



UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

DIPARTIMENTO DI FISICA

XXXVI Ciclo del Dottorato di Ricerca in Fisica

MEASUREMENT OF THE PROPERTIES OF $B^0 \rightarrow \pi^0 \pi^0$
DECAYS AT BELLE II

Settore scientifico-disciplinare: FIS/04

Candidato:
Sebastiano Raiz

Coordinatore:
Prof. Francesco Longo

Supervisore di tesi:
Dott. Diego Tonelli

ANNO ACCADEMICO 2022–2023

Abstract

This experimental particle-physics thesis reports measurements of the branching fraction and charge-parity (CP) violating decay-rate asymmetry of $B^0 \rightarrow \pi^0\pi^0$ decays reconstructed in electron-positron collisions with the Belle II detector. These quantities constrain parameters of potential processes not described by the Standard Model of particle physics; offer data-driven information to refine phenomenological models of hadronic bottom-meson amplitudes; and provide important inputs for the determination of α , a fundamental parameter of quark dynamics.

The analysis uses the full data set of electron-positron collisions at the $\Upsilon(4S)$ resonance produced as of June 2022 by the energy-asymmetric SuperKEKB collider and collected by the Belle II experiment as of this writing. Belle II is an hermetic solenoidal magnetic spectrometer surrounded by particle-identification detectors, calorimeter, and muon detectors, designed to reconstruct the decay products of 10 GeV electron-positron collisions. The data set contains 387 million pairs of bottom-antibottom mesons.

The analysis is developed using simulated and control-data samples and applied to the signal sample in data only when all procedures are established. Data are enriched in $B^0 \rightarrow \pi^0\pi^0$ events by minimizing the expected statistical variance of the quantities of interest as a function of selections on two multivariate statistical-learning selections that suppress the most prominent sources of background. The flavor of the neutral B meson, needed to measure the asymmetry, is determined by using information associated with the other, nonsignal B -meson produced in the $\Upsilon(4S)$ decay. A multidimensional fit of sample composition statistically identifies the signal and determines its CP -violating asymmetry.

The fit results, combined with acceptance and efficiency corrections determined from simulation and validated in control data, allow for determining the $B^0 \rightarrow \pi^0\pi^0$ branching fraction,

$$\mathcal{B}(B^0 \rightarrow \pi^0\pi^0) = (1.26_{-0.19}^{+0.20} \text{ (stat)} \pm 0.12 \text{ (syst)}) \times 10^{-6},$$

and the CP -violating decay-rate-asymmetry,

$$A_{CP}(B^0 \rightarrow \pi^0\pi^0) = 0.06 \pm 0.30 \text{ (stat)} \pm 0.05 \text{ (syst)}.$$

The results are compatible with previous results. Even though based on a smaller sample, they have precision superior to, or comparable with, the world-best results, and are expected to improve significantly several Standard-Model parameters and constraints associated with charmless B decays.

Contents

Introduction	1
1 Flavor physics to overcome the Standard Model	3
1.1 The Standard Model of particle physics	3
1.2 Where do we stand?	5
1.3 Flavor physics in the Standard Model	6
1.4 Time evolution of neutral flavored mesons	8
1.5 Flavor physics to overcome the Standard Model	12
1.6 Current flavor status	12
1.7 The $B^0 \rightarrow \pi^0\pi^0$ decay	13
1.7.1 $B^0 \rightarrow \pi^0\pi^0$ predictions	14
1.7.2 The angle α	15
1.7.2.1 Isospin analysis	18
1.7.2.2 The S_{CP}^{00} constraint	21
1.8 $B^0 \rightarrow \pi^0\pi^0$ experimental status and current constraints	22
1.8.1 Current constraints on $B \rightarrow \pi\pi$ isospin	23
1.8.2 Constraints on α	23
2 The Belle II experiment at the SuperKEKB collider	25
2.1 The SuperKEKB collider	25
2.2 The Belle II detector	29
2.2.1 Tracking system	32
2.2.1.1 Silicon-pixel vertexing detector	32
2.2.1.2 Silicon-strip vertexing detector	32
2.2.1.3 Central drift chamber	33
2.2.2 Electromagnetic calorimeter	34
2.2.3 Particle identification	37
2.2.3.1 Time-of-propagation detector	37
2.2.3.2 Aerogel ring-imaging Cherenkov counter	37
2.2.3.3 K_L^0 and muon detection system	38
2.2.4 Trigger and data acquisition system	38
2.3 Reconstruction of stable particles	39
2.3.1 Charged-particle reconstruction	40
2.3.2 Charged-particle identification	42
2.3.3 Photon reconstruction	46
2.3.3.1 Photon matching	48
2.3.3.2 Shower topologies	48
2.3.3.3 Neutral particle reconstruction performance	51

3	Experimental considerations	53
3.1	Generalities on $B^0 \rightarrow \pi^0\pi^0$ decays	53
3.2	B decay reconstruction at Belle II	54
3.2.1	Particle-level variables	55
3.2.2	Candidate-level variables	55
3.2.3	B_{tag}^0 -level variables	57
3.2.4	Event-level variables	58
3.3	Analysis overview	63
4	Signal reconstruction and selection	65
4.1	Data sample	65
4.1.1	Experimental data	65
4.1.2	Simulated data	65
4.1.3	Basic data structures	67
4.1.4	Skim selection	67
4.2	Signal candidate reconstruction	67
4.3	Photon selection	68
4.3.1	Photon backgrounds	68
4.3.1.1	Misreconstructed photons	68
4.3.1.2	Beam-induced background photons	68
4.3.2	Baseline photon selection	69
4.3.3	Optimized photon selection	70
4.3.3.1	Choice of discriminating observables	71
4.3.3.2	Classifier training and test	76
4.3.3.3	Optimization	76
4.3.3.4	Photon classifier performance validation	77
4.3.3.5	Photon classifier validation and checks for biases	79
4.4	Neutral pion selection and reconstruction	82
4.5	B^0 selection and reconstruction	84
4.5.1	B^0 baseline selection	86
4.5.2	Continuum suppression	86
4.5.2.1	Optimization	89
4.5.2.2	Dedicated ρ suppression	97
4.6	Candidate multiplicity and signal efficiency	98
4.7	Residual background	99
4.7.1	Self-cross-feed background	99
4.7.2	$B\bar{B}$ backgrounds	99
4.8	Reconstruction of the tag B meson	100
5	Determination of the sample composition	101
5.1	The fit	101
5.1.1	Minimizer	101
5.1.2	Fit components	101
5.1.3	Fit observables	102
5.1.4	Fit parameters	104
5.1.5	Model	105
5.2	Likelihood modeling	105
5.2.1	Identification of dependences between fit observables	106
5.2.2	Signal p.d.f.	110

5.2.3	$B\bar{B}$ background p.d.f.	111
5.2.4	Continuum background p.d.f.	112
5.2.5	P.d.f. summary	115
5.3	Model refinements	116
5.3.1	Wrong-tag calibration	116
5.3.2	Flavor asymmetry in background	118
5.3.2.1	$B\bar{B}$ background	118
5.3.2.2	Continuum background	118
5.4	Estimator properties	119
5.5	Fit to realistically simulated samples	121
5.6	Fit strategy choices	123
5.6.1	Previous analysis configuration	123
5.6.2	Continuum amplifier	126
6	Analysis validation on data	127
6.1	Introduction	127
6.2	Fit validation using $B^+ \rightarrow K^+\pi^0$ decays	127
6.2.1	$B^+ \rightarrow K^+\pi^0$ selection	127
6.2.2	$B^+ \rightarrow K^+\pi^0$ likelihood modeling	128
6.2.3	$B^+ \rightarrow K^+\pi^0$ results and model correction extraction	132
6.3	Fit validation using $B^0 \rightarrow \bar{D}^0\pi^0$ decays	134
6.3.1	$B^0 \rightarrow \bar{D}^0\pi^0$ selection	134
6.3.2	$B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ likelihood modeling	134
6.3.3	$B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ results and model correction extraction	139
6.4	Closure test on partial sample	142
7	Fit of $B^0 \rightarrow \pi^0\pi^0$ data	145
8	Systematic uncertainties	151
8.1	Introduction	151
8.2	Systematic uncertainties included in the statistical uncertainty	151
8.3	Uncertainty in the neutral-pion reconstruction efficiency	152
8.4	Uncertainty on the continuum-background suppression efficiency	153
8.5	Uncertainty on the $B\bar{B}$ pair yield	154
8.6	Assumptions and uncertainties in fit modeling	154
8.7	Assumption on w calibration	155
8.8	Uncertainty in the $\Upsilon(4S)$ branching fractions	156
8.9	Systematic uncertainty summary	156
9	Results	157
	Summary	159
A	Photon energy bias correction	162
B	Methods for the determination of the π^0 reconstruction efficiency	164
B.1	Relative efficiency from D meson decays	164
B.2	Relative efficiency from τ lepton decays	165
B.3	Relative efficiency from η meson decays	165
B.4	Absolute efficiency from the $e^+e^- \rightarrow \pi^+\pi^-\pi^0\gamma_{\text{ISR}}$ process	165

CONTENTS

B.5 Possible dependence on multiplicity	165
Bibliography	167

Introduction

The Standard Model of elementary particles and their interactions (SM) is the currently accepted theory of particle physics. It is widely recognized as the ultimate success of the reductionist paradigm for describing microphysics at its most fundamental level. By means of about twenty parameters, the Standard Model describes accurately thousands of measurements involving processes mediated by the electromagnetic, weak, and strong interactions that span more than ten orders of magnitude in energy.

However, theoretical considerations and, possibly, experimental inconsistencies support the general belief that the Standard Model might still be an *effective theory* — a theory valid at the energies probed so far, that is incorporated in a yet-unknown and more general theory spanning higher energies. Completing the Standard Model is the principal goal of today's particle physics.

Direct approaches, which broadly consist in searching for decay products of non-SM particles produced on mass-shell in high-energy collisions, have been traditionally fruitful. However, their current reach is limited by the collision energy of today's accelerators and by the large investments needed to further it in future. Complementary approaches consist in comparing with predictions precise measurements in lower-energy processes where virtual non-SM particles could contribute. The reach of such *indirect* approaches is not constrained by collision energy, but rather by the precision attainable, both in measurements and predictions.

The Belle II experiment is an international collaboration of about 700 physicists that *indirectly* tests the Standard Model by studying millions of decays of mesons containing the quarks b and c (heavier and longer-lived partners of the fundamental constituents of nuclear matter) and τ leptons (heaviest partners of the electron). These particles are pair-produced in electron-positron collisions at 10.58 GeV, which is the threshold energy for the production of the $\Upsilon(4S)$ particle (a bound states of two b quarks). Since the start of physics data taking in 2019, Belle II collected samples corresponding to 424 fb^{-1} of integrated luminosity. The 362.4 fb^{-1} portion collected at the $\Upsilon(4S)$ resonance contains 387 million pairs of quantum-correlated bottom-antibottom mesons (B mesons, bound states of a b quark and a lighter quark), which are the data used to reconstruct our signal.

This work focuses on the experimental determination of the branching fraction and the charge-parity-violating decay-rate asymmetry of charmless $B^0 \rightarrow \pi^0 \pi^0$ decays reconstructed in the full Belle II data set available as of this writing. These parameters play an important role in the current understanding of the weak interactions of quarks. They are sensitive to the existence of non-SM particles that could contribute to the $B^0 \rightarrow \pi^0 \pi^0$ amplitude; they offer data-driven constraints to test and improve existing phenomenological models of the SM amplitudes of B decays into hadronic final states; and they enter in the determination of the parameter α , a fundamental parameter of quark dynamics related to the couplings of up-type ($+2e/3$ charge) and down-type ($-e/3$ charge) quarks and whose precision is one of the limiting factors in current SM consistency-tests.

The main experimental challenge is the reconstruction of a rare signal, about one decay per million B produced, in a final state with no charged particles, $B^0 \rightarrow \pi^0(\rightarrow \gamma\gamma)\pi^0(\rightarrow \gamma\gamma)$. The reconstruction of photons, typically based on calorimetric information only, is much less precise than the reconstruction of charged-particles, which relies on fine and precise samplings of their trajectories in spectrometers. And photon-energy reconstruction is affected by energy leakages and beam-related backgrounds. In addition, lack of signal decay-vertex information deprives the analysis of distinctive lifetime-related information usually employed to suppress background, which in this analysis is initially 10^7 larger than the signal at production. These limitations lead to a small signal-to-background ratio. A further challenge is the determination of whether the signal meson is a B^0 or a \bar{B}^0 , which is needed for the measurement of the CP -violating asymmetry because the $\pi^0\pi^0$ final state is common to both. Distinguishing the signal flavor implies relying on partial and often ambiguous information from the rest of the event.

Interpretation of $B^0 \rightarrow \pi^0\pi^0$ results poses phenomenological challenges as well. These are nonleptonic B decays in which the weak transitions we are interested in occur in a cloud of low-energy gluon exchanges. Hence, unambiguous interpretation of measurements from these decays is spoiled by hadronic uncertainties, which are hardly tractable in perturbative quantum-chromodynamics calculations. Rarely a single measurement in a single channel enables major phenomenological advances. Most frequently, appropriate combinations of measurements from decays related by flavor symmetries, such as isospin, are needed reduce the impact of such unknowns, thus leading to progress.

An early Belle II measurement of $B^0 \rightarrow \pi^0\pi^0$ decay properties using half of the sample used here has been recently published. In this work I improved and refined many aspects of the previous analysis: I improved the suppression of backgrounds by increasing the discriminating power of dedicated classifiers; I made the sample composition fit less dependent on simulation by lifting constraints, and simplified it by including as observable the predicted flavor of the B meson, obtained by new algorithms with higher efficacy; I included in the fit background-enriched events to constrain the background models directly from data, largely reducing systematic uncertainties. The results of this work are under internal collaboration review, but the observed precision is superior to, or comparable with, with that of current world-best results based on much larger samples.

This thesis is structured as follows. Chapter 1 introduces the flavor sector of the Standard Model; Chapter 2 describes the Belle II experiment; Chapter 3 outlines the relevant experimental features of the measurement and provides an outline of the general principles of B meson reconstruction; Chapter 4, where the description of my direct, original contributions begins, reports on the reconstruction and selection of the signal; Chapter 5 describes the fit of sample composition, the core of this work; Chapter 6 documents the analysis validation and consistency checks on control data; Chapter 7 reports the results of the fit of $B^0 \rightarrow \pi^0\pi^0$ sample composition; Chapter 8 discusses sources of systematic effects and associated uncertainties; Chapter 9 summarizes the final results and implications; a final summary concludes the document. Two appendices offer an overview of work I did for this analysis that proved essential for the whole collaboration. This consists in calibrating the photon energy measurement in the calorimeter and determining π^0 reconstruction efficiencies.

Charge- and flavor-conjugate processes are implied throughout the document unless specified otherwise. Generic particle symbols (B , π , ...) indicate indistinctly charged or neutral particles.

Chapter 1

Flavor physics to overcome the Standard Model

This is a concise introduction to the weak interactions of quarks and how they are incorporated in the Standard Model of particle physics. Emphasis is on their role in searches for as-yet unknown particles that may complete the Standard Model at high energies. The discussion focuses on charmless decays of bottom mesons, targeting the decay $B^0 \rightarrow \pi^0 \pi^0$ and on how measurements of its branching fraction and decay-rate asymmetry provide constraints on the Standard Model.

1.1 The Standard Model of particle physics

The Standard Model (SM) is an effective quantum field theory that describes all fundamental interactions in nature but gravity [1–6].

The quantum-field-theory framework results from the unification of quantum mechanics with special relativity and offers the most fundamental description of nature known to date.

A field is a set of values, associated to certain physical properties, assigned to every point in space and time. Quantum fields are fields that pervade the whole spacetime and obey the rules of quantum mechanics. If a quantum field is modified by an appropriate perturbation, the resulting oscillatory states, called field excitations, carry more energy than the resting state and are called ‘particles’. For instance, the electron is the massive excitation of the electron field. The quantized nature of the description implies that only certain perturbations that satisfy precise energetic conditions are capable of generating field excitations. It is not possible, for example, to generate a wave in the electron field that corresponds to half an electron.

Quantum fields interact with each other. The Standard Model is the theory that describes their dynamics at energy scales relevant for the subnuclear world. Particles and their interactions are described in a Lagrangian formalism, in which every combination of fields and interaction operators that is not forbidden by the symmetries of the dynamics is, in principle, included. Local gauge symmetry, i.e., the invariance of the Lagrangian under space-time-dependent transformations applied to the phases of fields, is the key overarching concept. Interaction terms appear in the free-field Lagrangian after requiring it to be invariant under local gauge symmetries. The Standard Model is based on the symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y,$$

where $SU(3)_C$ is the standard unitary group that describes the strong interactions (quantum chromodynamics, QCD), and C stands for the color charge; $SU(2)_L \otimes U(1)_Y$ is the

product of groups that describe the combination of the weak and electromagnetic interactions, where $SU(2)_L$ is the standard unitary group of weak isospin doublets (L standing for *left*¹) and $U(1)_Y$ stands for the unitary group of *hypercharge* Y .

Spin-1 particles called *gauge bosons* mediate the interactions. Strong interactions are mediated by eight massless particles corresponding to the $SU(3)_C$ generators, called *gluons*: they carry a combination of two charges that can be each of three kinds, called *color* and *anti-color*. Weak interactions are mediated by two charged massive bosons, W^\pm , and a neutral massive boson, Z^0 . Electromagnetic interactions occur between particles carrying electric charge and are mediated by a neutral massless boson, the photon γ . The physical electroweak bosons (W^\pm , Z^0 , γ) arise from the following linear combinations of $SU(2)_L \otimes U(1)_Y$ generators:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \quad \text{and} \quad \begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix},$$

where θ_W is a free parameter, called *Weinberg angle*. The W^\pm mass depends on the Z mass via θ_W . Particles acquire mass via the interaction with the Higgs field, which is mediated by a spin-0 particle, the Higgs boson.

Matter particles correspond to excitations of spin- $\frac{1}{2}$ fields and are called *fermions*. Their masses are free parameters of the theory. Each fermion is also associated with an antiparticle that has the same mass and opposite internal quantum numbers. Fermions are further classified into two classes, quarks, which are the fundamental constituents of nuclear matter, and leptons, each organized in three weak-isospin doublets.

- Quark doublets are composed of an up-type quark, with charge $\frac{2}{3}e$, and a down-type quark, with charge $-\frac{1}{3}e$ each,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}.$$

They couple with both the strong and electroweak interactions. Each quark has color and a ‘flavor’ quantum number, which comes in six varieties and is conserved in the electromagnetic and strong interactions, but not in the weak interactions. Due to color confinement free quarks are not observable [7]. They are only observed in their colorless bound states, which include mesons, typically composed of a quark and an anti-quark, and baryons, composed of three quarks. Baryons are assigned a quantum number, called baryon number, found to be conserved even if no symmetry of the Lagrangian implies that.

- Lepton doublets are composed each by an almost massless neutral neutrino and a massive particle with electric charge $-e$;

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}.$$

They couple only with the electroweak interaction. Each lepton has a lepton-family quantum number; their sum in a process, called global lepton number, is found to be conserved in all interactions, although no symmetry of the dynamics prescribes that; individual lepton numbers are not conserved in neutrino oscillations.

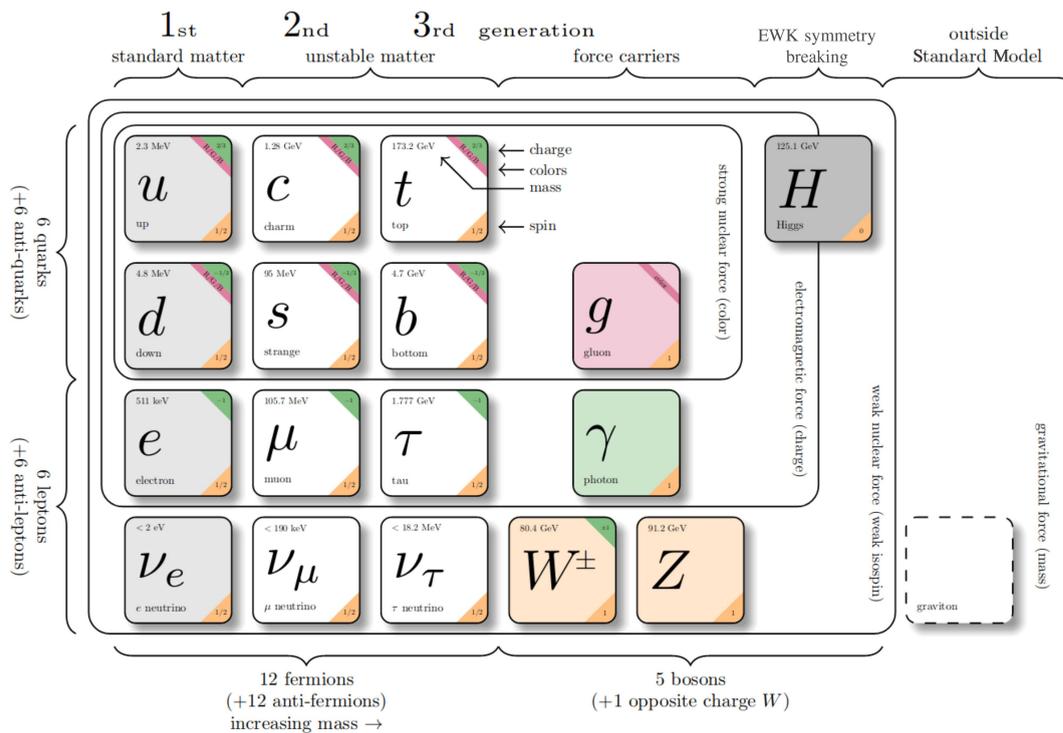


Figure 1.1: Scheme of particles and interactions in the Standard Model.

Figure 1.1 shows a scheme of the Standard Model particles and their interactions.

In addition to gauge symmetry, discrete symmetries are important in constraining the dynamics. Parity (P) is a transformation that inverts all spatial coordinates; charge conjugation (C) is the exchange of every particle with its own antiparticle; and time reversal (T) inverts the time axis. The product of these three discrete symmetry transformations is found to be conserved in all interactions, as prescribed by foundational axioms of field theory [8,9], but the symmetries are not conserved individually. Parity symmetry is maximally violated in the weak interactions, while the combined CP symmetry is violated in the weak interactions at the 0.1% level. In principle, the strong interaction too could violate CP symmetry, but no experimental evidence of that has ever been observed. The existence of as-yet unobserved particles, called axions, has been postulated to account for that [10].

1.2 Where do we stand?

The Standard Model was completed in the 1970's and has been successfully tested since, in thousands of measurements whose fractional precisions reach one part per trillion [11]. However, observations and theoretical considerations suggest that the Standard Model is likely to be an effective theory, valid at the eV–TeV energies probed so far, that should be completed by a more general full theory valid over a broader range of high energies. Open questions that support this interpretation include the lack of an explanation for a dynamical origin for the observed asymmetry between matter and antimatter in the

¹Only particles with *left* chirality are influenced by the weak interaction.

universe, the strikingly large differences observed between fermion masses, the possible instability of the Higgs vacuum, the conceptual and technical difficulties in achieving a description of gravity consistent with quantum mechanics, or the postulated large amounts of non-interacting matter (dark matter), introduced to justify cosmological observations.

Extending the Standard Model to higher energy-scales is the main goal of today's particle physics, in an attempt at addressing these and other open issues. Current strategies to extend the Standard Model can broadly be classified into two synergic approaches.

The energy-frontier, *direct* approach aims at using high-energy collisions to produce on-shell particles (that is, particles satisfying the energy-momentum conservation at production) not included in the Standard Model, and detect directly their decay products, thus gaining direct evidence of their existence.² Historically this offered striking experimental evidence of new phenomena, when energetically accessible, but its reach is limited by the maximum energy available at colliders.

The intensity-frontier, *indirect* approach broadly consists in searching for significant differences between precise measurements and equally precise SM predictions in lower-energy processes sensitive to non-SM contributions. A semi-intuitive, although simplified conceptual representation of the subtending idea is that exchanges of virtual (off-mass-shell) particles of arbitrary high mass, including those not described in the Standard Model, occur in the transition, thus altering the amplitudes in an observable manner. The presence of virtual particles, which may imply a temporary non-conservation of energy if interpreted classically, is allowed by Heisenberg's uncertainty principle $\Delta E \Delta t > \hbar/2$. Experimental evidence is typically harder to establish, but the reach is not bounded by the maximum collision energy reachable by experiments. A large portion of the effort in this approach is centered on the weak-interactions of quarks (so called 'flavor physics').

1.3 Flavor physics in the Standard Model

Although technically flavor physics includes also lepton interactions, I restrict the scope by referring solely to the quark interactions here.

The role of flavor in shaping the Standard Model has been central since the early days of particle physics. However, its prominence in determining the theory can perhaps be tracked down to the early 1960's with the apparent inconsistency between weak coupling constants measured in muon decay, neutron decay, and strange-particle decays. Such inconsistency was first addressed by Gell-Mann and Levy [12] and then Cabibbo [13], who postulated differing mass (d) and weak (d') eigenstates for down-type quarks. This was achieved by introducing a mixing angle (θ_C) between the s quark and d quark, the only two down-type quarks known at the time. While Cabibbo's theory addressed efficiently the difference of weak coupling constants, it also predicted a rate for the $K_L^0 \rightarrow \mu^+ \mu^-$ and other kaon decays inconsistent with the experimental exclusion limits at the time. Glashow, Iliopoulos, and Maiani addressed the conundrum by postulating the existence of a fourth quark (c) of $2 \text{ GeV}/c^2$ mass, whose contribution in the $K_L^0 \rightarrow \mu^+ \mu^-$ decay amplitude would cancel the u quark contribution, suppressing the branching fraction down to values consistent with experimental limits [14]. The charm quark was then discovered four years after the prediction, showing the compelling power of the indirect approach. In addition, in 1973, when only three quarks were known, Kobayashi and Maskawa generalized Cabibbo's theory from a four-quark model to a six-quark model to accommodate the phenomenon of CP

²*Mass shell* is jargon for mass hyperboloid, which identifies the hyperboloid in energy-momentum space describing the solutions to the mass-energy equivalence equation $E^2 = (pc)^2 + m^2c^4$. A particle *on-mass-shell* satisfies this relation.

violation observed in 1964 [15]. They introduced a complex unitary matrix to describe the relations between mass (unprimed) and weak (primed) interaction eigenstates of quarks as seen by W^\pm bosons. This is known as the Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix or V_{CKM} ,

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

Each V_{ij} matrix element encapsulates the weak-interaction coupling between an up-type i and down-type j quarks. It is a $N \times N$ CKM matrix with $(N-1)^2$ free parameters, where N is the number of quark families [16]. If $N = 2$, the only free parameter is the Cabibbo angle $\theta_C \approx 13^\circ$, whereas if $N = 3$, the free parameters are three Euler angles (θ_{12} , θ_{13} , and θ_{23}) and a complex phase (δ), which allows for CP -violating couplings [11]. The matrix is most conveniently written in the so-called *Wolfenstein parametrization*, an expansion in the small parameter $\lambda = \sin \theta_C \approx 0.23$ that makes explicit the observed hierarchy between its elements [17],

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) ,$$

where

$$\lambda = \frac{V_{us}}{\sqrt{V_{ud}^2 + V_{us}^2}}, \quad A\lambda^2 = \lambda \frac{V_{cb}}{V_{us}}, \quad \text{and} \quad A\lambda^3(\rho + i\eta) = V_{ub}^* .$$

The parameter λ expresses the mixing between the first and second quark generations, A and ρ are real parameters, and η is a complex phase that introduces CP violation. The unitarity condition $V_{\text{CKM}}V_{\text{CKM}}^\dagger = \mathbb{1}$ yields nine relations,

$$\begin{aligned} |V_{ud}|^2 + |V_{cd}|^2 + |V_{td}|^2 = 1 & \quad V_{us}^*V_{ud} + V_{cs}^*V_{cd} + V_{ts}^*V_{td} = 0 & \quad V_{ud}V_{cd}^* + V_{us}V_{cs}^* + V_{ub}V_{cb}^* = 0 , \\ |V_{us}|^2 + |V_{cs}|^2 + |V_{ts}|^2 = 1 & \quad V_{ub}^*V_{ud} + V_{cb}^*V_{cd} + V_{tb}^*V_{td} = 0 & \quad V_{ud}V_{td}^* + V_{us}V_{ts}^* + V_{ub}V_{tb}^* = 0 , \\ |V_{ub}|^2 + |V_{cb}|^2 + |V_{tb}|^2 = 1 & \quad V_{ub}^*V_{us} + V_{cb}^*V_{cs} + V_{tb}^*V_{ts} = 0 & \quad V_{cd}V_{td}^* + V_{cs}V_{ts}^* + V_{cb}V_{tb}^* = 0 , \end{aligned}$$

which are sums of three complex numbers each. The six equations summing to zero prompt a convenient geometric representation in the complex plane in terms of so-called *unitarity triangles*. A CP conserving theory would yield null-area triangles or, equivalently, a vanishing Jarlskog invariant $J = \Im(V_{us}V_{cb}V_{ub}^*V_{cs}^*)$ [18–20]. All elements of the second equation in the second row have similar magnitudes, yielding a notable triangle referred to as ‘the Unitarity Triangle’, shown in Fig. 1.2. Conventionally, side sizes are normalized to the length of the base, and the three angles are labelled α or ϕ_2 , β or ϕ_1 , and γ or ϕ_3 .

In addition to CP violation, the flavor-mixing phenomenon, which involves flavored neutral mesons $|M\rangle$, enriches significantly the phenomenology. Flavor quantum numbers are conserved in strong interactions and thus flavor eigenstates are eigenstates of strong interactions. Weak interactions do not conserve flavor, allowing $|M\rangle$ to undergo a transition into $|\bar{M}\rangle$ (or vice versa), which changes flavor by two units. Because the full Hamiltonian contains strong and weak interactions, its eigenstates (which are the particles we observe, with definite masses and lifetimes) are linear superpositions of flavor eigenstates $|M\rangle$ and $|\bar{M}\rangle$.

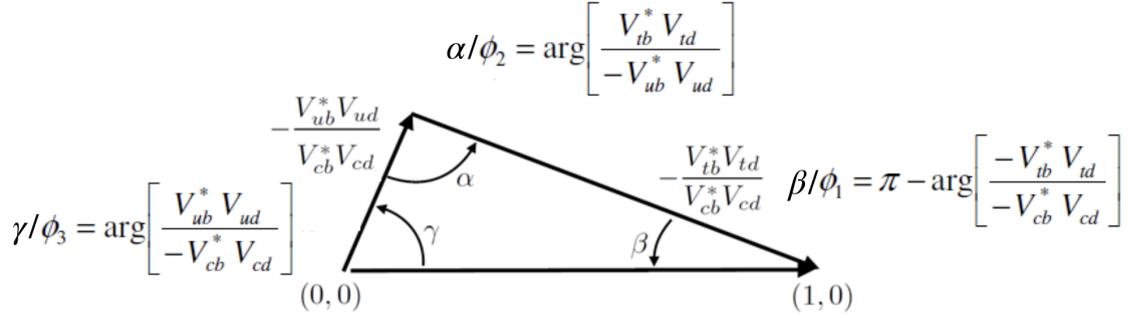


Figure 1.2: Graphical representation of the Unitarity Triangle.

1.4 Time evolution of neutral flavored mesons

The time evolution of neutral flavored mesons is approximated by the Schrödinger equation (in 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

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

and \mathbf{M} and $\mathbf{\Gamma}$ 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 mesons, which decay. The diagonal elements of \mathbf{M} and $\mathbf{\Gamma}$ are related respectively to the mass and the lifetime of the pure flavor eigenstates. Conservation of CPT symmetry 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 interactions. Thus, the diagonal elements of \mathcal{H}_{eff} can be identified as the strong and electromagnetic 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 of \mathbf{M} and $\mathbf{\Gamma}$. These elements involve weak interactions and can be considered as a weak perturbation Hamiltonian \mathcal{H}_w . One obtains

$$\mathcal{H}_{\text{eff}} = \begin{pmatrix} H_0 & 0 \\ 0 & H_0 \end{pmatrix} + \mathcal{H}_w = \begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{pmatrix}. \quad (1.1)$$

Since the phase factors of $|M\rangle$ and $|\bar{M}\rangle$ can be arbitrarily adjusted without changing the physics, only the difference between them, ϕ_{12} , is relevant. Flavor mixing is parametrized by five real parameters [21]

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

By diagonalizing the Hamilton operator in Eq.(1.1), the time evolution of the meson-antimeson system gets described in terms of its eigenvalues

$$\omega_{\pm} = M - \frac{i}{2}\Gamma \pm \sqrt{(M_{12} - \frac{i}{2}\Gamma_{12})(M_{12}^* - \frac{i}{2}\Gamma_{12}^*)} = M - \frac{i}{2}\Gamma \pm pq \quad (1.2)$$

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

$$\begin{aligned} |M_+(t)\rangle &\equiv p|M(t)\rangle + q|\bar{M}(t)\rangle, \\ |M_-(t)\rangle &\equiv p|M(t)\rangle - q|\bar{M}(t)\rangle. \end{aligned} \quad (1.3)$$

As an example, Fig. 1.3 shows the leading-order Feynman diagrams contributing to $B^0 - \bar{B}^0$ mixing. Flavor mixing induces flavor oscillations in the time evolution of neutral flavored mesons. Flavor oscillations generate additional time-evolution paths to the simple decay, which interfere with the decay thus enriching the dynamics and our opportunities to study it.

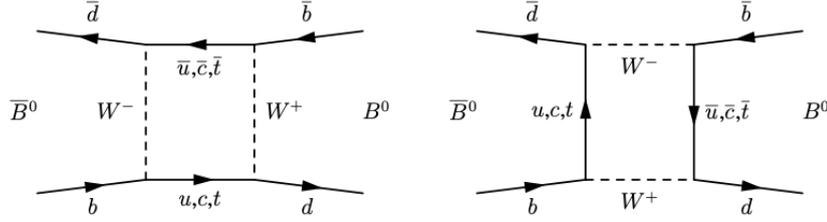


Figure 1.3: Feynman diagrams for leading-order amplitudes contributing to $B^0 - \bar{B}^0$ mixing.

The complex numbers p and q satisfy

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}}} = e^{-i\phi_M} \sqrt{\frac{|M_{12}| + \frac{i}{2}|\Gamma_{12}|e^{i\phi_{12}}}{|M_{12}| + \frac{i}{2}|\Gamma_{12}|e^{-i\phi_{12}}}} \quad \text{with } |p|^2 + |q|^2 = 1, \quad (1.4)$$

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 are derived from the eigenvalues in Eq. (1.2),

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

which can be recasted in terms of the following observables

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

The time evolution of $|M_{\pm}\rangle$ is given by

$$\begin{aligned} |M_+(t)\rangle &= e^{-i\omega_+t} (p|M\rangle + q|\bar{M}\rangle), \\ |M_-(t)\rangle &= e^{-i\omega_-t} (p|M\rangle - q|\bar{M}\rangle), \end{aligned} \quad (1.5)$$

where both mass eigenstates decay as $e^{-\Gamma_{\pm}t}$ with a modulation generated by the complex phase $e^{im_{\pm}t}$. Considering now Eq.(1.3), the time evolution of the pure flavor states reads

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

Insertion of the time evolution in Eq.(1.4) yields

$$\begin{aligned} |M(t)\rangle &= \frac{1}{2} \left((e^{-i\omega_+t} + e^{-i\omega_-t}) |M\rangle + \frac{q}{p} (e^{-i\omega_+t} - e^{-i\omega_-t}) |\bar{M}\rangle \right) \\ &= \frac{1}{2} e^{-i\omega_+t} \left(\left(1 + e^{-i(\Delta m - \frac{i}{2}\Delta\Gamma)t}\right) |M\rangle + \frac{q}{p} \left(1 - e^{-i(\Delta m - \frac{i}{2}\Delta\Gamma)t}\right) |\bar{M}\rangle \right), \end{aligned} \quad (1.6)$$

$$\begin{aligned} |\bar{M}(t)\rangle &= \frac{1}{2} \left(\frac{p}{q} (e^{-i\omega_+t} - e^{-i\omega_-t}) |M\rangle + (e^{-i\omega_+t} + e^{-i\omega_-t}) |\bar{M}\rangle \right) \\ &= \frac{1}{2} e^{-i\omega_+t} \left(\frac{p}{q} \left(1 - e^{-i(\Delta m - \frac{i}{2}\Delta\Gamma)t}\right) |M\rangle + \left(1 + e^{-i(\Delta m - \frac{i}{2}\Delta\Gamma)t}\right) |\bar{M}\rangle \right). \end{aligned} \quad (1.7)$$

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 $|\bar{M}\rangle$ [22]. 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

$$|\langle \bar{M} | M(t) \rangle|^2 = \frac{1}{2} \left| \frac{q}{p} \right|^2 e^{-\Gamma t} \left(\cosh\left(\frac{\Delta\Gamma t}{2}\right) - \cos(\Delta m t) \right), \quad (1.8)$$

and to observe $|M\rangle$ is

$$|\langle M | M(t) \rangle|^2 = \frac{1}{2} e^{-\Gamma t} \left(\cosh\left(\frac{\Delta\Gamma t}{2}\right) + \cos(\Delta m t) \right). \quad (1.9)$$

The time evolutions of the transition amplitudes are

$$A_f(t) = \langle f | \mathcal{H} | M(t) \rangle = A e^{-\frac{\Gamma t}{2}} \left(\cos\frac{\Delta m t}{2} + i\lambda \sin\frac{\Delta m t}{2} \right) \quad (1.10)$$

$$\bar{A}(t) = \langle f | \mathcal{H} | \bar{M}(t) \rangle = \bar{A} e^{-\frac{\Gamma t}{2}} \left(\cos\frac{\Delta m t}{2} + \frac{i}{\lambda} \sin\frac{\Delta m t}{2} \right), \quad (1.11)$$

and the time-dependent asymmetries are

$$a_f = \frac{\Gamma(M \rightarrow f) - \Gamma(\bar{M} \rightarrow f)}{\Gamma(M \rightarrow f) + \Gamma(\bar{M} \rightarrow f)} \quad (1.12)$$

$$= \frac{|\lambda|^2 - 1}{1 + |\lambda|^2} \cos(\Delta m t) + \frac{2\text{Im}(\lambda)}{1 + |\lambda|^2} \sin(\Delta m t) \quad (1.13)$$

$$= A_{CP} \cos(\Delta m t) + \mathcal{S}_{CP} \sin(\Delta m t) \quad (1.14)$$

where C is the direct CP -violation parameter, and S is the mixing-induced CP -violation parameter.

These are two of the three manifestations of nonconservation of charge-parity symmetry. Depending on how the CP -violating complex phase enters the amplitude of the dynamical evolution of a system, CP violation can be classified into three distinct phenomenologies. In the most general case, one considers the amplitudes of transitions governed by the effective Hamiltonian \mathcal{H}_{eff} , of a flavored meson $|M\rangle$ and of its antiparticle $|\bar{M}\rangle$, into a final state $|f\rangle$ and into the CP conjugate $|\bar{f}\rangle$, respectively,

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

Violation of CP symmetry *in decay* occurs if

$$\left| \frac{A_f}{\bar{A}_{\bar{f}}} \right| \neq 1 \quad \text{or} \quad \left| \frac{\bar{A}_{\bar{f}}}{A_f} \right| \neq 1, \quad (1.15)$$

and appears experimentally as a difference between the $|\bar{M}\rangle \rightarrow |\bar{f}\rangle$ and $|M\rangle \rightarrow |f\rangle$ decay rates. This is the only CP -violation phenomenology possible for both neutral and charged mesons, and for baryons. Since, in general, various amplitudes contribute to a single decay, the total decay amplitudes A_f and $\bar{A}_{\bar{f}}$ are written as sums of the individual contributions

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

where symbols δ_i indicate CP -conserving phases, and ϕ_i are CP -violating phases associated with the elements of the CKM matrix that appear in each amplitude. As CP -conserving couplings are real, the corresponding phases are invariant under CP transformation. Since a CP transformation turns CKM coefficients into their complex conjugate, the CP -violating phases flip its sign. The condition of CP violation in decay in Eq. (1.15) is satisfied if

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

Thus, CP violation in decay can occur only if at least two amplitudes with different CP -conserving and CP -violating phases contribute to a decay process. The squared magnitudes of the total decay amplitudes $|A_f|^2$ and $|\bar{A}_{\bar{f}}|^2$ are the quantities typically accessible experimentally since they are proportional to the total decay rates. The individual amplitudes A_i are often difficult to compute theoretically as they typically involve contributions from strong-interaction amplitudes at low energy. These are non-perturbative and therefore hard to calculate, leading to large uncertainties. Thus, observables that depend only on the weak phases, such as the CKM angles, allow to test SM predictions in a reliable way.

Violation of CP symmetry *in mixing* implies

$$\left| \frac{q}{p} \right| = \left| \frac{1 - \varepsilon}{1 + \varepsilon} \right| \neq 1 \quad \Rightarrow \quad |\varepsilon| \neq 0, \quad (1.16)$$

where q and p are introduced in Eq. (1.1). In this case, CP violation generates a difference between the flavor-oscillation rates $|M\rangle \rightarrow |\bar{M}\rangle$ and $|\bar{M}\rangle \rightarrow |M\rangle$, which can be observed as a charge-dependent asymmetry in the yields of charged leptons from semileptonic decays of oscillating B^0 mesons.

Violation of CP symmetry *in the interference* of mixing and decay is observed when the neutral mesons $|M\rangle$ and $|\bar{M}\rangle$ can decay into a common final state $|f\rangle$, preferentially a pure CP eigenstate f_{CP} ,

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

Even if CP is conserved in mixing and in decay separately, i.e., if $|\bar{A}_{f_{CP}}/A_{f_{CP}}| = |q/p| = 1$, the combination of the decay and mixing phases can generate a total phase difference and thus an interference between these two processes, generating a violation of CP symmetry. Introducing the complex quantity λ_{CP} , the condition for CP violation is

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

1.5 Flavor physics to overcome the Standard Model

Many physicists find the current understanding of flavor dynamics unsatisfactory. The observed hierarchies between quark masses and couplings seem too regular to be accidental and the abundance of free parameters (six quark masses and four couplings) suggests the possibility of a deeper, more fundamental theory possibly based on a reduced set of parameters. In addition, while the CKM mechanism offers a framework to include CP violation in the Standard Model, it does not really enlighten the origin for such a singular phenomenon.

But even in the absence of a deeper understanding of the origin of CP violation, naturalness arguments indicate that most generic extensions of the Standard Model would involve additional sources of CP violation. These and other considerations support the notion that a more detailed and complete study of the phenomenology of quarks dynamics may reveal useful information to guide searches for SM extensions.

The abundance and diversity of experimentally accessible processes to measure redundantly a reduced set of parameters makes indirect searches in the flavor sector a powerful option for exploring non-SM dynamics. In fact, even if no deviations from the Standard Model will be found, the resulting stringent constraints on SM extensions are expected to remain useful in informing future searches.

The two classes of flavor-physics processes most promising for probing contributions of non-SM particles are *flavor-changing-neutral-currents* and *CP -violating* processes.

Flavor-changing neutral currents (FCNC) are processes in which quark flavor changes in the transition, but quark electric charge does not. The processes are suppressed in the Standard Model, because they occur only through higher-order amplitudes involving the internal exchange of W^\pm bosons ('loop amplitudes'), as shown in Fig. 1.4. Such amplitudes are naturally sensitive to non-SM contributions, since any particle with proper quantum numbers and nearly arbitrary mass can replace the SM-quark closed-line in these diagrams thus altering the rate. Hence, FCNC are powerful in signaling contributions from non-SM particles if rate enhancements, or suppressions, with respect to Standard Model expectations are observed.

In addition to rate alterations, the phenomenon of CP violation offer additional avenues to uncover or characterize possible non-SM contributions. Alterations of the CP -violating phases with respect to those predicted by the SM are generically expected in a broad class of SM-extensions. Observing experimental evidence of those phases offers further opportunities to explore the dynamics, even if total rates are unaffected.

1.6 Current flavor status

Measurements of parameters associated with quark-flavor physics have been performed in many dedicated, or general-purpose, experiments in the last three decades, including CLEO, CPLEAR, NA32, NA48, NA62, KTeV, SLD, OPAL, L3, ALEPH, DELPHI, BaBar, Belle, CDF, CDFII, LHCb, BESIII, ATLAS, and CMS [23].

The current status of constraints on sides and angles of the Unitarity Triangle is shown in Fig. 1.5 [24]. Measurements of $\sin 2\beta$ reached a precision of 1%, mainly due to the availability of large samples of $B^0 \rightarrow J/\psi K^0$ decays in e^+e^- and pp collisions, while the angle α is known down to a 4% precision from $B \rightarrow hh$ decays, and h represents a charged or neutral π or ρ) in e^+e^- collisions. The angle γ is measured with 4% precision using combinations of several measurements involving $B \rightarrow DK$ decays reconstructed in e^+e^- and pp collisions. Discrepancies in the determinations of $|V_{cb}|$ and $|V_{ub}|$ are found between

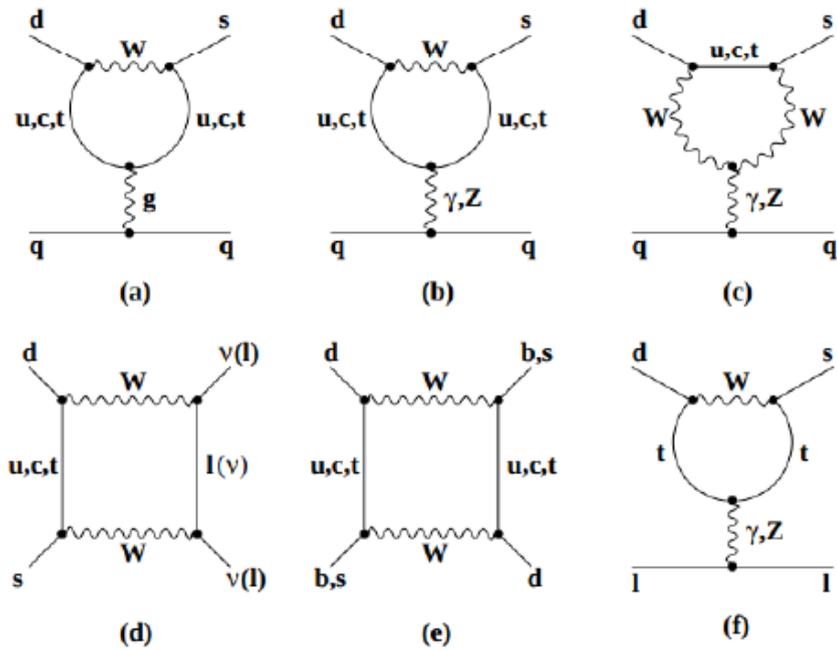


Figure 1.4: Examples of leading FCNC diagrams.

values measured using different analyses of semileptonic decays, mainly performed in e^+e^- collisions.

The decay-width difference of the $B_s^0 - \bar{B}_s^0$ system is determined with 5% precision in pp collisions, while measurements are not yet precise enough to discern the expected non-zero value for the $B^0 - \bar{B}^0$ system. Mass differences in both systems are known with better than 1% precision from pp and $p\bar{p}$ collisions. In addition, many other measurements in charm and kaon physics contribute that are not straightforwardly represented in the Unitarity Triangle.

The resulting global picture is that the CKM interpretation of quark-flavor phenomenology is the dominant mechanism at play in the dynamics. Despite the first-order consistency of the experimental flavor picture with the CKM theory, possible deviations of up to 10%-15% are still unconstrained, especially those associated with loop-mediated processes, leaving sufficient room for non-SM physics. It is especially promising that most of the relevant measurements are currently dominated by statistical uncertainties, offering therefore fruitful opportunities for the two experiments that will contribute the most in the next decade, LHCb and Belle II.

This is all the more attractive because recent direct searches for non-SM physics, mainly in pp collisions at the Large Hadron Collider (LHC) at CERN, excluded large portions of the parameter space for several proposed SM extensions, but showed no conclusive evidence of non-SM physics to date. Since plans for a higher-energy collider in the near future are still fluid, the systematic study of flavor physics emerges as a promising program to search for non-SM in the next decade.

1.7 The $B^0 \rightarrow \pi^0\pi^0$ decay

The study of $B \rightarrow \pi\pi$ decays, and of $B^0 \rightarrow \pi^0\pi^0$ decays among them, is of particular interest for searching for potential non-SM physics amplitudes that may compete with

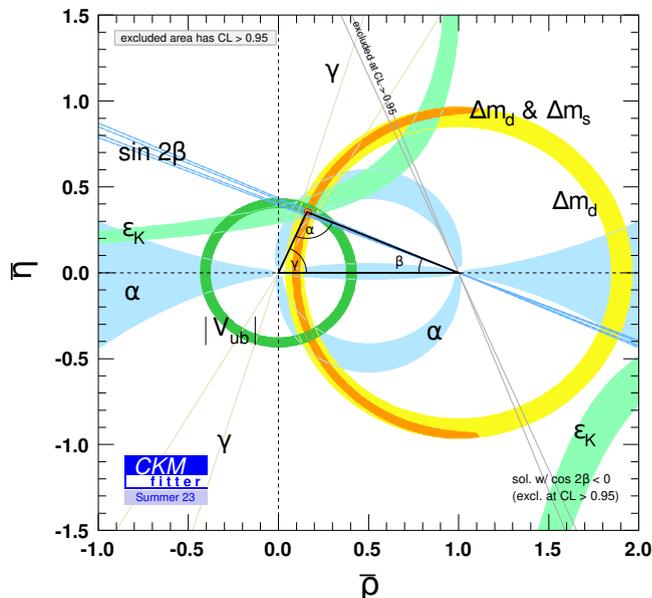


Figure 1.5: Current constraints on sides and angles of the Unitarity Triangle. Reproduced from Ref. [24].

suppressed SM decay amplitudes, for the data-driven constraints they offer to improve phenomenological models of the SM amplitudes of B decays into hadronic final states, and for their role in the determination of the parameter α .

1.7.1 $B^0 \rightarrow \pi^0\pi^0$ predictions

Quantitative predictions of $B^0 \rightarrow \pi^0\pi^0$ decay properties require the computation of hadronic matrix elements, which is challenging as it involves both short-distance perturbative and long-distance non-perturbative QCD. At long distances, the strong interaction – which cannot be treated perturbatively at the hadronic scale – is relevant for the formation of the hadronic final state f and for the initial bound state.

In the past years, predictions have exploited increasingly refined techniques and included higher-order amplitudes, yielding to various predictions between 0.2×10^{-6} and 1×10^{-6} for the branching fraction, and between 0.45 and 0.9 for the decay-rate asymmetry [25–28].

Three dominant amplitudes are expected to contribute with similar magnitude to the $B^0 \rightarrow \pi^0\pi^0$ decay: the color-suppressed tree amplitude, the QCD-penguin amplitude, and the electroweak penguin amplitude. The color-favored tree amplitude does not contribute. Any significant deviations from the decay rate, or any other observable of interest as predicted by the Standard Model, could indicate contributions from non-SM particles. A possible significant example is that, currently, the ratio between color-allowed and -suppressed tree contributions for $\pi\pi$ modes does not agree well with theoretical expectations.

Prior to the first measurements, in 2003, the $B^0 \rightarrow \pi^0\pi^0$ branching fraction was expected to be smaller than 1×10^{-6} [27, 29]. This is intuitively seen in the topological

amplitude parametrization [25],

$$\sqrt{2}A(B^+ \rightarrow \pi^+\pi^0) = -T \left[1 + \frac{C}{T} + \frac{P_{EW}}{T} e^{i\alpha} \right], \quad (1.17)$$

$$A(B^0 \rightarrow \pi^+\pi^-) = -T \left[1 + \frac{P}{T} e^{i\alpha} \right], \quad (1.18)$$

$$\sqrt{2}A(B^0 \rightarrow \pi^0\pi^0) = T \left[\left(\frac{P}{T} - \frac{P_{EW}}{T} \right) e^{i\alpha} - \frac{C}{T} \right], \quad (1.19)$$

$$(1.20)$$

where, T , C , P and P_{EW} stand for the color-allowed tree, color-suppressed tree, gluonic penguin, and electroweak penguin amplitudes, respectively. The color-suppressed electroweak penguin, exchange, and penguin-annihilation amplitudes are neglected. In a first approximation, one would expect the amplitudes to obey the hierarchy dictated by SM constraints [30],

$$\frac{P}{T} \approx \lambda, \quad \frac{C}{T} \approx \lambda, \quad \frac{P_{EW}}{T} \approx \lambda^2. \quad (1.21)$$

Hence, the main contributions to the $B^0 \rightarrow \pi^0\pi^0$ decays would be expected to come from color-suppressed tree and gluonic penguin amplitudes. Furthermore, the P_{EW} contribution to $B^+ \rightarrow \pi^+\pi^0$ and $B^0 \rightarrow \pi^+\pi^-$ would be expected to be small, especially in the latter case due to color suppression. The $B^0 \rightarrow \pi^0\pi^0$ branching ratio would be expected to be $\mathcal{O}(\lambda^2)$ of the $B^0 \rightarrow \pi^+\pi^-$, i.e., $\mathcal{B}(B^0 \rightarrow \pi^+\pi^-) \gg \mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$.

However, the $\mathcal{B}(B^0 \rightarrow \pi^+\pi^-) = (5.12 \pm 0.19) \times 10^{-6}$ and $\mathcal{B}(B^0 \rightarrow \pi^0\pi^0) = (1.59 \pm 0.26) \times 10^{-6}$ measurements contradicted this prediction. Even though there might be mundane explanations for this based on unaccounted for hadronic effects, this discrepancy could also be suggestive of the presence of anomalously enhanced P_{EW} or C amplitudes, possibly due to non-SM physics. Under flavor SU(3) symmetry, the amplitudes of charmless decay modes are related: if P_{EW} contributions in $B \rightarrow K\pi$ decays are large, then they are expected to be large in $B \rightarrow \pi\pi$ decays, too. However, measurements from other charmless decays disfavor large electroweak penguin amplitudes: for example, measurements show that $\mathcal{B}(B^0 \rightarrow \rho^+\rho^-) \gg \mathcal{B}(B^0 \rightarrow \rho^0\rho^0)$ [11]. Furthermore, leading-order perturbative QCD implies a negligible C amplitude, with next-to-leading-order and enhancement from vertex corrections being insufficient to compensate [28]. It seems therefore appropriate to consider additional factors, such as SU(3) breaking effects, which might imply non-SM physics.

In recent years, theoretical predictions based on pQCD or QCD factorization have been agreeing with the data [31, 32]. Predictions in theoretical frameworks and models exhibit considerable variance for the relative size and phase of penguins and colour-suppressed tree amplitudes. More precise measurements of these quantities would help to discriminate among the various solutions proposed to address this discrepancy.

1.7.2 The angle α

The properties of $B^0 \rightarrow \pi^0\pi^0$ decays are also an important input for the determination of the CKM angle $\alpha = \arg[-(V_{td}V_{tb}^*)/(V_{ud}V_{ub}^*)]$. With an uncertainty of about 4%, this parameter is currently among the major limiting factors on the global consistency check of the Unitarity Triangle and one of the primary goals of Belle II physics.

In this work I will not provide a direct measurement of the angle α , but given its importance in the Standard Model and the relevance $B^0 \rightarrow \pi^0\pi^0$ will have in its future, I provide here an overview on the matter.

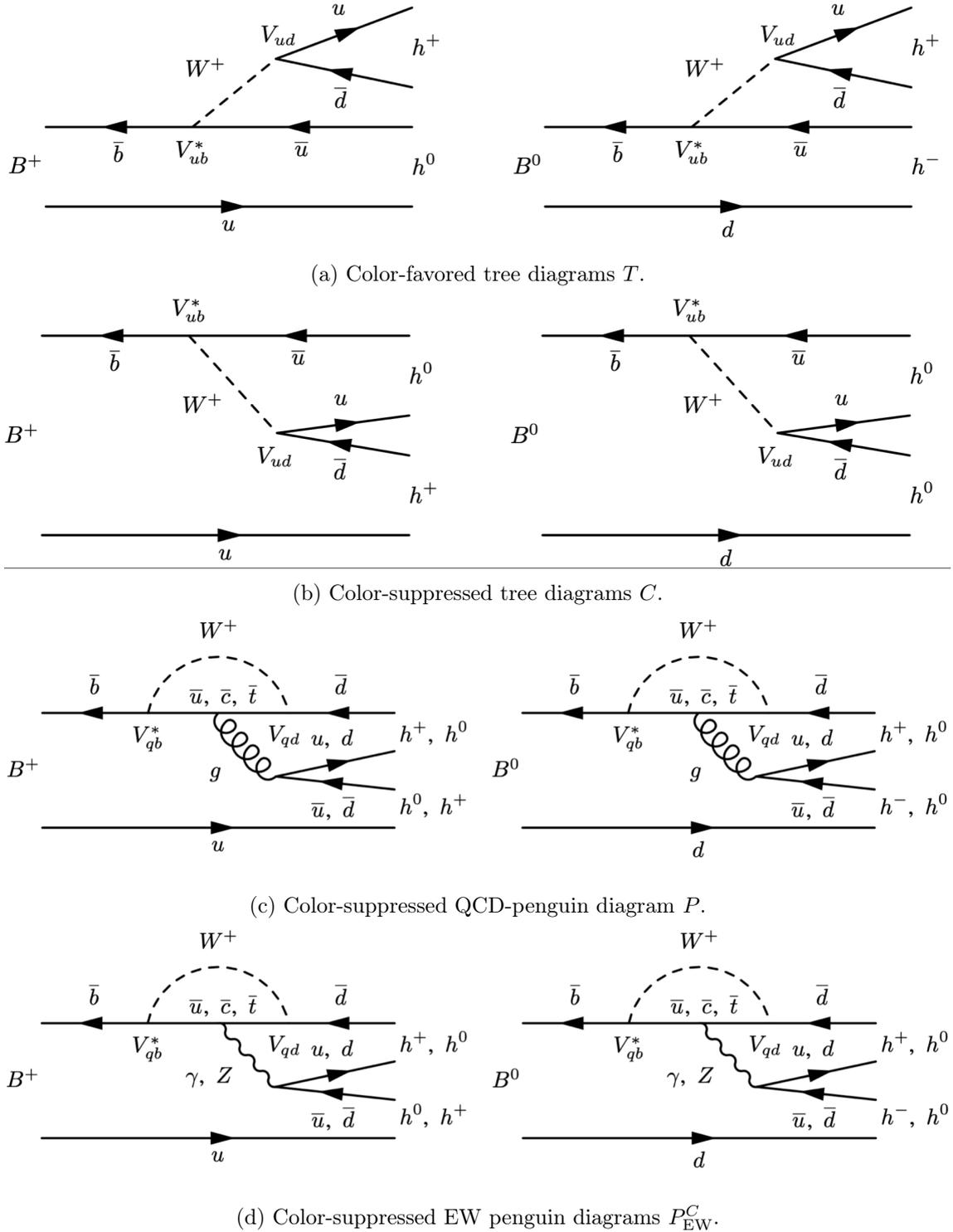


Figure 1.6: Feynman diagrams for the dominant amplitudes of tree, color-suppressed QCD-penguin, and color-suppressed EW-penguin transitions contributing to $B^+ \rightarrow h^- h^0$ (left) and to $B^0 \rightarrow h^+ h^-$ and $h^0 h^0$ (right). The generic shorthand h indicates a pion or ρ meson.

Any B decay induced by a $b \rightarrow u\bar{d}$ transition, which means decays of charged and neutral B mesons to two, three, or four pions, is sensitive to α . Sensitive decays with three and four final state pions occur through intermediate resonances, for example $B \rightarrow \rho(\rightarrow \pi\pi)\pi$, $B \rightarrow \rho(\rightarrow \pi\pi)\rho(\rightarrow \pi\pi)$ and $B \rightarrow a_1(\rightarrow \pi\pi\pi)\pi$. Figure 1.6 shows the leading-order tree and QCD-penguin Feynman diagrams, together with the color-suppressed electroweak (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 $h_1^i h_2^j$ indicates a pair of unflavoured light mesons such that h is π or ρ and i, j is $-$, 0 , or $+$. A more complete overview of higher-order amplitudes is in Refs. [33, 34].

If we consider, as an example, the decays of neutral B^0 mesons only, and we assume that only tree-level diagrams contribute (Fig. 1.6a and 1.6b), a measurement of the differential decay rate of B^0 and \bar{B}^0 mesons, and therefore of the CP asymmetry due to interference between direct decays and decays following mixing, is proportional to

$$\lambda_{hh} = \left(\frac{q}{p}\right)_{B^0} \left(\frac{\bar{A}^{hh}}{A^{hh}}\right) = \frac{V_{tb}^* V_{td} V_{ud}^* V_{ub}}{V_{tb} V_{td}^* V_{ud} V_{ub}^*} = e^{2i\alpha}. \quad (1.22)$$

where the mixing and decay amplitudes are transparently shown. The $(V_{ud}^* V_{ub}) / (V_{ud} V_{ub}^*)$ term corresponds to the tree decay amplitude, while the $(V_{tb}^* V_{td}) / (V_{tb} V_{td}^*)$ term corresponds to the mixing amplitude. In the mixing amplitude, we consider only contributions with an internal top quark neglecting those with a charm or up quark, as the mixing amplitude is proportional to the squared mass of the exchanged up-type quark.

Hence, if penguin amplitudes were nonexistent, a measurement of decay-rate asymmetry between flavor-tagged $B^0 \rightarrow hh$ decays would offer direct access to α through the corresponding CP -violating coefficients of the cosine and sine terms of the time evolution, $A_{CP} = 0$, $S_{CP} = \eta_{CP} \sin(2\alpha)$ (see Eq.(1.14)).

The value $A_{CP}^{\rho^+\rho^-} = 0.00 \pm 0.09$ observed in the decay $B^0 \rightarrow \rho^+\rho^-$ [11] shows that the approximation of tree-amplitude dominance might be reasonable up to a 10% uncertainty. If large penguin amplitudes would contribute, then the CP asymmetry in decay A_{CP} would sense them unless the difference in CP -conserving phases is small. However, the value $A_{CP}^{\pi^+\pi^-} = 0.31 \pm 0.03$ measured for the decay $B^0 \rightarrow \pi^+\pi^-$ differs considerably from zero [11], revealing that penguin contributions are not negligible in general, and especially so for decays $B^0 \rightarrow h^0 h^0$, where the leading-order tree contribution is color suppressed.

Hence, penguin contributions have to be considered. Since the fine-structure constant α is smaller than the strong coupling coefficient α_s , EW penguin amplitudes are expected to be $\mathcal{O}(0.1)$ of QCD penguin amplitudes and can be neglected to a good approximation [35, 36]. The $B \rightarrow h_1^i h_2^j$ decay amplitude, where CKM coefficients are factorized, reads [37]

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

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. The unitarity relation of the CKM matrix given in Sec. 1.3, allows for rewriting the decay amplitude A^{ij} as

$$A^{ij} = V_{ud} V_{ub}^* (T_{uc}^{ij} + P_u^{ij} - P_c^{ij}) + V_{td} V_{tb}^* (P_t^{ij} - P_c^{ij}).$$

By defining the tree and 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 squared magnitudes 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 -violating quantity λ_{hh} becomes

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

The amplitudes T^{ij} and P^{ij} can be written in terms of magnitude and hadronic phase, leading to

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

Charge-parity invariance of strong interactions requires the hadronic amplitudes T^{ij} and P^{ij} to be invariant under CP transformations. Thus, a CP transformation on A^{ij} applies complex conjugation on weak phases only. Rotating consistently all CP -transformed amplitudes \bar{A}^{ij} to absorb the mixing phase, i.e., $\tilde{A}^{ij} = e^{-2i\beta} \bar{A}^{ij}$, the ratio between amplitudes \tilde{A}^{ij} and A^{ij} yields

$$\frac{\tilde{A}^{ij}}{A^{ij}} = \lambda_{hh} = \frac{e^{-i(\beta+\gamma)} e^{i(\delta_T-\delta_P)} |T^{ij}| - |P^{ij}|}{e^{i(\beta+\gamma)} e^{i(\delta_T-\delta_P)} |T^{ij}| - |P^{ij}|} = \frac{e^{i(\delta+\alpha)} |T^{ij}| + |P^{ij}|}{e^{i(\delta-\alpha)} |T^{ij}| + |P^{ij}|}, \quad (1.23)$$

where $\delta = \delta_T - \delta_P$ and $\alpha - \pi = -\beta - \gamma$ (see Fig. 1.2). If penguin contributions are negligible, i.e., $|P^{ij}| \approx 0$, one obtains Eq. (1.22). If penguin contributions are non-negligible, a direct determination of α is not possible, and an effective angle α_{eff} , also referred to as penguin *polluted* angle,

$$\frac{\tilde{A}^{ij}}{A^{ij}} = \lambda_{hh} = \left| \frac{\tilde{A}^{ij}}{A^{ij}} \right| e^{2i\alpha_{\text{eff}}}, \quad (1.24)$$

is observed.

In 1990, Michael Gronau and David London proposed to determine the size of penguin pollution from data using known symmetry relations among the hadronic amplitudes, based on a $SU(2)$ isospin analysis of $B \rightarrow \pi\pi$ decays (Sec. 1.7.2.1) [38]. A later review clarified that the method applies to $B \rightarrow \rho\rho$ and $B \rightarrow \rho\pi$ too [39].

A further approach to determine the penguin contribution is to exploit the (approximate) $SU(3)$ flavor symmetry between the decays $B \rightarrow K\pi$, $B \rightarrow KK$, and $B \rightarrow \pi\pi$ [40, 41], or, similarly, for the decays $B \rightarrow K^*\rho$ and $B \rightarrow \rho\rho$ [42]. Isospin symmetry in a Dalitz plot analysis, and $SU(3)$ flavor symmetry have been applied to determine α with $B \rightarrow \rho\pi$ decays [43, 44]. A treatment based on $SU(3)$ was proposed also for the $B \rightarrow a_1(1260)\pi$ system [45]. Other approaches consider the extraction of α from $B \rightarrow a_0(980)(\rightarrow \eta\pi)\pi$, $B \rightarrow a_0(980)(\rightarrow \eta\pi)\rho(\rightarrow \pi\pi)$, and other B decays to resonances with isospin $I = 1$ [46]. However, the uncertainties associated with the assumption of $SU(3)$ flavor symmetry are large and poorly known, making those methods scarcely relevant for CKM parameter estimation.

In the following, the determination of α based on an isospin analyses of the systems $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ is outlined.

1.7.2.1 Isospin analysis

The π and ρ mesons are, respectively, the lightest pseudoscalar (spin-0) and vector (spin-1) mesons. Both particles form $SU(2)$ isospin triplets ($I = 1$),

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

The validity of the Gronau-London isospin analysis relies on the near exact conservation of isospin symmetry during the hadronization in $B \rightarrow \pi\pi$ and $B \rightarrow \rho\rho$ decays. Thus, relations between decay amplitudes derived from the isospin symmetry are used to separate the effects of the tree and the QCD penguin contributions to single out α [38].

Given a final state $h_1^i h_2^j$ with two identical mesons $h_1, h_2 = h = \pi$ or ρ , similarly as in Eq. (1.11) the decay amplitude A^{+0} , A^{+-} , and A^{00} is written as

$$A^{ij} \equiv \langle h^i h^j | \mathcal{H}_{\text{eff}} | B^{i+j} \rangle. \quad (1.25)$$

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 are

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

Since B mesons have spin-0, the total angular momentum of the final state is $J = 0$. For a single π or ρ the isospin is 1. Spin sum rules dictate that the total final-state isospin I_f for a $\pi\pi$ or $\rho\rho$ pair 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 prescribed by the symmetry $(-1)^{J+I}$ under particle exchange [39]. This is exact for final-state particles with equal masses. In the case of ρ mesons, which have a significant width, the possible mass difference between final-state particles could generate a final state with $I_f = 1$ [47]. Nevertheless, lack of a specific enhancement of the $I_f = 1$ amplitude shows that the results for α are insensitive to the ρ width [48]. The determination of any possible $I_f = 1$ contribution requires more data and is neglected in this discussion. The three relevant final states for $I_f = 0$ or 2 are

$$|h^+ h^0\rangle = \sqrt{\frac{1}{2}} \left(|h_1^+ h_2^0\rangle + |h_1^0 h_2^+\rangle \right) = |2, 1\rangle, \quad (1.26)$$

$$|h^+ h^-\rangle = \sqrt{\frac{1}{2}} \left(|h_1^+ h_2^-\rangle + |h_1^- h_2^+\rangle \right) = \sqrt{\frac{1}{3}} |2, 0\rangle + \sqrt{\frac{2}{3}} |0, 0\rangle, \quad (1.27)$$

$$|h^0 h^0\rangle = \sqrt{\frac{2}{3}} |2, 0\rangle - \sqrt{\frac{1}{3}} |0, 0\rangle. \quad (1.28)$$

The decay amplitudes in Eq. (1.25) can be factorized in the weak decay $b \rightarrow u\bar{u}d$ corresponding to a ΔI isospin transition, and the hadronization into two light mesons [37]. Following Ref. [39], the Hamiltonian for the quark transition $\bar{b} \rightarrow \bar{u}u\bar{d}$, in terms of $A_{\Delta I}$ amplitudes, is

$$\mathcal{H}_{\text{eff}} = A_{\frac{3}{2}} \left| \frac{3}{2}, +\frac{1}{2} \right\rangle + A_{\frac{1}{2}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle. \quad (1.29)$$

Application of the operator to the initial states B^+ and B^0 yields

$$\mathcal{H}_{\text{eff}} | B^+ \rangle = \mathcal{H}_{\text{eff}} \left| \frac{1}{2}, +\frac{1}{2} \right\rangle = \sqrt{\frac{3}{4}} A_{\frac{3}{2}} |2, 1\rangle + \left(A_{\frac{1}{2}} - \frac{1}{2} A_{\frac{3}{2}} \right) |1, 1\rangle, \quad (1.30)$$

$$\mathcal{H}_{\text{eff}} | B^0 \rangle = \mathcal{H}_{\text{eff}} \left| \frac{1}{2}, -\frac{1}{2} \right\rangle = \sqrt{\frac{1}{2}} A_{\frac{3}{2}} |2, 0\rangle + \sqrt{\frac{1}{2}} \left(A_{\frac{1}{2}} + A_{\frac{3}{2}} \right) |1, 0\rangle + \sqrt{\frac{1}{2}} A_{\frac{1}{2}} |0, 0\rangle. \quad (1.31)$$

The equations above correspond to four-quark states $|\bar{u}u\bar{d}, q\rangle$, with q being u or d , while the states in Eq. (1.26)–(1.28) are two-meson states. The transition between the two involves hadronization and other rescattering effects.

Following the Wigner-Eckhart theorem [49], 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 interaction rescattering effects that are absent in the amplitudes $A_{\Delta I}$ are included into the amplitudes $A_{\Delta I, I_f}$. The projection of the states $\mathcal{H}_{\text{eff}}|B^{i+j}\rangle$ onto the final states yields

$$A^{+0}{}^{(-)} = \langle h^{+(-)}h^0|\mathcal{H}_{\text{eff}}|B^{+(-)}\rangle = \sqrt{\frac{3}{4}}A_{\frac{3}{2}, 2}^{(-)}, \quad (1.32)$$

$$A^{+-}{}^{(-)} = \langle h^+h^-|\mathcal{H}_{\text{eff}}|B^0\rangle = \sqrt{\frac{1}{6}}A_{\frac{3}{2}, 2}^{(-)} + \sqrt{\frac{1}{3}}A_{\frac{1}{2}, 0}^{(-)}, \quad (1.33)$$

$$A^{00}{}^{(-)} = \langle h^0h^0|\mathcal{H}_{\text{eff}}|B^0\rangle = \sqrt{\frac{1}{3}}A_{\frac{3}{2}, 2}^{(-)} - \sqrt{\frac{1}{6}}A_{\frac{1}{2}, 0}^{(-)}. \quad (1.34)$$

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. Equations (1.32)–(1.34) yield two relations among the hadronic amplitudes,

$$\begin{aligned} A^{+0} - A^{00} &= \sqrt{\frac{1}{2}}A^{+-}, \\ \bar{A}^{+0} - \bar{A}^{00} &= \sqrt{\frac{1}{2}}\bar{A}^{+-}, \end{aligned} \quad (1.35)$$

which allow extracting the size of the penguin contribution, so-called penguin pollution. These are referred to as isospin triangles following their representation in the complex space (Fig. 1.7). The triangle relations hold also for consistently rotated amplitudes. I use the convention $\tilde{A}^{ij} = e^{-2i\beta}\bar{A}^{ij}$.

Amplitudes from QCD penguins can lead only to $\Delta I = \frac{1}{2}$ transitions. Since the amplitudes in Eq. (1.32) lack $\Delta I = \frac{1}{2}$ components, the decay $B^\pm \rightarrow h^\pm h^0$ occurs purely as a tree amplitude. Using Eq. (1.23), one obtains

$$\frac{\tilde{A}^{+0}}{A^{+0}} = \frac{e^{i(\delta+\alpha)}}{e^{i(\delta-\alpha)}} = e^{2i\alpha} \Rightarrow A_{CP}^{h^+h^0} = 0. \quad (1.36)$$

In this case, the imaginary component of the ratio cannot be measured since no mixing-induced CP parameter is present in charged B decays. However, the observed CP violating asymmetries, $A_{CP}^{\pi^+\pi^0} = 0.03 \pm 0.04$ and $A_{CP}^{\rho^+\rho^0} = -0.05 \pm 0.05$, agree with zero [11]. Therefore, $A_{CP}^{h^+h^0}$ is not taken into account for the determination of α . Using Eq. (1.24) for the penguin-polluted decay $B^0 \rightarrow h^+h^-$, the CKM triangle relations, and the relation in Eq. (1.36), yields all ingredients to determine α through isospin symmetry. By convention, the A^{+-} amplitude is chosen to be real, yielding [50],

$$\begin{aligned} A^{+0} &= |A^{+0}|e^{i(\delta-\alpha)}, & \tilde{A}^{+0} &= |A^{+0}|e^{i(\delta+\alpha)}, \\ A^{+-} &= |A^{+-}|, & \tilde{A}^{+-} &= |\bar{A}^{+-}|e^{2i\alpha_{\text{eff}}}, \\ A^{00} &= A^{+0} - \frac{A^{+-}}{\sqrt{2}} = e^{i(\delta-\alpha)} \left(|A^{+0}| - \frac{|A^{+-}|}{\sqrt{2}} e^{-i(\delta-\alpha)} \right), \\ \tilde{A}^{00} &= \tilde{A}^{+0} - \frac{\tilde{A}^{+-}}{\sqrt{2}} = e^{i(\delta+\alpha)} \left(|A^{+0}| - \frac{|\bar{A}^{+-}|}{\sqrt{2}} e^{-i(\delta+\alpha-2\alpha_{\text{eff}})} \right), \end{aligned} \quad (1.37)$$

where α_{eff} equals $\alpha + \Delta\alpha$ with $\Delta\alpha$ being the shift due to penguin amplitudes. For the branching fractions \mathcal{B}^{ij} and the CP parameters A_{CP}^{ij} and S_{CP}^{ij} ,

$$\begin{aligned}
 \mathcal{B}^{+0} &= \tau_{B^+} |A^{+0}|^2, & \mathcal{B}^{+-} &= \tau_{B^0} \frac{|A^{+-}|^2 + |\bar{A}^{+-}|^2}{2}, \\
 \mathcal{B}^{00} &= \tau_{B^0} \left(|A^{+0}|^2 + \frac{|A^{+-}|^2 + |\bar{A}^{+-}|^2}{4} - \frac{|A^{+0}|}{\sqrt{2}} (|A^{+-}|c + |\bar{A}^{+-}|\bar{c}) \right), \\
 A_{CP}^{+-} &= \frac{|\bar{A}^{+-}|^2 - |A^{+-}|^2}{|\bar{A}^{+-}|^2 + |A^{+-}|^2}, & S_{CP}^{+-} &= \frac{2|\bar{A}^{+-}||A^{+-}|\sin(2\alpha_{\text{eff}})}{|\bar{A}^{+-}|^2 + |A^{+-}|^2}, \\
 A_{CP}^{00} &= \frac{|\bar{A}^{+-}|^2 - |A^{+-}|^2 - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{c} - |A^{+-}|c)}{4|A^{+0}|^2 + |\bar{A}^{+-}|^2 + |A^{+-}|^2 - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{c} + |A^{+-}|c)}, \tag{1.38}
 \end{aligned}$$

where $c = \cos(\alpha - \delta)$ and $\bar{c} = \cos(\alpha + \delta - 2\alpha_{\text{eff}})$. The lifetimes of charged and neutral B mesons, τ_{B^+} and τ_{B^0} , differ and are included in the branching fractions. If all branching fractions and CP -violation parameters in Eq. (1.38) are measured, the system has six linear independent equations with six real non-negative variables: $|A^{+0}|$, $|A^{+-}|$, $|\bar{A}^{+-}|$, δ , α_{eff} , and α . The value of α is determined up to an eight-fold ambiguity in the range $[0, \pi]$ because each isospin triangle has two possible orientations, leading to a four-fold trigonometric ambiguity

$$(\alpha, \delta) \leftrightarrow (\delta, \alpha), (2\alpha_{\text{eff}} - \alpha, 2\alpha_{\text{eff}} - \delta), (2\alpha_{\text{eff}} - \delta, 2\alpha_{\text{eff}} - \alpha). \tag{1.39}$$

An additional symmetry exists involving also α_{eff} that prescribes the functions c , \bar{c} , and S_{CP}^{+-} to be invariant under the reflection

$$(\alpha_{\text{eff}}, \alpha, \delta) \leftrightarrow \left(\frac{\pi}{2} - \alpha_{\text{eff}}, \frac{\pi}{2} - \alpha, \frac{\pi}{2} - \delta \right). \tag{1.40}$$

An illustration of the eight-fold α ambiguity is shown in Fig. 1.7.

1.7.2.2 The S_{CP}^{00} constraint

An important additional constraint that can partially lift the eight-fold ambiguity in the determination of α is the mixing-induced CP -violation asymmetry in the decay $B^0 \rightarrow h^0 h^0$, that is $B^0 \rightarrow \pi^0 \pi^0$ or $B^0 \rightarrow \rho^0 \rho^0$,

$$S_{CP}^{00} = \frac{4|A^{+0}|^2 \sin(2\alpha) + 2|A^{+-}||\bar{A}^{+-}| \sin(2\alpha_{\text{eff}}) - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{s} + |A^{+-}|s)}{4|A^{+0}|^2 + |\bar{A}^{+-}|^2 + |A^{+-}|^2 - 2\sqrt{2}|A^{+0}| (|\bar{A}^{+-}|\bar{c} + |A^{+-}|c)}, \tag{1.41}$$

where $s = \sin(\alpha + \delta)$ and $\bar{s} = \sin(\alpha - \delta + 2\alpha_{\text{eff}})$ are not invariant under the (α, δ) transformations in Eq. (1.39), thus fixing the orientation of each isospin triangle.

No measurement of S_{CP}^{00} has been reported for the decay $B^0 \rightarrow \pi^0 \pi^0$. This measurement is a formidable challenge because it requires to sample the time evolution of $B^0 \rightarrow \pi^0 \pi^0$ decays which in turn requires a precise reconstruction of the B^0 -decay vertex, something that cannot be achieved in the dominant four-photons final state.

For the decay $B^0 \rightarrow \rho^0 \rho^0$, a single measurement is available by the BaBar experiment with large uncertainties, $S_{CP}^{\rho^0 \rho^0} = 0.3 \pm 0.7 \pm 0.2$ [51]. However, the sides A^{00} and \tilde{A}^{00} of the $B \rightarrow \rho\rho$ isospin triangles are much smaller than the other sides, and the triangles are squashed into lines. The eight-fold ambiguity is therefore already reduced by a factor four.

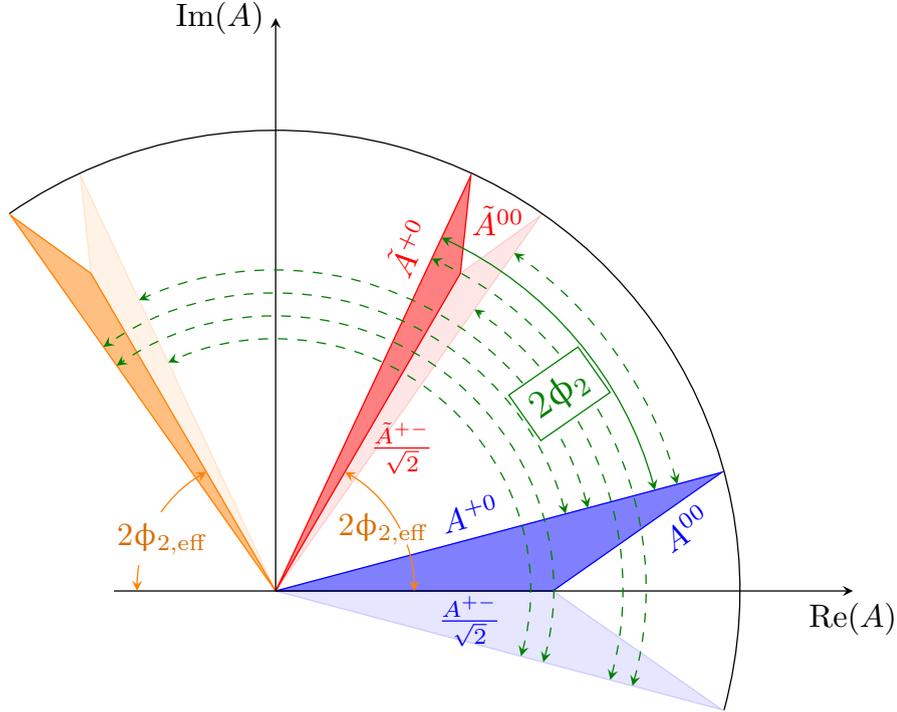


Figure 1.7: Geometrical representation of the isospin triangular relations (Eq. (1.35)) in the complex plane of $B^{i+j} \rightarrow h^i h^j$ amplitudes. Blue and red shaded areas correspond to the isospin triangles. The angle between the CP -conjugate amplitudes A^{+-} and \tilde{A}^{+-} corresponds to twice the weak phase α_{eff} (orange solid lines). The angle between the CP -conjugate amplitudes A^{+0} and \tilde{A}^{+0} corresponds to twice the CKM angle α (green solid line). Lighter-shaded triangles represent the mirror solutions allowed by the discrete ambiguities in Eq. (1.38), with the corresponding values for α represented by the green dashed lines.

1.8 $B^0 \rightarrow \pi^0 \pi^0$ experimental status and current constraints

Measurements of $B^0 \rightarrow \pi^0 \pi^0$ decay properties have been reported by previous experiments. The first observations of this decay were reported in 2003 by the Belle [52] and BaBar [53] experiments, which studied them since with increasingly larger samples. The most recent results, based on their full datasets, are summarized in Table 1.1. The Belle II experiment recently reported a first analysis of these decays using half of the currently available data [54]. The general picture is that significant uncertainties of 15 – 22%, dominated by statistical component, prevent from deriving stringent constraints on phenomenology as yet, thus offering important opportunities for experimental progress.

	Belle [55]	Belle II [54]	BaBar [56]	Average value [11]
$\mathcal{B}(B^0 \rightarrow \pi^0 \pi^0)$	$1.31 \pm 0.19 \pm 0.19$	$1.38 \pm 0.27 \pm 0.22$	$1.83 \pm 0.21 \pm 0.13$	1.59 ± 0.26
$A_{CP}(B^0 \rightarrow \pi^0 \pi^0)$	$0.14 \pm 0.36 \pm 0.10$	$-0.14 \pm 0.46 \pm 0.07$	$0.43 \pm 0.26 \pm 0.05$	0.33 ± 0.22
$B\bar{B}$ pairs	752×10^6	198×10^6	467×10^6	-

Table 1.1: Branching fraction ($\times 10^{-6}$) and A_{CP} of $B^0 \rightarrow \pi^0 \pi^0$ decays as measured at Belle, Belle II, and BaBar, and their global average values (only combining Belle and BaBar results) [11].

In addition, it is interesting to compare the constraints given by measurements of the $B \rightarrow \pi\pi$ system with the SM predictions, on both the isospin relations and the α parameter.

1.8.1 Current constraints on $B \rightarrow \pi\pi$ isospin

The consistency of the experimental $B \rightarrow \pi\pi$ data with isospin symmetry is assessed with the two-side sum

$$t = \frac{|a^{+-}|}{\sqrt{2}} + |a^{00}|, \quad (1.42)$$

where a^{ij} denotes the normalized amplitude $a^{ij} = A^{ij}/A^{+0}$. The sum of the two lengths, t , must be greater than one. The measured values, $t = 1.05 \pm 0.09$ and $t = 1.45 \pm 0.08$ [37], are consistent with an almost squashed and open triangle for $B \rightarrow \pi\pi$ and $\bar{B} \rightarrow \pi\pi$, respectively, and agree with expectations. This is shown in Fig. 1.8.

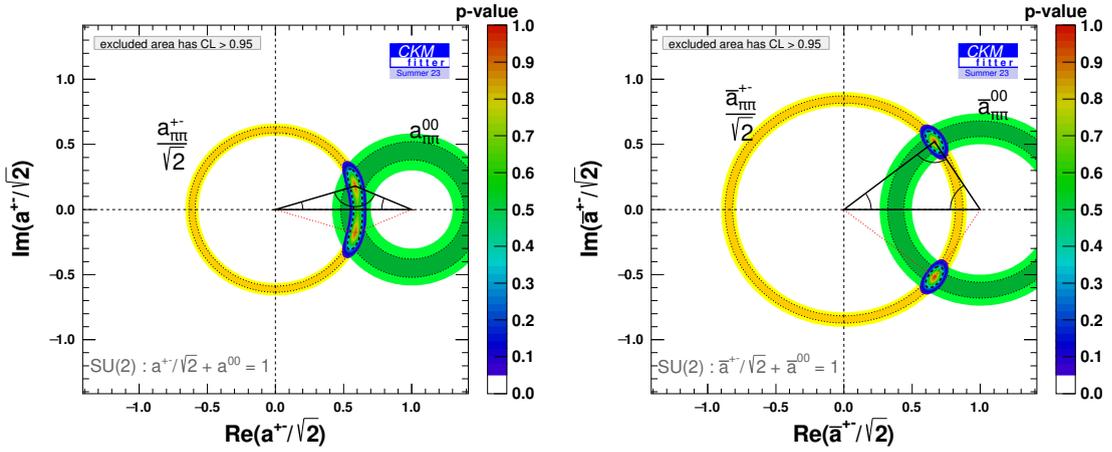


Figure 1.8: Constraints on the complex plane for the reduced isospin amplitude (left) $a^{+-} = A^{+-}/A^{+0}$ for B^0 mesons and (right) $\bar{a}^{+-} = \bar{A}^{+-}/\bar{A}^{+0}$ for \bar{B}^0 mesons in the $B \rightarrow \pi\pi$ system. Individual constraints from (yellow) $\pi^+\pi^-$ and (green) $\pi^0\pi^0$ observables are shown. The corresponding isospin triangular relation is represented by the black triangle. Reproduced from Ref. [37]. The latest Belle II results are not included [54, 57].

1.8.2 Constraints on α

Currently, the precision in the determination of $\alpha = (85.2^{+4.8}_{-4.3})^\circ$ is dominated by the $B \rightarrow \rho\rho$ system, and to a lesser extent, by the $B \rightarrow \pi\pi$ system [37].

The experimental extraction of the $B \rightarrow \pi\pi$ isospin parameters has been reported by BaBar, Belle and LHCb. Using the global averages of the measurements $\mathcal{S}_{CP}(\pi^+\pi^-) = -0.67 \pm 0.03$, $A_{CP}(\pi^+\pi^-) = -0.31 \pm 0.03$, the branching fractions of all three modes, and the direct CP -violating asymmetry $A_{CP}(\pi^0\pi^0) = 0.33 \pm 0.22$, one obtains the constraints on α shown in Fig. 1.9. Due to the large experimental uncertainties, some solutions are partially overlapping, leading to the following 68% C.L. constraints: $-13.5^\circ < \alpha < 15.7^\circ$, $74.3^\circ < \alpha < 105.6^\circ$, and $118.5^\circ < \alpha < 151.5^\circ$ [11]. The solution centered at 90° agrees with the indirect determination of $(91.9^{+1.4}_{-0.8})^\circ$ by a global fit of the flavor data performed by the CKMFitter group in Summer 2023 [37]. The largest source of uncertainty comes from the measurement of the properties of $B^0 \rightarrow \pi^0\pi^0$ decays, due to the low branching

ratio and the high background. Hence, a more precise determination of these properties, and in particular of A_{CP} , is of great importance to better constraint α .

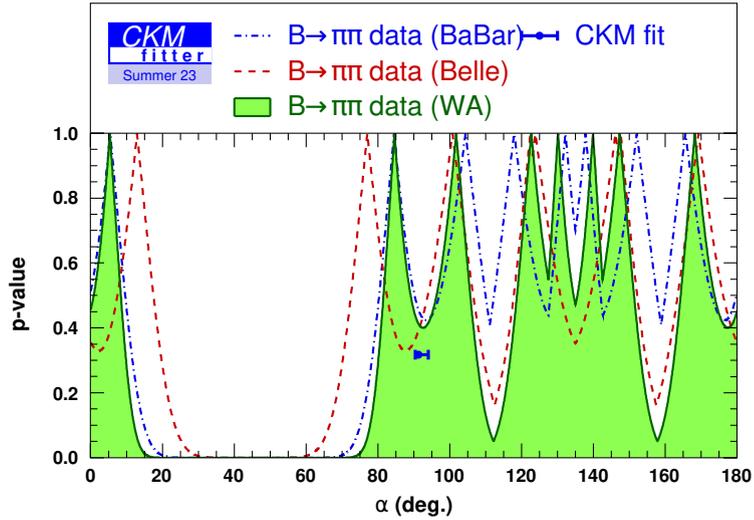


Figure 1.9: Constraints on α from $B \rightarrow \pi\pi$ isospin and the global CKM fit. Reproduced from Ref. [24]. The latest Belle II results [54, 57] are not included.

Chapter 2

The Belle II experiment at the SuperKEKB collider

The data used in this work are collected by the Belle II experiment. This chapter outlines the Belle II detector at the SuperKEKB accelerator, with emphasis on the subdetectors more relevant for the reconstruction of $B^0 \rightarrow \pi^0 \pi^0$ decays.

2.1 The SuperKEKB collider

SuperKEKB is an electron-positron (e^+e^-) energy-asymmetric collider, designed to produce more than 600 $B\bar{B}$ pairs per second ($B^0\bar{B}^0$ and B^+B^- in approximately equal proportions) via decays of $\Upsilon(4S)$ mesons produced at threshold [58]. Such colliders are called ‘ B -factories’, and were proposed in the 1990’s for the dedicated exploration of CP violation in B mesons. The main goal of B -factories is to produce low-background quantum-correlated $B\bar{B}$ pairs at high rates and with sufficient boost to study their time evolution.

Intense beams of electrons and positrons are brought to collision at the energy corresponding to the $\Upsilon(4S)$ meson mass, 10.58 GeV, which is just above the $B\bar{B}$ production kinematic threshold. The great majority of collisions yield electromagnetic processes ($e^+e^- \rightarrow e^+e^-$, $e^+e^- \rightarrow \gamma\gamma$, etc., see Fig. 2.1) that are scarcely interesting and straightforwardly discarded using global event quantities such as charged-particle multiplicity or total energy detected in the event. More interesting for flavor physics are the collisions that produce hadrons (henceforth called hadronic events). Figure 2.2 shows the hadron-production cross-section in e^+e^- collisions as a function of the final-state mass. The various peaks are radial excitations of the Υ meson overlapping the nearly uniform background at about 4 nb from so-called continuum of lighter-quark pair-production from the process $e^+e^- \rightarrow q\bar{q}$, where q identifies a u , d , c , or s quark. These are useful for charm physics, some selected topics in hadron physics, and as control channels. At the $\Upsilon(4S)$ collision energy, an about 70% fraction of the hadronic events involves the production of continuum. The rest are $\Upsilon(4S)$ events, which decay to $B\bar{B}$ pairs more than 96% of the time. At-threshold production implies little available energy to produce additional particles in the $B\bar{B}$ events, resulting in low-background conditions. These are the collisions relevant for the analysis described in this work as they produce $B^0 \rightarrow \pi^0 \pi^0$ signal, among many other processes. In addition, colliding beams of point-like particles imply precisely known collision energy, which sets stringent constraints on the collision’s kinematic properties, thus offering means of further background suppression. Since bottom mesons are produced in a strong-interaction decay, flavor is conserved, and the null net bottom content of the

initial state implies production of a flavorless $B\bar{B}$ pair. Even though B^0 and \bar{B}^0 undergo flavor oscillations before decaying, their time-evolution is quantum-correlated in such a way that no B^0B^0 or $\bar{B}^0\bar{B}^0$ pairs are present at any time. Angular-momentum conservation implies that the decay of the spin-1 $\Upsilon(4S)$ in the two spin-0 bottom mesons yields total angular momentum $J = 1$. Because the simultaneous presence of two identical particles in an antisymmetric state would violate Bose statistics, the system evolves coherently as an oscillating $B^0\bar{B}^0$ particle-antiparticle pair until either one decays. This allows efficient identification of the bottom (or antibottom) content of one meson at the time of decay of the other, if the latter decays in a final state accessible only by either bottom or antibottom states. This important capability is called ‘flavor tagging’ and allows measurements of flavor-dependent decay rates, as needed in many determinations of CP -violating quantities.

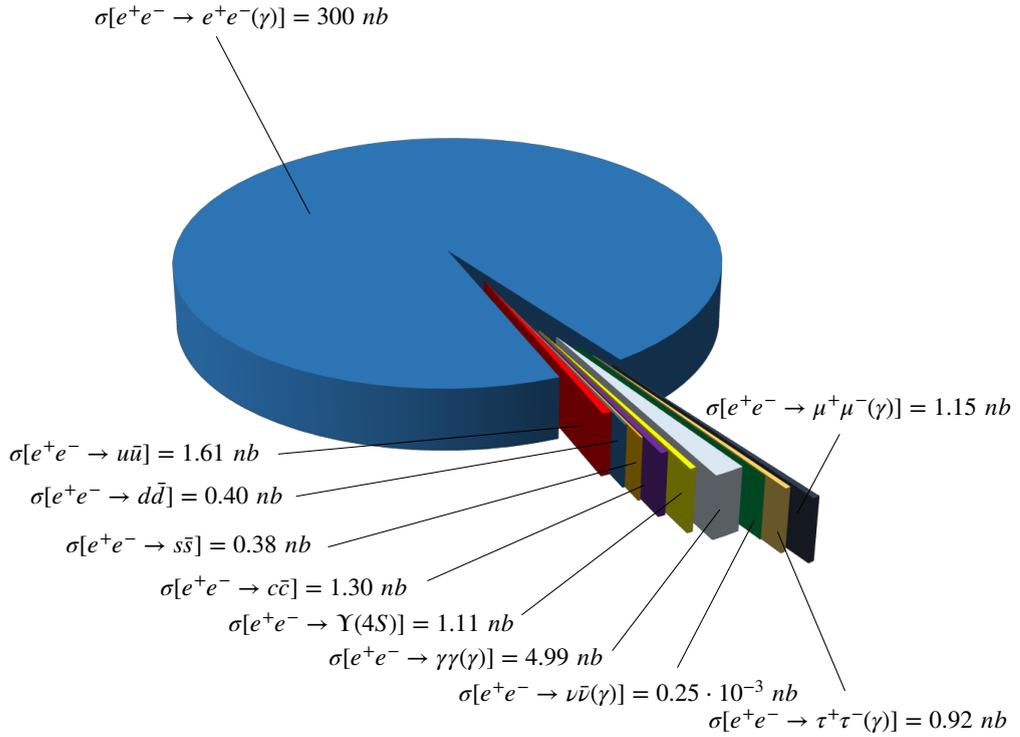


Figure 2.1: Cross sections of the main final states produced in e^+e^- collision at the $\Upsilon(4S)$ center-of-mass energy.

Because the $\Upsilon(4S)$ mesons are produced at threshold, they would be nearly at rest in the laboratory in an energy-symmetric collider. The resulting B mesons too would be produced with low momentum (about $10 \text{ MeV}/c$) in the laboratory, because of the $21 \text{ MeV}/c^2$ difference between the $\Upsilon(4S)$ mass and the $B\bar{B}$ pair mass. With such low momenta they would only travel approximately $1 \mu\text{m}$ before decaying, rendering the $10 \mu\text{m}$ typical spatial resolution of vertex detectors insufficient to separate B -decay vertices and study the decay-time evolution. Asymmetric beam energies are used to circumvent this limitation. By boosting the collision center-of-mass along the beam in the laboratory frame, B -decay vertex separations are achieved that are resolvable with current vertex detectors [59]. SuperKEKB (Fig. 2.3) implements a 7-on-4 GeV energy-asymmetric double-ring design, which achieves a vertex displacement of about $130 \mu\text{m}$.

Electrons are produced in a thermionic gun with a barium-impregnated tungsten cat-

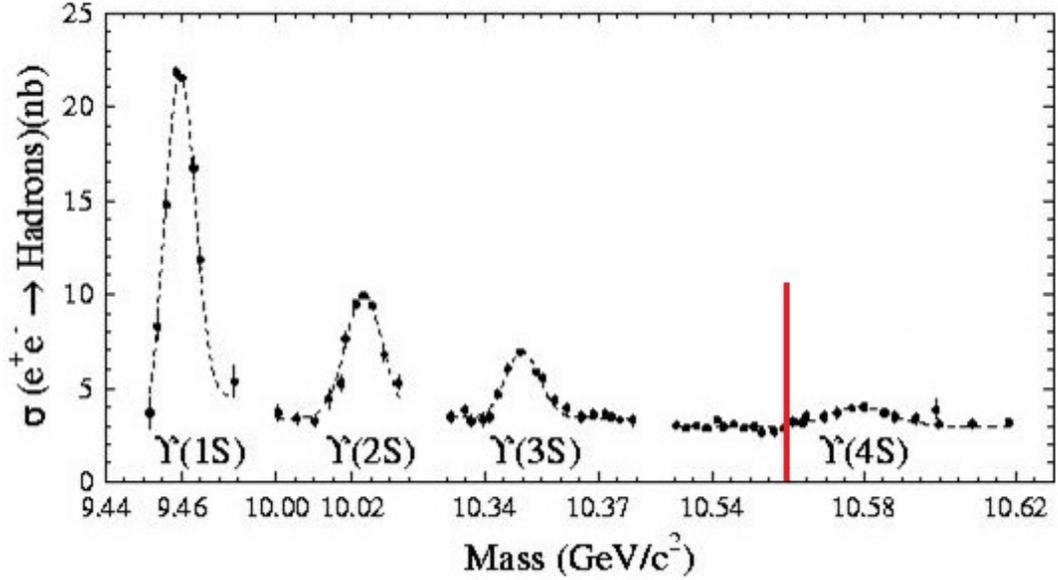


Figure 2.2: Hadron production cross section from e^+e^- collisions as a function of the final-state mass. The vertical red line indicates the $B\bar{B}$ production threshold.

hode, then accelerated to 7 GeV with a linear accelerator (linac) and injected in the high-energy ring (HER). Positrons are produced by colliding electrons on a tungsten target, then isolated by a magnetic field, accelerated to 4 GeV with the linac and injected in the low-energy ring (LER).

The electrons and positrons continuously collide at a single interaction point, surrounded by the Belle II detector. To achieve high luminosities, a nano-beam, large crossing-angle collision scheme is implemented [60]. This is an innovative configuration based on keeping small horizontal and vertical emittance, which is a measure of the spread and size of the particle beam in the phase space of position and momentum, and large crossing angle, as shown in Fig. 2.4. Such configuration is obtained with the production of low emittance beams, in addition to a sophisticated final-focus superconducting-quadrupole-magnet system, made of magnets, corrector coils, and compensation solenoids installed at each longitudinal end of the interaction region. Conceptually the nano-beam scheme mimics a collision with many short micro-bunches, allowing significant advantages in luminosity with respect to previous conventional schemes. The reduction of the luminous volume size to about 5% with respect to the predecessor KEKB, combined with doubled beam currents, is expected to yield a factor 40 gain in intensity. The penalty for such high intensities are significant challenges in achieving the design performance and operating steadily, and higher beam-induced backgrounds.

The performance of the SuperKEKB collider is mainly characterized in terms of the instantaneous luminosity \mathcal{L} , which is a measure of collision intensity,

$$\mathcal{L} = \frac{\gamma_{\pm}}{2er_e} \left(1 + \frac{\sigma_y^*}{\sigma_x^*} \right) \frac{I_{\pm}\xi_{y\pm}}{\beta_{y\pm}^*} \cdot \frac{R_{\mathcal{L}}}{R_{\xi_y}},$$

where γ is the relativistic Lorentz factor, e is the absolute value of the electron charge, r_e is the classical radius of the electron, σ_x^* and σ_y^* are the bunch widths at the interaction point (IP) in the plane orthogonal to the beam direction (transverse plane), I is the current of the beam, β_y^* is the vertical betatron function at the IP [61], ξ_y is the vertical beam-beam parameter, $R_{\mathcal{L}}$ and R_{ξ_y} are the luminosity reduction factors and the vertical beam-

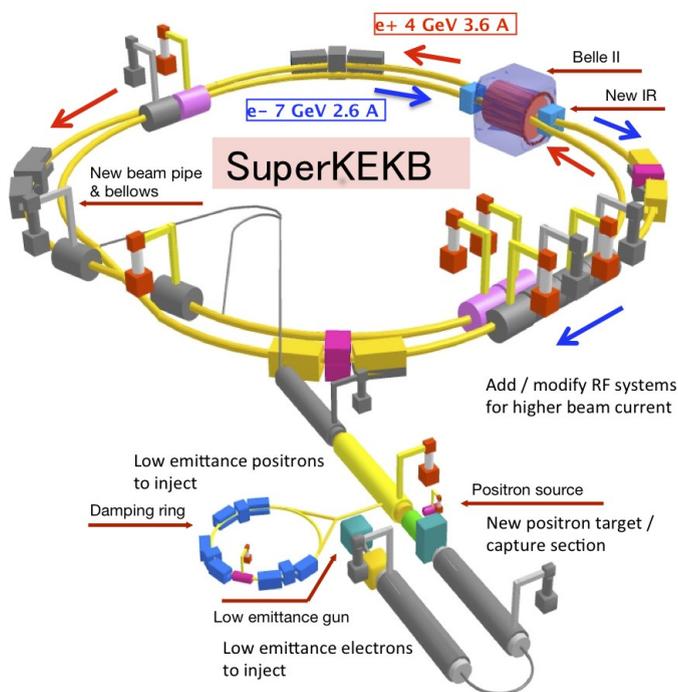


Figure 2.3: Illustration of the SuperKEKB collider.

beam parameter, respectively, due to non-vanishing crossing angle [62]. The ratio of these reduction factors is close to unity, while the design values for the other parameters are reported in Table 2.1.

The rate of any given process

$$\text{rate} [\text{events s}^{-1}] = \mathcal{L} [\text{cm}^{-2} \text{s}^{-1}] \times \sigma [\text{cm}^2],$$

is the product of its cross-section and \mathcal{L} .

	Design	Achieved (as of mid 2022)
Energy [GeV]	4.0/7.0	4.0/7.0
ξ_y	0.090/0.088	0.0407/0.0279
β_y^* [mm]	0.27/0.41	1.0/1.0
I [A]	3.6/2.62	1.321/1.099

Table 2.1: Design and achieved values for SuperKEKB fundamental parameters (LER/HER).

The integral of instantaneous luminosity over time T , called integrated luminosity,

$$\mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t') dt'$$

is a direct measure of the number of produced events of interest $N = \mathcal{L}_{\text{int}} \sigma$.

Physics data-taking started in March 2019, and Belle II has integrated $424 \pm 3 \text{ fb}^{-1}$ of luminosity at the time of this writing. In 2022, SuperKEKB achieved the instantaneous-luminosity world record, $4.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. In spite of these achievements, a number of technological and scientific challenges have significantly reduced SuperKEKB performance

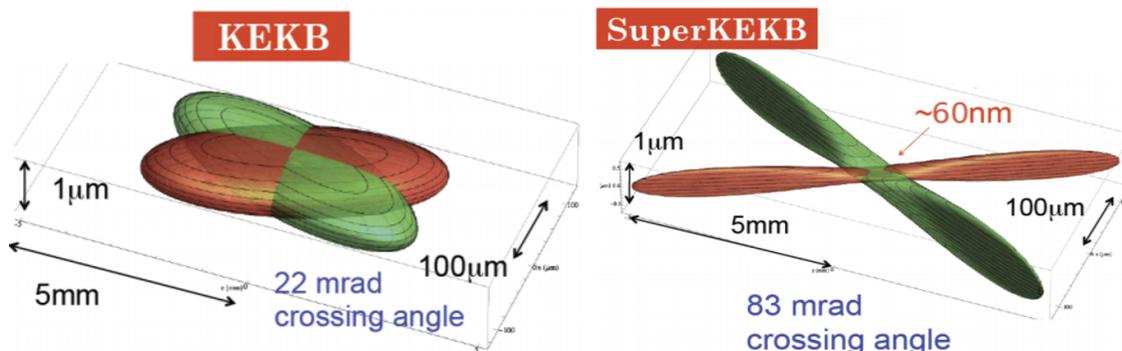


Figure 2.4: Two-dimensional sketch of (right) the nano-beam mechanism implemented in SuperKEKB compared with (left) the previous KEKB collision scheme.

compared to design. A number of issues associated with beam injection, collimation, and short beams lifetime due to the reduction of their dynamic aperture, which also causes high uncontrollable beam backgrounds, limited the capability to deliver the expected samples of data in its first four years. Consolidation, improvement and development work is ongoing to overcome these difficulties.

2.2 The Belle II detector

Belle II (Fig. 2.5) is a large-solid-angle, multipurpose magnetic spectrometer surrounded by a calorimeter and particle-identification systems, installed around the SuperKEKB interaction point. It is designed to determine energy, momentum, and identity of a broad range of particles produced in 10.58 GeV e^+e^- collisions. Belle II is approximately a cylinder of about 7 m in length and 7 m in diameter. It employs a right-handed Cartesian coordinate system with origin in the interaction point. The z axis corresponds to the principal axis of the solenoid, which is approximately parallel to the electron beam direction at the interaction point; the y axis points vertically upward, and the x axis is horizontal and pointing outward of the accelerator tunnel. The polar angle, θ , is referred to the positive z axis. The azimuthal angle, ϕ , is referred to the positive x axis in the xy plane. The radius, $r = \sqrt{x^2 + y^2}$, is defined in cylindrical coordinates and measured from the origin in the xy plane. Throughout this thesis, *longitudinal* means parallel to the electron beam direction (to the z axis), and *transverse* means perpendicular to the electron beam direction, i.e., in the xy plane.

Belle II comprises several subsystems, each dedicated to a specific aspect of event reconstruction. From the interaction point outward, a particle would traverse the beam pipe, a two-layer silicon-pixel vertex-detector (PXD), a four-layer silicon-strip vertex-detector (SVD), a central wire drift-chamber (CDC), a time-of-propagation central Cherenkov counter (TOP) or an aerogel threshold forward Cherenkov counter (ARICH), an array of CsI(Tl) crystals (ECL), a superconducting solenoidal magnet, and multiple layers of resistive plate counters (KLM).

The principal experimental strengths are hermetic coverage, which allows for reconstruction of final states involving neutrinos; efficient and precise reconstruction of charged-particle trajectories (tracks), which provide accurately reconstructed decay-vertices and good momentum resolution; high-purity charged-particle identification and neutral-particle reconstruction. A summary of the technological specifications of the Belle II subsystems is in Table 2.2. A detailed description of Belle II and its performance is given in Ref. [63].

In the following, I focus on the electromagnetic calorimeter, which is the subdetector more relevant for the analysis reported in this thesis.

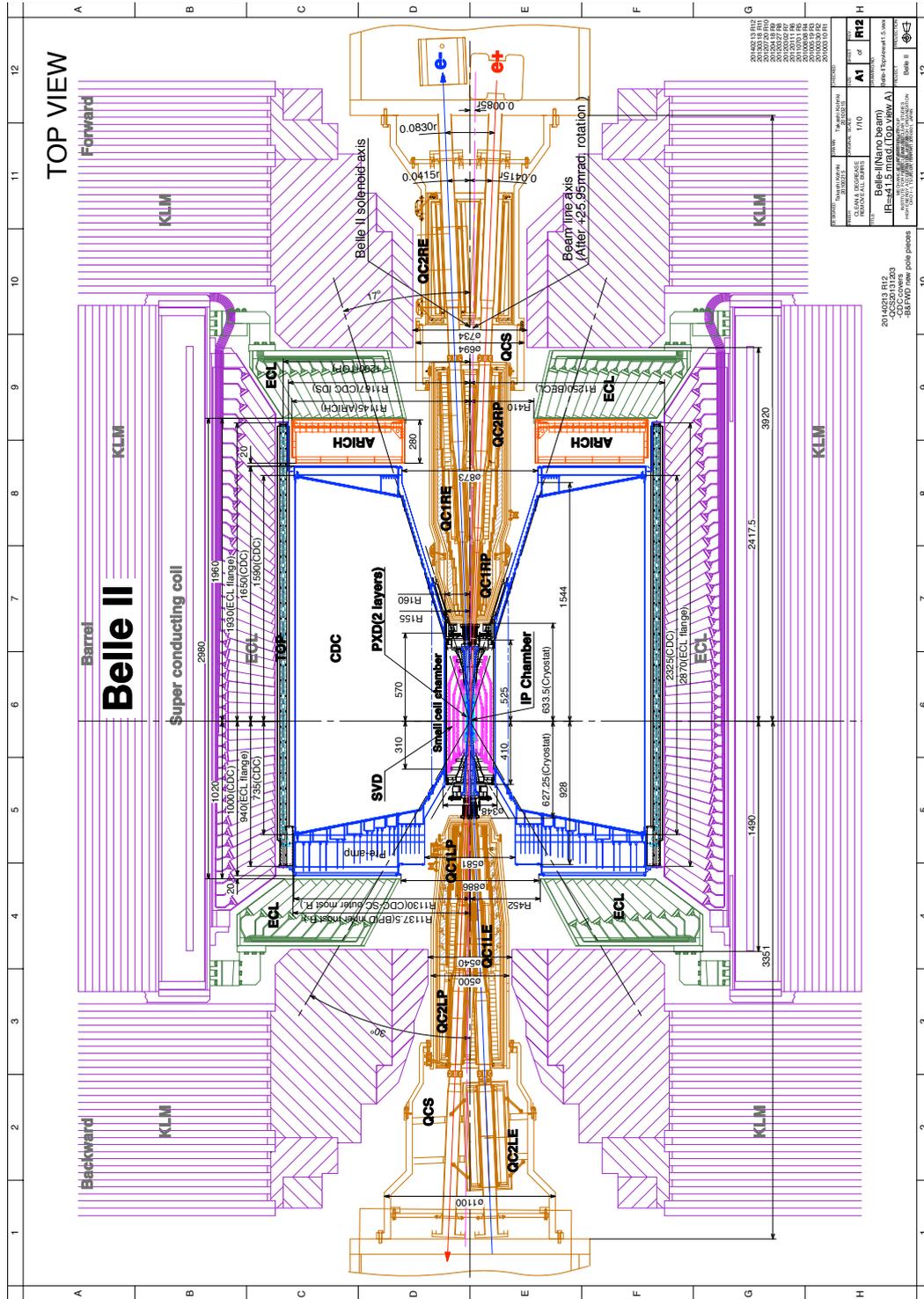


Figure 2.5: Top view of Belle II, the beam pipe at IP and final-focus magnets.

Purpose	Acronym	Technology	Configuration	Channels	Polar coverage (θ)
Beam pipe		Beryllium	Cylindrical, inner radius 10 mm, 10 μm Au, 0.6 mm Be, 1 mm paraffin, 0.4 mm Be		
Tracking	PXD	Silicon pixel	Sensor size: $15 \times (L1 \text{ } 136, L2 \text{ } 170)$ mm ² , Pixel size: $50 \times (L1a \text{ } 50, L1b \text{ } 60, L2a \text{ } 75, L2b \text{ } 85)$ μm^2 ; two layers at radii: 14, 22 mm	10^6	[17°;150°]
	SVD	Silicon strip	Rectangular and trapezoidal, strip pitch: $50(p)/160(n) - 75(p)/240(n)$ μm , with one floating intermediate strip; four layers at radii: 38, 80, 115, 140 mm	2.45×10^5	[17°;150°]
	CDC	Drift chamber with He-C ₂ H ₆ gas	14336 wires in 56 layers, inner radius of 160mm outer radius of 1130 mm	1.4×10^5	[17°;150°]
	Solenoid	Solenoid coil	Length: 4.41 m. Radius: 1.80 m		
Particle ID	TOP	RICH with quartz radiator	16 segments in ϕ at $r \approx 120$ cm, 275 cm long, 2cm thick quartz bars with 4×4 channel MCP PMTs	8×10^3	[31°;128°]
	ARICH	RICH with aerogel radiator	2×2 cm thick focusing radiators with different n , HAPD photodetectors	7.8×10^4	[14°;30°]
Calorimetry	ECL	CsI(Tl) crystals	Barrel: $r = 125 - 162\text{cm}$, end-cap: $z = -102 - +196\text{cm}$	6624 (Barrel), 1152 (FWD), 960 (BWD)	[12.4°;31.4°], [32.2°;128.7°], [130.7°;155.1°]
Muon ID	KLM	Barrel: RPCs and scintillator strips	2 layers with scintillator strips and 12 layers with 2 RPCs	$\theta \text{ } 1.6 \times 10^4$, $\phi \text{ } 1.6 \times 10^4$	[40°;129°]
	KLM	End-cap: scintillator strips	12 layers of (7-10) \times 40 mm ² strips	1.7×10^4	[25°;40°], [129°;155°]

Table 2.2: Summary of the Belle II subdetectors and their specifications.

2.2.1 Tracking system

At Belle II, reconstruction of charged particles and ensuing measurement of their momenta and charges is achieved through an integrated system consisting of six layers of silicon and a drift chamber, surrounding the beam pipe and immersed in a 1.5 T axial magnetic field maintained in a cylindrical volume 3.4 m in diameter and 4.4 m in length. The field is oriented along the z direction and provided by an aluminum-stabilized superconducting solenoid made of NbTi/Cu alloy. The solenoid surrounds all the subdetectors up to the KLM. The iron yoke of the detector serves as the return path of the magnetic flux. The beam pipe is a 3 km-long vacuum enclosure to allow beams circulating inside the detector. In the following, I refer only to the straight section of the beam pipe surrounding the interaction point. Multiple Coulomb scattering in the beam-pipe wall of the final-state charged particles would spoil the vertex-position resolution; this dictates a thin beam-pipe wall made of a low- Z material. Moreover, since the vertex resolution is inversely proportional to the distance between the interaction point and the first track sampling, the beam pipe has to be narrow. The possibility for beam-halo to interact with the beam pipe, thus inducing beam backgrounds, and heating of the pipe wall due to charge induction complicates the design. Hence, the beam pipe is constantly cooled and shielded from the vertex detector. The Belle II beam pipe is made of two beryllium cylinders, 0.6 mm thick at radius of 10 mm, and 0.4 mm thick at radius of 12 mm, respectively. A 1.0 mm gap between the inner and outer walls of the pipe is filled with paraffin for cooling. The beam pipe is coated with a 10 μm gold sheet that absorbs low-energy photons, which could damage the silicon detector.

2.2.1.1 Silicon-pixel vertexing detector

The innermost detector is a pixel vertex detector (PXD) [64]. Its goal is to sample the trajectories of final-state charged particles in the vicinity of the decay position (vertex) of their long lived ancestors, so that the decay point can be inferred by extrapolation inward.

PXD sensors are based on the technology of depleted field-effect transistors [64]. They are made of p-channel MOSFET integrated on a silicon substrate, which is fully depleted by applying an appropriate voltage. Incident particles generate electron-hole pairs in the depleted region. The charge carriers drift towards the minimum of potential placed under the transistor channel, and thus modulate a current passing through the MOSFET. Sensors are 75 μm thick.

The PXD has two layers at 14 mm and 22 mm radii, respectively, and a full length of 174 mm at the radius of the outer layer. It comprises around 8 million pixels, $50 \times (50 - 55)\mu\text{m}^2$ (inner layer) and $50 \times (70 - 85)\mu\text{m}^2$ (outer layer) each. The polar acceptance ranges from 17° to 150° . The design impact-parameter resolution is 12 μm , achieved by weighting the charge deposited in neighboring pixels. For the data used in this thesis, the full first pixel layer is used, along with a 1/6 azimuthal sector of the second layer, as completion of the pixel detector only happened in 2023.

2.2.1.2 Silicon-strip vertexing detector

Around the PXD is SVD, a silicon detector aimed at reconstructing decay vertices and low-momentum charged-particles at high resolution [65].

It uses double-sided silicon strip sensors. Each sensor is made of a silicon n-doped bulk on one side, and a perpendicular highly p-doped implant on the other side. This means that for each sensor, one side has strips parallel to the beams direction, and the other

perpendicular. A voltage is applied to enhance the depletion region at the p-n junction, and removes intrinsic charge-carriers from the region. Traversing charged particles ionize the silicon, freeing electron-hole pairs that drift due to the electric field thus inducing a signal in highly granular strip electrodes implanted at both ends of the depletion region. The fine segmentation and fast charge collection of SVD sensors make possible to deal with large track density environments.

The SVD structure consists of four concentric layers at radii of 39, 80, 104 and 135 mm, composed by, respectively, 7, 10, 12, and 16 independently-readout longitudinal modules called ladders, arranged in a cylindrical geometry. As shown in Fig. 2.6, SVD has a polar-asymmetric geometry that mirrors the asymmetry in particle density resulting from the center-of-mass boost. The polar acceptance ranges from 17° to 150° .

Sensors are $300\ \mu\text{m}$ thick, and the separation between adjacent strips (d_{pitch}) ranges from $50\ \mu\text{m}$ to $240\ \mu\text{m}$. Hence, the nominal spatial resolution $d_{\text{pitch}}/\sqrt{12}$ varies with the polar angle. Since the charge associated with an incident particle is usually distributed among several strips, position resolution is improved by interpolation.

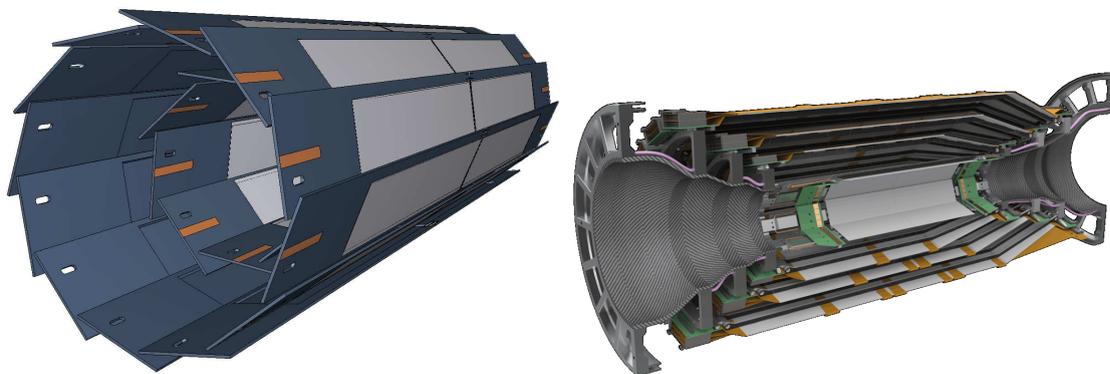


Figure 2.6: (Left) sketch of the PXD detector and (right) exploded view of a SVD detector half.

2.2.1.3 Central drift chamber

The CDC is a drift chamber [66]. It samples charged-particle trajectories at radii between 16 cm and 113 cm, thus providing accurate measurements of momentum and electric charge, trigger information for events containing charged particles, and information on identification of charged-particle species by measuring their specific-ionization energy-loss (dE/dx). Between SVD and CDC, a small section with 2.5 cm radius is present.

When a charged particle traverses the CDC volume, it ionizes the gas, freeing electrons and positive ions from gas atoms. A stationary electric field then accelerates these charges until they approach the sense wires. In their vicinity high field gradients cause an abrupt acceleration that causes secondary ionizations, which induce an electric signal whose time is digitized. The particle trajectory is inferred from the time between the collision and the signal.

The chamber volume contains 14336 $30\text{-}\mu\text{m}$ -diameter sense wires, divided in 56 layers, immersed in a gaseous mixture of 50% He and 50% C_2H_6 , while 42240 $126\text{-}\mu\text{m}$ -diameter aluminum wires shape the electric field. Layers of wires are installed with either "axial" orientation, i.e., aligned with the solenoidal magnetic field, or skewed with respect to the axial wires with a "stereo" orientation. The azimuthal acceptance ranges from 17° to 150° .

The spatial resolution is about $100\ \mu\text{m}$ and the dE/dx resolution is 11.9% for an incident angle of 90° . Figure 2.7 shows a sliced view of the CDC and the wire configurations.

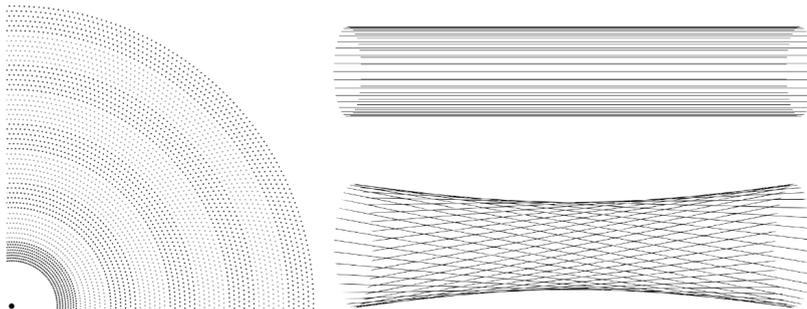


Figure 2.7: On the left, a transverse quadrant of the drift chamber, organized in layers, called *superlayers*; on the right, wire orientation for (top) axial and (bottom) stereo layers. Skew is exaggerated for visualization purposes.

2.2.2 Electromagnetic calorimeter

The ECL measures the energy of photons and electrons [67]. High energy photons and electrons entering the calorimeter initiate an electromagnetic shower through bremsstrahlung and electron-positron pair production. The energy is mostly converted to photons, which are collected by the photodiodes. In contrast to hadrons, which pass through the calorimeter with minimal energy loss, most photons and electrons dissipate their entire energy.

The configuration, mechanical structure, and crystals of Belle II ECL are those of the Belle’s calorimeter. The readout electronic boards have been upgraded to cope with SuperKEKB’s higher luminosity. The layout is shown in Fig. 2.8. The ECL consists of three polar compartments: the barrel, the forward endcap, and the backward endcap section. The barrel section is 3.0 m long with 1.25 m of inner radius; the endcaps are located at $z = +2.0$ m (forward) and -1.0 m (backward) from the interaction point. Table 2.3 summarizes the geometrical parameters of each section.

Item	θ coverage	θ segmentation	ϕ segmentation	Number of crystals
Forward endcap	12.4° – 31.4°	13	48–144	1152
Barrel	32.2° – 128.7°	46	144	6624
Backward endcap	130.7° – 155.1°	10	64–144	960

Table 2.3: Summary of ECL parameters.

High momentum π^0 detection is essential in this work and requires good separation of two nearby photons and a precise determination of the opening angle. This requires a segmented calorimeter. The ECL is a highly segmented array of 8736 cesium iodide crystals doped with thallium (CsI(Tl)). Thallium shifts the energy of the excitation light into the visible spectrum. The light is detected by a independent pair of silicon PIN photodiodes [67] and charge-sensitive preamplifiers installed at the outer end of each crystal.

A typical crystal in the barrel section has a $55 \times 55\ \text{mm}^2$ active surface on the front face and $65 \times 65\ \text{mm}^2$ on the rear face; the dimensions of the crystals in the endcap sections

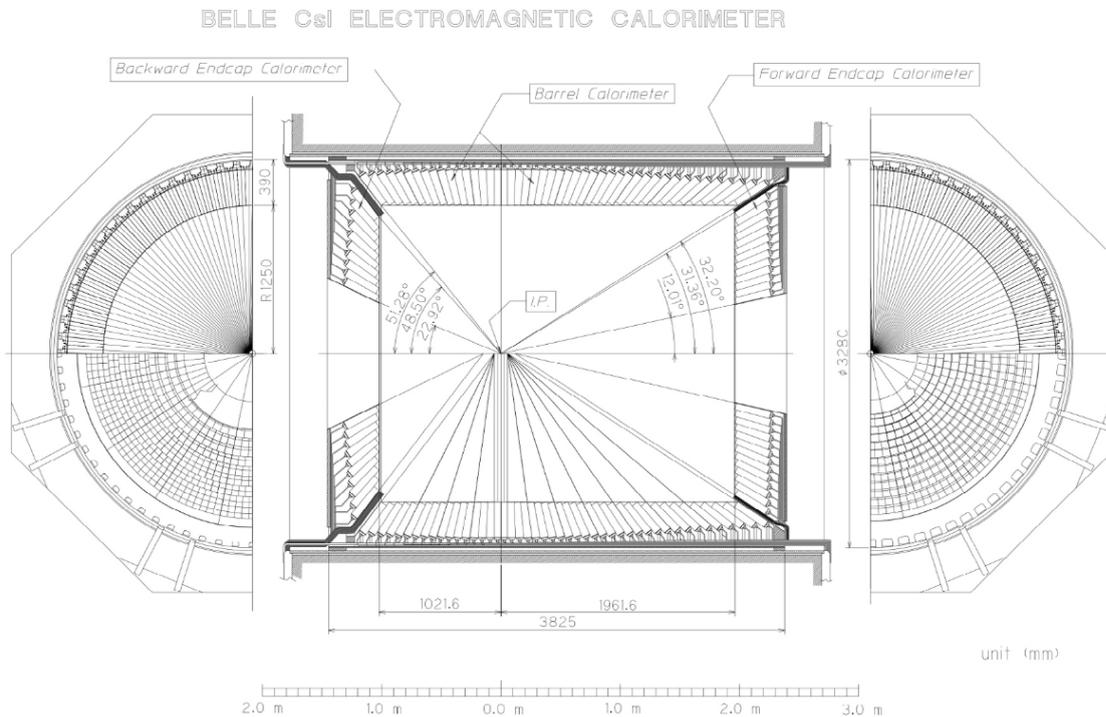


Figure 2.8: ECL layout.

vary from 44.5 to 70.8 mm and from 54 to 82 mm for front and rear faces, respectively. A diagram of an ECL crystal is shown in Fig. 2.9. The 30-cm crystal length, corresponding to $16.1X_0$, reduces the fluctuations of shower leakages out of the outermost end of the crystals, which spoils energy resolution. The crystals are designed in such a way that a photon injected at the center of the crystal would deposit 80% of its energy in the crystal on average. The crystals principal axes do not point exactly to the nominal interaction point, but they are inclined to prevent photons from escaping through gaps between crystals by about 1.3° in the θ and ϕ directions in the barrel section, and by about 1.5° and about 4° in the θ direction in the forward and backward sections.

Considering the ECL structure – gaps, crystal wrapping, mechanical structure – the fraction of photons that do not leave a detectable signal in the calorimeter is only 0.2%.

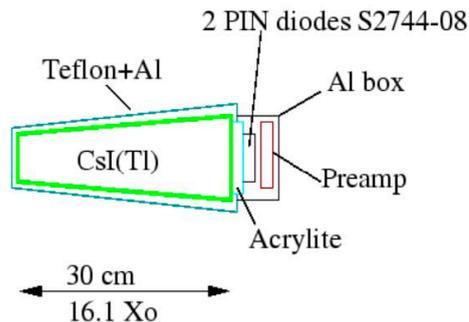


Figure 2.9: Schematic design of a CsI(Tl) crystal with attached readout electronic circuits.

Signals from the photodiodes are sent to two preamplifiers mounted on the rear of the crystal for charge integration. The two resulting signals are sent to a readout board [67] located outside the Belle II detector and containing 16 analog shaper circuits. These channels receive signals from up to 16 CsI(Tl) crystals. A total of 576 shaper modules are needed to process signals from all 8736 crystals. Shaper circuits amplify and shape the analog signal to provide a clean and well-defined pulse while removing noise, pile-up due to overlapping pulses from neighbouring events, and other unwanted features from the signal. The signal is sampled by a digitizer at 1.76 MHz, which corresponds to an interval between measurements of 567 ns. After collecting 31 samples, the signal waveform is processed using a photon template fit to compute the signal amplitude of the signal, the time relative to the trigger signal, and the χ^2 fit quality. The first 16 samples (pedestal) contain information about the baseline value, while the remaining 15 samples, from the 16th to the 31st sample, contain the signal waveform. In data, the pedestal is not uniform, but shows fluctuations due to electronic noise and backgrounds not coming from collisions. An example of a fit is shown in Fig. 2.10.

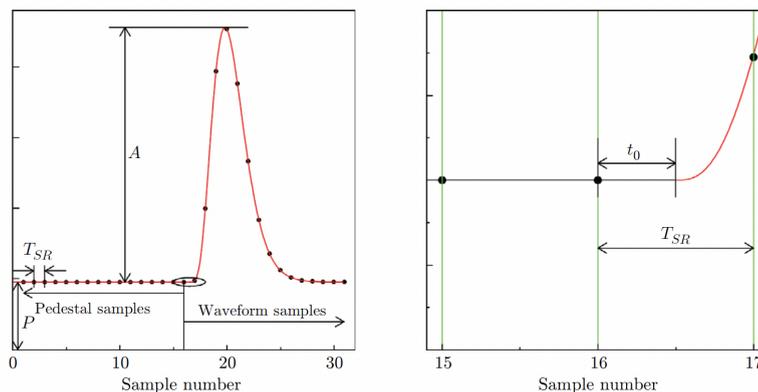


Figure 2.10: On the left, fit to the 31 samples containing the pedestal and the signal waveform, and on the right, focus on the 16th sample, where the signal starts, in simulation.

The signal waveforms are analyzed using pulse-shape discrimination to improve particle identification, since the signal decay time in CsI(Tl) includes a fast component of around $0.6 \mu\text{s}$ and a slow component at around $3.5 \mu\text{s}$. The fast component is associated with the scintillation response to electromagnetic interactions, while the slow component is associated with the scintillation response to hadronic (e.g., proton or neutron) interactions. This occurs because scintillation in CsI(Tl) for electromagnetic interactions involves the excitation and deexcitation of Tl atoms, while scintillation for hadronic interactions involves the excitation and deexcitation of both Tl and Cs atoms, leading to a longer decay time. The ratio between the intensity of these two decay components varies as a function of the ionizing power of the absorbed particle. When analyzing the shape of the scintillation pulse, the shape discrimination can help to distinguish between these different types of particles. This feature was not yet available in the data used for my analysis. The photon emission spectrum peaks at around 550 nm, which is convenient for photodiode readout. However, the time for the light in the crystals to decay is relatively long, increasing considerably the

overlap of pulses from neighboring (background) events. This means that scintillation light may be present when a particle from a later event arrives, generating pile-up background.

The ECL also uses Bhabha scattering to measure luminosity. Because the Bhabha cross section is predicted with high accuracy in QED, a precise inference of luminosity is achieved from the measured rate of Bhabha events in a volume of known acceptance.

2.2.3 Particle identification

Belle II combines measurements of time-of-propagation, Cherenkov radiation, and ionization energy loss in the tracker and drift chamber to identify charged particles.

2.2.3.1 Time-of-propagation detector

The TOP detector measures the time of propagation of the Cherenkov photons emitted from charged particles passing through its quartz bars and internally reflected within a radiator [68]. It is made of 16 quartz bars mounted at 1.2 m from the IP. Each bar has three main components (Fig. 2.11): a long bar acts as Cherenkov radiator, where photons are generated and propagated; a focusing mirror is mounted at the forward end; and a prism mounted at the backward end collects photons and guides them to a photomultiplier. The polar coverage ranges from 31° to 128° . On average, photons originated from slower particles take more time to reach the photomultipliers, because of the inverse proportionality between β and $\cos\theta_C$.

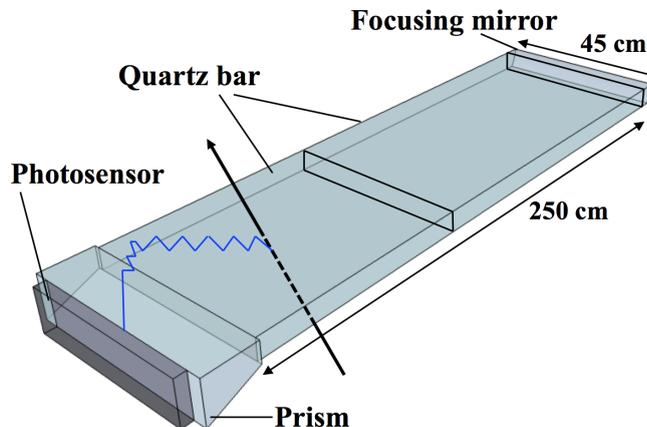


Figure 2.11: Scheme of a TOP bar. A charged particle crossing the radiator and emitting Cherenkov photons, which are collected at the photomultipliers, is also represented.

2.2.3.2 Aerogel ring-imaging Cherenkov counter

The ARICH detector identifies charged particles by measuring the Cherenkov ring produced when passing through a radiator [69]. It consists of 420 modules for photon detection in seven layers extending from 0.56 to 1.14 m radius, and 248 aerogel tiles installed on the detector endcaps. The aerogel radiator produces Cherenkov photons when traversed by charged particles of a certain momentum range. Next to the radiator is an expansion volume where photons are propagated, to form rings on position-sensitive photodiodes. Photocathodes then convert photons into photoelectrons and generate electric signals. Two adjacent radiators with different refraction indexes generate enough photons for achieving sufficient resolution, as shown in Fig. 2.12.

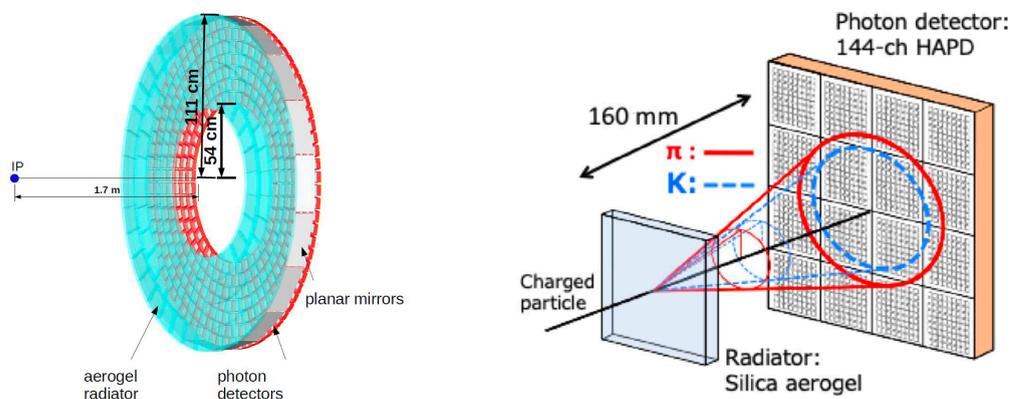


Figure 2.12: Sketch of (left) the ARICH with its main components and (right) diagram of the difference in the photon path for Cherenkov photons from kaons and pions.

2.2.3.3 K_L^0 and muon detection system

The KLM detects muons and neutral particles that do not get absorbed in the inner detectors, such as K_L^0 mesons [70]. It is made of alternating 4.7-cm-thick iron plates and active detector elements. Iron elements act also as magnetic flux returns for the tracking solenoid. In the inner layers, the active material is scintillator, in the outer layers are glass-electrode resistive-plates chambers, with a gas mixture filling the space between electrodes. When particles traverse the KLM, they produce charges that are collected by applying an appropriate voltage. The barrel section of the detector covers 45° to 125° in polar angle. The endcaps cover 20° to 45° and 125° to 155° .

2.2.4 Trigger and data acquisition system

The e^+e^- collisions at the $\Upsilon(4S)$ resonance produce a variety of processes. As the events of interest are only a fraction of the total cross section and it would be impossible to record all collisions on permanent memory, an online event-selection system (trigger) is used to distinguish them from background in real time, and to feed only the interesting events to the data acquisition system (DAQ), compatibly with data processing resources. The physics processes of interest include hadronic, μ/τ -pair, Bhabha, and two photon events. Accept rates of Bhabha and $\gamma\gamma$ events, which have high cross section and can be identified by their distinct signature, are artificially reduced by a factor of 100 to comply with the data acquisition limitations. Preferably discarded events include beam-related background resulting from synchrotron radiation, scattering of the beams on the residual gas, interactions in the beam pipe, and cosmic-ray events.

The Belle II trigger is organized according to a two-level logic, with a level 1 (L1) hardware trigger followed by a software-based, high-level trigger (HLT).

The L1 trigger, designed for a maximum rate of 30 kHz, uses input from four sub-detectors: (i) the CDC, that provides three-dimensional track information to suppress tracks not originating from the interaction point; (ii) the ECL, that gives information on total energy deposit and cluster multiplicity; (iii) the TOP, that provides timing and hit topology information; and (iv) the KLM, that gives high-efficiency trigger for muons. These are used to achieve a low-level reconstruction that is fed to the global decision logic, which sends the proper trigger signal if the event passes the selection requirements. The L1 logic is implemented using field-programmable gate arrays that have a fixed latency of 5 μ s, with

an uncertainty on the trigger timing (jitter) of approximately 10 ns.

Expected cross sections and trigger accept rates for physics processes of interest at the design instantaneous luminosity of $8 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$ are given in Table 2.4.

Process	σ [nb]	Rate [Hz]
$e^+e^- \rightarrow \Upsilon(4S)$	1.2	960
$e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$)	2.8	2200
$e^+e^- \rightarrow \mu^+\mu^-$	0.8	640
$e^+e^- \rightarrow \tau^+\tau^-$	0.8	640
$e^+e^- \rightarrow e^+e^-$ (Bhabha scattering) $\theta_{lab} > 17^\circ$	44	350*
$e^+e^- \rightarrow \gamma\gamma$ $\theta_{lab} > 17^\circ$	2.4	19*
Two photon events ($\theta_{lab} > 17^\circ$ and $p_T \geq 0.1 \text{ GeV}/c$)	≈ 80	≈ 1500

Table 2.4: Expected cross sections and trigger rates of various physics processes at $8 \times 10^{35} \text{cm}^{-2} \text{s}^{-1}$ luminosity [63]. Bhabha and $\gamma\gamma$ accept rates (*) are artificially reduced by a factor of 100 to comply with the data acquisition limitations.

Events selected by the L1 trigger are input to the HLT, that makes a decision using information from all the subdetectors except for PXD. The online software reconstruction is similar to that used offline. A first selection, performed after the first step of the reconstruction and aimed at discarding about half of the events, is based on requirements on track multiplicity, vertex position, and total ECL energy deposit. After the remaining steps of the standard reconstruction are completed, further physics-level selection are performed. After this stage, the number of events is reduced to about 1/5 of those passing the L1 trigger. The efficiency of the HLT for $\Upsilon(4S) \rightarrow B\bar{B}$ events is higher than 99%.

Data from the PXD for events that pass the L1 selection are stored in a dedicated online data reduction system. Once an event passes the selection, HLT extrapolates the tracks found by CDC and SVD to the PXD layers, defining regions of interest (ROIs). These are passed to the data reduction system, and only hits matching with a ROI are transmitted to the DAQ system. This keeps the PXD data size to about 100 kB/event.

Fully reconstructed events are stored in DST files. The size of a DST of a typical hadronic event is 100 kB. The large amount of information stored in DST files is reduced into mini-DST to isolate subsets of events of physics processes of interest like hadronic events. The size of a mini-DST of an hadronic event is around 40 kB.

2.3 Reconstruction of stable particles

Reconstruction is the process through which raw data collected by the detectors are transformed into manageable physics information, in terms of quantity, quality, and meaningfulness. Several algorithms use low-level objects (detector signals, alignment, and calibration information) combined with our knowledge of relativistic kinematics to produce higher-level objects (tracks, energy deposits, etc). In the work described in this thesis I use information associated to two types of stable particles, that are particles that do not decay prior to the tracking system, and that are reconstructed by the detector: charged particles, related for example to the second B meson produced in the collision or reconstructed in the final states of various control channels, and photons, that form the final state of my signal channel. An outline of the essential aspects of the reconstruction of these, along with the associated performance quantities follows.

2.3.1 Charged-particle reconstruction

The ideal trajectories of charged particles in a solenoidal magnetic field are helical, with radius proportional to their transverse momentum. This ideal configuration can be altered by effects such as Coulomb scattering or other energy losses. When reconstructing a track, that is, measuring its momentum and position of closest approach to the interaction point, we need to take into account for these possible effects.

Track reconstruction, or “tracking”, in Belle II [71] consists in the combination of sequences of hits (measurement space-points) into tracks (full trajectories) after a charged particle crosses multiple active layers. The first step is called track finding; the second, track fitting. Tracking relies on PXD, SVD, and CDC information. Due to the different properties of these detectors, specific algorithms are used for each.

As a first step of track finding, hits in the outer tracking volume (CDC), where lower occupancy aids track finding, are filtered and reconstructed by two independent algorithms. One is a global track finding based on the Legendre algorithm [72], that transforms the position of each hit into a (θ, ρ) pair, which represents all the circles traversing both the IP and the considered hit. Another is a local algorithm that takes into account possible non-circular trajectories. The global track finding searches for patterns of hits consistent with helical trajectories, accounting for layer inefficiencies, while local track finding detects extended patterns of nearby hits, to complement the global search and detect short tracks and tracks displaced from the IP. The results of both algorithms are merged and the resulting CDC-only tracks are fitted by an iterative fitter based on the Kalman filter technique, that accounts also for possible random perturbations on the trajectory due for example to multiple scattering or energy losses [73].

Then, tracks are extrapolated inward making sure to avoid duplications, and SVD information is added. They are fitted again, before being extrapolated further inward to the PXD to define regions of interest around their expected intersection points. If an excited pixel is found inside this region, it is included in the pattern recognition algorithm, otherwise it is discarded.

Finally, the parameters of the track are determined thanks to a fitting algorithm and by assuming a mass hypothesis (Fig. 2.13):

- d_0 , the distance of the point of the closest approach to the z axis;
- ϕ_0 , the angle between the transverse momentum and the x axis at the point of the closest approach;
- ω , the track curvature signed according to the particle charge;
- z_0 , the z coordinate at d_0 ;
- $\tan \lambda$, the tangent of the angle between track momentum and transverse momentum.

Track reconstruction is subjected to uncertainties and errors. A track might sometimes be a fake track, if it includes hits from beam-induced background or combines hits from two different particles, or a clone track, if other tracks are reconstructed from the same particle.

Tracking efficiency, that is the efficiency in reconstructing the track of a particle produced after a collision in the detector acceptance, varies from 75% at low transverse momenta ($\mathcal{O}(10)$ MeV) to 95% around $4 \text{ GeV}/c$. It degrades the closer the track is to the beam axis (small or large polar angles), while it is mostly constant around 90% regardless of the azimuthal angle.

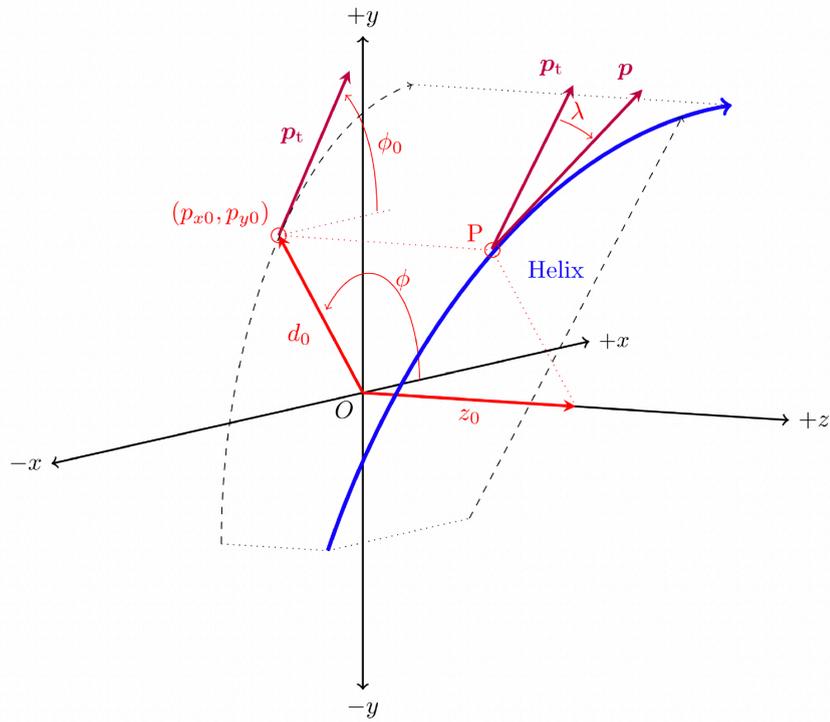


Figure 2.13: Three-dimensional representation of the helical trajectory of a track at the point of closest approach (P) to the IP (the origin O). Symbol p is the momentum of the charged particle at the point P , p_t its transverse momentum and λ is the angle between the two vectors.

The observed transverse momentum resolution is $\sigma(p_T)/p_T = 0.0011p_T[\text{GeV}/c] \oplus 0.0025/\beta$ as shown in Fig. 2.14. The momentum- and angle-dependent impact parameter resolutions are $\sigma_{xy} = 10 \oplus 25/(p\beta\sin^{3/2}\theta)\mu\text{m}$ and $\sigma_z = 15 \oplus 27/(p\beta\sin^{5/2}\theta)\mu\text{m}$ for the transverse and longitudinal projections, respectively.

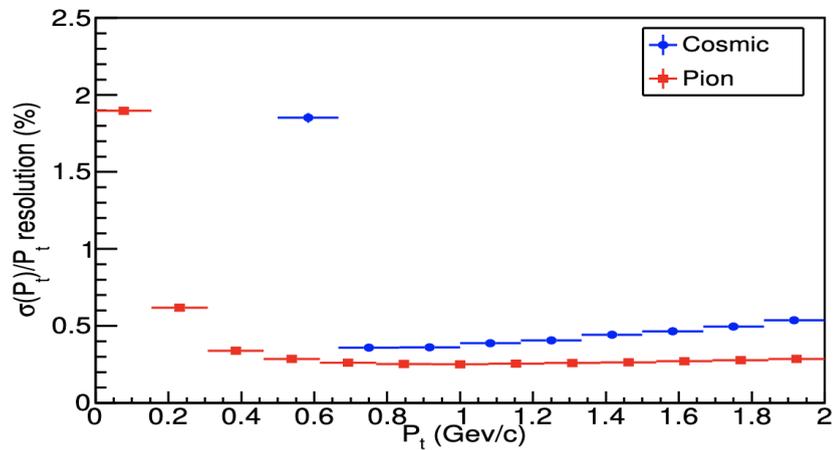


Figure 2.14: Transverse momentum resolution for collision and cosmic ray data.

2.3.2 Charged-particle identification

Particle identification is essential in flavor physics as most of the interesting channels are suppressed and therefore affected by signal-like backgrounds that only differ by the identity of some final-state hadrons. Particle identification (PID) at Belle II is achieved by combining information from several subdetectors. The trajectories of charged particles reconstructed by the tracking detectors, PXD, SVD, and CDC, are extrapolated outward to the TOP, ARICH, ECL and KLM detectors, where geometric matching between the tracks and observed signals is attempted. Offline reconstruction associates PID-detector information sensitive to its identity to each matching track. For example, the drift chamber output encodes information on the specific ionization energy loss associated with each track. The raw information is further processed to provide higher-level quantities that are more convenient for usage in analysis. These are typically ‘likelihood’ values associated to the track. For each of six possible mass hypotheses, kaon, pion, electron, muon, proton and deuteron, the likelihood expresses the probability to observe the reconstructed PID information if the mass hypothesis was true.

For each detector and particle-hypothesis, the likelihood is usually obtained by comparing the expected and the observed value of the raw information, taking into account the uncertainties. For instance, in the CDC such information is $dE/dx_{\text{obs}}(h)$, the specific-ionization energy-loss observed for a charged particle h , averaged across the CDC wires. The resulting (natural logarithm of) the likelihood is

$$\ln \mathcal{L}_{\text{hyp}}^{\text{CDC}}(h) = -\frac{1}{2} \left[\frac{\frac{dE}{dx}_{\text{obs}}(h) - \frac{dE}{dx}_{\text{exp-hyp}}(h)}{\sigma_{\text{obs}}(h)} \right]^2, \quad (2.1)$$

where ‘hyp’ represents the particle hypothesis and $\sigma_{\text{obs}}(h)$ is the observed uncertainty on $dE/dx_{\text{obs}}(h)$, which mainly depends on the number of CDC hits associated to h . The expected value $dE/dx_{\text{exp-hyp}}(h)$ is the average ionization-energy loss from a charged particle h that has the observed momentum, assuming the hypothesis ‘hyp’, calculated using the Bethe-Bloch equation [74, 75] modified according to minor empirical adjustments to adapt to the details of the CDC response. Figure 2.15 shows the $dE/dx_{\text{obs}}(h)$ distribution for various particle species in Belle II data and the expected energy loss for each of the six mass hypotheses considered.

In the TOP, the likelihood is calculated by comparing the observed number of detected photons associated to the charged particle with the photon yield expected from simulation [77, 78],

$$\ln \mathcal{L}_{\text{hyp}}^{\text{TOP}}(h) = \sum_{i=1}^N \ln \left[\frac{S_{\text{hyp}}(x_i, t_i, h) + B(x_i, t_i)}{N_e(h)} \right] + \ln P_N(N_e(h)), \quad (2.2)$$

where x_i and t_i are, respectively, the positions and times of arrival of the N Cherenkov photons excited by the charged hadron h . The term $S_{\text{hyp}}(x, t, h)$ is the signal distribution for the hypothesis ‘hyp’; $B(x, t)$ is the distribution of background; and $N_e(h) = N_{\text{hyp}}(h) + N_B$ is the expected number of detected photons, which is the sum of the expected number of signal photons $N_{\text{hyp}}(h)$ for hypothesis ‘hyp’ and background photons N_B . The second term in Eq. (2.2) is a probability for a Poisson with mean N_e to generate N photons [77]. Figure 2.16 shows an example of the identification of a kaon in the TOP detector: the positions and arrival times of Cherenkov photons are compared with the values expected for a pion or a kaon.

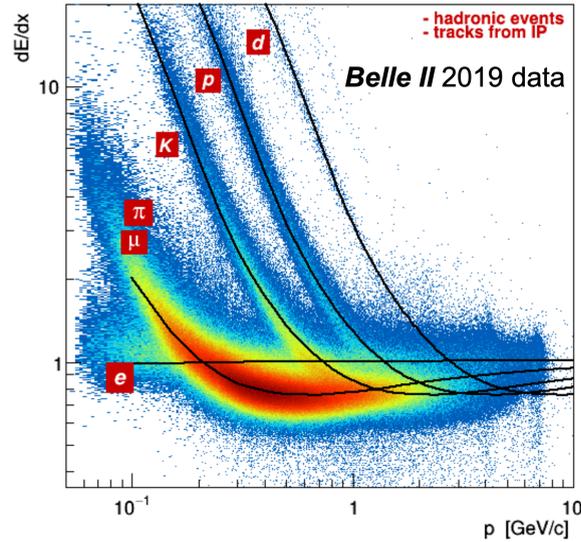


Figure 2.15: Distribution of (points) observed ionization-energy loss as a function of momentum for charged particles from hadronic events reconstructed in Belle II data, along with average expected values (solid lines). Reproduced from Ref. [76].

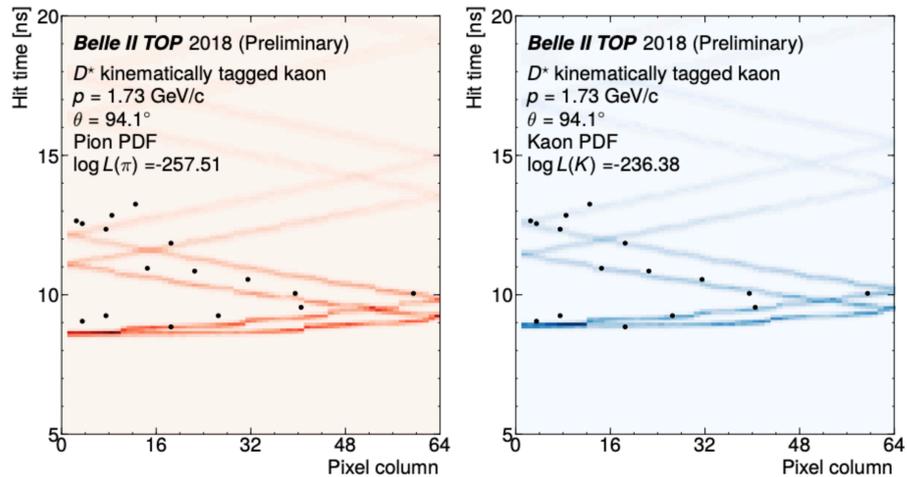


Figure 2.16: Example of kaon identification in the TOP detector. Arrival time of the Cherenkov photons as a function of position is compared with the expectations for (left) a pion and (right) a kaon passing in the TOP [76].

Using the likelihoods for the various mass hypotheses, Belle II algorithms construct a particle identification variable $\text{PID}_{\text{hyp}}^{\text{det}}$ for every detector,

$$\text{PID}_{\pi}^{\text{det}}(h) = \frac{\mathcal{L}_{\pi}^{\text{det}}(h)}{\mathcal{L}_{\pi}^{\text{det}}(h) + \mathcal{L}_K^{\text{det}}(h) + \mathcal{L}_e^{\text{det}}(h) + \mathcal{L}_{\mu}(h) + \mathcal{L}_p^{\text{det}}(h) + \mathcal{L}_d^{\text{det}}(h)}, \quad (2.3)$$

which is directly used in physics analyses. As an example, this was the $\text{PID}_{\text{hyp}}^{\text{det}}$ expression associated with the pion mass hypothesis¹, but the $\text{PID}_{\text{hyp}}^{\text{det}}$ values for other mass hypotheses are obtained by replacing the likelihood at the numerator $\mathcal{L}_{\pi}^{\text{det}}(h)$ with the corresponding hypothesis-specific value $\mathcal{L}_{\text{hyp}}^{\text{det}}(h)$. The quantity $\text{PID}_{\text{hyp}}^{\text{det}}(h)$ is defined similarly to a likelihood ratio $\mathcal{L}_0/\mathcal{L}_1$, which is the best-performing quantity to test two alternative simple statistical hypotheses [79]. The $\text{PID}_{\text{hyp}}^{\text{det}}$ variable assumes values from 0 to 1. The larger the $\text{PID}_{\text{hyp}}^{\text{det}}$, the higher the probability of observing the reconstructed track assuming true the chosen mass hypothesis.

Information from individual detectors is combined to improve the identification performance. The detector-specific likelihoods are combined together as a product,

$$\mathcal{L}_{\text{hyp}}(h) = \mathcal{L}_{\text{hyp}}^{\text{TOP}}(h)\mathcal{L}_{\text{hyp}}^{\text{CDC}}(h)\mathcal{L}_{\text{hyp}}^{\text{SVD}}(h)\mathcal{L}_{\text{hyp}}^{\text{ARICH}}(h)\mathcal{L}_{\text{hyp}}^{\text{ECL}}(h)\mathcal{L}_{\text{hyp}}^{\text{KLM}}(h), \quad (2.5)$$

and the result is used in Eq.(2.6) to obtain the detector-combined PID. If a particle does not get reconstructed in a detector because, for instance, it escapes its acceptance, no PID information from that detector is available and the corresponding individual likelihood is set to one.

Of the two main PID detectors, TOP allows separating pions from kaons at 0.4 – 4 GeV/ c momenta with kaon identification efficiency of 85% and pion misidentification rate of about 10%, while the ARICH separates pions from kaons across all their momentum spectrum and discriminates also pions, electrons, and muons below 1 GeV/ c with 4σ separation or more.

Combining information from all detectors, the electron and muon identification efficiencies are respectively 86% and 88.5% after requiring the binary PID to be larger than 0.9, with pion misidentification rates of 0.4% and 7.3%, respectively. Binary PID is an additional PID variable that compares only two mass hypotheses, for example

$$\text{PID}_{\mu,\pi}^{\text{det}}(h) = \frac{\mathcal{L}_{\mu}^{\text{det}}(h)}{\mathcal{L}_{\mu}^{\text{det}}(h) + \mathcal{L}_{\pi}^{\text{det}}(h)}. \quad (2.6)$$

Data and simulation agree, except at low momenta where discrepancies within 20% are observed. Performance of kaon identification for a threshold of 0.8 on the kaon-pion binary PID is summarized in Fig. 2.17. Efficiency varies from 95% to around 60%, depending on kaon momentum and polar angle. The pion misidentification varies from about 20% to less than 5%.

¹In practice, the Belle II software expresses PID_{hyp} using only the natural logarithm of the likelihood values,

$$\text{PID}_{\text{hyp}} = \frac{e^{\ln \mathcal{L}_{\text{hyp}} - \ln \mathcal{L}_{\text{MAX}}}}{\sum_i (e^{\ln \mathcal{L}_i - \ln \mathcal{L}_{\text{MAX}}})} = \frac{e^{\Delta \ln \mathcal{L}_{\text{hyp}}}}{\sum_i (e^{\Delta \ln \mathcal{L}_i})} \quad (2.4)$$

where \mathcal{L}_{MAX} is the largest of the likelihood values over the six hypotheses.

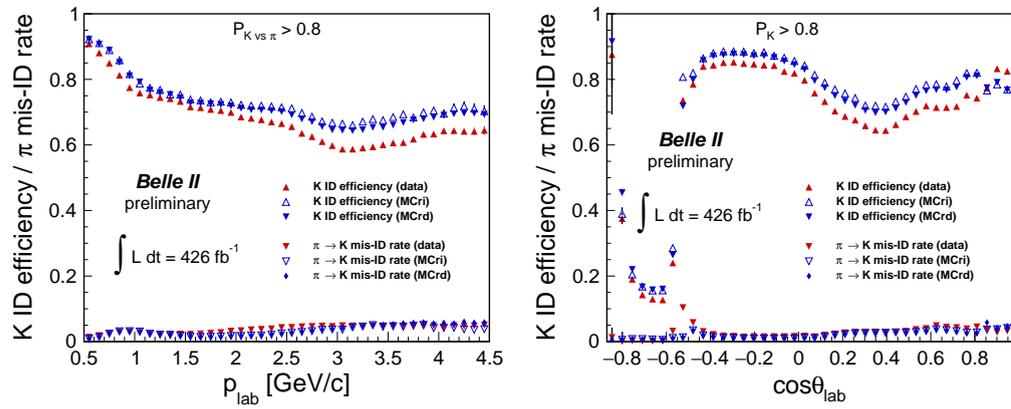


Figure 2.17: Kaon identification efficiencies and pion misidentification rates for events having a binary PID larger than 0.8, in data and simulation (MC) as functions of (left) kaon momentum, and (right) cosine of the polar angle in the laboratory frame.

2.3.3 Photon reconstruction

Photons are reconstructed from energy deposits in the ECL not associated with any outward CDC track extrapolation. When photons impinge on the ECL, they undergo electromagnetic interactions with the crystals thus depositing energy, which is commonly spread across multiple crystals and might overlap with energy deposits from other nearby photons. In each collision that passes the trigger, the information of all 8736 crystals in the calorimeter is recorded for offline analysis. Raw crystal-level information is translated into photon candidates used in Belle II analyses through the photon reconstruction algorithms.

Sets of calorimeter crystals showing energy deposits are called clusters, and the process by which the energies and times observed in each of the crystals are converted into a set of clusters is called cluster reconstruction. The cluster reconstruction algorithm, illustrated in Fig. 2.18, starts by dividing the calorimeter active surface into connected regions, which are contiguous sets of crystals containing significant energy, isolated from all other connected regions. A connected region is assembled starting from a seed crystal, whose energy must exceed 10 MeV. The eight immediate neighbors to the seed crystal are examined and included in the connected region if their energy exceeds 0.5 MeV. If any of these have energy greater than 10 MeV, their neighbors are also examined, and the process repeats. To form an idea of the chosen thresholds, the electronic noise is of order 0.35 MeV, while energy deposits from beam background typically range from 0.5 MeV to 1 MeV in the barrel and forward endcap, and up to 2 MeV in the backward endcap.

Each connected region is then divided into clusters, one per local maximum. A local maximum is a crystal whose energy is greater than the seed energy and is greater than its eight immediate neighbours. If there is more than one local maximum, the energy in each crystal is shared among the resulting clusters. For each local maximum i , each crystal j in the connected region is assigned an energy weight w_j^i equal to

$$w_j^i = E_{\text{sum}}^i e^{-R_j^i/R_w} / \sum_k E_{\text{sum}}^k e^{-R_j^k/R_w},$$

where R_j^i is the distance from the crystal j to the location of the centroid of the cluster associated with the local maximum i , $R_w = 1.43$ cm (40% of the Molière radius), and $E_{\text{sum}}^i = \sum_j w_j^i E_j$ is the total energy of cluster i deposited in the crystals, where E_j is the total energy in crystal j . The weight w_j^i is the fraction of the energy in crystal j assigned to each cluster, and does not depend on the energy in the crystal, but rather on the energy of the clusters and the distance between the crystals and cluster centroid.

The properties of the cluster, including the centroid position, are derived from the corresponding set of associated crystals. The centroid is the weighted sum of the positions of the crystals, where the position weight w_j^i for cluster i from crystal j is

$$w_j^i = 4 + \log(w_j^i E_j / E_{\text{sum}}^i).$$

Only crystals with positive weights (corresponding to crystals containing at least 1.8% of E_{sum}^i) are included in the sum. This is an iterative procedure. The centroid position is initially taken to be the center of the crystal exhibiting a local maximum, and cluster energy to be 50% more than the energy in the crystal showing the local maximum. The position and energy are updated after each iteration, and the process repeats until the centroid positions are stable. The process is similar for connected regions with only one local maximum, but in this case, the energy weights are all unity.

Cluster shape quantities are calculated using all crystals in the cluster. The energy is the sum of energies in the N most energetic crystals, where N depends on the location in

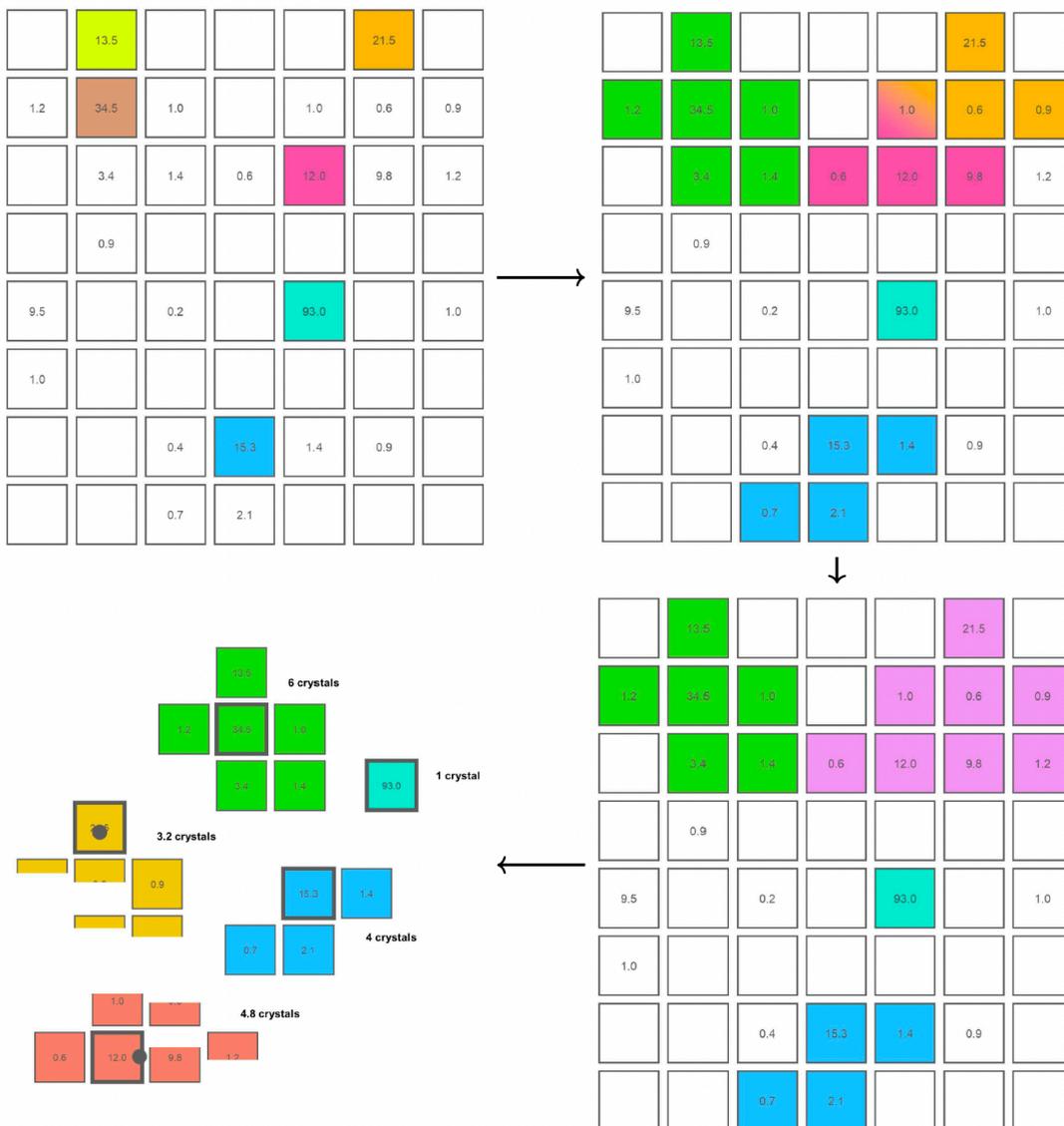


Figure 2.18: Scheme of the steps of cluster reconstruction. (Top left) seed crystals are identified, (top right) neighbouring crystals are attached if they have a large enough energy, (bottom right) connected regions are created, and (bottom left) if a connected region has more than one local maxima, it is split in more clusters.

the detector, the photon energy, and the level of beam-induced backgrounds in the event, estimated using out-of-time events, that are events with times far from the trigger time.

Finally, clusters associated with a CDC track, or with energy less than 50 MeV, or having times further away from the trigger time than (nominally) 99% of real photons of that energy, are discarded.

The energy of the surviving clusters is calibrated using simulated single photon events due to various effects that can modify its true energy: energy leakage out of the back of the calorimeter, energy deposited in inactive material between crystals or in front of the calorimeter, or energy deposited in crystals not included in the cluster energy calculation.

Finally, we associate to the cluster a reconstructed photon (or neutral hadron) candidate.

2.3.3.1 Photon matching

A particle reconstructed in simulation is considered “correctly reconstructed” if it is “matched” with the underlying generator-level particle. In the case of a photon, a cluster is correctly matched if a certain fraction of its reconstructed energy is attributable to the generated particle.

Each cluster can have weighted relations with up to 21 crystals. Each excited crystal, in turn, can have a weighted relation with none, one, or multiple generated particles. The weight between a cluster and a generated particle is given by (weight between the corresponding cluster and crystal) \times (weight between the crystal and the generated particle). The weight between the cluster and the crystal, as described above, is just the fraction of energy the crystal contributes to the cluster. The weight between the crystal and the generated particle is calculated using the total energy deposited by the generated particle in each crystal. If multiple relations exist between a given cluster and generated particle, only the relation with the largest weight is used for photon matching. Truth matching is set between the reconstructed cluster and the generated particle if the following conditions are met:

- the generated particle is actually a photon,
- $\text{weight}/E_{\text{rec}} > 0.2$, and
- $\text{weight}/E_{\text{true}} > 0.3$,

where E_{rec} is the reconstructed energy and E_{true} is the true energy in simulation. If the generated particle is not a photon, such as an electron, no match occurs. This applies even if one of the other lower-weighted relations for the particle is correct.

2.3.3.2 Shower topologies

Once clusters are reconstructed, the distribution of the detected energy within the crystals in each cluster provides information about the spatial distribution of the released energy, so-called "shower-shape". This in turn offers information useful to statistically identify the various phenomena that generated the interaction.

The simplest shower shape originates from a single photon where the maximal energy is deposited in the center crystal of a shower symmetric around the crystal main axis. The whole shower is typically contained in an array of about 5×5 crystals, even for high photon energies. Example shower shapes are shown in Fig. 2.19 for simulated events.

While electron-induced clusters are intrinsically similar to photon clusters, their shower shape is often different due to additional bremsstrahlung photons emitted in the interaction with material surrounded by the ECL. These photons lead to less well-defined cluster shapes for electrons when compared to photons. In addition, electrons are bent in the magnetic field, which result in a different entry angle into the ECL and a deviation of the shower from the radially symmetric shape even in the absence of additional radiation. Final state radiation photons collinear with the electron resulting from the collision can overlap or merge with the electron shower. Example shower shapes with and without radiated photons are shown in Fig. 2.20 for simulated events.

Highly-energetic π^0 mesons, like in $B^0 \rightarrow \pi^0\pi^0$ decays, yield two photon showers that may overlap or merge. To recover these single showers as π^0 mesons requires a dedicated

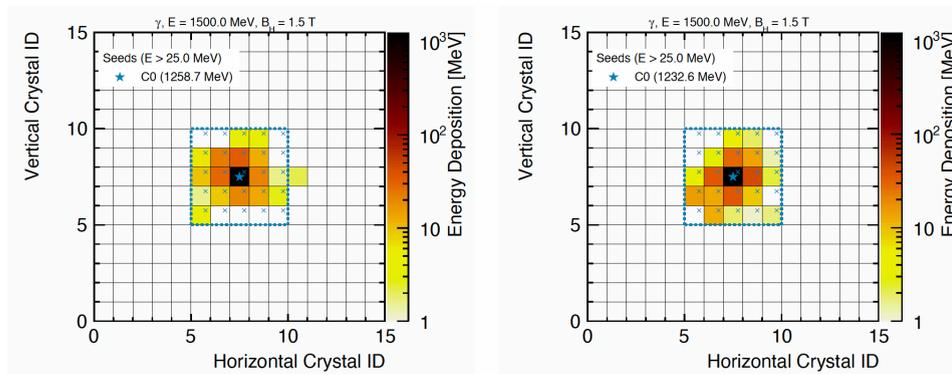


Figure 2.19: Simplified event displays of the energy deposition of two simulated photons with $E = 1.5$ GeV. In addition to the seed crystal, a 5×5 area around the seed is marked.

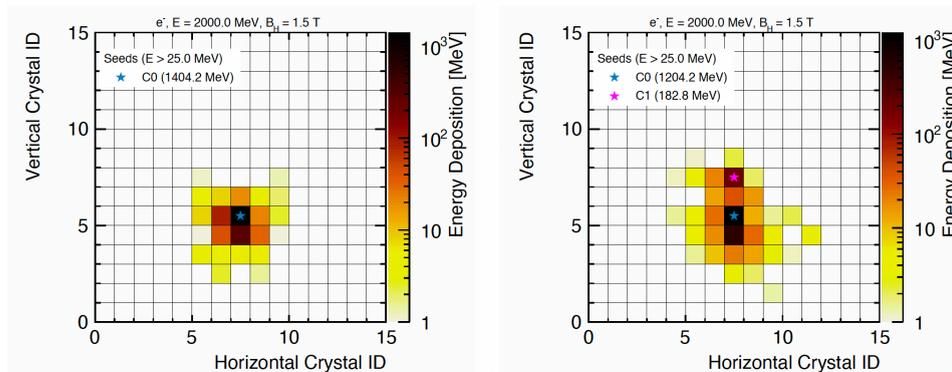


Figure 2.20: Simplified event displays of the energy deposition of (left) one simulated electron with $E = 2.0$ GeV and (right) one simulated electron with $E = 2.0$ GeV and an additional radiated photon.

reconstruction. Example shower shapes with overlapping and merged photons are shown in Fig. 2.21 for simulated events. Overlapping photon showers share crystals but have separate seeds, whereas merged showers have only one seed. In this analysis, the fraction of overlapping and merged photons is expected not to exceed 3%, and we did not include that in the reconstruction.

A fraction of neutral hadrons, such as neutrons and K_L^0 , undergo strong interactions in the ECL: the resulting shower shape is irregular and radial asymmetric. Example shower shapes are shown in Fig. 2.22 for simulated events. If charged hadrons interact strongly within the ECL, they produce irregularly shaped showers in addition to a tilted entry angle due to the magnetic field. If they do not interact strongly, they still leave a small, minimum-ionizing-like signal similar to that of muons. An example of shower shape is shown in Fig. 2.23 (left) for a simulated event. Additional FSR photons may lead to rather complicated topologies of overlapping hadron and electromagnetic showers.

Particles like muons and hadrons that do not interact strongly within the crystal are minimum-ionizing, depositing around 200 MeV almost solely in the crystals directly transversed by the particle. An example shower shape is shown in Fig. 2.23 (right) for a simulated event. Additional radiated photons may lead to an overlapping minimal ionizing signal with electromagnetic showers.

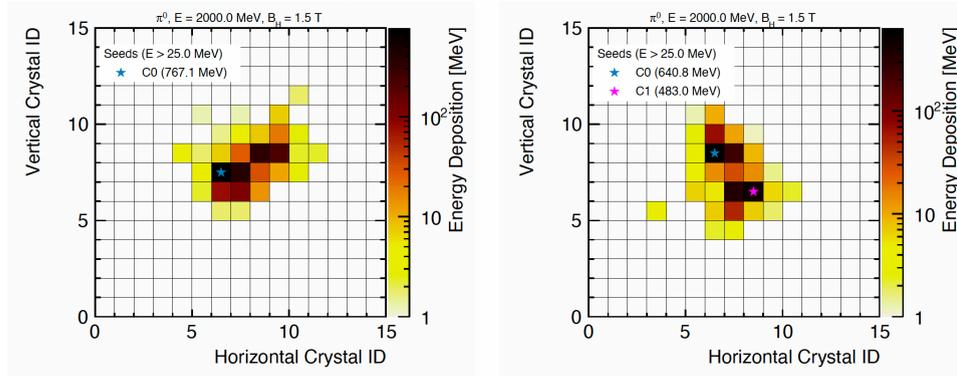


Figure 2.21: Simplified event displays of the energy deposition of (left) one simulated merged π^0 with $E = 2.0$ GeV and (right) one simulated π^0 with $E = 2.0$ GeV with overlapping photon showers.

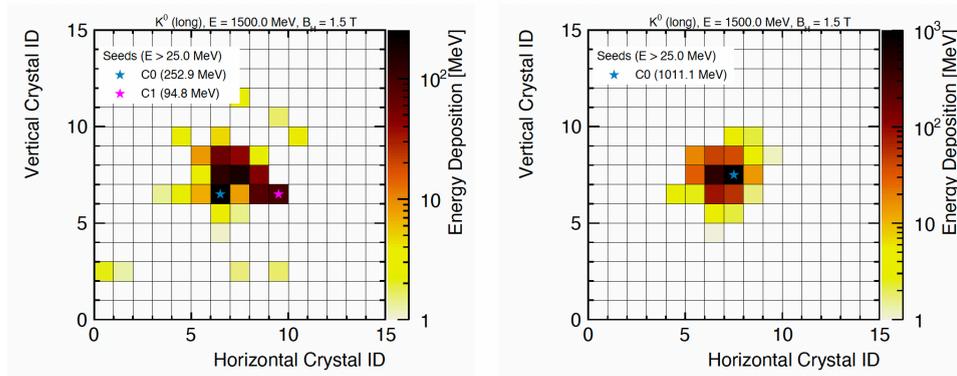


Figure 2.22: Simplified event displays of the energy deposition of (left) one simulated K_L^0 with $E = 1.5$ GeV with two seeds and (right) one simulated K_L^0 with $E = 1.5$ GeV.

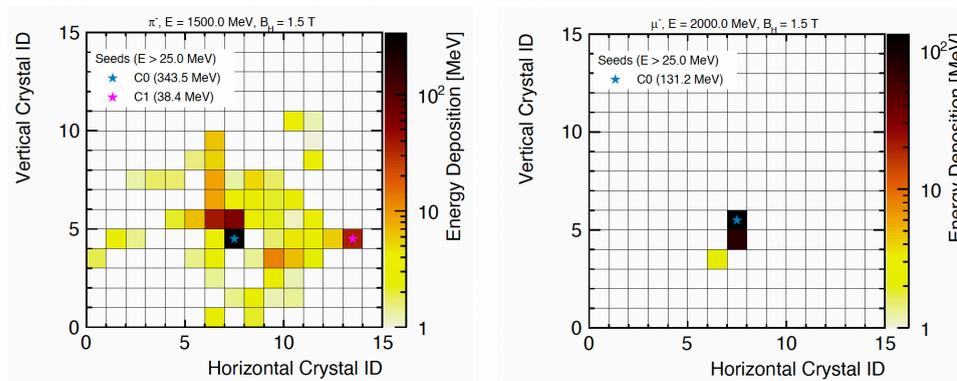


Figure 2.23: Simplified event displays of the energy deposition of (left) a simulated charged pion with $E = 1.5$ GeV and (right) a simulated muon with $E = 2.0$ GeV.

2.3.3.3 Neutral particle reconstruction performance

The energy resolution for photons ranges from $\sigma_E/E = 7.7\%$ at 100 MeV to 2.2% at 1 GeV. The resolution on the reconstructed π^0 mass is $5.4 \text{ MeV}/c^2$.

Chapter 3

Experimental considerations

This chapter discusses several experimental aspects of $B^0 \rightarrow \pi^0\pi^0$ decays and specific features of their analysis in a Belle II.

3.1 Generalities on $B^0 \rightarrow \pi^0\pi^0$ decays

The $B^0 \rightarrow \pi^0\pi^0$ decay mode has a small branching fraction, of the order of 10^{-6} , due to the smallness of the relevant CKM matrix element $|V_{ub}| = 3.82 \times 10^{-3}$. Even with 100% acceptance, selection, and reconstruction efficiency, one expects only about one signal event per fb^{-1} , or about 370 events in the whole Belle II sample. Neutral pions decay into photons or electrons: $\pi^0 \rightarrow \gamma\gamma$ with 98.8% branching fraction, or $\pi^0 \rightarrow e^+e^-\gamma$ with 1.17% branching fraction. In addition, the process $\pi^0 \rightarrow \gamma(\rightarrow e^+e^-\gamma)\gamma$ may occur when a photon interacts with the detector and converts into an electron-positron pair. However, in this analysis only the $B^0 \rightarrow \pi^0(\rightarrow \gamma\gamma)\pi^0(\rightarrow \gamma\gamma)$ decay is considered, happening about 97.6% of the times. The other two processes allow for measuring the decay position vertex, essential for a time-dependent measurement, but they are not used in the analysis due to the expected marginal signal yield. In all subsequent discussions, all mention of the signal decays refers to the $B^0 \rightarrow \pi^0(\rightarrow \gamma\gamma)\pi^0(\rightarrow \gamma\gamma)$ mode.

The native background rate, about 10^7 higher than signal, compounds the signal extraction problem. Continuum background generates large amounts of signal-like high-energy π^0 mesons, which can be randomly combined to resemble signal. A further challenge are the limited signal discriminating features due to a final state consisting of only four photons. Reconstruction of photons in the ECL is less precise than track reconstruction: the trajectory resolution is modest, because, contrary to charged particles, trajectories of photons are not finely sampled along their length using precise position sensitive detectors. We assume the photon to be originated in the interaction point and we use the leading ECL crystal polar angle for the direction. In addition, the energy measurement is affected by leakages in the calorimeter and photon backgrounds. Lack of charged particles on the signal final-state also implies no information on the B^0 decay vertex, which would be precise and distinctive to suppress background. Finally, the $\pi^0\pi^0$ final state is common to both B^0 and \bar{B}^0 mesons, so the measurement of A_{CP} has to rely on background-dominated information from the pair-produced B meson coming from the $\Upsilon(4S)$.

Belle II is currently the only active experiment capable of competitively studying this decay owing to backgrounds that are high, but still much smaller than in hadron collisions; known kinematic properties of the initial state; nearly-hermetic detector; and precise calorimeter.

3.2 B decay reconstruction at Belle II

Before delving into the details of my work, it helps introducing the general aspects of B decay reconstruction at Belle II.

Figure 3.1 shows a sketch of a $B^0 \rightarrow \pi^0\pi^0$ decay. Confined bunches of electrons and positrons are brought to collision in the interaction point (IP). Various final states are produced, with proportions given by the associated cross sections. In 0.4% of collisions,

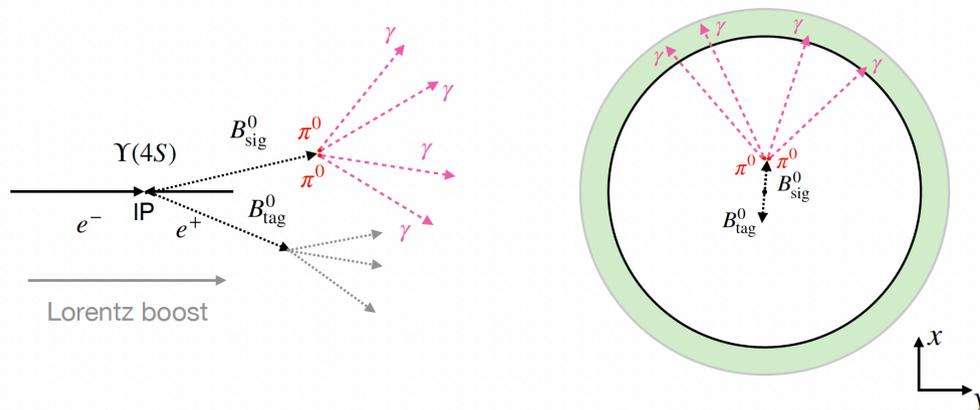


Figure 3.1: Sketch of a $B^0 \rightarrow \pi^0\pi^0$ decay in the (left) longitudinal and (right) transverse views. The green circle represents the ECL detector. Nothing is to scale.

a $\Upsilon(4S)$ meson is produced, and more than 96% of them decay into a pair of B mesons, labeled as signal B -meson, B_{sig} , and partner B -meson, B_{tag} , in Fig. 3.1. These events produce about ten tracks in acceptance on average. However, track multiplicity is lower in $B^0 \rightarrow \pi^0\pi^0$ decays because of the lack of tracks from the signal. The 1.5 ps B meson lifetime, combined with a typical B momentum of 1.5 GeV/ c in the laboratory frame, results in a flight length of about 130 μm for B mesons in general. However, this discriminating information is not available in $B^0 \rightarrow \pi^0\pi^0$ decays because the availability of only photons from the signal does not provide any information on its decay position.

My signal B meson decays into a pair of neutral π^0 mesons of 1.5–3.8 GeV/ c in the laboratory frame, each promptly decaying into pairs of photons. The kinematic properties of the neutral pions are reconstructed by combining the four-momenta of the two photons from their decay, each having typical energies up to 3 GeV in the laboratory frame. These are identified and reconstructed by using information on their energy depositions in the electromagnetic calorimeter. The trajectory is inferred from the position of the signal in the calorimeter and the IP, assumed to be the originating point of the photon. Reconstructed π^0 mesons are then used to determine the kinematic properties of the signal candidate of interest B_{sig}^0 . As the $\pi^0\pi^0$ final state is common to both B^0 and \bar{B}^0 , the flavor of the B_{sig}^0 candidate cannot be inferred from the signal's final-state particles, but from the flavor of the partner B meson, B_{tag}^0 , produced in the collision, if it decays into a flavor specific final state. The flavor is deduced by dedicated algorithms (flavor tagging) that use the properties of the charged decay products of B_{tag}^0 , such as leptons, kaons, and pions.

Every combination of final-state particles that meets the reconstruction quality-requirements is a possible signal candidate. However, signal candidates can originate from genuine signal events or background events. Background events can in turn be associated with three

broadly defined sources.

Misreconstructed π^0 mesons, these are π^0 candidates that are not real π^0 mesons. They can be formed by two photons not coming from the same π^0 , by one photon coming from the π^0 and one misreconstructed or beam background photon, or by two misreconstructed or beam background photons. Beam background and misreconstructed photons are described in Sec. 4.3.1.1;

Continuum, these are neutral pions originated from light quark-antiquark pairs that form random combinations accidentally meeting the signal reconstruction and selection requirements;

B decays (other than signal), these are candidates where a non-signal decay is misreconstructed as signal. A large fraction (85%) are $B^+ \rightarrow \rho^+(\rightarrow \pi^+\pi^0)\pi^0$ and $B^0 \rightarrow K_S^0(\rightarrow \pi^0\pi^0)\pi^0$ decays where a charged or neutral pion, respectively, is not included in the reconstruction.

Belle II exploits available event information to construct various discriminating variables at particle-, candidate-, and event-level to separate signal from background.

3.2.1 Particle-level variables

Several discriminating variables are specific to each reconstructed final-state particle. Relevant examples include

Photon energy. Real photons originating from the $B^0 \rightarrow \pi^0\pi^0$ decay typically have higher energies with respect to misreconstructed photons and beam background photons, originated from the interaction of the beam line with the detector material;

Cosine of the π^0 meson helicity angle. The cosine of the angle between the momentum difference of the two final-state photons in the B^0 rest frame and the boosted π^0 momentum is discriminating against misreconstructed photons. The distribution is uniform for correctly reconstructed π^0 mesons, and peaks at the extremes for wrong combinations. Figure 3.2 shows the distribution of this variable in simulation for real and misreconstructed photons.

Angle between the momenta of the final-state photons of the π^0 meson. In this analysis, real π^0 mesons have typically relatively high energies. Consequently, the laboratory-frame angle between the photon momenta is typically smaller than that of misreconstructed π^0 mesons. Figure 3.2 shows the distribution of this variable in simulation for real and misreconstructed photons. This variable is highly correlated to the π^0 invariant mass.

A complete list of photon and π^0 meson variables useful to discriminate signal from background is discussed in the next chapter.

3.2.2 Candidate-level variables

A class of higher-level discriminating features is associated with the reconstructed signal candidates.

Using distinctive kinematic information of the signal is a common approach to suppress background in many experimental environments. A widely used and effective discriminator

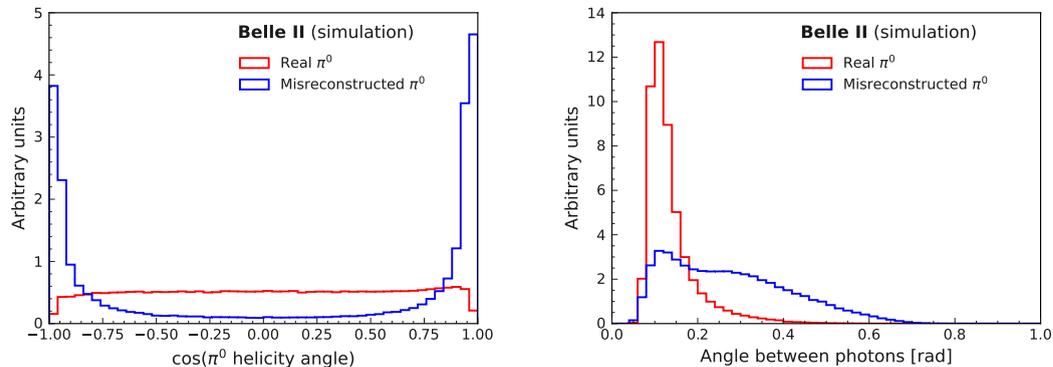


Figure 3.2: Distribution of (left) cosine of the helicity angle and (right) angle between photon momenta in the laboratory frame for real π^0 mesons originated from simulated signal $B^0 \rightarrow \pi^0\pi^0$, in red, and for candidate π^0 mesons reconstructed in background, in blue. Distributions normalized to each other.

is the candidate’s invariant mass, since fully reconstructed signal events cluster at a specific mass value and background shows typically broader distributions.

The peculiar kinematic environment of B -factory colliders provides additional constraints that further background separation. The $\Upsilon(4S)$ is produced almost at threshold and decays in two same-mass particles, B and \bar{B} . If the B meson is correctly reconstructed, the energy of its decay products equals half of the collision energy in the center-of-mass frame. This is optimally exploited by two variables specific of B -factories (all quantities in the $\Upsilon(4S)$ frame).

Beam-constrained mass,

$$M_{bc} = \sqrt{s/4 - |\vec{p}_B^*|^2}, \quad (3.1)$$

where \vec{p}_B^* is the momentum of the B meson reconstructed from the momenta of its decay products, and s is the squared collision energy. The M_{bc} variable is a distinctive reparametrization of the B meson momentum that has a narrower signal distribution than the B invariant mass since the beam-energy spread is smaller than the uncertainty on the reconstructed B -meson energy. All B decays fully reconstructed using their decay products peak at the B meson mass regardless if the products are correctly or incorrectly identified, while non- B events and partially reconstructed B decays have a smooth distribution. Both distributions drop to zero at the kinematic limit of half the collision energy. This makes M_{bc} powerful to separate B -events from $q\bar{q}$ continuum events.

Energy difference,

$$\Delta E = E_B^* - \sqrt{s}/2, \quad (3.2)$$

that is the difference between the reconstructed B -candidate energy and half of the collision energy, which is known with high precision. If the B meson is correctly reconstructed, the energy of the decay products equals approximately half of the collision energy. Therefore, B signals peak at zero, while continuum background follows a smooth distribution. In addition to discriminating against continuum, ΔE suppresses background from misidentified B decays. If a B final-state particle is misidentified as another, its reconstructed energy, and consequently that of the B

candidate, will depart from its true energy because of the mismatch in mass, resulting in a ΔE shift.

Figure 3.3 shows a sketch of the ΔE and M_{bc} distributions for signal decays, continuum background events, and candidates reconstructed in other $B\bar{B}$ events, demonstrating the discrimination of these variables. Since both ΔE and M_{bc} are computed using the momenta of final-state photons, the spread of both of the signal peaks is dominated by the energy resolution, yielding dependencies between their distributions.

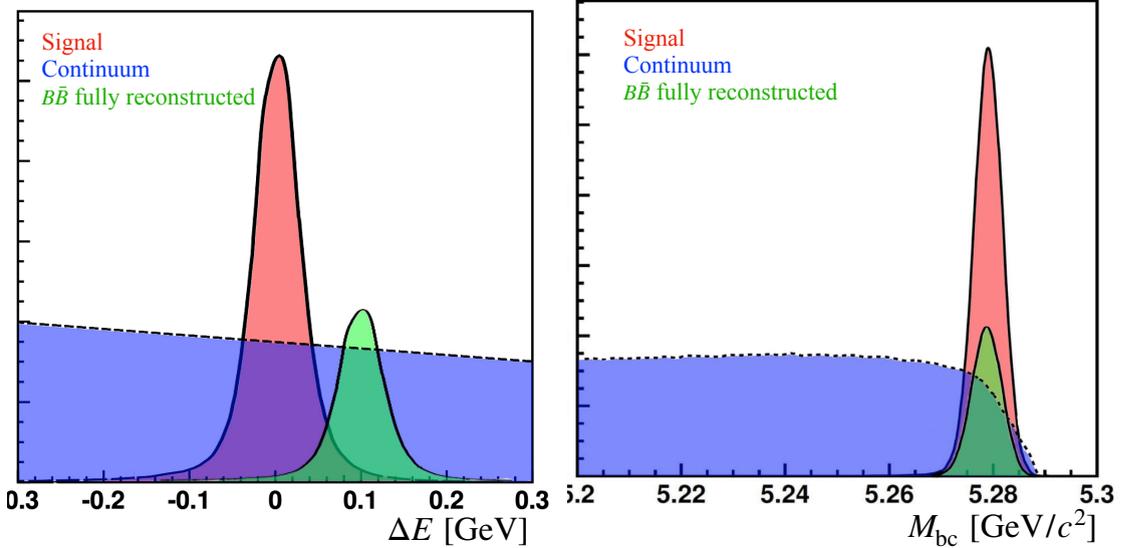


Figure 3.3: Sketched distributions of (left) ΔE and (right) M_{bc} for signal, continuum and fully reconstructed candidates in $B\bar{B}$ events.

3.2.3 B_{tag}^0 -level variables

Having no information on the B_{sig}^0 vertex, we lack the important portion of lifetime-related signal information, which is usually employed to distinguish between continuum and $B\bar{B}$ events. Still, we use observables related to the partner B_{tag}^0 to infer the production of a $B\bar{B}$ pair. If these observables indicate that a partner B meson is reconstructed in an event, that information obviously correlates with the presence of a B meson and therefore helps suppressing continuum:

Vertex displacement. In continuum events, all tracks typically originate from the interaction point, unlike in $B\bar{B}$ events, where the vertices of two B mesons are displaced due to the lifetime of B mesons and their momenta. Usually, the displacement between the decay position of the signal B vertex and that of the other B along the beam axis offers good discrimination of $B\bar{B}$ events from continuum [80]. In this analysis, only the vertex displacement of the B_{tag}^0 with respect to the production space-point along the beam axis is available. Figure 3.4 (left) shows that this observable has discriminating power.

Track displacement. Interactions between beam particles within the same bunch, or with residual gas, may result in interactions of the beam halo with the SuperKEKB or Belle II infrastructure, yielding intense showers of secondary particles that illuminate

the detector (referred to as beam backgrounds). Measurements of track displacement from the IP are effective to suppress such backgrounds, because beam-background tracks do not usually point back to the interaction point. The quantities typically used are the transverse (dr) and longitudinal (dz) distances of a track from the IP. In this analysis we use the displacement of the tracks originating from the B_{tag}^0 ;

Particle identification. Information associated with the identity of charged particles enhances discrimination against background, especially from candidates from misreconstructed B decays. As an example, Fig. 3.4 shows the distribution of the muon PID of the most energetic charged particle originating from the B_{tag}^0 .

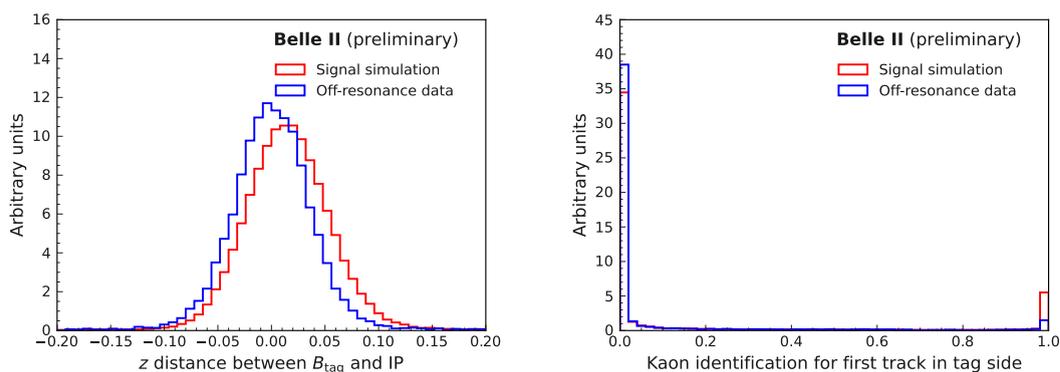


Figure 3.4: Distribution of (left) vertex displacement along the beam axis and (right) kaon PID of the most energetic charged particle originating from the B_{tag}^0 in simulated signal $B^0 \rightarrow \pi^0\pi^0$ and off-resonance data continuum events. Distributions normalized to each other.

Information on B_{tag}^0 is not only used to discriminate between signal and background, but also to determine the flavor of the neutral B_{sig} due to the quantum-entangled flavor discussed in Sec. 2.1. In this analysis, the $\pi^0\pi^0$ final state is common to both B^0 and \bar{B}^0 mesons, so flavor cannot be distinguished using signal information, and relies on the flavor of the pair-produced tag meson. This is accomplished using so-called flavor tagging algorithms, or flavor taggers. Flavor taggers search for flavor-specific final states, such as semileptonic B decays like $B^0 \rightarrow D^{*-}\ell^+\nu_\ell$ or hadronic decays like $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$, where the flavor is inferred from the charge of the charged lepton or K meson. Instead of reconstructing a limited set of specific decays, flavor tagging algorithms apply inclusive techniques to maximally exploit the information provided by an extended number of B_{tag}^0 decays. The Belle II algorithm employs particle identification information and kinematic properties of the charged particles coming from the B_{tag}^0 decay, as well as global information associated with all the tracks in the event to determine the flavor of the tag meson.

3.2.4 Event-level variables

At a further level of abstraction, global-event information also provides signal-from-background discrimination.

Hadronic e^+e^- cross-sections are dominated by continuum background, consisting in production of light $q\bar{q}$ pairs that mostly yield pions and kaons. The kinematic features

associated with at-threshold $B\bar{B}$ production make variables capable to capture the "shape" of the event, that is, the spatial and phase-space distributions of final-state particles, powerful discriminators of $B\bar{B}$ events from continuum.

Figure 3.5 shows an illustrative sketch of the event shapes of a continuum and a $B\bar{B}$ event. In a $B\bar{B}$ event, both B mesons are nearly at rest in the $\Upsilon(4S)$ frame. The B decay products are therefore emitted isotropically in that frame, unlike light quarks, which are produced with a comparatively large initial momentum due to their small mass compared to the total collision energy. This results in their fragmentation to develop along two collimated back-to-back jets of light hadrons. Hence, the spatial and energy-momentum distributions of $B\bar{B}$ decay products are approximately spherical, compared to pencil-like shapes for continuum.

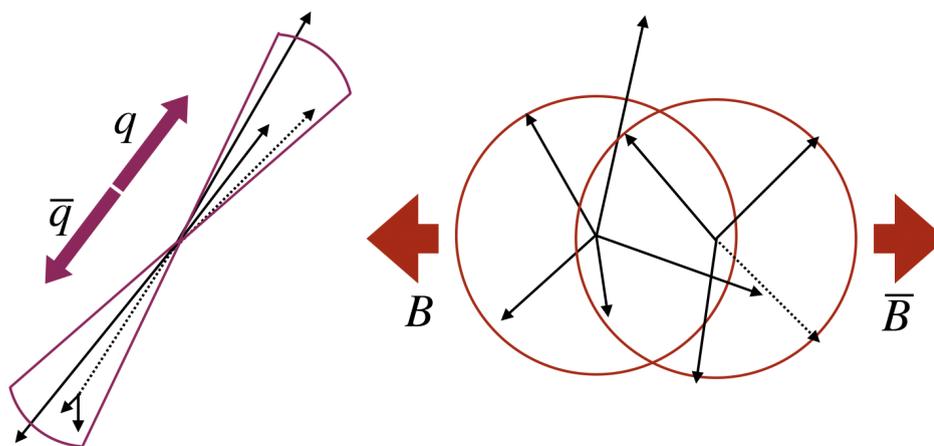


Figure 3.5: Event-shape sketch for a (left) pencil-like continuum and (right) sphere-like $B\bar{B}$ event in the $\Upsilon(4S)$ frame.

Information based on these distributions is therefore useful to discriminate signal events from continuum and it is exploited through several variables. Relevant examples are discussed and demonstrated in Figs. 3.6–3.9. In these and other plots, we use real off-resonance data to compare the continuum distributions to those of signal, which are simulated instead.

B meson direction. The momenta of spin-0 B mesons produced in a spin-1 $\Upsilon(4S)$ meson decay have a $\sin^2\theta^*$ angular distribution, where θ^* is the angle between the B momentum and the beam axis in the $\Upsilon(4S)$ frame. The momenta of spin-1/2 quarks, and their resulting jets follow a $1+\cos^2\theta^*$ angular distribution. The variable $|\cos\theta^*|$ allows therefore to distinguish signal B decays and background from continuum events (Fig. 3.6, left panel).

Thrust. This is a variable designed to capture the extent to which the distribution of particles in an event clusters along a principal axis. For N momenta in an event \mathbf{p}_i ($i = 1, \dots, N$), thrust T is defined as

$$T = \frac{\sum_{i=1}^N |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_{i=1}^N |\mathbf{p}_i|}, \quad (3.3)$$

where \mathbf{T} is the unit vector that maximizes the total momentum projection, and therefore, the longitudinal projection of particle's momenta [81]. In a continuum event, particles typically have larger longitudinal momentum, due to their smaller

masses, as they are produced in jets induced by initial quarks, and therefore show values of T closer to unity. The thrust is calculated for the momenta of particles produced in the signal B meson decay, T_{sig} , and for the momenta of all the other particles in the event, T_{tag} (Fig. 3.6, right panel). Continuum events tend to cluster at higher values of the T_{tag} distribution, unlike signal.

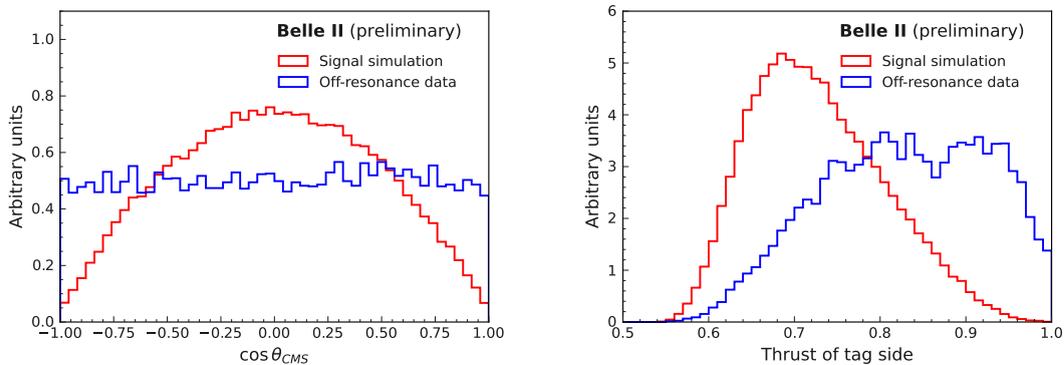


Figure 3.6: Distribution of (left) $|\cos\theta^*|$ and (right) T_{tag} for simulated signal and off-resonance data continuum events. Distributions normalized to each other.

Thrust angles. The cosine of the angles between the \mathbf{T}_{sig} and \mathbf{T}_{tag} axes, $|\cos\theta_{\text{TBTO}}|$, and between the \mathbf{T}_{sig} and the z axis, $|\cos\theta_{\text{TBz}}|$, are also used to discriminate between $B\bar{B}$ and continuum events offering two of the most powerful discriminators. Figure 3.7 (left) shows $|\cos\theta_{\text{TBTO}}|$ distributions for simulated signal and continuum events. Since the momenta of the decay products of B^0 and \bar{B}^0 are isotropically distributed, \mathbf{T}_{sig} and \mathbf{T}_{tag} are randomly distributed, leading to $|\cos\theta_{\text{TBTO}}|$ and $|\cos\theta_{\text{TBz}}|$ distributed uniformly. For $q\bar{q}$ events, particle momenta are collimated, resulting in strongly directional \mathbf{T}_{sig} and \mathbf{T}_{tag} vectors and yielding peaking $|\cos\theta_{\text{TBTO}}|$ and $|\cos\theta_{\text{TBz}}|$ distributions.

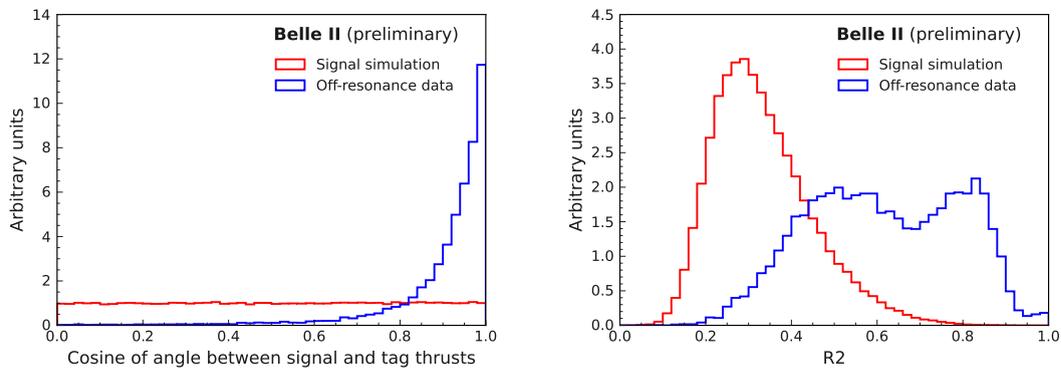


Figure 3.7: Distribution of (left) $|\cos\theta_{\text{TBTO}}|$ and (right) R_2 in simulated signal and off-resonance data continuum events. Distributions normalized to each other.

CLEO cones. The CLEO collaboration introduced nine variables capable of offering additional discrimination against continuum [82]. "CLEO cones" are a refinement of the concept of thrust: they are based on the sum of the absolute values of the mo-

menta of all particles within angular sectors around the thrust axis, in intervals of 10° . This results in nine concentric cones (Fig. 3.8, left panel). The event is “folded” such that the particle content of two cones of same aperture but pointing to opposite directions are combined. Since the flight directions of the B -mesons’ decay products are less correlated than those of continuum events, CLEO-cone distributions provide discrimination against continuum (Fig. 3.8, right panel).

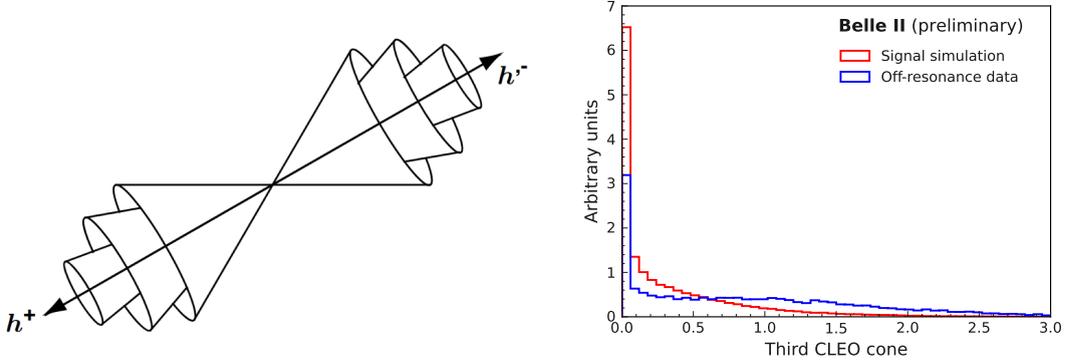


Figure 3.8: (Left) sketch of the first three CLEO cones, where h^\pm represent the two B mesons produced in the collision. (Right) distribution of the third CLEO cone for simulated signal and off-resonance data continuum events. Distributions normalized to each other.

Fox-Wolfram moments and modifications. These are variables designed to capture the principal geometric features of the spatial distributions of final state particles. 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 is defined as

$$H_\ell = \sum_{i,j}^N \frac{|\mathbf{p}_i^*| \cdot |\mathbf{p}_j^*|}{s} \cdot P_\ell(\cos\theta_{i,j}^*), \quad (3.4)$$

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 [83]. The most discriminating example is the normalized second Fox-Wolfram moment $R_2 = H_2/H_0$, which captures the "shape" of spatial distribution and energy flow. Events with collimated jets, and therefore more likely to originate from continuum, show values of R_2 close to 0.5, while $B\bar{B}$ events cluster at lower R_2 values (Fig. 3.7, right panel).

The discrimination power provided by the Fox-Wolfram moments deteriorates when particles are not reconstructed because they are neutrinos, or due to finite acceptances, or inefficiencies. To correct for this limitation, the Belle collaboration developed a variant called Kakuno-Super-Fox-Wolfram moments (KSFW) [52]. All reconstructed particles associated with the signal B (denoted as s for "signal") are treated separately from those in the rest-of-event (denoted as o for "others"). The $H_{x\ell}^{so}$ KSFW moments are sorted into three categories depending on whether the particle is charged ($x = c$), neutral ($x = n$), or missing ($x = m$). For even ℓ , they are defined as

$$H_{x\ell}^{so} = \sum_i \sum_{j_x} |\mathbf{p}_{j_x}^*| \cdot P_\ell(\cos\theta_{i,j_x}^*), \quad (3.5)$$

where i runs over the signal B decay products, and j_x over all other particles belonging to generic category x ; $\mathbf{p}_{j_x}^*$ is the momentum of the particle j_x in the $\Upsilon(4S)$ frame; and $P_\ell(\cos\theta_{i,j_x}^*)$ is the ℓ th order Legendre polynomial of the cosine of the angle between the momenta of particles i and j_x in the $\Upsilon(4S)$ frame. For odd ℓ one obtains $H_{nl}^{so} = H_{ml}^{so} = 0$ and

$$H_{cl}^{so} = \sum_i \sum_{j_x} q_i \cdot q_{j_x} \cdot |\mathbf{p}_{j_x}^*| \cdot P_\ell(\cos\theta_{i,j_x}^*), \quad (3.6)$$

where q_i and q_{j_x} are the charges of the particles i and j_x . There are eleven $H_{x\ell}^{so}$ moments in total, two for $\ell = 1, 3$ and nine (3×3) for $\ell = 0, 2, 4$. Other five KSFW moments are associated with the rest-of-event particles,

$$H_\ell^{oo} = \begin{cases} \sum_j \sum_k |\mathbf{p}_j^*| \cdot |\mathbf{p}_k^*| \cdot P_\ell(\cos\theta_{j,k}^*) & (\ell = \text{even}) \\ \sum_j \sum_k q_j \cdot q_k \cdot |\mathbf{p}_j^*| \cdot |\mathbf{p}_k^*| \cdot P_\ell(\cos\theta_{j,k}^*) & (\ell = \text{odd}), \end{cases} \quad (3.7)$$

where j and k extend over all rest-of-event particles.

To eliminate the dependence on ΔE , which would complicate usage of these variables in analyses based on ΔE for signal extraction, the $H_{x\ell}^{so}$ moments are normalized to H_0^{\max} , and the H_ℓ^{oo} moments to $(H_0^{\max})^2$, where $H_0^{\max} = 2(\sqrt{s} - \Delta E)$. Figure 3.9 shows the distribution of the KSFW moments for $\ell = 2$, which offer the highest discrimination power.

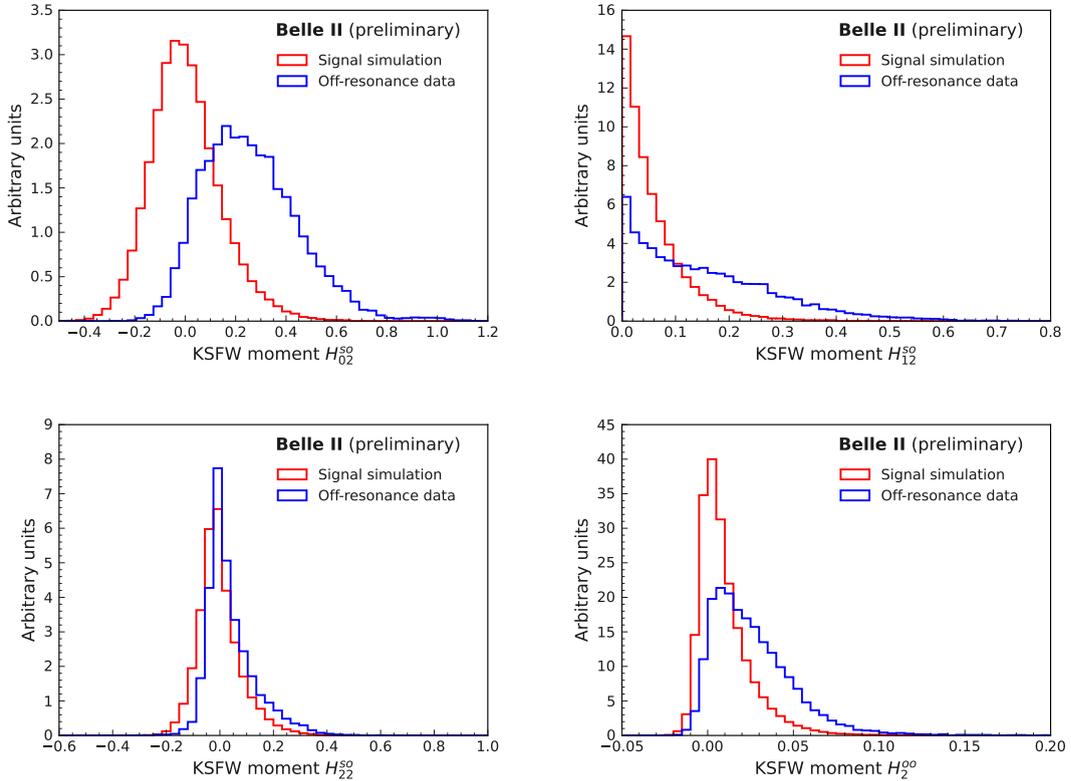


Figure 3.9: Distributions of the $\ell = 2$ KSFW moments for simulated signal and off-resonance data continuum events. Distributions normalized to each other.

3.3 Analysis overview

This work aims at measuring the branching fraction and time-integrated CP -violating asymmetry of $B^0 \rightarrow \pi^0\pi^0$ decays. This is a "blind" analysis; I develop the whole analysis using only simulated- and control-data samples without analyzing the signal-enriched region in data until all procedures are established and final. Such approach prevents from tuning the measurement procedure towards the expected or desired result, reducing the chances for bias.

The principal challenge is to overcome the initial 10^{-7} signal-to-background ratio with a selection sufficiently discriminating to isolate an abundant, low-background signal based only on four photons, and to measure the CP -violating asymmetry using B_{tag} information only. First, I design a multivariate statistical-learning method that combines nonlinearly about ten photon properties to suppress misreconstructed and beam-induced photons. Then, I suppress random combinations of photon pairs by applying requirements on the π^0 meson candidates, determined from simulation or control-data samples. To suppress the continuum background which is dominant, I design another multivariate method that combines nonlinearly about 20 kinematic, final-state topology, and B_{tag}^0 properties. Then I optimize the selection by maximizing the expected statistical precision on the physics parameters of interest. Proper modeling of all discriminating variables used in input to the above classifiers is verified in control samples in data.

After selection, continuum and B background events still dominate the sample composition. I therefore perform a multidimensional fit that allows determining sample composition and estimating A_{CP} . Developing the fit is the core of this analysis. The outcome is a robust model as shown by tests of the estimator properties using simulated data under various experimental conditions.

I study and correct possible effects due to discrepancies between simulation and data using abundant control samples that share relevant features with the signal decay, or signal-depleted control regions ("sidebands") of the signal sample.

Finally, I apply simulation-based efficiency and acceptance corrections to the fit results, calibrated on control samples of data, to determine the final $B^0 \rightarrow \pi^0\pi^0$ results.

Once every step is validated, I apply the analysis to the full Belle II data sample available as of this writing.

Chapter 4

Signal reconstruction and selection

This chapter describes signal reconstruction, in which I select and combine photons into π^0 mesons, and subsequently π^0 pairs into B^0 and \bar{B}^0 mesons applying requirements to suppress background.

4.1 Data sample

I use experimental and simulated data in the analysis.

4.1.1 Experimental data

The Belle II experiment started physics data taking in March 2019, accumulating electron-positron collision data corresponding to an integrated luminosity of $424 \pm 3 \text{ fb}^{-1}$. I use the full data set collected at the energy of the $\Upsilon(4S)$ resonance up to July 2022, corresponding to $362 \pm 2 \text{ fb}^{-1}$, to reconstruct the signal and control modes. I also use data collected at a collision energy 60 MeV below the $\Upsilon(4S)$ mass (called henceforth "off-resonance" data), and corresponding to $42.3 \pm 0.3 \text{ fb}^{-1}$ of integrated luminosity, to characterize and study continuum background. The analysis does not use data collected at a collision energy higher than the $\Upsilon(4S)$ resonance, used for exotic spectroscopy analyses.

4.1.2 Simulated data

I use simulated data to optimize the event selection, to estimate acceptances and efficiency for reconstructing and selecting signal and control modes, to study and model the relevant distributions used in the sample composition fit, and to validate the whole analysis.

Simulated samples are based on the Monte Carlo approach. Monte Carlo samples are produced using event generators, which are computer programs that use pseudorandom number generators to produce sets of four-vectors reproducing final states of e^+e^- collisions according to theoretical models of particle kinematic properties and interactions. Generated data are then subjected to detector simulation, where models of the detector geometry and material are interfaced with models of interactions of particles with matter and signal formation to reproduce the expected values of the raw quantities observed in the detector. These are then subjected to processing and event reconstruction as if they were collision data. The resulting simulated data contain information about reconstructed particles and about the generated *true* particles. By matching these sets of information, we understand whether particles are reconstructed properly, or what are the most frequent misreconstruction occurrences, and what are principal backgrounds. This "truth-matching"

procedure is useful to optimize selection requirements, calculate signal efficiency, classify sample components, and for many consistency checks.

Figure 4.1 shows a sketch of the generation sequence for an hadronic event in Belle II. The properties of virtual photons, created in the electron-positron annihilation, and their subsequent splitting into a quark-antiquark pair, which in turn produces the observed hadrons, is simulated by `Pythia8` [84]. The decay of the heavy hadron (top right corner in Fig. 4.1) is simulated according to a `EvtGen` model for known decays [85] and using `Pythia8` for unmeasured decays. The photon emission by final-state charged particles is simulated by `PHOTOS` [86].

For signal studies, I simulate exclusive samples of $e^+e^- \rightarrow B^0\bar{B}^0$ events, where one of the B mesons is forced to decay to the $\pi^0\pi^0$ final state. The `EvtGen` program simulates the time evolution and known decays of B -mesons according to the relevant experimental and theoretical knowledge. The decay model used in my signal is `SSS_CP`, which describes the decay of a spin-zero into two spin-zero particles, allowing for CP -violating asymmetries. Events are then fed to the standard Belle II detector simulation, based on the `GEANT4` package [87], which simulates interaction with matter and signal formation yielding simulated data in the same format as experimental data.

This analysis uses a centrally produced simulated sample of 2×10^6 $B^0 \rightarrow \pi^0\pi^0$ decays with true $A_{CP} = 0$, about a factor 10^3 larger than expected in data.

For background and validation studies, I use centrally produced simulated samples corresponding to four times the size of the data sample available as of this writing. These include $e^+e^- \rightarrow u\bar{u}$, $e^+e^- \rightarrow d\bar{d}$, $e^+e^- \rightarrow s\bar{s}$, $e^+e^- \rightarrow c\bar{c}$, and $e^+e^- \rightarrow \tau^+\tau^-$ events generated using `KKMC` [88] and `TAUOLA` [89] interfaced with `Pythia8`, and `PHOTOS`.

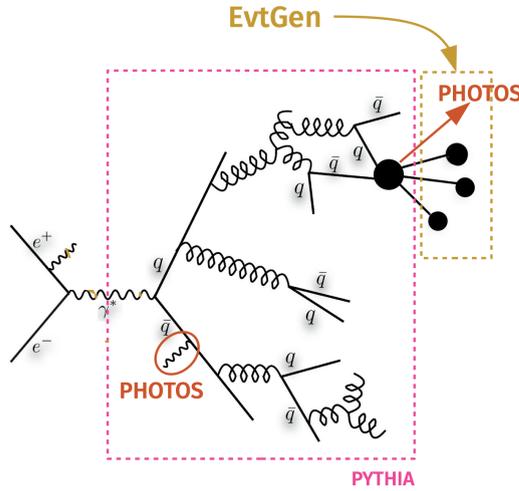


Figure 4.1: Sketch of the various portions of each simulation step of an hadronic event in Belle II.

For $B\bar{B}$ background studies, I use centrally simulated $e^+e^- \rightarrow B^+B^-$ and $e^+e^- \rightarrow B^0\bar{B}^0$ events based on `EvtGen` and `Pythia8`, where B mesons undergo all their allowed decays. The relative proportions among decay modes are based on known values or upper limits when available [11], and on arbitrary or educated guesses otherwise. The simulated $B\bar{B}$ samples correspond to ten times the size of the full Belle II data sample.

In addition to full-fledged, realistically simulated samples, I also use simplified simulated samples (“toy”) based on pseudorandom extraction of simulated values of data observables

based on the likelihood of the fit of sample composition (Chapter 5). These are used mostly for prototyping the analysis, studying estimator properties in the fit of sample composition, and assessing systematic uncertainties.

4.1.3 Basic data structures

As many physics topics are studied in Belle II, data are analyzed multiple times by hundreds of collaborators. To ease the simultaneous analysis of such a large amount of data, various centralized data processing steps are implemented.

Raw data are processed centrally to produce summary data, which are reduced in size and focus on higher-level information related to primitives for physics analysis, including four-momenta, vertices, particle-identification information, and others. A second centralized step consists in applying loose selection criteria on summary data to obtain analysis-specific subsets (*skims*) further reduced in size so that each collaborator can access them and quickly process them. For this analysis, we use a skim specifically designed for $B^0 \rightarrow \pi^0\pi^0$ decays.

4.1.4 Skim selection

For an event to be included in the skim, it must pass at least one of the HLT triggers (see section 2.2.4), which are designed to remove uninteresting events. To remove the high number of events with low-energy photons detected in the ECL detector, an energy threshold for photons is imposed. In the backward and barrel region, the energy is required to exceed 20 MeV; in the forward region, 22.5 MeV. The difference accounts for higher occurrence of forward-directed beam background from the electron beam, which is more intense than the positron beam.

Photons that meet these selection criteria are paired to form π^0 candidates. The diphoton mass is required to be between $0.105 \text{ GeV}/c^2$ and $0.150 \text{ GeV}/c^2$, which corresponds to a range of approximately $+2.0$ and -2.5 standard deviations in resolution about the known π^0 mass. The mass requirement is asymmetric, as the reconstructed π^0 lineshape has a negative skew due to ECL energy leakage. The absolute value of the cosine of the π^0 meson helicity angle is also required to be smaller than 0.98. Candidates B^0 are then reconstructed by combining two π^0 meson candidates, and required to have $M_{bc} > 5.20 \text{ GeV}/c^2$ and $|\Delta E| < 0.5 \text{ GeV}$.

After applying these criteria, 98.8% of the $B^0 \rightarrow \pi^0\pi^0$ signal events are retained, with a rejection of more than 99% of other background events. Running the signal reconstruction code on the skim for the full Belle II data set of 362 fb^{-1} takes about six hours on central processing computers.

4.2 Signal candidate reconstruction

The major challenge of this analysis is to isolate a clear signal based on photons only in a high-background environment. I study dedicated selections and requirements for photons candidates, π^0 meson candidates reconstructed by combining two photons, and B^0 candidates reconstructed by combining two π^0 mesons.

4.3 Photon selection

Before delving into the requirements applied to photon candidates, I introduce the main photon backgrounds and the means to distinguish them from the signal photons.

4.3.1 Photon backgrounds

Photon backgrounds can be broadly divided in two typologies: misreconstructed photons, and beam-induced background photons.

4.3.1.1 Misreconstructed photons

Misreconstructed photons are ECL energy deposits that are reconstructed split into multiple clusters. The two main sources of these multiple clusters are the following:

Hadronic split-offs. A charged hadron (e.g. π^\pm) generates an irregular shower-shape, different from that from photons or electrons (see Sec. 2.3.3.2). Its energy deposition is usually split into multiple clusters, and those sufficiently far from the related CDC track may be reconstructed as photons. Photon matching in simulation can indicate which is the generated particle that deposited energy in the ECL, or not work at all. Particles causing hadronic split-offs reconstructed as photons are mainly charged pions and protons. Figure 4.2 (left) shows a sketch illustrative example.

Low transverse momentum charged particles. Charged particles with low transverse momentum usually pass through many ECL crystals, approximately perpendicularly to the crystals main axis. The energy deposit is split into multiple clusters, but photon matching cannot indicate which was the original generated particle. Figure 4.2 (right) shows a sketch illustrative example.

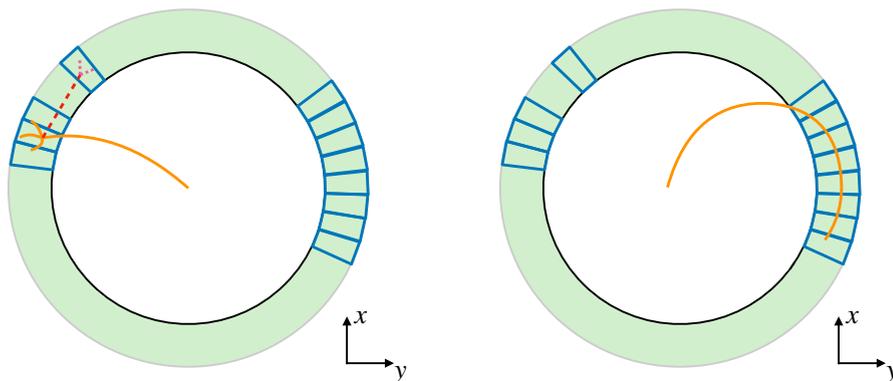


Figure 4.2: Simplified sketch of the ECL (in green) showing (left) a hadronic split-off and (right) low transverse momentum charged particle. ECL crystals are represented in blue.

4.3.1.2 Beam-induced background photons

Beam-induced background photons are photons originated from the electromagnetic interaction of the beam with the detector or associated infrastructure material. This is primarily the result of beam particles being deviated due to bremsstrahlung or Coulomb scattering

with residual gas molecules in the beam pipe or with other beam particles in the same bunch. The latter is known as the ‘Touschek scattering’ and occurs when two electrons in a bunch elastically scatter and transfer momentum from the longitudinal to transverse direction, causing them to depart from the intended trajectory. If their deviation occurs close to the detector, they can generate showers that reach its active parts. Heavy-metal shields in the tracking volume, superconducting final focus cryostats, and horizontal and vertical movable collimators are used to suppress shower particles entering the Belle II acceptance, but they are not fully efficient. Energy resolution in the ECL is expected to degrade with increasing beam background, as shown in Fig. 4.3. Beam background is hard

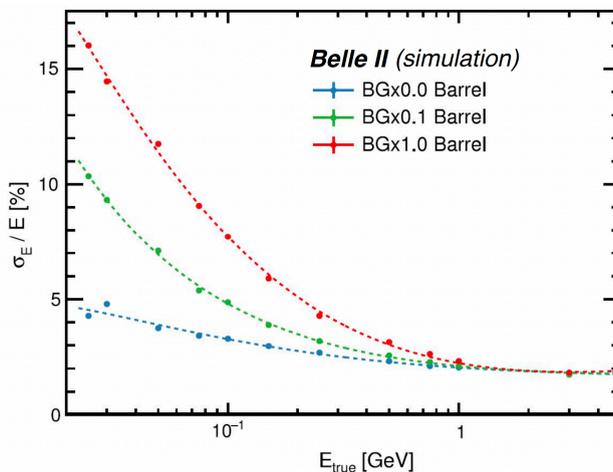


Figure 4.3: Relative ECL energy resolution as a function of real energy for data simulated with nominal (red), 10% (green), and no (blue) beam background.

to study using simulation, as photon matching does not work properly. Beam background is not generated, but only injected – so there is no generated particle to match. High beam background not only affects energy resolution, but also reconstruction. Figure 4.4 shows a quadrant of the calorimeter in the $\theta - \phi$ plane where each square represents a CsI(Tl) crystal. This simulated event contains a signal photon incident in the center of the image in addition to numerous energy deposits from beam-induced backgrounds. The input information to the cluster algorithm is illustrated by Fig. 4.4a showing the observed energy deposit in each crystal as indicated by the color scale. A large number of crystals show energy deposits due to beam-induced backgrounds. The results of the clustering algorithm are shown in Fig. 4.4b, where the cluster of crystals originate from the signal photon is isolated from the numerous additional energy deposits arising from beam backgrounds, which are rejected.

4.3.2 Baseline photon selection

To distinguish between photons produced by a neutral pion decay and beam-background or misreconstructed photons, I first apply baseline selections to the candidate photons, which are required to have energy greater than 30 MeV, a timing of the energy deposit with respect to the event time, recorded by different detectors, smaller than 200 ns, and to have deposited energy in at least two ECL crystals. In addition, I require the photon polar angle to be between 12° and 155° .

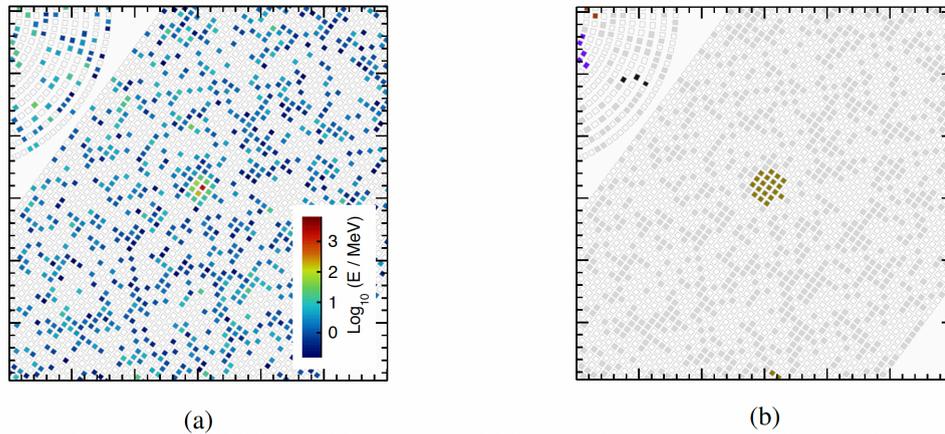


Figure 4.4: Sketch of an ECL quadrant in the $\theta - \phi$ plane for a simulated event containing a signal photon in the center of the image as well as scattered energy deposits from beam-induced radiation backgrounds. Each square represents a CsI(Tl) crystal. a) Crystal energy information used as input to the clustering algorithm. b) Result of the clustering algorithm.

4.3.3 Optimized photon selection

Then, I combine non-linearly various discriminating variables in a fast boosted decision tree [90], a supervised multivariate classifier. The idea is that nonlinear discriminators may achieve a better signal-to-background separation than sequential application of one-dimensional restrictions ("cuts") as they can capture and exploit statistical differences present in the multidimensional correlations over the space of discriminating observables.

A multivariate classifier estimates the probability of a datum to belong to a given class; this probability is inferred from a set of explanatory observables $\mathbf{x} = (x_1, \dots, x_n)$. The algorithm operates in two phases. In the fitting phase, the classifier is "trained" using data with known classification (training sample). In this "supervised" phase the internal configuration of the classifying function that maps the inputs into a classification output is adjusted to maximize the rate of successful classification. Successful classifications are known because the true classification is known for the training data. In the application phase, the resulting classifier is applied to new data with unknown classification (testing sample). In this phase, the internal configuration of the classifying function is established from the training and used to classify the test data. I use a stochastic gradient-boosted decision tree.

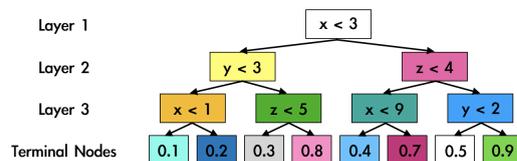


Figure 4.5: Schematic example of a three-layer decision tree. At each node of the tree a binary decision is made until a terminal node is reached. The numbers in the terminal node correspond to a probability of the test data-point to be signal.

The decision tree is a specific type of supervised classifier that approximates the optimal classifying function by applying a set of consecutive binary requirements on each of the

given discriminating observables. The maximum number of consecutive requirements is typically configurable and is called the depth of the tree. A schematic example of a decision tree is shown in Fig. 4.5 for a simple case of three discriminating observables and two classes.

The setting of the selection requirements at each node are determined during training. By varying the requirements at each node, the tree estimates the probability of a training data-point to belong to a certain class of events. These predictions are then compared with the known true classification. Only requirements that result in accurate predictions are implemented in the node. Hence each requirement maximizes locally the separation between classes of events on the given training sample.

Predictions of trees with many consecutive requirements (deep trees) are often driven by the statistical fluctuations of the training data-sample, rather than by genuinely significant distinctive features. This "over-fitting" reduces the predictive power of the tree.

To reduce overfitting, ensembles of shallow trees are combined into a "boosted" tree. While individually each shallow tree may give inaccurate predictions, combining them sequentially yields a model that is less likely to overfit, yielding good classification performance. Boosting proceeds by fitting an initial tree to the data; then a second tree is built targeted at classifying accurately only the events where the first tree performs poorly; and then the sequence is repeated many times. Each successive tree attempts to correct the shortcomings of the combination of previous trees. A common used boosting technique is gradient boosting [91].

The robustness of gradient-boosted decision-trees against overfitting is improved by using random subsamples of the training data set instead of the full training sample in each boosting step. The strength of possible correlations between trees is reduced thus achieving enhanced discriminating capabilities. This approach is called stochastic gradient boosted decision tree [92].

4.3.3.1 Choice of discriminating observables

I explore a broad set of discriminating observables to choose the classifier inputs that offer optimal separation between photons produced in the π^0 decay and beam-background or misreconstructed photons. The chosen discriminating observables need to be well-reproduced by simulation. Otherwise, the classifier trained on the simulated samples will be suboptimal when applied to data.

I explore various ECL-related variables, and converge to a subset of nine discriminating variables, whose distributions for misreconstructed and real photons reconstructed in simulation are shown in Figures 4.8 and 4.9:

- clusterHighestE.** This is the energy of the most energetic crystal in the cluster. Generally, events where the photon deposits almost all of its energy into a single crystal are more likely to be signal events rather than beam background or electron events, which tend to distribute their energy more broadly.
- pt.** Momentum component transverse to the beam line. This parameter is informative because the momentum perpendicular to the beam line is associated with the masses of the decaying particles yielding photons, which is likely to be different between true and misreconstructed or beam-induced background photons. It is calculated assuming the photon originated in the interaction point and multiplying the cluster energy by the sine of the leading crystal polar angle.
- clusterE1E9.** This shower shape variable is the ratio of the energy contained in the central crystal and the surrounding 3×3 grid crystal of a cluster (Fig. 4.6). Photon deposits,

which concentrate their energy in a single crystal, tend toward unity, while hadronic clusters, which distribute their energy across multiple crystals, tend toward smaller values.

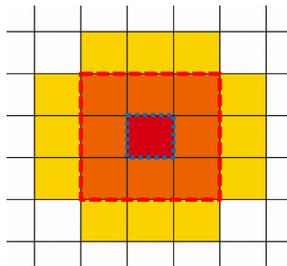


Figure 4.6: Sketch illustrative of `clusterE1E9`, where the blue dotted region encloses the central crystal and the red dotted region encloses the surrounding crystals used in the variable determination.

`clusterLAT`. Lateral energy distribution, defined as

$$\text{LAT} = \frac{\sum_{i=2}^n \omega_i E_i r_i^2}{(\omega_0 E_0 + \omega_1 E_1) r_0^2 + \sum_{i=2}^n \omega_i E_i r_i^2},$$

where E_i are the single crystal energies sorted by energy (E_0 is the highest energy and E_1 the second highest), ω_i is the crystal weight, r_i is the distance of the i th crystal to the shower center projected to a plane perpendicular to the shower axis, and $r_0 \approx 6$ cm is the average distance between two crystals. This variable peaks around 0.3 for radially symmetrical electromagnetic showers, while is larger for hadronic events and electrons with a close-by radiative or bremsstrahlung photon;

`minC2TDist`. "Minimum cluster-to-track distance", that is the shortest distance between an ECL cluster and the extrapolation to the ECL of the nearest track. If the distance exceeds 250 cm, it is capped at that value. If no extrapolated hits are found in the ECL for the event, its value is not a number. This variable is sensitive to identify hadronic split-off photons, which are often in close proximity to the generating track. For signal photons, the closest charged track originates solely from the decay of the tag B meson. Consequently, the distribution of `minC2TDist` is relatively uniform, with the most probable value being a considerable distance away. Meanwhile, ECL clusters that originate from a charged particle, but are misreconstructed into photons, preferably peak at lower values.

`clusterSecondMoment`. Second moment defined as

$$S = \frac{\sum_{i=0}^n \omega_i E_i r_i^2}{\sum_{i=0}^n \omega_i E_i},$$

where E_i are the single crystal energies, ω is the crystal weight, and r_i is the distance of the i th crystal to the shower center projected to a plane perpendicular to the shower axis. This variable peaks around unity for real photons, while is larger for beam background or misreconstructed photons;

`clusterZernikeMVA`. Zernike moments are calculated for each cluster to quantify the circularity of the distribution in a plane (Fig. 4.7) [93]. They are calculated per shower

in a plane perpendicular to the shower direction via

$$|Z_{nm}| = \frac{n+1}{\pi} \frac{1}{\sum_i \omega_i E_i} \left| \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} \omega_i E_i \right|,$$

where n, m are integers, i runs over the crystals in the shower, E_i is the energy of the i th crystal in the shower, R_{nm} is a polynomial of degree n , ρ_i is the radial distance of the i th crystal in the perpendicular plane, and α_i is the polar angle of the i th crystal in the perpendicular plane. As a crystal can relate to more than one shower, ω_i is the fraction of the energy of the i th crystal associated with the shower. Storing all the Zernike moments for each ECL shower would require an excessive amount of memory, so the variable `clusterZernikeMVA` is implemented, which is an artificial neural network combining eleven Zernike moments of the cluster into a single real number.

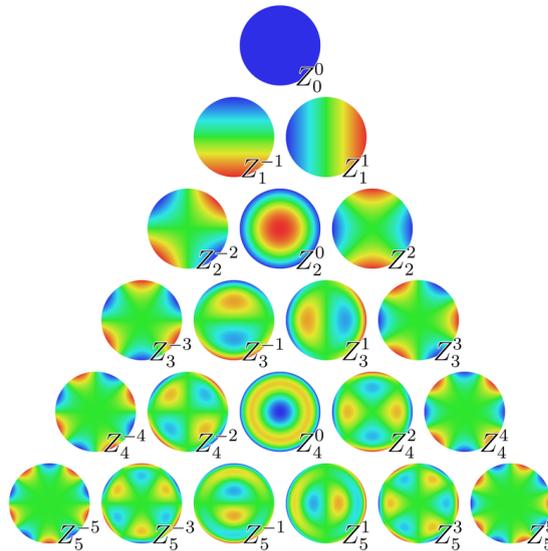


Figure 4.7: Representation of Zernike polynomials at various orders of n and m .

clusterNHits. Sum of weights ω_i ($\omega_i \leq 1$) of all crystals in an ECL cluster. For non-overlapping clusters this equals the number of crystals in the cluster. In case of energy splitting among nearby clusters, the sum can have a non-integer value. It has lower values for misreconstructed photons, and higher values for real photons;

clusterTheta. ECL cluster's polar angle (not generally equal to the photon polar angle). The direction of a cluster is given by the θ coordinate of the centroid position in the ECL. Cluster centroids are generally biased towards the centers of the highest energetic crystal. Beam-background photons are more frequently aligned with the beam line, that is, detected in the forward and backward ECL endcaps, while real photons are more frequent in the barrel.

The correlation matrix for these ECL variables for real photons is shown in Fig 4.10.

To ensure that the chosen observables are well-reproduced in the simulation, I compare their distributions for a sample dominated by real photons obtained in simulated and collision data. I use the $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0(\rightarrow \gamma\gamma))\pi^+$ mode as it is very abundant, straightforward to distinguish from background, and yielding photons over a broad energy

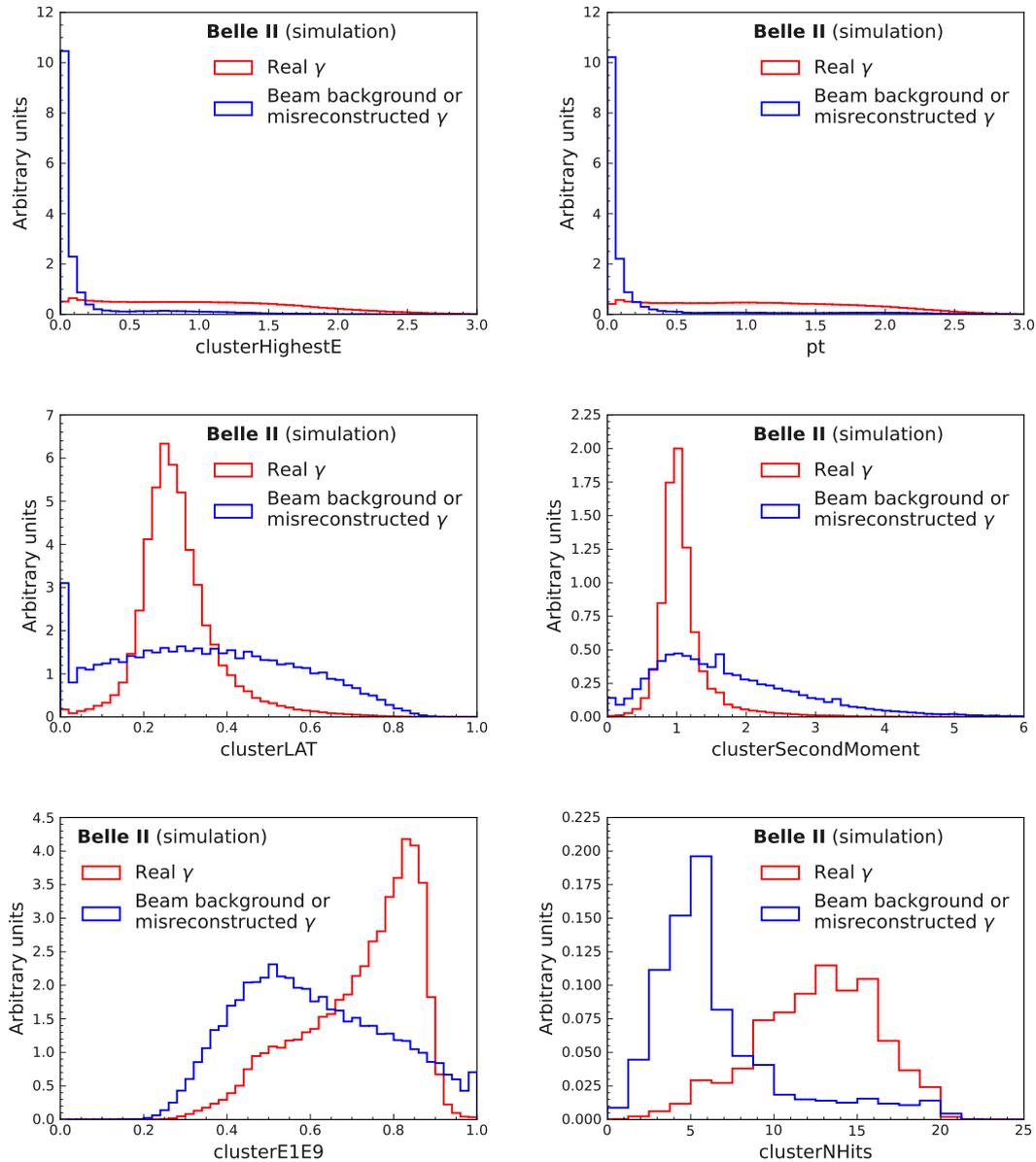


Figure 4.8: Comparison between (red) true and (blue) misreconstructed photons for the variables included in the photon discriminating algorithm, reconstructed in simulation. Distributions normalized to each other (1/2).

spectrum. I reconstruct $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0(\rightarrow \gamma\gamma))\pi^+$ decays in on-resonance simulated and collision data applying the same baseline photon selections used for the signal channel (Sec. 4.3.2), and I weight the photon energy distribution to that of the signal photons. The sample is dominated by true photons.

Figures 4.11 and 4.12 show a data-simulation comparison of the chosen nine discriminating observables. The agreement is high.

Table 4.1 shows the relative importance of the photon BDT input variables. The most important variable of the discrimination is the energy of the highest-energetic cluster. Beam-background and misreconstructed photons typically have low energies, while $B^0 \rightarrow$

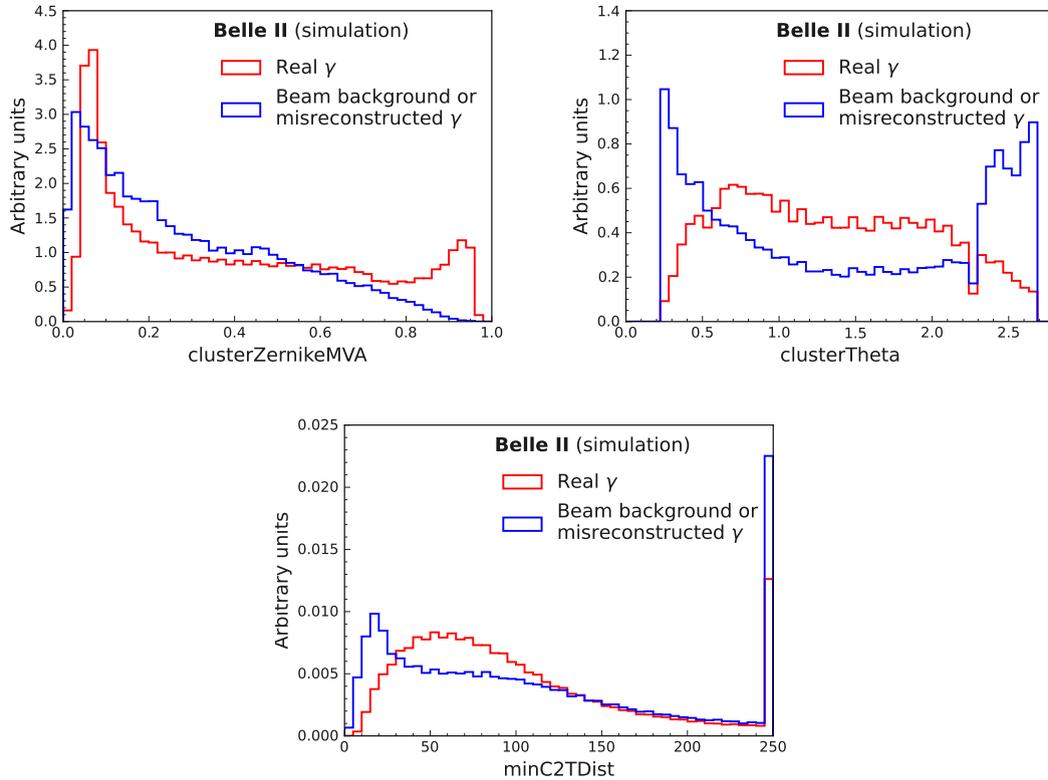


Figure 4.9: Comparison between (red) true and (blue) misreconstructed photons for the variables included in the photon discriminating algorithm, reconstructed in simulation. Distributions normalized to each other (2/2).

$\pi^0\pi^0$ decays have high-energy π^0 mesons and subsequently high-energy photons. The contamination from misreconstructed photons is thus expected to be small in the final sample.

ECL Variable	Relative importance in the FBDT.
clusterHighestE	100
clusterSecondMoment	47
pt	30
clusterLAT	25
minC2TDist	14
clusterZernikeMVA	13
clusterNHits	9
clusterTheta	4
clusterE1E9	1

Table 4.1: Variables used in photon BDT and relative importance.

	E_{ECL}	#Theta	NHits	ZeMVA	mc2TD	LAT	p_{ECL}	2Mom	$E_{\text{ECL}}^{\text{High}}$
$E_{\text{ECL}}^{\text{High}}$	32	-22	79	18	-1	-12	87	-20	100
2Mom	-42	3	11	-15	1	81	-11	100	-20
p_{ECL}	3	-2	79	8	-2	3	100	-11	87
LAT	-49	8	27	-15	-1	100	3	81	-12
mc2TD	1	6	-1	0	100	-1	-2	1	-1
ZeMVA	16	-4	9	100	0	-15	8	-15	18
NHits	5	-11	100	9	-1	27	79	11	79
#Theta	-2	100	-11	-4	6	8	-2	3	-22
E_{ECL}	100	-2	5	16	1	-49	3	-42	32

Figure 4.10: Correlation matrix for the ECL variables chosen to be inputs of the photon selection BDT, as resulting from simulation.

4.3.3.2 Classifier training and test

I train the photon selection BDT using 3×10^4 truth-matched photons and 3×10^4 simulated misreconstructed and beam-background photons passing the baseline selection. A fraction of 65% of the data is used for training while the remaining 35% is used for testing.

A convenient way to estimate the classifier performance is through a receiver operating characteristic (ROC) curve, which represents signal efficiency as a function of background rejection (defined as the complement to the background efficiency). The classifier performance improves as the ROC curve approaches the top-right corner of the two-dimensional space. This is expressed quantitatively by the area under the curve (AUC), which equals unity for the ideal case of 100% signal efficiency with 100% background rejection. The resulting ROC curve is in Fig. 4.13.

I test the photon classifier on an independent simulated sample with same composition as the training data to ensure that the model is not overfitted. Figure 4.13 and Fig. 4.14 compare the ROC curves and the distributions of classifier output for training and testing samples, respectively. The distributions obtained from training and testing data agree for both signal and background components, showing negligible, if any, overfitting.

4.3.3.3 Optimization

The ROC curve (Fig. 4.13) qualifies the classifier performance, but it does not indicate which requirement on the classifier output provides the best selection, as this depends on the physics goal at hand.

I optimize the photon selection by maximizing the ratio between the yield of signal $B^0 \rightarrow \pi^0\pi^0$ decays and the square root of its sum with the background yield, both observed in realistic simulation. I reconstruct photons using the baseline selection (Sec.4.3.2) and scan the classifier across the whole range. A selection at 0.25 is optimal (see Fig. 4.15). In a statistically independent simulated sample, this criterion removes 18.7% of background while keeping 90.8% of signal. A more sophisticated optimization is not necessary because

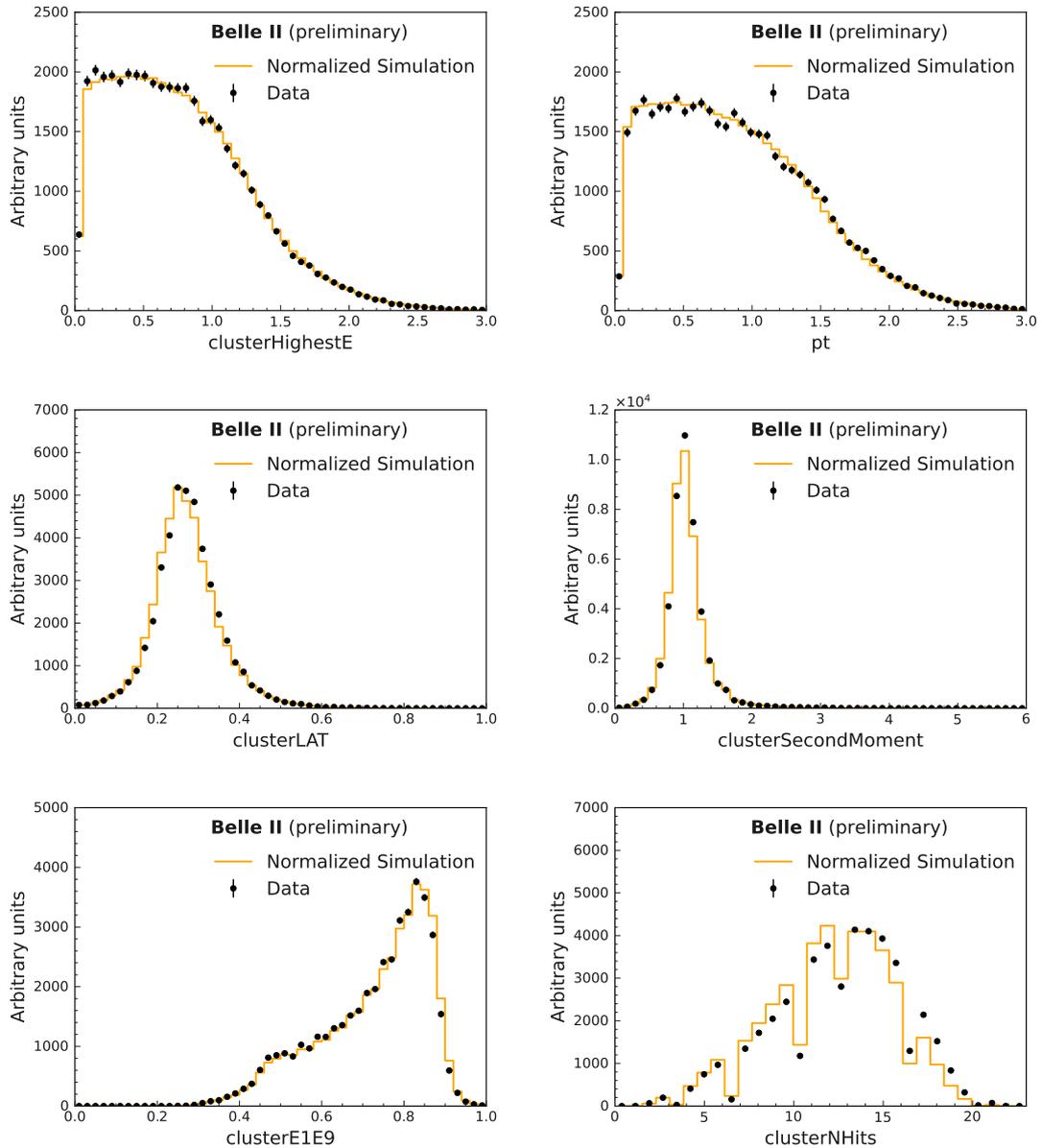


Figure 4.11: Comparison between photons reconstructed in reweighted $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0(\rightarrow \gamma\gamma))\pi^+$ decays (using the same photon and π^0 selections used in the signal channel) in (orange) simulation and (dots) data. Distributions normalized to each other (1/2).

impact on final precision is expected to be negligible due to the small fraction of misreconstructed and beam-background photons in the sample.

4.3.3.4 Photon classifier performance validation

To validate the photon classifier performance on data, I study the signal purity in $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0)\pi^+$ and $B^+ \rightarrow K^+\pi^0$ decays reconstructed in data after various selections on the photon classifier. In this test, π^0 's are reconstructed as in the signal $B^0 \rightarrow \pi^0\pi^0$ decay and their momenta weighted to match the signal π^0 momenta. For

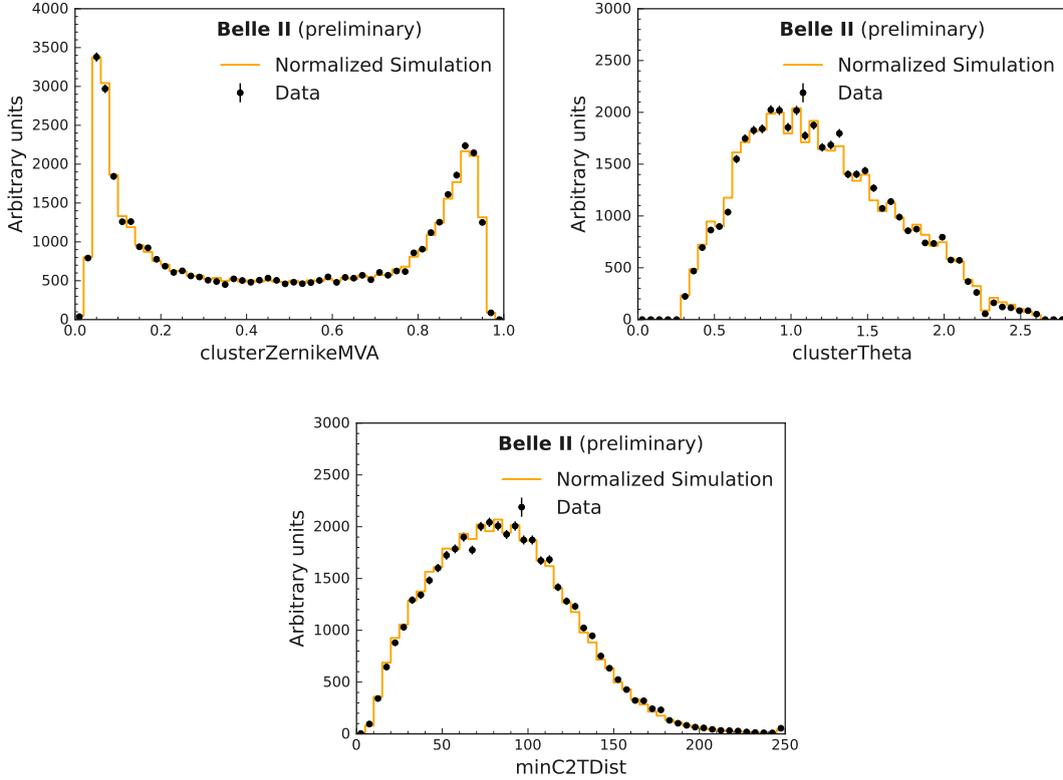


Figure 4.12: Comparison between photons reconstructed in reweighted $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0)\pi^+$ decays (using the same photon and π^0 selections used in the signal channel) in (orange) simulation and (dots) data. Distributions normalized to each other (2/2).

$D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0)\pi^+$, I first identify K^\mp and π^\pm candidates from charged-particle candidates reconstructed within the full CDC polar angle acceptance ($17^\circ < \theta < 150^\circ$), originating close to the interaction point in the longitudinal ($|dz| < 3.0$ cm) and radial ($|dr| < 0.5$ cm) directions, and associated with more than 20 CDC measurement points – to reduce beam-background. For D^0 reconstruction, I combine the K^- candidate with a π^+ and a π^0 candidates and require charmed meson candidates have masses in the range $1.82 - 1.90$ GeV/c^2 and center-of-mass momenta greater than 2.5 GeV/c . Finally, I reconstruct D^{*+} candidates by combining the D^0 and π^+ in a kinematic vertex fit. I randomly choose one candidate per event. The difference between the mass of the D^{*+} and D^0 , $\Delta M = m(K^-\pi^+\pi^0\pi^+) - m(K^-\pi^+\pi^0)$, is a powerful discriminator against background. I require $0.144 < \Delta M < 0.147$ GeV/c^2 to ensure the D^{*+} is reconstructed with high purity.

For $B^+ \rightarrow K^+\pi^0$ decays, kaon reconstruction mirrors that of the charm control channel. I combine kaons with π^0 's in a kinematic vertex fit to form B^+ candidates, and request them to satisfy $|\Delta E| < 0.3$ GeV and $M_{bc} > 5.27$ GeV/c^2

I then optimize individually the photon classifier selections in each control channel, and determine signal and background yields using fits to ΔM for the charm control channel and to ΔE for the $B^+ \rightarrow K^+\pi^0$ channel. Fits to 62.8 fb^{-1} of data for the charm control channel, and to 362.4 fb^{-1} of data for the $B^+ \rightarrow K^+\pi^0$ channel, are shown in Fig. 4.16 and 4.17. Signals in both cases are modelled with a Gaussian plus a Crystal Ball function, while the background is modelled with a polynomial function when fitting the charm decay,

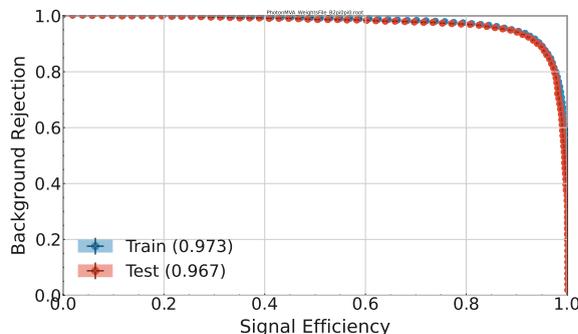


Figure 4.13: Receiver operating characteristic for the photon classifier for (blue) training and (red) testing simulated samples.

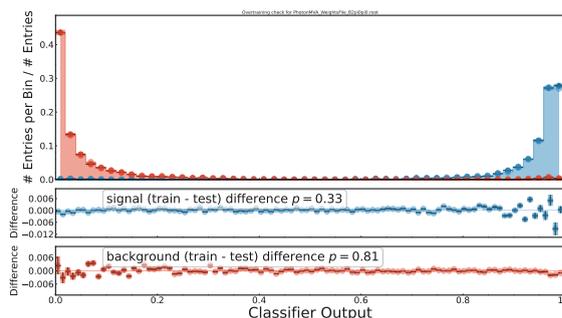


Figure 4.14: Output of the photon classifier on (solid) training and (circle) testing samples for simulated (red) background and (blue) signal events.

and with an exponential function when fitting the $B^+ \rightarrow K^+\pi^0$ decay. I fit data with and without the photon classifier selection applied, and compare the results (see Table 4.2). The background yield is reduced by the photon classifier selection of about 6 – 8%, depending on the channel, with a signal loss smaller than 2%. The impact of the photon classifier, although modest, is useful to partially suppress the high-level background of the $B^0 \rightarrow \pi^0\pi^0$ sample.

Decay channel	Photon BDT threshold	Signal yield	Background yield
$B^+ \rightarrow K^+\pi^0$	0.20	260 ± 30 (–0%)	680 ± 40 (–8.5%)
$D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0)\pi^+$	0.05	33422 ± 528 (–1.5%)	56066 ± 490 (–6.4%)

Table 4.2: Optimized photon classifier selection and corresponding signal and background yields in control samples. The fractional changes with respect to values observed without optimized selections are also shown.

4.3.3.5 Photon classifier validation and checks for biases

Even though individual validation of the simulated description of the each photon BDT input is successful, I also compare the photon BDT distribution of photons in $B^0 \rightarrow \pi^0\pi^0$ decays in simulation and data (see Fig. 4.18). No large disagreement is observed.

Since a sample-composition fit will follow the selection, any significant sculpting of observables used in the fit could complicate the next steps of the analysis and generate

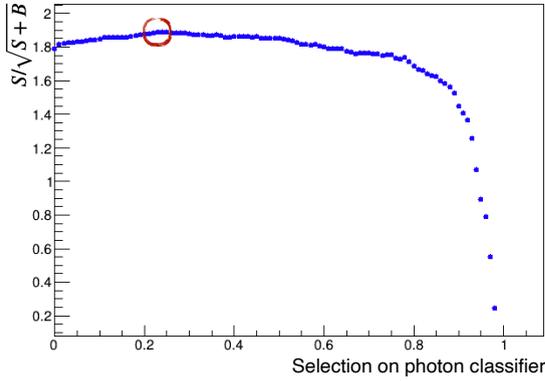


Figure 4.15: Ratio of signal yield over square root of signal-plus-background yields for simulated samples selected through various photon classifier criteria.

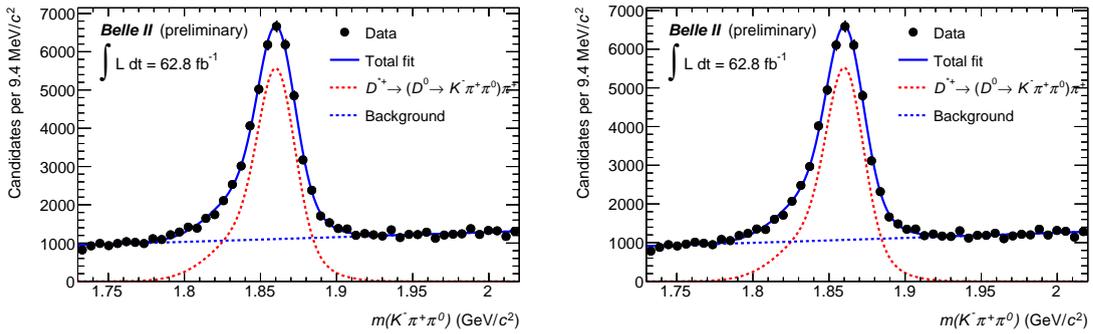


Figure 4.16: Distributions of $K^-\pi^+\pi^0$ mass for $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+\pi^0)\pi^+$ candidates reconstructed in a 62.8 fb^{-1} data sample with (left) no selection on the photon classifier and (right) an optimized selection on the photon classifier. Fit projections are overlaid.

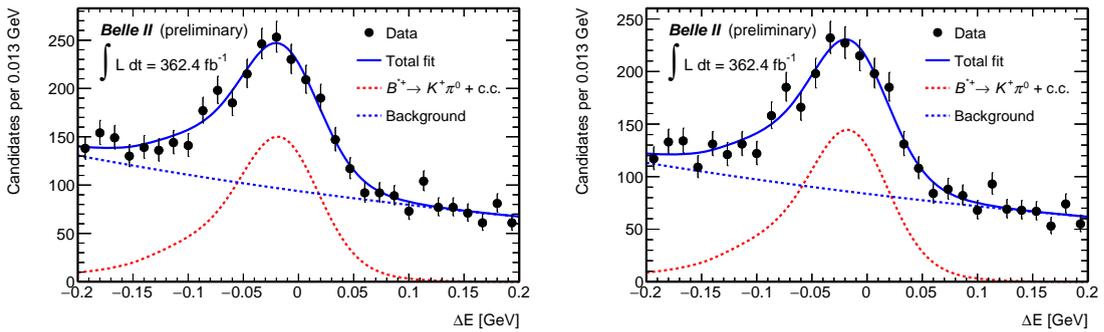


Figure 4.17: Distributions of ΔE for $B^+ \rightarrow K^+\pi^0$ candidates reconstructed in a 362.4 fb^{-1} sample of data with (left) no selection on the photon classifier and (right) an optimized selection on the photon classifier. Fit projections are overlaid.

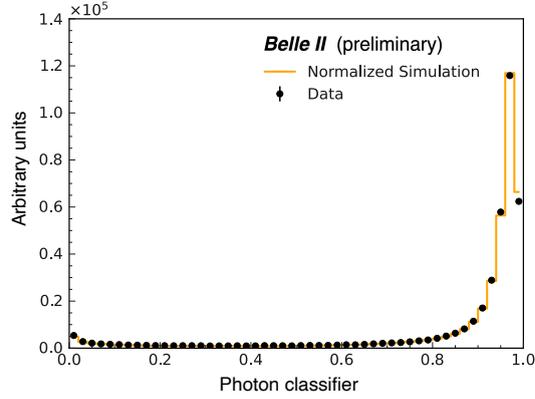


Figure 4.18: Photon selection classifier distribution for photons originating from the π^0 in the $B^0 \rightarrow \pi^0\pi^0$ decay for (dots) data and (histogram) simulation.

unwanted systematic uncertainties. I therefore check for possible sculpting of the fit observables M_{bc} , ΔE , output of the continuum suppression classifier (see Sec. 4.5.2), and output of the flavor tagger in the signal region (see Figs. 4.19–4.21) induced by the photon classifier selection. No sculpting is observed in any of the distributions, confirming that the use of the photon classifier will not complicate the subsequent fit.

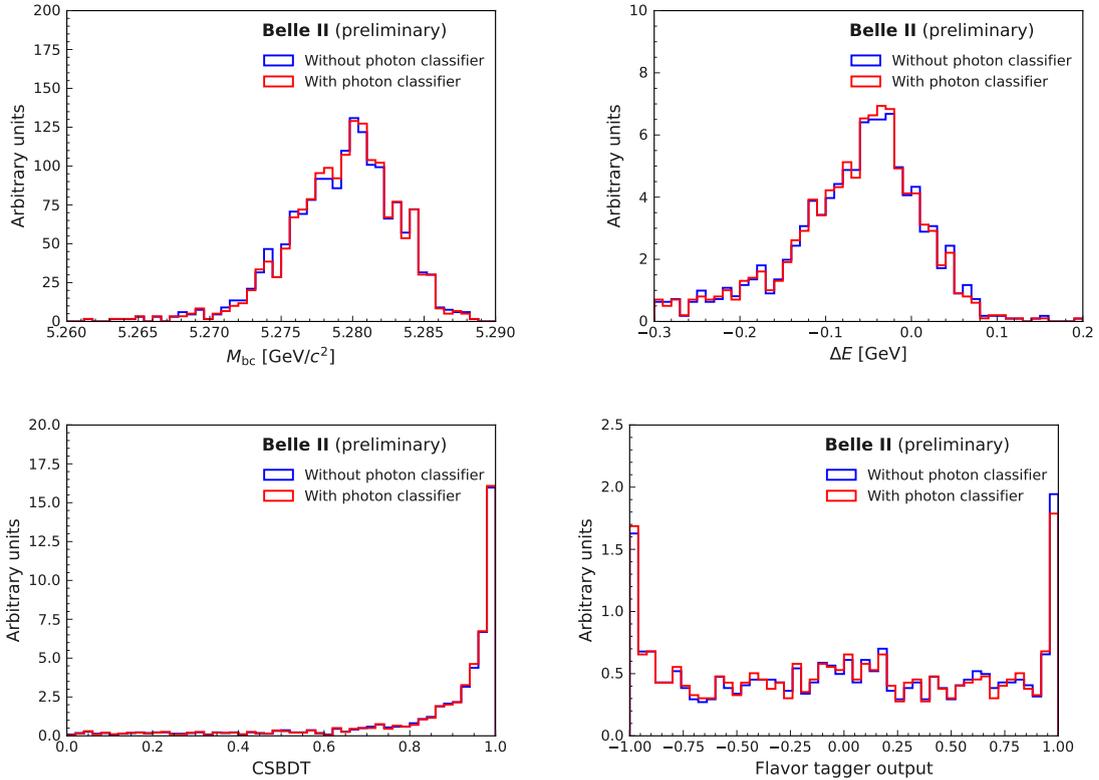


Figure 4.19: Distributions of (top left) M_{bc} , (top right) ΔE , (bottom left) output of the continuum suppression classifier and (bottom right) output of the flavor tagger for simulated signal, (red) with and (blue) without the optimized photon classifier selection applied. Distributions are normalized to each other. (1/3)

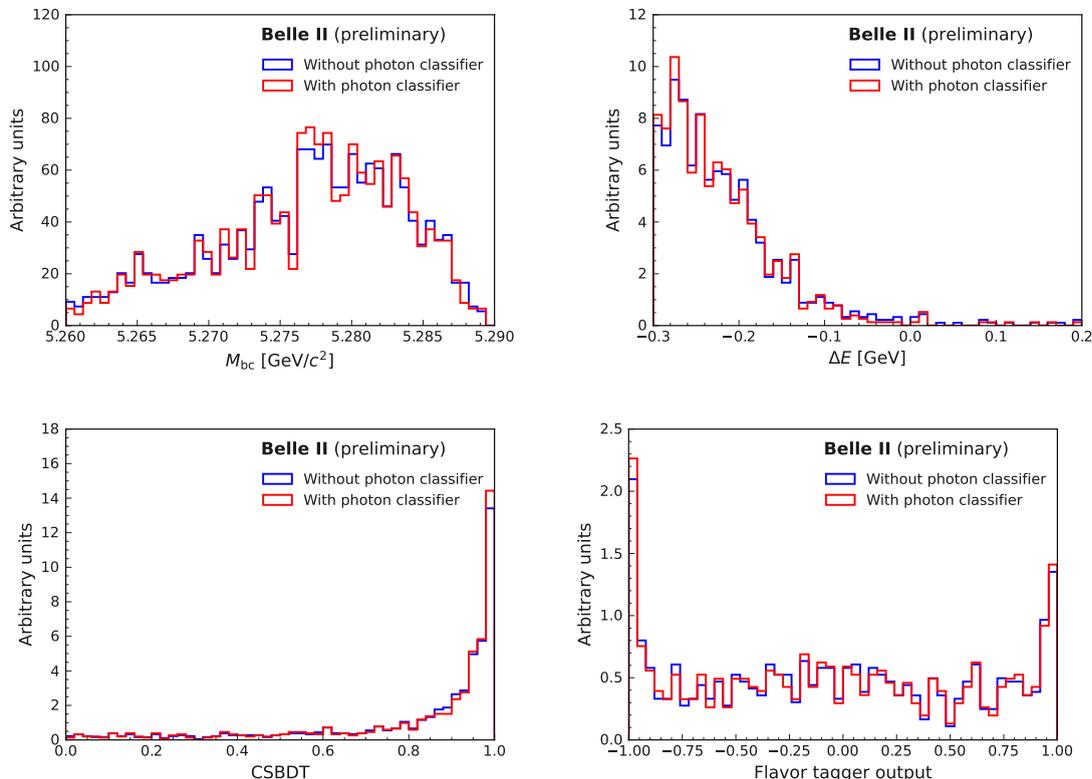


Figure 4.20: Distributions of (top left) M_{bc} , (top right) ΔE , (bottom left) output of the continuum suppression classifier and (bottom right) output of the flavor tagger for simulated $B\bar{B}$, (red) with and (blue) without the optimized photon classifier selection applied. Distributions are normalized to each other. (2/3)

4.4 Neutral pion selection and reconstruction

All pairs of photons that pass the baseline and optimized photon selections are used to form π^0 candidates. Table 4.3 shows the selections applied to the π^0 candidates.

π^0 meson selection	Signal efficiency (%)
daughterAngle < 0.4	54.18
daughterDiffOfPhi < 0.4	53.60
cosHelicityAngleMomentum < 0.98	53.52
p > 1.5 GeV/c	53.23
0.115 < InvM < 0.150 GeV/c ²	51.02

Table 4.3: Summary of π^0 selections and $B^0 \rightarrow \pi^0\pi^0$ simulated signal efficiency after each step.

The variable `daughterAngle` identifies the opening angle between the trajectories of the photons; `daughterDiffOfPhi` is the corresponding azimuthal opening angle; `cosHelicityAngleMomentum` is the cosine of the angle between the momentum difference of the two final-state photons in the B^0 rest frame and the boosted π^0 momentum; `p` is

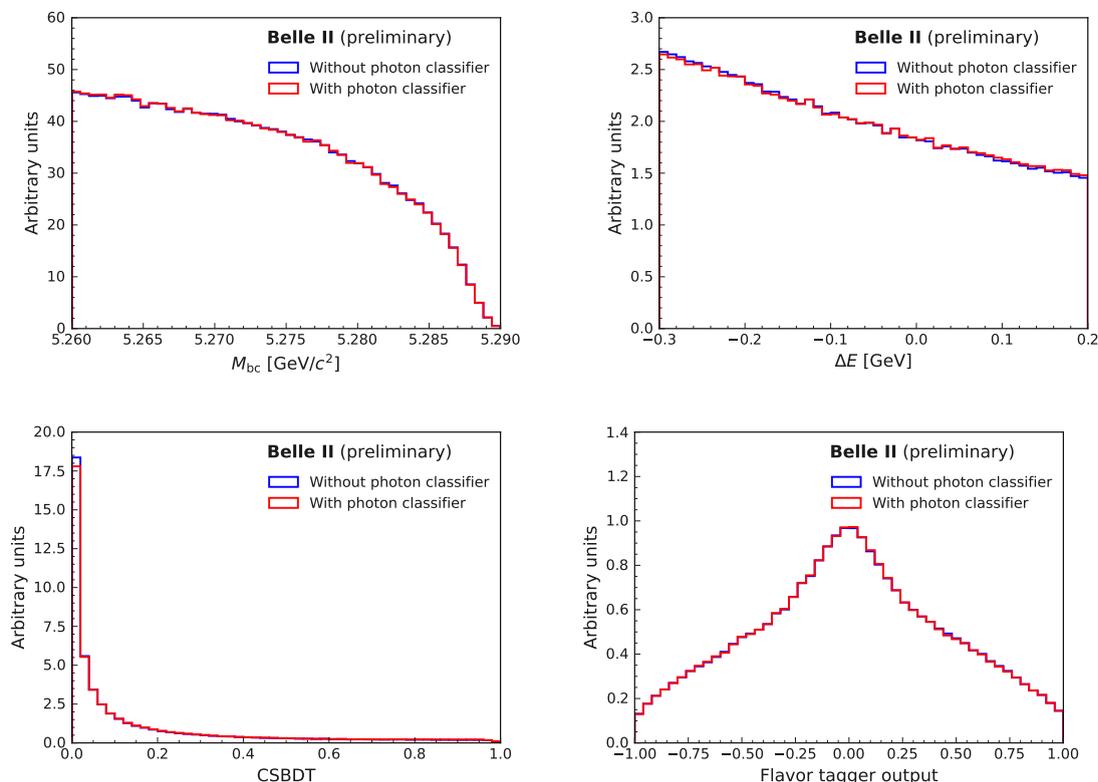


Figure 4.21: Distributions of (top left) M_{bc} , (top right) ΔE , (bottom left) output of the continuum suppression classifier and (bottom right) output of the flavor tagger for simulated continuum, (red) with and (blue) without the optimized photon classifier selection applied. Distributions are normalized to each other. (3/3)

the π^0 momentum; and InvM is the diphoton mass, which is restricted within two standard deviations of resolution around the known mass. All candidate π^0 are subjected to a mass-constrained fit, which imposes the value of the reconstructed π^0 meson mass to equal its known value [11], and corrects the photon energies accordingly. The relevant distributions are shown in Fig. 4.22 for real and misreconstructed π^0 candidates reconstructed in simulation and in Fig. 4.24 for π^0 candidates reconstructed in continuum-dominated data samples, defined as the sideband regions $M_{bc} < 5.26 \text{ GeV}/c^2$ or $0.2 < \Delta E < 0.5 \text{ GeV}$. Figure 4.23 shows the two-dimensional ΔE - M_{bc} distribution with highlighted the signal and sideband regions used for this check. From simulation, I expect the π^0 composition of the sample to be the same in signal and sideband regions.

All distributions reconstructed in sideband data shown in Fig. 4.24 resemble to the simulated distributions for real π^0 mesons of Fig. 4.22. This implies that the final sample has a small expected fraction of misreconstructed π^0 meson candidates (from simulation, less than 10%). This can be additionally seen in Fig. 4.25, where the diphoton mass distribution of candidates reconstructed in the data signal region is shown. A small fraction of flat, non- π^0 background is covered by a prominent peak of true π^0 mesons. Due to the small fraction of misreconstructed π^0 meson candidates, I do not optimize the π^0 selections. The impact of this choice on the final precision is expected to be negligible.

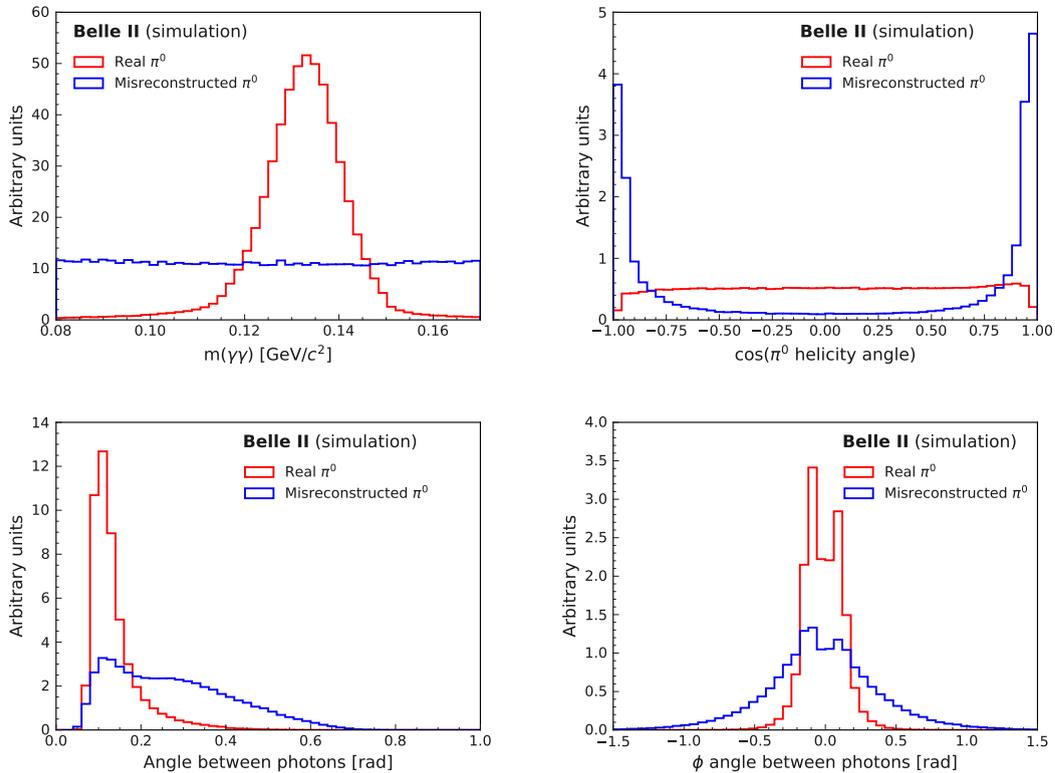


Figure 4.22: Comparison between distributions from simulated (red) real and (blue) misreconstructed π^0 candidates for (top left) mass, (top right) $|\cos\text{Helicity}|$, (bottom left) daughterAngle , and (bottom right) $|\text{daughterDiffOfPhi}|$. Distributions are normalized to each other.

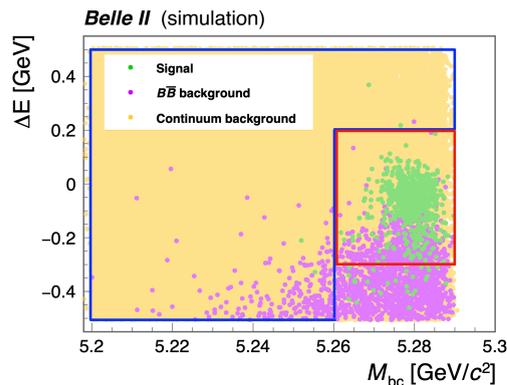


Figure 4.23: Two-dimensional ΔE - M_{bc} distribution of $B^0 \rightarrow \pi^0\pi^0$ candidates reconstructed in signal, $B\bar{B}$, and continuum simulation, with highlighted the (red) signal and (blue) sideband regions used for this check on the photon classifier.

4.5 B^0 selection and reconstruction

All π^0 pairs meeting the above selections are combined to form signal B^0 candidates. For correctly reconstructed B -meson candidates, ΔE should peak at zero. However, observed ΔE distributions peak at lower values, since energy is lost via either electromagnetic inter-

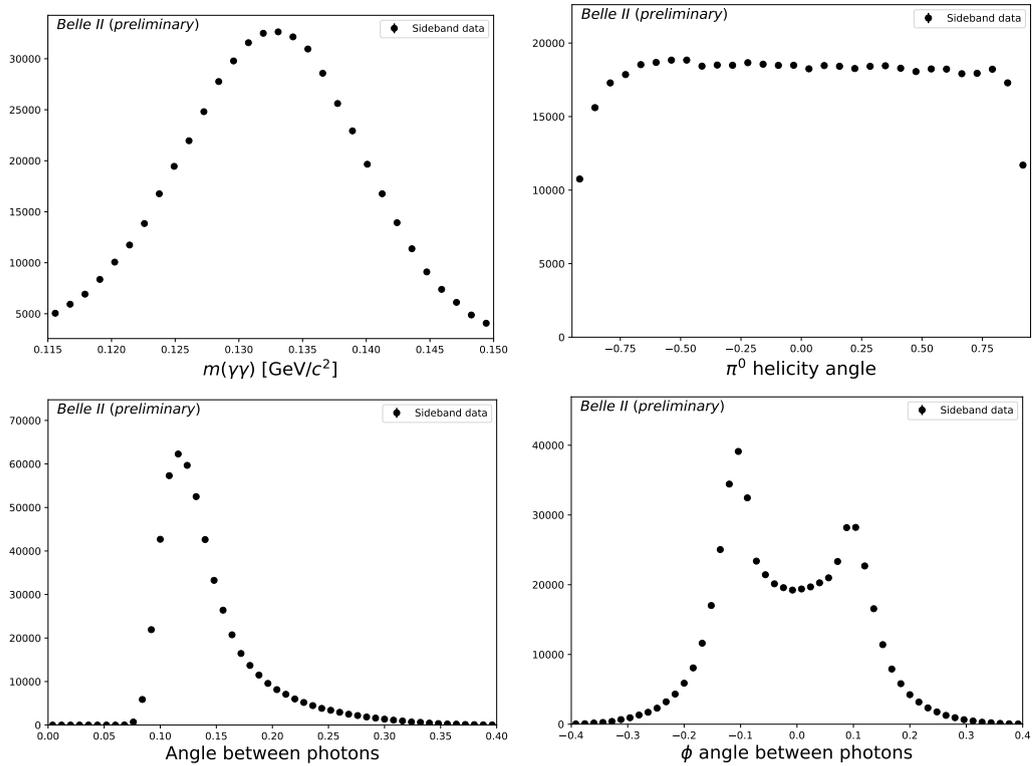


Figure 4.24: Distributions of (top left) mass, (top right) $|\cos\text{Helicity}|$, (bottom left) daughterAngle and (bottom right) $|\text{daughterDiffOfPhi}|$ for π^0 candidates reconstructed in continuum dominated data restricted to $M_{bc} < 5.26 \text{ GeV}/c^2$ or $0.2 < \Delta E < 0.5 \text{ GeV}$ (sidebands).

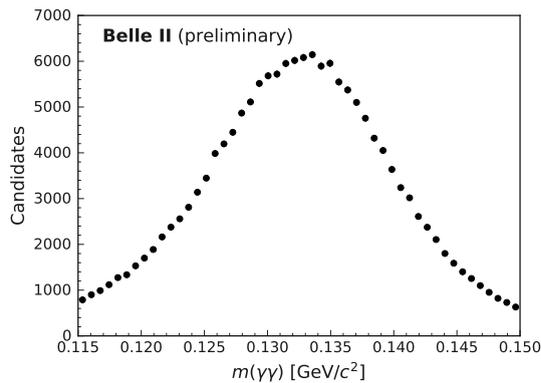


Figure 4.25: Distribution of the diphoton mass before the π^0 mass constraint for π^0 candidates reconstructed in the signal region in data.

actions in the material before the calorimeter or via energy leakage from the ECL cluster. The shower leaks into the surrounding crystals and may not be recovered. Continuum has a smooth distribution that decreases towards higher values of ΔE . For correctly reconstructed B -meson candidates M_{bc} peaks at the mass of the B meson at $5.28 \text{ GeV}/c^2$ while the continuum has a smooth distribution that rapidly decreases to zero at $5.28 \text{ GeV}/c^2$ (see Fig. 4.26).

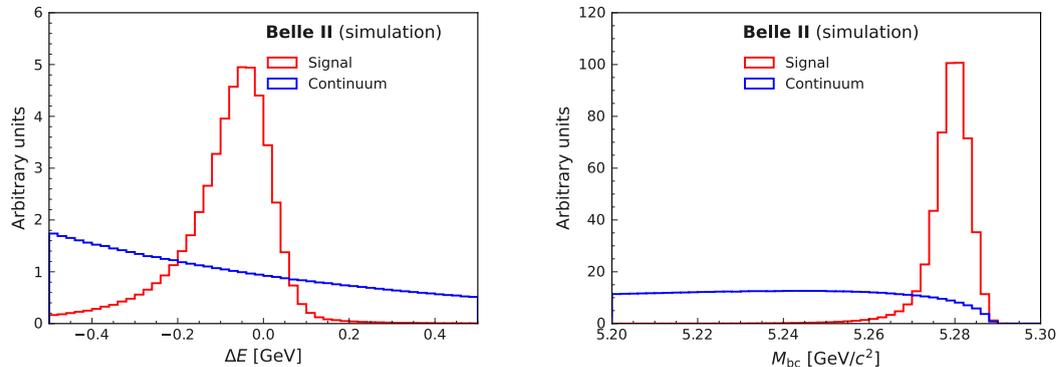


Figure 4.26: Normalized distribution of (left) ΔE and (right) M_{bc} reconstructed in simulated (red) signal and (blue) continuum decays.

4.5.1 B^0 baseline selection

B mesons candidates are required to meet a baseline selection listed in Table 4.4, where signal B^0 refers to correctly reconstructed $B^0 \rightarrow \pi^0\pi^0$ decays. The resulting sample after

Selection	Signal efficiency (%)
$-0.3 < \Delta E < 0.5$ GeV	43.7
$5.2 < M_{bc} < 5.29$ GeV/ c^2	43.7

Table 4.4: Summary of B^0 selection and associated absolute signal efficiency according to simulation.

this first selection is dominated (91%, according to simulation) by true, highly energetic π^0 mesons. However, only a 0.2% fraction of them comes from signal. The rest comes from direct production of light-pair of quarks and other B meson decays, as shown by the ΔE distribution in a realistically simulated data sample resulting from the baseline, photon, and π^0 meson selections (Fig. 4.27). No visible signal peak is observed at $\Delta E \approx 0$. Further background suppression is needed.

4.5.2 Continuum suppression

The chief challenge of the analysis is therefore to suppress background from real π^0 mesons produced in decays other than signal. About 43% of these is directly produced in the e^+e^- collision, with no intermediate particle involved. Another 25% comes from ρ^+ decays into $\pi^+\pi^0$ pairs.

To suppress continuum, I combine a set of event-shape variables known to provide statistical discrimination between signal and continuum in a binary boosted decision-tree classifier.

The discriminating variables are calculated using information from all particles not associated to the signal candidate, also called "rest of the event" (ROE), that pass a loose selection based on imposing $p < 3.2$ GeV/ c for charged particles and $0.5 < p < 3.2$ GeV/ c for photons. I choose a subset of 29 variables showing smaller-than-10% correlation with ΔE and M_{bc} , the principal observables used in the sample composition fit, to avoid sculpt-

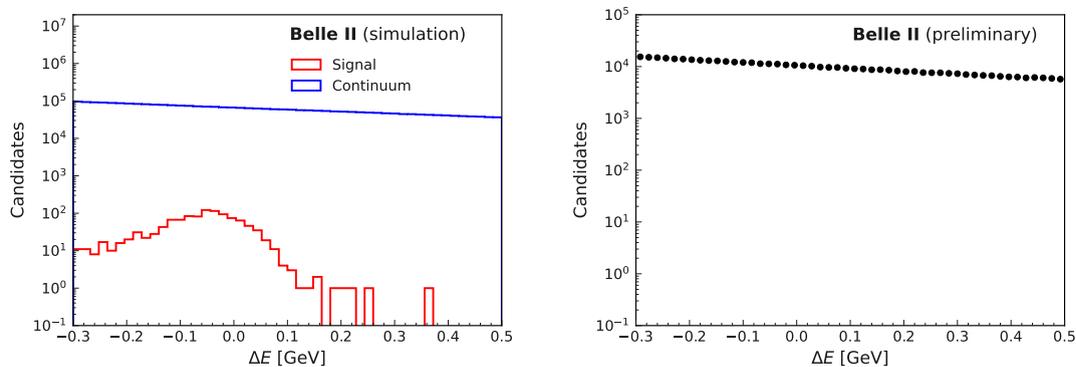


Figure 4.27: Distribution of ΔE in (left) realistically simulated data sample and (right) collision data, selected with the baseline, photon, and π^0 meson selections. The vertical axis is log-scaled.

ing that would complicate subsequent modeling. Many of these variables are introduced conceptually in Chapter 3.

- **R2** – ratio between the second and zeroth Fox-Wolfram moments. It is a quantity that captures the geometric features of the spatial distributions of final state particles to distinguish between the continuum collimated jets and the isotropically distributed final state particles from $B\bar{B}$ events;
- **cosTBT0** – cosine of the angle between the thrust axes of signal and ROE. For collimated jets, the cosine of the angle will peak at one, while for isotropically distributed decay products, it will have a uniform distribution;
- **thrust0m** – magnitude of ROE thrust. It is larger for strongly collimated continuum jets;
- **cosHelicityAngleMomentum** – cosine of the angle between the momentum difference of the two final-state photons and the boosted π^0 momentum;
- **thrustAxisCosTheta** – cosine of the polar angle component of the thrust axis, related to presence of a B meson produced from the $\Upsilon(4S)$ decay;
- **KSFVariables(mm2)** – missing mass squared, related to the rest of the event and to the presence of a B_{tag} ;
- 7 Kakuno-Super-Fox-Wolfram moments, related to the shape of the event and to the presence of a B_{tag} ;
- 1 CLEO cone constructed using only ROE particles to reduce correlation with ΔE and M_{bc} . It is related to the shape of the event;
- **useCMSFrame(cosTheta)** – polar angle of the signal- B candidate in the $\Upsilon(4S)$ frame. It is uniform for events where there is no B meson, and peaks at 0 for events where a B meson is produced in the $\Upsilon(4S)$ decay;
- **KSFVariables(et)** – transverse energy, related to the rest of the event and to the presence of a B_{tag} ;

- **DeltaZ** B_{tag} - longitudinal distance between the B_{Tag} and ROE B decay vertices. It is related to the presence of a B_{tag} in the ROE;
- **Deltar** B_{tag} - distance in the $x-y$ plane between the B_{Tag} and ROE B decay vertices. It is related to the presence of a B_{tag} in the ROE;
- **daughterAngle** - angle between the two π^0 mesons in the laboratory frame;
- **track_1_E** - energy of the most energetic charged particle in ROE. It is related to the presence of a B_{tag} in the ROE. Instead of using an exclusive reconstruction of the full event, we inclusively combine information on charged particles energy and PID to identify the B_{tag} mesons;
- **track_1_pt** - transverse momentum of the most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_1_muonID** - muon identification probability of the most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE. Semileptonic decays, which are the most probable B decays, have in the most cases the muon as the most energetic charged particle in the final state;
- **track_1_kaonID** - kaon identification probability of the most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_1_electronID** - electron identification probability of the most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_2_E** - energy of the second most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_2_pt** - transverse momentum of the second most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_2_muonID** - muon identification probability of the second most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_2_kaonID** - kaon identification probability of the second most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE;
- **track_2_electronID** - electron identification probability of the second most energetic charged particle in the ROE. It is related to the presence of a B_{tag} in the ROE.

Figures 4.28– 4.32 show the relevant distributions for signal simulation and off-resonance data. Figure 4.33 shows linear correlations between all the chosen inputs.

I train the continuum suppression algorithm using 1.2×10^3 truth-matched signal events and 1.2×10^3 off-resonance continuum events passing the baseline selection. I prefer off-resonance data to simulation for continuum because of large discrepancies observed in many event-shape variables. Due to the small number of off-resonance events passing the baseline selection, I do not split the off-resonance data sample into a training and a testing sample. I validate the classifier with a k -fold cross validation. I divide the sample in $k = 10$ equipopulated subsamples and perform k separated training-testing procedures: each combines nine samples into a training sample, and use the remaining subsample to test the resulting classifier. Then, I compare the resulting ROC curves obtained on the k testing samples in search for difference that could signal overtraining. Lacking those,

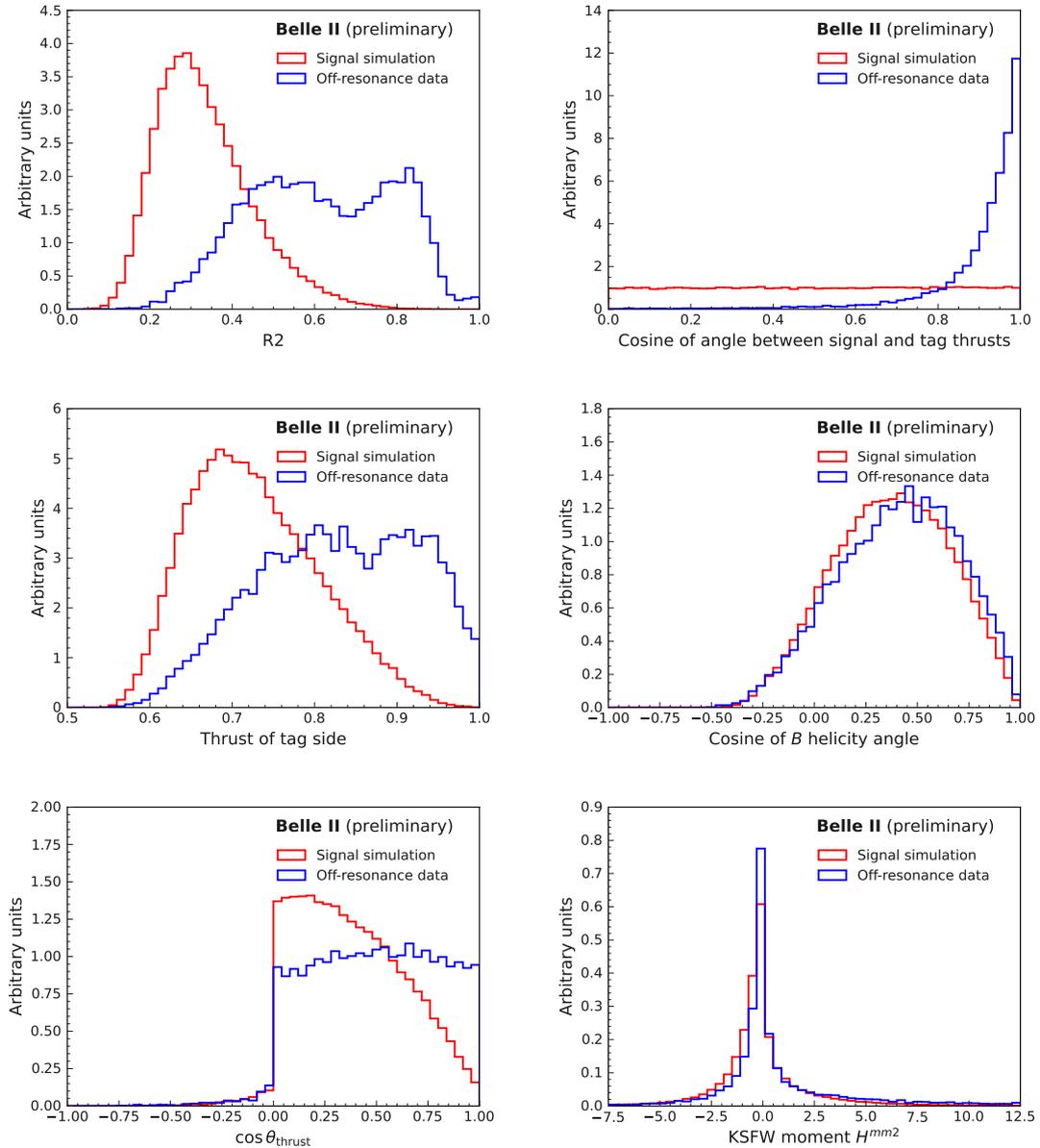


Figure 4.28: Comparison of continuum suppression inputs in (red) signal simulated data and (blue) off-resonance collision data for $B^0 \rightarrow \pi^0 \pi^0$ decays ($1/5$).

the training is performed employing the full data sample. Figure 4.34 shows the resulting ten ROC curves. No large difference is observed, reassuring us about lack of overtraining. Figure 4.35 shows the continuum-suppression output for signal and continuum-background events.

4.5.2.1 Optimization

I optimize the continuum suppression requirement by minimizing the expected average statistical variance on the asymmetry, as observed in ensembles of simplified simulated samples obtained by sampling the likelihood. I repeat the sample-composition fit in ensembles of simulated samples selected with continuum suppression thresholds scanning values from

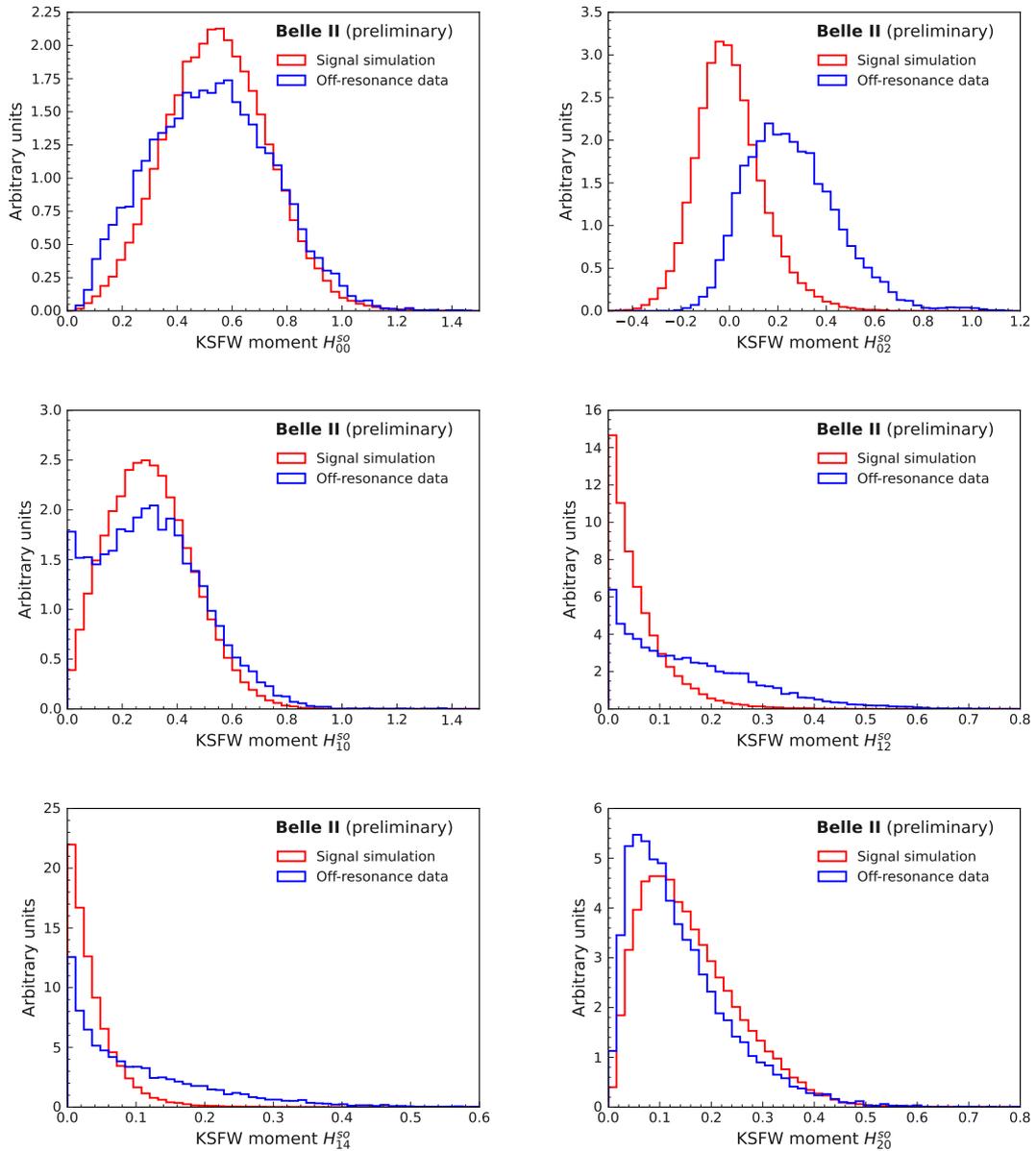


Figure 4.29: Comparison of continuum suppression inputs in (red) signal simulated data and (blue) off-resonance collision data for $B^0 \rightarrow \pi^0 \pi^0$ decays (2/5).

0.5 to 0.99 and choose the criterion that minimizes the average A_{CP} uncertainty in each ensemble. Each simulated sample contain signal, continuum, and $B\bar{B}$ components with expected proportions as shown in Fig. 5.1. I use probability integral transform to transform the continuum-background suppression output into a Gaussian shape to simplify fitting. Details on fitting are in Chapter 5.

For each continuum-suppression selection, I perform sample-composition fits to 1000 independent simulated data sets each corresponding to 362 fb^{-1} integrated luminosity. I record the mean of the signal-yield fractional uncertainty and the absolute A_{CP} uncertainty. The resulting average uncertainties as functions of the continuum-suppression criterion are shown in Fig. 4.36. The optimal values plateau at continuum-suppression thresholds of

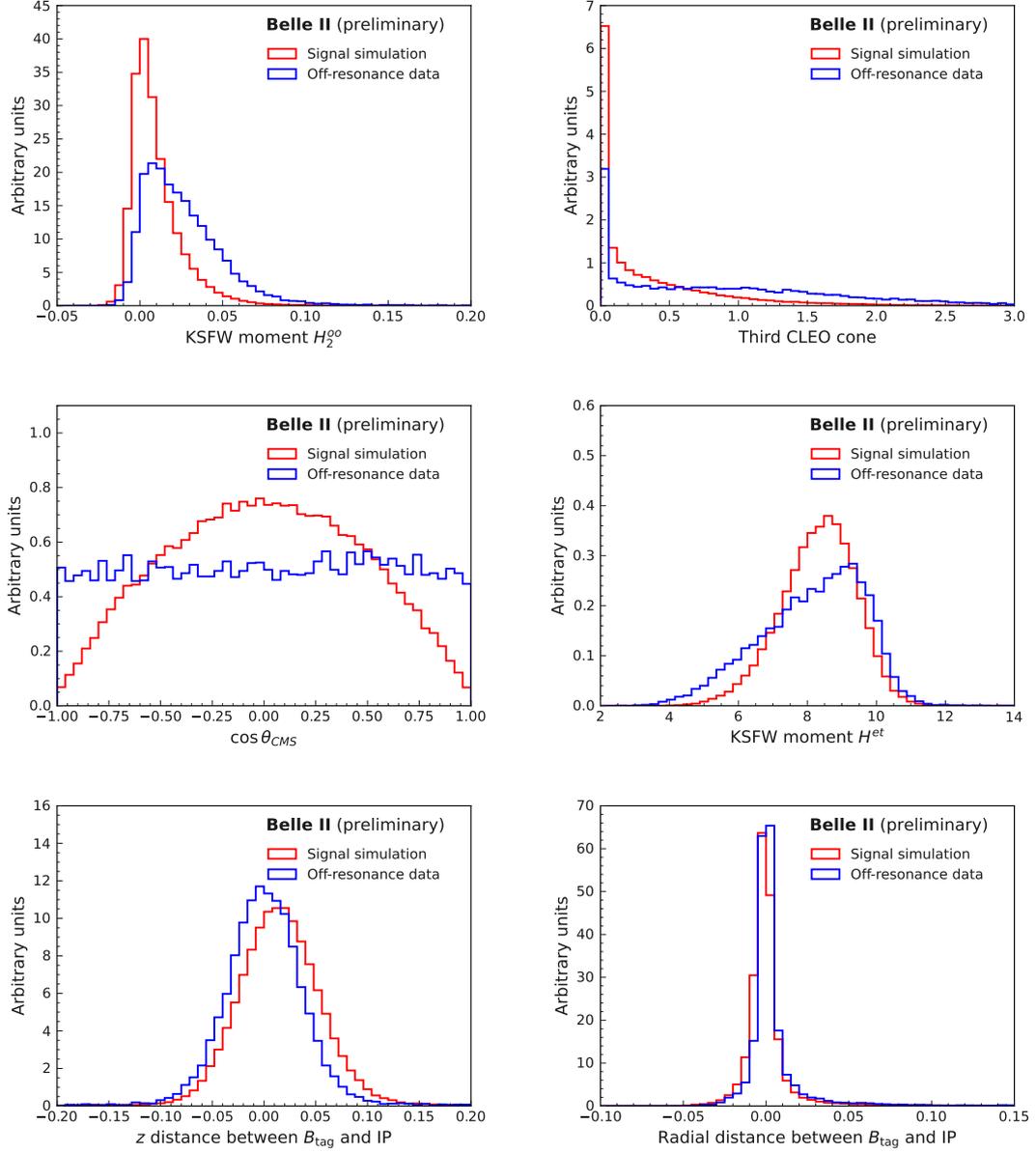


Figure 4.30: Comparison of continuum suppression inputs in (red) signal simulated data and (blue) off-resonance collision data for $B^0 \rightarrow \pi^0\pi^0$ decays (3/5).

around 0.5 extending to about 0.9, which offers the most stringent threshold that is still optimal. I choose the most stringent threshold, among optimal values, to minimize possible systematic uncertainties proportional to background contamination.

I also check for the possible generation of signal-like sculpting in the fit observables of background component, to avoid possible biases. Figure 4.37 shows the distributions of fit observables (see Chapter 5) M_{bc} , ΔE , and wrong-tag distributions of sideband data before and after the continuum background selection. Even if the shape gets modified by the selection due to dependencies, no peaking structure that could mimic signal emerges.

Using simulation, I obtain that the optimized selection removes 98% of background while keeping 64% of signal.

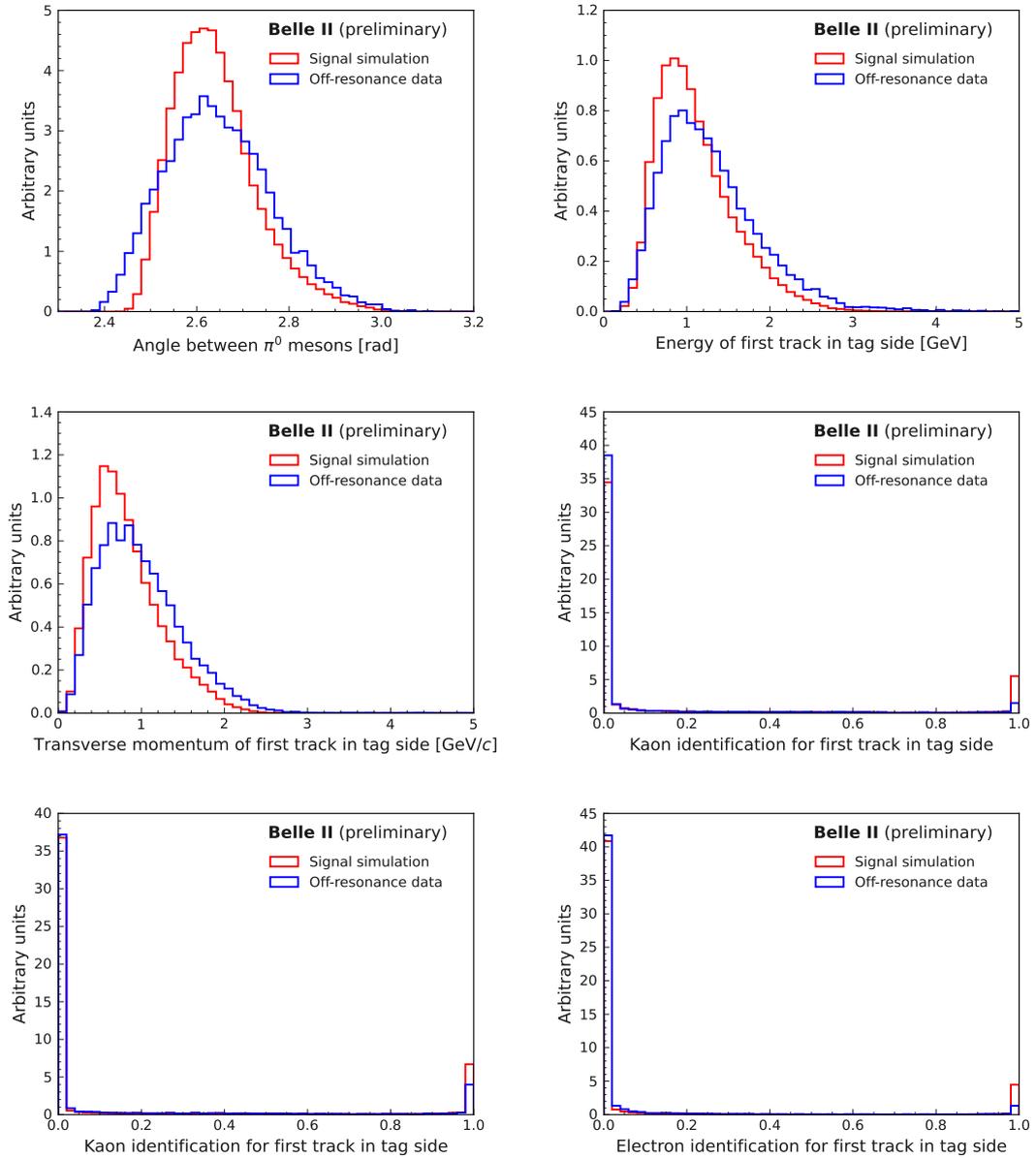


Figure 4.31: Comparison of continuum suppression inputs in (red) signal simulated data and (blue) off-resonance collision data for $B^0 \rightarrow \pi^0\pi^0$ decays (4/5).

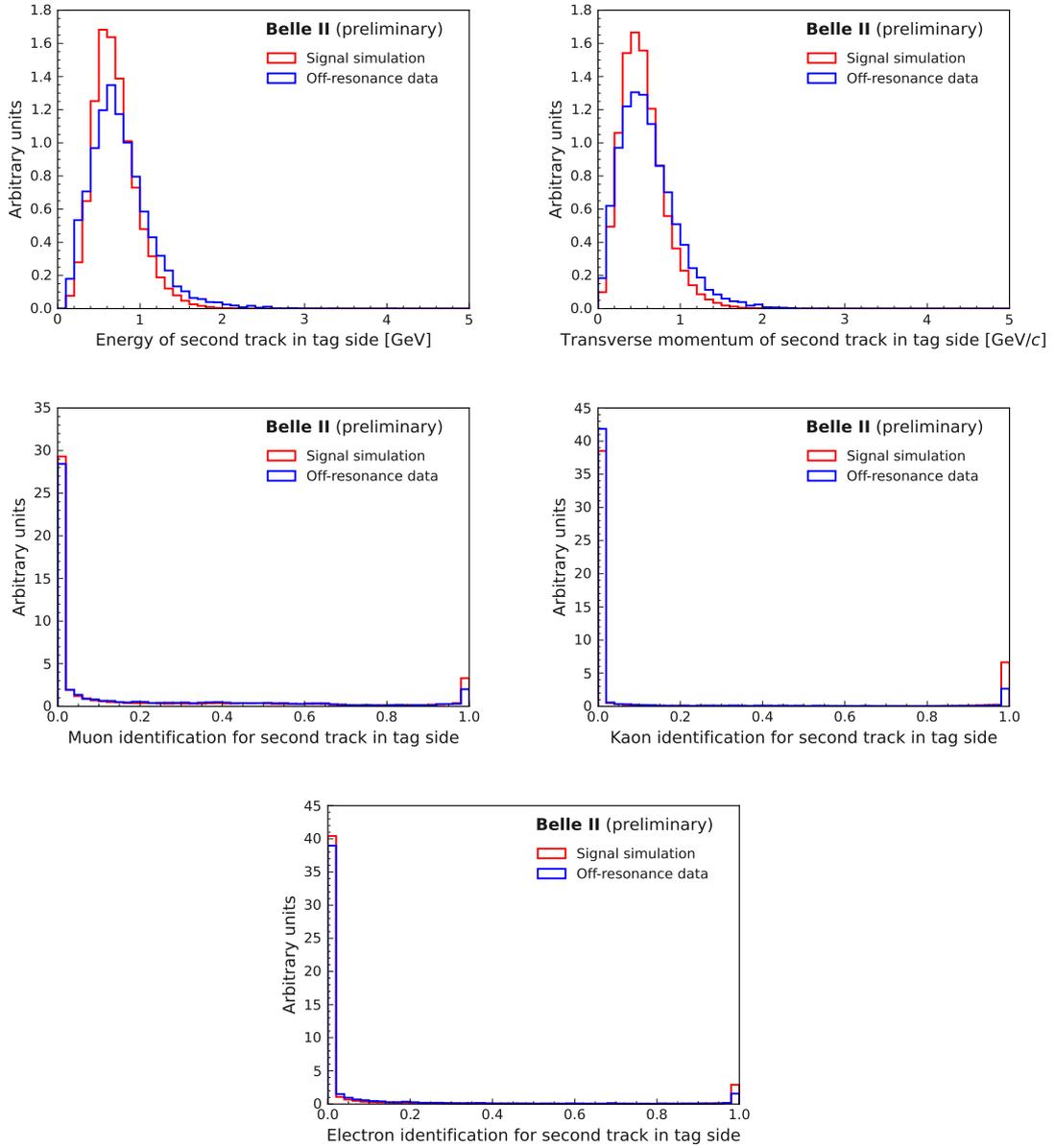


Figure 4.32: Comparison of continuum suppression inputs in (red) signal simulated data and (blue) off-resonance collision data for $B^0 \rightarrow \pi^0 \pi^0$ decays (5/5).

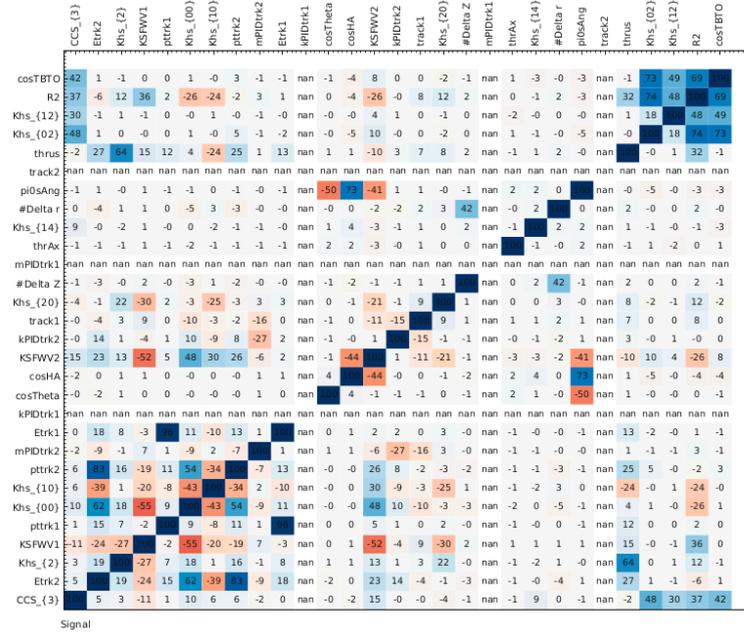


Figure 4.33: Linear correlations between the input variables of the continuum-suppression classifier for $B^0 \rightarrow \pi^0 \pi^0$ decays as resulting from simulation. Not all events have information on the B_{tag} side, hence the variables related to it (energy, PID of the B_{tag} tracks) can yield non numerical numbers and not be included in the classifier output for that event.

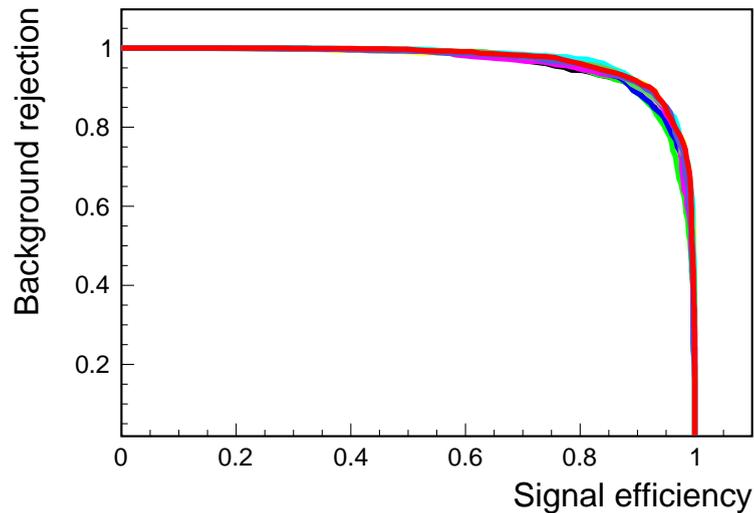


Figure 4.34: Receiver operating characteristic for ten different continuum-suppression classifiers obtained with the k -fold method.

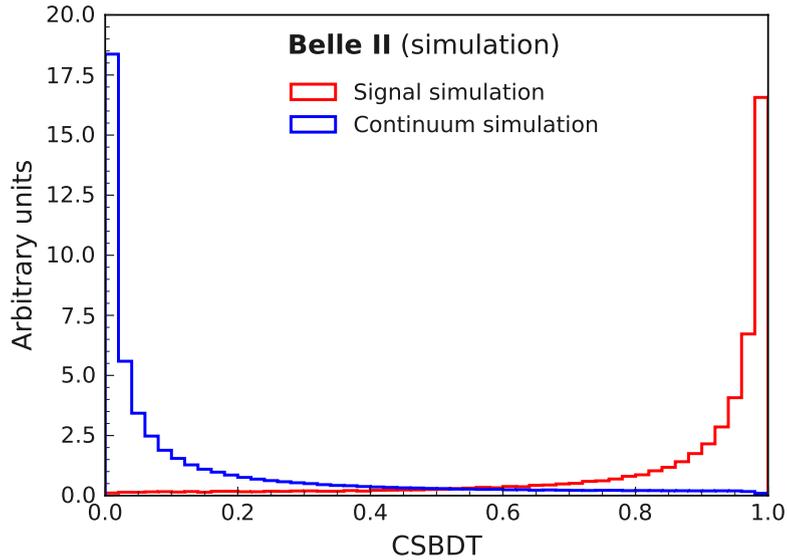


Figure 4.35: Output of the continuum-suppression classifier for (blue) continuum and (red) signal $B^0 \rightarrow \pi^0\pi^0$ events.

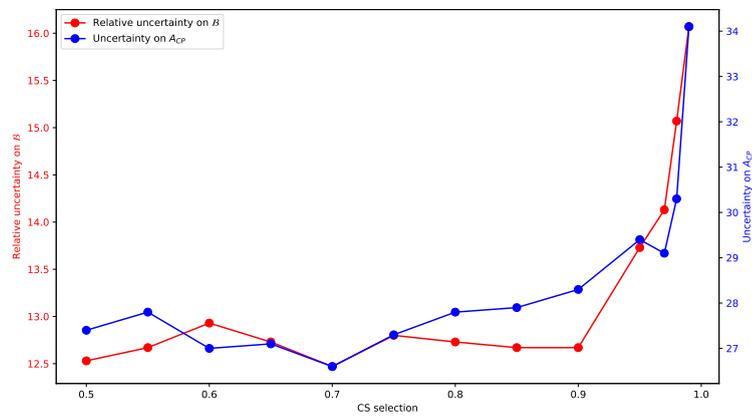


Figure 4.36: Average expected statistical uncertainty on \mathcal{B} and A_{CP} as a function of continuum suppression requirement as resulting from analyses of simplified simulated experiments.

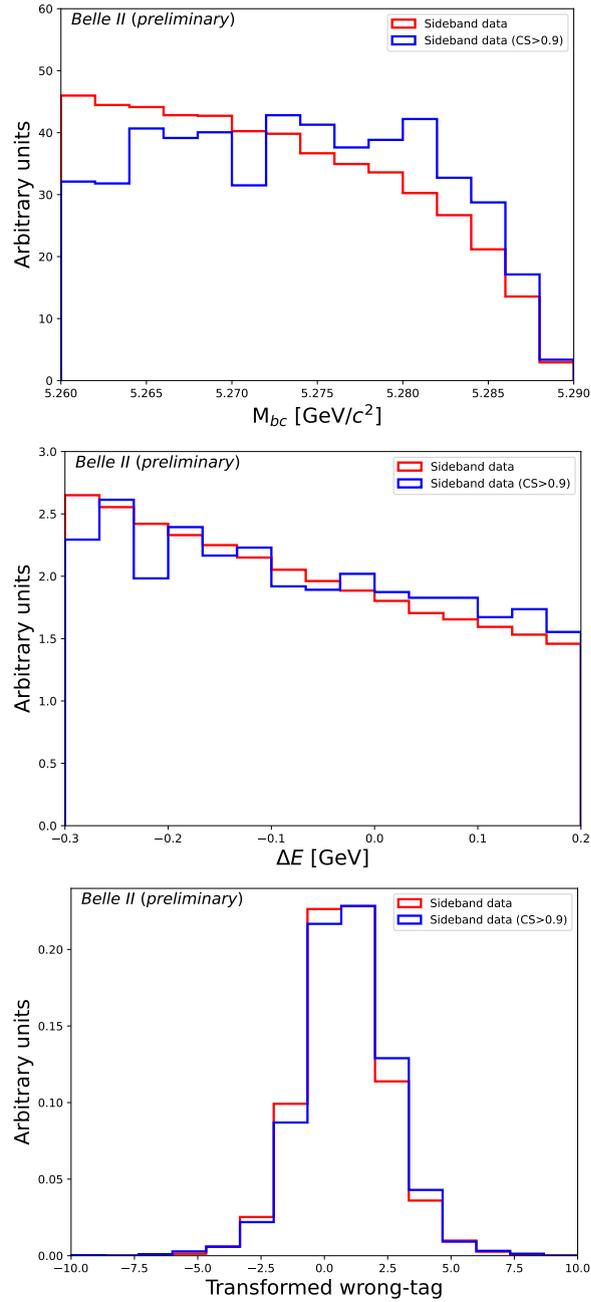


Figure 4.37: Distributions of the observables of the fit of composition (top) M_{bc} , (center) ΔE , and (bottom) wrong-tag fraction for $B^0 \rightarrow \pi^0 \pi^0$ candidates reconstructed in sideband collision data (red) before and (blue) after the optimized continuum suppression selection. Distributions are normalized to each other.

4.5.2.2 Dedicated ρ suppression

Continuum suppression is the main challenge of the analysis. In addition to the continuum suppression classifier, I explored another multivariate method specifically designed to suppress $e^+e^- \rightarrow X\rho^+(\rightarrow \pi^+\pi^0)$ decays, since 25% of the real π^0 mesons in the final sample comes from these decays and the presence of a resonant source might assist a more effective suppression. I form ρ^+ candidates by combining each candidate π^0 in the reconstructed event with each charged particle found in the rest of the event, assumed to be a pion.

Given the large ρ width, a veto on $m(\pi^+\pi^0)$ would remove a large fraction of signal events. Instead, I combine about ten variables related to the ρ meson kinematic and topological properties in a boosted decision tree in an attempt at identifying real ρ mesons, that should be discarded from the final sample. Due to the large number of ρ meson candidates for each event (about 20), I consider only the candidate classified as most-likely to be a ρ . After this choice, about 55% of the real ρ mesons are correctly identified by the BDT.

The resulting classifier output, called ρ meson-suppression BDT, is related to how likely a ρ meson is present in the event, and shows discrimination power between $B^0 \rightarrow \pi^0\pi^0$ events and background, as shown in Fig. 4.38 (left).

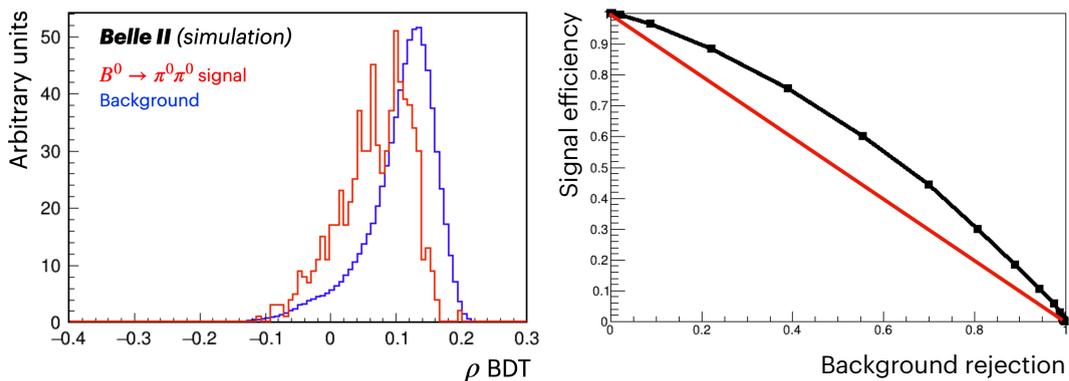


Figure 4.38: On the left, distribution of the ρ meson-suppression classifier for simulated (red) signal and (blue) background decays. On the right, receiver operating characteristic for the ρ meson BDT classifier. The red line corresponds to no discrimination between signal and background.

However, simulation shows that the discriminating power of this information is already captured by the general continuum suppression classifier, resulting in no significant improvement in final discrimination (Fig. 4.38 (right)). Hence, the dedicated ρ suppression is not included in the analysis, also for the possibly large associated systematic uncertainties caused by data-simulation discrepancies.

4.6 Candidate multiplicity and signal efficiency

Table 4.5 summarizes all the selections applied to the sample.

Object	Requirement
Photon	$E > 0.030 \text{ GeV}$
	$ \text{timing} < 200 \text{ ns}$
	Number of cluster hits > 1.5
	Photon classifier > 0.25
Neutral pion	$115 < m(\gamma\gamma) < 150 \text{ MeV}/c^2$
	$p > 1.5 \text{ GeV}/c$
	$ \cos \Theta_H < 0.98$
	Angle between photons < 0.4
	Angle between photons in ϕ plane < 0.4
	Converged vertex fit
B meson	$M_{bc} > 5.2 \text{ GeV}/c^2$
	$-0.3 < \Delta E < 0.5 \text{ GeV}$
Random candidate selection	-
Continuum suppression	$C_{\text{FBDT}} > 0.9$

Table 4.5: Summary of selection.

After these selections, the fraction of events with more than one candidate is 0.3% in data and 0.6% in simulation. Those events have, in average, two candidates. Multiplicity distributions are shown in Fig. 4.39. I restrict the samples to one candidate per event by randomly choosing it.

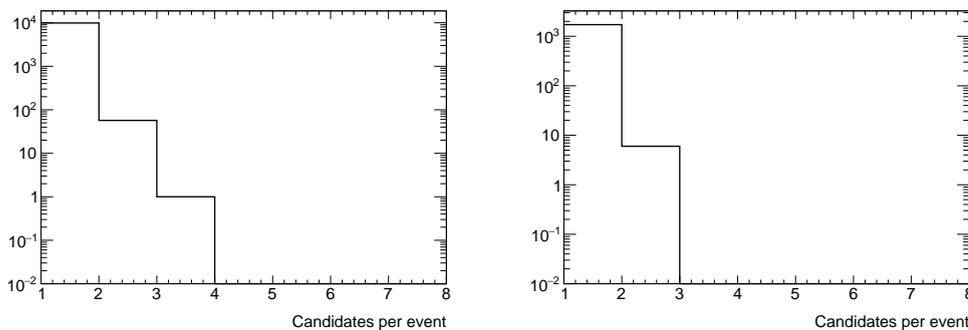


Figure 4.39: Distributions of candidate multiplicity in (left) simulated and (right) data $B^0 \rightarrow \pi^0 \pi^0$ candidates. The vertical axis is log-scaled.

Table 4.6 shows the absolute signal efficiency determined on signal-only simulation after the various reconstruction and selection steps. Data-simulation discrepancies that affect efficiencies are included in the systematic uncertainties based on control samples in data.

Table 4.6: Summary of absolute signal efficiency as determined in simulation.

	Signal efficiency
Reconstruction and preselection	63.40 %
Photon, π^0 , and B selections	43.65 %
CS requirements	27.80 %
Single candidate	27.62 %

4.7 Residual background

Two residual background components are not specifically targeted by any of the above selections and thus remain in the sample. Events containing genuine signal that are misreconstructed (self-cross feed) and $B\bar{B}$ events not yielding $B^0 \rightarrow \pi^0\pi^0$ decays. While a specific suppression of these can be attempted, we decide that their size and impact is sufficiently moderate to be properly addressed in the fit of composition.

4.7.1 Self-cross-feed background

Generated signal $B^0 \rightarrow \pi^0\pi^0$ events are classified into two categories after reconstruction: properly reconstructed ("true") $B^0 \rightarrow \pi^0\pi^0$ signal decays and misreconstructed $B^0 \rightarrow \pi^0\pi^0$ decays (*self-cross-feed*). Simulation shows that the ratio of self-cross-feed to signal candidates is about 3%. Given its small fraction, and the large expected statistical uncertainty on the final results, self-cross-feed is included in the signal in the further steps of the analysis.

4.7.2 $B\bar{B}$ backgrounds

Charmless B decays with same or similar final states as the $B^0 \rightarrow \pi^0\pi^0$ signal tend to peak under, or in proximity, of the $B^0 \rightarrow \pi^0\pi^0$ peaks in the distributions of the ΔE and M_{bc} observables. Such $B\bar{B}$ backgrounds are included and modeled in the fit of sample composition.

Generic $e^+e^- \rightarrow B\bar{B}$ simulated samples (see Sec. 4.1.2) provide information on the composition of $B\bar{B}$ backgrounds (Table 4.7). The most prominent decays are misreconstructed $B^+ \rightarrow \rho^+(\rightarrow \pi^+\pi^0)\pi^0$, where the π^+ is not reconstructed (81% of the total $B\bar{B}$ background). The second most important mode is $B^0 \rightarrow K_S^0(\rightarrow \pi^0\pi^0)\pi^0$, where a neutral meson is not associated to the candidate (3%). All other modes have contributions smaller than 1%.

Channel	Fraction
$B^+ \rightarrow \rho^+(\rightarrow \pi^+\pi^0)\pi^0$	81%
$B^0 \rightarrow K_S^0(\rightarrow \pi^0\pi^0)\pi^0$	3%
Others (< 1% each)	16%

 Table 4.7: $B\bar{B}$ backgrounds contaminating the $B^0 \rightarrow \pi^0\pi^0$ sample according to simulation.

4.8 Reconstruction of the tag B meson

Partial information from the pair-produced nonsignal B meson in the event is used in this analysis to suppress continuum background and, most importantly, to statistically determine the flavor of the signal by means of the flavor tagging algorithm. The flavor tagger does not aim at fully reconstructing the tag B meson, but uses the information from all charged particles not assigned to the signal candidate to infer its flavor. A baseline selection requires charged particles to be associated to a minimum number of CDC measurements points, be reconstructed within the full CDC acceptance and point back to the interaction point within loose requirements. Those tracks are then associated with particle-identification information and fed to a dynamic graph convolutional neural network trained in simulation to infer flavor by identifying relational patterns among its inputs, in addition to identifying the inputs as conventional taggers do. This results in a flavor tagging performance, that is the effective fraction of events correctly tagged, of 37%, which is equivalent to a 15% net increase in data sample size for tagged decays compared with the 32% tagging performance of the standard Belle II tagger. The algorithm provides a single output, the product qr of the tag meson flavor q and the dilution $r = 1 - 2w$, where w is the wrong-tag probability, for use in analysis. Introduction of this new tagger, and inclusion of its unbinned information in the fit, is an innovation of this analysis, which is only the second Belle II analysis in using it.

Chapter 5

Determination of the sample composition

The reconstructed sample of signal candidates is expected to contain a prominent fraction of background events. The individual sample components need to be separated for extracting the desired measurements of $B^0 \rightarrow \pi^0\pi^0$ properties. In this chapter, I describe how kinematic, topological, and flavor information is combined in a maximum likelihood fit to determine sample composition and therefore the signal branching fraction and CP-violating asymmetry.

5.1 The fit

The composition of the $B^0 \rightarrow \pi^0\pi^0$ sample is determined statistically with a multivariate maximum likelihood fit of the unbinned distributions of kinematic, topological, and flavor tagger discriminating observables. I write a likelihood $\mathcal{L}(\vec{\theta}|\vec{x})$, function of the n unknown parameters of the model $\vec{\theta} = (\theta_1, \dots, \theta_n)$ at the given set \vec{x} of observed data. For any specific set of values of the unknown parameters θ , \mathcal{L} is the joint probability density for obtaining the observed values of the discriminating observables in the sample, \vec{x} . In fitting, I maximize the likelihood function with respect to the unknown parameters by minimizing its negative natural logarithm by imposing

$$-2 \frac{\partial(\ln\mathcal{L})}{\partial\vec{\theta}} = 0. \quad (5.1)$$

5.1.1 Minimizer

I use the computer program MINUIT [94], which calculates numerically Eq. (5.1). In a first fit, I use the MIGRAD minimization algorithm, which features a variable-metric method with inexact line search, a stable metric updating scheme, and checks for positive-definiteness to identify coarsely the maximum. Then I use MINOS, which uses the likelihood ratio to account for non-linearities in the uncertainty calculation, therefore providing more accurate uncertainties.

5.1.2 Fit components

The sample of $B^0 \rightarrow \pi^0\pi^0$ candidates resulting from the final selection receives contributions from three main sources, signal, continuum background, and backgrounds from nonsignal $B\bar{B}$ events. Self-cross-feed is included in the signal component. Continuum

background dominates the sample composition, as expected from the simulation and shown in Fig. 5.1.

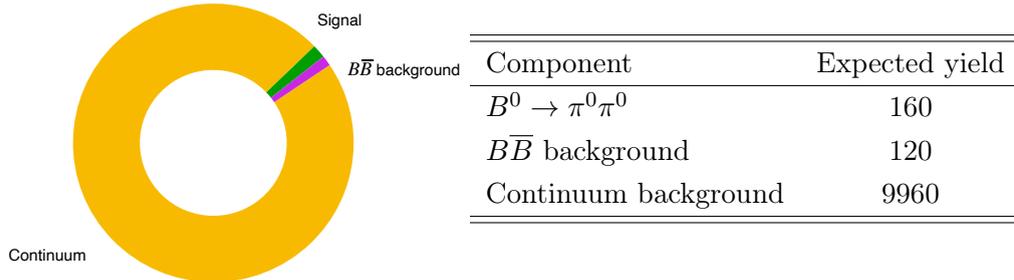


Figure 5.1: Summary of $B^0 \rightarrow \pi^0 \pi^0$ sample composition resulting from the final selection expected from simulation in the full Belle II data sample.

5.1.3 Fit observables

I choose a set of discriminating observables that allow distinguishing signal decays from background, and B^0 from \bar{B}^0 decays within signal. The goal is to achieve a balance between effective signal-to-background separation and reasonable fit complexity both in terms of number of observables and of their multidimensional dependences.

The energy difference ΔE is the most powerful observable to separate exclusively reconstructed signal from continuum events (Fig. 5.2 top left). It also has the advantage to discriminate signal from other misreconstructed B decays. The beam-constrained mass M_{bc} is also discriminating between $B\bar{B}$ and continuum events (Fig. 5.2 top right), since $B\bar{B}$ events cluster in a narrow peak near the B mass, while continuum has a smooth, nearly uniform distribution.

Another important discriminating observable against continuum background is the C_{FBDT} output of the continuum-background classifier (Fig. 5.2 bottom left). Its distribution has a nontrivial shape, so I transform it into $C'_{\text{FBDT}} = \log \frac{C_{\text{FBDT}} - C_{\text{FBDT}_{\min}}}{C_{\text{FBDT}_{\max}} - C_{\text{FBDT}}}$, which is described by an analytical model straightforwardly without degrading discriminating power. In what follows, we refer as continuum-suppression output to the transformed and untransformed observables indistinguishably for simplicity.

To measure A_{CP} I use the flavor-tagger output, qr , that provides for each event statistical information about the signal flavor. It is convenient to use the observable w , that is the probability for an event to be assigned a wrong flavor tag, referred to as “wrong-tag” in the following. Similarly to C_{FBDT} , I transform w into $w' = \log \frac{w}{0.5-w}$. This observable is included in the likelihood and provides some further signal-background separation in addition to flavor information (Fig. 5.2 bottom right).

Figure 5.3 shows the same distributions for the different components normalized to one, to emphasize the signal-background separation.

In summary, the fit uses four discriminating observables:

1. ΔE – difference between observed and expected B energy;
2. M_{bc} – beam-energy-constrained mass;

3. C'_{FBDT} – log-transformed continuum suppression classifier output;
4. w' – log-transformed wrong-tag fraction.

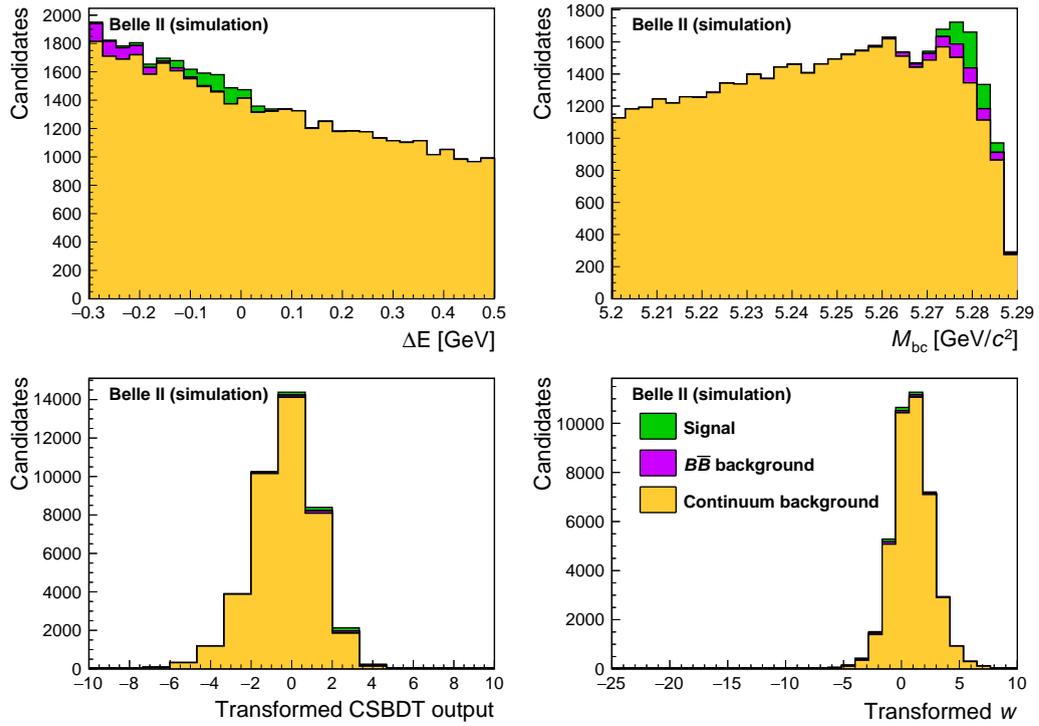


Figure 5.2: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) continuum-suppression output, and (bottom right) wrong-tag w for simulated $B^0 \rightarrow \pi^0 \pi^0$ components. Proportions mirror realistically expected yields.

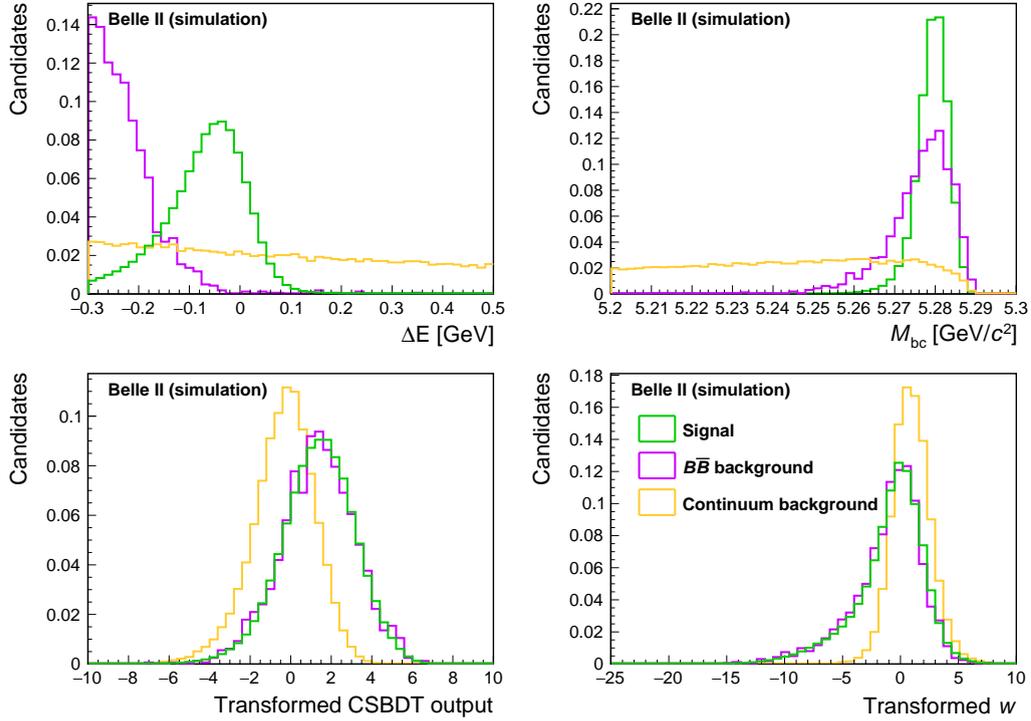


Figure 5.3: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) continuum-suppression output, and (bottom right) w for simulated $B^0 \rightarrow \pi^0 \pi^0$ components. All distributions are normalized to one.

5.1.4 Fit parameters

For convenience, I parametrize the likelihood directly in terms of the $B^0 \rightarrow \pi^0 \pi^0$ parameters of interest, branching fraction and A_{CP} , so that they are directly determined by the fit, along with their statistical uncertainties including correlations.

The branching fraction \mathcal{B} for a B decay into final state f is expressed as the ratio of the observed signal yield N , corrected by reconstruction and selection efficiency ε and normalized to the total number of collected $B\bar{B}$ pairs $N_{B\bar{B}}$, in turn corrected for the difference between the $B^+ B^-$ and $B^0 \bar{B}^0$ pair production rates,

$$\mathcal{B}(B \rightarrow f) = \frac{N(1 + f^{+-}/f^{00})}{2\varepsilon N_{B\bar{B}}}. \quad (5.2)$$

At the time of this writing Belle II has collected 387×10^6 $B\bar{B}$ pairs. The number of $B\bar{B}$ pairs is determined by counting the total number of events at the $\Upsilon(4S)$ energy (on-resonance). The same counting is performed also in off-resonance collisions, and then scaled to the on-resonance integrated luminosity. The number of $B\bar{B}$ pairs is obtained from the difference between these two counts. The term $f^{+-}/f^{00} = 1.065 \pm 0.012 \pm 0.019 \pm 0.047$ [95] is the ratio of charged to neutral $B\bar{B}$ pair production. The total number of $B^0 \bar{B}^0$ pairs used in this analysis is around 180×10^6 .

The CP -violating asymmetry A_{CP} for a B decay into final state f is expressed as the asymmetry between widths of $\bar{B} \rightarrow \bar{f}$ and $B \rightarrow f$ decays,

$$A_{CP} = \frac{\Gamma(\bar{B} \rightarrow \bar{f}) - \Gamma(B \rightarrow f)}{\Gamma(\bar{B} \rightarrow \bar{f}) + \Gamma(B \rightarrow f)} = \frac{N(\bar{B} \rightarrow \bar{f}) - N(B \rightarrow f)}{N(\bar{B} \rightarrow \bar{f}) + N(B \rightarrow f)}, \quad (5.3)$$

where the approximation holds since our final state is the same for B^0 and \bar{B}^0 , and therefore instrumental effects are only associated to tagging asymmetries.

In addition to the parameters of interest, the likelihood depends on a set of nuisance parameters whose values are not necessarily of interest, but influence the values of parameters of interest. Examples of nuisance parameters include the yields of the background components, the effective background asymmetries, etc.

5.1.5 Model

Distributions for some of the observables are known based on physics-motivated models. For instance, we know that the signal ΔE distribution is approximately Gaussian with a broad left tail due to energy leakage in the calorimeter. For other observables, excessive complexity prevents from identifying physics-motivated models. In those cases, I empirically find models that are capable of reproducing adequately the distribution shapes and include the effects of these approximation in the systematic uncertainty. For each component, shapes are extracted from the corresponding simulated samples or sideband in experimental data.

5.2 Likelihood modeling

The likelihood function \mathcal{L} is written as a product, over the N events in the sample assumed to be independent and identically distributed, of the single-event likelihoods \mathcal{L}_i

$$\mathcal{L}(\vec{\theta}|\vec{x}) \equiv \mathcal{L}(\vec{\theta}) = \prod_{i=1}^N \mathcal{L}_i(\vec{\theta}). \quad (5.4)$$

Each single event likelihood has the following expression

$$\mathcal{L}_i = N_{\text{tot}} \left[f_{\text{sig}}(1 + q(1 - 2\chi_d)(1 - 2w)A_{CP}^{\text{sig}})p_{\text{sig},i} + f_{B\bar{B}}p_{B\bar{B},i} + (1 - f_{\text{sig}} - f_{B\bar{B}})p_{\text{cont},i} \right], \quad (5.5)$$

where N_{tot} is the total number of observed candidates in the sample, and is fixed, each f_j is the fraction of the generic component j in the sample, and $p_{j,i}$ indicates the four-dimensional probability density function. The parameter χ_d is the time-integrated $B^0 - \bar{B}^0$ mixing probability, defined in terms of the mass difference Δm_d between the two B^0 mass eigenstates (light-heavy) and the B^0 lifetime τ_{B^0} as

$$\chi_d = \frac{1}{2} - \frac{1}{2(1 + (\tau_{B^0}\Delta m_d)^2)},$$

which accounts for the fraction of signal B that may have undergone flavor oscillation after production thus decaying with a different flavor that they had at production. The current known value is $\chi_d = 0.1858 \pm 0.0011$ [11].

I extract central values and uncertainties of branching ratio and CP -violating asymmetry from a non-extended unbinned fraction fit. I then add in quadrature a Poisson $\sqrt{N_{\text{tot}}}$ factor to the MINOS uncertainty of the branching fraction and all yields for estimating properly the uncertainty from the fluctuation of the total number of events.

5.2.1 Identification of dependences between fit observables

Special care is needed in writing a multivariate likelihood, since potentially large biases may arise if the probability densities of dependent observables are factorized improperly. To determine the proper probability density functions, dependences between the fit observables need to be identified and modeled.

For each component I identify the major dependences by studying distributions of each observable conditional on each of the others. This is achieved by inspecting the distributions of one variable for separate narrow intervals ("slices") of the others and checking whether the distribution of the probe observable depends on the range chosen in the other or not. For instance, Fig. 5.4 shows how the distribution of ΔE in slices of M_{bc} is obtained. The shape of the ΔE distribution changes as a function of M_{bc} , indicating a dependence between these two observables.

In principle one should study distributions of each variable sliced simultaneously in all three remaining dimensions. This approach is complicated by the reduction in statistical information that becomes critical if simulated samples are partitioned across too many dimensions. However, a coarse study of the full-fledged partitioning shows that all major features are mostly captured by one-dimensional slicings, where each observable is shown in slices of another, integrated over the remaining two observables. I therefore show here distributions obtained by one-dimensional slicings only.

Figure 5.5 shows such distributions for all the combinations of pairs of fit observables for the all components. For the signal, I observe dependences between ΔE and M_{bc} , and between C'_{FBDT} and w' . The dependence between ΔE and M_{bc} is due to the photon energy measurement entering in both the observables, while the dependence between C'_{FBDT} and w' is due to the common information on the B_{tag} meson that is used in both observables – in the first case to distinguish between $B\bar{B}$ and $q\bar{q}$ events, in the second to predict the flavor of the B meson.

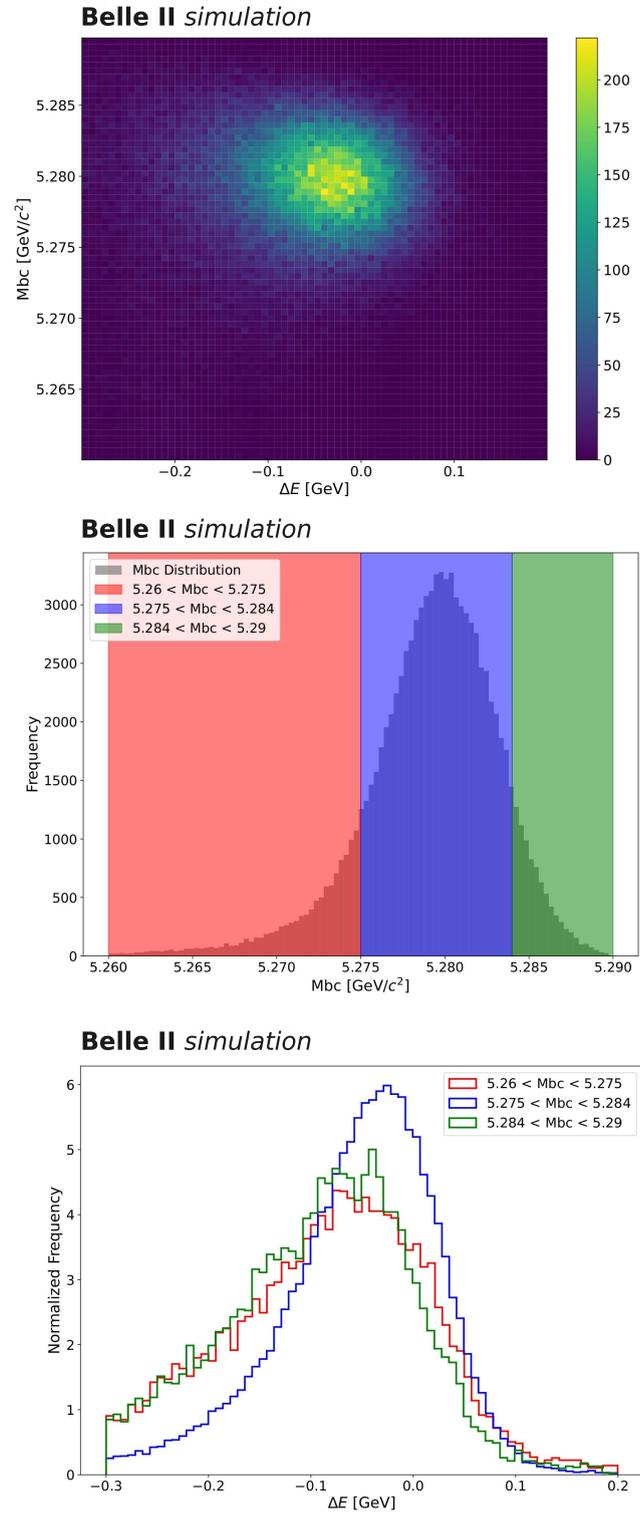


Figure 5.4: Distribution of (top) M_{bc} as a function of ΔE , (center) M_{bc} divided into three regions, and (bottom) ΔE for the three slices. All distributions are based on simulated signal events.

For $B\bar{B}$ background, there are no significant dependences among ΔE and M_{bc} with the available size of simulated samples, while a large dependence between C'_{FBDT} and w' exists. For continuum, I observe a dependence between ΔE and w' .

This study allows identification of which multidimensional p.d.f.'s can be factorized while still approximating reasonably the dependences present in the data. Not all dependences can be exactly reproduced in the model, since the finite size of the simulated or data samples used to identify them imposes a limit on the sensitivity to the effects probed. In addition, all of this study of dependences is based on simulation, but experimental data may exhibit additional or different dependences not described in simulation. The model is described in the following sections.

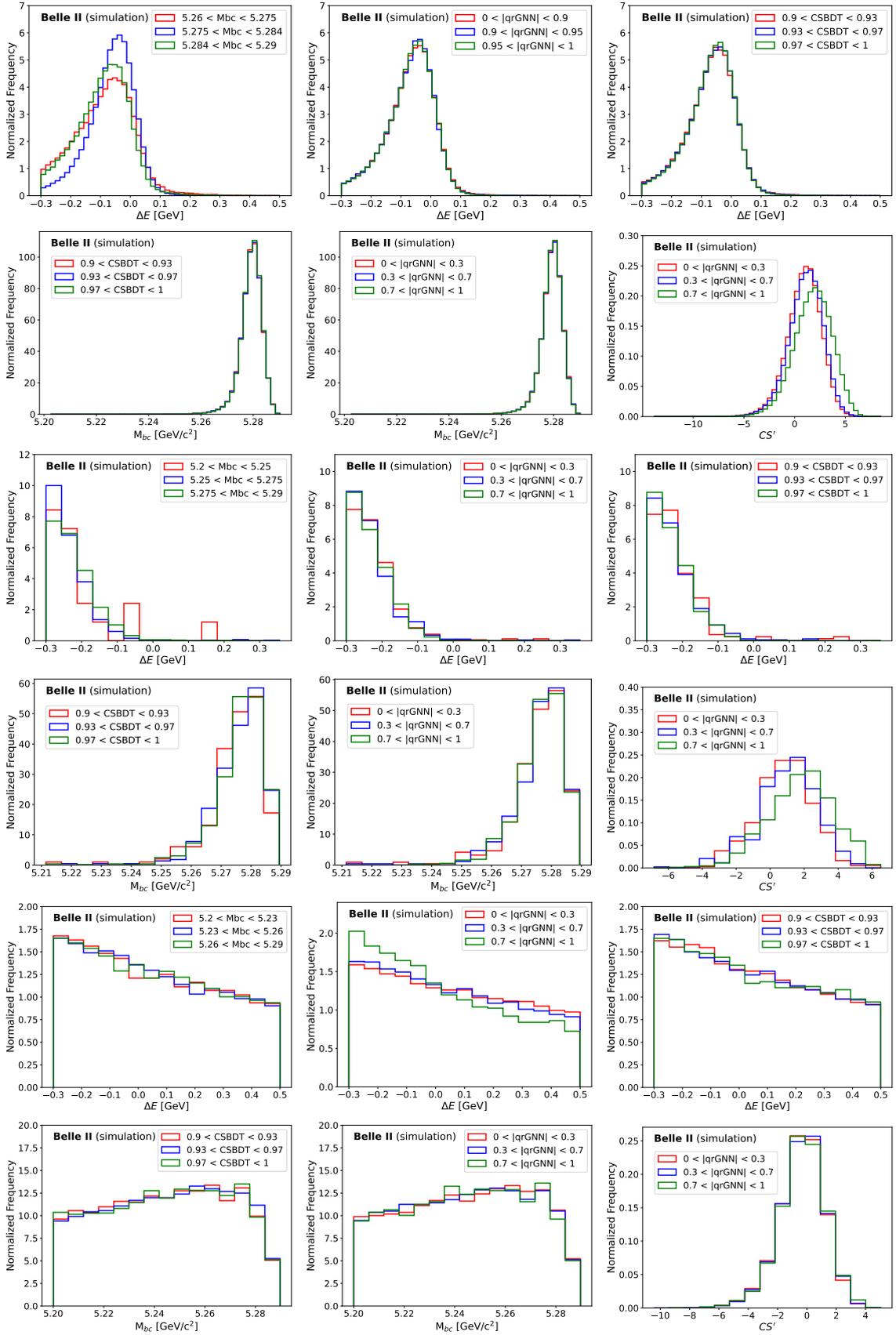


Figure 5.5: Dependence studies: distributions of fit observables in slices of other fit observables for simulated (top) signal, (center) $B\bar{B}$, and (bottom) continuum events. Different colors indicate different slices. Distributions within each panel are normalized to each other.

5.2.2 Signal p.d.f.

The probability density function that describes the signal component is factorized as the product of four terms,

$$p_{\text{sig}} = p_{\text{sig}}(\Delta E|M_{\text{bc}})p_{\text{sig}}(M_{\text{bc}})p_{\text{sig}}(C'_{\text{FBDT}}|w')p_{\text{sig}}(w'). \quad (5.6)$$

The function $p(\Delta E|M_{\text{bc}})$ is a conditional p.d.f. that accounts for the dependence between the two observables: ΔE is modeled empirically as the sum of a Gaussian function and a bifurcated Gaussian

$$p(\Delta E) \propto f_1 \frac{e^{-\frac{1}{2}\left(\frac{\Delta E - \mu_1}{\sigma_1}\right)^2}}{\sigma_1 \sqrt{2\pi}} + (1 - f_1) \frac{e^{-\frac{1}{2}\left(\frac{\Delta E - \mu_2}{\sigma_{L,R}}\right)^2}}{\sigma_{L,R} \sqrt{2\pi}} \quad (5.7)$$

where μ_1 and μ_2 are the means of the Gaussian function and of the bifurcated Gaussian, respectively, σ_1 is the width of the Gaussian, and σ_L (σ_R) is the width of the bifurcated Gaussian used for ΔE values smaller (larger) than μ_2 . The proportion f_i , the means μ_i , and widths σ_i are determined independently from modeling fits to simulation in three separate intervals of M_{bc} ; $p(M_{\text{bc}})$ is modeled empirically as the sum of two Gaussian functions:

$$p(M_{\text{bc}}) \propto f_1 \frac{e^{-\frac{1}{2}\left(\frac{M_{\text{bc}} - \mu_1}{\sigma_1}\right)^2}}{\sigma_1 \sqrt{2\pi}} + (1 - f_1) \frac{e^{-\frac{1}{2}\left(\frac{M_{\text{bc}} - \mu_2}{\sigma_2}\right)^2}}{\sigma_2 \sqrt{2\pi}} \quad (5.8)$$

The term $p(C'_{\text{FBDT}}|w)$ is a conditional p.d.f. that accounts for the dependence between continuum-suppression output and wrong-tag fraction. The continuum suppression is described with the sum of a Gaussian function and a bifurcated Gaussian similar to Eq.(5.7), where parameters are determined independently from modeling fits to simulation in three separate intervals of w' . The wrong-tag $p(w')$ is also described with the sum of a Gaussian function and a bifurcated Gaussian. Figure 5.6 shows the modeling fits to simulation and the conditional p.d.f. models chosen to describe the observables. Structures in the differences between simulated data and fit projections in ΔE , M_{bc} , and w' , show that our model does not fully capture all the detailed features of the distributions. Nevertheless, simulated samples used here contain 500 times the signal yield expected in data. Hence, I expect the impact of these mismodeling to be marginal, and fully captured by the systematic uncertainty associated with the signal modeling. All the signal model parameters are fixed in the composition fit to data, except for those associated to model refinements described in Sec. 6.2.3.

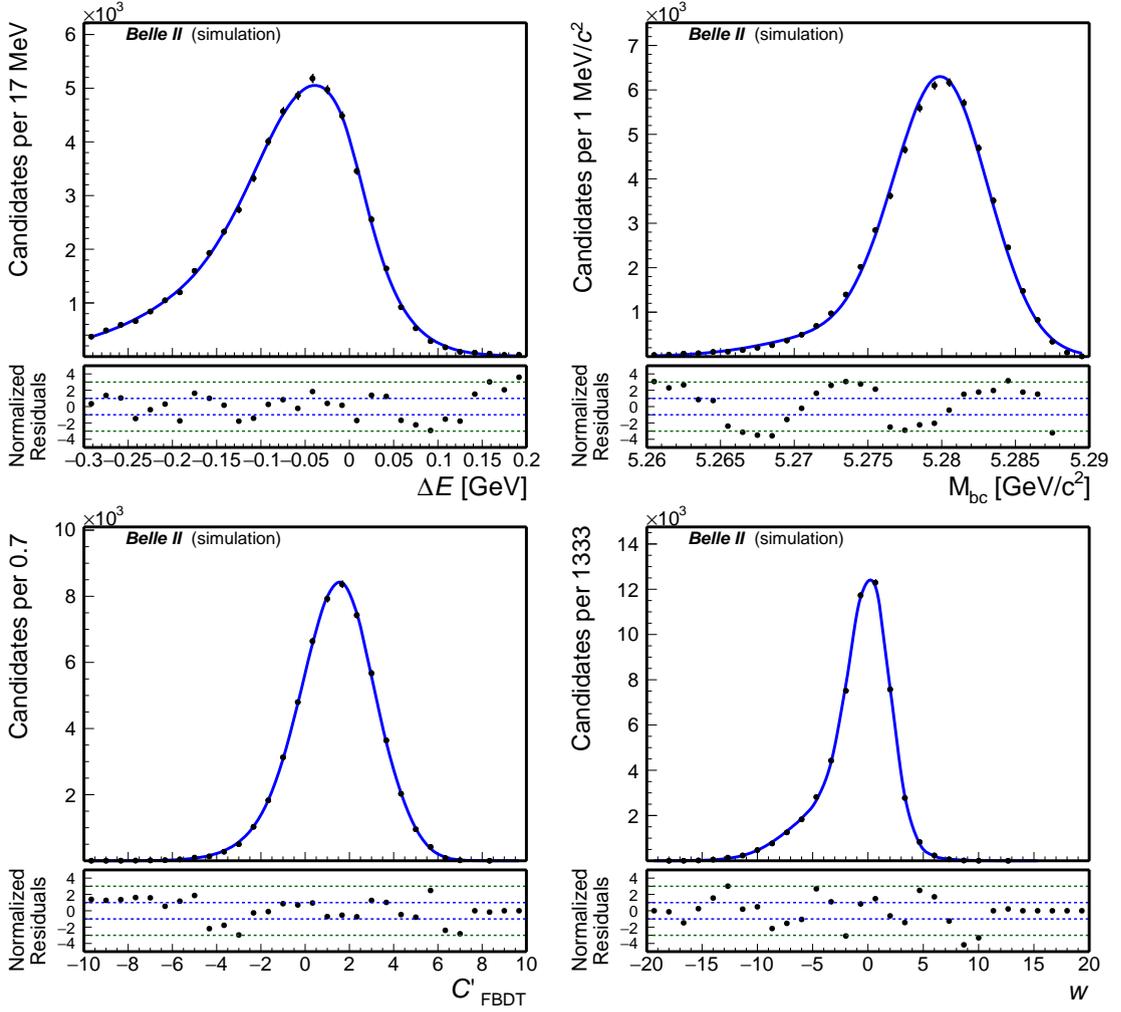


Figure 5.6: Fit modeling for signal: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated signal. The corresponding model shapes are overlaid (solid). Bottom panels show the uncertainty-weighted differences between distribution and fit.

5.2.3 $B\bar{B}$ background p.d.f.

The probability density function for the $B\bar{B}$ backgrounds has similar structure to the signal p.d.f.. It is factorized as the product of four terms,

$$p_{B\bar{B}} = p_{B\bar{B}}(\Delta E)p_{B\bar{B}}(M_{bc})p_{B\bar{B}}(C'_{\text{FBDT}}|w')p_{B\bar{B}}(w'). \quad (5.9)$$

The term in ΔE is modeled empirically as the sum of a Gaussian function and a bifurcated Gaussian similar to Eq.(5.7) but with independent parameters, while M_{bc} is modeled empirically as the sum of a Gaussian function and a Johnson function,

$$p(M_{bc}) \propto f_1 \frac{e^{-\frac{1}{2}\left(\frac{M_{bc}-\mu_1}{\sigma_1}\right)^2}}{\sigma_1\sqrt{2\pi}} + (1-f_1) \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+\left(\frac{M_{bc}-\xi}{\lambda}\right)^2}} e^{-\frac{1}{2}\left(\gamma+\delta \sinh^{-1}\left(\frac{M_{bc}-\xi}{\lambda}\right)\right)^2}. \quad (5.10)$$

The continuum suppression is described with the sum of a Gaussian function and a bifurcated Gaussian similar to Eq.(5.7), with parameters determined independently from

modeling fits to simulation in three separate intervals of w' ; the term $p(w')$ is also described with the sum of a Gaussian function and a bifurcated Gaussian. Figure 5.7 shows that all $B\bar{B}$ p.d.f. models are adequate.

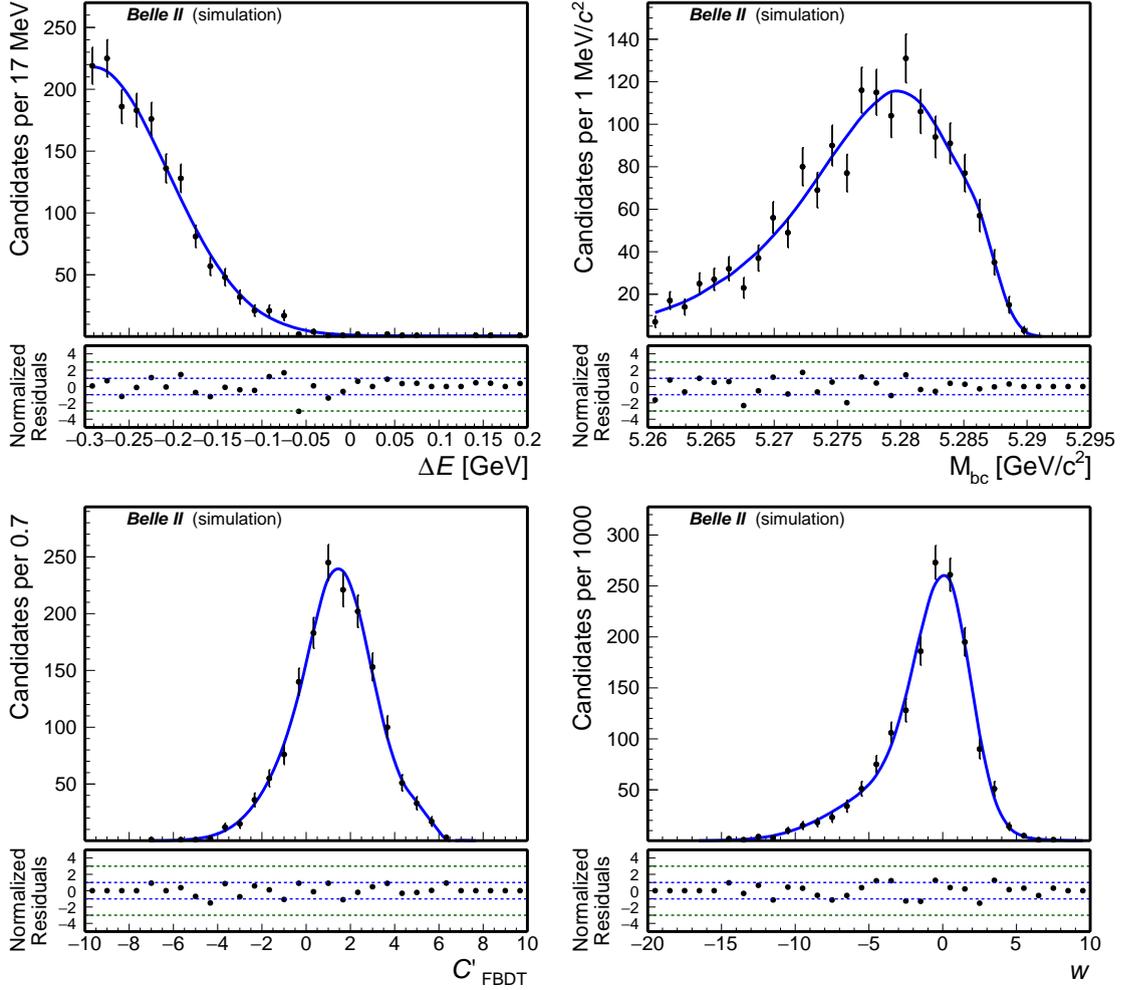


Figure 5.7: Fit modeling for $B\bar{B}$: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w for simulated signal. The corresponding model shapes are overlaid (solid). Bottom panels show the uncertainty-weighted differences between distribution and fit.

5.2.4 Continuum background p.d.f.

The continuum background is described with the product of four terms,

$$p_{\text{cont}} = p_{\text{cont}}(\Delta E|w)p_{\text{cont}}(M_{bc})p_{\text{cont}}(C'_{\text{FBDT}})p_{\text{cont}}(w). \quad (5.11)$$

The function $p(\Delta E|w')$ is a conditional p.d.f. that accounts for the dependence between the two observables. The ΔE distribution is modeled empirically as a straight line,

$$p(\Delta E) \propto a\Delta E + 1, \quad (5.12)$$

where a is determined from a modeling fit in the sideband region of experimental data. The term $p(w')$ is modeled with the sum of a Gaussian function and a bifurcated Gaussian

similar to Eq.(5.7), also modeled with independent parameters from sideband data. The beam-constrained mass is modeled with an empirical Argus function [96],

$$p(M_{bc}) \propto \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} \cdot \frac{M_{bc}}{c^2} \sqrt{1 - \frac{M_{bc}^2}{c^2}} \exp \left[-\frac{1}{2}\chi^2 \left(1 - \frac{M_{bc}^2}{c^2} \right) \right], \quad (5.13)$$

where c is the end-point of the distribution, χ is its curvature, and $\Psi(\chi) = \Phi(\chi) - \chi\phi(\chi) - \frac{1}{2}$, where $\Phi(\chi)$ and $\phi(\chi)$ are the cumulative distribution and probability density functions of the Gaussian distribution, respectively. Due to the slight changes and drifts in center-of-mass energies as functions of time, the endpoint of the M_{bc} distribution varies in time. I empirically take this into account by modeling the continuum with the sum of two Argus functions. The continuum suppression is modeled with the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7) but with independent parameters.

The models are determined from modeling fits to the sidebands of data. The sideband is the ΔE - M_{bc} region defined in Table 5.1 and shown in Fig. 5.8. The continuum parameters are then left free in the fit to signal sample composition.

Observable	Range
ΔE	$M_{bc} < 5.26 \text{ GeV}/c^2$
M_{bc}	$\Delta E > 0.2 \text{ GeV}$
C'_{FBDT}	$M_{bc} < 5.26 \text{ GeV}/c^2$ or $\Delta E > 0.2 \text{ GeV}$
w	$M_{bc} < 5.26 \text{ GeV}/c^2$ or $\Delta E > 0.2 \text{ GeV}$

Table 5.1: Sideband region definitions.

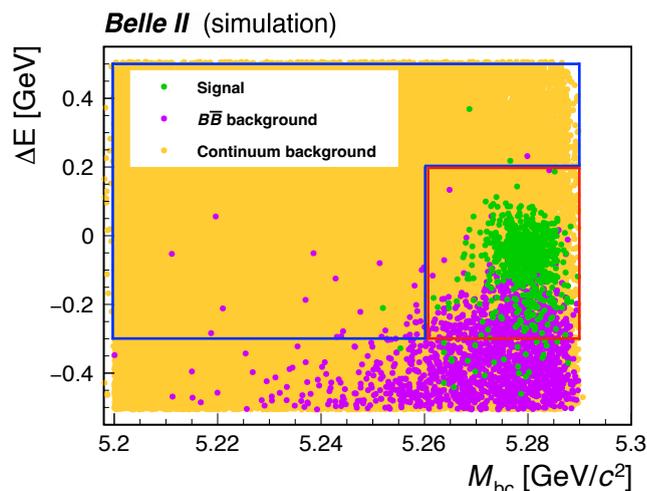


Figure 5.8: Distribution of ΔE as a function of M_{bc} for simulated $B^0 \rightarrow \pi^0\pi^0$ events, with signal and sideband regions highlighted.

Figure 5.9 shows one-dimensional projections of the continuum models.

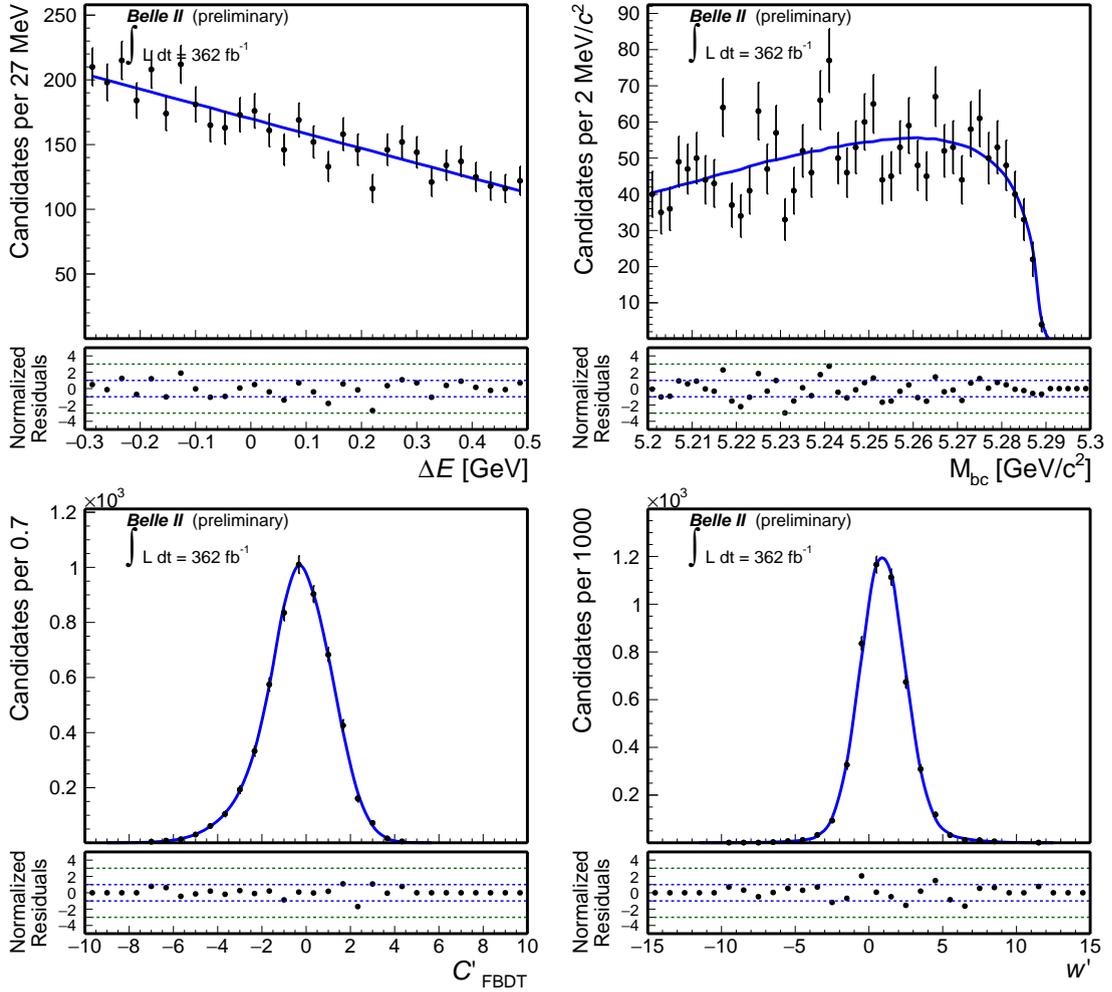


Figure 5.9: Fit modeling for the continuum: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for sideband events in data. The corresponding model shapes are overlaid (solid).

I assume that the distributions for the four observables in the sidebands correctly describe the distributions in the signal region. To partially verify this assumption, I compare the distribution of the fit observables in signal and sideband regions of simulated continuum. The result of the comparison is shown in Fig. 5.10. No major discrepancy is observed. It is sound to assume that the observed distributions do not change between signal and sideband region in data too.

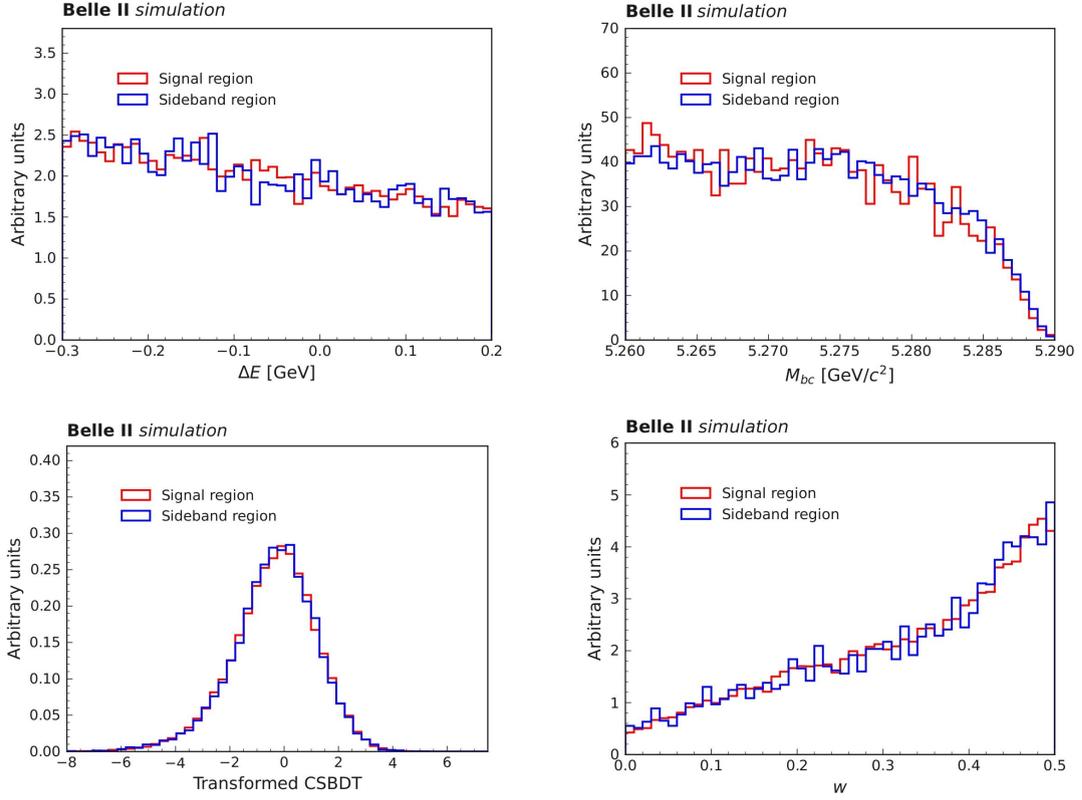


Figure 5.10: Distribution of (top left) ΔE , (top right) M_{bc} , (bottom left) continuum suppression, and (bottom right) w for candidates reconstructed in simulated continuum and restricted to the (red) signal and (blue) sideband regions.

5.2.5 P.d.f. summary

I summarize the model choices in Table 5.2. The resulting single-event likelihood consists of a combination of various analytical functions. The shape parameters for signal and $B\bar{B}$ background are fixed from modeling fits to simulation. For continuum, the model is determined in sideband data, but the shape parameters are free and directly determined in the fit of sample composition. I assume and verify in simulation that the probability density functions are adequate to describe distributions of both B^0 and \bar{B}^0 candidates.

Component	ΔE	M_{bc}	C'_{FBDT}	w'
$B^0 \rightarrow \pi^0\pi^0$	G+BG M_{bc}	2G	G+BG w'	G+BG
$B\bar{B}$	G+BG	G+J	G+BG w'	G+BG
Continuum	Pol1 w'	Double Argus	G+BG	G+BG

Table 5.2: Model summary. The shorthand "2G" means sum of two Gaussian functions, "G+BG" indicates the sum of a Gaussian function and a bifurcated Gaussian function, "G+J" indicates the sum of a Gaussian function and a Johnson function, and "Pol1" indicates a straight line.

5.3 Model refinements

The introduced likelihood describes an ideal and simplified model. In reality, when fitting data I need to take care of possible differences between data and the models I use, and of possible effects that can bias the measurement. In particular, I describe here model refinements related to the calibration of the flavor tagger algorithm, that can wrongly-describe data, and to the possible effective asymmetries in background that could bias the A_{CP} measurement.

5.3.1 Wrong-tag calibration

The flavor tagger provides the predicted flavor of the B_{tag} meson, q , and a dilution factor $r = 1 - 2w$ that is included in the likelihood function (Eq.(5.2)). Nevertheless, the wrong-tag fraction predicted by the algorithm can differ from the wrong-tag fraction in data: I need to calibrate the flavor tagger output to make these two quantities match, and to account for possible algorithm-generated asymmetries in the predicted rates of B^0 vs \bar{B}^0 . This is done by performing fits to the Δt distributions of $B^0 \rightarrow D^{(*)-}\pi^+$ decays, where Δt is the time between the decay of a B meson and the observation of the second B meson. This allows to measure the time-dependent CP -violating parameters, A_{CP} and \mathcal{S}_{CP} . $B^0 \rightarrow D^{(*)-}\pi^+$ are reconstructed by combining tracks and eventually π^0 mesons in a D meson candidate, required to have a mass similar to the known mass (between ± 2.5 and $\pm 5\sigma$ depending on the channel). In addition, if D^* candidates are reconstructed, requirements on their energy are applied. In these decays, the real flavor is known from the charge of the pion coming from the B meson and is therefore compared with the tagger prediction. In particular, I use two quantities, defined for B^0 and \bar{B}^0 mesons and in seven separate intervals of qr ,

- w_i , the wrong-tag fraction in the qr interval i for B_{tag}^0 and \bar{B}_{tag}^0 mesons,
- $\mu_i = (\varepsilon_{B_{\text{tag}}^0} - \varepsilon_{\bar{B}_{\text{tag}}^0}) / (\varepsilon_{B_{\text{tag}}^0} + \varepsilon_{\bar{B}_{\text{tag}}^0})$, the asymmetry of the tagging efficiency between B^0 and \bar{B}^0 .

To obtain an event-per-event calibration of w , I compare the values of w_i predicted by the flavor tagger and those obtained from the fit to $B^0 \rightarrow D^{(*)-}\pi^+$ candidates in data, separately for B_{tag}^0 and \bar{B}_{tag}^0 mesons. To a good approximation, the true wrong-tag w_{true} and its predicted value w are related by the linear function

$$w_{\text{true}} = k \cdot w.$$

I obtain the calibration parameter k for both data and simulation. If the calibrated wrong-tag value exceeds 0.5, the wrong-tag value is taken to be

$$w_{\text{true}}^{\text{new}} = 1 - w_{\text{true}},$$

and the sign of the predicted q in the likelihood switched.

I obtain a weighted μ value based on the signal distribution of qr in the control channel, finding a mean value $\langle \mu \rangle = 0.0015 \pm 0.0016$.

The calibrated likelihood becomes then

$$\begin{aligned} \mathcal{L}_i(f_{\text{sig}}, f_{B\bar{B}}, A_{CP}|q, w, \Delta E, M_{\text{bc}}, C'_{\text{FBDT}}) = & N_{\text{tot}} [f_{\text{sig}}(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))A_{CP}) \\ & \cdot p_{\text{sig},i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w') \\ & + f_{B\bar{B}} \cdot p_{B\bar{B},i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w') \\ & + (1 - f_{\text{sig}} - f_{B\bar{B}}) \cdot p_{\text{cont},i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w')], \end{aligned}$$

	k (data)	p-value (data)	k (sim.)	p-value (sim.)
B_{tag}^0	1.028 ± 0.012	0.150	0.981 ± 0.010	0.899
\bar{B}_{tag}^0	1.029 ± 0.012	0.217	0.993 ± 0.011	0.711

Table 5.3: Calibration factors in $B^0 \rightarrow D^{(*)-}\pi^+$ decays reconstructed in data and simulation.

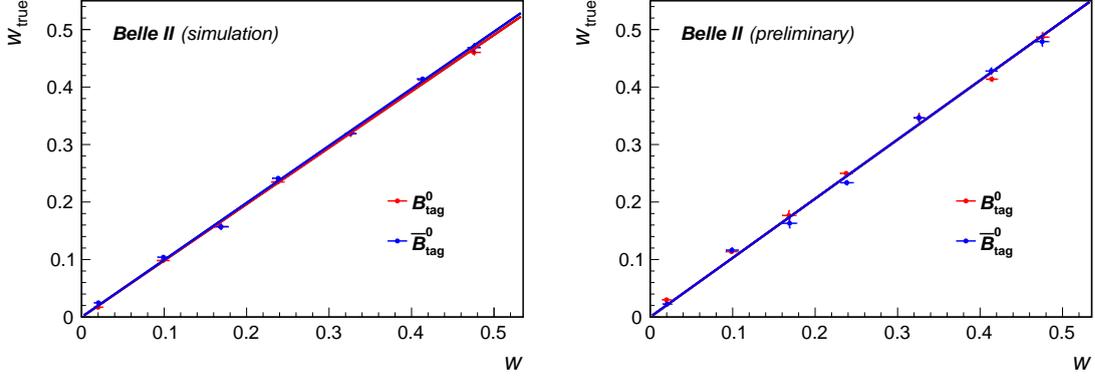


Figure 5.11: Real wrong-tag as function of predicted wrong-tag in $B^0 \rightarrow D^{(*)-}\pi^+$ decays, separately for (red) B^0 and (blue) \bar{B}^0 mesons, for (left) simulation and (right) data.

where N_{tot} is the total number of observed candidates, which is fixed; f_{sig} and $f_{B\bar{B}}$ are the fractions of signal and $B\bar{B}$ background in the sample, respectively; $q = \pm 1$ is the B_{tag}^0 flavor predicted by the flavor tagger (B^0 and \bar{B}^0 , respectively); k^q is the flavor tagger calibration factor; μ is the tagging efficiency asymmetry, $\chi_d = 0.1858 \pm 0.0011$, and $p_{j,i} = p_{j,i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w')$ are the appropriate combinations of p.d.f.s, assumed to be flavor-symmetric. In the fit composition likelihood likelihood, I impose a Gaussian constraint on all calibration parameters.

The calibration of mistag parameters is based on $B^0 \rightarrow D^{(*)-}\pi^+$ decays. To ensure that the calibration parameters can be applied to the signal channel, I perform the w calibration on simulated $B^0 \rightarrow \pi^0\pi^0$ decays, and compare the results with those obtained from $B^0 \rightarrow D^{(*)-}\pi^+$ simulation. Table 5.3 and Fig. 5.11 show the results.

The parameters are compatible with those of Table 5.4: the difference between measured and predicted wrong-tag values is consistent across channels, making calibration factors independent from the decay. Hence, I use the calibration parameters obtained from the $B^0 \rightarrow D^{(*)-}\pi^+$ control channel in data in the $B^0 \rightarrow \pi^0\pi^0$ analysis.

I also check that the calibration factors do not depend on the continuum suppression requirements by performing a w calibration on simulated $B^0 \rightarrow \pi^0\pi^0$ decays restricted in continuum suppression to exceed 0.9, and comparing the calibration results with those obtained previously. Table 5.4 and Fig. 5.12 show the results.

The parameters are compatible with those in Table 5.4 and with those obtained without any continuum suppression requirement. Hence, the the calibration factors are not only independent from the channel considered, but also from any restriction in continuum suppression.

	k (sig. sim.)	p-value (sig. sim.)	k (sig. sim. $C > 0.9$)	p-value (sig. sim. $C > 0.9$)
B_{tag}^0	0.987 ± 0.008	0.891	0.989 ± 0.010	0.634
\bar{B}_{tag}^0	0.989 ± 0.008	0.028	0.989 ± 0.010	0.005

Table 5.4: Calibration factors obtained from simulated signal $B^0 \rightarrow \pi^0\pi^0$ decays, and the same events restricted to have continuum suppression larger than 0.9.

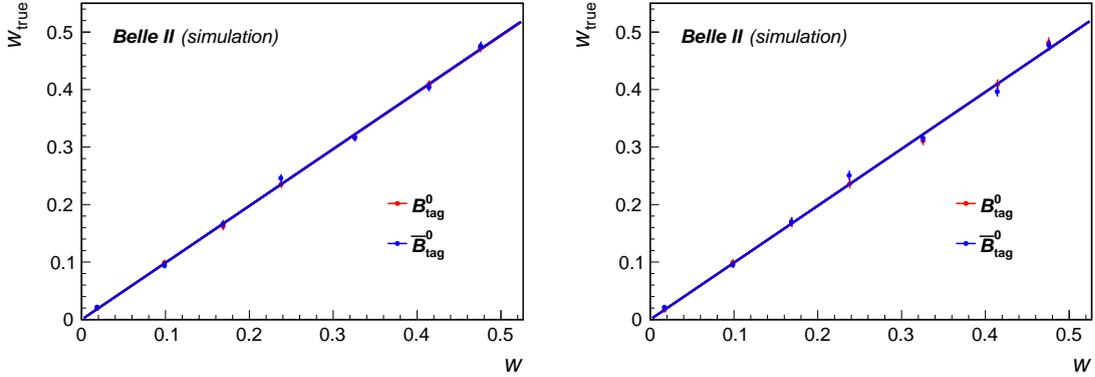


Figure 5.12: Real wrong-tag as function of predicted wrong-tag in simulated signal $B^0 \rightarrow \pi^0\pi^0$ decays, separately for (red) B^0 and (blue) \bar{B}^0 mesons, for (left) all events and (right) events restricted to have continuum suppression larger than 0.9.

5.3.2 Flavor asymmetry in background

5.3.2.1 $B\bar{B}$ background

The $B\bar{B}$ background is composed by misreconstructed B decays. Instrumental effects or CP -violating asymmetries of the involved B decays can mimic an effective asymmetry of the flavor tagger output that may bias the A_{CP} result. I include an effective asymmetry of the $B\bar{B}$ background, $A_{\text{eff}}^{B\bar{B}}$, as an additional fit parameter in the likelihood of the fit of sample composition. Since 85% of the $B\bar{B}$ background is composed by two specific decays, $B^+ \rightarrow \rho^+(\rightarrow \pi^+\pi^0)\pi^0$ and $B^0 \rightarrow K_S^0(\rightarrow \pi^0\pi^0)\pi^0$, as described in Sec. 4.8, I assume $A_{\text{eff}}^{B\bar{B}}$ to have the known value $0 \pm 11\%$ [11].

5.3.2.2 Continuum background

Continuum background is expected to be around 50 times larger than the signal. Hence, it is important to check for any continuum flavor asymmetry. Ideally, the qr distribution of the continuum should be symmetric. However, instrumental asymmetries in the detection of particles and antiparticles can introduce an effective asymmetry in the qr distribution, which in turn may bias the signal A_{CP} . The asymmetry of the continuum is

$$A_{\text{eff}}^{q\bar{q}} = \frac{N(\bar{B}^0) - N(B^0)}{N(\bar{B}^0) + N(B^0)} \quad (5.14)$$

where $N(\bar{B}^0)$ and $N(B^0)$ are the numbers of reconstructed neutral B meson candidates in continuum with $qr < 0$ and $qr > 0$, respectively. The qr distribution of the continuum background in both the sideband and the signal regions (as shown in Fig. 5.13) agree. A comparison of the continuum asymmetry in simulation between sideband and signal region,

finds compatible results. Therefore, I assume the asymmetry in the sideband to describe the asymmetry in the signal region also in data, as shown in Table 5.5, and account for this instrumental asymmetry by including in the likelihood the value of $A_{\text{eff}}^{q\bar{q}}$ measured in the data sideband.

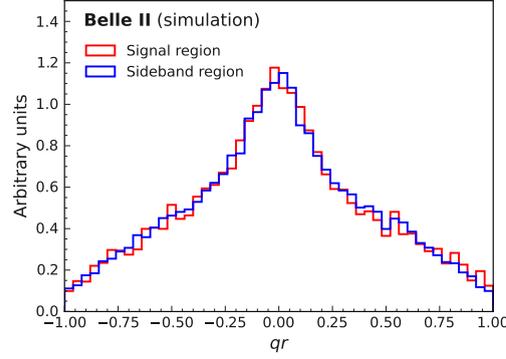


Figure 5.13: Distribution of qr for simulated continuum reconstructed in the (red) signal region and (blue) sideband regions.

Sample	$A_{\text{eff}}^{q\bar{q}}$
Signal region (simulation)	-0.0146 ± 0.0186
Sideband region (simulation)	-0.0151 ± 0.0105
Sideband region (data)	-0.0153 ± 0.0271

Table 5.5: Continuum asymmetry values for sideband and signal region obtained in simulation and experimental data. The error is statistical only.

After including the effective background asymmetries, the likelihood becomes

$$\begin{aligned}
 \mathcal{L}_i(f_{\text{sig}}, f_{B\bar{B}}, A_{CP}|q, w, \Delta E, M_{\text{bc}}, C'_{\text{FBDT}}) = & N_{\text{tot}} [f_{\text{sig}}(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))A_{CP}) \\
 & \cdot p_{\text{sig},i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w') \\
 & + f_{B\bar{B}} \left(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))A_{\text{eff}}^{B\bar{B}}\right) \\
 & \cdot p_{B\bar{B},i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w') \\
 & + (1 - f_{\text{sig}} - f_{B\bar{B}}) \left(1 + q(1 - 2(k^q w))A_{\text{eff}}^{q\bar{q}}\right) \\
 & \cdot p_{\text{cont},i}(\Delta E, M_{\text{bc}}, C'_{\text{FBDT}}, w')],
 \end{aligned}$$

I impose a Gaussian constraint on both the effective background asymmetries.

5.4 Estimator properties

Maximum likelihood estimates can show statistical bias and non-Gaussian dispersions because of multiple causes. The most common is that the model is too simplified and fails to capture some of the relevant data features. Both phenomena lead to unsatisfactory results. To investigate the presence and entity of these phenomena, I study the distributions of

estimators on simulated data. I use the pull value of each fit parameter θ_i , defined as the uncertainty-weighted residual,

$$\text{Pull}(\theta_i) = \frac{\hat{\theta}_i - \theta_i}{\hat{\sigma}_{\hat{\theta}_i}}, \quad (5.15)$$

where θ is the known true value of the parameter, $\hat{\theta}_i$ is the estimate, and $\hat{\sigma}_{\hat{\theta}_i}$ is the estimate of its uncertainty. Distributions of residuals are studied over ensembles of simplified simulated experiments ("toys") that simulate the experimental circumstances of the fit on data. The samples are obtained by mixing simulated signal, continuum, and $B\bar{B}$ background events in realistic proportions and sizes.

I generate 2000 toy samples by drawing the yield of each component according to a Poisson distribution centered at the true values expected from simulation. The proportion between B^0 and \bar{B}^0 candidates is given by the true A_{CP} value used in generation. Then, I fit the composition of each toy as if they were data, and construct the pull-value distributions of the fit results.

Ideally, for an unbiased and asymptotic maximum likelihood estimator, the pull distribution is a Gaussian distribution mean-valued at zero with unit variance. A mean differing from zero corresponds to a bias (in units of statistical uncertainty), while a variance smaller (larger) than one corresponds to an over-(under)estimation of the statistical uncertainty.

Figure 5.14 shows pull distributions for the $\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$ and $A_{CP}(B^0 \rightarrow \pi^0\pi^0)$ parameters. The means are compatible with zero, showing no intrinsic biases in the fit. The widths of the distributions are compatible with one, showing no significant over-(under-)estimation of the uncertainty.

Figure 5.15 shows residual distributions for the $\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$ and $A_{CP}(B^0 \rightarrow \pi^0\pi^0)$ parameters. The widths of the distributions correspond to the expected statistical uncertainties.

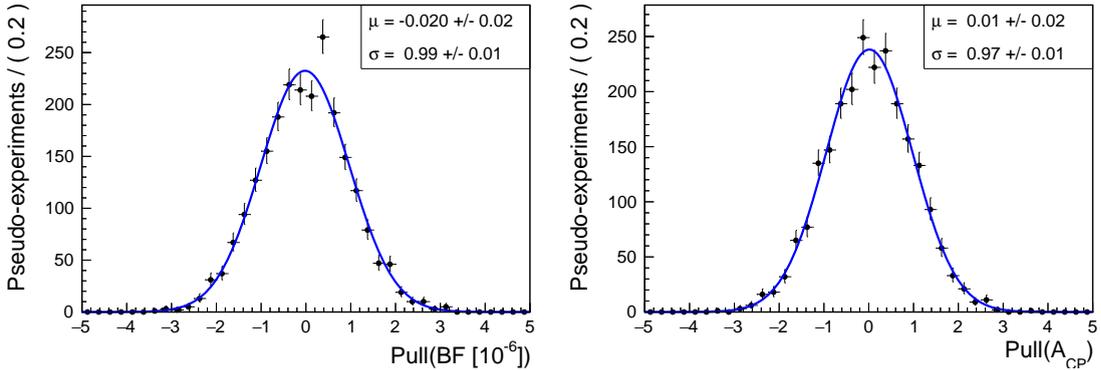


Figure 5.14: Pull distributions of (left) branching fraction and (right) CP -violating asymmetry of $B^0 \rightarrow \pi^0\pi^0$ decays. Results of fits to Gaussian functions are overlaid (solid).

To verify if the estimator properties depend on the true values, I repeat the toy study by generating 10 ensembles of 1000 toy samples each. Each ensemble simulates a combination of true continuum, $B\bar{B}$ background, B^0 , and \bar{B}^0 proportions sampled in the $\pm 3\sigma$ range from the expected values. Then, I fit the composition of each toy and construct the pull-value distributions of the fit results in each ensemble.

Figure 5.16 shows the means and the widths of the pull distributions for $\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$ and $A_{CP}(B^0 \rightarrow \pi^0\pi^0)$. The means are compatible with zero and the widths are compatible with one, showing no intrinsic biases in the fit even if the sample composition in our data

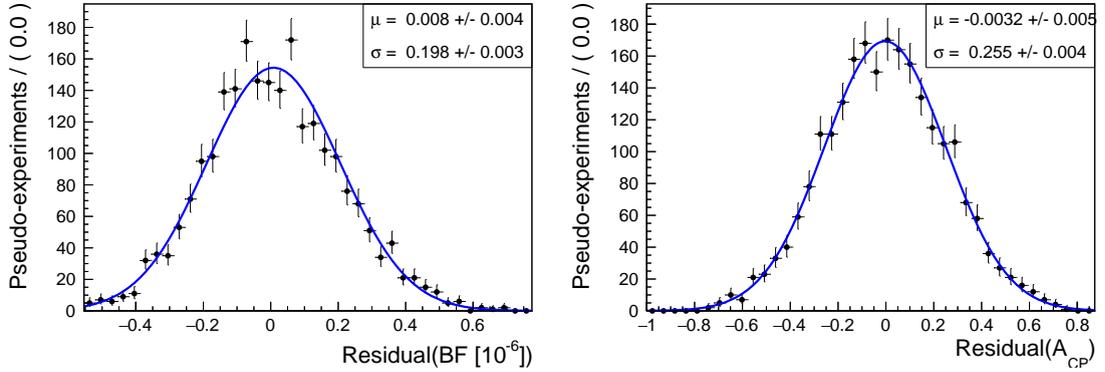


Figure 5.15: Residual distributions of (left) branching fraction and (right) CP -violating asymmetry of $B^0 \rightarrow \pi^0\pi^0$ decays. Results of fits to Gaussian functions are overlaid (solid).

differs from the expected one. The widths of the A_{CP} pulls distributions are frequently larger than one, but still compatible with unity. We don't investigate this effect further.

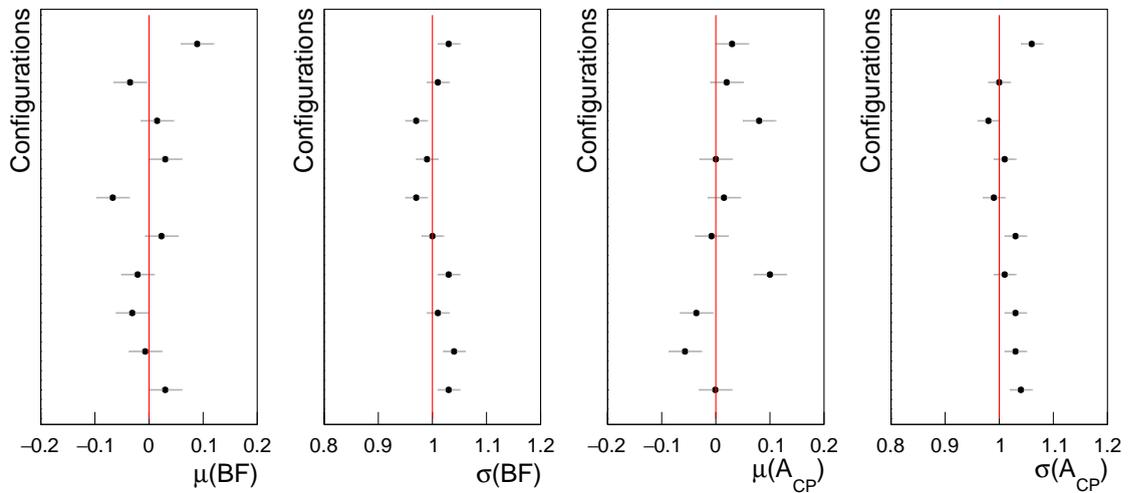


Figure 5.16: Means and widths of the pull distributions for branching fraction and CP -violating asymmetry of $B^0 \rightarrow \pi^0\pi^0$ decays using various combinations of true parameters.

5.5 Fit to realistically simulated samples

Prior to analyzing Belle II signal data, it is useful to test the fit procedure on a sample of simulated events having a composition similar to the composition we expect in data, and compare the results with the true values. Figure 5.17 shows example projections obtained from a fit to a sample simulated with realistic composition and corresponding to 1.4 ab^{-1} . A broad signal peak is observed at $\Delta E \approx -50 \text{ MeV}$, overlapping smooth continuum and non-peaking $B\bar{B}$ background distributions. In the M_{bc} distribution, a signal peak is observed at the mass of the B meson together with the $B\bar{B}$ background distribution, while continuum has a smoother shape. In the continuum-suppression-output distribution, continuum events tend to peak at $C'_{\text{FBDT}} \approx 0$, while the peak corresponding to $B\bar{B}$ events (including signal and background) is shifted toward $C'_{\text{FBDT}} \approx 1$. In the wrong-

tag distribution, continuum events tend to peak at $w' \approx 0$, while the peak corresponding to $B\bar{B}$ events (including signal and background) is shifted toward $w' \approx -1$.

Projections show that the fit reproduces adequately all the observed distributions and provides confidence on the accuracy of our model. The M_{bc} distribution shows an indication of mismodeling in the 5.27–5.275 GeV/ c^2 region, where a data deficit appears at the turn-on of signal and $B\bar{B}$ decays components. We do not investigate this effect further given that fit results agree with the true values and fluctuations in the data sample, which is four times smaller in size, would dilute any such mismodeling. Table 5.6 shows the results of the fit, compared with the known values of the simulated sample. All results are compatible with the expectations.

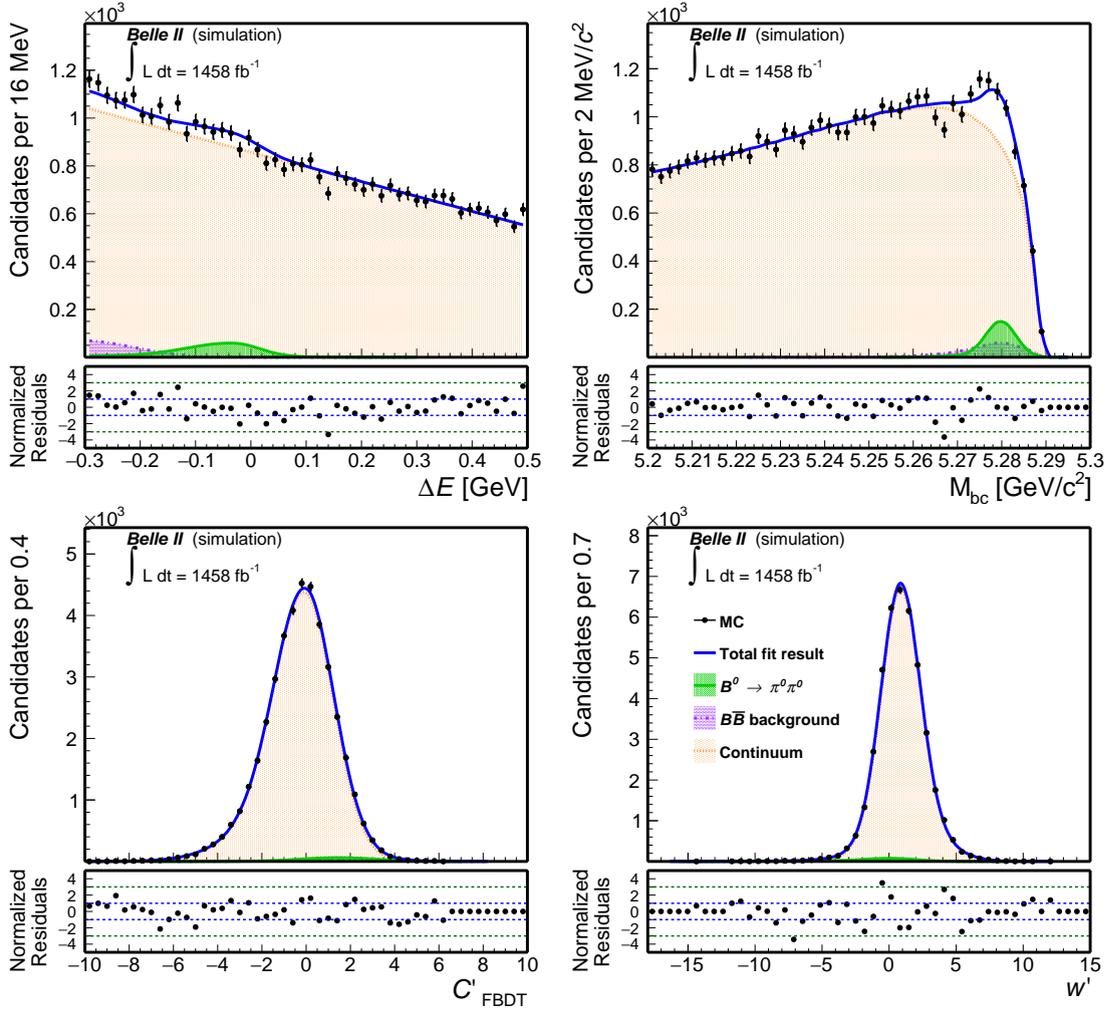


Figure 5.17: Example of fit projections: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) continuum suppression output, and (bottom right) w' for a realistic simulated Belle II sample corresponding to 1.4 ab^{-1} with fit projections overlaid.

In addition, I divide the available simulated sample in four independent subsamples, each having the same number of expected events of the real data set. The results of fits to each subsample are shown in Table 5.7 and Fig. 5.18. Results are compatible with the expected values from the simulation, and they show good statistical agreement between each other and with the result of the fit to the full sample.

Component	Fit result	Simulation value
$\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$	1.627 ± 0.099	1.670
A_{CP}	-0.028 ± 0.123	-0.038
$N_{B\bar{B}}$	491 ± 40	526

Table 5.6: Results of a fit to a sample simulated with realistic composition and corresponding to 1.4 ab^{-1} and true values generated in simulation.

Component	Total sample	Sample 1	Sample 2	Sample 3	Sample 4
$\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$	1.627 ± 0.099	1.641 ± 0.199	1.722 ± 0.201	1.400 ± 0.192	1.772 ± 0.204
A_{CP}	-0.028 ± 0.123	0.334 ± 0.241	0.090 ± 0.229	-0.264 ± 0.275	-0.292 ± 0.228
$N_{B\bar{B}}$	491.1 ± 40.4	103.7 ± 19.8	106.2 ± 19.4	135.1 ± 20.3	144.4 ± 21.4

Table 5.7: Results of fits to four independent samples simulated with realistic composition and corresponding to 362 fb^{-1} each.

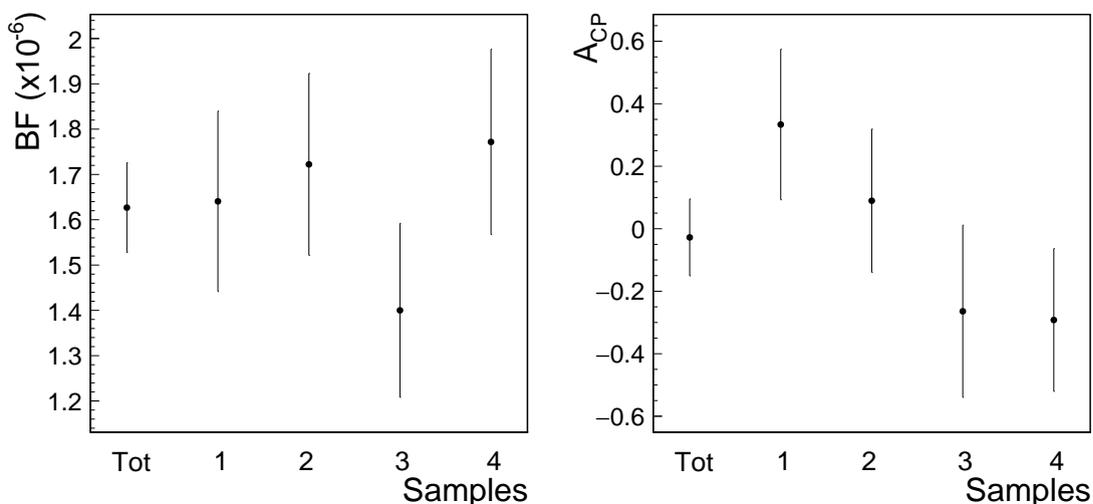


Figure 5.18: Results of fits to four independent, simulated samples corresponding to 362 fb^{-1} each, with realistic sample composition.

5.6 Fit strategy choices

During the analysis the fit strategy has undergone several changes. In the interest of full documentation, I briefly discuss here alternative strategies that were explored, but not pursued.

5.6.1 Previous analysis configuration

At the early stages of the analysis, I used a fit similar to the one of the previous Belle II analysis, that is, a simultaneous fit of the sample in seven separate intervals of the qr observable and with a range restricted to $-0.3 < \Delta E < 0.2 \text{ GeV}$ and $M_{bc} > 5.26 \text{ GeV}/c^2$. In this configuration, seven mean wrong-tag values, one for each bin, are included in the likelihood, together with an equal number of calibration factors. The fit uses the unbinned distributions of three observables, ΔE , M_{bc} , and continuum suppression. All models that

describe the observables are fixed from fits to simulation but continuum, which is based on fits to sideband data.

An example of a preliminary fit to the full simulated Belle II sample is shown in Fig. 5.19, where only the distributions of ΔE , together with the fit projections, for each qr bin and separately for $q < 0$ and $q > 0$ are shown. The continuum suppression requirement is looser than in the work of this thesis, so the continuum contamination is larger.

The simultaneous fit in separate intervals of qr is suboptimal, because of possible large deviations of qr in the same bin, which would be averaged over the bin, partially spoiling the statistical power of data. The use of an unbinned treatment for wrong-tag information, in addition to slightly improved tagging power, also helps in discriminating the signal from background. Moreover, the use of an enlarged ΔE - M_{bc} fit range, so that the fit can directly determine the continuum shape parameters from data, ensures a much smaller systematic contribution than the choice of fixing the model from fits to sideband data, that heavily depends on the availability of such data.

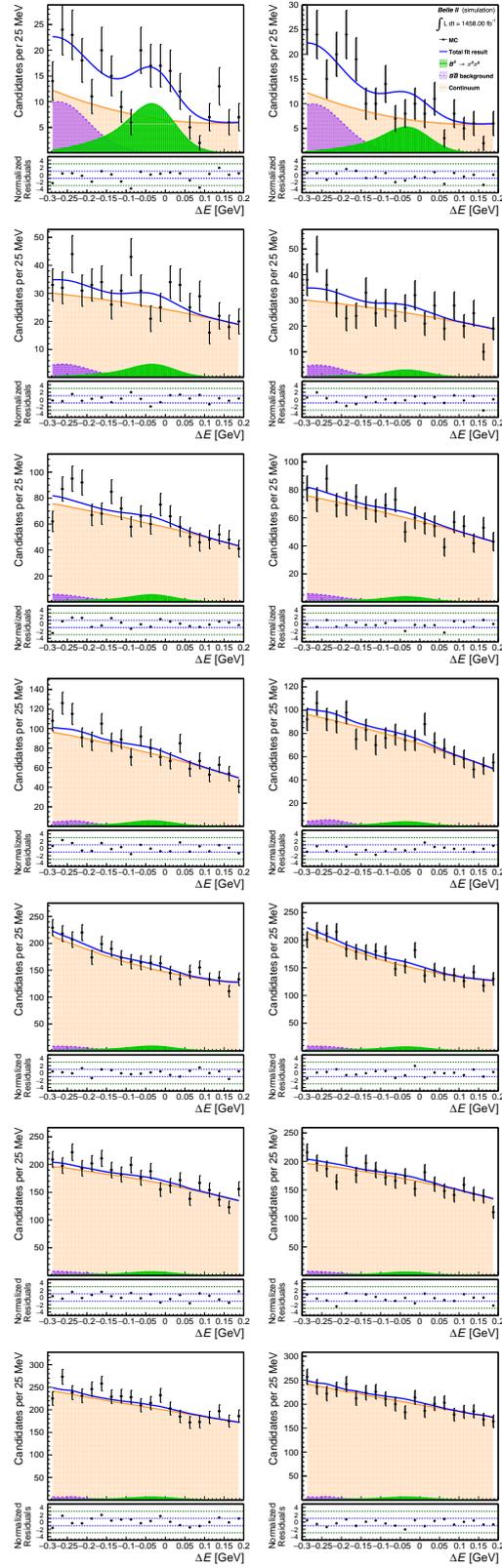


Figure 5.19: Distributions of ΔE for $B^0 \rightarrow \pi^0\pi^0$ candidates reconstructed in simulated Belle II data corresponding to 1458 fb^{-1} , for different bins of qr (top plots have higher absolute value of qr) and for (left) $q < 0$ and (right) $q > 0$. Fit projections are overlaid.

5.6.2 Continuum amplifier

The limited availability of off-resonance data to model continuum, which results in large ΔE and M_{bc} model uncertainties due to statistical fluctuations, prompted to explore novel methods based on mixing different data events to artificially increase the modeling sample size. By exploiting the symmetry and relative simplicity of two-body $B^0 \rightarrow \pi^0\pi^0$ candidates, I developed a "continuum amplifier". Figure 5.20 (left) shows a sketch of the concept by displaying two similar $\pi^0\pi^0$ candidates reconstructed from continuum represented with different colors. I search for pairs of $B^0 \rightarrow \pi^0\pi^0$ candidates reconstructed in continuum data where the reconstructed π^0 meson candidate from one event has similar momentum as a π^0 meson in any other event (red π_1^0 and blue π_1^0). This ensures that also the partner π^0 mesons in the decay, red π_2^0 and blue π_2^0 , have similar kinematic properties owing to energy-momentum conservation. I then combine the red π_1^0 with the blue π_2^0 , and – vice-versa – the red π_2^0 with the blue π_1^0 , to obtain two new pseudocandidates with similar kinematic properties of a genuine continuum event. When generalized to the full continuum sample from off-resonance data, this approach achieves an effective multiplication of the available continuum sample size, thus offering a convenient modeling solution that circumvents the difficulties in producing large samples of full-fledged simulated continuum.

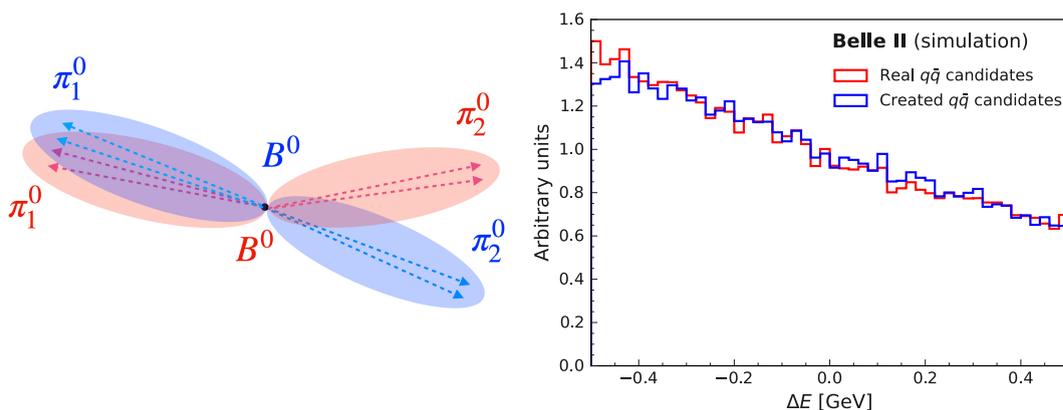


Figure 5.20: On the left, sketch of two $B^0 \rightarrow \pi^0\pi^0$ candidates (one in red, the second in blue) having a π^0 meson with similar momentum to the other. On the right, distribution of ΔE for (red) real continuum candidates and (blue) pseudocandidates in simulation. Distributions are normalized to each other.

Technically, to impose that π^0 mesons from different events have similar kinematic properties, I require their momenta to be within two standard deviations of resolution. I then combine the pions across events as if they were genuine candidates and calculate ΔE and M_{bc} , which only depend only on beam and π^0 energies. Simulation shows that the ΔE distribution for real continuum candidates and pseudocandidates are compatible (Fig. 5.20, right). This way an effective amplification by a factor 2.6 of the continuum sample size is obtained.

Given the reduction in modeling uncertainty already achieved by enlarging the fit range, this method is not adopted in the current analysis, but it may prove useful in future extensions to further shrink the associated systematic uncertainty. In addition, it can be applied to other fully reconstructed two-body decays in Belle II. On the other hand the method cannot be used to study continuum suppression and wrong-tag observables, because they are calculated using information from the rest of the event too.

Chapter 6

Analysis validation on data

Data and simulation are known to feature differences that can spoil the measurement results. This chapter discusses the validation of the analysis on data to account for such differences.

6.1 Introduction

The sample-composition fit is developed based on realistically simulated samples. However, potentially harmful discrepancies are known to exist between Belle II data and simulation. In order to ensure a reliable estimation of the desired parameters in data, it is essential to identify and correct such discrepancies. A variety of control samples in data are used to this end. I apply minimum variations of the $B^0 \rightarrow \pi^0\pi^0$ selection and fit to the abundant control channels $B^+ \rightarrow K^+\pi^0$ and $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ to search for possible undetected issues or unmodeled effects and possibly devise model corrections.

The $B^+ \rightarrow K^+\pi^0$ mode is a two-body decay, as the $B^0 \rightarrow \pi^0\pi^0$, and shares the π^0 kinematic properties of the signal decay. Hence, it probes reliably the reconstruction of a high-energy π^0 meson. In addition, this decay has a similar sample composition to the signal channel: hence, it provides validation of the continuum shape determination from data. The $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ mode is a multibody decay, with two π^0 mesons in the final state as in $B^0 \rightarrow \pi^0\pi^0$, albeit with lower energies. The decaying B meson is neutral, thus mirroring closely the tagging information of the signal channel and providing validation of flavor asymmetries. Finally, projections and uncertainties of fits of a $B^0 \rightarrow \pi^0\pi^0$ sample in half of the data set are inspected and compared with previous Belle II results to assess realistically the effects of various analysis choices.

6.2 Fit validation using $B^+ \rightarrow K^+\pi^0$ decays

A total of 1500 $B^+ \rightarrow K^+\pi^0$ events is expected in data, which is a factor of approximately ten larger than the expected $B^0 \rightarrow \pi^0\pi^0$ signal yield and ensures a precise assessment of various experimental aspects. I apply the full selection and fit to the $B^+ \rightarrow K^+\pi^0$ channel, adapted to the unavoidable differences in topology and final states.

6.2.1 $B^+ \rightarrow K^+\pi^0$ selection

The selection is the same as for the signal, except for an additional particle-identification selection for the kaon to suppress combinatorial background from pions. I choose randomly one candidate per event, as done in the $B^0 \rightarrow \pi^0\pi^0$ analysis.

6.2.2 $B^+ \rightarrow K^+\pi^0$ likelihood modeling

The $B^+ \rightarrow K^+\pi^0$ composition is statistically determined with a maximum likelihood fit to the unbinned distributions of discriminating observables as similar as possible to the fit of the signal sample. The single event likelihood is

$$\mathcal{L}_i = N_{\text{tot}} \left[f_{\text{sig}} p_{\text{sig},i} + f_{B\bar{B}} p_{B\bar{B},i} + (1 - f_{\text{sig}} - f_{B\bar{B}}) p_{\text{cont},i} \right], \quad (6.1)$$

where p_{sig} describes the probability density function of the observables for the signal events, $p_{B\bar{B}}$ for the $B\bar{B}$ background events, and p_{cont} for the continuum events; N_{tot} is the total number of observed candidates, which is fixed, and f_j corresponds to the fraction of each component, to be determined by the fit.

The shapes for signal and $B\bar{B}$ background are determined from modeling fits to simulated samples. Continuum shapes are determined from modeling fits to sideband data, but the shape parameters are left free in the fit to of sample composition.

The single event likelihood for signal is factorized, as for the $B^0 \rightarrow \pi^0\pi^0$ signal, as

$$p_{\text{sig}} = p_{\text{sig}}(\Delta E | M_{\text{bc}}) p_{\text{sig}}(M_{\text{bc}}) p_{\text{sig}}(C'_{\text{FBDT}} | w') p_{\text{sig}}(w'), \quad (6.2)$$

where the function $p(\Delta E | M_{\text{bc}})$ is a conditional p.d.f. that accounts for the dependence between the two observables: ΔE is modeled empirically as the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7) but with independent parameters in three separate intervals of M_{bc} ; $p(M_{\text{bc}})$ is modeled empirically as the sum of two Gaussian functions. The model is adequate for all observables, as shown in Fig. 6.1. A minor mismodeling is observed at low values of M_{bc} . The impact on our consistency check is expected to be negligible given the much smaller size of the data sample where this model is applied, compared to the size of the simulated sample used in the modeling fit.

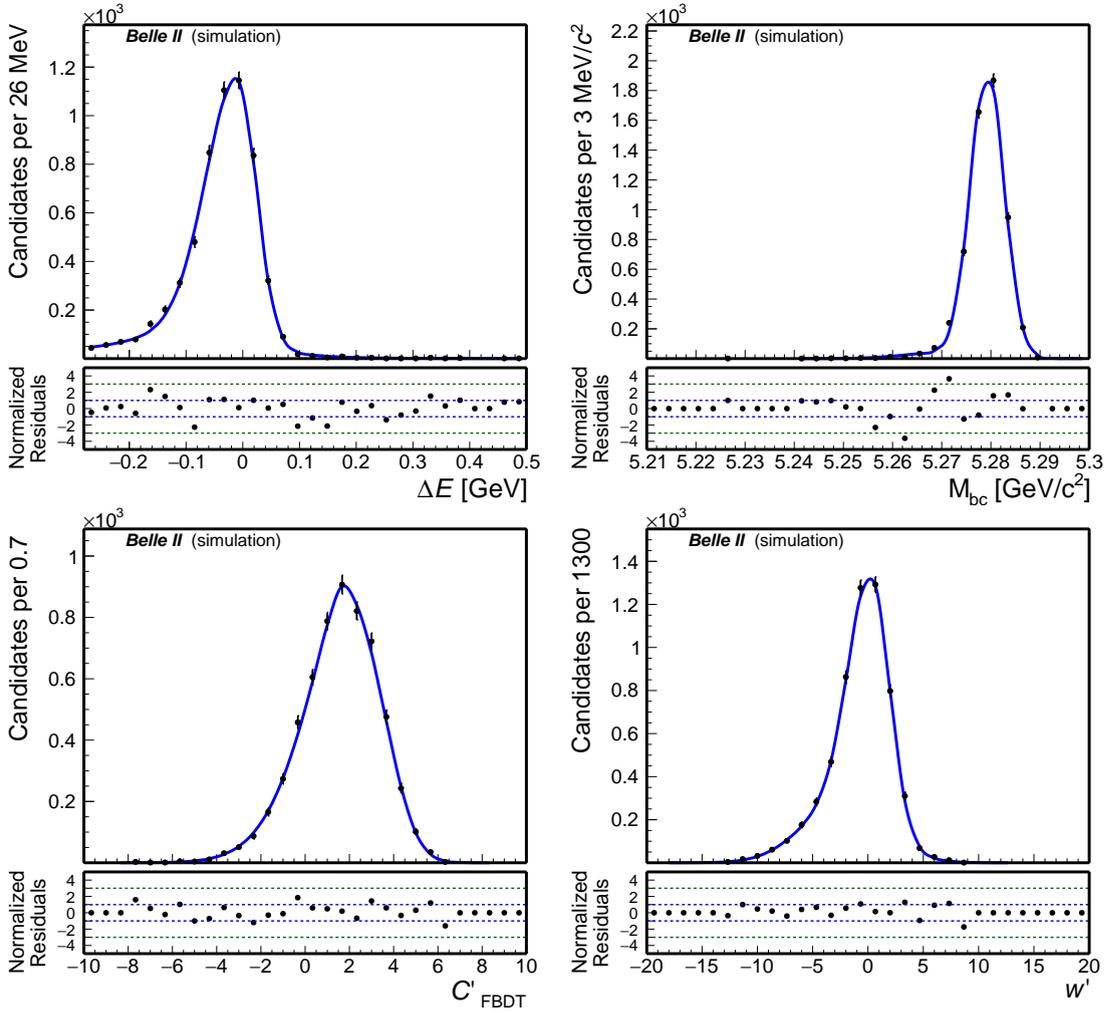


Figure 6.1: Control channel $B^+ \rightarrow K^+\pi^0$ signal modeling: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated signal. The corresponding model shapes are overlaid (solid).

For candidates reconstructed from $B\bar{B}$ background events, the single-event likelihood is factorized as

$$p_{B\bar{B}} = p_{B\bar{B}}(\Delta E)p_{B\bar{B}}(M_{bc})p_{B\bar{B}}(C'_{\text{FBDT}}|w')p_{B\bar{B}}(w'), \quad (6.3)$$

where the observable ΔE is modeled empirically as the sum of three Gaussian functions, while M_{bc} is modeled empirically as the sum of a Gaussian function, a Johnson function, and an Argus function. The continuum suppression output is described with the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7), with parameters determined independently from a modeling fit in simulation in three separate intervals of w' ; w' is also described with the sum of a Gaussian function and a bifurcated Gaussian. Figure 6.2 shows that the conditional p.d.f. model chosen to describe the observables are adequate. The ΔE distribution show a minor mismodeling, with various peaking contributions not precisely described by the model. We expect this to have negligible effect on the results, given the small expected contamination from $B\bar{B}$ decays in data.

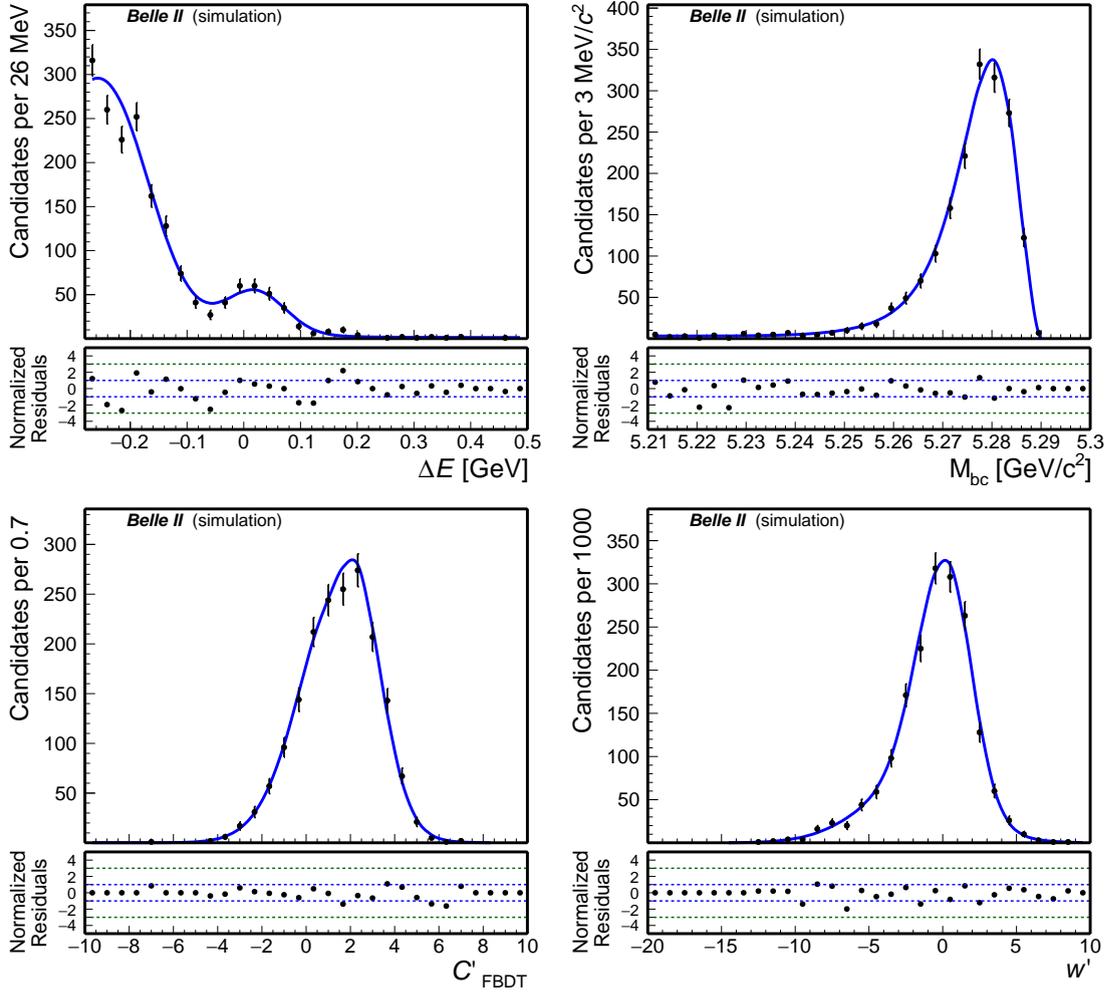


Figure 6.2: Control channel $B\bar{B}$ modeling: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated $B\bar{B}$ background. The corresponding model shapes are overlaid (solid).

Finally, for candidates reconstructed from continuum background, the single-event likelihood is factorized as

$$p_{\text{cont}} = p_{\text{cont}}(\Delta E|w')p_{\text{cont}}(M_{bc})p_{\text{cont}}(C'_{\text{FBDT}})p_{\text{cont}}(w'). \quad (6.4)$$

The function $p(\Delta E|w')$ is a conditional p.d.f. that accounts for the dependence between the two observables: the observable ΔE is modeled empirically as an exponential function, with parameters determined independently from a modeling fit in simulation in three separate intervals of w' ; $p(w')$ is described with the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7). The beam-constrained mass is modeled with the sum of two Argus functions. The continuum suppression is modeled with the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7). Figure 6.3 shows that the models chosen to describe the observables are adequate. The continuum model parameters are left free in data.

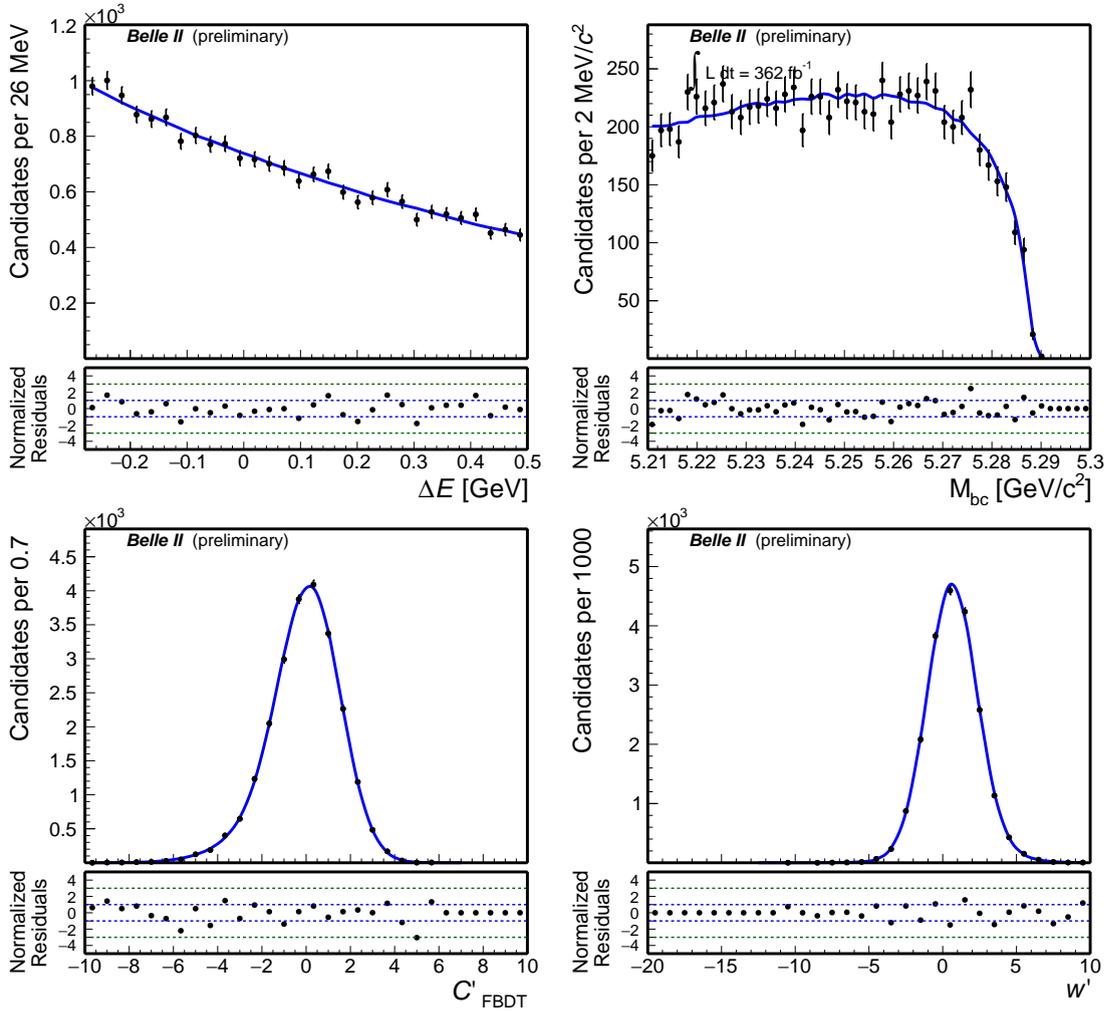


Figure 6.3: Control channel continuum modeling: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated continuum background. The corresponding model shapes are overlaid (solid).

Component	ΔE	M_{bc}	C'_{FBDT}	w'
$B^+ \rightarrow K^+\pi^0$	G+BG M_{bc}	2G	G+BG w'	G+BG
$B\bar{B}$	3G	G+J	G+BG w'	G+BG
Continuum	Exp w'	Double Argus	G+BG	G+BG

Table 6.1: Summary of fit models for the $B^+ \rightarrow K^+\pi^0$ control channel. The shorthand "2G" means sum of two Gaussian functions, "3G" means sum of three Gaussian functions, "G+BG" indicates the sum of a Gaussian function and a bifurcated Gaussian function, "G+J" indicates the sum of a Gaussian function and a Johnson function, and "Exp" indicates an exponential function.

6.2.3 $B^+ \rightarrow K^+\pi^0$ results and model correction extraction

The likelihood of the $B^+ \rightarrow K^+\pi^0$ channel is

$$\begin{aligned} \mathcal{L}_i(f_{\text{sig}}, f_{B\bar{B}}, \mathcal{A}_{CP}|q, w, \Delta E, M_{bc}, C'_{\text{FBDT}}) = & N_{\text{tot}} [f_{\text{sig}}(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))\mathcal{A}_{CP}) \\ & \cdot p_{\text{sig},i}(\Delta E, M_{bc}, C'_{\text{FBDT}}, w') \\ & + f_{B\bar{B}} \left(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))\mathcal{A}_{\text{eff}}^{B\bar{B}}\right) \\ & \cdot p_{B\bar{B},i}(\Delta E, M_{bc}, C'_{\text{FBDT}}, w') \\ & + (1 - f_{\text{sig}} - f_{B\bar{B}}) \left(1 + q(1 - 2(k^q w))\mathcal{A}_{\text{eff}}^{q\bar{q}}\right) \\ & \cdot p_{\text{cont},i}(\Delta E, M_{bc}, C'_{\text{FBDT}}, w')], \end{aligned}$$

To check the estimator properties, I generate 1000 toy samples by drawing the true yield of each component according to a Poisson distribution centered at the true values expected from simulation. The proportion between B^+ and B^- candidates is given by the generated \mathcal{A}_{CP} value. Then, I fit the composition of each toy as if they were data, and construct the pull-value distributions of the fit results.

Figure 6.4 shows pull distributions for $\mathcal{B}(B^+ \rightarrow K^+\pi^0)$ and $\mathcal{A}_{CP}(B^+ \rightarrow K^+\pi^0)$. The means are compatible with zero, showing no intrinsic bias in the fit. The widths of the distributions are compatible with unity, showing proper estimation of the uncertainty.

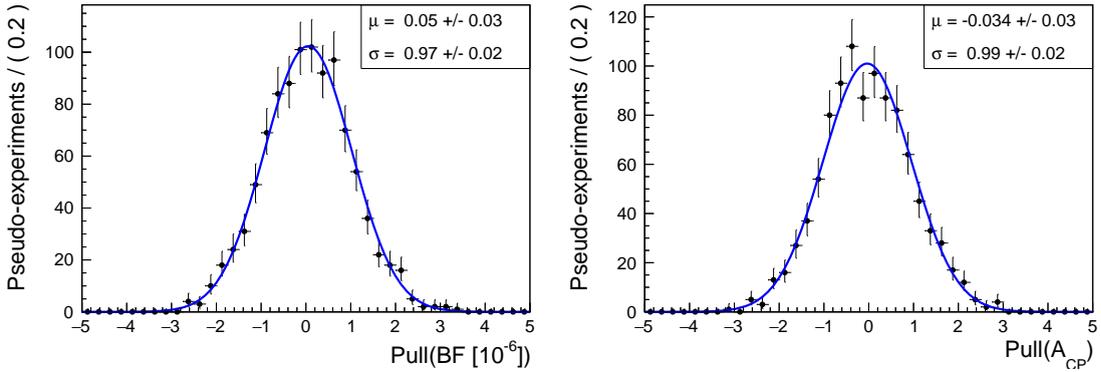


Figure 6.4: Pull distributions of (left) branching fraction and (right) CP -violating asymmetry of $B^+ \rightarrow K^+\pi^0$ decays. Results of fits to Gaussian functions are overlaid (solid).

I perform the fit to the $B^+ \rightarrow K^+\pi^0$ candidates reconstructed in realistic simulation. Figure 6.5 shows $B^+ \rightarrow K^+\pi^0$ distributions with fit projections overlaid. A signal peak is observed at $\Delta E \approx 0$, while a structure due to $B\bar{B}$ decays is visible at negative ΔE values. In the continuum suppression distribution, continuum tends to peak at $C'_{\text{FBDT}} \approx 0$, while $B\bar{B}$ tends to peak at $C'_{\text{FBDT}} \approx 2$. In the M_{bc} distribution, a clear B meson peak is visible, due to both signal and $B\bar{B}$ decays. Table 6.2 shows the results obtained in simulation, together with their true values.

A few model adjustments are needed before the fit on data. Known miscalibrations in the ECL energy-reconstruction result in deformations of the ΔE and M_{bc} shapes, which are correlated with the reconstructed π^0 mesons energies. Hence, I allow for additional model flexibility by introducing an additive constant μ_C and a multiplicative factor σ_C for all peaking components (Gaussians or Johnson) to enable the fit to capture differences between data and simulation in ΔE and M_{bc} resolutions and global scales; these constants are common to all components and incorporated in the models by transforming each mean

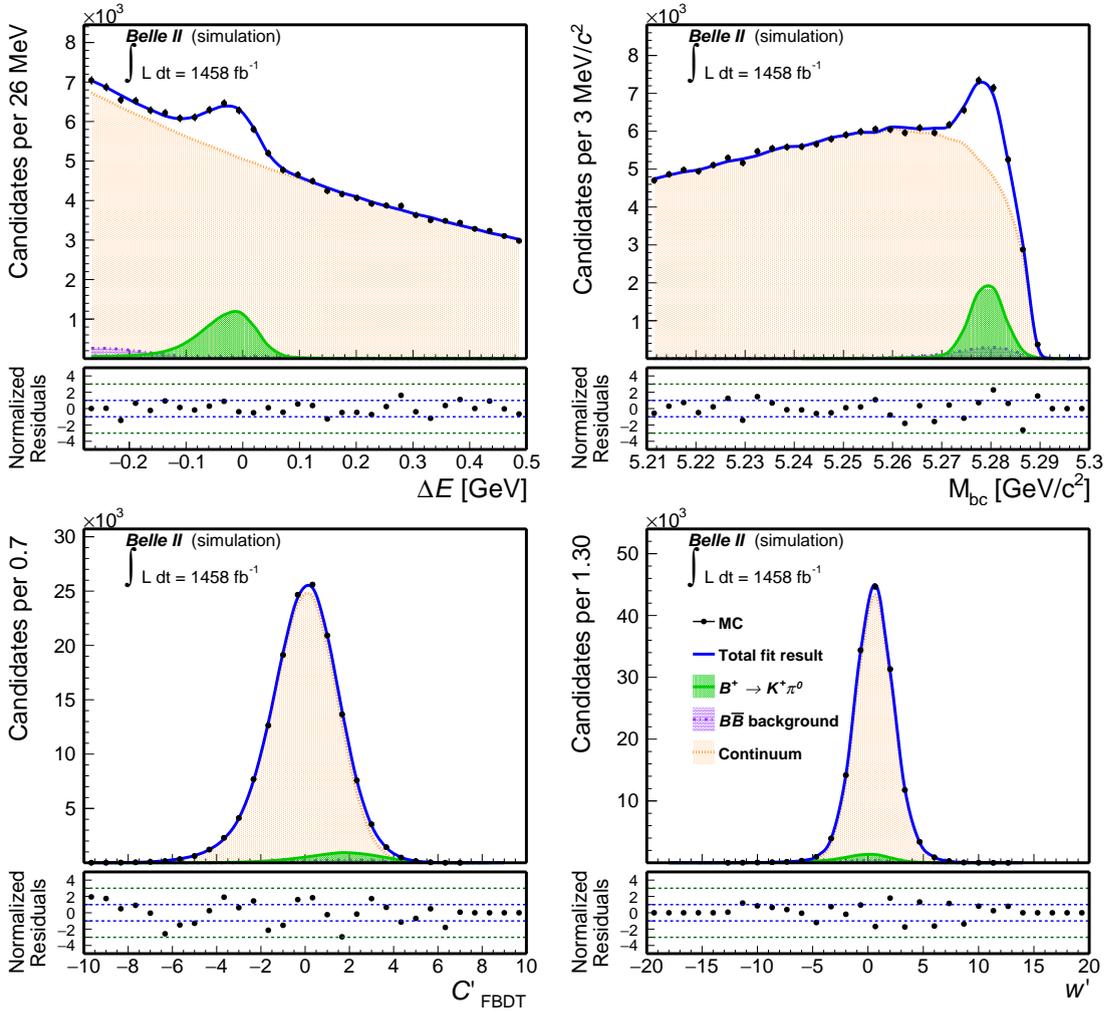


Figure 6.5: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for $B^+ \rightarrow K^+\pi^0$ decays reconstructed in Belle II simulated data corresponding to 1.4 ab^{-1} . Fit projections are overlaid (blue solid line).

and width as $\mu \rightarrow \mu + \mu_C(\Delta E, M_{bc})$ and $\sigma \rightarrow \sigma \cdot \sigma_C(\Delta E, M_{bc})$, and freely determined by the $B^+ \rightarrow K^+\pi^0$ fit.

In addition, I correct for possible differences between data and simulation for the π^0 efficiency and the continuum-suppression efficiency, as described in Sections 8.4 and 8.3 and for possible differences between data and simulation in particle identification selection, using $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ decays reconstructed in collision data and simulation.

Figure 6.6 shows $B^+ \rightarrow K^+\pi^0$ distributions in the full Belle II data sample with fit projections overlaid. The observed features are similar to those in simulation. The projections of the fit show good agreement with data after adding flexibility to the signal model. Table 6.3 shows the results together with their known values [11], with which they agree. This shows that the fit procedure, the models chosen to describe the observables, which also account for dependencies, and the continuum shape determination from data, are all working accurately on a sample mainly composed by continuum as my signal sample.

I use the calibration factors μ and σ obtained in the fit (shown in Table 6.4) to correct the model in the fit of $B^0 \rightarrow \pi^0\pi^0$ composition.

	Simulation fit	True value
$\mathcal{B} (\cdot 10^{-6})$	12.19 ± 0.23	11.97
A_{CP}	-0.042 ± 0.038	-0.028

Table 6.2: Comparison of results for $B^+ \rightarrow K^+\pi^0$ decays reconstructed in Belle II simulated data with true values. Uncertainties are statistical only.

	Data fit	PDG value
$\mathcal{B} (\cdot 10^{-6})$	14.3 ± 0.5	12.9 ± 0.5
A_{CP}	0.097 ± 0.072	0.037 ± 0.021

Table 6.3: Branching fraction and CP -violating asymmetry values obtained from $B^+ \rightarrow K^+\pi^0$ decays reconstructed in Belle II full data sample, together with their known values [11].

6.3 Fit validation using $B^0 \rightarrow \bar{D}^0\pi^0$ decays

A total of 600 $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays are expected in data, which is a factor of approximately five larger than our expected signal yield. I apply the full selection and fit to the $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ channel, adapted to the unavoidable differences in topology and final states.

6.3.1 $B^0 \rightarrow \bar{D}^0\pi^0$ selection

The selection is the same as for the signal, except for the following variations due to the differences between the two decays: I apply a particle-identification selection for the kaon to suppress combinatorial background from pions; I relax the selections on the π^0 meson coming from the B^0 meson, since its energy is lower than in $B^0 \rightarrow \pi^0\pi^0$ decays. I require the $K^+\pi^-\pi^0$ mass to be in the $1.84 < m(K^+\pi^-\pi^0) < 1.88$ GeV/ c^2 range; I require ΔE to be in the $-0.15 < \Delta E < 0.20$ GeV range, to remove as much $B\bar{B}$ background as possible. Finally, I fix the continuum background parameters in the fit. I choose randomly one candidate per event, as done in the $B^0 \rightarrow \pi^0\pi^0$ analysis.

6.3.2 $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ likelihood modeling

The $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ composition is statistically determined with a maximum likelihood fit to the unbinned distributions of discriminating observables as similar as pos-

	Shift	Resolution factor
ΔE	-5.94 ± 1.78 MeV	1.018 ± 0.040
M_{bc}	-0.01 ± 0.10 MeV/ c^2	0.976 ± 0.036

Table 6.4: Results for shifts and resolution factors of the signal peaks in the ΔE and M_{bc} obtained in the $B^+ \rightarrow K^+\pi^0$ control channel.

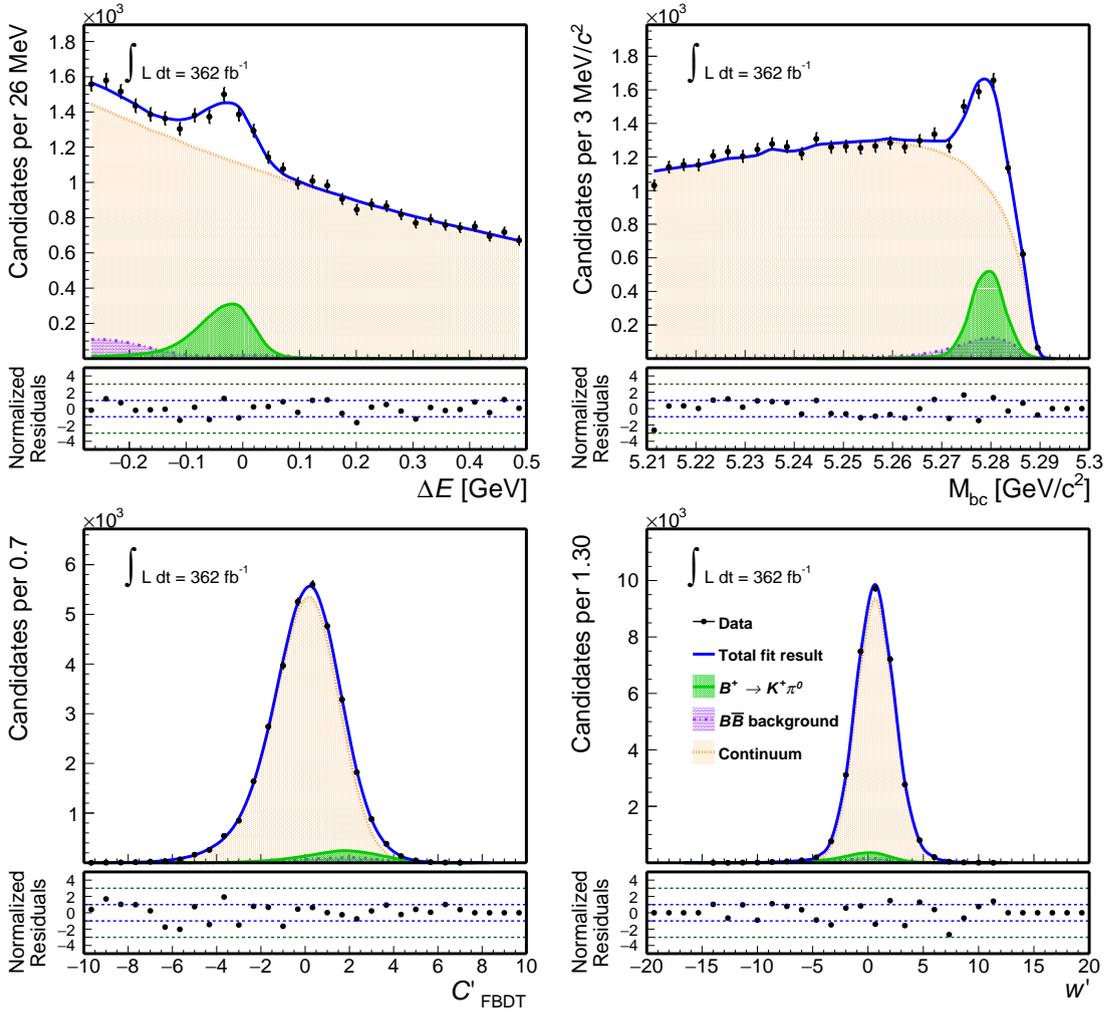


Figure 6.6: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for $B^+ \rightarrow K^+ \pi^0$ decays reconstructed in Belle II data. Fit projections are overlaid (blue solid line).

sible to the fit of the signal sample. The single event likelihood is

$$\mathcal{L}_i = N_{\text{tot}} \left[f_{\text{sig}} p_{\text{sig},i} + f_{B\bar{B}} p_{B\bar{B},i} + (1 - f_{\text{sig}} - f_{B\bar{B}}) p_{\text{cont},i} \right], \quad (6.5)$$

where p_{sig} describes the probability density function of the observables for the signal events, $p_{B\bar{B}}$ for the $B\bar{B}$ background events, and p_{cont} for the continuum events; N_{tot} is the total number of observed candidates, which is fixed, and f_j corresponds to the fraction of each component, to be determined by the fit.

The shape for each component is determined from modeling fits to realistically simulated samples. Continuum background shapes are obtained from the generic simulated sample, as the fraction of continuum in this sample is small and there is no need for a more refined determination of the parameters in sideband data.

The single event likelihood for signal is factorized, as for the $B^0 \rightarrow \pi^0 \pi^0$ signal, as

$$p_{\text{sig}} = p_{\text{sig}}(\Delta E | M_{bc}) p_{\text{sig}}(M_{bc}) p_{\text{sig}}(C'_{\text{FBDT}} | w') p_{\text{sig}}(w'), \quad (6.6)$$

where the function $p(\Delta E | M_{bc})$ is a conditional p.d.f. that accounts for the dependence between the two observables: the observable ΔE is modeled empirically as the sum of

a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7) but with different parameters in three separate intervals of M_{bc} ; $p(M_{bc})$ is modeled empirically as the sum of two Gaussian functions. The model is adequate for all observables, as shown in Fig. 6.7.

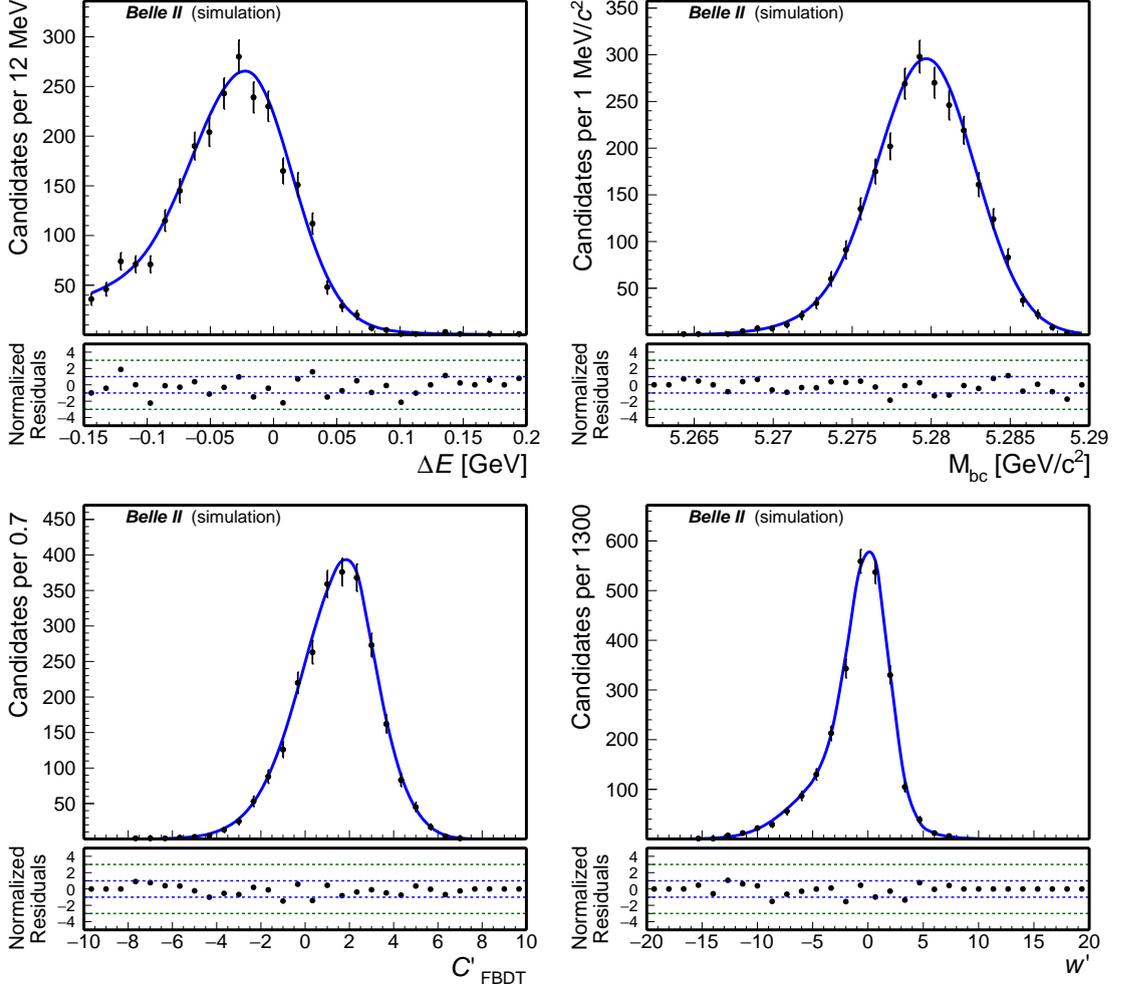


Figure 6.7: Control channel $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ signal modeling: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated signal. The corresponding model shapes are overlaid (solid).

For candidates reconstructed from $B\bar{B}$ background events, the single-event likelihood is factorized as

$$p_{B\bar{B}} = p_{B\bar{B}}(\Delta E|M_{bc})p_{B\bar{B}}(M_{bc})p_{B\bar{B}}(C'_{\text{FBDT}}|w')p_{B\bar{B}}(w'). \quad (6.7)$$

In this case, the dependence between ΔE and M_{bc} cannot be neglected, and is taken into account using conditional functions. The observable ΔE is modeled empirically as the sum of three Gaussian functions with parameters determined independently from fits to simulation in three separate intervals of M_{bc} , while M_{bc} is modeled empirically as the sum of a Gaussian function, a Johnson function, and an Argus function. The continuum suppression output is described with the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7) but with parameters determined independently from a modeling fit in simulation in three separate intervals of w' ; w' is also described with the sum of a Gaussian

function and a bifurcated Gaussian. Figure 6.8 shows that the p.d.f. models chosen to describe the observables are adequate.

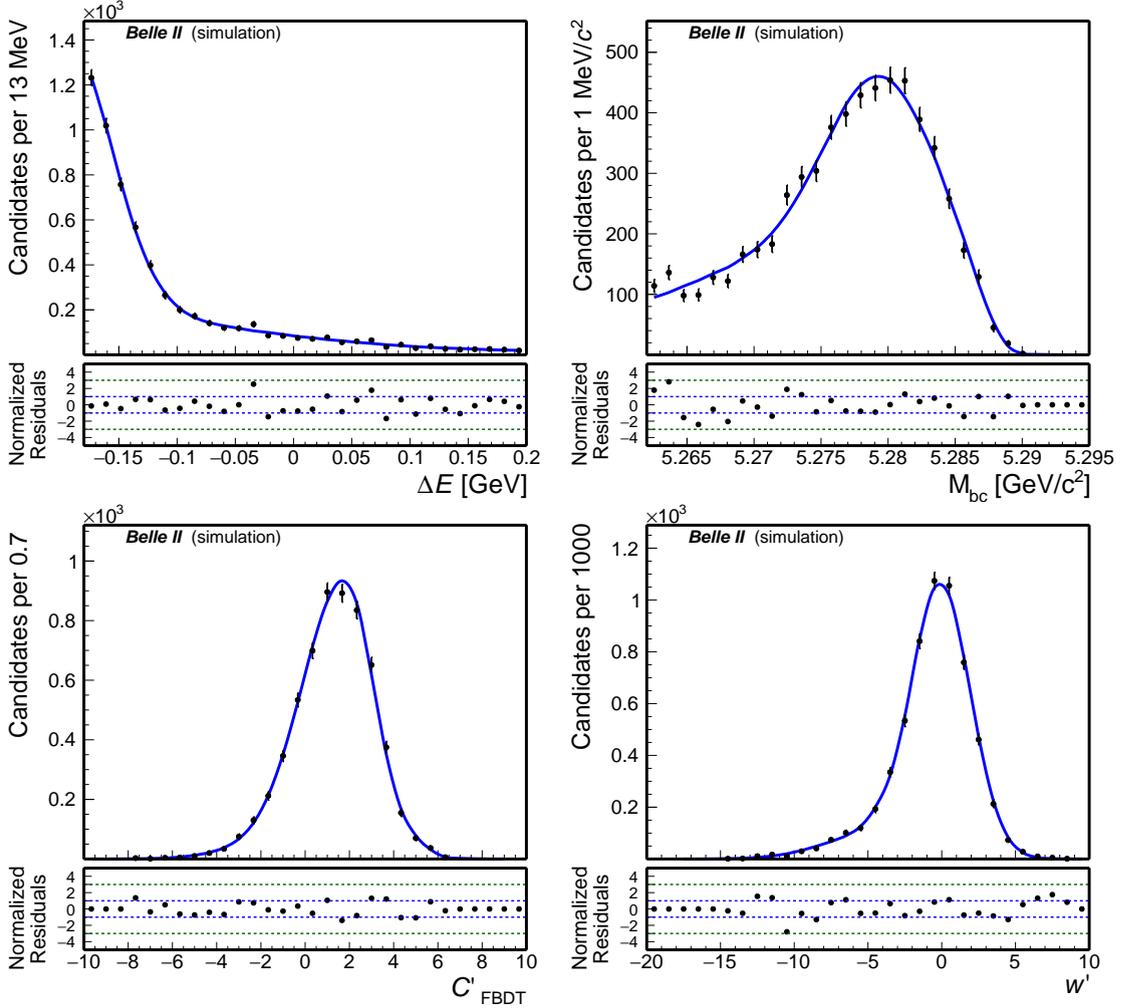


Figure 6.8: Control channel $B\bar{B}$ modeling: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated $B\bar{B}$ background. The corresponding model shapes are overlaid (solid).

Finally, for candidates reconstructed from continuum background, the single-event likelihood is factorized as

$$p_{\text{cont}} = p_{\text{cont}}(\Delta E|w)p_{\text{cont}}(M_{bc})p_{\text{cont}}(C'_{\text{FBDT}})p_{\text{cont}}(w). \quad (6.8)$$

The function $p(\Delta E|w')$ is a conditional p.d.f. that accounts for the dependence between the two observables: the observable ΔE is modeled empirically as a straight line, with parameters determined independently from a modeling fit in simulation in three separate intervals of w' ; $p(w')$ is described with the sum of a Gaussian function and a bifurcated Gaussian, similar to Eq.(5.7), but with independent parameters. The continuum suppression is modeled with the sum of a Gaussian function and a bifurcated Gaussian (Eq.(5.7)). Table 6.5 shows the models for all components for the fit to $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays. Figure 6.9 shows that the models chosen to describe the observables are adequate.

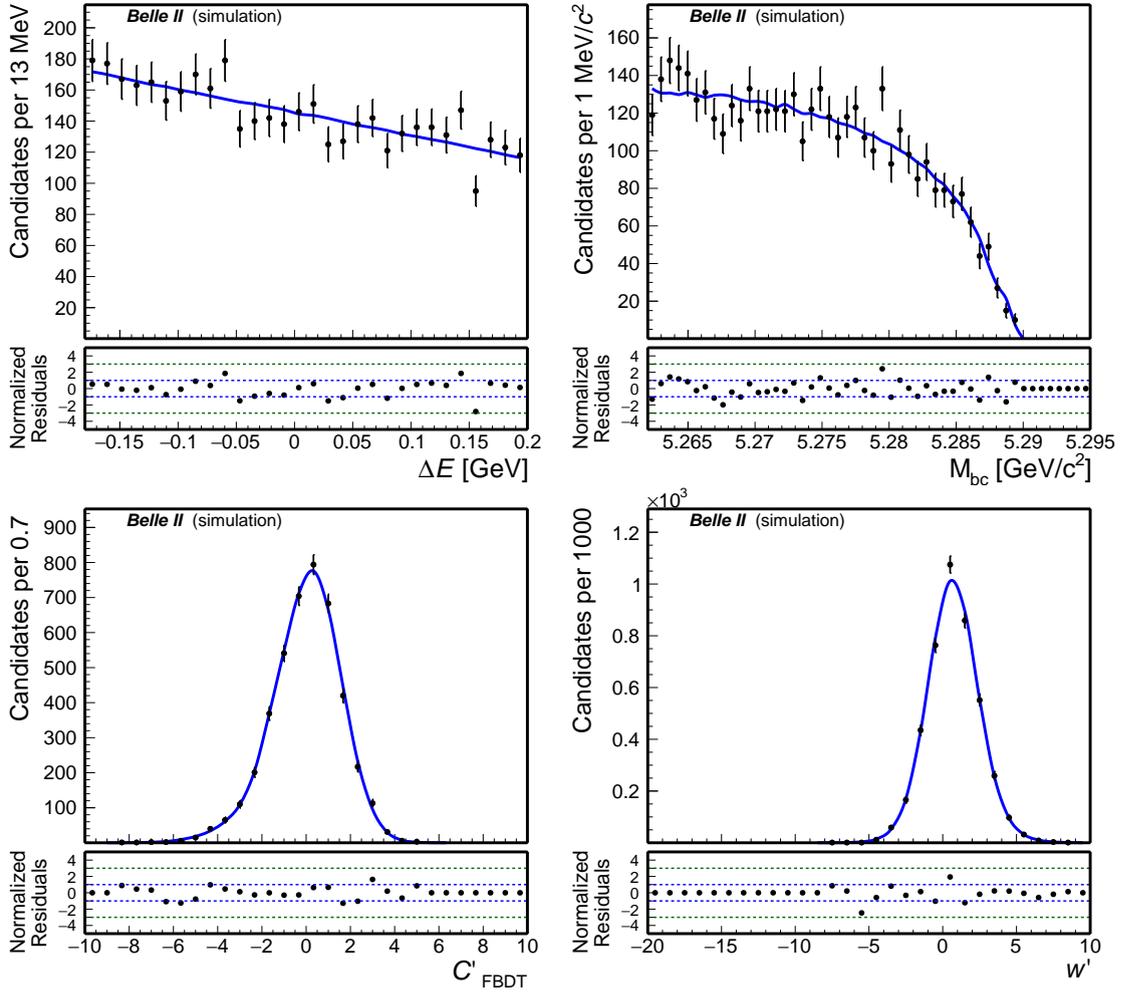


Figure 6.9: Control channel continuum modeling: distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for simulated continuum background. The corresponding model shapes are overlaid (solid).

Component	ΔE	M_{bc}	C'_{FBDT}	w'
$B^0 \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^- \pi^0) \pi^0$	G+BG M_{bc}	2G	G+BG w'	G+BG
$B\bar{B}$	G+BG M_{bc}	G+J	G+BG w'	G+BG
Continuum	Pol1 w'	Double Argus	G+BG	G+BG

Table 6.5: Summary of fit models for the $B^0 \rightarrow \bar{D}^0 (\rightarrow K^+ \pi^- \pi^0) \pi^0$ control channel. The shorthand "2G" means sum of two Gaussian functions, "G+BG" indicates the sum of a Gaussian function and a bifurcated Gaussian function, "G+J" indicates the sum of a Gaussian function and a Johnson function, and "Pol1" indicates a straight line.

6.3.3 $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ results and model correction extraction

The likelihood of the $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ channel is

$$\begin{aligned} \mathcal{L}_i(f_{\text{sig}}, f_{B\bar{B}}, A_{CP}|q, w, \Delta E, M_{bc}, C'_{\text{FBDT}}) = & N_{\text{tot}} [f_{\text{sig}}(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))A_{CP}) \\ & \cdot p_{\text{sig},i}(\Delta E, M_{bc}, C'_{\text{FBDT}}, w') \\ & + f_{B\bar{B}} \left(1 + q(1 + q\mu)(1 - 2\chi_d)(1 - 2(k^q w))A_{\text{eff}}^{B\bar{B}}\right) \\ & \cdot p_{B\bar{B},i}(\Delta E, M_{bc}, C'_{\text{FBDT}}, w') \\ & + (1 - f_{\text{sig}} - f_{B\bar{B}}) \left(1 + q(1 - 2(k^q w))A_{\text{eff}}^{q\bar{q}}\right) \\ & \cdot p_{\text{cont},i}(\Delta E, M_{bc}, C'_{\text{FBDT}}, w')], \end{aligned}$$

I check the estimator properties as done for the $B^+ \rightarrow K^+\pi^0$ channel.

Figure 6.10 shows pull distributions for $\mathcal{B}(B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0)$ and $A_{CP}(B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0)$ parameters. The widths of the distributions are compatible with one showing proper estimation of the uncertainty. The means are compatible with zero, showing no intrinsic bias in the fit.

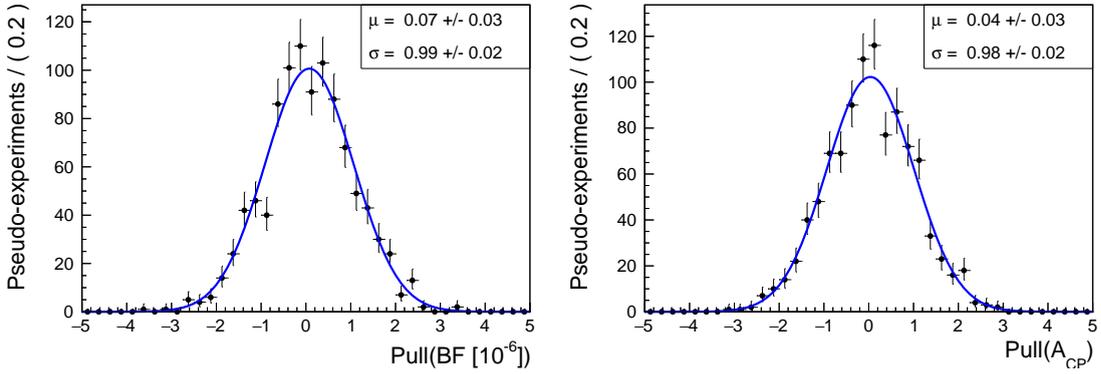


Figure 6.10: Pull distributions of (left) branching fraction and (right) CP -violating asymmetry of $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays. Results of fits to Gaussian functions are overlaid (solid).

Figure 6.11 shows $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ distributions with fit projections overlaid, for realistic Belle II simulation. A signal peak is observed at $\Delta E \approx 0$, while a structure due to background $B\bar{B}$ decays is visible at negative ΔE values. In the continuum suppression output distribution, continuum events tend to peak at $C'_{\text{FBDT}} \approx 0$, while signal events tend to peak at $C'_{\text{FBDT}} \approx 2$. In the M_{bc} distribution, a prominent B^0 meson peak is visible, due to both signal and $B\bar{B}$ decays. The $B\bar{B}$ background is the dominant component, and is mainly due to $B^+ \rightarrow \bar{D}^0(K^+\pi^-\pi^0)\rho^+(\rightarrow \pi^+\pi^0)$ decays where the charged pion from the ρ meson is not reconstructed (64.1%). Table 6.6 shows the results from the fit in simulation together with their true values, with which they agree.

I then apply the fit on the full Belle II dataset, allowing model flexibility through constants determined from data as done in Sec. 6.2. These factors, as determined by the fit, are compatible with those obtained in the $B^+ \rightarrow K^+\pi^0$ case, supporting our strategy of applying them to the signal $B^0 \rightarrow \pi^0\pi^0$ model.

As done for the $B^+ \rightarrow K^+\pi^0$ channel, I correct for possible differences between data and simulation on π^0 efficiency, continuum-suppression efficiency, and particle identification efficiency.

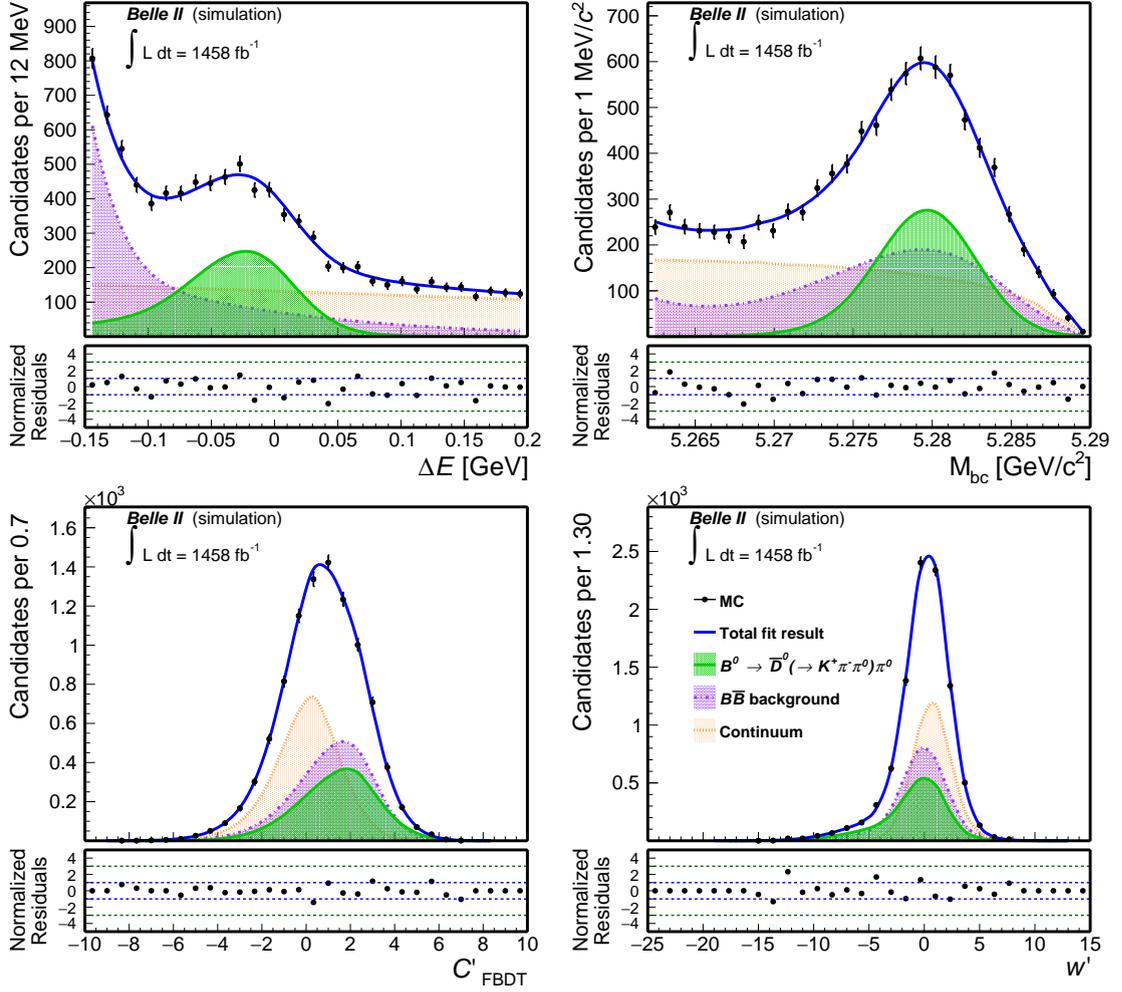


Figure 6.11: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays reconstructed in simulation corresponding to 1.4 ab^{-1} . Fit projections are overlaid (blue solid line).

Figure 6.12 shows $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ distributions in the full Belle II data sample with fit projections overlaid. The observed features are similar to those in simulation. The projections of the fit show good agreement with data after adding flexibility to the signal model. Table 6.7 shows the results together with their known values [11], with which they agree. This shows that the fit procedure, the models chosen to describe the observables, which also account for the additional dependencies for the $B\bar{B}$ component, are working as expected. Importantly, the agreement of A_{CP} with the known value provides proof of the validity of the flavor tagger and of the asymmetry determination in the fit.

	Simulation fit	True value
$\mathcal{B} (\cdot 10^{-6})$	35.9 ± 1.0	36.8
A_{CP}	-0.002 ± 0.059	-0.011

Table 6.6: Comparison of results on simulation with true values for $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays reconstructed in simulation. Uncertainties are statistical only.

	Data fit	PDG value
$\mathcal{B} (\cdot 10^{-6})$	41.40 ± 2.37	38.5 ± 1.1
A_{CP}	0.013 ± 0.102	0.004 ± 0.024

Table 6.7: Branching fraction and CP -violating asymmetry values for $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays reconstructed in the full Belle II data sample, together with their known values [11]. Uncertainties are statistical only.

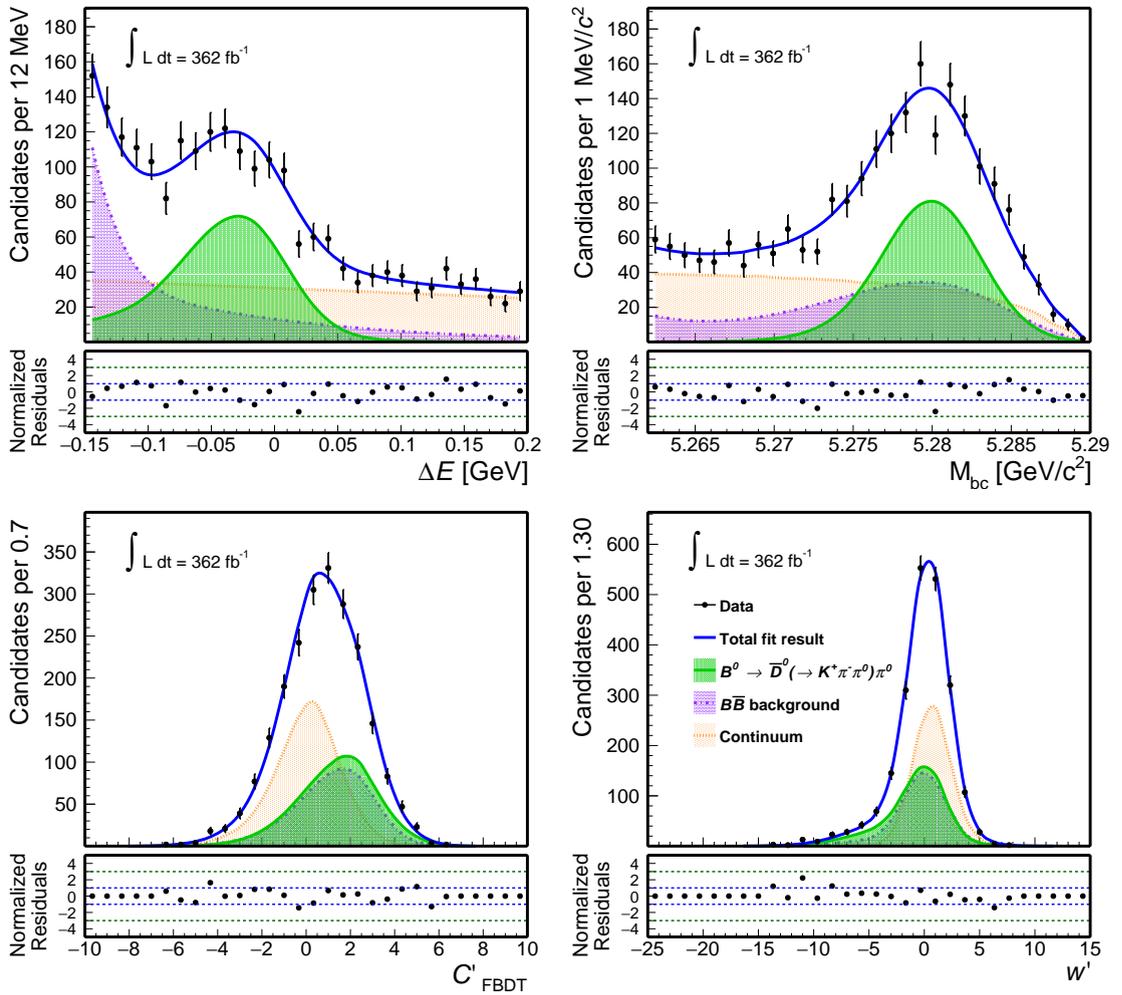


Figure 6.12: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for $B^0 \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^0$ decays reconstructed in the Belle II data set. Fit projections are overlaid (blue solid line).

6.4 Closure test on partial sample

To test the procedure without spoiling the blind status of this thesis' analysis, I perform the analysis on the subsample of signal $B^0 \rightarrow \pi^0\pi^0$ data used for the published Belle II result [54] collected until July 2021.

The data processing version and calibrations have changed between this analysis and the previous one, and the analysis choices depart significantly from those of the previous work. Differences in results induced by the former aspects are unlikely to be major whereas those from the latter aspect might be significant. To achieve the most thorough consistency check, I first try to reproduce, to the maximum extent possible, the previous analysis by mirroring its choices in terms of models etc, and using the same events.

The projections of this fit are shown in Fig. 6.13. The fit describes accurately the data in ΔE , M_{bc} , C'_{FBDT} , and w' . The results, compared with the results obtained in the previous analysis, are shown in Table 6.8.

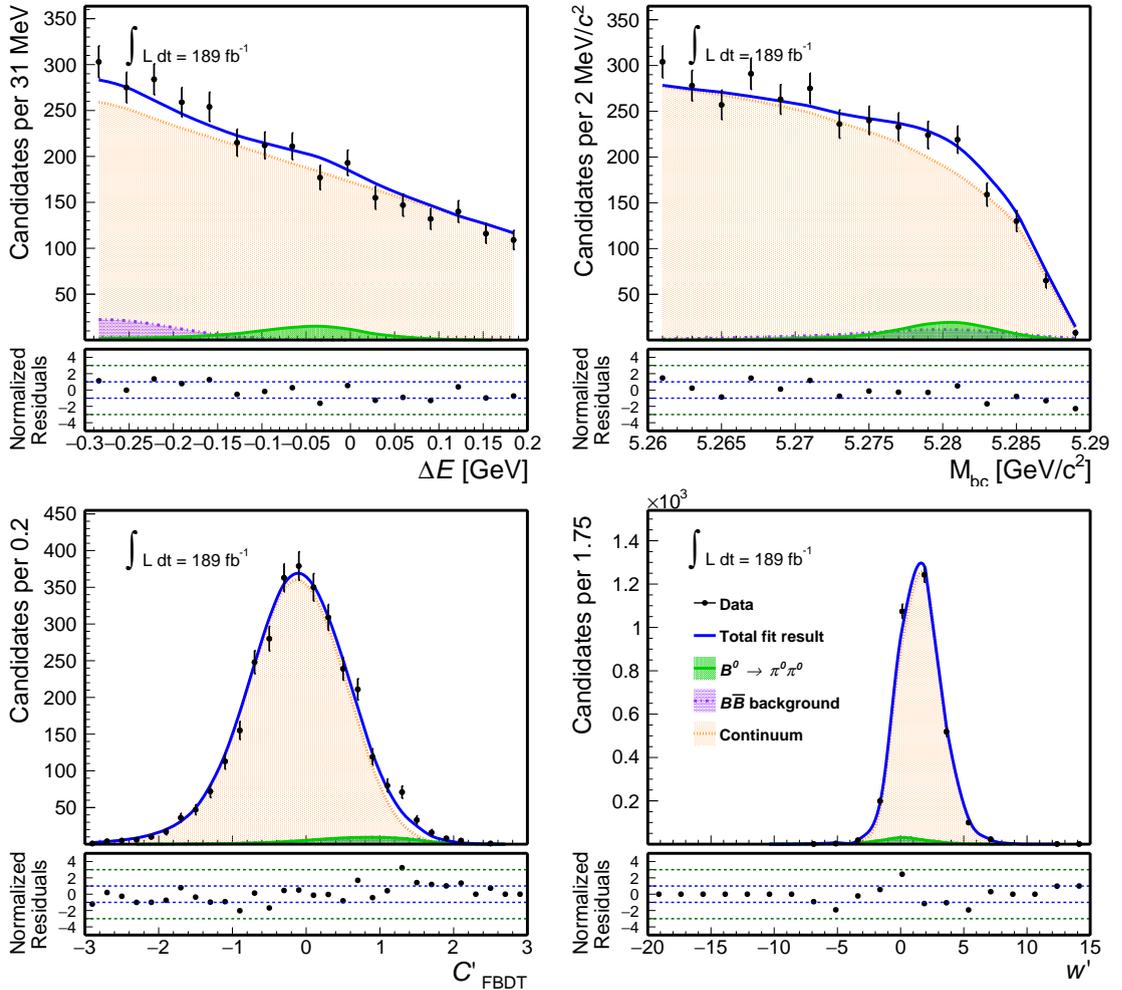


Figure 6.13: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for $B^0 \rightarrow \pi^0\pi^0$ decays reconstructed in the 189 fb^{-1} Belle II data set used for the previous analysis and analyzed mirroring the analysis of Ref. [54]. Fit projections are overlaid (blue solid line).

The results agree well, with small differences expected to be caused by the different

	Data fit	Previous analysis	PDG
$\mathcal{B} (\cdot 10^{-6})$	1.27 ± 0.25	$1.38 \pm 0.27 \pm 0.22$	1.59 ± 0.26
A_{CP}	-0.17 ± 0.46	$-0.14 \pm 0.46 \pm 0.07$	0.33 ± 0.22

Table 6.8: Branching fraction and CP -violating asymmetry results for $B^0 \rightarrow \pi^0\pi^0$ decays reconstructed in the 189 fb^{-1} Belle II data sample and analyzed mirroring the analysis of Ref. [54], together with the results of Ref [54] and their known values. For this work, uncertainties are statistical only.

treatment of the wrong-tag variable, and to the slightly different models used for signal and $B\bar{B}$ background.

As a second check, I perform the analysis of data used for the Belle II publication using my selections, fit strategy, and models, and compare the uncertainties on the results.

The projections of this fit are shown in Fig. 6.14. The fit describes accurately the data in ΔE , M_{bc} , C'_{FBDT} , and w' . The uncertainties on the results, compared with those obtained in the previous analysis, are shown in Table 6.9.

	Data fit	Previous analysis	PDG
$\mathcal{B} (\cdot 10^{-6})$	[xx] ± 0.25	$1.38 \pm 0.27 \pm 0.22$	1.59 ± 0.26
A_{CP}	[xx] ± 0.41	$-0.14 \pm 0.46 \pm 0.07$	0.33 ± 0.22

Table 6.9: Branching fraction and CP -violating asymmetry uncertainties for $B^0 \rightarrow \pi^0\pi^0$ decays reconstructed in the 189 fb^{-1} Belle II data sample and analyzed using my strategy, together with the results of Ref. [54] and known values. Central values are hidden, and uncertainties are statistical only.

The uncertainties on both branching fraction and CP -violating asymmetry are smaller than those of the previous Belle II measurement, showing better quality of my analysis.

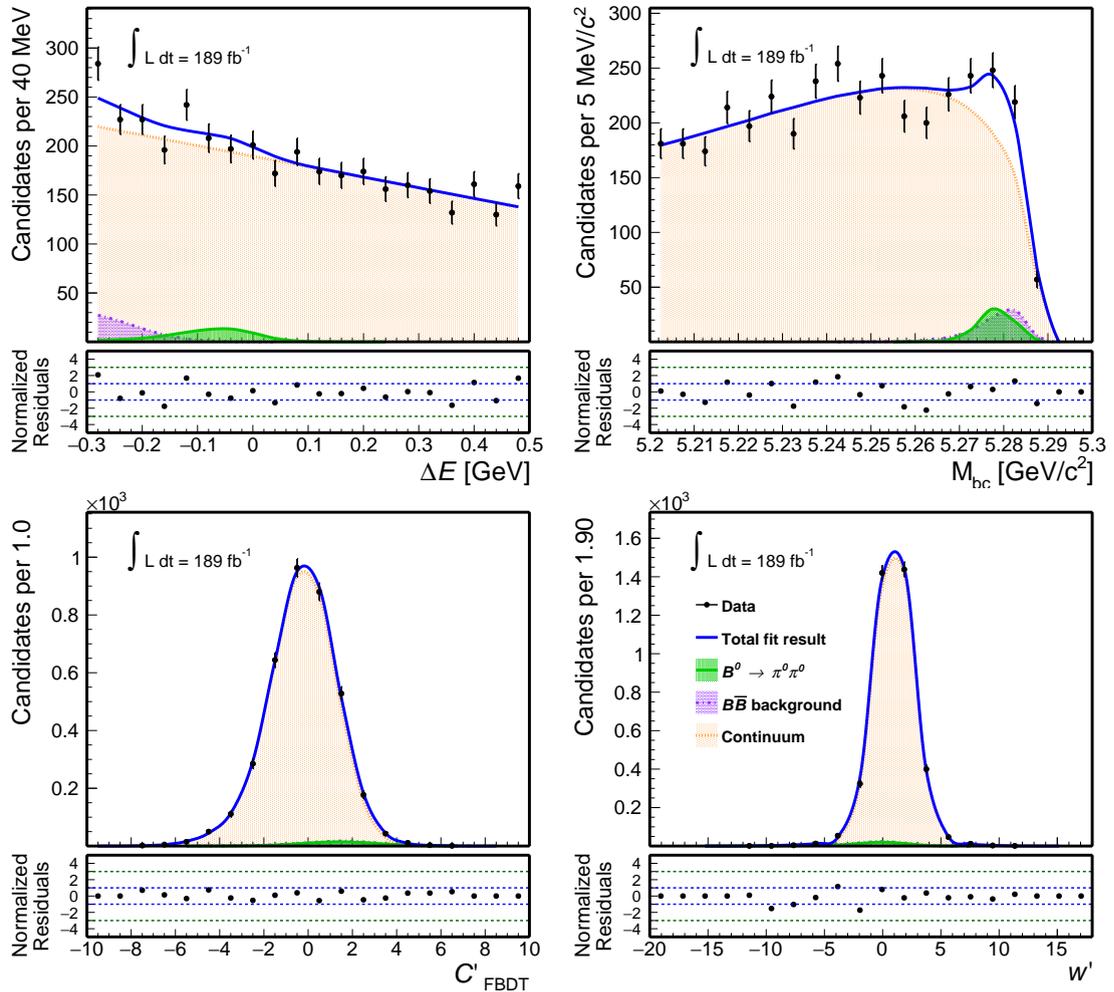


Figure 6.14: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) C'_{FBDT} , and (bottom right) w' for $B^0 \rightarrow \pi^0\pi^0$ decays reconstructed in the 189 fb^{-1} Belle II data set used for the previous analysis, but analyzed using this analysis's strategy. Fit projections are overlaid (blue solid line).

Chapter 7

Fit of $B^0 \rightarrow \pi^0\pi^0$ data

After validating the analysis on simulated samples and control data, I apply the fit to the full sample of signal candidates in data. I show fit projections and results, and compare the latter with existing measurements.

I perform a fit to the Belle II data set corresponding to 387 million of $B\bar{B}$ pairs and containing 7140 $B^0 \rightarrow \pi^0\pi^0$ candidates after the selection. As detailed in the previous sections, the fit features 3 components, 4 observables, and 33 variables. Before the fit I apply the ΔE and M_{bc} additive corrections and multiplicative factors obtained from the control samples as discussed in Sec. 6.2 to the relevant models, with Gaussian-constrained values. The fit runs for eight minutes with HESSE and for about one hour with MINOS on a Xeon E5-2640 v4 2.4 GHz processor.

Full projections onto the fit observables are shown in Fig. 7.1. The sample is dominated by large continuum background component. As expected, a signal peak is observed in the $\Delta E \approx 0$ region and in the M_{bc} mass distributions. The C'_{FBDT} and w' distributions show separation between the continuum and the other components. All distributions are adequately modeled. Projections onto the fit observables in the signal-enriched sample are shown in Fig. 7.2. Signal-enhanced regions are restrictions of the data that expose more transparently the signal region of the multidimensional space of observables the fit actually "sees". They are obtained by restricting the sample to $-0.10 < \Delta E < 0.05$ GeV (except when displaying ΔE), $5.275 < M_{bc} < 5.285$ GeV/ c^2 (except when displaying M_{bc}), and $C'_{\text{FBDT}} > 0$ (except when displaying C'_{FBDT}).

The full set of fit results is reported in Table 7.1. A list of Gaussian-constrained parameters and associated constraints is listed in Table 7.2. Table 7.3 summarizes the estimated correlations.

Table 7.4 shows the principal fit results, compared with previous results from Belle [55], Belle II [54], and BaBar [56]. Only statistical uncertainties are considered. The uncertainty on $\mathcal{B}(B^0 \rightarrow \pi^+\pi^0)$ is on par with the uncertainty of the world-best result by Belle, based on a much larger sample. The apparent lack of improvement in statistical precision over the previous Belle II result is expected because the current analysis includes as statistical uncertainties several effects that were ascribed as systematic uncertainties in the previous analysis. A fair comparison taking into account both statistical and systematic uncertainties is discussed in the next chapters. The statistical uncertainty on $A_{CP}(B^0 \rightarrow \pi^+\pi^0)$ is second only to that of the current world-best result by BaBar, and better than the uncertainties of the earlier Belle II and Belle results. The improvement is mainly due to the novel flavor tagger algorithm used in the analysis.

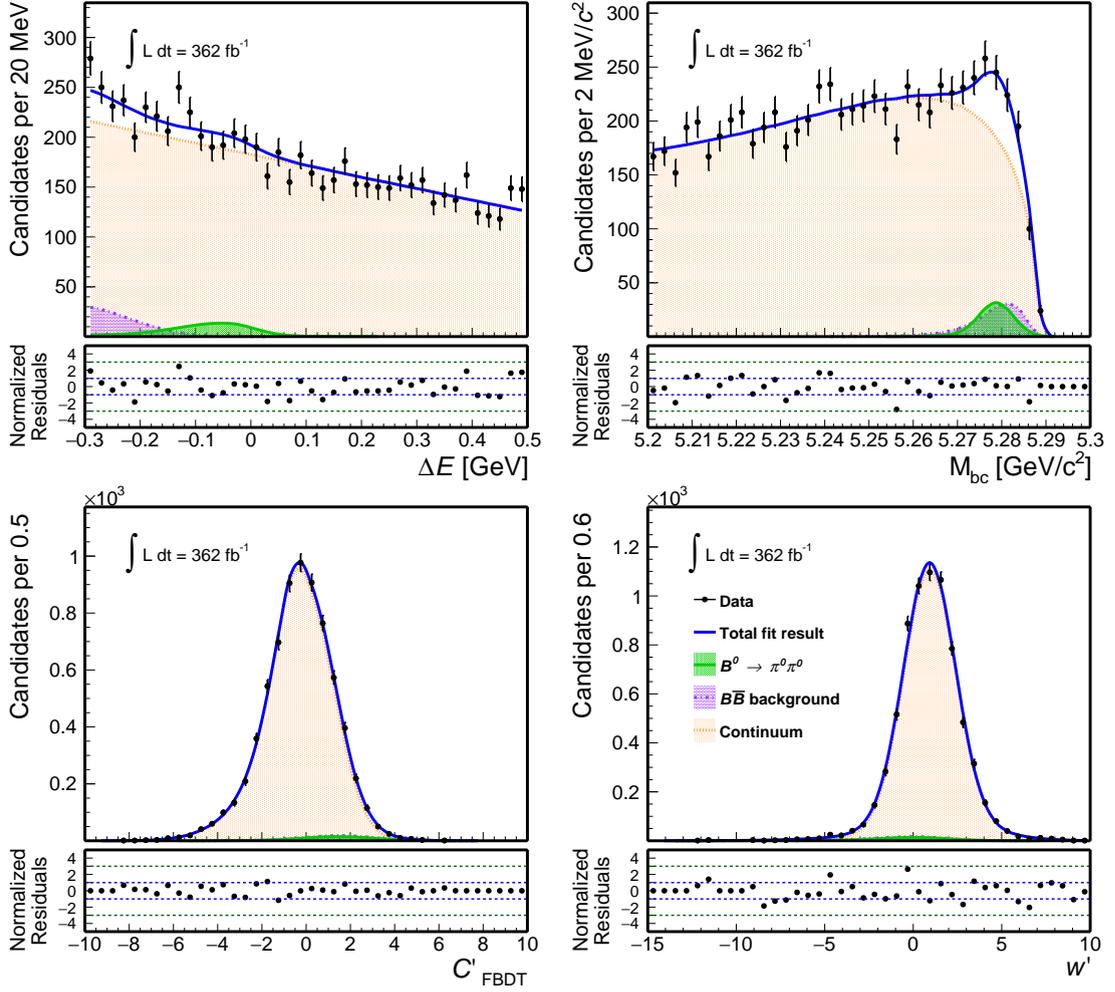


Figure 7.1: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) continuum-suppression output, and (bottom right) wrong-tag for $B^0 \rightarrow \pi^0 \pi^0$ candidates reconstructed in the full Belle II data set collected at the $\Upsilon(4S)$ resonance and corresponding to 387 million $B\bar{B}$ pairs. Fit projections are overlaid.

	This work	Belle [55]	Belle II [54]	BaBar [56]
$\mathcal{B}(B^0 \rightarrow \pi^0 \pi^0)$	$1.26^{+0.20}_{-0.19}$	$1.31 \pm 0.19(0.26)$	$1.38 \pm 0.27(0.19)$	$1.83 \pm 0.21(0.23)$
$A_{CP}(B^0 \rightarrow \pi^0 \pi^0)$	0.06 ± 0.30	$0.14 \pm 0.36(0.50)$	$-0.14 \pm 0.46(0.33)$	$0.43 \pm 0.26(0.29)$

Table 7.4: Comparison of results with previous measurements. Only statistical uncertainties are shown. Values in parentheses indicate projections of the statistical uncertainties in previous measurements to the sample size used in this work.

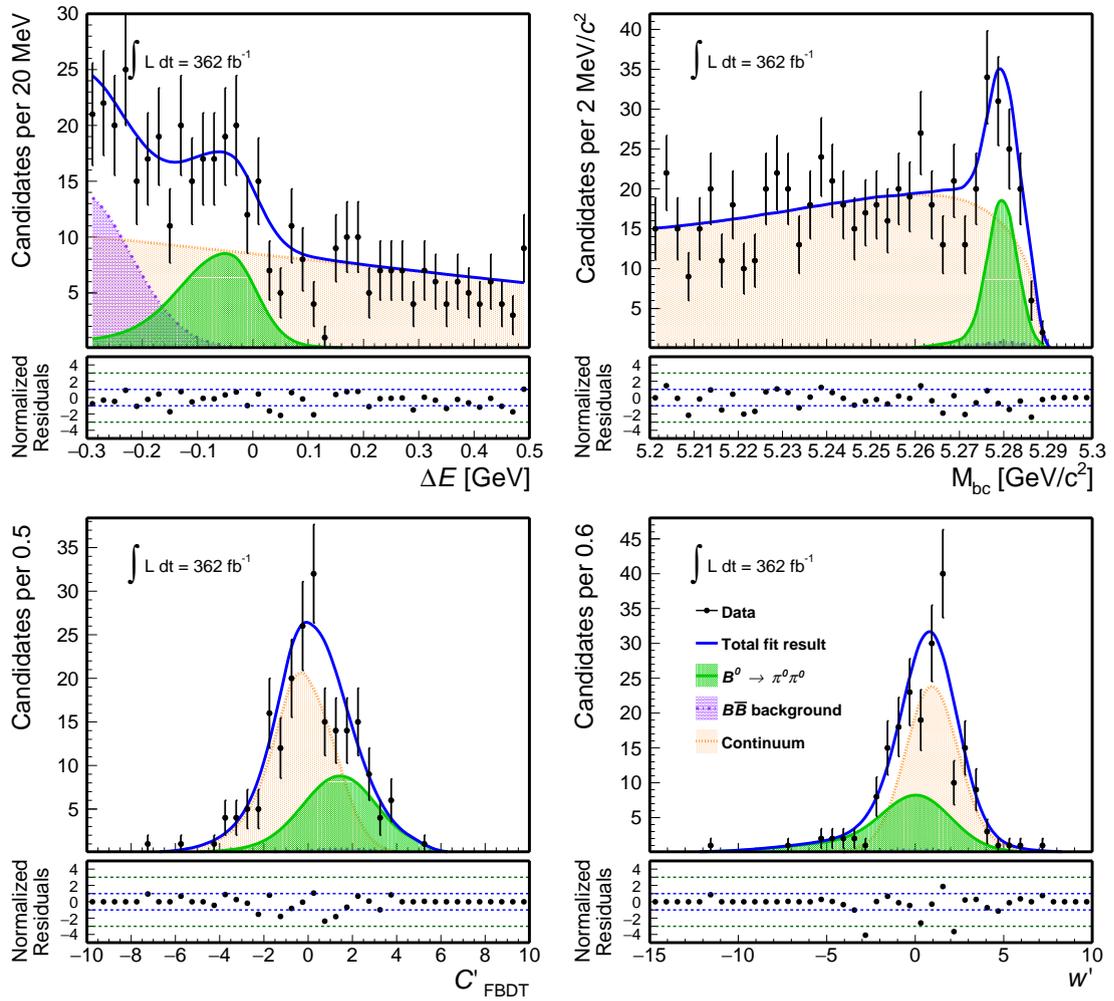


Figure 7.2: Distributions of (top left) ΔE , (top right) M_{bc} , (bottom left) continuum-suppression output, and (bottom right) wrong-tag for $B^0 \rightarrow \pi^0\pi^0$ candidates reconstructed in the full Belle II data set collected at the $\Upsilon(4S)$ resonance and corresponding to 387 million $B\bar{B}$ pairs, and restricted to the signal region to enhance signal visibility. Fit projections are overlaid.

Variable	Fit result
$\mathcal{B}(B^0 \rightarrow \pi^0 \pi^0)$	$(1.261_{-0.185}^{+0.197}) \times 10^{-6}$
Charge-parity violating asymmetry A_{CP}	$0.062_{-0.303}^{+0.300}$
Bottom-antibottom yield $N_{B\bar{B}}$	166_{-20}^{+21}
Effective $B\bar{B}$ flavor asymmetry $A_{\text{eff}}^{B\bar{B}}$	0.012 ± 0.074
Effective continuum flavor asymmetry $A_{\text{eff}}^{q\bar{q}}$	-0.0153 ± 0.0192
Wrong-tag calibration factor k^+	1.028 ± 0.008
Wrong-tag calibration factor k^-	1.029 ± 0.008
Tagging asymmetry μ	0.0015 ± 0.0011
Time-integrated mixing probability χ_d	0.1858 ± 0.0008
Additive correction factor to ΔE	-0.0118 ± 0.0025
Multiplicative correction factor to Δ	1.034 ± 0.039
Additive correction factor to M_{bc}	-0.00002 ± 0.00017
Multiplicative correction factor to M_{bc}	0.95 ± 0.04
Argus fraction f_{Argus}	0.408 ± 0.124
First Argus endpoint Argus	5.2897 ± 0.0003
c_{Argus1}	-2.3 ± 10.8
Second Argus endpoint	5.2883 ± 0.0001
c_{Argus2}	-66.7 ± 9.3
Slope p of ΔE continuum model in wrong-tag bin 1	-0.85 ± 0.08
Slope p of ΔE continuum model in wrong-tag bin 2	-0.64 ± 0.07
Slope p of ΔE continuum model in wrong-tag bin 3	-0.36 ± 0.10
f_{G+BG} of w'	0.100 ± 0.027
Bifurcated Gaussian mean μ in w' model	2.11 ± 0.47
Bifurcated Gaussian left-side width σ_L in w' model	3.46 ± 0.40
Bifurcated Gaussian right-side width σ_R in w' model	2.70 ± 0.26
Gaussian mean μ in w' model	0.93 ± 0.02
Gaussian width σ in w' model	1.45 ± 0.03
Fraction of Bifurcated plus Gaussian f_{G+BG} in C'_{FBDT} model	0.216 ± 0.032
Bifurcated Gaussian mean μ in C'_{FBDT} model	-0.78 ± 0.22
Bifurcated Gaussian left-side width σ_L in C'_{FBDT} model	2.07 ± 0.12
Bifurcated right-side width σ_R in C'_{FBDT} model	0.54 ± 0.19
Gaussian mean μ in C'_{FBDT} model	0.06 ± 0.06
Gaussian σ in C'_{FBDT} model	1.26 ± 0.03

Table 7.1: Results of the fit to data. The parameter f is used in models that are sums of different functions, and represent the fraction of the first function over the total. The symbol BG represents the Gaussian plus bifurcated Gaussian model, while G represents the single Gaussian model.

Variable	Constraint
Effective $B\bar{B}$ flavor asymmetry $A_{\text{eff}}^{B\bar{B}}$	0 ± 0.11
Effective continuum flavor asymmetry $A_{\text{eff}}^{q\bar{q}}$	-0.0153 ± 0.0271
Wrong-tag calibration factor k^+	1.028 ± 0.012
Wrong-tag calibration factor k^-	1.029 ± 0.012
Tagging asymmetry μ	0.0015 ± 0.0016
Time-integrated mixing probability χ_d	0.1858 ± 0.0011
Additive correction factor to ΔE	-0.0118 ± 0.0036
Multiplicative correction factor to Δ	1.035 ± 0.057
Additive correction factor to M_{bc}	-0.00002 ± 0.00026
Multiplicative correction factor to M_{bc}	0.953 ± 0.050

Table 7.2: Gaussian constrained parameters used in the fit, with corresponding constraints.

	$\mathcal{B}(B^0 \rightarrow \pi^0 \pi^0)$	A_{CP}	$N_{B\bar{B}}$	$A_{\text{eff}}^{B\bar{B}}$	$A_{\text{eff}}^{q\bar{q}}$	k^+	k^-	μ	χ_d
$\mathcal{B}(B^0 \rightarrow \pi^0 \pi^0)$	1	+1.6%	+0.6%	-3%	-6.5%	+2.5%	-0.1%	<0.1%	-1.9%
A_{CP}		1	<0.1%	-2.4%	+0.3%	-0.4%	-3.5%	+0.1%	-0.2%
$N_{B\bar{B}}$			1	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%
$A_{\text{eff}}^{B\bar{B}}$				1	+2.5%	-0.2%	+0.1%	<0.1%	<0.1%
$A_{\text{eff}}^{q\bar{q}}$					1	+0.6%	-0.8%	<0.1%	+1%
k^+						1	+1%	<0.1%	+0.1%
k^-							1	<0.1%	+0.1%
μ								1	<0.1%
χ_d									1

Table 7.3: Correlation matrix in data for all fit parameters but the shape parameters.

Chapter 8

Systematic uncertainties

This chapter describes the determination of the systematic uncertainties.

8.1 Introduction

The precision of our results is chiefly limited by the statistical uncertainties associated with the small signal yield and large background. However, it is important to understand and reduce systematic uncertainties for ensuring accurate results and prepare the collaboration for future updates on larger samples.

I consider various sources of systematic uncertainties. These are associated with assumptions made in the analysis, with possible biases due to discrepancies between relevant distributions as observed in collision data and simulation, or to intrinsic uncertainties of external analysis inputs. Whenever a systematic uncertainty is associated with an assumption or an arbitrary choice made in the analysis, I evaluate its impact by using ensembles of simplified simulated experiments, based on sampling the likelihood. Each ensemble contains many simulated experiments that realize a specific configurations for that choice. I repeat the analysis on them using a model implementing our default choice and use the maximum distance between the averages of the results from the alternative configurations and those from the default configuration as systematic uncertainty. This approach reproduces faithfully the process of fitting data generated with an unknown model with an assumed model while avoiding folding-in statistical uncertainties. To correct for possible data-simulation discrepancies, I use control samples reconstructed in collision data and in simulation to estimate correction factors and their uncertainties, which are then propagated in the signal-sample analysis as systematic uncertainties. The systematic uncertainties due to limited knowledge of external analysis inputs are the uncertainties associated to those inputs.

8.2 Systematic uncertainties included in the statistical uncertainty

I include directly in the fit model various nuisance parameters constrained using Gaussian penalties to values determined in data, simulation, or taken from literature, instead of considering them as separate systematic uncertainties. This allows for treating properly and conveniently correlations between sources. Gaussian-constrained parameters included in the fit are

- χ_d ;

- wrong tag calibration factors k^q ;
- tagging efficiency asymmetry μ ;
- effective continuum asymmetry $A_{\text{eff}}^{q\bar{q}}$;
- effective $B\bar{B}$ background asymmetry $A_{\text{eff}}^{B\bar{B}}$;
- additive and multiplicative factors for ΔE and M_{bc} .

8.3 Uncertainty in the neutral-pion reconstruction efficiency

I assess a systematic uncertainty associated with known data-simulation discrepancies in the π^0 reconstruction and selection using the yield of $D^{*-} \rightarrow \bar{D}^0 (\rightarrow K^+\pi^-\pi^0)\pi^-$ decays, normalized to that of $D^{*-} \rightarrow \bar{D}^0 (\rightarrow K^+\pi^-)\pi^-$ decays observed in data and compared to that in simulation.

These modes are chosen because they are low-background, straightforward to reconstruct, abundant, and provide π^0 mesons spanning a broad spectrum of energies. Other methods are described in Appendix B. The observed charm meson yields are expressed as

$$N(\bar{D}^0 \rightarrow K^+\pi^-\pi^0) \propto \mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)\mathcal{B}(\pi^0 \rightarrow \gamma\gamma)\varepsilon_{D^*,\bar{D}^0,K^+,\pi^-, \pi^0}$$

and

$$N(\bar{D}^0 \rightarrow K^+\pi^-) \propto \mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-)\varepsilon'_{D^*,\bar{D}^0,K^+,\pi^-}$$

in terms of known branching fractions and the joint efficiencies, ε and ε' for reconstructing and selecting the respective full decay chain. Here D^{*-} production cross sections and integrated luminosity are omitted as they are equal in both yields. I then assume that each joint efficiency can be approximated as the product of the individual efficiencies for reconstructing each particle in the chain and that the efficiencies for reconstructing charged particles do not depend strongly on their kinematic properties in the probed range,

$$\begin{aligned} N(\bar{D}^0 \rightarrow K^+\pi^-\pi^0) &\propto \mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)\mathcal{B}(\pi^0 \rightarrow \gamma\gamma)\varepsilon_{D^*}\varepsilon_{\bar{D}^0}\varepsilon_{K^+}\varepsilon_{\pi^-}\varepsilon_{\pi^0} \\ N(\bar{D}^0 \rightarrow K^+\pi^-) &\propto \mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-)\varepsilon_{D^*}\varepsilon_{\bar{D}^0}\varepsilon_{K^+}\varepsilon_{\pi^-}. \end{aligned}$$

Hence, the π^0 efficiency is the ratio of observed yields corrected by known branching fractions,

$$\epsilon_{\pi^0} \propto \frac{N(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)}{N(\bar{D}^0 \rightarrow K^+\pi^-)} \cdot \frac{\mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-)}{\mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)\mathcal{B}(\pi^0 \rightarrow \gamma\gamma)}.$$

I determine this efficiency in data and simulation and calculate the ratio between results, as a function of π^0 momentum and polar direction θ .

The selection of charged particles in both decays is identical. The selections on the π^0 mesons are the same as in the signal mode. Candidate K^+ , π^- , and π^0 , when appropriate, are combined to form D^0 meson candidates. The mass of the $\bar{D}^0 \rightarrow K^+\pi^-\pi^0$ candidates is required to be between $1.75 \text{ GeV}/c^2$ and $2.00 \text{ GeV}/c^2$, while the mass of the $\bar{D}^0 \rightarrow K^+\pi^-$ candidates is required to be between $1.80 \text{ GeV}/c^2$ and $1.92 \text{ GeV}/c^2$. Candidate \bar{D}^0 's are then combined with a charged pion to form D^{*-} candidates. The mass differences between D^{*-} and D^0 mesons are required to be $0.1448 < m(K^+\pi^-\pi^0\pi^+) - m(K^+\pi^-\pi^0) < 0.1461 \text{ GeV}/c^2$ and $0.1447 < m(K^+\pi^-\pi^+) - m(K^+\pi^-) < 0.1467 \text{ GeV}/c^2$. These selections

have the same signal efficiency in the two channels. The momentum of the D^{*+} candidate in the center-of-mass frame is required to exceed $2.5 \text{ GeV}/c$, to remove events where lower-energy D^{*+} candidates come from B meson decays.

Example of fits to the $m(K^- \pi^+ \pi^0)$ and $m(K^- \pi^+)$ distributions are shown in Fig. 8.1.

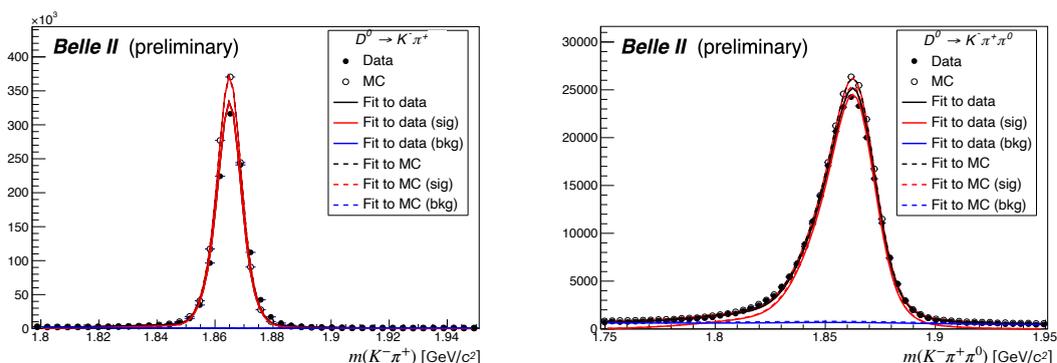


Figure 8.1: Example distributions of (left) $m(K^- \pi^+)$ and (right) $m(K^- \pi^+ \pi^0)$ in (full dots) data and (empty dots) simulation, with fit projections overlaid.

Since the π^0 -reconstruction efficiency is expected to depend on the π^0 kinematic properties and on the detector material traversed in its trajectory, efficiencies are studied as functions of π^0 momentum and polar angle. Figure 8.2 shows the data-to-simulation efficiency ratio resulting from the control channel as a function of momentum and polar-angle. The π^0 -momentum and polar-angle distributions of signal π^0 's implies correction factors of $(1.01 \pm 0.04)\%$ for the lower-energy and $(1.04 \pm 0.04)\%$ for the higher-energy π^0 meson. I correct the signal efficiency by these factors, and I use their summed uncertainty as systematic uncertainty.

8.4 Uncertainty on the continuum-background suppression efficiency

I investigate possible data-simulation discrepancies in the efficiency associated with the continuum-suppression criterion using the control channel $B^+ \rightarrow \bar{D}^0(\rightarrow K^+ \pi^-) \pi^+$. The efficiency results from a simultaneous fit of signal yields in the sample of candidates passing the selection and its complement. I select the control sample through the optimized criteria of the $B^0 \rightarrow \pi^0 \pi^0$ analysis applied on candidates meeting the $K\pi$ -mass restriction $1.84 < m(K\pi) < 1.89 \text{ GeV}/c^2$. Signal yields result from fits to the ΔE distribution, distinguishing signal, continuum background, and misreconstructed $B^+ \rightarrow \bar{D}^0(\rightarrow K^+ \pi^-) K^+$ decays. I model all distributions from simulation, with same shapes for the two subsamples.

Figure 8.3 shows the results of this fit. The resulting selection efficiencies in data and simulation are 0.542 ± 0.009 and 0.574 ± 0.005 , respectively, yielding 0.9420 ± 0.0177 as a ratio. Since the ratio differs from unity, I scale the branching-fraction result in data according to this ratio and take the uncertainty as a systematic uncertainty.

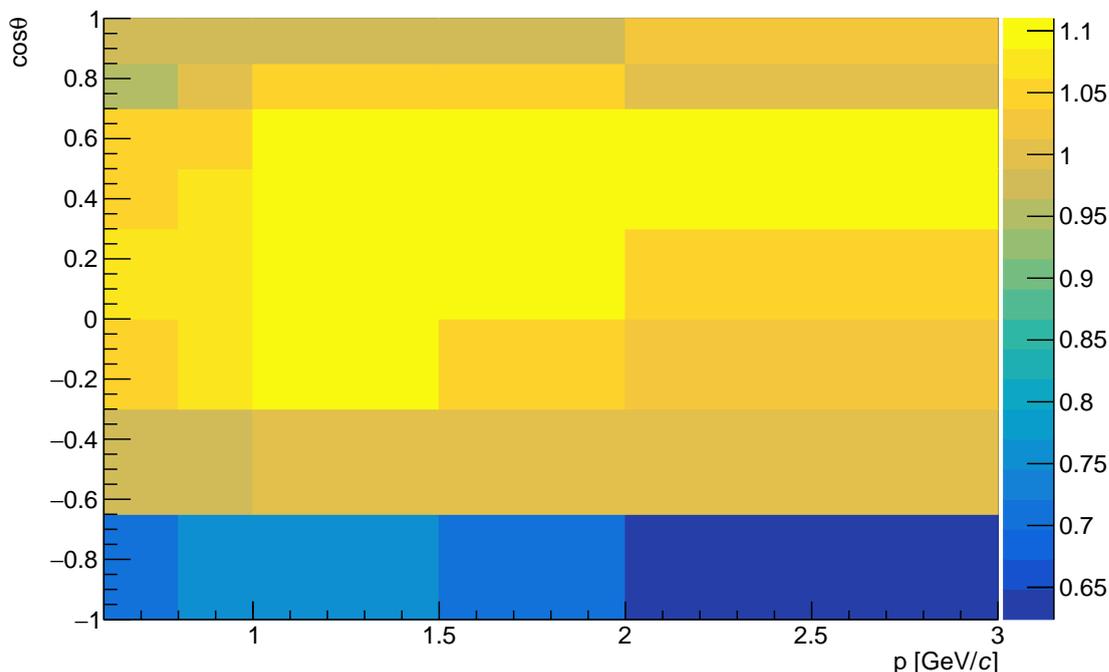


Figure 8.2: Ratios of data-to-simulation efficiency for reconstructing π^0 mesons as functions of momentum and polar angle after applying the π^0 selection of this analysis, except for the π^0 momentum criterion.

8.5 Uncertainty on the $B\bar{B}$ pair yield

The 1.55% uncertainty on the number of $B\bar{B}$ pairs, which includes uncertainties on the $\Upsilon(4S)$ branching fractions, on inputs to the counting (luminosity, efficiency, beam energy spread, etc...), and on limited size of simulated samples, reflects directly as a systematic uncertainty on the branching fraction.

8.6 Assumptions and uncertainties in fit modeling

The fit model is an approximation of the real mechanisms that generated the data. The associated inaccuracies and approximations are a source of systematic uncertainty. I assume that these have two sources (i) uncertainties in the parameters of the chosen model (ii) possibility that alternative empirical shapes that approximate equally well the data could yield different results. To account for the uncertainties in model parameters, I generate ensembles of simplified simulated experiments. Each ensemble implements an alternative model based on a specific variations of the shape parameters within their covariances. I then repeat the analysis on the alternate samples and take as systematic uncertainty the largest difference between the average of the results observed in the default model and in the alternatives.

A conceptually similar procedure is done for several choices of alternative shapes, including the sum of three Gaussians for the signal or $B\bar{B}$ functions for the ΔE and M_{bc} , exponential functions for the ΔE continuum, and others. Resulting uncertainties range between 0.9% and 1.7% (relative) for the branching fractions and from 0.021 to 0.034 for

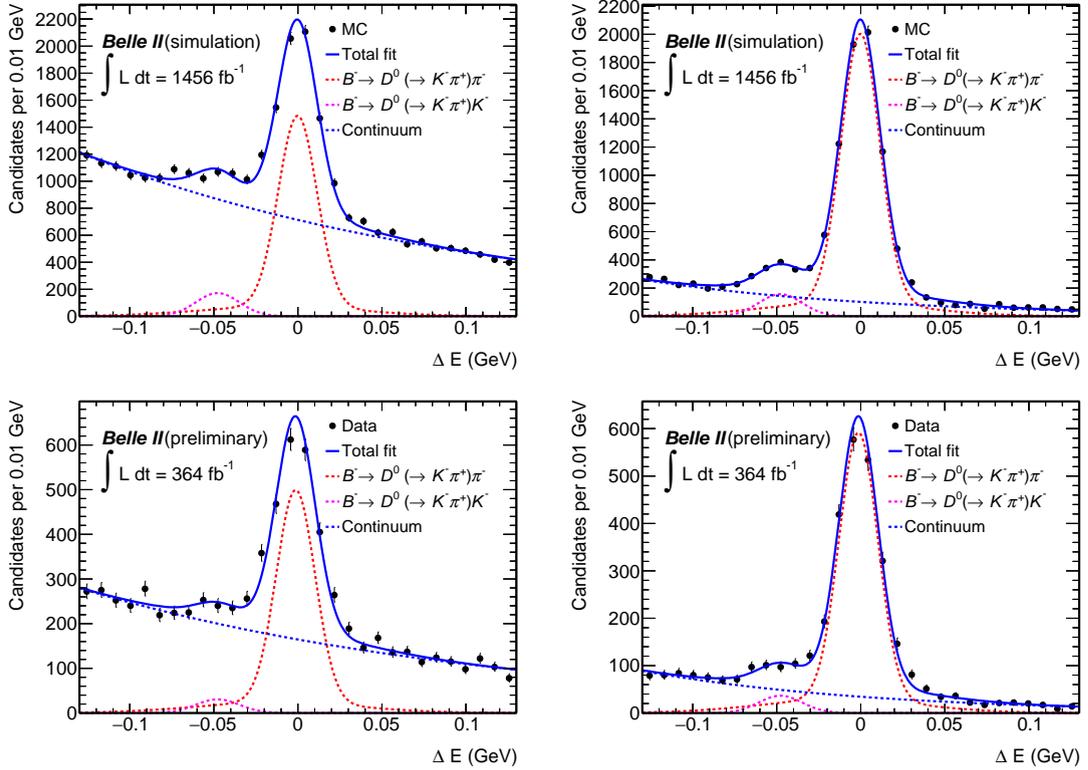


Figure 8.3: Distributions of ΔE for $B^+ \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\pi^+$ candidates reconstructed in Belle II (top) simulated and (bottom) collision data, (left) failing and (right) passing the continuum suppression requirement, with fit projections overlaid.

the CP -violating asymmetry, as detailed in the summary Table 8.2.

8.7 Assumption on w calibration

The uncertainty in the w calibration parameters is already incorporated in the fit results by means of Gaussian constraints. However, the w calibration relies on the empirical assumption of a linear relation between predicted and true wrong-tag probability with zero intercept. I estimate the associated systematic uncertainty by using an alternative, quadratic calibration function with fixed points at (0,0) and (0.5,0.5). The alternative calibration parameters measured in simulation are shown in Table 8.1. I generate simpli-

	$p1$	$p2$	Correlation	p-Value
B_{tag}^0	0.979 ± 0.043	0.004 ± 0.109	-0.97	0.82
\bar{B}_{tag}^0	0.999 ± 0.042	-0.018 ± 0.110	-0.97	0.71

Table 8.1: Alternative wrong-tag calibration factors from simulation.

fied simulated experiments according to that alternative model, and I take the difference between averages of results observed in the default model and in the alternative model. The difference 0.004 ± 0.008 is compatible with zero, so I take the uncertainty on the value as contribution to the systematic uncertainty.

8.8 Uncertainty in the $\Upsilon(4S)$ branching fractions

The ratio between the number of charged and neutral B mesons produced in $\Upsilon(4S)$ decays, denoted as f^{+-}/f^{00} , is expected to be close to unity due to charged and neutral B mesons having similar masses and (approximately valid) isospin symmetry. However, rather than relying on assumptions, we use a recent measurement by the Belle experiment based on $B^+ \rightarrow J/\psi(\rightarrow \ell^+\ell^-)K^+$ and $B^0 \rightarrow J/\psi(\rightarrow \ell^+\ell^-)K^0$ decays, with $\ell = e, \mu$ [95]. Belle determined $f^{+-}/f^{00} = 1.065 \pm 0.012 \pm 0.019 \pm 0.047$, where the first and second uncertainties are statistical and systematic, respectively, and the third is due to the assumption of isospin symmetry in $B \rightarrow J/\psi(\rightarrow \ell^+\ell^-)K$ decays. All uncertainties are added in quadrature leading to a systematic uncertainty for the branching fraction of 2.5%.

8.9 Systematic uncertainty summary

Table 8.2 reports a summary of all systematic uncertainties. Individual contributions are considered independent and summed in quadrature to determine the total systematic uncertainties.

Source	Uncertainty on \mathcal{B}	Uncertainty on A_{CP}
π^0 efficiency	8.00% (0.10)	N/A
Continuum-suppress efficiency	1.77% (0.02)	N/A
$N_{B\bar{B}}$	1.55% (0.02)	N/A
$1+f^{+-}/f^{00}$	2.50% (0.03)	N/A
Signal model	1.20% (0.02)	0.021
$B\bar{B}$ model	1.70% (0.02)	0.034
Continuum model	0.90% (0.01)	0.025
w calibration	-	0.010
Total systematic uncertainty	9.4% (0.12)	0.048
Total stat. uncertainty	16.0% (0.20)	0.30

Table 8.2: Relative (absolute) systematic uncertainties on the measurement of the $B^0 \rightarrow \pi^0\pi^0$ branching fraction and absolute systematic uncertainties on the measurement of the CP -violating asymmetry. Total systematic uncertainties, resulting from their sums in quadrature, are also reported, and compared with statistical uncertainties.

Chapter 9

Results

By combining fit results with systematic uncertainties, I report the final results for the measurement of the $B^0 \rightarrow \pi^0\pi^0$ branching fraction,

$$\mathcal{B}(B^0 \rightarrow \pi^0\pi^0) = (1.26_{-0.19}^{+0.20} \text{ (stat)} \pm 0.12 \text{ (syst)}) \times 10^{-6},$$

and CP -violating charge-asymmetry,

$$A_{CP}(B^0 \rightarrow \pi^0\pi^0) = 0.06 \pm 0.30 \text{ (stat)} \pm 0.05 \text{ (syst)}.$$

The results are compatible with the known values [11], and have precision superior or comparable with the precision of world-best results, despite our use of a smaller sample. For both results systematic uncertainties are smaller than the statistical uncertainties. Consistency with the previous Belle II measurement, based on a subset of our data, is checked using simplified simulated experiments to account for correlations.

The dominant contribution to the systematic uncertainty on the branching fraction is the 8% uncertainty on the π^0 -reconstruction efficiency. This is entirely driven by the uncertainty on the branching fraction of the control channel used to determine such efficiency in this analysis. As discussed in Appendix B, additional methods for determining the π^0 -reconstruction efficiency are being developed in Belle II and are expected to reduce this uncertainty by a factor of up to five in future. The second dominant systematic uncertainty, associated with our knowledge of the branching fractions of $\Upsilon(4S)$ decays, is also expected to reduce in future with dedicated measurements at Belle II.

Dominant systematic uncertainties on the measurement of the CP -violating asymmetry are those associated with the accuracy of fit modeling. As of this writing it would be speculative to assert obvious future opportunities of large reductions of those uncertainties. However, experience suggests that reductions are likely, as the collaboration matures and develops a more detailed understanding of its data and detector.

Table 9.1 shows comparison of these results with the legacy Belle results, based on a sample twice as large as ours [55]; with the earlier Belle II measurement, which uses half of our data [54]; and with the legacy BaBar results, based on a 20% larger sample [56].

	This work	Belle [55]	Belle II [54]	BaBar [56]
$\mathcal{B}(B^0 \rightarrow \pi^0\pi^0)$	$1.26_{-0.19}^{+0.20} \pm 0.12$	$1.31 \pm 0.19 \pm 0.19$	$1.38 \pm 0.27 \pm 0.22$	$1.83 \pm 0.21 \pm 0.13$
$A_{CP}(B^0 \rightarrow \pi^0\pi^0)$	$0.06 \pm 0.30 \pm 0.05$	$0.14 \pm 0.36 \pm 0.10$	$-0.14 \pm 0.46 \pm 0.07$	$0.43 \pm 0.26 \pm 0.05$

Table 9.1: Comparison of results with previous measurements. The first contribution to the uncertainties is statistical, the second is systematic.

The statistical uncertainty on the branching fraction is on par with the best one, which was reported by the Belle collaboration, using a sample twice as large as ours. It is also on par with the statistical precision of the earlier Belle II measurement scaled to our sample size, even though we increased our statistical uncertainty by incorporating several systematic effects through Gaussian constraints. The systematic precision on the branching fraction is world leading. It will further reduce once ongoing alternative methods to determine the π^0 efficiency will be finalized in Belle II.

The statistical precision on the CP -violating asymmetry is second only to the world-leading result by Babar, which is based on a 20% larger sample. It is superior to that of Belle, based on twice the data used in this work, and to that of the earlier Belle II measurement scaled to our sample size even though we increased our statistical uncertainty by incorporating in it several systematic effects through Gaussian constraints. The systematic precision on the asymmetry is on par to the world-leading one by Babar.

The quality of these results, which matches or surpasses that of the world-best results in spite of the disadvantage in sample size, is the result of a thoroughly optimized analysis of well understood data from an improved and well understood detector.

The fact that we further improve over the earlier Belle II measurement (at given sample size) is the product of two key innovations introduced in this analysis, (i) the adoption of a significantly more data-driven approach to fit modeling, entirely developed in this work and (ii) the introduction of a more performing flavor-identification algorithm, developed by others in Belle II. For instance, the earlier Belle II analysis modeled continuum by fixing the model parameters to the values observed in signal sidebands. This was suboptimal as it did not use the statistical information of the continuum present in the signal region thus subjecting the results to a 7% systematic uncertainty due to statistical fluctuations of the sideband population. Our approach extends the fit range to include the sidebands directly in the signal fit. This effectively adds a factor of up to 2.3 of continuum events that allow for a significantly more precise description and a corresponding model uncertainty reduced to 1%. Similar considerations apply to the $B\bar{B}$ model, which in the earlier Belle II analysis was fixed from simulation whereas here has its yield freely determined by the fit. These innovations improved chiefly the branching fraction result. The increased precision on the asymmetry is mostly ascribed to the use of an improved flavor tagging algorithm. This work is the second Belle II analysis that uses a newly introduced flavor-tagging algorithm and the first that incorporates its *unbinned* information in the analysis. The algorithm is based on a graph neural-network that identifies, and accounts for, the relationships between final-state particles, in addition to just identifying particles, as conventional taggers do. Thus the algorithm discerns better the origin of charged particles and correlates them more efficiently to useful flavor information. This results in an enhanced flavor tagging performance equivalent to a 15% net increase in data sample size for tagged decays with respect to the default Belle II tagger and previous B factories.

The achieved precision demonstrates that our work offers a compelling contribution to the global knowledge of $B^0 \rightarrow \pi^0\pi^0$ decays and its phenomenological implications.

Summary

This thesis reports on a measurement of the branching fraction and CP -violating asymmetry of $B^0 \rightarrow \pi^0\pi^0$ decays reconstructed in the full data set collected by the Belle II experiment as of this writing. The sample contains 387 million pairs of bottom-antibottom mesons from decays of $\Upsilon(4S)$ mesons produced in 10.58 GeV electron-positron collisions by the SuperKEKB accelerator. The chief challenge is the reconstruction, in a decay without charged particles, $B^0 \rightarrow \pi^0(\rightarrow \gamma\gamma)\pi^0(\rightarrow \gamma\gamma)$, of a rare signal immersed in 10^7 -times larger signal-like backgrounds of neutral pions from light-quark production. This poses two principal experimental demands.

Maximizing the signal-to-background ratio. I develop two multivariate statistical-learning classifiers, one targeted at separating genuine, collision-related photons from misreconstructed photons and photons not coming from the collision and another at suppressing the dominant background from lighter quark-pairs.

Determining accurately the sample composition. I devise a fit that properly accounts for multidimensional dependencies between observables, implementing multidimensional probability densities and studying carefully the distributions of estimators.

This analysis upgrades and improves an earlier Belle II analysis based on a smaller sample [54]. In addition to using a doubled sample size, I improve background suppression by increasing the discriminating power of dedicated and optimized classifiers; I render the sample-composition fit significantly less dependent on simulation and assumptions, thus achieving significantly reduced modeling uncertainties; and I simplify the fit by including as observable the B flavor as obtained by new algorithms with higher efficacy. The results are

$$\begin{aligned}\mathcal{B}(B^0 \rightarrow \pi^0\pi^0) &= (1.26_{-0.19}^{+0.20} \text{ (stat)} \pm 0.12 \text{ (syst)}) \times 10^{-6} \text{ and} \\ A_{CP}(B^0 \rightarrow \pi^0\pi^0) &= 0.06 \pm 0.30 \text{ (stat)} \pm 0.05 \text{ (syst)}.\end{aligned}$$

The results are compatible with the known values and have precision superior to, or comparable with, that of world-best results, despite our use of a smaller sample.

The statistical uncertainty on the branching fraction is on par with that of the world-best result by Babar, which is based on a 20% larger samples than ours. The systematic precision is world-leading. The statistical uncertainty on the CP -violating asymmetry is slightly inferior to the world-best one from Babar [56] while the systematic uncertainty is on par with its Babar counterpart. The quality of our results largely surpasses that of the corresponding Belle legacy results [55], in spite of the strong disadvantage in sample size, and that of the earlier Belle II measurement.

These results will contribute significantly to the knowledge of two-body charmless B decays and their implications on constraining SM extensions, improving phenomenological models of hadronic B decay amplitudes, and increasing the precision on the quark-mixing matrix parameter α .

Appendices

Appendix A

Photon energy bias correction

Miscalibrations in the photon energy-reconstruction in the calorimeter are known to affect Belle II reconstruction. By studying diphoton mass distributions in π^0 decays, these miscalibrations were estimated to be at the level of 1% by Belle II experts, and therefore neglected. However, in the early phase of this analysis, I observed for the first time in Belle II that photon-energy miscalibration caused shifts of the peaks in the ΔE and M_{bc} distributions of B decays into π^0 mesons, with much larger effects. For example, shifts of approximately -10 MeV appeared on ΔE peaks of 50 MeV resolution. I therefore performed dedicated studies using $B^+ \rightarrow K^+\pi^0$ and $B^+ \rightarrow \bar{D}^0(\rightarrow K^-\pi^+\pi^0)\pi^+$ decays in simulation and in collision data to assess the severity of the effect and understanding its major features.

Figure A.1 shows the ΔE distribution for reconstructed $B^+ \rightarrow \bar{D}^0(\rightarrow K^-\pi^+\pi^0)\pi^+$ candidates on simulation and on data. Evidence of a negative shift of data with respect to simulation is seen.

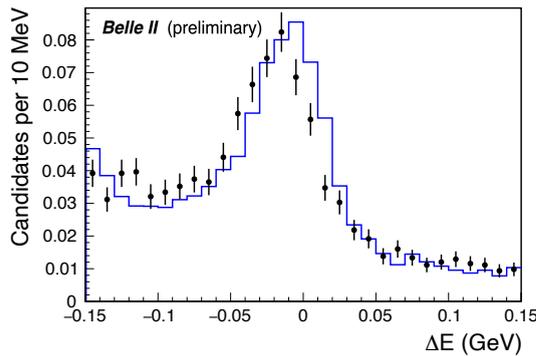


Figure A.1: Distribution of ΔE for $B^+ \rightarrow \bar{D}^0(\rightarrow K^-\pi^+\pi^0)\pi^+$ candidates reconstructed in (histogram) generic simulation and (circles) collision data, normalized to each other.

Assuming, as a working approximation, that the observed photon energy $E'_i = E_i - \delta$ receives an additive negative bias δ , the observed ΔE distribution is biased toward negative values by twice the bias,

$$\Delta E' = E_{B'}^* - \sqrt{s}/2 \approx (E_1 - \delta + E_2 - \delta + E_{\text{chrg}}) - \sqrt{s}/2 \approx -2\delta.$$

where E_{chrg} collectively refers to the energy of charged decay products of the B meson. The key finding is that such biases do not affect significantly the diphoton mass, that

was the observable typically used in neutral-energy calibration to monitor the quality of calibrations, as it gets suppressed by the inverse of the energy,

$$\begin{aligned}
 m'(\gamma_1\gamma_2) &= \sqrt{(E'_1 + E'_2)^2 - (\vec{p}'_1 + \vec{p}'_2)^2} = \sqrt{E_1'^2 + E_2'^2 + 2E_1'E_2' - (E_1'^2 + E_2'^2 + 2E_1'E_2' \cos \theta)} \\
 &= \sqrt{2E_1'E_2'(1 - \cos \theta)} = \sqrt{2[E_1E_2 - \delta(E_1 + E_2) + \mathcal{O}(\delta^2)](1 - \cos \theta)} \\
 &\approx \sqrt{2E_1E_2(1 - \cos \theta)} \left(1 - \frac{\delta}{E}\right) \\
 &= m(\gamma_1\gamma_2) \left(1 - \frac{\delta}{E}\right),
 \end{aligned}$$

where θ is the opening angle between photon directions in the lab frame. This explains why these biases went unnoticed until my study. Dedicated simulations showed that analytic calculations based on a negative photon-energy bias agreed well with the observations in various channels used in this work.

Following this observation, I implemented an empirical calibration to correct the photon energy based on high-level quantities like ΔE in addition to diphoton masses. This study prompted the preparation of Belle II calibrations applied to the photon energies, called 'photon-energy bias corrections', which after extended to include a variety of methods that covered satisfactory multiple effects. Figure A.2 shows a summary of the current status of photon-bias corrections.

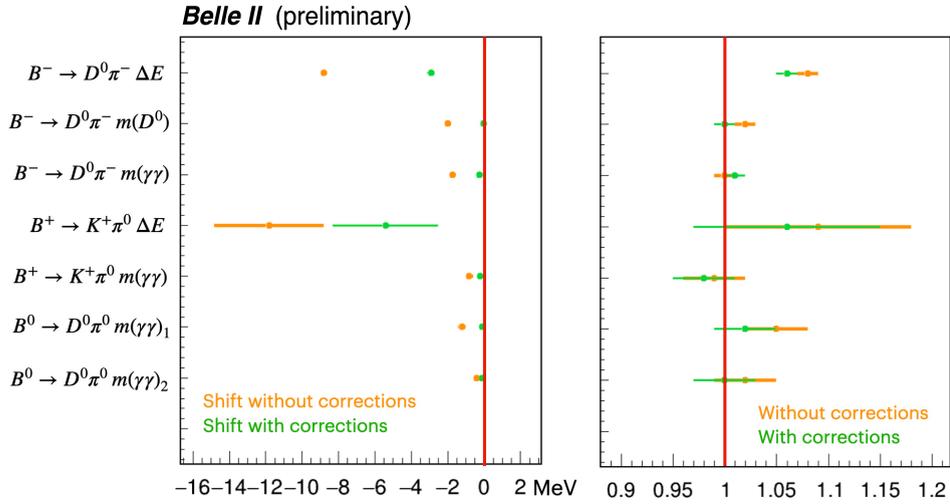


Figure A.2: Comparison between (left) additive shifts and (right) multiplicative factors applied to the signal peak-positions of ΔE , diphoton mass, and D^0 meson mass (orange) before and (green) after applying the photon corrections in the full Belle II data sample.

Appendix B

Methods for the determination of the π^0 reconstruction efficiency

The reconstruction of neutral particles is one of the unique strengths of the Belle II experiment. It is important to measure and understand potential differences between data and simulation in their reconstruction and selection efficiencies. Many processes are being used at Belle II to study the π^0 reconstruction efficiency. I am responsible for measuring relative π^0 efficiencies using D mesons decays. Others have done similar studies using τ lepton decays, η decays, and $e^+e^- \rightarrow \pi^+\pi^-\pi^0\gamma_{\text{ISR}}$ decays. I report here a brief description of each method and a concise discussion of their merits and shortcomings.

B.1 Relative efficiency from D meson decays

I measured the relative π^0 reconstruction efficiency between data and simulation using $D^{*-} \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^-$ decays, normalized to the $D^{*-} \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\pi^-$ yield, as described in Section 8.3,

$$\epsilon_{\pi^0} = \frac{N(K^+\pi^-\pi^0)}{N(K^+\pi^-)} \frac{\mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-)}{\mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)\mathcal{B}(\pi^0 \rightarrow \gamma\gamma)},$$

where $N(K^+\pi^-\pi^0)$ and $N(K^+\pi^-)$ are the observed signal yields of $D^{*-} \rightarrow \bar{D}^0(\rightarrow K^+\pi^-\pi^0)\pi^-$ and $D^{*-} \rightarrow \bar{D}^0(\rightarrow K^+\pi^-)\pi^-$ decays, respectively. Hence,

$$\frac{\epsilon_{\pi^0}}{\epsilon_{\pi^0}^{\text{MC}}} = \frac{N_D^0 \rightarrow K^-\pi^+\pi^0 / N^{\text{MC}}(D^0 \rightarrow K^-\pi^+\pi^0)}{N(D^0 \rightarrow K^-\pi^+) / N^{\text{MC}}(D^0 \rightarrow K^-\pi^+)}.$$

where the "MC" label dubs quantities determined in simulation. Large samples of pure D mesons decays are available, so the statistical uncertainty on this measurement is small. Nevertheless, the final accuracy of the method is limited by the 2.6% uncertainty on $\mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-) / \mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)$ [11], and by possible discrepancies between data and simulation in the Dalitz distribution of $\bar{D}^0 \rightarrow K^+\pi^-\pi^0$ decays, which are hard to decouple from π^0 -related effects. Unless more precise measurements of $\mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-) / \mathcal{B}(\bar{D}^0 \rightarrow K^+\pi^-\pi^0)$ will become available, this method will be limited by the irreducible systematic uncertainty on the branching-fraction ratio, and hence will be not suitable for future Belle II analyses.

B.2 Relative efficiency from τ lepton decays

In analogy with the approach above, the π^0 reconstruction-efficiency ratio between data and simulation using τ decays is measured as

$$\frac{\varepsilon_{\pi^0}}{\varepsilon_{\pi^0}^{\text{MC}}} = \frac{N(\tau^+ \rightarrow \pi^+ \pi^- \pi^+ \pi^0 \nu_\tau) / N^{\text{MC}}(\tau^+ \rightarrow \pi^+ \pi^- \pi^+ \pi^0 \nu_\tau)}{N(\tau^+ \rightarrow \pi^+ \pi^- \pi^+ X \nu_\tau) / N^{\text{MC}}(\tau^+ \rightarrow \pi^+ \pi^- \pi^+ X \nu_\tau)},$$

where the τ decay at the denominator is inclusive, and the three-pion mass distribution is weighted to match the distribution at the numerator. These samples have large size and high purities, and the intrinsic uncertainty on the known branching-fraction ratio is only 1.2%. However, the description of the kinematic properties of the multibody τ decay is uncertain and the study is still preliminary, so the resulting efficiencies are not used for this analysis.

B.3 Relative efficiency from η meson decays

The π^0 reconstruction-efficiency ratio between data and simulation is measured using $\eta \rightarrow \pi^0 \pi^0 \pi^0$ decays, normalized to the $\eta \rightarrow \gamma\gamma$ or $\eta \rightarrow \pi^+ \pi^- \pi^0$ channels, as

$$\frac{\varepsilon_{\pi^0}}{\varepsilon_{\pi^0}^{\text{MC}}} = \frac{N(\eta \rightarrow \pi^0 \pi^0 \pi^0) / N^{\text{MC}}(\eta \rightarrow \pi^0 \pi^0 \pi^0)}{N(\eta \rightarrow \gamma\gamma, \eta \rightarrow \pi^+ \pi^- \pi^0) / N^{\text{MC}}(\eta \rightarrow \gamma\gamma, \eta \rightarrow \pi^+ \pi^- \pi^0)}.$$

The uncertainty on the known branching-fraction ratio is about 0.7% for the $\eta \rightarrow \gamma\gamma$ case, and 1.6% for the $\eta \rightarrow \pi^+ \pi^- \pi^0$ case. However, signal-to-background ratio is poor due to the large combinatorial background making the uncertainty dominated by modeling uncertainties. Currently this is the least precise method among those listed here, and the study is still preliminary, so the resulting efficiencies are not used for this analysis.

B.4 Absolute efficiency from the $e^+ e^- \rightarrow \pi^+ \pi^- \pi^0 \gamma_{\text{ISR}}$ process

In all the above methods, the π^0 -reconstruction efficiency in data is determined relative to simulation. Unlike in those, the absolute π^0 efficiency is accessible by considering $e^+ e^- \rightarrow \omega(\rightarrow \pi^+ \pi^- \pi^0) \gamma_{\text{ISR}}$ decays, that are $e^+ e^-$ annihilation events in three pions, with an additional photon from initial state radiation. We use partially and fully reconstructed events,

$$\varepsilon_{\pi^0} = \frac{N_{\pi^0}(\omega(\rightarrow \pi^+ \pi^- \pi^0) \gamma_{\text{ISR}})}{N_{\omega}(\omega(\rightarrow \pi^+ \pi^- \pi^0) \gamma_{\text{ISR}})},$$

where N_{π^0} is the yield of selected events after explicit π^0 reconstruction and N_{ω} is the event yield determined without using any π^0 information, but using only properties of the beam, charged pions, and ISR photon. This method uses large and pure samples and provides an absolute π^0 efficiency, which makes it convenient.

However, the study is still preliminary and several aspects of the involved backgrounds are not fully understood, so the correction factors are not used for this analysis.

B.5 Possible dependence on multiplicity

The π^0 reconstruction efficiency methods described here are based on samples with largely different particle multiplicities per event. Decays of D mesons have high multiplicities,

while τ and $e^+e^- \rightarrow \omega\gamma_{\text{ISR}}$ events have lower multiplicities. Only if the π^0 efficiency does not depend on the event multiplicity, a method can be applied universally to all analyses. Inefficiencies can be caused by various factors: wrong estimation of the shower energy or direction, combination of a real photon from a π^0 with a misreconstructed photon, and other causes. Some of these depend on the event multiplicity, such as the overlap between a track and a photon, and the overlap between two photons (merged photon). Studies on this matter are still in progress.

Bibliography

- [1] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.
- [2] S. L. Glashow, *Partial-Symmetries of Weak Interactions*, Nuclear Physics **22** (1961) 579.
- [3] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519** (1968) 367.
- [4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.
- [5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.
- [6] G. Guralnik, C. Hagen, and T. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.
- [7] To a good approximation, the attractive force between quarks grows linearly with distance. As two quarks are separated, putting energy into the bounded system, it becomes energetically favorable for a new quark–antiquark pair to appear, rather than extending the distance further. As a consequence new bound states are generated instead of separated and free quarks.
- [8] G. Lüders, *Proof of the TCP Theorem*, Annals of Physics **2** (1957) 1.
- [9] W. Pauli, *Exclusion Principle, Lorentz Group, and Reversal of Space-time and Charge*, Niels Bohr and the Development of Physics (1955) 1.
- [10] C. B. Adams *et al.*, *Axion Dark Matter*, 2023. [arXiv:2203.14923](https://arxiv.org/abs/2203.14923) [hep-ex].
- [11] R. L. Workman and Others, Particle Data Group, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.
- [12] M. Gell-Mann and M. Lévy, *The Axial Vector Current in Beta Decay*, Il Nuovo Cimento **16** (1960) 705.
- [13] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.
- [14] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (1970) 1285.
- [15] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

- [16] A generic $N \times N$ complex matrix U is defined with $2N^2$ free parameters. The unitary condition $UU^\dagger = \mathbb{1}$ eliminates N^2 degrees of freedom. We can then redefine $(N - 1)$ quark fields to absorb $(N - 1)^2$ free parameters. We end up with $2N^2 - N^2 - (2N - 1) = (N - 1)^2$ free parameters.
- [17] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.
- [18] 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) 1039.
- [19] C. Jarlskog, *A Basis Independent Formulation of the Connection Between Quark Mass Matrices, CP Violation and Experiment*, Z. Phys. C **29** (1985) 491.
- [20] I. Dunietz, O. W. Greenberg, and D. Wu, *A Priori Definition of Maximal CP Nonconservation*, Phys. Rev. Lett. **55** (1985) 2935.
- [21] W. Hulsbergen, *Constraining New Physics in Meson Mixing*, Modern Physics Letters A **28** (2013) .
- [22] R. Lewis, *On the phenomenological description of CP violation for K-mesons and its consequences*, Phys. Rev. Letters **14** (1965) 749.
- [23] Y. Amhis *et al.*, HFLAV Group, *Averages of b-hadron, c-hadron, and τ -lepton properties as of 2021*, Phys. Rev. D **107** (2023) 052008, arXiv:2206.07501 [hep-ex].
- [24] J. Charles *et al.*, CKMfitter Group, *Unitarity Triangle Plot*, Eur. Phys. J. C **41** (2005, updated plots available at <http://ckmfitter.in2p3.fr>) .
- [25] L.-L. Chau, H.-Y. Cheng, W. K. Sze, H. Yao, and B. Tseng, *Charmless Nonleptonic Rare Decays of B Mesons*, Phys. Rev. D **43** (1991) 2176. [Erratum: Phys. Rev. D **58**, 019902 (1998)].
- [26] M. Beneke, G. Buchalla, M. Neubert, and C. T. Sachrajda, *QCD Factorization for $B \rightarrow \pi\pi$ Decays: Strong Phases and CP Violation in the Heavy Quark Limit*, Phys. Rev. Lett. **83** (1999) 1914.
- [27] C.-D. Lu, K. Ukai, and M.-Z. Yang, *Branching Ratio and CP Violation of $B \rightarrow \pi\pi$ Decays in Perturbative QCD Approach*, Phys. Rev. D **63** (2001) 074009.
- [28] Y.-L. Zhang, X.-Y. Liu, Y.-Y. Fan, S. Cheng, and Z.-J. Xiao, *$B \rightarrow \pi\pi$ Decays and Effects of the Next-to-Leading Order Contributions*, Phys. Rev. D **90** (2014) 014029.
- [29] M. Beneke and M. Neubert, *QCD factorization for $B \rightarrow PP$ and $B \rightarrow PV$ decays*, Nucl. Phys. B **675** (2003) 333.
- [30] M. Gronau, O. F. Hernandez, D. London, and J. L. Rosner, *Broken $SU(3)$ Symmetry in Two-Body B Decays*, Phys. Rev. D **52** (1995) 6356.
- [31] C.-F. Qiao, R.-L. Zhu, X.-G. Wu, and S. J. Brodsky, *A Possible Solution to the $B \rightarrow \pi\pi$ Puzzle Using the Principle of Maximum Conformality*, Phys. Lett. B **748** (2015) 422.
- [32] H.-Y. Cheng, C.-W. Chiang, and A.-L. Kuo Phys. Rev. D **91** (2015) 014011.

- [33] H.-Y. Cheng and S. Oh, *Flavor $SU(3)$ Symmetry and QCD Factorization in $B \rightarrow PP$ and PV Decays*, Journal of High Energy Physics **09** (2011) 024.
- [34] S.-H. Zhou, Q.-A. Zhang, W.-R. Lyu, and C.-D. Lü, *Analysis of Charmless Two-body B decays in Factorization Assisted Topological Amplitude Approach*, Eur. Phys. J. C **77** (2017) 125.
- [35] N. G. Deshpande and X.-G. He, *Isospin Structure of Penguins and Their Consequences in B Physics*, Phys. Rev. Lett. **74** (1995) 26. [Erratum: Phys. Rev. Lett.74, 4099 (1995)].
- [36] M. Gronau, O. F. Hernandez, D. London, and J. L. Rosner, *Electroweak Penguins and Two-Body B Decays*, Phys. Rev. D **52** (1995) 6374.
- [37] J. Charles, O. Deschamps, S. Descotes-Genon, and V. Niess, *Isospin Analysis of Charmless B -Meson Decays*, Eur. Phys. J. C **77** (2017) 574.
- [38] M. Gronau and D. London, *Isospin Analysis of CP Asymmetries in B Decays*, Phys. Rev. Lett. **65** (1990) 3381.
- [39] H. J. Lipkin, Y. Nir, H. R. Quinn, and A. Snyder, *Penguin Trapping with Isospin Analysis and CP Asymmetries in B Decays*, Phys. Rev. D **44** (1991) 1454.
- [40] M. Gronau, O. F. Hernandez, D. London, and J. L. Rosner, *Decays of B Mesons to Two Light Pseudoscalars*, Phys. Rev. D **50** (1994) 4529.
- [41] M. Gronau, O. F. Hernandez, D. London, and J. L. Rosner, *Broken $SU(3)$ Symmetry in Two-Body B Decays*, Phys. Rev. D **52** (1995) 6356.
- [42] M. Beneke, M. Gronau, J. Rohrer, and M. Spranger, *A Precise Determination of α Using $B^0 \rightarrow \rho^+ \rho^-$ and $B^+ \rightarrow K^* \rho^+$* , Phys. Lett. B **638** (2006) 68.
- [43] A. E. Snyder and H. R. Quinn, *Measuring CP Asymmetry in $B \rightarrow \rho\pi$ Decays Without Ambiguities*, Phys. Rev. D **48** (1993) 2139.
- [44] M. Gronau and J. Zupan, *On Measuring α in $B(t) \rightarrow \rho^\pm \pi^\mp$* , Phys. Rev. D **70** (2004) 074031.
- [45] M. Gronau and J. Zupan, *Weak Phase α from $B \rightarrow a_1(1260)\pi$* , Phys. Rev. D **73** (2006) 057502.
- [46] S. Laplace and V. Shelkov, *CP Violation and the Absence of Second Class Currents in Charmless B Decays*, Eur. Phys. J. C **22** (2001) 431.
- [47] A. F. Falk, Z. Ligeti, Y. Nir, and H. Quinn, *Comment on Extracting α from $B \rightarrow \rho\rho$* , Phys. Rev. D **69** (2004) 011502.
- [48] 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) 345.
- [49] R. Shankar, *Principles of Quantum Mechanics*. Springer, 2nd ed., 1994.
- [50] M. Pivk and F. R. Le Diberder, *Isospin Constraints from/on $B \rightarrow \pi\pi$* , Eur. Phys. J. C **39** (2005) 397.

- [51] B. Aubert *et al.*, BABAR Collaboration, *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) 071104.
- [52] S. H. Lee *et al.*, Belle Collaboration, *Evidence for $B^0 \rightarrow \pi^0 \pi^0$* , Physical Review Letters **91** (2003) 261801.
- [53] B. Aubert *et al.*, BABAR Collaboration, *Observation of a Significant Excess of $\pi^0 \pi^0$ Events in B Meson Decays*, Phys. Rev. Lett. **91** (2003) 241801.
- [54] F. Abudinén *et al.*, Belle II Collaboration, *Measurement of the Branching Fraction and CP Asymmetry of $B^0 \rightarrow \pi^0 \pi^0$ Decays Using $198 \times 10^6 B\bar{B}$ pairs in Belle II Data*, Phys. Rev. D **107** (2023) 112009.
- [55] T. Julius *et al.*, Belle Collaboration, *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** (2017) 032007.
- [56] J. P. Lees *et al.*, BABAR Collaboration, *Measurement of CP Asymmetries and Branching Fractions in Charmless Two-Body B-Meson Decays to Pions and Kaons*, Phys. Rev. D **87** (2013) 052009.
- [57] I. Adachi *et al.*, Belle II Collaboration, *Measurement of Branching Fractions and Airect CP Asymmetries for $B \rightarrow K\pi$ and $B \rightarrow \pi\pi$ Decays at Belle II*, Phys. Rev. D **109** (2024) 012001.
- [58] Y. Ohnishi *et al.*, *Accelerator Design at SuperKEKB*, Prog. Theo. Exp. Phys. (2013) 03A011.
- [59] P. Oddone, *Detector Considerations*, in D. H. Stork, proceedings of Workshop on Conceptual Design of a Test Linear Collider: Possibilities for a BB Factory (1987) 423.
- [60] P. Raimondi *et al.*, *Beam-Beam Issues for Colliding Schemes with Large Piwinski Angle and Crabbed Waist*, arXiv:physics/0702033.
- [61] D. R. Douglas, J. Kewisch, and R. C. York, *Betatron Function Parameterization of Beam Optics Including Acceleration*, in *Proceedings of the 1988 Linear Accelerator Conference*. (1988).
- [62] S. Verdu-Andres, *Luminosity Geometric Reduction Factor from Colliding Bunches with Different Lengths*, BNL Report BNL-114409-2017-IR (2017).
- [63] T. Abe *et al.*, Belle II Collaboration, *Belle II Technical Design Report*, arXiv:1011.0352.
- [64] C. Marinas, DEPFET, *The Belle II Pixel Detector: High precision with Low Material*, Nucl. Instrum. Meth. A **731** (2013) 31.
- [65] K. Adamczyk *et al.*, Belle II Collaboration, *The Belle II Silicon Vertex Detector Assembly and Mechanics*, Nucl. Instrum. Meth. A **845** (2017) 38.
- [66] H. Hirano *et al.*, *A High-Resolution Cylindrical Drift Chamber for the KEK B-Factory*, Nucl. Instrum. Meth. A **455** (2000) 294.

- [67] V. Aulchenko *et al.*, *Electromagnetic Calorimeter for Belle II*, J. Phys. Conf. Ser. **587** (2015) 012045.
- [68] K. Inami *et al.*, *TOP Counter for Particle Identification at the Belle II Experiment*, Nucl. Instrum. Meth. A **766** (2014) 5.
- [69] S. Iwata *et al.*, *Particle Identification Performance of the Prototype Aerogel RICH Counter for the Belle II Experiment*, Prog. Theo. Exp. Phys. **2016** (2016) 033.
- [70] A. Abashian *et al.*, Belle Collaboration, *The K_L/μ Detector Subsystem for the BELLE Experiment at the KEK B factory*, Nucl. Instrum. Meth. A **449** (2000) 112.
- [71] V. Bertacchi *et al.*, *Track Finding at Belle II*, Computer Physics Communications **259** (2021) 107610.
- [72] T. Alexopoulos, M. Bachtis, E. Gazis, and G. Tsipolitis, *Implementation of the Legendre Transform for Track Segment Reconstruction in Drift Tube Chambers*, Nucl. Instrum. Meth. A **592** (2008) 456.
- [73] R. Mankel, *Pattern Recognition and Event Reconstruction in Particle Physics Experiments*, Reports on Progress in Physics **67** (2004) 553.
- [74] H. Bethe, *Theory of the Passage of Fast Corpuscular Rays Through Matter*, Annalen Phys. **5** (1930) 325.
- [75] F. Bloch, *Stopping Power of Atoms with Several Electrons*, Z. Phys. **81** (1933) 363.
- [76] M. Milesi, *Lepton Identification in Belle II Using Observables from the Electromagnetic Calorimeter and Precision Trackers*, Talk at CHEP 2019, https://indico.cern.ch/event/773049/contributions/3476156/attachments/1932835/3217425/mmilesi_mva_CHEP_07_11_19.pdf.
- [77] M. Starič *et al.*, *Likelihood Analysis of Patterns in a Time-of-Propagation (TOP) Counter*, Nucl. Instrum. Meth. A **595** (2008) 252.
- [78] M. Starič, *Pattern recognition for the time-of-propagation counter*, Nucl. Instrum. Meth. A **639** (2011) 252.
- [79] J. Neyman and E. S. Pearson, *IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Phil. Trans. R. Soc. Lond. A. **231** (1933) 289.
- [80] F. Abudinén, *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*. PhD thesis, Munich, Max Planck Inst., 2018.
- [81] S. Brandt, C. Peyrou, R. Sosnowski, and A. Wroblewski, *The Principal Axis of Jets. An Attempt to Analyze High-Energy Collisions as Two-Body Processes*, Phys. Lett. **12** (1964) 57.
- [82] D. M. Asner *et al.*, CLEO collaboration, *Search for Exclusive Charmless Hadronic B Decays*, Phys. Rev. D **53** (1996) 1039.
- [83] 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) 1581.

- [84] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 Physics and Manual*, Journal of High Energy Physics **2006** (2006) 026.
- [85] D. J. Lange, *The EvtGen Particle Decay Simulation Package*, Nucl. Instrum. Meth. A **462** (2001) 152.
- [86] P. Golonka and Z. Was, *PHOTOS Monte Carlo: a Precision Tool for QED Corrections in Z and W Decays*, Eur. Phys. J. C **45** (2006) 97.
- [87] S. Agostinelli *et al.*, GEANT4 Collaboration, *GEANT4 – a Simulation Toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.
- [88] B. Ward, S. Jadach, and Z. Was, *Precision Calculation for $e^+e^- \rightarrow 2f$: the KK MC Project*, Nuclear Physics B - Proceedings Supplements **116** (2003) 73.
- [89] M. Chrzaszcz, T. Przedzinski, Z. Was, and J. Zaremba, *TAUOLA of τ lepton decays—framework for hadronic currents, matrix elements and anomalous decays*, Computer Physics Communications **232** (2018) 220.
- [90] T. Keck, *FastBDT: A Speed-Optimized and Cache-Friendly Implementation of Stochastic Gradient-Boosted Decision Trees for Multivariate Classification*, 2016. arXiv:1609.06119.
- [91] J. H. Friedman, *Greedy Function Approximation: A Gradient Boosting Machine*, Ann. Statist. **29** (2001) 1189.
- [92] J. H. Friedman, *Stochastic Gradient Boosting*, Computational Statistics & Data Analysis **38** (2002) 367.
- [93] V. F. Zernike, *Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode*, Physica **1** (1934) 689.
- [94] F. James and M. Roos, *MINUIT: a System for Function Minimization and Analysis of the Parameter Errors and Corrections*, Comput. Phys. Commun. **10** (1975) 343.
- [95] S. Choudhury *et al.*, Belle Collaboration, *Measurement of the B^+/B^0 Production Ratio in e^+e^- Collisions at the $\Upsilon(4S)$ Resonance using $B \rightarrow J/\psi(\ell\ell)K$ Decays at Belle*, Phys. Rev. D **107** (2023) L031102.
- [96] H. Albrecht *et al.*, ARGUS Collaboration, *Search for Hadronic $b \rightarrow u$ Decays*, Physics Letters B **241** (1990) 278.