{2} \varepsilon \left[1 - q \cdot \Delta w + q \cdot \mu \cdot (1 - 2w) \right. \\ \left. + [q \cdot (1 - 2w) + \mu \cdot (1 - q \cdot \Delta w)] \cdot \mathcal{A}_{CP} \frac{1}{1 + (\tau_{B^0} \Delta m)^2} \right],$$

where

$$\frac{1}{1 + (\tau_{B^0} \Delta m)^2} = 1 - 2 \cdot \chi_d.$$

B.2 Calculation of statistical uncertainties

The statistical uncertainties of the effective efficiencies and of the different quantities characterizing the performance of the flavor tagger are calculated via Gaussian error propagation. For an observable $f(x_1, x_2, \dots, x_n)$, depending on n variables x_1, x_2, \dots, x_n , the standard deviation δf in case that the variables are independent is

$$\delta f = \sqrt{\sum_i^n \left(\frac{\partial f}{\partial x_i} \cdot \delta x_i \right)^2},$$

where δx_i represents the standard deviation of the i -th variable x_i . Using this formula, the expressions used to calculate the statistical uncertainties are obtained.

Using MC information, one can determine in which events B_{tag}^0 is a B^0 and in which events B_{tag}^0 is a \bar{B}^0 . For events in which B_{tag}^0 is a B^0 , the effective efficiency $\varepsilon_{\text{eff}}(B^0)$, in terms of the effective efficiencies $\varepsilon_{\text{eff}}^i(B^0)$, is defined as

$$\varepsilon_{\text{eff}}(B^0) = \sum_i \varepsilon_{\text{eff}}^i(B^0) = \sum_i \varepsilon_{B^0}^i (r_{\text{MC},B^0})_i^2, \quad \text{using} \quad \varepsilon_{B^0}^i = \frac{n_{B^0}^i}{N_{B^0}},$$

where the sum extends over all r bins. The total number of events N_{B^0} in which B_{tag}^0 is a B^0 can be written as

$$N_{B^0} = \sum_j n_{B^0}^j + n_{B^0}^{\text{untagged}},$$

where the sum extends over all r bins, and $n_{B^0}^{\text{untagged}}$ is the number of events for which a flavor tag is not possible. Using this expression for N_{B^0} , one can rewrite the efficiencies $\varepsilon_{B^0}^i$ as

$$\varepsilon_{B^0}^i = \frac{n_{B^0}^i}{\sum_{j \neq i} n_{B^0}^j + n_{B^0}^i + n_{B^0}^{\text{untagged}}}.$$

Starting with the standard deviation of $\varepsilon_{B^0}^i$, one needs first to calculate the partial derivatives with respect to $n_{B^0}^i$, to $n_{B^0}^j$ for $i \neq j$, and to $n_{B^0}^{\text{untagged}}$. This gives,

$$\frac{\partial \varepsilon_{B^0}^i}{\partial n_{B^0}^i} = \left(\frac{1}{\sum_{j \neq i} n_{B^0}^j + n_{B^0}^i + n_{B^0}^{\text{untagged}}} + \frac{-n_{B^0}^i}{\left(\sum_{j \neq i} n_{B^0}^j + n_{B^0}^i + n_{B^0}^{\text{untagged}} \right)^2} \right) = \frac{N_{B^0} - n_{B^0}^i}{N_{B^0}^2},$$

$$\left. \frac{\partial \varepsilon_{B^0}^i}{\partial n_{B^0}^j} \right|_{i \neq j} = \frac{-n_{B^0}^i}{\left(\sum_{j \neq i} n_{B^0}^j + n_{B^0}^i + n_{B^0}^{\text{untagged}} \right)^2} = \frac{-n_{B^0}^i}{N_{B^0}^2} = \frac{\partial \varepsilon_{B^0}^i}{\partial n_{B^0}^{\text{untagged}}}.$$

Using for the standard deviations of the numbers $\delta n_{B^0}^i = \sqrt{n_{B^0}^i}$, and for $\delta n_{B^0}^{\text{untagged}} = \sqrt{n_{B^0}^{\text{untagged}}}$, the standard deviation of $\varepsilon_{B^0}^i$ is approximated as

$$\delta \varepsilon_{B^0}^i = \sqrt{\left(\frac{N_{B^0} - n_{B^0}^i}{N_{B^0}^2} \right)^2 \cdot n_{B^0}^i + \left(\frac{-n_{B^0}^i}{N_{B^0}^2} \right)^2 \cdot \left(\sum_{j \neq i} n_{B^0}^j + n_{B^0}^{\text{untagged}} \right)}$$

$$\begin{aligned}
&= \sqrt{\left(\frac{N_{B^0} - n_{B^0}^i}{N_{B^0}^2}\right)^2 \cdot n_{B^0}^i + \left(\frac{n_{B^0}^i}{N_{B^0}^2}\right)^2 \cdot (N_{B^0} - n_{B^0}^i)} \\
&= \frac{1}{N_{B^0}^2} \cdot \sqrt{(N_{B^0} - n_{B^0}^i) \cdot n_{B^0}^i \cdot N_{B^0}}.
\end{aligned}$$

The standard deviation of $\delta\varepsilon_{\overline{B}^0}^i$ is obtained by replacing B^0 with \overline{B}^0 in all the subscripts of the equation above. For the differences $\Delta\varepsilon_i = \varepsilon_{B^0}^i - \varepsilon_{\overline{B}^0}^i$ and the averages $\varepsilon_i = (\varepsilon_{B^0}^i + \varepsilon_{\overline{B}^0}^i)/2$ one obtains

$$\delta\Delta\varepsilon_i = \sqrt{(\delta\varepsilon_{B^0}^i)^2 + (\delta\varepsilon_{\overline{B}^0}^i)^2}, \quad \delta\varepsilon_i = \frac{1}{2}\delta\Delta\varepsilon_i.$$

And for the parameters $\mu_i = \Delta\varepsilon_i/(2\varepsilon_i)$ one obtains

$$\begin{aligned}
\delta\mu_i &= \sqrt{\left(\frac{1}{2\varepsilon_i}\right)^2 \cdot (\delta\Delta\varepsilon_i)^2 + \left(\frac{-\Delta\varepsilon_i}{2\varepsilon_i^2}\right)^2 \cdot (\delta\varepsilon_i)^2} \quad \Bigg| \quad \delta\varepsilon_i = \frac{1}{2}\delta\Delta\varepsilon_i \\
&= \frac{1}{2\varepsilon_i} \cdot \sqrt{(\delta\Delta\varepsilon_i)^2 + \left(\frac{\Delta\varepsilon_i}{\varepsilon_i}\right)^2 \cdot \left(\frac{1}{2}\delta\Delta\varepsilon_i\right)^2} \\
&= \frac{\delta\Delta\varepsilon_i}{2\varepsilon_i} \cdot \sqrt{\mu_i^2 + 1}.
\end{aligned}$$

The standard deviation of the wrong tag fraction $w_{\text{MC},B^0}^i = n_{B^0,\text{wrong}}^i/n_{B^0}^i$ is approximated as

$$\delta w_{\text{MC},B^0}^i = \frac{1}{n_{B^0}^i} \cdot \sqrt{\frac{(n_{B^0}^i - n_{B^0,\text{wrong}}^i) \cdot n_{B^0,\text{wrong}}^i}{n_{B^0}^i}}. \quad (\text{B.7})$$

The same formula above applies in the case of \overline{B}^0 by exchanging the subscripts. For the differences $\Delta w_{\text{MC},i} = w_{\text{MC},B^0}^i - w_{\text{MC},\overline{B}^0}^i$ and the averages $w_{\text{MC},i} = (w_{\text{MC},B^0}^i + w_{\text{MC},\overline{B}^0}^i)/2$ the standard deviations are

$$\delta\Delta w_{\text{MC},i} = \sqrt{(\delta w_{\text{MC},B^0}^i)^2 + (\delta w_{\text{MC},\overline{B}^0}^i)^2}, \quad \delta w_{\text{MC},i} = \frac{1}{2}\delta\Delta w_{\text{MC},i}.$$

For some figures, the asymmetry $A_w = (w_{B^0} - w_{\overline{B}^0})/(w_{B^0} + w_{\overline{B}^0})$ is calculated. Its standard deviation is

$$\delta A_w = \frac{2}{(w_{B^0} + w_{\overline{B}^0})^2} \cdot \sqrt{(w_{\overline{B}^0} \cdot \delta w_{B^0})^2 + (w_{B^0} \cdot \delta w_{\overline{B}^0})^2}.$$

Considering the dilution factor $(r_{\text{MC},B^0})_i = |1 - 2 \cdot w_{\text{MC},B^0}^i|$, its standard deviation is approximated as

$$\delta(r_{\text{MC},B^0})_i = 2 \cdot \delta w_{\text{MC},B^0}^i.$$

Coming back to the effective efficiencies $\varepsilon_{\text{eff}}^i(B^0) = \varepsilon_{B^0}^i \langle r_{B^0} \rangle_i^2$, the corresponding standard deviations are

$$\begin{aligned} \delta\varepsilon_{\text{eff}}^i(B^0) &= \sqrt{\left(2\varepsilon_{B^0}^i\right)^2 \cdot \delta\langle r_{\text{MC},B^0} \rangle_i^2 + \left(r_{\text{MC},B^0}\right)_i^4 \cdot \left(\delta\varepsilon_{B^0}^i\right)^2} \\ &= \sqrt{\left(2\frac{n_{B^0}^i}{N_{B^0}}\left(r_{\text{MC},B^0}\right)_i\right)^2 \cdot \delta\langle r_{\text{MC},B^0} \rangle_i^2 + \left(r_{\text{MC},B^0}\right)_i^4 \cdot \frac{1}{N_{B^0}^4} \cdot \left(N_{B^0} - n_{B^0}^i\right) \cdot n_{B^0}^i \cdot N_{B^0}} \\ &= \frac{\left(r_{\text{MC},B^0}\right)_i}{N_{B^0}^2} \cdot \sqrt{N_{B^0}^2(2n_{B^0}^i)^2 \cdot \left(\delta\langle r_{\text{MC},B^0} \rangle_i\right)^2 + \left(r_{\text{MC},B^0}\right)_i^2 \cdot \left(N_{B^0} - n_{B^0}^i\right) \cdot n_{B^0}^i \cdot N_{B^0}}. \end{aligned}$$

The same formula applies using $\delta\langle r_{B^0} \rangle_i$ in cases where $\left(r_{\text{MC},B^0}\right)_i$ is replaced by $\langle r_{B^0} \rangle_i$. For events where B_{tag}^0 is a \bar{B}^0 , one has to exchange the subscript. For the differences $\Delta\varepsilon_{\text{eff}}^i = \varepsilon_{\text{eff}}^i(B^0) - \varepsilon_{\text{eff}}^i(\bar{B}^0)$ one obtains

$$\delta\Delta\varepsilon_{\text{eff}}^i = \sqrt{\delta\varepsilon_{\text{eff}}^i(B^0)^2 + \delta\varepsilon_{\text{eff}}^i(\bar{B}^0)^2}.$$

For the total effective efficiencies $\varepsilon_{\text{eff}}(B^0)$ and $\varepsilon_{\text{eff}}(\bar{B}^0)$, the standard deviations are

$$\delta\varepsilon_{\text{eff}}(B^0) = \sqrt{\sum_i \delta\varepsilon_{\text{eff}}^i(B^0)^2}, \quad \delta\varepsilon_{\text{eff}}(\bar{B}^0) = \sqrt{\sum_i \delta\varepsilon_{\text{eff}}^i(\bar{B}^0)^2}.$$

For the difference $\Delta\varepsilon_{\text{eff}} = \varepsilon_{\text{eff}}(B^0) - \varepsilon_{\text{eff}}(\bar{B}^0)$ one obtains

$$\delta\Delta\varepsilon_{\text{eff}} = \sqrt{\delta\varepsilon_{\text{eff}}(B^0)^2 + \delta\varepsilon_{\text{eff}}(\bar{B}^0)^2}.$$

For studies on Belle data, the dilution factor is obtained from the mean values $\langle r \rangle_i$. Their standard deviations are

$$\delta\langle r \rangle_i = \frac{\langle (r - \langle r \rangle_i)^2 \rangle_i}{\sqrt{N - 1}} = \frac{\langle r^2 \rangle_i - \langle r \rangle_i^2}{\sqrt{N - 1}}.$$

Since $r_i = |1 - 2w_i|$, the error of the wrong tag fractions w_i from the mean values $\langle r \rangle_i$ is

$$\delta w_i = \frac{\delta\langle r \rangle_i}{2}.$$

For the effective efficiencies $\varepsilon_{\text{eff}}^i$, the standard deviation is

$$\delta\varepsilon_{\text{eff}}^i = \frac{\langle r \rangle_i}{N^2} \cdot \sqrt{N^2(2n_i)^2 \cdot \delta\langle r \rangle_i^2 + \langle r \rangle_i^2 \cdot (N - n_i) \cdot n_i \cdot N},$$

where n_i is the number of events in the i -th bin, and N is the total number of events. For the total average effective efficiency, it follows

$$\delta\varepsilon_{\text{eff}} = \sqrt{\sum_i \delta\varepsilon_{\text{eff}}^i{}^2}.$$

B.3 Input variables for each category

This section presents the MC distributions of all input variables, the correlations between them, and the ranking provided by the FBDT for each category. These are obtained from the Belle II MC sample which was used to train the flavor tagger (S. Sect. 4.5) during its development. This sample corresponds to about 2.6 million $B^0 \bar{B}^0$ events in which B_{sig}^0 decays to $J/\psi (\rightarrow \mu^+ \mu^-) K_S^0 (\rightarrow \pi^+ \pi^-)$ while B_{tag}^0 decays generically according to the known branching fractions. The sample contains only events where the decay channel B_{sig}^0 could be fully reconstructed and correctly matched with the MC decay chain. The signal distributions correspond to the target (or targets) of the respective category, while the background distributions (Bkgr.) correspond to all other tracks on the tag side. For a better visibility, the distributions are normalized such that the areas under the signal and the background distributions are equal to unity.

The correlations R_{ij} between two variables x_i and x_j correspond to

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii} \cdot C_{jj}}},$$

where C is the covariance matrix

$$C_{ij} = \langle (x_i - \mu_i)(x_j - \mu_j) \rangle = \langle x_i x_j \rangle - \mu_i \mu_j,$$

and μ_i, μ_j are the mean values $\langle x_i \rangle, \langle x_j \rangle$, respectively.

The variable ranking is provided by the FBDT algorithm [115]. The FBDT algorithm ranks the input variables by counting how often they are used to split decision tree nodes, and by weighing each split according to the separation gain and the number of events in the node [115, 124, 167]. The ranking is relative: the most important variable has measure 100, and the others are in the range between 0 and 100.

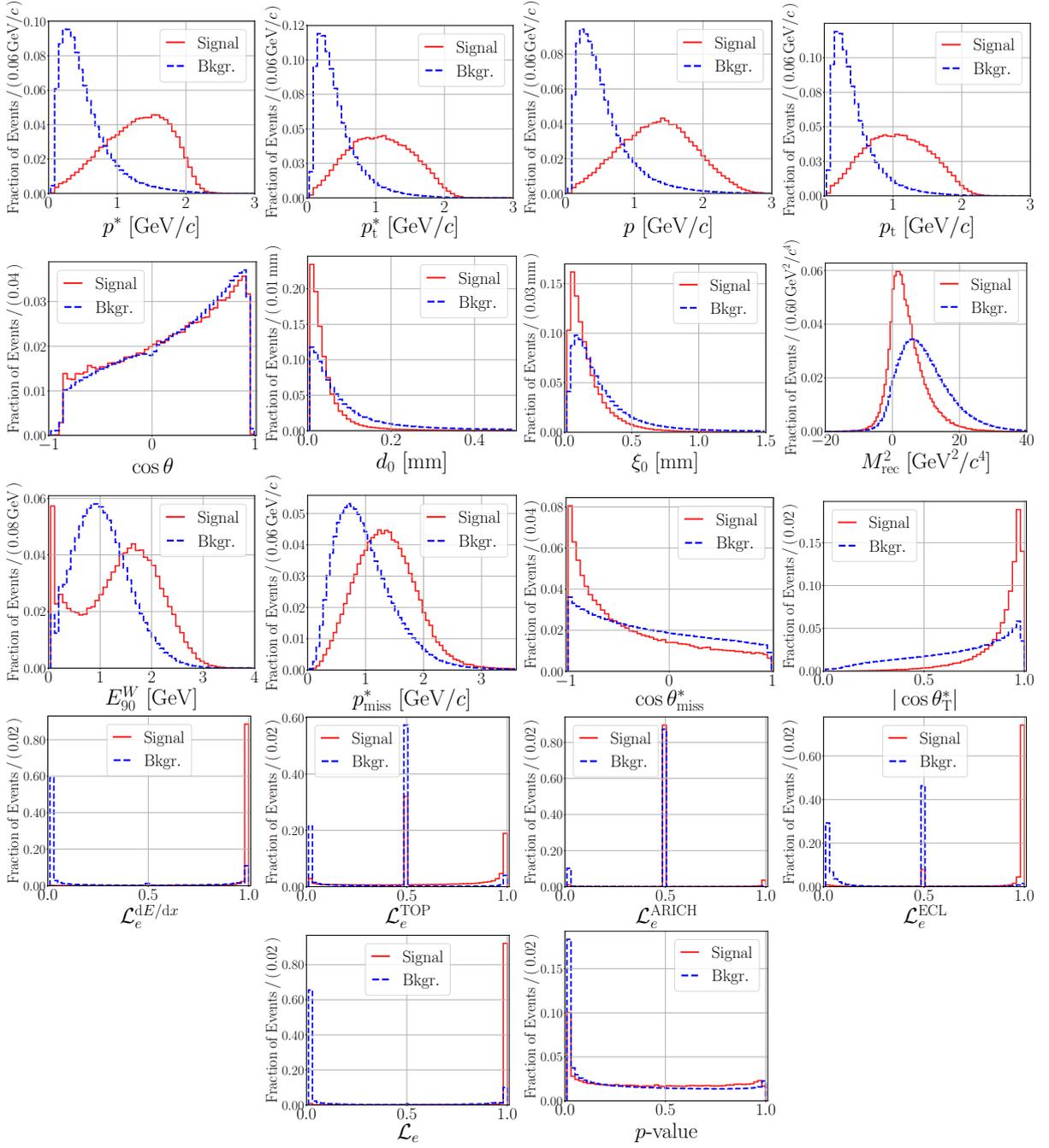


Figure B.1: Distributions of the input variables used for the Electron category.

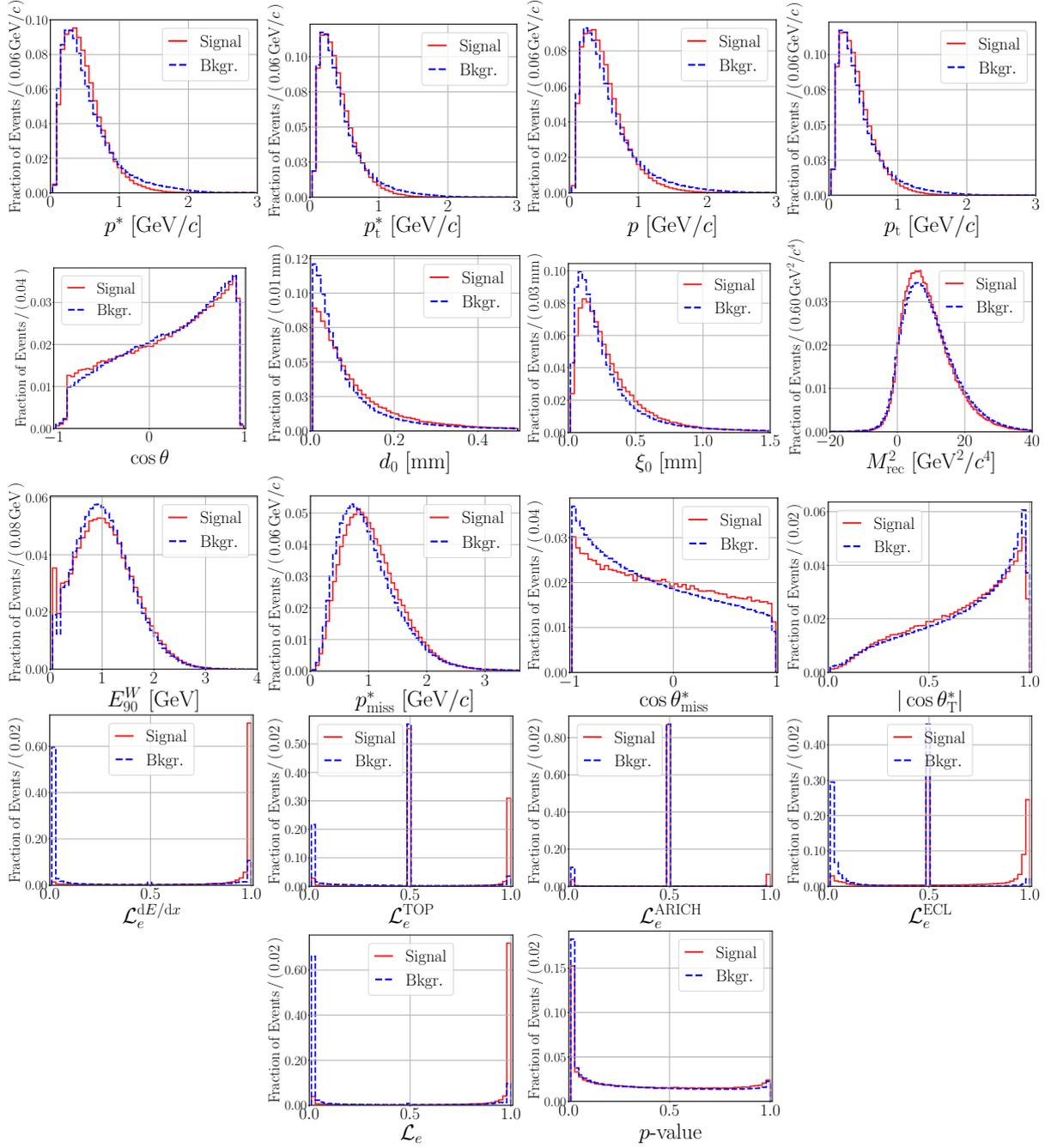


Figure B.2: Distributions of the input variables used for the Intermediate Electron category.

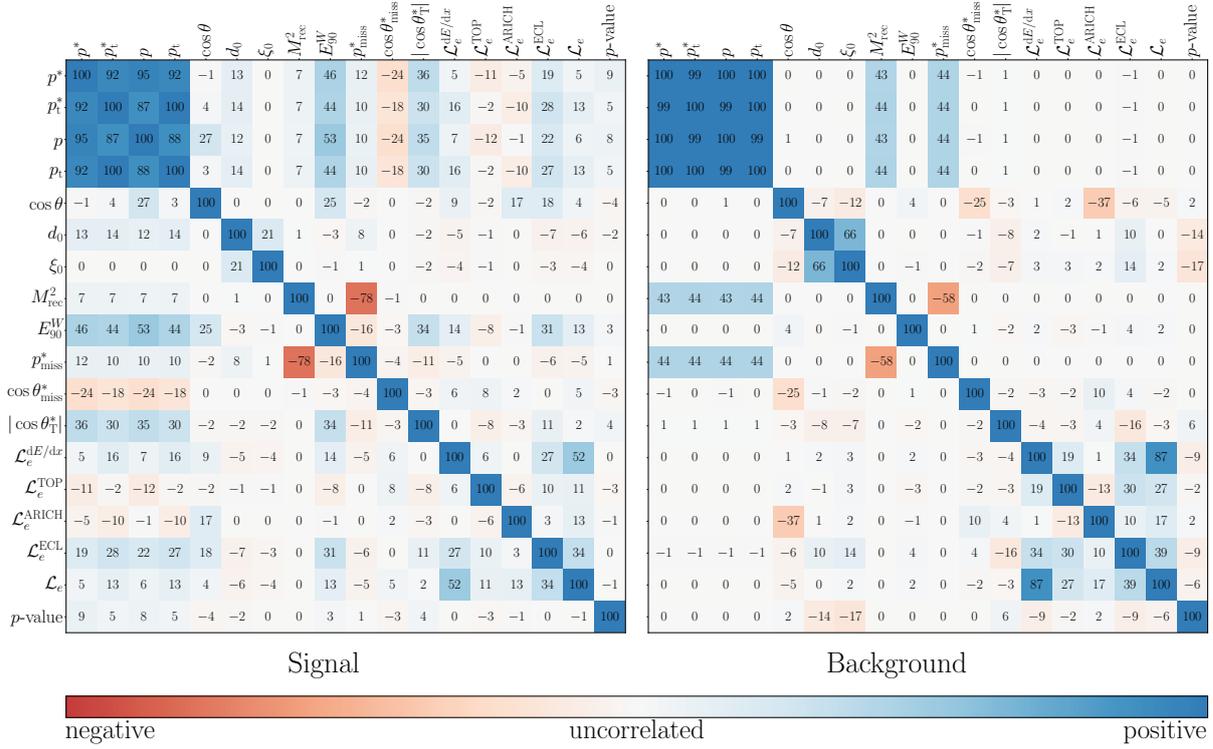


Figure B.3: Correlations between the input variables used for the Electron category.

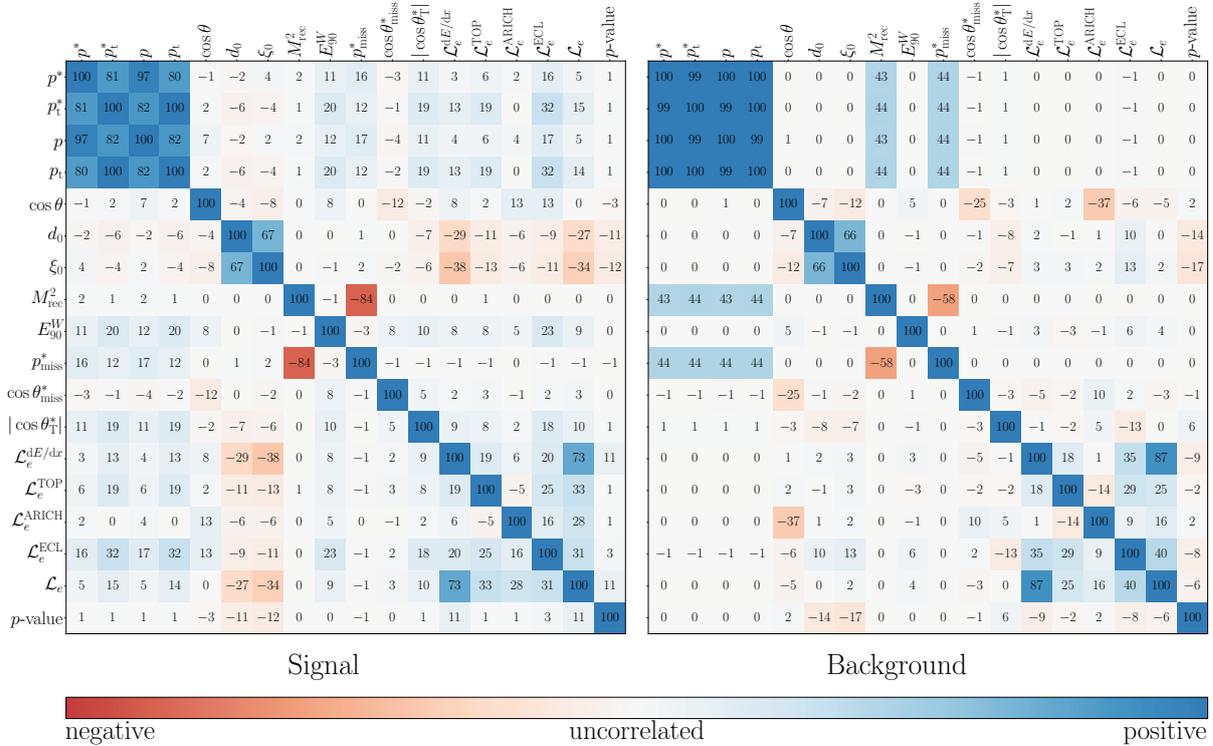


Figure B.4: Correlations between the input variables used for the Intermediate Electron category.

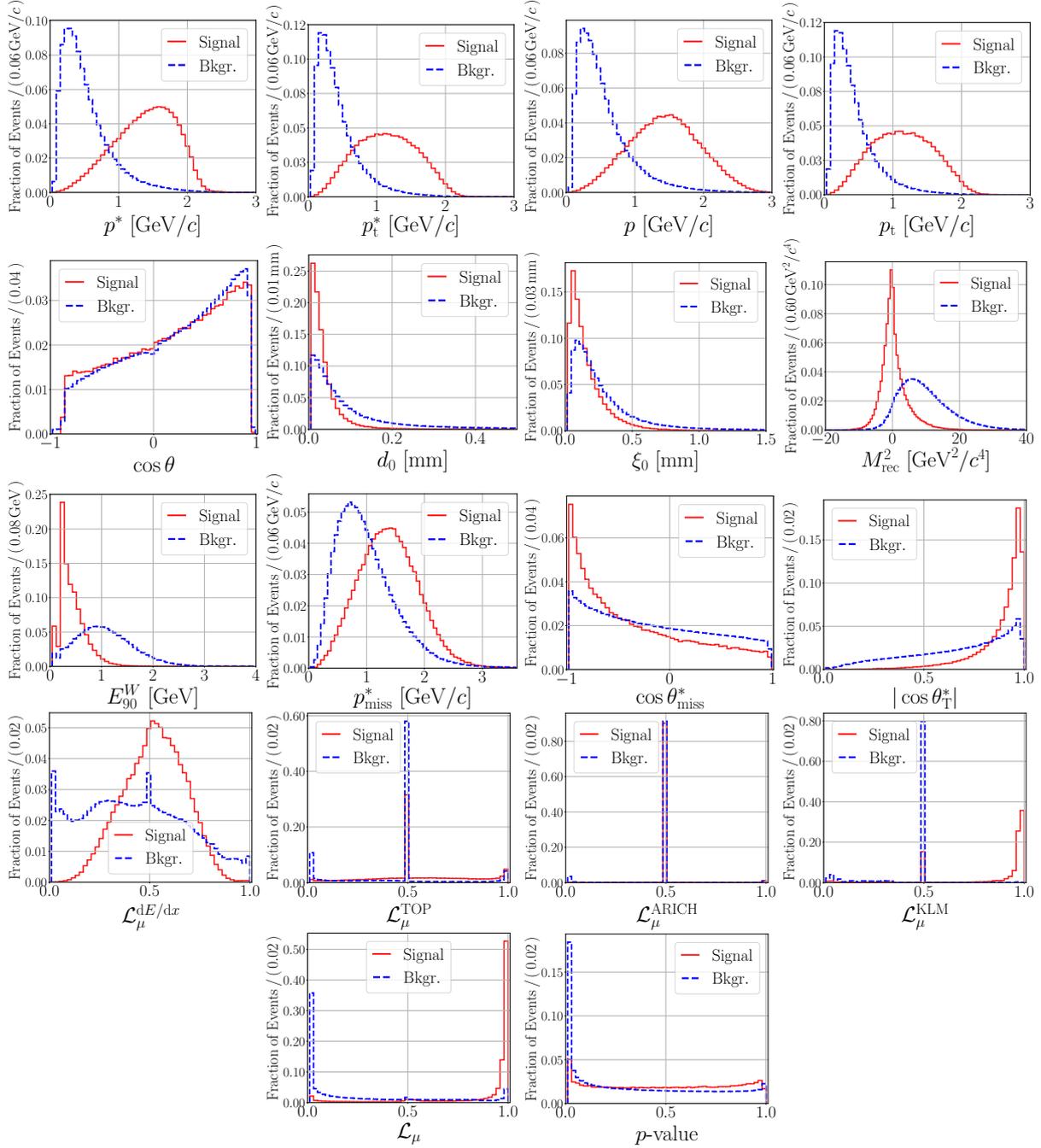


Figure B.5: Distributions of the input variables used for the Muon category.

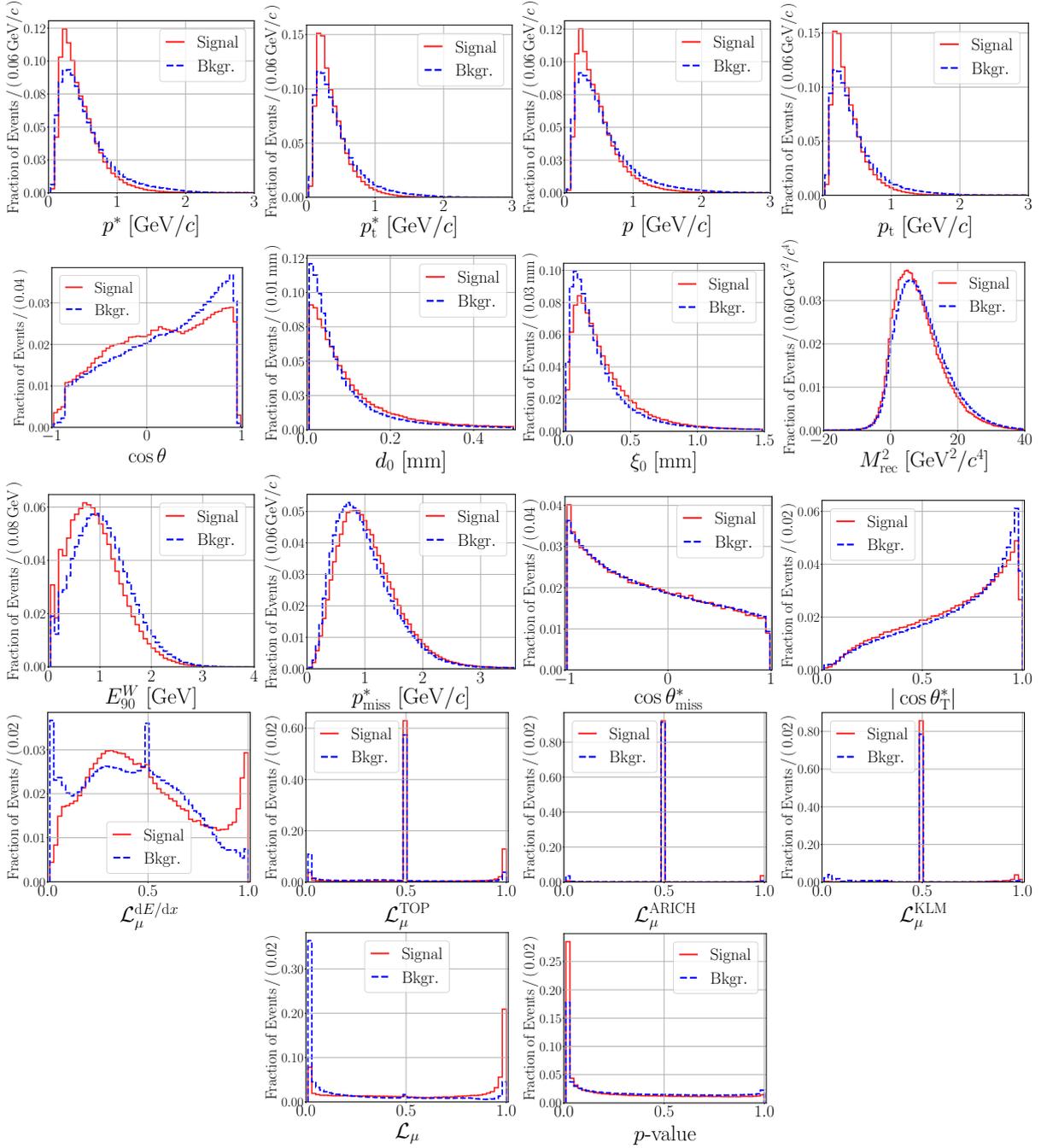


Figure B.6: Distributions of the input variables used for the Intermediate Muon category.

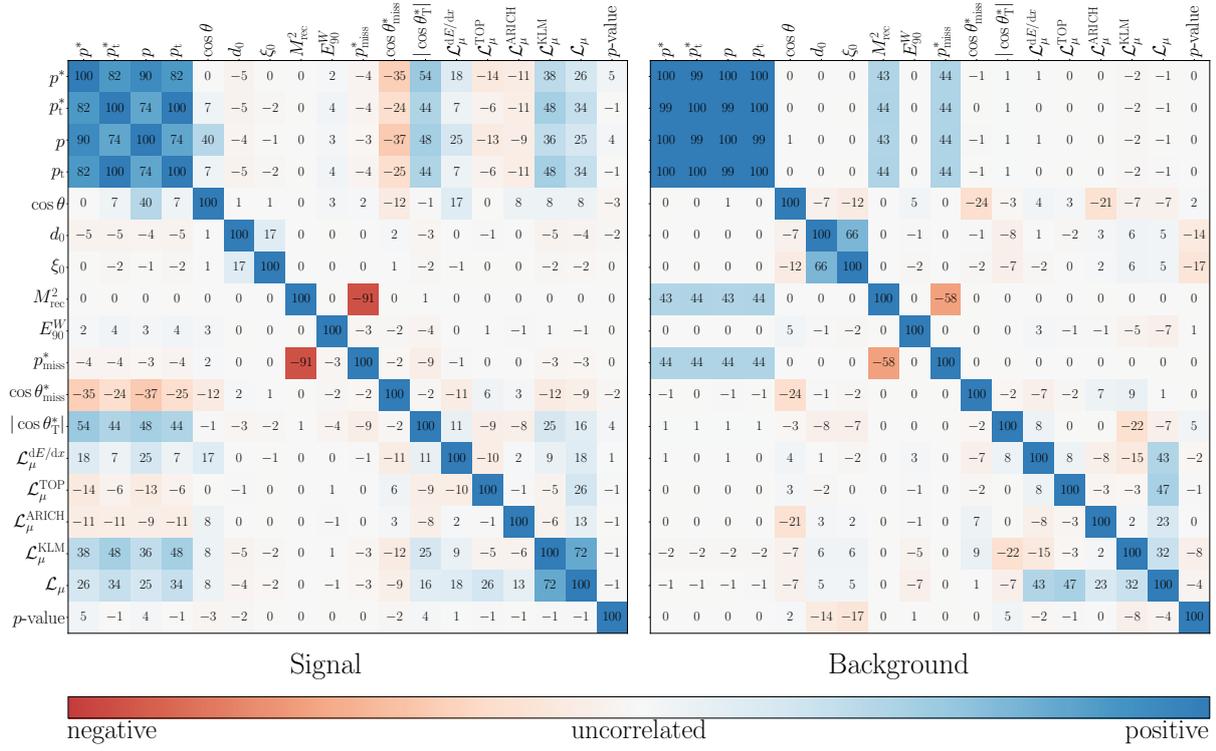


Figure B.7: Correlations between the input variables used for the Muon category.

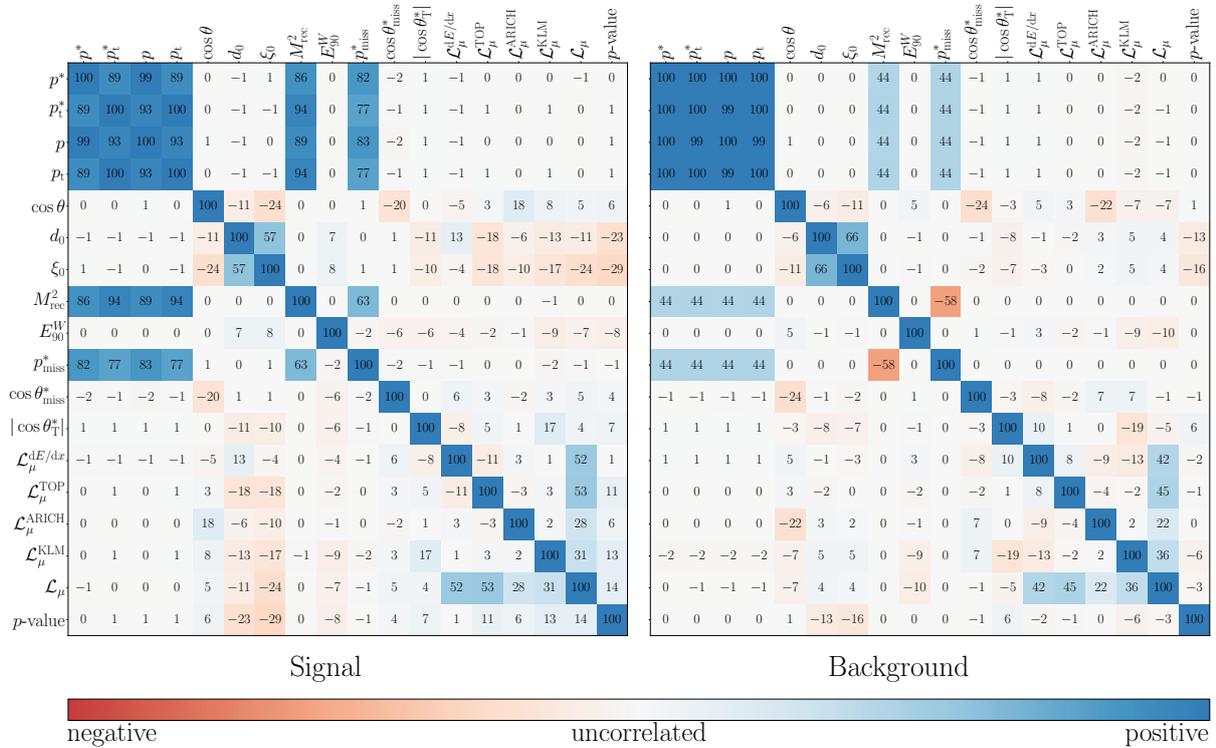


Figure B.8: Correlations between the input variables used for the Intermediate Muon category.

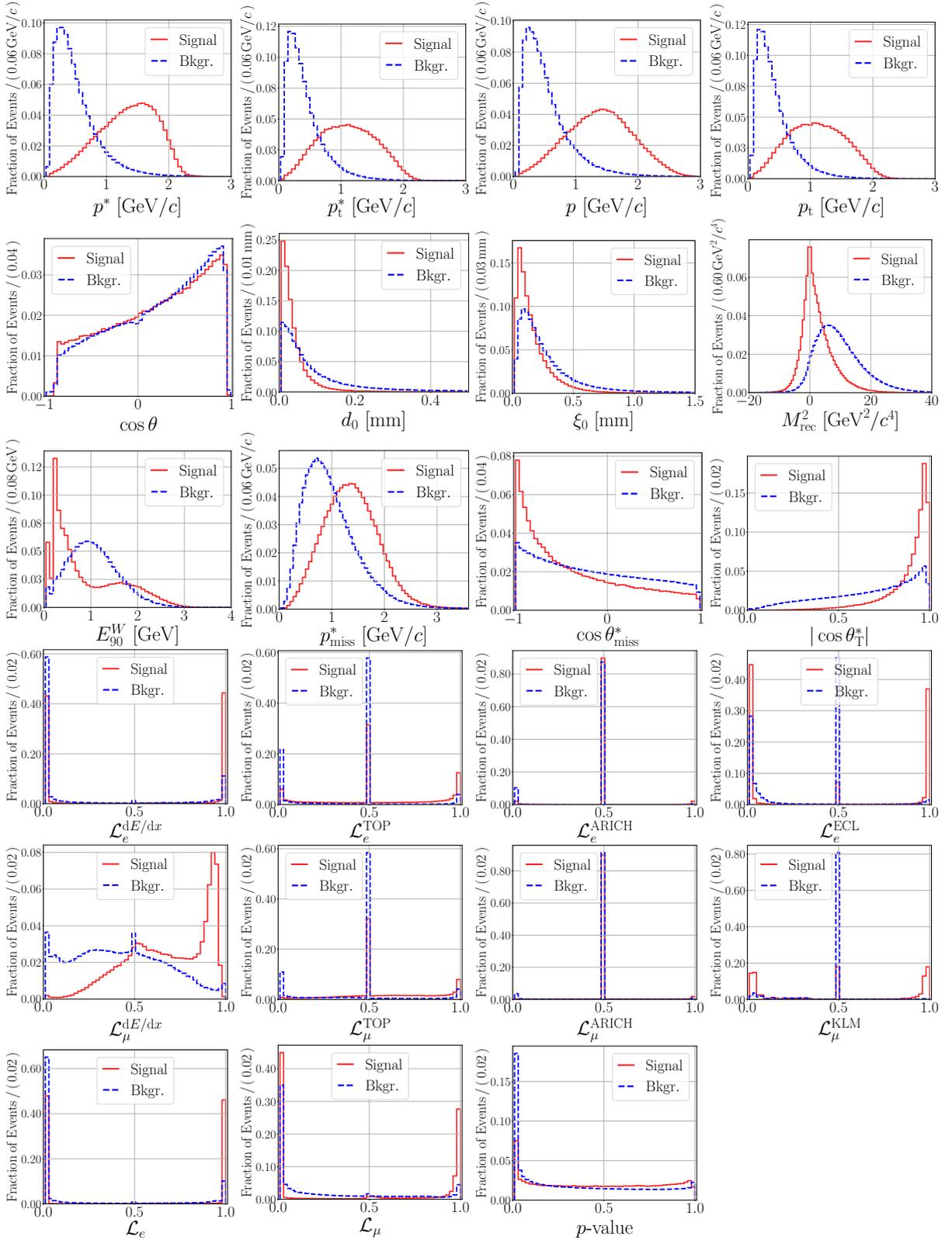


Figure B.9: Distributions of the input variables used for the Kinetic Lepton category.

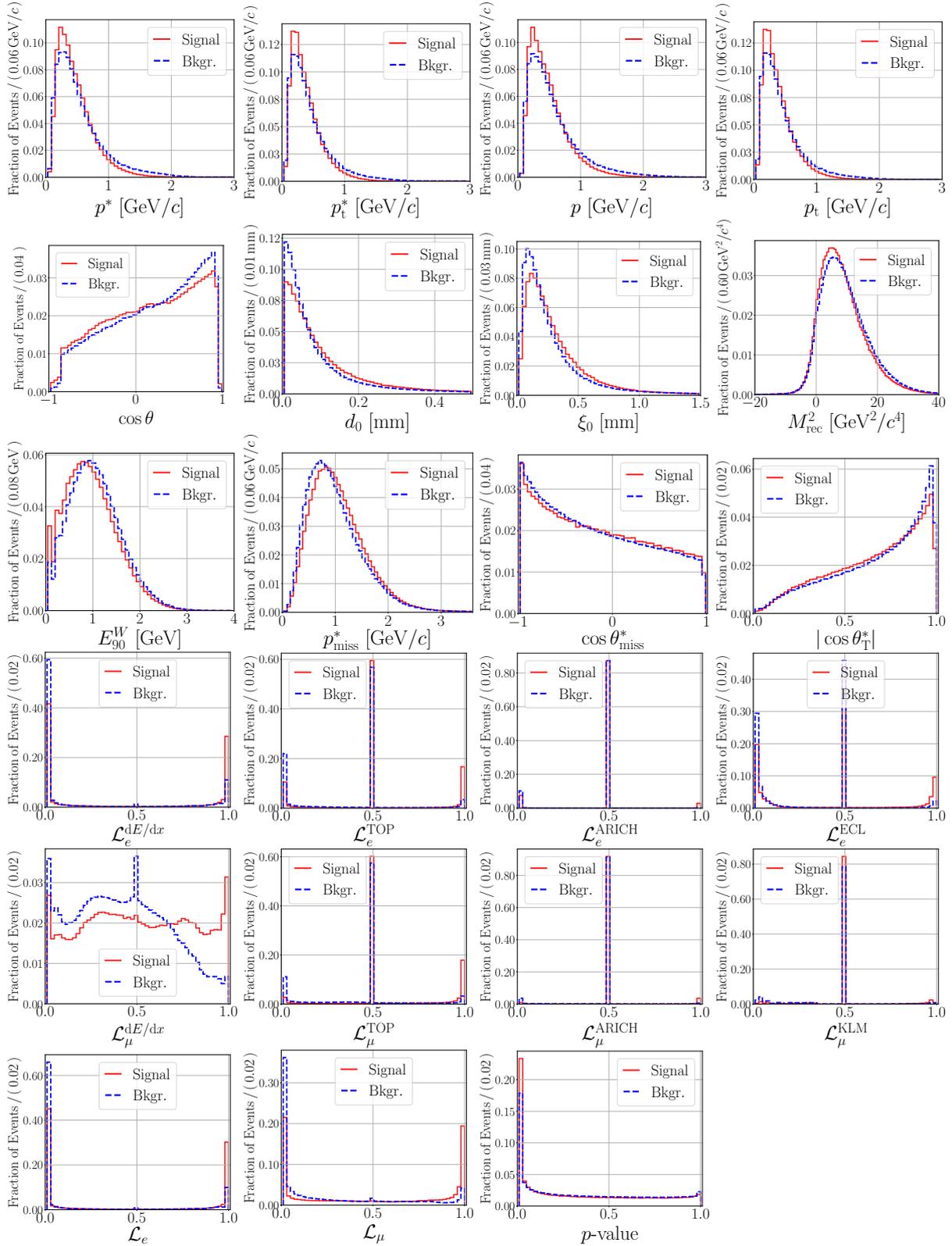


Figure B.10: Distributions of the input variables used for the Intermediate Kinetic Lepton category.

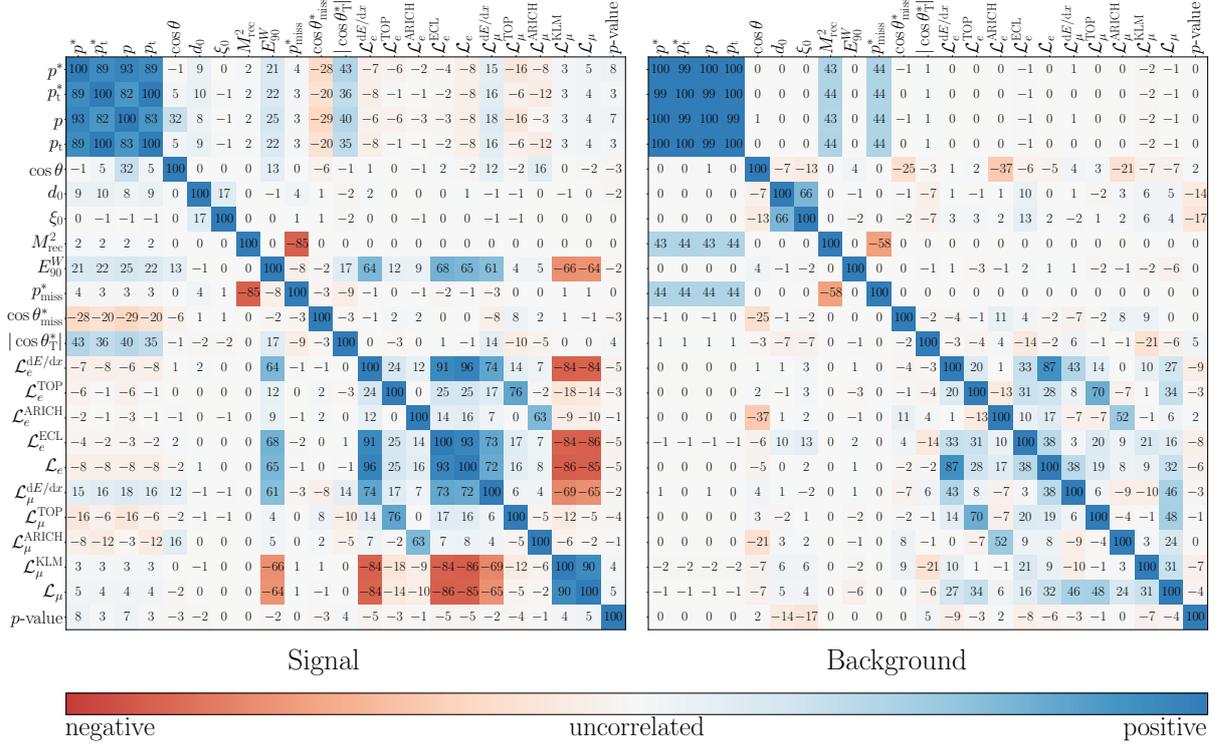


Figure B.11: Correlations between the input variables used for the Kinetic Lepton category.

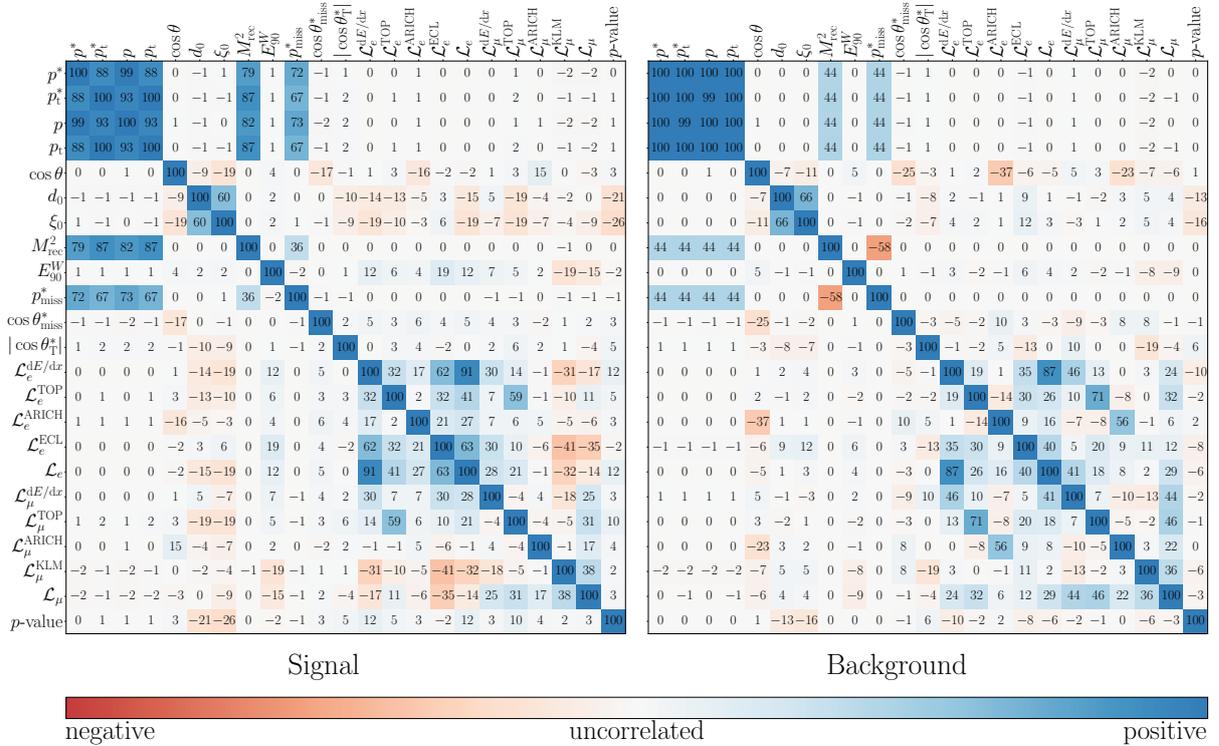


Figure B.12: Correlations between the input variables used for the Intermediate Kinetic Lepton category.

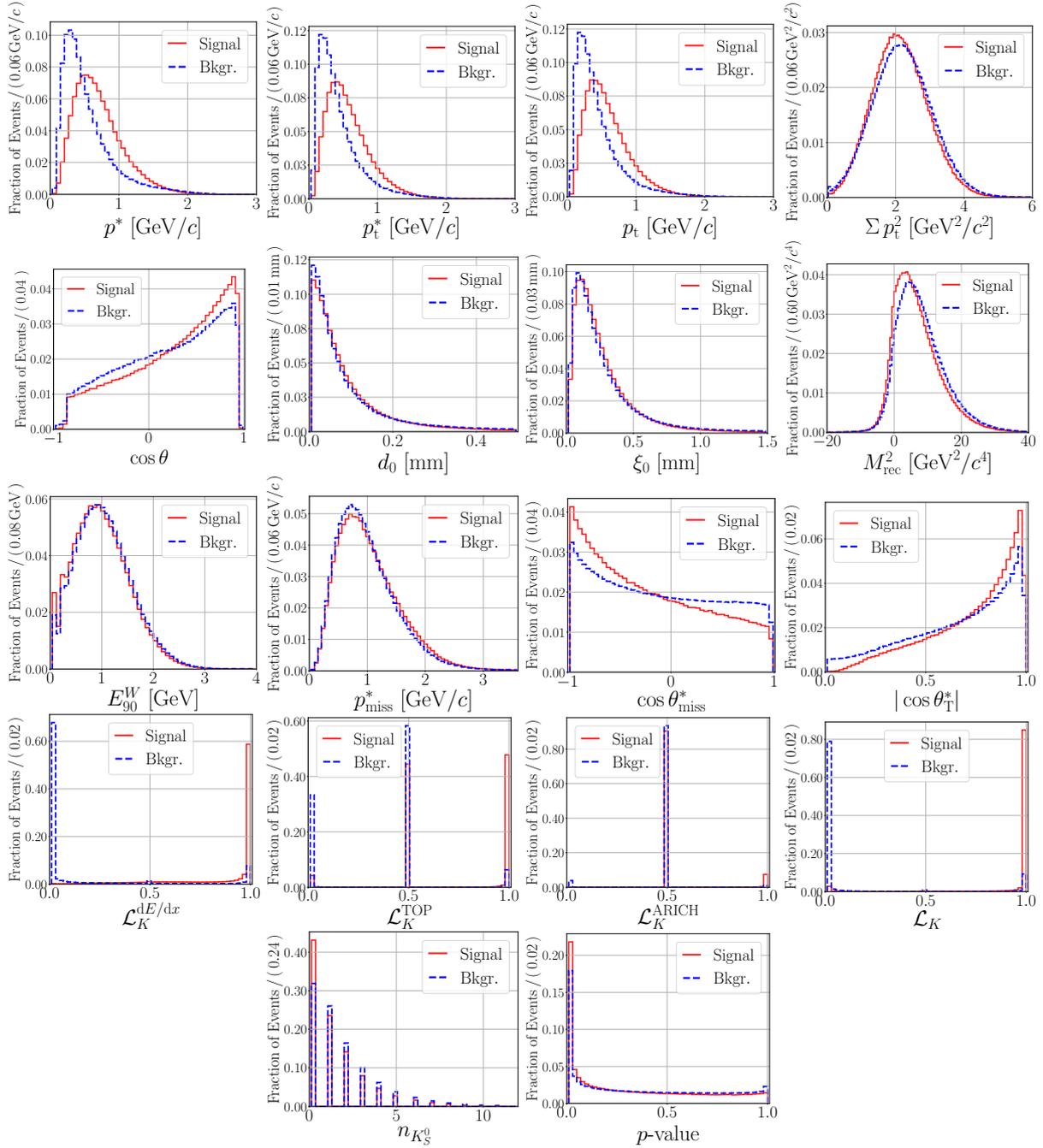


Figure B.13: Distributions of the input variables used for the Kaon category.

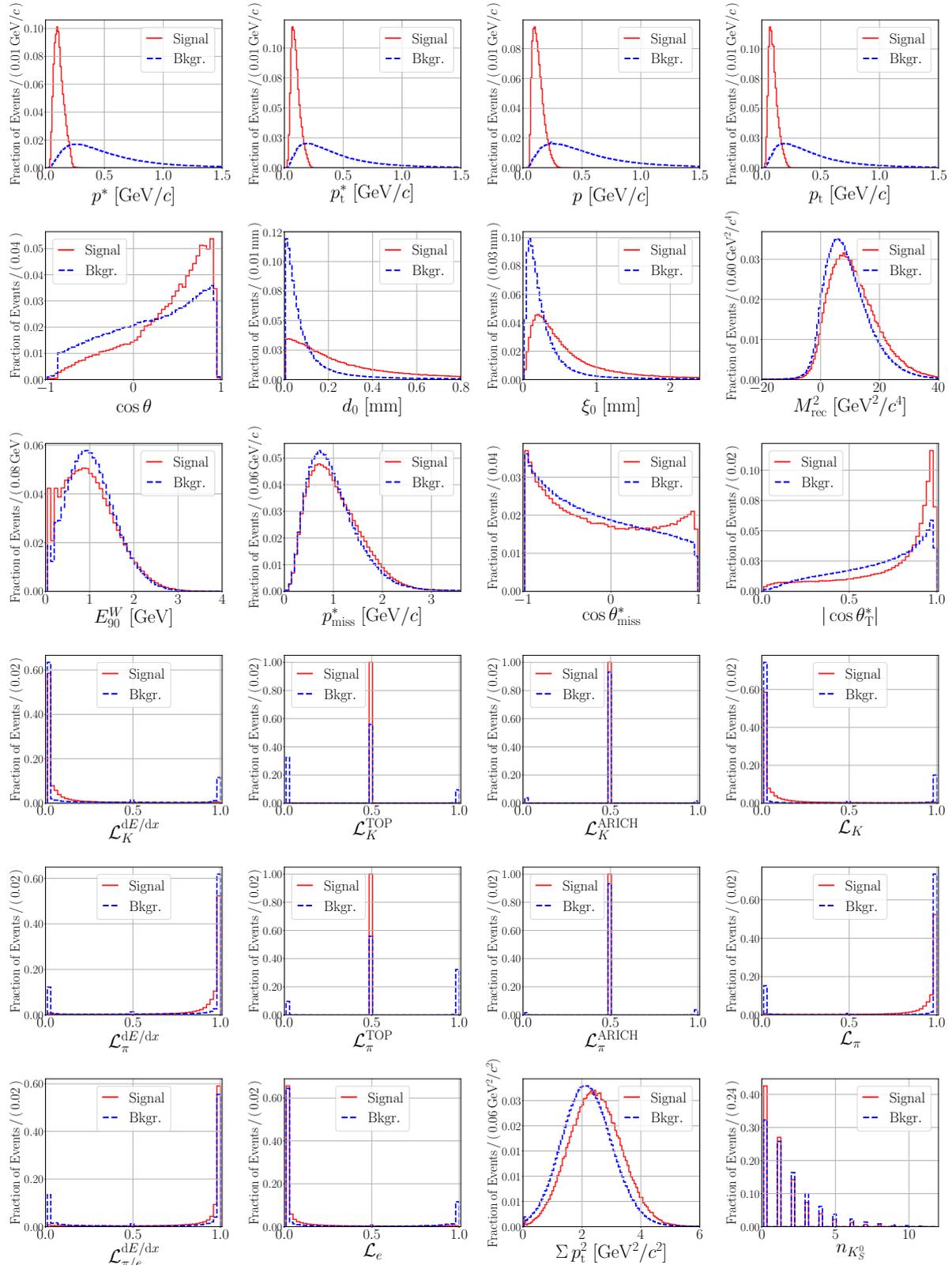


Figure B.16: Distributions of the input variables used for the Slow Pion category.

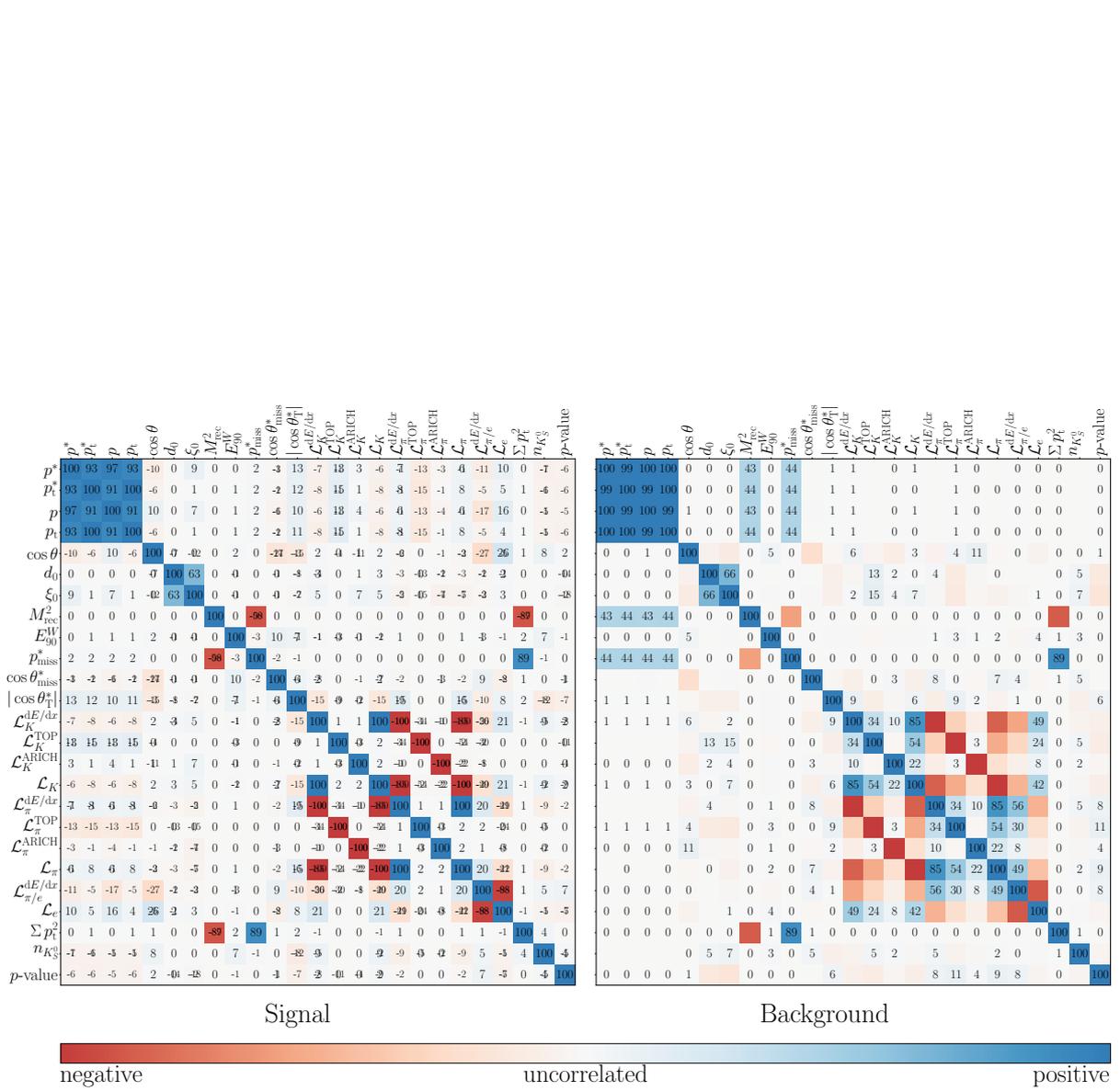


Figure B.17: Correlations between the input variables used for the Slow Pion category.

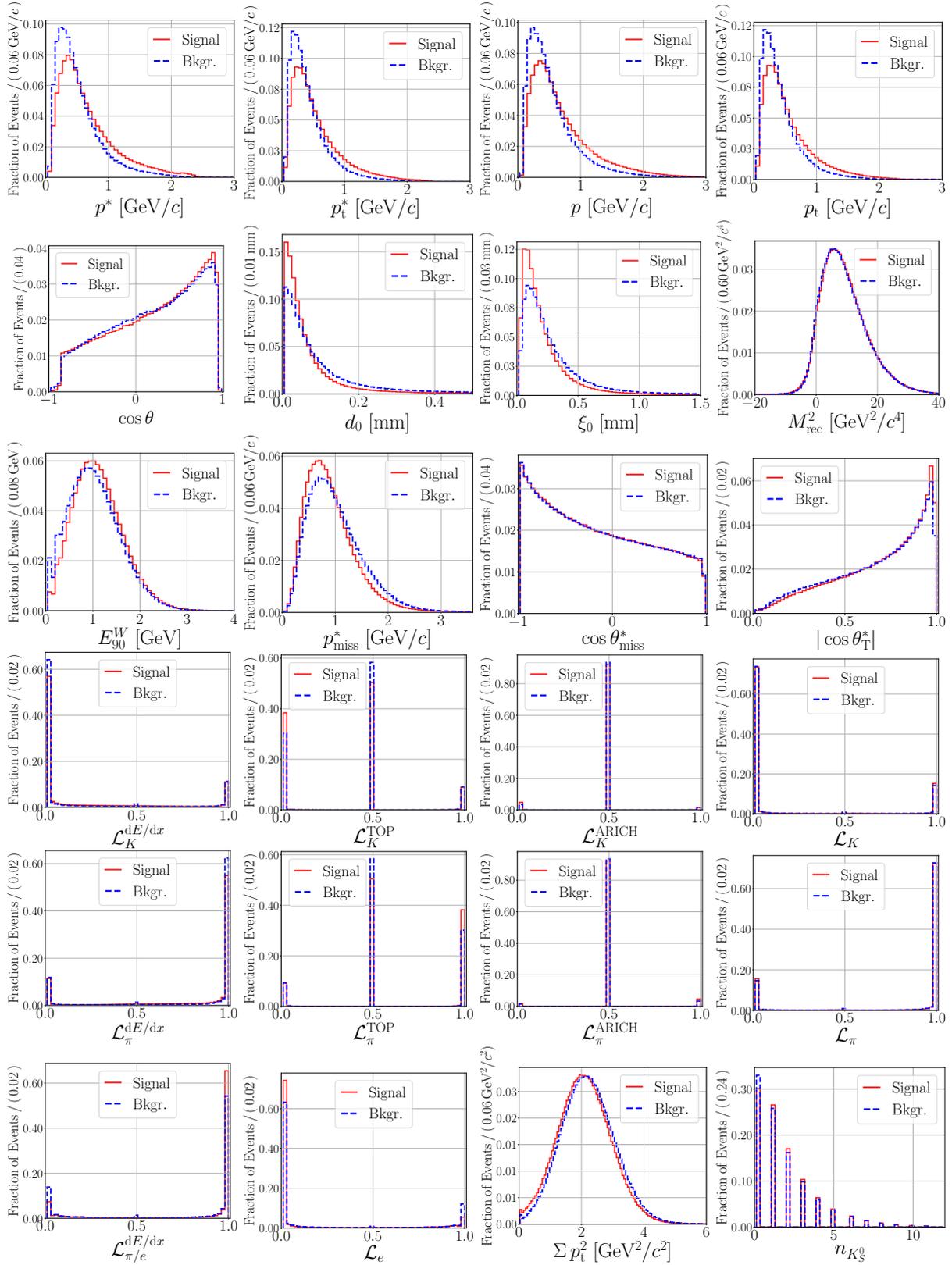


Figure B.18: Distributions of the input variables used for the Fast Hadron category.

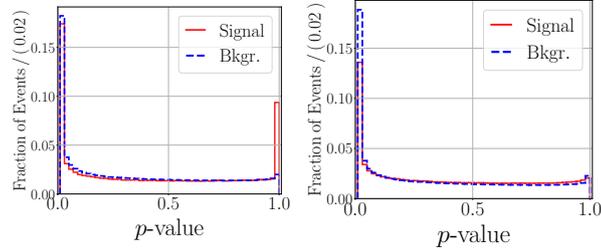


Figure B.19: Distributions of the track p -value for the Slow Pion category (left) and for the Fast Hadron category (right).

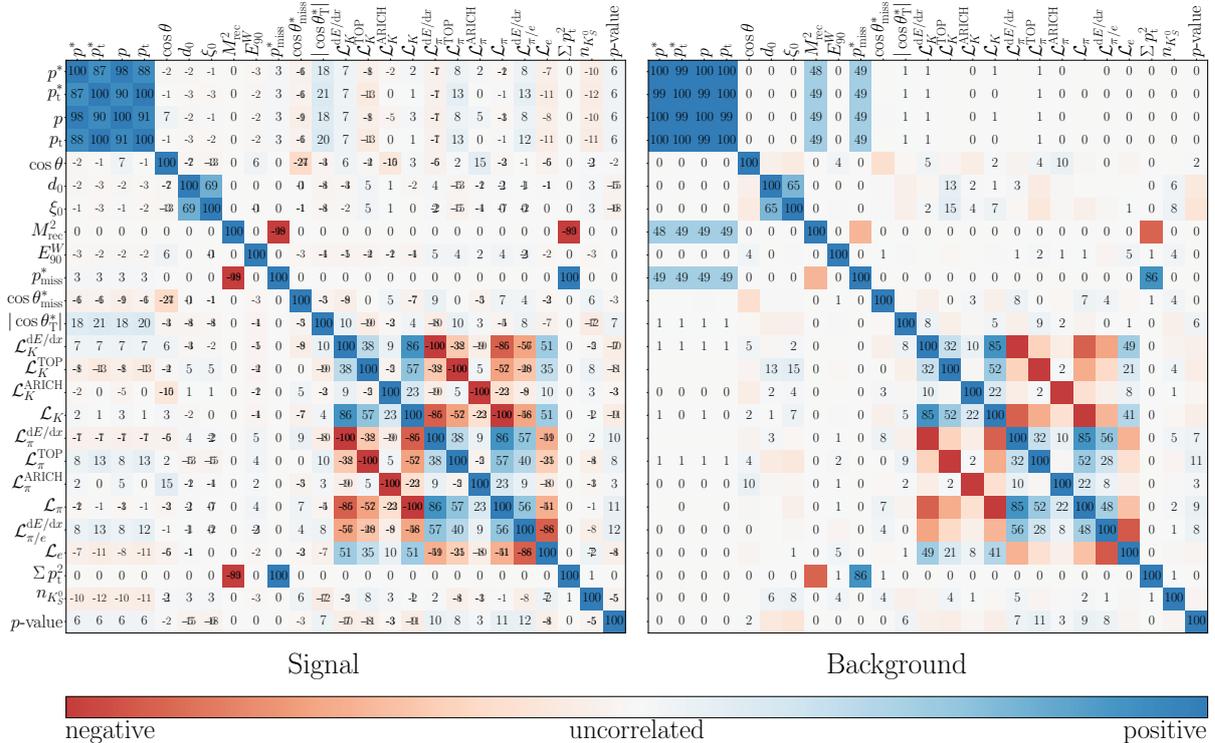


Figure B.20: Correlations between the input variables used for the Fast Hadron category.

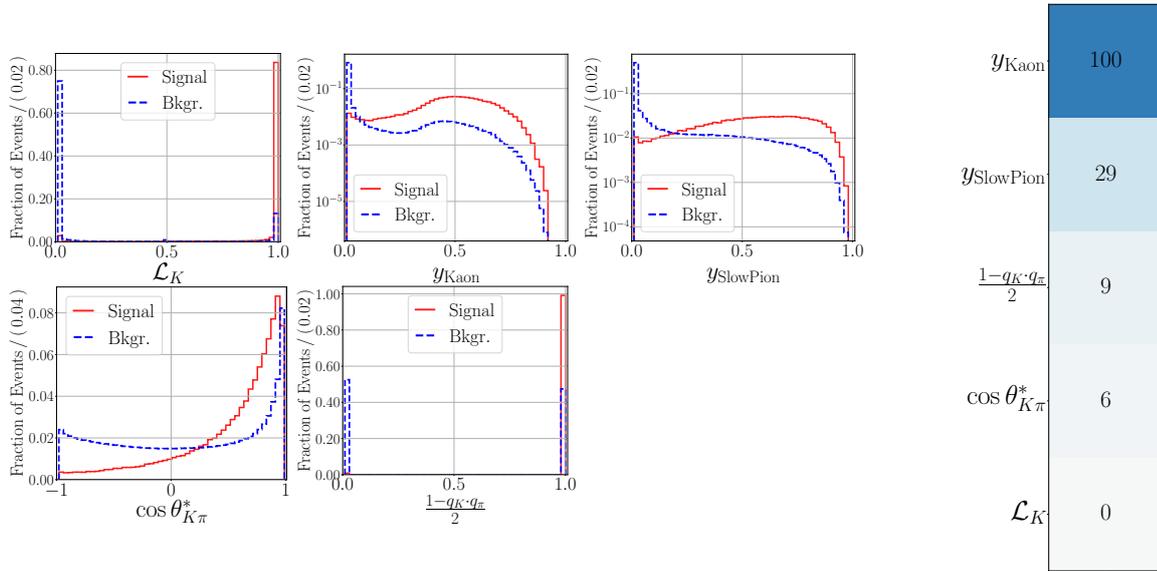


Figure B.21: Distributions and ranking of the input variables used for the Kaon-Pion category. The ranking is provided by the FBDT algorithm.

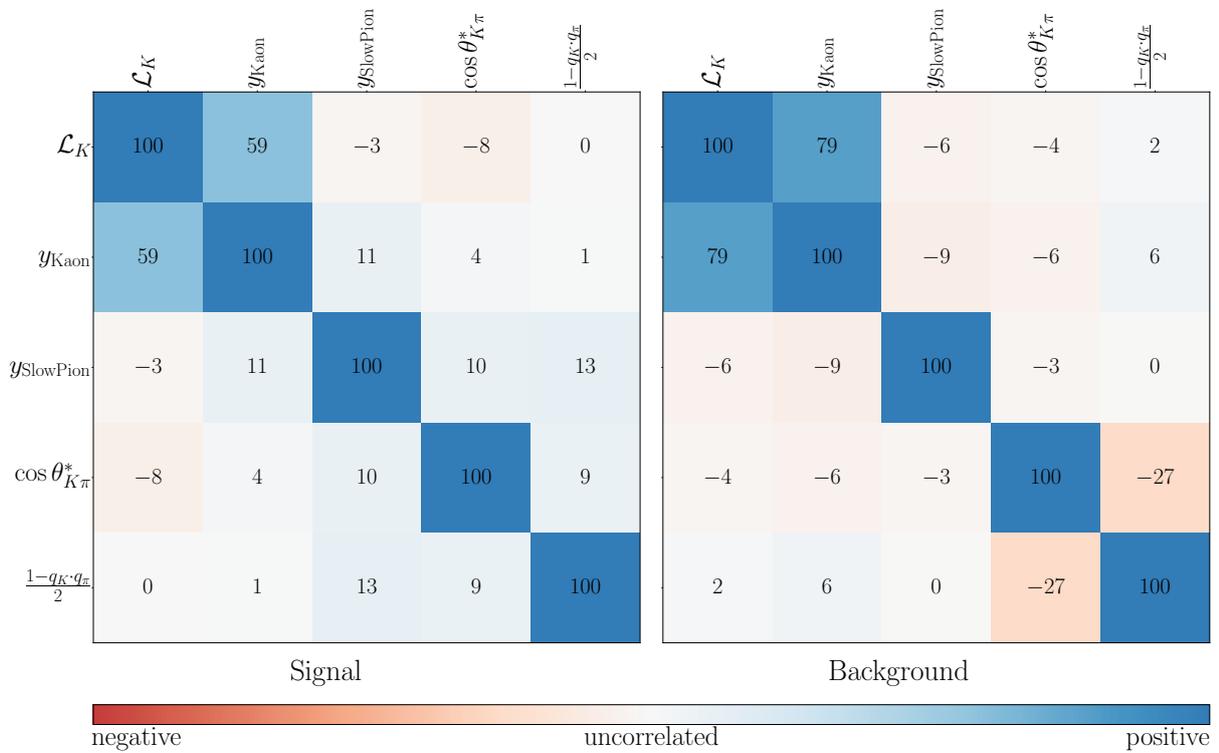


Figure B.22: Correlations between the input variables used for the Kaon-Pion category.

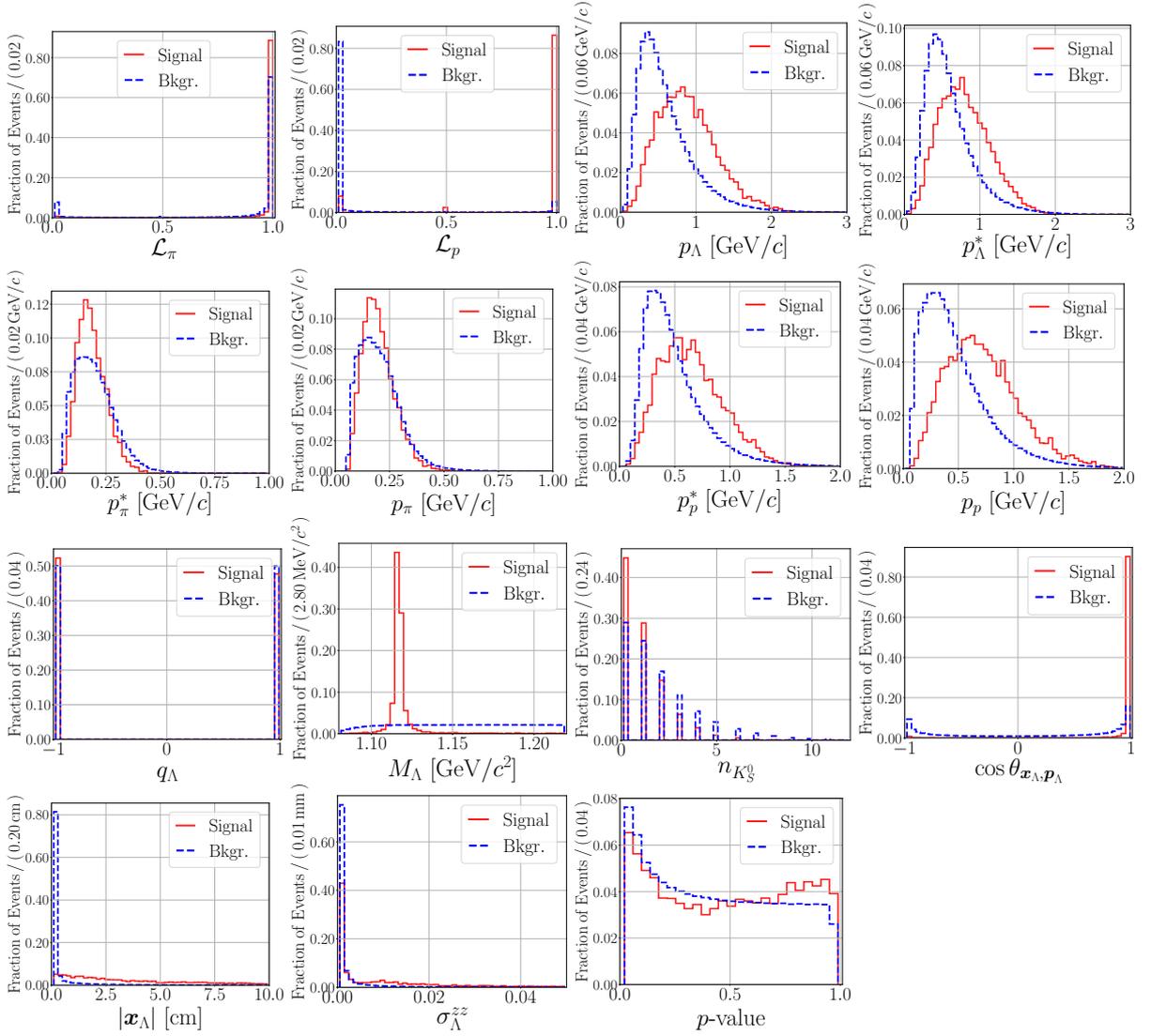


Figure B.27: Distributions of the input variables used for the Lambda category.

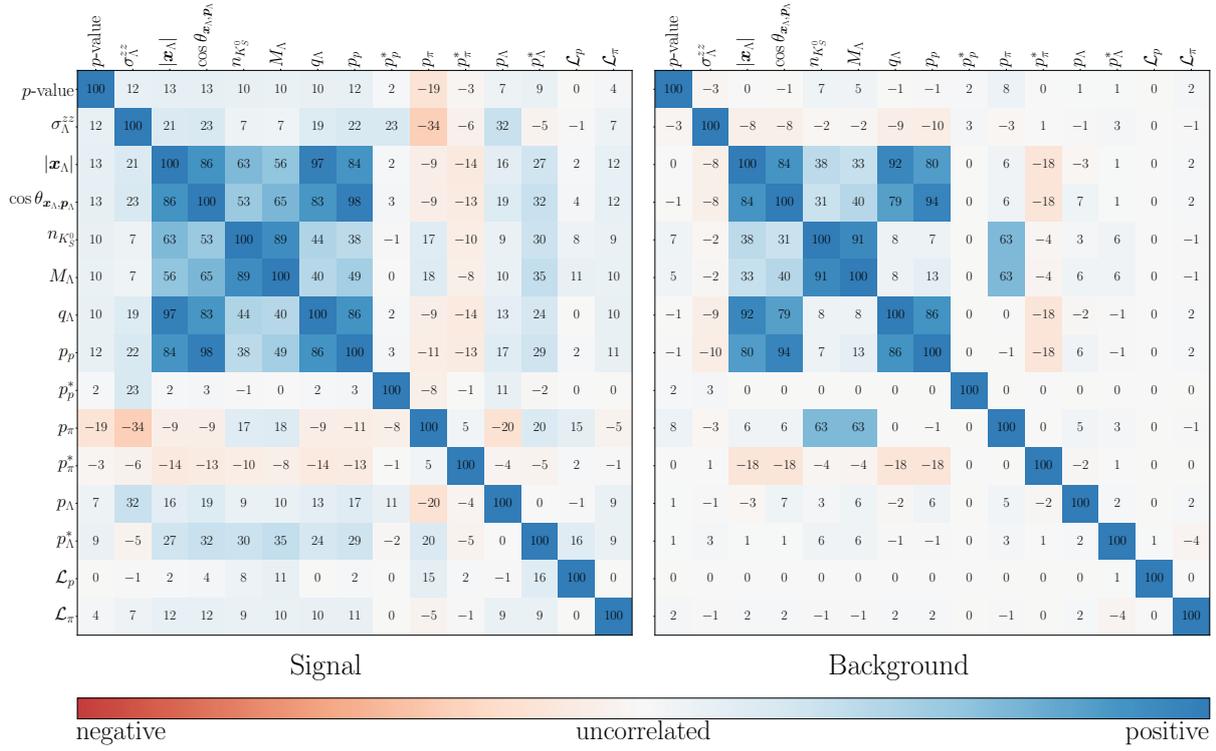


Figure B.28: Correlations between the input variables used for the Lambda category.

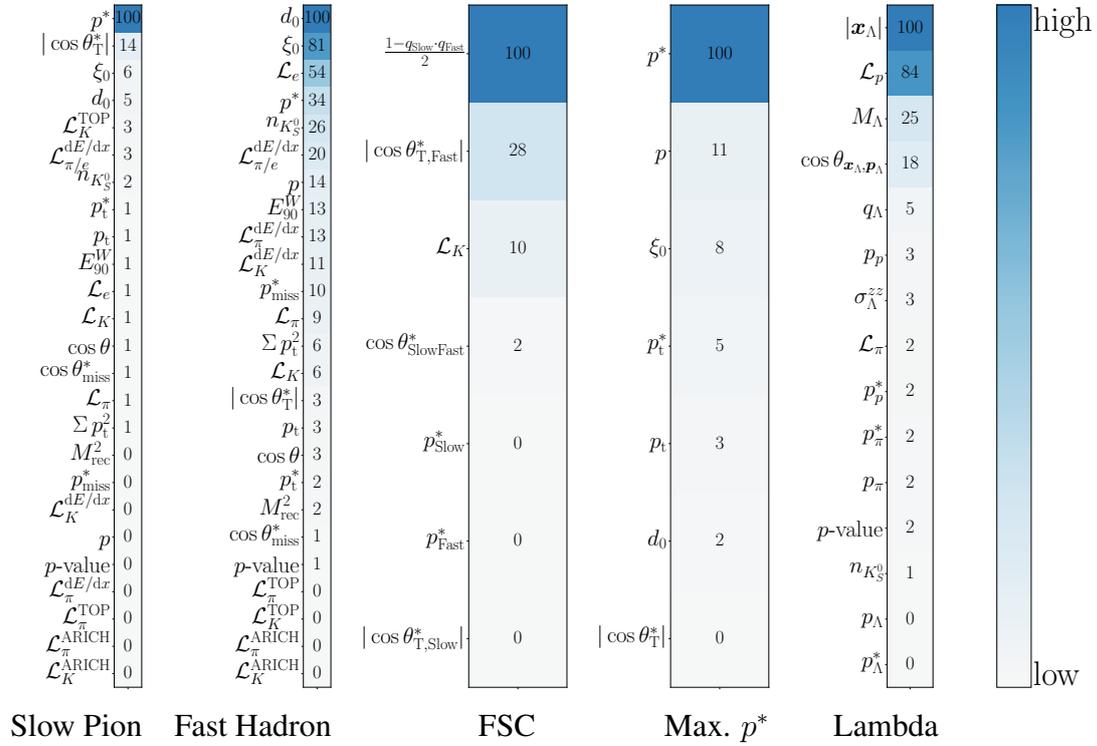


Figure B.29: Ranking of input variables provided by the FBDT algorithm for different categories.

B.4 Studies without impact parameters

This section presents the results of studies which are complementary to the ones discussed in Sect. 4.6 and 4.7. The aim of these studies was to investigate how the performance of the flavor tagger changes when it is trained without using the impact parameters d_0 and z_0 .

Figure B.30 (left) shows the output of the flavor tagger after being trained with the official Belle II MC sample for which $\mathcal{S}_{CP} = 0.69$. The trained flavor tagger was tested for the three different values of \mathcal{S}_{CP} . In all the cases, the central bump in the $q \cdot r$ distribution is centered around zero. The total effective efficiency has compatible values for the different values of \mathcal{S}_{CP} , but the difference $\Delta\varepsilon_{\text{eff}}$ shows a slightly difference for $\mathcal{S}_{CP} = 0$. In contrast, when the flavor tagger is trained with the same sample, but using the impact parameters (see Sect. 4.6), the central bump is shifted, and the effective efficiency and the difference $\Delta\varepsilon_{\text{eff}}$ have different values for the different values of \mathcal{S}_{CP} . However, with impact parameters, the effective efficiency is about 3 percent points higher. Figure B.30 (right) shows the output of the flavor tagger multiplied by the true flavor q_{MC} of B_{tag}^0 . For all values of \mathcal{S}_{CP} , the asymmetry between \overline{B}^0 ($q_{MC} = -1$) and B^0 ($q_{MC} = 1$) is similar. However, in contrast to the case with impact parameters, the asymmetry is prominent in r bins where the tagging power is high.

Figure B.31 presents the corresponding values of ε , μ , w and Δw for each r bin, as well as the linearity checks. As Fig. B.31(a) and Fig. B.31(b) show, the values of ε and μ change subtly for the different values of \mathcal{S}_{CP} , while the values of w and Δw remain practically the same. Fig. B.31(c) shows the linearity checks for the three different values of \mathcal{S}_{CP} . In average, the output dilution r_{FBDT} corresponds to the true dilution r_{MC} . However, this does not hold if one considers separately the events where B_{tag}^0 is a B^0 , and where B_{tag}^0 is a \overline{B}^0 : for the r bins between 0.25 and 0.625, the value of r_{FBDT} differs considerably from r_{MC} . In contrast, with impact parameters, the individual linearity checks do not hold only in the lowest two r bins, where the tagging power is low.

Figure B.32 shows the $q \cdot r$ output and the product $q_{MC} \cdot q \cdot r$ after training the flavor tagger with the private Belle II MC sample for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. The flavor tagger is tested for the three considered values of \mathcal{S}_{CP} . In this case also the effective efficiency reduces by about 3 percent points compared to the case with impact parameters. Additionally, the $q_{MC} \cdot q \cdot r$ distributions show an asymmetry in r bins with high tagging power. In contrast, with impact parameters, the distributions for B^0 and for \overline{B}^0 overlap almost perfectly except at the boundaries -1 and $+1$, where small asymmetries are present.

As Figure B.33 shows, even when the flavor tagger is trained with $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$ avoiding impact parameters, the values of μ still change slightly depending on the value of \mathcal{S}_{CP} . Considering the linearity plots, the dilution factor for B^0 and for \overline{B}^0 events is slightly non-calibrated in the r bin between 0.25 and 0.5. On the contrary, when the flavor tagger is trained with $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$ using the impact parameters, the dilution factor is correctly calibrated in average, and separately for B^0 and for \overline{B}^0 in the whole range and in all the considered CP -violation cases.

Considering now the correlations with Δt , Fig. B.34 shows the $q \cdot r$ output of the flavor tagger and the product $q_{MC} \cdot q \cdot r$ for positive and for negative values of Δt^{gen} . The flavor tagger was trained and tested with the Belle II MC samples for which $(\mathcal{A}_{CP}, \mathcal{S}_{CP}) = (0, 0)$. As Fig. B.34 shows, the distributions for positive and for negative values of Δt^{gen} overlap perfectly. In contrast, with impact parameters (see Sect. 4.7), they are slightly different. Even without impact parameters, the effective efficiency for positive values of Δt^{gen} is slightly higher than for negative values of Δt^{gen} . Considering the values of $\Delta \varepsilon_{\text{eff}}$, there is also a small difference between positive and negative Δt^{gen} ranges of about 0.2 percent points, which is the same difference as in the case with impact parameters.

Figure B.35 shows the respective values of ε , μ , w and Δw for each r bin, and for the positive and the negative Δt^{gen} ranges. For ε , w and Δw , there is no difference, but for the parameter μ , there is a small one.

In conclusion, avoiding impact parameters eliminates the shift of the central bump depending on the CP -violation parameters, and reduces the correlations with the generated Δt . However, a small correlation with the CP -violation parameters and with the sign of the generated Δt remain. Additionally, small asymmetries and miscalibration effects become prominent in r bins with high tagging power. And the effective efficiency drops by about 3 percent points. Therefore, avoiding impact parameters is not an optimal solution. The best solution is to use them, but training without CP -violation as shown in Sect. 4.6.

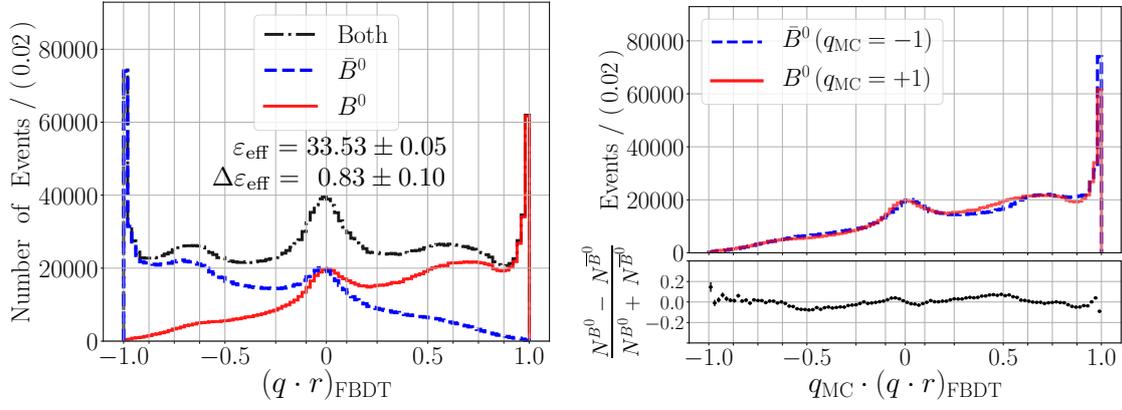
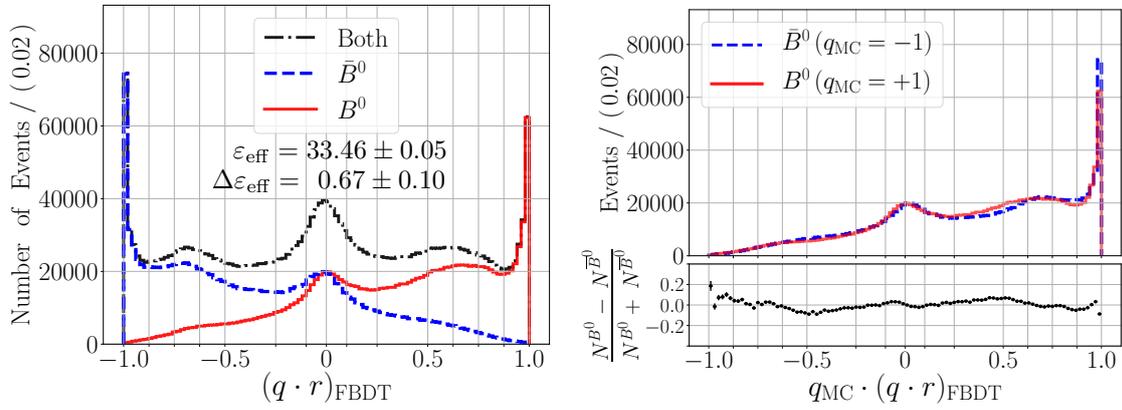
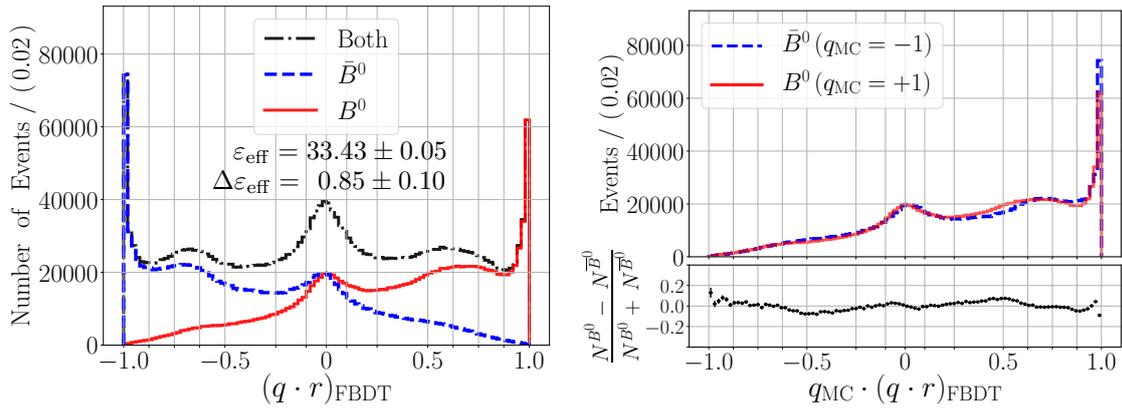
(a) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0.69$.(b) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = 0$.(c) $\mathcal{A}_{CP} = 0$, $\mathcal{S}_{CP} = -0.69$.

Figure B.30: Performance of the FBDT combiner avoiding the impact parameters d_0 and ξ_0 for three values of \mathcal{S}_{CP} : (left) combiner output; (right) combiner output multiplied by the true flavor q_{MC} of B_{tag}^0 . The effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained using the official Belle II MC sample for which $\mathcal{S}_{CP} = 0.69$.

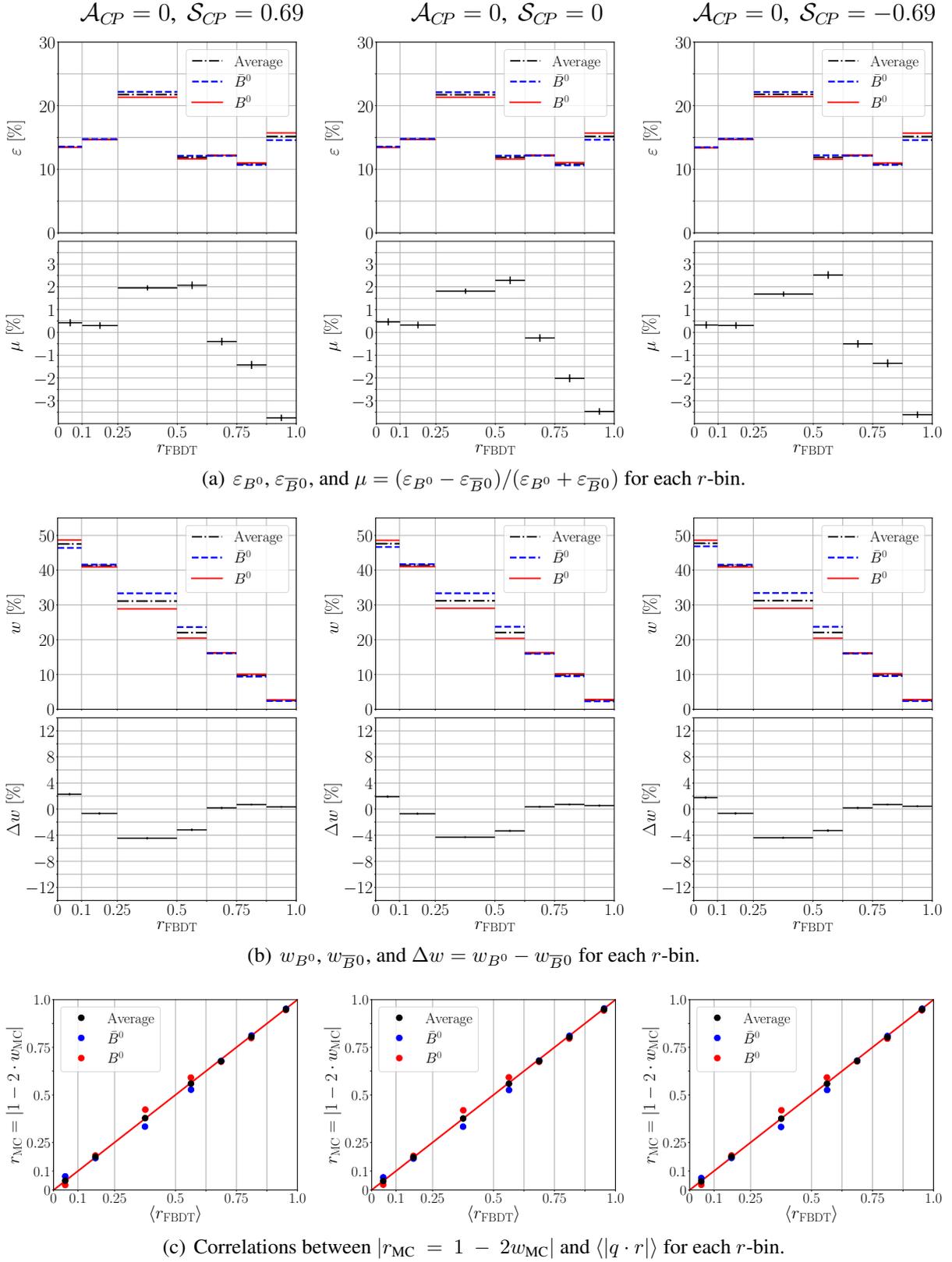


Figure B.31: Performance of the FBDT combiner avoiding the impact parameters d_0 and ξ_0 for three different values of \mathcal{S}_{CP} . The flavor tagger was trained using the official Belle II MC sample for which $\mathcal{S}_{CP} = 0.69$.

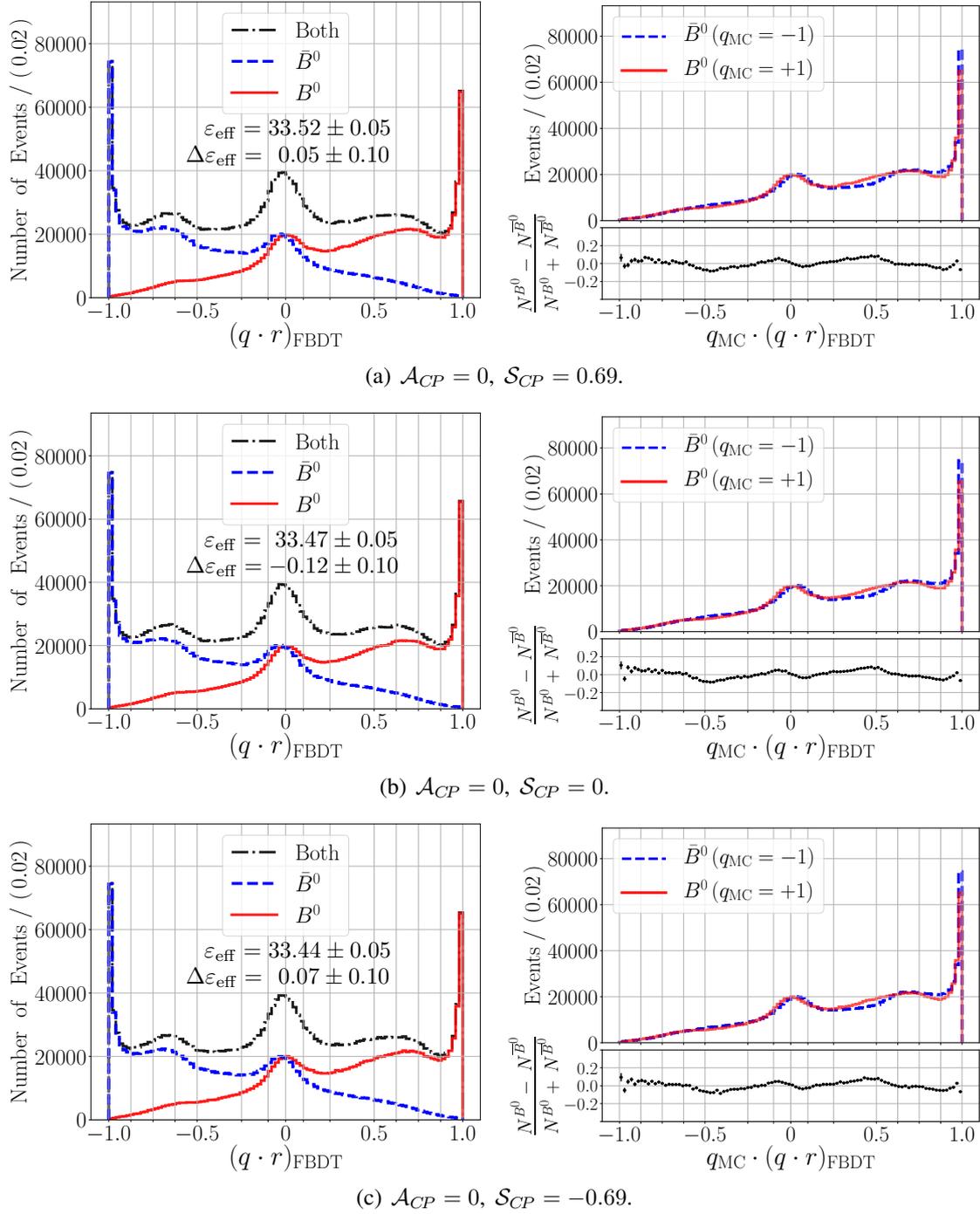


Figure B.32: Performance of the FBBDT combiner avoiding the impact parameters d_0 and ξ_0 for three values of \mathcal{S}_{CP} : (left) combiner output; (right) combiner output multiplied by the true flavor q_{MC} of B_{tag}^0 . The effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained using the private Belle II MC sample for which $\mathcal{S}_{CP} = 0$.

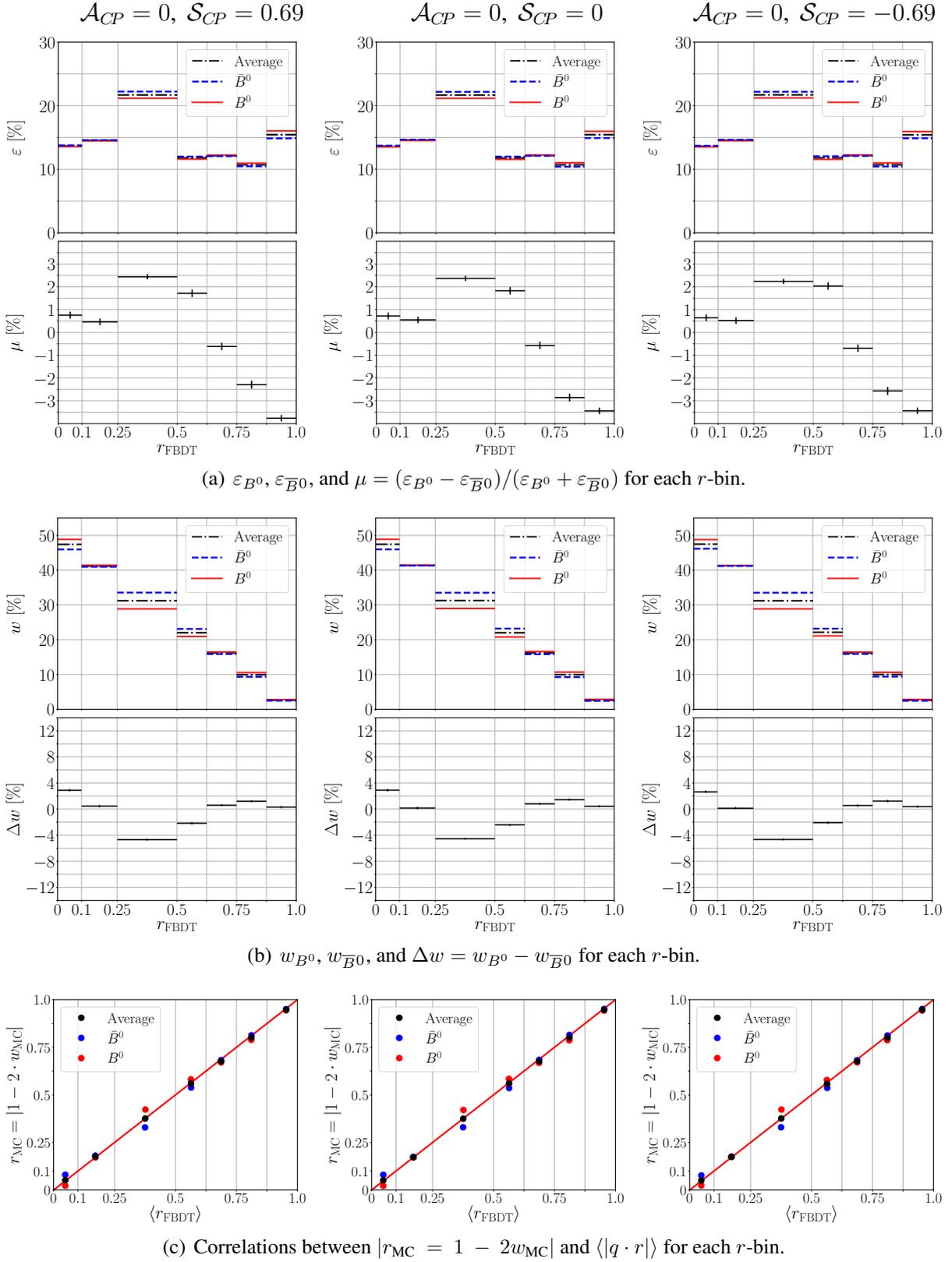


Figure B.33: Performance of the FBDT combiner avoiding the impact parameters d_0 and ξ_0 for three different values of \mathcal{S}_{CP} . The flavor tagger was trained using the private Belle II MC sample for which $\mathcal{S}_{CP} = 0$.

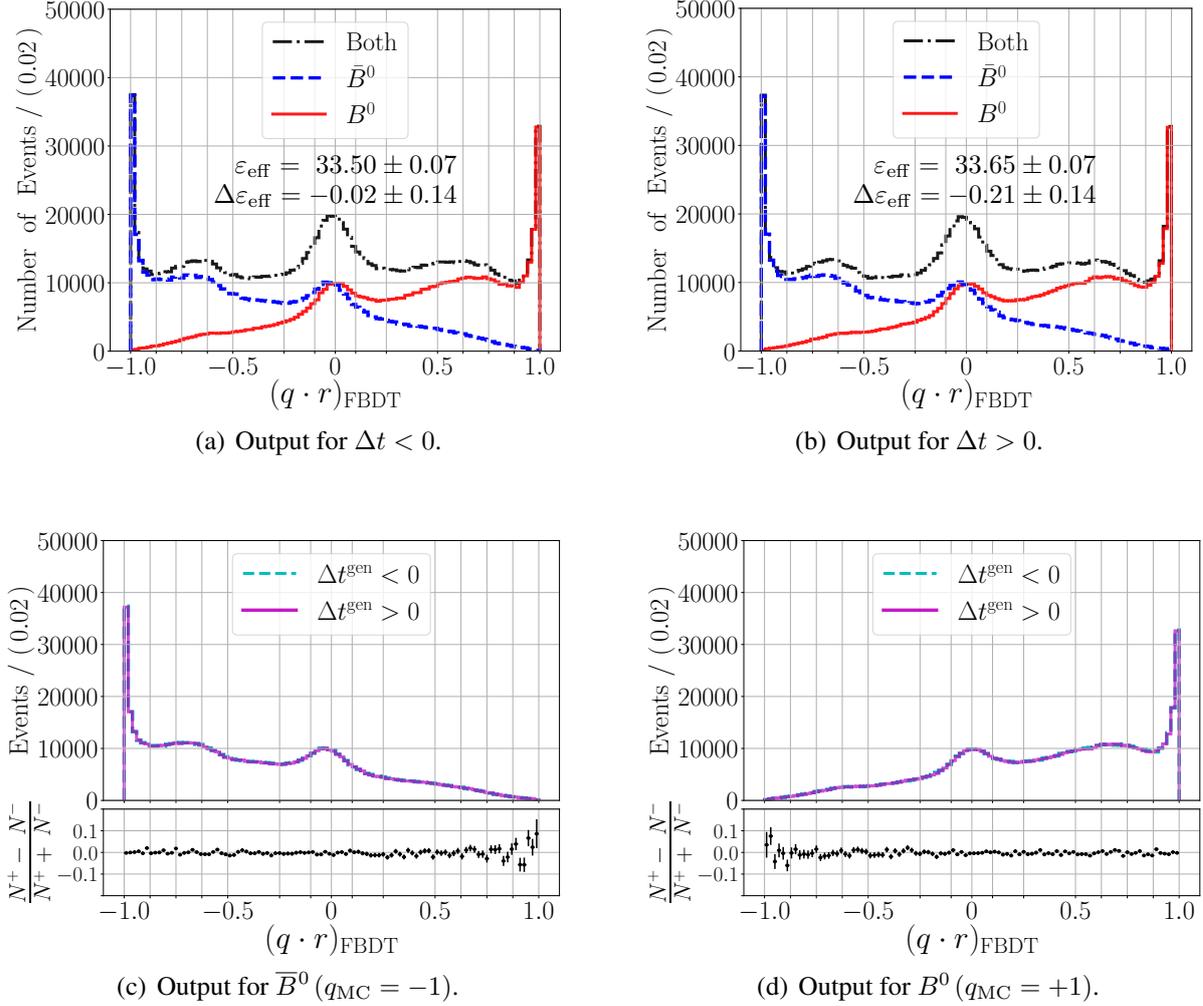


Figure B.34: Performance of the FBDT combiner avoiding the impact parameters d_0 and ξ_0 for different Δt ranges. The effective efficiencies ε_{eff} and the differences $\Delta\varepsilon_{\text{eff}}$ are given in percent. The flavor tagger was trained and tested with Belle II MC generated with $S_{CP} = 0$.

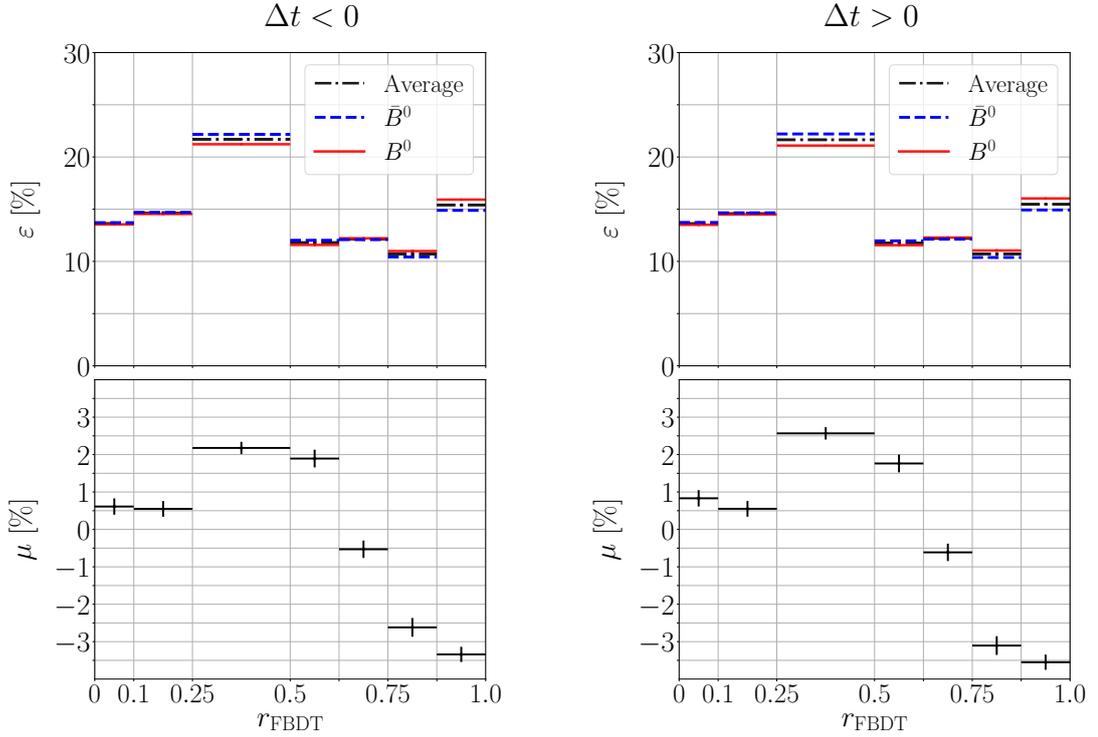
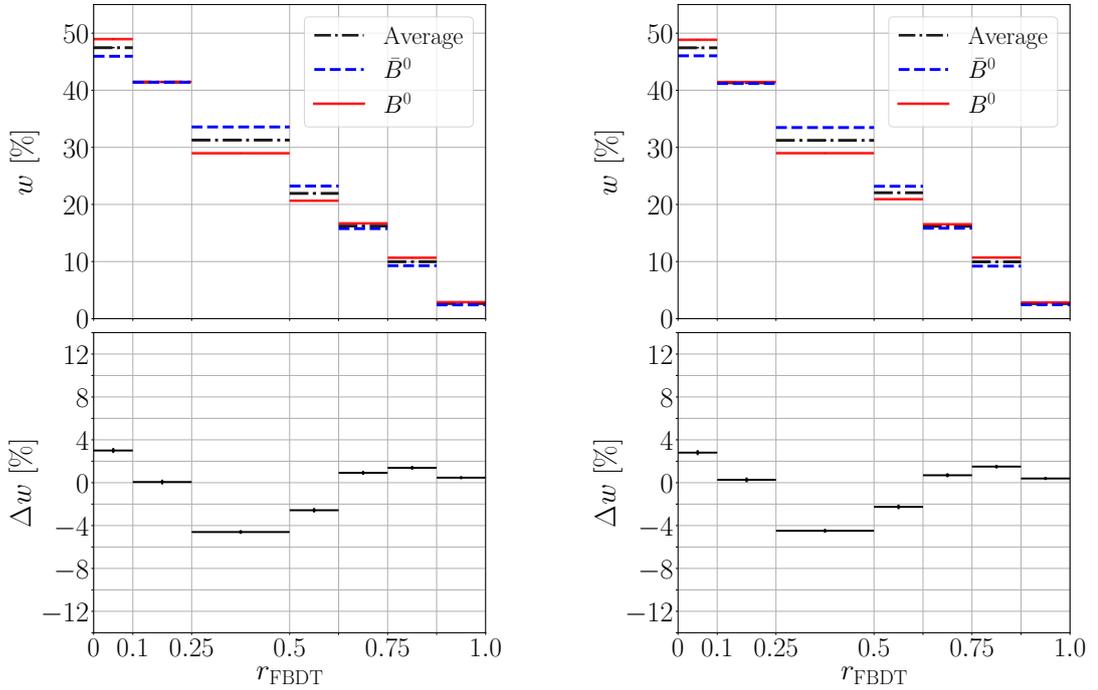
(a) ε_{B^0} , $\varepsilon_{\bar{B}^0}$, and $\mu = (\varepsilon_{B^0} - \varepsilon_{\bar{B}^0}) / (\varepsilon_{B^0} + \varepsilon_{\bar{B}^0})$ for each r -bin.(b) w_{B^0} , $w_{\bar{B}^0}$, and $\Delta w = w_{B^0} - w_{\bar{B}^0}$ for each r -bin.

Figure B.35: Performance of the FBDT combiner avoiding the impact parameters d_0 and ξ_0 for different Δt ranges. The flavor tagger was trained and tested with Belle II MC generated with $\mathcal{S}_{CP} = 0$.

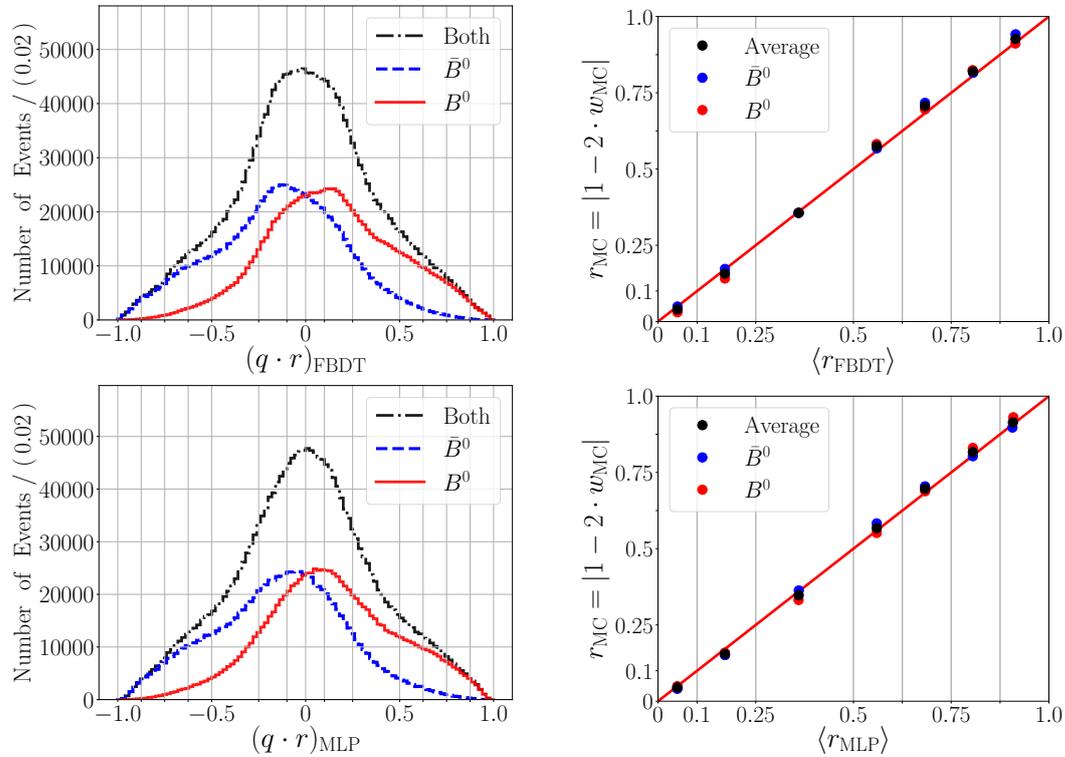
B.5 Test of Robustness

This section presents all the results of the test of robustness performed on Belle data, as described in Sect. 4.9. Using all input variables, the $q \cdot r$ distributions of the FBDT and the MLP combiners on Belle MC show good agreement with the distributions on Belle data. To evaluate the differences between the results on MC and on data using only parts of the input information, the input variables of the event-level multivariate methods were classified into four subsets according to their type (PID or kinematic), the ranking provided by the FBDT event-level methods (high rank or low rank), and the correlations between them. The subsets are: PID-high, PID-low, kinematic-high, and kinematic-low. The four single subsets, the six possible combinations of two subsets, and the four possible combinations of three subsets yield 14 different combinations of variables.

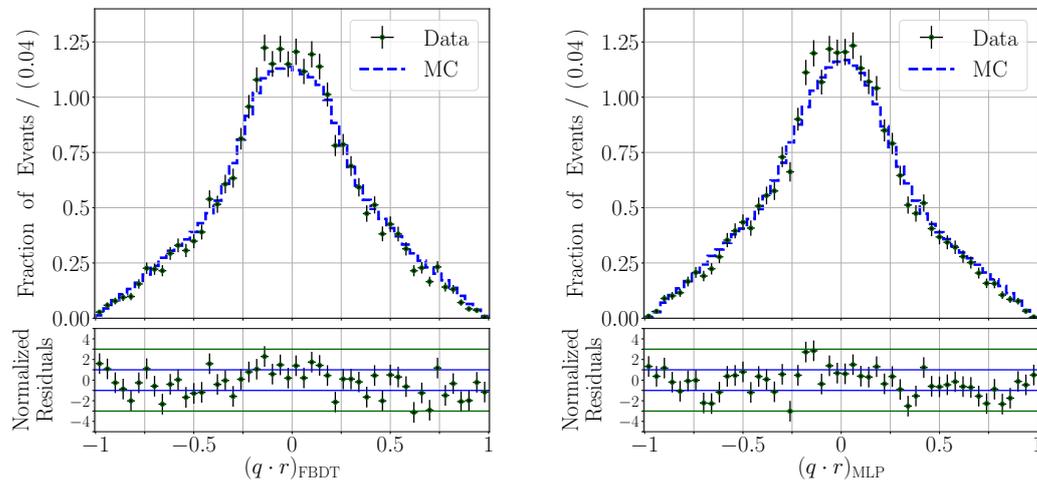
For each combination of variables, the flavor tagger is trained using Belle MC and tested on Belle MC and on Belle data. The results for the 14 different combinations of variables are compared with the results of the final algorithm which uses all the variables together. Figures B.36 to B.50 present the respective $q \cdot r$ output distributions on MC and on data, and the linearity checks between $\langle r \rangle$ and r_{MC} , for both combiners and for each combination of input variables. The results are ordered according to the increase in effective efficiency obtained by the FBDT combiner on data (see Tab. 4.8).

The linearity checks are important because the effective efficiencies on data are calculated assuming that the linearity between $\langle r \rangle$ and r_{MC} is also valid for $\langle r \rangle$ and r_{data} . As Figs. B.36 to B.50 show, for each possible combination of variables, this linearity holds.

The results show that the best performance is obtained using all the input variables as in the current algorithm. And that within the uncertainties, the $q \cdot r$ distributions of the FBDT and the MLP combiners on Belle MC agree with those obtained on Belle data for each combination of input variables.

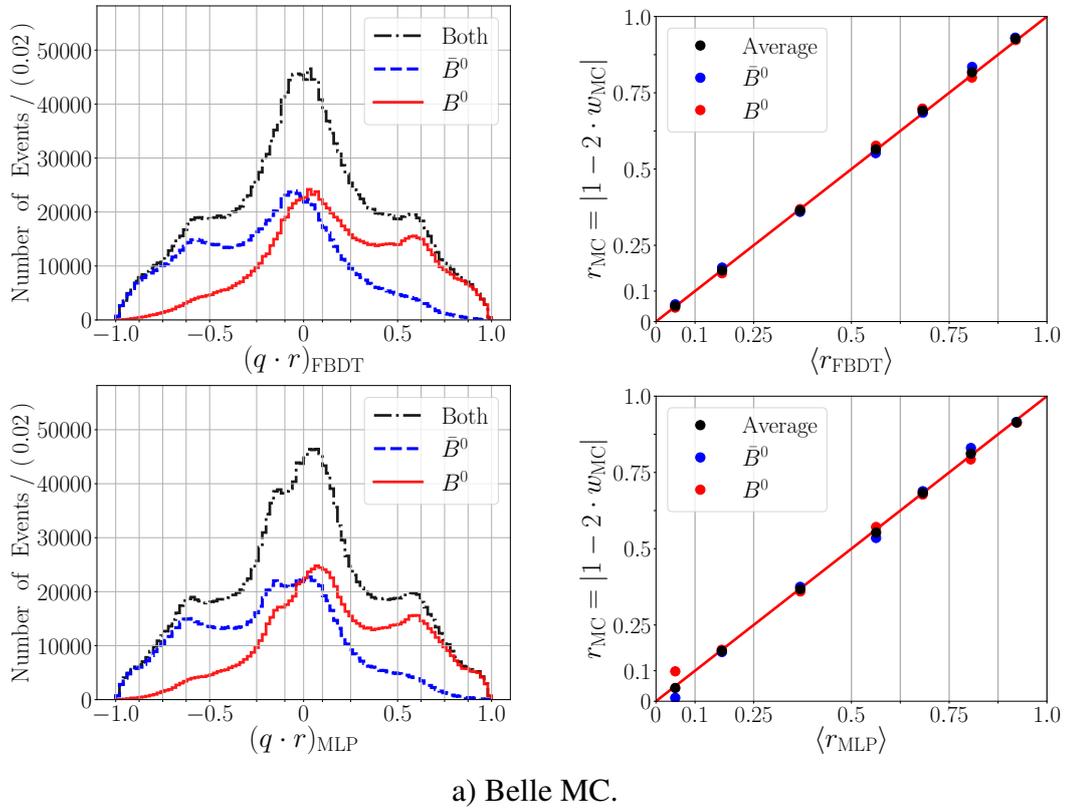


a) Belle MC.

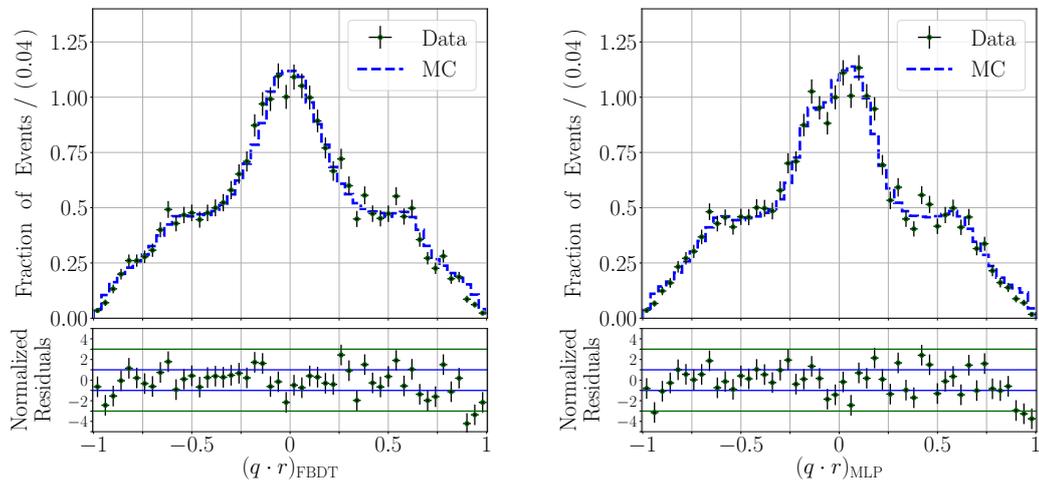


b) Belle data and Belle MC.

Figure B.36: Performance of FBDT and of MLP combiners using the kinematic-low variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.



a) Belle MC.



b) Belle data and Belle MC.

Figure B.37: Performance of FBDT and of MLP combiners using the PID-low variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

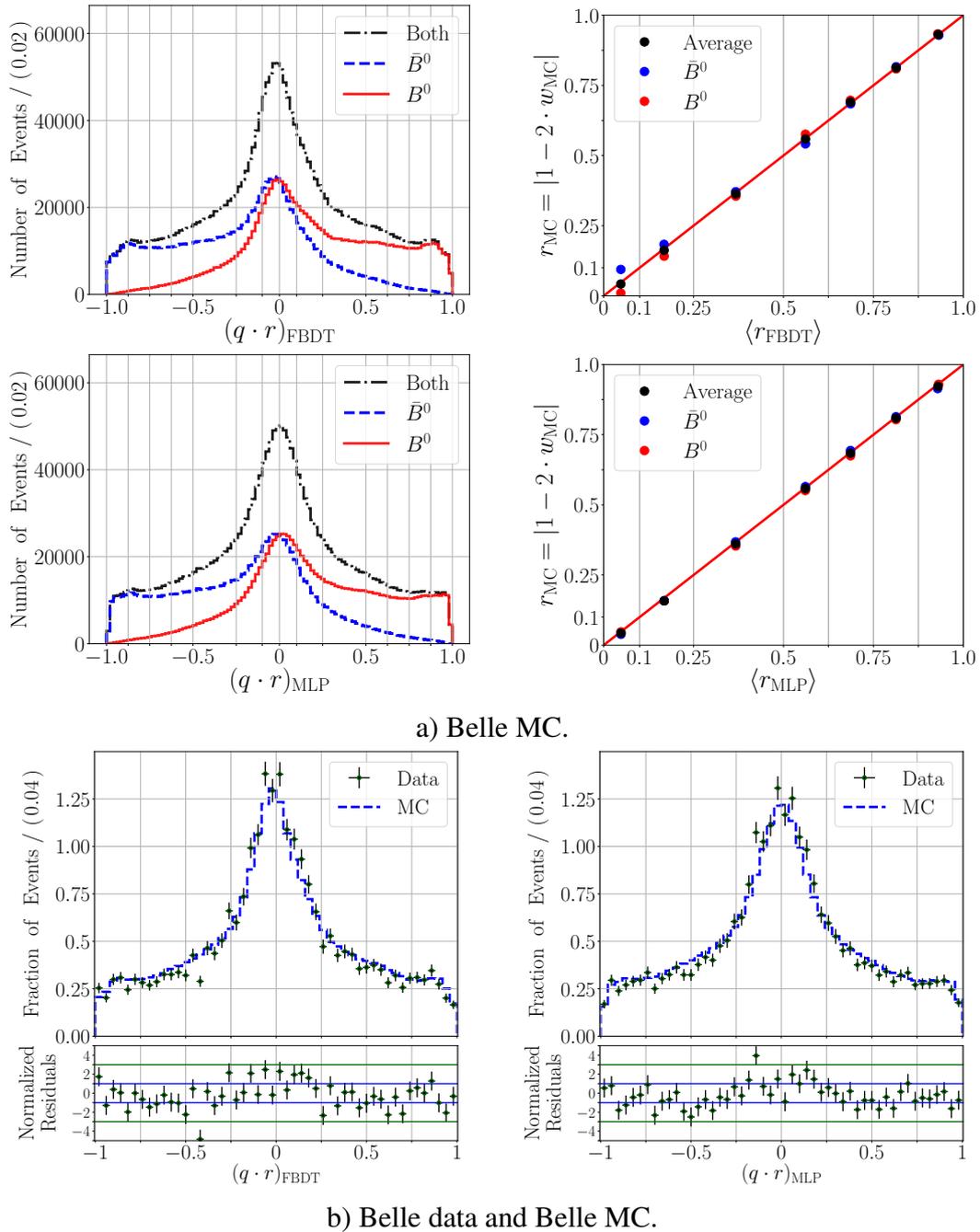


Figure B.38: Performance of FBDT and of MLP combiners using the kinematic high variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2w_{\text{MC}}|$ and $\langle q \cdot r \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

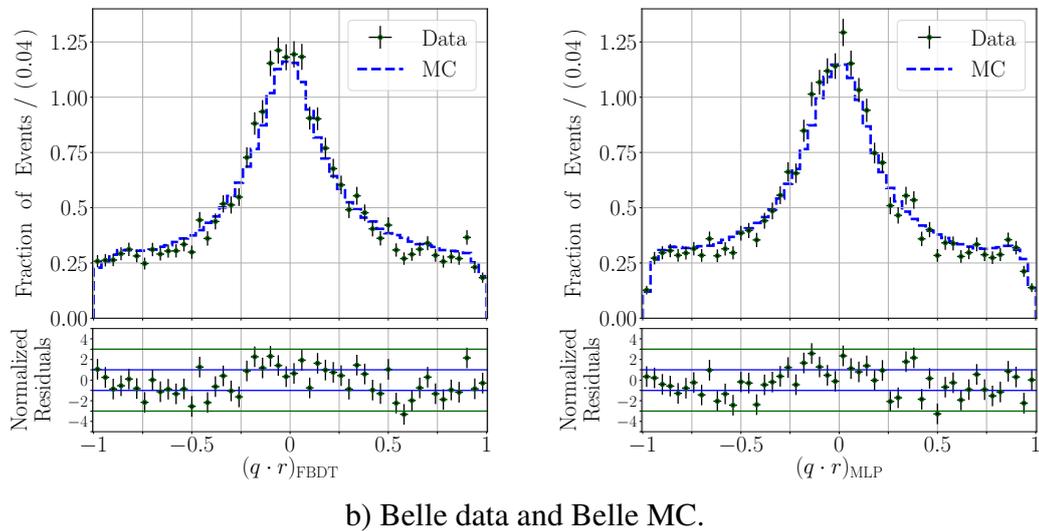
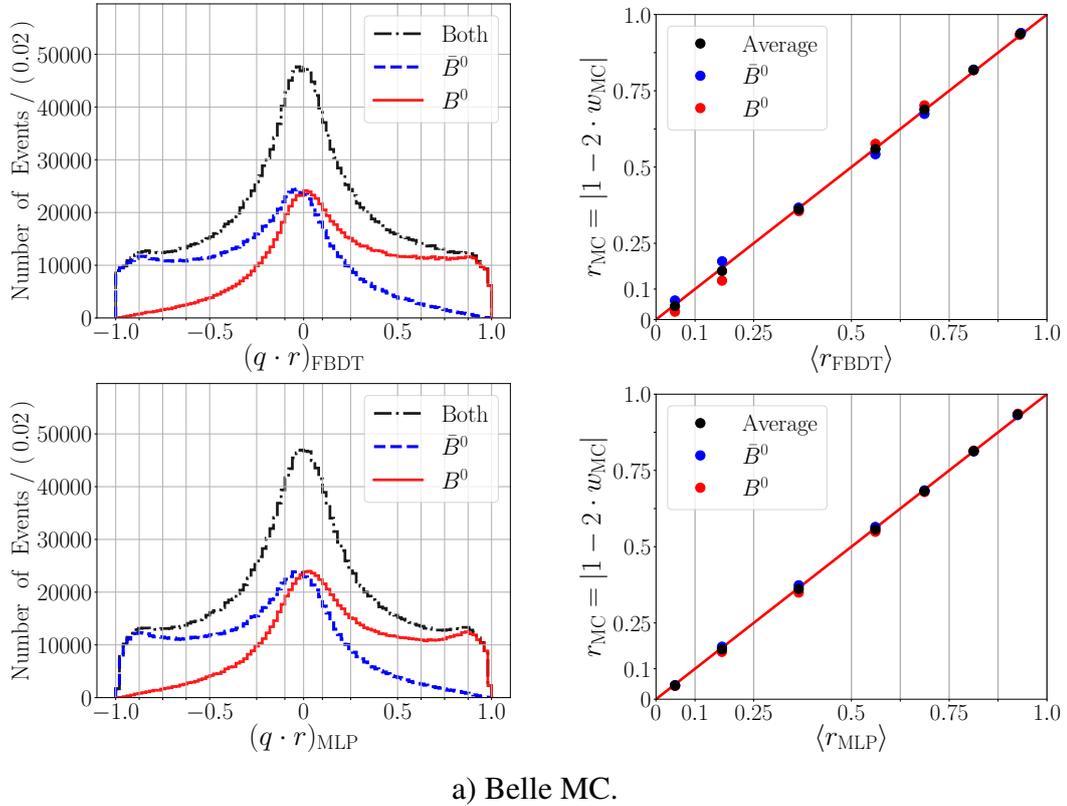


Figure B.39: Performance of FBDT and of MLP combiners using all kinematic variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

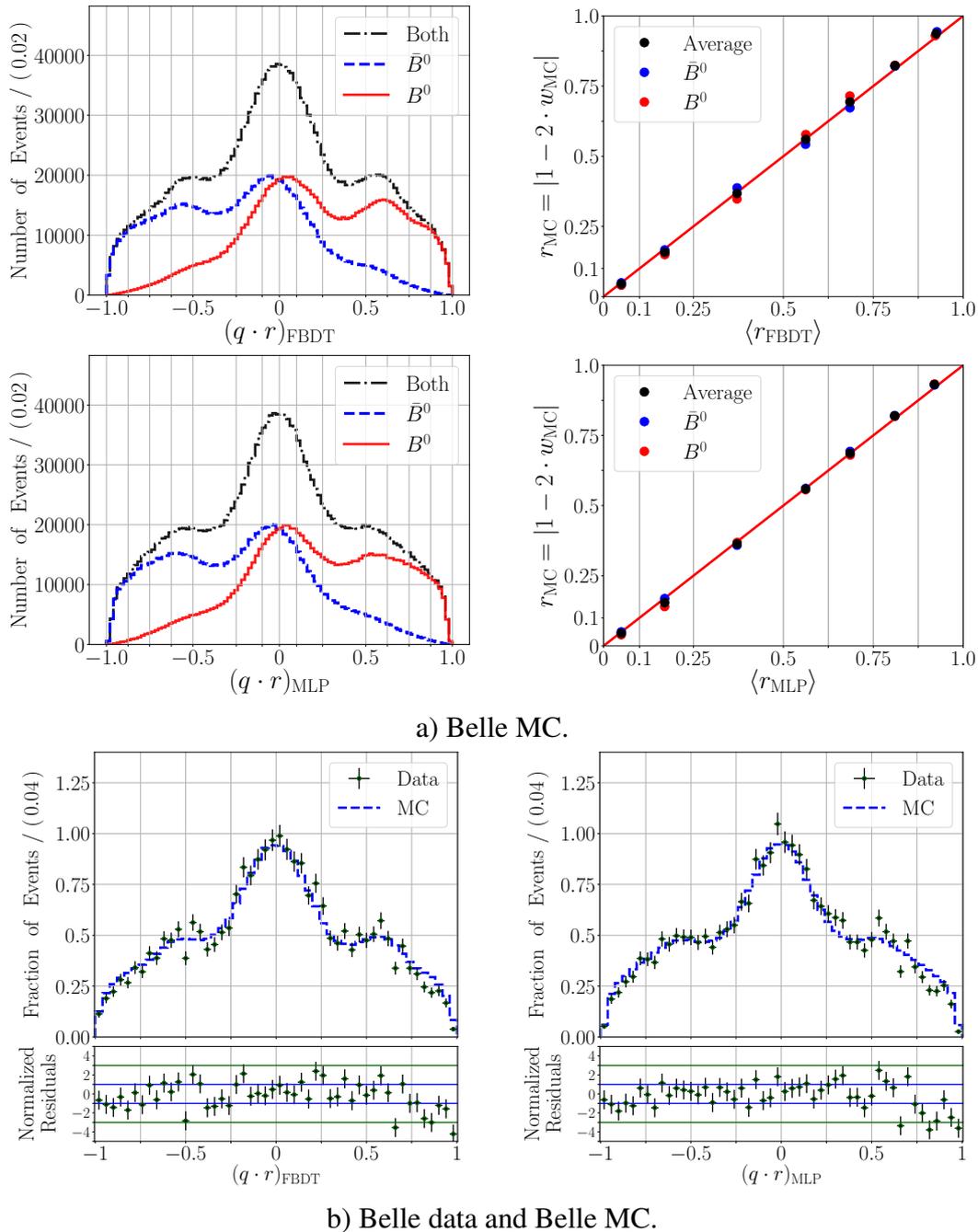
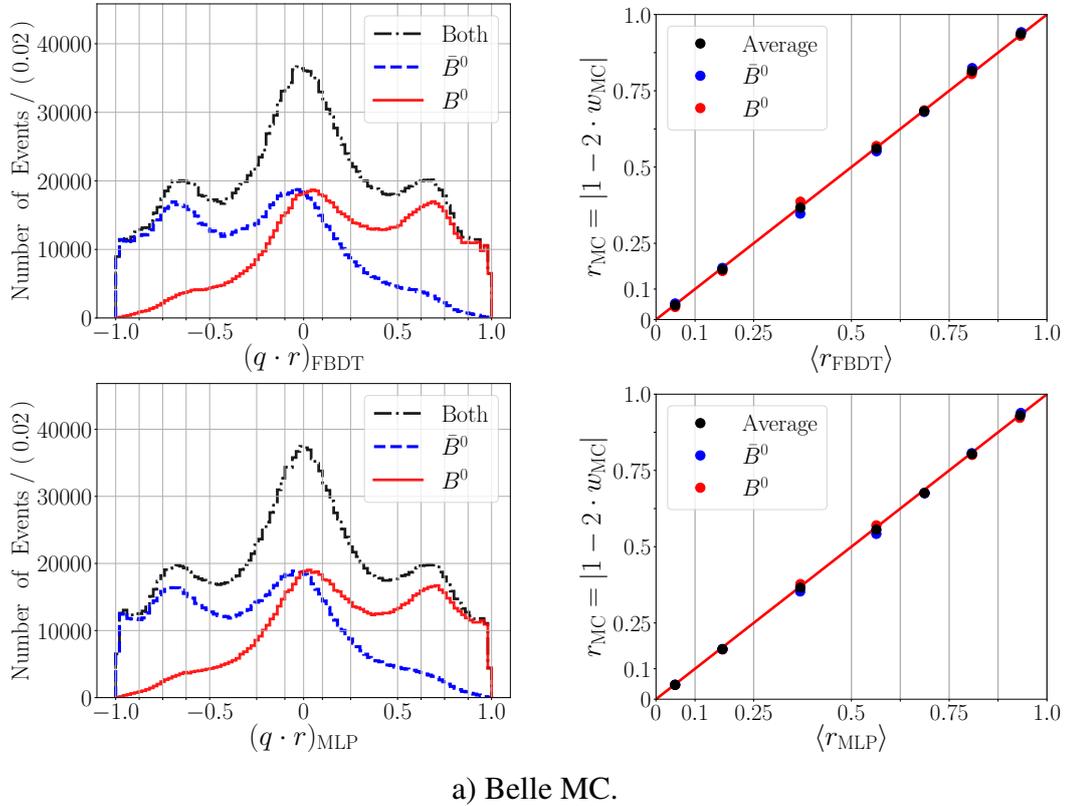
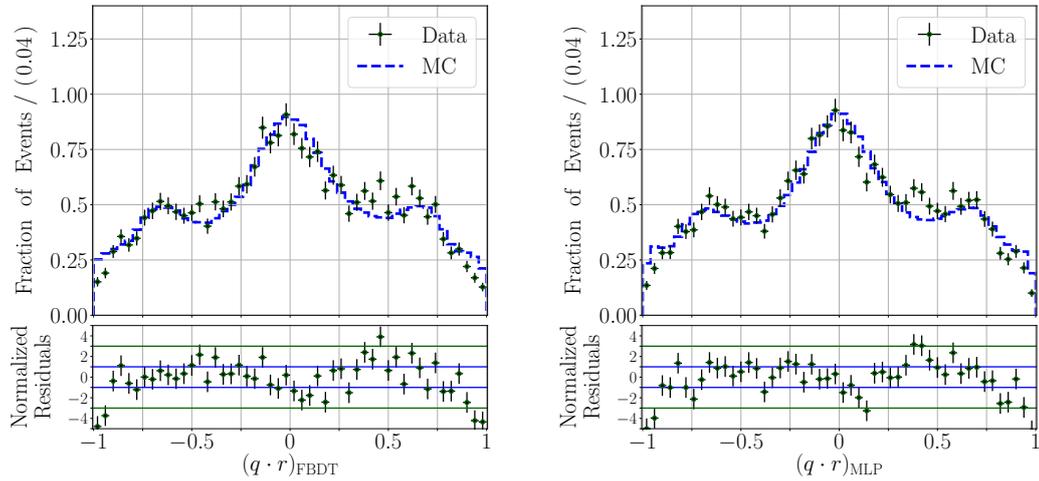


Figure B.40: Performance of FBDT and of MLP combiners using the PID-low variables together with the Kinematic-low variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2 \cdot w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.



a) Belle MC.



b) Belle data and Belle MC.

Figure B.41: Performance of FBDT and of MLP combiners using the PID high variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle q \cdot r \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

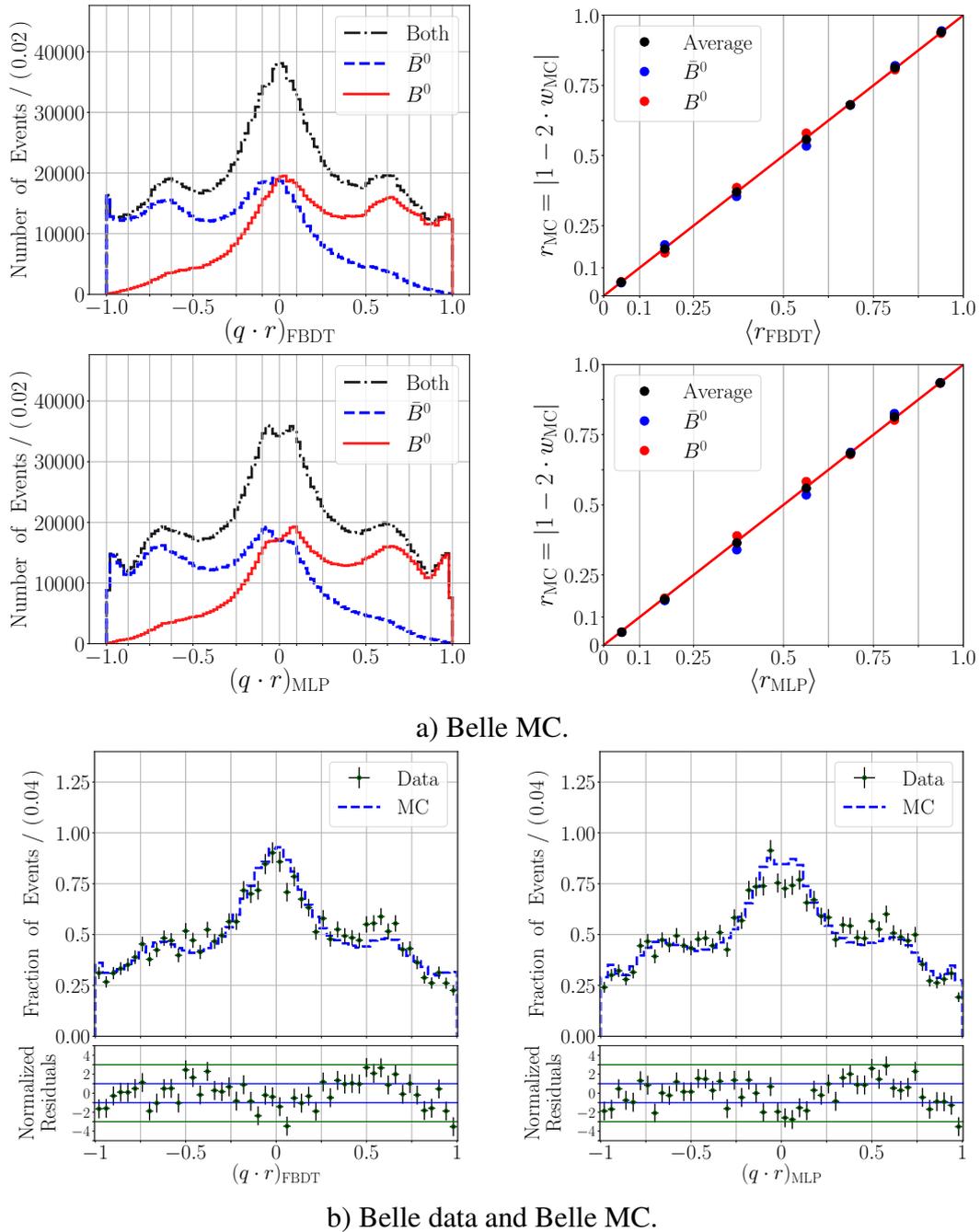


Figure B.42: Performance of FBDT and of MLP combiners using all PID variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

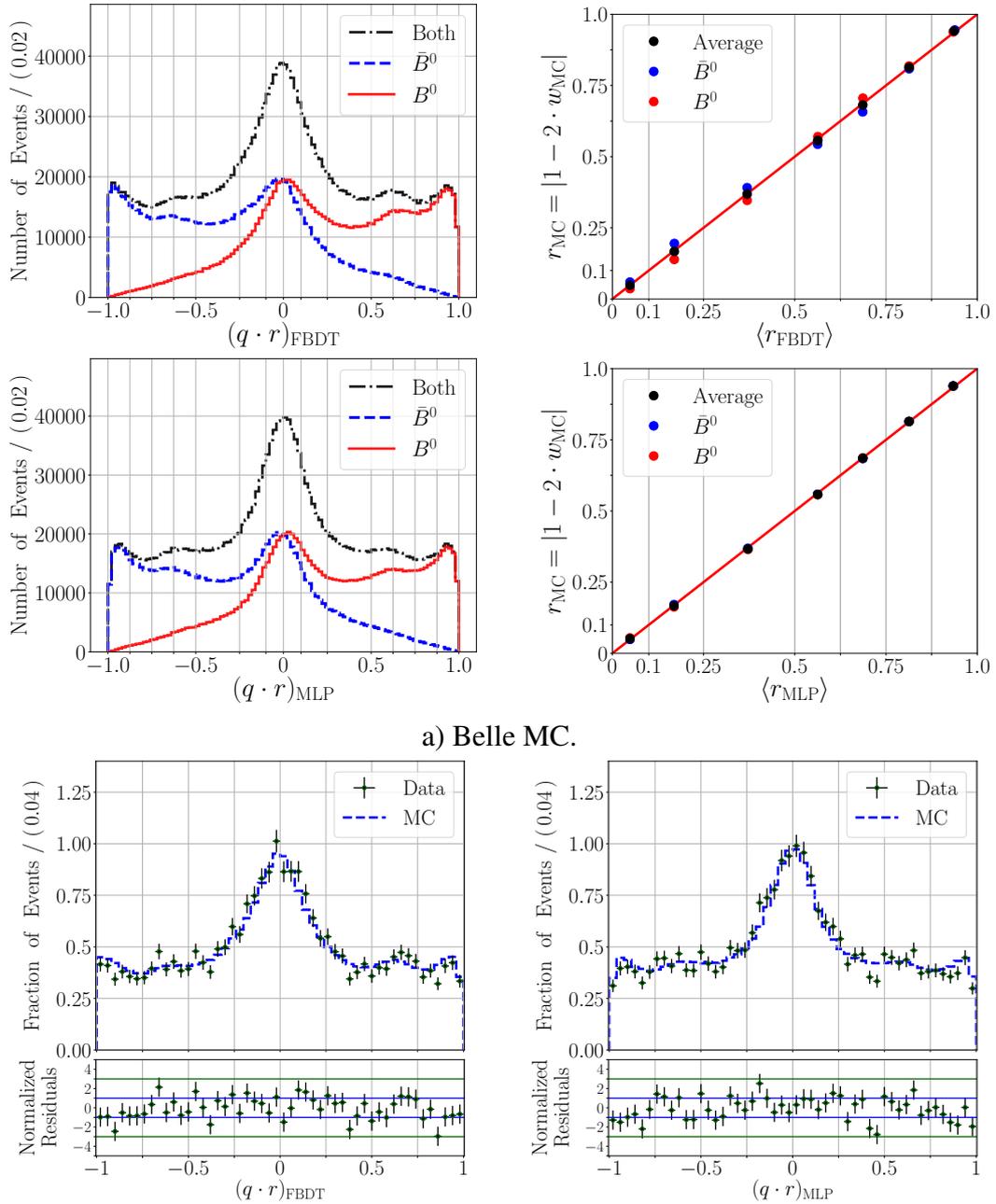
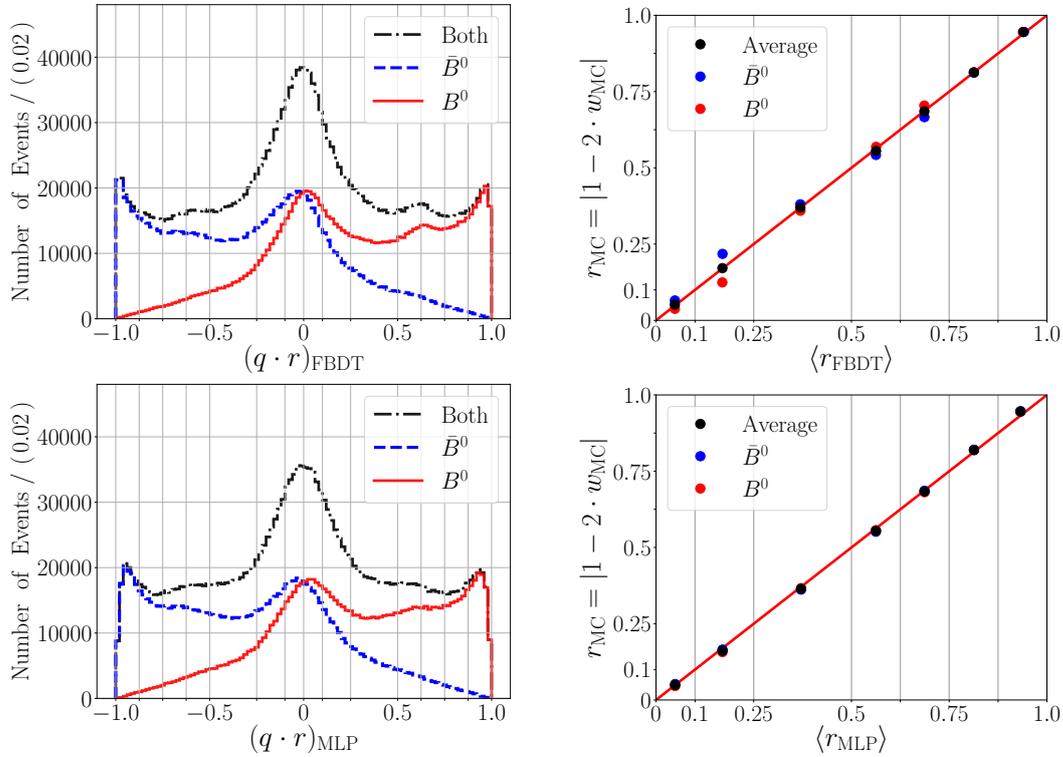
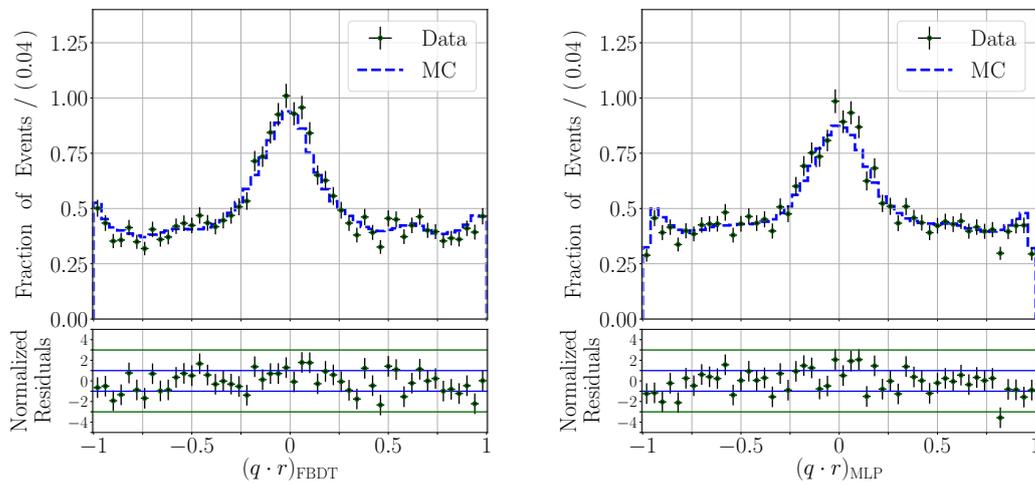


Figure B.43: Performance of FBDT and of MLP combiners using the PID-low variables together with the Kinematic high variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

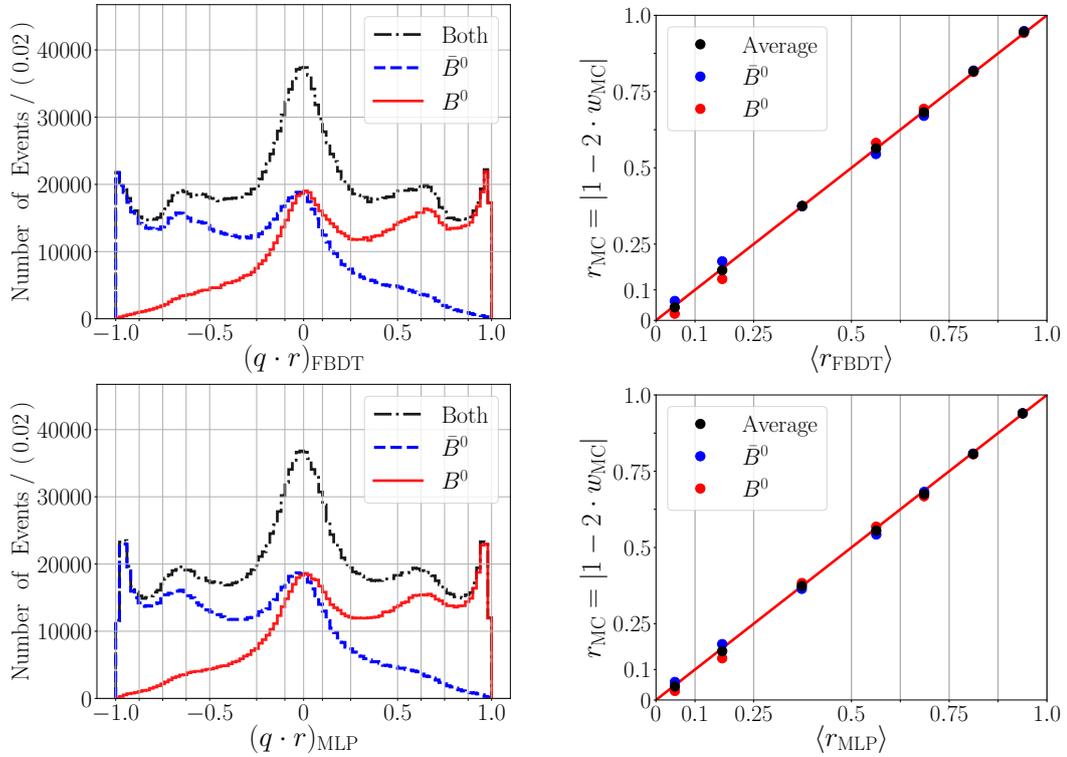


a) Belle MC.

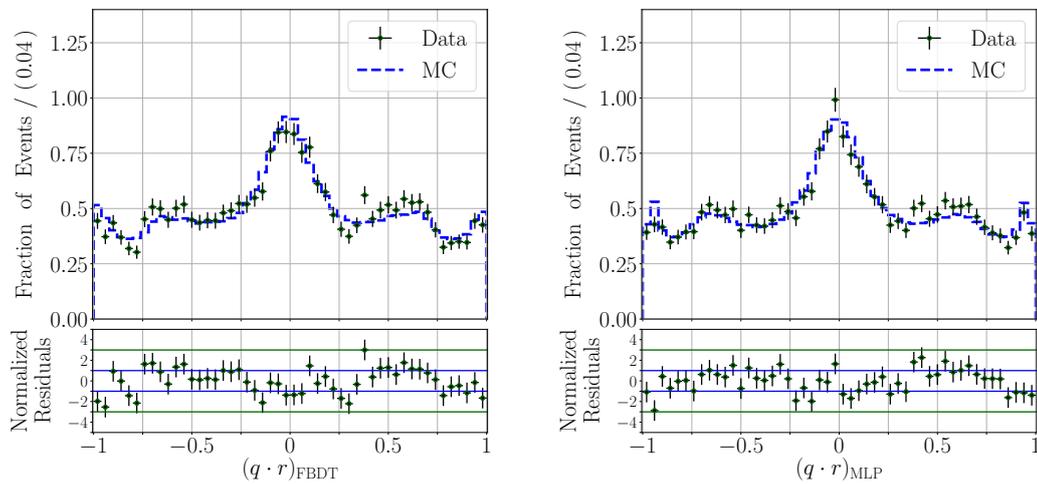


b) Belle data and Belle MC.

Figure B.44: Performance of FBDT and of MLP combiners using all kinematic variables together with the PID-low variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.



a) Belle MC.



b) Belle data and Belle MC.

Figure B.45: Performance of FBDT and of MLP combiners using the PID high variables together with the Kinematic-low variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

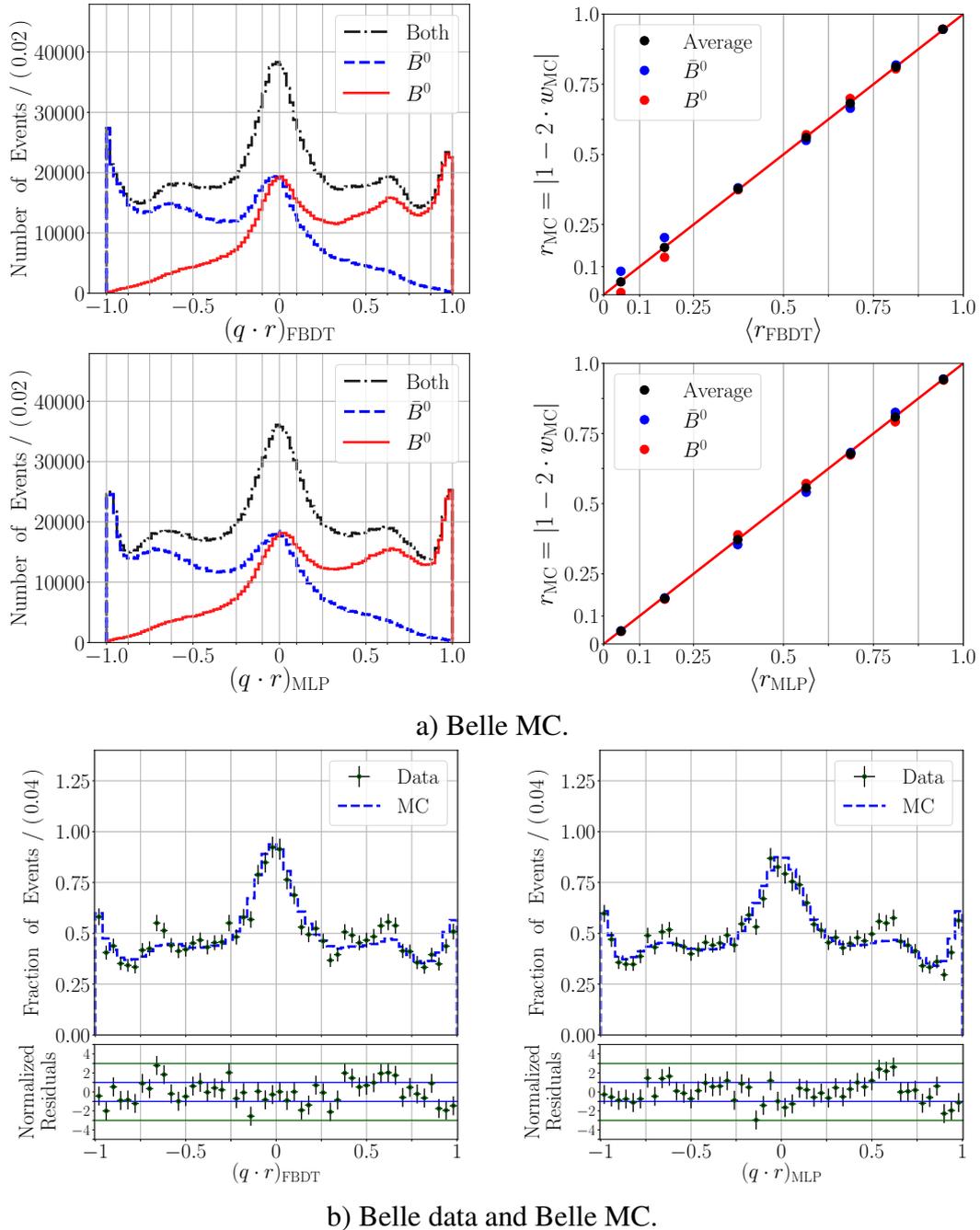
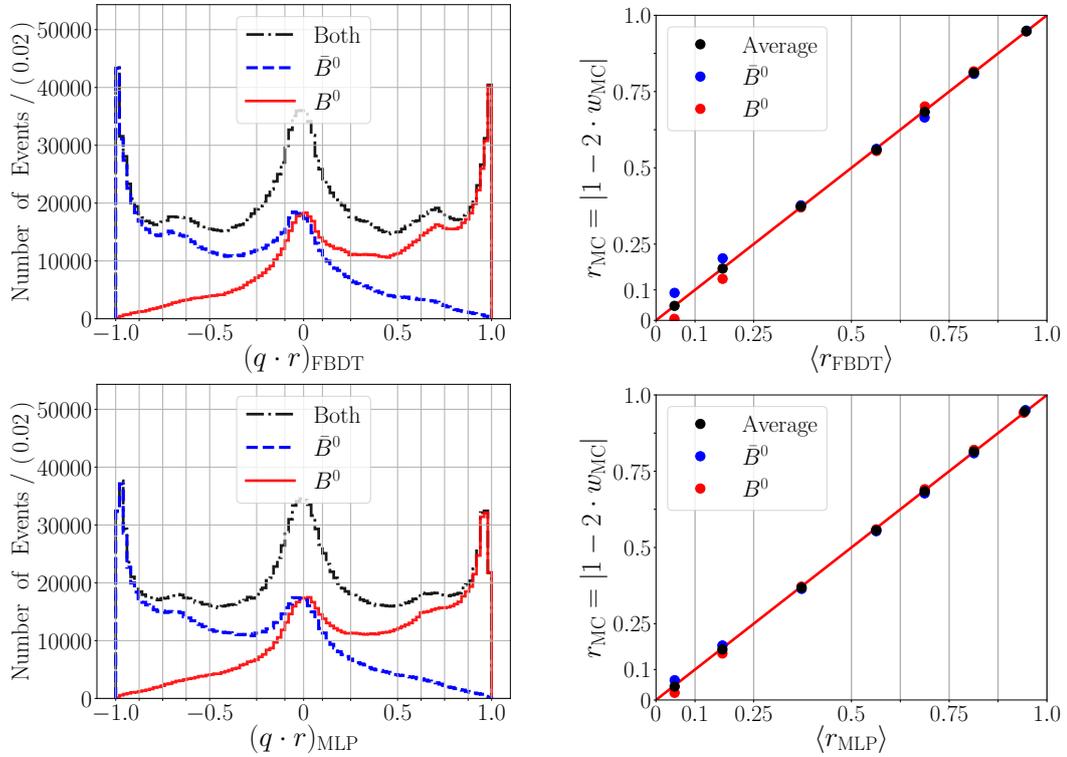
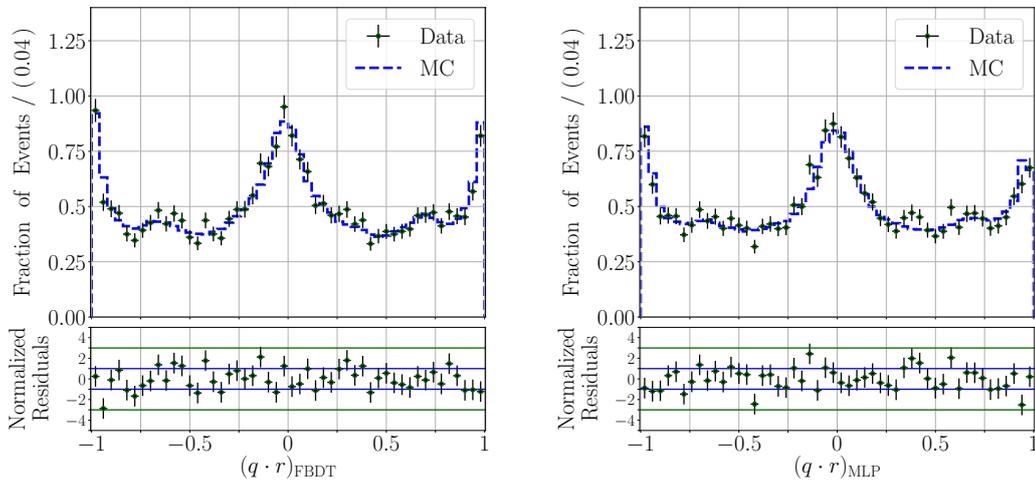


Figure B.46: Performance of FBDT and of MLP combiners using all PID variables together with the kinematic-low variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = |1 - 2 \cdot w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.



a) Belle MC.



b) Belle data and Belle MC.

Figure B.47: Performance of FBDT and of MLP combiners using the PID high variables together with the Kinematic high variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

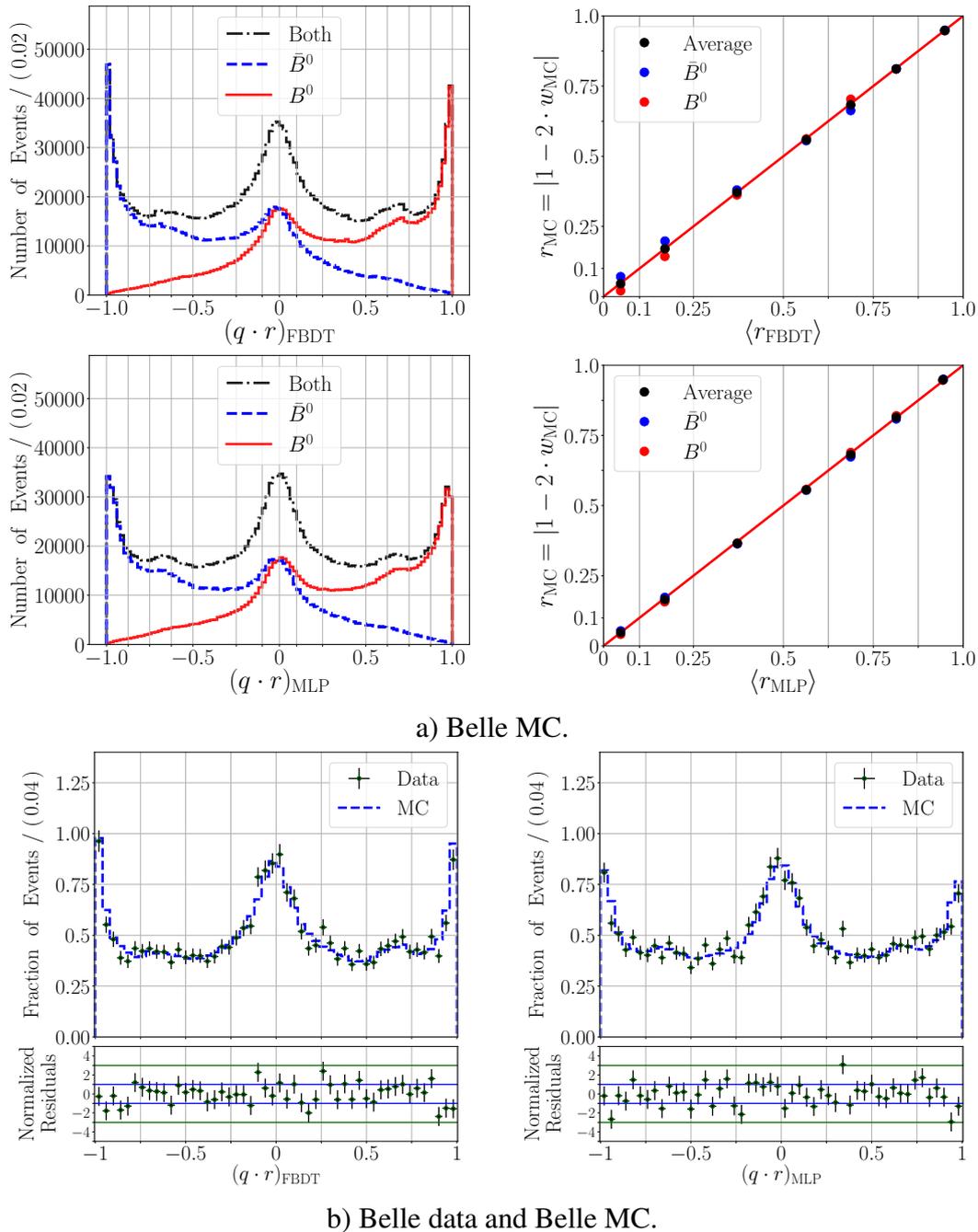
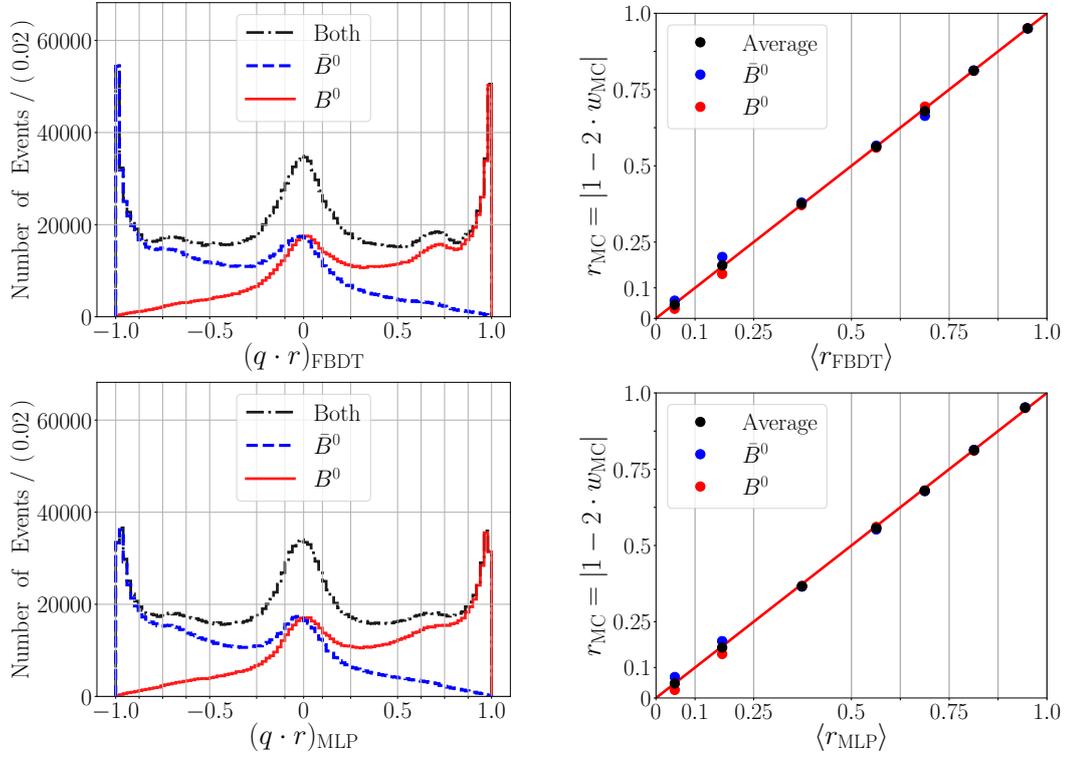
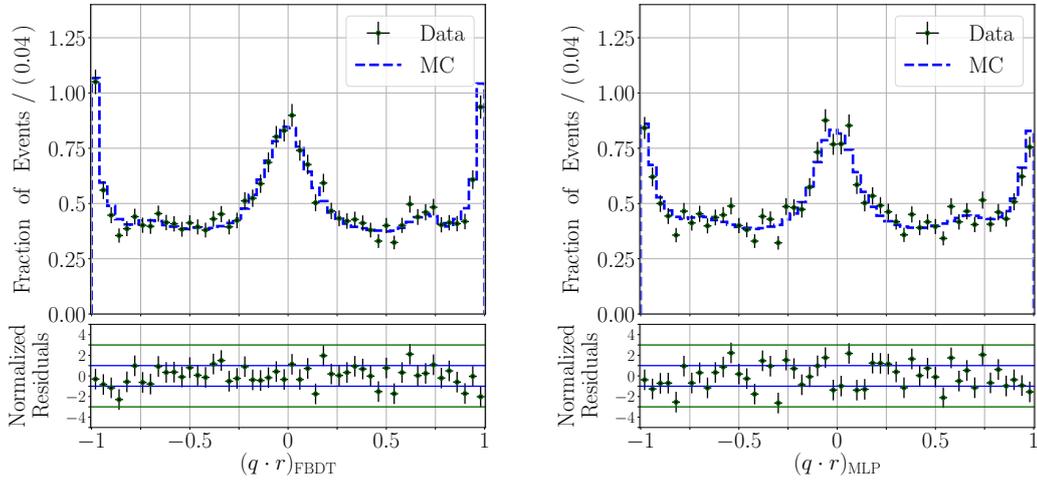


Figure B.48: Performance of FBDT and of MLP combiners using all PID variables together with the kinematic high variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{\text{MC}} = 1 - 2 \cdot w_{\text{MC}}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

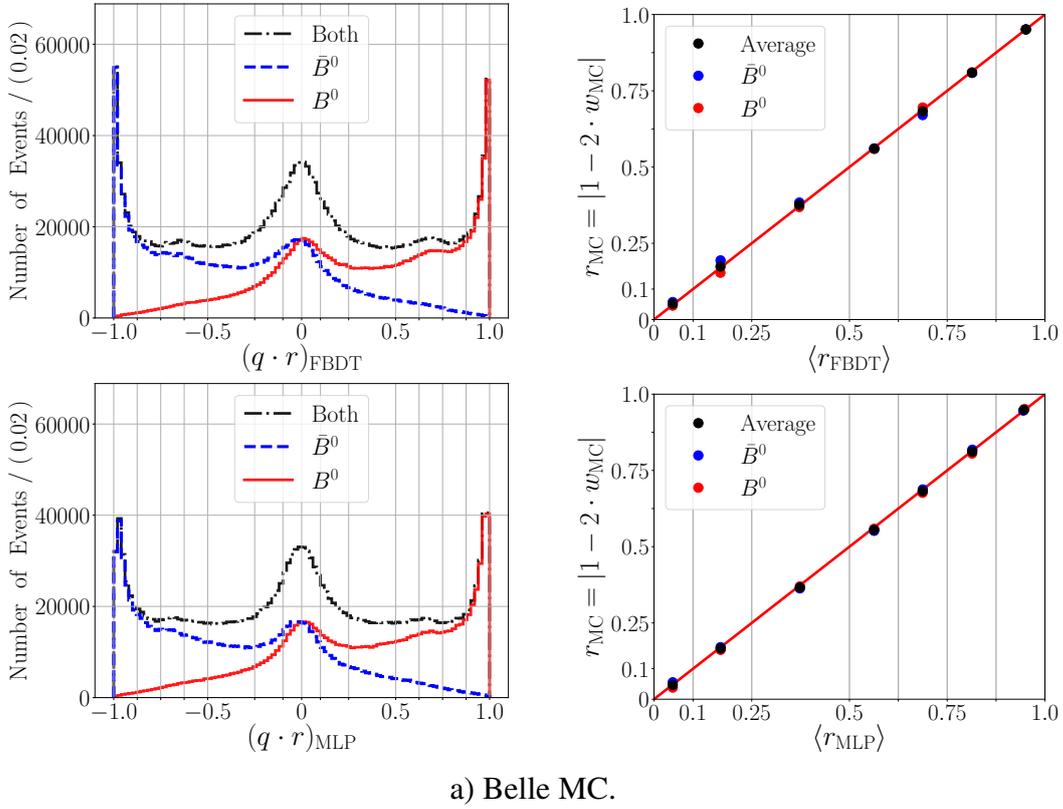


a) Belle MC.

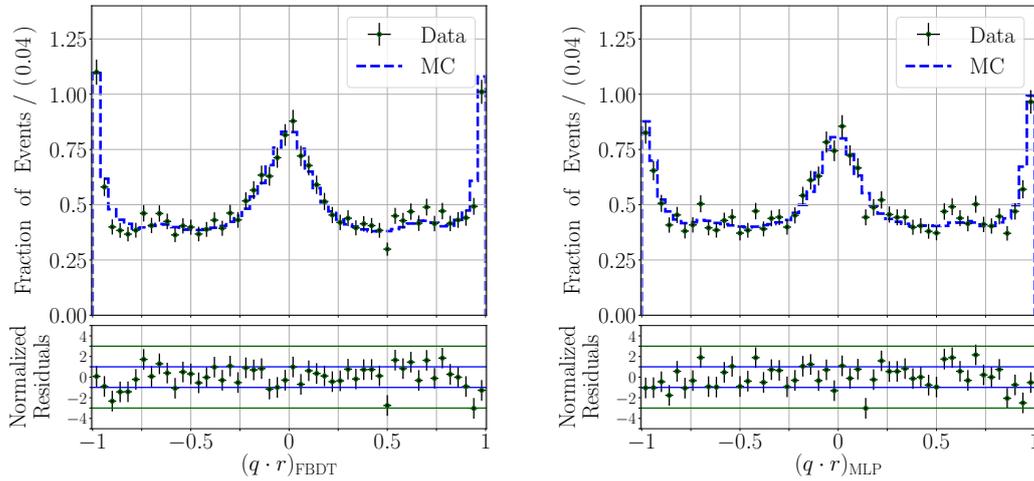


b) Belle data and Belle MC.

Figure B.49: Performance of FBDT and of MLP combiners using all kinematic variables together with the PID-high variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.



a) Belle MC.



b) Belle data and Belle MC.

Figure B.50: Performance of FBDT and of MLP combiners using all variables: a) $q \cdot r$ distributions (left) and correlations between $|r_{MC} = 1 - 2w_{MC}|$ and $\langle |q \cdot r| \rangle$ on Belle MC (right); b) normalized $q \cdot r$ distributions on Belle data and on Belle MC.

C Probability Density Functions

The probability density function (PDF) of a continuous random variable describes the relative probability that the random variable takes a given value within its phase space. A PDF is non-negative everywhere and its integral over the entire phase space is normalized to one. The PDFs used in this work are presented in the following.

C.1 Gaussian distribution

The standard Gaussian probability density function (PDF) is defined as

$$G(x; \mu, \sigma) = \frac{1}{N} \cdot \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right),$$

where μ is the mean and σ is the width of the distribution; N is a constant normalization factor. In some cases, a bifurcated Gaussian probability density function (PDF) is used

$$G_b(x; \mu, \sigma_L, \sigma_R) = \frac{1}{N} \cdot \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad \sigma = \begin{cases} \sigma_L & \text{if } x \leq \mu, \\ \sigma_R & \text{else,} \end{cases}$$

which is an asymmetric Gaussian distribution with widths σ_L and σ_R on the left or right side of the mean, respectively.

C.2 ARGUS distribution

The Argus distribution is an empirical function introduced by the ARGUS collaboration [168] to model the phase space of multi-body decays near threshold (end point). In the case of experiments at B -factories, it is commonly used to model the distribution of the beam-energy-constrained mass m_{bc} . ARGUS is defined as

$$\text{ARGUS}(x; x_0, a, c) = \frac{1}{N} \cdot x \cdot \left(1 - \frac{x^2}{x_0^2}\right)^a \exp\left(c \cdot \left(1 - \frac{x^2}{x_0^2}\right)\right) \quad \text{for } x \in [0, x_0],$$

where $x_0 > 0$ is the end point parameter, a is the power parameter (usually $a = \frac{1}{2}$), $c < 0$ is the curvature parameter, and N is a normalization factor. In the case of experiments at B -factories, the cutoff parameter for the distribution of the beam constrained mass m_{bc} is the total beam energy $E_{\text{beam}}^* = \sqrt{s}/2$ in the $\Upsilon(4S)$ frame, which is an event dependent variable. Thus, the

following notation is used

$$\text{ARGUS}(m_{bc}; a, c) = \text{ARGUS}(m_{bc}; E_{\text{beam}}^*, a, c)$$

C.3 Crystal Ball function

The Crystal Ball function, named after the Crystal Ball collaboration, is a probability density function (PDF) commonly used in physics to model distributions featuring a low-end tail. It consists of a Gaussian core component and a power-law tail, below a certain threshold $-\alpha$. The function is given by

$$\text{CB}(x; \alpha, n, \mu, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\mu}{\sigma})^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha \end{cases},$$

where

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right),$$

$$B = \frac{n}{|\alpha|} - |\alpha|,$$

and N is a normalization factor

$$N = \frac{1}{\sigma(C + D)},$$

$$C = \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot \exp\left(-\frac{|\alpha|^2}{2}\right),$$

$$D = \sqrt{\frac{\pi}{2}} \left(1 + \text{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right).$$

C.4 Kernel density estimation

The kernel density estimation is a non-parametric way to estimate the PDF of a random variable. In comparison with a histogram, a kernel density estimation has the advantage that it allows a continuous and statistically consistent estimation of a PDF through a series of density functions. Given a univariate independent and identically distributed sample x_1, \dots, x_n drawn from a distribution with unknown density f , the shape of f is estimated as

$$\hat{f}_0(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

where K is the kernel and $h > 0$ is a smoothing parameter called bandwidth. The kernel is a non-negative function that integrates to one. In the equation above, h is constant for all i . This approach is thus called fixed kernel estimation. In this work, the used algorithm performs a second iteration, where the bandwidth is modified depending on the local event density: an approach called adaptive kernel estimation. The kernels are Gaussian functions with equal surface. The estimated shape of f is then

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{x - x_i}{h_i}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \cdot \frac{(x - x_i)^2}{h_i^2}\right),$$

with

$$h_i = \left(\frac{4}{3}\right)^{\frac{1}{5}} \sqrt{\frac{\sigma}{\hat{f}_0(x_i)}} n^{-\frac{1}{5}},$$

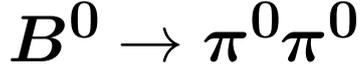
where σ is the standard deviation of the data. Further details of the general algorithm are explained in [169]. The estimated PDF of the random variable is

$$\text{Key}(x) = \frac{\hat{f}_1(x)}{N}, \quad (\text{C.1})$$

where N is a normalization factor.

Fixed and adaptive kernel estimates assume that the domain of f is all of \mathbb{R} . However, the output of a multi-variate method is usually bounded, e.g. $0 < x < 1$. Boundary conditions are then required to properly normalize $\hat{f}_1(x)$ without underestimating the value of the distribution close to the boundaries [169]. In this work, the boundary problem is solved by mirroring the data set about the boundaries.

D Performance study of the decay



D.1 Distributions of the pull and the residuals

This section shows the distributions of the pull and the residuals of the pseudo-experiments performed for the six different combinations of input values $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ (S. Sect. 6.10.6).

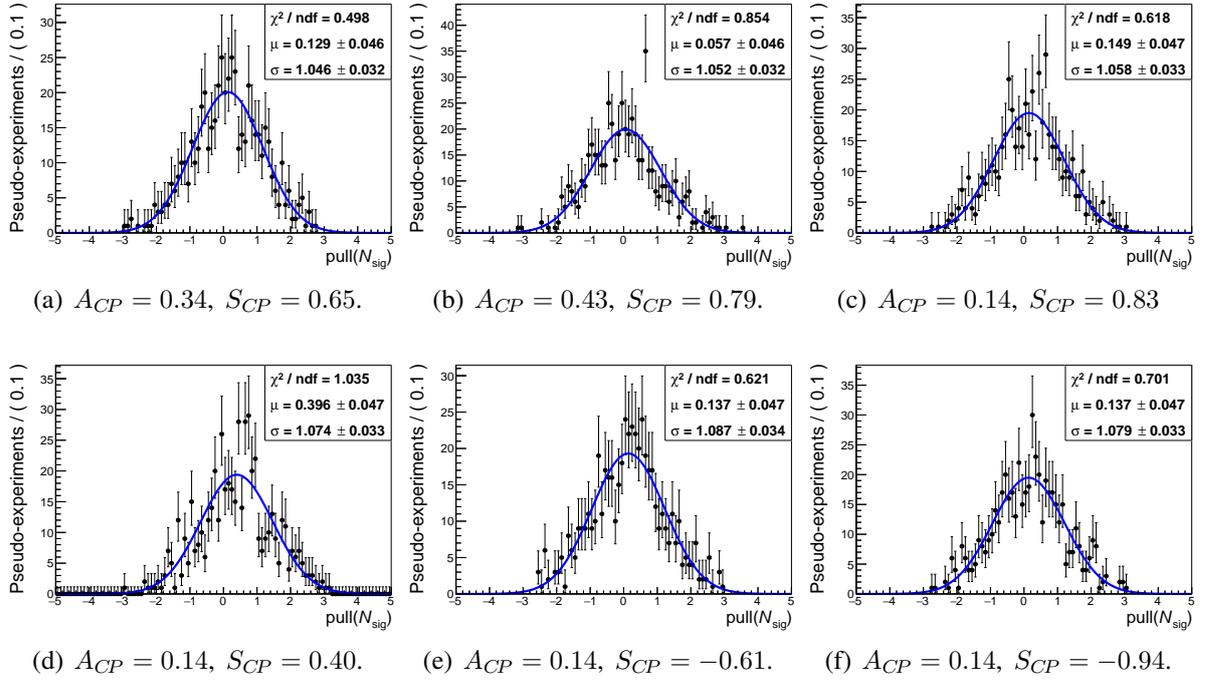


Figure D.1: Pull distributions for the yield of signal events. The input value for these pseudo-experiments is $N_{\text{sig}} = 15068$. The input values for $\mathcal{A}^{\pi^0\pi^0}$ and $\mathcal{S}^{\pi^0\pi^0}$ are shown below.

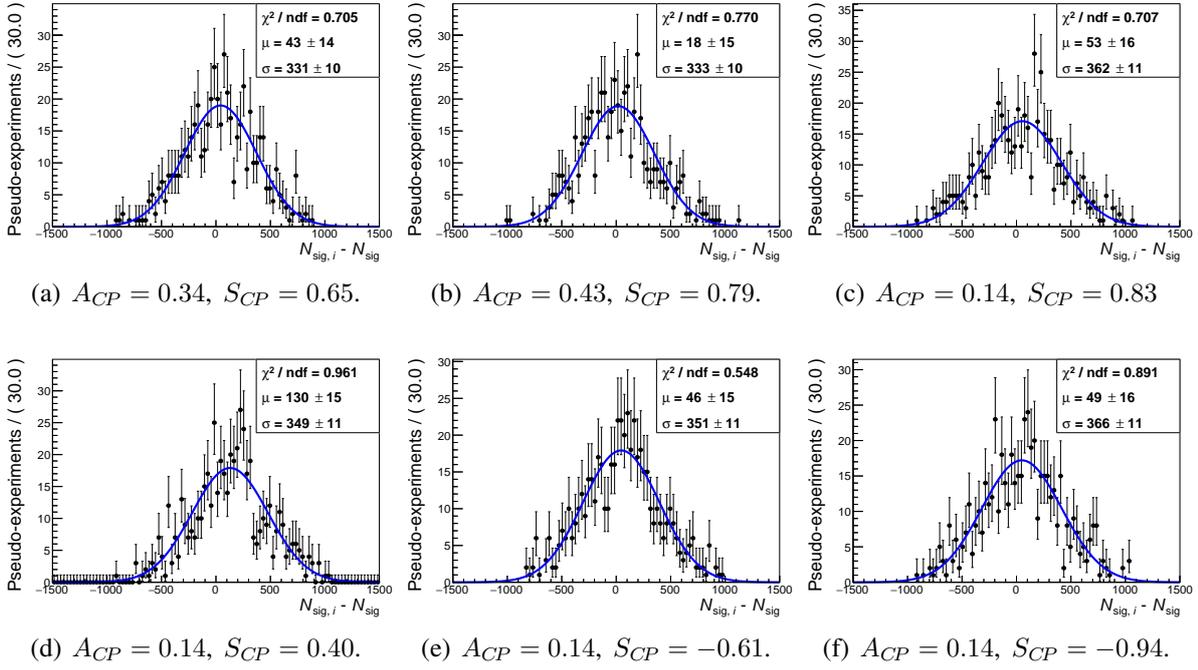


Figure D.2: Distributions of the residuals for the yield of signal events. The input value for these pseudo-experiments is $N_{\text{sig}} = 15068$.

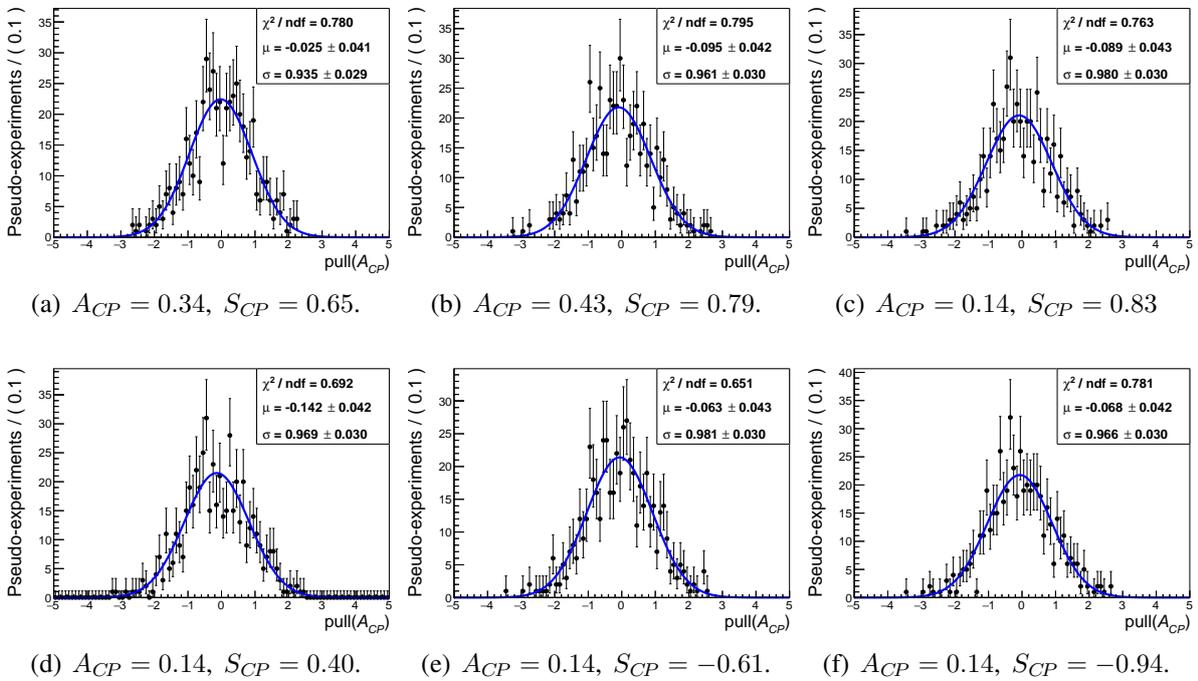


Figure D.3: Pull distributions for the CP -violation parameter $\mathcal{A}^{\pi^0 \pi^0}$ (time-integrated measurement). The input values for $\mathcal{A}^{\pi^0 \pi^0}$ and $\mathcal{S}^{\pi^0 \pi^0}$ are shown below.

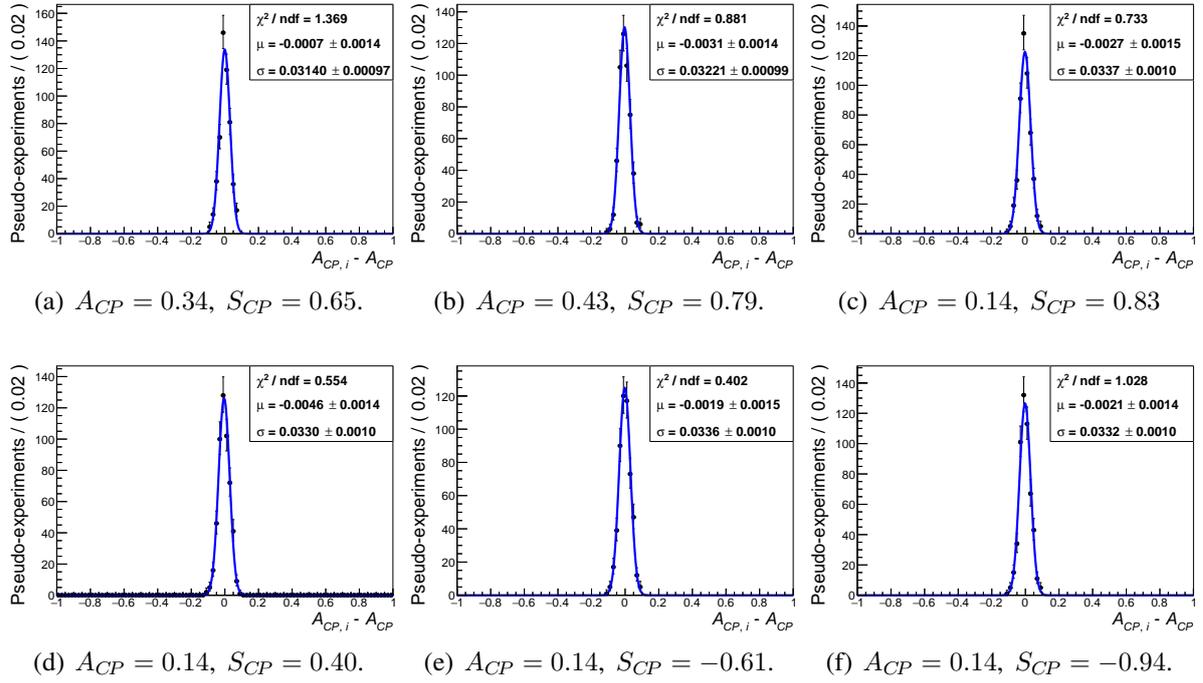


Figure D.4: Distributions of the residuals for the CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ (time-integrated measurement). The input values for $\mathcal{A}^{\pi^0\pi^0}$ and $S^{\pi^0\pi^0}$ are shown below.

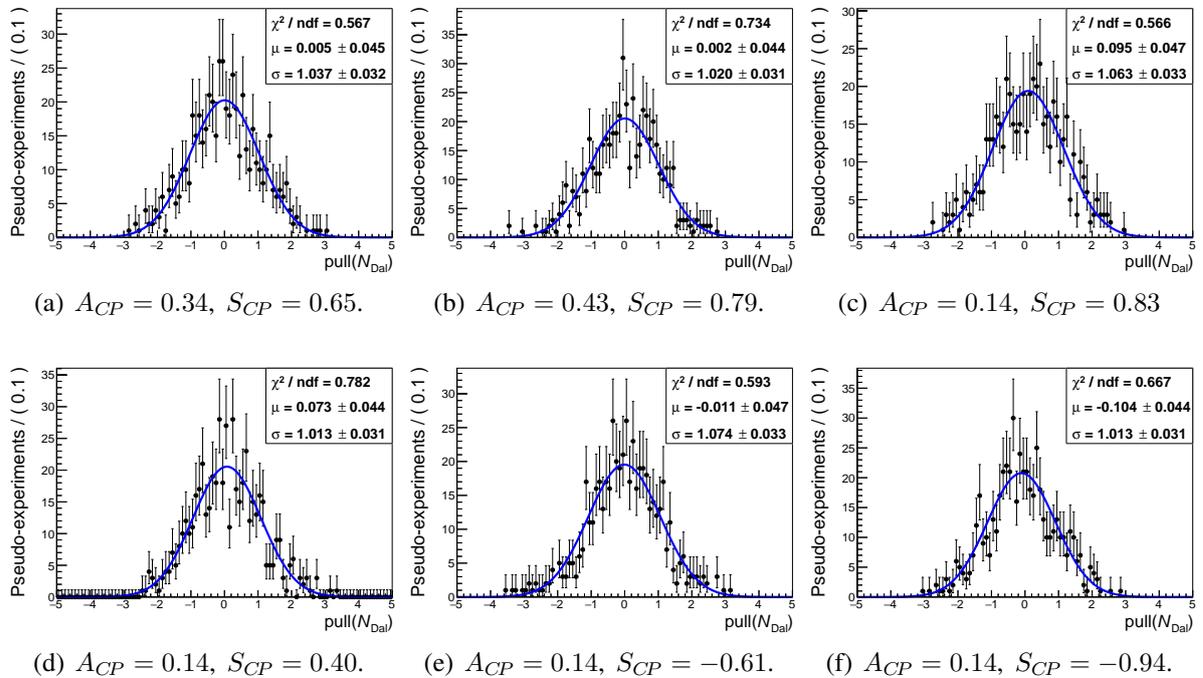


Figure D.5: Pull distributions for the yield of Dalitz events. The input value for these pseudo-experiments is $N_{\text{Dal}} = 147$.

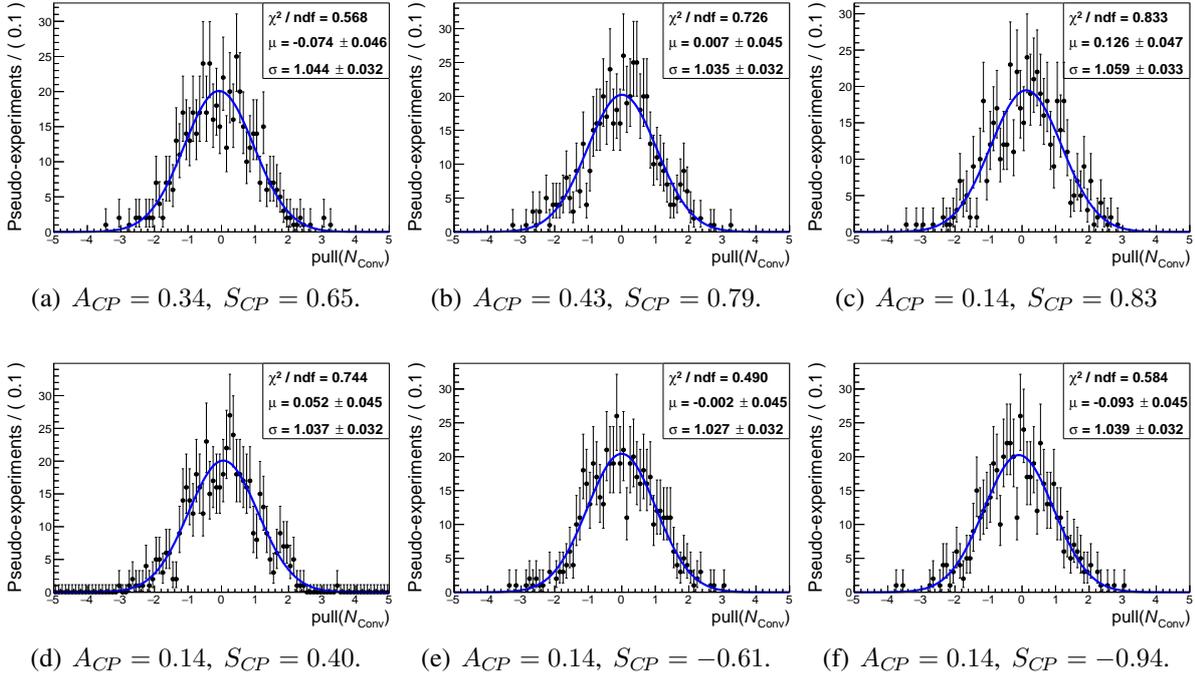


Figure D.6: Pull distributions for the yield of conversion events. The input value for these pseudo-experiments is $N_{\text{Conv}} = 124$.

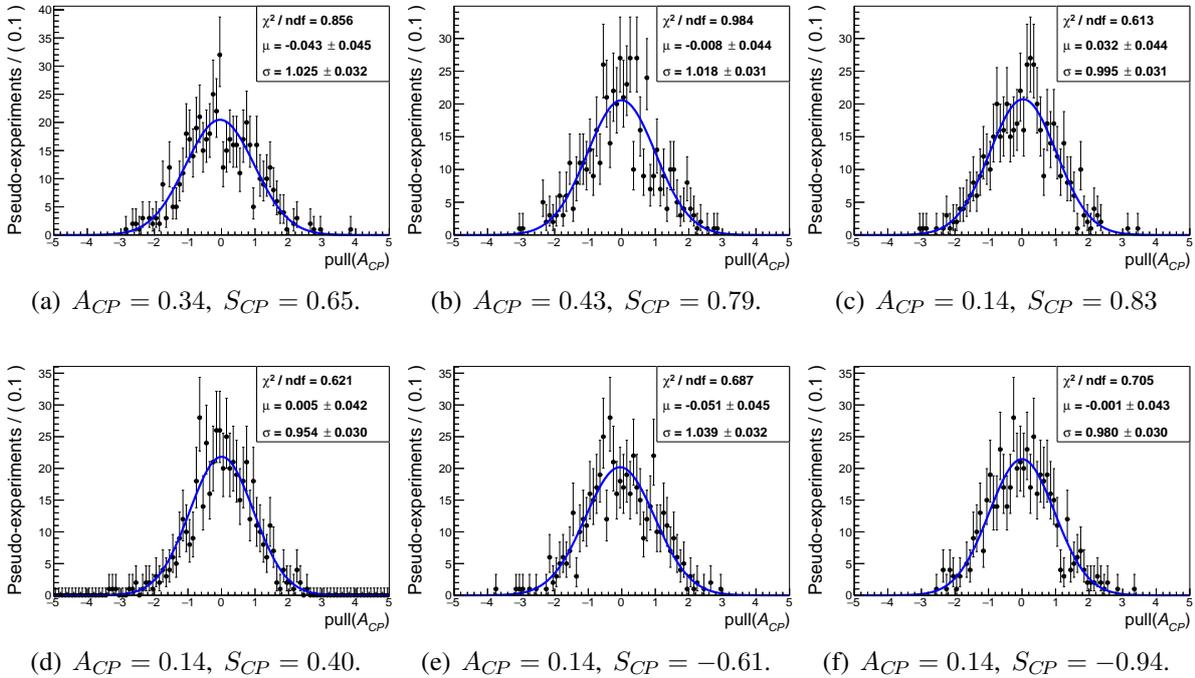


Figure D.7: Pull distributions for the CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ (time-dependent measurement). The input values for $\mathcal{A}^{\pi^0\pi^0}$ and $S^{\pi^0\pi^0}$ are shown below.

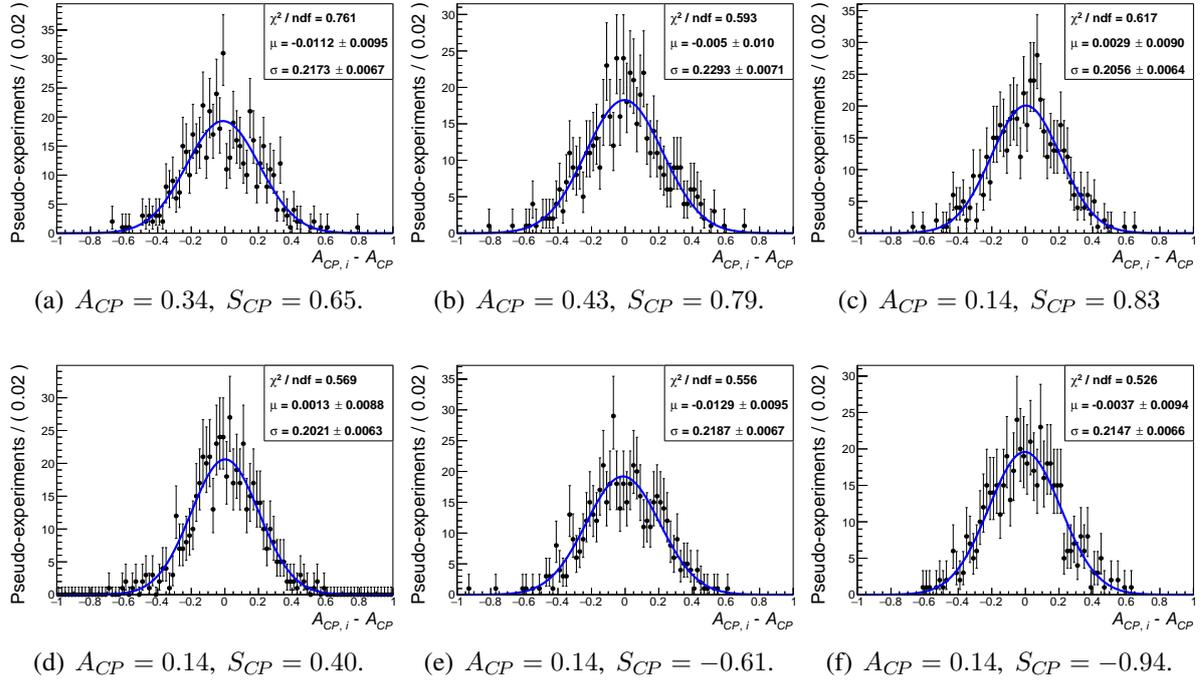


Figure D.8: Distributions of the residuals for the CP -violation parameter $\mathcal{A}^{\pi^0\pi^0}$ (time-dependent measurement). The input values for $\mathcal{A}^{\pi^0\pi^0}$ and $S^{\pi^0\pi^0}$ are shown below.

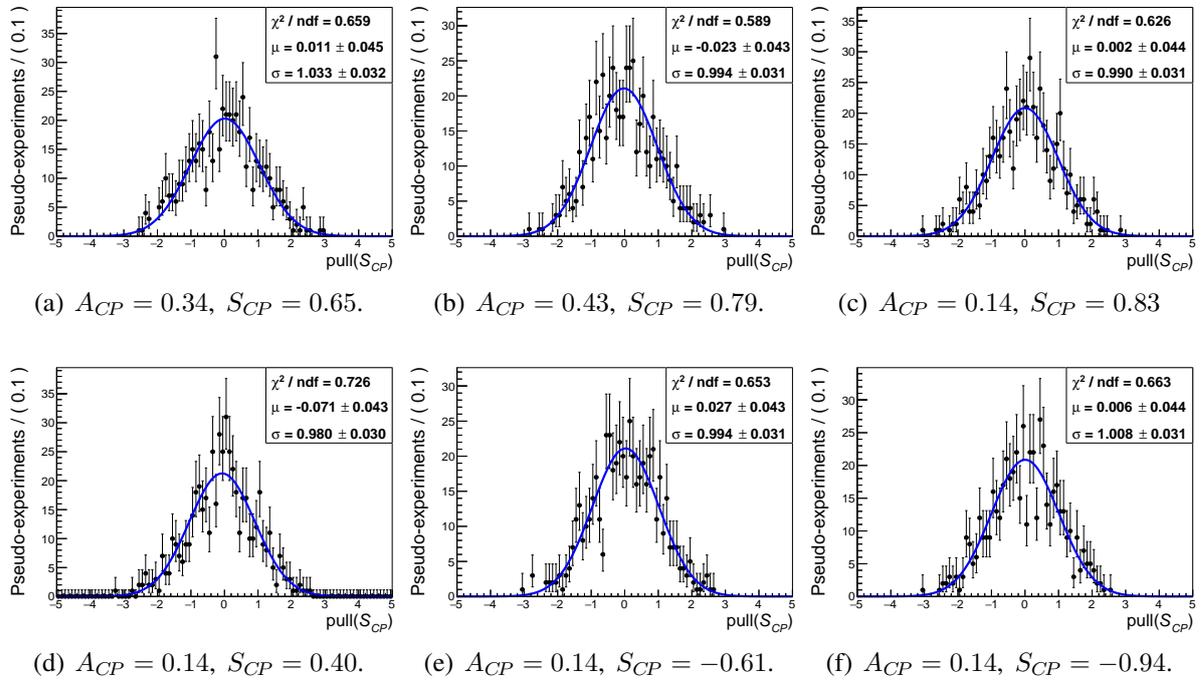


Figure D.9: Pull distributions for the CP -violation parameter $S^{\pi^0\pi^0}$. The input values for $\mathcal{A}^{\pi^0\pi^0}$ and $S^{\pi^0\pi^0}$ are shown below.

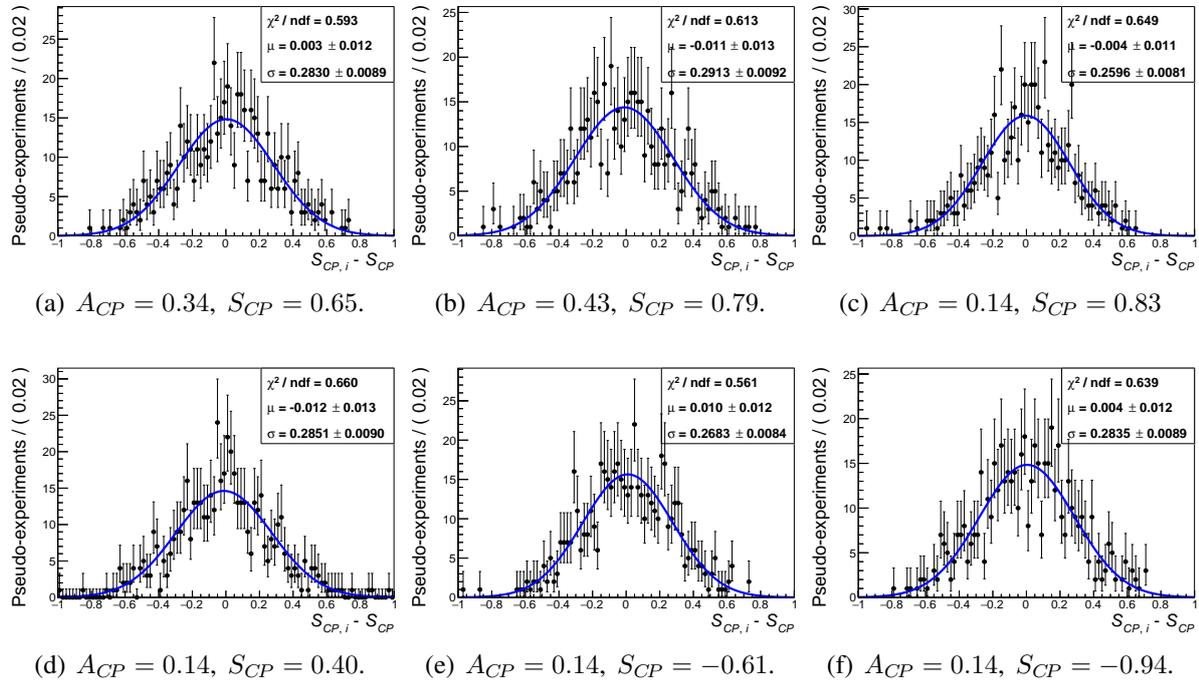


Figure D.10: Distributions of the residuals for the CP -violation parameter $S^{\pi^0\pi^0}$. The input values for $\mathcal{A}^{\pi^0\pi^0}$ and $S^{\pi^0\pi^0}$ are shown below.

Bibliography

- [1] E. Kou et al. “The Belle II Physics Book” (2018). arXiv: 1808.10567 [hep-ex].
- [2] H. Yukawa. “On the Interaction of Elementary Particles I”. *Proc. Phys. Math. Soc. Jap.* 17 (1935). [Prog. Theor. Phys. Suppl.1,1(1935)], pp. 48–57.
- [3] D. J. Gross and F. Wilczek. “Ultraviolet Behavior of Nonabelian Gauge Theories”. *Phys. Rev. Lett.* 30 (1973). [,271(1973)], pp. 1343–1346.
- [4] H. D. Politzer. “Reliable Perturbative Results for Strong Interactions?” *Phys. Rev. Lett.* 30 (1973). [,274(1973)], pp. 1346–1349.
- [5] S. L. Glashow. “Partial Symmetries of Weak Interactions”. *Nucl. Phys.* 22 (1961), pp. 579–588.
- [6] S. Weinberg. “A Model of Leptons”. *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266.
- [7] A. Salam. “Elementary particle theory”. *Ed. N. Svartholm, Stockholm. Almqvist and Wiksell* 367 (1968).
- [8] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. *Phys. Rev. Lett.* 13 (1964). [,157(1964)], pp. 321–323.
- [9] P. W. Higgs. “Broken symmetries, massless particles and gauge fields”. *Phys. Lett.* 12 (1964), pp. 132–133.
- [10] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. *Phys. Rev. Lett.* 13 (1964). [,160(1964)], pp. 508–509.
- [11] Wikipedia. *Standard Model — Wikipedia, The Free Encyclopedia*. [Online; accessed 2018-02-02] https://en.wikipedia.org/wiki/Standard_Model. 2018.
- [12] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. *Phys. Lett.* B716 (2012), pp. 1–29. arXiv: 1207.7214 [hep-ex].
- [13] S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. *Phys. Lett.* B716 (2012), pp. 30–61. arXiv: 1207.7235 [hep-ex].
- [14] E. Noether. “Invariante Variationsprobleme”. *Gott. Nachr.* 1918 (1918). [Transp. Theory Statist. Phys.1,186(1971)], pp. 235–257. arXiv: physics/0503066 [physics].

- [15] S. Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 2005. ISBN: 9780521670531, 9780511252044.
- [16] T. D. Lee and C.-N. Yang. “Question of Parity Conservation in Weak Interactions”. *Phys. Rev.* 104 (1956), pp. 254–258.
- [17] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. *Phys. Rev.* 105 (1957), pp. 1413–1414.
- [18] T. D. Lee, R. Oehme, and C.-N. Yang. “Remarks on Possible Noninvariance Under Time Reversal and Charge Conjugation”. *Phys. Rev.* 106 (1957), pp. 340–345.
- [19] A. Angelopoulos et al. “First direct observation of time reversal noninvariance in the neutral kaon system”. *Phys. Lett.* B444 (1998), pp. 43–51.
- [20] H. J. Gerber. “Evidence for time-reversal violation?” *Eur. Phys. J.* C35 (2004), pp. 195–196.
- [21] A. Alavi-Harati et al. “Observation of CP violation in $K_L^0 \rightarrow \pi^+\pi^-e^+e^-$ decays”. *Phys. Rev. Lett.* 84 (2000), pp. 408–411. arXiv: hep-ex/9908020 [hep-ex].
- [22] J. P. Lees et al. “Observation of Time Reversal Violation in the B^0 Meson System”. *Phys. Rev. Lett.* 109 (2012), p. 211801. arXiv: 1207.5832 [hep-ex].
- [23] L. Wolfenstein. “Violation of time reversal invariance in K_0 decays”. *Phys. Rev. Lett.* 83 (1999), pp. 911–912.
- [24] E. Applebaum et al. “Subtleties in the $BABAR$ measurement of time-reversal violation”. *Phys. Rev.* D89.7 (2014), p. 076011. arXiv: 1312.4164 [hep-ph].
- [25] H. Frauenfelder and E. M. Henley. *SUBATOMIC PHYSICS*. 1991. ISBN: 0-13-859430-9.
- [26] G. Luders. “On the Equivalence of Invariance under Time Reversal and under Particle-Antiparticle Conjugation for Relativistic Field Theories”. *Kong. Dan. Vid. Sel. Mat. Fys. Med.* 28N5.5 (1954), pp. 1–17.
- [27] W. Pauli. *Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of His Seventieth Birthday*. McGraw-Hill, 1955.
- [28] C. Patrignani et al. “Review of Particle Physics”. *Chin. Phys.* C40.10 (2016), p. 100001.
- [29] A. J. Bevan et al. “The Physics of the B Factories”. *Eur. Phys. J.* C74 (2014), p. 3026. arXiv: 1406.6311 [hep-ex].
- [30] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. *Phys. Rev. Lett.* 13 (1964), pp. 138–140.
- [31] A. Abashian et al. “Search for CP nonconservation in K -20 decays”. *Phys. Rev. Lett.* 13 (1964), pp. 243–246.
- [32] E. Fermi. “An attempt of a theory of beta radiation. 1.” *Z. Phys.* 88 (1934), pp. 161–177.

- [33] N. Cabibbo. “Unitary Symmetry and Leptonic Decays”. *Phys. Rev. Lett.* 10 (1963). [648(1963)], pp. 531–533.
- [34] A. Barbaro-Galtieri et al. “Review of particle properties”. *Rev. Mod. Phys.* 42 (1970), pp. 87–200.
- [35] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. *Phys. Rev. D* 2 (1970), pp. 1285–1292.
- [36] D. Boutigny et al. “The BABAR physics book: Physics at an asymmetric B factory”. *Workshop on Physics at an Asymmetric B Factory (BaBar Collaboration Meeting) Pasadena, California, September 22-24, 1997*. 1998. URL: <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=SLAC-R-504>.
- [37] L.-L. Chau and W.-Y. Keung. “Comments on the Parametrization of the Kobayashi-Maskawa Matrix”. *Phys. Rev. Lett.* 53 (1984), p. 1802.
- [38] L. Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. *Phys. Rev. Lett.* 51 (1983), p. 1945.
- [39] C. Jarlskog. “Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Violation”. *Phys. Rev. Lett.* 55 (1985), p. 1039.
- [40] W. Hulsbergen. “CONSTRAINING NEW PHYSICS IN MESON MIXING”. *Modern Physics Letters A* 28.27 (2013).
- [41] R. Lewis and C. Enz. “Phys. Letters 16, 73 (1965); RR Lewis”. *Phys. Rev. Letters* 14 (1965), p. 749.
- [42] Y. Amhis et al. “Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016”. *Eur. Phys. J. C* 77.12 (2017). updated results and plots available at <https://hflav.web.cern.ch>, p. 895. arXiv: 1612.07233 [hep-ex].
- [43] M. Gell-Mann and A. Pais. “Behavior of neutral particles under charge conjugation”. *Phys. Rev.* 97 (1955), pp. 1387–1389.
- [44] K. Lande et al. “Observation of Long-Lived Neutral V Particles”. *Phys. Rev.* 103 (1956), pp. 1901–1904.
- [45] H. Albrecht et al. “Observation of B^0 - anti- B^0 Mixing”. *Phys. Lett.* B192 (1987). [51(1987)], pp. 245–252.
- [46] M. V. Danilov. “B physics”. *Conf. Proc.* C930722 (1993), pp. 851–868.
- [47] A. Abulencia et al. “Observation of $B_s^0 - \bar{B}_s^0$ Oscillations”. *Phys. Rev. Lett.* 97 (2006), p. 242003. arXiv: hep-ex/0609040 [hep-ex].
- [48] A. Deur, S. J. Brodsky, and G. F. de Teramond. “The QCD Running Coupling”. *Prog. Part. Nucl. Phys.* 90 (2016), pp. 1–74. arXiv: 1604.08082 [hep-ph].

- [49] G. Buchalla. “Heavy quark theory”. *Heavy flavor physics: Theory and experimental results in heavy quark physics and CP violation. Proceedings, 55th Scottish Universities Summer School in Physics, SUSSP 2001, St. Andrews, UK, August 7-23, 2001*. 2002, pp. 57–104. arXiv: hep-ph/0202092 [hep-ph]. URL: <http://weblib.cern.ch/abstract?CERN-TH-2002-018>.
- [50] J. S. Hagelin and M. B. Wise. “Comment on CP Violation in $B^0 - \bar{B}^0$ Mixing”. *Nucl. Phys.* B189 (1981), pp. 87–92.
- [51] M. Beneke, G. Buchalla, and I. Dunietz. “Mixing induced CP asymmetries in inclusive B decays”. *Phys. Lett.* B393 (1997), pp. 132–142. arXiv: hep-ph/9609357 [hep-ph].
- [52] T. E. Browder and K. Honscheid. “B mesons”. *Progress in Particle and Nuclear Physics* 35 (1995), pp. 81–219. ISSN: 0146-6410.
- [53] A. A. Alves Jr. et al. “The LHCb Detector at the LHC”. *JINST* 3 (2008), S08005.
- [54] R. Aaij et al. “Measurement of the b -quark production cross-section in 7 and 13 TeV pp collisions”. *Phys. Rev. Lett.* 118.5 (2017), p. 052002. arXiv: 1612.05140 [hep-ex].
- [55] M. Pepe Altarelli and F. Teubert. “B Physics at LHCb”. *Int. J. Mod. Phys. A* 23 (2008). [299(2008)], pp. 5117–5136. arXiv: 0802.1901 [hep-ph].
- [56] T. Schörner-Sadenius, ed. *The Large Hadron Collider*. Berlin: Springer, 2015. ISBN: 9783319150000, 9783319150017. URL: <http://www.springer.com/gb/book/9783319150017>.
- [57] S. S. Malde. *Synergy of BESIII and LHCb physics programmes*. Tech. rep. LHCb-PUB-2016-025. CERN-LHCb-PUB-2016-025. Geneva: CERN, Oct. 2016. URL: <http://cds.cern.ch/record/2223391>.
- [58] J. Charles et al. “CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories”. *Eur. Phys. J.* C41.1 (2005). Updated results and plots available at: <http://ckmfitter.in2p3.fr>, pp. 1–131. arXiv: hep-ph/0406184 [hep-ph].
- [59] I. Adachi et al. “Precise measurement of the CP violation parameter $\sin(2\phi_1)$ in $B^0 \rightarrow (c\bar{c})K^0$ decays”. *Phys. Rev. Lett.* 108 (2012), p. 171802. arXiv: 1201.4643 [hep-ex].
- [60] B. Aubert et al. “Measurement of Time-Dependent CP Asymmetry in $B^0 \rightarrow c\bar{c}K^{(*)0}$ Decays”. *Phys. Rev.* D79 (2009), p. 072009. arXiv: 0902.1708 [hep-ex].
- [61] R. Aaij et al. “Measurement of CP violation in $B^0 \rightarrow J/\psi K_S^0$ decays”. *Phys. Rev. Lett.* 115.3 (2015), p. 031601. arXiv: 1503.07089 [hep-ex].
- [62] M. Jung. “Determining weak phases from $B \rightarrow J/\psi P$ decays”. *Phys. Rev.* D86 (2012), p. 053008. arXiv: 1206.2050 [hep-ph].

- [63] M. Gronau and D. London. “How to determine all the angles of the unitarity triangle from $B_d^0 \rightarrow DK_S^0$ and $B_s^0 \rightarrow D\phi$ ”. *Phys. Lett.* B253 (1991), pp. 483–488.
- [64] D. Atwood, I. Dunietz, and A. Soni. “Improved methods for observing CP violation in $B^\pm \rightarrow KD$ and measuring the CKM phase γ ”. *Phys. Rev.* D63 (2001), p. 036005. arXiv: hep-ph/0008090 [hep-ph].
- [65] A. Giri et al. “Determining γ using $B^\pm \rightarrow DK^\pm$ with multibody D decays”. *Phys. Rev.* D68 (2003), p. 054018. arXiv: hep-ph/0303187 [hep-ph].
- [66] R. Aaij et al. “Measurement of the CKM angle γ from a combination of LHCb results”. *JHEP* 12 (2016), p. 087. arXiv: 1611.03076 [hep-ex].
- [67] H.-Y. Cheng and S. Oh. “Flavor $SU(3)$ symmetry and QCD factorization in $B \rightarrow PP$ and PV decays”. *JHEP* 09 (2011), p. 024. arXiv: 1104.4144 [hep-ph].
- [68] S.-H. Zhou et al. “Analysis of Charmless Two-body B decays in Factorization Assisted Topological Amplitude Approach”. *Eur. Phys. J.* C77.2 (2017), p. 125. arXiv: 1608.02819 [hep-ph].
- [69] N. G. Deshpande and X.-G. He. “Isospin structure of penguins and their consequences in B physics”. *Phys. Rev. Lett.* 74 (1995). [Erratum: *Phys. Rev. Lett.* 74,4099(1995)], pp. 26–29. arXiv: hep-ph/9408404 [hep-ph].
- [70] M. Gronau et al. “Electroweak penguins and two-body B decays”. *Phys. Rev.* D52 (1995), pp. 6374–6382. arXiv: hep-ph/9504327 [hep-ph].
- [71] J. Charles et al. “Isospin analysis of charmless B-meson decays”. *Eur. Phys. J.* C77.8 (2017), p. 574. arXiv: 1705.02981 [hep-ph].
- [72] D. London, N. Sinha, and R. Sinha. “Can one measure the weak phase of a penguin diagram?” *Phys. Rev.* D60 (1999), p. 074020. arXiv: hep-ph/9905404 [hep-ph].
- [73] M. Gronau and D. London. “Isospin analysis of CP asymmetries in B decays”. *Phys. Rev. Lett.* 65 (1990), pp. 3381–3384.
- [74] H. J. Lipkin et al. “Penguin trapping with isospin analysis and CP asymmetries in B decays”. *Phys. Rev.* D44 (1991), pp. 1454–1460.
- [75] M. Gronau et al. “Decays of B mesons to two light pseudoscalars”. *Phys. Rev.* D50 (1994), pp. 4529–4543. arXiv: hep-ph/9404283 [hep-ph].
- [76] M. Gronau et al. “Broken $SU(3)$ symmetry in two-body B decays”. *Phys. Rev.* D52 (1995), pp. 6356–6373. arXiv: hep-ph/9504326 [hep-ph].
- [77] M. Beneke et al. “A Precise determination of α using $B^0 \rightarrow \rho^+\rho^-$ and $B^+ \rightarrow K^{*0}\rho^+$ ”. *Phys. Lett.* B638 (2006), pp. 68–73. arXiv: hep-ph/0604005 [hep-ph].
- [78] A. E. Snyder and H. R. Quinn. “Measuring CP asymmetry in $B \rightarrow \rho\pi$ decays without ambiguities”. *Phys. Rev.* D48 (1993), pp. 2139–2144.

- [79] M. Gronau and J. Zupan. “On measuring α in $B(t) \rightarrow \rho^\pm \pi^\pm$ ”. *Phys. Rev. D* 70 (2004), p. 074031. arXiv: hep-ph/0407002 [hep-ph].
- [80] M. Gronau and J. Zupan. “Weak phase α from $B \rightarrow a(1260)_1^{+-} \pi$ ”. *Phys. Rev. D* 73 (2006), p. 057502. arXiv: hep-ph/0512148 [hep-ph].
- [81] S. Laplace and V. Shelkov. “ CP violation and the absence of second class currents in charmless B decays”. *Eur. Phys. J. C* 22 (2001), pp. 431–438. arXiv: hep-ph/0105252 [hep-ph].
- [82] R. H. Dalitz. “On an alternative decay process for the neutral pi-meson, Letters to the Editor”. *Proc. Phys. Soc. A* 64 (1951), pp. 667–669.
- [83] A. F. Falk et al. “Comment on extracting α from $B \rightarrow \rho\rho$ ”. *Phys. Rev. D* 69 (2004), p. 011502. arXiv: hep-ph/0310242 [hep-ph].
- [84] M. Gronau and J. L. Rosner. “Controlling ρ width effects for a precise value of α in $B \rightarrow \rho\rho$ ”. *Phys. Lett. B* 766 (2017), pp. 345–350. arXiv: 1612.08524 [hep-ph].
- [85] M. Pivk and F. R. Le Diberder. “Isospin constraints from I on $B \rightarrow \pi\pi$ ”. *Eur. Phys. J. C* 39 (2005), pp. 397–409. arXiv: hep-ph/0406263 [hep-ph].
- [86] B. Aubert et al. “Measurement of the Branching Fraction, Polarization, and CP Asymmetries in $B^0 \rightarrow \rho^0 \rho^0$ Decay, and Implications for the CKM Angle α ”. *Phys. Rev. D* 78 (2008), p. 071104. arXiv: 0807.4977 [hep-ex].
- [87] P. Vanhoefer et al. “Study of $B^0 \rightarrow \rho^+ \rho^-$ decays and implications for the CKM angle ϕ_2 ”. *Phys. Rev. D* 93.3 (2016). [Addendum: *Phys. Rev. D* 94, no. 9, 099903 (2016)], p. 032010. arXiv: 1510.01245 [hep-ex].
- [88] J. Zhang et al. “Observation of $B^+ \rightarrow \rho^+ \rho^0$ ”. *Phys. Rev. Lett.* 91 (2003), p. 221801. arXiv: hep-ex/0306007 [hep-ex].
- [89] B. Aubert et al. “A Study of $B^0 \rightarrow \rho^+ \rho^-$ Decays and Constraints on the CKM Angle α ”. *Phys. Rev. D* 76 (2007), p. 052007. arXiv: 0705.2157 [hep-ex].
- [90] B. Aubert et al. “Improved Measurement of $B^+ \rightarrow \rho^+ \rho^0$ and Determination of the Quark-Mixing Phase Angle α ”. *Phys. Rev. Lett.* 102 (2009), p. 141802. arXiv: 0901.3522 [hep-ex].
- [91] A. J. Buras and R. Fleischer. “A General analysis of gamma determinations from $B \rightarrow \pi K$ decays”. *Eur. Phys. J. C* 11 (1999), pp. 93–109. arXiv: hep-ph/9810260 [hep-ph].
- [92] M. Gronau and J. Zupan. “Isospin-breaking effects on α extracted in $B \rightarrow \pi\pi, \rho\rho, \rho\pi$ ”. *Phys. Rev. D* 71 (2005), p. 074017. arXiv: hep-ph/0502139 [hep-ph].
- [93] E. Baracchini and G. Isidori. “Electromagnetic corrections to non-leptonic two-body B and D decays”. *Phys. Lett. B* 633 (2006), pp. 309–313. arXiv: hep-ph/0508071 [hep-ph].

- [94] F. J. Botella, D. London, and J. P. Silva. “Looking for Delta $I = 5/2$ amplitude components in $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ experiments”. *Phys. Rev. D* **73** (2006), p. 071501. arXiv: hep-ph/0602060 [hep-ph].
- [95] A. J. Buras and J. M. Gerard. “Isospin Breaking Contributions to Epsilon-prime / Epsilon”. *Phys. Lett. B* **192** (1987), pp. 156–162.
- [96] H. B. O’Connell et al. “ $\rho - \omega$ mixing, vector meson dominance and the pion form-factor”. *Prog. Part. Nucl. Phys.* **39** (1997), pp. 201–252. arXiv: hep-ph/9501251 [hep-ph].
- [97] M. Gronau and J. L. Rosner. “Improving the measurement of the CKM phase $\phi_2 = \alpha$ in $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ decays”. *Phys. Lett. B* **763** (2016), pp. 228–233. arXiv: 1608.06224 [hep-ph].
- [98] J. Zupan. “Determining alpha and gamma: Theory”. *3rd Conference on Flavor Physics and CP violation (FPCP 2004) Daegu, Korea, October 4-9, 2004*. 2004. arXiv: hep-ph/0410371 [hep-ph].
- [99] A. Kusaka et al. “Measurement of CP Asymmetry in a Time-Dependent Dalitz Analysis of $B^0 \rightarrow (\rho\pi)^0$ and a Constraint on the CKM Angle ϕ_2 ”. *Phys. Rev. Lett.* **98** (2007), p. 221602. arXiv: hep-ex/0701015 [hep-ex].
- [100] J. P. Lees et al. “Measurement of CP-violating asymmetries in $B^0 \rightarrow (\rho\pi)^0$ decays using a time-dependent Dalitz plot analysis”. *Phys. Rev. D* **88.1** (2013), p. 012003. arXiv: 1304.3503 [hep-ex].
- [101] H. R. Quinn and J. P. Silva. “The Use of early data on $B \rightarrow \rho\pi$ decays”. *Phys. Rev. D* **62** (2000), p. 054002. arXiv: hep-ph/0001290 [hep-ph].
- [102] A. Abashian et al. “The Belle Detector”. *Nucl. Instrum. Meth. A* **479** (2002), pp. 117–232.
- [103] J. Brodzicka et al. “Physics Achievements from the Belle Experiment”. *PTEP* **2012** (2012), p. 04D001. arXiv: 1212.5342 [hep-ex].
- [104] S. Kurokawa and E. Kikutani. “Overview of the KEKB accelerators”. *Nucl. Instrum. Meth. A* **499** (2003), pp. 1–7.
- [105] T. Abe et al. “Belle II Technical Design Report” (2010). arXiv: 1011.0352.
- [106] N. Braun et al. *Study of the collision point properties*. BELLE2-NOTE-PH-2018-008. May 2018.
- [107] *Belle II Detector 3D model*. <https://www.belle2.org/archives/>.
- [108] Z. Dolezal et al. “Belle II PXD Whitebook”. Version 0. 2012.
- [109] R. H. Richter et al. “Design and technology of DEPFET pixel sensors for linear collider applications”. *Proceedings, 11th International Workshop on Vertex Detectors (Vertex 2002): Kailua-Kona, USA, November 3-8, 2002*. Vol. A511. 2003, pp. 250–256.

- [110] F. Müller. “Characterization and Optimization of the Prototype DEPFET Modules for the Belle II Pixel Vertex Detector”. PhD thesis. Ludwig-Maximilians-Universität, 2017.
- [111] A. Moll et al. *The vertex detector numbering scheme*. BELLE2-NOTE-TE-2015-010. Aug. 2016.
- [112] S. Pohl. “Track Reconstruction at the First Level Trigger of the Belle II Experiment”. PhD thesis. Ludwig-Maximilians-Universität, 2017.
- [113] J. R. et al. “GENFIT - a Generic Track-Fitting Toolkit” (2014). arXiv: 1410.3698 [hep-ex].
- [114] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. *Nucl. Instrum. Meth. A*262 (1987), pp. 444–450.
- [115] T. Keck. “FastBDT: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification”. *CoRR* abs/1609.06119 (2016). URL: <http://arxiv.org/abs/1609.06119>.
- [116] A. Moll. “Comprehensive study of the background for the Pixel Vertex Detector at Belle II”. PhD thesis. Ludwig-Maximilians-Universität, 2015.
- [117] H. Ishino, A. Kibayashi, and K. Takahashi. “Alignment method for the SVD2 and its Performance”. *Belle internal note BN715* (2005).
- [118] J. Tanaka. *Kinematic fitting*.
- [119] W. Waltenberger. “RAVE: A detector-independent toolkit to reconstruct vertices”. *IEEE Trans. Nucl. Sci.* 58 (2011), pp. 434–444.
- [120] T. Kawasaki. “Simulation Study for Measurement of CP violation in $B \rightarrow \text{charmonium} + K_S^0$ at KEK B -factory.” PhD thesis. Osaka University, 1999.
- [121] M. J. Gelb. “Neutral B meson Flavor Tagging for Belle II”. MA thesis. Karlsruher Institut für Technologie, 2015.
- [122] J. F. Gemmler. “Study of B meson Flavor Tagging with Deep Neural Networks at Belle and at Belle II”. MA thesis. Karlsruher Institut für Technologie, 2016.
- [123] Y. LeCun, Y. Bengio, and G. Hinton. “Deep learning”. *nature* 521.7553 (2015), p. 436.
- [124] L. Breiman et al. *Classification and regression trees*. Chapman and Hall/CRC, 1984. ISBN: 0412048418, 9780412048418.
- [125] S. Nissen. “Implementation of a fast artificial neural network library (fann)”. *Report, Department of Computer Science University of Copenhagen (DIKU)* 31 (2003), p. 29.
- [126] F. Abudinén. “Studies on the neural z -Vertex-Trigger for the Belle II Particle Detector”. MA thesis. Ludwig-Maximilians-Universität, 2014.
- [127] T. Keck. “Machine Learning Algorithms for the Belle II experiment and their validation on Belle data”. PhD thesis. Karlsruher Institut für Technologie, 2017.

- [128] K. F. Chen et al. “Time-dependent CP -violating asymmetries in $b \rightarrow s\bar{q}q$ transitions”. *Phys. Rev. D* 72 (2005), p. 012004. arXiv: hep-ex/0504023 [hep-ex].
- [129] H. Sahoo et al. “Measurements of time-dependent CP violation in $B^0 \rightarrow \psi(2S)K_S^0$ decays”. *Phys. Rev. D* 77 (2008), p. 091103. arXiv: 0708.2604 [hep-ex].
- [130] H. Kakuno et al. “Neutral B flavor tagging for the measurement of mixing induced CP violation at Belle”. *Nucl. Instrum. Meth. A* 533 (2004), pp. 516–531. arXiv: hep-ex/0403022 [hep-ex].
- [131] A. Abashian et al. “Measurement of the CP violation parameter $\sin(2\phi_1)$ in B_d^0 meson decays”. *Phys. Rev. Lett.* 86 (2001), pp. 2509–2514. arXiv: hep-ex/0102018 [hep-ex].
- [132] K. Abe et al. “Observation of mixing induced CP violation in the neutral B meson system”. *Phys. Rev. D* 66 (2002), p. 032007. arXiv: hep-ex/0202027 [hep-ex].
- [133] M. Pivk and F. R. Le Diberder. “SPlot: A Statistical tool to unfold data distributions”. *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. arXiv: physics/0402083 [physics.data-an].
- [134] O. Long et al. “Impact of tag side interference on time dependent CP asymmetry measurements using coherent B^0 anti- B^0 pairs”. *Phys. Rev. D* 68 (2003), p. 034010. arXiv: hep-ex/0303030 [hep-ex].
- [135] I. J. Goodfellow et al. “Pylearn2: a machine learning research library”. *arXiv preprint arXiv:1308.4214* (2013).
- [136] R. Al-Rfou et al. “Theano: A Python framework for fast computation of mathematical expressions” (2016). arXiv: 1605.02688 [cs.SC].
- [137] A. Moll. “The Software Framework of the Belle II Experiment”. *Journal of Physics: Conference Series* 331.3 (2011), p. 032024.
- [138] H. Ishino et al. “New Measurements Using External Photon Conversion at a High Luminosity B Factory” (2007). arXiv: hep-ex/0703039 [HEP-EX].
- [139] D. J. Lange. “The EvtGen particle decay simulation package”. *Nucl. Instrum. Meth. A* 462 (2001), pp. 152–155.
- [140] T. Julius et al. “Measurement of the branching fraction and CP asymmetry in $B^0 \rightarrow \pi^0\pi^0$ decays, and an improved constraint on ϕ_2 ”. *Phys. Rev. D* 96.3 (2017), p. 032007. arXiv: 1705.02083 [hep-ex].
- [141] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303.
- [142] Y. Chao et al. “Observation of $B^0 \rightarrow \pi^0\pi^0$ ”. *Phys. Rev. Lett.* 94 (2005), p. 181803. arXiv: hep-ex/0408101 [hep-ex].

- [143] K. Abe et al. “Improved measurement of $B^0 \rightarrow \pi^0\pi^0$ ” (2006). arXiv: hep-ex/0610065 [hep-ex].
- [144] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. “Adaptive vertex fitting”. *J. Phys.* G34 (2007), N343.
- [145] D. M. Asner et al. “Search for exclusive charmless hadronic B decays”. *Phys. Rev.* D53 (1996), pp. 1039–1050. arXiv: hep-ex/9508004 [hep-ex].
- [146] G. C. Fox and S. Wolfram. “Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes”. *Phys. Rev. Lett.* 41 (1978), p. 1581.
- [147] S. H. Lee et al. “Evidence for $B^0 \rightarrow \pi^0\pi^0$ ”. *Phys. Rev. Lett.* 91 (2003), p. 261801. arXiv: hep-ex/0308040 [hep-ex].
- [148] T. B. Collaboration. *RooRarFit: a general Maximum Likelihood fitting package based on ROOT/RooFit*. [Online; accessed 2018-08-06] <http://rarfit.sourceforge.net/>. 2007.
- [149] W. Verkerke and D. P. Kirkby. “The RooFit toolkit for data modeling”. *eConf* C0303241 (2003). [186(2003)], MOLT007. arXiv: physics/0306116 [physics].
- [150] F. James and M. Roos. “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”. *Comput. Phys. Commun.* 10 (1975), pp. 343–367.
- [151] J. P. Lees et al. “Measurement of CP Asymmetries and Branching Fractions in Charmless Two-Body B -Meson Decays to Pions and Kaons”. *Phys. Rev.* D87.5 (2013), p. 052009. arXiv: 1206.3525 [hep-ex].
- [152] J. Charles et al. “Bayesian statistics at work: The Troublesome extraction of the CKM phase α ” (2006). arXiv: hep-ph/0607246 [hep-ph].
- [153] P. Golonka and Z. Was. “PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays”. *Eur. Phys. J.* C45 (2006), pp. 97–107. arXiv: hep-ph/0506026 [hep-ph].
- [154] T. B. Collaboration. *The number of $B\bar{B}$ events in the Belle hadronic skims*. <http://belle.kek.jp/secured/nbb/nbb.html>.
- [155] F. Abudinén. “Impact of the Belle II pixel detector on the analysis of CP-violation”. *JINST* 12.03 (2017), p. C03014.
- [156] R. Aaij et al. “Measurement of b -hadron branching fractions for two-body decays into charmless charged hadrons”. *JHEP* 10 (2012), p. 037. arXiv: 1206.2794 [hep-ex].
- [157] Y. -.-T. Duh et al. “Measurements of branching fractions and direct CP asymmetries for $B \rightarrow K\pi$, $B \rightarrow \pi\pi$ and $B \rightarrow KK$ decays”. *Phys. Rev.* D87.3 (2013), p. 031103. arXiv: 1210.1348 [hep-ex].

- [158] B. Aubert et al. “Improved Measurements of the Branching Fractions for $B^0 \rightarrow \pi^+\pi^-$ and $B^0 \rightarrow K^+\pi^-$, and a Search for $B^0 \rightarrow K^+K^-$ ”. *Phys. Rev. D* 75 (2007), p. 012008. arXiv: hep-ex/0608003 [hep-ex].
- [159] B. Aubert et al. “Study of $B^0 \rightarrow \pi^0\pi^0$, $B^\pm \rightarrow \pi^\pm\pi^0$, and $B^\pm \rightarrow K^\pm\pi^0$ Decays, and Isospin Analysis of $B \rightarrow \pi\pi$ Decays”. *Phys. Rev. D* 76 (2007), p. 091102. arXiv: 0707.2798 [hep-ex].
- [160] I. Adachi et al. “Measurement of the CP violation parameters in $B^0 \rightarrow \pi^+\pi^-$ decays”. *Phys. Rev. D* 88.9 (2013), p. 092003. arXiv: 1302.0551 [hep-ex].
- [161] T. L. Collaboration. “Measurement of time-dependent CP -violating asymmetries in $B^0 \rightarrow \pi^+\pi^-$ and $B_s^0 \rightarrow K^+K^-$ decays at LHCb” (2017).
- [162] K. A. Prothmann. “Measurement of Branching Ratios and CP Asymmetries for the decays $B^0 \rightarrow \pi^+\pi^-$, $B^0 \rightarrow K^+\pi^-$ and $B^0 \rightarrow K^+K^-$ ”. PhD thesis. Max-Planck Institut für Physik München, 2012.
- [163] I. Adachi et al. “Study of $B^0 \rightarrow \rho^0\rho^0$ decays, implications for the CKM angle ϕ_2 and search for other B^0 decay modes with a four-pion final state” (2012). [Addendum: *Phys. Rev. D* 89, no. 11, 119903 (2014)]. arXiv: 1212.4015 [hep-ex].
- [164] R. Aaij et al. “Observation of the $B^0 \rightarrow \rho^0\rho^0$ decay from an amplitude analysis of $B^0 \rightarrow (\pi^+\pi^-)(\pi^+\pi^-)$ decays”. *Phys. Lett. B* 747 (2015), pp. 468–478. arXiv: 1503.07770 [hep-ex].
- [165] I. N. Bronstein et al. *Taschenbuch der mathematik*. Vol. 1. Springer-Verlag, 2012.
- [166] V. Devanathan. *Angular momentum techniques in quantum mechanics*. Vol. 108. Springer Science & Business Media, 1999.
- [167] A. Hoecker et al. “TMVA-Toolkit for multivariate data analysis” (2007).
- [168] H. Albrecht et al. “Exclusive hadronic decays of B mesons”. *Z. Phys. C* 48.4 (1990), pp. 543–551. ISSN: 0170–9739.
- [169] K. S. Cranmer. “Kernel estimation in high-energy physics”. *Comput. Phys. Commun.* 136 (2001), pp. 198–207. arXiv: hep-ex/0011057 [hep-ex].

Acknowledgments

I feel honored to recognize all the support I have received during my Ph.D. studies. I had the privilege to learn a lot from and to work with so many interesting and talented people within an international environment. Now I have the opportunity to express my gratitude to all of them.

First, I want to thank my Ph.D. adviser, Professor Christian Kiesling, for his support and his supervision, and for always finding time for me, even during his most hectic moments. Christian, thank you for teaching me so much about physics, for your many lectures, and especially for sharing your optimism and your passion for science. Thanks also to my second adviser, Professor Thomas Kuhr, for his support and his advises. Especially, I am indebted to him for finding funds to send me to the CHEP conference in Sofia. My next thanks goes to Hans-Günther Moser, the head of our group, for his many recommendations and for his support. Thanks also to Professor Jochen Schieck, my master thesis adviser, for having encouraged me to join the Belle II project.

Now I want to express my deepest gratitude to my day-to-day supervisor Luigi Li Gioi. Luigi, I have no words to thank you for all I have learned from you. Thanks for sharing with me all your experience about data analysis. Thanks for all the physics you explained to me. Thanks for always encouraging me and helping me find my path without hiding to me the many difficulties that were about to come. Thank you also for all the discussions about economy, politics, European and Italian history, Italian and French language. And last but not least, thank you for your many driving tips! I will never forget when we were driving out of New York during rush hour on the way to the B2TIP meeting in Pittsburgh, with the French speaking Cadillac! I also want to thank your wife Daniela for being so kind to me. She brought always a bit of Sicilian sun from Palermo into the office when she came to visit, and she made me feel as if I would be for a second in my sunny hometown Barranquilla.

During my Ph.D., and during my master studies, I received help from a senior member of our group, Vladimir Chekelian. Vladimir, my work always interested you; and you have followed and reviewed almost all of it. You gave to me so many advises that were crucial to overcome the many obstacles, ultimately helping me accomplish my goals; and I have learned also a lot about data analysis from you. Simply, thank you!

I want to thank my Ph.D. colleagues, Felix Benjamin Müller, Sara Pohl, Marco Szalay, Benedikt Wach, Philipp Leitl, Sebastian Skambraks, Miroslav Gabriel, Hendrik Windel, Lorenz Emberger and Yasmine Israeli, for the nice time I had during my studies. Thank you Felix for giving me the nice lab insight behind the scenes of the PXD during my first six months of Ph.D.

studies, for all the great time during them, and especially for sharing your experience about handing in. Thank you Sara for all your help with *TikZ*, for accompanying me to the opera, for the nice physics discussions, and especially for reviewing my derivation of the isospin triangles. Marco! Thank you for all the discussions about the Higgs, about calorimetry, and especially, you know why, about boosted decision trees. I always admire your vast knowledgeability and especially your fantastic ability of making complicated things understandable. I also want to thank my office mate Benedikt who, although he started just one week before I handed in, gave me valuable motivation during those stressful days.

I am grateful to a former postdoc of our group, Manfred Valentan, who reviewed some of my early articles. To him, and to his wife Conny-san, I want to express my gratitude for all the fun during our Japanese lessons, for the fun I had during the parties at their homes (in Munich and in Vienna), and for all the fun that is about to come when I will visit them in future.

If they ever get to read these lines—I want to thank Martin Jung, for always finding time to solve my theoretical questions about flavor physics. Thanks to Martin Ritter for motivating me to apply to the International Max Planck Research School after I handed in my master thesis. Thanks to Thomas Keck, Alessandro Gaz and Stefano Lacaprra for the nice collaboration, especially to Alessandro and to Stefano for helping me with the technicalities of `ROOT`. Sincere thanks to Phillip Urquijo, the Belle II physics coordinator, for all the nice comments and advises he gave to me and especially for supporting my application for my first post-doc position at INFN in Trieste. A big thanks goes also to Daniel Greenwald for his nice lectures about how to write science in plain English.

Finally, yet very important, I want to thank Diego Tonelli for all his advises and for all his support. Diego, I am very happy and looking forward to work with your group in Trieste from November on.

Now I want to thank all those persons that have been the cornerstones of my life. First, I want to thank my parents Sara Inés Gallego and Fernando Julio Abudinén. Mami, gracias por todos tus esfuerzos, por todo tu apoyo, y en especial por transmitirme tu pasión por el trabajo, siempre te llevo en mi corazón. Papá, yo sé que allá en el cielo estás muy orgulloso de mí, gracias por toda tu dedicación y por transmitirme todo tu carisma. To my second father Jean André Robert, my mother's second husband, I want to express my undying gratitude. Jean, merci pour tout ton aide et spécialement pour m'avoir enseigné la langue et la cuisine française.

I want to thank my three half-sisters Lisbeth, Rosemary and Maria Fernanda Abudinén and their mother María Inés Velásquez, my father's first wife, for all the love they gave to me as I was a child and for helping me doing my homework during the elementary school. I also want to thank my parents' friend Clara Martínez who recommended them to send me to the German school in Barranquilla.

I want to thank all my schoolteachers for their guidance, especially Jaime Álvarez Llanos, my politics and history teacher, Daniel Casseres, my math teacher, Jorge Fajardo, my physics

teacher, who helped me taking the crucial decision to study physics, and my biology teacher Rolf Ixmeier, who motivated me to pursue a scientific career and supported my successful application for a DAAD scholarship. Deep thanks goes to my German teacher Carmen Cepeda, who strongly supported me during the last period of high school and transmitted to me her love for language.

I want to express my deepest gratitude to my host parents Georgia and Werner Schneider, who treated me as one of their children during my school exchange in Eisenach. Thanks to the wonderful time I had with them, I was encouraged to come to Germany for my university studies.

Since I arrived in Munich, I have received orientation and support from two special friends, Thurid Piehler-Braun and Beate Cochlovius. I want to thank them for this and for all the fun we had together, especially for the fun we had during hiking and swimming at Starnberg and at Tegernsee. Special thanks goes to Sister Rossana Sollai for all her spiritual guidance. Last but not least, I want to thank my godfather Peter Müller for his huge support during my whole university studies, especially during the crucial three months before handing in.

