

BELLE NOTE 1486

Measurement of the decay
 $B^+ \rightarrow K^+ K^- \ell^+ \nu_\ell$ with B2BII

Matic Lubej

Ljubljana, 2018

⁶ Changelog

- ⁷ • First submission of the note.

Contents

	Page
1 Introduction	1
2 Data and Monte-Carlo samples	6
2.1 Signal MC production	7
3 B2BII conversion	10
3.1 Validation	11
4 Event reconstruction	12
4.1 Final state particles selection	12
4.2 Combination of FSP particles	19
4.3 Loose neutrino reconstruction	21
4.4 Event categorization	25
4.4.1 Hadronic decay MVA training	26
4.5 Signal region definition	28
4.6 Selection summary	29
5 Rest of event clean-up	30
5.1 Setting up the MVA	30
5.2 Clusters clean-up	31
5.2.1 π^0 MVA training	31
5.2.2 γ MVA training	33
5.3 Tracks clean-up	35
5.3.1 Tracks from long-lived particles	37
5.3.2 Duplicate tracks	38
5.4 Belle clean-up	46
5.5 Clean-up results	47
5.6 ROE clean-up validation	50
6 Background suppression	51
6.1 Control decay and other resonant background	51
6.2 Continuum suppression	52
6.2.1 Characteristic variables	53
6.2.2 MVA training	55

38	6.3	$B\bar{B}$ suppression	57
39	6.3.1	Boosting to uniformity	59
40	6.4	Selection optimization	60
41	6.5	Data and MC agreement	63
42	7	Signal extraction	67
43	7.1	Fit templates	68
44	7.2	Adaptive binning algorithm	69
45	7.3	Signal MC fit results	71
46	7.3.1	Toy MC experiment	74
47	7.3.2	Toy MC linearity test	75
48	7.4	Control fit results	76
49	7.4.1	Fit yields and yield ratios	77
50	7.4.2	Double ratios	77
51	7.4.3	Branching ratio measurement and comparison	77
52	7.5	Signal fit to data	77
53	8	Systematics	78
54	8.1	Model uncertainty effects	78
55	8.2	PID efficiency correction	78
56	8.3	Bias	78

Chapter 1

Introduction

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

Combinations of quarks such as $q_1 q_2 q_3$ (hadrons) or $q_1 \bar{q}_2$ (mesons) can make up heavier particles that we see today. Such particles are protons and neutrons, but also heavier particles which can be produced in processes involving very high energies. Such heavy particles are unstable and decay into lighter ones via forces of nature. Together with the elementary particles and force carriers, three out of four of these forces are joined in a theoretical model called the Standard Model (SM), which is shown in Figure 1.1. They are the electromagnetic, weak nuclear and strong nuclear force. Gravity is not included in the current version of the Standard Model due to its nature. By researching such processes in large experiments it is possible to study the mechanism of how elementary particles interact. By doing so we are able to learn secrets of the universe and how it all began.



Figure 1.1: A schematic of the Standard Model.

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

One of the most important properties of the weak nuclear force is that it can change the flavor of particles. Flavor is a quantum number which is conserved for each type of quark, so changing a flavor of a quark means changing the quark itself. Such processes are forbidden for the electromagnetic and the strong nuclear force, but not for the weak one. All of the information regarding quark transitions and transition probabilities can be merged into a form of a complex matrix called the Cabibbo-Kobayashi-Maskawa (CKM) matrix [1, 2]

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

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

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

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

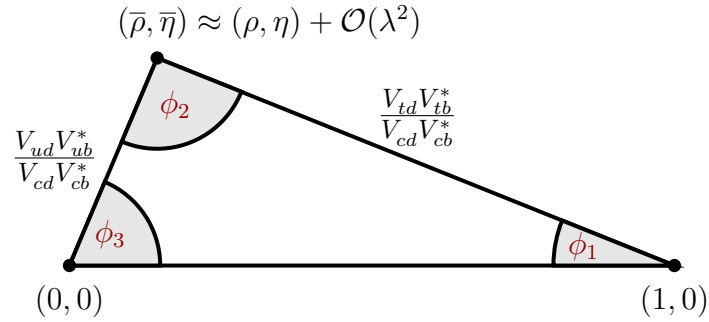


Figure 1.2: The unitarity triangle with λ , η , ρ and A (not shown) as free parameters of the CKM matrix.

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

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

where X_u represents a charmless hadron with a u quark and ℓ one of the charged leptons e , μ or τ . Measuring the decay rate of the B meson in such decays offers the way for

121 the CKM matrix element determination. Decay rates are directly connected to the V_{ub}
 122 element as

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

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

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

140 In the inclusive method one studies the decays of B mesons to any charmless hadronic
 141 final state $B \rightarrow X_u \ell \nu$. In this case, the total decay rate for $b \rightarrow u \ell \nu$ can be calculated
 142 accurately, since hadronization does not have to be taken into account. The greater
 143 challenge with this method is again the experimental measurement of the total decay
 144 rate due to the $B \rightarrow X_c \ell \nu$ background. Experimental sensitivity to V_{ub} is highest where
 145 $B \rightarrow X_c \ell \nu$ decays are less dominant. Theory and experiment have to compromise and
 146 limit the V_{ub} determination to a region where the signal-to-background ratio is good.
 147 Theory takes this into account by reliably calculating the partial decay rate $\Delta\Gamma$, which
 148 is more challenging than the total decay rate. One possible and relevant approach to
 149 reduce $b \rightarrow c$ background is to reject all events with K particles, or kaons, in the final
 150 state. The procedure is also called a K -veto. Kaons consist of an s quark, which is
 151 mainly produced in $c \rightarrow s$ transitions. This means that if a kaon is found, it is very
 152 likely that it originates from a particle with a c quark.

153 If V_{ub} is determined with both these methods, the values can be compared. It turns out
 154 that consistency between these two results is only marginal, where the difference is at
 155 a level of 3σ . In fact, inclusive measurements prefer higher values than exclusive ones.
 156 This is known as the V_{ub} puzzle. It is necessary to make further research as to why this
 157 difference occurs. The reason could be an unknown experimental or theoretical error, or

it is even possible that some NP contributions occur. Part of the reason 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. In this analysis 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. The results of this analysis should help us make a step closer to solving the V_{ub} puzzle.

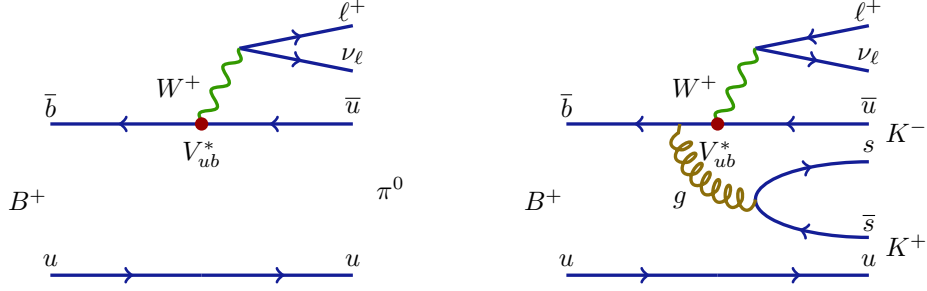


Figure 1.3: Feynman diagrams for the $B^+ \rightarrow \pi^0 \ell^+ \nu_\ell$ decay (left) and the $B^+ \rightarrow K^- K^+ \ell^+ \nu_\ell$ decay (right).

Even though the Belle experiment finished collecting data in 2010, the data is still relevant and available for analysis. On the other hand, Belle II experiment has entered its initial stages of operation. A large part of Belle II software is already available and working, so the practical part of this analysis is done in the framework of the Belle II experiment, called Belle II Analysis Framework (BASF2). The Belle II detector wasn't able to collect a significant amount of data for this analysis, however, the Belle II collaboration developed a data converter for Belle to Belle II format conversion, making it possible to perform analysis of Belle data with Belle II software and enables us to make a relevant contribution despite the transitional period.

Specifically, we will be focusing on decays of the charged B mesons of the form $B^+ \rightarrow K^+ K^- \ell^+ \nu$, since it includes two charged kaons, as opposed to the case of the neutral B meson decay. The reason for this is a simpler decay chain and a higher reconstruction efficiency. All further occurrences of $B \rightarrow KK\ell\nu$ automatically imply decays of the form $B^+ \rightarrow K^+ K^- \ell^+ \nu$ and its charge conjugate 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, which corresponds to about $771 \times 10^6 B\bar{B}$ meson pairs. Additionally, several streams of Monte-Carlo (MC) samples were produced, where each stream of MC corresponds to the same amount of data that was taken with the detector. The main focus of this and other similar analyses is to study a rare signal decay, which means that the amount of such decays in the existing MC is not abundant enough. In such cases, it is a common practice to produce specific samples of signal MC, where the abundance of signal decays is much larger, enabling us to study its properties in greater detail.

The following samples were used in this analysis

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

below $\Upsilon(4S)$ resonance energy.

2.1 Signal MC production

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

The relevant processes which contribute to our signal decay are

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

where a_{00} , a_{20} , f_2 and f_0 are some of the light unflavored states which in some cases further decay into a $K^+ K^-$ pair, while X_u^0 represents a generic $u\bar{u}$ quark pair, which further hadronizes based on the PYTHIA quark hadronization model [3]. 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 structure, while KK pairs from the X_u^0 state are more frequent and follow a wider and smoother distribution.

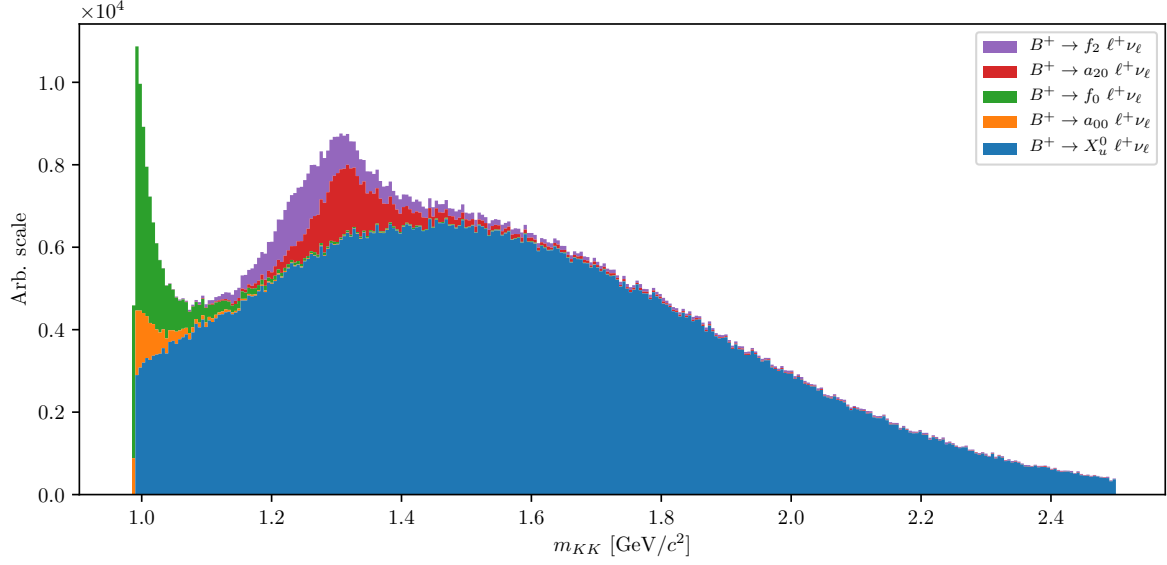


Figure 2.1: Invariant mass of the KK pair from various contributions of the MC generator. The light unflavored states have small contributions with resonant structure, while KK pairs from the X_u^0 state are more frequent and follow a wider and smoother distribution.

The produced signal MC sample contains decays of the form $B \rightarrow KK\ell\nu$ as well as $B \rightarrow KKKX\ell\nu$, where X can be any hadron as long as it satisfies all the selection rules of the decay. It is possible to calculate the MC branching ratios for each channel by making combinations of the particles directly from the generator. Table 2.1 shows some of the most prominent channels, which are similar to our signal decay, as long as their relative fraction. It is clear that our signal decay is the most abundant one, with a relative contribution of about 28.14 %, while other channels contribute only up to about 8 % or less. Additionally, our signal decay is also the cleanest, meaning that it has no neutral particles like π^0 , which are harder to reconstruct and suffer from decrease in efficiency due to reconstruction effects.

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

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

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

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

240 where ℓ is e or μ . During analysis the abundant signal MC sample is scaled down to
241 correspond to the amount of data taken with the Belle detector.

Chapter 3

B2BII conversion

The predecessor of the Belle II experiment was the Belle experiment, which finished its data taking run of 10 years at end of 2010 after collecting a dataset of about 1 ab^{-1} . That year the Belle detector was shut down and the Belle II experiment was born from the ashes, where even some of the old detector components were reused. The construction of the Belle II detector today is still an ongoing process, although first collisions are expected in 2018. By the year 2025 it is foreseen that Belle II will have recorded about 50 ab^{-1} of data, which is about 50 times more than in case of Belle. However, this is still a long time away, while the recorded dataset of the Belle experiment still has the potential for interesting physics analyses today. On the other hand, most of the people from Belle transitioned to Belle II and a lot of new institutes joined the Belle II Collaboration. This moved focus from Belle analyses and Belle Analysis Framework (BASF) to the construction of the Belle II detector and the development of Belle II Analysis Framework (BASF2), which was written completely from scratch, making the BASF2 software incompatible with the Belle data. This resulted in gradual loss of knowledge on the maintenance and operation of the BASF software.

In the Belle II Collaboration, a task force was created in order to convert Belle data into Belle II format (B2BII). The B2BII package was developed as a part of BASF2 in order to convert data and MC of the Belle experiment and make it available within BASF2. In addition to the convenience of Belle data being processed in the more intuitive and advanced BASF2 framework, B2BII allows for estimation and validation of performances of various advanced algorithms being developed for Belle II. The conversion itself, however, is considered non-trivial. Although the conversion of the raw detector data would be possible, the reconstruction algorithms of BASF2 are optimized for Belle II and cannot be effectively applied to Belle data. To bypass this problem, reconstructed objects from PANTHER tables, a custom solution of the Belle collaboration based on C/C++ and Fortran, are mapped to their corresponding representations in BASF2.

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

272 module applies various calibration factors, such as experiment and run dependent factors,
273 to the beam energy, particle identification information, error matrices of the fitted tracks,
274 etc. The module also applies some low-level cuts to reproduce removing background
275 events as done within BASF. The actual conversion and the mapping of reconstructed
276 objects is done in the last module. For more information see [4].

277 **3.1 Validation**

278 In order to make sure the conversion was successful and without errors, a thorough
279 validation should be performed. This is done by comparing histograms of all physical
280 quantities of the reconstructed objects on simulated and recorded events, processed with
281 BASF and BASF2. Figures X and X show some of the physical properties of the neutral
282 and charged particles, obtained with BASF and BASF2, and their difference. The plots
283 indicate that the conversion is successful and we can proceed with the analysis in the
284 framework of BASF2.

285 PLOT

286 PLOT

Chapter 4

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 B meson candidates.

4.1 Final state particles selection

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

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

Leptons

Figures 4.1 and 4.2 show the impact parameters d_0 and z_0 , the momentum in $\Upsilon(4S)$ center-of-mass system (CMS), and the PID information for true and fake electrons and muons, where an extra category for true electrons/muons from the signal decay is shown.

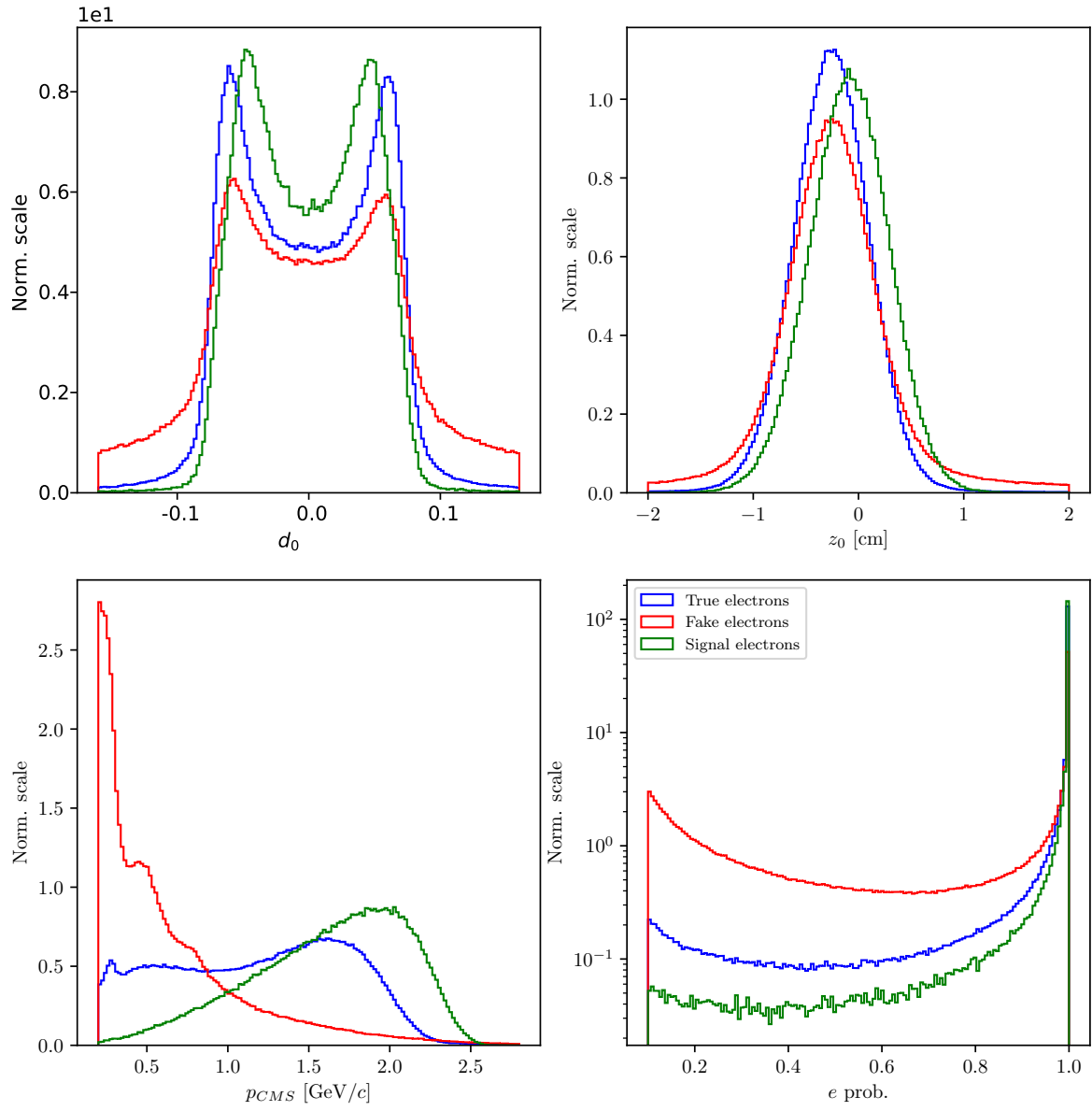


Figure 4.1: Normalized properties of true (blue), fake (red) and true electrons from signal decay (green).

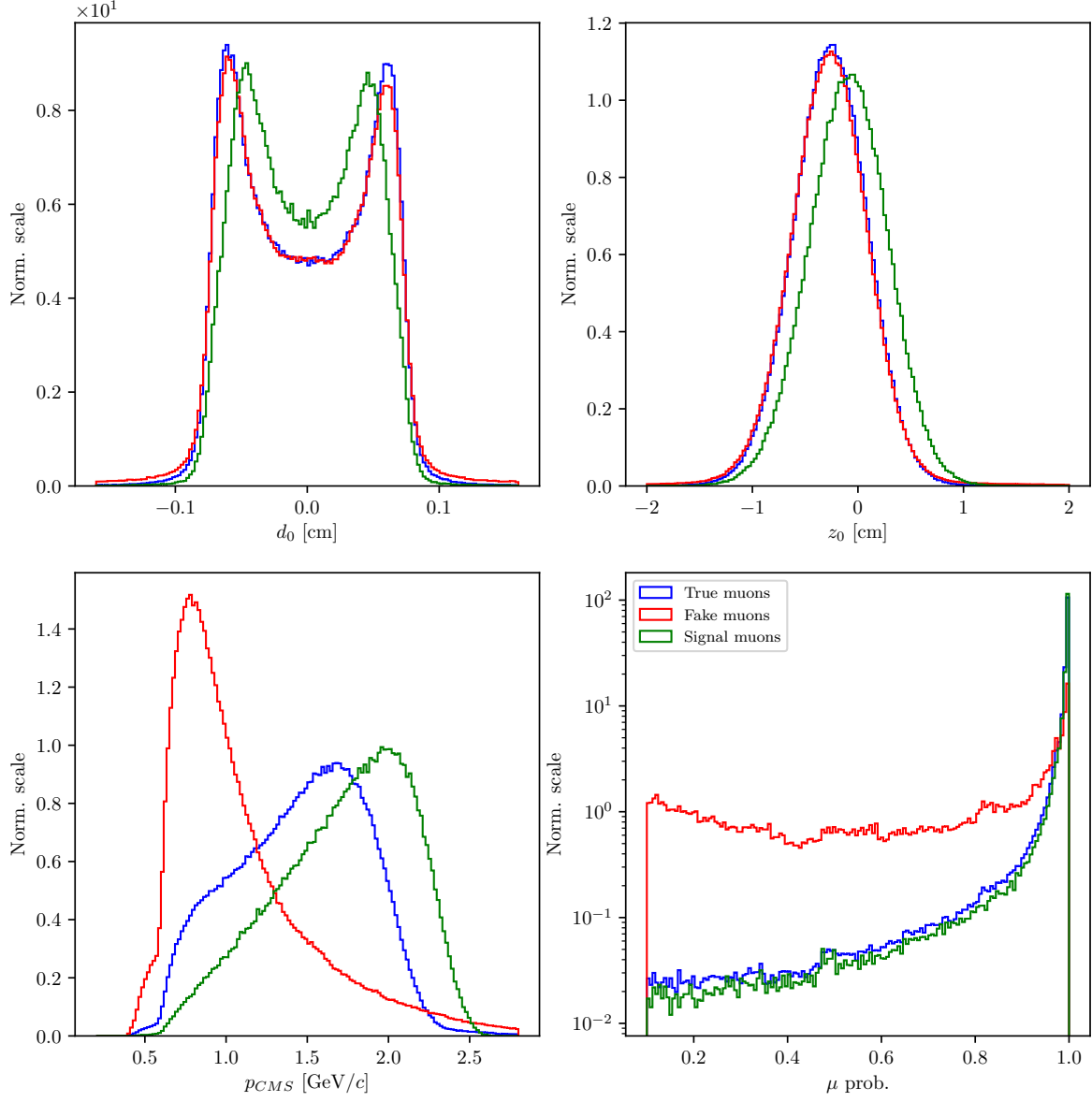


Figure 4.2: Normalized properties of true (blue), fake (red) and true muons from signal decay (green).

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

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

After this selection we can determine the optimal PID cuts for electrons and muons, where we optimize the selection by maximizing the standard definition of *figure of merit* (FOM), defined in Eq. (4.1)

$$FOM = \frac{S}{\sqrt{S+B}}. \quad (4.1)$$

The definition of S and 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 particles, while in FOM_2 the signal S represents correctly reconstructed particles, which come from a correctly reconstructed B meson candidate. In both cases B represents all other particle candidates which do not satisfy the conditions of S .

The FOM plots are shown in Figures 4.3 and 4.4. The cuts values are based on PID cuts used for PID efficiency calibration. The optimal value for the PID cuts is equal to the largest available value, regardless of the leptons coming from signal decays or not. The optimized PID cuts for leptons are

- e prob. > 0.9 for electrons,
- μ prob. > 0.97 for muons.

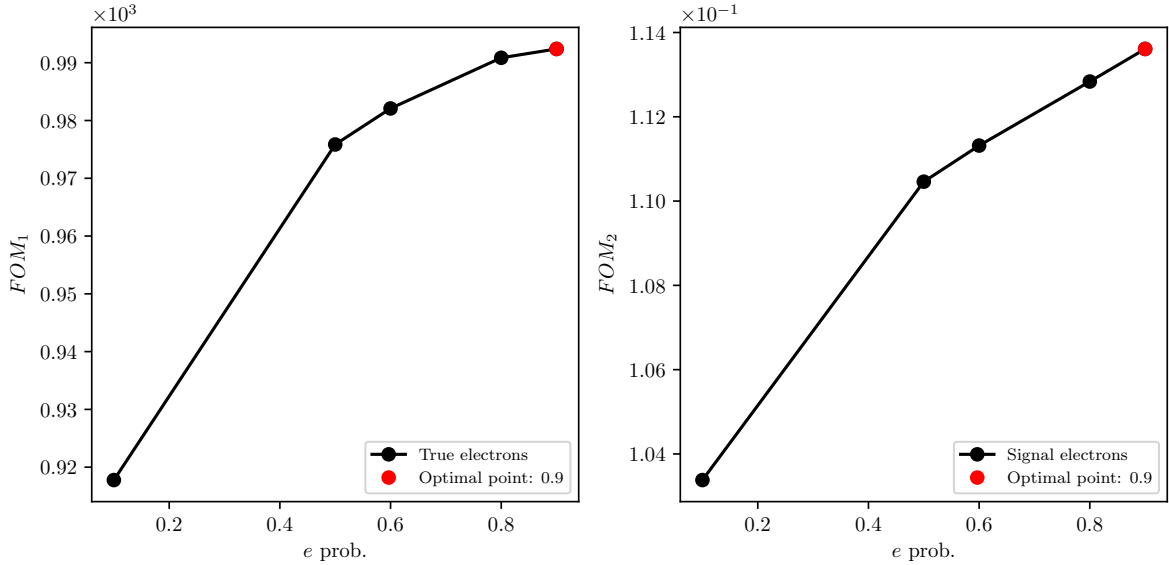


Figure 4.3: FOM optimizations of the PID probability cuts for true electrons (left) and true electrons from signal decays (right).

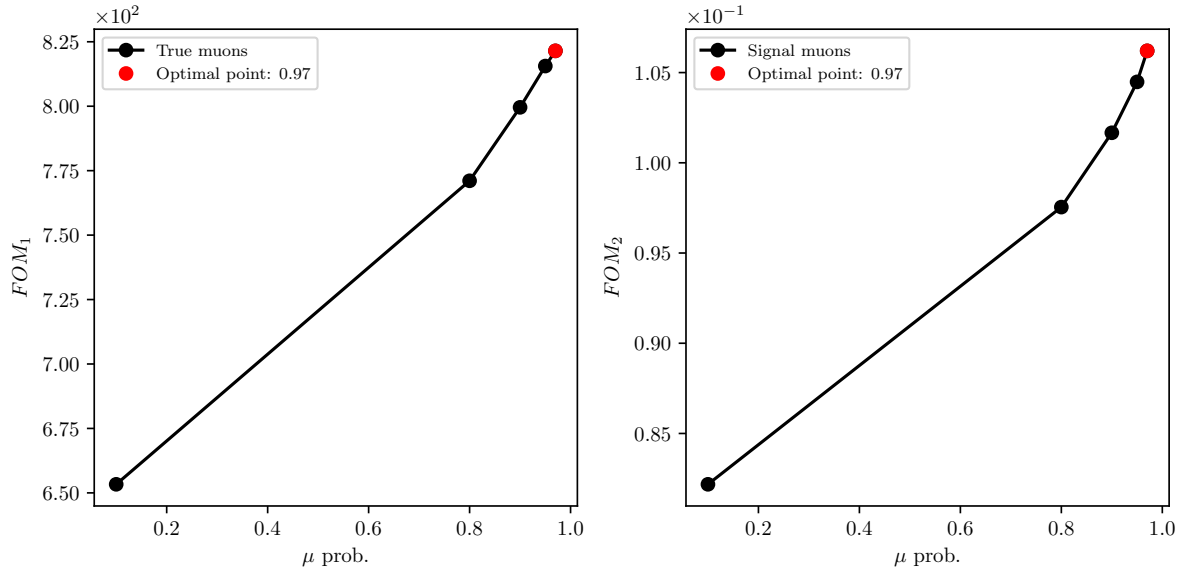


Figure 4.4: FOM optimizations of the PID probability cuts for true muons (left) and true muons from signal decays (right).

323 Kaons

324 We repeat the procedure for both kaons. Figure 4.5 shows the impact parameters d_0 and
 325 z_0 , the momentum in $\Upsilon(4S)$ center-of-mass system (CMS), and the PID information for
 326 true and fake kaons, where an extra category for true kaons from the signal decay is
 327 shown.

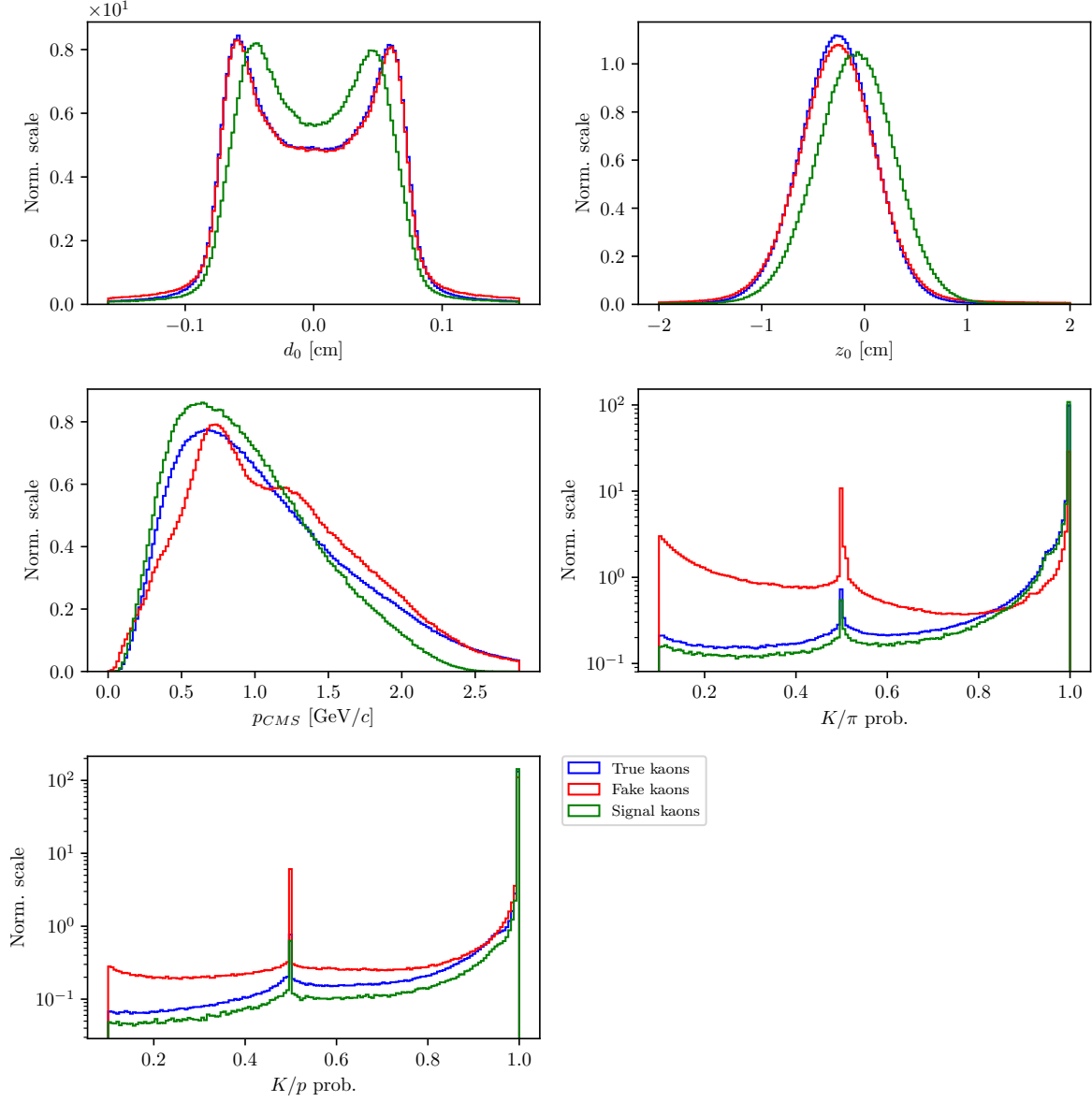


Figure 4.5: Normalized properties of true (blue), fake (red) and true kaons (green) from signal decay.

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

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

The PID optimization in this case is taken in two steps. First we optimize the cut on

333 K/π , and after that the K/p separation probability. Figure 4.6 shows the optimization
 334 procedure for PID cuts on kaon candidates.

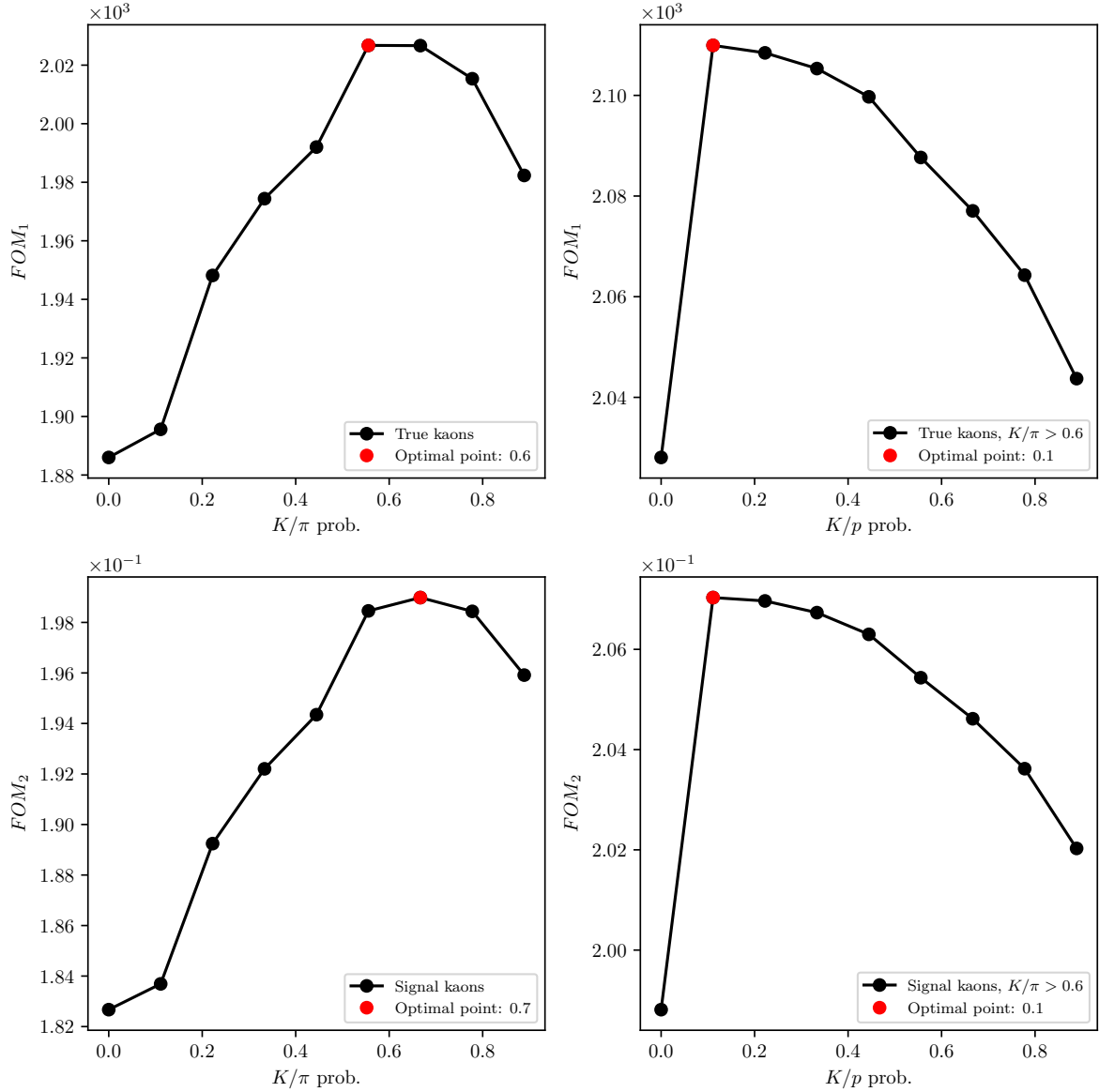


Figure 4.6: FOM optimizations of the PID probability cuts for true kaons (top) and true kaons from signal decays (bottom). The plots on the left show the optimization of the first step for the K/π probability cut, and the plot on the right the K/p probability cut.

335 The optimized PID cuts for kaons are

- 336 • $K/\pi > 0.6$,

- $K/p > 0.1$.

4.2 Combination of FSP particles

With the pre-selected kaon and lepton candidates we make combinations for potential B meson candidates. Since the missing neutrino escapes detection, we reconstruct the B mesons in the following two channels

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

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 beam direction. We demand that the fit converged and apply a cut on the minimal fit probability. The fit probability for signal and background B meson candidates is shown in Figure 4.7 (left). We perform a *FOM* cut optimization of this variable, which is shown in Figure 4.7 (right). In this and in the following cases, the definition of S from Eq. (4.1) are correctly reconstructed B meson candidates with a missing neutrino, with an additional constraint that the B meson did not decay via a $b \rightarrow c$ transition.

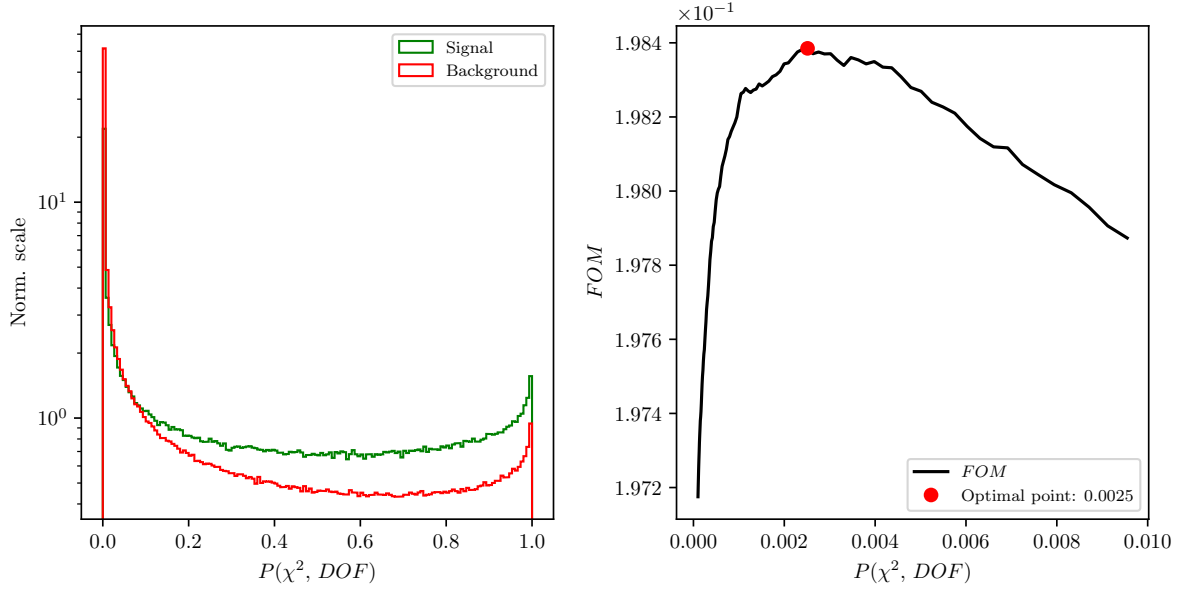


Figure 4.7: Normalized vertex fit probability distribution for signal and background B meson candidates in logarithmic scale (left) and *FOM* optimization of the vertex fit probability (right).

350 Even though that the vertex fit probability cut was optimized here, we choose a standard
 351 cut of

352 $\bullet P(\chi^2, NDF) > 1.0 \times 10^{-3}.$

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 true B , as

$$\mathbf{p}_\nu = \mathbf{p}_B - \mathbf{p}_Y, \quad (4.2)$$

$$p_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_BE_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \quad (4.3)$$

$$\cos(\theta_{BY}) = \frac{2E_BE_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \quad (4.4)$$

where all the energy and momenta above are calculated in the CMS frame. The mass of the neutrino is equal to 0 to a very good precision, so we use it in Eq. (4.3). In addition, we can substitute the unknown energy and momentum magnitude, E_B and $|\vec{p}_B|$, of the B meson in Eq. (4.4), with quantities from the well known initial conditions

$$E_B = E_{CMS}/2, \quad (4.5)$$

$$|\vec{p}_B| = p_B = \sqrt{E_B^2 - m_B^2}, \quad (4.6)$$

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

355 For the correctly reconstructed candidates, this variable lies in the $[-1, 1]$ region, though
 356 only to a certain precision, due to the finite detector resolution. For background candi-
 357 dates, however, the values populate also the non-physical regions, as is shown in Figure
 358 4.8 (left). We impose an optimized cut on this variable from Figure 4.8 (right) to discard
 359 a large amount of background.

360 $\bullet |\cos(\theta_{BY})| < 1.0.$

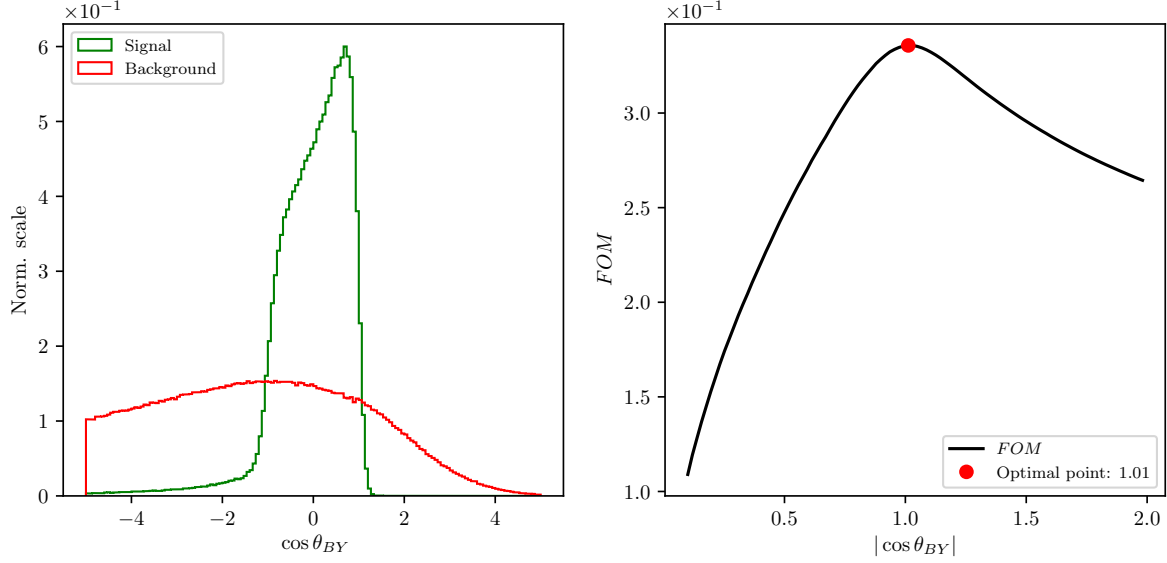


Figure 4.8: Normalized $\cos \theta_{BY}$ distribution for signal and background B meson candidates (left) and FOM optimization of the $\cos \theta_{BY}$ variable (right).

4.3 Loose neutrino reconstruction

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

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

mask is created for tracks and clusters, which narrates the use of this object in the final calculations of the ROE four-momentum. From this point on we assume the ROE to be efficiently cleansed of extra tracks and clusters. A more detailed description of the ROE clean-up can be found in Chapter 5.

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

$$p_{miss} = p_{\Upsilon(4S)} - \sum_i^{\text{Event}} (E_i, \vec{p}_i), \quad (4.7)$$

$$p_{miss} = p_{\Upsilon(4S)} - \left(p_Y - \sum_i^{\text{Rest of event}} (E_i, \vec{p}_i) \right), \quad (4.8)$$

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

$$p_i^{\text{neutral}} = (p_i, \vec{p}_i) \quad \text{and} \quad p_i^{\text{charged}} = \left(\sqrt{m_i^2 + p_i^2}, \vec{p}_i \right), \quad (4.9)$$

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

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

We define the square of the missing mass, m_{miss}^2 , which is zero, if signal-side neutrino is the only missing particle in the event, as shown in Eq. (4.11).

$$p_\nu = p_{miss} = (E_{miss}, \vec{p}_{miss}), \quad (4.10)$$

$$m_{miss}^2 = p_{miss}^2 = p_\nu^2 = m_\nu^2 \approx 0. \quad (4.11)$$

Since the detector is not perfect, the distribution of the m_{miss}^2 variable has a non-zero width. Additionally, tails are introduced as soon as we have missing particles such as extra missing neutrinos, other neutral undetected particles such as K_L^0 , or simply missing tracks due to detection failure. Figure 4.9 shows the distribution of m_{miss}^2 as defined with the missing four-momentum in Eq. (4.10). Correctly reconstructed candidates, which come from events where the other B meson decayed via a hadronic decay mode, should peak at zero. If this is not the case, candidates are shifted to larger values

of this variable. Due to this fact, we impose a cut 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

- $|m_{miss}^2| < 7 \text{ GeV}/c^2$.

This cut was not optimized, since the optimal case would result in a too strong threshold at this point in the analysis, since we still may want to retain as much signal candidates as possible, even if they are coming from events with semi-leptonic decays of the other B meson.

For further purposes in this analysis we also define a subset of all signal candidates, which come from events where the companion B meson decayed hadronically and all of its particles were taken into account correctly. We only allow for missing photons, since they are frequently irradiated due to brehmsstrahlung effects and don't have such a big impact on the 4-momentum of the final candidate. We denote this subset as *perfect* signal.

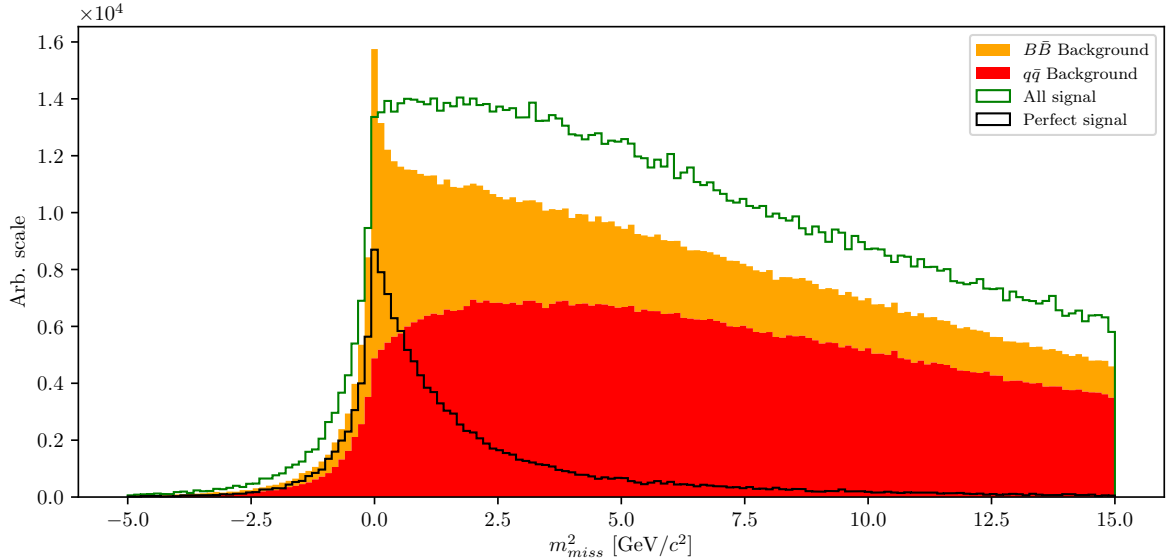


Figure 4.9: Squared missing mass distribution for $q\bar{q}$ and $B\bar{B}$ background in filled histograms. All signal (green) and perfect signal (black) are scaled up equally.

The main uncertainty in neutrino four-momentum, defined as in Eq. (4.10) comes from energy uncertainty. It is a common practice to substitute the missing energy with the magnitude of the missing momentum, since the momentum resolution from the measurement is much better, thus redefining the neutrino four-momentum to

$$p_\nu = (|\vec{p}_{miss}|, \vec{p}_{miss}), \quad (4.12)$$

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

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

$$\Delta E = E_B - E_{CMS}/2, \quad (4.13)$$

$$M_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}|^2}. \quad (4.14)$$

420 Since the final fit will be performed over ΔE and M_{BC} , we define the fit region

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

423 Figure 4.10 shows the distributions of ΔE (left) and M_{BC} (right) for signal and major
 424 types of background after the pre-cuts. Both signal components are scaled up with
 425 respect to the background components, but are in proper scale one to another. The
 426 effects of missing particles are clearly seen based on the shape difference between all and
 427 perfect signal.

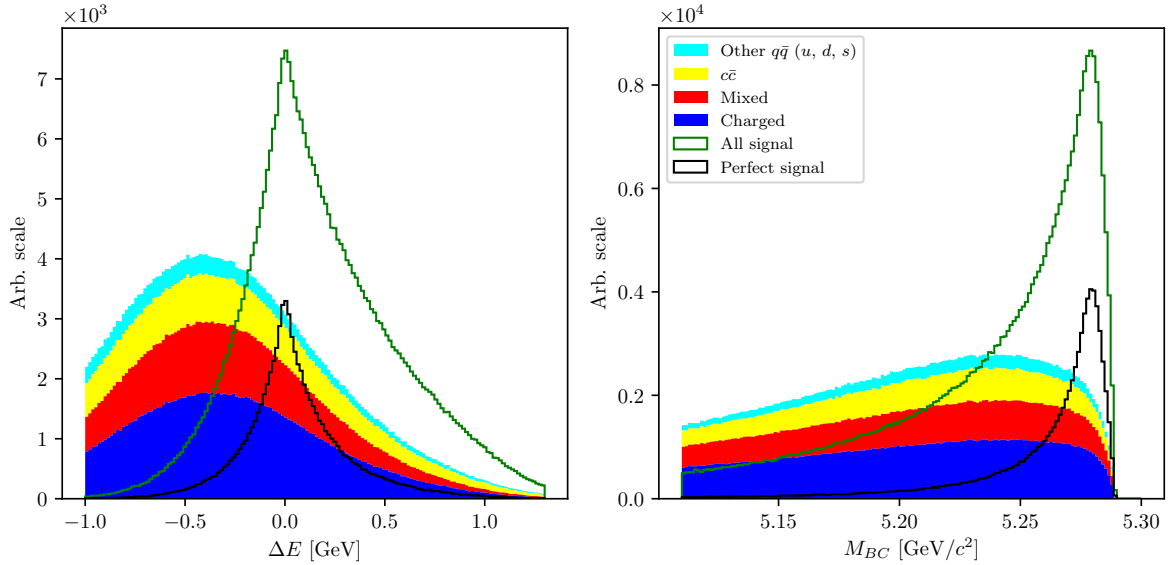


Figure 4.10: Distributions of ΔE (left) and M_{BC} (right) for signal and major types of background after the precuts. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution in both distributions, since the event is perfectly reconstructed.

4.4 Event categorization

The missing information due to an escaping neutrino in our reconstructed channel is replaced by information from the companion B meson. Since this is an untagged reconstruction, the quality of the companion B meson affects the properties of signal candidate. Perfect reconstruction of a hadronically decaying 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 long tails and/or a shift from the desired value in the mentioned distributions. These effects are undesired, since they make it harder to separate signal from background.

To remedy this, we split our signal candidates into 4 categories and work only with the best category. For categorization we use splitting in two ways. First, 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

$$q_{B_{sig}} q_{B_{comp}} = -1, \quad (4.15)$$

however, this value is distributed due to missing charged particles in the reconstruction. Secondly, we train an MVA classifier based on ROE object properties in order to recognize companion B mesons, which decayed hadronically. The details of the training are described in subsection 4.4.1.

We define the 4 categories in the following way

category I: $q_{B^\pm} q_{B^\mp} = -1$ and $BDT_{had.} > 0.57$,

category II: $q_{B^\pm} q_{B^\mp} \neq -1$ and $BDT_{had.} > 0.57$,

category III: $q_{B^\pm} q_{B^\mp} = -1$ and $BDT_{had.} \leq 0.57$,

category IV: $q_{B^\pm} q_{B^\mp} \neq -1$ and $BDT_{had.} \leq 0.57$,

with the relative ratios of 37.74 %, 40.50 %, 13.43 % and 8.32 % for categories I, II, III and IV, respectively. Different categories for all signal candidates after the pre-cuts are shown in Figure 4.11. Category I and II represent the majority of the candidates more or less equally, but category I has the best resolution, so we focus on candidates from category I in this analysis.

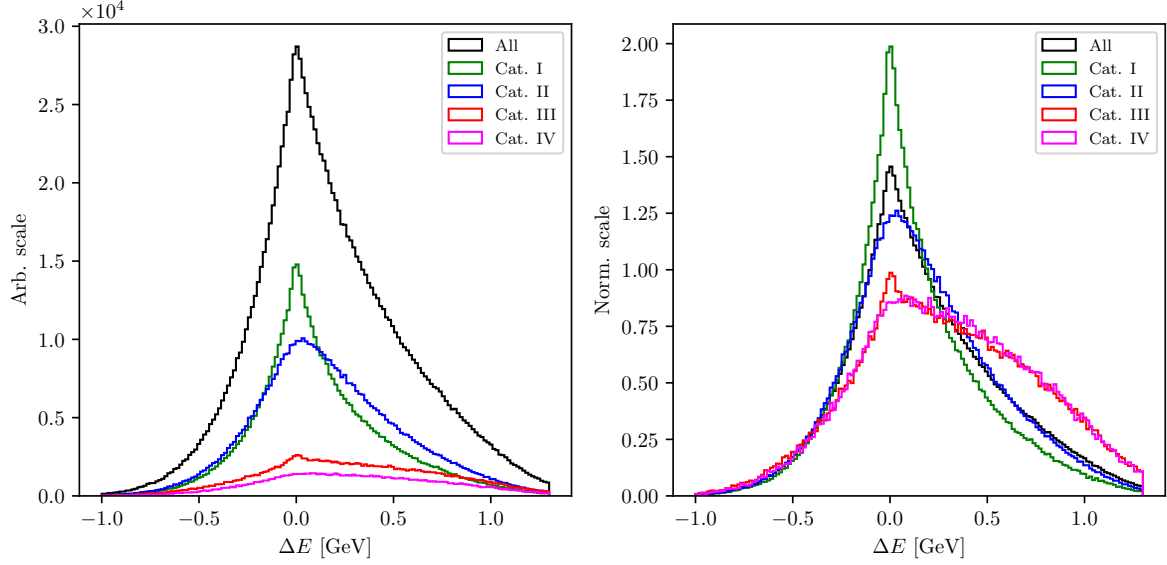


Figure 4.11: Categorization of signal candidates based on the charge product of both B mesons in the event and the MVA output for recognizing hadronic decays of the companion B meson. The plot on the left shows the distributions in an arbitrary scales, while the plot on the right shows the normalized distributions.

4.4.1 Hadronic decay MVA training

In order to train an MVA classifier to recognize events with hadronically decaying B mesons in the ROE side, we prepare a dataset of

- 249385 target candidates,
- 223681 background candidates,

where the definition of target is a ROE which corresponds to a hadronically decayed companion B meson.

The input variables used in this MVA are ROE specific and do not depend on the signal side. They are

- angle between tracks,
- track quantities
- $P(\chi^2, DOF)$ of the tracks fit from the ROE side,

- K and ℓ FlavorTagger variables,
- Charge of the ROE,
- $\cos \theta$ of the ROE momentum in the CMS frame,
- Number of tracks in ROE,
- Number of distant tracks in ROE, which don't pass requirements from Section 4.6.

The classifier output is shown in Figure 4.12 (left). Candidates which populate the region with low values of the classifier output are more likely to come from semi-leptonic decays, so we want to discard those candidates. When optimizing the FOM , we redefine the S in Eq. (4.1) to a correctly reconstructed signal candidate with a hadronically decayed companion B meson. This FOM optimization, shown in Figure 4.12 (right), yields the optimal cut

- $BDT_{had.} > 0.57$.

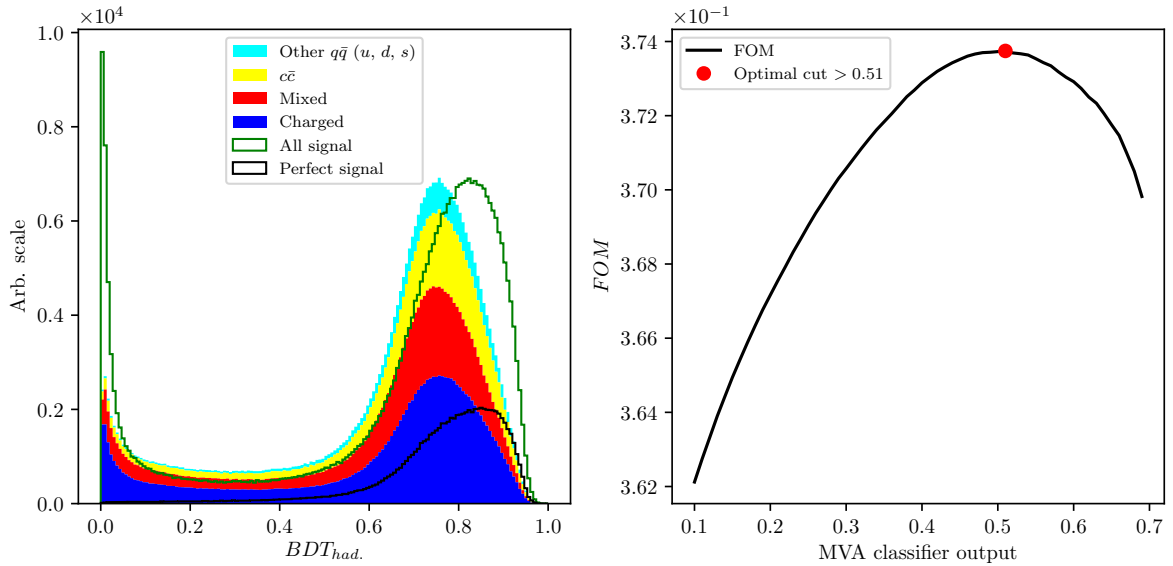


Figure 4.12: Hadronic MVA classifier output for major types of background and scaled-up signal (left), and FOM optimization on this cut for correctly reconstructed signal candidates with hadronically decayed companion B meson (right).

4.5 Signal region definition

Since signal candidates are now categorized, we can define a signal region, where most of our perfectly reconstructed candidates lie. We use this region for optimizations of all cuts in the following steps of background suppression in chapter 6. The 2D *FOM* optimization of the optimal M_{BC} and ΔE is shown in Figure 4.13. The signal region is defined as

- $M_{BC} > 5.270 \text{ GeV}/c^2$,
- $|\Delta E| < 0.166 \text{ GeV}$.

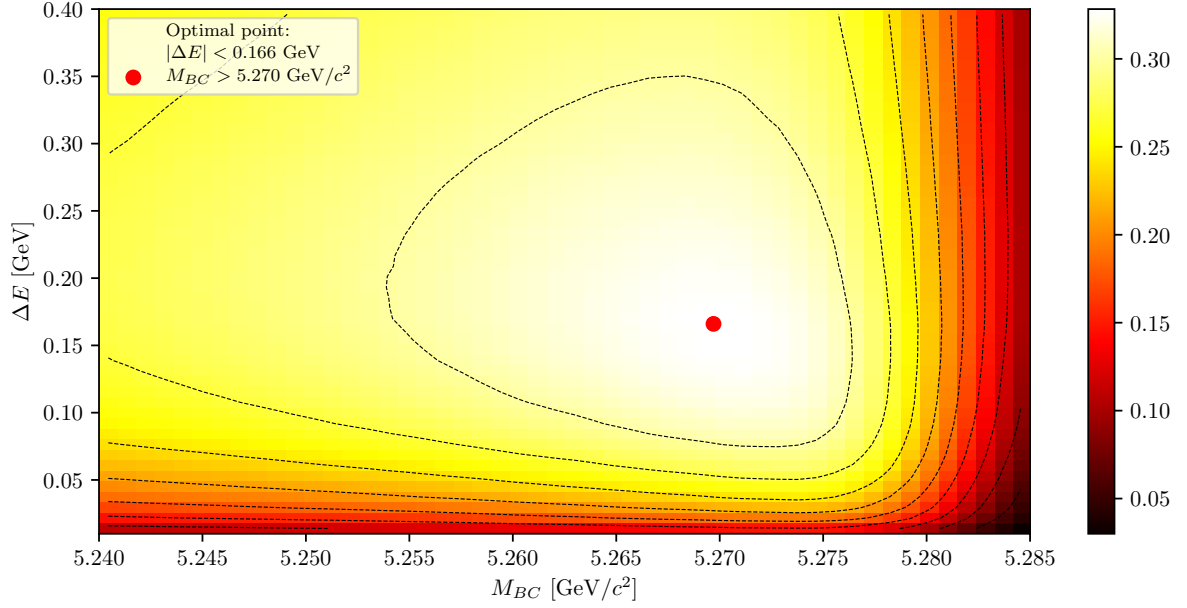


Figure 4.13: 2D *FOM* optimization of the signal region definition, where the signal in the optimization was represented by perfectly reconstructed candidates.

Lastly, we can tighten the cut 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 4.14, the optimal cut on m_{miss}^2 is

- $|m_{miss}^2| < 1.1 \text{ GeV}/c^2$.

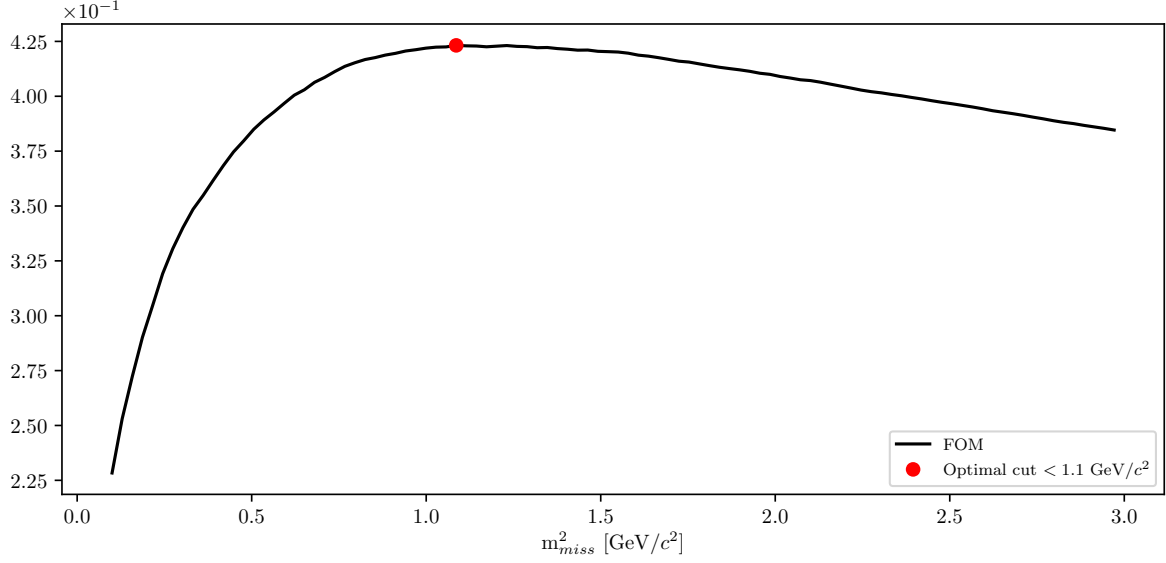


Figure 4.14: *FOM* optimization of the optimal m_{miss}^2 cut in the signal region.

4.6 Selection summary

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

- FSP particles:

- Electrons: $|d_0| < 0.1$ cm, $|z_0| < 1, 5$ cm, $p > 0.6$ GeV/ c ,
 $p_{CMS} \in [0.4, 2.6]$ GeV/ c , $eID > 0.9$,
- Muons: $|d_0| < 0.1$ cm, $|z_0| < 1, 5$ cm, $p_{CMS} \in [0.6, 2.6]$ GeV/ c ,
 $\mu ID > 0.9$,
- 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:

- Before ROE clean-up: $P(\chi^2, DOF) > 1 \times 10^{-3}$, $|\cos \theta_{BY}| < 1$, $|m_{miss}^2| < 7$ GeV/ c^2 ,
- After ROE clean-up: $\Delta E \in [-1.0, 1.3]$ GeV, $M_{BC} \in [5.1, 5.295]$ GeV/ c^2 ,
- After signal categorization: $q_{B_{sig}} q_{B_{comp}} = -1$, $BDT_{had.} > 0.57$, $|m_{miss}^2| < 1.1$ GeV/ c^2 .

Chapter 5

Rest of event clean-up

Continuing from Section 4.3, the description of the ROE clean-up process is described here.

5.1 Setting up the MVA

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

For each clean-up step a dataset was prepared for the training of "signal" (target) against "background". The Fast-BDT (FBDT) [5] algorithm was used as the MVA classifier and the following hyper-parameters were chosen for optimization

- **nTrees**: the number of trees in the FBDT forest,
- **nLevels**: the number of levels in each FBDT tree.

Figure 5.1 shows a graphical interpretation of the FBDT forest with **nTrees** and **nLevels**. In all cases the hyper-parameters were optimized with a grid-search method in the hyper-parameter phase-space. Additionally, we perform a k -fold cross-validation ($k = 5$) for each hyper-parameter set, where we cycle through the 5 parts of the dataset and use 4 parts of the dataset for training, whereas the remaining part is used for validation. This minimizes the bias towards the validation sample when optimizing the hyper-parameters.

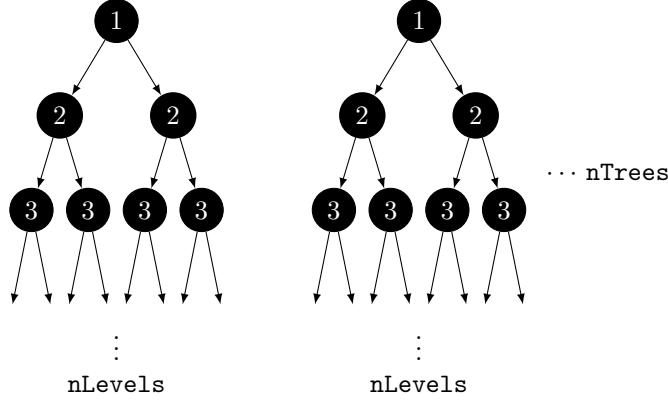


Figure 5.1: A schematic of a BDT forest with $nTrees$, each tree having a depth of $nLevels$.

5.2 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. (4.8).

Most of the photons in events with B mesons come from $\pi^0 \rightarrow \gamma\gamma$ decays. However, a lot of hits in the ECL are also created by photons coming from beam-induced background or secondary interactions with the detector material. Such photons add extra energy and momentum which spoils our measured quantities.

In the first step of the clusters clean-up we train an MVA which recognizes good π^0 candidates. The output of this classifier is then applied to photons and represents a sort of a π^0 origin probability, which is used as an additional classifier variable in the next step of the clean-up.

5.2.1 π^0 MVA training

The dataset of π^0 candidates contains

- 387125 target candidates,
- 416019 background candidates,

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

552 The input variables used in this MVA are

- 553 • p and p_{CMS} of π^0 and γ daughters,
- 554 • fit prob. of the mass-constrained fit, invariant mass and significance of mass before
 555 and after the fit,
- 556 • angle between the photon daughters in the CMS frame,
- 557 • cluster quantities for each daughter photon
 - 558 – E_9/E_{25} ,
 - 559 – theta angle,
 - 560 – number of hit cells in the ECL,
 - 561 – highest energy in cell,
 - 562 – energy error,
 - 563 – distance to closest track at ECL radius.

564 The classifier output variable is shown in Figure 5.2.

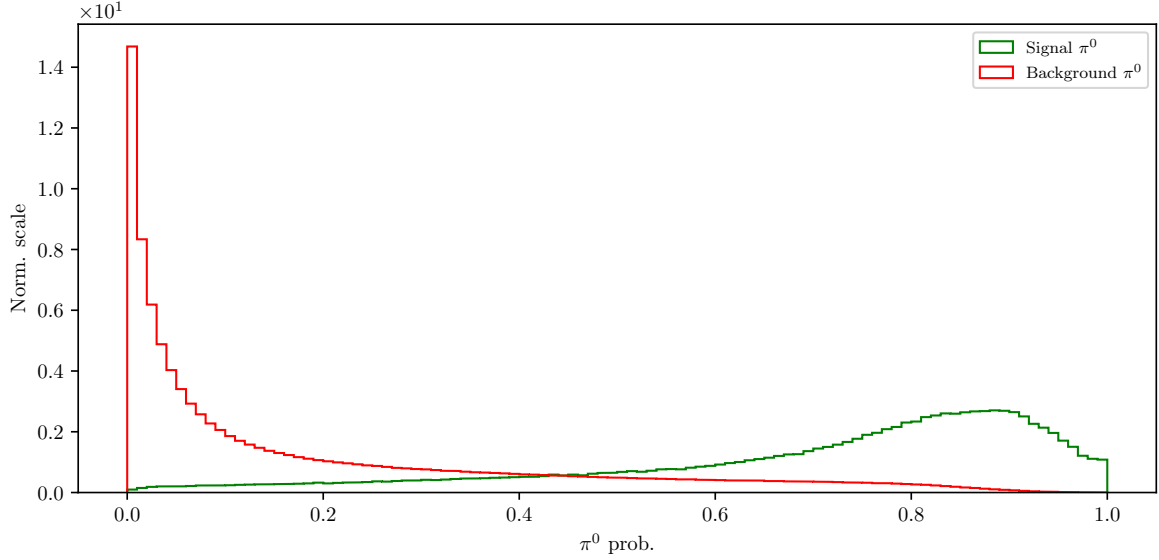


Figure 5.2: Classifier output of the π^0 training for signal and background π^0 candidates.

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

5.2.2 γ MVA training

In this MVA training we take the π^0 classifier output of the previous training as an input in order to train a classifier to distinguish between good and bad photons. The π^0 probability information from the previous step is applied to all photon pairs which pass the same π^0 cuts as defined in the previous step. Since it's possible to have overlapping pairs of photons, the π^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 π^0 selection, these candidates have a fixed value of π^0 probability equal to zero.

The dataset of γ candidates contains

- 324781 target candidates,
- 333353 background candidates,

where the definition of 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 γ candidates from the existing Belle particle

list.

The input variables used in this MVA are

- p and p_{CMS} of γ candidates,
- π^0 probability,
- cluster quantities
 - 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 5.3.

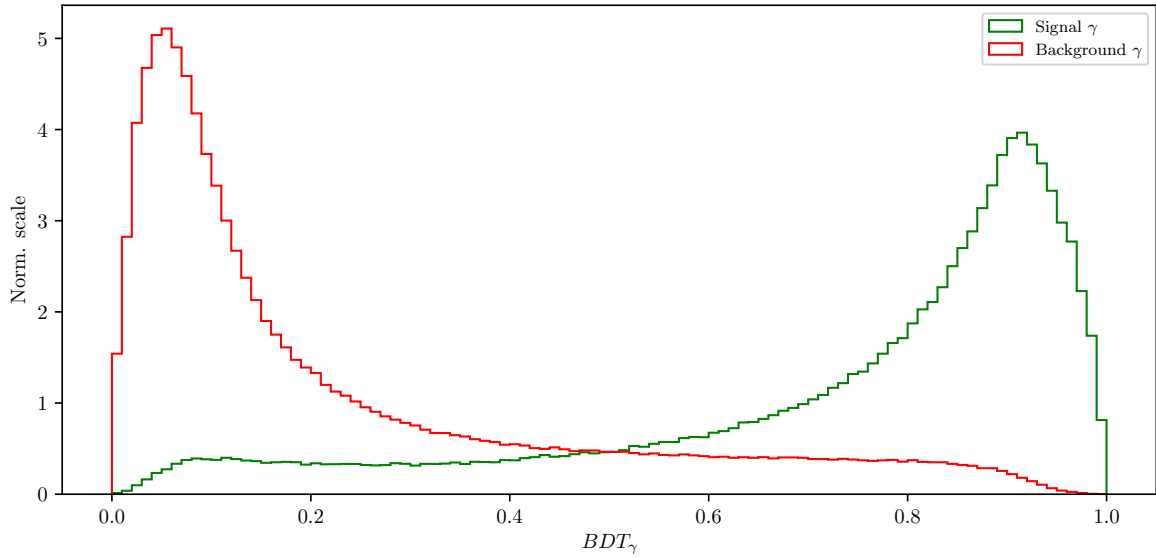


Figure 5.3: Classifier output of the γ training for signal and background γ candidates.

With the final weights for photon classification in hand, we apply them to the photon particle list. The cut optimization is shown in Figure 5.4 (left), with the optimal cut on

the γ classifier output at

- $BDT_\gamma > 0.519$.

Figure 5.4 (right) shows the LAB frame momentum of the photons before and after the cut in logarithmic scale. The signal efficiency and background rejection at this clean-up cut are

- Signal efficiency: $\epsilon_{SIG} = 82.6 \%$,
- Background rejection: $1 - \epsilon_{BKG} = 81.7 \%$.

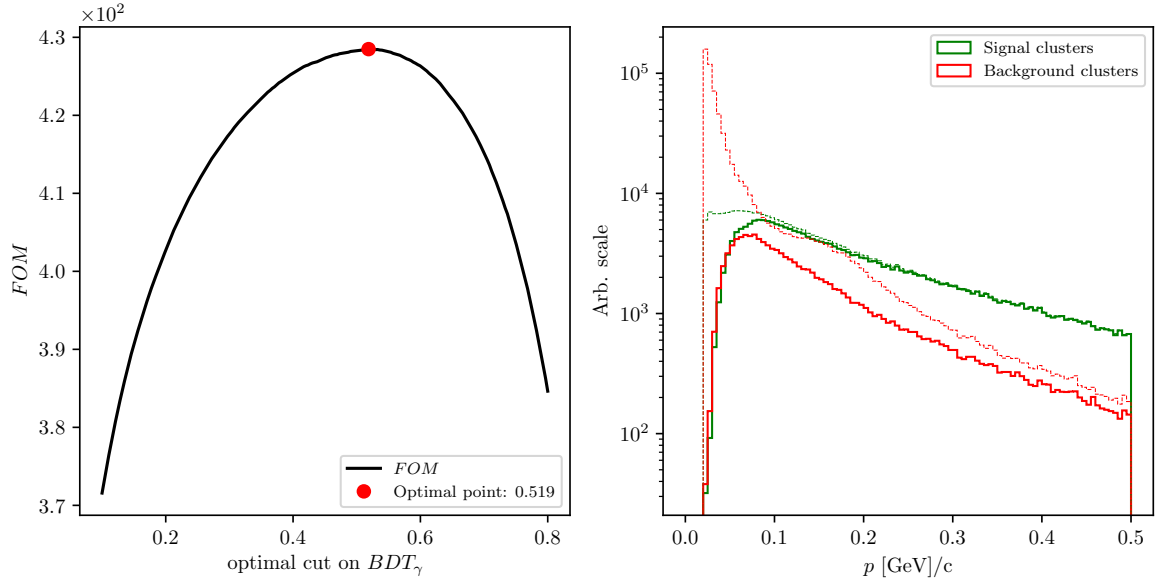


Figure 5.4: The FOM of the classifier output optimization (left) and momentum magnitude in the LAB frame of signal and background photon candidates before and after the optimal cut (right).

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

5.3 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 charged particle's four-momentum, which is then added in the loop in Eq. (4.8).

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 5.5.

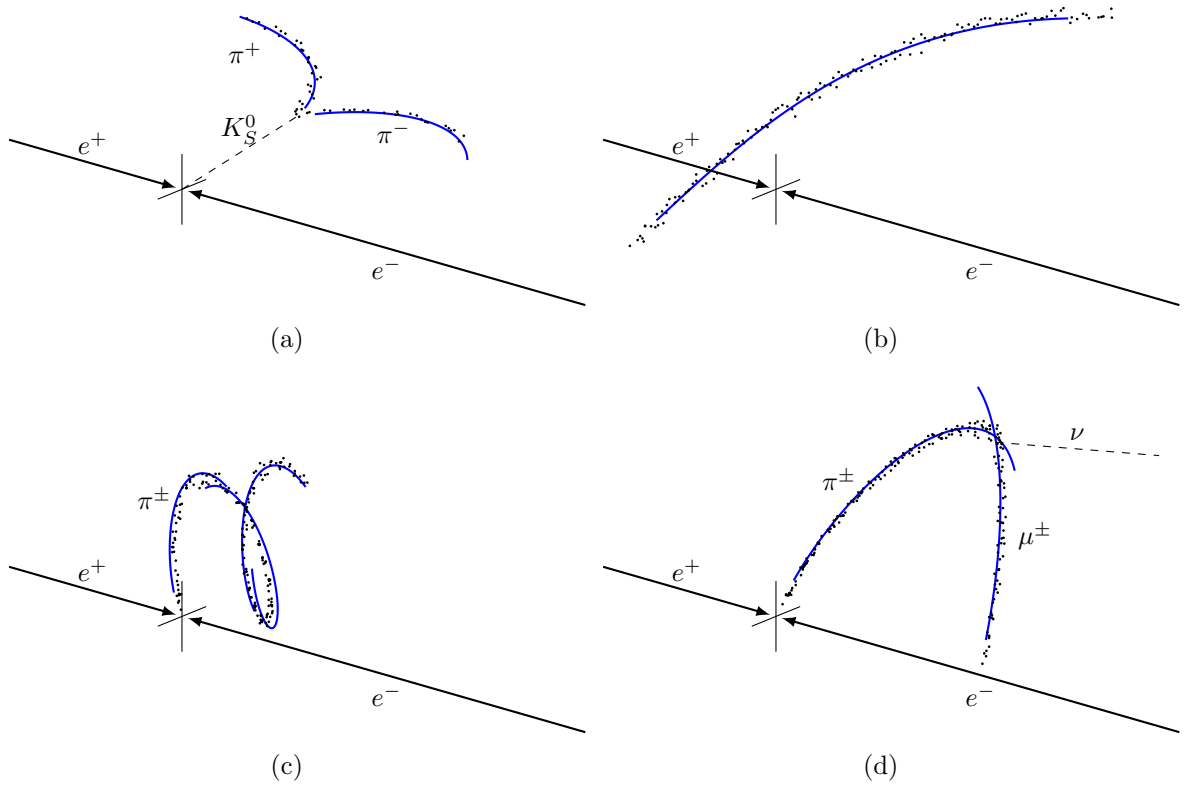


Figure 5.5: (a) Tracks from long-lived neutral particles, which decay away from the IP region, (b) Random tracks from background which are reconstructed, (c) Low-momentum particles which curl in the magnetic field, (d) in-flight decays of particles, which produce a kink in the trajectory.

It is obvious that tracks from the same momentum source should only be taken into account once, or, in case of background tracks, not at all. Such tracks will from this point on be denoted as *extra* tracks, because they add extra four-momentum to our final calculations in Eq. (4.8). At the same time, we have to take care that we don't identify *good* tracks as *extra* tracks. Both of these cases have negative impacts on the final resolution of all variables which depend on information from ROE.

5.3.1 Tracks from long-lived particles

The first step in tracks clean-up is taking care of tracks from long-lived particles. Here we only focus on K_S^0 , since they are the most abundant. This step is necessary because the π^\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 result in order to select only the good K_S^0 candidates. Figure 5.6 shows the distribution of the K_S^0 invariant mass for signal and background candidates, before and after the classifier cut. The momentum of selected K_S^0 candidates is added to the ROE, while the daughter tracks are discarded from our set.

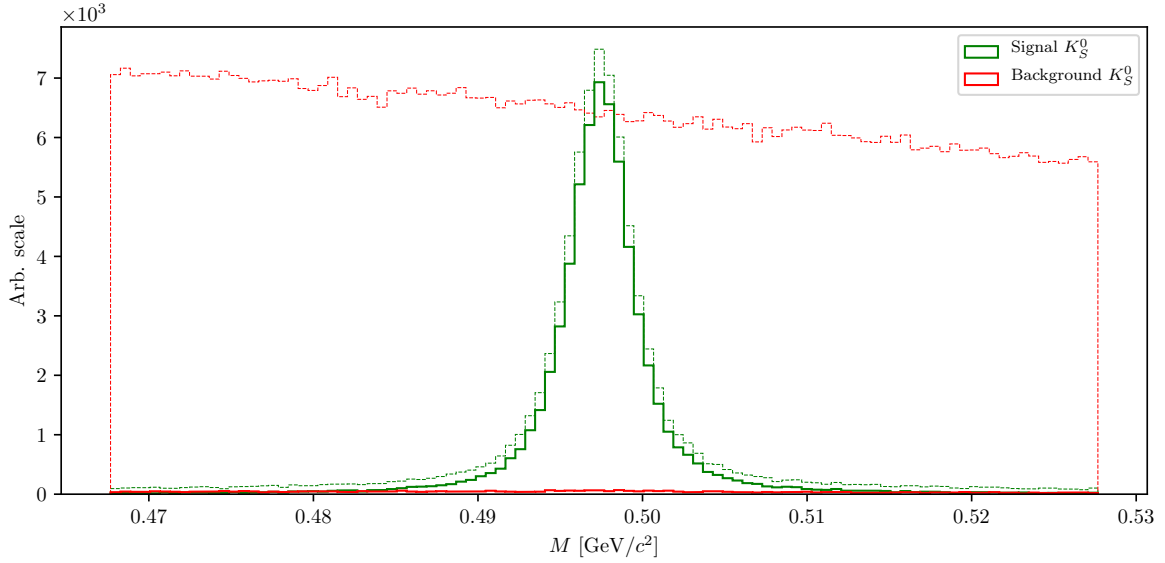


Figure 5.6: Invariant mass of the K_S^0 candidates before (dashed lines) and after (solid lines) the cut on the Neural Network classifier for signal (green) and background candidates (red). Signal peaks at nominal K_S^0 mass, while background covers a wider region.

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

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

644 5.3.2 Duplicate tracks

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

- 648 • $|d_0| < 10 \text{ cm}$ and $|z_0| < 20 \text{ cm}$

649 and proceed with the clean-up of track duplicates.

650 Defining a duplicate track pair

651 In this step we wish to find a handle on secondary tracks from low momentum curlers
652 and decays in flight. The main property for these cases is that the angle between such
653 two tracks is very close to 0° or 180° , since tracks deviate only slightly from the initial
654 direction, but can also be reconstructed in the opposite way. Figure 5.7 shows the
655 distribution of the angle between two tracks in a single pair for random track pairs and
656 duplicate track pairs, where the latter were reconstructed as two same-sign or opposite-
657 sign tracks.

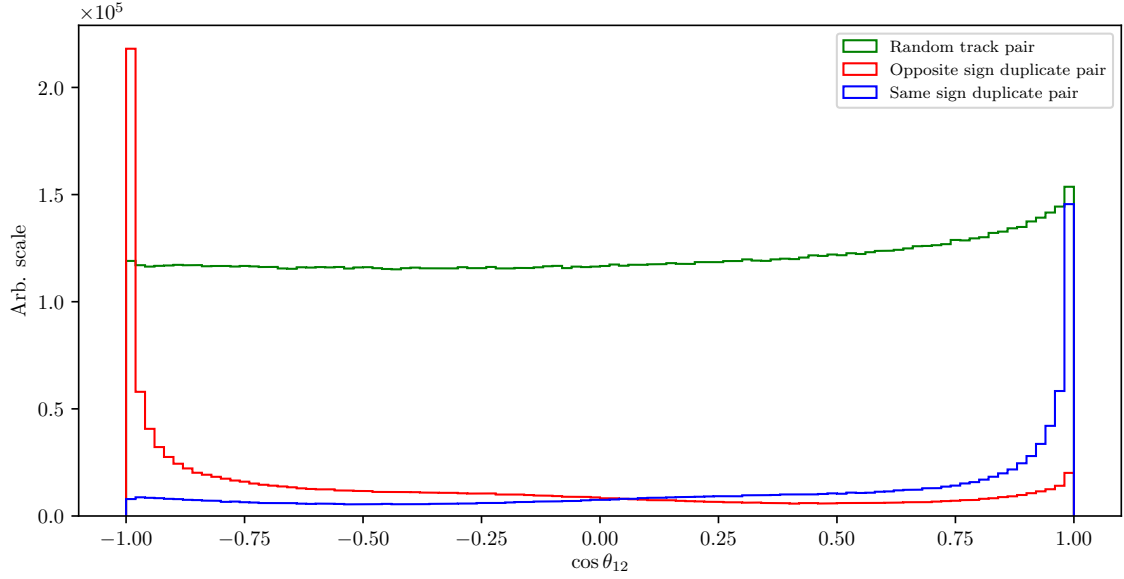


Figure 5.7: Distribution of the angle between two tracks in a single pair for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

658 If the particle decayed mid-flight or produced multiple tracks due to being a low-
 659 momentum curler, then, as the name suggests, these particles most likely had low mo-
 660 mentum in the transverse direction, p_T . Since both tracks originate from the same initial
 661 particle, the momentum difference should also peak at small values. Figure 5.8 shows the
 662 momentum and momentum difference of tracks which belong to a random or a duplicate
 663 track pair.

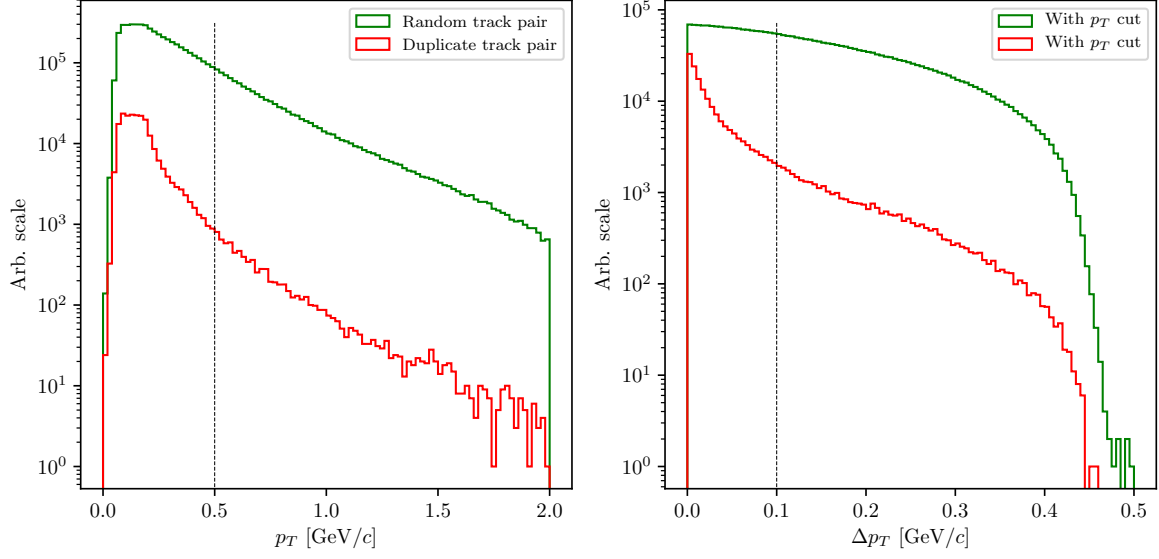


Figure 5.8: Distribution of transverse momentum p_T (left) and transverse momentum difference Δp_T (right) for all tracks coming from random (green) or duplicate track pairs (red). The plot on the right already includes the cut on p_T from the plot on the left.

We impose a cut of

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

in order to cut down the number of random track pairs, while retaining a high percentage of duplicate track pairs. After all the cuts defined in this chapter, the final distribution of the angle between two tracks is shown in Figure 5.9.

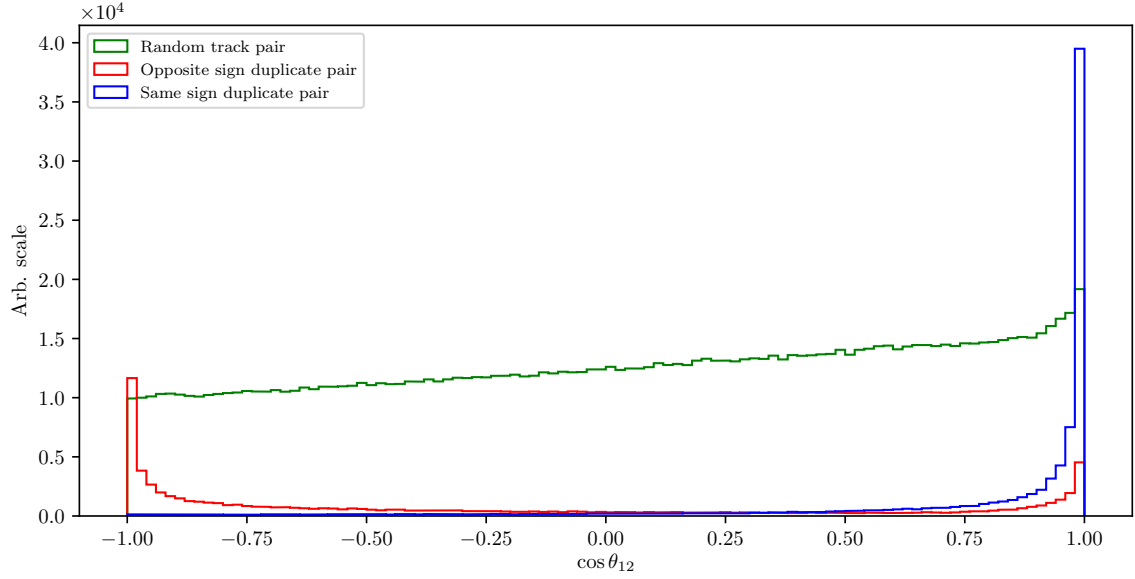


Figure 5.9: Distribution of the angle between two tracks in a single pair after applying the selection cuts defined in this subsection. The distributions are shown for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

Training the duplicate track pair MVA

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

- 215601 target candidates,
- 311124 background candidates,

where the definition of 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 5.10.

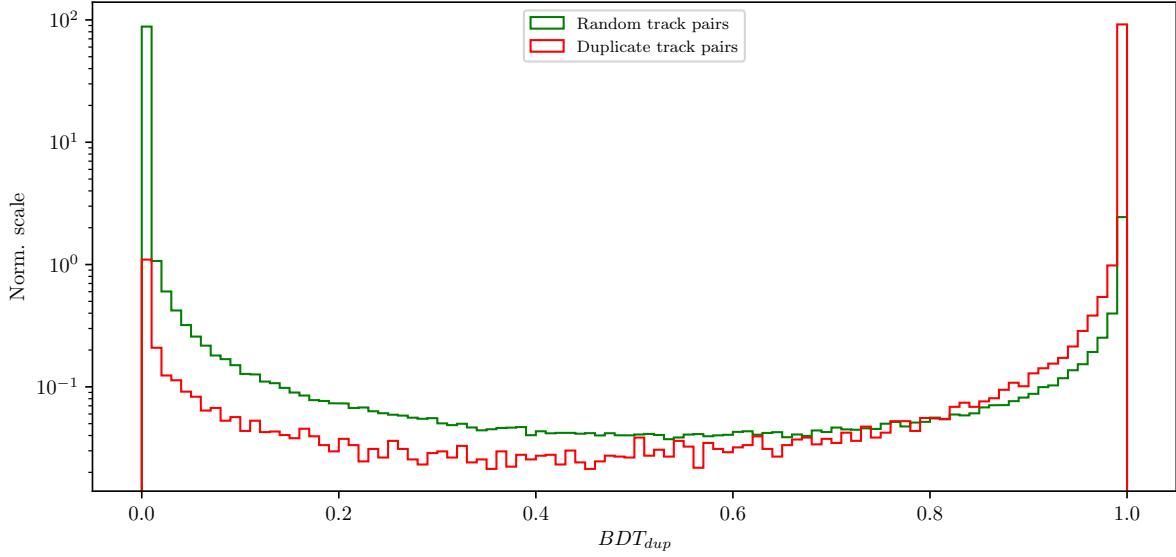


Figure 5.10: Classifier output of the track pair training for random track pairs and duplicate track pairs.

The FOM function for optimal cut selection is shown in Figure 5.11 (left), along with the angle between the two tracks before and after the optimal cut (right). The optimal cut for duplicate track selection is

- $BDT_{duplicate} > 0.99915$.

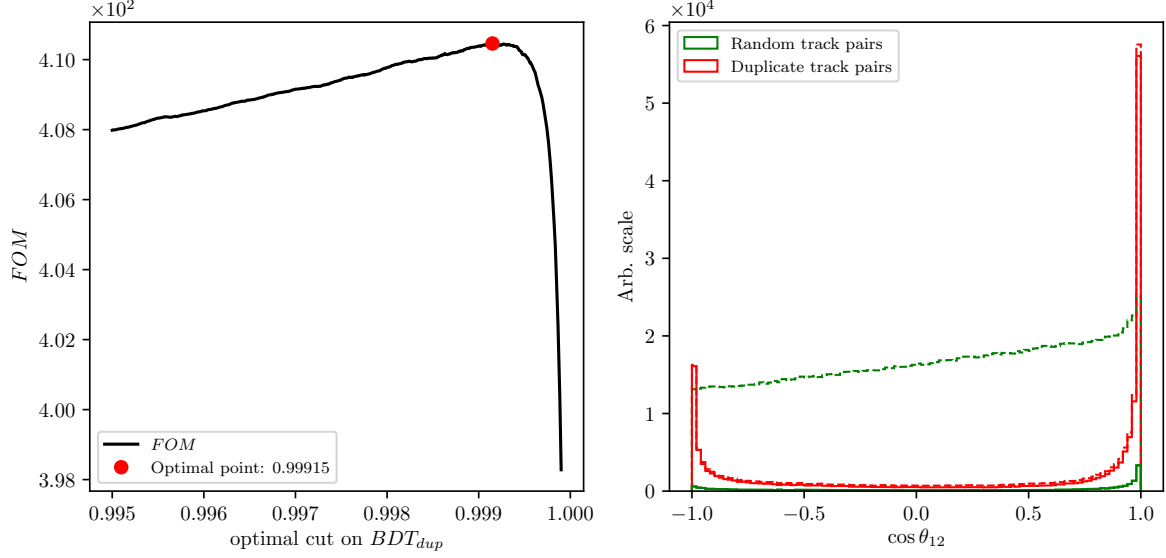


Figure 5.11: The optimization of the FOM function for the cut on classifier output (left) and distribution of the angle between two tracks in a single pair before (dashed) and after (solid) applying the optimal cut on the output classifier for random and duplicate track pairs (right).

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

$$\Delta f = f_{this} - f_{other}, \quad (5.1)$$

where f is an arbitrary variable from the list of track quantities in Subsection 5.3.2. 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),

700 Additionally we define an MC truth variable

$$\Delta\chi^2 = \chi_{this}^2 - \chi_{other}^2, \quad \chi^2 = \sum_{i=x,y,z} \frac{(p_i - p_i^{MC})^2}{\sigma(p_i)^2}, \quad (5.2)$$

701 where we compare all components of track momentum to the true values. If $\Delta\chi^2 >$
 702 0, then *this* track has a higher probability of being a duplicate track and should be
 703 discarded.

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

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

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

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

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

707 Training the duplicate track MVA

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

711 • 160146 target candidates,
712 • 128568 background candidates,
713 where the definition of target is that the track is a duplicate track.

714 The input variables used in this MVA are

- 715 • theta angle of the track momentum,
- 716 • track quantities
 - 717 – impact parameters d_0 and z_0 ,
 - 718 – momentum components p_T and p_z
 - 719 – number of hits in the SVD and CDC detectors
 - 720 – track fit p -value,
- 721 • pair-level information
 - 722 – $\Delta d_0, \Delta z_0, \Delta N_{CDC}, \Delta N_{SVD}, \Delta p_T, \Delta p_z$.

723 The classifier is shown in Figure 5.12. The weights from this training are applied to the
724 tracks, where now each track has a certain probability of being a duplicate track. We
725 now compare these values between both tracks in each track pair as

$$\Delta BDT_{final} = BDT_{final}^{this} - BDT_{final}^{other}, \quad (5.3)$$

726 which is again applied to all track pairs and overwritten for tracks which are more
727 duplicate-like.

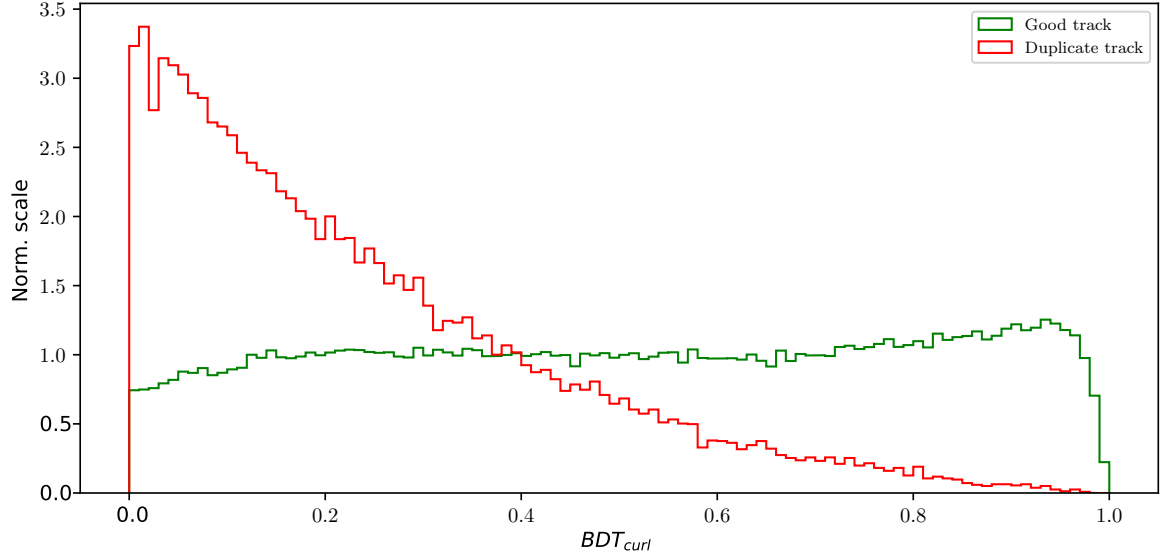


Figure 5.12: Classifier output of the MVA training for curling track recognition.

728 Finally, we select all duplicate tracks which survive the cut

$$\Delta BDT_{final} > 0 \quad (5.4)$$

729 and discard them from our ROE. We can check the performance of our duplicate track
 730 classifier by applying the procedure on a sample of duplicate track pairs and comparing
 731 the predicted result with the truth, based on Eq. (5.2). Table 5.1 shows the performance
 732 of the duplicate track recognition in the form of percentages of correctly and incorrectly
 733 identified duplicate and original tracks. The model seems to perform well and the event
 734 is now considered to be clean of duplicate tracks.

	Predicted duplicate track	Predicted good track
Duplicate track	84.63 %	20.95 %
Good track	15.37 %	79.05 %

Table 5.1: Ratios of correctly classified and misclassified tracks.

735 5.4 Belle clean-up

736 The clean-up, used standardly at Belle, is much simpler and relies only on a set of
 737 rectangular cuts for neutral particles as well as charged ones. In case of photons, a

single cut on photon energy is applied, depending on the region where the photon hit the relevant part of the detector. The photon cuts are summarized in Table 5.2.

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

Table 5.2: Photon selection for the Belle clean-up procedure. Different cuts are applied on photons in different parts of the detector

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

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

Of the two tracks, the one with a larger value of formula in Eq. 5.5 is discarded. The remaining tracks in the event then need to satisfy the conditions described in Table 5.3.

$$(\gamma|d_0|)^2 + |z_0|^2, \quad \gamma = 5. \quad (5.5)$$

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

Table 5.3: Photon selection for the Belle clean-up procedure. Different cuts are applied on photons in different parts of the detector

5.5 Clean-up results

In this section the results of the ROE clean-up are shown. It is clear that cleaning up the event affects the shape of various distribution, especially ΔE and M_{BC} , which we are most interested in. Since the reconstruction procedure includes cuts on the cleaned-up variables, this also results in an efficiency increase of the reconstructed sample.

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. Figure 5.13 shows distributions of ΔE and M_{BC} for various clean-up setups. We see an improvement in

resolution as well as an increase in efficiency in both observed variables for the case where the ROE clean-up is performed. Table 5.4 shows ratios of efficiencies and $FWHM$'s of ΔE with respect to the default case. While both the Belle and ROE clean-up increase efficiency, only ROE clean-up improves the resolution.

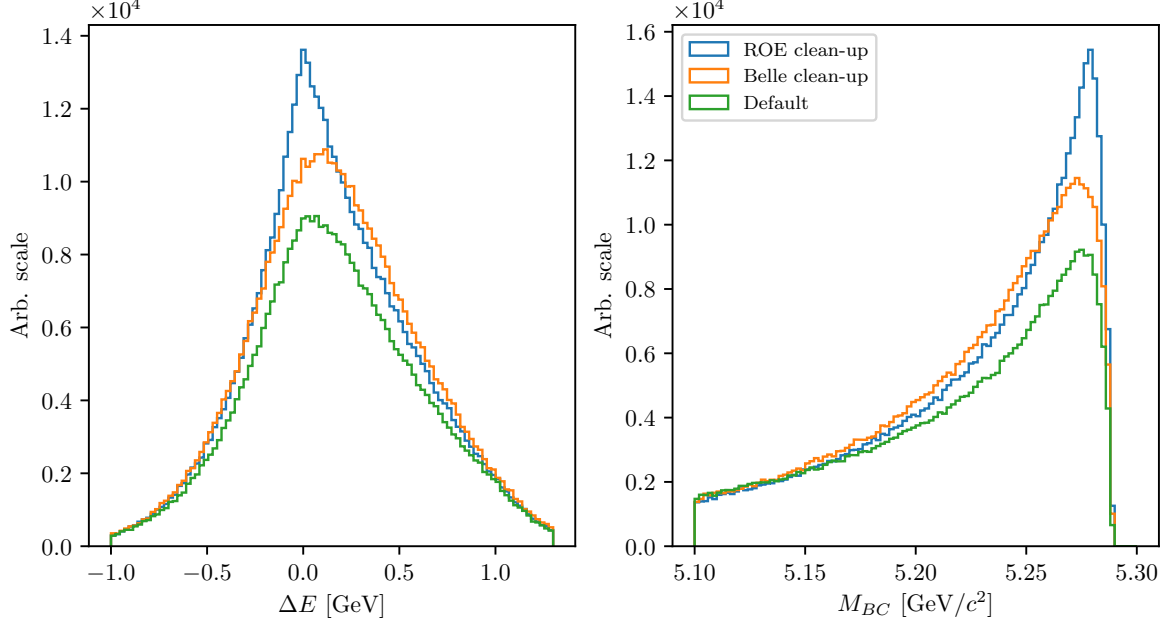


Figure 5.13: ΔE and M_{BC} distributions for various types of clean-up procedures. For ROE clean-up, the procedure seems to improve resolution as well as increase efficiency, compared to the default case.

	Efficiency ratio	FWHM ratio
Belle clean-up	121.90 %	102.60 %
ROE clean-up	123.40 %	77.92 %

Table 5.4: Comparison of efficiencies and $FWHM$'s of ROE and Belle clean-up setups with respect to the default case (no clean-up).

Another variable which heavily depends on the clean-up is the charge product of the signal and companion B meson candidate, already defined in Eq. (4.15), shown in Figure 5.14 for various clean-up procedures. As a cross-check, we can also look at ΔE and M_{BC} variables for each value of the charge product. These plots are shown in Figure 5.15 and they show a clear increase in efficiency and improvement in resolution 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 an increase in efficiency for both cases of

clean-up, but neither help improve the resolution. This supports our choice of signal categorization, defined in Section 4.4, where we select only candidates with the correct value of the charge product.

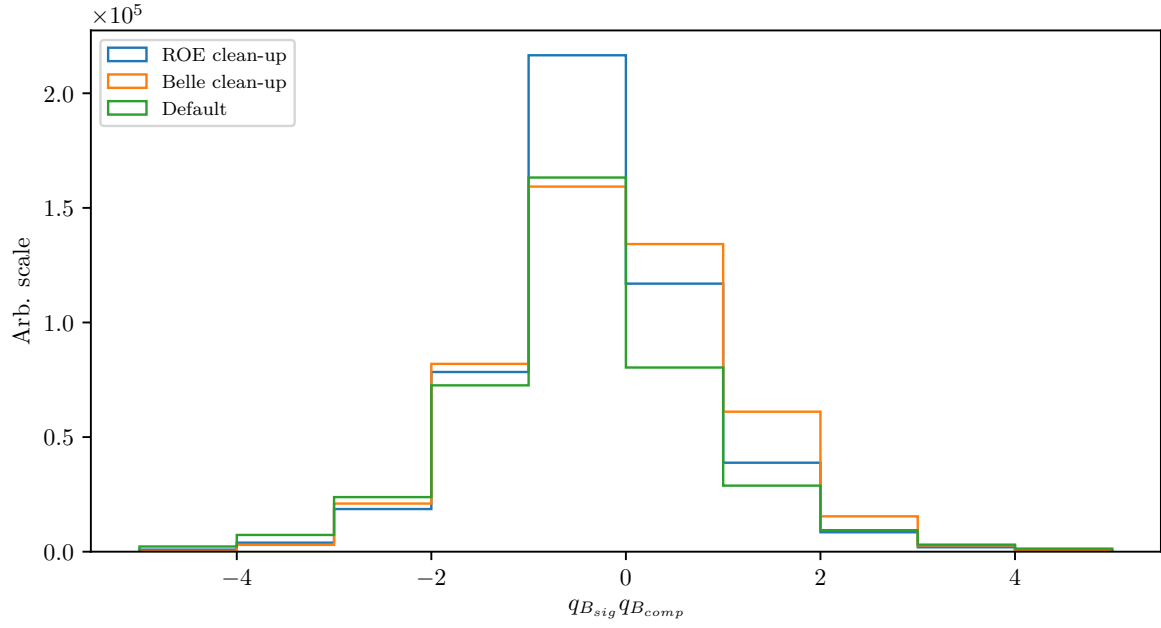


Figure 5.14: Distribution of B_{sig} and B_{comp} charge product for various types of clean-up procedures. For ROE clean-up, the procedure seems to increase the number of candidates with the correct value of the product.

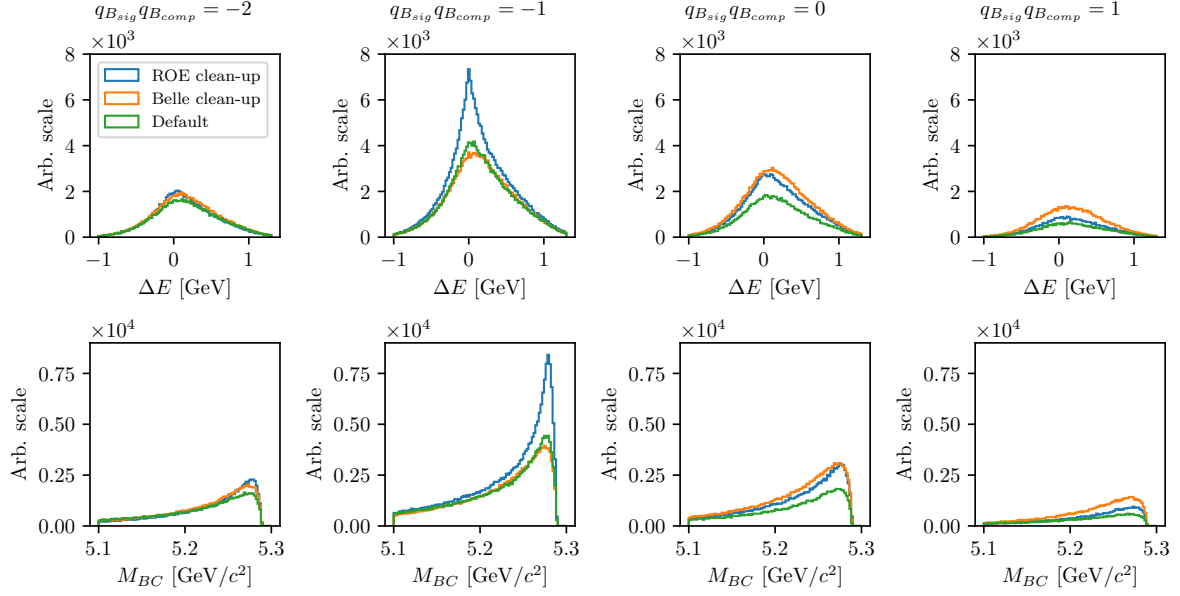


Figure 5.15: Distributions of ΔE (top) and M_{BC} (bottom) for various types of clean-up procedures, split by specific values of the charge product. While there seems to be an increase in efficiency for both clean-up procedures in general, ROE-cleanup makes the most improvement to resolution and efficiency for the correct value of the charge product.

5.6 ROE clean-up validation

The ROE clean-up seems to perform well, but we have to keep in mind that we are testing the procedure on simulated, not real data. Therefore we have to make sure that this procedure performs on simulated as well as real data in the same manner.

Will add more later.

Chapter 6

Background suppression

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

6.1 Control decay and other resonant background

In this 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 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 6.1 shows this invariant mass spectrum of two kaons, m_{KK} , where obvious resonant structures are present from sources like

- $\phi \rightarrow K^+ K^-$ (sharp resonance 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).

An opportunity for a control decay choice presents itself as

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

which has very similar properties as the studied signal decay, is much more abundant, and, most importantly, is easy to suppress, since it only populates a very narrow region in the m_{KK} invariant mass spectrum. In order to suppress these resonant backgrounds while studying the signal or control decay, we impose a set of the following cuts

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

794 where $m_\phi \approx 1.019 \text{ GeV}/c^2$ and $m_{D^0} \approx 1.864 \text{ GeV}/c^2$ are nominal masses of the ϕ and
 795 D^0 mesons, and $\Delta_\phi \approx 8 \times 10^{-3} \text{ GeV}/c^2$ and $\Delta_{D^0} \approx 1.2 \times 10^{-2} \text{ GeV}/c^2$ are symmetric
 796 cut widths around the nominal mass values for the ϕ and D^0 mesons, respectively.

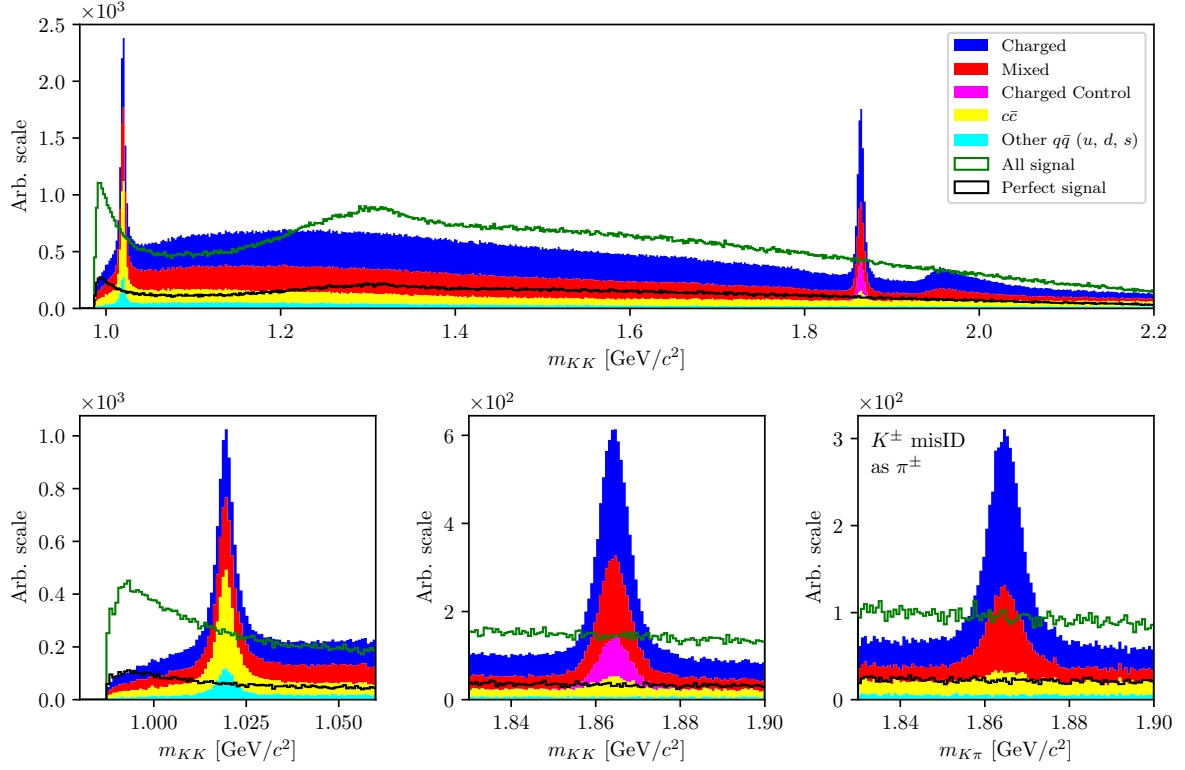


Figure 6.1: Invariant mass of two correctly reconstructed kaons (left)
 and invariant mass of two kaons, where one was mis-
 identified as a pion (right).

797 6.2 Continuum suppression

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

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

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

hand B mesons from $B\bar{B}$ events are produced almost at rest in the CMS frame. Their decay products from an isotropic distribution in the detector, which yields a spherical event shape.

6.2.1 Characteristic variables

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

Thrust and related variables

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

$$\mathbf{T} = \frac{\sum_i |\mathbf{T} \cdot \mathbf{p}_i|}{\sum_i |\mathbf{p}_i|}. \quad (6.1)$$

In this case, a related variable is $|\cos \theta_T|$, where θ_T is the angle between the thrust axis of the momenta from B meson decay particles, and the thrust axis of all particles in the ROE. Since both B mesons in $B\bar{B}$ events are produced at rest, their decay particles, and consequentially their thrust axes, are uniformly distributed in the range $[0, 1]$. On the other hand, decay particles from continuum events follow the direction of the jets in the event. As a consequence, the thrusts of both the B meson and the ROE are strongly directional and collimated, which results in a large peak at $|\cos \theta_T| \approx 1$. Additionally, one can also use the variable $|\cos \theta_{TB}|$, which is the thrust axis between the B candidate and the beam axis. For B candidates from $B\bar{B}$, this distributions is again uniformly distributed, while for candidates from continuum events this distribution follows the distribution of the jets with the function $1 + \cos^2 \theta_{T,B}$. Figure 6.2 shows the distributions of $|\cos \theta_T|$ (left) and $|\cos \theta_{T,B}|$ (right) for different B meson candidates.

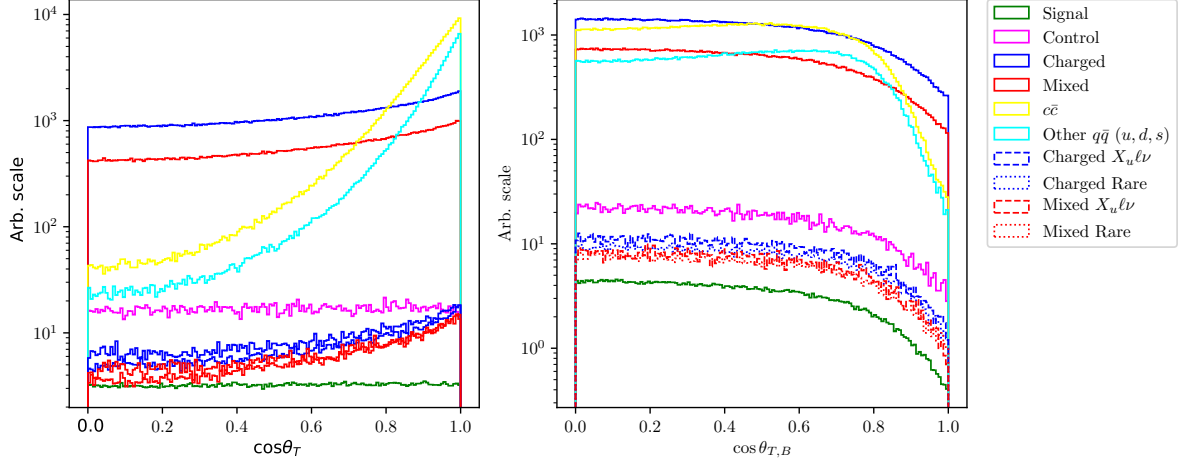


Figure 6.2: Distributions of $|\cos \theta_T|$ (left) and $|\cos \theta_{T,B}|$ (right) for different B meson candidates.

CLEO cones

CLEO cones have been introduced by the CLEO collaboration [6] and are an additional specific tool to provide optimal background discrimination. They are nine variables corresponding to the momentum flow around the thrust axis of the B meson candidate, binned in nine cones of 10° around the thrust axis, as illustrated in Figure 6.3.

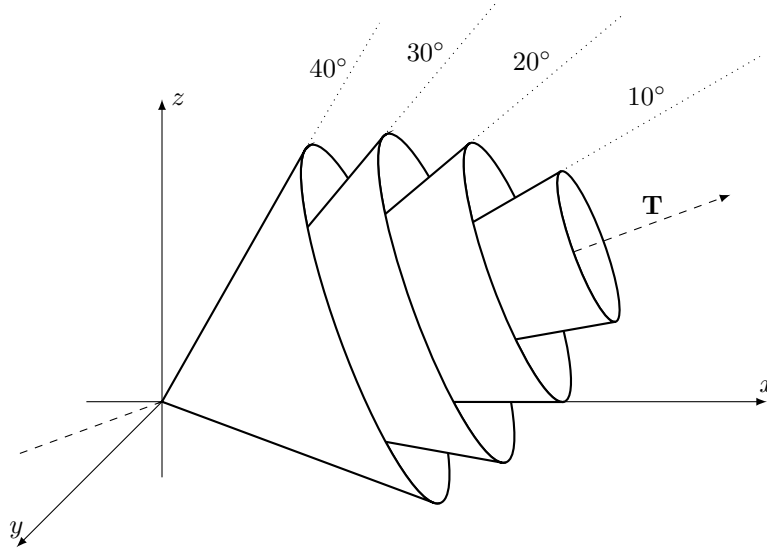


Figure 6.3: Concept of CLEO cones. \mathbf{T} denotes the thrust axis of the B meson candidates in the event. Each variable corresponds to a momentum flow around the thrust axis in steps of 10° .

834 KSW moments

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

$$R_k = \frac{H_k}{H_0} = \frac{1}{H_0} \sum_{i,j} |p_i||p_j| P_k(\cos \theta_{ij}), \quad (6.2)$$

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

842 Belle developed a refined generation of Fox-Wolfram moments, called Kakuno-Super-
 843 Fox-Wolfram (KSW) moments to further suppress the continuum background. There
 844 are 17 different KSW moments which are grouped into R_k^{so} , R_k^{oo} and R_k^{ss} [7]. The latter
 845 ones are excluded due to correlations with B meson specific variables.

846 6.2.2 MVA training

847 Most of the characteristic variables, described in subsection 6.2.1, were taken together
 848 in order to train a single MVA classifier for continuum suppression. All characteristic
 849 variables were checked for possible q^2 correlation. Variables with significant correlation
 850 or complex shapes in the 2D plot were discarded from the training set, since they would
 851 have introduced unwanted dependence on the unreliable model, ISGW2, used for signal
 852 MC generation. Additionally, all of the characteristic variables in our set do not depend
 853 on the signal mode, they only differ in the kinematic and topological aspects of $B\bar{B}$ and
 854 continuum background events.

855 The training dataset consisted of 2×10^5 candidates, where 50 % of the candidates are
 856 correctly reconstructed signal events, 25 % are $u\bar{u}$, $d\bar{d}$ and $s\bar{s}$ background with expected
 857 proportions, and 25 % is $c\bar{c}$ background. Since the full Belle dataset is experiment de-
 858 pendent, we construct the training dataset by proportionally sampling each MC dataset,
 859 corresponding to the appropriate experiment number.

860 The training variable set consisted of

- 861 • B meson direction and thrust related variables
 - 862 – magnitude of thrust axes of B and ROE ,
 - 863 – cosine of angle between thrust axis of B and thrust axis of ROE ,

- 864 – cosine of angle between thrust axis of B and beam direction,
- 865 – reduced Fox-Wolfram moment R_2 ,
- 866 • all 9 CLEO Cones
- 867 • KSFW Moments
 - 868 – $R_{01}^{so}, R_{02}^{so}, R_{03}^{so}, R_{04}^{so}$,
 - 869 – $R_{10}^{so}, R_{12}^{so}, R_{14}^{so}$,
 - 870 – $R_{20}^{so}, R_{22}^{so}, R_{24}^{so}$,
 - 871 – $R_0^{oo}, R_1^{oo}, R_2^{oo}, R_3^{oo}, R_4^{oo}$,
- 872 • FlavorTagging variables
 - 873 – qp of e, μ, ℓ ,
 - 874 – qp of intermediate e, μ, ℓ ,
 - 875 – qp of $K, K/\pi$, slow pion, fast hadron,
 - 876 – qp of maximum P^*, Λ , fast-slow-correlated (FSC),
- 877 • Other
 - 878 – $\Delta z, \Delta t$.

879 Figure 6.4 shows the classifier output for various types of background, all in expected
 880 MC proportions. B meson candidates from continuum background are dominant at
 881 lower values, while candidates from $B\bar{B}$ events populate the region with higher values.

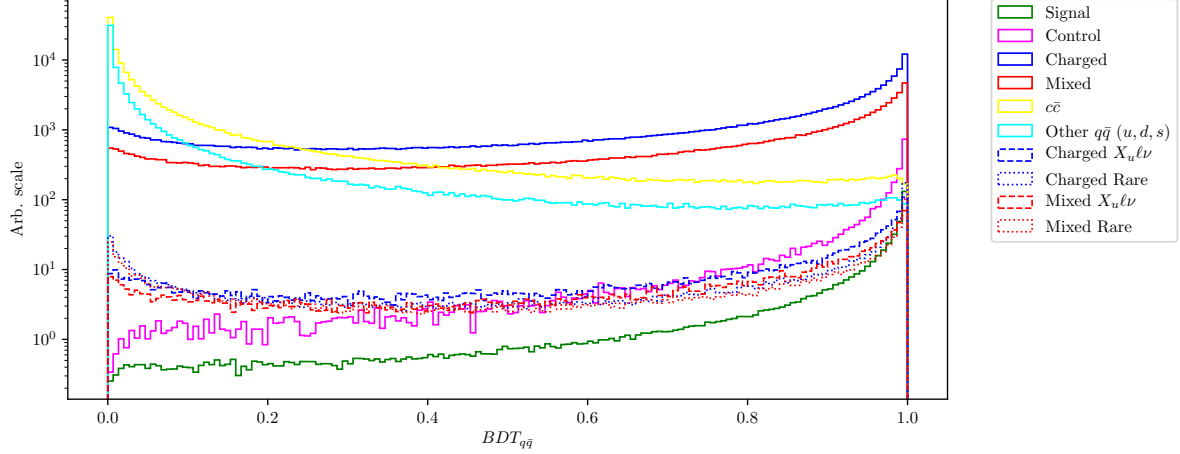


Figure 6.4: Continuum suppression classifier output for signal and various types of background. B candidates from continuum events dominate the lower region, while candidates from $B\bar{B}$ dominate in the upper region of the classifier output.

6.3 $B\bar{B}$ suppression

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

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

Similarly as before, the training dataset for this classifier consisted of 2×10^5 candidates, where 50 % of the candidates are correctly reconstructed signal events. The remaining part of the training dataset consists of all background, not including the control sample, because we are not interested in suppressing it directly. The background part of the dataset consists of 75 % charged and neutral $B\bar{B}$ events in equal proportions, whereas the remaining 25 % is equally populated with charged and neutral $B\bar{B}$ events from $b \rightarrow u\ell\nu$ and other rare decays. The training dataset was proportionally sampled in the same manner as described in subsection 6.2.2.

In order to separate this kind of background, we must be careful not to introduce correlations with the fit variables (ΔE , M_{BC}) or any kind of model dependence (correlation with q^2). This means that we can not use any information of the decay particles or the

candidate, which is of kinematics nature, such as decay particles momenta, decay angles or other variables with such behavior.

The training variable set consisted of

- fit probability of $P(\chi^2, DOF)$ of the signal candidate and the ROE side, separately,
- $\cos \theta_{BY}$ from Eq. (4.4),
- \cos of the angle between momentum and vertex of X , where $X \in [KK, KK\ell, KK\ell\nu]$,
- FlavorTagger variables for the two signal-side kaons,
- number of kaons, tracks and distant tracks in ROE,
- θ angle of the ROE momentum in CMS frame,
- ξ_Z from [8]
- Δz ,
- m_{miss}^2 from Eq. (4.11),
- $B \rightarrow D^*\ell\nu$ veto variables,

where distant tracks are all tracks in ROE which satisfy the condition of $|d_0| > 10.0$ cm or $|z_0| > 20.0$ cm. The last entry, $B \rightarrow D^*\ell\nu$ veto variables, are a set of variables where we partially reconstruct the D^* candidate 4-momentum via a linear combination of the π_s^\pm 4-momentum in the $D^* \rightarrow D\pi_s^\pm$ decay. It helps discard the most dominant $B \rightarrow D^*\ell\nu$ background. It is most efficient in the $B^0 \rightarrow D^{*-}\ell^+\nu$ decay, where D^{*-} further decays via $D^{*-} \rightarrow \bar{D}^0\pi_s^-$. Other decays do not contain a charged π_s particle and are harder to reconstruct with good precision. This results in larger suppression of the neutral $B\bar{B}$ background only. Figures 6.5 shows the veto variable with a partial reconstruction of a charged π_s^\pm .

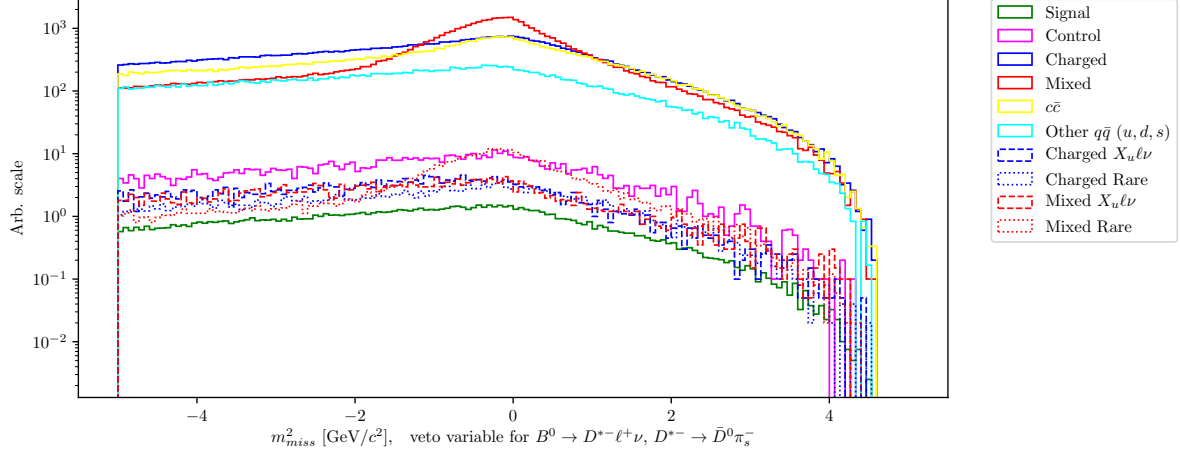


Figure 6.5: Distribution of m_{miss}^2 for partially reconstructed $B^0 \rightarrow D^{*-}\ell^+\nu$ decays, which serves as a veto.

When the training is finished and hyper-parameters of the classifier are optimized, the classifier output, shown in Figure 6.6 (left), can be used for background suppression. B meson candidates from $B\bar{B}$ background are dominant at lower values, while candidates from $B\bar{B}$ events populate the region with higher values. Since the differences between signal and background $B\bar{B}$ events are smaller than $B\bar{B}$ and $q\bar{q}$ events, the resulting classifier has a smaller separation power than in previous section.

6.3.1 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 ΔE 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 [9].

The training set used in this training is the same as described in the beginning of this chapter, along with the same set of training variables. It will be seen later that the standard BDT classifier shapes the background to look like signal mostly in the M_{BC} picture, therefore we train the uBDT classifier with a uniformity constraint in the M_{BC}

variable for background candidates with the uBoost algorithm. The resulting classifier output is shown in Figure 6.6 (right). For this classifier, the separation power between signal and background seems worse, however, the shapes of backgrounds can differ and the final conclusion will be shown only after the signal extraction.

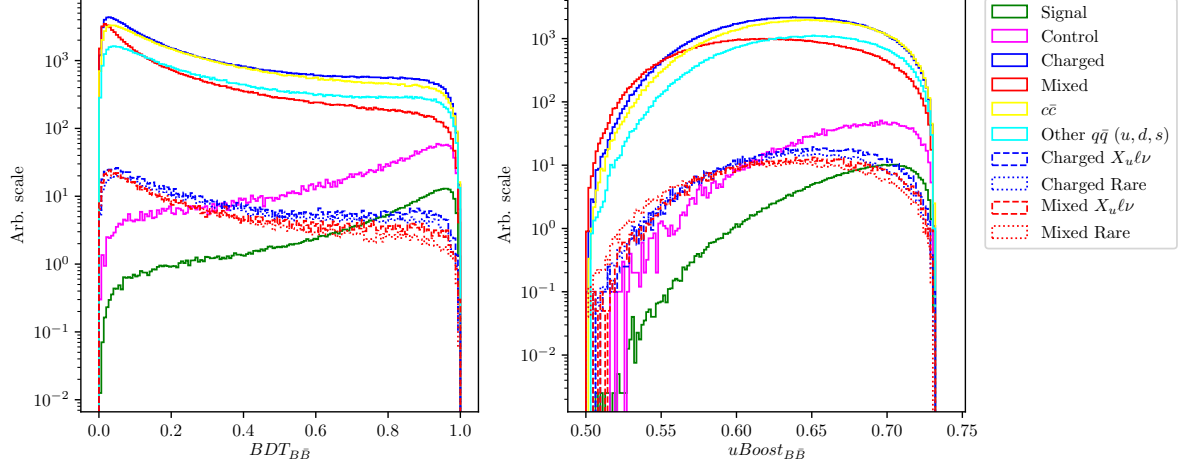


Figure 6.6: $B\bar{B}$ suppression classifier output for signal and various types of background for the standard BDT classifier (left) and the uBDT classifier (right). B candidates from $B\bar{B}$ background events dominate the lower region, while signal and control candidates dominate in the upper region of the classifier output.

6.4 Selection optimization

Instead of two separate $q\bar{q}$ and $B\bar{B}$ FOM optimizations, it is more efficient to do a simultaneous 2D FOM optimization, since the two classifiers are not completely uncorrelated. In the same manner as before, FOM is optimized for perfectly reconstructed signal candidates in the signal window, after all the pre-cuts, signal categorization, and after cutting out the background resonances and the control decay. The FOM plot with the optimal point for both $B\bar{B}$ MVA classifiers is shown in Figure 6.7.

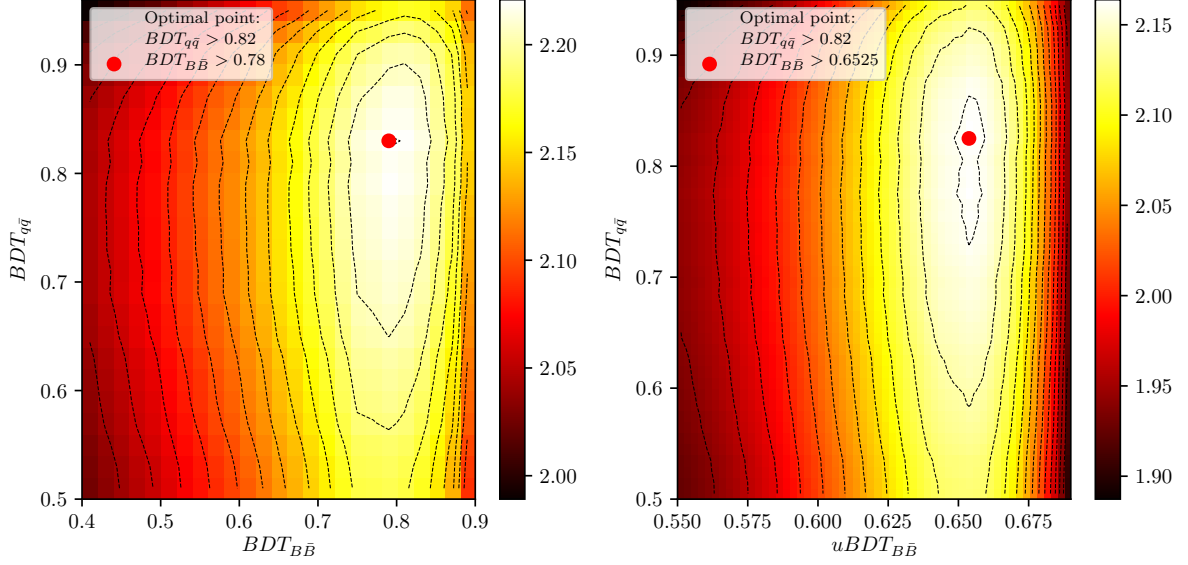


Figure 6.7: 2D *FOM* optimization of continuum suppression classifier and the standard BDT (left) and uBDT (right) $B\bar{B}$ suppression classifier.

We can compare signal and major background distributions of ΔE and M_{BC} after the 2D *FOM* optimization for both classifiers. Figure 6.8 shows the arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the standard BDT classifier, while Figure 6.9 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. 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. The total numbers of expected signal candidates and the signal-to-noise ratios for both classifiers are:

- Standard BDT: $N_{sig} = 152.82$, $N_{sig}/N_{bkg} = 3.70 \%$,

- uBDT: $N_{sig} = 241.56$, $N_{sig}/N_{bkg} = 1.09 \%$.

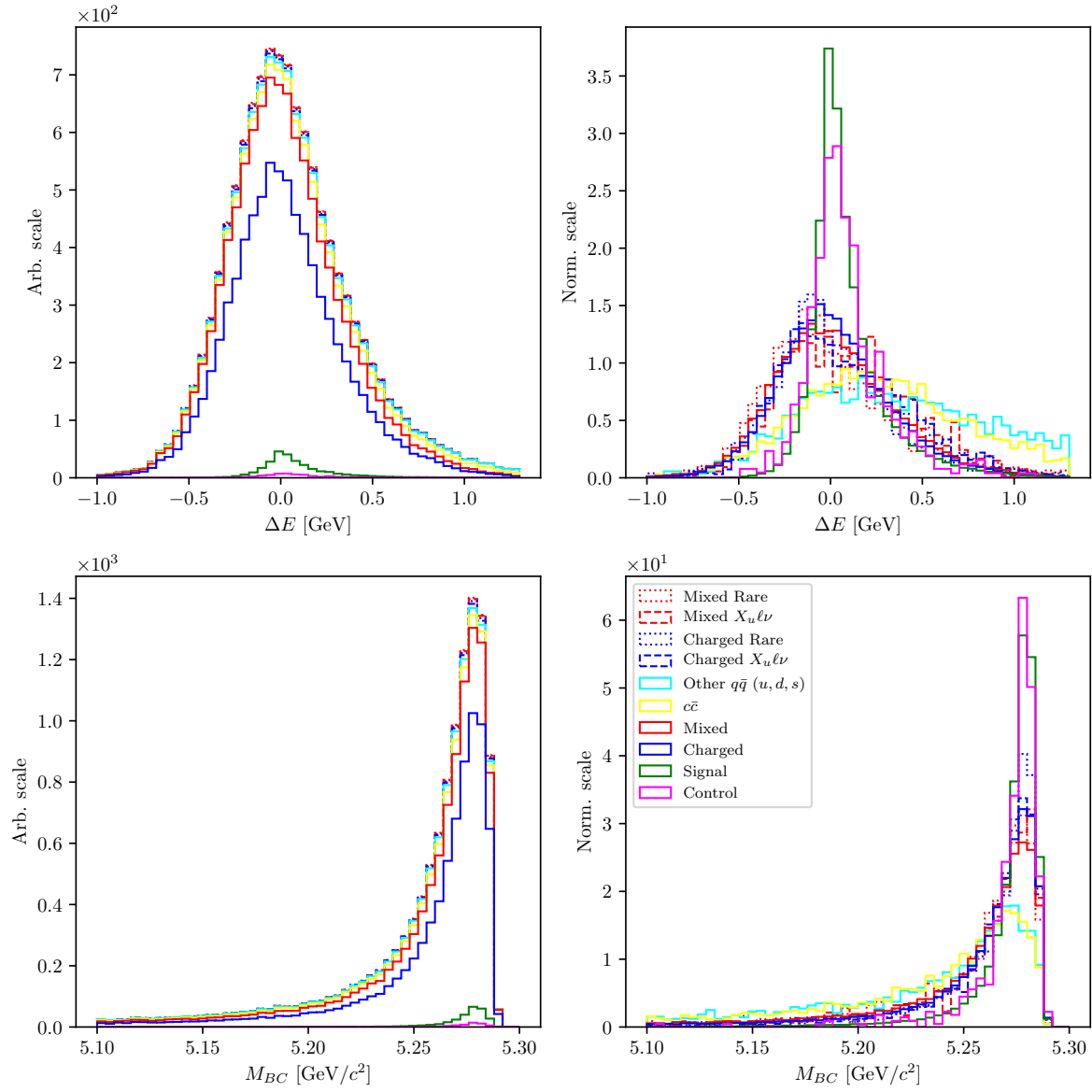


Figure 6.8: Arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the standard BDT classifier.

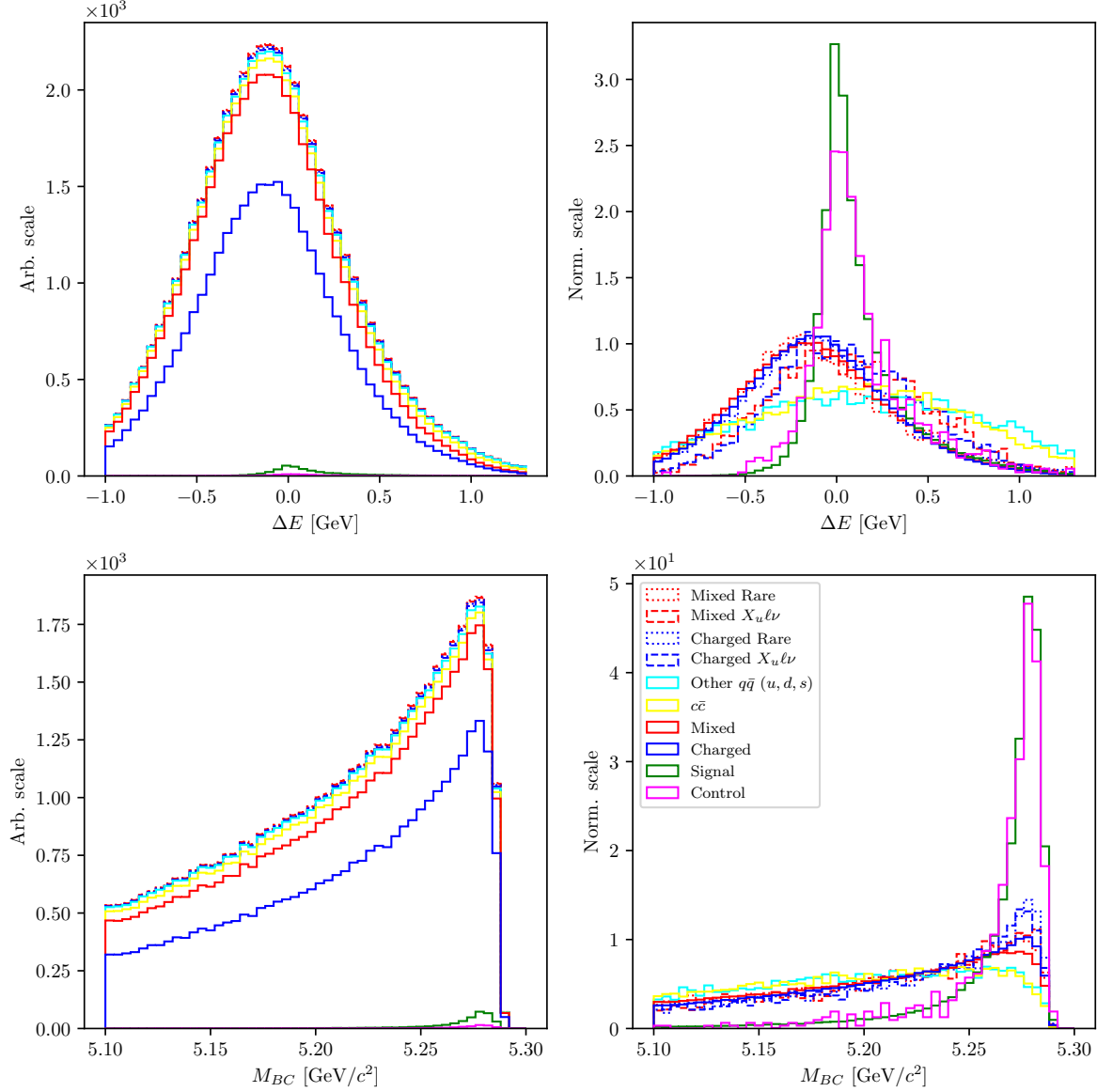


Figure 6.9: Arbitrary (left) and normalized scale (right) for ΔE (top) and M_{BC} (bottom) for the final sample optimized with the uBDT classifier.

6.5 Data and MC agreement

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

the signal MC sample.

The off-resonance data were collected at 60 MeV below the $\Upsilon(4S)$ resonance peak energy in order to determine the non- $B\bar{B}$ background. It therefore offers a direct view of the $q\bar{q}$ background data sample, which we can compare to the off-resonance MC sample. Figure 6.10 shows ΔE , M_{BC} and the $q\bar{q}$ classifier output, $BDT_{q\bar{q}}$, for off-resonance data and MC in the control region, before any MVA cuts, where the MC sample was scaled down by a factor of 6, due to having more streams of MC. Figures show a slight and a consistent overestimation of $q\bar{q}$ MC. However, looking at the shapes of the distributions, the only difference seems to be the normalization factor, so the distribution shapes can be left as they are, since the normalization will be automatically set in the template fit. There also seems to be a difference in the classifier performance for the continuum background suppression on data and MC. This leads to further differences between data and MC in the $q\bar{q}$ sample after the classifier cut, but we estimate that these differences are negligible, since a relatively small amount of continuum background passes the selection, compared to other background types.

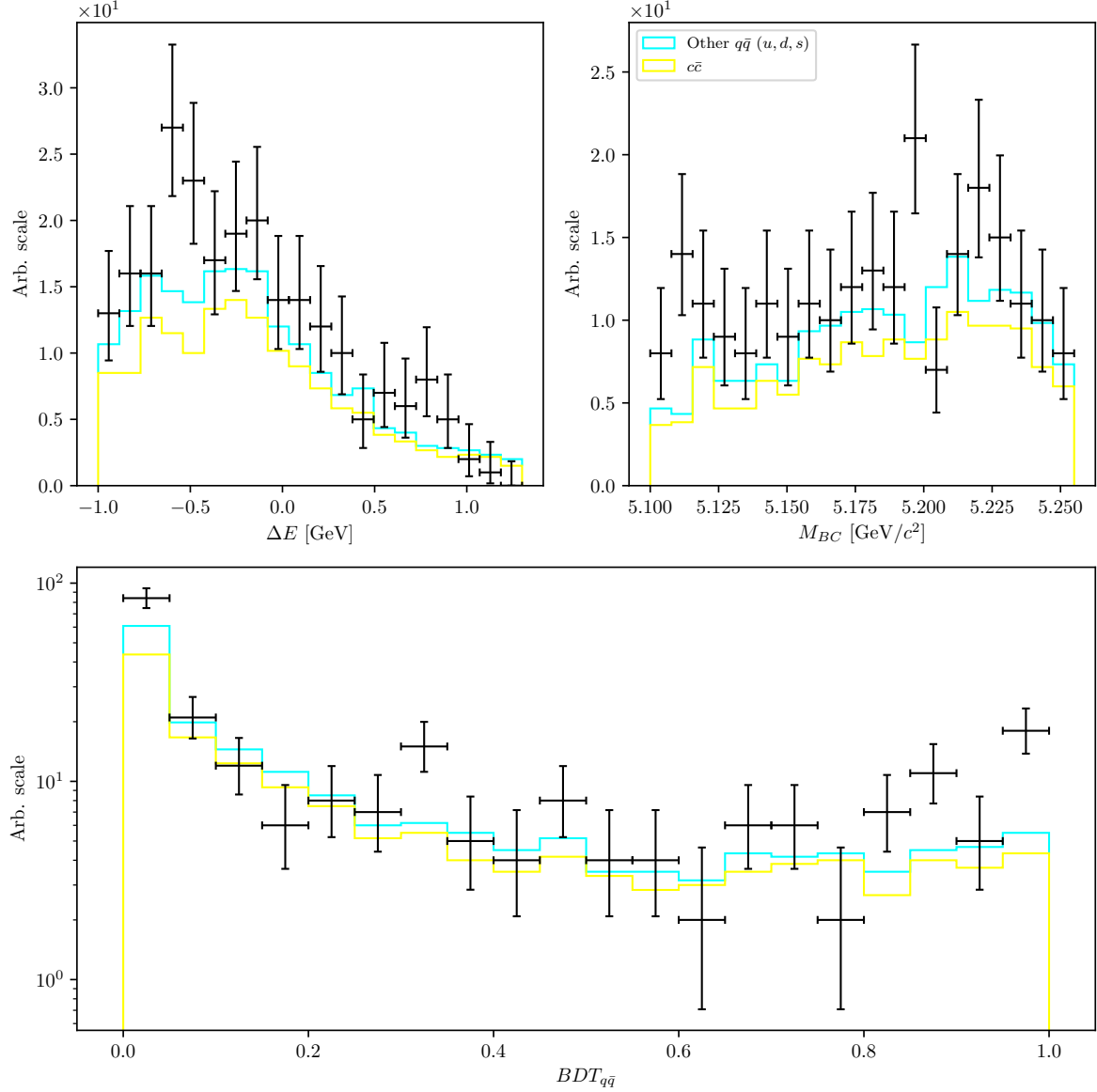


Figure 6.10: ΔE (left), M_{BC} (right) and the $q\bar{q}$ classifier output (bottom), for off-resonance data and MC in the control region before any MVA cuts.

We can repeat the check on on-resonance data. Figure 6.11 shows ΔE , M_{BC} and $BDT_{q\bar{q}}$, where one can see inconsistencies between data and MC on the lower spectrum, where continuum background is dominant. On the other hand, data and MC seem to agree well in the upper part of the spectrum, where $B\bar{B}$ events are dominant, which confirms our statement about data and MC differences in continuum background being negligible from before. Overall, data and MC seem to agree very well already off-the-shelf after all the pre-cuts and without any corrections. This means that the modeling of this MC

sample is very precise in this particular region of data and that there are no significant differences between data and MC for the control sample and the signal sample.

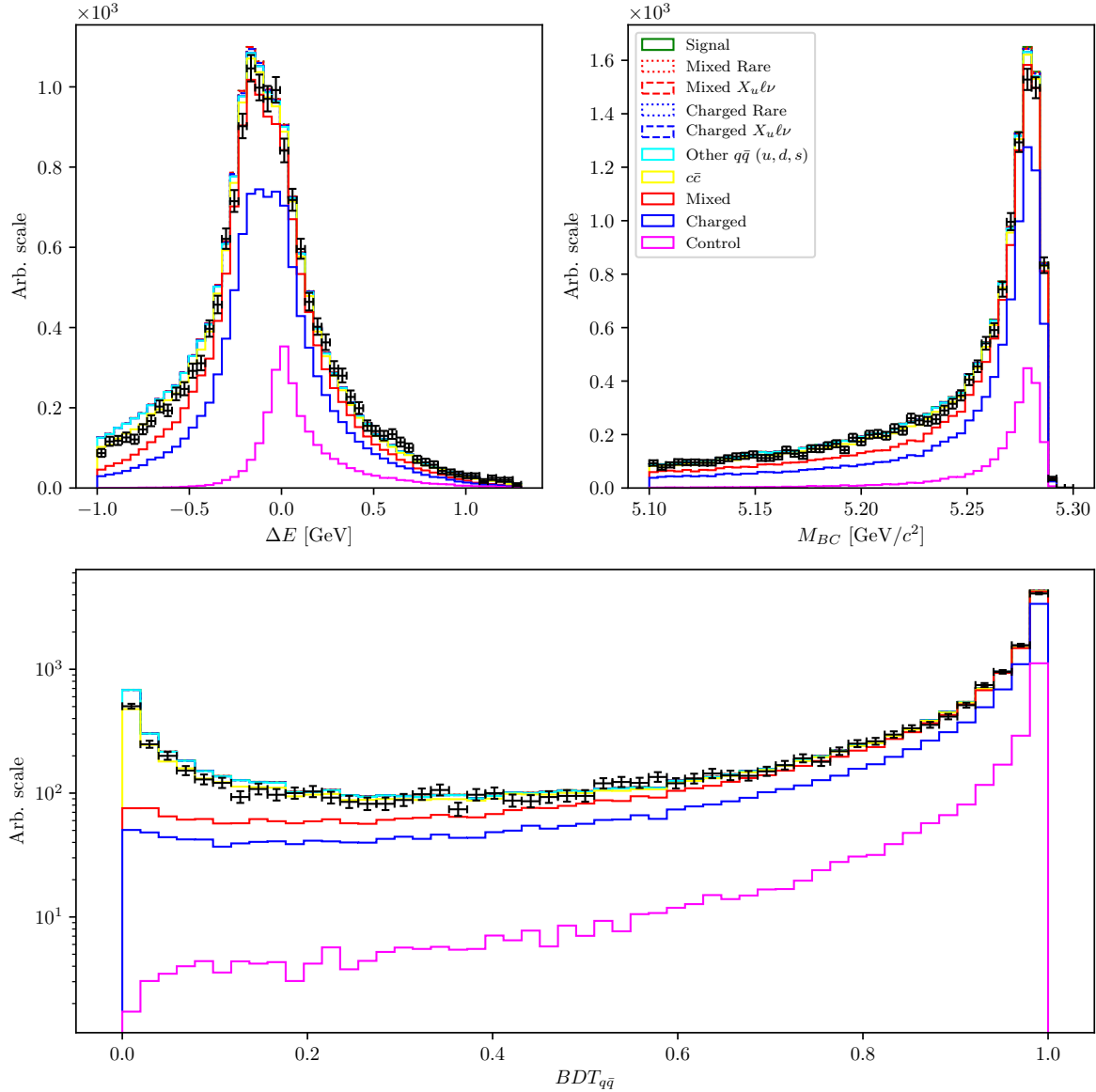


Figure 6.11: ΔE (left), M_{BC} (right) and the $q\bar{q}$ classifier output (bottom), for on-resonance data and MC in the control region before any MVA cuts.

Chapter 7

Signal extraction

In this chapter the procedure for signal yield extraction is presented. We use the framework of RooFit [10] where we define 2D histogram templates in ΔE and M_{BC} , based on MC, for signal and several types of background. Using these templates, the independent full sample is fitted with binned extended maximum likelihood (ML) fit, so that the individual template ratios and their sum describe the fitted sample as best as possible. In particle physics we are often dealing with low numbers of events and need to account for the Poisson nature of the data, therefore we use the likelihood fit, since it takes the Poisson errors into account, unlike the χ^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 on a given event, 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

$$L(\vec{\alpha}) = \prod_{i=1} P(n_i|\vec{\alpha}) = \prod_{i=1} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \quad (7.1)$$

where μ_i is the expected and n_i the measured 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

$$\mathcal{L}(\vec{\alpha}) = -\ln L(\vec{\alpha}) = -\sum_i \ln \left(\frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \right) = \sum_i \ln(n_i!) + \mu_i - n_i \ln(\mu_i). \quad (7.2)$$

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 [11], which is implemented in RooFit. The algorithm follows the log-likelihood function out of the

1019 minimum to find the appropriate intervals of confidence for each parameter, taking the
 1020 parameter correlations into account.

1021 To estimate the goodness of the likelihood fit, one option is to generate toy MC ex-
 1022 periments and obtain the expected log-likelihood distribution. Likelihood fits, however,
 1023 also offer another way to test the goodness of fit via the likelihood ratio (LR), where
 1024 we compare the likelihood obtained under the ML parameters $\vec{\alpha}_{ML}$, to the likelihood
 1025 obtained under the null hypothesis parameters $\vec{\alpha}_{H_0}$. This determines how likely the data
 1026 is under one model than the other. We define the LR test as

$$\lambda = -2 \ln \left(\frac{L(\vec{\alpha}_{ML})}{L(\vec{\alpha}_{H_0})} \right) = -2 [\ln L(\vec{\alpha}_{ML}) - L(\vec{\alpha}_{H_0})] \sim \chi_q^2, \quad (7.3)$$

1027 which asymptotically behaves as the χ_q^2 distribution with $q = m - n$ degrees of freedom,
 1028 where m and n are degrees of freedom of $L(\vec{\alpha}_{ML})$ and $L(\vec{\alpha}_{H_0})$, respectively. In particle
 1029 physics we usually study a specific decay and try to perform measurements of the signal
 1030 yield, so the null hypothesis in this case is that we expect to observe no signal. This
 1031 means that for the null hypothesis we fix the expected signal yield parameter to zero,
 1032 while leaving the other parameters of $\vec{\alpha}_{H_0}$ the same as in $\vec{\alpha}_{ML}$, which results in $n = m - 1$
 1033 degrees of freedom and in their difference $q = m - n = 1$. For such a simple LR test of a
 1034 single parameter the LR test then follows the χ^2 distribution with 1 degree of freedom.
 1035 In this case we can define the fit significance from the χ^2 value in units of σ as

$$\text{Significance} = \sqrt{\lambda} = \sqrt{\chi^2}. \quad (7.4)$$

1036 7.1 Fit templates

1037 The same MC samples are used for template construction as described in Chapter 2

- 1038 • Signal MC,
- 1039 • 10 streams of **charged** and **mixed** $B\bar{B}$ background,
- 1040 • 6 streams of $c\bar{c}$ (**charm**) and other $q\bar{q}$ (**uds**) background,
- 1041 • **ulnu** sample, corresponding to $20\times$ integrated luminosity of the full Belle dataset,
- 1042 • **rare** sample, corresponding to $50\times$ integrated luminosity of the full Belle dataset.

1043 We perform 10 fits to each stream of MC, where 9 streams were used for the creation of
 1044 the templates and the remaining stream was used as data. When fitting real data, all
 1045 available MC was used for creating the templates. The full signal MC sample was used
 1046 for the signal template definition in case of MC as well as data fits. The signal part of
 1047 the **ulnu** sample was not used in template construction, it was only used as a part of
 1048 the fitted sample. The following histogram templates were defined for the fits

- signal template,
- control template,
- $B\bar{B}$ template,
- $q\bar{q}$ template.

When fitting the signal sample, all template shapes were fixed and all template yields were left as floating parameters in the fit, except in the case of control template, where the yield was fixed to the expected MC value, since the sample size was very small and very similar to the signal template. In case of control sample fits, the yield of the control template was left as a floating parameter, while the yield of the signal template was fixed to the expected MC value for the same reasons as mentioned before. Figures 6.8 and 6.9 show ΔE and M_{BC} distributions for the signal template for both final samples, where pronounced peaks can be seen at $\Delta E \approx 0$ GeV and $M_{BC} \approx m_B$.

The remaining three templates for control, $q\bar{q}$ and $B\bar{B}$ sample all describe the background part and, with the exception of the control template, tend to be more widely distributed or even shifted from the nominal value. For the $q\bar{q}$ template the sample of $q\bar{q}$ background was used, whereas in case of the control and $B\bar{B}$ templates, they were constructed using $B\bar{B}$ background, along with the `rare` and `ulnu` background samples with expected relative ratios, based on MC. Figures 6.8 and 6.9 show ΔE and M_{BC} distributions for the three types of background for the final selection using the standard *BDT* and the *uBDT*, respectively.

7.2 Adaptive binning algorithm

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

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

We have devised an algorithm, which compares the templates and the fitted sample, and defines a variable binning so that there are no more problematic bins in the end. Figure 7.1 shows an example of how the procedure works. The algorithm takes an argument for the initial number of uniform bins in each dimension, and does the following

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

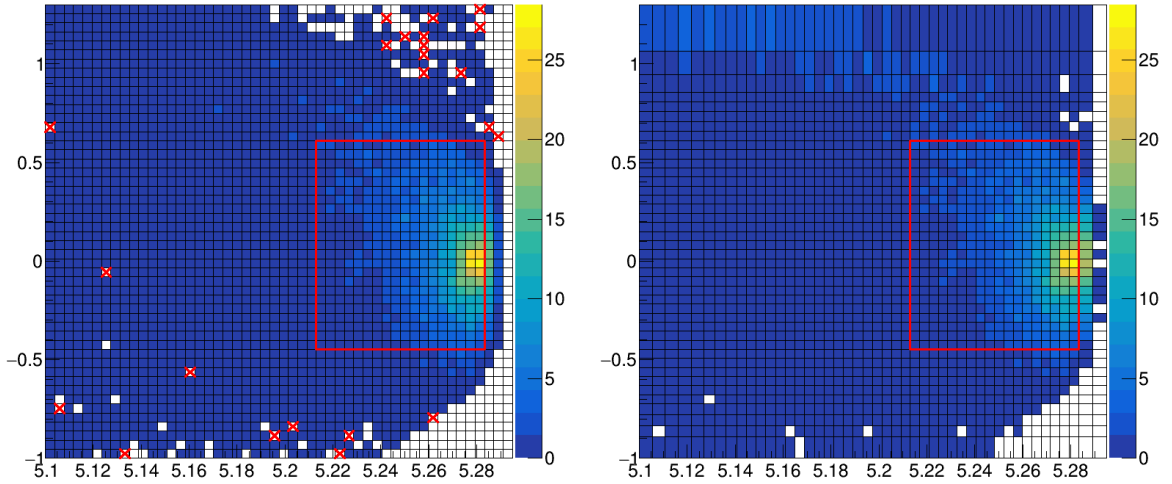


Figure 7.1: Steps taken in the adaptive binning algorithm. Left image shows the initial 2D histogram with the defined optimal region and the problematic bins, the right image shows the final binning with the unchanged optimal region, while the problematic bins are gone due to the new binning choice.

An additional problem seems to be the plotting of the fitted templates and the sample

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

7.3 Signal MC fit results

The fit setup was used on the final samples obtained with both versions of the final selection (standard BDT and uBDT). The choice of the initial uniform binning is not obvious, so we perform fits to all streams of MC for each initial binning choice and the fitted and expected yield difference, pulls and fit significance for both final samples, shown in Figure 7.2. Fits to final sample, obtained with the uBDT classifier generally seem to have a lower bias, pull distributions closer to the normal distribution and better significance than the fits in the case of the standard BDT final sample. This fixes our choice of the final selection. It is also possible to determine the optimal initial binning choice to be somewhere in the region of 22 bins in each dimension, since a higher number of initial bins yields biased results, while a lower one yields results with poorer significance.

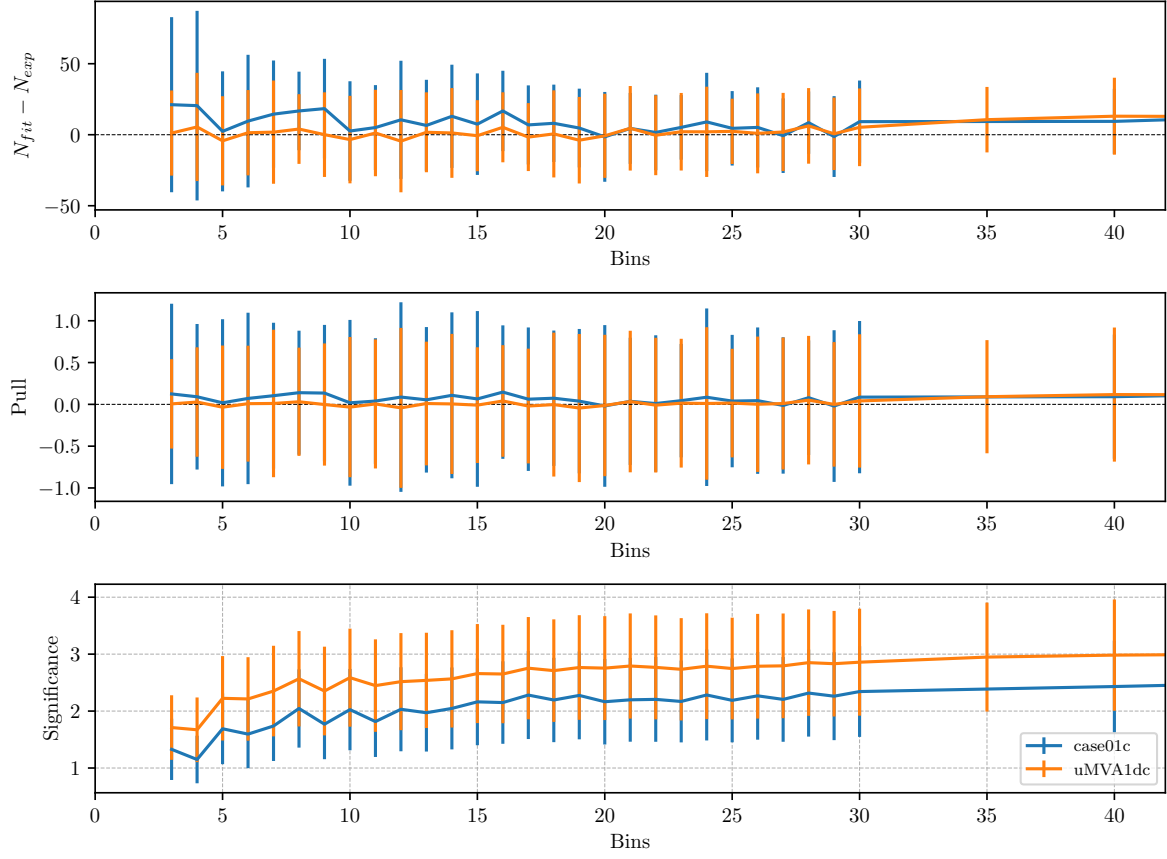


Figure 7.2: Fitted yield and expected yield difference (top), pulls (center) and fit significance (bottom) as a function of binning in ΔE and M_{BC} for the final sample, optimized with the standard BDT classifier (blue) and the uBDT classifier (orange).

Figure 7.3 shows the fit result and the fitted sample of an arbitrary stream for ΔE and M_{BC} , for the fit region and for the signal region. The fits seem stable and under control. Fit results for all streams of MC are shown in Figure 7.4, along with the global fit over all streams. The results seem to describe the expected value in a precise manner, with the bias much smaller than the average statistical error. The normalized χ^2 value with 9 degrees of freedom of the global fit is $\chi^2_9 = 0.72$, while the average significance of the fits is around 2.75σ .

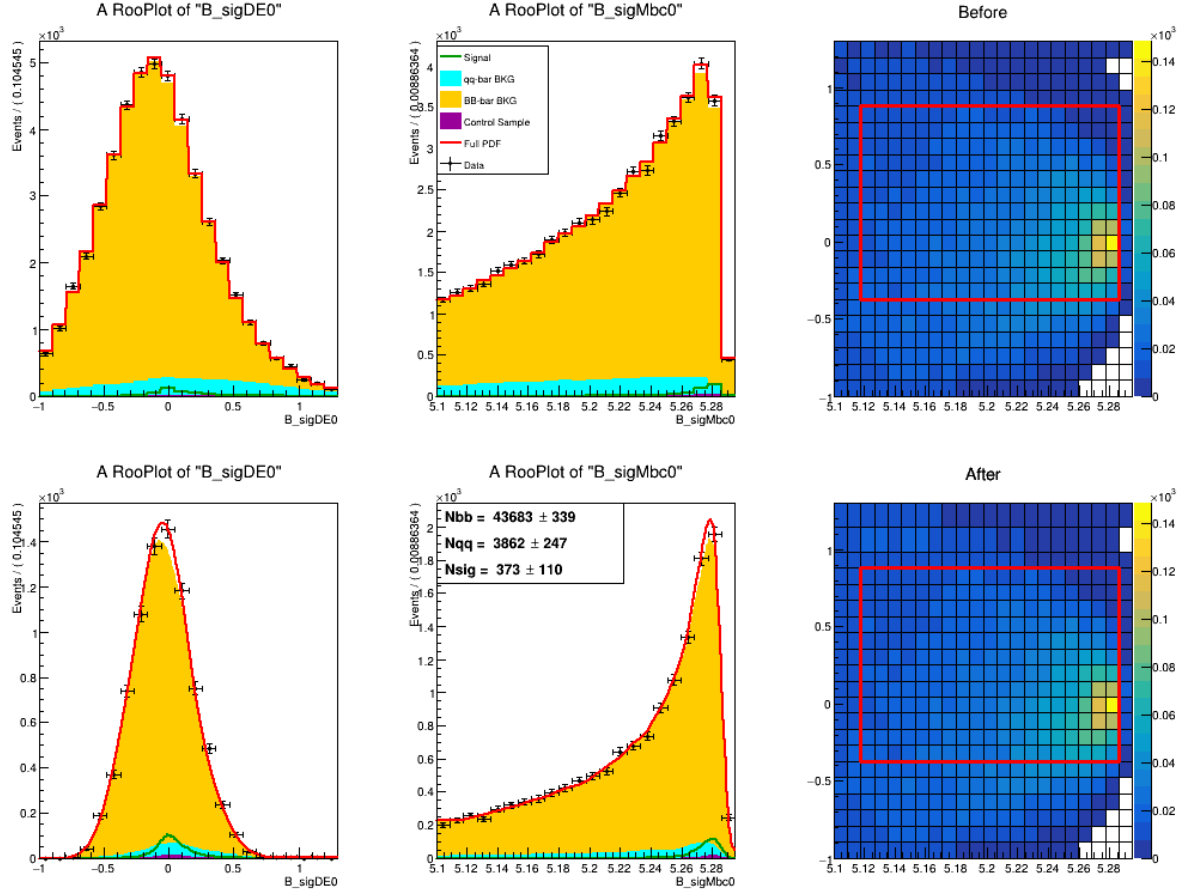


Figure 7.3: An example fit to one stream of MC. Left column shows the M_{BC} and the center column shows the ΔE distribution of the full fitted sample in the full fit region (top) and the signal region (bottom). The right column shows the expected 2D histogram in ΔE and M_{BC} with initial uniform binning (top) and the variable binning (bottom) after the binning algorithm.

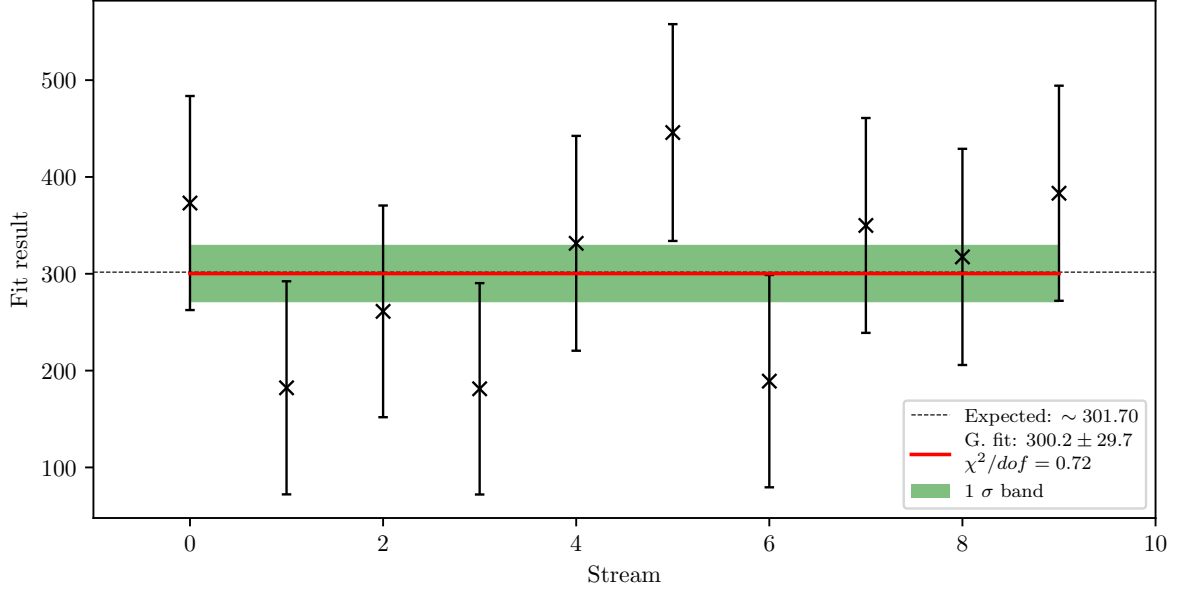


Figure 7.4: Fits to all 10 streams of MC and the global fit with a zero degree polynomial. The red line shows the mean value of the global fit and the green band shows the 1σ confidence interval.

7.3.1 Toy MC experiment

Toy MC experiments allow us to study the yields, errors and the pulls of the signal fit by generating our own pseudo datasets, according to the MC, and not producing it in the standard way, which consumes unaffordable amounts of CPU time. In order to test the fit behavior, a toy MC study was performed where all available MC was used for template creation. We constructed 3×10^3 pseudo datasets, where each dataset was generated expected amount of each template category, distributed according to the Poisson distribution. All fits were performed with the optimal initial uniform binning of 22×22 bins in ΔE and M_{BC} .

Figure 7.5 shows distributions of the fit yields, errors and the pull distribution of all pseudo fits. The fits seem to be under control since there is not any significant bias, and the pulls follow a normal distribution with a mean of $(-4 \pm 18) \times 10^3$ and standard deviation of (1.009 ± 0.013) . The mean (\bar{X}) and the standard deviation (S) were calculated in the usual way, while the errors of these statistics were obtained by calculating the standard error of the mean ($\sigma_{\bar{X}}$) and standard deviation (σ_S), taken from [12], as

$$\sigma_{\bar{X}} = \frac{S}{\sqrt{N}}, \quad \sigma_S = \frac{S}{\sqrt{2(N-1)}}, \quad (7.5)$$

1136 where N is the number of performed measurements.

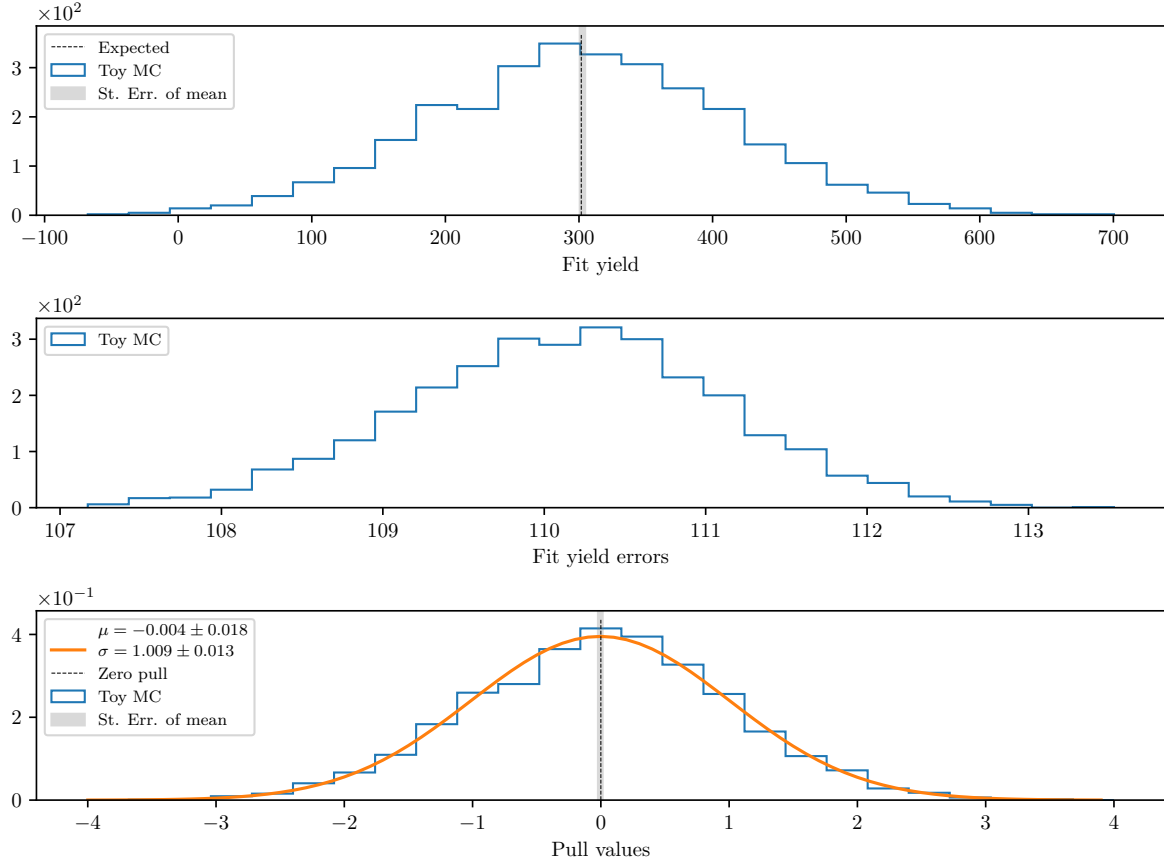


Figure 7.5: Toy MC fits of pseudo data showing the fit yield (top), fit errors (center) and the pull distribution of the fits (bottom).

1137 7.3.2 Toy MC linearity test

1138 Linearity test is used for determining sensitivity of the fit to the amount of signal in
 1139 the fitted sample. Since this is a first measurement of this decay channel, MC modeling
 1140 is not reliable and could be very different from reality, so we need to perform this test
 1141 in order to determine our sensitivity to smaller, as well as larger amounts of expected
 1142 signal.

1143 The pseudo datasets are generated in the same way as in the previous subsection, with
 1144 the exception of signal, which is generated with various amounts. 50 steps from $[0.1, 10]$
 1145 in the logarithmic scale are taken for fractions of signal amount and for each fraction we
 1146 generate 600 pseudo datasets according to Poisson statistics.

Figure 7.6 shows the mean fit yield and expected yield difference, mean pull and the mean significance at each signal fraction value. The expected MC result is shown at fraction value 1. The plots show no significant bias with respect to the signal fraction, while the pulls seems to be described by the normal distributions throughout the fraction range. At expected value we are just below 3σ significance, as already pointed out at the beginning of this section.

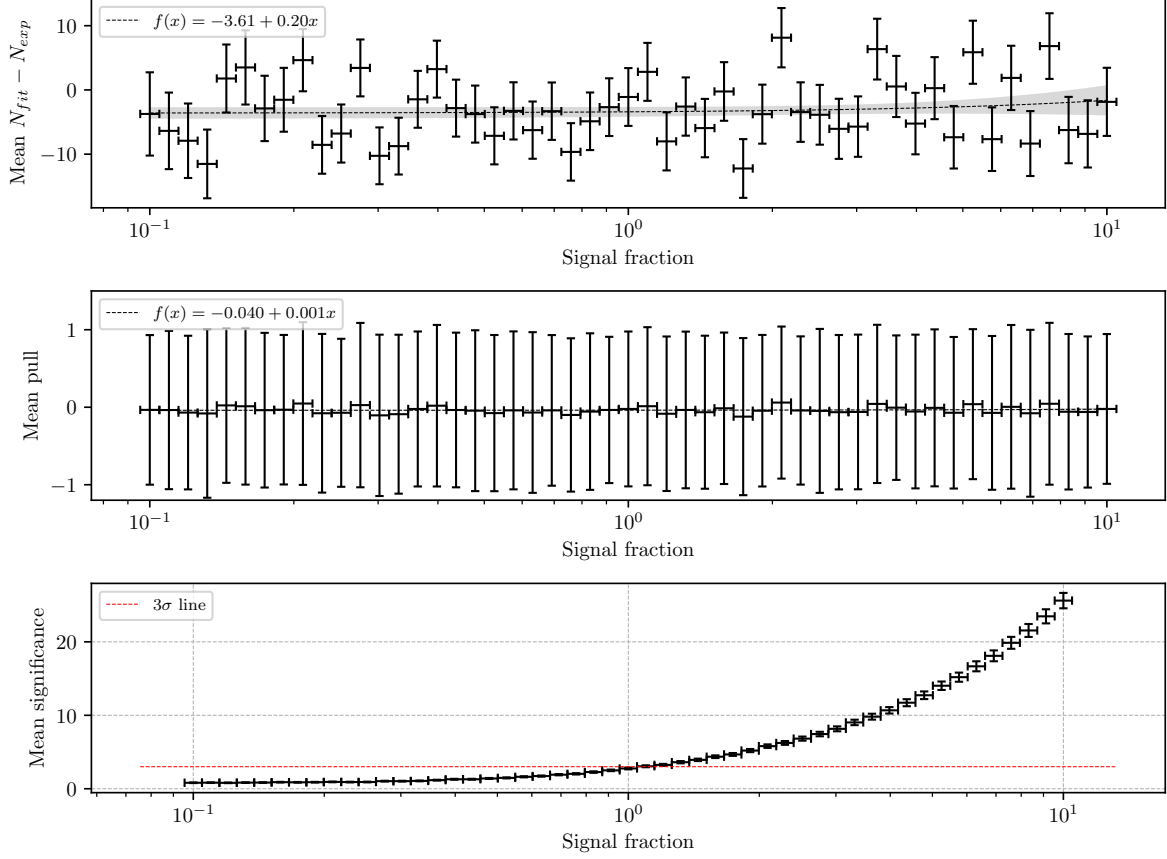


Figure 7.6: Mean fit yield and expected yield difference (top), mean pull (center) and mean significance (bottom) as a function of signal fraction.

7.4 Control fit results

The same fit procedure as in Section 7.3 is applied for the control decay. This enables us to compare the fit behavior on MC as well as data and check the MC/data agreement. The control sample is used to check the efficiency of the MVA classifier cuts for data and MC at various selection stages. We compare control yields, their ratios and ratios of cut efficiencies (double ratios). The following cut scenarios were studied

- 1159 (a) after pre-cuts, $|m_{miss}^2| < 1.1 \text{ GeV}/c^2$, $q_{B_{sig}} q_{B_{comp}} = -1$,
- 1160 (b) (a) + $BDT_{had.}$ cut (equal to signal category I),
- 1161 (c) (b) + $BDT_{q\bar{q}}$ cut,
- 1162 (d) (b) + $uBDT_{B\bar{B}}$ cut,
- 1163 (e) (b) + $BDT_{q\bar{q}} + uBDT_{B\bar{B}}$ cut.

1164 **7.4.1 Fit yields and yield ratios**

1165 **7.4.2 Double ratios**

1166 **7.4.3 Branching ratio measurement and comparison**

1167 **7.5 Signal fit to data**

Chapter 8

1168 Systematics

1169 8.1 Model uncertainty effects

1170 8.2 PID efficiency correction

1171 8.3 Bias

Bibliography

- [1] N. Cabibbo. “Unitary symmetry and leptonic decays”. In: *Physical Review L* 10.12 (1963), p. 531.
- [2] M. Kobayashi and T. Maskawa. “CP-violation in the renormalizable theory of weak interaction”. In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657.
- [3] T. Sjöstrand, S. Mrenna, and P. Skands. “PYTHIA 6.4 physics and manual”. In: *Journal of High Energy Physics* 2006.05 (2006), p. 026.
- [4] T. Keck. “Machine learning algorithms for the Belle II experiment and their validation on Belle data”. Karlsruhe Institut fr Technologie, Diss., 2017. Dr. Karlsruhe Institut fr Technologie, 2017, 240 pages. URL: <https://ekp-invenio.physik.uni-karlsruhe.de/record/48940>.
- [5] T. Keck. “Fastbdt: A speed-optimized multivariate classification algorithm for the belle ii experiment”. In: *Computing and Software for Big Science* 1.1 (2017), p. 2.
- [6] D. Asner et al. “Search for exclusive charmless hadronic B decays”. In: *Physical Review D* 53.3 (1996), p. 1039.
- [7] A. Bevan et al. “The physics of the B factories”. In: *The European Physical Journal C* 74.11 (2014), p. 3026.
- [8] P. del Amo Sanchez 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 2011), p. 032007. DOI: [10.1103/PhysRevD.83.032007](https://link.aps.org/doi/10.1103/PhysRevD.83.032007). URL: <https://link.aps.org/doi/10.1103/PhysRevD.83.032007>.
- [9] J. Stevens and M. Williams. “uBoost: A boosting method for producing uniform selection efficiencies from multivariate classifiers”. In: *Journal of Instrumentation* 8.12 (2013), P12013.
- [10] W. Verkerke and D. Kirkby. “The RooFit toolkit for data modeling”. In: *Statistical Problems in Particle Physics, Astrophysics and Cosmology*. World Scientific, 2006, pp. 186–189.
- [11] F. James. “MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1”. In: (1994).
- [12] S. Ahn and J. A. Fessler. “Standard errors of mean, variance, and standard deviation estimators”. In: *EECS Department, The University of Michigan* (2003), pp. 1–2.

Addendum A: ROE clean-up π^0 training

Correlations

Importance

Hyper-parameter optimization

Results

1209 **Addendum B: ROE clean-up γ** 1210 **training**

1211 **Correlations**

1212 **Hyper-parameter optimization**

1213 **Importance**

1214 **Validation**

1215 **Addendum C: ROE clean-up**
1216 **duplicate pair training**

1217 **Correlations**

1218 **Hyper-parameter optimization**

1219 **Importance**

1220 **Validation**

1221 **Addendum D: ROE clean-up**
1222 **duplicate track training**

1223 **Correlations**

1224 **Hyper-parameter optimization**

1225 **Importance**

1226 **Validation**

1227 Addendum E: $q\bar{q}$ suppression 1228 training

1229 Correlations

1230 Hyper-parameter optimization

1231 Importance

1232 Validation

1233 Addendum F: $B\bar{B}$ suppression 1234 training

1235 Correlations

1236 Hyper-parameter optimization

1237 Importance

1238 Validation