UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS

# DOKTORSKA DISERTACIJA

Matic Lubej

2018

UNIVERSITY OF LJUBLJANA
FACULTY OF MATHEMATICS AND PHYSICS
DEPARTMENT OF PHYSICS

Matic Lubej

# Measurement of the Decay $B^+ \to K^+K^-\ell^+\nu_\ell$ with the Belle Detector

Doctoral thesis

ADVISER: Assist. Prof. Dr. Anže Zupanc

Ljubljana, 2018

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
ODDELEK ZA FIZIKO

Matic Lubej

# Meritev razpada $B^+ \to K^+K^-\ell^+\nu_\ell$ z detektorjem Belle

Doktorska disertacija

MENTOR: doc. dr. Anže Zupanc

Ljubljana, 2018

## Zahvala

Na tem mestu zapišite, komu se zahvaljujete za pomoč pri nastanku doktorske disertacije.

# Izvleček

Kratek izvleček v slovenskem jeziku.

**Ključne besede:**
**PACS:**

## Abstract

Kratek izvle"cek v angle"skem jeziku.

**Keywords:**
**PACS:**

# Contents

# List of figures

# List of tables

# Chapter 1

# Introduction

Particle physics is an established branch of physics with a rich history in theory and experiments ever since the beginning of the 20th century. So far the experimental and theoretical research have shown us hand in hand that the universe consists of particles and force carriers. Particles of matter, or elementary particles, are divided into two groups, quarks and leptons. The quarks that we know today are called $u$ (up), $d$ (down), $s$ (strange), $c$ (charm), $b$ (bottom) and $t$ (top). Leptons are further split into two groups; charged leptons $e$ (electron), $\mu$ (muon), $\tau$ (tau lepton) and their corresponding neutrinos $\nu_e$ (electron neutrino), $\nu_\mu$ (muon neutrino), $\nu_\tau$ (tau neutrino). Particles of force are known as gauge bosons and they are $\gamma$ (photon), $g$ (gluon), $W^\pm$ (charged weak bosons) and $Z^0$ (neutral weak boson). Theory also predicted the recently discovered Higgs boson ($H$), which is responsible for the mass of all particles. Some of the particles above also have a mirrored version of themselves, called antiparticles, which exhibit somewhat different properties as their un-mirrored versions.

Combinations of quarks such as $q_1q_2q_3$ (hadrons) or $q_1\bar{q}_2$ (mesons) can make up heavier particles that we see today. Such particles are protons and neutrons, but also heavier particles which can be produced in processes involving very high energies. Such heavy particles are unstable and decay into lighter ones via forces of nature. Together with the elementary particles and force carriers, three out of four of these forces are joined in a theoretical model called the Standard Model (SM), which is shown in Figure 1.1. They are the electromagnetic, weak nuclear and strong nuclear force. Gravity is not included in the current version of the Standard Model due to its complex and weakly interacting nature. Researching such processes in large experiments enables us to study the mechanism of how elementary particles interact. By doing so we are able to learn the secrets of the universe and how it all began.

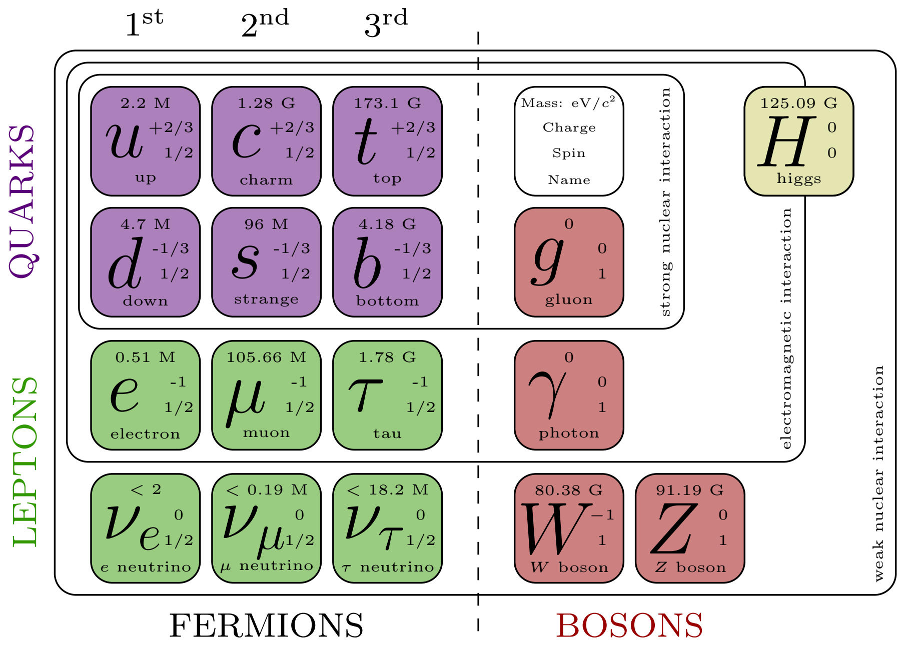



Figure 1.1: A schematic of the Standard Model.

This analysis revolves around decays of the so-called $B$ mesons, which are particles that consist of a $b$ quark and a light $\bar{u}$ or $\bar{d}$ quark (or vice-versa). One of the most surprising features of the universe that can be studied with decays of $B$ mesons is the $CP$ symmetry violation ($\not{C}\not{P}$). $CP$ symmetry is a combination of the $C$ symmetry (charge conjugation) and the $P$ symmetry (spatial inversion). It states that there is no reason why processes of particles and mirrored processes of antiparticles would be different. Today we know that this does not hold for all cases and we, in fact, find processes which violate this postulate. We also know that $\not{C}\not{P}$ is very closely related to the weak nuclear force. Here lies our motivation for studying decays of $B$ mesons, since they exhibit a rich spectrum of decays, many of which underway via the weak nuclear force.

One of the most important properties of the weak nuclear force is that it can change the flavor of particles. Flavor is a quantum number which is conserved for each type of quark, so changing a flavor of a quark means changing the quark itself. Such processes are forbidden for the electromagnetic and the strong nuclear force, but not for the weak one. All of the information regarding quark transitions and transition probabilities can be merged into a form of a complex matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [11, 12]

$$V_{CKM} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}. \tag{1.1}$$

The CKM matrix is a unitary matrix and has only four free parameters which are not described by theory. Its unitarity provides us with several mathematical identities, out of which the most famous one is

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.2}$$

It can be represented by a triangle in the complex plane, called the unitarity triangle, shown in Figure 1.2. The sides and the angles of the unitarity triangle are closely connected to the free parameters of the CKM matrix. It is important to mention that all experimental measurements depend only on these four parameters, so it is possible to determine them by measuring the angles and sides of the unitarity triangle. This way the unitarity triangle offers us a unique way to test the consistency of the SM. The ultimate goal is to then join all such measurements and overconstrain the unitarity triangle to check if all the sides meet. By improving such measurements one can check whether the SM is consistent, or if there are some contributing physics processes that we do not yet understand. Such processes are commonly referred to as "new physics" (NP). The measurements of the sides and angles of the triangle are done by using different decays of which a large portion are $B$ meson decays. Here lies another motivation for using $B$ mesons in the analysis.

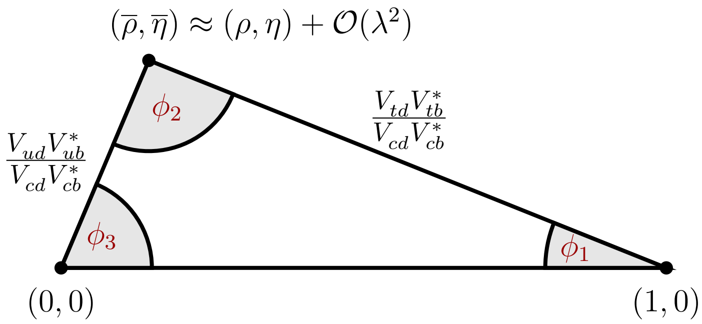



Figure 1.2: The unitarity triangle with $\lambda$, $\eta$, $\rho$ and $A$ (not shown) as free parameters of the CKM matrix.

In this analysis, we focus on the $V_{ub}$ CKM matrix element, which corresponds to $b \to u$ quark transitions. It has the smallest absolute value of all the CKM matrix elements and the largest error, so it offers the most room for improvement. Such quark transitions are present in charmless semi-leptonic $B$ meson decays of the form

$$B^+ \to X_u^0 \ell^+ \nu_\ell, \tag{1.3}$$

where $X_u^0$ represents a charmless hadron with a $u$ quark and $\ell$ is one of the charged leptons $e$, $\mu$ or $\tau$. Measuring the decay rate of the $B$ meson in such decays paves the way for the CKM matrix element determination. Decay rates are directly connected to the $V_{ub}$ element as

$$\mathrm{d}\Gamma \propto G_F^2 |V_{ub}|^2 |L^\mu \langle X_u | \bar{u} \gamma_u \frac{1}{2}(1-\gamma_5) b | B \rangle|^2, \tag{1.4}$$

where $\Gamma$ is the decay width, $G_F$ is the Fermi coupling constant, $L^\mu$ is the leptonic current and the expression in the Dirac brackets is the hadronic current. The factor $|V_{ub}|^2$ represents the probability for the $b \to u$ quark transition. Measurement of the $V_{ub}$ CKM matrix element can be performed in two possible ways. With the exclusive or the inclusive method, which are described below. Both methods require different experimental and theoretical techniques, so they provide largely independent determinations of $|V_{ub}|$. Currently, both methods also have comparable accuracies.

In the exclusive method, one studies the decays of $B$ mesons to a specific charmless hadronic final state, such as $B \to \pi \ell \nu$. Clean determination of the $V_{ub}$ is possible due to precise experimental measurements along with reliable theoretical calculations. However, theoretical calculations are more challenging for decays to a specific final state, since hadronization of quarks has to be taken into account. There are also two main experimental challenges in this method. One has to reduce the abundant background from $B \to X_c \ell \nu$ processes since the $b \to c$ quark transition is much more common. The second experimental challenge is to separate the $B$ meson decay with the specific charmless hadronic final state from other $B \to X_u \ell \nu$ decays since it roughly populates the same regions of the phase-space as the signal decay.

In the inclusive method, one studies the decays of $B$ mesons to any charmless hadronic final state $B \to X_u \ell \nu$. In this case, the total decay rate for $b \to u \ell \nu$ can be calculated accurately since hadronization does not have to be taken into account. The greater challenge with this method is again the experimental measurement of the total decay rate due to the $B \to X_c \ell \nu$ background. Experimental sensitivity to $V_{ub}$ is highest where $B \to X_c \ell \nu$ decays are less dominant. Theory and experiment have to compromise and limit the $V_{ub}$ determination to a region where the signal-to-background ratio is good. Theory takes this into account by reliably calculating the partial decay rate $\Delta\Gamma$, which is more challenging than the total decay rate. One possible and often used approach to reduce $b \to c$ background is to reject all events with $K$ particles, or kaons, present in the final particle selection. The procedure is called a $K$-veto. Kaons consist of an $s$ quark, which is mainly produced in $c \to s$ transitions. This means that if a kaon is found in the event, it is very likely that it originates from a particle with a $c$ quark, indicating the $b \to c$ process.

If $V_{ub}$ is determined with both these methods, the values can be compared. It turns out that consistency between these two results is only marginal, where the difference is at a level of $3\sigma$. The current world averages [13] of the exclusive (from $B^0 \to \pi^- \ell^+ \nu$) and inclusive (GGOU collab. [14]) are

$$|V_{ub}|_{\text{excl.}} = (3.65 \pm 0.09 \pm 0.11) \times 10^{-3}, \tag{1.5}$$

$$|V_{ub}|^{\text{GGOU}}_{\text{incl.}} = \left(4.52 \pm 0.15 \, ^{+0.11}_{-0.14}\right) \times 10^{-3}, \tag{1.6}$$

where the first and the second errors are the experimental and the theoretical error, respectively. We see that inclusive measurements prefer higher values than exclusive ones. This is known as the $V_{ub}$ puzzle. It is necessary to make further research as to why this difference occurs. The reason could be an unknown experimental or theoretical error, or it is even possible that some NP contributions occur. This analysis will focus on a possible reason that could be hidden in the selection mentioned before. By performing a $K$-veto, one discards all events with kaons in the final state in order to suppress $b \to c$ contributions. In this analysis, we focus on the charged $B \to KK\ell\nu$ decay, which is very similar to the $B \to \pi\ell\nu$, except for a production of an $s\bar{s}$ quark pair, which then combines with final state quarks to form kaons, as shown in Figure 10.2. In this case, we have kaons in the final state where the $B$ meson decayed via a $b \to u$ process. Such decays were discarded in previous $V_{ub}$ determinations with the inclusive method, but in principle, they contribute to the result and should be taken into account. The results of this analysis should help us make a step closer to solving the $V_{ub}$ puzzle.

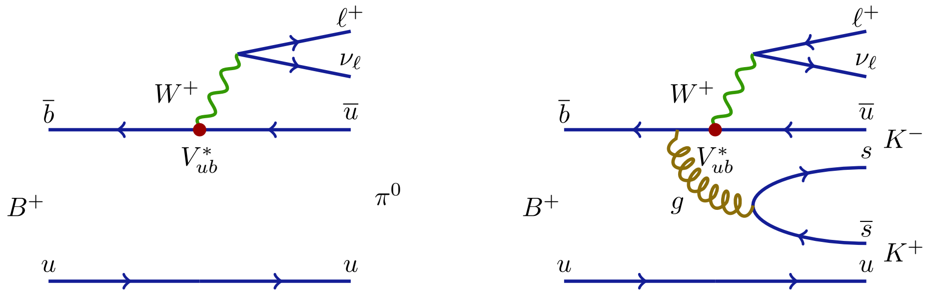



Figure 1.3: Feynman diagrams for the $B^+ \to \pi^0 \ell^+ \nu_\ell$ decay (left) and the $B^+ \to K^- K^+ \ell^+ \nu_\ell$ decay (right).

Specifically, we will be focusing on decays of the charged $B$ mesons of the form $B^+ \to K^+ K^- \ell^+ \nu$, since it includes two charged kaons, as opposed to the case of the neutral $B$ meson decay. The reason for this is a simpler decay chain and a higher reconstruction efficiency. All further occurrences of $B \to KK\ell\nu$ automatically imply decays of the form $B^+ \to K^+ K^- \ell^+ \nu$ and its charge conjugated counterpart.

# Chapter 2

# Data and Monte-Carlo Samples

The Belle detector acquired a dataset of about $L_0 \approx 710\ \mathrm{fb}^{-1}$ of integrated luminosity in it's lifetime, which corresponds to about $771 \times 10^6$ $B\bar{B}$ meson pairs. Additionally, several streams of Monte-Carlo (MC) samples were produced, where each stream of MC corresponds to the same amount of data that was taken with the detector. The main focus of this and other similar analyses is to study a rare signal decay, which means that the amount of such decays in the existing MC is not abundant enough. In such cases, it is a common practice to produce specific samples of signal MC, where the abundance of signal decays is much larger, enabling us to study its properties in greater detail.

The following samples were used in this analysis

- data
  - Belle on-resonance dataset of about $L_0$ integrated luminosity, measured at $\Upsilon(4S)$ resonance energy,
  - Belle off-resonance dataset of about $1/10 \times L_0$ integrated luminosity, measured at 60 MeV below $\Upsilon(4S)$ resonance energy,
- signal MC, corresponding to about $400 \times L_0$,
- other MC
  - generic on-resonance, 10 streams of $B\bar{B}$ (denoted as `charged` and `mixed`) and 6 streams of $q\bar{q}$ produced at $\Upsilon(4S)$ resonance energy, where each stream corresponds to $L_0$,
  - generic off-resonance, 6 streams of $q\bar{q}$ produced at 60 MeV below $\Upsilon(4S)$ resonance energy, where each stream corresponds to $1/10 \times L_0$,
  - $B \to X_u \ell \nu$ (denoted as `ulnu`), not included in previous MC samples, equal to an amount of $20 \times L_0$,
  - other rare $B$ meson decays (denoted as `rare`), not included in previous MC samples, equal to an amount of $50 \times L_0$.

## 2.1 Signal MC Production

The signal MC sample of $B^+ \to K^+ K^- \ell \nu_\ell$ and the charge conjugated $B^-$ decays was produced using the `mcproduzh` package for producing Belle MC. The package accepts a decay file, which describes the decays to be generated. The decay file used for signal MC generation was the same as for the `ulnu` sample since it includes the decays of interest. An additional skim was applied in order to select only events of interest with at least 2 kaons and a light lepton, all coming from the same particle. This decreases the CPU consumption during the detector simulation and reconstruction.

The relevant processes which contribute to our signal decay are

- $B^+ \to a_{00}\ell^+\nu_\ell$,
- $B^+ \to a_{20}\ell^+\nu_\ell$,
- $B^+ \to f_2\ell^+\nu_\ell$,
- $B^+ \to f_0\ell^+\nu_\ell$,
- $B^+ \to X_u^0\ell^+\nu_\ell$,

where $a_{00}$, $a_{20}$, $f_2$ and $f_0$ are light unflavored states which include further decays into a $K^+K^-$ pair, and $X_u^0$ represents a generic $u\bar{u}$ quark pair, which further hadronizes based on the `PYTHIA` quark hadronization model [1]. Figure 2.1 shows the invariant mass of the $KK$ pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structures, while $KK$ pairs from the $X_u^0$ state are more frequent and follow a wider and smoother distribution.

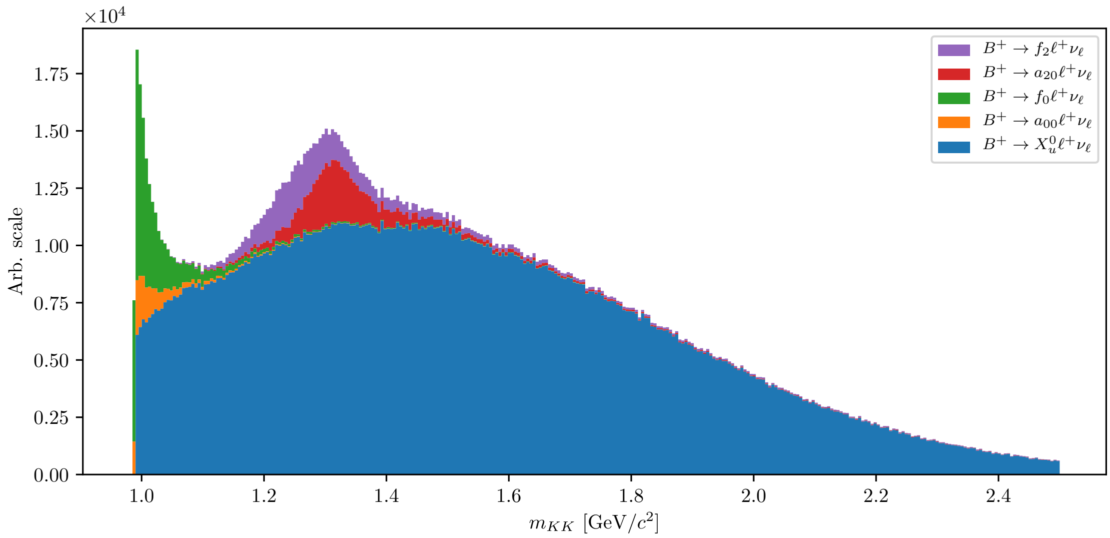



Figure 2.1: Invariant mass of the $KK$ pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structurer, while $KK$ pairs from the $X_u^0$ state are more frequent and follow a wider and smoother distribution.

The produced signal MC sample contains decays of the form $B \to KK\ell\nu$ as well as $B \to KKX\ell\nu$, where $X$ can be any hadron as long as it satisfies all the selection rules of the decay. It is possible to calculate the MC branching ratios for each channel by making combinations of the particles directly from the generator. Table 2.1 shows some of the most prominent channels, which are similar to our signal decay, as well as their relative fraction. It is clear that our signal decay is the most abundant one, with a relative contribution of about 28.14 %, while other channels contribute only up to about 8 % or less. Additionally, our signal decay is the cleanest, while other decays include neutral particles like $\pi^0$, which are harder to reconstruct and suffer from decrease in efficiency due to reconstruction effects.

| Channel | Ratio [%] |
|---|---|
| $K^+K^-$ | 28.14 |
| $K^+K^-\pi^0$ | 8.94 |
| $K^+\bar{K}^0\pi^-$ | 8.71 |
| $K^0K^-\pi^+$ | 8.70 |
| $K^+K^-\pi^+\pi^-$ | 4.15 |
| $K^0\bar{K}^0$ | 3.32 |
| $K^0\bar{K}^0\pi^0$ | 3.26 |
| $K^+K^-\rho^0$ | 1.93 |
| $K^+\bar{K}^0\rho^-$ | 1.84 |
| $K^0K^-\rho^+$ | 1.83 |

| Channel | Ratio |
|---|---|
| $K^+K^-\pi^0\pi^0$ | 0.86 |
| $K^+K^-\pi^+\rho^-$ | 0.69 |
| $K^+K^-\rho^+\pi^-$ | 0.68 |
| $K^0\bar{K}^0\rho^0$ | 0.00 |
| $K\bar{K}$ pair with $\eta$ | 7.08 |
| $K\bar{K}$ pair with $\omega$ | 5.33 |
| Other | 14.53 |

Table 2.1: Relative branching ratios of $B \to KKX\ell\nu$ decays by channel.

We generate about $1.3 \times 10^9$ events of the form $B \to X_u\ell\nu$, which corresponds to an integrated luminosity of about $L = 400 \times L_0$, where this value was obtained by normalizing the signal MC to the amount of signal in the $B \to X_u\ell\nu$ MC sample. This amounts to a total of about $9.37 \times 10^6$ generated signal events, and to a branching ratio

$$\mathcal{B}\left(B^+ \to K^+K^-\ell^+\nu_\ell\right)_{MC} = 1.53 \times 10^{-5}, \tag{2.1}$$

where $\ell$ is $e$ or $\mu$. During analysis, the abundant signal MC sample is scaled down to correspond to the amount of data taken with the Belle detector.

## 2.2 Control Decay

In this analysis, we are also able to define another $B$ meson decay which occupies almost the same phase space as our signal decay. This process can be used for the monitoring of our analysis steps, which are applied to both measured and simulated data. Any kind of difference between the two might indicate our procedure to be fine-tuned to simulated data, or some other similar problem.

We define a control decay of the form

$$B^+ \to \bar{D}^0\ell^+\nu, \quad D^0 \to K^+K^-,$$

which is much more abundant and, most importantly, easy to suppress since it only populates a very narrow region in the kaon invariant mass spectrum. Due to no extra particles in the $D^0$ decay, the kaon invariant mass is equal to $m_{KK} \approx m_{D^0}$ up to very good precision. By excluding this narrow region we discard the majority of the control candidates while discarding only a small amount of the signal candidates. A more quantitative description of suppressing control and other background candidates is written in chapter 7.

# Chapter 3

# Experimental Setup

The data used in this analysis were produced in $e^+e^-$ collisions at the KEKB accelerator and collected with the Belle detector. The experiment was hosted at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. The experiment ran from years 1999 to 2010, collecting data at and near the energy of the $\Upsilon(4S)$ resonance. This chapter briefly describes the accelerator and the detector, based on detailed reports from [16] and [17], respectively.

## 3.1 KEKB Accelerator

KEKB is an asymmetric $e^+e^-$ collider, composed roughly of an electron source and a positron target, a linear accelerator (Linac) and two separate main rings with a circumference of about 3 km as shown in Figure 3.1. Electrons are first produced by a thermal electron gun and accelerated in the Linac to an energy of about 8 GeV. Part of the electrons collide with a tungsten target to produce positrons, which are accelerated in the Linac to an energy of about 3.5 GeV. Electron and positron beams are injected into the high- (HER) and low energy ring (LER) where they collide as bunches of particles at a single interaction point (IP) at an angle of about 22 mrad. The combined center-of-mass (CM) energy of the collision corresponds to the mass of the $\Upsilon(4S)$ resonance

$$E_{CM} = 2\sqrt{E_{e^+}E_{e^-}} = m_{\Upsilon(4S)}c^2 \approx 10.58 \text{ GeV}. \tag{3.1}$$

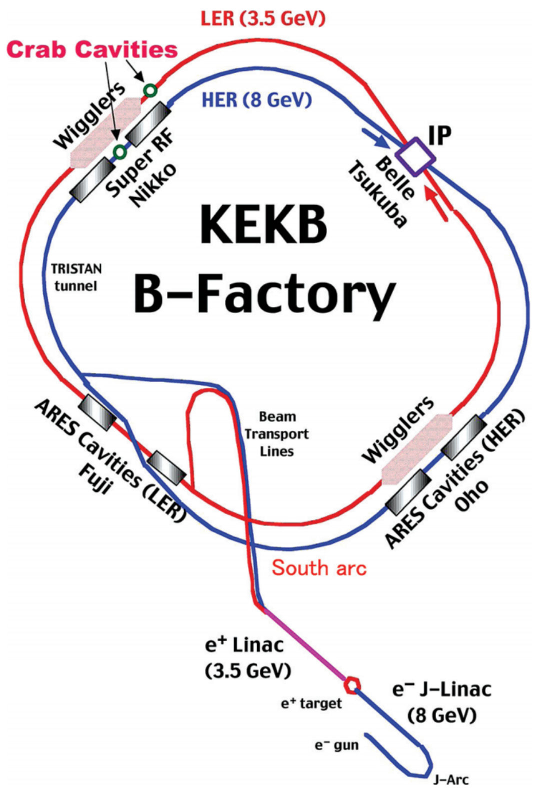



Figure 3.1: Schematic layout of the KEKB accelerator. The HER and the LER are the $e^-$ and the $e^+$ beams, respectively. Four experimental halls, FUJI, NIKKO, OHO and TSUKUBA are shown.

The $\Upsilon(4S)$ state is produced only in a fraction of all collisions, but when it is produced, it predominantly decays to a pair of charged or neutral $B$ mesons. This setup was chosen in accordance with the main goal of the experiment, which was to study CP violation in the $B$ meson system. In other cases, the processes include $e^+e^-$ scattering, also known as Bhabha scattering, two-photon events, muon or tau lepton pair production, and production of $q\bar{q}$, where $q = u$, $d$, $s$ or $c$. Table 3.1 shows the cross-sections for all mentioned interactions in collisions of $e^+e^-$. In addition to the nominal CM energy, the experiment collected data also at energies corresponding to other $\Upsilon(nS)$ resonances, where $n = 1$, 2, 3, 5, and also at energies below the resonances.

| Interaction | Cross-section [nb] |
|---|---|
| $\Upsilon(4S) \to B\bar{B}$ | 1.2 |
| $q\bar{q},\ q \in [u, d, s, c]$ | 2.8 |
| $\mu^+\mu^-,\ \tau^+\tau^-$ | 1.6 |
| Bhabha scattering (within detector acceptance) | 44 |
| Other QED processes (within detector acceptance) | $\sim 17$ |
| Total | $\sim 67$ |

Table 3.1: Cross-sections with $L = 10^{34}$ cm$^{-2}$s$^{-1}$ for various physics processes at $\Upsilon(4S)$ resonance energy [17].

KEKB achieved the world-record for the peak luminosity of $2.11 \times 10^{34}$ cm$^{-2}$s$^{-1}$, twice as much as the designed prediction, and the total integrated luminosity of

1041 $\text{fb}^{-1}$. Of the full Belle dataset, about 711 $\text{fb}^{-1}$ of data were taken at the $\Upsilon(4S)$ energy of 10.58 GeV, which corresponds to about $771 \times 10^6$ $B\bar{B}$ meson pairs.

## 3.2 Belle Detector

The Belle detector is a magnetic mass spectrometer which covers a large solid angle. It is designed to detect remnants of $e^+e^-$ collisions. The detector is configured around a 1.5 T superconducting solenoid and iron structure surrounding the interaction point (IP). The 4-momentum of the decaying $B$ mesons and it's decayed daughter particles are determined via a series of sub-detector systems, which are installed in an onion-like shape. Short-lived particle decay vertices are measured by the silicon vertex detector (SVD) situated outside of a cylindrical beryllium beam pipe. Long-lived charged particle momentum is measured via tracking, which is performed by a wire drift chamber (CDC). Particle identification is provided by energy-loss measurements in CDC, aerogel Cherenkov counters (ACC) and time-of-flight counters (TOF), situated radially outside of CDC. Particles producing electromagnetic showers deposit energy in an array of CsI(Tl) crystals, known as the electromagnetic calorimeter (ECL), which is located inside the solenoid coil. Muons and $K_L$ mesons (KLM) are identified by arrays of resistive plate counters in the iron yoke on the outside of the coil.

The coordinate system of the Belle detector originates at the IP, with the $z$ axis pointing in the opposite direction of the positron beam, the $x$ axis pointing horizontally out of the ring, and the $y$ axis being perpendicular to the aforementioned axes. Th electron beam crosses the positron beam at an angle of about 22°. The polar angle $\theta$ covers the region between $17° \leq \theta \leq 150°$, while the cylindrical angle $\varphi$ covers the full 360° range, amounting to about 92% coverage of the full solid angle.

### 3.2.1 Silicon Vertex Detector

SVD is the inner-most part of the Belle detector and serves the purpose of measuring the decay vertices of decaying particles. The precision of the subsystem is about 100 $\mu$m, which is important for measuring the difference in $z$-vertex positions of the $B$ mesons in time-dependent CP violation studies. The main parts of the SVD are the double-sided silicon detectors (DSSD). With their thin profile and parallel silicon strips on both sides they provide 2D hit information of charged particle and are perfect for a small-scale device which acts with high precision.

During the data taking period, two configurations of the SVD have been used. The first, SVD1, has three layers of DSSD detectors, positioned at 30, 45.5 and 60 mm away from the IP. They compose a ladder-like structure, covering the polar angle of $23° < \theta < 140°$. This configuration was used from the beginning of the experiment until 2003 when a dataset of about $1.52 \times 10^8$ pairs of $B\bar{B}$ mesons were recorded. Due to problems with radiation hardness, a new configuration was used, SVD2, which was operational until the end of data taking, measuring about $6.20 \times 10^8$ pairs of $B\bar{B}$ mesons. The SVD2 has 4 layers of DSSD detectors positioned at 20, 43.5, 70 and 80 mm away from the IP and covered the polar angle of $17° < \theta < 150°$. The first layer was moved closer to the IP, which greatly improved the sub-system precision, due to multiple-Coulomb scattering affecting resolution more as the distance from the IP increases. The front and side view of the SVD2 are shown in Figure 3.2.

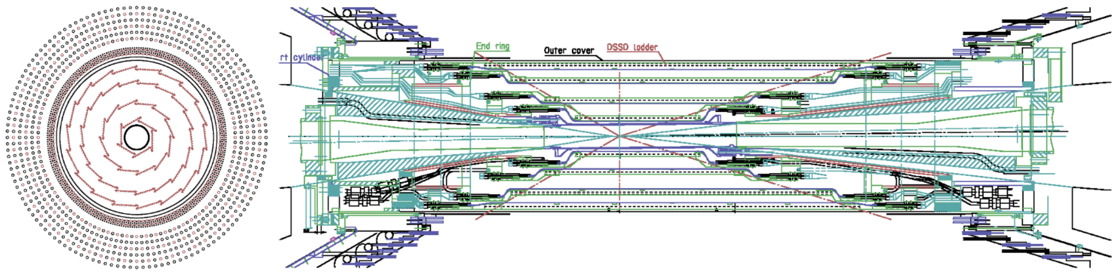



Figure 3.2: Front (left) and side (right) view of the SVD detector with the SVD2 configuration. The front view also shows the inner wires of the Drift Chamber [18].

The efficiency of the SVD was determined as a fraction of CDC tracks within the SVD acceptance that have associated SVD hits, needed for the $B$ meson reconstruction. The average efficiency is found to be around 98% and is in agreement with simulation. SVD performance is also determined via the impact parameter $z$ and $r\phi$ resolution, which was obtained from cosmic ray data. The momentum and angular dependence of the impact parameters is shown in Figure 3.3 and is well represented by the following parametrization for the SVD2

$$\sigma_z = 28\ \mu\text{m} \oplus \frac{32\ \mu\text{m}}{(p/(1\ \text{GeV}/c))} \frac{1}{\beta \sin^{5/2}\theta}, \tag{3.2}$$

$$\sigma_{r\phi} = 22\ \mu\text{m} \oplus \frac{36\ \mu\text{m}}{(p/(1\ \text{GeV}/c))} \frac{1}{\beta \sin^{3/2}\theta}, \tag{3.3}$$

where $p$ is the particle momentum, $\theta$ is the polar angle, and $\beta = v/c$. An advantage of the smaller distance between the IP and the first DSSD layer in SVD2 is clearly seen.

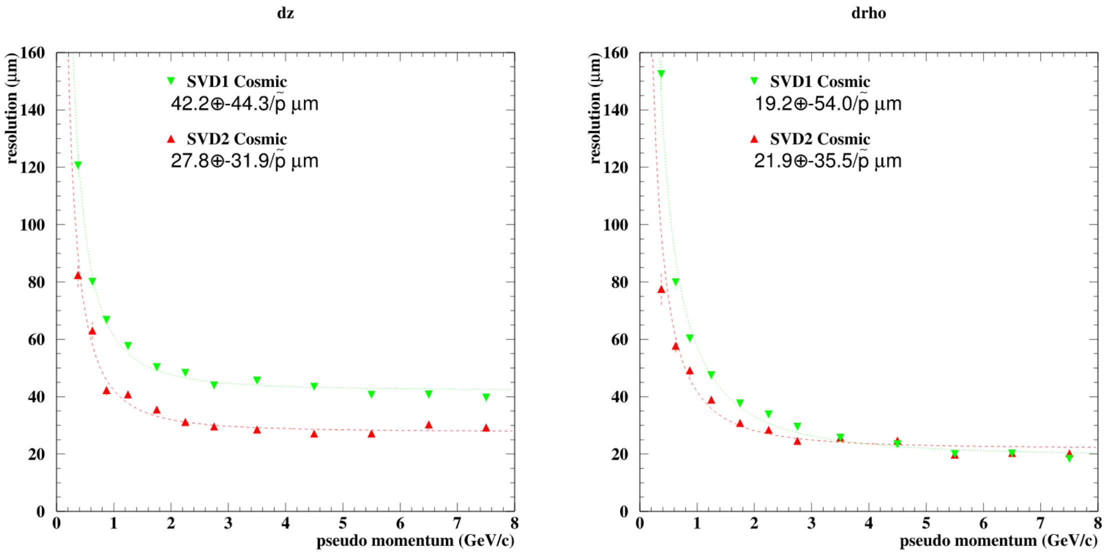



Figure 3.3: Impact parameter resolutions of $z$ (left) and $r\phi$ (right) coordinates for the SVD1 and SVD2 configuration of the vertex detector [18].

### 3.2.2 Central Drift Chamber

CDC is a large-volume tracking device located at the central part of the Belle detector and is crucial for measurements of the particle trajectories and momenta, but also serves as a particle identification device (PID). It has a cylindrical structure with a radius of 88 cm, length of 2.4 m and acceptance equal to the one of SVD2. The chamber has a total of 8400 wires, which are positioned in 50 layers and describe a nearly square wire configuration. There are two types of wires – field wires for producing the electrical field, and sense wires for detecting the particles. Odd-numbered wire layers are oriented in the $z$ direction and provide a measurement of the transverse momentum $p_t$, while even-numbered wires are inclined with respect to the $z$ axis by a small angle of $\pm 50$ mrad to allow for measuring of the polar angle of the track. The wire configuration is shown in Figure 3.4. The space between the wires is filled with a gas mixture of 1 : 1 helium-ethane, a low-$Z$ gas in order to minimize multiple-Coulomb scattering contributions to momentum resolution, since the majority of particles in $B$ events have a momentum lower than 1 GeV/$c$. It also has a small cross section of the photoelectric effect, which is important to reduce background electrons induced by the synchrotron radiation from the beam.

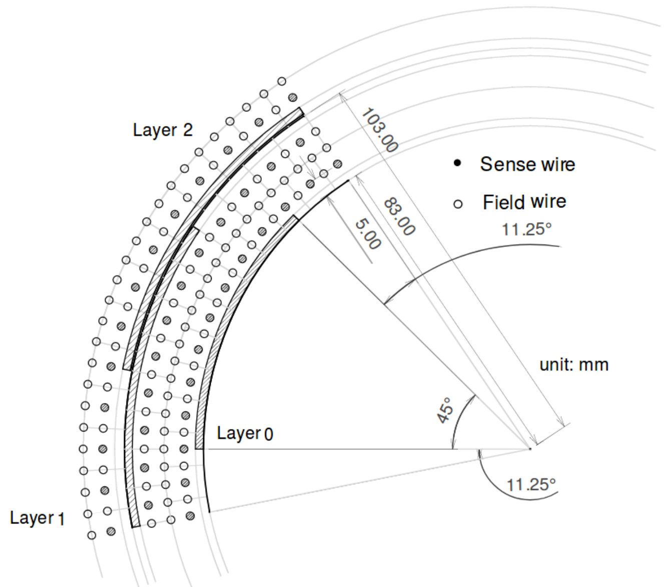



Figure 3.4: Cell structure of CDC [17].

Charged particles which pass the CDC wire frame cause gas ionization. The produced electrons drift toward the sense wires with great acceleration due to the strong electric field close to the wire. The accelerated electrons collide with gas molecules and produce secondary, tertiary etc. ionizations, which result in an electron avalanche, a process which increases the signal by many orders of magnitude. The primary electrons also have a specific drift velocity, which allows us to relate the measured pulse height and drift time to the energy deposit of the particle as well as the distance from the sense wire. This information is important for calculating the energy loss $\mathrm{d}E/\mathrm{d}x$. $\mathrm{d}E/\mathrm{d}x$ as a function of momentum differs for different particles, as shown in Figure 3.5. This allows for identification purposes of, specifically for kaons and pions. In the momentum region less than 0.8 GeV/$c$, $\mathrm{d}E/\mathrm{d}x$ enables a separation between kaons and pions up to $3\sigma$. The resolution of the transverse momentum measurement with the CDC is a function of the transverse momentum itself, as well as the particle velocity, and is parametrized as

$$\sigma(p_T)/p_T = \frac{0.201\%\ p_T}{1\ \mathrm{GeV}/c} \oplus \frac{0.290\%}{\beta}. \tag{3.4}$$

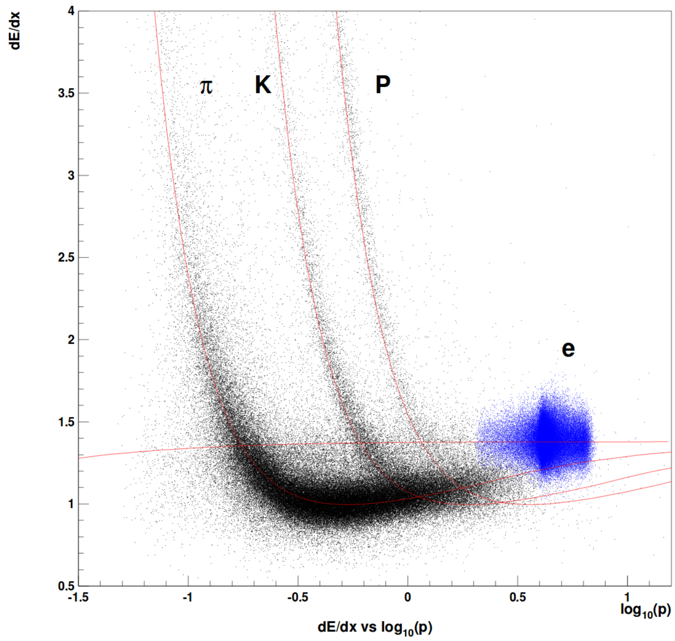



Figure 3.5: Measured d$E$/d$x$ as a function of particle momentum. The red lines show the expected distribution for different types of particles [17].

### 3.2.3 Time-of-Flight Counter

The purpose of the TOF subdetector is particle identification in the momentum region 0.8 GeV/$c$ $< p <$ 1.2 GeV/$c$, especially for kaons and pions. There are 64 TOF modules in the barrel region, covering the polar angle of $33° < \theta < 121°$. One TOF module consists of two long polyvinyl toluene-based plastic scintillator bars, 4 fine-mesh photo-multiplier tubes (PMT) at the 4 ends of the bars, and a trigger scintillation counter, where the latter provides additional trigger information. TOF measures the time interval between the $e^+e^-$ collision and the passage of the particle through it. The mass of a particle can be inferred via the relation

$$m^2 = \left(\frac{1}{\beta^2} - 1\right) p^2 = \left(\frac{T^2c^2}{L^2} - 1\right) p^2, \tag{3.5}$$

where $T$ is the measured time interval, $L$ is the charged particle trajectory length from the IP to TOF and $p$ is the charged particle momentum, determined by SVD and CDC. The resulting mass distribution for charged tracks measured by TOF in hadron events is shown in Figure 3.6, where clear peaks corresponding to pions, kaons and protons can be seen. To achieve the good discrimination between kaons and pions, a time-of-flight resolution of less than 100 ps is needed for particles with momentum below about 1.2 GeV/$c$, which encompasses 90% of the particles produced in $\Upsilon(4S)$ decays. The identification power can also be determined in the form of $\pi^\pm/K^\pm$ separation significance as a function of particle momentum, shown in Figure 3.7. A clear separation of about $2\sigma$ is achieved for particle momenta up to 1.25 GeV/$c$.

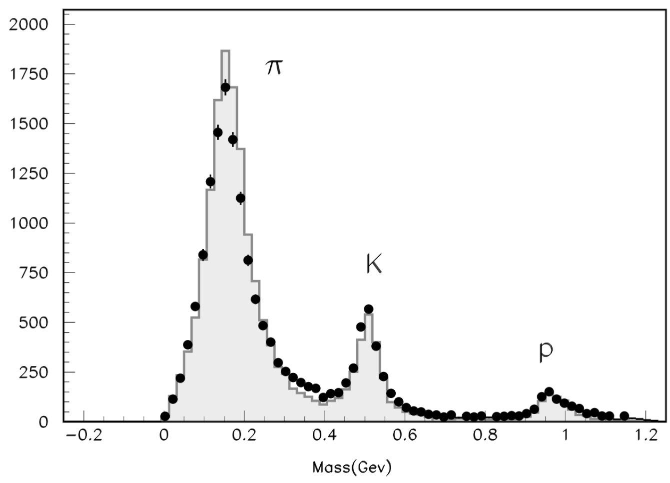



Figure 3.6: Mass distribution from TOF measurements for particle momenta below 1.2 GeV/$c$ [17].

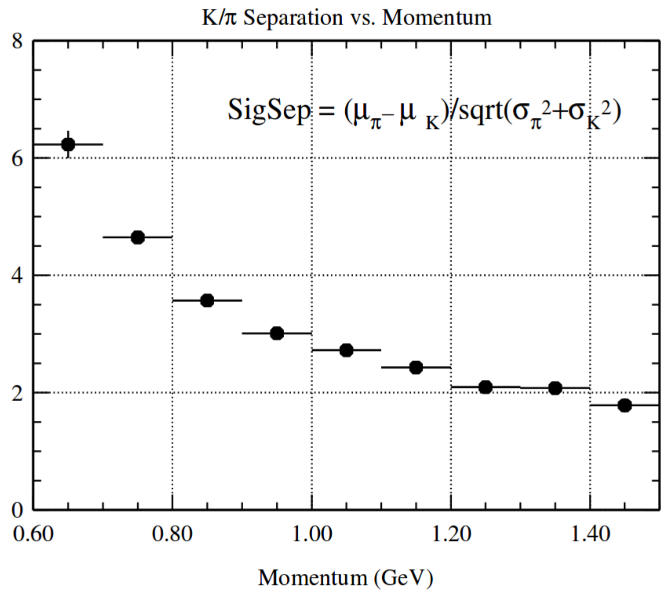



Figure 3.7: $\pi^{\pm}/K^{\pm}$ separation by TOF [17].

### 3.2.4 Aerogel Cherenkov Counter

TOF is not capable of performing good PID above 1.2 GeV/$c$ momentum since $\beta$ is almost equal to 1. For higher momenta in the region 1.0 GeV/$c$ < 4.0 GeV/$c$, the ACC is introduced. It is a threshold-type Cherenkov counter which utilizes the fact that particles emit Cherenkov light if the particle speed is greater than the speed of light in the passing medium. ACC is introduced in the barrel region with 960 separate modules, covering a polar angle of $34° < \theta < 127°$ and 228 modules in the forward end-cap region, with the polar angle coverage of $17° < \theta < 34°$. Each ACC module consists of an aluminum encased block of silica aerogel and one or two fine-mesh PMTs encased on each block to detect Cherenkov light pulses. Due to the polar angle dependence of the particle momentum, 6 different refractive indices are chosen for the aerogel material, ranging from 1.010 up to 1.030 and are controlled within 3% precision. The layout of the ACC is shown in Figure 3.8.

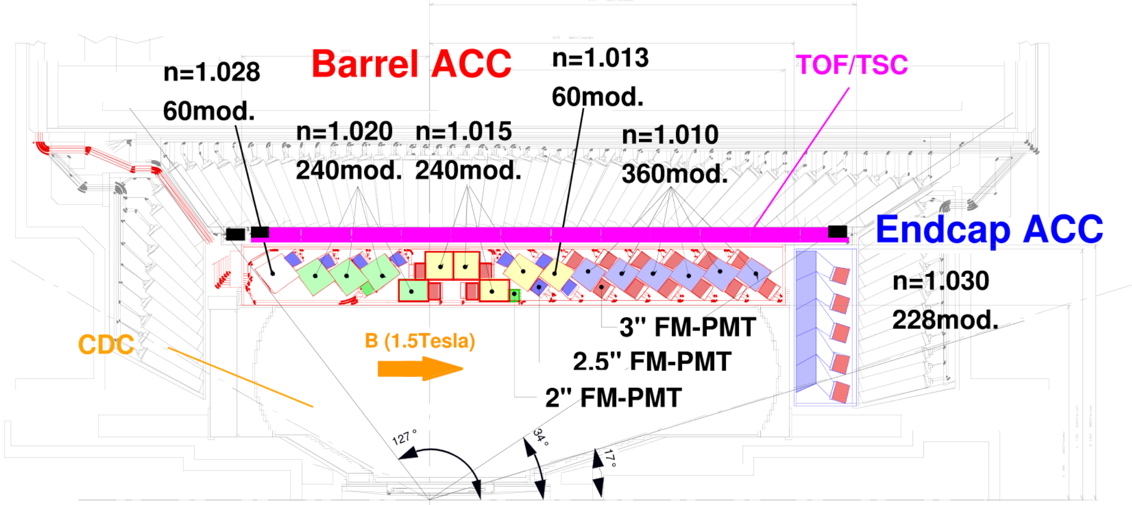



Figure 3.8: Cross-sectional view of the CDC (inner most), ACC and TOF (outer most) detectors [17].

The threshold velocity $\beta$ of a given particle for Cherenkov radiation is

$$\beta \leq \frac{1}{n}, \tag{3.6}$$

where $n$ is the refractive index of the medium. The refractive indices in the ACC are such that, due to different masses, pions will emit Cherenkov light and kaons will not, due to different masses of the particles. Using the PID of ACC, along with other sub-system PID info, the electron identification efficiency in the momentum range above 1 GeV/$c$ is equal to or above 90% while the pion fake rate, the probability of wrongly identifying pions as electrons, to be around 0.2 - 0.3%. Similarly for kaons, kaon ID efficiency is equal to 80% for most of the momentum region up to 4 GeV/$c$, while pion fake rate remains below 10%. Figure 3.9 shows the electron and kaon efficiencies and the corresponding pion fake rates as a function of particle momenta.

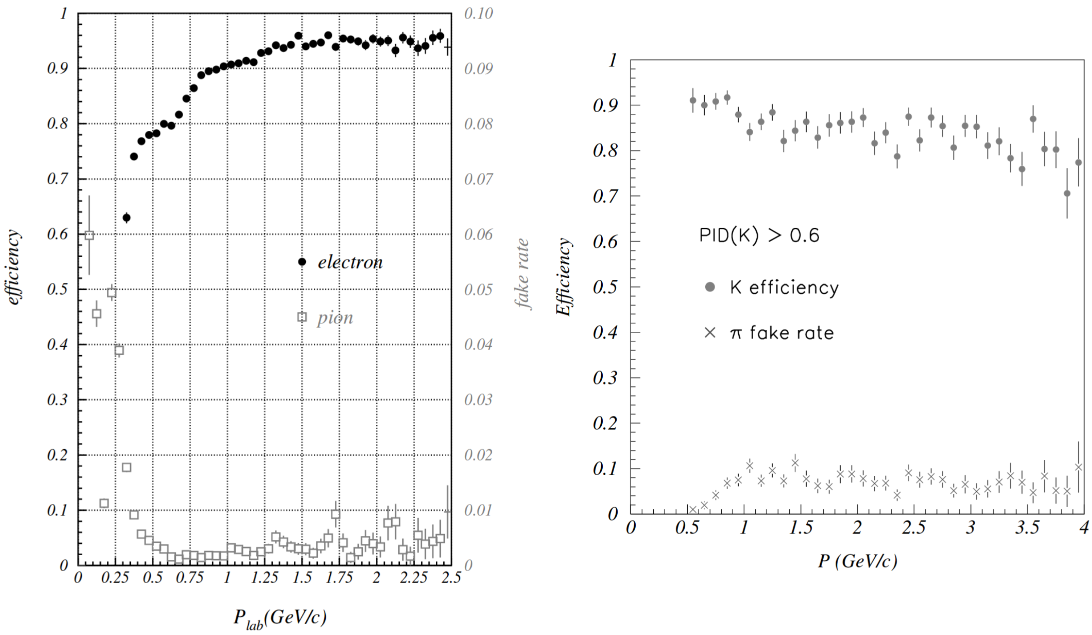



Figure 3.9: Electron identification efficiency and fake rate for charged pions (left) and similarly for kaons (right). Note the different scales for the electron efficiency and fake rate in the former case [17].

### 3.2.5 Electromagnetic Calorimeter

Measurement of position and energy deposit of particles is performed in the ECL, especially for electrons and photons, where the latter are not measured by any of the subsystems described so far. It also provides complimentary particle identifications for electrons versus pions. The calorimeter consists of a highly segmented array of thallium-doped cesium iodide (CsI(Tl)) in the form of tower-shaped crystals, each pointing towards the IP. Each crystal is about 30 cm long with a width from 44.5 mm to 65 mm in the barrel, and from 44.5 mm to 82 mm in the end-caps. Out of a total of 8736 crystals with a total mass of about 43 tons, 6624 of them are positioned in the barrel region and 1152 (960) in the forward (backward) end-caps. The inner radius of the barrel section is about 1.25 m, while the end-caps are positioned at $-1.0$ m and 2.0 m from the IP in direction of the $z$ axis. The polar angle coverage of the barrel region is $32.2° < \theta < 128.7°$, and for the end-caps $12.4° < \theta < 31.4°$ and $130.7° < \theta < 155.1°$. Figure 3.10 shows the layout of the barrel and end-caps ECL.

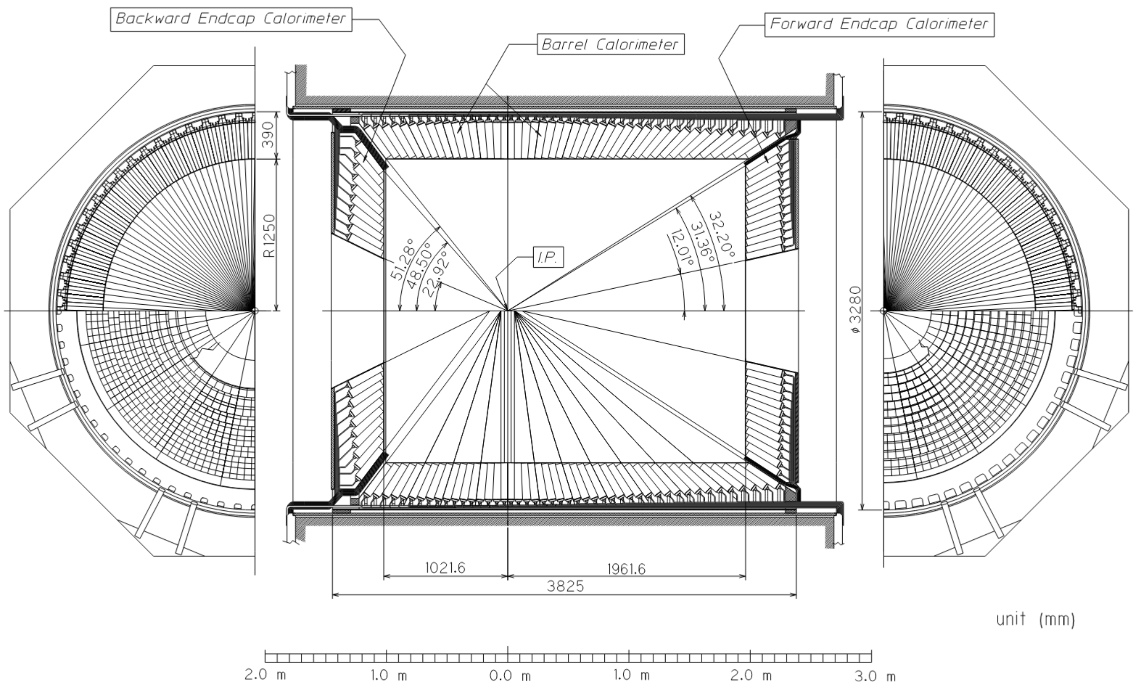



Figure 3.10: Overall configuration of the ECL [17].

When an electron or a photon hits a crystal, it produces an electromagnetic shower, a result of the bremsstrahlung and pair-production effects. Heavier charged particles do not interact in the same way and deposit only a small amount of energy by ionization effects. The information from the ECL, compared with momentum measurements provided by the CDC, enables the identification of electrons. The distribution of the deposited energy for different particles is shown in Figure 3.11. The probability of misidentifying an electron as a pion is approximately 5% for momenta less than 1 GeV/$c$, and less than 1% for momenta above 2 GeV/$c$.

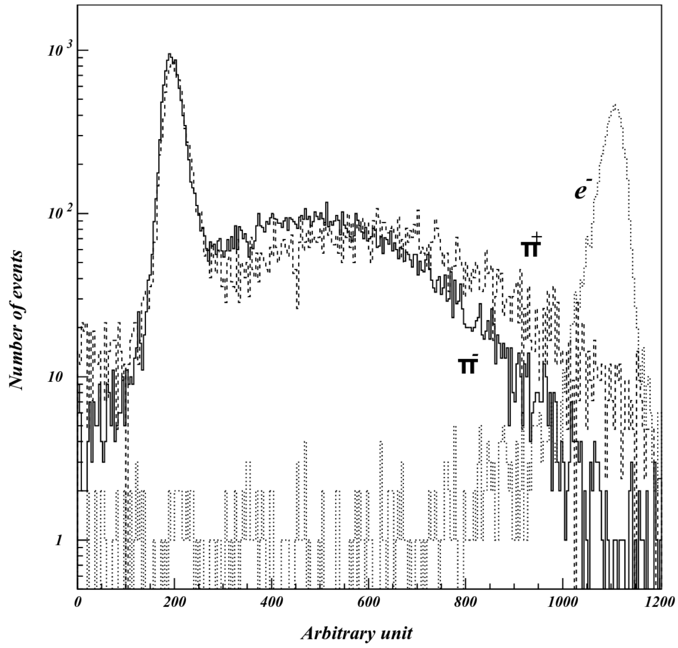



Figure 3.11: Distribution of the energy deposit by electrons and charged pions at 1 GeV/$c$ momentum [17].

For ECL calibration, $e^+e^- \rightarrow e^+e^-$ and $e^+e^- \rightarrow \gamma\gamma$ events were used. The average energy resolution was achieved to be 1.7% for the barrel ECL, and 1.74% and 2.85% for the forward and backward ECL, respectively, as shown in Figure 3.12. These value are in good agreement with Monte Carlo predictions. Worse energy resolution in backward end-cap is due to the lower photon energy, which results in larger effects of passive material in front of the calorimeter [18]. The energy resolution as a function of energy can be obtained via the following relation

$$\frac{\sigma_E}{E} = \frac{0.0066\%}{(E/1\text{ GeV})} \oplus \frac{1.53\%}{(E/1\text{ GeV})^{1/4}} \oplus 1.18\%, \tag{3.7}$$

while the resolution of the position measurement is

$$\sigma_{pos} = 0.27\text{ mm} + \frac{3.4\text{ mm}}{(E/1\text{ GeV})^{1/2}} + \frac{1.8\text{ mm}}{(E/1\text{ GeV})^{1/4}}. \tag{3.8}$$

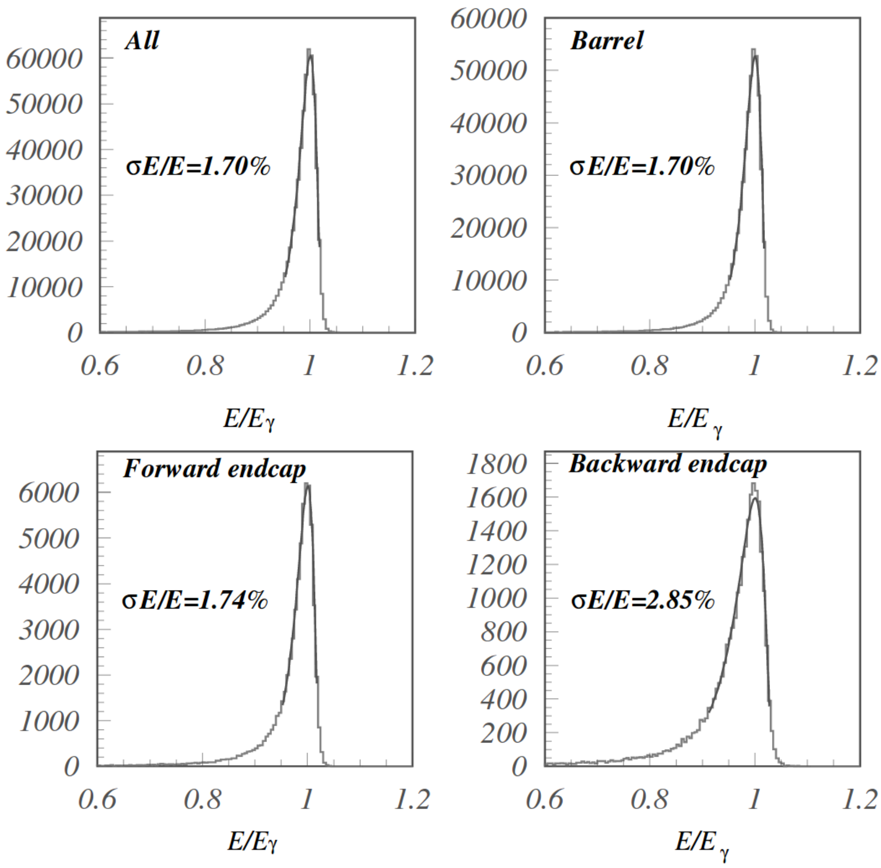



Figure 3.12: Reconstructed energy distribution for $e^+e^- \to \gamma\gamma$ events for overall, barrel, forward and backward end-cap calorimeters [17].

### 3.2.6 $K_L^0/\mu$ Detector

The KLM detector is used for detection of high-penetration particles such as $K_L^0$ and $\mu$ for momenta larger than 0.6 GeV/$c$. The setup covers the polar angle of $20° < \theta < 155°$. Detection of $K_L^0$ particles is troublesome since they are neutral and have a small material interaction probability, therefore a lot of material is needed in the KLM. To provide detection of both kinds of particles, hadronic and neutral, as well as electromagnetically and hadronically interacting, the KLM is constructed as a sampling calorimeter, which consists of 15 layers of 3.7 cm thick resistive-plate counters (RPC) with 14 layers of 4.7 cm thick iron plates between them. A single RPC module consists of two parallel plate electrodes, two glass panels, and gas in between. A charged particle passing the gas gap initiates a local discharge of the plates, which in turn induces signal to record the time and location of ionization. This is possible since the resistivity of the glass surface is high, so the discharge occurs locally. Hadrons interacting with the iron plates may produce a shower of ionizing particles, which are then also detected by the RPCs. The KLM is located outside of the superconducting solenoid and the iron plates of the KLM serve a dual role as the flux return for the magnetic field. Figure 3.13 shows a cross-section of an RPC superlayer, consisting of an RPC pair.

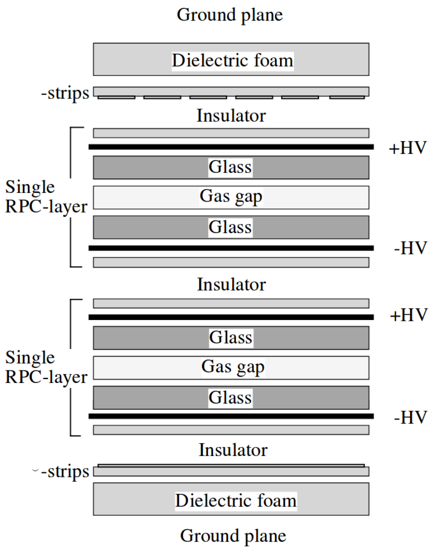



Figure 3.13: Cross-section of an RPC superlayer, consisting of an RPC pair [17].

The $K_L^0$ particle can be distinguished from other charged hadrons because they have no matched track in the CDC. The flight direction can also be inferred from the hit locations in the consecutive RPCs. Tracks of charged particles measured in CDC are extrapolated into KLM and clusters within 15° of an extrapolated charged particle track are excluded from $K_L^0$ cluster candidates. On the other hand, muons with matched CDC tracks are able to reach the KLM if their momentum is larger than 0.5 GeV/$c$. They do not interact strongly and do not produce hadronic showers in the KLM, which serves as a handle on the muon identification. Figure 3.14 (left) shows the number of neutral clusters per event and a Monte Carlo simulation of the predicted number of $K_L^0$ clusters per event. The average number of $K_L^0$ clusters per event is 0.5. The agreement with the prediction gives us the confidence that the detector and our reconstruction software are performing correctly. Figure 3.14 (right) shows the muon detection efficiency as a function of momentum and shown for a likelihood cut of 0.66, where muon likelihood is based on the comparison of the measured range of a particle with the predicted range for a muon. Based on $K_S \to \pi^+\pi^-$ events, a muon identification efficiency of better than 90% is determined, with a pion fake rate of less than 5% for particles with momenta more than 1.5 GeV/$c$ and a likelihood cut of 0.66.

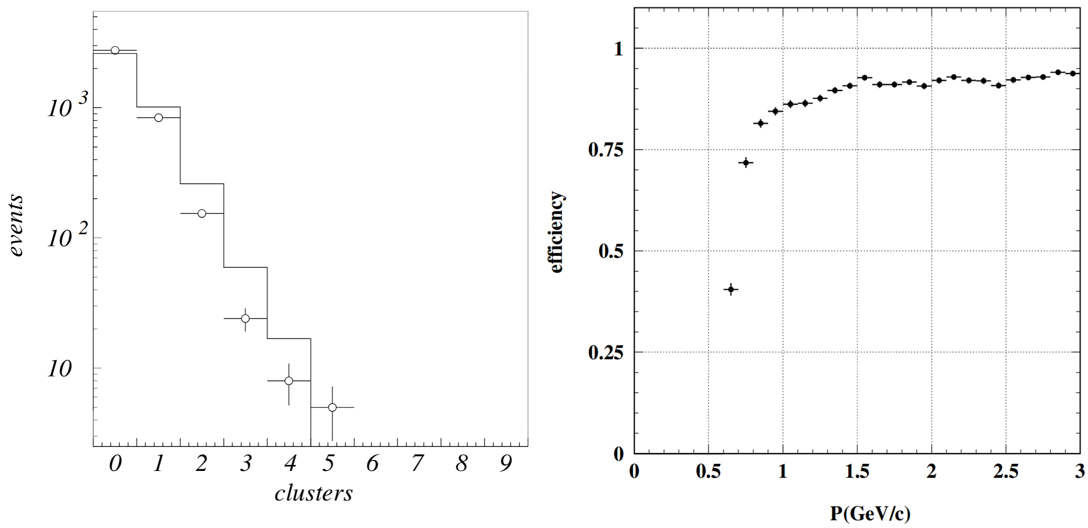



Figure 3.14: Number of neutral clusters per event in KLM (left) and muon detection efficiency as a function of momentum in KLM (right) [17].

Cosmic ray events have been used to determine the efficiency and resolution of the KLM, with an overall efficiency typically over 98%. The temporal and spatial resolutions of the KLM are few ns and about 1.2 cm, respectively. The latter corresponds to an angular resolution from the interaction point of better than 10 mrad.

In order to do detector calibration and proper luminosity measurements, we need to accumulate samples of Bhabha and $\gamma\gamma$ scattering. Otherwise, as shown in Table 3.1, the cross-section for physics events of interest is reasonably small. During normal operation (luminosity of $L = 10 \times 10^{34}$ cm$^{-2}$s$^{-1}$) the total event rate is around 200 Hz, which is well below the data acquisition (DAQ) limit of 500 Hz. Out of this rate, 100 Hz are physically interesting events, which include also two-photon events, Bhabha scattering, and $\mu$ pair production, besides hadronic events from $B\bar{B}$ pair events. In order to discard events which are not interesting for physics analyses, we use a trigger system by appropriately applying restrictive conditions. This section describes the necessary procedures and equipment to successfully do so.

### 3.2.7 Trigger System

The trigger system operates by immediately eliminating events that are not of interest, so that the amount of stored data is within the 500 Hz frequency limit, while the efficiency for physics events of interest is kept high. Events which pass the triggers are then stored, otherwise discarded. The Belle trigger system consists of three stages, Level-1 (L1) online hardware trigger, Level-3 (L3) online software trigger and Level-4 (L4) offline software trigger.

L1 trigger is the first stage of the trigger system, which consists of multiple sub-detector triggers, all connected to a central trigger system called the Global Decision Logic (GDL), as schematically shown in Figure 3.15. Each sub-detector trigger works on a principle of either a track trigger or an energy trigger. In the former case, the triggers discard events not meeting conditions based on the number of reconstructed tracks or track hits, while the latter is based on the total energy deposit and counting of crystal hits. Each sub-detector processes the event information and provides it to the GDL, where all the information is combined and the current event is characterized. The information from the sub-detector triggers reaches the GLD within 1.85 $\mu$s after the collision, and the final trigger signal is provided within at a fixed 2.2 $\mu$s latency. The combined efficiency from the L1 trigger is greater than 99.5% for hadronic events.

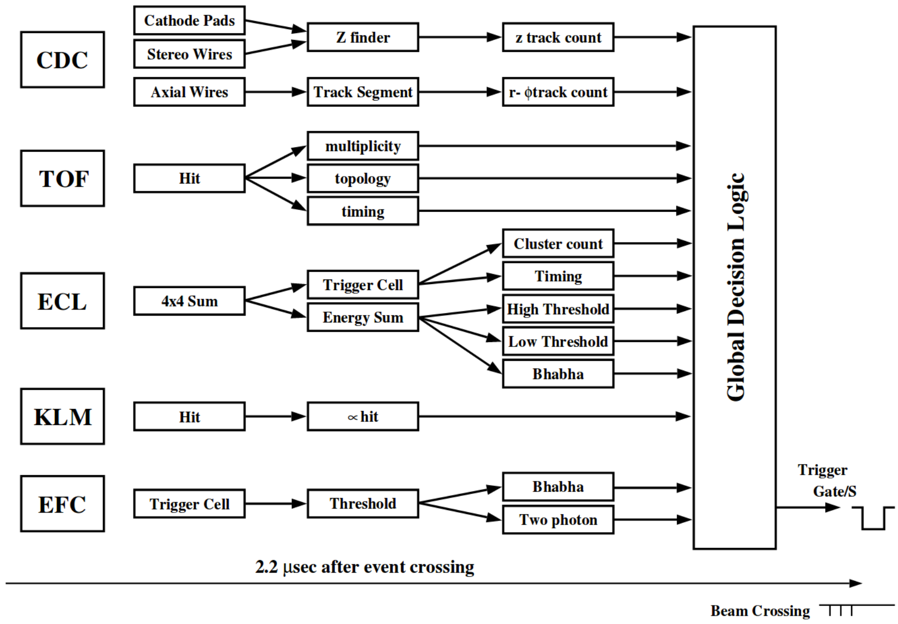



Figure 3.15: The Level-1 trigger system for the Belle detector [17].

After passing L1 trigger, the L3 discards background events from the software-wise perspective. L3 is an online software trigger which performs a simple, but fast reconstruction of the event. Events with at least one track satisfying the impact parameter condition $|\mathrm{d}z| < 5.0$ cm and with a total energy deposit in the ECL more than 1 GeV/$c$ are selected. The L3 trigger reduces the event rate by 50%, with a 99% efficiency for hadronic events.

After passing the L3 trigger, the events are recorded on tapes. However, these data still contain many events from the beam background. To reduce the background events even further, they are required to pass the L4 offline software filtering. At the same time, high efficiencies for signal events is still required. Events must satisfy the following conditions

- have at least one track with $p_T > 300$ MeV/$c$ and impact parameters $|\mathrm{d}r| < 1.0$ cm and $|\mathrm{d}z| < 4.0$ cm,
- have total energy deposit in the ECL must greater than 4 GeV.

Approximately 27% of triggered events are passed through L4 while keeping an almost full efficiency for hadronic events. Events that pass the L4 trigger are fully reconstructed and stored to the DST. Overall, the efficiency of hadronic events after all trigger stages is measured to be more than 99%, which is more than the requirements from physics analyses.

# Chapter 4

# B2BII Conversion

The predecessor of the Belle II experiment was the Belle experiment, which finished its data-taking run of 10 years at end of 2010 after collecting a dataset of about 1 $\mathrm{ab}^{-1}$. That year the Belle detector was shut down and the Belle II experiment was born from the ashes, where even some of the old detector components were reused. This moved focus from Belle analyses and Belle Analysis Framework (BASF) to the construction of the Belle II detector and the development of Belle II Analysis Framework (BASF2), which was written completely from scratch, making the BASF2 software incompatible with Belle data. This resulted in gradual loss of knowledge on the maintenance and operation of the BASF software. The construction of the Belle II detector today is still an ongoing process, although first collisions were already recorded in April 2018. By the year 2025, it is foreseen that Belle II will have recorded about 50 $\mathrm{ab}^{-1}$ of data, which is about 50 times more than in the case of Belle.

However, this is still in the distant future and in principle, we need to wait for data in order to start doing analyses. On the other hand, even though the Belle experiment finished collecting data, the data itself is still relevant and has the potential for interesting physics analyses today. In the Belle II Collaboration, a task force was created in order to convert Belle data into Belle II format (`B2BII`). The `B2BII` package was developed as a part of BASF2 in order to convert data and MC of the Belle experiment and make it available within BASF2. In addition to the convenience of Belle data being processed in the more intuitive and advanced BASF2 framework, `B2BII` allows for estimation and validation of performances of various advanced algorithms being developed for Belle II. The conversion itself, however, is considered non-trivial. Although the conversion of the raw detector data would be possible, the reconstruction algorithms of BASF2 are optimized for Belle II and cannot be effectively applied to Belle data. To bypass this problem, reconstructed objects from `PANTHER` tables, a custom solution of the Belle collaboration based on C/C++ and Fortran, are mapped to their corresponding representations in BASF2. In this analysis, we use the developed converter package in order to analyze Belle data with the Belle II software.

The conversion in the `B2BII` package is divided into three BASF2 modules. The first module opens the Belle input files and reads the events into memory in the form of `PANTHER` tables. This module consists predominantly of reused BASF code. The second module applies various calibration factors, such as experiment and run dependent factors, to the beam energy, particle identification information, error matrices of the fitted tracks, etc. The module also applies some low-level cuts to reproduce removing background events as done within BASF. The actual conversion and the mapping of reconstructed objects are done in the last module. For more information see [3].

## 4.1 Validation

In order to make sure the conversion was successful and without errors, a thorough validation should be performed. This is done by comparing histograms of all physical quantities of the reconstructed objects on simulated and recorded events, processed with BASF and BASF2.

Our signal decay mode consists of three charged tracks, track conversion should perform flawlessly. Additionally, energy measurement is also very important in our untagged analysis in order to successfully determine the missing 4-momentum in the event, which is why we also need a correct conversion of the ECL clusters for photons and $\pi^0$ particles. Figures 4.1 to 4.3 show the basic physical properties of converted tracks, photons and $\pi^0$ particles, obtained with BASF and BASF2. The plots indicate that the conversion is successful in all aspects and we can proceed with the analysis in the framework of BASF2.

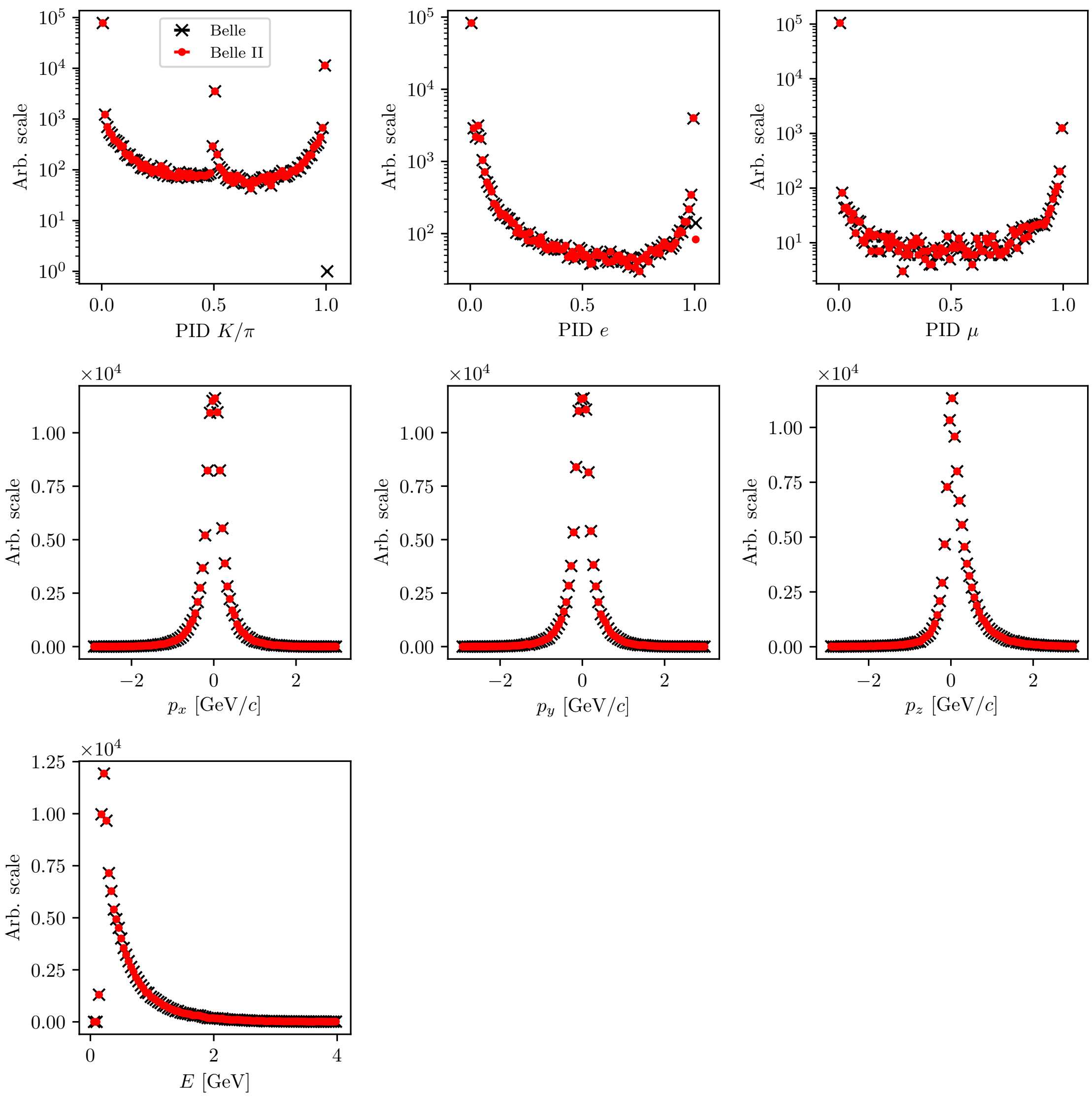



Figure 4.1: Some of the more important physical properties of tracks for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

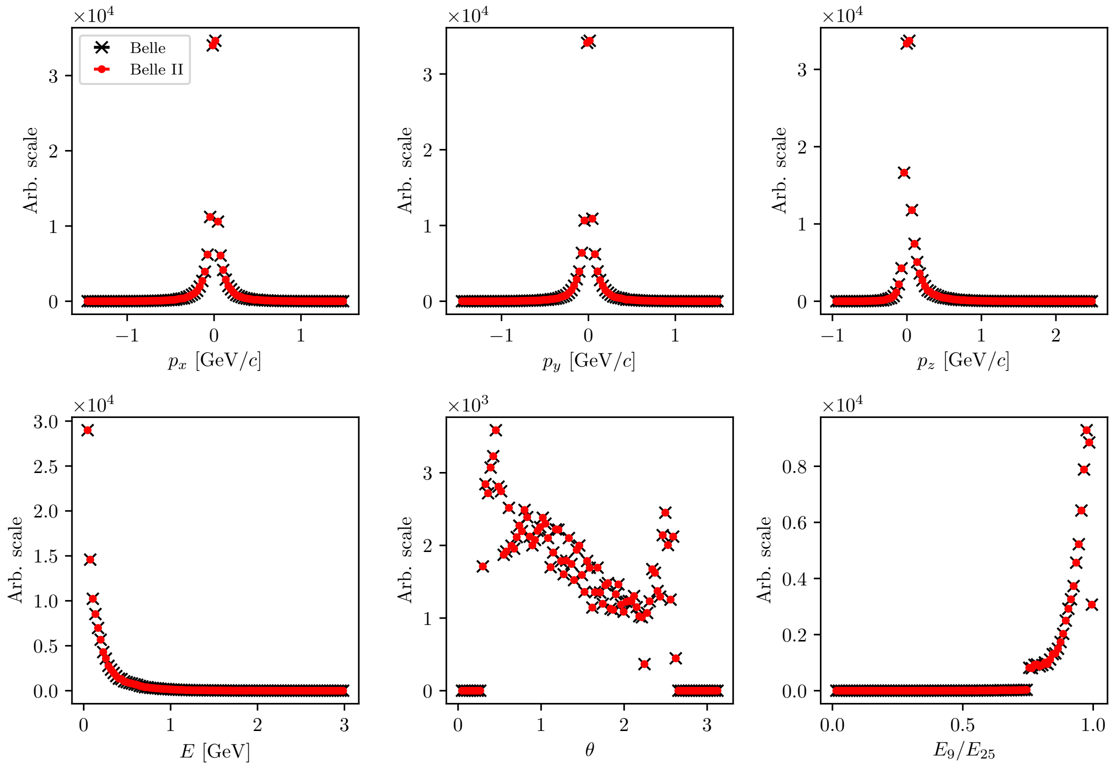



Figure 4.2: Some of the more important physical properties of photons for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

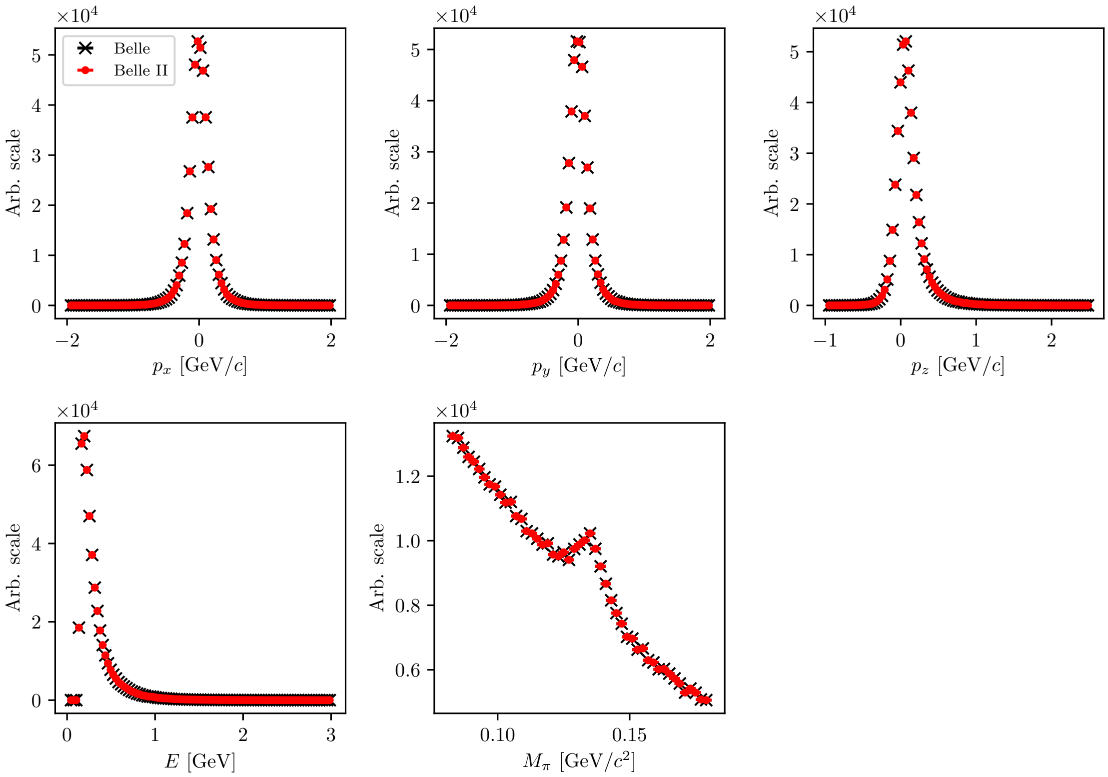



Figure 4.3: Some of the more important physical properties of $\pi^0$ particles for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

# Chapter 5

# Event Reconstruction

In this chapter the procedure for event reconstruction of the $B$ meson decay $B \to KK\ell\nu$ is shown, starting with final state particle selection and then combining them to obtain $B$ meson candidates.

## 5.1 Final State Particles Selection

Since the neutrino escapes detection, we can only reconstruct the charged tracks in the decay, which are the two charged kaons ($K$) and the light lepton, which is the electron ($e$) or muon ($\mu$). These are some of the particles which are commonly referred to as final state particles (FSP). Final state particles have a long lifetime and are usually the particles that we detect when they interact with the material in the detector.

It is important to limit our selection of FSP particles in order to cut down the number of particle combinations, and consequentially computation time and file sizes.

At this point in the analysis, we do not apply any intelligent cuts yet, which results in a large number of available particles and their combinations. In order to minimize this effect, we perform this part of the study on a smaller subset of the available generic MC, experiment no. 23 and 31, which correspond to an integrated luminosity of 6.273 $\text{fb}^{-1}$ and 17.725 $\text{fb}^{-1}$, respectively. We chose these two experiments to get closer to the appropriate ratio of SVD1 and SVD2 data in the full Belle MC.

#### Leptons

Figures 5.1 and 5.2 show the impact parameters $d_0$ and $z_0$, the momentum in $\Upsilon(4S)$ center-of-mass system (CMS), and the PID information for true and fake electrons and muons, where an extra category for true electrons/muons from the signal decay is shown.

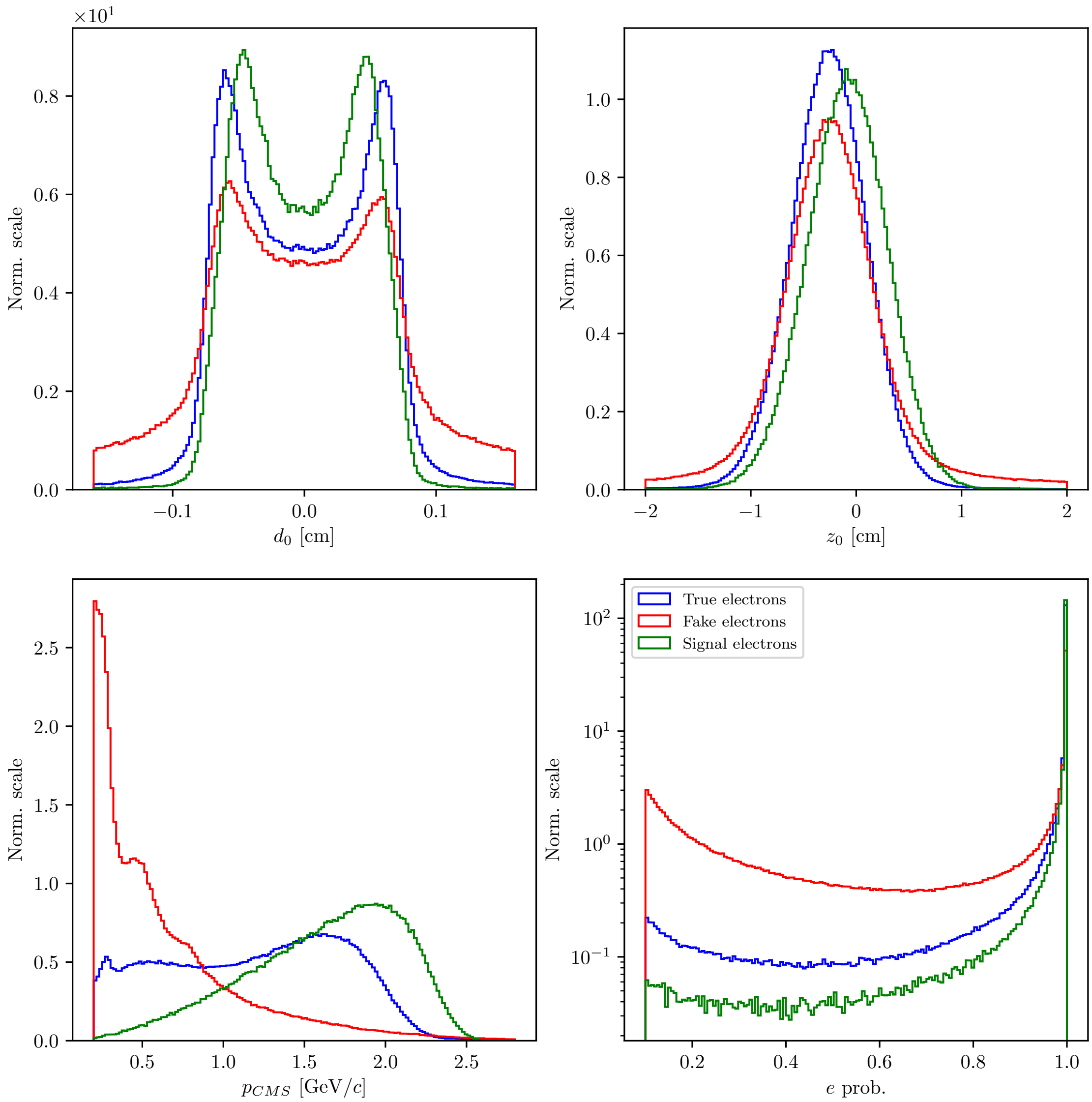



Figure 5.1: Normalized properties of true (blue), fake (red) and true electrons from signal $B$ candidates (green).

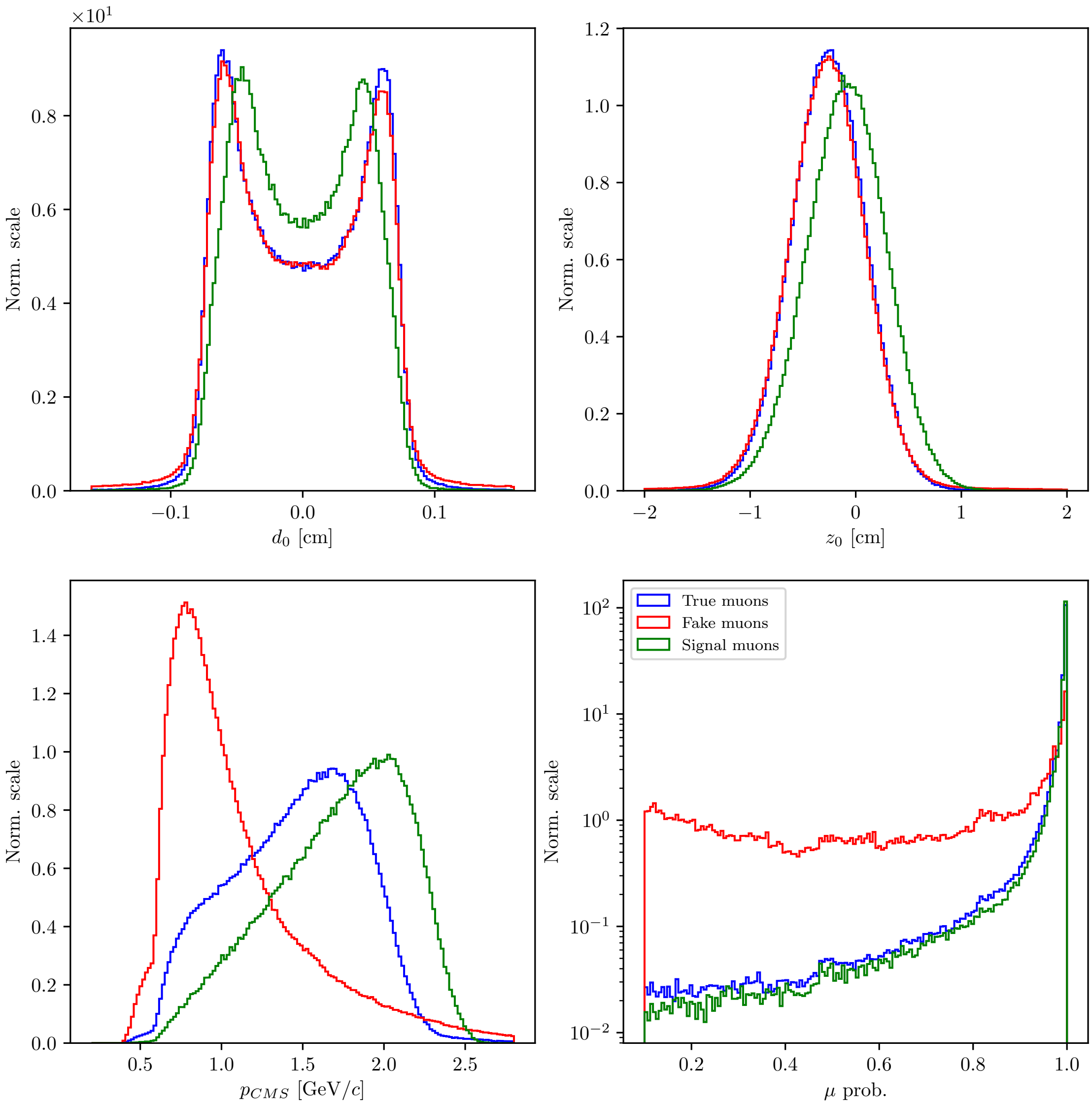



Figure 5.2: Normalized properties of true (blue), fake (red) and true muons from signal $B$ candidates (green).

Based on these distributions, we can define a set of cuts

- $|d_0| < 0.1$ cm,
- $|z_0| < 1.5$ cm,
- $p_{CMS} \in [0.4, 2.6]$ GeV/$c$ for electrons,
- $p_{CMS} \in [0.6, 2.6]$ GeV/$c$ for muons.

After this selection we can determine the optimal PID cuts for electrons and muons, where we optimize the selection by maximizing the standard definition of *figure of merit* (FOM), defined in Eq. (5.1)

$$\text{FOM} = \sqrt{\mathcal{EP}} \propto \frac{S}{\sqrt{S+B}}, \tag{5.1}$$

where the argument in the square root is the product of the efficiency ($\mathcal{E}$) and the purity ($\mathcal{P}$) function. The definitions of signal ($S$) and background ($B$) are somewhat fluid throughout the analysis and need to be defined for each FOM separately. In this section we define two representations of $S$ and $B$. In $\text{FOM}_1$ the signal $S$ represents correctly reconstructed final state particles, while in $\text{FOM}_2$ the signal $S$ represents correctly reconstructed final state particles which also come from a correctly reconstructed $B$ meson candidate. In both cases $B$ represents all other particle candidates which do not satisfy the conditions of $S$.

The FOM plots are shown in Figures 5.3 and 5.4. The cut values are based on PID cuts used for PID efficiency calibration. The optimal value for the PID cuts is equal to the largest available value, regardless of the leptons coming from signal decays or not. The optimized PID cuts for leptons are

- $e$ prob. $> 0.9$ for electrons,
- $\mu$ prob. $> 0.97$ for muons.

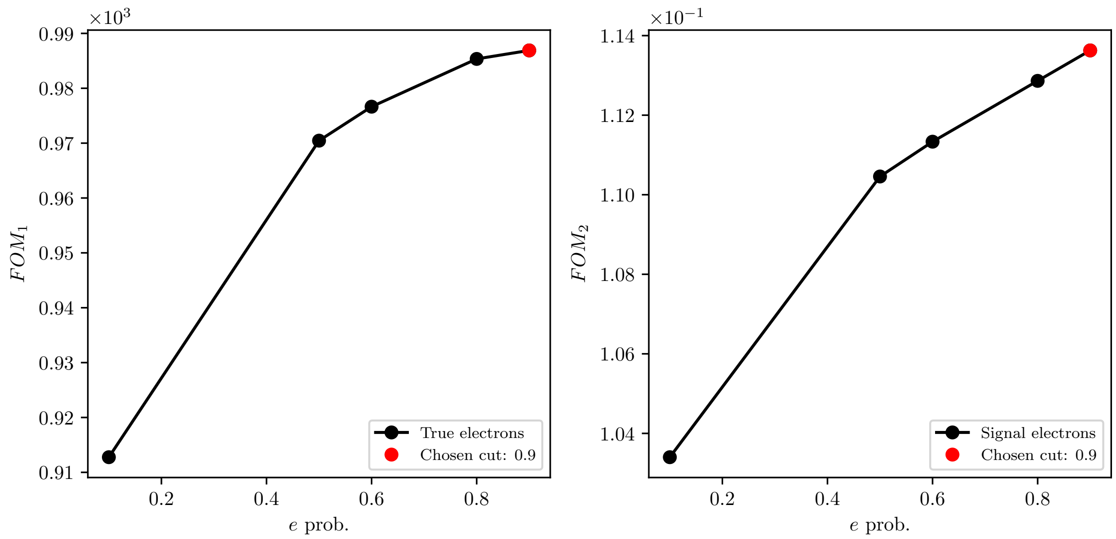



Figure 5.3: FOM optimizations of the PID probability cuts for true electrons (left) and true electrons from signal $B$ candidatess (right).

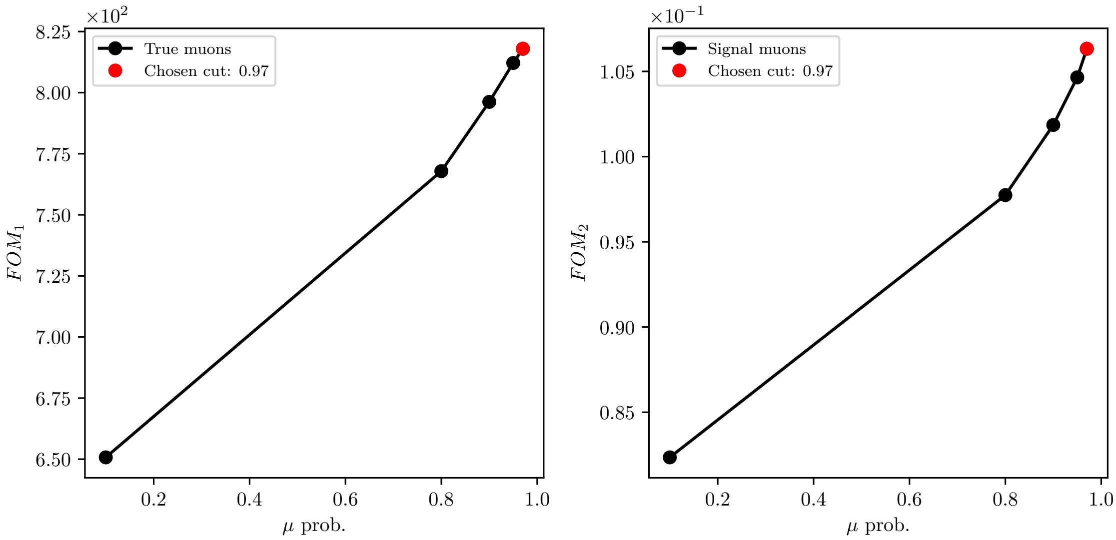



Figure 5.4: FOM optimizations of the PID probability cuts for true muons (left) and true muons from signal $B$ candidates (right).

### Kaons

We repeat the procedure for both kaons. Figure 5.5 shows the impact parameters $d_0$ and $z_0$, the momentum in $\Upsilon(4S)$ center-of-mass system (CMS), and the PID information for true and fake kaons, where an extra category for true kaons from the signal decay is shown.

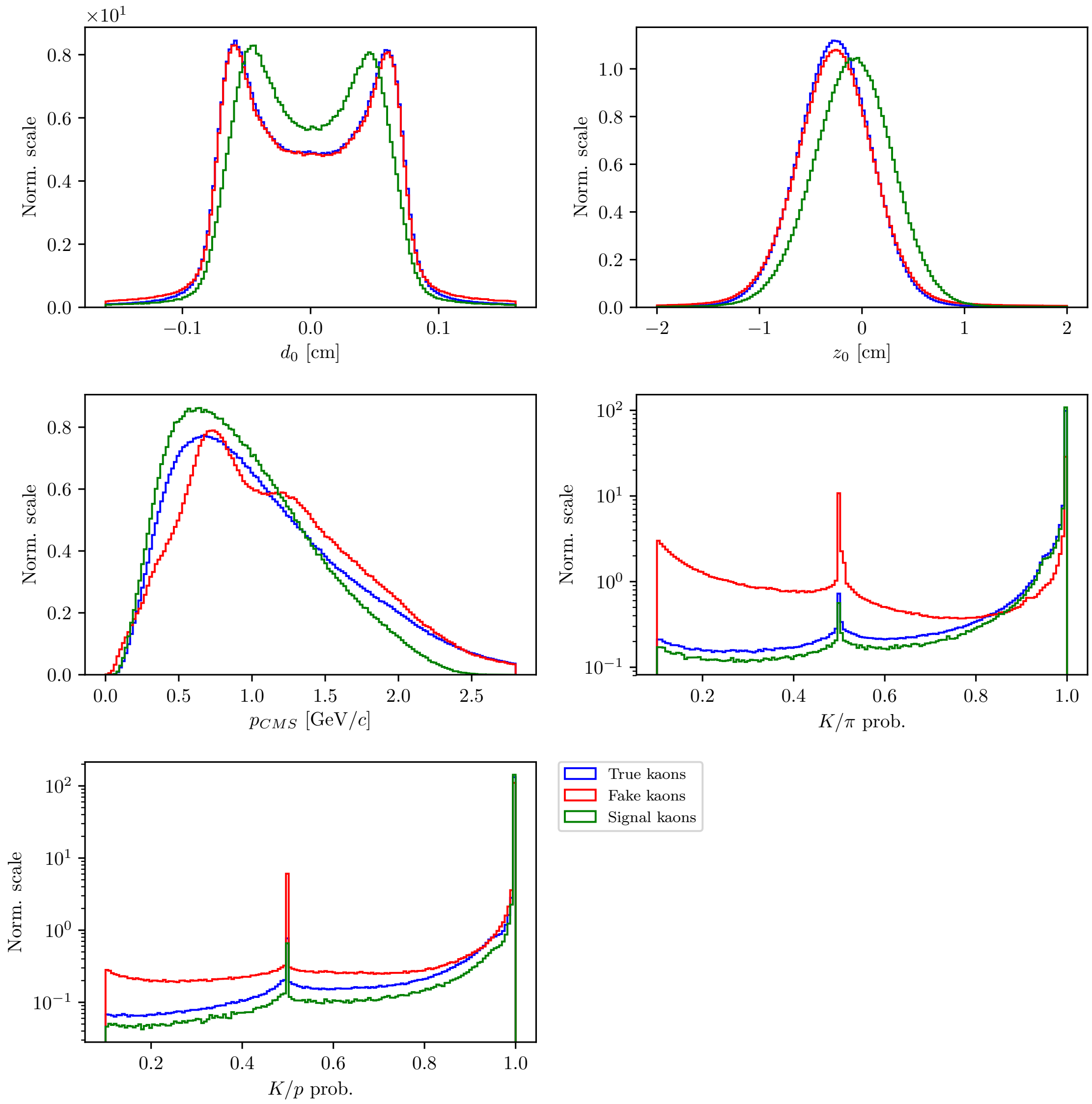



Figure 5.5: Normalized properties of true (blue), fake (red) and true kaons (green) from signal $B$ candidates.

We define the kaon cuts in the same manner as in the case for leptons

- $|d_0| < 0.15$ cm,
- $|z_0| < 1.5$ cm,
- $p_{CMS} \in [0,\, 2.5]$ GeV$/c$.

The PID optimization, in this case, is taken in two steps. First, we optimize the cut on $K/\pi$, and after that the $K/p$ separation probability. Figure 5.6 shows the optimization procedure for PID cuts on kaon candidates.

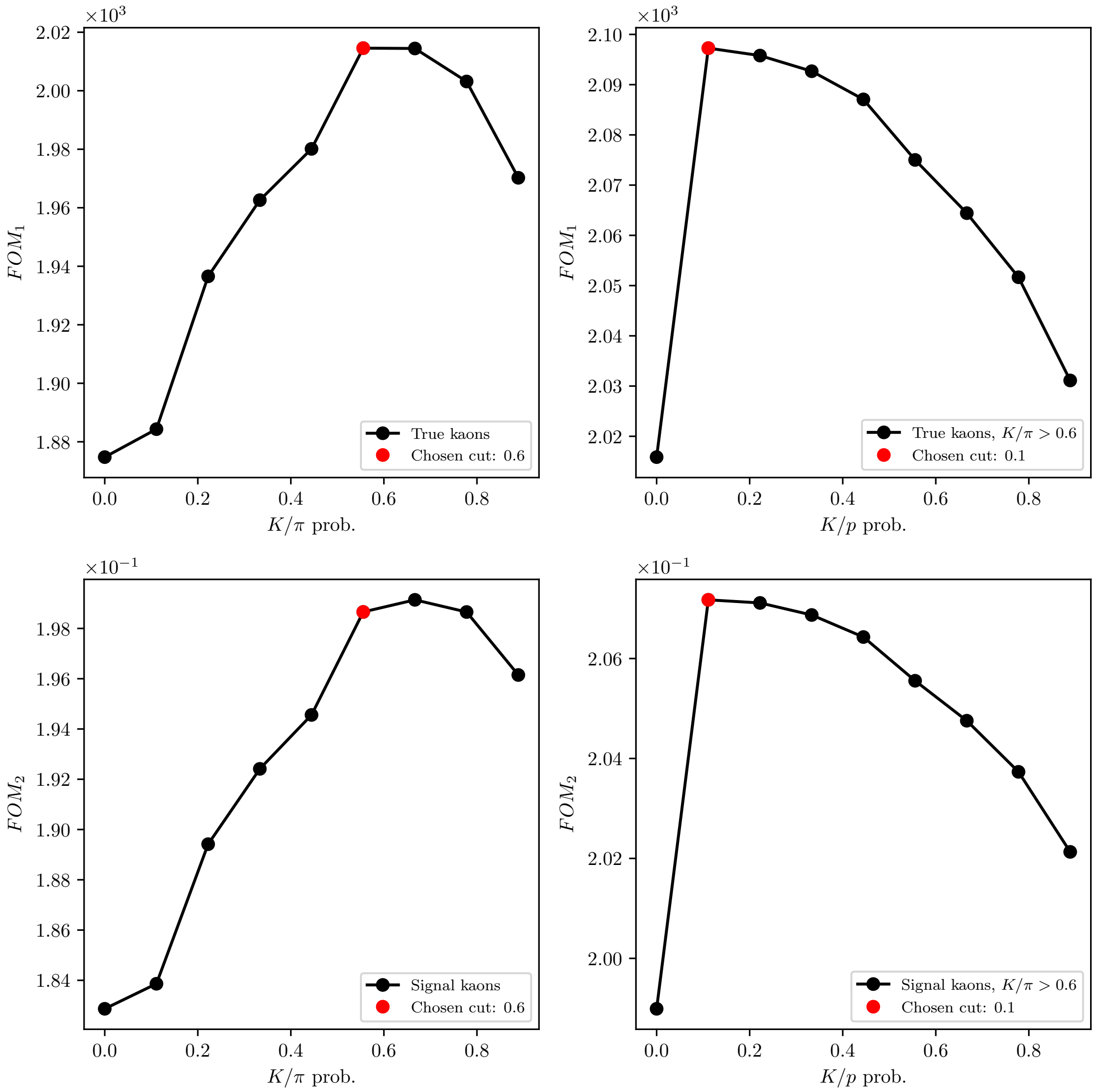



Figure 5.6: FOM optimizations of the PID probability cuts for true kaons (top) and true kaons from signal $B$ candidates (bottom). The plots on the left show the optimization of the first step for the $K/\pi$ probability cut, and the plot on the right the $K/p$ probability cut.

The optimized PID cuts for kaons are

- $K/\pi > 0.6$,
- $K/p > 0.1$.

## 5.2 Pre-selection of First $B$ Meson Candidates

In this section, we use the charged particle candidates from the previous section to make particle combinations, which correspond to $B$ meson candidates. When a $B$ meson candidate is selected, additional features can be calculated and used for background rejection. Since we are still performing this part of the study on a smaller subset of the full available MC sample, we will perform under-optimized cuts based on the FOM optimization in order to optimize them later on the full Belle MC sample.

Since the missing neutrino escapes detection, we reconstruct the $B$ mesons in the following two channels

$$B^+ \to K^+K^-e^+,$$
$$B^+ \to K^+K^-\mu^+,$$

and similarly for $B^-$. When an arbitrary combination is obtained, we perform a vertex fit of the three tracks in order to discard combinations with a low probability of tracks coming from the same point. $B$ mesons have a relatively long lifetime and decay along the $z$ axis of the detector in the direction of the boost, so the vertex fit is enforced with an `IPTUBE` constraint, which constrains the vertex to an elongated ellipsoid along beam direction. We demand that the fit converged and apply a cut on the minimal fit probability. The fit probability for signal and background $B$ meson candidates is shown in Figure 5.7 (left). We perform a FOM cut optimization of this variable, which is shown in Figure 5.7 (right) for the subset of the Belle MC sample. In this and in the following cases, the definition of $S$ from Eq. (5.1) are correctly reconstructed $B$ meson candidates with a missing neutrino which are not coming from a $b \to c$ transition.

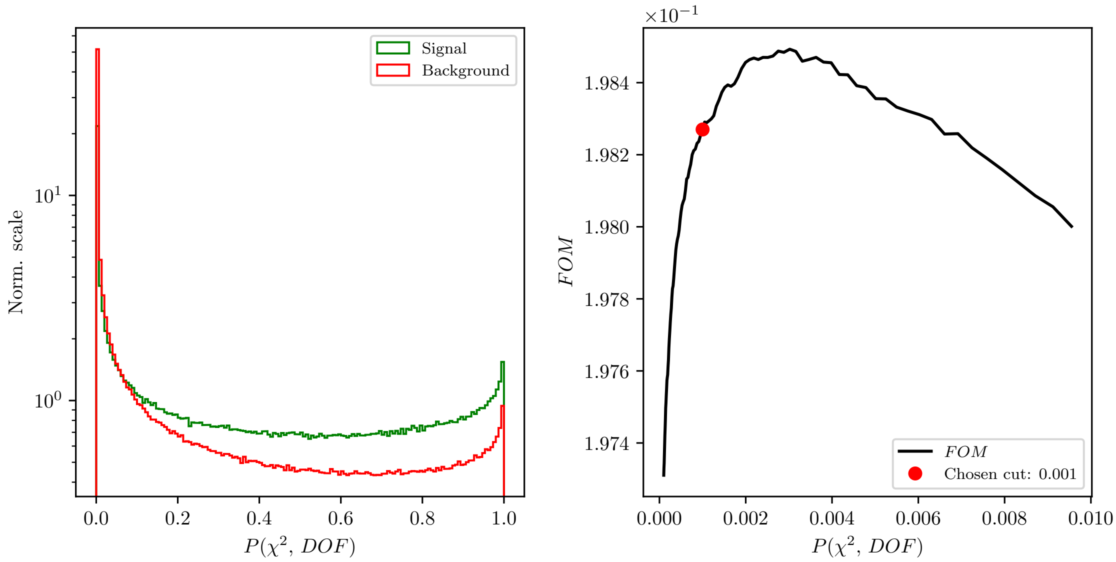



Figure 5.7: Normalized vertex fit probability distribution for signal and background $B$ meson candidates in logarithmic scale(left) and FOM optimization of the vertex fit probability (right) for the subset of the full Belle MC sample.

The chosen pre-cut on the fit probability is

- $P(\chi^2, NDF) > 1.0 \times 10^{-3}$.

With the neutrino being the only missing particle on the reconstructed side, it is possible to determine the angle between the direction of the reconstructed $B$ (denoted as $Y \to KK\ell$) and the nominal $B$, as

$$\mathrm{p}_\nu = \mathrm{p}_B - \mathrm{p}_Y, \tag{5.2}$$
$$\mathrm{p}_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_B E_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \tag{5.3}$$
$$\cos(\theta_{BY}) = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \tag{5.4}$$

where all the energy and momenta above are calculated in the CMS frame. The mass of the neutrino is equal to 0 to a very good precision, so we use it in Eq. (5.3). In addition, we can substitute the unknown energy and momentum magnitude, $E_B$ and $|\vec{p}_B|$, of the $B$ meson in Eq. (5.4), with quantities from the well known initial conditions

$$E_B = E_{CMS}/2, \tag{5.5}$$
$$|\vec{p}_B| = p_B = \sqrt{E_B^2 - m_B^2}, \tag{5.6}$$

where $E_{CMS}$ is the total energy of the $e^+e^-$ collision in the CMS frame and $m_B$ is the nominal mass of the $B$ meson.

For the correctly reconstructed candidates, this variable lies in the $[-1, 1]$ region, though only to a certain precision, due to the finite detector resolution. Background candidates, however, populate also the non-physical regions, as shown in Figure 5.8 (left).

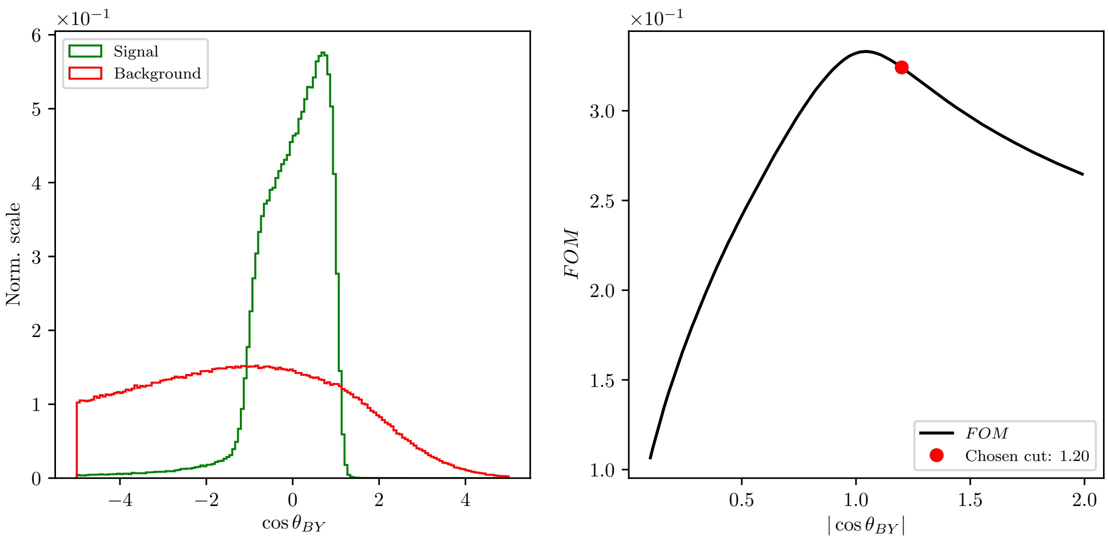



Figure 5.8: Normalized $\cos\theta_{BY}$ distribution for signal and background $B$ meson candidates (left) and FOM optimization of the $\cos\theta_{BY}$ variable (right) for the subset of the full Belle MC.

We again impose an under-optimized cut on this variable from Figure 5.8 (right) to discard a large amount of background on this subset of the full Belle MC

- $|\cos(\theta_{BY})| < 1.20$.

## 5.3 Loose Neutrino Reconstruction

The signal-side neutrino escapes detection, so we cannot directly determine it's four-momentum. However, due to the detectors geometry, which almost completely covers the full solid angle, and due to well known initial conditions of the $\Upsilon(4S)$ meson, it is possible to determine the kinematics of the missing neutrino via indirectly reconstructing the companion $B$ meson by summing up the four-momenta of all the FSP particles in the event which were not used in the reconstruction of the signal side $B$ meson. This is known as the *untagged* method since we are not using any kind of tagging method to reconstruct the companion $B$ meson. The particles used in the indirect companion $B$ meson reconstruction are also said to belong to the *rest of the event* (ROE).

Due to the beam background in the detector, material interactions, or other processes, random tracks and clusters enter our event and get reconstructed as part of the physics process we want to study. These tracks and clusters are not interesting and further spoil the data we measure. In order to remedy this, we perform an extensive clean-up of the tracks and clusters in the ROE side before calculating the four-momentum of the missing part of the event. Here we see the motivation for the ROE clean-up since our signal candidate reconstruction depends on tracks and clusters in the ROE side. The clean-up procedure is performed separately on tracks and clusters and uses multiple steps with multivariate analysis (MVA) algorithms in order to separate good tracks and clusters from the bad ones, which populate the ROE. Then, for each ROE object, a ROE mask is created for tracks and clusters, which narrates the use of this object in the final calculations of the ROE four-momentum. From this point on we assume the ROE to be efficiently cleansed of extra tracks and clusters. A more detailed description of the ROE clean-up can be found in Chapter 6.

The total missing four-momentum in the event can be determined as

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \sum_{i}^{\text{Event}} (E_i,\, \vec{p}_i)\,, \tag{5.7}$$

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \left( \mathrm{p}_Y - \sum_{i}^{\text{Rest of event}} (E_i,\, \vec{p}_i) \right), \tag{5.8}$$

where the summation runs over all charged and neutral particles in the defined set with

$$\mathrm{p}_i^{\text{neutral}} = (p_i,\, \vec{p}_i) \quad \text{and} \quad \mathrm{p}_i^{\text{charged}} = \left( \sqrt{m_i^2 + p_i^2},\, \vec{p}_i \right), \tag{5.9}$$

where we assumed all neutral particles to be massless photons. For charged tracks in the ROE a mass hypothesis needs to be defined in order to determine the track's energy. After the ROE clean-up we make the following procedure of choosing the mass hypothesis

1. $e$, if $e$ prob. $>$ $\mu$ prob. and $e$ prob. $> 0.9$,
2. otherwise $\mu$, if $\mu$ prob. $>$ $e$ prob. and $\mu$ prob. $> 0.97$,
3. otherwise $K$, if $K/\pi$ prob. $> 0.6$,

4. otherwise $\pi$.

We define the square of the missing mass, $m^2_{miss}$, which is consistent with zero, if signal-side neutrino is the only missing particle in the event, as shown in Eq. (5.11).

$$\mathrm{p}_\nu = \mathrm{p}_{miss} = (E_{miss}, \vec{p}_{miss}) , \tag{5.10}$$
$$m^2_{miss} = \mathrm{p}^2_{miss} = \mathrm{p}^2_\nu = m^2_\nu \approx 0. \tag{5.11}$$

Since the detector is not perfect, the distribution of the $m^2_{miss}$ variable has a non-zero width. Additionally, a tail is introduced due to missing particles like neutrinos, other neutral undetected particles such as $K^0_L$, or simply missing tracks due to detection failure. Figure 5.9 shows the distribution of $m^2_{miss}$ as defined with the missing four-momentum in Eq. (5.10). Correctly reconstructed candidates, which come from events where the other $B$ meson decayed via a hadronic decay mode, peak at zero. If this is not the case, candidates are shifted to larger values of this variable, even if the event in question is a signal event. For this purpose, we define a subset of all signal candidates, which come from events where the companion $B$ meson decayed hadronically and all of its particles were taken into account correctly. We only allow for missing photons, since they are frequently irradiated due to bremsstrahlung effects and they do not have such a big impact on the 4-momentum of the final candidate. We denote this subset as *perfect* signal.

Due to this fact, we impose a cut on the $m^2_{miss}$ variable in order to partially discard candidates with spoiled properties, even if it was in principle a correct combination of FSP particles on the signal side. The cut was chosen based on the optimization of FOM, where in this case the definition of $S$ were perfectly reconstructed signal candidates. The chosen cut value is

- $|m^2_{miss}| < 3.0$ GeV$/c^2$,

which is also under-optimized at this point due to the same reasons as in the cases above.

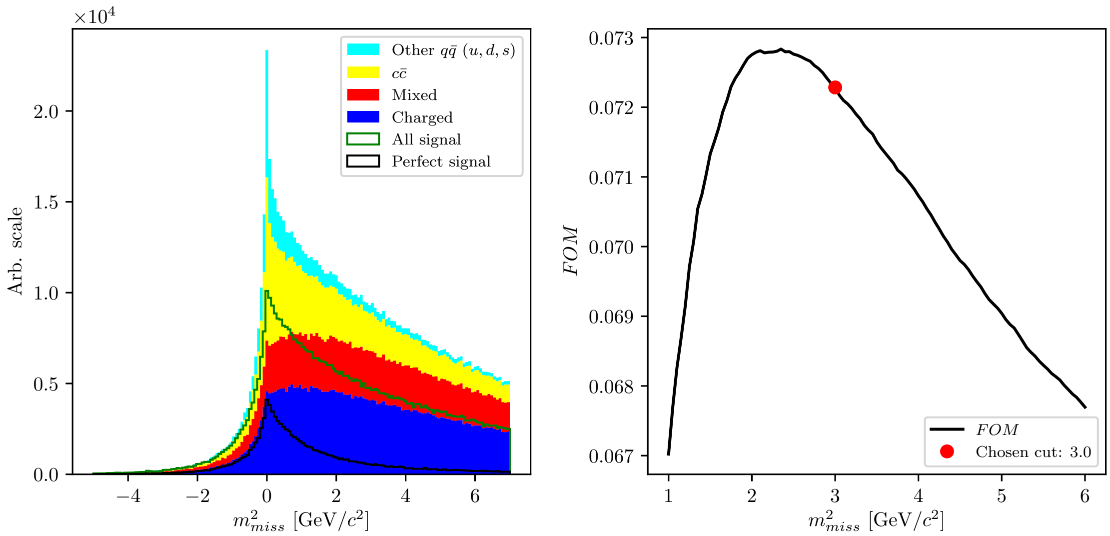



Figure 5.9: Squared missing mass distribution for signal and various types of background. All signal (green) and perfect signal (black) are scaled up equally.

The main uncertainty of the neutrino four-momentum, defined in Eq. (5.10), comes from energy uncertainty. It is a common practice to substitute the missing energy with the magnitude of the missing momentum, since the momentum resolution from the measurement is much better, thus redefining the neutrino four-momentum to

$$\mathrm{p}_\nu = (|\vec{p}_{miss}|, \vec{p}_{miss}) , \tag{5.12}$$

which fixes the neutrino mass to 0 GeV/$c^2$.

The newly defined neutrino four-momentum can be added to the four-momentum of the $Y(KK\ell)$ candidate to obtain the full $B$ meson four-momentum and calculate the traditional $M_{BC}$ and $\Delta E$ variables

$$\Delta E = E_B - E_{CMS}/2, \tag{5.13}$$

$$M_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}|^2}. \tag{5.14}$$

Since the final fit will be performed over $\Delta E$ and $M_{BC}$, we define the fit region

- $M_{BC} \in [5.1,\ 5.295]$ GeV/$c^2$,
- $\Delta E \in [-1.0,\ 1.3]$ GeV.

Figure 5.10 shows the distributions of $\Delta E$ (left) and $M_{BC}$ (right) for signal and major types of background after the pre-cuts. Both signal components are scaled up with respect to the background components but are in proper scale one to another. The effects of missing particles are clearly seen based on the shape difference between all and perfect signal.

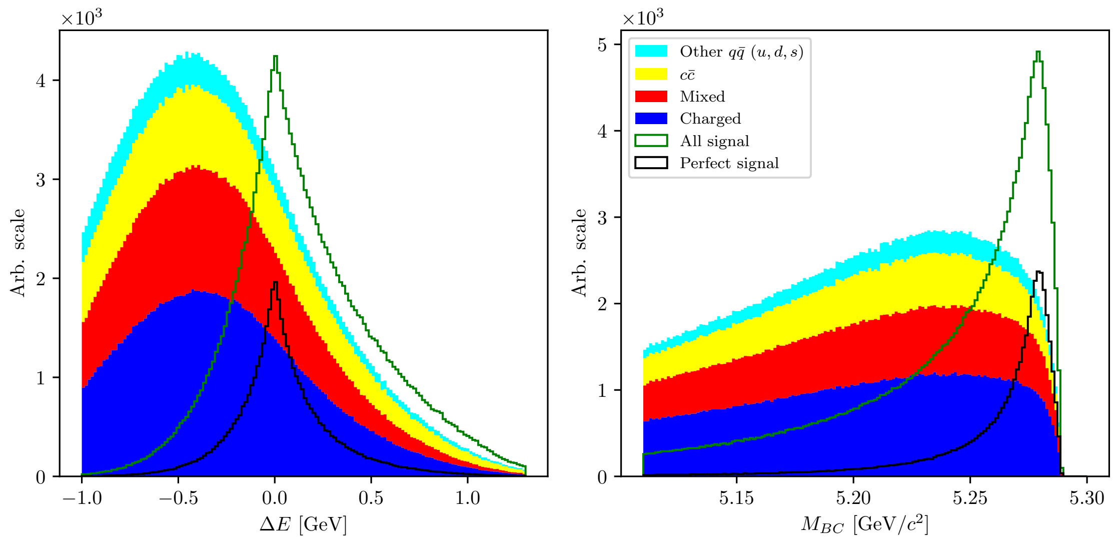



Figure 5.10: Distributions of $\Delta E$ (left) and $M_{BC}$ (right) for signal and major types of background after the precuts. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution in both distributions, since the event is perfectly reconstructed.

## 5.4 Final Stage Optimization

With the charge particle selection and a rough selection of the $B$ meson particles in place, we can now afford to run the reconstruction procedure over all 10 streams of the full available Belle generic MC. After obtaining the full reconstructed sample, the first task is to optimize the under-optimized cuts from the pre-selection stage. Repeating the procedure on the full sample results in the FOM shapes shown in Figure 5.11, while the optimal cut values are

- $P(\chi^2, NDF) > 6.0 \times 10^{-3}$,
- $|\cos{(\theta_{BY})}| < 1.05$.

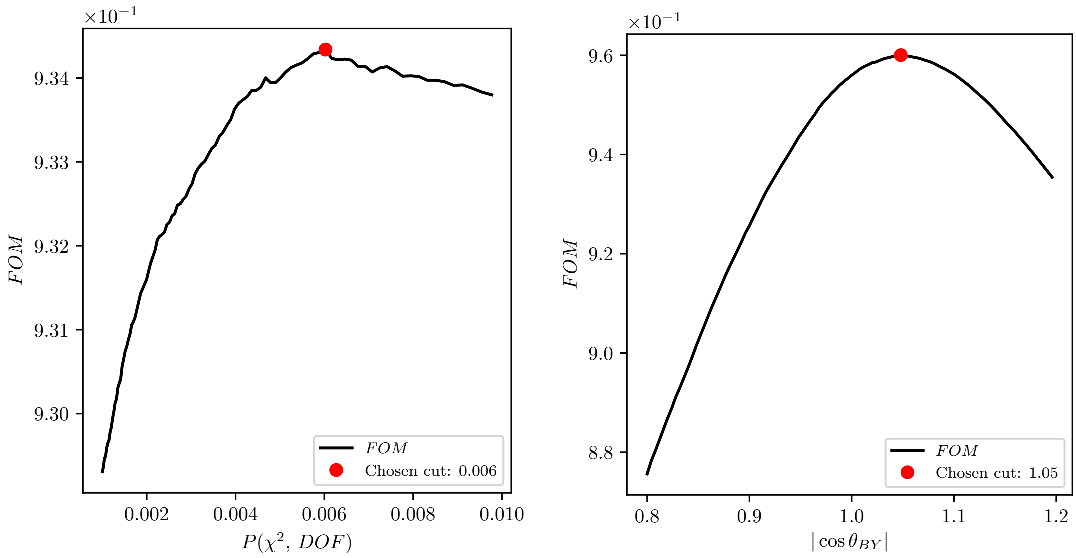



Figure 5.11: FOM optimization of the vertex fit probability (left) and the $\cos\theta_{BY}$ variable (right) for the full Belle MC sample.

With further optimizations, we will be fine-tuning the signal to background ratio. The most prominent and distinguishable part of our signal is the perfectly reconstructed signal. For this purpose, we define a signal region in $\Delta E$ and $M_{BC}$, where most of our perfectly reconstructed candidates lie. We use this region for all of the following optimization steps in this chapter and also in the background suppression chapter 7 since this allows us to better improve the signal to background ratio where it counts most. The 2D FOM optimization of the optimal $\Delta E$ and $M_{BC}$ signal region is shown in Figure 5.12.

The signal region is defined as

- $M_{BC} > 5.271$ GeV$/c^2$.,
- $|\Delta E| < 0.126$ Gev.

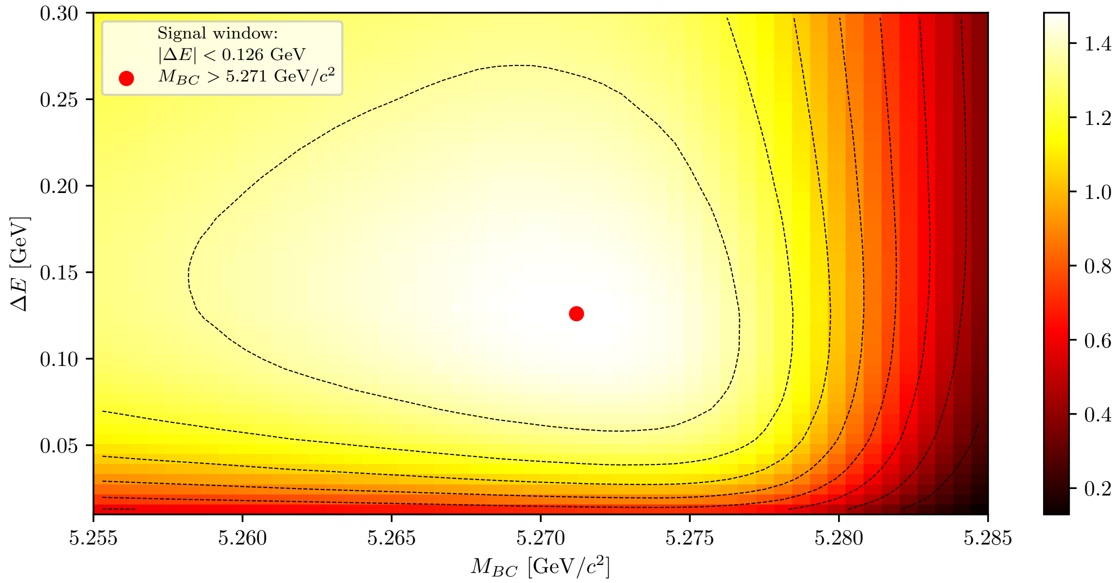



Figure 5.12: 2D FOM optimization of the signal region definition, where the signal in the optimization was represented by perfectly reconstructed candidates.

With the signal window defined, we can tighten the cut on $m^2_{miss}$, which we intentionally left loose before the signal categorization. With the FOM optimization of perfectly reconstructed candidates inside the signal region, shown in Figure 5.13, the optimal cut on $m^2_{miss}$ is

- $|m^2_{miss}| < 0.975$ GeV/$c^2$.

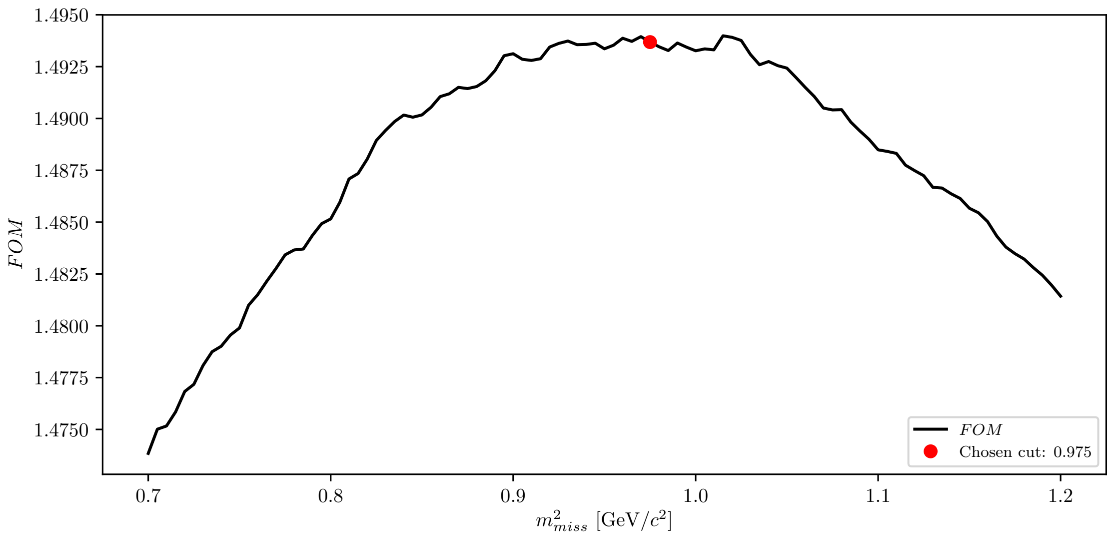



Figure 5.13: FOM optimization of the optimal $m^2_{miss}$ cut in the signal region.

## 5.5 Charge product categorization

The missing information due to an escaping neutrino in our reconstructed channel is replaced by information from the companion $B$ meson. Since this is an untagged reconstruction, the quality of the companion $B$ meson affects the properties of the signal candidate. Perfect reconstruction of a hadronically decayed companion $B$ meson results in pronounced peaks at $\Delta E \approx 0$, $m^2_{miss} \approx 0$ and $M_{BC} \approx m_B$, while imperfect reconstruction due to any kind of missing particles produces tails, shift or simply a worse resolution of the mentioned distributions. These effects are undesired since they make it harder to separate signal from background.

To remedy this, we look at the charge product of the reconstructed $B$ meson and the ROE object. For correctly reconstructed events, this should have a value of

$$q_{B^\pm} q_{B^\mp} = -1, \tag{5.15}$$

however, this value is distributed due to missing charged particles in the reconstruction. Figure 5.14 shows various signal distributions of $\Delta E$ and $M_{BC}$ in arbitrary (left) and normalized (right) scales, with the relative ratios of 67.86 % and 32.14 % for correct and wrong values of the charge product, respectively. We see that correctly reconstructed events represent the majority of signal and also have the best resolution in $\Delta E$ and $M_{BC}$, so we proceed with the analysis by imposing the cut in Eq. 5.15.

While this cut introduces a drop in the signal efficiency of about 32.14 %, it improves the resolution of our signal $\Delta E$ and $M_{BC}$ distributions and also the signal to background ratio, where the latter changes from $0.95 \times 10^{-3}$ to $1.09 \times 10^{-3}$.

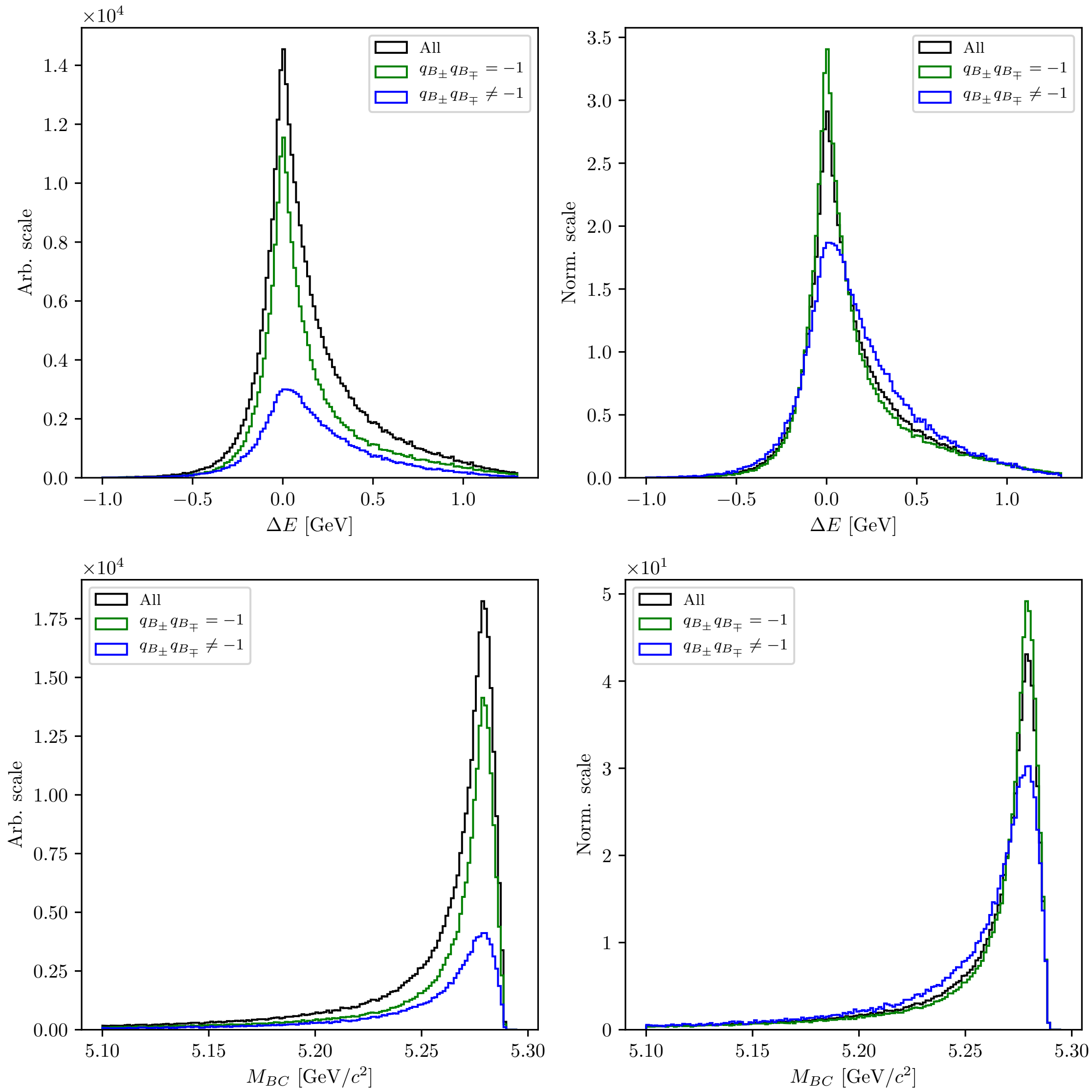



Figure 5.14: Signal distributions of $\Delta E$ and $M_{BC}$ based on the charge product of both $B$ mesons in the event. The plot on the left shows the distributions in an arbitrary scales, while the plot on the right shows the normalized distributions.

## 5.6 Selection Summary

In this section, one can find the full summary of all final selection cuts in the event reconstruction, from FSP particles up to the $B$ meson.

- FSP particles:
  - electrons: $|d_0| < 0.1$ cm, $|z_0| < 1,5$ cm, $p > 0.6$ GeV/$c$, $p_{CMS} \in [0.4, 2.6]$ GeV/$c$, $eID > 0.9$,
  - muons: $|d_0| < 0.1$ cm, $|z_0| < 1,5$ cm, $p_{CMS} \in [0.6, 2.6]$ GeV/$c$, $\mu ID > 0.97$,
  - kaons: $|d_0| < 0.15$ cm, $|z_0| < 1,5$ cm, $p_{CMS} < 2.5$ GeV/$c$, $K/\pi\ ID > 0.6$, $K/p\ ID > 0.1$,

- $B$ meson candidates:
  - standard cuts: $P(\chi^2, DOF) > 6 \times 10^{-3}$, $|\cos\theta_{BY}| < 1.05$, $|m^2_{miss}| < 0.975$ GeV/$c^2$,
  - fit region cuts: $\Delta E \in [-1.0, 1.3]$ GeV, $M_{BC} \in [5.1, 5.295]$ GeV/$c^2$,
  - signal region cuts: $|\Delta E| < 0.126$ GeV, $M_{BC} > 5.271$ GeV/$c^2$,
  - charge categorization: $q_{B^\pm} q_{B^\mp} = -1$.

# Chapter 6

# Rest of Event Clean-up

Continuing from section 5.3, the description of the ROE clean-up process is described here.

Training the MVA classifiers follows the same recipe for all the steps in this chapter. For each step, we run $B$ meson reconstruction on Signal MC with a generic companion $B$ meson. This way the produced weight files are less likely to be signal-side dependent and can be used also for untagged analyses of other decays. For every correctly reconstructed signal $B$ meson we save the necessary information for each MVA step (i.e. properties of ROE clusters). Only correctly reconstructed $B$ candidates are chosen here, to prevent leaks of information from the signal side to the ROE side.

## 6.1 Clusters Clean-up

Photons originate from the IP region, travel to the ECL part of the detector in a straight line and produce a cluster. The direction of the photon is determined via the location of the cluster hit in the ECL and the energy of the photon is directly measured via the deposited energy. This way the four-momentum of photons is determined and used in Eq. (5.8).

Most of the photons in events with $B$ mesons come from $\pi^0 \to \gamma\gamma$ decays. However, a lot of hits in the ECL are also created by photons coming from the beam-induced background or secondary interactions with the detector material. Photons of the first kind should be taken into account when calculating the missing 4-momentum, while the latter kind adds extra energy and momentum which spoils our measured quantities. In the first step of the clusters clean-up, we train an MVA which recognizes good $\pi^0$ candidates and apply this information to the daughter photons. This represents a sort of a $\pi^0$ origin probability, which peaks at or is equal to 0 for photons not coming from $\pi^0$ particles, and peaks at 1 otherwise. This information is used as an additional classifier variable in the next step of the clean-up, where we train to recognize good photons in an event.

### 6.1.1 $\pi^0$ MVA Training

The training dataset of $\pi^0$ candidates contains

- 183255 target candidates,

- 200000 background candidates,

where the definition of a target is that both photon daughters that were used in the reconstruction of the $\pi^0$ are actual photons and real daughters of the $\pi^0$ particle. We use $\pi^0$ candidates from the converted Belle particle list and select those with invariant mass in the range of $M \in [0.10,\ 0.16]$ GeV. After that we perform a mass-constrained fit on all candidates, keeping only the ones for which the fit converged.

The input variables used in this MVA are

- $p$ and $p_{CMS}$ of $\pi^0$ and $\gamma$ daughters,
- fit prob. of the mass-constrained fit, invariant mass and significance of mass before and after the fit,
- angle between the photon daughters in the CMS frame,
- cluster quantities for each daughter photon
  - $E_9/E_{25}$,
  - theta angle,
  - number of hit cells in the ECL,
  - highest energy in cell,
  - energy error,
  - distance to closest track at ECL radius.

The classifier output variable is shown in Figure 6.1.

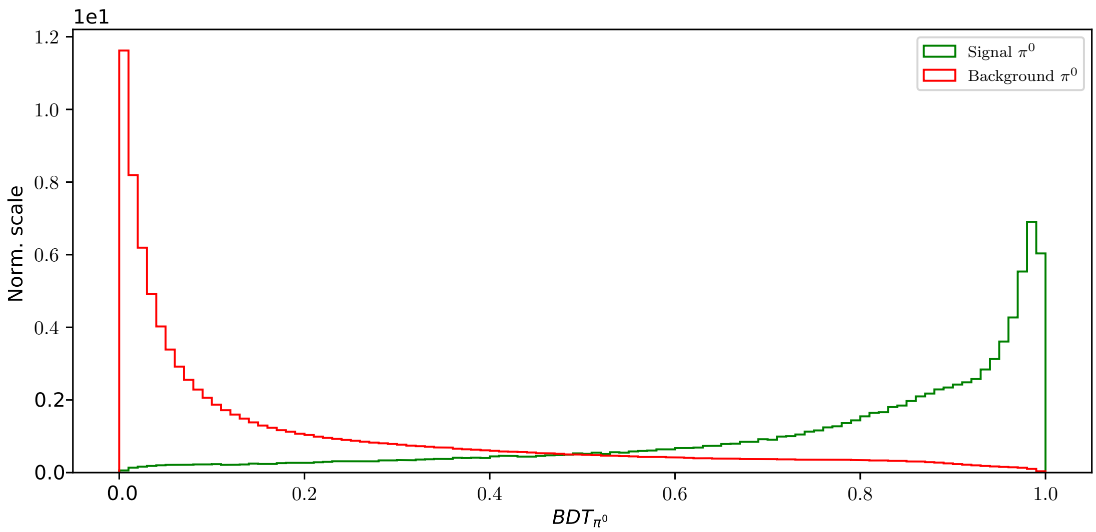



Figure 6.1: Classifier output of the $\pi^0$ training for signal and background $\pi^0$ candidates.

The distributions for all input variables and their correlations for signal and background candidates can be found in Appendix A for all steps of the ROE clean-up.

### 6.1.2 $\gamma$ MVA Training

In this MVA training, we take the $\pi^0$ classifier output of the previous training as an input in order to train a classifier to distinguish between good and bad photons. The $\pi^0$ probability information from the previous step is applied to all photon pairs which pass the same $\pi^0$ cuts as defined in the previous step. Since it's possible to have overlapping pairs of photons, the $\pi^0$ probability is overwritten in the case of a larger value, since this points to a greater probability of a correct photon combination. On the other hand, some photon candidates fail to pass the $\pi^0$ selection, these candidates have a fixed value of $\pi^0$ probability equal to zero.

The training dataset of $\gamma$ candidates contains

- 171699 target candidates,
- 177773 background candidates,

where the definition of a target is that the photon is an actual photon which is related to a primary MC particle. This tags all photon particles from secondary interactions as background photons. We use the converted $\gamma$ candidates from the existing Belle particle list.

The input variables used in this MVA are

- $p$ and $p_{CMS}$ of $\gamma$ candidates,
- $\pi^0$ probability,
- cluster quantities
  - $E_9/E25$,
  - theta angle,
  - number of hit cells in the ECL,
  - highest energy in cell,
  - energy error,
  - distance to closest track at ECL radius.

The classifier output variable is shown in Figure 6.2.

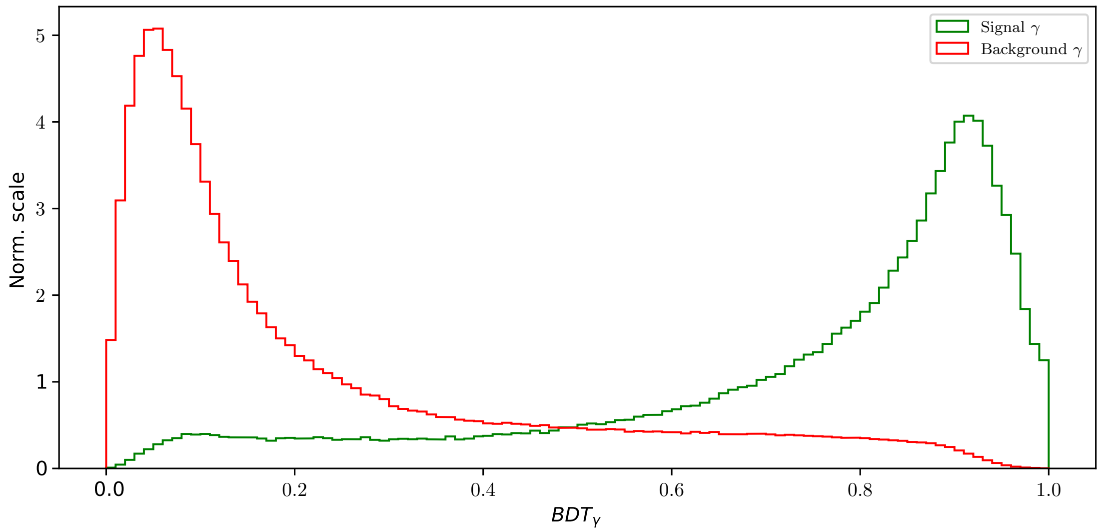



Figure 6.2: Classifier output of the $\gamma$ training for signal and background $\gamma$ candidates.

With the final weights for photon classification in hand, we apply them to the photon particle list. The cut optimization is shown in Figure 6.3 (left), with the optimal cut on the $\gamma$ classifier output at

- $BDT_\gamma > 0.5045$.

Figure 6.3 (right) shows the LAB frame momentum of the photons before and after the cut in logarithmic scale. The signal efficiency and background rejection at this clean-up cut are

- Signal efficiency: $\epsilon_{SIG} = 83.2$ %,
- Background rejection: $1 - \epsilon_{BKG} = 81.2$ %.

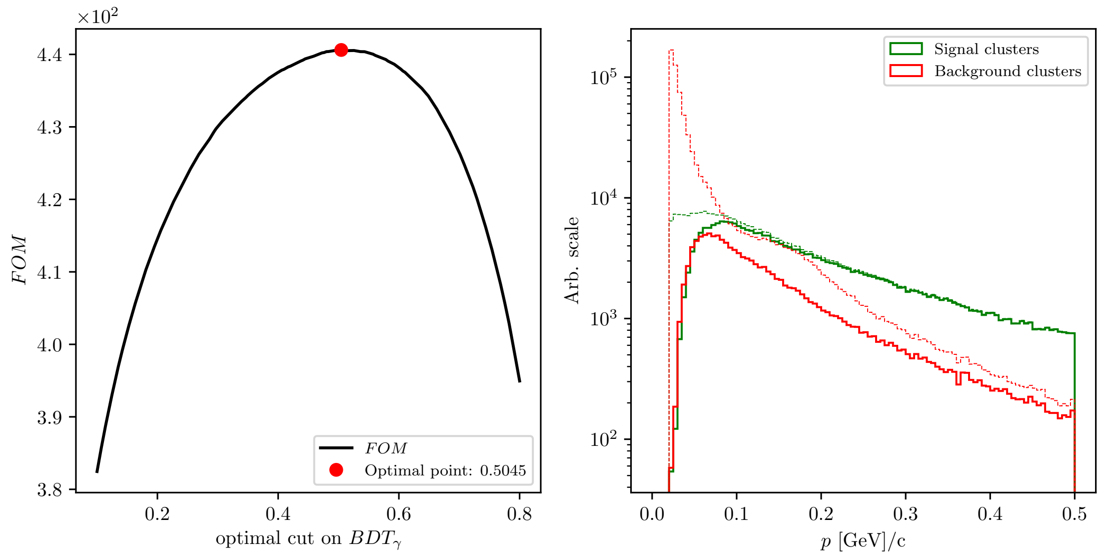



Figure 6.3: The FOM of the classifier output optimization (left) and momentum magnitude in the LAB frame of signal and background photon candidates before and after the optimal cut (right).

The event is now considered to be clean of extra clusters.

## 6.2 Tracks Clean-up

Charged particles leave hits in the detector, which are then grouped into tracks by advanced tracking algorithms. The track is fitted and the track momentum is determined. With the help of particle identification information (PID), we are able to make an intelligent decision about the mass hypothesis of the particle and thus reconstruct the charged particle's four-momentum, which is then added in the loop in Eq. (5.8).

Most of the quality (good) tracks, which come from physics event of interest, come from the IP region, where the collisions occur. Cleaning up the tracks is a more complex procedure than cleaning up the clusters. The following facts need to be taken into account

(a) good tracks can also originate away from the IP region, due to decays of long-lived particles, such as $K_S^0 \to \pi^+\pi^-$,

(b) charged particles from background sources produce extra tracks or duplicates,

(c) low momentum charged particles can curl in the magnetic field and produce multiple tracks,

(d) secondary interactions with detector material or decays of particles in flight can produce "kinks" in the flight directory, resulting in multiple track fit results per track.

Schematics of all the cases mentioned above are shown in Figure 6.4.

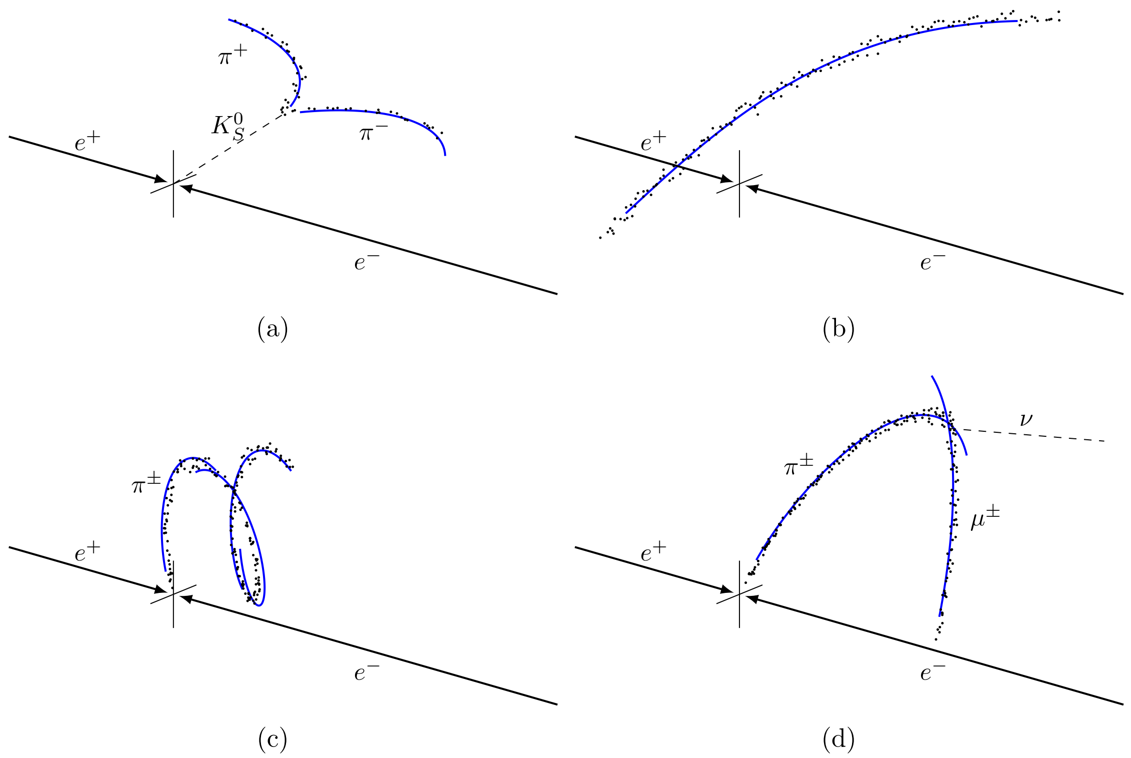



Figure 6.4: (a) Tracks from long-lived neutral particles, which decay away from the IP region, (b) Random tracks from background which are reconstructed, (c) Low-momentum particles which curl in the magnetic field, (d) in-flight decays of particles, which produce a kink in the trajectory.

It is obvious that tracks from the same momentum source should only be taken into account once, or, in case of background tracks, not at all. Such tracks will from this point on be denoted as *extra* tracks, because they add extra four-momentum to our final calculations in Eq. (5.8). At the same time, we have to take care that we don't identify *good* tracks as *extra* tracks. Both of these cases have negative impacts on the final resolution of all variables which depend on information from ROE.

### 6.2.1 Tracks from Long-lived Particles

The first step in tracks clean-up is taking care of tracks from long-lived particles. Here we only focus on $K_S^0$, since they are the most abundant. This step is necessary because the $\pi^\pm$ particles, coming from the $K_S^0$ decays, have large impact parameters, which is usually a trait of background particles. In order to minimize confusion from the MVA point-of-view, these tracks are taken into account separately.

We use the converted $K_S^0$ candidates from the existing Belle particle list and use a pre-trained Neural Network classifier in order to select only the good $K_S^0$ candidates. Figure 6.5 shows the distribution of the $K_S^0$ invariant mass for signal and background candidates, before and after the classifier cut. The momentum of selected $K_S^0$ candidates is added to the ROE, while the daughter tracks are discarded from our set.

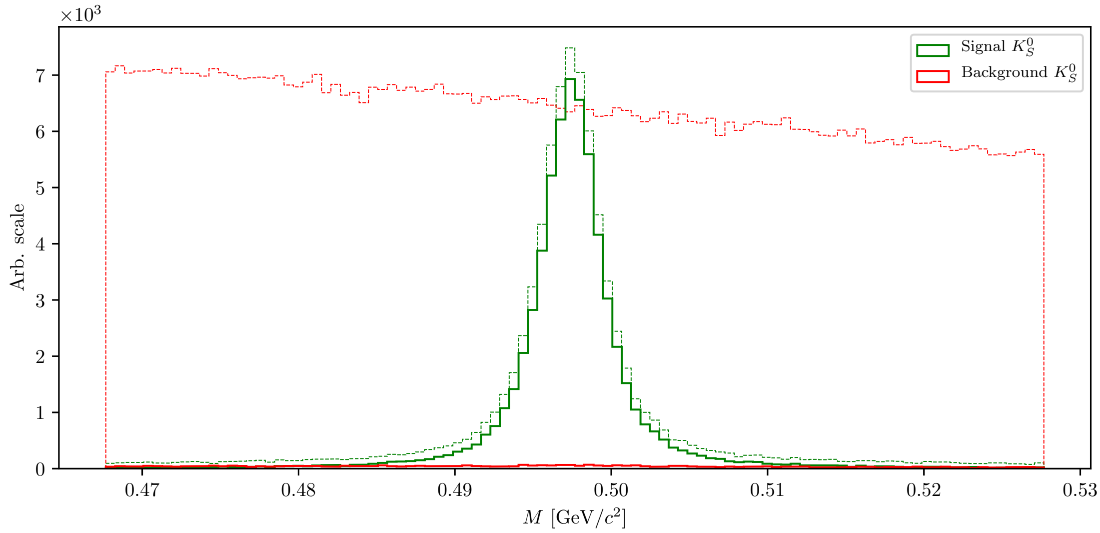



Figure 6.5: Invariant mass of the $K_S^0$ candidates before (dashed lines) and after (solid lines) the cut on the Neural Network classifier for signal (green) and background candidates (red). Signal peaks at nominal $K_S^0$ mass, while background covers a wider region.

The signal efficiency and background rejection for $K_S^0$ candidates after this cut and on the full range are

- Signal efficiency: $\epsilon_{SIG} = 80.7$ %,
- Background rejection: $1 - \epsilon_{BKG} = 99.4$ %.

### 6.2.2 Duplicate Tracks

All good tracks at this point should be coming from the IP region, since we took care of all the good tracks from long-lived particle decays, therefore we apply a cut on impact parameters for all the remaining tracks

- $|d_0| < 10$ cm and $|z_0| < 20$ cm

and proceed with the clean-up of track duplicates.

#### Defining a duplicate track pair

In this step, we wish to find a handle on secondary tracks from low momentum curlers and decays in flight. The main property for these cases is that the angle between such two tracks is very close to 0° or 180°, since tracks deviate only slightly from the initial direction, but can also be reconstructed in the opposite way. Figure 6.6 shows the distribution of the angle between two tracks in a single pair for random track pairs and duplicate track pairs, where the latter were reconstructed as two same-sign or opposite-sign tracks.

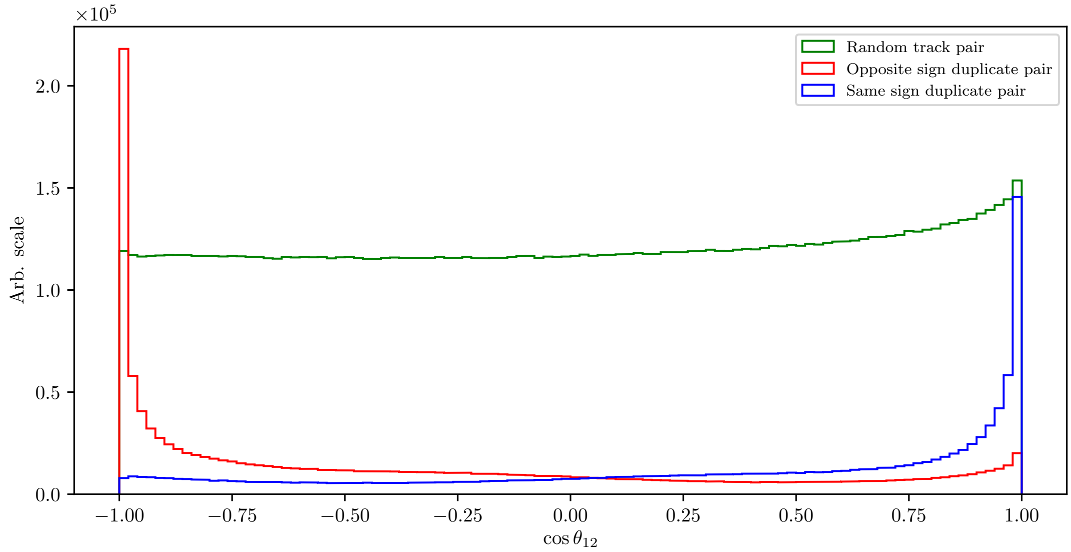



Figure 6.6: Distribution of the angle between two tracks in a single pair for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

If the particle decayed mid-flight or produced multiple tracks due to being a low-momentum curler, then, as the name suggests, these particles most likely had low momentum in the transverse direction, $p_T$. Since both tracks originate from the same initial particle, the momentum difference should also peak at small values. Figure 6.7 shows the momentum and momentum difference of tracks which belong to a random or a duplicate track pair.

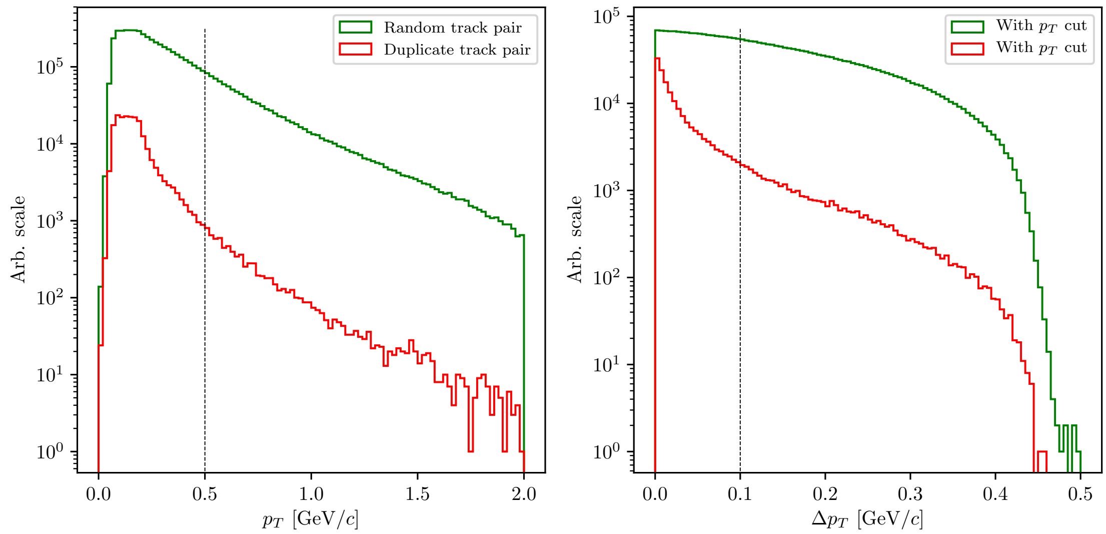



Figure 6.7: Distribution of transverse momentum $p_T$ (left) and transverse momentum difference $\Delta p_T$ (right) for all tracks coming from random (green) or duplicate track pairs (red). The plot on the right already includes the cut on $p_T$ from the plot on the left.

We impose a cut of

- $p_T < 0.5$ GeV/$c$,
- $|\Delta p_T| < 0.1$ GeV/$c$,

in order to cut down the number of random track pairs, while retaining a high percentage of duplicate track pairs. After all the cuts defined in this chapter, the final distribution of the angle between two tracks is shown in Figure 6.8.

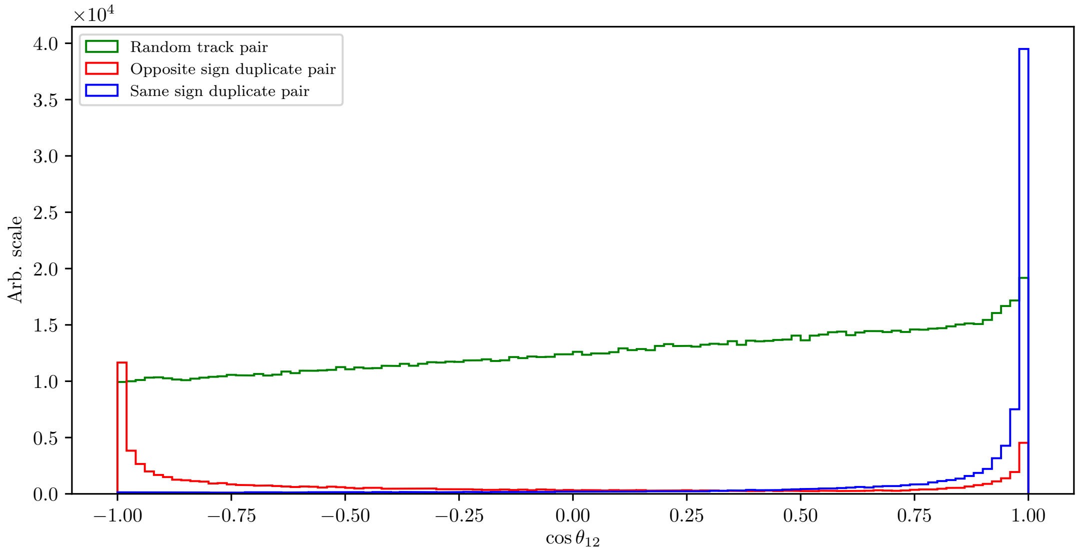



Figure 6.8: Distribution of the angle between two tracks in a single pair after applying the selection cuts defined in this subsection. The distributions are shown for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

### Training the duplicate track pair MVA

This final sample of track pairs is now fed into an MVA, which is trained to recognize duplicate track pairs over random ones. The training dataset contains

- 113707 target candidates,
- 190314 background candidates,

where the definition of a target is that the track pair is a duplicate track pair.

The input variables used in this MVA are

- angle between tracks,
- track quantities
  - impact parameters $d_0$ and $z_0$,
  - transverse momentum $p_T$,
  - helix parameters and helix parameter errors of the track,

- track fit $p$-value,
- number of hits in the SVD and CDC detectors

The classifier is able to distinguish between random and duplicate track pairs in a very efficient manner, as shown in Figure 6.9.

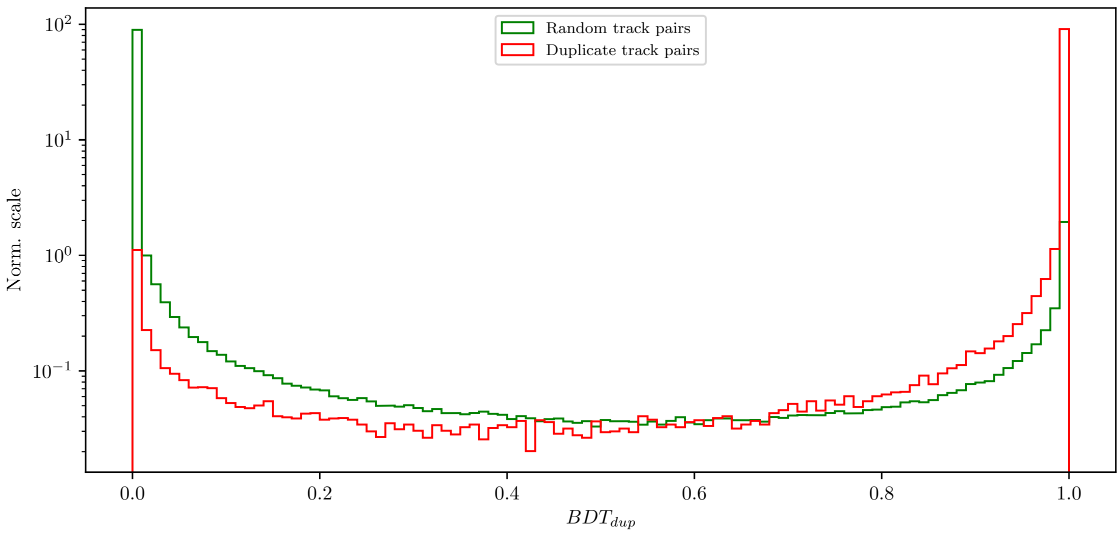



Figure 6.9: Classifier output of the track pair training for random track pairs and duplicate track pairs.

The FOM function for optimal cut selection is shown in Figure 6.10 (left), along with the angle between the two tracks before and after the optimal cut (right). The optimal cut for duplicate track selection is

- $BDT_{duplicate} > 0.9985$.

The signal efficiency and background rejection for duplicate pair candidates after this cut is

- Signal efficiency: $\epsilon_{SIG} = 87.2$ %,
- Background rejection: $1 - \epsilon_{BKG} = 98.8$ %,

where signal and background represent duplicate and random track pairs, respectively.

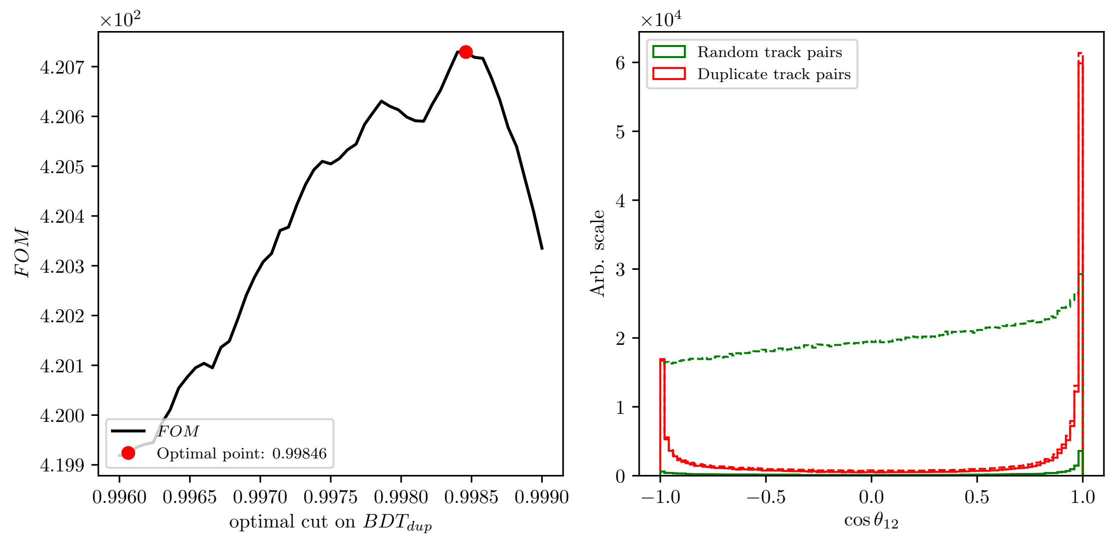



Figure 6.10: The optimization of the FOM function for the cut on classifier output (left) and distribution of the angle between two tracks in a single pair before (dashed) and after (solid) applying the optimal cut on the output classifier for random and duplicate track pairs (right).

### Defining duplicate tracks

What remains now is to decide which track from the duplicate track pair to keep and which to discard. For this purpose, we apply duplicate pair-level information to each track in the pair in the form of

$$\Delta f = f_{this} - f_{other}, \tag{6.1}$$

where $f$ is an arbitrary variable from the list of track quantities in section 6.2.2. From the point-of-view of *this* track, a track is more *duplicate*-like if the following is true

- $\Delta d_0$, $\Delta z_0 > 0$ (*this* track further away from the IP region),
- $\Delta p_T$, $\Delta p_Z < 0$ (*this* track has lower momentum),
- $\Delta N_{SVD}$, $\Delta N_{CDC} < 0$ (*this* track has less hits in the SVD and CDC),

Additionally we define an MC truth variable

$$\Delta\chi^2 = \chi^2_{this} - \chi^2_{other}, \quad \chi^2 = \sum_{i=x,y,z} \frac{\left(p_i - p_i^{MC}\right)^2}{\sigma(p_i)^2}, \tag{6.2}$$

where we compare all components of track momentum to the true values. If $\Delta\chi^2 > 0$, then *this* track has a higher probability of being a duplicate track and should be discarded.

However, it turns out that solving this problem is not as simple as discarding one track and keeping the other one. An additional complication here is that we can have more than one extra track from the same initial particle, which leads to track pairs where both tracks are track duplicates. For example, if we have the following case

$$\begin{aligned} t_1 &: \text{good track}, \\ t_2 &: \text{extra track}, \\ t_3 &: \text{extra track}, \\ \text{pair}_1 &: (t_1, t_2)\,, \\ \text{pair}_2 &: (t_1, t_3)\,, \\ \text{pair}_3 &: (t_2, t_3)\,, \end{aligned}$$

where $t_1$ is the original track and $t_2$ and $t_3$ are extra tracks, with $t_3$ being even more duplicate-like with respect to $t_2$. Here tracks $t_2$ and $t_3$ should be discarded while $t_1$ should be kept. We can achieve this if we overwrite existing pair-level information in the tracks for cases where the variable difference $\Delta f$ is more duplicate-like. If we follow the same example, we could fill information about the property $f$ in six different orders.

$$\begin{aligned} &1.\ (t_1, t_2*) && \rightarrow && (t_1, t_3*) && \rightarrow && (t_2*, t_3*)\,, \\ &2.\ (t_1, t_2*) && \rightarrow && (t_2*, t_3*) && \rightarrow && (t_1, t_3*)\,, \\ &3.\ (t_1, t_3*) && \rightarrow && (t_2, t_3*) && \rightarrow && (t_1, t_2*)\,, \\ &4.\ (t_1, t_3*) && \rightarrow && (t_1, t_2*) && \rightarrow && (t_2*, t_3*)\,, \\ &5.\ (t_2, t_3*) && \rightarrow && (t_1, t_3*) && \rightarrow && (t_1, t_2*)\,, \\ &6.\ (t_2, t_3*) && \rightarrow && (t_1, t_2*) && \rightarrow && (t_1, t_3*)\,, \end{aligned}$$

where the "*" symbol denotes when a track is recognized as a duplicate track with respect to the other track. We see that no matter the order, both $t_2$ and $t_3$ get recognized as duplicate tracks correctly.

### Training the duplicate track MVA

The training procedure is similar as before. The sample of tracks from duplicate track pairs is now fed into an MVA, which is trained to distinguish duplicate tracks from good tracks. The training dataset contains

- 84339 target candidates,
- 68280 background candidates,

where the definition of a target is that the track is a duplicate track.

The input variables used in this MVA are

- theta angle of the track momentum,
- track quantities
  - impact parameters $d_0$ and $z_0$ and their errors,
  - CMS frame momentum $p_{CMS}$ and momentum components $p_T$ and $p_z$
  - number of hits in the SVD and CDC detectors
  - track fit $p$-value,

- pair-level information
  - $\Delta d_0$, $\Delta z_0$, $\Delta N_{CDC}$, $\Delta N_{SVD}$, $\Delta p_T$, $\Delta p_z$, $\Delta p$-value.

The weights from this training are applied to the tracks, where now each track has a certain probability of being a duplicate track. We then compare these values between both tracks in each track pair as

$$\Delta BDT_{final} = BDT^{this}_{final} - BDT^{other}_{final}, \tag{6.3}$$

which is again applied to all track pairs and overwritten for tracks which are more duplicate-like. The classifier output and the classifier output difference for each track are shown in Figure 6.11.

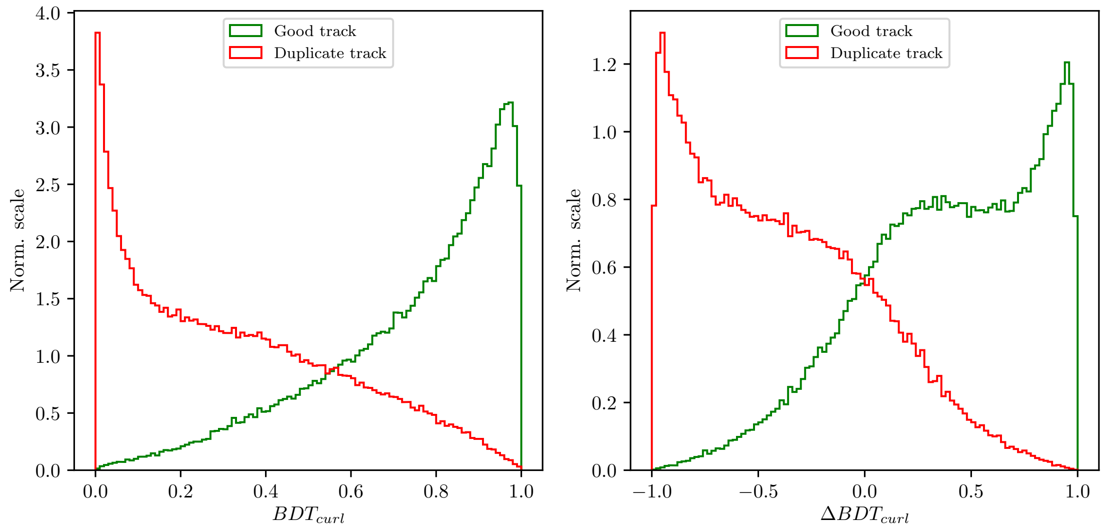



Figure 6.11: Classifier output of the MVA training for curling track recognition (left) and difference of the classifier output, calculated for each track in a track pair (right).

Finally, we select all duplicate tracks which survive the cut

$$\Delta BDT_{final} > 0 \tag{6.4}$$

and discard them from our ROE. We can check the performance of our duplicate track classifier by applying the procedure on a validation sample of duplicate track pairs and compare the predicted result with the truth, based on Eq. (6.2). Table 6.1 shows the performance of the duplicate track recognition in the form of percentages of correctly and incorrectly identified duplicate and original tracks. The model seems to perform well and the event is now considered to be clean of duplicate tracks.

| | Predicted duplicate track | Predicted good track |
|---|---|---|
| Duplicate track | 83.07 % | 22.62 % |
| Good track | 16.93 % | 77.38% |

Table 6.1: Ratios of correctly classified and misclassified tracks.

## 6.3 Belle Clean-up

The clean-up, used standardly at Belle, is much simpler and relies only on a set of rectangular cuts for neutral particles as well as charged ones. In case of photons, a single cut on photon energy is applied, depending on the region where the photon hit the relevant part of the detector. The photon cuts are summarized in Table 6.2.

| | $17^\circ < \theta < 32^\circ$ | $32^\circ < \theta < 130^\circ$ | $130^\circ < \theta < 150^\circ$ |
|---|---|---|---|
| $E_\gamma$ | $> 100$ MeV | $> 50$ MeV | $> 150$ MeV |

Table 6.2: Photon selection for the Belle clean-up procedure. Different cuts are applied on photons in different parts of the detector

In case of tracks, pairs are selected which satisfy the following criteria:

- $p_T < 275$ MeV/$c$,
- $\Delta p = |\mathbf{p}_1 - \mathbf{p}_2| < 100$ MeV/$c$,
- $\cos\theta(\mathbf{p}_1, \mathbf{p}_2) < 15^\circ$ for same sign,
- $\cos\theta(\mathbf{p}_1, \mathbf{p}_2) > 165^\circ$ for opposite sign.

Of the two tracks, the one with a larger value of formula in Eq. 6.5 is discarded. The remaining tracks in the event then need to satisfy the conditions described in Table 6.3.

$$(\gamma|d_0|)^2 + |z_0|^2, \quad \gamma = 5. \tag{6.5}$$

| | $p_T < 250$ MeV/$c$ | $250$ MeV/$c < p_T < 500$ MeV/$c$ | $p_T > 500$ MeV/$c$ |
|---|---|---|---|
| $\|d_0\|$ | $< 20$ cm | $< 15$ cm | $< 10$ cm |
| $\|z_0\|$ | $< 100$ cm | $< 50$ cm | $< 20$ cm |

Table 6.3: Photon selection for the Belle clean-up procedure. Different cuts are applied on photons in different parts of the detector

## 6.4 Clean-up Results

In this section, the results of the ROE clean-up are shown. It is obvious that cleaning up the event affects the shape of various distributions, especially $\Delta E$ and $M_{BC}$, which we are most interested in. Since the reconstruction procedure includes cuts on the cleaned-up variables, the clean-up also affects the efficiency of the reconstructed sample, not only the resolution.

We compare the clean-up setup, defined in this analysis, to the standard clean-up used by Belle, and to a default case, where no clean-up was applied at all. We apply the clean-up procedure to our signal MC sample with all the applied cuts, defined in section 5.6, except for the signal categorization cuts. Figure 6.12 (left) shows signal candidate distributions of $\Delta E$ and $M_{BC}$ for various clean-up setups. Focusing on the ROE clean-up, we see an improvement in resolution in both observed variables and an overall decrease in efficiency. The efficiency decrease is expected since the cleaned-up variables are able to better isolate the perfectly reconstructed candidates and discard the non-perfect candidates. In fact, the efficiency of the perfectly reconstructed candidates increases after the ROE clean-up, as shown in Figure 6.12 (right). The signal MC sample after in case of the Belle clean-up also shows a slight improvement in the resolution, but looking at the perfectly reconstructed candidates we see that this clean-up procedure is not optimal. Table 6.4 shows ratios of efficiencies and $FWHM$'s of the clean-up procedures for the perfect signal with respect to the default case, based on the $\Delta E$ distribution. While both the Belle and ROE clean-up improve the resolution, ROE clean-up performs significantly better and also increases the amount of the perfectly reconstructed candidates in the final sample.

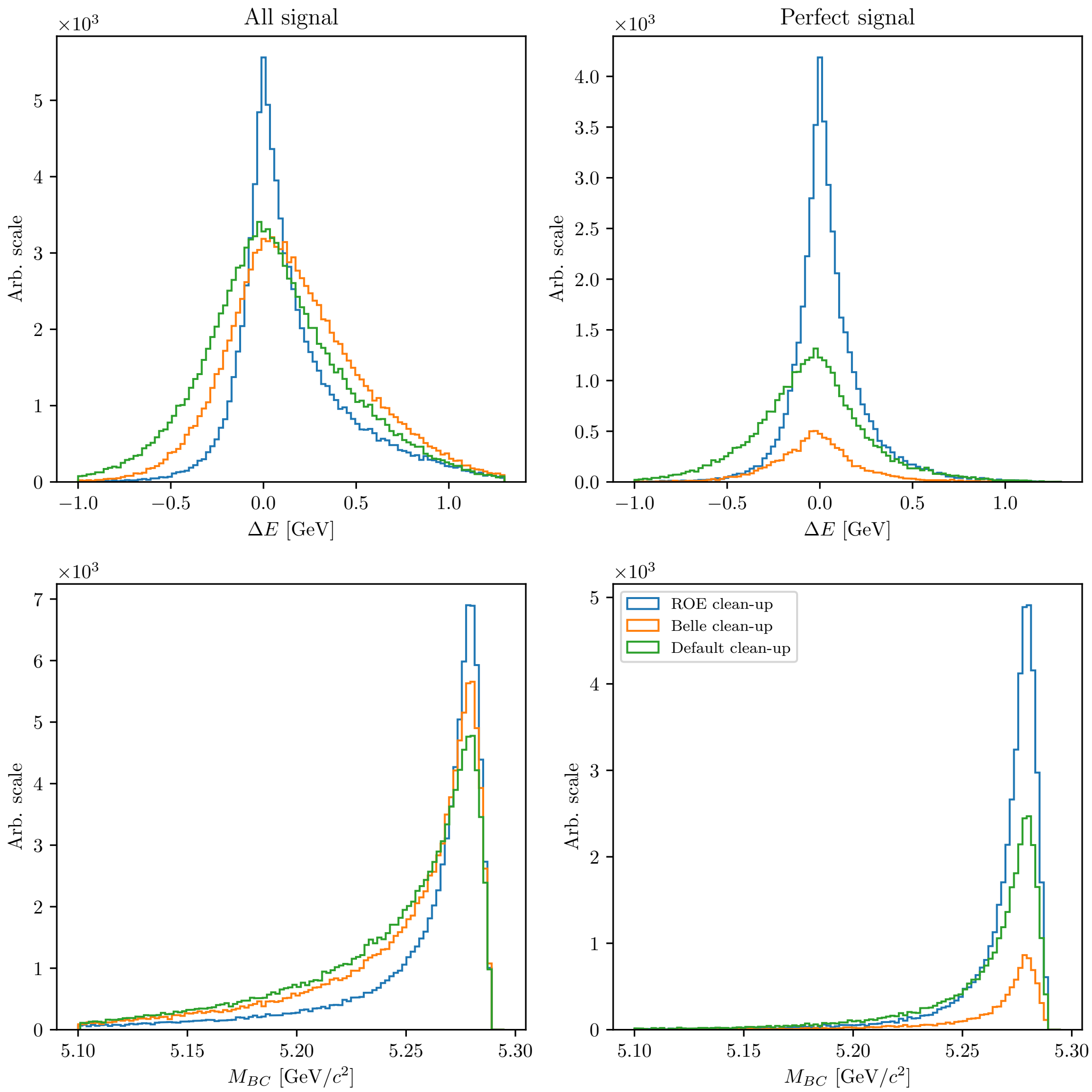



Figure 6.12: $\Delta E$ and $M_{BC}$ distributions for various types of clean-up procedures. The figures on the left are shown for the full signal sample after the stated cuts, while the figures on the right are shown for the perfectly reconstructed signal candidates. For ROE clean-up, the procedure seems to improve resolution as well as increase the amount of perfectly reconstructed candidates, relative to the default case.

| | Efficiency ratio | FWHM ratio |
|---|---|---|
| Belle clean-up | 28.5 % | 75.0 % |
| ROE clean-up | 140.1 % | 35.0 % |

Table 6.4: Comparison of efficiencies and $FWHM$'s of ROE and Belle clean-up setups with respect to the default case (no clean-up).

Another variable which heavily depends on the clean-up is the charge product of the signal and companion $B$ meson candidate, already defined in Eq. (5.15), shown in Figure 6.13 for various clean-up procedures. The figure shows an improved resolution of the distribution, which means that candidates migrate to the correct value of the charge product after the clean-up. Looking at the perfectly reconstructed candidates we again see the increase in the bin corresponding to the correct charge product. As a cross-check, we can also look at $\Delta E$ and $M_{BC}$ variables for each value of the charge product. These plots are shown for the full signal MC sample in Figure 6.14 and they show a clear resolution improvement for the correct value of the charge product in the case of the ROE clean-up. For other values of the charge product, there also seems to be a small improvement for both cases of clean-up, but it is negligible compared to the plots in the second column. This supports our choice of signal categorization, defined in section 5.5, where we select only candidates with the correct value of the charge product.

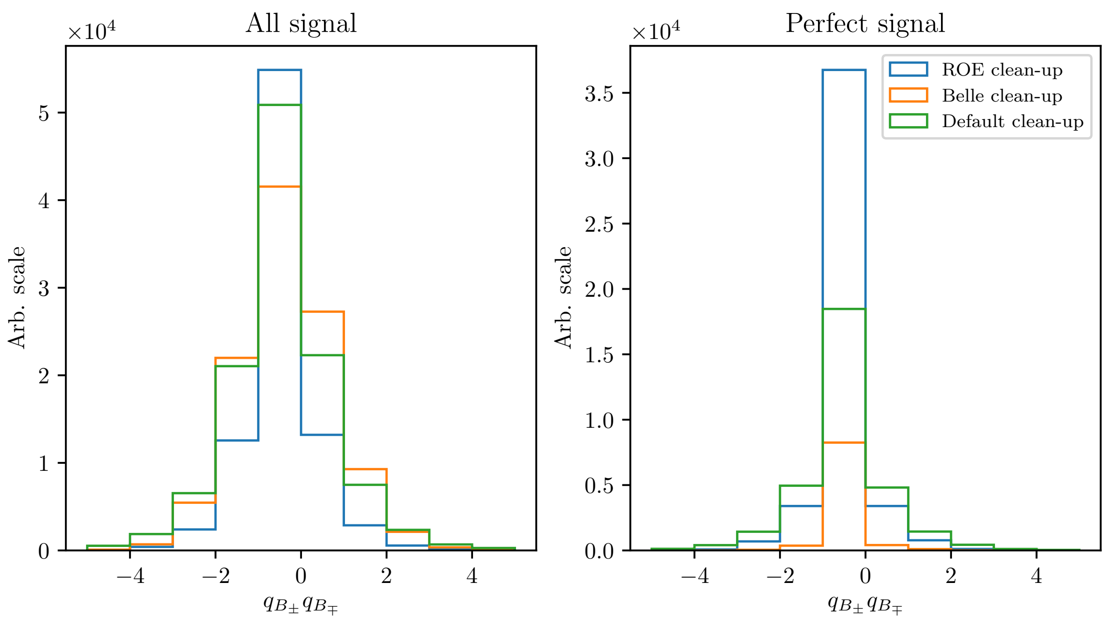



Figure 6.13: Distribution of the charge product of both $B$ mesons for various types of clean-up procedures, shown on the full signal MC (left) and for the perfectly reconstructed signal candidates (right). For ROE clean-up, the procedure seems to increase the number of perfectly reconstructed candidates.

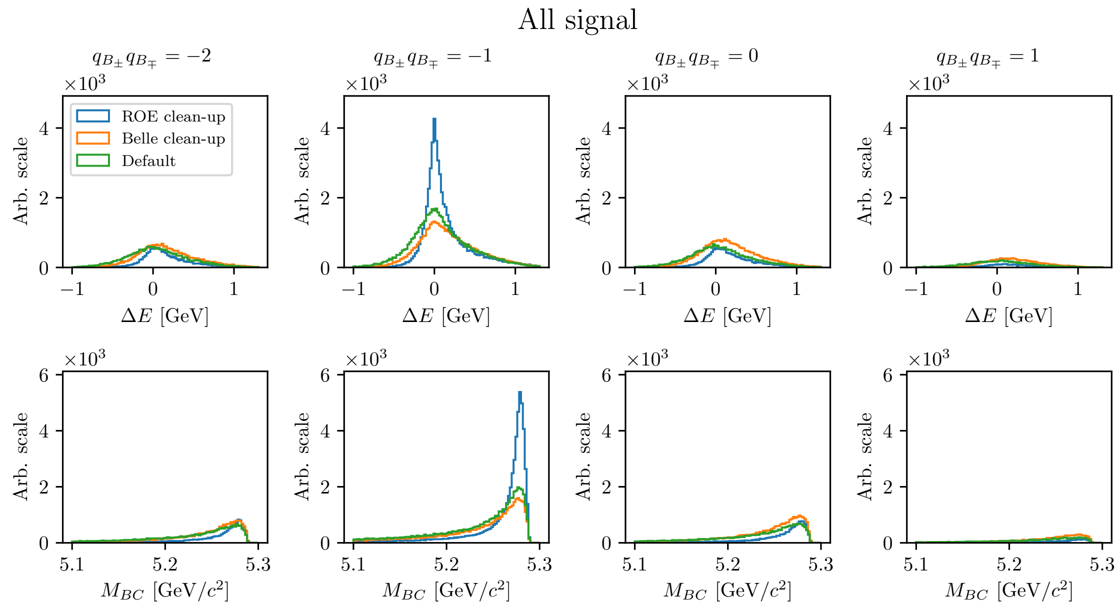



Figure 6.14: Distributions of $\Delta E$ (top) and $M_{BC}$ (bottom) for various types of clean-up procedures, split by specific values of the charge product, shown for the full signal MC. There is a significant improvement in resolution after ROE-cleanup for the case of the correct value of the charge product.

## 6.5 ROE Clean-up Validation

The ROE clean-up seems to perform well on signal MC based on the results in the previous section. However, it is necessary to make sure that this procedure performs as well on other simulated and measured data, which is done in this section. The clean-up procedure is validated on the control sample,

$$B^+ \to \bar{D}^0 \ell^+ \nu, \quad D^0 \to K^+ K^-,$$

which was already defined in section 2.2. The control candidates are reconstructed in the same manner as the signal candidates. In addition to the same cuts applied as in the previous section we also apply a selection to make the control sample more significant. We discard all candidates which fail to pass the cut on invariant mass of the two kaons

$$1.849 \text{ GeV}/c^2 < m_{KK} < 1.879 \text{ GeV}/c^2 \tag{6.6}$$

as shown in Figure 6.15.

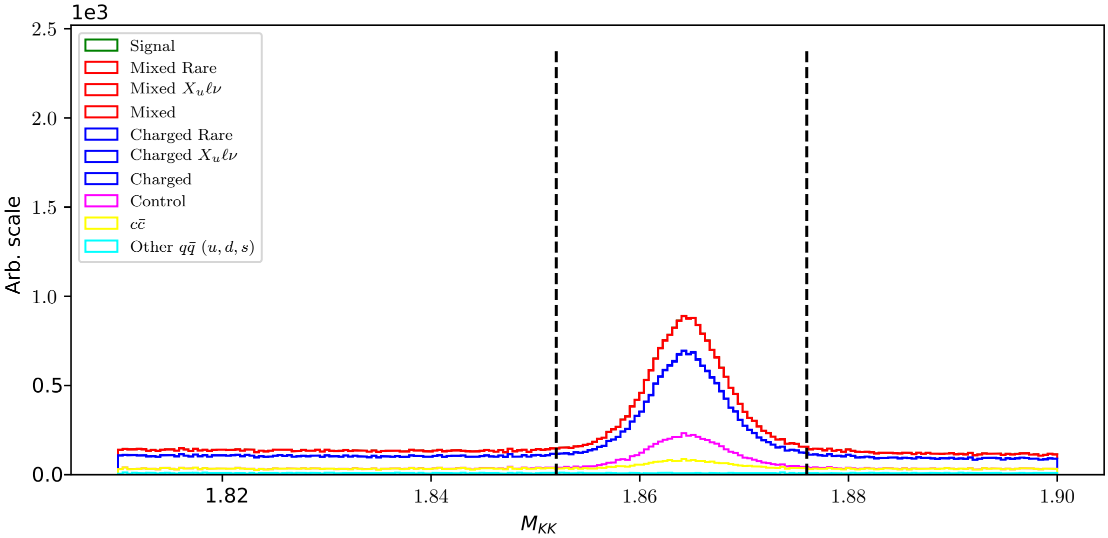



Figure 6.15: Distributions of $m_{KK}$ for the full MC dataset. The black lines represent the cut on region in the $m_{KK}$ distribution where the control sample is enhanced.

With the control sample selection determined, we now run the reconstruction with and without the ROE clean-up procedure on MC and data. For the purpose of this validation, we only run the reconstruction over 1 stream of the full available generated MC.

The effects of the ROE clean-up are shown in Figure 6.16. We see that data and MC agree well, and evidently even better after the clean-up procedure. The control sample resolution seems very poor in the case without the clean-up, but it improves significantly if the clean-up procedure is applied, as expected. The simulated background also seems to gain an improvement in the resolution, but this is likely due to the background consisting of similar candidates as the control sample. This means that the clean-up performs as expected due to the nature of the decays and does not arbitrarily shape the background to be more signal like. Additionally, it should be pointed out that, after the clean-up, the simulated background resolution is worse compared to the control decay resolution, while this is not the case if the clean-up procedure is not performed.

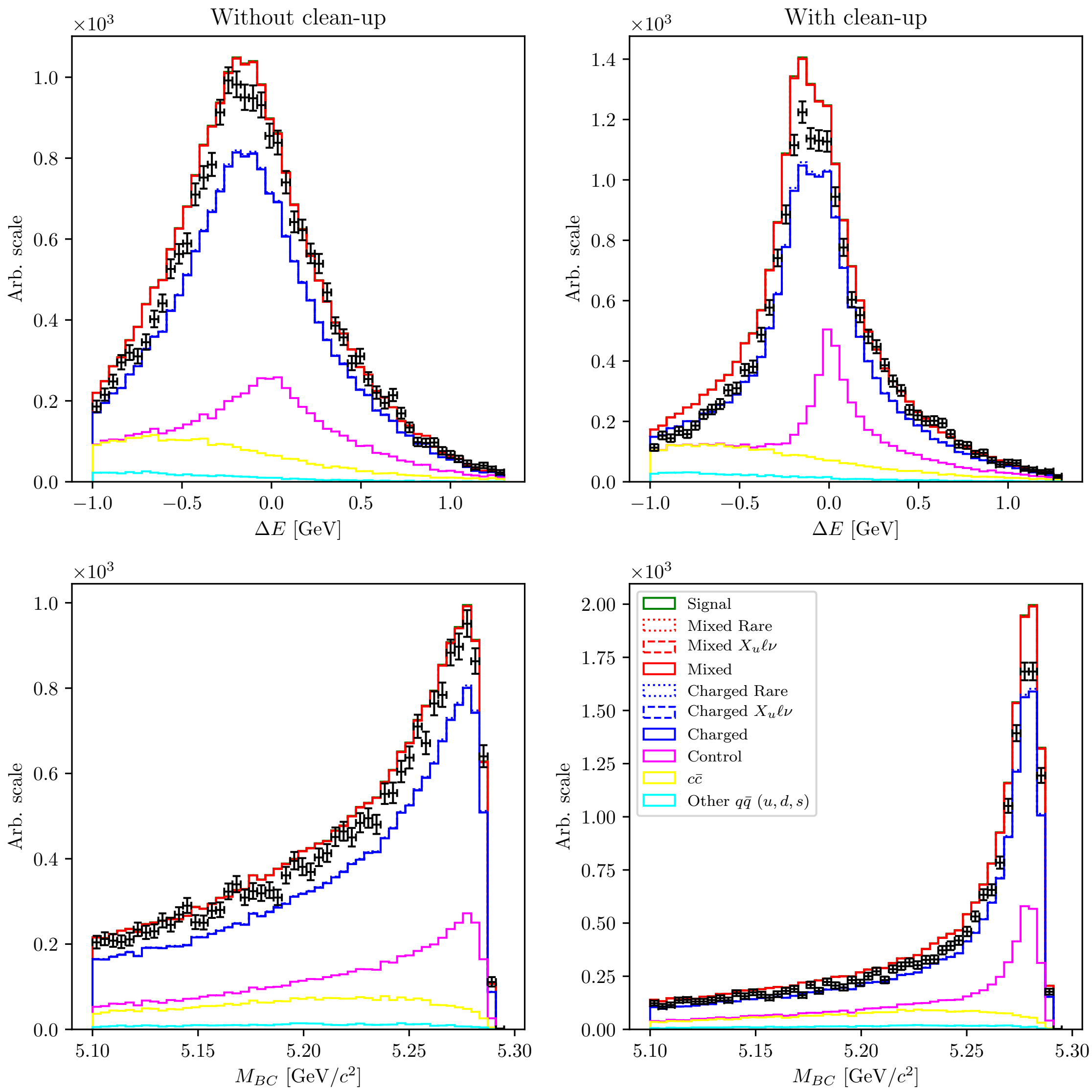



Figure 6.16: Distributions of $\Delta E$ (top) and $M_{BC}$ bottom for the case without (left) and with ROE clean-up (right). The resolution of the control sample is improved and the MC and data agree well in all aspects. While the simulated background resolution is also improved, it is worse compared to the resolution of the control sample.

To perform the clean-up validation in greater detail we also compare the data and MC agreement in bins of the charge product of the two $B$ mesons. Figure 6.17 shows the cleaned-up versions of $\Delta E$ and $M_{BC}$ for each charge product bin in the same manner as shown in the previous section. We see that the MC and data agreement persists in all cases.

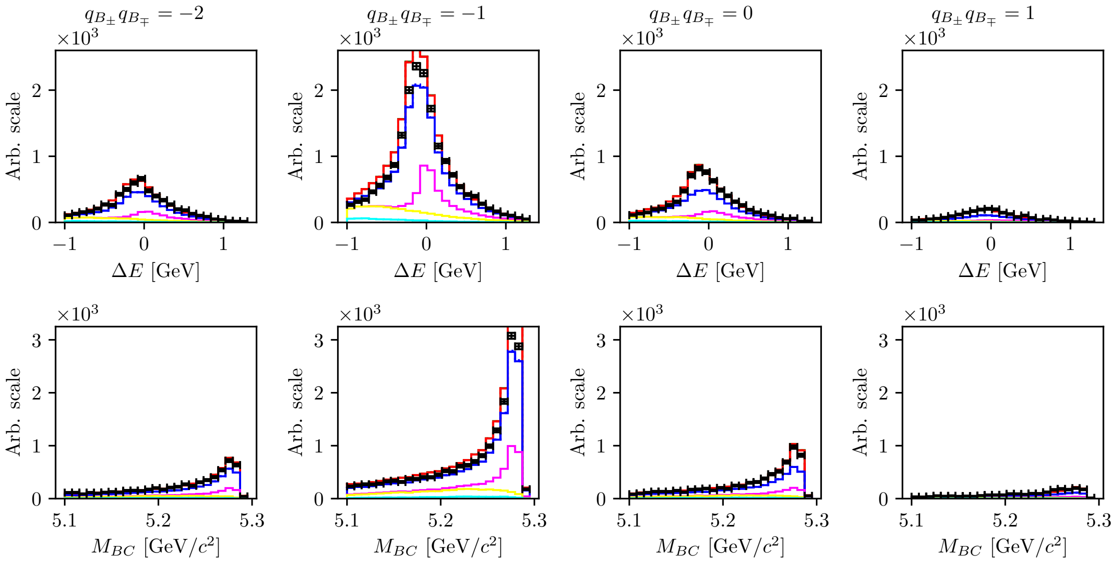



Figure 6.17: Distributions of $\Delta E$ (top) and $M_{BC}$ (bottom) split in bins of the charge product of the two $B$ mesons.

The ROE clean-up procedure seems to perform well. It significantly improves the resolution of the signal/control candidates and increases the amount of perfectly clean events. The clean-up procedure was also applied to data and no disagreement with respect to the simulated MC samples was found. This means that the procedure does not differ between MC and data and does not affect them differently. The procedure was therefore validated in great detail and is suitable to be used in this analysis.

# Chapter 7

# Background Suppression

This chapter shows the procedure in suppressing various kinds of background by applying cuts on MVA classifier outputs.

## 7.1 Resonant Background

In this analysis we study decays with kaons in the final state. This means that standard procedures in $b \to u$ analyses in order to suppress $b \to c$ backgrounds, such as $K$-veto, are not possible. As a consequence, our final sample consists of combinations of $K$ pairs coming also from $b \to c$ sources, such as $D^0 \to K^+K^-$. Such candidates usually have resonance-like properties in the two-kaon invariant mass spectrum. Figure 7.1 shows this invariant mass spectrum of two kaons, $m_{KK}$, where obvious resonant structures are present from sources like

- $\phi \to K^+K^-$ (sharp resonance at $\sim 1.019$ GeV/$c^2$),
- $D^0 \to K^+K^-$ (sharp peak at $\sim 1.864$ GeV/$c^2$),
- $D^0 \to K^+\pi^-$ (wide, shifted peak, due to kaon miss-identification).

In order to suppress these resonant backgrounds, while studying signal or control decay, we impose a set of the following cuts

- Signal cut: $|m_{KK} - m_\phi| > \Delta_\phi$, $|m_{KK} - m_{D^0}| > \Delta_{D^0}$, $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$,
- Control cut: $|m_{KK} - m_{D^0}| \leq \Delta_{D^0}$, $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$,

where $m_{KK}$ is the $KK$ invariant mass and $m_{K\pi}$ is the invariant mass of $KK$, where the kaon's mass, which has the same charge as the $B$ meson, was given the mass of the $\pi^0$ particle, and where $m_\phi \approx 1.019$ GeV/$c^2$ and $m_{D^0} \approx 1.864$ GeV/$c^2$ are nominal masses of the $\phi$ and $D^0$ mesons, and $\Delta_\phi \approx 8 \times 10^{-3}$ GeV/$c^2$ and $\Delta_{D^0} \approx 1.5 \times 10^{-2}$ GeV/$c^2$ are symmetric cut widths around the nominal mass values for the $\phi$ and $D^0$ mesons, respectively. By imposing the signal or control cut on our data, we are able to efficiently isolate the desired subset, which is very useful for further studies of the control decay. Table 7.1 shows the subsample efficiency after applying either of the cuts.

| | $\epsilon$(Signal cand.) | $\epsilon$(Control cand.) | $\epsilon$($\phi$ resonance cand.) |
|---|---|---|---|
| Signal cut | 95.4% | 4.0% | 13.6% |
| Control cut | 1.9% | 96.0% | 0.0% |

Table 7.1: Various subset efficiencies after imposing the signal or control cut on the $KK$ invariant mass.

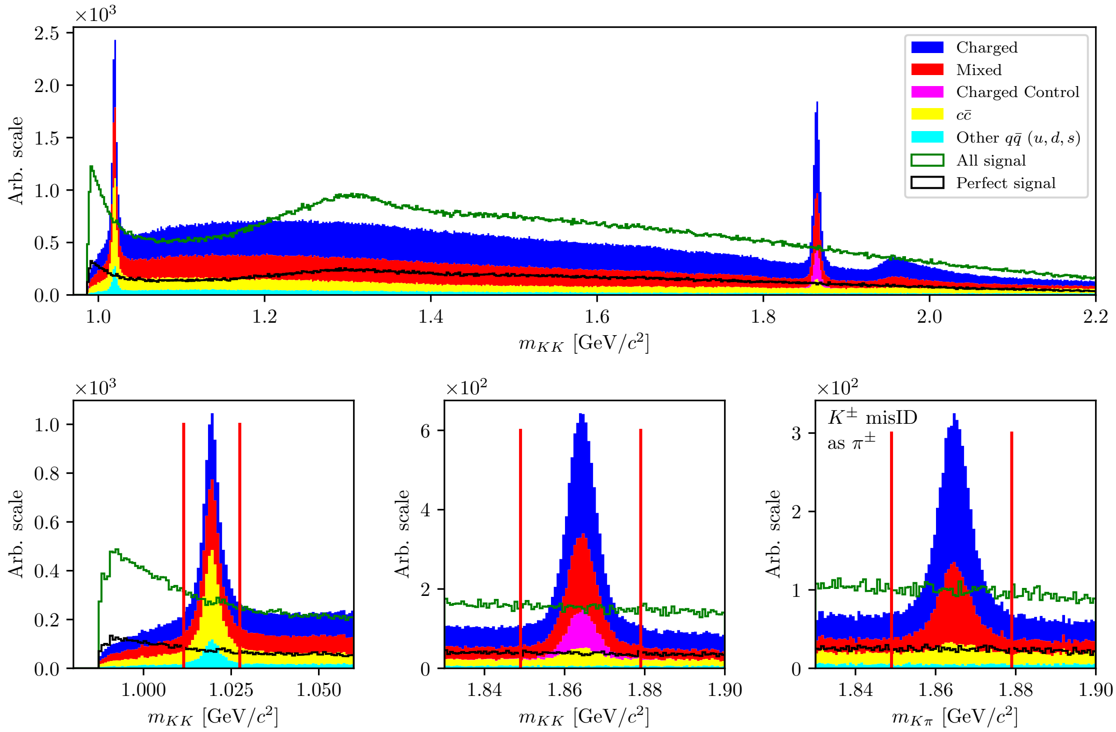



Figure 7.1: Invariant mass of two correctly reconstructed kaons (left) and invariant mass of two kaons, where one was miss-identified as a pion (right). Signal (green) and perfect signal (black) are equally scaled up.

## 7.2 Continuum Cuppression

Continuum background are physics processes where continuum states are produced in electron and positron collisions

$$e^+e^- \to q\bar{q},$$

where $q = u,\ d,\ s$ or $c$, and are a sizable contribution to $B\bar{B}$ events. Additionally to kinematic constraints to separate $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ decays from $e^+e^- \to q\bar{q}$, properties of the "event shape" are also often used, because phase-space distributions of decayed particles differ for these two processes. Continuum background events are generated in a back-to-back way in the CMS frame, so hadrons produced in the quark fragmentation possess only a small transverse momentum compared to the initial momentum magnitude. This leads to a spatially confined, jet-like structure. On the other hand, $B$ mesons from $B\bar{B}$ events are produced almost at rest in the

CMS frame. Their decay products from an isotropic distribution in the detector, which yields a spherical event shape.

### 7.2.1 Characteristic Variables

Information on the phase-space distribution of decay particles is obtained in a number of different ways. In this subsection different characteristic variables are presented which are used in the MVA training. They all focus on kinematic and shape differences between the two processes, which we want to discriminate.

#### Thrust and Related Variables

It is possible to define a thrust axis $\mathbf{T}$ for a collection of $N$ momenta $p_i$ as a unit vector along which their total projection is maximal. Thrust axis $\mathbf{T}$ can be obtained by maximizing the expression

$$\mathbf{T} = \frac{\sum_i |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_i |\mathbf{p}_i|}. \tag{7.1}$$

In this case, a related variable is $|\cos\theta_T|$, where $\theta_T$ is the angle between the thrust axis of the momenta from $B$ meson decay particles and the thrust axis of all particles in the ROE. Since both $B$ mesons in $B\bar{B}$ events are produced at rest, their decay particles, and consequentially their thrust axes, are uniformly distributed in the range $[0,\ 1]$. On the other hand, decay particles from continuum events follow the direction of the jets in the event. As a consequence, the thrusts of both the $B$ meson and the ROE are strongly directional and collimated, which results in a large peak at $|\cos\theta_T| \approx 1$. Additionally, one can also use the variable $|\cos\theta_{TB}|$, which is the thrust axis between the $B$ candidate and the beam axis. For $B$ candidates from $B\bar{B}$, this distributions is again uniformly distributed, while for candidates from continuum events this distribution follows the distribution of the jets with the function $1 + \cos^2\theta_{T,B}$. Figure 7.2 shows the distributions of $|\cos\theta_T|$ (left) and $|\cos\theta_{T,B}|$ (right) for different $B$ meson candidates.

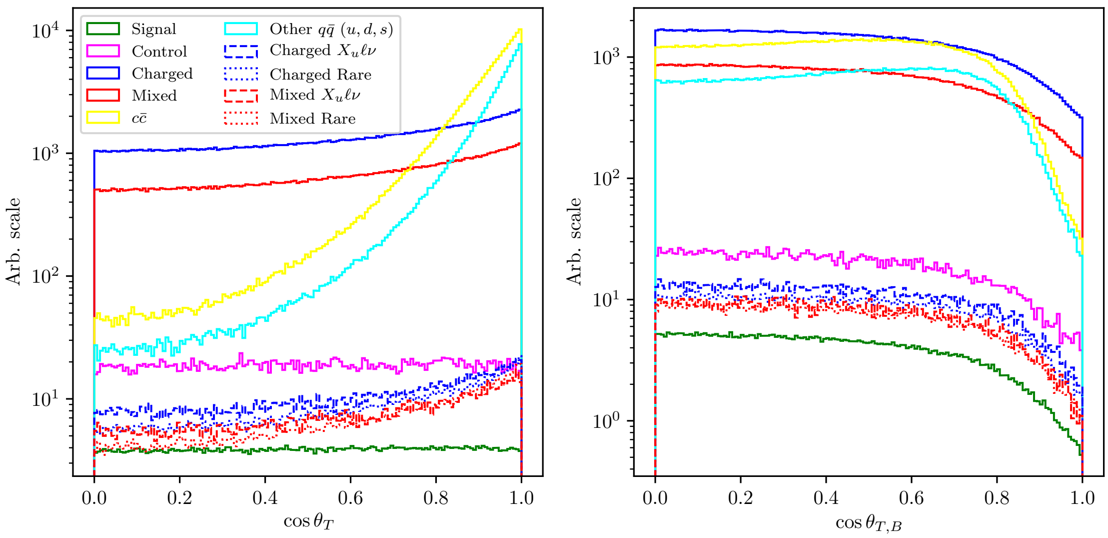



Figure 7.2: Distributions of $|\cos\theta_T|$ (left) and $|\cos\theta_{T,B}|$ (right) for different $B$ meson candidates.

### CLEO Cones

CLEO cones have been introduced by the CLEO collaboration [4] and are an additional specific tool to provide optimal background discrimination. They are nine variables corresponding to the momentum flow around the thrust axis of the $B$ meson candidate, binned in nine cones of 10° around the thrust axis, as illustrated in Figure 7.3.

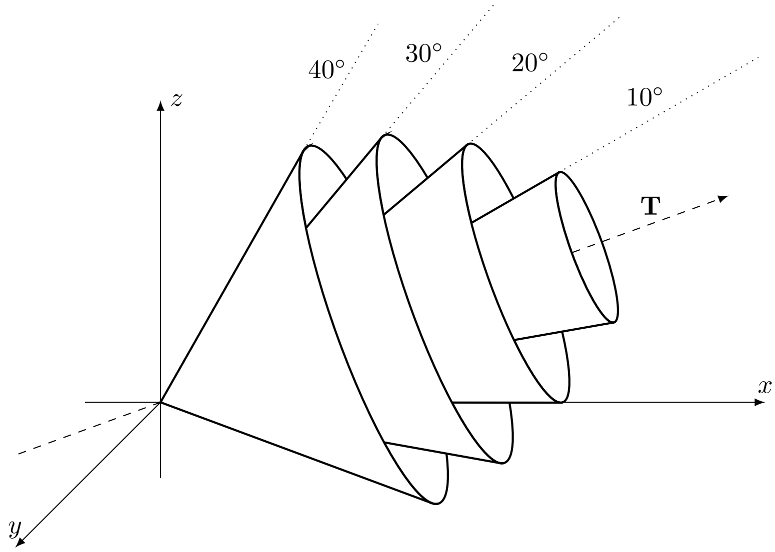



Figure 7.3: Concept of CLEO cones. **T** denotes the thrust axis of the $B$ meson candidates in the event. Each variable corresponds to a momentum flow around the thrust axis in steps of 10°.

### KSFW Moments

Fox-Wolfram moments are another useful parametrization of phase-space distribution of energy flow and momentum in an event. For a collection of $N$ momenta $p_i$, the $k$-th order normalized Fox-Wolfram moment $R_k$ is defined as

$$R_k = \frac{H_k}{H_0} = \frac{1}{H_0} \sum_{i,j} |p_i||p_j| P_k(\cos\theta_{ij}), \tag{7.2}$$

where $\theta_{ij}$ is the angle between $p_i$ and $p_j$, and $P_k$ is the $k$-th order Legendre polynomial. For events with two strongly collimated jets, $R_k$ takes values close to 0 (1) for odd (even) values of $k$, so these moments provide a convenient discrimination between $B\bar{B}$ and continuum events.

Belle developed a refined generation of Fox-Wolfram moments, called Kakuno-Super-Fox-Wolfram (KSFW) moments to further suppress the continuum background. There are 17 different KSFW moments which are grouped into $R_k^{so}$, $R_k^{oo}$ and $R_k^{ss}$ [5]. The latter ones are excluded due to correlations with $B$ meson specific variables.

## 7.2.2 MVA Training

Most of the characteristic variables, described in section 7.2.1, were taken together in order to train a single MVA classifier for continuum suppression. All characteristic variables were checked for possible $q^2$ correlation. Variables with significant correlation or complex shapes in the 2D plot were discarded from the training set, since they would have introduced unwanted dependence on the unreliable model, ISGW2, used for signal MC generation. Additionally, all of the characteristic variables in our set do not depend on the signal mode, they only differ in the kinematic and topological aspects of $B\bar{B}$ and continuum background events.

The training dataset consisted of $2\times10^5$ candidates, where 50 % of the candidates are correctly reconstructed signal events, 25 % are $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ background with expected proportions, and 25 % is $c\bar{c}$ background. Since the full Belle dataset is experiment dependent, we construct the training dataset by proportionally sampling each MC dataset, corresponding to the appropriate experiment number.

The training variable set consisted of

- $B$ meson direction and thrust related variables
  - magnitude of thrust axes of $B$ and $ROE$,
  - cosine of the angle between the thrust axis of $B$ and thrust axis of ROE,
  - cosine of the angle between the thrust axis of $B$ and beam direction,
  - reduced Fox-Wolfram moment $R_2$,
- all 9 CLEO Cones
- KSFW Moments
  - $R^{so}_{01}$, $R^{so}_{02}$, $R^{so}_{03}$, $R^{so}_{04}$,
  - $R^{so}_{10}$, $R^{so}_{12}$, $R^{so}_{14}$,
  - $R^{so}_{20}$, $R^{so}_{22}$, $R^{so}_{24}$,
  - $R^{oo}_{0}$, $R^{oo}_{1}$, $R^{oo}_{2}$, $R^{oo}_{3}$, $R^{oo}_{4}$,
- FlavorTagging variables
  - $qp$ of $e$, $\mu$, $\ell$,
  - $qp$ of intermediate $e$, $\mu$, $\ell$,
  - $qp$ of $K$, $K/\pi$, slow pion, fast hadron,
  - $qp$ of maximum $P^*$, $\Lambda$, fast-slow-correlated (FSC),
- Other
  - $\Delta z$, $\Delta t$.

Figure 7.4 shows the classifier output for various types of background, all in expected MC proportions. $B$ meson candidates from continuum background are dominant at lower values, while candidates from $B\bar{B}$ events populate the region with higher values.

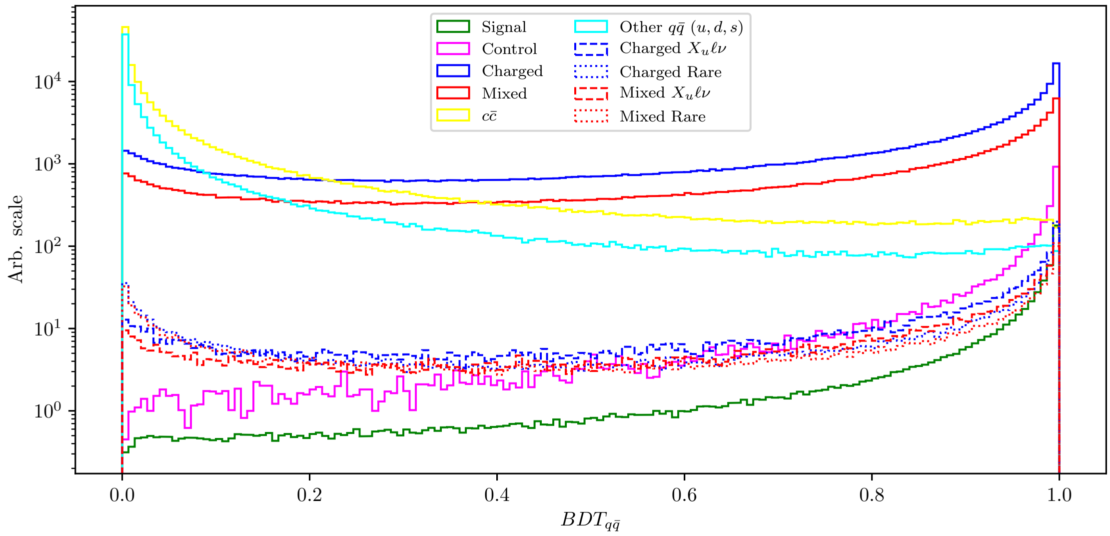



Figure 7.4: Continuum suppression classifier output for signal and various types of background. $B$ candidates from continuum events dominate the lower region, while candidates from $B\bar{B}$ dominate in the upper region of the classifier output.

## 7.3 $B\bar{B}$ Suppression

After separating continuum background from $B\bar{B}$ events, the next step is to train an MVA classifier to recognize our signal candidates amongst the candidates of other $B\bar{B}$ background. $B\bar{B}$ background consists of

- $b \to c\ell\nu$ background,
- $b \to u\ell\nu$ background,
- Other rare decays (radiative, penguin, rare 2- and 3-body decays, ...).

Similarly, the training dataset for this classifier consisted of $2 \times 10^5$ candidates, where 50 % of the candidates are correctly reconstructed signal events. The remaining part of the training dataset consists of all background, not including the control sample, because we are not interested in suppressing it directly. The background part of the dataset consists of 75 % charged and neutral $B\bar{B}$ events in equal proportions, whereas the remaining 25 % is equally populated with charged and neutral $B\bar{B}$ events from $b \to u\ell\nu$ and other rare decays. The training dataset was proportionally sampled in the same manner as described in section 7.2.2.

In order to separate this kind of background, we must be careful not to introduce correlations with the fit variables ($\Delta E$, $M_{BC}$) or any kind of model dependence (correlation with $q^2$). This means that we can not use any information of the decay particles or the candidate, which is of kinematics nature, such as decay particles momenta, decay angles or other variables with such behavior.

The training variable set consisted of

- fit probability of $P(\chi^2, DOF)$ of the signal candidate and the ROE side, separately,

- $\cos\theta_{BY}$ from Eq. (5.4),
- cos of the angle between momentum and vertex of $X$, where $X \in [KK,\ KK\ell,\ KK\ell\nu]$,
- FlavorTagger variables for the two signal-side kaons,
- number of kaons, tracks and distant tracks in ROE,
- $\theta$ angle of the ROE momentum in CMS frame,
- $\xi_Z$ from [6]
- $\Delta z$,
- $m^2_{miss}$ from Eq. (5.11),
- $B \to D^*\ell\nu$ veto variables,

where distant tracks are all tracks in ROE which satisfy the condition of $|d_0| >$ 10.0 cm or $|z_0| >$ 20.0 cm. The last entry, $B \to D^*\ell\nu$ veto variables, are a set of variables where we partially reconstruct the $D^*$ candidate 4-momentum via a linear combination of the $\pi_s^\pm$ 4-momentum in the $D^* \to D\pi_s^\pm$ decay. It helps discard the most dominant $B \to D^*\ell\nu$ background. It is most efficient in the $B^0 \to D^{*-}\ell^+\nu$ decay, where $D^{*-}$ further decays via $D^{*-} \to \bar{D}^0\pi_s^-$. Other decays do not contain a charged $\pi_s$ particle and are harder to reconstruct with good precision. This results in larger suppression of the neutral $B\bar{B}$ background only. Figures 7.5 shows the veto variable with a partial reconstruction of a charged $\pi_s^\pm$.

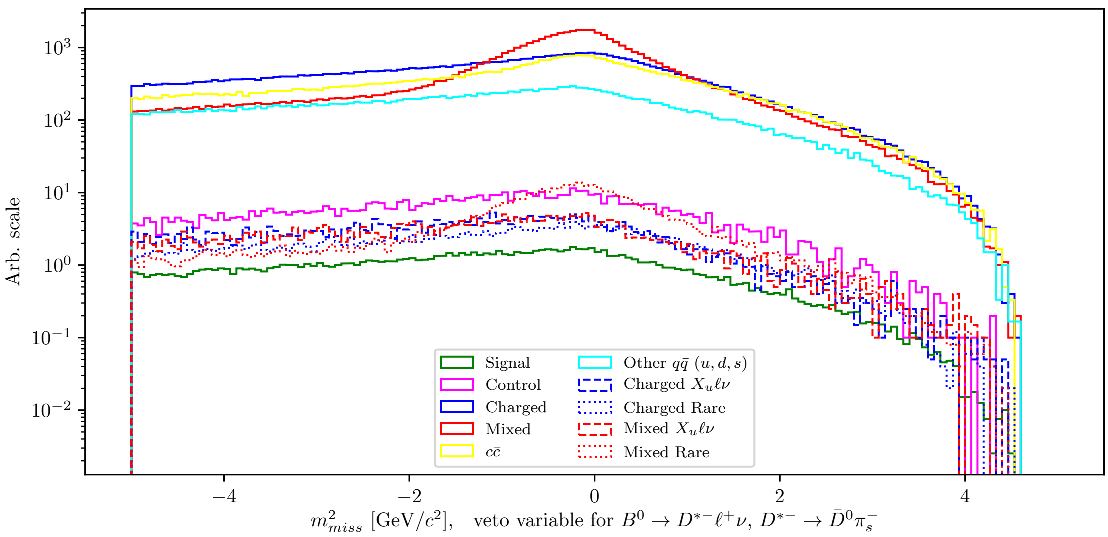



Figure 7.5: Distribution of $m^2_{miss}$ for partially reconstructed $B^0 \to D^{*-}\ell^+\nu$ decays, which serves as a veto.

When the training is finished and the hyper-parameters of the classifier are optimized, the classifier output, as shown in Figure 7.6 (left), can be used for background suppression. $B$ meson candidates from $B\bar{B}$ background are dominant at lower values, while candidates from $B\bar{B}$ events populate the region with higher values. Since the differences between signal and background $B\bar{B}$ events are smaller than $B\bar{B}$ and $q\bar{q}$ events, the resulting classifier has a smaller separation power than in the previous section.

### 7.3.1 Boosting to Uniformity

The selection approach with standard classifiers is optimal for counting experiments, as it by construction produces the optimal selection for observing an excess of signal over background events. Today's BDT algorithms, which work in this way, produce non-uniform selection efficiencies and may, as a consequence, shape background distributions to look like signal. In order to minimize such behavior, it is possible to discard variables, which are correlated with the variable of interest (in our case $\Delta E$ and $M_{BC}$), from the training set. This, however, decreases the classifiers discriminating power. Another approach is to use a novel boosting method, uBoost, which is trained to optimize an integrated FOM under the constraint that the BDT selection efficiency for the desired class must be uniform. The uBoost algorithm balances the biases to produce the optimal uniform selection [7].

The training set used in this training is the same as described at the beginning of this chapter, along with the same set of training variables. It will be seen later that the standard BDT classifier shapes the background to look like signal mostly in the $M_{BC}$ picture, therefore we train the uBDT classifier with a uniformity constraint in the $M_{BC}$ variable for background candidates with the uBoost algorithm. The resulting classifier output is shown in Figure 7.6 (right). For this classifier, the separation power between signal and background seems worse, however, the shapes of backgrounds differ significantly, which greatly affects the performance of signal extraction.

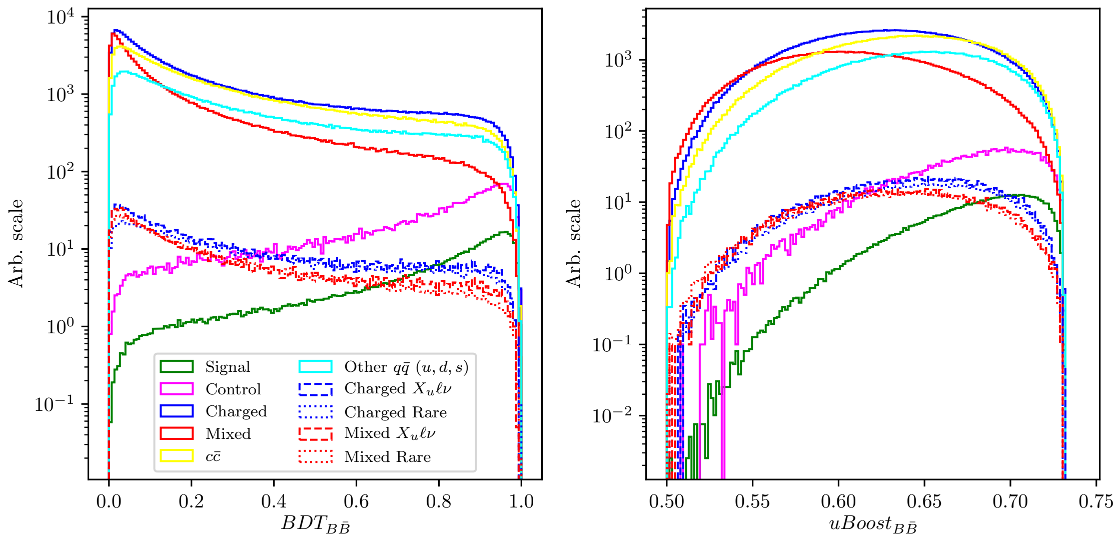



Figure 7.6: $B\bar{B}$ suppression classifier output for signal and various types of background for the standard BDT classifier (left) and the uBDT classifier (right). $B$ candidates from $B\bar{B}$ background events dominate the lower region, while signal and control candidates dominate in the upper region of the classifier output.

## 7.4 Selection Optimization

Instead of two separate $q\bar{q}$ and $B\bar{B}$ FOM optimizations, it is more efficient to do a simultaneous 2D FOM optimization, since the two classifiers are not completely uncorrelated. In the same manner, as before, FOM is optimized for perfectly reconstructed signal candidates in the signal window, after all the pre-cuts, signal categorization, and after cutting out the background resonances and the control decay. The FOM plot with the optimal point for both $B\bar{B}$ MVA classifiers is shown in Figure 7.7.

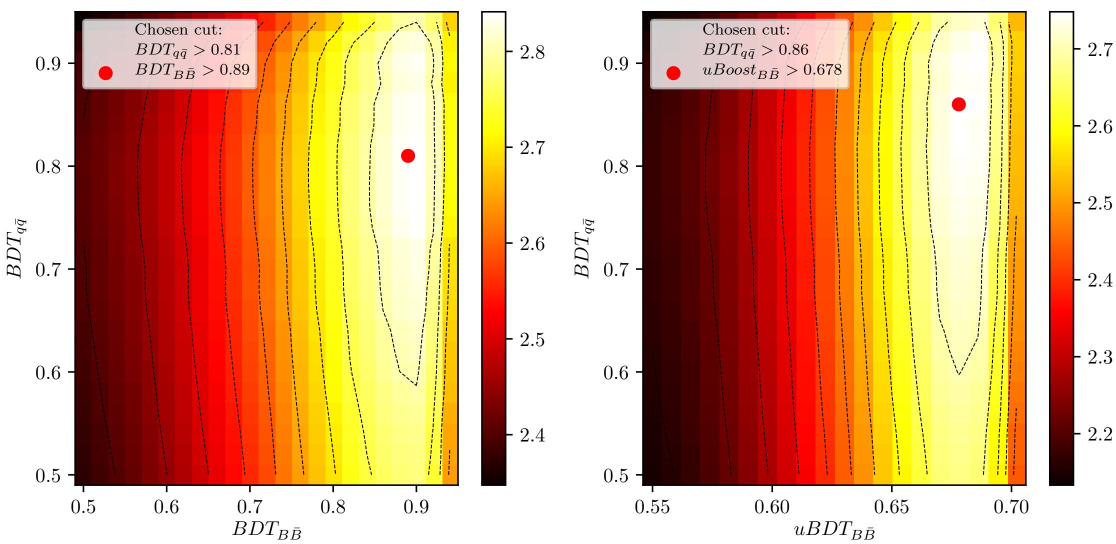



Figure 7.7: 2D FOM optimization of continuum suppression classifier and the standard BDT (left) and uBDT (right) $B\bar{B}$ suppression classifier.

We can compare signal and major background distributions of $\Delta E$ and $M_{BC}$ after the 2D FOM optimization for both classifiers. Figure 7.8 shows the arbitrary (left) and normalized scale (right) for $\Delta E$ (top) and $M_{BC}$ (bottom) for the final sample optimized with the standard BDT classifier, while Figure 7.9 shows similarly for the final sample optimized with uBDT classifier. We can see that there is considerably more background in the latter case, however, also shapes of background and signal distributions differ greatly, meaning there is less room for correlation. The biggest change seems to be in the shape of the $M_{BC}$ distribution, where the background component is much more signal like in the final sample optimized with the standard BDT classifier than in the other case. Additionally, the shapes are more easily constrained in the latter case, since they are present in regions where no signal is expected. The total numbers of expected signal candidates and the signal-to-noise ratios for both classifiers are:

- Standard BDT: $N_{sig} = 176, \quad N_{sig}/N_{bkg} = 4.83\ \%$,
- uBDT: $N_{sig} = 264, \quad N_{sig}/N_{bkg} = 1.33\ \%$.

Due to the large difference in $\Delta E$ and $M_{BC}$ shape, we will continue the analysis with the uBDT classifier, although the comparison between both methods will be shown for the final fit result in the next chapter.

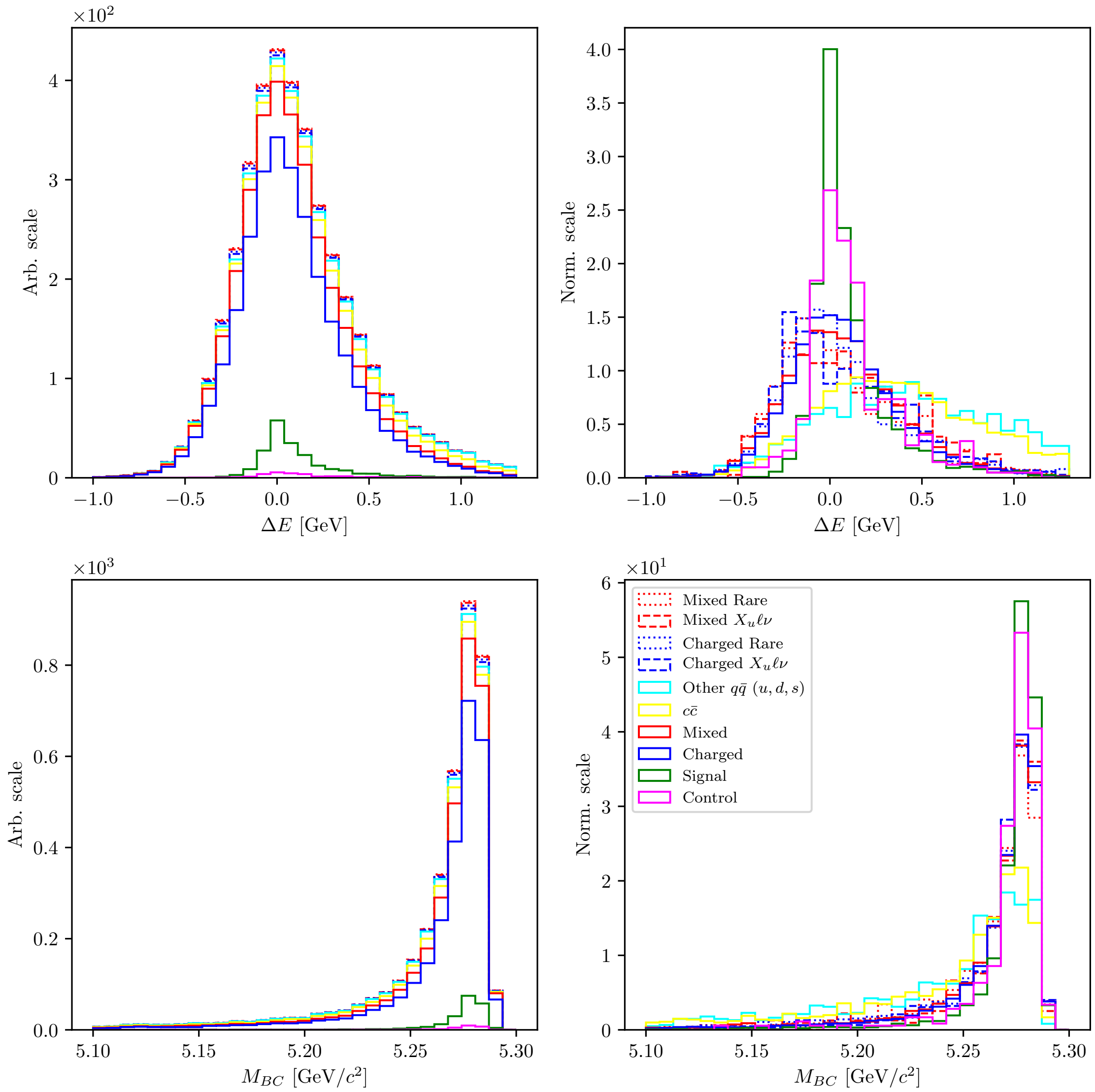



Figure 7.8: Arbitrary (left) and normalized scale (right) for $\Delta E$ (top) and $M_{BC}$ (bottom) for the final sample optimized with the standard BDT classifier.

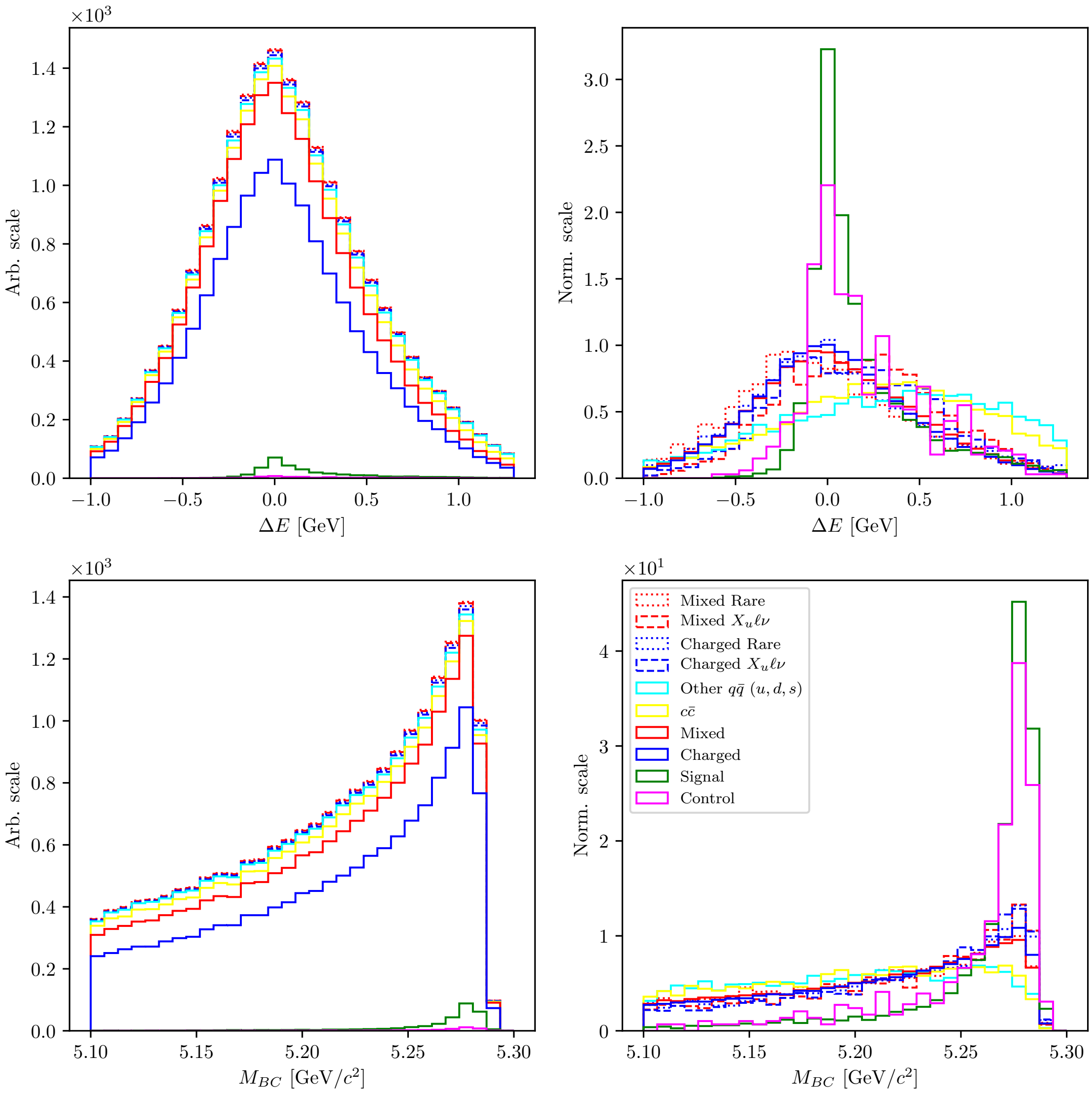



Figure 7.9: Arbitrary (left) and normalized scale (right) for $\Delta E$ (top) and $M_{BC}$ (bottom) for the final sample optimized with the uBDT classifier for $B\bar{B}$ suppression.

### 7.4.1 $B\bar{B}$ Background Composition and Lepton Veto

The majority of background candidates after the final selection is represented by candidates from $B\bar{B}$ events. In order to suppress this background even further, we need to take a look at its structure and recognize various contributions to this part of the background. Figure 7.10 shows $\Delta E$ and $M_{BC}$ for the most significant contributions, along with $m_{KK}$. While most of the candidates come from events where all reconstructed charged particles in the signal decay do not come from one $B$ meson, but both of them, these candidates are not so problematic, because their collaborative distribution is rather smooth and frequent in regions where we expect no signal. On the other hand, there are also contributions from $B$ meson decays which produce more signal-like distributions. We will denote the first kind of background as $\Upsilon(4S)$-matched and the second kind as the $B$-matched $B\bar{B}$ background. Fortunately, these decays are well known and well measured, so their yields can be constrained. Especially problematic is the double semileptonic decay $B \to \bar{D}^{(*)}\ell^{+}\nu,\ D^{+} \to \bar{K}^{-}\ell^{+}\nu$, where the secondary lepton is misidentified as a kaon. Even though the decay has two neutrinos, these events survive the $m^{2}_{miss}$ selection cut and produce peaks at the same positions as signal distributions, while exhibiting only a slightly worse resolution.

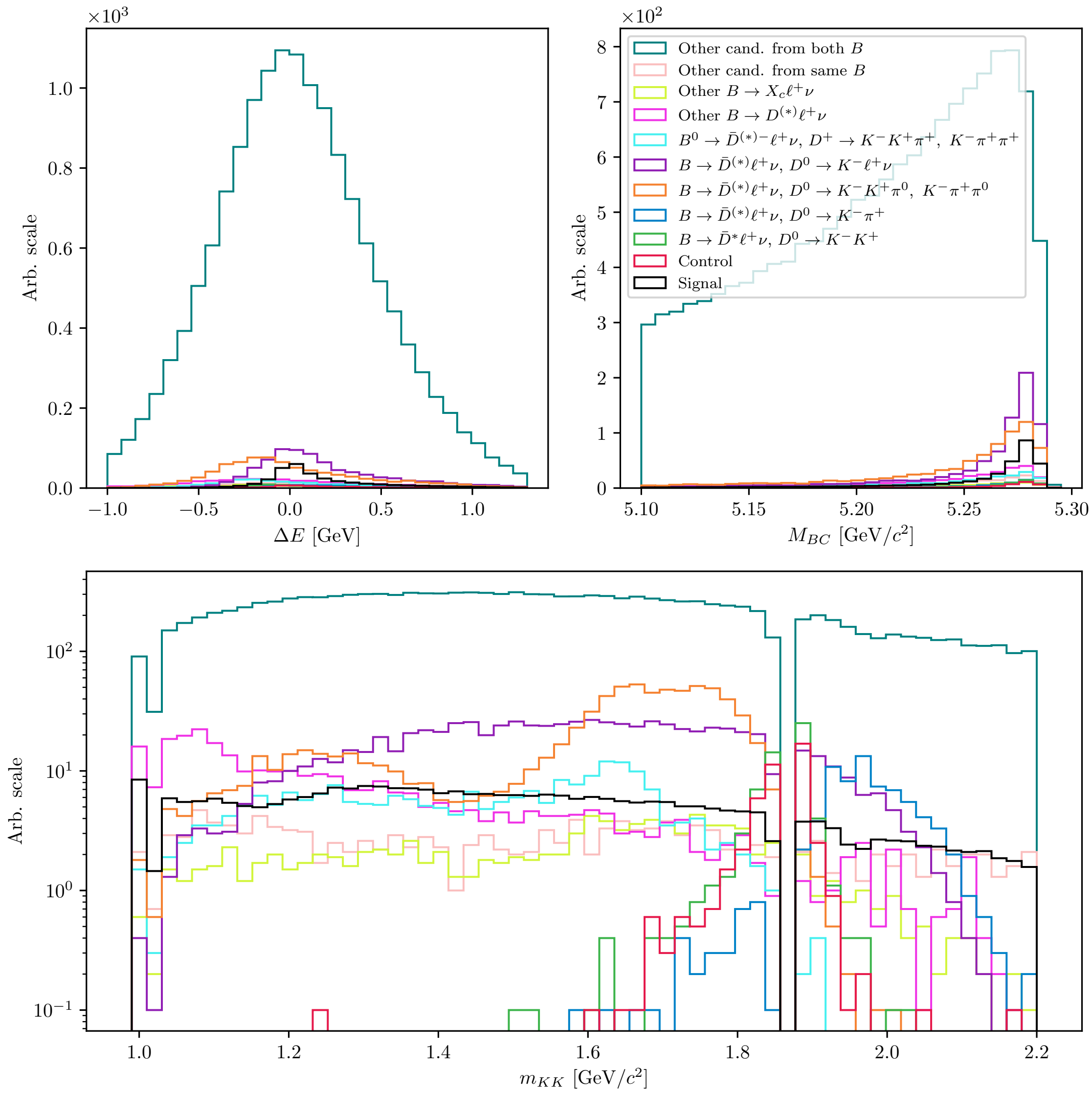



Figure 7.10: $\Delta E$ (left), $M_{BC}$ (right) and $m_{KK}$ (bottom) for major contributions to the $B\bar{B}$ background in the signal cut region. $\Upsilon(4S)$-matched backgrounds represents the majority, but has a smooth and wide background, which has a distinguishable shape from signal. $B$-matched contributions show a peak in $M_{BC}$ and sometimes in $\Delta E$, but can be constrained using existing measurements.

In order to suppress these candidates, a lepton veto can be applied to both kaons, stating that neither of the kaons should exhibit lepton-like properties. On the candidates passing the final selection, we optimize the $eID$ and $\mu ID$ PID cuts, where $S$ and $B$ in Eq. 5.1 are represented by perfect signal candidates and by background candidates, respectively, whereas in the latter case a lepton has been misidentified as a kaon. 2D FOM plots for both kaons are shown in Figure 7.11, where it can be seen that in the majority of the cases, an electron is misidentified as the kaon with the opposite charge to the $B$ meson. With the optimal cuts of

- $K_0: \; eID < 0.7,$

- $K_1: \; eID < 0.1, \; \mu ID < 0.8,$

we reject 77.5% of candidates from the double semileptonic decays, while efficiency loss of signal and other types of $B\bar{B}$ background is about $5-6\%$. The $B\bar{B}$ background after the lepton veto cuts is shown in Figure 7.12 for the signal cut region, and in Figure 7.13 for the control cut region.

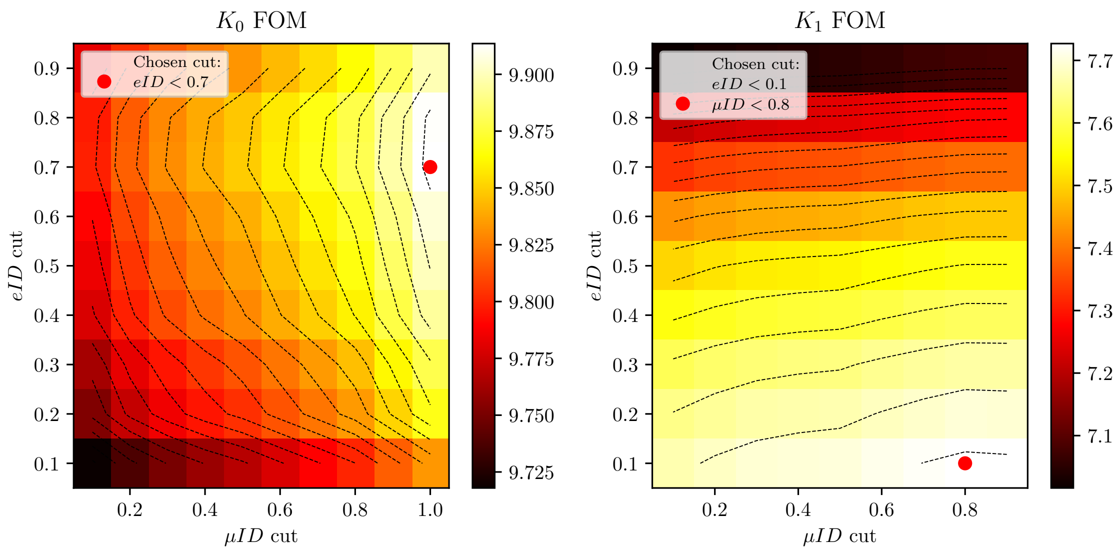



Figure 7.11: 2D FOM for optimal $eID$ and $\mu ID$ cuts on same-sign (left) and opposite-sign (right) kaons with respect to the $B$ meson charge. For double semileptonic background component, in most cases an electron is missidentified as the opposite-sign kaon in the reconstruction chain.

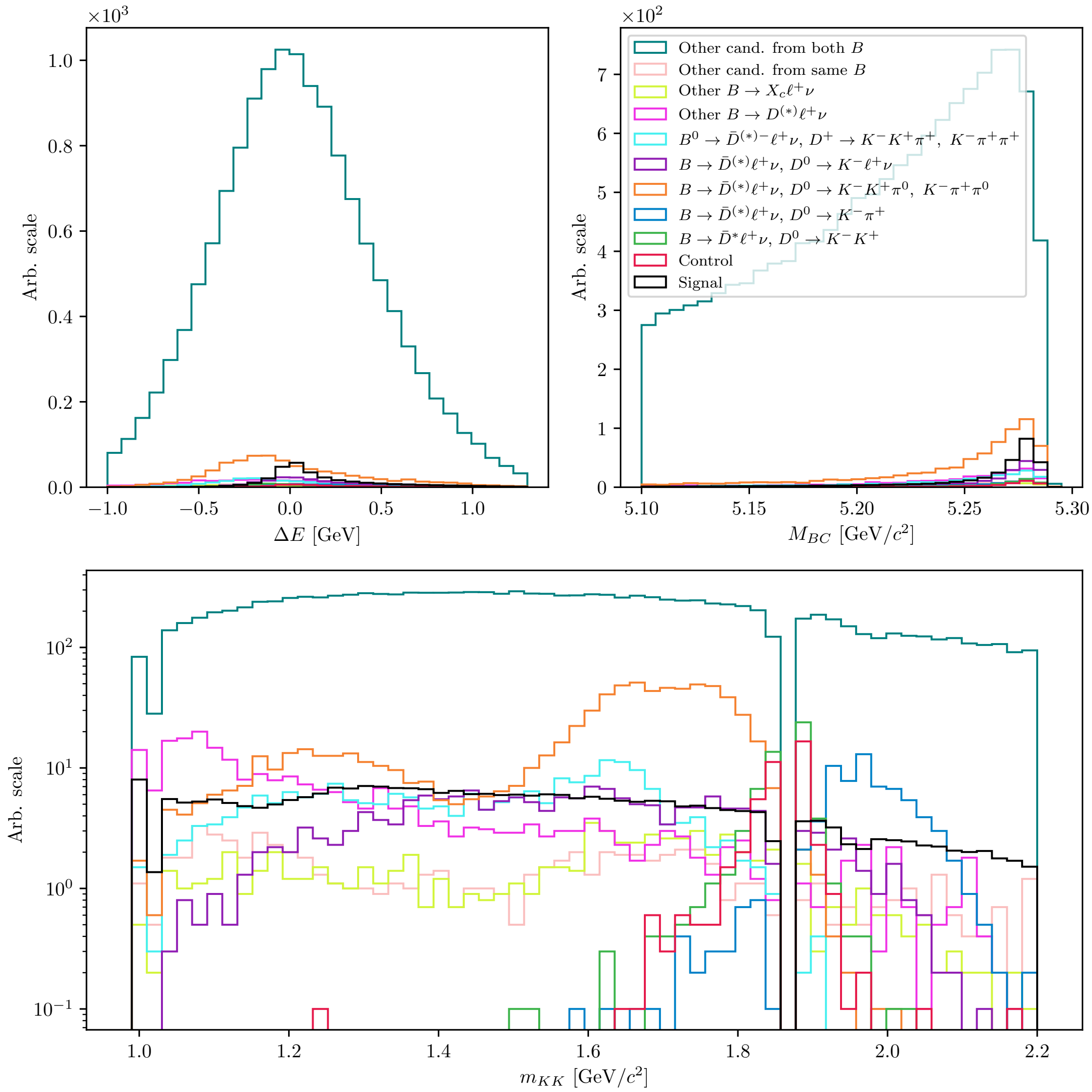



Figure 7.12: $\Delta E$ (left), $M_{BC}$ (right) and $m_{KK}$ (bottom) for major contributions to the $B\bar{B}$ background in the signal cut region after the lepton veto. The double semileptonic background component is suppressed by a factor of $4 - 5$.

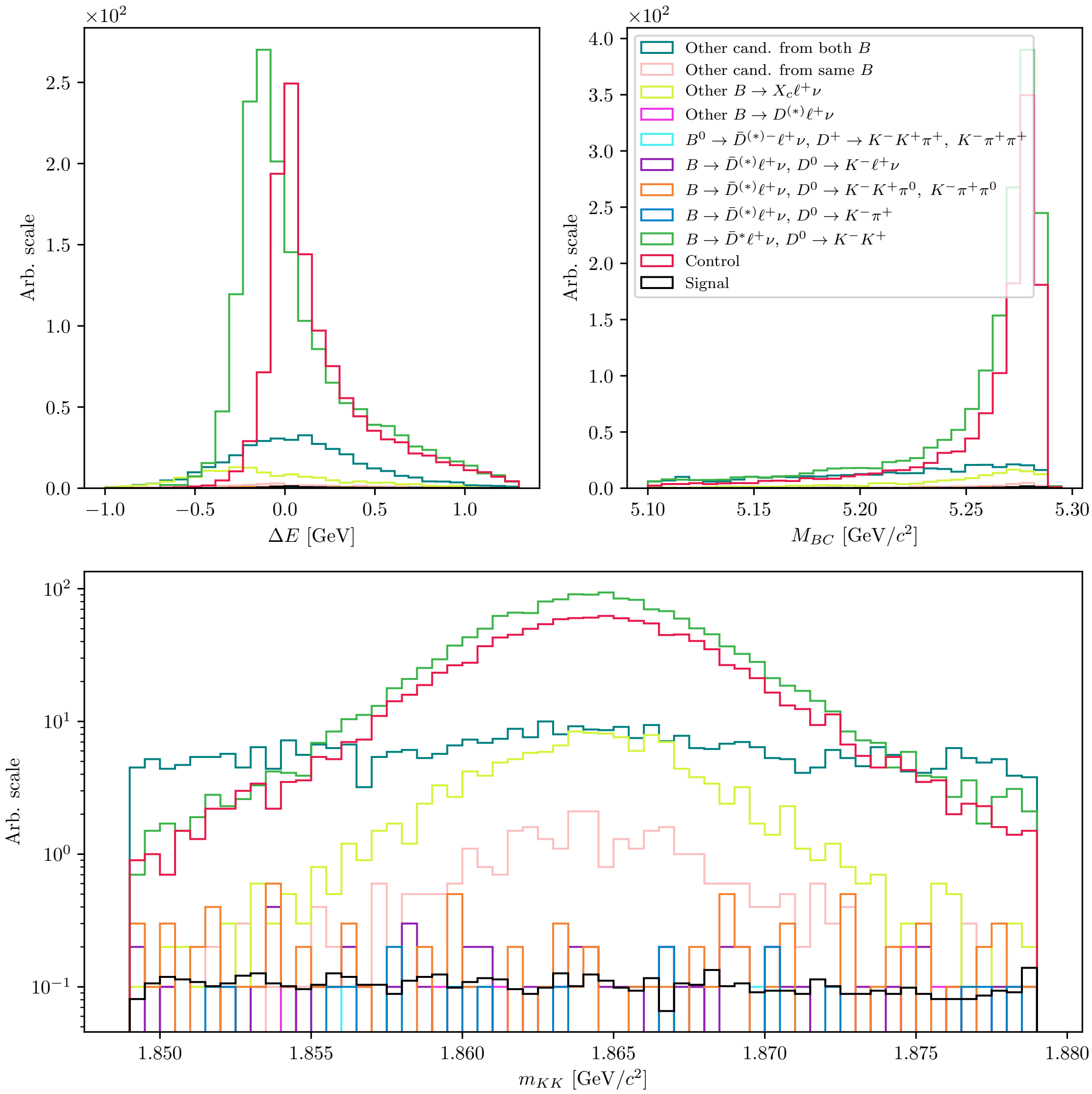



Figure 7.13: $\Delta E$ (left), $M_{BC}$ (right) and $m_{KK}$ (bottom) for major contributions to the $B\bar{B}$ background in the control cut region after the lepton veto. The major component in this case are other $B \to D^*\ell + \nu,\ D \to K^+K^-$ decays, besides the control decay.

# 7.5 Data and MC Agreement

With the final selection in place, we can check the data and MC agreement by checking the control decay region in on- and off- resonance data. Off-resonance samples provide the ability to check the agreement of the $q\bar{q}$ background component, while on-resonance samples can be used to check the validity of the control MC sample and, consequentially, the signal MC sample.

## 7.5.1 Off-resonance Data

The off-resonance data were collected at 60 MeV below the $\Upsilon(4S)$ resonance peak energy in order to determine the non-$B\bar{B}$ background. It, therefore, offers a direct view of the $q\bar{q}$ background data sample, which we can compare to the off-resonance MC sample. Figure 7.14 shows $\Delta E$, $M_{BC}$ and the $q\bar{q}$ classifier output, $BDT_{q\bar{q}}$, for off-resonance data and MC in the control region, before any MVA cuts, where the MC sample was scaled down by a factor of 6, due to 6 streams of MC. Figures show good data and MC agreement of the off-resonance sample. More importantly than the normalization, the shape of data and MC also seems to be the same, so further corrections of $\Delta E$ and $M_{BC}$ on MC are not necessary. This is also consistent by the flatness of the ratio function for $\Delta E$ and $M_{BC}$, shown in the same Figure. There seems to be a difference in the classifier performance for the continuum background suppression on data and MC. This leads to further differences between data and MC in the $q\bar{q}$ sample after the classifier cut, but we estimate that these differences are negligible, since a relatively small amount of continuum background passes the selection, compared to other background types.

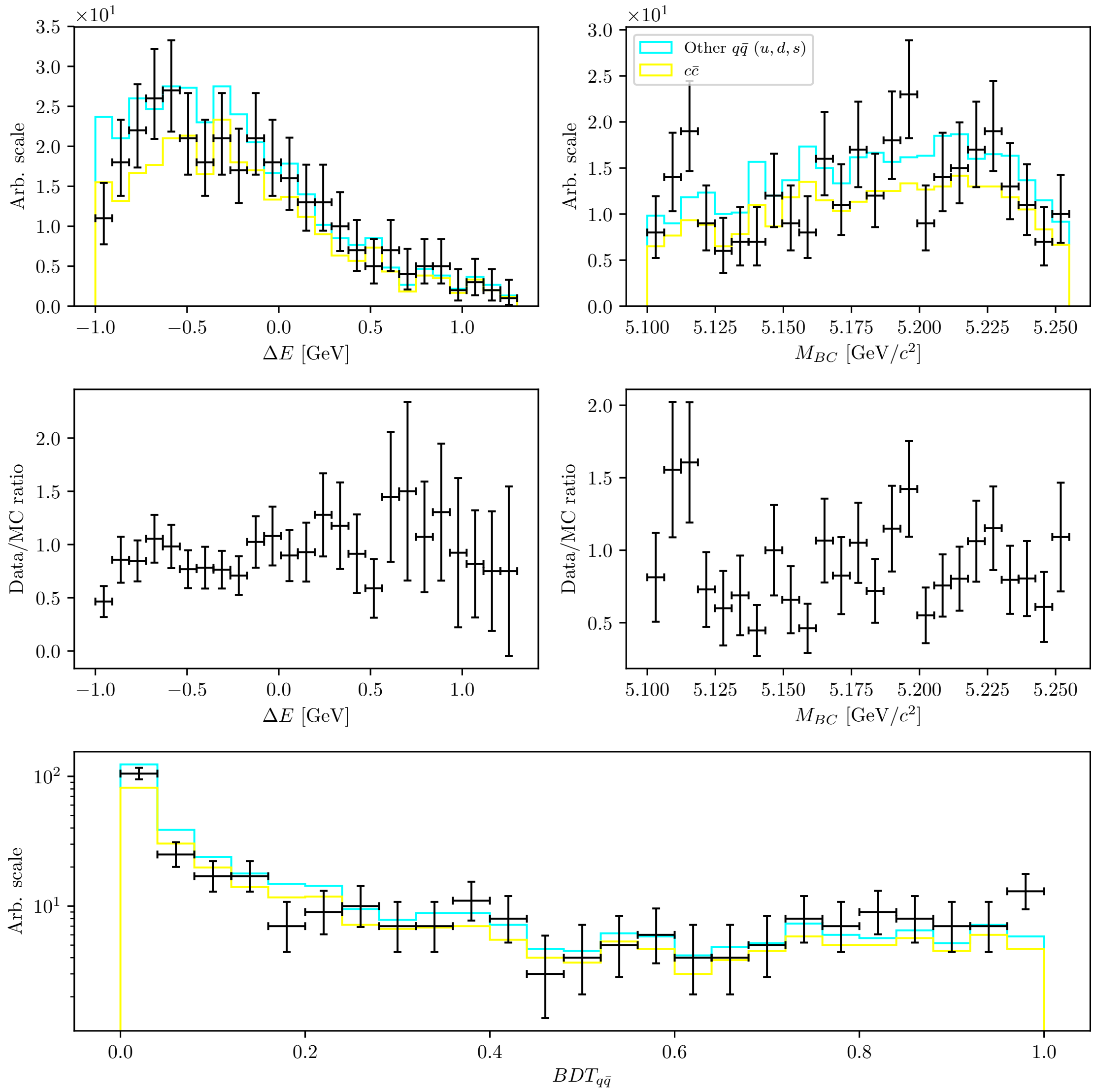



Figure 7.14: $\Delta E$ (left), $M_{BC}$ (right) and the $q\bar{q}$ classifier output (bottom), for off-resonance data and MC in the control region before any MVA cuts.

### 7.5.2 On-resonance Data

We can repeat the check on on-resonance data. Figure 7.15 shows $\Delta E$, $M_{BC}$ and $BDT_{q\bar{q}}$, where one can see inconsistencies between data and MC on the lower spectrum, where continuum background is dominant. We see that MC is over-estimated in this region, most likely due to additional disagreements from other sources. On the other hand, data and MC seem to agree well in the upper part of the spectrum, where $B\bar{B}$ events are dominant. Overall, data and MC seem to agree well already off-the-shelf after all the pre-cuts and without any corrections. This means that the modeling of this MC sample is very precise in this particular region of data and that there are no significant differences between data an MC for the control sample and the signal sample.

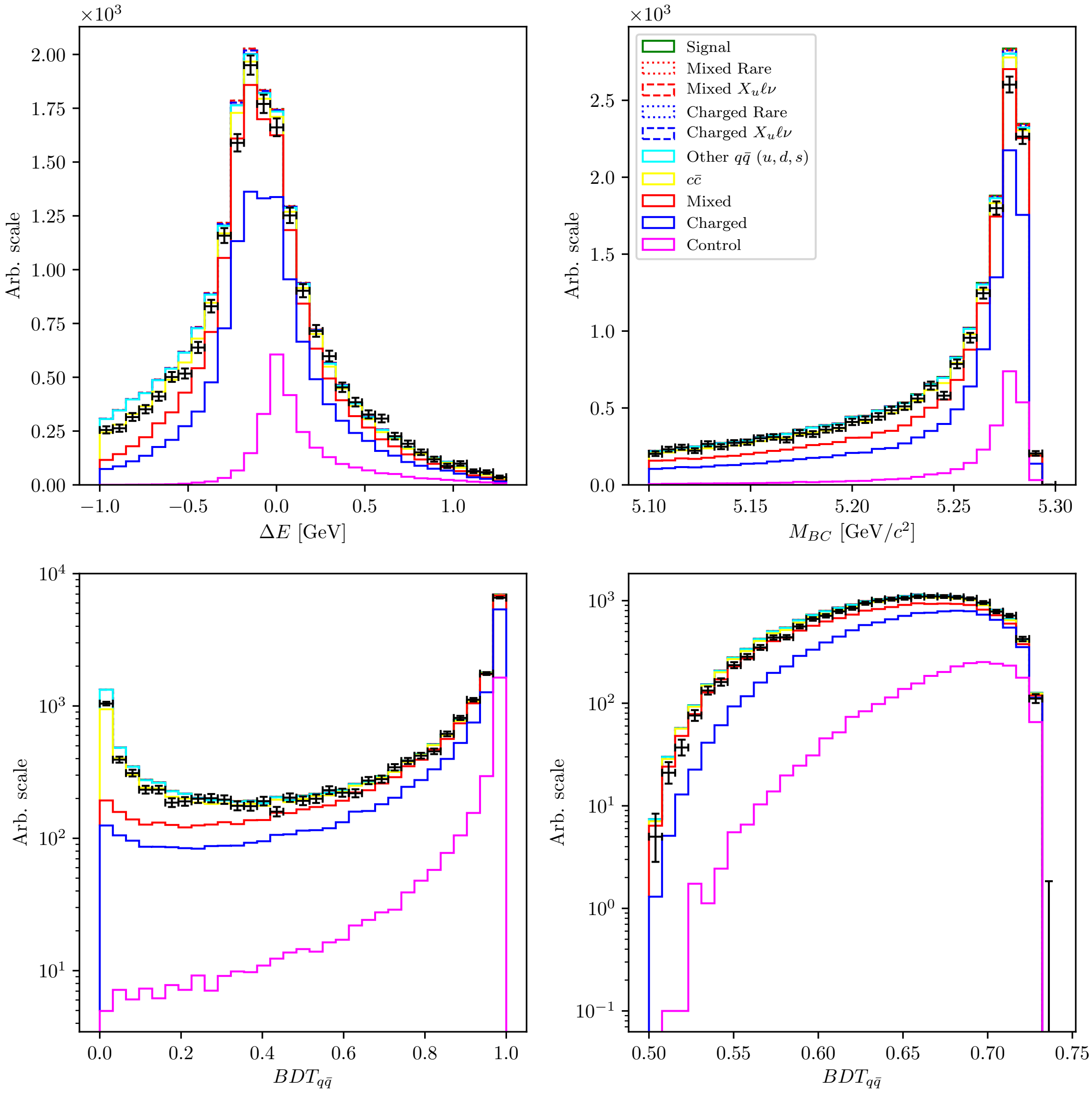



Figure 7.15: $\Delta E$ (left), $M_{BC}$ (right) and the $q\bar{q}$ classifier output (bottom), for on-resonance data and MC in the control region before any MVA cuts.

# Chapter 8

# Signal Extraction

In this chapter, the procedure for signal yield extraction is presented. We use the framework of `RooFit` [8] where we define 2D histogram templates in $\Delta E$ and $M_{BC}$, based on MC, for signal and several types of background. Using these templates, the independent full sample is fitted with the binned extended maximum likelihood (ML) fit, so that the individual template ratios and their sum describe the fitted sample as best as possible. In particle physics we are often dealing with low numbers of events and need to account for the Poisson nature of the data, therefore we use the likelihood fit, since it takes the Poisson errors into account, unlike the $\chi^2$ fit, where the errors are assumed to be Gaussian. In this procedure, we attempt to find the parameter values that maximize the likelihood function, given the observations.

If $P(n|\vec{\alpha})$ is the probability of measuring $n$ candidates, where $\vec{\alpha}$ is a set of parameters on which $P$ depends, we can define the likelihood function $L$ for a series of such measurements (i.e., bins in histogram) $n_i$ based on Poisson statistics as

$$L(\vec{\alpha}) = \prod_{i=1} P(n_i|\vec{\alpha}) = \prod_{i=1} \frac{\mu_i^{n_i} \mathrm{e}^{-\mu_i}}{n_i!}, \tag{8.1}$$

where $\mu_i$ is the expected value for each measurement. It is also common to search for the minimum of the negative value of $\ln L$, or negative log-likelihood (NLL), as

$$\mathcal{L}(\vec{\alpha}) = -\ln L(\vec{\alpha}) = -\sum_i \ln\left(\frac{\mu_i^{n_i} \mathrm{e}^{-\mu_i}}{n_i!}\right) = \sum_i \ln(n_i!) + \mu_i - n_i \ln(\mu_i). \tag{8.2}$$

Maximizing $L$ or minimizing $\mathcal{L}$ gives us a maximum likelihood estimate of the set of parameters $\vec{\alpha}_{ML}$ which best describe the observed data.

The ML method provides a method to estimate the fit uncertainty. This is especially useful if the log-likelihood has a non-parabolic shape, which leads to asymmetric errors. We calculate the errors using the `MINOS` algorithm from the `MINUIT` package [9], which is implemented in RooFit. The algorithm follows the log-likelihood function out of the minimum to find the appropriate intervals of confidence for each parameter, taking the parameter correlations into account.

To estimate the goodness of the likelihood fit, one option is to generate toy MC experiments and obtain the expected log-likelihood distribution. Likelihood fits, however, also offer another way to test the goodness of fit via the likelihood ratio (LR), where we compare the likelihood obtained under the ML parameters $\vec{\alpha}_{ML}$, to the likelihood obtained under the null hypothesis parameters $\vec{\alpha}_{H_0}$. This determines how likely the data is under one model than the other. We define the LR test as

$$\lambda = -2\ln\left(\frac{L(\vec{\alpha}_{ML})}{L(\vec{\alpha}_{H_0})}\right) = -2\left[\ln L(\vec{\alpha}_{ML}) - L(\vec{\alpha}_{H_0})\right] \sim \chi^2_q, \tag{8.3}$$

which asymptotically behaves as the $\chi^2_q$ distribution with $q = m - n$ degrees of freedom, where $m$ and $n$ are degrees of freedom of $L(\vec{\alpha}_{ML})$ and $L(\vec{\alpha}_{H_0})$, respectively. In particle physics we usually study a specific decay and try to perform measurements of the signal yield, so the null hypothesis in this case is that we expect to observe no signal. This means that for the null hypothesis we fix the expected signal yield parameter to zero, while leaving the other parameters of $\vec{\alpha}_{H_0}$ the same as in $\vec{\alpha}_{ML}$, which results in $n = m-1$ degrees of freedom and in their difference $q = m - n = 1$. For such a simple $LR$ test of a single parameter the LR test then follows the $\chi^2$ distribution with 1 degree of freedom. In this case we can define the fit significance from the $\chi^2$ value in units of $\sigma$ as

$$\text{Significance} = \sqrt{\lambda} = \sqrt{\chi^2}. \tag{8.4}$$

## 8.1 Fit Setup

We perform 10 fits to each stream of MC, where 9 streams were used for the creation of the templates and the remaining stream was used as fitted data. When fitting real data, all available MC was used for creating the templates. The full signal MC sample was used for the signal template definition in case of MC as well as data fits. The signal part of the `ulnu` sample was not used in template construction, it was only used as a part of the fitted sample.

The same MC samples are used for template construction as described in chapter 2,

- signal MC,
- 10 streams of `charged` and `mixed` $B\bar{B}$ background,
- 6 streams of $c\bar{c}$ (`charm`) and other $q\bar{q}$ (`uds`) background,
- `ulnu` sample, corresponding to 20× integrated luminosity of the full Belle dataset,
- `rare` sample, corresponding to 50× integrated luminosity of the full Belle dataset.

### 8.1.1 Control Fit

$B\bar{B}$ background composition in control region is shown in Figure 7.13. Due to the strict cut of the $m_{KK}$ around the $D^0$ mass window, most of the decays with a $D^0$ proceed via $D^0 \to K^+K^-$ decay. In this case, the following fit templates are chosen

- signal template,
- $q\bar{q}$ template,

- $C_0: \ B^+ \to \bar{D}^0 \ell^+ \nu, \ D^0 \to K^- K^+$ (control decay),
- $C_1: \ B \to \bar{D}^* \ell^+ \nu, \ D^0 \to K^- K^+$,
- other $B\bar{B}$ BKG template.

In control fits, all template shapes are fixed and the yields of all templates are floated, except for the signal template, which is expected to be very close to zero and is fixed to the expected MC yield. Additionally, since the $C_0$ and $C_1$ decays are well known and measured, we make use of this fact in the form of a ratio $N_{C_1}/N_{C_0}$, which is fixed to the MC value in case of the MC fit and constrained to the measured value in case of fits to real data. The ratio is implemented based on the decay channels shown in Table 8.1 and is defined as

$$r_1 = \frac{\left(\sum_j N_{1j} \times \rho_{1j}\right)}{N_{00} \times \rho_{00}}, \tag{8.5}$$

where $j$ runs over all channels in the category $C_1$ and where $\rho_{ij}$ is the branching ratio correction factor for the specific channel $N_{ij}$, which incorporates information from world measurements. It is defined as

$$\rho = \frac{\mathcal{B}^{PDG}}{\mathcal{B}^{GEN}}, \tag{8.6}$$

where $\mathcal{B}^{PDG}$ is the measured branching ratio and $\mathcal{B}^{GEN}$ is the branching ratio value used in MC generation. The branching ratio correction factor has been implemented due to differences between measured and MC branching ratio values. Each branching ratio measurement serves as a constraint used in the fit. All branching ratio constraints in the control fit are shown in Table 8.2. The measured values are cited only for the $B^0$ decay mode, where isospin symmetry has been assumed. The corresponding $B^+$ branching ratios can be calculated as

$$\mathcal{B}(B^+) = \mathcal{B}(B^0) \times \tau_{B^+/B^0}, \tag{8.7}$$

where $\tau_{B^+/B^0}$ is the ratio of $B$-meson decay times, which is measured to be [13]

$$\tau_{B^+/B^0} = 1.076 \pm 0.004. \tag{8.8}$$

| Category | Channel | $B$ Decay Mode | $D$ Decay Mode | $N_{MC}$ |
|---|---|---|---|---|
| $C_0$ | $N_{00}$ | $B^+ \to \bar{D}^0 \ell^+ \nu$ | $D^0 \to K^- K^+$ | $1182 \pm 34$ |
| $C_1$ | $N_{10}$ | $B^+ \to \bar{D}^{*0} \ell^+ \nu$ | $D^0 \to K^- K^+$ | $1455 \pm 38$ |
| | $N_{11}$ | $B^0 \to D^{*-} \ell^+ \nu$ | $D^0 \to K^- K^+$ | $186 \pm 16$ |

Table 8.1: Well defined decay channels used for constraining the control fits.

In case of MC fits, the fitted sample is also generated with MC, so $\mathcal{B}_i^{PDG} = \mathcal{B}_i^{GEN}$ and Eq. (8.5) simplifies to a simple MC yield ratio. On fits to real data, expected MC yields and branching ratio measurements are implemented as independent Gaussian constraints in order to properly account for correlations in Eq. (8.5).

| $\mathcal{B}$ ID | Decay | $\mathcal{B}_{GEN}$ | $\mathcal{B}_{PDG}$ | $\rho$ | Reference |
|---|---|---|---|---|---|
| 0 | $B^0 \to D^- \ell^+ \nu$ | $2.13 \times 10^{-2}$ | $(2.13 \pm 0.09) \times 10^{-2}$ | $1.00 \pm 0.04$ | [13] |
| 1 | $B^0 \to D^{*-} \ell^+ \nu$ | $5.33 \times 10^{-2}$ | $(4.88 \pm 0.11) \times 10^{-2}$ | $0.92 \pm 0.02$ | |
| 2 | $B^+ \to \bar{D}^0 \ell^+ \nu$ | $2.31 \times 10^{-2}$ | $(2.29 \pm 0.10) \times 10^{-2}$ | $0.99 \pm 0.04$ | [calc.] |
| 3 | $B^+ \to D^{*0} \ell^+ \nu$ | $5.79 \times 10^{-2}$ | $(5.25 \pm 0.12) \times 10^{-2}$ | $0.91 \pm 0.02$ | |
| 4 | $D^0 \to K^- K^+$ | $3.90 \times 10^{-3}$ | $(3.97 \pm 0.07) \times 10^{-3}$ | $1.02 \pm 0.02$ | [15] |

Table 8.2: MC and measured values of branching ratios along with the calculated correction factors used for constraining the control fit.

### Smearing and Offset Parameters

With simulated data, we are able to perform detailed studies prior to looking at the measured data. However, simulated data often does not describe real data perfectly. Out of variables $\Delta E$ and $M_{BC}$, $\Delta E$ is especially prone to a lack of precision in energy measurements. This can introduce either overestimation of resolution on MC as well as a possible shift in the measured energy in either direction. Due to this fact, we perform a scan over two additional parameters of offset and smearing, applied on the $\Delta E$ variable. In case of the $M_{BC}$ variable the mentioned effects are not as prominent, so the smearing and offset for the latter variable are omitted.

The following parameter phase space is scanned

- smearing factor in range $[0.0, 0.08]$ GeV in steps of $8 \times 10^{-3}$,
- offset in range $[0.0, 0.003]$ GeV in steps of $1.5 \times 10^{-4}$,

where for each parameter pair the likelihood ratio test is performed to estimate the goodness of the fit. Figure 8.1 shows the contour plot of the likelihood ratio $\lambda$, as defined in Eq. (8.3), for 2 d.o.f., for MC (left) and data (right). The scan over MC serves the purpose of a consistency check, where we expect the best fit to occur in the phase space where neither smearing nor offset are applied. In the case of data, we see that we obtain a better fit by introducing some level of smearing and offset. In both cases, the two parameters have shown no sign of significant correlation, so we treat them independently. The likelihood ratio test allows us to estimate the parameter values in the $1\sigma$ confidence interval, where we obtain the optimal parameter set

- Smearing: $40^{+15}_{-17}$ MeV,
- Offset: $6^{+4.6}_{-6}$ MeV.

We apply this transformation to our MC samples in all cases when fitting real data.

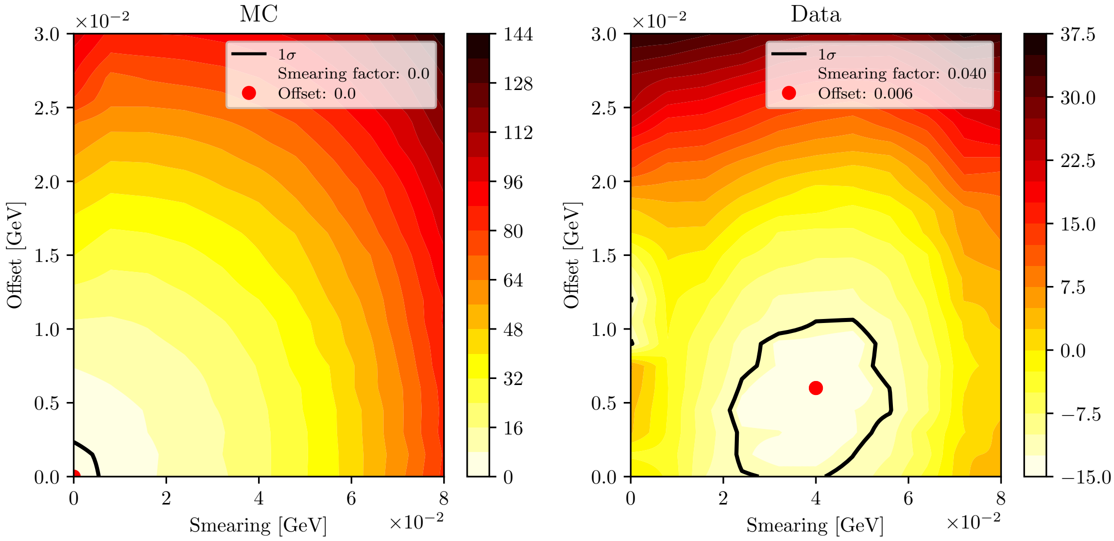



Figure 8.1: Likelihood ratio test of an additional smearing and offset parameter to MC (left) and data (right).

### 8.1.2 Signal Fit

The motivation for the choice of signal fit templates comes from Figure 7.12. The following histogram templates were defined

- signal template,
- $q\bar{q}$ template,
- a series of well defined templates from $B\bar{B}$ background:
  - $C_0 : \; B^+ \to \bar{D}^0\ell^+\nu, \; D^0 \to K^-K^+$ (control decay),
  - $C_1 : \; B \to \bar{D}^*\ell^+\nu, \; D^0 \to K^-K^+$,
  - $C_2 : \; B \to \bar{D}^{(*)}\ell^+\nu, \; D^0 \to K^-\pi^+$,
  - $C_3 : \; B \to \bar{D}^{(*)}\ell^+\nu, \; D^0 \to K^-K^+\pi^0, \; K^-\pi^+\pi^0$,
  - $C_4 : \; B \to \bar{D}(*)\ell^+\nu, \; D^0 \to K^-\ell^+\nu$,
  - $C_5 : \; B^0 \to D^{(*)-}\ell^+\nu, \; D^+ \to K^-K^+\pi^+, \; K^-\pi^+\pi^+$,
  - $C_6 :$ other $B \to \bar{D}^{(*)}\ell^+\nu$ decays,
- remaining $B\bar{B}$ background template.

As mentioned in chapter 7, the majority of the background comes from $B\bar{B}$ events. Various processes ($C_0$ to $C_6$) contribute to this background which are well known and measured, so we make use of these measurements by fixing their yields in MC fits and appropriately constraining them in real data fits. The remaining $B\bar{B}$ background is merged into a single template. In this case, the shape of all templates is fixed as well, while the yields are floated for all templates except for the constrained background templates. The yield constraints are based on the channels shown in Table 8.3 and defined for each template category as

$$Y_i = \eta_{\text{norm.}} \times \frac{\left(\sum_j N_{ij} \times \rho_{ij}\right)}{\rho_{00}}, \tag{8.9}$$

where $j$ runs over all channels in the category $C_i$ and where $\rho_{ij}$ is defined in Eq. (8.6). The first factor, $\eta_{\text{norm.}}$, serves as a normalization factor in order to scale the number of generated $B\bar{B}$ events to the number of $B\bar{B}$ events in measured data. We define it as

$$\eta_{\text{norm.}} = \frac{N^{D}_{\text{control}}}{N^{MC}_{\text{control}}}, \tag{8.10}$$

where $N^{D}_{\text{control}}$ and $N^{MC}_{\text{control}}$ are control yields in the control fit for data and MC, respectively.

In addition to branching ratio constraints in Table 8.2, further constraints are defined in Table 8.4 due to more decay channels. In case of the category $C_6$, we have no firm handle on the $D$ meson decay, therefore no correction for this branching ratio can be introduced, so we set a correction factor of 1 with a 100% error for the $D$ meson decay branching ratio. As most of the correction factors used for constraints have deviations (including the errors) from nominal values well below 100% this value is very conservative.

| Category | Channel | $B$ Decay Mode | $D$ Decay Mode | Expected MC Yield |
|---|---|---|---|---|
| $C_0$ | $N_{00}$ | $B^+ \to \bar{D}^0\ell^+\nu$ | $D^0 \to K^-K^+$ | $44 \pm 7$ |
| $C_1$ | $N_{10}$ | $B^+ \to \bar{D}^{*0}\ell^+\nu$ | $D^0 \to K^-K^+$ | $53 \pm 7$ |
| | $N_{11}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^0 \to K^-K^+$ | $6 \pm 2$ |
| $C_2$ | $N_{20}$ | $B^+ \to \bar{D}^0\ell^+\nu$ | $D^0 \to K^-\pi^+$ | $23 \pm 5$ |
| | $N_{21}$ | $B^+ \to D^{*0}\ell^+\nu$ | $D^0 \to K^-\pi^+$ | $41 \pm 6$ |
| | $N_{22}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^0 \to K^-\pi^+$ | $6 \pm 2$ |
| $C_3$ | $N_{30}$ | $B^+ \to \bar{D}^0\ell^+\nu$ | $D^0 \to K^-K^+\pi^0$ | $103 \pm 10$ |
| | $N_{31}$ | $B^+ \to D^0\ell^+\nu$ | $D^0 \to K^-\pi^+\pi^0$ | $211 \pm 15$ |
| | $N_{32}$ | $B^+ \to D^{*0}\ell^+\nu$ | $D^0 \to K^-K^+\pi^0$ | $135 \pm 12$ |
| | $N_{33}$ | $B^+ \to D^{*0}\ell^+\nu$ | $D^0 \to K^-\pi^+\pi^0$ | $267 \pm 16$ |
| | $N_{34}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^0 \to K^-K^+\pi^0$ | $19 \pm 4$ |
| | $N_{35}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^0 \to K^-\pi^+\pi^0$ | $36 \pm 6$ |
| $C_4$ | $N_{40}$ | $B^+ \to \bar{D}^0\ell^+\nu$ | $D^0 \to K^-e^+\nu$ | $48 \pm 7$ |
| | $N_{41}$ | $B^+ \to D^0\ell^+\nu$ | $D^0 \to K^-\mu^+\nu$ | $7 \pm 3$ |
| | $N_{42}$ | $B^+ \to \bar{D}^{*0}\ell^+\nu$ | $D^0 \to K^-e^+\nu$ | $98 \pm 10$ |
| | $N_{43}$ | $B^+ \to \bar{D}^{*0}\ell^+\nu$ | $D^0 \to K^-\mu^+\nu$ | $10 \pm 3$ |
| | $N_{44}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^0 \to K^-e^+\nu$ | $14 \pm 4$ |
| | $N_{45}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^0 \to K^-\mu^+\nu$ | $3 \pm 2$ |
| $C_5$ | $N_{50}$ | $B^0 \to D^-\ell^+\nu$ | $D^+ \to K^-K^+\pi^+$ | $103 \pm 10$ |
| | $N_{51}$ | $B^0 \to D^-\ell^+\nu$ | $D^+ \to K^-\pi^+\pi^+$ | $63 \pm 8$ |
| | $N_{52}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^+ \to K^-K^+\pi^+$ | $31 \pm 6$ |
| | $N_{53}$ | $B^0 \to D^{*-}\ell^+\nu$ | $D^+ \to K^-\pi^+\pi^+$ | $21 \pm 5$ |
| $C_6$ | $N_{60}$ | $B^+ \to \bar{D}^0\ell^+\nu$ | Other $D^0$ and $D^+$ decays | $69 \pm 8$ |
| | $N_{61}$ | $B^+ \to \bar{D}^{*0}\ell^+\nu$ | | $95 \pm 10$ |
| | $N_{62}$ | $B^0 \to D^-\ell^+\nu$ | | $63 \pm 8$ |
| | $N_{63}$ | $B^0 \to D^{*-}\ell^+\nu$ | | $36 \pm 6$ |

Table 8.3

| $\mathcal{B}$ ID | Decay | $\mathcal{B}_{GEN}$ | $\mathcal{B}_{PDG}$ | $\rho$ | Reference |
|---|---|---|---|---|---|
| 5 | $D^0 \to K^-\pi^+$ | $3.82 \times 10^{-2}$ | $(3.89 \pm 0.04) \times 10^{-2}$ | $1.02 \pm 0.01$ | |
| 6 | $D^0 \to K^-K^+\pi^0$ | $2.36 \times 10^{-3}$ | $(3.37 \pm 0.15) \times 10^{-3}$ | $1.43 \pm 0.06$ | |
| 7 | $D^0 \to K^-\pi^+\pi^0$ | $13.08 \times 10^{-2}$ | $(14.2 \pm 0.5) \times 10^{-2}$ | $1.09 \pm 0.04$ | |
| 8 | $D^0 \to K^-e^+\nu$ | $3.41 \times 10^{-2}$ | $(3.53 \pm 0.028) \times 10^{-2}$ | $1.04 \pm 0.01$ | [15] |
| 9 | $D^0 \to K^-\mu^+\nu$ | $3.41 \times 10^{-2}$ | $(3.31 \pm 0.13) \times 10^{-2}$ | $0.97 \pm 0.04$ | |
| 10 | $D^+ \to K^-K^+\pi^+$ | $9.06 \times 10^{-3}$ | $(9.51 \pm 0.34) \times 10^{-3}$ | $1.05 \pm 0.04$ | |
| 11 | $D^+ \to K^-\pi^+\pi^+$ | $9.51 \times 10^{-2}$ | $(8.98 \pm 0.28) \times 10^{-2}$ | $0.94 \pm 0.03$ | |

Table 8.4: Additional MC and measured values of $D$ meson branching ratios along with the calculated correction factors used for constraining the signal fit.

## 8.2 Adaptive Binning Algorithm

The fit templates contain areas of low statistics, which are populated with bins with zero content. This is a direct consequence of having a finite MC sample and represent a liability in ML fits. Due to low statistics in the edge regions, the locations of these empty bins can vary for the templates and the fitted sample. A problem occurs if all templates have an empty bin where the fitted sample does not. In the scope of ML fits, this effectively means that there are entries in bins where the probability of having them is 0. We will call such bins *problematic* because in these cases the fit does not converge.

The ideal solution for this problem would be to increase the MC statistics. Since this is not an option, we pursue other solutions, such as decreasing the number of bins. While this solves the problem, the drawback of it is a decrease in the template resolution in densely populated regions, where good resolution is most needed. The optimal solution here seems to be a choice of variable bins, with fine binning in the densely populated regions and larger bins in the regions with low statistics.

We have devised an algorithm, which compares the templates and the fitted sample, and defines a variable binning so that there are no more problematic bins in the end. Figure 8.2 shows an example of how the procedure works. The algorithm takes an argument for the initial number of uniform bins in each dimension and does the following

1. define uniform binning in both dimensions with the provided argument,
2. create a 2D histogram from MC templates with expected yields,
3. define an *optimal* region, where most of the 2D integral is contained and where all bins have a non-zero content (this region does not change throughout the process),
4. compare the histograms for the expected and the fitted sample, find the problematic bins,
5. loop until all problematic bins disappear
   (a) find the problematic bin, which is nearest to the maximum bin,
   (b) change the binning from $N$ to $N-1$ from that bin and in the direction away from the maximum bin.

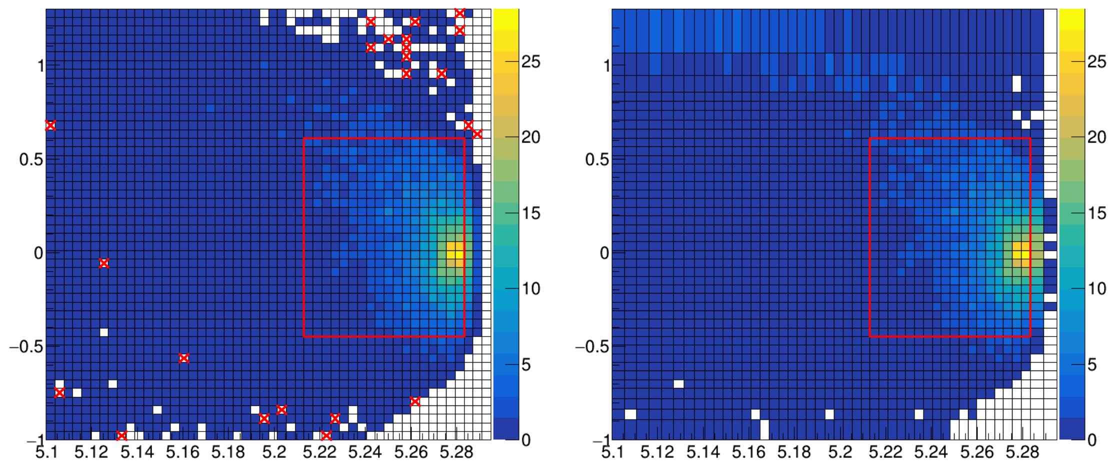



Figure 8.2: Steps taken in the adaptive binning algorithm. Left image shows the initial 2D histogram with the defined optimal region and the problematic bins, the right image shows the final binning with the unchanged optimal region, while the problematic bins are gone due to the new binning choice.

An additional problem occurs in the plotting of the fitted templates and the sample with variable binning. It would seem that RooFit does not take the bin widths into account when plotting, while everything works as expected for the fit itself. This was bypassed by extracting the fitted yields and applying them to templates and samples with uniform binning, which were then used for drawing.

## 8.3 Toy MC Experiments

For the chosen final selection and fit procedure, toy MC pseudo-experiments were performed in order to confirm the behavior of the fit setup. The fit behavior is also checked as a function of the signal yield in the form of a linearity test. A detailed description of toy MC experiments is written in this section.

With toy MC experiments in this section, we study the yields, errors and the pulls of the signal fit by generating our own pseudo data sets, according to the MC. This significantly reduces the time we would need to produce the data sets in the standard way, while still reliably describing the underlying physics behind the pseudo data set. All available MC was used for pseudo data set generation as well as creating templates.

### 8.3.1 Pseudo Experiment: Expected Signal Yield

We constructed $3 \times 10^3$ pseudo data sets, where each data set was generated with the expected amount of each template category, distributed according to the Poisson distribution. All fits were performed with the optimal initial uniform binning of $19 \times 19$ bins in $\Delta E$ and $M_{BC}$.

Figure 8.3 shows distributions of the fit yields, errors and the pull distribution of all pseudo fits. The fits seem to be under control, although there is a slight bias present in the negative direction, which can also be seen in the pull distribution plot. The latter follows a normal distribution with a mean of $(-0.11 \pm 0.02)$ and standard deviation of $(1.01 \pm 0.01)$. The mean $(\bar{X})$ and the standard deviation $(S)$ were calculated in the usual way, while their errors $\sigma_{\bar{X}}$ and $\sigma_S$ were calculated as [10]

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{N}}, \quad \sigma_S = \frac{S}{\sqrt{2(N-1)}}, \tag{8.11}$$

where $N$ is the number of performed measurements.

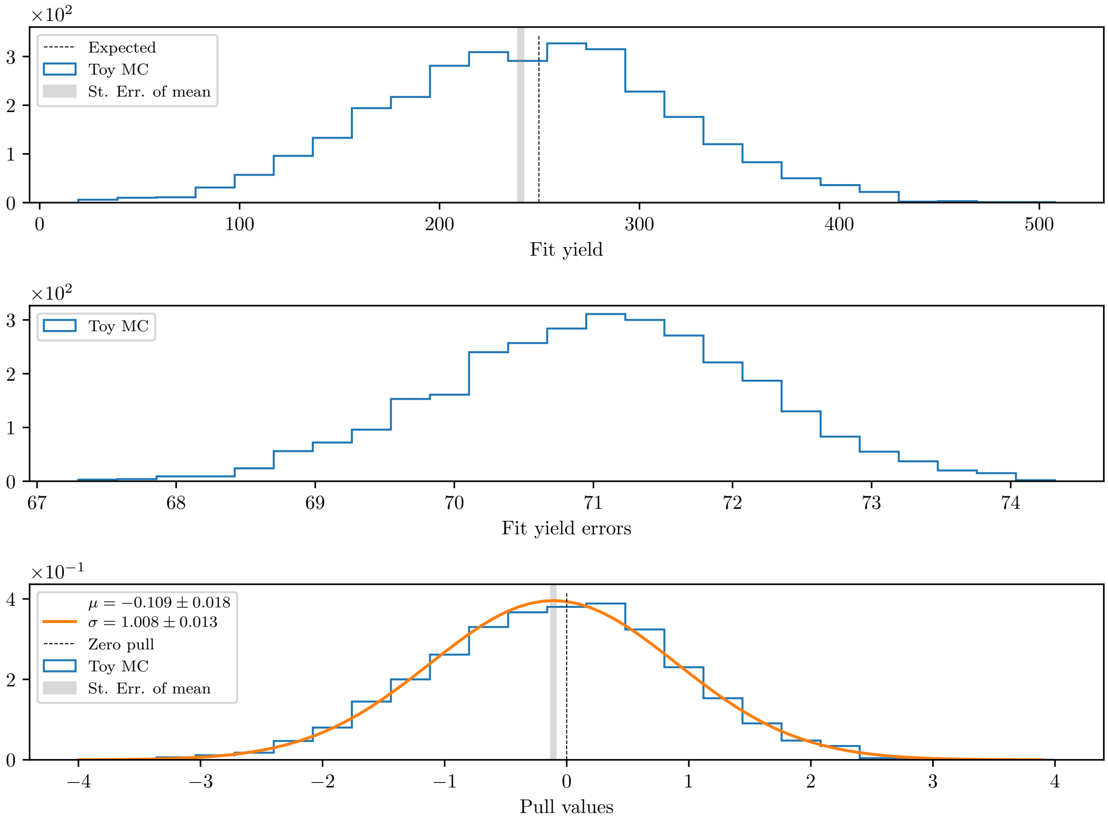



Figure 8.3: Toy MC fits of pseudo data showing the fit yield (top), fit errors (center) and the pull distribution of the fits (bottom).

### 8.3.2 Pseudo Experiment: Linearity Test

Linearity test is used for determining the sensitivity of the fit to the amount of signal in the fitted sample. Since this is the first measurement of this decay channel, MC modeling is not reliable and could be very different from reality, so we need to perform this test in order to determine our sensitivity to smaller, as well as larger amounts of expected signal.

The pseudo data sets are generated in the same way as in the previous subsection, with the exception of signal, which is generated in various amounts. 50 steps from $[0.1,\ 10]$ in the logarithmic scale are taken for fractions of signal amount and for each fraction we generate 500 pseudo data sets according to Poisson statistics.

Figure 8.4 shows the difference between mean fit yield and the expected yield, mean pull and the mean significance at each signal fraction value. The expected

MC result lies at the fraction value $10^0 = 1$. The plots show no significant bias with respect to the signal fraction, while the pulls seem to be described by the normal distributions throughout the fraction range. At expected value, we are at about $3.58\sigma$ significance.

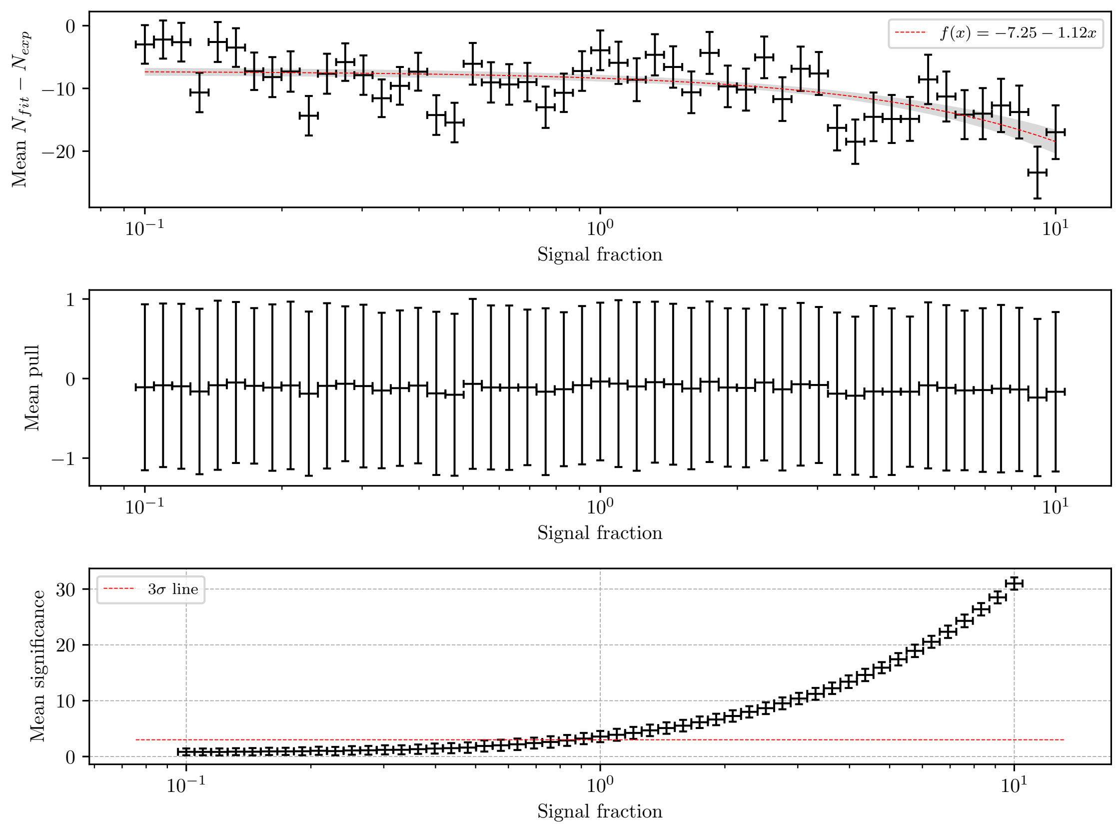



Figure 8.4: Mean fit yield and expected yield difference (top), mean pull (center) and mean significance (bottom) as a function of signal fraction.

## 8.4 Fit Results

In this section, we present the first results of signal and control fits on MC as well as data, along with the control decay branching ratio measurement. We also show results of the signal fit on MC and data.

### 8.4.1 Signal MC Fit Results

With the signal fit setup described in section 8.1.2, we proceed to fit the 10 streams of MC. To compare both methods of $B\bar{B}$ bar suppression, two different samples were prepared and used in the fit. Since the choice of initial uniform binning is not obvious, we perform fits to all streams of MC for each initial binning choice in the range $N \times N,\ N \in [4, 30]$. Figure 8.5 shows the expected yield differences, pulls and fit significances for both final samples for each binning case. The difference between fitted an expected signal yield should be equal to 0 to ensure no bias is present in the fit, while the average pull distribution for each bin case should have a central value at 0, with a width of 1. While both fit results seem to have no significant bias with respect to the binning choice, the pull distribution seems to be closer to the normal distribution. The uBDT classifier fit setup outperforms the standard BDT fit setup in terms of significance by $1\sigma$. This determines our choice of the final selection. The binning in $\Delta E$ and $M_{BC}$ is chosen at the plateau of the significance, where no significant bias is present and is somewhere in the region of 20 bins in each dimension.

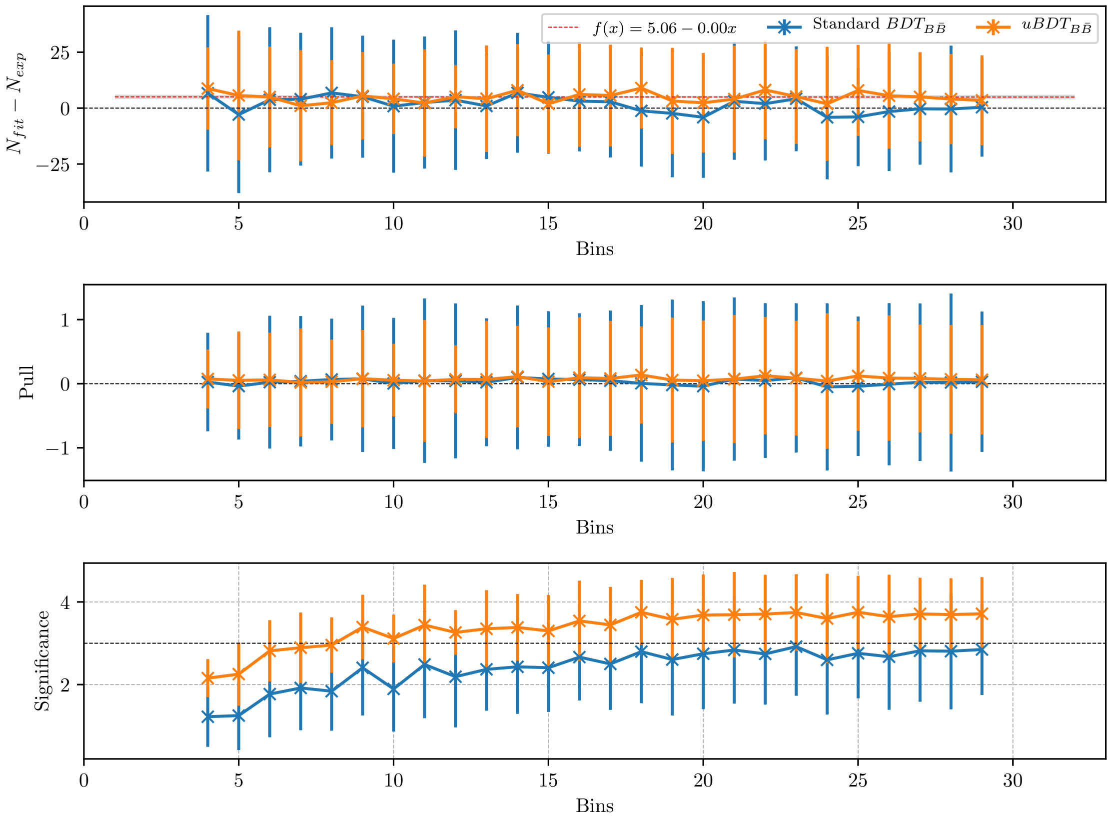



Figure 8.5: Fitted yield and expected yield difference (top), pulls (center) and fit significance (bottom) as a function of binning in $\Delta E$ and $M_{BC}$ for the final sample, optimized with the standard BDT classifier (blue) and the uBDT classifier (orange).

For the chosen binning of $19 \times 19$ in $\Delta E$ and $M_{BC}$ we perform the 10 stream MC fits, where an average stream fit is shown in Figure 8.6, while all fit results are shown in Figure 8.7. All stream fit results were fitted with a 0th-degree polynomial. The global result seems to describe the expected value in a precise manner, with the bias much smaller than the average statistical error. The normalized $\chi^2$ value with $10 - 1 = 9$ degrees of freedom of the global fit is $\chi^2_9 = 0.98$, while the average significance of the fits is around $3.56\sigma$.

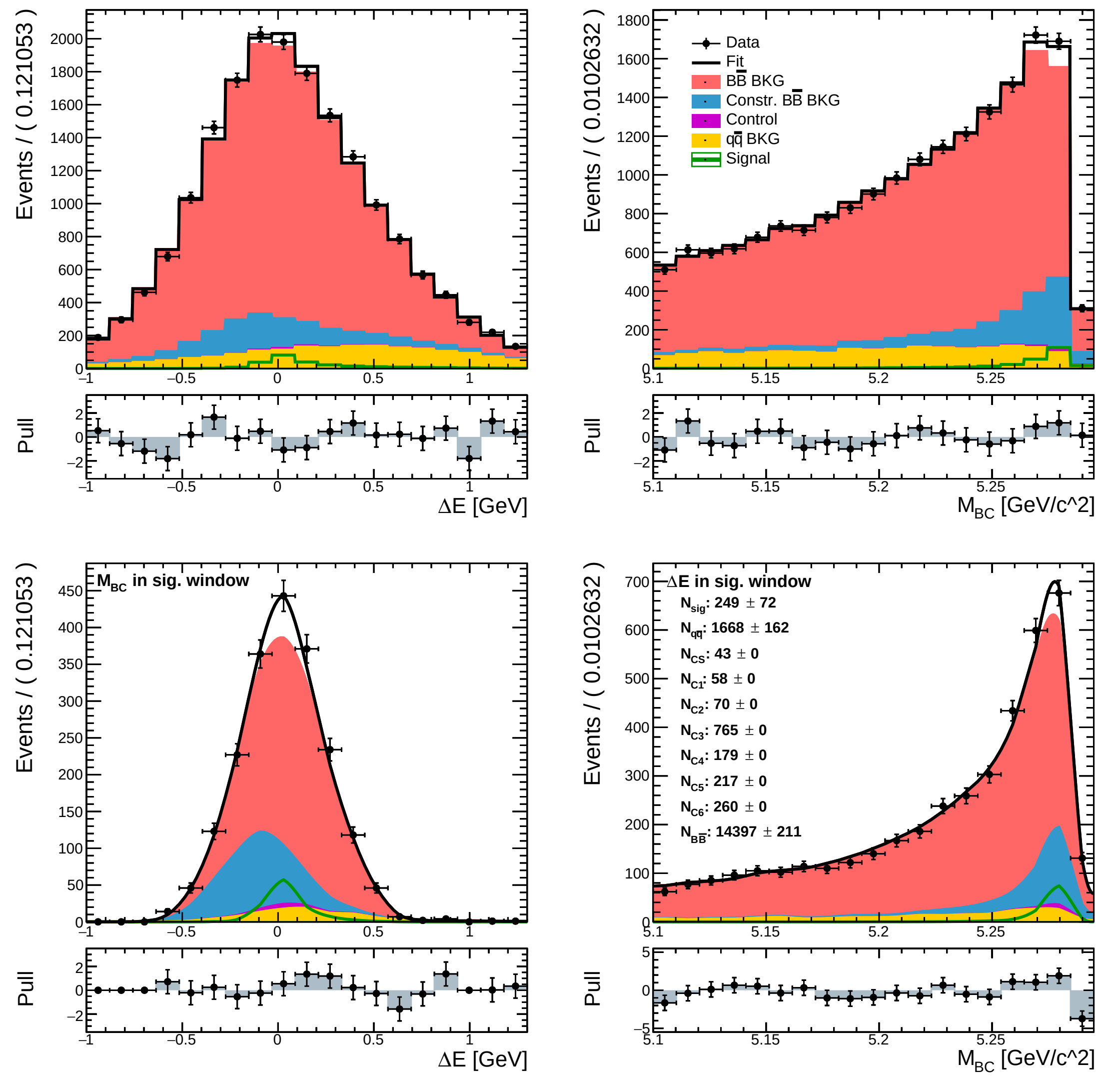



Figure 8.6: An example fit to one stream of MC. Left column shows the $M_{BC}$ and the right column shows the $\Delta E$ distribution of the full fitted sample in the full fit region (top) and the signal region (bottom).

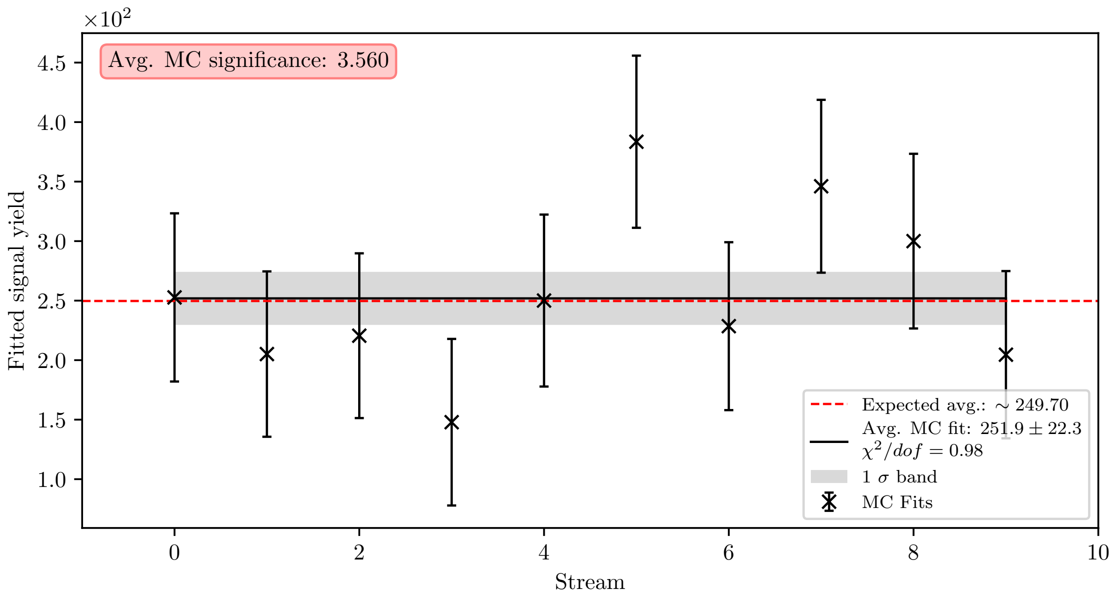



Figure 8.7: Fits to all 10 streams of MC and the global fit with a zero degree polynomial. The red line shows the mean value of the global fit and the gray band shows the $1\sigma$ confidence interval.

### 8.4.2 Control Fit Result

With the control fit setup described in section 8.1.1, we proceed to fit the control sample after the final selection for 10 streams of MC and 1 stream of data. An average MC stream fit is shown in Figure 8.8 for MC and Figure 8.9 for data, while all fit results are shown in Figure 8.10, where all streams of MC are fitted with a 0th degree polynomial. The control fit results for split and joined lepton modes are shown in Table 8.5.

| | $N^{MC}$ | $N^{\text{data}}$ |
|---|---|---|
| $\ell = e$ or $\mu$ | $1180 \pm 11$ | $1192 \pm 44$ |
| $\ell = e$ | $590 \pm 9$ | $588 \pm 28$ |
| $\ell = \mu$ | $591 \pm 7$ | $610 \pm 30$ |

Table 8.5: Control sample fit results for MC and data for various lepton final state modes.

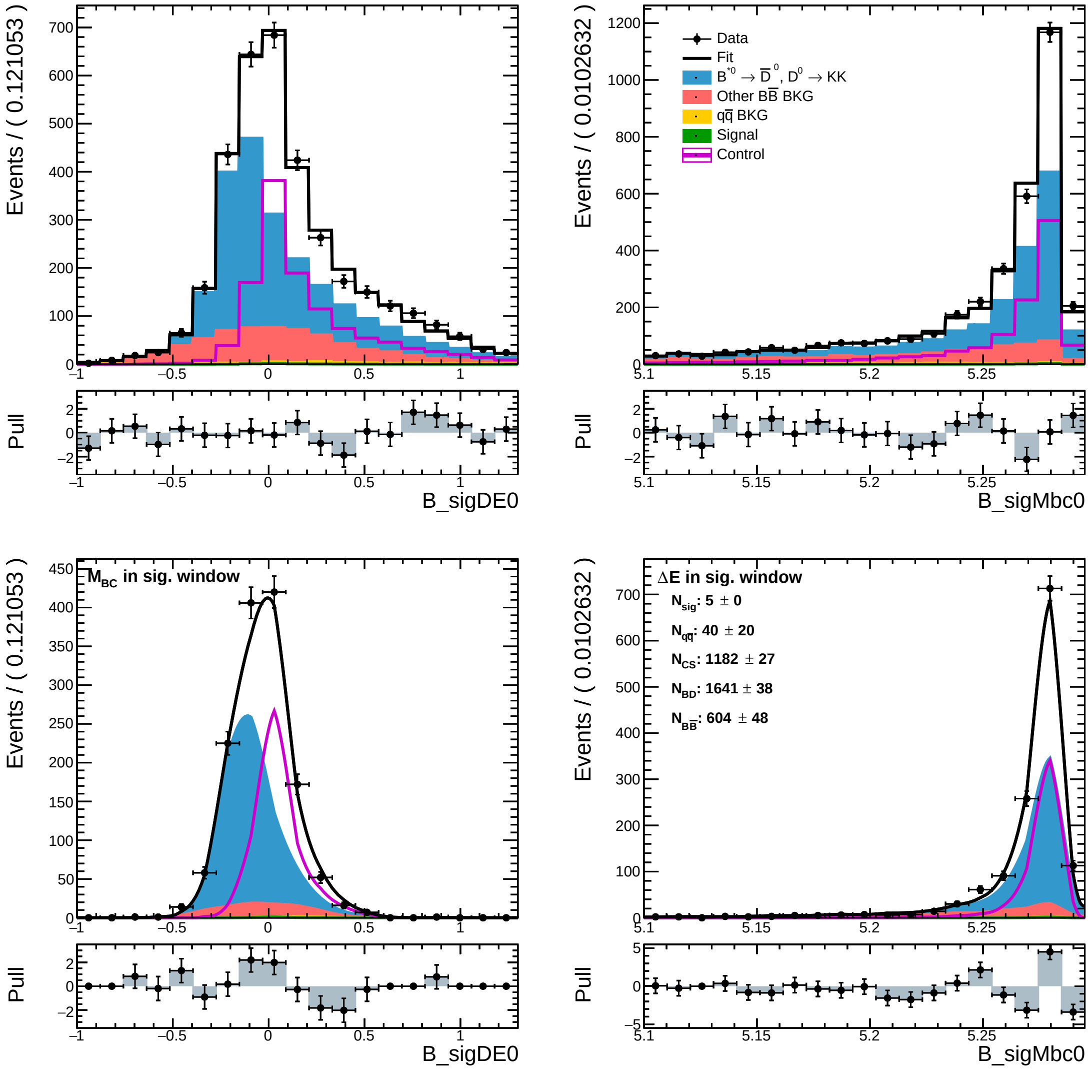



Figure 8.8: Control fit result on stream 9 of MC. Left column shows the $M_{BC}$ and the right column shows the $\Delta E$ distribution in the full fit window (top) and the signal window (bottom).

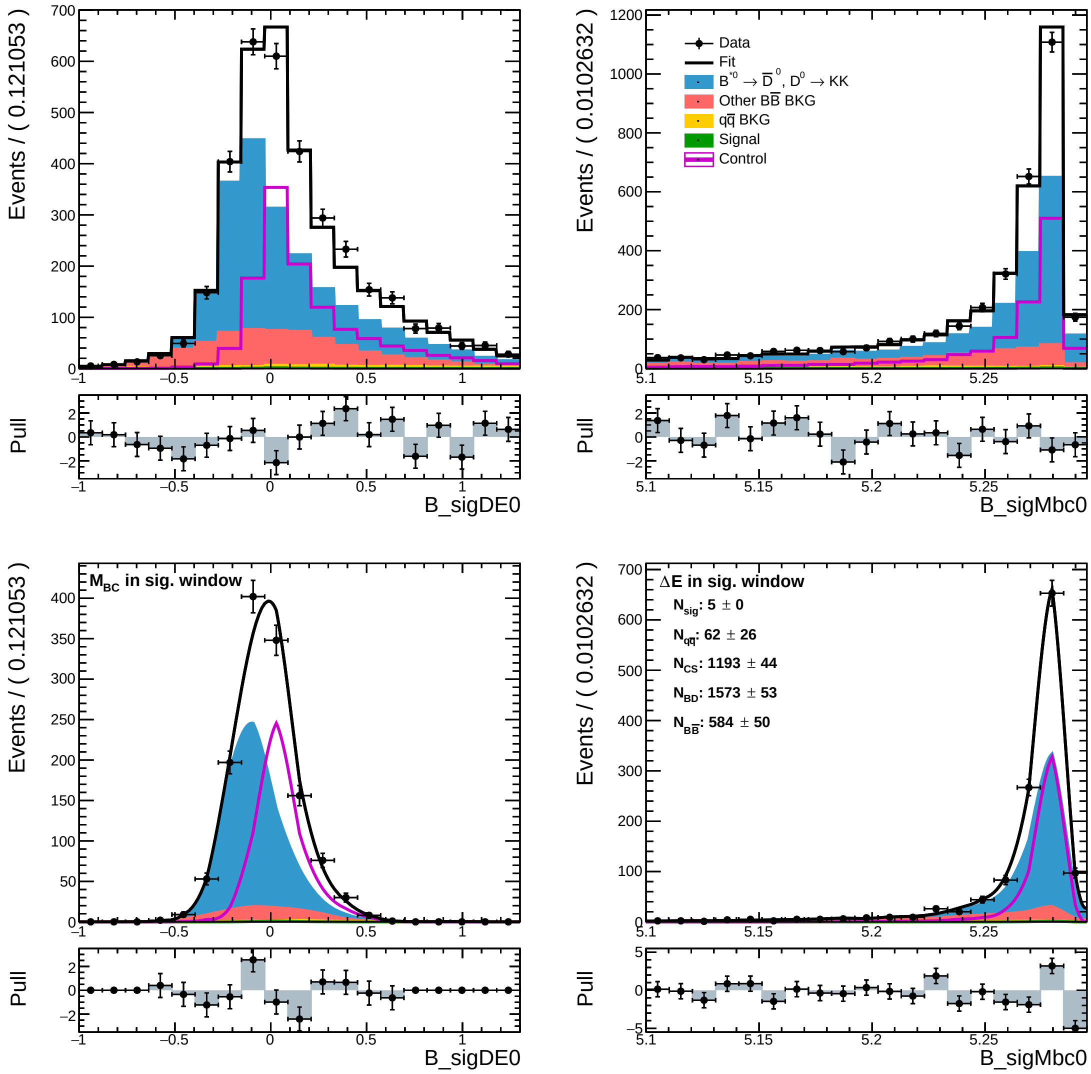



Figure 8.9: Control fit result on real data. Left column shows the $M_{BC}$ and the right column shows the $\Delta E$ distribution in the full fit window (top) and the signal window (bottom).

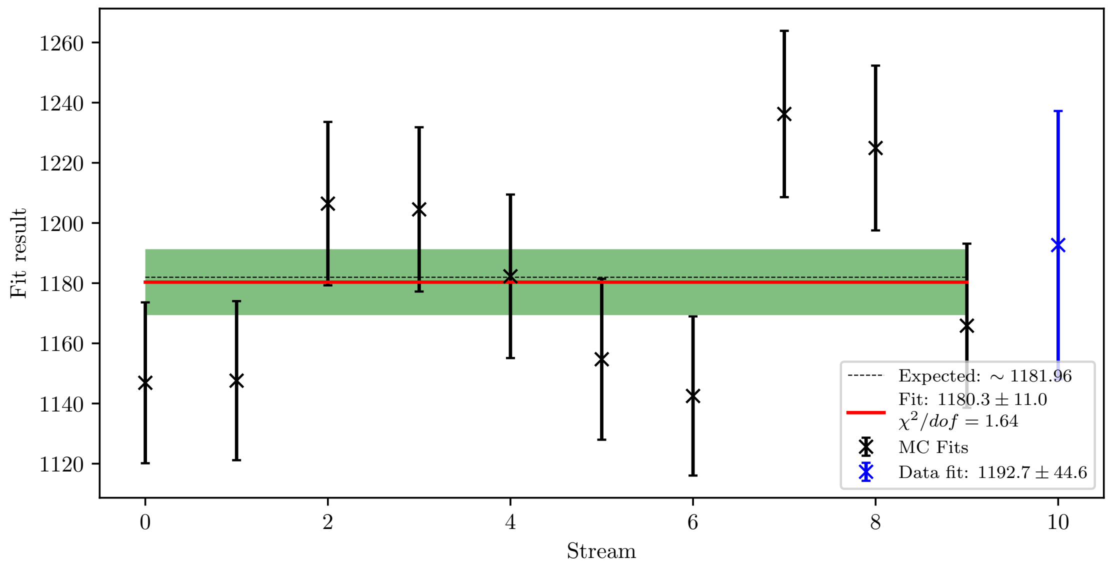



Figure 8.10: Control fit to data and to all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0th degree polynomial. The gray band shows the $1\sigma$ confidence interval around the global MC fit.

### Branching Ratio Measurement for Control Decay

After acquiring the fit results on MC and data, we are able to determine the branching ratio of the control decay, which is defined as

$$\mathcal{B}^{MC}_{\text{control}} = \frac{N^{\text{MC}}_{\text{control}} \times \epsilon_{MC}}{2N^{MC}_{B\bar{B}}}, \tag{8.12}$$

$$\mathcal{B}_{\text{control}} = \frac{N_{\text{control}} \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \tag{8.13}$$

where $N^{\text{MC}}_{\text{control}}$ and $N_{\text{control}}$ are yields of the control fit on MC and data, $\epsilon_{MC}$ is the MC efficiency of the control sample, $\rho_{PID}$ the PID correction factor, and $N^{MC}_{B\bar{B}}$ and $N_{B\bar{B}}$ are the numbers of $B\bar{B}$ meson pairs on MC and data, respectively. The factor of 2 in the denominator comes from the fact that there are 2 $B$ mesons in each $B$ meson pair ($\times 1/2$), where only about 50% of the $B$ meson pairs are charged $B^+B^-$ meson pairs ($\times 2$), and from the fact that we are interested in the branching ratio to the lepton final state of either $e$ or $\mu$, and not their sum ($\times 1/2$).

The control sample efficiency was determined on a separate signal MC sample of the control decay, where we generated $5 \times 10^6$ $B^+\bar{B}^-$ pairs, with one $B$ always decaying via $B^+ \to \bar{D}^0\ell^+\nu$, $D^0 \to K^+K^-$. After applying the final selection, the full and split efficiencies with regard to the lepton final state were determined to be

$$\epsilon_{MC} = (8.89 \pm 0.04) \times 10^{-3},$$
$$\epsilon^{e}_{MC} = (4.40 \pm 0.03) \times 10^{-3},$$
$$\epsilon^{\mu}_{MC} = (4.49 \pm 0.03) \times 10^{-3},$$

where the efficiency error was estimated based on a formula from [19].

$$\epsilon_{MC} = \frac{1}{N}\sqrt{n(1 - \frac{n}{N})},$$

where $n$ is a subset of the full set $N$.

The PID correction factor is obtained by taking into account the known PID efficiency differences between data and MC. It is described in detail in sec:pid-efficiency-correction and is determined to be

$$\rho_{PID} = 0.99 \pm 0.02$$

for the $e$ and $\mu$ mode, as well as both of them together.

The number of $B\bar{B}$ meson pairs can be counted on MC and has been measured for data by the collaboration. The values are

$$N_{B\bar{B}}^{MC} = 765.98 \times 10^6,$$
$$N_{B\bar{B}} = (771.58 \pm 10.56) \times 10^6.$$

Finally, we can determine the branching ratios based on the calculations in Eq. (8.13). The obtained values are shown in Table 8.6 and graphically shown in Figure 8.11, along with the MC generated value and the current PDG world average. Both MC and measured determinations of the control decay branching ratio are in agreement with the expected and the world average values. One should note that the black error bars correspond to statistical uncertainty. Of all the systematic uncertainties, only the PID systematics are included in this results. Other contributions of systematics are not included since this measurement is not the goal of our analysis.

| | | |
|---|---|---|
| $\mathcal{B}_{PDG}$ | $(9.10 \pm 0.42) \times 10^{-5}$ | |
| $\mathcal{B}_{GEN}$ | $9.01 \times 10^{-5}$ | |
| | $\mathcal{B}^{MC}$ | $\mathcal{B}^{\text{data}}$ |
| $\ell = e$ or $\mu$ | $(8.94 \pm 0.09) \times 10^{-5}$ | $(9.06 \pm 0.4) \times 10^{-5}$ |
| $\ell = e$ | $(9.04 \pm 0.15) \times 10^{-5}$ | $(9.07 \pm 0.49) \times 10^{-5}$ |
| $\ell = \mu$ | $(8.85 \pm 0.12) \times 10^{-5}$ | $(9.12 \pm 0.50) \times 10^{-5}$ |

Table 8.6: Control sample fit results for MC and data for various lepton final state modes.

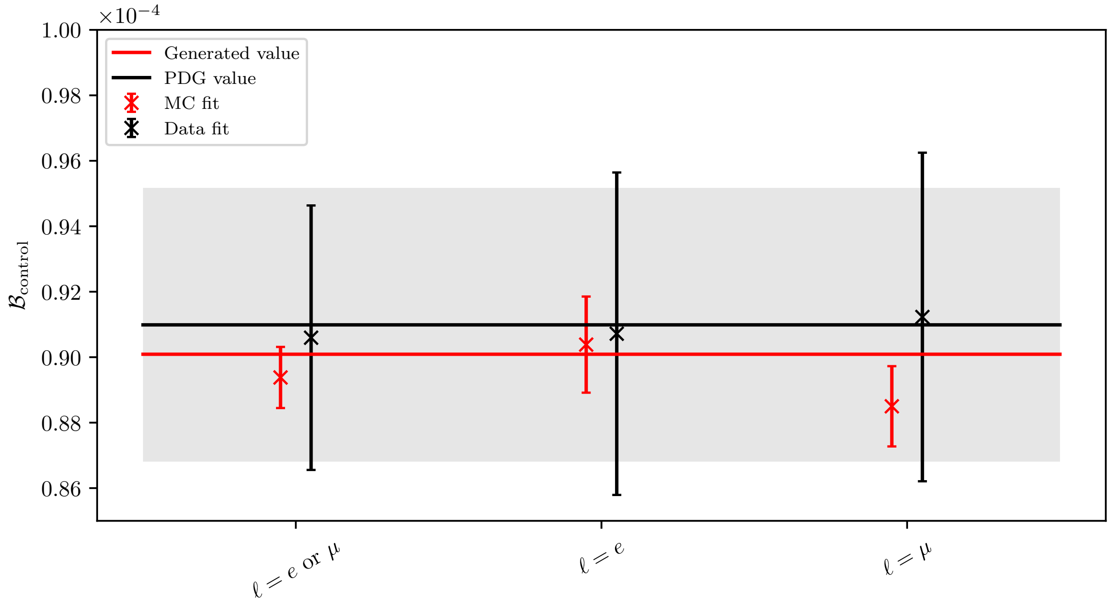



Figure 8.11: Various branching ratio determinations for the control decay of our analysis.

### 8.4.3 Final Results from Data

Finally, after making sure that the analysis procedure makes sense on the control MC, control data and signal MC sample, we can continue to perform the signal extraction process on the full Belle $\Upsilon(4S)$ data sample. Figure 8.12 shows the fit result in projections of $\Delta E$ and $M_{BC}$ for both fit regions. The extracted signal yield on data as well as the yields of the remaining contributions are shown in Table 8.7. Values of all the constraints are shown in Table 8.8.

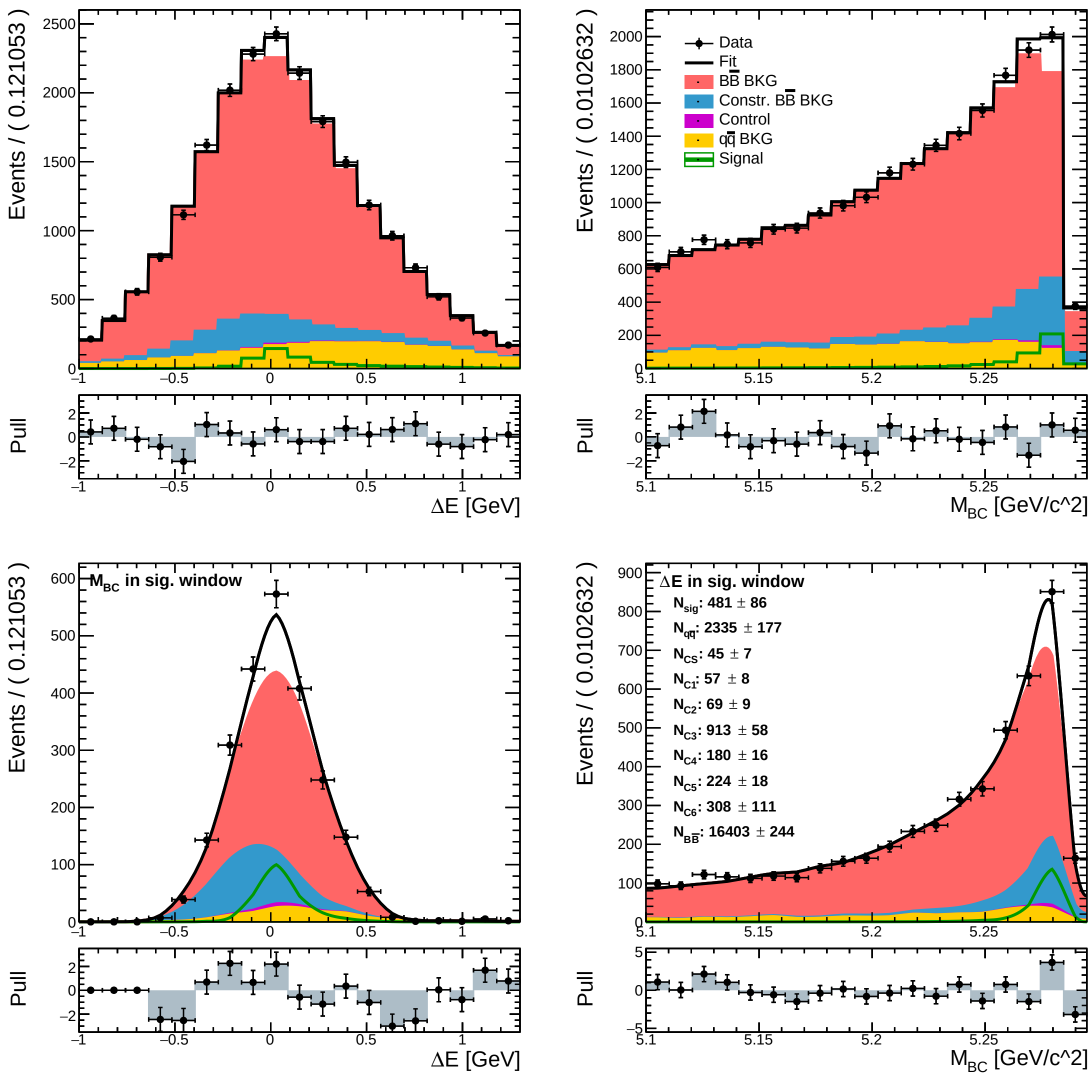



Figure 8.12: Signal fit result on real data. Left column shows the $M_{BC}$ and the right column shows the $\Delta E$ distribution in the full fit window (top) and the signal window (bottom).

| Category | Fit Yield |
|---|---|
| Signal | $476 \pm 86$ |
| $q\bar{q}$ background | $2315 \pm 179$ |
| $C_0$ | $45 \pm 7$ |
| $C_1$ | $57 \pm 8$ |
| $C_2$ | $69 \pm 9$ |
| $C_3$ | $910 \pm 57$ |
| $C_4$ | $180 \pm 16$ |
| $C_5$ | $223 \pm 18$ |
| $C_6$ | $341 \pm 107$ |
| Other $B\bar{B}$ background | $16399 \pm 246$ |

Table 8.7: Yields of all signal fit contributions on data.

| Constraint | Value | Constraint | Value |
|---|---|---|---|
| $\mathcal{B}_0$ | $0.049 \pm 0.001$ | $N_{30}$ | $103 \pm 10$ |
| $\mathcal{B}_1$ | $1.076 \pm 0.004$ | $N_{31}$ | $211 \pm 14$ |
| $\mathcal{B}_2$ | $0.021 \pm 0.001$ | $N_{32}$ | $136 \pm 12$ |
| $\mathcal{B}_3$ | $1.014 \pm 0.018$ | $N_{33}$ | $267 \pm 16$ |
| $\mathcal{B}_4$ | $1.018 \pm 0.010$ | $N_{34}$ | $19 \pm 4$ |
| $\mathcal{B}_5$ | $1.439 \pm 0.063$ | $N_{35}$ | $35 \pm 6$ |
| $\mathcal{B}_6$ | $1.092 \pm 0.038$ | $N_{40}$ | $48 \pm 7$ |
| $\mathcal{B}_7$ | $1.035 \pm 0.008$ | $N_{41}$ | $7 \pm 3$ |
| $\mathcal{B}_8$ | $0.971 \pm 0.038$ | $N_{42}$ | $99 \pm 10$ |
| $\mathcal{B}_9$ | $1.052 \pm 0.038$ | $N_{43}$ | $10 \pm 3$ |
| $\mathcal{B}_{10}$ | $0.945 \pm 0.029$ | $N_{44}$ | $14 \pm 4$ |
| $\mathcal{B}_{11}$ | $1.331 \pm 0.441$ | $N_{45}$ | $3 \pm 2$ |
| $N^{MC}_{\text{control}}$ | $1179 \pm 11$ | $N_{50}$ | $103 \pm 10$ |
| $N_{\text{control}}$ | $1215 \pm 43$ | $N_{51}$ | $63 \pm 8$ |
| $N_{00}$ | $44 \pm 7$ | $N_{52}$ | $31 \pm 6$ |
| $N_{10}$ | $54 \pm 7$ | $N_{53}$ | $21 \pm 5$ |
| $N_{11}$ | $6 \pm 2$ | $N_{60}$ | $69 \pm 8$ |
| $N_{20}$ | $23 \pm 5$ | $N_{61}$ | $94 \pm 10$ |
| $N_{21}$ | $41 \pm 6$ | $N_{62}$ | $63 \pm 8$ |
| $N_{22}$ | $6 \pm 2$ | $N_{63}$ | $35 \pm 6$ |

Table 8.8: Mean values and standard deviations of constraints after the fit.

In this fit setup we perform a random smearing of the $\Delta E$ variable according to the Poisson distribution. The result is similar to a convolution with a Gaussian function, but the application is different in the sense that it returns a slightly different result after each fit. We perform 500 data fits in order to obtain the mean values of the signal yield and fit error. The averaged result is

$$\bar{N}_{sig} = 487 \pm 86,$$

where the error represents the joint value of the statistical error as well as the systematic error of the Gaussian constraints used in the fit. This means that our initial fit is a good representation of the average fit result.

It is possible to estimate the contribution of the systematic uncertainty to the fit error above, which is found to be

$$\sigma_{stat} = 81, \quad \sigma^{GC}_{sys} = 5, \quad \delta^{GC}_{sys} = 1\%, \tag{8.14}$$

where GC stands for Gaussian constraint. Further details can be found in chapter 9.

### Branching Ratio Calculation for Signal Decay

Similarly as for the control decay, we are able to calculate the branching ratio of the signal decay via the formulas

$$\mathcal{B}_{\text{sig}}^{MC} = \frac{N_{\text{sig}}^{\text{MC}} \times \epsilon_{MC}}{2N_{B\bar{B}}^{MC}}, \tag{8.15}$$

$$\mathcal{B}_{\text{sig}} = \frac{N_{\text{sig}} \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \tag{8.16}$$

where $N_{\text{sig}}^{\text{MC}}$ and $N_{\text{sig}}$ are yields of the signal fit on MC and data, $\epsilon_{MC}$ is the MC efficiency of the signal sample, $\rho_{PID}$ the PID correction factor, and $N_{B\bar{B}}^{MC}$ and $N_{B\bar{B}}$ are the numbers of $B\bar{B}$ meson pairs on MC and data, respectively.

The signal efficiency was determined on the same signal MC sample as was used throughout the analysis. The full signal efficiency is determined to be

$$\epsilon_{MC} = (1.030 \pm 0.007) \times 10^{-2},$$

where the efficiency error was calculated in the same manner as in section 8.4.2. The PID correction factors for signal and the numbers of $B\bar{B}$ meson pairs on MC and data are the same as in the case of control decay branching ration measurement

Finally, we can determine the branching ratios based on the calculations in Eq. (8.16). The obtained values are shown in Table 8.9. We see that the measured value is twice as large as the MC value. Of all the systematic uncertainties, only the Gaussian constraint uncertainties are included in the statistical error, while other sources are grouped separately. Further details about the systematic uncertainties estimation can be found in chapter 9.

With the statistical significance of X this is the first observation of the charmless semileptonic decay $B^+ \to K^+K^-\ell^+\nu$. it indicates that the MC contribution to our simulated samples is underestimated. This analysis has shown that the branching fraction of the decay seems considerable enough in order to affect results in precision physics in cases where they are ignored.

| $\mathcal{B}_{GEN}$ | $1.57 \times 10^{-5}$ | |
|---|---|---|
| | $\mathcal{B}^{MC}$ | $\mathcal{B}^{\text{data}}$ |
| $\ell = e$ or $\mu$ | $(1.60 \pm 0.14 \pm X) \times 10^{-5}$ | $(3.09 \pm 0.52 \pm X) \times 10^{-5}$ |

Table 8.9: Signal decay fit results for MC and data.

### Signal Distribution in $m_{KK}$

# Chapter 9

# Systematic Uncertainty

In this chapter, the systematic errors of the analysis are discussed. These uncertainties arise due to various reasons, some of them being the difference between real and simulated data, or due to the nature of the approaches taken in a specific analysis. Depending on their type, some uncertainties are generic and prepared beforehand in order to be used in all analyses, while others are analysis specific and possible sources need to be thought through thoroughly. All systematic uncertainties contributions which are cited in the final result in section 8.4.3 are summarized in the end of this chapter.

## 9.1 PID Efficiency Correction

The PID selection efficiency for the three charged particles in our signal decay needs to be corrected on MC due to various differences when comparing to data. The Belle PID group has prepared correction factors and systematics tables for PID efficiencies for all charged particles. In case of kaon ID and lepton ID, the tables are binned in experiment numbers, particle momentum and in $\cos\theta$ of the particle direction, where, for each bin, a ratio of efficiencies between MC and data is provided, as well as the systematic errors. Each particle's correction factor and error is shown in Table 9.1, as well as the corresponding entry for all 3 particles. The entries are shown for both signal and control region.

The central values were obtained with a weighted average over all experiments, where 100% correlation for error calculation was assumed. A full correlation was also assumed when calculating the $KK$ correction, as both $K$ use the same PID information.

The final PID efficiency systematic error on the full MC sample is determined to be

$$\sigma_{\text{sys}}^{\text{PID}} = 10, \quad \delta_{\text{sys}}^{\text{PID}} = 2\%, \tag{9.1}$$

for the signal as well as the control decay.

| PID correction and systematics | Control region | Signal region |
|---|---|---|
| Same sign $K$ (w.r.t the $B$ meson) | $1.005 \pm 0.009$ | $1.007 \pm 0.010$ |
| Opposite sign $K$ (w.r.t the $B$ meson) | $1.004 \pm 0.009$ | $1.006 \pm 0.009$ |
| $e$ | $0.977 \pm 0.011$ | $0.976 \pm 0.011$ |
| $\mu$ | $0.985 \pm 0.009$ | $0.980 \pm 0.009$ |
| $\ell$ | $0.981 \pm 0.007$ | $0.980 \pm 0.007$ |
| $KKe$ | $0.986 \pm 0.021$ | $0.988 \pm 0.022$ |
| $KK\mu$ | $0.994 \pm 0.020$ | $0.993 \pm 0.021$ |
| $KK\ell$ | $0.990 \pm 0.019$ | $0.990 \pm 0.020$ |

Table 9.1: PID correction factors and systematic error for various charged particles and their combinations.

## 9.2 Fit Bias

Signal and background templates in our analysis are not perfectly distinct from one another and may potentially cause some over- or underestimation of the signal fit yield. In order to study this problem, we estimate the bias from the binning study performed in section 8.4.1 as well as the linearity test toy MC study in section 8.3.2. The two bias functions describe a bias in each direction and are approximated as

$$f_{\min}(x) = -7.25 - 1.12x - \sigma_{f_{\min}}(x), \tag{9.2}$$

$$\sigma_{f_{\min}}(x) = \sqrt{0.050x^2 - 0.175x + 0.410}, \tag{9.3}$$

$$f_{\max}(N_b) = 5.06 + \sigma_{f_{\max}}(N_b), \tag{9.4}$$

$$\sigma_{f_{\max}}(N_b) = \sqrt{0.004N_b^2 - 0.113N_b + 1.112} \tag{9.5}$$

where $x$ represents the signal yield fraction of the data fit and $N_b$ represents the binning choice of the fit. Values of $1\sigma$ intervals have been added to the bias functions in order to be more conservative. The extracted signal yield on data with the fit setup of $N_b = 19$ bins was determined to be $N_{sig} = 487$, which leads to $x = N_{sig}/N_{sig}^{MC} = 487/250 \approx 2$. The bias interval is therefore

$$\sigma_{\text{sys}}^{bias} = {}^{+6}_{-10}, \quad \delta_{\text{sys}}^{bias} = {}^{+1.23\%}_{-2.05\%} \tag{9.6}$$

## 9.3 Gaussian Constraints

As mentioned in section 8.4.3, it is possible to estimate the size of the systematic error of the Gaussian constraints. By fixing the constraints to the nominal values, determined by the data fit, we obtain the pure statistical error, which can then be subtracted from the average fit error in order to determine the systematic uncertainty contribution due to using Gaussian constraints. For the fixed and non-fixed case we perform 500 fits and calculate with the mean values. The split errors are then

$$\bar{N}_{sig} = 487 \pm 86,$$
$$\bar{N}_{sig}^{fix} = 485 \pm 81,$$
$$\sigma_{stat} = 81, \quad \sigma_{sys}^{GC} = 5, \quad \delta_{sys}^{GC} = 1\%,$$

where GC stands for Gaussian constraint. We see that the constrained channels are very well defined and introduce little to none additional uncertainty.

## 9.4 Fit Template Smearing and Offset

The smearing and offset of the $\Delta E$ variable was discussed in section 8.1.1, where we have estimated the central value of the parameters as well as their range in the $1\sigma$ confidence level. We have to perform a study of the effects of different smearing and offset parameter values on the final value of the signal yield. From section 8.1.1, the parameter values are

- Smearing: $40^{+15}_{-17}$ MeV,
- Offset: $6^{+4.6}_{-6}$ MeV.

We perform signal fits for all four different combinations of parameters in the given ranges, where for each parameter setting X fits are performed.

`Wait for referees to do the fits.`

## 9.5 Effects of a Finite MC sample

The shape of signal and backgrounds templates in our analysis is fixed and only their normalization is considered as a floating parameter in the fit. Due to the finite size of the MC sample, the template shape introduces an additional source of uncertainty, as it may differ if produced in a separate, equal-sized MC sample. Since the bins in these 2D histogram templates are statistically independent, we take the content of each bin and vary the value according to the Poisson distribution. This procedure is repeated for X times and the width of the fit yield distribution is taken as the uncertainty estimate.

`Wait for referees to do the fits.`

## 9.6 MVA Selection Efficiencies

Control sample fits allow us to check the behavior of optimized MVA cuts on MC as well as data and see if any of the MVA steps introduce a possible disagreement between the two. We compare control yields, their ratios and ratios of cut efficiencies (double ratios). The following cut scenarios are studied

(a) final selection before any MVA step,

(b) (a) + $BDT_{q\bar{q}}$ cut,

(c) (a) + $uBDT_{B\bar{B}}$ cut,

(d) (a) + $BDT_{q\bar{q}} + uBDT_{B\bar{B}}$ cut (final selection).

The results for control fit yields, their ratios, and double ratios are shown in Figure 9.1. The plot shows that the yield ratios and cut efficiency ratios are consistent with 1. This means that data and MC are in agreement before as well as after applying the final selection cuts. This is an important check since the behavior of our analysis on the control sample suggests that the final selection is not over-optimized to signal MC.

We estimate the systematic error due to the MVA selection steps as the standard deviation of double ratio entries around the nominal value for each step in the MVA selection, except for the final two values for $e$ and $\mu$ modes, since we are performing the inclusive fit. The systematic error estimation is

$$\sigma_{\text{sys.}}^{\text{MVA}} = 1\% \tag{9.7}$$

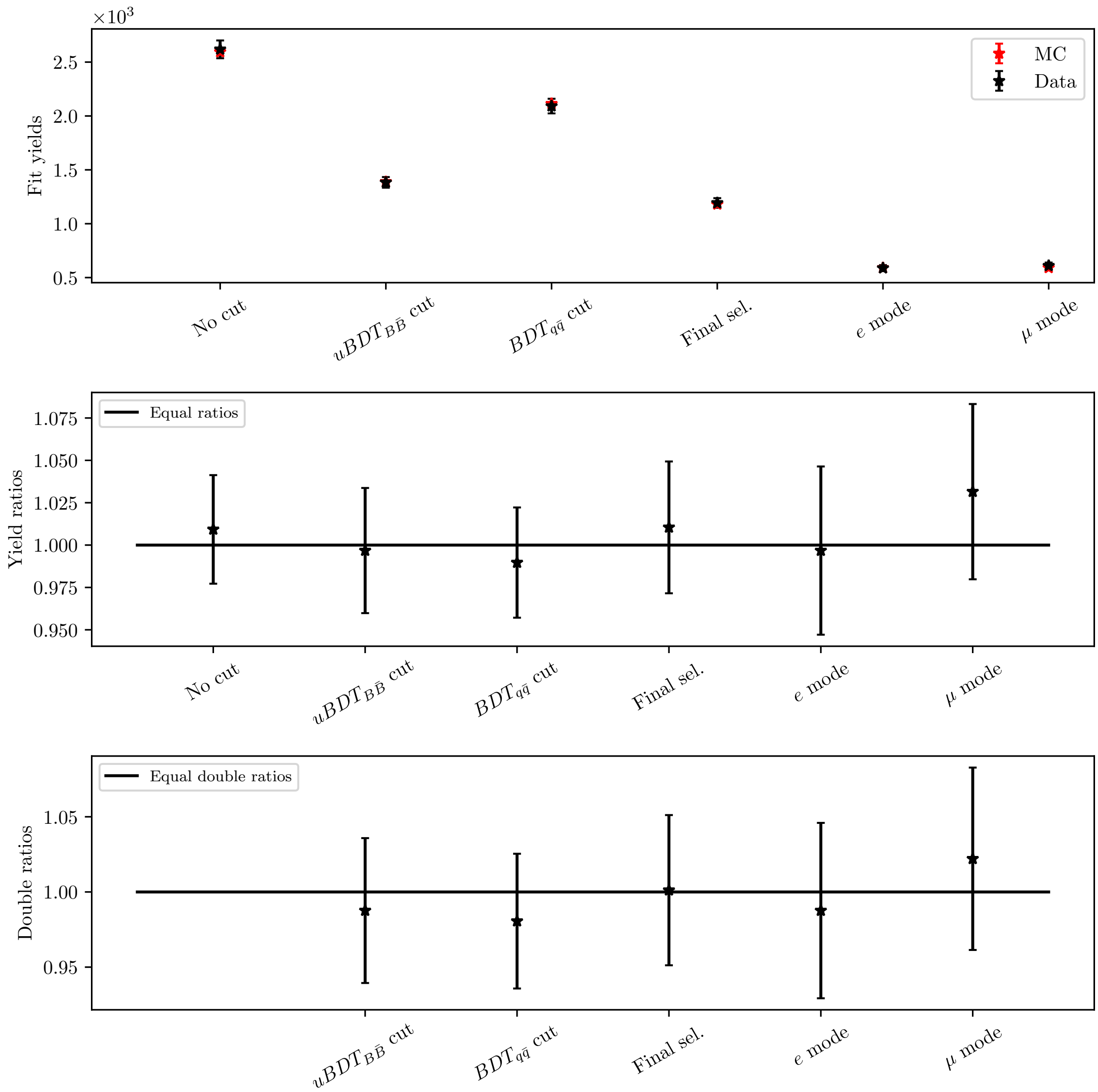



Figure 9.1: Fit yields, their ratios and ratios of cut efficiencies (double ratios) for the control sample fits to data and MC.

## 9.7 Model Uncertainty Effects

The used signal decay model in the generation step was ISGW2 [20], which is known to result in unrealistic predictions and poor agreement with data, so it is not the most precise model for our signal. Due to this model unreliability, our analysis has been made as model-independent as possible via means of not using variables, which exhibit model dependence. Such variables are i.e. squared momentum transfer to the lepton pair ($q^2$), the invariant mass of the two kaon daughters ($m_{KK}$) or decay angle between any two charged particles in the final state. In order to test the effects of model dependency on our final result, we prepare two additional signal MC samples, produced with two extreme scenarios of decay model choice. In the first scenario we generate the signal MC sample with a generic phase-space decay mode `PHSP` [21], which results in continuum-like distributions of $q^2$ and $m_{KK}$. In the other scenario, only resonant-like contributions of $m_{KK}$ are used. These two scenarios act as extreme cases of decay model choice and present a reasonable measure of the model uncertainty. Figure 9.2 shows the generated $m_{KK}$ and $q^2$ distributions of the three mentioned decay models as well as distributions of $\Delta E$ and $M_{BC}$ after the final selection. The different signal MC samples are used for templates in the signal fit where X fits are performed in each case. The differences between mean values of fit yields serve as a measure of model uncertainty.

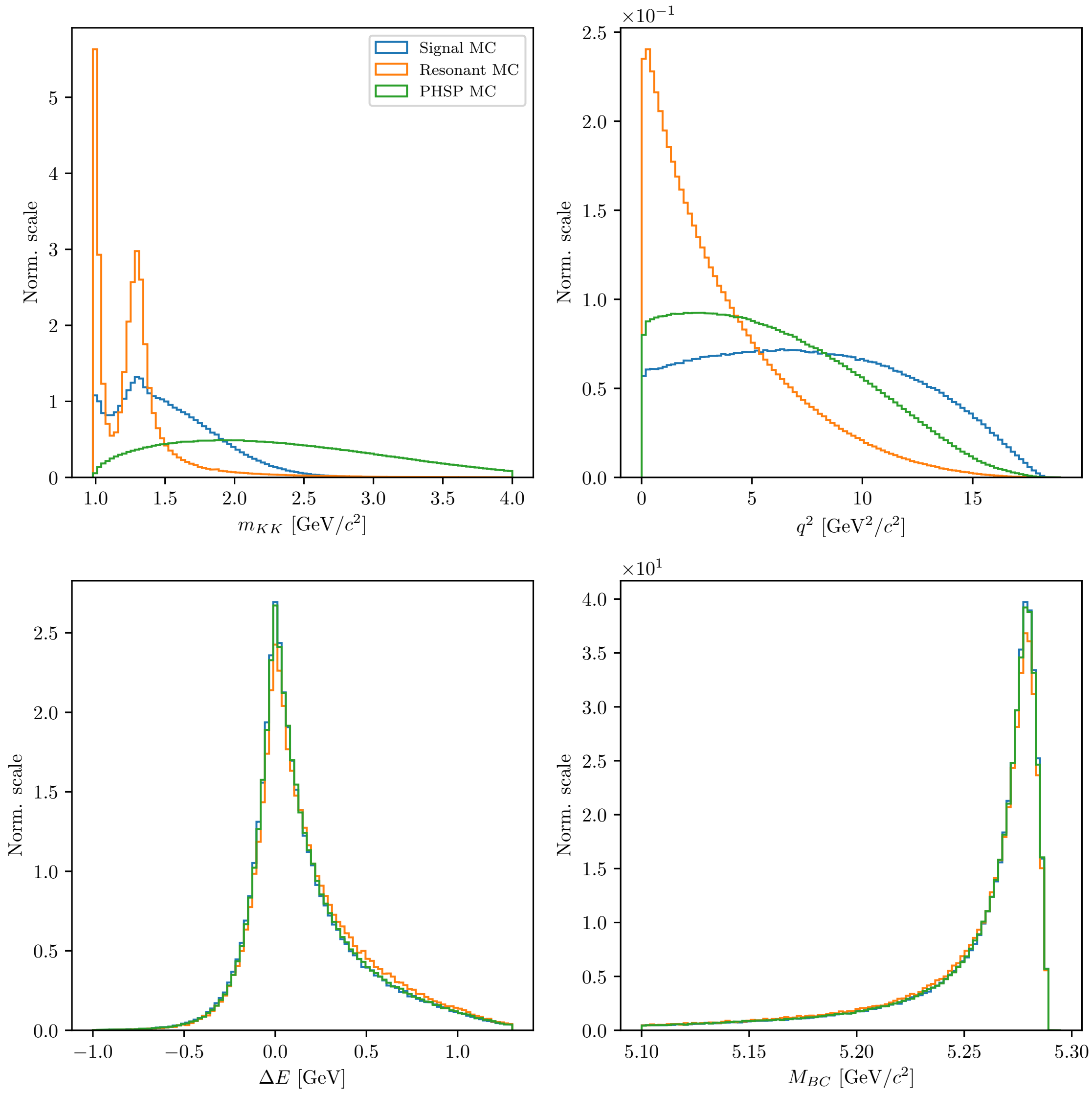



Figure 9.2: $m_{KK}$ (top left), $q^2$ (tip right), $\Delta E$ (bottom left) and $M_{BC}$ (bottom right) for three different signal MC data sets with the standard ISGW2 decay model and two extreme cases of decay models, phase-space (PHSP) and resonant (RES) modes.

`Wait for referees to do the fits.`

## 9.8 Summary of Systematics

# Poglavje 10

# Povzetek doktorskega dela

## 10.1 Uvod

Fizika delcev je eden od stebrov fizike, z močnimi koreninami, ki segajo vse do začetka 20. stoletja. Natančni eksperimenti in preverljiva teorija so pokazali, da vesolje sestoji iz osnovnih delcev in nosilcev interakcij. Osnovne delce delimo na kvarke ($u$, $d$, $s$, $c$, $b$, $t$) in leptone, ki so nadaljnje razdeljeni na nabite leptone ($e$, $\mu$, $\tau$) in pa nevtrine ($\nu_e$, $\nu_\mu$, $\nu_\tau$). Nosilci treh (od štirih) osnovnih interakcij, s katerimi se ukvarjamo na tem področju, so fotoni ($\gamma$) za elektromagnetno, gluoni ($g$) za močno in nabiti- ($W^\pm$) ter nevtralni ($Z^0$) bozoni za šibko interakcijo. Vsi delci in njihovi zrcalni partnerji, antidelci (označeni z $\bar{\ }$), imajo maso, ki jim jo določa Higgsov bozon ($H$). Vse delce ter interakcije med njimi opisuje Standardni model, ki je osrednja teorija fizike visokih energij. Kvarke lahko združujemo v kombinacije oblike $q_1q_2q_3$ (hadroni) ali pa $q_1\bar{q}_2$ (mezoni), med katere sodijo tudi protoni in nevtroni, ki jih opazimo v naravi. Poleg omenjenih dolgo-živečih delcev pa obstajajo tudi težji, manj stabilni delci, ki preko zgoraj naštetih interakcij razpadajo v lažje, stabilnejše. Raziskovanje takšnih procesov s pomočjo pospeševalnikov in trkalnikov nam omogoča spoznavanje zakonov vesolja danes pa vse do njegovega začetka.

Osrednji del doktorske disertacije predstavljajo meritve razpadov mezonov $B$, delcev, ki so sestavljeni iz težkega kvarka $b$ in enega od lahkih kvarkov $u$ ali $d$. Ena bolj presenetljivih lastnosti vesolja je kršitev simetrije $CP$, t.j. kombinacije simetrij konjugacije naboja ($C$) in prostorske inverzije ($P$). Simetrija $CP$ nakazuje, da so fizikalni procesi delcev in zrcalni procesi antidelcev enaki, kar pa danes vemo, da ne drži v celoti, in poznamo procese, ki to simetrijo kršijo. Kršitev simetrije $CP$ je tesno povezana s šibko interakcijo, to pa predstavlja našo motivacijo za študijo mezonov $B$, saj šibki razpadi predstavljajo večji del vseh razpadov mezonov $B$.

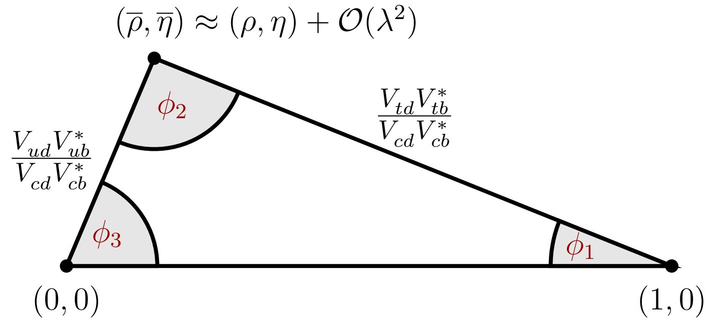
Edinstvena lastnost šibke interakcije je, da lahko spreminja tip oziroma t.i. okus kvarkov, medtem ko ga ostale interakcije ohranjajo. Takšni procesi so opisani s prehodno matriko CKM (Cabibbo-Kobayashi-Maskawa) [11, 12]

$$V_{CKM} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} . \tag{10.1}$$

Unitarnost matrike CKM nam omogoča, da iz nje izluščimo matematične identitete, od katerih je ena pomembnejših

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \tag{10.2}$$

poznana pod imenom unitarni trikotnik, saj predstavlja zaključen vektor treh točk v kompleksni ravnini, kot prikazuje Slika 10.1. Parametri matrike CKM niso določljivi s strani teorije, temveč jih moramo določiti z eksperimentalnimi meritvami tako, da najdemo procese, ki so tesno povezani s stranicami in koti unitarnega trikotnika. Na tak način lahko preverimo, če je oblika trikotnika konsistentna, kar predstavlja dober test Standardnega modela. V primeru, da opisana enaba ne bi opisala trikotnika, bi to nakazovalo na potencialne nove procese, ki jih še ne poznamo, in jih kolektivno imenujemo "nova fizika". Dodatna motivacija za študijo mezonov $B$ je ta, da velik delež njihovih razpadov predstavlja procese, potrebne za meritev unitarnega trikotnika.




Slika 10.1: Unitarni trikotnik s parametri $\lambda$, $\eta$, $\rho$ and $A$ (slednji ni prikazan), ki predstavljajo proste parametre matrike CKM. Trikotnik je prikazan v Wolfensteinovi parametrizaciji [22].

Procesi, ki jih študiramo v tej analizi, so tesno povezani z elementom $V_{ub}$ matrike CKM, saj le-ta opisuje prehode kvarkov $b \to c$. Od vseh elementov, je absolutna vrednost tega elementa najmanjša, relativna napaka pa največja, zato meritve iz tega področja potencialno omogočajo največjo izboljšavo. Takšni prehodi kvarkov so prisotni v nečarobnih (t.j. brez kvarkov $c$) semi-leptonskih razpadih mezonov $B$ oblike

$$B^+ \to X_u^0 \ell^+ \nu_\ell, \tag{10.3}$$

kjer $X_u^0$ predstavlja nečarnobne mezone, $\ell$ pa je eden od nabitih leptonov. Frekvenco razpadov, ki je tesno povezana z elementom $V_{ub}$, opišemo z enačbo

$$\mathrm{d}\Gamma \propto G_F^2 |V_{ub}|^2 |L^\mu \langle X_u | \bar{u}\gamma_u \frac{1}{2}(1-\gamma_5) b | B \rangle|^2, \tag{10.4}$$

kjer $G_F$ predstavlja Fermijevo konstanto, $L^\mu$ leptonski tok, izraz v Diracovih oklepajih pa hadronski tok. V takšnih prehodih $|V_{ub}|^2$ predstavlja verjetnost za prehod $b \to u$.

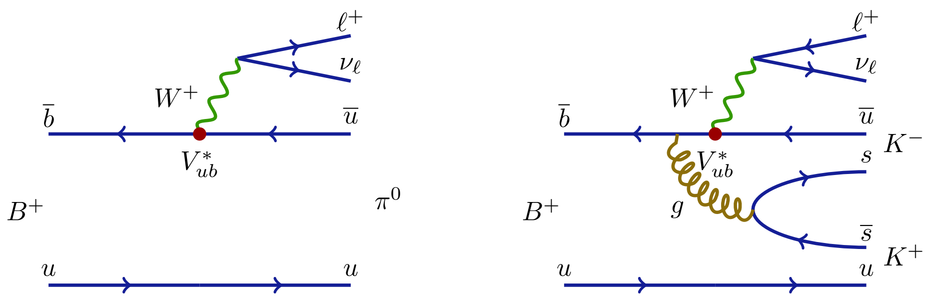
Meritev elementa $V_{ub}$ je možna na ekskluziven in inkluziven način, kjer pri prvi metodi opravljamo meritve v specifično definirana končna stanja, kot na primer $B \to \pi\ell\nu$, pri drugi metodi pa opravljamo meritev s kupno končno stanje oblike $B \to X_u \ell\nu$. Obe metodi potekata preko različnih pristopov in se soočata z različnimi tezavami, kar pomeni, da sta oba končna rezultata nekorelirana. Rezultata obeh meritev imata tudi zelo podobno natančnost, medtem ko se srednja vrednost le deloma ujema. Rezultata se razlikujeta s signifikanco $3\sigma$, kar predstavlja večjo težavo znotraj področja. Trenutni svetovni povprečji [13] ekskluzivne (iz razpadov $B^0 \to \pi^- \ell^+ \nu$) in inkluzivne meritve (GGOU kolaboracija [14]) sta

$$|V_{ub}|_{\mathrm{e.}} = (3.65 \pm 0.09 \pm 0.11) \times 10^{-3}, \tag{10.5}$$

$$|V_{ub}|_{\mathrm{i.}}^{\mathrm{GGOU}} = \left(4.52 \pm 0.15\ ^{+0.11}_{-0.14}\right) \times 10^{-3}, \tag{10.6}$$

kjer prva in druga napaka predstavljata eksperimentalno in teoretsko napako. Rezultati inkluzivnih meritev so praviloma večji kot rezultati ekskluzivnih. Razlogov za neujemanje je lahko več, od nepoznanih napak pri eksperimentu ali teoriji, do prispevkov nove fizike.

V tej analizi se osredotočamo na enega od možnih razlogov za zgoraj omenjeno neujemanje, konkretneje za razpad $B^+ \to K^+ K^- \ell^+ \nu$, ki je strukturno precej podoben razpadu $B \to \pi \ell \nu$ za razliko produkcije para kvarkov $s\bar{s}$ ki se potem hadronizira v nove delce, kot prikazuje Slika 10.2. V inkluzivnih meritvah nečarobnih semi-leptonskih razpadov mezonov $B$ se standardno uporablja $K$-veto, t.j. selekcija, kjer zahtevamo, da v končnem stanju nimamo mezonov $K$ (sestava $q\bar{s}$, $q \in [u, d]$), poznanih tudi pod imenom kaoni. Kaoni v končnem stanju nakazujejo na pogost prehod kvarkov $b \to c \to s$, ki pa jih hočemo v analizah prehodov $b \to u$ zatreti. V primeru naše analize imamo v končnem stanju 2 kaona s prehodom $b \to u$, kar pomeni, da takšni razpadi niso upoštevani v inkluzivnih meritvah, čeprav bi morali biti. Cilj študije je določiti pogostost razpadov $B^+ \to K^+ K^- \ell^+ \nu$ s prehodom $b \to u$ in s tem oceniti, kakšen potencialen efekt ima lahko neupoštevanje teh razpadov na inkluzivno meritev elementa $V_{ub}$. V nadaljevanju bo razpad $B^+ \to K^+ K^- \ell^+ \nu$ zaradi enostavnosti zapisan kot $B \to K K \ell \nu$.




Slika 10.2: Feynmanovi diagrami za razpada $B^+ \to \pi^0 \ell^+ \nu_\ell$ (levo) in $B^+ \to K^- K^+ \ell^+ \nu_\ell$ (desno).

## 10.2 Experimentalna postavitev

Podatki, uporabljeni v tej analizi, so bili ustvarjeni pri trkih elektronov $e^-$ in pozitronov $e^+$ v pospeševalniku KEKB in zajeti z detektorjem Belle. Eksperiment je trajal od leta 1999 do 2010 pod okriljem znanstvene organizacije KEK v mestu Tsukuba na Japonskem. Trki delcev so se dogajali pri energiji, ki je ustrezala masi resonance $\Upsilon(4S)$, (sestava $b\bar{b}$). V tem delu disertacije sta opisana pospeševalnik in detektor, podrobnejši opis pa se nahaja v literaturah [16] in [17].

### 10.2.1 Trkalnik KEKB

KEKB je asimetričen trkalnik delcev $e^+e^-$, ki potujejo po obročih s premerom 3 km v gručah. V središču detektorja gruči elektronov z energijo 8 GeV in pozitronov z energijo 3.5 GeV trčita pod kotom 22 mrad. Skupna invariantna masa trka ustreza masi resonance $\Upsilon(4S)$

$$E_{CM} = 2\sqrt{E_{e^+}E_{e^-}} = m_{\Upsilon(4S)}c^2 \approx 10.58 \text{ GeV}. \tag{10.7}$$

Delež mezonov $\Upsilon(4S)$ razpade preko zelo čistega kanala v dva praktično mirujoča mezona $B$, kar tej in v podobnih analizah pogosto izkoriščamo, saj je začetno stanje dobro poznano.

Trkalnik je v času obratovanja zajel količino podatkov, ki ustreza integrirani luminoznosti 1041 fb$^{-1}$, od katere okoli 711 fb$^{-1}$ predstavlja podatke, zajete pri energiji 10.58 GeV, t.j. masi resonance $\Upsilon(4S)$. Slednja vrednost integrirane luminoznosti ustreza številu $771 \times 10^6$ parov $B\bar{B}$ mezonov.

### 10.2.2 Detektor Belle

Detektor Belle je magnetni masni spektrometer, ki pokriva večji del prostorskega kota. Njegov namen je, da detektira delce, ki se gibljejo v magnetnem polju 1.5 T in so potomci trkov $e^+e^-$. Cilj je določiti energijo in gibalno količino delcev, kar dosežemo preko detektorskih podsistemov, ki so okoli interacijske točke postavljeni v plasteh. Detektor pokriva polarni kot med $17° \leq \theta \leq 150°$, med tem ko je azimutni kot pokrit v celoti, kar skupaj predstavlja 92% pokritost polnega prostorskega kota.

#### Silikonski detektor verteksov

Silikonski detektor verteksov je postavljen najbli—zje interakcijski točki. Sestavljen je iz dvostranskih silikonskih detektorjev, ki podajajo 2D informacijo o prehodih nabitih delcev z natančnostjo okoli 100 $\mu$m. To nam omogoča določitev točk razpada (verteksov) kratko-živecih delcev.

#### Osrednja potovalna komora

Osrednja potovalna komora je sestavljena iz mnogo žic, napeljanih skozi mešanico plina. Komora tako meri sledi nabitih delcev, ki potujejo skozi magnetno polje v detektorju. Preko sledi lahko določimo informacijo o gibalni količini delca, hkrati pa v območju gibalne količine pod 0.8 GeV/$c$ služi tudi za njihovo identifikacijo.

#### Merilec časa preleta

Merilec časa preleta meri časovno razliko od trka pa do preleta delca skozi enega od scintilatorjev tega podsistema. Namen je identifikacija delcev v območju gibalnih količin 0.8 GeV/$c < p <$ 1.2 GeV/$c$, še posebej kaonov $K^\pm$ in pionov $\pi^\pm$. Pri isti gibalni količini zaradi različnih mas delcev dobimo različne case preleta, kar lahko uporabimo za določitev njihove mase. Časovna resolucijo tega podsistema je boljša kot ali enaka 100 ps.

#### Pragovni števec sevanja Čerenkova

Števec sevanja Čerenkova se prav tako uporablja za identifikacijo delcev, deluje pa v višjih območjih gibalne količine, kjer merilec časa preleta ni več zadosten, t.j. v območju 1.0 GeV/$c$ < 4.0 GeV/$c$. Silikatni aerogel z dobro določenim lomnim količnikom predstavlja osrednjo strukturo podsistema, seva Čerenkovo svetlobo, če ga preletijo delci, ki se gibljejo hitreje od svetlobne hitrosti v tej snovi. Pragovni števec deluje na osnovi, da prelet lažjih delcev povzroči sevanje Čerenkova, prelet težjih delcev pa ne.

#### Elektromagnetni kalorimeter

Elektromagnetni kalorimeter služi za detekcijo delcev, ki interagirajo elektromagnetno, predvsem elektronov in fotonov. Z njim lahko izmerimo pozicijo in energijo delca, ko zadane kalorimeter. Ko elektroni ali fotoni zadanejo kristalne celice kalorimetra, povzročijo t.i. elektromagnetni tuš, medtem ko drugi, težji delci, ne interagirajo na enak način in v kalorimetru pustijo le majhen delež energije. Energijska ločljivost kalorimetra je približno 1.7%.

#### Detektor mezonov $K_L^0$ in mionov

Za elektromagnetnim kalorimetrom, na drugi strani magnetnega jedra, je postavljen detektor mezonov $K_L^0$ in mionov za gibalno količino večjo od 0.6 GeV/$c$. Ti delci so so visoko penetrirajoči, saj lahko preletijo vse do sedaj opisane podsisteme. Prvi so nevtralni in jih lahko določimo preko hadronske interakcije v detektorju in preko manjkajoče nabite sledi, medtem ko so drugi nabiti in jih identificiramo že samo z njihovo prisotnostjo.

## 10.3 Analizni postopek

Analizni postopek je doloen na podlagi simuliranih podatkov, oziroma Monte Carlo (MC) simulacije. Ta nam omogoa, da na podlagi teoretinega modela razpadov, dobro opiemo realnost, dodatno pa nam je na voljo "resnica", kot na primer generirane lastnosti delcev in njihova identiteta, ki je bila doloena ob generaciji.

Za pripravo analiznega postopka imamo na voljo $6 - 10\times$ ve podatkov kot jih je izmerjenih, s imer poveamo natannost analiznih korakov in zmanjamo monost statistinih fluktuacij.

Poleg signalnega razpada $B^+ \to K^+K^-\ell^+\nu$, v tudiji rekonstruiramo tudi t.i. kontrolni razpad $B^+ \to \bar{D}^0\ell^+\nu$, $\bar{D}^0 \to K^+K^-$. Drugi ima enako konno stanje kot prvi, le da se zgodi pri prehodu kvarkov $b \to c$, za razliko prehoda $b \to u$ pri prvem razpadu. Loimo jih lahko zelo dobro preko invariantne mase dveh kaonov, ki je v primeru kontrolnega razpada zelo omejena okoli mase mezona $D^0$, v primeru signalnega razpada pa je razporejena po celotnem obmoju, kot prikazuje Slika X.

PLOT

### 10.3.1 Rekonstrukcija razpada

Postopek rekonstrukcije prinemo z izbiro dolgo-iveih stabilnih delcev, ki so v naem primeru elektroni $e^{\pm}$, mioni $\mu^{\pm}$ ter kaoni $K^{\pm}$. Vsi so nabiti in v detektorju pustijo sled. Nevtrino $\nu$ je nevtralen in interagira le preko ibke interakcije, zato jih s taknim detektorjem ne moremo opaziti, kar predstavlja manjkajoo energijo in gibalno koliino v dogodku trka $e^+e^-$.

Selekcija poteka na podlagi reza spremenljivk, kjer je izbrano obmoje doloeno na podlagi optimizacije metrike FOM (ang. *figure of merit*), definirane kot

$$FOM = \frac{N_S}{\sqrt{N_S + N_O}}, \tag{10.8}$$

kjer $N_S$ predstavlja pravilno rekonstruirane kandidate (signal), $N_O$ pa nepravilno rekonstruirane (ozadje).

Povzeta selekcija dolgo-iveih stabilnih delcev je

- elektroni: $|d_0| < 0.1$ cm, $|z_0| < 1.5$ cm, $p_{CMS} \in [0.4, 2.6]$ GeV/$c$, $PID_e > 0.9$,
- mioni: $|d_0| < 0.1$ cm, $|z_0| < 1.5$ cm, $p_{CMS} \in [0.6, 2.6]$ GeV/$c$, $PID_\mu > 0.97$,
- kaoni: $|d_0| < 0.15$ cm, $|z_0| < 1.5$ cm, $p_{CMS} \in [0, 2.5]$ GeV/$c$, $PID_{K/\pi} > 0.6$,

kjer $d_0$ in $z_0$ predstavljata vpadne parametre nabitih delcev, $p_{CMS}$ gibalno koliino v teinem koordinatnem sistemu, $PID_e$ in $PID_\mu$ metriko identifikacije delcev za elektrone in mione, $PID_{K/\pi}$ pa metriko separacije med kaoni in pioni.

Iz izbranih kandidatov nato naredimo kombinacije $Y = KKe,\ KK\mu$, ki sluijo kot kandidati mezonov $B$, z izjemo manjkajoih nevtrinov. Na podlagi dejstva, da je detektor Belle hermetino zaprt in pokriva veino prostorskega kota, in da dobro poznamo zaetno stanje $\Upsilon(4S)$, lahko doloimo etverec manjkajoe (ang. *missing*) gibalne koliine kot

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \sum_i^{\text{Dogodek}} (E_i,\, \vec{p}_i)\,, \tag{10.9}$$

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \left(\mathrm{p}_Y - \sum_i^{\text{ROE}} (E_i,\, \vec{p}_i)\right), \tag{10.10}$$

kjer p predstavlja etverec gibalne koliine, indeks $i$ tee po vseh delcih znotraj mnoice, ROE (ang. *rest of event*) pa predstavlja podmnoico celotnega dogodka, ki vsebuje vse delce, ki niso bili uporabljeni v rekonstrukciji kandidata $Y$.

Tudi na tej stopnji je prisotnih veliko napanih kombinacij kandidatov $Y$, zato po enakem postopku optimiziramo nadaljnjo selekcijo

- mezoni $B$: $P(\chi^2, NDF) > 6.0{\times}10^{-3}$, $|\cos\theta_{BY}| < 1.05$, $m^2_{miss} < 0.975$ GeV/$c^2$, $5.1$ GeV/$c^2 < M_{BC} < 5.295$ GeV/$c^2$, $-1.0$ GeV $< \Delta E < 1.3$ GeV,

kjer $P(\chi^2, NDF)$ predstavlja kvaliteto razpadne toke mezona $B$, $m^2_{miss}$ pa invariantno maso etverca manjkajoe gibalne koliine v dogodku. Ostali izrazi za $\cos\theta_{BY}$, $M_{BC}$ in $\Delta E$ so

$$\cos(\theta_{BY}) = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \tag{10.11}$$

$$M_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}_B|^2}, \tag{10.12}$$

$$\Delta E = E_B - E_{CMS}/2 \tag{10.13}$$

in po vrsti predstavljajo kot med nominalnim ($B$) in rekonstruiranim ($Y$) mezonom $B$, maso, vezano na energijo arka v teinem koordinatnem sistemu, in razliko energije kandidata in polovice teine energije $E_{CMS}$. Za pravilne kombinacije mezonov $B$ ima porazdelitev po $M_{BC}$ vrh pri $m_B$, porazdelitev $\Delta E$ pa okoli $\Delta E \approx 0$ GeV.

Potrebno je omeniti, da so v posameznem dogodku lahko prisotni tako nevtralni kot nabiti delci, ki ne prihajajo neposredno iz trka, temve so lahko bodisi produkti sekundarnih interakcij v detektorju, bodisi delci, ki izhajajo iz ozadja na raun potovanja arkov po obroih pospeevalnika. Takne delce je potrebno odstraniti iz En. (10.10), emur pravimo "ienje dogodka". V tej analizi je bilo opravljeno temeljito ienje tako nevtralnih kot nabitih delcev, ki so terjali razline pristope, pri tem pa smo uporabili zelo uspene metode strojnega uenja za prepoznavanje taknih neelenih delcev. Slika X prikazuje primerjavo med oienim in neoienim dogodkom, kjer primerjamo porazdelitvi $\Delta E$ and $M_{BC}$.

PLOT

### 10.3.2 Odstranjevanje ozadja

Ozadje v takni analizi predstavljajo napani kombinacije razpadne verige signalnega kandidata. Napana kombinacija lahko pomeni v smislu napane kombinatorike ali pa konno stanje drugih razpadnih kanalov posnema konno stanje signalnega razpada, z izjemo potencialnih manjkajoih delcev. Takne kombinacije v splonem nimajo enakih lastnosti kot signalne kombinacije, zato skuamo najti naine, kako takno ozadje odstraniti na najbolj optimalen nain.

Odstranjevanja ozadja se lotimo v treh korakih, v prvem koraku uporabimo enostavne reze na invariantni masi kaonskega para, saj priakujemo, da veliko parov $KK$ pride iz resonancam podobnih struktur, kot na primer $\phi \to KK$ ali $D^0 \to KK$, kjer za slednjo e vemo, da je prisotna v kontrolnem razpadu. Prav tako se lahko zgodi, da je eden od pionov napano identificiran kot kaon. Slika X prikazuje vse reze, ki jih uporabimo za odstranjevanje omenjenih kandidatov, rezi pa so

- signalni razpad: $|m_{KK} - m_\phi| > \Delta_\phi$, $|m_{KK} - m_{D^0}| > \Delta_{D^0}$, $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$,

kjer $m_{KK}$ predstavlja invariantno maso kaonskega para $KK$, $m_{K\pi}$ pa invariantno maso kaonskega para $KK$, kjer je bila masa kaona, katerega naboj je nasproten naboju $B$ mezona, zamenjana z maso delca $\pi^0$. Ostali parametri so $m_\phi \approx 1.019\ \mathrm{GeV}/c^2$, $m_{D^0} \approx 1.864\ \mathrm{GeV}/c^2$, $\Delta_\phi \approx 8 \times 10^{-3}\ \mathrm{GeV}/c^2$ in $\Delta_{D^0} \approx 1.5 \times 10^{-2}\ \mathrm{GeV}/c^2$. V primeru tudije kontrolnega razpada je bil uporabljen drugaen rez

- kontrolni razpad: $|m_{KK} - m_{D^0}| < \Delta_{D^0}$, $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$,

kjer se osredotoimo na ozko okno okoli mase mezona $D^0$.

V drugem koraku se lotimo odstranjevanja t.i. kontinuumskega ozadja, kjer kandidati prihajajo iz procesov $e^+e^- \to q\bar{q}$, $q \in [u, d, s, c]$. Posluimo se metod strojnega uenja, ki prepoznajo kandidate iz kontinuumskih procesov od signalnih kandidatov. Za ta namen potrebujemo spremenljivke, ki opazujejo sferine momente fizikalnih dogodkov, saj so le-ti zelo razlini med procesi $e^+e^- \to q\bar{q}$ in $e^+e^- \to B\bar{B}$.

V tretjem koraku se na podoben nain lotimo odstranjevanja ostalih kandidatov iz procesov $e^+e^- \to B\bar{B}$, za kar uporabimo vse ostale lastnosti kandidatov, razen $\Delta E$ and $M_{BC}$, ker le-te potrebujemo za luenje tevila signalnih kandidatov. Pri odstranjevanju ozadja te vrste uporabimo posebno metodo strojnega uenja, ki ohranja obliko porazdelitve spremenljivke $M_{BC}$ za ozadje, kar prepreuje, da bi optimizacija preoblikovala obliko porazdelitve ozadja tisti od signala.

Kot v prejnjih optimizacijah optimiziramo metriko FOM za odstranjevanje ozadja v drugem in tretjem koraku. Konni vzorec za signalni razpad je prikazan na Sliki X, za kontrolni razpad pa na Sliki X.

PLOT

PLOT

### 10.3.3 Luščenje fizikalnih parametrov

Po selekciji konnega vzorca lahko zanemo luiti fizikalne parametre iz podatkov. Za to uporabimo orodje RooFit [X], ki nam omogoa, da teoretini model prilagodimo izmerjenim podatkom, in na tak nain doloimo fizikalne parametre, ki jih iemo. Na podlagi MC vzorca doloimo porazdelitve za kandidate signalne kategorije in veih tipov ozadja. Te porazdelitve sluijo kot predloge, ki jih ustrezno setejemo skupaj, da dobimo teoretien model, ki dobro opie podatke.

Vsaka predloga je predstavljena kot 2-dimenzionalen histogram v spremenljivkah $\Delta E$ in $M_{BC}$, in sicer v $20\times$ razredih v zgoraj definiranem obmoju. Predloge posameznih kategorij so si med seboj razline, kar programu omogoa, da z visoko verjetnostjo pravilno doloi posamezne prispevke.

Luščenja parametrov se lotimo z metodo najveje zanesljivosti (ang. *maximum likelihood method*), saj nam omogoa zanesljiveje rezultate, ko so signalni vzorci majhni, kot v naem primeru. Za vsakega od 10 vzorcev MC podatkov izvrimo prilagajanje v namen preverjanja metode, na koncu pa enako ponovimo e na pravih podatkih. V nadaljevanju so prikazani rezultati prilagajanja za signalni in kontrolni razpad.

#### Kontrolni razpad

Fizikalne parametre v primeru kontrolnega razpada izluimo v ozkem oknu okoli mase mezona $D^0$. Pri postopku luenja uporabimo naslednje predloge

- kontrolni razpad,
- signalni razpad,
- kontinuumsko ozadje,
- ozadje razpada $B \to D^* \ell \nu,\ D^0 \to K^+ K^-$,
- ostalo $B\bar{B}$ ozadje.

Slika X prikazuje rezultate luenja na pravih podatkih in na vseh MC podatkih, Slika X pa prikazuje primer posameznega prilagajanja predlog MC podatkom. tevilo kandidatov kontrolnega razpada na podlagi luenja je

REZULTATI.

Pri luenju parametrov na pravih podatkih smo uporabili dodatno informacijo, in sicer meritev razpada $B \to D^* \ell \nu,\ D^0 \to K^+ K^-$ [X], ki smo jo uporabili v obliki omejitve vrednosti razmerja tevila kandidatov omenjenega in kontrolnega razpada.

Na podlagi konnega tevila kandidatov kontrolnega razpada $N$ lahko doloimo tudi razvejitveno razmerje, ki je doloeno kot

$$\mathcal{B} = \frac{N \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \tag{10.14}$$

kjer $\epsilon_{MC}$ predstavlja izkoristek kontrolnega razpada, doloenega na MC vzorcu, $\rho_{PID}$ je korekcijski faktor na raun razlike identifikacij nabitih delcev na MC in na pravih podatkih, $N_{B\bar{B}}$ pa je izmerjeno tevilo parov mezonov $B$. Razvejitveno razmerje lahko doloimo tako na podatkih kot na MC vzorcu, rezultati obeh pa so prikazani na Sliki X. Rezultati so konsistentni s priakovanimi in izmerjenimi vrednostmi, kar potrjuje zanesljivost nae analize.

PLOT

### Signalni razpad

Rezultati prilagajanja MC in pravim podatkov potrjujejo konsistentnost našega analiznega postopka, tako da jih lahko ponovimo še na signalnem razpadu. V tem primeru smo uporabili naslednje predloge

- signalni razpad,
- kontinuumsko ozadje,
- dobro poznana ozadja
  - $C_0: \; B^+ \to \bar{D}^0 \ell^+ \nu, \; D^0 \to K^- K^+$ (kontrolni razpad),
  - $C_1: \; B \to \bar{D}^* \ell^+ \nu, \; D^0 \to K^- K^+,$
  - $C_2: \; B \to \bar{D}^{(*)} \ell^+ \nu, \; D^0 \to K^- \pi^+,$
  - $C_3: \; B \to \bar{D}^{(*)} \ell^+ \nu, \; D^0 \to K^- K^+ \pi^0, \; K^- \pi^+ \pi^0,$
  - $C_4: \; B \to \bar{D}(*) \ell^+ \nu, \; D^0 \to K^- \ell^+ \nu,$
  - $C_5: \; B^0 \to D^{(*)-} \ell^+ \nu, \; D^+ \to K^- K^+ \pi^+, \; K^- \pi^+ \pi^+,$
  - $C_6:$ ostali $B \to \bar{D}^{(*)} \ell^+ \nu$ razpadi,
- ostalo $B\bar{B}$ ozadje.

V primeru dobro poznanih razpadov zopet uporabimo informacije o najnovejših meritveh in jih uporabimo za omejitev števila kandidatov posamezne kategorije. Pri prilagajanju predlog pravim podatkom smo uporabili kontorlni razpad za kalibracijo števila parov mezonov $B$. Slika X prikazuje rezultate luenja na pravih podatkih in na vseh MC podatkih, Slika X pa prikazuje primer posameznega prilagajanja predlog MC podatkom. tevilo kandidatov signalnega razpada na podlagi luenja je

REZULTATI.

## 10.4 Sistematske negotovosti

## 10.5 Končni rezultat

# Bibliography


[1] T. Sjöstrand, S. Mrenna, and P. Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (2006), p. 026.

[2] T. Keck. "Fastbdt: A speed-optimized multivariate classification algorithm for the belle ii experiment". In: *Computing and Software for Big Science* 1.1 (2017), p. 2.

[3] T. Keck. "Machine learning algorithms for the Belle II experiment and their validation on Belle data". Karlsruher Institut fr Technologie, Diss., 2017. Dr. Karlsruher Institut fr Technologie, 2017, 240 pages. URL: https://ekp-invenio.physik.uni-karlsruhe.de/record/48940.

[4] D. Asner et al. "Search for exclusive charmless hadronic B decays". In: *Physical Review D* 53.3 (1996), p. 1039.

[5] A. Bevan et al. "The physics of the B factories". In: *The European Physical Journal C* 74.11 (2014), p. 3026.

[6] P. del Amo Sanchez et al. "Study of $B \to \pi l \nu$ and $B \to \rho l \nu$ decays and determination of $|V_{ub}|$". In: *Phys. Rev. D* 83 (3 2011), p. 032007. DOI: 10.1103/PhysRevD.83.032007. URL: https://link.aps.org/doi/10.1103/PhysRevD.83.032007.

[7] J. Stevens and M. Williams. "uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers". In: *Journal of Instrumentation* 8.12 (2013), P12013.

[8] W. Verkerke and D. Kirkby. "The RooFit toolkit for data modeling". In: *Statistical Problems in Particle Physics, Astrophysics and Cosmology*. World Scientific, 2006, pp. 186–189.

[9] F. James. "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1". In: (1994).

[10] S. Ahn and J. A. Fessler. "Standard errors of mean, variance, and standard deviation estimators". In: *EECS Department, The University of Michigan* (2003), pp. 1–2.

[11] N. Cabibbo. "Unitary symmetry and leptonic decays". In: *Physical Review L* 10.12 (1963), p. 531.

[12] M. Kobayashi and T. Maskawa. "CP-violation in the renormalizable theory of weak interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657.

[13] Y. Amhis et al. "Averages of b-hadron, c-hadron, and -lepton properties as of summer 2016". In: *Eur. Phys. J.* C77 (2017). updated results and plots available at https://hflav.web.cern.ch, p. 895. DOI: 10.1140/epjc/s10052-017-5058-4. arXiv: 1612.07233 [hep-ex].

[14] P. Gambino et al. "Inclusive semileptonic B decays and the determination of —V(ub)—". In: *JHEP* 10 (2007), p. 058. DOI: 10.1088/1126-6708/2007/10/058. arXiv: 0707.2493 [hep-ph].

[15] M Tanabashi. "Review of particle physics". In: *Phys. Rev. D* 98 (2018), p. 030001.

[16] T. Abe et al. "Achievements of KEKB". In: *Progress of Theoretical and Experimental Physics* 2013.3 (2013), 03A001. DOI: 10.1093/ptep/pts102. eprint: /oup/backfile/content_public/journal/ptep/2013/3/10.1093/ptep/pts102/2/pts102.pdf. URL: http://dx.doi.org/10.1093/ptep/pts102.

[17] A. Abashian et al. "The Belle detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 479.1 (2002). Detectors for Asymmetric B-factories, pp. 117 –232. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)02013-7. URL: http://www.sciencedirect.com/science/article/pii/S0168900201020137.

[18] J Haba. "Letter of Intent for KEK Super B Factory, Part II: Detector". In: *KEK Report* (2004), pp. 04–4. URL: http://superb.kek.jp/documents/loi/img/LoI_detector.pdf.

[19] M. Paterno. *Calculating efficiencies and their uncertainties*. Tech. rep. Fermi National Accelerator Lab.(FNAL), Batavia, IL (United States), 2004.

[20] D. Scora and N. Isgur. "Semileptonic meson decays in the quark model: An update". In: *Phys. Rev.* D52 (1995), pp. 2783–2812. DOI: 10.1103/PhysRevD.52.2783. arXiv: hep-ph/9503486 [hep-ph].

[21] D. Lange. "DJ Lange, Nucl. Instrum. Methods Phys. Res., Sect. A 462, 152 (2001)." In: *Nucl. Instrum. Methods Phys. Res., Sect. A* 462 (2001), p. 152.

[22] L. Wolfenstein. "Parametrization of the Kobayashi-Maskawa Matrix". In: *Phys. Rev. Lett.* 51 (21 1983), pp. 1945–1947. DOI: 10.1103/PhysRevLett.51.1945. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1945.

# A: MVA Control Plots

## ROE Clean-up $\pi^0$ Training

### Variable Importance

| | Name | Alias | Importance |
|---|---|---|---|
| 0 | `chiProb` | $v_0$ | 0.280 |
| 1 | `useCMSFrame(daughterAngleInBetween(0,1))` | $v_1$ | 0.203 |
| 2 | `daughter(0,useCMSFrame(p))` | $v_2$ | 0.073 |
| 3 | `InvM` | $v_3$ | 0.072 |
| 4 | `daughter(1,clusterHighestE)` | $v_4$ | 0.061 |
| 5 | `daughter(1,clusterTheta)` | $v_5$ | 0.049 |
| 6 | `daughter(1,p)` | $v_6$ | 0.047 |
| 7 | `daughter(0,clusterHighestE)` | $v_7$ | 0.029 |
| 8 | `daughter(0,clusterTheta)` | $v_8$ | 0.024 |
| 9 | `daughter(0,clusterE9E25)` | $v_9$ | 0.018 |
| 10 | `daughter(0,minC2HDist)` | $v_{10}$ | 0.018 |
| 11 | `daughter(1,minC2HDist)` | $v_{11}$ | 0.017 |
| 12 | `daughter(1,clusterE9E25)` | $v_{12}$ | 0.016 |
| 13 | `useRestFrame(daughterAngleInBetween(0,1))` | $v_{13}$ | 0.014 |
| 14 | `daughter(1,clusterNHits)` | $v_{14}$ | 0.013 |
| 15 | `daughter(0,clusterNHits)` | $v_{15}$ | 0.011 |
| 16 | `daughter(0,clusterErrorE)` | $v_{16}$ | 0.009 |
| 17 | `daughter(1,clusterErrorE)` | $v_{17}$ | 0.009 |
| 18 | `SigMBF` | $v_{18}$ | 0.007 |
| 19 | `useCMSFrame(p)` | $v_{19}$ | 0.006 |
| 20 | `daughter(0,p)` | $v_{20}$ | 0.005 |
| 21 | `SigM` | $v_{21}$ | 0.005 |
| 22 | `daughter(1,useCMSFrame(p))` | $v_{22}$ | 0.005 |
| 23 | `useLabFrame(daughterAngleInBetween(0,1))` | $v_{23}$ | 0.005 |
| 24 | `p` | $v_{24}$ | 0.003 |

Table 10.1: Variable names, aliases and importance in the scope of $\pi^0$ MVA training for ROE clean-up.

## Variable Distributions

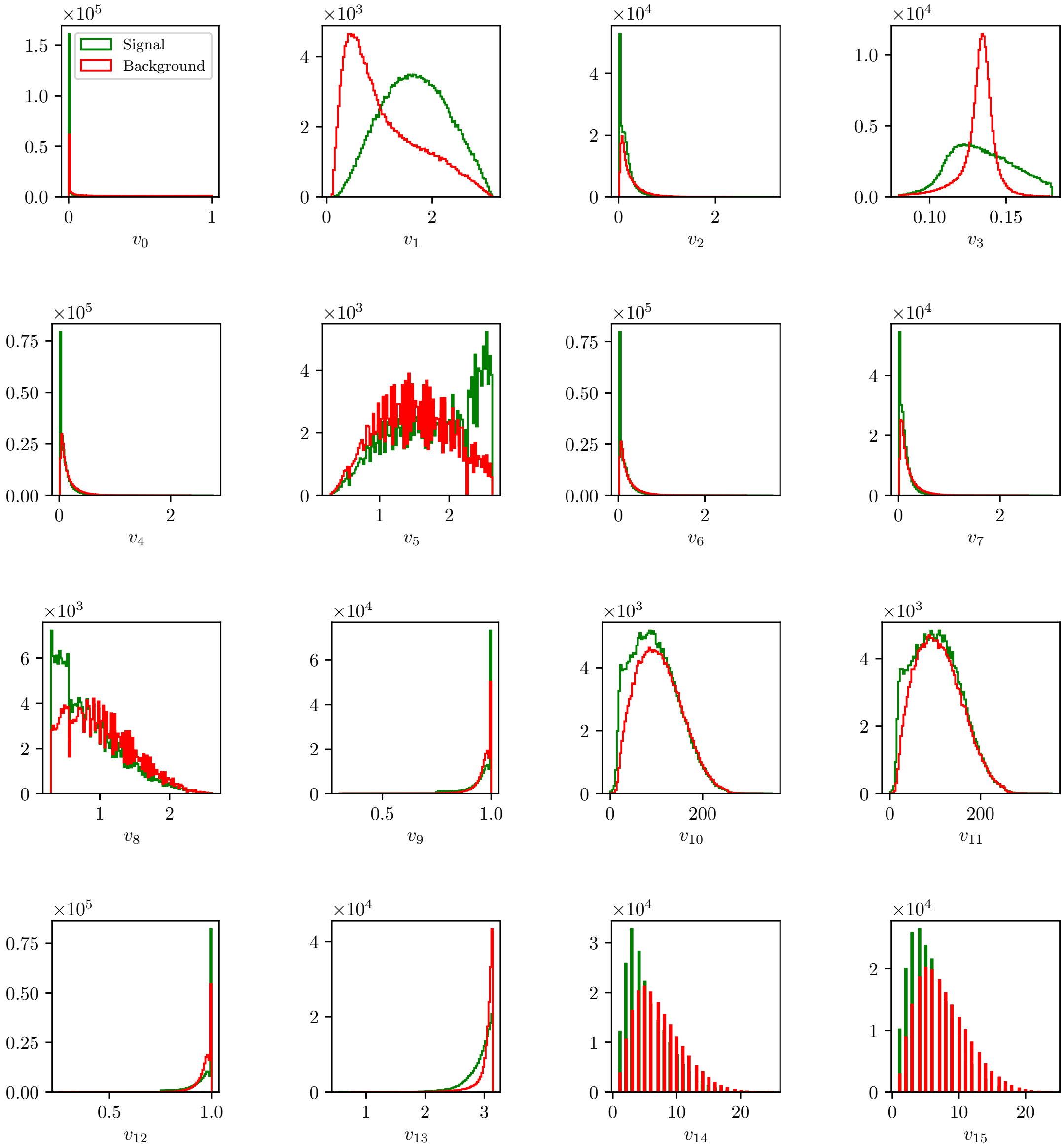



Figure 10.3: Feature distributions for MVA training of $\pi^0$ candidates in the scope of ROE clean-up.

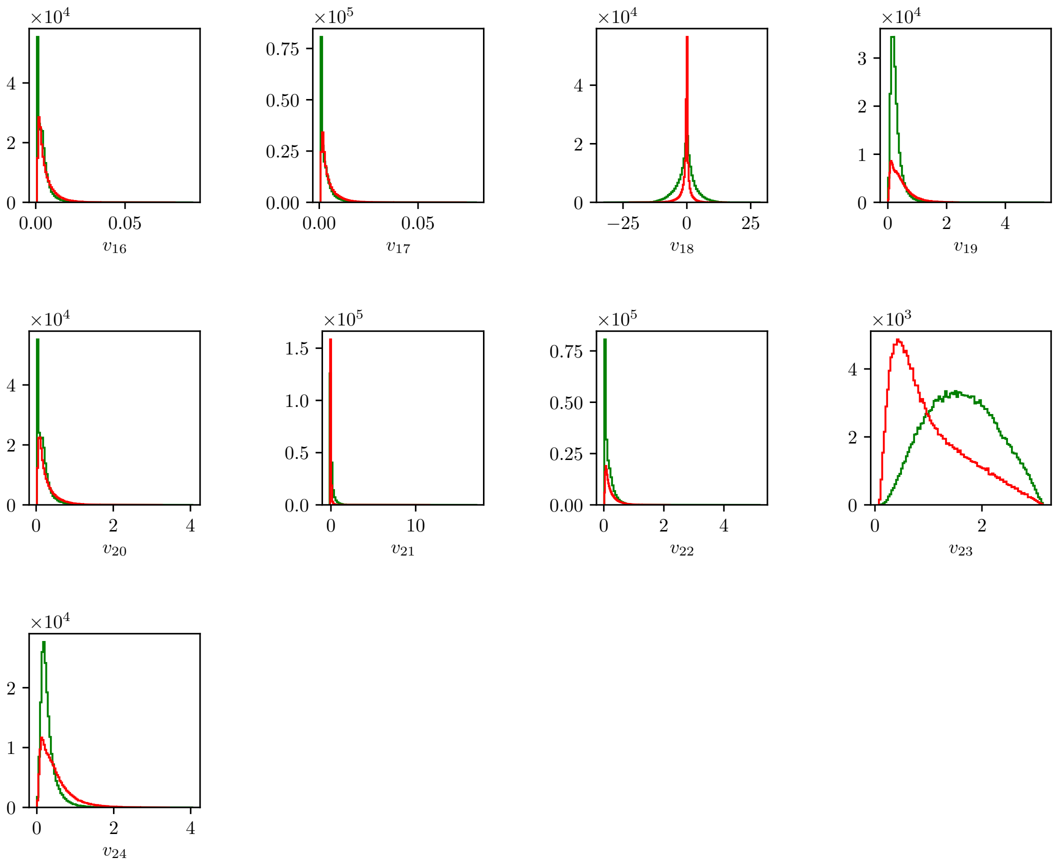


Figure 10.3: Feature distributions for MVA training of $\pi^0$ candidates in the scope of ROE clean-up.

## Hyper-parameter Optimization

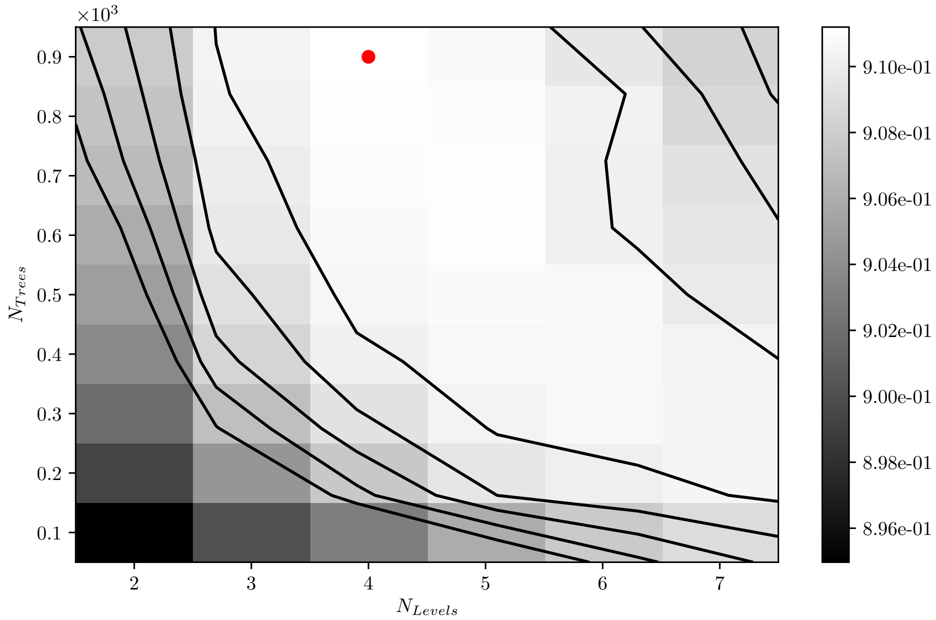



Figure 10.4: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $\pi^0$ candidates in the scope of the ROE clean-up.

## Results

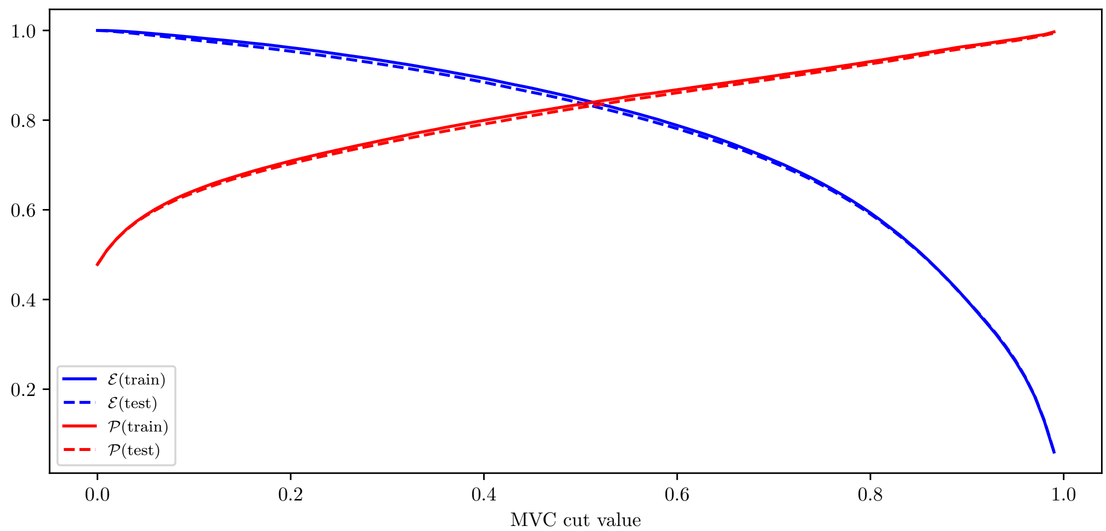



Figure 10.5: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for $\pi^0$ candidates training on the train (solid) and test (dashed) samples.

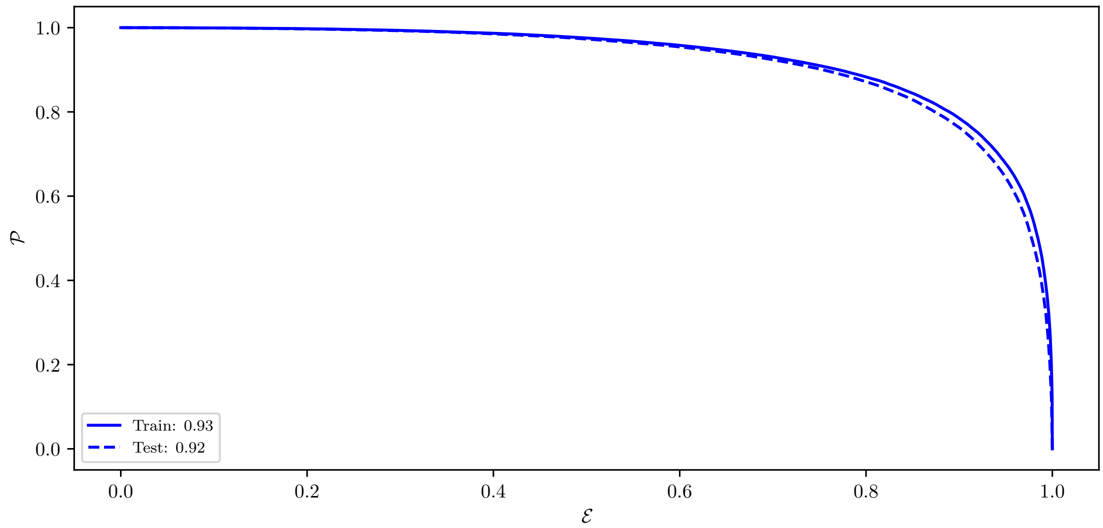



Figure 10.6: ROC curves of the MVA classifier output for $\pi^0$ candidates training on the train (solid) and test (dashed) samples.

# ROE Clean-up $\gamma$ Training

## Variable Importance

| | Name | Alias | Importance |
|---|---|---|---|
| 0 | `p` | $v_0$ | 0.327 |
| 1 | `pi0p` | $v_1$ | 0.243 |
| 2 | `clusterHighestE` | $v_2$ | 0.226 |
| 3 | `minC2HDist` | $v_3$ | 0.052 |
| 4 | `cosTheta` | $v_4$ | 0.036 |
| 5 | `clusterE9E25` | $v_5$ | 0.031 |
| 6 | `clusterNHits` | $v_6$ | 0.025 |
| 7 | `clusterUncorrE` | $v_7$ | 0.022 |
| 8 | `clusterR` | $v_8$ | 0.015 |
| 9 | `useCMSFrame(p)` | $v_9$ | 0.013 |
| 10 | `clusterErrorE` | $v_{10}$ | 0.010 |
| 11 | `clusterReg` | $v_{11}$ | 0.000 |

Table 10.2: Variable names, aliases and importance in the scope of $\gamma$ MVA training for ROE clean-up.

## Variable Distributions

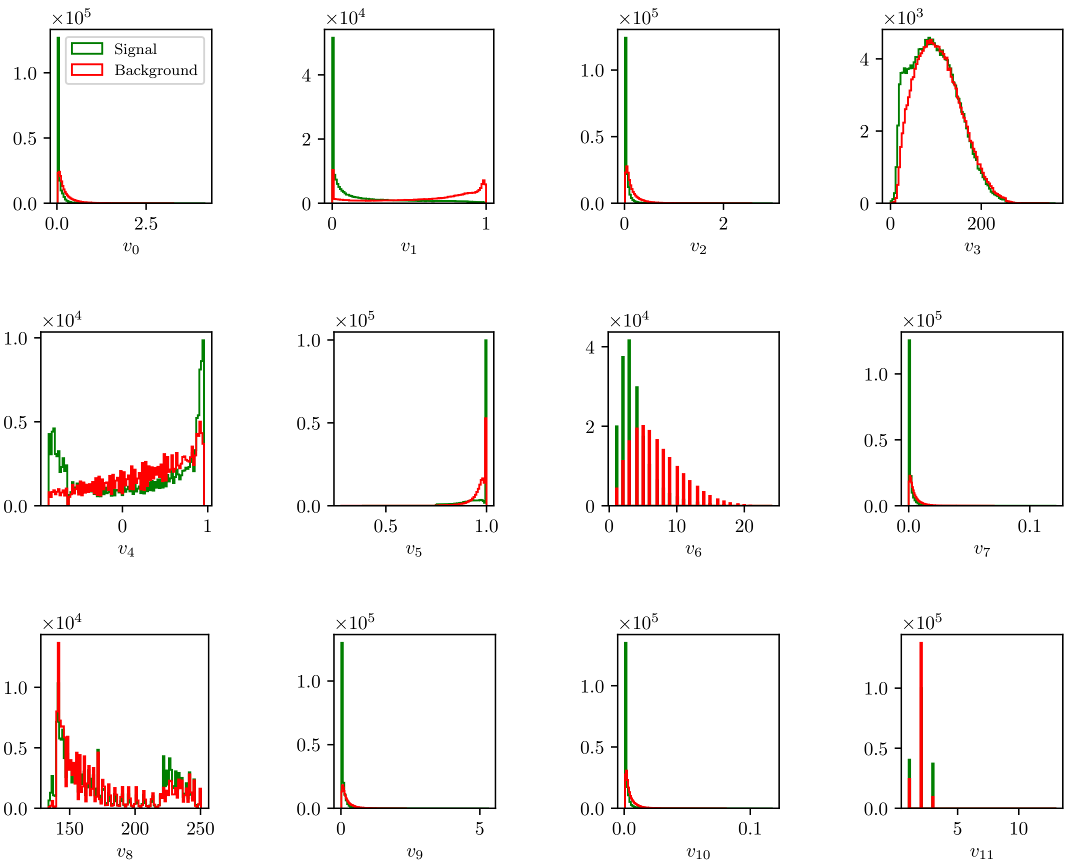



Figure 10.7: Feature distributions for MVA training of $\gamma$ candidates in the scope of ROE clean-up.

## Hyper-parameter Optimization

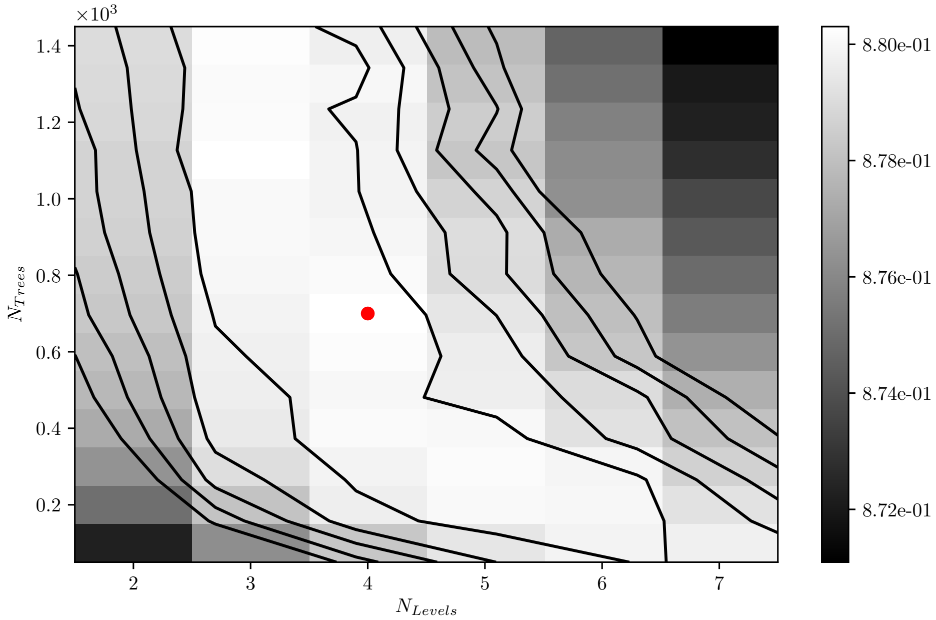



Figure 10.8: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $\gamma$ candidates in the scope of the ROE clean-up.

## Results

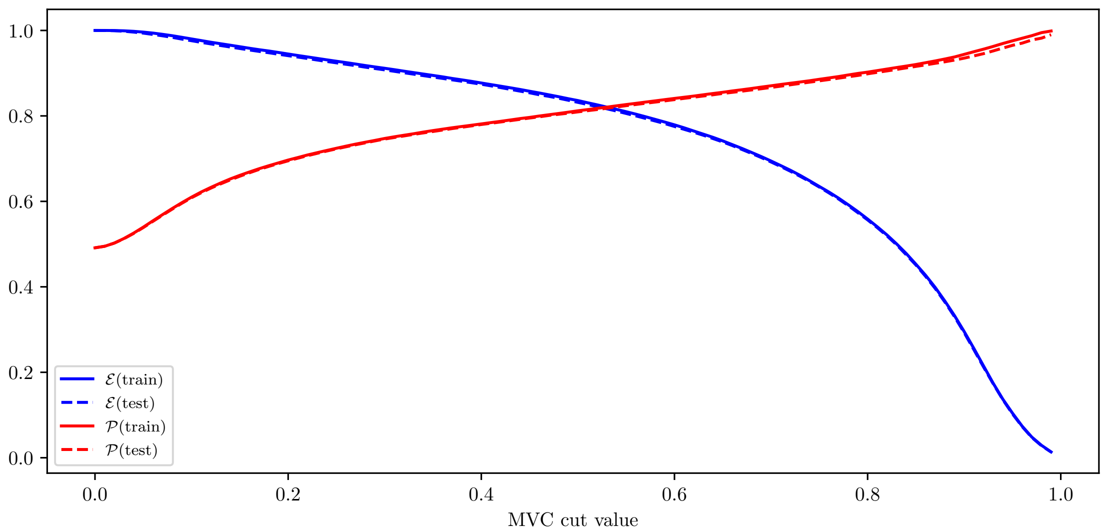



Figure 10.9: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for $\gamma$ candidates training on the train (solid) and test (dashed) samples.

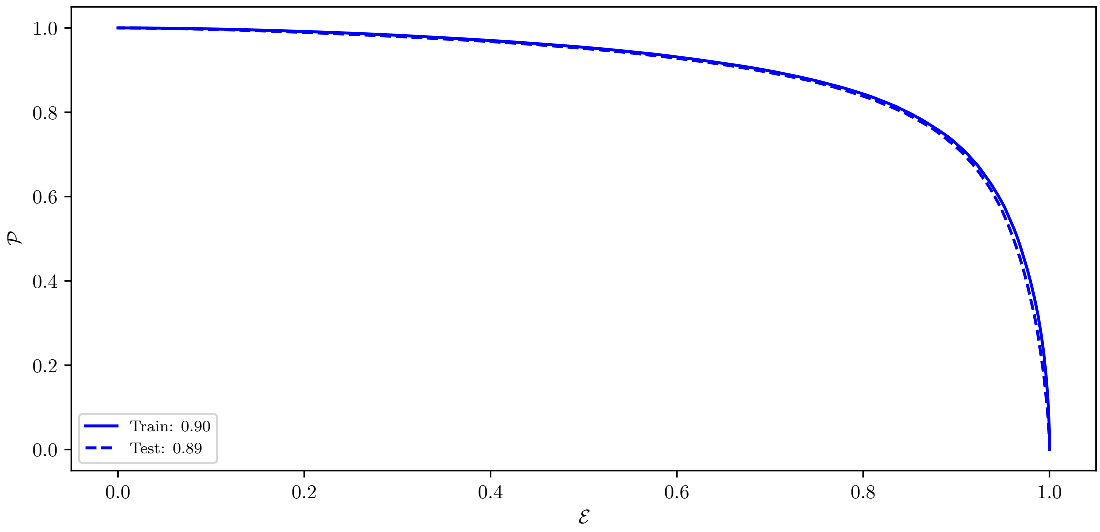



Figure 10.10: ROC curves of the MVA classifier output for $\gamma$ candidates training on the train (solid) and test (dashed) samples.

# ROE Clean-up Duplicate Pair Training

## Variable Importance

| | Name | Alias | Importance |
|---|---|---|---|
| 0 | `useCMSFrame(daughterAngleInBetween(0,1))` | $v_0$ | 0.132 |
| 1 | `daughter(0,phi0Err)` | $v_1$ | 0.082 |
| 2 | `useLabFrame(daughterAngleInBetween(0,1))` | $v_2$ | 0.055 |
| 3 | `daughter(1,d0)` | $v_3$ | 0.051 |
| 4 | `daughter(1,phi0Err)` | $v_4$ | 0.051 |
| 5 | `daughter(0,d0)` | $v_5$ | 0.050 |
| 6 | `daughter(1,nCDCHits)` | $v_6$ | 0.040 |
| 7 | `daughter(1,d0Err)` | $v_7$ | 0.037 |
| 8 | `daughter(0,nCDCHits)` | $v_8$ | 0.034 |
| 9 | `daughter(1,z0)` | $v_9$ | 0.032 |
| 10 | `daughter(0,z0)` | $v_{10}$ | 0.030 |
| 11 | `daughter(0,d0Err)` | $v_{11}$ | 0.028 |
| 12 | `daughter(0,nSVDHits)` | $v_{12}$ | 0.028 |
| 13 | `daughter(1,pz)` | $v_{13}$ | 0.027 |
| 14 | `daughter(1,useCMSFrame(p))` | $v_{14}$ | 0.024 |
| 15 | `extraInfo(decayModeID)` | $v_{15}$ | 0.023 |
| 16 | `daughter(0,pz)` | $v_{16}$ | 0.020 |
| 17 | `daughter(1,nSVDHits)` | $v_{17}$ | 0.020 |
| 18 | `daughter(0,pValue)` | $v_{18}$ | 0.020 |
| 19 | `daughter(1,tanlambda)` | $v_{19}$ | 0.018 |
| 20 | `daughter(1,pValue)` | $v_{20}$ | 0.018 |
| 21 | `daughter(0,tanlambda)` | $v_{21}$ | 0.017 |
| 22 | `daughter(0,phi0)` | $v_{22}$ | 0.016 |

| | | | |
|---|---|---|---|
| 23 | `daughter(1,phi0)` | $v_{23}$ | 0.016 |
| 24 | `daughter(0,useCMSFrame(p))` | $v_{24}$ | 0.015 |
| 25 | `daughter(0,z0Err)` | $v_{25}$ | 0.014 |
| 26 | `daughter(1,omega)` | $v_{26}$ | 0.013 |
| 27 | `daughter(0,omega)` | $v_{27}$ | 0.013 |
| 28 | `daughter(1,z0Err)` | $v_{28}$ | 0.012 |
| 29 | `daughter(0,pt)` | $v_{29}$ | 0.011 |
| 30 | `daughter(0,omegaErr)` | $v_{30}$ | 0.011 |
| 31 | `daughter(1,omegaErr)` | $v_{31}$ | 0.010 |
| 32 | `daughter(1,pt)` | $v_{32}$ | 0.009 |
| 33 | `daughter(0,tanlambdaErr)` | $v_{33}$ | 0.009 |
| 34 | `daughter(1,tanlambdaErr)` | $v_{34}$ | 0.009 |
| 35 | `useRestFrame(daughterAngleInBetween(0,1))` | $v_{35}$ | 0.003 |
| 36 | `daughter(1,charge)` | $v_{36}$ | 0.000 |
| 37 | `daughter(0,charge)` | $v_{37}$ | 0.000 |

Table 10.3: Variable names, aliases and importance in the scope of duplicate track pair MVA training for ROE clean-up.

# Variable Distributions

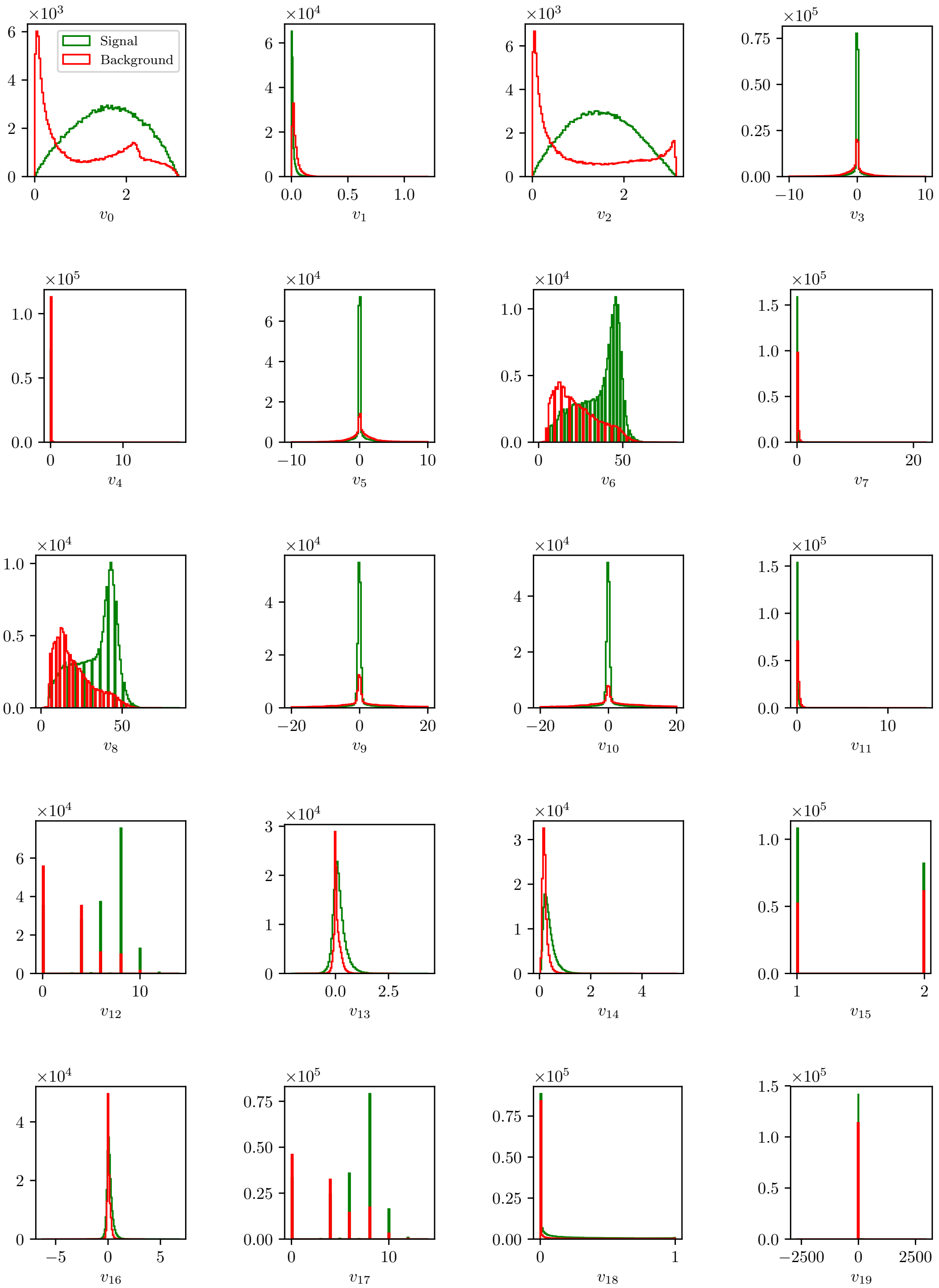



Figure 10.11: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

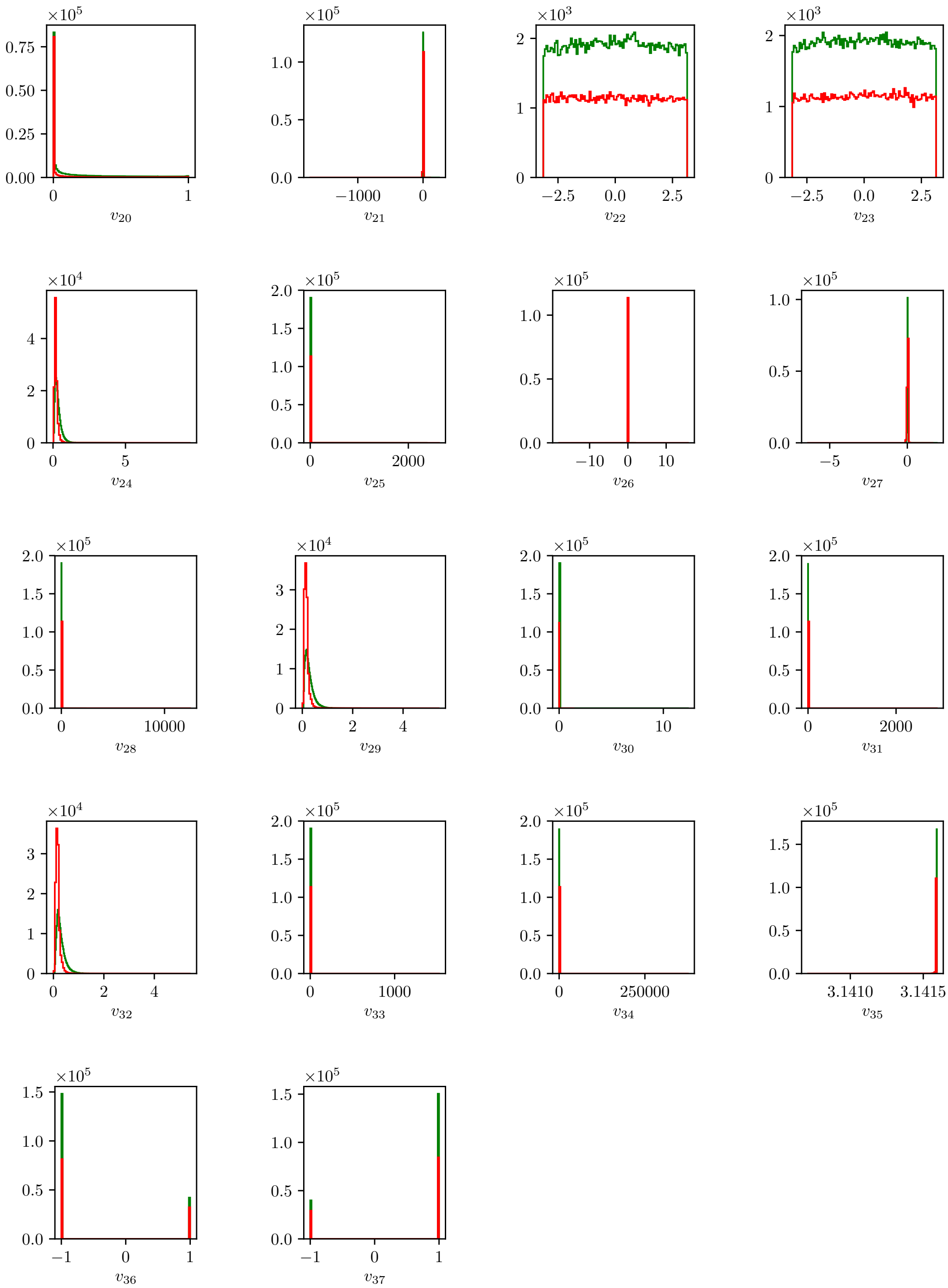



Figure 10.11: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

## Hyper-parameter Optimization

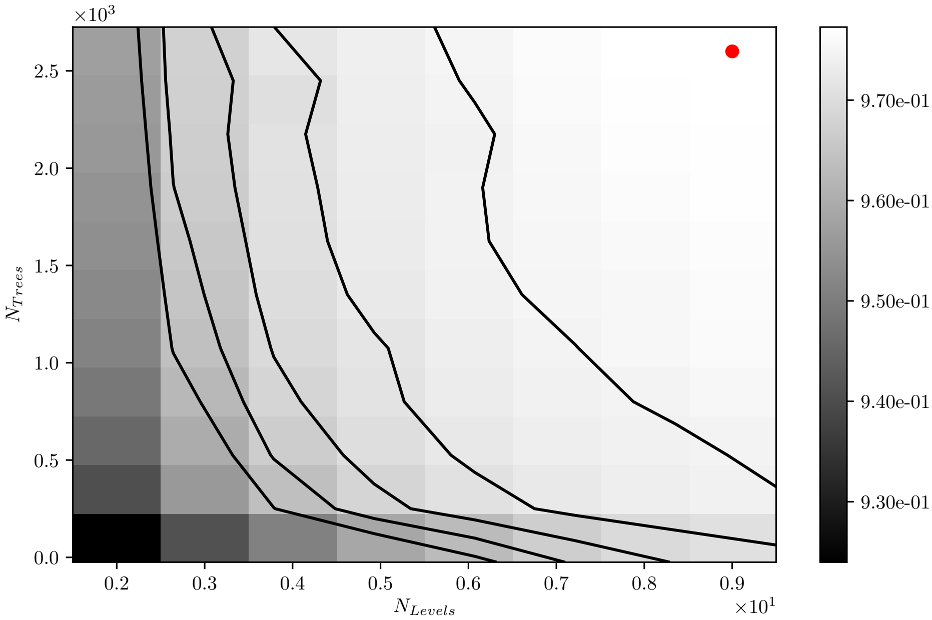



Figure 10.12: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of duplicate track pair candidates in the scope of the ROE clean-up.

## Results

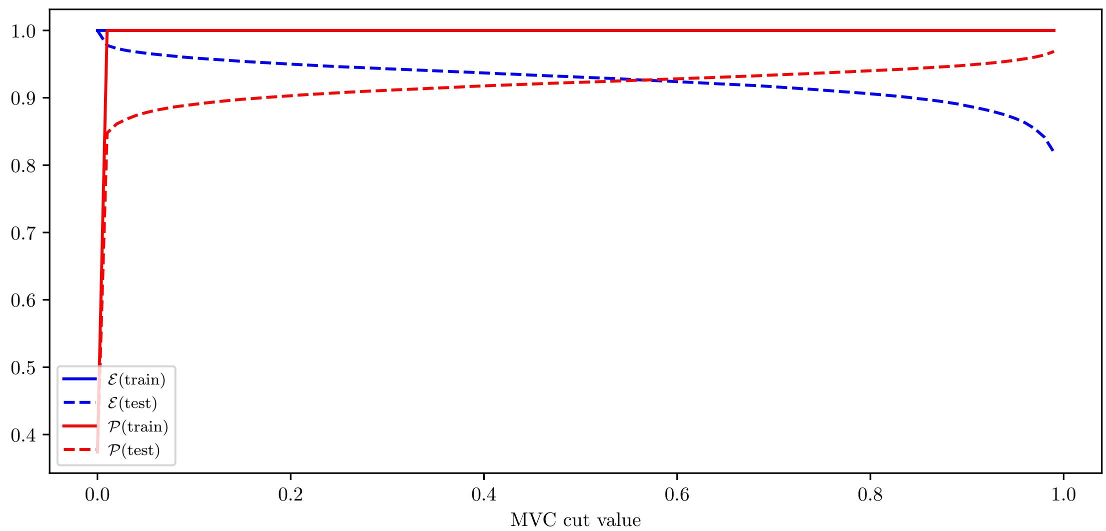



Figure 10.13: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

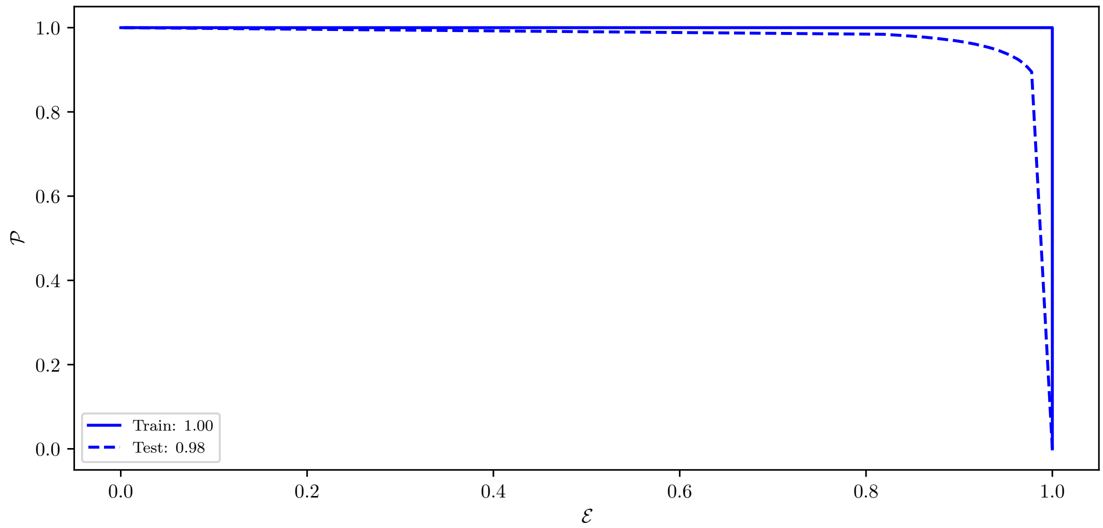



Figure 10.14: ROC curves of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

# ROE Clean-up Duplicate Track Training

## Variable Importance

| | Name | Alias | Importance |
|---|---|---|---|
| 0 | `extraInfo(d0Diff)` | $v_0$ | 0.214 |
| 1 | `extraInfo(z0Diff)` | $v_1$ | 0.087 |
| 2 | `d0` | $v_2$ | 0.069 |
| 3 | `extraInfo(pValueDiff)` | $v_3$ | 0.060 |
| 4 | `z0` | $v_4$ | 0.058 |
| 5 | `phi0Err` | $v_5$ | 0.056 |
| 6 | `extraInfo(pzDiff)` | $v_6$ | 0.055 |
| 7 | `extraInfo(ptDiff)` | $v_7$ | 0.045 |
| 8 | `z0Err` | $v_8$ | 0.043 |
| 9 | `extraInfo(nCDCHitsDiff)` | $v_9$ | 0.037 |
| 10 | `extraInfo(nSVDHitsDiff)` | $v_{10}$ | 0.034 |
| 11 | `pt` | $v_{11}$ | 0.032 |
| 12 | `d0Err` | $v_{12}$ | 0.030 |
| 13 | `pValue` | $v_{13}$ | 0.029 |
| 14 | `nCDCHits` | $v_{14}$ | 0.028 |
| 15 | `nSVDHits` | $v_{15}$ | 0.028 |
| 16 | `pz` | $v_{16}$ | 0.025 |
| 17 | `cosTheta` | $v_{17}$ | 0.024 |
| 18 | `phi0` | $v_{18}$ | 0.023 |
| 19 | `useCMSFrame(p)` | $v_{19}$ | 0.021 |

Table 10.4: Variable names, aliases and importance in the scope of duplicate track MVA training for ROE clean-up.

# Variable Distributions

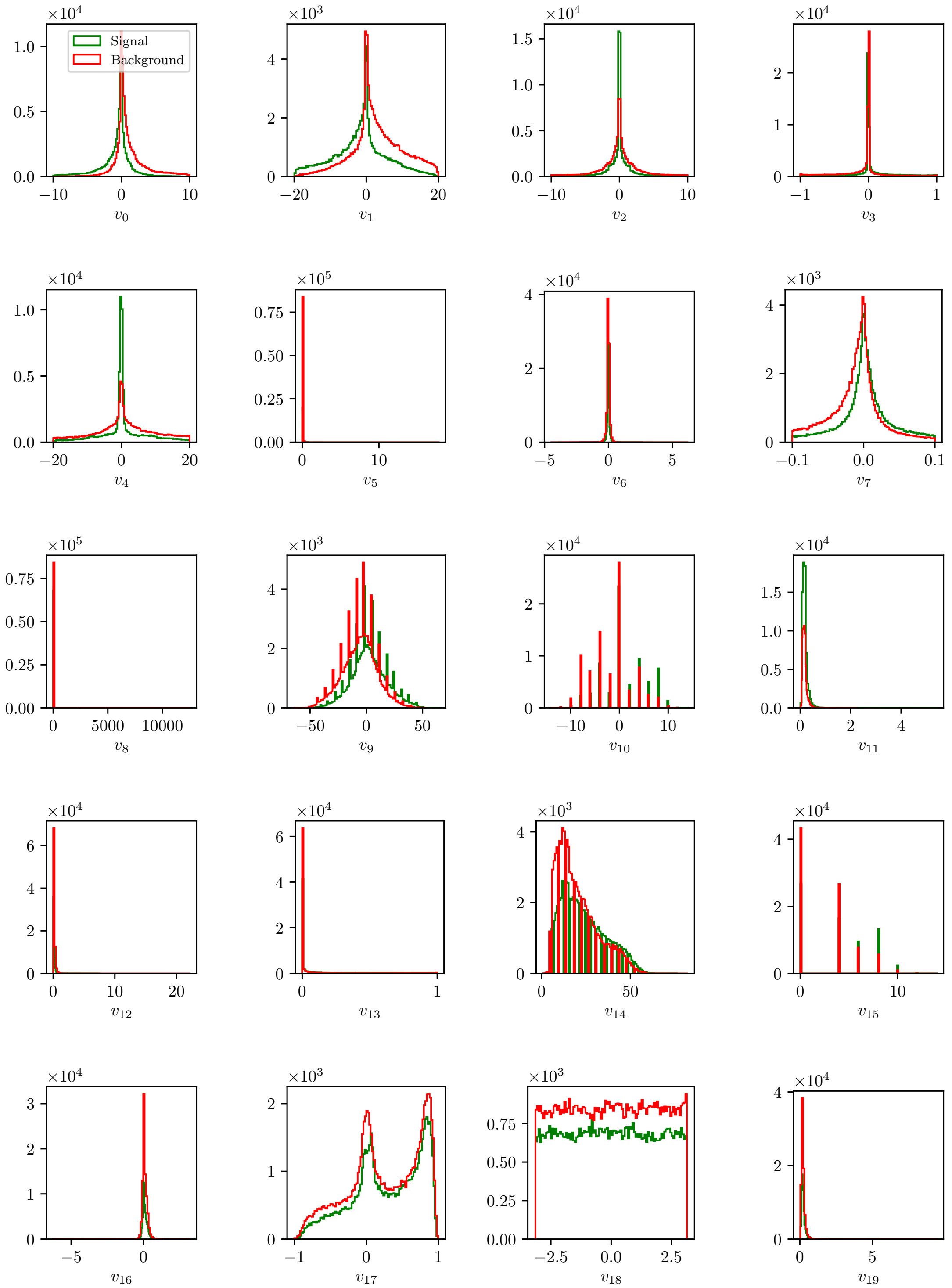



Figure 10.15: Feature distributions for MVA training of duplicate track candidates in the scope of ROE clean-up.

## Hyper-parameter Optimization

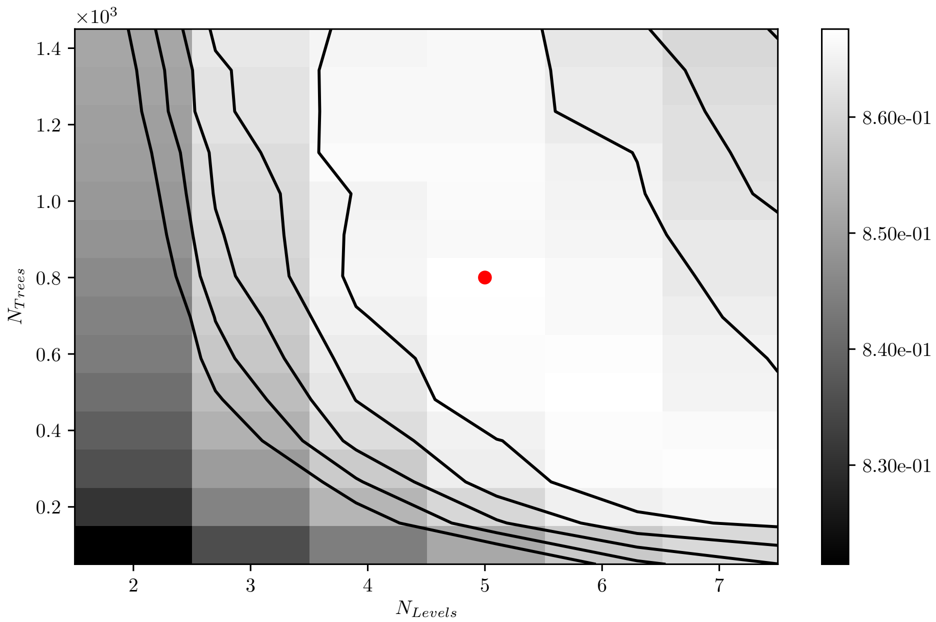



Figure 10.16: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of duplicate track candidates in the scope of the ROE clean-up.

## Results

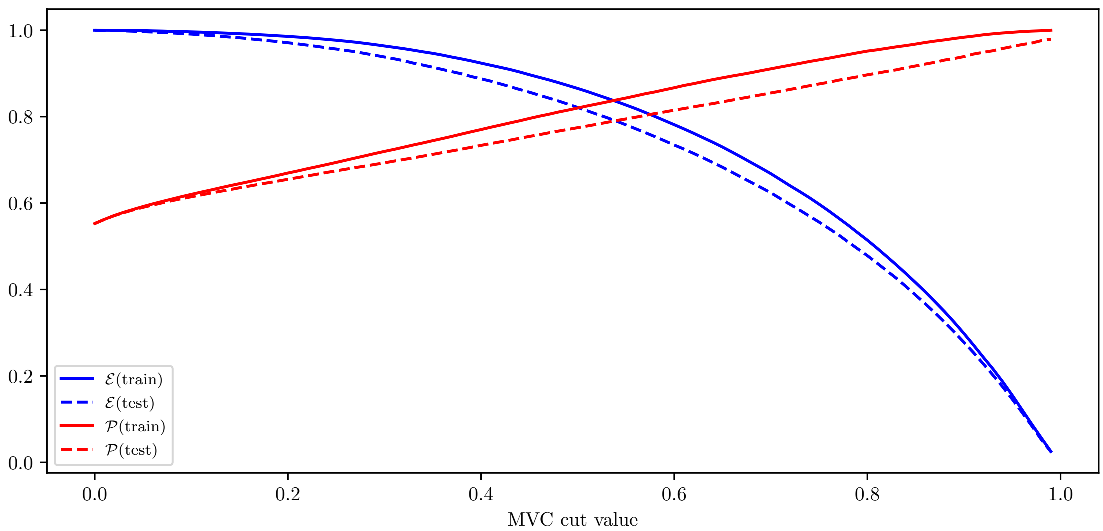



Figure 10.17: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.

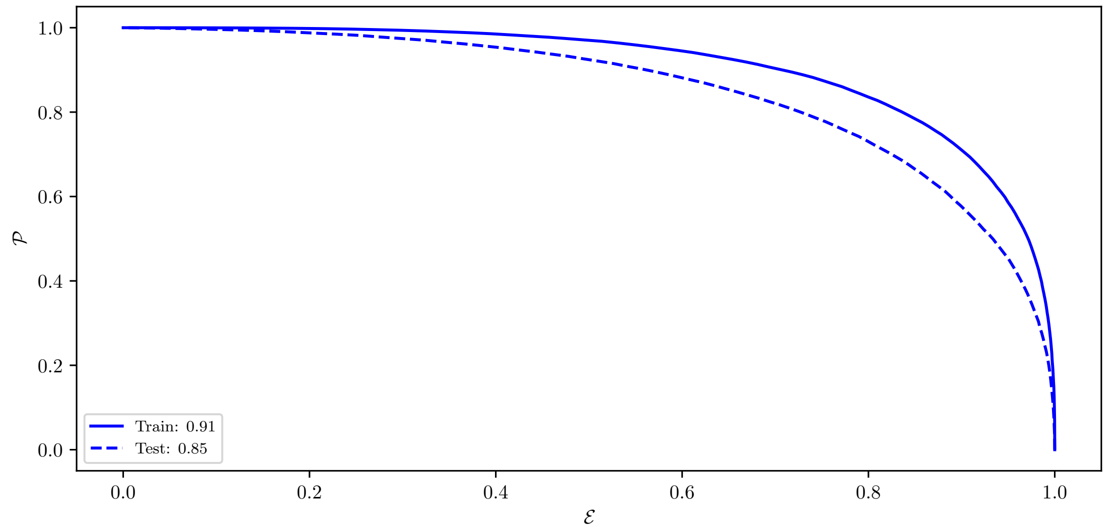



Figure 10.18: ROC curves of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.

# $q\bar{q}$ Suppression Training

## Variable Importance

| | Name | Alias | Importance |
|---|---|---|---|
| 0 | `B_CosTBTO` | $v_0$ | 0.291 |
| 1 | `B_hso02` | $v_1$ | 0.096 |
| 2 | `B_ThrustB` | $v_2$ | 0.096 |
| 3 | `B_roeFit_dz` | $v_3$ | 0.075 |
| 4 | `B_R2` | $v_4$ | 0.054 |
| 5 | `B_hso12` | $v_5$ | 0.044 |
| 6 | `B_hoo2` | $v_6$ | 0.032 |
| 7 | `B_ThrustO` | $v_7$ | 0.027 |
| 8 | `B_qpKaon` | $v_8$ | 0.024 |
| 9 | `B_cc2_CcROE` | $v_9$ | 0.023 |
| 10 | `B_hoo0` | $v_{10}$ | 0.019 |
| 11 | `B_cc3_CcROE` | $v_{11}$ | 0.019 |
| 12 | `B_cc4_CcROE` | $v_{12}$ | 0.016 |
| 13 | `B_CosTBz` | $v_{13}$ | 0.015 |
| 14 | `B_hso01` | $v_{14}$ | 0.015 |
| 15 | `B_cc1_CcROE` | $v_{15}$ | 0.015 |
| 16 | `B_cc5_CcROE` | $v_{16}$ | 0.013 |
| 17 | `B_cc6_CcROE` | $v_{17}$ | 0.012 |
| 18 | `B_qpFastHadron` | $v_{18}$ | 0.012 |
| 19 | `B_cc7_CcROE` | $v_{19}$ | 0.010 |
| 20 | `B_cc9_CcROE` | $v_{20}$ | 0.010 |
| 21 | `B_cc8_CcROE` | $v_{21}$ | 0.010 |

| | | | |
|---|---|---|---|
| 22 | `B_qpMaximumPstar` | $v_{22}$ | 0.008 |
| 23 | `B_hso10` | $v_{23}$ | 0.008 |
| 24 | `B_hso04` | $v_{24}$ | 0.007 |
| 25 | `B_qpLambda` | $v_{25}$ | 0.006 |
| 26 | `B_hoo1` | $v_{26}$ | 0.006 |
| 27 | `B_qpKaonPion` | $v_{27}$ | 0.006 |
| 28 | `B_hoo4` | $v_{28}$ | 0.006 |
| 29 | `B_qpSlowPion` | $v_{29}$ | 0.006 |
| 30 | `B_hso03` | $v_{30}$ | 0.005 |
| 31 | `B_hso14` | $v_{31}$ | 0.004 |
| 32 | `B_qpFSC` | $v_{32}$ | 0.004 |
| 33 | `B_hoo3` | $v_{33}$ | 0.004 |

Table 10.5: Variable names, aliases and importance in the scope of $q\bar{q}$ suppression MVA training.

# Variable Distributions

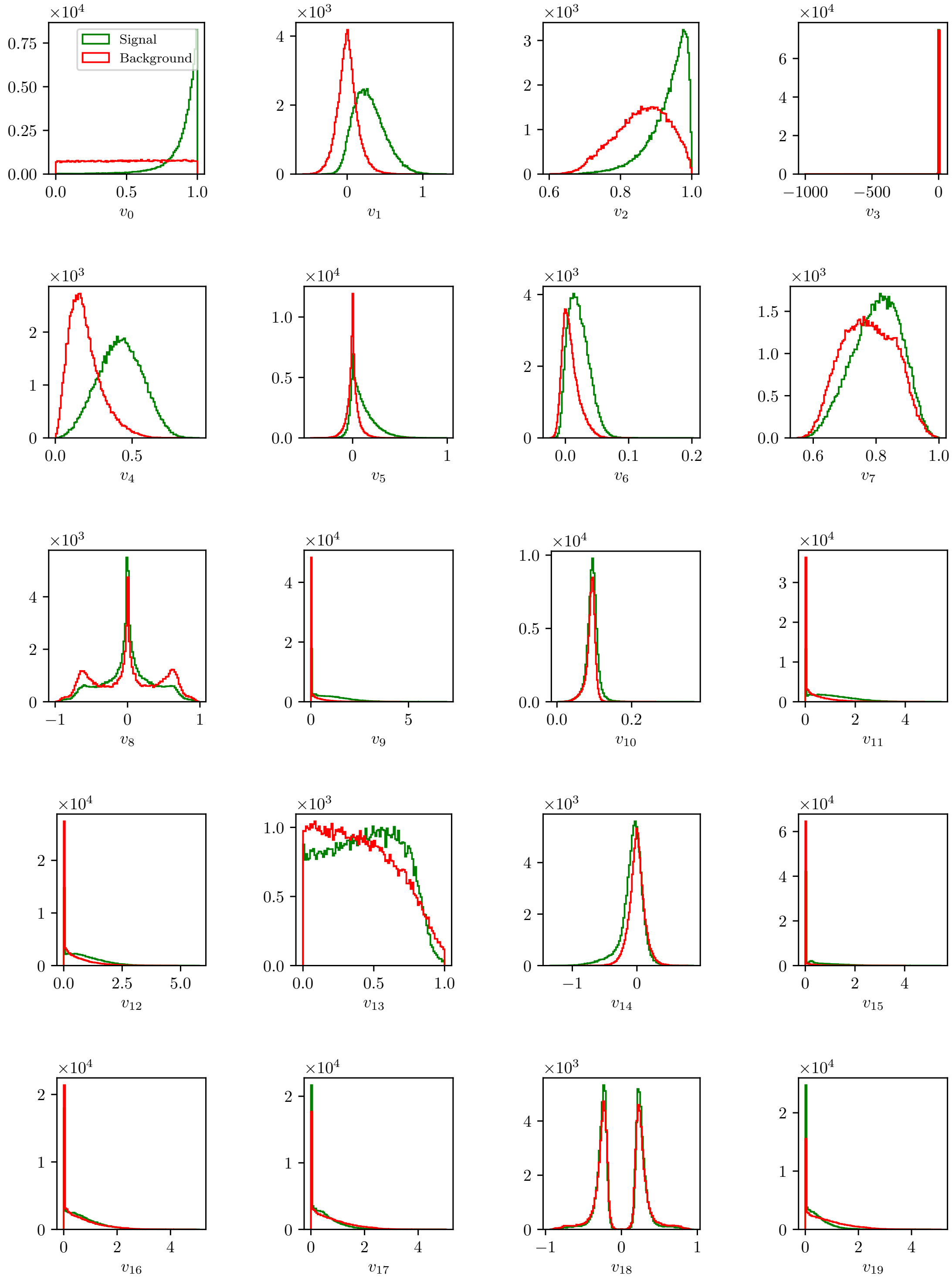



Figure 10.19: Feature distributions for MVA training of $q\bar{q}$ background suppression.

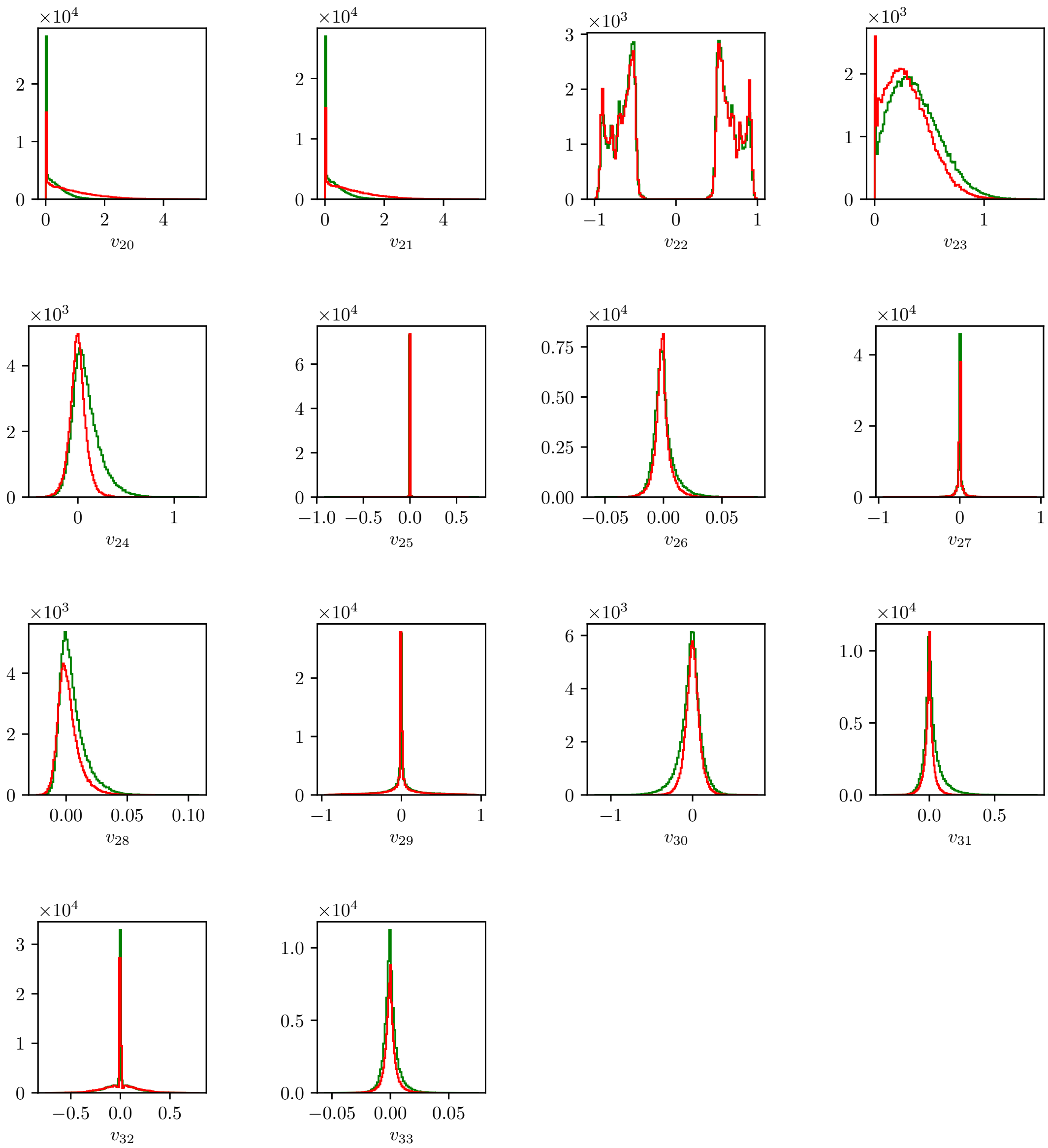



Figure 10.19: Feature distributions for MVA training of $q\bar{q}$ background suppression.

## Hyper-parameter Optimization

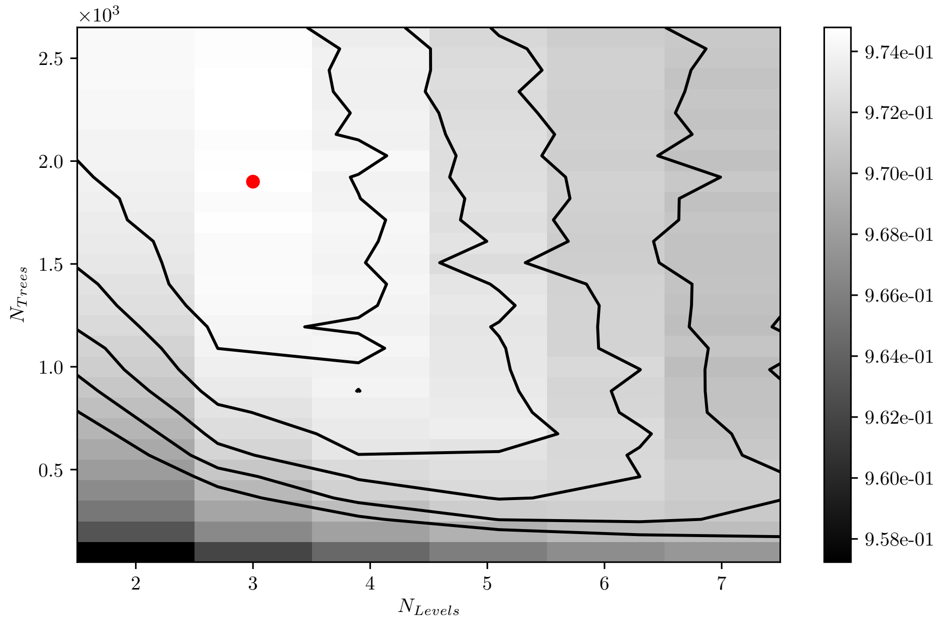



Figure 10.20: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $q\bar{q}$ background suppression.

## Results

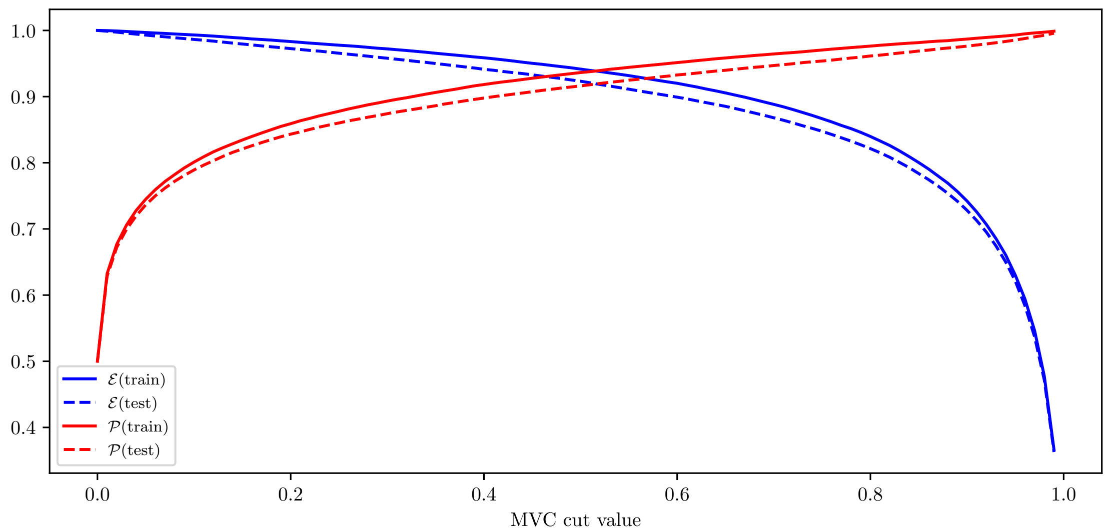



Figure 10.21: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples.

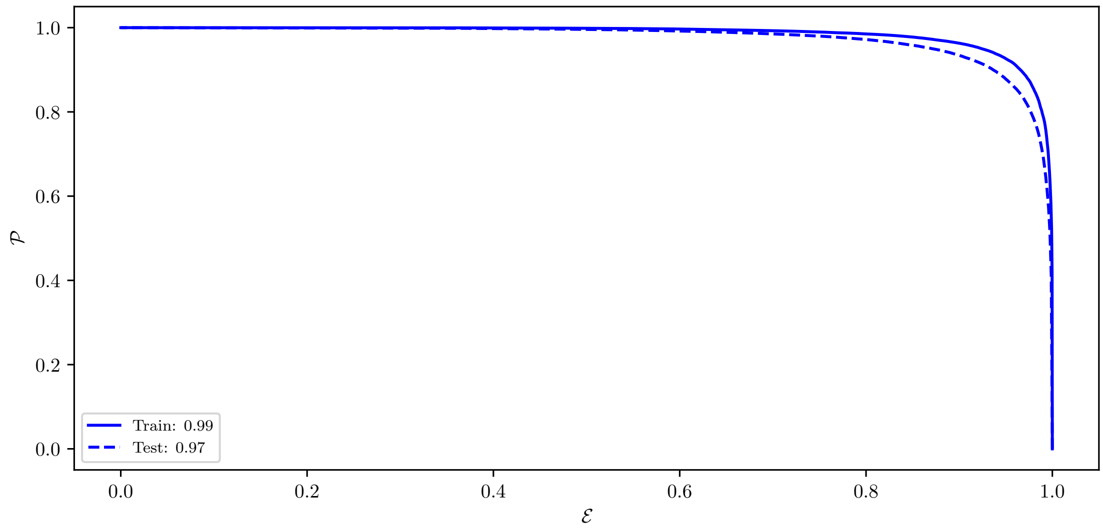



Figure 10.22: ROC curves of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples.

# Standard $B\bar{B}$ Suppression Training

## Variable Importance

| | Name | Alias | Importance |
|---|---|---|---|
| 0 | `B_cosMomVtxKKlnu` | $v_0$ | 0.372 |
| 1 | `B_ROE_PThetacms0` | $v_1$ | 0.096 |
| 2 | `B_nROETrk0` | $v_2$ | 0.079 |
| 3 | `B_K1FT` | $v_3$ | 0.063 |
| 4 | `B_cosBY` | $v_4$ | 0.051 |
| 5 | `B_roeFit_dz` | $v_5$ | 0.047 |
| 6 | `B_xiZ0` | $v_6$ | 0.043 |
| 7 | `B_cosMomVtx` | $v_7$ | 0.038 |
| 8 | `B_chiProb` | $v_8$ | 0.031 |
| 9 | `B_nKaonInROE` | $v_9$ | 0.028 |
| 10 | `B_missM2Veto1` | $v_{10}$ | 0.026 |
| 11 | `B_missM2Veto2` | $v_{11}$ | 0.021 |
| 12 | `B_nROEDistTrk` | $v_{12}$ | 0.018 |
| 13 | `B_cosMomVtxKK` | $v_{13}$ | 0.018 |
| 14 | `B_K0FT` | $v_{14}$ | 0.017 |
| 15 | `B_QVeto1` | $v_{15}$ | 0.016 |
| 16 | `B_missM20` | $v_{16}$ | 0.015 |
| 17 | `B_TagVPvalue` | $v_{17}$ | 0.012 |
| 18 | `B_QVeto2` | $v_{18}$ | 0.010 |

Table 10.6: Variable names, aliases and importance in the scope of $B\bar{B}$ background suppression.

## Variable Distributions

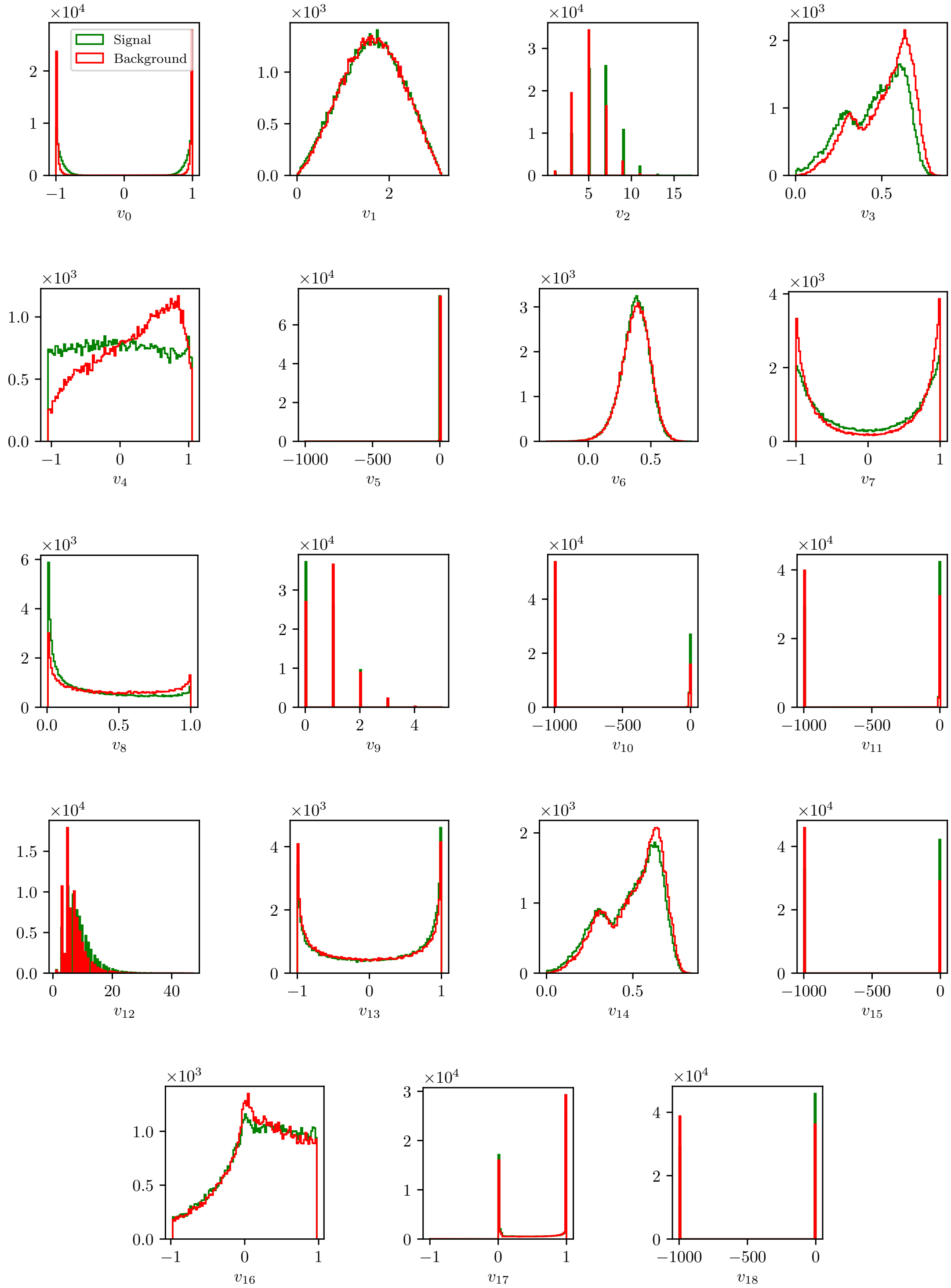



Figure 10.23: Feature distributions for MVA training of $B\bar{B}$ background suppression.

## Hyper-parameter Optimization

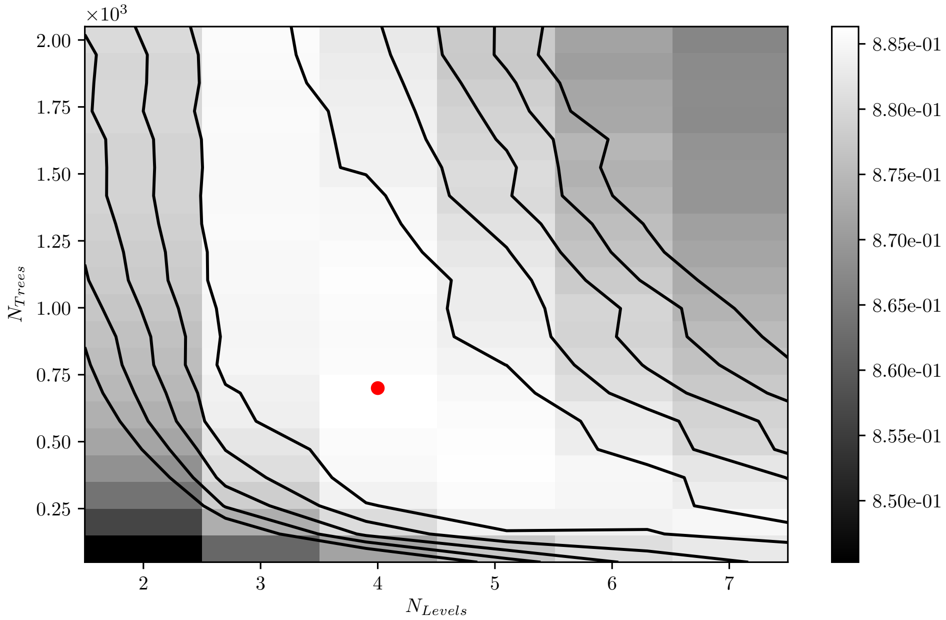



Figure 10.24: Hyper-parameter optimization of `nTrees` and `nLevels` in the BDT forest training of $B\bar{B}$ background suppression.

## Results

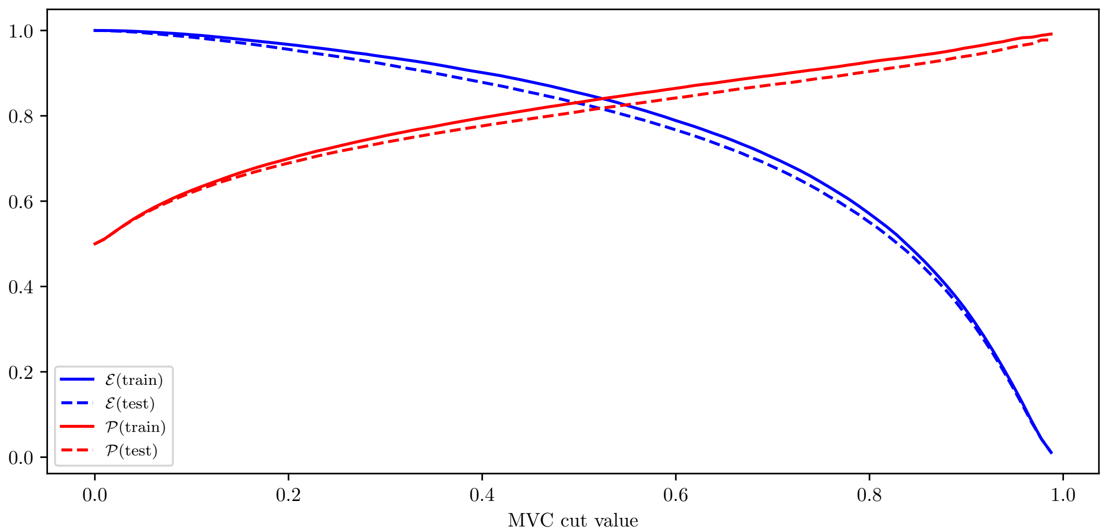



Figure 10.25: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

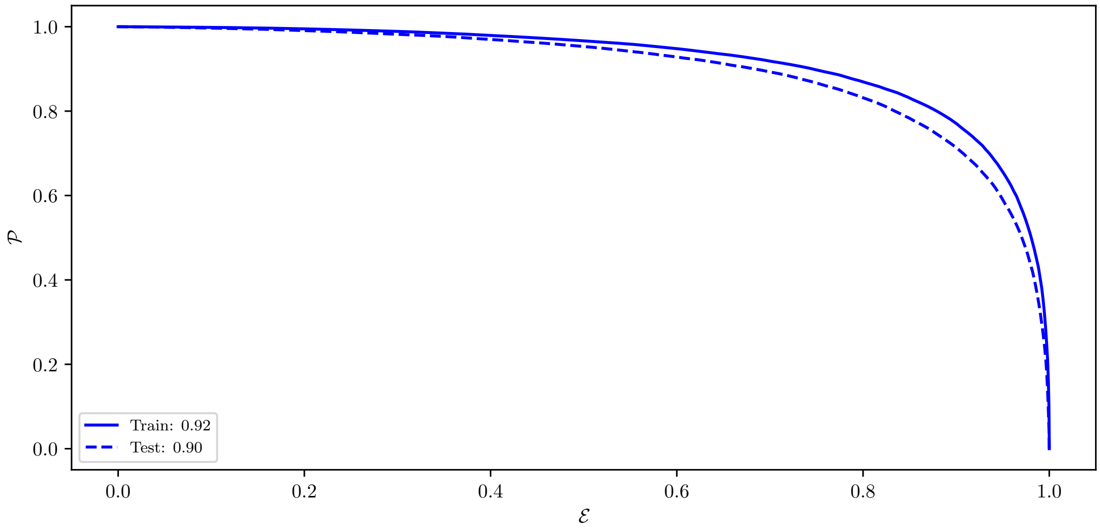



Figure 10.26: ROC curves of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

# Uniformity Boosted $B\bar{B}$ Suppression Training

## Hyper-parameter Optimization

Hyper-parameters were not optimized due to the large CPU time consumption of the algorithm. The following set up of the hyper-parameters was chosen

- `nTrees`: 300

- `nLevels`: 4

## Results

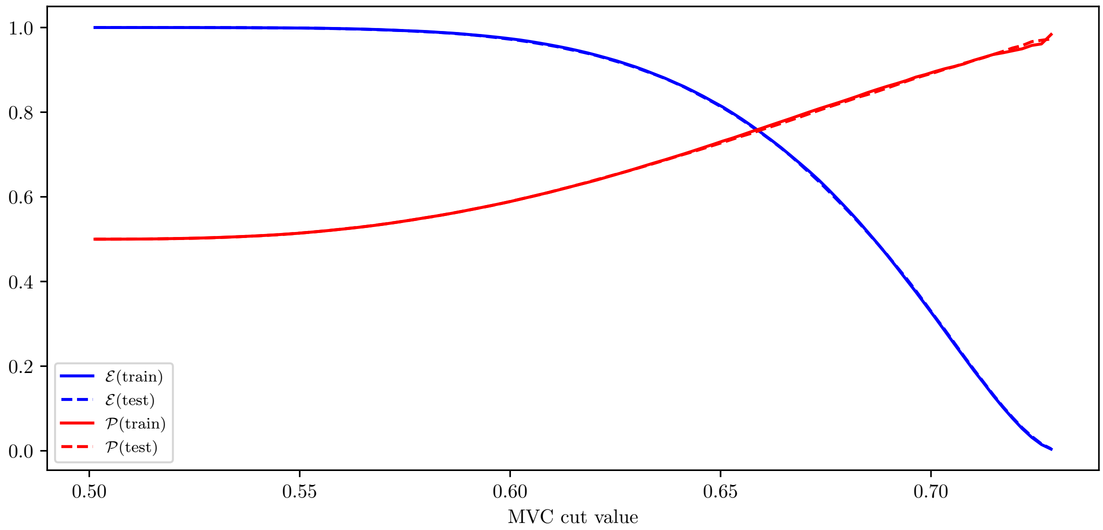



Figure 10.27: Efficiency ($\mathcal{E}$) and purity ($\mathcal{P}$) of the uniformity boosted MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.

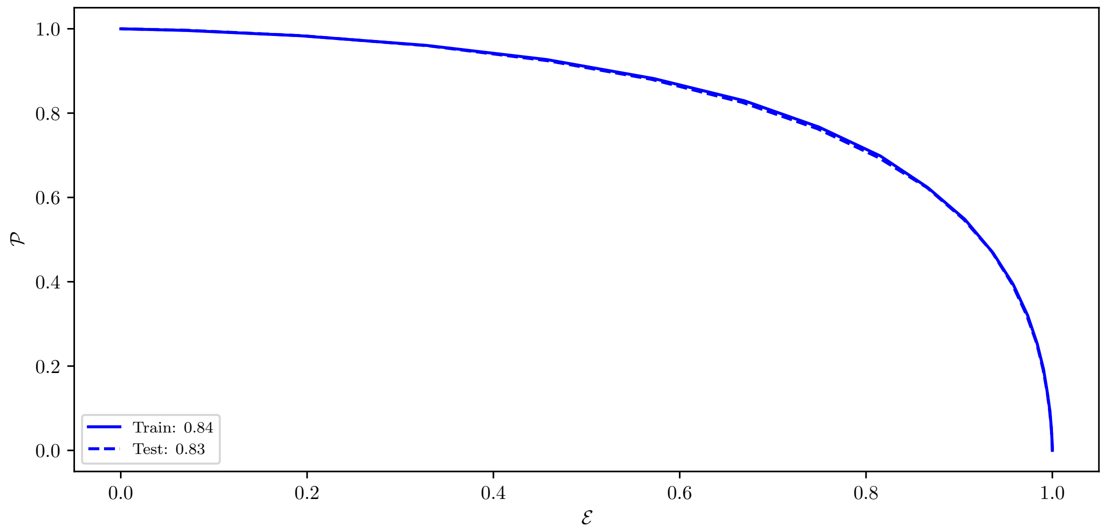



Figure 10.28: ROC curves of the uniformity boosted MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples.