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

Matic Lubej

Ljubljana, 2018

# Changelog

# Contents

# 1 Introduction

# 2 Data and MC

## 2.1 MC

### 2.1.1 Used for precuts

### 2.1.2 Used for MVA training

# 3 Event reconstruction

In this chapter the procedure for event reconstruction of the decay Decay is shown, starting from final state particle selection and combining particles up the chain all the way to the $B$ meson.

## 3.1 Final state particles selection

Since the neutrino escapes detection, we can only reconstruct the charged tracks of our 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 to reduce computation time and file sizes.

### Leptons

Figures 3.1 and 3.2 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.
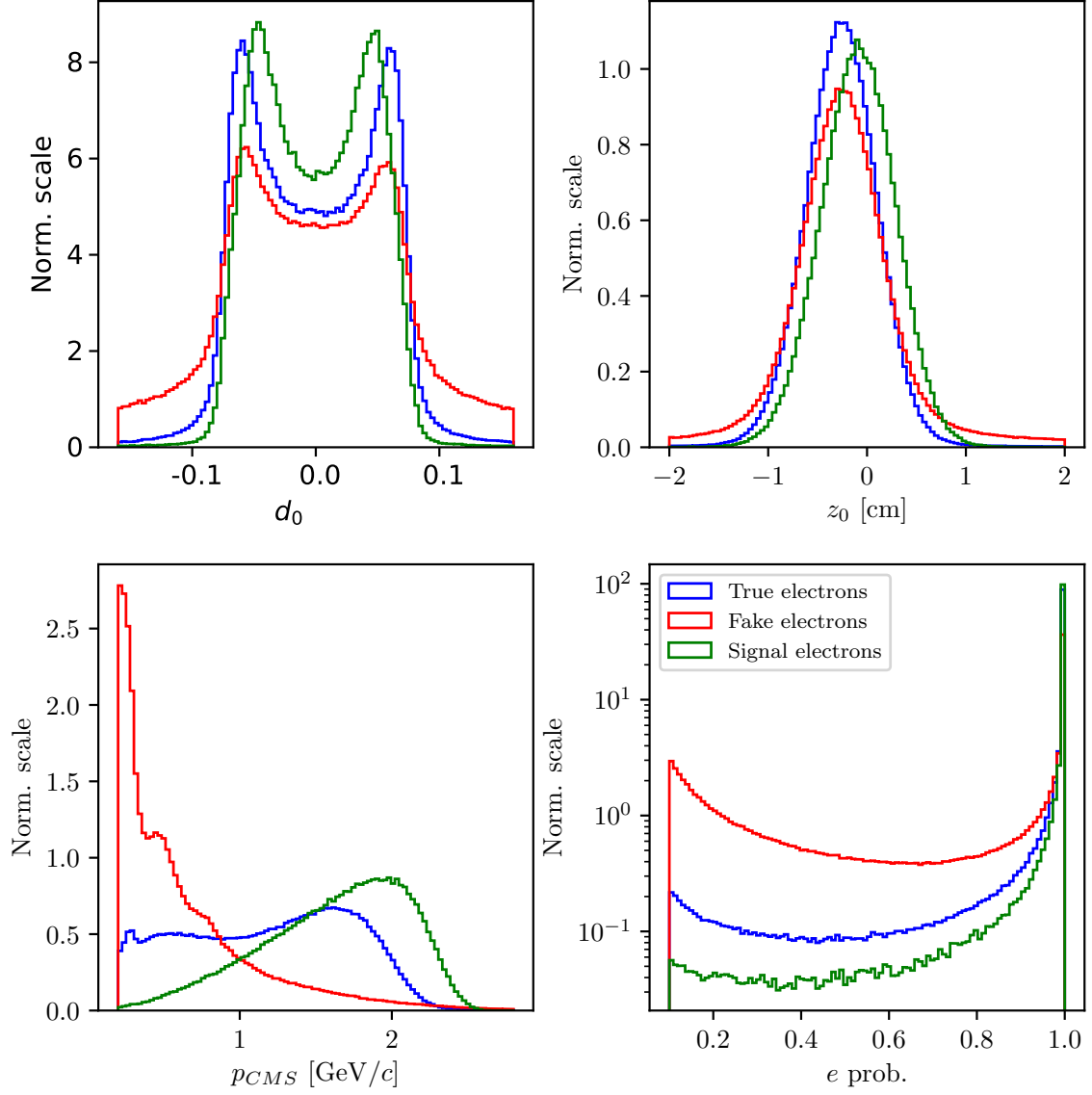
Figure 3.1: Normalized properties of true (blue), fake (red) and true electrons (green) from signal decay
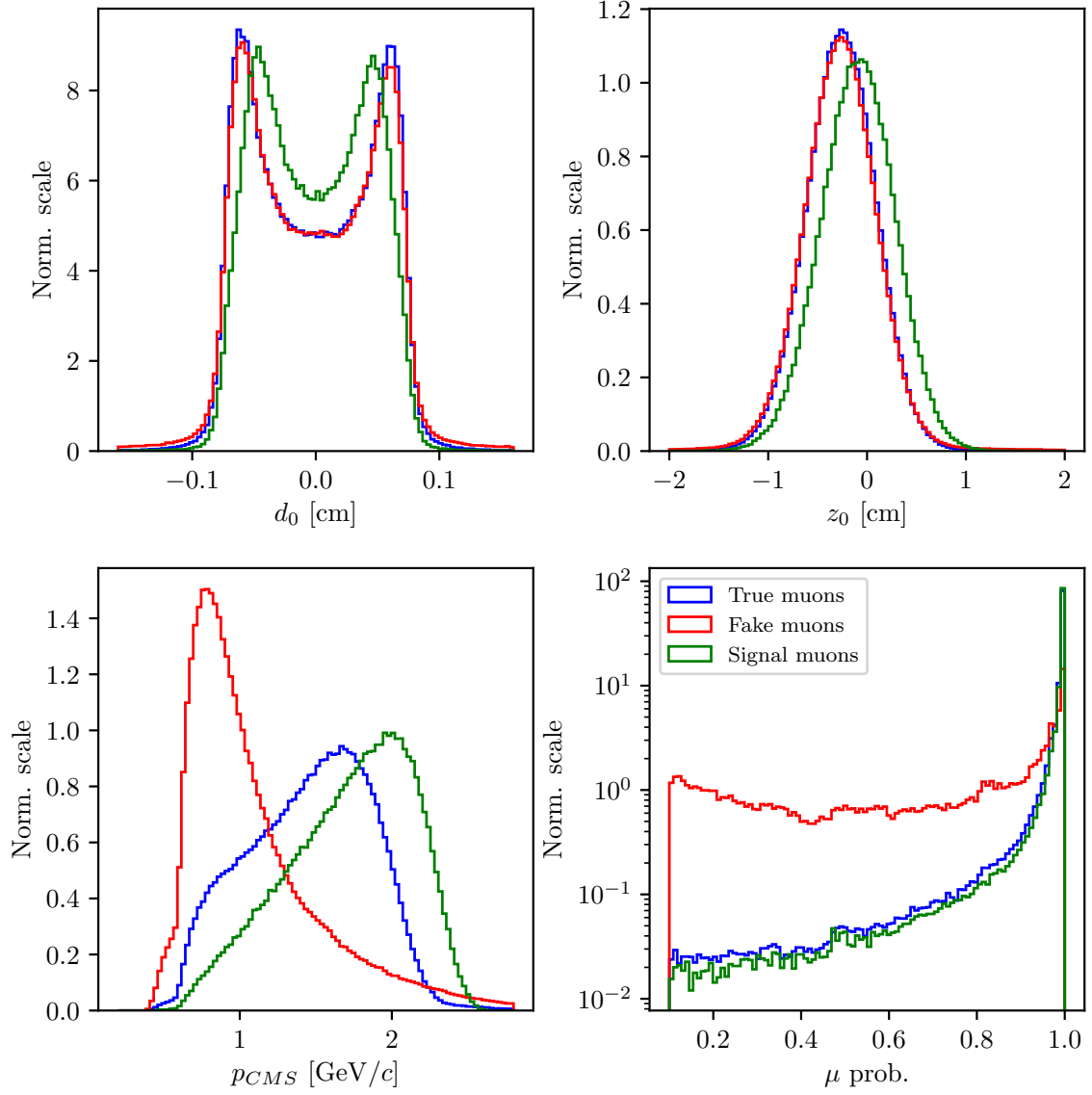
.

Figure 3.2: Normalized properties of true (blue), fake (red) and true muons (green) from signal decay
.

Based on the first plots, we can define a set of cuts

- $|d_0| < 0.1$ cm,

- $|z_0| < 1.5$ cm,

- $p_{LAB} > 0.6$ GeV/$c$ and $p_{CMS} \in [0.4, 2.6]$ GeV/$c$ for electrons,

6

- $p_{CMS} \in [0.6, 2.6]$  GeV/$c$ for muons,

where the $p_{LAB}$ momentum cut for the electron case is chosen to discard a region with a sharp jump, which is assumed to come from sources like hard coded values in Belle software.

With this selection we can now determine the optimal PID cuts for electrons and muons, where we optimize with the standard definition of *figure of merit* (FOM), shown in Figures 3.3 and 3.4. The cuts are optimized in steps of 0.1, since we will later need to apply the PID correction factors, which are only available for cuts of these values.

$$FOM = \frac{S}{\sqrt{S+B}}, \tag{3.1}$$

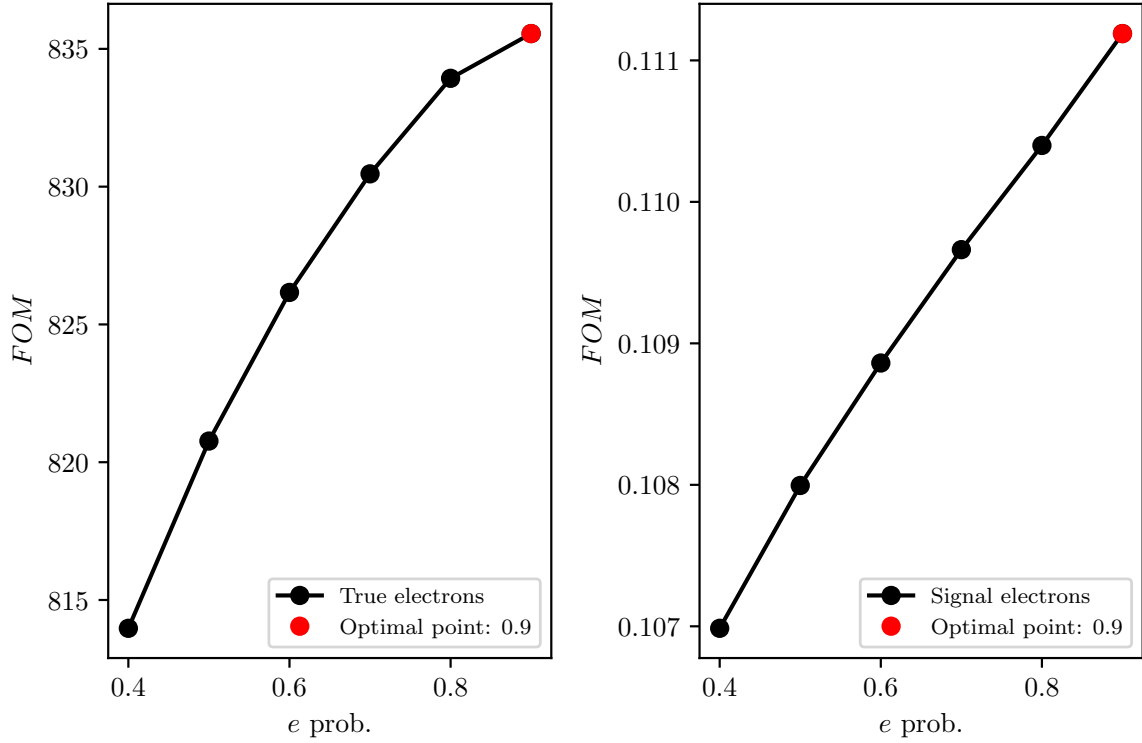where $S$ represents number of signal and $B$ the number of background candidates.



Figure 3.3: *FOM* optimizations of the PID probability cuts for true electrons (left) and true electrons from signal decays (right).

Figure 3.4: $FOM$ optimizations of the PID probability cuts for true muons (left) and true muons from signal decays (right).

The optimized PID cuts for leptons are

- $e$ prob. $> 0.9$ for electrons,

- $\mu$ prob. $> 0.9$ for muons.

The optimal value for the PID cuts is the same for all true leptons, if they are coming from signal decays or not, so this cut can be widely used on leptons coming from either $B$ meson in our analysis.

**Kaons**

For the case of kaons, the procedure is the same and the cuts are very similar. Figure 3.5 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.
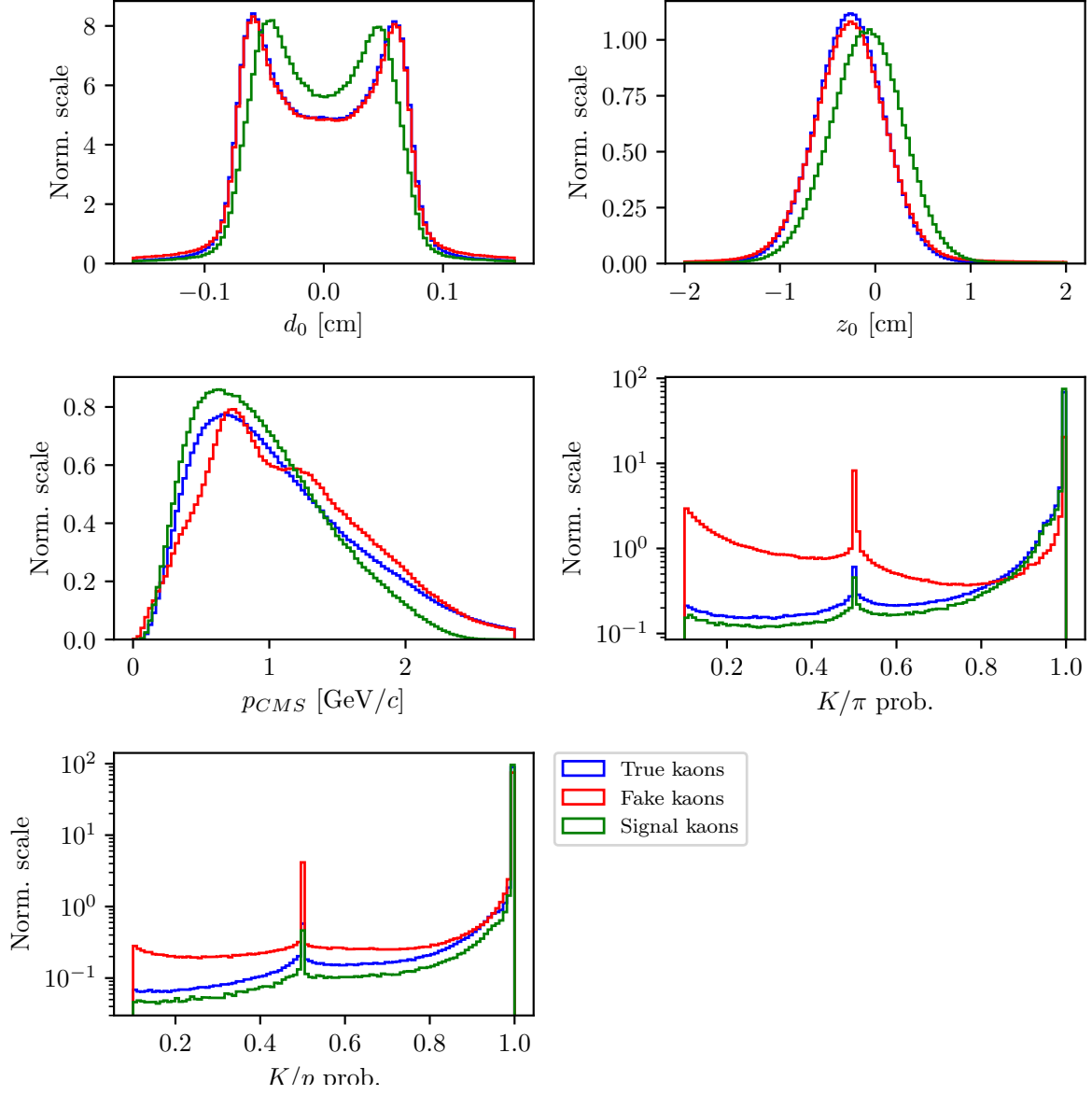
Figure 3.5: Normalized properties of true (blue), fake (red) and true kaons (green) from signal decay

.

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 3.6 shows the optimization procedure for PID cuts on kaon candidates.



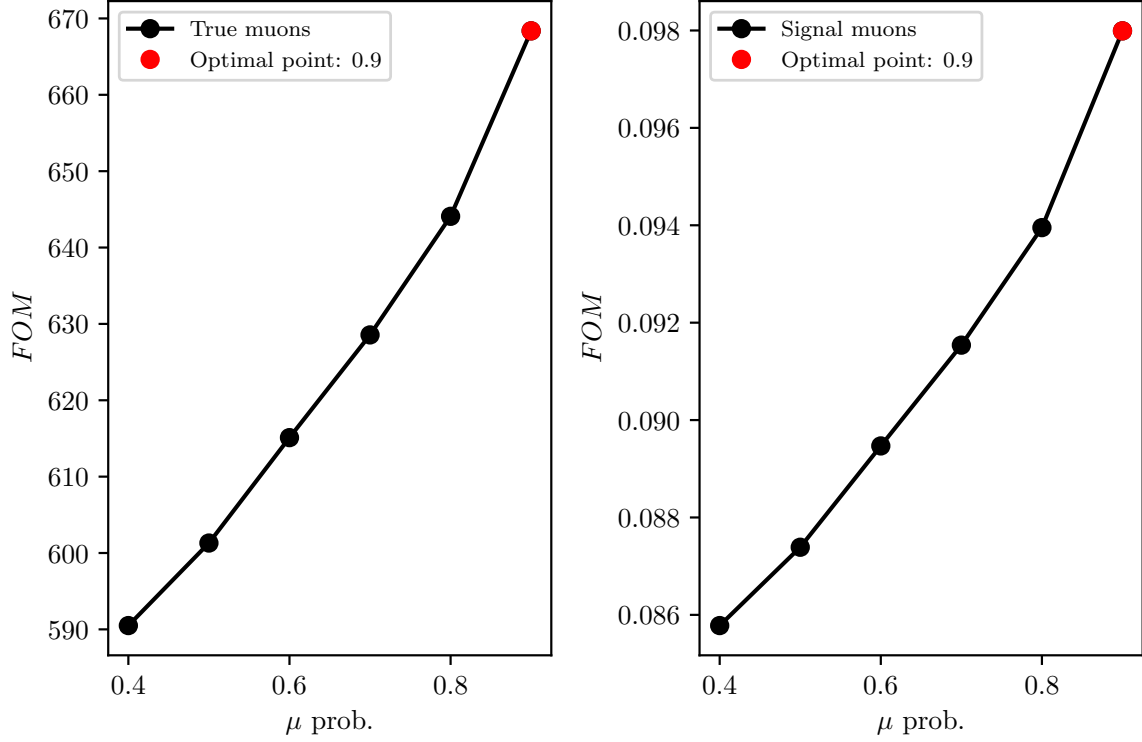Figure 3.6: $FOM$ optimizations of the PID probability cuts for true kaons (top) and true kaons from signal decays (bottom). The plots on the left show the optimization of the first step for the $K/\pi$ probability cut, and the plot on the right the $K/p$ probability cut.

The optimized PID cuts for kaons are

- $K/\pi > 0.6$,

- $K/p > 0.1$,

where we chose a cut of 0.6 even for the signal kaons, because there is a small difference and since we can always make a tighter cut later in the analysis.


## 3.2 Combination of FSP particles

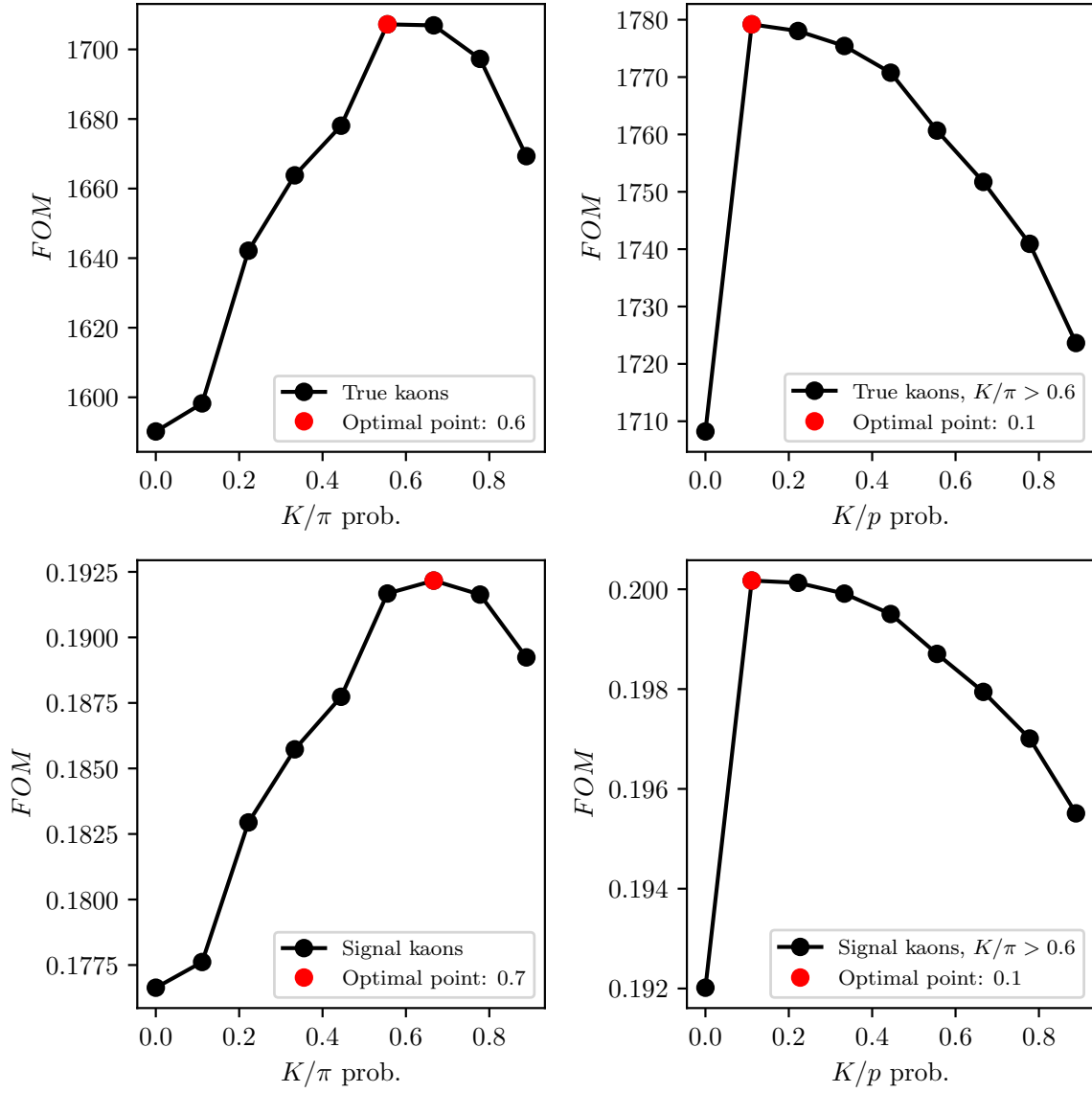With the pre-selected kaon and lepton candidates we make appropriate combinations for potential $B^+$ meson candidates. Due to our inability to reconstruct the neutrino, we are only able, at this point, to reconstruct the $B$ mesons in the following two channels

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

and similarly for $B^-$. In order to further guarantee the proper combination of FSP particles, we perform a vertex fit of the three tracks. $B$ mesons have a relatively long lifetime, so they travel and decay along the z-axis of the detector in the direction of the boost, so we perform the vertex fit 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 fit probability. The fit probability for signal and background $B$ meson candidates is shown in Figure 3.7 (left). We perform a $FOM$ cut optimization of this variable, which is shown in Figure 3.7 (right).
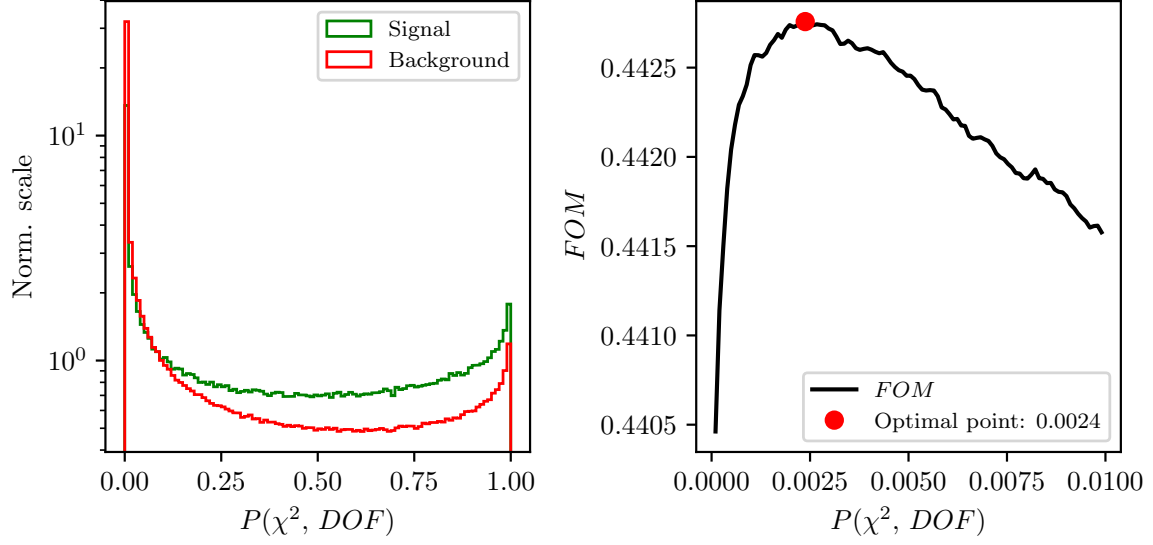
Figure 3.7: Normalized vertex fit probability distribution for signal and background $B$ meson candidates in logarithmic scale (left) and $FOM$ optimization of the vertex fit probability (right).

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

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

since it's on the lower side.

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

$$\mathrm{p}_\nu = \mathrm{p}_B - \mathrm{p}_Y, \tag{3.2}$$

$$\mathrm{p}_\nu^2 = m_\nu^2 = m_B^2 + m_Y^2 - 2E_B E_Y + 2\vec{p}_B \cdot \vec{p}_Y \approx 0, \tag{3.3}$$

$$\cos\left(\theta_{BY}\right) = \frac{2E_B E_Y - m_B^2 - m_Y^2}{2|\vec{p}_B||\vec{p}_Y|}, \tag{3.4}$$

where all the energy and momenta above are calculated in the CMS frame. The mass of the neutrino equals to 0 to a very good precision, so we used it in Eq. (3.3). In addition, we can substitute the unknown energy and momentum magnitude, $E_B$ and $|\vec{p}_B|$, of the $B$ meson in Eq. (3.4), with quantities from the well known initial conditions

$$E_B = E_{CMS}/2, \tag{3.5}$$

$$|\vec{p}_B| = p_B = \sqrt{E_B^2 - m_B^2}, \tag{3.6}$$

12

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 3.8 (left). We impose an optimized cut on this variable from Figure 3.8 (right) to discard a large amount of background.

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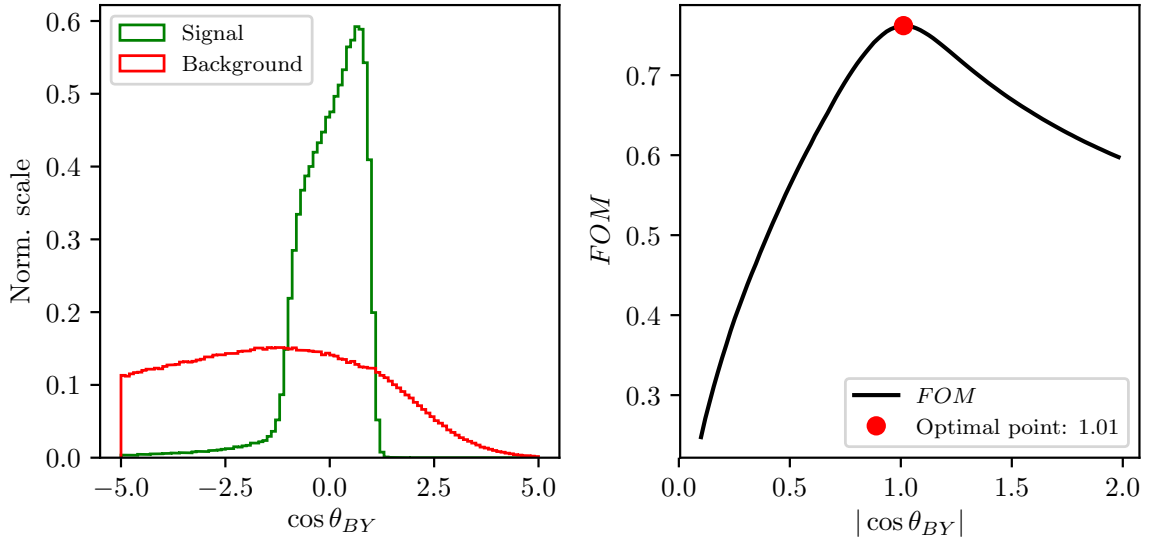


Figure 3.8: Normalized $\cos\theta_{BY}$ distribution for signal and background $B$ meson candidates (left) and $FOM$ optimization of the $\cos\theta_{BY}$ variable (right).

## 3.3 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 neutrino. 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 is used to narrate 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 4. From this point on in this chapter the reader can assume the ROE to be clean.

## 3.4 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 other $B$ 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. If we did not depend on the other $B$ meson for our signal candidate reconstruction, this would not have been an issue.

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

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \sum_{i}^{\mathrm{Event}} (E_i, \vec{p}_i), \tag{3.7}$$

$$\mathrm{p}_{miss} = \mathrm{p}_{\Upsilon(4S)} - \left( \mathrm{p}_Y - \sum_{i}^{\mathrm{Rest\ of\ event}} (E_i, \vec{p}_i) \right). \tag{3.8}$$

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

$$\mathrm{p}_i^{neutral} = (p_i, \vec{p}_i) \quad \text{and} \quad \mathrm{p}_i^{charged} = \left( \sqrt{m_i^2 + p_i^2}, \vec{p}_i \right). \tag{3.9}$$

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. $> 0.9$,

3. otherwise $K$, if $K/\pi$ prob. $> 0.6$,

4. otherwise $\pi$.

We define the square of the missing mass, $m^2_{miss}$, which is zero, if signal side neutrino is the only missing particle in the event, as shown in Equation (3.11).

$$p_\nu = p_{miss} = (E_{miss}, \vec{p}_{miss}),  \tag{3.10}$$
$$m^2_{miss} = p^2_{miss} = p^2_\nu = m^2_\nu \approx 0.  \tag{3.11}$$

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 additional missing particles such as extra missing neutrinos or other neutral undetected particles such as $K^0_L$, or simply missing tracks due to detection failure. Figure X shows the distribution of $m^2_{miss}$ as defined with the missing four-momentum in Eq. (3.10). 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, as shown in Figure 3.9. Due to this fact, we impose a cut on the $m^2_{miss}$ variable in order to discard all candidates with spoiled properties, even if it was in principle a correct combination of FSP particles on the signal side, in order to cut away events with a higher probability of a semi-leptonic decay of the other $B$ meson

- $|m^2_{miss}| < 7$ GeV/$c^2$.

This cut was not optimized, since the optimal case is considered to be too strong at this point in the analysis, since we still may want to attain 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 other $B$ meson decayed hadronically and all of it's particles were taken into account correctly[1]. We denote this subset as *perfect* signal and the rest as *non-perfect* signal.

---

[1]We only allow for missing photons, since all photons are much harder to correctly take into account.
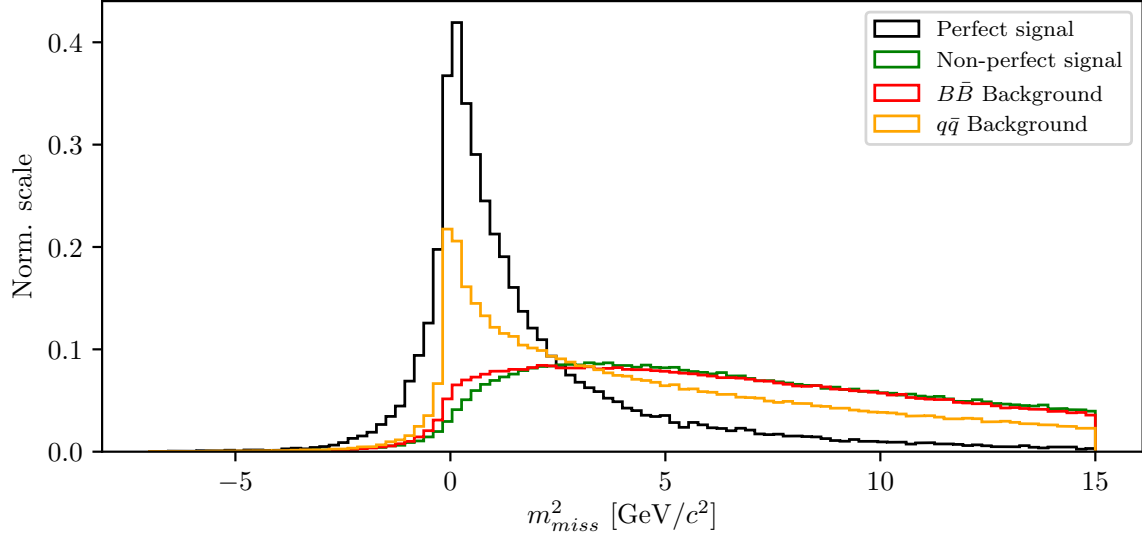
Figure 3.9: Normalized $\cos\theta_{BY}$ distribution for signal and background
$B$ meson candidates (left) and $FOM$ optimization of the
$\cos\theta_{BY}$ variable (right).

The main uncertainty in neutrino four-momentum, defined as in Eq. (3.10) 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

$$p_\nu = \left(|\vec{p}_{miss}|,\, \vec{p}_{miss}\right). \tag{3.12}$$

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

$$\Delta E = E_B - E_{CMS}/2, \tag{3.13}$$

$$M_{BC} = \sqrt{\left(E_{CMS}/2\right)^2 - |\vec{p}|^2}. \tag{3.14}$$

Since the final fit will be performed on these two variables, we define two regions

- Fit region: $M_{BC} \in [5.1,\, 5.3]$ GeV/$c^2$ and $\Delta E \in [-1.3,\, 1,\, 5]$ GeV,

- Signal enhanced region: $M_{BC} \in [5.27,\, 5.3]$ GeV/$c^2$ and $|\Delta E| < 0.143$ GeV,

where the signal enhanced region was defined via a 2D $FOM$ optimization on the per-
fectly reconstructed signal candidates, shown on Figure X. Figure 3.11 shows the distri-
butions of $\Delta E$ (left) and $M_{BC}$ (right) for signal and major types of background after the

16

precuts. 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 on both distributions even for signal candidates.
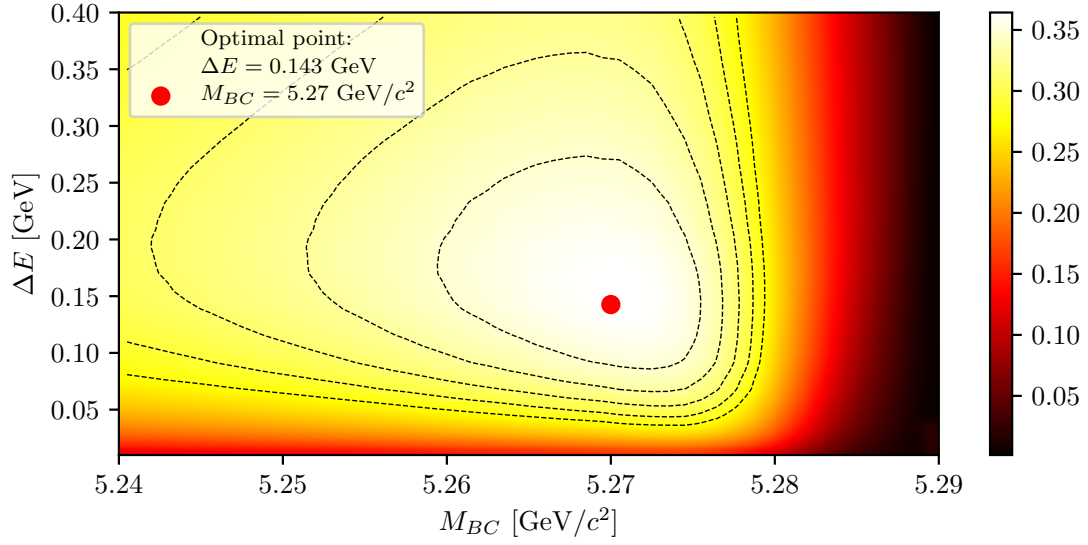


Figure 3.10: 2D *FOM* optimization of the signal window definition, where the signal in the optimization was represented by perfectly reconstructed candidates.
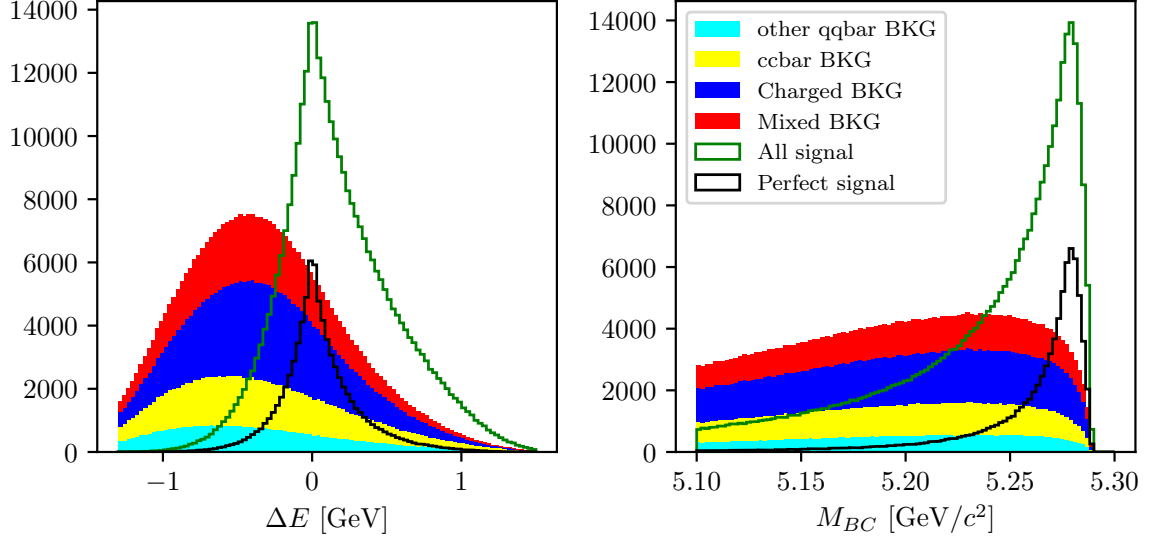
Figure 3.11: Distributions of $\Delta E$ (left) and $M_{BC}$ (right) for signal and major types of background after the precuts. Both signal components are scaled up with respect to the background components, but are in proper scale one to another. The perfect signal has a much better resolution in both distributions, since the event is perfectly reconstructed.

## 3.5 $q^2$ 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. (3.13) should satisfy 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

$$\Delta E' = (E_Y + \alpha E_\nu) - E_{CMS}/2 = 0. \tag{3.15}$$

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

$$M'_{BC} = \sqrt{(E_{CMS}/2)^2 - |\vec{p}_Y + \alpha \vec{p}_\nu|^2}. \tag{3.16}$$

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

18

between $\Delta E$ and $M'_{BC}$, so the latter variable might be the better choice thank $M_{BC}$ in the final extraction fit.

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

$$q^2 = \mathrm{q}^2 = (\mathrm{p}_\ell + \mathrm{p}_\nu)^2 \,. \tag{3.17}$$

The $q^2$ distribution and its resolution are shown in Figure 3.12, along with additional versions of $q^2$, with details of the calculation method in the caption.
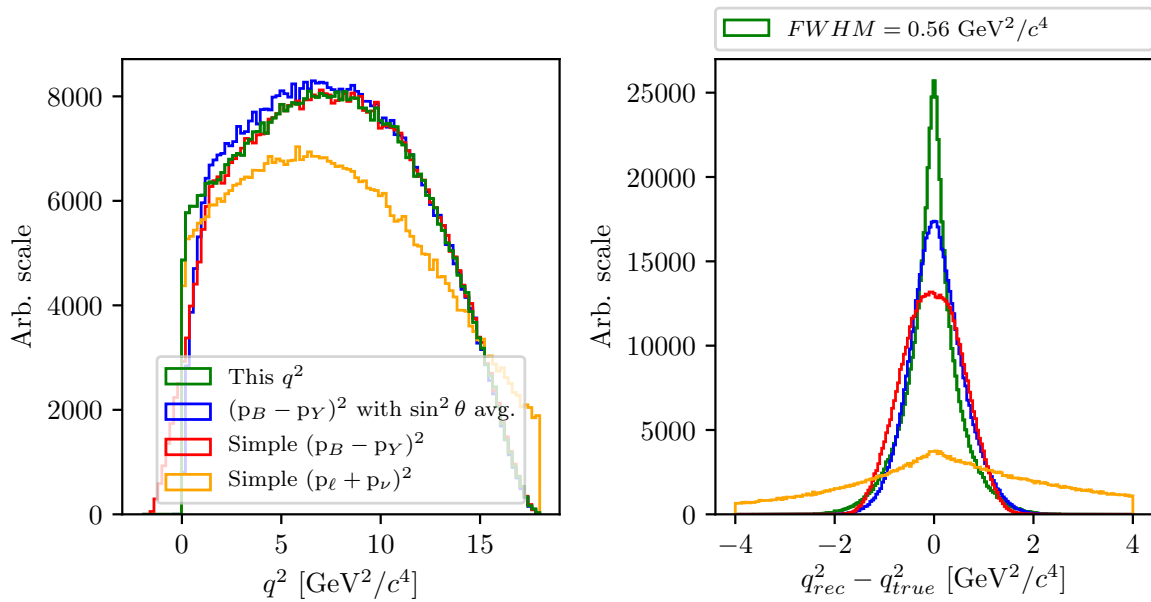


Figure 3.12: Distributions of $q^2$ (left) and $q^2$ resolution (right) for various methods of $q^2$ calculation. The blue distribution follows the procedure in [], whereas the red and the orange ones are straight-forward calculations with available information in the reconstruction. The $q^2$ calculation in red assumes a resting $B$ meson in the CMS frame, and the calculation in orange uses the neutrino four-momentum from Eq. (3.10).

However, 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.

## 3.6 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/$c$, $K/\pi\ ID > 0.6$, $K/p\ ID > 0.1$,

- $B$ mesons:

  - $P(\chi^2,\ DOF) > 1 \times 10^{-3}$, $|\cos \theta_{BY}| < 1$, $|m_{miss}^2| < 7$ GeV/$c^2$, $\Delta E \in [-1.3, 1.5]$ GeV, $M_{BC} > 5.1$ GeV/$c^2$.

# 4 Rest of event clean-up

Continuing from Section 3.3, the description of the ROE clean-up process is described here.

## 4.1 Setting up the MVA

Training the MVA classifiers follows the same recipe for all the steps in this chapter. For each step we run our $B$ meson reconstruction on Signal MC with a generic companion $B$ meson. This way the produced weight files are less likely to be dependent on the signal side 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 leakage of information from the signal side to the ROE side.

For each clean-up step a dataset was prepared for the training of "signal" against "background". The dataset for each case was split into 5 parts for reasons which are explained later in the note. The Fast-BDT (FBDT) [] algorithm was used as the MVA method and the following hyper-parameters were chosen for hyper-parameter optimization

- `nTrees`: the number of trees in the FBDT forest,

- `nLevels`: the number of levels in each FBDT tree.

Figure 4.1 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 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, taken from [].
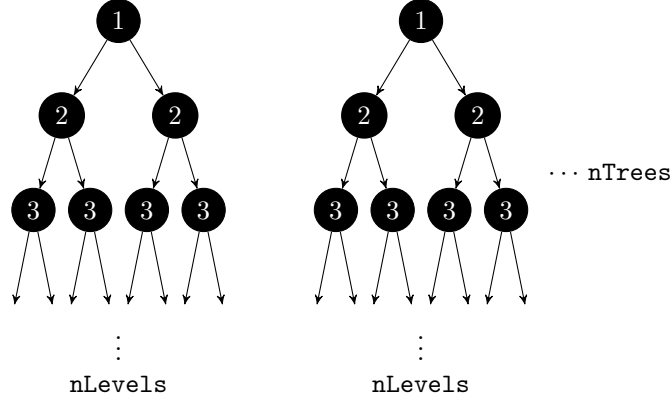
Figure 4.1: A schematic of a BDT forest with `nTrees`, each tree having a depth of `nLevels`.

## 4.2 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 in the ECL, and the energy of the cluster is directly measured via the deposited energy. This way the four-momentum of photons is determined an used in Eq. (3.8).

Generally speaking, most of the photons in events with $B$ mesons should come from $\pi^0 \to \gamma\gamma$ decays. However, a lot of hist in the ECL are also created by photons coming from the beam background or secondary interactions with the detector material. Such photons add extra energy and momentum which spoils our measured quantities if we don't discard them.

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, where we train an MVA to separate good photon clusters from bad ones.

### 4.2.1 $\pi^0$ MVA training

The dataset of $\pi^0$ candidates contains

- 387125 signal candidates,

- 416019 background candidates,

where the definition of signal is that both photon daughters that were used in the reconstruction of the $\pi^0$ are actual photons and actual daughters of the particle, based on MC truth information. We use the converted $\pi^0$ candidates from the existing 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 cases for which the fit converged.

The input variables used in this MVA are

- $p$ and $p_{CMS}$ of $\pi^0$ and $\gamma$ daughters,

- $P(\chi^2, DOF)$ 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 4.2.
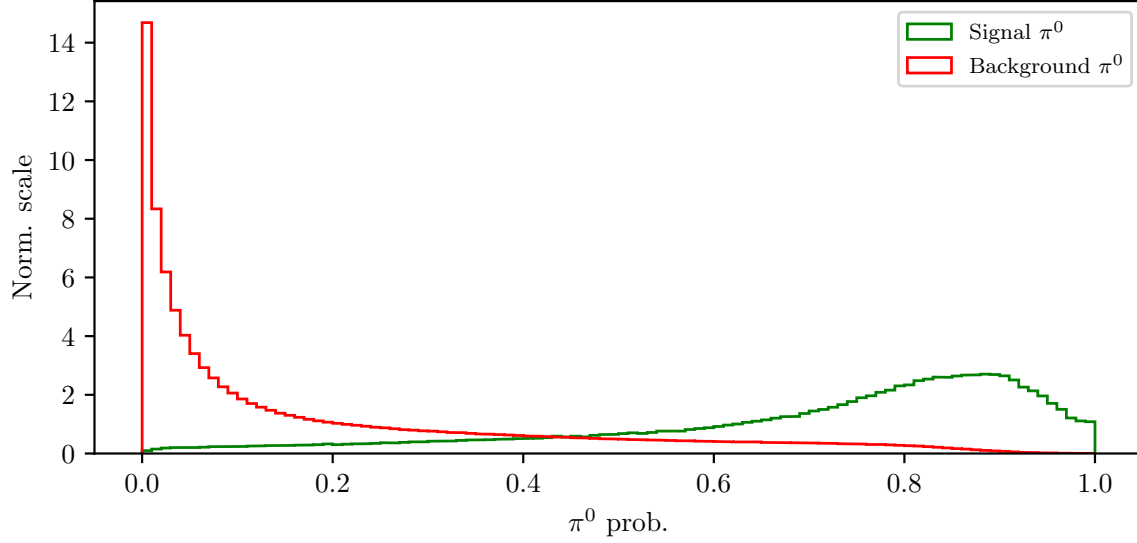
Figure 4.2: Classifier output of the $\pi^0$ training for signal and background $\pi^0$ candidates.

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.

## 4.2.2 $\gamma$ MVA training

The dataset of $\gamma$ candidates contains

- 324781 signal candidates,

- 333353 background candidates,

where the definition of signal 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 $\pi^0$ 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 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.

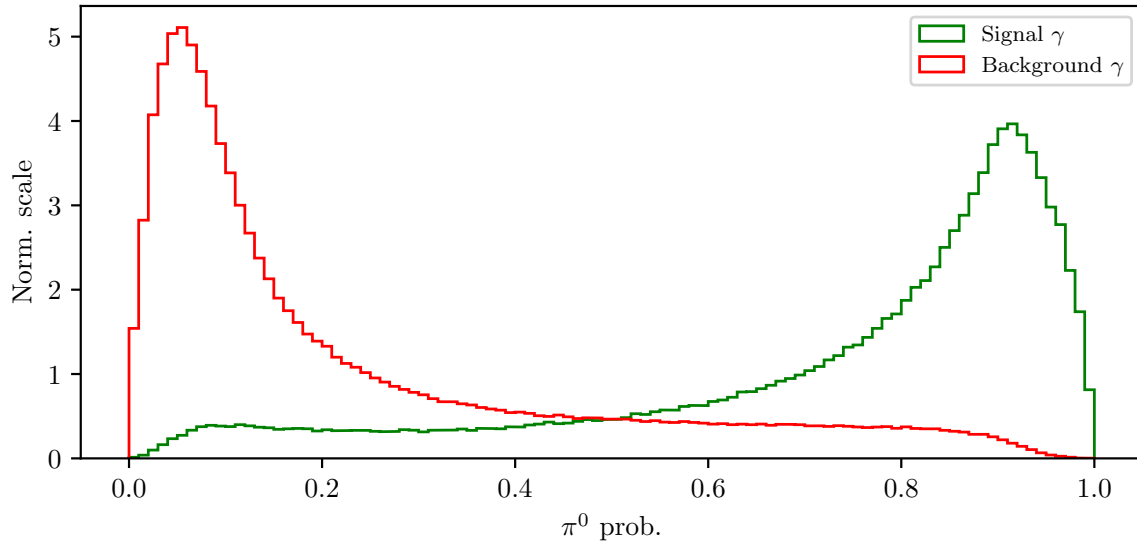The classifier output variable is shown in Figure 4.3.



Figure 4.3: Classifier output of the $\gamma$ training for signal and background $\gamma$ candidates.

## 4.2.3 Clusters clean-up optimization

With the final weights for photon classification in hand, we apply them to the photon particle list. For the purpose of cut optimization, we use only events with correctly

reconstructed signal $B$ mesons. The cut optimization is shown in Figure 4.4 (right), with the optimal cut on the $\gamma$ classifier output at

- $BDT_\gamma > 0.519$.

Figure 4.4 (left) 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: $1 - \epsilon_{BKG} = 81.7$ %.
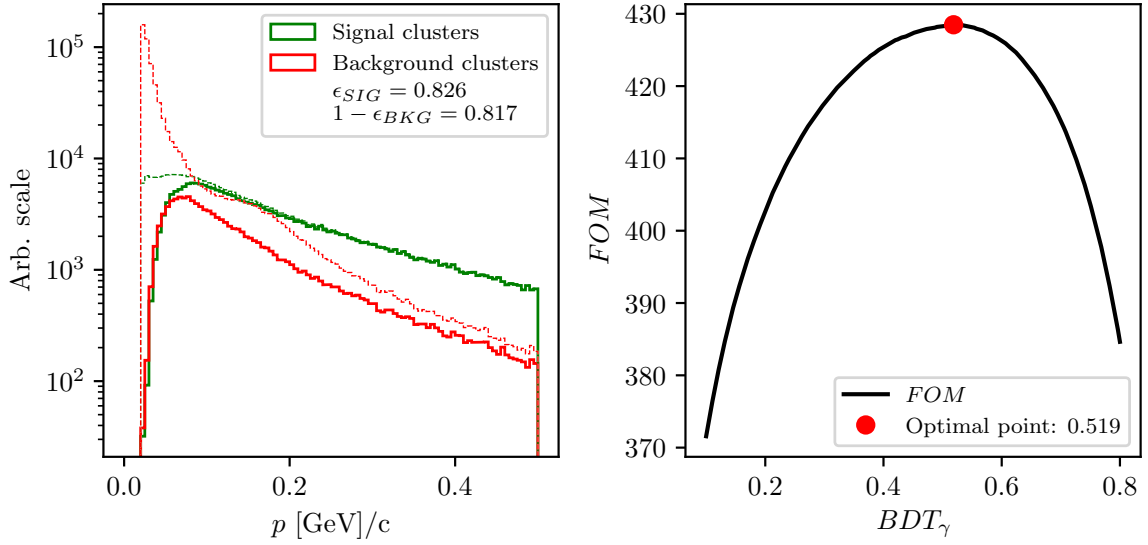


Figure 4.4: Momentum magnitude in the LAB frame of signal and background photon candidates before and after the optimal cut (left) on the classifier output variable (right).

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

## 4.3 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 measured. With the help of particle ID (PID) information, 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. (3.8).

Most of the quality tracks, which come from physics event of interest, come from the IP region, where the collision occurred. Such tracks will from this point on be denoted as *good* tracks. However, cleaning up the tracks is a more complex procedure, due to the following facts which 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,

(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 produce kinks in the flight directory, resulting in multiple tracks.

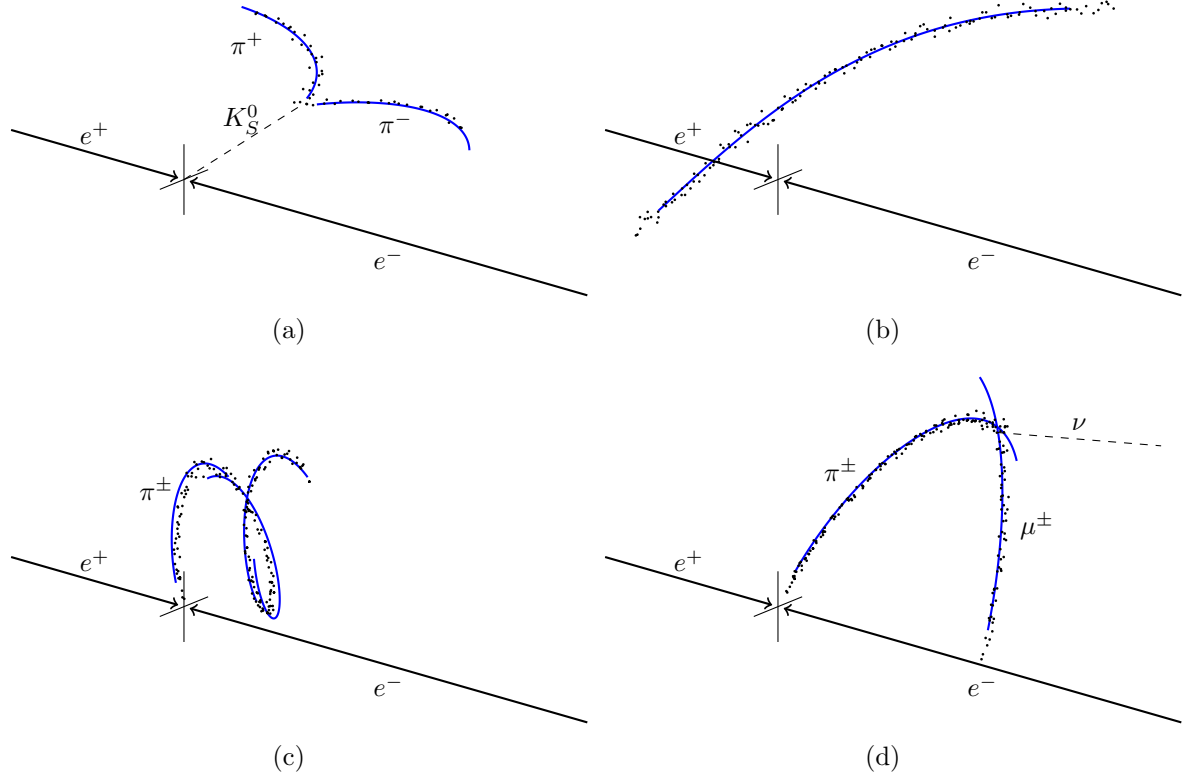Schematics of all the cases mentioned above are shown in Figure 4.5.

Figure 4.5: Schematics of all the special cases of tracks in the detector which need to be considered. (a) shows the tracks from long-lived neutral particles, which decay away from the IP region, (b) shows a random track from background which was reconstructed, (c) shows a case of a low-momentum particle which curls in the magnetic field and (d) shows a case of a particle decay in flight, which produced 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 which add extra four-momentum to our final calculationsin Eq. (3.8). At the same time, we have to take care that we don't discard *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.

## 4.3.1 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 charged $\pi$ 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

28

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 an existing Neural Network classifier result in order to select only the good $K_S^0$ candidates. Figure X shows the distribution of the $K_S^0$ invariant mass for signal and background candidates, before and after the classifier cut.

PLOT

The momentum of selected $K_S^0$ candidates is added to the ROE, while the daughter tracks are discarded from our set.

## 4.3.2 Duplicate tracks

All good tracks at this point should be coming from the IP region, therefore we apply a cut on impact parameters for all the remaining tracks

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

### Defining a duplicate pair tracks

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 X shows the distribution of the angle between two track pairs for random track pairs and duplicate track pairs, where the latter were reconstructed as two same-sign or opposite-sign tracks.

PLOT

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 transverse momentum $p_T$. Since both tracks originate from the same initial particle, the momentum difference should also peak at small values. Figure X shows the momentum and momentum difference of tracks which belong to a random or a duplicate track pair.

PLOT

29

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 number 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 X.

PLOT

**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. The dataset contains

- 215601 signal candidates,

- 311124 background candidates,

where the definition of signal 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 X (left). We optimize the cut and show the $FOM$ function in Figure X (right). The optimal cut for duplicate track selection is

- $BDT_{duplicate} > X$.

**Defining duplicate tracks**

What remains now is to decide which track from the duplicate track pairs 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

$$\Delta f = f_{this} - f_{other}, \tag{4.1}$$

where $f$ is an arbitrary variable from the list of track quantities in Subsection 4.3.2. We denote that a track is more *duplicate*-like, if the following holds from the point-of-view of *this* track

- $\Delta d_0,\ \Delta z_0 > 0$,

- $\Delta p_T,\ \Delta p_Z < 0$,

- $\Delta N_{SVD},\ \Delta N_{CDC} < 0$,

Additionaly we also define an MC truth variable

$$\Delta \chi^2 = \chi^2_{this} - \chi^2_{other}, \quad \chi^2 = \sum_{i=x,y,z} \frac{\left(p_i - p_i^{MC}\right)^2}{\sigma(p_i)^2}. \tag{4.2}$$

If $\Delta \chi^2 > 0$, then *this* track has a higher probability of being a duplicate track and should be discarded.

One additional complication is that we can have more than one extra track from the same particle, which leads to track pairs which share the same tracks. For example, if we have the following case

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

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 impact parameter $d_0$ 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, this way 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 signal candidates,

- 128568 background candidates,

where the definition of signal is that the track 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 X. The weights from this training are applied to the tracks, where now each track has a certain probability of being a duplicate track. However, this is not optimal. We must compare this number to other tracks in duplicate pairs with a new variable

$$\Delta BDT_{final} = BDT_{final}^{this} - BDT_{final}^{other}, \tag{4.3}$$

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

PLOT

Finally, we select all duplicate tracks with a cut

$$\Delta BDT_{final} > 0 \tag{4.4}$$

and discard this set of duplicate tracks from our ROE. The event is now considered to be clean of duplicate tracks.

## 4.4 Clean-up results

## 4.5 ROE clean-up validation

# 5 Background suppression

## 5.1 Event categorization

## 5.2 Continuum suppression

## 5.3 $B\bar{B}$ suppression

## 5.4 Selection optimization