

UNIVERZA V LJUBLJANI  
FAKULTETA ZA MATEMATIKO IN FIZIKO

DOKTORSKA DISERTACIJA

Matic Lubej

2018





UNIVERSITY OF LJUBLJANA  
FACULTY OF MATHEMATICS AND PHYSICS  
DEPARTMENT OF PHYSICS

Matic Lubej

**Study of  $B \rightarrow KK\ell\nu_\ell$  Decays at Belle**

Doctoral Thesis

ADVISER: Assist. Prof. Dr. Anže Zupanc

Ljubljana, 2018





UNIVERZA V LJUBLJANI  
FAKULTETA ZA MATEMATIKO IN FIZIKO  
ODDELEK ZA FIZIKO

Matic Lubej

**Študija razpadov  $B \rightarrow KK\ell\nu_\ell$  pri Belle**

Doktorska disertacija

MENTOR: doc. dr. Anže Zupanc

Ljubljana, 2018



# Izjava o avtorstvu

Spodaj podpisani Matic Lubej, avtor doktorske disertacije (v nadaljevanju: pisnega zaključnega dela študija) z naslovom:

Meritev razpada  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  z detektorjem Belle,

## IZJAVLJAM

1. da sem pisno zaključno delo študija izdelal samostojno;
2. da je tiskana oblika pisnega zaključnega dela študija istovetna elektronski obliki pisnega zaključnega dela študija;
3. da sem pridobil vsa potrebna dovoljenja za uporabo podatkov in avtorskih del v pisnem zaključnem delu študija in jih v pisnem zaključnem delu študija jasno označil;
4. da sem pri pripravi pisnega zaključnega dela študija ravnal v skladu z etičnimi načeli in, kjer je to potrebno, za raziskavo pridobil soglasje etične komisije;
5. da soglašam, da se elektronska oblika pisnega zaključnega dela študija uporabi za preverjanje podobnosti vsebine z drugimi deli s programsko opremo za preverjanje podobnosti vsebine, ki je povezana s študijskim informacijskim sistemom fakultete;
6. da na UL neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve avtorskega dela v elektronski obliki, pravico reproduciranja ter pravico dajanja pisnega zaključnega dela študija na voljo javnosti na svetovnem spletu preko Repozitorija UL;
7. da dovoljujem objavo svojih osebnih podatkov, ki so navedeni v pisnem zaključnem delu študija in tej izjavi, skupaj z objavo pisnega zaključnega dela študija.

Ljubljana, 26. november 2018,

Matic Lubej





# Zahvala

Anže – hvala za strokovno in moralno pomoč. Bil si, in ostajaš, več kot le moj mentor. V tem času si postal tudi moj prijatelj in vzor.

Družina in bližnji prijatelji – hvala, da ste konstanta v mojem življenju, na katero sem se vedno lahko zanesel. Srečen sem, da vas imam ob sebi.

Ina – največja zahvala pripada tebi. Tvoji objemi so skrbeli za moje zdravje in razum. Oprosti, da je bilo mojih vedno manj.

Mami – to delo posvečam tebi, ki si vame posadila takšno željo po znanju. Vedno sem se trudil zaradi tebe.

Brez vseh vas mi ne bi uspelo.



# Izvleček

V disertaciji predstavljamo meritev razvejitenega razmerja ne-čarobnega semileptonskega razpada  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$ . Meritev je bila opravljena na vzorcu podatkov, ki ustreza integrirani luminoznosti  $710 \text{ fb}^{-1}$ , zbranim z detektorjem Belle na asimetričnem trkalniku delcev  $e^+e^-$  KEKB v mestu Tsukuba, na Japonskem. V delu predstavimo rezultate, ki so bili pridobljeni s konverzijo B2BII. To je prva meritev tega razpada, kjer izmerimo razvejitveno razmerje  $\mathcal{B}(B^+ \rightarrow K^+ K^- \ell^+ \nu) = (3.04 \pm 0.51 \pm {}^{+0.67}_{-0.66}) \times 10^{-5}$ . S signifikanco  $4.6\sigma$  ta meritev šteje kot prvi dokaz za ta razpad.

## Ključne besede:

- detektor Belle
- mezoni  $B$
- semileptonski razpadi
- razvejitveno razmerje

## PACS:

- 11.30.Er Konjugacija naboja, parnost, obrat časa in ostale diskretne simetrije
- 13.20.-v Leptonski, semileptonski in radiativni razpadi mezonov
- 13.20.He Razpadi mezonov s kvarkom  $b$
- 14.40.Nd Mezoni s kvarkom  $b$  ( $|B| > 0$ )



# Abstract

We present the branching fraction measurement of the charmless semileptonic decay  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$ . The measurement has been performed on a data sample corresponding to  $710 \text{ fb}^{-1}$  of integrated luminosity, collected with the Belle detector at the KEKB asymmetric-energy  $e^+e^-$  collider in Tsukuba, Japan. We present the results obtained with the B2BII data format converter. This is the first measurement of the decay, where we obtain the branching fraction of  $\mathcal{B}(B^+ \rightarrow K^+ K^- \ell^+ \nu) = (3.04 \pm 0.51 \pm_{-0.66}^{+0.67}) \times 10^{-5}$ . With the signal significance of  $4.6\sigma$ , this measurement counts as the first evidence for the decay.

## Keywords:

- Belle detector
- $B$  mesons
- semileptonic decays
- branching fraction
- inclusive tagging
- untagged
- rest of event

## PACS:

- 11.30.Er Charge conjugation, parity, time reversal, and other discrete symmetries
- 13.20.-v Leptonic, semileptonic, and radiative decays of mesons
- 13.20.He Decays of bottom mesons
- 14.40.Nd Bottom mesons ( $|B| > 0$ )



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data and Monte-Carlo Samples</b>	<b>7</b>
2.1	Signal MC Production . . . . .	8
2.2	Control Decay . . . . .	10
<b>3</b>	<b>Experimental Setup</b>	<b>11</b>
3.1	KEKB Accelerator . . . . .	11
3.2	Belle Detector . . . . .	13
3.2.1	Silicon Vertex Detector . . . . .	14
3.2.2	Central Drift Chamber . . . . .	15
3.2.3	Time-of-Flight Counter . . . . .	17
3.2.4	Aerogel Cherenkov Counter . . . . .	18
3.2.5	Electromagnetic Calorimeter . . . . .	20
3.2.6	$K_L^0/\mu$ Detector . . . . .	23
3.2.7	Trigger System . . . . .	25
<b>4</b>	<b>Belle to Belle II Format Conversion</b>	<b>27</b>
4.1	Conversion Procedure . . . . .	27
4.2	Validation . . . . .	28
<b>5</b>	<b>Event Reconstruction</b>	<b>33</b>
5.1	Final State Particles Selection . . . . .	33
5.2	Pre-selection of First $B$ Meson Candidates . . . . .	39
5.3	Loose Neutrino Reconstruction . . . . .	42
5.3.1	$q^2$ calculation . . . . .	45
5.4	Final Stage Optimization . . . . .	47
5.5	Charge product categorization . . . . .	49
5.6	Selection Summary . . . . .	51
<b>6</b>	<b>Rest of Event Clean-up</b>	<b>53</b>
6.1	Machine Learning Setup . . . . .	53
6.2	Clusters Clean-up . . . . .	54
6.2.1	$\pi^0$ MVA Training . . . . .	55
6.2.2	$\gamma$ MVA Training . . . . .	56
6.3	Tracks Clean-up . . . . .	58
6.3.1	Tracks from Long-lived Particles . . . . .	59
6.3.2	Duplicate Tracks . . . . .	60
6.4	Belle Clean-up . . . . .	67

6.5	Clean-up Results . . . . .	68
6.6	ROE Clean-up Validation . . . . .	71
<b>7</b>	<b>Background Suppression</b>	<b>75</b>
7.1	Resonant Background . . . . .	75
7.2	Continuum Suppression . . . . .	76
7.2.1	Characteristic Variables . . . . .	77
7.2.2	MVA Training . . . . .	80
7.3	$B\bar{B}$ Suppression . . . . .	82
7.3.1	Boosting to Uniformity . . . . .	83
7.4	Selection Optimization . . . . .	84
7.4.1	$B\bar{B}$ Background Composition and Lepton Veto . . . . .	87
7.5	Data and MC Agreement . . . . .	91
7.5.1	Off-resonance Data . . . . .	91
7.5.2	On-resonance Data . . . . .	93
<b>8</b>	<b>Extraction of Physical Parameters</b>	<b>95</b>
8.1	Fit Setup . . . . .	96
8.1.1	Control Fit . . . . .	96
8.1.2	Signal Fit . . . . .	99
8.2	Adaptive Binning Algorithm . . . . .	102
8.3	Toy MC Experiments . . . . .	103
8.3.1	Pseudo-Experiment: Expected Signal Yield . . . . .	103
8.3.2	Pseudo-Experiment: Linearity Test . . . . .	105
<b>9</b>	<b>Fit Results</b>	<b>107</b>
9.1	Signal MC Fit Results . . . . .	107
9.2	Control Fit Result . . . . .	110
9.2.1	Branching Ratio Measurement for Control Decay . . . . .	113
9.3	Results of Signal Fit in Data . . . . .	115
9.3.1	Signal Yield . . . . .	115
9.3.2	Branching Ratio . . . . .	119
9.3.3	Signal Distribution in bins of $m_{KK}$ . . . . .	119
9.3.4	Signal Distribution in bins of $q^2$ . . . . .	120
<b>10</b>	<b>Systematic Uncertainty</b>	<b>123</b>
10.1	Contributions to the Systematic Uncertainty . . . . .	123
10.1.1	PID Efficiency Correction . . . . .	123
10.1.2	Fit Bias and Binning Effects . . . . .	124
10.1.3	Gaussian Constraints . . . . .	124
10.1.4	Fit Template Smearing and Offset . . . . .	125
10.1.5	Effects of a Finite MC sample . . . . .	125
10.1.6	MVA Selection Efficiencies . . . . .	126
10.1.7	Model Uncertainty Effects . . . . .	127
10.2	Summary of Systematic Uncertainties . . . . .	131
<b>11</b>	<b>Final Results and Conclusions</b>	<b>133</b>
11.1	Signal Significance . . . . .	133
11.2	Branching Fraction . . . . .	134



<b>12 Povzetek doktorskega dela</b>	<b>135</b>
12.1 Uvod . . . . .	135
12.2 Experimentalna postavitve . . . . .	137
12.2.1 Trkalnik KEKB . . . . .	138
12.2.2 Detektor Belle . . . . .	138
12.3 Analizni postopek . . . . .	140
12.3.1 Rekonstrukcija razpada . . . . .	140
12.3.2 Odstranjevanje ozadja . . . . .	142
12.3.3 Luščenje fizikalnih parametrov . . . . .	144
12.4 Sistematske negotovosti . . . . .	151
12.4.1 Posamezni prispevki . . . . .	151
12.4.2 Povzetek sistematskih negotovosti . . . . .	153
12.5 Končni rezultat in zaključek . . . . .	154
<b>A ROE MVA Control Plots</b>	<b>159</b>
A.1 ROE Clean-up $\pi^0$ Training . . . . .	159
A.2 ROE Clean-up $\gamma$ Training . . . . .	163
A.3 ROE Clean-up Duplicate Pair Training . . . . .	166
A.4 ROE Clean-up Duplicate Track Training . . . . .	171
<b>B MVA Control Plots</b>	<b>177</b>
B.1 $q\bar{q}$ Suppression Training . . . . .	177
B.2 Standard $B\bar{B}$ Suppression Training . . . . .	182
B.3 Uniformity Boosted $B\bar{B}$ Suppression Training . . . . .	186
<b>C Other Plots</b>	<b>189</b>
C.1 Signal Fits in $m_{KK}$ . . . . .	189
C.2 Signal Fits in $q^2$ . . . . .	193



# 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 20<sup>th</sup> century. So far the experimental and theoretical research have shown us hand in hand that the universe consists of particles and interaction carriers. Particles of matter, or elementary particles, are divided in 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 in charged leptons;  $e$  (electron),  $\mu$  (muon),  $\tau$  (tau lepton), and their corresponding neutrinos;  $\nu_e$  (electron neutrino),  $\nu_\mu$  (muon neutrino) and  $\nu_\tau$  (tau neutrino). Interaction carriers are known as gauge bosons and they are  $\gamma$  (photon),  $g$  (gluon),  $W^\pm$  (charged weak bosons) and  $Z^0$  (neutral weak boson). Theoretical calculations also predicted the recently discovered Higgs boson ( $H$ ), which is responsible for the mass of all particles. Some of the particles above also have mirrored versions of themselves, called antiparticles, which exhibit somewhat different properties compared to their un-mirrored versions.

Combinations of quarks such as  $q_1q_2q_3$  (hadrons) or  $q_1\bar{q}_2$  (mesons) make up heavier particles. Examples of such particles are not only protons and neutrons, but also heavier particles which can be produced in processes involving high enough energies. Such heavy particles are unstable and decay into lighter ones. Together with the elementary particles and interaction carriers, three (out of four) of these interactions are joined in a theoretical model called the Standard Model (SM) [1, 2, 3, 4] (see Figure 1.1). Standard Model describes the electromagnetic, weak nuclear and strong nuclear interaction. General relativity – the theory of gravity – is not included in the Standard Model, since the two are incompatible on a mathematical level. However, due to its low coupling constant, gravity does not play a significant role in the world of subatomic particles. Experimental studies of particle processes give an insight into the mechanisms of basic interactions between them. By doing so, we are able to learn the secrets of the universe and how it all began.

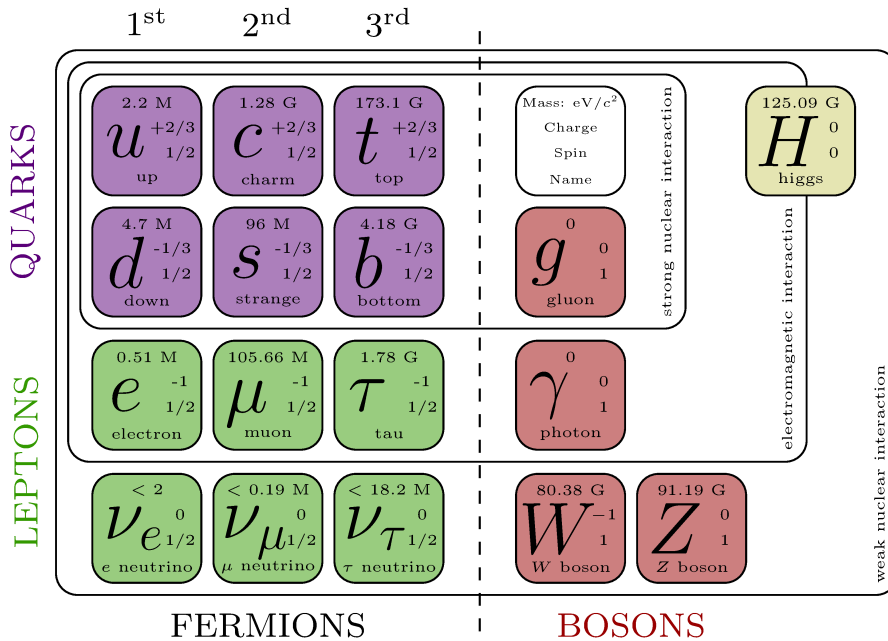


Figure 1.1: A schematic representation of particles in the Standard Model.

This analysis revolves around decays of the so-called  $B$  mesons, which are particles consisting of a  $b$  quark and a light  $u$  or a  $d$  quark. The charged  $B^+$  and the neutral  $B^0$  meson have a structure of  $(\bar{b}, u)$  and  $(\bar{b}, d)$ , respectively, while the anti-particle  $B$  mesons are  $B^-(b, \bar{u})$  and  $\bar{B}^0(b, \bar{d})$ . Perhaps one of the most surprising features of nature that can be studied with decays of  $B$  mesons is the  $CP$  symmetry violation ( $\mathcal{CP}$ ).  $CP$  symmetry is a combination of the  $C$  symmetry (charge conjugation) and the  $P$  symmetry (spatial inversion). A conservation of the  $CP$  symmetry would mean that the mirrored processes, in which all particles are exchanged by the corresponding anti-particles, proceed in exactly the same manner as the original processes. Today we know that this does not hold true for all cases and we, in fact, find processes which violate the  $CP$  symmetry. We also know that  $\mathcal{CP}$  is related to the weak nuclear interaction. Here lies our motivation for studying decays of  $B$  mesons, since they exhibit a rich spectrum of decays which proceed via the weak nuclear interaction.

One of the most important properties of the weak interaction is the fact that it can change the flavor of particles. Such processes are forbidden for the electromagnetic and the strong nuclear interaction, but not for the weak one. Information about the quark transition probabilities is merged into a form of a complex matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [5, 6]

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

The CKM matrix is a unitary matrix and has only four parameters, which are free parameters of the theory and hence must be experimentally determined. The unitarity of the CKM matrix provides us with several mathematical identities, out

of which the most relevant one for  $B$  meson physics is

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

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 related to the free parameters of the CKM matrix. All measurements of weak interaction processes involving  $B$  mesons depend on the four free parameters of the CKM matrix. Results of such measurements hence determine the sides and angles of the unitarity triangle. The goal is to then combine all such measurements and overconstrain the sides and angles of 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 using different decays, with the most important input from  $B$  meson decays. This fact represents another motivation to study the  $B$  meson decays.

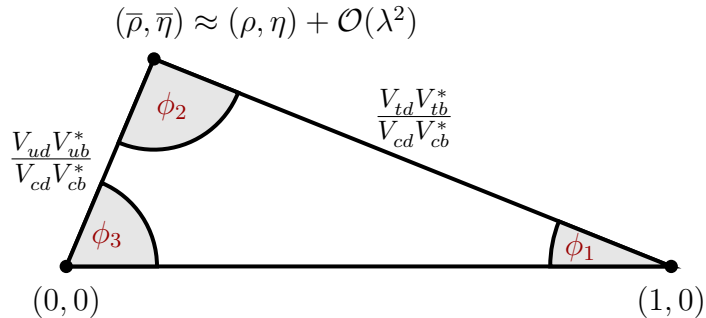


Figure 1.2: The unitarity triangle in the Wolfenstein parametrization [7].

In this analysis, we focus on the  $V_{ub}$  CKM matrix element, which corresponds to  $b \rightarrow u$  quark transitions. It has the smallest absolute value of all the CKM matrix elements and is currently determined with the largest uncertainty. Such quark transitions are present in charmless semileptonic  $B$  meson decays of the form

$$B^+ \rightarrow X_u^0 \ell^+ \nu_\ell, \quad (1.3)$$

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

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

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  $|V_{ub}|^2$  is the CKM element describing the  $b \rightarrow u$  quark transition. Measurement of the  $V_{ub}$  CKM matrix element can be performed using two general approaches, with the exclusive or inclusive method, which are described below. Both methods

require an application of 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  $B \rightarrow \pi \ell \nu$ . 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  $B \rightarrow X_c \ell \nu$  processes since the  $b \rightarrow c$  quark transition is much more probable than the  $b \rightarrow u$  transition. The second experimental challenge is to separate  $B$  meson decays into a specific charmless hadronic final state from other  $B \rightarrow X_u \ell \nu$  decays, since they populate roughly 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  $B \rightarrow X_u \ell \nu$ . In this case, the total decay rate for  $b \rightarrow u \ell \nu$  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  $B \rightarrow X_c \ell \nu$  background. Experimental sensitivity to  $V_{ub}$  is highest where  $B \rightarrow X_c \ell \nu$  decays are less dominant. Theory and experiment have to compromise and limit the  $|V_{ub}|$  determination to a region of phase-space where the signal-to-background ratio is good. Theoretical calculations take this into account by calculating the partial decay rate  $\Delta\Gamma$ , which is more challenging to determine than the total decay rate. One possible and often used approach to reduce  $b \rightarrow c$  background is to reject all events with 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 the dominant  $b \rightarrow c \rightarrow s$  transition chain. This means that if a kaon is found in an event, it is very likely that it originates from a particle with a  $c$  quark, indicating the  $b \rightarrow c$  process.

If  $|V_{ub}|$  is determined with both of these methods, the values can be compared and potentially combined. It turns out that the consistency between the two results is only marginal, the difference is at a level of  $3\sigma$ . The current world averages [8] of the exclusive (from  $B^0 \rightarrow \pi^- \ell^+ \nu$ ) and inclusive (GGOU collab. [9]) methods are

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

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

where the first and the second uncertainties are the experimental and the theoretical, 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  $b \rightarrow c$  contributions. We focus on the charged  $B \rightarrow KK \ell \nu$  decay, which is very similar to the  $B \rightarrow \pi \ell \nu$ , 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 1.3. In this case, we have kaons in the final state where the  $B$  meson decayed via a  $b \rightarrow u$  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 take a step closer towards solving the  $V_{ub}$  puzzle.

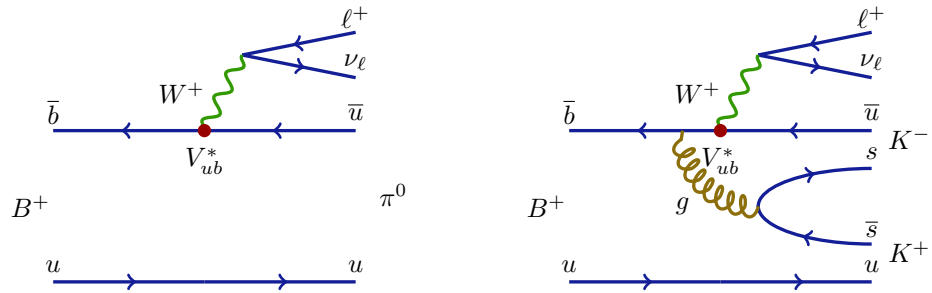


Figure 1.3: Feynman diagrams for the  $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$  decay (left) and the  $B^+ \rightarrow K^- K^+ \ell^+ \nu_\ell$  decay (right).

Specifically, we will be focusing on the decays of the charged  $B$  mesons of the form  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$ , 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  $B \rightarrow K K \ell \nu$  imply decays of the form  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  and its charge conjugated counterpart.





# Chapter 2

## Data and Monte-Carlo Samples

The Belle detector acquired a dataset of about  $L_0 \approx 710 \text{ fb}^{-1}$  of integrated luminosity in its lifetime at the  $\Upsilon(4S)$  energy of 10.58 GeV, which corresponds to about  $771 \times 10^6 B\bar{B}$  meson pairs. Additionally, several streams of Monte-Carlo (MC) simulated samples were produced, where each stream of MC corresponds to the same amount of data that was recorded the detector. The main focus of this work is to study a rare signal decay that is not necessarily produced abundantly or at all in the existing MC samples. In such cases, it is a common practice to produce specific samples of signal MC, where the abundance of the signal decay is much larger, enabling us to study its properties in greater detail.

The following samples were used in 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  $B^+B^-$  and  $B^0\bar{B}^0$  (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  $1/10 \times L_0$ ,
  - $B \rightarrow X_u \ell \nu$  (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  $B^+ \rightarrow K^+K^-\ell\nu_\ell$ , including the charge conjugated  $B^-$  mesons, was produced using the `mcproduzh` [10, 11] 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 MC 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  $B$  meson. This decreases the CPU consumption during the detector simulation and reconstruction.

The relevant processes which contribute to our signal decay are

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

where  $a_{00}$ ,  $a_{20}$ ,  $f_2$ , and  $f_0$  are light, unflavored states which include further decays into a  $K^+K^-$  pair. The  $X_u^0$  state represents a generic  $u\bar{u}$  quark pair, which further hadronizes based on the PYTHIA quark hadronization model [12]. 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 abundant and follow a wider and smoother distribution.

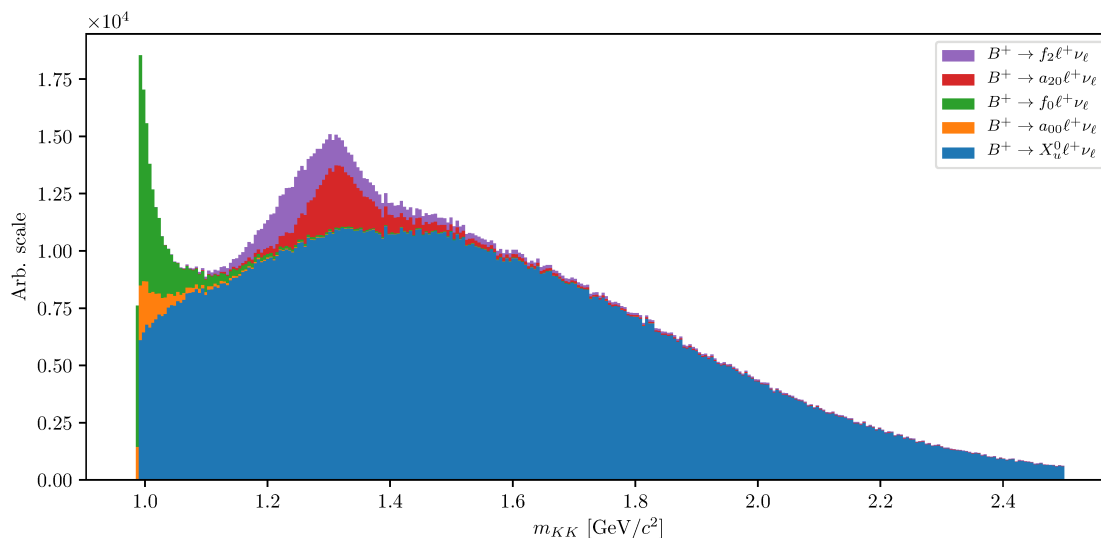


Figure 2.1: Invariant mass of the  $KK$  pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structure, while  $KK$  pairs from the  $X_u^0$  state are more abundant and follow a wider and smoother distribution.

The produced signal MC sample contains decays of the form  $B \rightarrow KK\ell\nu$ , as well as  $B \rightarrow KKX\ell\nu$ , 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 fractions for each channel by making combinations of the particles directly from the generator output. Table 2.1 shows some of the most prominent  $B \rightarrow KKX\ell\nu$  channels and their relative fractions. It is clear that our signal decay is the most abundant one, with a relative contribution of about 28 %, 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 a decrease in efficiency due to reconstruction effects.

Channel	Ratio [%]	Channel	Ratio [%]
$K^+K^-$	28.14	$K^+K^-\rho^0$	1.93
$K^+K^-\pi^0$	8.94	$K^+\bar{K}^0\rho^-$	1.84
$K^+\bar{K}^0\pi^-$	8.71	$K^0K^-\rho^+$	1.83
$K^0K^-\pi^+$	8.70	$K^0\bar{K}^0\rho^0$	0.00
$K^+K^-\pi^+\pi^-$	4.15	$K^+K^-\pi^0\pi^0$	0.86
$K^0\bar{K}^0$	3.32	$K^+K^-\pi^+\rho^-$	0.69
$K^0\bar{K}^0\pi^0$	3.26	$K^+K^-\rho^+\pi^-$	0.68
$K\bar{K}$ pair with $\eta$	7.08		
$K\bar{K}$ pair with $\omega$	5.33		
Other	14.53		

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

We generate about  $1.3 \times 10^9$  events of the form  $B \rightarrow X_u\ell\nu$ , 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 candidates in the `ulnu` MC sample. This amounts to a total of about  $9.37 \times 10^6$  generated signal events, and to a branching fraction

$$\mathcal{B}(B^+ \rightarrow K^+K^-\ell^+\nu_\ell)_{GEN} = 1.53 \times 10^{-5}, \quad (2.1)$$

where  $\ell$  is  $e$  or  $\mu$ . During the 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 also 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 the control decay of the form

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

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 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 [13] and [14], 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. Some of the electrons collide with a tungsten target to produce positrons, which are accelerated in the LINAC to an energy of around 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 a crossing angle of about 22 mrad. The center-of-mass (CMS) energy of the collision corresponds to the mass of the  $\Upsilon(4S)$  resonance as

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

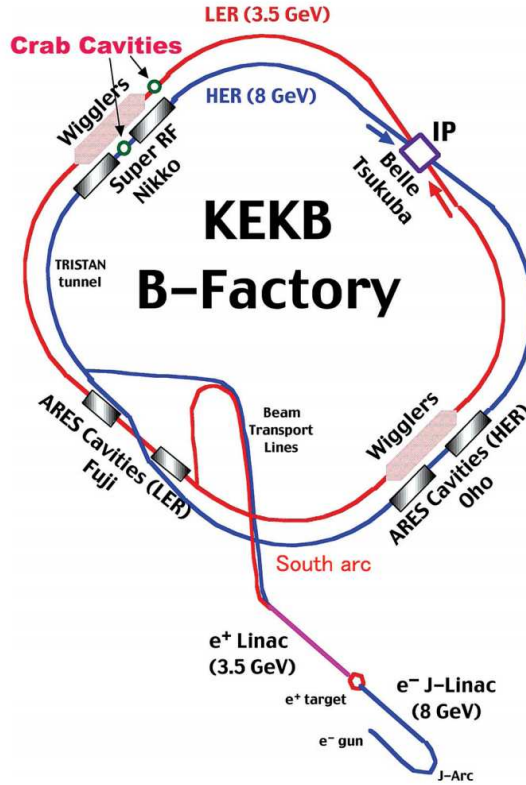


Figure 3.1: Schematic layout of the KEKB accelerator. The HER and the LER are the  $e^-$  and the  $e^+$  beams, respectively. Four experimental halls (FUJI, NIKKO, OHO and TSUKUBA) are shown.

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 CMS energy), the experiment collected data also at energies corresponding to other  $\Upsilon(nS)$  resonances, where  $n = 1, 2, 3, 5$ , and at energies below the resonances.

Interaction	Cross-section [nb]
$\Upsilon(4S) \rightarrow B\bar{B}$	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 for various physics processes at  $\Upsilon(4S)$  resonance energy [14].

KEKB achieved the world-record in the instantaneous luminosity of  $2.11 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , twice as large as the design value. The total dataset, collected with the Belle detector, amounts to the total integrated luminosity of  $1041 \text{ fb}^{-1}$ . Out of the full Belle dataset about  $711 \text{ fb}^{-1}$  of data were taken at the  $\Upsilon(4S)$  energy of  $10.58 \text{ GeV}$ , 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 products of  $e^+e^-$  collisions. The detector is configured around a  $1.5 \text{ T}$  superconducting solenoid and iron structure surrounding the interaction point (IP). The four-momentum of the decaying  $B$  mesons and its daughter particles are determined via a series of sub-detector systems, which are installed in an onion-like shape, as shown in Figure 3.2. 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.

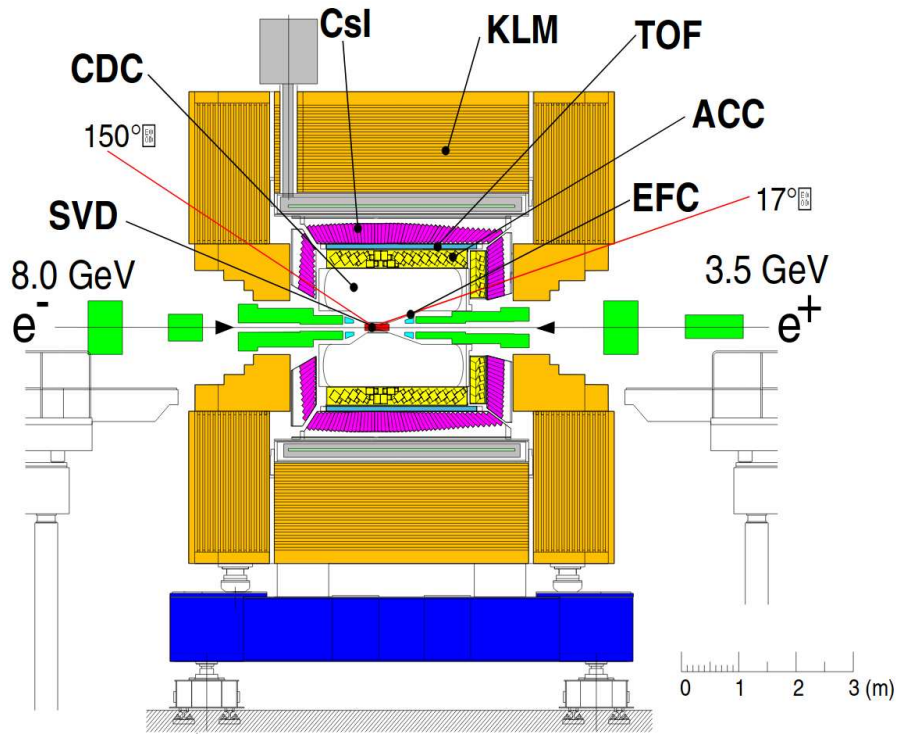


Figure 3.2: Schematic layout of the Belle detector with the corresponding sub-detectors [14].

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. The electron beam crosses the positron beam at an angle of about  $22^\circ$ . The polar angle  $\theta$  covers the region between  $17^\circ \leq \theta \leq 150^\circ$ , while the cylindrical angle  $\varphi$  covers the full  $360^\circ$  range, amounting to a solid angle coverage of about 92%.

### 3.2.1 Silicon Vertex Detector

SVD is the inner-most part of the Belle detector and its purpose is the determination of the decay vertices of decaying particles. The precision of the subsystem is about  $100 \mu\text{m}$ , which is important for measuring the difference in  $z$ -vertex positions of the  $B$  mesons in time-dependent  $CP$  violation studies. The main part of the SVD are the double-sided silicon detectors (DSSD). With their thin profile and parallel silicon strips on both sides, they provide two-dimensional hit information of the 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  $23^\circ < \theta < 140^\circ$ . 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 was 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 covers the polar angle of  $17^\circ < \theta < 150^\circ$ . The first layer was moved closer to the IP, which greatly improved the sub-system precision, since the multiple-Coulomb scattering affects the resolution more as the distance from the IP increases. The front and the side views of the SVD2 are shown in Figure 3.3.

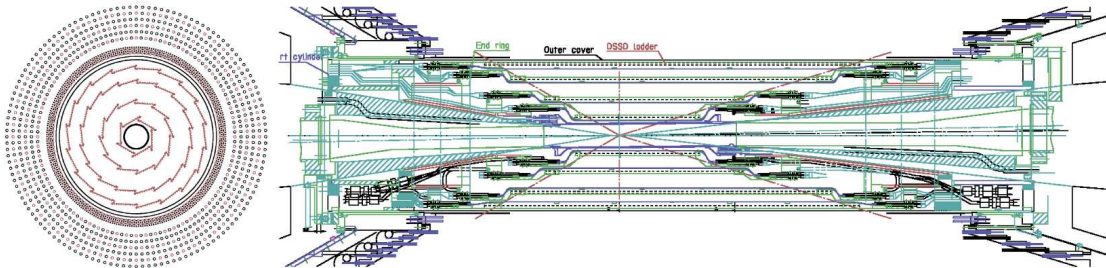


Figure 3.3: Front (left) and side (right) view of the SVD detector with the SVD2 configuration. The front view also shows the inner wires of the Central Drift Chamber [15].

The efficiency of the SVD was determined as a fraction of CDC tracks within the SVD acceptance with 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 based on the resolutions of the impact parameter  $z$  and  $r\phi$ , obtained from cosmic ray data. The momentum and angular dependence of the impact parameters resolution is well represented as a function of



the pseudo-momentum for the SVD2 by the following parametrization

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

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

where pseudo-momentum is defined as  $p\beta \sin^{5/2} \theta$  for the  $z$  direction and as  $p\beta \sin^{3/2} \theta$  for the  $r\phi$  plane,  $p$  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 in Figure 3.4.

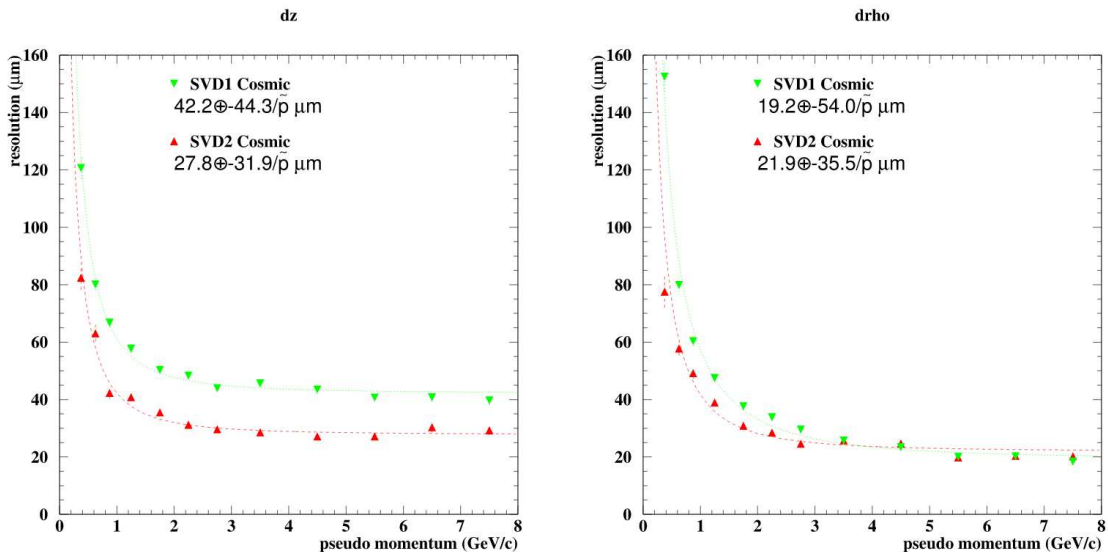


Figure 3.4: Impact parameter resolutions of  $z$  (left) and  $r\phi$  (right) coordinates for the SVD1 and SVD2 configuration of the vertex detector [15].

### 3.2.2 Central Drift Chamber

CDC is a large-volume tracking device located in the central part of the Belle detector. It 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 the measuring of the polar angle of the track. The wire configuration is shown in Figure 3.5. 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 the 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 for the photoelectric effect, which is important for reducing background electrons induced by the synchrotron radiation from the beam.

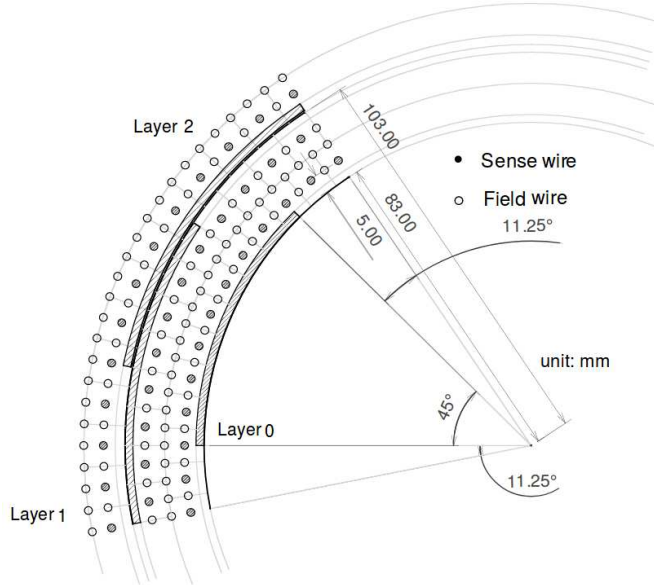


Figure 3.5: Cell structure of CDC [14].

Charged particles, which pass the CDC wire frame, cause gas ionization. The produced electrons drift toward the sense wires with high acceleration, due to the strong electric field close to the wire. The accelerated electrons collide with the gas molecules and produce secondary, tertiary, and so on, ionizations, which results 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 enables 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  $dE/dx$ .  $dE/dx$ , as a function of momentum, differs for different particles, as shown in Figure 3.6. This allows for identification purposes, specifically for kaons and pions. In the momentum region of less than  $0.8 \text{ GeV}/c$ ,  $dE/dx$  enables a separation between kaons and pions with at least  $3\sigma$ . 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

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

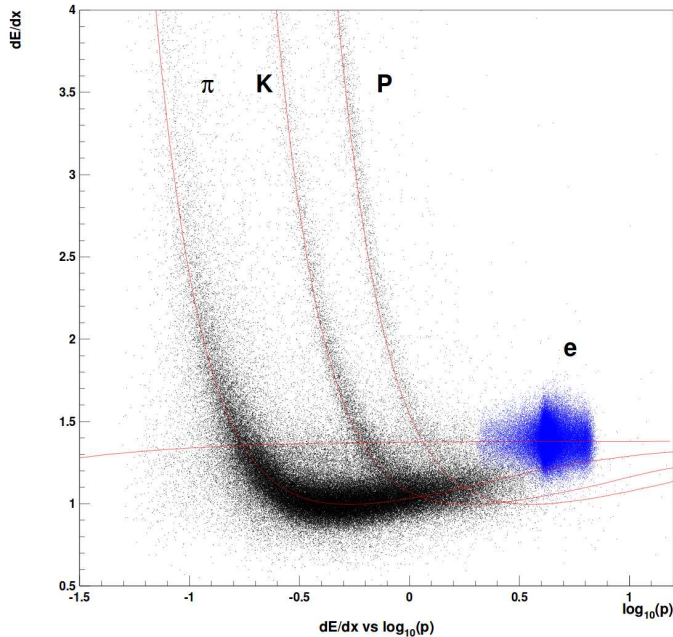


Figure 3.6: Measured  $dE/dx$  as a function of particle momentum. The red lines show the expected distribution for different types of particles [14].

### 3.2.3 Time-of-Flight Counter

The purpose of the TOF subdetector is particle identification in the momentum region of  $0.8 \text{ GeV}/c < p < 1.2 \text{ GeV}/c$ , especially for kaons and pions. There are 64 TOF modules in the barrel region, covering the polar angle of  $33^\circ < \theta < 121^\circ$ . 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

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

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 identified by TOF in hadron events is shown in Figure 3.7, 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  $1.2 \text{ GeV}/c$ , 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 the particle momentum, shown in Figure 3.8. A clear separation of about  $2\sigma$  is achieved for particle momenta up to  $1.25 \text{ GeV}/c$ .

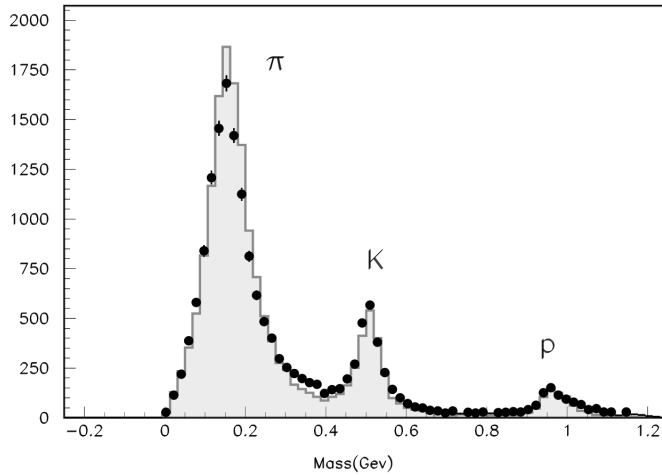


Figure 3.7: Mass distribution from TOF measurements for particle momenta below 1.2 GeV/c [14].

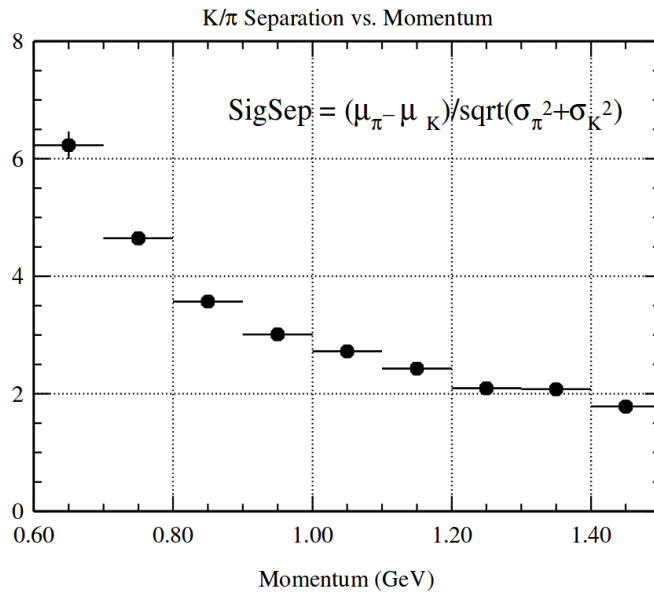


Figure 3.8:  $\pi^\pm/K^\pm$  separation by TOF [14].

### 3.2.4 Aerogel Cherenkov Counter

TOF is not capable of performing efficient PID above 1.2 GeV/c momentum since  $\beta$  is almost equal to 1. For higher momenta in the region of 1.0 GeV/c < 4.0 GeV/c the ACC is introduced. It is a threshold-type Cherenkov counter, which utilizes the fact that a particle emits Cherenkov light if the speed of the particle exceeds the speed of light in the passing medium. ACC is introduced in the barrel region with 960 separate modules, covering a polar angle of  $34^\circ < \theta < 127^\circ$ , and with 228 modules in the forward end-cap region, with the polar angle coverage of  $17^\circ < \theta < 34^\circ$ . 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 particles momenta, 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 3.9.

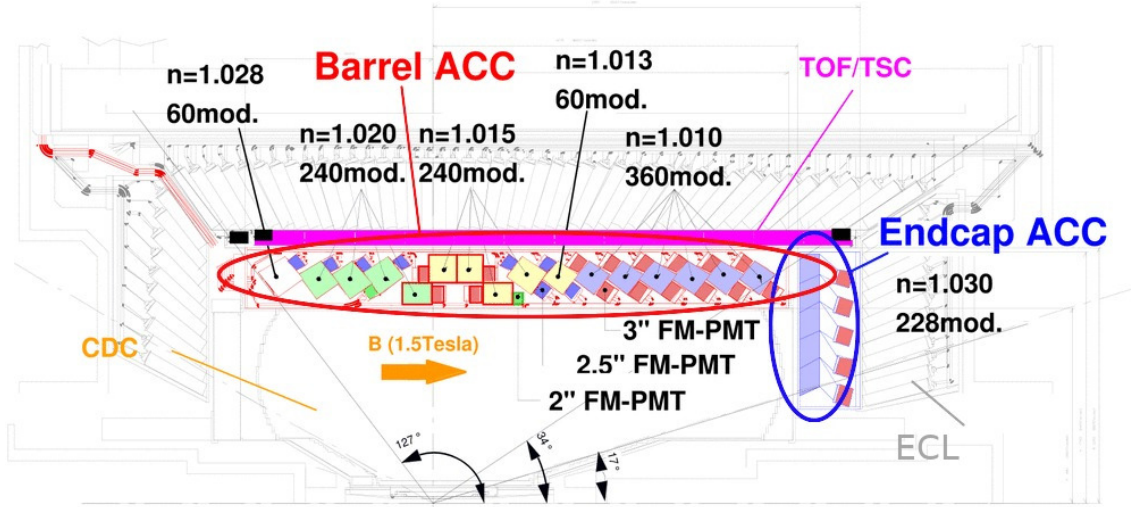


Figure 3.9: Cross-sectional view of the Belle detector with some of the highlighted detector components. The barrel ACC is located between the CDC and TOF, while the endcap ACC is located in the forward endcap, behind the ECL crystals [14].

The threshold velocity  $\beta$  to irradiate Cherenkov light, for a given particle of mass  $m$ , is given by

$$\beta \geq \frac{1}{n}, \quad (3.6)$$

where  $n$  is the refractive index of the medium. The refractive indices in the ACC are chosen such that, due to a difference in masses, for a given interval of momenta, pions emit Cherenkov light, while kaons do not. Using the PID information from ACC, as well as other sub-systems, the electron identification efficiency, in the momentum range above 1 GeV/ $c$ , is larger than 90%, while the pion fake rate – the probability of wrongly identifying pions as electrons – is around 0.2 - 0.3%. Similarly for kaons, kaon PID efficiency is equal to 80% for most of the momentum region up to 4 GeV/ $c$ , while pion fake rate remains below 10%. Figure 3.10 shows the electron and kaon efficiencies, and the corresponding pion fake rates as a function of the particle's momentum.

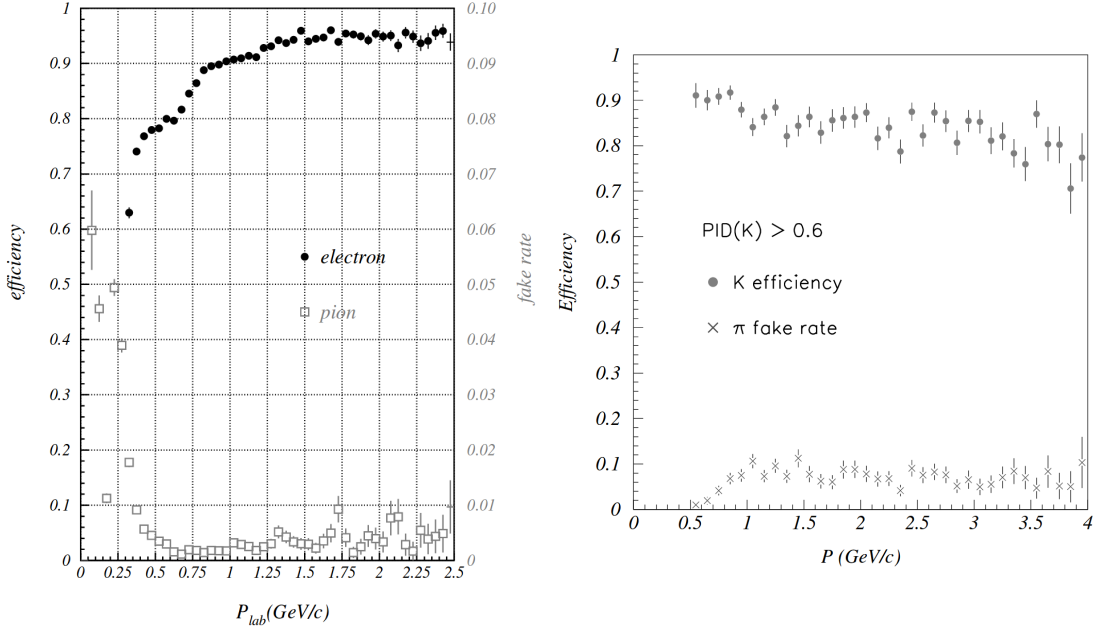


Figure 3.10: Electron identification efficiency and fake rate for charged pions (left) and similarly for kaons (right). Note the different scales for the electron efficiency and fake rate in the left plot [14].

### 3.2.5 Electromagnetic Calorimeter

Measurement of the position and energy deposit of particles is performed in the ECL, especially for electrons and photons. The latter are electrically neutral and, therefore, not detected in any of the previously mentioned subdetectors. The ECL also provides complimentary particle identifications for the separation of electrons and 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  $32.2^\circ < \theta < 128.7^\circ$ , and for the forward and backward end-caps  $12.4^\circ < \theta < 31.4^\circ$  and  $130.7^\circ < \theta < 155.1^\circ$ , respectively. Figure 3.11 shows the layout of the barrel and end-cap ECL.

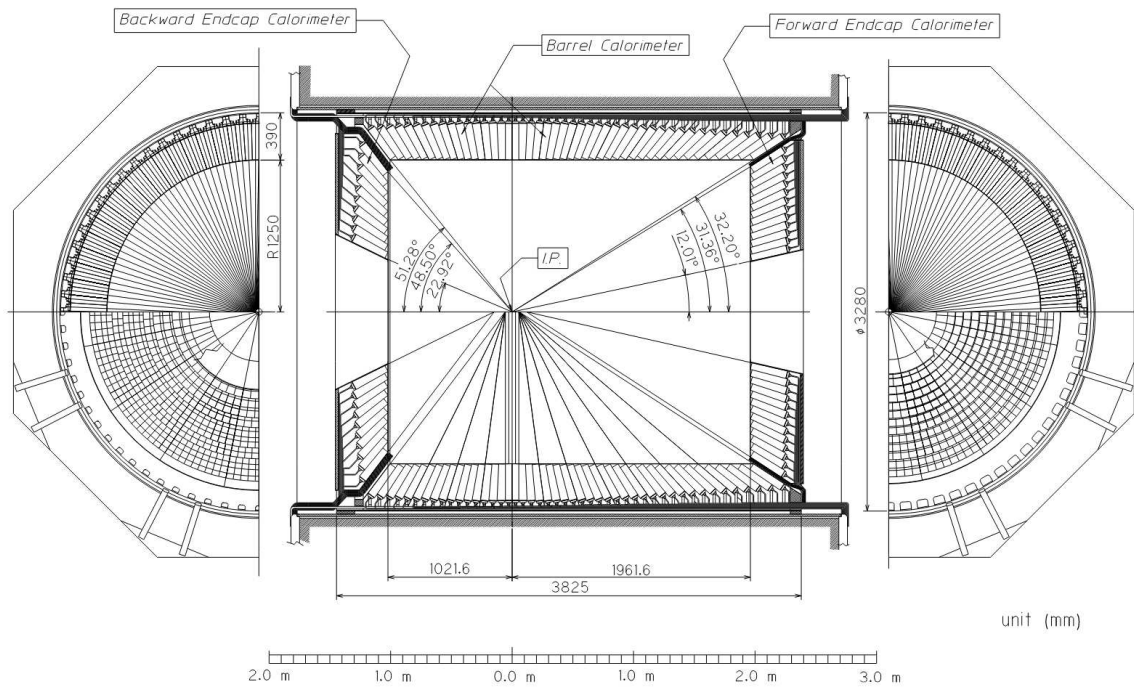


Figure 3.11: Overall configuration of the ECL [14].

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 3.12. 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/ $c$ .

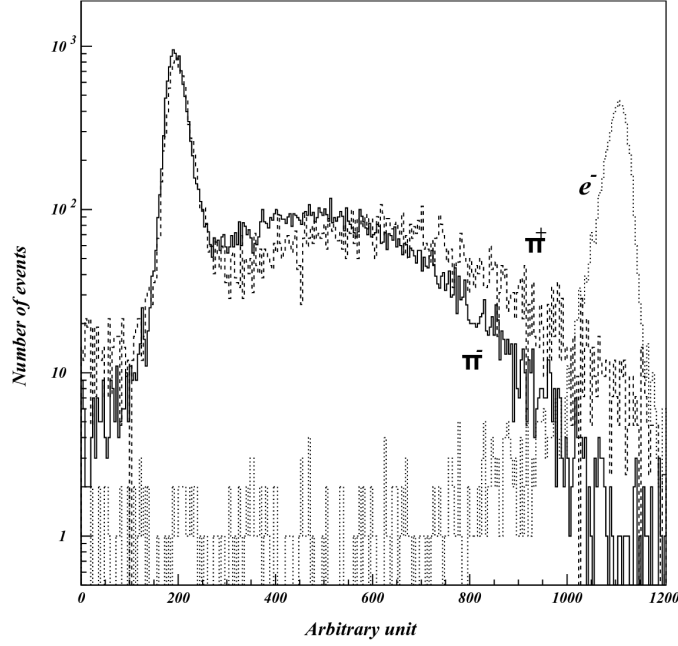


Figure 3.12: Distribution of the energy deposit by electrons and charged pions at 1 GeV/ $c$  momentum [14].

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 3.13. These values are in good agreement with Monte Carlo predictions. Worse energy resolution in the backward end-cap is due to the lower photon energy, which results in larger scattering in the material in front of the calorimeter [15]. The energy resolution as a function of the energy can be described via the following relation

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

while the resolution of the position measurement is

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



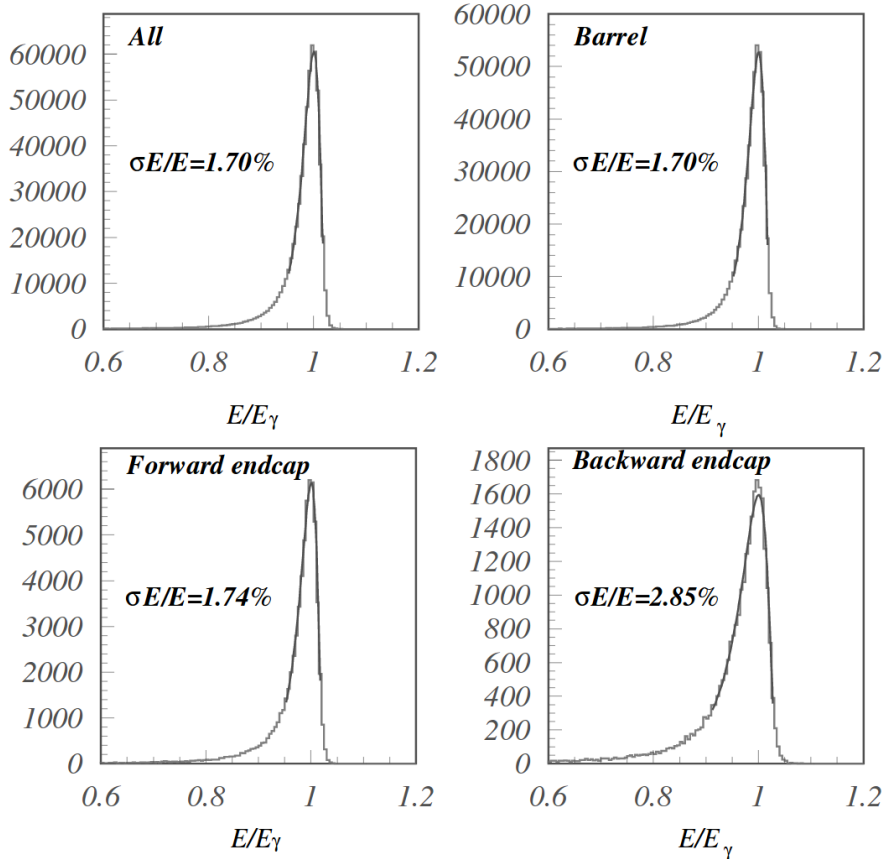


Figure 3.13: Reconstructed energy distribution for  $e^+e^- \rightarrow \gamma\gamma$  events for overall, barrel, forward, and backward end-cap ECL [14].

### 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  $\mu$  with momenta larger than 0.6 GeV/ $c$ . The setup covers the polar angle of  $20^\circ < \theta < 155^\circ$ . 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. The iron plates of the KLM serve a dual role; as the necessary material for hadronic interactions, as well as the flux return for the magnetic field. Figure 3.14 shows a cross-section of an RPC superlayer, consisting of an RPC pair.

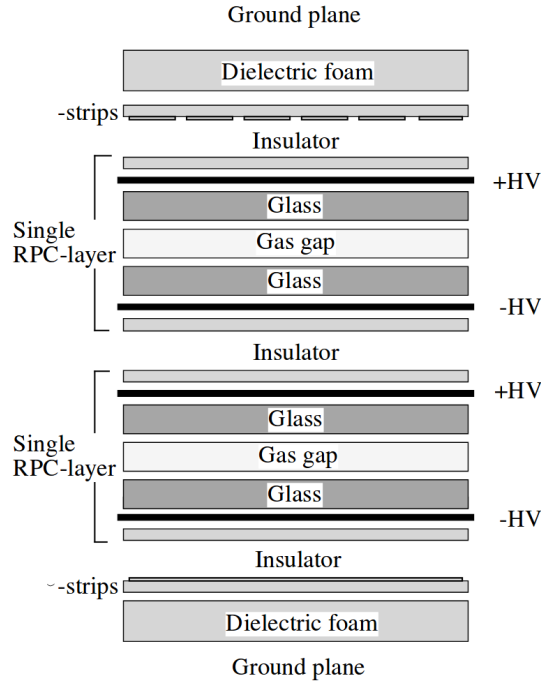


Figure 3.14: Cross-section of an RPC superlayer, consisting of an RPC pair [14].

The  $K_L^0$  particle can be distinguished from other charged hadrons by the fact that they have no matched track in the CDC. The flight direction can be inferred from the hit locations in the consecutive RPCs. Tracks of charged particles measured in CDC are extrapolated into KLM, and clusters within  $15^\circ$  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  $0.5 \text{ GeV}/c$ . They do not interact strongly and do not produce hadronic showers in the KLM. Figure 3.15 (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 3.15 (right) shows the muon detection efficiency as a function of the momentum, 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. Using  $K_S \rightarrow \pi^+\pi^-$  events, a muon identification efficiency of better than 90% is determined, with a pion fake rate of less than 5% for particles with momenta larger than  $1.5 \text{ GeV}/c$  and an applied likelihood cut of 0.66.

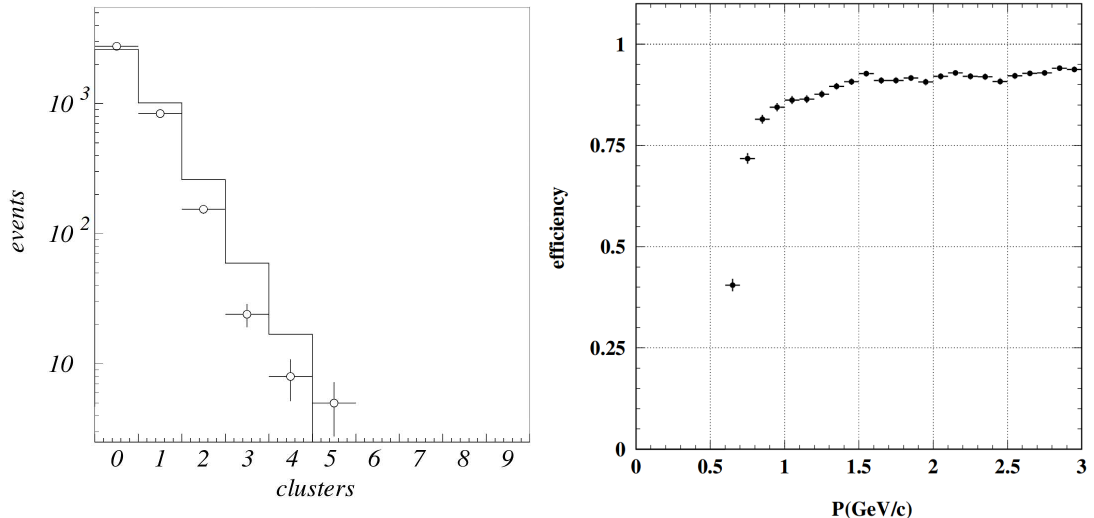


Figure 3.15: Number of neutral clusters per event in KLM (left) and muon detection efficiency as a function of momentum in KLM (right) [14].

Cosmic ray events have been used to determine the efficiency and resolution of the KLM, with an overall efficiency typically above 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 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. The next 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 3.16. 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 GDL within  $1.85 \mu\text{s}$  after the collision, and the final trigger signal is provided within a fixed  $2.2 \mu\text{s}$  latency. The combined efficiency from the L1 trigger is greater than 99.5% for hadronic events.

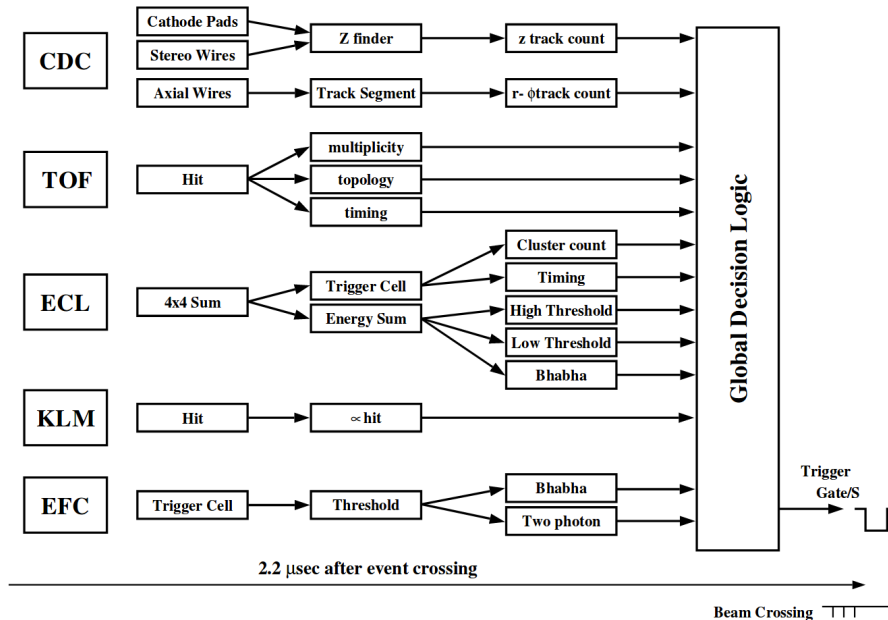


Figure 3.16: The Level-1 trigger system for the Belle detector [14].

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  $|dz| < 5.0 \text{ cm}$  and with a total energy deposit in the ECL more than  $1 \text{ GeV}/c$  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, a high efficiency for signal events is still required. Events must satisfy the following conditions

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

Approximately 27% of triggered events are passed through L4 while keeping an almost 100% efficiency for hadronic events. Events that pass the L4 trigger are fully reconstructed and stored to the file system. Overall, the efficiency of hadronic events after all trigger stages is measured to be more than 99%.

# Chapter 4

## Belle to Belle II Format Conversion

### 4.1 Conversion Procedure

The Belle experiment finished its data-taking run of 10 years at the end of 2010, after collecting a dataset of about  $1 \text{ ab}^{-1}$ . That year the Belle detector was shut down, the Belle II experiment started in its place and the focus moved to the construction of the Belle II detector and the development of the Belle II Analysis Framework (BASF2) [16]. However, Belle analyses are still on-going and Belle data is still being used today. BASF2 software, with its modular structure, has a more intuitive approach to performing analyses, however, since it was rewritten completely from scratch, it was designed for the incoming Belle II data, therefore usage of Belle data is outside of its scope.

In the Belle Collaboration, a task force was created in order to convert Belle data into Belle II format (B2BII) [17]. The B2BII package was developed as a part of BASF2 in order to convert the data and MC of the Belle experiment and make it available within BASF2. In addition to the convenience of the 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 selection criteria 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 [18].

## 4.2 Validation

In order to make sure that the conversion was successful, a thorough validation is 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, so track conversion should perform flawlessly. Additionally, energy measurement is also important in our analysis. In order to successfully determine the missing four-momentum in the event, 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, and their difference, which is (up to numerical precision) equal to 0. The plots indicate that the conversion is successful in all aspects, so we can proceed with the analysis in the framework of BASF2.

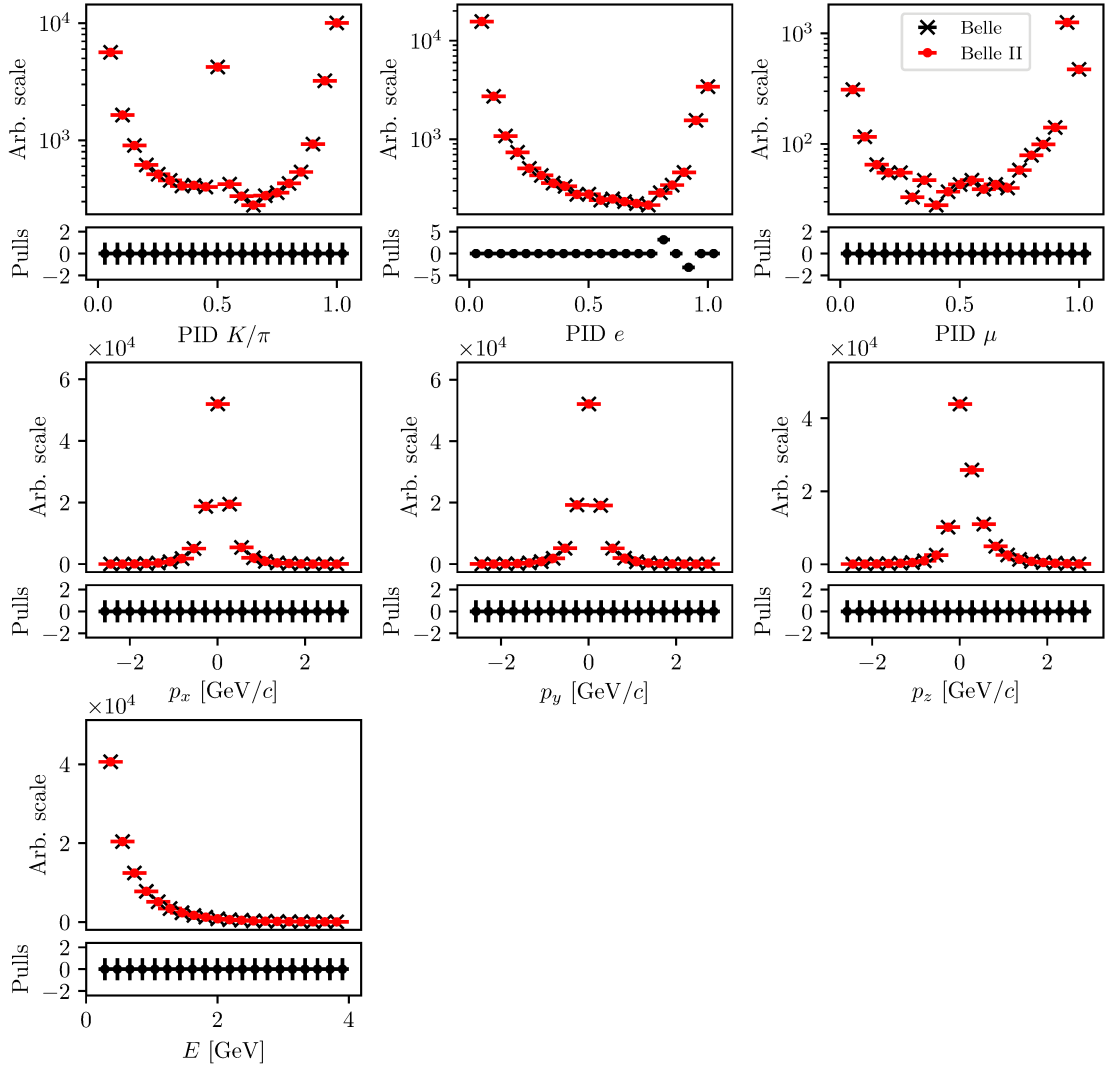


Figure 4.1: Some of the more important physical properties of tracks for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.

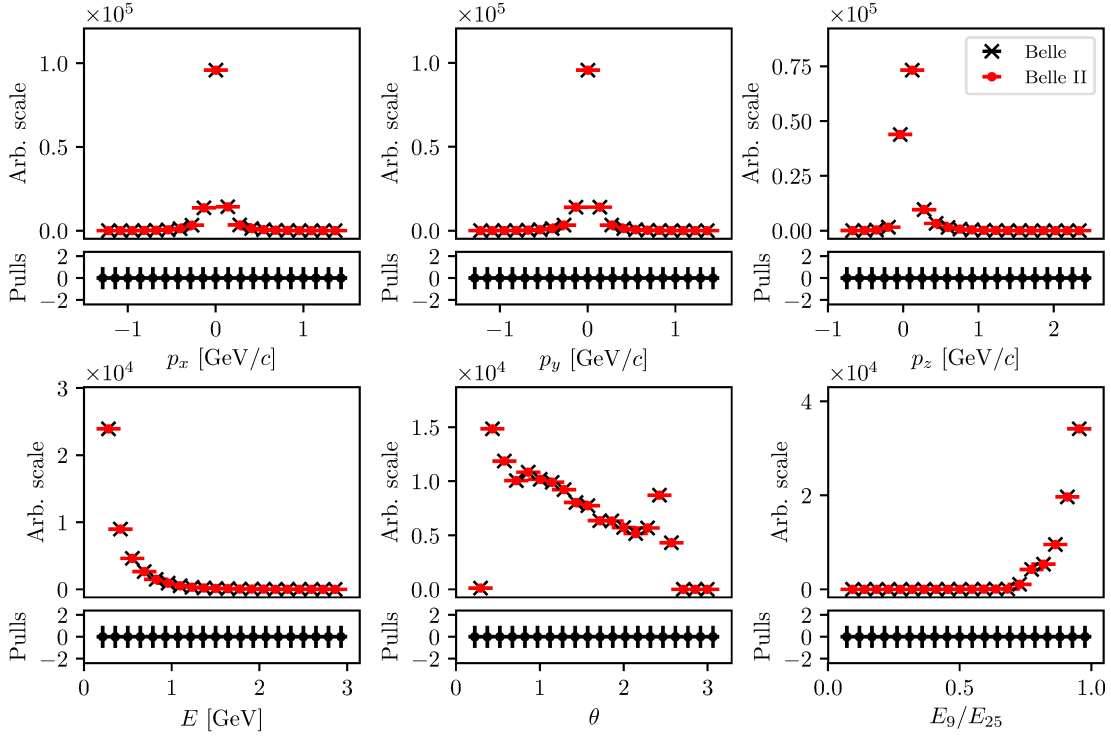


Figure 4.2: Some of the more important physical properties of photons for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.



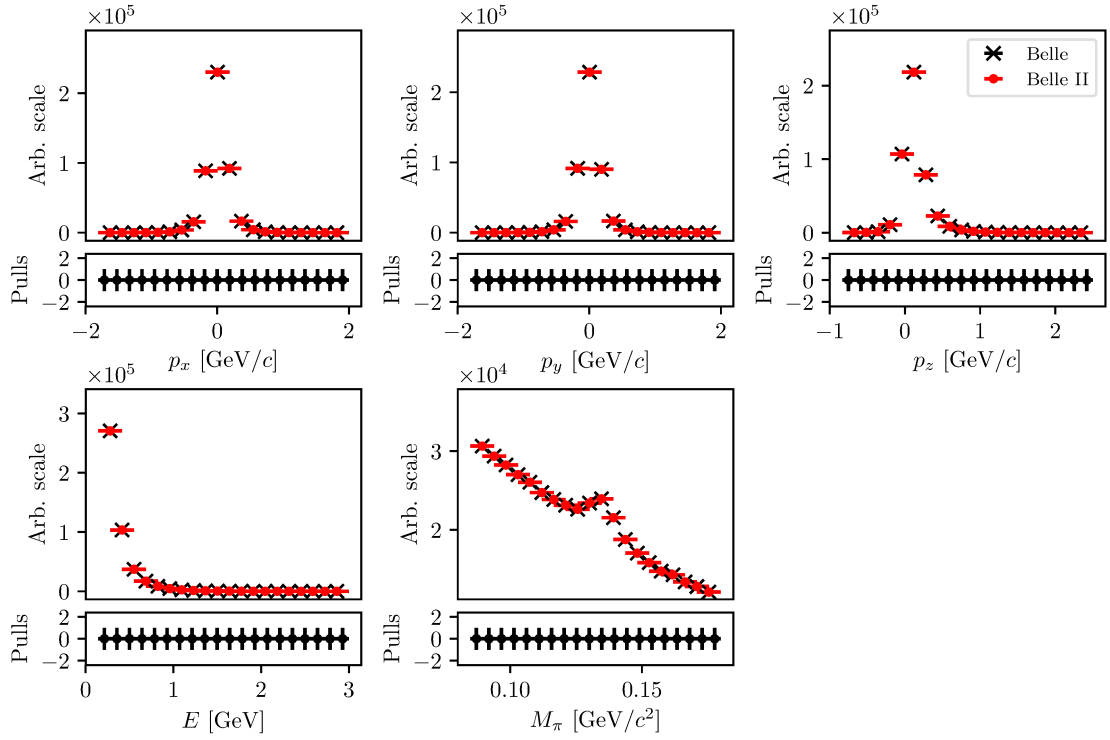


Figure 4.3: Some of the more important physical properties of  $\pi^0$  particles for Belle and Belle II in the conversion process. The histograms seem to overlap and the conversion is assumed to be successful.



# Chapter 5

## Event Reconstruction

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

### 5.1 Final State Particles Selection

The Belle detector is not able to detect all kinds of particles. Neutrinos are one such example, since they escapes detection, therefore we can only reconstruct the charged tracks in the decay, which are the two charged kaons ( $K$ ) and the light lepton ( $e$  or  $\mu$ ). These are some of the particles which are commonly referred to as the 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.

At this point in the analysis, we do not apply any specific particle selections yet, which results in a large number of available particles and their combinations, and, consequently, the computation time. In order to minimize this effect, we perform this part of the study on a smaller subset of the available generic MC, experiments 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 approximate 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 from any source, where true electrons/muons from the signal  $B$  meson decay are shown separately. *True* and *fake* implies that the particles are correctly or wrongly reconstructed, with respect to the generated MC truth. The difference between the true leptons from any source and those from signal decays is due to the distinct kinematics of the parent's decay.

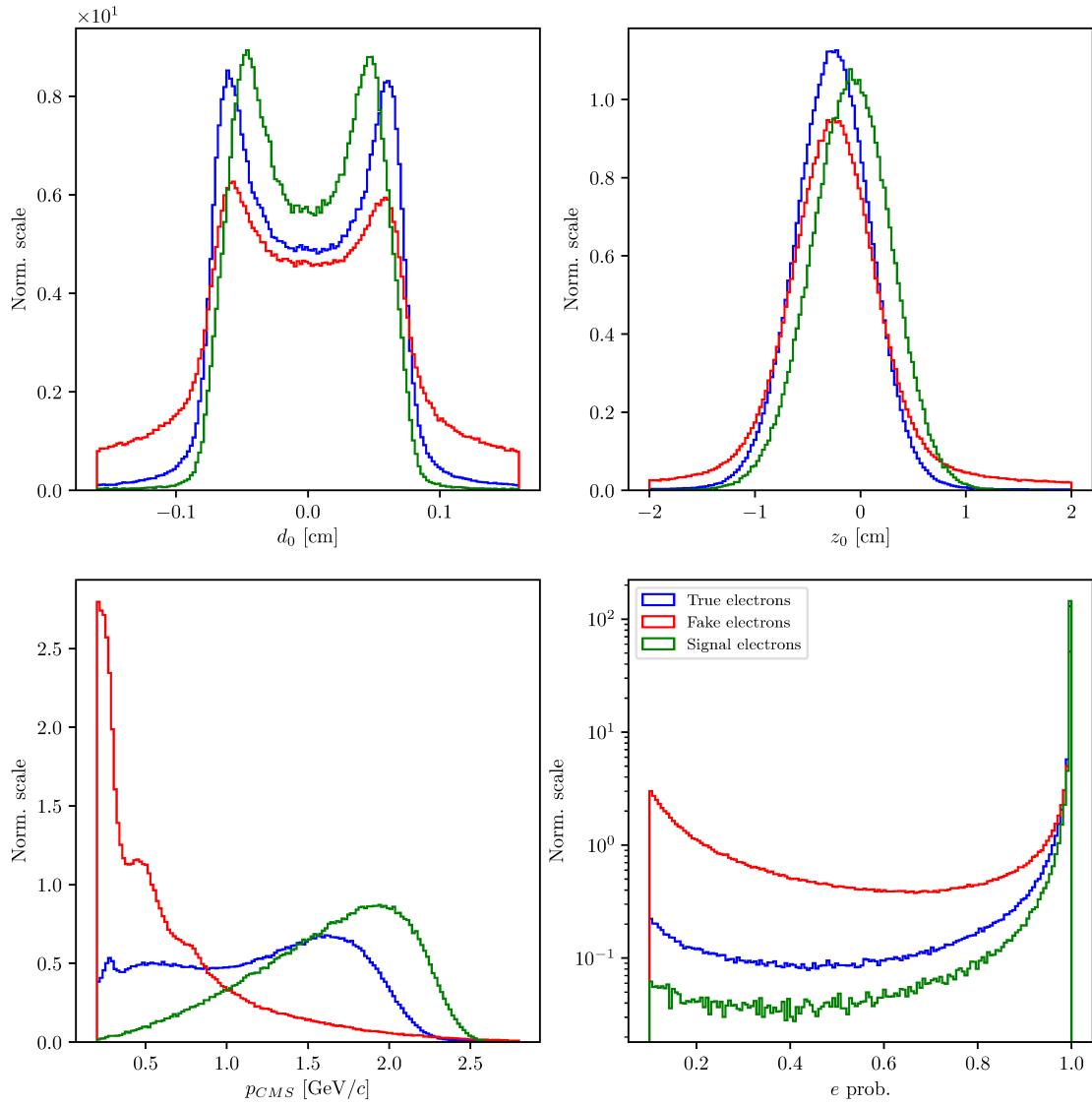


Figure 5.1: Normalized properties of true (blue), fake (red) electrons from any source, and true electrons from signal  $B$  candidates (green).

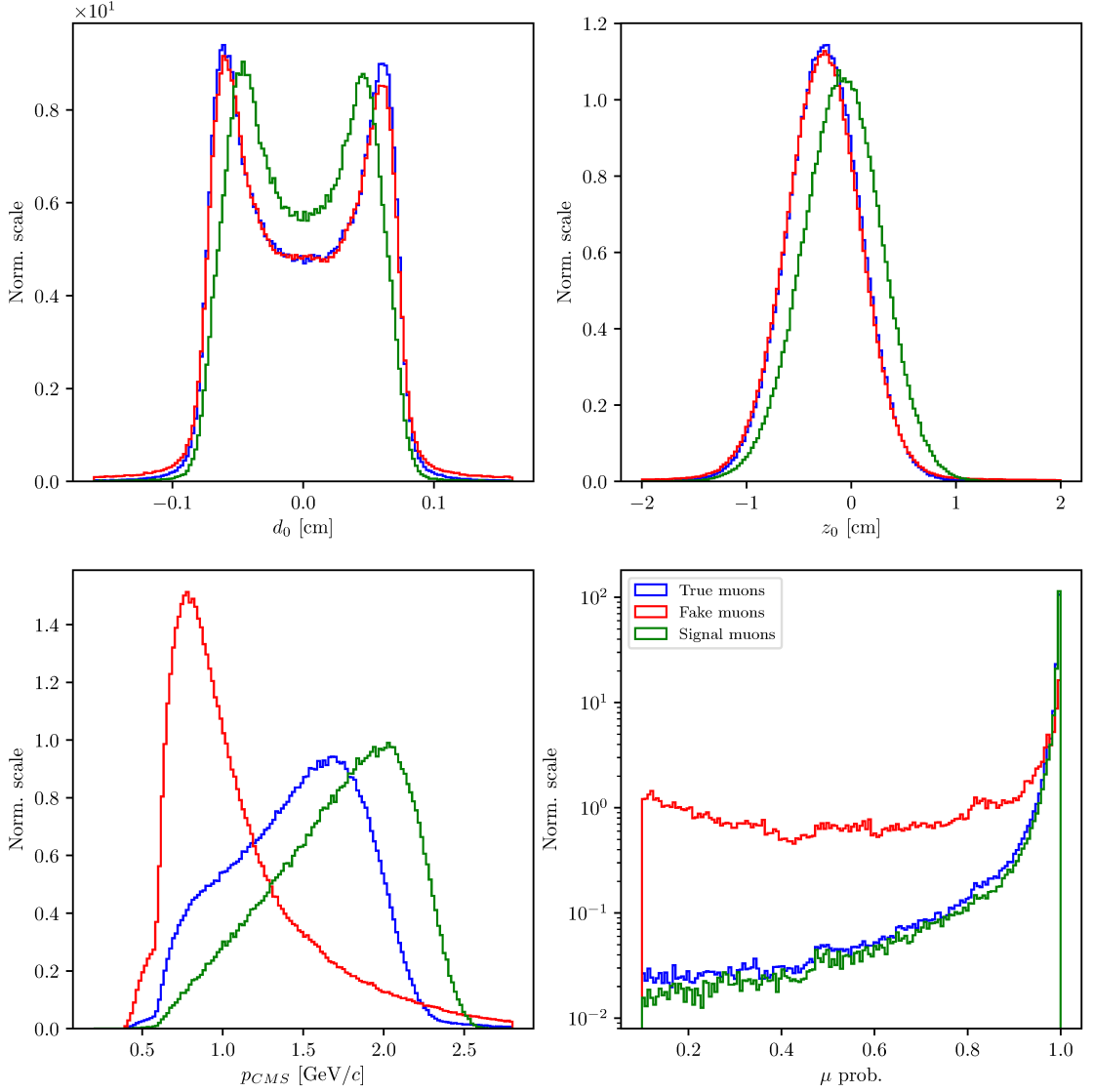


Figure 5.2: Normalized properties of true (blue), fake (red) muons from any source, and true muons from signal  $B$  candidates (green).

Based on these distributions, we can define a selection criteria

- $|d_0| < 0.1$  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 we can determine the optimal PID selection for electrons and muons, where we optimize the criteria by maximizing the standard definition of *figure of merit* ( $FOM$ ), defined in Eq. (5.1)

$$FOM = \sqrt{\mathcal{E}\mathcal{P}} \propto \frac{S}{\sqrt{S+B}}, \quad (5.1)$$

where the argument in the square root is the product of the efficiency ( $\mathcal{E}$ ) and the purity ( $\mathcal{P}$ ). 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  $FOM_1$  the signal  $S$  represents correctly reconstructed final state particles, while in  $FOM_2$  the signal  $S$  represents correctly reconstructed final state particles coming from correctly reconstructed  $B$  meson candidates. In both cases the background 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 selection criteria are based on PID cuts used for PID efficiency calibration. The optimal value for the PID cuts is equal to the largest available value for true leptons ( $FOM_1$ ), as well as for true leptons from signal  $B$  decays ( $FOM_2$ ), so selections via both methods are the same. The optimal PID selection criteria for leptons are then

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

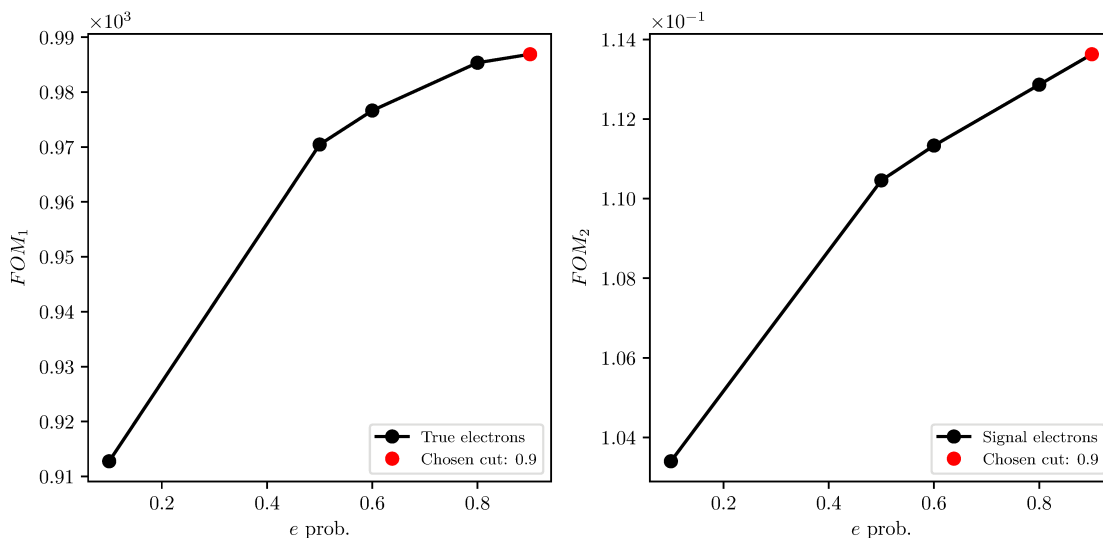


Figure 5.3:  $FOM$  optimizations of the PID selection for true electrons (left) and true electrons from signal  $B$  candidates (right).

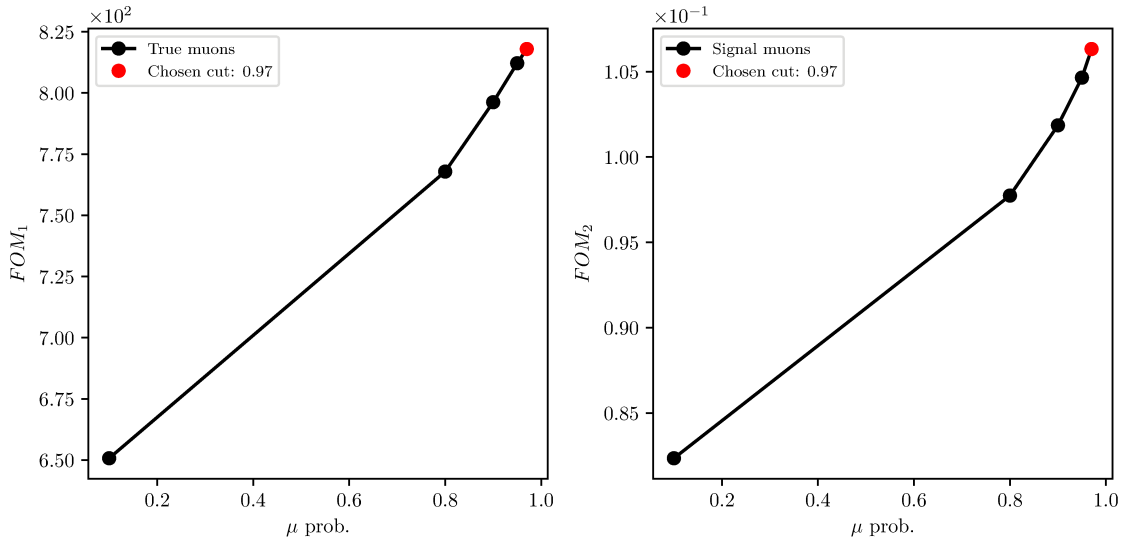


Figure 5.4:  $FOM$  optimizations of the PID selection for true muons (left) and true muons from signal  $B$  candidates (right).

## 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 true kaons from the signal decay are shown separately.

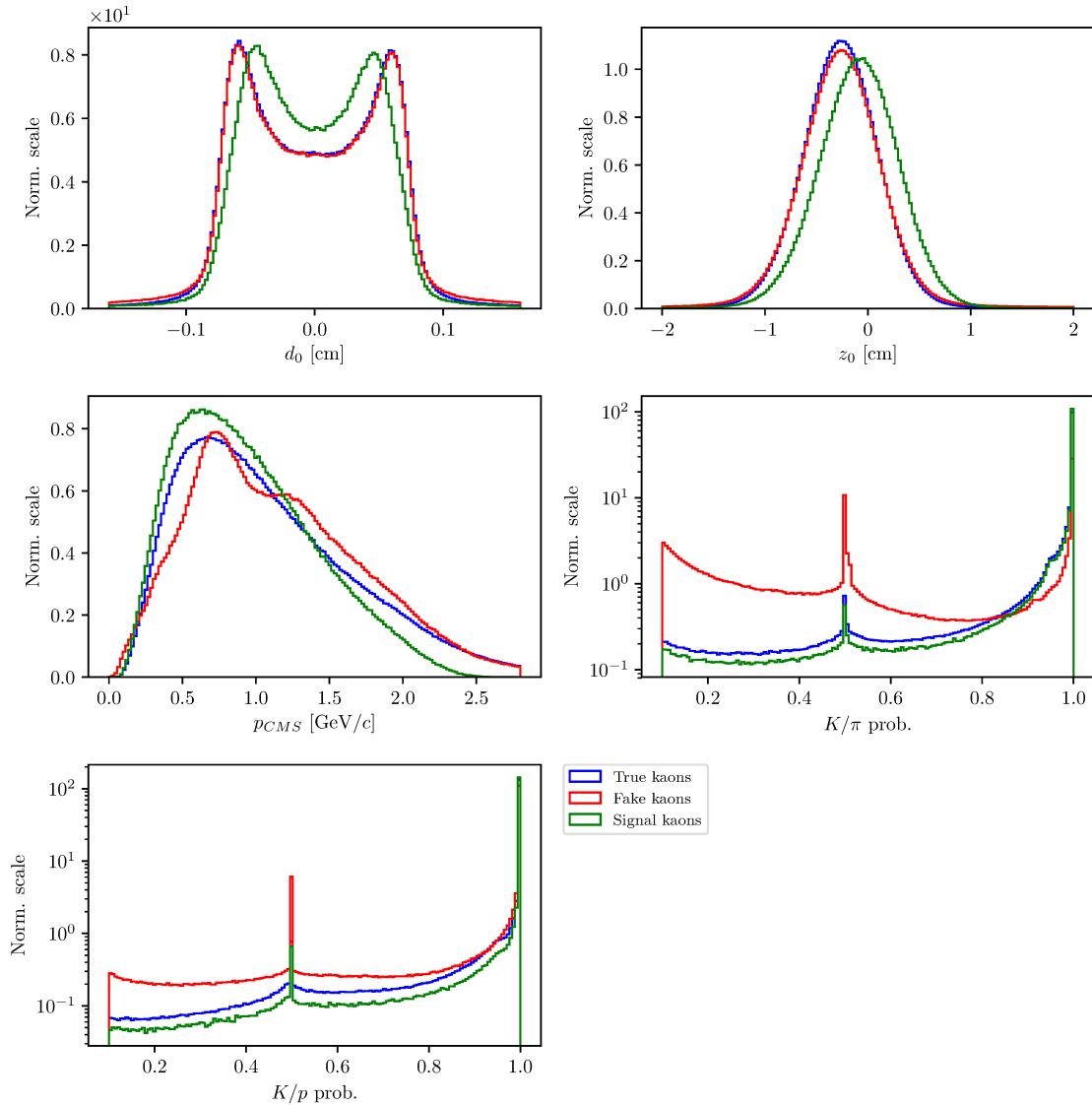


Figure 5.5: Normalized properties of true (blue), fake (red) and true kaons (green) from signal  $B$  candidates.

We define the kaon selection criteria 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 selection on  $K/\pi$ , and after that, the  $K/p$  separation probability. Figure 5.6 shows the optimization procedure for PID cuts on kaon candidates.



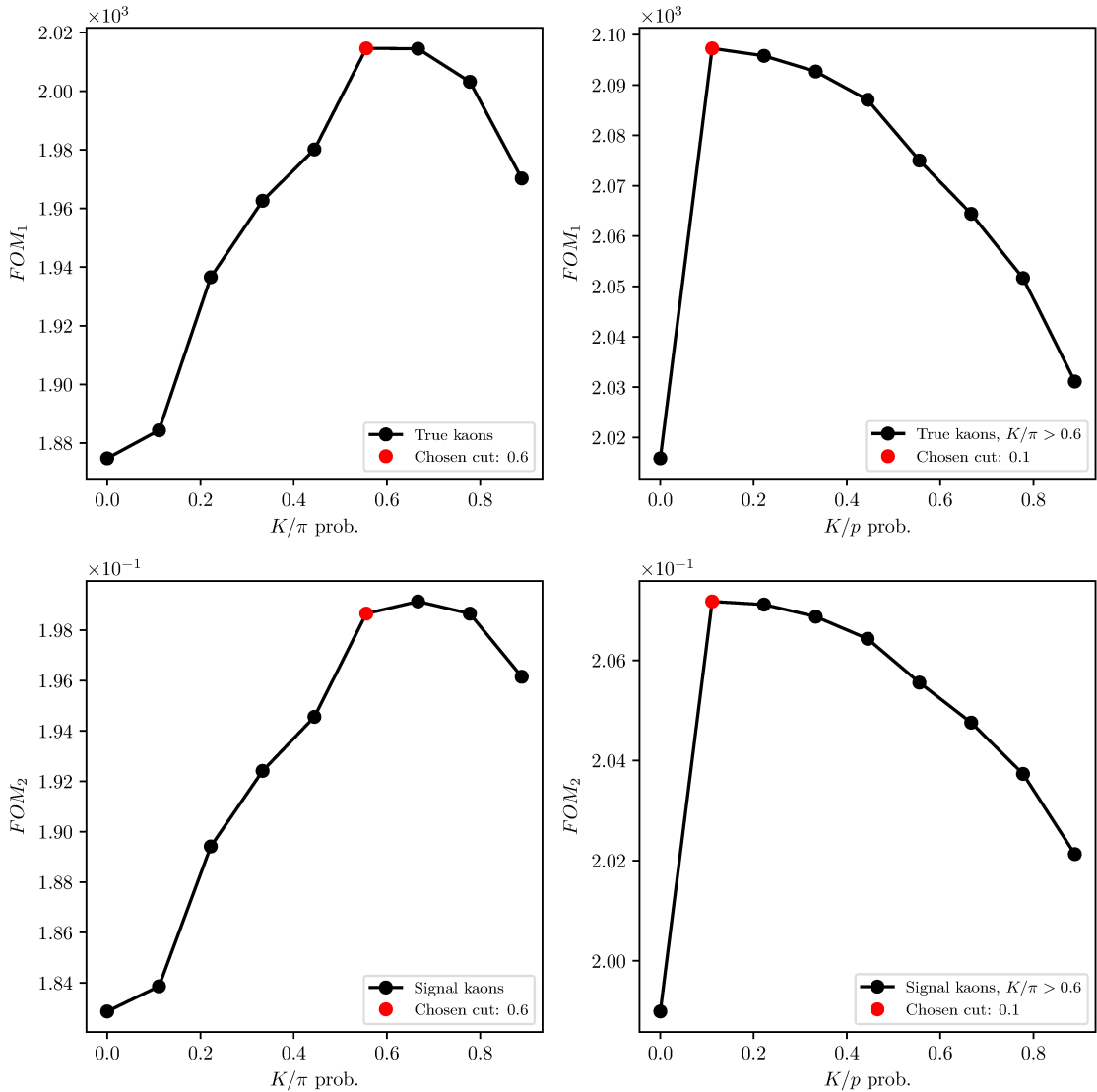


Figure 5.6:  $FOM$  optimizations of the PID selection for true kaons (top) and true kaons from signal  $B$  candidates (bottom). The plots on the left show the optimization of the first step for the  $K/\pi$  probability cut, and the plot on the right the  $K/p$  probability cut.

The optimal PID selection for kaons is

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

## 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 an under-optimized

selection based on the  $FOM$  optimizations, in order to optimize them later on the full Belle MC sample.

Since the missing neutrino escapes detection, we reconstruct the  $B$  mesons using the following final state particles

$$\begin{aligned} B^+ &\rightarrow K^+ K^- e^+, \\ B^+ &\rightarrow K^+ K^- \mu^+, \end{aligned}$$

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 the beam direction. We require that the fit converged and apply a selection on the minimal fit probability. The fit probability for signal and background  $B$  meson candidates is shown in Figure 5.7 (left). We perform an  $FOM$  optimization of this selection, 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  $b \rightarrow c$  transition.

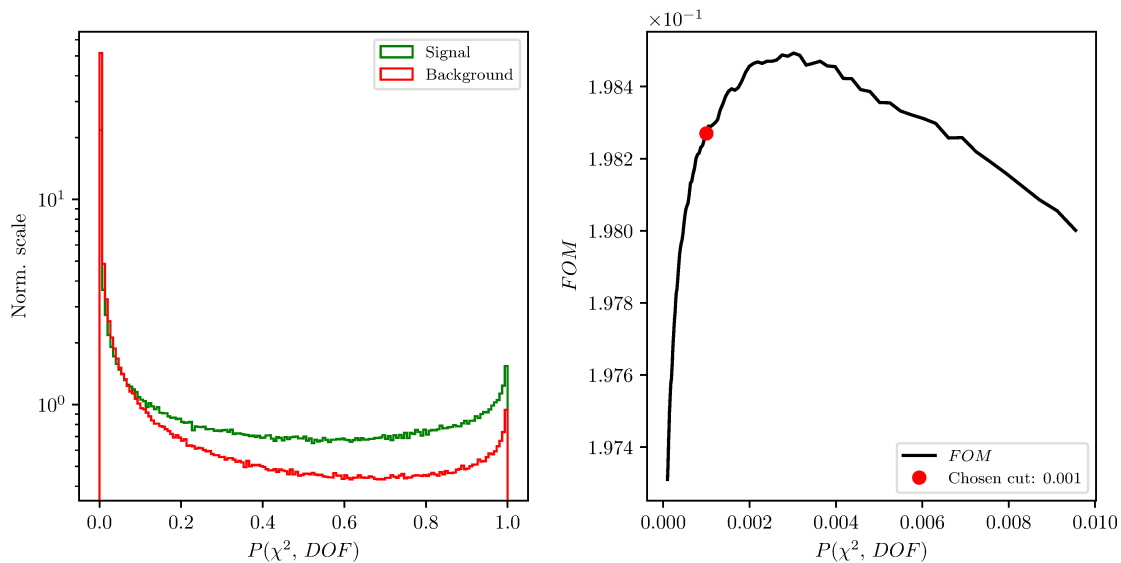


Figure 5.7: Normalized vertex fit probability distribution for signal and background  $B$  meson candidates in the logarithmic scale (left) and the  $FOM$  optimization of the vertex fit probability (right) for the subset of the full Belle MC sample.

The chosen pre-selection 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  $Y \rightarrow KK\ell$ ) and the nominal  $B$ , as

$$\mathbf{p}_\nu = \mathbf{p}_B - \mathbf{p}_Y, \quad (5.2)$$

$$\mathbf{p}_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_B E_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \quad (5.3)$$

$$\cos(\theta_{BY}) = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \quad (5.4)$$

where  $p$  denotes a scalar,  $\vec{p}$  a vector, and  $\mathbf{p}$  a four-vector. All the energy and momenta above are calculated in the CMS frame. The mass of the neutrino is equal to 0 up to a very good precision, so we use it in Eq. (5.3). Additionally, we 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

$$E_B = E_{CMS}/2, \quad (5.5)$$

$$|\vec{p}_B| = p_B = \sqrt{E_B^2 - m_B^2}, \quad (5.6)$$

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. This improves the resolution of  $\cos(\theta_{BY})$ , which leads to better signal-to-background discrimination.

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).

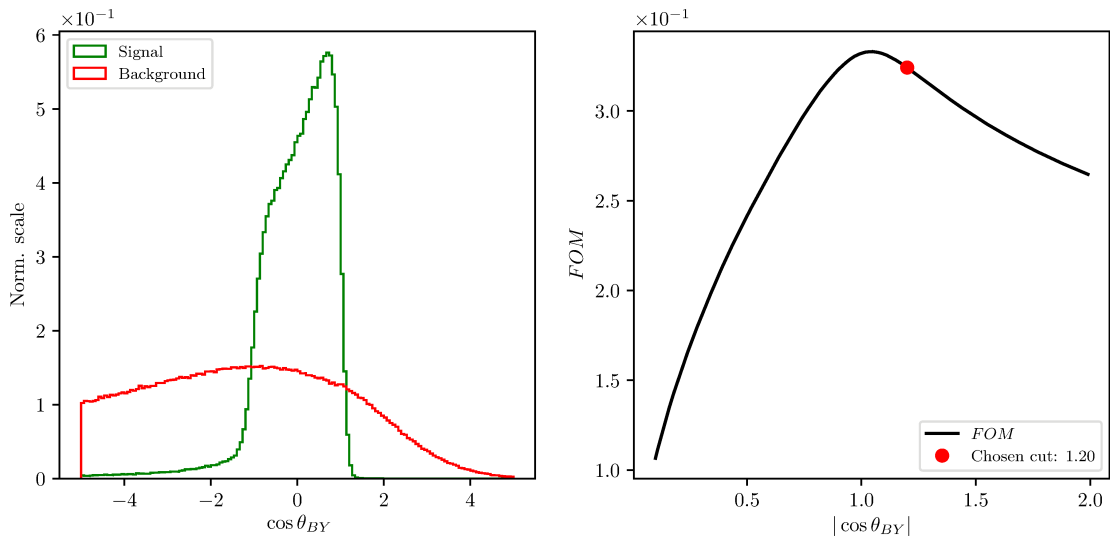


Figure 5.8: The normalized  $\cos\theta_{BY}$  distribution for signal and background  $B$  meson candidates (left) and the  $FOM$  optimization of the  $\cos\theta_{BY}$  variable (right) for the subset of the full Belle MC.

We again impose an under-optimized selection 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

We are not able to directly determine the four-momentum of the missing neutrino. However, due to the detectors geometry, which almost completely covers the full solid angle, and due to the well known initial conditions of the  $\Upsilon(4S)$  meson, it is possible to determine the kinematics of the missing neutrino indirectly. Specifically, this is performed by reconstructing the companion  $B$  meson via 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. 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. 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 background ones, which also 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

$$p_{miss} = p_{\Upsilon(4S)} - \sum_i^{\text{Event}} (E_i, \vec{p}_i), \quad (5.7)$$

$$p_{miss} = p_{\Upsilon(4S)} - \left( p_Y - \sum_i^{\text{Rest of event}} (E_i, \vec{p}_i) \right), \quad (5.8)$$

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

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

We assume all neutral particles to be massless. For charged tracks in the ROE, a mass hypothesis needs to be defined in order to determine the energy of the track. 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 calculate the square of the missing mass,  $m_{miss}^2$ , which is consistent with zero, if signal-side neutrino is the only missing particle in the event. The  $m_{miss}^2$  distribution is shown in Eq. (5.11).

$$p_\nu = p_{miss} = (E_{miss}, \vec{p}_{miss}), \quad (5.10)$$

$$m_{miss}^2 = p_{miss}^2 = p_\nu^2 = m_\nu^2 \approx 0. \quad (5.11)$$

Since the detector performance is not perfect, the distribution of the  $m_{miss}^2$  variable has a non-zero width. Additionally, a tail is introduced due to missing particles like neutrinos, other neutral undetected particles such as  $K_L^0$ , or simply missing tracks, due to detection failure. Figure 5.9 shows the distribution of  $m_{miss}^2$ , 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 the companion  $B$  meson decayed (semi-)leptonically, 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 radiated due to bremsstrahlung effects from final-state electrons and they typically do not have a big impact on the four-momentum of the final candidate. We denote this subset as the *perfect* signal. This subset is used to correctly define the clean-up parameters and is not used in any reconstruction steps, since we cannot know in data which neutral particles are actually missing.

Due to this fact, we impose a selection on the  $m_{miss}^2$  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 selection was chosen based on the optimization of the  $FOM$ , where in this case the definition of  $S$  were perfectly reconstructed signal candidates. The chosen selection is

- $|m_{miss}^2| < 3.0 \text{ GeV}/c^2$ ,

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

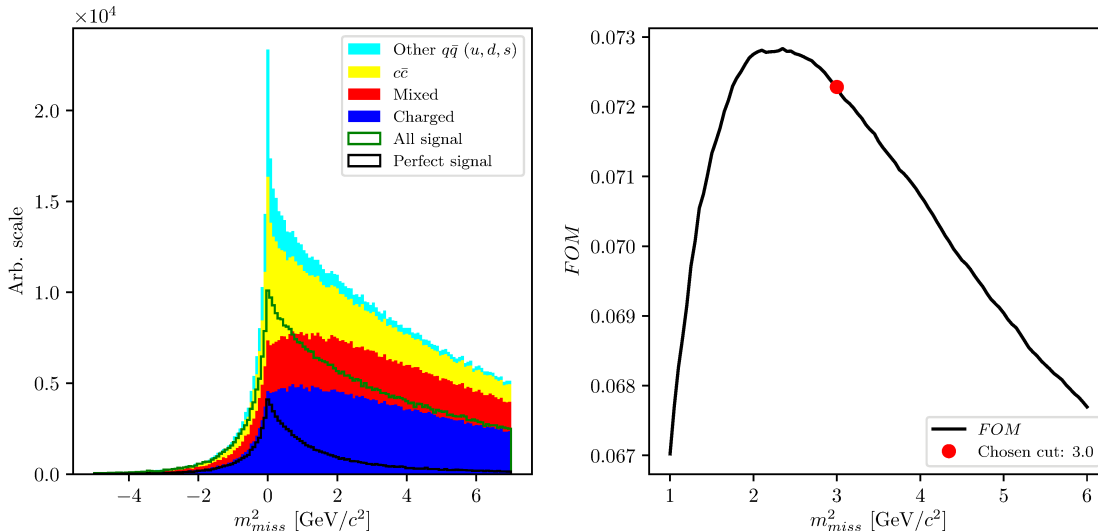


Figure 5.9:  $m_{miss}^2$  distribution for signal and various types of background. All signal (green) and perfect signal (black) are scaled up equally.

The main uncertainty of the neutrino four-momentum, defined in Eq. (5.10), comes from the 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,

$$p_\nu = (|\vec{p}_{miss}|, \vec{p}_{miss}). \quad (5.12)$$

This also fixes the neutrino mass to 0  $\text{GeV}/c^2$ . 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

$$\Delta E = E_B - E_{CMS}/2, \quad (5.13)$$

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

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

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

Figure 5.10 shows the distributions of  $\Delta E$  (left) and  $M_{BC}$  (right) for signal and major types of background after the pre-selection. Both signal components (all signal and perfect signal) 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 full and perfect signal.

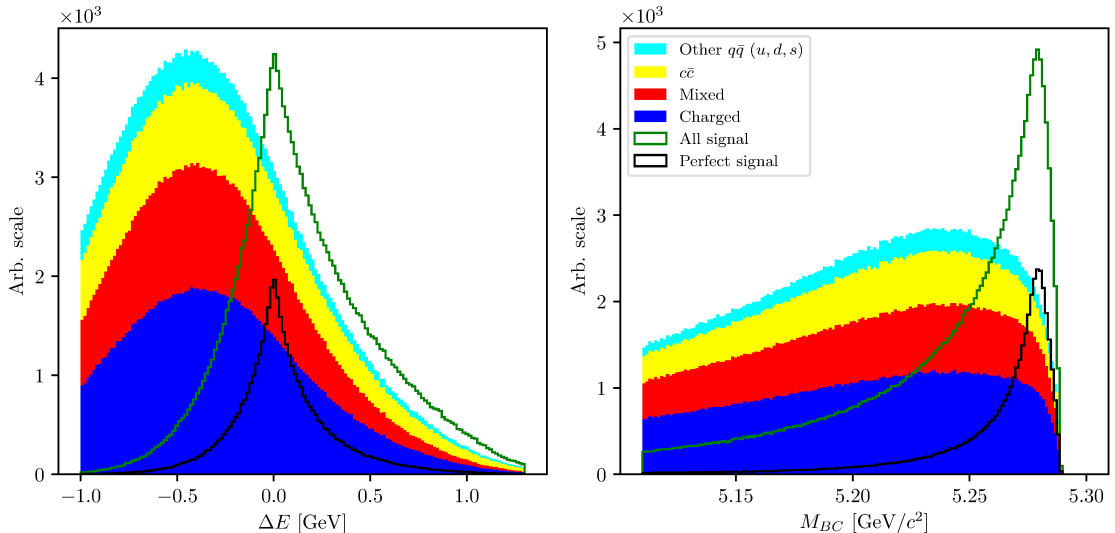


Figure 5.10: Distributions of  $\Delta E$  (left) and  $M_{BC}$  (right) for signal and major types of background after the pre-selection. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution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

### 5.3.1 $q^2$ calculation

$q^2$  is the squared Lorentz invariant of the four-momentum, also known as *momentum transfer squared*, which is transferred from the  $B$  meson to the  $W$  boson. There are several possible calculations of this variable which offer a different resolution. In this analysis, we follow the calculation procedure of  $q^2$  from [19], which yields the best resolution.

For correctly reconstructed events, Eq. (5.13) satisfies the condition  $\Delta E \approx 0$  within precision. It is possible to rescale the neutrino energy in such way that we fix  $\Delta E$  to zero, meaning

$$\Delta E' = (E_Y + \alpha E_\nu) - E_{CMS}/2 = 0. \quad (5.15)$$

and calculate a corrected value of  $M_{BC}$

$$M'_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}_Y + \alpha \vec{p}_\nu|^2}. \quad (5.16)$$

The neutrino momentum resolution dominates the  $\Delta E$  uncertainty, so the correction factor  $\alpha$  reduces this effect.

A second correction can be applied by rotating the direction of the neutrino momentum by a small angle with respect to the reconstructed one. Such an angle is chosen in order fix the value of  $M'_{BC}$  to the nominal mass of the  $B$  meson,  $m_B$ .

The corrected neutrino momentum is then fixed to expected values of  $\Delta E$  and  $M_{BC}$ , and is solely used for the  $q^2$  calculation. With  $p_\ell$  as the reconstructed lepton four-momentum, we define the  $q^2$  as

$$q^2 = q'^2 = (p_\ell + p_\nu)^2. \quad (5.17)$$

The  $q^2$  distributions and the corresponding resolutions are shown in Figure 5.11, along with other methods of  $q^2$  determination.

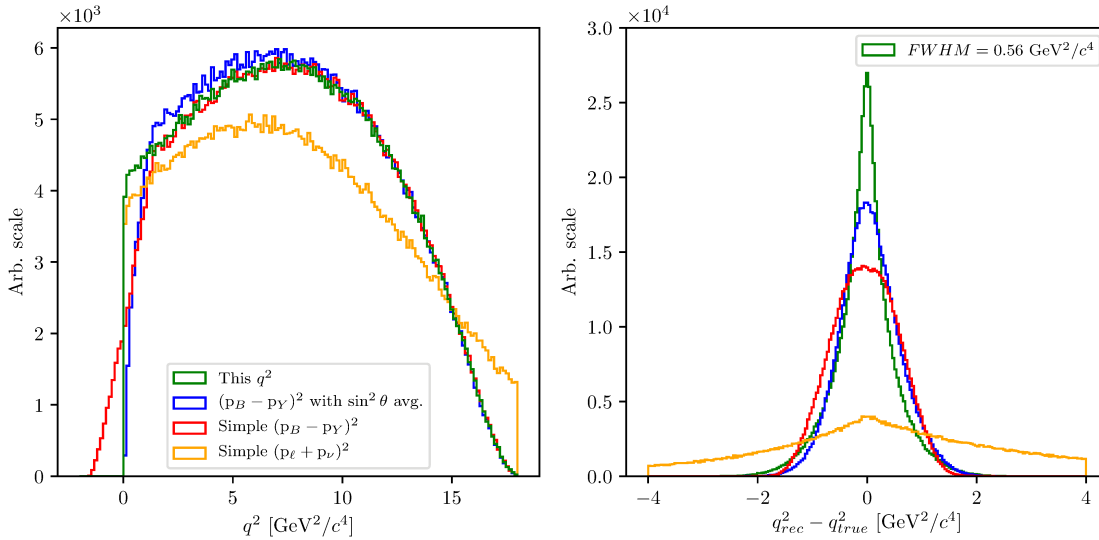


Figure 5.11: Distributions of  $q^2$  (left) and  $q^2$  resolution (right) for various methods of  $q^2$  calculation. The green distribution follows the procedure in [19], the blue distribution takes into account the weighted average of the  $B$  meson direction [20], and the red and orange distributions are straight-forward calculations with available information in the reconstruction. The  $q^2$  calculation in red assumes a resting  $B$  meson in the CMS frame, and the calculation in orange uses the neutrino four-momentum from Eq. (5.10).

One must bear in mind that even though the chosen calculation yields the most precise result, this does not affect the correctness of the  $q^2$  model, which was used in MC simulation (ISGW2 [21]). Since the signal decay has not been observed yet, we do not have a good description of the decay model, and we treat this as a source of systematic uncertainty in this analysis.



## 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 of the 10 streams of the full available Belle generic MC. After obtaining the full reconstructed sample, the first task is to optimize the under-optimized selection criteria from the previous stage. Repeating the procedure on the full sample results in the  $FOM$  shapes shown in Figure 5.12, with the optimal selection criteria

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

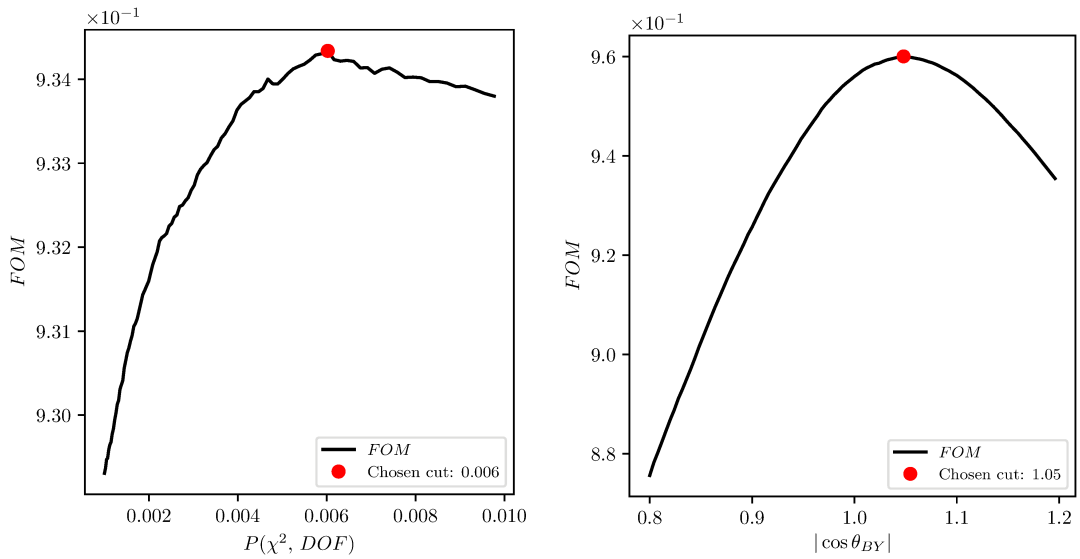


Figure 5.12: The  $FOM$  optimization of the vertex fit probability (left) and the  $\cos \theta_{BY}$  variable (right) for the full Belle MC sample.

With further optimizations, we will be fine-tuning the signal-to-background ratio. The most prominent and distinguishable part of our signal is the perfect 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. The 2D  $FOM$  optimization of the optimal  $\Delta E$  and  $M_{BC}$  signal region is shown in Figure 5.13.

The signal region is defined as

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

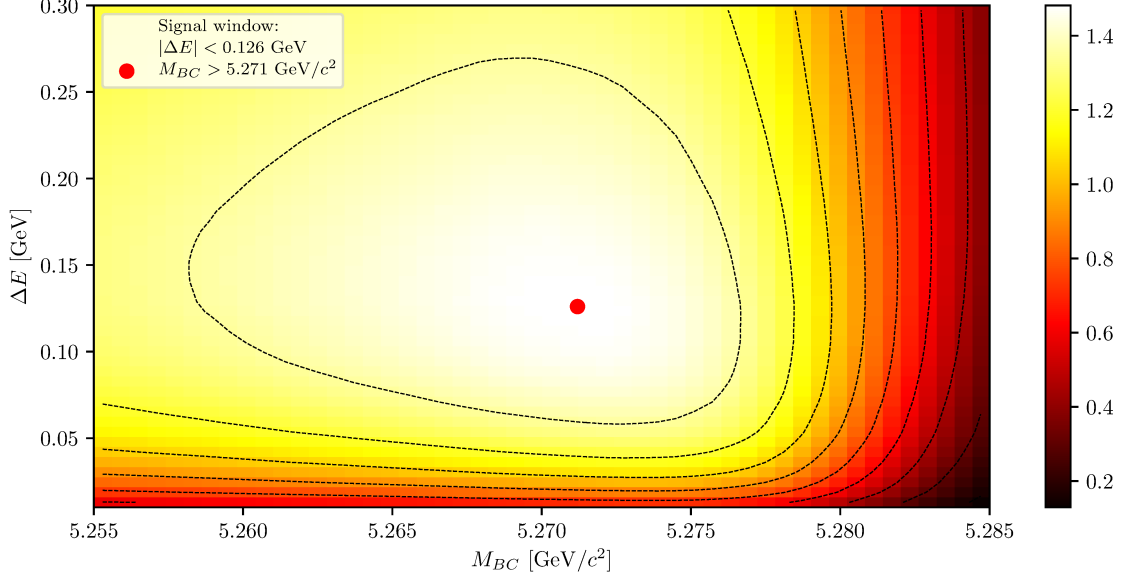


Figure 5.13: 2D *FOM* optimization of the signal region definition, where most of the perfectly reconstructed candidates are located.

With the signal window defined, we can tighten the selection on  $m_{miss}^2$ , which we intentionally left loose before the signal categorization. With the *FOM* optimization of perfectly reconstructed candidates inside the signal region, shown in Figure 5.14, the optimal selection of  $m_{miss}^2$  range is

- $|m_{miss}^2| < 0.975 \text{ GeV}/c^2$ .

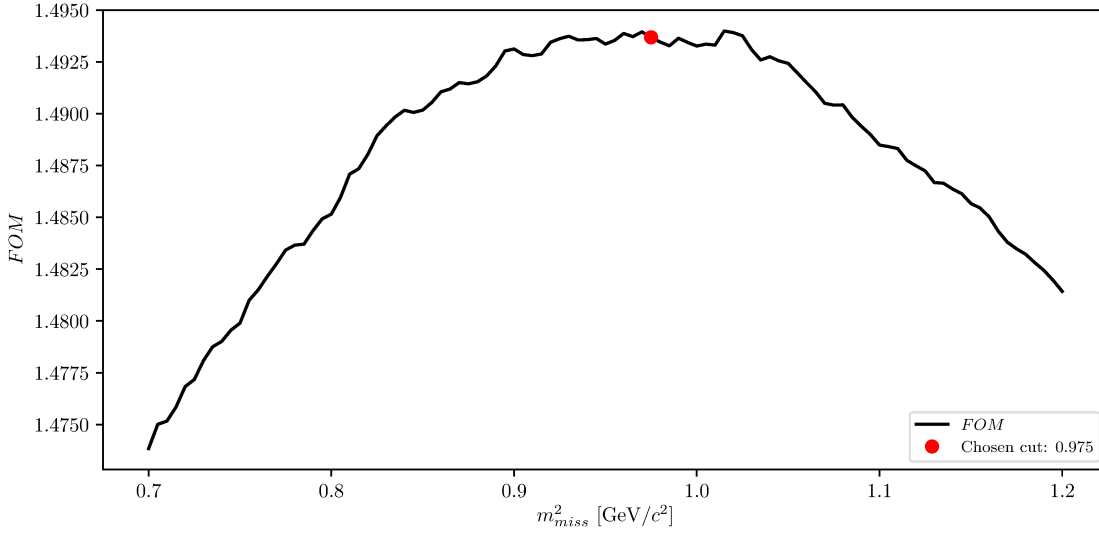


Figure 5.14: *FOM* optimization of the  $m_{miss}^2$  selection in the signal region.

## 5.5 Charge product categorization

The missing information, due to an escaping neutrino in our reconstructed channel, is replaced by the 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  $\Delta E \approx 0$ ,  $m_{miss}^2 \approx 0$ , 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

$$q_{B^\pm} q_{B^\mp} = -1, \quad (5.18)$$

however, this value is distributed due to missing, or additional background charged particles in the ROE. Figure 5.15 shows various signal distributions of  $\Delta E$  and  $M_{BC}$  in arbitrary (left) and normalized (right) scales. We find the relative ratios of 67.86 % and 32.14 % for correct and wrong values of the charge product. Correctly reconstructed events represent the majority of the signal candidates and also have the best resolution in  $\Delta E$  and  $M_{BC}$ , hence we proceed with the analysis by imposing the selection in Eq. 5.18.

While this selection 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  $1.09 \times 10^{-3}$ .

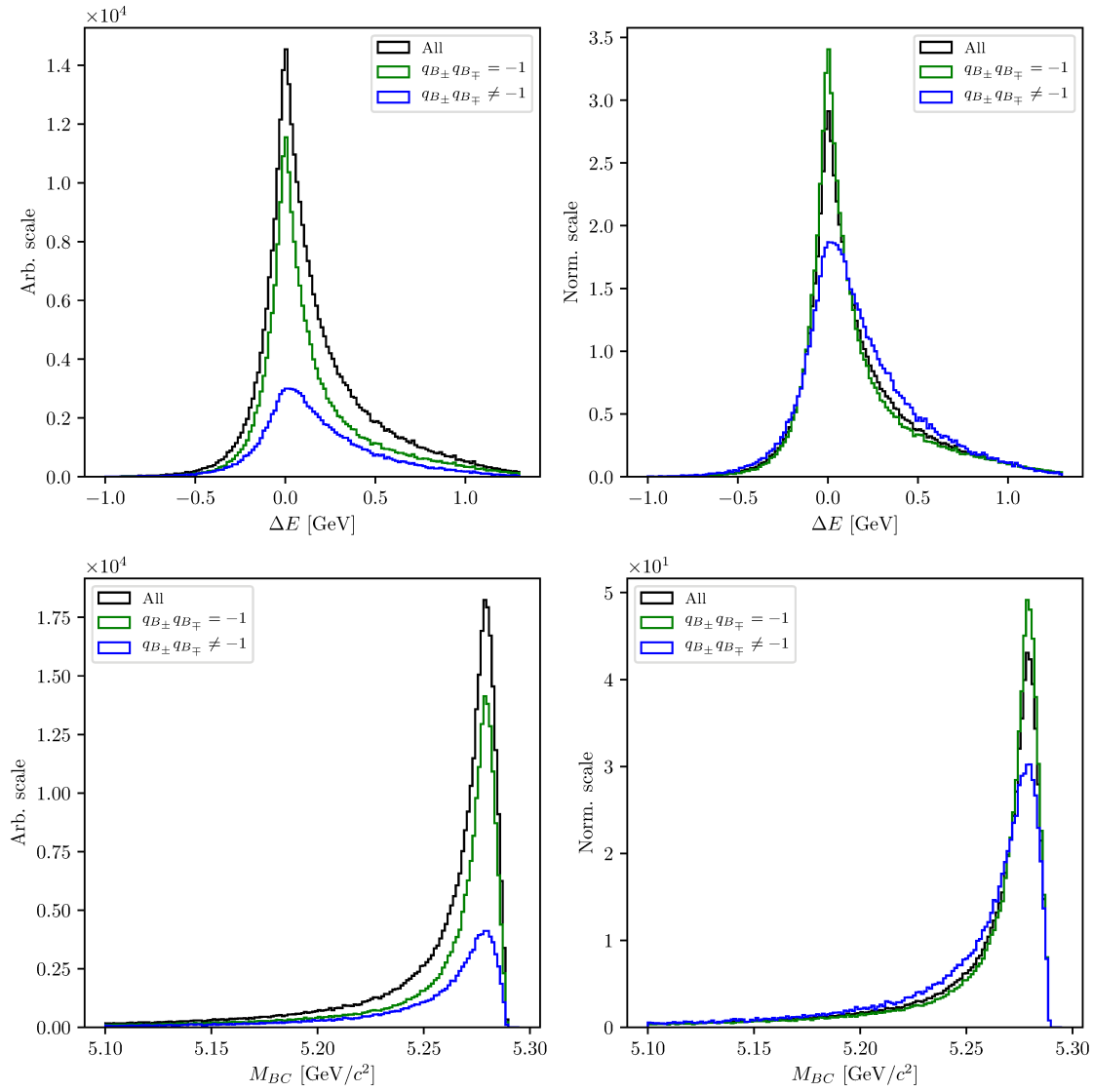


Figure 5.15: Signal distributions of  $\Delta E$  and  $M_{BC}$  based on the charge product of both  $B$  mesons in the event. The plots on the left show the distributions in an arbitrary scales, while the plots on the right show the normalized distributions.

## 5.6 Selection Summary

In this section, one can find the full summary of final selection criteria 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 selection:  $P(\chi^2, DOF) > 6 \times 10^{-3}$ ,  $|\cos \theta_{BY}| < 1.05$ ,  $|m_{miss}^2| < 0.975$  GeV/ $c^2$ ,
  - fit region selection:  $\Delta E \in [-1.0, 1.3]$  GeV,  $M_{BC} \in [5.1, 5.295]$  GeV/ $c^2$ ,
  - signal region selection:  $|\Delta E| < 0.126$  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 the  $B$  meson reconstruction on Signal MC with a generic companion  $B$  meson. For every correctly reconstructed signal  $B$  meson candidate we save the necessary information for each MVA step (e.g. 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 Machine Learning Setup

Throughout the analysis we use the Fast-BDT algorithm as the main machine learning model, in order to perform the multivariate analysis steps, unless explicitly stated otherwise. The Fast-BDT algorithm [22] is an optimized version of the *boosted decision trees* (BDT) algorithm, which is one of the most commonly used in the world of machine learning. As the name suggests, it is based on the method of decision trees. The decision tree algorithm generates decision trees from the data to solve classification and regression problems, where each decision tree splits the data in at least two groups. A schematic of such a *decision forest* is shown in Figure 6.1, where the forest consists of several trees, each with a certain depth.

The number of trees in the classifier and the depth level of a single tree are two hyper-parameters of the machine learning classifier, which we optimize in each case of applying the MVA step. More information about the MVA training, hyper-parameter optimization and feature importance for each MVA step in this chapter can be found in Appendix A.

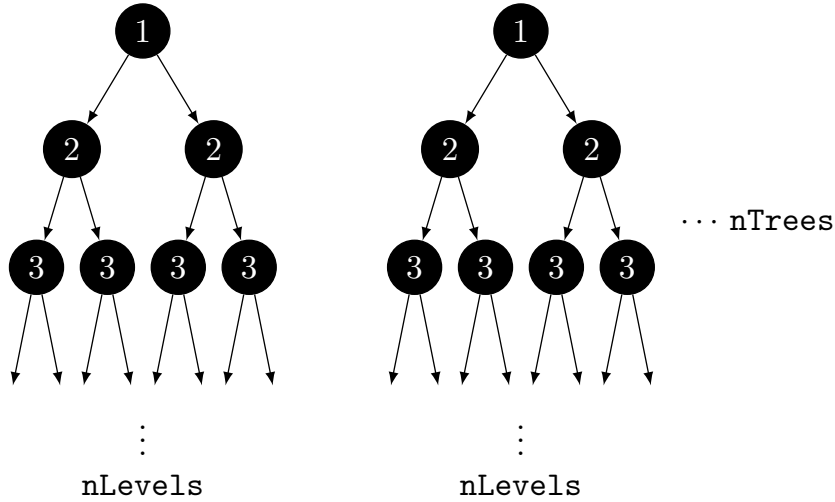


Figure 6.1: A schematic representation of decision trees in a decision forest.

## 6.2 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 processes such as  $\pi^0 \rightarrow \gamma\gamma$  decays, which are interesting for physics. 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, which we are usually not interested in, except in cases involving material studies. Photons of the first kind should be taken into account when calculating the missing four-momentum, while photons, which are not directly related to the collision, add extra energy and momentum to the event. 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.



## 6.2.1 $\pi^0$ 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, which 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 the invariant mass in the range of  $M \in [0.10, 0.16]$  GeV. After that, we perform a mass-constrained fit on all  $\pi^0$  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 probability 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 6.2.

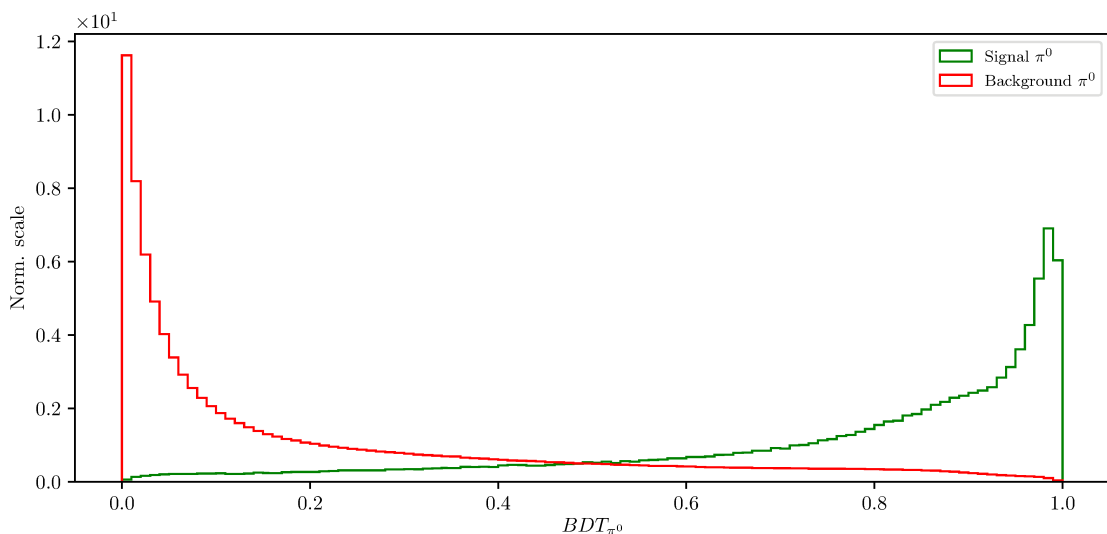


Figure 6.2: Classifier output of the  $\pi^0$  training for signal and background  $\pi^0$  candidates.

### 6.2.2 $\gamma$ 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  $\pi^0$  selection criteria, as defined in the previous step. Since it is 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 6.3.

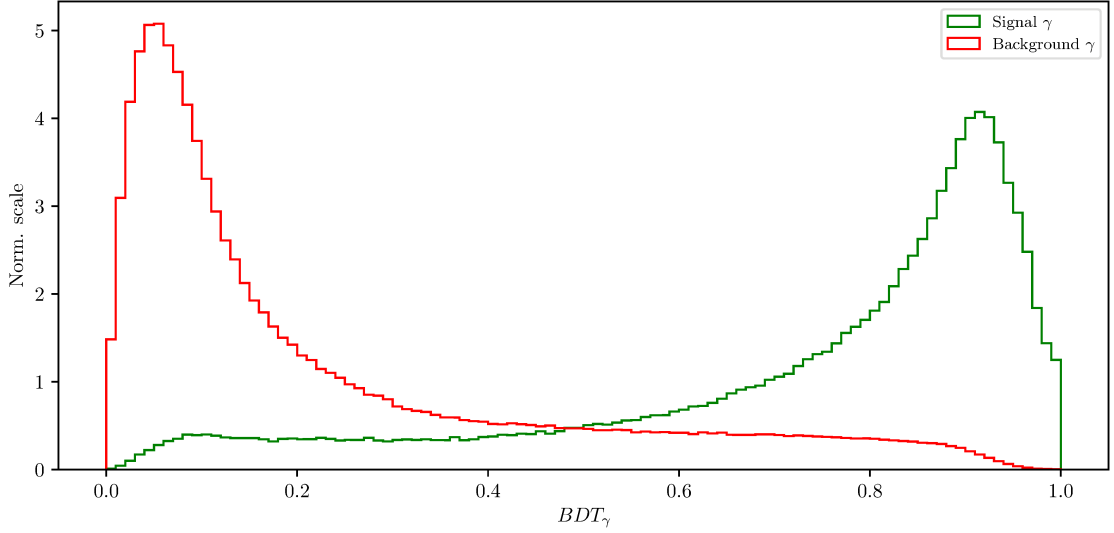


Figure 6.3: Classifier output of the  $\gamma$  training for signal and background  $\gamma$  candidates.

With the classifier output at hand, we apply a selection to the photon particle list. The selection optimization is shown in Figure 6.4 (left), with the optimal selection on the  $\gamma$  classifier output at

- $BDT_\gamma > 0.5045$ .

Figure 6.4 (right) shows the LAB frame momentum of the photons in the logarithmic scale, before and after the selection. The signal efficiency and background rejection of this clean-up step are

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

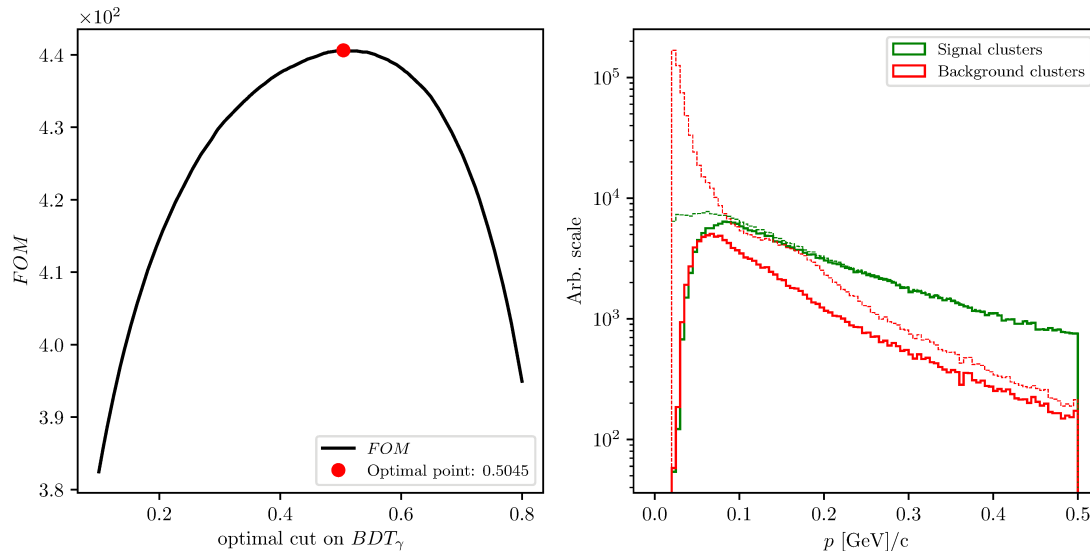


Figure 6.4: The  $FOM$  of the classifier output optimization (left) and momentum magnitude in the LAB frame (right) of signal and background photon candidates before (dashed) and after (solid) the optimal selection.

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

### 6.3 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 four-momentum of the charged particle, which is then taken into account 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  $K_S^0 \rightarrow \pi^+\pi^-$ ,
- (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 6.5.

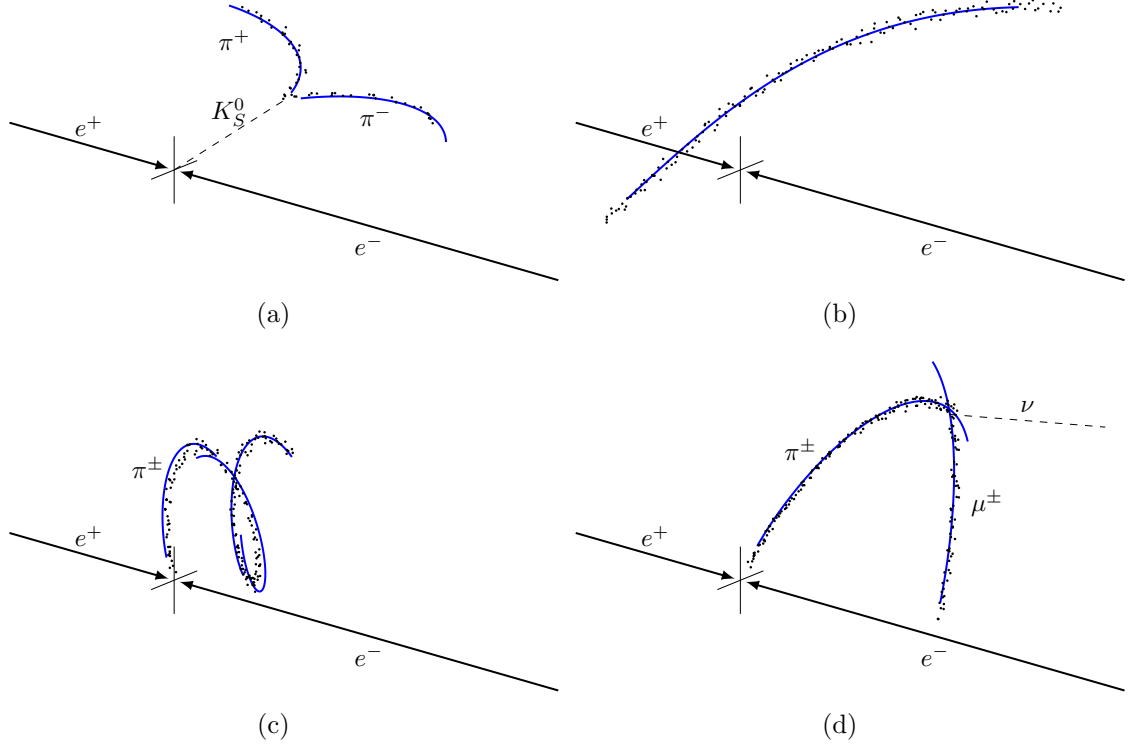


Figure 6.5: (a) Tracks from long-lived neutral particles, which decay away from the IP region, (b) random reconstructed background tracks, (c) low-momentum particles, which curl in the magnetic field, (d) in-flight decays of particles, which produce a "kink" in the trajectory.

It is obvious that tracks from the same momentum source should only be taken into account once, or, in the 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 do not 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.

### 6.3.1 Tracks from Long-lived Particles

The first step in the tracks clean-up is taking care of the tracks from long-lived particles, such as  $K_S^0 \rightarrow \pi^+\pi^-$ ,  $\gamma \rightarrow e^+e^-$  and decays of  $\Lambda$  baryons. Here we only focus on  $K_S^0$ , since they are the most abundant. This step is necessary because of 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 6.6 shows the distribution of the  $K_S^0$  invariant mass for signal and background candidates, before and after the selection cut on the classifier output. The momentum of selected  $K_S^0$  candidates is added to the ROE, while the daughter tracks are discarded from our set.

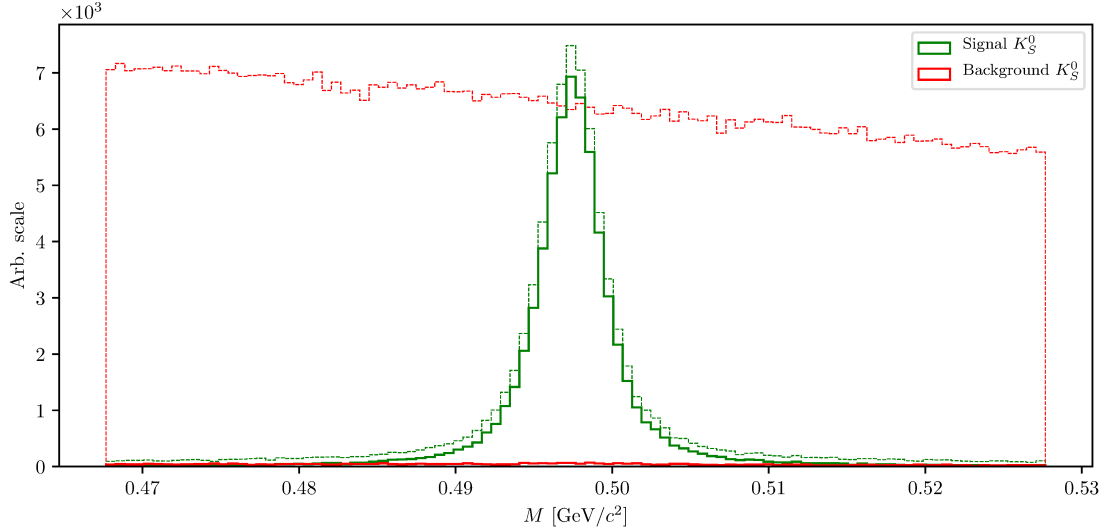


Figure 6.6: Invariant mass of the  $K_S^0$  candidates before (dashed) and after (solid) the selection on the Neural Network classifier for signal (green) and background candidates (red). Signal peaks at nominal  $K_S^0$  mass, while background covers a wider region.

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

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

### 6.3.2 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 selection 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 3D opening angle between such two tracks is very close to  $0^\circ$  or  $180^\circ$ , since the tracks deviate only slightly from the initial direction, but can also be reconstructed in the opposite way. Figure 6.7 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.

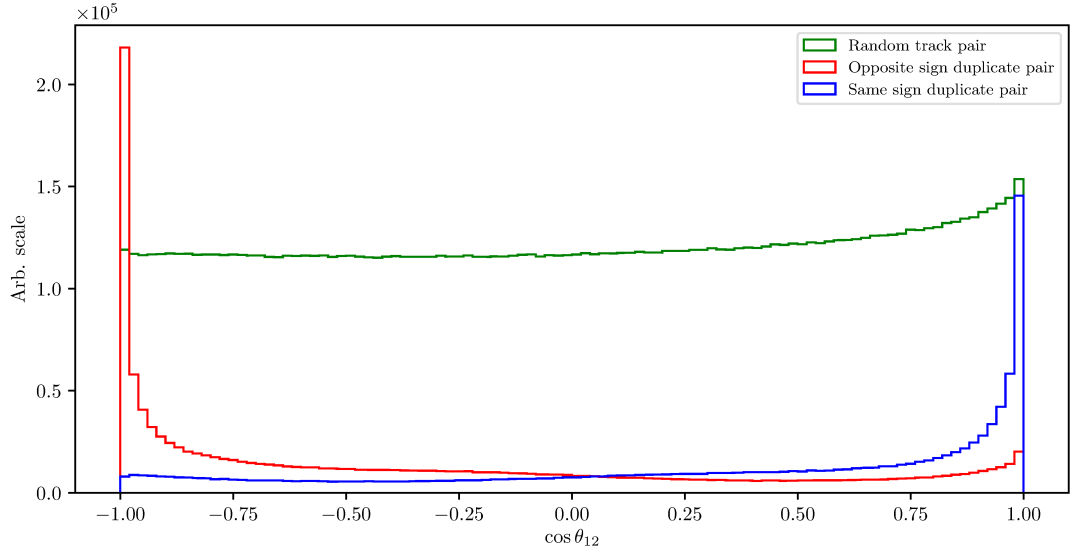


Figure 6.7: Distribution of the angle between two tracks in a single pair for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

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 6.8 shows the momentum and momentum difference of tracks which belong to a random or a duplicate track pair.

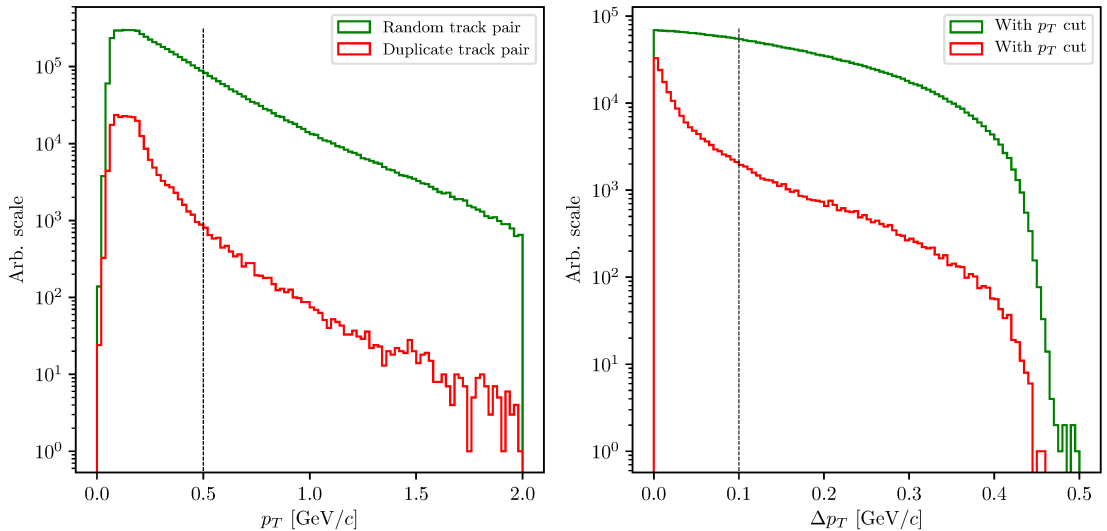


Figure 6.8: Distribution of transverse momentum  $p_T$  (left) and transverse momentum difference  $\Delta p_T$  (right) for all tracks coming from random (green) or duplicate track pairs (red). The plot on the right already includes the selection on  $p_T$  from the plot on the left.

We impose a selection of

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

in order to minimize the number of random track pairs, while retaining a high percentage of duplicate track pairs. After applying the selection criteria defined in this chapter, the final distribution of the angle between two tracks is shown in Figure 6.9.

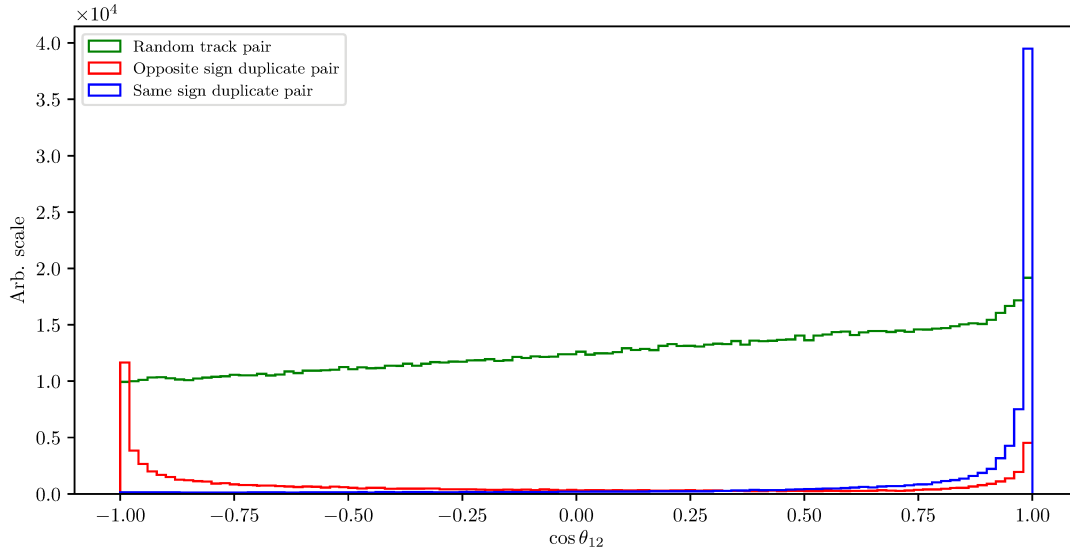


Figure 6.9: Distribution of the angle between two tracks in a single pair after applying the selection defined in this section. The distributions are shown for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

### Training the duplicate track pair MVA

This final sample of track pairs is used in the MVA training to recognize duplicate track pairs. 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 6.10.

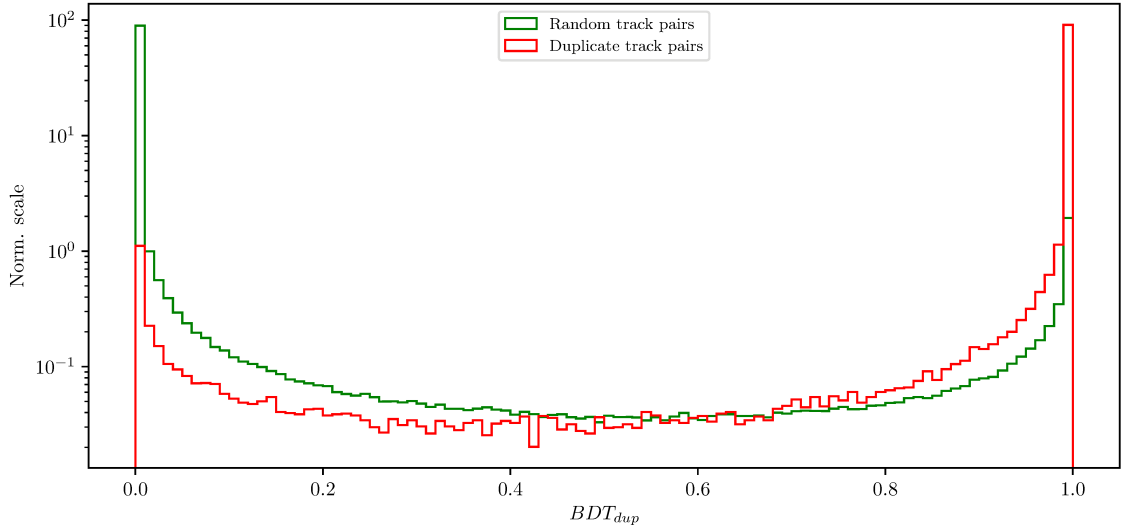


Figure 6.10: Classifier output of the track pair training for random track pairs and duplicate track pairs.

The  $FOM$  function for optimal selection is shown in Figure 6.11 (left), along with the angle between the two tracks before and after the optimal selection (right). The optimal duplicate track selection is

- $BDT_{duplicate} > 0.9985$ .

The signal efficiency and background rejection for duplicate pair candidates after this selection 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.

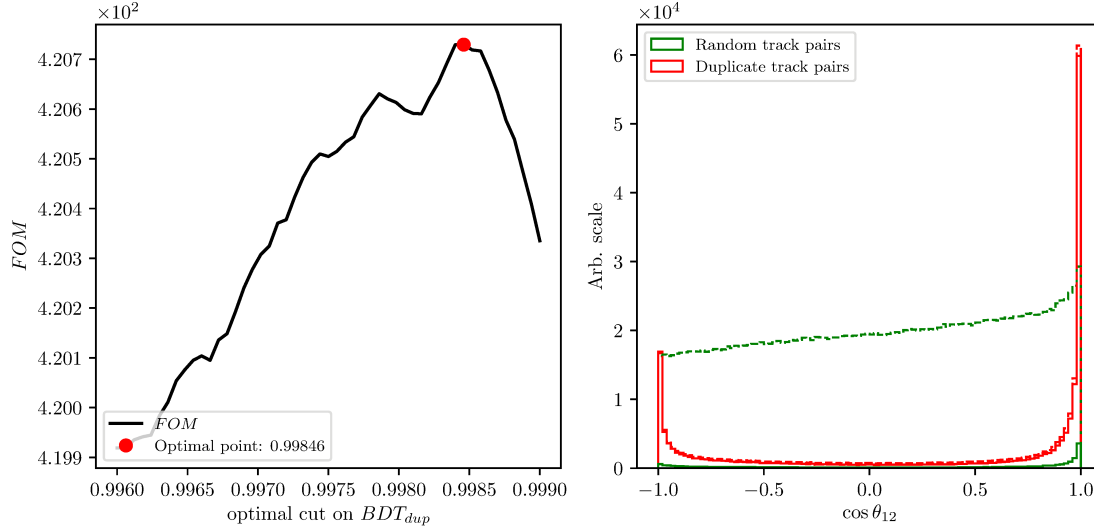


Figure 6.11: The optimization of the  $FOM$  function for the cut on classifier output (left) and distribution of the angle between two tracks in a single pair before (dashed) and after (solid) applying the optimal cut on the output classifier for random and duplicate track pairs (right).

### 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

$$\Delta f = f_{this} - f_{other}, \quad (6.1)$$

where  $f$  is an arbitrary variable from the list of track quantities in Section 6.3.2. 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

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

where we compare all components of track momentum to the true values. If the condition  $\Delta \chi^2 > 0$  is satisfied, then *this* track is probably 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 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. If you are a part of my defense committee and actually read this before my thesis defense, let me know, I owe you a bottle of whiskey. 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 used in the MVA training 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, based on the  $\Delta\chi^2 > 0$  condition.

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 output classifier information is added 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

$$\Delta BDT_{final} = BDT_{final}^{this} - BDT_{final}^{other}, \quad (6.3)$$

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 6.12.

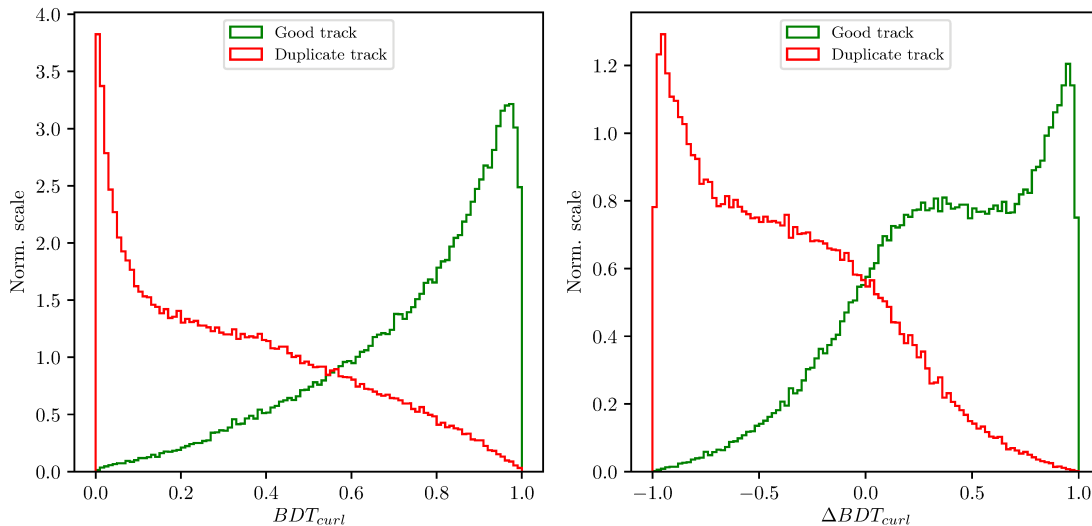


Figure 6.12: Classifier output of the MVA training for curling track recognition (left) and difference of the classifier output, calculated for each track in a track pair (right).

Finally, we select all duplicate tracks, which survive the selection

$$\Delta BDT_{final} > 0, \quad (6.4)$$

and discard them from our ROE. We can check the performance of our duplicate track classifier by applying the procedure to 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.

## 6.4 Belle Clean-up

For comparison, we define the Belle clean-up, used standardly at Belle, which is much simpler and relies only on a set of basic selection criteria for neutral and charged particles. This clean-up procedure is not applied in addition to our ROE clean-up, but separately, only for comparison.

In the case of photons, only a selection on the photon energy is applied, depending on the region where the photon hit the relevant part of the detector. The photon selection is summarized in Table 6.2.

	$17^\circ < \theta < 32^\circ$	$32^\circ < \theta < 130^\circ$	$130^\circ < \theta < 150^\circ$
$E_\gamma$	$> 100 \text{ MeV}$	$> 50 \text{ MeV}$	$> 150 \text{ MeV}$

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

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

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

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

	$p_T < 250 \text{ MeV}/c$	$250 \text{ MeV}/c < p_T < 500 \text{ MeV}/c$	$p_T > 500 \text{ MeV}/c$
$ d_0 $	$< 20 \text{ cm}$	$< 15 \text{ cm}$	$< 10 \text{ cm}$
$ z_0 $	$< 100 \text{ cm}$	$< 50 \text{ cm}$	$< 20 \text{ cm}$

Table 6.3: Track selection for the Belle clean-up procedure. Different selection criteria are applied to tracks in different  $p_T$  regions.

## 6.5 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 applying selection criteria 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 selection criteria, defined in Section 5.6, except for the signal categorization. Figure 6.13 (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 6.13 (right). The signal MC sample in case of the Belle clean-up also shows a slight improvement in the resolution after the procedure, 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.

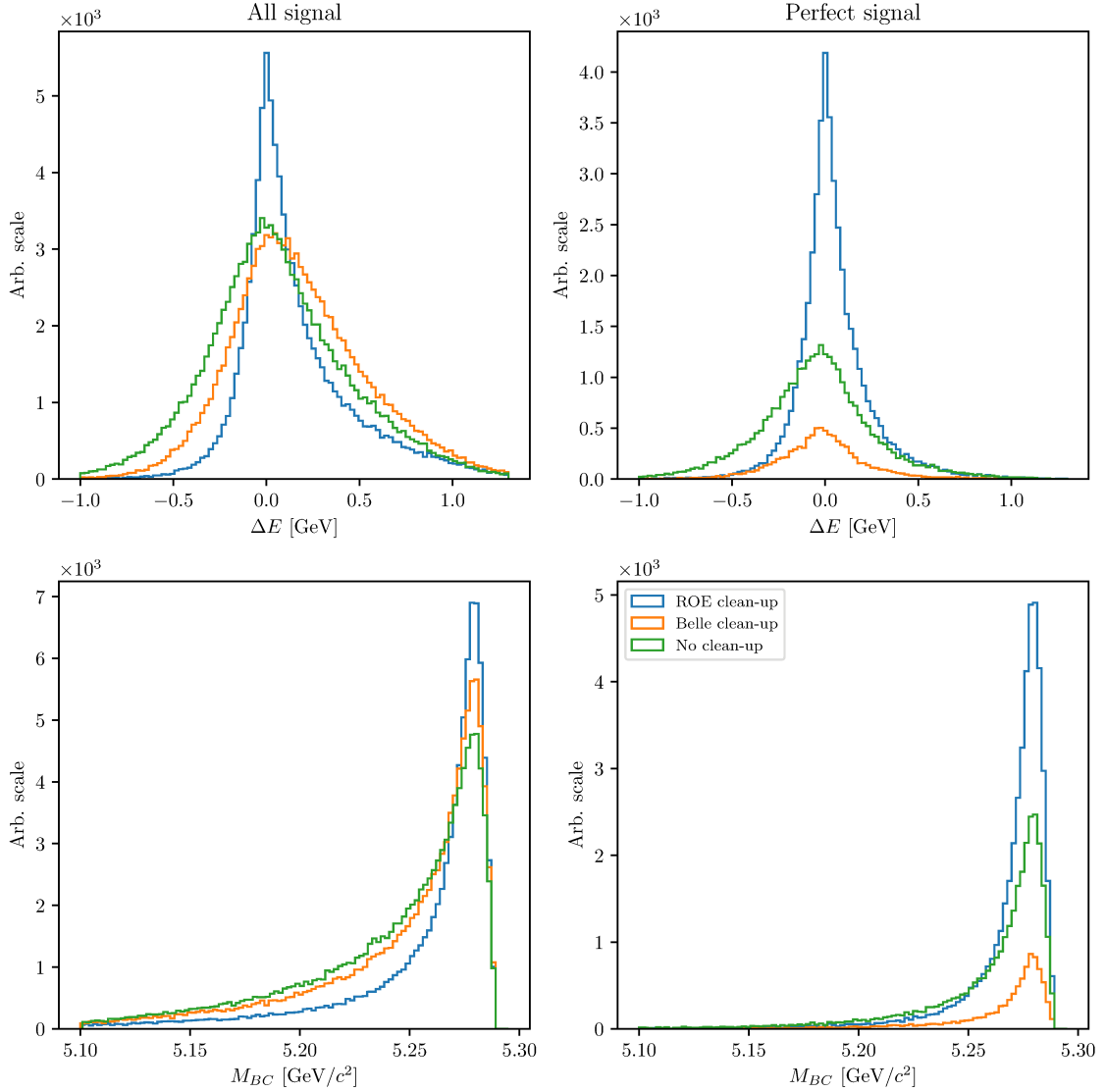


Figure 6.13:  $\Delta E$  and  $M_{BC}$  distributions for various types of clean-up procedures. The figures on the left are shown for the full signal sample after the stated selection criteria, while the figures on the right are shown for the perfectly reconstructed signal candidates. For ROE clean-up, 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) for the perfect signal.

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. (5.18), shown in Figure 6.14 for various clean-up procedures. The figure shows an improved resolution of the distribution of the charged product, 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 6.15 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 for the optimal charge product value. 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.

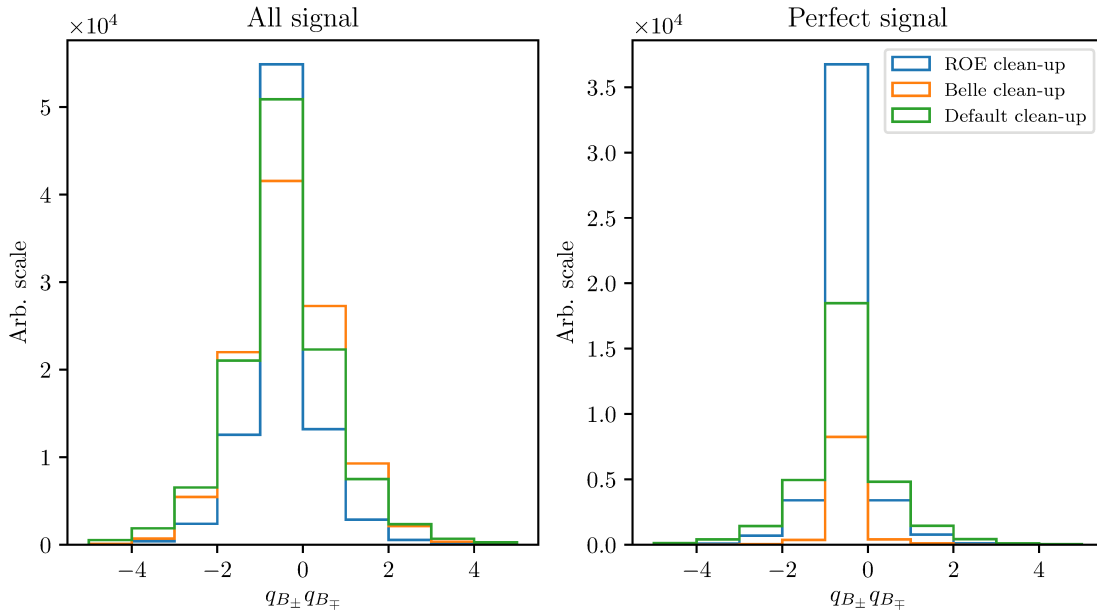


Figure 6.14: Distribution of the charge product of both  $B$  mesons for various types of clean-up procedures, shown on the full signal MC (left) and for the perfectly reconstructed signal candidates (right). For ROE clean-up, the procedure seems to increase the number of perfectly reconstructed candidates.



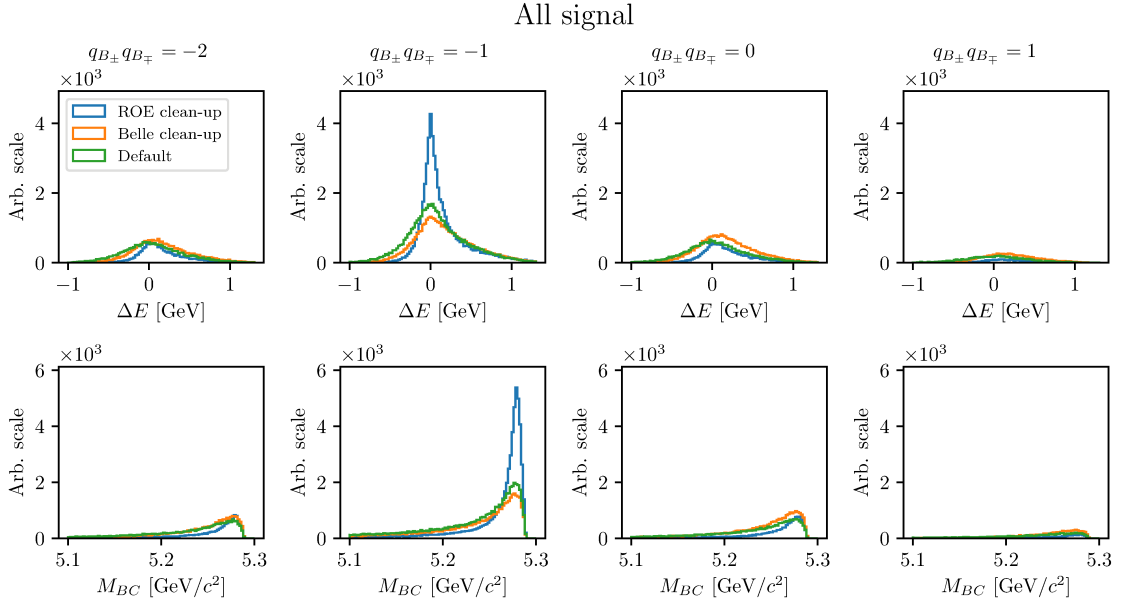


Figure 6.15: Distributions of  $\Delta E$  (top) and  $M_{BC}$  (bottom) for various types of clean-up procedures, split by specific values of the charge product, shown for the full signal MC. There is a significant improvement in resolution after ROE-cleanup, for the case of the correct value of the charge product.

## 6.6 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,

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

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 selection criteria applied, as in the previous section, we also apply a selection to make the control sample more significant. We keep only the candidates passing the following selection cut on the invariant mass of the two kaons

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

as shown in Figure 6.16. Further detail about this cut can be found in Section 7.1.

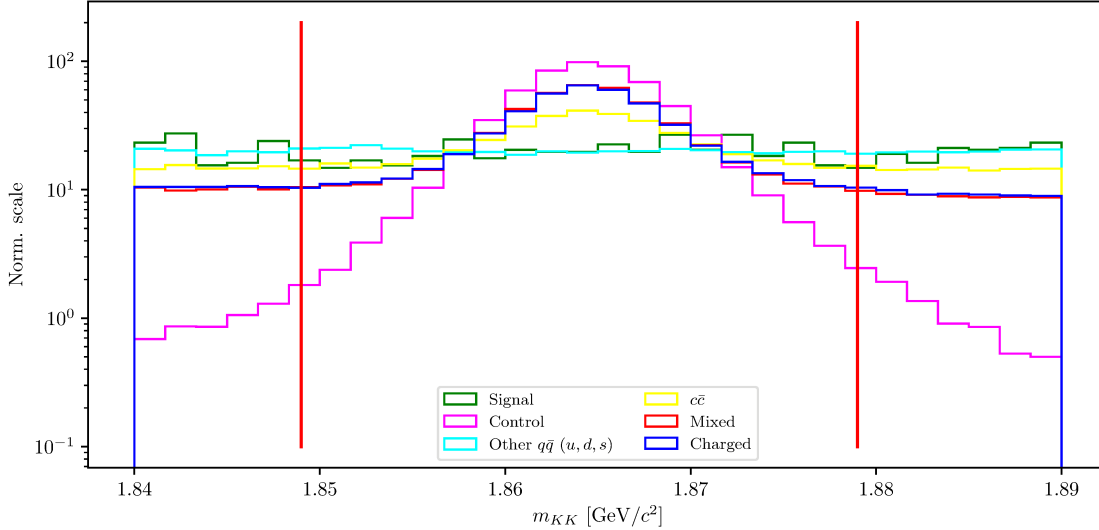


Figure 6.16: Normalized distributions of  $m_{KK}$  for the full MC dataset. The red lines represent the edges of the selection in the  $m_{KK}$  distribution where the control sample is enhanced. The  $m_{KK}$  distribution drops quickly for the case of the control decay, while staying uniform for other contributions.

With the control sample selection determined, we now run the reconstruction with and without the ROE clean-up procedure on all available MC and data.

The effects of the ROE clean-up are shown in Figure 6.17, where we overlay the data points to a stacked histogram of MC contributions for  $\Delta E$  and  $M_{BC}$ . We see that the data and MC agree well. A slight systematic trend can be seen in the  $\Delta E$  variable, which is addressed in Section 8.1.1. 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.

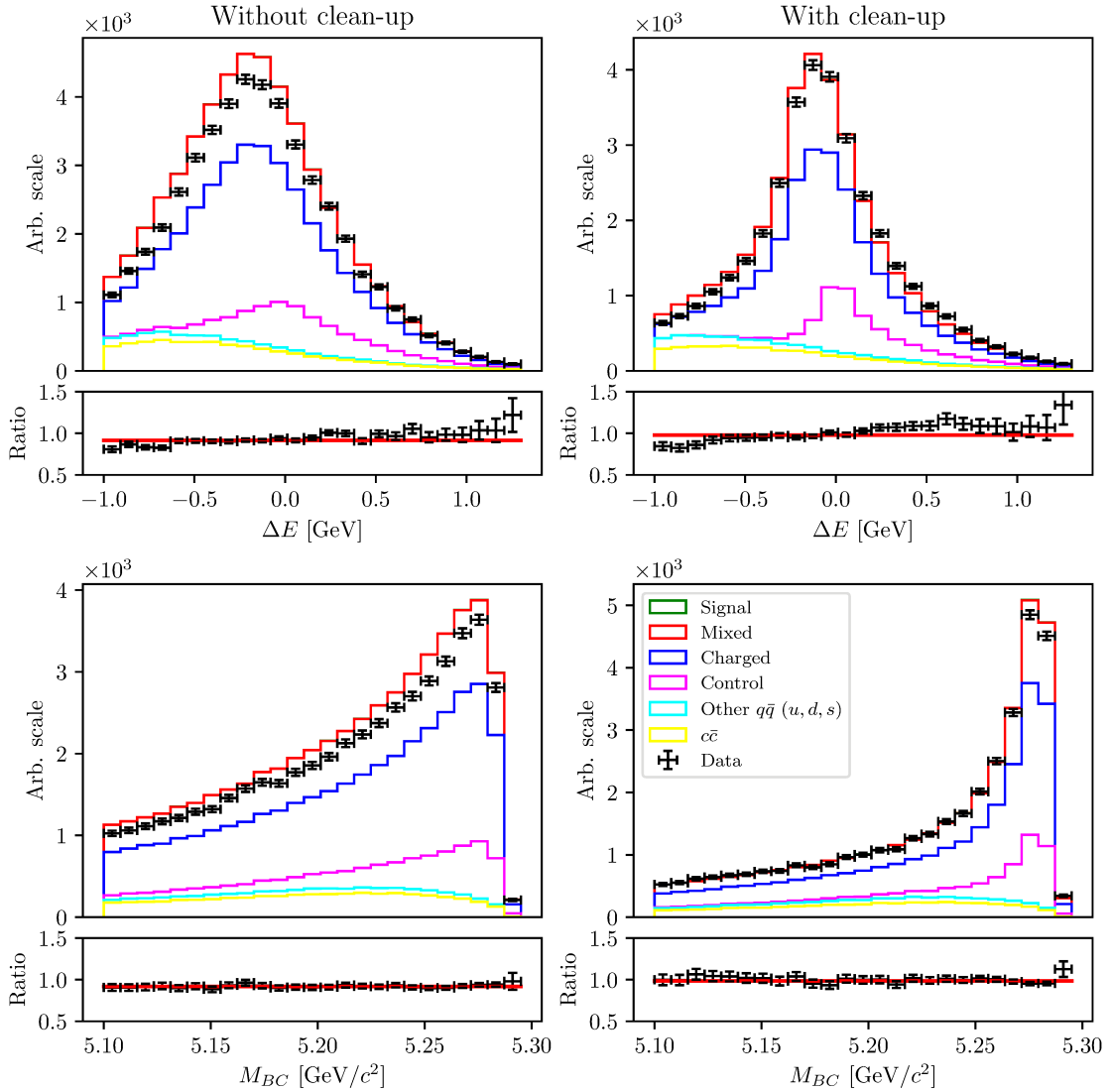


Figure 6.17: Distributions of  $\Delta E$  (top) and  $M_{BC}$  bottom for the case without (left) and with ROE clean-up (right). The resolution of the control sample is improved and the MC and data agree well in all aspects. While the simulated background resolution is also improved, it is worse compared 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 6.18 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.

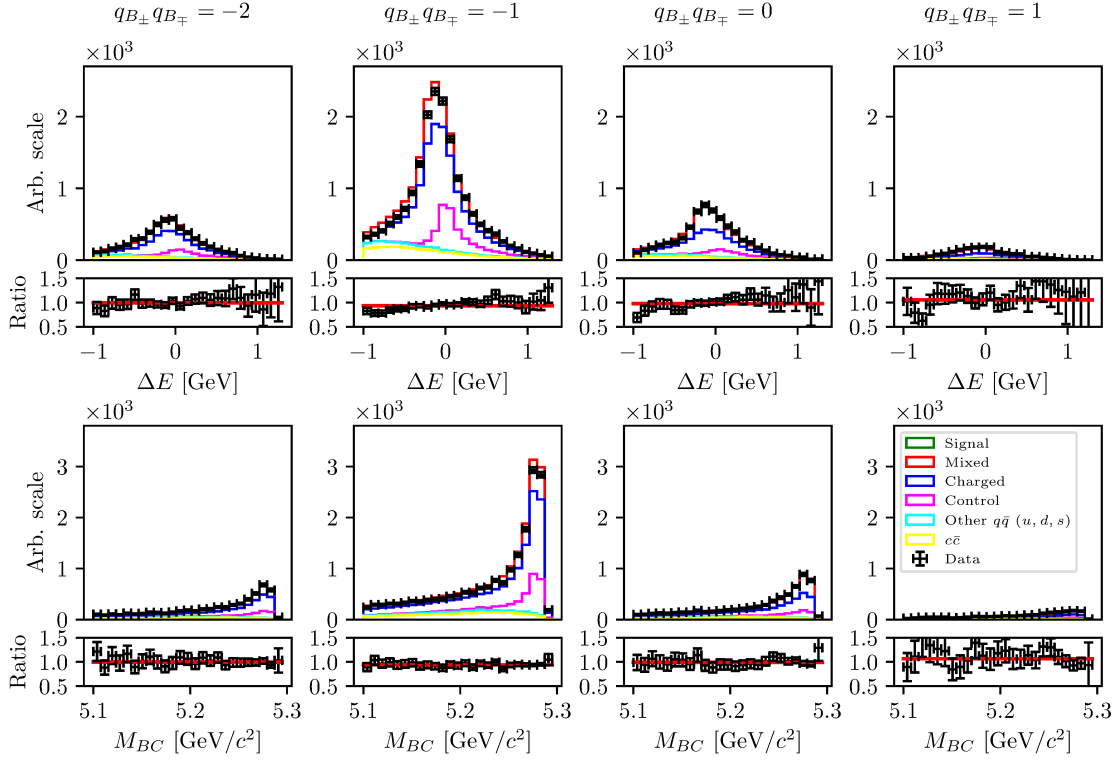


Figure 6.18: Distributions of  $\Delta E$  (top) and  $M_{BC}$  (bottom) split in bins of the charge product of the two  $B$  mesons.

The ROE clean-up procedure seems to perform well. It significantly improves the resolution of the control and, therefore, signal candidates and increases the amount of perfectly reconstructed 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 perform on 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 explains the procedure of suppressing various kinds of backgrounds by using MVA classifiers. More information about the MVA training, feature importance, and hyper-parameter optimization for each MVA step in this chapter can be found in Appendix B.

### 7.1 Resonant Background

In the analysis we study decays with kaons in the final state. This means that standard procedures in  $b \rightarrow u$  analyses, in order to suppress  $b \rightarrow c$  backgrounds, such as  $K$ -veto, are not possible. As a consequence, our final sample consists of combinations of  $K$  pairs coming also from  $b \rightarrow c$  sources, such as  $D^0 \rightarrow K^+K^-$ . 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

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

In order to suppress these resonant backgrounds, while studying signal or control decays, we define two regions

- signal region:  
 $|m_{KK} - m_\phi| > \Delta_\phi, |m_{KK} - m_{D^0}| > \Delta_{D^0}, |m_{K\pi} - m_{D^0}| > \Delta_{D^0},$
- control region:  
 $|m_{KK} - m_{D^0}| \leq \Delta_{D^0}, |m_{K\pi} - m_{D^0}| > \Delta_{D^0},$

where  $m_{KK}$  is the  $KK$  invariant mass and  $m_{K\pi}$  is the invariant mass of  $KK$  candidates, where the kaon with the same charge as the  $B$  meson was given the mass of the charged  $\pi$ .  $m_\phi \approx 1.019 \text{ GeV}/c^2$  and  $m_{D^0} \approx 1.864 \text{ GeV}/c^2$  are nominal masses of the  $\phi$  and  $D^0$  mesons, and  $\Delta_\phi \approx 8 \times 10^{-3} \text{ GeV}/c^2$  and  $\Delta_{D^0} \approx 1.5 \times 10^{-2} \text{ GeV}/c^2$  are the widths around the nominal mass values for the  $\phi$  and  $D^0$  meson, respectively. By selecting the signal or control region, we are able to efficiently isolate the desired subset. Table 7.1 shows the subsample efficiency after selecting either of the regions.

	$\epsilon(\text{Signal cand.})$	$\epsilon(\text{Control cand.})$	$\epsilon(\phi \text{ resonance cand.})$
Signal-specific	95.4%	4.0%	13.6%
Control-specific	1.9%	96.0%	0.0%

Table 7.1: Efficiencies after selecting the signal or control region

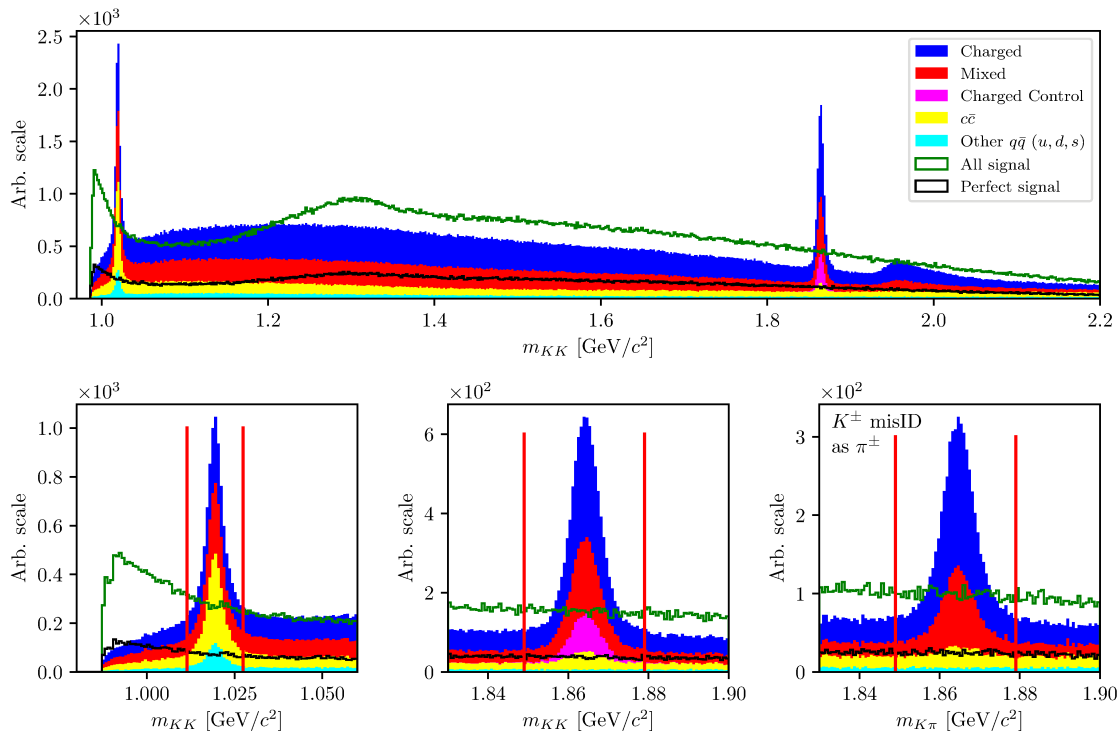


Figure 7.1: Invariant mass of  $KK$  candidates over a wider region (top) and for a specific region around the  $\phi$  peak (bottom left), the  $D^0$  peak (bottom center), and again for the  $D^0$  peak, but with the  $m_{K\pi}$  mass of the  $KK$  candidates (bottom right). Signal (green) and perfect signal (black) are equally scaled up.

## 7.2 Continuum Suppression

Physics processes where continuum states are produced in electron and positron collisions

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

with  $q = u, d, s$  or  $c$ , represent an important category of backgrounds, called the continuum background. In addition to the previously used kinematic constraints, used to separate  $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$  decays from  $e^+e^- \rightarrow q\bar{q}$ , properties of the “event shape” are also used, since phase-space distributions of produced particles differ for these two processes. Continuum background events are produced back-to-back 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 form an isotropic distribution in the detector, which yields a spherical event shape.

## 7.2.1 Characteristic Variables

Information on the phase-space distribution of produced particles can be obtained in a number of different ways. In this section, 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 wish to discriminate.

### Thrust and Related Variables

It is possible to define a thrust axis  $\vec{T}$  for a collection of  $N$  momenta  $p_i$  as a unit vector, along which their total projection is maximal. The thrust scalar  $T$  (or thrust) is a derived quantity, defined as

$$T = \frac{\sum_i |\vec{T} \cdot \vec{p}_i|}{\sum_i |\vec{p}_i|}. \quad (7.1)$$

In this case, a related variable is  $|\cos \theta_T|$ , where  $\theta_T$  is the angle between the thrust axis of the particles from the  $B$  meson candidate 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 isotropically distributed. On the other hand, particles in continuum events follow the direction of the jets in the event. As a consequence, both thrust axes are strongly directional and aligned in the opposite direction, 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 cosine of the angle between the thrust axis of the  $B$  candidate and the beam axis. For  $B$  candidates from  $B\bar{B}$  events this distribution is again uniformly distributed, while for candidates from continuum events this distribution follows the distribution of the initially produced quark pairs,  $1 + \cos^2 \theta_{T,B}$ . In practice, such a distribution exhibits a drop at  $|\cos \theta_{TB}| \approx 1$ , due to the acceptance loss of the detector in the direction of the beam pipes. Figure 7.2 shows the distributions of  $|\cos \theta_T|$  (left) and  $|\cos \theta_{TB}|$  (right) for  $B$  meson candidates from various sources.

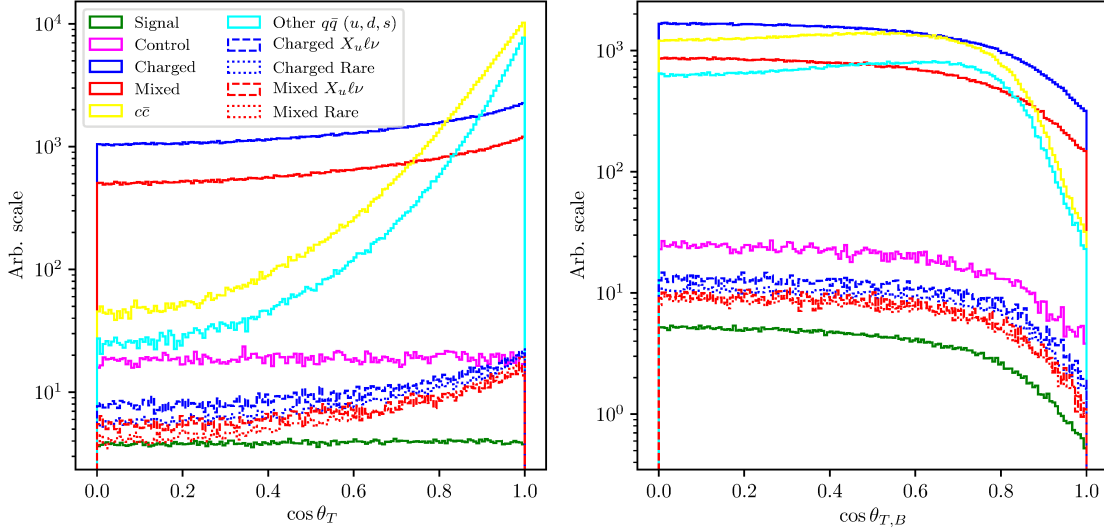


Figure 7.2: Distributions of  $|\cos \theta_T|$  (left) and  $|\cos \theta_{T,B}|$  (right) for  $B$  meson candidates from various sources.

### CLEO Cones

CLEO cones have been introduced by the CLEO collaboration and are an additional specific tool to provide optimal continuum background discrimination. They consist of nine variables corresponding to the momentum flow around the thrust axis of the  $B$  meson candidate, binned in nine cones of  $10^\circ$  around the thrust axis, as illustrated in Figure 7.3. Momentum flow is defined as the scalar sum of the momenta of all charged tracks and neutral particles, pointing in a specific angle interval. A Fisher discriminant is formed from the nine momentum flow variables and from  $|\cos \theta_T|$  for  $B$  candidate and the ROE. The Fisher discriminant,  $\mathcal{F}$ , is the linear combination of the input variables, which maximizes the separation between signal and background. Additional information is provided in [23].



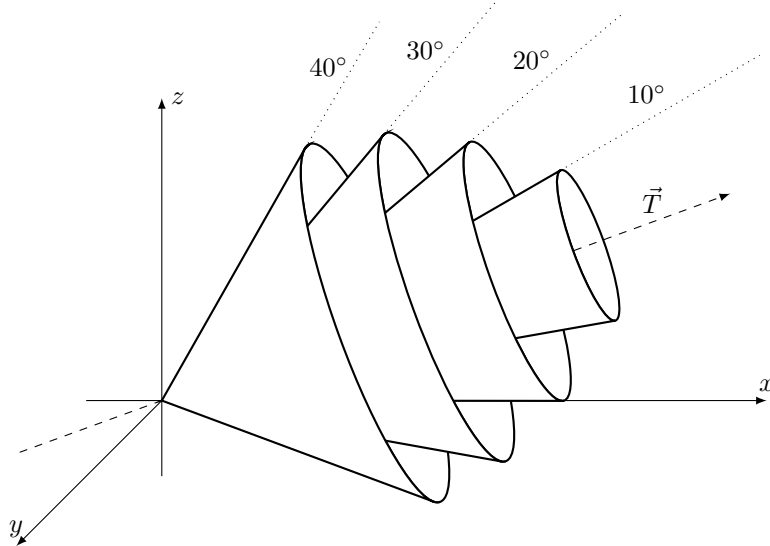


Figure 7.3: Concept of CLEO cones.  $\vec{T}$  denotes the thrust axis of the  $B$  meson candidate in an event. Each variable corresponds to a momentum flow around the thrust axis in steps of  $10^\circ$ .

### KSFW Moments

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

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

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. They are described in detail in [24].

### $B$ -Flavor Tagging

While the goal of  $B$ -flavor tagging is to determine the flavor of a  $B$  meson, the variables used for flavor tagging also potentially contribute to background suppression. Flavor tagging relies on the fact that a large fraction of  $B$  mesons decay to a final state that is flavor specific, and can only be reached either through the decay of a  $b$  or a  $\bar{b}$  quark. Because of the large number of  $B$  meson decay channels, full reconstruction of a sufficiently large number of flavor-specific  $B$  candidates is not feasible. Instead, inclusive techniques are employed, that make use of different flavor-specific signatures of  $B$  decays.

The flavor tagging algorithm proceeds in two stages. In the first stage, individual flavor-specific signatures are analyzed, each of which provides a signature-specific flavor tag. In the second stage, the results from the first stage signatures are combined into a final flavor tag. Both stages heavily rely on MVA methods in order to

optimally combine all available information. The final result of each stage is  $qp$ , a product of the flavor sign,  $q$ , and the probability of a correct flavor tag,  $p$ . Further details can be found in [24].

### ROE fit information

Most  $B$  tag mesons decay via  $b \rightarrow c$  transitions with at least one additional vertex at a distance comparable to the decay length of a  $B$  meson. These vertices introduce a bias in the measurement of the companion  $B$  meson vertex position, which degrades the vertex resolution.

The strategy to select the optimal set of tracks for the vertex determination is to first select a subset of the tracks in the ROE, which satisfy some requirements, such as a minimum number of vertex detector hits and a maximum transverse distance to the interaction region. All tracks, which do not pass the selection criteria, are removed. In the end, all tracks are combined in a single vertex using the interaction region as a constraint. If the goodness of the vertex fit is not good enough, the worst track is removed and the vertex refitted. This procedure is repeated until the fit is satisfactory or no tracks are left. After the vertex of the ROE has been determined,  $\Delta z$  can be calculated as the distance of vertices between the signal  $B$  meson candidate and the ROE in the  $z$  direction. More information is available in [24].

## 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  $q^2$ ,  $M_{BC}$  or  $\Delta E$  correlation. Variables with significant correlation or complex shapes in the correlation distribution were discarded from the training set, since they would have introduced unwanted dependence on the unreliable model, 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  $2 \times 10^5$  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. Such a composition is chosen so that there are enough signal and background samples for the MVA model training, and to avoid any of the background contributions being under-represented. Since the full Belle dataset is experiment-dependent, we construct the training dataset by sampling MC events from appropriate experiments, proportionally to the size of the dataset per each experiment.

The training variable set consisted of

- $B$  meson direction and thrust related variables
  - magnitude of thrust axes of the  $B$  and  $ROE$  candidates,
  - cosine of the angle between the thrust axis of the  $B$  candidate and thrust axis of the  $ROE$  candidate,

- cosine of the angle between the thrust axis of the  $B$  candidate and the beam direction,
- reduced Fox-Wolfram moment  $R_2$ ,
- all 9 CLEO Cones
- KSWF Moments
  - $R_{01}^{so}, R_{02}^{so}, R_{03}^{so}, R_{04}^{so}$ ,
  - $R_{10}^{so}, R_{12}^{so}, R_{14}^{so}$ ,
  - $R_{20}^{so}, R_{22}^{so}, R_{24}^{so}$ ,
  - $R_0^{oo}, R_1^{oo}, R_2^{oo}, R_3^{oo}, R_4^{oo}$ ,
- $B$ -flavor tagging 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
  - $\Delta z$ .

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.

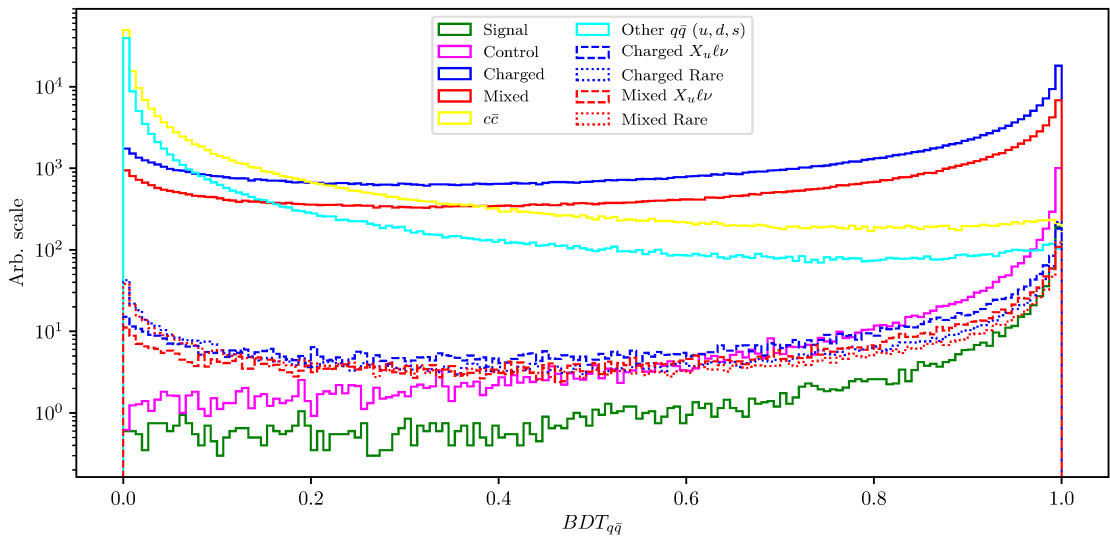


Figure 7.4: Continuum suppression classifier output for signal and various types of background.  $B$  candidates from continuum events dominate the lower region, while candidates from  $B\bar{B}$  dominate in the upper region of the classifier output.

### 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 among the candidates from other  $B\bar{B}$  events.  $B\bar{B}$  events consists of

- $b \rightarrow c\ell\nu$  background,
- $b \rightarrow u\ell\nu$  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 % generically decaying 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  $b \rightarrow u\ell\nu$  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 ( $\Delta E$  and  $M_{BC}$ ) 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 kinematic nature, such as decay particles momenta, decay angles or other similar variables.

The training variable set consisted of

- vertex fit probability of  $P(\chi^2, DOF)$  of the signal  $B$  meson candidate
- vertex fit probability of  $P(\chi^2, DOF)$  of the ROE side,
- $\cos\theta_{BY}$  from Eq. (5.4),
- cos of the angle between the momentum vector and vector joining the IP and the production vertex of the  $KK\ell$  candidate,
- $B$ -flavor tagging variables for the two signal-side kaons,
- numbers of kaons, tracks and distant tracks in ROE,
- $\theta$  angle of the ROE momentum in CMS frame,
- $\xi_Z$  from [25]
- $\Delta z$ ,
- $m_{miss}^2$  from Eq. (5.11),
- $m_{miss}^2$  for partial reconstruction of  $B^0 \rightarrow D^{*-}\ell\nu$ ,

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 is a veto variable where we partially reconstruct the  $D^*$  candidate four-momentum via a linear combination of the  $\pi_s^\pm$  four-momentum in the  $D^* \rightarrow D\pi_s^\pm$  decay. It contributes to discarding the  $B^0 \rightarrow D^{*-}\ell^+\nu$  decay, where  $D^{*-}$  further decays to  $D^{*-} \rightarrow \bar{D}^0\pi_s^-$ . Figure 7.5 shows the veto variable with a partial reconstruction of a charged  $\pi_s^\pm$ .

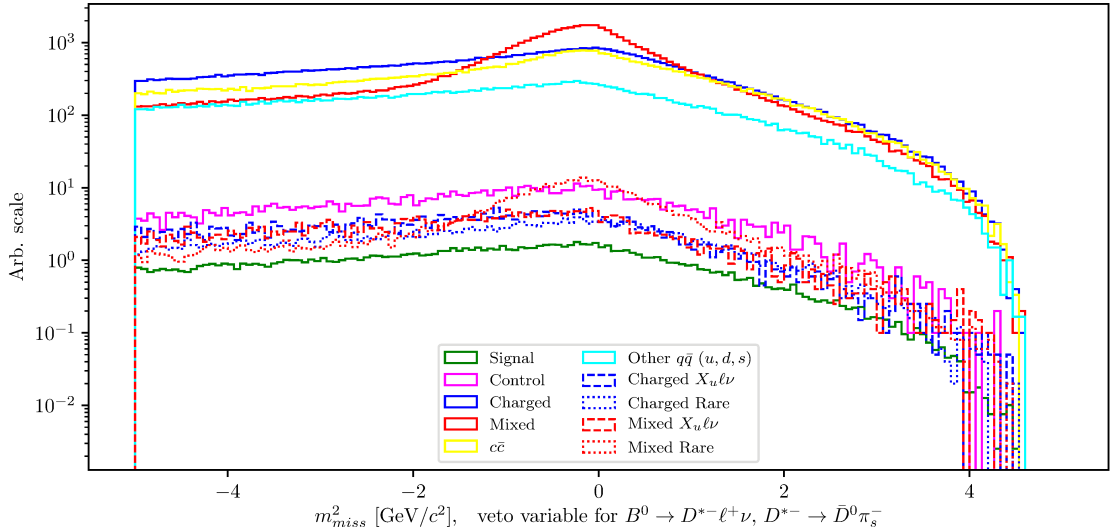


Figure 7.5: Distribution of  $m^2_{miss}$  for partially reconstructed  $B^0 \rightarrow D^{*-}\ell^+\nu$  decays.

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 the one described 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 [26].

The training set used 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  $M_{BC}$  distribution, therefore we train the  $uBDT$  classifier with a uniformity constraint on the  $M_{BC}$  variable of the background candidates. 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 aids in the performance of signal extraction.

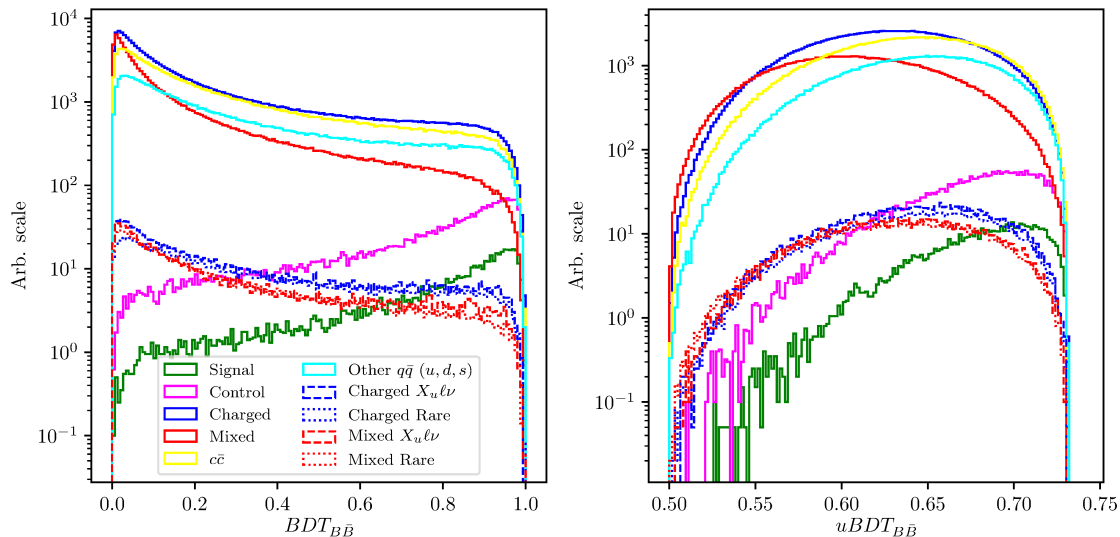


Figure 7.6:  $B\bar{B}$  suppression classifier output for signal and various types of background for the standard  $BDT$  classifier (left) and the  $uBDT$  classifier (right).  $B$  candidates from  $B\bar{B}$  background events dominate the lower region, while signal and control candidates dominate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

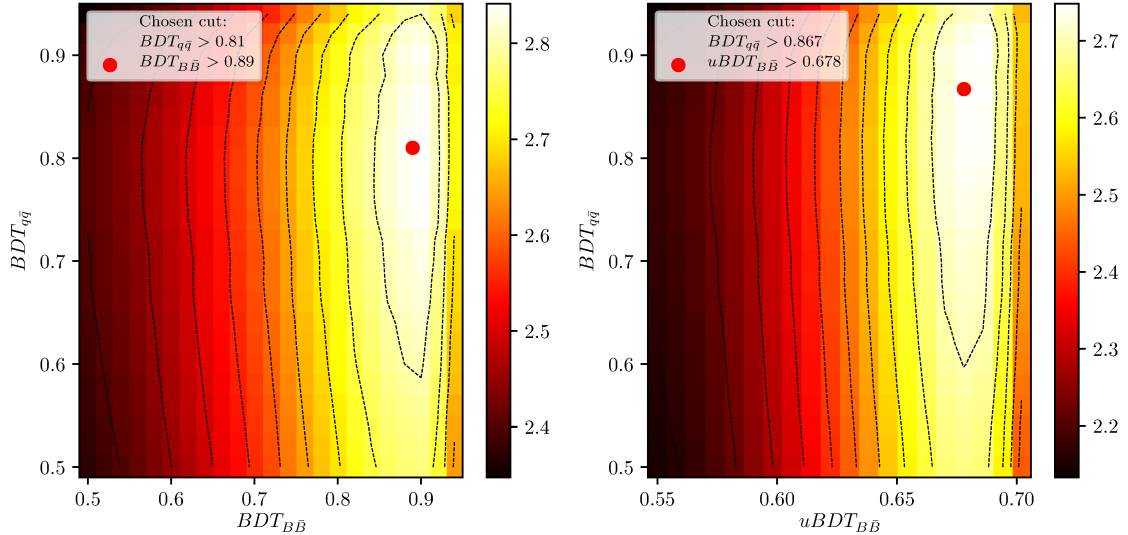


Figure 7.7: 2D *FOM* optimization of continuum suppression classifier and the standard *BDT* (left) and *uBDT* (right)  $B\bar{B}$  suppression classifier.

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 in the extraction process. 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*:  $N_{sig} = 176$ ,  $N_{sig}/N_{bkg} = 4.83 \%$ ,
- *uBDT*:  $N_{sig} = 264$ ,  $N_{sig}/N_{bkg} = 1.33 \%$ .

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.

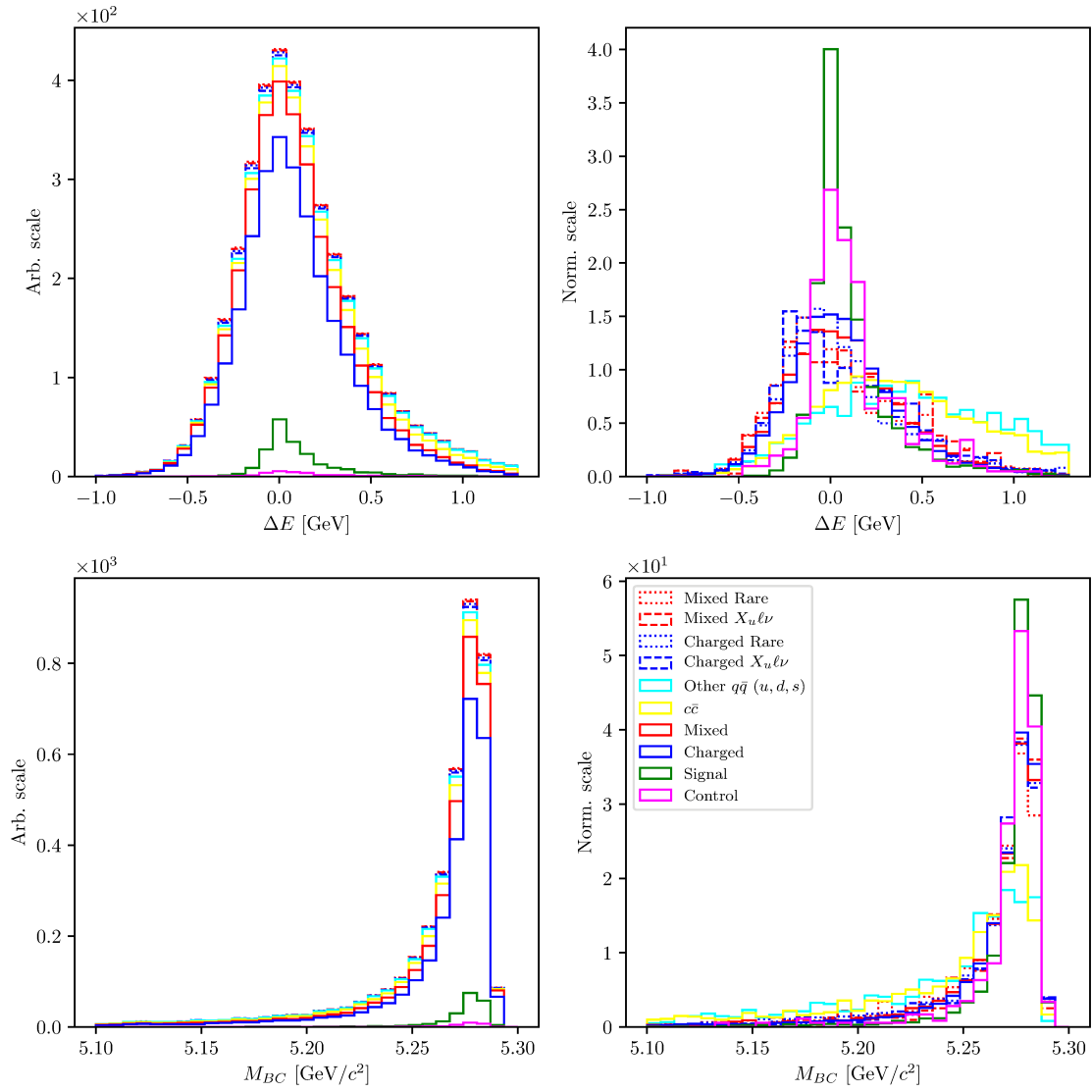


Figure 7.8: Arbitrary (left) and normalized scale (right) for  $\Delta E$  (top) and  $M_{BC}$  (bottom) for the final sample optimized with the standard  $BDT$  classifier.



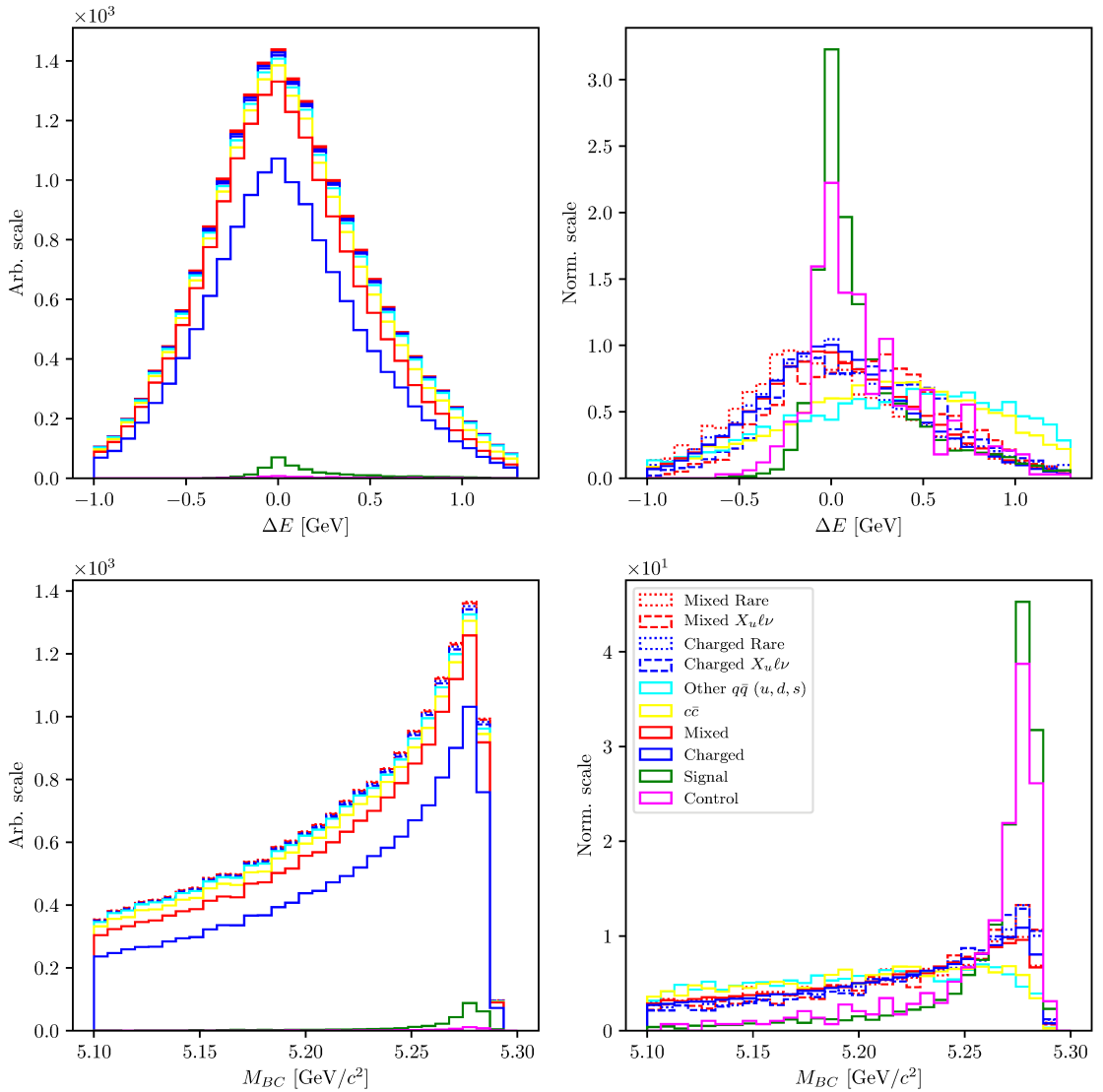


Figure 7.9: Arbitrary (left) and normalized scale (right) for  $\Delta E$  (top) and  $M_{BC}$  (bottom) for the final sample optimized with the  $uBDT$  classifier for  $B\bar{B}$  suppression.

### 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  $\Delta E$ ,  $M_{BC}$ , and  $m_{KK}$  for the most significant contributions in the signal region. While most of the candidates come from events, where all reconstructed charged particles in the signal decay do not come from a single  $B$  meson, but both of them, these candidates are not so problematic. Their distribution is rather smooth and frequent in regions where we expect no signal. On the other hand, there are also contributions from some specific  $B$  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  $B \rightarrow \bar{D}^{(*)}\ell^+\nu$ ,  $D^+ \rightarrow \bar{K}^-\ell^+\nu$ , where the secondary lepton is misidentified as a kaon. Even though the decay has two neutrinos, these events survive the  $m_{miss}^2$  selection and produce peaks at the same positions as the signal distributions, while exhibiting only a slightly worse resolution.

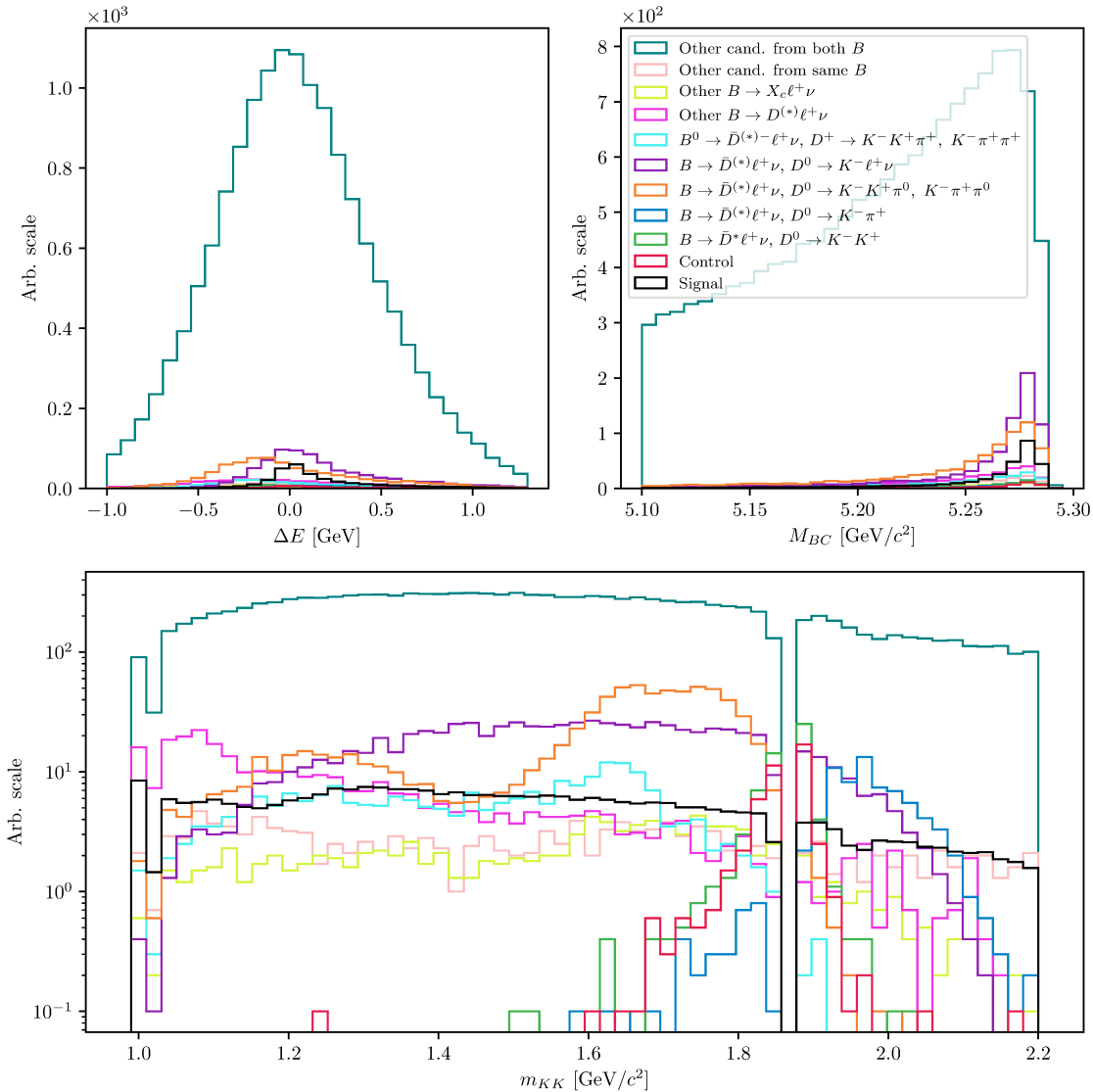


Figure 7.10:  $\Delta E$  (left),  $M_{BC}$  (right) and  $m_{KK}$  (bottom) for major contributions to the  $B\bar{B}$  background in the signal region.  $\Upsilon(4S)$ -matched backgrounds represents the majority, but have a smooth and wide distribution, distinguishable from signal.  $B$ -matched contributions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

In order to suppress the latter source of background, a lepton veto is applied to both kaons, requiring that neither of the kaons should exhibit lepton-like properties. On the candidates, which pass 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. Background in this case is represented by

events in which a lepton has been misidentified as a kaon. 2D *FOM* plots for both kaons are shown in Figure 7.11, where  $K_0$  denotes same-sign, and  $K_1$  the opposite-sign kaon, with respect to the  $B$  meson which they are part of. It can be seen that in the majority of cases an electron is misidentified as  $K_1$ . With the optimal selection

- $K_0$  :  $eID < 0.8$ ,
- $K_1$  :  $eID < 0.1$ ,  $\mu ID < 0.8$ ,

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

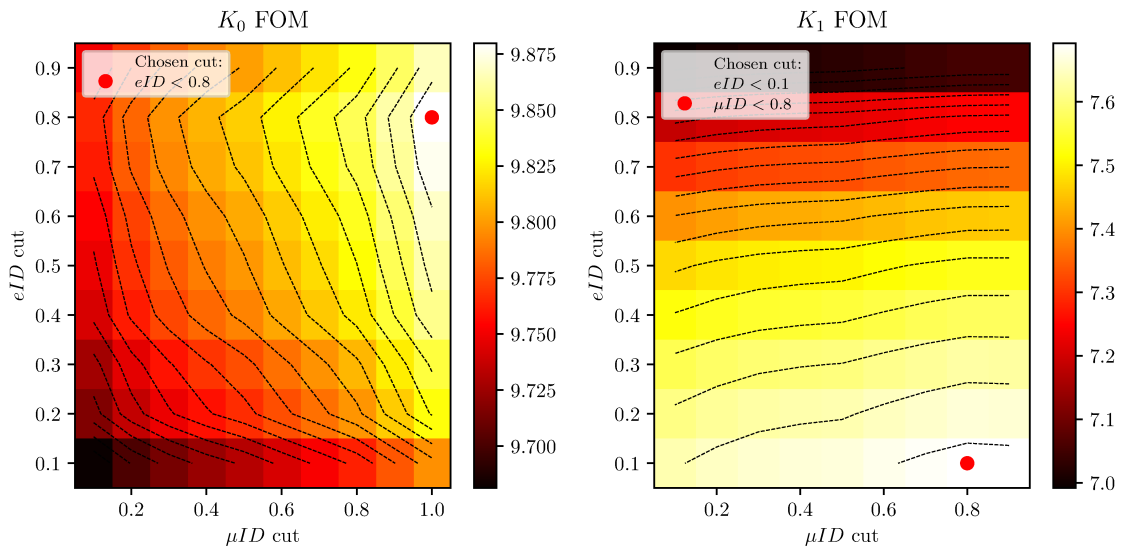


Figure 7.11: 2D *FOM* for optimal  $eID$  and  $\mu ID$  selection on same-sign (left) and opposite-sign (right) kaons with respect to the  $B$  meson charge. For double semileptonic background component, in most cases an electron is misidentified as the opposite-sign kaon in the reconstruction chain.

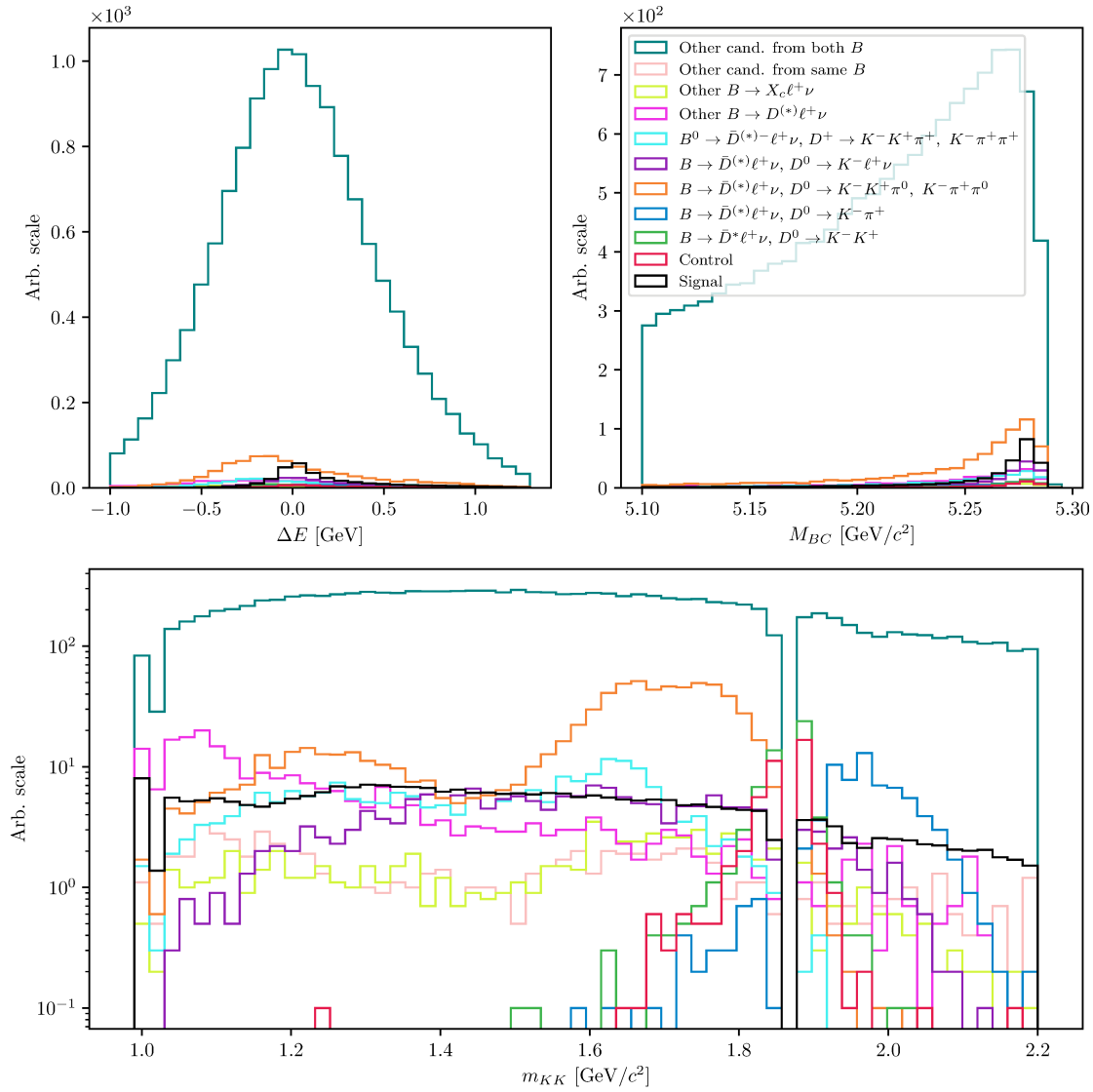


Figure 7.12:  $\Delta E$  (left),  $M_{BC}$  (right) and  $m_{KK}$  (bottom) for major contributions to the  $B\bar{B}$  background in the signal region after the lepton veto. The double semileptonic background component is suppressed by a factor of 4 – 5.

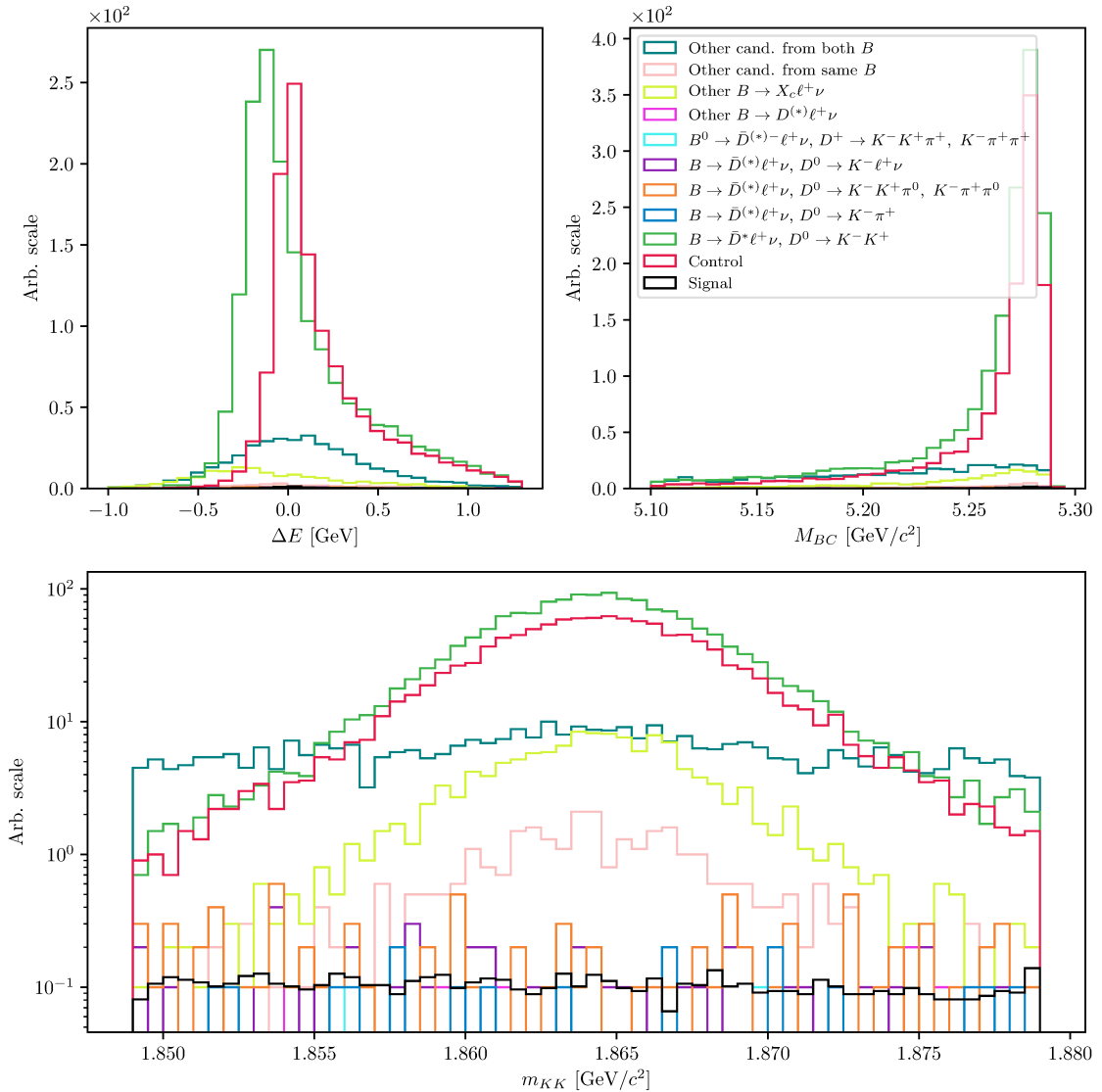


Figure 7.13:  $\Delta E$  (left),  $M_{BC}$  (right) and  $m_{KK}$  (bottom) for major contributions to the  $B\bar{B}$  background in the control region after the lepton veto. The major component in this case are other  $B \rightarrow D^* \ell + \nu$ ,  $D \rightarrow K^+ K^-$  decays, besides the control 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, consequently, 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 continuum 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 stacked MC in the control region, before the MVA selection, where the MC sample was scaled down by a factor of 6, due to having 6 streams of MC. These figures do not show a fit to the data, but merely an overlay of the data and stacked MC distributions. They show good data and MC agreement for the off-resonance sample already before the fit. More importantly than the normalization, the shape of data and MC also seems to match, so further corrections of  $\Delta E$  and  $M_{BC}$  on MC are not necessary. This is also demonstrated by the flatness of the ratio of  $\Delta E$  and  $M_{BC}$  distributions, shown in the same Figure. There seems to be a difference in the classifier performance for the continuum background suppression in data and on MC in the lower region of  $BDT_{q\bar{q}}$ , where classifier efficiency on MC is overestimated. However, after the selection on classifier output, these differences are negligible, since a relatively small amount of continuum background passes the selection.

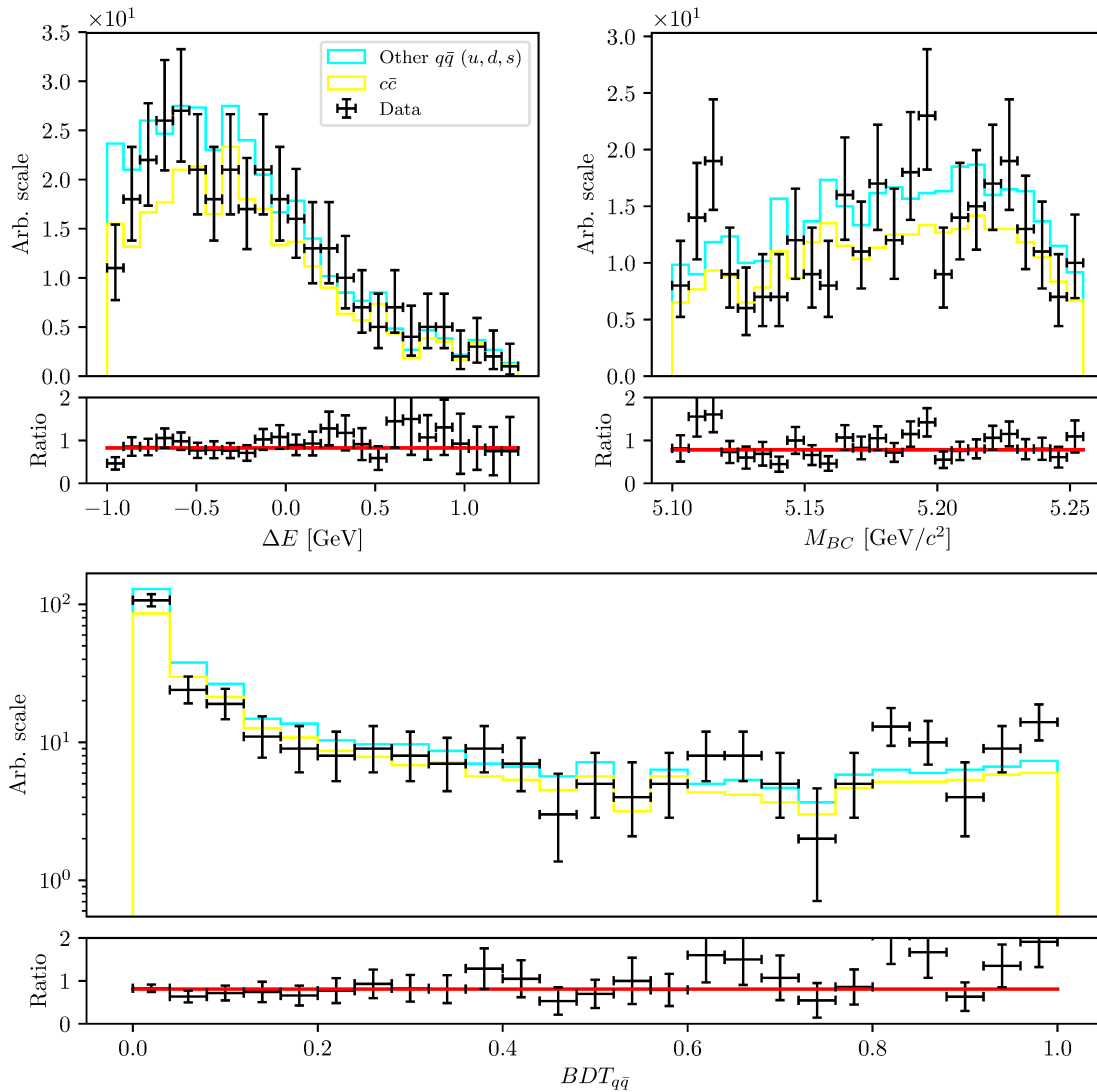


Figure 7.14:  $\Delta E$  (left),  $M_{BC}$  (right), and the  $q\bar{q}$  classifier output (bottom) for off-resonance data and MC in the control region prior to any MVA selection.

## 7.5.2 On-resonance Data

We repeat the check on on-resonance data. Figure 7.15 shows  $\Delta E$ ,  $M_{BC}$ ,  $BDT_{q\bar{q}}$ , and  $uBDT_{B\bar{B}}$ , where one can see inconsistencies between data and stacked MC on the lower parts of all  $BDT$  spectra. These figures again do not show the fit, but merely an overlay of the data to the stacked MC distributions. On the other hand, the data and MC seem to agree well in the upper parts of the spectra. Overall, data and MC seem to agree well after the pre-selection and without any corrections. This means that the modeling of the MC sample is precise and that the MVA selection does not introduce any additional differences between the data and MC for the control and, therefore, the signal sample.

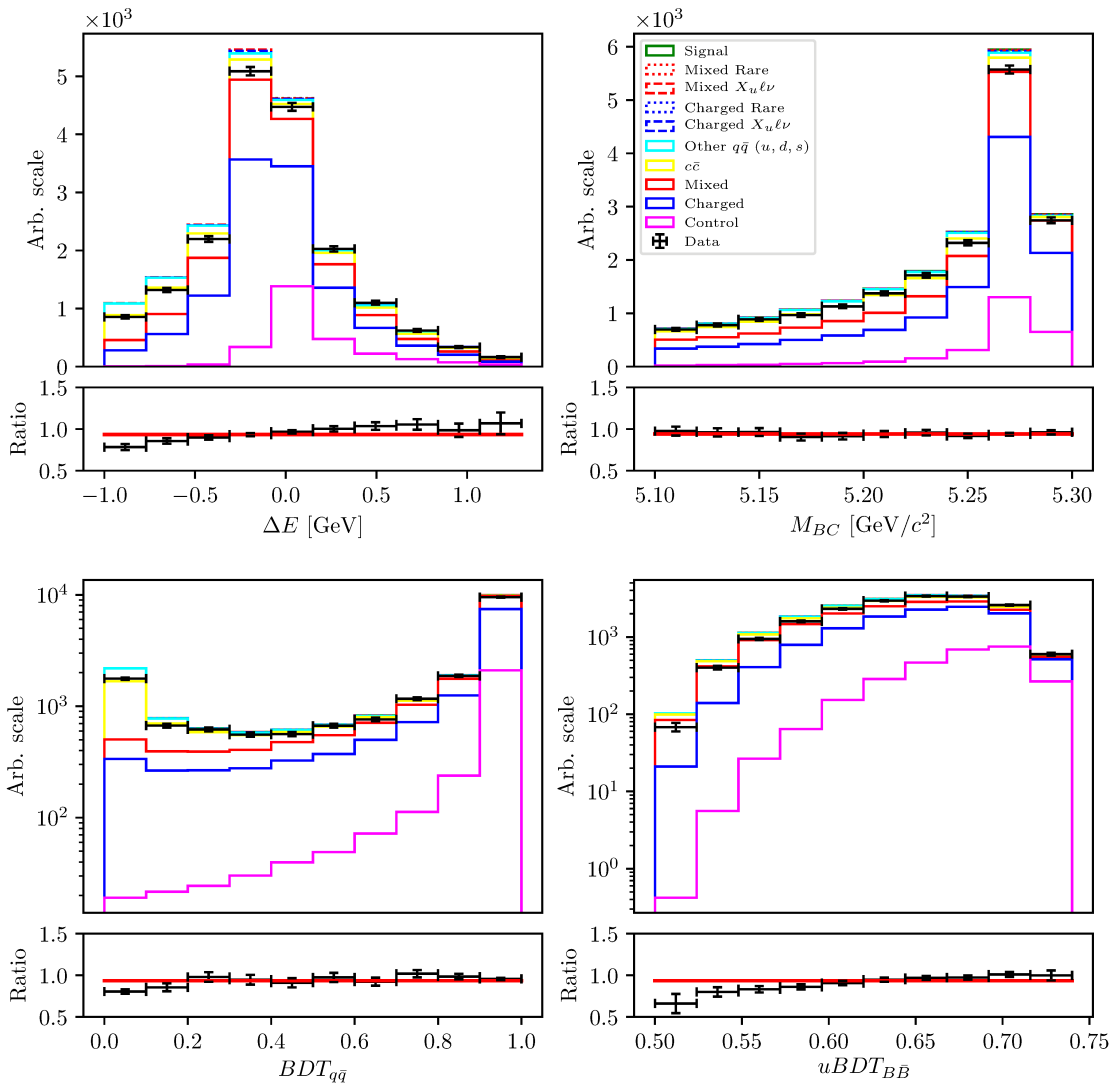


Figure 7.15:  $\Delta E$  (top left),  $M_{BC}$  (top right), the  $q\bar{q}$ - (bottom left) and the  $B\bar{B}$  classifier output (bottom right), for on-resonance data and MC in the control region prior to the MVA selection.





# Chapter 8

## Extraction of Physical Parameters

In this chapter, the procedure for signal yield extraction is presented. We use the framework of `Roofit` [27], where we define 2D histogram templates in  $\Delta E$  and  $M_{BC}$ , for signal and several types of background, based on MC. 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 Poissonian 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

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

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

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

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 [28], 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 the 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

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

which asymptotically behaves as the  $\chi_q^2$  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

$$\text{Significance} = \sqrt{\lambda} = \sqrt{\chi^2}. \quad (8.4)$$

## 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 the fitted data. When fitting the measured 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 the data fit. 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  $20\times$  integrated luminosity of the full Belle dataset,
- `rare` sample, corresponding to  $50\times$  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 selection of the  $m_{KK}$  around the  $D^0$  mass window, most of the decays with a  $D^0$  proceed via  $D^0 \rightarrow K^+K^-$  decay. In this case, the following fit templates are chosen

- signal template,

- $q\bar{q}$  template,
- $C_0$  :  $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$ ,  $D^0 \rightarrow K^- K^+$  (control decay),
- $C_1$  :  $B \rightarrow \bar{D}^* \ell^+ \nu$ ,  $D^0 \rightarrow K^- K^+$ ,
- 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  $q\bar{q}$  template in cases after the  $BDT_{q\bar{q}}$  cut, and the signal template in all cases, since they both have expected yields close to zero and are instead fixed to the expected MC values. 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. In case of fits to the data, we constrain the ratio to the measured value in the form of a Gaussian function with the width corresponding to the measurement uncertainty. These constraints are implemented in the fit, so the systematic effects are taken into account in the fit error. The ratio is implemented based on the decay channels shown in Table 8.1 and is defined as

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

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

$$\rho = \frac{\mathcal{B}^{PDG}}{\mathcal{B}^{GEN}}, \quad (8.6)$$

where  $\mathcal{B}^{PDG}$  is the measured branching fraction and  $\mathcal{B}^{GEN}$  is the branching fraction value used in MC generation. The branching fraction correction factor has been implemented due to differences between measured and MC branching fraction values. Each branching fraction measurement serves as a constraint used in the fit. All branching fraction 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 fractions were calculated as

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

where  $\tau_{B^+/B^0}$  is the ratio of  $B$ -meson decay times, which is measured to be [8]

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

Category	Channel	$B$ Decay Mode	$D$ Decay Mode	$N_{MC}$
$C_0$	$N_{00}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	$1184 \pm 34$
$C_1$	$N_{10}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	$1458 \pm 38$
	$N_{11}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	$186 \pm 16$

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

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

ID	Decay	$\mathcal{B}_{GEN}$	$\mathcal{B}_{PDG}$	$\rho$	Ref.
0	$B^0 \rightarrow D^- \ell^+ \nu$	$2.13 \times 10^{-2}$	$(2.13 \pm 0.09) \times 10^{-2}$	$1.00 \pm 0.04$	[8]
1	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$5.33 \times 10^{-2}$	$(4.88 \pm 0.11) \times 10^{-2}$	$0.92 \pm 0.02$	[calc.]
2	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$2.31 \times 10^{-2}$	$(2.29 \pm 0.10) \times 10^{-2}$	$0.99 \pm 0.04$	
3	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$5.79 \times 10^{-2}$	$(5.25 \pm 0.12) \times 10^{-2}$	$0.91 \pm 0.02$	[29]
4	$D^0 \rightarrow K^- K^+$	$3.90 \times 10^{-3}$	$(3.97 \pm 0.07) \times 10^{-3}$	$1.02 \pm 0.02$	

Table 8.2: MC and measured values of branching fractions 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 introduce a modification of the  $\Delta E$  variable by applying a smearing factor and an offset. These are applied by simple transformations of

$$f_{\text{offset}} : x \mapsto x + a, \quad (8.9)$$

$$f_{\text{smearing}} : x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (8.10)$$

where  $x$  goes over all entries in the  $\Delta E$  distribution,  $a$  is the energy offset and  $f_{\text{smearing}}$  corresponds to the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . Since this operation is applied on the candidate level, the smearing introduces some randomness to the fit procedure, which is the reason why we repeat the fit procedure several times in order to reliably extract the parameters of interest. In the 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 in order to determine the best parameter values

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

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

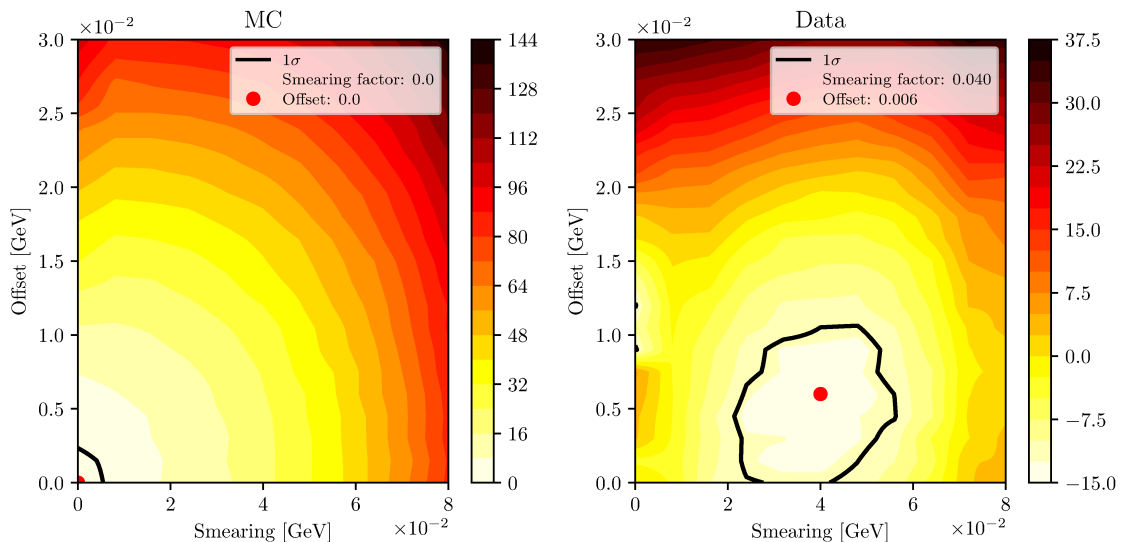


Figure 8.1: Likelihood ratio test of an additional smearing and offset parameter to MC (left) and data (right).

### 8.1.2 Signal Fit

The motivation for the choice of the 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:
  - $C_0$  :  $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$ ,  $D^0 \rightarrow K^- K^+$  (control decay),
  - $C_1$  :  $B \rightarrow \bar{D}^* \ell^+ \nu$ ,  $D^0 \rightarrow K^- K^+$ ,
  - $C_2$  :  $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$ ,  $D^0 \rightarrow K^- \pi^+$ ,
  - $C_3$  :  $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$ ,  $D^0 \rightarrow K^- K^+ \pi^0$ ,  $K^- \pi^+ \pi^0$ ,
  - $C_4$  :  $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$ ,  $D^0 \rightarrow K^- \ell^+ \nu$ ,

- $C_5$  :  $B^0 \rightarrow D^{(*)-} \ell^+ \nu$ ,  $D^+ \rightarrow K^- K^+ \pi^+$ ,  $K^- \pi^+ \pi^+$ ,
- $C_6$  : other  $B \rightarrow \bar{D}^{(*)} \ell^+ \nu$  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 the data fits. Similarly, as described in the control fits, these constraints are implemented in the fit process. The remaining  $B\bar{B}$  background is merged into a single template. The shape of all templates is fixed, 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

$$Y_i = \eta_{\text{norm.}} \times \frac{\left( \sum_j N_{ij} \times \rho_{ij} \right)}{\rho_{00}}, \quad (8.11)$$

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

$$\eta_{\text{norm.}} = \frac{N_{\text{control}}^D}{N_{\text{control}}^{MC}}, \quad (8.12)$$

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

In addition to the branching fraction constraints in Table 8.2, further constraints are defined in Table 8.4. In case of the category  $C_6$ , we have no firm handle on the  $D$  meson decay, therefore no correction for this branching fraction can be introduced, so we set a correction factor of 1 with a 100% error for the  $D$  meson decay branching fraction. 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
$C_0$	$N_{00}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	$44 \pm 7$
$C_1$	$N_{10}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	$53 \pm 7$
	$N_{11}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- K^+$	$6 \pm 2$
$C_2$	$N_{20}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+$	$23 \pm 5$
	$N_{21}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+$	$41 \pm 6$
	$N_{22}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+$	$6 \pm 2$
$C_3$	$N_{30}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- K^+ \pi^0$	$102 \pm 10$
	$N_{31}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+ \pi^0$	$210 \pm 15$
	$N_{32}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- K^+ \pi^0$	$135 \pm 12$
	$N_{33}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+ \pi^0$	$266 \pm 16$
	$N_{34}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- K^+ \pi^0$	$19 \pm 4$
	$N_{35}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- \pi^+ \pi^0$	$35 \pm 6$
$C_4$	$N_{40}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- e^+ \nu$	$47 \pm 7$
	$N_{41}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	$D^0 \rightarrow K^- \mu^+ \nu$	$7 \pm 3$
	$N_{42}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- e^+ \nu$	$98 \pm 10$
	$N_{43}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$	$D^0 \rightarrow K^- \mu^+ \nu$	$10 \pm 3$
	$N_{44}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- e^+ \nu$	$14 \pm 4$
	$N_{45}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^0 \rightarrow K^- \mu^+ \nu$	$3 \pm 2$
$C_5$	$N_{50}$	$B^0 \rightarrow D^- \ell^+ \nu$	$D^+ \rightarrow K^- K^+ \pi^+$	$102 \pm 10$
	$N_{51}$	$B^0 \rightarrow D^- \ell^+ \nu$	$D^+ \rightarrow K^- \pi^+ \pi^+$	$63 \pm 8$
	$N_{52}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^+ \rightarrow K^- K^+ \pi^+$	$31 \pm 6$
	$N_{53}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$	$D^+ \rightarrow K^- \pi^+ \pi^+$	$21 \pm 5$
$C_6$	$N_{60}$	$B^+ \rightarrow \bar{D}^0 \ell^+ \nu$	Other $D^0$ and $D^+$ decays	$69 \pm 8$
	$N_{61}$	$B^+ \rightarrow \bar{D}^{*0} \ell^+ \nu$		$94 \pm 10$
	$N_{62}$	$B^0 \rightarrow D^- \ell^+ \nu$		$63 \pm 8$
	$N_{63}$	$B^0 \rightarrow D^{*-} \ell^+ \nu$		$35 \pm 6$

Table 8.3: Well defined decay channels used for constraining the signal fits.

ID	Decay	$\mathcal{B}_{GEN}$	$\mathcal{B}_{PDG}$	$\rho$	Ref.
5	$D^0 \rightarrow K^- \pi^+$	$3.82 \times 10^{-2}$	$(3.89 \pm 0.04) \times 10^{-2}$	$1.02 \pm 0.01$	[29]
6	$D^0 \rightarrow K^- K^+ \pi^0$	$2.36 \times 10^{-3}$	$(3.37 \pm 0.15) \times 10^{-3}$	$1.43 \pm 0.06$	
7	$D^0 \rightarrow K^- \pi^+ \pi^0$	$13.08 \times 10^{-2}$	$(14.2 \pm 0.5) \times 10^{-2}$	$1.09 \pm 0.04$	
8	$D^0 \rightarrow K^- e^+ \nu$	$3.41 \times 10^{-2}$	$(3.53 \pm 0.028) \times 10^{-2}$	$1.04 \pm 0.01$	
9	$D^0 \rightarrow K^- \mu^+ \nu$	$3.41 \times 10^{-2}$	$(3.31 \pm 0.13) \times 10^{-2}$	$0.97 \pm 0.04$	
10	$D^+ \rightarrow K^- K^+ \pi^+$	$9.06 \times 10^{-3}$	$(9.51 \pm 0.34) \times 10^{-3}$	$1.05 \pm 0.04$	
11	$D^+ \rightarrow K^- \pi^+ \pi^+$	$9.51 \times 10^{-2}$	$(8.98 \pm 0.28) \times 10^{-2}$	$0.94 \pm 0.03$	

Table 8.4: Additional MC and measured values of  $D$  meson branching fractions 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 the 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 compromise solution 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 does the following

1. define uniform binning in both dimensions,
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 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.

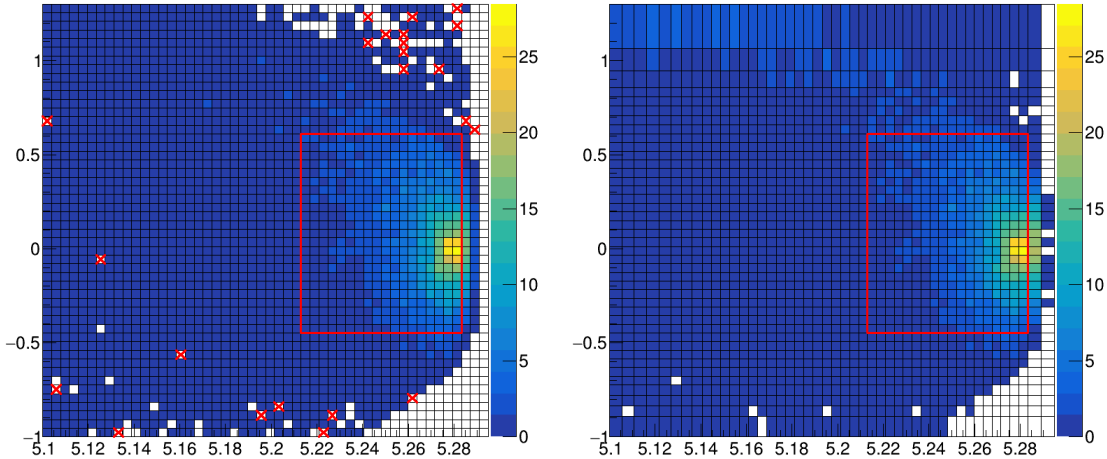


Figure 8.2: Steps taken in the adaptive binning algorithm. Left image shows the initial 2D histogram with the defined optimal region and the problematic bins, the right image shows the final binning with the unchanged optimal region, while the problematic bins are gone due to the new binning choice.

An additional problem occurs in the plotting of the fitted templates 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 the templates, defined with uniform binning.

## 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 we study the yields, errors and the pulls of the signal fit by generating our own pseudo-datasets, according to the MC. This significantly reduces the time we would need to produce the datasets in the standard way, while still reliably describing the underlying physics behind the pseudo-dataset. All available MC was used for pseudo-dataset generation, as well as creating the templates.

### 8.3.1 Pseudo-Experiment: Expected Signal Yield

We constructed  $3 \times 10^3$  pseudo-datasets, where each dataset 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  $M_{BC}$ . The optimal choice of binning is based on the binning test in Section 9.1.

Figure 8.3 shows distributions of the fit yields, errors and the pull distribution of all pseudo-fits. The pull distribution  $g$  of a normally distributed variable  $x$  with a mean  $\mu$  and a width  $\sigma$  is defined as

$$g = \frac{x - \mu}{\sigma}, \quad (8.13)$$

and is also distributed normally, with mean zero and unit width. The fits in Figure 8.3 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 a 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 [30]

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{N}}, \quad \sigma_S = \frac{S}{\sqrt{2(N-1)}}, \quad (8.14)$$

where  $N$  is the number of performed measurements. A decision has been made to put this yield difference as one of the contributions to the overall systematic uncertainty.

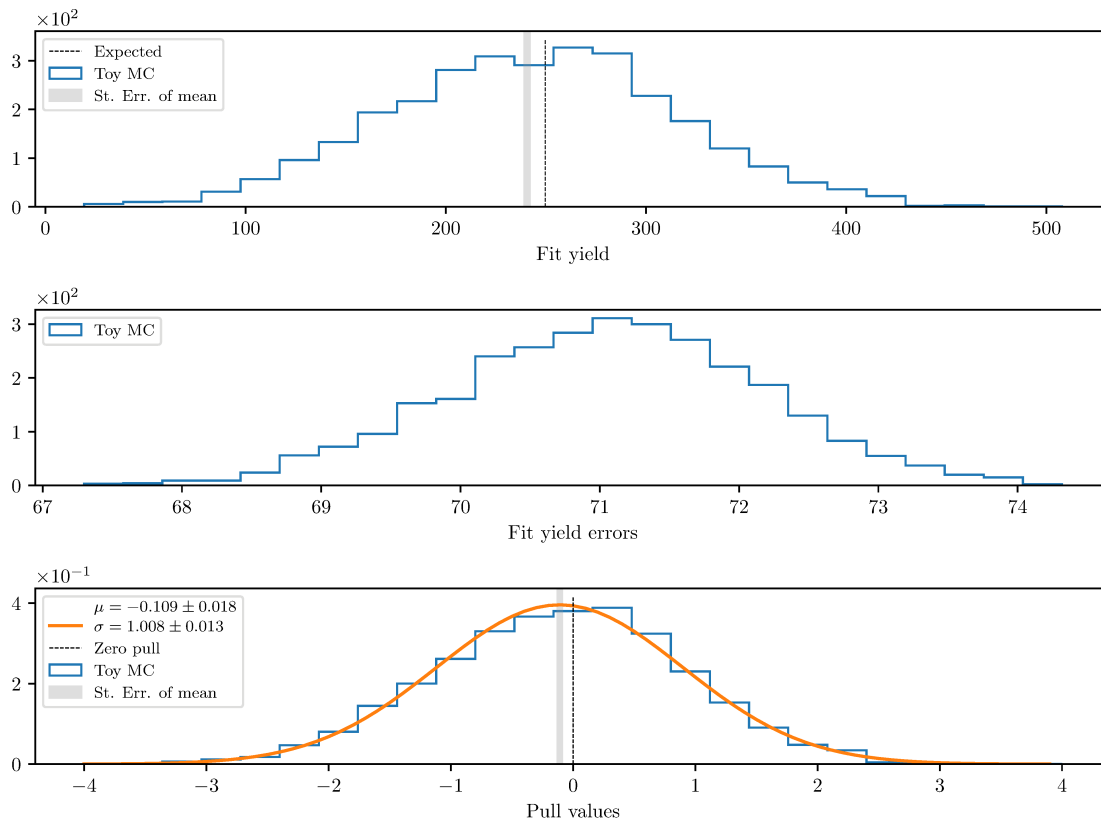


Figure 8.3: Toy MC fits of pseudo-data showing the fit yield (top), the fit errors (center) and the pull distribution of the fits (bottom).

### 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-datasets are generated in the same way as in the previous section, 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, where, for each fraction, we generate 500 pseudo-datasets according to Poisson statistics.

Figure 8.4 shows the difference between the mean fit yield and the expected yield, the 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 a small bias with respect to the expected signal yield, while the pulls seem to be described by the normal distributions throughout the fraction range with a slight offset, which will be incorporated as a contribution to the overall systematic uncertainty. At expected value, we are at about  $3.6\sigma$  statistical significance.

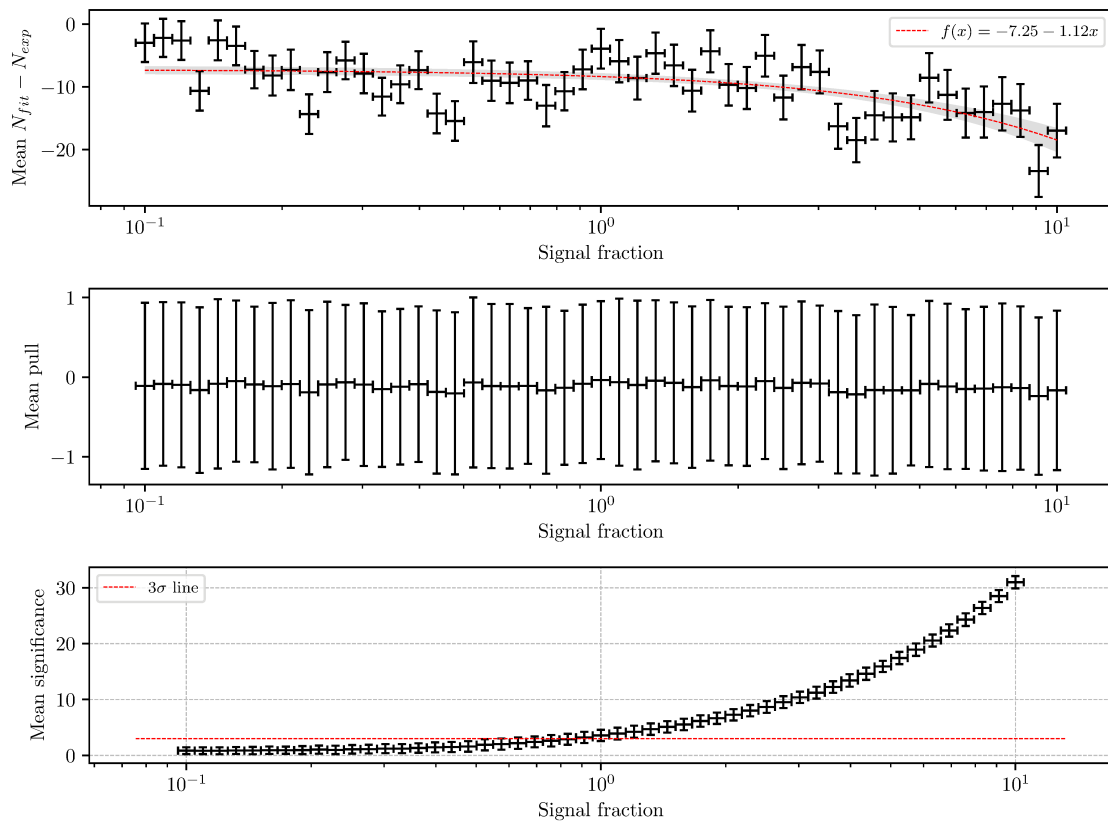


Figure 8.4: The mean fit yield and expected yield difference (top), the mean pull (center) and the mean significance (bottom) as a function of the signal fraction.



# Chapter 9

## Fit Results

In this chapter, we present the results of signal and control fits on MC, as well as in data, along with the control decay branching fraction measurement.

### 9.1 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  $B\bar{B}$  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 binning choice in the range  $N \times N$ ,  $N \in [4,30]$ . Figure 9.1 shows the fitted and the true MC matched yield differences, the pulls, and fit significances for both final samples for each binning case. The difference between fitted and 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 of 0, and a width of 1. From the top plot in Figure 9.1, we see that the yield difference is consistent with 0, while exhibiting a slight, but a consistent overestimation of the fitted yield. This difference is accounted for as a separate contribution to the overall systematic uncertainty. The pull distribution seems to be closer to the normal distribution for the case of the  $uBDT$  classifier. Also the fit significance is larger by approximately  $1\sigma$ , so we conclude that the  $uBDT$  classifier outperforms the standard  $BDT$  setup. 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.

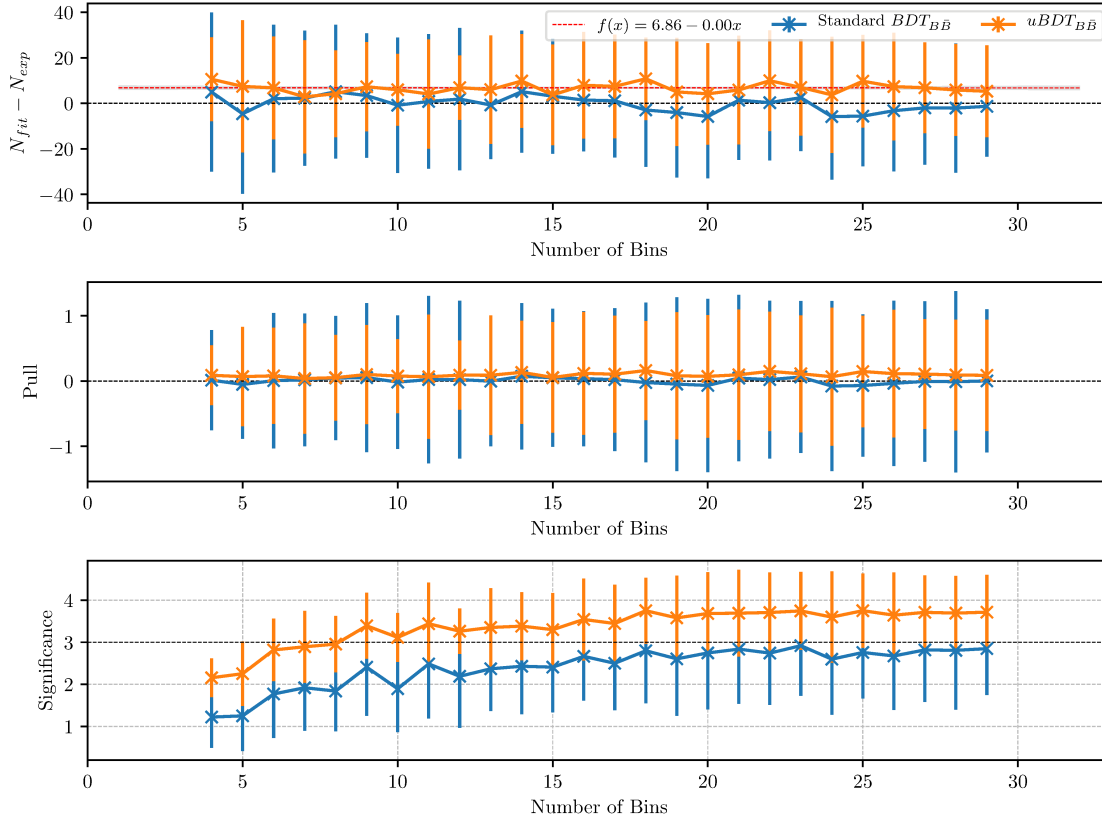


Figure 9.1: Expected yield difference (top), pulls (center) and fit significance (bottom) as a function of binning in  $\Delta E$  and  $M_{BC}$  for the final sample, optimized with the standard  $BDT$  classifier (blue) and the  $uBDT$  classifier (orange).

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 9.2, while all fit results are shown in Figure 9.3. All stream fit results were fitted with a 0<sup>th</sup> degree polynomial. The global result seems to describe the expected value precisely, with the residual bias much smaller than the average statistical error. The normalized  $\chi^2$  value with  $10 - 1 = 9$  degrees of freedom for the global fit is  $\chi_9^2 = 1.23$ , while the average statistical significance of the fits is around  $3.5\sigma$ .

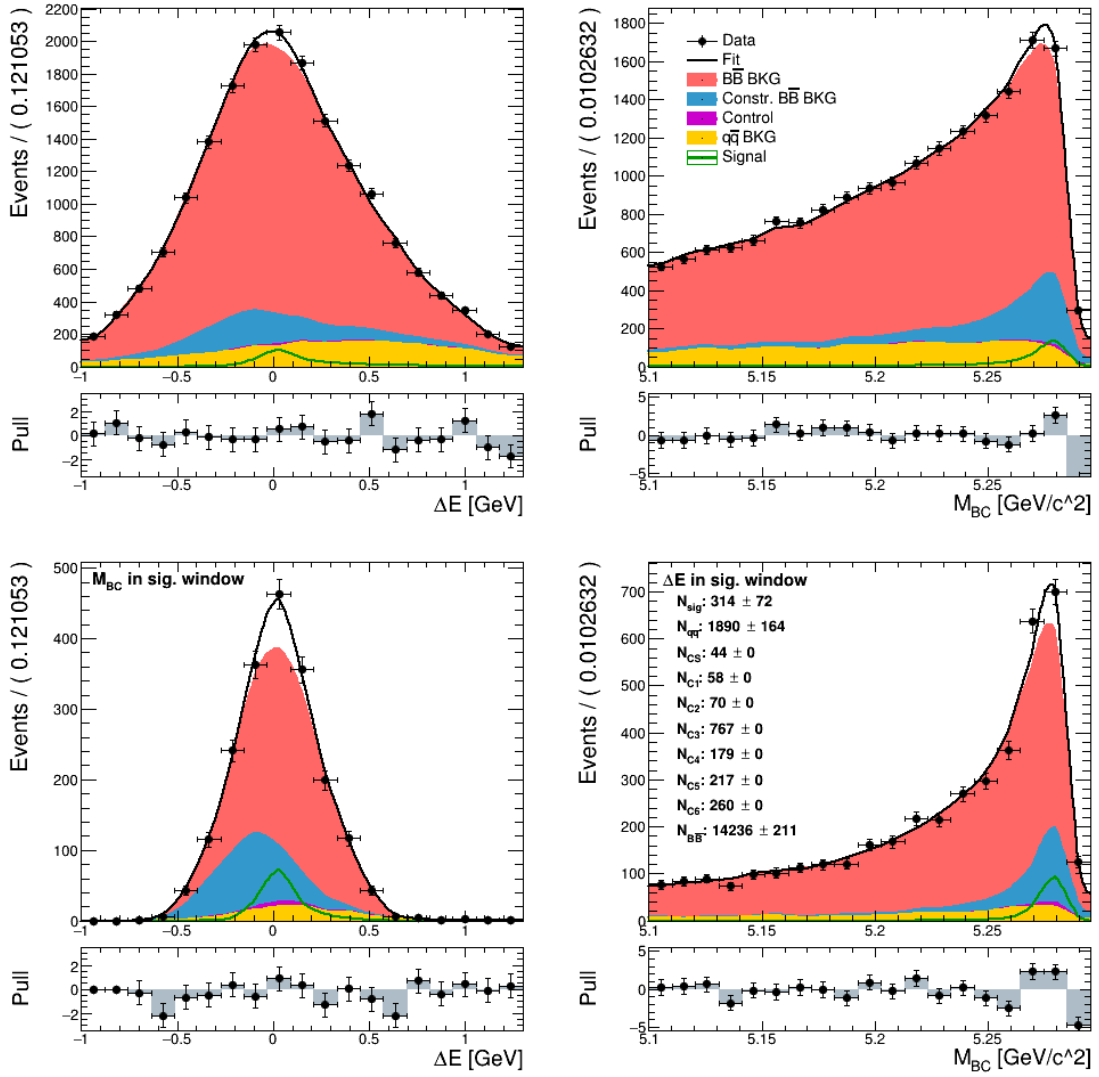


Figure 9.2: An example fit to one stream of MC. Left column shows the  $\Delta E$  and the right column shows the  $M_{BC}$  distribution of the full fitted sample in the full fit region (top) and the in signal enhanced region (bottom).

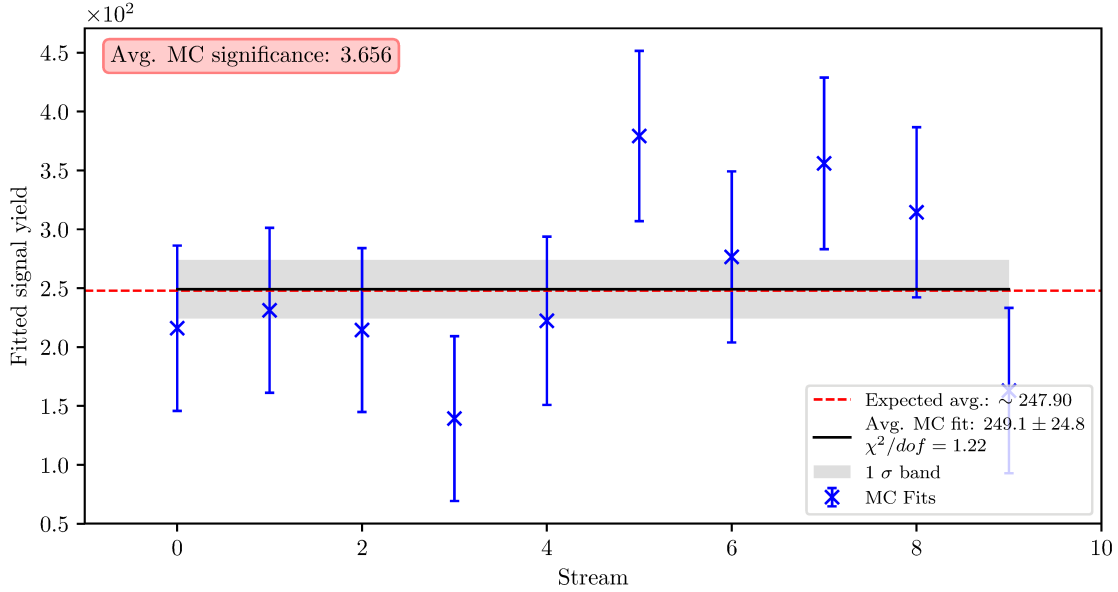


Figure 9.3: Fits to all 10 streams of MC and the global fit with a 0<sup>th</sup> degree polynomial. The red line shows the mean value of the global fit and the gray band shows the 1 $\sigma$  confidence interval.

## 9.2 Control Fit Result

With the control fit setup described in Section 8.1.1, we proceed to fit the control sample for the 10 streams of MC and 1 stream of data. A typical MC stream fit is shown in Figure 9.4 for MC and in Figure 9.5 for the data, while all fit results are shown in Figure 9.6, where all streams of MC are fitted with a 0<sup>th</sup> degree polynomial. The control fit results for the split and joined lepton modes are shown in Table 9.1.

	$N^{MC}$	$N^{\text{data}}$
$\ell = e \text{ or } \mu$	$1182 \pm 11$	$1187 \pm 44$
$\ell = e$	$591 \pm 8$	$583 \pm 28$
$\ell = \mu$	$592 \pm 7$	$613 \pm 30$

Table 9.1: Control sample fit results for MC and data for various lepton final state modes.



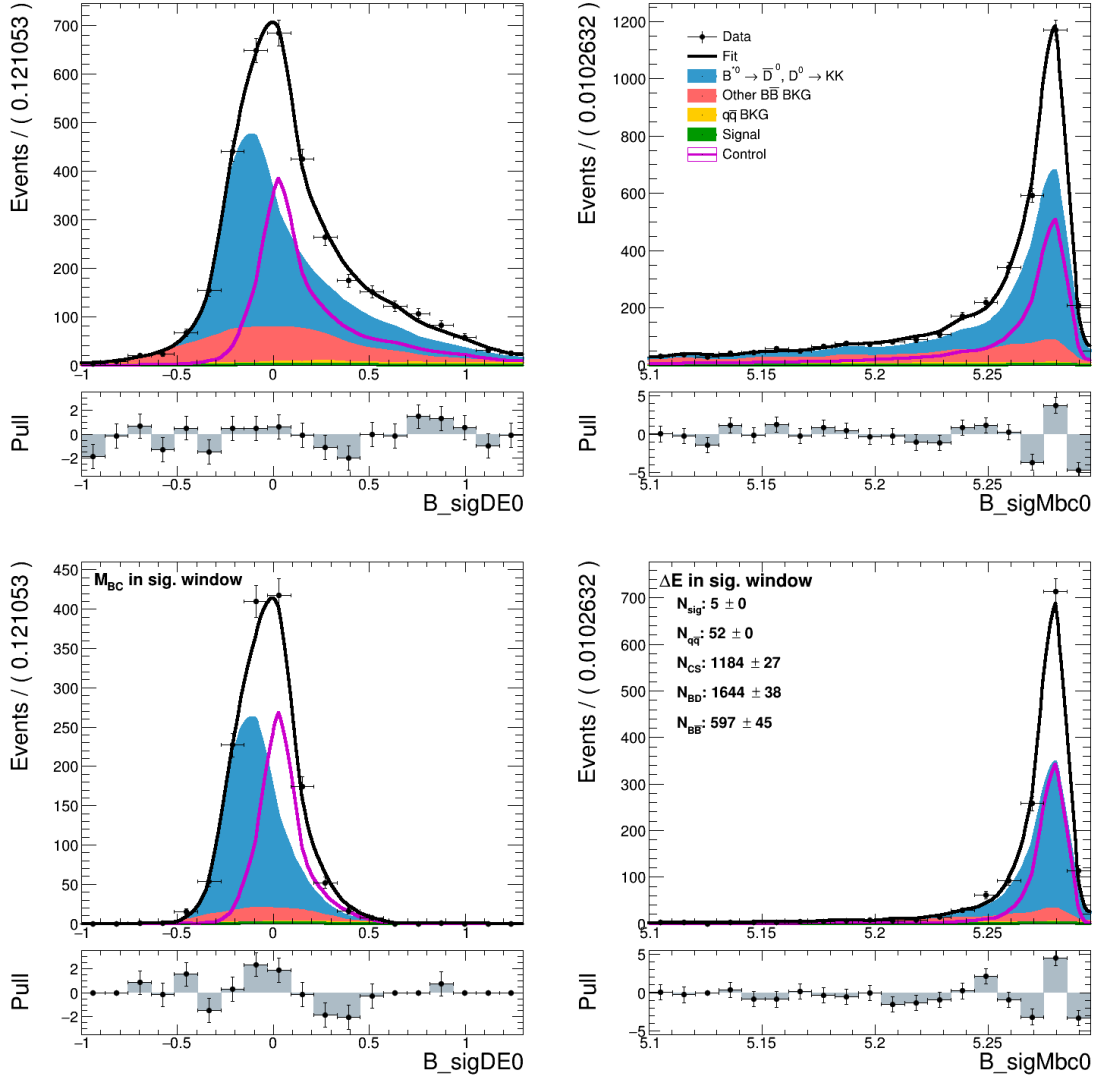


Figure 9.4: Control fit result on one stream of MC. Left column shows the  $\Delta E$  and the right column shows the  $M_{BC}$  distribution in the full fit window (top) and the in signal enhanced window (bottom).

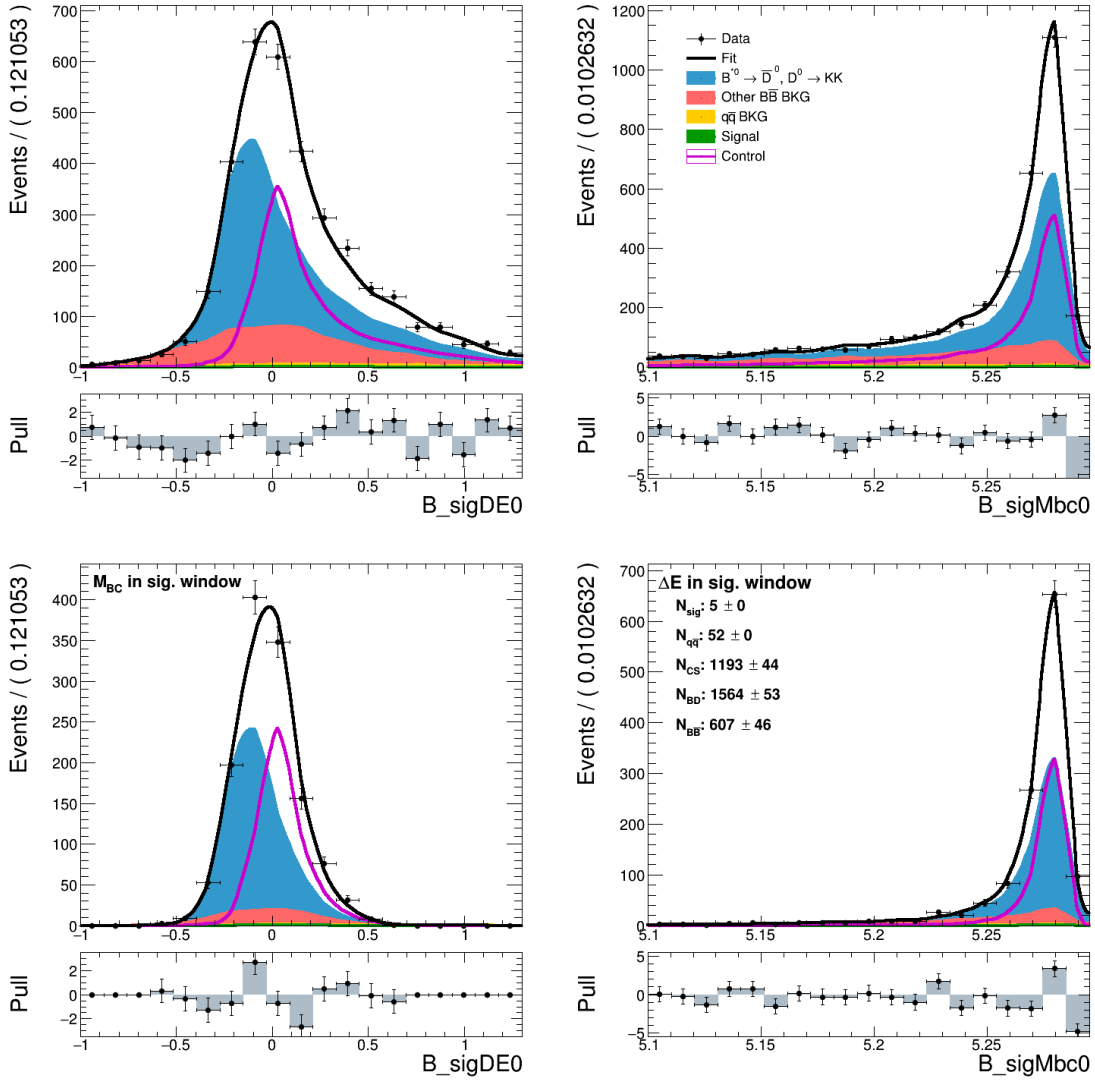


Figure 9.5: Control fit result on real data. Left column shows the  $\Delta E$  and the right column shows the  $M_{BC}$  distribution in the full fit window (top) and the signal enhanced window (bottom).

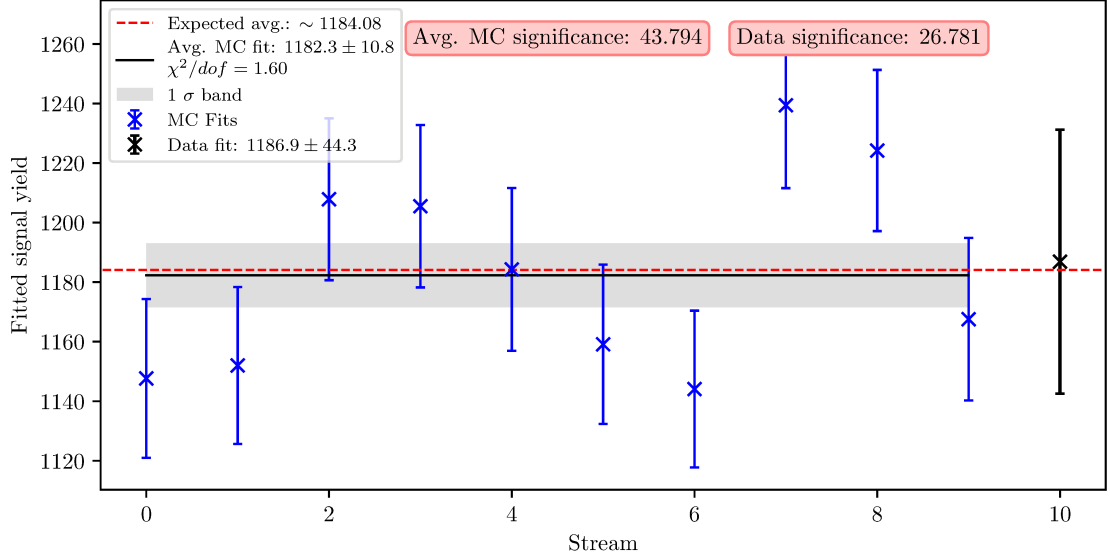


Figure 9.6: Control fit to the data and all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0<sup>th</sup> degree polynomial. The gray band shows the 1 $\sigma$  confidence interval around the global MC fit.

### 9.2.1 Branching Ratio Measurement for Control Decay

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

$$\mathcal{B}_{\text{control}}^{MC} = \frac{N_{\text{control}}^{MC} \times \epsilon_{MC}}{2N_{B\bar{B}}^{MC}}, \quad (9.1)$$

$$\mathcal{B}_{\text{control}} = \frac{N_{\text{control}} \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \quad (9.2)$$

where  $N_{\text{control}}^{MC}$  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  $N_{B\bar{B}}^{MC}$  and  $N_{B\bar{B}}$  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 fraction 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  $B^+ \rightarrow \bar{D}^0 \ell^+ \nu$ ,  $D^0 \rightarrow K^+ K^-$ . After applying the final selection, the full and split efficiencies, with respect to the lepton final state, were determined to be

$$\begin{aligned} \epsilon_{MC} &= (8.61 \pm 0.04) \times 10^{-3}, \\ \epsilon_{MC}^e &= (4.25 \pm 0.03) \times 10^{-3}, \\ \epsilon_{MC}^\mu &= (4.35 \pm 0.03) \times 10^{-3}. \end{aligned}$$

The efficiency error was estimated according to the binomial distribution

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

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 the data and MC. It is described in detail in Section 10.1.1 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 the data sample by the collaboration. The values are

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

Finally, we can determine the branching fractions based on the calculations in Eq. (9.2). The obtained values are shown in Table 9.2 and graphically presented in Figure 9.7, along with the MC generated value and the current PDG world average. Both MC and the data results for the control decay branching fraction are in agreement with the expected and the world average values. One should note that the error bars correspond to statistical uncertainty and the PID systematic uncertainty only. Other systematic uncertainties are not included, since this measurement is not the goal of our analysis.

	$\mathcal{B}_{GEN} [\times 10^{-5}]$	$\mathcal{B}_{PDG} [\times 10^{-5}]$	$\mathcal{B}^{MC} [\times 10^{-5}]$	$\mathcal{B}^{data} [\times 10^{-5}]$
$\ell = e \text{ or } \mu$			$8.97 \pm 0.09$	$9.03 \pm 0.40$
$\ell = e$	9.01	$9.10 \pm 0.42$	$9.07 \pm 0.14$	$9.02 \pm 0.49$
$\ell = \mu$			$8.88 \pm 0.12$	$9.18 \pm 0.50$

Table 9.2: Control sample fit results for MC and data for various lepton final state modes.

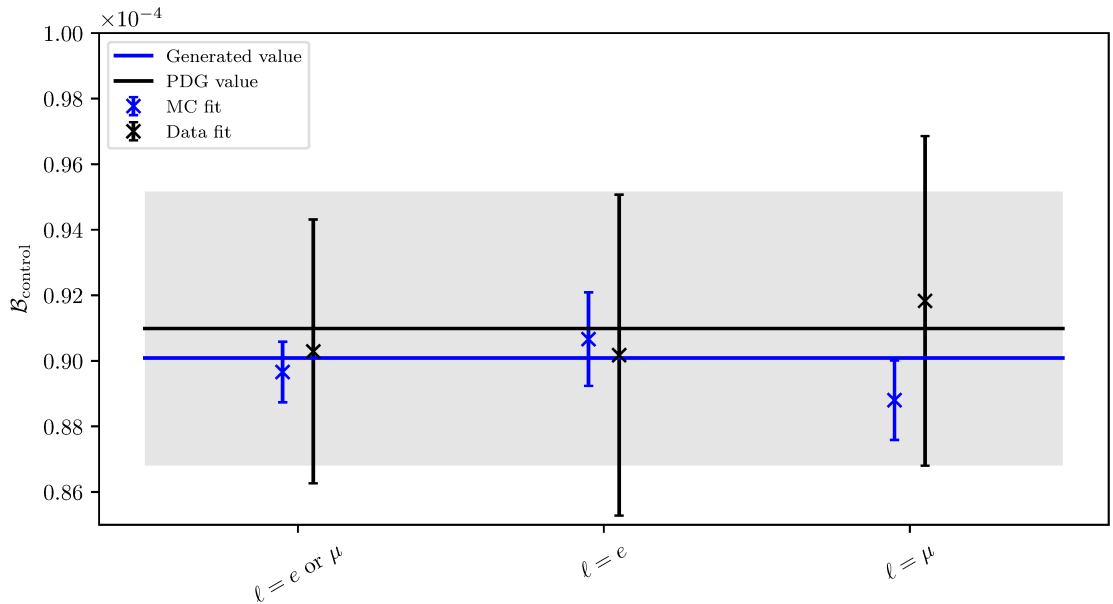


Figure 9.7: Various branching fraction determinations for the control decay of our analysis.

## 9.3 Results of Signal Fit in Data

### 9.3.1 Signal Yield

Finally, after validating the analysis procedure using the signal MC sample, control MC sample and the control data sample, we can continue to perform the signal extraction process on the full Belle  $\Upsilon(4S)$  data sample. Figure 9.8 shows the fit result in projections of  $\Delta E$  and  $M_{BC}$  for both fit regions. The extracted signal yield in data, as well as the yields of the remaining contributions, are shown in Table 9.3. Values of all the constraints are shown in Table 9.4. As the main results, we obtain the signal yield of

$$N_{\text{sig}} = 491 \pm 86, \quad \text{fit significance} = 6.0\sigma. \quad (9.3)$$

The uncertainty includes the statistical uncertainty, as well as the partial systematic uncertainty, due to the limited knowledge of various branching fractions, entering the fit in a form of Gaussian constraints. The significance of the result, taking into account only the mentioned uncertainties, is  $6.0\sigma$ . The latter is obtained via the likelihood-ratio test, as previously described in Eq. (8.4).

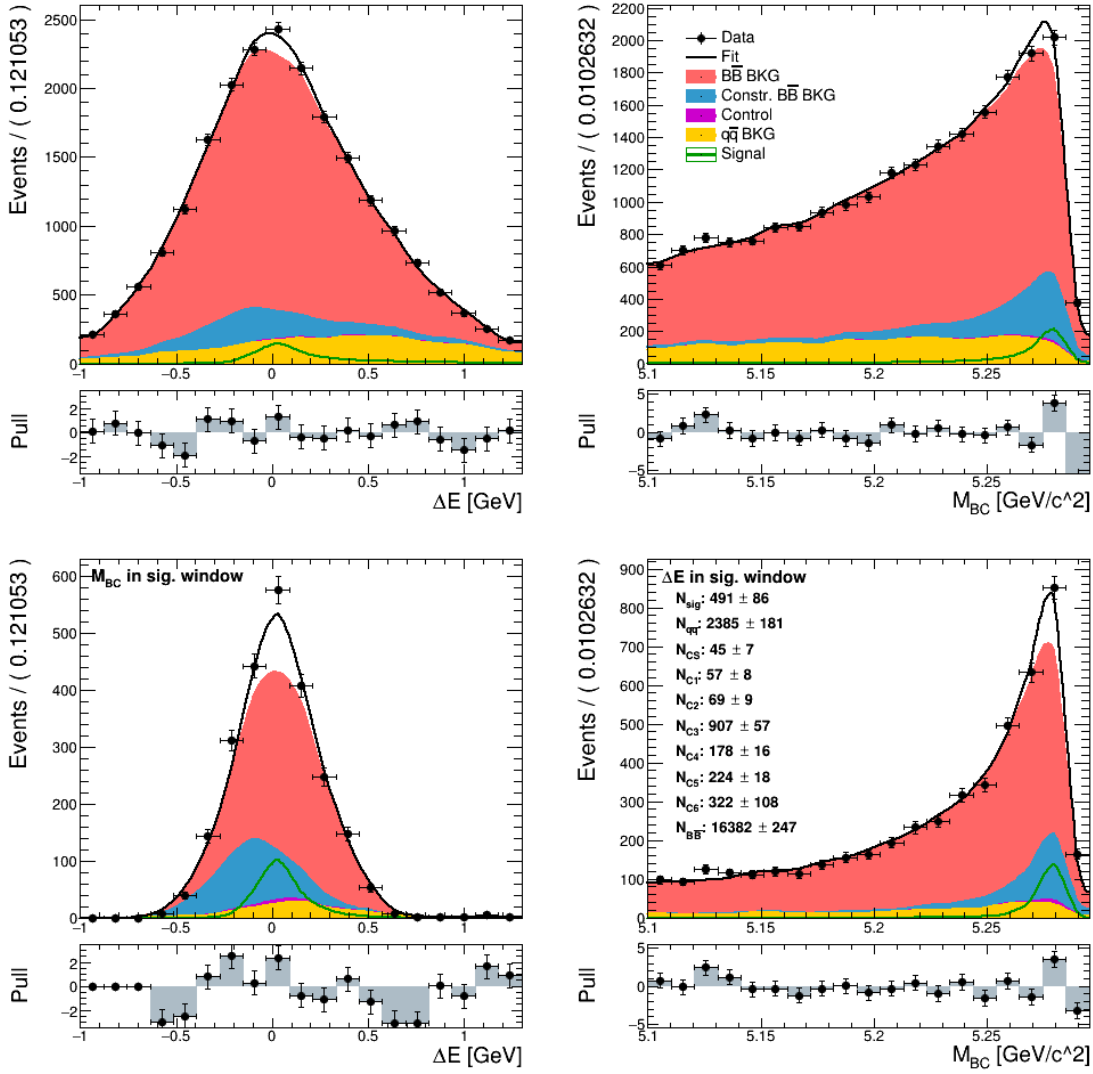


Figure 9.8: Signal fit result on real data. Left column shows the  $\Delta E$  and the right column shows the  $M_{BC}$  distribution in the full fit window (top) and in the signal enhanced window (bottom).

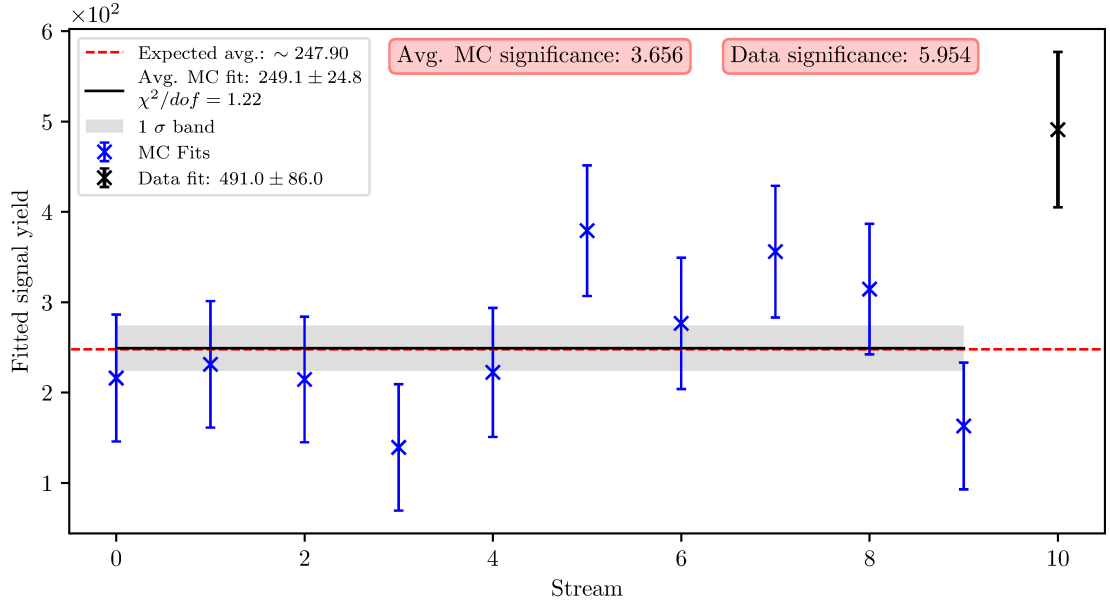


Figure 9.9: Signal fit to the data and all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0<sup>th</sup> degree polynomial. The gray band shows the  $1\sigma$  confidence interval around the global MC fit.

Category	Fit Yield
Signal	$491 \pm 86$
$q\bar{q}$ background	$2385 \pm 181$
$C_0$	$45 \pm 7$
$C_1$	$57 \pm 8$
$C_2$	$69 \pm 9$
$C_3$	$907 \pm 57$
$C_4$	$178 \pm 16$
$C_5$	$224 \pm 18$
$C_6$	$322 \pm 108$
Other $B\bar{B}$ background	$16382 \pm 247$

Table 9.3: Yields of all signal fit contributions in data.

Constraint	Value	Constraint	Value
$\mathcal{B}_0$	$0.049 \pm 0.001$	$N_{30}$	$103 \pm 10$
$\mathcal{B}_1$	$1.076 \pm 0.004$	$N_{31}$	$212 \pm 14$
$\mathcal{B}_2$	$0.021 \pm 0.001$	$N_{32}$	$136 \pm 12$
$\mathcal{B}_3$	$1.014 \pm 0.018$	$N_{33}$	$268 \pm 16$
$\mathcal{B}_4$	$1.018 \pm 0.010$	$N_{34}$	$20 \pm 4$
$\mathcal{B}_5$	$1.439 \pm 0.063$	$N_{35}$	$35 \pm 6$
$\mathcal{B}_6$	$1.092 \pm 0.038$	$N_{40}$	$48 \pm 7$
$\mathcal{B}_7$	$1.035 \pm 0.008$	$N_{41}$	$7 \pm 3$
$\mathcal{B}_8$	$0.971 \pm 0.038$	$N_{42}$	$99 \pm 10$
$\mathcal{B}_9$	$1.053 \pm 0.038$	$N_{43}$	$10 \pm 3$
$\mathcal{B}_{10}$	$0.946 \pm 0.029$	$N_{44}$	$14 \pm 4$
$\mathcal{B}_{11}$	$1.264 \pm 0.446$	$N_{45}$	$3 \pm 2$
$N_{\text{control}}^{MC}$	$1181 \pm 11$	$N_{50}$	$104 \pm 10$
$N_{\text{control}}^{Data}$	$1210 \pm 43$	$N_{51}$	$64 \pm 8$
$N_{00}$	$44 \pm 7$	$N_{52}$	$31 \pm 6$
$N_{10}$	$54 \pm 7$	$N_{53}$	$22 \pm 5$
$N_{11}$	$6 \pm 2$	$N_{60}$	$69 \pm 8$
$N_{20}$	$23 \pm 5$	$N_{61}$	$94 \pm 10$
$N_{21}$	$41 \pm 6$	$N_{62}$	$63 \pm 8$
$N_{22}$	$6 \pm 2$	$N_{63}$	$35 \pm 6$

Table 9.4: Mean values and standard deviations of constraints after the fit.

In this fit setup, we perform a random smearing of the  $\Delta E$  (see Eq. 8.9 and description there). Because of the randomness involved in the smearing, such a method results in small random fluctuations of the fitted central values. In order to check the consistency of the fit result because of the  $\Delta E$  smearing, we perform 500 fit repetitions. The result in Eq. (9.3) represents the average result, corresponding to the central values of the yield, uncertainty, and the significance of the signal fit in data.

The systematic part of the fit uncertainty can be estimated by fixing the Gaussian constraints to central values, presented in Table 9.4, and repeating the fit. The resulting uncertainty and significance are purely statistical and are found to be

$$\sigma_{\text{stat}} = 81, \quad \text{fit significance} = 6.3\sigma. \quad (9.4)$$

Other sources of systematic uncertainties and their estimations are presented in details in Chapter 10.



### 9.3.2 Branching Ratio

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

$$\mathcal{B}_{\text{sig}}^{MC} = \frac{N_{\text{sig}}^{\text{MC}} \times \epsilon_{MC}}{2N_{B\bar{B}}^{MC}}, \quad (9.5)$$

$$\mathcal{B}_{\text{sig}} = \frac{N_{\text{sig}} \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \quad (9.6)$$

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 in 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

$$\epsilon_{MC} = (1.052 \pm 0.003) \times 10^{-2},$$

where the efficiency error was calculated in the same manner as in Section 9.2.1. 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 fraction measurement.

Finally, we can determine the branching fractions based on the calculations in Eq. (9.6). The obtained values are shown in Table 9.5. The measured value is almost twice as large as the MC value. The errors in all cases are statistical only. The result shows that the MC contribution to our simulated samples is underestimated and indicates that the branching fraction of the decay may be large enough to affect results in precision physics, in cases where it is ignored.

	$\mathcal{B}_{GEN} [\times 10^{-5}]$	$\mathcal{B}^{MC} [\times 10^{-5}]$	$\mathcal{B}^{\text{data}} [\times 10^{-5}]$
$\ell = e \text{ or } \mu$	1.57	$1.55 \pm 0.15$	$3.04 \pm 0.51$

Table 9.5: Signal decay branching fraction results on MC and in data.

### 9.3.3 Signal Distribution in bins of $m_{KK}$

It is possible to take a deeper look in the signal distribution over the  $m_{KK}$  variable by performing the signal fit in bins of the  $m_{KK}$  distribution, instead of a single fit over the whole region. This offers a deeper insight into the decay process and provides more details about the reliability of our MC samples. Results can be used to check theoretical predictions or to update existing MC generators. Table 9.6 shows the selected regions in  $m_{KK}$ , along with the corresponding signal yields on MC and in data. The results are graphically presented in Figure 9.10. Figures for each  $m_{KK}$  window fit can be found in Appendix C.1. In the given results, the  $m_{KK}$  regions around the  $\phi$  and the  $D^0$  resonances have been excluded (see Section 7.1).

Region	Expected	MC fit	Data fit
$0.980 < m_{KK} < 1.232$	44	$42 \pm 11$	$113 \pm 29$
$1.232 < m_{KK} < 1.483$	66	$64 \pm 12$	$185 \pm 40$
$1.483 < m_{KK} < 1.735$	67	$67 \pm 15$	$166 \pm 47$
$1.735 < m_{KK} < 1.987$	38	$35 \pm 9$	$87 \pm 46$
$1.987 < m_{KK} < 2.238$	24	$26 \pm 7$	$35 \pm 25$
$2.238 < m_{KK} < 2.490$	9	$8 \pm 4$	$-10 \pm 13$
$2.490 < m_{KK} < 2.742$	4	$7 \pm 4$	$-8 \pm 7$
$2.742 < m_{KK} < 2.993$	1	$0 \pm 1$	$-2 \pm 5$
$2.993 < m_{KK} < 3.245$	0	$-3 \pm 1$	$-2 \pm 9$
$3.245 < m_{KK} < 3.497$	0	$-0 \pm 1$	$-1 \pm 6$
$3.497 < m_{KK} < 3.748$	0	$-0 \pm 1$	$-2 \pm 8$
$3.748 < m_{KK} < 4.000$	0	$-1 \pm 1$	$-1 \pm 224$

Table 9.6: Various signal fit yields for each of the defined  $m_{KK}$  windows.

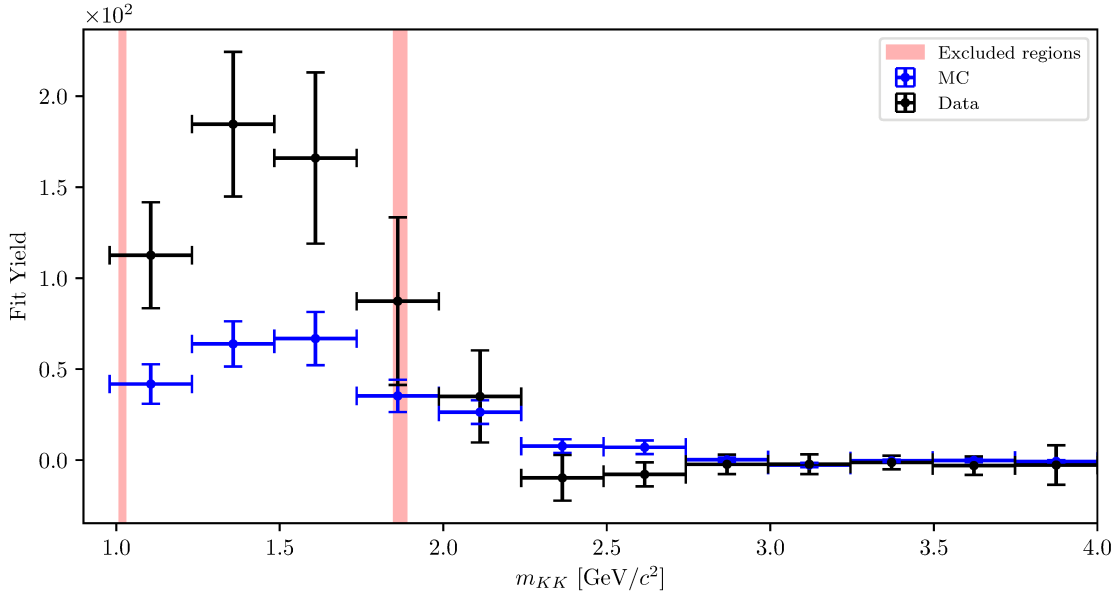


Figure 9.10: Signal yield distribution as a function of  $m_{KK}$ .

### 9.3.4 Signal Distribution in bins of $q^2$

Similarly, we look at the signal distribution over the square momentum transfer to the lepton pair,  $q^2$ , by performing the signal fit in bins of the  $q^2$  distribution. In addition to the possibility of improved modeling of the MC generators, this enables the extraction of the  $V_{ub}$  parameter of the CKM matrix, provided that the appropriate form factors of the decay are available. Table 9.7 shows the selected regions in  $q^2$ , along with the corresponding signal yields on MC and data. The results are

graphically presented in Figure 9.11. Figures for each  $q^2$  window fit can be found in Appendix C.2.

Region	Expected	MC fit	Data fit
$0.000 < q^2 < 1.500$	27	$22 \pm 11$	$-19 \pm 31$
$1.500 < q^2 < 3.000$	31	$30 \pm 4$	$25 \pm 29$
$3.000 < q^2 < 4.500$	30	$30 \pm 5$	$83 \pm 32$
$4.500 < q^2 < 6.000$	28	$28 \pm 6$	$124 \pm 31$
$6.000 < q^2 < 7.500$	31	$29 \pm 8$	$96 \pm 31$
$7.500 < q^2 < 9.000$	25	$24 \pm 8$	$83 \pm 30$
$9.000 < q^2 < 10.500$	24	$22 \pm 8$	$37 \pm 28$
$10.500 < q^2 < 12.000$	19	$19 \pm 5$	$57 \pm 24$
$12.000 < q^2 < 13.500$	16	$15 \pm 5$	$17 \pm 18$
$13.500 < q^2 < 15.000$	9	$5 \pm 3$	$15 \pm 14$
$15.000 < q^2 < 16.500$	5	$3 \pm 2$	$17 \pm 3$
$16.500 < q^2 < 18.000$	1	$-3 \pm 1$	$2 \pm 4$

Table 9.7: Various signal fit yields for each of the defined  $q^2$  windows.

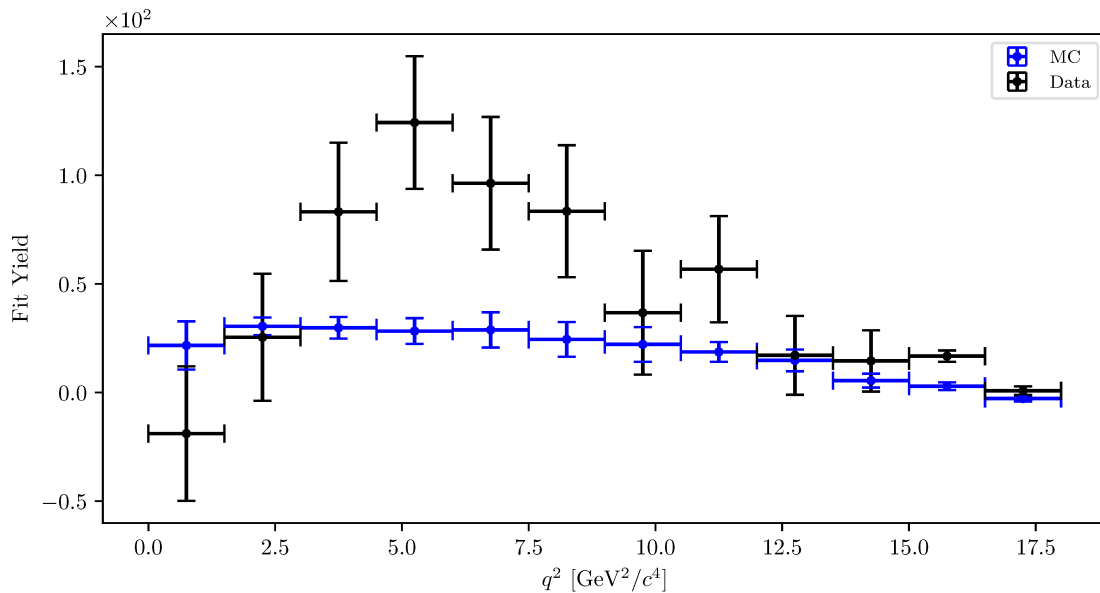


Figure 9.11: Signal yield distribution as a function of  $q^2$ .

The measured  $q^2$  and  $m_{KK}$  distributions exhibit differences with respect to the MC models used. Differences are larger in the  $q^2$  distribution, where the measured signal appears to fall more rapidly than in simulation in the region below  $2.5 \text{ GeV}^2/c^4$ , and is more concentrated in the region of  $2.5 \text{ GeV}^2/c^4 < q^2 < 10 \text{ GeV}^2/c^4$ . In  $m_{KK}$ , both data and simulation are concentrated in the region below  $2 \text{ GeV}/c^2$ .



# Chapter 10

## 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.

### 10.1 Contributions to the Systematic Uncertainty

#### 10.1.1 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 a set of correction factors and tables of systematic uncertainties 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 10.1, as well as the corresponding entry for all 3 particles. The entries are shown for both the signal and the control region, where the differences are related to the kinematic variations between the two decay modes.

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  $KK$  PID error, as both  $K$  use the same PID information.

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

$$\sigma_{\text{sys}}^{\text{PID}} = 10, \quad \delta_{\text{sys}}^{\text{PID}} = 2.0\%, \quad (10.1)$$

for the signal, as well as the control decay.

PID correction and systematic uncertainties	Control decay	Signal decay
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)	$1.004 \pm 0.009$	$1.006 \pm 0.009$
$e$	$0.977 \pm 0.011$	$0.976 \pm 0.011$
$\mu$	$0.985 \pm 0.009$	$0.980 \pm 0.009$
$\ell$	$0.981 \pm 0.007$	$0.980 \pm 0.007$
$KKe$	$0.986 \pm 0.021$	$0.988 \pm 0.022$
$KK\mu$	$0.994 \pm 0.020$	$0.993 \pm 0.021$
$KK\ell$	$0.991 \pm 0.019$	$0.990 \pm 0.020$

Table 10.1: PID correction factors and systematic uncertainties for various charged particles and their combinations.

### 10.1.2 Fit Bias and Binning Effects

Signal and background templates in our analysis are not perfectly distinct from one another and may potentially cause some over- or underestimation of the fitted signal yield. In order to study this problem, we estimate the bias from the binning study performed in Section 9.1, 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

$$f_{\min}(x) = -7.25 - 1.12x - \sigma_{f_{\min}}(x), \quad (10.2)$$

$$\sigma_{f_{\min}}(x) = \sqrt{0.050x^2 - 0.175x + 0.410}, \quad (10.3)$$

$$f_{\max}(N_b) = 6.86 + \sigma_{f_{\max}}(N_b), \quad (10.4)$$

$$\sigma_{f_{\max}}(N_b) = \sqrt{0.004N_b^2 - 0.113N_b + 1.112} \quad (10.5)$$

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 in data, with the fit setup of  $N_b = 19$  bins, was determined to be  $N_{sig} = 491$ , which leads to  $x = N_{sig}/N_{sig}^{MC} = 491/249 \approx 2$ . The bias interval is therefore

$$\sigma_{\text{sys}}^{\text{bias}} = \begin{matrix} +7 \\ -10 \end{matrix}, \quad \delta_{\text{sys}}^{\text{bias}} = \begin{matrix} +1.5\% \\ -2.0\% \end{matrix}. \quad (10.6)$$

### 10.1.3 Gaussian Constraints

As mentioned in Section 9.3, it is possible to estimate the size of the systematic error of the Gaussian constraints. By fixing the constraints to the central values in Table 9.4, we obtain the pure statistical error, which can then be subtracted from the average fit error in order to determine the systematic uncertainty contribution, arising from the Gaussian constraints. Due to the nature of implementing the smearing of the  $\Delta E$  variable, there is some randomness involved in our fits, so we perform 500 fits for the fixed and non-fixed case, in order to determine the uncertainties. The

split errors are then

$$\bar{\sigma}_{\text{fit}} = 86, \quad (10.7)$$

$$\bar{\sigma}_{\text{stat}} = 81, \quad (10.8)$$

$$\sigma_{\text{sys}}^{GC} = 26, \quad \delta_{\text{sys}}^{GC} = 5.4\%, \quad (10.9)$$

where GC stands for the Gaussian constraints. We see that the constrained channels are very well defined and introduce a relatively small level of uncertainty.

### 10.1.4 Fit Template Smearing and Offset

The smearing and offset of the  $\Delta E$  variable was discussed in Section 8.1.1, where we estimated the central value of the parameters, as well as their range in the  $1\sigma$  confidence interval. We have performed a study of the effects of different smearing and offset parameter values in data. From Section 8.1.1, the parameter values are

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

Since the two parameters are largely uncorrelated, we are able to perform the study in the form of signal fits with four different combinations of parameters in the given range. For each parameter setting, the Gaussian constraints are fixed and 500 fits are performed to obtain the following results

- set: [smearing, offset]: [23 MeV, 6 MeV], Result:  $\bar{N}_{sig} = 458$ ,
- set: [smearing, offset]: [55 MeV, 6 MeV], Result:  $\bar{N}_{sig} = 532$ ,
- set: [smearing, offset]: [40 MeV, 0 MeV], Result:  $\bar{N}_{sig} = 532$ ,
- set: [smearing, offset]: [40 MeV, 12.6 MeV], Result:  $\bar{N}_{sig} = 460$ ,

which results in the following estimate of systematic uncertainties for smearing and offset parameters

$$\sigma_{\text{sys}}^{\text{sm.}} = \begin{matrix} +41 \\ -33 \end{matrix}, \quad \delta_{\text{sys}}^{\text{sm.}} = \begin{matrix} +8.3\% \\ -6.7\% \end{matrix}, \quad (10.10)$$

$$\sigma_{\text{sys}}^{\text{off.}} = \begin{matrix} +41 \\ -31 \end{matrix}, \quad \delta_{\text{sys}}^{\text{off.}} = \begin{matrix} +8.4\% \\ -6.3\% \end{matrix}. \quad (10.11)$$

### 10.1.5 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 values according to the Poisson distribution. This procedure is repeated 500 times with the Gaussian constraints fixed to the central values in Table 9.4 and with a fixed random seed for the application of smearing in  $\Delta E$ . To estimate the size of this uncertainty, we take the width of the fit yield distribution. The resulting finite MC sample contribution of the systematic uncertainty is

$$\sigma_{\text{sys}}^{MC} = 26, \quad \delta_{\text{sys}}^{MC} = 5.3\%. \quad (10.12)$$

### 10.1.6 MVA Selection Efficiencies

Control sample fits allow us to check the behavior of the optimized MVA selection on MC, as well as in 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 selection efficiencies (double ratios). The following selection criteria 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 10.1. The plot shows that the yield ratios and selection efficiency ratios are consistent with 1. This means that data and MC are in agreement before, as well as after applying the final selection. 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 values, for each step in the MVA selection, except for the results coming from measurements with specific lepton states. The systematic error estimate for this contribution is

$$\sigma_{\text{sys.}}^{\text{MVA}} = 5, \quad \sigma_{\text{sys.}}^{\text{MVA}} = 1.0\%. \quad (10.13)$$



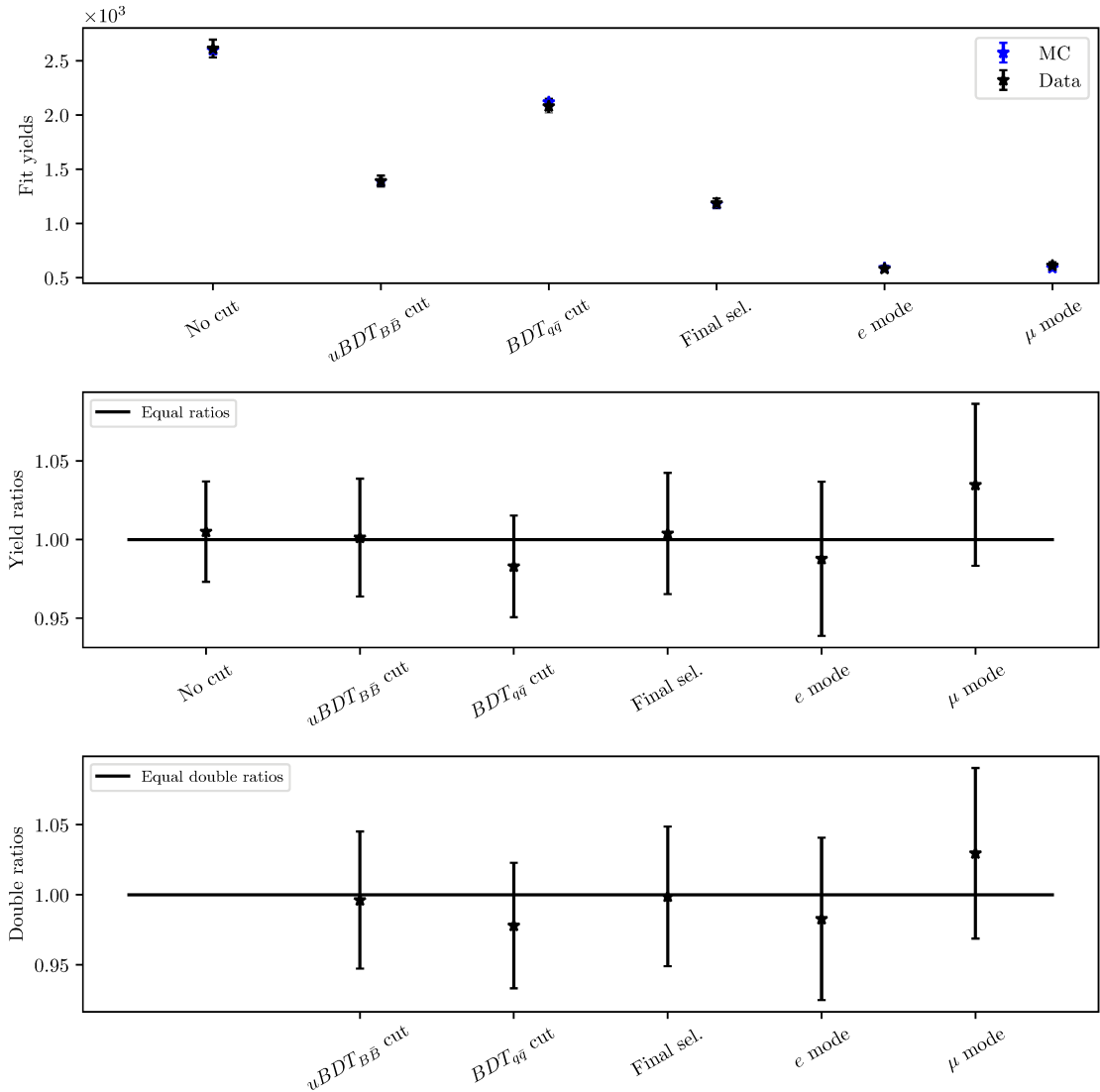


Figure 10.1: Fit yields, their ratios and ratios of selection efficiencies (double ratios) for the control sample fits in data and on MC.

### 10.1.7 Model Uncertainty Effects

The signal decay model used in the MC generation is ISGW2 [21], which is known to result in unrealistic predictions and poor agreement with data, so it is not the most precise model for our signal MC sample. Due to this model unreliability, our analysis has been set up as model-independent as possible via means of not using variables, which exhibit model dependence. Such variables are, for example, the 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 three additional signal MC samples, produced with three extreme scenarios of decay model choice. In the first scenario, we generate the signal MC sample with a generic phase-space decay mode, PHSP [10], which results in a continuum-like distributions of  $q^2$  and  $m_{KK}$ . In the other scenario, only resonant-like contributions of  $m_{KK}$  are

used. The third scenario is obtained by sampling the first model in a non-uniform way, so that the final distribution is more dominant in the higher region of the  $q^2$  distribution. These scenarios act as extreme cases of decay model choice and present a reasonable, if not a conservative measure of the model uncertainty.

### Shape Effects of the Model

The shape effect is evaluated by using signal templates, constructed from different signal MC samples. We perform 500 fits for each case, with the Gaussian constraints fixed to the central values, presented in Table 9.4. The differences between the mean values of fit yields serve as an estimate of the model uncertainty. Figure 10.2 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. Figures show a good agreement of  $\Delta E$  and  $M_{BC}$  for different cases of the generator model.

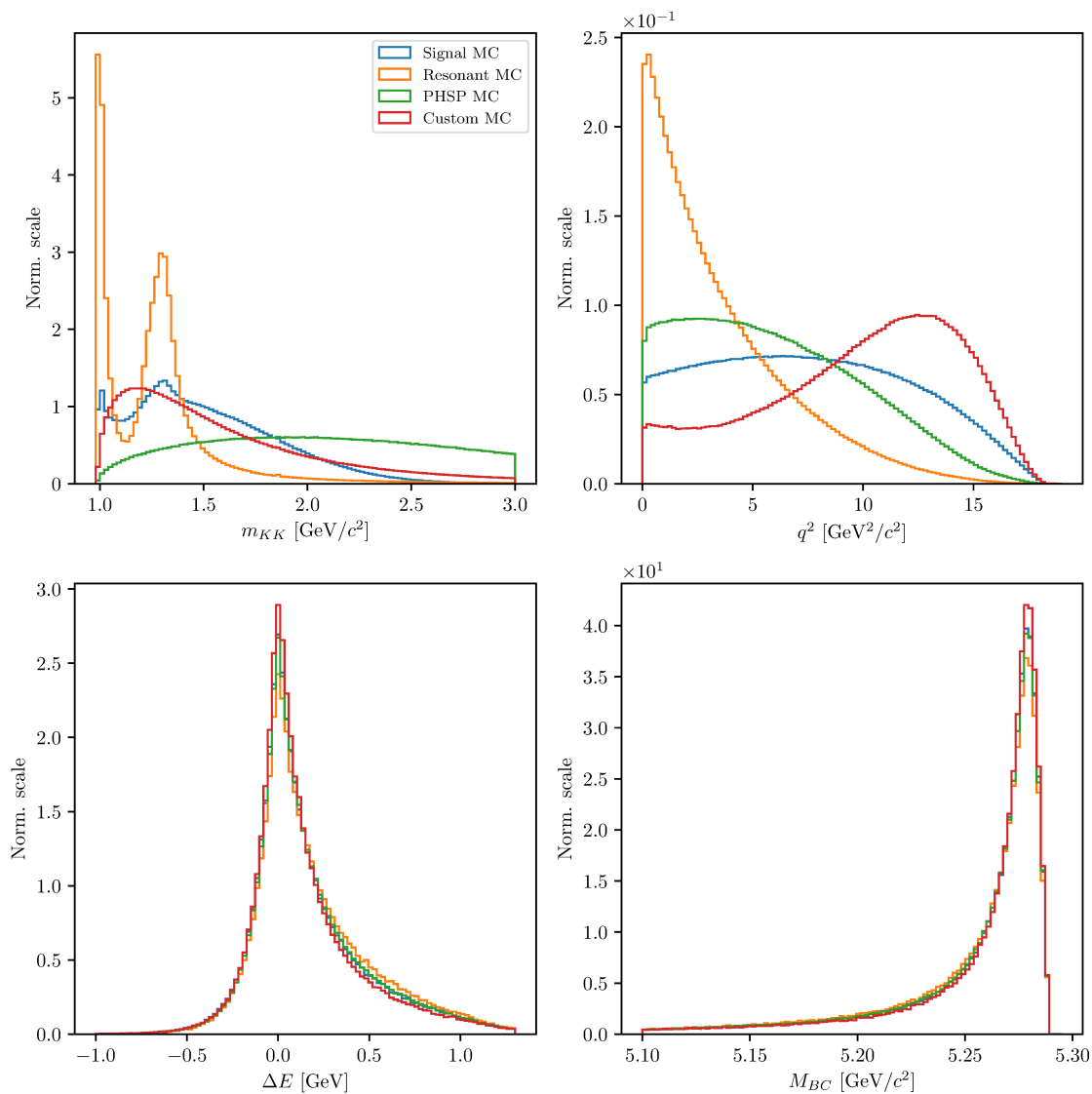


Figure 10.2:  $m_{KK}$  (top left),  $q^2$  (top right),  $\Delta E$  (bottom left), and  $M_{BC}$  (bottom right) for the main and the three extreme cases of signal MC generator choice.

The resulting average signal yields for the three model choices are

$$\bar{N}_{sig} = 491, \quad (10.14)$$

$$\bar{N}_{sig}^{\text{res.}} = 494, \quad (10.15)$$

$$\bar{N}_{sig}^{\text{PHSP}} = 536, \quad (10.16)$$

$$\bar{N}_{sig}^{\text{Custom}} = 451. \quad (10.17)$$

We see that, in general, the model with a worse resolution in  $\Delta E$  and  $M_{BC}$  will likely result in a larger yield, and vice versa for a model with a better resolution. Overall, these models give a conservative estimation of the systematic uncertainty with a value of

$$\sigma_{\text{sys}}^{\text{mod.}} = \begin{matrix} +45 \\ -39 \end{matrix}, \quad \delta_{\text{sys}}^{\text{mod.}} = \begin{matrix} +9.3\% \\ -8.0\% \end{matrix}. \quad (10.18)$$

### Model Efficiency

The second way that a model can affect our analysis is the efficiency of the model. Due to the different properties of the model, we can expect different efficiencies in model dependent variables like  $m_{KK}$  and  $q^2$ . The efficiency as a function of these variables is shown in Figure 10.3 for all discussed generator models. In a perfect scenario, a model-independent analysis should produce a flat efficiency with respect to the model dependent variables. While this is rarely achieved in practice, the overall discrepancy of the efficiency functions from very different models is relatively small.

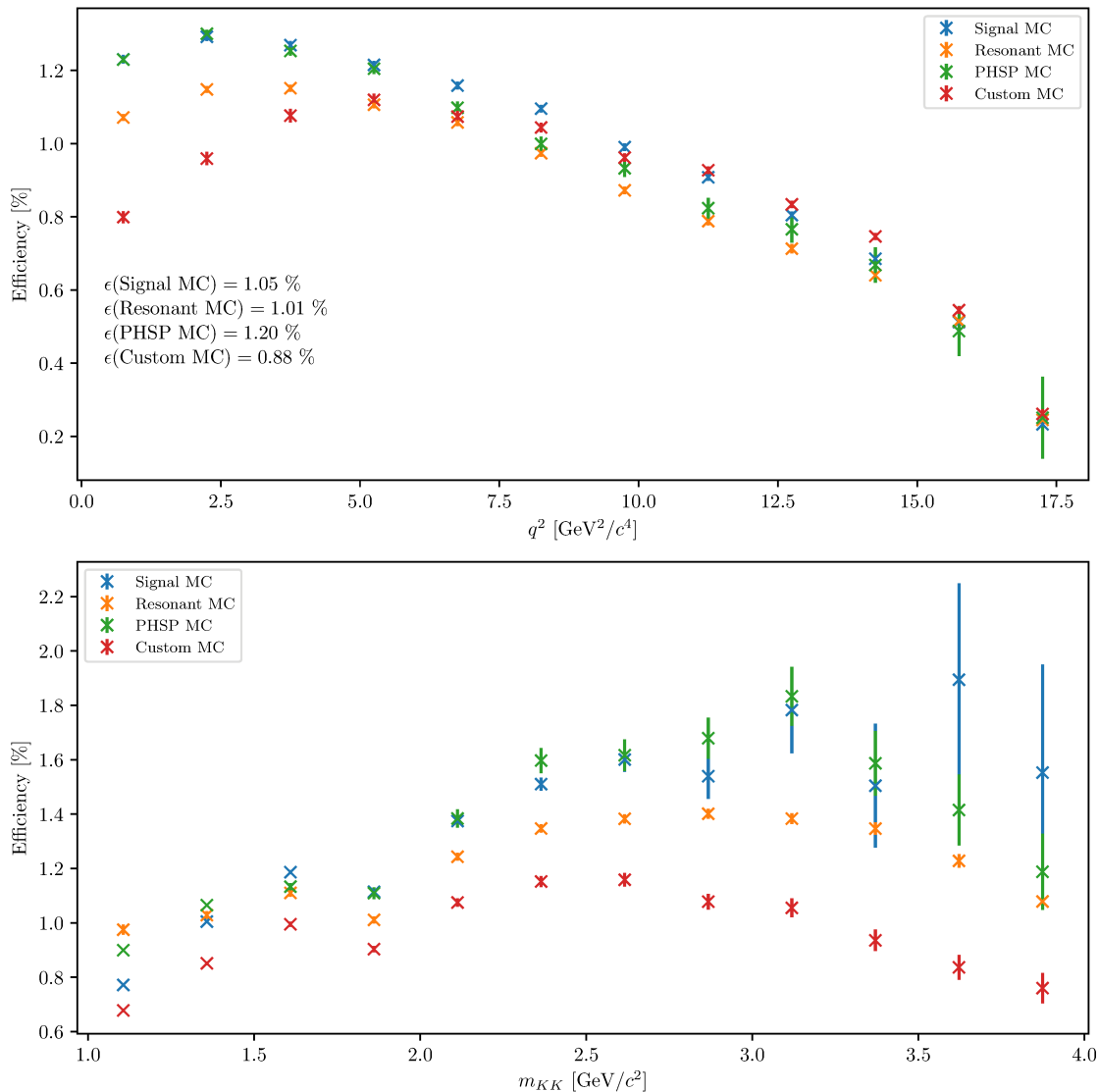


Figure 10.3: Efficiency as a function of  $q^2$  (top) and  $m_{KK}$  (bottom) for the main and three extreme cases of signal MC generator choice.

To estimate the systematic uncertainty due to the efficiency effects of the model, we compute the overall efficiencies and take the extreme cases of relative differences with regard to the main signal MC model. This results in

$$\sigma_{\text{sys}}^{\text{mod.}} = \begin{matrix} +70 \\ -79 \end{matrix}, \quad \delta_{\text{sys}}^{\text{mod.}} = \begin{matrix} +14.3\% \\ -16.2\% \end{matrix}. \quad (10.19)$$

As mentioned in Section 9.3.4, the measured  $q^2$  and  $m_{KK}$  distributions differ to some extent from the simulated ones. Since the reconstruction efficiency is not constant as a function of these variables, this could present an additional source of the systematic uncertainty. However, in  $m_{KK}$ , the measured signal is concentrated in the region of  $m_{KK} < 2$  GeV/c<sup>2</sup>, where the efficiency is relatively constant and within the range of the already estimated systematic uncertainty of around  $\pm 15\%$ . In  $q^2$ , the data distribution is wider and encompasses the region of a more pronounced efficiency dependence. We estimated the branching fraction using the signal yield as obtained in bins of  $q^2$  (see Table 9.7) and the efficiency as a function of  $q^2$ . The result, although with a larger statistical uncertainty, due to low statistics in

$q^2$  bins, is within the assigned systematic uncertainty, and therefore consistent with the branching fraction, found using the average reconstruction efficiency.

## 10.2 Summary of Systematic Uncertainties

The summary of all systematic uncertainties is shown in Table 10.2. The full estimate of the systematic uncertainty is summed up in quadrature and applied to the result in Section 9.3.2.

Source	$\sigma$	$\delta$ [%]
PID	10	2.0
Fit Bias	+7 -10	+1.5 -2.0
Gaussian Constraints	26	5.4
Template Smearing	+41 -33	+8.3 -6.7
Template Offset	+41 -31	+8.4 -6.3
Finite MC Effects	26	5.3
MVA Selection	5	1.0
Model Shape	+45 -39	+9.3 -8.0
Model Efficiency	+70 -79	+14.3 -16.2
Total	+109 -107	+22.2 -21.9

Table 10.2: Summary of systematic uncertainties in this analysis.



# Chapter 11

## Final Results and Conclusions

In this work, we measured the branching fraction of the charmless semileptonic  $B$  meson decay  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  on a data sample corresponding to  $710 \text{ fb}^{-1}$  of integrated luminosity. We present the final result, shown in Chapter 9.3, with the statistical and systematic uncertainties determined in Chapter 10.

### 11.1 Signal Significance

In order to determine the signal significance, the profile likelihood function is obtained by performing the signal fit to the data, with a fixed signal yield. The yield is fixed to values in the range  $[0, 1000]$  and, for each fit, the maximum likelihood of the fit is extracted. The signal yield at the maximum of the profile likelihood corresponds to the optimal fit value. In order to incorporate systematic uncertainties in the profile likelihood, the latter is convoluted with a Gaussian function. The width of the Gaussian function, used in the convolution, corresponds to the systematic uncertainty of the signal yield. We exclude those sources of systematic uncertainty, listed in Table 10.2, which do not affect the yield, but only the calculation of the branching fraction. The total systematic uncertainty on the signal yield is

$$\sigma_{\text{sys}}^{\text{yield}} = {}^{+85}_{-71}. \quad (11.1)$$

The asymmetric systematic error is taken into account by using the negative error in the convolution for the left side of the profile likelihood, and the positive error for the right side. Figure 11.1 (left) shows the profile likelihood as a function of the signal yield, before and after incorporating the systematic uncertainties, while Figure 11.1 (right) shows the profile negative-log-likelihood, which is more commonly used. As mentioned in Eq. (8.3 – 8.4), the signal significance is calculated as the square-root value of  $\sqrt{-2 \log(\mathcal{L}/\mathcal{L}_{\text{max}})}$ , where  $\mathcal{L}$  is the value of the likelihood at the signal yield, fixed to 0. In this measurement, the statistical significance of the signal yield is equal to about  $6.3\sigma$ , while the total significance amounts to about  $4.6\sigma$ .

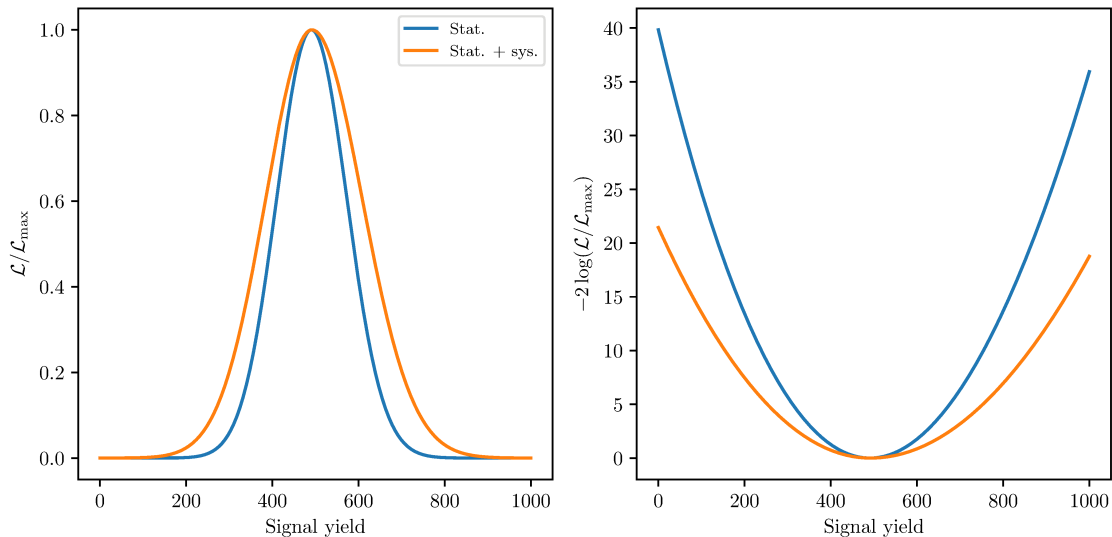


Figure 11.1: The profile likelihood function (left) and the profile negative-log-likelihood function (right) for the case with statistical error only (blue) and convoluted with a Gaussian function in order to incorporate the systematic uncertainties on the signal yield (orange).

## 11.2 Branching Fraction

This work presents the first measurement of the  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  decay. Including the total systematic uncertainties from Table 10.2, the final result for the signal branching fraction is

$$\mathcal{B}(B^+ \rightarrow K^+ K^- \ell^+ \nu) = (3.04 \pm 0.51 \pm {}^{+0.67}_{-0.66}) \times 10^{-5}, \quad (11.2)$$

where the first and the second uncertainty are statistical and systematic, respectively. With a signal significance of  $4.6\sigma$ , this measurement represents the first evidence for the  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  decay.



# Poglavje 12

## Povzetek doktorskega dela

### 12.1 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 med njimi. 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  $\bar{\phantom{x}}$ ), 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  $q_1 q_2 q_3$  (hadroni) ali pa  $q_1 \bar{q}_2$  (mezoni), kamor med prve uvrščamo tudi protone in nevtrone. Poleg omenjenih dolgoživečih delcev pa obstajajo tudi težji, manj stabilni delci, ki preko zgoraj naštetih interakcij razpadajo v lažje in stabilnejše. Raziskovanje takšnih procesov s pomočjo pospeševalnikov in trkalnikov nam danes omogoča spoznavanje zakonov vesolja vse do njegovega začetka.

Osrednji del doktorske disertacije predstavljajo meritve razpadov mezonov  $B$ , t.j. delcev, ki so sestavljeni iz težkega kvarka  $b$  in enega od lahkkih 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 razpadajo preko velike množice šibkih razpadov.

Edinstvena lastnost šibke interakcije je, da lahko spreminja tip oziroma kvarkov, medtem ko ga ostale interakcije ohranjajo. Takšni procesi so opisani s prehodno matriko CKM (Cabibbo-Kobayashi-Maskawa) [5, 6]

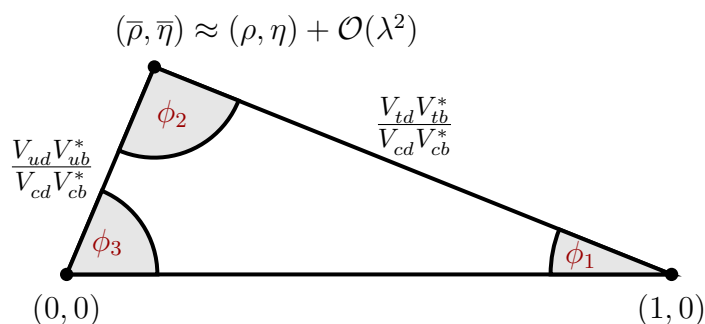
$$V_{CKM} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix}. \quad (12.1)$$

Unitarnost matrike CKM nam omogoča, da iz nje izluščimo matematične identitete,

od katerih je ena pomembnejših

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \quad (12.2)$$

poznana pod imenom *unitarni trikotnik*, saj predstavlja zaključen vektor treh točk v kompleksni ravnini, kot prikazuje Slika 12.1. 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 enačba ne bi opisala trikotnika, bi to nakazovalo na potencialne nove procese, ki jih še ne poznamo, in jih kolektivno imenujemo "nova fizika".



Slika 12.1: Unitarni trikotnik, prikazan v Wolfensteinovi parametrizaciji [7].

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

$$B^+ \rightarrow X_u^0 \ell^+ \nu_\ell, \quad (12.3)$$

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

$$d\Gamma \propto G_F^2 |V_{ub}|^2 |L^\mu \langle X_u | \bar{u} \gamma_u \frac{1}{2} (1 - \gamma_5) b | B \rangle|^2, \quad (12.4)$$

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  $b \rightarrow u$ .

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  $B \rightarrow \pi \ell \nu$ , pri drugi metodi pa opravljamo meritev v skupno končno stanje oblike  $B \rightarrow X_u \ell \nu$ . Obe metodi potekata preko različnih pristopov in se soočata z različnimi težavami, kar pomeni, da sta oba končna rezultata v večji meri neodvisna. Izmerjeni vrednosti imata zelo podobno natančnost, medtem ko se srednja vrednost le deloma ujema. Rezultata se razlikujeta s signifikanco  $3\sigma$ , kar predstavlja večjo težavo znotraj

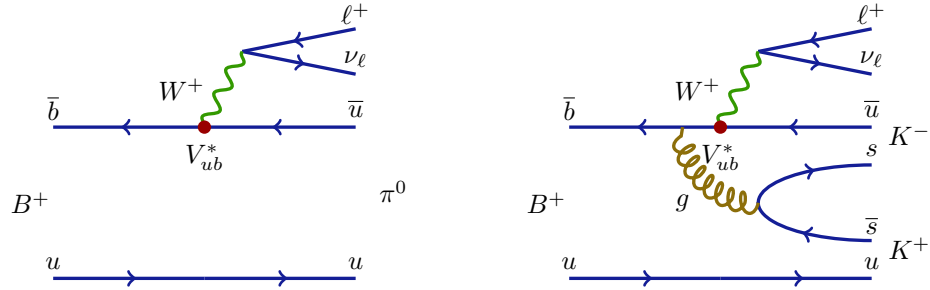
področja. Trenutni svetovni povprečji [8] ekskluzivne (iz razpadov  $B^0 \rightarrow \pi^- \ell^+ \nu$ ) in inkluzivne meritve (GGOU kolaboracija [9]) sta

$$|V_{ub}|_e = (3,65 \pm 0,09 \pm 0,11) \times 10^{-3}, \quad (12.5)$$

$$|V_{ub}|_i^{\text{GGOU}} = (4,52 \pm 0,15 \begin{smallmatrix} +0,11 \\ -0,14 \end{smallmatrix}) \times 10^{-3}, \quad (12.6)$$

kjer prva in druga napaka predstavljata eksperimentalno in teoretsko negotovost. Rezultati inkluzivnih meritev so praviloma večjih vrednosti 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, konkretnije za razpad  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$ , ki je strukturno precej podoben razpadu  $B \rightarrow \pi \ell \nu$ , za razliko produkcije para kvarkov  $s\bar{s}$ , ki se potem hadronizira v nove delce, kot prikazuje Slika 12.2. V inkluzivnih meritvah ne-čarobnih semileptonskih 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  $b \rightarrow c \rightarrow s$ , ki pa jih hočemo v analizah prehodov  $b \rightarrow u$  zatreti. V primeru te analize imamo v končnem stanju 2 kaona pri prehodu  $b \rightarrow u$ , kar pomeni, da takšni razpadi niso upoštevani v inkluzivnih meritvah, čeprav bi morali biti. Cilj študije je določiti pogostost razpadov  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  s prehodom  $b \rightarrow u$  in s tem oceniti, kakšen potencialen efekt ima lahko neupoštevanje teh razpadov na inkluzivno meritev elementa  $V_{ub}$ . V nadaljevanju bo razpad  $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$  zaradi enostavnosti zapisan kot  $B \rightarrow KK\ell\nu$ .



Slika 12.2: Feynmanovi diagrami za razpada  $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$  (levo) in  $B^+ \rightarrow K^- K^+ \ell^+ \nu_\ell$  (desno).

## 12.2 Experimentalna postavitve

Podatki, uporabljeni v tej analizi, so bili ustvarjeni pri trkih elektronov  $e^-$  in pozitronov  $e^+$  v pospeševalniku KEKB ter 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  $b\bar{b}$ ). Podrobnejši opis pospeševalnika in detektorja se nahaja v literaturah [13] in [14].

### 12.2.1 Trkalnik KEKB

Trkalnik KEKB je asimetričen trkalnik delcev  $e^+e^-$ , ki potujejo po obročih s premerom 3 km v grućah. V središču detektorja grući elektronov z energijo 8 GeV in pozitronov z energijo 3,5 GeV trćita pod kotom 22 mrad. Skupna invariantna masa trka ustreza masi resonance  $\Upsilon(4S)$

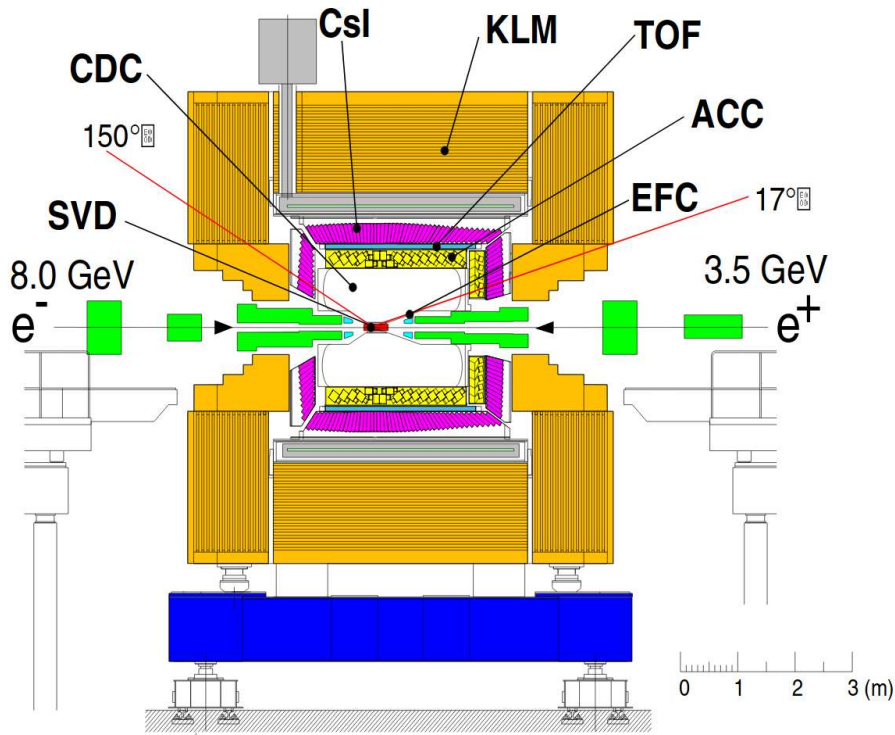
$$E_{CM} = \sqrt{2E_{e^+}E_{e^-}} = m_{\Upsilon(4S)}c^2 \approx 10,58 \text{ GeV}. \quad (12.7)$$

DeleŇ mezonov  $\Upsilon(4S)$  razpade preko zelo ćistega kanala v dva praktićno mirujoća mezona  $B$  v teŇiŇnem sistemu, kar v tej in v podobnih analizah pogosto izkoriŇćamo, saj je zaćetno stanje dobro poznano.

Trkalnik je v ćasu obratovanja zajel kolićino podatkov, ki ustreza integrirani luminoznosti  $1041 \text{ fb}^{-1}$ , od katere okoli  $711 \text{ fb}^{-1}$  predstavlja podatke, zajete pri energiji 10,58 GeV, t.j. masi resonance  $\Upsilon(4S)$ . Slednja vrednost integrirane luminoznosti ustreza Ňtevilu  $771 \times 10^6$  parov  $B\bar{B}$  mezonov.

### 12.2.2 Detektor Belle

Detektor Belle je magnetni masni spektrometer, ki pokriva većji 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 doloćiti energijo in gibalno kolićino delcev, kar doseŇemo preko detektorskih podsistemov, ki so okoli interacijske toćke postavljeni v plasteh, kot je prikazano na Sliki 12.3. Detektor pokriva polarni kot med  $17^\circ \leq \theta \leq 150^\circ$ , medtem ko je azimutni kot pokrit v celoti, kar skupaj predstavlja 92% pokritost polnega prostorskega kota.



Slika 12.3: Shematski prikaz detektorja Belle in ustreznih podsistemov [14].

## Silicijev detektor verteksov

Silicijev detektor verteksov je postavljen najbližje interakcijski točki. Sestavljen je iz dvostranskih silicijevih detektorjev, ki podajajo dvodimenzionalno informacijo o prehodnih nabitih delcev z natančnostjo okoli  $100 \mu\text{m}$ . To nam omogoča določitev točk razpada (verteksov) kratkoživečih delcev.

## Osrednja potovalna komora

Osrednja potovalna komora je sestavljena iz mnogih žic, napeljanih skozi skrbno izbrano 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  $0,8 \text{ GeV}/c$  komora 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  $0,8 \text{ GeV}/c < p < 1,2 \text{ GeV}/c$ , še posebej kaonov  $K^\pm$  in pionov  $\pi^\pm$ . Pri isti gibalni količini zaradi različnih mas delcev dobimo različne čase preleta, kar lahko uporabimo za določitev njihove mase. Časovna resolucijo tega podsistema ima zgornjo mejo  $100 \text{ ps}$ .

## Pragovni števec sevanja Čerenkova

Števec sevanja Čerenkova se prav tako uporablja za identifikacijo delcev, deluje pa v višjih območjih gibalne količine  $1,0 \text{ GeV}/c < 4,0 \text{ GeV}/c$ , kjer učinkovitost merilca časa preleta ni več zadostna. Silikatni aerogel, ki z dobro določenim lomnim količnikom predstavlja osrednjo strukturo podsistema, seva svetlobo Čerenkova, 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. Karakteristično so to elektroni in fotoni. Z njim lahko izmerimo pozicijo in energijo delca, ko le-ta zadane kalorimeter. Ko elektroni ali fotoni zadenejo 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  $0,6 \text{ GeV}/c$ . Ti delci so visokopenetrirajoči, saj lahko preletijo vse do sedaj opisane podsisteme. Prvi so nevtralni in jih lahko določimo preko hadronske interakcije v detektorju in preko

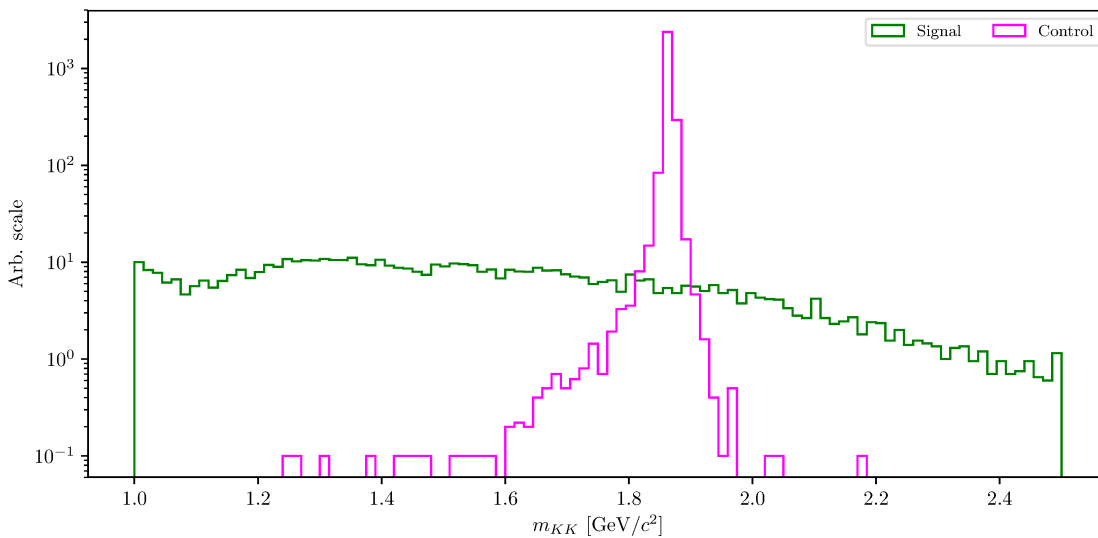
manjkajoče nabite sledi, medtem ko so drugi nabiti in jih identificiramo na pogladi ionizacije.

## 12.3 Analizni postopek

Analizni postopek je določen na podlagi simuliranih podatkov, oziroma Monte Carlo (MC) simulacije. Ta nam omogoča, da na podlagi teoretičnega modela razpadov dobro opišemo realnost, dodatno pa nam je na voljo "resnica", kot na primer generirane lastnosti delcev in njihova identiteta, ki je bila določena pri generaciji podatkov.

Za pripravo analiznega postopka imamo na voljo  $6 - 10 \times$  več podatkov kot jih je izmerjenih, s čimer povečamo natančnost analiznih korakov in zmanjšamo možnost statističnih fluktuacij.

Poleg signalnega razpada  $B^+ \rightarrow K^+K^-\ell^+\nu_\ell$  v študiji rekonstruiramo tudi kontrolni razpad  $B^+ \rightarrow \bar{D}^0\ell^+\nu$ ,  $\bar{D}^0 \rightarrow K^+K^-$ . Drugi ima enako končno stanje kot prvi, le da poteka preko prehoda kvarkov  $b \rightarrow c$  in ne  $b \rightarrow u$ , kot pri signalnem razpadu. Uspešno jih lahko ločimo preko invariantne mase dveh kaonov, ki je v primeru kontrolnega razpada zelo omejena na območje okoli mase mezona  $D^0$ , v primeru signalnega razpada pa je razporejena po celotnem območju, kot prikazuje Slika 12.4.



Slika 12.4: Porazdelitev mase dveh kaonov ( $m_{KK}$ ) za signalni in kontrolni razpad. Porazdelitev  $m_{KK}$  kontrolnega razpada je prisotna samo v območju mase mezona  $D^0$ , medtem ko je porazdelitev  $m_{KK}$  signalnega razpada prisotna po širšem območju.

### 12.3.1 Rekonstrukcija razpada

Postopek rekonstrukcije se prične z izbiro dolgoživečih stabilnih delcev, ki so v našem 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 šibke interakcije, zato jih s takšnim detektorjem ne moremo opaziti, kar predstavlja manjkajočo energijo in gibalno količino v dogodku trka  $e^+e^-$ .

Selekcija poteka na podlagi rezov spremenljivk, kjer je izbrano območje določeno na podlagi optimizacije metrike  $FOM$  (ang. *figure of merit*), definirane kot

$$FOM = \frac{N_S}{\sqrt{N_S + N_O}}, \quad (12.8)$$

kjer  $N_S$  predstavlja število pravilno rekonstruiranih kandidatov (signal),  $N_O$  pa število nepravilno rekonstruiranih kandidatov.

Povzeta selekcija dolgo-živečih stabilnih delcev je

- elektroni:  $|d_0| < 0,1 \text{ cm}$ ,  $|z_0| < 1,5 \text{ cm}$ ,  $p_{CMS} \in [0,4, 2,6] \text{ GeV}/c$ ,  $PID_e > 0,9$ ,
- mioni:  $|d_0| < 0,1 \text{ cm}$ ,  $|z_0| < 1,5 \text{ cm}$ ,  $p_{CMS} \in [0,6, 2,6] \text{ GeV}/c$ ,  $PID_\mu > 0,97$ ,
- kaoni:  $|d_0| < 0,15 \text{ cm}$ ,  $|z_0| < 1,5 \text{ cm}$ ,  $p_{CMS} \in [0,0, 2,5] \text{ GeV}/c$ ,  $PID_{K/\pi} > 0,6$ ,

kjer  $d_0$  in  $z_0$  predstavljata vpadne parametre nabitih delcev,  $p_{CMS}$  gibalno količino v težiščnem 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  $Y = KKe$  in  $KK\mu$ , ki služijo kot kandidati mezonov  $B$ , z izjemo manjkajočih nevtrinov. Na podlagi dejstva, da je detektor Belle hermetično zaprt in pokriva večino prostorskega kota ter da dobro poznamo začetno stanje  $\Upsilon(4S)$ , lahko določimo četverec manjkajoče (ang. *missing*) gibalne količine kot

$$P_{miss} = P_{\Upsilon(4S)} - \sum_i^{\text{Dogodek}} (E_i, \vec{p}_i), \quad (12.9)$$

$$P_{miss} = P_{\Upsilon(4S)} - \left( P_Y - \sum_i^{\text{ROE}} (E_i, \vec{p}_i) \right), \quad (12.10)$$

kjer  $p$  predstavlja četverec gibalne količine, indeks  $i$  teče po vseh delcih znotraj množice, ROE (ang. *rest of event*) pa predstavlja podmnožico celotnega dogodka trka  $e^+e^-$ , ki vsebuje vse delce, ki niso bili uporabljeni v rekonstrukciji kandidata  $Y$ .

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

- mezoni  $B$ :
  - $P(\chi^2, NDF) > 6,0 \times 10^{-3}$ ,
  - $|\cos \theta_{BY}| < 1,05$ ,
  - $m_{miss}^2 < 0,975 \text{ GeV}/c^2$ ,
  - $5,1 \text{ GeV}/c^2 < M_{BC} < 5,295 \text{ GeV}/c^2$ ,
  - $-1,0 \text{ GeV} < \Delta E < 1,3 \text{ GeV}$ ,

kjer  $P(\chi^2, NDF)$  predstavlja kvaliteto rekonstrukcije verteksa mezona  $B$ ,  $m_{miss}^2$  pa invariantno maso četverca manjkajoče gibalne količine v dogodku. Ostali izrazi za

$\cos \theta_{BY}$ ,  $M_{BC}$  in  $\Delta E$  so

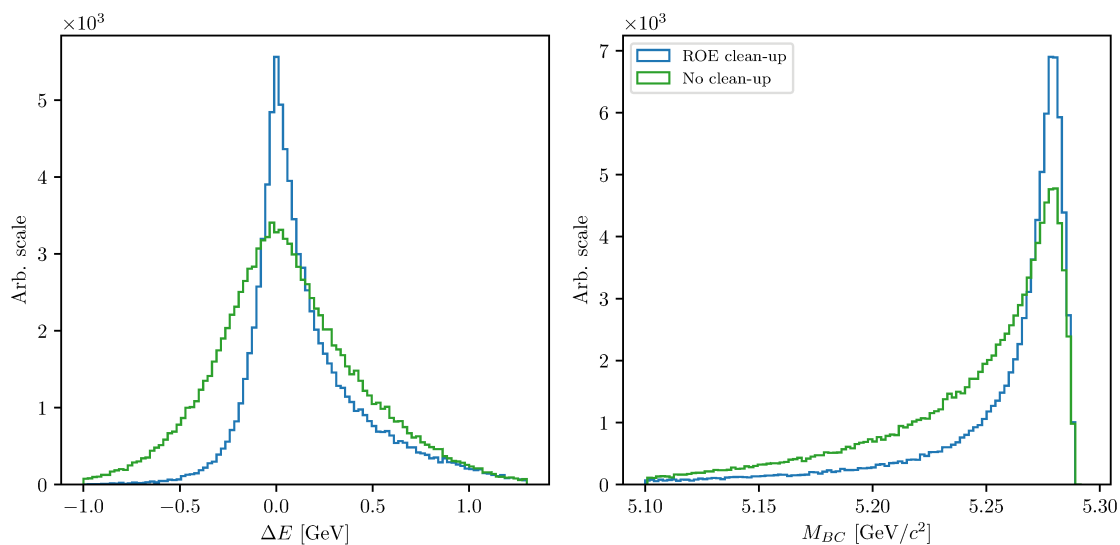
$$\cos(\theta_{BY}) = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \quad (12.11)$$

$$M_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}_B|^2}, \quad (12.12)$$

$$\Delta E = E_B - E_{CMS}/2 \quad (12.13)$$

in po vrsti predstavljajo kot med nominalnim ( $B$ ) in rekonstruiranim ( $Y$ ) mezonom  $B$ , maso, vezano na energijo žarka v težiščnem koordinatnem sistemu, in razliko energije kandidata in polovice težiščne energije  $E_{CMS}$ . Za pravilne kombinacije mezonov  $B$  se porazdelitev po  $\cos(\theta_{BY})$  nahaja na intervalu  $[-1, 1]$ , porazdelitvi  $M_{BC}$  in  $\Delta E$  pa imata vrh okoli  $m_B$  in  $0$  GeV, kjer  $m_B$  predstavlja nominalno maso mezona  $B$ .

Potrebno je omeniti, da so v posameznem dogodku lahko prisotni tako nevtralni kot nabiti delci, ki ne prihajajo neposredno iz trka, temveč so lahko bodisi produkti sekundarnih interakcij v detektorju, bodisi delci, ki izhajajo iz ozadja na račun raznih interakcij žarkov med potovanjem po obročih pospeševalnika. Takšne delce je potrebno odstraniti iz En. (12.10), čemur pravimo čiščenje dogodka. V tej analizi je bilo opravljeno temeljito čiščenje tako nevtralnih kot nabitih delcev, ki je terjalo različne pristope. V ta namen so bile uporabljene metode strojnega učenja za prepoznavanje takšnih neželenih delcev. Slika 12.5 prikazuje primerjavo med očiščenim in neočiščenim dogodkom, kjer primerjamo porazdelitvi  $\Delta E$  in  $M_{BC}$ .



Slika 12.5: Primerjava očiščenega in neočiščenega dogodka za porazdelitvi  $\Delta E$  in  $M_{BC}$ . Porazdelitve iz očiščenega dogodka so bolj ostre in predstavljajo boljše možnost za ločitev od ozadja.

### 12.3.2 Odstranjevanje ozadja

Ozadje v takšni analizi predstavljajo napačne kombinacije razpadne verige signalnega kandidata. Napačno kombinacijo lahko predstavlja napačna kombinatorika ali pa primer končnega stanja drugih razpadnih kanalov, ki posnema končno stanje signalnega razpada. Takšne kombinacije v splošnem nimajo enakih lastnosti



kot signalne, zato skušamo najti načine, kako takšno ozadje odstraniti na najbolj optimalen način.

Odstranjevanja ozadja se lotimo v treh korakih, v prvem koraku uporabimo enostavne reze na invariantni masi kaonskega para, saj pričakujemo, da veliko parov  $KK$  pride iz resonancam podobnih struktur, kot na primer  $\phi \rightarrow KK$  ali  $D^0 \rightarrow KK$ , kjer za slednjo že vemo, da je prisotna v kontrolnem razpadu. Prav tako se lahko zgodi, da je eden od pionov napačno identificiran kot kaon in tako dobimo vrh porazdelitve, ki je zamaknjen za razliko mas masnih hipotez. Rezi, ki jih uporabimo za odstranjevanje omenjenih kandidatov, so

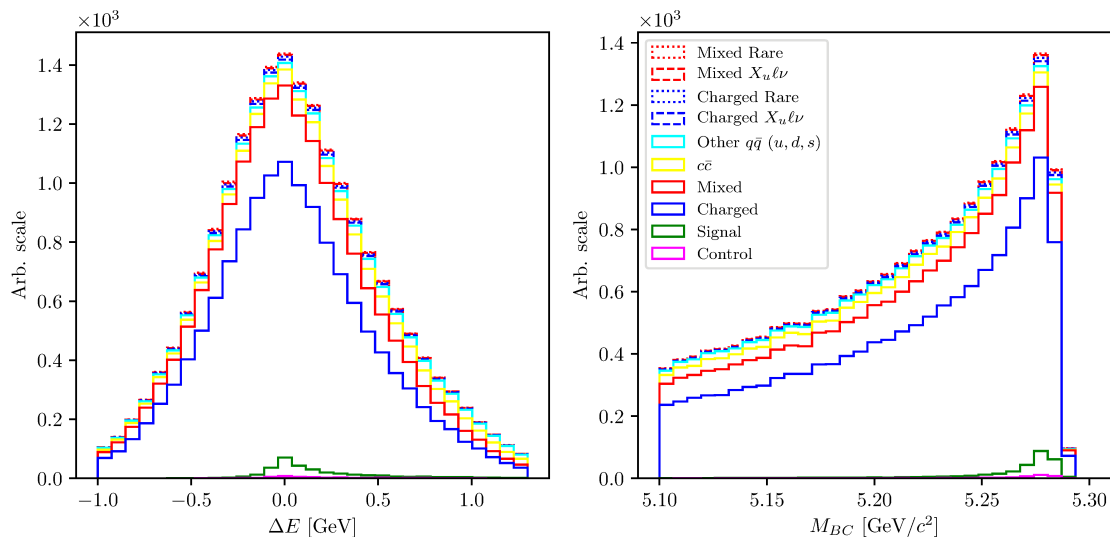
- signalni rez:  $|m_{KK} - m_\phi| > \Delta_\phi$ ,  $|m_{KK} - m_{D^0}| > \Delta_{D^0}$ ,  $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$ ,
- kontrolni rez:  $|m_{KK} - m_{D^0}| < \Delta_{D^0}$ ,  $|m_{K\pi} - m_{D^0}| > \Delta_{D^0}$ ,

kjer  $m_{KK}$  predstavlja invariantno maso kaonskega para  $KK$ ,  $m_{K\pi}$  pa invariantno maso kaonskega para  $K\pi$ , kjer je bila masa kaona, katerega naboj je nasproten naboju  $B$  mezona, zamenjana z maso delca  $\pi$ . Ostali parametri so  $m_\phi \approx 1.019 \text{ GeV}/c^2$ ,  $m_{D^0} \approx 1.864 \text{ GeV}/c^2$ ,  $\Delta_\phi \approx 8 \times 10^{-3} \text{ GeV}/c^2$  in  $\Delta_{D^0} \approx 1.5 \times 10^{-2} \text{ GeV}/c^2$ . V primeru študije kontrolnega razpada se osredotočimo na ozko okno okoli mase mezona  $D^0$ .

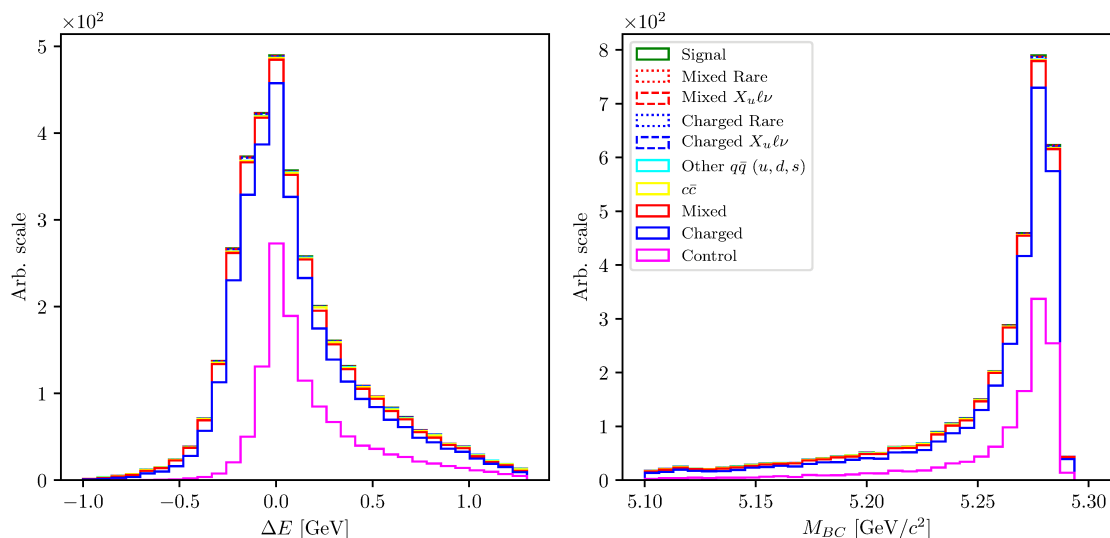
V drugem koraku se lotimo odstranjevanja t.i. kontinuumskega ozadja, kjer kandidati prihajajo iz procesov  $e^+e^- \rightarrow q\bar{q}$ ,  $q \in [u, d, s, c]$ . Poslužimo se metod strojnega učenja, ki prepoznajo kandidate iz kontinuumskih procesov od signalnih kandidatov. Za ta namen potrebujemo spremenljivke,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

V tretjem koraku se na podoben način lotimo odstranjevanja ostalih kandidatov iz procesov  $e^+e^- \rightarrow B\bar{B}$ , za kar uporabimo vse ostale lastnosti kandidatov, razen  $\Delta E$  in  $M_{BC}$ , ker le-te potrebujemo za luščenje števila signalnih kandidatov. Pri odstranjevanju ozadja te vrste uporabimo posebno metodo strojnega učenja, ki ohranja obliko porazdelitve spremenljivke  $M_{BC}$  za ozadje, kar preprečuje, da bi optimizacija preoblikovala obliko porazdelitve ozadja v tisto od signala.

Kot v prejšnjih optimizacijah optimiziramo metriko  $FOM$  za odstranjevanje ozadja v drugem in tretjem koraku. Končni vzorec za signalni razpad je prikazan na Sliki 12.6, za kontrolni razpad pa na Sliki 12.7.



Slika 12.6: Porazdelitvi spremenljivk  $\Delta E$  in  $M_{BC}$  za končni signalni vzorec.



Slika 12.7: Porazdelitvi spremenljivk  $\Delta E$  in  $M_{BC}$  za končni kontrolni vzorec.

### 12.3.3 Luščenje fizikalnih parametrov

Po selekciji končnega vzorca lahko začnemo luščiti fizikalne parametre iz podatkov. Za to uporabimo orodje RooFit [27], ki nam omogoča, da teoretični model prilagodimo izmerjenim podatkom, in na tak način določimo fizikalne parametre, ki jih iščemo. Na podlagi MC vzorca določimo porazdelitve za kandidate signalne kategorije in večih tipov ozadja. Te porazdelitve služijo kot predloge, ki jih ustrezno seštejemo skupaj, da dobimo teoretičen model, ki dobro opiše podatke. Vsaka predloga je predstavljena kot 2-dimenzionalen histogram v spremenljivkah  $\Delta E$  in  $M_{BC}$ , in sicer v  $19 \times 19$  razredih v območju, definiranim v Sekciji 12.3.1. Predloge posameznih kategorij so si med seboj različne, kar programu omogoča, da z visoko

verjetnostjo pravilno določi prispevke posameznih komponent.

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

### Kontrolni razpad

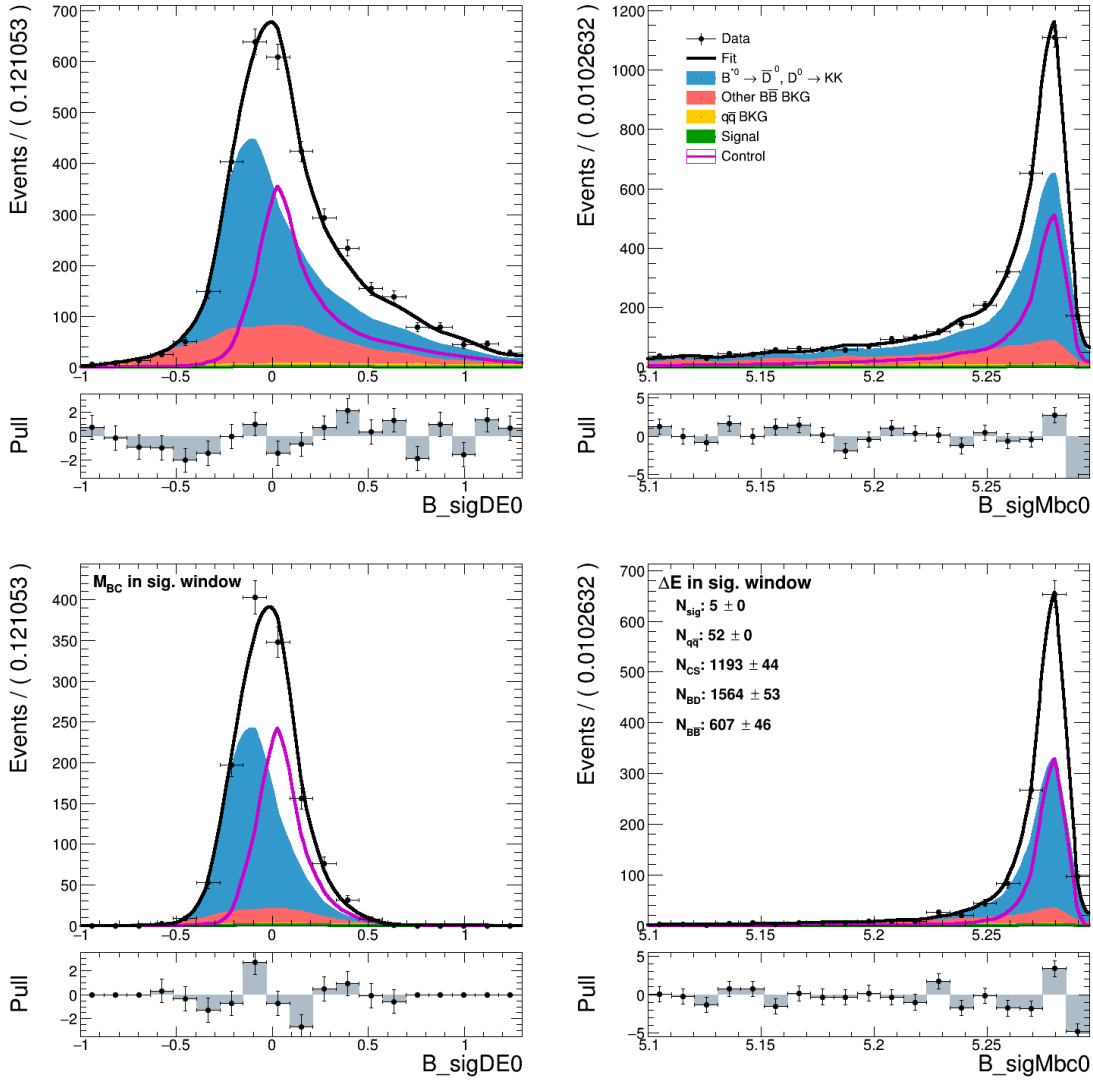
Fizikalne parametre v primeru kontrolnega razpada izluščimo v ozkem oknu okoli mase mezona  $D^0$ . Pri postopku luščenja uporabimo naslednje predloge

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

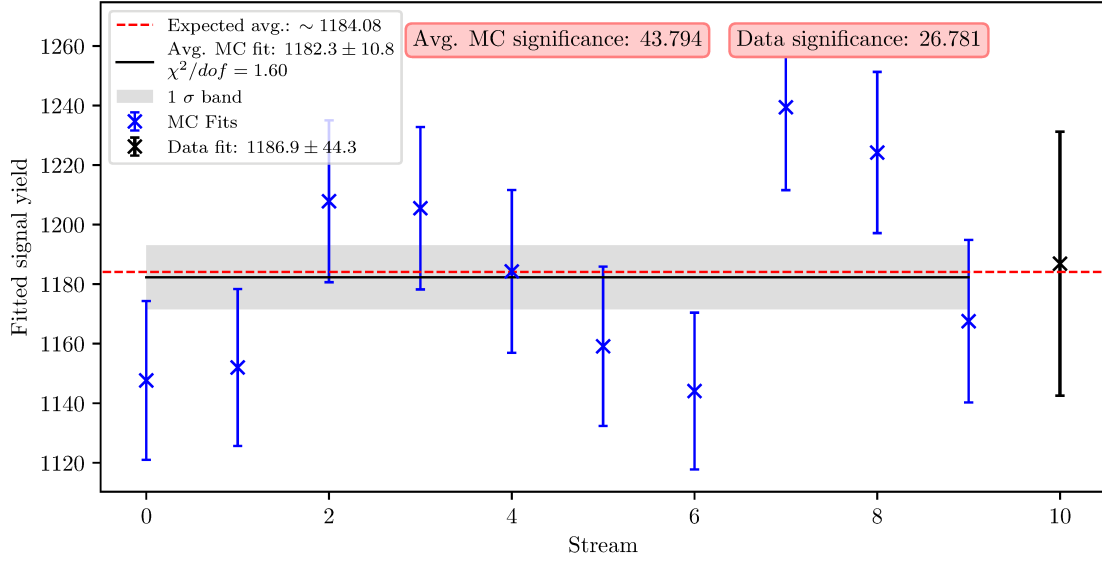
Slika 12.8 prikazuje primer prilagajanja predlog izmerjenim podatkom, Slika 12.9 pa rezultate luščenja na vseh MC podatkih in na pravih podatkih. Število kandidatov kontrolnega razpada na podlagi luščenja je

	$N^{MC}$	$N^{\text{podatki}}$
$\ell = e$ ali $\mu$	$1182 \pm 11$	$1187 \pm 44$
$\ell = e$	$591 \pm 8$	$583 \pm 28$
$\ell = \mu$	$592 \pm 7$	$613 \pm 30$

Tabela 12.1: Rezultati luščenja števila kontrolnih kandidatov za različna končna leptonska stanja.



Slika 12.8: Primer luščenja števila kontrolnih kandidatov na pravih podatkih. Lev stolpec prikazuje  $\Delta E$ , desni pa  $M_{BC}$ , medtem ko zgornja vrstica prikazuje porazdelitvi na celotnem definiranim območju, spodnja pa projekcija na ozko okno okoli vrha signalne porazdelitve.

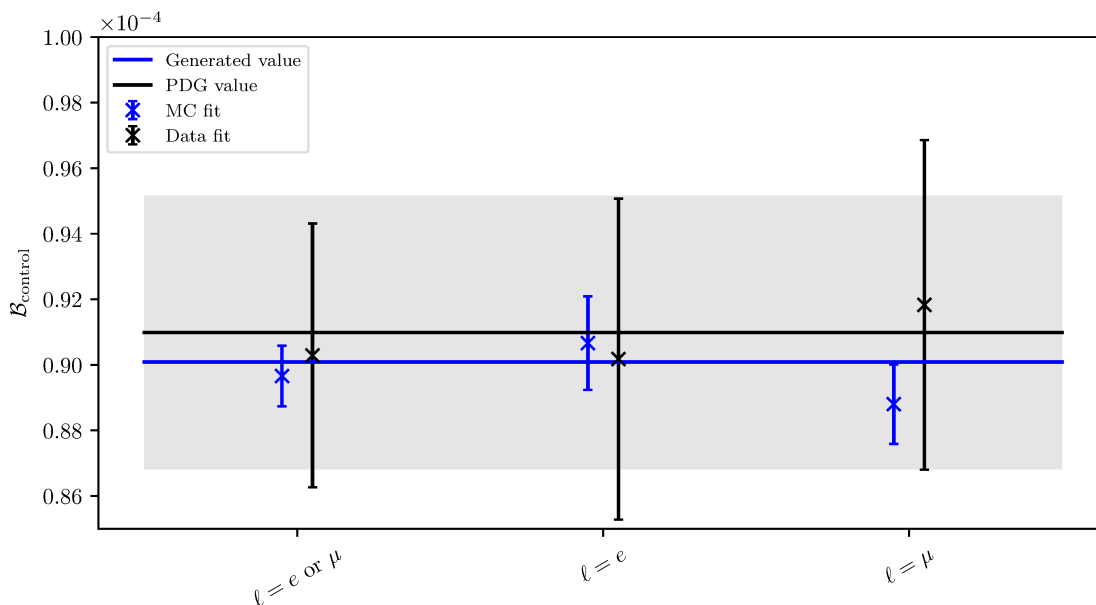


Slika 12.9: Število kontrolnih kandidatov za vseh 10 vzorcev MC podatkov in njihovo uteženo povprečje, ter za izmerjene podatke.

Pri luščenju parametrov na pravih podatkih smo uporabili dodatno informacijo, in sicer meritev razpada  $B \rightarrow D^* \ell \nu$ ,  $D^0 \rightarrow K^+ K^-$  [8, 29], ki smo jo uporabili v obliki omejitve vrednosti razmerja števila kandidatov omenjenega ter kontrolnega razpada. Na podlagi končnega števila kandidatov kontrolnega razpada  $N$  lahko določimo tudi razvejitevno razmerje, ki je določeno kot

$$\mathcal{B} = \frac{N \times \epsilon_{MC} \times \rho_{PID}}{2N_{B\bar{B}}}, \quad (12.14)$$

kjer  $\epsilon_{MC}$  predstavlja izkoristek kontrolnega razpada, določenega na MC vzorcu,  $\rho_{PID}$  je korekcijski faktor na račun razlike identifikacij nabitih delcev na MC in na pravih podatkih,  $N_{B\bar{B}}$  pa je izmerjeno število generiranih parov mezonov  $B$ . Razvejitevno razmerje lahko določimo tako na podatkih kot na MC vzorcu, rezultati obeh pa so prikazani na Sliki 12.10. Rezultati so konsistentni s pričakovanimi in izmerjenimi vrednostmi, kar potrjuje zanesljivost naše analize.



Slika 12.10: Rezultati meritev razvejitenih razmerij kontrolnega razpada za MC in za podatke za različna končna leptonska stanja.

## Signalni razpad

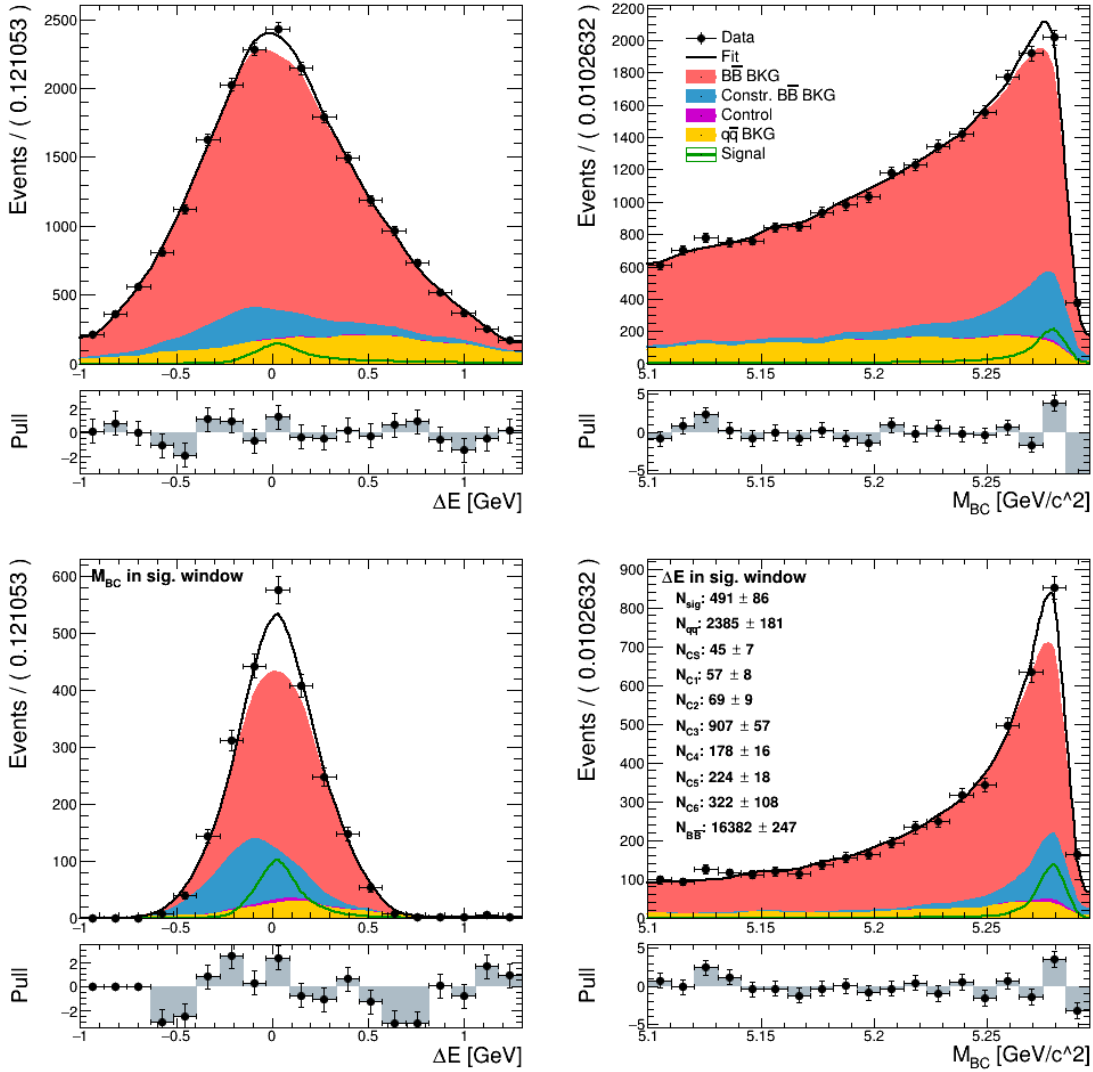
Rezultati prilagajanja MC in pravih podatkov za kontrolni razpad potrjujejo konsistentnost našega analiznega postopka, tako da jih lahko ponovimo še na signalnem razpadu. V tem primeru smo uporabili naslednje predloge

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

V primeru dobro poznanih razpadov zopet uporabimo informacije o najnovejših meritvah in jih uporabimo za omejitev števila kandidatov posamezne kategorije. Pri prilagajanju predlog pravih podatkom smo uporabili kontrolni razpad za kalibracijo števila parov mezonov  $B$ . Slika 12.11 prikazuje primer prilagajanja predlog izmerjenim podatkom, Slika 12.12 pa prikazuje rezultate luščenja na vseh MC in na pravih podatkih skupaj. Števila kandidatov signalnega razpada in ostalih prispevkov na podlagi luščenja so

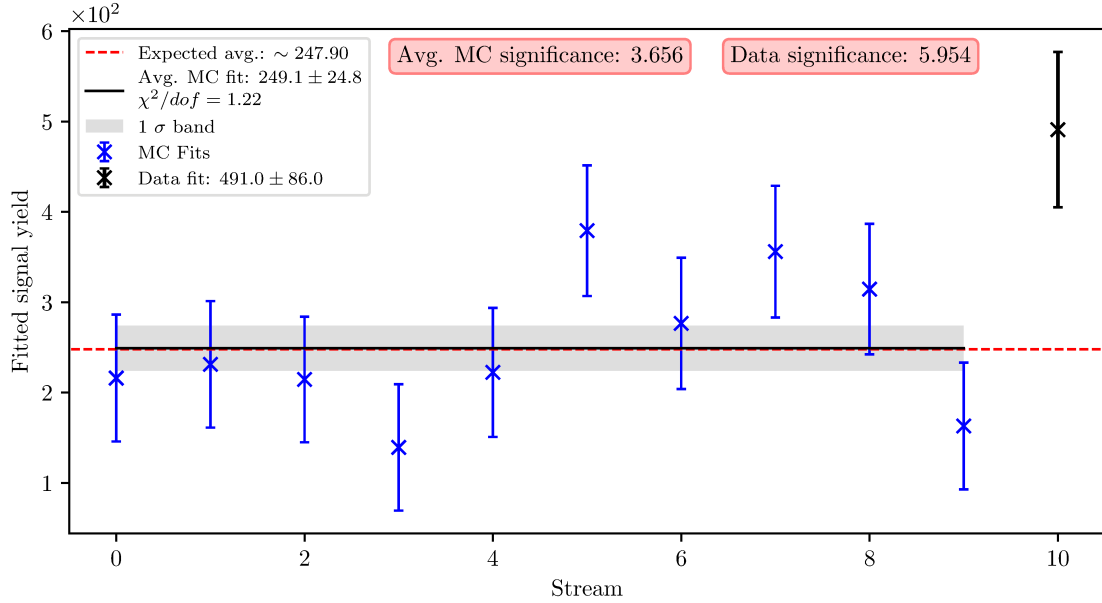
Kategorija	Število kandidatov
Signal	$491 \pm 86$
$q\bar{q}$ ozadje	$2385 \pm 181$
$C_0$	$45 \pm 7$
$C_1$	$57 \pm 8$
$C_2$	$69 \pm 9$
$C_3$	$907 \pm 57$
$C_4$	$178 \pm 16$
$C_5$	$224 \pm 18$
$C_6$	$322 \pm 108$
Ostalo $B\bar{B}$ ozadje	$16382 \pm 247$

Tabela 12.2: Števila kandidatov vseh prispevkov, določenih z luščenjem na signalnem vzorcu.



Slika 12.11: Primer luščenja števila signalnih kandidatov na podatkih. Lev stolpec prikazuje  $\Delta E$ , desni pa  $M_{BC}$ , medtem ko zgornja vrstica prikazuje porazdelitvi na celotnem definiranim območju, spodnja pa projekcija na ozko okno okoli vrha signalne porazdelitve.





Slika 12.12: Število kontrolnih kandidatov za vseh 10 vzorcev MC podatkov in njihovo uteženo povprečje, ter za izmerjene podatke.

Tabela 12.3 prikazuje vrednosti razvejitenega razmerja za signalni razpad na MC in na pravih podatkih. V vseh primerih je rezultat prikazan le s statistično napako.

	$\mathcal{B}_{GEN} [\times 10^{-5}]$	$\mathcal{B}^{MC} [\times 10^{-5}]$	$\mathcal{B}^{\text{podatki}} [\times 10^{-5}]$
$\ell = e$ ali $\mu$	1.57	$1.55 \pm 0.15$	$3.04 \pm 0.51$

Tabela 12.3: Vrednost razvejitenega razmerja signalnega razpada za MC in za podatke.

## 12.4 Sistematske negotovosti

Sistematske negotovosti vstopijo v analizo zaradi različnih razlogov, bodisi zaradi poznanih razlik med MC in med podatki, bodisi zaradi pomanjkljivosti pristopov v posameznih analiznih korakih. Nekatere negotovosti so splošne in vnaprej pripravljene za vse analize znotraj posameznih eksperimentov, medtem ko so druge specifične za vsako posamezno analizo in jih je potrebno temeljito preveriti.

### 12.4.1 Posamezni prispevki

#### Identifikacija delcev

Selekcija delcev na podlagi njihove metrike identifikacije (PID) se razlikuje na MC in na podatkih. Razlike so bile izračunane v za to namenjenih študijah znotraj kolaboracije. V tej analizi obravnavamo kaone, elektrone in mione, končna sistematska negotovost za obravnavan izvor pa je

$$\sigma_{\text{sis}}^{\text{PID}} = 10, \quad \delta_{\text{sis}}^{\text{PID}} = 2.0\%, \quad (12.15)$$

## Pristranskost postopka luščenja parametrov

Zanesljivost postopka luščenja parametrov je odvisna od kvalitete predlog posameznih kategorij, saj so si lahko nekatere predloge v nekaterih pogledih podobne in lahko na tak način pod- ali precenimo njihovo amplitudo (ang. *bias*). Na podlagi dveh različnih študij ocenimo, da je sistematska negotovost končne vrednosti na račun izbire postopka luščenja parametrov enaka

$$\sigma_{\text{sis}}^{\text{bias}} = \begin{matrix} +7 \\ -10 \end{matrix}, \quad \delta_{\text{sis}}^{\text{bias}} = \begin{matrix} +1.5\% \\ -2.0\% \end{matrix}. \quad (12.16)$$

## Omejitev dobro poznane ozadja

V postopku luščenja parametrov uporabimo informacijo o meritvah dobro poznanih razpadnih kanalov za omejitev števila kandidatov v obliki Gaussove porazdelitve. Te omejitve niso del statistične napake, ampak spadajo med prispevke sistematskih negotovosti. V tej študiji smo ta prispevek ločili in določili vrednost te sistematske negotovosti, ki je enaka

$$\sigma_{\text{sis}}^{\text{gaus}} = 26, \quad \delta_{\text{sis}}^{\text{GC}} = 5.3\%, \quad (12.17)$$

## Zamik in razširitev porazdelitve $\Delta E$

Da bi MC porazdelitve boljše približali tistim iz pravih podatkov, sta v tej analizi bila predstavljena dva nova parametra, ki na porazdelitev  $\Delta E$  delujeta kot zamik in razširitev. Centralna vrednost parametrov in intervali zanesljivosti so bili določeni kot

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

Sistematska negotovost končnega rezultata na podlagi izbire teh dveh parametrov je tako

$$\sigma_{\text{sis}}^{\text{raz.}} = \begin{matrix} +41 \\ -33 \end{matrix}, \quad \delta_{\text{sis}}^{\text{raz.}} = \begin{matrix} +8.3\% \\ -6.7\% \end{matrix}, \quad (12.18)$$

$$\sigma_{\text{sis}}^{\text{zam.}} = \begin{matrix} +41 \\ -31 \end{matrix}, \quad \delta_{\text{sis}}^{\text{zam.}} = \begin{matrix} +9.3\% \\ -8.0\% \end{matrix}. \quad (12.19)$$

## Vpliv velikosti MC vzorca

V analizi uporabimo porazdelitve na podlagi MC vzorca in na podlagi teh porazdelitev zgradimo cel analizni postopek. MC vzorec je končne velikosti in lahko statistično fluktuiira, kar lahko spremeni vrednost rezultatov. Na podlagi simulacij statističnih fluktuacij določimo prispevek sistematske negotovosti

$$\sigma_{\text{sis}}^{\text{MC}} = 26, \quad \delta_{\text{sis}}^{\text{MC}} = 2.4\%. \quad (12.20)$$

## Izkoristek multivariatnih postopkov

Kot rečeno, kontrolni razpad služi namenu, da preverimo analizne korake na MC in na izmerjenih podatkih, brez da bi pri tem tvegali pristranskost pri signalnem razpadu. Na tak način lahko preverimo tudi kompleksnejše postopke, ki vključujejo strojno učenje, tako da primerjamo razmerje števila kandidatov kontrolnega razpada

na MC in na podatkih za vsak posamezen korak aplikacije multivariatnih postopkov. Sistematska negotovost tega prispevka je določena kot standardna deviacija teh razmerij, s končno vrednostjo

$$\sigma_{\text{sis.}}^{\text{MVA}} = 5, \quad \sigma_{\text{sis.}}^{\text{MVA}} = 1.0\%. \quad (12.21)$$

### Negotovost signalnega modela

V tej analizi je za generacijo signalnih kandidatov bil uporabljen model ISGW2 [21], za katerega je znano, da se njegova napoved slabše ujema z meritvami. Na podlagi te zanesljivosti je postopek analize bil zastavljen čimbolj neodvisno od modela. Kvalitativno je to sistematsko negotovost na račun odvisnosti modela težko oceniti, zato so v ta namen bili uporabljeni trije dodatni modeli, katerih lastnosti so precej različne od pričakovanih, in tako služijo za oceno prispevka sistematske negotovosti. Prvi del prispevka prihaja na račun vpliva izbire modela na obliko  $\Delta E$  in  $M_{BC}$ . Z uporabo porazdelitev, pridobljenih z omenjenimi modeli, smo določili nove vrednosti izluščenih parametrov, njihovo razliko med glavno vrednostjo pa uporabili za oceno tega dela sistematske negotovosti, ki je

$$\sigma_{\text{sis}}^{\text{mod. obl.}} = \begin{matrix} +45 \\ -39 \end{matrix}, \quad \delta_{\text{sis}}^{\text{mod. obl.}} = \begin{matrix} +9.3\% \\ -8.0\% \end{matrix}. \quad (12.22)$$

Drugi del prispevka določimo na podlagi povprečnega izkoristka različnih modelov. Do različnih izkoristkov lahko pride zaradi tega, ker imajo različni modeli različne lastnosti razpada, kar vpliva na učinkovitost analiznega postopka. Na podlagi razlik izkoristkov različnih modelov v primerjavi z glavnim določimo oceno tega dela sistematske negotovosti, ki je

$$\sigma_{\text{sis}}^{\text{mod. izk.}} = \begin{matrix} +70 \\ -79 \end{matrix}, \quad \delta_{\text{sys}}^{\text{mod. izk.}} = \begin{matrix} +14.3\% \\ -16.2\% \end{matrix}. \quad (12.23)$$

### 12.4.2 Povzetek sistematskih negotovosti

Tabela 12.4 prikazuje posamezne prispevke sistematskih negotovosti in njihovo skupno vrednost, ki je bila uporabljena pri določitvi končnega rezultata.

Prispevek	$\sigma$	$\delta$ [%]
Identifikacija delcev	10	2
Pristranskost postopka	+7 -10	+1.5 -2.1
Omejitev poznanega ozadja	26	5.3
Zamik $\Delta E$	+41 -33	+8.3 -6.7
Razširitev $\Delta E$	+41 -31	+8.4 -6.3
Velikost MC vzorca	26	5.3
Izkoristek multiv. postopkov	5	1.0
Oblika signalnih modelov	+45 -39	+9.3 -8.0
Izkoristek signalnih modelov	+70 -79	+14.3 -16.2
Skupaj	+109 -107	+22.2 -21.9

Tabela 12.4: Povzetek sistematskih negotovosti te analize.

## 12.5 Končni rezultat in zaključek

Delo predstavlja prvo meritev razpada  $B^+ \rightarrow K^+K^-\ell^+\nu_\ell$ . Z upoštevanjem vseh sistematičnih negotovosti lahko določimo končni rezultat za razvejitevno razmerje razpada, ki znaša

$$\mathcal{B}(B^+ \rightarrow K^+K^-\ell^+\nu) = (3.04 \pm 0.51 \pm {}^{+0.67}_{-0.66}) \times 10^{-5}, \quad (12.24)$$

kjer prva napaka predstavlja statistično negotovost, druga pa sistematično. Statistična signifikanca signala v tem delu je enaka  $6.3\sigma$ , skupna signifikanca pa  $4.6\sigma$ , s čimer meritev pridobi status dokaza za signalni razpad  $B^+ \rightarrow K^+K^-\ell^+\nu_\ell$ .

# Bibliography

- [1] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL: <http://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [2] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [3] Abdus Salam. “Weak and electromagnetic interactions”. In: *Selected Papers Of Abdus Salam: (With Commentary)*. World Scientific, 1994, pp. 244–254.
- [4] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: [10.1103/PhysRevD.2.1285](https://doi.org/10.1103/PhysRevD.2.1285). URL: <https://link.aps.org/doi/10.1103/PhysRevD.2.1285>.
- [5] Nicola Cabibbo. “Unitary symmetry and leptonic decays”. In: *Physical Review L* 10.12 (1963), p. 531.
- [6] Makoto Kobayashi and Toshihide Maskawa. “CP-violation in the renormalizable theory of weak interaction”. In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657.
- [7] Lincoln Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. In: *Phys. Rev. Lett.* 51 (21 Nov. 1983), pp. 1945–1947. DOI: [10.1103/PhysRevLett.51.1945](https://doi.org/10.1103/PhysRevLett.51.1945). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.51.1945>.
- [8] Y. Amhis et al. “Averages of  $b$ -hadron,  $c$ -hadron, and  $\tau$ -lepton properties as of summer 2016”. In: *Eur. Phys. J. C* 77 (2017). updated results and plots available at <https://hflav.web.cern.ch>, p. 895. DOI: [10.1140/epjc/s10052-017-5058-4](https://doi.org/10.1140/epjc/s10052-017-5058-4). arXiv: [1612.07233](https://arxiv.org/abs/1612.07233) [hep-ex].
- [9] P. Gambino, P. Giordano, 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](https://doi.org/10.1088/1126-6708/2007/10/058). arXiv: [0707.2493](https://arxiv.org/abs/0707.2493) [hep-ph].
- [10] David J Lange. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1-2 (2001), pp. 152–155.
- [11] Sea Agostinelli, John Allison, et al. “GEANT4 — a simulation toolkit”. In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

- [12] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “PYTHIA 6.4 physics and manual”. In: *Journal of High Energy Physics* 2006.05 (2006), p. 026.
- [13] Tetsuo Abe, Kazunori Akai, et al. “Achievements of KEKB”. In: *Progress of Theoretical and Experimental Physics* 2013.3 (2013), 03A001. DOI: [10.1093/ptep/pts102](https://doi.org/10.1093/ptep/pts102). eprint: [/oupanbackfile/content\\_public/journal/ptep/2013/3/10.1093/ptep/pts102/2/pts102.pdf](http://oupanbackfile/content_public/journal/ptep/2013/3/10.1093/ptep/pts102/2/pts102.pdf). URL: <http://dx.doi.org/10.1093/ptep/pts102>.
- [14] A. Abashian, K. Gotow, 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](https://doi.org/10.1016/S0168-9002(01)02013-7). URL: <http://www.sciencedirect.com/science/article/pii/S0168900201020137>.
- [15] 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](http://superb.kek.jp/documents/loi/img/LoI_detector.pdf).
- [16] T. Kuhr, C. Pulvermacher, et al. “The Belle II Core Software”. In: (2018). arXiv: [1809.04299](https://arxiv.org/abs/1809.04299) [[physics.comp-ph](https://arxiv.org/abs/1809.04299)].
- [17] Moritz Gelb et al. “B2BII - Data conversion from Belle to Belle II”. In: (2018). arXiv: [1810.00019](https://arxiv.org/abs/1810.00019) [[hep-ex](https://arxiv.org/abs/1810.00019)].
- [18] Thomas Keck. “Machine learning algorithms for the Belle II experiment and their validation on Belle data”. Karlsruhe Institut für Technologie, Diss., 2017. Dr. Karlsruhe Institut für Technologie, 2017, 240 pages. URL: <https://ekp-invenio.physik.uni-karlsruhe.de/record/48940>.
- [19] D. M. Asner, K. W. Edwards, and others. “Study of exclusive charmless semileptonic  $B$  decays and extraction of  $|V_{ub}|$  at CLEO”. In: *Phys. Rev. D* 76 (1 July 2007), p. 012007. DOI: [10.1103/PhysRevD.76.012007](https://doi.org/10.1103/PhysRevD.76.012007). URL: <https://link.aps.org/doi/10.1103/PhysRevD.76.012007>.
- [20] H. Ha et al. “Measurement of the decay  $B^0 \rightarrow \pi^- \ell^+ \nu$  and determination of  $|V_{ub}|$ ”. In: *Phys. Rev. D* 83 (2011), p. 071101. DOI: [10.1103/PhysRevD.83.071101](https://doi.org/10.1103/PhysRevD.83.071101). arXiv: [1012.0090](https://arxiv.org/abs/1012.0090) [[hep-ex](https://arxiv.org/abs/1012.0090)].
- [21] Daryl Scora and Nathan Isgur. “Semileptonic meson decays in the quark model: An update”. In: *Phys. Rev. D* 52 (1995), pp. 2783–2812. DOI: [10.1103/PhysRevD.52.2783](https://doi.org/10.1103/PhysRevD.52.2783). arXiv: [hep-ph/9503486](https://arxiv.org/abs/hep-ph/9503486) [[hep-ph](https://arxiv.org/abs/hep-ph/9503486)].
- [22] Thomas 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. ISSN: 2510-2044. DOI: [10.1007/s41781-017-0002-8](https://doi.org/10.1007/s41781-017-0002-8). URL: <https://doi.org/10.1007/s41781-017-0002-8>.
- [23] D. M. Asner, M. Athanas, et al. “Search for exclusive charmless hadronic  $B$  decays”. In: *Physical Review D* 53.3 (1996), p. 1039.
- [24] A. J. Bevan, B. Golob, et al. “The physics of the B factories”. In: *The European Physical Journal C* 74.11 (2014), p. 3026.

- [25] P. del Amo Sanchez, J. P. Lees, et al. “Study of  $B \rightarrow \pi l \nu$  and  $B \rightarrow \rho l \nu$  decays and determination of  $|V_{ub}|$ ”. In: *Phys. Rev. D* 83 (3 Feb. 2011), p. 032007. DOI: [10.1103/PhysRevD.83.032007](https://doi.org/10.1103/PhysRevD.83.032007). URL: <https://link.aps.org/doi/10.1103/PhysRevD.83.032007>.
- [26] Justin Stevens and Mike Williams. “uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers”. In: *Journal of Instrumentation* 8.12 (2013), P12013.
- [27] Wouter Verkerke and David Kirkby. “The RooFit toolkit for data modeling”. In: *Statistical Problems in Particle Physics, Astrophysics and Cosmology*. World Scientific, 2006, pp. 186–189.
- [28] F. James. “MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1”. In: (1994).
- [29] M. Tanabashi. “Review of particle physics”. In: *Phys. Rev. D* 98 (2018), p. 030001.
- [30] Sangtae Ahn and Jeffrey A Fessler. “Standard errors of mean, variance, and standard deviation estimators”. In: *EECS Department, The University of Michigan* (2003), pp. 1–2.





# Appendix A

## ROE MVA Control Plots

### A.1 ROE Clean-up $\pi^0$ Training

#### A.1.1 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 A.1: Variable names, aliases and importance in the scope of  $\pi^0$  MVA training for ROE clean-up.

### A.1.2 Variable Distributions

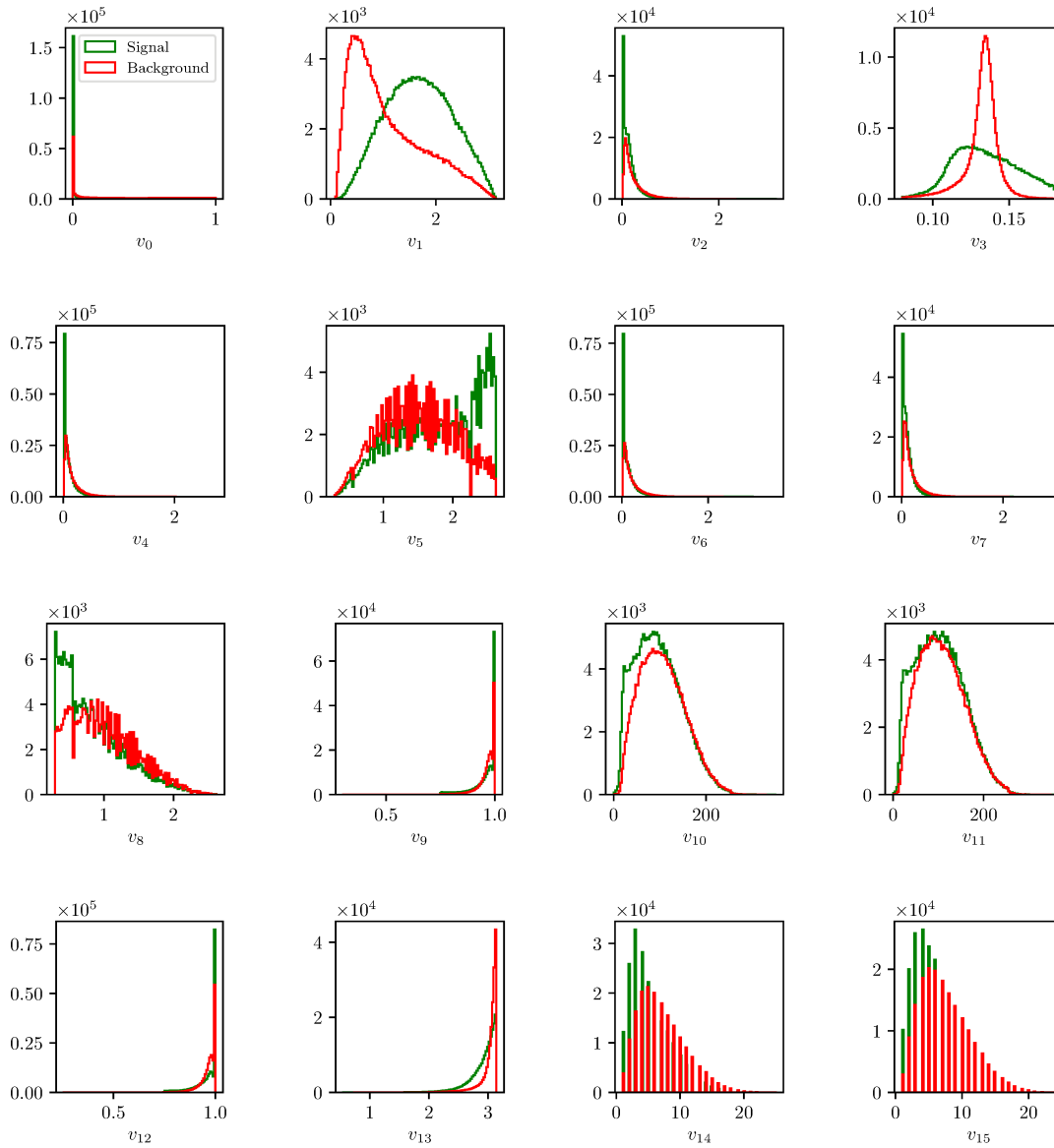


Figure A.1: Feature distributions for MVA training of  $\pi^0$  candidates in the scope of ROE clean-up.

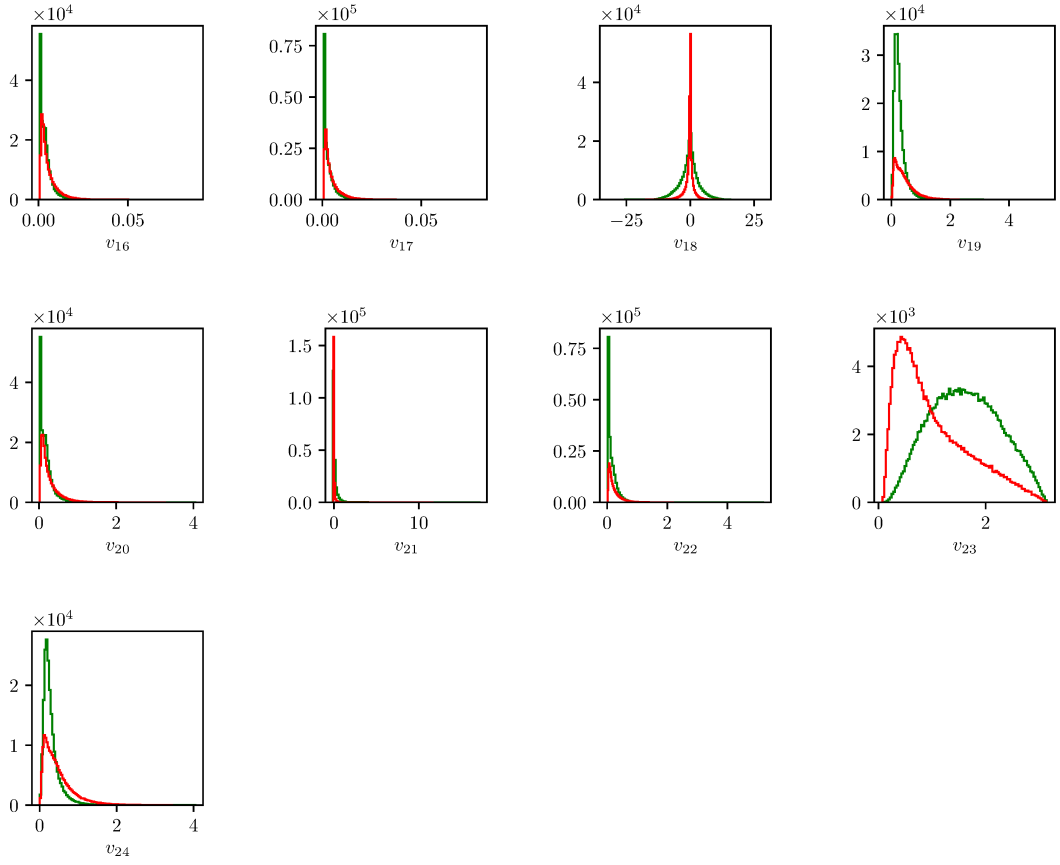


Figure A.1: Feature distributions for MVA training of  $\pi^0$  candidates in the scope of ROE clean-up.

### A.1.3 Hyper-parameter Optimization

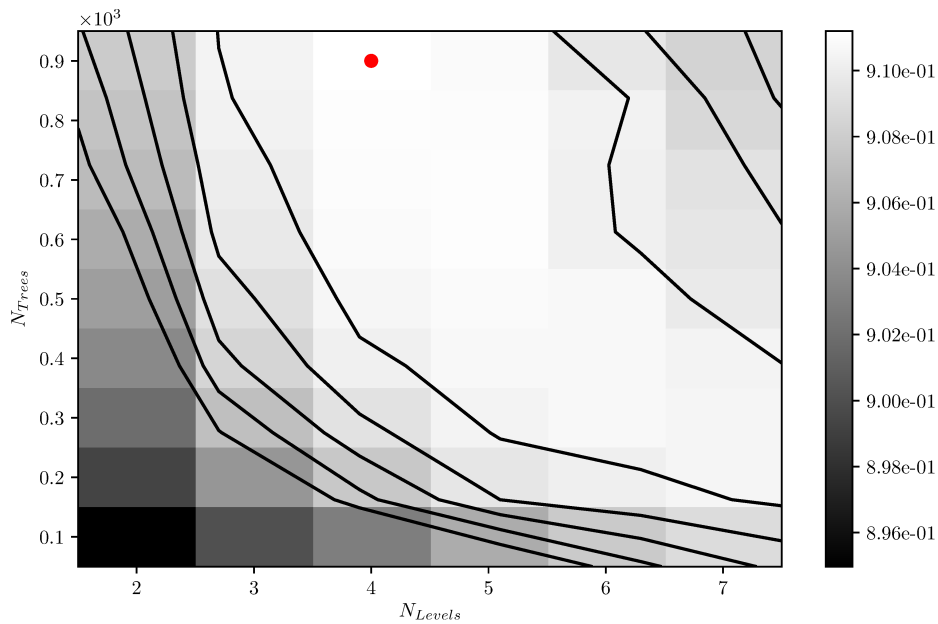


Figure A.2: Hyper-parameter optimization of  $nTrees$  and  $nLevels$  in the  $BDT$  forest training of  $\pi^0$  candidates in the scope of the ROE clean-up.

### A.1.4 Results

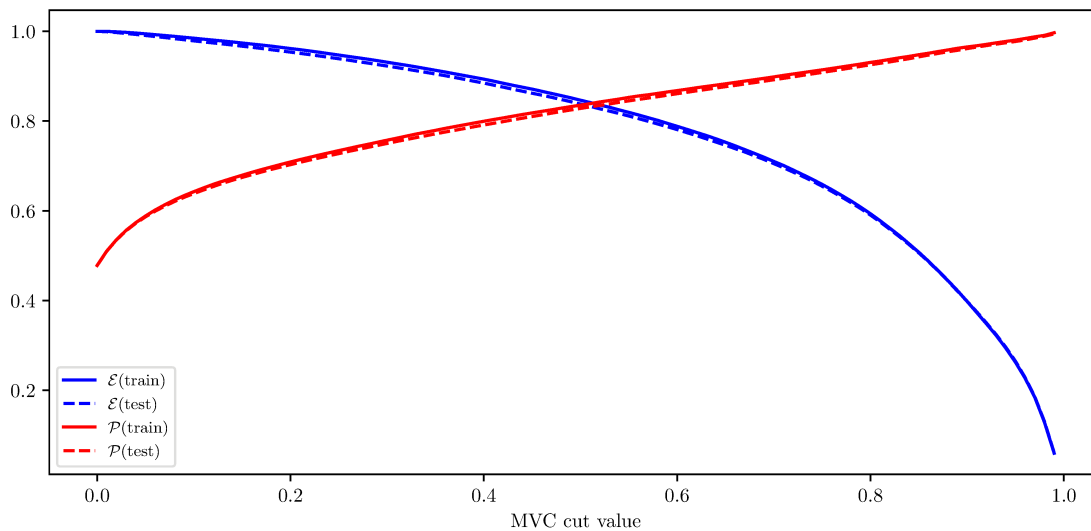


Figure A.3: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for  $\pi^0$  candidates training on the train (solid) and test (dashed) samples.

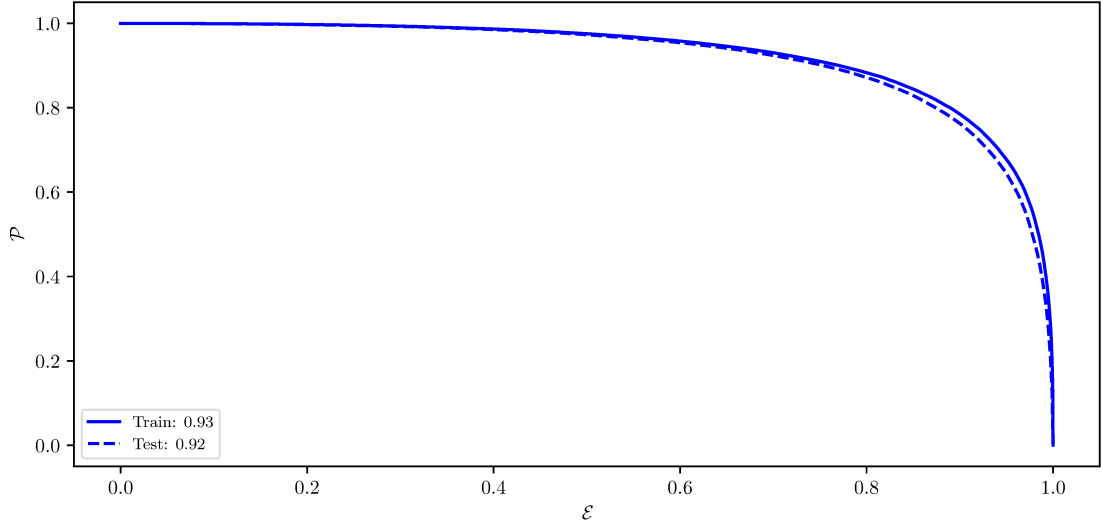


Figure A.4: ROC curves of the MVA classifier output for  $\pi^0$  candidates training on the train (solid) and test (dashed) samples.

## A.2 ROE Clean-up $\gamma$ Training

### A.2.1 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 A.2: Variable names, aliases and importance in the scope of  $\gamma$  MVA training for ROE clean-up.

## A.2.2 Variable Distributions

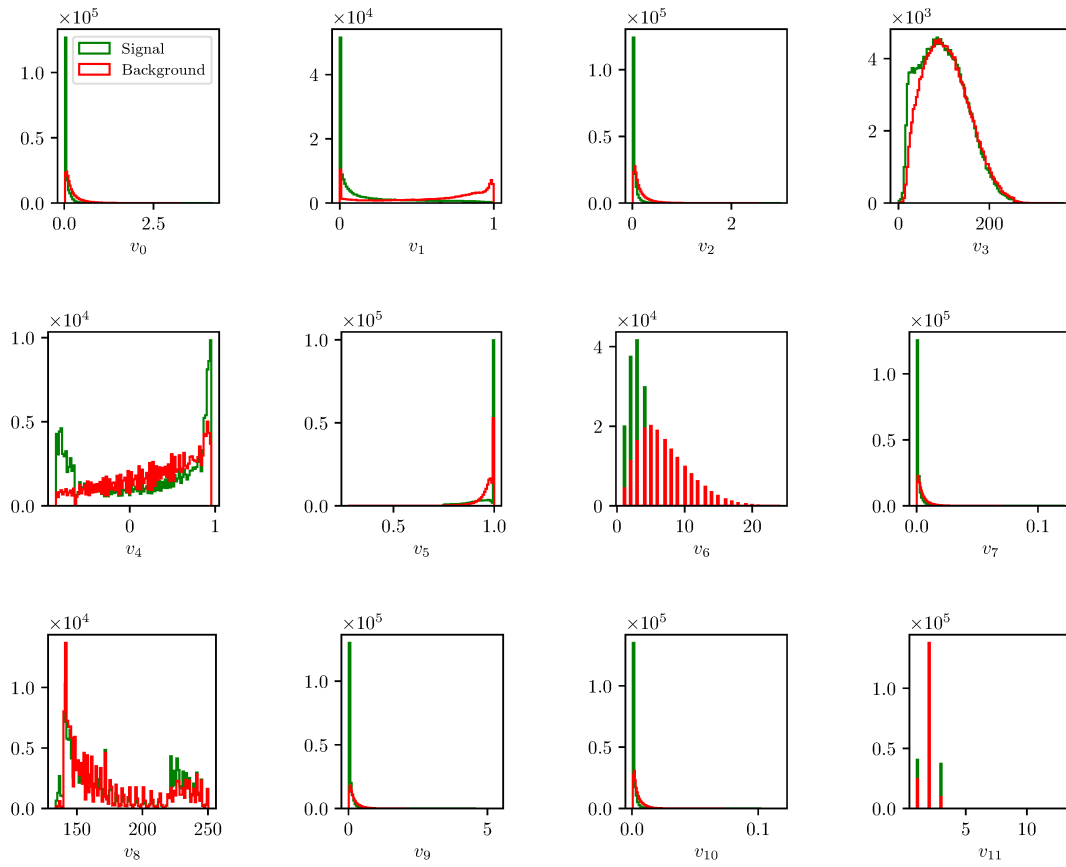


Figure A.5: Feature distributions for MVA training of  $\gamma$  candidates in the scope of ROE clean-up.

### A.2.3 Hyper-parameter Optimization

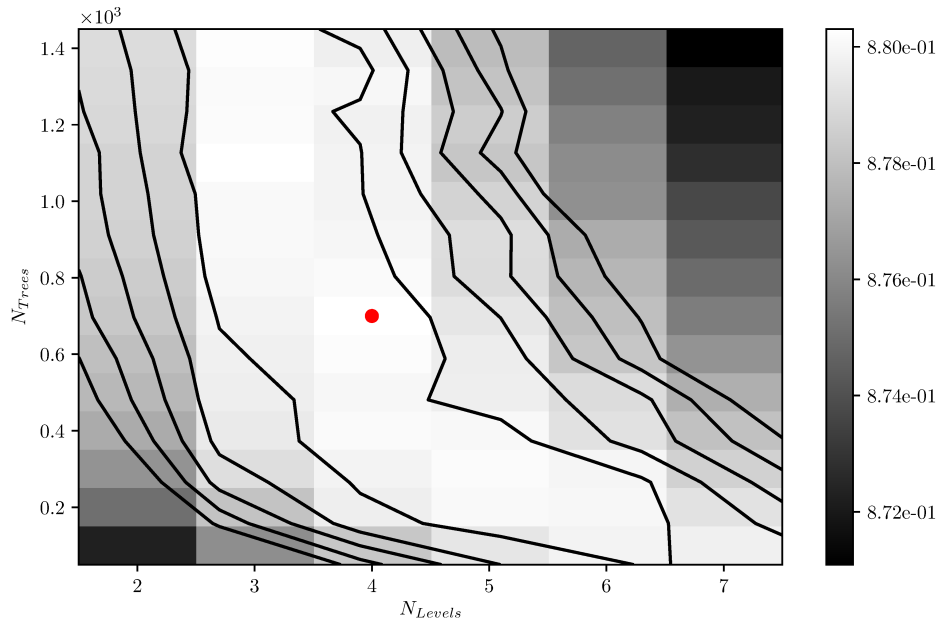


Figure A.6: Hyper-parameter optimization of  $n_{Trees}$  and  $n_{Levels}$  in the  $BDT$  forest training of  $\gamma$  candidates in the scope of the ROE clean-up.

### A.2.4 Results

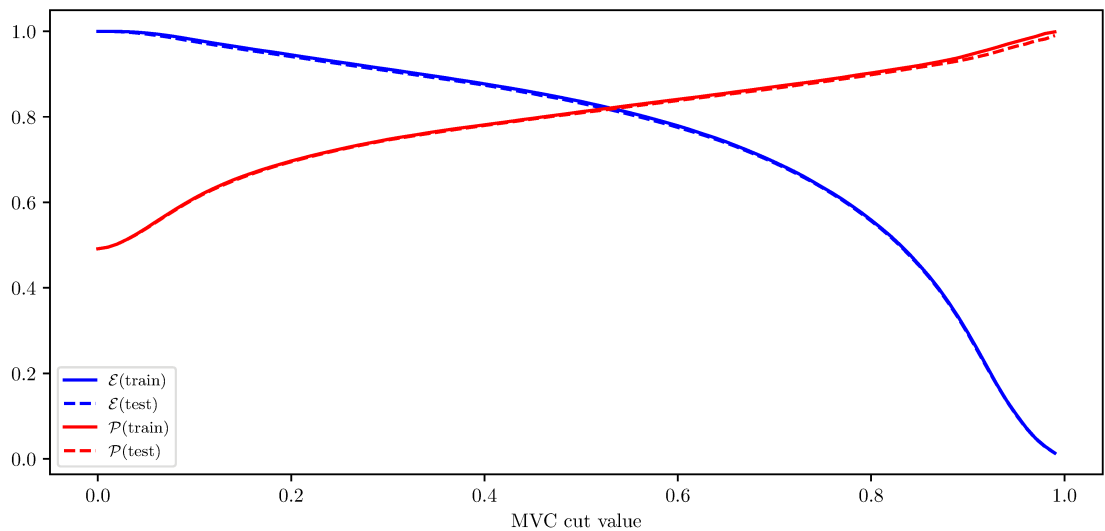


Figure A.7: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for  $\gamma$  candidates training on the train (solid) and test (dashed) samples.

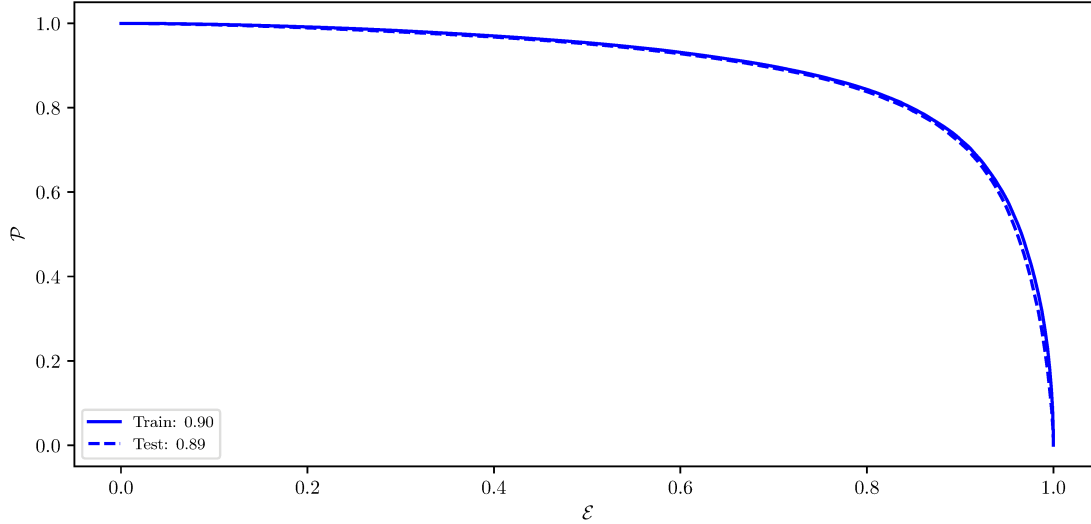


Figure A.8: ROC curves of the MVA classifier output for  $\gamma$  candidates training on the train (solid) and test (dashed) samples.

## A.3 ROE Clean-up Duplicate Pair Training

### A.3.1 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	<code>daughter(0,pValue)</code>	$v_{18}$	0.020
19	<code>daughter(1,tanlambda)</code>	$v_{19}$	0.018
20	<code>daughter(1,pValue)</code>	$v_{20}$	0.018
21	<code>daughter(0,tanlambda)</code>	$v_{21}$	0.017
22	<code>daughter(0,phi0)</code>	$v_{22}$	0.016
23	<code>daughter(1,phi0)</code>	$v_{23}$	0.016
24	<code>daughter(0,useCMSFrame(p))</code>	$v_{24}$	0.015
25	<code>daughter(0,z0Err)</code>	$v_{25}$	0.014
26	<code>daughter(1,omega)</code>	$v_{26}$	0.013
27	<code>daughter(0,omega)</code>	$v_{27}$	0.013
28	<code>daughter(1,z0Err)</code>	$v_{28}$	0.012
29	<code>daughter(0,pt)</code>	$v_{29}$	0.011
30	<code>daughter(0,omegaErr)</code>	$v_{30}$	0.011
31	<code>daughter(1,omegaErr)</code>	$v_{31}$	0.010
32	<code>daughter(1,pt)</code>	$v_{32}$	0.009
33	<code>daughter(0,tanlambdaErr)</code>	$v_{33}$	0.009
34	<code>daughter(1,tanlambdaErr)</code>	$v_{34}$	0.009
35	<code>useRestFrame(daughterAngleInBetween(0,1))</code>	$v_{35}$	0.003
36	<code>daughter(1,charge)</code>	$v_{36}$	0.000
37	<code>daughter(0,charge)</code>	$v_{37}$	0.000

Table A.3: Variable names, aliases and importance in the scope of duplicate track pair MVA training for ROE clean-up.

### A.3.2 Variable Distributions

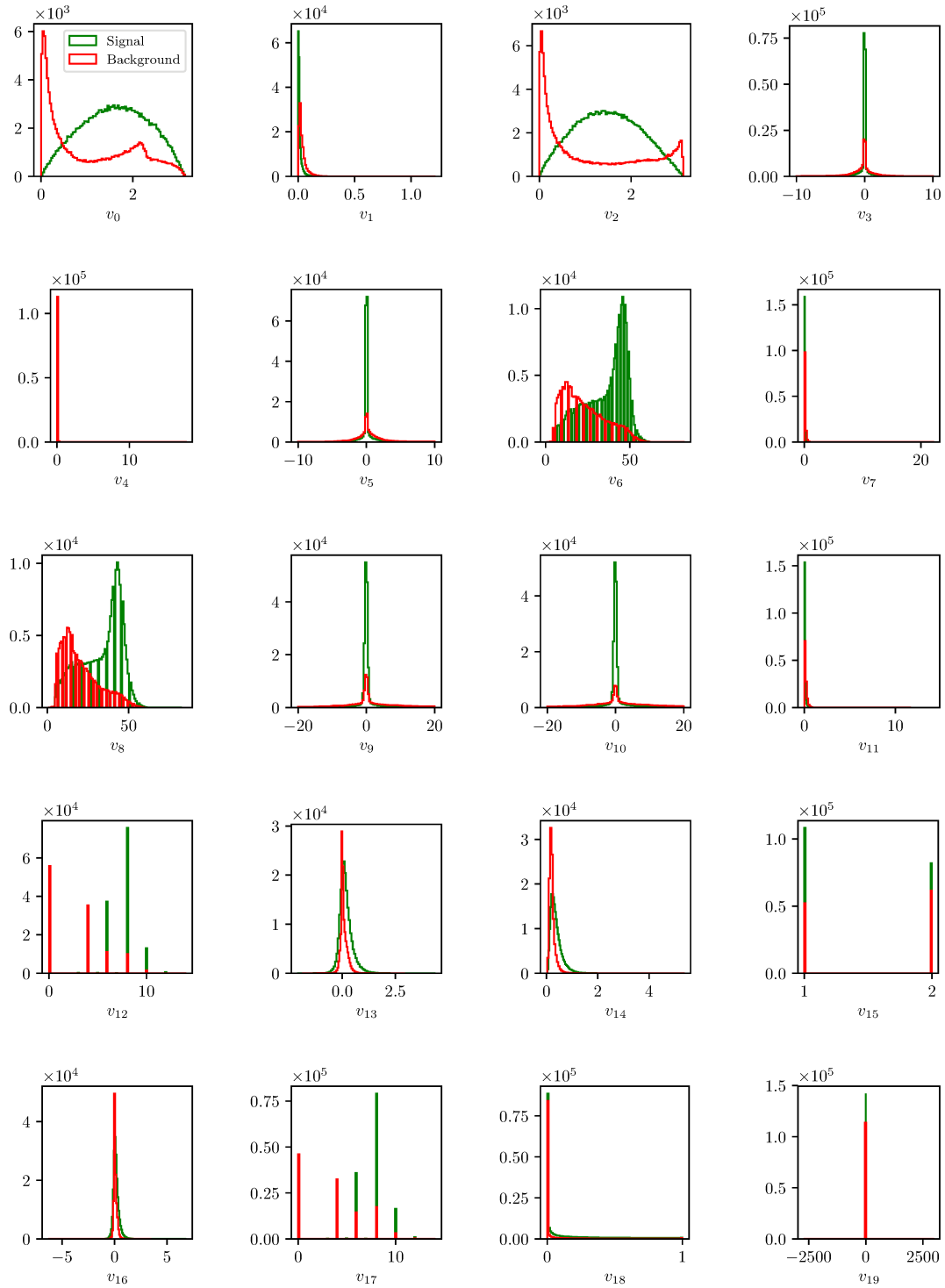


Figure A.9: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

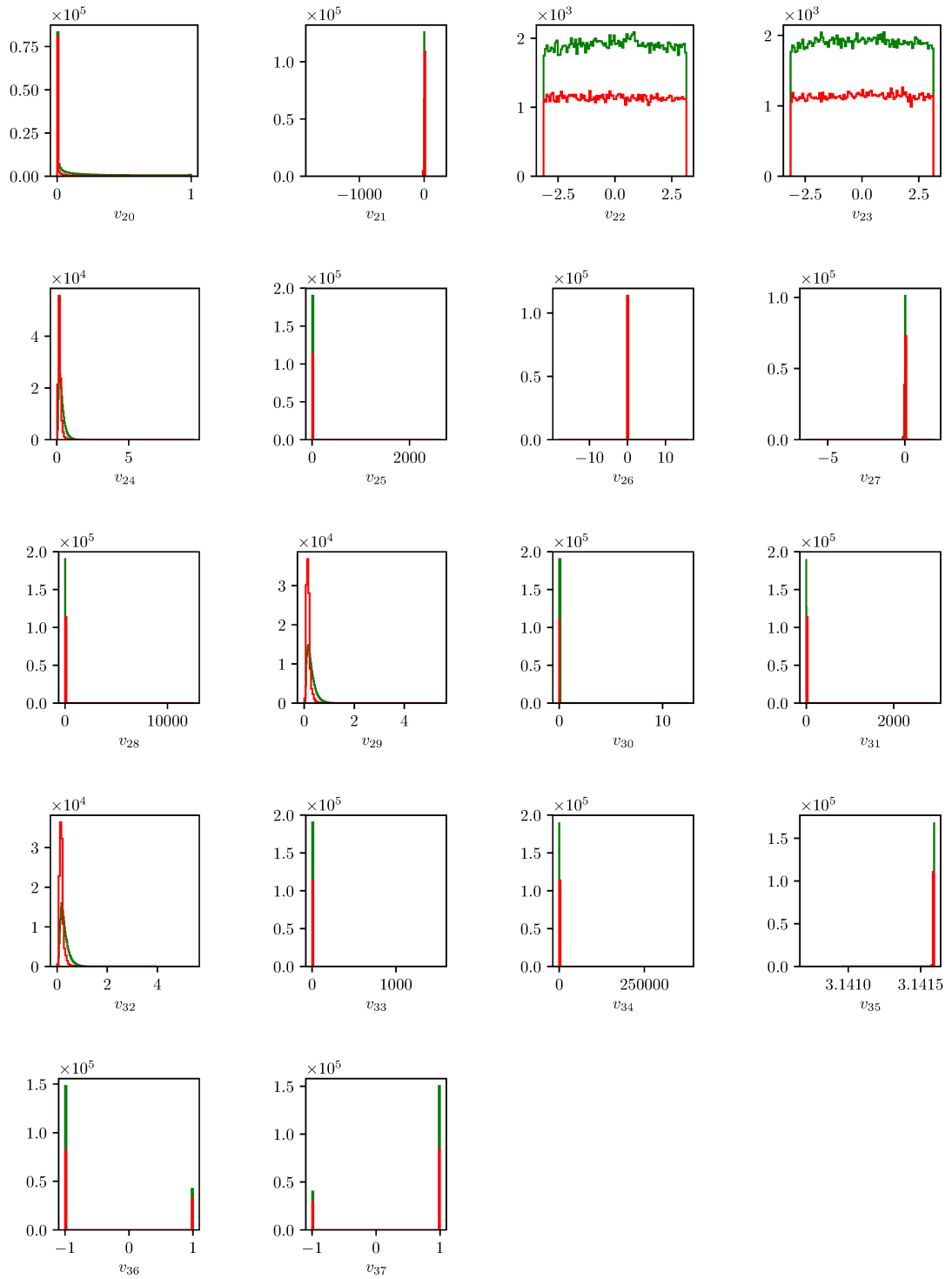


Figure A.9: Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up.

### A.3.3 Hyper-parameter Optimization

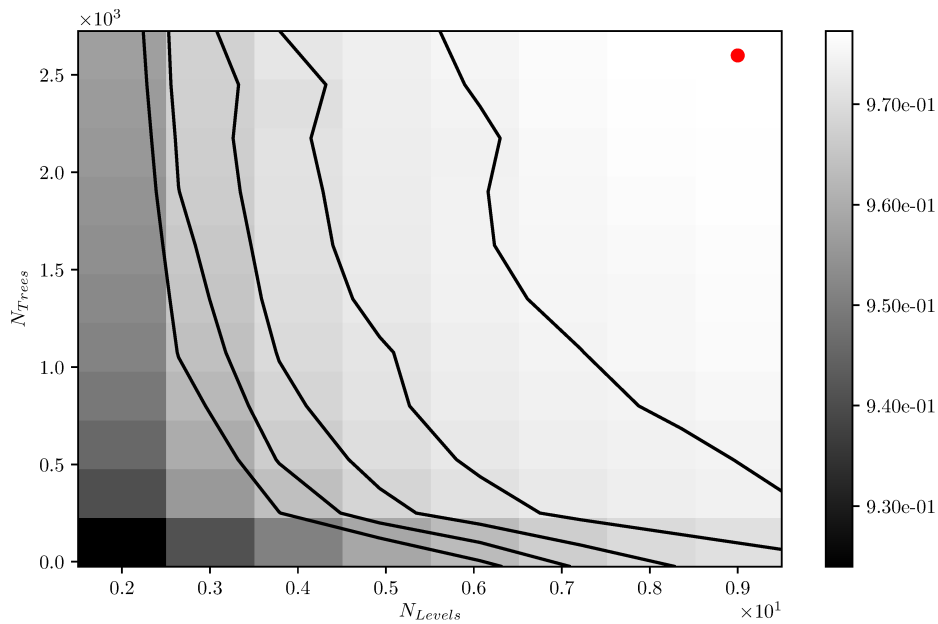


Figure A.10: Hyper-parameter optimization of  $nTrees$  and  $nLevels$  in the *BDT* forest training of duplicate track pair candidates in the scope of the ROE clean-up.

### A.3.4 Results

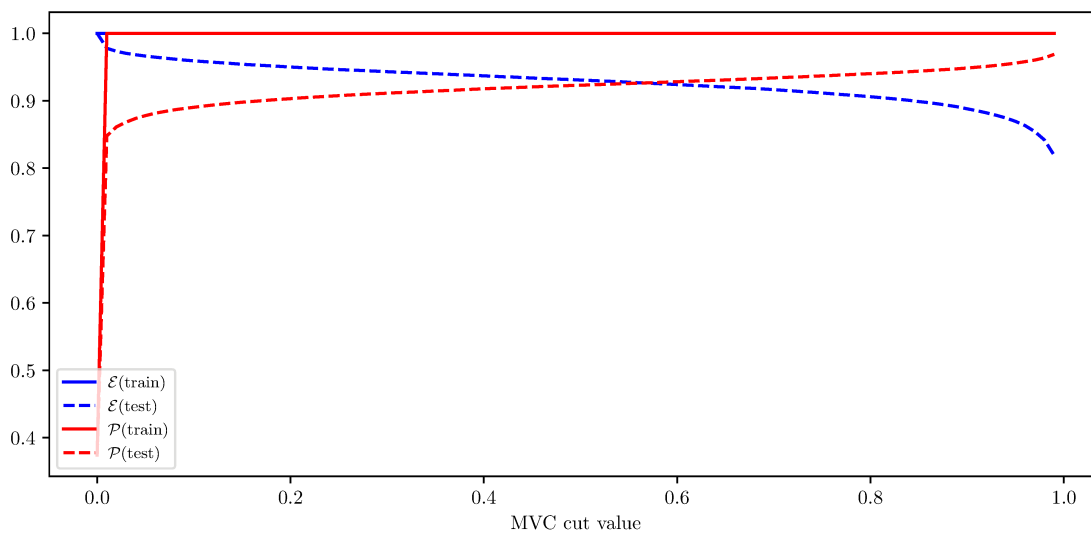


Figure A.11: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

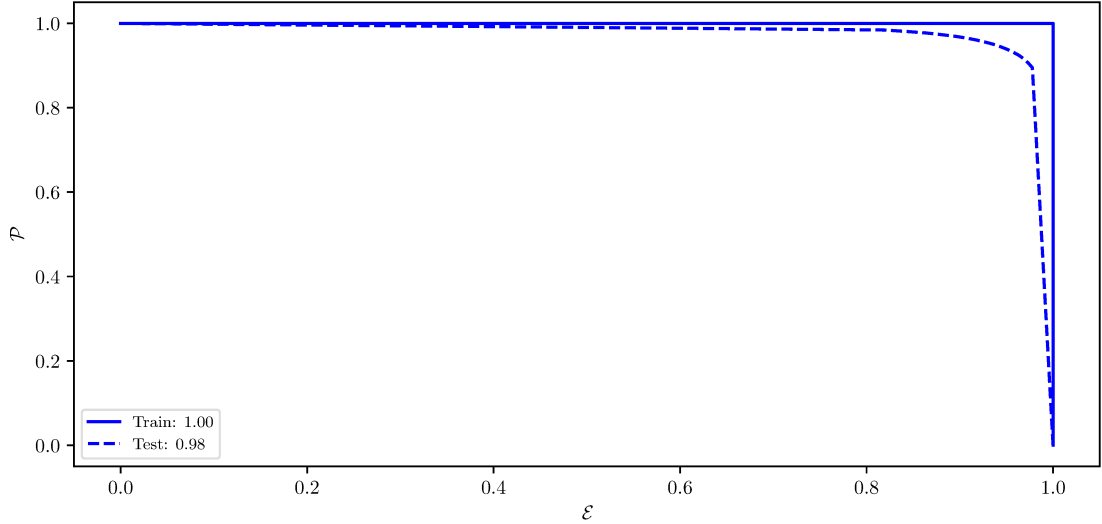


Figure A.12: ROC curves of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples.

## A.4 ROE Clean-up Duplicate Track Training

### A.4.1 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	<code>phi0</code>	$v_{18}$	0.023
19	<code>useCMSFrame(p)</code>	$v_{19}$	0.021

Table A.4: Variable names, aliases and importance in the scope of duplicate track MVA training for ROE clean-up.

## A.4.2 Variable Distributions

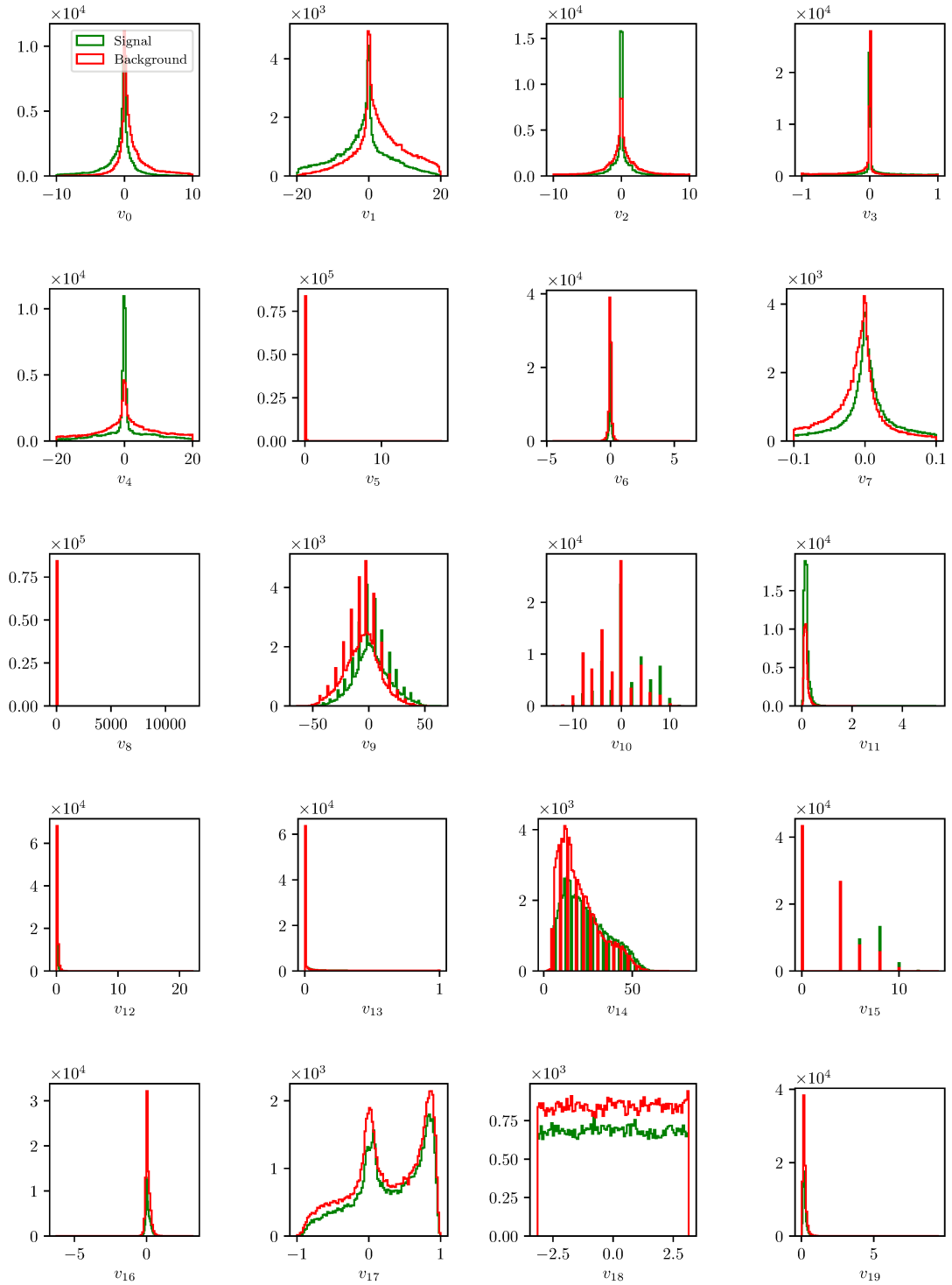


Figure A.13: Feature distributions for MVA training of duplicate track candidates in the scope of ROE clean-up.

### A.4.3 Hyper-parameter Optimization

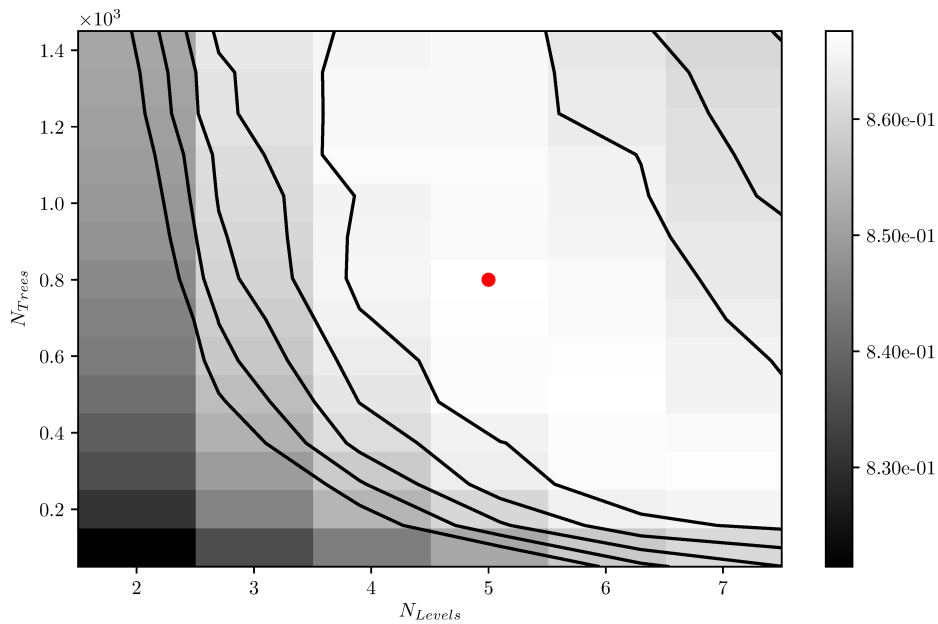


Figure A.14: Hyper-parameter optimization of `nTrees` and `nLevels` in the *BDT* forest training of duplicate track candidates in the scope of the ROE clean-up.

### A.4.4 Results

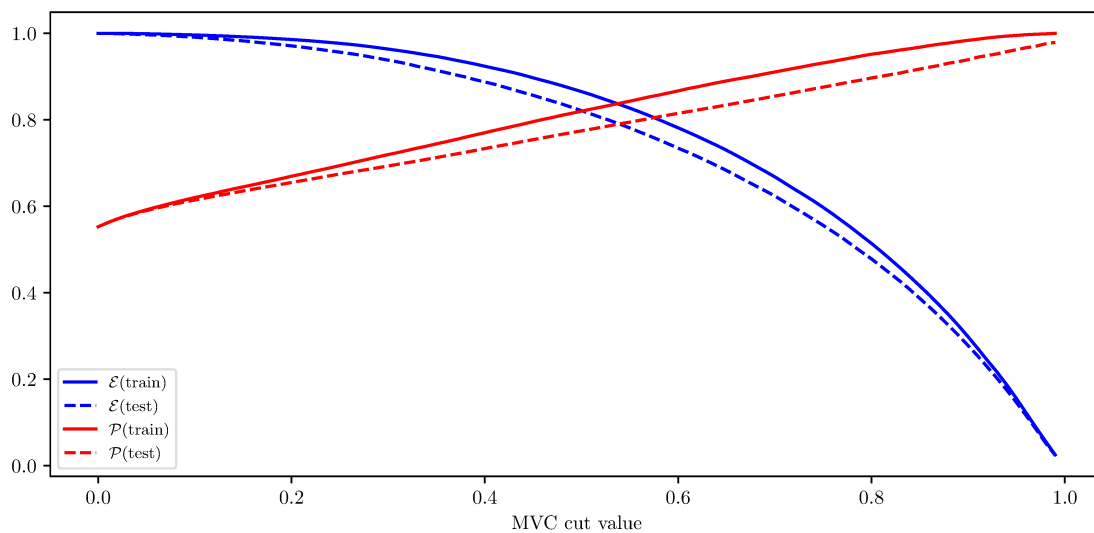


Figure A.15: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.



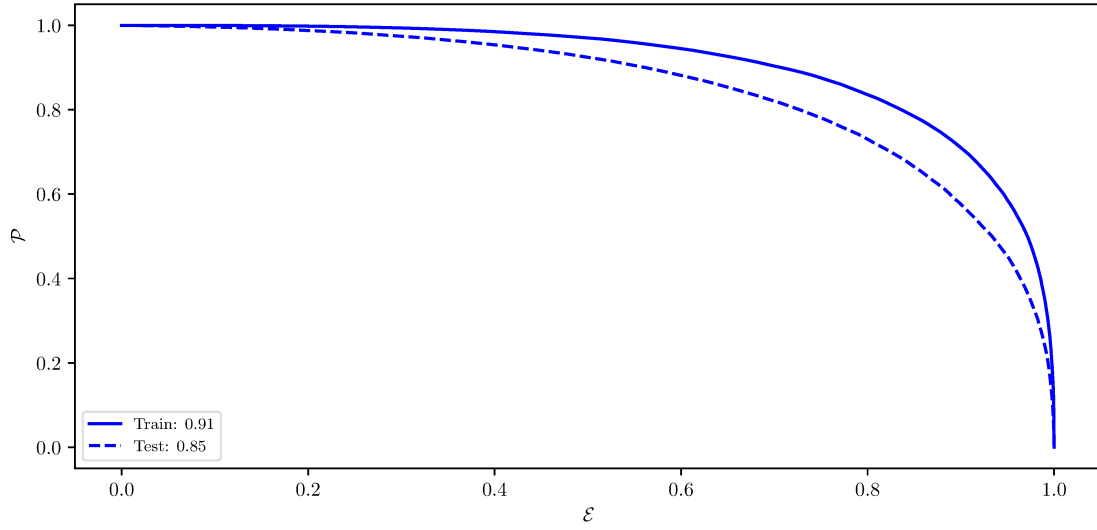


Figure A.16: ROC curves of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples.



# Appendix B

## MVA Control Plots

### B.1 $q\bar{q}$ Suppression Training

#### B.1.1 Variable Importance

	Name	Alias	Importance
0	B_CosTBT0	$v_0$	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_Thrust0	$v_7$	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 B.1: Variable names, aliases and importance in the scope of  $q\bar{q}$  suppression MVA training.

## B.1.2 Variable Distributions

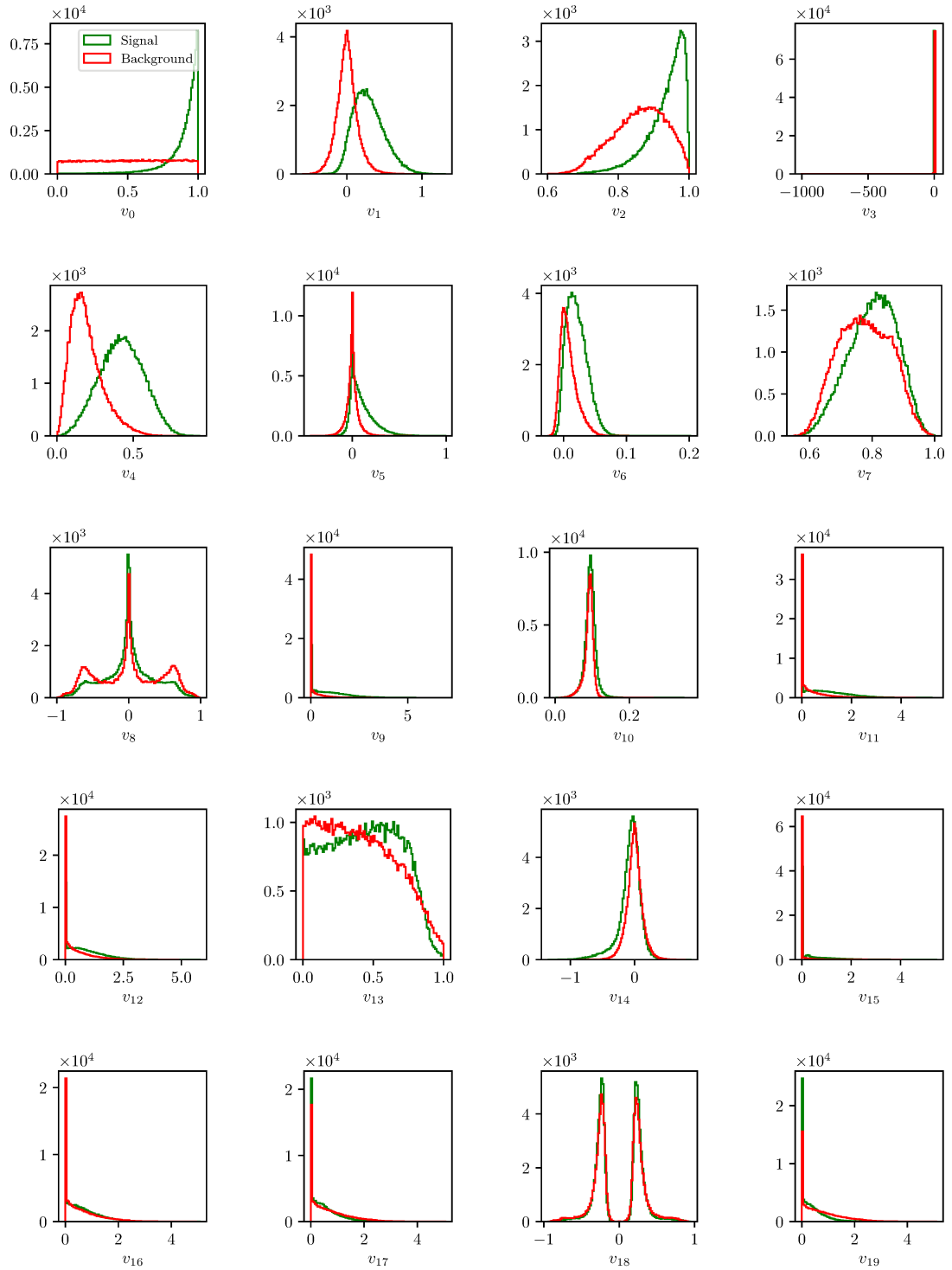


Figure B.1: Feature distributions for MVA training of  $q\bar{q}$  background suppression.

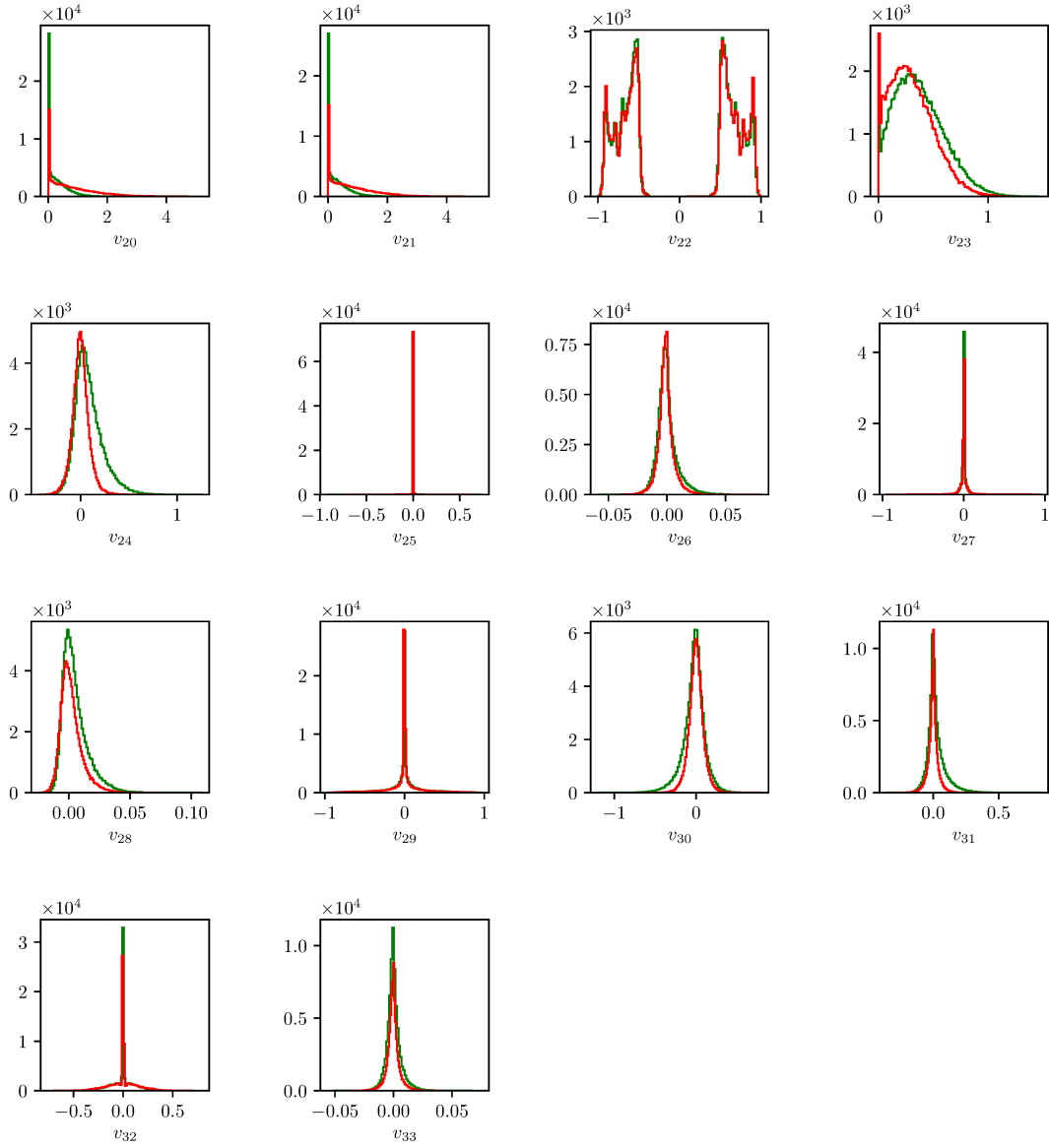


Figure B.1: Feature distributions for MVA training of  $q\bar{q}$  background suppression.

### B.1.3 Hyper-parameter Optimization

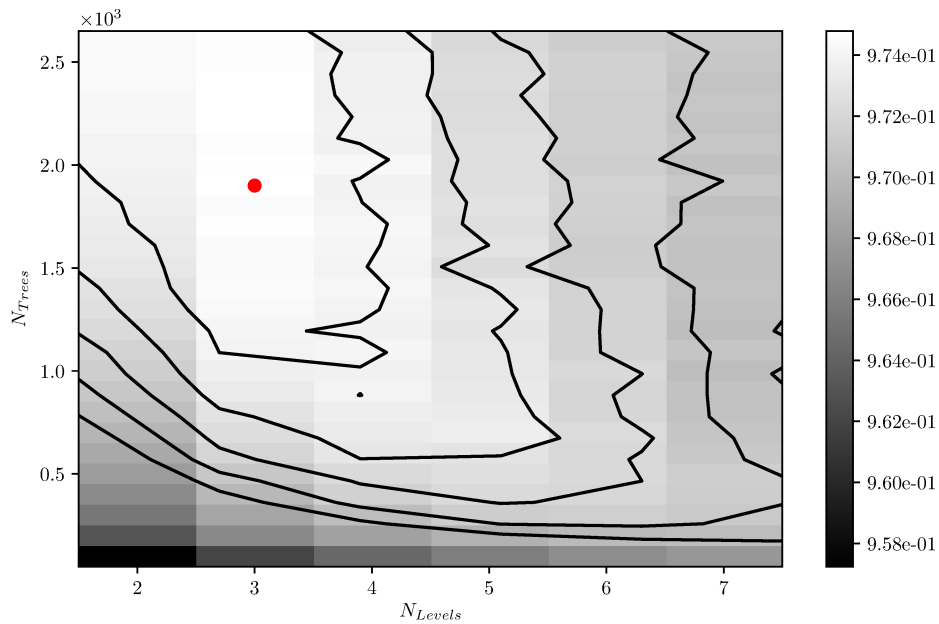


Figure B.2: Hyper-parameter optimization of `nTrees` and `nLevels` in the *BDT* forest training of  $q\bar{q}$  background suppression.

### B.1.4 Results

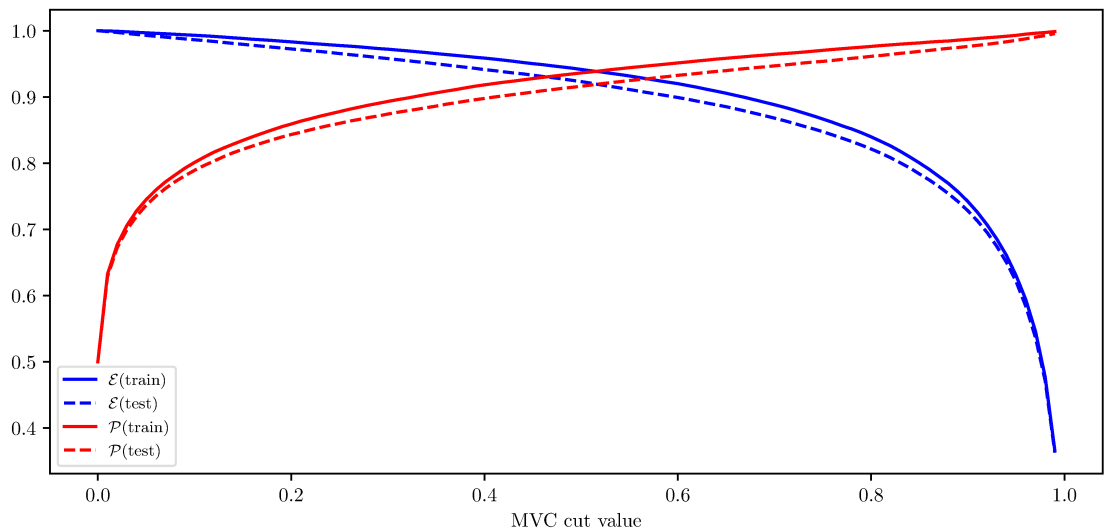


Figure B.3: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for  $q\bar{q}$  background suppression training on the train (solid) and test (dashed) samples.

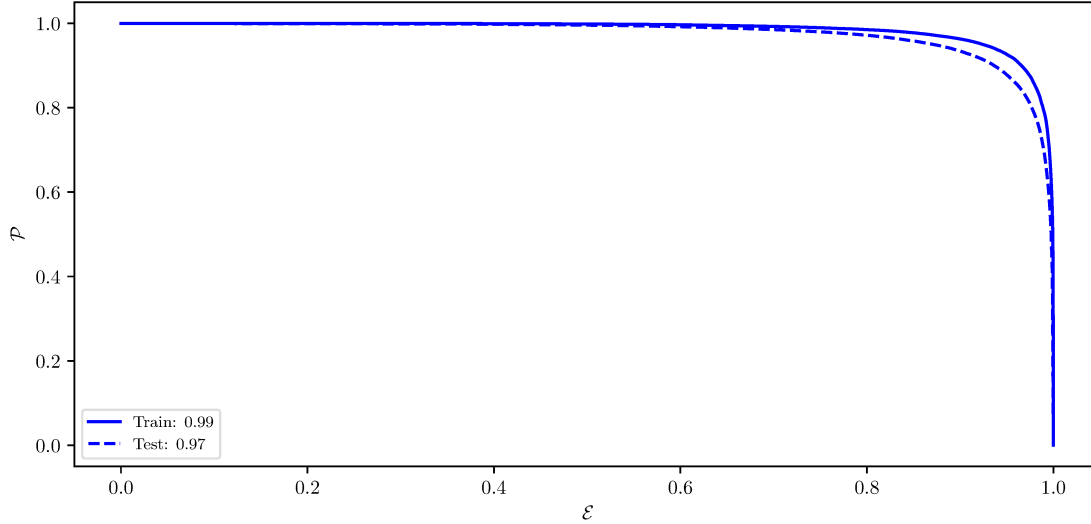


Figure B.4: ROC curves of the MVA classifier output for  $q\bar{q}$  background suppression training on the train (solid) and test (dashed) samples.

## B.2 Standard $B\bar{B}$ Suppression Training

### B.2.1 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_KOFT	$v_{14}$	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 B.2: Variable names, aliases and importance in the scope of  $B\bar{B}$  background suppression.

## B.2.2 Variable Distributions

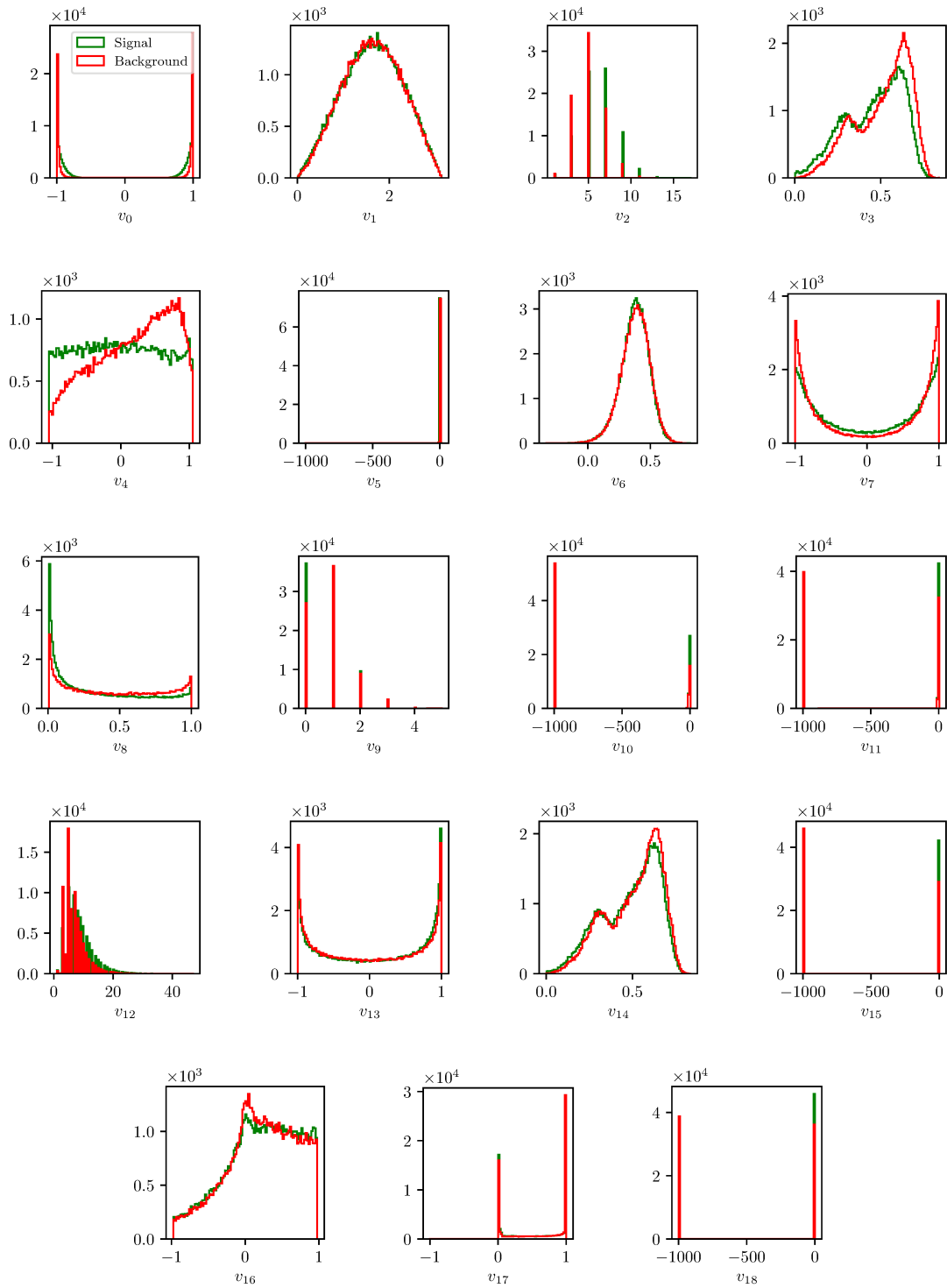


Figure B.5: Feature distributions for MVA training of  $B\bar{B}$  background suppression.

### B.2.3 Hyper-parameter Optimization

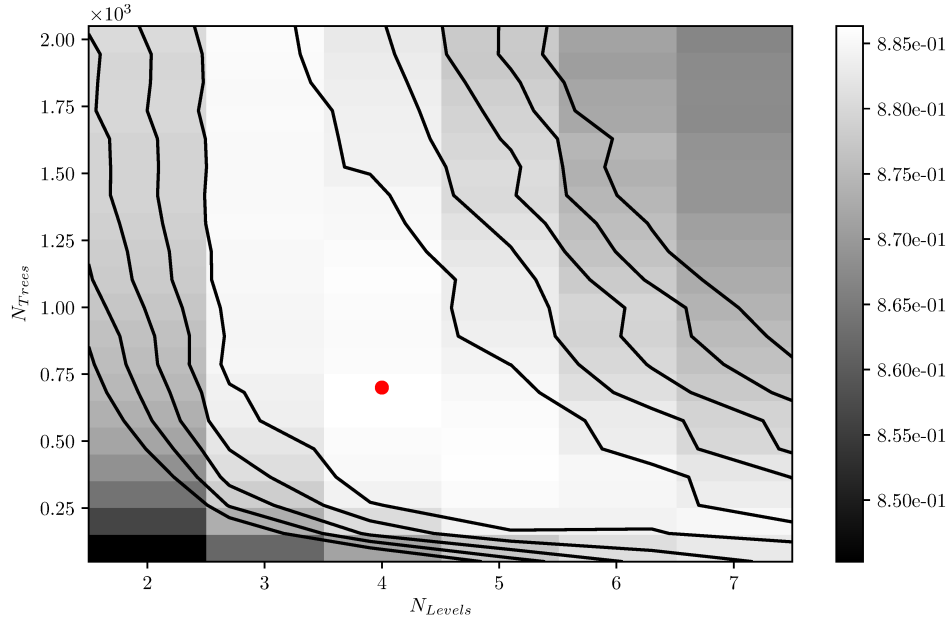


Figure B.6: Hyper-parameter optimization of `nTrees` and `nLevels` in the *BDT* forest training of  $B\bar{B}$  background suppression.

### B.2.4 Results

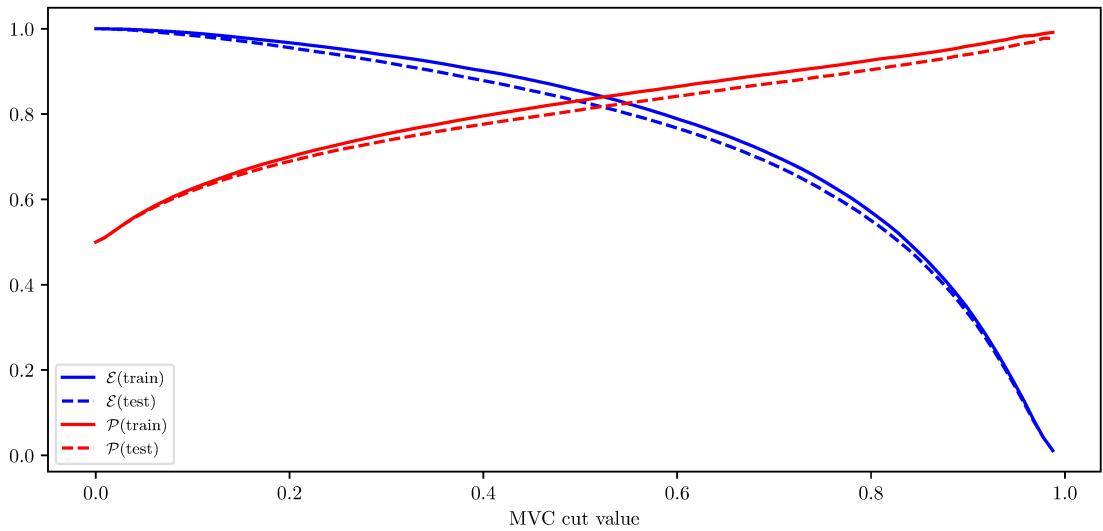


Figure B.7: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for  $B\bar{B}$  background suppression training on the train (solid) and test (dashed) samples.

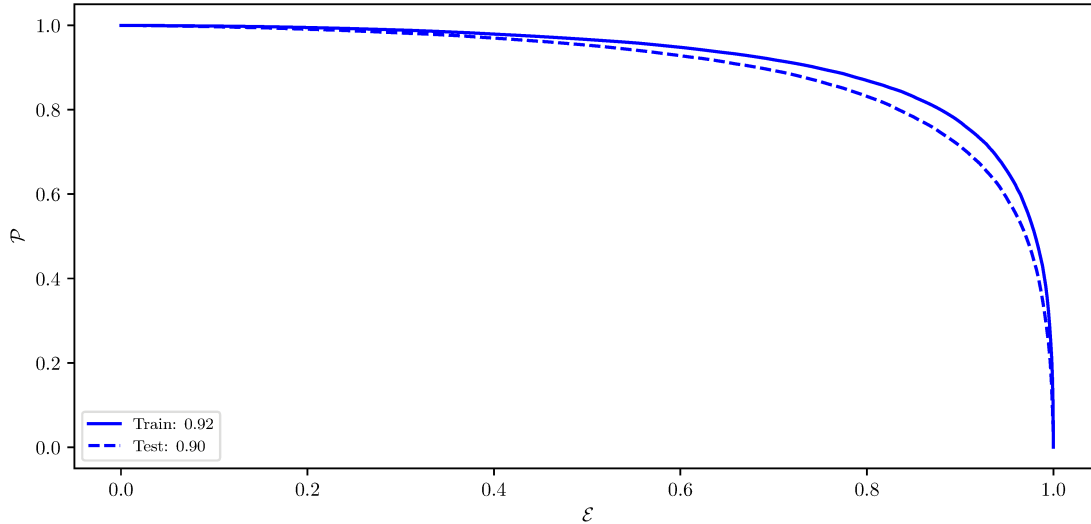


Figure B.8: ROC curves of the MVA classifier output for  $B\bar{B}$  background suppression training on the train (solid) and test (dashed) samples.

## B.3 Uniformity Boosted $B\bar{B}$ Suppression Training

### B.3.1 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

### B.3.2 Results

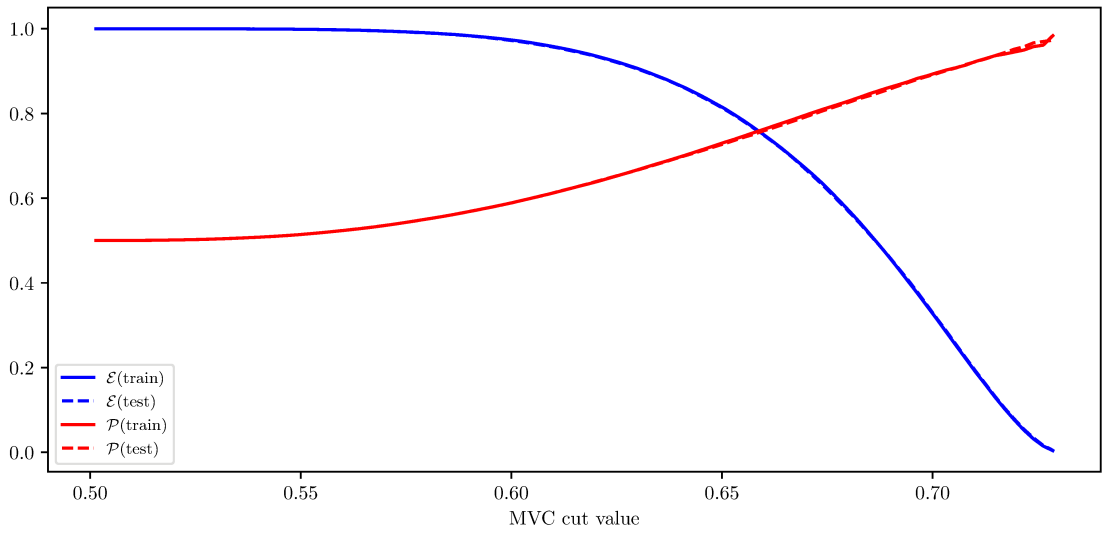


Figure B.9: Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the uniformity boosted MVA classifier output for  $B\bar{B}$  background suppression training on the train (solid) and test (dashed) samples.

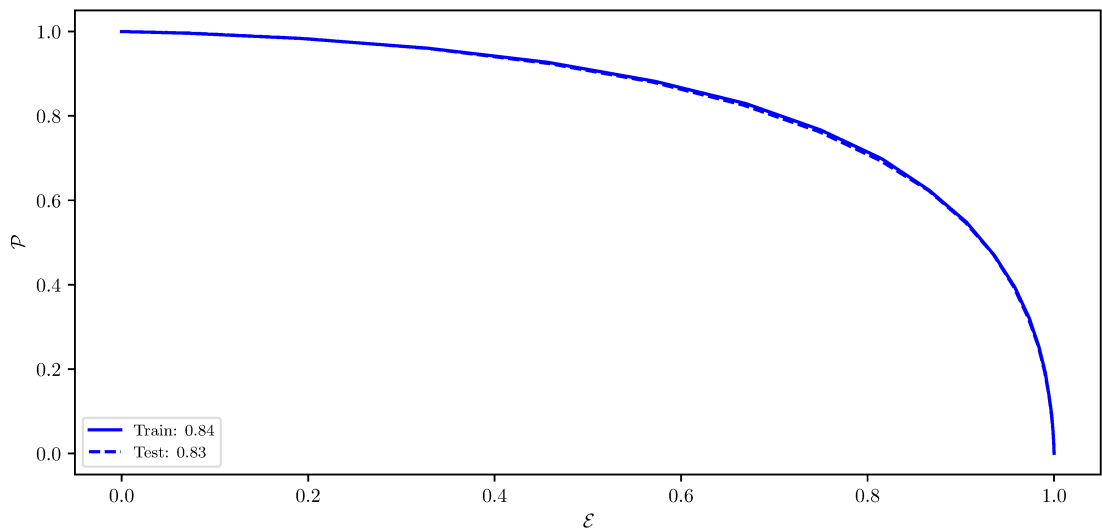


Figure B.10: ROC curves of the uniformity boosted MVA classifier output for  $B\bar{B}$  background suppression training on the train (solid) and test (dashed) samples.



# Appendix C

## Other Plots

### C.1 Signal Fits in $m_{KK}$

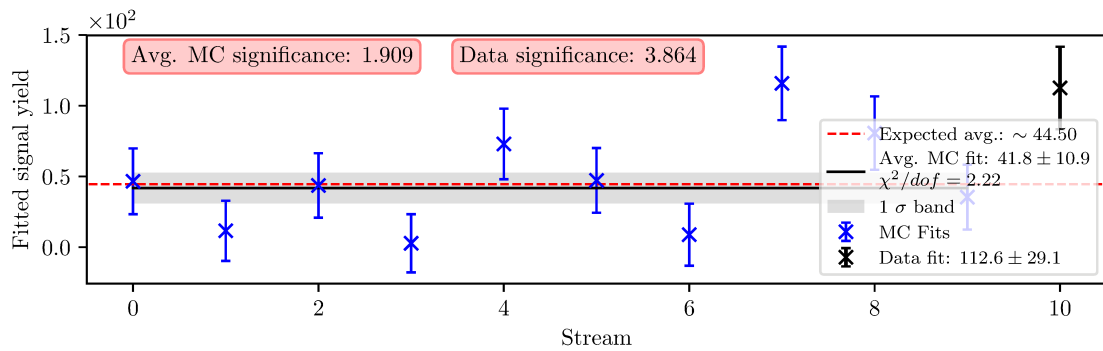


Figure C.1: Signal fit result for the 1<sup>st</sup>  $m_{KK}$  window for MC and data in the range  $0.980 < m_{KK} < 1.232$ .

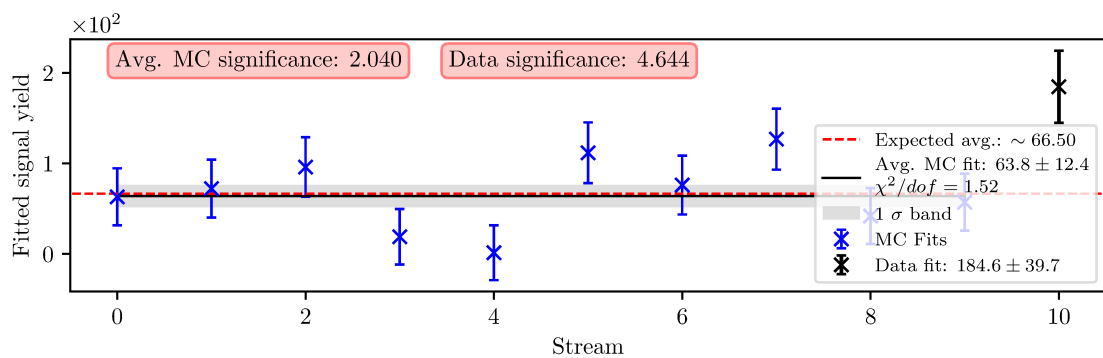


Figure C.2: Signal fit result for the 2<sup>nd</sup>  $m_{KK}$  window for MC and data in the range  $1.232 < m_{KK} < 1.483$ .

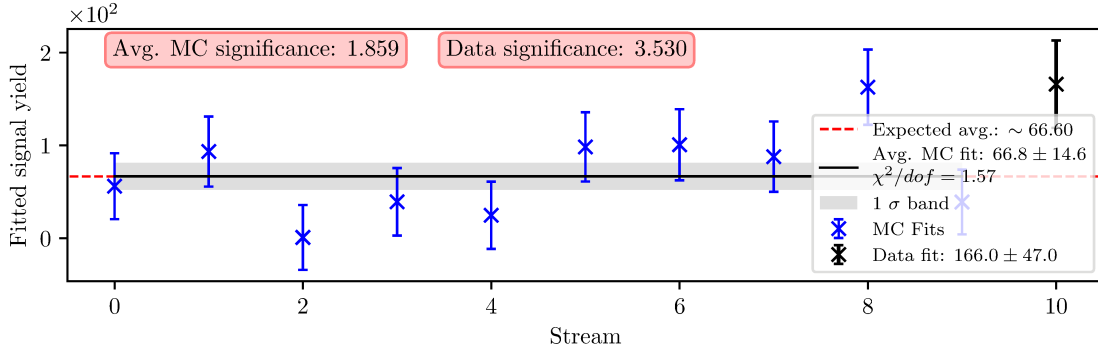


Figure C.3: Signal fit result for the 3<sup>rd</sup>  $m_{KK}$  window for MC and data in the range  $1.483 < m_{KK} < 1.735$ .

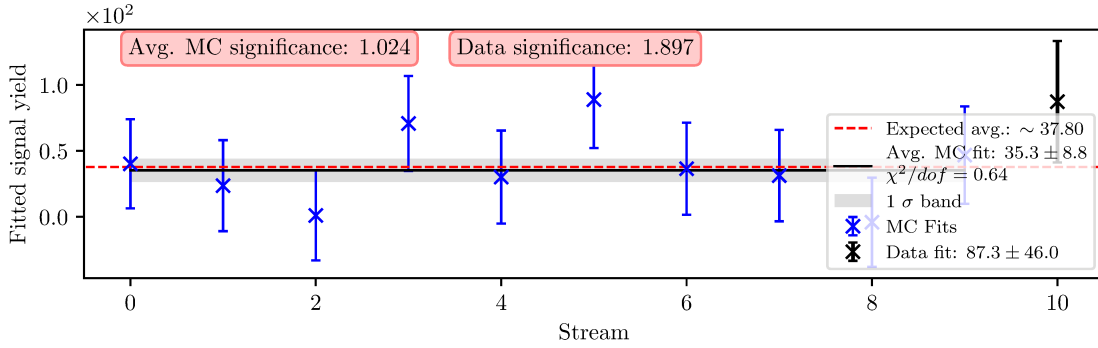


Figure C.4: Signal fit result for the 4<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $1.735 < m_{KK} < 1.987$ .

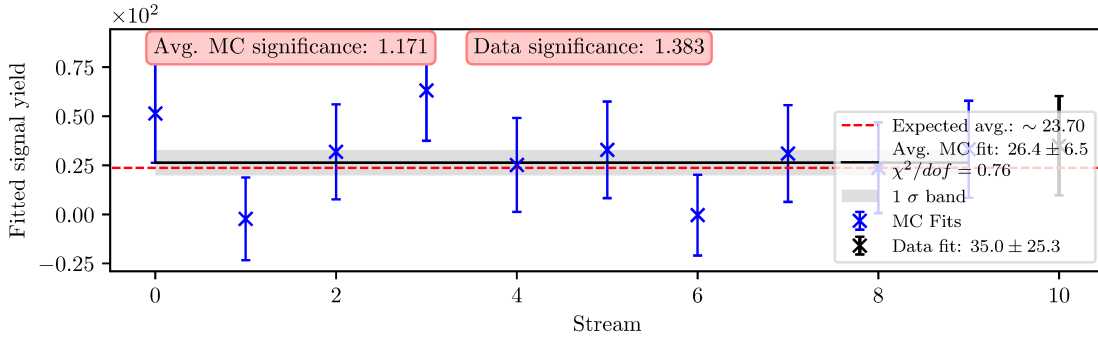


Figure C.5: Signal fit result for the 5<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $1.987 < m_{KK} < 2.238$ .



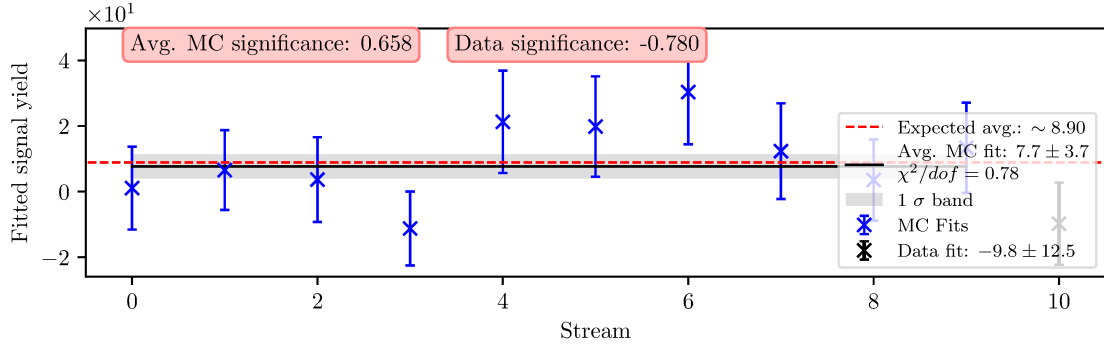


Figure C.6: Signal fit result for the 6<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $2.238 < m_{KK} < 2.490$ .

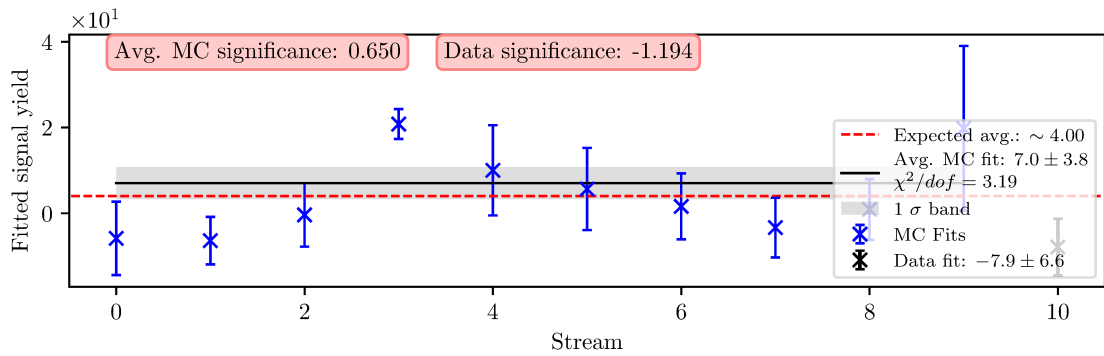


Figure C.7: Signal fit result for the 7<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $2.490 < m_{KK} < 2.742$ .

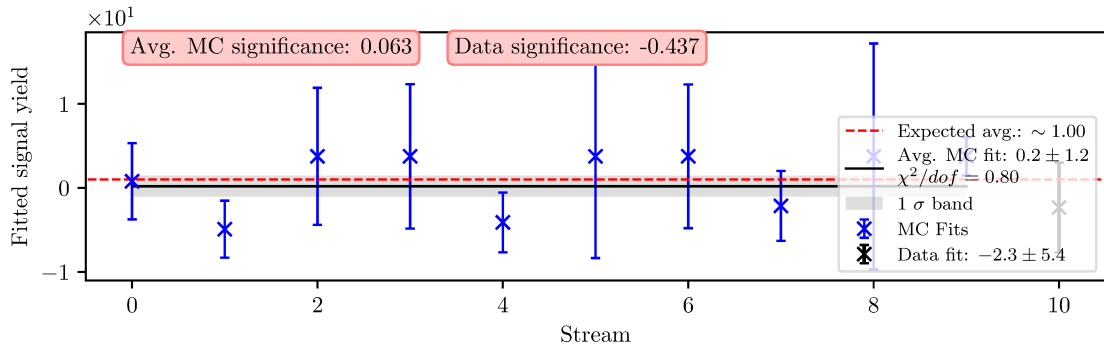


Figure C.8: Signal fit result for the 8<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $2.742 < m_{KK} < 2.993$ .

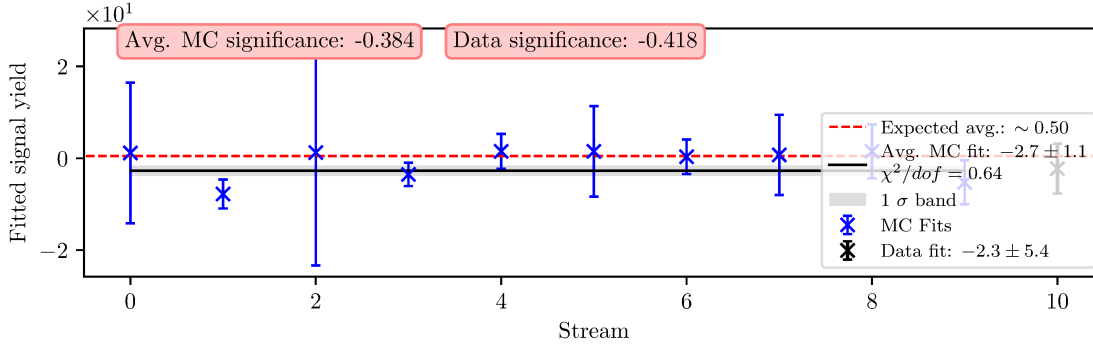


Figure C.9: Signal fit result for the 9<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $2.993 < m_{KK} < 3.245$ .

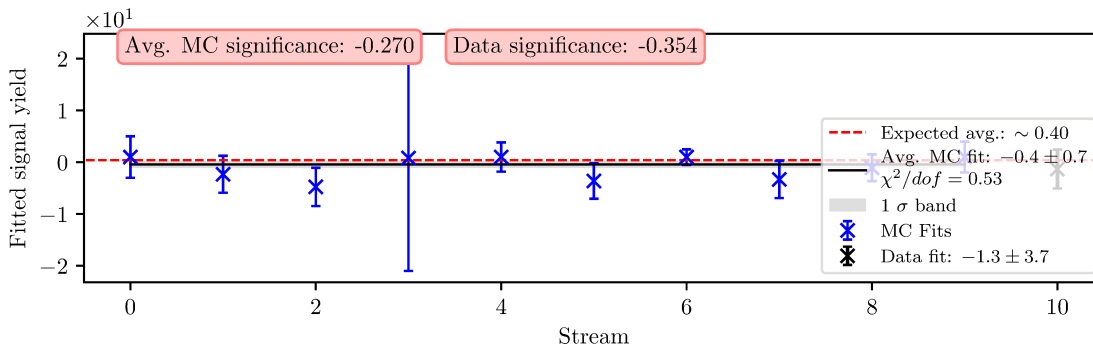


Figure C.10: Signal fit result for the 10<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $3.245 < m_{KK} < 3.497$ .

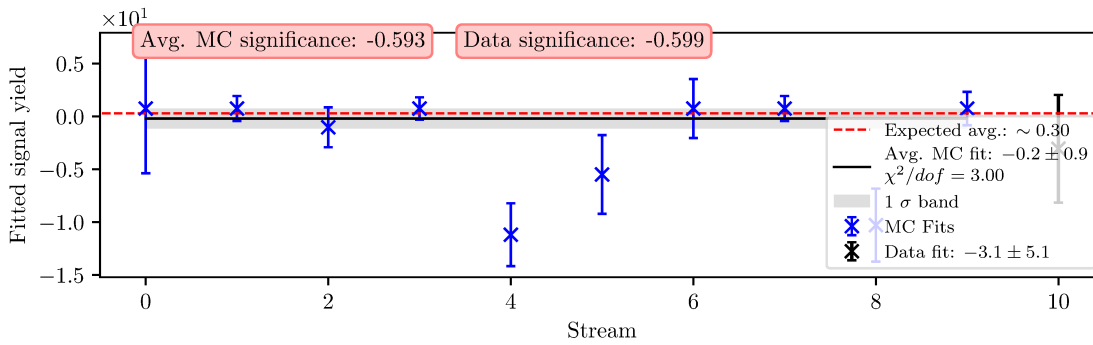


Figure C.11: Signal fit result for the 11<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $3.497 < m_{KK} < 3.748$ .

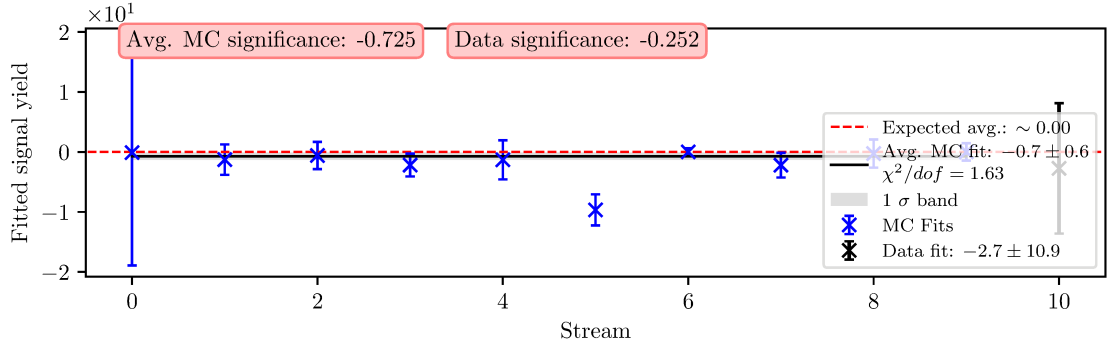


Figure C.12: Signal fit result for the 12<sup>th</sup>  $m_{KK}$  window for MC and data in the range  $3.748 < m_{KK} < 4.000$ .

## C.2 Signal Fits in $q^2$

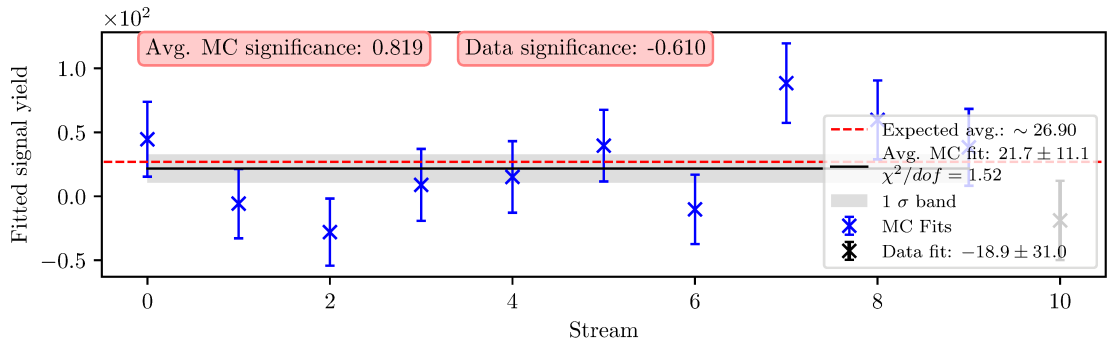


Figure C.13: Signal fit result for the 1<sup>st</sup>  $q^2$  window for MC and data in the range  $0.000 < q^2 < 1.500$ .

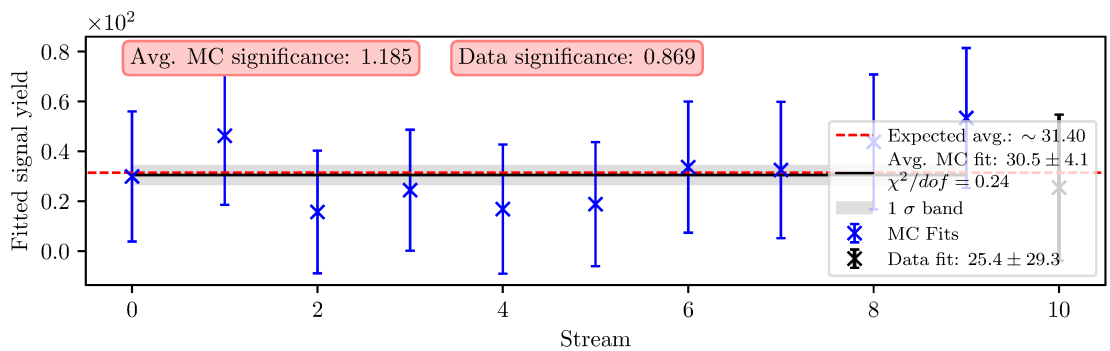


Figure C.14: Signal fit result for the 2<sup>nd</sup>  $q^2$  window for MC and data in the range  $1.500 < q^2 < 3.000$ .

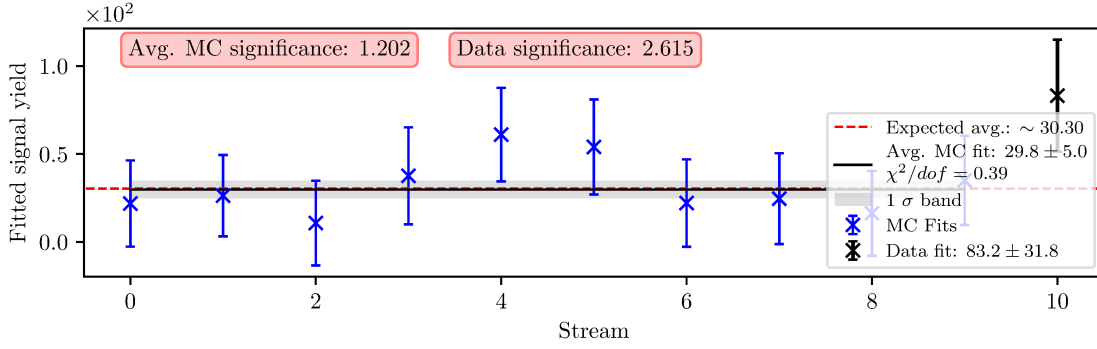


Figure C.15: Signal fit result for the 3<sup>rd</sup>  $q^2$  window for MC and data in the range  $3.000 < q^2 < 4.500$ .

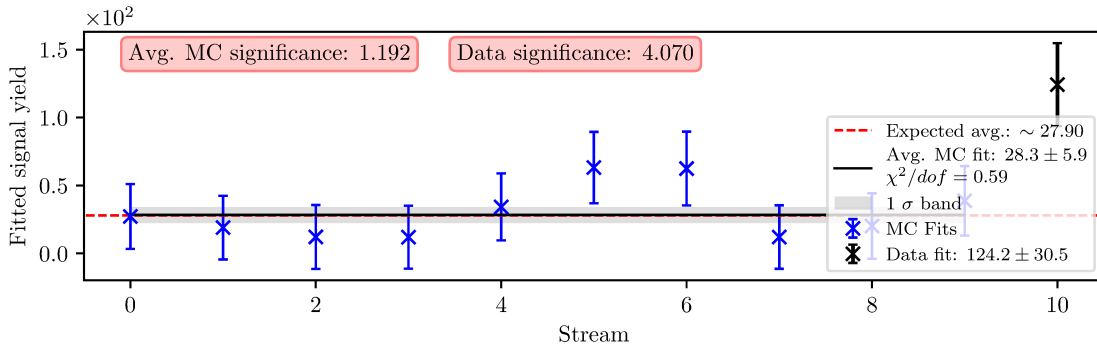


Figure C.16: Signal fit result for the 4<sup>th</sup>  $q^2$  window for MC and data in the range  $4.500 < q^2 < 6.000$ .

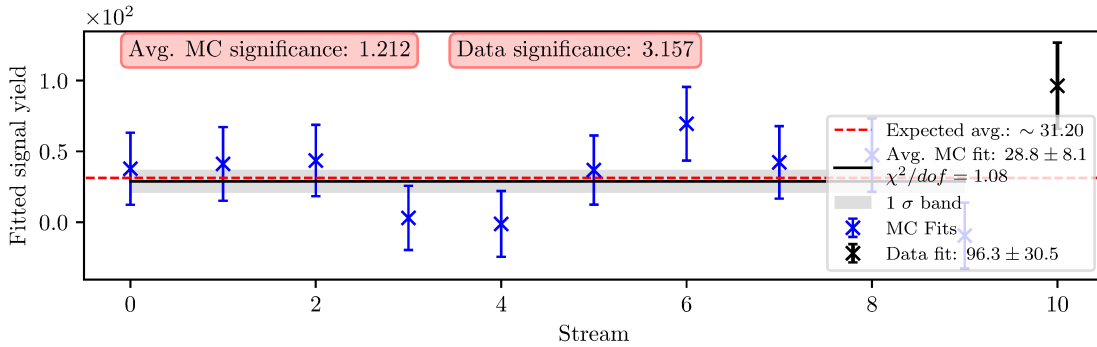


Figure C.17: Signal fit result for the 5<sup>th</sup>  $q^2$  window for MC and data in the range  $6.000 < q^2 < 7.500$ .

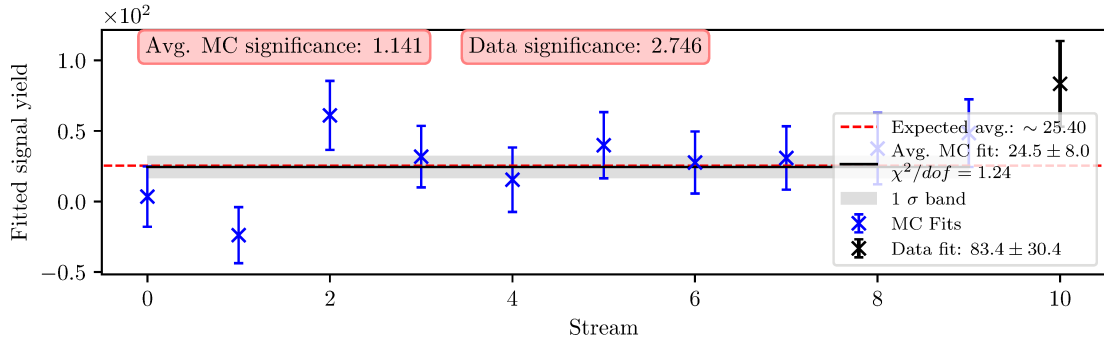


Figure C.18: Signal fit result for the 6<sup>th</sup>  $q^2$  window for MC and data in the range  $7.500 < q^2 < 9.000$ .

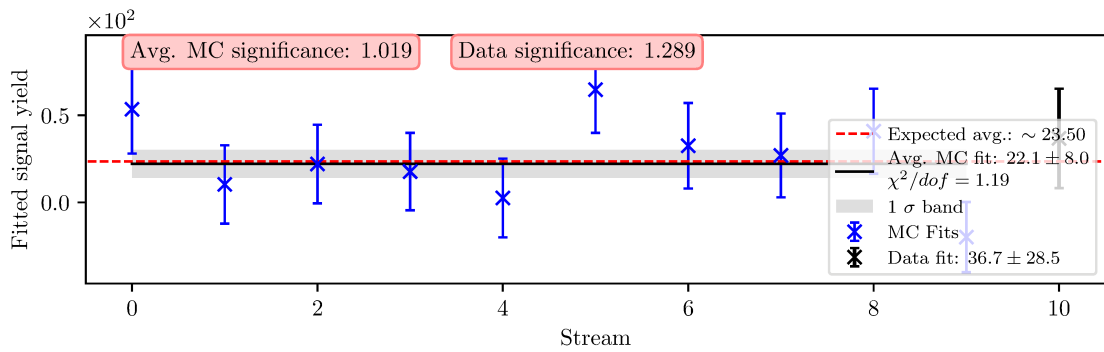


Figure C.19: Signal fit result for the 7<sup>th</sup>  $q^2$  window for MC and data in the range  $9.000 < q^2 < 10.500$ .

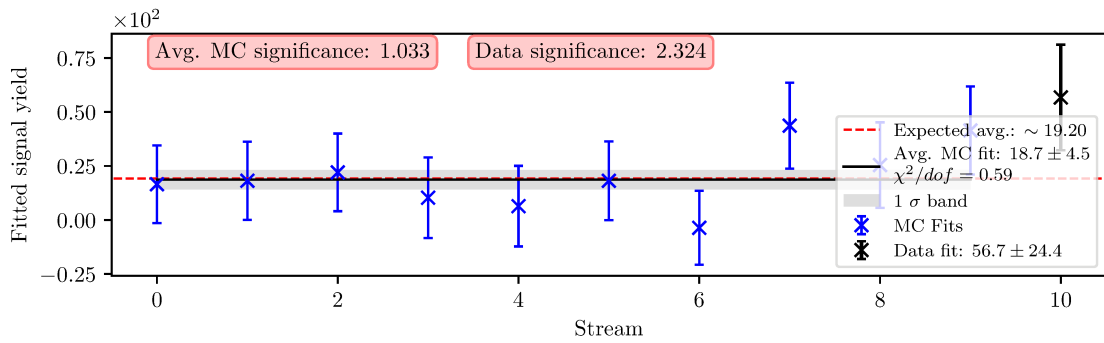


Figure C.20: Signal fit result for the 8<sup>th</sup>  $q^2$  window for MC and data in the range  $10.500 < q^2 < 12.000$ .

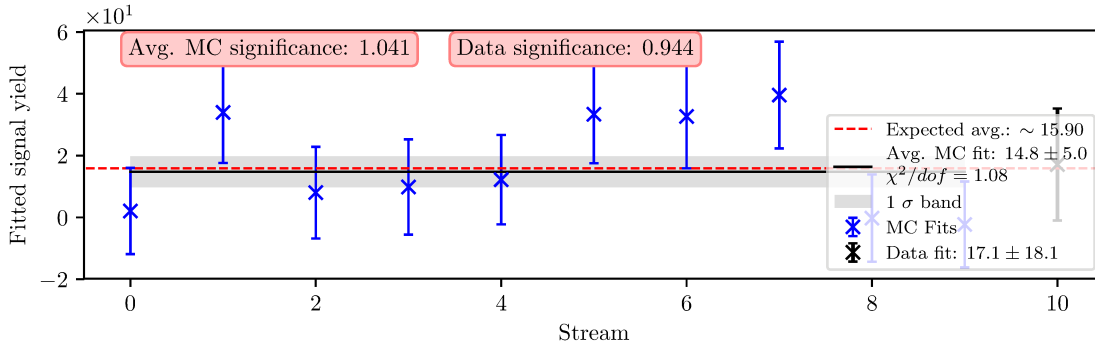


Figure C.21: Signal fit result for the 9<sup>th</sup>  $q^2$  window for MC and data in the range  $12.000 < q^2 < 13.500$ .

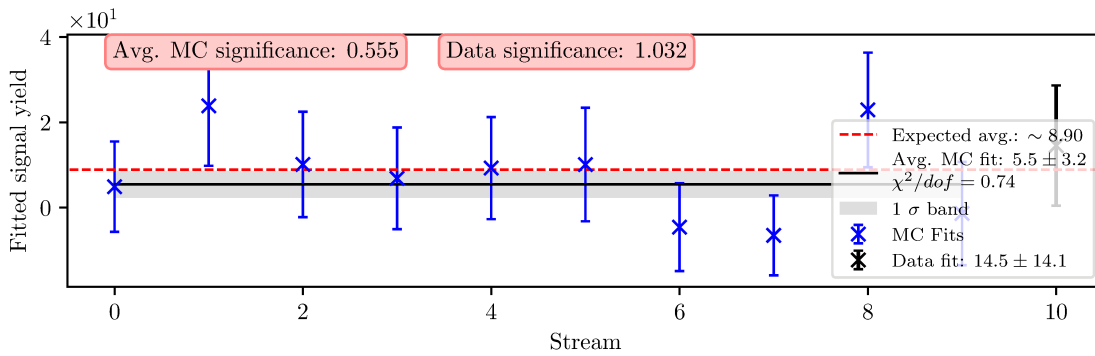


Figure C.22: Signal fit result for the 10<sup>th</sup>  $q^2$  window for MC and data in the range  $13.500 < q^2 < 15.000$ .

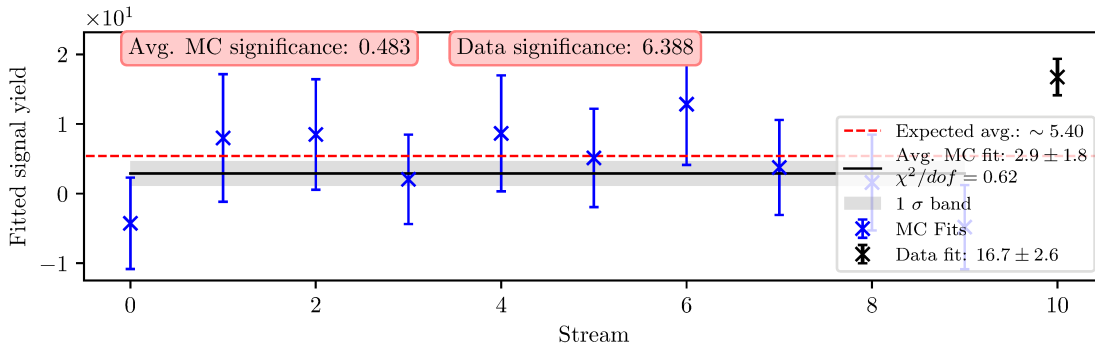


Figure C.23: Signal fit result for the 11<sup>th</sup>  $q^2$  window for MC and data in the range  $15.000 < q^2 < 16.500$ .

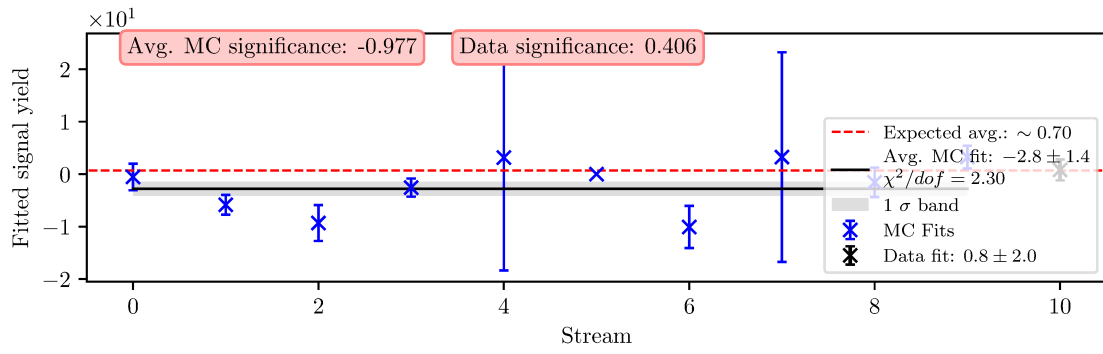


Figure C.24: Signal fit result for the 12<sup>th</sup>  $q^2$  window for MC and data in the range  $16.500 < q^2 < 18.000$ .





# List of Figures

1.1	A schematic representation of particles in the Standard Model. . . .	2
1.2	The unitarity triangle in the Wolfenstein parametrization [7]. . . .	3
1.3	Feynman diagrams for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay (left) and the $B^+ \rightarrow K^- K^+ \ell^+ \nu_\ell$ decay (right). . . . .	5
2.1	Invariant mass of the $KK$ pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structure, while $KK$ pairs from the $X_u^0$ state are more abundant and follow a wider and smoother distribution. . . . .	8
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. . . . .	12
3.2	Schematic layout of the Belle detector with the corresponding sub-detectors [14]. . . . .	13
3.3	Front (left) and side (right) view of the SVD detector with the SVD2 configuration. The front view also shows the inner wires of the Central Drift Chamber [15]. . . . .	14
3.4	Impact parameter resolutions of $z$ (left) and $r\phi$ (right) coordinates for the SVD1 and SVD2 configuration of the vertex detector [15]. . .	15
3.5	Cell structure of CDC [14]. . . . .	16
3.6	Measured $dE/dx$ as a function of particle momentum. The red lines show the expected distribution for different types of particles [14]. . .	17
3.7	Mass distribution from TOF measurements for particle momenta below 1.2 GeV/c [14]. . . . .	18
3.8	$\pi^\pm/K^\pm$ separation by TOF [14]. . . . .	18
3.9	Cross-sectional view of the Belle detector with some of the highlighted detector components. The barrel ACC is located between the CDC and TOF, while the endcap ACC is located in the forward endcap, behind the ECL crystals [14]. . . . .	19
3.10	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 left plot [14]. . . . .	20
3.11	Overall configuration of the ECL [14]. . . . .	21
3.12	Distribution of the energy deposit by electrons and charged pions at 1 GeV/c momentum [14]. . . . .	22
3.13	Reconstructed energy distribution for $e^+e^- \rightarrow \gamma\gamma$ events for overall, barrel, forward, and backward end-cap ECL [14]. . . . .	23
3.14	Cross-section of an RPC superlayer, consisting of an RPC pair [14]. .	24

3.15	Number of neutral clusters per event in KLM (left) and muon detection efficiency as a function of momentum in KLM (right) [14]. . . . .	25
3.16	The Level-1 trigger system for the Belle detector [14]. . . . .	26
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. . . . .	29
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. . . . .	30
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. . . . .	31
5.1	Normalized properties of true (blue), fake (red) electrons from any source, and true electrons from signal $B$ candidates (green). . . . .	34
5.2	Normalized properties of true (blue), fake (red) muons from any source, and true muons from signal $B$ candidates (green). . . . .	35
5.3	$FOM$ optimizations of the PID selection for true electrons (left) and true electrons from signal $B$ candidates (right). . . . .	36
5.4	$FOM$ optimizations of the PID selection for true muons (left) and true muons from signal $B$ candidates (right). . . . .	37
5.5	Normalized properties of true (blue), fake (red) and true kaons (green) from signal $B$ candidates. . . . .	38
5.6	$FOM$ optimizations of the PID selection 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. . . . .	39
5.7	Normalized vertex fit probability distribution for signal and background $B$ meson candidates in the logarithmic scale (left) and the $FOM$ optimization of the vertex fit probability (right) for the subset of the full Belle MC sample. . . . .	40
5.8	The normalized $\cos\theta_{BY}$ distribution for signal and background $B$ meson candidates (left) and the $FOM$ optimization of the $\cos\theta_{BY}$ variable (right) for the subset of the full Belle MC. . . . .	41
5.9	$m_{miss}^2$ distribution for signal and various types of background. All signal (green) and perfect signal (black) are scaled up equally. . . . .	44
5.10	Distributions of $\Delta E$ (left) and $M_{BC}$ (right) for signal and major types of background after the pre-selection. 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. . . . .	45

5.11	Distributions of $q^2$ (left) and $q^2$ resolution (right) for various methods of $q^2$ calculation. The green distribution follows the procedure in [19], the blue distribution takes into account the weighted average of the $B$ meson direction [20], and the red and orange distributions are straight-forward calculations with available information in the reconstruction. The $q^2$ calculation in red assumes a resting $B$ meson in the CMS frame, and the calculation in orange uses the neutrino four-momentum from Eq. (5.10). . . . .	46
5.12	The $FOM$ optimization of the vertex fit probability (left) and the $\cos\theta_{BY}$ variable (right) for the full Belle MC sample. . . . .	47
5.13	2D $FOM$ optimization of the signal region definition, where most of the perfectly reconstructed candidates are located. . . . .	48
5.14	$FOM$ optimization of the $m_{miss}^2$ selection in the signal region. . . . .	48
5.15	Signal distributions of $\Delta E$ and $M_{BC}$ based on the charge product of both $B$ mesons in the event. The plots on the left show the distributions in an arbitrary scales, while the plots on the right show the normalized distributions. . . . .	50
6.1	A schematic representation of decision trees in a decision forest. . . . .	54
6.2	Classifier output of the $\pi^0$ training for signal and background $\pi^0$ candidates. . . . .	55
6.3	Classifier output of the $\gamma$ training for signal and background $\gamma$ candidates. . . . .	57
6.4	The $FOM$ of the classifier output optimization (left) and momentum magnitude in the LAB frame (right) of signal and background photon candidates before (dashed) and after (solid) the optimal selection. . . . .	58
6.5	(a) Tracks from long-lived neutral particles, which decay away from the IP region, (b) random reconstructed background tracks, (c) low-momentum particles, which curl in the magnetic field, (d) in-flight decays of particles, which produce a "kink" in the trajectory. . . . .	59
6.6	Invariant mass of the $K_S^0$ candidates before (dashed) and after (solid) the selection 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. . . . .	60
6.7	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). . . . .	61
6.8	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 selection on $p_T$ from the plot on the left. . . . .	61
6.9	Distribution of the angle between two tracks in a single pair after applying the selection defined in this section. 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). . . . .	62

6.10	Classifier output of the track pair training for random track pairs and duplicate track pairs. . . . .	63
6.11	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). . .	64
6.12	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). . . . .	66
6.13	$\Delta E$ 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 selection criteria, 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. . . . .	69
6.14	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. . . . .	70
6.15	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. . . . .	71
6.16	Normalized distributions of $m_{KK}$ for the full MC dataset. The red lines represent the edges of the selection in the $m_{KK}$ distribution where the control sample is enhanced. The $m_{KK}$ distribution drops quickly for the case of the control decay, while staying uniform for other contributions. . . . .	72
6.17	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. . . . .	73
6.18	Distributions of $\Delta E$ (top) and $M_{BC}$ (bottom) split in bins of the charge product of the two $B$ mesons. . . . .	74
7.1	Invariant mass of $KK$ candidates over a wider region (top) and for a specific region around the $\phi$ peak (bottom left), the $D^0$ peak (bottom center), and again for the $D^0$ peak, but with the $m_{K\pi}$ mass of the $KK$ candidates (bottom right). Signal (green) and perfect signal (black) are equally scaled up. . . . .	76
7.2	Distributions of $ \cos \theta_T $ (left) and $ \cos \theta_{T,B} $ (right) for $B$ meson candidates from various sources. . . . .	78

7.3	Concept of CLEO cones. $\vec{T}$ denotes the thrust axis of the $B$ meson candidate in an event. Each variable corresponds to a momentum flow around the thrust axis in steps of $10^\circ$ . . . . .	79
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. . . . .	81
7.5	Distribution of $m_{miss}^2$ for partially reconstructed $B^0 \rightarrow D^{*-} \ell^+ \nu$ decays.	83
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. . . . .	84
7.7	2D $FOM$ optimization of continuum suppression classifier and the standard $BDT$ (left) and $uBDT$ (right) $B\bar{B}$ suppression classifier. . . . .	85
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. . . . .	86
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. . . . .	87
7.10	$\Delta E$ (left), $M_{BC}$ (right) and $m_{KK}$ (bottom) for major contributions to the $B\bar{B}$ background in the signal region. $\Upsilon(4S)$ -matched backgrounds represents the majority, but have a smooth and wide distribution, distinguishable from signal. $B$ -matched contributions show a peak in $M_{BC}$ and sometimes in $\Delta E$ , but can be constrained using existing measurements. . . . .	88
7.11	2D $FOM$ for optimal $eID$ and $\mu ID$ selection 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 misidentified as the opposite-sign kaon in the reconstruction chain. . . . .	89
7.12	$\Delta E$ (left), $M_{BC}$ (right) and $m_{KK}$ (bottom) for major contributions to the $B\bar{B}$ background in the signal region after the lepton veto. The double semileptonic background component is suppressed by a factor of 4 – 5. . . . .	90
7.13	$\Delta E$ (left), $M_{BC}$ (right) and $m_{KK}$ (bottom) for major contributions to the $B\bar{B}$ background in the control region after the lepton veto. The major component in this case are other $B \rightarrow D^* \ell + \nu$ , $D \rightarrow K^+ K^-$ decays, besides the control decay. . . . .	91
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 prior to any MVA selection. . . . .	92
7.15	$\Delta E$ (top left), $M_{BC}$ (top right), the $q\bar{q}$ - (bottom left) and the $B\bar{B}$ classifier output (bottom right), for on-resonance data and MC in the control region prior to the MVA selection. . . . .	93

8.1	Likelihood ratio test of an additional smearing and offset parameter to MC (left) and data (right). . . . .	99
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. . . . .	103
8.3	Toy MC fits of pseudo-data showing the fit yield (top), the fit errors (center) and the pull distribution of the fits (bottom). . . . .	104
8.4	The mean fit yield and expected yield difference (top), the mean pull (center) and the mean significance (bottom) as a function of the signal fraction. . . . .	105
9.1	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 <i>BDT</i> classifier (blue) and the <i>uBDT</i> classifier (orange). . . . .	108
9.2	An example fit to one stream of MC. Left column shows the $\Delta E$ and the right column shows the $M_{BC}$ distribution of the full fitted sample in the full fit region (top) and the in signal enhanced region (bottom). . . . .	109
9.3	Fits to all 10 streams of MC and the global fit with a 0 <sup>th</sup> degree polynomial. The red line shows the mean value of the global fit and the gray band shows the $1\sigma$ confidence interval. . . . .	110
9.4	Control fit result on one stream of MC. Left column shows the $\Delta E$ and the right column shows the $M_{BC}$ distribution in the full fit window (top) and the in signal enhanced window (bottom). . . . .	111
9.5	Control fit result on real data. Left column shows the $\Delta E$ and the right column shows the $M_{BC}$ distribution in the full fit window (top) and the signal enhanced window (bottom). . . . .	112
9.6	Control fit to the data and all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0 <sup>th</sup> degree polynomial. The gray band shows the $1\sigma$ confidence interval around the global MC fit. . . . .	113
9.7	Various branching fraction determinations for the control decay of our analysis. . . . .	115
9.8	Signal fit result on real data. Left column shows the $\Delta E$ and the right column shows the $M_{BC}$ distribution in the full fit window (top) and in the signal enhanced window (bottom). . . . .	116
9.9	Signal fit to the data and all 10 streams of MC. The red line shows the mean value of the global MC fit with a 0 <sup>th</sup> degree polynomial. The gray band shows the $1\sigma$ confidence interval around the global MC fit. . . . .	117
9.10	Signal yield distribution as a function of $m_{KK}$ . . . . .	120
9.11	Signal yield distribution as a function of $q^2$ . . . . .	121
10.1	Fit yields, their ratios and ratios of selection efficiencies (double ratios) for the control sample fits in data and on MC. . . . .	127

10.2	$m_{KK}$ (top left), $q^2$ (top right), $\Delta E$ (bottom left), and $M_{BC}$ (bottom right) for the main and the three extreme cases of signal MC generator choice. . . . .	128
10.3	Efficiency as a function of $q^2$ (top) and $m_{KK}$ (bottom) for the main and three extreme cases of signal MC generator choice. . . . .	130
11.1	The profile likelihood function (left) and the profile negative-log-likelihood function (right) for the case with statistical error only (blue) and convoluted with a Gaussian function in order to incorporate the systematic uncertainties on the signal yield (orange). . . . .	134
12.1	Unitarni trikotnik, prikazan v Wolfensteinovi parametrizaciji [7]. . . . .	136
12.2	Feynmanovi diagrami za razpada $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ (levo) in $B^+ \rightarrow K^- K^+ \ell^+ \nu_\ell$ (desno). . . . .	137
12.3	Shematski prikaz detektorja Belle in ustreznih podsistemov [14]. . . . .	138
12.4	Porazdelitev mase dveh kaonov ( $m_{KK}$ ) za signalni in kontrolni razpad. Porazdelitev $m_{KK}$ kontrolnega razpada je prisotna samo v območju mase mezona $D^0$ , medtem ko je porazdelitev $m_{KK}$ signalnega razpada prisotna po širšem območju. . . . .	140
12.5	Primerjava očiščenega in neočiščenega dogodka za porazdelitvi $\Delta E$ in $M_{BC}$ . Porazdelitve iz očiščenega dogodka so bolj ostre in predstavljajo boljšo možnost za ločitev od ozadja. . . . .	142
12.6	Porazdelitvi spremenljivk $\Delta E$ in $M_{BC}$ za končni signalni vzorec. . . . .	144
12.7	Porazdelitvi spremenljivk $\Delta E$ in $M_{BC}$ za končni kontrolni vzorec. . . . .	144
12.8	Primer luščenja števila kontrolnih kandidatov na pravih podatkih. Lev stolpec prikazuje $\Delta E$ , desni pa $M_{BC}$ , medtem ko zgornja vrstica prikazuje porazdelitvi na celotnem definiranem območju, spodnja pa projekcija na ozko okno okoli vrha signalne porazdelitve. . . . .	146
12.9	Število kontrolnih kandidatov za vseh 10 vzorcev MC podatkov in njihovo uteženo povprečje, ter za izmerjene podatke. . . . .	147
12.10	Rezultati meritev razvejitenih razmerij kontrolnega razpada za MC in za podatke za različna končna leptonska stanja. . . . .	148
12.11	Primer luščenja števila signalnih kandidatov na podatkih. Lev stolpec prikazuje $\Delta E$ , desni pa $M_{BC}$ , medtem ko zgornja vrstica prikazuje porazdelitvi na celotnem definiranem območju, spodnja pa projekcija na ozko okno okoli vrha signalne porazdelitve. . . . .	150
12.12	Število kontrolnih kandidatov za vseh 10 vzorcev MC podatkov in njihovo uteženo povprečje, ter za izmerjene podatke. . . . .	151
A.1	Feature distributions for MVA training of $\pi^0$ candidates in the scope of ROE clean-up. . . . .	160
A.1	Feature distributions for MVA training of $\pi^0$ candidates in the scope of ROE clean-up. . . . .	161
A.2	Hyper-parameter optimization of <code>nTrees</code> and <code>nLevels</code> in the <i>BDT</i> forest training of $\pi^0$ candidates in the scope of the ROE clean-up. . . . .	162
A.3	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. . . . .	162
A.4	ROC curves of the MVA classifier output for $\pi^0$ candidates training on the train (solid) and test (dashed) samples. . . . .	163

A.5	Feature distributions for MVA training of $\gamma$ candidates in the scope of ROE clean-up. . . . .	164
A.6	Hyper-parameter optimization of <code>nTrees</code> and <code>nLevels</code> in the <i>BDT</i> forest training of $\gamma$ candidates in the scope of the ROE clean-up. . .	165
A.7	Efficiency ( $\mathcal{E}$ ) and purity ( $\mathcal{P}$ ) of the MVA classifier output for $\gamma$ candidates training on the train (solid) and test (dashed) samples. . .	165
A.8	ROC curves of the MVA classifier output for $\gamma$ candidates training on the train (solid) and test (dashed) samples. . . . .	166
A.9	Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up. . . . .	168
A.9	Feature distributions for MVA training of duplicate track pair candidates in the scope of ROE clean-up. . . . .	169
A.10	Hyper-parameter optimization of <code>nTrees</code> and <code>nLevels</code> in the <i>BDT</i> forest training of duplicate track pair candidates in the scope of the ROE clean-up. . . . .	170
A.11	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. . . . .	170
A.12	ROC curves of the MVA classifier output for duplicate track pair candidates training on the train (solid) and test (dashed) samples. . .	171
A.13	Feature distributions for MVA training of duplicate track candidates in the scope of ROE clean-up. . . . .	173
A.14	Hyper-parameter optimization of <code>nTrees</code> and <code>nLevels</code> in the <i>BDT</i> forest training of duplicate track candidates in the scope of the ROE clean-up. . . . .	174
A.15	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. . . . .	174
A.16	ROC curves of the MVA classifier output for duplicate track candidates training on the train (solid) and test (dashed) samples. . . . .	175
B.1	Feature distributions for MVA training of $q\bar{q}$ background suppression. . . . .	179
B.1	Feature distributions for MVA training of $q\bar{q}$ background suppression. . . . .	180
B.2	Hyper-parameter optimization of <code>nTrees</code> and <code>nLevels</code> in the <i>BDT</i> forest training of $q\bar{q}$ background suppression. . . . .	181
B.3	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. . . . .	181
B.4	ROC curves of the MVA classifier output for $q\bar{q}$ background suppression training on the train (solid) and test (dashed) samples. . . . .	182
B.5	Feature distributions for MVA training of $B\bar{B}$ background suppression. . . . .	184
B.6	Hyper-parameter optimization of <code>nTrees</code> and <code>nLevels</code> in the <i>BDT</i> forest training of $B\bar{B}$ background suppression. . . . .	185
B.7	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. . . . .	185
B.8	ROC curves of the MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples. . . . .	186



B.9	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. . . . .	187
B.10	ROC curves of the uniformity boosted MVA classifier output for $B\bar{B}$ background suppression training on the train (solid) and test (dashed) samples. . . . .	187
C.1	Signal fit result for the 1 <sup>st</sup> $m_{KK}$ window for MC and data in the range $0.980 < m_{KK} < 1.232$ . . . . .	189
C.2	Signal fit result for the 2 <sup>nd</sup> $m_{KK}$ window for MC and data in the range $1.232 < m_{KK} < 1.483$ . . . . .	189
C.3	Signal fit result for the 3 <sup>rd</sup> $m_{KK}$ window for MC and data in the range $1.483 < m_{KK} < 1.735$ . . . . .	190
C.4	Signal fit result for the 4 <sup>th</sup> $m_{KK}$ window for MC and data in the range $1.735 < m_{KK} < 1.987$ . . . . .	190
C.5	Signal fit result for the 5 <sup>th</sup> $m_{KK}$ window for MC and data in the range $1.987 < m_{KK} < 2.238$ . . . . .	190
C.6	Signal fit result for the 6 <sup>th</sup> $m_{KK}$ window for MC and data in the range $2.238 < m_{KK} < 2.490$ . . . . .	191
C.7	Signal fit result for the 7 <sup>th</sup> $m_{KK}$ window for MC and data in the range $2.490 < m_{KK} < 2.742$ . . . . .	191
C.8	Signal fit result for the 8 <sup>th</sup> $m_{KK}$ window for MC and data in the range $2.742 < m_{KK} < 2.993$ . . . . .	191
C.9	Signal fit result for the 9 <sup>th</sup> $m_{KK}$ window for MC and data in the range $2.993 < m_{KK} < 3.245$ . . . . .	192
C.10	Signal fit result for the 10 <sup>th</sup> $m_{KK}$ window for MC and data in the range $3.245 < m_{KK} < 3.497$ . . . . .	192
C.11	Signal fit result for the 11 <sup>th</sup> $m_{KK}$ window for MC and data in the range $3.497 < m_{KK} < 3.748$ . . . . .	192
C.12	Signal fit result for the 12 <sup>th</sup> $m_{KK}$ window for MC and data in the range $3.748 < m_{KK} < 4.000$ . . . . .	193
C.13	Signal fit result for the 1 <sup>st</sup> $q^2$ window for MC and data in the range $0.000 < q^2 < 1.500$ . . . . .	193
C.14	Signal fit result for the 2 <sup>nd</sup> $q^2$ window for MC and data in the range $1.500 < q^2 < 3.000$ . . . . .	193
C.15	Signal fit result for the 3 <sup>rd</sup> $q^2$ window for MC and data in the range $3.000 < q^2 < 4.500$ . . . . .	194
C.16	Signal fit result for the 4 <sup>th</sup> $q^2$ window for MC and data in the range $4.500 < q^2 < 6.000$ . . . . .	194
C.17	Signal fit result for the 5 <sup>th</sup> $q^2$ window for MC and data in the range $6.000 < q^2 < 7.500$ . . . . .	194
C.18	Signal fit result for the 6 <sup>th</sup> $q^2$ window for MC and data in the range $7.500 < q^2 < 9.000$ . . . . .	195
C.19	Signal fit result for the 7 <sup>th</sup> $q^2$ window for MC and data in the range $9.000 < q^2 < 10.500$ . . . . .	195
C.20	Signal fit result for the 8 <sup>th</sup> $q^2$ window for MC and data in the range $10.500 < q^2 < 12.000$ . . . . .	195

C.21	Signal fit result for the 9 <sup>th</sup> $q^2$ window for MC and data in the range 12.000 < $q^2$ < 13.500. . . . .	196
C.22	Signal fit result for the 10 <sup>th</sup> $q^2$ window for MC and data in the range 13.500 < $q^2$ < 15.000. . . . .	196
C.23	Signal fit result for the 11 <sup>th</sup> $q^2$ window for MC and data in the range 15.000 < $q^2$ < 16.500. . . . .	196
C.24	Signal fit result for the 12 <sup>th</sup> $q^2$ window for MC and data in the range 16.500 < $q^2$ < 18.000. . . . .	197

# List of Tables

2.1	Relative branching fractions of $B \rightarrow KKX\ell\nu$ decays by channel. . . .	9
3.1	Cross-sections for various physics processes at $\Upsilon(4S)$ resonance energy [14]. . . . .	12
6.1	Ratios of correctly classified and misclassified tracks. . . . .	67
6.2	Photon selection for the Belle clean-up procedure. Different selection criteria are applied on photons in different parts of the detector. . . .	67
6.3	Track selection for the Belle clean-up procedure. Different selection criteria are applied to tracks in different $p_T$ regions. . . . .	67
6.4	Comparison of efficiencies and $FWHM$ 's of ROE and Belle clean-up setups with respect to the default case (no clean-up) for the perfect signal. . . . .	69
7.1	Efficiencies after selecting the signal or control region . . . . .	76
8.1	Well defined decay channels used for constraining the control fits. . .	97
8.2	MC and measured values of branching fractions along with the calculated correction factors used for constraining the control fit. . . . .	98
8.3	Well defined decay channels used for constraining the signal fits. . . .	101
8.4	Additional MC and measured values of $D$ meson branching fractions along with the calculated correction factors used for constraining the signal fit. . . . .	102
9.1	Control sample fit results for MC and data for various lepton final state modes. . . . .	110
9.2	Control sample fit results for MC and data for various lepton final state modes. . . . .	114
9.3	Yields of all signal fit contributions in data. . . . .	117
9.4	Mean values and standard deviations of constraints after the fit. . . .	118
9.5	Signal decay branching fraction results on MC and in data. . . . .	119
9.6	Various signal fit yields for each of the defined $m_{KK}$ windows. . . . .	120
9.7	Various signal fit yields for each of the defined $q^2$ windows. . . . .	121
10.1	PID correction factors and systematic uncertainties for various charged particles and their combinations. . . . .	124
10.2	Summary of systematic uncertainties in this analysis. . . . .	131
12.1	Rezultati luščenja števila kontrolnih kandidatov za različna končna leptonska stanja. . . . .	145

12.2	Števila kandidatov vseh prispevkov, določenih z luščenjem na signalnem vzorcu. . . . .	149
12.3	Vrednost razvejitenega razmerja signalnega razpada za MC in za podatke. . . . .	151
12.4	Povzetek sistematskih negotovosti te analize. . . . .	153
A.1	Variable names, aliases and importance in the scope of $\pi^0$ MVA training for ROE clean-up. . . . .	160
A.2	Variable names, aliases and importance in the scope of $\gamma$ MVA training for ROE clean-up. . . . .	163
A.3	Variable names, aliases and importance in the scope of duplicate track pair MVA training for ROE clean-up. . . . .	167
A.4	Variable names, aliases and importance in the scope of duplicate track MVA training for ROE clean-up. . . . .	172
B.1	Variable names, aliases and importance in the scope of $q\bar{q}$ suppression MVA training. . . . .	178
B.2	Variable names, aliases and importance in the scope of $B\bar{B}$ background suppression. . . . .	183