Università degli Studi di Perugia

DIPARTIMENTO DI FISICA
Corso di Dottorato in Fisica

XXXVI CICLO

Search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay with an hadronic tagged approach at the Belle II experiment.

Candidato:
Stefano Moneta

Relatore:
Prof. Claudia Cecchi

Correlatore:
Dott. Elisa Manoni

Anno Accademico 2023-2024

Abstract

This document describes the search for the rare flavor changing neutral current decay $B^0 \to K^{*0}\tau^+\tau^-$ at the Belle II experiment. This decay is forbidden at tree level in the Standard Model, occurring only at loop level, and it is expected to be sensible to possible new physics extensions. The analysis exploits the 362.2 fb^{-1} of data collected in the 2019-2022 period at the center of mass energy of the $\Upsilon(4S)$ by the SuperKEKB electron-positron collider. One of the B mesons, produced in the $\Upsilon(4S) \to B\overline{B}$ decay, is fully reconstructed in hadronic decay modes, while the partner B meson is required to decay in a $K^{*0} \to K^+\pi^-$ and in two τ leptons of opposite signs, reconstructed in their 1-prong modes. A multivariate approach with a binary classifier is adopted to separate signal from background events. The classifier is trained on simulated events with data-driven corrections applied, and used as main tool for the separation between signal and mis-reconstructed events. The signal branching ratio is directly extracted through a binned maximum likelihood fit to the output of the classifier.

Based on the reconstructed $\tau^+\tau^-$ decay modes ($\tau \to \ell\nu\overline{\nu}$, $\tau \to \pi\nu$, $\tau \to \pi\pi^0\nu$, being ℓ an electron or a muon), signal events are categorized into different groups. The background suppression stage is optimized separately for each group, while they are combined to extract the signal branching fraction. Including all the systematic uncertainties, the expected upper limit on the branching ratio is determined to be $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2.7 \times 10^{-3}$ at 90% confidence level. This expectation is already competitive with the current world-best upper limit $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 3.1 \times 10^{-3}$, at 90% confidence level, set by the Belle collaboration with about a factor $\times 2$ larger dataset of 711 fb^{-1}.

Contents

1 Introduction

1.1 The Standard Model

The Standard Model (SM) of particle physics describes the known fundamental constituents of the Universe and their interactions. The SM framework (Figure 1.1) includes 12 spin-1/2 matter constituents (6 leptons and 6 quarks), four vector gauge bosons and one scalar Higgs boson [1].

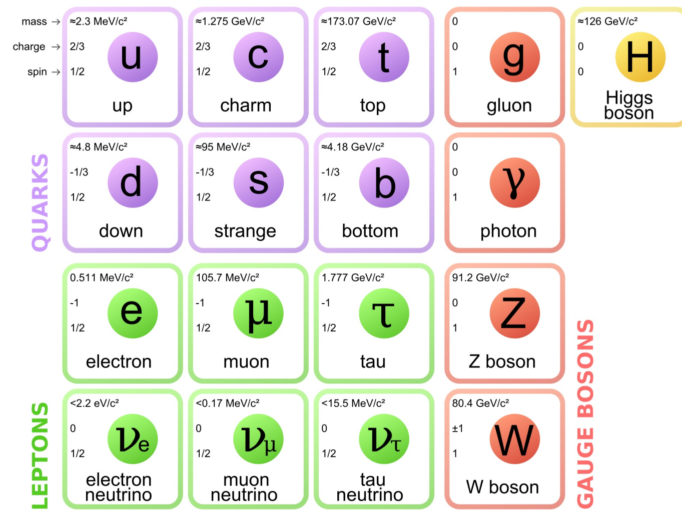


Figure 1.1: The elementary particles of the Standard Model (SM). The model contains 12 spin-1/2 fermions: 6 quarks and 6 leptons, divided into three flavor families (columns). Four gauge bosons account for strong (gluon), electromagnetic (photon) and weak (Z^0, $W^{\pm}$) interactions. Finally, the Higgs scalar boson generates the masses of leptons and gauge bosons through the spontaneous symmetry breaking mechanism.

In quantum field theory, the SM is described by a Lagrangian invariant under the action of the local gauge symmetry group $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$. $\mathrm{SU}(3)_C$ is the symmetry group of quantum chromodynamics (QCD) while $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ determines the electroweak theory (EW). Hence the Standard Model Lagrangian can be factorized as [2]:

$$\mathscr{L}_{SM} = \mathscr{L}_{QCD} + \mathscr{L}_{EW}$$

The QCD Lagrangian $\mathscr{L}_{QCD}$ describes strong interactions. In the fundamental representation of $\mathrm{SU}(3)_C$, quark fields transform as triplets and thus they assume 3 possible strong charge values, called color charge, usually denoted as R, G and B (Red, Green and Blue). There are 8 independent group generators to which correspond 8 spin-1 massless gauge fields called gluons.

The electroweak Lagrangian $\mathscr{L}_{EW}$ is built requiring local gauge symmetry $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$, i.e. invariance under weak isospin and weak hypercharge transformations. In the fundamental representation, the $SU(2)_L$ group generator is a weak isospin triplet $\vec{\tau} = (\tau^1, \tau^2, \tau^3)$, which couples to the 3 massless vector gauge bosons $\vec{W}_\mu = (W^1_\mu, W^2_\mu, W^3_\mu)$ with a coupling constant g. All fermions are represented by left-handed weak isospin doublets. Explicitly, leptons and quarks can be grouped as:

$$\begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L \quad \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L \quad \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L \quad \begin{pmatrix}u\\ d'\end{pmatrix}_L \quad \begin{pmatrix}c\\ s'\end{pmatrix}_L \quad \begin{pmatrix}t\\ b'\end{pmatrix}_L$$

The weak isospin eigenvalues are $+1/2$ for the upper component in the doublet, and $-1/2$ for the lower one. The weak-isospin eigenstates for "down" quarks, d', s' and b' are combination of the mass eigenstates d, s, b via the Cabibbo-Kobayashi-Maskawa (CKM) unitary matrix:

$$\begin{pmatrix}d'\\ s'\\ b'\end{pmatrix} = \begin{pmatrix}V_{ud} & V_{us} & V_{ub}\\ V_{cd} & V_{cs} & V_{cb}\\ V_{td} & V_{ts} & V_{tb}\end{pmatrix} \begin{pmatrix}d\\ s\\ b\end{pmatrix} \tag{1.1}$$

Because of the unitarity condition, the CKM matrix contains 4 independent parameters: three angles and one phase which is responsible for CP violation in the quark sector.

Similarly, the neutrino $SU(2)_L$ eigenstates are related to the mass eigenstates through the unitary Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix:

$$\begin{pmatrix}\nu_e\\ \nu_\mu\\ \nu_\tau\end{pmatrix} = \begin{pmatrix}U_{1e} & U_{2e} & U_{3e}\\ U_{1\mu} & U_{2\mu} & U_{3\mu}\\ U_{1\tau} & U_{2\tau} & U_{3\tau}\end{pmatrix} \begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}$$

which also contains other 4 independent parameters (3 mixing angles and a phase).

Finally, in the fundamental representation of $U(1)_Y$, there is only one massless vector gauge boson B_μ, coupling to the generator singlet Y with a constant g'. The elements are singlet of $U(1)_Y$ and the weak hypercharge operator Y acts like:

$$Y = 2(Q - T^3)$$

where Q is the electromagnetic charge and T^3 is the eigenvalue of τ^3.

The gauge symmetry in the Lagrangian is obtained substituting the derivatives ∂^μ with the covariant ones. For the electroweak $SU(2)_L \otimes U(1)_Y$ gauge group:

$$\partial^\mu \longrightarrow D^\mu = \partial^\mu + ig\vec{\tau} \cdot \vec{W}^\mu + ig'\frac{Y}{2}B^\mu$$

1.1.1 Higgs mechanism

Masses of fermions and gauge bosons are generated by the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge invariant coupling with the Higgs field. The SM predicts one Higgs boson consisting of two complex scalar fields ϕ^+, ϕ^0 placed in a weak isospin doublet with weak hypercharge $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

The Higgs Lagrangian, written such as to be $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ local gauge invariant, is:

$$\mathscr{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi) \quad ; \qquad V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2$$

$V(\phi)$ is the Higgs potential, with a mass term μ and a four particle vertex $(\phi^\dagger\phi)^2$ scaled by the parameter $\lambda > 0$, defining a self-interacting field. In order to apply perturbation theory, the potential should be expanded around its minimum. However, choosing $\mu^2 < 0$, $V(\phi)$ has not a minimum for $\phi = (0, 0)$, but a set of infinite degenerate minima for:

$$\phi^\dagger\phi = -\frac{\mu^2}{2\lambda}$$

as depicted in Figure 1.2. The arbitrary choice of a minimum leads to a spontaneous symmetry breaking. The chosen minimum is:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}$$

such that $\langle 0|\phi^+|0\rangle = 0$ and $\langle 0|\phi^0|0\rangle = v/\sqrt{2}$ where $v \equiv \sqrt{-\mu^2/\lambda}$ is the vacuum expectation value.

The Higgs field can be expanded around this minimum and redefined in unitary gauge as:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{1.2}$$

Substituting inside $\mathscr{L}_{Higgs}$, the kinetic part can be rearranged to have explicit mass terms:

$$\left(\frac{1}{2}vg\right)^2 W_\mu^+ W^{-\mu} + \frac{1}{8}v^2 \begin{pmatrix} A^\mu & Z^\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & g^2 + g'^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix}$$

Here the physical weak and electromagnetic fields $(W^\pm)^\mu$, Z^μ, A^μ appeared with their respective mass terms. Regarding the charged weak boson:

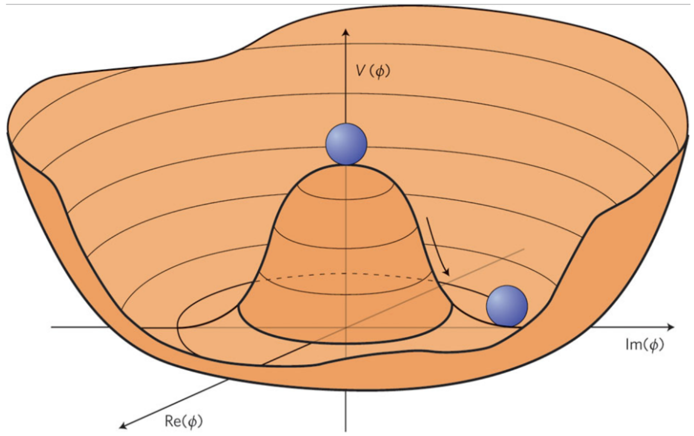


Figure 1.2: Profile of the Higgs potential $V(\phi)$ and spontaneous symmetry breaking. The set of degenerate minima satisfies the condition $\phi^\dagger\phi = -\mu^2/2\lambda$.

$$W^\pm = \frac{1}{\sqrt{2}}\left(W^1 \mp iW^2\right) \qquad M_W = \frac{1}{2}vg$$

While the physical electromagnetic field A^μ and neutral weak field Z^μ are linear combinations of B^μ and W_3^μ through the mixing matrix:

$$\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^\mu \\ W_3^\mu \end{pmatrix}$$

their masses are $M_A = 0$ and $M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}$. θ_W is known as Weinberg angle and $\tan\theta_W = g'/g$. The Weinberg angle can also be expressed as a function of weak boson masses:

$$\cos\theta_W = \frac{M_W}{M_Z}$$

The remaining terms of $(D_\mu\phi)^\dagger(D^\mu\phi)$ determine the coupling between Higgs field and weak bosons. From the potential $V(\phi)$ the Higgs mass $M_H = \sqrt{2\lambda}v$ is obtained as well as the self coupling terms.

Fermion masses are also generated from the interaction with the Higgs field. In particular, for charged leptons and down type quarks, an interaction Lagrangian invariant under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ can be written as:

$$\mathscr{L} = -g_f\,\overline{L}\phi R \tag{1.3}$$

where g_f is the Yukawa coupling constant for the fermion f, whose value is not a priori predicted by the SM. L and R are left-handed and right-handed chiral spinors, i.e. a doublet and a singlet under the $\mathrm{SU}(2)_L$ symmetry. After the spontaneous symmetry breaking, the Higgs isospin doublet ϕ can be expressed as (1.2).

Inserting all terms in the Lagrangian of Eq. (1.3), gives:

$$\mathscr{L} = -\frac{g_f}{\sqrt{2}} v \bar{f} f - \frac{g_f}{\sqrt{2}} \bar{f} f h \tag{1.4}$$

The first term is the fermion mass term with $m_f = g_f v/\sqrt{2}$, while the second one accounts for the interaction between the fermionic current and the Higgs boson and it is proportional to m_f/v. An analogous Lagrangian is obtained for the up-type quarks.

The first observation of the SM Higgs boson candidate was published in 2012 by the ATLAS and CMS collaborations [3, 4]. Exploiting *pp* collisions at the Large Hadron Collider (LHC) with 7 and 8 TeV center of mass energy, the most pure decay channels $H \to \gamma\gamma$ and $H \to Z^+ Z^{-*} \to \ell^+ \ell^+ \ell^+ \ell^-$ led to the discovery and allowed to identify the new particle as a scalar neutral boson.

1.1.2 Possible extensions

The minimal SM described above contains 25 independent free parameters:

- 12 Yukawa terms determining the fermion masses.
- 3 coupling constants for the gauge interactions: α_s, g and g'.
- 2 parameters μ and λ describing the Higgs potential.
- 4 parameters for the mixing of down type quarks (CKM matrix).
- 4 parameters accounting for neutrino oscillations (PMNS matrix).

Some subtended patterns, such as the fact that fermions are grouped into three mass-hierarchy families, or that the coupling constants α_s, g and g' are approximately of the same order of magnitude, seem to suggest an underlining symmetry beyond the SM. The introduction of a new and larger (broken) symmetry could reduce the high number of free parameters and possibly explain some of the open issues left in the SM. One of the most intriguing is the experimental evidence, at the cosmological level, of the presence of a dark matter, incompatible with any of the SM particles [5].

A possible extension of SM is the Supersymmetry (SUSY) theory, which, for each SM particle, predicts the existence of a supersymmetric partner with different spin. SUSY partners of leptons and quarks are scalar sleptons ans squarks, the correspondent of gauge bosons are spin-1/2 gauginos while the vector Higgsino is the partner of the Higgs. There are several SUSY models where the lightest supersymmetric particle is a suitable dark matter candidate. To date, SUSY particles have been directly searched at the Large Hadron Collider (LHC) excluding masses below $\sim 1\,\mathrm{TeV}$ [6].

1.2 Flavor changing neutral current in the SM

Transitions between "up" type and "down" type quarks are described in the SM Lagrangian with the coupling of $W^{\pm\mu}$ filed with the charged weak current J^{μ}:

$$J^{\pm\mu} = (\overline{u}, \overline{c}, \overline{t})_L \, \gamma^{\mu} \, V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{1.5}$$

where γ^{μ} are the Dirac matrices and V_{CKM} is the quark mixing matrix defined in (1.1). For a transition between two quarks q_i, q_j the amplitude depends on the CKM matrix element V_{ij} and since off-diagonal elements have non zero values, transitions between different flavors are allowed in the charged coupling.

On the other hand, the CKM mechanism does not allow flavor-changing neutral currents (FCNC). To show this, a transitions between two "down" type quarks would be described by a neutral current such as:

$$J^{0\mu} = (\overline{d}, \overline{s}, \overline{b})_L \, V_{CKM}^{\dagger} \, \gamma^{\mu} \, V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L$$

but the unitarity of CKM matrix leads to $V_{CKM}^{\dagger} \, V_{CKM} = \mathbf{1}$ implying that no FCNC is possible at the tree level. However, FCNC can still occur in SM at loop level through the so-called penguin or box Feynman diagrams. Figure 1.3 shows some examples of such Feynman diagrams in the $b \to s$ FCNC transitions. These processes are suppressed by a factor $\sim 10^2 \div 10^3$ with respect to the charged current flavor changing $b \to c$ transitions. Nevertheless, FCNC are particularly appealing for beyond SM searches, since new physics could potentially produce these processes at tree level, concurring with the loop level SM processes.

For B physics, decays containing $b \to s, d$ transitions are classified as rare, in comparison to the more abundant $b \to c$.

1.3 Motivation of $b \to s\tau^+\tau^-$ searches

Charged current semileptonic transitions $b \to c\ell^-\overline{\nu}$ are tree level processes in the SM, therefore they are not expected to be very sensitive to new physics. Nevertheless, different experiments have independently found consistent anomalies in the ratios between semi-tauonic and semileptonic B decays, defined as:

$$R(X) \equiv \frac{\mathcal{B}(B \to X\tau^-\overline{\nu}_{\tau})}{\mathcal{B}(B \to X\ell^-\overline{\nu}_{\ell})}\,; \qquad \text{with } X = D,\, D^*,\, J/\psi \qquad \textit{and } \ell = e,\, \mu \tag{1.6}$$

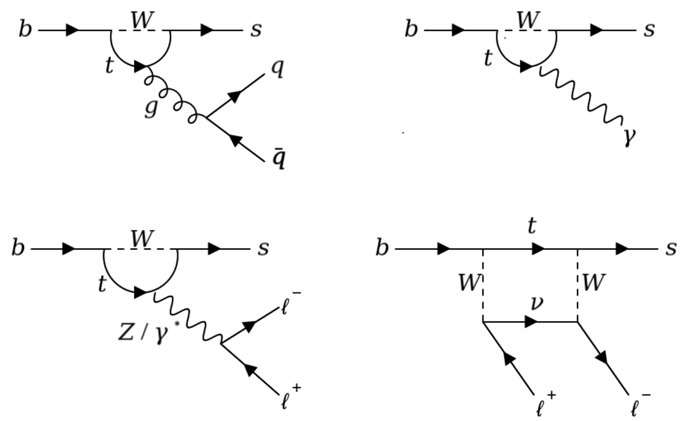


Figure 1.3: Examples of FCNC Feynman diagrams for FCNC $b \to s$ processes in the SM. Different name conventions are aopted depending on the final states: gluonic penguin (top left), radiative penguin (top right), electroweak penguin (bottom left) and box diagram (bottom right).

Such observables are theoretically clean, because the systematic uncertainties related to hadronic form factors cancel in the ratio. Combining together all the measurements of $R(D)$ and $R(D^*)$ from Belle, BaBar and LHCb (see Figure 1.4), and considering their correlation, a difference at the level of $3.2\,\sigma$ is found with respect to the SM theoretical prediction, based mainly on lattice calculations [7]. Furthermore, LHCb has measured an excess of about 2 σ in the ratio $R(J/\psi) = \mathcal{B}(B_c \to J/\psi\,\tau\nu)/(B_c \to J/\psi\,\mu\nu)$, with respect to the SM expectation, consistently with the R(D) and R(D*) analyses. [8].

These anomalies seem to suggest a violation of the flavor universality of the SM couplings. However, similar observables, testing the FCNC $b \to s$ ratios with light leptons only (i.e. e/μ), are now fully consistent with SM predictions [9, 10], pointing towards a LFU violation involving only the third flavor generation.

1.3.1 B decay anomalies in $b \to s$ neutral currents

The decay $B \to K^*\tau^+\tau^-$ is allowed in the Standard Model only at loop level through a penguin-like or a box diagram. It is therefore highly suppressed, with an expected branching ratio [11] of:

$$\mathrm{BR}_{\mathrm{SM}}(B \to K^*\tau^+\tau^-) = (0.98 \pm 0.10) \times 10^{-7} \tag{1.7}$$

averaging over the charged and the neutral modes. The leading theoretical uncertainties comes from the hadronic form factors, which here are computed in the limited $\tau^+\tau^-$ invariant mass range $15\,\mathrm{GeV}/c^4 < q^2 < 19\,\mathrm{GeV}/c^4$, in order to exclude the region of the $\Psi(2S)$ resonance.

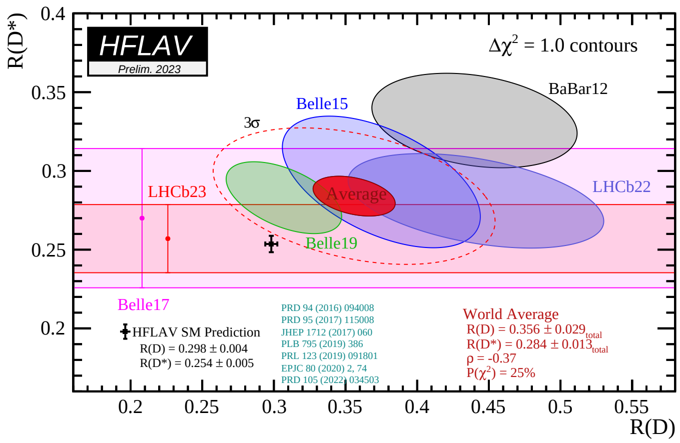


Figure 1.4: Status of $R(D)$ and $R(D^*)$ measurements (defined in eq. (1.6)). The red ellipse represents the world average, including both the separate measurements of $R(D^*)$ and results from simultaneous measurements of $R(D)$ and $R(D^*)$. The black point with error bars indicates the average of SM calculations. Plot from HFLAV preliminary averages for Winter 2023 [7].

Several new physics models, which account for the observed LFU anomalies, predict an enhancement by a factor up to 10^3 of this branching ratio [12]. As shown in Figure 1.5, the observed anomalies on the $R(X)$ ratio of equation (1.6) can be correlated to an approximately quadratic enhancement in the $b \to s\,\tau^+\tau^-$ transitions.

The BaBar experiment set the most stringent limit on the decay $B^+ \to K^+\tau^+\tau^-$ using 424 fb^{-1} of integrated luminosity collected at the resonance $e^+e^- \to \Upsilon(4S) \to B\overline{B}$. Tagging one B from the $\Upsilon(4S)$ in its hadronic decay and reconstructing the partner B as $B^+ \to K^+\tau^+\tau^-$ with $\tau \to \ell\nu\overline{\nu}$ ($\ell = e, \mu$), a 90% confidence level upper limit on the branching ratio at $\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 2.25 \times 10^{-3}$ was set [13].

Similarly, the Belle experiment, exploiting 424 fb^{-1} of integrated luminosity at the $\Upsilon(4S)$, has searched for the $B^0 \to K^{*0}\tau^+\tau^-$ decay. Tagging a B meson in a fully hadronic decay and looking, in the rest of the event, for all the possible combinations of $\tau \to \pi\nu$ and $\tau \to \ell\nu$ ($\ell = e, \mu$), a 90% confidence level upper limit on the branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 3.1 \times 10^{-3}$ was set [14].

Finally, another search in $b \to s\tau^+\tau^-$ was published by the LHCb collaboration for the decay of the strange-beauty meson $B^0_s \to \tau^+\tau^-$. Using 3 fb^{-1} of integrated luminosity collected at pp collisions, an upper limit on the branching ratio $\mathcal{B}(B^0_s \to \tau^+\tau^-) < 6.8 \times 10^{-3}$ was set at 95% confidence level [15].

None of this measurements observed an indication of a $b \to s\,\tau^+\tau^-$ signal, and all the upper limits set are consistent with the SM. However, they are still few order of magnitude above the predicted branching ratio of equation (1.7), and the space for the new physics scenarios, as those shown in Figure 1.5, is still open.

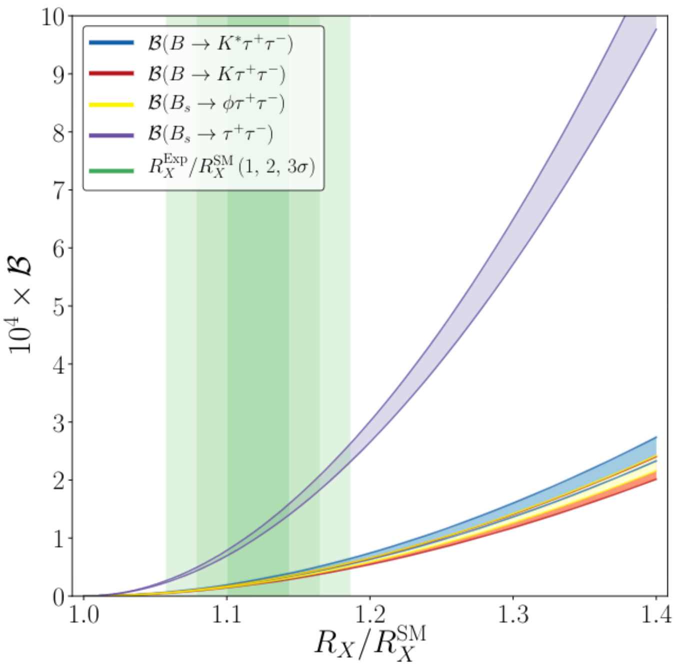


Figure 1.5: Predicted branching ratios for $b \to s\ \tau^+\tau^-$ processes as a function of the anomaly, with respect to the SM, of the $R(X)$ observables, defined in equation (1.6). The widths of the green vertical bands are the current experimental uncertainties at 1, 2 and 3 standard deviations, respectively. The widths of the parabolic bands represent only the theoretical uncertainties. No correlation is assumed. Plots from [12].

2 The Belle II experiment

In the present chapter, after a brief historical overview on the previous B factory experiments in section 2.1, the SuperKEKB collider and Belle II detector are outlined across sections 2.2 and 2.3, respectively. Section 2.4 contains a general description of the software tools implemented for the Belle II experiment and a summary of the physics performance of interest for the $B^0 \to K^{*0}\tau^+\tau^-$ analysis. Finally, the status of the operations of the SuperKEKB collider and of the Belle II collider are reported in section 2.5, together with the plans for the next years of data-taking.

2.1 B Factories

Starting from the end of the 1980s, a new concept of a e^+e^- collider machine operating at the center of mass (CM) energy of the $\Upsilon(4S)$ resonance started to be pursued [16]. As the $\Upsilon(4S)$ decays about 100% of the times to a pair of $B\overline{B}$ mesons, such a machine would have allowed to study in detail the properties of the heavy b quark in a clean environment, with a special focus on the CP violation predicted by the CKM mechanism [17]. The CLEO experiment at the CESR storage ring in Cornell (USA) and the ARGUS experiment at the DORIS-II accelerator at DESY (Germany) at that time had already collected a significant sample of B meson decays at the $\Upsilon(4S)$ CM energy, though not sufficient for CP violation measurements. The B factory concept was then designed as an asymmetric e^+e^- collider at the $\Upsilon(4S)$ [18], aiming to exploit the technical progress in e^+e^- storage rings to increase the instantaneous luminosity up to $10^{34}\,\mathrm{cm^{-2}s^{-1}}$. The detector requires high hermeticity, an inner light tracker with good vertex capabilities (to resolve the decay length of the B meson), a particle identification system to separate between charged pions and kaons, an electromagnetic calorimeter to measure the energy of electrons and neutral particles, and a sub-detector dedicated to the identification of muons and K_L^0 mesons.

In 1993, the first experiment adopting the B factory concept was approved: the BaBar experiment at SLAC (Stanford Linear Accelerator Center). In the following year, the Belle experiment at KEK (High Energy Accelerator Research Organization), in Japan, was approved as well. Both experiments started taking collision data in 1999 and in 2001 they reported the first evidence of CP violation in the $B^0\overline{B}^0$ system, in agreement with the prediction from the CKM model [19, 20]. These results eventually led to the assignation of the Physics Nobel Prize to Kobayashi and Maskawa in 2008.

BaBar and Belle stopped data-taking in 2008 and in 2010, respectively, having reached their integrated luminosity goals. BaBar (Belle) collected a total of $424.2\,\mathrm{fb^{-1}}$ ($711.0\,\mathrm{fb^{-1}}$) at the $\Upsilon(4S)$ resonance ("on-resonance" sample) plus about $94\,\mathrm{fb^{-1}}$ ($277\,\mathrm{fb^{-1}}$) scans at others $\Upsilon(nS)$ resonances and about 60 MeV below the $\Upsilon(4S)$ resonance ("off-resonance" sample). The plot in Figure 2.1 summarizes the integrated luminosity collected by the two experiments.

In the meantime, a different approach, exploiting b hadrons produced from proton-proton collisions, was developed for the LHCb experiment at CERN [21]. A single-arm spectrometer

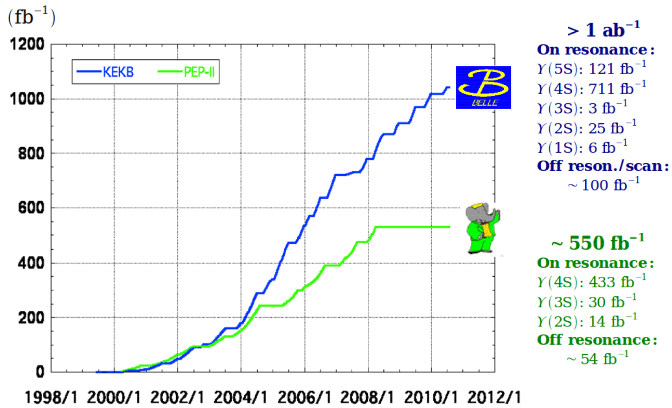


Figure 2.1: Total integrated luminosity collected by the two first generation B factories as a function of time. The e^+e^- colliders are KEKB and PEP-II for Belle and BaBar, respectively.

covering the forward direction exploits the very large b hadrons cross section $\sim \mathcal{O}(\text{mb})$ in the forward-backward direction. The main advantages of this approach over B factories are the higher event rate and longer decay length of the b hadrons, due to the larger boost, and the fact that all b hadrons species are produced at the LHC. On the other hand, the background rate and the pileup in a hadron collider is much larger with respect to a e^+e^- machine and the energy of the initial states is in general unknown.

Regarding B factories, the idea of an upgrade to a second generation super-B factory was put forward [22]. The core of these upgrade projects was to increase the instantaneous luminosity by at least an order of magnitude compared to the $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ designed for the first generation. Among the two proposed super-B factory projects, the SuperB experiment, to be realized at the National Laboratory of Frascati (Italy) [23], and the Belle II experiment, an upgrade of the Belle experiment, only the latter was eventually carried on.

2.2 The SuperKEKB collider

SuperKEKB is a 3 km length electron-positron collider located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan [24]. It is an upgrade of the previous generation B factory collider KEKB, which achieved the instantaneous luminosity record of $2.11 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ during the data-taking of the Belle experiment. SuperKEKB is designed to reach an instantaneous luminosity 30 times larger than its predecessor, thanks to a major redesign of the injection system and of the interaction point.

A schematic view of the SuperKEKB facility is shown in Figure 2.2. Electron and positron beams circulate in opposite directions in two adjacent storage rings, respectively called High Energy Ring (HER) and Low Energy Ring (LER), and collide in their only interaction point, where the Belle II detector is located. Particles are accelerated to their nominal energies by the Linear Accelerator (Linac) and injected into bunches in the main rings.

The major upgrades with respect to the KEKB involve the injector Linac, to reduce the beam emittance by more than an order of magnitude and at the same time double the intensity of the beam currents. For electrons, this is achieved through a new photocatode radio-frequency gun which generates a low-emittance short-pulse beam. Positrons are produced as secondary particles converting the electrons on a scattering target. Their luminosity is increased exploiting a capture section with a larger acceptance. The collected positrons are then cooled in a new 1.1 GeV damping ring to reduce their emittance up to the required level (vertical emittance to about 5 nm). Regarding the main rings, the beam optics is completely redesigned with new dipole magnets, and the interaction region is updated including super-conducting quadrupole magnets for a strong final focusing of the beams. Finally, the beam pipe is renovated to increase the vacuum level.

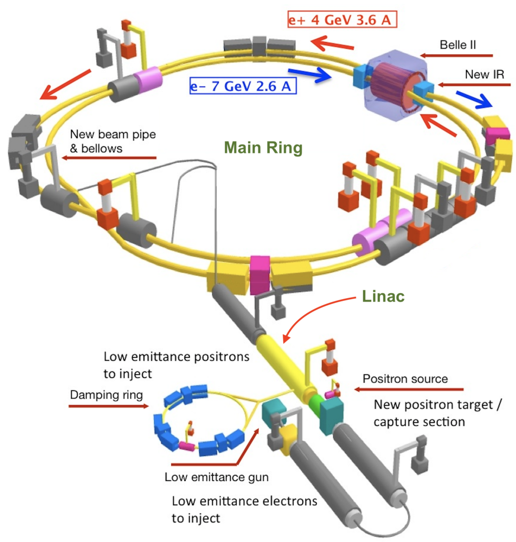


Figure 2.2: Schematic view of the SuperKEKB accelerator complex located at the High Energy Accelerator Research Organisation (KEK) in Tsukuba, Japan. Electron and positron beams are accelerated in the LINAC and injected into two circular storage rings, so that they circulate in opposite directions. A Damping Ring was built to reduce the positron emittance, as required for the nano-beam collision scheme.

Since $B\bar{B}$ mesons are produced almost at rest in the $\Upsilon(4S)$ frame ($p_B^{\text{CM}} \simeq 330\,\text{MeV}$), for precise measurements of time-dependent CP-violation in B decays the center of mass needs to be boosted in the laboratory. The boost is obtained by using asymmetric beam energies E_{e^+}, E_{e^-}, which must satisfy:

$$4E_{e^+}E_{e^-} = m^2_{\Upsilon(4S)}$$

neglecting the electron mass. The desired boost factor is given by:

$$\beta\gamma \simeq \frac{E_{e^-} - E_{e^+}}{m_{\Upsilon(4S)}}$$

At SuperKEKB, $E_{e^+} = 4\,\mathrm{GeV}$ and $E_{e^-} = 7\,\mathrm{GeV}$, resulting in $\beta\gamma \simeq 0.28$ and a B meson mean decay length of $\lambda_B \approx 130\,\mu\mathrm{m}$. The CM boost is about two-third of what experienced in Belle, and half of the boost designed for BaBar, thus requiring a better vertex resolution. The symmetry reduction with respect to the predecessors is necessary to further limit beam emittance and develop the high-luminosity scheme.

2.2.1 Nano-beam scheme

The luminosity upgrade of the KEKB facility to SuperKEKB is based on the nano-beam collision scheme developed for the SuperB project [23].

For an electron-positron collider, the instantaneous luminosity can be expressed as [25]:

$$\mathcal{L} = \frac{\gamma_{e^\pm}}{2er_e}\left(1 + \frac{\sigma_y^*}{\sigma_x^*}\right)\left(\frac{I_{e^\pm}\cdot\xi_{y,e^\pm}}{\beta_y^*}\right)\left(\frac{R_L}{R_{\xi_y}}\right) \tag{2.1}$$

where $\gamma_{e^\pm}$ is the Lorentz factor for the colliding $e^\pm$, r_e is the classical electron radius, σ_x^* and σ_y^* are the horizontal and vertical beam sizes at the interaction point (IP), assumed to be equal for electron and positron. $I_{e^\pm}$ are the beam currents and β is the function describing the amplitude of betatron oscillations in the transverse plane while β_y^* is its restriction on the y direction calculated at the IP. ξ_y is the vertical beam-beam tune shift, accounting for resonances due to bunch crossing. R_L and R_{ξ_y} account for the reduction of the luminosity and the vertical beam-beam parameter respectively, due to the actual crossing angle of the beams. The main parameters to be optimized for achieve high luminosity are β_y^*, $I_\pm$, and ξ_y.

At SuperKEKB, the vertical beam size will be eventually squeezed approximately by a factor $\times 15$ more than at the previous B factory, going from $\beta_y^* = 5.9\,\mathrm{mm}$ obtained at KEKB, to the designed value $\beta_y^* = 0.4\,\mathrm{mm}$. At the same time, the beam currents $I_{e^\pm}$ are doubled, thus leading to an expected increase in the instantaneous luminosity by a factor of about 30 with respect to KEKB.

The value of β function increases quadratically moving away from the interaction point, therefore long bunches are actually less focused than what predicted by the β^* entering equation (2.1). This effect, known as hourglass effect, lowers the expected instantaneous luminosity [26]. To avoid this, the longitudinal beam size should be kept smaller than β_y^*. Within nano-beam scheme, such condition is fulfilled making the two beams collide with an incident angle ϕ. In this way the actual beam length d, relevant for the hourglass effect, is determined by the overlapping region, which, as clear from Figure 2.3, is:

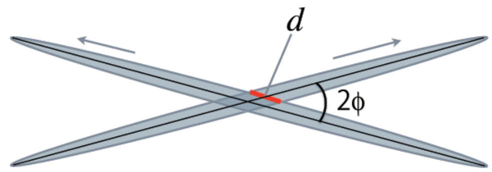


Figure 2.3: Principle of the nano-beam collision scheme. The two bunches overlap with a half-crossing angle ϕ, hence the effective beam length is $d \simeq \sigma_x^*/\phi$. With this scheme the luminosity can be increased limiting the hourglass effect.

Table 2.1: Designed machine parameters [27] for the SuperKEKB e^+e^- collider. Beam energy, bunch density, and optics differs between the High Energy Ring (HER) and the Low Energy Ring (LER). The * symbol indicate quantities at the interaction point.

	LER	HER
Beam energy [GeV]	4.007	7.000
Beam current [A]	3.60	2.60
Crossing angle ϕ [mrad]	41.5	
Bunch number	2500	
Particles per bunch	9.04×10^{10}	6.53×10^{10}
β_y^* [mm]	0.30	0.27
Horizontal beam size σ_x^* [µm]	10.1	10.7
Vertical beam size σ_y^* [nm]	48	62
Beam–beam parameter ξ_y	0.088	0.081
Luminosity $\mathcal{L}$ [cm^{-2}s^{-1}]	6×10^{35}	

$$d = \frac{2\,\sigma_x^*}{\sin 2\phi} \simeq \frac{\sigma_x^*}{\phi}$$

thus the requirement becomes:

$$d < \beta_y^* \qquad \longrightarrow \qquad \frac{\sigma_x^*}{\phi} < \beta_y^*$$

In table 2.1 some design parameters of SuperKEKB collider are reported. During the first three running period of SuperKEKB (2019 - 2022), the instantaneous luminosity of 4.71×10^{34}cm^{-2}s^{-1} was reached. The planned luminosity roadmap is detailed in section 2.5.

2.3 The Belle II detector

The Belle II detector [25], situated at the SuperKEKB interaction region, is designed to take data in a high background environment resulting from the increased beam currents and a stronger beam focusing with respect to Belle.

A Cartesian right-handed system of coordinates of Belle II is defined with the z axis forming the same angle ϕ with the two nominal beams directions, such that in the CM it comes to point towards the electron beam direction. The transverse plane is defined by the x axis lying on the machine plane and pointing towards the center, and the y axis pointing upwards. θ is the polar angle respect to the positive z direction and φ the azimuthal angle in the transverse plane.

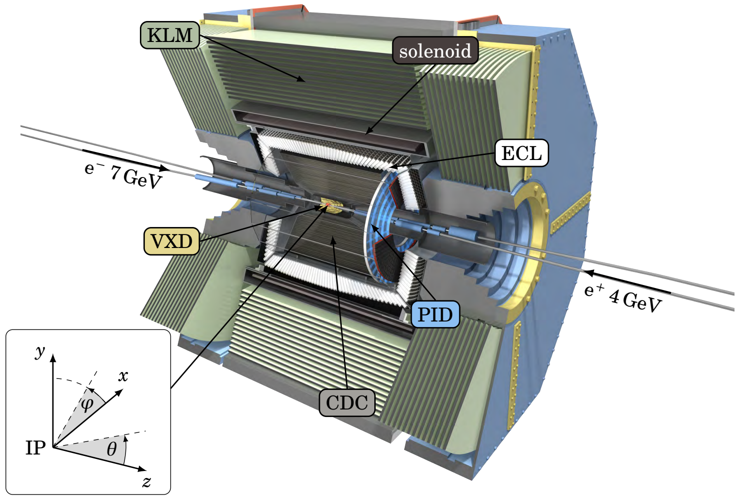


Figure 2.4: Cross section view of the full Belle II detector. The superconducting quadrupoles (QCS) are placed along the beam line on the two sides.

The Belle II detector is characterized by high hermeticity, with a full coverage in the azimuthal angle φ and an asymmetric acceptance in the polar angle $17^\circ < \theta < 150^\circ$. Considering the forward boost, the acceptance in the CM system becomes $21^\circ < \theta^{\mathrm{CM}} < 156^\circ$, which is almost symmetric.

A superconducting magnet generates a uniform solenoidal magnetic field of 1.5 T in the central part of the detector. For precise momentum measurement the magnetic field is carefully mapped taking into account contributions from the focusing superconducting quadrupoles.

Figure 2.4 shows a cross section on the x-z plane of the full detector. The tracking system, composed of a silicon vertex detector (VXD) and a gaseous drift chamber (CDC), measures charged particles momenta and reconstructs production and decay vertexes. Photons and electrons are then stopped making them showering in a homogeneous electromagnetic calorimeter (ECL). Muons and K_L are discriminated in the outermost subdetector (KLM), which exploits the iron return yoke of the magnetic field to detect long lived neutral hadrons. Two non-destructive Cherenkov imaging particle identification detectors (PID) are placed between the CDC and the ECL, in the barrel and in the forward region respectively.

In the following, starting from the center and going outward, each subdetector is briefly described. A partial intermediate upgrade of the Belle II detector, involving mainly the VXD and PID, has been carried on during the first long shutdown (LS1) period, between 2022 and 2023. The specific interventions are described separately for each subsystem.

2.3.1 Vertex Detector

The Belle II innermost detector, located just outside the Beryllium beam pipe and surrounding the IP, is a 6-layers cylindrical vertex detector (VXD) composed of a inner pixel detector and an outer silicon strip detector. The VXD task is to measure, with high precision, tracks of charged particles and reconstruct the decay vertexes, e.g. of B-mesons, D-mesons, τ-leptons and K_S^0.

In order to improve vertex reconstruction in comparison to Belle, the first VXD layer is placed closer to the IP halving the beam pipe radius to 10 mm. At the same time the VXD external radius is increased from 60.5 mm to 135 mm.

A picture of the full Belle II VXD after the 2022/23 LS1 upgrade, is shown in Figure 2.5.

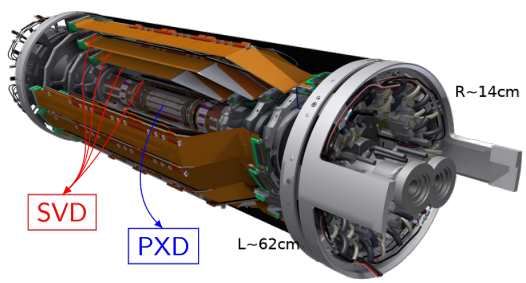


Figure 2.5: Picture of the full Belle II vertex detector as it is after the installation of the updated PXD in the summer 2023.

Pixel Detector

At high luminosity the detector layers closest to the interaction region face an elevated hit rate due to beam background and low momentum QED processes. Hence, to limit channel occupancy, the two inner layers of VXD placed at $r = 14$ mm and $r = 22$ mm are a DEPFET (DEPleted Field Effect Transistor) Pixel Detector (PXD) [28]. A rendering of the PXD layers is shown in Figure 2.6. In the DEPFET technology the readout and the cooling system can be placed away from the 50 µm-thick active sensors, thus minimizing the material budget responsible for multiple scattering.

During the first period of Belle II data taking (2019 - 2022), the PXD was operated in a preliminary setup with an entire first layer and half of the second layer. This first detector, called PXD1, was replaced in the LS1 period by a brand new pixel detector (PXD2) having two entire layers. A picture of the replaced VXD with PXD2 is shown in Figure 2.5.

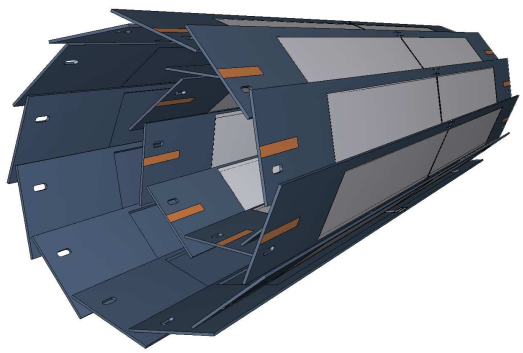


Figure 2.6: Schematic view of the two-layers DEPFET Pixel Detector.

Silicon Vertex Detector

The remaining 4 layers are based on 320 µm thick double sided silicon strip detectors (SVD) placed at 39 mm, 80 mm, 104 mm, and 135 mm distance from z axis [29]. For the sensors of the inner layer (L3), strips on n-side are along the z direction with a 50 µm pitch, while on the p-side strips are oriented on $r\varphi$ and have a 160 µm pitch. For the other layers (L4, L5 and L6) the pitches increase to 75 µm on the p-side and 240 µm on the n-side. To reduce the number of wafers, in the forward region the 3 external layers use trapezoidal slanted sensors 300 µm thick (see Figure 2.7). The total material budget of each layer is about 0.7% radiation lengths. The readout electronics is implemented using the APV25 amplifier chip having a shaping time of 50 ns.

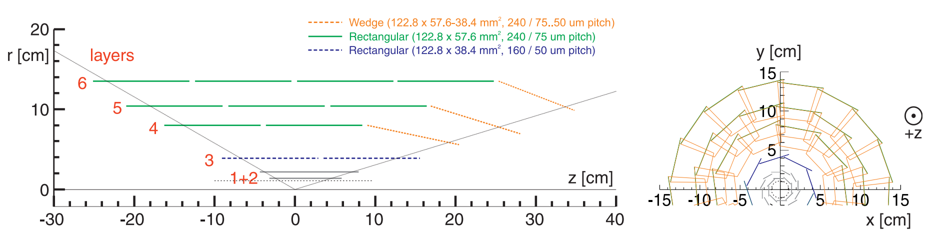


Figure 2.7: Schematic longitudinal (left) and transverse (right) view of the SVD layers. Note the geometrical arrangement of the slanted sensors in the 3 outermost layers.

2.3.2 Central Drift Chamber

The tracking system is completed by the Central multi-wire Drift Chamber (CDC). This subdetector, besides tracking, provides also particle identification information measuring the dE/dx energy loss for tracks not reaching the PID detector ($p_t \lesssim 500\,\mathrm{MeV}$) and has an important role for the first level trigger as well. CDC is filled with a gas mixture of He (50%) and C_2H_6 (50%) having a drift velocity of 4 cm/µs. Wires are arranged in drift cells grouped radially on 56 layers alternating axial layers, with wires directed along z, and stereo layers, with skewed wires, making possible a 3D track reconstruction. Each cell is defined as one sense wire surrounded by 8 field wires. In total 14336 Tungsten sens wires and 42240 Aluminum field wires are present, respectively of diameter 30 µm and 126 µm.

In comparison to Belle, a better momentum resolution is obtained through a larger external radius (113 cm instead of 88 cm) and smaller drift cells make possible to operate under higher event rate.

2.3.3 Particle Identification Detector

The Particle Identification Detector (PID) is deputed mainly to discriminate kaons, pions and protons in a wide energy range. A good K-π separation is of particular importance for tagging different B decay modes. PID use two distinct detectors depending on the acceptance region: the barrel is provided with a Time of Propagation counter (TOP) while in the forward region the particle identification is guaranteed by a Ring Imaging Cherenkov Detector (ARICH). Both work measuring the angle of the emitted light Cherenkov cone, which is sensible to the particle velocity.

Time of Propagation

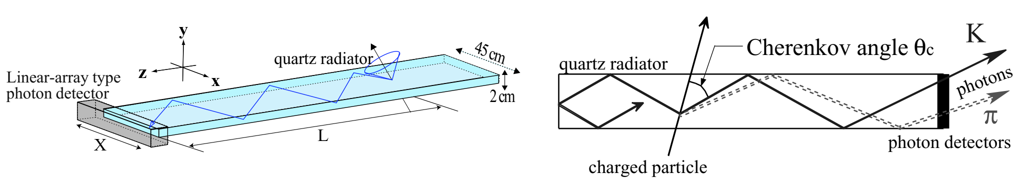


Figure 2.8: Scheme of the quartz bar used for the TOP detector. Cherenkov photons, subjected to total internal reflection, reach the photon detectors placed at one edge of the bar. The x, y positions and the time of arrival of the photons are used to reconstruct the Cherenkov angle.

The Time of Propagation (TOP) counter is composed of 16 quartz bars, 2 cm thick, 45 cm wide and 2.75 m long, arranged to cover the full φ angle around the CDC external radius. The polar angle acceptance is $31^\circ < \theta < 128^\circ$ [30]. The Cherenkov photons generated by the passage of a charged particle propagate through the bar via internal reflection and are collected at one edge (see Figure 2.8). For the photodetectors located on the edge of the bars, a minimum of 100 ps time resolution on a single photon is required for K/π particle identification. An array of microchannel plate photomultipliers (MCP-PMTs) is chosen to measure position and time of arrival reconstructing the 3D Cherenkov image.

During the LS1 upgrade, all the MCP-PMTs have been replaced with new ones to overcome degradation in quantum efficiency after three years of operation.

Ring Imaging Cherenkov Detector

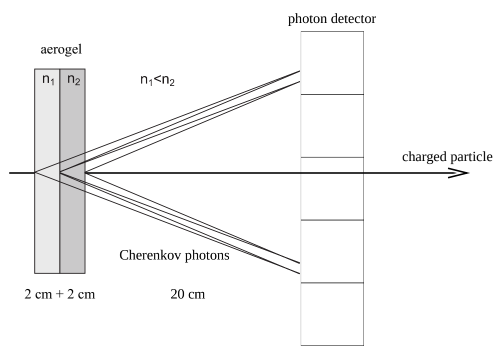


Figure 2.9: Operating principle of the ARICH detector. The charged particle crosses the non-homogeneous aerogel and the emitted Cherenkov photons are detected by the HAPDs, which reconstruct the Cherenkov ring.

In the forward endcap a proximity focusing aerogel Ring Imaging Cherenkov Detector (ARICH) is installed, covering the region $14^{\circ} < \theta < 30^{\circ}$. It is designed to provide a good photon separation in the momentum range from 400 MeV to 4 GeV. It consists of a 4 cm thick non-homogeneous aerogel Cherenkov radiator and an array of hybrid avalanche photon detectors (HAPD), separated by a 20 cm expansion gap. The operating principle of the ARICH is shown in Figure 2.9. The HAPD are sensible to the single photon position and estimate the Cherenkov angle reconstructing the full ring. In order to increase the Cherenkov photon yield maintaining a good angle resolution, the ARICH employs two materials with different refraction index $n_1 < n_2$ so that a focusing effect is obtained.

2.3.4 Electromagnetic Calorimeter

An electromagnetic calorimeter at Belle II should be able to detect photons in a broad spectrum with good energy resolution, and also to discriminate electrons from charged hadrons.

The Belle II Electromagnetic Calorimeter (ECL) maintains the same structure of the Belle ECL, with a slightly improved readout, to cope with the expected increase in beam background. It is made by a total of 8736 CsI(Tl) crystals, approximately $(5 \times 5 \times 30)\text{cm}^3$

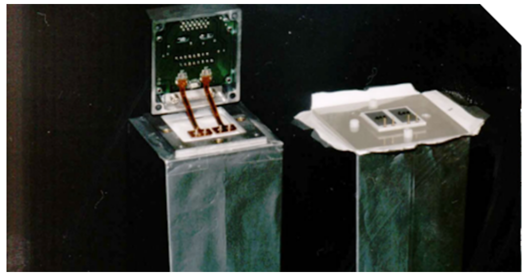


Figure 2.10: Picture of Belle II CsI(Tl) crystals showing the pin-diode photo sensors and the pre-amplifier.

each, with the longitudinal dimension corresponding to 16.2 radiation lengths. They are arranged in a projective geometry covering three regions in the polar angle: the barrel for the central region and two endcaps for the forward and backward region. The acceptance of each region is reported in table 2.2, note that also in this detector case the coverage is asymmetric to account for the boost in the detector reference frame.

Table 2.2: Polar angle coverage of the ECL.

Forward endcap	$12.01^\circ < \theta < 31.36^\circ$
Barrel	$32.2^\circ < \theta < 128.7^\circ$
Backward endcap	$131.5^\circ < \theta < 155.03^\circ$

The scintillation light emitted by each CsI(Tl) crystal is read by two pin-diodes attached on the rear face of each crystal and equipped with a charge sensitive pre-amplifier (see Figure 2.10). The signal is then sent to a board with a 500 ns shaper, a 1.7 MHz waveform digitizer and an FPGA for online fit of the shaped waveform. For all triggered events, the waveform amplitude and time resulting from each crystal are saved.

The intrinsic energy measured on a Belle prototype can be parameterized as [31]:

$$\frac{\sigma_E}{E} = \frac{0.066\%}{E} \oplus \frac{0.81\%}{E^{1/4}} \oplus 1.34\%$$

where E is the photon energy in GeV and the first term comes from the electronic noise.

2.3.5 K_L and μ detector

The outermost detector identifies long lived particles that do not undergo electromagnetic showering and separate muons from charged hadrons (pions in particular).

The K_L and muon detector (KLM), located outside the ECL and the superconducting magnet, consists of a sandwich of active detectors and 4.7 cm thick iron, used as return yoke for the magnetic field. The total thickness is equivalent to about 3.9 interaction lengths, to be summed to the ≈ 0.8 interaction lengths of the ECL. K_L mesons are hence detected through hadronic showers which can start either from ECL or from KLM. Muons and non-showering charged hadrons cross the detector as minimum ionizing particles and are identified from the range information.

The KLM is divided in three sections, one for the barrel and two for the endcaps, in total covering a polar angle $20^\circ < \theta < 155^\circ$. The barrel and the endcaps are segmented into 14 and 12 active layers, respectively. In the barrel the active detectors are glass-electrode resistive plate chambers (RPC), already utilized at Belle. However, the long RPC dead time make them not optimal for the Belle II endcaps, where a much higher background rate due to neutrons coming from the beam line is expected. Therefore, in the forward and backward sections and in the first 2 layers of the barrel, RPCs have been substituted by scintillator strips read out by silicon photomultipliers [32].

2.3.6 Trigger

The Belle II trigger system is divided into two levels. A hardware-based trigger, called Level 1 (L1), reduces the event rate to 30 kHz, so that the collected information can be transferred to the data acquisition system. The L1 trigger should maintain a high efficiency for both low and high multiplicity processes. It mainly exploits the CDC and the ECL subsystems: in the CDC a dedicated trigger line performs a fast track finding to estimate the number of high momentum tracks, while for the ECL the total deposited energy and the number of reconstructed clusters are provided to the trigger. A veto for Bhabha events is also implemented to get rid of the background process $e^+e^- \to e^+e^- \ (\gamma)$, which would trigger both the CDC and the ECL, and it has a much higher cross section with respect to the physics processes of interest, as can be seen from table 2.3.

The second level of the trigger system is called High Level Trigger (HLT) [33]. It receives the signal from the L1 trigger, further reducing the final rate to a maximum of 10 kHz, for offline storage. The HLT exploits the information from all subsystems, except for the PXD, to perform a full reconstruction of the event. As for the L1, the HLT trigger lines are designed to keep a high efficiency on all the channels studied in the Belle II physics program. In particular, for $B\overline{B}$ events, the total trigger efficiency can be considered to be 100%.

2.4 Software and reconstruction

Belle II needs an efficient and reliable software for online and offline data handling. The Belle II software analysis framework (`basf2`) is implemented with processing blocks called *modules*, each executing a defined task [34, 35]. A sequence of modules makes up a *path*, hence, when a path is processed, the modules attached to it are executed in order. Different modules communicate sharing data and variables in a common object store.

Table 2.3: e^+e^- cross section of different processes at $\sqrt{s} = 10.58\,\mathrm{GeV}$.

Process	Cross section [nb]
$e^+e^- \to \Upsilon(4S)$	1.110 ± 0.008
$e^+e^- \to u\bar{u}\ (\gamma)$	1.61
$e^+e^- \to d\bar{d}\ (\gamma)$	0.40
$e^+e^- \to s\bar{s}\ (\gamma)$	0.38
$e^+e^- \to c\bar{c}\ (\gamma)$	1.30
$e^+e^- \to \tau^+\tau^-\ (\gamma)$	0.919
$e^+e^- \to e^+e^-\ (\gamma)$	$\sim$300
$e^+e^- \to e^+e^-\ e^+e^-$	$\sim$40
$e^+e^- \to e^+e^-\ \mu^+\mu^-$	$\sim$20
$e^+e^- \to \gamma\gamma\ (\gamma)$	$\sim$5
$e^+e^- \to \mu^+\mu^-\ (\gamma)$	1.148

Basf2 handles the full Belle II software, including the reconstruction of high-level objects from raw dataset and the analysis tools for the user.

2.4.1 Monte Carlo generators

An accurate simulation for particle production and decay is fundamental for precision measurements. Beyond the analysis tools, `basf2` includes all the modules dedicated to the Monte Carlo (MC) simulation of physics processes and detector response. The physics process to be simulated for e^+e^- collisions at $\sqrt{s} = 10.58\,\mathrm{GeV}$ are not limited to the $\Upsilon(4S)$ since the cross sections into lepton pairs and lighter hadrons are still $\sim$ nb. The main simulated processes are summarized in table 2.3, they also include Bhabha and radiative Bhabha scattering. Several Monte Carlo (MC) generators are inherited from external packages, in particular to simulate production and particle decays of all the relevant processes at the $\Upsilon(4S)$ energy.

The `PYTHIA` generator is used for the production of light quark continuum $e^+e^- \to q\bar{q}$ ($q = u, d, s$) and inclusive decays of light quarks [36]. The production $e^+e^- \to c\bar{c}$, $b\bar{b}$ and the exclusive decay chain of B and D mesons are handled by `EvtGen` generator [37]. Bhabha events and other QED background processes such as two photons are generated by `BABAYAGA.NLO`. Low multiplicity $e^+e^- \to \tau^+\tau^-(n\gamma)$ and $e^+e^- \to \mu^+\mu^-(n\gamma)$ processes are simulated by the `KK2f` generator, included in the `KKMC` module [38]. Tau decays are handled by the TAUOLA package, taking into account spin polarization effects [39, 40].

The detector simulation is implemented in the `basf2` software by the `Geant4` package which simulates the interaction of long lived and secondary particles with the subdetectors

[41]. Other `basf2` modules are responsible for the subsequent hit digitization. The beam energy and all the data taking condition parameters are provided by a central database.

Beam background

Another important component for the MC simulation of the Belle II environment is the beam-induced background. Especially for a high luminosity collider, this component cannot be neglected and it may strongly impact the physics results. The main processes contributing to this kind of background in the Belle II detector are Touschek scattering, beam-gas scattering, radiative Bhabha, and synchrotron radiation.

The Touschek process is an intra-bunch Coulomb scattering, i.e. a scattering between two particles belonging to the same bunch. These particles are thus deviated from the beam direction, and they may interact with the beam pipe material, causing a shower of secondary particles in the interaction region. The rate of Touschek scattering is inversely proportional to the beam size, therefore it is particularly harsh around the interaction point. It can be mitigated locating some collimators just before the Belle II detector.

The beam-gas scattering is due to the interaction of beam particles with residual gas molecules in the beam pipe. The interaction can consist either in Coulomb a scattering or in a Bremsstrahlung and the effect is similar to the Touschek scattering. This source of beam background is proportional to the beam currents. As for the Touschek, beam-gas scattering can be reduced through some collimators and improving the vacuum level as well.

Radiative Bhabha $e^+e^- \to e^+e^- \gamma$ occurs in the interaction point with a wide scattering angle. The radiated photon is emitted along the beam pipe direction and it may interact with the iron of the magnets, with a high neutrons production cross section. Such neutrons are the major background source for the KLM detector. Similarly, the scattered electron and positron are slightly deviated from their original direction, therefore, after the interaction point they are over-bent by the focusing magnets and they may interact with the wall of the magnets as well. In this case, an electromagnetic shower is generated, which affects mainly the ECL detector. The radiative Bhabha, and the secondary backgrounds originating from that, scales linearly with the instantaneous luminosity.

Finally, the synchrotron radiation is emitted by beam particles and it has a spectrum in the X-rays region. It is proportional to the square of the beam energy, hence it is more relevant for electrons in the HER. The beam pipe is specifically designed in order to minimize the synchrotron radiation impinging on the VXD, in particular it is coated with a gold layer in the inner surface.

To account for these processes in the simulation, the detector responses from beam background events are directly superimposed to the simulated physics processes. However, an accurate simulation of all the beam-induced background processes is not straightforward and it can be computationally intensive. To deal with this, the amount of beam background processes is increased by reshuffling the same events over different MC generated physics events, so that the same beam background is reused multiple times.

Belle II adopts two complementary approaches to superimpose beam background in the simulated samples, called run-dependent and run-independent, respectively. The beam background rate depends indeed on the machine running condition, such as the luminosity value, the beam currents, the β^* focusing and the setting of collimators. The run-independent approach neglects all these dependencies using a fixed reference. Though less accurate, this approach has the advantage to be computationally more affordable. On the opposite side, in the run-dependent approach, different parameters for beam background are used, based on the actual running conditions and on their evolution during the full data-taking period. Therefore, the run-dependent MC sample is made of multiple subsamples, each corresponding to an actual running condition. Furthermore, the run-dependent samples are generated reproducing the variations in the detector conditions, as well. For instance, information about dead channels, calibrations and variations in temperature can be taken into account when adopting the run-dependent approach.

2.4.2 Tracking

Track objects are determined connecting the hits of charged particles in VXD and CDC detectors [42]. At this aim, a different pattern recognition algorithm is exploited for each subdetector. The VXD track finder is based on a cellular automaton model, while for the CDC a global track finder algorithm which exploits a Hough transformation is combined with a local track finder using a cellular automaton to deal with energy losses and displaced tracks [43, 44]. The track candidates from VXD and CDC are then merged together according to their relative distance.

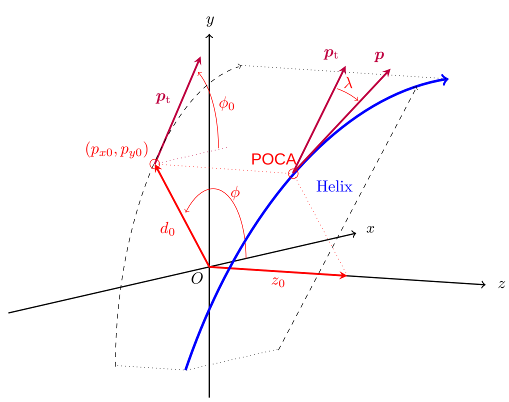


Figure 2.11: Scheme showing the tracking parameters used in the `basf2` software. The blue arrow represents the reconstructed helix. Credits: `github.com/ndawe/tikz-track`.

The coordinates of the POCA (point of closest approach to the origin of the coordinate system) and the momentum vector at the POCA are associated to every reconstructed track. The trajectory of a charged particle in a uniform magnetic field is described by a helix, defined by 5 independent parameters. The following parametrization is adopted for tracks at Belle II (see Figure 2.11):

- d_0: the signed distance of the POCA respect to the origin in the transverse plane. The sign of d_0 depends on the sign of the vector product $\vec{d_0} \times \vec{B}$, where $\vec{B}$ is the magnetic field.
- z_0: the z coordinate of the POCA.
- ϕ_0: the angle between the transverse momentum p_t at the POCA and the x axis.
- $\tan\lambda$: λ is the angle between the momentum at the POCA and the transverse plane.
- ω: the curvature of the helix, with the sign given by the charge of the track.

The track expressed under such parameterization is then fitted using an algorithm based on a Kalman filter [45].

The two plots in Figure 2.12 show the tracking efficiency as a function of transverse momentum p_T, measured from $e^+e^- \to \tau^+\tau^-$ events. For a single track in the high transverse momentum range $0.2\,\mathrm{GeV}/c < p_T < 5.0\,\mathrm{GeV}/c$ the agreement between data and simulation is very good with a 0.3% systematic uncertainty. While for slow pions in the momentum range $50\,\mathrm{MeV}/c < p_T < 200\,\mathrm{MeV}/c$ the systematic uncertainty increases to 2.1%. This second measurement, however, is statistically limited and it is expected to improve with more collected data.

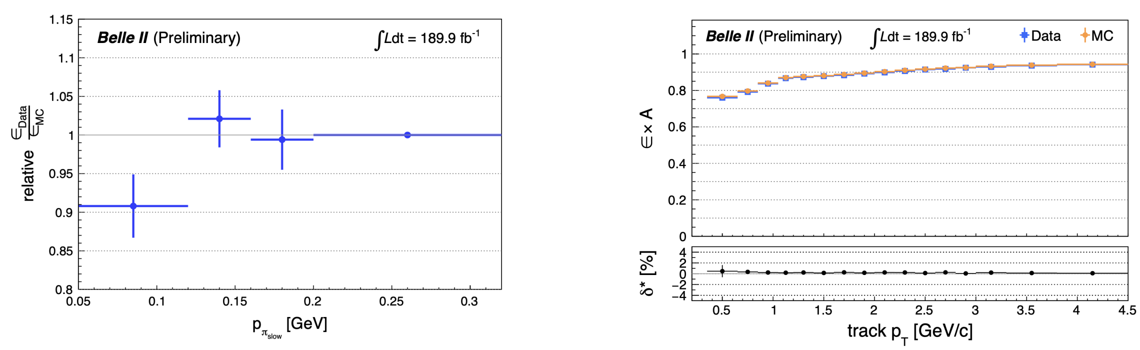
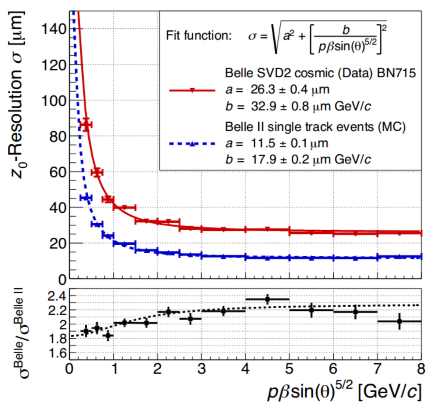


Figure 2.12: Ratio between tracking efficiency measured on data and on simulation for slow pions (left plot). Tracking efficiency for transverse momentum $p_T > 500\,\mathrm{MeV}/c$ on data and simulation (right plot).

Another interesting metric to evaluate tracking performance and vertex capabilities, is the impact parameter resolution, i.e. the spatial resolution of the d_0 and z_0 track parameters. The left plot of Figure 2.13 shows the impact parameter resolution in the longitudinal plane z as a function of the particle momentum, multiplied by $\sin^{5/2}\theta$, to factorize the dependence from multiple scattering, which dominates in the low $p\beta \sin^{5/2}\theta$ region, where θ is the polar angle of the particle. The z_0 resolution is plotted separately for Belle II MC simulation and Belle data to emphasize the improvement stemming from the new Belle II VXD. A z_0 resolution below 15 µm is reached for high momentum tracks.

Regarding the transverse plane, the right plot of Figure 2.13 shows the measured d_0 resolution as a function of the track polar angle ϕ_0 on Belle II simulation and an early collected dataset of 21.1 pb^{-1}. As for z_0, the impact parameter resolution in the transverse plane d_0 is below the 15 µm threshold.



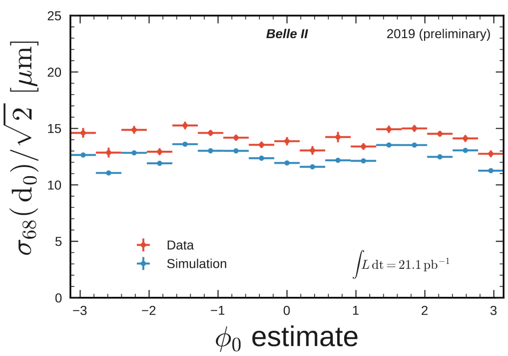


Figure 2.13: Impact parameter resolution in the z direction (left plot) and in the transverse $x - y$ plane (right plot). The left plot compares Belle II simulation against Belle data, showing a factor of about ×2 improvement in the resolution. In the right plot, Belle II data and simulation are shown together as a function of the azimuthal angle ϕ_0.

2.4.3 Clustering

The main task of the electromagnetic calorimeter is to detect photons and neutral hadrons with a good energy and position resolution. From a software point of view, physical objects are reconstructed in the calorimeter as clusters. A cluster is built summing up the energy deposits in contiguous ECL crystals. To decide which crystals to include in a cluster, a cluster reconstruction algorithm is used.

Firstly, connected regions are formed starting from a crystal with an energy deposit above 10 MeV (called local maximum), and including the neighboring crystals with energy deposit over a background threshold of 0.5 MeV. Connected regions sharing crystals are merged. Clusters are then formed using as seed the local maximum crystals and adding iteratively the other crystals of the connected region. The energy of the i-crystal assigned to the cluster is weighted by:

$$w_i = \frac{E_i\, e^{-Cd_i}}{\sum_k E_k\, e^{-Cd_k}} \tag{2.2}$$

where C is a constant determined from simulation, and the index k runs over all the crystals in the cluster. E_i is the energy deposited in the i-crystal and d_i is the distance from its geometric center $\vec{x}_i$ to the cluster's centroid $\vec{x}$, defined as:

$$\vec{x} = \frac{\sum_i w_i' \vec{x}_i}{\sum_i w_i'} \qquad \text{with} \qquad w_i' = 4.0 + \ln\left(w_i + \frac{E_i}{E_{\text{all}}}\right)$$

where E_{all} is the sum of all the weighted energies in the cluster. The centroid position and the weights w_i are updated each time a new crystal is added to the cluster. The equation (2.2) is motivated by the lateral shape of an electromagnetic shower, which decreases exponentially from the cluster center.

If the same connected regions contain multiple local maxima, each one generated a distinct cluster and the energy of a single crystal can be shared among the different clusters. The iteration terminates when the centroid positions of all clusters in the connected region are stable within 1 mm.

Eventually, the total energy E_{cluster} of the reconstructed cluster is corrected against any bias, using simulated events. In particular, the energy leakage of the shower and the additional energy from beam background photons are taken into account for the corrections.

All the ECL clusters which cannot be matched to a charged track are classified as neutral clusters. Neutral clusters are then classified as photons based on their lateral shower shape development. Figure 2.14 shows the photon efficiency reconstruction as a function of the polar angle θ, obtained from radiative $e^+e^- \to \mu^+\mu^-\ \gamma$ events, where the recoil momentum p_{Recoil} of the two muons is measured [46]. The photon reconstruction is reproduced by MC simulation with a systematic uncertainty at 0.3% level in the barrel region for $p_{\text{Recoil}} > 2.0\,\text{GeV}/c$.

2.4.4 Identification of charged particles

The long-lived charged particles to be separately identified by the Belle II detector are $\pi^\pm$, $K^\pm$, $e^\pm$, $\mu^\pm$, (anti-)protons and (anti-)deuterons. The PID system, composed by the TOP and the ARICH detectors, provides specific information for particle identification, using time of propagation and Cherenkov emission. This can be combined with measurements from other subdetectors, such as dE/dx from CDC and SVD, ECL cluster shapes and the KLM energy deposits. For a given particle hypothesis α, an independent likelihood $\mathcal{L}_\alpha^{\text{det}}$ is built for each subdetector. The information from all subdetectors is then combined into a global likelihood:

$$\mathcal{L}_\alpha = \prod_{\text{det}} \mathcal{L}_\alpha^{\text{det}} \qquad\qquad \alpha = \pi^\pm,\ K^\pm,\ e^\pm,\ \mu^\pm \tag{2.3}$$

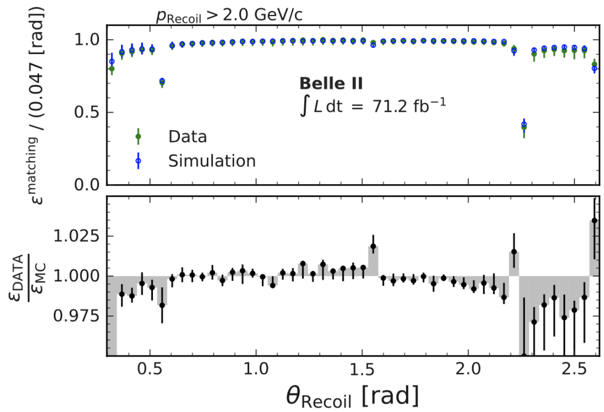


Figure 2.14: Photon reconstruction efficiency for photon energy above 2.0 GeV. The efficiency is measured exploiting the $e^+e^- \to \mu^+\mu^- \gamma$ events, on both Belle II data and simulation [46].

The particle-ID variable, for the particle α, is thus defined as:

$$\text{PID}(\alpha) = \frac{\mathcal{L}_\alpha}{\sum_\beta \mathcal{L}_\beta} \tag{2.4}$$

while the binary particle-ID between two particle hypothesis α and β, can be defined as:

$$\text{PID}(\alpha;\, \beta) = \frac{\mathcal{L}_\alpha}{\mathcal{L}_\alpha + \mathcal{L}_\beta} \tag{2.5}$$

As an example, in Figure 2.15 the kaon-ID efficiency and pion to kaon mis-ID rate are shown for the binary selection criteria $\text{PID}(K;\, \pi) > 0.5$, for different bins of particle momentum. Here the efficiency is defined as the fraction of true $K^\pm$ correctly identified, while the mis-ID is the fraction of $\pi^\pm$ identified as charged kaons. Such quantities are estimated both on Belle II data and MC simulation, looking for a charged K in the decay chain $D^{*+} \to D^0\pi^+$, $D^0 \to K^-\pi^+$ [47].

Similarly, efficiency and mis-ID rates can be estimated for lepton-ID. Several control channels are employed for this purpose, nominally, for measuring the efficiency, the $J/\psi \to \ell^+\ell^-$ decay and the $e^+e^- \to \ell^+\ell^-(\gamma)$, $e^+e^- \to e^+e^-\ell^+\ell^-$ processes. $K_S \to \pi^+\pi^-$ and $\tau^+ \to \pi^+\pi^+\pi^-\overline{\nu}$ are used to extract the pion mis-ID rate while the $D^0 \to K^-\pi^+$ for the kaon mis-ID. The left plot of Figure 2.16 shows the results from these different channels on Belle II data for the muon-ID, in bins of momentum. The $\text{PID}(\mu)$ cut is chosen such that the efficiency is 95% for each bin of the phase space $(p,\, \theta)$. The differences observed among different control channels can be taken as systematic uncertainties on these measurements.

On the right plot of Figure 2.16 the same quantities are shown for electron-ID. However, in this case, instead of the global likelihood as defined in equation (2.3), a multivariate approach is exploited, where all the detector-level variables potentially discriminating for

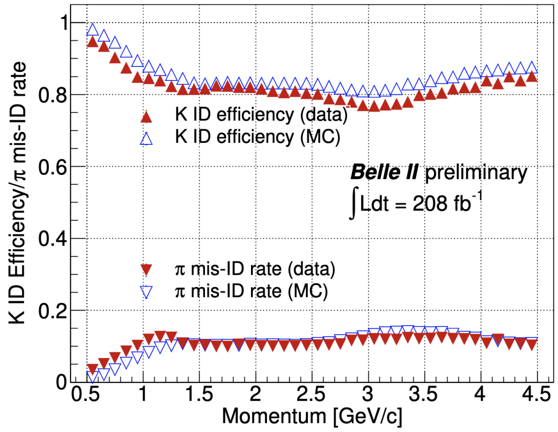


Figure 2.15: Charged K-ID efficiency and π mis-ID rate as a function of momentum for the binary particle-ID cut PID$(K;\pi) > 0.5$, estimated from $D^0 \to K^-\pi^+$ events in the decay chain $D^{*+} \to D^0\pi^+$.

electrons are combined in a boosted decision tree (BDT, see section 3.1.2 for a general overview of this multivariate technique). With respect to the standard likelihood approach, the BDT has an improved performance in the full momentum range, and especially for $p < 0.6\,\mathrm{GeV}/c$, where the fake rate is reduced by up to a factor of 10.

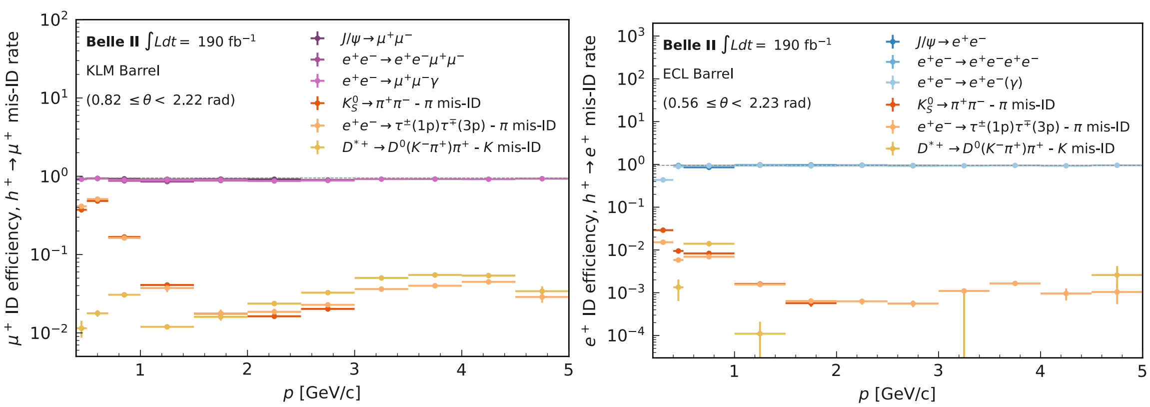


Figure 2.16: Lepton-ID efficiencies and mis-ID rates for muons (left plot) and electrons (right plot), using several control channels. The muon-ID is computed combining the likelihoods from all subdetectors, while for electron-ID a BDT approach is shown.

2.5 Operations and future plans

At Belle II, first electron-positron collisions started in April 2018. On 15th June 2020, SuperKEKB achieved the world's highest instantaneous luminosity for a colliding-beam accelerator, equal to $2.22\times10^{34}\text{cm}^{-2}\text{s}^{-1}$, exceeding the previous record of $2.14\times10^{34}\text{cm}^{-2}\text{s}^{-1}$ detained by the Large Hadron Collider. The current record is set to $4.71\times10^{34}\text{cm}^{-2}\text{s}^{-1}$, obtained on June 2022, just before starting a first long shutdown (LS1) lasting one and a half year.

As shown in the left plot of Figure 2.17, during the first three years of operation a total integrated luminosity of $428\,\text{fb}^{-1}$ was recorded by the Belle detector, a sample size comparable to the full BaBar dataset. Overall, $364\,\text{fb}^{-1}$ of data were recorded at the $\Upsilon(4S)$ resonance, while the remaining $\sim 64\,\text{fb}^{-1}$ was taken either operating at other $\Upsilon(nS)$ peaks or from off-resonance scans.

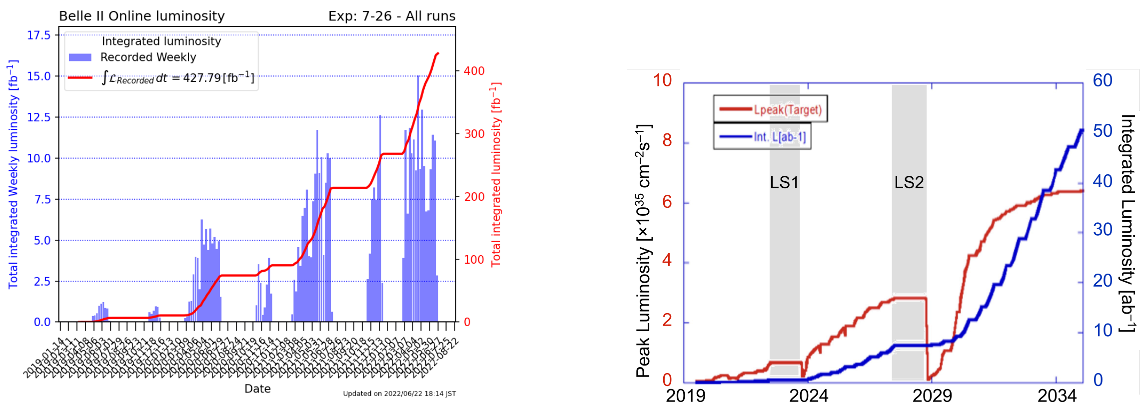


Figure 2.17: Daily integrated luminosity for the first period of Belle II data taking (2019 - 2022) and roadmap for Belle II operations up to 2034. The first long shutdown (LS1) has been completed with a minor upgrade of the detector and interaction region needed to reach $10^{35}\,\text{cm}^{-2}\text{s}^{-1}$ instantaneous luminosity. A second upgrade is scheduled around 2027 (LS2)

Once concluded the detector upgrades scheduled during the LS1, the machine operation resumes in 2024 with a continuous rise of the instantaneous luminosity. The path is defined to reduce the vertical betatron oscillation at the interaction point up to $\beta_y^* = 0.8\,\text{mm}$. As depicted in the right plot of Figure 2.17, a second long shutdown (LS2) is scheduled to start around 2027 to allow the installation of a new superconducting final focusing magnets (QCS), which should permit to squeeze β_y^* up to $0.3\,\text{mm}$ and further improve the luminosity performances reaching the target luminosity of $6\times10^{35}\text{cm}^{-2}\text{s}^{-1}$.

3 Analysis techniques

The search for rare $B^0 \to K^{*0}\tau^+\tau^-$ decay, described in the present work, exploits several advanced analysis techniques, some of them specific for B physics environment. They are briefly outlined in this chapter, starting from section 3.1, with an overview of multivariate classifiers. Here a special focus is put on boosted decision trees (BDT), used in different steps of this analysis, including the signal-to-background discrimination. Section 3.2 is devoted to the introduction of the B-tagging approach, i.e. a reconstruction technique for $e^+e^- \to B\overline{B}$ events, which is crucial for this analysis and it is based on a BDT classifier. Finally, section 3.3 is dedicated to describe the main statistics tools employed for the signal extraction.

3.1 Multivariate analysis

In this section, details on some specific machine-learning techniques are given, focusing on multivariate classification approach, nowadays largely employed in high energy physics. For the $B^0 \to K^{*0}\tau^+\tau^-$ analysis, in particular, such approach is exploited to separate signal events from background (section 4.4), for electron-ID (section 2.4.4) and for B-tag reconstruction (section 3.2).

3.1.1 Binary classification

Let's consider a dataset of n elements $\vec{x}_1, \ldots, \vec{x}_n$, each having m features $\vec{x}_i = (x_i^1, \ldots, x_i^m)$, which can be seen as the values of m observed variables for each element. The multivariate classification consists in assigning to each element $\vec{x}_i$ a label y^k among a class of categories $y^1, \ldots, y^N$, with a certain probability. When just the distinction between two categories is relevant, the multivariate problem reduces to a binary classification ($N = 2$) and the targets can be labeled as $y = 1$ (signal) and $y = 0$ (background).

In the supervised machine-learning approach, a function mapping $\vec{x} \to y$ is built exploiting an independent training dataset with n_{train} elements $\vec{x}_1, \ldots, \vec{x}_{n_{\text{train}}}$ and having human-assigned labels $y_1, \ldots, y_{n_{\text{train}}}$ [48]. The simplest case of this approach would be to use a linear model, i.e. to approximate the mapping function as:

$$\hat{y} = \sum_j \theta_j \, x^j \tag{3.1}$$

where $\vec{\theta} = (\theta_1, \ldots, \theta_m)$ are the parameters of the model and the index j runs over all the m features. In this framework, the output of the linear classifier $\hat{y}$ is a real number, with $0 \leq \hat{y} \leq 1$. Values closer to 1 (0) can be interpreted as more likely to be signal (background). In an actual analysis, the variable $\hat{y}$ should therefore separate signal from background and the signal yield can be directly extracted from its distribution.

The model is fitted finding the best parameters $\vec{\theta}$ which minimize, on the training sample, a so-called objective function:

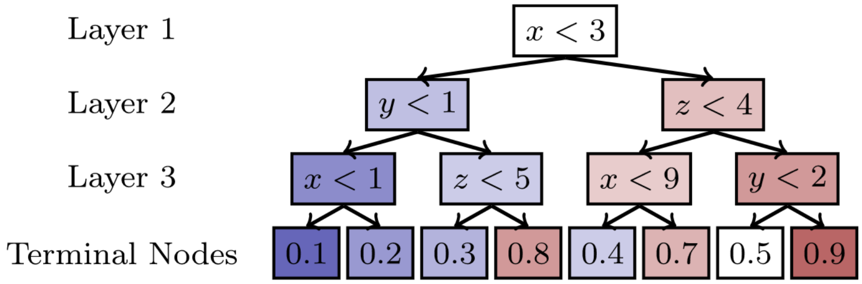


Figure 3.1: Simple example of a decision tree for a binary classifier. Here the features of the dataset are called x, y, z. Decision trees are structured with nodes followed by binary splits, arranged in multiple layers. The terminal node gives the output of the tree. Credits from [51].

$$\mathrm{Obj}(\vec{\theta}) = \mathrm{Loss}(\vec{\theta}) + \Omega(\vec{\theta}) \tag{3.2}$$

The first term, known as loss function, measures the distance between the predicted $\hat{y}$ and the assigned y on the training sample. A popular loss function is the cross-entropy function, defined as:

$$\mathrm{Loss}(\vec{\theta}) = \mathrm{Loss}(y_i, \hat{y}_i) \equiv -\sum_{i=1}^{n_{\mathrm{train}}} \left[y_i \ln \hat{y}_i + (1 - y_i) \ln(1 - \hat{y}_i) \right] \tag{3.3}$$

The term $\Omega(\vec{\theta})$ in equation (3.2) is called regularization function and helps in mitigating the over-training of the model: too complex model $\hat{y}$, which would under-perform on samples independent from the training one, are penalized in the minimization by the regularization term [49].

3.1.2 Boosted Decision Trees (BDT)

Decision trees are a supervised learning algorithm used for multivariate classification [50]. As depicted in Figure 3.1, they consist of a chain of multiple nodes with binary splits, each determined by a cut on the features of the event. The terminal node is the predicted classification of the event. The maximum number of layers allowed for a decision tree is called “depth”.

Despite being based on single cuts, decision trees can be very sensitive to statistical fluctuations of the training sample as their depth increases. For this reason, they are not exploited for multivariate classification in high energy physics [52]. On their place, the boosted decision trees (BDT) are largely used. A BDT combines together a “forest” of many different decision trees, each with few layers, to avoid over-training [53]. The classification of an event is then decided by combining the output of all the decision trees, which are trained via a boosted algorithm. This consists in reweighting the training sample after each iteration, so that mis-classified events receive a larger weight. Explicitly, at the step t of the training, a weight w_t is assigned to the output $f_t(\vec{x})$ of the tree f_t. Therefore, the predicted output is:

$$\hat{y}_i^{(t)} = \sum_{j=1}^{t} w_j \, f^j(\vec{x}_i) \tag{3.4}$$

The tree to be added at the step t is built by minimizing the objective function:

$$\mathrm{Obj}^{(t)} = \sum_{i=1}^{n_{\mathrm{train}}} \left[\mathrm{Loss}(y_i,\, \hat{y}_i^{t-1} + f_t(\vec{x}))\right] + \Omega(f_t) \tag{3.5}$$

where Ω is the regularization function and an example of loss function was defined in equation (3.3).

Among the different existing boosting algorithms, the gradient-boosting is considered to be the most robust [54]. It exploits the gradient descent optimization to identify the tree which is closest to the gradient of the loss function. In the $B^0 \to K^{*0}\tau^+\tau^-$ analysis the XGBoost library is used, which implements a version of gradient-boosting algorithm called extreme gradient boosting [55, 56].

3.2 Reconstruction of tag B mesons

The B mesons produced from $e^+e^- \to \Upsilon(4S)$ are always in either $B^0\overline{B}^0$ or B^+B^- flavor pairs. This peculiarity of a B factory can be exploited when searching for a rare signal B decay, with neutrinos or other missing particles in the final state (e.g. beyond-SM candidates). The principle, known as B-tagging, is to reconstruct the partner B, called "tag", in a well-known decay mode and thus isolating the signal B in the event. Furthermore, if the tag is fully reconstructed in exclusive hadronic decay modes, its four-momentum $(E(B_{\mathrm{tag}}), \vec{p}(B_{\mathrm{tag}}))$ is measurable and the kinematic information of the signal B_{sig} can be inferred from the CM energy constraint:

$$E(B_{\mathrm{sig}}) = \sqrt{s} - E(B_{\mathrm{tag}}) \qquad\qquad \vec{p}(B_{\mathrm{sig}}) = \vec{p}(e^+e^-) - \vec{p}(B_{\mathrm{tag}}) \tag{3.6}$$

where $\sqrt{s} = 10.58\,\mathrm{GeV}$ is the e^+e^- collision CM energy and $\vec{p}(e^+e^-)$ is the total momentum of the two colliding beams. In general, the energy of a particle produced in an asymmetric collider is not known a priori in the detector reference frame, because it depends on the flight direction with respect to the Lorentz boost direction. Hence, the constraint in (3.6) is fundamental to recover this basic information which, combined with the high hermeticity of the detector, makes possible the indirect reconstruction of undetected neutrinos.

In addition, all the activities in the detector (tracks and clusters) which are not associated to the B_{tag} can be assigned to the signal B, provided that they do not originate from beam background, nor from mis-reconstructed B_{tag} decay chains. This additional information

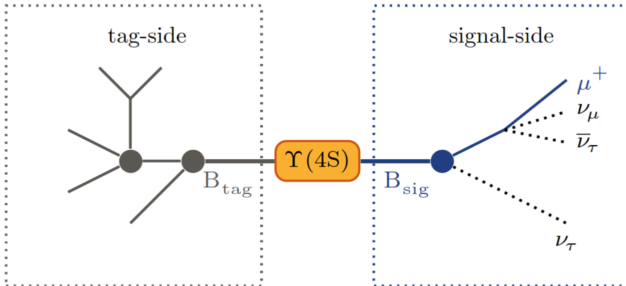


Figure 3.2: Scheme of a full B-tagging for a signal $B \to \tau\nu$ decay. If the tag-side B is fully reconstructed, then the remaining particles should all be assigned to the signal B. Credits to [57].

yields to a high purity signal reconstruction, even for final states with multiple neutrinos (see Figure 3.2).

Depending on the targeted B_{tag} decay modes, two different approaches are typically considered. Beyond the hadronic tag aforementioned, an alternative is represented by the semi-leptonic tag, where semi-leptonic decays $B \to D^{(*)}\ell\nu$, $\ell = e, \mu$, are reconstructed. The main difference stems from higher branching ratio of semi-leptonic B decays with respect to hadronic modes and an easier identification due to the high momentum lepton. This leads to an efficiency up to $\times 5$ larger. However, because of the missing neutrino, the purity of a semi-leptonic tag is much lower and, in particular, the kinematic information is degraded.

The present analysis, as well as all the published $B \to K\tau^+\tau^-$ searches [13, 14], exploits a fully hadronic tag reconstruction. Though more challenging from an experimental point of view, the semi-leptonic tagging technique can in future provide independent results while further reducing the statistical uncertainties on the observables.

In the B factory framework, some kinematic variables can be used to discriminate correctly identified B-tags from mis-reconstructed candidates. In particular, for fully hadronic tagging, the so-called energy difference ΔE and beam-constrained mass M_{bc} are largely exploited. In the CM frame, the e^+e^- collision is symmetric, then $E_{\text{beam}}^{\text{CM}} = \sqrt{s}/2$ and the expected energy of both B mesons is $E_B^{\text{CM}} = E_{\text{beam}}^{\text{CM}}$. Therefore, the energy difference variable of a reconstructed B is defined as:

$$\Delta E \equiv E_B^{\text{CM}} - E_{\text{beam}}^{\text{CM}} \tag{3.7}$$

and it should be peaked at zero for correctly identified tags. Similarly, the invariant mass of the reconstructed B can be redefined substituting its measured energy with the nominal beam energy:

$$M_{\text{bc}} \equiv \sqrt{\left(\frac{E_{\text{beam}}^{\text{CM}}}{c^2}\right)^2 - \left(\frac{|\vec{p}_B^{\,\text{CM}}|}{c}\right)^2} \tag{3.8}$$

and this variable is expected to peak at the B mass $M_B \simeq 5.28\,\text{GeV}/c^2$.

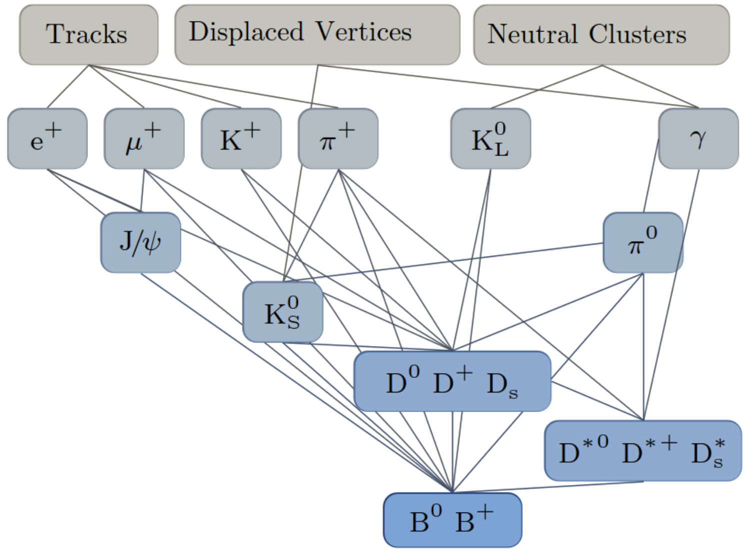


Figure 3.3: Schematic representation of the FEI algorithm. Starting from basic detector-level objects, final states particle are reconstructed and then combined to form intermediate resonances and eventually the final B-tag candidate. Credits to [57].

3.2.1 Hadronic B-tagging algorithms

Several B-tagging algorithms have been developed at the B factories. In the following, an overview of the concept is given, focusing on the most recent Full Event Interpretation (FEI) algorithm, adopted in the Belle II experiment and for the $B^0 \to K^{*0}\tau^+\tau^-$ search [57].

The BaBar experiment implemented and employed a semi-exclusive approach called Semi-Exclusive B Reconstruction (SER) algorithm [16]. This algorithm consists in searching the event for all the candidate charmed hadrons reconstructed in a well-known decay, e.g. $D^+ \to K^-\pi^+\pi^+$ or $D^0 \to K^-\pi^+$. Each of these candidates is then coupled with other pions and kaons in order to form a B meson candidate, which is eventually selected using ΔE and $M_{\rm bc}$ variables as criteria. The efficiency of the SER algorithm, evaluated across different BaBar analysis, is about 0.2% for $B^0\overline{B}^0$ and 0.4% for B^+B^-, with a 30% purity.

Regarding the Belle experiment, the Full Reconstruction (FR) algorithm was developed for hadronic tagging [58]. The strategy is to reconstruct a list of exclusive hadronic B decay modes, covering about 10% of the total branching ratio. To select them, the FR uses a multivariate classifier (nominally, a NeuroBayes neural network [59]) with a hierarchical approach, i.e. it firstly reconstructs single particles from low level objects and then combines them into intermediate resonances up to the B meson candidate. The hadronic tag efficiency achieved by the FR algorithm is 0.18% for $B^0\overline{B}^0$ and 0.28% for B^+B^-, with a corresponding purity of about 10%.

The novel approach adopted for B tagging at the Belle II experiment, called Full Event

Interpretation (FEI), is a substantial improvement of FR, algorithm, following a similar strategy of a hierarchical neural network, sketched in Figure 3.3 [57]. A total of about 10000 combinations of different objects giving a B meson are considered for training the FEI neural network. The output of the multivariate classifier gives the probability that the final B candidate is correctly identified. This variable, here called $\mathcal{P}_{\mathrm{FEI}}$, can be used in addition to ΔE and M_{bc} as a further discriminant to select an higher tag purity. In general, the FEI algorithm outperform the previous FR for any chosen level of tag purity. The maximum tagging efficiency reached by the FEI is 0.46% for $B^0\overline{B}^0$ and 0.76% for B^+B^-.

3.3 Statistical methods

For the $B^0 \to K^{*0}\tau^+\tau^-$ search, the physical observable of interest is the branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$, which is inferred via a binned maximum likelihood fit to the output distribution of a BDT classifier. In the following, after an introduction on binned likelihoods, more details on the fit setup are given.

3.3.1 Maximum likelihood

Given a set of observed data $\vec{x} = x_1, \ldots, x_n$ and a set of model parameters $\vec{\theta} = \theta_1, \ldots, \theta_m$, the likelihood function is defined as [60, 61]:

$$L(\vec{\theta}) \equiv p(\vec{x};\, \vec{\theta}) \tag{3.9}$$

where $p(\vec{x};\, \vec{\theta})$ is the probability density function (PDF) of $\vec{x}$ given $\vec{\theta}$, satisfying $p(\vec{x};\, \vec{\theta}) \geq 0$ and $\int p(\vec{x};\, \vec{\theta}) d\vec{x} = 1$. The Poisson and Gaussian PDFs will be used across this work:

$$\mathrm{Pois}(k;\, \lambda) = e^{-\lambda}\,\frac{\lambda^k}{k!} \tag{3.10}$$

$$\mathrm{Gaus}(x;\, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.11}$$

If the random variables $\vec{x} = x_1, \ldots, x_n$ are all independent and identically distributed, then the joint PDF can be factorized as the product of single PDFs:

$$p(\vec{x};\, \vec{\theta}) = \prod_{i=1}^{n} p(x_i;\, \vec{\theta}) \tag{3.12}$$

Among the parameters $\vec{\theta}$ in (3.9), one (or more) so-called parameter of interest (POI) can be distinguished. That is the parameter which one wants to infer out from the model, for example the rate of the signal process. All the other unknown $\vec{\theta}$ parameters are called nuisance parameters. When measuring branching ratios, a choice of POI is the signal strength μ, which scales the signal rate so that $\mu = 0$ corresponds to the background-only hypothesis, while $\mu = 1$ stays when the signal branching ratio is the expected one.

The likelihood function $L(\mu, \vec{\theta})$ is used to infer information about the parameters μ and $\vec{\theta}$. A point estimation is given by the maximum likelihood (ML) estimator, here denoted as $\hat{\mu}$, $\hat{\vec{\theta}}$, corresponding to the value of the parameter which maximize the likelihood function:

$$\left.\frac{\partial L(\mu, \vec{\theta})}{\partial \mu}\right|_{\mu=\hat{\mu}} = 0 \qquad \left.\frac{\partial L(\mu, \vec{\theta})}{\partial \theta_i}\right|_{\theta_i=\hat{\theta}_i} = 0 \tag{3.13}$$

It is often convenient to consider the function $-2\ln L$, which is a monotonic function of L and it has a minimum for the ML estimator. For a large sample ($n \to \infty$) the likelihood converges to a Gaussian function and $-2\ln L$ has a parabolic profile. The standard deviation of the ML estimator is then the σ of the Gaussian, which can be easily extrapolated noting that:

$$-2\ln L(\hat{\mu} + \sigma(\hat{\mu})) = -2\ln L(\hat{\mu}) + 1 \tag{3.14}$$

If the likelihood differs from a Gaussian, the profile is non-parabolic and the $\hat{\mu}$ uncertainty can be asymmetric. In these cases, using the expression (3.14) does not provide in general a 68% uncertainty coverage.

In many practical situations, instead of a point estimation of μ, an interval estimation may result more useful, that is the case when setting an upper limit on a branching ratio of a signal. In a frequentist framework, the so-called CL$_s$ technique, outlined in the following, is often adopted at the scope [62, 63]. Firstly, the profile likelihood ratio function is defined as:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\vec{\theta}}})}{L(\hat{\mu}, \hat{\vec{\theta}})} \tag{3.15}$$

where $\hat{\hat{\vec{\theta}}}$ is the ML estimator of $\vec{\theta}$ for fixed μ. It is straightforward to show that $0 \leq \lambda(\mu) \leq 1$ and that the value of λ is higher when the given μ is more compatible with the $\hat{\mu}$ estimated from the data. The function λ can then be used to build a test statistics for the parameter μ. In particular, if the signal hypothesis to test is $\mu > 0$, the so-called q_μ test statistics is defined:

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{3.16}$$

The compatibility between the hypothesis of a signal strength μ and the value of q_μ observed from the data can be quantified through the p-value:

$$\mathrm{CL}_{s+b} \equiv p\text{-value}(\mu) = \int_{q_\mu^{\mathrm{obs}}}^{\infty} p(q_\mu;\, \mu) dq_\mu \tag{3.17}$$

where $p(q_\mu;\, \mu)$ is the PDF of q_μ, depending only on the hypothesized μ. Similarly, to test the background-only hypothesis, the following p-value can be evaluated:

$$\mathrm{CL}_{b} \equiv p\text{-value}(\mu = 0) = \int_{q_\mu^{\mathrm{obs}}}^{\infty} p(q_\mu;\, \mu = 0) dq_\mu \tag{3.18}$$

The CL_s is defined as the ratio of these p-values:

$$\mathrm{CL}_s \equiv \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} \tag{3.19}$$

The upper limit at confidence level α is set by determining the value of μ for which the $\mathrm{CL}_s = 1 - \alpha$.

3.3.2 Dedicated tools

When measuring branching ratios, the total expected event yield λ_{tot} is a parameter that needs to be inferred, therefore, it should be included in the model. This is done multiplying the total PDF by a Poisson PDF $\mathrm{Pois}(N_{\mathrm{tot}};\, \lambda_{\mathrm{tot}})$, where N_{tot} is the total number of observed events. Hence, the so-called extended likelihood is defined as follows:

$$L(\lambda_{\mathrm{tot}})^{\mathrm{ext}} = \mathrm{Pois}(N_{\mathrm{tot}};\, \lambda_{\mathrm{tot}})\, L \tag{3.20}$$

The present work exploits the HistFactory tool for statistical models, which implements likelihoods suited for high energy physics searches [64]. For the $B^0 \to K^{*0}\tau^+\tau^-$ analysis, a binned extended likelihood is adopted to fit the signal region. The observed events in each bin $b = 1, \ldots, n$, where n is the number of bins of the histogram, are counted as $\vec{k} = k_1, \ldots, k_n$. They are independently distributed as Poisson PDFs with rates given by the expected yield in each bin: $\vec{\lambda} = \lambda_1, \ldots, \lambda_n$. In the general case of multiple samples, i.e. one signal component and one or more background components, the expected yields for each bin λ_b can be decomposed as:

$$\lambda_b = \sum_{s \in \mathrm{samples}} \lambda_{sb} \qquad \text{where} \qquad \sum_{s \in \mathrm{samples}} \sum_{b \in \mathrm{bins}} \lambda_{sb} = \lambda_{\mathrm{tot}} \tag{3.21}$$

and the index s runs over all the components of the model (signal and backgrounds). Similarly, multiple independent channels can be included in the same model, for instance different decay channels of the same particle (the τ decay modes for $B^0 \to K^{*0}\tau^+\tau^-$). This can be implemented through an additional index c running over all the channels considered. After some manipulation, the likelihood can be written in the form:

$$L(\vec{\lambda}) = \prod_{c\in\text{channels}} \prod_{b\in\text{bins}} \text{Pois}(k_{cb};\, \lambda_{cb}) \qquad \text{where} \qquad \lambda_{cb} = \sum_{s\in\text{samples}} \lambda_{scb} \tag{3.22}$$

At this point, the model can be further enriched including a set of constrained parameters $\vec{\chi}$ which can modify the nominal yields $\vec{\lambda}^0$. Each of these parameters enters the likelihood via a constraint PDF term $c_\chi(a_\chi;\chi)$ containing the systematic uncertainties. The variables a_χ are called auxiliary data, as they are determined from auxiliary measurements, typically from control channels or external inputs. A simple example of a constraint term would be a Gaussian PDF $\text{Gaus}(\alpha; 1, \sigma_{\text{sys}}(\lambda))$ multiplying the nominal rate λ^0 by a factor α, with a standard deviation $\sigma_{\text{sys}}(\lambda)$ corresponding to the systematic uncertainty.

Beyond the constrained parameters $\vec{\chi}$, the rates may depend also on a set of free parameters $\vec{\theta}$, thus $\vec{\lambda} = \vec{\lambda}(\vec{\theta},\, \vec{\chi})$. The POI μ (signal strength) is an example of unconstrained parameter.

The likelihood of equation (3.22) then becomes:

$$L(\vec{\theta},\, \vec{\chi}) = \prod_{c\in\text{channels}} \prod_{b\in\text{bins}} \text{Pois}(k_{cb};\, \lambda_{cb}(\vec{\theta},\, \vec{\chi})) \prod_{\chi\in\vec{\chi}} c_\chi(a_\chi;\chi) \tag{3.23}$$

In the HistFactory framework, the expected yields for each bin and for each channel are explicitly depending on the constraint parameters as follows:

$$\lambda_{cb}(\vec{\theta},\, \vec{\chi}) = \sum_{s\in\text{samples}} \left(\prod_{\kappa\in\vec{\kappa}} \kappa_{scb}(\vec{\theta},\, \vec{\chi}) \right) \left(\lambda^0_{scb} + \sum_{\Delta\in\vec{\Delta}} \Delta_{scb}(\vec{\theta},\, \vec{\chi}) \right) \tag{3.24}$$

The terms κ_{scb} and Δ_{scb} are, respectively, multiplicative and additive modifiers of the nominal yields λ^0_{scb}. They are functions of the model parameters $\vec{\theta}$, $\vec{\chi}$ (typically just one parameter). Some examples of HistFactory standard modifiers, which are used in the $B^0 \to K^{*0}\tau^+\tau^-$ analysis, are listed in table 3.1.

Table 3.1: Examples of histogram modifiers, as implemented in the HistFactory framework. The indexes c, s and b run over channels, sample components, and bins, respectively. a_χ are the auxiliary data for the constraint χ. δ_{bs} is the MC statistical uncertainty for the bin b of sample s. The so-called correlated shape term scales with the same factor α across different bins.

Modifier	Histogram variation	Constraint term $c_\chi(a_\chi;\, \chi)$
Correlated shape	$\Delta_{scb}(\alpha)$	$\text{Gaus}(a_\chi = 0;\, \alpha,\, \sigma = 1)$
Normalization uncertainty	$\kappa_{scb}(\alpha)$	$\text{Gaus}(a_\chi = 0;\, \alpha,\, \sigma = 1)$
MC Stat. Uncertainty	$\kappa_{scb}(\gamma_b)$	$\prod_b \text{Gaus}(a_{\gamma_b} = 1;\, \gamma_b,\, \sqrt{\sum_s \delta^2_{bs}})$
Luminosity	$\kappa_{scb}(\eta) = \eta$	$\text{Gaus}(l = \eta_0;\, \eta,\, \sigma_\eta)$

The fit model for the $B^0 \to K^{*0}\tau^+\tau^-$ analysis, detailed in section 4.6, is implemented using Pyhf, a python implementation of the HistFactory PDF template which provides a collection of built in functions for statistical analysis [65, 66]. Pyhf is used in combination with the Cabinetry python package [67], which makes possible to handle easily the modifiers and includes additional methods for histogram visualization.

4 Search for rare $B^0 \to K^{*0}\tau^+\tau^-$ decay

This chapter describes in details the analysis for the search of $B^0 \to K^{*0}\tau^+\tau^-$ decay, using 362.2 fb^{-1} of Belle II data collected at the $\Upsilon(4S)$ resonance during the first period of data taking, between 2019 and 2022.

After a brief introduction of the analysis strategy in section 4.1, section 4.2 outlines the first stage of the event selection. Section 4.3 lists the corrections applied to simulated events for improving the agreement with data, while section 4.4 describes the setup of the multivariate classifier, which is validated on several data control channels, as detailed in section 4.5. The systematic uncertainties and the fit framework for signal extraction are finally reported in the sections 4.6.1 and 4.6, respectively.

4.1 Analysis strategy

The decay $B^0 \to K^{*0}\tau^+\tau^-$ includes a large combination of possible final states, determined by the different decay modes of the K^{*0} meson and of the two τ leptons.

The K^{*0} is a wide resonance state with strangeness flavor, having a mass peak $M(K^{*0}) = (895.6 \pm 0.2)\,\mathrm{MeV}$ and a decay width $\Gamma(K^{*0}) = (47.3 \pm 0.5)\,\mathrm{MeV}$ (all values are from PDG [6]). It decays almost 100% of the times as $K^{*0} \to K\pi$, to be split in 2/3 as charged $K^{*0} \to K^+\pi^-$ and 1/3 as neutral $K^{*0} \to K^0\pi^0$, where the K^0 is either a K_S or a K_L[1]. In this analysis, only the charged decay mode of the K^{*0} is exclusively reconstructed, being about 66% of the total signal events and having the cleanest signature. In contrast, the neutral mode is experimentally challenging since it requires an efficient π^0 reconstruction in the ECL, in combination with a $K_S \to \pi\pi$ displaced vertex or a deposit in the KLM detector compatible with a K_L. Considering the lower branching fraction ($\simeq$33%), the actual gain coming from the $K^{*0} \to K^0\pi^0$ mode can be considered negligible.

The τ lepton can decay either to hadrons, with one neutrino, or to a lighter lepton and two neutrinos. Therefore, for a signal with a pair of τ, many different combination of final states are possible, and the number of neutrinos in the final state can vary between two and four. Since neutrinos escape the detector, the invariant mass of the signal B^0 cannot be fully reconstructed. Furthermore, each τ from a $B^0 \to K^{*0}\tau^+\tau^-$ signal event at Belle II is expected to have a decay vertex displaced by a decay length of about $\mathcal{O}(100\,\mu\mathrm{m})$. Hence, for this analysis, the decay vertices cannot be easily constrained as well. This means that there is not a specific signature targeting all the signal events, except for the presence of undetected neutrinos, leading to a missing energy with respect to the initial state.

As described in section 3.2, in a B factory the B-tagging can be exploited as a powerful tool for isolating the signal B and recovering its kinematic, thus allowing to estimate missing energy and momentum. The present analysis takes advantage of the FEI algorithm to reconstruct a hadronic B-tag [57]. Previous $B \to K\tau^+\tau^-$ searches were based on hadronic tag techniques as well, however an improved efficiency is expected from the new algorithm developed for Belle II (see section 3.2.1) [13, 14].

[1]Note that charge-conjugate modes are always implied throughout this chapter.

Table 4.1: Branching ratios of the τ-lepton decay modes targeted in this analyis [6]. The $\tau \to \pi\pi^0\nu$ decay includes the $\tau \to \rho(770)\nu$ mode and the non resonant one, with an estimated branching ratio $\mathrm{BR}(\tau \to \pi\pi^0(\mathrm{non}\,\rho)\nu) = (3.0 \pm 3.2) \times 10^{-3}$.

Signal category	$BR(\%)$
$\tau \to \mu\nu\nu$	17.39 ± 0.04
$\tau \to e\nu\nu$	17.82 ± 0.04
$\tau \to \pi\nu$	10.82 ± 0.05
$\tau \to \pi\pi^0\nu$	25.49 ± 0.09

Aside from neutrinos, a deficit in the total energy of the event could also originate from particles not reconstructed due to detector inefficiencies. In these cases, background events can mimic the signal. An important variable to deal with such background is the energy of the ECL clusters which was not associated to any particle, in this analysis denoted as E_{ECL}^{extra}. Such variable should be exactly zero for correctly reconstructed signal events, and it reaches energies up to few GeV for background events.

The main 1-prong (i.e. with one charged track) τ decay modes, listed with their respective branching ratios in table 4.1, are exclusively reconstructed in this analysis. Considering all the possible $\tau^+\tau^-$ combinations of these decay modes, there are 10 different final states, accounting for 51.2% of all the total $\tau^+\tau^-$ decays. The inclusion of the $\tau \to \rho\nu$ decay mode for the signal side in the present analysis is an important difference with respect to the previous Belle strategy, covering only 21.2% of all the $\tau^+\tau^-$ final states.

Figure 4.1 shows a sketch of the decay modes reconstructed to identify the $B^0 \to K^{*0}\tau^+\tau^-$ signal. An additional factor to be considered for a neutral B meson is the $B^0 - \overline{B}^0$ flavor oscillation [68]. The flavor of the signal B in the decay can be inferred from the charge of the kaon in the reconstructed $K^{*0} \to K^+\pi^-$. Nominally, a K^+ identifies a B^0 while a K^- comes from a $\overline{B}^0$. In a tagged approach, the tag candidate flavor is reconstructed as well, and since the two B mesons from a $\Upsilon(4S)$ decay are entangled states, the signal B must have an opposite flavor until one of the two decays [69]. The flavor oscillation probability is then determined by the time difference between the B_{tag} decay and the $B \to K^*\tau^+\tau^-$ decay, which is not explicitly measured. In this case, the time-integrated $B^0 - \overline{B}^0$ mixing probability χ_d can be taken as a reference. Its world average value is $\chi_d = 0.1858 \pm 0.0011$ [6], implying that $1 - \chi_d \simeq 80\%$ of the reconstructed B meson pairs have actually opposite flavor ($B^0\overline{B}^0$), while about 20% are reconstructed in the same flavor states (B^0B^0, $\overline{B}^0\overline{B}^0$). In this analysis, the latter case (B^0B^0, $\overline{B}^0\overline{B}^0$) is used as a control sample, having approximately four times lower expected signal efficiency but a relatively higher contamination from mis-reconstructed events.

The separation between signal and background events is based on a multivariate approach: a binary classifier is trained on simulated events to distinguish the signal against all the backgrounds. The classifier combines features related to signal kinematics, missing energy

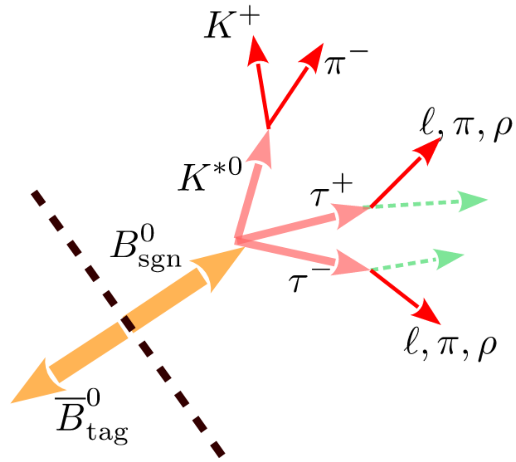


Figure 4.1: Schematic view of the signal $B^0 \to K^{*0}\tau^+\tau^-$ decay chain as reconstructed in this analysis. The K^{*0} is reconstructed in its charged decay mode while the two $\tau^\pm$ are considered in their main 1-prong modes, with a neutrino multiplicity varying between two and four (the dashed arrows stand for two neutrinos in a leptonic τ and one neutrino in a hadronic τ). The partner B^0 is tagged in a hadronic decay by the FEI algorithm [57].

and momentum due to neutrinos escaping the detector, and signatures targeting specific background topologies.

Prior to the training, a pre-selection is performed to increase the purity of the reconstructed $B_{\rm tag}$ and to reduce the background rate. Furthermore, several control channels are exploited to correct mis-modeling between data and simulation and to validate both the input features and the output of the binary classifier, while still keeping blind the data used to extract the signal yield.

The $B \to K^*\tau^+\tau^-$ branching ratio is finally determined through a binned maximum likelihood fit, performed directly on the output of the classifier, using the simulated distributions as templates and including all the systematic uncertainties, previously estimated from control channels.

4.2 Event selection and reconstruction

4.2.1 Dataset and simulated samples

Data samples

This analysis profits of the full dataset recorded by the Belle II detector in the period 2019 - 2022 at the center of mass energy of $\Upsilon(4S)$ resonance (on-resonance data), corresponding to 362.2 fb^{-1}, plus and additional 42.3 fb^{-1} sample collected 60 MeV below the resonance (off-resonance data). Being below the $b\bar{b}$ production threshold, the off-resonance sample does not contain any B meson, while all the other processes (e.g. $e^+e^- \to q\bar{q}$, $q = u, d, s, c$, $e^+e^- \to \ell^+\ell^- \; (\gamma)$) are expected to be the same as in the on-resonance sample, though with a slightly lower total CM energy $\delta E = 60\,\mathrm{MeV}/10.58\,\mathrm{GeV} \simeq 0.6\%$. This results also in a smaller cross-section for the on-resonance case by a factor $(1 - \delta E)^2 \simeq 0.989$, which can be neglected for the present analysis, while the kinematic variables M_{bc} and ΔE, exploited in the selection, are adjusted for the off-resonance data.

All data considered were selected by the HLT trigger (see section 2.3.6), whose efficiency for $B\overline{B}$ events is close to 100%.

Monte Carlo samples

The simulated Monte Carlo (MC) sample is generated to be a factor from three to about eight times larger than the available data sample. Physics processes of interest are generated starting from e^+e^- collisions, with beam background superimposed via a run-independent approach, thus neglecting variations of the machine background parameters during the data-taking period, as described in section 2.4.1.

The MC samples exploited throughout this analysis are listed in table 4.2, together with their corresponding integrated luminosity and the actual number of generated events. The equivalent of 3 ab^{-1} of generic $B\overline{B}$ background ($B^0\overline{B}^0$ and B^+B^-) is generated via `EvtGen`. However, a subset corresponding to 200 fb^{-1} is used to train the FEI classifier, while the remaining 2.8 ab^{-1} is actually exploited in the analysis. The simulated background from continuum production of light quark pairs ($q\overline{q}$) corresponds to 1 ab^{-1} of integrated luminosity and it is simulated using `PYTHIA`. No simulated MC sample is included for $e^+e^- \to \tau^+\tau^-$ and low multiplicity physics processes, since such events would be already rejected by the FEI tag-reconstruction algorithm.

Regarding the signal MC sample, 50 millions of events are generated with `EvtGen`, where one B meson decays as $B^0 \to K^{*0}\tau^+\tau^-$ with $B^0 \to K^{*0}$ SM form factors computed through light-cone sum rules [70]. Both τ leptons decay generically with TAUOLA, while the K^{*0} is forced to decay as $K^+\pi^-$, to increase the efficiency on the generated events. To take this into account, signal events will be then multiplied by a weight $\mathcal{B}(K^{*0} \to K^+\pi^-) = 0.66$. The partner B in the event decays generically and both B mesons undergo flavor oscillation.

Table 4.2: Simulated MC datasets used in the $B^0 \to K^{*0}\tau^+\tau^-$ analysis, with the number of total generated events and the equivalent integrated luminosity. For all the samples the beam background is superimposed in a run-independent way.

Process	Generated events $\times 10^6$	Integrated luminosity [fb^{-1}]
$e^-e^- \to \Upsilon(4S) \to B^0\overline{B}^0$	1607	2800
$e^-e^- \to \Upsilon(4S) \to B^+B^-$	1701	2800
$e^-e^- \to u\overline{u}$	1605	1000
$e^-e^- \to d\overline{d}$	401	1000
$e^-e^- \to s\overline{s}$	383	1000
$e^-e^- \to c\overline{c}$	1329	1000
$B^0 \to K^{*0}\tau^+\tau^-$	50	

4.2.2 Event reconstruction

A typical reconstructed event can be ideally split into three parts, corresponding to three separate steps of the event-reconstruction scheme:

- The hadronic B^0_{tag} candidate, reconstructed via the FEI algorithm.
- The signal $B^0 \to K^{*0}\tau^+\tau^-$ candidate, reconstructed with all the detector-level objects not used for the B^0_{tag} candidate.
- The so-called rest of the event (ROE), i.e. all the detector-level objects not associated neither to the B^0_{tag} candidate, nor to the B^0_{sig} candidate.

Tag-side reconstruction

The FEI algorithm is applied on all data and simulated samples to tag a B meson in a hadronic decay [57]. For such decays, the B_{tag} energy and momentum can be estimated as the sum of all the corresponding final states, as highlighted in section 3.2. For the most recent version of the FEI algorithm, used in the present analysis, the estimated efficiency of the hadronic tag on a generic $B^0\overline{B}^0$ at Belle II is about 0.4%.

Before running the FEI tag reconstruction, to save computing time and reject non-$B\overline{B}$ low multiplicity processes, the following skimming selection is applied on all the samples:

1. Each event should have at least 3 tracks with $|z_0| < 2$ cm, $|d_0| < 0.5$ cm and $p_T > 0.1\,\mathrm{GeV}/c$. Here $|z_0|$ and $|d_0|$ are the distances of the track POCA (point of closest approach) with respect to the origin, in the longitudinal z direction and in the r, ϕ transverse plane, respectively (see section 2.4.2).
2. Each event should have at least 3 ECL clusters with $E > 0.1$ GeV and within the acceptance region of the CDC ($17^\circ < \theta < 150^\circ$).

3. The visible energy of the event should satisfy $E_{\text{vis}} > 4$ GeV, in order to reject di-photon and low multiplicity processes. E_{vis} is computed in the center of mass (CM) frame, summing up the energy of all tracks and neutral clusters satisfying requirements 1. and 2. above.

To increase the purity of the sample, additional requirements are imposed to the B_{tag} candidate reconstructed by the FEI. The deviation from the nominal energy $\Delta E = E^{\text{CM}}_{B_{\text{tag}}} - E^{\text{CM}}_{\text{beam}}$ and the beam-constrained mass $M_{\text{bc}} = \sqrt{(E^{\text{CM}}_{\text{beam}}/c^2)^2 - (|\vec{p}^{\,\text{CM}}_{B_{\text{tag}}}|/c)^2}$, defined in section 3.2, can be used as kinematic constraints for evaluating the quality of the B tags. In addition, the output of the FEI classifier gives an independent probability that the B_{tag} candidate is correctly reconstructed, hereafter called "signal probability" $\mathcal{P}_{FEI}$. The following cuts are then requested on these three variables:

$$M_{\text{bc}} > 5.24 \text{ GeV}/c^2 \tag{4.1}$$

$$-0.2 < \Delta E < 0.2 \text{ GeV} \tag{4.2}$$

$$\mathcal{P}_{\text{FEI}} > 0.001 \tag{4.3}$$

Only the B_{tag} candidates surviving these selection are considered in the analysis. For the off-resonance data, before applying the above requirements, the energy difference and beam-constrained mass are redefined to correct for the lower overall energy and momentum of the reconstructed B_{tag}:

$$\Delta E^{\text{off}} \equiv \frac{E^{\text{CM, on}}_{\text{beam}}}{E^{\text{CM, off}}_{\text{beam}}} E^{\text{CM}}_{B_{\text{tag}}} - E^{\text{CM}}_{\text{beam}} \tag{4.4}$$

$$M^{\text{off}}_{\text{bc}} \equiv \sqrt{\left(\frac{E^{\text{CM, on}}_{\text{beam}}}{c^2}\right)^2 - \left(\frac{E^{\text{CM, on}}_{\text{beam}}}{E^{\text{CM, off}}_{\text{beam}}} \frac{|\vec{p}^{\,\text{CM}}_{B_{\text{tag}}}|}{c}\right)^2} \tag{4.5}$$

where $E^{\text{CM, on}}_{\text{beam}} = \sqrt{s}/2$ is the CM energy of each beam at the $\Upsilon(4S)$ resonance, while $E^{\text{CM, off}}_{\text{beam}} \simeq \sqrt{s}/2 - 60\,\text{MeV}$ is the CM beam energy for the off-resonance data-taking.

At this stage, 3.4% of the $B^0 \to K^{*0}\tau^+\tau^-$ signal events have at least one reconstructed B_{tag} after the FEI reconstruction and skimming selection, with an average multiplicity of 1.9 B_{tag} candidates per event. The fraction of events with a correctly reconstructed B_{tag} candidate (defined as a candidate with all the particles in the decay chain matched to MC simulation) is 16.9%. For each event, only the B_{tag} candidate with highest value of $\mathcal{P}_{FEI}$ is taken among the ones passing the FEI skim, so that the tag-side is unambiguously defined. For the simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal, the efficacy of this single-tag selection is evaluated to be 86.7%, and after its application 14.2% of the selected B_{tag} are correctly reconstructed (all the particles in the decay chain are truth-matched).

Signal-side reconstruction

The detector objects which are not associated with the B_{tag} candidate, are used to reconstruct a $B^0 \to K^{*0}\tau^+\tau^-$ candidate. The minimal requirements for all tracks on the signal side are:

- Originating closely to the interaction point: $|d_0| < 2\,\text{cm}$ and $|z_0| < 4\,\text{cm}$, to exclude beam background tracks.
- Polar angle in the CDC acceptance: $17^\circ < \theta < 150^\circ$.
- At least 20 hits in the CDC, to have a reliable particle-ID from dE/dx measurement.
- Transverse momentum $p_T > 100$ MeV$/c$, to reject low-momentum beam background tracks.
- Energy associated to the track $E < 5.5$ GeV, to remove electrons from Bhabha scattering.

Only events with exactly four tracks in the signal side, passing the selection criteria mentioned above, are retained as possible signal candidates.

Depending on the binary π, K charged particle-ID, defined in section 2.4.4, tracks are classified as pions if $\text{PID}(\pi;\, K) > 0.6$ and as kaons if $\text{PID}(K;\, \pi) > 0.6$. A loose selection is here preferred in order to keep a high signal efficiency and separate background afterwards with a multivariate approach.

Muons are identified among tracks having $\text{PID}(\mu) > 0.5$, using the likelihood-based definition of equation (2.3), while tracks identified as electrons and positrons must satisfy $\text{PID}(e) > 0.5$, using instead the BDT-based PID, which outperforms the corresponding likelihood electron-ID in the full phase-space region.

Neutral clusters used to identify photons coming from π^0 decays, are selected via the following criteria:

- Number of (weighted) crystals in the ECL cluster to be $N_{\text{crystals}} > 1.5$. Where $N_{\text{crystals}} = \sum_{i \in \text{cluster}} w_i$ and the weights w_i, defined in equation (2.2), account of overlapping clusters.
- Cluster polar angle in the CDC acceptance ($17^\circ < \theta < 150^\circ$), to reject clusters originating from missing tracks.
- Cluster energy threshold depending on the ECL region, requiring tighter thresholds on the endcaps, where the background level is harsher:
  - Forward ECL: $E > 80$ MeV
  - Barrel ECL: $E > 30$ MeV

– Backward ECL: $E > 60$ MeV

and then $\pi^0 \to \gamma\gamma$ are reconstructed from a pair of photons satisfying:

- $120\,\mathrm{MeV}/c^2 < M(\gamma\gamma) < 145\,\mathrm{MeV}/c^2$.
- Azimuthal angle between the two photons $-1.5\,\mathrm{rad} < \varphi < 1.5\,\mathrm{rad}$.
- Angle between the two photons < 1.4 rad.

Such reconstruction criteria for π^0 are chosen in order to give a good agreement between data and simulation, while retaining a reasonable (about 30%) π^0 reconstruction efficiency.

The π^0's are used to identify charged ρ meson candidates decaying as $\rho^\pm \to \pi^\pm\pi^0$. The invariant mass of a ρ candidate must be in the range $0.65 < M(\rho) < 0.9$ GeV/c^2.

Pairs of charged pions and kaons having opposite charges and invariant mass $M(K\pi) \in$ (0.6, 1.3) GeV/c^2, are combined as a $K^{*0} \to K^+\pi^-$ candidate, or its flavor conjugate. A vertex fit is performed on the two tracks and the K^{*0} candidates whose p-value of the vertex fit is lower than 10^{-3} are rejected [71].

For each reconstructed K^{*0} on the signal-side, a B^0_{sig} candidate is formed by adding the daughters of two opposite charge 1-prong taus in decays $\tau \to \pi\nu$, $\tau \to \rho\nu$, $\tau \to \ell\nu\nu$, with $\ell = e$, μ. There are in total 10 possible combinations of $\tau^+\tau^-$ decay modes. The 1-prong daughters of the two τ leptons are denoted henceforth as t_1 and t_2, where t_1 has the same charge sign of the $K^\pm$ from K^{*0}, while t_2 has the opposite sign. The invariant mass of K^{*0} and τ daughters should be $M(K^{*0}t_1t_2) < 6.0$ GeV/c^2, to reject background from fully hadronic B decays.

Multiple K^{*0} candidates and tau-pair final states could be reconstructed on the same signal side, therefore multiple B^0_{sig} candidates are possible for a single event. At this stage, all the B^0_{sig} candidates are kept and combined with the unique B^0_{tag}, to form one $\Upsilon(4S)$ candidate for each B^0_{sig}. Pairs of $B_{\mathrm{tag}}B_{\mathrm{sig}}$ mesons having the same flavor (B^0B^0 or $\overline{B}^0\overline{B}^0$) are allowed in the analysis, since they can be due to flavor mixing. However, a distinction is done between the "opposite-flavor" candidates ($\Upsilon(4S) \to B^0\overline{B}^0$), which are used for signal extraction, and the "same-flavor" candidates, i.e. $\Upsilon(4S) \to B^0B^0$ and $\Upsilon(4S) \to \overline{B}^0\overline{B}^0$, having the same b flavor, which are instead exploited as a control sample for the analysis validation.

Rest of the event

For each $\Upsilon(4S)$ candidate, the rest of the event (ROE) is defined as all the detector-level objects not used neither for B_{tag} nor for B_{sig} reconstruction. In the ROE, charged tracks are defined via the following selection:

- Track from the interaction point region: $|d_0| < 2\,\mathrm{cm}$ and $|z_0| < 4\,\mathrm{cm}$.
- Polar angle in the CDC acceptance: $17^\circ < \theta < 150^\circ$.

- $p_T > 100\,\mathrm{MeV}/c$.

Exactly zero charged tracks passing such criteria are require to be in the ROE: if one or more track is present, the corresponding $\Upsilon(4S)$ candidate is rejected. The rate of fake tracks was evaluated on $e^+e^- \to \tau^+\tau^-$ events, with a tag and probe method, to be $2.0\% \pm 0.3\%$ on simulation and $2.2\% \pm 0.2\%$ on data, therefore the reduction of signal efficiency due to fake tracks can be neglected.

At this reconstruction level, the average $\Upsilon(4S)$ candidate multiplicity is 5.6 for signal events, 7.0 for $B\overline{B}$ background and 5.2 candidates for $q\overline{q}$ background events ($q = u, d, s, c$).

The extra ECL energy (E_{ECL}^{extra}) is defined as the sum of the energy of all clusters in the ROE passing the following selection:

- Number of weighted crystals in the ECL cluster: $N_{\text{crystals}} > 1.5$.
- Cluster polar angle in the CDC acceptance ($17^\circ < \theta < 150^\circ$).
- Energy cuts depending on the ECL region:
  - Forward ECL: $E > 100$ MeV
  - Barrel ECL: $E > 60$ MeV
  - Backward ECL: $E > 150$ MeV
- Minimum distance between the ECL cluster and any charged track > 20 cm.

The minimum distance between cluster and tracks avoids to consider background clusters due to secondary particles, generated from nuclear interactions of hadronic tracks with the material of the calorimeter. Furthermore, only the "neutral" component is included in the E_{ECL}^{extra} evaluation, i.e. all the clusters which can be associated to a charged track are excluded in the E_{ECL}^{extra} calculation.

No extra π^0 veto is applied, in order to keep, at this stage, signal events with additional π^0 from τ decays.

4.2.3 Event pre-selection

The number of events passing the reconstruction requirements described in the previous sections are reported in the second column of table 4.3. In this analysis, the signal yield is counted considering any of the generated $B^0 \to K^{*0}\tau^+\tau^-$ event passing the selection, even those which are not correctly identified (for example a mis-reconstructed τ decay mode). The signal efficiency $\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$ is defined as the ratio between the signal yield and the number of generated $B^0 \to K^{*0}\tau^+\tau^-$ events. Since in the generate signal sample, the K^{*0} is forced to decay exclusively into $K^+\pi^-$, with $\mathcal{B}(K^{*0} \to K^+\pi^-) = 0.66$, then the signal yield should be adjusted by applying a weight of 0.66 to all the signal events.

A loose pre-selection is applied in order to clean up background events before checking the agreement between data and simulation. As described in the following, a set of four

Table 4.3: Number of MC events after reconstruction (second column) and pre-selection (third column) for the different background components and signal efficiency, as estimated from simulation. The background yields are weighted assuming an integrated luminosity of 362.2 fb^{-1}.

Process	After reconstruction	After pre-selection
$e^+e^- \to u\overline{u}$	984×10^3	23.6×10^3
$e^+e^- \to d\overline{d}$	230×10^3	5.7×10^3
$e^+e^- \to s\overline{s}$	229×10^3	5.9×10^3
$e^+e^- \to c\overline{c}$	1416×10^3	39.6×10^3
$e^+e^- \to B^0\overline{B}^0$	398×10^3	41.0×10^3
$e^+e^- \to B^+B^-$	374×10^3	14.3×10^3
Total Background	3.631×10^6	130.1×10^3
$\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$	$(1.76 \pm 0.01) \times 10^{-3}$	$(5.76 \pm 0.05) \times 10^{-4}$

variables is exploited for the pre-selection. Their MC distributions, for both signal and all the generated background processes, are shown in Figure 4.2, and the cut values are defined as follows:

- To suppress continuum and $B\overline{B}$ combinatorial background, a cleaning on the reconstructed B_{tag} is performed requiring $M_{\text{bc}} > 5.27$ GeV$/c^2$ and $\mathcal{P}_{\text{FEI}} > 0.01$ in order to increase the tag quality.

- To avoid events with particles escaping the tracking detectors, the polar angle of the missing momentum $\theta(\vec{p}_{\text{miss}})$ is requested to stay inside the CDC angular acceptance: $17^\circ < \theta(\vec{p}_{\text{miss}}) < 150^\circ$. The missing momentum of the event is defined as $\vec{p}_{\text{miss}} = \vec{p}_{\text{init}} - \sum_i \vec{p}_i$ where $\vec{p}_{\text{init}}$ is the initial total momentum of the colliding beams and the sum of the $\vec{p}_i$ is done over all the particles in the event.

- To further clean the signal side, the reconstructed $K^{*0} \to K^+\pi^-$ invariant mass is restricted in a window $800 < M(K\pi) < 990\,\text{MeV}/c^2$, corresponding to about two decay widths around the mass peak of the K^{*0} resonance [6].

Furthermore, the flavor of the reconstructed B^0_{sig}, determined by the charge of the K from K^{*0}, is required to be opposite to the flavor of the partner B^0_{tag} candidate. This allows to define the "same-flavor" control sample, with reconstructed B^0B^0 and $\overline{B}^0\overline{B}^0$ pairs. The opposite flavor request also reduces mis-reconstructed events and combinatorial background, excluding about 13% of the $B^0\overline{B}^0$-like signal events which are wrongly reconstructed as a same-flavor, mainly because of mis-tagging.

The yields of background events and signal efficiency after applying the pre-selection is reported in the third column of table 4.3. After the pre-selection, the total background yield is reduced by about a factor $\times 30$, with $B^0\overline{B}^0$ being the most abundant component.

The dominant $e^+e^- \to q\bar{q}$ processes are from up-type quarks (i.e. $u\bar{u}$ and $c\bar{c}$, having a higher cross section with respect to $d\bar{d}$ and $s\bar{s}$), and in particular from $c\bar{c}$ events, due to their higher multiplicity. The signal efficiency is reduced of about a factor $\times 3$, mainly due to the tag-side purity requirements.

4.2.4 Signal categories

After the pre-selection, the main analysis steps are optimized differentiating the events into four signal categories, in order to exploit the different properties of the final states on the signal side. The signal category classification depends on the possible τ decay mode combinations. Nominally they are:

1. $\ell\ell$: both τ decaying as $\tau \to \ell\nu\nu$, where $\ell = e$, μ
2. $\pi\ell$: one τ decaying as $\tau \to \pi\nu$ and the other decaying leptonically ($\tau \to \ell\nu\nu$)
3. $\pi\pi$: both τ decaying as $\tau \to \pi\nu$
4. ρ: one τ decaying as $\tau \to \rho\nu$, the other one to any of the reconstructed modes ($\ell\nu\nu$, $\pi\nu$, $\rho\nu$)

The first three categories are distinguished by the number of neutrinos in the final state. The ρ category can have one or two reconstructed π^0, depending on the number of reconstructed $\tau \to \rho\nu$ decays, while the number of neutrinos is not fixed. These four categories have different final states which may result in very different kinematic distributions. Therefore, a separated optimization is performed for each of them.

4.2.5 Single candidate selection

For each event, multiple $\Upsilon(4S)$ candidates may pass the reconstruction and pre-selection stages. A unique B_{tag} candidate is selected during the reconstruction (choosing the one with highest FEI probability). However, the particles not associated to the tag-side, in some cases, can be combined to form different B_{sig} candidates, each having its own ROE.

The candidate multiplicity after pre-selection is shown in Figure 4.3 for simulated signal and background. All the events having a reconstructed ρ on the signal-side, have also an additional B_{sig} candidate with instead a $\tau \to \pi$ decay and an extra π^0 in the ROE. This leads to a high number of events with even candidate multiplicity. Furthermore, because of the loose pion-ID choice, the same particle can be reconstructed twice as a lepton and a pion. For simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal, the average candidate multiplicity per event is 4.4 and the fraction of events with one correctly reconstructed B_{sig} candidate is 76.6%. Here, a correctly reconstructed B_{sig} is defined as a candidate for which the K^{*0} and both the $\tau^\pm$ daughter particles (t_1, t_2) are all truth matched.

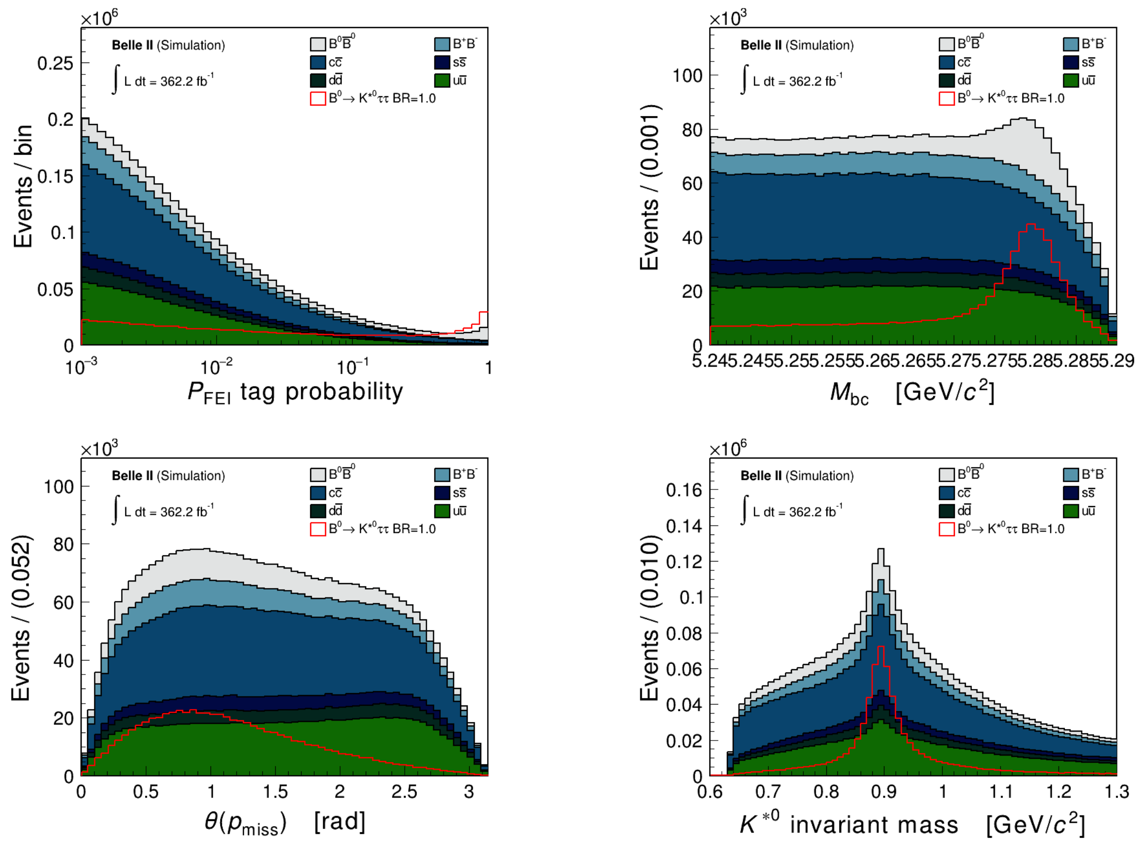


Figure 4.2: Distributions of continuum and $B\overline{B}$ backgrounds for the variables exploited in the pre-selection clean-up, right after the event reconstruction. Simulated backgrounds are normalized assuming an integrated luminosity of 362.2 fb^{-1}. The red histogram shows the expected signal distribution assuming an hypothetical branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 100\%$. A single $\Upsilon(4S)$ candidate per event is shown in the plot to avoid multiple counting of the same events.

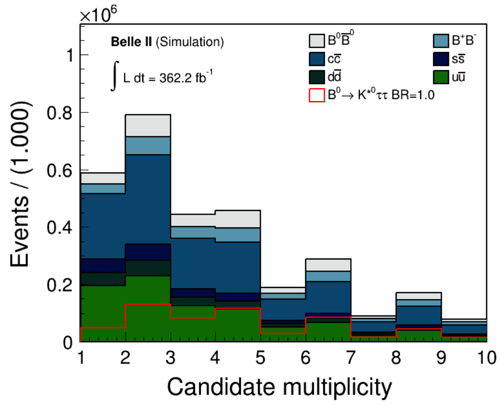


Figure 4.3: Distribution of per-event candidate multiplicity, on simulated backgrounds and signal after the pre-selection. The simulated backgrounds are normalized assuming an integrated luminosity of 362.2 fb^{-1}, while the signal is shown assuming a branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 100\%$.

In order to avoid any cross-feed among the signal categories defined in section 4.2.4, a selection is done to pick a single candidate. The following criteria are applied, in decreasing order of priority:

1. Select the $B^0 \to K^{*0}\tau^+\tau^-$ candidate(s) with a reconstructed K^{*0} mass closer to the nominal $M(K^{*0})$ value. After this request, the average candidate multiplicity for simulated signal becomes 3.66.
2. If multiple candidates are still in the event, select the ones with more identified $\tau^\pm \to \rho^\pm\nu$ on the signal side. This ensures that all the events with a reconstructed $\rho^\pm \to \pi^\pm\pi^0$ will be classified into the ρ signal category, and all the other signal categories do not have any reconstructed ρ.
3. Higher number of leptons (e, μ) identified as coming from τ decays, to prioritize leptonic final states, having a better sensitivity.
4. If multiple candidates are still left, choose one candidate randomly among the remaining ones. Before this last stage, the average candidate multiplicity for signal sample is 1.32.

These criteria for single candidate selection are applied right after the pre-selection. The same event yields listed in table 4.3, can be then divided by signal categories, as summarized in table 4.4. The $\ell\ell$ category, having only leptonic τ decay modes, has the lowest background yields, though having the lowest signal efficiency as well. In general, the background composition differs across the different signal categories, being $\ell\ell$ and $\pi\ell$ dominated by the $B\overline{B}$ component, while $\pi\pi$ and ρ categories, at the pre-selection stage, contain an higher fraction of $q\overline{q}$ backgrounds.

Table 4.4: Signal efficiencies and background yields after the pre-selection, divided by the four signal categories defined. The background yields are grouped into $B\overline{B}$ and continuum $q\overline{q}$ and they are weighted assuming an integrated luminosity of 362.2 fb^{-1}.

Signal category	$\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$	$B\overline{B}$ background	$q\overline{q}$ background
$\ell\ell$	0.83×10^{-4}	4.6×10^3	1.4×10^3
$\pi\ell$	1.48×10^{-4}	13.0×10^3	9.4×10^3
$\pi\pi$	0.70×10^{-4}	7.2×10^3	18.0×10^3
ρ	2.75×10^{-4}	30.5×10^3	46.1×10^3

For simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal, after applying the single candidate selection defined above, the fraction of correctly reconstructed B_{sig} candidates is 30.9%, to be compared against a completely random single candidate selection, resulting in a 26.9% of correctly reconstructed signal candidates. Here, a correctly reconstructed candidate is defined as a candidate for which the K^{*0} and both the $\tau^\pm$ daughter particles (t_1, t_2) are all identified and correctly assigned in the decay chain. A more detailed comparison between the random candidate selection and the nominal one, shows that the criteria chosen does not introduce any bias in any of the observables used for the signal extraction.

4.3 Comparison between data and simulation

To account for different responses between data and simulation, a set of correction factors is evaluated on specific control channels, and then applied on the simulated MC events.

4.3.1 Particle identification corrections

The identification efficiencies of charged particles and π^0 are corrected on the simulated samples using a variety of data control channels, which provides, beyond the correction values, also statistical and systematics uncertainties.

Hadron-ID

The reconstruction efficiency and the mis-identification (mis-ID) rate of charged pions and kaons are evaluated on specific control channels. Nominally, for kaon-ID efficiency the decay chain $D^{*+} \to \left[D^0 \to K^-\pi^+\right]\pi^+$ is exploited (see also section 2.4.4), while the decay $K^0_S \to \pi^+\pi^-$ is used for pion-ID efficiency [47]. From these same channels, the mis-ID rate of a kaon into a pion and vice-versa are derived. In addition, using a $J/\psi \to \ell^+\ell^-$ control channel, the mis-ID rates for leptons mis-identified as pions or kaons are estimated as well. Both the efficiency and the mis-ID rate, can be computed differentially in bins of particle momentum p and polar angle θ, to keep under control the variations on different regions of the detector.

The differences in efficiency and mis-ID rate between data and simulation, are also quantified through the control channels aforementioned, just by comparing the data against the MC simulation of the control channel. In particular, the data/MC ratio provides a correction value for simulated events, to be applied in each of the phase-space bin where the efficiencies are computed.

The tables shown in Figure 4.4 collect the data/MC efficiency ratios for pions and kaons, computed on the control channels in bins of reconstructed particle momentum p and polar angle θ. A hadron-ID efficiency correction weight, taken from such tables, is applied to each simulated event having a reconstructed pion or kaon on the signal-side, based on the generator-level p and θ of that reconstructed particle. Analogously, a suitable mis-ID correction weight is applied to the simulated events when the reconstructed hadron is actually mis-identified (based on the generator-level particle identity). After pre-selection, the mis-ID rate for $\pi \to K$ is 0.32% for the signal events and 7.3% for all the background events. The opposite mis-ID rate, $K \to \pi$, on the other hand, happens for 0.3% of the signal events and 0.9% of the background events.

In this analysis, the hadron-ID corrections are applied on all events to the $K^\pm$ and $\pi^\pm$ forming the reconstructed K^{*0}. In addition, for $\pi\ell$, $\pi\pi$ and ρ signal categories, the hadron-ID weights are applied also to the charged pions associated to $\tau \to \pi\nu$ and $\tau \to \rho\nu$

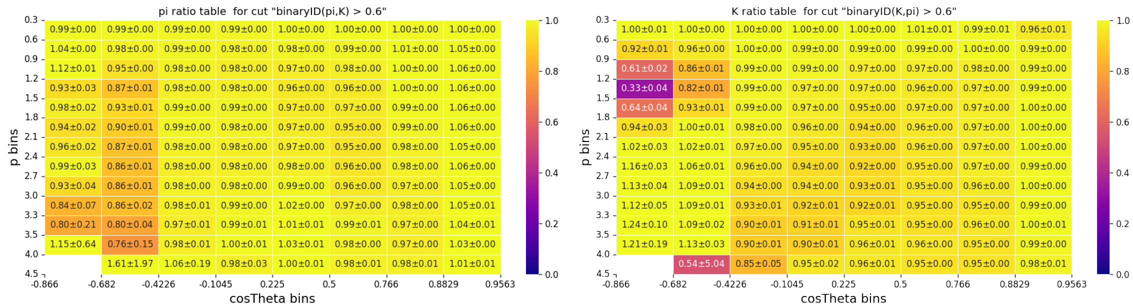


Figure 4.4: Efficiency corrections applied to charged pions (left table) and kaons (right table) for the simulated samples. The values are computed for bins of momentum and polar angle of the reconstructed particles comparing data and MC on specific control channels. The total uncertainty (statistical and systematics) of each correction is reported as well. Depending on the generator-level identity of the particle, either the efficiency or mis-ID rate correction weights are applied.

decays. If the signal-side contains multiple charged hadrons, as it always happens because of the reconstructed $K^{*0} \to K^+\pi^-$, all the correction weights are multiplied each others.

Lepton-ID

The selection efficiency for electrons and muons, and the mis-identification rate of pions and kaons as leptons, are corrected in bins of charge, particle momentum and polar angle, using the same scheme adopted for hadron-ID. Multiple control channels are used to extract the data/MC efficiency and mis-ID ratios in the different regions of phase space (see section 2.4.4 and Figure 2.16). For evaluating the lepton-ID efficiencies, $J/\psi \to \ell^+\ell^-$, $e^+e^- \to \ell^+\ell^-(\gamma)$ and $e^+e^- \to e^+e^-\ell^+\ell^-$ channels are exploited. The decays $K_S \to \pi^+\pi^-$ and $\tau^+ \to \pi^+\pi^+\pi^-\overline{\nu}$ are used for pion mis-ID rate, while the $D^0 \to K^-\pi^+$ for kaon mis-ID.

The efficiency correction weights are applied on simulated events for each correctly identified lepton on the signal-side. On the other hand, the mis-ID rate correction weights are assigned whenever a generated charged pion or kaon is identified as a lepton. No corrections instead are applied for the mis-identified $e \to \mu$ and $\mu \to e$, which occur for less than 0.4% of the reconstructed leptons after the pre-selection step.

The lepton-ID correction weights are applied only to the events belonging in the $\ell\ell$, $\pi\ell$ and ρ (if one τ decays leptonically) signal categories, to correct for the e or μ identified as a $\tau \to \ell\nu\nu$ candidate.

π^0 efficiency

For the ρ signal category, one or two neutral π^0 are reconstructed on the signal-side. The π^0 efficiency is measured via two distinct group of control channels, covering different regions

Table 4.5: Momentum dependent π^0 efficiency corrections. The calibration factors are derived from control channels with π^0 from D meson decays and, in the region $p_{\pi^0} > 1.0\,\mathrm{GeV}/c$, from τ decays. The last row shows the systematic uncertainty evaluated from the difference of the two control channels.

p_{π^0} [GeV/c]	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0	1.0-1.5	1.5-2.0	2.0-3.0
Eff. corr.	0.917	0.965	0.988	1.013	1.042	1.044	1.011
stat.	$\pm$ 0.004	$\pm$ 0.004	$\pm$ 0.004	$\pm$ 0.005	$\pm$ 0.004	$\pm$ 0.005	$\pm$ 0.005
D meson syst.	$\pm$ 0.049	$\pm$ 0.036	$\pm$ 0.079	$\pm$ 0.058	$\pm$ 0.045	$\pm$ 0.041	$\pm$ 0.040
D vs τ syst.					$\pm$ 0.039	$\pm$ 0.051	$\pm$ 0.030

of the π^0 spectrum. A first one exploits π^0 from D-meson decays and includes also slow π^0 from $D^* \to D\pi^0$, then in the low momentum region. The other control channel exploits the decays $\tau \to \pi\pi^0\nu$ in τ pairs promptly produced from the collision $e^+e^- \to \tau^+\tau^-$.

Table 4.5 lists the π^0 efficiency correction factors for the momentum bins considered, with their relative statistical and systematic uncertainties. A correction factor is applied to the simulated events for each reconstructed π^0, depending on its momentum.

4.3.2 Signal and background normalization corrections

Discrepancies in the overall data/MC yields can be due to mis-modeling effects in the pre-selection variables. However, most of the discrepancies are embedded into the FEI tagging reconstruction and depend on the cleanness of the reconstructed B_{tag}. Based on the quality of the tag-side, different complementary classes of simulated events are considered for evaluating the overall data/MC discrepancies:

- "Continuum" $e^+e^- \to q\bar{q}$ events, where $q = u, d, s, c$ are light quarks. Although the FEI may reconstruct a B_{tag}, no B meson is actually present in such events.
- $B^0 \to K^{*0}\tau^+\tau^-$ signal events. Due to the pre-selection requirements on M_{bc} and on the FEI probability, the tag-side has a very high purity: for about 64% of the signal events the B_{tag} decay chain is completely reconstructed with all the final state particles correctly identified.
- "Peaking" $e^+e^- \to B^0\overline{B}^0$ events. This class includes the generic $B^0\overline{B}^0$ where the B^0_{tag} candidate is correctly reconstructed (hence "Peaking" in M_{bc} and ΔE, like the signal), meaning again that all the particles in the B_{tag} decay chain are correctly reconstructed and identified. About 53% of the total simulated $B^0\overline{B}^0$ satisfy this requirement after the pre-selection.
- "Combinatorial" $e^+e^- \to B\overline{B}$ events. Defined as all the B^+B^- events, together with the $B^0\overline{B}^0$ not entering the "peaking" class (then not having the reconstructed

$B_{\rm tag}$ decay chain fully matched to the generated one). In these cases, an incorrect $B^0_{\rm tag}$ candidate is reconstructed by the FEI algorithm combining particles from two different B mesons, and by chance the combination gives a candidate surviving the pre-selection.

In the following, the procedure to derive an overall correction factor for each of these classes of events is described.

Continuum background

The continuum $e^+e^- \to q\overline{q}$ component, where $q = u, d, s, c$, can be separated from the rest of the physics processes exploiting, as a control channel, the 42.3 fb^{-1} off-resonance sample, collected at a CM energy 60 MeV below the $\Upsilon(4S)$ resonance, and thus below the $B\overline{B}$ production threshold.

To compare the off-resonance data against the $e^+e^- \to q\overline{q}$ MC sample (generated at the on-resonance beam energy), the energy and momentum of the reconstructed particles needs to be down-scaled, since the decay products share a lower CM energy. At the pre-selection stage of the analysis, the beam-constrained mass $M_{\rm bc}$ is the only variable depending on the reconstructed momenta, and it is therefore redefined for off-resonance events as in equation (4.5), to correct for the lower $B_{\rm tag}$ momentum.

Here, the 0.989 factor of difference in the cross sections (see section 4.2.1), can be neglected if compared to the Poisson statistical uncertainty of the off-resonance data sample.

After the FEI tagging, the signal-side reconstruction and the pre-selection, the overall ratio between off-resonance data and $q\overline{q}$ MC is computed for each signal category independently:

- $\ell\ell$: Data/MC $= 0.69 \pm 0.06$
- $\pi\ell$: Data/MC $= 0.68 \pm 0.02$
- $\pi\pi$: Data/MC $= 0.81 \pm 0.02$
- ρ: Data/MC $= 0.70 \pm 0.01$

where only statistical uncertainty, dominated by the number of off-resonance data events, is considered. Such factors should account for the overall mis-modeling of the continuum background, and they are hereafter applied to all $q\overline{q}$ simulated events.

Signal events and peaking $B^0\overline{B}^0$

The simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal and the peaking $B^0\overline{B}^0$ background (i.e. with the $B_{\rm tag}$ decay chain correctly reconstructed) can be treated with the same approach: although peaking $B^0\overline{B}^0$ are different with respect to the signal, they have the same multiplicity on the signal-side, and therefore the FEI tagging efficiency should be the same as well.

A specific control sample, called "embedded" sample, is built to estimate the data/MC differences for these classes of events. It consists of generic $B\overline{B}$ events where the signal-side $B^0 \to K^{*0}\tau^+\tau^-$ is artificially inserted.

Embedded sample The embedded sample is based on the clean and abundant $B^0 \to K^{*0}J/\psi$ decay channel, with $J/\psi \to \mu^+\mu^-$, being $\mathcal{B}(B^0 \to K^{*0}J/\psi) = 1.27 \times 10^{-3}$ and $\mathcal{B}(J/\psi \to \mu^+\mu^-) = 5.96\%$ [6]. This decay mode can be exclusively reconstructed both on data and on simulated $B^0\overline{B}^0$ sample. As shown in the sketch of Figure 4.5, the reconstructed signal-side $B^0_{\text{sig}} \to K^{*0}J/\psi$ is removed from the event, and replaced by a simulated $B^0_{\text{sig}} \to K^{*0}\tau^+\tau^-$. The ROE and the generic tag-side are still the ones from the $B^0 \to K^{*0}J/\psi$ event.

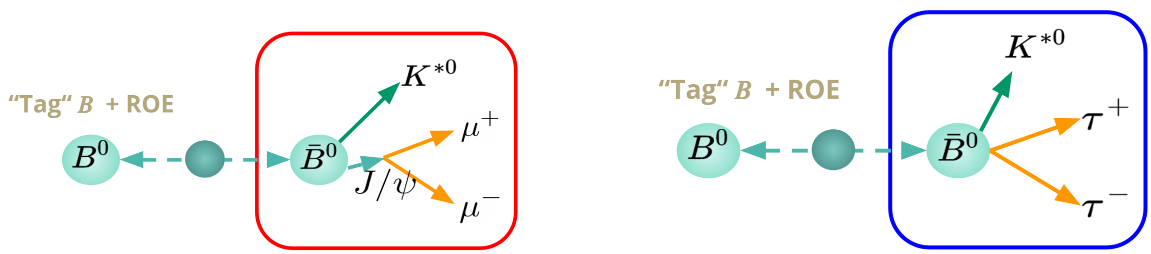


Figure 4.5: Simple sketch of the method to build the embedded control sample. First, $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ decay is reconstructed on a generic sample (left figure). The reconstructed B^0, with all the particles in its decay chain, is then replaced by a simulated $B^0 \to K^{*0}\tau^+\tau^-$ (right figure).

The $B^0 \to K^{*0}\tau^+\tau^-$ signal-side, together with all the particles in its decay chain, is in turn extracted from the simulated signal sample. To combine this signal side with the tag-side and the ROE of the $B^0 \to K^{*0}J/\psi$ event, the detector level objects (tracks, clusters, etc) associated to the simulated $K^{*0}\tau^+\tau^-$ particles must be isolated. They are reconstructed with the same criteria adopted in the main analysis (see section 4.2).

The two simulated τ leptons on the signal-side are free to decay inclusively. However, to build the embedded sample, only a fraction of their decay modes is exclusively reconstructed. Nominally, the same 1-prong decays considered in the present analysis: $\tau^\pm \to \ell^\pm\nu\overline{\nu}$, $\tau^\pm \to \pi^\pm\nu$ and $\tau^\pm \to \pi^\pm\pi^0\nu$, adding up to about 51.2% of all the combinations of generated $\tau^+\tau^-$.

To increase the statistics, the same B_{tag} and ROE are allowed to be merged multiple times with $B^0_{\text{sig}} \to K^{*0}\tau^+\tau^-$ having different τ decay modes. This implies that different reconstructed candidates of the embedded sample could have exactly the same B^0_{tag} and ROE.

Before merging the two halves of the events, the momenta of the simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal-side must be adjusted to account for the different kinematics, so that the two B mesons have opposite momenta in the CM frame. The combinations of $B_{\rm tag}$, $B_{\rm sig}$ pairs which do not satisfy this requirement are rejected.

In total, 77712 events are formed by merging a $B_{\rm tag}$ and ROE from simulated $B^0 \to K^{*0}J/\psi\,(\to \mu^+\mu^-)$, with a simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal (embedded MC sample). A specific MC sample with one B^0 decaying as $K^{*0}J/\psi$ and the other one decaying generically, is simulated at the scope. In addition, exploiting the full 362.2 fb^{-1} on-resonance data set, 7524 events are obtained merging the tag-side and ROE, extracted from $B^0 \to K^{*0}J/\psi\,(\to \mu^+\mu^-)$, with the MC $B^0 \to K^{*0}\tau^+\tau^-$ signal-side (embedded data sample).

These events undergo the reconstruction described in section 4.2.2, starting from the FEI tagging. After applying the pre-selection and the single candidate selection, 1982 events remain for the embedded MC sample, while 121 are left on the embedded data sample, corresponding to the efficiencies $\varepsilon_{\rm MC} = (2.55 \pm 0.06) \times 10^{-2}$ and $\varepsilon_{\rm data} = (1.6 \pm 0.1) \times 10^{-2}$. This translates to a data/MC efficiency ratio for the embedded sample of $\varepsilon_{\rm data}/\varepsilon_{\rm MC} = 0.63 \pm 0.06$, integrated over all the four signal categories. However, such ratio is consistent within errors across all the signal categories:

- $\ell\ell$: $\varepsilon_{\rm data}/\varepsilon_{\rm MC} = 0.50 \pm 0.10$
- $\pi\ell$: $\varepsilon_{\rm data}/\varepsilon_{\rm MC} = 0.64 \pm 0.12$
- $\pi\pi$: $\varepsilon_{\rm data}/\varepsilon_{\rm MC} = 0.63 \pm 0.17$
- ρ: $\varepsilon_{\rm data}/\varepsilon_{\rm MC} = 0.71 \pm 0.10$

Therefore, the unique correction $\varepsilon_{\rm data}/\varepsilon_{\rm MC} = 0.63 \pm 0.06$ can be used for all the signal categories.

This calibration factor is applied on all the simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal events and on the peaking $B^0\overline{B}^0$ MC events (i.e. with a good reconstructed $B_{\rm tag}$, as defined at the beginning of this section). The fraction of peaking $B^0\overline{B}^0$, with respect to the total $B^0\overline{B}^0$ yield, varies across different signal categories, being higher for the categories involving leptons. After the pre-selection, such fractions are 64% for $\ell\ell$, 60% for $\pi\ell$ and 51% for both $\pi\pi$ and ρ signal category.

Combinatorial $B\overline{B}$

As for the continuum component, the total yield of the $B\overline{B}$ combinatorial background, i.e. $B\overline{B}$ with at least one mis-reconstructed or missing particle from the tag-side, can be fixed in a data-driven way. The same-flavor control sample, defined by inverting the opposite B flavor requirement at the pre-selection level (see section 4.2.3), is exploited at the scope. This region is orthogonal by construction to all the events passing the standard pre-selection

Table 4.6: Signal efficiencies and background yields expected after the pre-selection for the same-flavor control sample, separated by signal category. The same-flavor sample is defined by inverting the requirement on the opposite flavor of reconstructed B mesons. Thus it contains pairs of B^0B^0 and $\overline{B}^0\overline{B}^0$ candidates. The background yields are grouped into $B\overline{B}$ and continuum $q\overline{q}$ and they are scaled assuming an integrated luminosity of 362.2 fb^{-1}.

Signal category	$\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$	$B\overline{B}$ background	$q\overline{q}$ background
$\ell\ell$	2.5×10^{-5}	1.43×10^3	0.76×10^3
$\pi\ell$	6.0×10^{-5}	4.52×10^3	5.20×10^3
$\pi\pi$	4.6×10^{-5}	3.14×10^3	9.72×10^3
ρ	11.3×10^{-5}	13.4×10^3	24.9×10^3

and, comparing its yields (table 4.6) to the ones computed on the opposite-flavor sample (table 4.4), the expected signal efficiency is between two and three times lower and the expected background yield is about 70% lower, with similar relative composition.

All the PID efficiency corrections described in section 4.3.1 are also applied to the same-flavor sample, while the B^0B^0 or $\overline{B}^0\overline{B}^0$ peaking events are weighted by the factor 0.63 ± 0.06, derived from the embedded sample. The continuum $q\overline{q}$ scaling can be computed specifically for the same-flavor, by inverting the $B\overline{B}$ flavor requirement on the off-resonance control sample as well. As done for the opposite-flavor sample, the following calibration factors are found in the off-resonance same-flavor sample:

- $\ell\ell$: Data/MC $= 0.45 \pm 0.07$
- $\pi\ell$: Data/MC $= 0.68 \pm 0.03$
- $\pi\pi$: Data/MC $= 0.77 \pm 0.03$
- ρ: Data/MC $= 0.66 \pm 0.02$

considering the statistical uncertainties, these factors are all compatible with the ones derived from the opposite-flavor off-resonance sample.

In the same-flavor sample, the number of observed data (N_{data}) is compared with the yield of simulated background events, where the continuum and the peaking BB are fixed to their expected values $N_{\text{MC}}^{q\overline{q}}$ and $N_{\text{MC}}^{\text{peak}}$. A calibration factor is thus derived for the combinatorial $B\overline{B}$ component as:

$$\text{C.F.}^{\text{comb}} = \frac{N_{\text{data}} - N_{\text{MC}}^{q\overline{q}} - N_{\text{MC}}^{\text{peak}}}{N_{\text{MC}}^{\text{comb}}} \tag{4.6}$$

$N_{\text{MC}}^{\text{comb}}$ is the number of combinatorial $B\overline{B}$ expected from the simulation, with only the PID corrections applied. For each signal category, the following calibration factors are obtained:

- $\ell\ell$: C.F.$^{\text{comb}} = 0.93 \pm 0.02$

- $\pi\ell$: $\mathrm{C.F.}^{\mathrm{comb}} = 0.78 \pm 0.01$
- $\pi\pi$: $\mathrm{C.F.}^{\mathrm{comb}} = 0.94 \pm 0.01$
- ρ : $\mathrm{C.F.}^{\mathrm{comb}} = 0.769 \pm 0.005$

The uncertainties associated here are the data statistical uncertainty of the same-flavor control sample. The factors, derived for each signal category, are applied as correction weights on all the combinatorial $B\overline{B}$ events.

4.3.3 Extra photons multiplicity

In order to check for possible mis-modeling of the E_{ECL}^{extra} variable (defined in section 4.2.2), distribution of simulated events are compared with data, using again the same-flavor control sample. With all the previous corrections in, Figures 4.6 and 4.7 show, respectively, the E_{ECL}^{extra} distribution and the multiplicity of photons associated to E_{ECL}^{extra}, comparing the same-flavor data and simulation for each signal category. In the E_{ECL}^{extra} distribution, the discrepancy between data and simulation is clearly visible for all signal categories. In particular, for low values, a relative difference of about 40% is observed in the first bins. Those bins of E_{ECL}^{extra} are the most important for the signal extraction, because background events could have some extra energy deposits in the ECL, which are not expected for well-reconstructed signal events, hence the signal sensitivity is higher for low values of E_{ECL}^{extra}. It is then crucial to understand and correct for these data/MC discrepancies.

The photon multiplicity distributions show a similar pattern for the ratio data/MC, indicating that more extra photons are reconstructed in the simulation. Such feature is interpreted as a mis-modeling in the simulated cluster reconstruction. A correction for this mis-modeling can then be directly derived from the extra photons multiplicity distribution. The data/MC ratio value is taken for each bin of photon multiplicity, and applied as a weight to all the events having the corresponding number of extra photons. Since the ROE distributions could depend not only on the detector response and on the event selection, but also on the background composition of the samples, therefore, the extra photons corrections are computed and applied independently for each signal category.

The effect of applying such weights can be observed in Figures 4.8 and 4.9. By construction, the corrected extra photons distributions has a flat data/MC ratio (Figure 4.9). Note that the weights applied for this correction are normalized separately for $B\overline{B}$ and $q\overline{q}$ simulated distributions. In this way the overall data/MC ratio, fixed by the corrections determined in section 4.3.2, is unaffected by this additional correction, which modifies only the shapes of the simulated distributions.

Regarding E_{ECL}^{extra}, the discrepancies present in the first bins of the distributions, are significantly reduced when applying the extra photons multiplicity correction, as shown in Figure 4.8.

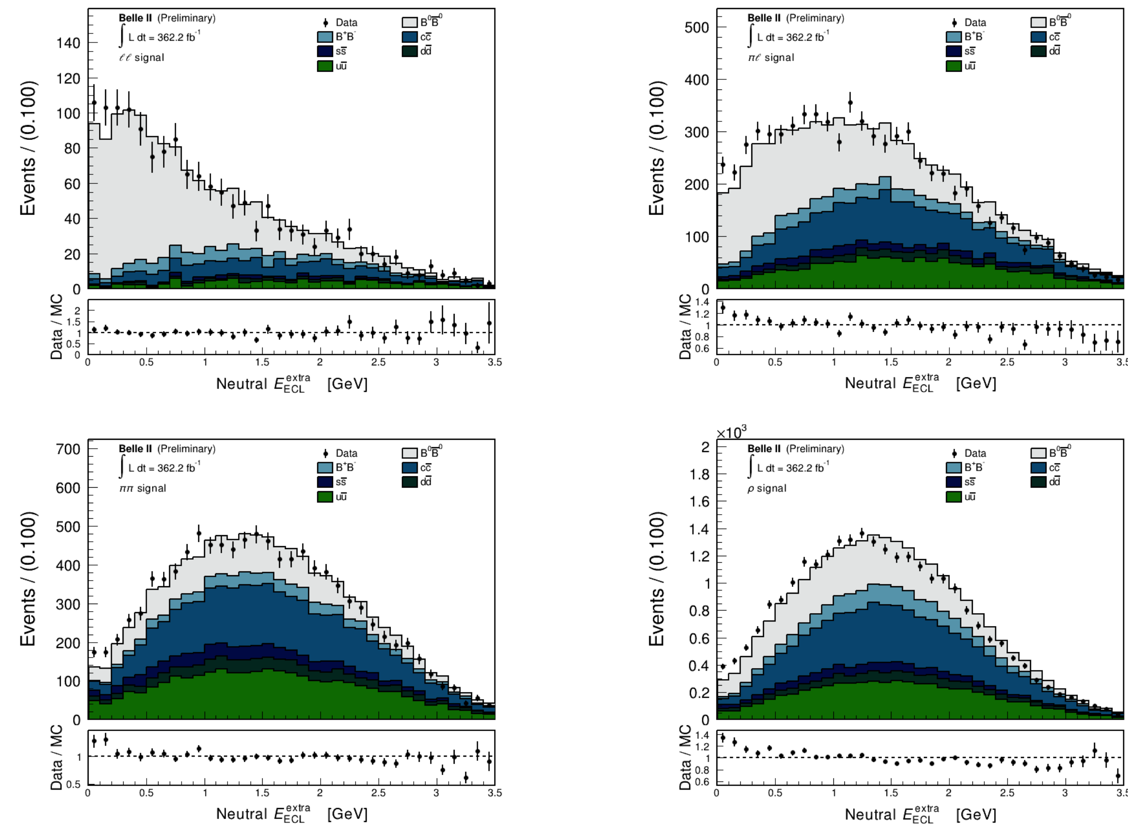


Figure 4.6: Distribution of E_{ECL}^{extra} for the same-flavor control sample before the extra photon multiplicity corrections. Particle-ID and background re-scaling weights are applied to the simulated events, which are also normalized to data integrated luminosity. Data/MC discrepancies are evident for all signal categories, especially at low values of E_{ECL}^{extra}.

Table 4.7: Signal efficiencies and background yields after the pre-selection with all the data/MC corrections applied, for each signal category. The background yields are grouped into $B\overline{B}$ and continuum $q\overline{q}$ and they are given assuming an integrated luminosity of 362.2 fb^{-1}.

Signal category	$\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$	$B\overline{B}$ background	$q\overline{q}$ background
$\ell\ell$	5.0×10^{-5}	3.5×10^{3}	1.0×10^{3}
$\pi\ell$	9.5×10^{-5}	9.5×10^{3}	6.7×10^{3}
$\pi\pi$	4.7×10^{-5}	6.2×10^{3}	14.4×10^{3}
ρ	16.9×10^{-5}	21.7×10^{3}	31.9×10^{3}

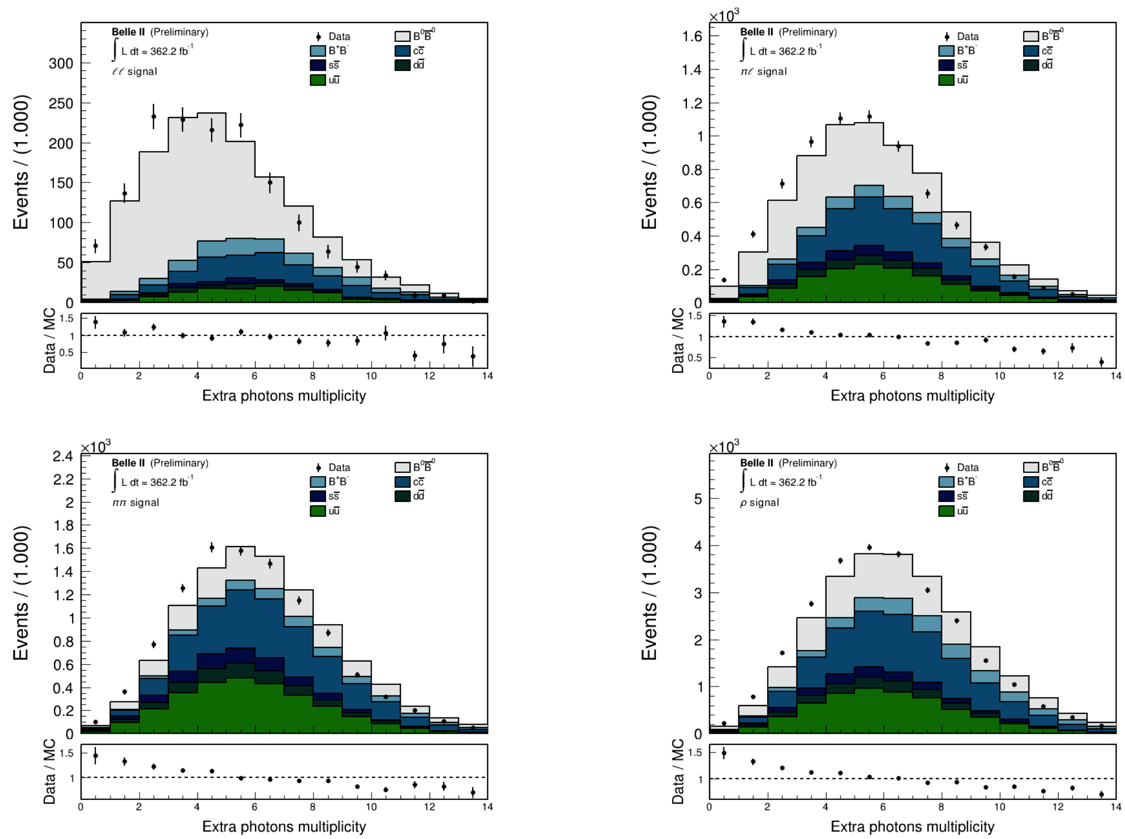


Figure 4.7: Extra photons multiplicity associated to E_{ECL}^{extra} for the four signal categories. The corrections are derived from the data/MC ratio of this distribution. Particle-ID and background re-scaling weights are applied to simulated distributions, which are also normalized to data integrated luminosity. Simulated events are weighted in order to have the same shape of data in this variable for each signal category.

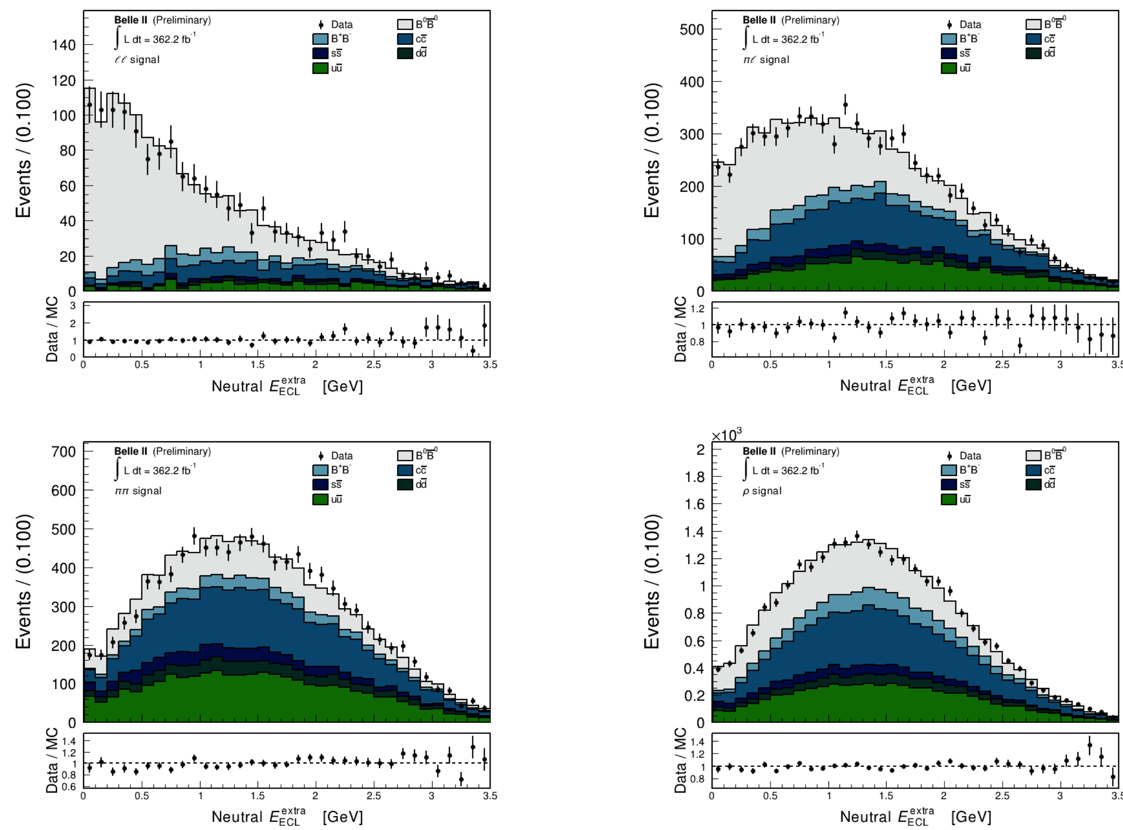


Figure 4.8: Distribution of E_{ECL}^{extra} for the same-flavor control sample after applying the corrections from extra photons multiplicity. Particle-ID and background re-scaling weights were already applied to the simulated events, which are also normalized to data integrated luminosity.

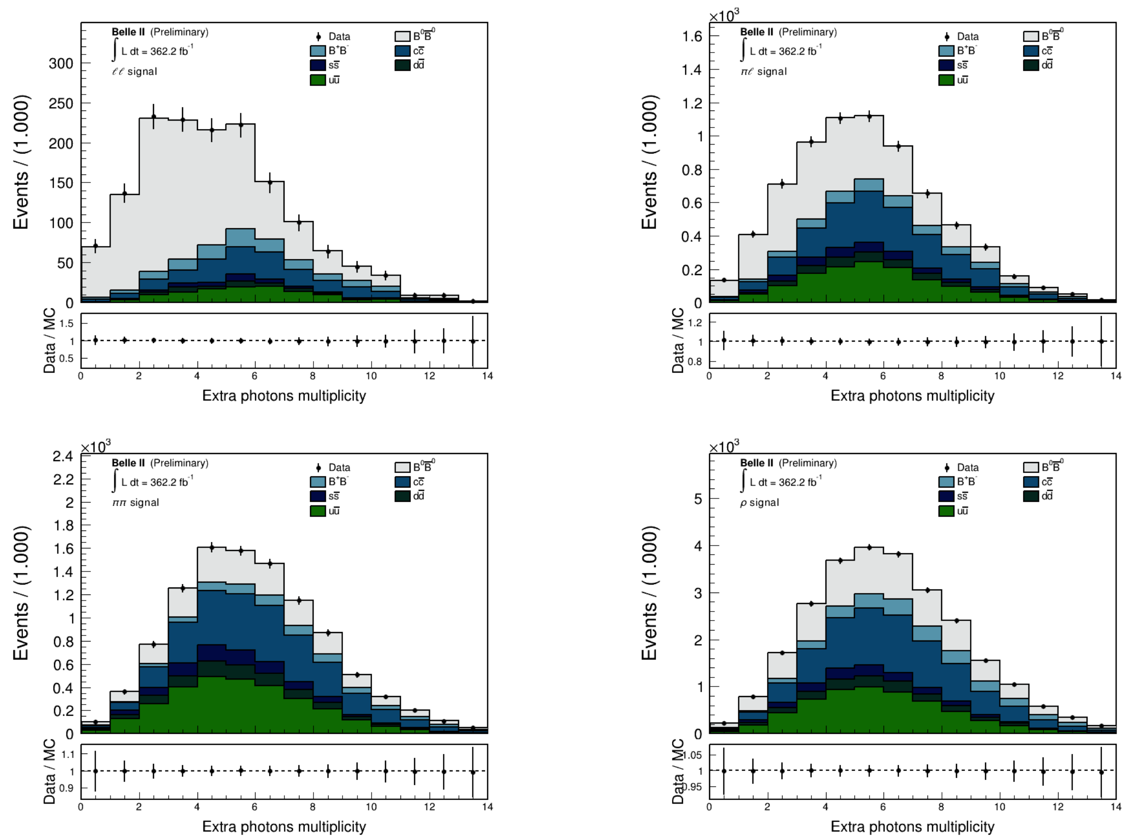


Figure 4.9: Extra photons multiplicity associated to E_{ECL}^{extra} for the four signal categories in the same-flavor control sample after the data/MC corrections derived for these same distributions. Particle-ID and background re-scaling weights are also applied to simulated distributions, which are normalized to data integrated luminosity. The data/MC ratio is perfectly flat, as expected by construction.

4.3.4 Corrected samples

After applying all the corrections described on the simulated sample, the yields of table 4.4 can be updated to the ones listed in table 4.7. The signal efficiency is about halved due to the calibration factor 0.63 ± 0.06 derived from the embedded sample. The overall background yields are reduced by about 30%, driven by continuum and combinatorial scaling factors.

4.4 Background Suppression

After the pre-selection step, a further separation of signal events from background is done exploiting a multivariate approach. With respect to a standard "rectangular-cut", this approach aims to optimize the full available information. For this analysis, in particular, the signal-to-background separation is obtained by training a binary classifier to recognize signal events against background. Applying the data-driven corrections derived in section 4.3, the distributions of simulated samples are modified to look more similar to data. Hence, the BDT can be directly trained on these corrected MC events, to be eventually applied on the Belle II dataset.

4.4.1 BDT classifier

As a binary classifier for this analysis, a BDT based on the XGBoost implementation of the gradient-boosting algorithm is chosen (see section 3.1). The four signal categories defined in section 4.2.4 may differ significantly among each other: the signal-sides are reconstructed targeting very heterogeneous final states, with different τ decay daughters t_1, t_2, as well as having different neutrino multiplicities. Therefore, to exploit at the most the power of a multivariate approach, one distinct BDT is devoted to each of the four signal categories. The four BDT classifiers are then trained on the simulated events passing the pre-selection, with a single candidate per event (selected as described in section 4.2.5), and applying all the data/MC correction weights (listed in section 4.3). Except for the weights, which may vary from event to event, all the backgrounds are treated at the same way, with no distinction among different components (as $q\bar{q}$, $B\overline{B}$ combinatorial and peaking).

Input variables

The BDT is trained exploiting as features several variables sensible to signal-background separation, including, as much as possible, complementary type of information. A total of 14 variables, listed and described in the following, are chosen and exploited in the training of all the four BDT classifiers. Their relative importance may vary a lot between different categories, depending on the dominant background components. The most important variables considered as training features are related to missing energy and ROE extra energy, followed by kinematic properties of the signal-side, variables sensible to peaking background from D mesons and information on the overall event topology, to suppress continuum $q\bar{q}$ background. Beyond their importance in the training, these features are chosen considering also the agreement between their simulated distributions and data, studied on control samples, as detailed in section 4.5.

The distributions for simulated signal and backgrounds are shown for all the signal categories in the figures A.1 - A.14 collected in appendix A.1.

Missing energy and ROE related variables

- Missing energy $E_{\rm miss}^{\rm cms}$ (Figure A.9), and missing momentum $p_{\rm miss}^{\rm cms}$ (Figure A.8) of the event, both computed in the CM system.
- $E_{ECL}^{\rm extra}$, defined as the extra energy deposited in the ECL not associated to any charged track (Figure A.10). $E_{ECL}^{\rm extra}$ is computed using the selection defined in section 4.2.
- Angle $\phi(K^{*0}; p_{\rm miss}^{\rm cms})$ between the reconstructed K^{*0} flight direction and the missing momentum direction in the CM system (Figure A.11).

Kinematic variables

- Invariant mass-squared of the $\tau^+\tau^-$ system, q^2 (Figure A.6). It is computed from the reconstructed four-momentum of the $B_{\rm tag}$ and the K^{*0}:

  $$q^2 = (p_{\Upsilon(4S)} - p_{B_{\rm tag}} - p_{K^{*0}})^2$$

  The expected shape of this variable for the signal directly depends on the form factor modeling.
- Invariant mass $M(K^{*0})$ of the reconstructed K^{*0} (Figure A.1). A window of width of $290\,\mathrm{MeV}/c^2$, around the nominal mass peak, was already selected at the pre-selection stage. However this information can be further exploited by the classifier.
- Momentum modules of the charged kaon, $p_K^{\rm cms}$, and of the charged pion, $p_\pi^{\rm cms}$, computed in the CM system (figures A.2 and A.3, respectively) and coming from the K^{*0} decay.
- Cosine $\cos\theta_{K;\pi}$ of the angle between the two K^{*0} daughters (Figure A.7).

Variables for D meson veto

- Invariant mass $M(K^{*0}; t_i)$, $i = 1, 2$ of the K^{*0} and the daughter of one τ. t_1 and t_2 have, respectively, the same and the opposite sign with respect to the charged kaon from K^{*0}. These variables are sensitive to possible D mesons on the signal side, such as a $B \to D\ell\nu$. If the D decays as $D \to K^{*0}t_i$, these variables would peak at the D mass value.

Continuum suppression variables Specific variables for rejecting the continuum $e^+e^- \to q\bar{q}$ background are exploited at the B factory experiments [22]. They are based on the principle that the production of a pair of light quarks is expected to be more jet-like, with respect to $B\overline{B}$ events, where each B meson has a momentum of about $330\,\mathrm{MeV}/c$ in the CM frame. The $B\overline{B}$ events would then have a more spherical spatial distribution. Therefore, so-called shape variables can be built to address these properties.

For a set of N particles having momenta $p_i, \ldots p_N$, the Fox-Wolfram moment of i-th order is defined as:

$$H_i \equiv \sum_{j,k}^{N} |\vec{p}_j||\vec{p}_k| P_i(\cos\theta_{jk}) \tag{4.7}$$

where P_i is the Legendre polynomial of i-th order, computed on the cosine of θ_{jk} angle between two particles j and k. For the shape variables adopted in this work, the indexes j and k run over the particles associated to the reconstructed B_{sig} candidate, and all the other detector-level objects not associated to the B_{sig}, respectively.

Two so-called Fisher discriminants h_{22}^{so} and h_{24}^{so}, made from modified Fox-Wolfram moments, are exploited for the BDT training (see Figures A.12 and A.13).

A different continuum-suppression feature, included as well in the BDT training, is the cosine $\cos(T_{B_{\text{sig}}}; T_{\text{ROE}})$ of the angle between the thrust axis of the B_{sig} and the thrust axis of the ROE and the B_{tag} (Figure A.14). The thrust is defined as the direction $\hat{n}$ which maximizes:

$$\frac{\sum_i |\vec{p}_i \cdot \hat{n}|}{\sum_i |\vec{p}_i|}$$

where the sum runs over the momenta of all the particles considered.

BDT settings

A total of about 54 k signal events, for all the signal categories, survive the pre-selection requirements. This is much less than the overall background events both for $B\overline{B}$ (about 427 k events, with 2.8 ab^{-1} of generated MC) and for $q\overline{q}$ (about 206 k events for 1.0 ab^{-1} of MC). All the available signal events are exploited in the training. For the $\ell\ell$ category all the background events are also used (about $\times 2$ the signal events), while for $\pi\ell$, $\pi\pi$ and ρ categories the number of background events in the training is limited to a maximum factor $\times 5$ larger than the number of signal events. The relative fraction of background types is kept the same as right after the pre-selection stage and all the corrections described in section 4.3 to weight the events are applied, both on signal and background. The XGBoost library makes possible to deal with unbalanced training datasets, as in this case signal and background are, by assigning a higher penalty, in the training, for the mis-classification of events which are in the less populated sample.

To avoid any bias from re-using the events on which the BDT was trained on, a two-fold cross-training is performed: samples are randomly split into two halves (here labeled as "even" and "odd"), and the same classifier is trained on both. The output weights of either the two trainings are then applied to the complementary half sample. This procedure allows to exploit the full available simulated signal sample.

The parameters of the BDT classifiers are tuned to maximize the area under the Receiver Operating Characteristic (ROC) curve [48], while at the same time limiting a possible

Table 4.8: Total number of events used in the training for each BDT (to be split into even and odd parts) and list of main hyper parameters optimized by a cross-validated grid scan.

Signal category	$\ell\ell$	$\pi\ell$	$\pi\pi$	ρ
Signal events	7682	12373	6548	25578
Background events	16405	61866	32740	127890
Max. $\frac{\mathrm{AUC_{train}}-\mathrm{AUC_{test}}}{\mathrm{AUC_{test}}}$	1%	0.8%	1%	0.5%
Number of trees	140	160	110	370
Learning rate	0.1	0.2	0.2	0.1
Tree depth	2	2	2	2
Subsample	0.5	0.5	0.8	0.5
Instances for new node	10	14	10	9

over-training. The ROC gives the rate of mis-classification as a function of the correct-classification efficiency. Its integral quantifies the overall performance of the classifier. The parameters which are optimized are the number of trees in the model, the maximum depth of each tree and the learning rate. This last parameter, scales down the trees added at each iteration of the boosting algorithm in order to reduce the impact of each individual tree.

The tuning is performed via a grid scan on multiple values of these parameters. At each point of the grid, a five-fold cross-training is performed, i.e. the sample is split into five sub-samples and a different classifier is tested on each sub-sample, while trained on the remaining four sub-samples. The performance is evaluated by considering the mean values of the areas under the ROC curve, $\mathrm{AUC_{train}}$ and $\mathrm{AUC_{test}}$, for those five classifiers. To limit the over-training, a threshold in the relative difference between train and test AUC is set for each signal category, and the scanned BDT parameters resulting in an highest $\mathrm{AUC_{test}}$ value below that threshold are chosen. The scans of parameters are shown in Figure 4.10 for the different signal categories.

Table 4.8 summarizes the values chosen for the most important BDT parameters, together with the total number of simulated events exploited in the training for each signal category.

After the training, the BDT output for each category is finally transformed in order to have a flat distribution in the interval $[0, 1]$ for the signal. The transformation is performed on 1000 quantiles of the BDT output variable, with an equal number of signal events for each quantile. For simplicity, the correction weights are not considered in this transformation. As a result, the final shape of the signal distribution is not exactly flat: the events with smaller extra photon multiplicity have an higher correction weight and tends to have an higher BDT value.

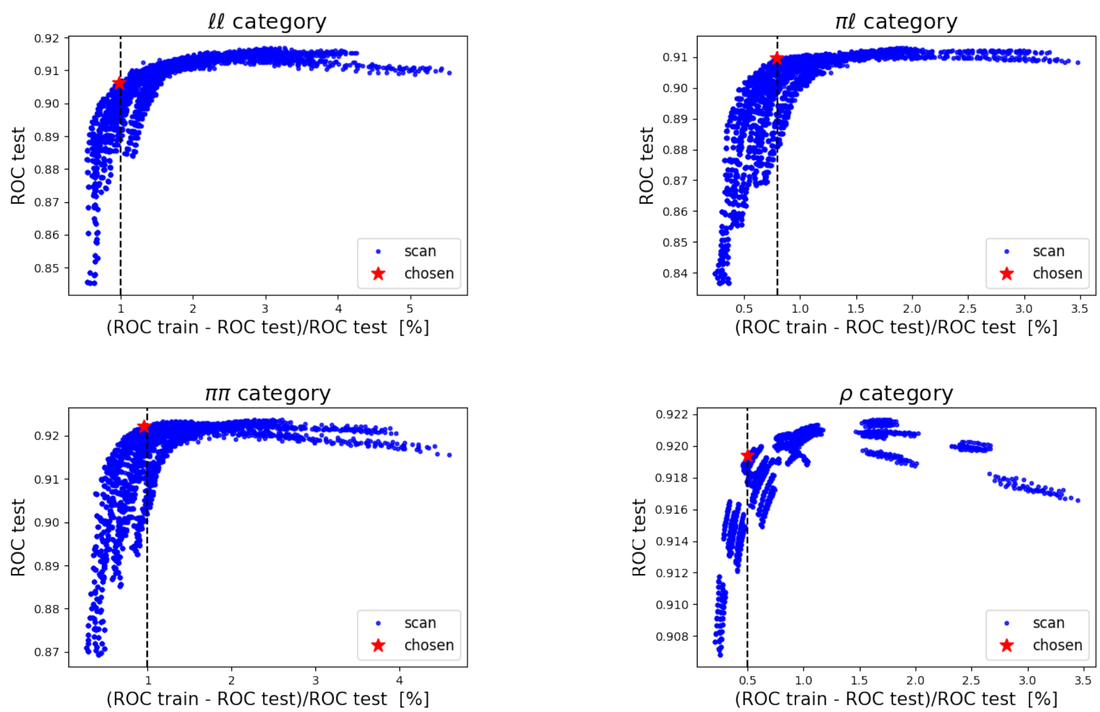


Figure 4.10: Scan of the BDT parameters for each signal category with a five-fold cross-validation. The values of the chosen parameters are summarized in table 4.8.

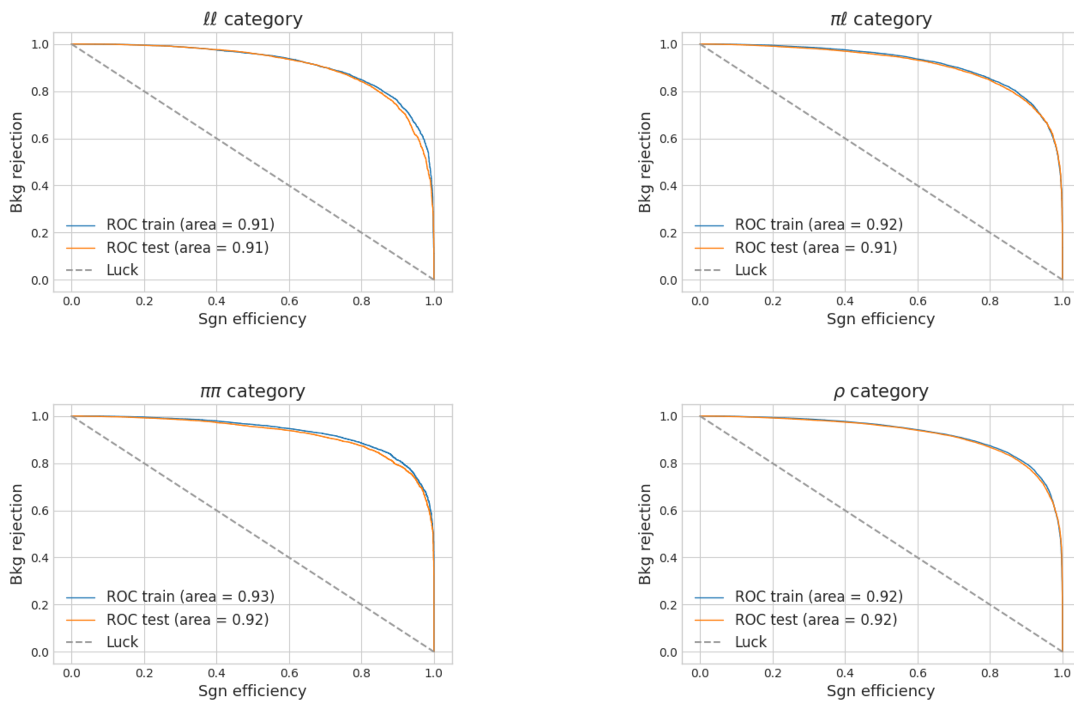


Figure 4.11: Receiver Operating Characteristic curves (ROC) evaluated for the train and test sample on each signal category. In this case the even subsample corresponds to the train, while the odd subsample to the test. No significant over-training is observed.

Training performance

For each signal category, the performance of the trained classifier can be evaluated by the ROC curve. This is shown in the plots of Figure 4.11, where the ROC is evaluated both on the training and on the complementary (test) sample. No visible over-training features are observed and the AUC values are consistent between test and train. Furthermore, the AUC values are about 90% for all the categories, indicating that the available information is exploited at the best for all the decay topologies.

The training importance of the features is shown in Figure 4.12 for each signal category. The importance is evaluated by summing up the gains from all the nodes in the tree where that variable is used [48]. The ranking varies depending on the signal category, anyway, the missing energy $E_{\text{miss}}^{\text{cms}}$ and the E_{ECL}^{extra} have a high score for all categories. In particular E_{ECL}^{extra} is more important for the categories with hadronic τ decays. The continuum suppression variable $\cos\theta(T_B; T_{\text{ROE}})$ is also more important for $\pi\pi$ and ρ categories, where the $q\bar{q}$ contribution to the background is dominant. On the other side, D veto variables have a higher raking for the $\ell\ell$ and the $\pi\ell$ categories, due to peaking backgrounds from semi-leptonic D decays.

In order to check that the splitting of the training between the even and odd subsamples are consistent, and no bias is introduced by this procedure, the BDT outputs distribution of the two complementary samples are compared both for signal (Figure 4.13) and the total backgrounds (Figure 4.14). Conservatively, only the statistical uncertainty of the output sample is considered, neglecting the uncertainties from the training sample and the BDT random sampling. However, the two distributions are in a reasonable agreement for all the four signal categories, and in particular for the $\ell\ell$ and $\pi\ell$ ones.

4.4.2 Signal region

Figure 4.15 shows the transformed BDT outputs for all the signal categories, computed on all the simulated events passing the pre-selection. In principle, these distributions could be taken as they are to extract the signal yield. However, an alternative choice, here pursued, is to restrict the signal (extraction) region, avoiding to include in the fit the bins with highest background yields. Since the ROC performances are similar for all the signal categories (as can be seen from Figure 4.11), an equal cut BDT > 0.4 is chosen to define such signal region. This choice also allows to use the sideband of the signal region (BDT < 0.4) as an additional control sample to validate the data/MC agreement for backgrounds.

The expected signal efficiencies and background yields in the signal region, are summarized in table 4.9. Figure 4.16 shows the simulated distributions of the BDT output variable by zooming in this region. The binning of those histograms will be used in the binned maximum likeihood fit (see section 4.6). The width is chosen by requiring a minimum of ten expected background events in the most sensitive bin.

The choice of such signal region cut and binning, can be justified by comparing the

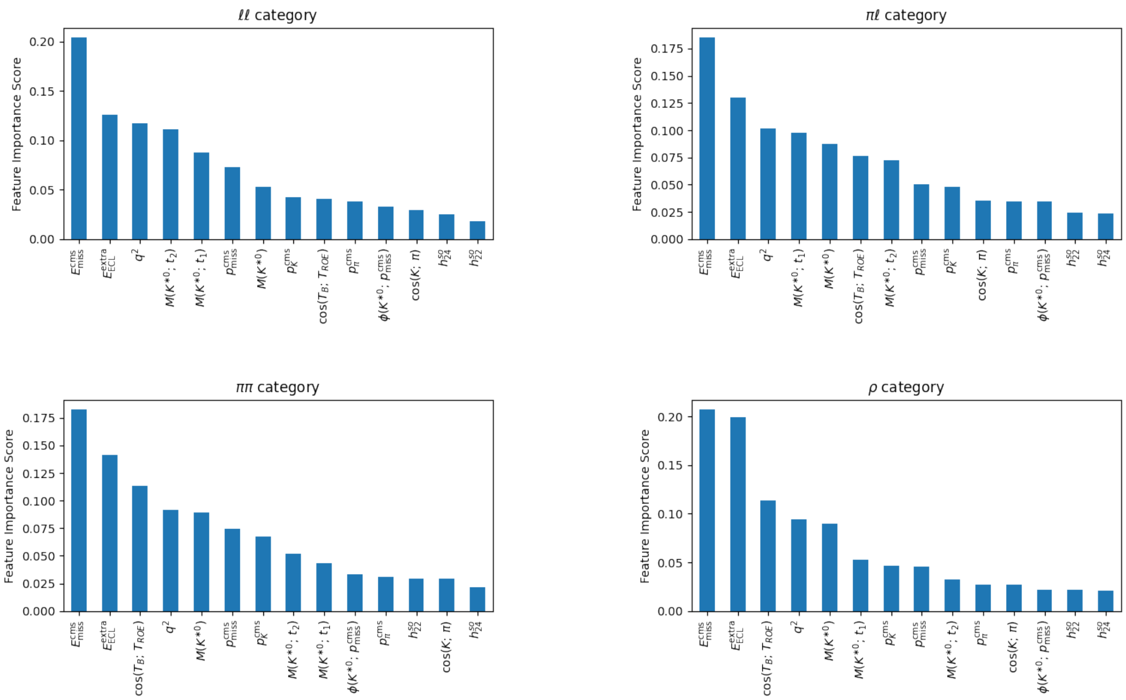


Figure 4.12: Ranking of the 14 BDT training features for the four signal categories. Only the results for even subsamples are shown here, however the feature rankings for the odd training are consistent.

Table 4.9: Signal efficiencies and background yields computed in the signal region, defined by the selection on the classifier output BDT>0.4. The background yields are grouped into $B\overline{B}$ and continuum $q\bar{q}$ and they are weighted assuming an integrated luminosity of 362.2 fb^{-1}.

Signal category	$\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$	$B\overline{B}$ background	$q\bar{q}$ background
$\ell\ell$	$(3.06 \pm 0.05) \times 10^{-5}$	281.8	41.0
$\pi\ell$	$(5.94 \pm 0.07) \times 10^{-5}$	1022.0	237.9
$\pi\pi$	$(3.01 \pm 0.05) \times 10^{-5}$	1025.1	413.0
ρ	$(10.6 \pm 0.1) \times 10^{-5}$	2925.6	802.8

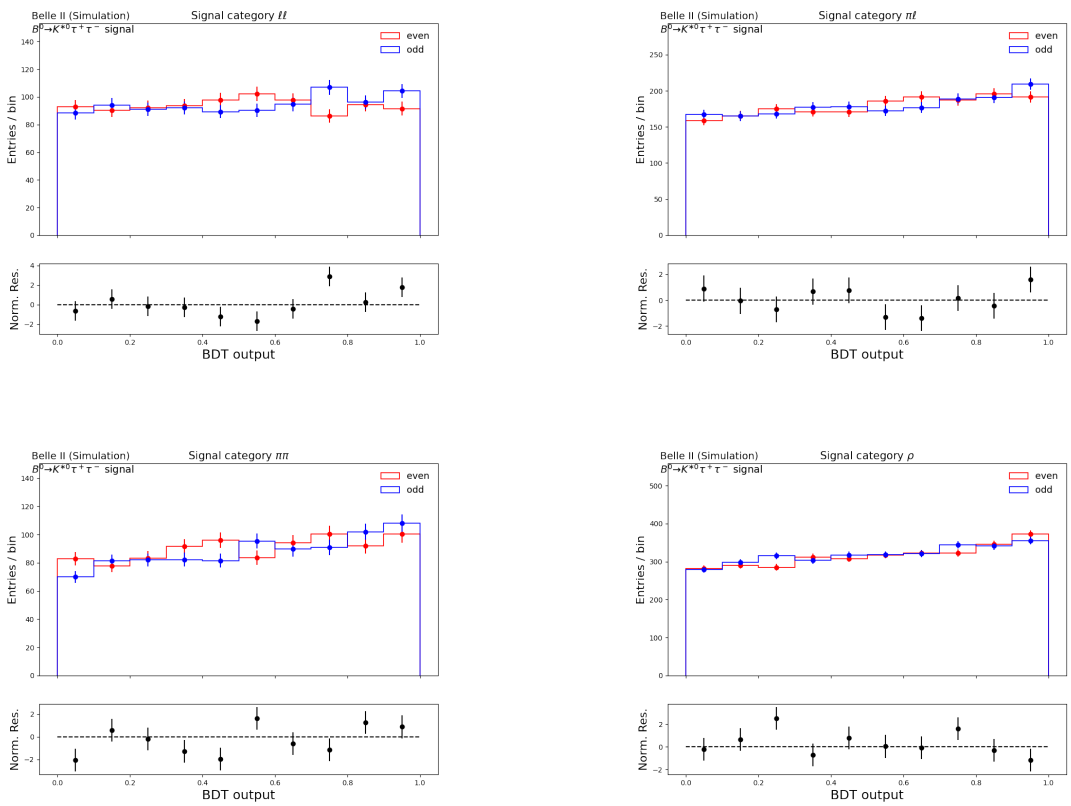


Figure 4.13: Comparison of the BDT output signal distributions for the odd subsample (with BDT weights from the even training) and the even subsample (with BDT weights from the odd training). Only the statistical uncertainty of the output sample is considered for the error bars. All the correction weights derived in section 4.3 are also applied.

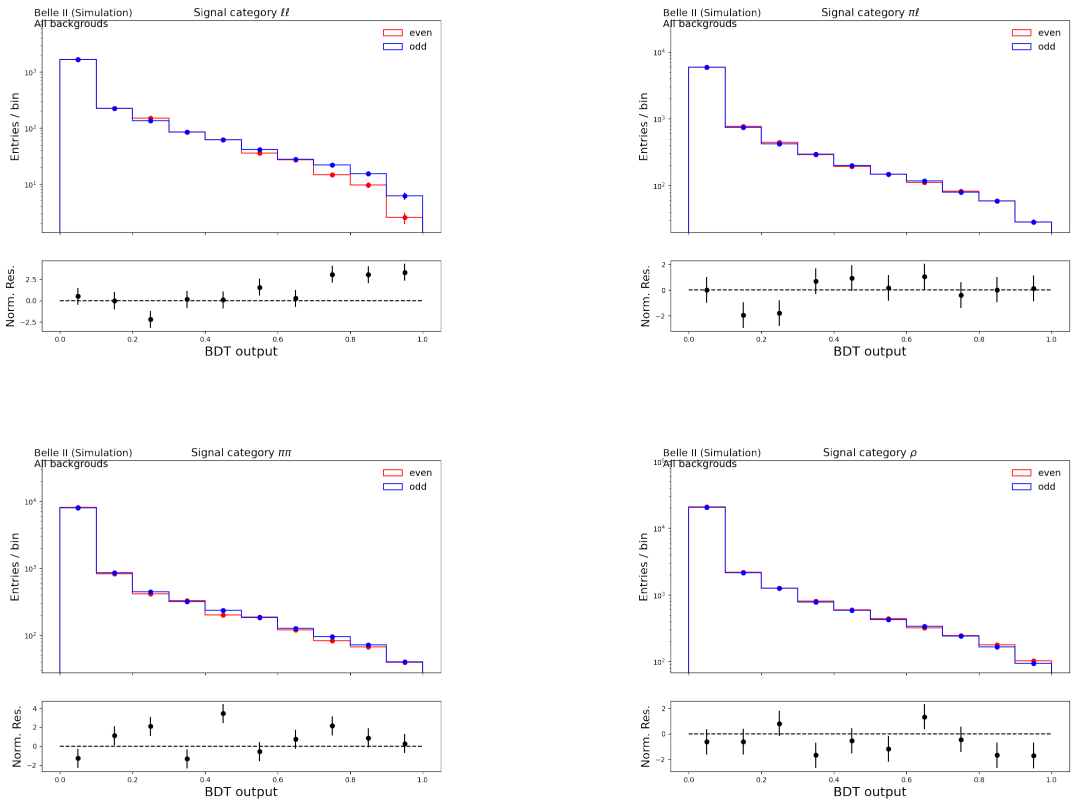


Figure 4.14: Comparison of the BDT output background distributions for the odd subsample (with BDT weights from the even training) and the even subsample (with BDT weights from the odd training). Only the statistical uncertainty of the output sample is considered for the error bars. All the correction weights derived in section 4.3 are also applied.

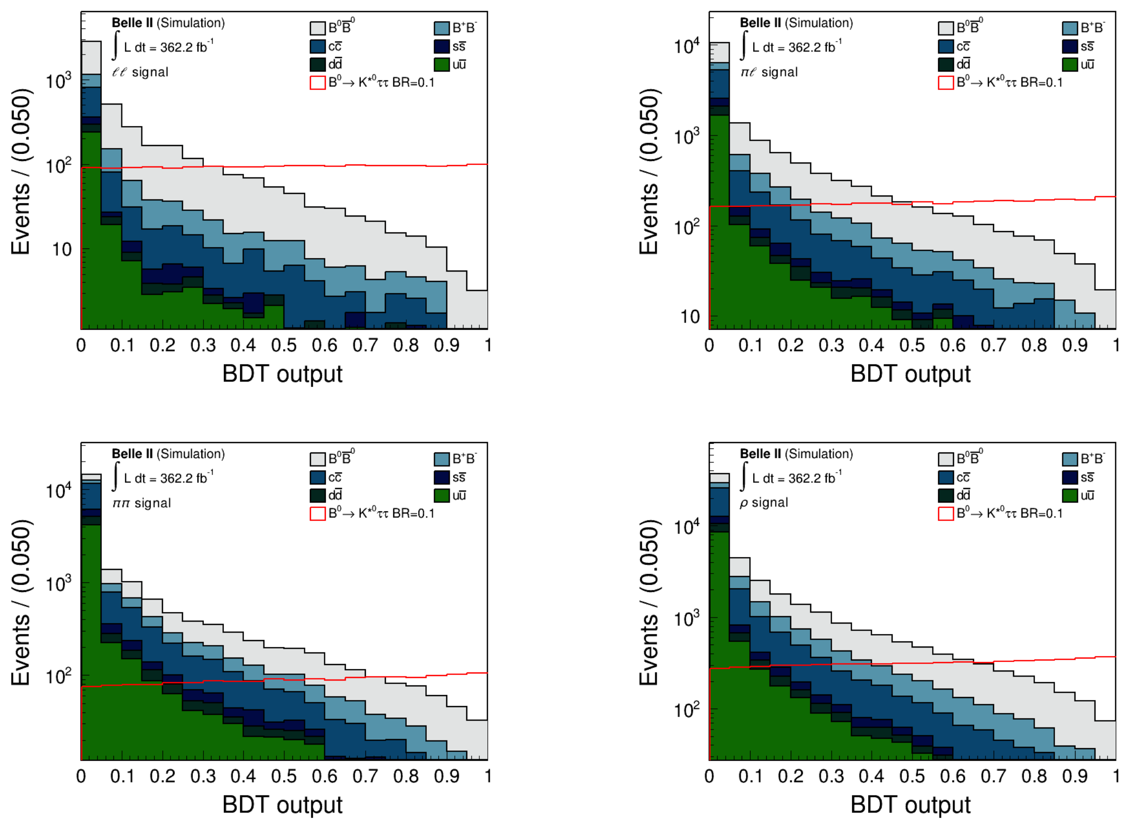


Figure 4.15: BDT output (after flat-signal transformation) for all signal categories. All available corrections are applied and the signal MC histogram is scaled to a branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

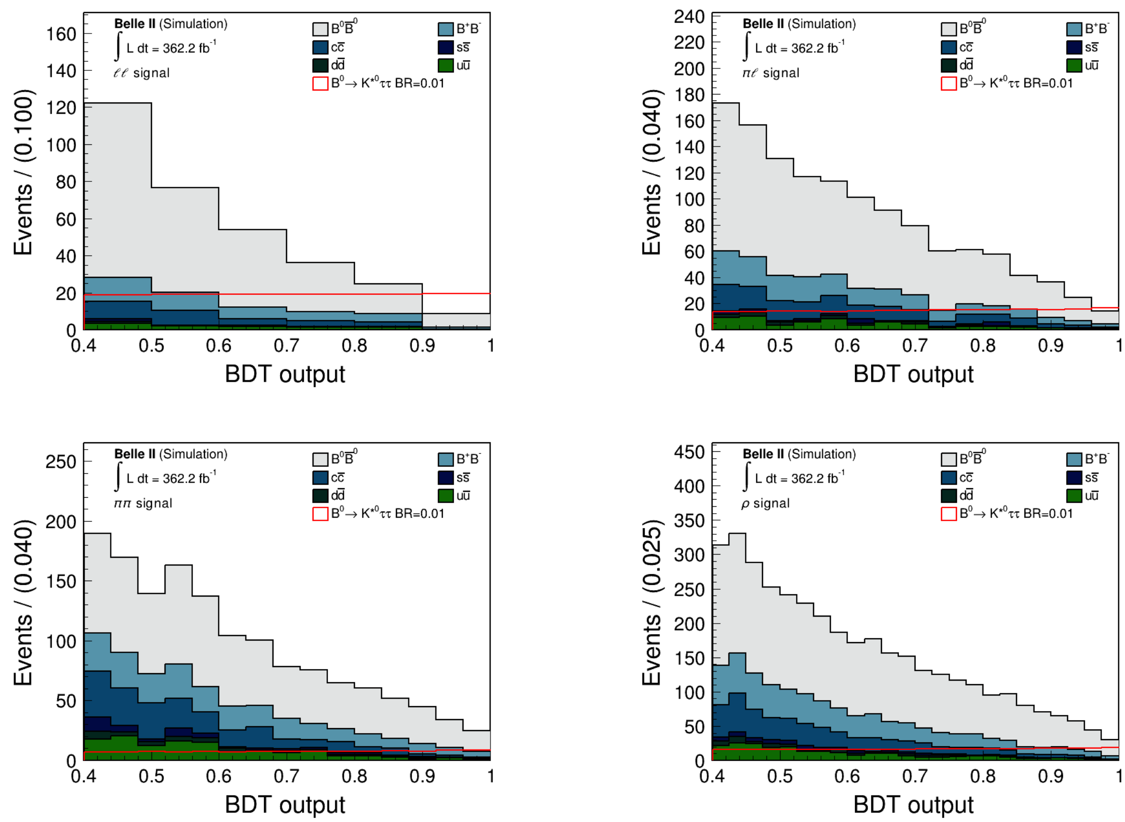


Figure 4.16: BDT output (after flat-signal transformation) for all categories with the optimized signal region cuts. All available corrections are applied and the simulated signal is scaled to a branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.01$.

discrepancies in the data/MC shapes from control sample, with the variations in the shape of the histograms deriving from the systematic uncertainties. Indeed, in section 4.6.1, it is shown that, for the chosen signal region, the systematics cover the observed data/MC differences.

Residual backgrounds characterization

As clear from table 4.9, the $\ell\ell$ signal category is the one with the highest purity and thus the signal extraction will be mostly driven by it. Considering the simulated background events in the last BDT bin of the $\ell\ell$ category (BDT$>$ 0.9), about 87% of them are of $B^0\overline{B}^0$ type and 50% of the $B^0\overline{B}^0$ have a B_{tag} completely truth-matched (then are classified as peaking). Based on their generator-level decays, the $B^0\overline{B}^0$ can be roughly grouped as follows:

- 19% of the $B^0\overline{B}^0$ have a signal-side $B^0 \to D^{(*)}\tau\nu$ decay, with the $D^{(*)}$ subsequently decaying semi-leptonically as $D^{(*)} \to h\ell\nu$.
- Similarly, 49% of the $B^0\overline{B}^0$ residual background events are from a chain of semi-leptonic decays of the type: $B \to D^{(*)}\ell\nu$, $D^{(*)} \to h\ell\nu$ with $\ell = e$, μ.
- A fraction of 23% of the $B^0\overline{B}^0$ events has just one semi-leptonic $D^{(*)}$ decay and a particle mis-identified as the second lepton in the signal-side (mostly $\pi \to \mu$ mis-ID).
- The remaining 12% of the $B^0_{\rm sig}$ decays are not truth matched to a specific decay mode, and they appear to have protons or other baryons on the signal side.

Furthermore, about 6% of the $B^0\overline{B}^0$ events contain a K^0_L from a $D^{(*)} \to XK^0_L$ decay or from $B^0 \to DDK^0_L$.

4.5 Validation

The simulated background and signal distributions of the variables used for the training of the classifier are validated on data control samples, keeping blind the data events entering the signal region. These same control samples can be used to validate the BDT output distribution as well, hence to estimate a set of systematic uncertainties which will be eventually associated to the simulated events in the signal region.

4.5.1 Validation of continuum background

The off-resonance sideband, with 42.3 fb^{-1} of data available, is used to validate the simulated continuum $e^+e^- \to q\overline{q}$ backgrounds. Once all the corrections described in section 4.3 are in place, the agreement between continuum simulation and off-resonance data is checked for all the input features of the BDT, independently for each signal category. The plots showing the comparisons are collected in the Figures from A.15 to A.28, in the appendix A.2.

Similarly, the BDT output can be computed on the off-resonance data and compared to the continuum simulation, as shown in Figure 4.17. The distributions for simulated $q\overline{q}$ and off-resonance data, zooming into the signal region BDT > 0.4, are shown in Figure 4.18.

Taking the ratio between observed off-resonance data and expected $q\overline{q}$ backgrounds in the signal region, one finds:

- $\ell\ell$: Data/MC $= 0.42 \pm 0.30\,(71\%)$
- $\pi\ell$: Data/MC $= 0.83 \pm 0.17\,(20\%)$
- $\pi\pi$: Data/MC $= 0.93 \pm 0.14\,(15\%)$
- ρ: Data/MC $= 0.82 \pm 0.09\,(11\%)$

where only the statistical uncertainties on the number of data is considered for all the ratios. The relative uncertainty reported is taken as the normalization systematic for the expected $q\overline{q}$ yield in the signal region. Since such systematics cover the observed discrepancy of these ratios with respect to one, no further correction is applied on the continuum yield.

4.5.2 Signal validation

As described in section 4.3.2, the embedded control sample is built to validate and correct the signal efficiency. The same correction factor is then used for the $B^0\overline{B}^0$ peaking background component, as well. Beyond the overall data/MC yield, the distributions of the BDT features can be also validated on the embedded control sample.

The plots collected in Figures from A.57 to A.59 in the appendix A.4, show the 14 input variables of the BDT training for the data and simulated embedded sample. To get

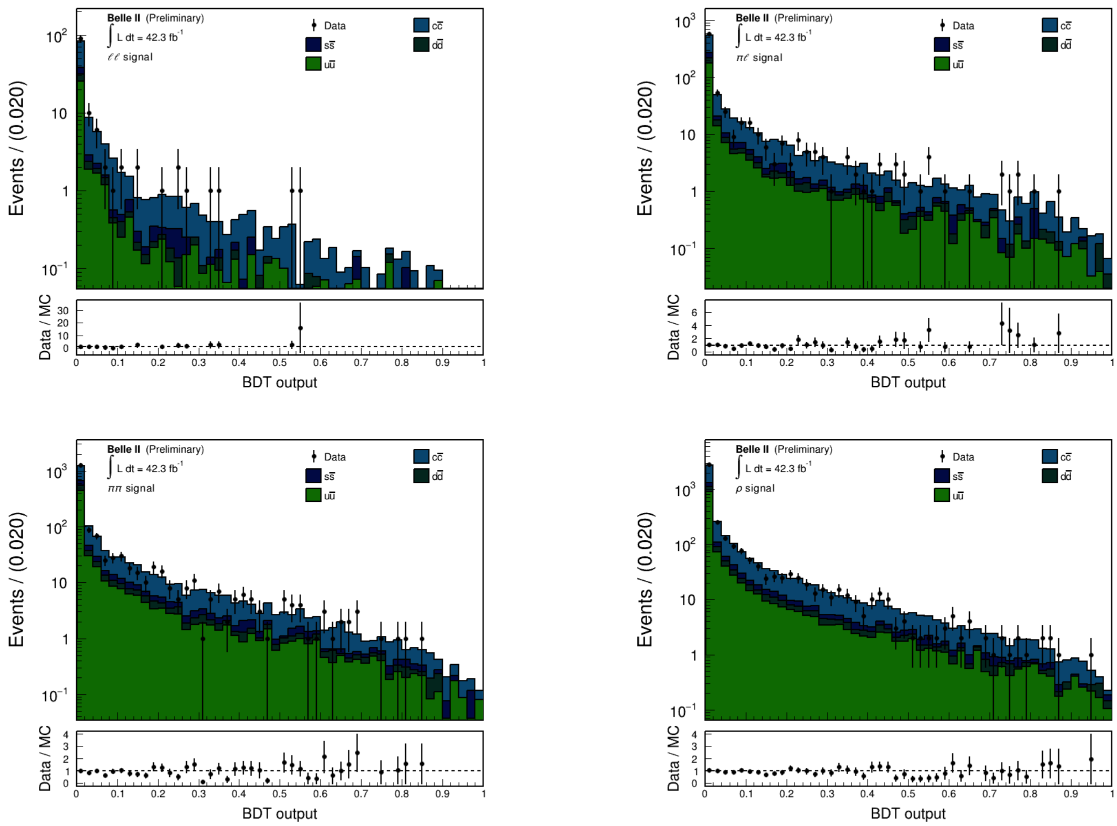


Figure 4.17: BDT output for off-resonance data and continuum simulation. All available corrections are applied.

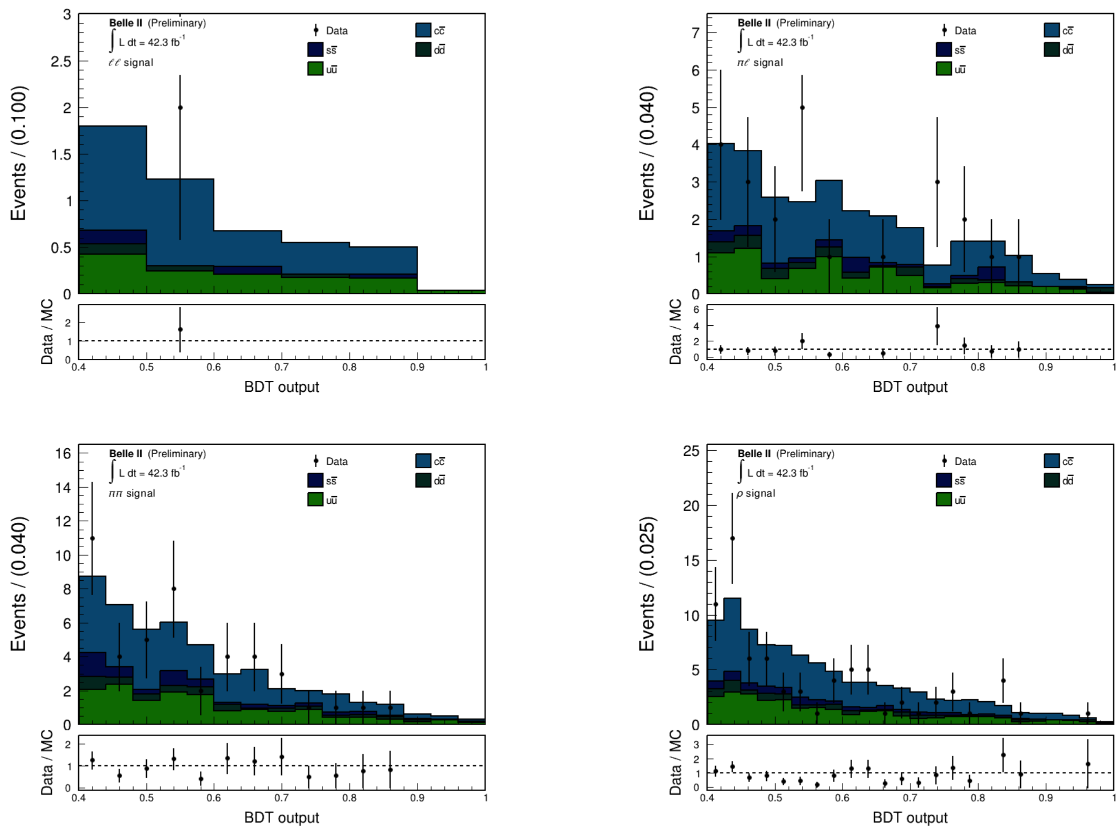


Figure 4.18: Zoom of the BDT output in the optimized signal region: comparison between off-resonance data and continuum simulation. All available corrections are applied.

Table 4.10: Summary of the data/MC efficiency for the embedded sample, computed at the pre-selection stage and in the signal region. The values are consistent at both steps of the analysis and across different signal categories as well.

Signal category	Pre-selection $\varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$	BDT $>$ 0.4 , $\varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$
$\ell\ell$	0.50 ± 0.10	0.45 ± 0.12
$\pi\ell$	0.64 ± 0.12	0.49 ± 0.12
$\pi\pi$	0.63 ± 0.17	0.53 ± 0.17
ρ	0.71 ± 0.10	0.77 ± 0.12
All	0.63 ± 0.06	0.60 ± 0.07

rid of the low data statistics, the distributions were plotted summing over all the signal categories. The embedded MC is scaled by the 0.63 correction factor, derived in section 4.3.2. Only the extra photon multiplicity weights are applied to the MC histograms, which mainly affect the shape of E_{ECL}^{extra}. The PID corrections are not applied since the needed truth-level particle identity is lost during the embedding procedure.

In Figure A.60 the distribution of the BDT outputs for the four signal categories are shown as well. Applying the signal region BDT > 0.4 selection for all the signal categories, the total number of events in the embedded data sample reduces to $N_{\text{data}}^{\text{BDT}>0.4} = 85$, while for the simulated embedded sample $N_{\text{MC}}^{\text{BDT}>0.4} = 1474$. The ratio of data/MC efficiencies, instead, is 0.60 ± 0.07, still in excellent agreement with the value at the pre-selection stage. The 12% statistical uncertainty of this correction in the signal region, will be assigned as a systematic uncertainty for the embedded calibration factor, acting on signal and peaking background events. All the yields for embedded samples, and the data/MC efficiency ratios, divided by signal categories, are summarized in table 4.10.

4.5.3 Validation of generic background

The same-flavor control sample, exploited in section 4.3 for correcting the combinatorial background and adjusting the distribution of the extra photons multiplicity, can be also used to validate the shapes of the BDT output and of its input features.

The shapes of the simulated backgrounds can be directly compared between the same-flavor and the opposite flavor samples. The plots showing such comparison for the input variables of the BDT are collected in appendix A.3 (figures from A.29 to A.42). The simulated same-flavor events are hereafter corrected with all the factors derived in section 4.3. For the comparison, the histograms are scaled in order to have the same total background normalization on both samples. The shapes of most of the BDT training features are in a good agreement for the two complementary samples.

The data/MC comparisons for the same-flavor control sample are collected in appendix A.3 (Figures from A.43 to A.56), for the training inputs of the BDT and separately for all

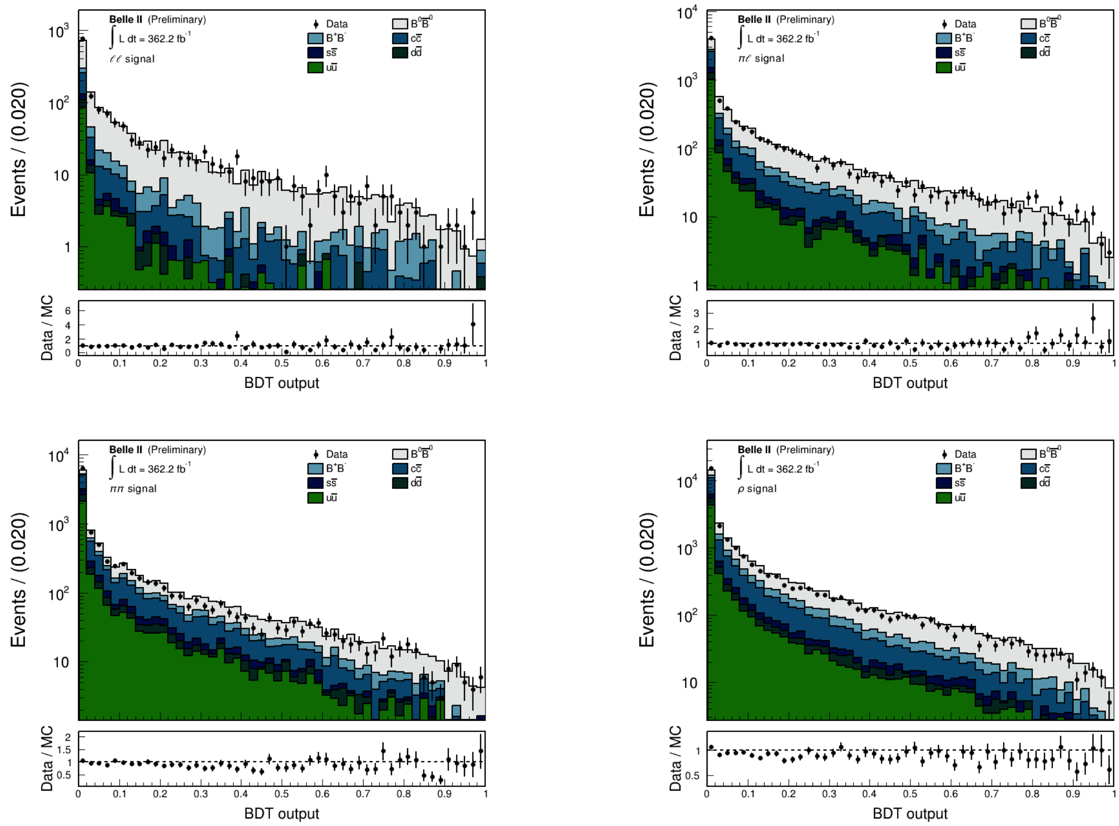


Figure 4.19: BDT output for all signal categories compared between data and simulation in the same-flavor control sample. All available corrections are applied.

signal categories. For all the variables a reasonable agreement in the shapes is observed.

Figure 4.19 compares the full BDT output distribution for the same-flavor data and simulation, showing a good data/MC agreement on this control sample.

Restricting to the signal region, the expected signal efficiency and background yields for the same-flavor sample are reported in table 4.11. Compared to the corresponding table 4.9 for the opposite-flavor selection, the signal efficiencies are about four times lower in the same-flavor sample, with total background yields still about a factor of two lower. The comparison of the BDT output distributions, in Figure 4.20, highlights again a reasonable agreement in the signal region as well, consistently with what observed for the input features.

The data/MC comparisons in the full BDT output and restricted to the signal region is shown in Figure 4.21. Finally, the comparison for the BDT output in the same-flavor signal region is shown in Figure 4.19, indicating an overall good agreement in the shapes. Regarding the background normalization in this region, the overall ratio between observed

Table 4.11: Expected signal efficiencies and background yields in the BDT output signal region for the same-flavor sideband sample. The background yields are grouped into $B\overline{B}$ and continuum $q\overline{q}$ and they are weighted assuming an integrated luminosity of 362.2 fb^{-1}.

Signal category	$\varepsilon(B^0 \to K^{*0}\tau^+\tau^-)$	$B\overline{B}$ background	$q\overline{q}$ background
$\ell\ell$	$(0.8 \pm 0.1) \times 10^{-6}$	150.9	38.8
$\pi\ell$	$(1.7 \pm 0.2) \times 10^{-6}$	541.9	212.9
$\pi\pi$	$(3.3 \pm 0.2) \times 10^{-6}$	1435.5	817.6
ρ	$(1.7 \pm 0.1) \times 10^{-6}$	905.9	483.9

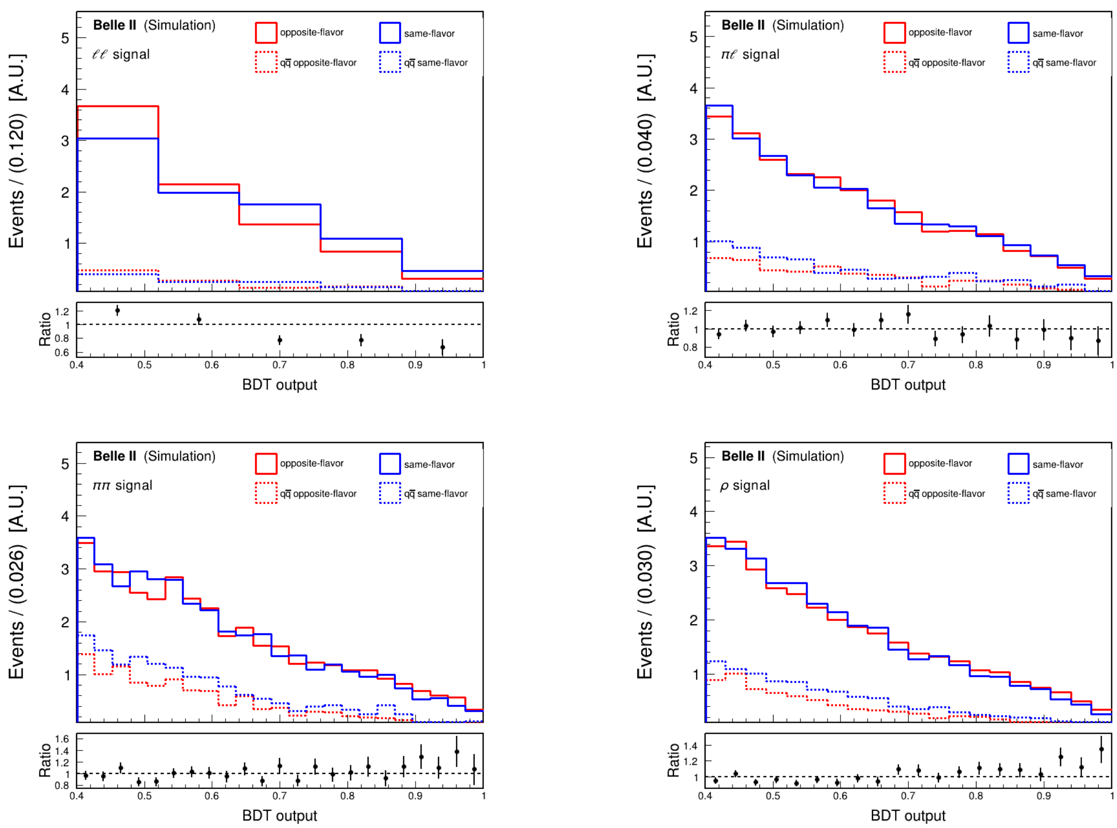


Figure 4.20: Expected BDT output distributions of the total background yields compared between the nominal sample (opposite-flavor) and the same-flavor control sample. The histograms are normalized to the total background yields in the signal region, so that the differences in the shapes can be evaluated. The $q\overline{q}$ simulated components are shown with dashed lines. All the available corrections are applied on both simulations.

data and simulated background events, with all the corrections applied, is (for each signal category):

- $\ell\ell$: Data/MC $= 0.88 \pm 0.08(9\%)$
- $\pi\ell$: Data/MC $= 0.96 \pm 0.04(4\%)$
- $\pi\pi$: Data/MC $= 0.87 \pm 0.04(4\%)$
- ρ: Data/MC $= 0.89 \pm 0.02(3\%)$

where only the statistical uncertainties on the number of data events are reported. Since these uncertainties do not cover the 10% normalization discrepancies between data and simulation in the signal region, a conservative 15% systematic uncertainty is assigned to the combinatorial $B\overline{B}$ component. The peaking $B\overline{B}$ and the continuum $q\overline{q}$ components already have relative uncertainties above 10%, derived as described in sections 4.5.2 and 4.5.1.

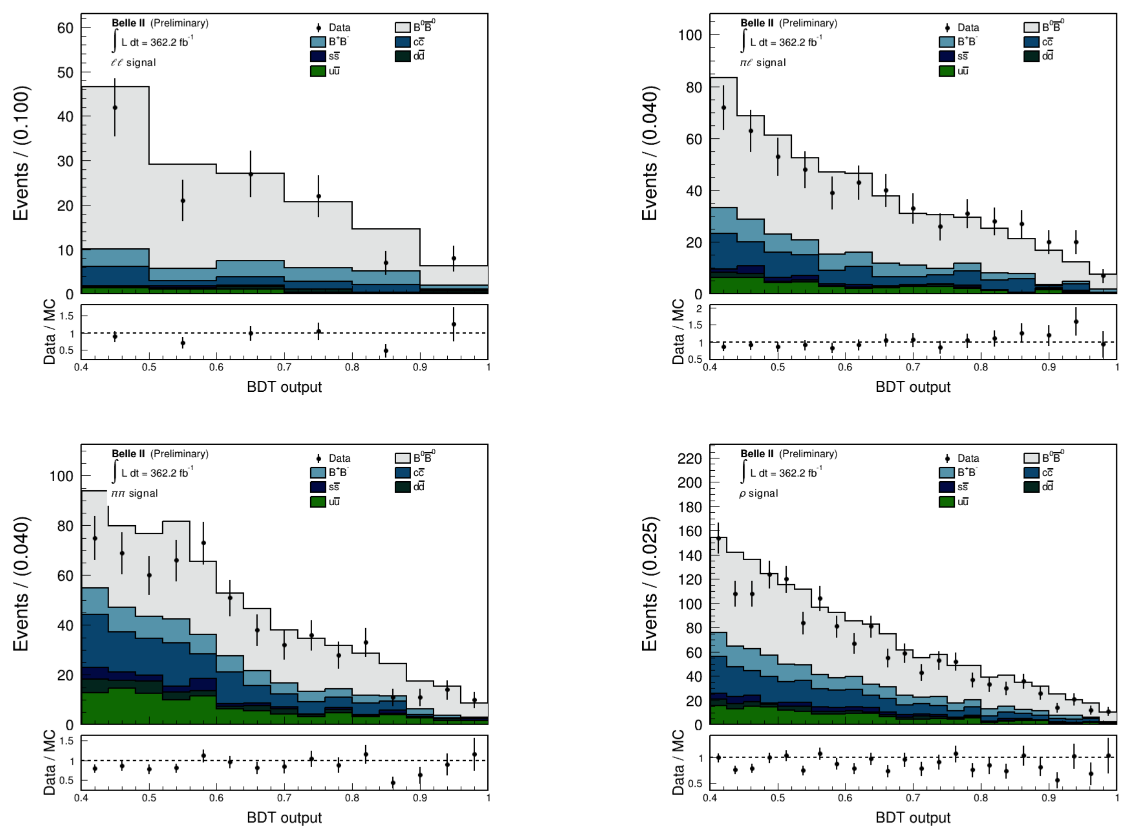


Figure 4.21: BDT output for all categories with the optimized signal region cuts: comparison between data and simulation in the same-flavor sideband. All available corrections are applied.

4.6 Signal extraction framework

The signal is extracted through a binned maximum likelihood fit performed in the signal regions of the BDT output, for all four signal categories simultaneously. An upper limit is then set scanning the CL_s built with the same likelihood used for signal extraction (section 3.3.1). The Pyhf and the Cabinetry libraries are exploited to build the binned likelihood model, perform the fit and compute the upper limit (see section 3.3.2).

Three separate components are considered:

- Signal $B^0 \to K^{*0}\tau^+\tau^-$.
- $B\overline{B}$ generic background, including B^+B^- and $B^0\overline{B}^0$ (both peaking and combinatorial).
- $q\overline{q}$ continuum background.

All the systematics and the MC statistical uncertainties are included as nuisance parameters in the likelihood.

The fit observable is the signal strength μ, which multiplies the signal yield N_{sig}, that can be written as:

$$N_{\mathrm{sig}} = 2\,\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)\, N_{B\overline{B}}\, f_{00/+-}\, \varepsilon_{\mathrm{sig}} \tag{4.8}$$

where $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$ is the signal branching ratio. $N_{B\overline{B}} = 387 \times 10^6$ is the number of produced $B\overline{B}$ pairs, estimated from a data-driven approach in which non-$B\overline{B}$ events are subtracted from on-resonance data ("B-counting"). While $f_{00/+-} = 0.484$ is the ratio between $\Upsilon(4S) \to B^0\overline{B}^0$ and $\Upsilon(4S) \to B^+B^-$ decays, taken from PDG. The signal efficiency $\varepsilon_{\mathrm{sig}}$ is the one expected in the signal region from simulation. In the fit framework the branching ratio is fixed to $\mathcal{B} = 10^{-3}$, so that the nominal N_{sig} is fixed as well to $N_{\mathrm{sig}} = 3.75 \cdot \varepsilon_{\mathrm{sig}} \times 10^5$.

The expected branching for a SM signal is $\mathcal{B} \simeq 10^{-7}$, then the nominal value for SM signal strength can be safely set to $\mu = 0$.

4.6.1 Systematic uncertainties

Main systematic uncertainties are included in the fit model as nuisance parameters. Sources may change depending on the event type (signal, $B\overline{B}$ or $q\overline{q}$) and on the signal category. They can be classified into two classes:

- Normalization systematics, i.e. overall variation of the yields.
- Shape systematics, i.e. variations of the histogram shapes, with full correlation across different BDT output bins.

In addition, the statistical uncertainty of the simulated MC sample is also included as a systematic, and it is modeled by independent bin by bin variations.

Overall background systematics

The yield of $B\overline{B}$ and signal events depends on the measured number of $B\overline{B}$ pairs $N_{B\overline{B}}$, which is taken as an external input and it is known with a precision of 1.45%, included as a normalization systematic for signal and all $B\overline{B}$ components, combined with the uncertainty of the $f_{00/+-}$ factor.

Continuum background For the continuum $q\overline{q}$ component a 2% normalization systematic is assigned to cover the uncertainty on the total integrated luminosity. On top of that, the systematic uncertainty of the overall background normalization, checked on the off-resonance control sample (section 4.5.1), is added:

- $\ell\ell$: 71%
- $\pi\ell$: 20%
- $\pi\pi$: 15%
- ρ: 15%

where the uncertainty for ρ category is conservatively increased from 11% to 15%, in order to fully cover the 0.82 ± 0.09 data/MC ratio in the off-resonance control sample. Such systematics are considered as fully uncorrelated for different signal categories, since they were computed independently for each category.

Combinatorial $B\overline{B}$ background For the combinatorial $B\overline{B}$ background events, a conservative 15% uncertainty is taken as a systematic for their correction weights, derived from the same-flavor sample. This uncertainty covers the differences observed between same-flavor data and simulation for the total background yield (see section 4.5.3). Since the combinatorial corrections were derived and validated independently for each signal category, the corresponding systematic is treated as fully uncorrelated among different categories.

In the signal region, the fractions of simulated $B\overline{B}$ events which are classified as combinatorial are 42.2%, 51.9%, 58.1%, and 55.5% for $\ell\ell$, $\pi\ell$, $\pi\pi$ and ρ signal categories, respectively. Since the $B\overline{B}$ are fitted as a unique component, these combinatorial systematics result in a shape variation of the full $B\overline{B}$ histogram.

Signal and peaking $B^0\overline{B}^0$ background The main systematics for the signal and the peaking $B^0\overline{B}^0$ events, come from the embedded data/MC ratio uncertainties. The correction, computed in the signal region, has a 12% relative uncertainty and it is consistent with the value obtained after the pre-selection (sections 4.3.2 and 4.5.2). This 12% is taken as a systematic uncertainty, and it is treated as fully correlated among signal and peaking $B^0\overline{B}^0$, and across all the signal categories.

The embedded uncertainty affects only the normalization of the signal component, while for the $B\overline{B}$, being split into combinatorial and peaking component, the effect is a variation of the histogram shape.

Particle-ID systematics

Hadron identification The error on the hadron-ID correction weights combines the statistical and systematic uncertainty, derived from the control channels. To evaluate the effect on the histograms in the signal region, all the weights are simultaneously varied up and down, and the corresponding variations are taken as systematic uncertainty from the hadron-ID corrections, affecting all the components in the fit.

Lepton identification Similarly to hadron-ID, the weights applied to correct for the data/MC differences in the lepton identification, are associated to a statistical and systematic uncertainty. The squared sum of the two is taken as the total uncertainty. A fully correlated scenario is assumed: the total uncertainty is either added or subtracted simultaneously to all the weights. A lepton-ID systematic is assigned considering the variation of the BDT output distribution within these two scenarios.

This systematic is taken into account for all the components in the $\ell\ell$, $\pi\ell$ and ρ signal categories. In the final fit region, for the $\ell\ell$ category, the lepton-ID corrections lead to 1.6% uncertainty on the signal yield and 2.8% on the $B\overline{B}$ background.

π^0 efficiency The efficiency corrections for π^0 have a statistical and systematic uncertainties associates, as reported on table 4.5. They are combined together and the total uncertainty is taken as variation for each weight, assuming again a fully correlated scenario.

For this analysis, in the signal extraction region, the ρ category has overall 6.2% and 6.4% uncertainties, for signal and $B\overline{B}$ background respectively, due to the π^0 efficiency corrections.

Extra photons multiplicity (E_{ECL}^{extra})

The extra photon multiplicity corrections defined in section 4.3 from the same-flavor control sample have a statistical uncertainty given by the number of data events in each bin of figure 4.7.

The correction on the extra photon multiplicity is derived from the same-flavor sample and validated on the BDT$<$ 0.4 sideband of the signal region. This sideband, for each signal category, is expected to contain about 40% of the signal events and approximately 99% of the background events, with respect to the pre-selection stage (see table 4.7).

The photon multiplicity in this region is shown in figure 4.22 for each signal category, with the total data events normalized to the total MC yields on each category. The residual discrepancies in the data/MC ratio give an estimate of the systematic uncertainty (hereafter

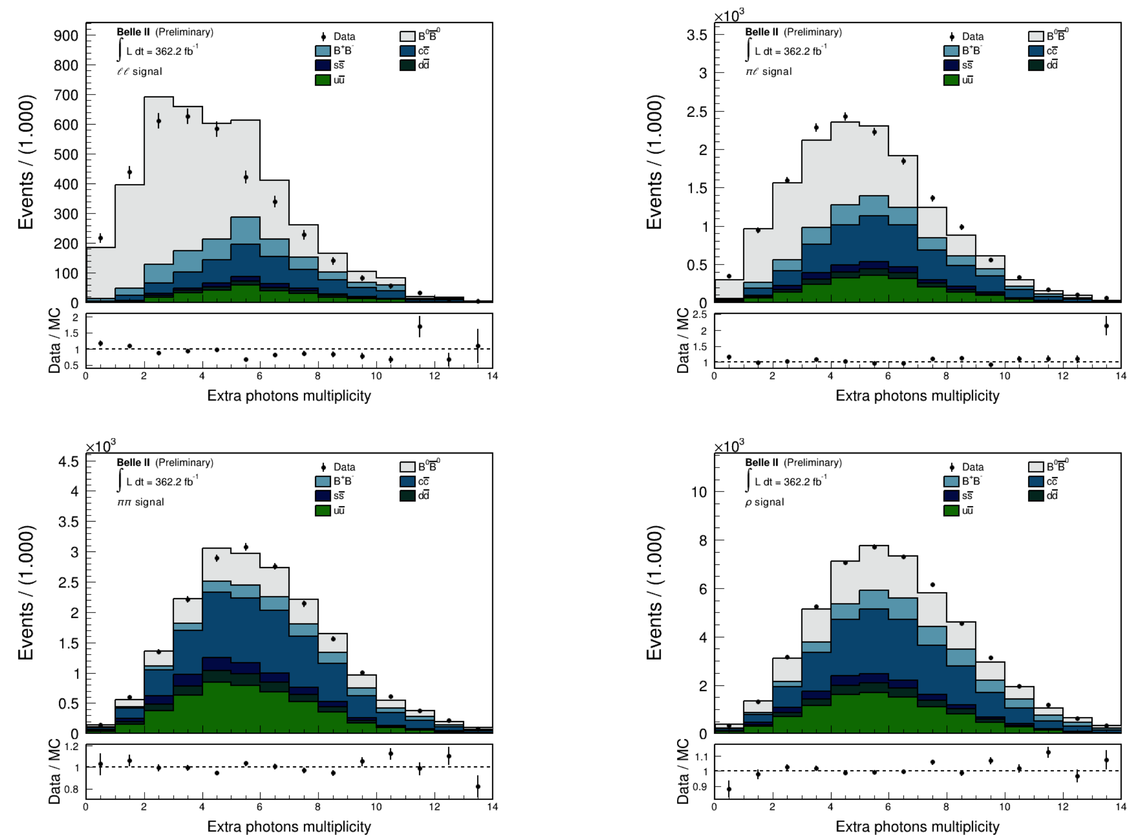


Figure 4.22: Extra photons multiplicity associated to E_{ECL}^{extra} for the four signal categories in the BDT<0.4 sideband. All the MC corrections discussed in section 4.3, included the ones on extra photons, are applied.

called E_{ECL}^{extra} systematic) coming from the background compositions. The module of the data/MC ratio is summed in quadrature, for each bin of photon multiplicity, with the statistical uncertainty from the same-flavor sideband.

Total systematics

Figure 4.23 collects all the modifiers entering the fit. Shape variations are split in a baseline overall variation of the normalization ("normsys") and in a pure shape variation of the histogram ("histosys").

The normalization variations which modify either the $q\bar{q}$ component only, or the $B\overline{B}$ only, cause actually a shape variation of the total background. As shown in Figure 4.24, the overall shape variations due to all the systematics included in the fit model can be compared

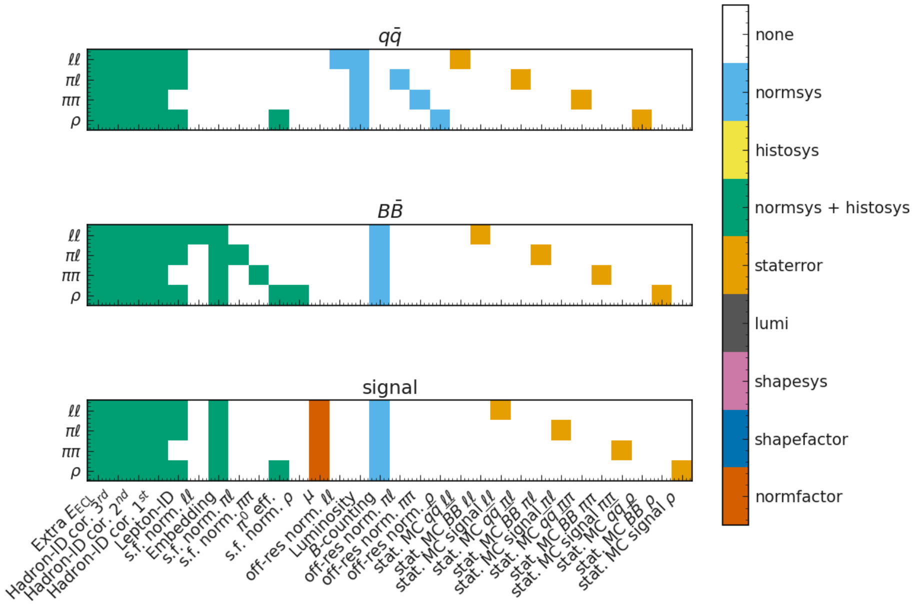


Figure 4.23: Scheme with all the modifiers of the model used for the fit, split for event types and signal categories.

with the data/MC shape variation computed on the same-flavor control sample. Both variations are normalized to the first bin. These plots show that the shape disagreement observed in the control sample is covered by the systematics included in the fit model, in particular for the $\ell\ell$, $\pi\ell$ and $\pi\pi$ categories, while a worse agreement is observed for the ρ category, which has the highest number of bins.

4.6.2 Fit validation

Asimov fit

The so-called Asimov fit is a method to test the consistency of the results by fitting the simulated histograms as if they were data [63]. An Asimov fit with $\mu = 0$, i.e. background-only hypothesis, is performed in order to test the fitting tool. Figure 4.25 shows the pre-fit distribution of the BDT output for each signal category. The total uncertainties (MC statistic and systematics) of the model are drawn.

As expected, the fit on the Asimov dataset gives zero pulls for all the nuisance parameters (see figure 4.26, left). The best-fit value for the parameter of interest is $\mu = (0.0 \pm 1.11)$, while the expected upper limit with 90% confidence level is $\mu < 2.12$ (figure 4.26, right), equivalent to $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2.12 \times 10^{-3}$.

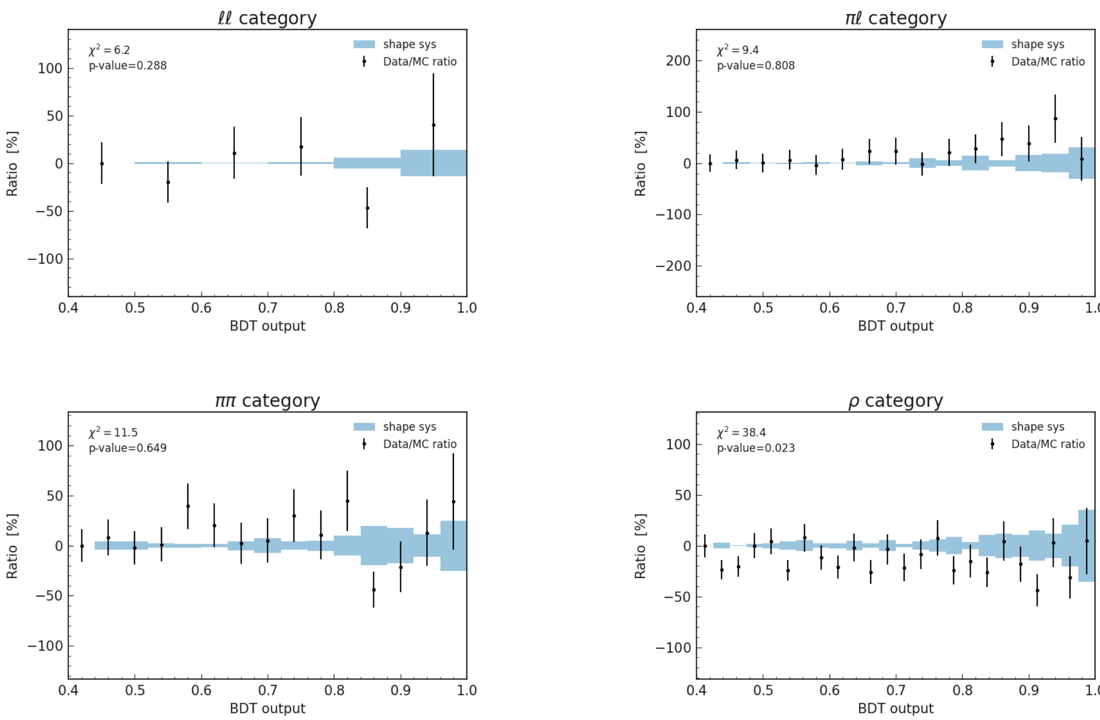


Figure 4.24: Comparison between the data/MC shape variations in the same-flavor control sample and the shape variation of the total yields, as determined by the systematic uncertainties included in the fit model.

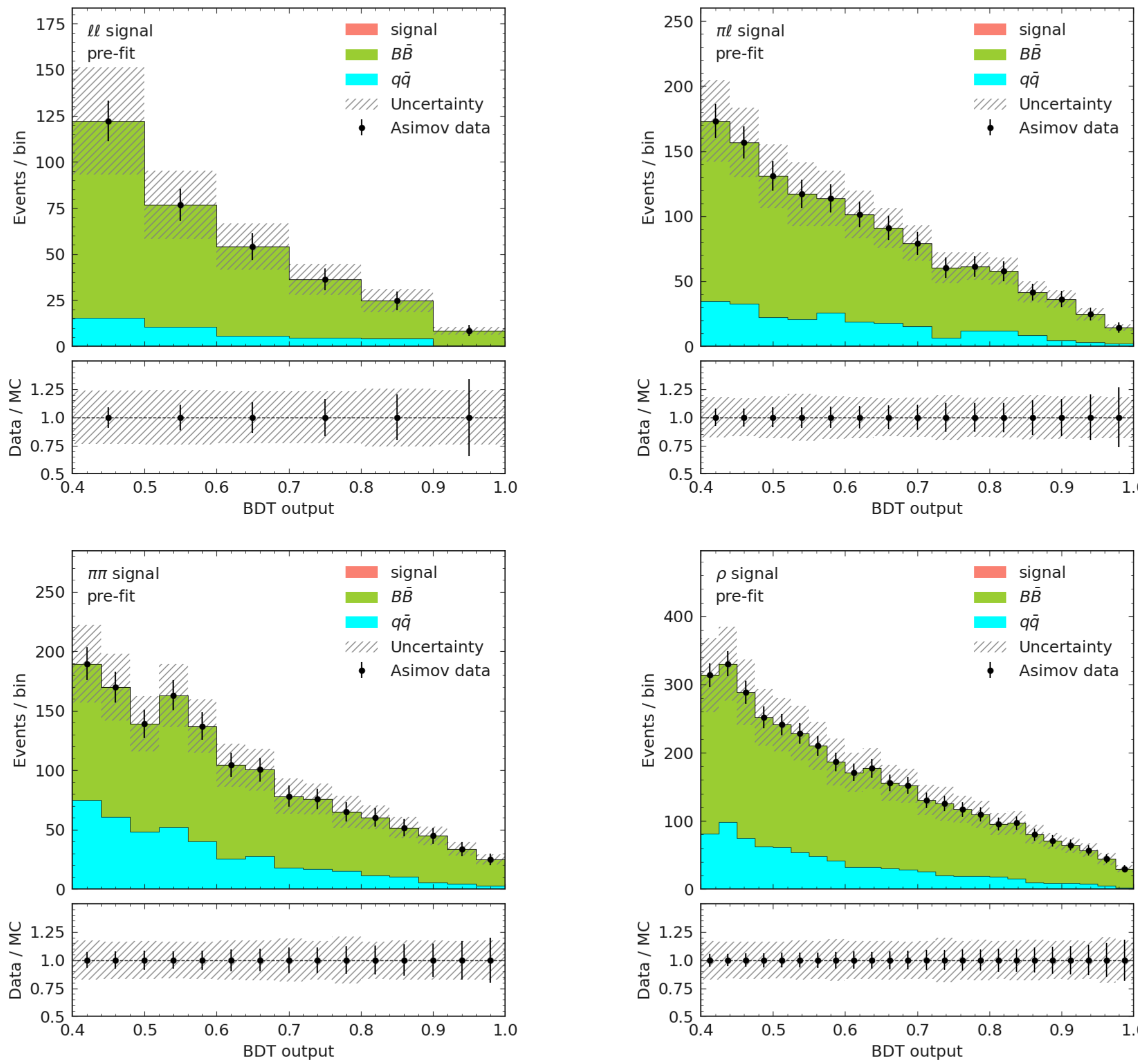


Figure 4.25: Pre-fit distribution of the BDT output in the signal region with background-only hypothesis ($\mu = 0$) and corresponding Asimov data.

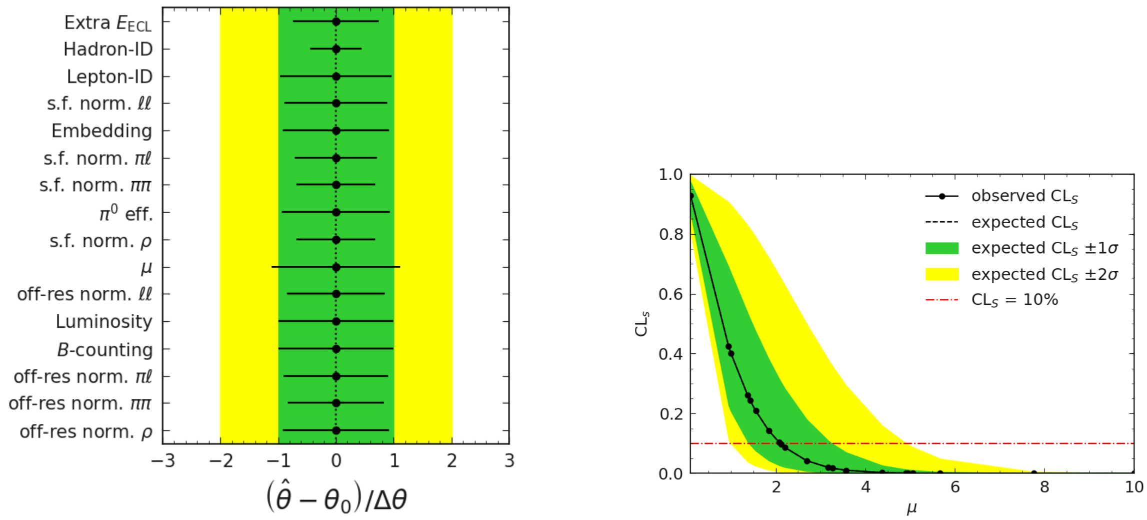


Figure 4.26: Post-fit pulls and CL_s scan from the Asimov fit with background-only hypothesis.

MC toys

To validate the signal extraction, 10 k MC toys are randomly generated and then fitted. For every toy, the events in each bin of the signal region are sampled from a Poisson distribution with mean value equal to the expected yields in that bin. Four hypothesis of signal strength are tested: $\mu = 0$, 1, 2, 3.

Figure 4.27 shows the distributions of fitted μ for each signal hypothesis. Since the most sensitive bins have a small number of expected events, the profile likelihood is non Gaussian and therefore the pull distribution of the MC toys will slightly differ from a Normal distribution, as shown in figure 4.28 for the three different values of injected μ. Hence, the compatibility between the injected μ and the fitted one is checked through the median and the central 68% quantile of the fitted μ distribution. Such values are compared to the profile likelihood scans done via an Asimov fit in the same four signal hypothesis (figure 4.28).

Ranking systematics

The MC toys generated for the $\mu = 0$ hypothesis can be used to separate the distinct contributions of the nuisance parameters to the total systematic uncertainty. First of all, the statistical uncertainty is evaluated by fitting the toys with all the nuisance parameters fixed. The 68% quantile on the distribution of the parameter of interest μ represents the statistical uncertainty, being $\sigma_{\mathrm{stat}} = 0.66$. Then, for each nuisance parameter θ, the fit on the toy samples is repeated keeping all the other nuisance parameters fixed. In this case, the 68% quantile in the distribution of the fitted μ represents the combination

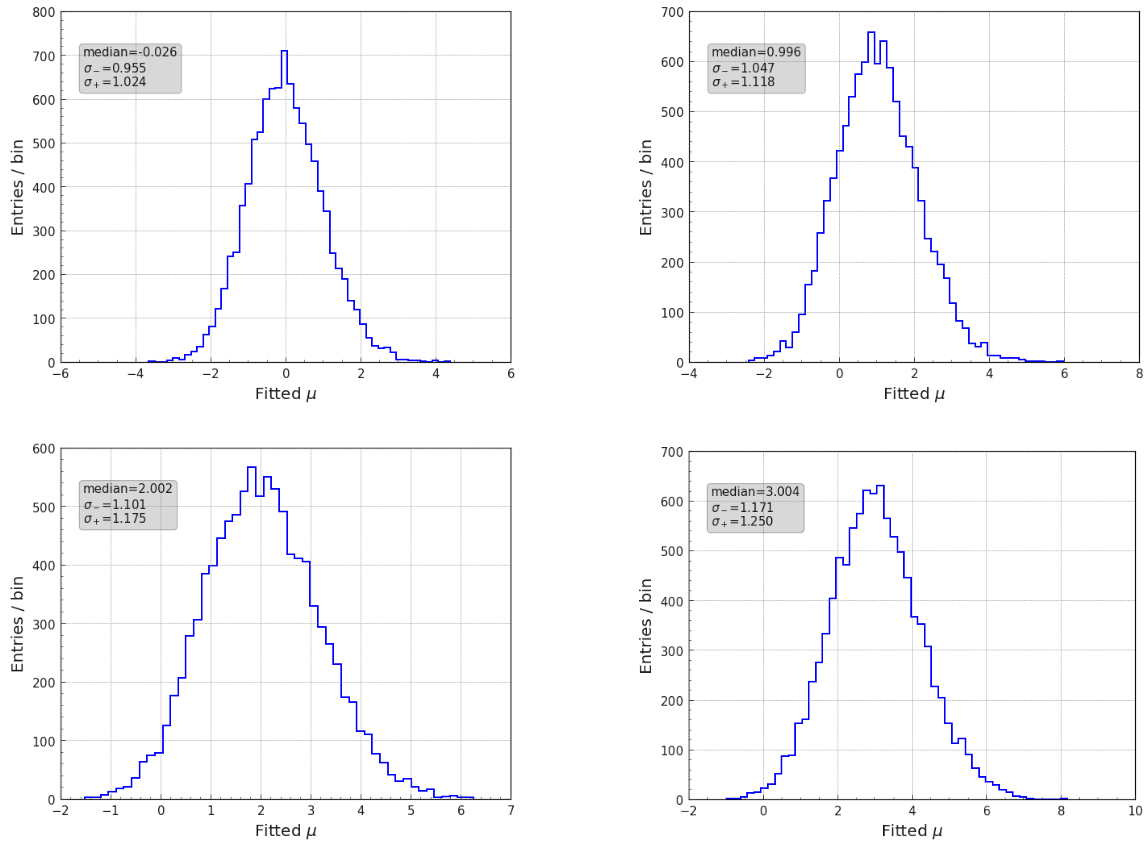


Figure 4.27: Distribution of the fitted μ on the MC toys for four signal strength hypothesis: $\mu = 0$ (top left), $\mu = 1$ (top right), $\mu = 2$ (bottom left), $\mu = 3$ (bottom right). The median and the σ_+, σ_-, estimated with the 68% central quantile, are reported.

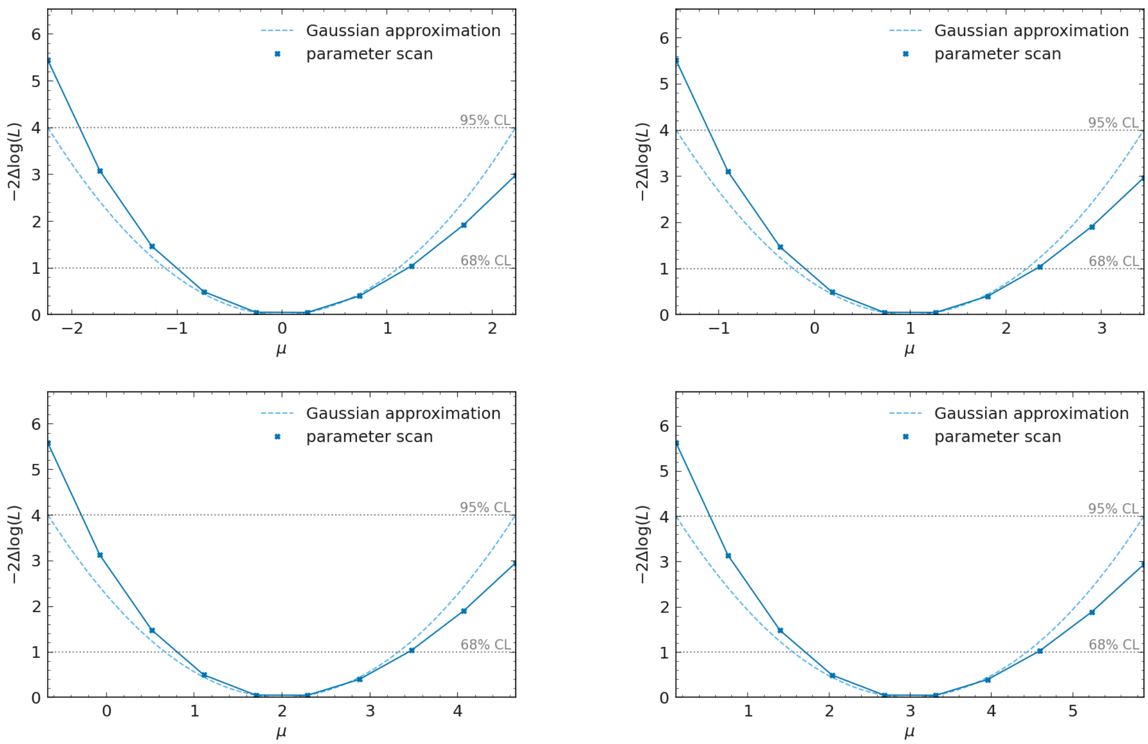


Figure 4.28: Profile likelihood scan in the fitted parameter μ on the Asimov dataset for four different signal hypothesis: $\mu = 0$ (top left), $\mu = 1$ (top right), $\mu = 2$ (bottom left), $\mu = 3$ (bottom right). The Gaussian profile likelihood approximation is drawn as a reference.

Table 4.12: Uncertainty on the fitted μ for each of the nuisance parameters included in the fit, computed from 10 k MC toys with the $\mu = 0$ hypothesis.

Source	Uncertainty on fitted μ
E_{ECL}^{extra} corrections	0.73
Embedding correction	0.59
Hadron-ID	0.58
Same-flavor norm. ρ	0.29
Same-flavor norm. $\pi\ell$	0.24
π^0 efficiency	0.23
B-counting	0.21
Lepton-ID	0.19
Same-flavor norm. $\ell\ell$	0.18
Off-resonance norm. $\ell\ell$	0.17
Off-resonance norm. ρ	0.17
Off-resonance norm. $\pi\ell$	0.17
Same-flavor norm. $\pi\pi$	0.13
Off-resonance norm. $\pi\pi$	0.08
Luminosity	0.05
MC statistics	0.04
Total systematics (including correlations)	0.74
Statistical uncertainty	0.66

of statistical uncertainty and of the systematic uncertainty $\sigma_{\text{syst}}(\theta)$, stemming from the nuisance parameter θ. Statistical and systematic uncertainties can be treated as uncorrelated, hence the systematic contribution is determined as:

$$\sigma_{\text{syst}}(\theta) = \sqrt{\sigma_{\text{tot}}^2(\theta) - \sigma_{\text{stat}}^2} \tag{4.9}$$

The systematics uncertainties evaluated with this approach are listed in table 4.12. To compute the total systematic uncertainty, the toys are fitted letting free all the parameters. The statistical uncertainty is subtracted from the 68% quantile (top left plot of Figure 4.27), obtaining $\sigma_{\text{syst}} = 0.74$. This value is lower with respect to what results from a naive combination of all the single systematic sources listed in table 4.12. Indeed, the assumption of uncorrelated systematics is not satisfied, as shown by the correlation table of Figure 4.29, computed for the Asimov $\mu = 0$ fit.

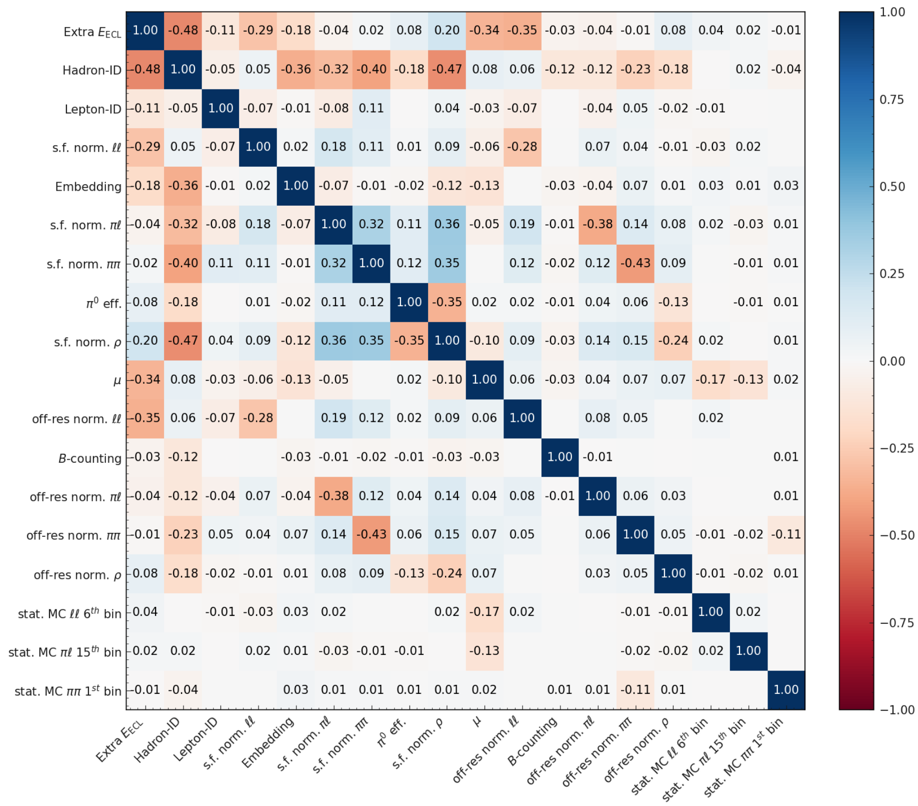


Figure 4.29: Correlation table for the parameters in the fit model, computed on the $\mu = 0$ Asimov fit. Only parameters with at least one correlation coefficient > 10% are shown in the table.

4.6.3 Unblinding strategy

Before the full un-blinding of the signal region, several cross-checks can be performed on the control samples and on subsets of the signal region itself. The following cross-checks will be performed on the control samples and on subsets of the signal region itself:

1. Fit to the same-flavor sideband data, with expected yields shown in table 4.11.
2. Blind fit on half-split signal region datasets, where only the uncertainty and the μ differences between the two split samples are compared. Possible criteria to divide the entire dataset are: random split, split based on data-taking period, split based on a pre-selection variable (e.g. polar angle of the missing momentum).
3. Partial fit on the signal region restricted to a specific signal category. Starting from the less sensitive ones (ρ and $\pi\pi$ categories).
4. Partial fit on the signal region restricted to the lowest sensitive BDT bins with all signal categories.
5. Full fit to the signal region keeping all results blinded except for the goodness of fit and the uncertainty on the parameter of interest μ.

5 Conclusion

The search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay at the Belle II experiment has been documented throughout this work. This decay channel, with all the possible combinations of $b \to s\ \tau^+\tau^-$ transitions, is attracting a rising interest from a theoretical perspective, since many new physics model, explaining the observed anomalies in the b flavor sector, would also predict a strong enhancement in this FCNC transitions.

The main challenges in measuring the $b \to s\ \tau^+\tau^-$ modes stem from the missing neutrinos in the final states, whose multiplicity depends on the decay modes of the $\tau^+\tau^-$ system.

This analysis exploits the novel FEI algorithm, developed for the Belle II experiment, to fully reconstruct a B meson in a hadronic decay. The partner B is then reconstructed looking for a signal $B^0 \to K^{*0}\tau^+\tau^-$ decay. Ten different combinations of $\tau^+\tau^-$ final states are exclusively reconstructed, including the main 1-prong decay modes: $\tau \to \ell\nu\overline{\nu}$, $\tau \to \pi\nu$ and $\tau \to \rho\nu$, covering about 51% of the total final states. The K^{*0} is also exclusively reconstructed as $K^{*0} \to K^+\pi^-$.

After the event reconstruction, the selected candidates are divided into four categories depending on the $\tau^+\tau^-$ decay modes. Each category has different final state particles, neutrino multiplicity and dominant background component. Therefore, the signal-to-background discrimination is optimized separately for each of the $\tau^+\tau^-$ categories.

The separation between signal and background events is performed by training a BDT binary classifier on simulated events. The signal branching ratio is then extracted through a maximum binned likelihood fit directly on the classifier output, simultaneously on the four $\tau^+\tau^-$ categories and with all the systematic uncertainties included as nuisance parameters in the fit model. The fit is restricted in a signal region of the BDT output.

Prior to the BDT training, the simulated events are validated and corrected through a data driven approach, exploiting several control channels. In particular, the off-resonance, the same-flavor, the embedded control samples and the BDT<0.4 side-band, were built to estimate correction factors targeting the specific selection of this analysis.

The dominant systematics in the signal extraction is the corrections of extra photon multiplicity, computed on the same-flavor sample and validated on the BDT signal region side-band. This correction is needed to account for the mis-modeling in the simulated extra photon multiplicity, affecting the important variable E_{ECL}^{extra}. This, as well as other missing energy analysis, will then strongly benefit from an improvement of the ECL clustering simulation.

Other leading systematics are uncertainty on the signal efficiency correction derived from the embedded control sample, the hadron-ID corrections and the limited statistics of the MC samples, which are expected to be reduced with more collected data.

The expected upper limit from Asimov fit, assuming the current 362.2 fb^{-1} of data recorded by the Belle II detector, is determined to be $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2.7 \times 10^{-3}$ at 90% confidence level. This should be compared with the world-leading upper limit $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 3.1 \times 10^{-3}$, at 90% confidence level, set by Belle using 711 fb^{-1} of data, about a factor $\times 2$ of the statistics exploited in the present analysis [14].

A Additional plots

A.1 BDT features

Figures from A.1 to A.14 show the simulated signal and background distributions for the variables used as input features in the BDT training, described in details in section 4.4.

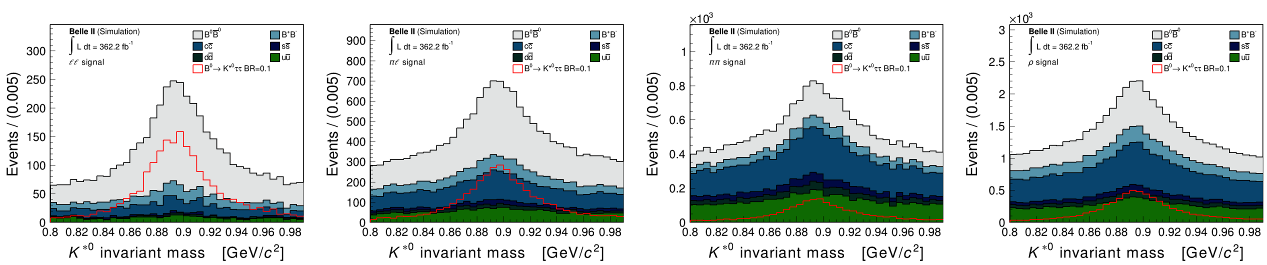


Figure A.1: Distribution of the invariant mass of reconstructed K^{*0}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

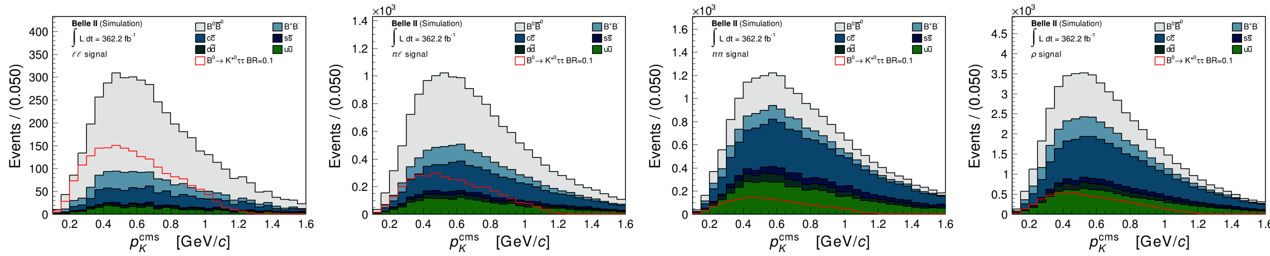


Figure A.2: Distribution of the charged K (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

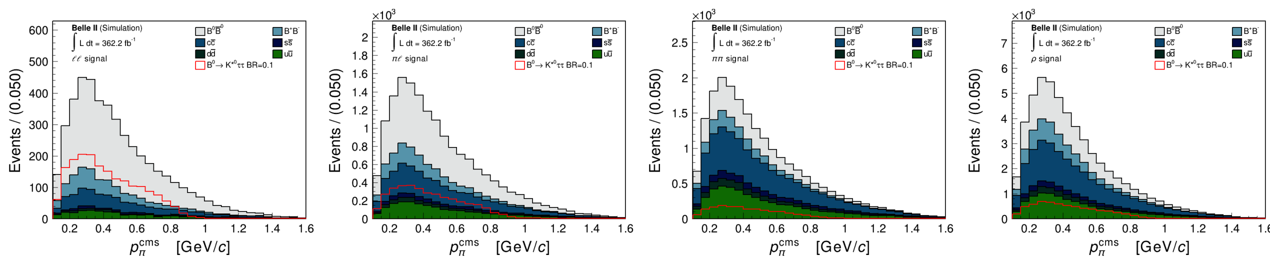


Figure A.3: Distribution of the charged π (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

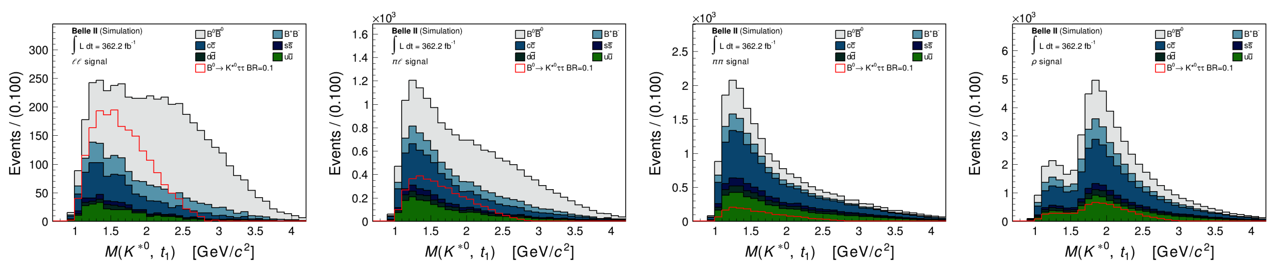


Figure A.4: Distribution of the invariant mass of the K^{*0} and the τ daughter t_1 (same sign of the K from K^{*0}), used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

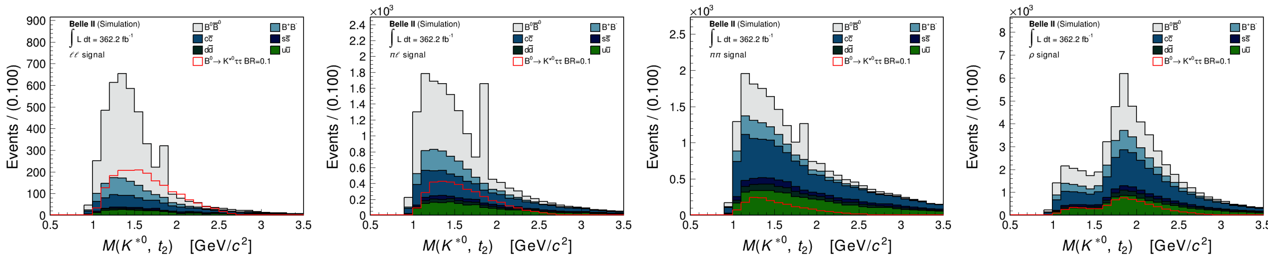


Figure A.5: Distribution of the invariant mass of the K^{*0} and the τ daughter t_2 (having opposite sign with respect to the K from K^{*0}), used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

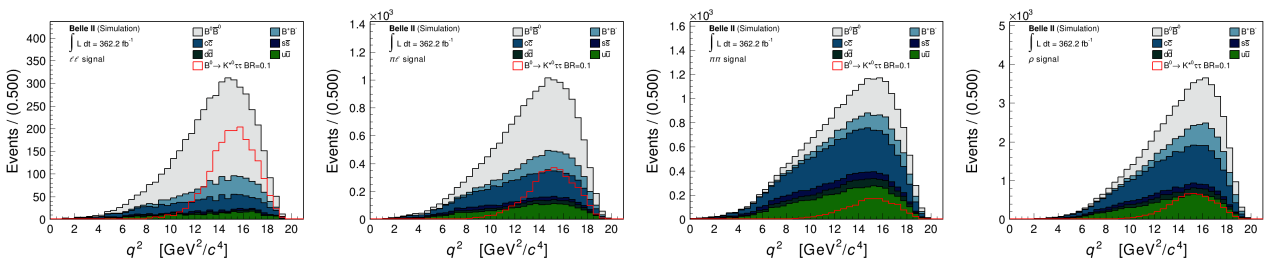


Figure A.6: Distribution of the signal side q^2, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

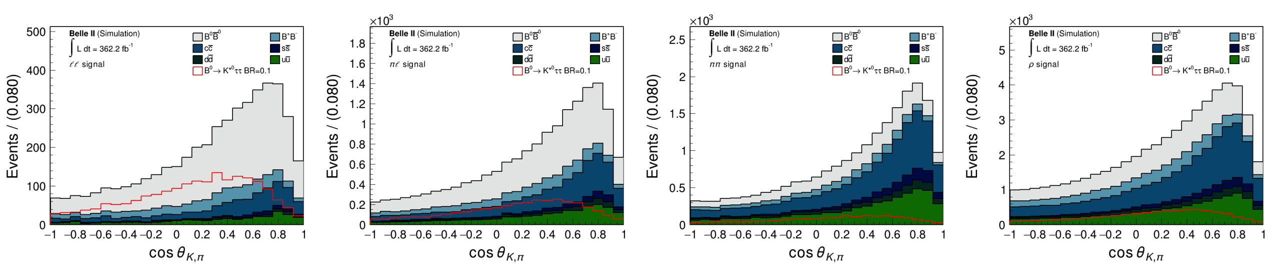


Figure A.7: Distribution of the $\cos\theta_{K;\pi}$ of the angle between the K^{*0} daughters, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

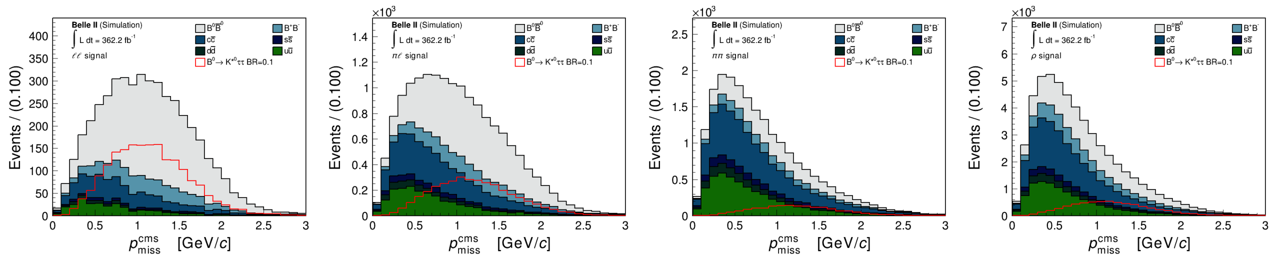


Figure A.8: Distribution of the missing momentum of the event in the CM frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

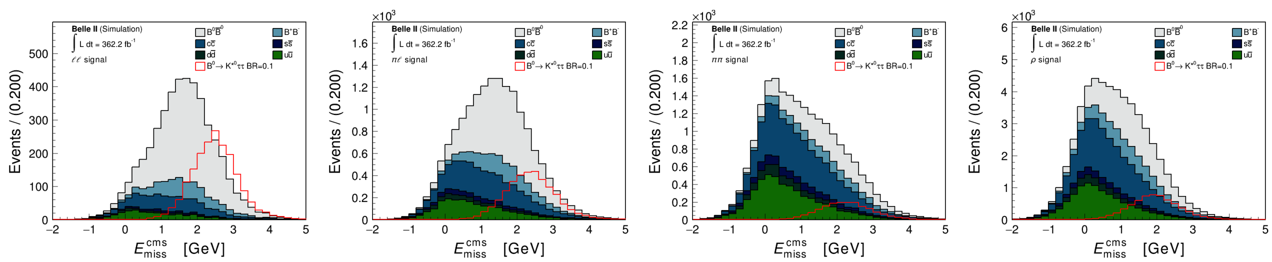


Figure A.9: Distribution of the missing energy of the event in the CM frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

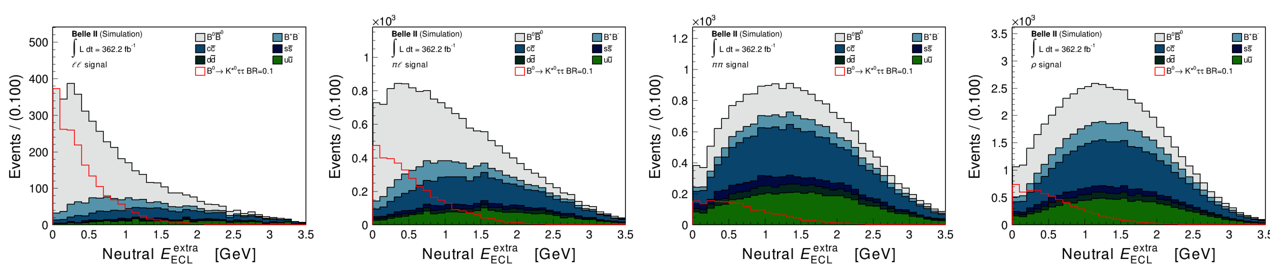


Figure A.10: Distribution of the neutral extra energy in the event, as defined by the mask in section 4.2, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

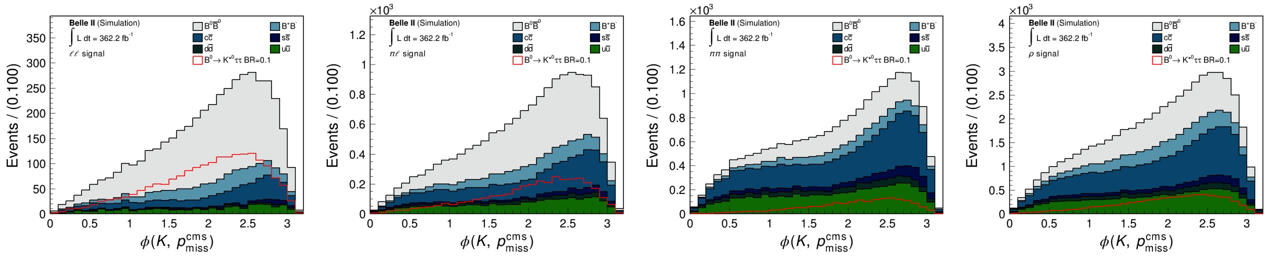


Figure A.11: Distribution of the angle φ between the K^{*0} and the missing momentum directions, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

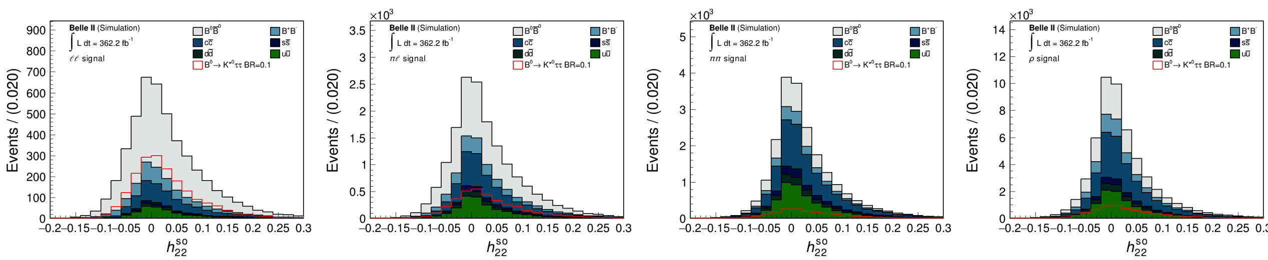


Figure A.12: Distribution of the KSFW variable h_{22}^{so}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

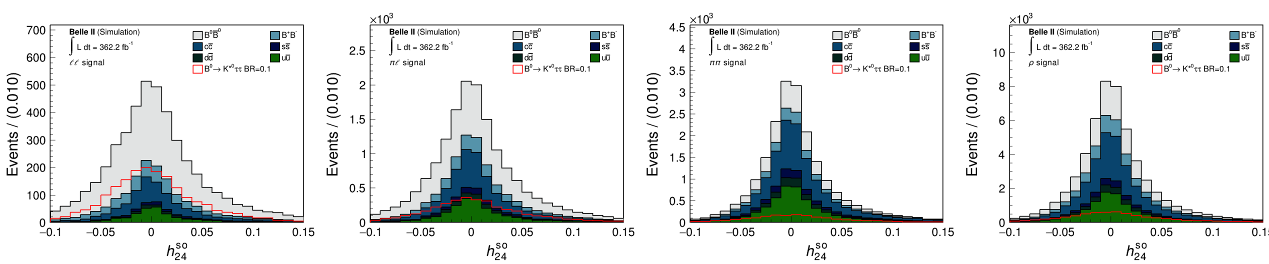


Figure A.13: Distribution of the KSFW variable h_{22}^{so}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

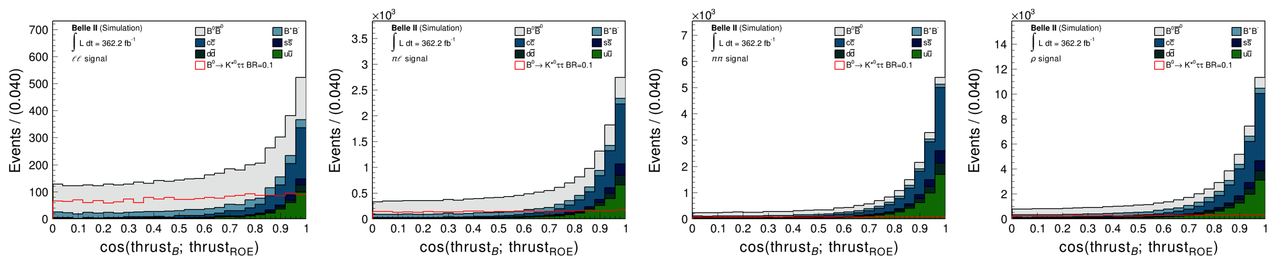


Figure A.14: Distribution of the continuum suppression variable $\cos\theta(T_B; T_{\mathrm{ROE}})$, where θ is the angle between the B_{sgn} axis and the thrust axis. It is used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied. The red histogram is the simulated signal distribution, scaled to an equivalent branching ratio $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = 0.1$.

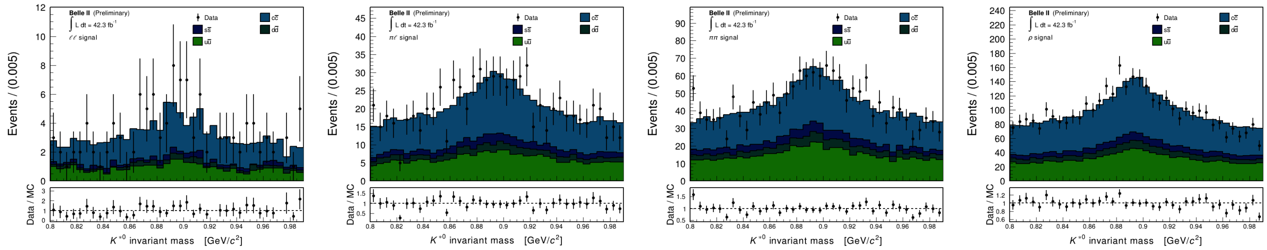


Figure A.15: Comparison between simulated $q\overline{q}$ and off-resonance data for the invariant mass of reconstructed K^{*0}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

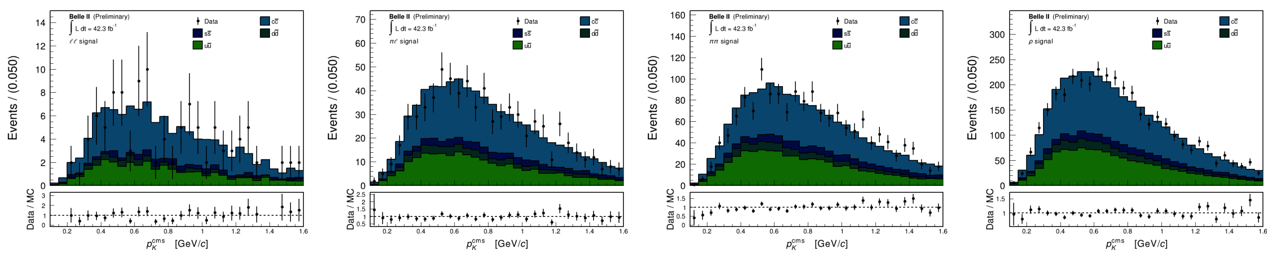


Figure A.16: Comparison between simulated $q\overline{q}$ and off-resonance data for the charged K (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

A.2 Continuum validation plots

Figures from A.15 to A.28 show the data/MC comparison for the off-resonance control sample against the simulated $e^+e^- \to q\overline{q}$ continuum background, as mentioned in section 4.5.1.

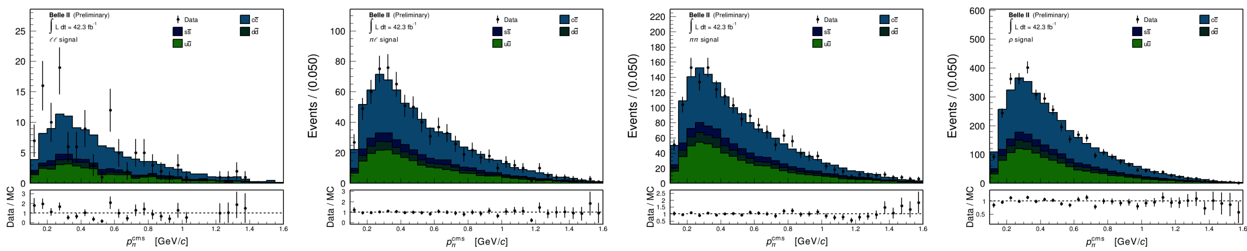


Figure A.17: Comparison between simulated $q\overline{q}$ and off-resonance data for the charged π (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

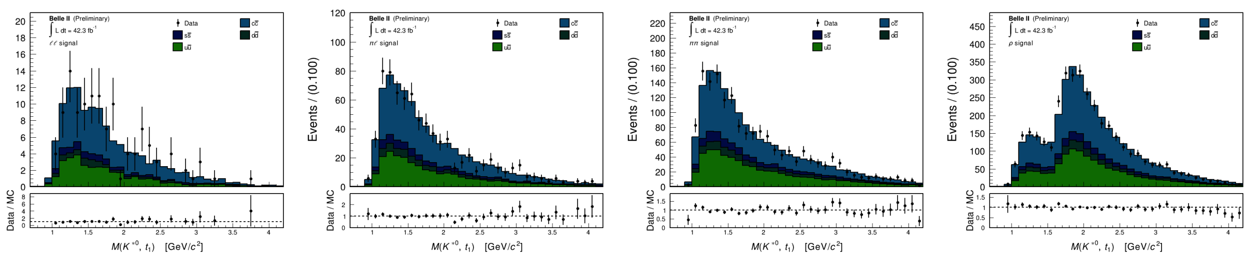


Figure A.18: Comparison between simulated $q\overline{q}$ and off-resonance data for the invariant mass of the K^{*0} and the daughter of the first τ $t1$, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

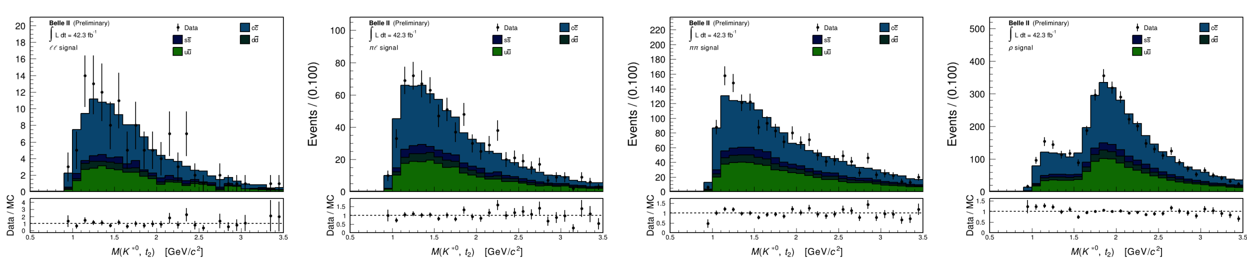


Figure A.19: Comparison between simulated $q\overline{q}$ and off-resonance data for the invariant mass of the K^{*0} and the daughter of the second τ $t2$, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

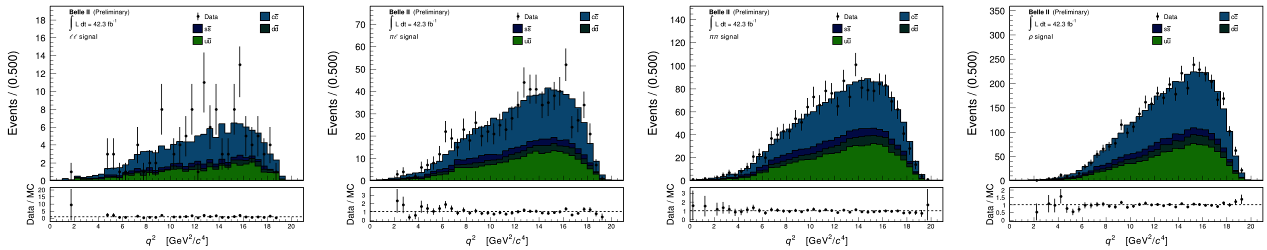


Figure A.20: Comparison between simulated $q\overline{q}$ and off-resonance data for the signal side q^2, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

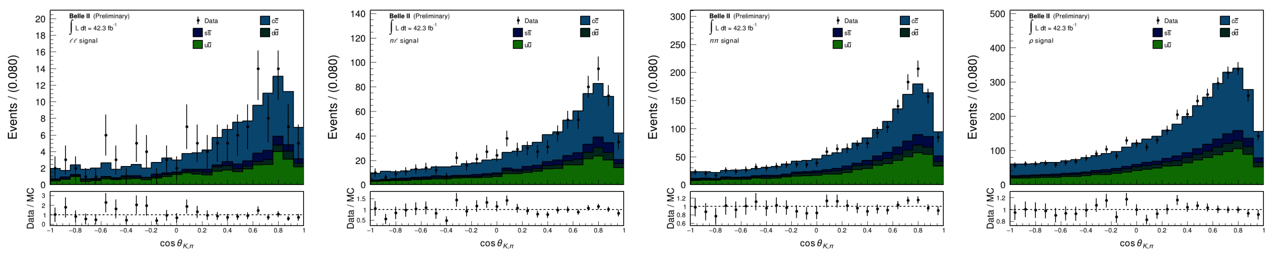


Figure A.21: Comparison between simulated $q\overline{q}$ and off-resonance data for the $\cos\theta_{K;\pi}$ of the angle between the K^{*0} daughters, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

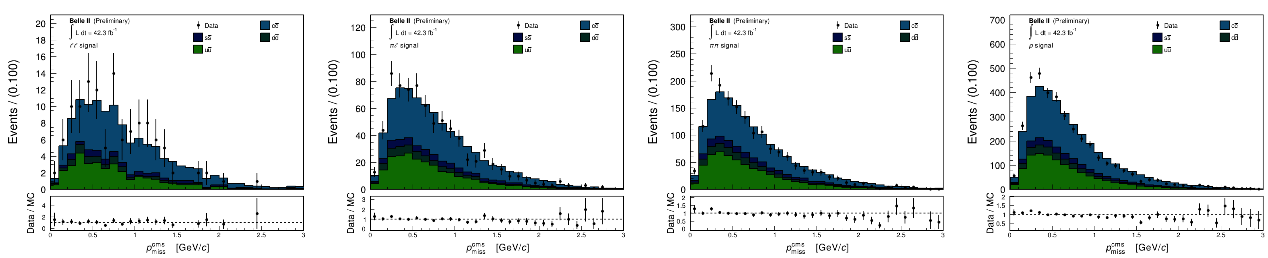


Figure A.22: Comparison between simulated $q\overline{q}$ and off-resonance data for the missing momentum of the event in the CM frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

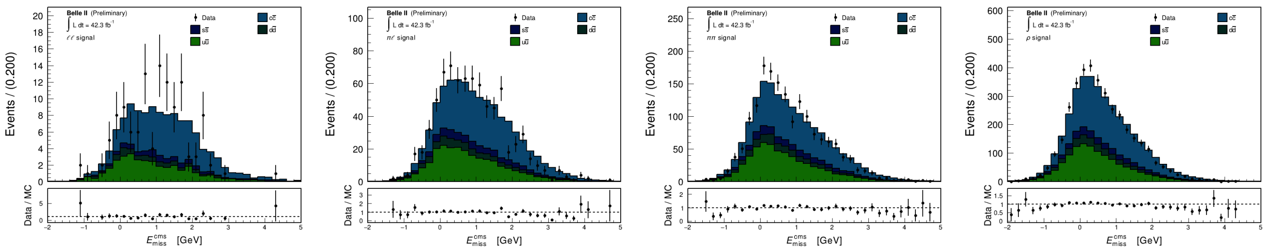


Figure A.23: Comparison between simulated $q\overline{q}$ and off-resonance data for the missing energy of the event in the CM frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

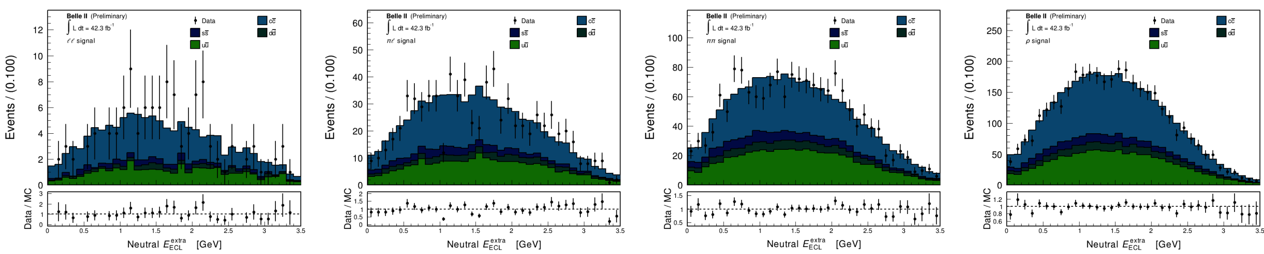


Figure A.24: Comparison between simulated $q\overline{q}$ and off-resonance data for the neutral extra energy in the event, as defined by the mask in section 4.2, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

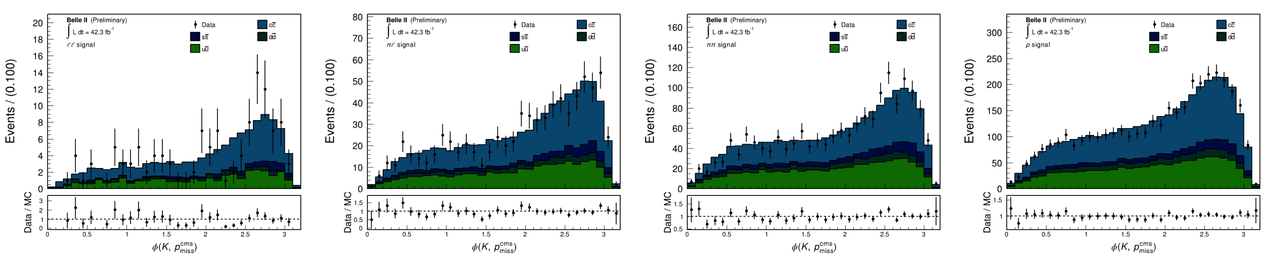


Figure A.25: Comparison between simulated $q\overline{q}$ and off-resonance data for the angle φ between the K^{*0} and the missing momentum directions, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

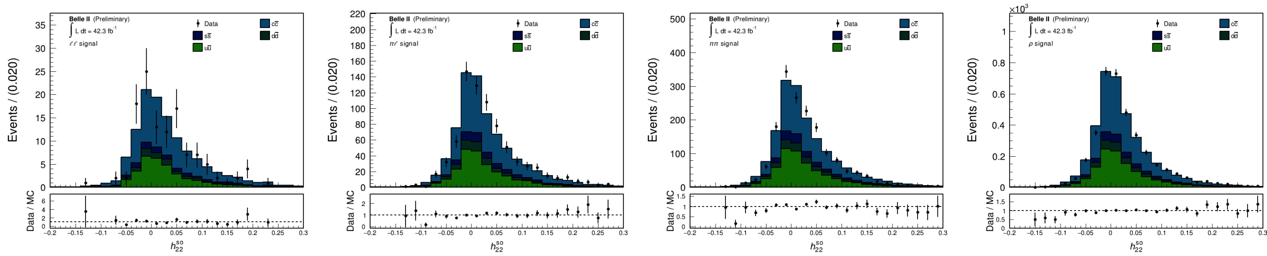


Figure A.26: Comparison between simulated $q\bar{q}$ and off-resonance data for the KSFW variable h_{22}^{so}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

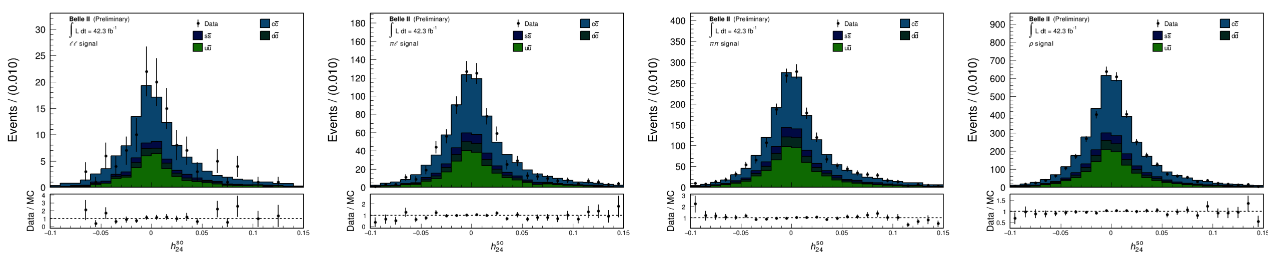


Figure A.27: Comparison between simulated $q\bar{q}$ and off-resonance data for the KSFW variable h_{22}^{so}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

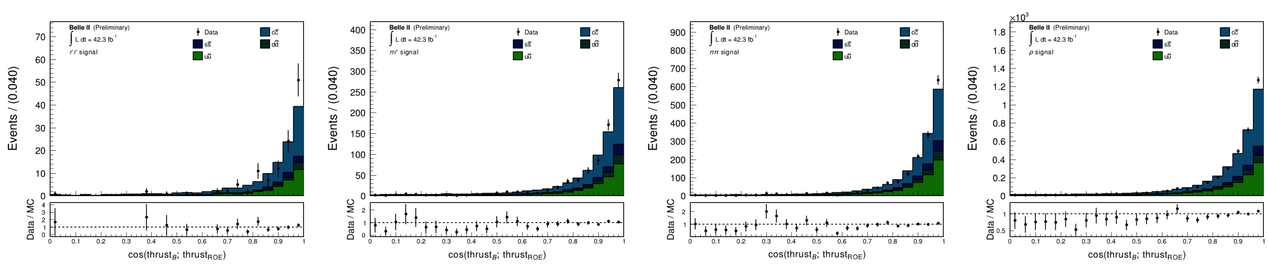


Figure A.28: Comparison between simulated $q\bar{q}$ and off-resonance data for the continuum suppression variable $\cos\theta(T_B; T_{\mathrm{ROE}})$, where θ is the angle between the B_{sgn} axis and the thrust axis. It is used as input features in the BDT training for all the signal categories. Simulated backgrounds are re-scaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

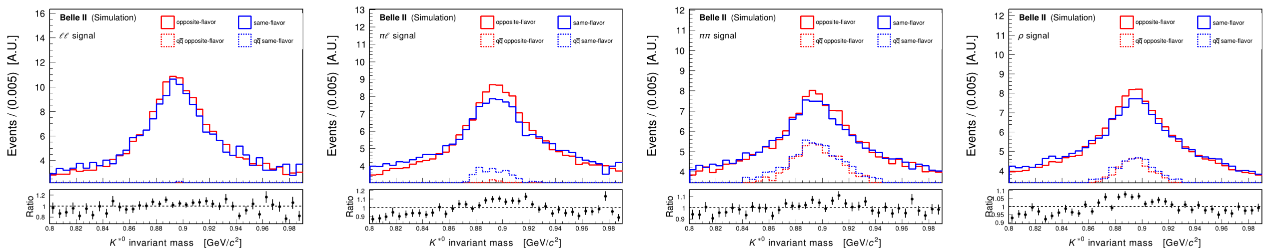


Figure A.29: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the invariant mass of reconstructed K^{*0}, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

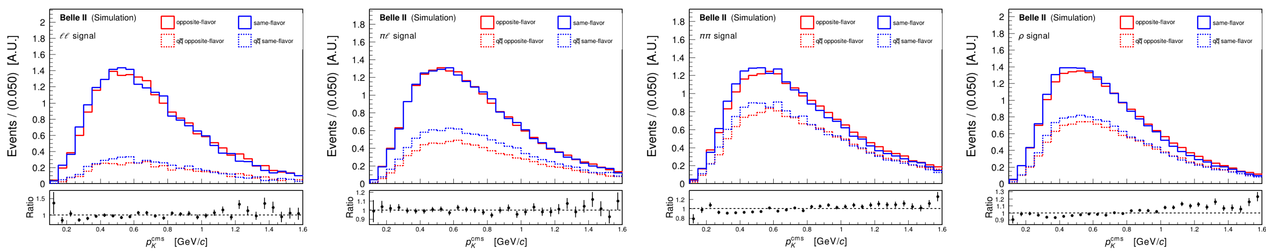


Figure A.30: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the charged K (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

A.3 Same-flavor validation plots

Figures from A.29 to A.42 show the comparisons between the simulated backgrounds after the pre-selection and the same backgrounds for the same-flavor control sample, as described in section 4.5.3. All the variables exploited in the BDT training and all the signal categories are shown.

Figures from A.43 to A.56 show the data/MC comparison form the same-flavor sideband described in section 4.5.3.

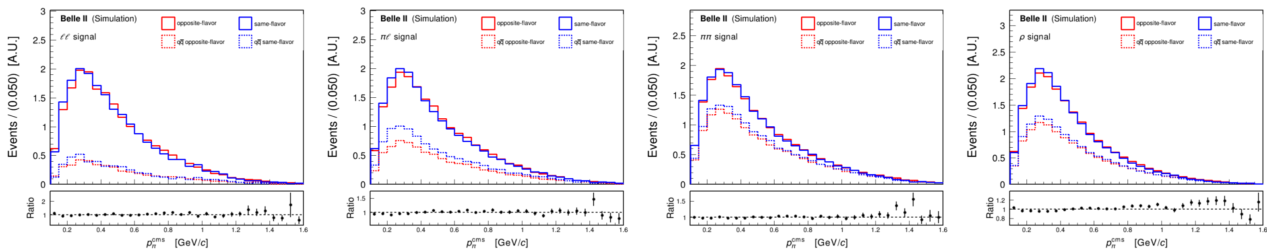


Figure A.31: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the charged π (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

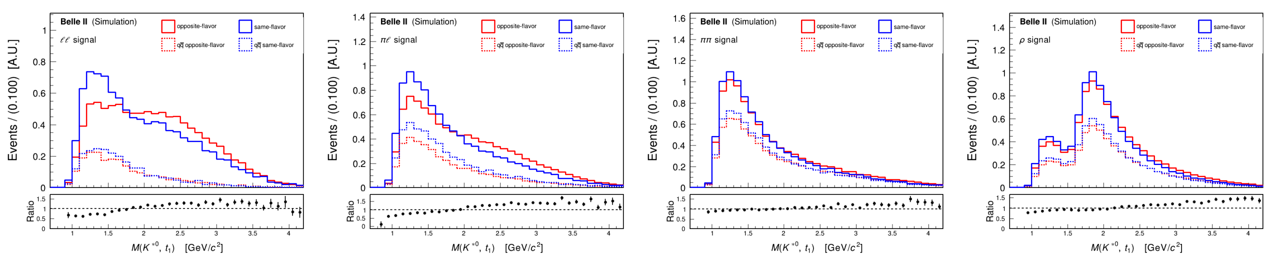


Figure A.32: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the invariant mass of the K^{*0} and the τ daughter t_1 (having the same sign of the K from K^{*0}), used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

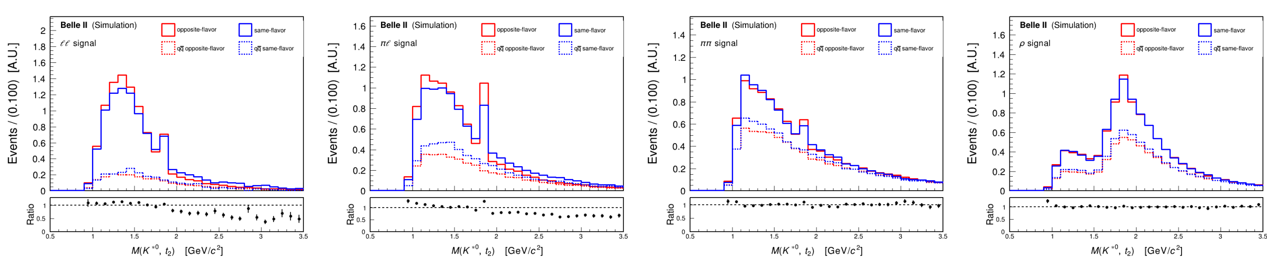


Figure A.33: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the invariant mass of the K^{*0} and the τ daughter t_2 (having an opposite sign of the K from K^{*0}), used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

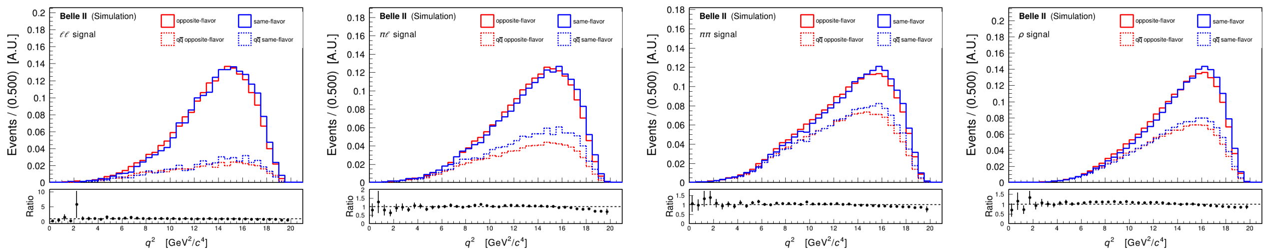


Figure A.34: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the signal side q^2, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

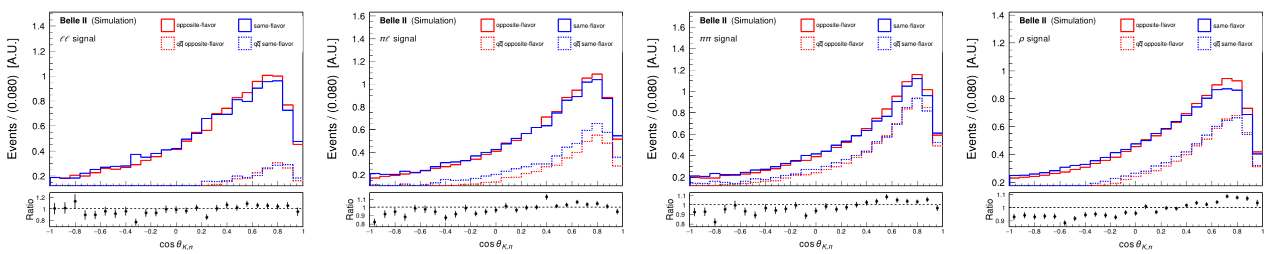


Figure A.35: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the $\cos\theta_{K;\pi}$ of the angle between the K^{*0} daughters, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

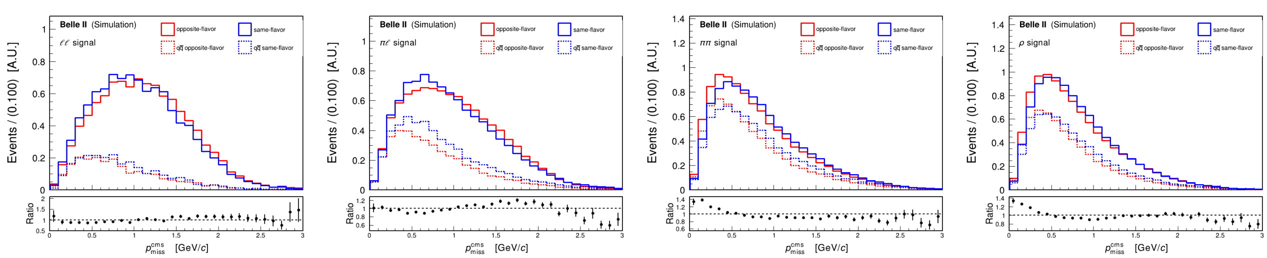


Figure A.36: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the missing momentum of the event in the CM frame, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

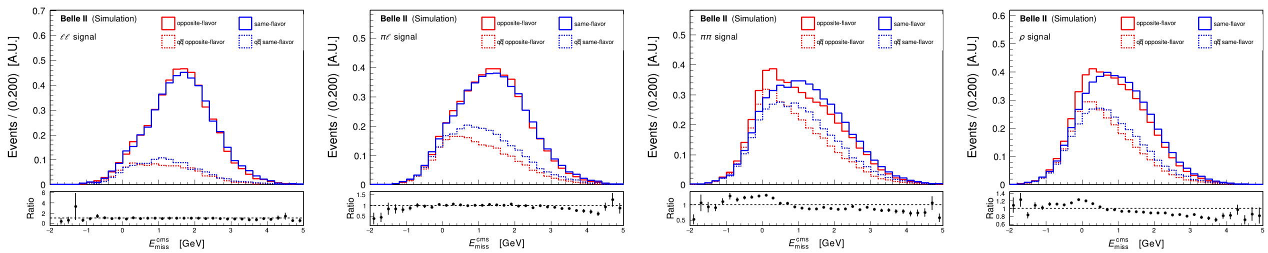


Figure A.37: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the missing energy of the event in the CM frame, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

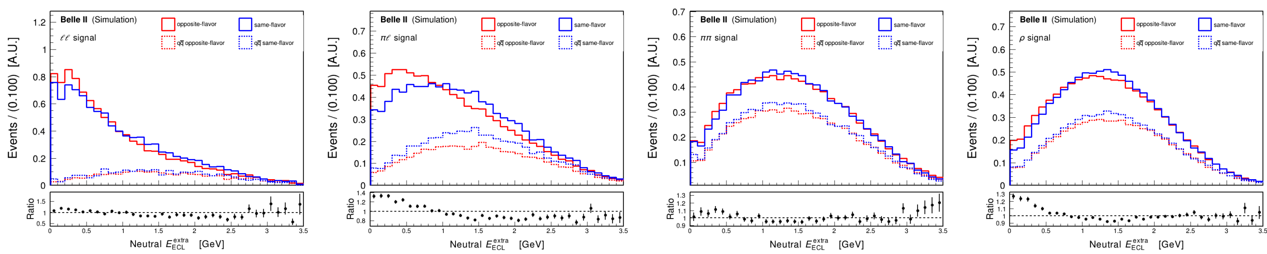


Figure A.38: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the neutral extra energy in the event, as defined by the mask in section 4.2, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

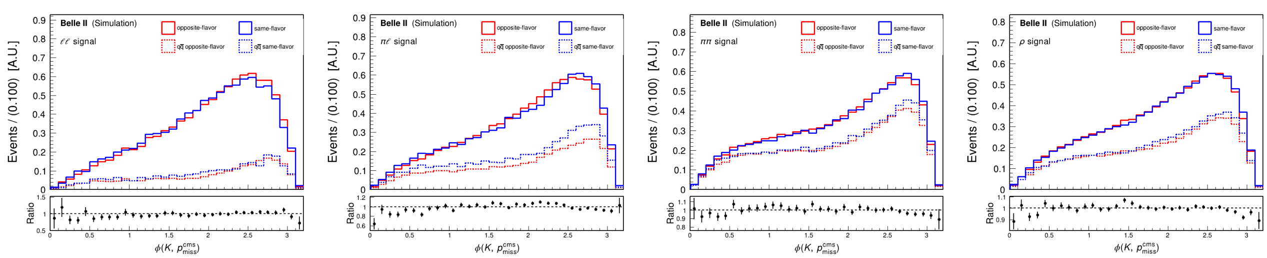


Figure A.39: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the angle φ between the K^{*0} and the missing momentum directions, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

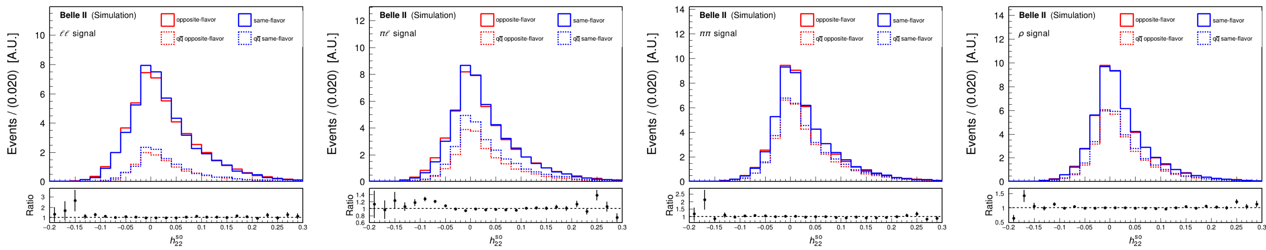


Figure A.40: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the KSFW variable h^{so}_{22}, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

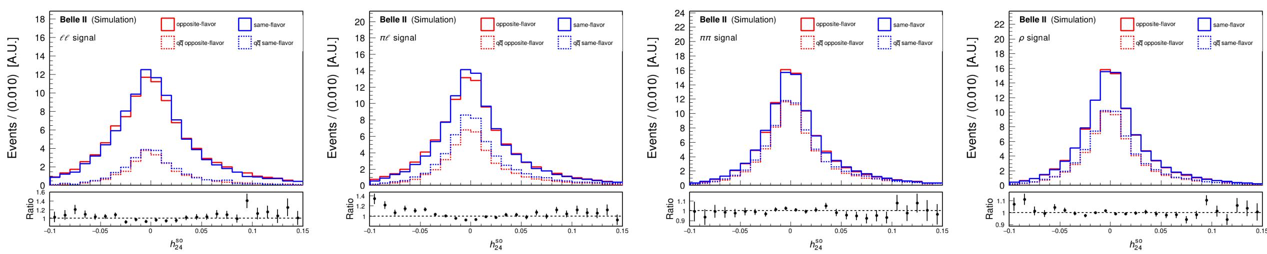


Figure A.41: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the KSFW variable h^{so}_{22}, used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

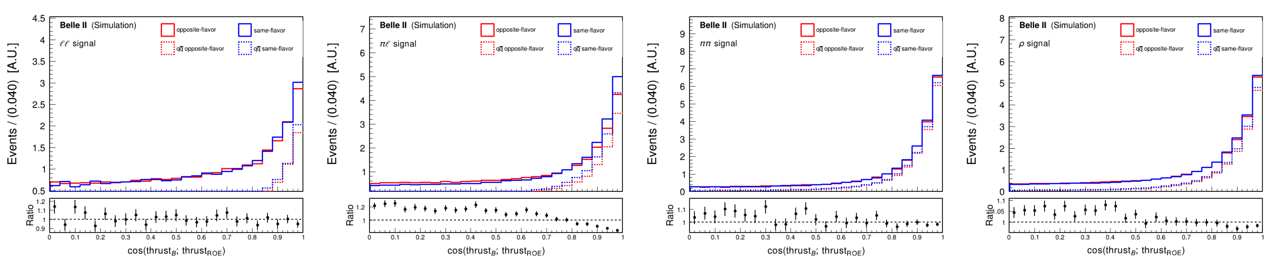


Figure A.42: Comparison between simulated background after the pre-selection (red histograms) and the expected backgrounds in the same-flavor sidebands (blue histograms). The two histograms have the same normalization. Distribution of the continuum suppression variable $\cos\theta(T_B; T_{\mathrm{ROE}})$, where θ is the angle between the B_{sgn} axis and the thrust axis. It is used as input features in the BDT training for all the signal categories. All the corrections described in section 4.3 are applied.

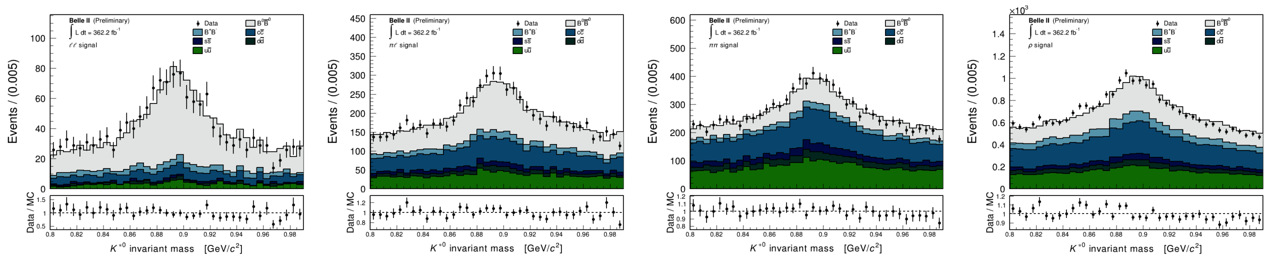


Figure A.43: Data/MC comparison on the same-flavor sideband for the invariant mass of reconstructed K^{*0}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

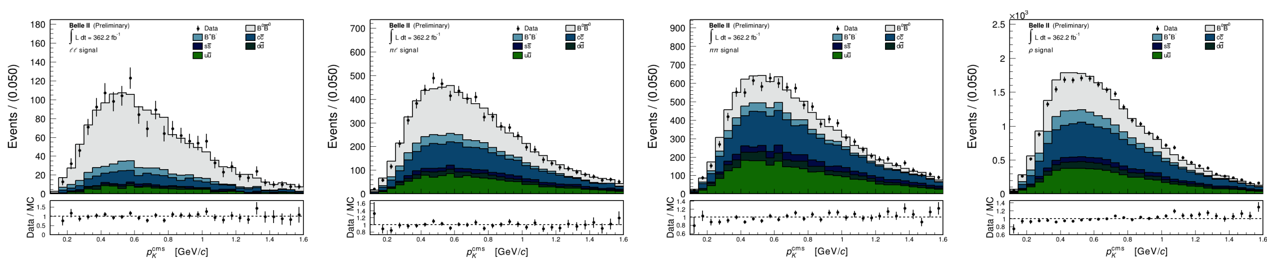


Figure A.44: Data/MC comparison on the same-flavor sideband for the charged K (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

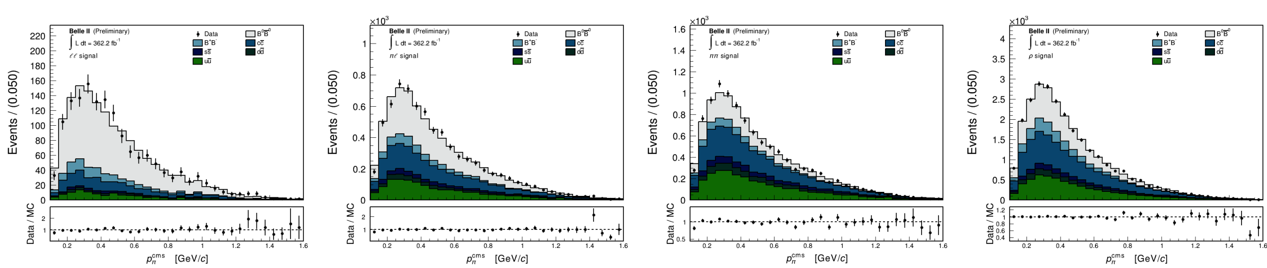


Figure A.45: Data/MC comparison on the same-flavor sideband for the charged π (associated to the K^{*0}) momentum measured in the center of mass frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

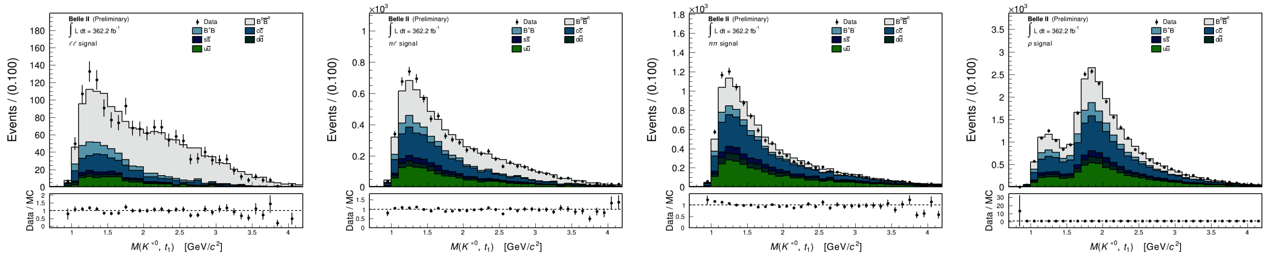


Figure A.46: Data/MC comparison on the same-flavor sideband for the invariant mass of the K^{*0} and the τ daughter t_1 (having the same sign of the K from K^{*0}), used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

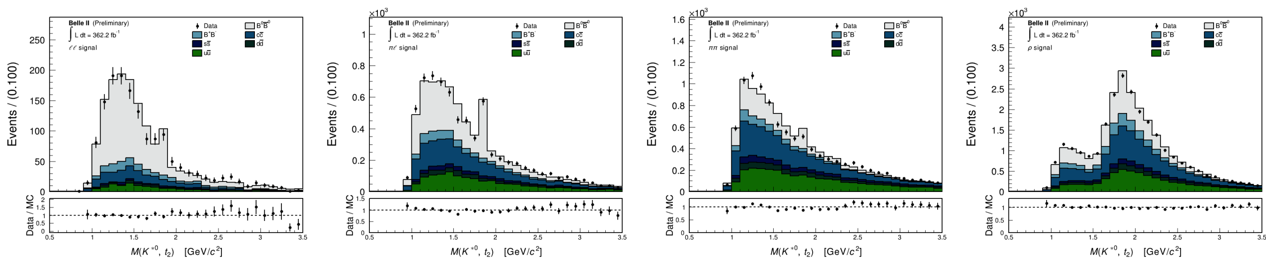


Figure A.47: Data/MC comparison on the same-flavor sideband for the invariant mass of the K^{*0} and the τ daughter t_2 (having an opposite sign of the K from K^{*0}), used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

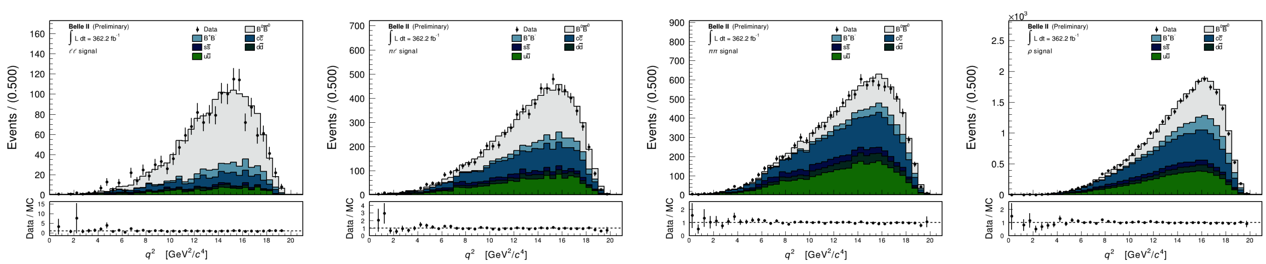


Figure A.48: Data/MC comparison on the same-flavor sideband for the signal side q^2, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

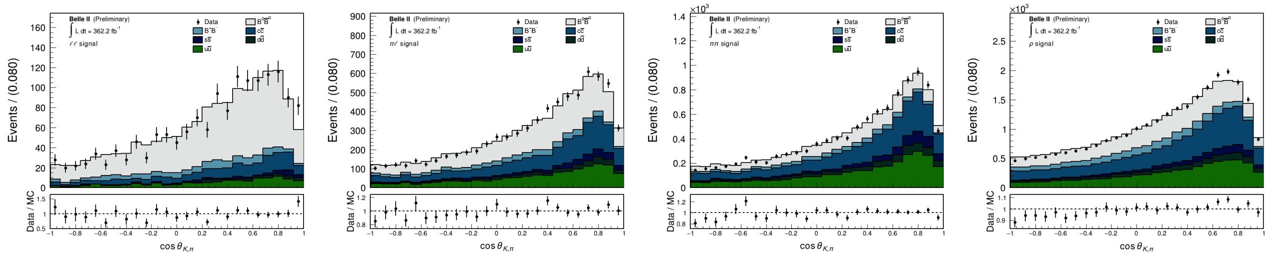


Figure A.49: Data/MC comparison on the same-flavor sideband for the $\cos\theta_{K;\pi}$ of the angle between the K^{*0} daughters, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

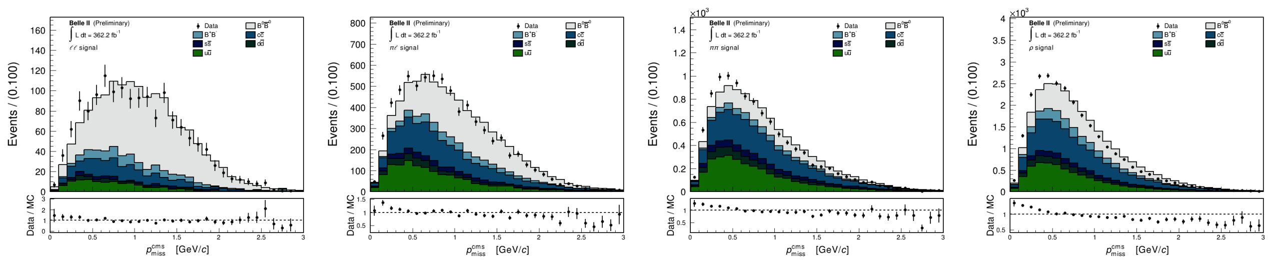


Figure A.50: Data/MC comparison on the same-flavor sideband for the missing momentum of the event in the CM frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

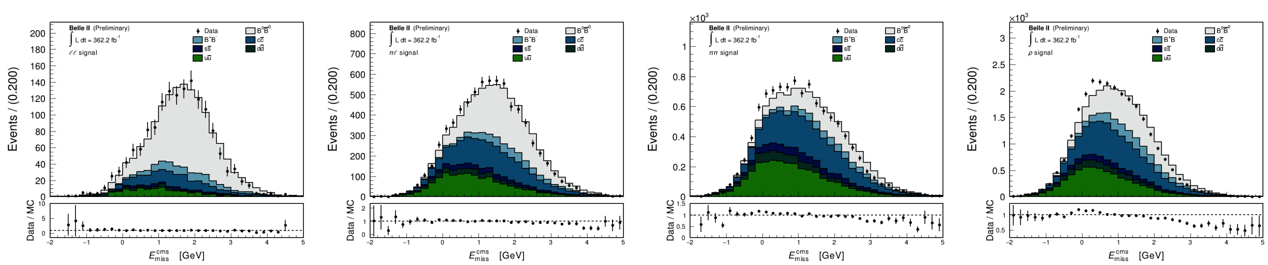


Figure A.51: Data/MC comparison on the same-flavor sideband for the missing energy of the event in the CM frame, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

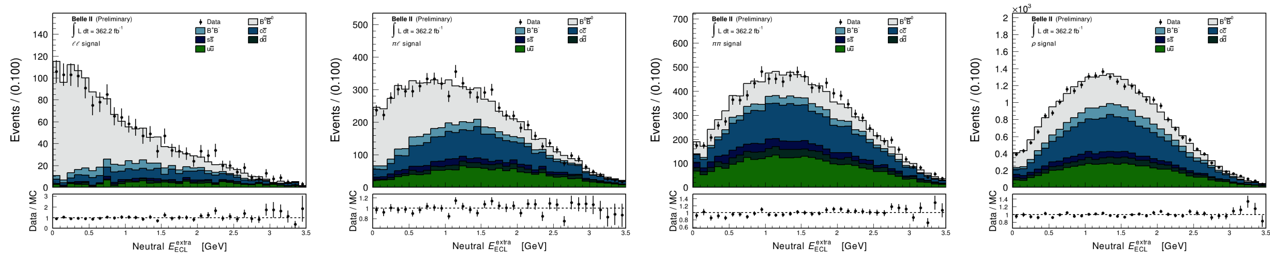


Figure A.52: Data/MC comparison on the same-flavor sideband for the neutral extra energy in the event, as defined by the mask in section 4.2, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

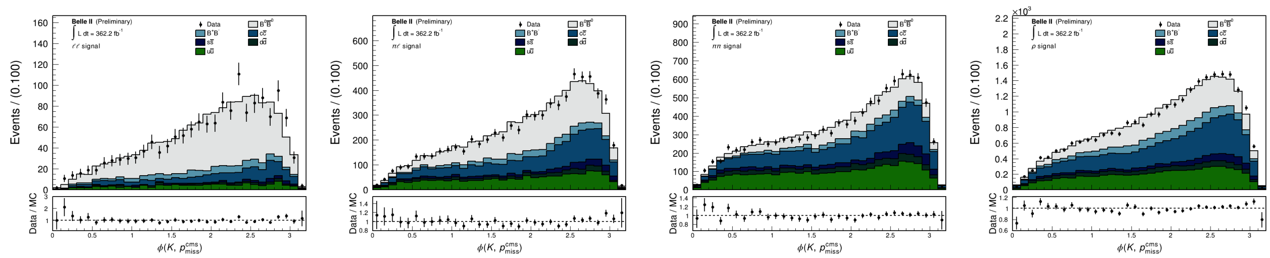


Figure A.53: Data/MC comparison on the same-flavor sideband for the angle φ between the K^{*0} and the missing momentum directions, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

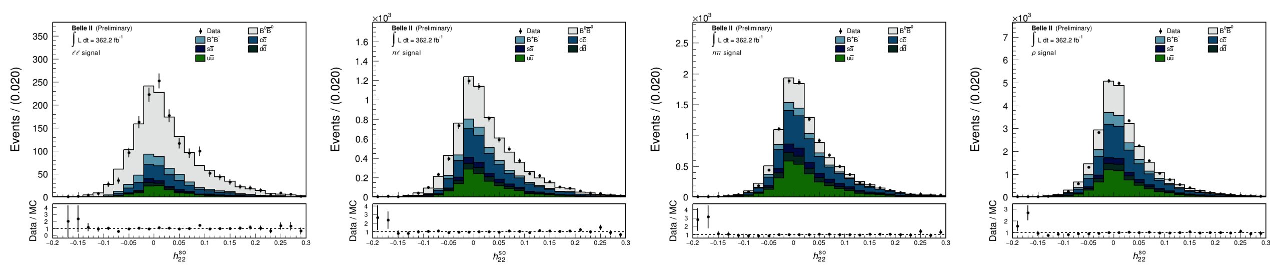


Figure A.54: Data/MC comparison on the same-flavor sideband for the KSFW variable h_{22}^{so}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

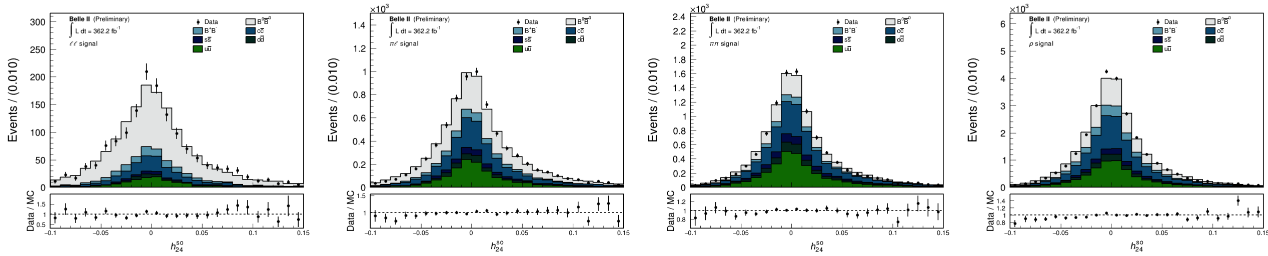


Figure A.55: Data/MC comparison on the same-flavor sideband for the KSFW variable h_{22}^{so}, used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

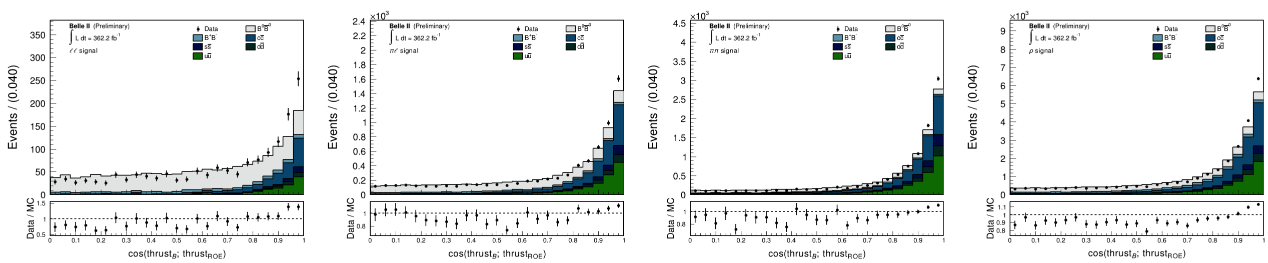


Figure A.56: Data/MC comparison on the same-flavor sideband for the continuum suppression variable $\cos\theta(T_B; T_{\mathrm{ROE}})$, where θ is the angle between the B_{sgn} axis and the thrust axis. It is used as input features in the BDT training for all the signal categories. Simulated backgrounds are rescaled based on data integrated luminosity and all the corrections described in section 4.3 are applied.

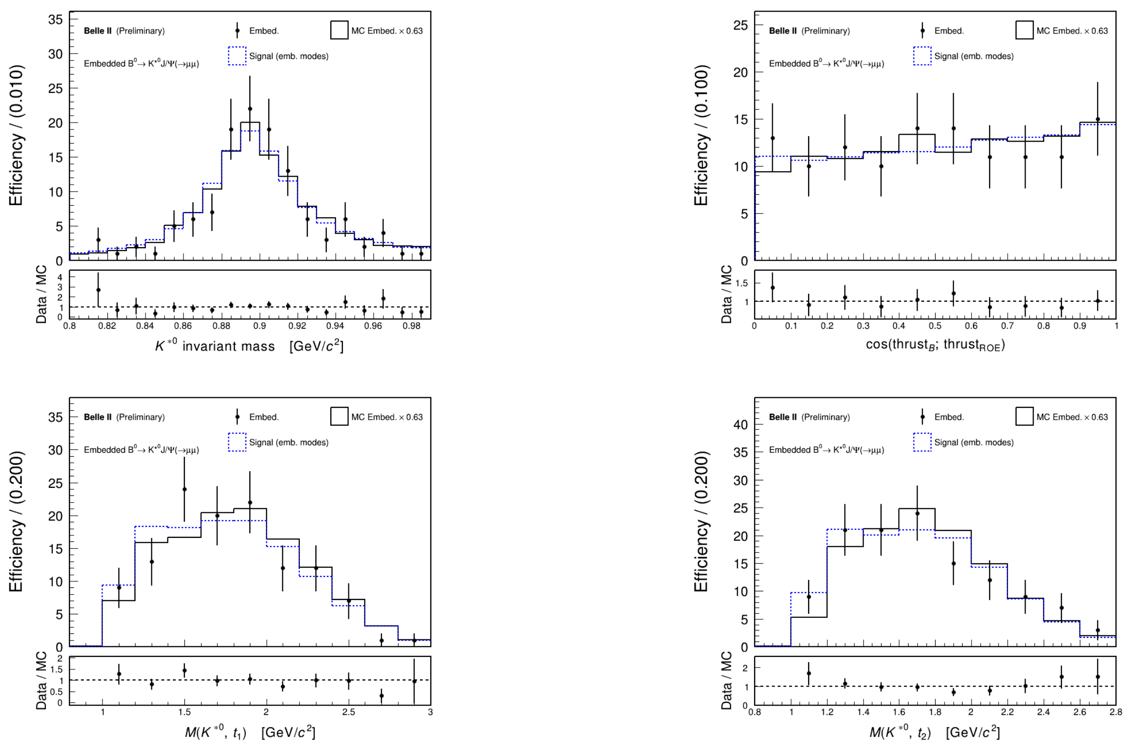


Figure A.57: Efficiency distributions of BDT input features obtained on the embedded sample using a 362.2 fb^{-1} data sample (error-bars) and a simulated $B^0 \to K^{*0}J/\psi$ sample (solid black line, scaled by the calibration factor $\times 0.53$). Dashed histogram corresponds to simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal distribution restricted exclusively to the 1-prong τ decay modes used for the embedded sample.

A.4 same-flavor validation plots

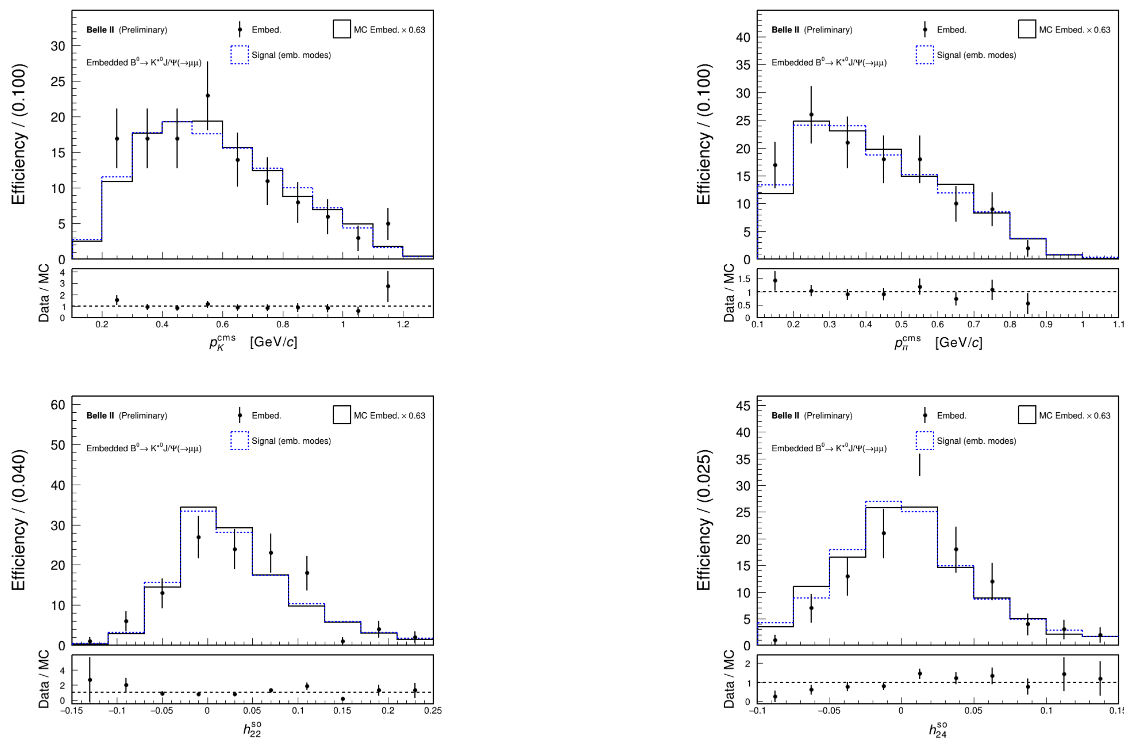


Figure A.58: Efficiency distributions of BDT input features obtained on the embedded sample using a 362.2 fb^{-1} data sample (error-bars) and a simulated $B^0 \to K^{*0} J/\psi$ sample (solid black line, scaled by the calibration factor $\times 0.53$). Dashed histogram corresponds to simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal distribution restricted exclusively to the 1-prong τ decay modes used for the embedded sample.

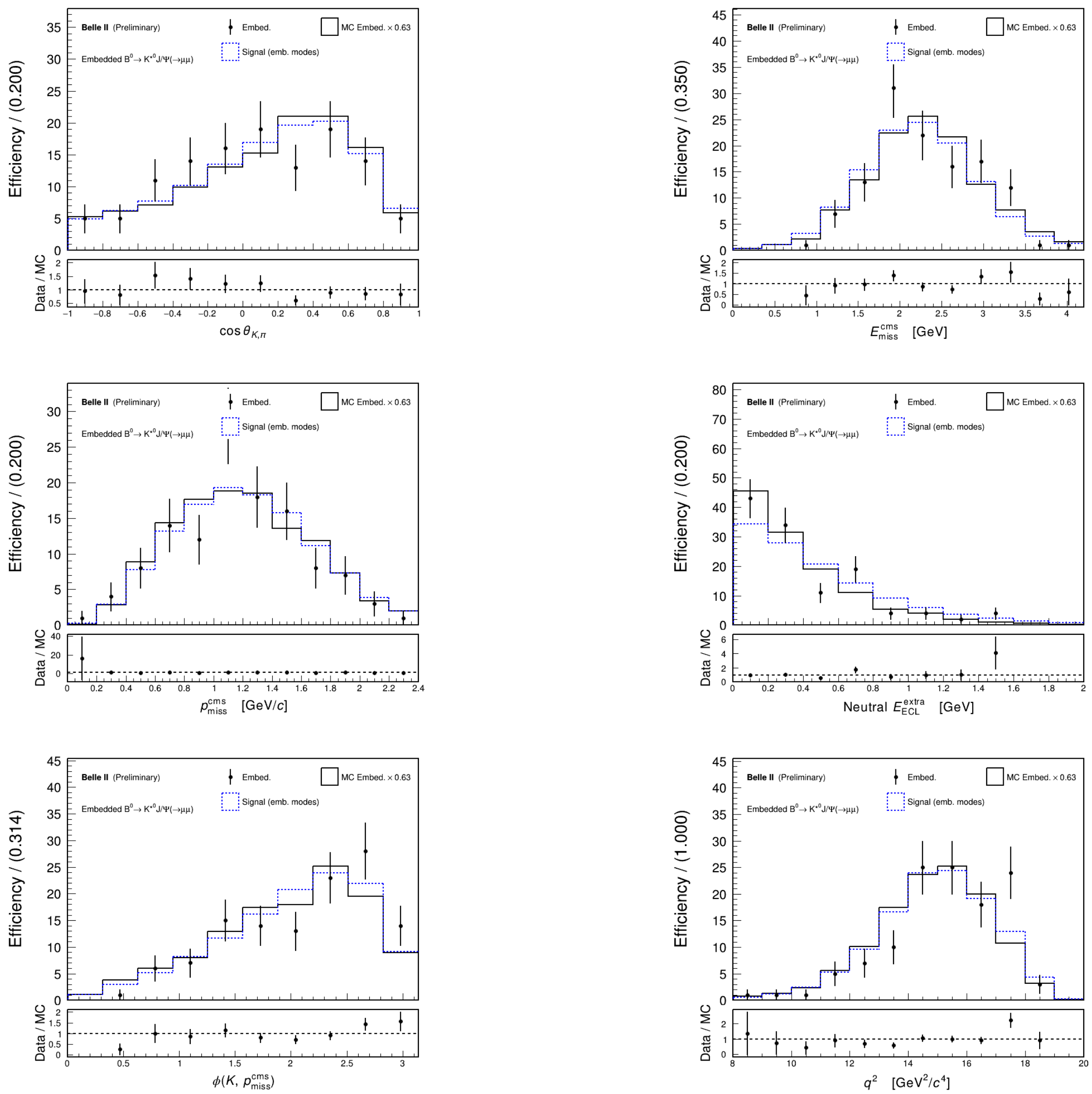


Figure A.59: Efficiency distributions of BDT input features obtained on the embedded sample using a 362.2 fb^{-1} data sample (error-bars) and a simulated $B^0 \to K^{*0} J/\psi$ sample (solid black line, scaled by the calibration factor $\times 0.53$). Dashed histogram corresponds to simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal distribution restricted exclusively to the 1-prong τ decay modes used for the embedded sample.

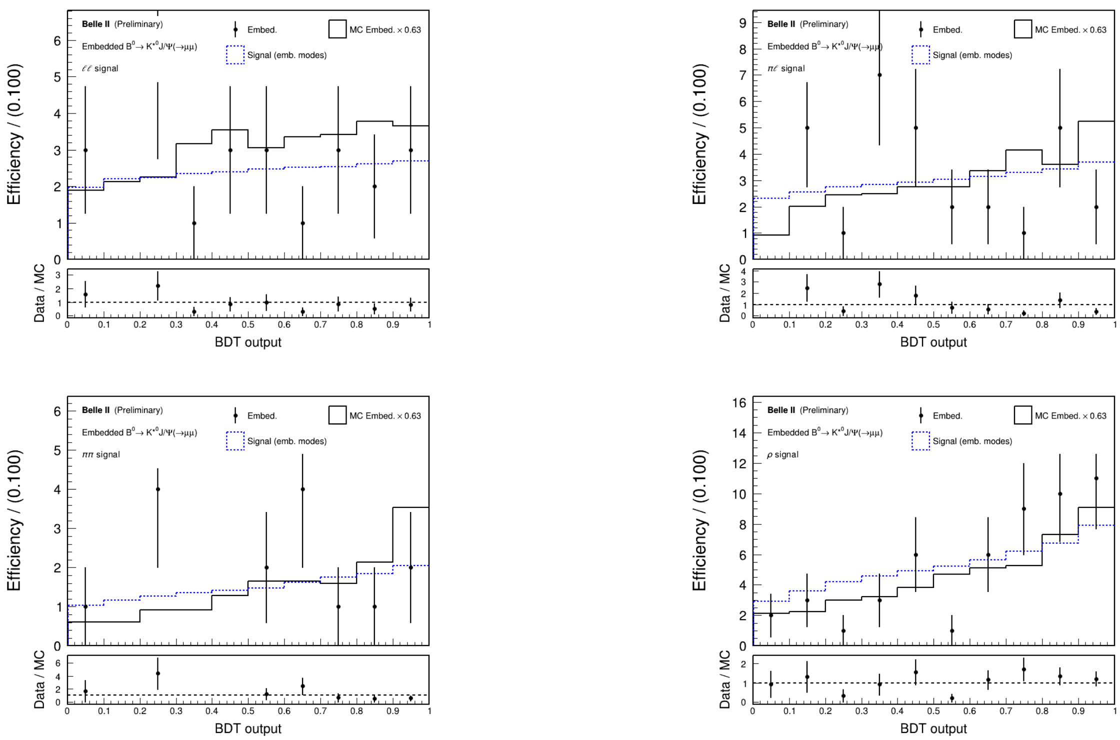


Figure A.60: Efficiency distributions of the BDT outputs obtained on the embedded sample using a 362.2 fb^{-1} data sample (error-bars) and a simulated $B^0 \to K^{*0} J/\psi$ sample (solid black line, scaled by the calibration factor $\times 0.63$). Dashed histogram corresponds to simulated $B^0 \to K^{*0}\tau^+\tau^-$ signal distribution restricted exclusively to the 1-prong τ decay modes used for the embedded sample.

Bibliography

[1] Mark Thomson. *Modern Particle Physics.* Cambridge University Press, 2013.

[2] F. Halzen and Alan D. Martin. *Quarks and Leptons: an introductory course in Modern Particle Physics.* New York: Wiley, Jan. 1984.

[3] G. Aad, T. Abajyan, B. Abbott, et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29.

[4] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61.

[5] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle dark matter: evidence, candidates and constraints". In: *Physics Reports* 405.5-6 (Jan. 2005), pp. 279–390.

[6] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01.

[7] Y. Amhis et al. "(HFLAV), Preliminary average of $R(D)$ and $R(D*)$ for Winter 2023". In: (2023).

[8] R. Aaij, B. Adeva, M. Adinolfi, et al. "Measurement of the Ratio of Branching Fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$". In: *Phys. Rev. Lett.* 120 (12 Mar. 2018), p. 121801.

[9] R. Aaij, A. S. W. Abdelmotteleb, C. Abellan Beteta, et al. "Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *Phys. Rev. D* 108 (3 Aug. 2023), p. 032002.

[10] R. Aaij, A. S. W. Abdelmotteleb, C. Abellan Beteta, et al. "Test of Lepton Universality in $b \to s\ell^+\ell^-$ Decays". In: *Phys. Rev. Lett.* 131 (5 Aug. 2023), p. 051803.

[11] JoAnne L. Hewett. "τ polarization asymmetry in $B \to X_s\tau^+\tau^-$". In: *Phys. Rev. D* 53 (9 May 1996), pp. 4964–4969.

[12] Bernat Capdevila et al. "Searching for New Physics with $b \to s\tau^+\tau^-$ Processes". In: *Phys. Rev. Lett.* 120 (18 May 2018), p. 181802.

[13] J. P. Lees, V. Poireau, V. Tisserand, et al. "Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar Experiment". In: *Phys. Rev. Lett.* 118 (3 Jan. 2017), p. 031802.

[14] T. V. Dong, T. Luo, et al. "Search for the decay $B^0 \to K^{*0}\tau^+\tau^-$ at the Belle experiment". In: *Phys. Rev. D* 108 (1 July 2023), p. L011102.

[15] R. Aaij, B. Adeva, M. Adinolfi, et al. "Search for the Decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$". In: *Phys. Rev. Lett.* 118 (25 June 2017), p. 251802.

[16] A. J. Bevan et al. "The Physics of the B Factories". In: *The European Physical Journal C* 74.11 (Nov. 2014).

[17] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. eprint: `https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf`.

[18] R. Aleksan et al. "Measuring CP violation in the $B^0 B^0$ system with asymmetric-energy e^+e^- beams". In: *Phys. Rev. D* 39 (5 Mar. 1989), pp. 1283–1296.

[19] B. Aubert et al. "Observation of CP violation in the B^0 meson system". In: *Physical Review Letters* 87.9 (Aug. 2001).

[20] K. Abe et al. "Observation of Large CP Violation in the Neutral B Meson System". In: *Physical Review Letters* 87.9 (Aug. 2001).

[21] Dario B. Crosetto et al. *LHCb: Technical Proposal. A Large Hadron Collider Beauty experiment for precision measurements of CP violation and rare decays.* Geneva: CERN, Feb. 1998.

[22] T. Aushev et al. *Physics at Super B Factory.* 2010. arXiv: `1002.5012 [hep-ex]`.

[23] M. Bona et al. "SuperB: A High-Luminosity Asymmetric e^+ e^- Super Flavor Factory. Conceptual Design Report". In: *INFN Publishing Services* (Sept. 4, 2007). arXiv: `0709.0451v2 [hep-ex]`.

[24] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. "SuperKEKB collider". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (2018). Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn), pp. 188–199.

[25] T. Abe et al. *Belle II Technical Design Report.* 2010. arXiv: `1011.0352 [physics.ins-det]`.

[26] M.A. Furman. "Hourglass effects for asymmetric colliders". In: *Conference Record of the 1991 IEEE Particle Accelerator Conference.* 1991, 422–424 vol.1.

[27] Yukiyoshi Ohnishi et al. "Accelerator design at SuperKEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 2013), 03A011. eprint: `https://academic.oup.com/ptep/article-pdf/2013/3/03A011/4439973/pts083.pdf`.

[28] B. Wang et al. "Operational experience of the Belle II pixel detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1032 (2022), p. 166631.

[29] F. Forti et al. "The design, construction, operation and performance of the Belle II silicon vertex detector". In: *Journal of Instrumentation* 17.11 (Nov. 2022), P11042.

[30] Umberto Tamponi. *The TOP counter of Belle II: status and first results*. 2018. arXiv: 1811.04532 [hep-ex].

[31] H. Ikeda, A. Satpathy, B. S. Ahn, et al. "A detailed test of the CsI(Tℓ) calorimeter for BELLE with photon beams of energy between 20 MeV and 5.4 GeV". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 441.3 (Mar. 2000), pp. 401–426.

[32] T. Aushev et al. "A scintillator based endcap K_L and muon detector for the Belle II experiment". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 789 (2015), pp. 134–142.

[33] Satoru Yamada et al. "Data Acquisition System for the Belle II Experiment". In: *IEEE Transactions on Nuclear Science* 62.3 (2015), pp. 1175–1180.

[34] T. Kuhr et al. "The Belle II Core Software". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 1. arXiv: 1809.04299 [physics.comp-ph].

[35] Belle II collaboration. *Belle II Analysis Software Framework (basf2)*. https://doi.org/10.5281/zenodo.5574115.

[36] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177.

[37] David J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155.

[38] S. Jadach, B. F. L. Ward, and Z. Wa. "The precision Monte Carlo event generator for two-fermion final states in collisions". In: *Computer Physics Communications* 130.3 (Aug. 2000), pp. 260–325.

[39] Stanisław Jadach, Johann H. Kühn, and Zbigniew Was. "TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons". In: *Computer Physics Communications* 64.2 (1991), pp. 275–299.

[40] N. Davidson et al. "Universal interface of TAUOLA: Technical and physics documentation". In: *Computer Physics Communications* 183.3 (2012), pp. 821–843.

[41] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl.Instrum.Meth.* A506 (2003), pp. 250–303.

[42] Valerio Bertacchi et al. "Track finding at Belle II". In: *Computer Physics Communications* 259 (2021), p. 107610.

[43] Tobias Schlüter. "Vertexing and Tracking Software at Belle II". In: *PoS (Vertex2014)* (Nov. 13, 2014). arXiv: 1411.3485v1 [physics.ins-det].

[44] P. V. C. Hough. *Method and Means for Recognizing Complex Patterns*. Dec. 1962.

[45] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45.

[46] Henrikas Svidras et al. "Measurement of the data to MC ratio of photon reconstruction efficiency of the Belle II calorimeter using radiative muon pair events". In: (Sept. 2021).

[47] Saurabh Sandilya and Alan Schwartz. "Kaon and Pion Identification Performances in Phase III data for TOP detector". In: (July 2019).

[48] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. Cambridge, MA, USA: MIT Press, 2016.

[49] J. Sjonberg and L. Ljung. "Overtraining, regularization and searching for a minimum, with application to neural networks". In: *International Journal of Control* 62.6 (1995), pp. 1391–1407. eprint: `https://doi.org/10.1080/00207179508921605`.

[50] L. Breiman et al. *Classification and Regression Trees*. Taylor & Francis, 1984.

[51] Thomas Keck. "FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment". In: *Computing and Software for Big Science* 1.2 (2017).

[52] Olaf Behnke et al. *Data analysis in high energy physics: a practical guide to statistical methods*. Weinheim: Wiley-VCH, 2013.

[53] Leo Breiman. "Arcing Classifiers". In: *The Annals of Statistics* 26.3 (1998), pp. 801–824.

[54] Jerome H. Friedman. "Stochastic gradient boosting". In: *Computational Statistics and Data Analysis* 38.4 (2002). Nonlinear Methods and Data Mining, pp. 367–378.

[55] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, Aug. 2016.

[56] J. Brownlee. *XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn*. Machine Learning Mastery, 2016.

[57] T. Keck et al. "The Full Event Interpretation". In: *Computing and Software for Big Science* 3.1 (Feb. 2019).

[58] M. Feindt et al. "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (2011), pp. 432–440.

[59] M. Feindt and U. Kerzel. “The NeuroBayes neural network package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (2006). Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research, pp. 190–194.

[60] G. Cowan. *Statistical data analysis*. Oxford University Press, USA, 1998.

[61] F. James. *Statistical Methods In Experimental Physics (2nd Edition)*. World Scientific Publishing Company, 2006.

[62] A. L. Read. “Presentation of search results: the CLs technique”. In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), p. 2693.

[63] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011).

[64] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012.

[65] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *pyhf: v0.7.5*. Version 0.7.5. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.5.

[66] Lukas Heinrich et al. “pyhf: pure-Python implementation of HistFactory statistical models”. In: *Journal of Open Source Software* 6.58 (2021), p. 2823.

[67] Kyle Cranmer and Alexander Held. “Building and steering binned template fits with cabinetry”. In: *25th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2021)* (2021).

[68] H. Albrecht, A.A. Andam, U. Binder, et al. “Observation of $B^0 - \overline{B}^0$ mixing”. In: *Physics Letters B* 192.1 (1987), pp. 245–252.

[69] B. Ananthanarayan and Kshitij Garg. “Quantum mechanical entanglement and tests of CPT theorem with neutral mesons at e^+e^- colliders”. In: *Current Science* 93.8 (2007), pp. 1055–1057.

[70] Patricia Ball and Roman Zwicky. “$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form factors from light-cone sum rules reexamined”. In: *Phys. Rev. D* 71 (1 Jan. 2005), p. 014029.

[71] J.-F. Krohn et al. “Global decay chain vertex fitting at Belle II”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 976 (2020), p. 164269.