Aix Marseille Université

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE
CPPM

Master de Physique fondamentale et Applications

# Search for $B \to \tau\tau$ decays with Belle II


Author : Elvin MORACCHINI

Tutors : Giampiero MANCINELLI & Justine SERRANO

Host Institution: Aix-Marseille Université

Laboratory: Centre de Physique des Particules de Marseille (CPPM)


July 16, 2026


**Abstract**

The decay $B^0 \to \tau^+\tau^-$ is a highly suppressed flavor-changing neutral-current process with a Standard Model branching fraction of approximately $2 \times 10^{-8}$. Owing to its strong helicity suppression and its sensitivity to additional scalar contributions, this decay is a particularly interesting probe of physics beyond the Standard Model. This report presents a search for the decay $B^0 \to \tau^+\tau^-$ using simulated Belle II samples. The analysis relies on the hadronic Full Event Interpretation algorithm to reconstruct one of the two $B$ mesons produced in $\Upsilon(4S)$ decays, while the signal-side candidate is reconstructed from the visible decay products of the two $\tau$ leptons. The large background contribution is reduced using Boosted Decision Trees trained separately for four signal categories. The multivariate analysis was optimized by selecting category-dependent input variables and tuning the classifier configuration. The performance of the optimized classifiers was assessed through feature importance studies, BDT output distributions, and the Punzi figure of merit. Compared with the baseline analysis, the optimized strategy improves the expected sensitivity and yields an expected upper limit of $\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.76 \times 10^{-4}$ at the 90% confidence level using the available simulated samples.

# Acknowledgements

I would like to express my sincere gratitude to my supervisors, Giampiero Mancinelli and Justine Serrano, for their guidance, support, and availability throughout this internship. Their expertise and advice were invaluable in helping me understand both the physics motivations and the experimental challenges.

I am also grateful to the members of the Belle II group at CPPM for the many fruitful discussions and for their willingness to share their knowledge and experience.

In particular, I would like to express my sincere gratitude to Mattia Marfoli, whose PhD thesis is dedicated to the analysis on which this work is based. His guidance was invaluable throughout the internship. He was always available to answer questions, explain the analysis framework, discuss the physics motivations behind the different analysis choices, and help me navigate the many technical challenges associated with this search. His patience, support, and willingness to share his expertise greatly contributed to my understanding of both the Belle II software environment and the methodology of modern particle physics analyses.

Finally, what would an internship be without other students?

A huge thank you to all the students of office 126 for the countless coffee breaks, lunches, discussions, and moments of laughter that made these months so enjoyable.

Matteo, Nicolas, after five years together, our Master's journey is finally coming to an end. It has been quite a ride at times, but we made it. I sincerely hope that all of us will find the PhD positions we have been working towards for so long.

Lucile, Laure, we met during this internship, and even if I do not say it very often, I truly value your friendship. I am very grateful for all the discussions, shared experiences, and good moments we had together. I wish you both all the best for the future.

And Lucile, I am still sorry for not coming to the apéro in the Calanques. In my defense, being constantly criticized on my driving did not help.

**Use of Artificial Intelligence Tools:** Artificial intelligence tools were used exclusively to assist with grammar correction, language polishing, and text formatting. All scientific content, analysis methods, results, interpretations, figures, and conclusions were produced and validated by the author.

# 1 Introduction

Rare decays of heavy mesons provide a powerful framework for testing the Standard Model (SM) and searching for possible contributions from physics beyond it. Because these processes are highly suppressed in the Standard Model, even small deviations from the predicted branching fractions could provide evidence for new particles or interactions beyond the energy reach of current colliders.

Among these rare decays, $B^0 \to \tau^+\tau^-$ is of particular interest. It is suppressed both by the flavor changing neutral current structure of the process and by the helicity structure of the weak interaction, resulting in a predicted branching fraction of about $2 \times 10^{-8}$ in the Standard Model. Several extensions of the Standard Model, such as models with an extended Higgs sector, predict sizable enhancements of this branching fraction, making this decay an excellent channel in which to search for physics beyond the Standard Model.

Experimentally, however, the search for $B^0 \to \tau^+\tau^-$ is particularly demanding. Depending on the $\tau$ decay modes, the final state contains at least two and up to four neutrinos, which prevents a complete reconstruction of the signal decay. The clean experimental environment of Belle II, operating at the $\Upsilon(4S)$ resonance, together with advanced reconstruction algorithms such as the Full Event Interpretation (FEI), makes it possible to tackle this challenging search.

The work presented in this report focuses on the development and optimization of a search strategy for $B^0 \to \tau^+\tau^-$ using simulated Belle II samples. Particular emphasis is placed on the use of multivariate techniques based on Boosted Decision Trees (BDTs) to suppress the large background contributions while maintaining a high signal efficiency. Different training strategies, including the use of truth-matched signal samples, are investigated and their impact on the expected sensitivity is evaluated.

The report is organized as follows. Chapter 2 introduces the theoretical framework of the decay and its sensitivity to physics beyond the Standard Model. Chapter 3 presents the Belle II experiment and the SuperKEKB accelerator. The event reconstruction and selection strategy are described in Chapters 4 and 5. Chapter 6 details the multivariate analysis and optimization procedure, while Chapter 7 presents the resulting sensitivities and compares the different training strategies considered in this study.

# 2 The decay $B^0 \to \tau^+\tau^-$

The SM of particle physics has been developed and tested extensively over the past decades and provides an accurate description of elementary particles and their interactions. Despite its remarkable experimental success, it is generally regarded as an effective theory valid at the energy scales explored so far. A more fundamental theory is expected to emerge at higher energies, introducing new particles and interactions beyond those described by the SM. If these new particles are much heavier than the energy scale of the process being studied, they can be integrated out together with the heavy SM particles, leading to an effective field theory that describes the low-energy dynamics.

## 2.1 SM framework

Within the SM, flavor-changing neutral currents (FCNC) are absent at tree level due to the gauge structure of the theory and the unitarity of the CKM matrix. They can only arise at higher orders through loop-induced processes, such as box and penguin diagrams, where the Glashow–Iliopoulos–Maiani (GIM) mechanism leads to strong suppressions. As a consequence, processes involving the transition of a b quark into a d or s quark are strongly suppressed in the SM. Such FCNCs transitions occur in the decay of a neutral $B^0 = |\bar{b}d\rangle$ meson into a pair of oppositely charged leptons, $B^0 \to \ell^+\ell^-$ and are sensitive to contributions from physics beyond the SM. In the Standard Model, the dominant contributions to the decay $B^0 \to \ell^+\ell^-$ arise from W-box and Z-penguin diagrams, illustrated in Fig.1. Contributions involving internal u and c quark loops are strongly suppressed and can be neglected compared to the top-quark contribution, owing to the much smaller masses $m_u$ and $m_c$ relative to $m_t$ [1], [2]. The branching ratio $\mathcal{B}(B^0 \to \ell^+\ell^-)$ cannot be calculated directly from the full SM Lagrangian, as the decay involves two widely separated energy scales. These scales correspond to the electroweak scale, characterized by the W boson mass $M_W$, and the hadronic scale associated with strong interactions, $\Lambda_{\text{QCD}}$ [3].

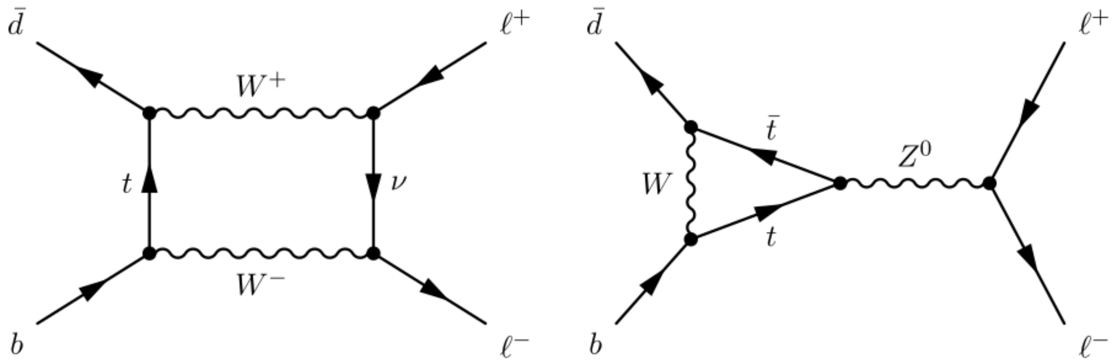



Figure 1: Leading Standard Model contributions to the decay $B^0 \to \ell^+\ell^-$. Left: box diagram involving virtual $W^\pm$ bosons and a top quark. Right: electroweak penguin diagram mediated by a virtual $Z^0$ boson. These flavor-changing neutral-current processes occur only at loop level and are therefore strongly suppressed in the Standard Model.

A widely used framework in flavor physics for describing decays such as $B^0 \to \ell^+\ell^-$ is the Effective Field Theory (EFT). By integrating out the heavy Standard Model particles at the electroweak scale, namely the W and Z bosons and the top quark, a low-energy effective theory is obtained in which only the light Standard Model fields remain [3]. A detailed introduction to EFT can be found in [1]. The effective Hamiltonian $\mathcal{H}_{\text{eff}}$, which describe the processes at low energies, can be written as,

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \sum_i V^i_{\text{CKM}} C_i(\mu) Q_i, \tag{1}$$

with $G_F$ the Fermi constant, $V_{\text{CKM}}$ the Cabibbo-Kobayashi-Maskawa (CKM) factors, $Q_i$ the local operator, and their Wilson coefficients $C_i(\mu)$, at a renormalisation scale $\mu$. The local operators describing the decay $B^0 \to \ell^+\ell^-$ can be written as,

$$Q_S = m_b \left(\bar{b} P_L d\right)\left(\bar{\ell}\,\ell\right), \tag{2}$$

$$Q_P = m_b \left(\bar{b} P_L d\right)\left(\bar{\ell}\,\gamma_5 \ell\right), \tag{3}$$

$$Q_A = \left(\bar{b}\,\gamma_\mu P_L d\right)\left(\bar{\ell}\,\gamma^\mu \gamma_5 \ell\right). \tag{4}$$

Here, $P_L = (1 - \gamma_5)/2$ denotes the left-handed projection operator, $m_b$ is the $b$-quark mass, $b$ and $d$ represent the quark fields, while $\ell$ denotes the lepton field. The matrices $\gamma_\mu$ and $\gamma_5$ are the usual Dirac matrices. The operators $Q_S$, $Q_P$, and $Q_A$ correspond to scalar, pseudoscalar, and axial-vector interactions, respectively. Within this framework, the effective Hamiltonian governing the decay $B^0 \to \ell^+\ell^-$ in the Standard Model can be written as

$$\mathcal{H}_{\rm eff} = -\frac{G_F \alpha_{em}}{\sqrt{2}\sin^2\theta_{\rm W}} V_{tb}^* V_{td} \left[C_S(\mu)Q_S + C_P(\mu)Q_P + C_A(\mu)Q_A\right]. \tag{5}$$

The contributions from the right-handed scalar quark operators are neglected, as they are proportional to the $d$ quark mass and are therefore strongly suppressed. The contributions to $C_S$ and $C_P$ from Higgs and neutral Goldstone boson exchange are suppressed by $\mathcal{O}(M_B^2/M_W^2)$ and are therefore negligible [2]. The Wilson coefficient $C_A$ receives its dominant contribution from the $W$ box and $Z$ penguin diagrams shown in Fig.1. Hence $\mathcal{H}_{\rm eff}$ reduces to,

$$\mathcal{H}_{\rm eff} = -\frac{G_F \alpha_{em}}{\sqrt{2}\sin^2\theta_{\rm W}} V_{tb}^* V_{td} C_A(\mu) \left(\bar{b}\,\gamma_\mu P_L d\right) \left(\bar{\ell}\,\gamma^\mu\gamma_5 \ell\right). \tag{6}$$

The decay amplitude is obtained from the matrix element of the effective Hamiltonian between the initial $B^0$ meson state and the final dilepton state,

$$\mathcal{M} = \langle \ell^+\ell^- | \mathcal{H}_{\rm eff} | B^0 \rangle. \tag{7}$$

Substituting the effective Hamiltonian yields

$$\mathcal{M} = -\frac{G_F \alpha_{\rm em}}{\sqrt{2}\sin^2\theta_W} V_{tb}^* V_{td}\, C_A(\mu)\, \langle 0 | \bar{b}\gamma_\mu P_L d | B^0(p)\rangle\ \langle \ell^+\ell^- |\, \bar{\ell}\gamma^\mu\gamma_5\ell\, | 0\rangle \tag{8}$$

The hadronic matrix element can be expressed in terms of the $B$-meson decay constant $f_B$ [2],

$$\langle 0 | \bar{b}\gamma_\mu\gamma_5 d | B^0(p)\rangle = i f_B p_\mu, \tag{9}$$

while the vector current vanishes,

$$\langle 0 | \bar{b}\gamma_\mu d | B^0(p)\rangle = 0. \tag{10}$$

Using the definition of the left-handed projector,

$$P_L = \frac{1-\gamma_5}{2}, \tag{11}$$

one finds

$$\langle 0 | \bar{b}\gamma_\mu P_L d | B^0(p)\rangle = -\frac{i}{2} f_B p_\mu. \tag{12}$$

The decay amplitude therefore becomes

$$\mathcal{M} = i\frac{G_F\alpha_{\rm em}}{2\sqrt{2}\sin^2\theta_W} V_{tb}^* V_{td}\, C_A(\mu) f_B p_\mu \, \langle \ell^+\ell^- | \, \bar{\ell}\gamma^\mu\gamma_5\ell \, |0\rangle \, . \tag{13}$$

The leptonic current can be expressed in terms of the external spinors as,

$$\langle \ell^+\ell^- | \, \bar{\ell}\gamma^\mu\gamma_5\ell \, |0\rangle = \bar{u}(p_{\ell^-})\gamma^\mu\gamma_5 v(p_{\ell^+}) \tag{14}$$

Using momentum conservation, $p_\mu = (p_{\ell^+} + p_{\ell^-})_\mu$, together with the Dirac equations for the outgoing leptons, one obtains

$$p_\mu \bar{u}(p_{\ell^-})\gamma^\mu\gamma_5 v(p_{\ell^+}) = 2m_\ell \, \bar{u}(p_{\ell^-})\gamma_5 v(p_{\ell^+}). \tag{15}$$

The amplitude can thus be written as

$$\mathcal{M} = i\frac{G_F\alpha_{\rm em}}{\sqrt{2}\sin^2\theta_W} V_{tb}^* V_{td}\, C_A(\mu) f_B m_\ell \, \bar{u}(p_{\ell^-})\gamma_5 v(p_{\ell^+}). \tag{16}$$

This explicit dependence on the lepton mass gives rise to the helicity suppression characteristic of the decay $B^0 \to \ell^+\ell^-$ as we will see later. To derive the decay width, the squared matrix element is summed over the spins of the final-state leptons,

$$\sum_{\rm spins} |\mathcal{M}|^2 = \frac{G_F^2\alpha_{\rm em}^2}{2\sin^4\theta_W} |V_{tb}^* V_{td}|^2 |C_A(\mu)|^2 f_B^2 m_\ell^2 \, {\rm Tr}\Big[(\not{p}_{\ell^-} + m_\ell)\gamma_5(\not{p}_{\ell^+} - m_\ell)\gamma_5\Big] \, , \tag{17}$$

$$= \frac{2G_F^2\alpha_{\rm em}^2}{\sin^4\theta_W} |V_{tb}^* V_{td}|^2 |C_A(\mu)|^2 f_B^2 m_\ell^2 \, (p_{\ell^-}\cdot p_{\ell^+} - m_\ell^2), \tag{18}$$

$$= \frac{G_F^2\alpha_{\rm em}^2}{\sin^4\theta_W} |V_{tb}^* V_{td}|^2 |C_A(\mu)|^2 f_B^2 m_\ell^2 \, (M_B^2 - 4m_\ell^2), \tag{19}$$

where the relation

$$2\, p_{\ell^-}\cdot p_{\ell^+} = M_B^2 - 2m_\ell^2 \tag{20}$$

has been used. The decay width is then obtained from the standard two-body phase-space formula [4]

$$\Gamma(B^0 \to \ell^+\ell^-) = \frac{1}{16\pi M_B}\sqrt{1 - \frac{4m_\ell^2}{M_B^2}} \sum_{\rm spins} |\mathcal{M}|^2 \tag{21}$$

$$= \frac{G_F^2\alpha_{\rm em}^2}{16\pi\sin^4\theta_W} f_B^2 M_B m_\ell^2 |V_{tb}^* V_{td}|^2 |C_A(\mu)|^2 \sqrt{1 - \frac{4m_\ell^2}{M_B^2}}. \tag{22}$$

Finally, using

$$\mathcal{B}(B^0 \to \ell^+\ell^-) = \tau_B\, \Gamma(B^0 \to \ell^+\ell^-) = \frac{\Gamma(B^0 \to \ell^+\ell^-)}{\Gamma_B}, \tag{23}$$

one obtains

$$\mathcal{B}(B^0 \to \ell^+\ell^-) = \frac{G_F^2 \alpha_{\mathrm{em}}^2}{16\pi \sin^4\theta_W \Gamma_B} f_B^2 M_B m_\ell^2 |V_{tb}^* V_{td}|^2 |C_A(\mu)|^2 \sqrt{1 - \frac{4m_\ell^2}{M_B^2}}. \tag{24}$$

Using

$$\frac{\alpha_{\mathrm{em}}}{\sin^2\theta_W} = \frac{\sqrt{2}\, G_F M_W^2}{\pi}, \tag{25}$$

the branching ratio can be rewritten as

$$\mathcal{B}(B^0 \to \ell^+\ell^-) = \frac{G_F^4 M_W^4 M_B^3}{8\pi^5 \Gamma_B} f_B^2 |V_{tb}^* V_{td}|^2 \frac{4m_\ell^2}{M_B^2} \sqrt{1 - \frac{4m_\ell^2}{M_B^2}} |C_A(\mu)|^2. \tag{26}$$

where $\Gamma_B$ is the total decay width of the $B^0$ meson, and $V_{tb}^*$ and $V_{td}$ are the relevant elements of the CKM matrix. The decay constant $f_B$ parametrizes the non-perturbative strong-interaction effects binding the quarks inside the $B$ meson and can be determined using lattice QCD calculations [5]. Finally, the factor $\frac{m_\ell^2}{M_B^2}$ leads to a strong helicity suppression of the decay rate. Fig.2 illustrates the possible spin and momentum configurations of the final-state leptons in the $B$-meson rest frame. Since the $B^0$ meson is a spin-zero particle, angular momentum conservation requires the two leptons to be produced with opposite helicities. For light leptons, this configuration is strongly suppressed, resulting in the characteristic helicity suppression factor proportional to $m_\ell^2$.

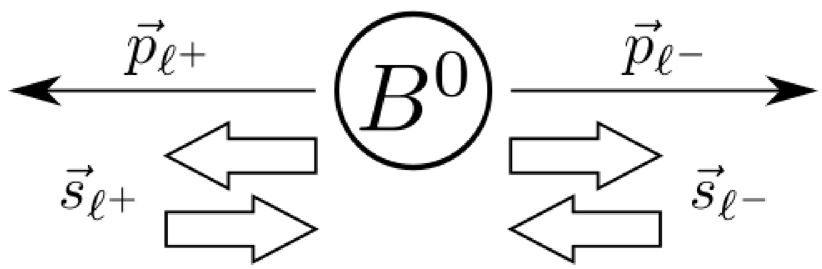



Figure 2: Illustration of the helicity configurations in the decay $B^0 \to \ell^+\ell^-$ in the $B^0$ rest frame. The leptons are produced with equal and opposite momenta in order to conserve momentum. Because the $B^0$ meson has total angular momentum $J = 0$, angular momentum conservation requires the spins of the two leptons to combine to a total spin of zero. In the limit of massless leptons, the required helicity configuration is forbidden by the chiral structure of the weak interaction. A helicity flip proportional to the lepton mass is therefore needed, resulting in the characteristic suppression of the decay rate by a factor $m_\ell^2/M_B^2$. The figure is taken from [6].

The helicity suppression becomes less severe as the lepton mass increases. As a consequence, the SM predictions for the branching fractions of the decays $B^0 \to \ell^+\ell^-$ span several orders of magnitude depending on the lepton flavor. The most precise theoretical predictions currently available are [7]

$$\mathcal{B}(B^0 \to e^+e^-)_{\text{SM}} = (2.48 \pm 0.21) \times 10^{-15}, \tag{27}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-)_{\text{SM}} = (1.06 \pm 0.09) \times 10^{-10}, \tag{28}$$
$$\mathcal{B}(B^0 \to \tau^+\tau^-)_{\text{SM}} = (2.22 \pm 0.19) \times 10^{-8}. \tag{29}$$

The dominant theoretical uncertainties originate from the CKM matrix elements and the decay constant $f_B$[7]. Owing to the helicity suppression factor proportional to $m_\ell^2$, the branching fractions increase by approximately seven orders of magnitude between the electron and tau channels. As a consequence of helicity suppression, the branching fraction scales strongly with the lepton mass, leading to substantial differences between the electron, muon, and tau channels, even though lepton flavor universality is preserved in the underlying weak interactions.

## 2.2 Effects from physics beyond the standard model

The strong suppression of the decay $B^0 \to \ell^+\ell^-$ in the SM makes it particularly sensitive to contributions from physics beyond the SM. New particles can contribute to the loop diagrams mediating the decay, modifying the Wilson coefficients of the effective Hamiltonian or generating additional scalar and pseudoscalar operators. Examples include extended Higgs sectors, such as two-Higgs-doublet models (2HDM). These new contributions can significantly enhance the branching fraction with respect to the SM prediction.

In the type-II 2HDM, the Higgs sector is extended by the introduction of a second scalar doublet, leading to additional charged and neutral Higgs bosons, $H^\pm$ and $A^0$. These particles can contribute to the loop diagrams mediating the decay $B^0 \to \ell^+\ell^-$, as illustrated in Fig.3. In particular, they generate additional scalar and pseudoscalar contributions to the effective Hamiltonian through the Wilson coefficients $C_S$ and $C_P$, while the coefficient $C_A$ remains unchanged. As a result, the branching fraction can be significantly enhanced with respect to the SM prediction. The magnitude of this enhancement depends mainly on the charged-Higgs mass $M_{H^\pm}$ and on the ratio of the vacuum expectation values of the two Higgs doublets, $\tan\beta$ [2]

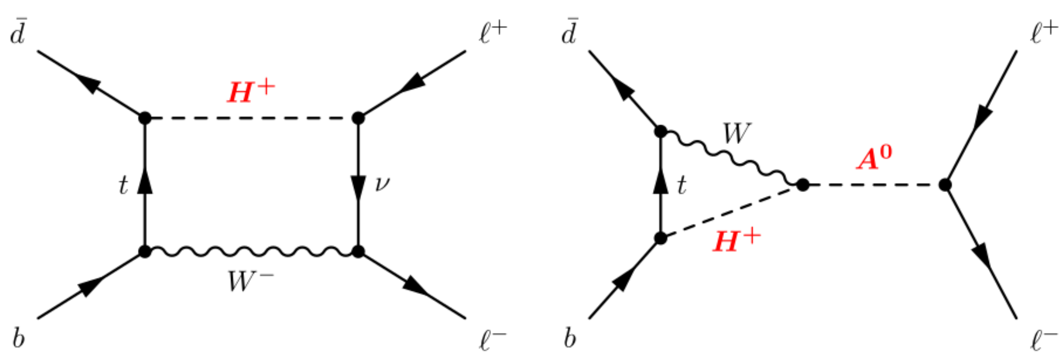



Figure 3: Representative contributions to the decay $B^0 \to \ell^+\ell^-$ in the type-II two-Higgs-doublet model (2HDM). Left: box diagram with the charged Higgs boson $H^\pm$ replacing one of the $W$ bosons. Right: penguin diagram involving the neutral Higgs boson $A^0$ and a charged Higgs boson. These additional contributions generate scalar and pseudoscalar operators that can enhance the branching fraction with respect to the SM prediction.

In summary, the strong suppression of $B^0 \to \tau^+\tau^-$ in the SM makes this decay an excellent probe of effects from physics beyond the SM. Precise measurements and improved experimental limits on this decay therefore provide powerful constraints on extensions of the SM.

## 2.3 Previous searches

To date, the most stringent published experimental limit on the decay $B^0 \to \tau^+\tau^-$ has been obtained by the LHCb Collaboration [8]. The search was performed using proton–proton collision data collected during Run 1 of the LHC corresponding to 3 fb$^{-1}$ and targeted final states in which both $\tau$ leptons decay hadronically through $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$. No significant excess above the expected background was observed, leading to an upper limit on the branching fraction of

$$\mathcal{B}(B^0 \to \tau^+\tau^-) < 1.6 \times 10^{-3}, \tag{30}$$

at the 90% confidence level. This result improves upon the previous limit set by the BaBar Collaboration[9] using the $\Upsilon(4S)$ data sample (210 fb$^{-1}$) and a hadronic tagging technique,

$$\mathcal{B}(B^0 \to \tau^+\tau^-) < 4.1 \times 10^{-3}, \tag{31}$$

at the 90% confidence level. Despite this improvement, the experimental sensitivity remains about five orders of magnitude above the Standard Model prediction, which is of the order of $10^{-8}$. The substantially larger data sample expected at Belle II, together with improved detector performance and advanced reconstruction techniques such as the Full Event Interpretation (FEI), will significantly help to enhance the sensitivity to this decay and provide a powerful probe of physics beyond the Standard Model.

# 3 The Belle II experiment

The Belle II experiment is located at the High Energy Accelerator Research Organisation (KEK) in Tsukuba, Japan. It operates at the SuperKEKB electron–positron collider and has been collecting data since 2018, with a target integrated luminosity of 50ab$^{-1}$. Belle II is dedicated to the study of B mesons, which are mainly produced through the $\Upsilon(4S)$ resonance. In this chapter the experimental setup of the SuperKEKB accelerator facility and the Belle II detector is described.

## 3.1 The $\Upsilon(4S)$ Resonance

The Upsilon meson ($\Upsilon$) is a neutral flavorless bound state of $\mathrm{b\overline{b}}$ quarks. It can be produced directly in $e^+e^-$ collisions. Several resonances exist, corresponding to different excited states of the $\mathrm{b\overline{b}}$ system, among which, the $\Upsilon(4S)$ resonance plays a central role in B physics. Its mass is slightly above the $\mathrm{B\overline{B}}$ production threshold, causing it to decay 96% of the cases into $B^0\overline{B^0}$ and $B^+B^-$ pairs. The cross section for $e^+e^- \to q\,\overline{q}$ with $q = u,\ d,\ s,\ c$, is shown Fig. 4. These non-resonant processes are referred to as continuum background in B analysis. Due to the small mass difference between the $\Upsilon(4S)$ resonance and the $\mathrm{B\overline{B}}$ production threshold, the decay produces only two B mesons with no additional particles. This clean environment makes it possible, using dedicated analysis techniques, to study B-meson decays containing multiple neutrinos in the final state.

## 3.2 SuperKEKB Accelerator

The SuperKEKB accelerator is an asymmetric-energy electron–positron collider operating at a center-of-mass energy of $\sqrt{s} = 10.58$ GeV, which corresponds to the mass of the $\Upsilon(4S)$ resonance. It consists of two storage rings with a circumference of approximately 3 km: the high energy ring (HER), which stores 7 GeV electrons, and the low energy ring (LER), which stores 4 GeV positrons. The facility also

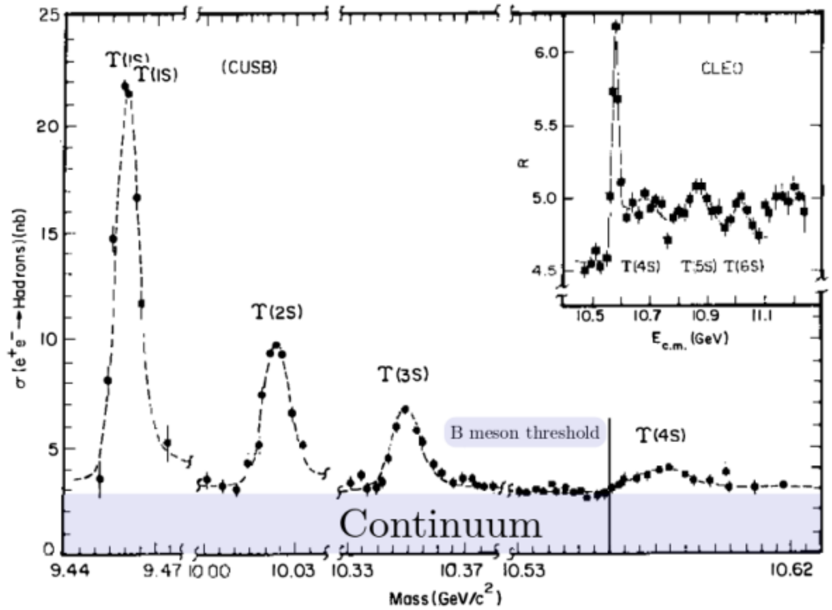



Figure 4: Cross section for $e^+e^- \to$ hadrons in the energy region of the $\Upsilon(1S)$ - $\Upsilon(4S)$ resonances. The vertical line indicates the $\mathrm{B}\overline{\mathrm{B}}$ production threshold. The continuum corresponds to the non-resonant process $e^+e^- \to q\,\overline{q}$ with $q = u,\ d,\ s,\ c$, while the peaks arise from the production of $\Upsilon$ resonances. The figure is taken from [10].

includes an injector linear accelerator (linac) and a 1.1 GeV positron damping ring (DR), as shown in Fig. 5. The entire complex is located about eleven meters below ground level. The design luminosity of SuperKEKB is $8 \times 10^{35}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, about 40 times higher than the peak luminosity achieved by KEKB. Its physics program focuses on precision tests of the Standard Model and indirect searches for New Physics through the study of rare decays, CP violation, and flavor-changing processes.

### 3.3 Belle II Detector

The Belle II detector consist of several sub-detectors which are arranged around the beam axis. The most important components of the Belle II detector are described below. A schematic overview is shown in Fig. 6.

- **The vertex detector system (VXD)**

  The VXD is the innermost detector surrounding the interaction point. It consists of two layers of silicon pixel detectors (PXD) and four layers of double-sided silicon strip detectors (SVD). Its primary purpose is the precise reconstruction of charged-particle trajectories close to the interaction point, allowing accurate determination of decay vertices and impact parameters. This capability is essential for identifying short-lived particles such as $B$ mesons, $D$ mesons, and $\tau$ leptons.

- **The central drift chamber (CDC)**

  The CDC is a large cylindrical gaseous detector located outside the VXD. It consists of approximately 14,000 sense wires arranged in several layers filled with a helium–ethane gas mixture. The CDC measures the momentum of charged particles through the curvature of their trajectories in the 1.5 T magnetic field generated by the superconducting solenoid. In addition, it provides information on the specific ionization energy loss ($dE/dx$), which contributes to particle identification.

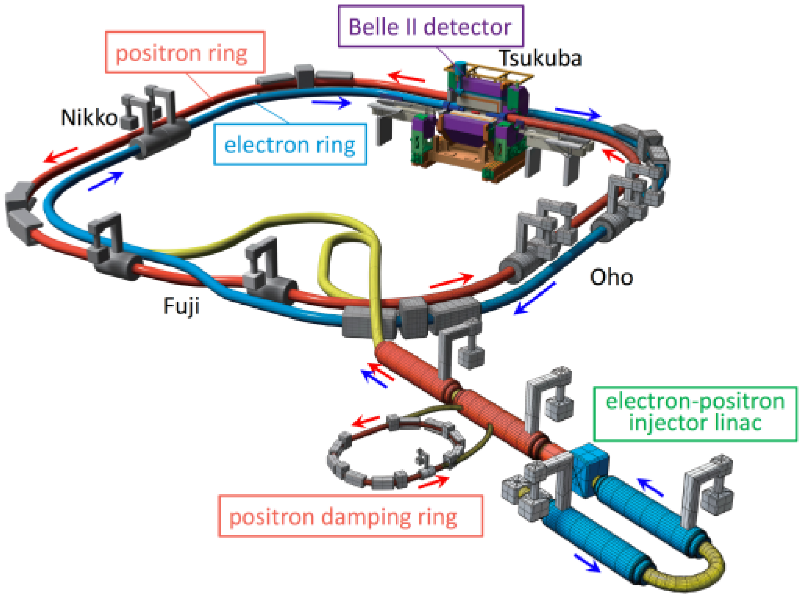



Figure 5: Schematic overview of SuperKEKB showing the high energy electron ring (HER, blue), the low energy positron ring (LER, red), the injector linac (green), and the positron damping ring. The storage rings include four main straight sections Tsukuba, Oho, Fuji, and Nikko while the electron and positron beams collide at the interaction point inside the Belle II detector (purple), located in the Tsukuba section. The figure is taken from [11].

- **The time of propagation system (TOP)**

  The TOP detector is the particle identification system used in the barrel region. It consists of quartz radiator bars in which charged particles emit Cherenkov photons. These photons propagate through the quartz by total internal reflection before being detected by micro-channel plate photomultipliers (MCP-PMTs). The arrival times and spatial distributions of the photons are used to reconstruct the Cherenkov angle, allowing efficient separation between pions, kaons, and protons over a wide momentum range.

- **The aerogel ring imaging Cherenkov system (ARICH)**

  The ARICH detector provides particle identification in the forward endcap region. It employs layers of silica aerogel radiators with different refractive indices to generate Cherenkov light. The emitted photons form characteristic ring patterns that are detected by hybrid avalanche photodetectors (HAPDs). By measuring the Cherenkov angle, the ARICH distinguishes between charged hadrons, particularly pions and kaons.

- **The electromagnetic calorimeter (ECL)**

  The ECL consists of 8736 CsI(Tl) scintillating crystals arranged in a barrel and two endcaps. It is designed to measure the energies and positions of electromagnetic showers produced by electrons, positrons, and photons. The ECL plays a crucial role in photon reconstruction, electron identification, and the reconstruction of neutral mesons such as $\pi^0$ and $\eta$. It also contributes to the determination of missing energy in events containing neutrinos.

- **The $K_L$ and muon detector (KLM)**

  The KLM forms the outermost layer of Belle II and is located outside the superconducting solenoid.

In the barrel region, it uses resistive plate chambers (RPCs), while the endcaps and inner barrel layers employ scintillator strips. The KLM identifies muons through their ability to penetrate large amounts of material and detects long-lived neutral kaons ($K_L$) via their hadronic interactions in the detector material.

Charged particles are first reconstructed using the VXD and CDC tracking detectors. Particle identification is then performed using information from the CDC, TOP, and ARICH systems. Electrons and photons deposit most of their energy in the ECL, while muons and $K_L^0$ mesons are identified using the KLM detector.

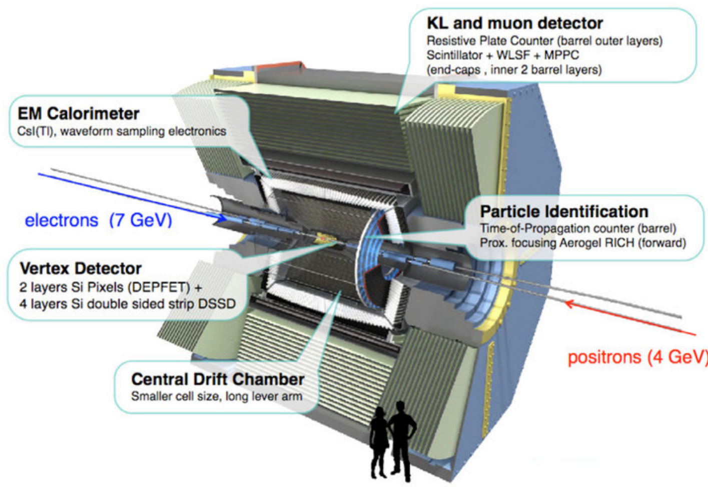



Figure 6: Schematic overview of the Belle II detector. The figure is taken from [11].

Despite the sophisticated design of the Belle II detector, only a limited number of particles can be observed directly. Most unstable particles produced in the collision decay long before reaching any detector component. As an example, the $\tau$ lepton has a mass of approximately $1.777\,\mathrm{GeV}/c^2$ and carries an electric charge of $\pm e$. The extremely short lifetime of the $\tau$ lepton further illustrates this challenge. The mean proper decay length of the $\tau$ is only

$$c\tau \simeq 87\ \mu\mathrm{m}, \tag{32}$$

whereas the first layer of the Belle II pixel detector is located at a radius of about 14 mm from the interaction point. This distance is more than two orders of magnitude larger than the characteristic $\tau$ decay length. As a result, $\tau$ leptons decay long before reaching any detector component and cannot be observed directly. Their reconstruction relies exclusively on the detection of their decay products and on the missing-energy signature associated with the neutrinos produced in the decay.

The same situation occurs for many heavy particles, including $B$ mesons, $D$ mesons, and numerous short-lived resonances. Their presence is inferred indirectly from the properties of their decay products and from the reconstruction of their decay vertices and invariant masses.

Table 1 summarizes the stable or sufficiently long-lived particles that can be detected directly in Belle II, together with the detector subsystems primarily used for their reconstruction and identification.

Table 1: Belle II subdetectors involved in the reconstruction and identification of different particle species.

| Particle | Main subdetectors |
|---|---|
| Electron ($e^{\pm}$) | VXD, CDC, TOP/ARICH, ECL |
| Photon ($\gamma$) | ECL |
| Pion ($\pi^{\pm}$) | VXD, CDC, TOP/ARICH |
| Kaon ($K^{\pm}$) | VXD, CDC, TOP/ARICH |
| Muon ($\mu^{\pm}$) | VXD, CDC, ECL, KLM |
| Proton ($p$) | VXD, CDC, TOP/ARICH |
| Neutral kaon ($K_L^0$) | KLM, ECL |

For a detailed description of the Belle II detector and its performance, the reader is referred to the Belle II Technical Design Report [12].

# 4 Analysis Procedure

The search for the rare decay $B^0 \to \tau^+\tau^-$ is experimentally challenging. Depending on the decay modes of the two $\tau$ leptons, the final state contains between two and four neutrinos. Since neutrinos escape detection, a significant fraction of the event kinematics cannot be measured directly. As a result, the signal decay cannot be fully reconstructed from the observed final-state particles alone.

To overcome this difficulty, the analysis exploits the clean environment provided by the Belle II experiment operating at the $\Upsilon(4S)$ resonance. Since the $\Upsilon(4S)$ decays almost exclusively into a pair of $B$ mesons and no additional particles are produced in the primary interaction, the reconstruction of one $B$ meson provides valuable constraints on the properties of the second one. The Full Event Interpretation (FEI) algorithm is therefore used to reconstruct one of the two B mesons [13], referred to as the tag side candidate $B_{\text{tag}}$. In the hadronic tagging approach, the $B_{\text{tag}}$ meson is reconstructed exclusively through hadronic decay channels. Since only fully reconstructed hadronic final states are considered, no neutrinos are present on the tag side, allowing the four momentum of the $B_{\text{tag}}$ candidate to be determined with good precision. The four momentum of the signal side meson, $B_{\text{sig}}$, can then be inferred from energy momentum conservation.

After the reconstruction of the tag side candidate, all remaining reconstructed particles are assigned to the signal side. Dedicated selection criteria are applied to identify the visible decay products of the two $\tau$ leptons and suppress the dominant background sources. These backgrounds originate primarily from generic $B\bar{B}$ decays and continuum processes $e^+e^- \to q\bar{q}$, where $q = u, d, s, c$. Particular attention is devoted to rejecting events containing additional particles incompatible with the signal hypothesis, as well as backgrounds arising from misidentified leptons and hadrons.

The selected events are subsequently divided into several categories according to the visible decay products of the two $\tau$ leptons. This categorization allows the analysis to exploit the specific kinematic properties of each final state while accounting for their different background compositions. Separate treatments are therefore performed for purely leptonic, semi leptonic, and hadronic decay channels.

Although the event selection significantly reduces the background contamination, the expected signal yield remains extremely small compared to the remaining background contribution. Simple one dimensional selections are therefore insufficient to achieve the required level of background rejection. To improve the discrimination power, a multivariate tool based on Boosted Decision Trees (BDTs) is employed. The classifiers combine information from a large set of observables describing the missing momentum, event topology, vertex properties, and kinematic characteristics of the reconstructed particles.

Dedicated BDTs are trained independently for each signal category using simulated signal and background samples. Their hyperparameters and input variable sets are optimized using the Optuna framework [14]. The performance of the resulting classifiers is evaluated using Receiver Operating Characteristic (ROC) curves, which represent the signal efficiency as a function of the background acceptance for all possible classifier thresholds, and by the corresponding Area Under the Curve (AUC), which provides a threshold independent measure of the classifier's discrimination power. The optimized classifiers provide a single discriminating variable that maximizes the separation between signal and background events. The final selection on the BDT output is subsequently determined by maximizing the Punzi figure of merit.

The events passing this final selection are subsequently used in a fit to extract the signal yield. The fitted yields are finally used to evaluate the expected sensitivity of the analysis and, in the presence of a signal, to determine the corresponding branching fraction. The overall workflow of the analysis can be summarized as follows:

$\Upsilon(4S)$ production $\rightarrow$ FEI hadronic tagging $\rightarrow$ event selection $\rightarrow$ channel categorization $\rightarrow$ BDT training and optimization $\rightarrow$ signal selection and BCS $\rightarrow$ signal extraction.

The following sections describe each of these steps in detail, starting with the data samples used throughout the analysis and the event-selection criteria applied to reconstruct signal candidates.

## 4.1 Data Samples

The analysis is based on the Belle II Run 1 and ongoing run 2 data sample collected at the $\Upsilon(4S)$ resonance between 2019 and end 2024, corresponding to an integrated luminosity of $\sim 490$ fb$^{-1}$. At this center-of-mass energy, the $\Upsilon(4S)$ resonance decays almost exclusively into pairs of B mesons, providing a clean environment for the study of rare B meson decays. The final analysis will also use the Belle sample, corresponding to an integrated luminosity of $\sim 710$ fb$^{-1}$.

Monte Carlo (MC) simulated samples are used to develop and validate the analysis strategy, estimate reconstruction efficiencies, optimize the event selection, and model the expected background contributions. Three categories of MC samples are considered:

- **Signal MC**, corresponding to simulated ($B^0 \rightarrow \tau^+\tau^-$) decays;
- **Generic $B\bar{B}$ MC**, which represents the dominant background originating from B meson decays produced through $\Upsilon(4S) \rightarrow B\bar{B}$;
- **Continuum MC**, consisting of $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$) and $e^+e^- \rightarrow \tau^+\tau^-$ processes.

The simulated background samples correspond to approximately four times the integrated luminosity of the data sample. This larger effective luminosity reduces the statistical uncertainty associated with the MC predictions and improves the modelling of background distributions. For the signal, a dedicated sample of approximately 60 millions $B^0 \rightarrow \tau^+\tau^-$ decays was generated, with the $\tau$ leptons allowed to decay generically according to their known branching fractions.

Throughout this analysis, Monte Carlo samples are used for the training and optimization of the multivariate classifiers, while the final signal extraction will be later performed using data.

# 5 Event Selection

The search for the decay $B^0 \to \tau^+\tau^-$ is performed using the hadronic tagging technique provided by the Full Event Interpretation (FEI) algorithm. After the reconstruction of a $B_{\text{tag}}$ candidate, all remaining reconstructed particles in the event are assigned to the signal side and used to reconstruct the $B_{\text{sig}}$ candidate. Since the decay contains at least two neutrinos and up to four neutrinos depending on the $\tau$ decay modes, the signal cannot be fully reconstructed. The event selection therefore aims at maximizing the signal efficiency while suppressing continuum and generic $B\bar{B}$ backgrounds.

## 5.1 Tag-side reconstruction

One of the two $B$ mesons produced in the $\Upsilon(4S)$ decay is reconstructed using the hadronic FEI algorithm (Fig.7). The FEI reconstructs hundreds of exclusive decay chains and assigns a signal probability to each candidate based on a multivariate classification procedure.

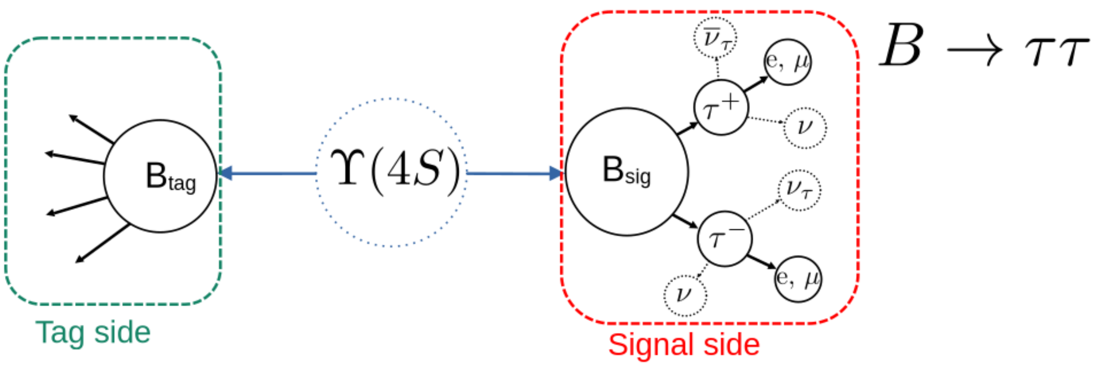



Figure 7: Illustration of the hadronic tagging strategy employed in the search for $B^0 \to \tau^+\tau^-$ decays at Belle II. The $\Upsilon(4S)$ resonance produces a pair of $B$ mesons, one of which is fully reconstructed as $B_{\text{tag}}$ using the FEI algorithm. The remaining particles are assigned to the signal-side candidate $B_{\text{sig}}$, which decays into a $\tau^+\tau^-$ pair. The subsequent $\tau$ decays produce visible charged particles and multiple neutrinos, resulting in significant missing energy in the final state.

The hadronic tagging technique provides a powerful method for studying decays involving multiple neutrinos in the final state. Since the four momentum of the initial $e^+e^-$ system is precisely known and the $B_{\text{tag}}$ meson is fully reconstructed, the four-momentum of the signal side meson can be inferred from energy momentum conservation. This information makes it possible to construct variables related to the missing energy and missing momentum of the event, which are essential for discriminating signal from background.

To ensure a well reconstructed tag-side candidate and to remain consistent with the available FEI calibration factors, the standard FEI selection criteria recommended for the latest MC campaign are applied.

The quality of the reconstructed $B_{\text{tag}}$ candidate is assessed using three quantities. The beam-constrained mass,

$$M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - |\vec{p}_{B\text{tag}}|^2}, \tag{33}$$

is computed in the center of mass frame using the beam energy $E_{\text{beam}}$ and the reconstructed momentum of the $B_{\text{tag}}$ candidate. Correctly reconstructed B mesons are expected to peak near the nominal B meson mass. The energy difference,

$$\Delta E = E_{B_{\text{tag}}} - E_{\text{beam}}, \tag{34}$$

measures the consistency between the reconstructed candidate energy and the beam energy, and is expected to be close to zero for correctly reconstructed candidates. Finally, FEIProb is the output of the final multivariate classifier of the Full Event Interpretation algorithm. This variable combines information from the entire reconstructed decay chain and provides a measure of the quality of the $B_{\mathrm{tag}}$ candidate. Higher values correspond to candidates that are more likely to be correctly reconstructed. The reconstructed $B_{\mathrm{tag}}$ candidate is required to satisfy

$$M_{\mathrm{bc}} > 5.27\ \mathrm{GeV}/c^2, \tag{35}$$

$$-0.15\ \mathrm{GeV} < \Delta E < 0.10\ \mathrm{GeV}, \tag{36}$$

$$\mathrm{FEIProb} > 0.001. \tag{37}$$

Several signal and tag side candidates may be reconstructed in the same event. To avoid multiple counting and simplify the statistical interpretation of the sample, a single candidate is retained for each event.

The selected candidate corresponds to the $B_{\mathrm{tag}}$ reconstruction with the highest FEI probability. This choice maximizes the probability of selecting the correctly reconstructed tag-side candidate and provides a unique event representation for the subsequent analysis steps.

After the reconstruction of the $B_{\mathrm{tag}}$ candidate, all remaining reconstructed particles are assigned to the signal side. This collection of particles, commonly referred to as the Rest Of Event (ROE), is used to reconstruct the visible decay products of the two $\tau$ leptons. Signal candidates are required to be consistent with one of the decay channels described in Section 5.6.

## 5.2 Signal-side charged particle selection

After the reconstruction of the tag side candidate, the remaining charged particles are used to reconstruct the visible decay products of the two $\tau$ leptons. All charged tracks are required to originate from the vicinity of the interaction point. The following requirements are applied:

- $|dz| < 4$ cm,
- $dr < 2$ cm,
- track reconstructed within the CDC acceptance,
- momentum $p > 0.2$ GeV/$c$.

Particle identification is based on the Belle II Deep Neural Network (DNN) classifiers. Dedicated criteria are applied to select electrons, muons and charged pions while suppressing particle misidentification.

Muon candidates are required to satisfy

$$\mathrm{muID} > 0.9, \tag{38}$$

Electron candidates are selected through

$$\mathrm{euID} > 0.9. \tag{39}$$

while simultaneously failing the muon selection

$$\text{muID} < 0.9. \quad (40)$$

Charged pion candidates are required to satisfy

$$\text{eID} < 0.9, \qquad \text{muID} < 0.9, \qquad \text{piID} > 0.1. \quad (41)$$

These requirements define mutually exclusive particle categories and reduce cross-feed among the different $\tau$ decay channels.

## 5.3 Neutral particle reconstruction

Neutral pions are reconstructed from pairs of photons belonging to the standard Belle II particle list. The invariant mass of the photon pair is required to be compatible with the nominal $\pi^0$ mass,

$$0.121 < M_{\gamma\gamma} < 0.142 \text{ GeV}/c^2. \quad (42)$$

Additional requirements are imposed on the photon energy, cluster timing, cluster multiplicity and polar angle in the electromagnetic calorimeter in order to suppress beam backgrounds and poorly reconstructed clusters.

The reconstructed $\pi^0$ candidates are subsequently used for the reconstruction of $\rho^\pm \to \pi^\pm\pi^0$ candidates. In addition, they are employed in the $\pi^0$ veto procedure described in the next section, which is designed to suppress background events containing neutral pions.

## 5.4 Rest-of-event requirements

The hadronic tagging technique provides access to the Rest Of Event (ROE), defined as all reconstructed particles not associated with the selected $B_{\text{tag}}$ and $B_{\text{sig}}$ candidates. The ROE is used to suppress backgrounds that contain additional particles not expected in signal events.

The following requirements are applied:

- no additional $K_S^0$ candidates in the ROE,
- no additional $\Lambda$ candidates in the ROE,
- no additional $\pi^0$ candidates in the ROE.

These criteria significantly reduce backgrounds from generic $B\bar{B}$ decays containing additional neutral hadrons.

## 5.5 Background vetoes

Several dedicated vetoes are applied to suppress peaking backgrounds that can mimic the signal signature.

In channels containing electrons, photon conversions $\gamma \to e^+e^-$ are rejected. In addition, candidates compatible with $J/\psi \to e^+e^-$ decays are removed.

Similarly, channels containing muons are subjected to a veto against $J/\psi \to \mu^+\mu^-$ decays. These requirements suppress easy to reject background originating from charmonium decays while preserving a high signal efficiency.

## 5.6 Signal categories

The analysis is performed separately for different visible $\tau$ decay modes. This strategy takes advantage of the distinct kinematic properties and background compositions of the various final states.

The reconstructed channels are grouped into four categories:

- leptonic channels: $ee$, $e\mu$, $\mu\mu$,
- lepton–pion channels: $e\pi$, $\mu\pi$,
- lepton–rho channels: $e\rho$, $\mu\rho$,
- hadronic channels: $\pi\pi$, $\pi\rho$, $\rho\rho$.

Following this preselection, a multivariate classifier based on Boosted Decision Trees (BDTs) is trained independently for each category in order to further separate signal events from the remaining background contributions.

### 5.6.1 Background Processes

Several processes can mimic the signature of the $B^0 \to \tau^+\tau^-$ decay and therefore constitute important sources of background. The dominant contributions arise from generic $B\bar{B}$ decays produced through $\Upsilon(4S) \to B\bar{B}$ as well as continuum processes $e^+e^- \to q\bar{q}$, where $q = u, d, s, c$.

Generic $B\bar{B}$ events often contain semileptonic decays involving neutrinos, resulting in missing momentum signatures similar to those expected from the signal decay. In addition, the large multiplicity of particles produced in $B$-meson decays can lead to incorrectly reconstructed candidates that satisfy the signal selection criteria.

Continuum events represent another major source of background. In contrast to $B\bar{B}$ events, which are produced nearly at rest in the center of mass frame, continuum events exhibit a more jet like topology resulting from the hadronization of energetic quark pairs. Although they rarely mimic the full signal signature, their production cross section is significantly larger than that of $\Upsilon(4S)$ decays, making their rejection an essential part of the analysis.

The event selection and multivariate classification procedures are therefore designed to exploit the kinematic and topological differences between signal and background events in order to maximize the sensitivity to the $B^0 \to \tau^+\tau^-$ decay.

# 6 Multivariate Analysis

Although the event selection significantly reduces the background contribution, the expected signal yield remains extremely small compared to the remaining background. Additional discrimination is therefore obtained using a multivariate classifier based on Boosted Decision Trees (BDTs).

Separate classifiers are trained for each signal category using simulated signal and background samples. This section describes the multivariate analysis used to discriminate signal from background. It presents the selection of the input variables, the training of the Boosted Decision Tree classifiers, and the optimization of the final event selection. The performance of the classifiers is then assessed through their signal–background separation and the expected sensitivity of the analysis.

## 6.1 Training Samples

The signal sample used for the training is obtained from simulated $B^0 \to \tau^+\tau^-$ events, while the background sample consists of generic $B\bar{B}$ and continuum Monte Carlo events. To ensure that the classifier learns the characteristics of correctly reconstructed signal candidates, truth matching is applied to the signal sample. A candidate is considered truth matched when all reconstructed final-state particles are correctly associated with their generated counterparts.

The application of truth matching provides a clean signal sample for the training of the classifier. By restricting the training sample to correctly reconstructed signal candidates, the multivariate algorithm learns the characteristic features of genuine signal events rather than reconstruction artefacts arising from incorrect particle assignments.

Truth matching information is available only in simulated samples and cannot be used in data. Furthermore, events failing the truth-matching requirements are not discarded from the analysis itself. They are excluded only from the training procedure and remain part of the signal Monte Carlo sample used for efficiency studies and sensitivity estimates.

To assess the impact of this choice, a dedicated comparison between classifiers trained with and without truth-matching requirements is performed and discussed in section 6.4. This study allows us to quantify the effect of reconstruction ambiguities on the classifier performance and on the expected analysis sensitivity.

The selected events are divided into statistically independent training and testing samples. The training sample is used to construct the classifier, while the testing sample is used exclusively to evaluate its performance and verify the absence of overtraining.

## 6.2 Choice of Input Variables

The performance of a multivariate classifier strongly depends on the quality of the input observables. For the search for $B^0 \to \tau^+\tau^-$, the selected variables are designed to exploit the characteristic features of signal events arising from the presence of multiple neutrinos and from the production topology of $B\bar{B}$ events at the $\Upsilon(4S)$ resonance.

Several classes of observables are considered.

The first class consists of missing energy and missing momentum variables. Since the signal decay contains at least two neutrinos and up to four neutrinos depending on the reconstructed $\tau$ decay modes, a significant fraction of the event energy remains undetected. Variables such as the missing mass squared $M^2_{\text{miss}}$, the missing momentum magnitude $p_{\text{miss}}$, and the missing energy related quantity $U$, defined as

$$U = E_{\text{miss}} - |\vec{p}_{\text{miss}}|, \tag{43}$$

where $E_{\text{miss}}$ and $\vec{p}_{\text{miss}}$ are the missing energy and missing momentum computed in the center of mass frame, provide strong discrimination against backgrounds with fewer neutrinos.

A second class of observables characterizes the global event topology. Continuum processes $e^+e^- \to q\bar{q}$ typically produce jet like events, whereas $B\bar{B}$ events are more spherical due to the low momentum of the $B$ mesons in the center of mass frame. This difference is exploited through event-shape variables such as the Fox–Wolfram moments [15] and the event sphericity.

Additional discrimination is obtained from variables related to the reconstructed decay products of the two $\tau$ leptons. These include momenta, angular correlations, invariant masses, and vertex quantities.

Such observables provide sensitivity to the decay kinematics and help reject incorrectly reconstructed candidates.

Finally, several variables exploiting the inferred kinematics of the signal side $B$ meson are considered. These quantities are reconstructed using the information provided by the FEI tagged $B_{\text{tag}}$ candidate and can probe the consistency of an event with the signal hypothesis.

The following subsections discuss some of the most important variables used in the multivariate analysis and illustrate their discriminating power against the dominant background processes.

### 6.2.1 Missing Energy and Missing Momentum Variables

Figure 8 shows the distributions of the missing-mass squared variable $M^2_{\text{miss}}$ for signal and background events in the four analysis categories. This observable is computed from the missing four-momentum of the event, obtained through energy momentum conservation after reconstruction of the $B_{\text{tag}}$ candidate. The determination of the missing four-momentum is possible thanks to both the precise knowledge of the initial $e^+e^-$ collision four momentum and the reconstruction of the tag side B meson. The missing four momentum is therefore defined as

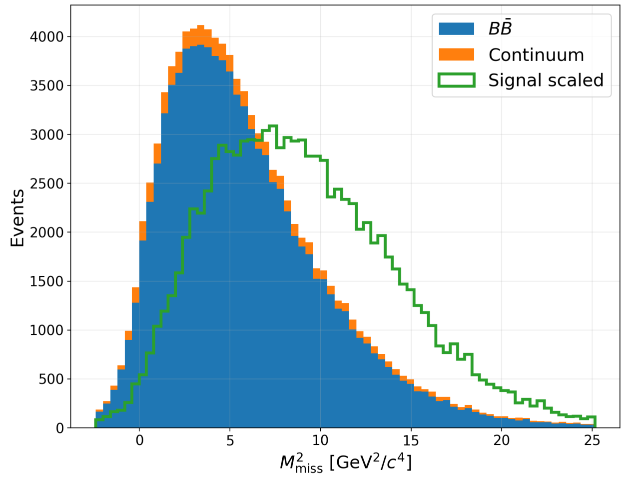
$$p_{\text{miss}} = p_{e^+e^-} - p_{B_{\text{tag}}} - \sum_i p_i, \tag{44}$$

where the sum runs over all reconstructed particles associated with the signal side decay.

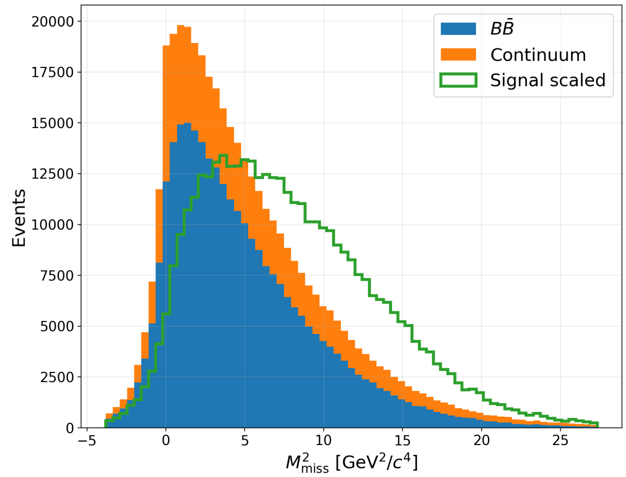
The variable $M^2_{\text{miss}}$ is particularly sensitive to the presence of undetected neutrinos. In the signal decay $B^0 \to \tau^+\tau^-$, each $\tau$ lepton decays into at least one neutrino, resulting in a final state containing between two and four neutrinos depending on the decay channel. Consequently, signal events are expected to exhibit larger missing energy and larger values of $M^2_{\text{miss}}$ than many background processes.

This behaviour is observed in all four categories, where the signal distributions are shifted towards larger values of $M^2_{\text{miss}}$ compared to the background. The effect is particularly pronounced in the purely leptonic channels, which contain the largest number of neutrinos in the final state. In contrast, hadronic channels involve fewer neutrinos and therefore exhibit a smaller separation.

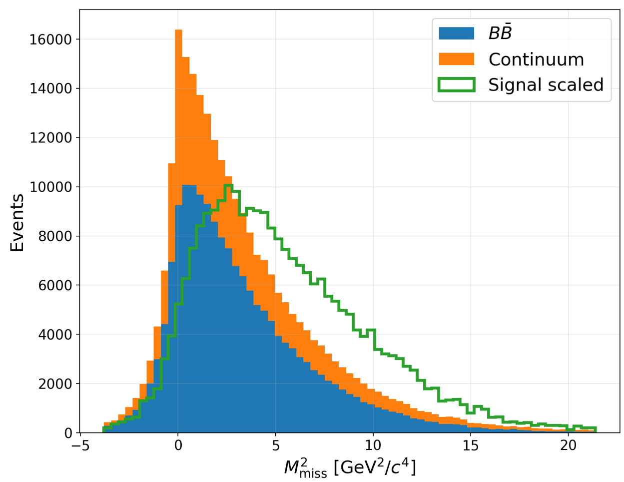
Although the overlap between signal and background remains significant, the missing-mass squared provides one of the most physically motivated observables of the analysis and constitutes an important input to the BDT classifier.



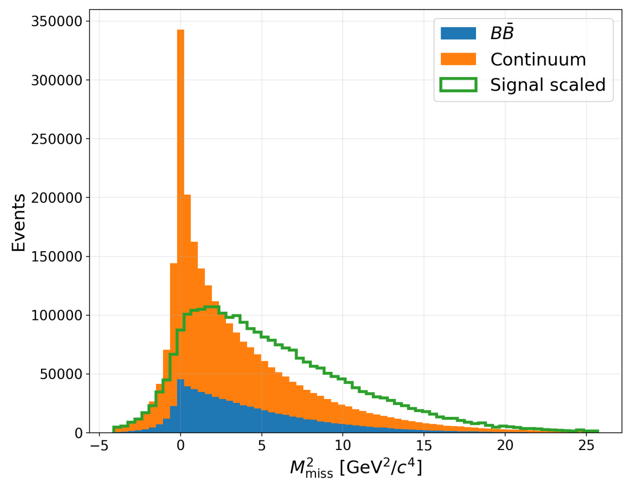

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

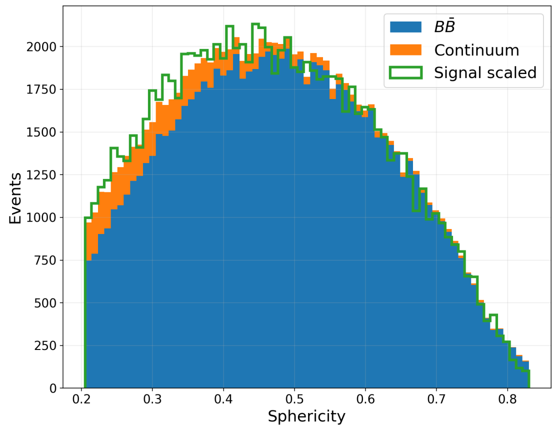
Figure 8: Distributions of the missing-mass squared variable $M^2_{\text{miss}}$ for signal and background events in the four analysis categories.

### 6.2.2 Event Shape Variables

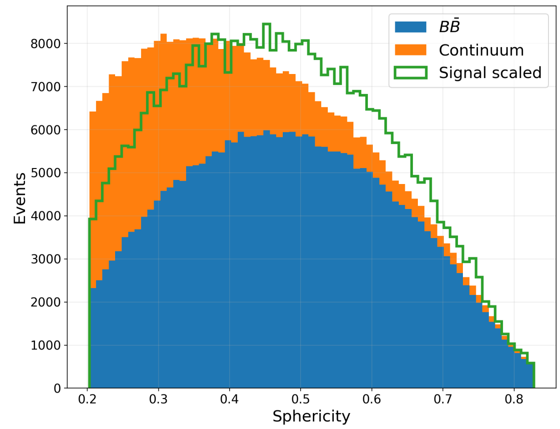
Figure 9 shows the distribution of the event sphericity for signal and background events in the four analysis categories. A cut at sphericity $> 0.2$ is applied as part of the FEI calibration requirements. This selection is introduced at the tagging stage and is therefore retained throughout the analysis. The sphericity is an event shape observable that quantifies how isotropic the reconstructed event topology is. Low values correspond to more collimated, jet like events, while larger values indicate a more spherical topology.

This variable is particularly useful for separating signal events from continuum background. Continuum processes, such as $e^+e^- \to q\bar{q}$, tend to produce two back-to-back jets in the center-of-mass frame and therefore populate the low-sphericity region. In contrast, signal events originate from $\Upsilon(4S) \to B\bar{B}$ production, where the two $B$ mesons are produced almost at rest. Their decay products are therefore distributed more isotropically, leading to larger sphericity values on average.

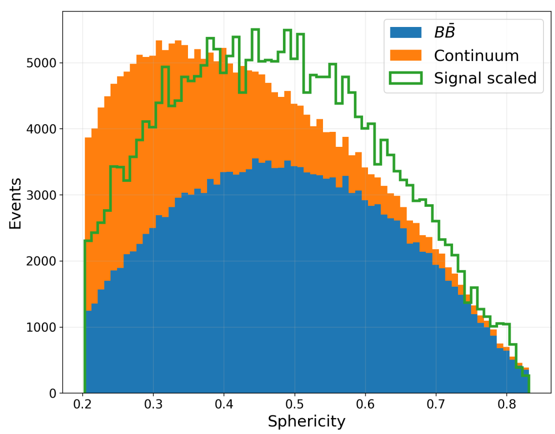
The separation is especially visible in the hadronic category, where the signal distribution is shifted towards higher sphericity compared to the background. In the other categories, the signal and background distributions overlap more strongly, indicating that sphericity alone is not sufficient to provide a strong discrimination. Nevertheless, it still contributes useful information when combined with the other input variables in the BDT classifier.



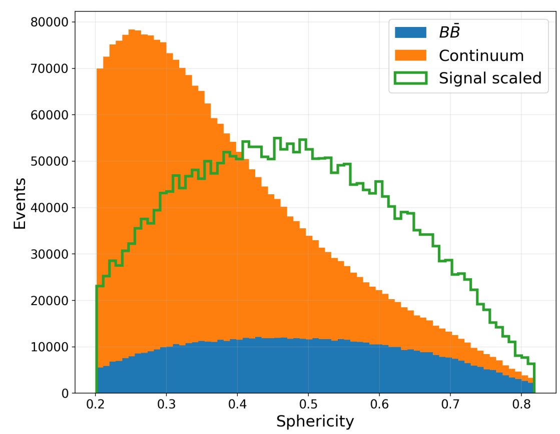

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

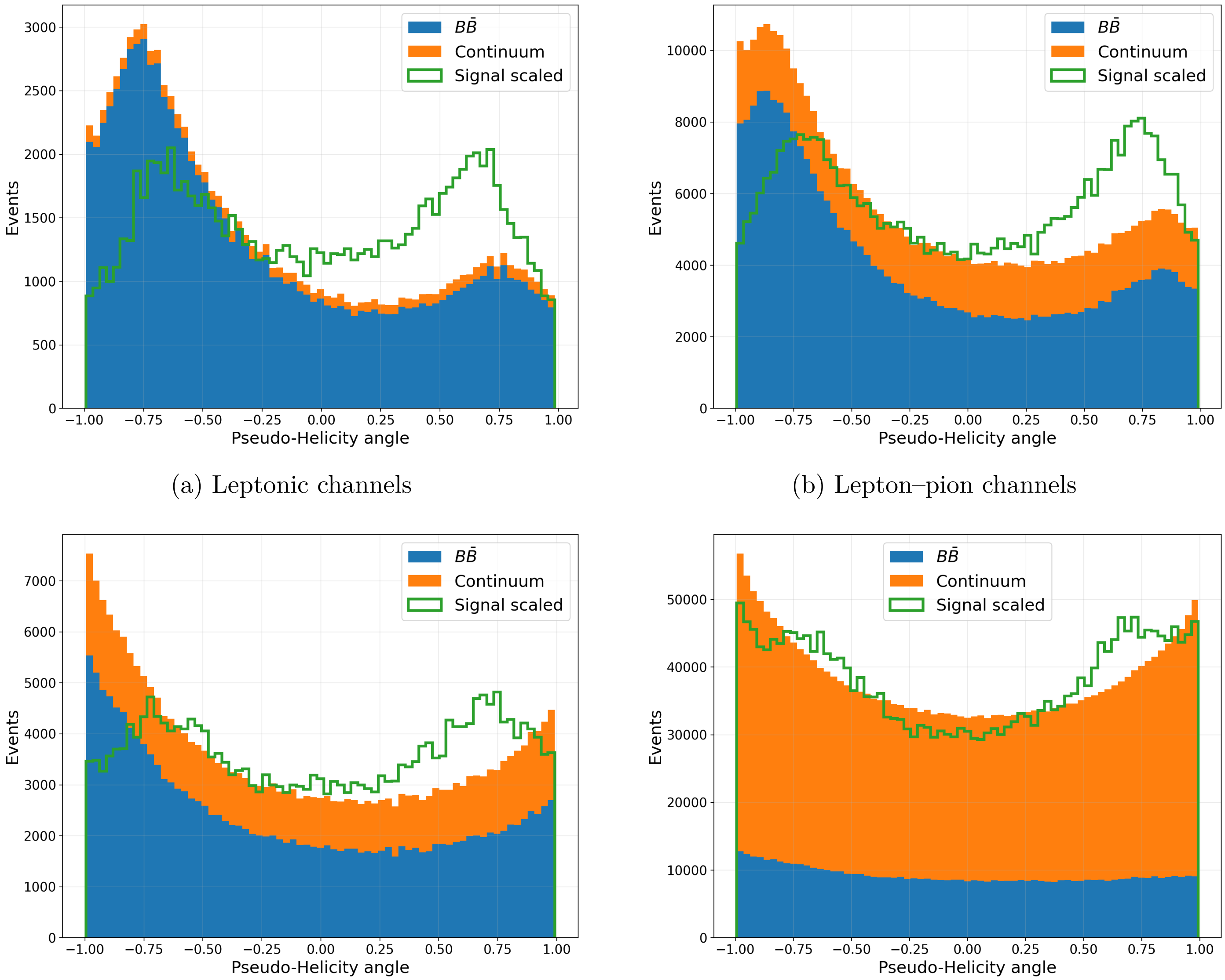
Figure 9: Distributions of the event sphericity for signal and background events in the four analysis categories.

### 6.2.3 Angular Variables

Among the other observables used in the multivariate analysis are pseudo-helicity variables constructed from the visible decay products of the two $\tau$ leptons (Fig.10). Since the neutrinos produced in the $\tau$ decays are not reconstructed, the true helicity angles cannot be determined experimentally. Instead, approximate angular observables are built using the reconstructed kinematics of the signal-side particles. For a given visible daughter particle, its four-momentum is boosted into the reconstructed $B_{\text{sig}}$ rest frame. The pseudo-helicity angle is then defined as the angle between the daughter momentum in this frame and the direction of the $B_{\text{sig}}$ meson in the center-of-mass frame,

$$\cos\theta_{pseudo\ hel} = \frac{\vec{p}_d \cdot \vec{p}_{B_{\text{sig}}}}{|\vec{p}_d||\vec{p}_{B_{\text{sig}}}|}. \tag{45}$$

This variable probes the angular structure of the decay and is sensitive to the kinematic correlations between the visible decay products and the missing momentum carried away by the neutrinos.




(a) Leptonic channels

(b) Lepton–pion channels

(c) Lepton–rho channels

(d) Hadronic channels

Figure 10: Distributions of the cosin of pseudo-helicity angle for signal and background events in the four analysis categories.

Since the signal process $B^0 \to \tau^+\tau^-$ produces two heavy leptons which subsequently decay into visible particles and multiple neutrinos, the resulting topology exhibits strong angular correlations. In the reconstructed $B_{\mathrm{sig}}$ rest frame, the visible decay products are preferentially emitted either along or opposite to the $B_{\mathrm{sig}}$ flight direction in the center-of-mass frame. As a result, signal events exhibit a characteristic double-peaked structure near PseudoHel $\approx \pm 1$, with relatively few events in the central region.

This feature originates from the two-body kinematics of the $B^0 \to \tau^+\tau^-$ decay. Although part of the $\tau$ momentum is carried away by neutrinos, the visible decay products remain strongly correlated with the $\tau$ direction. Depending on the decay kinematics, they are reconstructed either parallel or anti-parallel to the $B_{\mathrm{sig}}$ flight direction, producing the two peaks observed at the edges of the distribution.

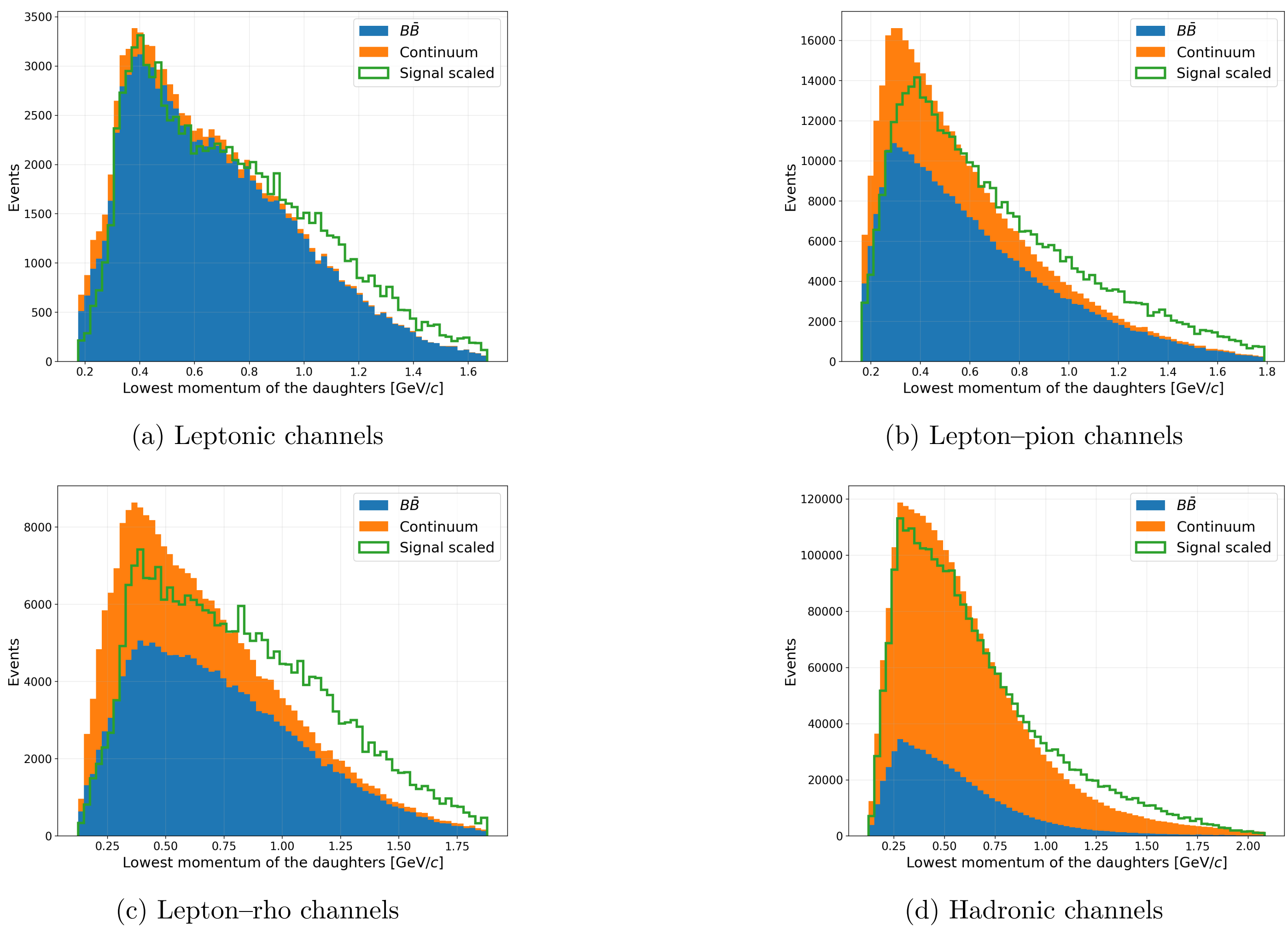
Background events, on the other hand, arise from multibody $B\bar{B}$ decays and continuum processes with more complex topologies. Their visible particles exhibit much weaker angular correlations with the reconstructed $B_{\mathrm{sig}}$ direction, leading to broader and smoother pseudo-helicity distributions without the pronounced double-peaked structure characteristic of the signal.

The separation is particularly visible for the higher-momentum daughter, where the correlation with the parent $\tau$ direction is stronger, making the pseudo-helicity variable a useful discriminator in the multivariate classification.

### 6.2.4 Kinematic Variables of the Visible Daughters

The visible decay products of the two $\tau$ leptons carry information about the underlying decay kinematics. To define observables that are independent of the particle ordering, the daughter particles are ranked according to their momentum. The momentum of the lower-momentum daughter and that of the higher-momentum daughter are then used as input variables.

Figures 11 and 12 show the momentum distributions of the lower and higher momentum visible daughters for signal and background events. In signal events, the visible particles originate from the decay of two $\tau$ leptons and are therefore constrained by the underlying two body decay kinematics. Background events, in contrast, arise from a wide range of $B\bar{B}$ decays and continuum processes, resulting in broader momentum distributions. The observed differences between signal and background distributions provide useful discrimination power and are exploited by the multivariate classifier.



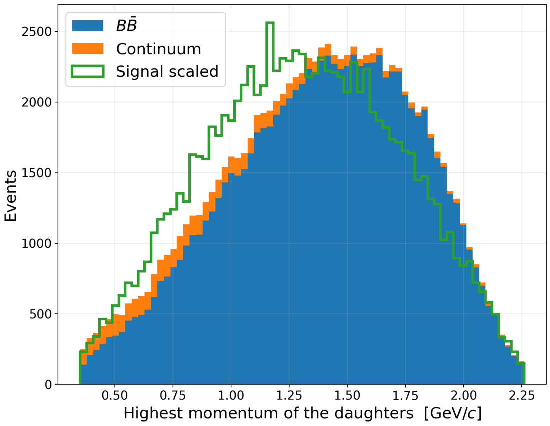

(a) Leptonic channels

(b) Lepton–pion channels

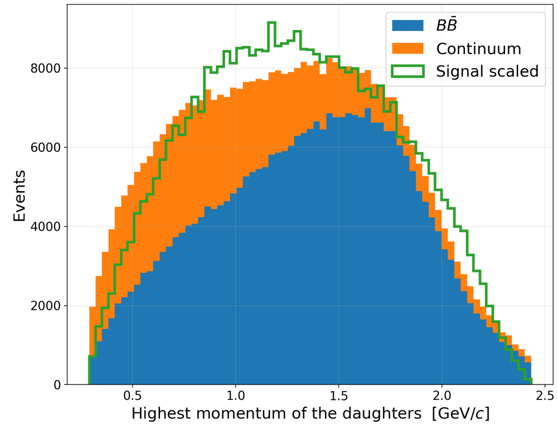
(c) Lepton–rho channels

(d) Hadronic channels

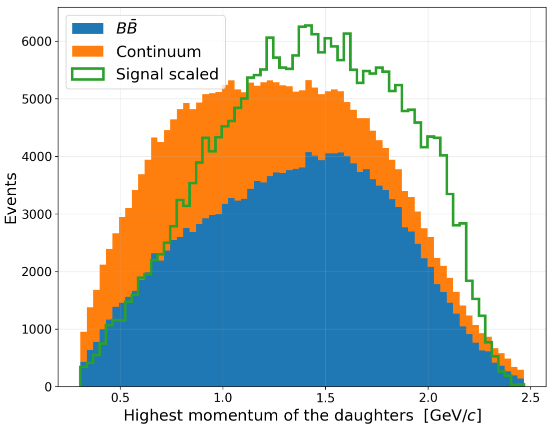
Figure 11: Distributions of the momentum of the lowest-momentum visible daughter particle, $p_{\text{low}}$, for signal and background events in the four analysis categories.



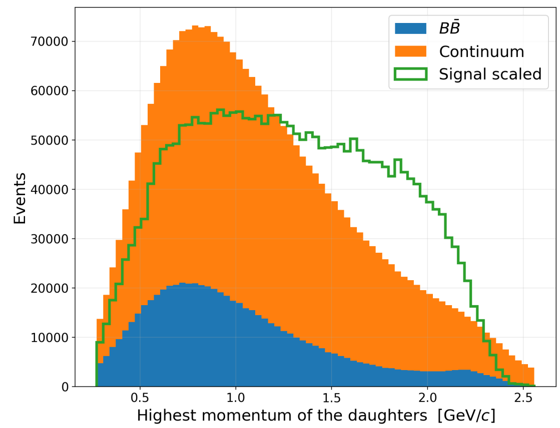

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

Figure 12: Distributions of the momentum of the highest-momentum visible daughter particle, $p_{\text{high}}$, for signal and background events in the four analysis categories.

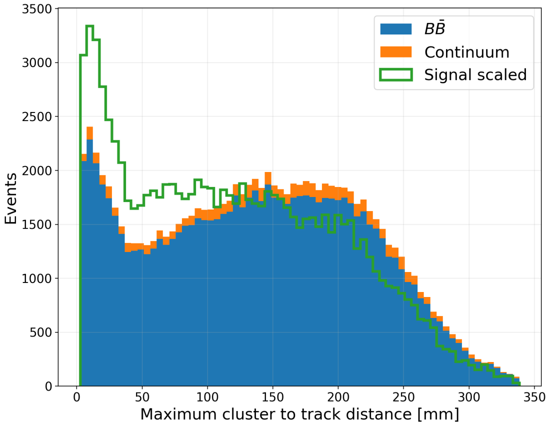
### 6.2.5 KLM variables

Figure 13 shows the distribution of the maximum distance between reconstructed charged particle trajectories and KLM clusters. For each event, the ten most energetic KLM clusters are considered and the largest cluster to track distance is retained.

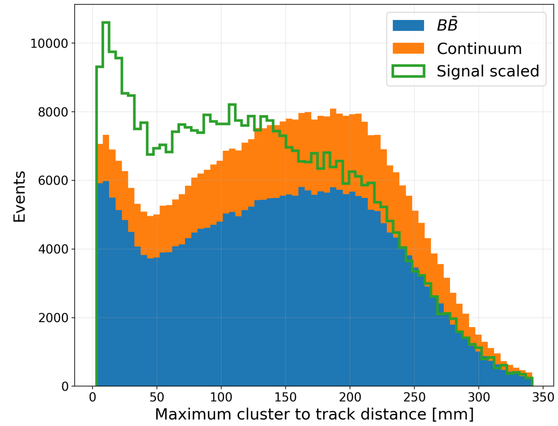
The KLM detector is primarily used for the identification of muons and long lived neutral kaons ($K_L^0$). In generic $B\bar{B}$ events, additional neutral hadrons frequently produce energy deposits in the KLM detector that are spatially separated from the reconstructed charged tracks. Such activity tends to generate large cluster to track distances.

In contrast, signal events are expected to contain only the visible products of the two $\tau$ decays together with multiple neutrinos. Consequently, the probability of observing isolated KLM activity unrelated to reconstructed charged particles is reduced.

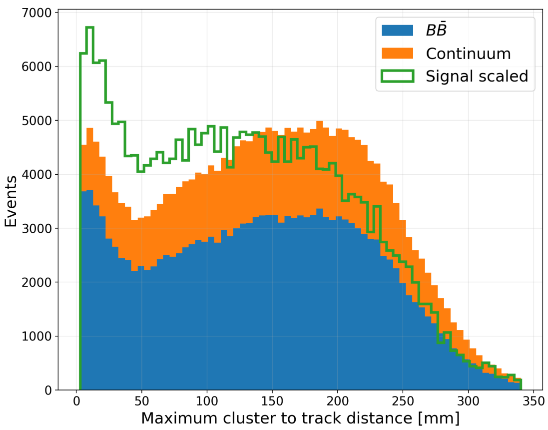
This behaviour is visible in (Fig. 13), where signal events exhibit an enhanced population at small cluster track distances and a reduced contribution at large values compared to the background. The effect is observed in all analysis categories and reflects the lower level of residual neutral hadron activity expected in signal events.



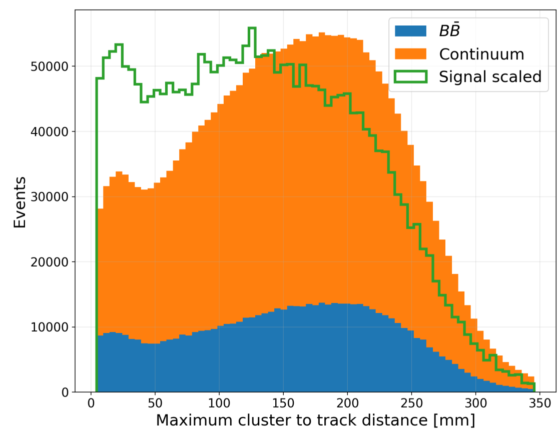

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

Figure 13: Distributions of the maximum KLM cluster–track distance for signal and background events in the four analysis categories.

#### 6.2.6 Summary of Selected Variables

The final set of input variables was designed to exploit several complementary aspects of the signal topology (Tab 2). Missing-energy observables are sensitive to the multiple neutrinos produced in the $\tau$ decays, while kinematic and angular variables characterize the visible decay products. Event shape variables help reject continuum backgrounds, and detector-level observables, such as KLM related variables, provide additional discrimination through the neutral hadron activity in the event.

Although none of these variables alone is sufficient to separate signal from background, their combination provides strong discrimination when used in a multivariate classifier. In particular, the missing-mass squared, pseudo helicity observables, visible particle momenta, and KLM variables were found to be among the most informative quantities according to the permutation importance studies showed below. The combination of all these observables within a multivariate classifier therefore provides a substantially improved signal to background discrimination compared to any individual variable.

Table 2: Main categories of observables used as inputs to the BDT classifier. Here, $\Delta z_{t_1 t_2}$ denotes the distance along the beam axis between the trajectories of the two visible daughter tracks, while the KLM distance variable corresponds to the maximum distance between a KLM cluster and the extrapolated charged-particle trajectory associated with the signal candidate. Finally $\cos \tau_{\mathrm{BMS}}$ denotes the cosine of the angle between the reconstructed $\tau$ direction and its visible decay products in the inferred $B$ meson rest frame.

| Category | Representative variables |
|---|---|
| Missing energy and momentum | $M^2_{\mathrm{miss}}$, $U$, $p_{\mathrm{miss}}$, $\cos\theta_{\mathrm{miss}}$, $p_{\mathrm{miss},d_0}$, $p_{\mathrm{miss},d_1}$ |
| Visible-particle kinematics | $d_{\mathrm{lowmom}}$, $d_{\mathrm{highmom}}$, $B^{\mathrm{CMS}}_{\mathrm{sgn}}\, p$, $d^{\mathrm{CMS}}_0\, p$, $d^{\mathrm{CMS}}_1\, p$, $d^{\mathrm{CMS}}_0\, p_T$, $d^{\mathrm{CMS}}_1\, p_T$ |
| Angular observables | $\cos(B_{\mathrm{sgn}}, d_0)$, pseudo-helicity, $\cos\tau_{\mathrm{BMS}}$ |
| Vertex-related quantities | significance of distance, vertex $\chi^2$ probability, $\Delta z_{t_1 t_2}$ |
| Invariant masses | $B_{\mathrm{sgn}}$ invariant mass, visible daughter invariant masses |
| Event topology | sphericity, Fox–Wolfram moments $R_2$ and $R_3$, KSFW variables |
| KLM observables | max_klmClusterTrackDistance, $n_{\mathrm{klmClusters}}$ |

#### 6.2.7 BDT Training

To exploit the discriminating power of the selected observables, a Boosted Decision Tree classifier is trained separately for each signal category. BDTs are supervised machine learning algorithms that combine a large number of decision trees into a single classifier. During the training process, successive trees are constructed to improve the classification of events that were incorrectly identified by the previous iterations.

In this analysis, the BDTs are implemented using the XGBoost package [16]. XGBoost is based on gradient boosting and includes regularization techniques that improve both the performance and the stability of the classifier. Its flexibility and efficiency have made it a standard choice for multivariate analyses in particle physics.

The classifier returns a continuous score between 0 and 1, where values close to 0 correspond to background-like events and values close to 1 correspond to signal like events. A selection on this score is then applied to retain events with a high signal probability while rejecting as much background as possible.

## 6.3 Variable Importance

After the training of the BDT classifier, the relative importance of the input variables is evaluated using a permutation based method. This approach provides a model independent way of estimating how much each observable contributes to the final discrimination power.

The primary metric used to assess the discrimination power of the classifier is the Receiver Operating Characteristic (ROC) curve, which represents the signal efficiency as a function of the background acceptance for all possible values of the BDT threshold, and the area under the ROC curve (AUC). Then, for a given input variable, its values are randomly shuffled among the events while all other variables are kept unchanged. This preserves the one dimensional distribution of the variable but destroys its correlation with the event class and with the other observables.

The trained classifier is then evaluated again on this modified sample. If the shuffled variable contains important discriminating information, the classifier performance decreases significantly. Conversely, if the variable is weakly informative or largely redundant with other observables, the decrease in performance remains small. The importance of each variable is quantified by the corresponding reduction in AUC.

The importance of a variable is therefore quantified as

$$I_i = \mathrm{AUC}_{\mathrm{nominal}} - \mathrm{AUC}^{(i)}_{\mathrm{shuffled}}, \tag{46}$$

where $\mathrm{AUC}_{\mathrm{nominal}}$ is the nominal AUC obtained with the original testing sample, and $\mathrm{AUC}^{(i)}_{\mathrm{shuffled}}$ is the AUC obtained after randomly permuting the values of variable $i$.

This procedure is repeated for each input variable and several random permutations are used in order to reduce the impact of statistical fluctuations. The resulting ranking allows the most discriminating observables to be identified and provides a useful diagnostic tool for the construction of the final BDT variable set.

The variable importance ranking is then used as a guide for the optimization of the input variable set and for identifying redundant or weakly discriminating observables.

Figure 14 shows the resulting importance rankings for the four analysis categories. Several observables consistently appear among the most discriminating variables. In particular, the missing mass squared $M^2_{\mathrm{miss}}$ and the missing energy variable $U$ are among the most important inputs in the leptonic and semi-leptonic channels, reflecting the presence of multiple neutrinos in the signal decay.

Angular observables such as the pseudo helicity variables and the $\tau$ polarization related quantities also provide significant discrimination power. In contrast, the hadronic channels rely more strongly on visible kinematic quantities, including the momentum of the highest momentum daughter particle, event shape observables, and missing transverse momentum.

The importance rankings confirm that the selected variable set captures complementary aspects of the signal topology and justify their inclusion in the final BDT training.

### 6.3.1 BDT Optimization

The performance of the classifier depends on several hyperparameters controlling the complexity of the boosted decision trees, including the maximum tree depth, learning rate, number of trees, subsampling fractions, and regularization parameters.

The optimization of these hyperparameters is performed using the Optuna [14] framework. Optuna employs a Bayesian optimization strategy to efficiently explore the multidimensional parameter space and identify configurations that maximize the classifier performance.

For each trial proposed by Optuna, a BDT is trained and evaluated using a three-fold cross-validation procedure. The available training sample is divided into three independent subsets. At each iteration, two subsets are used for training while the remaining subset is used for validation. The process is repeated three times such that each subset is used once as the validation sample.

The performance of a given hyperparameter configuration is quantified using the AUC. The final score assigned to a trial corresponds to the average validation AUC obtained over the three folds,

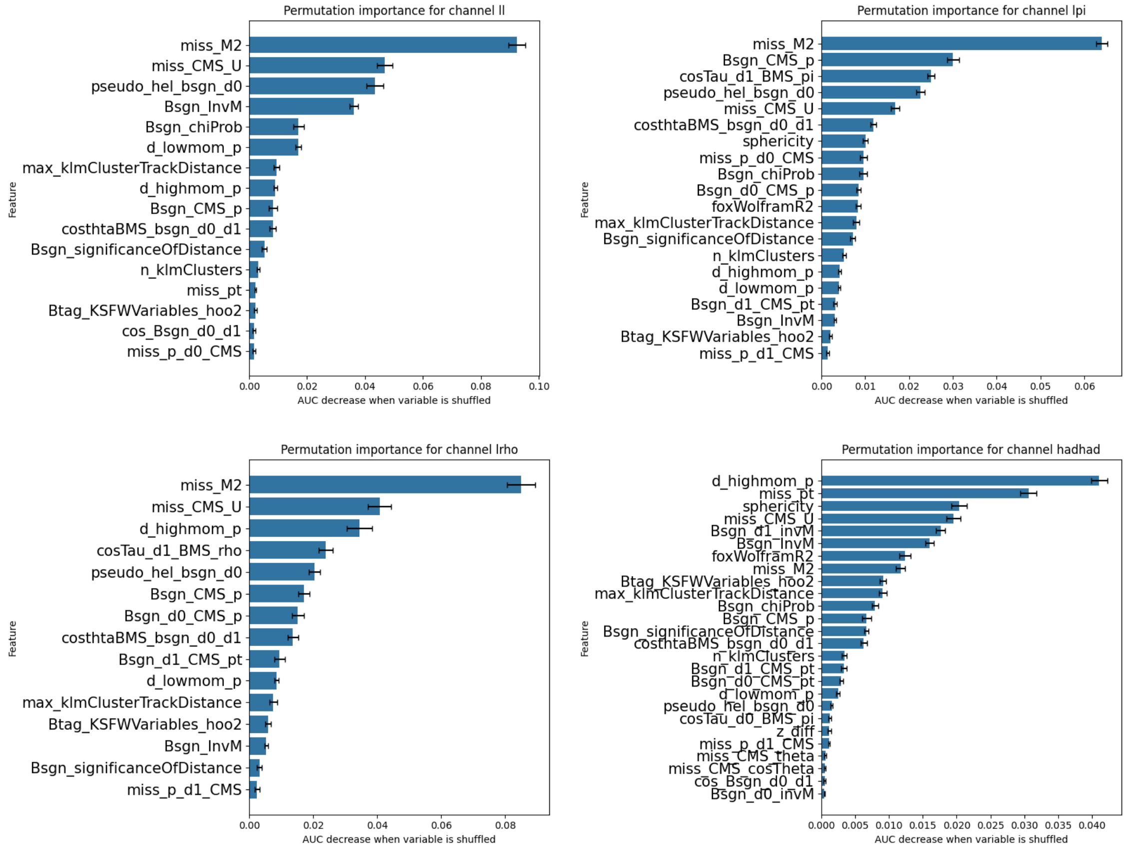



Figure 14: Permutation importance ranking obtained for the four analysis categories. The importance is defined as the decrease in the ROC AUC obtained after randomly shuffling the values of a given variable in the testing sample.

$$\mathrm{AUC}_{\mathrm{CV}} = \frac{1}{3}\sum_{i=1}^{3} \mathrm{AUC}_i. \tag{47}$$

In addition to maximizing the validation performance, a penalty is introduced to suppress configurations exhibiting signs of overtraining. This penalty is based on the difference between the AUC obtained on the training sample and the validation sample. Hyperparameter configurations producing large train–test discrepancies are therefore disfavoured, even if their validation performance is high.

The optimization score can therefore be written schematically as

$$\mathcal{S} = \mathrm{AUC}_{\mathrm{CV}} - \lambda \left|\mathrm{AUC}_{\mathrm{train}} - \mathrm{AUC}_{\mathrm{valid}}\right|, \tag{48}$$

where $\lambda$ controls the strength of the overtraining penalty.

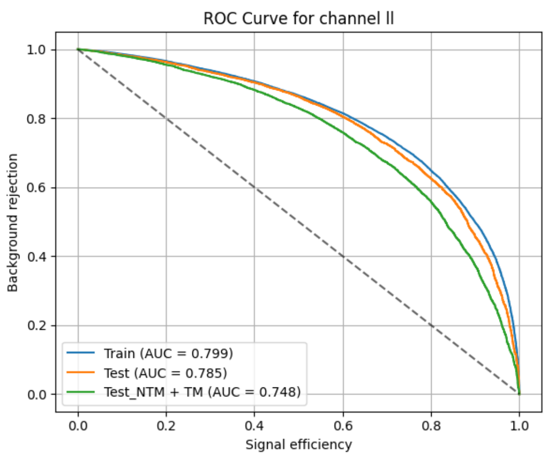
This strategy favours classifiers that simultaneously achieve strong discrimination power and good generalization performance.

The input variables used for the optimization are selected beforehand based on their physical motivation, their individual discriminating power, and their permutation-importance ranking. The hyperparameter optimization is therefore performed using a fixed set of observables for each analysis category.

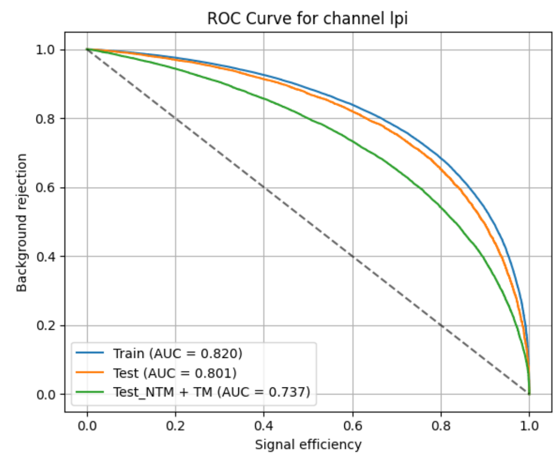
For every Optuna trial, a new BDT is trained and evaluated through the cross-validation procedure described above. The optimization algorithm progressively identifies regions of the parameter space that provide improved discrimination power while avoiding excessive overtraining. The best-performing configuration is then retained for the final classifier training.

## 6.4 BDT performance

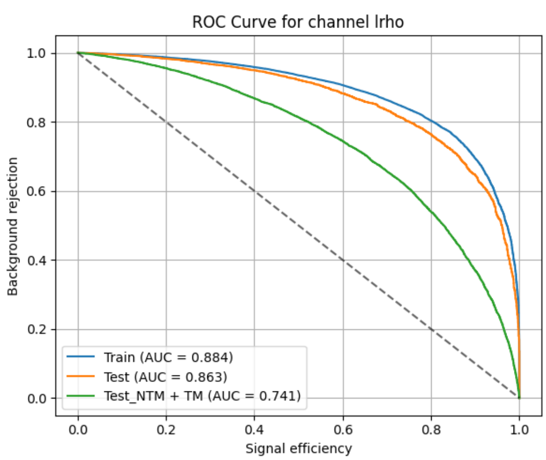
After the optimization procedure, the performance of the BDT classifier is evaluated using statistically independent testing samples. The classifier performance is assessed through the AUC, which provides a threshold-independent measure of its discriminating power. In addition, the classifier responses obtained for the training and testing samples are compared to verify the absence of significant overtraining.



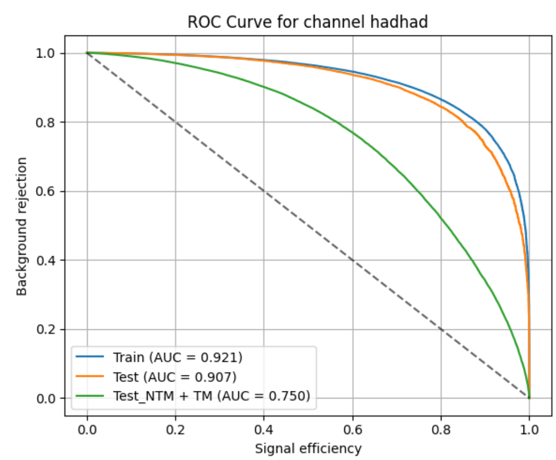

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

Figure 15: ROC curves obtained for the four analysis categories. The blue and orange curves correspond to the training and testing samples used during the BDT optimization (i.e truth matched signal candidate), while the green curve is obtained by evaluating the classifier on the full signal sample including both truth-matched (TM) and non-truth-matched (NTM) candidates. Good agreement between the training and testing curves is observed in all categories, indicating the absence of significant overtraining. The degradation of the performance when NTM candidates are included highlights the impact of reconstruction ambiguities on the classifier response. The hadronic category exhibits the highest discrimination power, reaching an AUC above 0.9 for the truth-matched sample.

Figure 15 shows the ROC curves obtained for the different signal categories when the classifier is trained using only truth matched (TM) signal candidates. In all cases, the classifier achieves a significant separation between signal and background events, demonstrating that the selected observables contain substantial discriminating information. The corresponding AUC values provide a threshold independent measure of the classifier performance and are used to compare the different category specific trainings.

A good agreement is observed between the training and testing samples, indicating that the classifiers generalize well and do not exhibit significant overtraining. The figure also compares the performance obtained when the classifiers are evaluated on the full signal sample including both TM and non truth matched (NTM) candidates. In all categories, the inclusion of NTM candidates leads to a reduction of the AUC, reflecting the impact of reconstruction ambiguities on the discriminating power of the input observables. Nevertheless, the classifiers retain a significant separation power, demonstrating that the BDT remains effective under realistic reconstruction conditions.

The best performance is achieved in the hadronic category, where the test AUC reaches 0.907, while the leptonic and semileptonic categories achieve AUC values between 0.785 and 0.863. The degradation observed when including NTM candidates is particularly pronounced in the hadronic category, highlighting the importance of reconstruction effects in channels containing hadrons.

Although training on truth matched candidates provides a clean signal sample, such information is not available in data and incorrectly reconstructed signal candidates remain present in the final event sample. To evaluate the impact of these reconstruction ambiguities, an alternative training strategy is considered in which both truth matched and non truth matched signal candidates are included in the training sample.

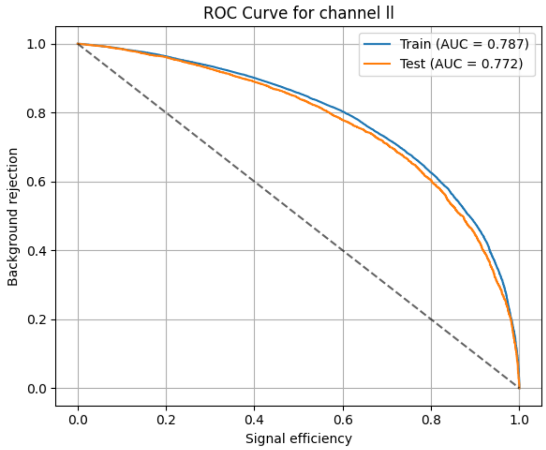
Figure 16 shows the corresponding ROC curves. Comparing the two approaches provides a quantitative assessment of the effect of reconstruction ambiguities on the classifier performance and allows to evaluate the robustness of the multivariate selection under realistic reconstruction conditions.

As expected, the inclusion of non truth-matched candidates leads to a reduction in the classifier performance in all categories. It should be noted, however, that this comparison is not strictly equivalent, since the ROC curves obtained with truth-matched training are evaluated using only correctly reconstructed signal candidates, whereas the full-sample training includes both correctly and incorrectly reconstructed signal events. The effect is particularly pronounced for channels containing hadrons, where mis-reconstructions and cross-feed between signal categories are more frequent. Nevertheless, the degradation remains moderate and the classifiers retain a significant discrimination power, demonstrating that the selected observables remain effective even when imperfectly reconstructed signal candidates are included.

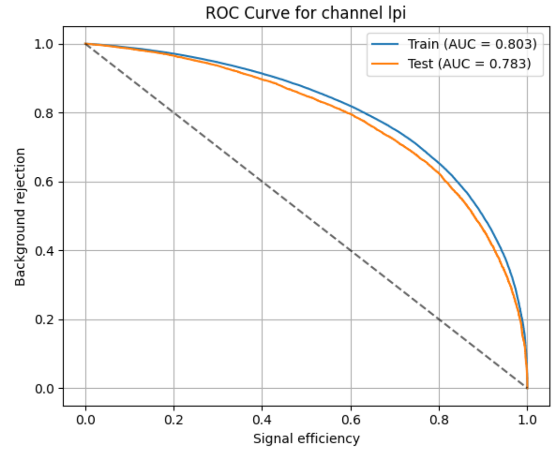
To verify that the classifier generalizes correctly and does not simply learn statistical fluctuations of the training sample, the BDT response is evaluated on the training and testing datasets.

Figure 17 shows the distribution of the BDT output score for signal and background events when the classifier is trained using only truth matched signal candidates. A clear separation between the two classes is observed in all categories, with signal events preferentially populating the high-score region and background events concentrated at low BDT values. The separation is particularly pronounced in the hadronic and lepton–rho categories, consistent with the larger AUC values obtained from the ROC curves.

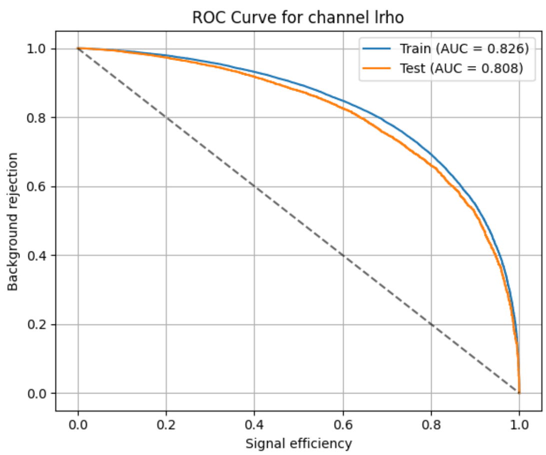
The agreement between the training and testing samples is generally good for both signal and background events. This behaviour is confirmed by the pull distributions shown in the lower panels, where the pull is defined as the difference between the testing and training histograms divided by the statistical



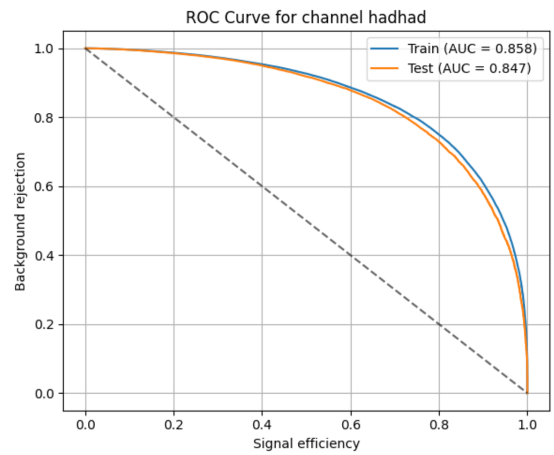

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels



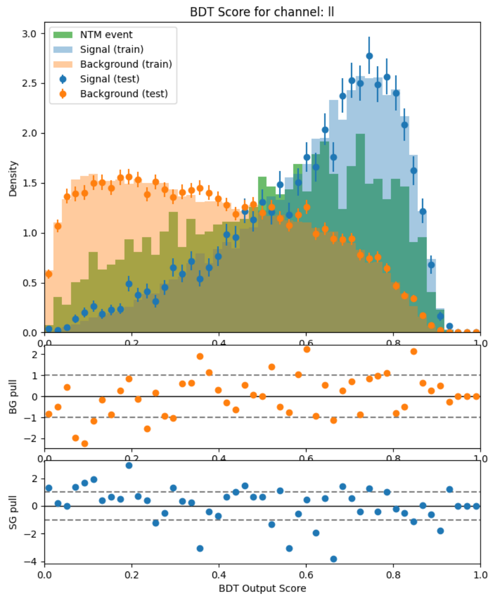

(d) Hadronic channels

Figure 16: ROC curves obtained for the four analysis categories when the BDT is trained using the full signal sample, including both truth-matched and non-truth-matched candidates. The blue and orange curves correspond to the training and testing samples, respectively. Good agreement between the two samples is observed in all categories, indicating the absence of significant overtraining. The corresponding AUC values are reported in each panel and provide a quantitative measure of the classifier discrimination power.

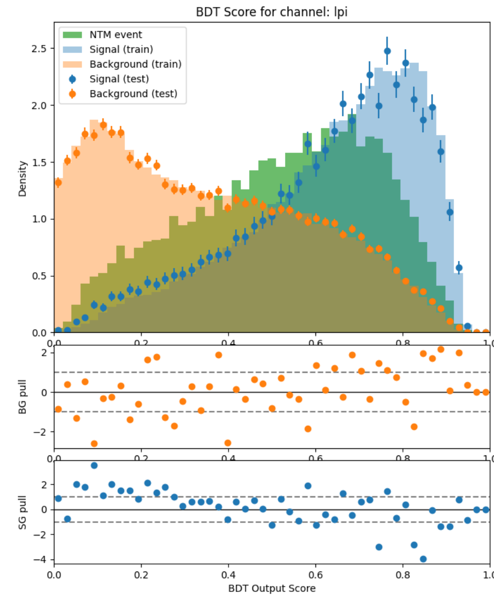
uncertainty of the testing sample. Most bins remain within the $\pm 1\sigma$ band, and no significant deviations are observed, indicating that the classifier does not suffer from substantial overtraining.

For comparison, the distributions of non truth matched signal candidates are also shown. These events populate intermediate BDT values and exhibit a reduced signal like behaviour compared to correctly reconstructed candidates. This effect suggests that reconstruction ambiguities partially dilute the discriminating power of the input observables and motivate the study of an alternative training strategy including these candidates.

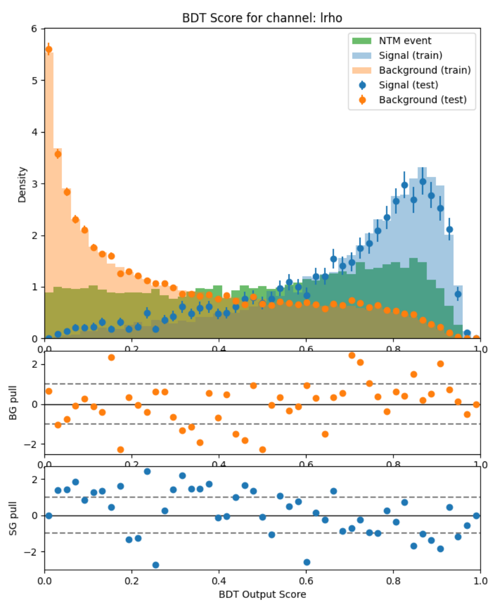
Figure 18 shows the BDT output distributions obtained when the classifier is trained using the full signal sample, including both truth matched and non truth matched candidates. Signal and background events remain well separated in all categories, demonstrating that the classifier retains a significant discrimination power under realistic reconstruction conditions.



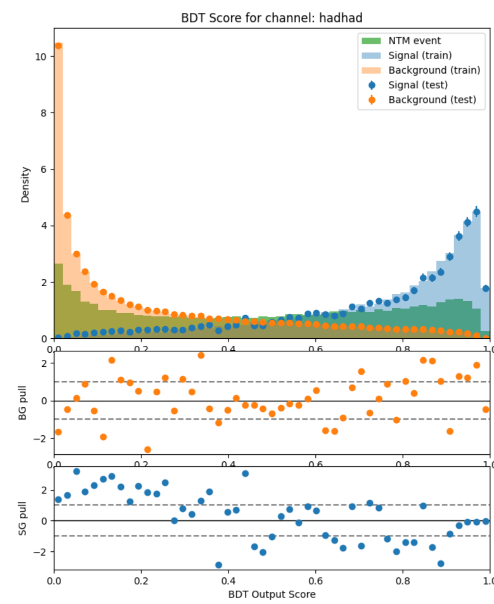

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

Figure 17: Distribution of the BDT output score for signal and background events in the four analysis categories when the classifier is trained using only truth matched signal candidates. The filled histograms correspond to the training samples, while the markers represent the statistically independent testing samples. The lower panels show the pull distributions between the training and testing samples for signal and background events. Good agreement is observed in all categories, indicating the absence of significant overtraining. Non truth matched signal candidates (green histograms) populate intermediate BDT values and exhibit a reduced signal-like behaviour compared to truth-matched candidates.

Compared to the truth-matched training, the signal distributions are broader and extend further towards intermediate BDT values. This behaviour is expected, since the training sample now includes imperfectly reconstructed signal candidates whose kinematic properties are less characteristic of the signal process. As a consequence, the separation between signal and background is reduced, in agreement with the lower AUC values observed in the ROC curves.

The agreement between the training and testing samples remains good across the full range of the classifier output. The pull distributions do not reveal any significant systematic discrepancies, confirming that the inclusion of non-truth-matched candidates does not introduce noticeable overtraining and that the classifier generalizes well to independent data samples.

The BDT output provides a powerful single discriminating variable, combining the information from all input observables into a single score. Signal events are preferentially located at large BDT values, while background events accumulate at lower scores. This separation forms the basis of the final event selection described in the next section.

## 6.5 Signal Selection Optimization

The output of the optimized BDT classifier is used to define the final event selection. Since the expected branching fraction of the decay $B^0 \to \tau^+\tau^-$ is extremely small, the selection must provide a strong background rejection while preserving as much signal efficiency as possible.

The optimal requirement on the BDT output is determined using the Punzi figure of merit (FOM) [17], which is widely used in searches for rare processes where the expected signal yield is small. The Punzi FOM is defined as

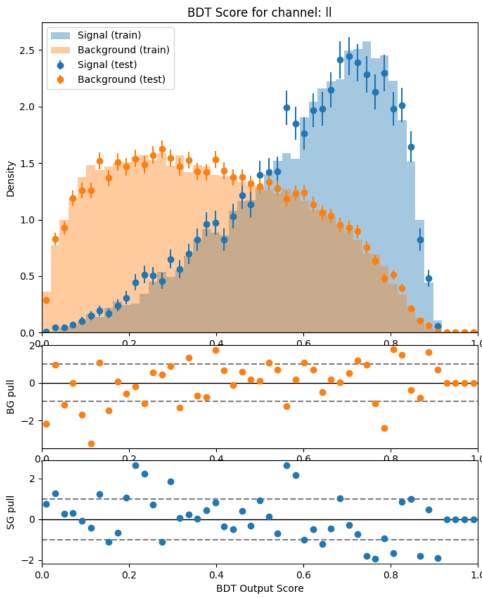
$$P = \frac{\varepsilon_{\text{sig}}}{a/2 + \sqrt{N_{\text{bg}}}}, \tag{49}$$

where $\varepsilon_{\text{sig}}$ is the signal efficiency after the selection, $N_{\text{bg}}$ is the expected number of background events, and $a$ corresponds to the target significance level. In this analysis, a value of $a = 3$ is used.

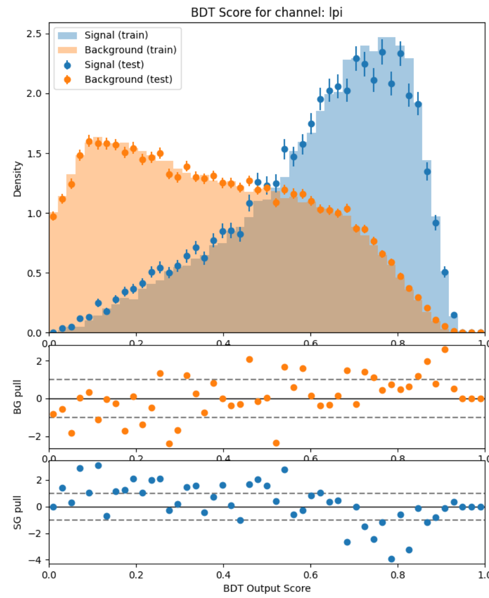
The signal efficiency is always evaluated using the full signal Monte Carlo sample, including both truth-matched and non-truth-matched candidates. While some BDT trainings are performed using only truth-matched signal events in order to provide a cleaner training sample, the final efficiency and sensitivity estimates are computed using all reconstructed signal candidates. This choice ensures that the performance evaluation remains representative of the conditions expected in real data, where truth-matching information is not available.

The BDT threshold is scanned over the full range of classifier outputs and the corresponding Punzi figure of merit is evaluated for each value. The threshold maximizing the figure of merit is selected as the optimal operating point for the analysis.

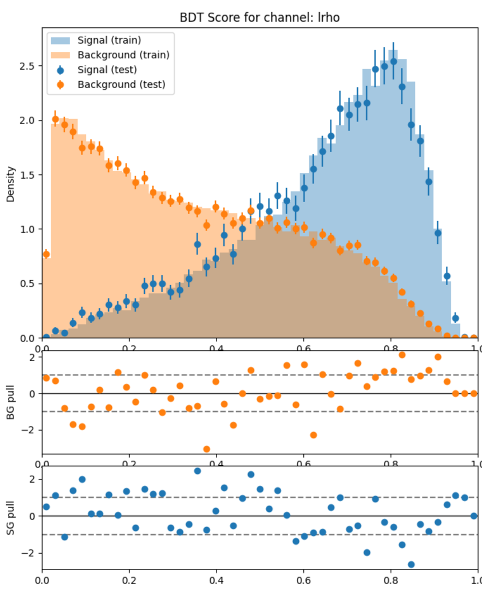
Figure 19 shows the evolution of the Punzi figure of merit as a function of the BDT threshold for the four analysis categories. As expected, the figure of merit initially increases as the background is progressively rejected while maintaining a large signal efficiency. Beyond the optimal working point, the signal efficiency decreases more rapidly than the background contribution, leading to a reduction of the expected sensitivity.
Figure 20 shows the evolution of the Punzi figure of merit as a function of the BDT threshold for the classifiers trained on the full signal sample. As in the truth matched training scenario, the figure of merit increases as the BDT requirement suppresses an increasing fraction of the background while retaining a



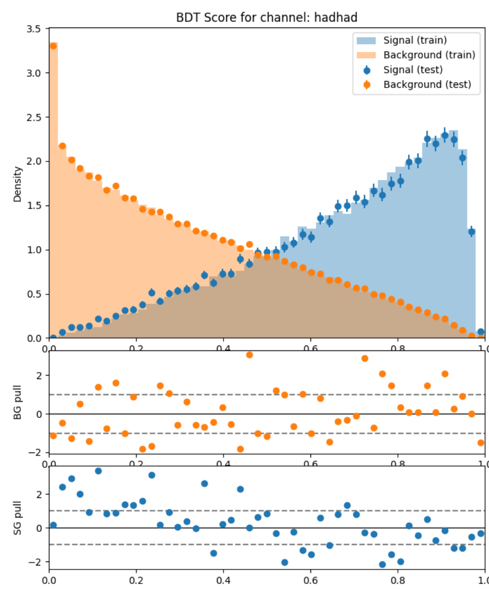

(a) Leptonic channels




(b) Lepton–pion channels

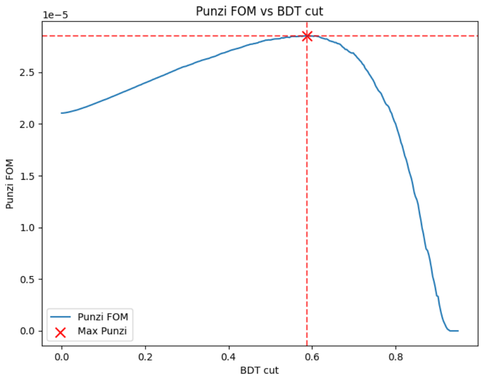



(c) Lepton–rho channels

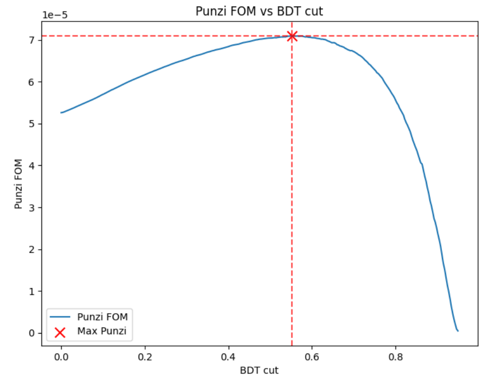



(d) Hadronic channels

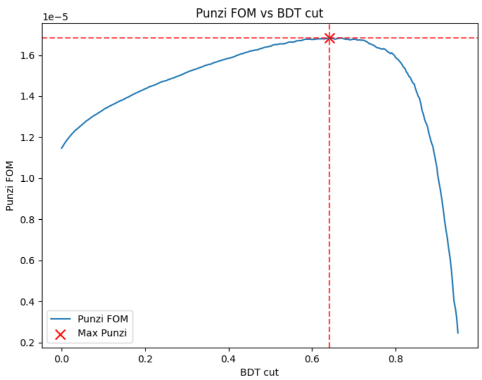
Figure 18: Distribution of the BDT output score for signal and background events in the four analysis categories when the classifier is trained using the full signal sample, including both truth-matched and non truth matched candidates. The filled histograms correspond to the training samples, while the markers represent the statistically independent testing samples. The lower panels show the pull distributions between the training and testing samples for signal and background events. Good agreement is observed in all categories, indicating the absence of significant overtraining. Signal events preferentially populate the high-score region, while background events accumulate at low BDT values, demonstrating the discriminating power of the classifier under realistic reconstruction conditions.



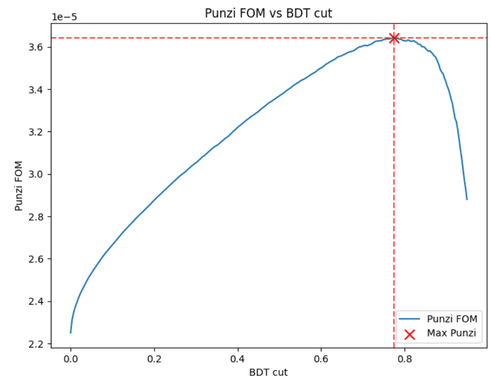

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

Figure 19: Punzi figure of merit as a function of the BDT output requirement for the four analysis categories using the classifier trained on truth-matched signal candidates. The red marker indicates the optimal operating point corresponding to the maximum value of the figure of merit, while the dashed lines highlight the associated BDT threshold and Punzi value. The optimal cut is determined by maximizing the expected sensitivity of the analysis, balancing signal efficiency against residual background contamination.

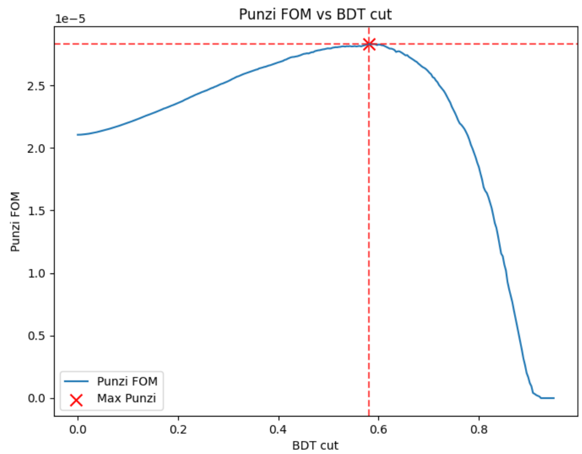
large fraction of signal events. Beyond the optimal working point, the loss in signal efficiency becomes dominant and the expected sensitivity decreases.

Compared with the truth-matched training, the positions of the maxima remain broadly similar, indicating that the optimal operating point is relatively insensitive to whether truth-matching requirements are applied during training. These optimized thresholds define the final event selections used in the subsequent stages of the analysis.

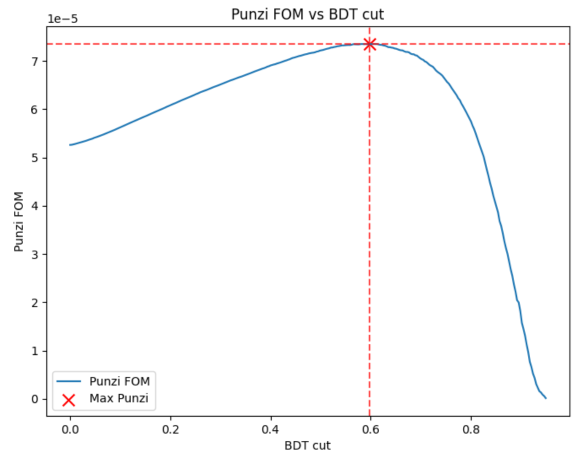
The optimal BDT thresholds differ among the analysis categories and reflect the different levels of signal-to-background separation achieved by the corresponding classifiers. In particular, the hadronic category reaches its maximum sensitivity for a relatively tight BDT requirement, while the leptonic and semileptonic categories favour slightly looser selections.

Tables 3 and 4 summarize the final selection performance obtained after the optimized BDT requirement and the random candidate selection for the two training strategies considered in this study.

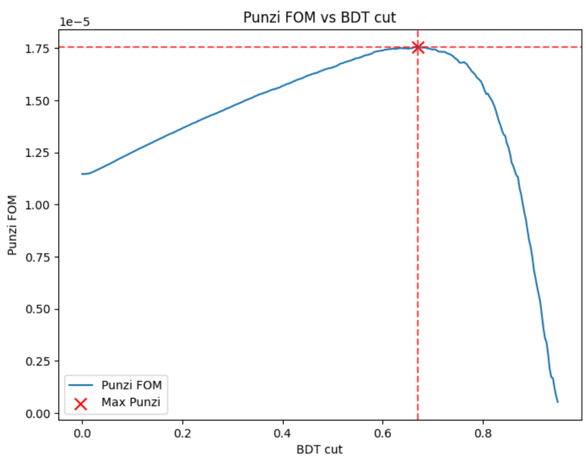
Multiple candidates can be reconstructed in a single event as different combinations of tracks and neutral clusters may satisfy the signal reconstruction requirements. To prevent double counting and maintain



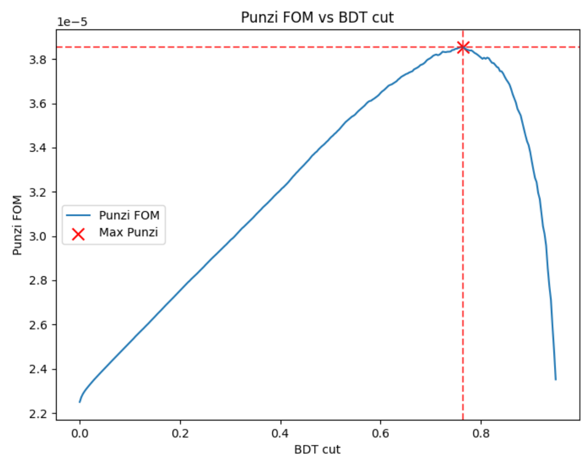

(a) Leptonic channels




(b) Lepton–pion channels




(c) Lepton–rho channels




(d) Hadronic channels

Figure 20: Punzi figure of merit as a function of the BDT output requirement for the four analysis categories using the classifier trained on the full signal sample, including both truth matched and non-truth-matched candidates. The red marker indicates the optimal operating point corresponding to the maximum value of the figure of merit, while the dashed lines highlight the associated BDT threshold and Punzi value. The selected thresholds maximize the expected sensitivity of the analysis by balancing signal efficiency against the residual background contribution.

a statistically well-defined event sample, a best-candidate selection is applied. In this analysis, one candidate is chosen randomly among the candidates reconstructed in the event.

As expected from the ROC curves presented in the previous section, the classifier trained using only truth matched signal candidates generally achieves a better separation between signal and background. However, after optimizing the BDT threshold using the Punzi figure of merit, the final performances obtained with the two approaches are found to be remarkably similar.

For the leptonic and lepton–rho categories, the signal efficiencies differ by only a few percent between the two training strategies, while the expected background yields remain of the same order. In the hadronic category, the training performed on the full signal sample even results in a slightly higher signal efficiency together with a reduced background yield. A similar behaviour is observed in the lepton–pion category, where the decrease in signal efficiency is compensated by a substantial reduction of the expected background contribution.

Although the classifier trained on truth matched candidates achieves a better intrinsic discrimination power, this improvement translates only into a modest gain in the final analysis performance. After optimization of the BDT selection, both approaches yield similar signal efficiencies, background rejections, and expected sensitivities. This result indicates that the multivariate selection remains robust against signal mis-reconstructions and category cross-feed, and that a training based on the full signal sample provides a realistic and competitive alternative for the final analysis.

Table 3: Final number of signal events, signal efficiency, and expected background yield after the optimized BDT selection and random candidate selection (BCS), for the BDT trained using truth matched signal candidates. The quoted signal yields are given for the available simulated signal sample and are not scaled to the expected signal yield in data.

| Category | $N_{\text{sig}}^{\text{final}}$ | $\varepsilon_{\text{sig}}$ [%] | $N_{\text{bg}}^{\text{final}}$ | Best Punzi |
|---|---|---|---|---|
| Leptonic | 13284 | 0.179 | 13611 | $2.9e^{-5}$ |
| Lepton–pion | 34425 | 0.755 | 38263 | $7.1e^{-5}$ |
| Lepton–rho | 11001 | 0.102 | 13176 | $1.7e^{-5}$ |
| Hadronic | 27025 | 0.343 | 30976 | $3.6e^{-5}$ |

Table 4: Final number of signal events, signal efficiency, and expected background yield after the optimized BDT selection and random candidate selection (BCS), for the BDT trained using the full signal sample, including both truth-matched and non truth matched candidates. The quoted signal yields are given for the available simulated signal sample and are not scaled to the expected signal yield in data.

| Category | $N_{\text{sig}}^{\text{final}}$ | $\varepsilon_{\text{sig}}$ [%] | $N_{\text{bg}}^{\text{final}}$ | Best Punzi |
|---|---|---|---|---|
| Leptonic | 13421 | 0.181 | 14079 | $2.8e^{-5}$ |
| Lepton–pion | 32822 | 0.720 | 32414 | $7.3e^{-5}$ |
| Lepton–rho | 10624 | 0.099 | 11359 | $1.8e^{-5}$ |
| Hadronic | 27928 | 0.354 | 29482 | $3.9e^{-5}$ |

Because of the limited size of the available signal sample, the final BDT selection is applied to the full dataset, including both the training and testing subsamples. Strictly speaking, this procedure introduces a bias, since events used during training are also used in the efficiency evaluation. As a consequence, the quoted signal efficiencies and expected sensitivities are likely to be overestimated.

Nevertheless, all training strategies considered in this work are evaluated using exactly the same procedure. The relative comparisons between different variable sets and training configurations therefore remain meaningful and provide a consistent measure of their respective performances. A more rigorous evaluation would require the use of a $k$-folding strategy, in which independent classifiers are applied to events that were not used during their training. Such an approach would remove the residual training bias while preserving the full statistical power of the available samples.

The events passing the optimized BDT requirement define the final analysis sample used for the signal extraction procedure described in the next section.

# 7 Results and comparison with previous work

This section is divided into two parts. First, the performance of the baseline analysis, developed prior to this internship, is presented and used as a reference. This initial analysis relies on the original set of discriminating variables and provides the benchmark sensitivity for the search.

The second part presents the results obtained using the analysis strategy developed during this internship, as described in Section 6. These improvements include the optimization of the BDT training procedure, the introduction of additional discriminating observables such as the KLM variables, pseudo helicity angles, and the higher and lower momentum particle variables, as well as an optimized event selection. Finally, the performance of classifiers trained using only truth matched signal candidates is compared with that of classifiers trained on the full signal sample, allowing the impact of the training strategy to be evaluated.

## 7.1 Baseline Analysis

The first stage of the analysis was based on an initial set of discriminating variables developed during the early studies of the $B^0 \to \tau^+\tau^-$ search prior to this internship. The objective of this baseline configuration was to evaluate the achievable sensitivity using standard kinematic and event-topology observables before investigating more dedicated variables related to the $\tau$ decay structure.

The classifiers were trained using simulated signal and background samples without applying any truth-matching requirement on the signal candidates. Consequently, both correctly reconstructed and misreconstructed signal candidates were included in the training sample. This approach provided a realistic first estimate of the performance that could be achieved under the nominal reconstruction conditions.

The baseline variable set was primarily composed of missing-energy observables, event-shape quantities, and basic kinematic properties of the reconstructed final-state particles. These variables were chosen for their expected sensitivity to the presence of multiple neutrinos in the signal decay and to the topological differences between signal and background events.

The performance of this initial configuration was evaluated through the BDT output distributions, the optimization of the Punzi figure of merit, and variable-importance studies. The results obtained with this baseline setup established the reference performance against which all subsequent improvements were compared.

## 7.2 Baseline Variable Set

The baseline set of input variables used for the BDT training is shown Fig. 21. These observables were selected based on previous studies and physics intuition, without performing a dedicated variable optimization procedure.

The baseline feature set included observables related to the missing momentum of the event, event-shape quantities, kinematic properties of the visible $\tau$ decay products, and variables describing the reconstructed $B_{\text{sig}}$ candidate.

A common feature observed across all categories is the dominant role played by missing-energy observables. In particular, the missing-mass squared $M^2_{\text{miss}}$ appears among the most important variables in all channels and constitutes the leading source of discrimination in the leptonic, lepton–pion, and lepton–rho categories. This behaviour is expected since signal events contain between two and four neutrinos in the final state, resulting in significantly different missing-energy distributions compared to most background processes.

Using this baseline configuration, dedicated BDT classifiers were trained independently for each signal category. The resulting classifier performances constitute the reference against which the optimized strategy developed in this work will be compared.

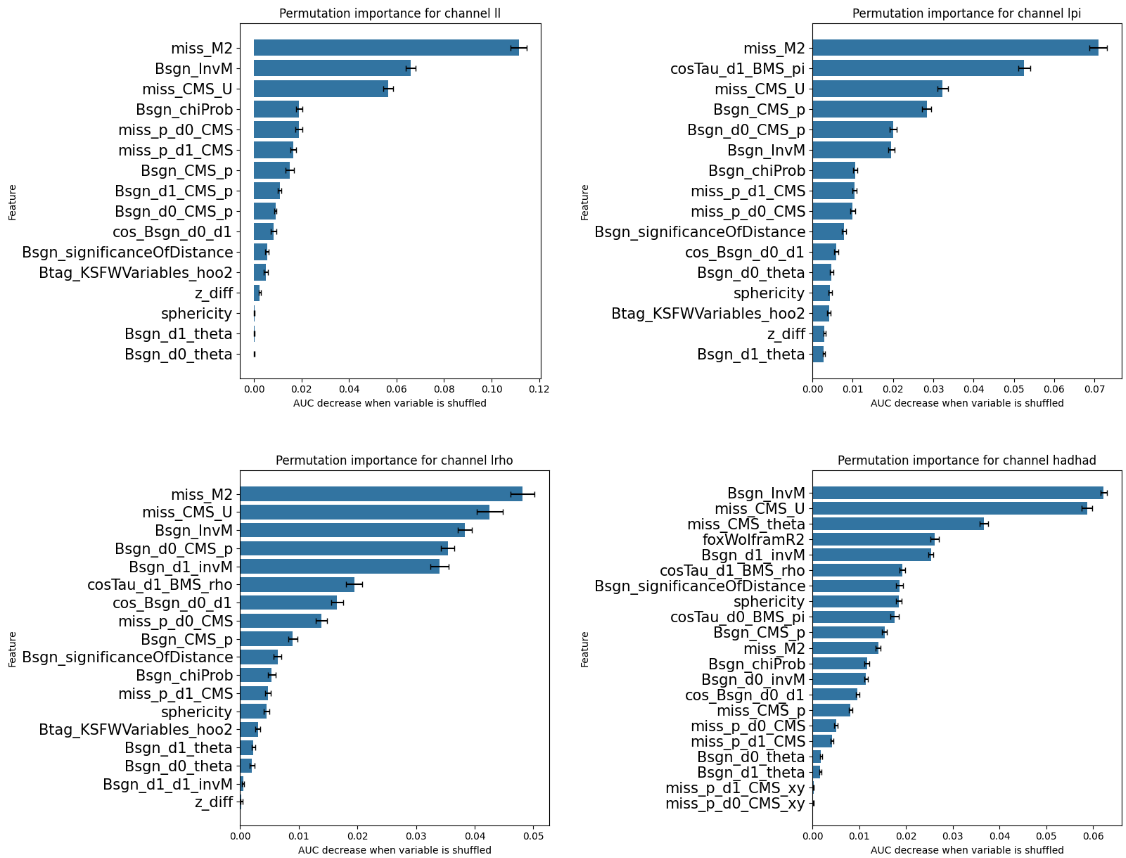



Figure 21: Permutation importance obtained for the baseline BDT configuration in the four analysis categories. The importance is quantified by the decrease in the AUC when a given variable is randomly shuffled. Larger values indicate a stronger contribution to the classifier performance. The distributions reveal that the discrimination is largely driven by missing-energy observables and a limited number of kinematic quantities, while several variables contribute only marginally to the final classifier response.

## 7.3 Baseline BDT Output Distribution

The resulting BDT output distributions, using the baseline variables set, for signal and background events are shown in Fig. 22.

At this stage, no truth-matching requirement is applied during the training procedure. Both correctly and incorrectly reconstructed signal candidates are therefore included in the signal sample used for the BDT training.

Despite the visible overlap between the two distributions, the classifier succeeds in enhancing the signal purity in the high-score region. This behavior demonstrates that the baseline variable set already contains discriminating information that can be exploited by the multivariate algorithm. The optimal BDT selection threshold is determined in the next section using the Punzi figure of merit.

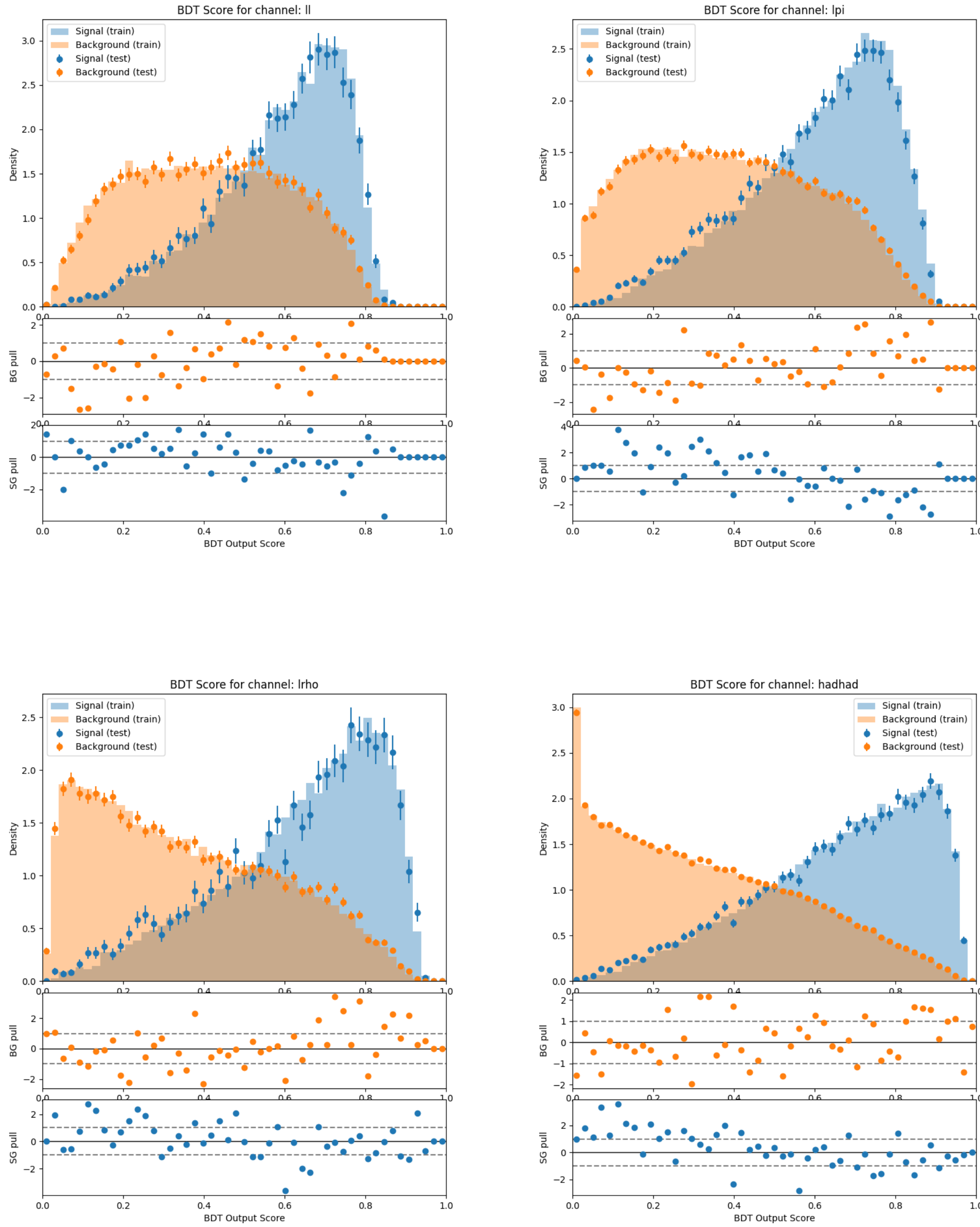



Figure 22: BDT output distributions obtained with the baseline variable set for the four analysis categories. The histograms correspond to the training samples, while the markers represent the independent test samples. The lower panels show the pull distributions between training and testing samples for signal and background events. A clear separation between signal and background is observed in all categories, with no significant indication of overtraining. The hadronic category exhibits the strongest discrimination power, while the leptonic channels remain the most challenging due to the larger overlap between signal and background distributions.

## 7.4 Baseline Punzi Optimization

The optimal BDT selection threshold is determined by maximizing the Punzi figure of merit introduced in Section 6.5. The corresponding distributions are shown in Fig. 23 for the four analysis categories.

For each channel, the Punzi figure of merit exhibits a clear maximum corresponding to the best compromise between signal efficiency and background rejection. At low BDT thresholds, the large background contamination reduces the sensitivity, while at very high thresholds the signal efficiency decreases significantly. The optimal working point is therefore defined by the maximum of the Punzi curve and is subsequently used to evaluate the expected sensitivity of the baseline analysis. The BDT thresholds

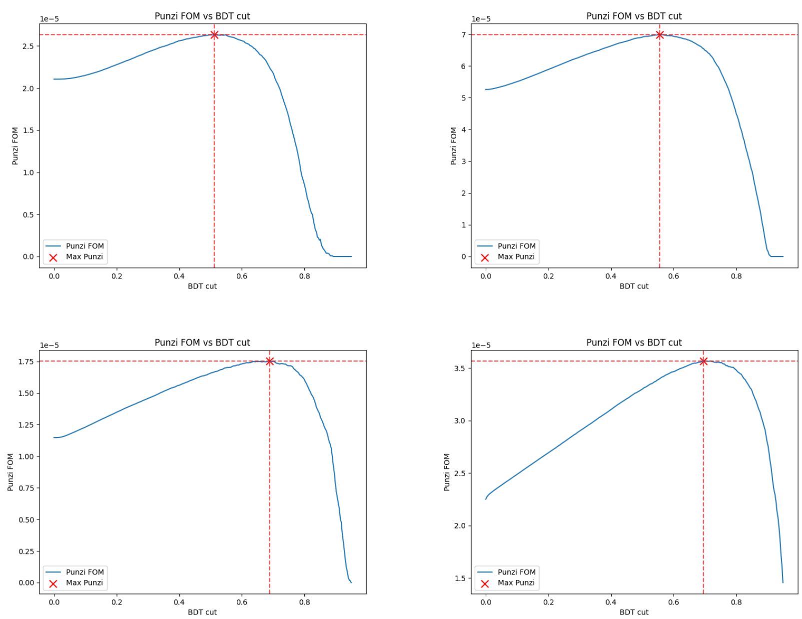



Figure 23: Evolution of the Punzi figure of merit as a function of the BDT score requirement for the baseline classifier configuration. The red marker indicates the optimal working point corresponding to the maximum Punzi value. The optimization balances signal efficiency against the expected background yield and is used to determine the final BDT selection threshold for each analysis category.

obtained from the Punzi optimization define the final baseline selection. The corresponding signal efficiencies and expected sensitivities, shown Table 5, provide the reference performance against which the optimized analysis developed in this work will be compared.

Table 5 summarize the performance of the baseline analysis. These results serve as a reference and are directly compared with those obtained using the optimized BDT trained on the full signal sample (without truth matching), presented in Section 6.5 and summarized in Table 4. Overall, the optimization procedure leads to a significant reduction of the expected background yield while preserving a comparable signal efficiency in most categories. The improvement is particularly visible in the lepton–pion and hadronic channels, where the lower background level results in a higher Punzi figure of merit. The lepton–rho category remains largely unchanged, indicating that the baseline variable set already captured most of the available discriminating information. In contrast, the purely leptonic category does not exhibit a noticeable gain in sensitivity. These results demonstrate that the optimized feature selection improves the signal-to-background separation and provides a modest but measurable increase in the overall expected sensitivity of the analysis.

Table 5: Final number of signal events, signal efficiency, and expected background yield after the optimized BDT selection and random candidate selection (RCS), for the BDT trained using the baseline set of variables 21.

| Category | $N_{\text{sig}}^{\text{final}}$ | $\varepsilon_{\text{sig}}$ [%] | $N_{\text{bg}}^{\text{final}}$ | Best Punzi |
|---|---|---|---|---|
| Leptonic | 15713 | 0.211 | 22302 | $2.6e^{-5}$ |
| Lepton–pion | 34712 | 0.761 | 41715 | $6.9e^{-5}$ |
| Lepton–rho | 10049 | 0.093 | 10057 | $1.8e^{-5}$ |
| Hadronic | 31142 | 0.394 | 42278 | $3.6e^{-5}$ |

### 7.5 Baseline Limit

The baseline variable set described in the previous section was used to train dedicated BDT classifiers for each signal category. The resulting selected events were then used into a simultaneous fit of the $E_{\text{extra}}^{\text{ECL}}$ distributions in order to obtain the signal yield. Figure 24 shows the corresponding pre-fit distributions for the four analysis categories.

The variable $E_{\text{extra}}^{\text{ECL}}$ is defined as the sum of the energy deposited in electromagnetic calorimeter (ECL) clusters that are not associated with any reconstructed particle belonging to either the signal side or tag side $B$ meson candidate. Before computing this quantity, a dedicated cluster mask is applied to reject clusters likely originating from beam induced background or fake photons. This selection typically relies on requirements on the cluster energy and timing with respect to the event time $T_0$.

Signal events are expected to populate the low-$E_{\text{extra}}^{\text{ECL}}$ region, since all visible particles are reconstructed and only neutrinos escape detection. In contrast, background events often contain additional neutral particles or residual calorimeter activity, leading to larger values of $E_{\text{extra}}^{\text{ECL}}$. To suppress poorly reconstructed events and focus on the region with the highest signal sensitivity, only events satisfying $E_{\text{extra}}^{\text{ECL}} < 1$ GeV are retained for the fit. In the statistical interpretation, the signal contribution is generated assuming an expected branching fraction of $10^{-3}$.

The points shown in Fig. 24 correspond to Asimov pseudodata [18]. The Asimov dataset is constructed from the expected background contribution, while the signal component corresponding to the assumed branching fraction hypothesis is injected into the expected model. By construction, Asimov pseudodata represent the expected event yields in the absence of statistical fluctuations and are commonly used to evaluate the expected sensitivity of an analysis.

These pre-fit distributions illustrate the expected composition of signal and background events before the statistical fit.

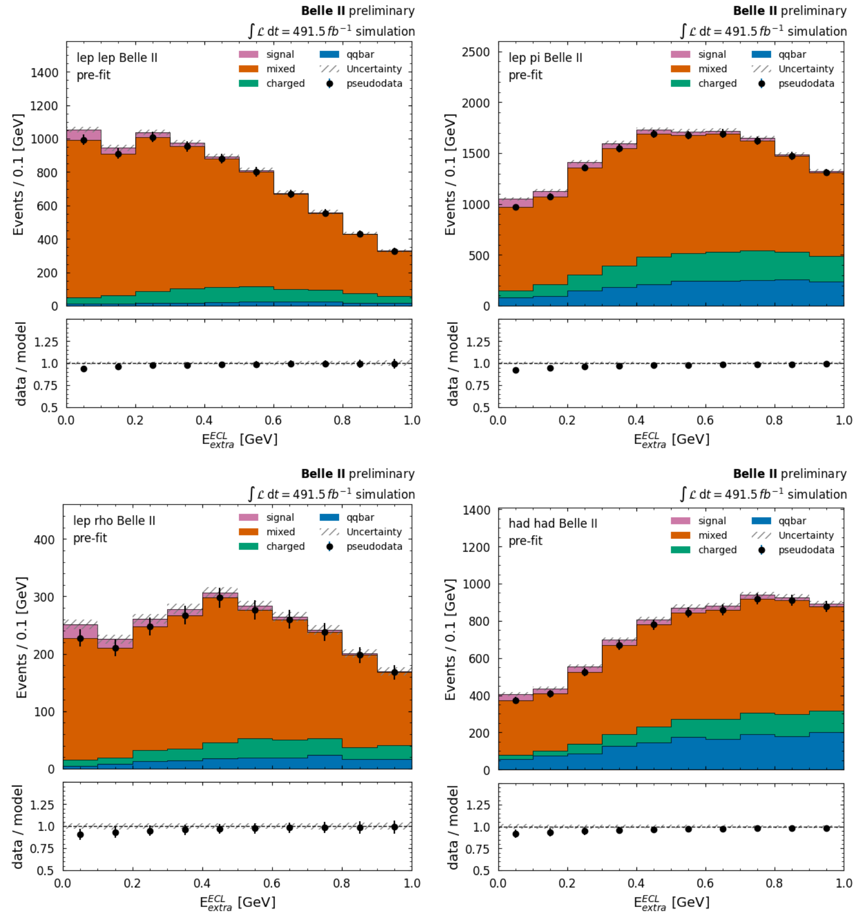



Figure 24: Pre-fit distributions of $E_{\mathrm{extra}}^{\mathrm{ECL}}$ obtained in the baseline analysis for the four signal categories. The stacked histograms show the expected signal and background contributions, where the signal corresponds to an assumed branching fraction of $10^{-3}$. The black markers represent the Asimov pseudodata constructed from the expected model. The lower panels show the ratio between the pseudodata and the model prediction.

The dominant background originates from generic $B\bar{B}$ decays, while the signal represents only a small fraction of the selected events. The relative contributions of the different background sources vary across the four signal categories, making category-specific classifiers and selection criteria necessary.

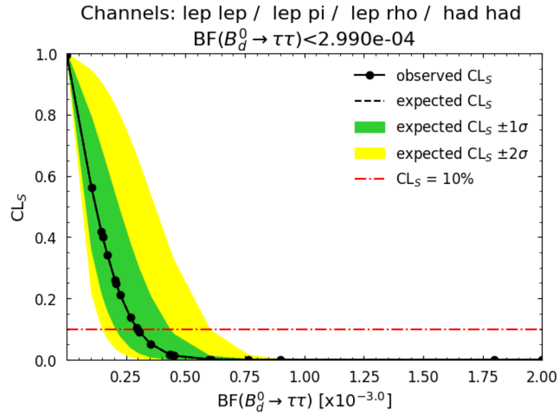



Figure 25: Expected and observed CL$_s$ curves obtained with the baseline analysis. Since the statistical interpretation is performed using an Asimov dataset, the observed and expected CL$_s$ curves coincide by construction. The dashed horizontal line indicates the 10% confidence level used to determine the upper limit on the branching fraction. The resulting expected sensitivity corresponds to an upper limit of $\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.99 \times 10^{-4}$ at the 90% confidence level.

The selected events are subsequently used in a statistical interpretation based on the CL$_s$ method [18]. The CL$_s$ procedure is used to determine the largest branching fraction value that remains compatible with the background-only hypothesis at a given confidence level, thereby providing an upper limit on the signal strength in the absence of a significant excess.

The resulting expected upper limit obtained with the baseline analysis is shown in Fig. 25. Since the statistical interpretation is based on an Asimov dataset, the observed and expected CL$_s$ curves overlap by construction. An expected upper limit of $\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.99 \times 10^{-4}$ is obtained at the 90% confidence level for an integrated luminosity of 491.5 fb$^{-1}$. This result is obtained from a statistical-only fit, with no systematic uncertainties included, and defines the reference sensitivity of the baseline analysis against which the optimized strategies developed in this work will be compared.

## 7.6 Final Sensitivity without Truth Matching

The optimization procedure and the variable sets shown in Chapter 6 are used to train dedicated BDT classifiers for each signal category. The training is performed using the full reconstructed signal sample without applying any truth-matching requirement. The selected events are subsequently incorporated into a simultaneous fit of the $E_{\text{extra}}^{\text{ECL}}$ distributions. Figure 26 shows the resulting pre-fit distributions.

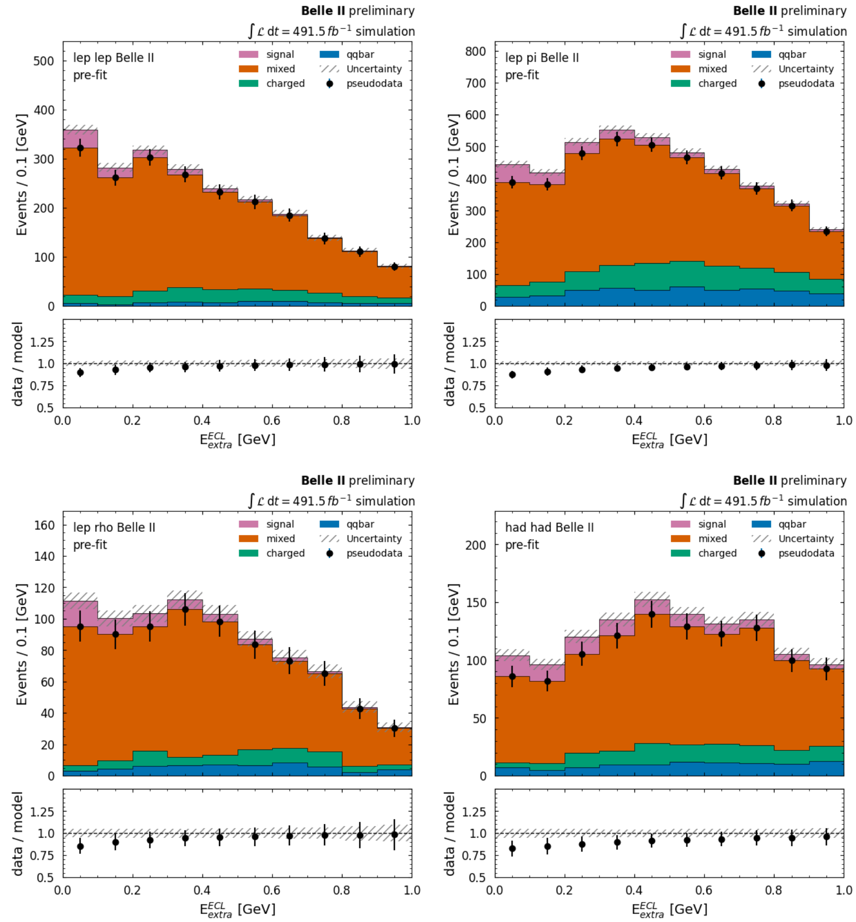



Figure 26: Pre-fit distributions of $E_{\mathrm{extra}}^{\mathrm{ECL}}$ obtained using the optimized BDT trained without truthmatching requirements. The stacked histograms show the expected signal and background contributions, where the signal corresponds to an assumed branching fraction of $10^{-3}$. The black markers represent the Asimov pseudodata constructed from the expected model. The lower panels show the ratio between the pseudodata and the model prediction.

The statistical interpretation is performed using the $\mathrm{CL}_s$ method. Figure 27 shows the resulting confidence level as a function of the assumed branching fraction. An expected upper limit of $\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.76 \times 10^{-4}$ is obtained at the 90% confidence level for an integrated luminosity of 491.5 $\mathrm{fb}^{-1}$. Compared to the baseline sensitivity of $2.99 \times 10^{-4}$, this corresponds to an improvement of approximately 7.7%, demonstrating the benefit of the optimization procedure, the variable selection and the multivariate training procedure.

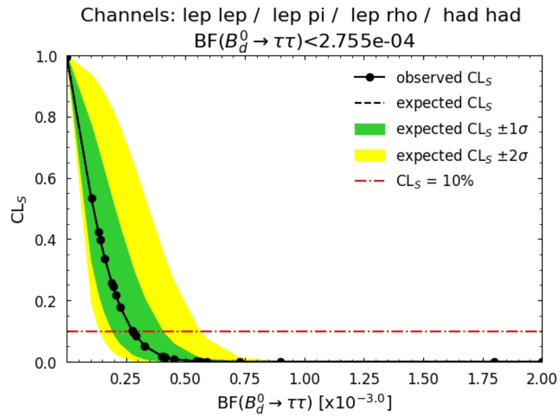



Figure 27: Expected and observed $\mathrm{CL}_s$ curves obtained using the optimized BDT trained without truth-matching requirements. Since the statistical interpretation is performed using an Asimov dataset, the observed and expected $\mathrm{CL}_s$ curves coincide by construction. The dashed horizontal line indicates the 10% confidence level used to determine the upper limit on the branching fraction. The resulting expected upper limit is $\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.76 \times 10^{-4}$ at the 90% confidence level.

## 7.7 Final Sensitivity with Truth Matching

To finalize this study, the expected sensitivity to the $B^0 \to \tau^+\tau^-$ decay is evaluated using the events passing the optimized BDT selection, where the selection threshold is optimized by maximizing the Punzi figure of merit and the classifier is trained on truth-matched signal candidates.

Figure 28 shows the pre-fit $E_{\mathrm{extra}}^{\mathrm{ECL}}$ distributions obtained from the pseudodata sample after the optimized event selections.

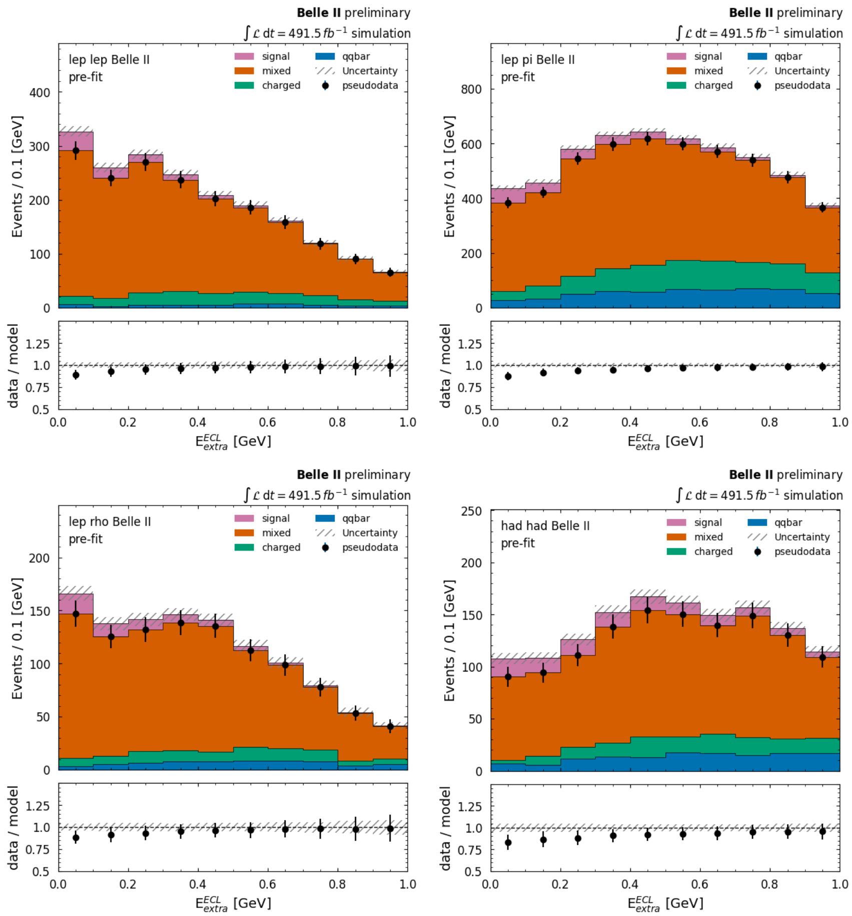



Figure 28: Pre-fit distributions of $E_{\text{extra}}^{\text{ECL}}$ obtained using the optimized BDT trained with truthmatching requirements. The stacked histograms show the expected signal and background contributions, where the signal corresponds to an assumed branching fraction of $10^{-3}$. The black markers represent the Asimov pseudodata constructed from the expected model. The lower panels show the ratio between the pseudodata and the model prediction.

The statistical interpretation is performed using the $\text{CL}_s$ method. The resulting expected and observed $\text{CL}_s$ curves are shown in Figure 29. The upper limit is determined from the branching fraction value corresponding to $\text{CL}_s = 0.1$, indicated by the horizontal red line. An expected upper limit of

$$\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.88 \times 10^{-4} \tag{50}$$

is obtained at the 90% confidence level for an integrated luminosity of 491.5 $\text{fb}^{-1}$. This expected sensitivity is slightly weaker than that obtained with the classifier trained on the full signal sample, indicating that the truth-matched training strategy does not provide the best performance.

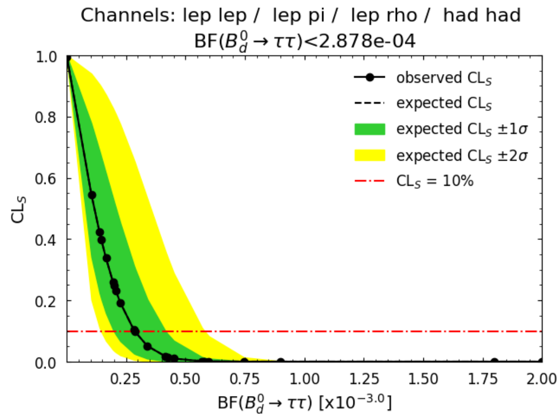



Figure 29: Expected CL$_s$ scan as a function of the branching fraction $\mathcal{B}(B^0 \to \tau^+\tau^-)$. The green and yellow bands correspond to the $\pm1\sigma$ and $\pm2\sigma$ expected intervals, respectively. The 90% confidence level upper limit is extracted from the intersection with the CL$_s = 0.1$ threshold.

Overall, the comparison of the two training strategies shows that including non-truth-matched signal candidates during training leads to a modest but measurable improvement in the expected sensitivity. This result suggests that a classifier trained on a more realistic representation of the reconstructed signal sample is better suited for the final statistical interpretation than one optimized exclusively on truth-matched candidates.

# 8 Discussion and Conclusions

In this work, a search strategy for the rare decay $B^0 \to \tau^+\tau^-$ has been developed using simulated Belle II samples. The analysis relies on the hadronic Full Event Interpretation algorithm to reconstruct one of the two $B$ mesons produced in the $\Upsilon(4S)$ decay and on a multivariate classification based on Boosted Decision Trees to suppress the dominant background contributions.

Starting from the baseline analysis, the input variables were re-optimized independently for each signal category. The resulting classifiers were evaluated using feature importance studies, BDT output distributions, and the Punzi figure of merit. Compared with the baseline configuration, the optimized selections achieve a better balance between signal efficiency and background rejection, resulting in an improved expected sensitivity.

When applied to the full reconstructed signal sample, the optimized analysis yields an expected upper limit of

$$\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.76 \times 10^{-4} \tag{51}$$

at the 90% confidence level, compared to

$$\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.99 \times 10^{-4} \tag{52}$$

for the baseline analysis. This corresponds to an improvement of approximately 7.7% demonstrating the benefit of the optimization strategy developed in this work.

A second aspect of this study concerns the use of truth-matched signal candidates during the BDT training. Since truth matching is available only in simulated samples and cannot be applied to real data, it is important to assess whether restricting the training sample to correctly reconstructed candidates can improve the final analysis sensitivity.

The comparison between the two training strategies shows that the classifier trained on the full signal sample achieves the best overall sensitivity. While the truth matched training yields an expected upper limit of

$$\mathcal{B}(B^0 \to \tau^+\tau^-) < 2.88 \times 10^{-4}, \tag{53}$$

the training performed on all reconstructed signal candidates results in a stronger expected limit eq.51.

This corresponds to an improvement of approximately 4% in sensitivity. Although the truth-matched sample provides a cleaner representation of the signal topology, the final analysis is performed on reconstructed candidates that include both correctly reconstructed and misreconstructed signal events. Training directly on the full signal sample therefore produces a classifier that is better adapted to the composition of the events entering the final fit. These results indicate that maximizing the purity of the training sample does not necessarily lead to the best expected sensitivity, and that including non-truth-matched signal candidates allows a more efficient selection of the signal population relevant for the measurement.

Beyond the comparison between the two training strategies, it is also instructive to compare the overall performance of the present analysis with that of the previous BaBar search [9]. Although this comparison remains approximate, as no systematic uncertainties are included at this stage, it provides a useful benchmark. Thanks to the improved FEI hadronic tagging algorithm, the number of selected events is increased by approximately a factor of four to five with respect to BaBar. In addition, by also considering a larger number of signal categories in this analysis, the combined figure of merit, $\varepsilon_{\text{sig}}/\sqrt{\varepsilon_{\text{sig}} + N_{\text{bg}}}$, is found to be about 5.6 times larger when comparing the two analyses at the same integrated luminosity. Ultimately, these improvements translate into an expected upper limit that is approximately one order of magnitude more stringent than the one achieved by BaBar.

# References


[1] Andrzej J. Buras. *Weak Hamiltonian, CP Violation and Rare Decays*. 1998. arXiv: `hep-ph/9806471 [hep-ph]`. URL: `https://arxiv.org/abs/hep-ph/9806471`.

[2] Heather E. Logan and Ulrich Nierste. "Bs,d l+l- in a two-Higgs-doublet model". In: *Nuclear Physics B* 586.1 (2000), pp. 39–55. ISSN: 0550-3213. DOI: `https://doi.org/10.1016/S0550-3213(00)00417-X`. URL: `https://www.sciencedirect.com/science/article/pii/S055032130000417X`.

[3] Gino Isidori. "Flavor physics and CP violation". In: *2012 European School of High-Energy Physics*. 2014, pp. 69–105. DOI: `10.5170/CERN-2014-008.69`. arXiv: `1302.0661 [hep-ph]`.

[4] Particle Data Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2024.8 (2024), p. 083C01.

[5] Christine Davies. *Standard Model Heavy Flavor physics on the Lattice*. 2012. arXiv: `1203.3862 [hep-lat]`. URL: `https://arxiv.org/abs/1203.3862`.

[6] Michael Ziegler. "Search for the Decay $B^0 \to \tau^+\tau^-$ with the Belle Experiment". PhD thesis. KIT, Karlsruhe, 2016. DOI: `10.5445/IR/1000063630`.

[7] Christoph Bobeth et al. "$B_{s,d} \to \ell^+\ell^-$ in the Standard Model with Reduced Theoretical Uncertainty". In: *Phys. Rev. Lett.* 112 (10 Mar. 2014), p. 101801. DOI: `10.1103/PhysRevLett.112.101801`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.112.101801`.

[8] R. Aaij et al. "Search for the decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$". In: *Phys. Rev. Lett.* 118.25 (2017), p. 251802. DOI: `10.1103/PhysRevLett.118.251802`. arXiv: `1703.02508 [hep-ex]`.

[9] B. Aubert et al. “Search for the rare decay $B^0 \to \tau^+\tau^-$ at BABAR”. In: *Phys. Rev. Lett.* 96 (2006), p. 241802. DOI: 10.1103/PhysRevLett.96.241802. arXiv: hep-ex/0511015 [hep-ex].

[10] D Besson and T Skwarnicki. “Upsilon Spectroscopy: Transitions in the Bottomonium System”. In: *Annual Review of Nuclear and Particle Science* 43.Volume 43, 1993 (1993), pp. 333–378. ISSN: 1545-4134. DOI: https://doi.org/10.1146/annurev.ns.43.120193.002001. URL: https://www.annualreviews.org/content/journals/10.1146/annurev.ns.43.120193.002001.

[11] Kazunori Akai, Kazuro Furukawa, and Haruyo Koiso. “SuperKEKB Collider”. In: *Nucl. Instrum. Meth. A* 907 (2018), pp. 188–199. DOI: 10.1016/j.nima.2018.08.017. arXiv: 1809.01958 [physics.acc-ph].

[12] T. Abe et al. *Belle II Technical Design Report*. 2010. arXiv: 1011.0352 [physics.ins-det]. URL: https://arxiv.org/abs/1011.0352.

[13] T. Keck et al. “The Full Event Interpretation: An Exclusive Tagging Algorithm for the Belle II Experiment”. In: *Computing and Software for Big Science* 3.1, 6 (Dec. 2019), p. 6. DOI: 10.1007/s41781-019-0021-8. arXiv: 1807.08680 [hep-ex].

[14] Yoshihiko Ozaki, Shuhei Watanabe, and Toshihiko Yanase. “OptunaHub: A Platform for Black-Box Optimization”. In: *arXiv preprint arXiv:2510.02798* (2025).

[15] Geoffrey C. Fox and Stephen Wolfram. “Observables for the Analysis of Event Shapes in $e^+e^-$ Annihilation and Other Processes”. In: *Phys. Rev. Lett.* 41 (23 Dec. 1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581. URL: https://link.aps.org/doi/10.1103/PhysRevLett.41.1581.

[16] Tianqi Chen et al. *xgboost: Extreme Gradient Boosting*. R package version 3.3.0.0. 2026. URL: https://github.com/dmlc/xgboost.

[17] Giovanni Punzi. “Sensitivity of Searches for New Signals and Its Optimization”. In: *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*. Ed. by L. Lyons, R. Mount, and R. Reitmeyer. Jan. 2003, p. 79. DOI: 10.48550/arXiv.physics/0308063. arXiv: physics/0308063 [physics.data-an].

[18] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *European Physical Journal C* 71.2, 1554 (Feb. 2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].