# Measurement of the decay $B \to KK\ell\nu$ with B2BII

Matic Lubej

Ljubljana, 2018

# Changelog

- First submission of the note.

# Contents

# 1 Introduction

Will add more info later

# 2 Data and MC

Will add more info later

## 2.1 MC

- Belle generic MC, on Y(4S) resonance energy, converted with B2BII
- Belle generic MC, off resonance energy, converted with B2BII (for checks)
- Belle rare + ulnu MC, converted with B2BII
- Signal MC, produced with similar script to charged ulnu, converted with B2BII

## 2.2 Data

- Full Belle dataset, on resonance, converted with B2BII
- Full Belle dataset, off resonance, converted with B2BII (for checks)

# 3 B2BII conversion

## 3.1 Description

## 3.2 Validation

# 4 Event reconstruction

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

## 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 kaons ($K$) and the light lepton, which is the electron ($e$) or muon ($\mu$). These are some of the particles which are commonly referred to as final state particles (FSP). Final state particles have a long lifetime and are usually the particles that we detect when they interact with the material in the detector.

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

#### 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.

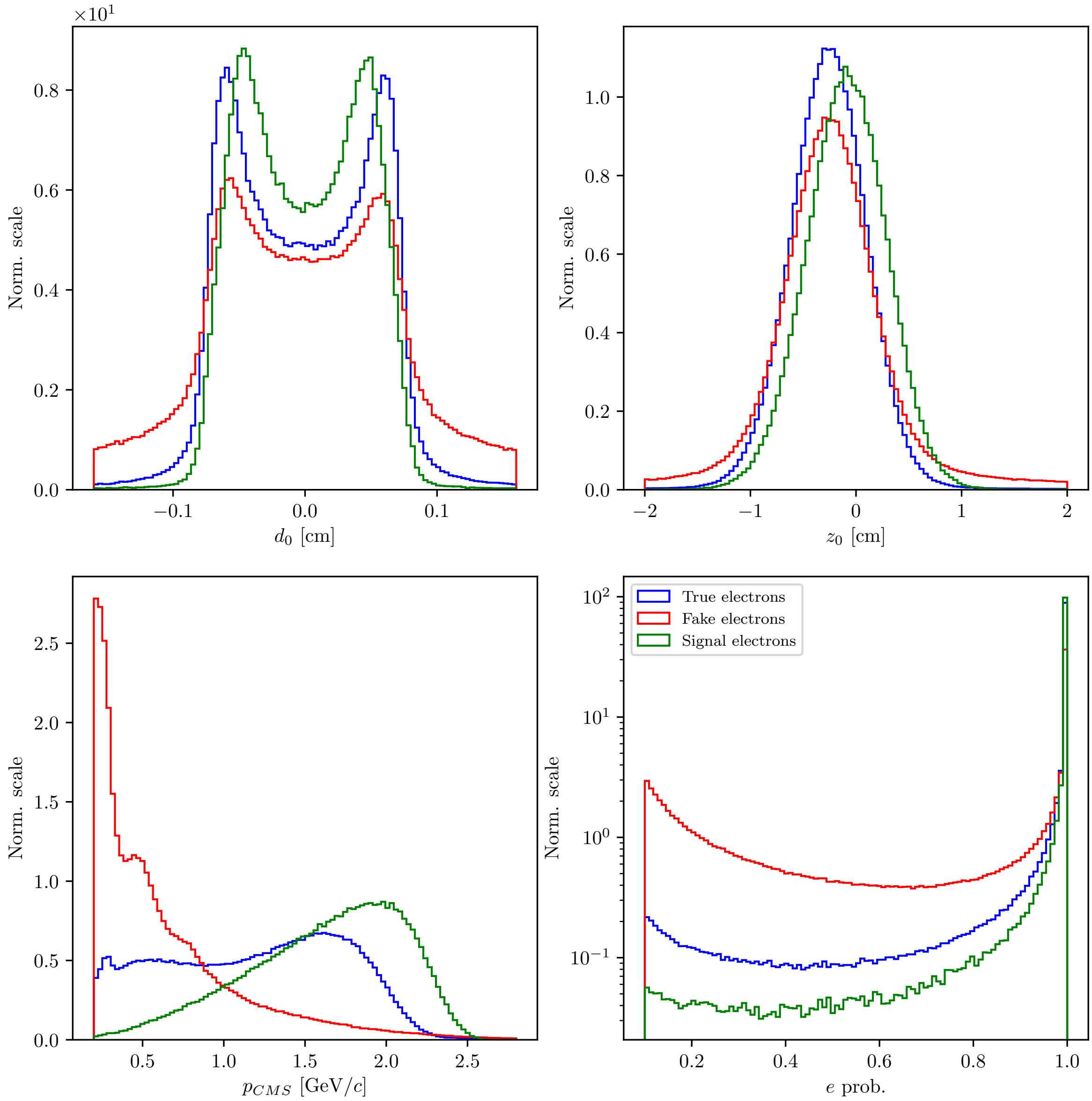


Figure 4.1: Normalized properties of true (blue), fake (red) and true electrons from signal decay (green)

.

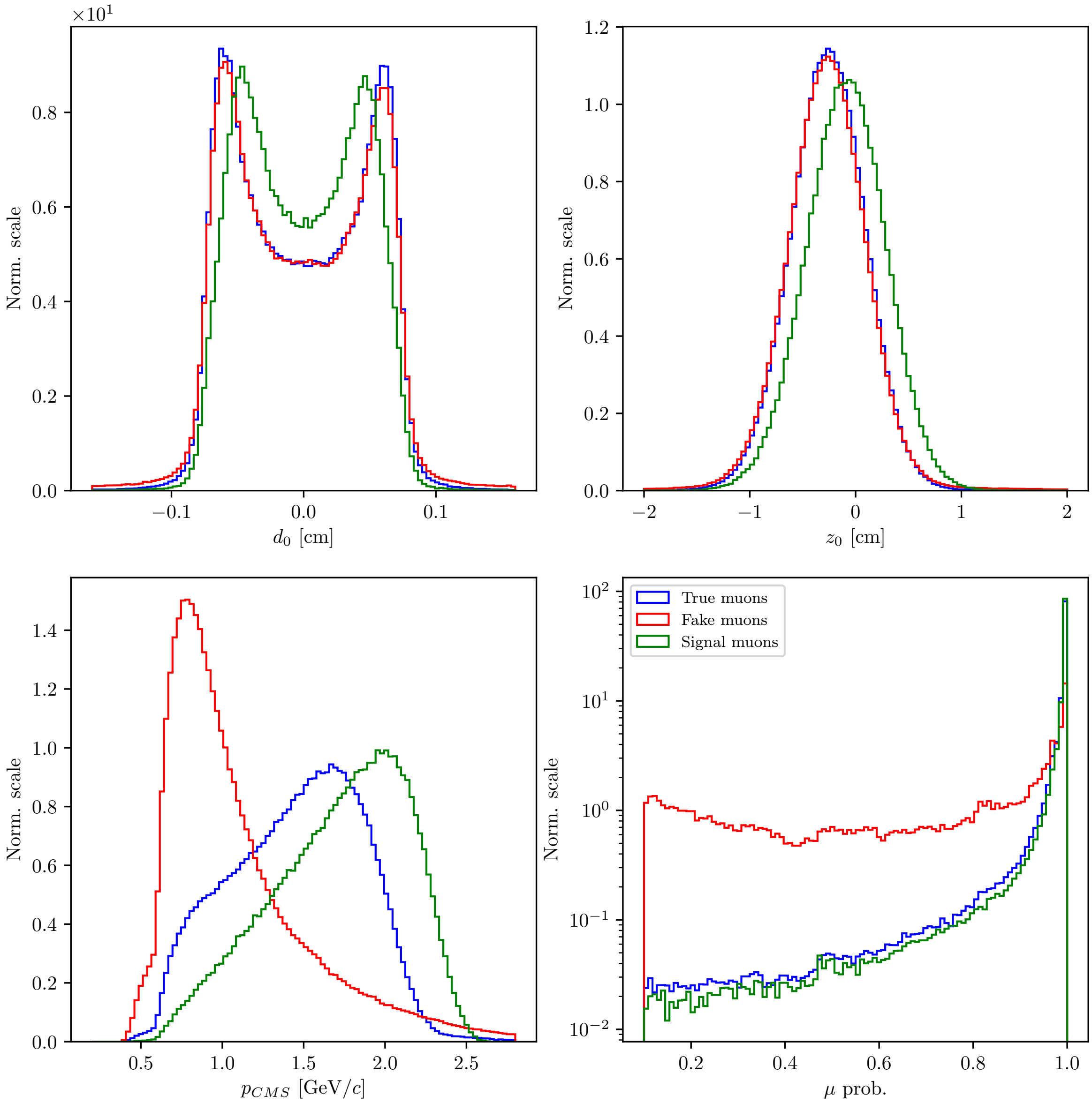


Figure 4.2: Normalized properties of true (blue), fake (red) and true muons from signal decay (green)

.

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 by maximizing the standard definition of *figure of merit* (FOM), defined in Eq. (4.1)

$$FOM = \frac{S}{\sqrt{S+B}}, \tag{4.1}$$

where $S$ represents number of signal and $B$ the number of background candidates.

The $FOM$ plot is 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,
- $\mu$ prob. $> 0.97$ for muons.

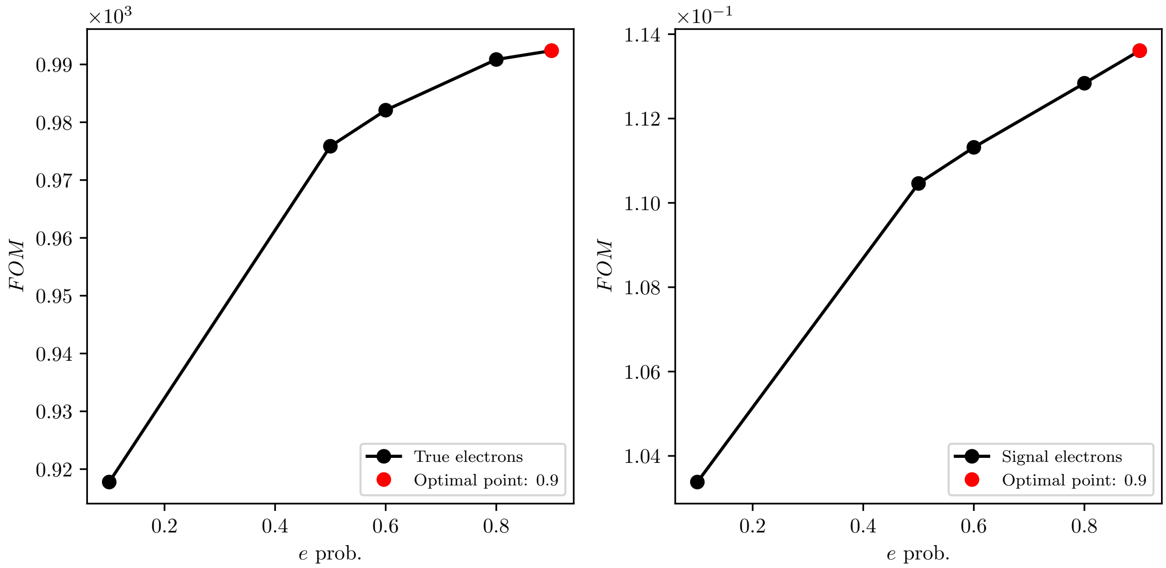


Figure 4.3: $FOM$ optimizations of the PID probability cuts for true electrons (left) and true electrons from signal decays (right).

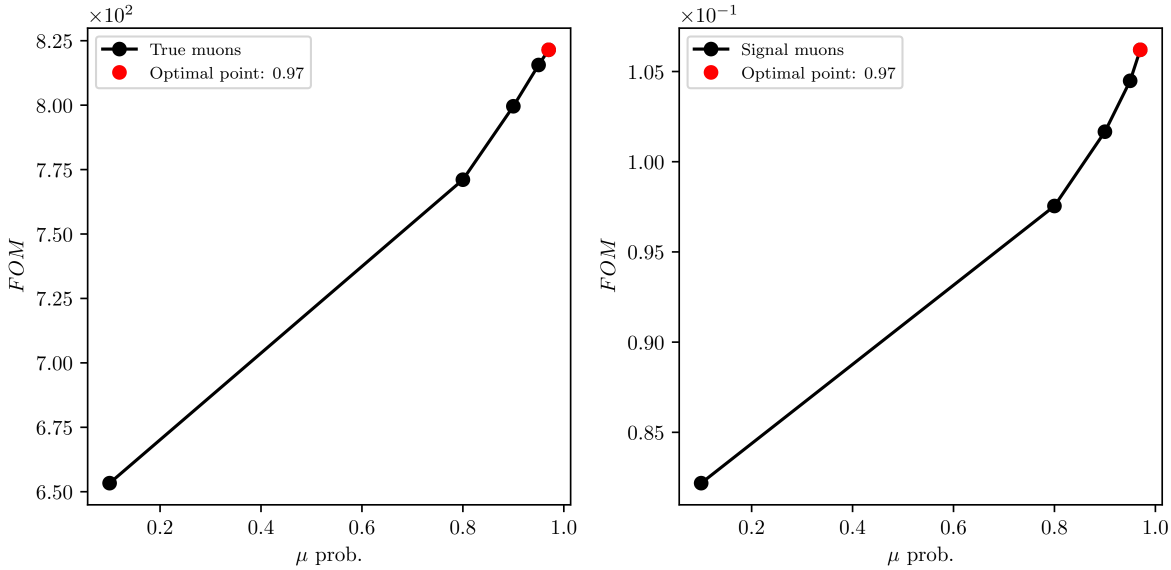


Figure 4.4: $FOM$ optimizations of the PID probability cuts for true muons (left) and true muons from signal decays (right).

### Kaons

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

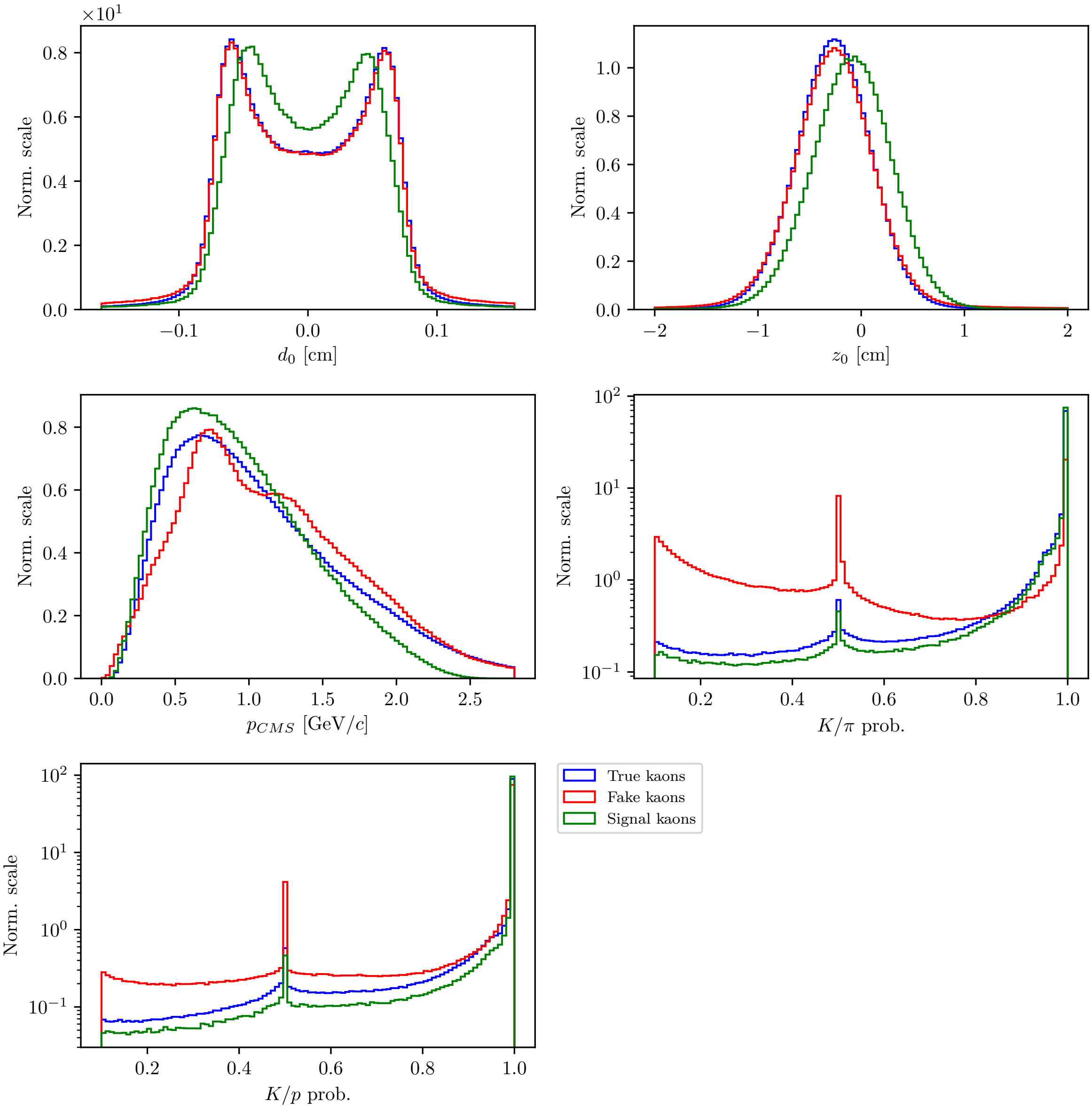


Figure 4.5: Normalized properties of true (blue), fake (red) and true kaons (green) from signal decay

.

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 $K/\pi$, and after that the $K/p$ separation probability. Figure 4.6 shows the optimization procedure for PID cuts on kaon candidates.

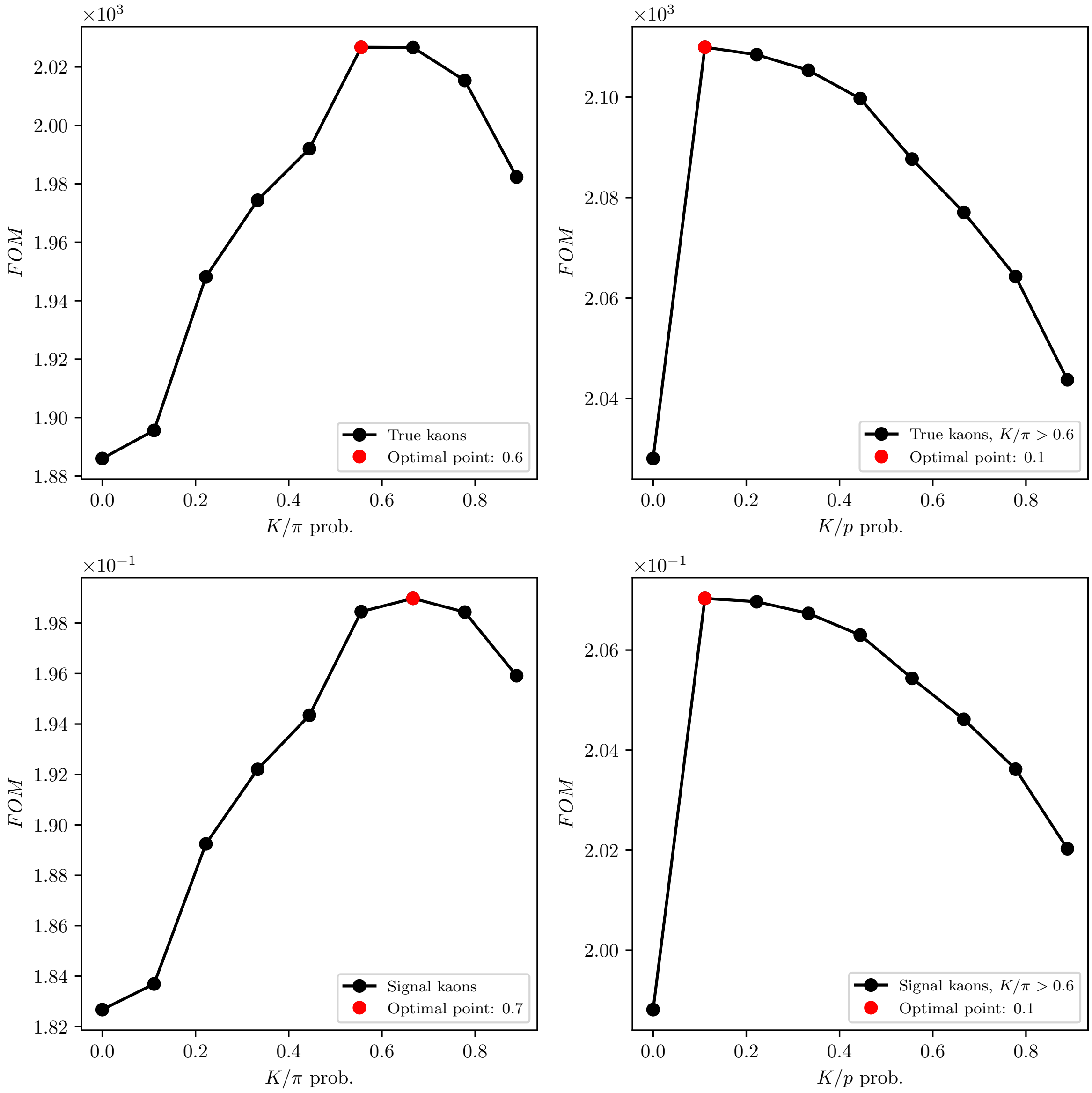


Figure 4.6: $FOM$ optimizations of the PID probability cuts for true kaons (top) and true kaons from signal decays (bottom). The plots on the left show the optimization of the first step for the $K/\pi$ probability cut, and the plot on the right the $K/p$ probability cut.

The optimized PID cuts for kaons are

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

## 4.2 Combination of FSP particles

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

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

and similarly for $B^-$. When an arbitrary combination is obtained, we perform a vertex fit of the three tracks in order to discard combinations with a low probability of tracks coming from the same point. $B$ mesons have a relatively long lifetime and decay along the z-axis of the detector in the direction of the boost, so the vertex fit is enforced with an *iptube* constraint, which constrains the vertex to an elongated ellipsoid along beam direction. We demand that the fit converged and apply a cut on the minimal fit probability. The fit probability for signal and background $B$ meson candidates is shown in Figure 4.7 (left). We perform a $FOM$ cut optimization of this variable, which is shown in Figure 4.7 (right).

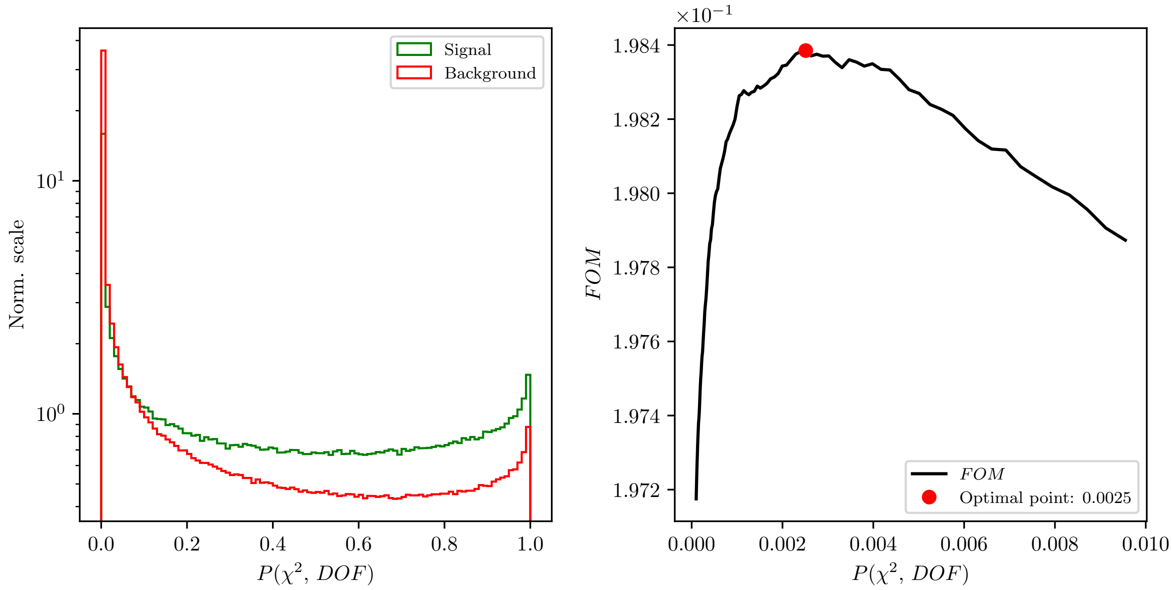


Figure 4.7: Normalized vertex fit probability distribution for signal and background $B$ meson candidates in logarithmic scale (left) and $FOM$ optimization of the vertex fit probability (right).

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

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

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

$$\mathrm{p}_\nu = \mathrm{p}_B - \mathrm{p}_Y, \tag{4.2}$$

$$\mathrm{p}_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_B E_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \tag{4.3}$$

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

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

$$E_B = E_{CMS}/2, \tag{4.5}$$

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

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

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

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

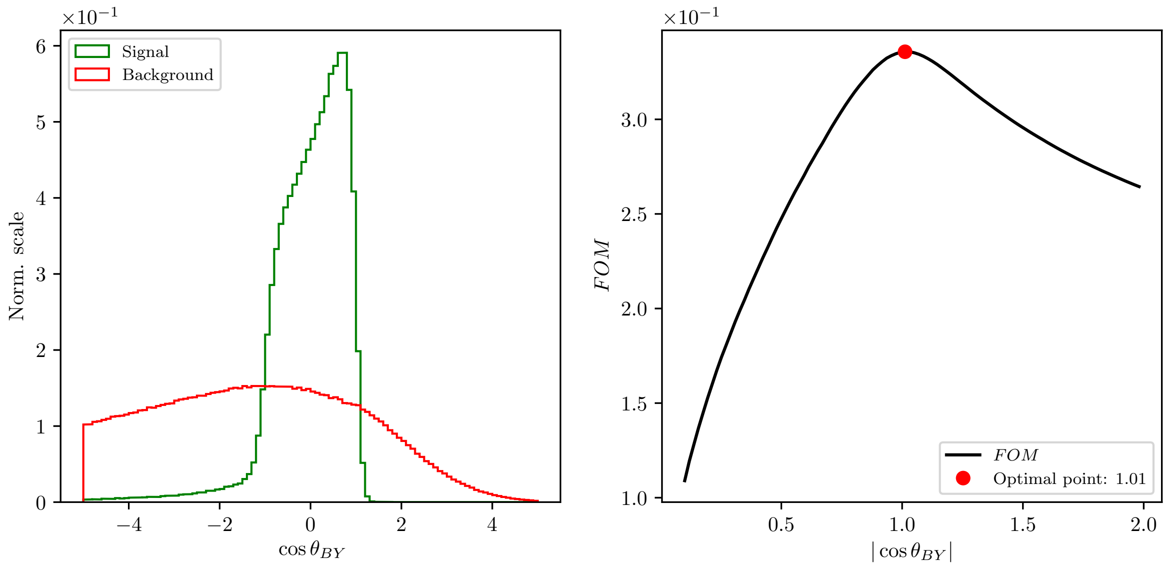


Figure 4.8: Normalized $\cos\theta_{BY}$ distribution for signal and background $B$ meson candidates (left) and $FOM$ optimization of the $\cos\theta_{BY}$ variable (right).

## 4.3 Rest of event clean-up

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. 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 track or cluster in the final calculations of the ROE four-momentum.

In order to preserve the continuity of this chapter, a more detailed description of the ROE clean-up can be found in Chapter 5. From this point on we assume the ROE to be efficiently cleansed of extra tracks and clusters .

## 4.4 Loose neutrino reconstruction

The signal-side neutrino escapes detection, so we cannot directly determine it's four-momentum. However, due to the detectors geometry, which almost completely covers the full solid angle, and due to well known initial conditions of the $\Upsilon(4S)$ meson, it is possible to determine the kinematics of the missing neutrino via indirectly reconstructing the companion $B$ meson by summing up all the FSP particles four-momenta in the ROE. This is known as the *untagged* method. Here we see the motivation for the ROE clean-up, since our signal candidate reconstruction depends on the other $B$ meson.

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

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \sum_{i}^{\text{Event}} (E_i, \vec{p}_i) \,, \tag{4.7}$$

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \left( \mathrm{p}_Y - \sum_{i}^{\text{Rest of event}} (E_i, \vec{p}_i) \right) , \tag{4.8}$$

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

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

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

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

$$\mathrm{p}_{\nu} = \mathrm{p}_{miss} = (E_{miss}, \vec{p}_{miss}) \,, \tag{4.10}$$
$$m^2_{miss} = \mathrm{p}^2_{miss} = \mathrm{p}^2_{\nu} = m^2_{\nu} \approx 0. \tag{4.11}$$

Since the detector is not perfect, the distribution of the $m^2_{miss}$ 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 $K^0_L$, or simply missing tracks due to detection failure. Figure 4.9 shows the distribution of $m^2_{miss}$ 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 $m^2_{miss}$ variable in order to partially discard candidates with spoiled properties, even if it was in principle a correct combination of FSP particles on the signal side

- $|m^2_{miss}| < 7$ 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 it's particles were taken into account correctly[1]. We denote this subset as *perfect* signal.

[1]We only allow for missing photons, since all photons are much harder to correctly take into account.

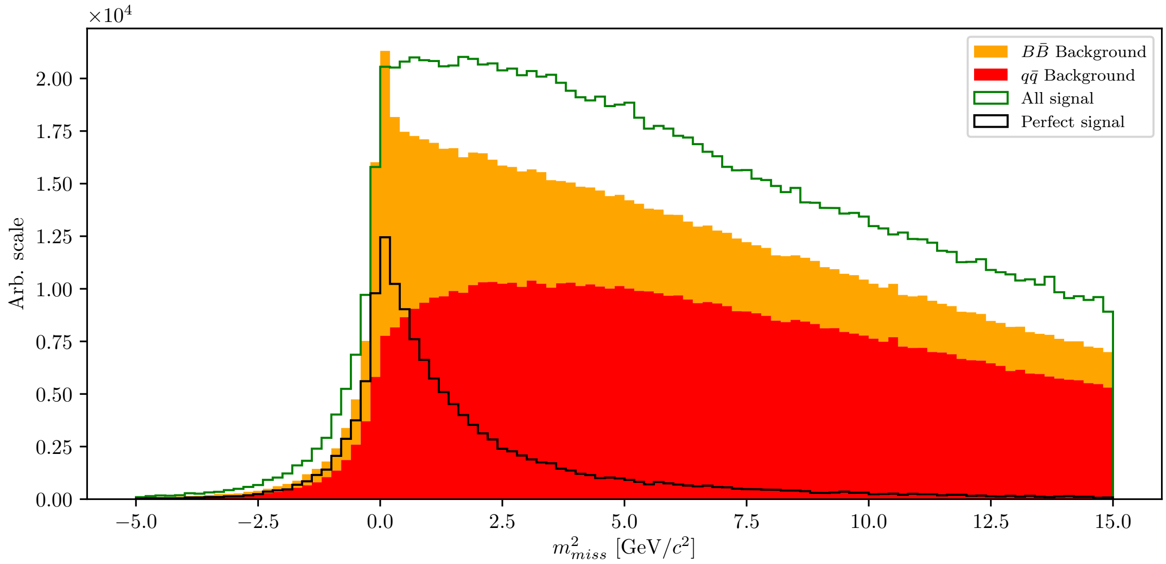


Figure 4.9: Squared missing mass distribution for $q\bar{q}$ and $B\bar{B}$ background in filled histograms. All signal (green) and perfect signal (black) are scaled up equally.

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

$$\mathrm{p}_\nu = (|\vec{p}_{miss}|, \vec{p}_{miss}) \,, \tag{4.12}$$

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

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

$$\Delta E = E_B - E_{CMS}/2, \tag{4.13}$$

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

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

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

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

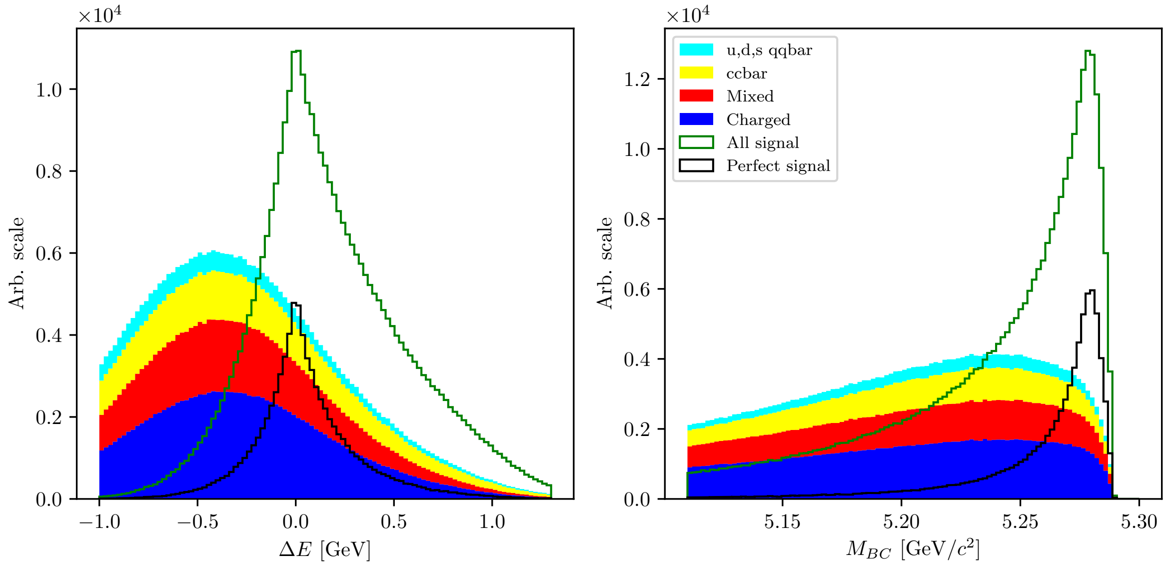


Figure 4.10: Distributions of $\Delta E$ (left) and $M_{BC}$ (right) for signal and major types of background after the precuts. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution in both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

## 4.5 $q^2$ calculation

Momentum transfer squared, $q^2$, is the squared Lorentz invariant of the four-momentum which is transferred from the $B$ meson to the $W$ boson. There are several possible calculations of this variable which offer different resolution. The following describes the calculation of $q^2$ which follows the calculation from [] and yields the best resolution.

For correctly reconstructed events Eq. (4.13) satisfies the condition $\Delta E \approx 0$ within precision. It is possible to rescale the neutrino energy in such way that we fix $\Delta E$ to zero, meaning

$$\Delta E' = (E_Y + \alpha E_\nu) - E_{CMS}/2 = 0. \tag{4.15}$$

and calculate and adapted version of $M_{BC}$

$$M'_{BC} = \sqrt{\left(E_{CMS}/2\right)^2 - |\vec{p}_Y + \alpha\vec{p}_\nu|^2}. \tag{4.16}$$

The neutrino momentum resolution dominates the $\Delta E$ uncertainty [], so the correction factor $\alpha$ reduces this source of uncertainty.

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

The corrected neutrino momentum is then solely used for the $q^2$ calculation, alongside the reconstructed lepton four-momentum, as

$$q^2 = \mathrm{q}^2 = (\mathrm{p}_\ell + \mathrm{p}_\nu)^2 \, . \tag{4.17}$$

The $q^2$ distribution and its resolution are shown in Figure 4.11, along with additional versions of $q^2$, with details of the calculation method in the caption.

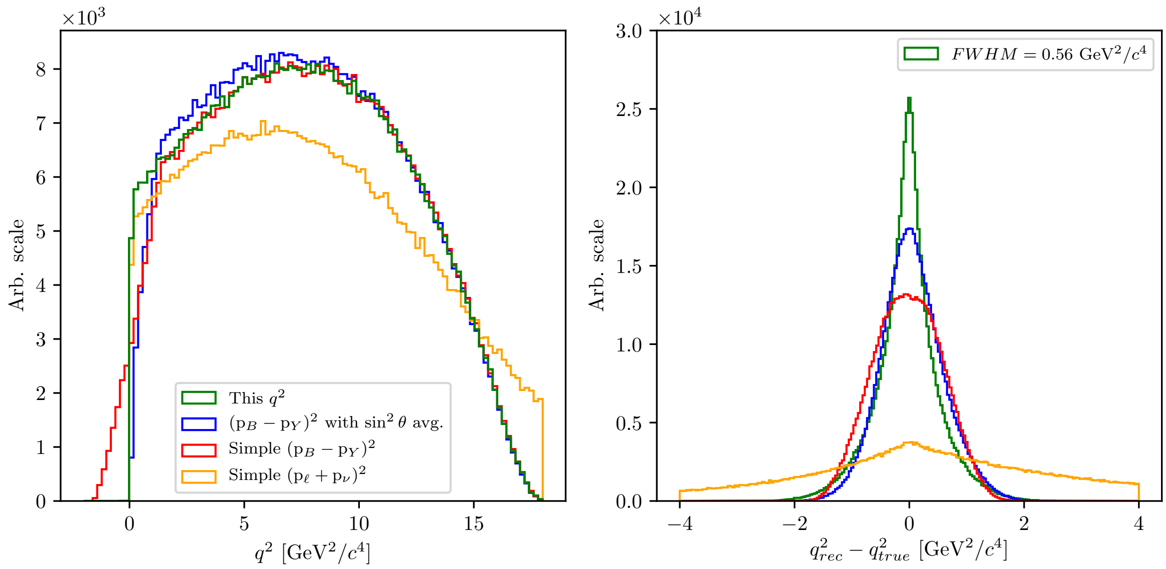


Figure 4.11: Distributions of $q^2$ (left) and $q^2$ resolution (right) for various methods of $q^2$ calculation. The blue distribution follows the procedure in [], whereas the red and the orange ones are straight-forward calculations with available information in the reconstruction. The $q^2$ calculation in red assumes a resting $B$ meson in the CMS frame, and the calculation in orange uses the neutrino four-momentum from Eq. (4.10).

One must bare in mind that even though this calculation yields a precise result, this says nothing about the correctness of the $q^2$ model which was used (ISGW2) []. Since this decay has not been observed yet, we do not have a good description of the decay model, which is also a source of systematics in this analysis.

## 4.6 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 out 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

$$q_{B_{sig}}\, q_{B_{comp}} = -1, \tag{4.18}$$

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 4.6.1.

We define the 4 categories in the following way

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

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

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

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

Different categories for all signal candidates after the pre-cuts are shown in Figure 4.12. Category I has the best resolution, so we use it in our analysis from this point on.

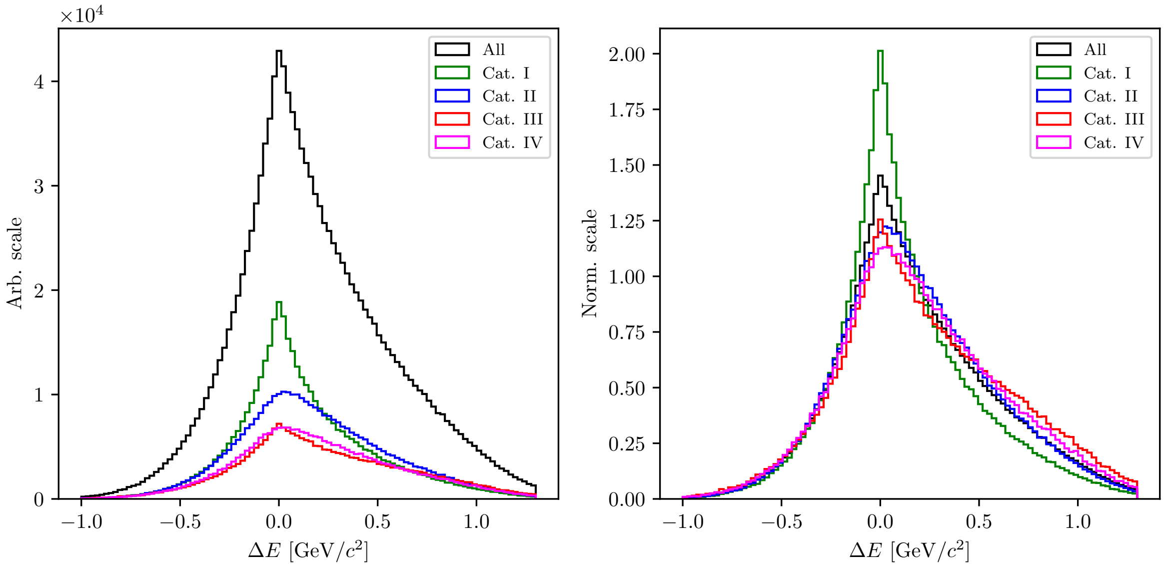


Figure 4.12: Categorization of signal candidates based on the charge product of both $B$ mesons in the event and the MVA output for recognizing hadronic decays of the companion $B$ meson. The plot on the left shows the distributions in an arbitrary scales, while the plot on the right shows the normalized distributions.

### 4.6.1 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 $\ell$ 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 4.8.

The classifier output is shown in Figure 4.13 (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 4.13 (right), yields the optimal cut

- $BDT_{had.} > 0.57$.

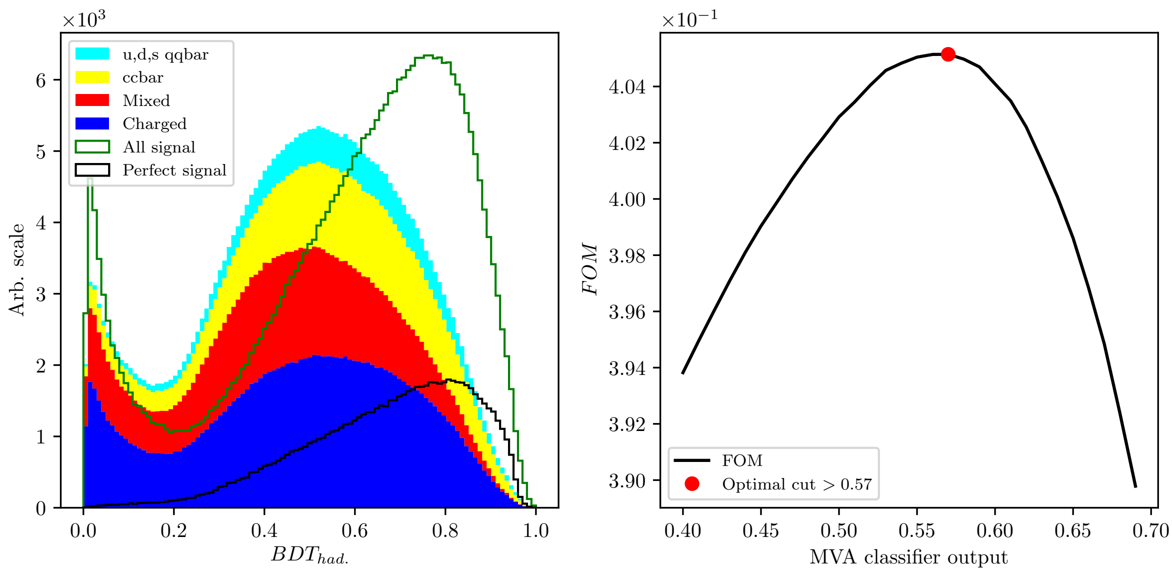


Figure 4.13: Hadronic MVA classifier output for major types of background and scaled-up signal (left), and $FOM$ optimization on this cut for correctly reconstructed signal candidates with hadronically decayed companion $B$ meson (right).

## 4.7 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 X. The 2D $FOM$ optimization of the optimal $M_{BC}$ and $\Delta E$ is shown in Figure 4.14. The signal region is defined as

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

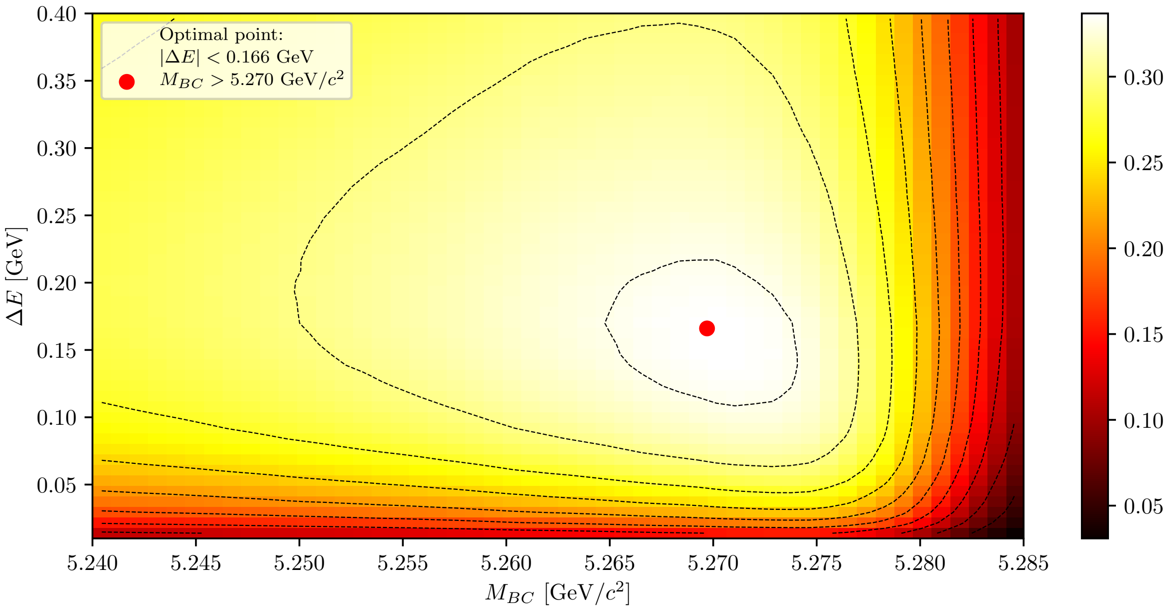


Figure 4.14: 2D $FOM$ optimization of the signal region definition, where the signal in the optimization was represented by perfectly reconstructed candidates.

Lastly, we can tighten the cut on $m^2_{miss}$, which we intentionally left loose before the signal categorization. With the $FOM$ optimization of perfectly reconstructed candidates inside the signal region, shown in Figure 4.15, the optimal cut on $m^2_{miss}$ is

- $|m^2_{miss}| < 1.1$ GeV$/c^2$.

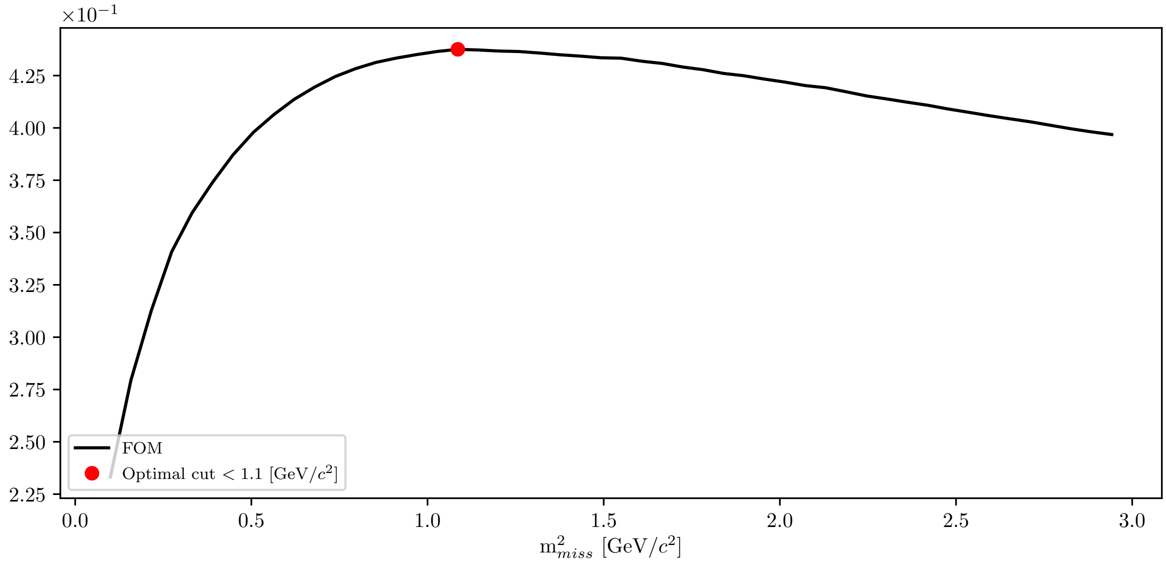


Figure 4.15: $FOM$ optimization of the optimal $m^2_{miss}$ cut in the signal region.

## 4.8 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: $P(\chi^2,\ DOF) > 1 \times 10^{-3}$, $|\cos\theta_{BY}| < 1$, $|m^2_{miss}| < 7$ 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: $q_{B_{sig}}\, q_{B_{comp}} = -1$, $BDT_{had.} > 0.57$, $|m^2_{miss}| <$ 1.1 GeV/$c^2$.

# 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) [] 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 X shows a schematic procedure of a $k$-fold cross-validation.

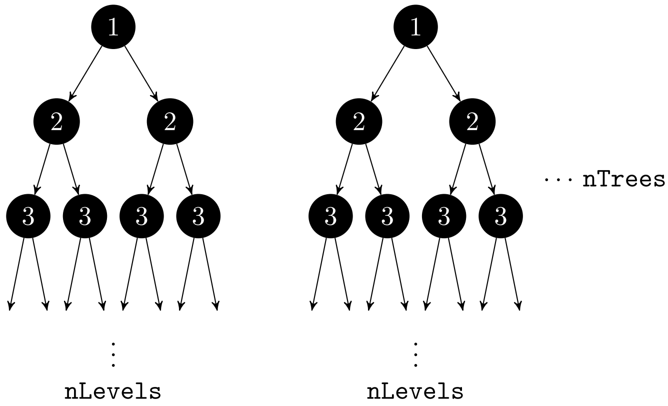


Figure 5.1: A schematic of a BDT forest with `nTrees`, each tree having a depth of `nLevels`.

PLOT

## 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 $\pi^0 \to \gamma\gamma$ 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 $\pi^0$ candidates. The output of this classifier is then applied to photons and represents a sort of a $\pi^0$ origin probability, which is used as an additional classifier variable in the next step of the clean-up.

### 5.2.1 $\pi^0$ MVA training

The dataset of $\pi^0$ candidates contains

- 387125 target candidates,
- 416019 background candidates,

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

The input variables used in this MVA are

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

The classifier output variable is shown in Figure 5.2.

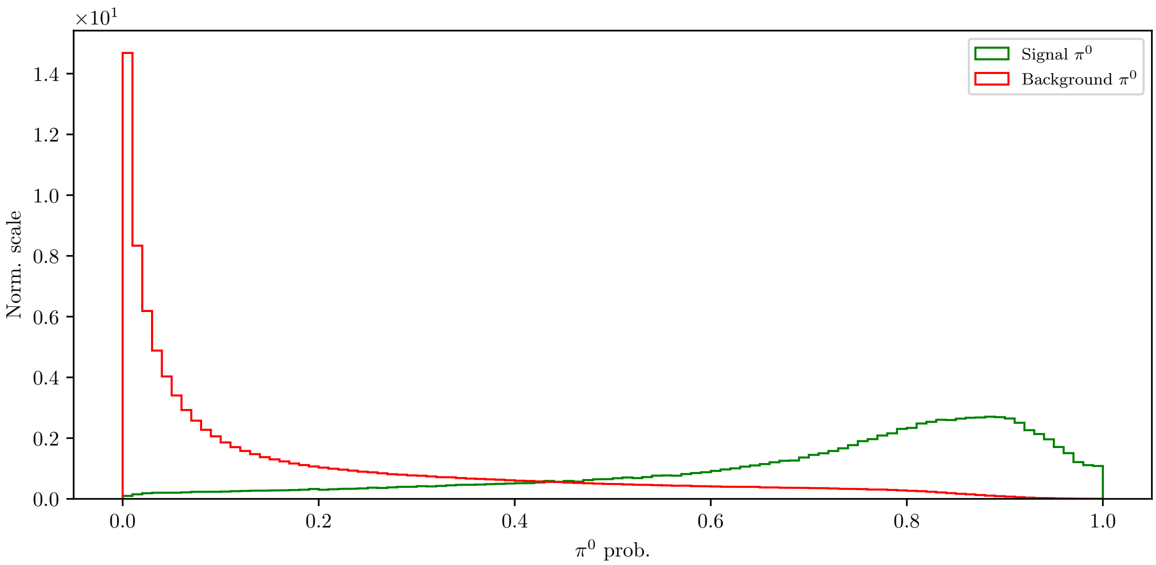


Figure 5.2: Classifier output of the $\pi^0$ training for signal and background $\pi^0$ candidates.

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

### 5.2.2 $\gamma$ MVA training

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

The dataset of $\gamma$ 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 $\gamma$ candidates from the existing Belle particle list.

The input variables used in this MVA are

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

The classifier output variable is shown in Figure 5.3.

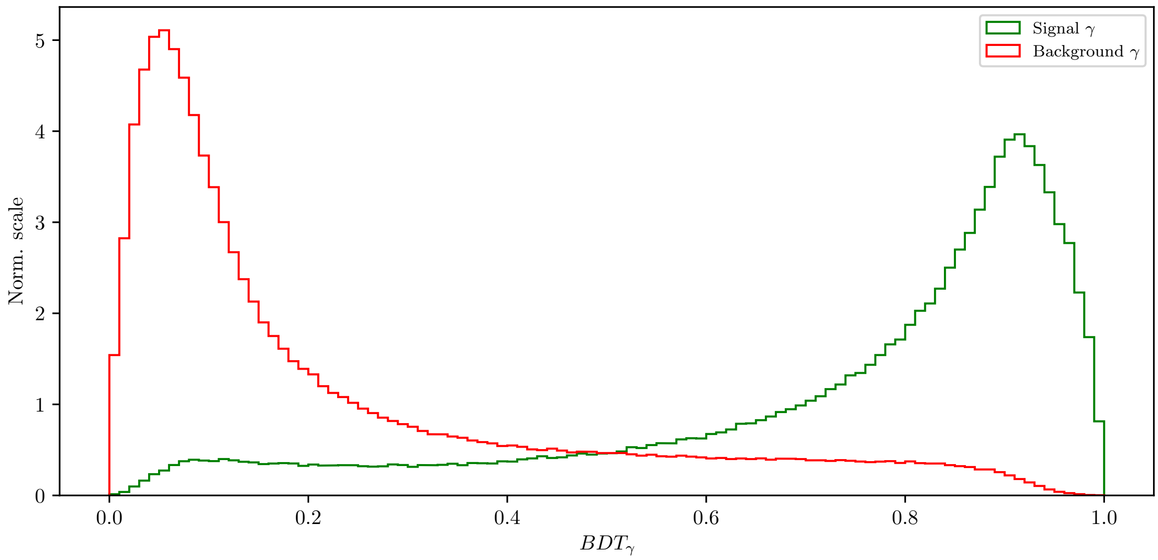


Figure 5.3: Classifier output of the $\gamma$ training for signal and background $\gamma$ candidates.

With the final weights for photon classification in hand, we apply them to the photon particle list. The cut optimization is shown in Figure 5.4 (left), with the optimal cut on the $\gamma$ 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: $1 - \epsilon_{BKG} = 81.7$ %.

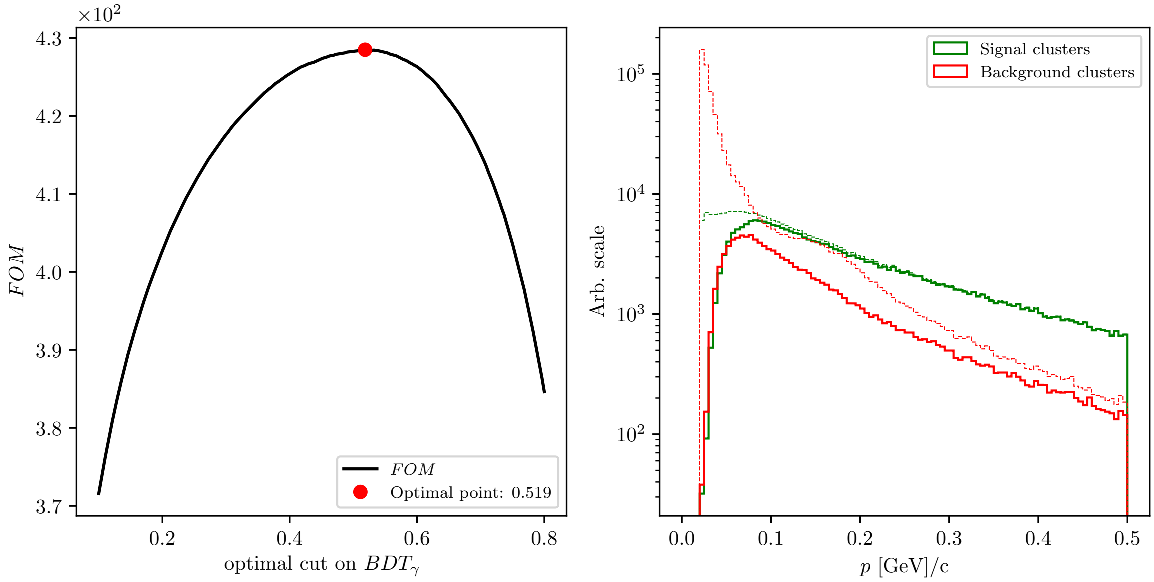


Figure 5.4: The $FOM$ of the classifier output optimization (left) and momentum magnitude in the LAB frame of signal and background photon candidates before and after the optimal cut (right).

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 $K_S^0 \to \pi^+\pi^-$,

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

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

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

Schematics of all the cases mentioned above are shown in Figure 5.5.

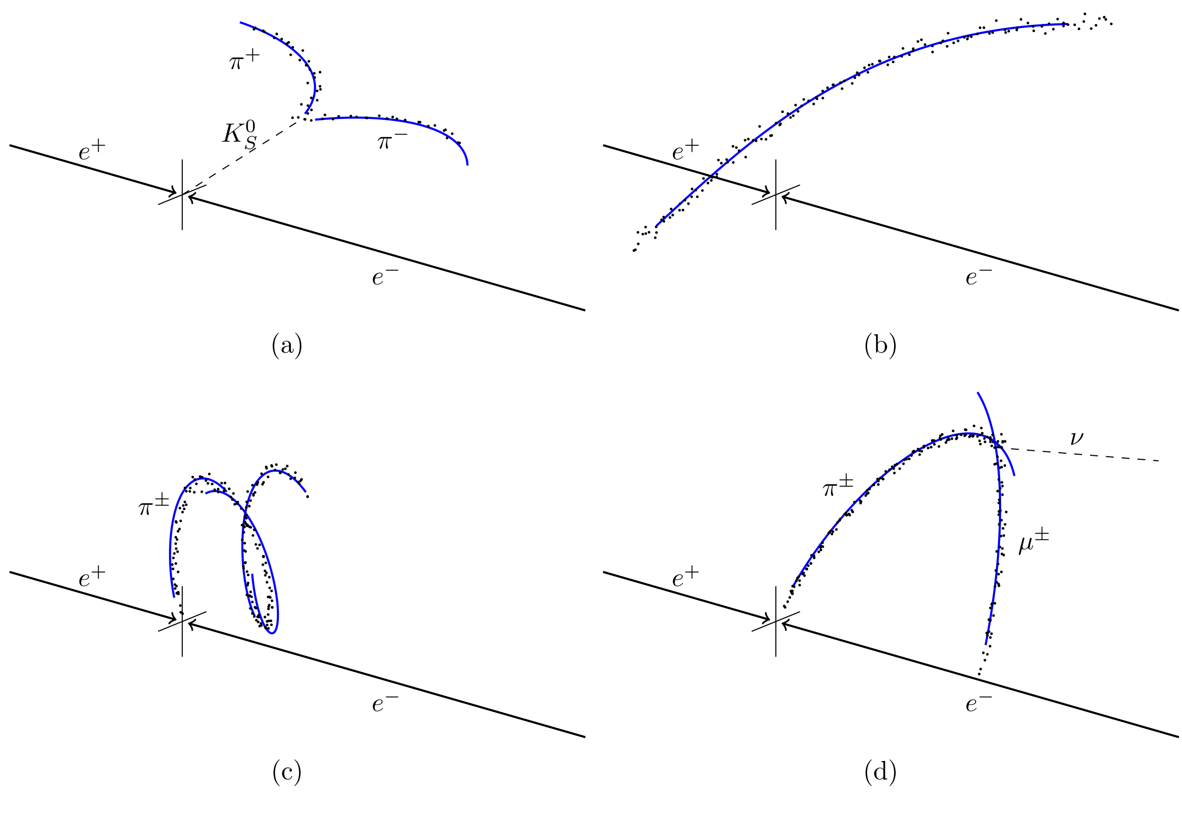


Figure 5.5: (a) Tracks from long-lived neutral particles, which decay away from the IP region, (b) Random tracks from background which are reconstructed, (c) Low-momentum particles which curl in the magnetic field, (d) in-flight decays of particles, which produce a kink in the trajectory.

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 $\pi^{\pm}$ particles, coming from the $K_S^0$ decays, have large impact parameters, which is usually a trait of background particles. In order to minimize confusion from the MVA point-of-view, these tracks are taken into account separately.

We use the converted $K_S^0$ candidates from the existing Belle particle list and use a pre-trained Neural Network classifier 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.

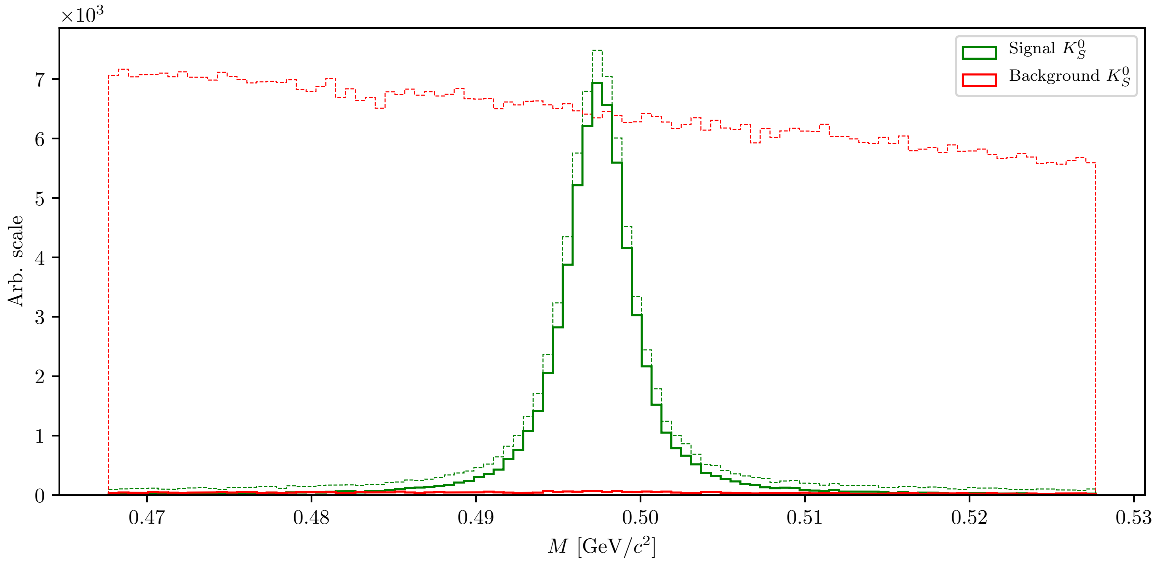


Figure 5.6: Invariant mass of the $K_S^0$ candidates before (dashed lines) and after (solid lines) the cut on the Neural Network classifier for signal (green) and background candidates (red). Signal peaks at nominal $K_S^0$ mass, while background covers a wider region.

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

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

### 5.3.2 Duplicate tracks

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

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

and proceed with the clean-up of track duplicates.

#### Defining a duplicate track pair

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

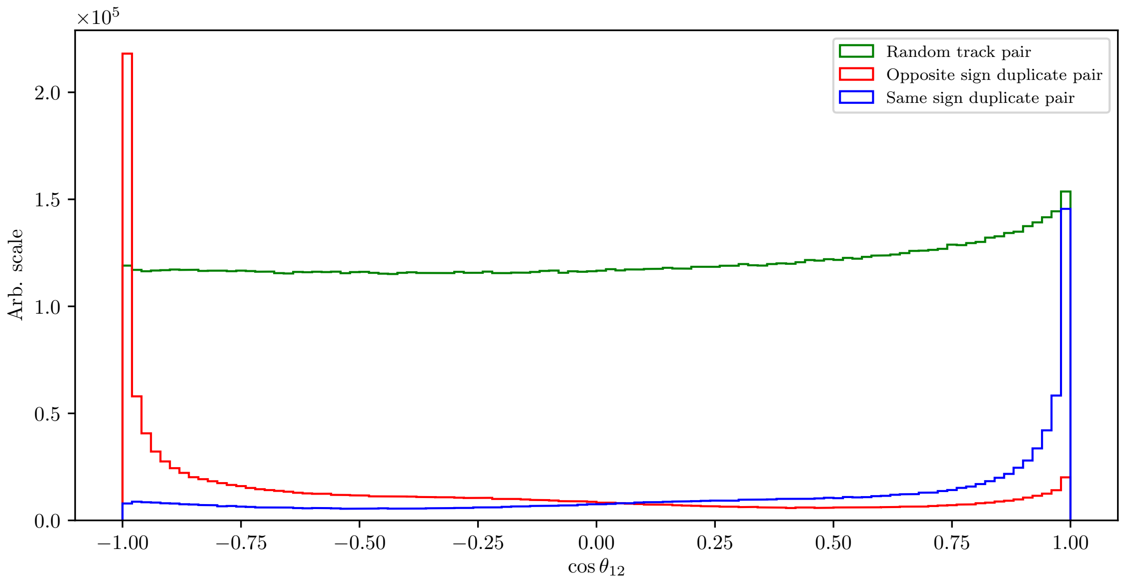


Figure 5.7: Distribution of the angle between two tracks in a single pair for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks (red).

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

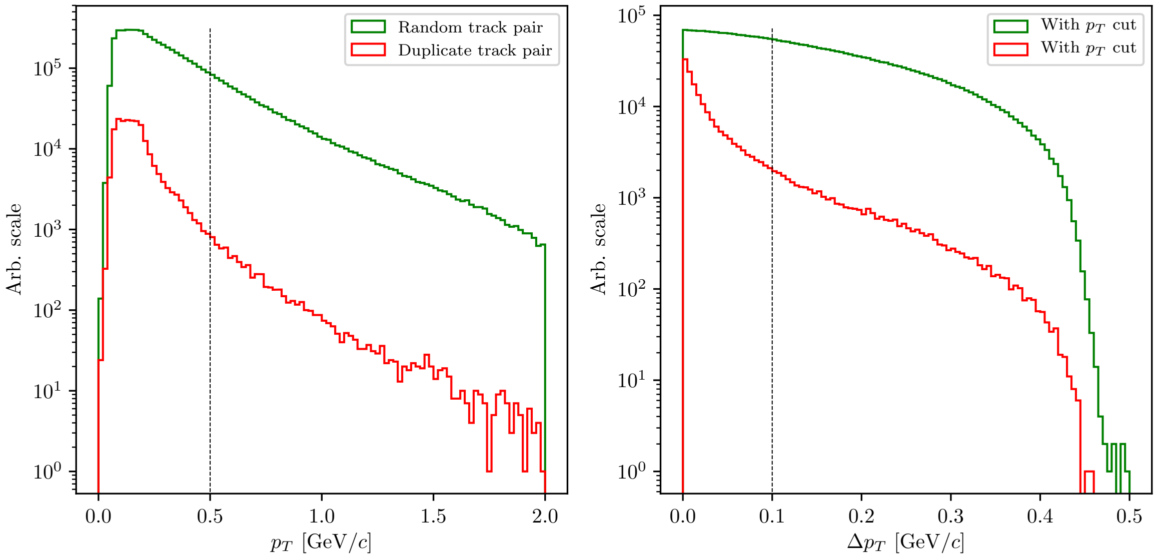


Figure 5.8: Distribution of transverse momentum $p_T$ (left) and transverse momentum difference $\Delta p_T$ (right) for all tracks coming from random (green) or duplicate track pairs (red). The plot on the right already includes the cut on $p_T$ from the plot on the left.

We impose a cut of

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

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

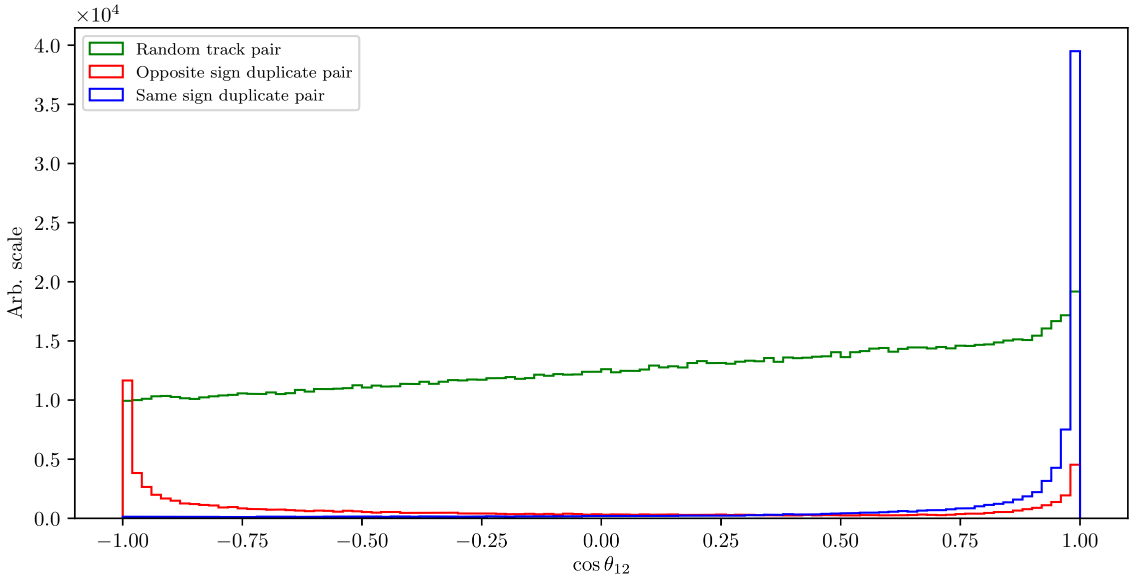


Figure 5.9: Distribution of the angle between two tracks in a single pair after applying the selection cuts defined in this subsection. The distributions are shown for random track pairs (green) and duplicate track pairs, where the latter were reconstructed as two same-sign (blue) or opposite-sign tracks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

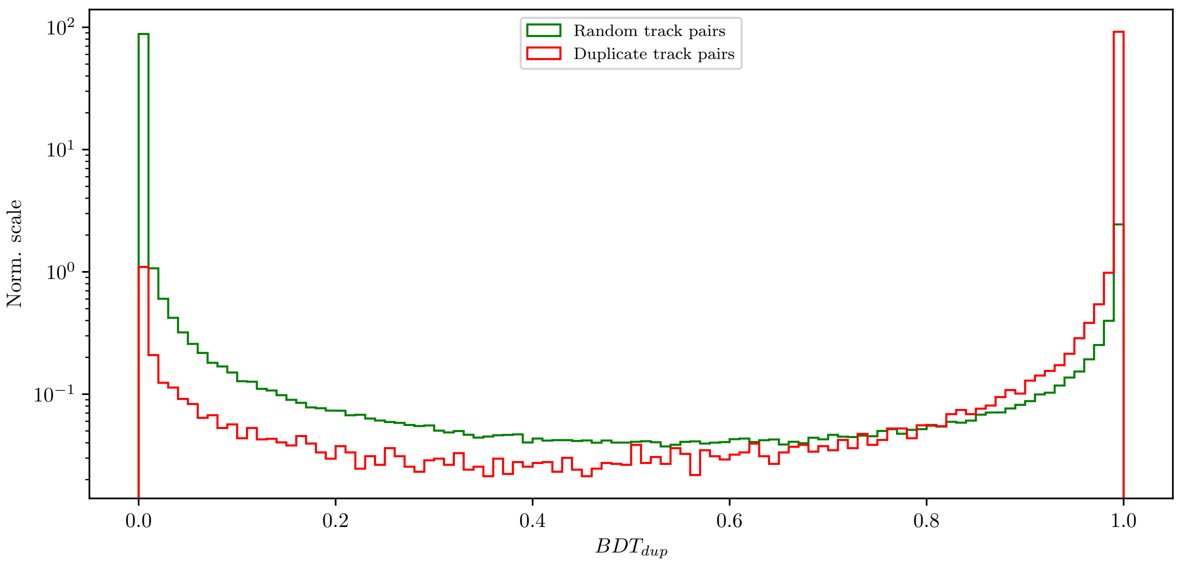


Figure 5.10: Classifier output of the track pair training for random track pairs and duplicate track pairs.

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

- $BDT_{duplicate} > 0.99915$.

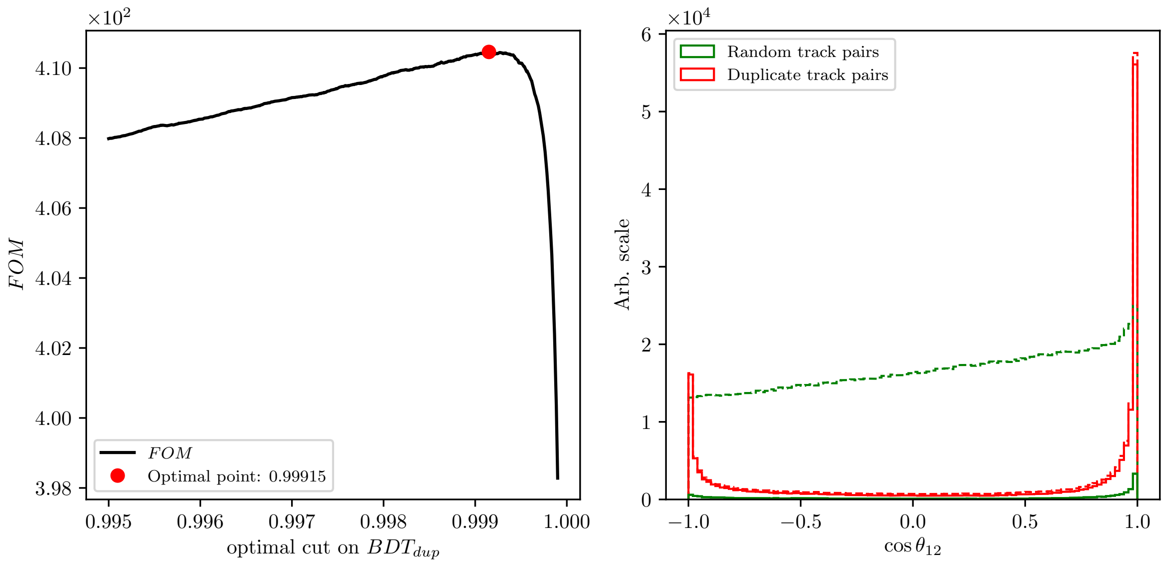


Figure 5.11: The optimization of the FOM function for the cut on classifier output (left) and distribution of the angle between two tracks in a single pair before (dashed) and after (solid) applying the optimal cut on the output classifier for random and duplicate track pairs (right).

### 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

$$\Delta f = f_{this} - f_{other}, \tag{5.1}$$

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),

Additionally we define an MC truth variable

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

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

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

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

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

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

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

### Training the duplicate track MVA

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

- 160146 target candidates,
- 128568 background candidates,

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

The input variables used in this MVA are

- theta angle of the track momentum,
- track quantities
  - impact parameters $d_0$ and $z_0$,
  - momentum components $p_T$ and $p_z$
  - number of hits in the SVD and CDC detectors
  - track fit $p$-value,
- pair-level information
  - $\Delta d_0$, $\Delta z_0$, $\Delta N_{CDC}$, $\Delta N_{SVD}$, $\Delta p_T$, $\Delta p_z$.

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

$$\Delta BDT_{final} = BDT_{final}^{this} - BDT_{final}^{other}, \tag{5.3}$$

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

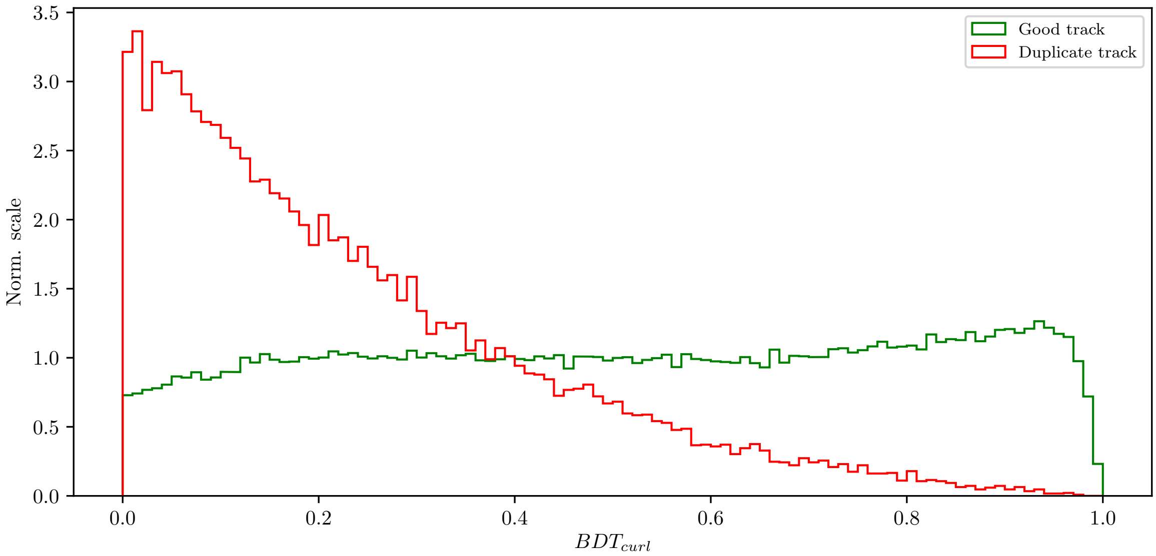


Figure 5.12: Classifier output of the MVA training for curling track recognition.

Finally, we select all duplicate tracks which survive the cut

$$\Delta BDT_{final} > 0 \tag{5.4}$$

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

TABLE

## 5.4 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 $\Delta E$ and $M_{BC}$, which we are most interested in. Since the reconstruction procedure also applies cuts on these cleaned-up variables, clean-up procedure consequentially also increases the efficiency 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 case where no clean-up was applied at all. Figure X shows distributions of $\Delta E$ and $M_{BC}$ for various clean-up setups. We see a large increase in efficiency and resolution in both observed variables for the case where the ROE clean-up is performed. Table X shows relative efficiencies and resolutions of $\Delta E$ and $M_{BC}$ based on the "perfect" ROE clean-up, which is the same as our optimal clean-up, with the exception of the masses of FSP particles in Eq. (4.9), which are, in this case, taken from MC. Mass hypothesis determination of charged FSP particles in Eq. (4.9) based on PID information seems to be quite successful, since the benefits of using true masses from Monte Carlo are very small.

PLOT

TABLE

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. (4.18), shown in Figure X (left) for various clean-up procedures. Figure X (right) shows the same distribution for the optimal ROE clean-up, but split into two groups. The first group are all the candidates which are present in the sample if no clean-up is applied, and the second group are all the new entries which enter our sample due to increased clean-up efficiency. This enables us to see the effects of the clean-up on already existing entries. As a cross-check, we can also look at $\Delta E$ and $M_{BC}$ variables for each value of the charge product. These plots are shown in Figure X and they show a clear increase in efficiency and resolution for the correct value of the charge product, with a slight efficiency increase also for other values.

PLOT

## 5.5 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.

# 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 $b \to u$ analyses in order to suppress $b \to c$ backgrounds, such as $K$-veto, are not possible. As a consequence, our final sample consists of combinations of $K$ pairs coming from $b \to c$ sources, such as $D^0 \to K^+K^-$. Such candidates usually have resonance-like properties in the two-kaon invariant mass spectrum. Figure 6.1 shows this invariant mass spectrum of two kaons, $m_{KK}$, where obvious resonant structures are present from sources like

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

An opportunity for a control decay choice presents itself as

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

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

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

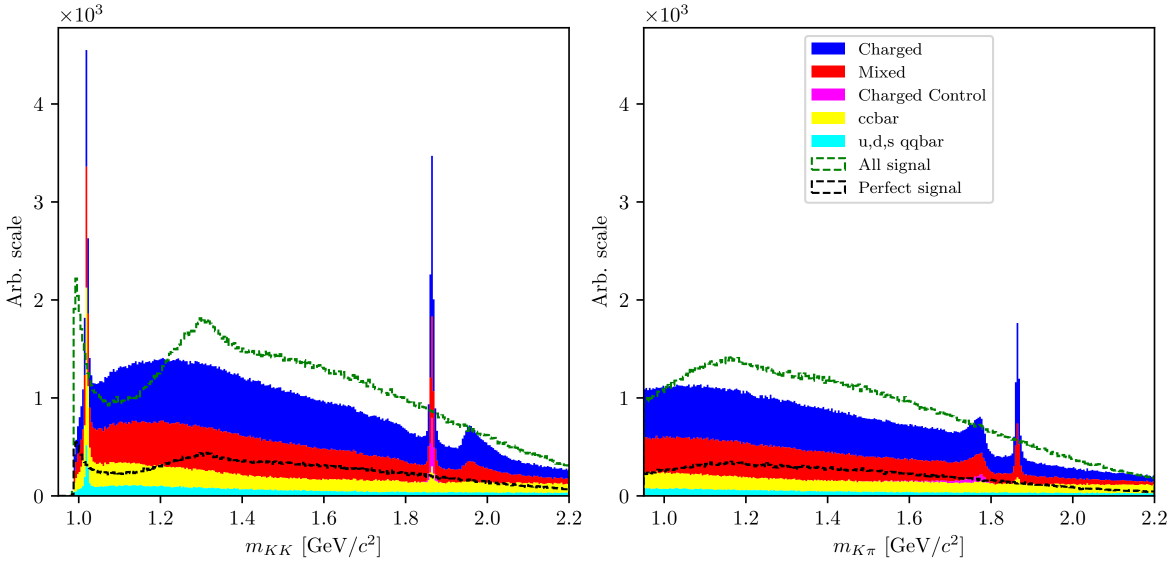


Figure 6.1: Invariant mass of two correctly reconstructed kaons (left) and invariant mass of two kaons, where one was miss-identified as a pion (right).

## 6.2 Continuum suppression

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

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

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

### 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.

### $B$ meson direction

Two $B$ mesons, coming from a spin-1 $\Upsilon(4S)$ meson, both have 0 spin, which results in a $\sin^2\theta_B$ angular distribution of the $B$ meson direction with respect to the beam axis. On the other hand, $q\bar{q}$ final states are represented by two half-spin fermions, which results in two jets, following a $1+\cos^2\theta_B$ distribution. The variable $|\cos\theta_B|$ allows one to discriminate between $B$ candidates from $B\bar{B}$ decays and continuum background. Figure X shows the distribution of $|\cos\theta_B|$ for different $B$ meson candidates.

### Thrust and related variables

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

$$T = \frac{\sum_i |T \cdot p_i|}{\sum_i |p_i|}. \tag{6.1}$$

In this case, a related variable is $|\cos\theta_T|$, where $\theta_T$ is the angle between the thrust axis of the momenta from $B$ meson decay particles, and the thrust axis of all particles in the ROE. Since both $B$ mesons in $B\bar{B}$ events are produced at rest, their decay particles, and consequentially their thrust axes, are uniformly distributed in the range [0, 1]. On the other hand, decay particles from continuum events follow the direction of the jets in the event. As a consequence, the thrusts of both the $B$ meson and the ROE are strongly directional and collimated, which results in a large peak at $|\cos\theta_T| \approx 1$. Additionally, one can also use the variable $|\cos\theta_{TB}|$, which is the thrust axis between the $B$ candidate and the beam axis. For $B$ candidates from $B\bar{B}$, this distributions is again uniformly distributed, while for candidates from continuum events this distribution follows the distribution of the jets with the function $1+\cos^2\theta_{T,B}$. Figure X shows the 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 [] is 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 $10^\circ$ around the thrust axis, as illustrated in Figure X.

### KSFW moments

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

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

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

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

## 6.2.2 MVA training

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

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

The training variable set consisted of

- $B$ meson direction and thrust related variables
  - Magnitude of thrust axes of $B$ and $ROE$
  - Cosine of angle between thrust axis of $B$ and thrust axis of ROE

- Cosine of angle between thrust axis of $B$ and beam direction
- Reduced Fox-Wolfram moment $R_2$

- All 9 CLEO Cones
- KSFW Moments
  - $R^{so}_{01}$, $R^{so}_{02}$, $R^{so}_{03}$, $R^{so}_{04}$,
  - $R^{so}_{10}$, $R^{so}_{12}$, $R^{so}_{14}$,
  - $R^{so}_{20}$, $R^{so}_{22}$, $R^{so}_{24}$,
  - $R^{oo}_{0}$, $R^{oo}_{1}$, $R^{oo}_{2}$, $R^{oo}_{3}$, $R^{oo}_{4}$,
- FlavorTagging variables
  - $qp$ of $e$, $\mu$, $\ell$,
  - $qp$ of intermediate $e$, $\mu$, $\ell$,
  - $qp$ of $K$, $K/\pi$, slow pion, fast hadron
  - $qp$ of maximum $P^*$, $\Lambda$, fast-slow-correlated (FSC)
- Other
  - $\Delta z$, $\Delta t$

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

PLOT

## 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

- $b \to c\ell\nu$ background,
- $b \to u\ell\nu$ background,

- Other rare decays (radiative, penguin, rare 2- and 3-body decays, . . . ).

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

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

The training variable set consisted of

- fit probability of $P(\chi^2, DOF)$ of the signal candidate and the ROE side, separately,
- $\cos\theta_{BY}$ from Eq. (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,
- $\theta$ angle of the ROE momentum in CMS frame,
- $\xi_Z$ from []
- $\Delta z$,
- $m^2_{miss}$ from Eq. (4.11),
- $B \to D^*\ell\nu$ veto variables,

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

PLOT

When the training is finished and hyper-parameters of the classifier are optimized, the classifier output, shown in Figure X, 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.

PLOT

### 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 $\Delta E$ and $M_{BC}$), from the training set. This, however, decreases the classifiers discriminating power. Another approach is to use a novel boosting method, uBoost, which is trained to optimize an integrated FOM under the constraint that the BDT selection efficiency for the desired class must be uniform. The uBoost algorithm balances the biases to produce the optimal uniform selection [].

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 uBoost classifier with a uniformity constraint in the $M_{BC}$ variable for background candidates. The resulting classifier output is shown in Figure X. The range of the classifier is more restrained with respect to the standard classifier, and separation between signal and background seems worse, however, the final conclusion will be shown after the signal extraction.

PLOT

## 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 X.

PLOT

We can compare signal and major background distributions of $\Delta E$ and $M_{BC}$ after the 2D $FOM$ optimization for both classifiers. Figure X shows major components in the background in arbitrary (left) and normalized scale (right) for $\Delta E$ (top) and $M_{BC}$ (bottom) for the final sample optimized with the standard BDT classifier, while Figure X shows similarly for the final sample optimized with uBoost classifier. We can see that there is considerably more background in the latter case, however, also shapes ob background and signal distributions differ greatly. The biggest change seems to be in the shape of the $M_{BC}$ distribution for the background components, where the background 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: $N_{sig} = 152.82, \quad N_{sig}/N_{bkg} = 3.70$ %,
- uBoost: $N_{sig} = 241.56, \quad N_{sig}/N_{bkg} = 1.09$ %.

PLOT

PLOT

## 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 sample 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 X shows $\Delta E$, $M_{BC}$ and $BDT_{q\bar{q}}$ for off-resonance data and MC, before any MVA cuts, where the MC sample was scaled down by a factor of 6, due having 6 streams of MC. Figures show a consistent overestimation of MC, which indicated that the $\Delta E$, $M_{BC}$ templates would in principle need to be fixed for the signal extraction with the template fit. However, looking at the ratios of MC and data for these variables before and after the $q\bar{q}$ MVA cut, shown in Figure X, one can conclude that, based on the flatness of the ratio function, the only difference is the normalization factor, so the distribution shapes can be left as they are, since the normalization will be automatically set in the template fit.

PLOT

PLOT

We can repeat the check on on-resonance data. Figure X shows the $q\bar{q}$ classifier output, $BDT_{q\bar{q}}$, 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. One can then make the same conclusion about overestimation of the $q\bar{q}$ MC component, as was seen on the off-resonance sample.

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 the control decay.

# 7 Signal extraction

## 7.1 Fit templates

## 7.2 Adaptive binning algorithm

## 7.3 Signal MC fit results

## 7.4 Control fit results

## 7.5 Signal fit to data

# 8 Systematics

## 8.1 Model uncertainty effects

## 8.2 PID efficiency correction

## 8.3 Bias

# Addendum A: ROE clean-up $\pi^0$ training

## Correlations

## Hyper-parameter optimization

## Importance

## Validation

# Addendum B: ROE clean-up $\gamma$ training

## Correlations

## Hyper-parameter optimization

## Importance

## Validation

# Addendum C: ROE clean-up duplicate pair training

## Correlations

## Hyper-parameter optimization

## Importance

## Validation

# Addendum D: ROE clean-up duplicate training

## Correlations

## Hyper-parameter optimization

## Importance

## Validation

# Addendum E: $q\bar{q}$ suppression training

## Correlations

## Hyper-parameter optimization

## Importance

## Validation

# Addendum F: $B\bar{B}$ suppression training

## Correlations

## Hyper-parameter optimization

## Importance

## Validation