Optimization of long-lived neutral kaon detection and rejection at the Belle II experiment

Lucas Stötzer

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

November 2022

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, 21.11.2022

Date   Signature

1. Gutachter: Prof. Dr. Jochen Dingfelder
2. Gutachter: Prof. Dr. Florian Bernlochner

Acknowledgement

Contents

Introduction

1.1 Fundamentals of particle physics

The standard model in particle physics is one of the best established theories. It describes the elementary particles and forces between them. It consists of two main categories into which the elementary particles are classified, namely fermions and bosons. They are distinguished by their spin, a quantum mechanical attribute. Fermions have half integer, while bosons have integer spins. A picture of the standard model is shown in Fig. 1.1.
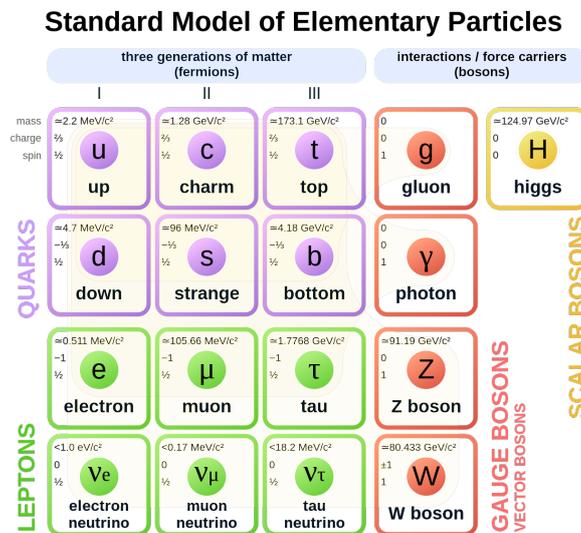
Figure 1.1: The standard model of elementary particle physics. [MIs19]

Fermions are separated in quarks and leptons, which are further classified in three families. Within these families the particles are distinguished by its charge and mass. The Lepton families consist of the electron, muon and tau lepton together with their equally named neutrinos (electron-neutrino, muon-neutrino and tau-neutrino).
The quarks are the elementary constituents of hadrons, like protons, neutrons, pions etc. We categorise

the quarks in three different generations/families by their charge and from light to heavy. A family consists always of two quarks, one with +2/3 the other with −1/3 of charge. The first family consists of the up and down quark, which have masses of only a few MeV. The up-quark has charge of +2/3, while the down-quark has −1/3. The charge stays the same for each quark family. The second family consists of the charm- and strange quark with masses of 1.27 GeV and 93.4 MeV [Zyl+20], respectively. The last family includes the top- and bottom quark, with a mass of 172.69 GeV and 4.18 GeV [Zyl+20], respectively.

The bosons are the force carriers of the different interactions. We observed five bosons namely the photon, the $W^\pm$ and Z^0, the gluon and the Higgs. The photon is the force carrier of the electromagnetic interaction and has a spin of one. The $W^\pm$ and Z^0 interaction particles of the weak interaction and have in contrat to the photon and the gluon, describing the strong interaction, a mass of several GeV. In 2012 the long predicted Higgs boson H was found. However, the Higgs has a spin of zero and thus is called a scalar boson, while the others are vector bosons. With the discovery of the Higgs boson one verified the existance of the Higgs mechanism, which is needed in the standard model. It describes the origin of the elementary particles mass by their interaction with the Higgs field through the Higgs boson. The Higgs boson it self has a mass of 125.25 GeV [Zyl+20]. While the origin of the $W\pm$, Z^0 and elementary particles could be described by the Higgs mechanism the mass of the Higgs itself can not be explained, yet.

The standard model is well proven and succeeds in combining the electromagnetic and weak interaction to one electroweak interaction. However, it also has its limitations. Primary, the unification of all three interactions can not be explained yet and the gravitational force is not considered. Thus, physicists seek to solve those problems and many more by doing high energy colliding experiments. One of those is the Belle II experiment, which we will discuss in the following.

1.1.1 CKM-Matrix

Quarks can change their flavour within or between families by charged weak decays involving $W^\pm$ bosons. Flavour changing neutral current (FCNC) decays ($b \rightarrow s, d$) are highly suppressed. The Cabibbo-Kobayashi-Maskawa (CKM) matrix describes these flavour changes between the different families.

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \text{[Tho19]} \tag{1.1}$$

The probability of a flavor change is proportional to the squared matrix element ($|V_{ij}|^2$). Thus, the most likely flavour change is within a quark family, while the probability that a quark changes the family is reduced, as we can see in Eq. (1.1) on the diagonal and off-diagonal elements, respectively.

The magnitudes of the CKM-matrix elements containing a b-quark are measured in B-Factories. Thus, by carrying out collider experiments and observing the B-meson decays we can test our experimental findings of $|V_{ub}|$ and $|V_{cb}|$ against theory predictions of the standard model.

1.1.2 Semileptonic Decays

In semileptonic B decays, the b-quark will change its flavor to a c- or u-quark via a $W^\pm$ vertex. The $W^\pm$ boson will then decay to a lepton and its corresponding anti neutrino. Measurements, of $R(D) - R(D^*)$

show a deviation of 3.2σ to the standard model [Gro22]. This difference is measured by the ratio of semileptonic B decays of

$$R(D^*) = \frac{\mathcal{BF}\left(B \to D^*\tau\nu\right)}{\mathcal{BF}\left(B \to D^*l\nu\right)} \tag{1.2}$$

Given the deviation to the standard model this hints a strong potential of new physics. Thus, if this is true we should observe a similar deviation in $b \to u\tau\bar{\nu}$ [TW16] events. Especially, measurements of

$$R_\pi = \frac{\mathcal{BF}\left(B \to \pi\tau\bar{\nu}\right)}{\mathcal{BF}\left(B \to \pi l\bar{\nu}\right)} \tag{1.3}$$

can be used to find new physics. However, the identification is problematic due to the small amount of signal in comparison to the large amount of background events. Here one of the most problematic backgrounds are events with a produced long lived kaon (K_L^0). Background events with produced K_L^0's are able to mirror our signal event and show the same event structure. However, a veto of K_L^0 events would lead to a decrease in efficiency since the identification of K_L^0's is challenging due to their high fake-rate. Neutral particles, like neutrons, are likely to be identified as K_L^0 clusters in the KLM and thus are falsify identified as K_L^0 events. Therefore, a highly efficient identification of K_L^0 events is necessary to to improve the sensitivity in R_π or $R(D)$ and $R(D^*)$ measurements.

1.1.3 Long- and short-living kaon states

Neutral kaons are the lightest mesons composed of a s-quark. There are two compositions either $K^0 := d\bar{s}$ or $\overline{K}^0 := s\bar{d}$, with a mass of $m_K = 0.497611\,\text{GeV}$ [Zyl+20]. Thus, they only decay via the weak interaction in either pions or leptons. Further, a K^0 can transform to a $\overline{K}^0$ via box Feynman diagrams shown in Fig. 1.2. This is called neutral kaon mixing.
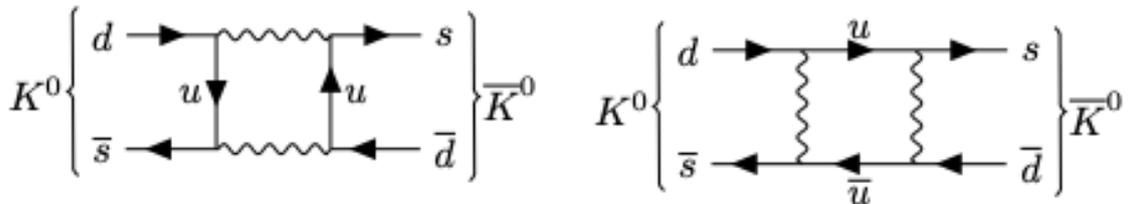
Figure 1.2: Box diagrams of neutral kaon mixing.

However, through this mixing process the K^0 and $\overline{K}^0$ are no charge-parity (CP) eigenstates. Instead one can build linear combinations of those to form CP eigenstates, denoted as K_1 and K_2. Applying the CP operator to those we see that the K_1 state is a CP-even, while K_2 is a CP-odd eigenstates.
In experiments two types of physic state neutral kaons with different lifetimes of $\tau_{K_S^0} = (8.954 \pm 0.004) \cdot 10^{-11}\,\text{s}$ and $\tau_{K_L^0} = (5.116 \pm 0.021) \cdot 10^{-8}\,\text{s}$ [Zyl+20] were discovered. Thus, those two states are named short-living (K_S^0) and long-living kaon (K_L^0), respectively. If CP is conserved in weak decays we could

identify the physic states as

$$|K_S^0> \approx |K_1> \quad \text{and} \tag{1.4}$$

$$|K_L^0> \approx |K_2>. \tag{1.5}$$

However, Cronin and Fitch discovered decays of a K_L^0 to two pions suggesting that the K_L^0 has a CP-even component. This was the first observation of direct CP violation in neutral kaon systems [Tho19]. The correct quantum mechanical picture of the K_S^0 and K_L^0 is derived in [Tho19] and given by

$$|K_S(t)> = \frac{1}{\sqrt{1+|\epsilon|^2}} \left[|K_1> +\epsilon|K_2>\right] e^{-i\lambda_S t} \quad \text{and} \tag{1.6}$$

$$|K_L(t)> = \frac{1}{\sqrt{1+|\epsilon|^2}} \left[|K_2> +\epsilon|K_1>\right] e^{-i\lambda_L t}. \tag{1.7}$$

Here, the exponential function gives the time evolution of the system and ϵ is the CP violating factor.

1.2 The Belle II experiment

The Belle II experiment is a high energy collider experiment at the SuperKEKB accelerator in Tsukuba, Japan. It is the successor to Belle. Belle II is also referred to as a B-Factory experiment, since many B-mesons are created. In Fig. 1.3 a sketch of the experiment is shown.
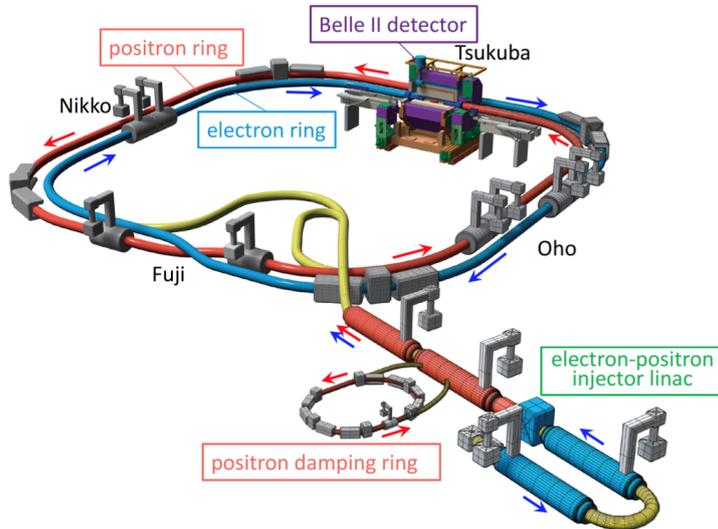

Figure 1.3: Schematic view of the Belle II experiment at the SuperKEKB accelerator. [AFK18]

As it is shown, electrons and positrons are accelerated up to energies of 4.0 GeV and 7.0 GeV [Abe+10],

respectively. They collide inside the Belle II detector at the interaction point (IP), with a center of mass energy of $\sqrt{s} = 10.58$ GeV. This energy is exactly at the $\Upsilon(4S)$ resonance, which will decay to a neutral or charged pair of B-mesons. The $\Upsilon(4S)$ resonance is the fourth resonance of $b\bar{b}$-quark pair and thus a bottomonium. Since the mass of the $\Upsilon(4S)$ lies above the $B\bar{B}$-threshold it will most likely decay to either neutral or charged B-mesons.

The Belle II detector is an onion-shaped detector with near-4π coverage with different layers of detector types. The closest to the interaction point (IP) is the Pixel Detector (PXD) and a Sillicon Vertex Detector (SVD). The main purpose of both is to measure the B-meson decay vertex. Then a Central Drift Chamber (CDC) for the track reconstruction and momentum measurements, particle identification and as trigger for signals of charged particles is installed. Followed by that is a Time-Of-Propagation (TOP) and an Aerogel Ring Imaging Cherenkov (ARICH) counter for particle identification. Both are Cherenkov detectors measuring the angle between the particle and radiated photon called the Cherenkov angle. Thus, by measuring the momentum of the particle we can determine its mass and therefore identify the particle. Thereafter, an Electromagnetic Calorimeter (ECL) is placed. In the ECL we are using scintillation crystals to detect photons and electrons efficiently. Last but not least, the Klong-Muon-Detector (KLM), which we will discuss in Section 1.2.1 in detail, is installed. [Ada+18][Abe+10] For a more detail on the other detector parts see [Ada+18] and [Abe+10].

1.2.1 Klong-Muon-Detector

The Klong-Muon-Detector (KLM) is the most outer detector of the Belle II experiment. It consists of a barrel, and forward and backward endcap, shown in Fig. 1.4.
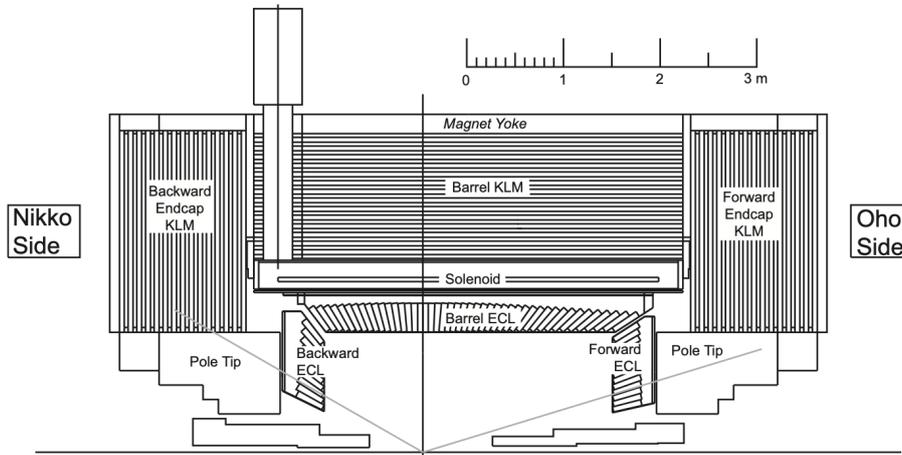
Figure 1.4: Profile of Belle II KLM upper half. [Abe+10]

The main purpose of the KLM is to detect muons, which leave charged tracks in the CDC, and K_L^0's, that nearly[1] do not interact in any other detector part.

The barrel KLM (BKLM) covers a region of $45°$ to $125°$ in the polar angle. We use alternating glass-electrode resistive plate chambers (RPC) and iron plates for the detection, as in Belle since they

---

[1] Except of the ECL, where a slight chance of an interaction exists.

showed good performance [Abe+10]. However, in the first two layers of the BKLM, scintillators with wavelength-shifting fibers are used, since the high deadtime of the RPC modules would result in major performance losses. The iron-plates are there to detect neutral particles, which will hadronically interact with the iron creating a hadronic shower. [Ada+18].

RPCs are build of two parallel glass plates with a gas gap in between. The glass plates are used as electrodes, where a high voltage is applied, forming an electric field. If a charged particle transverses through a RPC, it will ionize the gas. Those ions and free electrons will then be accelerated by the electric field towards the electrodes. Hereby, it is possible that additional interactions with the gas atoms occur, resulting in an avalanche of charge carriers. A pulse is then measured by readout strips mounted on top of the electrodes. Through the avalanche a decreasing potential between the electrodes can be observed, leading to a deadtime in which this segment can not detect further signals.

In order to obtain 3D information and a high detection efficiency orthogonal readout stripes are installed on both sides of the RPC, forming a so called superlayer [Abe+10]. Thus the overall detection efficiency is increased to roughly 99% [Abe+10]. The advantage of this configuration is that both RPC modules have independent electrodes and gas chambers. So if one module fails the other one still works with a muon detection efficiency of roughly $(90 - 95)\%$ [Abe+10]. A sketch of a superlayer RPC is shown in Fig. 1.5.
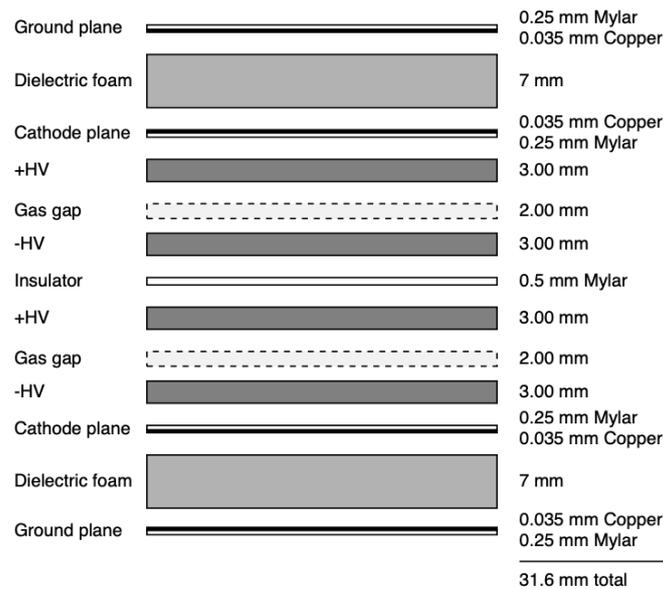

Figure 1.5: Profile of a RPC superlayer used in the barrel of the Belle II KLM. [Abe+10]

For the endcaps of the KLM (EKLM), the RPC performance would suffer due to their high deadtime and the higher luminosity of the SuperKEKB accelerator [Ada+18]. The background rate of neutrons and other particles created along the beam line is increased in the SuperKEKB accelerator and would therefore cause a reduction of our signal detection efficiency, if using RPC modules. Thus, scintillators with wavelength-shifting fibers are used together with Silicon PhotoMultipliers (SiPMs) as readout [Aus+15].

1.3 K_L^0 Identification

In order to understand how K_L^0's are identified, we first have to understand how a cluster is built in the KLM. A cluster consists of multiple hits[2] lying in a $5°$ cone. The cone is build from the interaction point, by a straight line to the center of the possible cluster candidate.

In order to identify now a cluster as a K_L^0 we will be checking if the cluster is considered to be charged or neutral. Therefore, a second cone surrounding the cluster from the interaction point is applied, with an opening angle of $15°$. Measured tracks in the CDC are extended to their entrance point in the KLM. If a line, drawn from this entrance point to the interaction point, lies then within the $15°$ cone, the cluster is not considered to be a K_L^0 cluster. Last but not least, we require the K_L^0 cluster to have at least two hits in different layers. For a K_L^0 cluster build from the KLM and ECL we require at least one hit in the KLM and ECL, respectively [Abe+10].

Nevertheless, often clusters are identified as K_L^0 clusters although they originate from other particles, like beam induced fast neutrons created by beam background effects. Thus, the efficiency of identifying a K_L^0 meson properly is not great. This is shown in [Tch+02] that the identification efficiency of good K_L^0 clusters is approximately 50% for K_L^0's with an momentum between 2 and 4 GeV. While in [Aus+15] it is shown that the identification efficiency in the Belle II experiment is approximately between 60% and 70% for a momentum region of 2 to 3 GeV. However, the number of fake clusters is high, which decreases the efficiency of for example $R(\pi)$ analyses discussed in Section 1.1.2.

1.4 Search for K_L^0 in BELLE II

We aim to build a new analysis tool, which helps us to reject fake K_L^0 clusters and identifying good K_L^0 clusters efficiently. Thereby, an increase in the efficiency of for example $R(\pi)$ analyses shall be obtained (see Section 1.1.2). Thus, we perform a data-driven Boosted-Decision-Tree training on good and bad K_L^0 clusters to identify K_L^0 clusters and separate them from fake clusters.

1.4.1 Initial state Radiation in ϕ Production

For this training we need to find a pure sample of good and bad K_L^0 cluster candidates to train on. The decay channel of

$$e^+ e^- \to \gamma_{\text{ISR}} \left[\phi \to K_L^0 K_S^0 \right]$$

offers the right conditions for this. Here, a photon gets radiated from the electron or positron beam, reducing the energy of the collision to an arbitrary value including the mass of a ϕ meson, a vector meson composed of a $s\bar{s}$ quark pair. The ϕ decays, with a branching ratio of $\mathcal{BF}(\phi \to K_S^0 K_L^0) = 33.9\%$ [Zyl+20], to a K_S^0 and K_L^0, respectively. The K_S^0 has a much lower lifetime of $\tau_{K_S^0} = 8.954 \cdot 10^{-11}$s than the K_L^0 with $\tau_{K_L^0} = 5.116 \cdot 10^{-8}$s [Zyl+20]. Thus, the K_S^0 will further decay, within the detector, to two charged pions, with a probability of $(69.20 \pm 0.05)\%$ [Zyl+20], while the decay of the K_L^0 can only rarely be observed. A sketch of this process is shown in Fig. 1.6.

[2] At least two hits in different KLM layers.
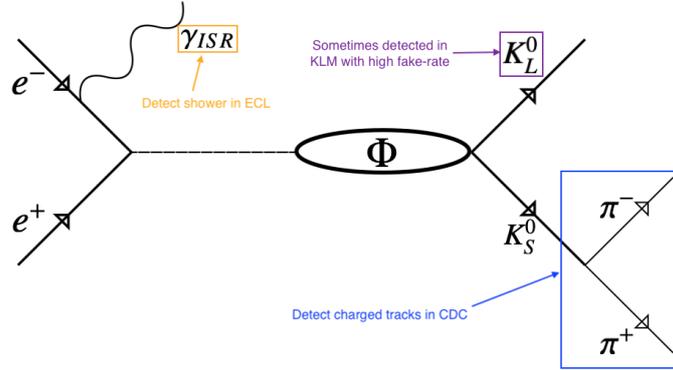
Figure 1.6: Sketch of our decay process used for K_L^0 identification. We detect the ISR photon (orange) as a high energy shower in the ECL and the two pions (blue) as tracks in the CDC. The K_L^0 is sometimes detected in the KLM, but with a high fake-rate.

We choose this channel since it is easy to identify and reconstruct, due to the fact that we expect two charged pions leaving tracks in the CDC originating from the same vertex displaced from the IP, a high-energy neutral shower in the ECL from the high energy photon and one neutral cluster in the KLM. Thus, by reconstructing our K_S^0 from the two pions and the high energy photon we can estimate the missing four momentum vector and the direction of the K_L^0 without reconstructing it explicitly.

In this analysis we will first look on signal Monte Carlo (MC) and then data. Signal MC is a MC simulation of this decay only, without any background. The reason why we are looking into signal MC is to evaluate whether we can identify good K_L^0 candidates, before investigating data. Thus, we simulate events where a virtual photon decays in a high energy photon (γ_{ISR}) and a ϕ meson in 100% of the time. The ϕ meson will further decay as shown in Fig. 1.6. We generate around 50000 events, so that we have sufficient statistics.

For the reconstruction of the events we make some loose pre-selections. First, we reconstruct our K_S^0 from the two charged pions and define a selection on the reconstructed mass of the two pions by $0.45 < m_{\pi\pi} < 0.55$ GeV. We choose this lower and upper limits since we know that the pions originate from the K_S^0, with a mass of $m_{K_S^0} = 0.4976$ GeV [Zyl+20]. Secondly, we require the high energy photon to have an energy between $3 < E_\gamma < 9$ GeV in the lab frame. We choose these boundries since we expect to observe a high asymmetric resolution for high-energy photons in the measured energy by the ECL, due to energy leakage. Further, one can validate that those limits are reasonable by calculating the energy analytically. We can do this by looking on a two body decay, where either the positron or electron beam radiates a ISR photon and produces a ϕ meson in the collision. We know that the center of mass energy[3] is $\sqrt{s} = 10.58$ GeV, as this is the operation energy of the Belle II experiment. Further, we know the mass of the ϕ meson. Thus, we can calculate the energy of the photon to be

$$E_\gamma^\phi = |p_\gamma^\phi| = \frac{\sqrt{s}^2 - m_\phi^2}{2\sqrt{s}} \tag{1.8}$$

$$E_\gamma^\phi = 5.24083 \, \text{GeV}. \tag{1.9}$$

---

[3] Here also referred to as the beam energy.

1.4.2 Initialstate Radiation in J/ψ Production

Another decay mode used for the identification of K_L^0 mesons is:

$$e^+e^- \rightarrow \gamma_{\mathrm{ISR}} \left[J/\psi \rightarrow \pi^+\pi^- K_L^0 K_S^0 \right].$$

Here the J/ψ is a meson composed of a $c\bar{c}$ and thus a resonance state. Special to this process is that we expect to have four charged pions in the final state. This is leading to an increased number of multiple K_S^0 candidates in an event. Thus, we introduce a best candidate selection, ranking the possible K_S^0 candidates by the significance of distance (σ_d), giving the significant distance of the K_S^0 decay vertex to the interaction point. Hence, we neglect bad candidates and use for the analysis only the best one. Our selection for the invariant mass of the two pions made in Section 1.4.1 can be transferred. However, for the high energy photon we calculate the analytical energy, by Eq. (1.8) to assure that we are still covering a good energy range. We get an energy of

$$E_\gamma^{J/\psi} = 4.83674 \, \mathrm{GeV}, \tag{1.10}$$

using the mass of the J/ψ in Eq. (1.8) instead of the ϕ mass. Thus, we can use the same pre-selection as in Section 1.4.1.

1.4.3 Fast Boosted-Decision-Tree

Once we have a highly pure sample of K_L^0 cluster candidates we aim to train a fast Boosted-Decision-Tree (BDT), which is a Belle II analysis tool, to identify good K_L^0 clusters and separate them from falsely identified K_L^0 clusters. The BDT is an advanced machine learning algorithm of the Decision-Tree (DT). A Decision-Tree is a supervised technique used in machine learning. With such an algorithm we can train a machine to optimize the separation between signal and background events based on some features, also called nodes. A DT consists of multiple layers of these nodes. Based on those a number of cuts and a probability for each cut is determined, by fitting the feature. In Fig. 1.7 a DT is illustrated.
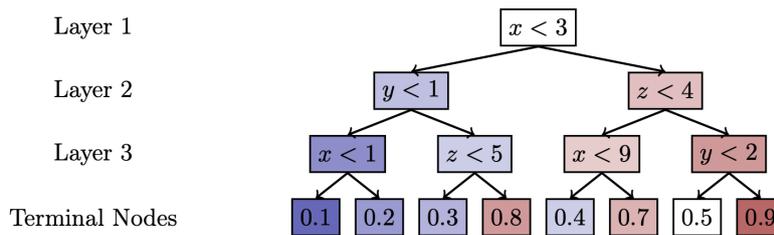

Figure 1.7: Illustration of a Decision-Tree. The DT consists of multiple layers and nodes. At each node a decision is made to advance in the layers. Finally we categorize the outcome in the terminal node giving the probability of the sample to be signal [Kec16].

The separation power is determined and maximised. In the end the highest power will be used as the final cut. This power gives us how good we can distinguish between signal and background by applying a certain cut. Still, a normal decision tree is susceptible to over-training, which means that the determined cuts are too strict to allow statistical fluctuations. Thus, a DT is called a weak learner. Hence, we use a

Boosted-Decision-Tree (BDT) with multiple single Decision-Trees (weak learners). Each of these DT's are fitted and weighted by the minimization of a loss function [Kec16]. For features that are difficult to categorize the weight is then increased. Thus, by multiple iterations of the algorithm a decrease in importance for the BDT is observed. Hence they can be neglected [Kec14].

After the training the classification is applied to a testing sample. Thus, we test the classification on an unknown sample for over- or under-fitting or in other words, if our model of the BDT is to complex or simple. For a more detailed explanation please see [Kec16; Cor+22].

K_L^0 identification in initial state radiation channels

We first look at the ϕ channel, since this is the easier one to identify and reconstruct of the two decay channels introduced in Section 1.4. Thus, as described in Section 1.4.1 we first generate signal MC events. Afterwards we search for this mode in a data sample, with an integrated luminosity of $L_{\text{int}} = 125.7907\,\text{fb}^{-1}$. We use signal MC to validate our calculations and to estimate what number of signal events to expect in data. Our aim is to perform a data-driven BDT-study in order to distinguish between good and bad K_L^0 cluster candidates in the KLM detector. Therefore, we first need a high purity selection of K_L^0 candidates in the first place. We can achieve this by selecting high purity K_S^0-γ_{ISR} candidates, since in our decay mode $e^+e^- \rightarrow \gamma_{\text{ISR}}\left[\phi \rightarrow K_L^0 K_S^0\right]$ we expect a K_L^0, once we find a K_S^0 in combination with a high-energy photon.

2.1 High purity K_S^0 selection

We want to define selections to collect a highly pure K_S^0-γ_{ISR} sample. Therefore, we first have to validate that our reconstruction of K_S^0 candidates works. Here, we first require that no tracks shall be left in the rest of event and first ignore all KLM clusters. We select only one K_S^0-γ_{ISR} candidate in an event. The selections discussed in this section and the following subsections are applied in the rest of the analysis. Later we allow multiple K_S^0-γ_{ISR} candidates per event. The selection of one K_S^0-γ_{ISR} candidate per event is used to estimate how many events as opposed to candidates we should expect.

We first take a look at the invariant mass of the two daughter pions. We expect a sharp peak at $m_{\pi\pi} = 0.497611\,\text{GeV}$ [Zyl+20]. The distribution is shown in Fig. 2.1, for signal MC (left) and for data (right).
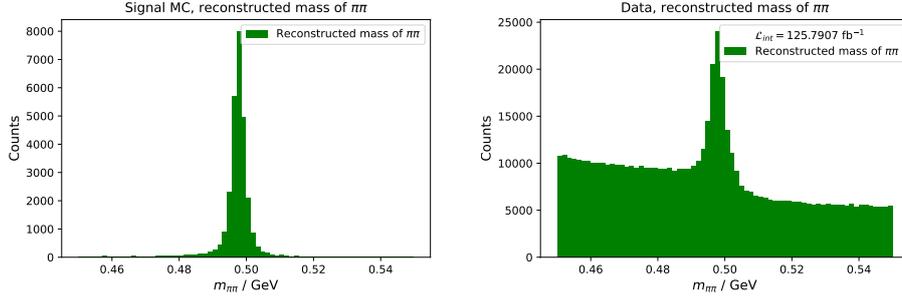
Figure 2.1: Distribution of the reconstructed mass of the two daughters of the K_S^0 for signal MC (left) and data (right). In both a sharp peak at the nominal K_S^0 mass is observed. In data of $L_{int} = 125.7907\,\text{fb}^{-1}$ a large amount of background is observed as well.

In both signal MC and data, we observe a sharp peak at the K_S^0 mass. Thus, we see that our reconstruction works. However, the purity of the K_S^0-γ_{ISR} sample is low due to the amount of background events. Thus, we aim to enhance the purity, meaning the percentage of good candidates in a defined region, by applying certain selections discussed in the following subsections. In Table 2.1 a list of these and their meaning is given. We use these in order to obtain a high purity and detection efficiency.

Selection Name	Description
Number of track candidates in the event beside the two pions (nROE)	To neglect events with background tracks. We only expect two tracks from the two pions, which are reconstructed as daughters of the K_S^0.
Reconstructed mass of the two pions ($m_{\pi\pi}$)	Reconstructed mass of the two charged pions. We use this selection to define a signal region of good K_S^0 candidates.
Significance of distance (σ_d)	The significance of distance gives the significant distance of the K_S^0 decay vertex to the interaction point (IP). We use this selection to enhance the K_S^0 purity.
ISR photon energy (E_γ)	Photon energy in the center of mass frame (CMS). We use this selection to enhance the K_S^0 purity.

Table 2.1: List of selections we use to obtain a high purity K_S^0 selection.

A detailed description of the calculation of the efficiency, purity and the algorithm to obtain the optimal values is given in the following Section 2.1.1.

2.1.1 Selection efficiency and purity

In order to obtain a high purity sample without reducing the selection efficiency, in data, we choose the optimal selection value by minimizing the distance between the calculated efficiency and purity and the optimal one at $\{\epsilon, P\} = \{1.0, 1.0\}$. Therefore, we use the **scipy.optimize.minimize** package [com22a]. We always calculate the selection efficiency ϵ and purity P simultaneously by defining a signal region in

the spectrum of the reconstructed mass of the two pions. We define this region by performing a global fit of a Gaussian as signal, combined with a polynomial of 1st order as background, shown in Eq. (2.1).

$$h(x) = \frac{A}{\sqrt{2\pi\sigma}} \cdot \exp - \left(\frac{(x-\mu)}{\sigma}\right)^2 + a \cdot x + b \tag{2.1}$$

Thus, we define the signal region as

$$\mu \pm 5 \cdot \sigma,$$

where μ is the mean and σ the width of the Gaussian in Eq. (2.1). The definition of our signal region is illustrated in Fig. 2.2.
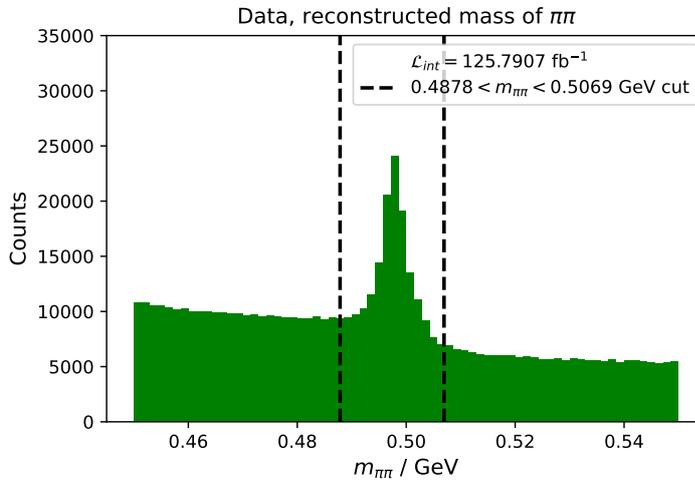
Figure 2.2: Definition of the signal region for the K_S^0 selection, through a global fit to the $m_{\pi\pi}$ spectrum in form of a Gaussian plus a polynomial function of the first order (see Eq. (2.1)). We use the signal f_{Sig} and background f_{Back} function to calculate the selection efficiency and purity of K_S^0 candidates.

By using the signal f_{Sig} and background f_{Back} function of the fit we determine the selection efficiency and purity by integrating over the signal region[1] to

$$\epsilon = \frac{\int_l^u dx\, f_{\text{Sig}_i}}{\int_l^u dx\, f_{\text{Sig}_0}} \text{ and,} \tag{2.2}$$

$$P = \frac{\int_l^u dx\, f_{\text{Sig}_i}}{\int_l^u dx\, f_{\text{Sig}_i} + \int_l^u dx\, f_{\text{Back}_i}}. \tag{2.3}$$

Here, the index i refers to the selection for which the efficiency and purity shall be calculated and f_{Sig_0} is the signal function before applying any selections[2]. We determine the distance δ between the optimal

---

[1] $l \mathrel{\hat{=}}$ lower boundary and $u \mathrel{\hat{=}}$ upper boundary.

[2] Except for the charged track cleaning selection

point (ϵ_{opt}, P_{opt}) and our calculated selection efficiency and purity as

$$\delta = \sqrt{\left(P - P_{\text{opt}}\right)^2 + \left(\epsilon - \epsilon_{\text{opt}}\right)^2}.$$

The minimization algorithm from **scipy.optimize.minimize** [com22a] uses then different cut values and therefore new f_{Sig_i} to search for the shortest distance between the calculated and optimal point. In this way we obtain the selection values for the significance of distance of the K_S^0 decay vertex towards the interaction point σ_d, and the ISR photon energy E_γ.

2.1.2 Significance of distance σ_d

As shown in Table 2.1 the significance of distance is the significant distance between the IP and the K_S^0 decay vertex. We use it to identify a shared displaced vertex of the two pion tracks, which are reconstructed to find K_S^0's. In Fig. 2.3 the distribution is shown for signal MC (left) and data (right).
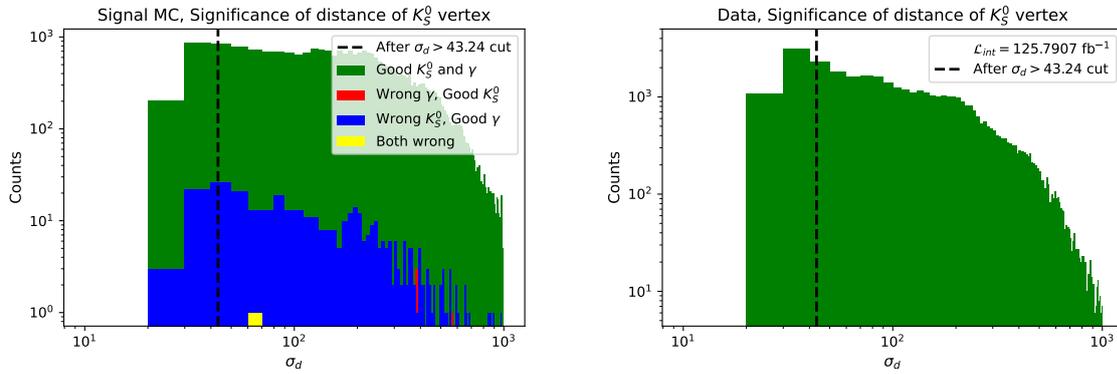
Figure 2.3: Distribution of the significance of distance of the K_S^0 decay vertex from the interaction point for signal MC (left) and data (right). In signal MC we use the cut defined by the minimization done in data Section 2.1.1 and select candidates with $\sigma_d > 43.24$. Further, events are shown where either the K_S^0(blue), the ISR photon (red) or both (yellow) are not reconstructed correctly. In data we apply the cut by the algorithm described in Section 2.1.1. Thus, we define a cut at $\sigma_d > 43.24$ in data.

In signal MC the distribution is divided into signal and background distributions. We observe many candidates close to the interaction point. With increasing distance the amount of K_S^0 candidates decreases, as expected. Close to the interaction point particles are often falsely identified as K_S^0 mesons. Thus, we select only events above $\sigma_d = 43.24$ in signal MC in order to reject falsely reconstructed K_S^0 candidates. This selection is done using the minimization algorithm on data.

In data we observe the same behaviour as in MC. In order to select a pure and efficient sample of K_S^0 candidates we use the minimization algorithm described in Section 2.1.1 in order to find the best selection to obtain a high purity sample of K_S^0 candidates. Thus, we select candidates with a significance of distance above $\sigma_d = 43.24$. By Eqs. (2.2) and (2.3) we achieve a purity of 90.18% with an efficiency of 89.98%.

2.1.3 ISR photon energy E_γ

The next variable of interest is the ISR photon energy. As shown in Section 1.4.1, we expect to find a peak at the analytical energy of $E_\gamma^\phi = 5.24083$ GeV (cf. Eq. (1.8)) for this decay. In Fig. 2.4 the photon energy distribution is shown in the CMS frame, on the left for signal MC and on the right for data.
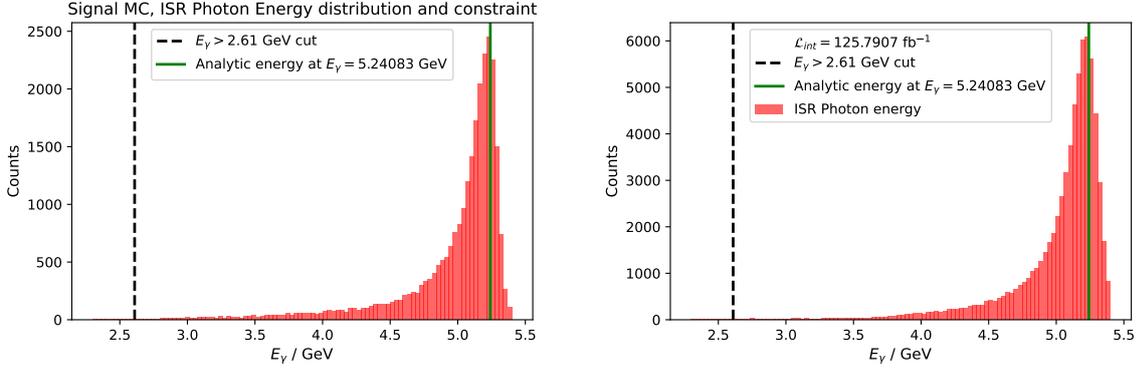

Figure 2.4: Distribution of the ISR photon energy after the significance of distance selection for signal MC (left) and data (right). In signal MC we apply a selection at $E_\gamma > 2.61$ GeV, using the same selection as in data. In data we determine a selection by the minimization algorithm resulting in a cut at $E_\gamma = 2.61$ GeV. The green line shows the analytical value for the energy of the photon in an $e^+e^- \to \gamma_{\mathrm{ISR}}\phi$ decay. Due to an energy leakage of the ECL shown by the long tail to lower energies we introduce the analytical value of the photon energy as a constraint.

In both figures the analytical value of the photon energy is shown as the green line. At this value we observe a peak of the measured photon energy by the ECL. Thus, the reconstructed photon energy matches our expectations at first sight.

Investigating both distributions more closely we observe a long tail towards lower energies in signal MC and in data. This is caused by an energy leakage of the ECL crystals. Thus, the energy measurement from the ECL has a high asymmetric resolution for high-energy photons. Therefore, to not rely on the large energy resolution asymmetry we introduce an energy constraint for the photon, by using the analytical value in Eq. (1.8) for $e^+e^- \to \gamma_{\mathrm{ISR}}\phi$ events, in all further calculations. This energy constraint is important for the analysis as we will see in Section 2.2.1. Furthermore, due to using the constraint we have to scale the momentum components of the photon in the following way:

$$p_\gamma'^i = p_\gamma^i \cdot \frac{E_\gamma^\phi}{E_\gamma}. \tag{2.4}$$

We get the scaled momenta $p_\gamma'^i$ by multiplying the measured momentum components p_γ^i with the ratio of the analytical energy value from Eq. (1.8) and the measured energy E_γ by the ECL.

In data (Fig. 2.4 right) we use the minimization algorithm to find the optimal selection. One observes that the minimization algorithm chooses a rather loose cut. Since, we already reduced the background by the significance of distance selection it seems that there are no background candidates left. Thus, we assume that the high energy photon selection is not very powerful to enhance our purity. In signal MC we apply the same selection.

The selection efficiency and purity is now calculated in combination with the selection applied to the

significance of distance. In Fig. 2.5 the selection efficiency and purity for each cut is illustrated. Due to the lack of background candidates, the significance of distance and photon energy selection overlap in Fig. 2.5. This validates our expectation that the photon energy selection shows only a small effect for selecting a high purity K_S^0-γ_{ISR} sample, due to the lack of further background events.
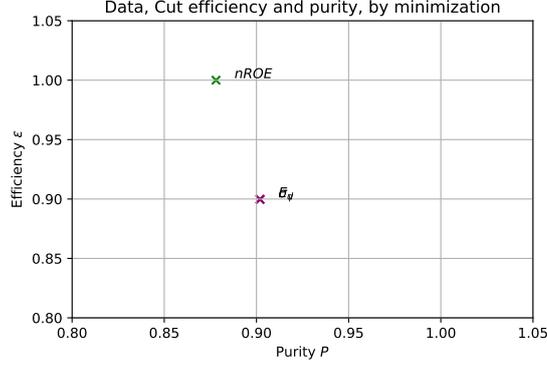

Figure 2.5: Selection efficiency ϵ and purity P of K_S^0 candidates without (nROE) and after significance of distance (σ_d) and photon energy selection (E_γ). The error bars are to small to display. An overlap of the σ_d and E_γ selection is visible, validating our expectation that a selection has only a small impact on the selection efficiency and purity of our K_S^0-γ_{ISR} sample.

We find a sample purity of 90.20%. In comparison to the selection efficiency and purity achieved by applying only the significance of distance selection we do not observe a large improvement, gaining 0.02% in total. Overall, with the selections applied so far a highly pure sample of K_S^0-γ_{ISR} candidates is created. Thus, we expect to find K_L^0 candidates with a similar purity.
Before we try to find good K_L^0 candidates we take a look at the two pion mass after all selections, in Fig. 2.6.
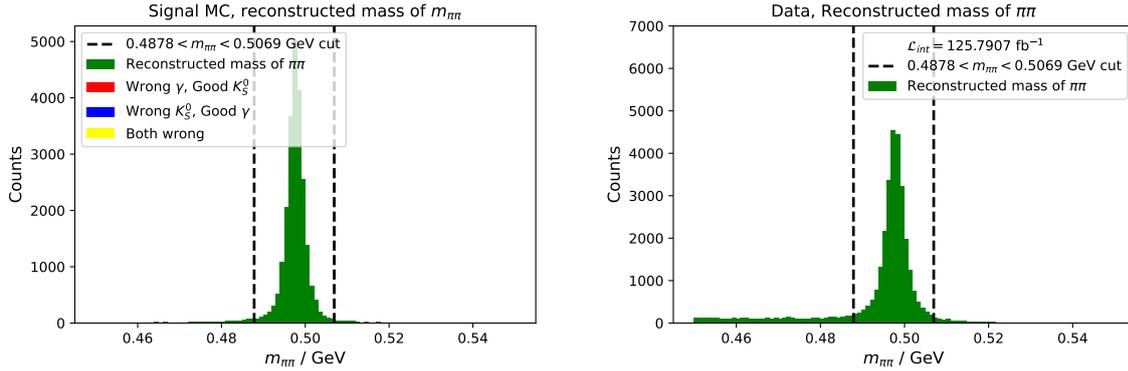
Figure 2.6: Distribution of the reconstructed mass of the two charged pions for signal MC (left) and data (right) after selections on σ_d and E_γ. We choose the same signal region in signal MC as in data, between $0.4878 < m_{\pi\pi} < 0.5069\,\mathrm{GeV}$.

We observe in both signal MC and data a sharp peak. In comparison to Fig. 2.1 we see the background is

significantly smaller. We apply a selection to the two pion mass defined by the signal region defined before (see Section 2.1.1), rejecting side band background events with clearly no good K_S^0 candidates. In Fig. 2.6 this is shown by the black dashed lines.

Using this highly pure and efficient sample of K_S^0-γ_{ISR} candidates we now want to find good K_L^0 candidates and predict their momentum and energy.

2.2 K_L^0 prediction and definition of good and bad KLM clusters

2.2.1 Missing mass selection

We now investigate, if we are able to predict the presence of good K_L^0 candidates, based on our highly pure K_S^0-γ_{ISR} selection. So far we have not reconstructed any K_L^0. Instead, we use the four-momenta of the beam, the ISR photon and the K_S^0 to determine the four-momentum of the missing particle. If the event is reconstructed correctly this particle is a K_L^0. Thus we calculate the momentum components from

$$p_{miss} = p_{beam} - p_\gamma - p_{K_S^0}. \tag{2.5}$$

We use the energy constraint introduced in Section 2.1.3 for the ISR photon energy. In the missing mass spectrum:

$$m_{\text{miss}} = \sqrt{E_{\text{miss}}^2 - \vec{p}_{\text{miss}}^2} \tag{2.6}$$

we expect a peak at the K_L^0 mass of $m_{K_L^0} = 0.497611\,\text{GeV}$ [Zyl+20].The distribution is shown in Fig. 2.7.
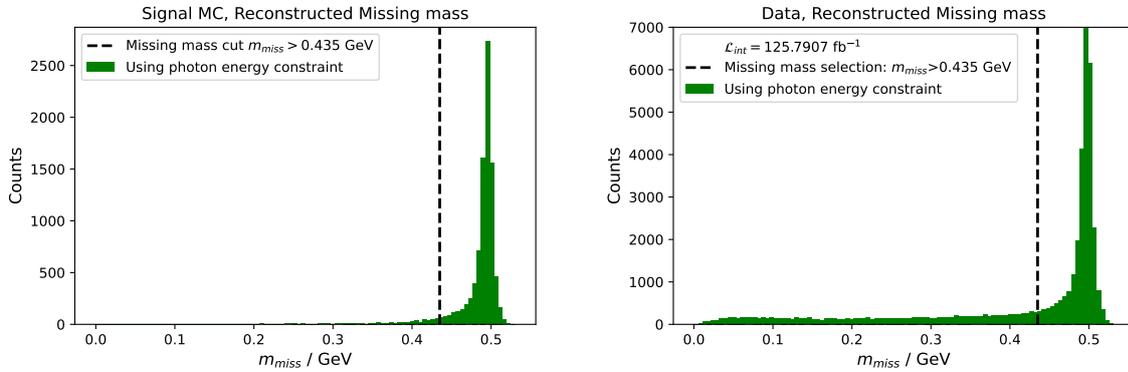
Figure 2.7: Distribution of the missing mass for signal MC (left) and data (right) after selections on σ_d, E_γ, $m_{\pi\pi}$ and using the photon energy constraint, in the m_{miss} calculation. A sharp peak can be observed at the nominal K_L^0 mass. To select good K_L^0 candidates we apply a cut at $m_{\text{miss}} > 0.435\,\text{GeV}$, by eye.

We observe a sharp peak close to the K_L^0 mass in signal MC and data, respectively. Thus, we are able to predict good K_L^0 candidates for each event and apply a selection, by eye, on the missing mass in signal MC and data at $m_{miss} > 0.435\,\text{GeV}$.

The reason why we need to use the photon energy constraint can be seen upon taking a closer look at the missing mass distribution without any constraint. This is shown in Fig. 2.8.
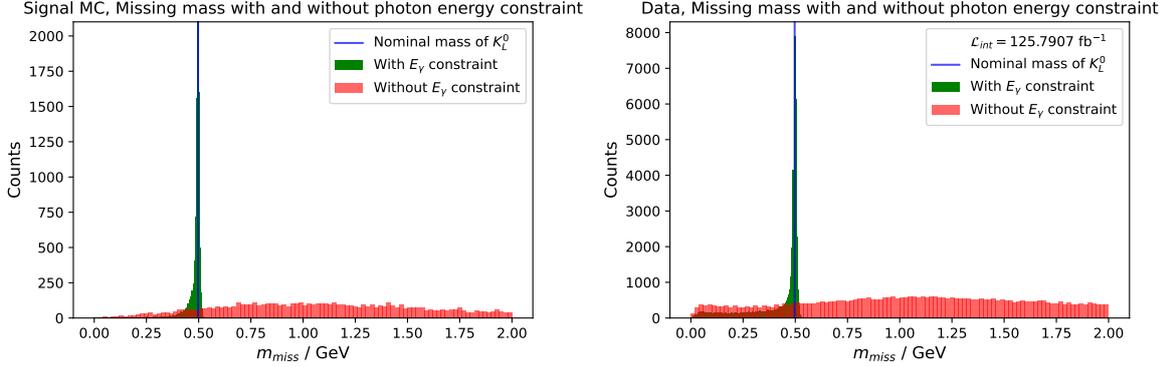
Figure 2.8: Comparison of the distribution of the missing mass with the photon energy constraint (green) and without the constraint (red) for signal MC truth matched (left) and data (right). Both distributions are calculated by Eq. (2.5). The blue line shows the nominal mass of the K_L^0.

We observe a broad distribution peaking at $\approx 1\,\text{GeV}$ without using the photon energy constraint. This would suggest that we are not able to identify good K_L^0 candidates based on the missing mass in our pure sample of K_S^0-γ_{ISR} candidates. With the application of the constraint a sharp peak at the K_L^0 mass is observed.

We perform a study of generator level signal MC to understand this broadening we observe upon not using the constraint. We calculate the missing mass by adding a small variation to the photon energy. We use a Gaussian with a mean of zero and a width of $\sigma = 0.1\,\text{GeV}$ to produce our variation, which is created randomly 10000 times and added to the photon energy. We calculated then our missing mass and illustrate the distribution in a histogram in Fig. 2.9.
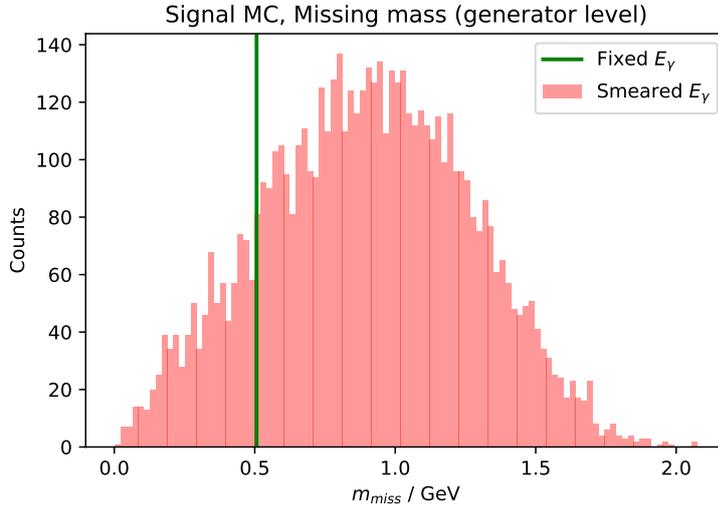
Figure 2.9: Distribution of m_{miss} after varying of the photon energy using a Gaussian distribution with mean $\mu = 0\,\text{GeV}$ and standard deviation $\sigma = 0.1\,\text{GeV}$ for generator level signal MC. A large broadening in m_{miss} is observed. We further plot mark the m_{miss} value, if we use the constraint for the photon energy (green).

If we compare this to the histograms in Fig. 2.8 we also observe a broadening in the missing mass spectrum peaking at around 1 GeV. However, a small difference can be seen. In Fig. 2.8 we obtain missing masses of up to 3 GeV in signal MC. In our simulation of generator level signal MC we observe an upper threshold of the missing mass at 2 GeV. This is very likely to come from the modeling of the variation of the photon energy, specifically the bad modeling of the low energy tail observed in the photon energy distribution in Fig. 2.4 by our Gaussian. Nevertheless, the modeling of the energy variation of the photon with a Gaussian function shows sufficient evidence that a small variation of a few 100 MeV results in a large uncertainty in the missing mass. Thus, we have enough evidence that a constraint to the photon energy is needed in order to identify K_L^0 candidates. We will use this constraint for all subsequent calculations that include the ISR photon four-momentum.

2.2.2 K_S^0 cut flow, step-wise and complete efficiency

We provide a brief summary of the selections applied so far. Thus, we want to display the cut flow of all selections and determine the step-wise and complete efficiency after each selection step. The efficiency will again be based on the two pion mass distribution. The step-wise and complete efficiency are determined from

$$\epsilon_{\text{Step}} = \frac{N_{\text{i}} + \frac{1}{2}}{N_{\text{i-1}} + 1} \text{ and} \tag{2.7}$$

$$\epsilon_{\text{Comp}} = \frac{N_{\text{i}} + \frac{1}{2}}{N_0 + 1}, \tag{2.8}$$

using the Bayesian approach [Cas12]. After each selection we count the number of remaining candidates N_{i} and divide it by the number of candidates remaining after the previous selection $N_{\text{i-1}}$ to determine the step-wise efficiency. For the complete efficiency divide the number of candidates after each selection by the number candidates before applying any selections N_0. Only candidates within the signal region defined by the two pion mass selection are counted. In data we use the background subtraction method in order to reject background candidates that are still left in the signal region. Therefore, we fit a polynomial of 1st order to our side regions of the two pion mass distribution. By an extrapolation inside our signal region we can determine the number of background candidates and reject them. Thus, we obtain for $N_0 = 106033 \pm 326$, where the error is calculated as the Poissonian error $\sqrt{N_0}$. The uncertainty for the other parameters is as well given by $\sqrt{N_i}$. The uncertainty of the step-wise and complete efficiency is calculated using the Bayesian approach [Cas12].

In Table 2.2 the step-wise and complete efficiencies, together with all selections performed so far, are shown for data and signal MC. The K_S^0 cut flow, for data, is illustrated in Fig. 2.10 showing the $m_{\pi\pi}$ distribution after each selection step. For truth matched signal MC the cut flow is given in Fig. 2.11.

Variable Name	Selection	Efficiency			
		signal MC		Data	
		Step: ϵ_{Step}	Complete: ϵ_{Comp}	Step: ϵ_{Step}	Complete: ϵ_{Comp}
Missing mass	$m_{\text{miss}} > 0.435\,\text{GeV}$	0.8399(25)	0.8399(25)	0.7880(22)	0.7880(22)
Signif. of Distance	$\sigma_d > 43.24$	0.9542(16)	0.8015(27)	0.9299(14)	0.7328(22)
Photon Energy	$E_\gamma > 2.61\,\text{GeV}$	0.9998(01)	0.8014(27)	0.9994(01)	0.7324(22)

Table 2.2: Step-wise and complete efficiency for each selection from signal MC and data.
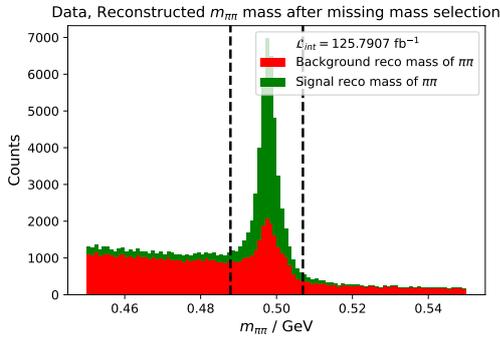
(a) Distribution of the two pion mass after missing mass selection.
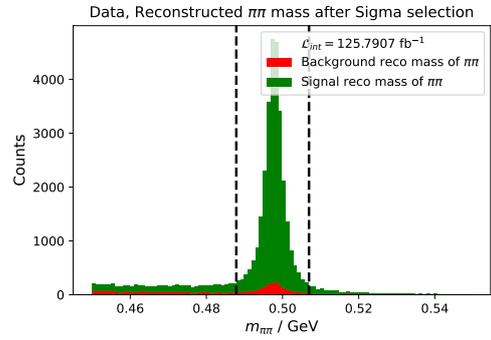
(b) Distribution of the two pion mass after significance of distance selection.
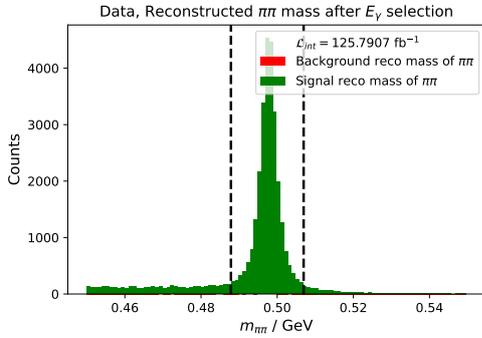
(c) Distribution of the two pion mass after photon energy selection.
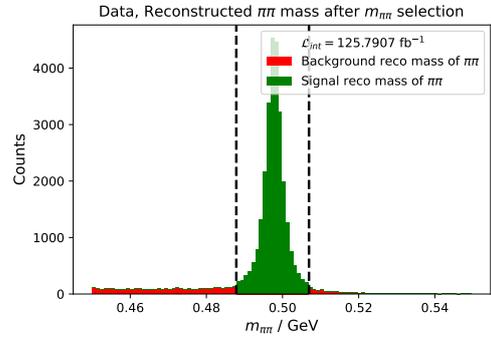
(d) Distribution of the two pion mass after two pion mass selection.

Figure 2.10: Distribution of two pion mass after each selection step from data.
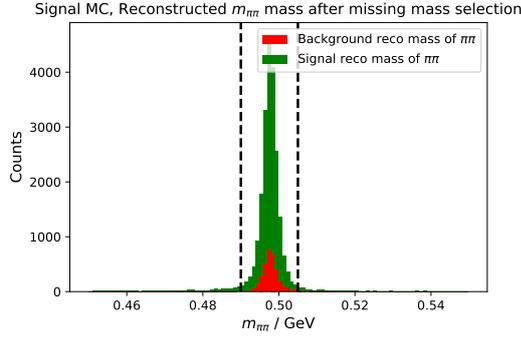
(a) Distribution of the two pion mass after missing mass selection.
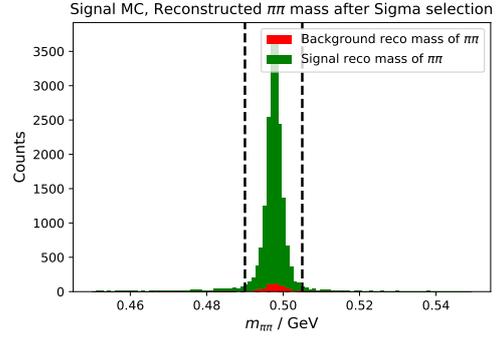
(b) Distribution of the two pion mass after significance of distance selection.
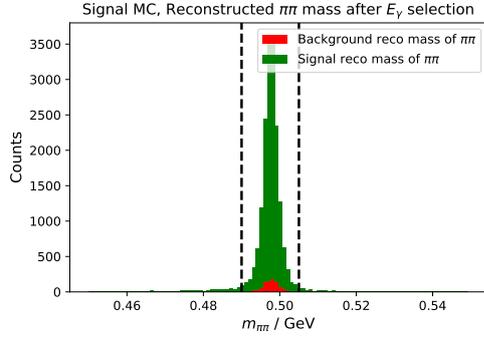
(c) Distribution of the two pion mass after photon energy selection.
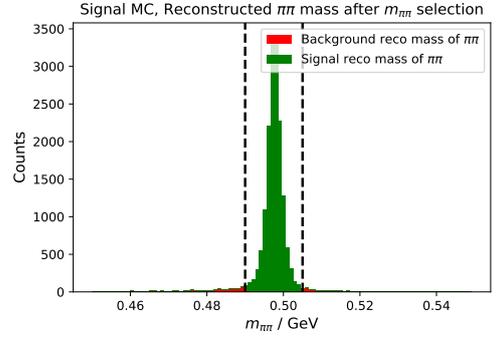
(d) Distribution of the two pion mass after two pion mass selection.

Figure 2.11: Distribution of two pion mass after each selection step from signal MC.

With each of the selections the step-wise efficiency increases, while the complete efficiency decreases, as expected. This can also be seen in Fig. 2.10 and Fig. 2.11 for data and signal MC, respectively. Ultimately, we achieve a complete efficiency in data of $\epsilon_{Comp}^{Data} = (73.24 \pm 0.22)\%$ and $\epsilon_{Comp}^{MC} = (80.14 \pm 0.27)\%$. One can see that the numbers between signal MC and data differ by $\approx 7\%$ from each other. This is caused by a small missmodeling in singal MC.

With this highly efficient and pure sample of K_L^0 candidates we next want to determine the expected number of K_L^0 events in data. Therefore, we simply count our signal candidates after all selections. Thus, we expect $N_{exp} = 26956 \pm 164$ K_L^0 events in the complete data set of $125.7907\,\mathrm{fb}^{-1}$. The error is calculated $\sqrt{N_{exp}}$.

2.2.3 Definition of Good K_L^0 cluster candidates

So far we found a highly purity and efficient sample of K_S^0-γ_{ISR} and thus also K_L^0 candidates, by introducing some simple selections on the $e^+e^- \rightarrow \gamma_{ISR}\left[\phi \rightarrow K_L^0 K_S^0\right]$ channel. However, we do not know which cluster is indeed a real K_L^0 cluster and which ones are fake. Thus, we have to define what a good K_L^0 cluster candidate in the KLM is. Here, we will allow now multiple candidates to occur in one event to select the best one. The definition of a good and bad K_L^0 cluster candidate is done by the

determination of the opening angle (α) between the missing momentum vector, from which we know that we select possible K_L^0 candidates, and the KLM cluster position, consisting of the x, y and z component of a cluster. To calculate the opening angle we have to do a Galilean transformation to get the missing momentum components in the lab frame as shown in Eqs. (2.9) to (2.12).

$$p_x^{lab} = p_x^{cms} \tag{2.9}$$

$$p_y^{lab} = p_y^{cms} \tag{2.10}$$

$$p_z^{lab} = \gamma(p_z^{cms} + \beta E^{cms}) \tag{2.11}$$

$$E^{lab} = \gamma(E^{cms} + \beta p_z^{cms}) \tag{2.12}$$

With these, we now use the **vector** python package to build vector arrays of our missing momentum and KLM position vector. Then the opening angle can be determined by the function **deltaangle** of this package[Esc+22]. The definition is illustrated in Fig. 2.12.
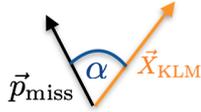
Figure 2.12: Definition of the opening angle α between the missing momentum and KLM cluster position vector.

We use the opening angle between those vectors since if we predict the momentum correctly we should observe a peak at one in $\cos(\alpha)$ for good K_L^0 cluster candidates, while bad K_L^0 cluster candidates not correlated to our signal should have a uniform distribution. The distributions is shown in Fig. 2.13 for truth matched MC.
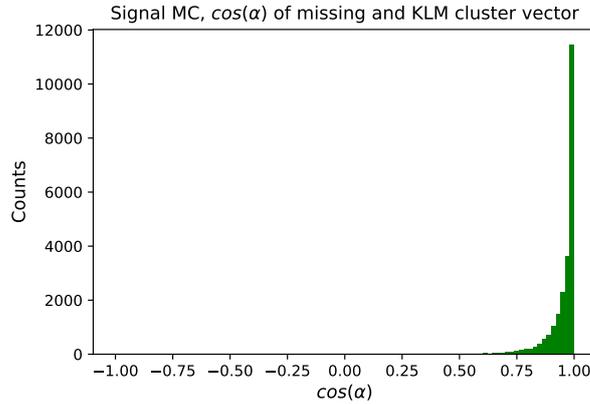
Figure 2.13: Distribution of the cosine of the opening angle $\cos(\alpha)$ between the missing momentum and KLM cluster position vector for signal MC (truth matched). A peak at one is observed with a long tail towards higher opening angles for truth matched clusters.

Here, we use truth-matched KLM clusters, which means we obtain the true values of the generated MC particles. The KLM truth-matching algorithm determines then the likelihood of an KLM cluster

to be generated by a generated K_L^0. We observe a sharp peak at one as expected. However, a tail to lower values of $\cos(\alpha)$ is visible. Thus, the distribution has a much higher width than expected for truth matched candidates. It suggests that there are truth matched vectors with opening angles up to $60°$ towards the truth matched KLM cluster. We investigate this further using event displays, to rule out calculation errors. In Fig. 2.14 we observe a K_L^0 matched to a cluster in the KLM with an opening angle of $\approx 60°$. Furthermore, one can see that this cluster (pink in Fig. 2.14) is likely to originate from an electron (blue dotted line) created through the decay of the K_L^0.
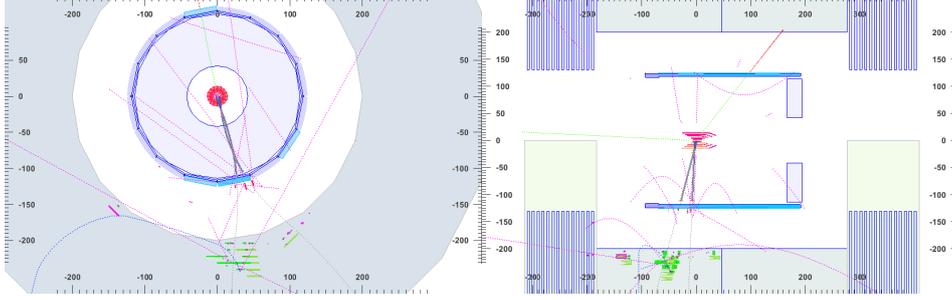
Figure 2.14: Picture of an event display of a K_L^0 matched cluster with an opening angle of $\approx 60°$ for the xy- (left) and xz-plane (right), respectively. Here an electron (blue dotted line) originates from K_L^0 decay and creates a cluster (pink) in the KLM, which is matched to the K_L^0.

These events, where the K_L^0 decays and the decay particles reach the KLM can explain a broadening in $\cos(\alpha)$. However, we also observe events where no MC particle is close to the K_L^0 matched cluster, as shown in Fig. 2.15.
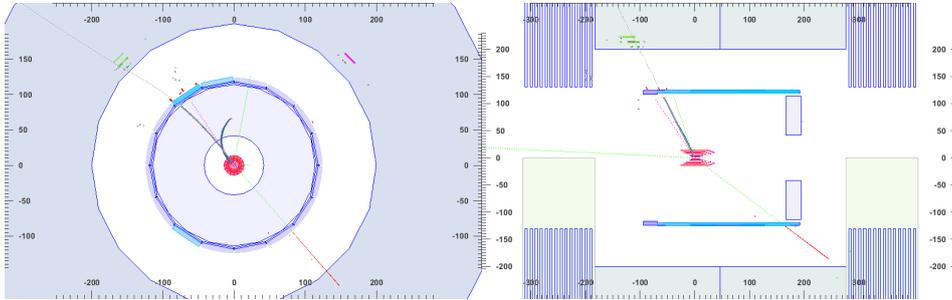
Figure 2.15: Event display of a K_L^0 matched cluster with an opening angle of $\approx 70°$ for the xy- (left) and xz-plane (right), respectively. Here no particle is close to the matched K_L^0 cluster cluster (pink).

One explanation could be that these clusters are created by beam background. Another, explanation could be that we observe hadronic splitoffs of the K_L^0. This means that the K_L^0 will interact with any detector material producing a hadronic shower. Thus, high energy neutrons could be produced and detected as clusters in the KLM with high opening angles towards the original K_L^0. Last but not least, it it could that the truth-matching algorithm is failing for KLM clusters. We investigate this more closely by using the particle gun, generating only one K_L^0 per event uniformly distributed in momentum, θ and ϕ in the detector. Thereby, we do not include any background files in the generation. Still, we observe the same behaviour as before. As an example, in Fig. 2.16 an event is shown where the generated K_L^0 should

not be match with the KLM cluster, referring to the opening angle between both.
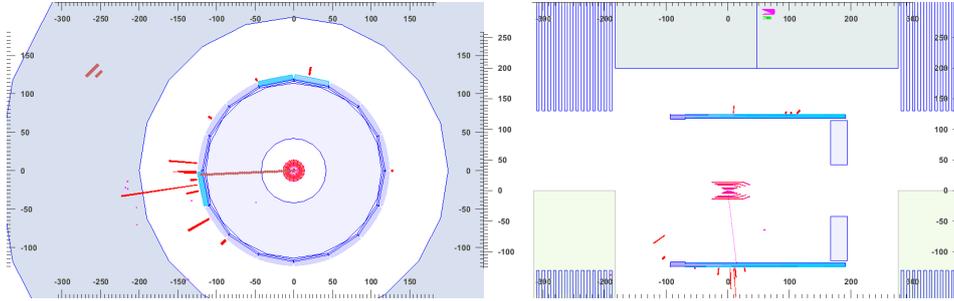
Figure 2.16: Event display of a K_L^0 matched cluster with an opening angle of $\propto 70°$ for the xy- (left) and xz-plane (right), respectively. Here no particle is close to the cluster (pink). The K_L^0 was generated by the particle gun module.

Thus, either we observe a hadronic splitoff of the K_L^0 or the KLM truth-matching algorithm is not working.

In addition to that we observe events where a truth matched cluster is associated to a charged track far away from the cluster. This however, is a known problem as shown in [PO19]. Here it was reported that a track is matched to a cluster, if the distance between both is below 150 cm. Considering that the Belle II detector measures about 8 m in length and height and that the KLM is the most outer and biggest detector component it is likely to find a track within 150 cm distance of a cluster. Thus, the definition of neutral clusters is not suitable and actual neutral clusters are considered charged.

We see that the truth matching seems to fail for some clusters and the track matching algorithm is not working properly as well. Both topics deserve a much more detailed study to identify the reason for this and improve the algorithms. However, this would exceed the scope of this study as we only use signal MC for testing purposes. Nevertheless, the distribution in Fig. 2.13 meets our expectations that we are able to find K_L^0 candidates pointing towards a KLM cluster. Thus, we define good and bad K_L^0 cluster candidates by applying a simple selection.

Since, we want to do a data-driven Fast-BDT training on KLM clusters we repeat the calculation of the cosine of the opening angle between the missing momentum and the KLM cluster position vector in data of 125.7907 fb^{-1}. The distribution is shown in Fig. 2.17.
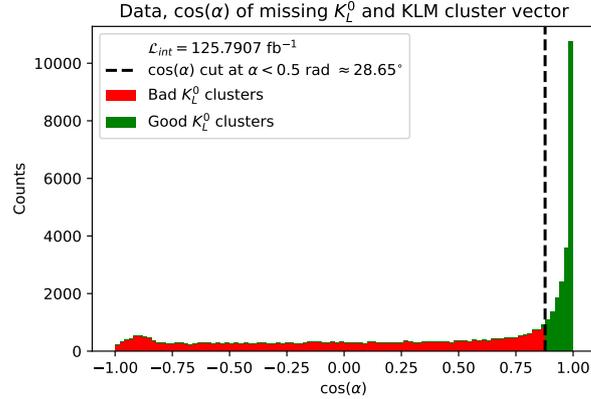
Figure 2.17: Distribution of the cosine of the opening angle $\cos(\alpha)$ for the missing momentum and KLM cluster position vector for $L_{\text{int}} = 125.7907\,\text{fb}^{-1}$ of data. A peak at one is observed. We define good K_L^0 clusters (green) to have an opening angle $\alpha > 0.5\,\text{rad} \rightarrow \cos(\alpha) > 0.8778$). Every candidate below that is considered to be background and thus not a K_L^0 cluster candidate (red).

We observe, as in signal MC, a peak at one with a long tail towards higher opening angles of α. In contrast to signal MC the higher width as well as the long tail is to be expected due to the background in data. With this plot we can define whether a KLM cluster shall be considered to be a good or bad K_L^0 cluster candidate by accounting each KLM candidate with an opening angle of less then $0.5\,\text{rad} \propto 28.65^{\circ}$[3] to be a good K_L^0 cluster candidate, shown in green. The rest is considered to be bad K_L^0 cluster candidates (red).

If one examines Fig. 2.17 closer a small peak at roughly $\cos(\alpha) \propto -0.9$ is visible. An error in the calculation could not be verified after cross checking it with different vector packages of python and pyroot and ROOT. We can not give a sufficient explanation why there seems to be a significant amount of clusters with opening angles of 180° towards the predicted K_L^0 flight direction.

As shown in [Pii21], beam induced fast neutrons are likely to hit in the forward endcap. Thus, we take a closer look on this region for our defined bad K_L^0 cluster candidates (red in Fig. 2.17). In Fig. 2.18 we can see the $x - y$ plane of the forward KLM endcap for bad K_L^0 cluster candidates.

---

[3] This means $\cos(\alpha) > 0.8776$.
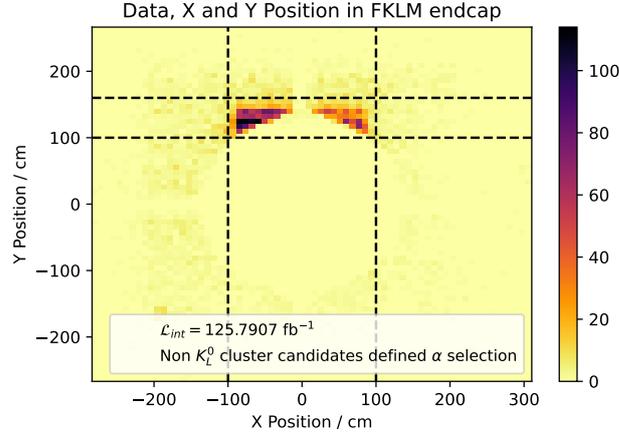
Figure 2.18: Two dimensional distribution of the bad K_L^0 cluster candidates of their X and Y position in the forward endcap of the KLM using $L_{\text{int}} = 125.7907\,\text{fb}^{-1}$ of data. We observe a significant amount of background clusters in the same endcap region as in [Pii21]. We reject these clusters since we already now that they are most likely background clusters created by beam induced fast neutrons [Pii21].

We observe that hits of bad K_L^0 cluster candidates occur in the same region as shown in [Pii21], giving strong evidence that those clusters are most likely beam induced fast neutrons. Thus, we reject the candidates in this region since we already know that these are fake clusters.
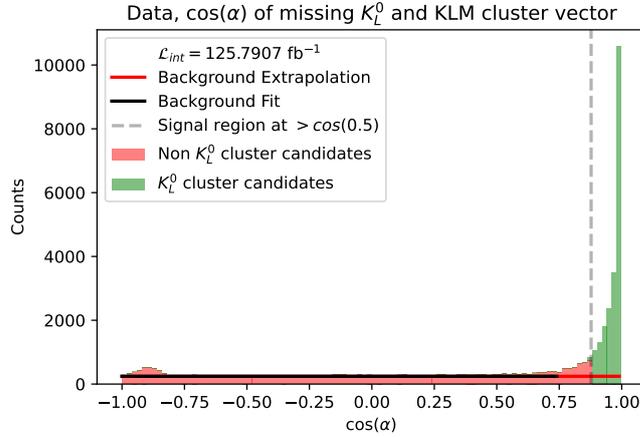
Figure 2.19: Distribution of the cosine of the opening angle $\cos(\alpha)$ between the missing momentum and KLM cluster position vector after rejecting forward KLM fast neutrons, for $L_{int} = 125.7907\,\text{fb}^{-1}$ of data. In order to determine the identification efficiency and the fake-rate we fit a constant function to our background (black line) from -1 to 0.75 and extrapolate it to higher values of $\cos(\alpha)$. By then taking the integral of our total signal region and subtract the integral of the extrapolated background we determine the efficiency and fake-rate.

By rejecting these fake clusters and our definition of good and bad K_L^0 cluster candidates we calculate the identification efficiency and fake-rate. Therefore, we fit a constant function to the background region and extrapolate it in the signal region, illustrated in Fig. 2.19. Thus, we can describe the background

candidates that pass our selections. If the missing vector has a very small opening angle towards the cluster position it becomes more unlikely that the cluster is a fake one. Therefore, we assume that the background decreases around $\cos(15°) \approx 0.966$. We choose for $\alpha \approx 15°$ since this is the opening angle needed towards charged tracks to count the cluster as a neutral one (c.f. [Abe+10]). Then we have to calculate our number of signal candidates, corrected by this falloff in the extrapolated background ($N_{\text{Sig}}^{\text{Cor}}$) to be

$$N_{\text{Sig}}^{\text{Cor}} = N_{\text{Sig}} - N_{\text{Ext}} + N_{\text{Ext}}^{15°} \cdot \epsilon. \tag{2.13}$$

Here, N_{Sig} is the amount of candidates passing our good and bad K_L^0 cluster candidate definition. N_{Ext} is the number of assumed background candidates by extrapolating the constant function into the signal region. We can calculate it to be

$$N_{\text{Ext}} = N_{\text{Sig}} \cdot \frac{I_{\text{Ext}}}{I_{\text{Sig}}^{\text{Total}}}, \tag{2.14}$$

where I_{Ext} is the integral of the extrapolated background function and $I_{\text{Sig}}^{\text{Total}}$ the total integral of the histogram within the signal region. The last term in Eq. (2.13) is $N_{\text{Ext}}^{15°} \cdot \epsilon$. ϵ is the identification efficiency and $N_{\text{Ext}}^{15°}$ is the number of estimated background candidates where the missing and KLM position vector has a smaller opening angle than $15°$. We can calculated it by

$$N_{\text{Ext}}^{15°} = N_{\text{Sig}}^{15°} \cdot \frac{I_{\text{Ext}}^{15°}}{I_{\text{All}}^{15°}}, \tag{2.15}$$

where $N_{\text{Sig}}^{15°}$ is the number of candidates within the $15°$ region and $I_{\text{Ext}}^{15°}$ and $I_{\text{All}}^{15°}$ are the integrals of the extrapolated background function and the histogram in that region, respectively. If we then have calculated $N_{\text{Sig}}^{\text{Cor}}$ by Eq. (2.13) we can also determine the identification efficiency by

$$\epsilon = \frac{N_{\text{Sig}}^{\text{Cor}}}{N_{\text{exp}}}. \tag{2.16}$$

Here N_{exp} is the number we determined at the end of Section 2.2.2. If we insert Eq. (2.13) in Eq. (2.16) and transform the equation we get

$$\epsilon = \frac{N_{\text{Sig}} - N_{\text{Ext}}}{N_{\text{exp}} \cdot \left(1 - \frac{N_{\text{Ext}}^{15°}}{N_{\text{exp}}}\right)}. \tag{2.17}$$

For the fake-rate (f) we have to determine

$$f = \frac{N_{\text{Back}}}{N_{\text{Exp}}}. \tag{2.18}$$

Here, N_{Back} is the total amount of background candidates and is determined by

$$N_{\text{Back}} = N_{\text{out}} + N_{\text{Ext}} - N_{\text{Ext}}^{15°} \cdot \epsilon. \tag{2.19}$$

N_{out} is the number of bad K_L^0 cluster candidates. Inserting this in Eq. (2.18) we get

$$f = \frac{N_{\text{out}} + N_{\text{Ext}} - N_{\text{Ext}}^{15^\circ} \cdot \epsilon}{N_{\text{exp}}}. \tag{2.20}$$

If we calculate now the identification efficiency and fake-rate we get

$$\epsilon = 0.7325 \pm 0.0067 \text{ and}$$
$$f = 1.0353 \pm 0.0087.$$

The statistical uncertainty for the fake-rate and efficiency is calculated by the Gaussian error propagation. We see that we achieve a rather high identification efficiency with roughly 73%. However, we also have a rather high fake-rate of 1. This means that we there is on average at least one fake cluster in an event that is falsely identified as a K_L^0 cluster. We aspire to reduce this fake rate, but keeping a high identification efficiency. Thus, we perform on this final sample of bad and good K_L^0 cluster candidates a MVA study in the following.

2.3 MVA training for K_L^0 identification

As we saw in the previous section we are able to define good and bad K_L^0 cluster candidates. Thus we could determine the identification efficiency and fake-rate, which showed that we can identify K_L^0 clusters with a high efficiency but also high fake-rate. In order to decrease the fake-rate, but without reducing the efficiency too much, we train a FastBDT on our data sample. Therefore, we use our selection on $\cos(\alpha)$ and create the target variable to separate good and bad clusters, which will be one for $\cos(\alpha) > \cos(0.5)$ and zero for $\cos(\alpha) < \cos(0.5)$. We split our final sample of good and bad K_L^0 cluster candidates in a training and testing file. We assign 80% of our total candidates to the training and the remaining 20% to the testing file. As training variables we use all KLM shape variables that are accessible in the Belle II framework (c.f. [Col22]). In the end to simplify the training for the FastBDT we neglect KLM variables with low importance and high correlations. Thus we train our final FastBDT on 11 variables, which are listed in Table 2.3. The variables are listed from least to most important.

Variable Name	Description
Cluster X position	Returns the x coordinate of the associated KLM cluster.
Inner most Layer	Gives the innermost number of layer where a 2-dimensional hit was found.
Theta	Gives the polar angle (θ) of the associated KLM cluster.
Maximal KLM Angle	Gives the maximum angle between all KLM clusters and the particle we are looking at, in an event.
ECL cluster matches to KLM cluster	Gives the number of ECL clusters that are matched to the particle which is associated to the KLM cluster.
Cluster timing	Gives the timing information of the KLM cluster.
Belle ECL Flag	Returns the Belle style ECL flag.
Track-Cluster distance	Gives the distance between the KLM cluster and the track It will return NAN if no Trakc-to-KLM cluster relation is found.
Number of Layers	Gives the number of KLM layers with a 2-dimensional hits.
K_L^0-ID	Gives the KlId classifier output that is related to the KLM cluster.
Number of tracks matched to KLM Cluster	Gives the number of Track matches to the KLM cluster. It can be greater than zero for K_L^0's or neutrons. It returns NAN if the particle is not related to any KLM cluster.

Table 2.3: KLM shape variables and their descriptions to train the BDT on [Col22].

In Fig. 2.20 the classifier output of the FastBDT is shown. As one can see we are able to separate fake and good K_L^0 clusters by those 11 variables shown in Table 2.3.
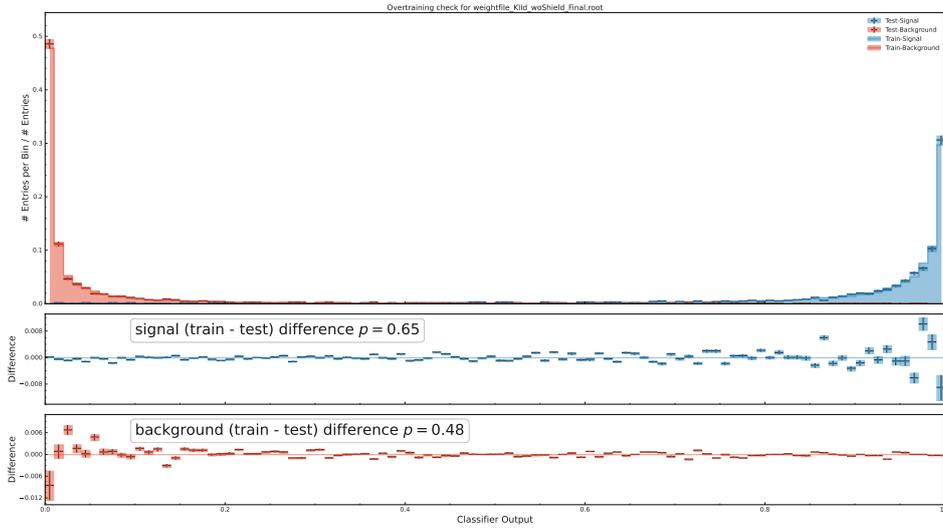
Figure 2.20: Distribution of the FastBDT classifier output. A separation of signal and background is observed. We see an agreement between the training and testing sample for signal of 65% and for background 48%.

We add the classifier output to our total sample of good and bad K_L^0 clusters. Thus, we can scan through the output classifier and calculate for each step the identification efficiency and fake-rate, as shown at the end of Section 2.2.3. We expect, by scanning through the FastBDT output classifier, first a strong decrease in the fake-rate, while the identification efficiency is only slowly decreasing. At high selection values for the classifier however, this will change and we will observe a major efficiency loss, while the fake-rate does not improve by much. The identification efficiency against the fake-rate is shown in Fig. 2.21.
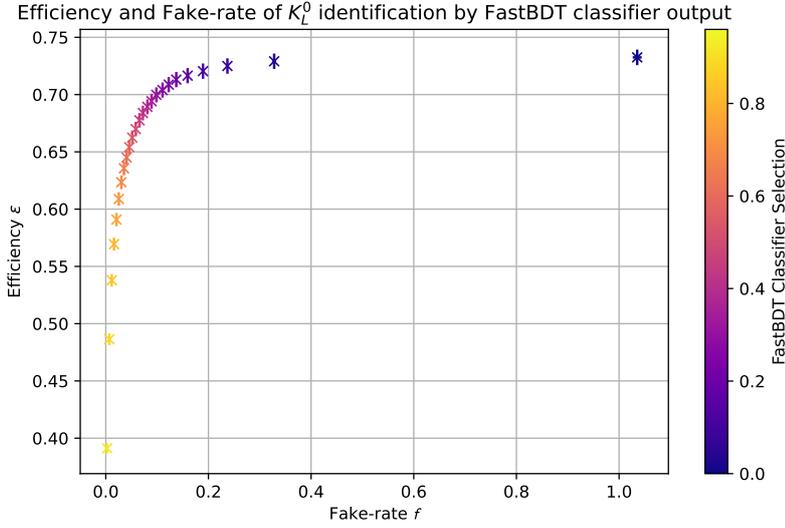
Figure 2.21: Identification efficiency plotted against the fake-rate for different selections on the FastBDT classifier output. We observe a strong decrease in the fake-rate for low cuts on the FastBDT classifier. For high selection values the efficiency decreases strongly, while the fake rate decreases only slowly.

The errors in Fig. 2.21, are calculated by the Gaussian error propagation. As we expected we observe a point, close to ($\epsilon = 1$, $f = 0$), at

$$\epsilon = 0.6998 \pm 0.0065 \text{ and}$$
$$f = 0.0985 \pm 0.0020.$$

We will call this point the optimal working-point. Furthermore, by finding this point we showed that by training a simple FastBDT on data a better identification of K_L^0 clusters in the KLM detector is achieved. If we compare our identification efficiency to the one from Belle, of roughly 50% [Tch+02], we see that in Belle II an enhancement in efficiency is achieved. Furthermore, we see good agreement with previous studies of Belle II, for example [Aus+15]. Here an identification efficiency between 60% and 70% was determined for K_L^0's with momenta between 2 GeV and 3 GeV in the endcaps of the KLM detector. This suggests that we are only able to observe high momentum K_L^0's. However, we also aim to identify low momentum K_L^0 candidates. Therefore, we search in other decay channels for low momentum K_L^0's.

We should be able to observe low momentum K_L^0 candidates in decays of $e^+e^- \to \gamma_{\text{ISR}}\left[J/\psi \to \pi^+\pi^- K_L^0 K_S^0\right]$. We discussed in Section 1.4.2 that in order to produce a J/ψ, using a center of mass energy of $\sqrt{s} = 10.58\,\text{GeV}$, we need a ISR photon of around $E_\gamma^{J/\psi} = 4.83674\,\text{GeV}$ and to distinguish the two pions from the J/ψ and K_S^0 decay we introduce a best K_S^0 selection. We use the same analysis procedure as before, resulting in no evidence for good K_L^0 cluster candidates. Thus, we are not able to use the J/ψ decay mode (described in Section 1.4.2) to identify low momentum K_L^0's.

Another approach to identify low momentum K_L^0's could be using MC signal events for the training. Thus, we investigate if we are able to use signal MC with low momentum K_L^0 candidates and train the FastBDT on MC signal and data background events, simultaneously. For this however, the training

variables (KLM shape variables) have to match in signal MC and data. A short validation for this can be done if we compare the signal MC and data KLM shape variables. In Fig. 2.22 the four most important variables are displayed for the FastBDT training.
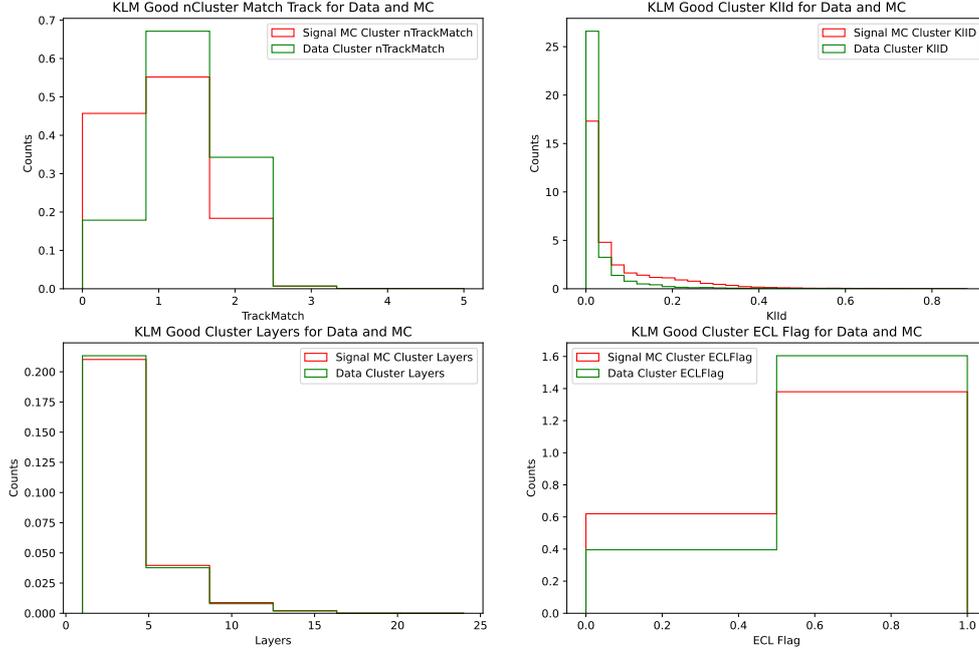
Figure 2.22: Comparison of the signal distributions of the KLM shape variables for signal MC and data in order to validate the assumption that we can do a combined training to predict low momentum K_L^0's. Large discrepancies between signal MC and data are observed.

As one can see, there are major discrepancies between important variables[4] for the training. The other shape variables are illustrated in Fig. A.8 and show similar discrepancies between data and signal MC. Thus, a combined training to predict low momentum K_L^0's would be possible. However, due to the discrepancies a reduction in the final efficiency of K_L^0 cluster identification could be seen, if we apply this FastBDT only on data.

In total we have now developed a way to select good K_L^0 candidates, with high energy, in an event. Low energy K_L^0's could be found in different decay modes like the $e^+ e^- \rightarrow \gamma_{\mathrm{ISR}} J/\psi$. However, here the background suppression is difficult to achieve. Thus we can not report of any good K_L^0 cluster candidates in the J/ψ mode. Nevertheless, a combined training of MC signal events and data background events is possible, but expected to be not efficient. Thus, we we first aim to use our identification done in the ϕ-mode and test it in a missing energy analysis in order to prove, if it is already efficient and usable. We aim to test our identification algorithm in an analysis of $B \rightarrow D l \nu$, with $l \in \{e, \mu\}$ in the following.

---

[4] Identified in the pure data-driven FastBDT training.

K_L^0 identification in $B \rightarrow Dl\nu$ decays

We aim to test our K_L^0 identification as described in Chapter 2 using the $B \rightarrow \left[D \rightarrow \pi K_L^0 \right] l\nu$ decay. Further, we use the $B \rightarrow \left[D \rightarrow \pi K_S^0 \right] l\nu$ decay as a normalization mode to determine the K_L^0 identification efficiency. This is possible because the branching ratios of those two decays only differ by 0.102% [Zyl+20]. Thus, we determine the identification efficiency by

$$\epsilon = \frac{N_{K_L}}{N_{K_S}^{Norm}}. \tag{3.1}$$

Here, $N_{K_S}^{\mathrm{Norm}}$ is the number of identified K_S^0 events and N_{K_L} is the number of observed K_L^0 clusters. By this test we could verify the usability in other missing energy decays, like $B \rightarrow \pi\tau\nu$ decays discussed in Section 1.1.2.

For the identification of both K_S^0 and K_L^0 events we use the hadronic tagging algorithm of the FEI, which will be briefly discussed in the following. We first start with the identification of the K_S^0 channel using run dependent mixed MC14 (MC14rd), of $L_{\mathrm{int}} = 213.3848\,\mathrm{fb}^{-1}$, in order to define tag selections. Later on we add charged and continuum run dependent MC14 as background sources. In order not to train the FEI itself we use skimmed[1] MC14rd files on the **Grid**.

3.1 Full Event Interpretation

A short introduction to the Full Event Interpretation shall be given. The Full Event Interpretation (FEI) is an important algorithm used for example in missing energy studies or in the measurement of the $|V_{ub}|$ CKM-matrix element. The FEI itself knows a variety of different decay channels. Thus, the FEI is trained via multiple multivariate classifiers (MVCs)[2] to identify each decay channel [Kec14]. In Fig. 3.1 the reconstruction algorithm is shown graphically.

---

[1] Files where the FEI was already trained on and applied.
[2] These are multiple BDTs.
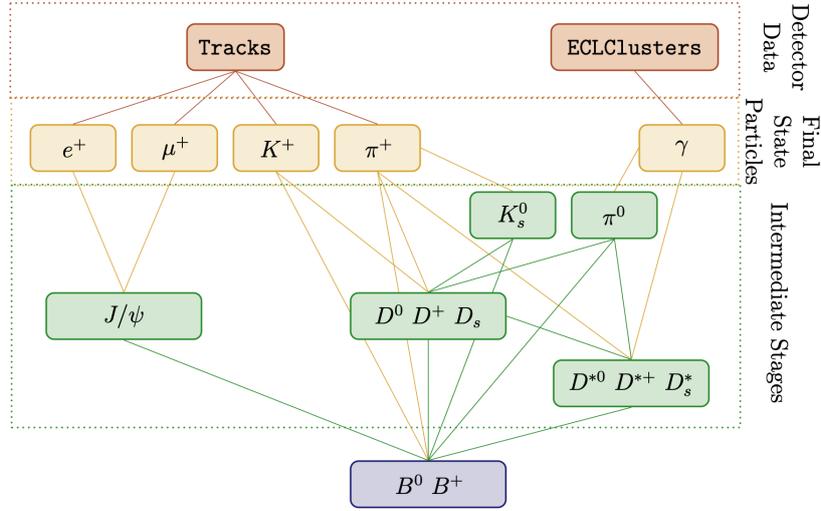
Figure 3.1: The FEI algorithm graphically visualised. From final-state particles we reconstruct intermediate particle candidates. These are then trained to build a pair of B-mesons. By using MVCs a probability P_{sig}, if the particle is reconstructed correctly, is determined. [Kec14]

Thus, in a $\Upsilon(4S) \to B^+ B^-$ event[3] the FEI algorithm reconstructs one B-meson, called B_{tag} using multiple MVCs. The remaining B-meson, called B_{sig}, underlies specific conditions, since the initial state of both is a $\Upsilon(4S)$ [Kec14]. Further, we get a probability P_{sig} if the B_{tag} is reconstructed correctly
There are different approaches for reconstructing the tag side called hadronic, semileptonic and inclusive tagging, which are described in the following.

- Hadronic tagging- Only uses hadronic decays to reconstruct the tag side. Thus, we achieve a good highly pure tag side with known kinematics. However, the tag efficiency is very low. [Kec14]

- Semileptonic tagging - Here we use semileptonic B decays for reconstructing the tag side. The tagging efficiency is therefore higher than in hadronic tagging, due to the higher branching fractions. Instead, the kinematics of the tag side are not reliable, since it suffers from missing energy from neutrinos. Thus, the tag side purity is low. [Kec14]

- Inclusive tagging - In inclusive tagging the four momenta of the rest of event particles of the signal side are used for the tag side reconstruction. The tagging efficiency is therefore higher than for the others. But the purity is rather low since there is also a lot of background, which comes from the high amount of background that has to be dealt with. [Kec14]

The decay on which the analysis will be performed is on the signal side. In Fig. 3.2 an example of an $B^- B^+$ decay is illustrated using the FEI to reconstruct the tag side.

---

[3] This is also valid for $\Upsilon(4S) \to B\bar{B}$. We choose $B^+ B^-$ only as an example.
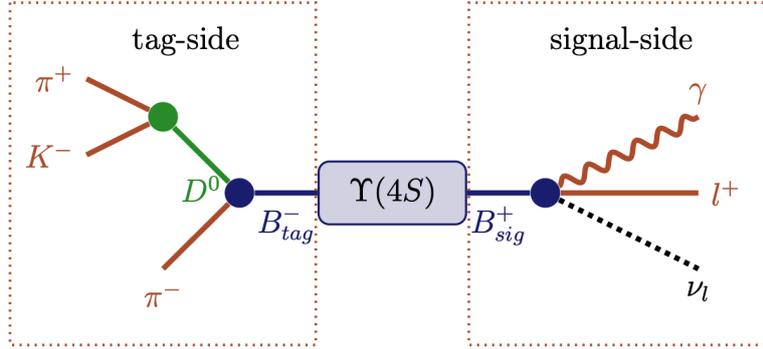
Figure 3.2: Illustration of a $\Upsilon(4S) \to B^- B^+$ decay. The tag side is reconstructed using the hadronic tagging algorithem [Kec14].

3.2 $D \to \pi K_S^0$ identification

The $B \to \left[D \to \pi K_S^0 \right] l\nu$ decay is as easy to identify as the ISR decay channel used in Chapter 2. We define some pre-selections for the leptons and pions, by requiring that the electron, muon and pion ID is greater than 0.9. Since we reconstruct the K_S^0 from two charged pions we introduce a best K_S^0 selection as in our J/ψ-study. Furthermore, a best candidate selection for the B_{tag} based on the signal probability of the same is applied. Thus, we assure that only one K_S^0 and B_{tag} candidate is considered. Thereby, we do not count multiple candidates for these, for example by switching pions in the K_S^0 reconstruction. We then reconstruct the D-meson from the K_S^0 and the remaining pion. The D-meson is finally combined with the lepton to create the signal side B-meson, without considering the missing neutrino. With the B_{tag} and B_{sig} we reconstruct the $\Upsilon(4S)$.

The advantage of using the tagging algorithm is that we know the kinematics of the B_{tag} and thus also of the B_{sig}. The only unknown particle we have no information on yet is the neutrino. Since we use this decay mode as a normalization for the K_L^0 mode we again need to obtain a highly pure sample of good K_S^0 candidates. We achieve this by introducing selections to the tag side. Further, for the signal side we need to select a pure sample of D-mesons decaying into a pion and K_S^0. Doing so, we can determine the number of $B \to \left[D \to \pi K_S^0 \right] l\nu$ events in our sample. However, in order to obtain a pure sample of D-mesons we first need to select a highly pure sample of K_S^0 candidates. Since, the background already reduces by selections on the B_{tag} we take a closer look on those in the following subsections.

3.2.1 B_{tag} selections

If the tagging algorithm works properly, no particle should be left in the rest of event, as explained in Section 3.1. Thus, we require that no charged tracks should be left in the rest of event. This is commonly referred to as the **completeness constraint** [Kec+19]. Furthermore, in MC we distinguish signal from background by demanding that everything is reconstructed correctly accept for a missing neutrino. This variable tells us, if a decay is reconstructed correctly except for some missing energy of a neutrino. Thus, we apply selections to the tag side, so that the signal is dominating the background. Therefore, we use three tagging variables, which are introduced in the following.

Beam-constraint mass (M_{bc})

The first of these variables is the beam-constraint mass (M_{bc}), which is defined as

$$M_{bc} = \sqrt{E_{\text{Beam}}^{*2} - \vec{p}_B^{*2}}. \tag{3.2}$$

Here, E_{Beam}^{*2} is the energy of the beam, while $\vec{p}_B^{*2}$ is the three momentum of the B-meson, in the center of mass frame. If the reconstruction is correct and if there is no neutrino in the decay the beam-constraint mass is equal to the B-meson mass [Kou+18]. Thus, we expect to observe a peak around $M_{bc} = 5.279\,\text{GeV}$ [Zyl+20]. In Fig. 3.3 the distribution of the beam-constraint mass for hadronically tagged run dependent mixed MC14 is shown.
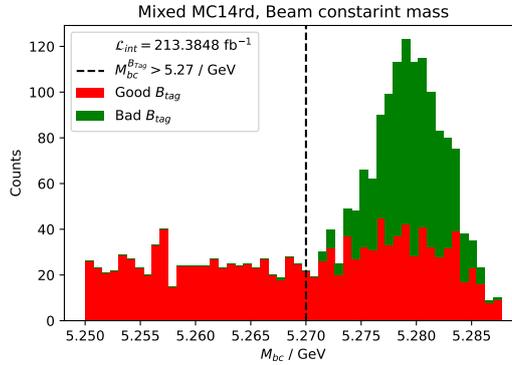

Figure 3.3: Distribution of the beam-constrained mass of hadronically tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$. We observe a peak at the nominal B mass as expected. The selection is optimised by eye, selecting events with $M_{bc} > 5.27\,\text{GeV}$.

We observe that our signal is peaking around the mass of the B-meson and thus matches our expectations. We apply a selection, seen above, at $M_{bc} > 5.27\,\text{GeV}$. Thus, we reject a great amount of wrongly reconstructed B_{tag} candidates.

Energy difference ΔE

In the center of mass frame the energy of B-mesons should be $E_B^* = \sqrt{s}/2$ and thus equivalent to the beam energy E_{Beam}^* [Kou+18]. Therefore, the difference in energy is defined as

$$\Delta E = E_{Beam}^* - E_B^*. \tag{3.3}$$

By this definition we expect ΔE to peak at zero, if there is no neutrino in the decay. In our reconstruction we use a pre-selection for this variable of $|\Delta E| < 0.1\,\text{GeV}$. In Fig. 3.4 is shown.
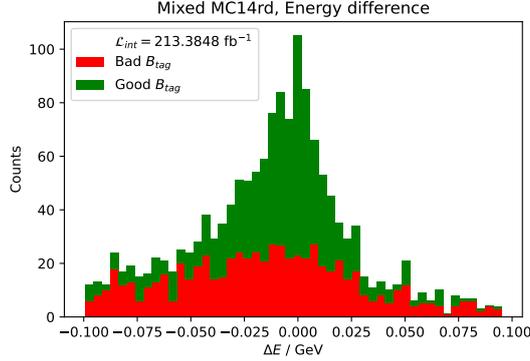
Figure 3.4: Distribution of the energy difference of the B-meson and the beam for hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$. A peak at zero for signal events is observed.

A peak at zero is observed, validating our expectation. Since we already applied a tight pre-selection, we do not tighten it further. We keep the pre-selection of $|\Delta E| < 0.1\,\text{GeV}$.

FEI signal probability P_{sig}

Last but not least, we take a closer look at the FEI signal probability, which we use as a ranking for the best B_{tag} selection, as mentioned previously. The variable tells us the probability that the B_{tag}-meson has been reconstructed correctly. Thus, we illustrate the distribution in Fig. 3.5. We plot the signal probability with a logarithm of basis ten.
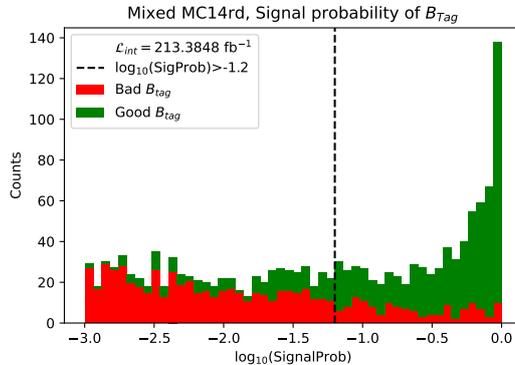
Figure 3.5: Distribution of the signal probability to reconstruct the B_{tag}-meson correctly of hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$. Here shown as $\log_{10}(P_{\text{sig}})$. A peak at zero is observed. We select events with $log_{10}(P_{\text{sig}}) > -1.2 \to P_{\text{sig}} \approx 0.06$. The selection is applied by eye.

We observe a good separation between signal and background candidates. As expected, signal candidates peak at zero, which corresponds to a probability of one. Thus we introduce a cut at -1.2 corresponding to $P_{\text{sig}} \approx 0.06$ and reject everything below. We chose this selection by eye. With those selections we now can search for good D- and K_S^0-mesons.

3.2.2 K_S^0 identification

In order to obtain a highly pure D-meson selection we need to select K_S^0 candidates first. To validate that our reconstruction, done so far, is working, we plot the two-pion mass distribution, expecting a peak at the K_S^0 mass. Therefore, we apply the previously introduced selections as well the best K_S^0 selection, in order to use the best match of pions for the reconstruction. Since we have to define a signal region for good K_S^0 candidates we fit a Gaussian in addition with a constant function for the signal and background, respectively. The fit function is shown in Eq. (3.4).

$$h(x) = \frac{A}{\sqrt{2\pi\sigma}} * \exp - \left(\frac{(x - \mu)}{\sigma} \right)^2 + a \tag{3.4}$$

The plot is shown in Fig. 3.6.
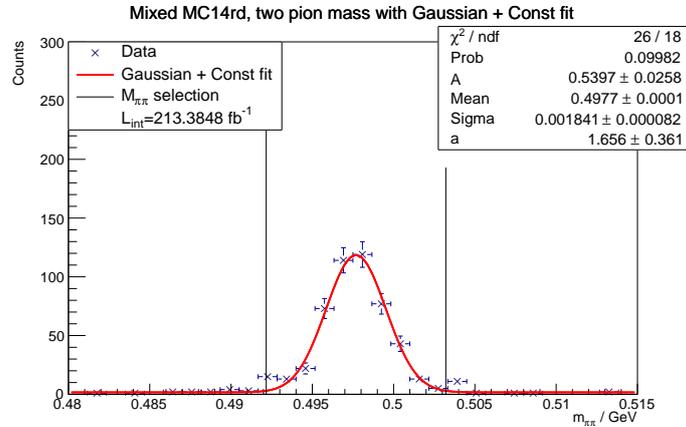
Figure 3.6: Distribution of the reconstructed mass of the two pions for hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$. We perform a global fit of a Gaussian plus a constant function Eq. (3.4). The fit result is shown on the top right. A peak at the nominal K_S^0 mass is observed. We define our signal region as the mean plus/minus three times the width of the Gaussian.

We observe a peak of the reconstructed mass of the two pions at the nominal K_S^0 mass. We define our signal region for good K_S^0 candidates by the mean plus/minus three times the width of the signal Gaussian, in our global fit. We, expect good K_S^0 candidates to occur within $0.492\,\text{GeV} < m_{\pi\pi} < 0.503\,\text{GeV}$.
We now enhance the purity of our K_S^0 sample. For this, we take a look at the significance of distance (σ_d), as done before in the ϕ-analysis. We expect the same behaviour as in Fig. 2.3 and define a selection by using the minimizing algorithm introduced in Section 2.1.1. In Fig. A.10 the result of the minimization and the distribution is shown. We define a cut at $\sigma_d^{K_S^0} > 7.09375$.
Applying this selection to our K_S^0 sample we can determine its purity and efficiency by fitting again Eq. (3.4), for signal and background respectively, to the reconstructed two pion mass distribution. The fit is shown in Fig. 3.7.
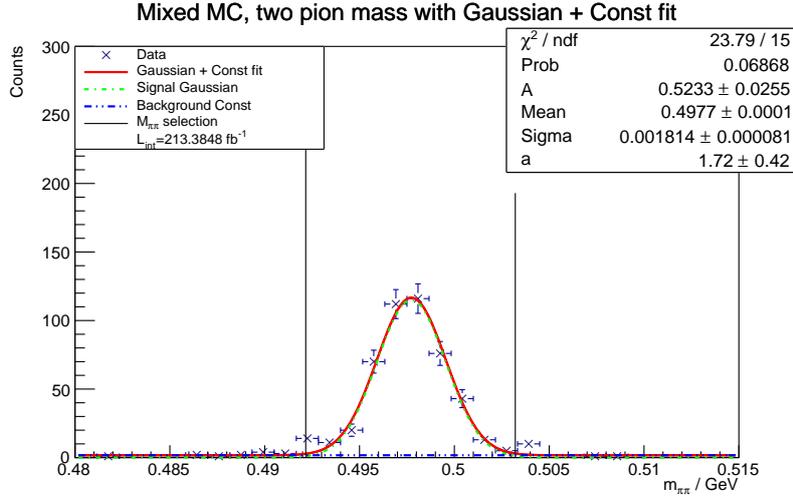
Figure 3.7: Distribution of the reconstructed mass of the two pions after the significance of distance selection for hadronic tagged run dependent mixed MC14 of $L_{int} = 213.3848 \, \text{fb}^{-1}$. A global fit using Eq. (3.4) is performed to determine the purity and efficiency of our K_S^0 selection.

We integrate our signal and background functions within the signal region, as explained in Eqs. (2.2) and (2.3), we determine the selection efficiency and purity to be $\epsilon = 97.00\%$ and $P = 96.49\%$.
In conclusion, we achieve a highly pure and efficient K_S^0 selection by applying a best K_S^0- and significance of distance selection.

3.2.3 D-meson identification

We combine the K_S^0 and remaining pion to find good D-meson candidates. Therefore, we apply the selections on the reconstructed two pion mass and the significance of distance and define an additional signal region for the D-mesons. For clarity reasons we will call the signal regions K_S^0-region and D-region. Since we already reconstructed the D online, using all events, we first want to see, if we observe D-candidates without applying any selections, except for the online selections. Therefore, we look at the invariant mass of the K_S^0 and pion. We expect to observe a sharp peak at the nominal D mass $m_D = 1.86966 \, \text{GeV}$ [Zyl+20]. In Fig. 3.8 the distribution is shown.
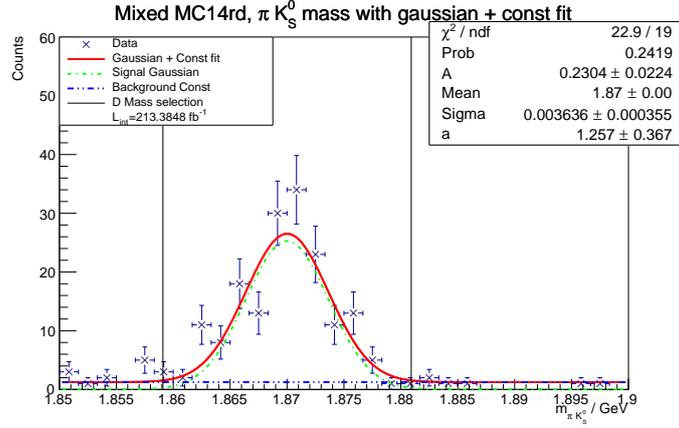
Figure 3.8: Distribution of the reconstructed mass of the K_S^0 and pion for hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$. A peak at the nominal D mass is observed. We perform a global fit (red) using Eq. (3.4) to define the D-region. The fit result is shown at the top right. We define the D-region by the mean plus/minus three time the width of the Gaussian function.

As expected, a peak around the nominal D mass can be observed. In order to define the D-region, we fit a Gaussian plus a constant function, shown in Eq. (3.4), to our distribution. The fit, with a reduced χ^2 of 1.205, shows a good result. The D-region is the defined by the mean plus/minus three times the width of the fit function. We select good D candidates with an invariant mass of $1.859 < M_{\pi K_S^0} < 1.881\,\text{GeV}$. We determine the selection efficiency and purity of the D-mesons as explained before. Therefore, we also apply our K_S^0 selections. The distribution, together with the fit function, is shown in Fig. A.11. We determine the selection efficiency and purity to be $\epsilon_D = 96.13\%$ and $P_D = 89.74\%$. Hence, we achieve to select an efficient and pure selection of D-mesons.

3.2.4 Determination of $N_{K_S^0}^{Norm}$

We aim to determine the number of $B \to \left[D \to \pi K_S^0 \right] l\nu$ events in run dependent MC14. Since we know that there is one missing particle in the decay chain, namely the neutrino, we determine the missing mass squared M_{miss}^2 of each event. Thus, we expect the signal to peak at zero while other background will be shifted to higher values. For this step in the analysis we add charged ($B^+ B^-$) and continuum (coming from $q\bar{q}$ decays) background. The missing mass squared is easy to determine since we know all the kinematics of the process except for the neutrino itself. The four-momentum kinematic equation is defined as follows:

$$p_{\Upsilon(4S)} = p_{B_{tag}} + p_l + p_{\nu_l} + p_\pi + p_{K_S^0} \tag{3.5}$$

$$\Rightarrow p_\nu = p_{\Upsilon(4S)} - p_{B_{tag}} - p_l - p_\pi - p_{K_S^0} \tag{3.6}$$

If we square the last equation we get

$$M_{miss}^2 := 0 = \left(p_{\Upsilon(4S)} - p_{B_{tag}} - p_l - p_\pi - p_{K_S^0} \right)^2 . \tag{3.7}$$

Using the center of mass frame, the energy of the $\Upsilon(4S)$ is equal to the collision energy of $\sqrt{s} = 10.58$ GeV and the momentum components are zero. Thus, we have all necessary values to calculate M_{miss}^2. Furthermore, we apply all defined selections and use truth matching to separate signal from background. Since the lepton is the most reliable particle in the decay chain we apply the truth matching to it and ask for the PDG codes of its mother, sisters and nieces. For this however, we have to think of possible background decays in mixed MC. Since in the charged and continuum file no signal is present we do not need to use truth matching. A list of expected background decays is given in Table 3.1 together with .

#	Definition	Name of decay modes
0.	Signal	$B \to \left[D \to \pi K_S^0 \right] l\nu$
1.	D^* downfeed	$B \to \left[D^* \to \left\{ D \to \pi K_S^0 \right\} X \right] l\nu$
2.	Background	$B \to D^* +$ anything except 1.
3.	Background	$B \to D+$ anything except 0.
4.	Background	Anything else that is not 0. - 3.

Table 3.1: Signal and background decays that we expect in Mixed MC, with number code.

The D^* downfeed decay mode gives us our signal as well as 0. However, since the B decays to a D^* in addition to a missing particle $X \in \{\pi^0, \gamma\}$ we expect a second peak at the nominal mass of the π^0 at $m_{\pi^0} = 134.9768$ MeV [Zyl+20]. The final missing mass squared histogram is shown in Fig. 3.9.
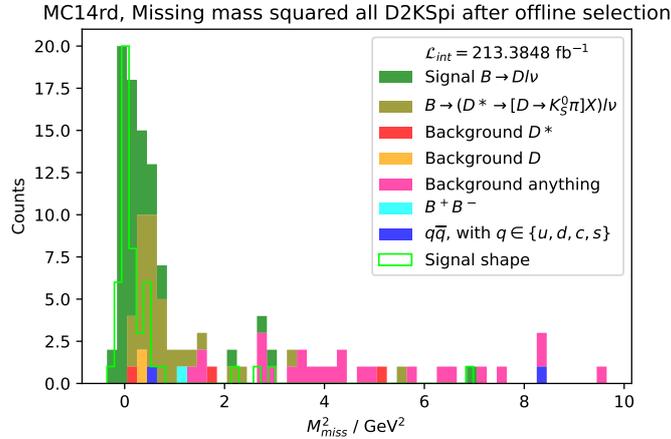
Figure 3.9: Distribution of the reconstructed missing mass squared using selections and truth matching to distinguish between signal and different background decay modes, for hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$. We observe a signal peak at zero.

As expected a peak for our signal decay is visible at $0\,\text{GeV}^2$. Furthermore, we observe a second peak of 1. D^* downfeed events at approximately the π^0 mass squared. Due to our selections other background of D^*, D, charged and continuum is rejected well. We count $N_{K_S} = 50 \pm 7$ signal candidates. We norm our counted signal candidates by the previous determined selection efficiency. Thus, we get $N_{K_S}^{\text{Norm}} = 52 \pm 10$. We compare this number to the expected number of signal events N_{exp}. For this, we take the cross section

of $\Upsilon(4S) \to B\bar{B}$ decays as $\sigma_{B\bar{B}} = 0.5346$ nb [Cun+20], the branching ratios of $\mathcal{BF}(B \to Dl\nu) = 0.0224$ where $l \in \{e, \mu\}$ and $\mathcal{BF}\left(D \to \pi K_S^0\right) = 0.01562$ [Zyl+20]. We assume that the hadronic tagging efficiency amounts to $\epsilon_{tag} = 0.002$ [Kec+19]. However, this number is not well known and thus we expect to see a discrepancy between N_{exp} and $N_{K_S}^{\text{Norm}}$ we derive the expected number of signal events to be

$$N_{\text{exp}} = \sigma_{B\bar{B}} \cdot \mathcal{BF}(B \to Dl\nu) \cdot \mathcal{BF}\left(D \to \pi K_S^0\right) \cdot \epsilon_{tag} \cdot 2 \cdot L_{\text{int}} \tag{3.8}$$

$$= 160 \tag{3.9}$$

We multiply a factor of two, since we have to consider the charge conjugate of the decay. Compared to the normalized number of found signal events we see that the theoretical predictions do not agree well with hadronically tagged run dependent MC14, with regard to the error limits. We observe a discrepancy of 67.5%. As expected this is most likely caused by the not well known tagging efficiency.

In conclusion, we showed a way to select $B \to \left[D \to \pi K_S^0\right] l\nu$ events with a high efficiency and purity. Nevertheless, in total a discrepancy of 67.5% between run dependent MC14 and theory is observed. Thus, our normalized number of K_S^0 events has to be treated carefully.

3.3 $B \to l\pi K_L^0 \nu$ identification

We aim to identify a signal region for good K_L^0 cluster candidates in our $B \to \left[D \to \pi K_L^0\right] l\nu$ sample. Therefore, we do not reconstruct our K_L^0. In doing so, we have two unknown particles within our decay. However, this means that we do not obtain any information about the kinematics of the D-meson. Thus, we reconstruct only $B \to l\pi$, online. We aim to use the missing mass squared to separate signal from background. However, since we have multiple missing particles in our decay we have to predict one of those missing particles in order to define the missing mass squared. Furthermore, we expect that the background of charged and continuum is higher than in the $D \to \pi K_S^0$ events, due to the additional missing K_L^0.

Since we have two missing particles, the missing mass can not be defined that easily. Thus, we need to predict the energy and momentum of the K_L^0 in order to define the predicted missing mass squared based on the neutrino to

$$p_\nu^2 = m_{\text{miss}}^2 = \left(p_{\Upsilon(4S)} - p_{B_{\text{tag}}} - p_l - p_\pi - p_{K_L^0}\right)^2. \tag{3.10}$$

We need to apply constraints to get information about the K_L^0 momentum and energy. Knowing that the K_L^0 and pion should be produced by a D-meson decay, we apply a constraint to the K_L^0 energy and momentum so that if combining the pion and K_L^0, a D-meson with the nominal D mass is created. We get the following formula:

$$m_D^2 = m_\pi^2 + m_{K_L}^2 + 2\left(E_\pi E_{K_L} - \vec{p}_\pi \vec{p}_{K_L}\right) \tag{3.11}$$

Breaking down the momentum vectors into the different momentum components we get

$$m_D^2 = m_\pi^2 + m_{K_L}^2 + 2\left(E_\pi E_{K_L} - p_\pi^x p_{K_L}^x - p_\pi^y p_{K_L}^y - p_\pi^z p_{K_L}^z\right) \tag{3.12}$$

Thus we have now four unknown parameters (E_{K_L}, $p_{K_L}^x$, $p_{K_L}^y$ and $p_{K_L}^z$) to solve for. Our momentum components can be solved for by the azimuthal and zenith angle ϕ_{K_L} and θ_{K_L} of the K_L^0, respectively. θ_{K_L} and ϕ_{K_L} are unknown variables as well, thus we add in total two additional unknown and four equations of

$$\cos(\theta_{K_L}) = \frac{p_{K_L}^z}{\sqrt{p_{K_L}^{x2} + p_{K_L}^{y2} + p_{K_L}^{z2}}} \tag{3.13}$$

$$\sin(\theta_{K_L}) = \frac{\sqrt{p_{K_L}^{x2} + p_{K_L}^{y2}}}{\sqrt{p_{K_L}^{x2} + p_{K_L}^{y2} + p_{K_L}^{z2}}} \tag{3.14}$$

$$\cos(\phi_{K_L}) = \frac{p_{K_L}^x}{\sqrt{p_{K_L}^{x2} + p_{K_L}^{y2}}} \tag{3.15}$$

$$\sin(\phi_{K_L}) = \frac{p_{K_L}^y}{\sqrt{p_{K_L}^{x2} + p_{K_L}^{y2}}} \text{ [TKU11]} \tag{3.16}$$

In total we have now six variables that we have to determine and five equations. We can add a sixth constraint, since we know the mass of the K_L^0 sufficiently well, by

$$m_{K_L} = \sqrt{E_{K_L}^2 - p_{K_L}^{x2} - p_{K_L}^{y2} - p_{K_L}^{z2}}. \tag{3.17}$$

Using the python package **scipy.optimize.fsolve** [com22b] we solve for the K_L^0 energy and momentum. Thus, we are then able to define the predicted missing mass squared. We expecting a peak at zero, since the neutrino is the last particle we have no information about. The distribution is shown in Fig. 3.10.

As expected we observe a peak at zero for our signal. However, we observe a long tail to negative values of our predicted missing mass squared. This suggests that we are reconstructing more than we should. Most likely this is caused by our constraints we introduced earlier. Furthermore, all the background is peaking at zero as well. By our constraints we use that we know the mass of the K_L^0, which results in forcing a particle of the mass of a K_L^0 in each background event. Through this background sculpting it is not possible to perform a fit and defining a good K_L^0 cluster candidates signal region.

Investigating this background sculpting closer by changing some of our constraints shows similar results. The difficulty is that if we use constraints on the B mass we also have to consider the neutrino in our equations. Thus, we would introduce additional unknowns to our system we can not solve for.
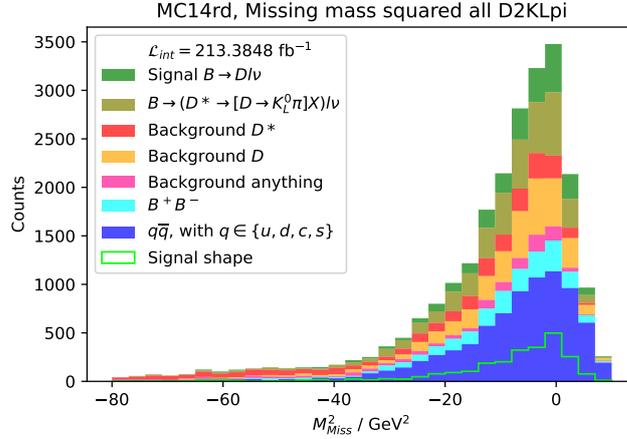
Figure 3.10: Distribution of the reconstructed missing mass squared, predicting the momentum and energy of the K_L^0, for hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848 \, \text{fb}^{-1}$. We observe a peak of signal events at zero, as well as for background events.

In conclusion, we can not report of a definition for a K_L^0 candidate signal region in a $B \to \left[D \to \pi K_L^0 \right] l \nu$ analysis. A massive background sculpting makes it impossible to define a K_L^0 signal region with our constraints used. Thus, we are not able to test our FastBDT-training done in Section 2.3.

Nevertheless, we want to give analysts a first way to identify K_L^0 clusters in the KLM. We mentioned at the end of Section 2.3 that it is possible to train a FastBDT on MC and data simultaneously to identify low momentum K_L^0's. However, due to poor MC modeling of the KLM shape variables a decrease in efficiency for data is expected. Still, this could give analysts a first conditionally useful module for the identification of K_L^0-mesons in the KLM. Thus, we aim to use the particle gun and generated specified K_L^0 particles in MC to train signal events from it against background data from our ϕ-study. The procedure will be discussed in detail in Chapter 4.

K_L^0 identification using the particle gun

We use the particle gun to generate and propagate specified particles through the simulated Belle II detector. Here we will generate K_L^0's with self chosen momentum and direction in the detector and simulate the detector output. We use this to obtain a pure sample of good K_L^0 cluster candidates. We then use this sample to train our FastBDT against bad cluster candidates as defined in Chapter 2. This chapter describes how we determine the momentum, θ and ϕ distribution for our simulated K_L^0 sample and how we train a FastBDT against data background events.

4.1 Creating K_L^0 events with the particle gun

With the particle gun we are able to simulate the momentum, θ and ϕ distribution by hand. In order to precisely simulate the K_L^0 distributions we model those distributions based on run dependent MC K_L^0 clusters. Therefore, we take a look on generator information of a small sample of mixed run dependent MC14. We select only true K_L^0 candidates using the PDG code of the generated particle. For those events, we illustrate the momentum p, θ and ϕ in histograms. Thus, we determine the bin centers and heights of each histogram and give these as a sequence of points to the particle gun. With these pdfs we generate 50000 events and compare the distributions of mixed MC and particle gun. Since, we model each pdf separately for p, θ and ϕ, assuming they are one dimensional and there are no correlations between the distributions, a small deviation in the shape is expected. The distributions are shown in Fig. 4.1. As expected we observe small deviations in the shape. This shows that there are correlations between the three variables. Still the agreement is sufficient enough to us the particle gun simulated K_L^0 candidates for our FastBDT training.

We use the generated K_L^0 four vector and the KLM cluster position of the particle gun to determine the cosine of the opening angle and thus define good and bad K_L^0 cluster candidates, as previously done in experimental data (c.f. Section 2.2.3). These signal events are combined with experimental data background of our ϕ study.
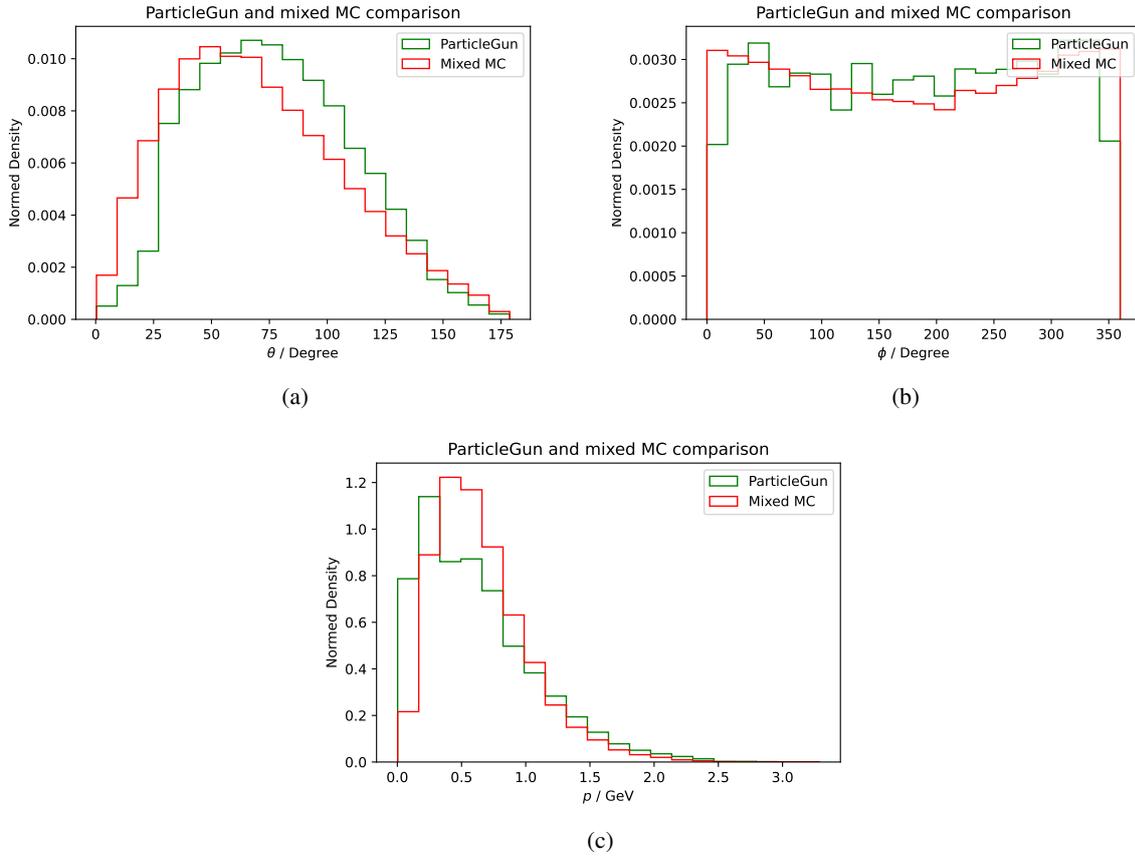
Figure 4.1: Distributions of θ (a), ϕ (b) and momentum (c) for the generated particle using the particle gun (green). We compare the distributions to mixed MC14 data (red) for (a) θ, (b) ϕ and (c) momentum. We observe a similar shape of the distributions.

4.2 Background selection in data

For the background in experimental data we use only events with an opening angle of $\alpha < 0.5\,\text{rad}$ between the predicted and KLM cluster vector. We want to focus on distinguishing beam background clusters, coming from fast neutrons, and real K_L^0 clusters in the KLM. To select those events we have to take a closer look at the KLM Timing variable. The distribution for it is shown in Fig. 4.2. We expect one peak around zero as this is where we expect physics events.
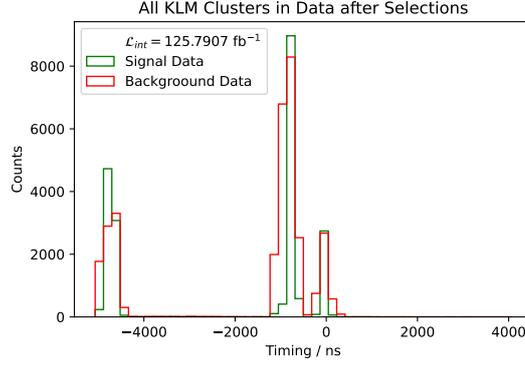
Figure 4.2: Distribution of the KLM timing for real data of $e^+e^- \rightarrow \gamma ISR \left[\phi \rightarrow K_S^0 K_L^0\right]$ decays. Three peaks of signal and background candidates are observed, suggesting that the KLM has no time calibration, yet.

We observe three peaks. One as expected around zero, the other two shifted to negative times of ≈ -730 ns and $\approx -4\,740$ ns. This indicates that there are events occurring before the collision. This comes due to the fact that the KLM time is not calibrated. However, this makes the variable difficult to use. We investigate these events closer by plotting the z and ϕ variable of the KLM cluster (c.f. Fig. 4.3).
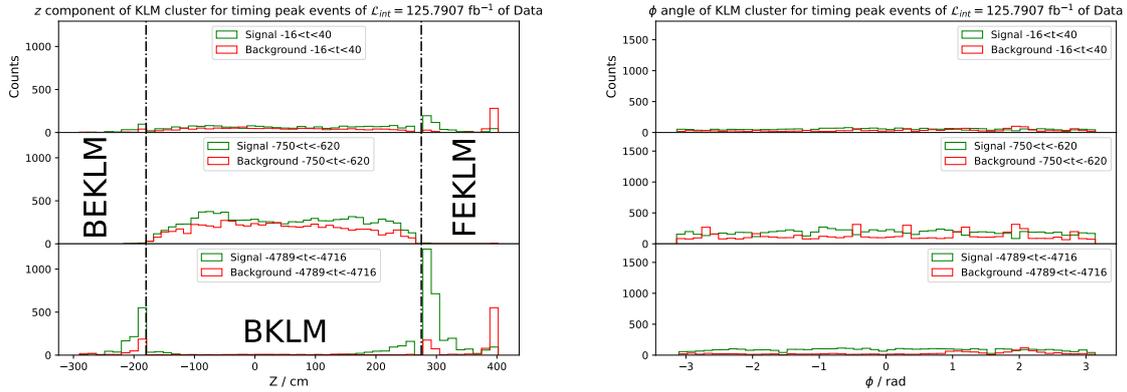
Figure 4.3: Distribution of the KLM z position (left) and ϕ angle (right) for all peaks in KLM timing for real data of $e^+e^- \rightarrow \gamma ISR \left[\phi \rightarrow K_S^0 K_L^0\right]$ decays. We observe in z that the especially the endcap and barrel have different timing calibration errors. In ϕ no pattern is visible.

We observe in z that events in the barrel occur mainly between $-750 < t < -620$ ns, while endcap events are shifted to $-4789 < t < -4716$ ns. In ϕ we do not observe a clear pattern. A calibration of the KLM timing would exceed the scope of the thesis. Nevertheless, since we expect that these two additional peaks and the zero peak are physics events we select only those regions for our background and combine them with our signal from the particle gun.

4.3 Pre-selections on combined particle gun signal and data background events

Before we train the FastBDT on particle gun signal and data background events we take a look at the KLM shape variables. Here we reject each discrete variable. Furthermore, variables we expect to have a high correlation with other variables are rejected as well. We plot the distribution for the remaining variables, in Fig. A.12.

We define some additional selections to reduce the background as much as possible. The first selection we apply is on the cluster z position. The distribution is once more shown in Fig. 4.4.
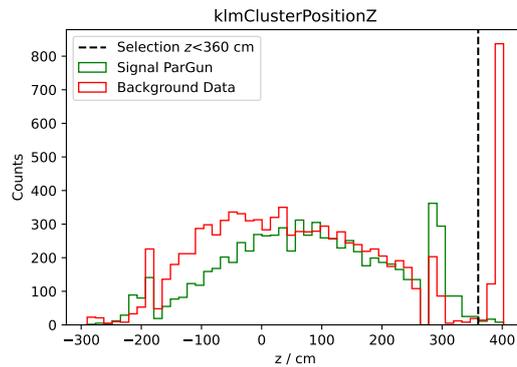
Figure 4.4: Distribution of the KLM cluster z position for particle gun signal (green) and data background (red). A majority of background events peaks in the forward endcap. We apply a cut by eye and select events with $z < 360$ cm.

We observe a high background peak above 360 cm. In contrast, we do not observe a significant amount of signal in the same region. This is a strong indicator that those background clusters originate from beam induced fast neutrons coming from the beam pipe and hitting the outer part of the forward endcap KLM. Thus, we select only event below 360 cm. Further, we take a closer look at the distribution of the number of the inner most layers that registered the first hit in Fig. 4.5.
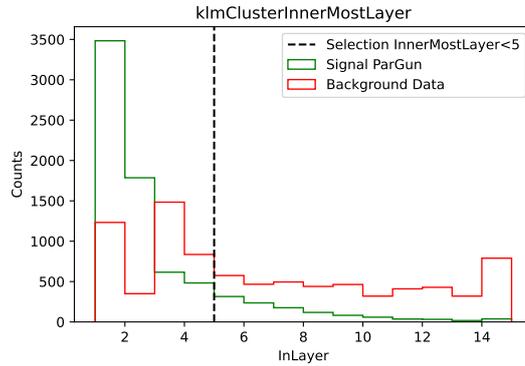
Figure 4.5: Distribution of the first layer that detected a hit in the KLM for particle gun signal (green) and data background (red). Signal events seem to produce hits already in the first two layers, while background events are more uniformly distributed. Thus we apply a selection at Inner most layer < 5, by eye.

The probability of the detection of the first hit of a signal K_L^0 drops with respect to the number of layers in the KLM. Most of the signal events are first detected in the first two layers. This shows that the K_L^0 candidates do not have enough energy to penetrate the detector material and therefore interact strongly with the first iron plates of the detector. Meanwhile, most of the background events are as well detected in the first four layers. However, a drop as in signal for the back layers is not observed. Instead, the amount of background events stays constant at ≈ 600 events per layer, with a small peak in the last layer. This small peak is likely to be produced by beam induced fast neutrons again coming from the beam pipe, which we did not reject with our selection on the z position of the cluster. Thus, we reduce the background by accepting only events where the assumed K_L^0 has its first hit within the first four layers. This selection is chosen by eye.

We now apply these selections and we take another look at the numbers of matched clusters and tracks. Since, K_L^0's produce neutral clusters we expect that the signal peaks at zero[1]. The distribution in Fig. 4.6 validates our expectation.

---

[1] With respect to the track cluster matching algorithm described in Section 2.2.3 which is again not very reliable as a track is matched to a cluster, if one is found with in a 150 cm radius around the KLM cluster.
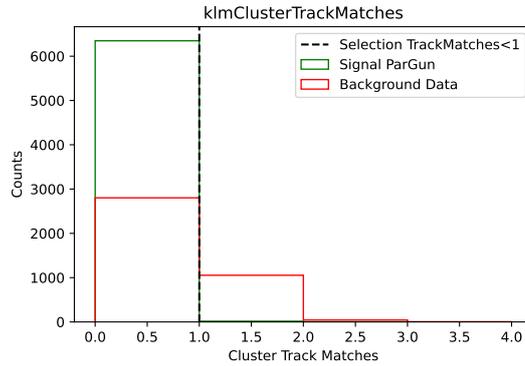
Figure 4.6: Distribution of the KLM clusters that are matched to a track for particle gun signal (green) and data background (red). Since the K_L^0 is a neutral particle we expect and see that signal events peak at zero. We select candidate below one for the number of matched tracks to a cluster.

Thus, we select only events with no Cluster-Track match for the FastBDT training and testing sample. If we once more examine at Fig. A.12 one could suggest to cut on the K_L^0-ID of the KLM as well, since it shows high separation power. However, we would have to cut very tight to zero and thus loose a great amount of background events. This would then result in even worse training due to the fact that we would train on approximately 6000 signal and only 500 background events. Thus we do not do a selection on the K_L^0-ID variable before the training.

4.4 FastBDT training on paritcle gun and data events

For the final FastBDT training we split our sample in a training and testing sample. We use 80% of our events for the training and the remaining 20% for the testing sample. As our final training variables we use:

- Inner most Layer - see Table 2.3.

- Cluster energy - gives the energy of a detected cluster in the KLM.

- Phi - gives the azimuthal angle of the cluster with regard to the interaction point.

- Cluster Z position - gives the z position of the detected KLM cluster.

- K_L^0-ID - see Table 2.3.

We exclude the variable Layers, because it shows high correlations with the cluster energy. Thus, we can enhance the performance of the FastBDT in the end. We show the result of the FastBDT training in Fig. 4.7, in form of the BDT output classifier.
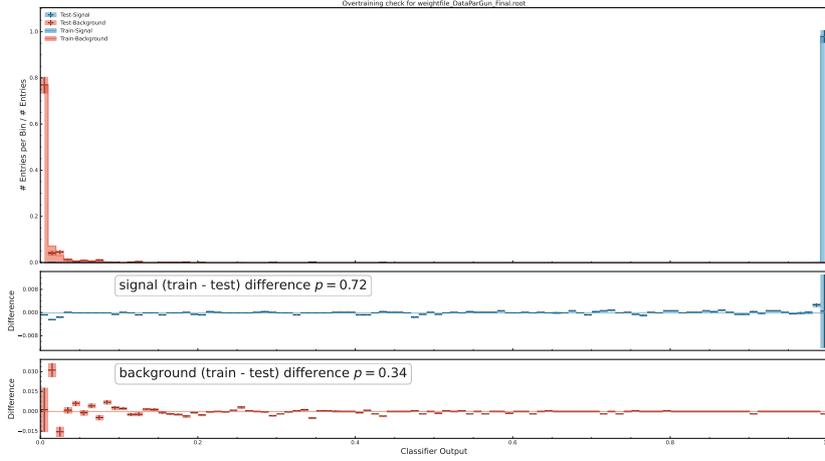
Figure 4.7: Distribution of the classifier output for the FastBDT training on particle gun signal and data background. A clear separation between signal form a particle gun and background events form real data is observed.

The FastBDT is showing the ability to separate signal form background, for particle gun signal and data background events. We do not observe signs of overtraining.

Still, we need to evaluate the efficiency of the BDT training on a sample of particle gun signal and data background events (we will call this "combined sample" for simplicity in the following), as well as on experimental data only of our ϕ study. Therefore, we derive the efficiency and fake-rate as shown at the end of Section 2.2.3. We plot the efficiency against the fake-rate for both the combined and only experimental data sample. The plots are shown in Fig. 4.8, left for the combined and right for the ϕ data sample.
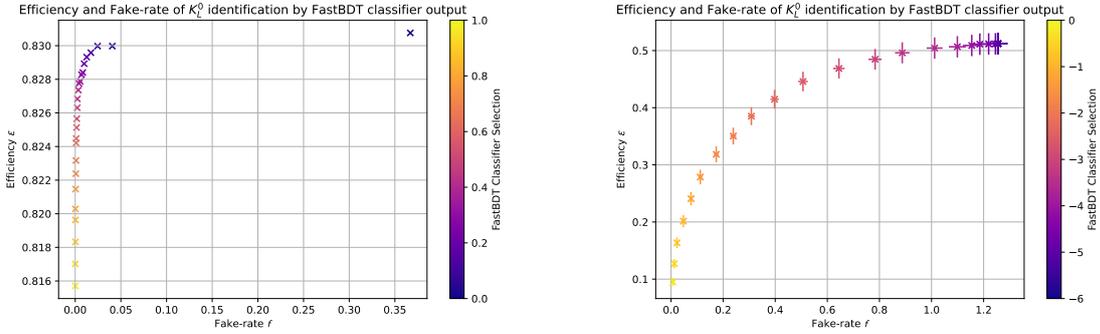
Figure 4.8: Efficiency and fake-rate for scanning through the output classifier of the combined FastBDT training. Applied to a combined sample of events (left) and to ϕ data only (right). The error for the combined sample is for all points of the order of $1 - 2\%$. Thus, for clarity we do not plot the errorbars. In the ϕ data plot we use a logarithmic FastBDT classifier ourput.

If we apply our training on a combined sample we observe a starting efficiency of $\epsilon_{\text{comb}} = (83.1 \pm 1.4)\%$ with a fake-rate of $f_{\text{comb}} = 0.367 \pm 0.008$, in the left plot of Fig. 4.8. Further, we observe a decrease in the fake-rate to $f_{\text{comb}} = 0.010 \pm 0.001$ with an efficiency of $\epsilon_{\text{comb}} = (82.8 \pm 1.4)\%$, which we define as the optimal working point. After that a fast decrease in the efficiency is visible. However, we have

to keep in mind that we still achieve an identification efficiency, for K_L^0's in a combined sample, of $\epsilon_{\text{comb}} = (81.6 \pm 1.3)\%$, while the fake-rate reduces to approximately zero. This is to be expected, since we apply here our combined training to a combined independent sample of MC signal and data background events. Thus, we already know if there is a true K_L^0 cluster or not.

Otherwise, if we apply our FastBDT output to ϕ data only (c.f. Fig. 4.8 right) we observe a starting identification efficiency of $\epsilon_{\text{Data}} = (51.2 \pm 1.9)\%$ and a fake-rate of $f_{\text{Data}} = 1.255 \pm 0.036$ average fake clusters per event. Regarding the efficiency, this seems to match previous studies of K_L^0 identification in Belle, see [Tch+02]. We normalize the efficiency by the initial efficiency to

$$\epsilon_{\text{Data}}^{\text{Norm}} = \frac{\epsilon_{\text{Data}}^i}{\epsilon_{\text{Data}}^0}. \tag{4.1}$$

The uncertainty is calculated using the Bayesian approach [Cas12]. We plot the normalized efficiency against the fake-rate in the histogram in Fig. 4.9.
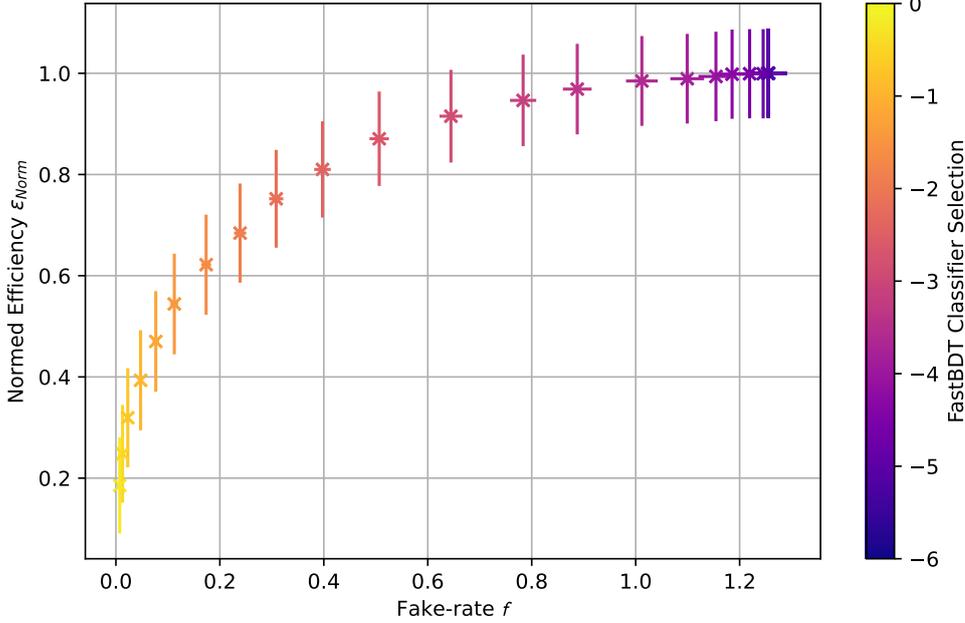
Figure 4.9: Efficiency and fake-rate for scanning through the output classifier of the combined FastBDT training. The efficiency is normalized by the initial value. We observe that a combined training applied to real data reduces the fake rate of K_L^0 clusters in the KLM, but suffers in efficiency. The FastBDT classifier output is given in a logarithmic scale.

By scanning through the FastBDT output classifier we observe a rapid decrease in efficiency and fake-rate at the same time. This validates our concerns from the end of Section 2.3, where we discussed the potential influence of the not well modeled MC KLM shape variables in comparison to experimental data. Due to not sufficient modeling of KLM variables between MC and experimental data, an identification

of true and fake K_L^0 clusters in the KLM by a FastBDT suffers from low efficiency when applied to experimental data. On the other hand, we applied the FastBDT to our ϕ data sample. Those events have only high momentum K_L^0's. The FastBDT however is modeled with low momentum K_L^0 candidates from $B \to Dl\nu$ decays. Thus, a decrease in efficiency has to be expected. Still we observe a reduction of average fake clusters per event to $f_{\text{Data}} = 0.239 \pm 0.012$ with an identification efficiency of $\epsilon_{\text{Data}} \approx (35.0 \pm 1.5)\% \sim \epsilon_{\text{Data}}^{\text{Norm}} \approx (68.4 \pm 9.8)\%$. Thus, we showed that the developed algorithm reduces the fake-rate significantly, while reducing the efficiency by approximately half with regard to the estimated KLM identification efficiency of $\sim 50\%$ [Tch+02].

In order to get a more precise picture on the performance of our combined FastBDT training we apply the training to a experimental data KLM only sample. In this data sample we only look at KLM clusters that have no track matches. In Fig. 4.10 the FastBDT output classifier is shown.
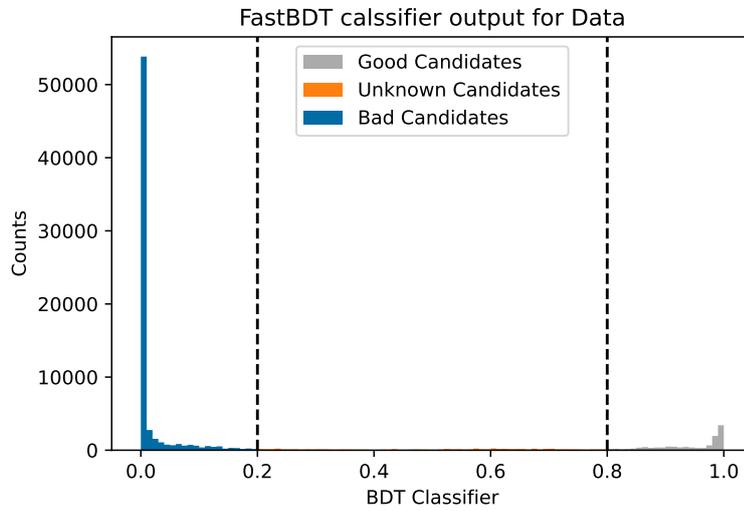
Figure 4.10: Distribution of the FastBDT output classifier of a combined training applied to KLM data only. We define three regions of bad (< 0.2), good (> 0.8) and unknown K_L^0 candidates in between. We observe a good separation between possible bad and good K_L^0 candidates.

We observe a high peak at zero and a small peak at one. This shows that the FastBDT is able to distinguish between good and bad K_L^0 cluster candidates. We define three regions of bad (< 0.2), good (> 0.8) and unknown K_L^0 candidates in between. We calculate an estimate of the fake-rate to

$$f_{\text{KLM}} = \frac{N_{\text{exp}} - N_{\text{removed}}}{N_{\text{events}}}. \tag{4.2}$$

Here N_{removed} is the number of removed bad K_L^0 candidates and N_{events} is the number of totally observed KLM data events. N_{exp} is the number of expected fake clusters in our KLM only data sample. We approximate the number of expected fake clusters by multiplying the number of totally observed KLM data events N_{events} with the initial fake-rate of our ϕ-study f_ϕ. Thus, we get

$$N_{\text{exp}} = f_\phi \cdot N_{\text{events}}. \tag{4.3}$$

Since we approximate the number of expected fake clusters in our KLM only data sample we do not defined an uncertainty. We apply a selection to the FastBDT output classifier and remove all candidates below 0.2. Thus, we calculate the approximate fake-rate of the number of average fake clusters after this cut to $f_{KLM} = 0.25$. This shows that our combined training reduces the fake-rate assuming an initial fake-rate of the ϕ-study.

We can visualize the performance of our FastBDT training by looking at the current K_L^0 classifier classifier, the K_L^0-ID (K_L^0 identification). The distribution is shown in Fig. 4.11.
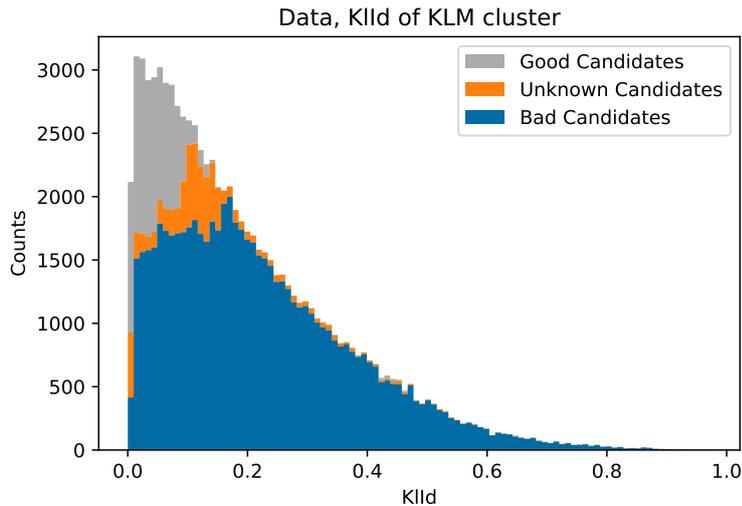
Figure 4.11: Distribution of the current K_L^0 identification classifier, the KLM K_L^0-ID. The distribution is separated in our FastBDT output classifier. Without our FastBDT separation we are not able to make assumptions of good and bad K_L^0 candidates.

We observe a broad distribution showing no separation power. However, using our classification in Fig. 4.10 we observe that we are able to distinguish good and bad K_L^0 cluster candidates.

Thus, we conclude that our new developed method of a combined MC-data training identifies K_L^0 cluster candidates while reducing the fake-rate. Still, a study of our K_L^0 identification is needed to study its efficiency and fake-rate more precisely.

Summary

To summarize, a new approach to identify K_L^0 clusters in the KLM is presented. We used a data-driven FastBDT training to identify K_L^0 clusters the $e^+e^- \to \gamma_{\text{ISR}} \left[\phi \to K_S^0 K_L^0 \right]$ decay. Since we know that the K_S^0 decays to two charged pions and the photon has to have an energy of $\sim 5.24083\,\text{GeV}$ due to the collision kinematics, we could predict the momentum and energy of the K_L^0 without reconstructing it. For a FastBDT training a sample of K_L^0 candidates with high purity has to be selected. Thus, by enhancing the purity of the K_S^0-γ_{ISR} sample increases the purity of the K_L^0 sample for this process as well. Therefore, we applied selections to the significance of distance (σ_d) of the K_S^0 vertex decay and the photon energy. We used signal MC to test our mathematical approaches. In data we used the minimization algorithm Section 2.1.1 to define the best selection. This is done by defining a signal region for good K_S^0 candidates by fitting a Gaussian plus a polynomial of first order (see Eq. (2.1)) to our reconstructed mass of the two pions. Then we calculate the selection efficiency and purity by Eqs. (2.2) and (2.3) and minimizing the distance to the optimal point of $(1, 1)$. Thus, we achieve a purity of $\sim 90.20\%$ and an selection efficiency of $\sim 89.96\%$.

With this sample of high purity K_S^0-γ_{ISR} candidates we then predict the four-momentum vector of our K_L^0 candidates by using the four-momenta of the beam, the photon and the K_S^0. Due to an energy leakage of the ECL crystals for high energy photons a constraint to the photon energy is applied, by using the analytical value for the process. Without this process no evidence for good K_L^0 candidates is found in our sample based on the missing mass, as shown in Fig. 2.8. Hence, with the energy constrain we observe a peak at the nominal K_L^0 mass. A study of generator MC events showed that a small error in the photon energy results in a large error in the missing mass. Still, with the constraint we define a signal region for good K_L^0 candidates by eye at $m_{\text{miss}} > 0.435\,\text{GeV}$.

We use this sample of good K_L^0 candidates to define good K_L^0 cluster candidates in the KLM. Therefore, we calculate the opening angle between the missing momentum and KLM cluster position vector. Good cluster candidates should show only a small opening angle or in a $\cos(\alpha)$ distribution values close to one. In signal MC a broad distribution is observed, although truth matching is applied. In event displays large opening angles between truth K_L^0 vector and truth matched cluster are observed. In some of those cases secondary particles originating form the decay of the K_L^0 caused these clusters. However, often no generated particle is found close to the cluster. Thus, we assume that we either observe hadronic splitoffs or that the KLM cluster truth-matching is not working. Further, we observed that neutral K_L^0 clusters are matched to tracks. Is this the case the definition of neutral cluster is not applicable anymore. The track matching failing could be traced back to the fact that a track is matched to a cluster if its extrapolation,

inside the KLM, is found within 150 cm of the cluster. Due to the detector measurements this selection seems to be not sufficient. Anyway, solving those problems of truth and track matching exceeds the scope of this analysis, but have to be considered.

In data we observe for $\cos(\alpha)$ a peak at one with a long tail towards higher opening angles (lower $\cos(\alpha)$ values). Thus we define good K_L^0 cluster candidates in the KLM to have an opening angle between missing momentum and KLM position vector of less than 0.5 rad. With this definition we then train the FastBDT on the KLM shape variables shown in Table 2.3. The training showed a strong separation power. Applied to data we scan through the FastBDT classifier output and calculate the identification efficiency and fake-rate defined in Eqs. (2.17) and (2.20). Thus, we observe a strong decrease in the fake-rate with a good efficiency (see Fig. 2.21). We report a reduction in the fake-rate from initial $f = 1.0353 \pm 0.0087$ average fake clusters per event to $f = 0.0985 \pm 0.0020$ fake clusters per event with an identification efficiency of $\epsilon = 0.6998 \pm 0.0065$. Comparing this to the overall efficiency of $\sim 50\%$ [Tch+02] an increase in efficiency is seen. If we compare it to the identification efficiency of different K_L^0 momenta, as in [Aus+15], a good agreement is found with high momentum K_L^0's between $2 - 3$ GeV. However, we identify K_L^0 candidates with a high efficiency. We are not able to identify low momentum K_L^0's. Studies of different decay modes for example $e^+ e^- \to \gamma_{\text{ISR}} \left[J/\psi \to \pi^+ \pi^- K_S^0 K_L^0 \right]$ showed no evidence of good K_L^0 candidates due to an overwhelming background.

We still have found a highly efficient way to identify K_L^0 clusters in the KLM reducing the fake-rate significantly. We test our identification in a $B \to Dl\nu$ analysis of hadronic tagged run dependent MC14. Here, due to their nearly identical branching ratios we use the decay modes $D \to \pi K_S^0$ and $D \to \pi K_L^0$. The $D \to \pi K_S^0$ is used as a normalization mode to define the K_L^0 identification efficiency. Thus, we start by selecting a high purity sample of K_S^0 and D candidates. Afterwards, we determine the missing mass squared of the event and observe a peak at zero. We count and normalize the number of signal events to $N_{K_S}^{Norm} = 50 \pm 10$. Compared to theory predictions we observe a discrepancy of $\approx 67.5\%$, which is expected since the tagging efficiency is not well known.

Knowing we have found good $D \to \pi K_S^0$ events, we investigate the $D \to \pi K_L^0$ channel. Since we have two missing particles, the neutrino and K_L^0 respectively, we determine the missing mass squared of each event. We expect to observe a peak at the nominal K_L^0 mass squared. However, no peak in our signal decay is observed due to the expected dominance of background. Thus, we predict the K_L^0 momentum and energy by applying the constraint that the missing K_L^0 and pion are daughters of the D-meson. Further, the momentum components are calculated by the azimuthal and zenith angle. Hence, we observe a peak at zero (see Fig. 3.10), which suggests that the K_L^0 four momentum is reconstructed correctly. However, all background components are peaking at zero as well. We observe a background sculpting at zero due to the applied constraints. Thus, we are not able to define a signal region for good K_L^0 mesons and do not report of any good K_L^0 cluster candidates.

Up to this point we are able to identify high momentum K_L^0 clusters of $\sim 2 - 3$ GeV, by the FastBDT trained on the ϕ-data. Still, in missing energy analysis low momentum K_L^0's are an important background source. Thus, we want to train a FastBDT using particle gun signal and ϕ-data background events. Therefore, we generate K_L^0 particles with the momentum, θ and ϕ distributions seen in run dependent mixed MC14 for truth matched K_L^0's. Thus, we get signal K_L^0 events of low and high momentum, respectively (see Fig. 4.1). For the background data we use the KLM timing to select physic events. Since the KLM has no time calibration the variable has to be treated with care. Still, we combine the signal particle gun and background data events and apply further selections to neglect obvious background events. Finally, we train a FastBDT on our combined sample and apply it to a combined and ϕ data

sample afterwards. We calculate the identification efficiency and fake-rate for our combined and data only sample, respectively. Applied to a combined sample we observe a high efficient identification of $\epsilon_{\text{comb}} = (83.1 \pm 1.4)\%$, while reducing the fake-rate to $f_{\text{comb}} = 0.367 \pm 0.008$ average fake clusters per event (see Fig. 4.8 left).

More important however is the performance applied to experimental ϕ-data only. Due to modelling discrepancies between data and MC for the KLM shape variables a reduced efficiency is expected. Further, it must be taken into account that we apply our trained FasdBDT to ϕ-data. Thus, a reduce in the efficiency is expected as well, due to the momentum limits in the ϕ mode. We observe an identification efficiency of $\epsilon_{\text{Data}} = (51.2 \pm 1.9)\%$ with a fake-rate of $f_{\text{Data}} = 1.255 \pm 0.036$ average fake cluster per event. The efficiency agrees well with the overall KLM identification efficiency of $\sim 50\%$ [Tch+02] for K_L^0 clusters. Applying selections to the FastBDT classifier output we see a decrease in efficiency and fake-rate. We are able to reduce the fake-rate to $f_{\text{Data}} = 0.239 \pm 0.012$, while reducing the efficiency to $\epsilon_{\text{Data}} = (35.0 \pm 1.5)\%$. Thus, the identification efficiency is still $(68.4 \pm 9.8)\%$ of the initial efficiency. Since we identify here high momentum K_L^0 clusters by training on low momentum K_L^0 clusters we test our identification on a KLM data only sample with low momentum K_L^0's. We observe a good separation power in comparison to the current K_L^0 identification by the KLM K_L^0-ID variable (see Fig. 4.11). Further, we calculated an approximated fake-rate of $f_{\text{KLM}} = 0.25$ average fake clusters per event after rejecting candidates with a FastBDT output classifier of less than 0.2.

Conclusion and Outlook

In conclusion we, report of a new approach to identify K_L^0 clusters in the KLM by training a FastBDT on particle gun signal and data background events. We showed that the training applied to a experimental ϕ-data sample reduces the average number of fake K_L^0 clusters per event to $f_{\text{Data}} = 0.239 \pm 0.012$ with an identification efficiency of $\epsilon_{\text{Data}}^{\text{Norm}} = (68.4 \pm 9.8)\%$ of the initial efficiency. Further, the combined FastBDT training showed a higher separation power than the current classifier, if applied to a KLM data sample only. We report of an approximated reduction of the fake-rate to 0.25 average fake clusters per event.

One could enhance the identification efficiency, by a better modelling of the KLM shape variables between data and MC. Further, a calibration of the KLM timing is essential to enhance the sensitivity of the FastBDT training. In addition, a test of the combined training in a missing energy analysis would be important to validate the usability of such a training in order to identify true K_L^0 clusters in the KLM more precisely.

Further, the KLM cluster-truth and -track matching needs to be examined in more detail.

Finally, a study of our combined training in an analysis where we find K_L^0 candidates of low and high momentum together is needed to evaluate the efficiency and fake-rate reduction more precisely. Especially suitable for this study could be J/ψ decays to K_L^0 final states.

Appendix

A.1 Analysing $\gamma_{\mathsf{ISR}}\left[J/\psi \to \pi^+\pi^- K_S^0 K_L^0\right]$

We discussed in Section 1.4.2 that in order to produce a J/ψ, using a center of mass energy of $\sqrt{s} = 10.58\,\text{GeV}$, we need a ISR photon of around $E_\gamma^{J/\psi} = 4.83674\,\text{GeV}$ and to distinguish the two pions from the J/ψ and K_S^0 decay we introduce a best K_S^0 selection. Thus, we are now able to follow the same analysis steps as discussed above.

First, we want to see if we can reconstruct K_S^0-mesons from two of the four pions. The two pion mass distribution is shown in Fig. A.1.
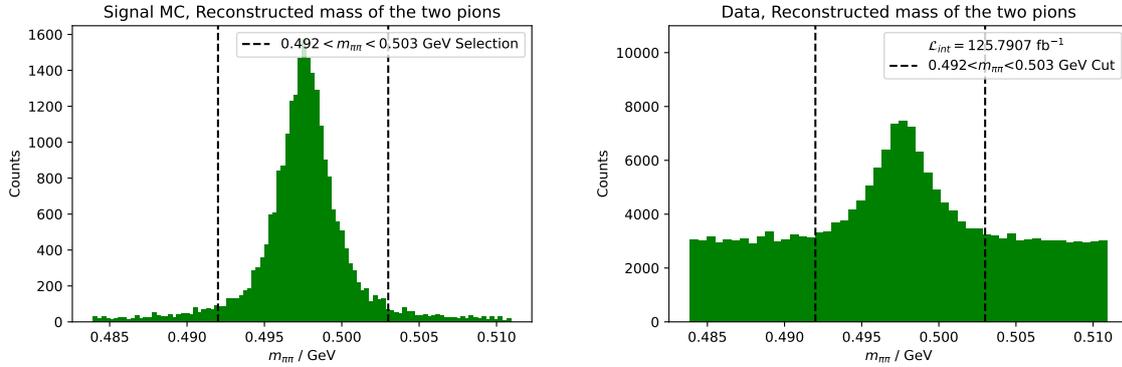
Figure A.1: Distribution of the reconstructed mass of two of the four charged pions for signal MC (left) and data (right) after best K_S^0 selection. In both a peak at the nominal mass of the K_S^0 is observed. We define the signal region based on Fig. A.2.

We observe a peak at the theoretical mass of the K_S^0. In signal MC (c.f. Fig. A.1 left) we define the signal region, for potential good K_S^0 candidates as $0.492 < m_{\pi\pi} < 0.503\,\text{GeV}$. This definition is based on a Gaussian fit for the signal and a constant function for the background shown in Eq. (A.1). The parameters are floating.

$$h(x) = \frac{A}{\sqrt{2\pi}\sigma} * \exp{-\left(\frac{(x-\mu)}{\sigma}\right)^2} + a \tag{A.1}$$

59

The fit result is shown in Fig. A.2. To get the signal region boundaries we take the mean value of the fit and add/subtract three times the width of the Gaussian fit.
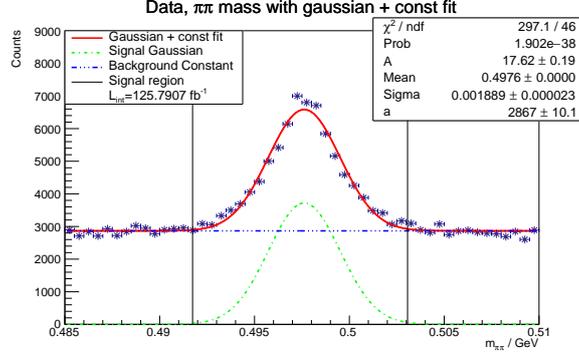
Figure A.2: Distribution of the reconstructed two pion mass with global Gaussian plus constant fit as signal and background (red), respectively. We plot the signal Gaussian and the background constant separately in green and blue, respectively. The signal region is defined as $\mu \pm 3 \cdot \sigma$ of the signal Gaussian.

Having defined our signal region we now want to select K_S^0 candidates with a high purity and efficiency. Therefore, we define selections to the significance of distance of the K_S^0 vertex σ_d (Fig. A.3) and the ISR photon energy E_γ (Fig. A.4) as previously done in the ϕ-study. For data we determine these selection values again by the minimization algorithm explained in Section 2.1.1.
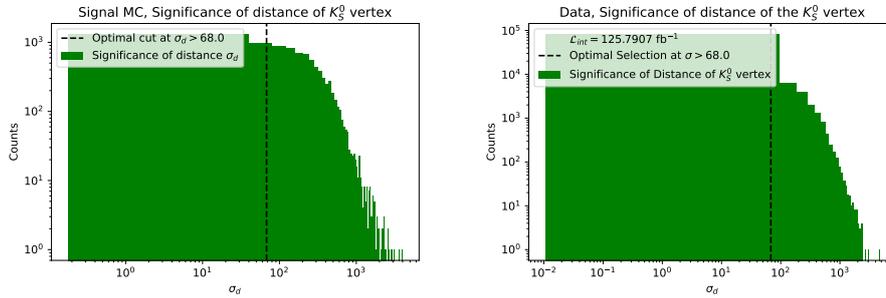
Figure A.3: Distribution of the significance of distance of the K_S^0 vertex decay for signal MC (left) and data (right). We use the minimization algorithm to define a high purity cut in data at $\sigma_d > 68$. In signal MC we define a selection by eye: $\sigma_d > 20$.
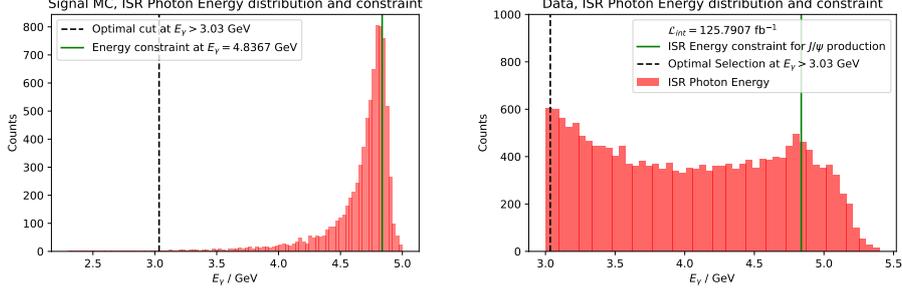
Figure A.4: Distribution of the ISR photon energy for signal MC (left) and data (right). The green line shows analytical number to create a J/ψ in the decay process $e^+e^- \to \gamma_{\mathrm{ISR}}J/\psi$. In signal MC we define a selection by eye at $E_\gamma > 4\,\mathrm{GeV}$. In data we use the minimization algorithm and define a selection at $E_\gamma > 3.03\,\mathrm{GeV}$.

Investigating signal MC and data for the ISR photon energy we observe a peak of the distribution at the calculated energy needed. Furthermore, in signal MC a long tail towards lower energies is visible as well as in Fig. 2.4 of the ϕ-study. This again comes from the possible energy leakage of high energy photons in the ECL. We apply a cut based on the minimization algorithm in data at $E_\gamma > 3.03\,\mathrm{GeV}$. In data we observe a small peak at the analytical photon energy. As in signal MC a tail to lower energies is visible. In contrast to signal MC we expect a large amount of background for this process. Still, we introduce a constraint, using the analytically calculated energy for all following calculations and correcting the momentum components as well, as shown in Section 2.1.3. Since in data we use the minimization algorithm to achieve the best selection values for σ_d and E_γ we can determine the selection purity and efficiency by Eq. (2.3) and Eq. (2.2), respectively. We observe the same behaviour for the ISR energy selection as in Section 2.1.3. This is shown in Fig. A.5.
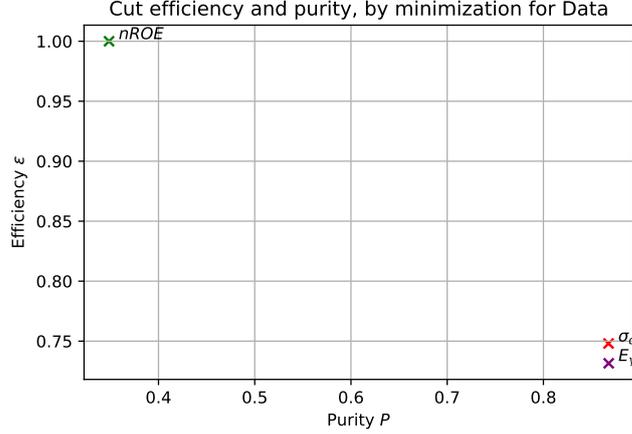
Figure A.5: Selection efficiency ϵ and purity P of K_S^0 candidates, in in the decay process $e^+e^- \to \gamma_{\mathrm{ISR}}\big[J/\psi \to \pi^+\pi^- K_S^0 K_L^0\big]$, without ($n$ROE) and after significance of distance (σ_d) and photon energy selection (E_γ). The error bars are to small to display. The E_γ selection shows only a small impact on the purity as discussed in Section 2.1.3.

With the selections chosen by the minimization algorithm we achieve a selection purity of $P = 86.78\%$ and an efficiency of $\epsilon = 73.15\%$. The E_γ selection is not very efficient as we loose 1% in efficiency

and only gain 0.01% in purity. Nevertheless, to follow the steps of the ϕ-study we decided to keep the selection.

With this we can calculat the missing four momentum vector and thus the missing mass and momentum by using Eq. (2.6). We first take a look at the missing momentum distribution, since we investigate this decay channel for the purpose of finding low momentum K_L^0's. The momentum distribution for signal MC and data is shown in Fig. A.6.
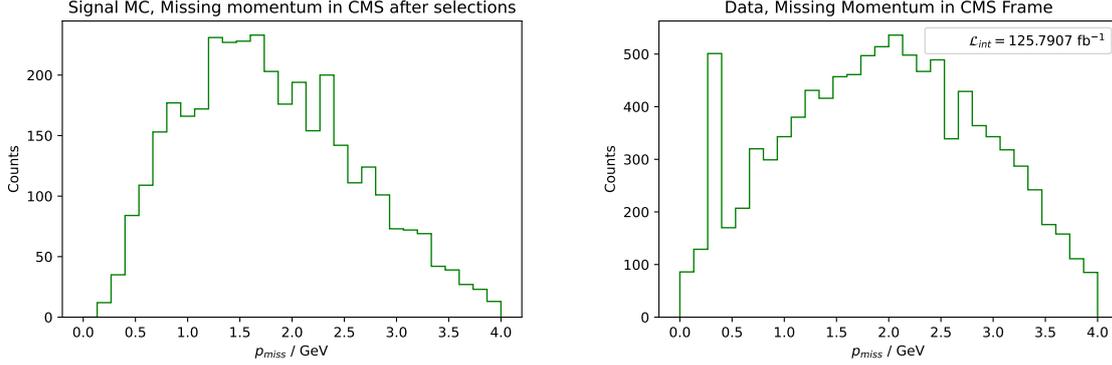
Figure A.6: Distribution of the missing momentum for signal MC (left) and data (right). The momentum is shown in the CMS frame after all our introduced selections. We observe a significant amount low momentum candidates in our selected sample.

We observe a significant number of low momentum K_L^0's in our sample of $e^+e^- \rightarrow \gamma_{\mathrm{ISR}}\left[J/\psi \rightarrow \pi^+\pi^-K_S^0K_L^0\right]$ events. Next we have to investigate the missing mass spectrum, if we can find candidates with a mass of a K_L^0. The histogram is shown in Fig. A.7.
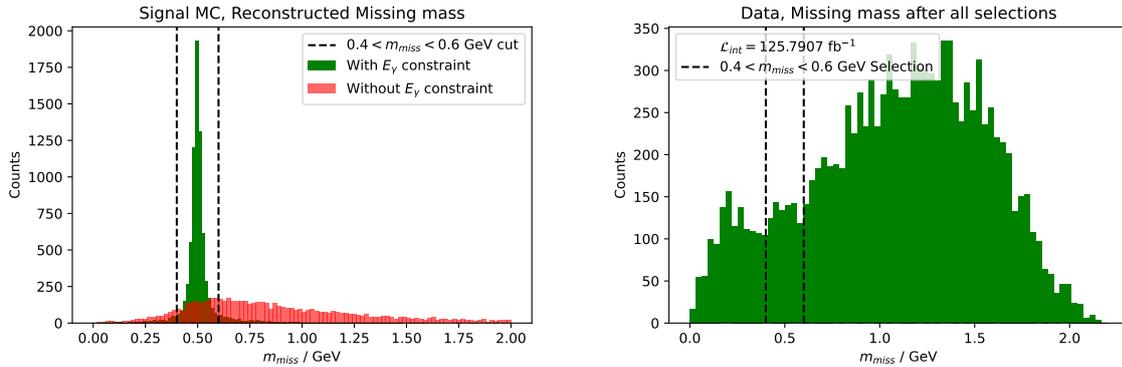
Figure A.7: Distribution of the missing mass for signal MC (left) and data (right). In signal MC we observe a peak at the nominal K_L^0 mass using the photon energy constraint (green). Without the constraint (red) we have no evidence for possible K_L^0 candidates. We select a signal region of potential K_L^0 candidates by eye (dashed line) between $0.4\,\mathrm{GeV} < m_{\mathrm{miss}} < 0.6\,\mathrm{GeV}$. In data we do not observe a peak at the nominal K_L^0 mass. We apply the same signal region selection as in MC.

We observe again a nice peak at the K_L^0 mass in signal MC, if we are using the photon energy constraint.

Thus, we can define a signal region for K_L^0 candidates of $0.4\,\mathrm{GeV} < m_{\mathrm{miss}} < 0.6\,\mathrm{GeV}$, by eye. Meanwhile, in data we do not observe a peak at the K_L^0 mass. Most likely, we have a lot of background in our sample, which overlays our signal candidates that we could not suppress with our defined selections. Still, we define the same signal region as in signal MC, by eye. Nevertheless, we are not able to report of good K_L^0 candidates in the data sample, which makes a BDT-training impossible.

Finally, we have to report that the $e^+e^- \rightarrow \gamma_{\mathrm{ISR}}\left[J/\psi \rightarrow \pi^+\pi^- K_S^0 K_L^0\right]$ channel is a promising decay mode for the K_L^0 identification, as we could show in signal MC. However, since we want to do a data driven BDT-study and the background is dominating any signal, we are not able to use it to identify low momentum K_L^0's. A more detailed study of the background in this channel would be necessary, which however exceeds the scope of this thesis.

A.2 KLM shape variable comparison signal MC and data
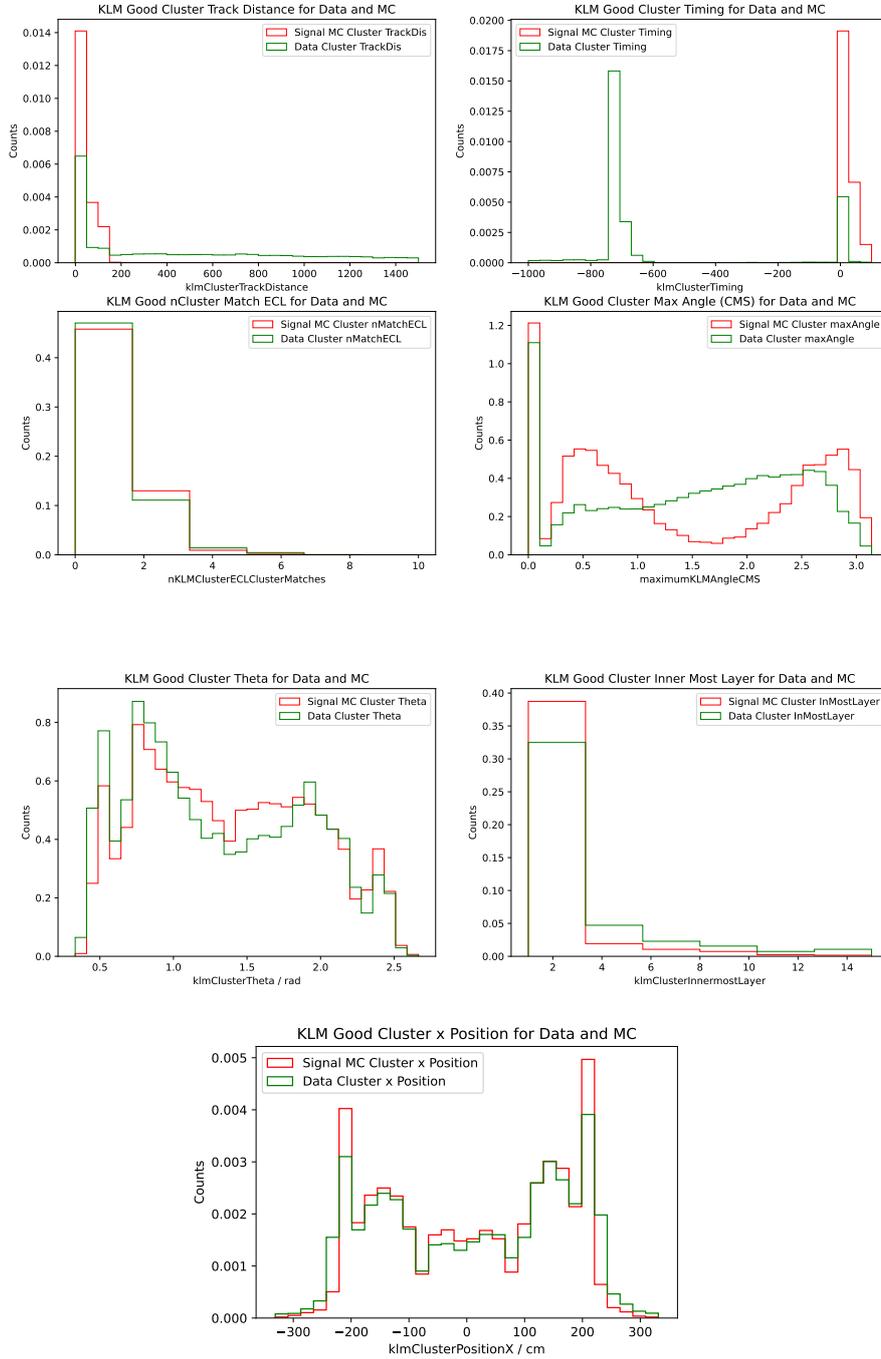
Figure A.8: Comparison of KLM shape variables for signal MC and data in order to validate the assumption that we can do a combined training to predict low momentum K_L^0's.

A.3 Selection Efficiency and Purity J/ψ-Study
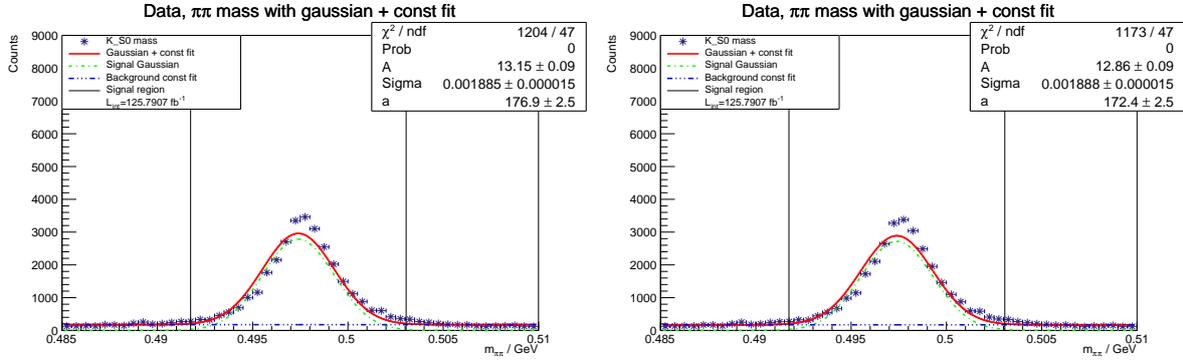
Figure A.9: Distribution of the reconstructed two pion mass with global Gaussian plus constant fit as signal and background (red) for σ_d (left) and E_γ (right) selection., respectively. We plot the signal Gaussian and the background constant separately in green and blue, respectively. The signal region is defined as $\mu \pm 3 \cdot \sigma$ of the signal Gaussian.

A.4 $B \to Dl\nu$ and $D \to \pi K_S^0$
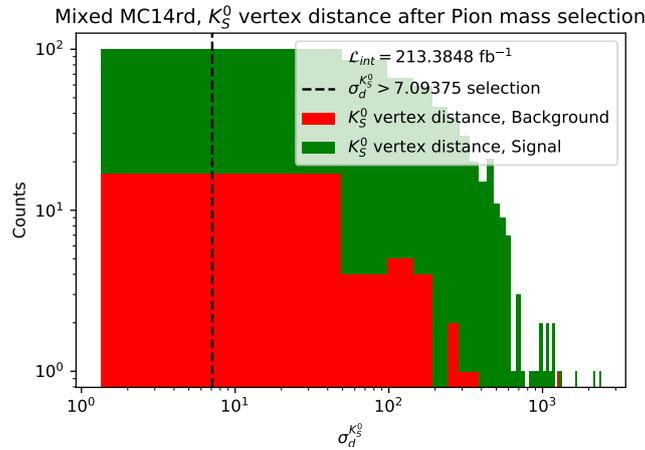
Figure A.10: Significance of distance of the K_S^0 decay vertex to the interaction point for tagged run dependent mixed MC14 of $\int \mathcal{L} = 213.3848\,\mathrm{fb}^{-1}$.
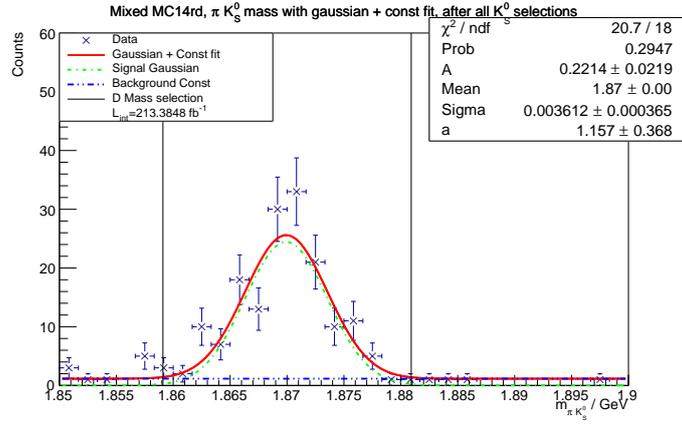
Figure A.11: Distribution of the reconstructed mass of the K_S^0 and pion for hadronic tagged run dependent mixed MC14 of $L_{\text{int}} = 213.3848\,\text{fb}^{-1}$ with K_S^0 selections. A peak at the nominal D mass is observed. We perform a global fit (red) using Eq. (3.4) to define the D-region. The fit result is shown at the top right. We define the D-region by the mean plus/minus three time the width of the Gaussian function.

A.5 BDT Training with ParticleGun signal and data background events
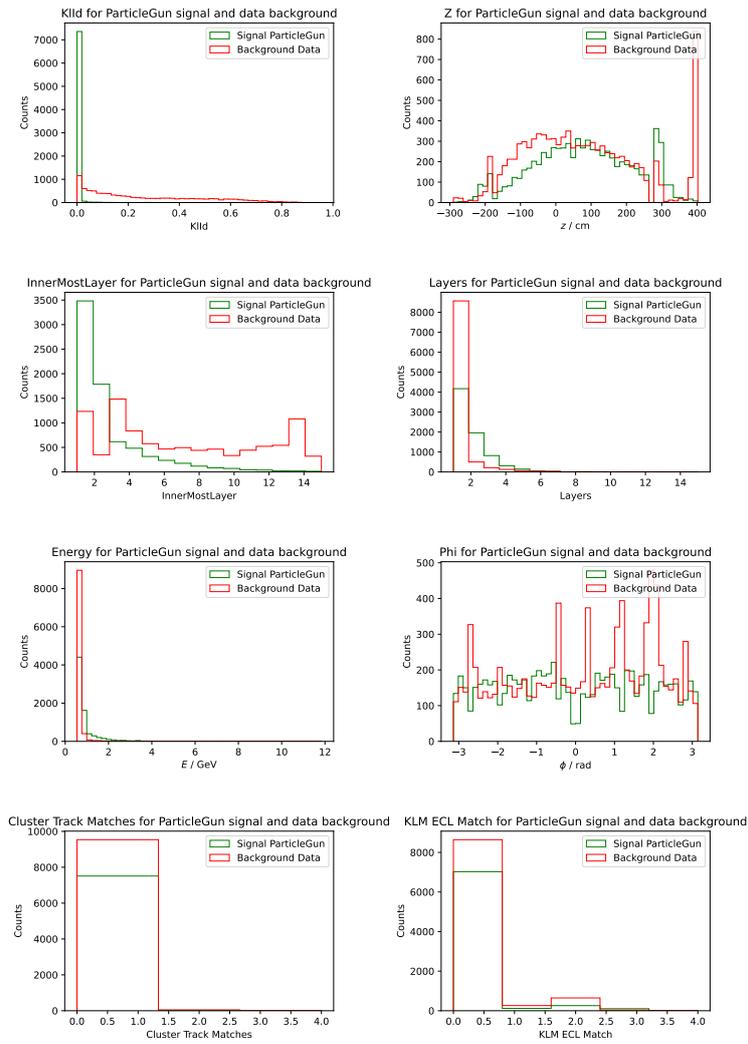
Figure A.12: Distribution for all KLM variables used for the FastBDT training.

Bibliography

[MIs19]   C. MIssMJ, *Standard Model of Elementary Particles*, 2019,
            URL: https://en.wikipedia.org/wiki/File:
            Standard_Model_of_Elementary_Particles.svg (cit. on p. 1).

[Zyl+20]   P. Zyla et al., *Review of Particle Physics*, PTEP **2020** (2020) 083C01, and 2021 update
            (cit. on pp. 2, 3, 7, 8, 11, 17, 33, 36, 39, 41, 42).

[Tho19]   M. Thomson, *Modern Particle Physics*, 5th, Cambridge University Press, 2019
            (cit. on pp. 2, 4).

[Gro22]   H. Group, *Average of $R(D)$ and $R(D*)$ for End of 2021*, 2022, URL: https://hflav-
            eos.web.cern.ch/hflav-eos/semi/fall22/html/RDsDsstar/RDRDs.html
            (cit. on p. 3).

[TW16]   M. Tanaka and R. Watanabe, *New physics contributions in $B \to \pi\tau\bar{\nu}$ and $B \to \tau\bar{\nu}$*, 2016,
            URL: https://arxiv.org/pdf/1608.05207.pdf (cit. on p. 3).

[AFK18]   K. Akai, K. Furukawa and H. Koiso, *SuperKEKB Collider*,
            KEK (High Energy Accelerator Research Organization), 2018,
            URL: https://arxiv.org/pdf/1809.01958.pdf (cit. on p. 4).

[Abe+10]   T. Abe et al., *Belle II Technical Design Report*, KEK (High Energy Accelerator Research
            Organization) Institute of Particle and Nuclear Studies, 2010,
            URL: https://arxiv.org/pdf/1011.0352.pdf (cit. on pp. 4–7, 27).

[Ada+18]   I. Adachi et al., *Detectors for extreme luminosity: Belle II*,
            Nuclear Inst. and Methods in Physics Research, A **907** (2018) 46,
            URL: https://www.phys.hawaii.edu/~teb/BelleII_NIM_special.pdf
            (cit. on pp. 5, 6).

[Aus+15]   T. Aushev et al., *A scintillator endcap K_L and muon detector for the Belle II experiment*,
            National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
            2015, URL: https://arxiv.org/pdf/1406.3267.pdf (cit. on pp. 6, 7, 31, 56).

[Tch+02]   A. Tchouvikov et al., *Using $e^+e^- \to \gamma_{ISR}\phi$ followed by $\phi \to K_S K_L$ for K_L calibration*,
            Belle Note #580 (Internal), 2002 (cit. on pp. 7, 31, 52, 53, 56, 57).

[Kec16]   T. Keck, *FastBDT A speed-optimization and cache-friendly implementation of stochastic
            gradient-boosted decision trees for multivariate classification*, 2016,
            URL: https://arxiv.org/pdf/1609.06119.pdf (cit. on pp. 9, 10).

[Kec14]   T. Keck, *The Full Event Interpretation for Belle II*, 2014, URL:
            https://docs.belle2.org/record/275/files/BELLE2-MTHESIS-2015-001.pdf
            (cit. on pp. 10, 33–35).

[Cor+22] A. S. Cornell et al.,
Boosted decision trees in the era of new physics: a smuon analysis case study, 2022,
URL: `https://arxiv.org/pdf/2109.11815.pdf` (cit. on p. 10).

[com22a] S. community, *scipy.optimize.minimize*, 2008-2022, URL: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html`
(cit. on pp. 12, 14).

[Cas12] D. Casadei, *Estimating the selection efficiency*, 2012,
URL: `https://arxiv.org/pdf/0908.0130.pdf` (cit. on pp. 19, 52).

[Esc+22] J. Eschle, J. Pivarski, E. Rodrigues and H. Schreiner, *Introduction to Vector*, 2019-2022,
URL: `https://vector.readthedocs.io/en/develop/usage/intro.html`
(cit. on p. 22).

[PO19] G. D. Pietro and B. Oberhof, *Report on K_L status*, Physics Performance Meeting, 2019,
URL: `https://confluence.desy.de/pages/viewpage.action?spaceKey=BI&title=Physics+Performance+Meetings+2019&preview=/60630391/118272168/PPM_Klong.pdf` (cit. on p. 24).

[Pii21] L. Piilonen, *KLM Work Plan for Summer 2022*,
February 2021 B2GM Technical Board Meeting, 2021,
URL: `https://indico.belle2.org/event/4201/contributions/20443/attachments/10434/16032/KLM_Summer2022Plans.pdf` (cit. on pp. 25, 26).

[Col22] B. I. Collaboration, *basf2: KLM Cluster and K_L^0 Identification*, 2022,
URL: `https://software.belle2.org/sphinx/release-06-00-03/analysis/doc/Variables.html?highlight=klm#klm-cluster-and-k-l-0-identification` (cit. on pp. 28, 29).

[Kec+19] T. Keck et al., *The Full Event Interpretation*, 2019,
URL: `https://arxiv.org/pdf/1807.08680.pdf` (cit. on pp. 35, 42).

[Kou+18] E. Kou et al., *The Belle II Physics Book*, KEK Preprint, 2018 (cit. on p. 36).

[Cun+20] S. Cunliffe et al., *Physics Cross Sections Table*, 2020, URL:
`https://confluence.desy.de/display/BI/Physics+Cross+Sections+Table`
(cit. on p. 42).

[TKU11] T.Hara, T. Kuhr and Y. Ushiroda,
Belle II Coordinate System and Guideline of Belle II Numbering Scheme, 2011,
URL: `https://indico.mpp.mpg.de/event/2308/contributions/4092/attachments/3414/3799/Belle2NumberingScheme.pdf` (cit. on p. 43).

[com22b] S. community, *scipy.optimize.fsolve*, 2008-2022, URL: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.fsolve.html` (cit. on p. 43).

List of Figures

List of Tables