Karlsruhe Institute of Technology

IEKP-KA/2011-36

Search for $B \to K^{(*)}\nu\bar{\nu}$ Decays Using a New Probabilistic Full Reconstruction Method

Sebastian Neubauer

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Phys. Sebastian Neubauer

aus Schwäbisch Gmünd

Tag der mündlichen Prüfung:	18. November 2011
Referent:	Prof. Dr. M. Feindt
Korreferent:	Prof. Dr. G. Quast

Contents

Part I.

Introduction

1. Introduction

There is an intrinsic dilemma in high energy physics. On the one hand, the ultimate goal is to find a model which describes all observable processes in nature and gives conclusive answers to all remaining questions. On the other hand, such a model would of course question the right to exist for this field of physics. As of today, the situation is even worse. After an explosion of surprising new findings from the 50's to the 70's a visionary called *Standard Model* was able to explain all observations up to that time and was established in the mid 70's. Furthermore, all predictions made by this model are found to be correct inside the errors until today. But for all that we know that this model is not the *theory of everything* as it does not answers all the questions we have. For this reason we are obliged to find experimental deviations from the Standard Model predictions in order to gain information about the next steps towards our final goal.

In particle physics, there are two frontiers where to search for such a deviation. In the high energy frontier one tries to find new heavy particles produced in high energetic particle collisions. In the precision frontier on tries to find quantitative deviations from the Standard Model predictions. Hints for physics beyond the Standard Model will show up indirectly in tiny deviations from the Standard Model predictions. The Belle experiment, and also this thesis, is part of the latter. In the field of B-physics one tries to find deviations in the B meson sector by precisely measuring decay rates, CP-asymmetries and other parameters.

The special setup of the Belle experiment enables a unique method to measure otherwise inaccessible decay channels, called the full reconstruction method. In this thesis such a method was developed introducing a new hierarchical reconstruction approach with extensive usage of multivariate analysis algorithms. With this new method we could double the efficiency compared to an existing cut-based method. For this we reconstruct B mesons in 1104 exclusive decay channels, employing nearly hundred neural networks in a hierarchical structure based on each other, in order to be able to reduce the processing time by many orders of magnitude. This new method is now officially used by the Belle collaboration for many interesting analyses and increases the sensitivity for the search for tiny deviations from the Standard Model predictions.

In the second part of my thesis I performed an analysis using the new full reconstruction method as a hadronic tag. I searched for the decays $B \to K^{(*)}\nu\bar{\nu}$. These decays are very interesting because they are predicted by the Standard Model to be highly suppressed, but the theoretical error on the branching ratio is very small. As we know that contributions from physics beyond the Standard Model are small, these rare decays are interesting, because here, contributions from the Standard Model and from new physics can be in the same order of magnitude. Depending on the new physics model, the branching ratio of such a rare decay can even be increased by several orders of magnitude. Because of the two neutrinos in the final state of the decays $B \to K^{(*)}\nu\bar{\nu}$ it is not possible to measure the quantities of these decays with conventional reconstruction methods. It is only possible using the full reconstruction method, where the missing information from the escaping neutrinos is compensated by fully reconstructing the entire event. The tag-side is reconstructed using the full reconstruction tool and the rest of the event is used as the signal-side $B \to K^{(*)}\nu\bar{\nu}$ decays in several modes.

It turned out that there is a significant improvement on the expected upper limit of the branching ratio using the new full reconstruction method on the tag-side, in comparison to the predecessor

analysis by the Belle collaboration utilizing the classical cut-based full reconstruction method. As the same cuts are applied and the increased data statistics is only 20%, this improvement arises mainly due to the higher efficiency of the new full reconstruction method.
Another significant improvement could be achieved by utilizing NeuroBayes instead of the cut-based selection on the signal-side of the analysis. At the same time, a higher efficiency and a higher purity of the selection could be achieved. The balance between purity and efficiency was optimized in order to reach the most stringent expected upper limit on the branching ratio.

Part II.

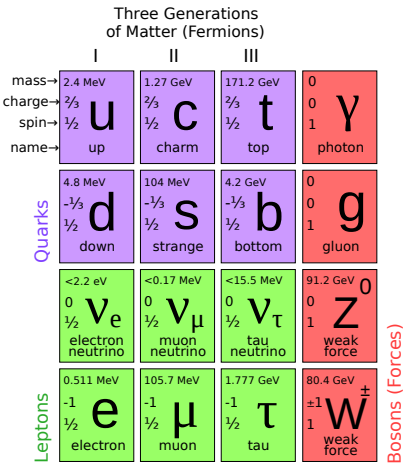
Theoretical Foundations

2. The Standard Model of Particle Physics

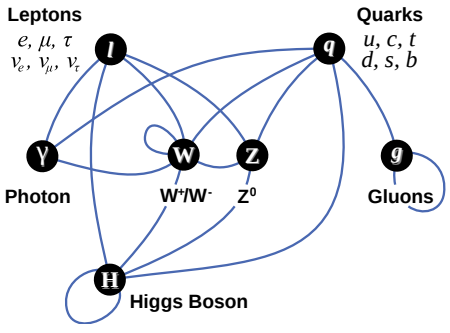
In the last century a new field of physics emerged after the observation of new particles. Only the stable particles proton, neutron and electron were known until the discovery of the unstable muon in 1936 and the charged pion in 1946. Already in 1935, Yukawa predicted the existence of the pion as a massive exchange boson, binding neutrons and protons in the nucleus [Yuk55]. Since then, many new findings lead to a well-established model of particle physics, the so-called Standard Model. This model currently describes all observations and measurements within the uncertainties.

2.1. Structure of the Standard Model

The Standard Model is a relativistic quantum field theory, describing the interactions between the elementary matter particles. All the elementary matter particles are fermions, interacting among themselves by the exchange of gauge bosons. The Standard Model describes all known fundamental forces, electromagnetic force, weak force and strong force. Only gravity is currently not described by the Standard Model. The electromagnetic and the weak force are unified to the electroweak force. There are two types of elementary matter particles, leptons and quarks. Whereas the quarks couple to all forces, the leptons do not couple to the strong force. There are six quarks and six leptons, both grouped in 3 generations (see figure 2.1a).



(a) The elementary particles of the Standard Model.



(b) The interactions between the elementary particles of the Standard Model.

Figure 2.1.

Mathematically, the Standard Model is described by the local gauge symmetry $SU(3) \times SU(2) \times U(1)$. The three gauge groups represent the three fundamental forces. The $U(1)$ represents the electromagnetic force with the weak hyper-charge, $SU(2)$ the weak force with the weak isospin and the $SU(3)$ the strong force with the color charge.

2.1.1. The Leptons

Leptons are elementary particles that do not couple to the strong force. There are six types of leptons, known as flavors, three charged particles, the electron, the muon, the tau and three neutral particles, the neutrinos. The six leptons form three generations containing each a pair of a charged lepton and a neutrino (see table 2.1a). Because the weak force only couples to left-handed particles and right-handed anti-particles, these generations are weak isospin doublets for the left-handed leptons. The right-handed particles are grouped in isospin singlets.
The first lepton, discovered by Thomson in 1897 is the stable electron with a mass of 511 keV. The muon, discovered by Anderson and Neddermeyer at Caltech in 1936 was at first thought to be the π-meson predicted by Hideki Yukawa but then identified as a lepton. It is unstable with a lifetime of $2.2 \cdot 10^{-6}$ s and a mass of 105.7 MeV. Because of the high mass of 1777 MeV, the tau lepton was discovered by Perl et al. not until 1975 [Per+75]. This discovery was a big surprise, because there wasn't any hint of a third generation of leptons up to that time.
The existence of neutral leptons which only interact weakly was postulated in 1930 by Wolfgang Pauli in order to explain the continuous electron spectrum of β-decays as a three body decay. The existence was then confirmed by Cowan, Reines, Harrison, Kruse, and McGuire in 1956 [Cow+56]. Figure 2.2 shows the first event showing a neutrino that hits a proton in a hydrogen bubble chamber. The neutrinos were always thought to be massless. But then, the Super-Kamiokande experiment found evidence that neutrinos can oscillate [Fuk+98]. That implies, the neutrinos must have a non-zero mass. It is still unknown, whether neutrinos are their own anti-particles, like the photon. If they would be their own anti-particles, they are called Majorana neutrinos. If the neutrinos would be Majorana neutrinos, one could for example observe the *neutrino-less double-beta decay*. There are several experiments searching for this decay. In 2006 Klapdor-Kleingrothaus et al. claimed to have found a more than 6σ evidence for this decay [KK+06], but as the results are heavily contested [Aal+02] the question is still not answered.

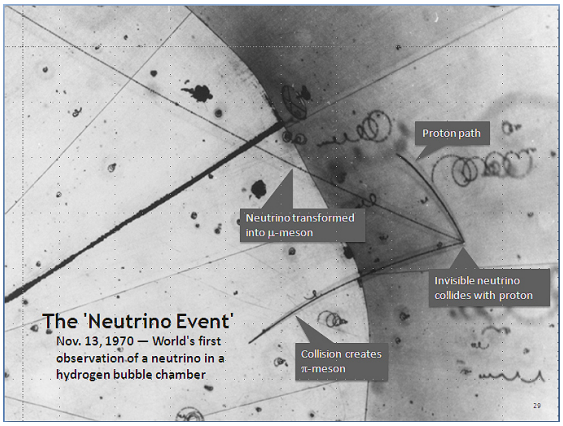


Figure 2.2. The first observation of a neutrino in a hydrogen bubble chamber. (Image courtesy of Argonne National Laboratory)

2.1.2. The Quarks

In 1964, Murray Gell-Mann proposed the existence of quarks as constituents of mesons and baryons [GM64]. He found out that, if mesons are built up of $(q\bar{q})$ constituents and baryons out of (qqq) one could explain many observations in a natural way. Subsequently, this model was confirmed mainly in deep elastic proton scattering experiments. The most convincing argument was the finding that the structure functions do not depend on Q^2 and therefore the proton has to have point-like constituents, called partons, or later quarks. Today we know that there are six quarks grouped in 3 families. They are spin 1/2 particles and the up-type quarks carry electrical charge of $+2/3$ and the down-type quarks $-1/3$. They also carry weak isospin and therefore couple to the weak force. They are the only matter particles carrying color charge and therefore interact with the strong force. Because of the properties of the strong force (see section 2.2.2) quarks are always enclosed in colorless hadrons. It is not possible to observe free quarks (one could argue that the top quark is free, but it is only because it decays before it can build a bound state). The charm quark was discovered in 1974 simultaneously by Richter [Aug+74] and Ting [Aub+74] in the famous *November Revolution* with the discovery of the J/ψ resonance. This resonance was found to be a $c\bar{c}$ state.
The b quark was discovered in 1977 by the Fermilab E288 experiment, led by Leon M. Lederman [Her+77]. Similar to the discovery of the J/ψ, they discovered a $b\bar{b}$-resonance at 9.5 GeV called Υ.
This was very exciting because it was not known before that there are 3 families of quarks. But knowing that it was obvious that there has to be also a sixth quark called the top quark. The prediction of the existence was indeed correct, but not the prediction of the mass of this quark. When it was discovered in 1995 at the Fermilab (once again published in the same journal issue by two groups CDF [Abe+95] and D$\emptyset$ [Aba+95]) with a mass of 173 GeV it was much heavier than expected.

2.2. Forces of the Standard Model

Forces are described in the Standard Model by the exchange of bosonic force particles. There are four known fundamental forces, the electromagnetic force, the weak force, the strong force and gravity. The gravitational force is not included in the Standard Model but can also be described by the exchange of a spin 2 *graviton*. As the electro-magnetic force is unified in the Standard Model with the weak force, there are only two forces left in the Standard Model, the electro-weak force and the strong force.

2.2.1. The Electroweak Force

Einstein was able to explain the photo-electrical effect by postulating that light consist of quantum particles. Later, with quantum mechanics, Dirac was able to formulate a quantum field theory for the electro-magnetic field, called quantum electro-dynamics. In the beginning, the weak interaction was described by the V-A theory which accounted for the fact that the weak interaction only couples to left-handed particles (and right-handed antiparticles) which was discovered in 1957 by Wu et al. [Wu+57] in the beta decay of cobalt-60. But it was known that this theory doesn't behave very well at high energies. There was a theory by Sheldon Glashow, Abdus Salam and Steven Weinberg unifying the electro-magnetic and weak theory. As the weak force is much weaker than the electro-magnetic force, this symmetry needs to be broken. This spontaneous symmetry breaking rotates the electro-weak gauge bosons into a massless purely electromagnetic photon and the heavy $W^{\pm}$ and Z^0 bosons. The confirmation of this theory was

the discovery of the neutral weak currents, mediated by the Z^0 boson in neutrino scattering by the Gargamelle collaboration in 1973 and the discovery of the $W^{\pm}$ and Z^0 bosons itself in 1983. According to this theory, the weak force is not weaker because of a smaller coupling constant, but it is only weaker due to the huge mass of the gauge bosons. Actually it is not possible to assign a mass to gauge bosons in a quantum field theory. In order to be able to assign a mass to the $W^{\pm}$ and Z^0, currently the Higgs mechanism is used. One introduces a scalar Higgs field doublet, which breaks since a phase transition in the early universe the electro-weak symmetry spontaneously by obtaining a non-zero vacuum expectation value. It is broken such that it gives a mass to the three $W^{\pm}$ and Z^0 bosons, while the remaining degree of freedom is the physical Higgs particle which is not discovered yet. This Higgs field also couples to the fermions through the Yukawa coupling and gives them a mass. The stronger this coupling is, the higher is the mass of the fermion.

By measuring the decay rates of pions, kaons, muons, it was found that kaon decay rates are four times lower. This fact could be explained by Cabibbo in 1963 [Cab63]. He introduced a mixing angle, the *Cabibbo angle*, which rotates the up quark with respect to the down type quarks u, s. By choosing the right angle, one could explain the suppression of the kaon decays.

Another unexplainable observation was the lack of flavor changing neutral currents that is to say transitions from a strange quark to a down quark. This could be explained by the work of Glashow, Iliopoulos and Maiani [Gla+70] in 1970. By postulating a fourth quark in the second quark family, these flavor changing neutral currents are highly suppressed because of cancellations. While it was known that the weak interaction violates parity, it was thought that

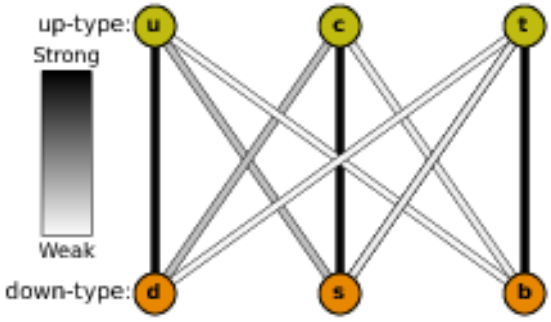


Figure 2.3. Pictorial description of quark weak interactions. The color of the lines between the quarks (u, c, t, d, s, b) indicates the strength of the weak coupling between these flavors. I.e. they represent the magnitudes of the components of the CKM matrix. The CKM matrix values were obtained from [Nak+10].

the combined CP is conserved. But then surprisingly, in 1964, James Cronin and Val Fitch discovered a clear CP violation in the neutral kaon system. There were several attempts to solve this problem, but only one was proven to be right many years later. In 1973 Makoto Kobayashi and Toshihide Maskawa pointed out [Kob+73], if one extends the Standard Model by a third quark generation, the resulting unitary 3×3 mixing matrix has a non-vanishing imaginary part which can give rise to CP violation. Their theory not noticed till the discovery of the b quark (and therefore the third family), was finally confirmed with the discovery of CP violation in the neutral B meson system by Belle [Abe+01] and BaBar [Aub+01] in 2001, which was amazingly

accurately predicted by this theory. This complex *CKM matrix* [Nak+10] is unitary

$$\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.00347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{bmatrix},$$

which gives additional constraints for the measured values of the matrix elements. These constraints can be visualized in the complex plane where the matrix is parametrized in the *Wolfenstein parametrization*

$$\begin{bmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{bmatrix},$$

which is shown in figure 2.4. If the Standard Model is correct, all constraints must be fulfilled that is the triangle in figure 2.4 must be closed exactly. To verify this or to find significant deviations is one of the mayor fields of B-physics. The so-called e^+e^- *B-factories* PEP-II in the USA and KEKB in Japan were mainly built to investigate this field. Furthermore there is LHCb at the LHC and the plans for an upgrade of the Belle experiment to the Belle II experiment with the ambitious goal to accumulate fifty times the data of the predecessor experiment.

2.2.2. The Strong Force

In order to explain the bindings of protons and neutrons in the nucleus, Yukawa formulated his theory of the exchange of massive particles (pions) which results in an attractive potential, analogous to the van der Waals forces between neutral atoms. Later with the discovery of the quarks, it was a severe problem to interpret the Δ^{++} resonance as a baryon, containing three up-quarks with parallel spins as it violates the Pauli principle. This problem was (once again) solved independently in 1965 by Moo-Young Han with Yoichiro Nambu and Oscar W. Greenberg. By introducing a new SU(3) gauge symmetry, the quarks gain enough degrees of freedom in order to not violate the Pauli principle any more. At the same time this new symmetry gives rise to a new force, carried by new gauge particles. From formal considerations, there have to be eight of these particles, the gluons. The properties of this interaction is described by quantum chromodynamics in dependence on the three charges, referred to as colors.

The main difference to the other forces is that the gluons are carrying color charge and therefore are interacting with themselves. This gives rise to some unusual properties of the strong force. In the electro-magnetic force, the virtual photons around a charge can produce e^+e^- pairs which are attracted or repelled depending on the sign of the charge. This results in a screening of the charge, that is, the coupling seems to be stronger with smaller distances. In the strong force, this effect is turned around into an anti-screening effect due to the self coupling of the gluons. The coupling becomes weaker with smaller distances. This effect is called the asymptotic freedom and was discovered by Frank Wilczek [Gro+73].

The other unusual effect is the linear increase of the strong potential with large distances. This implies that, with increasing distance of two color charges, the energy in the field rises until the energy is sufficient to create a quark anti-quark pair out of the vacuum. This is the reason why we cannot observe free color charges in nature and also the reason for the typical jet-like structure in high energy collisions, called the hadronization.

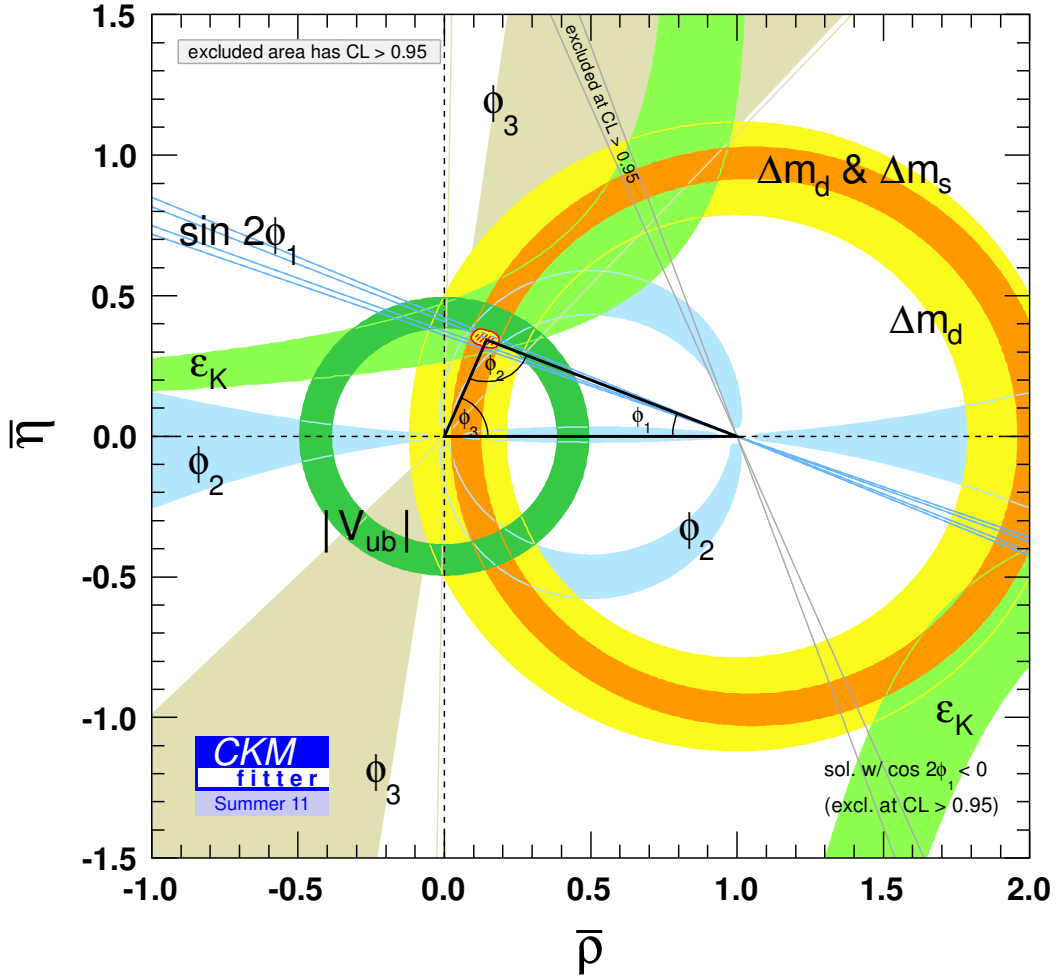


Figure 2.4. Constraints in the $(\bar{\rho},\bar{\eta})$ plane including the most recent ϕ_2 and ϕ_3 related inputs in the global CKM fit. The red hashed region of the global combination corresponds to 68% CL. (Preliminary results as of Summer 2011) [CKM05]

2.3. Beyond the Standard Model

There are several arguments why the Standard Model cannot be the correct description of the nature even as it describes all current observations.

- Baryon anti-baryon asymmetry
  Our universe consists of matter and not of anti-matter. The Standard Model cannot describe this impressive discrepancy.
- Hierarchy problem
  The weak force is 10^{32} times stronger than gravity. Expressed in other words, the Higgs boson is way too light compared to theory expectation. This discrepancy has to be "fine-tuned" in the calculations of quantum corrections and is therefore a weak point of the Standard Model.
- Massive neutrinos
  The fact that neutrinos can oscillate implies that they have a non-zero mass. Even if this could be added to the Standard Model, a satisfactory explanation for the mass is still unknown.
- Dark matter
  We know from rotation curves of galaxies that there has to be much more invisible dark matter in the universe than we can explain. The Standard Model has no satisfactory explanation for the origin of the dark matter.
- Energy scale
  Some interactions are diverging for energies much higher than we can currently observe. It is therefore obvious that there needs to be something beyond the Standard Model which canceles these divergencies.

There are of course plenty of theories going beyond the Standard Model. One of the most seriously discussed is supersymmetry (SUSY). In this theory each elementary fermion has a bosonic supersymmetric partner and each boson has a fermionic supersymmetric partner as shown in figure 2.5. As there were not found any of these supersymmetric partners yet, they

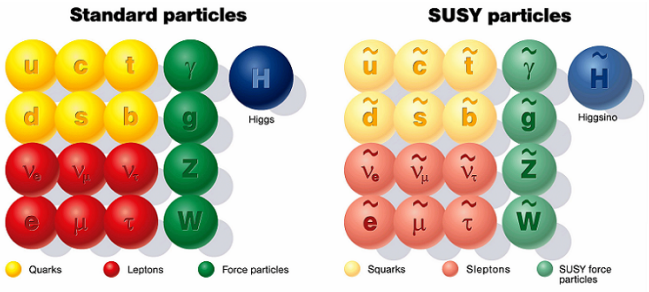


Figure 2.5. The particles of the supersymmetric theory.

have to be extremely heavy. Therefore this symmetry between fermions and bosons needs to be broken. Quite a lot of the arguments above can be naturally explained by this theory. Just as an example, in order to keep the proton stable one has to introduce the R-parity which needs to be conserved. This immediately gives rise to a lightest stable supersymmetric particle which is then a natural candidate for dark matter.

2.4. Rare B Decays

So-called rare B-decays are a powerful probe for physics beyond the Standard Model. One can either try to produce new, heavier particles in highest energy experiments, or one exploits the fact that heavy particles can emerge as virtual particles in decay amplitudes. As they are much heavier than other known particles, their contribution will be in general unmeasurable tiny compared to the Standard Model process.
But there are decays in the Standard Model which are heavily suppressed or even completely forbidden. A prominent example of such decays are flavor changing neutral currents (FCNC). They are extremely suppressed in the Standard Model because they can only decay via so-called penguin diagrams as shown in figure 2.6, a flavor changing neutral current decay of a b quark. While this process is suppressed in the Standard Model it could be allowed in physics beyond

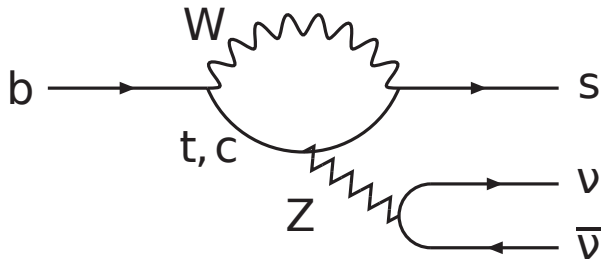


Figure 2.6. An example of a penguin diagram. It shows a flavor changing neutral current decay of a b quark.

the Standard Model. Not withstanding the high-mass suppression, it could be compatible or even larger than the Standard Model process. The relative discrepancy where experiments are sensitive to, is therefore possibly large. Some decays of the shown transition will be analyzed in part V. These are the decays $B \to K^{(*)}\nu\bar{\nu}$ which are very well suited for probing new physics models, as discussed in part V.
The importance of these rare decays is the major motivation for the upgrade of the Belle experiment to the Belle II experiment with the goal of increasing the luminosity by a factor of 50.
The improvement of the full reconstruction algorithm explained in part IV is a huge step forward towards the exciting sensitivity frontier of probing physics beyond the Standard Model.

3. Parameter Estimation

One fundamental task of statistical inference is the estimation of parameters. These parameters are not measured directly, but only through the data, distributed according to the probability density function defined by the true parameters. The more accurate the underlying probability density function can be estimated, the more accurate the true parameters are known. This chapter is about parameter estimation in counting experiments.

3.1. Counting Experiments

In particle physics, one often has to deal with counting experiments. This means that one counts the occurrence of statistically independent events in a given time interval. Such a process can be described as a Bernoulli-process. A Bernoulli-process is a chain of independent discrete trials, each with two possible outcomes, success and no success, with a fixed probability p. After n trials, the total count of successes $k = 0, \ldots, n$ is described by the binomial distribution:

$$P(r) = \binom{n}{k} p^k q^{n-k} \ . \tag{3.1}$$

Mostly, in physical experiments we do not have discrete trials, but infinitesimal small time intervals in which an event occurs or not with a very small probability p. In that case, the binomial distribution converges to the Poisson distribution:

$$P_\lambda(k) = \frac{\lambda^k}{k!} \, \mathrm{e}^{-\lambda} \ , \tag{3.2}$$

where λ is a parameter which describes the mean and the variance of the Poisson distribution at the same time. The distributions for three different parameters λ are shown in figure 3.1. Often, in physics experiments we are interested in the unknown amount of signal events N_S and in addition we have a certain amount of expected background events N_B. What we can observe is only the sum of both

$$N_O = N_S + N_B \ , \tag{3.3}$$

where N_O is the observed number of events. Those two types of events are independent of each other and therefore N_O is distributed as the sum of two Poisson distributions with the parameters

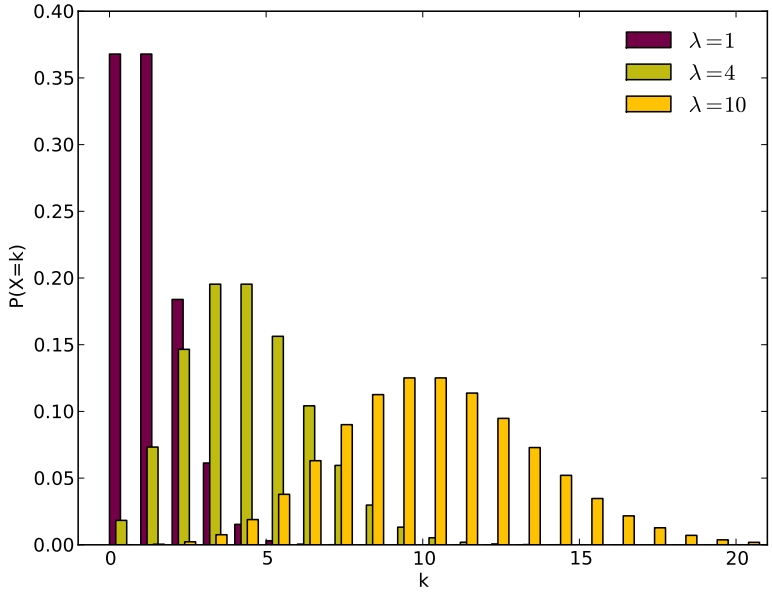


Figure 3.1. The Poisson distribution for three different parameters λ.

λ_S and λ_B. As

$$P(N_O = N_S + N_B) = \sum_{k=0}^{N_O} P(N_S = k)\, P(N_B = N_O - k) \tag{3.4}$$

$$= \sum_{k=0}^{N_O} \frac{\lambda_S^k}{k!}\, \mathrm{e}^{-\lambda_S}\, \frac{\lambda_B^{N_O-k}}{(N_O-k)!}\, \mathrm{e}^{-\lambda_B} \tag{3.5}$$

$$= \frac{1}{N_O!}\, \mathrm{e}^{-(\lambda_S+\lambda_B)} \sum_{k=0}^{N_O} \binom{N_O}{k} \lambda_S^k\, \lambda_B^{N_O-k} = \frac{(\lambda_S+\lambda_B)^{N_O}}{N_O!}\, \mathrm{e}^{-(\lambda_S+\lambda_B)} \tag{3.6}$$

shows, the sum of two Poisson distributions is once again a Poisson distribution with $\lambda_O = \lambda_S + \lambda_B$. The expectation for the background λ_B must be known, either from Monte Carlo simulations, or from data sidebands. Therefore, as we can treat λ_B as constant, we can estimate λ_S by maximizing the likelihood function

$$\mathcal{L}(N_O|\lambda_S) = \frac{(\lambda_S+\lambda_B)^{N_O}}{N_O!}\, \mathrm{e}^{-(\lambda_S+\lambda_B)}\ , \tag{3.7}$$

as we have typically only one measurement. Our best estimator $\hat{\lambda}_S$ is then just

$$\hat{\lambda}_S = N_O - \lambda_B\ . \tag{3.8}$$

The main issue is now to find the correct uncertainty of this point estimation. For this we first have to discuss what uncertainty denotes. There are two possible approaches to define uncertainty, one from frequentist statistics and one from Bayesian statistics.

3.1.1. Frequentist Approach

In frequentist statistics, the parameters which have to be estimated are unknown, but they have fixed values. It is not possible to treat them as random variables. Therefore, it is not possible to assign probabilities to these parameters. If we define an arbitrary interval $[s_1, s_2]$ in one parameter space, then the probability is either 0 (the true parameter lies outside this interval), or 1 (the true parameter lies inside the interval). The only valid interpretation is, if we repeat that experiment over and over again, we can give a frequentist probability of how often that interval included the true parameter. For a frequentist definition of uncertainty we need to define such an interval with appropriate features. It makes sense to define such an interval in a way that the probability that the constructed interval contains the true value, is given by a certain level $1 - \alpha$. If this interval is determined with the Neyman construction (see e.g [Ney37; Nak+10]), it is called a confidence interval with the confidence level $1 - \alpha$. With the probability density function $f(N|\lambda_S)$ we can find a set of $[N_1(\lambda_S, \alpha), N_2(\lambda_S, \alpha)]$ for each λ_S and a pre-defined α such that

$$P(N_1 < N < N_2, \lambda_S) = 1 - \alpha = \int_{N_1}^{N_2} f(N|\lambda_S) dN \ . \tag{3.9}$$

This is illustrated in figure 3.2. The intervals $[N_1, N_2]$ (denoted in the figure as $[X_1, X_2]$) are illustrated as horizontal lines for each λ_S (denoted as Θ) in figure 3.2. The set of all parameter intervals is called the confidence belt. For a given observation N_O (X in the figure) the confidence

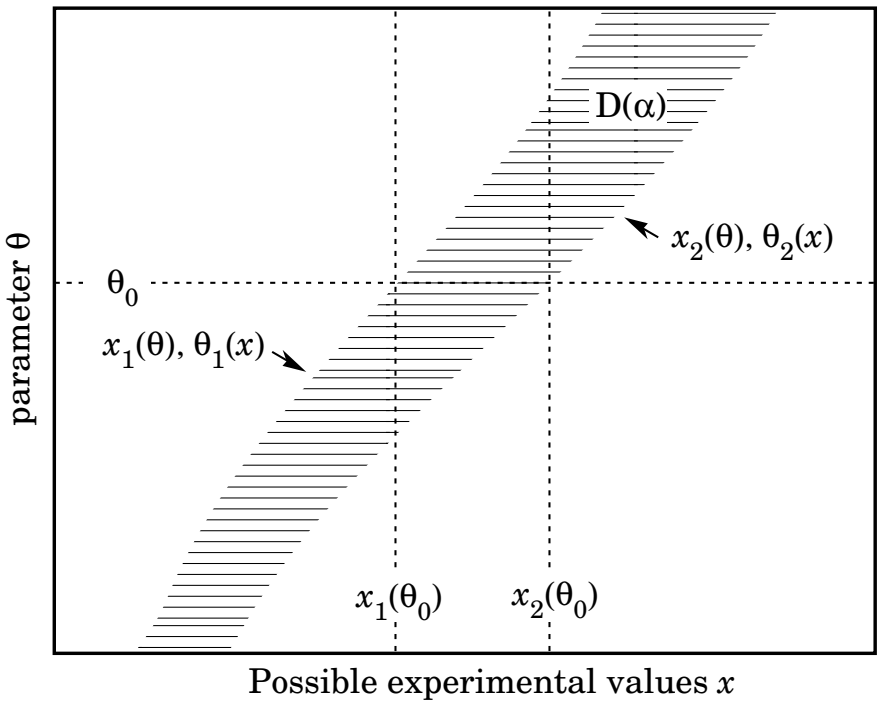


Figure 3.2. Illustration of the Neyman construction (from [Nak+10]).

interval is defined as the set of all parameters whose horizontal intervals intercept the vertical line of this measurement. This set of parameters needs not to be at a stretch because it is only

defined as the intersection with the confident belt which can be bent. Its boundaries are the parameters $\lambda_1(N_O)$ and $\lambda_2(N_O)$. If the experiment would be repeated a large number of times, the interval $[\lambda_1(N_O), \lambda_2(N_O)]$ would vary, covering the fixed value λ_{true} in a fraction $1-\alpha$ of the experiments. This is called coverage and is fulfilled for intervals determined with the Neyman construction. The coverage condition in eq. 3.9 does not determine the N_1 and N_2 completely, but only the enclosed probability. In [Ney37] it is suggested to take a central interval with $\alpha/2$ probability on each side or for an upper (lower) limit, take $N2 = \infty$ ($N1 = -\infty$).

Feldman-Cousins Method

As reported in [Fel+98] there is a problem with the coverage in the changeover from calculating a central interval instead of a previously planned limit after observing the data ("flip-flopping"). Feldman and Cousins introduced a different ordering principle to choose the interval $[N_1, N_2]$ in equation 3.9. This is illustrated with a counting experiment in the presence of background described by the likelihood function equation 3.7. For each possible true parameter λ_S and each possible experimental result N we can calculate the likelihood ratio:

$$R = \frac{f(N|\lambda_S)}{f(N|\hat{\lambda}_S)} \,, \tag{3.10}$$

where $f(N|\lambda_S)$ is the likelihood of that measurement given the true parameter λ_S and $f(N|\hat{\lambda}_S)$ the likelihood of that measurement given the parameter with the highest likelihood $\hat{\lambda}_S$. Each possible experimental result is then ranked according to this ratio and added to the acceptance region until the probability of the entire acceptance region supersedes the desired confidence level. This is shown in table 3.1 for $\lambda_B = 3$ and one parameter $\lambda_S = 0.5$. This calculation has to

Table 3.1. Illustrative calculation of the acceptance region for $\lambda_S = 0.5$ and known background $\lambda_B = 3$. The measurement with an entry in the rank column are added to the acceptance region.

N	$f(N\|\lambda_S)$	$\hat{\lambda_S}$	$f(N\|\hat{\lambda}_S)$	R	rank
0	0.03	0	0.05	0.607	6
1	0.106	0	0.149	0.708	5
2	0.185	0	0.224	0.826	3
3	0.216	0	0.224	0.963	2
4	0.189	1	0.195	0.966	1
5	0.132	2	0.175	0.753	4
6	0.077	3	0.161	0.48	7
7	0.039	4	0.149	0.259	
8	0.017	5	0.14	0.121	
9	0.007	6	0.132	0.05	
10	0.002	7	0.125	0.018	
11	0.001	8	0.119	0.006	

be repeated for a sufficient number of λ_S to reach the desired granularity. The confidence region is then defined as the set of all λ_S where the observed value N_O was in the acceptance region. A confidence belt constructed with the Feldman-Cousins method for this example can be seen in figure 3.3.

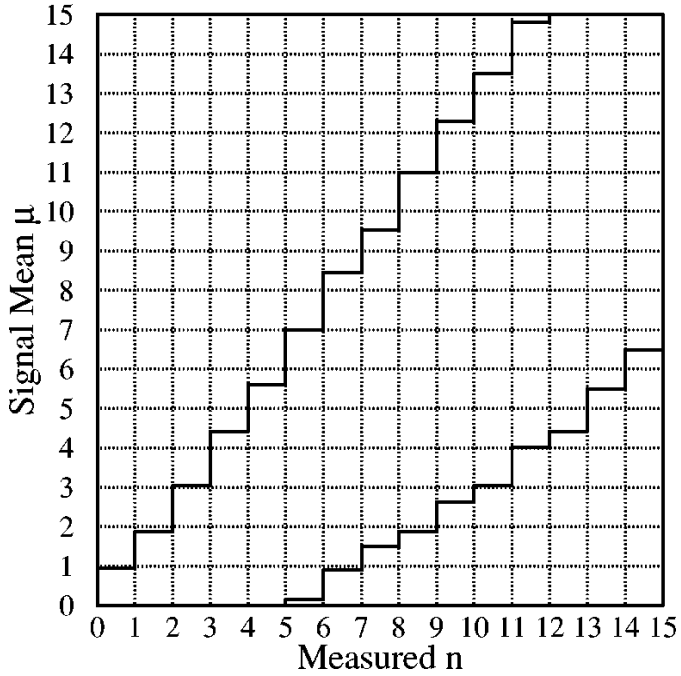


Figure 3.3. Confidence belt based on the Feldman-Cousins ordering principle, for 90% C.L. confidence intervals for unknown Poisson signal mean λ_S in the presence of a Poisson background with known mean $\lambda_B = 3$ (from [Fel+98]).

Profile Likelihood Method

This method can be generally used in the presence of nuisance parameters in the likelihood function. The main idea is to divide the parameters in two groups, the the parameter-of-interest (POI) $\vec{\lambda}$ and the nuisance parameters $\vec{\omega}$. Now we can fix $\vec{\lambda}$ and reduce the likelihood $L(\vec{x}|\vec{\lambda},\vec{\omega})$ to $L(\vec{x}|\vec{\omega})$ and maximize it. Now we repeat this for all values of $\vec{\lambda}$. Our best estimate and an estimation of the confidence interval can be deduced by the maximal likelihood function values with respect to $\vec{\omega}$ in dependence of $\vec{\lambda}$. Hence, we have profiled out the nuisance parameters $\vec{\omega}$. But this method also meets with criticism as it is only an approximation and theoretically not well funded (see e.g. [Mon+09]).

3.1.2. Bayesian Approach

As described above, in the frequentist approach, the true parameters are unknown, but fixed. The parameters therefore cannot be treated as random variables. In contrast, in the Bayesian approach the parameters are treated as random variables. Therefore it is possible to calculate the posterior density function for the parameters with the Bayes Theorem:

$$f_\Theta(\theta \mid X) = \text{constant} \cdot f_\Theta(\theta)\mathcal{L}(X \mid \theta), \tag{3.11}$$

where $\mathcal{L}(X \mid \theta)$ is the likelihood function, $f_\Theta(\theta)$ is the prior distribution of the parameters and constant is a normalization constant that ensures that the posterior distribution $f_\Theta(\theta \mid X)$ is a proper probability distribution. This is the conditional probability density of the parameters given this measurement X. It is not necessary to calculate a p-value or construct a confidence interval with the Neyman construction or the Feldman-Cousins method. One can either give the whole probability density function as a result, or the median, the mean, or any arbitrary quantile, known as Bayesian credibility interval. It is important to note that this credibility interval is not the same as the frequentist confidence interval. In contrast to the frequentist confidence interval, the Bayesian credibility interval can be interpreted as the probability that the true parameter lies inside this interval.

But there are two major problems of the Bayesian approach. The first one is that one needs a prior distribution to calculate the posterior density. That prior changes in general the result of the measurement. Sometimes one can use already existing measurements as a prior, but then it is not possible to average both measurements anymore. Another possibility is to put objective information, like physical boundaries into the prior distribution. But in general one would avoid to bias the result in any way. For this, one needs an uninformative prior, which is not always a uniform prior. There are many approaches to construct such a prior. For example the Jeffreys prior [Jef46], which is invariant under re-parameterization and constructed as the square root of the determinant of the Fisher information:

$$p(\theta) \propto \sqrt{I(\theta)} = \sqrt{\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2 \middle| \theta\right]}\,. \tag{3.12}$$

As an example, the Jeffreys prior for the rate parameter λ of a Poisson distribution

$$f(n|\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}\,, \tag{3.13}$$

is

$$p(\lambda) \propto \sqrt{I(\lambda)} = \sqrt{\mathrm{E}\left[\left(\frac{d}{d\lambda}\ln f(x|\lambda)\right)^2\right]} = \sqrt{\mathrm{E}\left[\left(\frac{n-\lambda}{\lambda}\right)^2\right]} \tag{3.14}$$

$$= \sqrt{\sum_{n=0}^{+\infty} f(n|\lambda)\left(\frac{n-\lambda}{\lambda}\right)^2} = \sqrt{\frac{1}{\lambda}}\,. \tag{3.15}$$

The second major problem is that in general one is interested in one parameter, called the parameter-of-interest (POI) but the likelihood and the prior may depend on many more parameters, the so called nuisance parameters, for example background shape parameters. In order to obtain the posterior distribution of the POI, one needs to get rid of the nuisance parameters. Therefore one has to integrate

$$f_\Theta(\theta)\mathcal{L}(X \mid \theta), \tag{3.16}$$

over all nuisance parameters, which might be impossible analytically, depending on the actual prior and likelihood function.

3.1.3. Markov-Chain Monte Carlo Method

In the last ten years, the Markov-chain Monte Carlo method (MCMC) has received an increasing popularity due to increasing computational power. Instead of integrating the Bayesian posterior distribution, one creates a Monte Carlo sample, distributed according to the posterior probability density function (PDF). Getting the distribution of the POI is then called marginalization, which just means to take the projection of the POI of the posterior Monte Carlo sample. The MCMC method is just a very efficient way to produce the Monte Carlo sample according to an arbitrary distribution of prior times likelihood. The main idea behind this algorithm is to do not produce independent random numbers, but produce random numbers, depending on the last random number according to special rules. Therefore one gets a random walk through the PDF, where the special rules ensure that the resulting sample is distributed according to the posterior PDF. One of such MCMC algorithms is the Metropolis-Hastings algorithm [HAS70].

3.1.4. Limit Estimation with Systematic Uncertainties

In order to estimate an upper limit on a branching fraction

$$\mathcal{B} = \frac{N_{\mathrm{signal}}}{\mathcal{N}}\,, \tag{3.17}$$

where the normalization $\mathcal{N} = N_{B\bar{B}} \cdot \epsilon_{sig}$ and the estimated number of signal events $N_{\mathrm{signal}} = N_{\mathrm{obs}} - N_{\mathrm{b}}$ is needed. $N_{B\bar{B}}$ denotes the number of all produced events and ϵ_{sig} the signal efficiency. N_{obs}, N_{b} are the number of observed events and the number of expected background events, respectively. The limit on the branching fraction $\mathcal{B}$ is estimated by taking as N_{signal} the upper limit of the parameter estimated using the methods explained above.
In order to take systematic uncertainties on the background estimation and on the signal efficiency into account, these parameters fitted with gaussian constraints corresponding to the estimated systematic uncertainty (see e.g. [Con+03]). All the methods explained above are able to take such systematic uncertainties as gaussian constraints into the limit estimation into account.

4. Multivariate Analysis Algorithms

This chapter is about so-called multivariate analysis algorithms. With the huge increase of computational power in the last decandes, these algorithms received a growing popularity as they are in general computationally intensive. Out of the many algorithms, this chapter will highlight neural networks, in particular the multivariate analysis package NeuroBayes.

4.1. Introduction

A wide and important field in statistical data analysis is classification. The common goal is to separate two or more types of events by analyzing the distributions of their observables. If there are many observables one refers to this methods as multivariate analysis algorithms. In general, these algorithms project a N-dimensional vector of variables

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \tag{4.1}$$

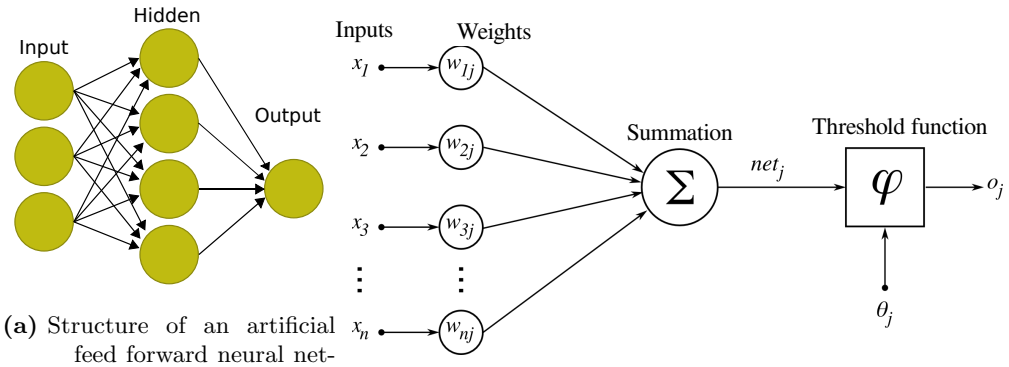
into a single scalar variable o

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \rightarrow o \; . \tag{4.2}$$

As an analogy we can think of a movie where several quantities like the actors performance, the complexity of the story and so on should be projected into a single rating from 1 to 10. The human brain is able to make such a projection. Therefore it is a natural approach to try to adopt the biological mechanism into a mathematical algorithm.

4.2. Artificial Neural Networks

The mathematical description of the mechanisms in the human brain is called artificial neural network. By the works of Warren McCulloch, Walter Pitts and Donald O. Hebb the fundamentals of these artificial neural networks are known. As shown in figure 4.1b the neurons are connected with many other neurons and, depending on the input gives one output to another neuron. Training such a neural network means to adjust the weights of the connections between the neurons. A special kind are the so-called feed-forward neural networks as shown in figure 4.1a. The information in such feed forward neural network flows only in one direction, any kind of loop is not allowed. For this special kind of a neural network structure it was possible to find an



(a) Structure of an artificial feed forward neural network.

(b) Structure of an artificial neuron.

Figure 4.1.

algorithm called back-propagation [Rum+86] which can be used to adjust the weights such that a error function E, for example

$$E = \frac{1}{2}\sum_{i=1}^{n}(t_i - o_i)^2$$

is minimized, where t_i is the target of event i and o_i the output of the network. It could be shown [Ruc+90] that as this error function is in its minimum with respect to the weights the output o_i is a real probability. This is a very important feature and plays a very important role in the new full reconstruction approach, explained in part IV.

4.3. Preprocessing in NeuroBayes

The NeuroBayes package is a multivariate analysis tool utilizing amongst other algorithms a feed-forward neural network. This neural network is embedded in a collection of pre- and post-processing algorithms.
The training of neural networks depends on various calculations including critical operations like computing ratios from potential very small numbers. Therefore, in NeuroBayes [Fei+06] the input variables are preprocessed such that numerical instabilities are minimized.

- Flatten the distribution
  The distribution of each input variable is flattened by a transformation using the cumulative distribution as shown in figure 4.2a.

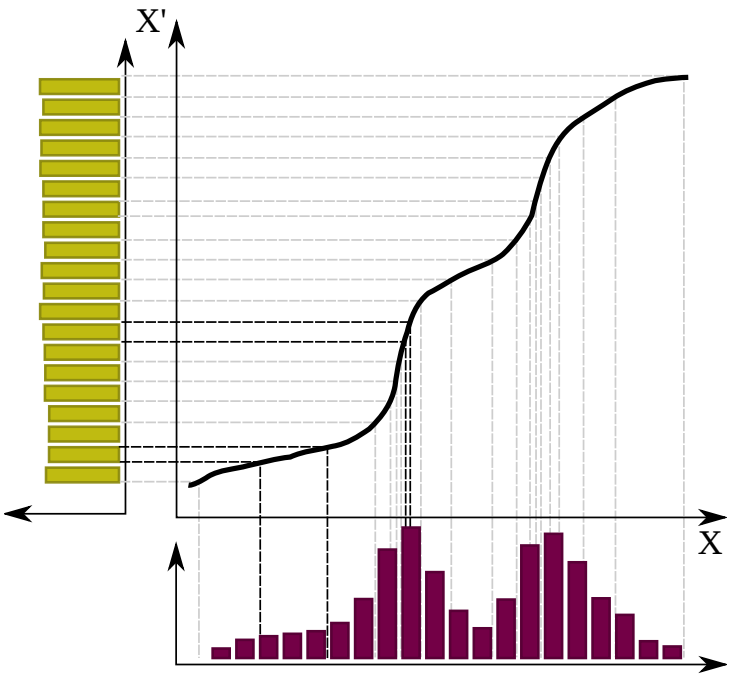
- Tranform into purity
  In the flattened distribution, in each bin the purity is estimated. A spline function is fitted to these purity values. Each bin is then transformed into the corresponding purity using this spline function. This regularizes the distribution and helps to get rid of statistical fluctuations. Moreover, the values are already sorted by their purity.

- Transformation to the normal distribution
  The flattened distribution is then transformed into a normal distribution using the cumulative Gauss distribution.

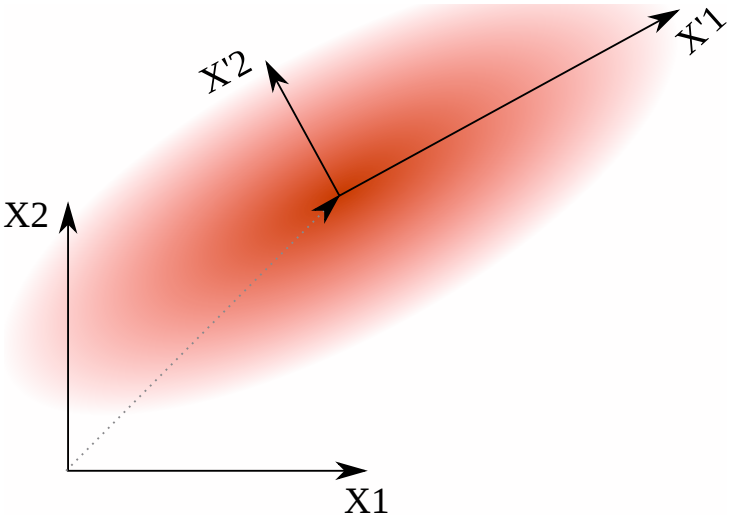
- De-correlation
  As there are only normal distributions in all dimensions left, one can rotate them such that all variables are de-correlated from each other as shown for two dimensions in figure 4.2b.

- Rotating into target
  As now all variables are linearily independent of each other, one can rotate the variables such that the whole information about the target variable is left in one variable. Only the deviation from this simple monotonous ansatz needs to be learned by the neural network.

After applying all the preprocessing procedures, the training is extremely robust and finds in most of the cases the optimum quite fast and reliable.
To be sure that the NeuroBayes output can be interpreted as a proper Bayesian probability, it is possible to transform the output such that it is by construction interpretable as a Bayesian probability.



(a) Any distribution can be transformed into a flat distribution using the cumulative distribution.



(b) By scaling and rotating the variables into new variables, any linear correlation can be removed.

Figure 4.2. Transformations of input variables used in the NeuroBayes preprocessing.

Part III.

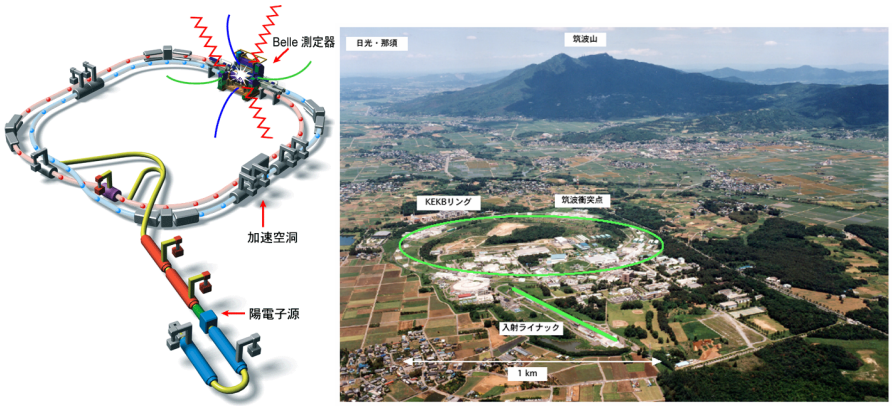
The Experimental Setup

5. The Belle Experiment

The next two chapters are about the experimental setup of the Belle experiment. There is the KEKB accelerator and the multipurpose Belle detector situated in the high energy research facility KEK in Tsukuba, Japan. The Belle experiment has stopped data acquisition in June 2010 and is now being upgraded to the high luminosity Belle II experiment.

5.1. The KEKB Accelerator

The KEKB accelerator [Kur+03] is a two-ring, asymmetric-energy, e^+e^- collider with one interaction point (IP) where 8 GeV e^- are colliding with 3.5 GeV e^+ resulting in 10.58 GeV center-of-mass energy, corresponding to the $\Upsilon(4S)$ resonance with a Lorentz boost factor of 0.425. It utilizes the tunnel of the predecessor 30 GeV e^+e^- accelerator TRISTAN with a circumference of 3016 m. A panoramic view of the accelerator and a drawing is shown in figure 5.1. The KEKB



(a) **(b)**

Figure 5.1. The KEKB accelerator as a drawing and a panoramic view.

accelerator currently holds two world records in the luminosity, a measure of number of collisions per time and beam size. First, the peak luminosity of

$$L = 21.083 \cdot 10^{33} \frac{1}{cm^2 \cdot s} = 21.083 \frac{1}{nb \cdot s}$$

and second, the world's largest integrated luminosity of

$$\int L = 1040 \cdot 10^{39} \frac{1}{cm^2} = 1040 \frac{1}{fb}$$

This high luminosity was achieved, amongst other things, by a small 22 mrad opening angle instead of head on collisions of the e^+e^- bunches. Later on, a new technique was introduced, called *crab cavities*, where the e^+e^- bunches are rotated shortly before the collision in order to hit head on once again. The design peak luminosity was excelled by a factor of 2.

5.2. The Belle Detector

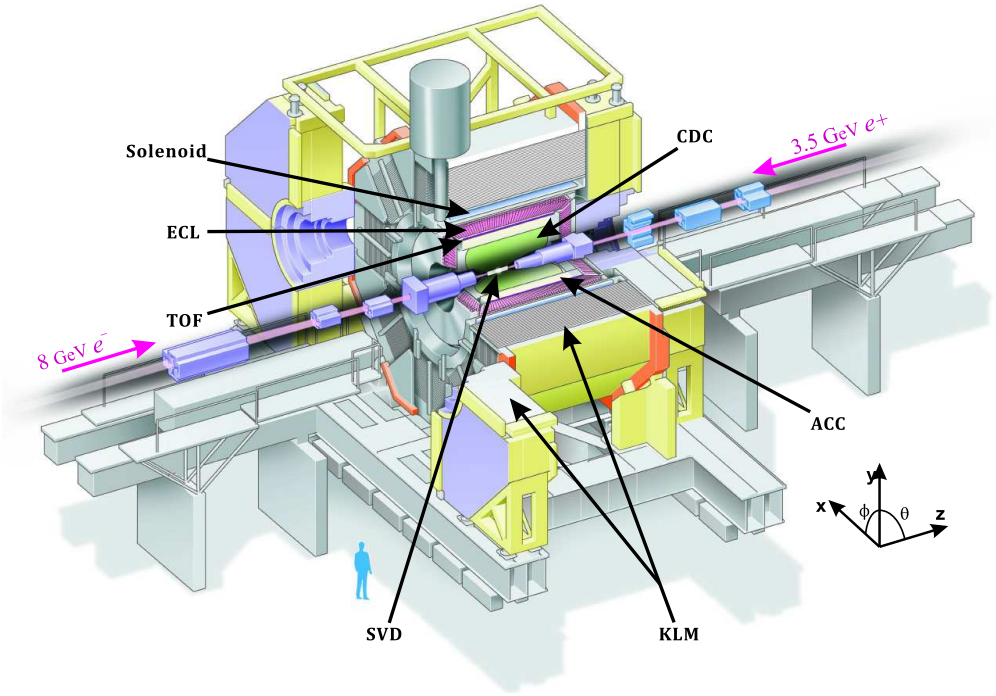
The Belle detector [Aba+02] is situated in the high energy research facility KEK in Tsukuba, Japan, using the KEKB accelerator described in chapter 5.1. It is a multi purpose detector which encloses the interaction point (IP) in all directions with several different detector components arranged like onion scales around the IP. Its purpose is to detect the visible final state decay products of the particles, produced in the e^+e^- collisions. Figure 5.2a shows a perspective view of the detector, whereas figure 5.2b shows a side-view along the beam-pipe of the detector. One can see that the detector is slightly asymmetric which reflects the resulting boost of the center-of-mass system due to the asymmetric beam energies.
In the following, the detector components are explained in more detail, starting from the innermost part to the outermost.

- Silicon Vertex Detector
  The silicon vertex detector (SVD) is the innermost detector component. It is a silicon strip detector with four layers around the beam-pipe. Its purpose is to precisely measure the origin of charged tracks and therefore finding decay vertices of the decaying particles. The silicon detector in the first period (called SVD1) covered a polar angle of $23° < \Theta < 139°$ with three layers. During an update (called SVD2), the SVD was replaced by a stip detector with four layers, increasing the coverage to $17° < \Theta < 150°$.

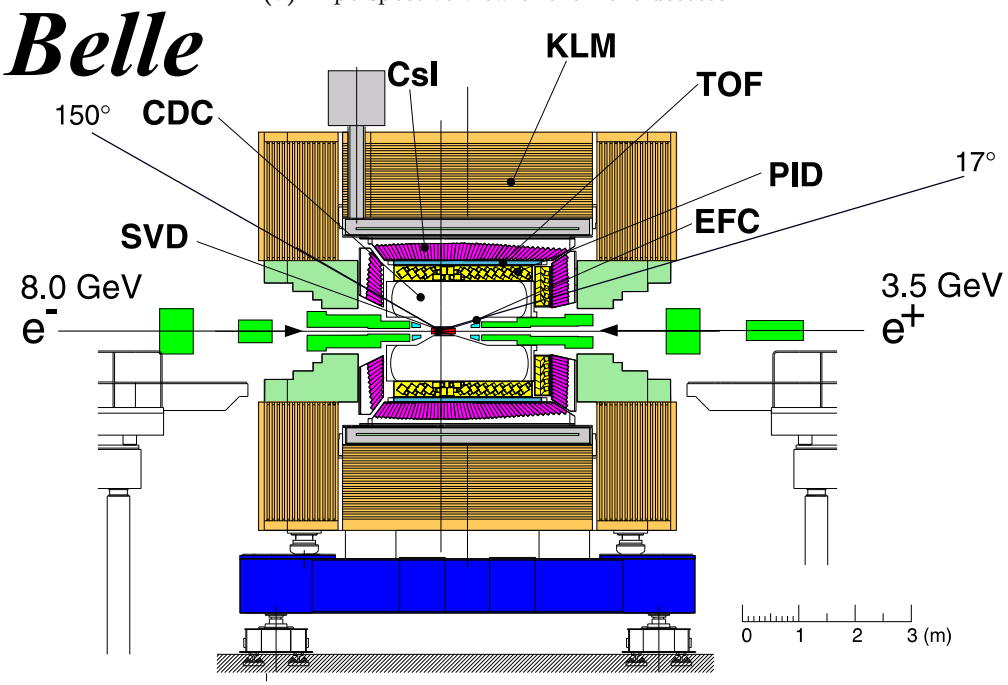
- Central Drift Chamber
  The central drift chamber (CDC) is a cylindrical volume with 8400 wires distributed in 50 layers, starting with a radius of 103 mm and ending with a radius of 874 mm. It covers a polar angle of $17° < \Theta < 150°$. Its main purpose is to measure precisely the momentum of charged tracks. For this a homogeneous magnetic field of 1.5 T is created by a superconducting solenoid coil. Depending on the momentum of the particles, they follow circular paths. The larger the momentum is, the larger the radius of these paths are. The volume is filled with gas which is ionized by the charged particles and the emitted electrons are detected by the wires, measuring the amount of emitted electrons and their drift time. With this information, pattern recognition algorithms try to find CDC hits belonging to one track and the track fit algorithms are fitting the momentum. The amount of emitted electrons is a measure of the energy loss dE/dx which is an important quantity to distinguish different particle types. The measured tracks are visualized in figure 5.3.

- Aerogel Cerenkov Counter
  The aerogel Cerenkov counter (ACC) is a Cerenkov light threshold counter. When charged particles traveling through matter with a higher velocity as the speed of light in that medium, they emit light in cones around the flight directions, with the opening angle proportional to the mass of the particle. Therefore, we can use this information for particle identification of high momentum charged particles as the particles need to have more than speed of light in the medium. The diffraction index is chosen such that pions are above the speed of light and other particles with the same momentum below this threshold.

- Time-of-Flight Detector
  The time-of-flight detector (TOF) measures the flight time of the particles. Together with



(a) A perspective view of the Belle detector.



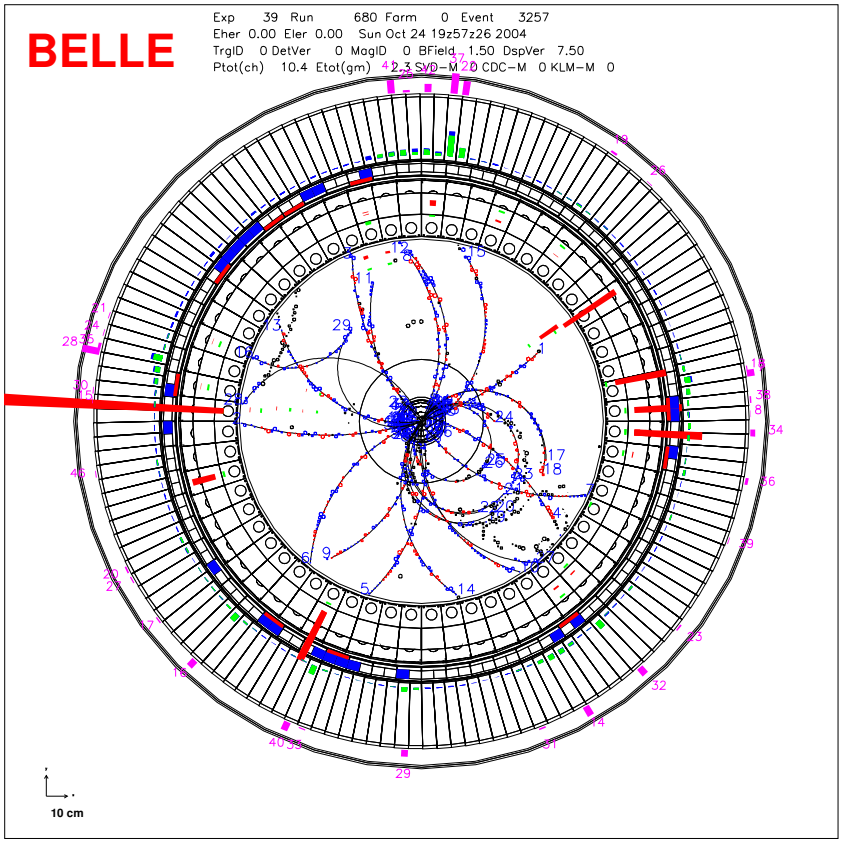
(b) The Belle detector from a side view.

Figure 5.2. The Belle detector

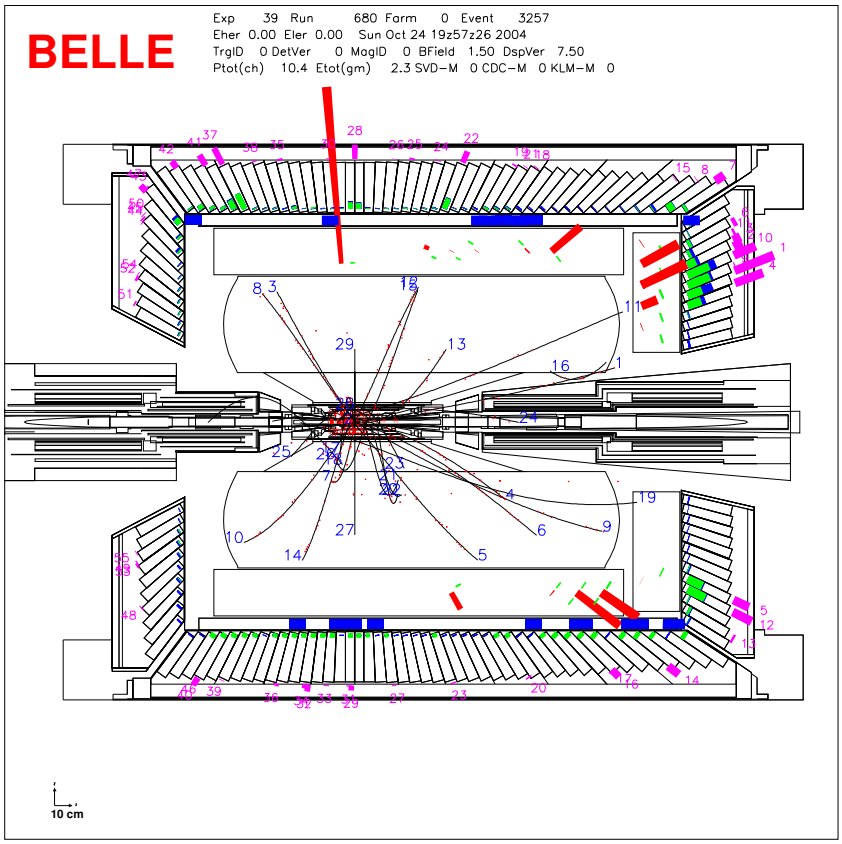
the momentum measured in the CDC we can therefore infer the mass of the particle. Therefore, we can use this information for particle identification of low momentum charged particles as the timing measurement has a too large error for very fast particles.

- Electro-Magnetic Calorimeter
  The electro-magnetic calorimeter (ECL) is composed of 8736 Thallium doped Cesium iodide (CsI(Tl)) scintillator crystals with silicon photo diode readout. It covers a polar angle of $17° < \Theta < 150°$. There are additional calorimeters in the extreme forward and backward directions (EFC), covering the polar angles $6.4° < \Theta < 11.5°$ and $163.3° < \Theta < 171.2°$. The ECL is shown in figure 5.3. The scintillating material gets excited while electrons are passing through. The intensity of the emitted visible light during the de-excitation is proportional to the energy of the charged particle. As we mainly want to measure photons, they first have to shower up, mainly due to pair production inside the crystals. As the crystal have a total length of 16 times the radiation length, most of the energy of the photons are contained and measured in the crystals.

- K_L and Muon Detector
  The K_L and muon detector (KLM) is an instrumented iron flux return for the magnetic field. It consists of iron plates, interlaced with glass resistive plate counters, detecting charged particles local discharges of the plates. Its main purpose is to detect K_L and muons. As K_L particles are neutral, the ECL is only 0.8 times the radiation length, while the iron KLM compares to 3.9 times the radiation length. As there are huge fluctuations, the energy of the K_L can not be reliable estimated, but only the direction. Muons with an energy above 600 MeV are able to reach the KLM and are identified by their unique signature in the KLM and ECL and with a charged track in the CDC pointing to the KLM cluster.



(a) View of the x-y plane

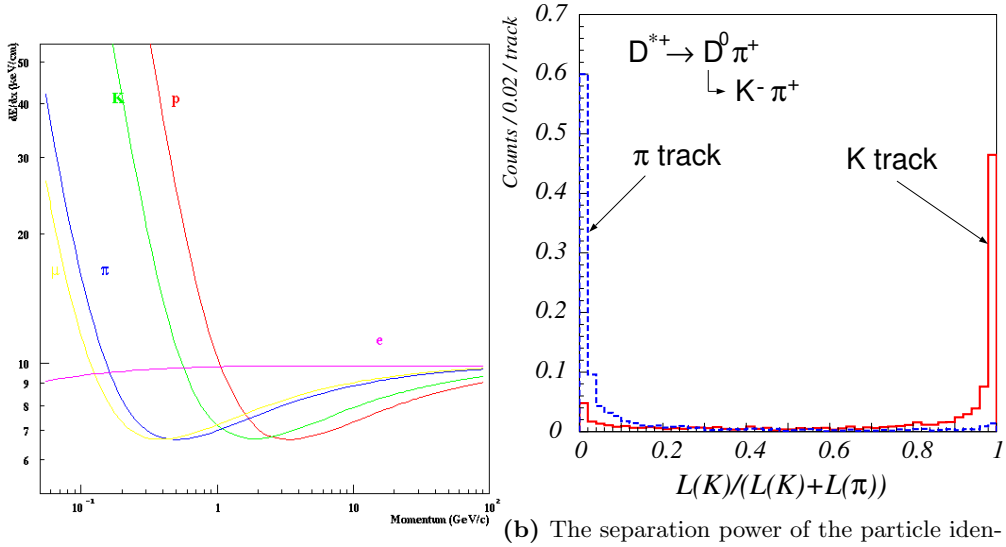


(b) View of the y-z plane

Figure 5.3. A typical event recored by the Belle detector. You can see the charged tracks in the CDC and the clusters in the ECL.

6. Particle Identification

For most of the analyses performed with the Belle detector, a good identification of final state particles is very important. Most important is the separation between kaons and pions. For this purpose, three detector components contribute with independent measurements to the particle identification. As already mentioned, the CDC measures the specific dE/dx of a charged track, the ACC the specific Cerenkov light emission and the TOF the specific time-of-flight. The specific energy loss dE/dx for different charged particles is shown in figure 6.1a.
The independent measurements are then combined into a likelihood ratio between the hypothesis of being a kaon track or a pion track. The separation power of this likelihood ratio is shown in figure 6.1b. This particle identification (PID) is described in detail in [Aba+02].



(a) Energy loss dE/dx for different particle types.

(b) The separation power of the particle identification likelihood ratio for pions and kaons.

Figure 6.1.

Part IV.

The Full Reconstruction Method

7. Introduction

In this part a unique analysis tool, the full reconstruction method, is presented. The major difficulty in this method is the huge number of combinations. It is therefore crucial to reduce the number of combinations to satisfy the limitations due to computational resources. In order to reduce the number of combinations by several orders of magnitude, the multivariate analysis algorithm NeuroBayes is used. Probabilities are estimated using NeuroBayes in a hierarchical structure. It is possible to postpone the decisions whether a combination is used, in higher stages in this hierarchy which maximizes the efficiency while the complexity can be chosen freely to satisfy the computational limitations.

7.1. Overview

One common feature of e^+e^- colliders compared to hadron colliders is the well known initial state of the reaction. The entire e^+e^- center-of-mass energy is carried away by the reaction products, in contrast to hadron colliders, where only an unknown part of the center-of-mass energy is available for the hard reaction. In e^+e^- colliders it is therefore possible to choose a well known initial state. In B-factories such as the Belle experiment, the e^+e^- center-of-mass energy is chosen such that the $\Upsilon(4S)$ resonance will be produced in roughly every fifth event. In the rest of the events with hadronic final states, continuum events $e^+e^- \to q\bar{q}(q = u, d, s, c)$ are produced. The $\Upsilon(4S)$ resonance lies just above the kinematical threshold of producing a $B\bar{B}$ pair and is therefore able to decay into a $B\bar{B}$ pair. All other decay channels are highly suppressed by the OZI-rule and so, the $\Upsilon(4S)$ resonance decays into a $B\bar{B}$ pair in over 96% of the cases. An overview of the total cross-section of e^+e^- into hadrons in the $\Upsilon(4S)$ region and a schematic illustration of the continuum and $B\bar{B}$ events is shown in figure 7.1.
Due to the close-by threshold, it is obvious that, if there was an $\Upsilon(4S)$ resonance produced, it always decays in exactly a $B\bar{B}$ pair and nothing else. This gives a unique constraint on the reconstruction of the entire event. Everything in the detector has its origin in the two B mesons in the event.
Employing momentum and energy conservation we even do not have to detect all final state particles. The known initial state gives us the opportunity to measure many decay channels which otherwise would be inaccessible. The most prominent example is the decay

$$B^0 \to \nu\bar{\nu}$$

which clearly is not possible to measure by reconstructing the final state particles, because both neutrinos cannot be detected. Nevertheless, it is possible to measure at least the branching ratio of this decay with the special setting mentioned above. If we fully reconstruct one B meson in one event completely correctly, we know that there is exactly one other B meson in this event. None of the final state particles used for this B meson does originate from the other B. Moreover, due to the fact that the detector encloses the interaction point nearly hermetically, we know that all remaining final state particles detected in this event must stem from the other B meson. In the

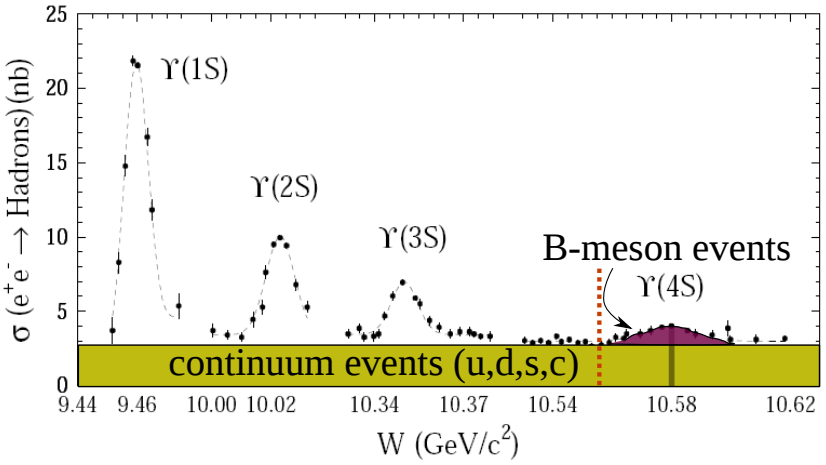


Figure 7.1. The e^+e^- to hadrons cross section in the $\Upsilon(1S)$ - $\Upsilon(4S)$ region. The orange dashed line marks the kinematic threshold for the production of a $B\bar{B}$ pair. (source: [Bes+93])

example above we can measure the branching ratio by counting all events, in which we found a fully reconstructed B meson but nothing else. In the following, the fully reconstructed B meson is called the tag-side $\mathrm{B_{Tag}}$ and the remaining B meson is called the signal-side $\mathrm{B_{Sig}}$ candidate. An illustration of this can be seen in figure 7.2. The aim of the full reconstruction tool is to

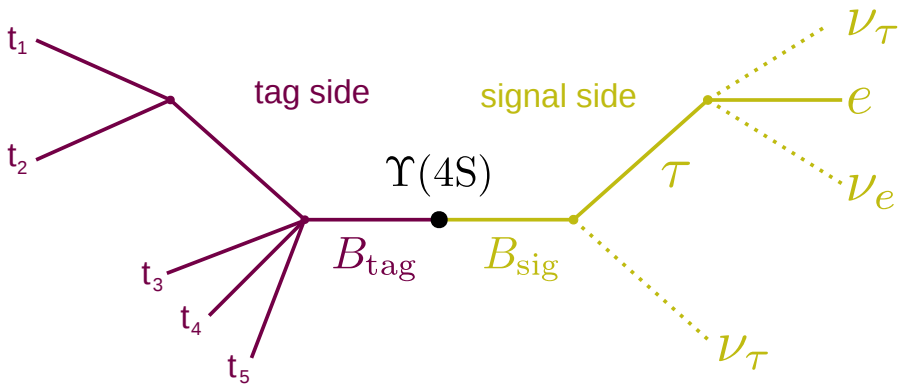


Figure 7.2. An illustration of the tag-side (purple) and an exemplary $B \to \tau\nu$ decay as the signal-side (green).

reconstruct the $\mathrm{B_{Tag}}$ candidate. How this can be done will be explained in the following sections.

8. Functional Requirements

8.1. Reconstruction

What the detector can measure are only tracks, caused by long-lived charged particles ($\pi^{\pm}$, $K^{\pm}$, $e^{\pm}$, $\mu^{\pm}$, $p^{\pm}$) and clusters in the calorimeter, caused by photons and neutral hadrons (and also charged particles). All other particles have to be reconstructed from these decay products. A visualization of two events simulated by the Belle detector simulation are shown in figure 8.3. The full reconstruction tool should be able to reconstruct, what happened on the tag-side. The signal-side is a decay of $B^+ \to K^+\nu\bar{\nu}$ (see part V) in both events. The B_{Tag} was simulated in the decay chain shown in figure 8.1 for the event shown in figure 8.3b and the decay chain shown in figure 8.2 for the event shown in figure 8.3a. Only the underlined particles in figure 8.2 and 8.3a can be found in the detector, all other particles have to be reconstructed from these final state particles. There is no information available which measured particle corresponds to which particle in such a decay chain. The only possibility is to try all combinations and then check which combinations match the selection criteria. The tracks marked green in figure 8.3 are correctly matched final state particles. The combination found in figure 8.3a was correctly reconstructed as the decay that was generated. The combination found in figure 8.3b used the track marked in red, but this is not a final state particle generated by the simulation. Therefore this combination is wrong.

The challenging task for the full reconstruction tool will be to sort out as many wrong combinations as possible and, at the same time find as many correct combinations as possible.

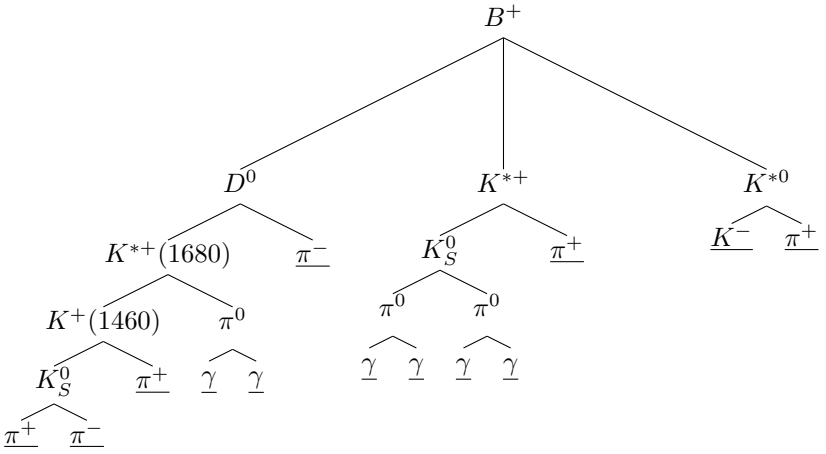


Figure 8.1. The simulated decay chain in the event shown in figure 8.3b. Only the underlined particles can be found in the detector.

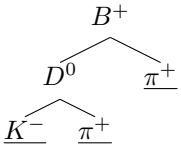
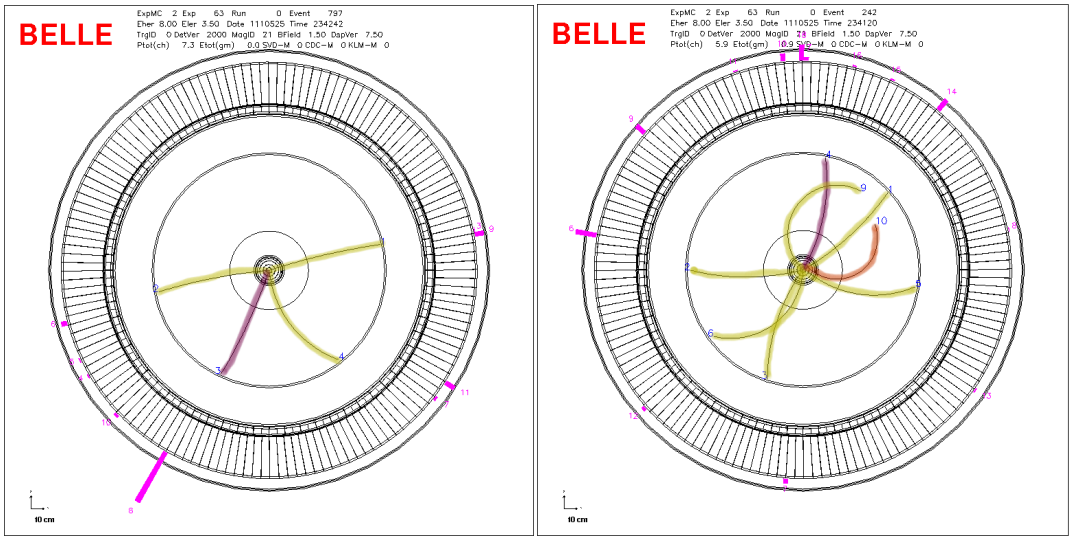


Figure 8.2. The simulated decay chain in the event shown in figure 8.3a. Only the underlined particles can be found in the detector.



(a) **(b)**

Figure 8.3. A visualization of two events measured by the Belle detector. The final state particles of the B_{Tag} are marked with green color and the single track of the B_{Sig} is marked purple. In the outermost part, calorimeter clusters can be seen which are not used here. The event shown in the left figure was reconstructed correctly. In the event shown in the right figure, the red marked track was falsely reconstructed.

8.2. Efficiency Maximization

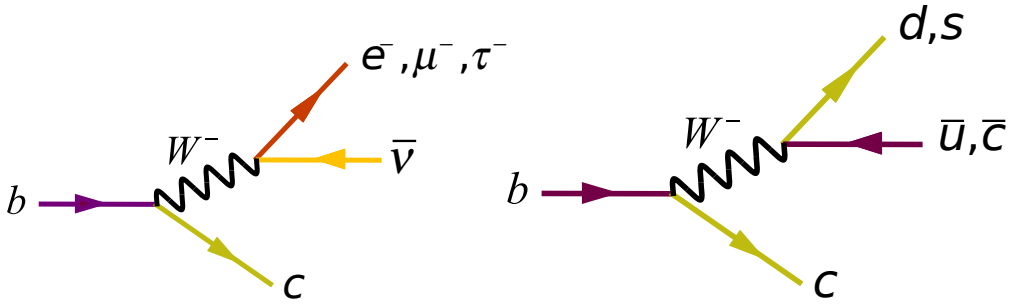
It is crucial to reconstruct one B meson correctly in as many events as possible. This means that we have to maximize the efficiency

$$\varepsilon_{tot} = \sum_{i}^{N} \varepsilon_i \cdot \mathcal{B}_i \ , \tag{8.1}$$

where ε_i is the reconstruction efficiency in the decay channel i and $\mathcal{B}_i$ the branching ratio of the decay channel i. While the branching ratios are fixed by nature, we have two possibilities to maximize this efficiency ε_{tot}:

- Increase the number of reconstructed decay channels N.
- Increase the reconstruction efficiency ε_i.

In principle this would be a trivial task, but there are further constraints that expand this problem to a complicated optimization problem.
B mesons are decaying in several hundred different decay channels with very small branching ratios each. We are restricted to those decay channels where all final state particles are detectable by the detector. Therefore all decay channels with one or more neutrinos in the final state are ruled out. These are all the semi-leptonic decays (see figure 8.4a), which sum up to a total fraction of $\sim$ 25%. The same holds for all the decay channels of the daughter particles like D mesons. These branching fractions are treated as reconstruction efficiencies of the B meson decay channels in the following. The Feynman graph of the $b \to c$ transition which can be used for the full reconstruction is shown in figure 8.4b.



(a) Feynman graph of a semi-leptonic $b \to \ell\nu c$ transition. **(b)** Feynman graph of a hadronic $b \to c$ transition.

Figure 8.4. Feynman graphs of the dominant $b \to c$ transition.

8.3. Adjustable Purity

The purity of the tag-side B meson reconstruction is important for several reasons. The first reason is that the analysis of signal-side B meson might highly depend on the purity of the tag-side B meson. But since this is not true for all analyses there is a need for the possibility to choose a different efficiency and purity of the tag-side B meson selection for different analyses. Another reason is that the processing time for the user depends on the number of events where at least one B meson candidate is found. By adjusting the purity such that only in every second event a B meson candidate is found, the processing time for the user will be halved.

8.4. Throughput Constraint

As already explained before it would be a trivial task to increase the efficiency of the B_{Tag} reconstruction to the detector-limited possible maximum by adding all possible decay channels. But this maximum cannot be reached as the combinatorial complexity increases exponentially with the number of final state particles. In contrast to a purely academic problem, we want to use this tool for analyses. This forces us to be able to process the whole data sample recorded by the Belle experiment together with all Monte Carlo simulated samples in a few months at the most. As there are 3 billion data events recorded and more than 5 times of the data luminosity is available as Monte Carlo samples we need to be able to process more than 1000 events per second. With the trivial solution by just doing naively each possible combination, one event takes up to several minutes of computing time. In order to reach the desired throughput, the tool has to speed up the processing time by many orders of magnitude.

9. The Hierarchical Approach

9.1. Exploiting the Natural Hierarchical Structure

As explained before it would not be possible to calculate all possible combinations, especially if we want to reconstruct the B_{Tag} in decay chains with high multiplicity. But this is also desired in order to achieve a high efficiency. Looking at the structure of an exemplary decay chain like in figure 8.2 one immediately recognizes a hierarchical structure. Due to energy conservation, particles can only decay into lighter particles. Together with other physical constraints like momentum conservation or angular momentum conservation we can reconstruct the intermediate particles in several stages up to the B mesons. This gives us the possibility to switch over from consecutively reconstructing exclusive decay channels to the reconstruction of semi-inclusive decay channels.
In a classical cut-based approach this would be done in the same way, but in order to keep the combinatorial complexity under control, one needs to cut away most of the intermediate candidates. This reduces of course the efficiency but there is another serious problem. The acceptable amount of candidates for the intermediate particles depends on the complexity of the decays above. As an example, the amount of combinations of the decay $B \to D\pi\pi\pi$ is much higher than for the decay $B \to D\pi$ with the same amount of D candidates. To account for this, one is forced to consecutively reconstruct exclusive decay channels and optimize the cuts separately.
The new hierarchical approach explicitly accounts for this. The idea is to find a way to compare all the different intermediate candidates, regardless of their individual decay channels. We decided to estimate posterior signal probabilities and use this probability to decide whether a candidate is used for further combinations or not, depending on the complexity of the decay in the higher stages. A candidate is not thrown away per se, but for some channels he is accepted, for others he is rejected.

9.2. Estimate Posterior Probabilities

As explained before, we need the posterior probabilities for each candidate. A trivial solution would be to just estimate the frequentist probabilities for each exclusive decay channel by counting signal and background candidates in the Monte Carlo samples. Depending on the complexity of the B decay channel one could then just require a minimum probability for the intermediate particles.
But as explained in chapter 4 we can do much better than this simple approach. By training a NeuroBayes neural network we can obtain the proper posterior probability for each candidate individually. This probability is not decay channel dependent, but candidate dependent. This will result in a huge improvement, as we are now able to also include decay channels for intermediate particles with huge amounts of background and use only those candidates which have a high probability.
But of course this probability is worth much more than just a cut. This information will be passed through the complete hierarchy up to the B mesons. As this probability contains all

the information about being signal or background, it is sufficient to pass this probability to combinations of heavier particles. Everything below that particle is combined in this probability. As there is in some decay channels such a vast amount of background candidates, one needs to find a method to train NeuroBayes efficiently with such a small signal fraction. Even if there is from a theoretical point of view not a problem, one expects to get numerical instabilities if the signal fraction is so tiny. We found that a signal fraction below 10% might lead to not optimal NeuroBayes trainings. One possibility would be to divide the separation procedure into two steps. In the first step, the biggest part of the background is reduced while not learning all details. In a second step, NeuroBayes learns the remaining differences between signal and background candidates. This is done by applying individual weights to the training samples, where the weights reflect the "wrongness" of the first step. This is called a *boost* procedure.
As it turned out, a much simpler procedure is enough for our purpose. We just artificially increased the signal fraction by dropping background events, until the signal fraction reaches 10%. The only problem that occurs with this procedure is that the probability of being signal or background is only valid if the signal fraction is the same in the training and in the data sample. This is not the case any more, but we can recalculate the proper probability if we know the signal fraction in the training sample and in the data sample.
To calculate this correction, we need Bayes' theorem, which is defined for two types of events, X and Y, as

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \ . \tag{9.1}$$

For our purposes, it is preferable to use Bayes' theorem in terms of the likelihood ratio

$$\Lambda(Y|X) = \frac{P(Y|X)}{P(Y|\neg X)} \ , \tag{9.2}$$

which leads to prior odds of

$$O(X) = \frac{P(X)}{P(\neg X)} \ , \tag{9.3}$$

and posterior odds given by

$$O(X|Y) = O(X) \cdot \Lambda(Y|X) \ , \tag{9.4}$$

In our example X and $\neg X$ are signal events (S) and background events (B) and Y is the output (o_t) from a network trained with the training dataset (denoted with the subscript t). The likelihood ratio is

$$\Lambda(Y|X) = \frac{P(o_t|S)}{P(o_t|B)} \ , \tag{9.5}$$

where $P(o_t|S)$ is the likelihood to get a network output, o_t, given a signal event S and $P(o_t|B)$ is the same for a background event. Given a network output o_t the conditional probability of being a signal event S, is

$$o_t = P_t(S|o_t) \ , \tag{9.6}$$

and the corresponding probability of being a background event B is given by

$$(1 - o_t) = P_t(B|o_t) \ . \tag{9.7}$$

By applying Bayes theorem as follows

$$\frac{P_t(S|o_t)}{P_t(B|o_t)} = \frac{P_t(S)}{P_t(B)} \cdot \Lambda(o_t|S) \tag{9.8}$$

we can write the likelihood ratio as

$$\begin{aligned} \Lambda(o_t|S) &= \frac{P(o_t|S)}{P(o_t|B)} \\ &= \frac{o_t}{1-o_t} \cdot \frac{P_t(B)}{P_t(S)} \, . \end{aligned} \tag{9.9}$$

This likelihood ratio does not depend on the signal to background ratio because it only contains measured information of one given event. We now can calculate, for any other signal to background ratio in the prediction dataset (denoted with the subscript p), the posterior odds with Bayes theorem:

$$\frac{P_p(S|o_p)}{P_p(B|o_p)} = \frac{P_p(S)}{P_p(B)} \cdot \Lambda(o_t|S) \, . \tag{9.10}$$

Because the transformed probability o_p has to satisfy

$$\frac{P_p(S|o_p)}{P_p(B|o_p)} = \frac{o_p}{1-o_p} \tag{9.11}$$

to be the correct probability, we get:

$$o_p = \frac{1}{1 + (\frac{1}{o_t} - 1)\frac{P_p(B)}{P_p(S)}\frac{P_t(S)}{P_t(B)}} \, . \tag{9.12}$$

This formula is used in the full reconstruction algorithm to correct for the artificially increased signal fraction and is also published [Fei+11].

10. The Framework Structure

10.1. The `DecayChannel` and `DigitalPhysicist` Classes

As we figured out the functional requirements in chapter 8 and an idea how to solve this problem with the hierarchical approach in chapter 9 we need to design a framework which is able to satisfy the requirements. As there are hundreds of decay channels, the framework has to have an easy interface to include new decay channels. For this, the `DecayChannel` class is created, which does not hold information about a specific candidate, but only information which is needed do the actual combinations. An object of a `DecayChannel` class represents a particle (e.g. a $D^{\pm}$ meson) and its specific decay channel. An example of the definition of such a `DecayChannel` object is shown in listing 10.1. The CERN Monte Carlo particle numbering scheme [Nak+10] is used to identify particle types.

```
1  //D*+ -> D0 Pi+
2  //particle type: D*+
3  DecayChannel* channel = new DecayChannel(413);
4  //child 1: D0
5  channel->addChild(421);
6  //child 2: Pi+
7  channel->addChild(211);
```

Listing 10.1 An example for a `DecayChannel` object construction.

In this example the decay of a D^{*+} meson into a D^0 meson and a π^+ is defined. As can be seen, the D^0 child is not further specified. That means for the combination, each candidate which was combined to a D^0 defined in a D^0 `DecayChannel` object, no matter in which decay channel, is combined with an object defined as a π^+ which is a final state particle.
In order to keep track and collect information about all the different decay channels which should be reconstructed, the `DigitalPhysicist` class is created. This class stores all the different `DecayChannel` objects. It is checked that there is at least one decay channel defined for each particle type which is used as a child. In the example above it would be automatically checked whether there is a D^0 `DecayChannel` object. As heavier particles only decay into lighter particles, it is further checked that all combinations of lighter particles are done before the combination to heavier particles.

10.2. The `Particle` Class

For each combination instructed by the several `DecayChannel` objects a `Particle` object is created and filled into a corresponding particle list. This `Particle` holds all the information for the individual particle candidates. For convenience, these `Particle` lists have overloaded operators for the actual combinatorial loop structure. The code snippet 10.2

```
1  ParticleList A,B,C;
2  fill(A);
3  fill(B);
4  C = A * B;
```

Listing 10.2 An example for a `DecayChannel` object construction.

fills the particle list A and B with some arbitrary `Particle` objects for example π^+ and π^- respectively. The particle list C will then be filled with all combinations of A and B. The actual combinations are dynamically generated using the `DecayChannel` instructions stored in the `DigitalPhysicist` object.

10.3. The `Variable` Class

As we want to train NeuroBayes neural networks for each particle in each decay channel separately it is crucial to keep track of the variables which should be used in the neural network. The general procedure is to write out ROOT files [Bru+97], train the NeuroBayes neural network off-line on these root files and then implement the final NeuroBayes expert into the framework. This is a very critical point, as NeuroBayes is trained with pure float type arrays and therefore a variable is only determined by the position in this array. In order to fill the right variable inside the full reconstruction framework into the same position in the NeuroBayes array as in the training, one has to look inside the code and count the positions in the array. As all variables are of the type float, NeuroBayes will not give any warning at all if one mixes up something here, also the output might be unsuspicious. To avoid this source of errors, the `Variable` class is created. This class defines what a variable is. A variable has to have a unique name and should be able to return a float number if given a `Particle` object. All the variables which are important for a specific decay channel, like invariant mass, number of daughter particles, angles and so on are added to the corresponding `DecayChannel` object.
The definition of such a class inheriting from the virtual `Variable` class is shown in the code snippet 10.3 and the usage in 10.4.

```
1  class Variable {
2   public:
3    Variable(std::string aName) : name(aName){};
4    virtual float calculate(Belle::Particle* p) const = 0;
5    std::string   getName() const {return name;};
6   protected:
7    std::string name;
8  };
9
10 class NChildrenMCVariable: public Variable {
11  public:
12    NChildrenMCVariable() : Variable("nchildrenMC") {};
13    virtual float calculate(Belle::Particle* p) const;
14 };
15
16 float NChildrenMCVariable::calculate(Belle::Particle* p) const
17 {
18   float N_Children= p.MCChildren();
19   return N_Children;
20 }
```

Listing 10.3 An example for a variable class.

```
1  //D*+ -> D0 Pi+
2  //particle type: D*+
3  DecayChannel* channel = new DecayChannel(413);
4  //child 1: D0
```

```
5  channel->addChild(421);
6  //child 2: Pi+
7  channel->addChild(211);
8  //Add variables
9  channel->addVariable(new NChildrenMCVariable());
```

Listing 10.4 An example for the usage of the variable class.

10.4. The NeuroBayes Interface

As all the decay channels have a list of the variables, they can automatically create ROOT n-tuples out of this list, with the columns named after the corresponding `Variable` objects. With these automatically generated ROOT n-tuples we can then train NeuroBayes neural networks. As the transition from the training to the expert inside the full reconstruction module is a big source of errors, this transition needs to automatized. For this, the necessary information about the training, that is, the variable names and their individual preprocessing is stored in a training file. This training file is read out in the training and the variables are automatically extracted from the ROOT files by their unique name and passed to NeuroBayes in the order defined in the training file. When implemented in the full reconstruction module, the NeuroBayes expert once again reads out the same training file, but instead of extracting the variables from a ROOT file, the `Variable` objects with the appropriate unique name are used. This procedure completely removes this source of errors and also makes it very convenient and fast to train and implement new NeuroBayes networks in the full reconstruction framework.

10.5. The Four Stages

By exploring the different possible B decay channels, it turned out that a structure with four stages as shown in figure 10.1 is sufficient to reconstruct most of the decays. A stage is defined as a group of particles which only decays into particles of a stage below. The lowest stage are therefore the final state particles plus K_S^0 and π^0 as these candidates were already combined during the data processing and stored.
As already explained in section 9.2 the probabilities of all candidates will be passed over to higher stages. Because of that, all NeuroBayes trainings of one stage can be done in one turn as they all depend only on lower stages. Only after implementing the experts for all particles in this stage, the next stage can be processed and trained.

10.6. The Decay Channels

Of course, the number of decay channels we add to the full reconstruction module is virtually not limited. But as this module is needed by many analyses in the Belle collaboration one has to stop developing at some point and release the package as it is. This argument might sound a little bit nonscientific but it should not. Science can of course not neglect real world's necessities such as time limitations.
Anyway, the channels are added one after another according to the expected gain in efficiency. Therefore one would not expect to gain much more efficiency by adding more channels than there are in the current version. The reconstructed modes for the stage 2 particles are listed in table 10.1 and the ones for the stage 3 particles in table 10.2. Finally, the B decay channels are listed in table 10.3.

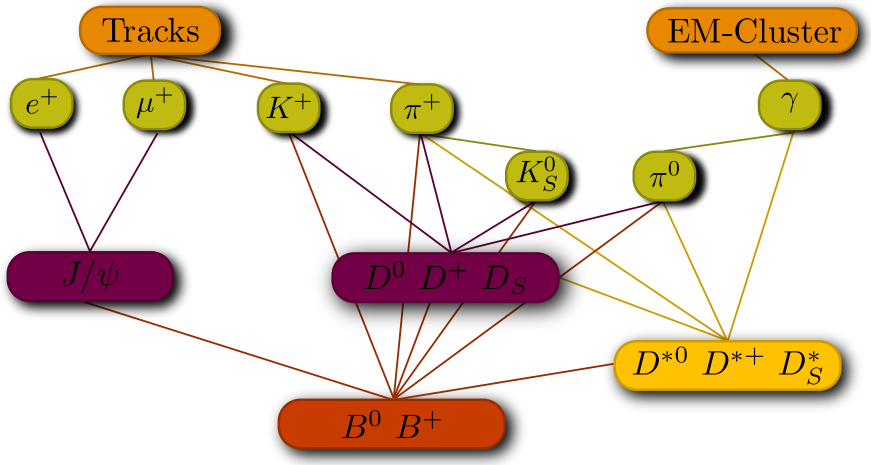


Figure 10.1. The 4 stages of the full reconstruction

There are automatically unique hash tags generated for each of the decay channels. They consist of 6 digit numbers, where the first 3 digits represent the CERN Monte Carlo particle numbering scheme [Nak+10] and the last 3 digits are the sum of the children particle numbers plus additional numbers for special cases.
At first glance it might look like a manageable amount of decay channels, but note that these are not the exclusive modes we are actually reconstructing. Counting all the branchings, there are 1104 exclusive decay channels recontructed.

10.7. The NeuroBayes Trainings

For the trainings we used variables with a good separation power for signal and background. The most powerful variables for the different particle types are listed in appendix 22.1.
We explicitly paid attention that the variables have similar distributions on the Monte Carlo samples and on the data sample. This was done by NeuroBayes trainings where the Monte Carlo samples are trained against the data samples. Variables with a high separation power between both samples have different distributions on data and Monte Carlo and therefore are removed from the signal to background trainings.
Moreover, to be able to estimate the efficiency of the full reconstruction tool, a fit to the resulting distribution of the beam-constraint mass

$$M_{bc} \equiv \sqrt{E_{\text{beam}}^2 - p_B^2} \tag{10.1}$$

was performed where E_{beam} is the beam energy and p_B the momentum of the reconstructed candidate. For this, the background must not be biased such that it looks similar to the signal distribution. If the B meson trainings would include variables which are highly correlated to $\mathrm{M_{bc}}$ this would be the case. Therefore, we trained networks between different windows of $\mathrm{M_{bc}}$ and removed all variables which were found to have a good separation power between the two $\mathrm{M_{bc}}$ windows.

D^0				D^+			
mode			BR	mode			BR
D^0	$\rightarrow$	$K^-\pi^+$	3.89%	D^+	$\rightarrow$	$K^-\pi^+\pi^+$	9.40%
D^0	$\rightarrow$	$K^-\pi^+\pi^+\pi^-$	8.09%	D^+	$\rightarrow$	$K^0_S\pi^+$	1.49%
D^0	$\rightarrow$	$K^-\pi^+\pi^0$	13.90%	D^+	$\rightarrow$	$K^0_S\pi^+\pi^0$	6.90%
D^0	$\rightarrow$	$\pi^+\pi^-$	0.14%	D^+	$\rightarrow$	$K^-\pi^+\pi^+\pi^0$	6.08%
D^0	$\rightarrow$	$\pi^+\pi^-\pi^0$	1.44%	D^+	$\rightarrow$	$K^0_S\pi^+\pi^+\pi^-$	3.10%
D^0	$\rightarrow$	$K^0_S\pi^0$	1.22%	D^+	$\rightarrow$	$K^+K^-\pi^+$	0.98%
D^0	$\rightarrow$	$K^0_S\pi^+\pi^-$	2.94%	D^+	$\rightarrow$	$K^+K^-\pi^+\pi^0$	1.50%
D^0	$\rightarrow$	$K^0_S\pi^+\pi^-\pi^0$	5.40%				
D^0	$\rightarrow$	K^+K^-	0.39%				
D^0	$\rightarrow$	$K^+K^-K^0_S$	0.47%				
D_s				J/ψ			
mode			BR	mode			BR
D^+_s	$\rightarrow$	$K^+K^0_S$	1.49%	J/ψ	$\rightarrow$	e^-e^+	5.94%
D^+_s	$\rightarrow$	$K^+\pi^+\pi^-$	0.69%	J/ψ	$\rightarrow$	$\mu^-\mu^+$	5.93%
D^+_s	$\rightarrow$	$K^+K^-\pi^+$	5.50%				
D^+_s	$\rightarrow$	$K^+K^-\pi^+\pi^0$	5.60%				
D^+_s	$\rightarrow$	$K^+K^0_S\pi^+\pi^-$	0.96%				
D^+_s	$\rightarrow$	$K^-K^0_S\pi^+\pi^+$	1.64%				
D^+_s	$\rightarrow$	$K^+K^-\pi^+\pi^+\pi^-$	0.88%				
D^+_s	$\rightarrow$	$\pi^+\pi^+\pi^-$	1.10%				

Table 10.1. Stage 2 - Reconstructed D and J/ψ modes. Branching ratios are from Ref. [Nak+10].

D^{*+}				D^{*0}			
mode			BR	mode			BR
D^{*+}	$\rightarrow$	$D^0\pi^+$	67.70%	D^{*0}	$\rightarrow$	$D^0\pi^0$	61.90%
D^{*+}	$\rightarrow$	$D^+\pi^0$	30.70%	D^{*0}	$\rightarrow$	$D^0\gamma$	38.10%
D^*_s							
mode			BR	mode			BR
D^{+*}_s	$\rightarrow$	$D^+_s\gamma$	94.20%				

Table 10.2. Stage 3 - all D^* modes (BR from [Nak+10]).

B^+				B^0			
mode			BR	mode			BR
B^+	$\rightarrow$	$\bar{D}^0\pi^+$	0.484%	B^0	$\rightarrow$	$D^-\pi^+$	0.268%
B^+	$\rightarrow$	$D^0\pi^+\pi^0$	1.340%	B^0	$\rightarrow$	$D^-\pi^+\pi^0$	0.760%
B^+	$\rightarrow$	$D^0\pi^+\pi^+\pi^-$	1.100%	B^0	$\rightarrow$	$D^-\pi^+\pi^+\pi^-$	0.800%
B^+	$\rightarrow$	$D_s^+D^0$	1.000%	B^0	$\rightarrow$	$D^0\pi^0$	0.026%
B^+	$\rightarrow$	$D^{0*}\pi^+$	0.519%	B^0	$\rightarrow$	$D_s^+D^-$	0.720%
B^+	$\rightarrow$	$\bar{D}^{0*}\pi^+\pi^0$	0.980%	B^0	$\rightarrow$	$D^{*-}\pi^+$	0.276%
B^+	$\rightarrow$	$D^{0*}\pi^+\pi^+\pi^-$	1.030%	B^0	$\rightarrow$	$D^{*-}\pi^+\pi^0$	1.500%
B^+	$\rightarrow$	$D^{0*}\pi^+\pi^+\pi^-\pi^0$	1.800%	B^0	$\rightarrow$	$D^{*-}\pi^+\pi^+\pi^-$	0.700%
B^+	$\rightarrow$	$D_s^{+*}\bar{D}^0$	0.760%	B^0	$\rightarrow$	$D^{*-}\pi^+\pi^+\pi^-\pi^0$	1.760%
B^+	$\rightarrow$	$D_s^+\bar{D}^{0*}$	0.820%	B^0	$\rightarrow$	$D_s^{+*}D^-$	0.740%
B^+	$\rightarrow$	$D_s^{+*}D^{0*}$	1.710%	B^0	$\rightarrow$	$D_s^+D^{*-}$	0.800%
B^+	$\rightarrow$	$\bar{D}^0K^+$	0.037%	B^0	$\rightarrow$	$D_s^{+*}D^{*-}$	1.770%
B^+	$\rightarrow$	$D^-\pi^+\pi^+$	0.107%	B^0	$\rightarrow$	$J/\psi K_S^0$	0.087%
B^+	$\rightarrow$	$J/\psi K^+$	0.101%	B^0	$\rightarrow$	$J/\psi K^+\pi^-$	0.120%
B^+	$\rightarrow$	$J/\psi K^+\pi^+\pi^-$	0.107%	B^0	$\rightarrow$	$J/\psi K_S^0\pi^+\pi^-$	0.100%
B^+	$\rightarrow$	$J/\psi K^+\pi^0$	0.047%				
B^+	$\rightarrow$	$J/\psi K_S^0\pi^+$	0.094%				

Table 10.3. Stage 4 - All B modes (BR from [Nak+10]).

11. Performance Optimizations

As already mentioned, it is crucial that the throughput of the module is inside an acceptable limit. For this we need to cut away enough combinations to match this requirement while at the same time keep as much efficiency as possible. As we have the probabilities of all candidates, we found that the product of the probabilities of the child particles

$$p_{\text{prod}} = \prod_{i}^{N_{\text{child}}} p_{\text{child i}} \tag{11.1}$$

is a very powerful quantity to cut on, in order to reduce the number of candidates drastically while keeping a fair amount of signal candidates. We could, of course, also cut on the probability of the particle itself, but the advantage is that it is orders of magnitude faster calculated than the probability obtained by NeuroBayes, as for this, all variables need to be calculated. By cutting on the product of the probabilities of the child particles p_{prod} beforehand, one avoids these expensive calculations for the majority of obviously wrong combinations.
The only remaining question is, where to place the cut for all the individual decay channels as this quantity is not comparable between different decay channels. As the purities vastly differ in different decay channels, the criteria for the cut depends on how many right combinations compared to how many wrong combinations drop out with a harder cut. This can be visualized as the slope by plotting the number of remaining signal candidates against the number of remaining background events after applying several cuts as shown in figure 11.1 exemplary for the D^0 modes. The cuts are chosen such that the slope of all the lines corresponding to the different decay channels are the same. The slopes for the different particle types are then chosen such that the total efficiency is as high as possible while the throughput is still acceptable.
Another constraint that needs to be mentioned is the average number of events where the full reconstruction module finds at least one B meson candidate, the so-called full reconstruction skim efficiency. For the usage afterwards the user needs to process each event in which the full reconstruction module found at least one candidate. By reducing the skim efficiency, the average processing time for all analyses using this tool will be reduced by this factor.

11.1. The Best $\mathrm{B}_{\mathrm{Tag}}$ Candidate Selection

As the chance to correctly reconstruct two B mesons in one event is negligible, it makes sense to assume that there can be only one correct combination in one event. Knowing this, if we select the correct candidate and reject all other combinations in one event, immediately reduces the background. As we do not know which combination is the correct one, we are forced to select the best one. For this we sort the candidates according to their probability and suggest the candidate with the highest probability as the best candidate for the user. In some analyses, the background level doesn't play a role and therefore we save the five best candidates in order to increase the efficiency for such analyses further.

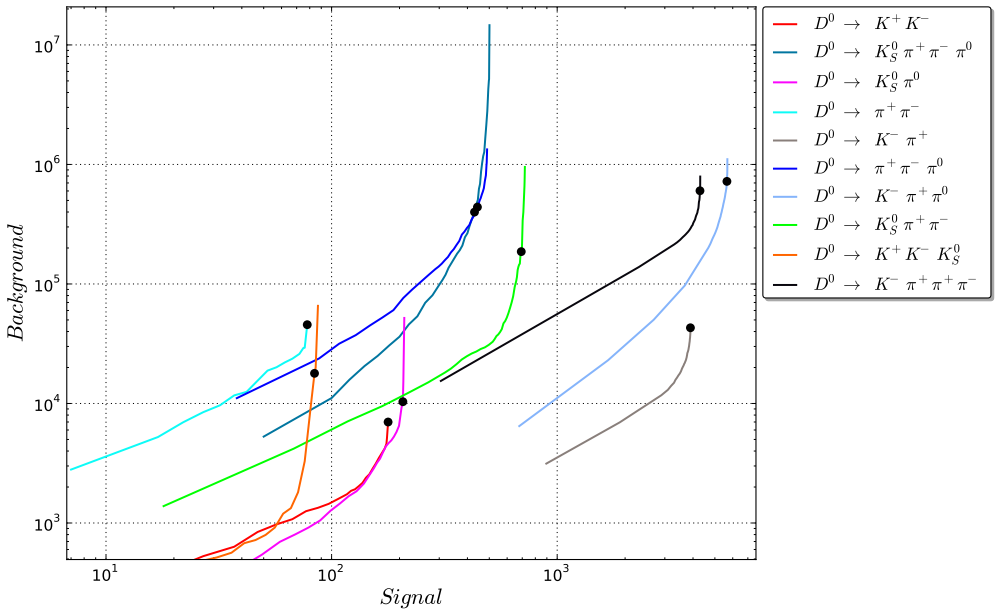


Figure 11.1. The signal-background plots for the D^0 cut determination. The black dots are our cutting points and all the lines have the same slope in these points.

12. Suppression of non $B\bar{B}$ Background

So far we only trained the networks on information about being signal or background. But the background is divided in two subclasses. There are $\Upsilon(4S)$ events decaying into $B\bar{B}$ pair and there are events of the type $e^+e^- \to u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ where no $B\bar{B}$ pair is produced. Of course candidates reconstructed in continuum events are also background as well as wrong combinations in the $\Upsilon(4S)$ events, but the kinematics of these continuum events are completely different. Therefore it is possible to separate to some extent between continuum events and $\Upsilon(4S)$ events independent of any B meson reconstruction. This, of course, helps to further suppress the background in the full reconstruction sample using independent information.

12.1. Topological Variables

12.1.1. Fox-Wolfram Moments

The so-called Fox-Wolfram moments [Fox+78] characterize the shape of an event. As shown in figure 12.1, the $\Upsilon(4S)$ events are much more spherical than the jet-like topologies of the continuum events. The reason for this is that the B mesons are nearly at rest due to the nearby threshold, while the lighter mesons have a large momentum and therefore a jet-like structure. The Fox-Wolfram moments are able to parametrize this topological structure.
In addition, there is a method called the super Fox-Wolfram moments [Lee+03] (SFWM) where the moments for $\mathrm{B_{Tag}}$ side and the $\mathrm{B_{Sig}}$ side are calculated separately and combine them using a fisher discriminant. As this might introduce a efficiency bias because the super Fox-Wolfram moments explicitly depend on the signal-side signature, these moments are treated separately.

12.1.2. Thrust Angle

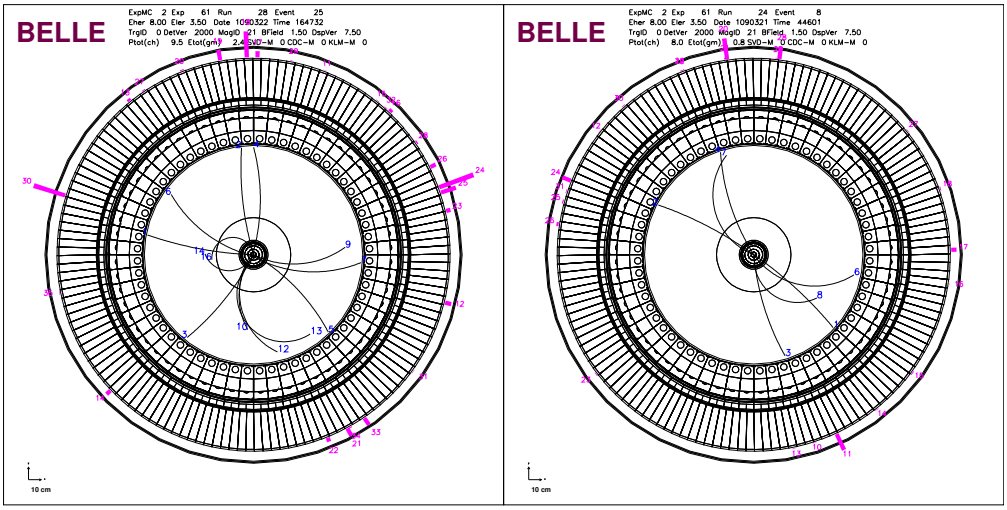
The thrust axis is defined as the vector where the alignment of all other tracks is maximal. The thrust angle is the angle between the thrust axis of the $\mathrm{B_{Tag}}$ candidate and the thrust axis of the rest of the event. As $B\bar{B}$ events are spherical, we expect this angle to be distributed flat, while for jet-like continuum events we expect this angle to be small.

12.1.3. Momentum Direction of the $\mathrm{B_{Tag}}$ candidate

The angle $cos\Theta_B$ is the angle between the z-axis and the momentum of the $\mathrm{B_{Tag}}$ candidate in the center-of-mass system. This angle is expected to be flat for continuum events, but peaks at zero for B mesons due to angular momentum conservation.

12.2. Continuum Suppression Module

In order to take the topological information into account, but still be able to avoid possible biases in special analyses, a completely separated module was built which can be optionally used by the user, depending on his analysis. There are three possible usage modes for the full reconstruction module.



(a) A typical $\Upsilon(4S)$ event with a spherical topology.

(b) A typical continuum event with a jet-like topology.

Figure 12.1.

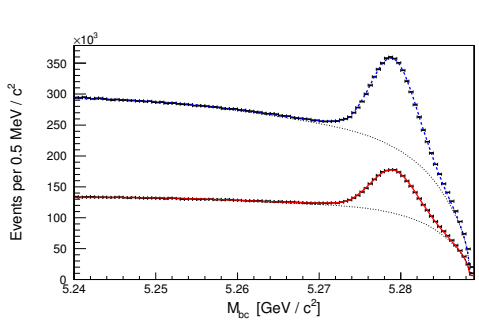
1. Without any continuum suppression.
2. With continuum suppression, only depending on the B_{Tag} candidate, i.e. without the SFWM.
3. With continuum suppression, including the SFWM.

The resulting improvements due to the additional information are discussed in chapter 13.

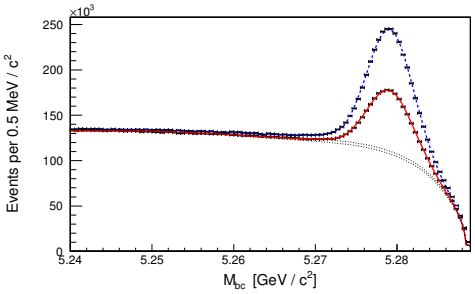
13. Resulting Performance

13.1. Fits to the M_{bc} Distribution

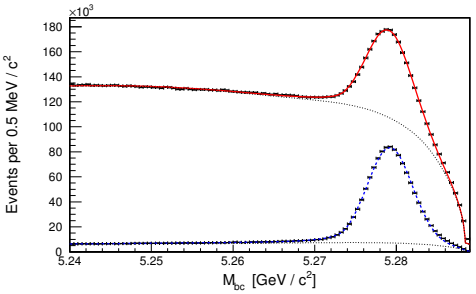
We are now able to compare the efficiency of the new full reconstruction approach to the classical cut-based tool, already existing and used in the Belle collaboration (e.g. [Lee+03; Liv+08; Che+07; Ika+06]). We can estimate the efficiency and the purity using a fit to the M_{bc} distribution defined in equation 10.1. The results are shown in figure 13.1. It can be seen that we have indeed reached a much higher efficiency, even with the same purity of the classical tool. These M_{bc} fits are done with the complete data sample recorded by the Belle detector. According to the fit, we reconstructed over 2 million B^0 mesons and over 3 million B^+ mesons. We can perform



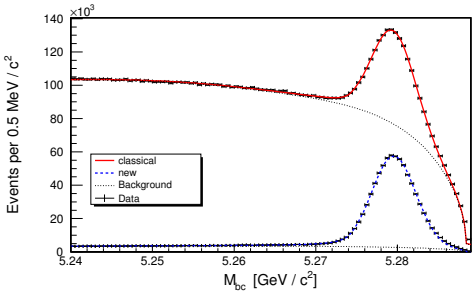
(a) B^+ selection with roughly equal purity



(b) B^+ selection with roughly equal background level



(c) B^+ selection with roughly equal efficiency



(d) B^0 selection with roughly equal efficiency

Figure 13.1. M_{bc} plots for different selections: The dashed blue line is a fit of the M_{bc} distributions for the new full reconstruction algorithm, the solid red line to the classical one. The network cuts are chosen to have (a) roughly equal purity, (b) roughly equal background level, (c), (d) roughly equal efficiency compared to the classical one.

these fits for several different cuts on the NeuroBayes probability, resulting in a purity-efficiency curve. In figure 13.2 the purity-efficiency curves for charged and in figure 13.3 for the neutral B mesons are shown for the three different modes of the full reconstruction module explained in

section 12.2. In addition, the single (as one cannot choose different purities in the classical tool) purity-efficiency point for the classical tool is shown. We observe a factor of two higher efficiency for the new method compared to the classical one. For the same efficiency, the purity rises from 25% to over 80%.

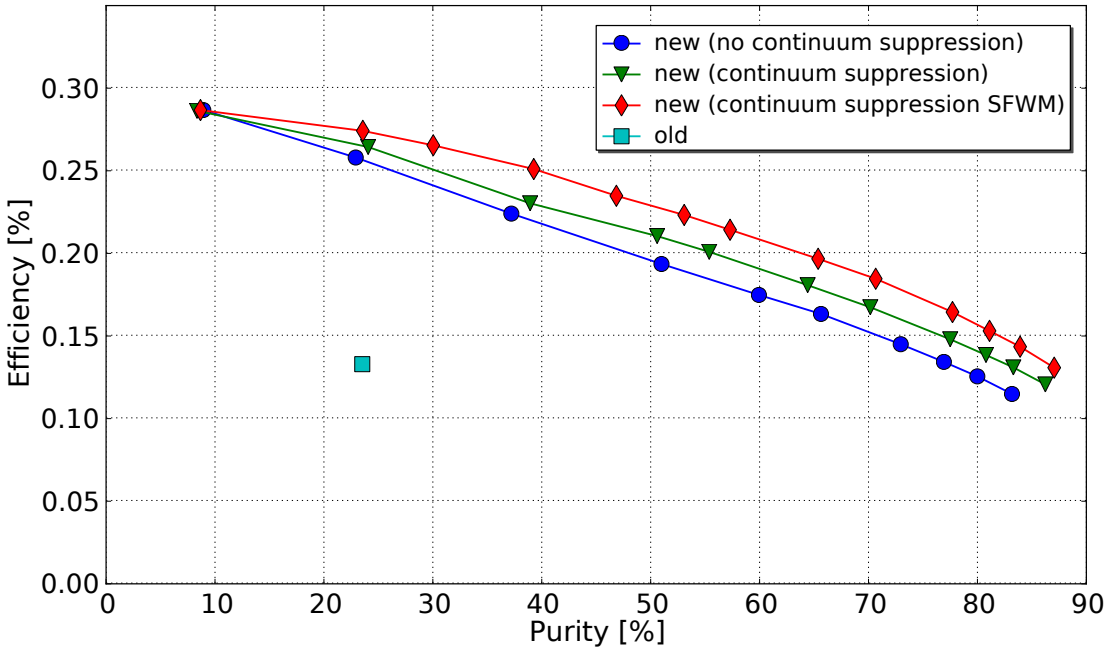


Figure 13.2. Purity-efficiency plot for B^+ mesons.

13.1.1. Without New Channels

If we exclude the newly added D and B decay channels from the full reconstruction and choose a network output cut to achieve the same background level as in the classical full reconstruction, the efficiency is increased by approximately 50% for B^0 mesons and 60% for B^+ mesons. A comparison of the individual B decay channels revealed that the largest improvement was achieved in modes with two or more light mesons, where the new full reconstruction does not impose any phase-space limits. The newly added channels make a valuable contribution of approximately 20% of the entire signal sample for both B^0 and B^+ mesons.

13.2. Efficiency Estimation Using a Signal-Side Analysis

Up to now, we only compared the efficiency of the $\mathrm{B_{Tag}}$ reconstruction alone. But in the end, the efficiency increase of a possible signal-side analysis matters. For this, we reconstruct the decay

$$B^0 \to D^{*-}\ell^+\nu_\ell \tag{13.1}$$

on the signal-side. A kinematic variable used to distinguish correctly reconstructed signal candidates from background candidates is the missing mass, defined as

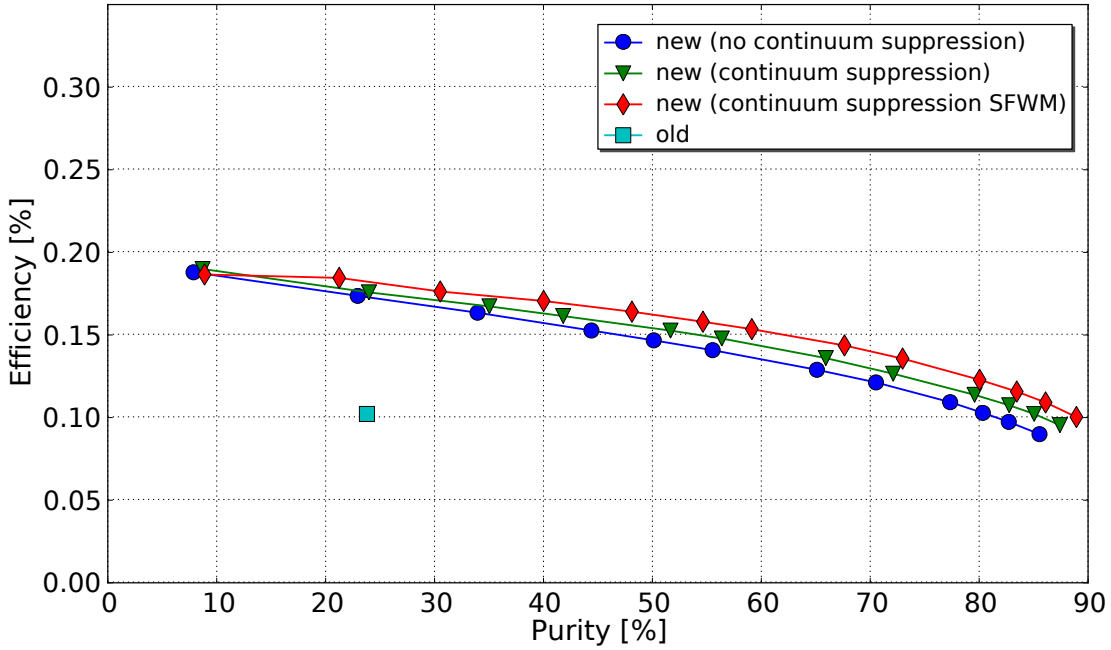
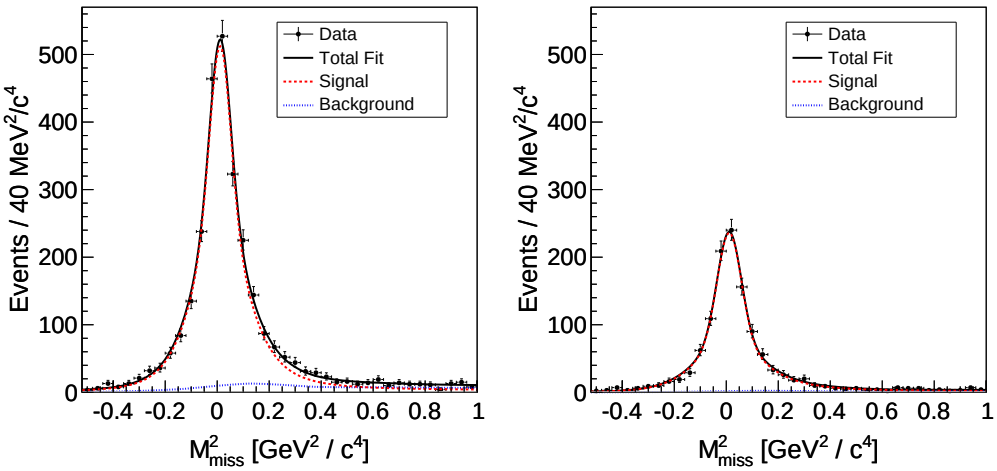


Figure 13.3. Purity-efficiency plot for B^0 mesons.

$$M^2_{\text{miss}} = |p_{\Upsilon(4S)} - (\sum_i p_i + p_{B_{\text{tag}}})|^2, \tag{13.2}$$

where $p_{\Upsilon(4S)}$ denotes the four-momentum of the $\Upsilon(4s)$ resonance, $p_{B_{\text{tag}}}$ is the four-momentum of the B_{tag} and $\sum_i p_i$ is the sum of the four-momenta of the reconstructed particles on the signal side. Because the neutrino is the only missing particle in this decay, we expect the missing mass to be zero for signal events. The result can be seen in figure 13.4a for the new full reconstruction algorithm and as a comparison in figure 13.4b the result for the classical full reconstruction algorithm. A clear peak is observed at the expected position with similar resolutions for new and classical full reconstruction. Thus despite the addition of less clean decay modes, the momentum resolution of the fully reconstructed B meson is preserved. As expected we also observe in this applied example a significant increase of efficiency.



(a) $B^0 \to D^{*-}\ \ell^+\ \nu_\ell$ - new full reconstruction **(b)** $B^0 \to D^{*-}\ \ell^+\ \nu_\ell$ - classical full reconstruction

Figure 13.4. Missing mass distributions for $B^0 \to D^{*-}\ \ell^+\ \nu_\ell$ decays of the new and classical full reconstruction tool.

14. The `ekpturbo` Module

The data, recorded by the Belle detector is stored in a row-wise data structure and consists of several petabytes. Because of the row-wise structure, it is not possible to add additional information, like our B_{Tag} candidate, for each event. Due to the limited capacity of the storage, it would also not be possible to copy the whole data sample including the new items. This forces each user to rerun the full reconstruction module over and over again.
In order to avoid this expensive waste of computation time we developed the `ekpturbo` module which saves the information of the found B_{Tag} candidates separately. If they are already stored for a given event, the `ekpturbo` module loads it, instead of recalculating it as shown in figure 14.1. This is completely transparent for the user, he only recognizes an order of magnitude faster processing time.

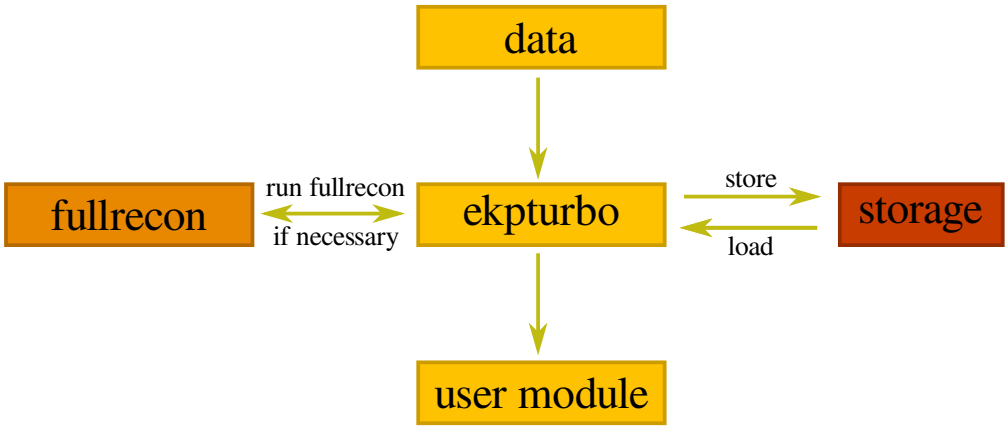


Figure 14.1. The drafted work flow of the `ekpturbo` module

15. Conclusion

We have developed an improved full reconstruction algorithm for the Belle experiment by introducing a hierarchical selection procedure. Instead of cutting away candidates at the early stages, we postpone the decision to later stages by very soft selections on the product of their Bayesian signal probability and giving this probability as an input for the higher stages networks. Together with a higher separation power of the neural networks compared to a cut based selection, this enabled us to reconstruct more decay channels with an acceptable computation time. Depending on the analysis, we expect an overall improvement of the effective luminosity of roughly a factor of 2 for a large number of analyses relying on the full reconstruction and could be also verified on data using a $D^{(*)}\ell\nu$ selection on the signal side.
This remarkable improvement correspons directly to several years of data taking and therefore saves a lot of money. Moreover, the physics results can reach a sensitivity which could otherwise be reached only by waiting for an equivalent amount of Belle II data for several years from now on. This is a priceless competitive advantage with respect to other experiments like LHCb.
The developed module is now part of the official Belle software package and is already used for many analyses.

Part V.

The Analysis $B \to K^{(*)}\nu\bar{\nu}$

16. Introduction

This part describes the search for the rare B meson decays $B \to K^{(*)}\nu\bar{\nu}$ using the new probabilistic full reconstruction approach explained before. Significant improvement in the limit estimation of the branching ratios of these decays can be observed mainly due to the higher efficiency of the full reconstruction method. Another improvement can be achieved by using NeuroBayes neural networks for the selection.

16.1. Overview

The Standard Model gives in general the biggest contributions to tree-level decay amplitudes and is precisely measured to be consistent with the predictions inside the uncertainties in many analyses. However, if a process is highly suppressed in the Standard Model, there is a chance that contributions from physics beyond Standard Model can be as large or even larger than the contribution from the Standard Model. Therefore, deviations of the order of percent up to orders of magnitude could be possible. That is the reason why the so-called rare decays are so interesting. One such interesting process is the quark transition $b \to s\nu\bar{\nu}$. The Standard Model processes facilitating such a transition are shown in figure 16.2. As these are higher order processes they are highly suppressed compared to tree level processes, which is forbidden in this case as it is a flavor changing neutral current. This quark process occurs in the decays $B^+ \to K^+\nu\bar{\nu}$, $B^+ \to K^{*+}\nu\bar{\nu}$, $B^0 \to K^0\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$. The theoretical predictions for these decays are very precise because there is only one hadron in the final state and no charged lepton as in the $b \to s\ell^+\ell^-$ transition. Moreover, these decays are enhanced by a factor of three due to the three neutrino generations, compared to $b \to s\ell^+\ell^-$ transition. Theoretical predictions for the branching ratios are listed in table 16.1.

decay	branching ratio	cite
$B \to K\nu\bar{\nu}$	$< 5 \cdot 10^{-5}$	[Buc+00]
$B \to K^*\nu\bar{\nu}$	$1.3 \cdot 10^{-5}$	[Buc+00]
$B^+ \to K^+\nu\bar{\nu}$	$4.5 \cdot 10^{-6}$	[Alt+09]
$B \to K^*\nu\bar{\nu}$	$6.8 \cdot 10^{-6}$	[Alt+09]

Table 16.1. Theoretical Standard Model predictions for the $b \to s\nu\bar{\nu}$ decay modes.

In models beyond the Standard Model, these decays can be enhanced by orders of magnitude. There are models with a nonstandard Z coupling [Buc+00], fourth quark models [Bur+10] and models with extra dimensions [Col+06] which can have significant contributions to the branching ratios of these transitions. In non minimal flavor violating supersymmetric models, the decay can be mediated by a light neutralino. The various models are briefly summarized in [Bev11]. Once these decays are discovered it would be very interesting to measure the longitudinal polarization f_L as this has huge discrimination power between the various extensions of the Standard Model.

It is possible to paramtetrize the contributions with the two parameters in terms of left and right handed Wilson coefficients, $C^{\nu}_{L,R}$ via ϵ and η, where

$$\epsilon = \frac{\sqrt{|C^{\nu}_{L}|^2 + |C^{\nu}_{R}|^2}}{|C^{\nu}_{L}|_{SM}} \tag{16.1}$$

and

$$\eta = -\frac{\Re(C_L C_R)}{\sqrt{|C_L|^2 + |C_R|^2}} \, . \tag{16.2}$$

For the Standard Model one expects $\epsilon = 1$ and $\eta = 0$.
A precise measurement would give a strong constraint to these models. This can be seen in figure 16.1 where the ϵ and η plane is cutted already with the current experimentsl limits. With a potential high luminosity B-factory such as the Belle II experiment one can probe the Standard Model expectation with a high precision. There are already measurements done by Belle, BaBar

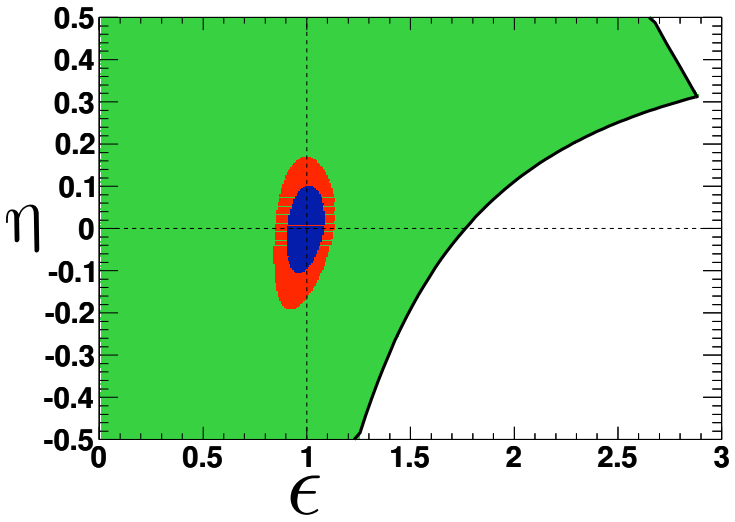


Figure 16.1. The constraint on ϵ and η expected using exclusive branching fraction and f_L measurements made with data sample of $75\,ab^{-1}$ at a potential high luminosity B-factory. The central two contours represent the 68% and 95% confidence level (C.L.) constraint obtained at a potential high luminosity B-factory, while the light (green) contour indicates the existing constraint obtained using limits on the $B \to K^{(*)}\nu\bar{\nu}$ modes.

and Cleo and listed in table 16.2.
The obvious disadvantage of these decays is that they are rare and therefore one needs very high statistics and efficiencies. In order to be able to detect some few signal events it is crucial to suppress most of the background. But, as there are two undetectable neutrinos in the final state, we cannot exclusively reconstruct these decays and therefore cannot distinguish efficiently the signal decays from background. But with the full reconstruction method explained in part IV and in [Fei+11] we can compensate the missing neutrinos and suppress the background by many orders of magnitude.

Decay	Branching ratio	Experiment	Cite
$B \to K^+ \nu\bar{\nu}$	$< 1.3 \cdot 10^{-5}$	BaBar	[AS+10]
	$< 1.4 \cdot 10^{-5}$	Belle	[Che+07]
	$< 2.4 \cdot 10^{-4}$	Cleo	[Bro+01]
$B \to K^0 \nu\bar{\nu}$	$< 5.6 \cdot 10^{-5}$	BaBar	[AS+10]
	$< 1.6 \cdot 10^{-4}$	Belle	[Che+07]
$B \to K^{*+} \nu\bar{\nu}$	$< 8 \cdot 10^{-5}$	BaBar	[Aub+08]
	$< 1.4 \cdot 10^{-4}$	Belle	[Che+07]
$B \to K^{*0} \nu\bar{\nu}$	$< 1.2 \cdot 10^{-4}$	BaBar	[Aub+08]
	$< 3.4 \cdot 10^{-4}$	Belle	[Che+07]

Table 16.2. Current experimental results.

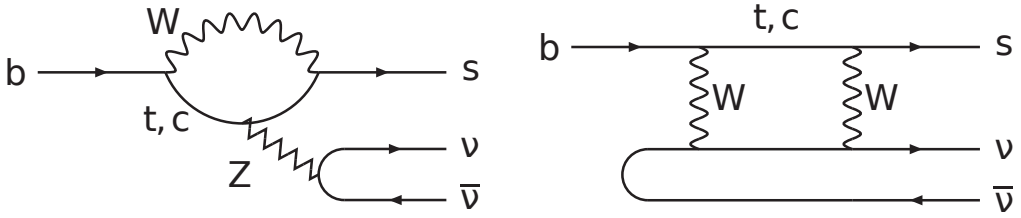


Figure 16.2. The Feynman diagrams for the $b \to s\nu\bar{\nu}$ transition in the Standard Model.

16.2. Outline of the Analysis

The general procedure will be, to reconstruct a B_{Tag} candidate using the full reconstruction method and then reconstruct the signal-side in the modes $B^+ \to K^+\nu\bar{\nu}$, $B^0 \to K^0_S\nu\bar{\nu}$, $B^0 \to K^{*0}\nu\bar{\nu}$ and $B^+ \to K^{*+}\nu\bar{\nu}$. The sum of the remaining energy measured by the calorimeter E_{ECL} is then used to estimate the number of signal events, because we expect a peak at low E_{ECL} values for signal events. We perform a counting experiment where we count the number of events in the signal-box $E_{ECL} < 0.3\,\text{MeV}$ and estimate the number of expected background events using Monte Carlo simulated events together with the sideband region $0.45\,\text{MeV} < E_{ECL} < 1.2\,\text{MeV}$ in the data sample.
In order to avoid a bias in the selection, this analysis is performed 'blind', i.e the signal-box in the data sample is opened only after finishing the whole procedure and selection criteria.

17. Candidate Reconstruction

17.1. Used Data Samples

For this analysis, the complete data sample of 702.6 fb^{-1} collected with the Belle experiment on the $\Upsilon(4S)$ resonance is used, which corresponds to $771 \cdot 10^6 \pm 1.4\%$ $B\bar{B}$ pairs. Because of a detector upgrade, the data sample is divided into two parts, called SVD1 and SVD2. In order to take advantage of the improved detector, a better track finding algorithm is used for a complete processing of the data, called CASE-B data sample. In this analysis, CASE-B data is used.

There are officially produced Monte Carlo samples corresponding to 5 times the amount of recorded data (5 streams). It is called generic Monte Carlo because all the particles produced in the e^+e^- collision are decayed generically with the dominating quark-level process $b \to c + W$ or the continuum processes $e^+e^- \to (u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c})$. These production processes are grouped into 4 categories (see table 17.1). Two other official Monte Carlo samples are used, the *rare* Monte

name	Description	rel. amount
charged	$\Upsilon(4S)$ decays into generically decaying charged B mesons	12 %
mixed	$\Upsilon(4S)$ decays into generically decaying neutral B mesons	12 %
charm	continuum $e^+e^- \to c\bar{c}$ decays	30 %
uds	continuum $e^+e^- \to u\bar{u}, dd, s\bar{s}$ decays	46 %

Table 17.1. Quality cuts for charged tracks

Carlo sample which covers all B decays with non $b \to c + W$ transitions corresponding to 50 times the data luminosity and the *ulnu* Monte Carlo sample which is a dedicated $b \to ul\nu$ sample corresponding to 20 times the data luminosity.

For each mode, a signal Monte Carlo sample was produced where one of the B mesons decays into the respective specific signal final state, while the other one decays via generic $b \to c + W$ transitions. The decay is generated using EvtGen [Lan01] with the modified decay amplitudes described in [Buc+00] and the form factors calculated in [Bal+05b] and [Bal+05a].

All the samples are processed with the full reconstruction module described in chapter 7. It is known from studies in chapter 7, using the full reconstruction module with the continuum suppression switched on, including the super-fox-wolfram moments (SFWM) gives the highest efficiency. Therefore this mode is used for the processing of the data. By applying the same cut on the B_{Tag} NeuroBayes output for example in the K^+ mode an increase in efficiency of 50% was found, compared to the processing without the continuum suppression in the full reconstruction module. But it has to be noted that the absolute value of this cut is not really comparable between both modes (with and without continuum supression) but at least, the efficiency is higher with this cut and can therefore be changed later on if necessary. Also the purity should be higher at a cut with the same efficiency because of less wrongly selected B_{Tag} candidates. This mode is used for the following analysis.

17.2. Track Selection Criteria

The tracks used to reconstruct the signal-side have to originate from a region around the interaction point (IP). In order to have at least the chance to measure the track momentum in the drift chamber, the tracks need to have a transversal momentum p_t of at least 0.1 GeV. All the

Cut (short name)	Description
$\lvert dr \rvert$ <2 cm	impact parameter in radial direction
$\lvert dz \rvert$ < 5 cm	impact parameter in beam direction
$p_t > 0.1$ GeV	lower cut on the transversal momentum

Table 17.2. Quality cuts for charged tracks

applied cuts are summarized in table 17.2

17.3. Candidate Selection Criteria

There are five different decay modes, K^+,K^0_S, $K^{*+} \to K^0_S\pi^+$,$K^{*+} \to K^+\pi^0$ and $K^{*0} \to K^+\pi^-$. The reconstruction criteria for these modes are:

- $K^\pm$ **candidates** are tracks with a particle identification likelihood ratio $\mathcal{R}_K \equiv \mathcal{L}_K/(\mathcal{L}_K + \mathcal{L}_\pi)$ of greater than 0.6. This particle identification is explained in chapter 6.
- $\pi^\pm$ **candidates** are tracks with a particle identification likelihood ratio $\mathcal{R}_K \equiv \mathcal{L}_K/(\mathcal{L}_K + \mathcal{L}_\pi)$ of smaller than 0.4.
- π^0 **candidates** are formed from two clusters in the electro-magnetic calorimeter that are not matched with a charged track and have an energy of at least 50 MeV. The IP is taken as the decay vertex, which results in a tail to lower invariant masses for π^0 candidates decaying further in the detector [Che+05]. The invariant mass has to lie between 117.8 MeV and 150.2 MeV. In order to reduce combinatorial background from combinations with noise clusters, the asymmetry
$$a_\gamma = \left| \frac{E_{\gamma 1} - E_{\gamma 2}}{E_{\gamma 1} + E_{\gamma 2}} \right|$$
between both clusters should be smaller than 0.9.
- K^0_S **candidates** are formed from two oppositely charged tracks. The so-called `GoodKS` selection criteria is required. Their invariant mass must be within $\pm 15\,\mathrm{MeV/c^2}$ of the nominal mass [Nak+10]. The distance of closest approach of the candidate charged tracks to the IP in the plane perpendicular to z axis is required to be larger than 0.02 cm for high momentum ($> 1.5\,\mathrm{GeV/c}$) K^0_S candidates and 0.03 cm for those with momentum less than 1.5 GeV/c. The $\pi^+\pi^-$ vertex is required to be displaced from the IP by a minimum transverse distance of 0.22 cm for high-momentum candidates and 0.08 cm for the remaining candidates. The mismatch in the z direction at the K^0_S vertex point for the $\pi^+\pi^-$ tracks must be less than 2.4 cm for high-momentum candidates and 1.8 cm for the remaining candidates. The direction of the pion pair momentum must also agree with the direction of the vertex point from the IP to within 0.03 rad for high-momentum candidates, and to within 0.1 rad for the remaining candidates [Che+05]. The invariant mass distribution can be seen in figure 17.1a.

- $K^{*\pm}$ **candidates** are reconstructed in two different decay modes:

  $$K^{*\pm} \to K^{+}\pi^{0}$$

  and

  $$K^{*\pm} \to K^{0}_{s}\pi^{+} \ .$$

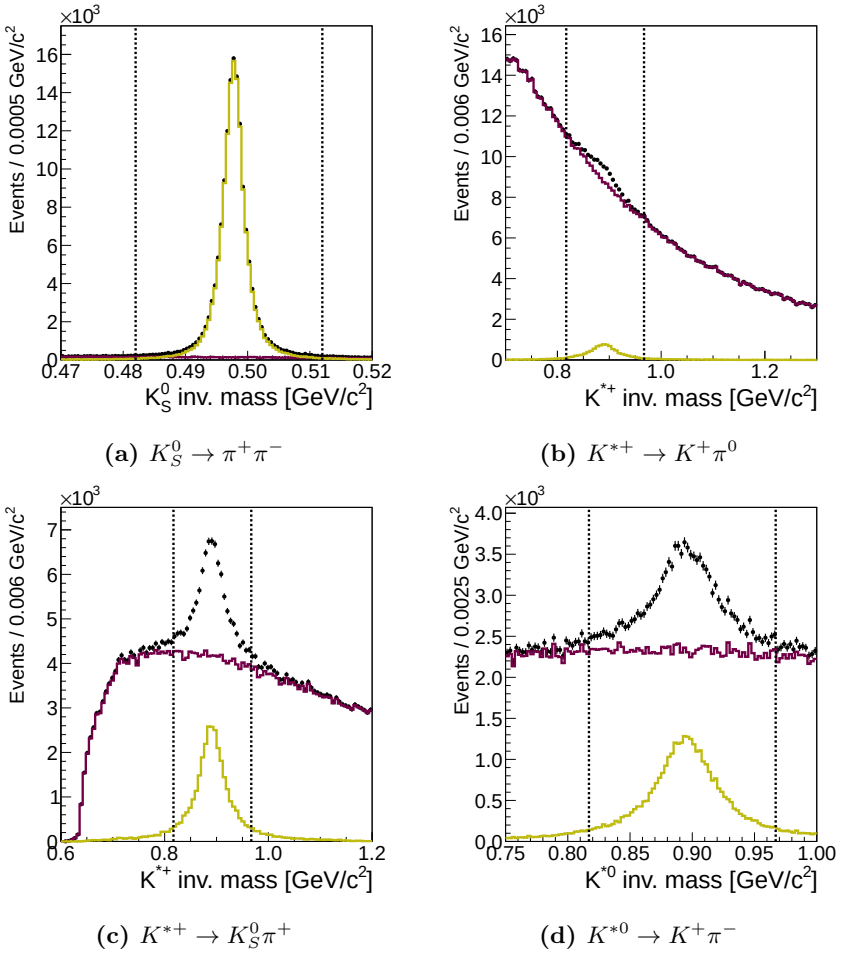
  Their invariant mass must be within $\pm 75\,\mathrm{MeV/c^2}$ of the nominal mass [Nak+10]. The invariant mass distribution of both decays can be seen in figure 17.1b and figure 17.1c respectively.

- K^{*0} **candidates** are formed from a charged kaon and an oppositely charged pion. Their invariant mass must be within $\pm 75\,\mathrm{MeV/c^2}$ of the nominal mass [Nak+10]. The invariant mass distribution can be seen in figure 17.1d.

All the applied cuts are summarized in table 17.3. As be can seen in figure 17.1, the cuts are chosen such that only a small fraction of signal is lost, without further optimizing these cuts.

Particle	Cut (short name)	Description
K^+	atc_PID > 0.6	lower cut on the particle identification likelihood ratio
π^+	atc_PID < 0.4	upper cut on the particle identification likelihood ratio
π^0	117.8 MeV< $m(\pi^0)$ <150.2 MeV	invariant mass cut
	E_γ >50 MeV	lower cut on the gamma energy
	a_γ <0.9	upper cut on the energy asymmetry of the two photons
K^0_S	$\lvert m(K^0_S) - M_{K^0_S}\rvert$ <15 MeV	invariant mass cut on the difference to the nominal mass $M_{K^0_S}$
	GoodK_S > 0	A set of cuts addressing the displaced vertex of the K^0_S from the IP (see text)
K^{*+}	$\lvert m(K^{*+}) - M_{K^{*+}}\rvert$ <75 MeV	invariant mass cut on the difference to the nominal mass $M_{K^{*+}}$
K^{*0}	$\lvert m(K^{*0}) - M_{K^{*0}}\rvert$ <75 MeV	invariant mass cut on the difference to the nominal mass $M_{K^{*0}}$

Table 17.3. Summary of the selection criteria for the reconstruction of light meson candidates



(a) $K_S^0 \to \pi^+\pi^-$ **(b)** $K^{*+} \to K^+\pi^0$

(c) $K^{*+} \to K_S^0\pi^+$ **(d)** $K^{*0} \to K^+\pi^-$

Figure 17.1. The invariant mass distributions for the different strange mesons decays. The black dots show the reconstructed invariant mass from the generic MC sample corresponding to 5 times the data. The green line shows the distribution of the correct reconstructed mesons, the purple line the background distribution.

18. Cut-based Analysis

This cut-based analysis is based on the analysis of Kai-Feng Chen [Che+07]. All the cuts are taken from there and not further optimized as most of them are physically motivated and still give optimal results.

18.1. Signal-Side Selection Criteria

In order to reduce the number of reconstructed candidates drastically, some selection criteria need to be fulfilled by each candidate. As already mentioned above, there needs to be a B_{Tag} candidate with the correct charge combination together with the signal-side candidate (neutral-neutral, positive-negative, negative-positive). If there are more than one B_{Tag} candidate in one event, the one with the highest NB_{out} is taken.

- **B_{Tag}**
  The NB_{out} for the B_{Tag} is calculated, taking information for continuum suppression into account, as explained before. In addition, this B_{Tag} candidate has to have a NB_{out} of at least 0.02. This cut is chosen because it gives a good limit near the optimum for all modes and also gives a similar background level as in the analysis of Kai-Feng Chen [Che+07]. The distribution is shown in figure 23.2.

  For the final selection, also a cut on M_{bc} and ΔE of the B_{Tag} is performed as seen in table 18.1.

- **Veto**
  Because we expect nothing else in an event than the B_{Tag} and the signal-side B_{Sig} candidate, we veto candidates where additional tracks, or additional π^0 candidates are left over. For π^0 candidates, the same quality cuts are applied as mentioned above. For the veto track candidates, all detected tracks are taken, without any quality cuts.

- **Missing momentum**
  The missing momentum $\mathbf{p}_{Mis}$, defined as

$$\mathbf{p}_{Mis} = \mathbf{p}_{Beam} - (\mathbf{p}_{BTag} + \mathbf{p}_{Sig}) \,, \tag{18.1}$$

  where $\mathbf{p}_{Beam}$ is the four-momentum of the beam and $\mathbf{p}_{BTag}$, $\mathbf{p}_{Sig}$ are the four-momenta of the B_{Tag} and the B_{Sig} respectively, should not point into the direction of the beam-pipe, in order to reduce background from particles escaping through the beam-pipe. The angle θ_{Mis} between the missing-momentum and the beam-pipe in the laboratory rest frame should therefore satisfy $-0.86 < cos\theta_{Mis} < 0.95$.

- **Extra energy in the calorimeter**
  The summed extra energy in the calorimeter E_{ECL} should not exceed 1.2 GeV. The energies of the clusters in the calorimeter added to E_{ECL} have to exceed the thresholds
  - Forward end-cap: 100 MeV.

- Barrel: 50 MeV.
- Backward end-cap: 150 MeV.

These thresholds are found to separate well between real photons and noise and also reject the beam background in forward and backward direction in which we are not interested in this analysis.

This window is further divided into the signal-box $E_{ECL} < 0.3\,\text{GeV}$ and the sideband $0.45\,\text{GeV} < E_{ECL} < 1.2\,\text{GeV}$, where only a fraction of signal signal events are found as shown in figure 18.1.

- **Momentum of the light meson**
  In order to reduce background from $b \to c$ decays, we perform a lower cut on the momentum of the light meson $K^{(*)}$ in the B_{Sig} rest frame $\vec{p}^{*} > 1.6\,\text{GeV}$. To reduce backgrounds from two-body B decays (e.g $B \to K\gamma$) an upper cut of $\vec{p}^{*} < 2.5\,\text{GeV}$ is performed.

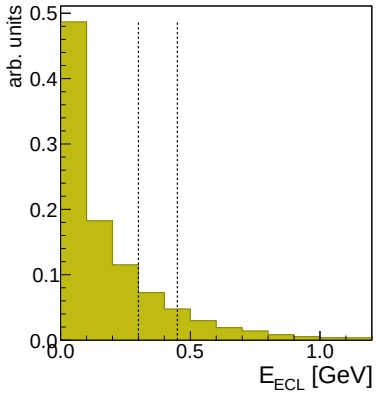
  All the applied cuts are summarized in table 18.1. The so called N-1 plots are shown in the appendix 23.1. In these plots, all cuts are applied, except for the cut on the plotted variable. In addition, the cut values for this variable are shown.

Cut (short name)	Description
`BTagChargeCombination`	B_{Tag} candidate with the corresponding charge combination (neutral-neutral, positive-negative, negative-positive)
$\text{M}_{\text{bc}} > 5.27\,\text{GeV}$	lower cut on M_{bc} of the B_{Tag} candidate
$-0.08\,\text{GeV} < \Delta\text{E} < 0.06\,\text{GeV}$	cut on ΔE of the B_{Tag} candidate
`BTagNBout` > 0.02	lower cut on the NB_{out} of the B_{Tag} candidate
`NRemainingPi0` $= 0$	no additional Pi^0 candidate should be left, standard Pi^0 cuts applied
`NRemainingTracksAll` $= 0$	no additional track should be left, no quality cuts on the veto tracks applied
$-0.86 < cos\theta_{mis} < 0.95$	the missing-momentum should not point into the beam-pipe
$E_{ECL} < 1.2\,\text{GeV}$	upper cut on the extra energy in the calorimeter
$\vec{p}^{*} > 1.6\,\text{GeV}$	lower cut on the momentum in B_{Sig} rest frame to reduce background from charm decays
$\vec{p}^{*} < 2.5\,\text{GeV}$	upper cut on the momentum in B_{Sig} rest frame to reduce background from two-body B decays

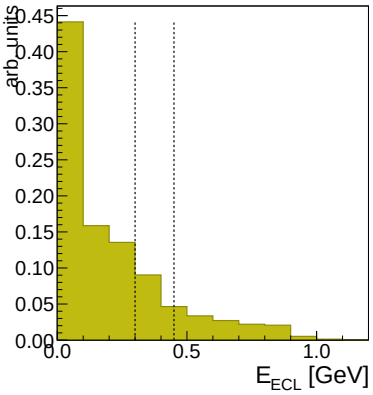
Table 18.1. A list of cuts for the cut-based selection

18.2. Comparison Between Monte Carlo and Data

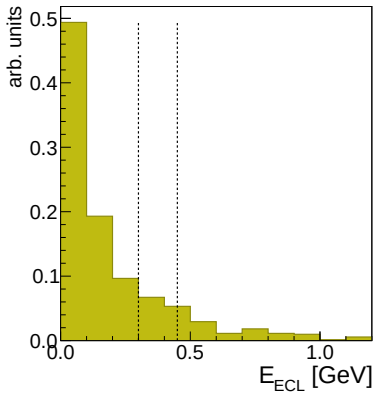
A comparison between the generic Monte Carlo sample and the data with these cuts applied is shown in figure 18.2. Since this is a blind analysis, these events are from the B_{Tag} M_{bc} sideband $M_{\text{bc}} < 5.27\,\text{GeV}$, where one expects no signal events. The data Monte Carlo agreement is really good. One has to keep in mind that the Monte Carlo is scaled by 0.2 since it corresponds to the luminosity of 5 times the data. It is not scaled to an equal number of entries as in the data sample.



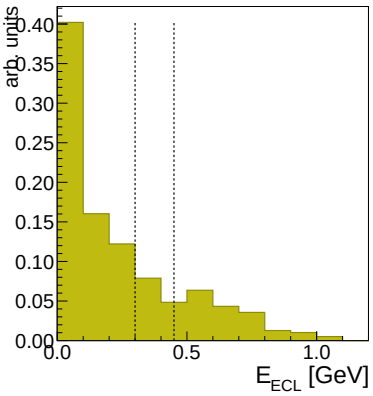
(a) K^+ mode



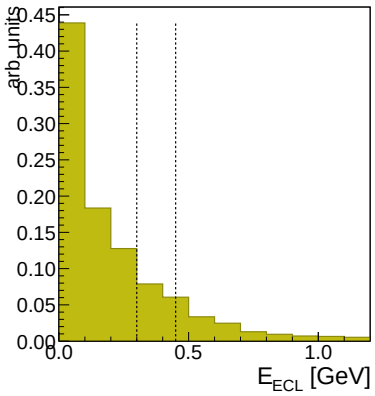
(b) K_S^0 mode



(c) $K^{*+} \to K^+\pi^0$ mode



(d) $K^{*+} \to K_S^0\pi^+$ mode



(e) $K^{*0} \to K^+\pi^-$ mode

Figure 18.1. The E_{ECL} distributions for signal events. The dashed lines border the signal-box on the left and the sideband region on the right.

There is also a quantitative result shown in table 18.2. In order to compare the two histograms a p-value was calculated, based on an algorithm [Gag06] which can also handle event-by-event weights. Here the Monte Carlo histogram is scaled to have the same total number of entries as the data histogram. A low p-value below 5% would indicate that both distributions are not compatible, but this is not the case. All decay modes agree well.
The full reconstruction method was not especially optimized in a way to obtain equal efficiencies on data and on Monte Carlo samples. Therefore, as for example some decay modes are experimentally only poorly known, one would not even expect to find such a good agreement. In order to find out if this is just pure luck, the ratios of the individual decay channels of the full reconstruction method are plotted in figure 18.3. The Monte Carlo samples for this comparison consist of generic and rare Monte Carlo, scaled to the data luminosity. For this comparison, the selection criteria explained above are applied, except that the $\mathrm{E_{ECL}}$ sideband is taken and the cuts on $cos\theta_{mis}$ and $\vec{p}^*$ are released. The different $\mathrm{B_{Tag}}$ decay channels are classified by their decay hash, explained in section 10.6. It seems that most of the decay channels are indeed compatible with one within their errors. Note that one cannot compare the yields with figures 18.2 as there different cuts were taken. The good agreement in figure 18.2 indicates that the $\mathrm{B_{Tag}}$ background level ($\mathrm{M_{bc}}$ sideband) is compatible between data and Monte Carlo, whereas the $\mathrm{B_{Tag}}$ signal efficiency is lower on data. The dependency of this ratio to $\mathrm{M_{bc}}$ can be seen in figure 18.4. For this comparison the $\mathrm{E_{ECL}}$ sideband is used and to increase statistics the cuts $cos\theta_{mis}$, $\vec{p}^*$ from table 18.1 are removed. The ratio between Monte Carlo and data is slightly higher in the $\mathrm{M_{bc}}$ signal window.
In order to take this into account, an $\mathrm{M_{bc}}$ fit has to be performed on data and Monte Carlo. From these fits, we obtain the number of $\mathrm{B_{Tag}}$ signal candidates $N_{B_{Tag}data}$ and $N_{B_{Tag}MC}$ for data and Monte Carlo respectively. The signal efficiency on data is then defined by

$$\epsilon_{Sigdata} = \frac{N_{B_{Tag}data}}{N_{B_{Tag}MC}} \times \epsilon_{SigMC} \ , \tag{18.2}$$

where ϵ_{SigMC} is the signal efficiency obtained by the signal Monte Carlo sample. The fits are shown in figure 18.5 and the resulting fit parameters are shown in table 18.3. These fits are done in the $\mathrm{E_{ECL}}$ sideband with the same selection criteria applied respectively. In order to increase the statistics, the cuts on $cos\theta_{mis}$ and $\vec{p}^*$ are removed which should not affect the $\mathrm{B_{Tag}}$ efficiency. The resulting ratios are listed in table 18.4 and visualized in figure 18.6. The uncertainties on the ratio is the quadratic sum of the relative uncertainties on the number of $\mathrm{B_{Tag}}$ candidates obtained by the fit.
If this approach really works is checked by comparing the normal MC samples with the same, but modified samples, where one arbitrary B_{Tag} decay channel is excluded ($B^+ \to \bar{D}^0\pi^+\pi^0$). The correction factor was obtained by M_{bc} fits of the B_{Tag} in the E_{ECL} sideband as explained above for the normal and the modified generic MC sample as

$$\frac{N_{B_{Tag}data}}{N_{B_{Tag}MC}} = 0.82 \pm 3.00\% \tag{18.3}$$

We can estimate the true efficiency of the modified sample with the signal MC by excluding the same B_{Tag} decay channel

$$\epsilon_{\mathrm{mod}} = 5.929 \cdot 10^{-04} \ . \tag{18.4}$$

As explained above, we can also estimate it using the normal signal MC sample and the B_{Tag} efficiency correction, obtained by the M_{bc} fits

$$\epsilon_{\mathrm{corr}} = \frac{N_{B_{Tag}data}}{N_{B_{Tag}MC}} \times \epsilon_{SigMC} = 5.965 \cdot 10^{-04} \ . \tag{18.5}$$

The double ratio is very well consistent with one

$$\frac{\epsilon_{\mathrm{mod}}}{\epsilon_{\mathrm{corr}}} = 0.994 \pm 0.035 \,, \tag{18.6}$$

This proofs that it is possible to correct for differences in the B_{Tag} efficiencies by fitting the M_{bc} distributions in the $\mathrm{E_{ECL}}$ sidebands on data and Monte Carlo separately.

mode	p-value
K^+	0.555228
K^0_S	0.631279
$K^{*+} \to K^0_S\pi^+$	0.74524
$K^{*+} \to K^+\pi^0$	0.8432
$K^{*0} \to K^+\pi^-$	0.6706

Table 18.2. Resulting p-values on how compatible the data distributions with the background hypotheses are. There is no significant deviation between Monte Carlo and data.

mode	data		MC	
	purity	N_{sig}	purity	N_{sig}
K^+	0.67	1158.9 ± 60.6	0.68	1449.6 ± 28.9
K^0_S	0.84	1463.9 ± 52.4	0.80	1918.4 ± 28.6
$K^{*+}(K^0_S\pi^+)$	0.66	1526.1 ± 73.2	0.72	2541.5 ± 37.4
$K^{*+}(K^+\pi^0)$	0.62	2374.4 ± 90.6	0.66	3001.0 ± 44.0
$K^{*0}(K^+\pi^-)$	0.68	1061.3 ± 55.5	0.68	1521.7 ± 29.0

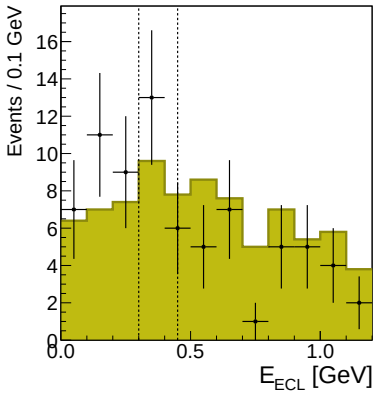
Table 18.3. Parameters of the fits to the $\mathrm{M_{bc}}$ distributions of the $\mathrm{B_{Tag}}$ candidates for data and Monte Carlo.

mode	ratio	error
K^+	0.80	$\pm$5.6%
K^0_S	0.76	$\pm$3.9%
$K^{*+} \to K^0_S\pi^+$	0.60	$\pm$5.0%
$K^{*+} \to K^+\pi^0$	0.79	$\pm$4.1%
$K^{*0} \to K^+\pi^-$	0.70	$\pm$5.6%

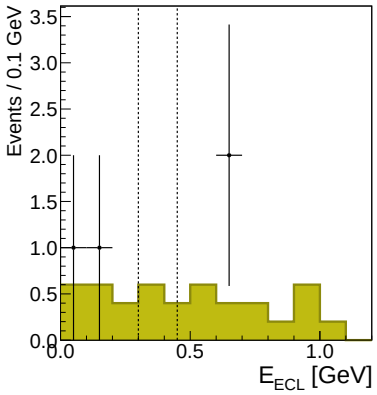
Table 18.4. Ratios of the $\mathrm{B_{Tag}}$ efficiencies between data and Monte Carlo.

18.3. Background Composition

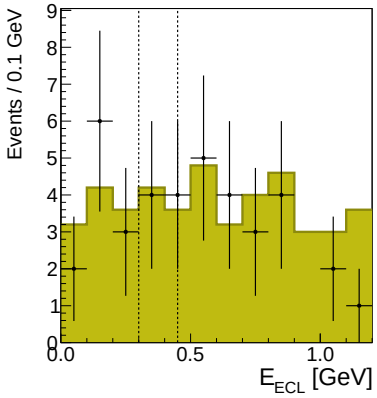
It is interesting to know the composition of the contributing background. In figure 18.7 one can see the different background contributions in the different decay modes. For charged B modes (K^+,K^{*+}) the biggest contribution comes from generic charged B decays, whereas for



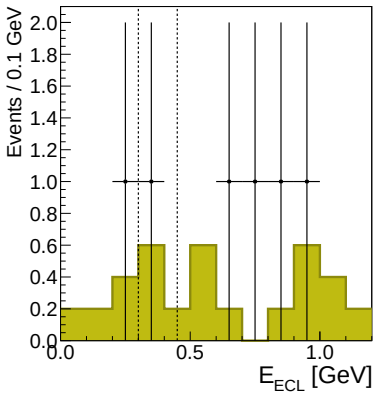
(a) K^+ mode



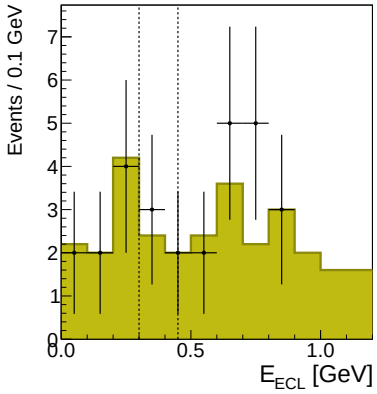
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

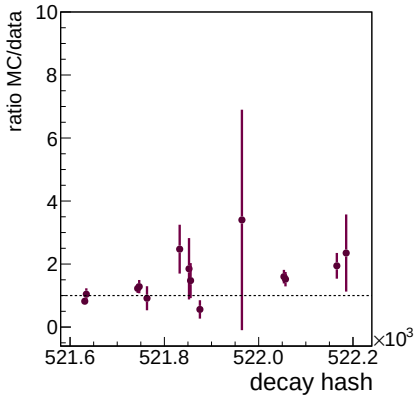


(d) $K^{*+} \to K^0_S\pi^+$ mode

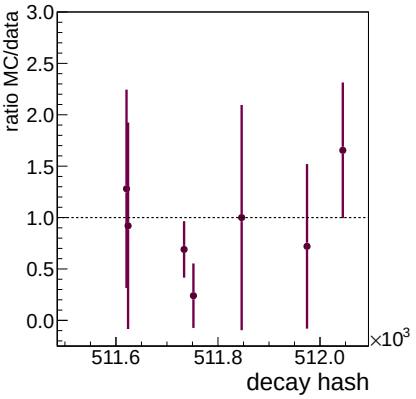


(e) $K^{*0} \to K^+\pi^-$ mode

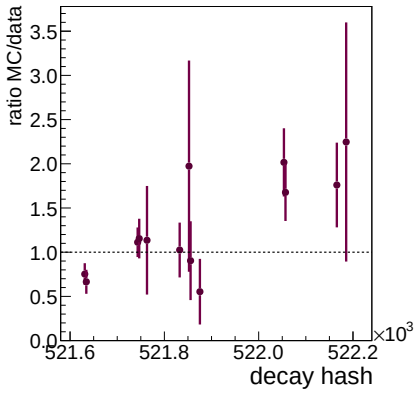
Figure 18.2. The E_{ECL} distributions for the different strange mesons decays in the M_{bc} sideband with the cut-based method. The black dots show the E_{ECL} distribution from data, and the purple solid line from the generic MC sample scaled to the luminosity of the data. The dashed lines mark the boundaries of the signal window (left) and sideband window (right).



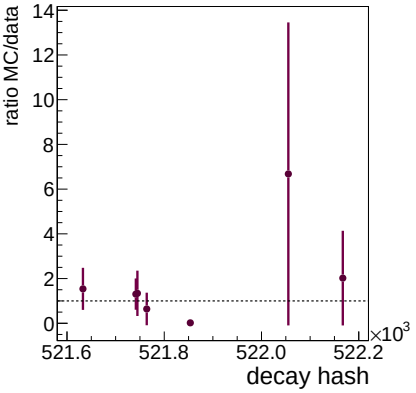
(a) K^+ mode



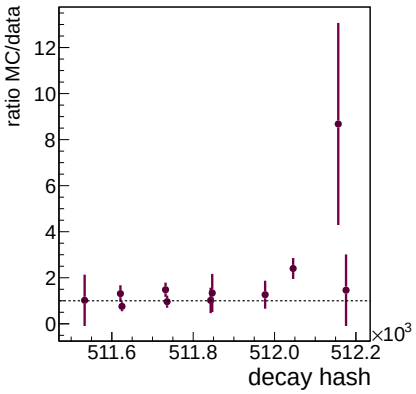
(b) K_S^0 mode



(c) $K^{*+} \to K^+\pi^0$ mode

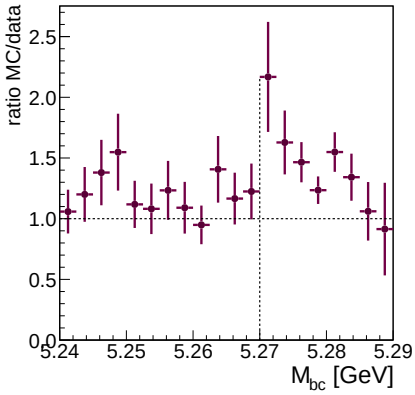


(d) $K^{*+} \to K_S^0\pi^+$ mode

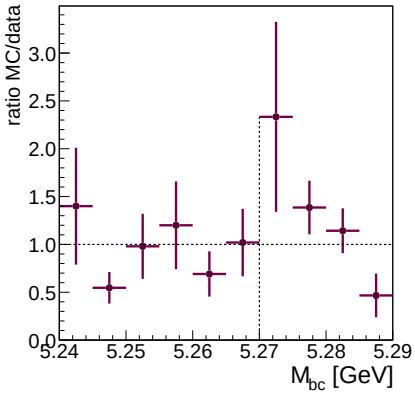


(e) $K^{*0} \to K^+\pi^-$ mode

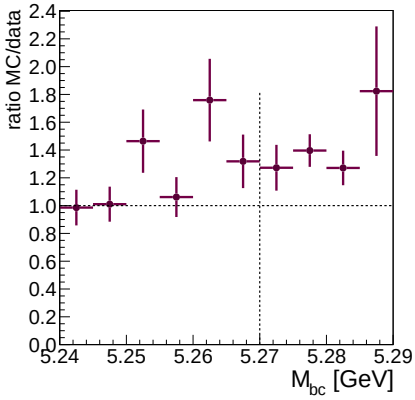
Figure 18.3. The ratio of the Monte Carlo sample divided by the data sample for the individual B_{Tag} decay channels



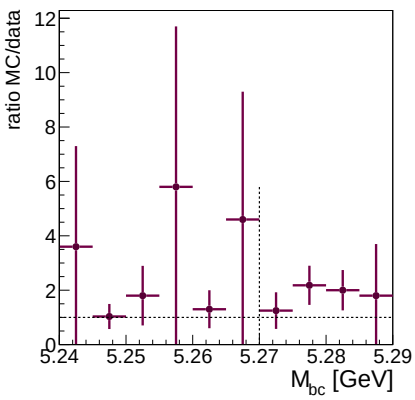
(a) K^+ mode



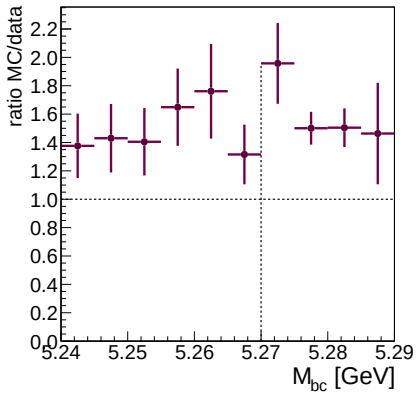
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

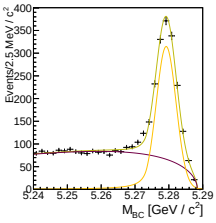


(d) $K^{*+} \to K^0_S\pi^+$ mode

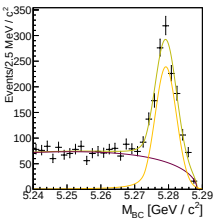


(e) $K^{*0} \to K^+\pi^-$ mode

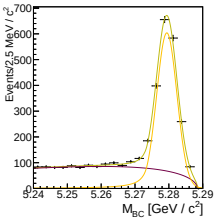
Figure 18.4. The ratio of the Monte Carlo sample divided by the data sample in the E_{ECL} sideband as a function of M_{bc}



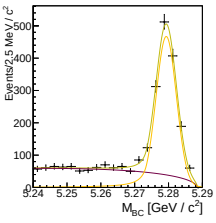
(a) K^+ mode MC



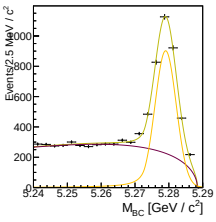
(b) K^+ mode data



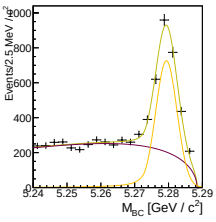
(c) K^0_S mode MC



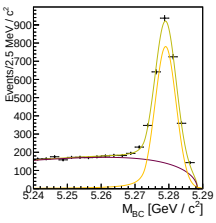
(d) K^0_S mode data



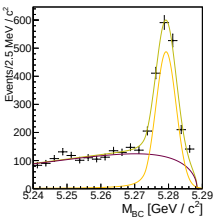
(e) $K^{*+} \to K^+\pi^0$ mode MC



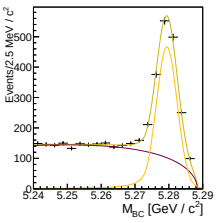
(f) $K^{*+} \to K^+\pi^0$ mode data



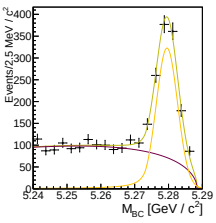
(g) $K^{*+} \to K^0_S\pi^+$ mode MC



(h) $K^{*+} \to K^0_S\pi^+$ mode data



(i) $K^{*0} \to K^+\pi^-$ mode MC



(j) $K^{*0} \to K^+\pi^-$ mode data

Figure 18.5. Fits to the B_{Tag} M_{bc} distributions for the data and Monte Carlo sample in the E_{ECL} sideband for the different modes separately. The black dots are the data points, the signal distribution is marked yellow, the background distribution is marked with a purple line and the green line is the sum of signal and background.

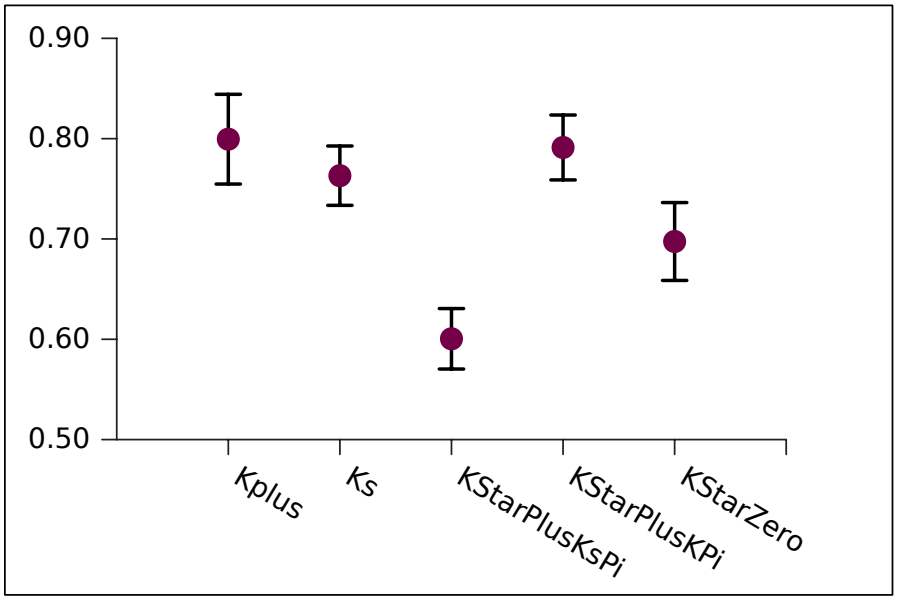


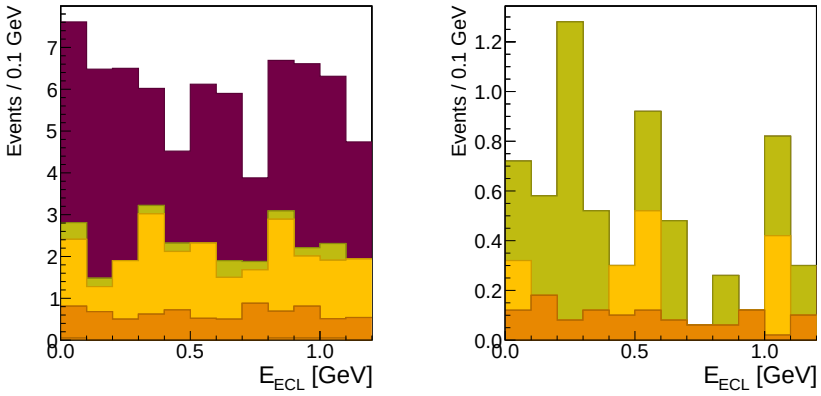
Figure 18.6. Ratios of the B_{Tag} efficiencies between data and Monte Carlo.

the neutral B modes (K^0_S,K^{*0}), the biggest contribution comes from generic neutral B decays. The contributions from the non-B Monte Carlo sample and the rare B decays Monte Carlo sample are smaller than the contributions from charmed B decays in all modes.
It is further interesting if there are dominant decays in the background samples which mimic the signal. For this, the Monte Carlo events are sorted by their real, generated decay channels. The most frequent decay channels are listed in table 18.5 and ordered by their frequency. There is no single dominant cross-feed from another decay. In general, all decays with only additional photons or neutral hadrons in the final state contribute. For example, if in the K^+ mode, the generated decay was $B \to D^0 K^+$, the K^+ is taken as the signal side. Then the D^0 most often decayed into a 0-prong final state, for example $D^0 \to \eta\pi^0$ or $D^0 \to K_L\pi^0$. Sometimes also a track was just not reconstructed, or it is used in the B_{Tag} candidate. Another example is the decay $B \to \eta_c K^+$, where the η_c goes to $K^0\bar{K}^0$ or nn. In figures 23.3e and 23.3d one can see a peak in the distribution of the rare Monte Carlo sample. This comes from decays $B \to f_2' K^*$ where the f_2' decays into two K^0_L mesons in 22% of the cases. Of course, a K_L veto would help to suppress such backgrounds, but as the K_L efficiency is known to be much worse on data (in the order of 50%) than on the Monte Carlo sample, the systematic uncertainties introduced (since the reason for the efficiency difference is yet unknown) would at least compensate the gain in background reduction.

18.4. Background Estimation and Signal Efficiency

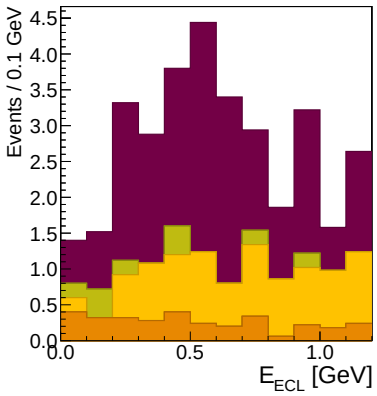
Using the signal Monte Carlo samples it is now possible to estimate the raw signal efficiencies in the signal-box by counting the number of correctly selected candidates N_{sig}

$$\epsilon_{sig} = \frac{N_{sig}}{N_{gen}} \,, \tag{18.7}$$

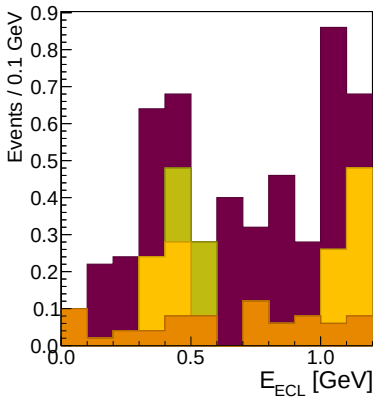


(a) K^+ mode

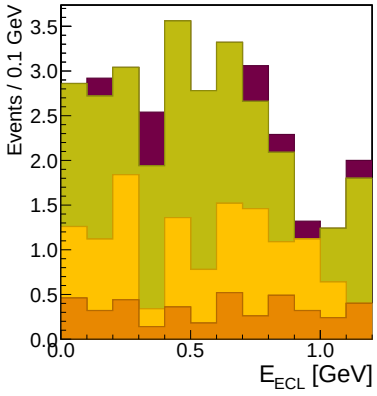
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

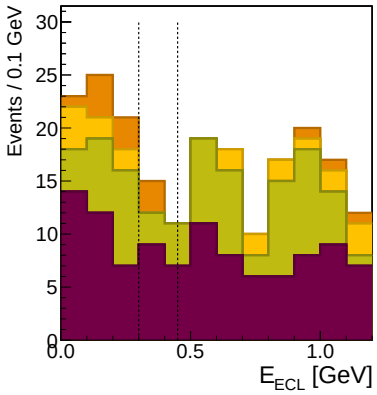


(d) $K^{*+} \to K^0_S\pi^+$ mode

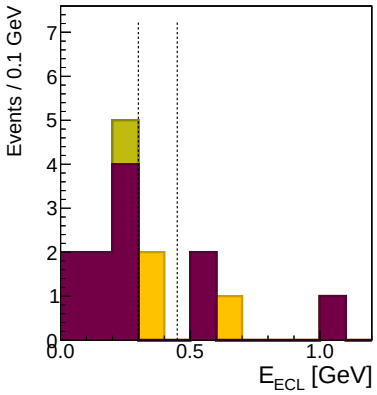


(e) $K^{*0} \to K^+\pi^-$ mode

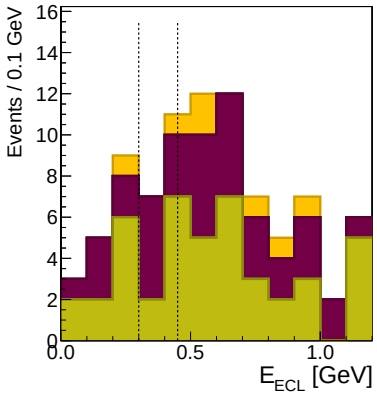
Figure 18.7. The E_{ECL} distribution of the different types of Monte Carlo generated events. The contributions from charged B decays are marked purple, from neutral B decays green, from continuum events yellow and the contributions from rare B decays are marked orange.



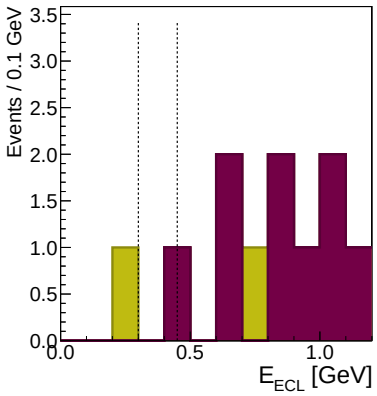
(a) K^+ mode



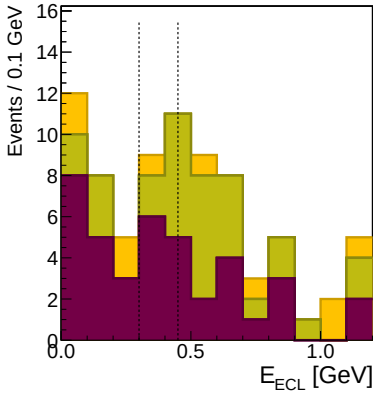
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode



(d) $K^{*+} \to K^0_S\pi^+$ mode



(e) $K^{*0} \to K^+\pi^-$ mode

Figure 18.8. The E_{ECL} distribution of the Monte Carlo generated decays for each mode. The purple histogram corresponds to the most frequent decay in table 18.5, the green one to the second most frequent, the yellow one to the third most frequent and orange to the fourth most frequent. The sample corresponds to five times the data luminosity.

mode	dominant background contributions	relative frequency
K^+	$D^0 + (e^\pm\nu, \mu^\pm\nu, K^\pm, \pi^\pm, \rho^\pm)$	42%
	$D^{*0} + (e^\pm\nu, \mu^\pm\nu, K^\pm, \pi^\pm)$	30%
	$\eta_c + K^\pm$	10%
	$J/\psi + K^\pm$	6%
K_S^0	$D^\mp + (e^\pm\nu, \mu^\pm\nu)$	41%
	$\eta_c + K_S^0$	7%
	$D^{*\mp} + (e^\pm\nu, \mu^\pm\nu)$	1%
$K^{*+} \to K_S^0\pi^+$	$D^{*0} + (e^\pm\nu, \mu^\pm\nu, K^{*+})$	48%
	$D^0 + (e^\pm\nu, \mu^\pm\nu, K^{*+})$	12%
$K^{*+} \to K^+\pi^0$	$D^0 + (e^\pm\nu, \mu^\pm\nu, \rho^\pm, K^{*\pm}, nn)$	39%
	$D^{*0} + (e^\pm\nu, \mu^\pm\nu, \rho^\pm, K^\pm K^0, nn, \tau^\pm\nu)$	32%
	$J/\psi + K^\pm$	5%
$K^{*0} \to K^+\pi^-$	$D^\mp + (e^\pm\nu, \mu^\pm\nu, \tau^\pm\nu, nn)$	30%
	$D^{*\mp} + (e^\pm\nu, \mu^\pm\nu, nn)$	22%
	$\eta_c + (K^{*0}, K_L)$	10%

Table 18.5. The Monte Carlo generated decays contributing to the background and their relative frequency for each mode.

where N_{gen} is the number of generated events. The raw efficiencies are corrected with the particle identification correction factors, obtained in section 18.5 and the correction factors for the B_{Tag} efficiencies, obtained in section 18.2. One has to be careful to take all intermediate branching fractions into account in order to obtain the correct efficiencies. The branching ratio correction factors for this analysis, the generated events and the yields are listed in table 18.6.
The uncertainty on the raw efficiency estimation is a binomial error because of the two possibilities, selected or not selected

$$\sigma_\epsilon = \frac{\sqrt{\frac{N_{sig}(N_{gen}-N_{sig})}{N_{gen}}}}{N_{gen}} \; . \tag{18.8}$$

But for the small efficiencies encountered here, the differences to just Poisson errors are tiny. The resulting efficiencies and their uncertainties are listed in table 18.6. The raw efficiencies are then

mode	ϵ_{sig} $[10^{-5}]$	stat. error	generated	branching fraction factor
K^+	72.8	$\pm$1.2%	10M	-
K_S^0	8.2	$\pm$4.2%	10M	0.692
$K^{*+}(K_S^0\pi^+)$	11.7	$\pm$6.1%	10M	$0.692 \times 0.666 \times 0.5$
$K^{*+}(K^+\pi^0)$	21.9	$\pm$3.7%	10M	0.33
$K^{*0}(K^+\pi^-)$	19.2	$\pm$2.8%	10M	0.66

Table 18.6. Signal efficiencies obtained from the signal Monte Carlo samples.

corrected by factors for the particle identification and the B_{Tag} efficiencies, listed in table 18.7.

mode	PID correction	B_{Tag} correction	ϵ_{sig} $[10^{-5}]$
K^+	0.991	0.80	57.7
K^0_S	1.000	0.76	6.3
$K^{*+}(K^0_S\pi^+)$	0.970	0.60	6.8
$K^{*+}(K^+\pi^0)$	1.003	0.79	17.4
$K^{*0}(K^+\pi^-)$	0.972	0.70	13.0

Table 18.7. Signal efficiencies corrected by the correction factor for the particle identification and the correction factors for the B_{Tag} efficiencies.

The background level in the data signal-box $E_{ECL} < 0.3\,\text{GeV}$ is estimated using the generic and rare Monte Carlo samples and the data sideband $E_{ECL} > 0.45\,\text{GeV}$. The expected number of background events is

$$N_b = N_{\text{sideband}} \cdot R_{MC} \ , \tag{18.9}$$

where N_{sideband} is the number of observed events in the sidebands of the data sample and R_{ECL} is the ratio

$$R_{MC} = \frac{N_{\text{MC,signal-box}}}{N_{\text{MC,sideband}}} \ , \tag{18.10}$$

where $N_{\text{MC,signal-box}}$ and $N_{\text{MC,sideband}}$ are the number of events from the Monte Carlo samples in the signal-box and sidebands respectively. The estimated ratios and the statistical errors are listed in table 18.8. The error is calculated as

$$\sigma_{R_{MC}} = \sqrt{\left(\frac{\sigma_{N_{\text{MC,sideband}}}}{N_{\text{MC,sideband}}}\right)^2 + \left(\frac{\sigma_{N_{\text{MC,signal-box}}}}{N_{\text{MC,signal-box}}}\right)^2} \ . \tag{18.11}$$

mode	$N_{\text{MC,signal-box}}$	$N_{\text{MC,sideband}}$	R_{ECL}	stat. error
K^+	20.54	41.92	0.49	±11.5%
K^0_S	2.58	3.22	0.80	±34.4%
$K^{*+}(K^0_S\pi^+)$	0.56	3.5	0.16	±56.1%
$K^{*+}(K^+\pi^0)$	6.24	22.32	0.28	±18.9%
$K^{*0}(K^+\pi^-)$	8.82	17.7	0.50	±17.2%

Table 18.8. The estimated signal-box to sideband ratios and the statistical errors obtained from the generic and rare Monte Carlo samples.

With equation 18.9, the observed events in the sideband N_{sideband} and the estimated ratios in table 18.8, we obtain N_b. The Poisson errors of the observed events in the sideband $\sigma_{N_{\text{sideband}}}$ has to be taken into account for the error on N_b. The observed N_{sideband} and the resulting expected number of background events in the signal region N_b and the error (only statistical error from N_{sideband}) is listed in table 18.9. For the K^0_S mode, no events were observed in the sideband. Therefore, the central value for N_b would be zero, in this case the expectation from the Monte Carlo sample is used.

mode	N_{sideband}	N_b	stat. error
K^+	31	15.19	$\pm$18.0%
K^0_S	2.58 (MC)	2.58 (MC)	-
$K^{*+}(K^0_S\pi^+)$	2	0.32	$\pm$70.7%
$K^{*+}(K^+\pi^0)$	23	6.44	$\pm$20.9%
$K^{*0}(K^+\pi^-)$	5	2.5	$\pm$44.7%

Table 18.9. The estimated number of background events N_b in the signal-box and the statistical errors obtained from the generic and rare Monte Carlo samples.

18.5. Systematic Uncertainties

There are two parts where systematic uncertainties occur. There is the uncertainty on the expected background level N_b and the uncertainty on the signal normalization

$$\mathcal{N} = N_{B\bar{B}} \cdot \epsilon_{sig} \; . \tag{18.12}$$

Both parts are needed to calculate the branching ratio limit and therefore also both sources of systematic uncertainties enter the limit calculation.

18.5.1. Systematic Uncertainties on the Signal Normalization

- **The number of $B\bar{B}$-pairs**
  There is an official estimation of the produced number of $B\bar{B}$-pairs and the error on that number:
  $$N_{B\bar{B}} = 771 \cdot 10^6 \pm 1.4\%$$

- **Track reconstruction efficiency**
  According to [Bhu] the systematic uncertainty on each single charged track is estimated to be 0.35%. Therefore the systematic uncertainty depends on the mode.

- **K^0_S reconstruction efficiency**
  According to [Whi], the overall systematic uncertainty on the reconstruction efficiency is 2.23%

- **π^0 reconstruction efficiency**
  The systematic error on the π^0 reconstruction efficiency is estimated in Belle note [Lin+] using η decays to be 4%.

- **K/π particle identification**
  The differences of the particle identification between data and Monte Carlo is estimated on an inclusive D^* sample in the unpublished Belle note [M.N+]. This study gives correction factors depending on the $cos\theta$ and the momentum of each single charged track. The correction factors and uncertainties obtained are listed in table 18.11. The correction factors are applied to the efficiency.

- **Signal reconstruction efficiency**
  The statistical error evaluated in section 18.4 is taken into account as a systematic uncertainty.

- **B_{Tag} efficiency**
  As already pointed out in section 18.2, there is a difference in the B_{Tag} efficiency between data and Monte Carlo. The statistical error on this ratio estimated by the M_{bc} fit is taken as a systematic uncertainty. The difference in efficiency is applied as a correction factor to the signal efficiency.

- **Veto efficiency**
  The difference in the charged track and π^0 veto efficiency between data and Monte Carlo is estimated using a $D^{(*)}\ell\nu$ sample. This sample has very little background and is therefore well suited for this veto efficiency check. The decay channels used for this sample are listed in table 18.9. The cuts applied are summarized in table 18.10

B Channels
$B^0 \to D^-\ell\nu$
$B^0 \to D^{*-}\ell\nu$
$B^+ \to D^0\ell\nu$
$B^+ \to D^{*0}\ell\nu$
D^* Channels
$D^{*+} \to D^0\pi^+$
$D^{*+} \to D^+\pi^0$
$D^{*0} \to D^0\pi^0$
$D^{*0} \to D^0\gamma$

D Channels
$D^+ \to K_S\pi^+$
$D^+ \to K_S\pi^+\pi^0$
$D^+ \to K_S\pi^+\pi^+\pi^-$
$D^+ \to K^-\pi^+\pi^+$
$D^+ \to K^-\pi^+\pi^+\pi^0$
$D^0 \to K^-\pi^+$
$D^0 \to K^-\pi^+\pi^0$
$D^0 \to K^-\pi^+\pi^+\pi^-$
$D^0 \to K_S\pi^0$
$D^0 \to K_S\pi^+\pi^-$

Figure 18.9. The decay channels for the $D^{(*)}\ell\nu$ sample

$dr < 2.0$ cm and $dz < 4.0$ cm for charged tracks.
$eID > 0.1$, $\mu ID > 0.1$, $\pi ID > 0.1$, $KID > 0.1$
$110\,\mathrm{MeV/c^2} < M(\pi^0) < 150\,\mathrm{MeV/c^2}$
`goodKs==1`
$E(\gamma \leftarrow \pi^0) > 50$ MeV/c^2 for all angles
$1.84\,\mathrm{GeV/c^2} < M(D^{+/0}) < 1.90\,\mathrm{GeV/c^2}$
$\lvert\vec{p^*}(D^{0/+})\rvert < 3.0$ GeV/c
$0.143\,\mathrm{GeV/c^2} < \Delta M(D^{*+/0} - D^{0/+}) < 0.148\,\mathrm{GeV/c^2}$
`BTagMbc` $> 5.27\,\mathrm{GeV/c^2}$
$-0.08\,\mathrm{GeV} <$ `BTagDE` $< 0.06\,\mathrm{GeV}$
$-0.2\,\mathrm{GeV} <$ `MissingMass`$^2 < 0.2\,\mathrm{GeV}$
$E_{ECL} < 0.3\,\mathrm{GeV}$

Table 18.10. Summary of the cuts applied for the $D^{(*)}\ell\nu$ sample.

The ratio

$$R = \frac{N_{veto}}{N_{noveto}}$$

is calculated separately for the data and the Monte Carlo sample in the E_{ECL} signal-box. The resulting ratios are

$$R_{MC} = 0.314 \pm 0.005$$

$$R_{data} = 0.335 \pm 0.004 \ .$$

The deviation from unity of the double ratio

$$\Delta = (1 - \frac{R_{data}}{R_{MC}}) = 0.0636 \pm 0.019$$

plus $1\,\sigma$ statistical uncertainty is used as an systematic error of $\pm 8.3\,\%$ on the signal efficiency. The $\mathrm{E_{ECL}}$ distributions of the $D^{(*)}\ell\nu$ for data and Monte Carlo are shown in figure 18.10 together with the distributions with the veto applied and the double ratio in bins of the $\mathrm{E_{ECL}}$ distribution. The double ratios separately for the D^* and the D samples were also checked. Both double ratios

$$R_D = 0.963 \pm 0.022$$

$$R_{D^*} = 0.915 \pm 0.027$$

agree with each other within $2\,\sigma$. That means we can use the single ratio from above.

Forthermore, the distribution of the remaining tracks is checked in the $\mathrm{E_{ECL}}$ signal-box below 0.3 GeV. As can be seen in figure 18.11 there is a slight dependency of the number of remaining tracks to the ratio of MC and data but this is alredy included in the systematic uncertainty estimated above.

All the systematic errors for the signal normalization are summarized in table 18.12.

mode	ratio(pion)	ratio(kaon)
K^+	-	$0.99 \pm 0.86\%$
$K^{*+} \to K^0_S\pi^+$	$0.97 \pm 1.19\%$	-
$K^{*+} \to K^+\pi^0$	-	$1.00 \pm 0.89\%$
$K^{*0} \to K^+\pi^-$	$0.97 \pm 1.11\%$	$1.01 \pm 0.88\%$

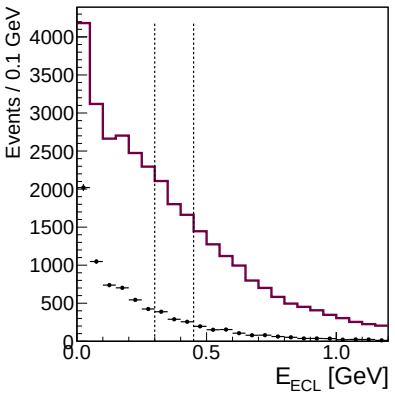
Table 18.11. Correction factors and uncertainties for the differences of the K/π particle identification between data and Monte Carlo.

18.5.2. Systematic Uncertainties on the Background Level

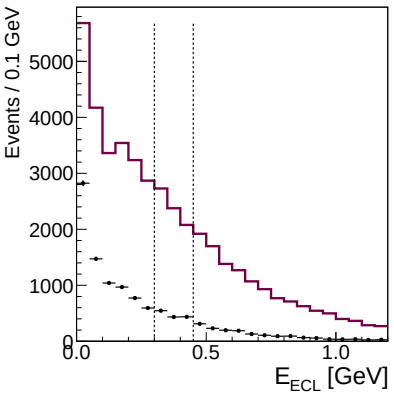
- **Statistical uncertainty of the ratio estimation**
  The statistical uncertainty ΔR_{ECL} in the estimation of the ratio R_{ECL} defined in equation 18.10 is taken as a systematic uncertainty. The error is listed in table 18.8.

- **Differences of R_{ECL} in data and Monte Carlo**
  The differences of data and Monte Carlo are evaluated using wrong-tag samples. In these samples, the $\mathrm{B_{Tag}}$ doesn't match to the signal side. The ratios R_{MC} and R_{data} are estimated for this wrong-tag sample and the mean shift plus 1σ statistical error is used as a systematic error. The cuts on $\vec{p}^*$ and $cos\theta_{mis}$ are removed in the wrong-tag sample in order to gain statistics and because they are not correlated to the $\mathrm{E_{ECL}}$ distribution. The ratios R_{MC} and R_{data} are averaged over all modes. Both ratios R_{MC} and R_{data} deviate by less then 2σ. The ratios are estimated to be

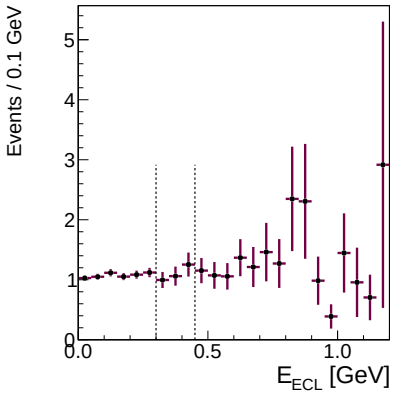
$$R_{MC} = 0.36 \pm 0.02$$



(a) data sample

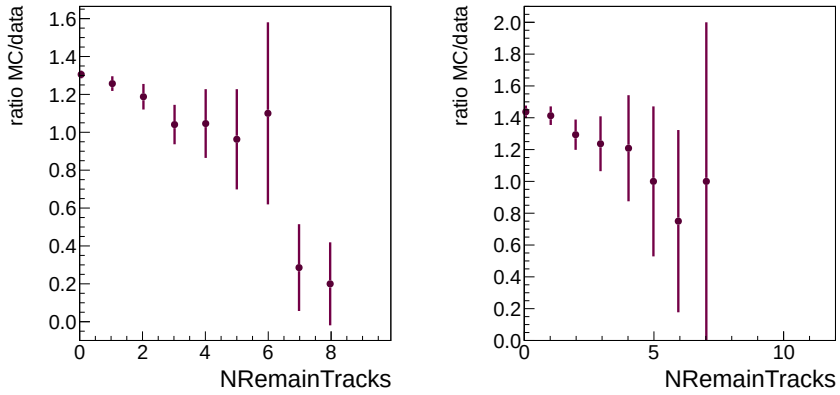


(b) Monte Carlo sample



(c) double ratio in bins of E_{ECL}

Figure 18.10. The E_{ECL} distributions of the $D^{(*)}\ell\nu$ sample. The solid line is the distribution without the vetos applied and the dots with errorbars with the vetos applied. In figure 18.10c, the double ratio in bins of the E_{ECL} distribution is shown.



(a) charged sample **(b)** neutral sample

Figure 18.11. The ratio between MC and data of the number of remaining tracks in the E_{ECL} signal-box for the charged and the neutral sample separately.

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
$N_{B\bar{B}}$	1.4	1.4	1.4	1.4	1.4
Tracks	0.35	-	0.35	0.35	0.7
K^0_S	-	2.23	2.23	-	-
π^0	-	-	-	4.0	-
K/π PID	0.86	-	0.90	0.87	1.83
signal MC statistics	1.17	4.19	6.10	3.70	2.80
B_{Tag} efficiency	5.60	3.88	5.02	4.09	5.57
veto efficiency	8.3	8.3	8.3	8.3	8.3
Sum	10.19	10.39	11.77	10.84	10.63

Table 18.12. Summary of the systematic uncertainties on the signal normalization. Each number is in units %.

$$R_{data} = 0.47 \pm 0.05 \ ,$$

and therefore 47.35 % is taken as the systematic uncertainty evoked by differences of the data and Monte Carlo samples.

- **Statistical uncertainty of the number of events in the sideband N_{sideband}**
  The statistical uncertainty of the number of events in the sideband N_{sideband} of the data sample need to be taken into account. This error is listed in table 18.9

All the systematic errors for the expected number of background events in the signal-box N_b are summarized in table 18.14.

source	K^+	K_S^0	$K^{*+}(K_S^0\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
ΔR_{ECL}(stat. only)	11.49	34.44	56.13	18.86	17.24
Data MC differences	47.35	47.35	47.35	47.35	47.35
Sum	48.72	58.55	73.43	50.96	50.39

Table 18.13. Summary of the systematic uncertainties on the ratio R_{ECL}. Each number is in units %.

source	K^+	K_S^0	$K^{*+}(K_S^0\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
central value	15.19	2.58	0.32	6.44	2.50
ΔR_{ECL}	7.40	1.51	0.23	3.28	1.26
$\Delta N_{\text{sideband}}$	2.73	0.58	0.23	1.34	1.12
Sum	7.89	1.62	0.33	3.55	1.68

Table 18.14. Summary of the absolute systematic uncertainties on the expected number of background events N_b.

18.6. Limit Estimation

As explained in section 3.1, by counting the observed numbers in the signal region $N_{\text{data,signal-box}}$ we can now calculate the limits on the branching ratio

$$\mathcal{B} = \frac{N_{\text{signal}}}{\mathcal{N}} \ , \tag{18.13}$$

where $\mathcal{N} = N_{B\bar{B}} \cdot \epsilon_{sig}$ and $N_{\text{signal}} = N_{\text{obs}} - N_{\text{b}}$, the estimated number of signal events in the data signal-box. The number of $B\bar{B}$ pairs $N_{B\bar{B}}$ is taken, because indeed there are only half of the $B\bar{B}$ pairs charged or neutral respectively, but there are 2 B mesons in one event and therefore cancels down. N_{obs}, N_{b} are the number of observed events in the data signal-box and the number of expected background events in the data signal-box respectively.
For evaluation of the sensitivity of this measurement, it is interesting to estimate the expectation value of the limit, given the hypothesis of only background. As we would expect that the number

of observed events fluctuates Poisson distributed around the expected background level N_b, we calculate the weighted mean of all limits, estimated for N_{obs} around N_b. The weights are the Poisson probabilities

$$w = Poisson(N_{obs}|N_b) \; . \tag{18.14}$$

There are many possible methods to calculate limits as already explained in section 3.1. For a comparison, five different methods were applied, profile likelihood [Mur+00], Feldman-Cousins [Fel+98], Markov-Chain Monte Carlo (MCMC) [HAS70] and the Bayesian approach [Yon+07] with a flat prior and a $1/s$ prior, i.e. the non-informative prior. In all methods, the systematic uncertainties are treated as Gaussian constraints with a width according to the systematic uncertainty and the parameters N_b and N_{signal} are fitted to N_{obs} with these constraints. The results are shown in figure 18.12. Depending on the background level and the systematic uncertainty on that, there are significant deviations between the different methods. As the Feldman-Cousins approach has

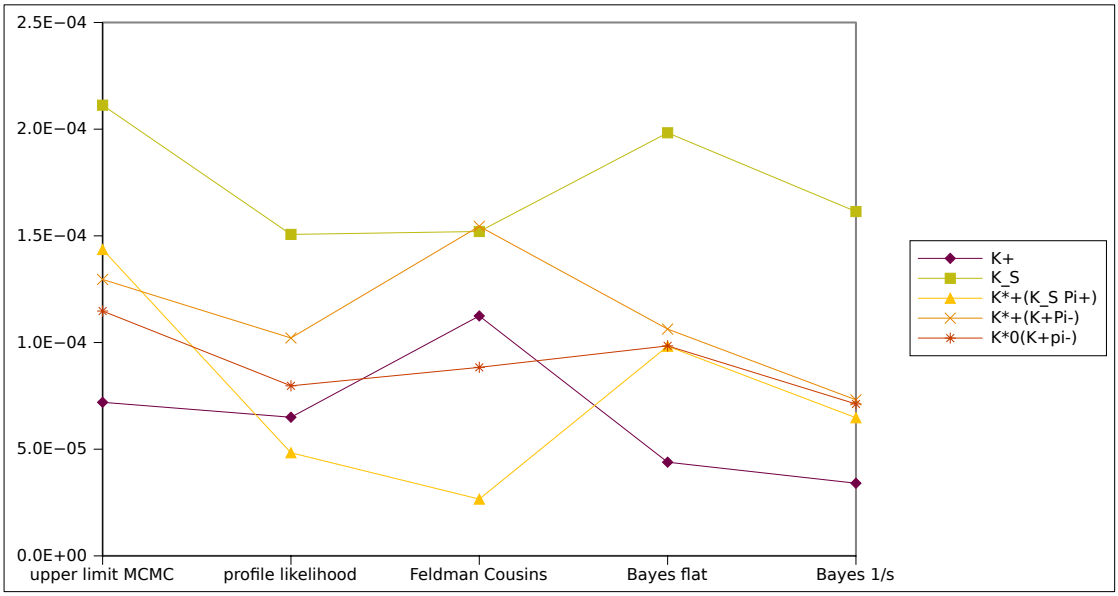


Figure 18.12. A comparison of five different methods to estimate the limit for all five modes.

several advantages like a proper coverage for all cases particularly in the transition from one-sided to a two-sided limit, this approach is taken to estimate the limit in the following. All estimated limits are limits with a two-sided 90% confidence level (as Feldman-Cousins is per construction a two-sided limit).
The resulting expected upper limits are listed in table 18.15
It is now possible to estimate the increase in sensitivity of this analysis to the predecessor analysis of Kai-Feng Chen [Che+07]. For each mode, both limits are calculated in the same way, but for the limits of the old analysis the number of $B\bar{B}$-pairs $N_{B\bar{B}} = 534.5 \cdot 10^6$ is taken. The result is shown in figure 18.13. A summary of all the numbers, relevant to estimate the limit are listed in table 18.16 for this cut-based analysis and in table 18.17 for the analysis of Kai-Feng Chen [Che+07].
The increase in sensitivity is not the same for all modes, but, as the limit highly depends on the background level, small deviations can lead to different limits. Note that this analysis comprises a completely new full reconstruction algorithm and the data is reprocessed with a new tracking

mode	exp. upper limit $[10^{-5}]$
K^+	5.1
K^0_S	9.0
$K^{*+}(K^0_S\pi^+)$	5.1
$K^{*+}(K^+\pi^0)$	7.2
$K^{*0}(K^+\pi^-)$	4.8

Table 18.15. The expected upper limts for each mode.

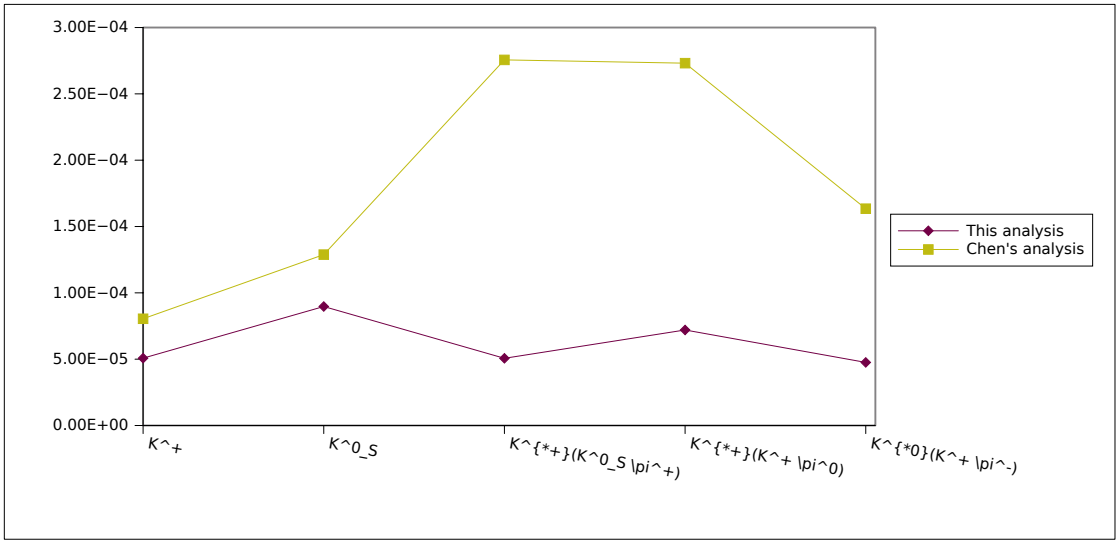


Figure 18.13. A comparison of the expected limits of this analysis and the predecessor analysis of Kai-Feng Chen [Che+07]

algorithm compared to the data of the predecessor analysis. This new tracking algorithm has a higher tracking efficiency. This analysis should be very sensitive on such an increase in track multiplicity, as there is a veto against additional tracks in an event.

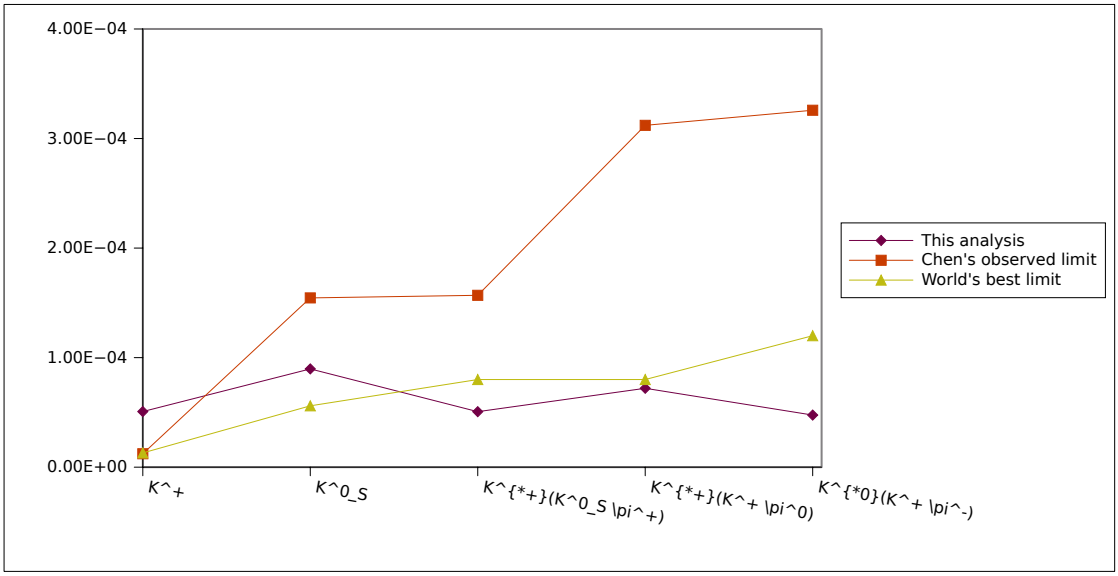


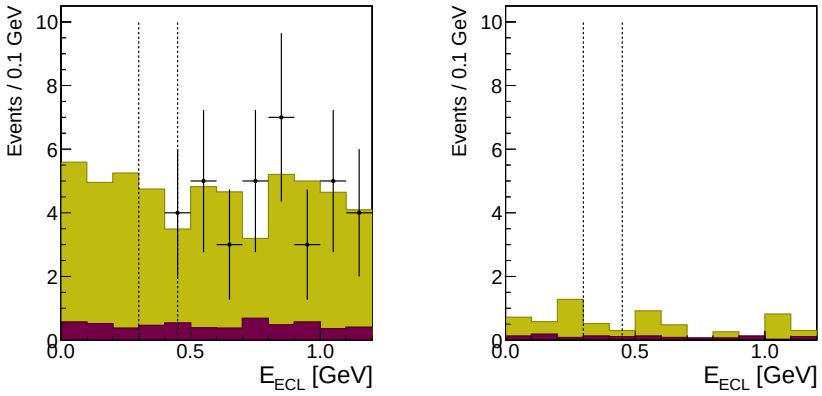
Figure 18.14. A comparison of the expected limits of this analysis, the observed limits in the predecessor analysis of Kai-Feng Chen [Che+07] and the world best limits from [Nak+10].

In figure 18.14, a comparison of the expected limits of this analysis compared to the observed limits in Chen's analysis [Che+07] and the world best limits from [Nak+10]. Except for the K^+ mode and the K^0_S mode, all of the expected limits are better than the current world best limits on the decays respectively. Of course, the actual observed limits can be quite different from the expectations. This is quite remarkable, as the world best limits are estimated by BaBar using a combined hadronic and semi-leptonic tag and the semi-leptonic tag has much more statistics.

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
N_b	15.19	2.58	0.32	6.44	2.50
$\epsilon_{sig}[10^{-5}]$	57.68	6.29	6.80	17.37	13.00
ΔN_b	7.89	1.62	0.33	3.55	1.68
$\Delta\epsilon_{sig}[\%]$	10.19	10.39	11.77	10.84	10.63

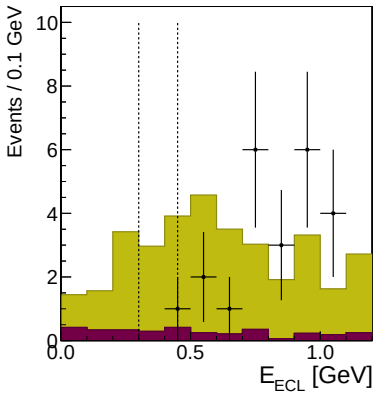
Table 18.16. Summary of the relevant information for the limit estimation for the cut-based selection.

The final blinded data E_{ECL} distributions and the scaled background expectations are shown in figure 18.15.

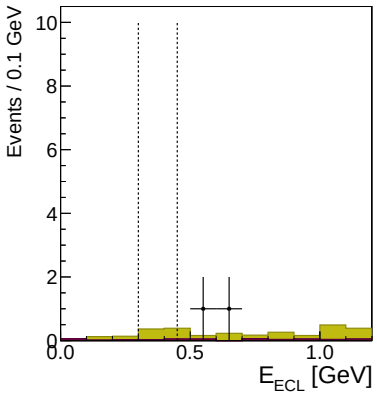


(a) K^+ mode

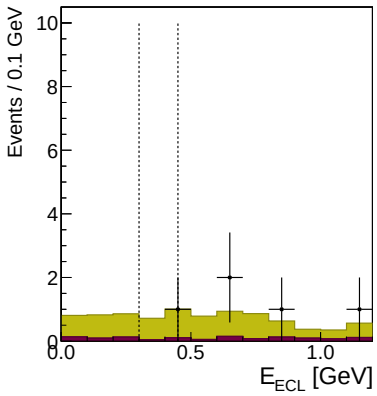
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode



(d) $K^{*+} \to K^0_S\pi^+$ mode



(e) $K^{*0} \to K^+\pi^-$ mode

Figure 18.15. The blinded E_{ECL} distribution and the Monte Carlo expectation (data: dots with error-bars, generic MC: green, rare MC: purple)

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
N_b	19.96	1.99	2.29	3.33	4.17
$\epsilon_{sig}[10^{-5}]$	26.70	4.99	2.77	3.01	5.12
ΔN_b	3.96	0.94	1.19	1.37	1.34
$\Delta\epsilon_{sig}[\%]$	10.90	6.90	12.10	12.20	5.90

Table 18.17. Summary of the relevant information for the limit estimation from the analysis of Kai-Feng Chen [Che+07].

19. Neural-Network-based Analysis

19.1. Selection Criteria

As already explained in chapter 4, a cut-based method can only reach optimal background suppression if there are no correlations among the variables. In order to increase the separation power between signal and background events, the multivariate analysis package NeuroBayes [Fei+06] is used. This package comprises a feed-forward neural network together with a series of sophisticated preprocessing algorithms.
The candidates are reconstructed in the same way as explained in 17.3. In order to reduce the most obvious background, some of the cuts of table 18.1 are applied and listed in table 19.1. The momentum of the child meson in the B_{Sig} rest frame $\vec{p}^*$ is highly correlated to the q^2 of the decay, and this distribution might be model-dependent. As NeuroBayes would learn the exact distribution it is safer to remove this observable from the training and just apply the cuts listed in table 19.1 because then the dependency on this distribution is weaker. The cut `NRemainingTracks` is also a veto against additionally charged track in the event, but, in contrast to the cut-based selection not on all tracks, but only on tracks satisfying the track selection criteria from section 17.2.

Cut (short name)	Description
`BTagChargeCombination`	B_{Tag} candidate with the corresponding charge combination (neutral-neutral, positive-negative, negative-positive)
$M_{bc} > 5.27\,\text{GeV}$	lower cut on M_{bc} of the B_{Tag} candidate
`BTagNBout` > 0.02	lower cut on the NB_{out} of the B_{Tag} candidate
`NRemainingPi0` $= 0$	no additional π^0 candidate should be left
`NRemainingTracks` $= 0$	no additional track should be left
$E_{ECL} < 2.0\,\text{GeV}$	upper cut on the extra energy in the calorimeter
$\vec{p}^* > 1.6\,\text{GeV}$	lower cut on the momentum in B_{Sig} rest frame to reduce background from charm decays
$\vec{p}^* < 2.5\,\text{GeV}$	upper cut on the momentum in B_{Sig} rest frame to reduce background from two-body B decays

Table 19.1. A list of pre-cuts for the NeuroBayes selection

19.2. The Set of Variables

The main advantage of multivariate analysis is that one is not restricted in the number of variables to use. As all the correlations among different variables are handled correctly by the NeuroBayes package one does not have to worry about this issue neither. There are only two issues one has to take care (but this also holds for the cut-based and any other method):

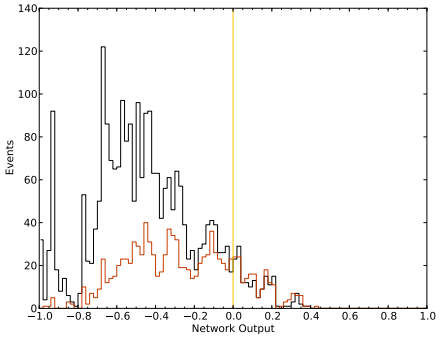
- **Correlation to the quantity to be measured**:
  In the general case one trains the classification tool with a sample where the truth is known, in most cases Monte Carlo samples. These classification tools are then optimized to separate signal from background. But later, on data, one tries to measure the amount of signal using a quantity which separates between signal and background, in this case the E_{ECL} distribution. If the classification tool can learn this distribution one will loose this separation power for the measurement. One therefore has to avoid that the tool can learn the distribution of the quantity to be measured.

- **Differences in the distributions of the variables between data and Monte Carlo**:
  While the idea of Monte Carlo simulation is to replicate the real data, there will be always differences between the simulation and real data. If there are significant differences, one can either not rely on the efficiencies and separation power obtained from the Monte Carlo samples, or one has to reduce the impact of these differences on the classification tool.

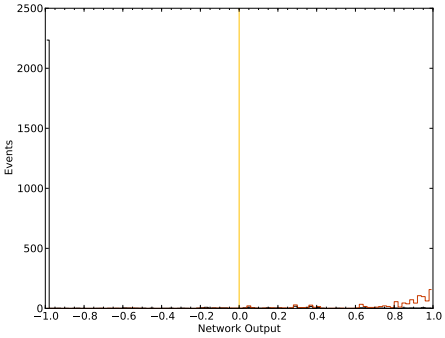
Both issues can be addressed also with a NeuroBayes training. For the correlation to the quantity to be measured one can for example train two E_{ECL} windows against each other only for background events. NeuroBayes will then calculate the significance of the correlation of each variable to the target, in this case E_{ECL} . The separation power of such a network gives a feeling how correlated the input variables are altogether to the E_{ECL} distribution. The separation of such a training is shown exemplary for the K^+ mode in figure 19.1a. It can be seen that there is a separation power within these variables, but it is so small that the training cannot learn the exact distribution of E_{ECL} . In figure 19.1b is shown the separation of the same training with one additional variable which is highly correlated to E_{ECL} . This variable is the difference between the overall energy measured in the calorimeter and the energy of all particles from B_{Tag} and B_{Sig} (`RemainingEcl`). In the distribution for lower and upper E_{ECL} window for the highest correlated variable in the weakly correlated training 19.1c and in the highly correlated training 19.1d one can clearly see that `AngleMMtoEvtMMinCMS` is only weakly correlated to E_{ECL} . The weak correlated variable set, listed in table 19.3 can safely be used for the separation of signal and background events.

The other issue of possible differences in the data and Monte Carlo sample can also be addressed with a NeuroBayes training. Possible difference can for example arise, if the momentum resolution is not correctly described by the Monte Carlo simulation. Similar to the method above we can train a signal free sideband from the data sample against the same sideband from the Monte Carlo sample. The separation of a training with the variable set listed in table 19.2 and 19.3 is shown in figure 19.2a together with distributions for the most significant variable `DistToOtherBdz` in figure 19.2b. The most significant variable was found to be `DistToOtherBdz`, the difference of the mean dz of the B_{Tag} and B_{Sig} candidate. As shown in figure 19.2c there is only a tiny difference between the data sample and the Monte Carlo sample. Another training for comparison is shown in figure 19.2b where in addition the variable `NRemainKL` is used, which is the number of remaining K_L^0 candidates in the event. This variable is known to have significant differences between data and Monte Carlo and is indeed found to be the most significant variable, as shown in figure 19.2d. The variable set used for the first training and listed in table 19.3 can safely be used for the analysis.

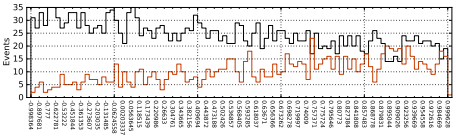
Note that both, the correlation to E_{ECL} network and the data Monte Carlo difference network are only used as checks whether there are strong dependencies. It would be possible to use NeuroBayes to also solve both problems by reweighting the Monte Carlo training samples according to the NeuroBayes outputs of the trainings explained above.



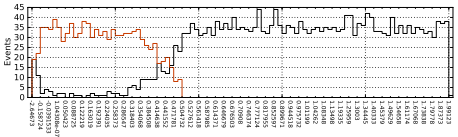
(a) The separation of lower E_{ECL} window (red) and upper E_{ECL} window (black) in the K^+ mode with weak correlation of the input variables to E_{ECL} .



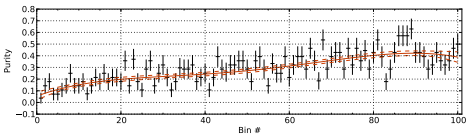
(b) The separation of lower E_{ECL} window (red) and upper E_{ECL} window (black) in the K^+ mode with strong correlation of the input variables to E_{ECL} .



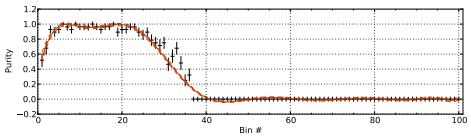
(c) The distribution of lower E_{ECL} window (red) and upper E_{ECL} window (black) for the variable with the highest correlation to E_{ECL} (`AngleMMtoEvtMMinCMS`).



(d) The distribution of lower E_{ECL} window (red) and upper E_{ECL} window (black) for the variable with the highest correlation to E_{ECL} (`RemainingEcl`).

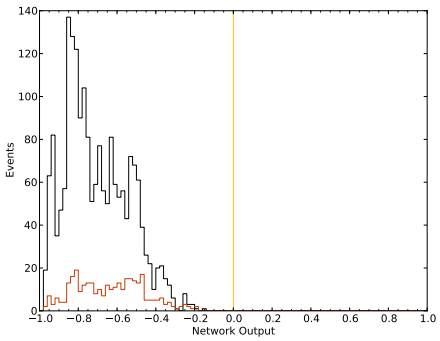


(e) The purity of lower E_{ECL} window (red) and upper E_{ECL} window (black) for the variable with the highest correlation to E_{ECL} (`AngleMMtoEvtMMinCMS`).

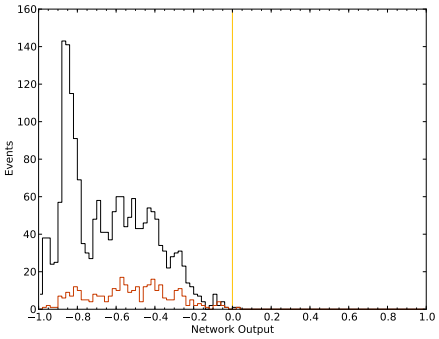


(f) The purity of lower E_{ECL} window (red) and upper E_{ECL} window (black) for the variable with the highest correlation to E_{ECL} (`RemainingEcl`).

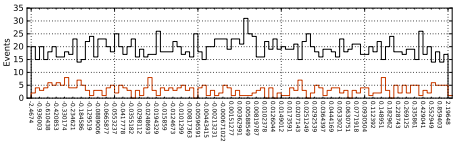
Figure 19.1. Correlation to the quantity to be measured



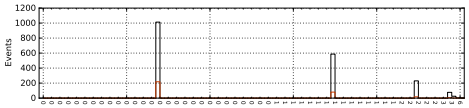
(a) The separation of data (red) and Monte Carlo (black) in the $K^{*0}(K^+\pi^-)$ mode.



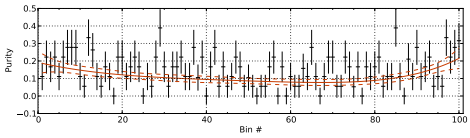
(b) The separation of data (red) and Monte Carlo (black) in the $K^{*0}(K^+\pi^-)$ mode with the additional variable `NRemainKL`.



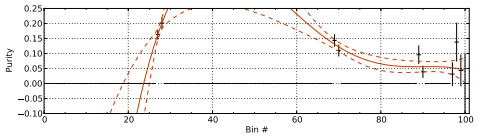
(c) The distribution of data (red) and Monte Carlo (black) for the variable with the highest significance (`DistToOtherBdz`).



(d) The distribution of data (red) and Monte Carlo (black) for the variable with the highest significance (`NRemainKL`).



(e) The purity of data (red) and Monte Carlo (black) for the variable with the highest significance (`DistToOtherBdz`).



(f) The purity of data (red) and Monte Carlo (black) for the variable with the highest significance (`NRemainKL`).

Figure 19.2. Data and Monte Carlo differences

No.	Cut (short name)	Description
1	`MCinfo`	This is the target variable, the information whether a candidate was correctly reconstructed or not.
2	ΔE	ΔE of the B_{Tag} candidate
3	`BTagNBout` > 0.02	the NB_{out} of the B_{Tag} candidate
4	`NRemainingTracksAll`	number additional tracks in the event without any qulity cuts applied on the tracks
5	`RatioToOtherType`	the ratio of this B_{Tag} candidates NB_{out} and the best of the other type if there is one
6	`RatioToSecondBest`	the ratio of this B_{Tag} candidates NB_{out} to the second best of this type if there is one
7	$cos\theta_{mis}$	the angle between the missing-momentum and the beam-pipe
8	`AngleMMtoEvtMMinCMS`	the ϕ angle angle between the missing-momentum of the signal candidate and the whole event in the center-of-mass rest frame
9	`EclEnergyInMMDir`	the total energy measured by the calorimeter in a small cone around the direction of the missing-momentum
10	`DistToOtherBdz`	the distance in z-direction between the mean |dz| of the signal candidate and the B_{Tag} candidate
11	`DistToOtherBdz_signi`	the significance of the distance in z-direction between the mean |dz| of the signal candidate and the B_{Tag} candidate

Table 19.2. A list of variables used for the NeuroBayes training for the K^+ mode.

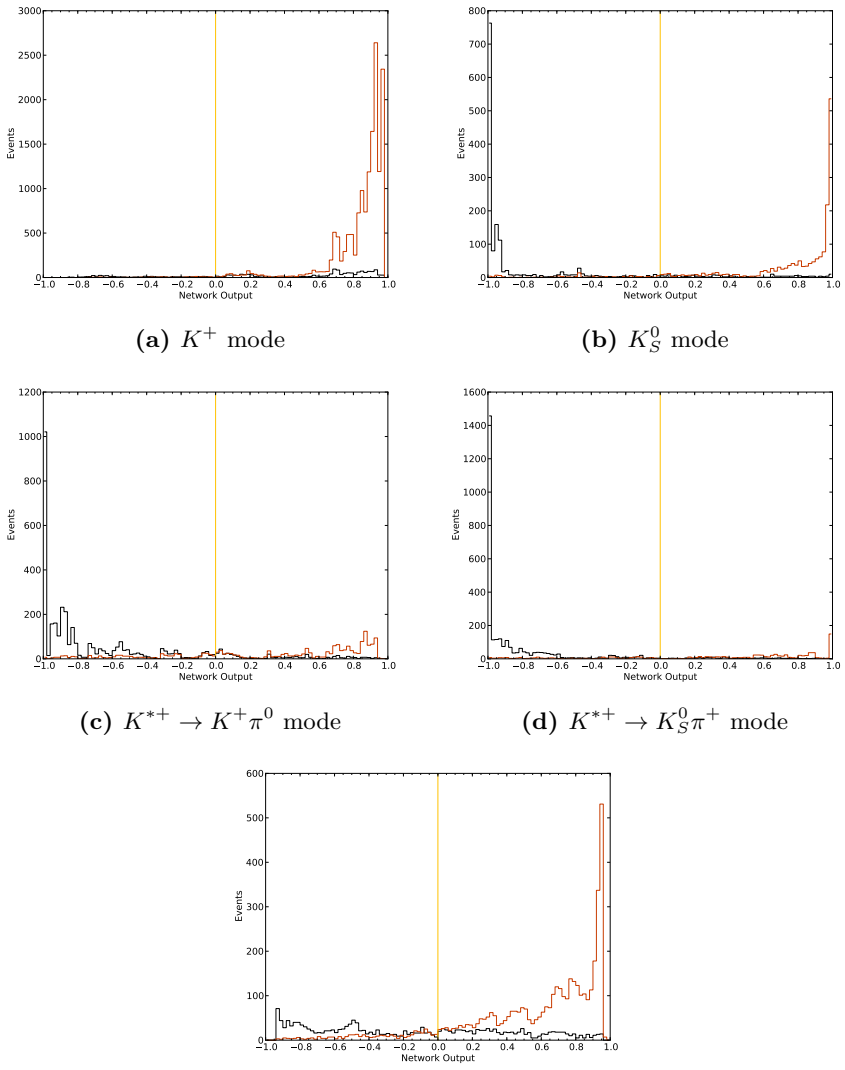
No.	Cut (short name)	Description
1	`MCinfo`	This is the target variable, the information whether a candidate was correctly reconstructed or not.
2	`ChildMass`	the invariant mass of the reconstructed child meson
3	ΔE	ΔE of the B_{Tag} candidate
4	`BTagNBout` > 0.02	the NB_{out} of the B_{Tag} candidate
5	`NRemainingTracksAll`	number of additional tracks in the event
6	`RatioToOtherType`	the ratio of this B_{Tag} candidates NB_{out} and the best of the other type if there is one
7	`RatioToSecondBest`	the ratio of this B_{Tag} candidates NB_{out} to the second best of this type if there is one
8	$cos\theta_{mis}$	the angle between the missing-momentum and the beam-pipe
9	`AngleMMtoEvtMMinCMS`	the ϕ angle between the missing-momentum of the signal candidate and the whole event in the center-of-mass rest frame
10	`EclEnergyInMMDir`	the total energy measured by the calorimeter in a small cone around the direction of the missing-momentum
11	`DistToOtherBdz`	the distance in z-direction between the mean |dz| of the signal candidate and the B_{Tag} candidate
12	`DistToOtherBdz_signi`	the significance of the distance in z-direction between the mean |dz| of the signal candidate and the B_{Tag} candidate

Table 19.3. A list of variables used for the NeuroBayes training all modes, except for the K^+ mode.

19.3. The Training Results

The set of variables listed in table 19.3 is then used to perform a training for all modes. For the background candidates, wrongly reconstructed candidates from the generic Monte Carlo sample are used. For the signal candidates, correctly reconstructed candidates from the respective signal Monte Carlo sample are used.

The resulting separations are shown in figure 19.3. The purity in bins of the NeuroBayes output is shown in figure 19.4 and a visualization of the correlation matrix of the input variables is shown in figure 19.5.



(a) K^+ mode

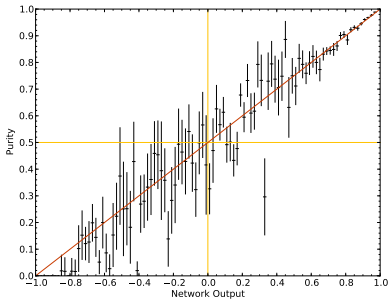
(b) K^0_S mode

(c) $K^{*+} \to K^+\pi^0$ mode

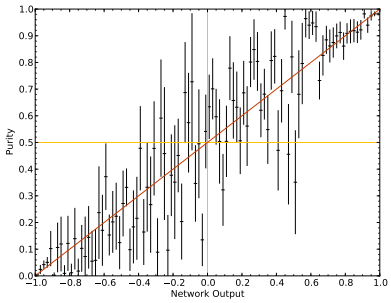
(d) $K^{*+} \to K^0_S\pi^+$ mode

(e) $K^{*0} \to K^+\pi^-$ mode

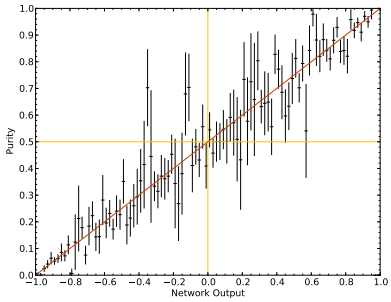
Figure 19.3. Separation of signal (red) and background (black) events for the different modes.



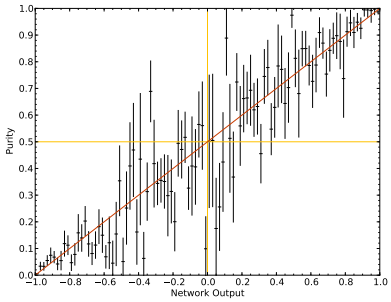
(a) K^+ mode



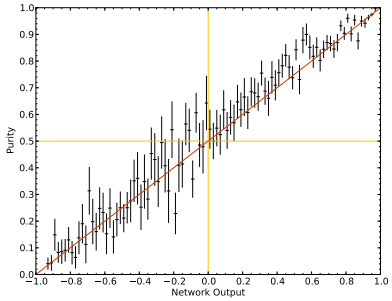
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

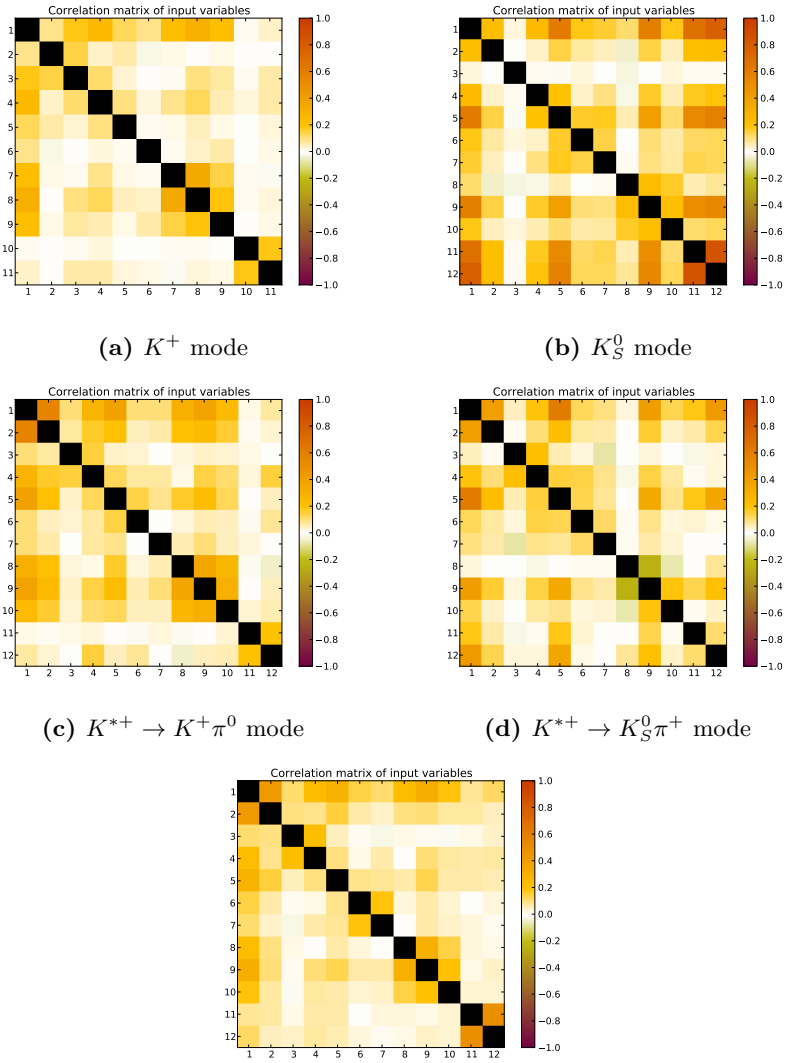


(d) $K^{*+} \to K^0_S\pi^+$ mode



(e) $K^{*0} \to K^+\pi^-$ mode

Figure 19.4. Purity as a function of NeuroBayes output.



(a) K^+ mode

(b) K^0_S mode

(c) $K^{*+} \to K^+\pi^0$ mode

(d) $K^{*+} \to K^0_S\pi^+$ mode

(e) $K^{*0} \to K^+\pi^-$ mode

Figure 19.5. Visualization of the correlation matrix. The numbers correspond to the numbers in table 19.2 and table 19.3. The first row respectively column is the target variable.

19.4. Optimization of the NeuroBayes Cut

Having the signal probabilities obtained by NeuroBayes for each event, one is free to choose any cut value, each with different purities and efficiencies. Therefore one has to choose the cut value such that the measurement has the highest sensitivity. As we are aiming for a limit on the branching fractions of these modes, it is best to optimize the cut according to the lowest expected limit on a signal free sample. This optimization is shown in figure 19.6 for all modes. The main contributions to the systematic uncertainties on the expected number of background events N_b is taken into account, as this depends on the statistics and therefore on the NeuroBayes cut. N_b is estimated without taking the data sideband, as this would bias the optimization towards lower limits. N_b is estimated using the Monte Carlo sample, scaled to the data sideband with no cut on the NeuroBayes selection applied. The steps in some of the plots are just artifacts of the fractional number of events estimated on Monte Carlo in contrast to the integer type number for the observed number. The estimated cut values are taken for the further analysis.

19.5. Background Estimation and Signal Efficiency for the NeuroBayes Selection

The background is estimated in the same way as explained in the cut-based analysis in section 18.4. First, the ratio R_{ECL} between the $\mathrm{E_{ECL}}$ signal-box and sideband is estimated using the Monte Carlo samples. The results are listed in table 19.4.

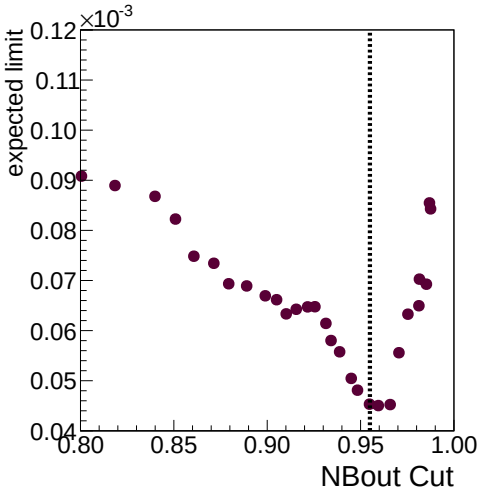
mode	$N_{\mathrm{MC,signal\text{-}box}}$	$N_{\mathrm{MC,sideband}}$	R_{ECL}	stat. error
K^+	7.62	13.52	0.56	±19.2%
K^0_S	4.10	11.78	0.35	±23.3%
$K^{*+}(K^0_S\pi^+)$	1.04	7.26	0.14	±42.2%
$K^{*+}(K^+\pi^0)$	3.08	10.82	0.28	±26.3%
$K^{*0}(K^+\pi^-)$	3.72	6.60	0.56	±26.4%

Table 19.4. The estimated signal-box to sideband ratios and the statistical errors obtained from the generic and rare Monte Carlo samples for the NeuroBayes selection.

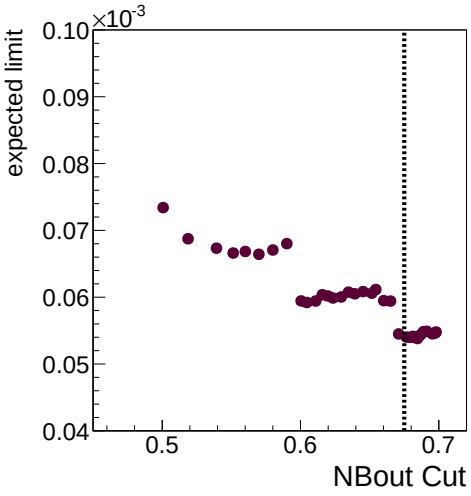
mode	N_{sideband}	N_b	stat. error
K^+	10	5.64	±31.62%
K^0_S	4.00	1.39	±50.00%
$K^{*+}(K^0_S\pi^+)$	2.00	0.29	±70.71%
$K^{*+}(K^+\pi^0)$	15.00	4.27	±25.82%
$K^{*0}(K^+\pi^-)$	4.00	2.25	±50.00%

Table 19.5. The estimated number of background events N_b in the signal-box and the statistical errors obtained from the generic and rare Monte Carlo samples for the NeuroBayes selection.

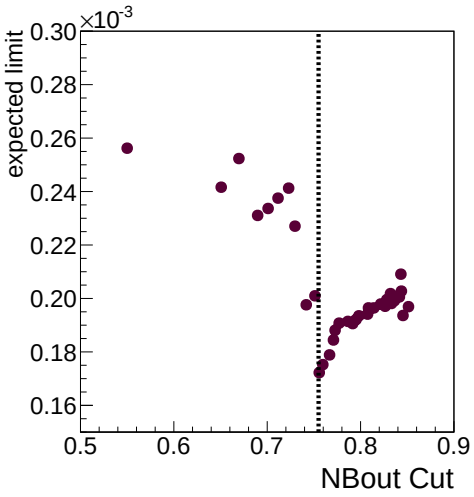
The signal efficiencies are estimated by applying the NeuroBayes cut on the signal Monte Carlo samples. The resulting raw efficiencies are listed in table 19.6. These raw efficiencies are then



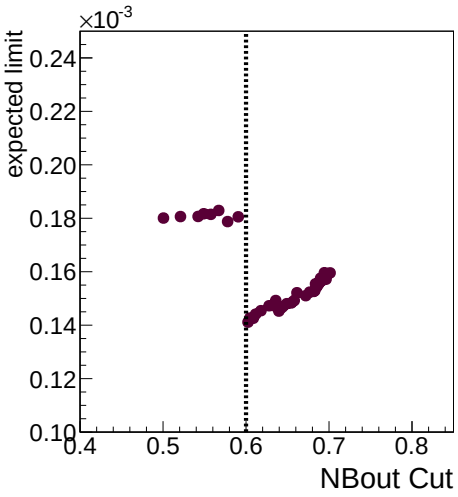
(a) K^+ mode



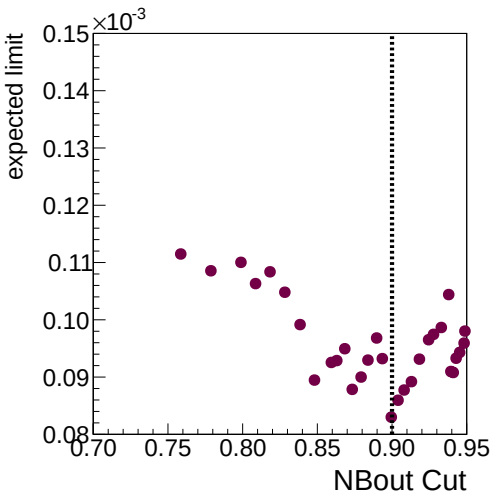
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode



(d) $K^{*+} \to K^0_S\pi^+$ mode



(e) $K^{*0} \to K^+\pi^-$ mode

Figure 19.6. The expected limit as a function of the NeuroBayes output. The purple dots are the expected limits and the dashed lines mark the optimal cut values.

corrected by the correction factor for the particle identification and the correction factors for the B_{Tag} efficiencies (see table 18.4) listed in table 19.7.

mode	ϵ_{sig} $[10^{-5}]$	stat. error	generated	branching fraction factor
K^+	53.17	$\pm 1.37\%$	10m	-
K^0_S	15.81	$\pm 3.02\%$	10m	0.692
$K^{*+}(K^0_S\pi^+)$	17.05	$\pm 5.04\%$	10m	$0.692 \times 0.666 \times 0.5$
$K^{*+}(K^+\pi^0)$	22.01	$\pm 3.69\%$	10m	0.33
$K^{*0}(K^+\pi^-)$	17.79	$\pm 2.91\%$	10m	0.66

Table 19.6. Signal efficiencies obtained from the signal Monte Carlo samples.

mode	PID correction	B_{Tag} correction	ϵ_{sig} $[10^{-5}]$
K^+	0.991	0.82	43.4
K^0_S	1.000	0.78	12.3
$K^{*+}(K^0_S\pi^+)$	0.970	0.61	10.0
$K^{*+}(K^+\pi^0)$	1.003	0.80	17.7
$K^{*0}(K^+\pi^-)$	0.972	0.71	12.3

Table 19.7. Signal efficiencies corrected by the correction factor for the particle identification and the correction factors for the B_{Tag} efficiencies.

19.6. Systematic Uncertainties for the NeuroBayes Selection

19.6.1. Systematic Uncertainties on the Signal Normalization

The systematic uncertainties for the signal normalization is carried over from the cut-based selection except for the uncertainty originating from the signal Monte Carlo statistical uncertainty. All the uncertainties are listed in table 19.8.

19.6.2. Systematic Uncertainties on the Background Level

The systematic uncertainty on the background level is obtained in the same way as for the cut-based selection. For the differences between data and Monte Carlo, also the wrong-tag sample is used. The respective NeuroBayes cuts are applied to this wrong-tag sample and then the difference of R_{MC} and R_{data} plus 1σ of the statistical uncertainty is used as a systematic uncertainty on the ratio R_{ECL}. The results from the wrong-tag sample are

$$R_{data} = 0.32 \pm 0.14 \tag{19.1}$$

$$R_{MC} = 0.36 \pm 0.04 \tag{19.2}$$

and the double-ratio

$$R_{\mathrm{MC/data}} = 0.88 \pm 0.4 \tag{19.3}$$

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
$N_{B\bar{B}}$	1.4	1.4	1.4	1.4	1.4
Tracks	0.35	-	0.35	0.35	0.7
K^0_S	-	2.23	2.23	-	-
π^0	-	-	-	4.0	-
K/π PID	0.86	-	0.90	0.87	1.83
signal MC statistics	1.37	3.02	5.04	3.69	2.91
B_{Tag} efficiency	5.60	3.88	5.02	4.09	5.57
veto efficiency	8.3	8.3	8.3	8.3	8.3
Sum	10.22	9.97	11.26	10.84	10.66

Table 19.8. Summary of the systematic uncertainties on the signal normalization for the NeuroBayes selection. Each number is in units %.

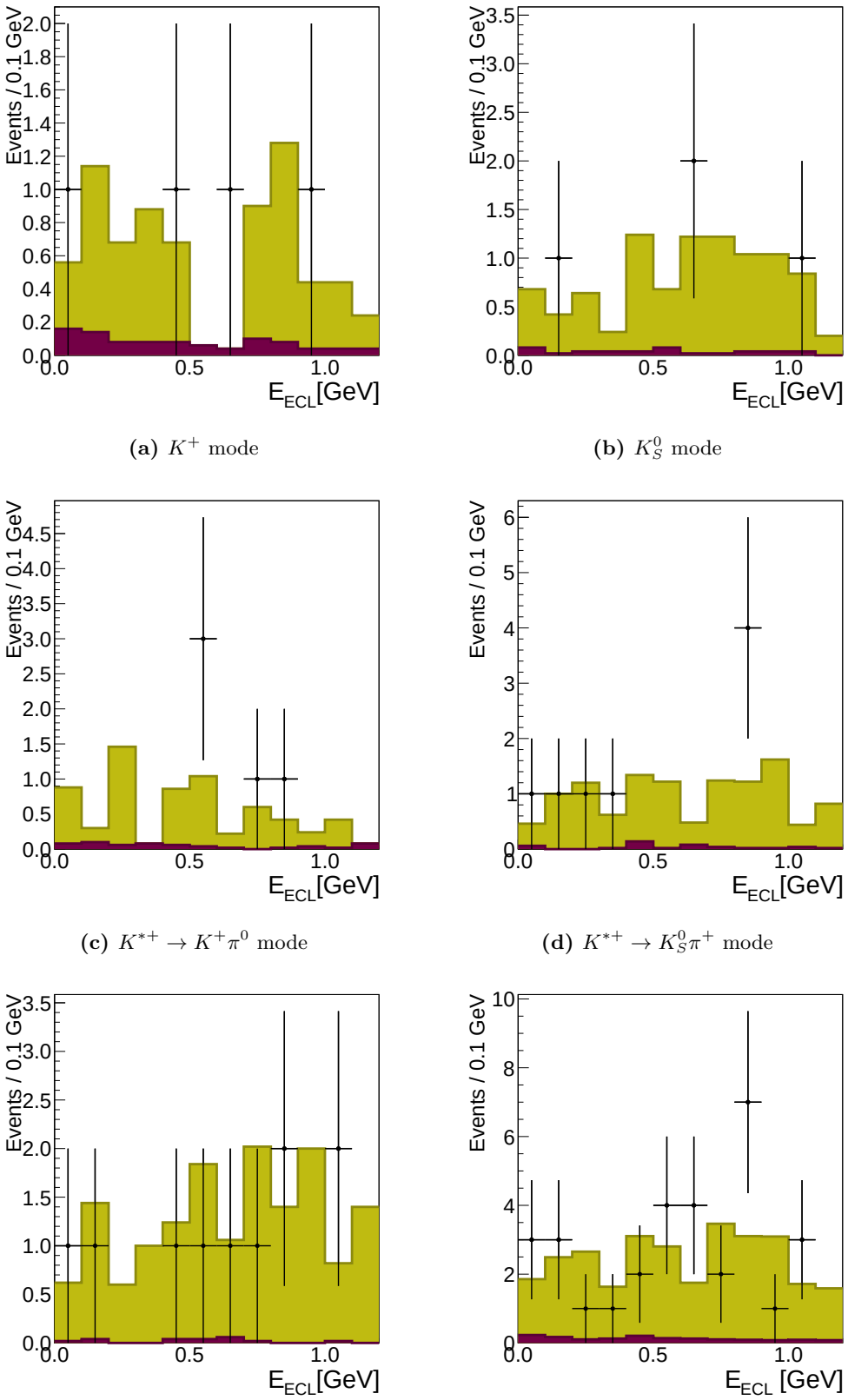
which gives an overall uncertainty of 51.66%. The distributions of the wrong-tag events for the data and the Monte Carlo sample are shown in figure 19.7. Even if the statistics is very low in some modes, it is interesting to see that there is no peaking structure in the data signal-box.
The errors on the ratio R_{ECL} are summarized in table 19.9 and the absolute uncertainties on the number of expected background events N_b are summarized in table 19.10.

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
ΔR_{ECL}(stat. only)	19.24	23.31	42.15	26.26	26.41
Data MC differences	51.66	51.66	51.66	51.66	51.66
Sum	55.13	56.68	66.67	57.95	58.02

Table 19.9. Summary of the absolute systematic uncertainties on the ratio R_{ECL}. Each number is in units %.

19.7. Limit Estimation

We can now estimate the expected limits assuming only background. This is done in the same way as for the cut-based method in section 18.6.
A summary of all the numbers, relevant to estimate the limit are listed in table 19.11.
The resulting expected upper limits are listed in table 19.12. A comparison to the expected limits obtained with the cut-based selection is visualized in figure 19.8 and with Chen's analysis in figure 19.9. There is a significant improvement for all modes with the NeuroBayes selection.
In figure 19.10, a comparison of the expected limits of this analysis with the NeuroBayes selection compared to the observed limits in Chen's analysis [Che+07] and the world best limits from [Nak+10]. Except for the K^+ mode, all of the expected limits are better than the current world best limits on the decays respectively. Of course, the actual observed limits can be different from the expectations.



(a) K^+ mode

(b) K^0_S mode

(c) $K^{*+} \to K^+\pi^0$ mode

(d) $K^{*+} \to K^0_S\pi^+$ mode

(e) $K^{*0} \to K^+\pi^-$ mode

(f) All modes together

Figure 19.7. The E_{ECL} distribution of the wrong-tag sample. The purple histogram shows the distribution of the rare Monte Carlo sample and the green histogram the generic Monte Carlo sample. The black dots with error bars are the data points.

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
central value	5.64	1.39	0.29	4.27	2.25
ΔR_{ECL}	3.11	0.79	0.19	2.47	1.31
$\Delta N_{\text{sideband}}$	1.78	0.70	0.20	1.10	1.13
Sum	3.58	3.00	0.28	2.71	1.73

Table 19.10. Summary of the absolute systematic uncertainties on the expected number of background events N_b.

source	K^+	K^0_S	$K^{*+}(K^0_S\pi^+)$	$K^{*+}(K^+\pi^0)$	$K^{*0}(K^+\pi^-)$
N_b	5.64	1.39	0.29	4.27	2.25
$\epsilon_{sig}[10^{-5}]$	42.11	12.06	9.93	17.47	12.07
ΔN_b	3.58	3.00	0.28	2.71	1.73
$\Delta\epsilon_{sig}[\%]$	10.22	9.97	11.26	10.84	10.66

Table 19.11. Summary of the relevant information for the limit estimation for the NeuroBayes selection.

mode	exp. upper limit $[10^{-5}]$
K^+	3.1
K^0_S	3.7
$K^{*+}(K^0_S\pi^+)$	3.4
$K^{*+}(K^+\pi^0)$	5.6
$K^{*0}(K^+\pi^-)$	4.9

Table 19.12. The expected upper limits for each mode.

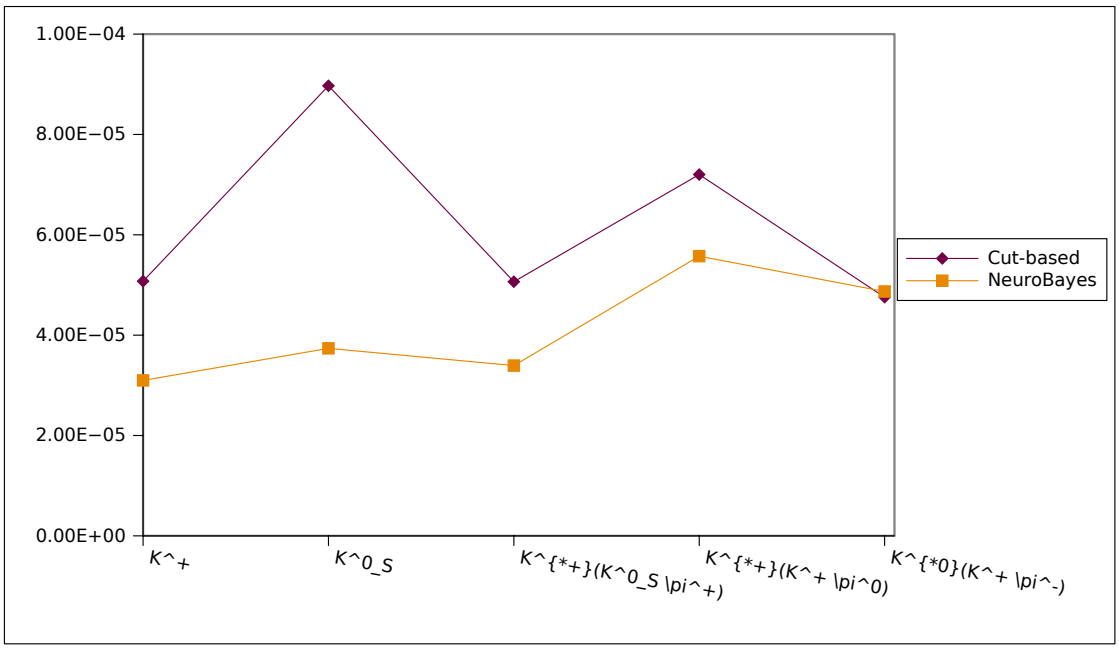


Figure 19.8. A comparison of the expected limits of the NeuroBayes selection compared to the cut-based selection.

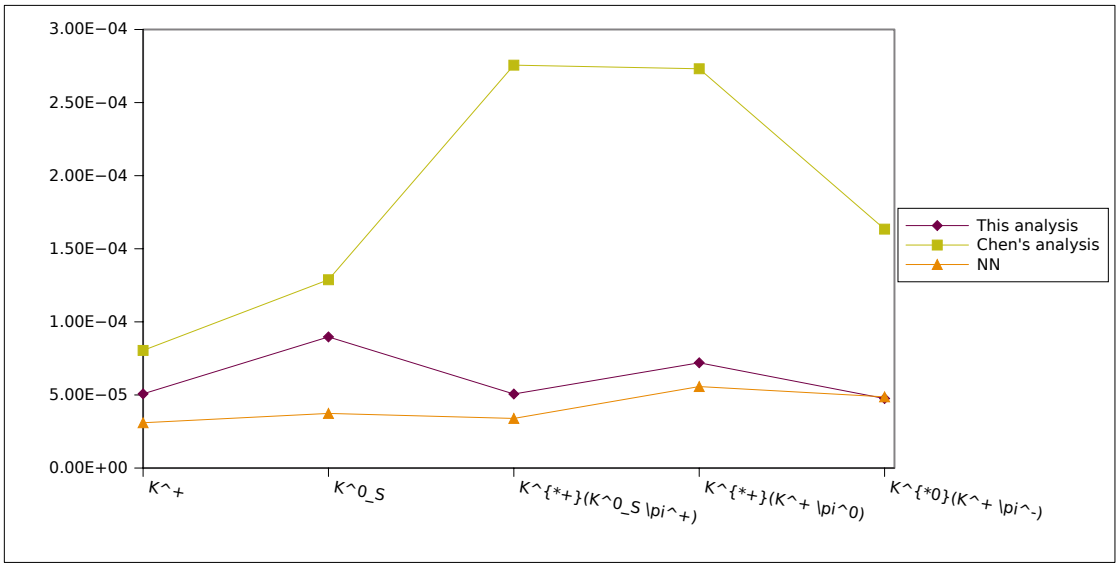


Figure 19.9. A comparison of the expected limits of the NeuroBayes selection compared to the cut-based selection and the expected limits for Chen's analysis.

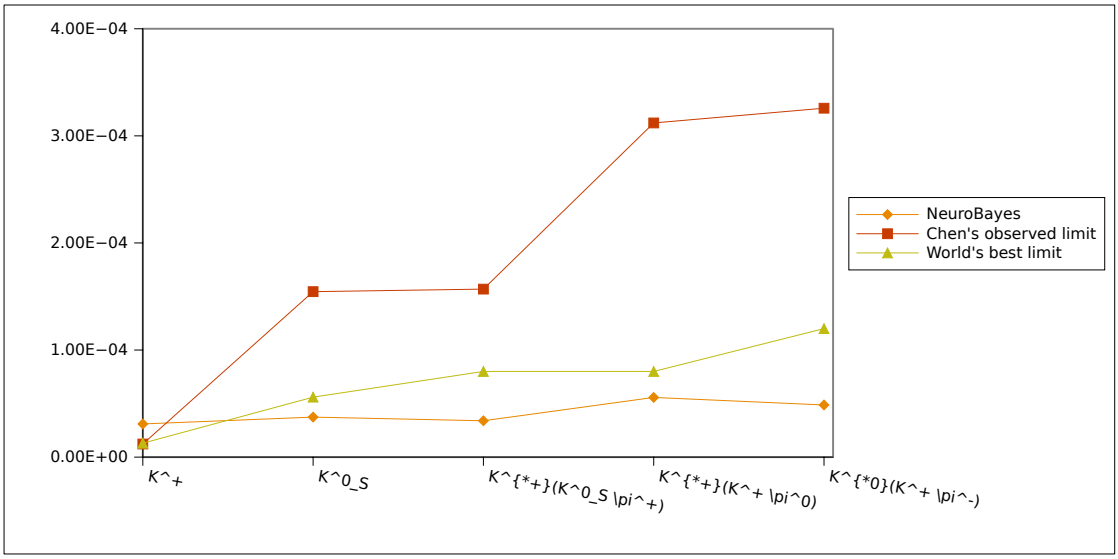
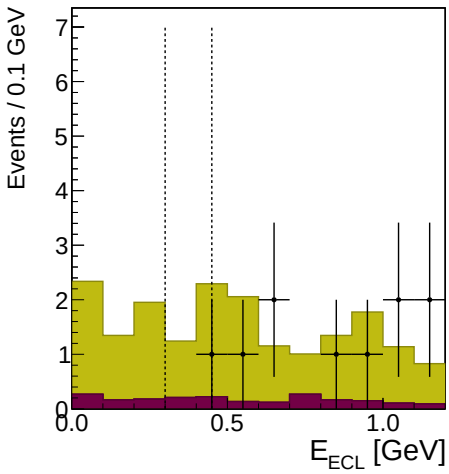
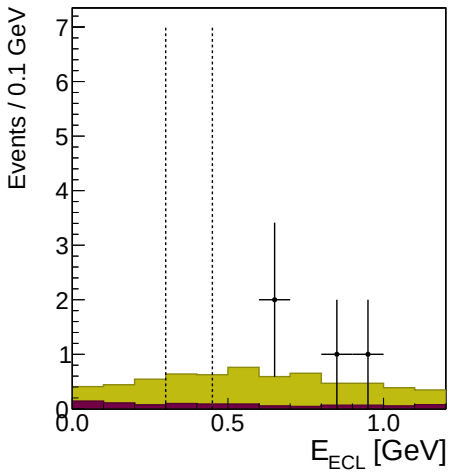


Figure 19.10. A comparison of the expected limits of this analysis with the NeuroBayes selection, the observed limits in the predecessor analysis of Kai-Feng Chen [Che+07] and the world best limits from [Nak+10].

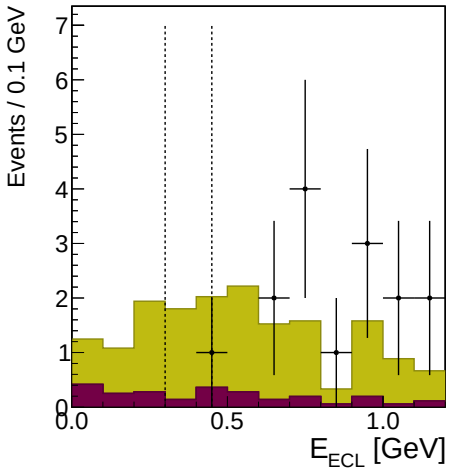
The final blinded data E_{ECL} distributions and the scaled background expectations are shown in figure 19.11.



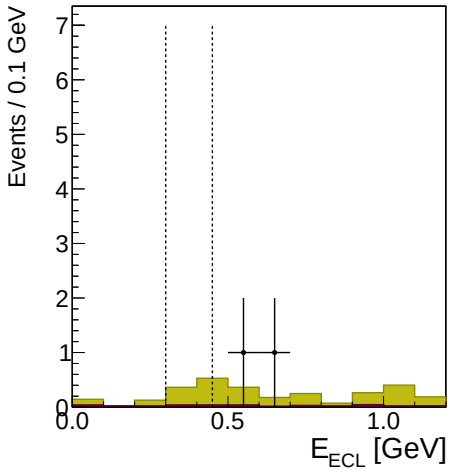
(a) K^+ mode



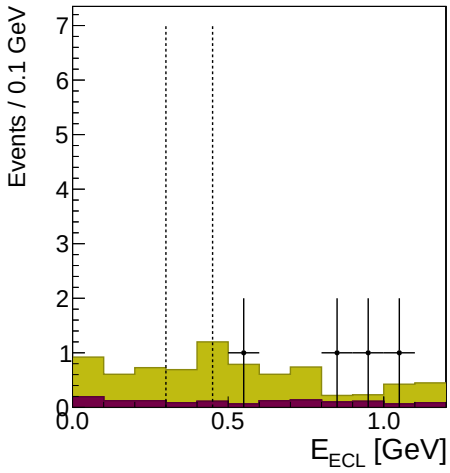
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode



(d) $K^{*+} \to K^0_S\pi^+$ mode



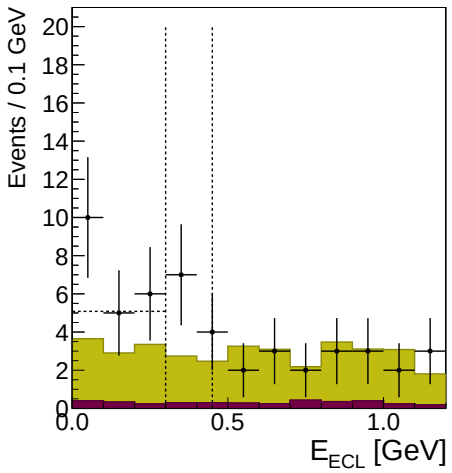
(e) $K^{*0} \to K^+\pi^-$ mode

Figure 19.11. The blinded E_{ECL} distribution and the Monte Carlo expectation (data: dots with error-bars, generic MC: green, rare MC: purple)

20. Partial Unblinding

20.1. Cut-based Selection

For the full unblinding, the whole collaboration needs to be satisfied with the analysis procedure which will take some time. However, because of the huge expected improvements and the sounding analysis, I was allowed to do a partial unblinding of the K^+ mode with the data used in Chen's analysis corresponding to 534.586 million $B\bar{B}$ pairs.
The resulting E_{ECL} distributions are shown in figure 20.1. The observed and expected numbers together with the estimated upper limit is shown in table 20.1. In this table, also the 90% lower limit is shown. If it is 0 then the observed number of signal events has less than $2\,\sigma$ evidence.



(a) K^+ mode

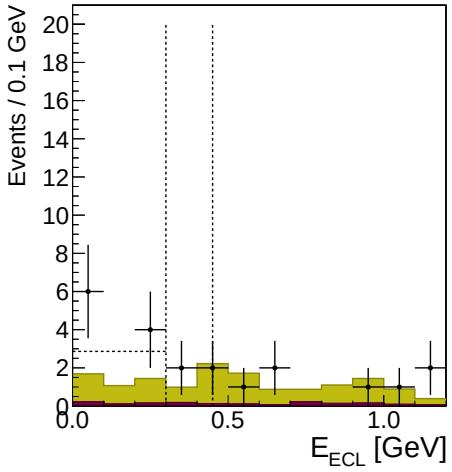
Figure 20.1. The unblinded E_{ECL} distribution and the scaled Monte Carlo expectation (data: dots with error-bars, generic MC: green, rare MC: purple) for the data sample used in the predecessor analysis. The expected background plus $1\,\sigma$ systematic uncertainty is shown as a horizontal dashed line.

20.2. NeuroBayes Selection

Also the NeuroBayes selection is partially ublinded. The resulting E_{ECL} distributions are shown in figure 20.2. The observed and expected numbers together with the estimated upper limit is shown in table 20.2.
The observed limits in both selections are weaker than the expected limits, because there are more events observed as expected with the background only hypothesis. This is further discussed

source	K^+
N_b	9.91
$\epsilon_{sig}[10^{-5}]$	57.68
ΔN_b	5.34
$\Delta\epsilon_{sig}[\%]$	10.19
Obs. events	21
Obs. limit$[10^{-5}]$	8.79
Obs. lower limit$[10^{-5}]$	0

Table 20.1. Summary of the relevant information for the limit estimation for the cut-based selection for the partial data-sample.



(a) K^+ mode

Figure 20.2. The unblinded E_{ECL} distribution and the scaled Monte Carlo expectation (data: dots with error-bars, generic MC: green, rare MC: purple) for the data sample used in the predecessor analysis. The expected background plus $1\,\sigma$ systematic uncertainty is shown as a horizontal dashed line.

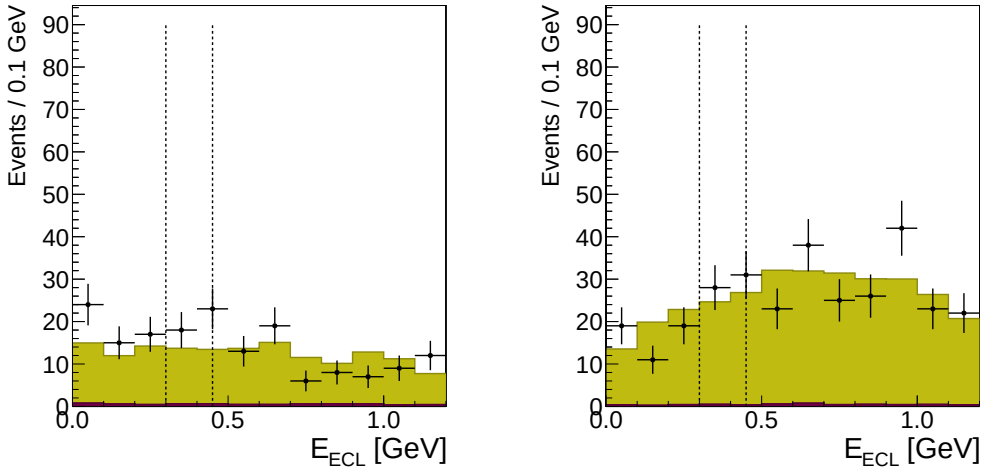
source	K^+
N_b	5.25
$\epsilon_{sig}[10^{-5}]$	42.11
ΔN_b	3.34
$\Delta\epsilon_{sig}[\%]$	10.22
Obs. events	10
Obs. limit$[10^{-5}]$	6.9
Obs. lower limit$[10^{-5}]$	0

Table 20.2. Summary of the relevant information for the limit estimation for the NeuroBayes selection for the partial data-sample.

in the next section.

20.3. Further Checks

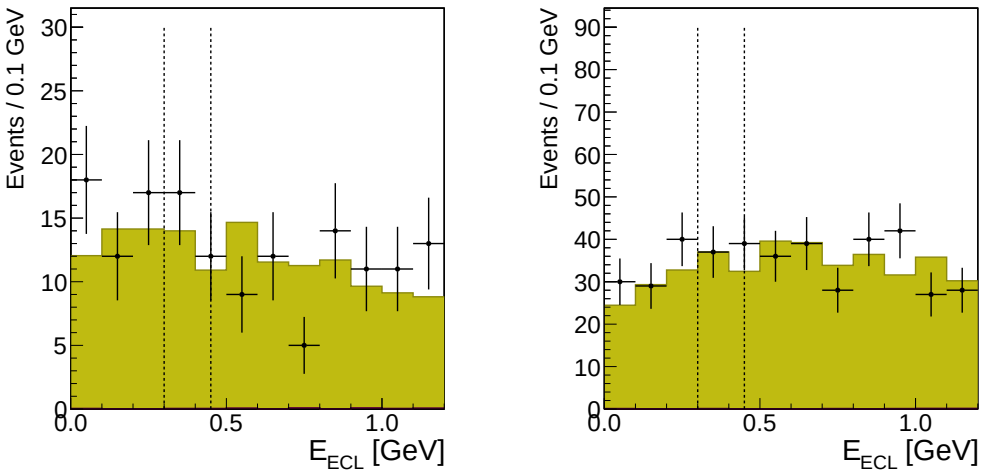
As can be seen in figure 20.1a and figure 20.2a there is a non-significant excess of data events in the signal-box. This needs extremely careful further examination. It is of course possible that it is a upward fluctuation of the background. The shape seems not to be described by the expected signal-shape, shown in figure 18.1a. The most dangerous case would be, if it is a bias in the data E_{ECL} distribution of the background. This could be either a peaking background which is not described by the Monte Carlo, or the E_{ECL} distribution itself is not well described by the Monte Carlo. This examination needs to be done before the box opening of the complete data sample, as it is a blind analysis.
In figures 20.3 one can see a comparison of the E_{ECL} distribution for the veto against all tracks and the anti-veto with only events where at least one track remains for the cut-based selection in the K^+ mode. For both distributions, no good track, passing the quality criteria is allowed and the lower cut on $\vec{p}^*$ is removed.



(a) track veto **(b)** track anti-veto

Figure 20.3. The unblinded E_{ECL} distribution for the veto and the anti-veto for data and the scaled Monte Carlo expectation (data: dots with error-bars, generic MC: green, rare MC: purple) for the data sample used in the predecessor analysis.

One can see that the distribution in figure 20.3a has more data events in the signal-box as expected from Monte Carlo and in the anti-veto sample it is the other way round, having too few events in the signal-box. That might be a hint that the veto against all tracks with no quality cuts applied might be the reason for the excess. A discrepancy in this variable was already seen in figure 18.11.
In addition, the E_{ECL} distribution for data and Monte Carlo in the M_{bc} sideband can be seen in figure 20.4a once with the veto on all tracks applied and once with a veto only against good tracks, matching the quality criteria. Both distributions seems to be fine, which would point to a background component from real B meson decays.



(a) K^+ mode with veto

(b) K^+ mode without veto

Figure 20.4. The unblinded E_{ECL} distribution of the M_{bc} sideband and the scaled Monte Carlo expectation (data: dots with error-bars, generic MC: green, rare MC: purple) for the data sample used in the predecessor analysis.

21. Further Improvements

The currently most stringent limits on the branching ratios of $B \to K^{(*)}\nu\bar{\nu}$ decays are set by the BaBar collaboration. In their analysis [Aub+08] and [AS+10] they not only used a hadronic full reconstruction as done in this analysis, but they also included a semi-leptonic full reconstruction tag. While the background level is expected to rise, the efficiency of the full reconstruction tag will also be significantly increased. This semi-leptonic tag is currently under construction and, with regards to the success of the BaBar results, a significant improvement can be expected.
Furthermore, as shown in section 18.3, an additional veto on neutral hadrons (e.g. K_L mesons) will help to reduce the expected background further and therefore, the reconstruction efficiency can be increased.
Currently, the signal shape is not used beside the division into a signal-free sideband and a signal region. It could help to fit the signal and background distributions in oder to separate signal from background further.

Part VI.

Conclusion

Conclusion

After ten years of extremely successful running of the Belle detector we achieved an improvement of a factor of two in efficiency of the full reconstruction technique used by the collaboration for many interesting analyses. The fundamental innovation was the turning away from cut-based binary decisions towards probability-based decisions. Exploiting the hierarchical structure, where heavy particles decay into lighter particles, we propagate the signal to background probability for each candidate upwards the hierarchy from lighter to heavier particles, postponing the decisions of keeping or discarding the candidate to higher stages. Of course, keeping all reconstructed candidates would result in a maximal efficiency, but this would result in so many possible combinations that the processing of the whole dataset would take several years. Therefore, we had to reduce the number of combinations by reducing the intermediate particle candidates. As already mentioned, we cut away candidates in higher stages, more detailed, we cut softly on the product of the children's probability. This means that a child with a very high probability can compensate a child with a lower probability. This results in a much higher efficiency compared to cut-away the children based on their own probability alone.
We could show that the multivariate analysis package NeuroBayes is able to estimate the proper Bayesian probability for each candidate which is crucial for our approach. In order to be able to train and deploy hundreds of NeuroBayes experts, with each of them using different variables, we developed a highly flexible reconstruction framework. By full automatization of the reconstruction procedure the human error rate is minimized as well as the time consumption for the implementation of new decay channels to be reconstructed. In the end we reconstruct 1104 exclusive B meson decay channels.
Altogether, as we are able to freely choose the total number of combinations, we found a good balance between highest possible efficiency and acceptable processing time for the whole data sample. As already confirmed by many analyses in the Belle collaboration, the efficiency is doubled compared to the previously used cut-based full reconstruction tool. This efficiency increase is comparable to several years of data taking and will improve the sensitivity of many future analyses.
In order to take advantage of this remarkable improvement I updated a previous measurement by K.F. Chen for the Belle collaboration, the search for the decays $B \to K^{(*)}\nu\bar{\nu}$. By redoing this analysis, a significant improvement of the expected limit showed up due to the higher efficiency of the full reconstruction tool. Depending on the mode, the expected limit could be lowered by a factor of 2 to 5 compared to the predecessor analysis (see figure 18.13).
By exchanging the cut-based selection of this analysis with a NeuroBayes-based selection, another significant improvement could be achieved because the efficiency and the purity increased at the same time using the multivariate technology. Another reduction by up to a factor of 2 of the expected limit could be achieved compared to the cut-based selection (see figure 19.8).
These amazing improvements reveal the importance of the efficiency enhancement of our new full reconstruction method for the sensitivity of future analyses, including the upcoming unblinding of the analysis presented in this thesis.

Part VII.

Appendix

22. Appendix for the Full Reconstruction

22.1. Variables Used for the NeuroBayes Trainings

22.1.1. Stage 1 Trainings

Variable name	Description
`energy`	measured energy
`pt`	measured p_t
`ptot`	measured p_tot
`dEdx_ratio`	measured dE/dx ratio
`dEdx_pull`	measured dE/dx pull
`trk_pid`	particle identification based on track measurements
`tof_pid`	particle identification based on time-of-flight measurements
`acc_pid`	particle identification based on aerogel-Cherenkov measurements
`PID_eid`	electron identification
`PID_eid_flag`	electron identification
`PID_muid`	muon identification
`PID_muid_flag`	muon identification
`ATC_PID_binned_01`	combined particle identification binned for different hypothesis
`ATC_PID_binned_02`	combined particle identification binned for different hypothesis
`ATC_PID_binned_03`	combined particle identification binned for different hypothesis
`ATC_PID_binned_04`	combined particle identification binned for different hypothesis

Table 22.1. A list of variables used for the NeuroBayes training for electrons

Variable name	Description
energy	measured energy
pt	measured p_t
ptot	measured p_tot
dEdx_ratio	measured dE/dx ratio
dEdx_pull	measured dE/dx pull
trk_pid	particle identification based on track measurements
tof_pid	particle identification based on time-of-flight measurements
acc_pid	particle identification based on aerogel-Cherenkov measurements
PID_eid	electron identification
PID_eid_flag	electron identification
PID_muid	muon identification
PID_muid_flag	muon identification
ATC_PID_binned_10	combined particle identification binned for different hypothesis
ATC_PID_binned_11	combined particle identification binned for different hypothesis
ATC_PID_binned_12	combined particle identification binned for different hypothesis
ATC_PID_binned_13	combined particle identification binned for different hypothesis
KLM_likelihood1	information about the muon likelihood from the KLM
KLM_likelihood2	information about the muon likelihood from the KLM
KLM_likelihood3	information about the muon likelihood from the KLM
KLM_likelihood4	information about the muon likelihood from the KLM
KLM_likelihood5	information about the muon likelihood from the KLM
KLM_Chi2	χ^2 for the muon hypothesis from the KLM
KLM_Outcome	information about the muon from the KLM

Table 22.2. A list of variables used for the NeuroBayes training for muons

Variable name	Description
energy	measured energy
pt	measured p_t
ptot	measured p_tot
dEdx_ratio	measured dE/dx ratio
dEdx_pull	measured dE/dx pull
trk_pid	particle identification based on track measurements
tof_pid	particle identification based on time-of-flight measurements
acc_pid	particle identification based on aerogel-Cherenkov measurements
PID_eid	electron identification
PID_eid_flag	electron identification
PID_muid	muon identification
PID_muid_flag	muon identification
ATC_PID_binned_20	combined particle identification binned for different hypothesis
ATC_PID_binned_21	combined particle identification binned for different hypothesis
ATC_PID_binned_23	combined particle identification binned for different hypothesis
ATC_PID_binned_24	combined particle identification binned for different hypothesis

Table 22.3. A list of variables used for the NeuroBayes training for pions

Variable name	Description
energy	measured energy
pt	measured p_t
ptot	measured p_tot
dEdx_ratio	measured dE/dx ratio
dEdx_pull	measured dE/dx pull
trk_pid	particle identification based on track measurements
tof_pid	particle identification based on time-of-flight measurements
acc_pid	particle identification based on aerogel-Cherenkov measurements
PID_eid	electron identification
PID_eid_flag	electron identification
PID_muid	muon identification
PID_muid_flag	muon identification
ATC_PID_binned_30	combined particle identification binned for different hypothesis
ATC_PID_binned_31	combined particle identification binned for different hypothesis
ATC_PID_binned_32	combined particle identification binned for different hypothesis
ATC_PID_binned_34	combined particle identification binned for different hypothesis

Table 22.4. A list of variables used for the NeuroBayes training for kaons

Variable name	Description
energy	measured energy
pt	measured p_t
ptot	measured p_tot
Pi0Mass	invariant mass
Pi0GammaMatch	booleaen whether the gammas are matched with tracks
Pi0ChildrenAngle	angle between gammas
Pi00ECL_en	energy gamma 1
Pi01ECL_en	energy gamma 2
Pi00EA_e9	gamma 1 shower shape
Pi01EA_e9	gamma 2 shower shape
Pi00EA_e9unf	gamma 1 shower shape
Pi01EA_e9unf	gamma 2 shower shape
Pi00EA_width	gamma 1 shower shape
Pi01EA_width	gamma 2 shower shape
Pi00EA_nhits	gamma 1 shower shape
Pi01EA_nhits	gamma 2 shower shape
Pi00EA_seed	gamma 1 initial christal
Pi01EA_seed	gamma 2 initial christal
Pi00EA_cID	gamma 1 christal ID
Pi01EA_cID	gamma 2 christal ID

Table 22.5. A list of variables used for the NeuroBayes training for π^0

Variable name	Description
Mass	invariant mass
energy	measured energy
px	momentum in x direction
py	momentum in y direction
pz	momentum in z direction
pt	transversal momentum
ptot	total momentum
Z_dist	difference of the impact parameters in z direction
Intersect	tracks intersect
Chi2	χ^2 of the fit
dr	radius of the nearest approach to the IP
dPhi	phi of the nearest approach to the IP

Table 22.6. A list of variables used for the NeuroBayes training for K_S^0

Variable name	Description
pt	transversal momentum
ptot	total momentum
EA_e9	shower shape
EA_e9unf	shower shape
EA_width	shower shape
EA_nhits	shower shape
EA_seed	initial christal
EA_cID	christal ID
ECL_en	Energy in the ECL

Table 22.7. A list of variables used for the NeuroBayes training for gamma

22.1.2. Stage 2 Trainings

Variable name	Description
pt	transversal momentum
ptot	total momentum
Ch0_NBout	NeuroBayes output of child 0
Ch0_PseudoHelAng	pseudo helicity angle of child 0
Ch0_pt	transversal momentum of child 0
Ch0_ptot	total momentum of child 0
Ch01_Angle	angle between of children
Ch1_NBout	NeuroBayes output of child 1
Ch1_energy	energy of child 1
Ch1_PseudoHelAng	pseudo helicity angle of child 1
Ch1_pt	transversal momentum of child 1
Ch1_ptot	total momentum of child 1
sumChldNB	sum of the childrens NeuroBayes ouputs
prodChildNB	product of the childrens NeuroBayes ouputs

Table 22.8. A list of variables used for the NeuroBayes training for J/ψ

Variable name	Description
ptot	total momentum
dist_to_IP	distance to the IP
mom_dir_dev	deviation of the momentum from the direction to the IP
prodChildNB	product of the childrens NeuroBayes ouputs
sig_dist_to_IP	significance of the distance to the IP
sumChldNB	sum of the childrens NeuroBayes ouputs
ChX_NBout	NeuroBayes output of child X
ChX_PseudoHelAng	pseudo helicity angle of child X
ChX_ptot	total momentum of child X
ChX_Mass	invariant mass of child X
ChXY_InvMassScaled	scaled invariant mass of child pair X,Y with no mass information for the mother particle
ChXY_Angle	angle of child pair X,Y

Table 22.9. A list of variables used for the NeuroBayes training for D^+,D^0 ,D_s mesons

22.1.3. Stage 3 Trainings

Variable name	Description
ptot	total momentum
prodChildNB	product of the childrens NeuroBayes ouputs
sumChldNB	sum of the childrens NeuroBayes ouputs
ChX_NBout	NeuroBayes output of child X
ChX_PseudoHelAng	pseudo helicity angle of child X
ChX_ptot	total momentum of child X
ChX_Mass	invariant mass of child X

Table 22.10. A list of variables used for the NeuroBayes training for D^{*+},D^{*0} ,D_s^* mesons

22.1.4. Stage 4 Trainings

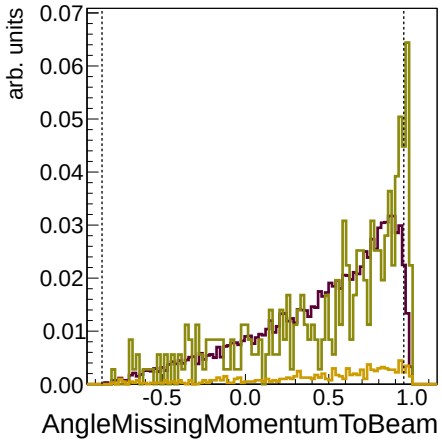
Variable name	Description
deltaE	ΔE of the B candidate
prodChildNB	product of the childrens NeuroBayes ouputs
sumChldNB	sum of the childrens NeuroBayes ouputs
CosThetaB	$cos\theta$ of the B candidate to the beam pipe
Dstar_D_massdiff	$D^* - D$ mass difference of the first D^*
D_hash_from_1st_dstar	decay hash of the first D^*
Dstar2_D_massdiff	$D^* - D$ mass difference of the second D^*
D_hash_from_2nd_dstar	decay hash of the first D^*
ChX_NBout	NeuroBayes output of child X
ChX_PseudoHelAng	pseudo helicity angle of child X
ChX_ptot	total momentum of child X
ChX_Mass	invariant mass of child X
ChX_hash	decay hash of child X
ChXY_InvMassScaled	scaled invariant mass of child pair X,Y with no mass information for the mother particle
ChXY_Angle	angle of child pair X,Y

Table 22.11. A list of variables used for the NeuroBayes training for B^+,B^0 mesons

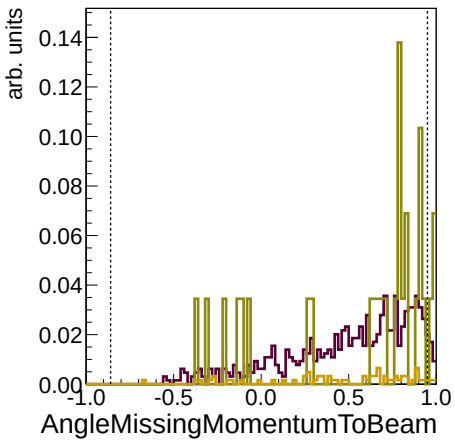
23. Appendix for the Analysis $B \to K^{(*)}\nu\bar{\nu}$

23.1. N-1 Plots

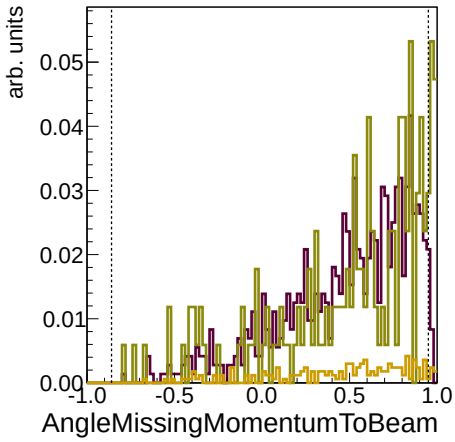
In the following you can see the so-called N-1 plots of the cut variables for the different modes separately. In these plots, all cuts are applied, except for the plotted one. This gives a direct feeling about the cut efficiencies for the different cut values. Because of correlations between the cut variables, it might be misleading to look only to the distribution of the individual variables, without the other cuts applied.



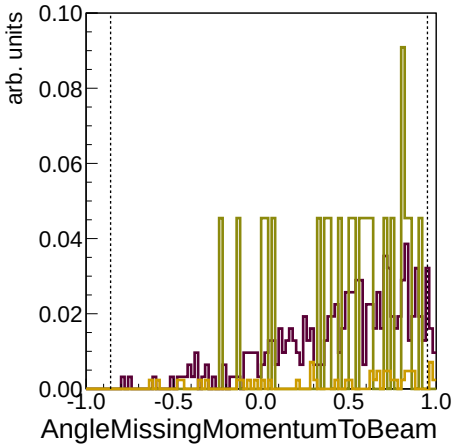
(a) K^+ mode



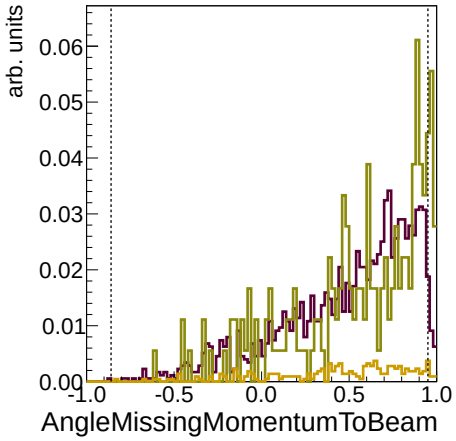
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

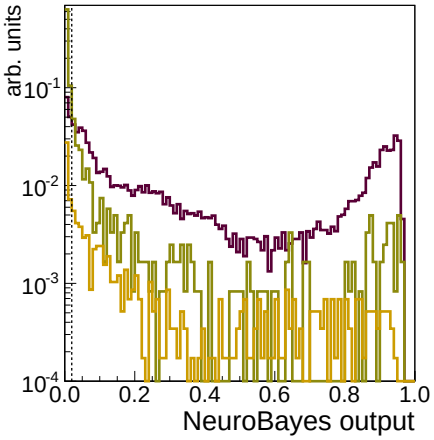


(d) $K^{*+} \to K^0_S\pi^+$ mode

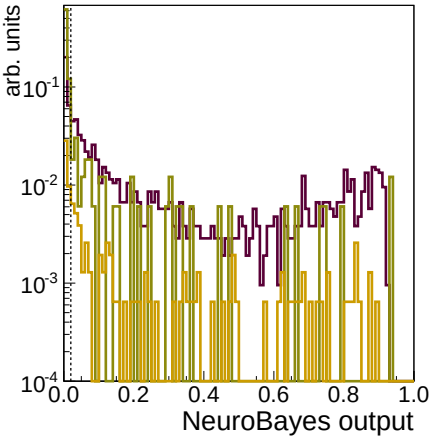


(e) $K^{*0} \to K^+\pi^-$ mode

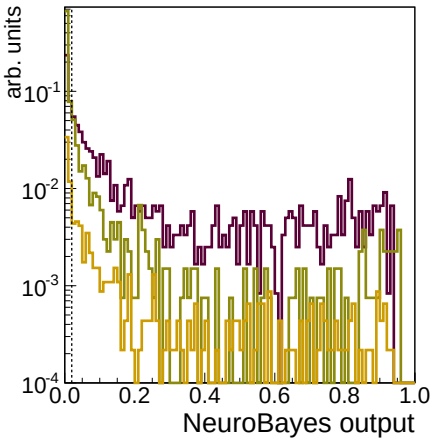
Figure 23.1. N-1 plots for AngleMissingMomentumToBeam. This variable is the angle of the missing momentum to the beam pipe. The purple histogram shows the signal distribution and is normalized to one. The green and the yellow histograms show the distributions for the generic and the rare Monte Carlo sample repectively and the sum of both is normalized to one. The cut values for this variable are marked with dashed lines.



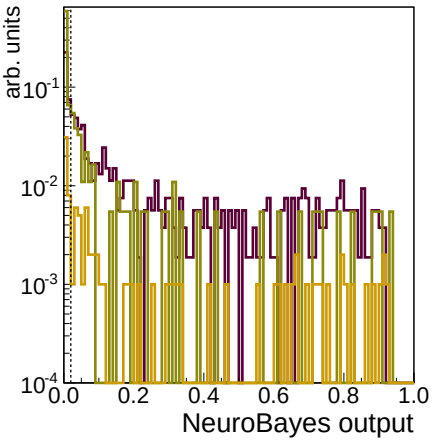
(a) K^+ mode



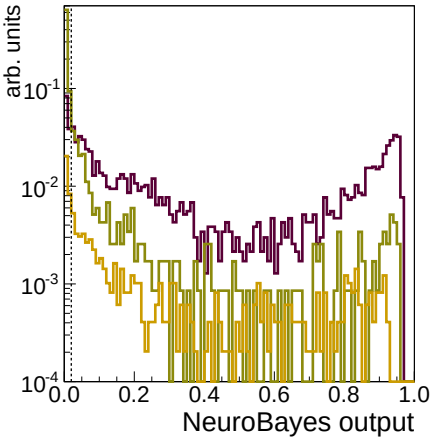
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

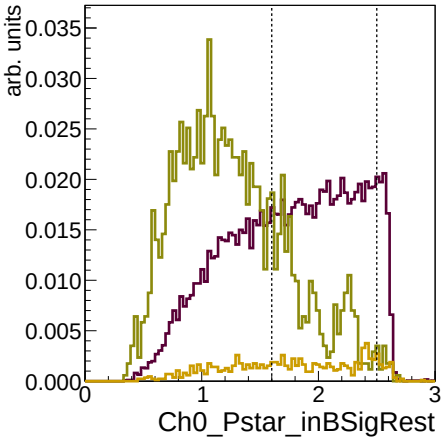


(d) $K^{*+} \to K^0_S\pi^+$ mode

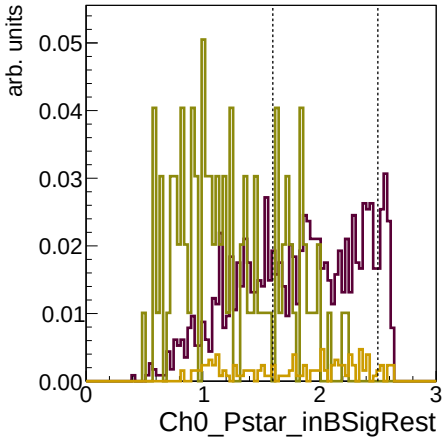


(e) $K^{*0} \to K^+\pi^-$ mode

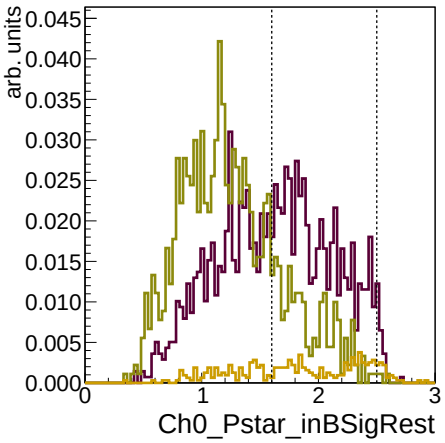
Figure 23.2. N-1 plots for BTagNBout. This variable is the NeuroBayes output of the B_{Tag} canditate. The purple histogram shows the signal distribution and is normalized to one. The green and the yellow histograms show the distributions for the generic and the rare Monte Carlo sample repectively and the sum of both is normalized to one. The cut values for this variable are marked with dashed lines.



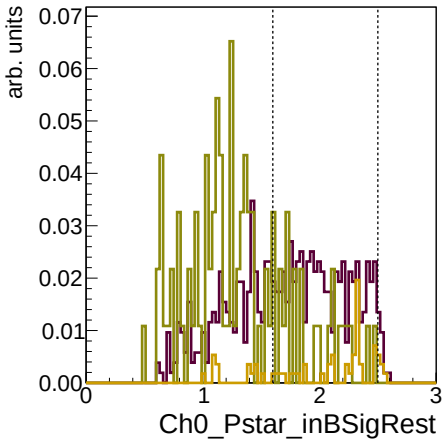
(a) K^+ mode



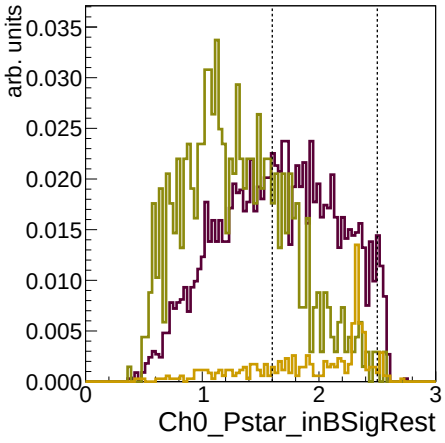
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

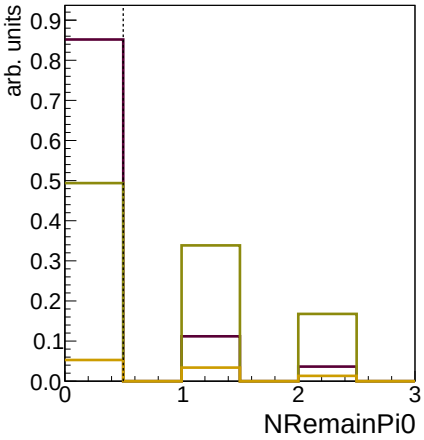


(d) $K^{*+} \to K^0_S\pi^+$ mode

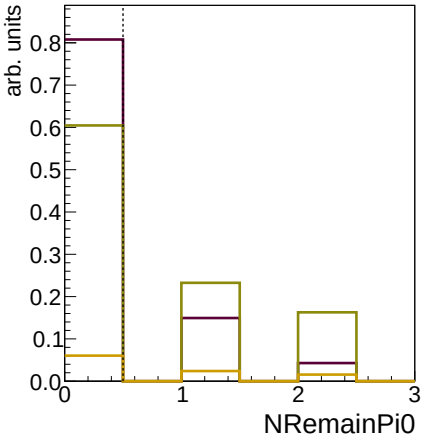


(e) $K^{*0} \to K^+\pi^-$ mode

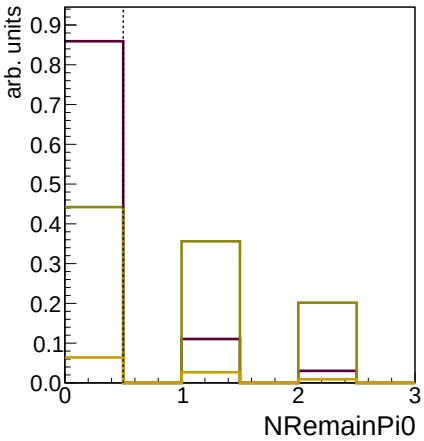
Figure 23.3. N-1 plots for Ch0PstarinBSigRest. This variable is the momentum of the light meson in the B_{Sig} rest frame. The purple histogram shows the signal distribution and is normalized to one. The green and the yellow histograms show the distributions for the generic and the rare Monte Carlo sample repectively and the sum of both is normalized to one. The cut values for this variable are marked with dashed lines.



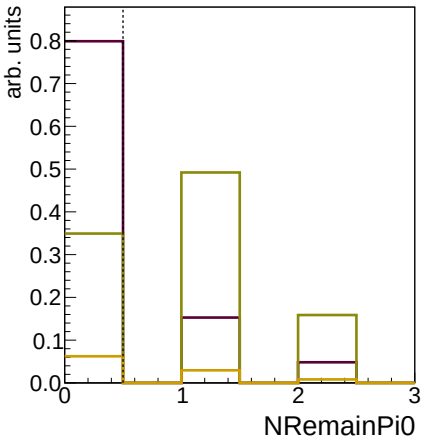
(a) K^+ mode



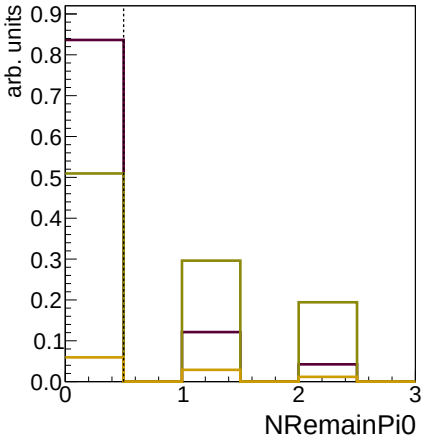
(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

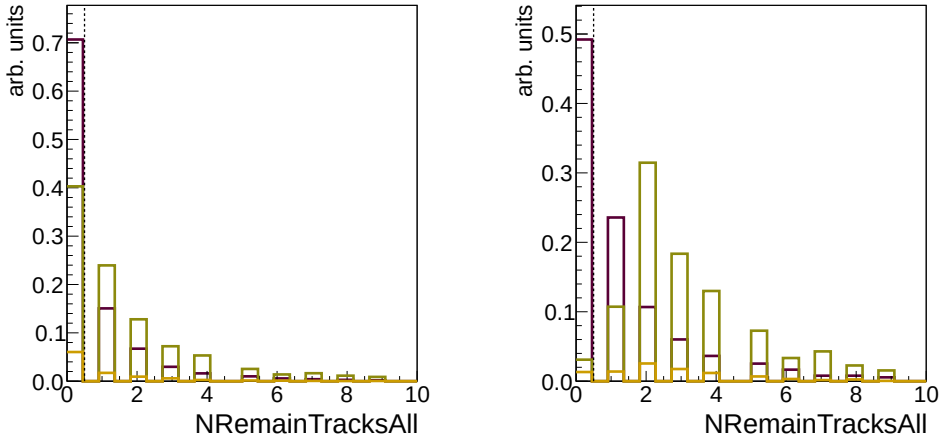


(d) $K^{*+} \to K^0_S\pi^+$ mode



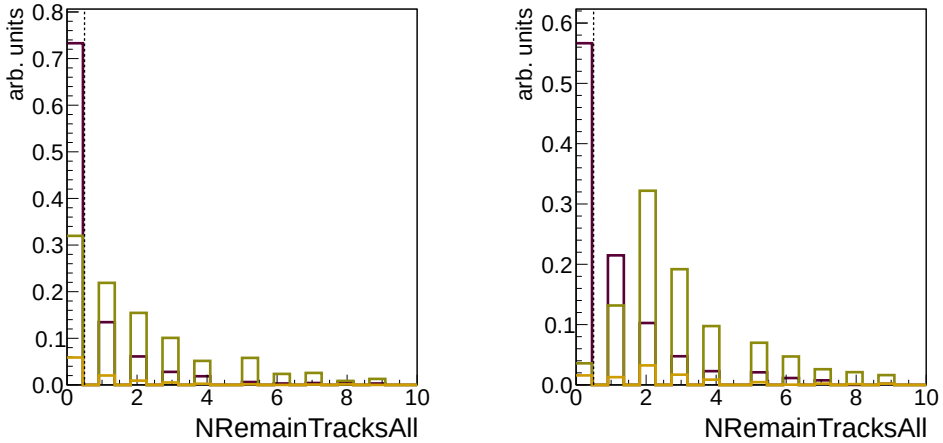
(e) $K^{*0} \to K^+\pi^-$ mode

Figure 23.4. N-1 plots for NRemainPi0. This variable is the number of remaining π^0 candidates. The purple histogram shows the signal distribution and is normalized to one. The green and the yellow histograms show the distributions for the generic and the rare Monte Carlo sample repectively and the sum of both is normalized to one. The cut values for this variable are marked with dashed lines.



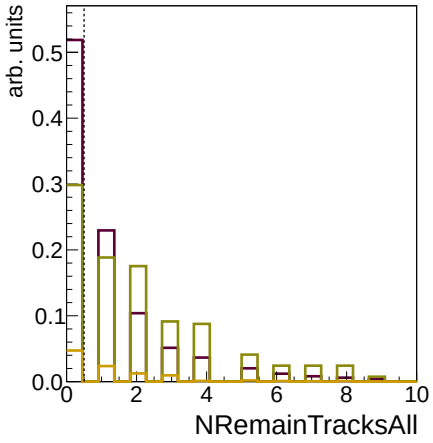
(a) K^+ mode

(b) K^0_S mode



(c) $K^{*+} \to K^+\pi^0$ mode

(d) $K^{*+} \to K^0_S\pi^+$ mode

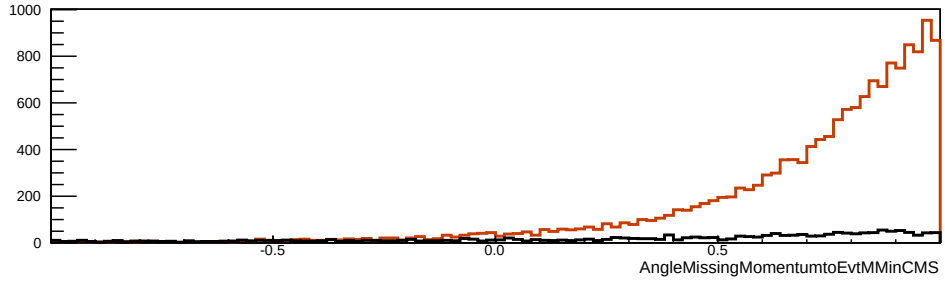


(e) $K^{*0} \to K^+\pi^-$ mode

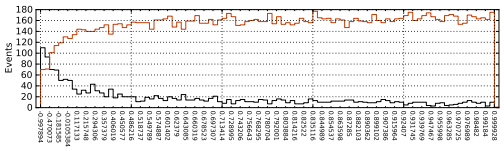
Figure 23.5. N-1 plots for NRemainTracksAll. This variable is number of remaining tracks. The purple histogram shows the signal distribution and is normalized to one. The green and the yellow histograms show the distributions for the generic and the rare Monte Carlo sample repectively and the sum of both is normalized to one. The cut values for this variable are marked with dashed lines.

23.2. Variables Used for the NeuroBayes training

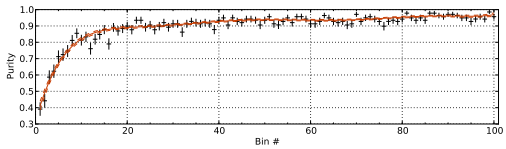
In the following you can see a summary of all the variables used for the trainings. For each mode, all variables are shown, ordered by the significance, contributed to the total separation of the training. For each variable, the distribution is shown for background (black) and signal (red) in the big figure (e.g figure 23.6a). In the smaller figure on the left (e.g figure 23.6b) you can see the flattened distribution for background (black) and signal (red) and in the small left figure (e.g figure 23.6c) the purity in bins of this variable. If there is no information available for a candidate in one vaiable, they are filled in an extra bin, marked yellow.



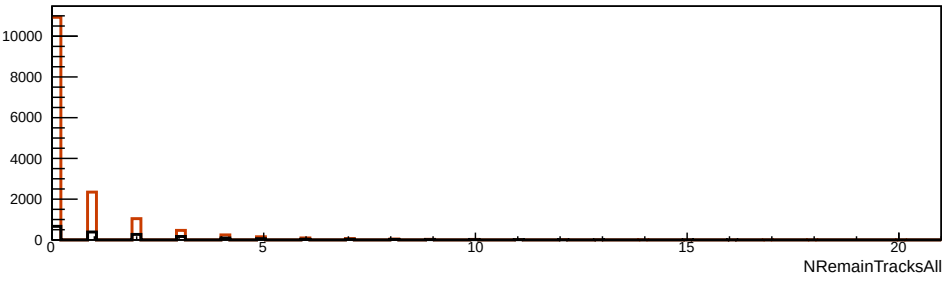
(a) AngleMissingMomentumtoEvtMMinCMS



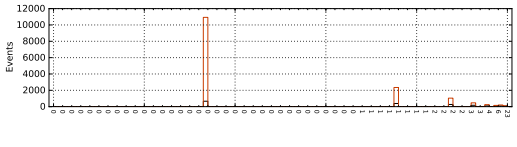
(b) Flat distribution



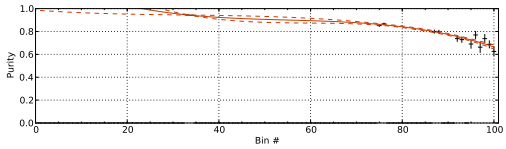
(c) Purity



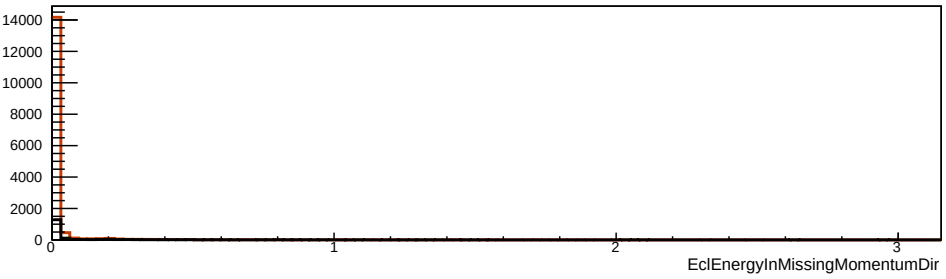
(d) NRemainTracksAll



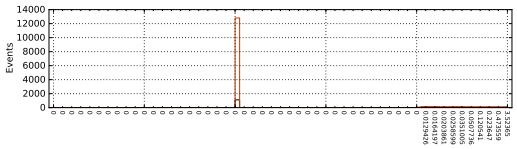
(e) Flat distribution



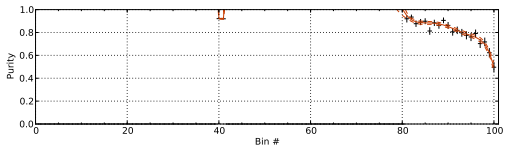
(f) Purity



(g) EclEnergyInMissingMomentumDir

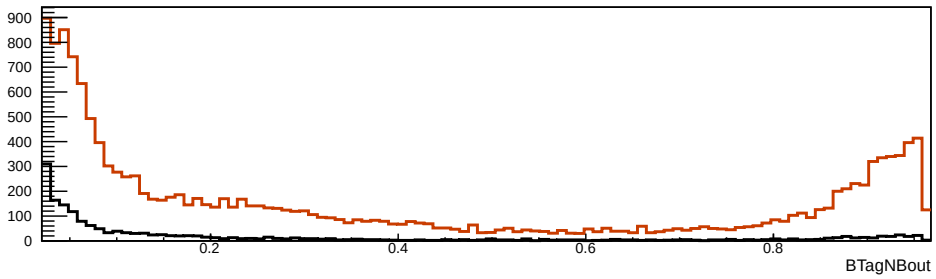


(h) Flat distribution

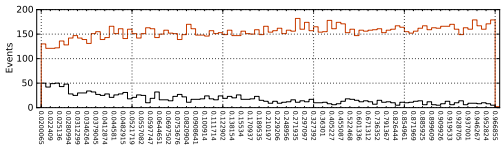


(i) Purity

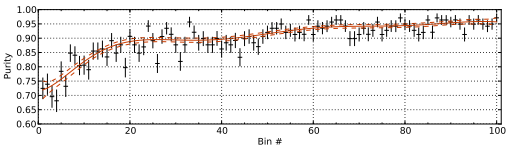
Figure 23.6. K^+ mode



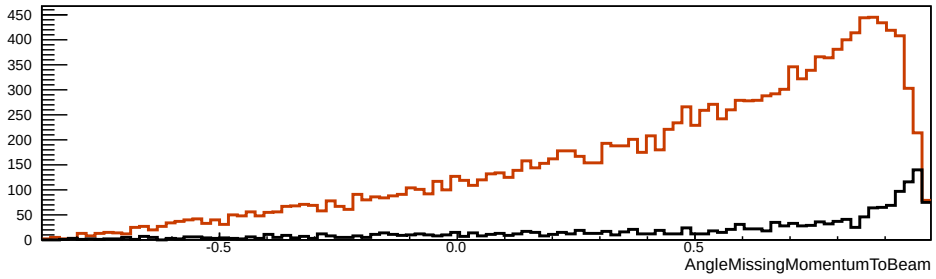
(a) BTagNBout



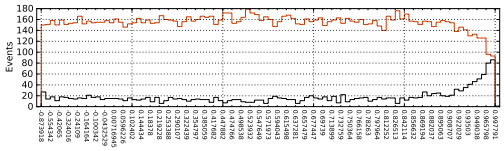
(b) Flat distribution



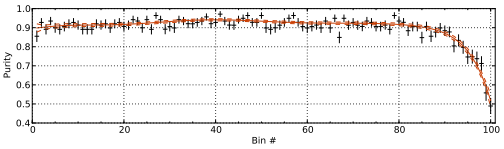
(c) Purity



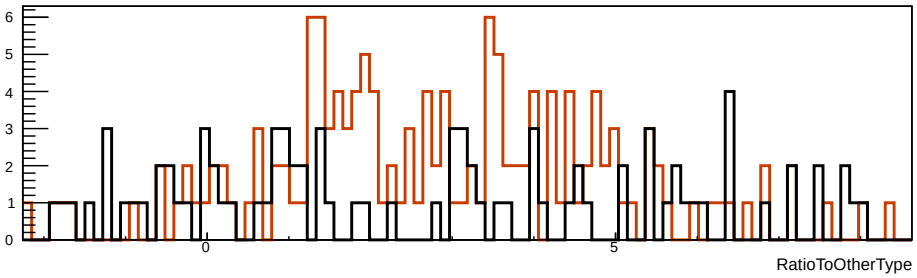
(d) AngleMissingMomentumToBeam



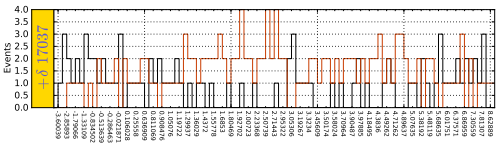
(e) Flat distribution



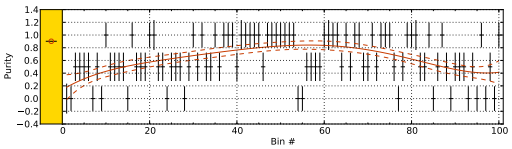
(f) Purity



(g) RatioToOtherType

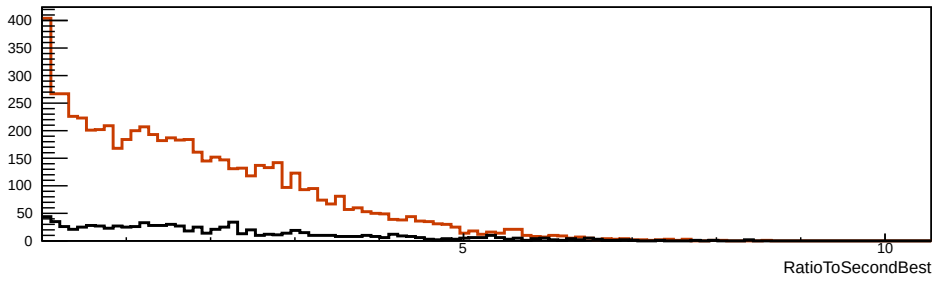


(h) Flat distribution

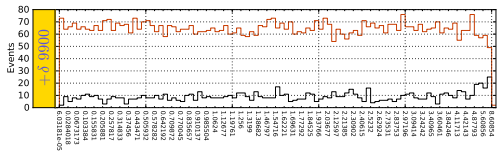


(i) Purity

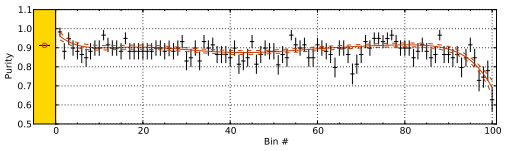
Figure 23.7. K^+ mode



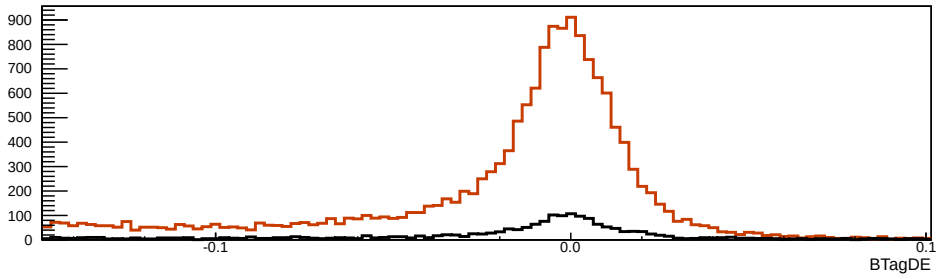
(a) RatioToSecondBest



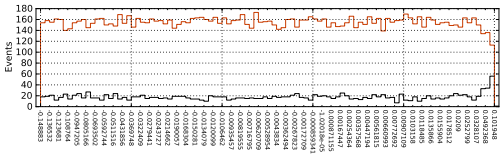
(b) Flat distribution



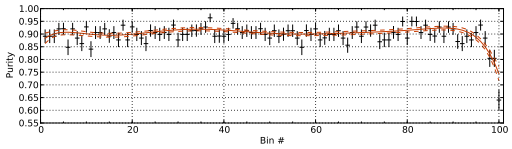
(c) Purity



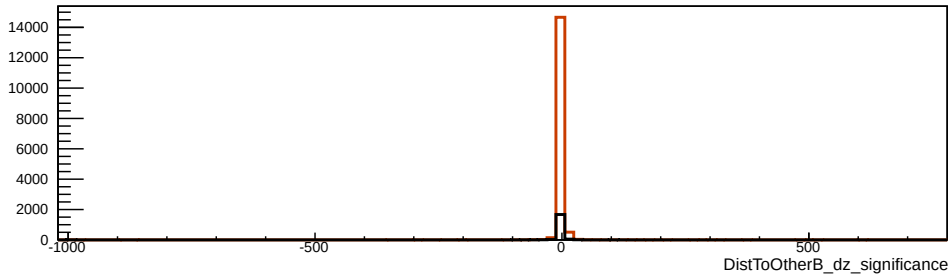
(d) BTagDE



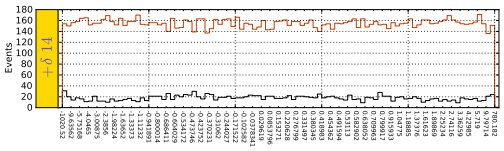
(e) Flat distribution



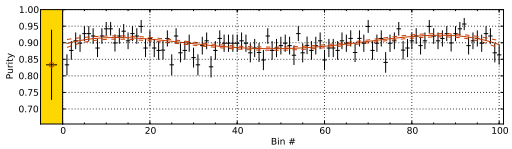
(f) Purity



(g) DistToOtherBdzsignificance

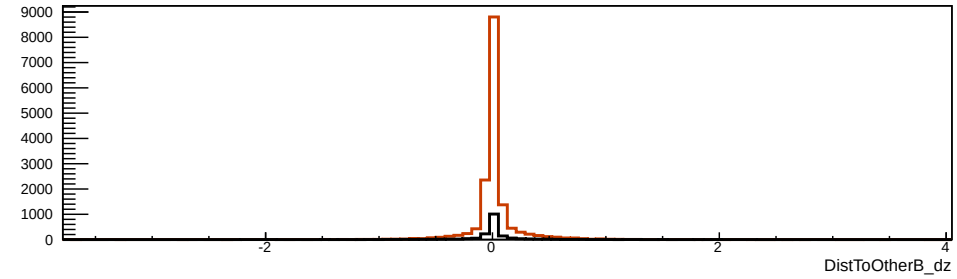


(h) Flat distribution

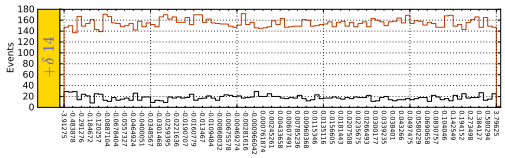


(i) Purity

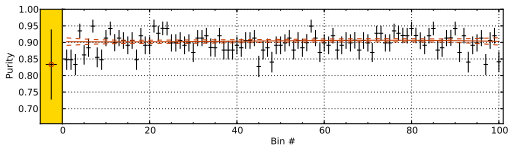
Figure 23.8. K^+ mode



(a) DistToOtherBdz

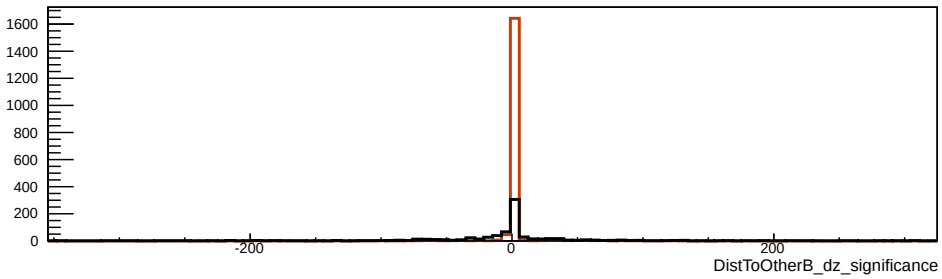


(b) Flat distribution

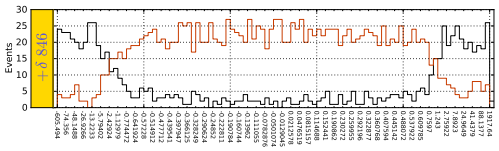


(c) Purity

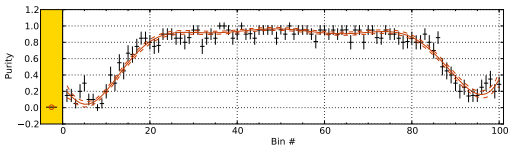
Figure 23.9. K^+ mode



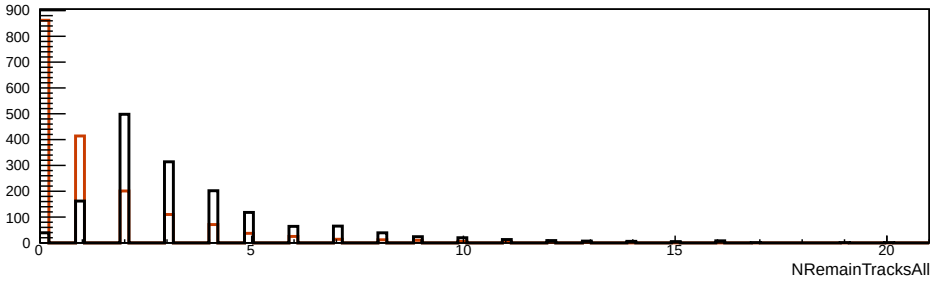
(a) DistToOtherBdzsignificance



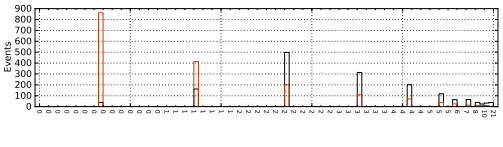
(b) Flat distribution



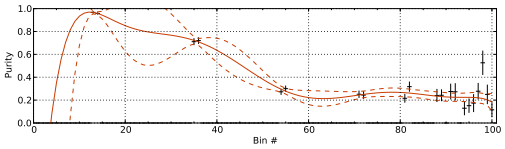
(c) Purity



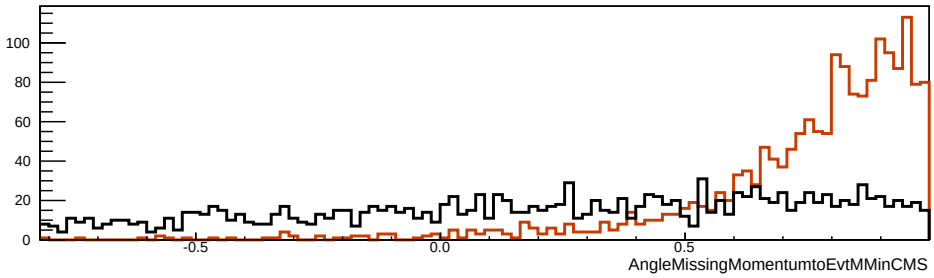
(d) NRemainTracksAll



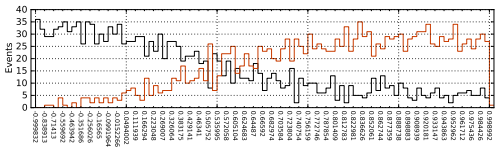
(e) Flat distribution



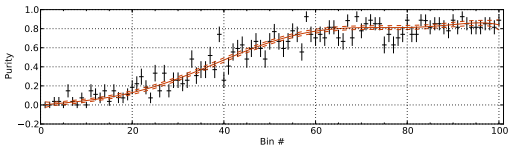
(f) Purity



(g) AngleMissingMomentumtoEvtMMinCMS

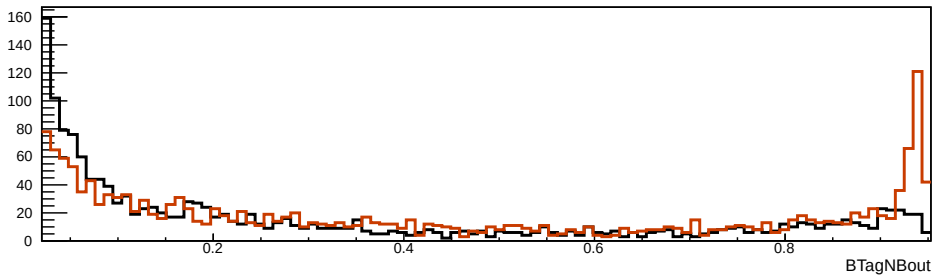


(h) Flat distribution

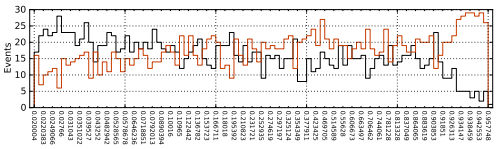


(i) Purity

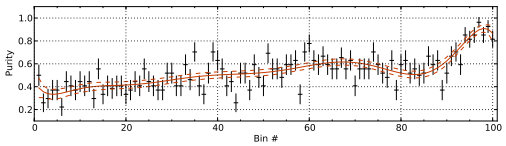
Figure 23.10. K^0_S mode



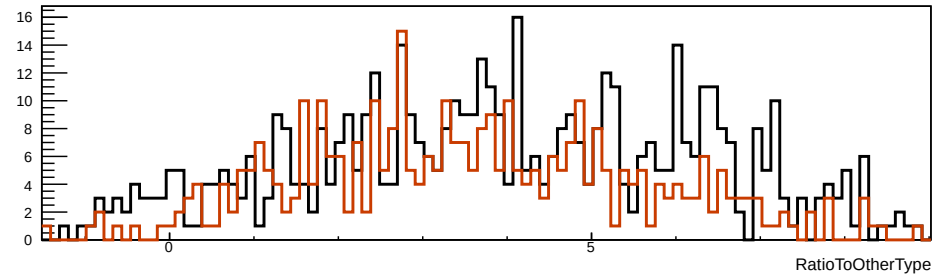
(a) BTagNBout



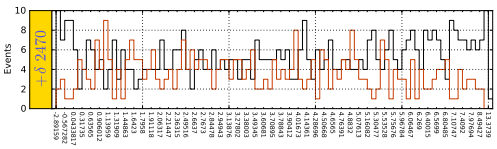
(b) Flat distribution



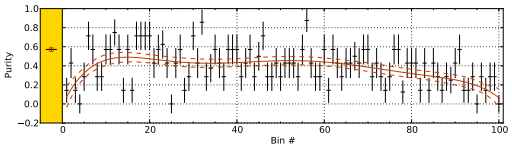
(c) Purity



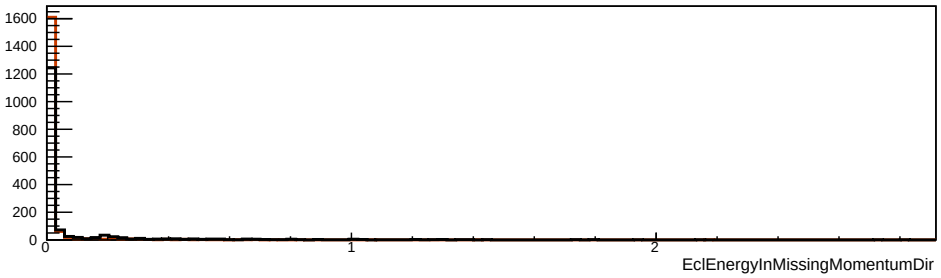
(d) RatioToOtherType



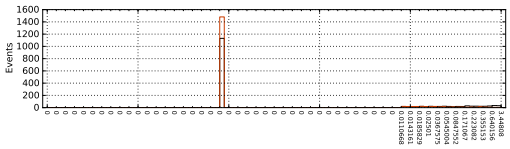
(e) Flat distribution



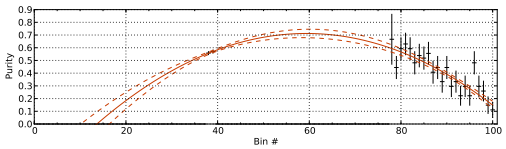
(f) Purity



(g) EclEnergyInMissingMomentumDir

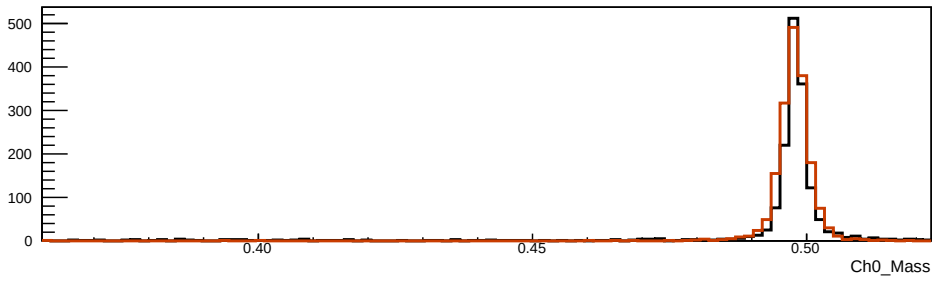


(h) Flat distribution

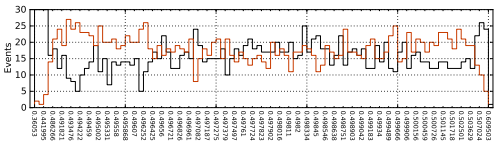


(i) Purity

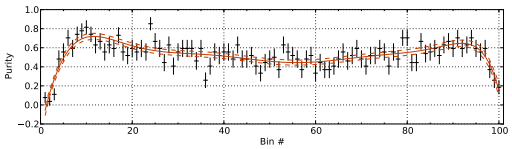
Figure 23.11. K^0_S mode



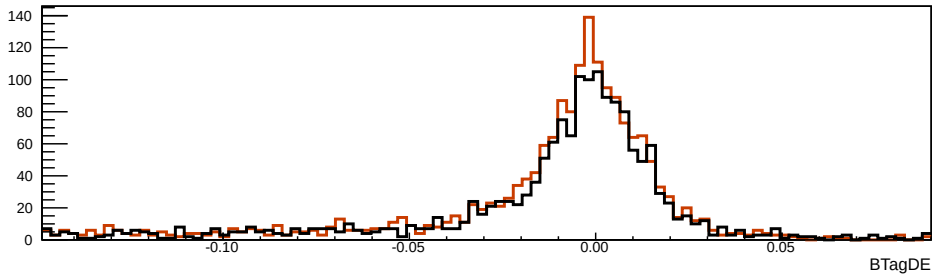
(a) Ch0Mass



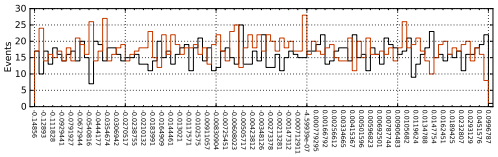
(b) Flat distribution



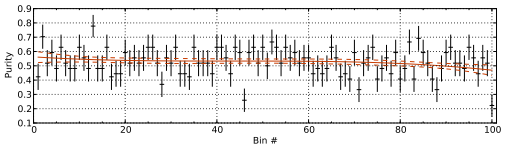
(c) Purity



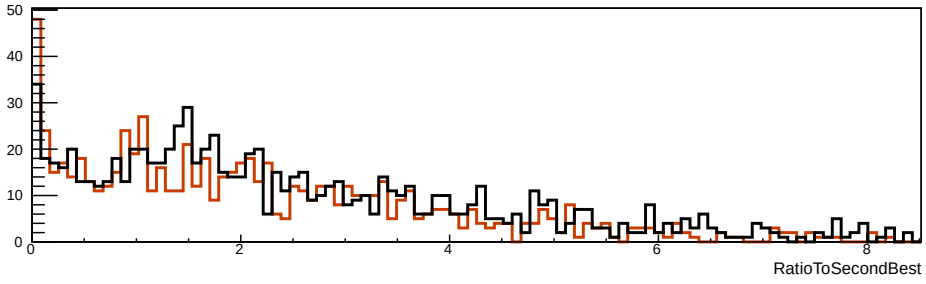
(d) BTagDE



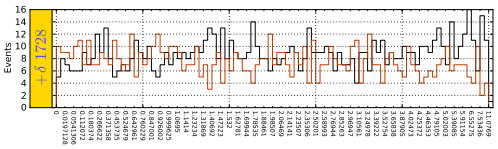
(e) Flat distribution



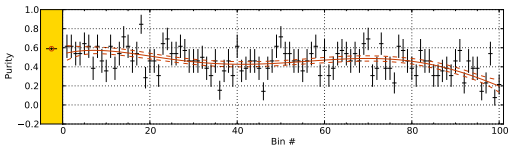
(f) Purity



(g) RatioToSecondBest

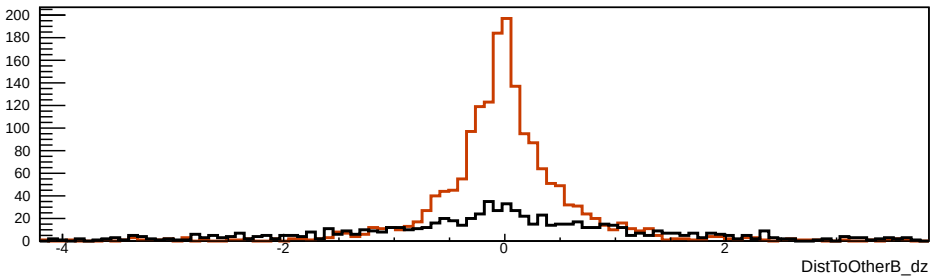


(h) Flat distribution

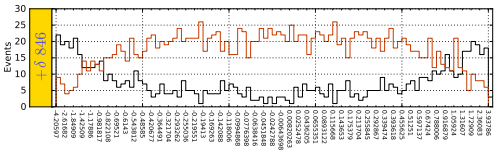


(i) Purity

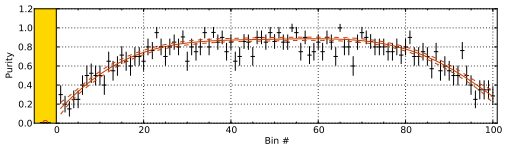
Figure 23.12. K^0_S mode



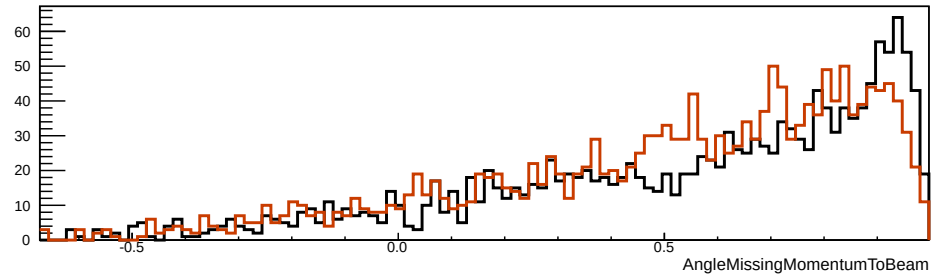
(a) DistToOtherBdz



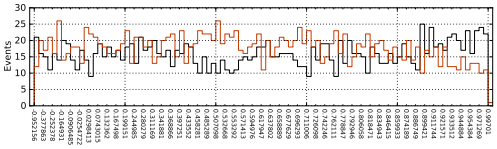
(b) Flat distribution



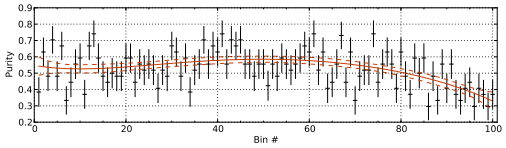
(c) Purity



(d) AngleMissingMomentumToBeam

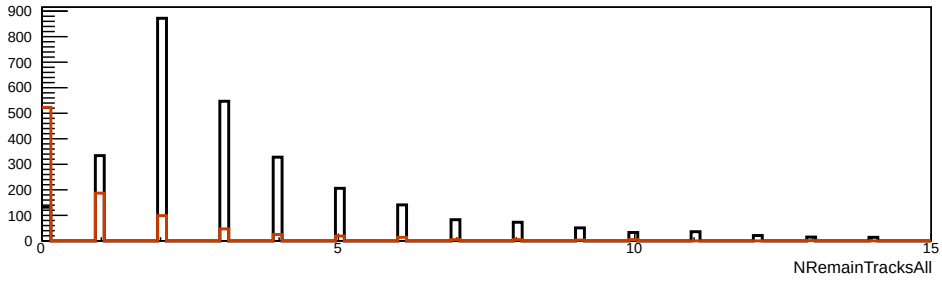


(e) Flat distribution

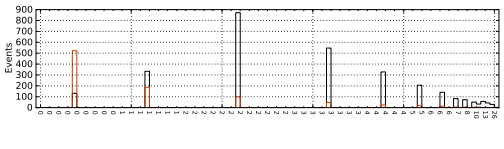


(f) Purity

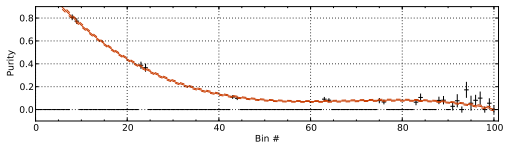
Figure 23.13. K_S^0 mode



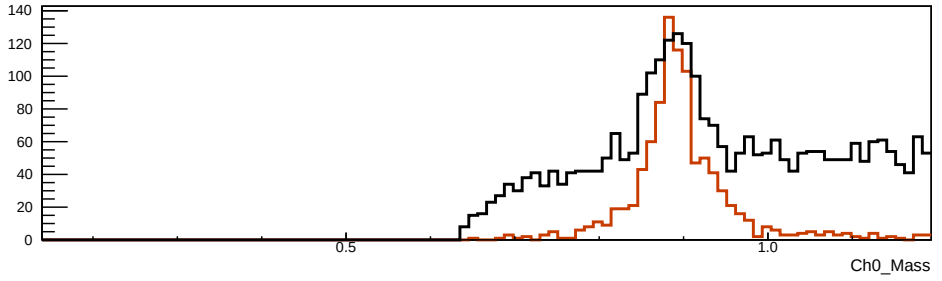
(a) NRemainTracksAll



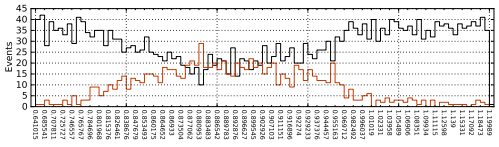
(b) Flat distribution



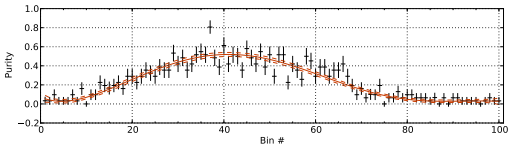
(c) Purity



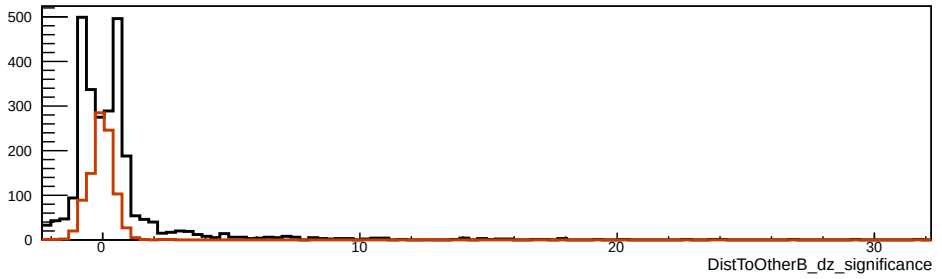
(d) Ch0Mass



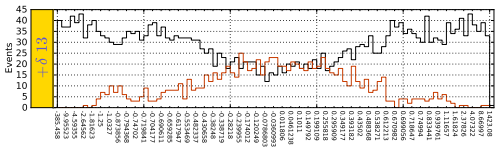
(e) Flat distribution



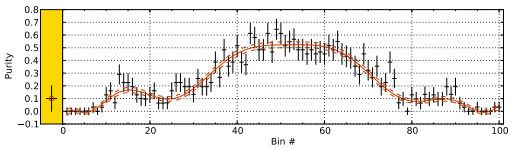
(f) Purity



(g) DistToOtherBdzsignificance

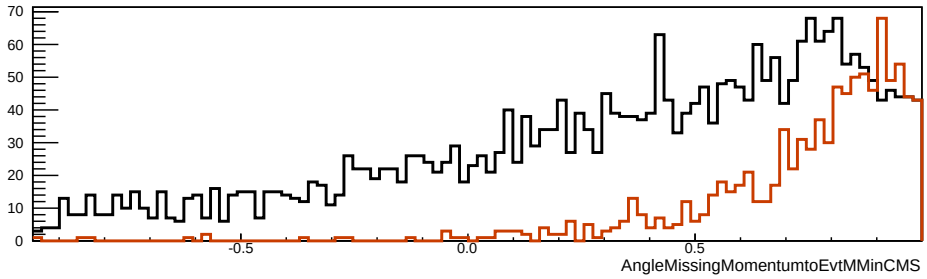


(h) Flat distribution

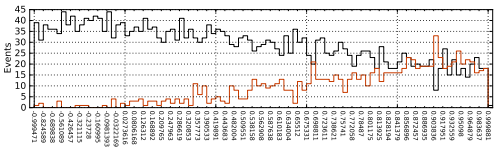


(i) Purity

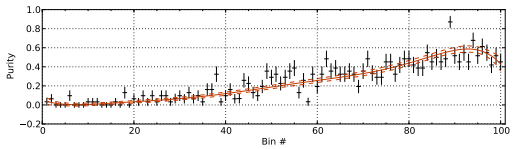
Figure 23.14. $K^{*+} \to K^0_S\pi^+$ mode



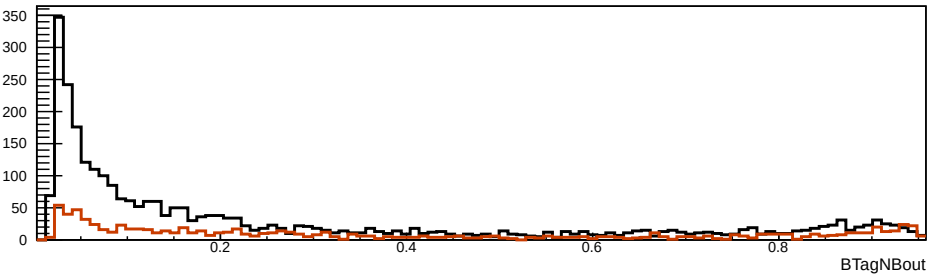
(a) AngleMissingMomentumtoEvtMMinCMS



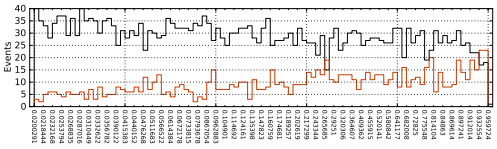
(b) Flat distribution



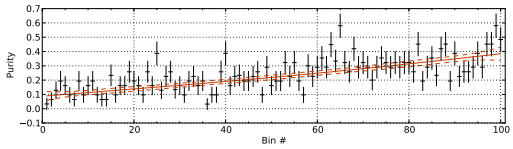
(c) Purity



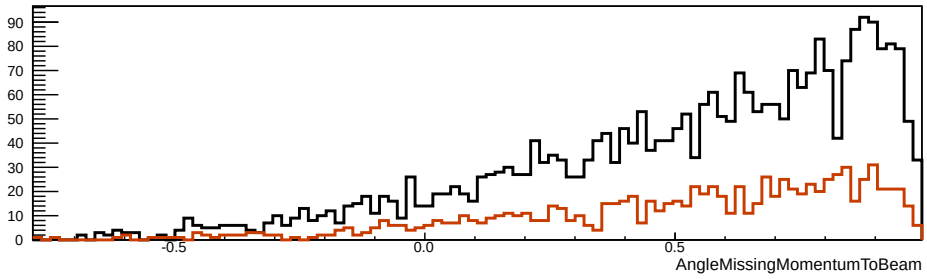
(d) BTagNBout



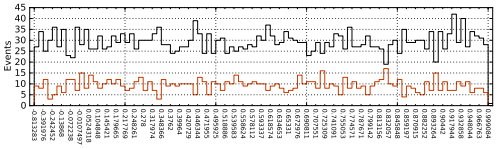
(e) Flat distribution



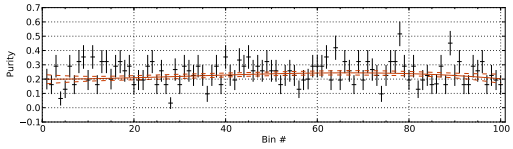
(f) Purity



(g) AngleMissingMomentumToBeam

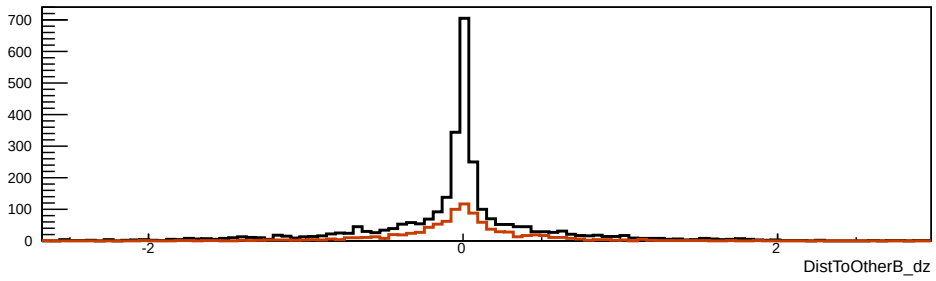


(h) Flat distribution

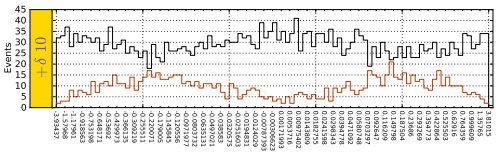


(i) Purity

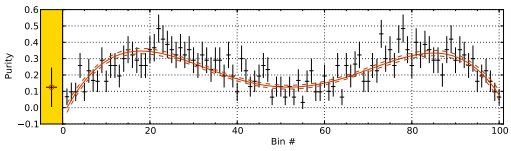
Figure 23.15. $K^{*+} \to K^0_S \pi^+$ mode



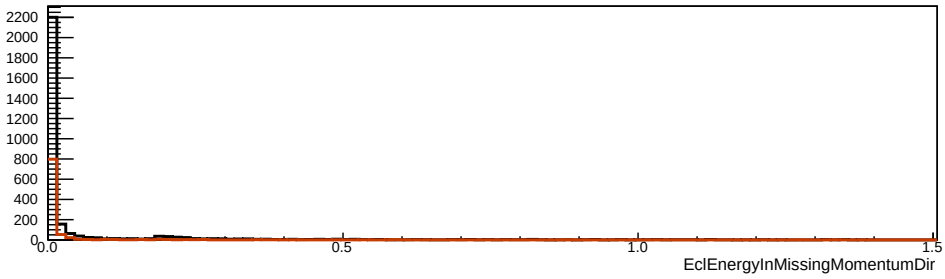
(a) DistToOtherBdz



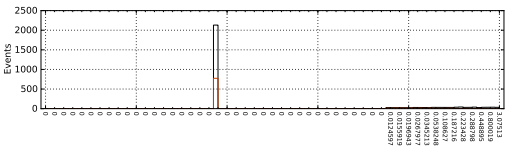
(b) Flat distribution



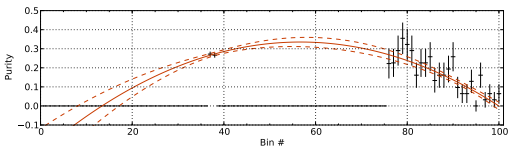
(c) Purity



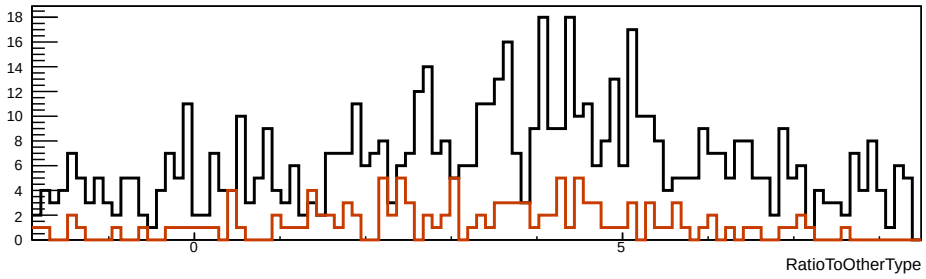
(d) EclEnergyInMissingMomentumDir



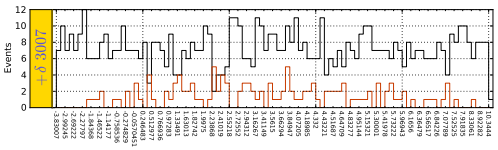
(e) Flat distribution



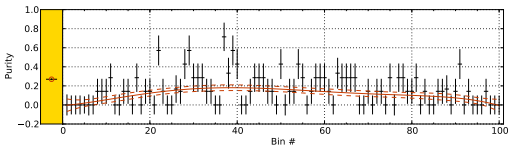
(f) Purity



(g) RatioToOtherType

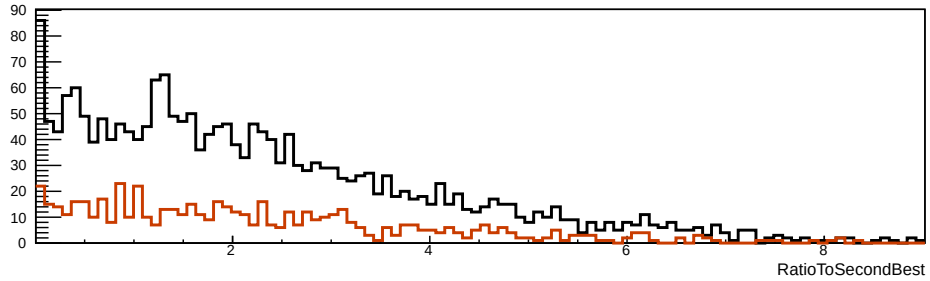


(h) Flat distribution

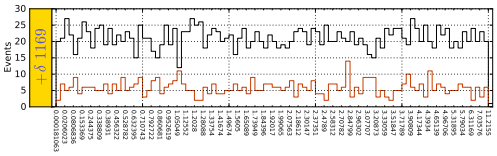


(i) Purity

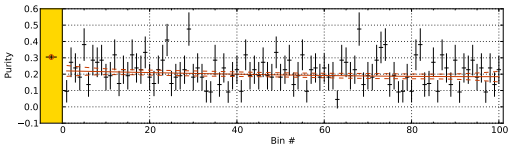
Figure 23.16. $K^{*+} \to K^0_S\pi^+$ mode



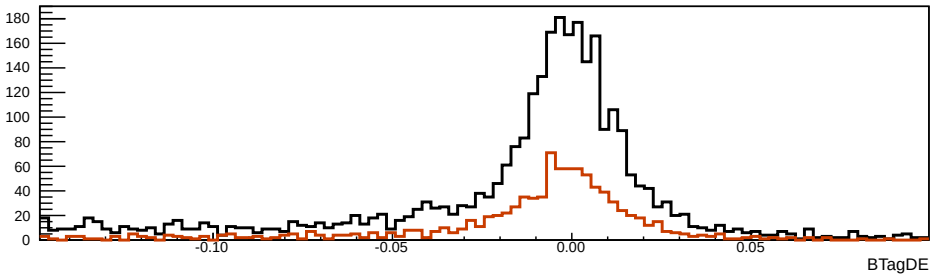
(a) RatioToSecondBest



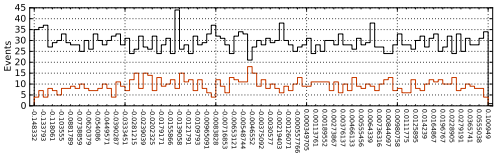
(b) Flat distribution



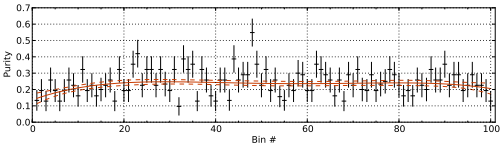
(c) Purity



(d) BTagDE

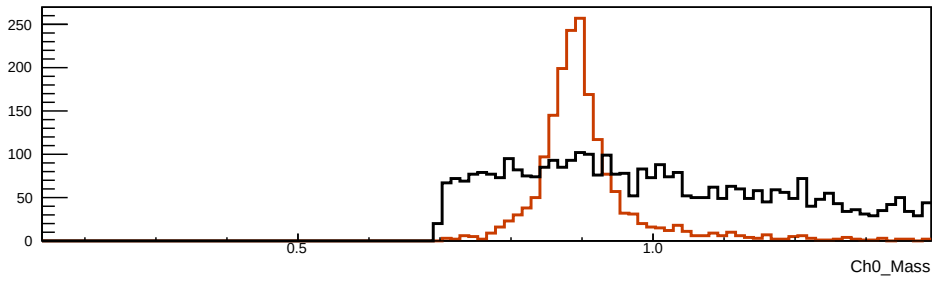


(e) Flat distribution

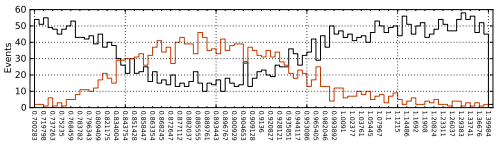


(f) Purity

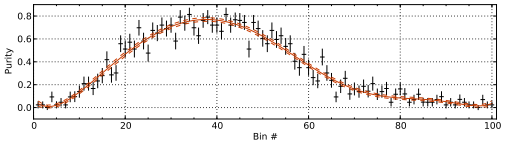
Figure 23.17. $K^{*+} \to K^0_S \pi^+$ mode



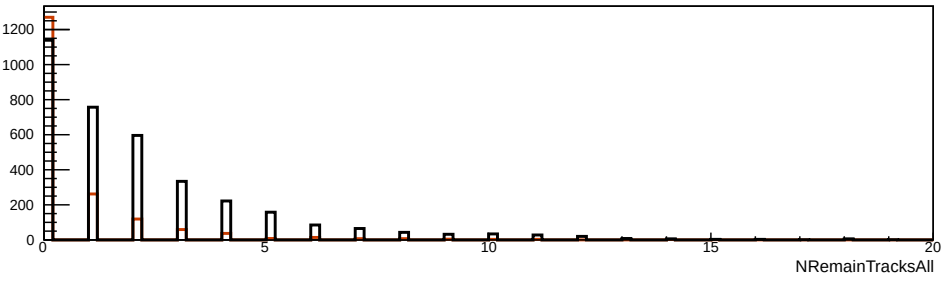
(a) Ch0Mass



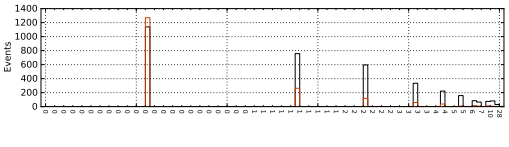
(b) Flat distribution



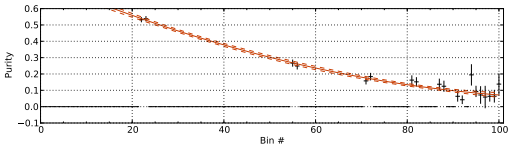
(c) Purity



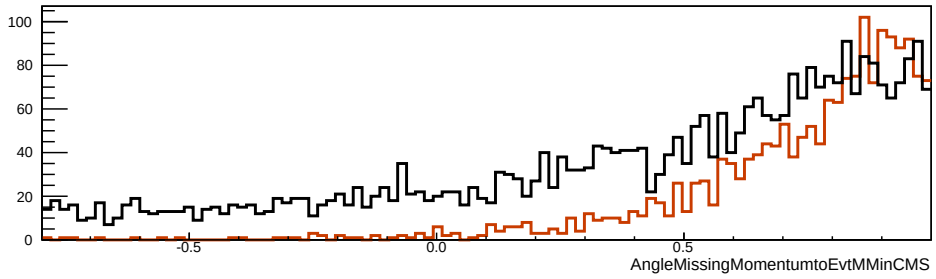
(d) NRemainTracksAll



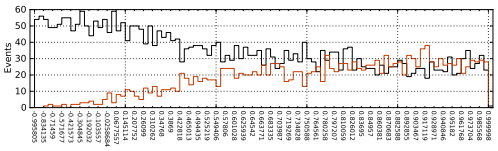
(e) Flat distribution



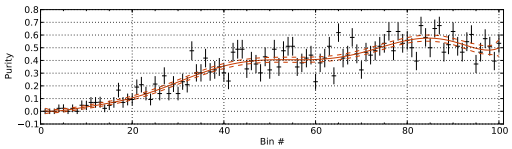
(f) Purity



(g) AngleMissingMomentumtoEvtMMinCMS

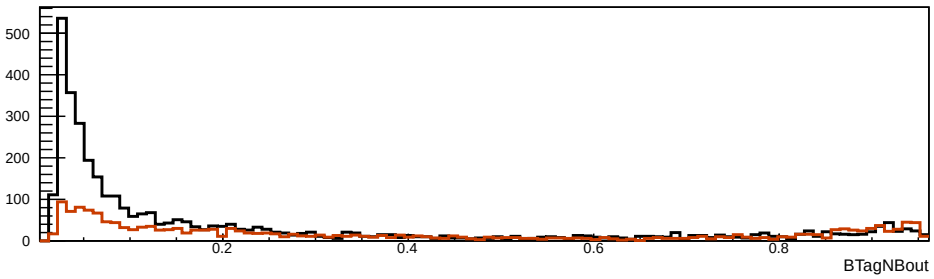


(h) Flat distribution

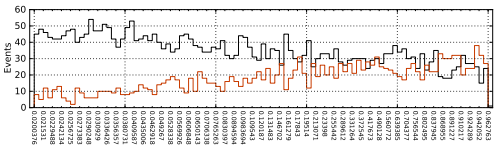


(i) Purity

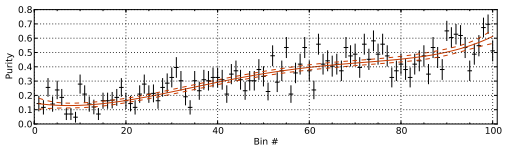
Figure 23.18. $K^{*+} \to K^+\pi^0$ mode



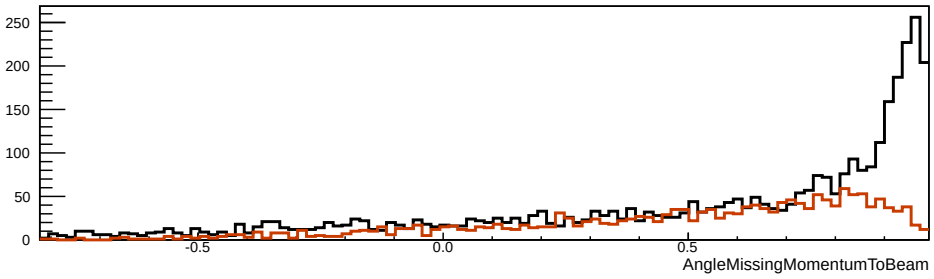
(a) BTagNBout



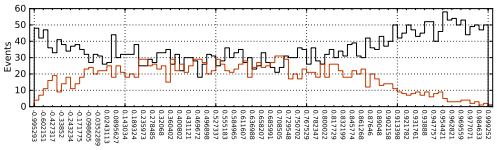
(b) Flat distribution



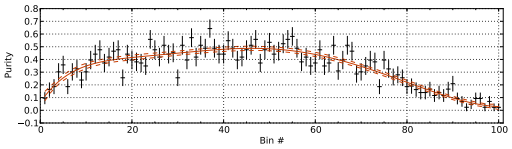
(c) Purity



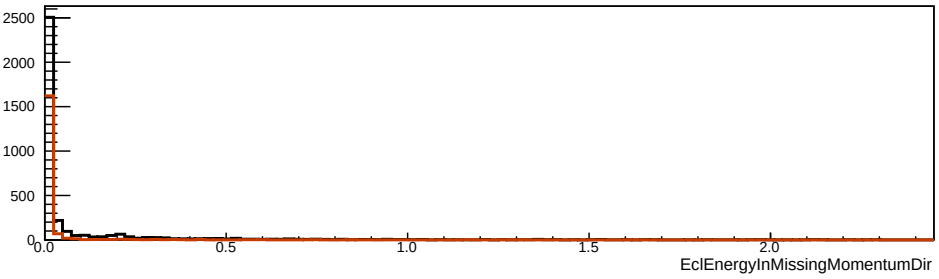
(d) AngleMissingMomentumToBeam



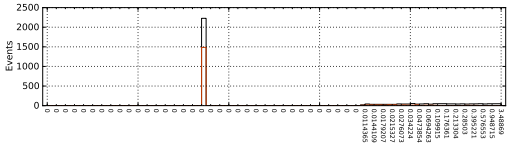
(e) Flat distribution



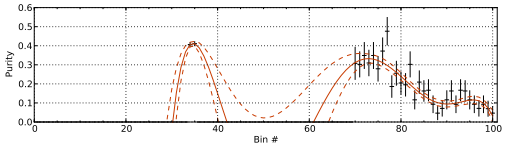
(f) Purity



(g) EclEnergyInMissingMomentumDir

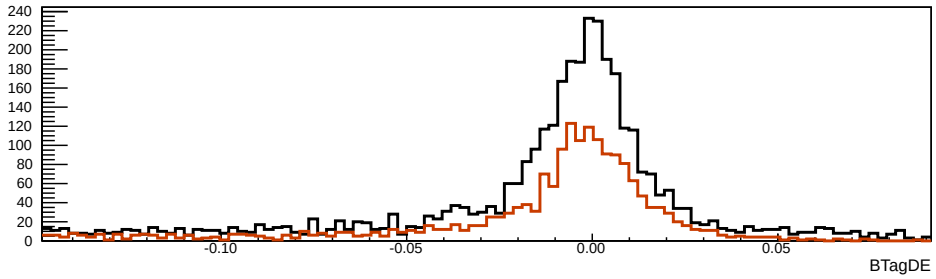


(h) Flat distribution

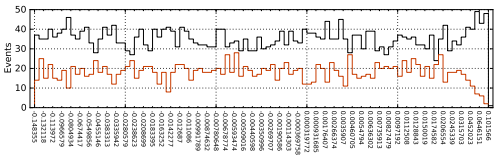


(i) Purity

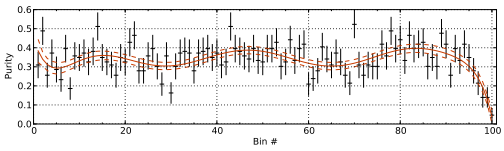
Figure 23.19. $K^{*+} \to K^+\pi^0$ mode



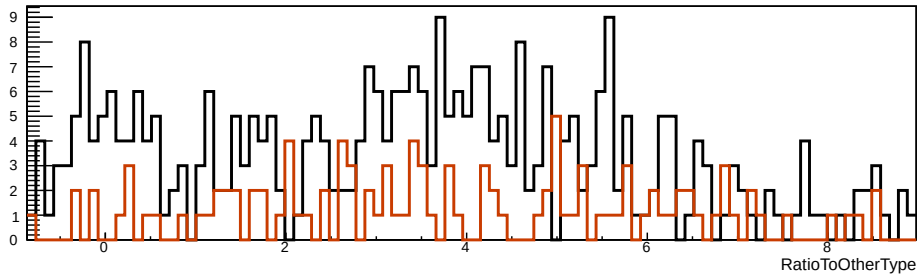
(a) BTagDE



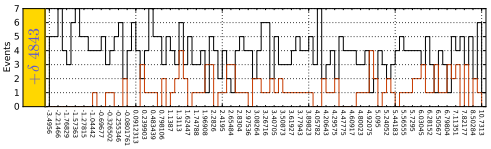
(b) Flat distribution



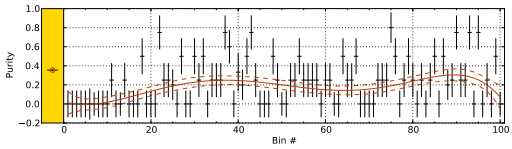
(c) Purity



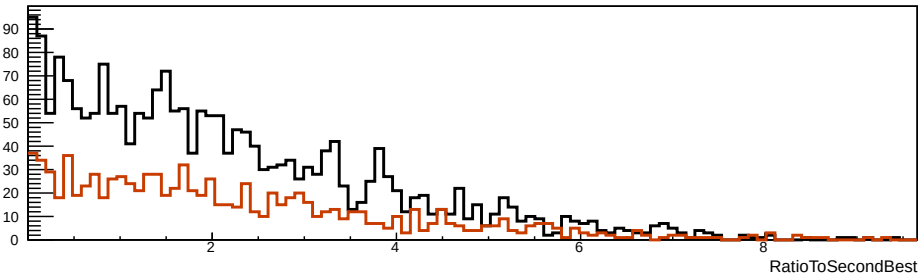
(d) RatioToOtherType



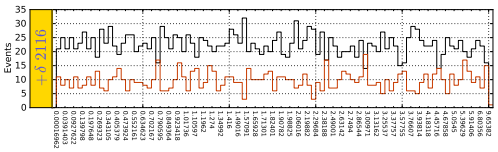
(e) Flat distribution



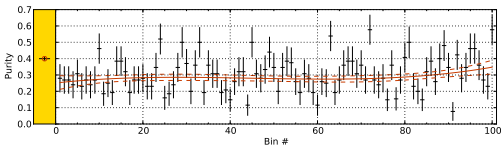
(f) Purity



(g) RatioToSecondBest

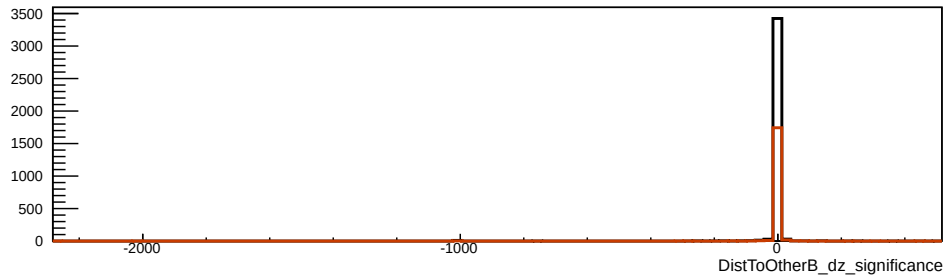


(h) Flat distribution

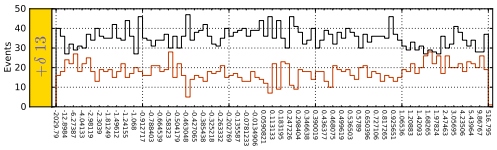


(i) Purity

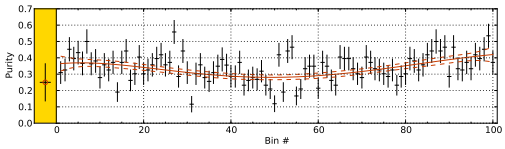
Figure 23.20. $K^{*+} \to K^+\pi^0$ mode



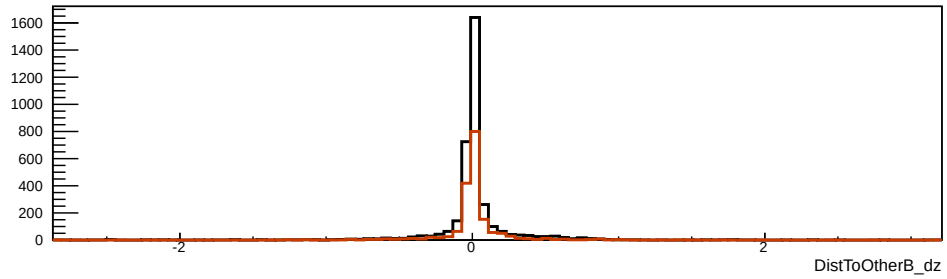
(a) DistToOtherBdzsignificance



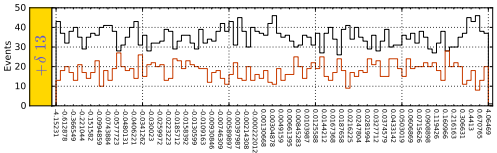
(b) Flat distribution



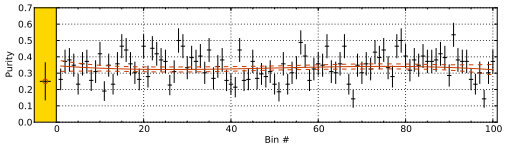
(c) Purity



(d) DistToOtherBdz

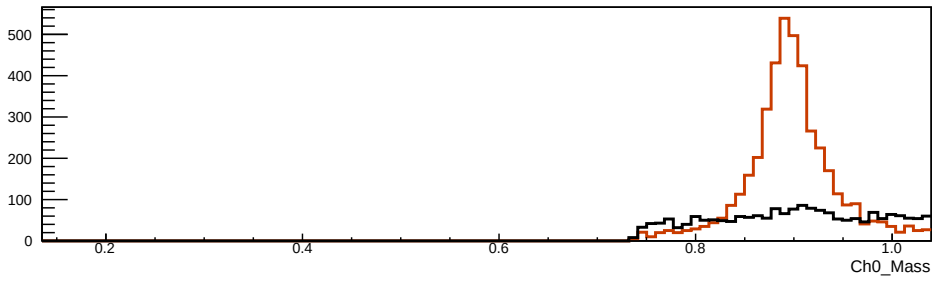


(e) Flat distribution

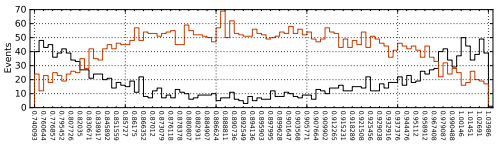


(f) Purity

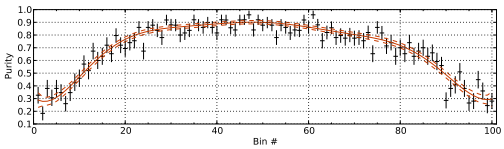
Figure 23.21. $K^{*+} \to K^+\pi^0$ mode



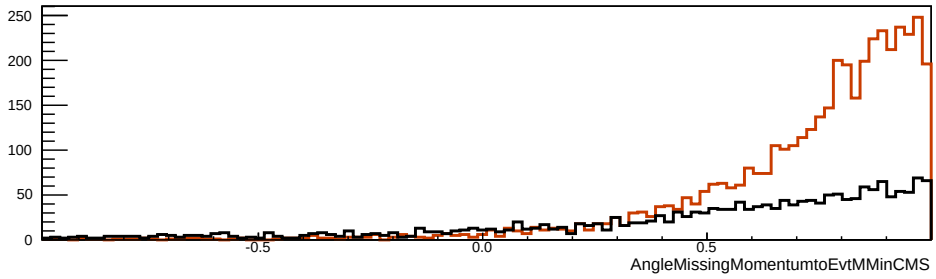
(a) Ch0Mass



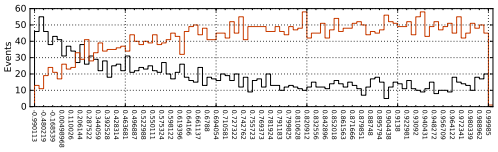
(b) Flat distribution



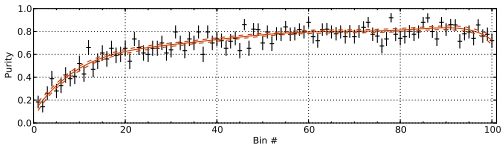
(c) Purity



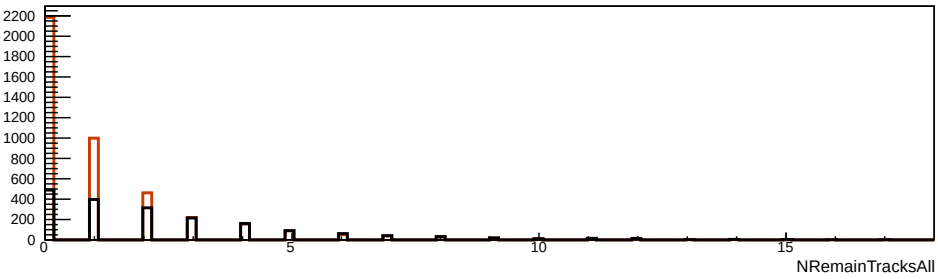
(d) AngleMissingMomentumtoEvtMMinCMS



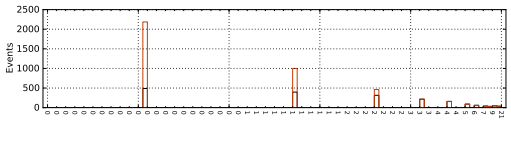
(e) Flat distribution



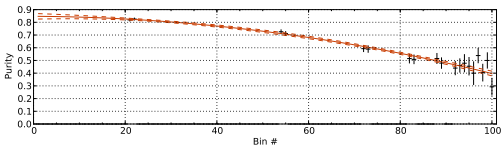
(f) Purity



(g) NRemainTracksAll

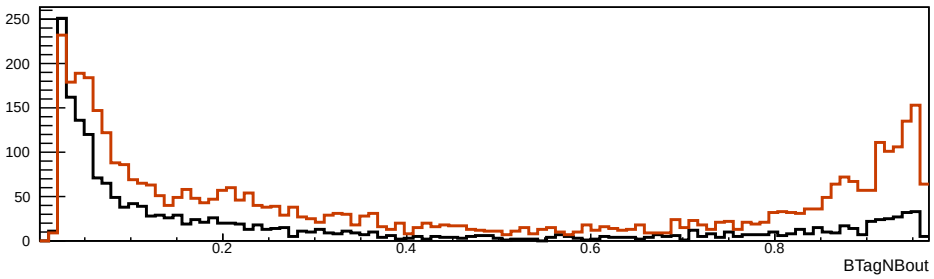


(h) Flat distribution

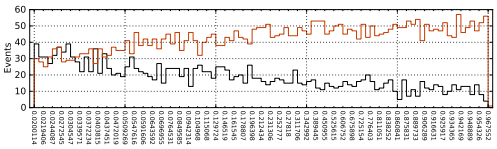


(i) Purity

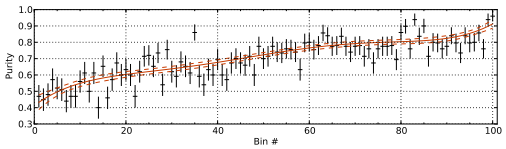
Figure 23.22. $K^{*0} \to K^+\pi^-$ mode



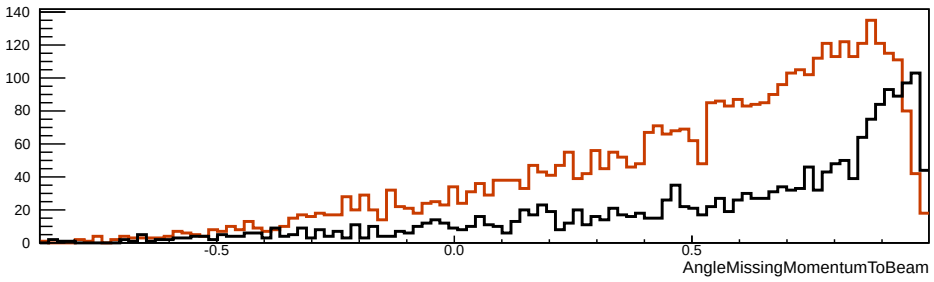
(a) BTagNBout



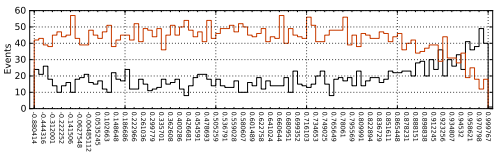
(b) Flat distribution



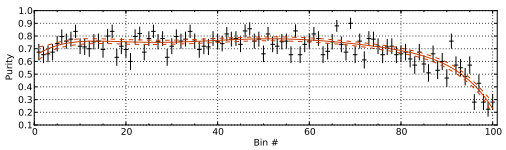
(c) Purity



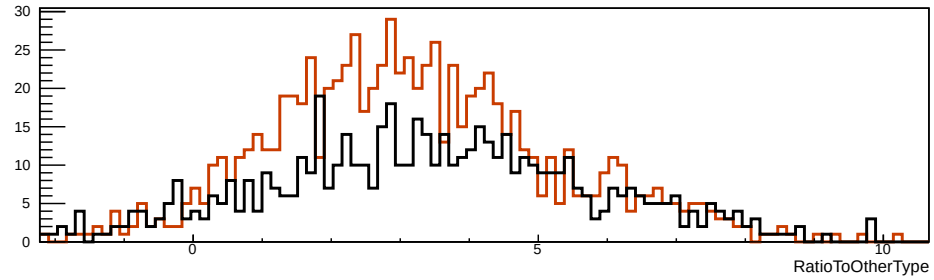
(d) AngleMissingMomentumToBeam



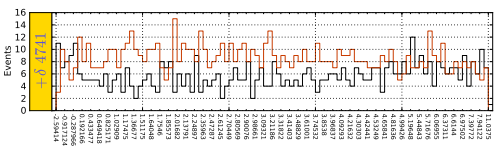
(e) Flat distribution



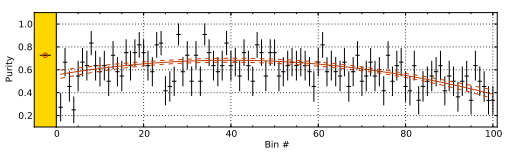
(f) Purity



(g) RatioToOtherType

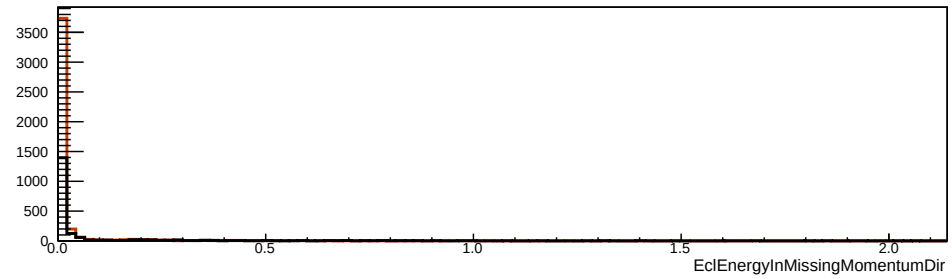


(h) Flat distribution

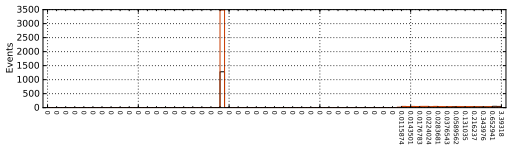


(i) Purity

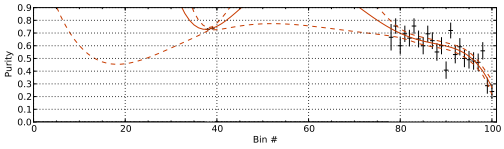
Figure 23.23. $K^{*0} \to K^+\pi^-$ mode



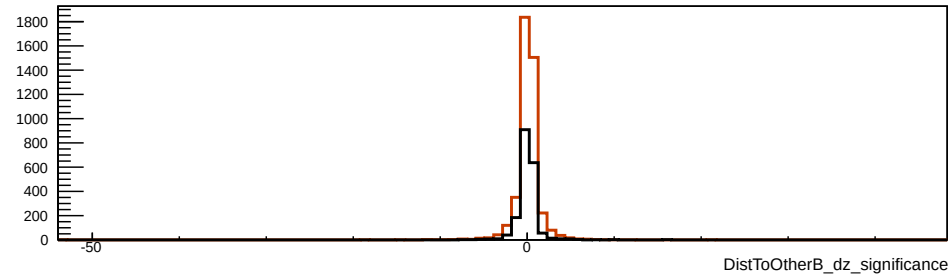
(a) EclEnergyInMissingMomentumDir



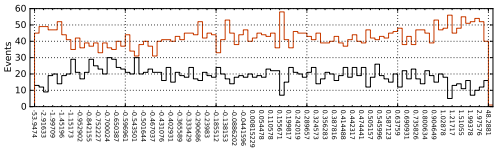
(b) Flat distribution



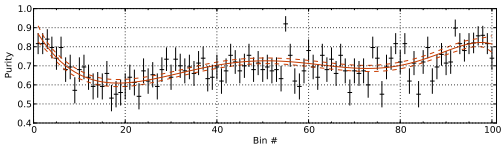
(c) Purity



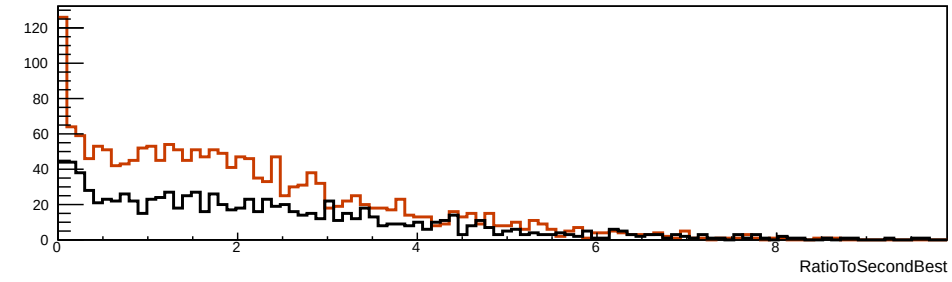
(d) DistToOtherBdzsignificance



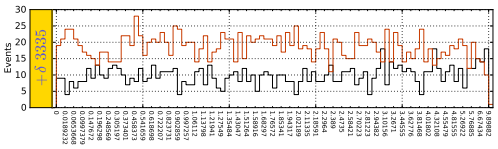
(e) Flat distribution



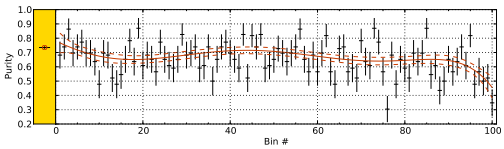
(f) Purity



(g) RatioToSecondBest

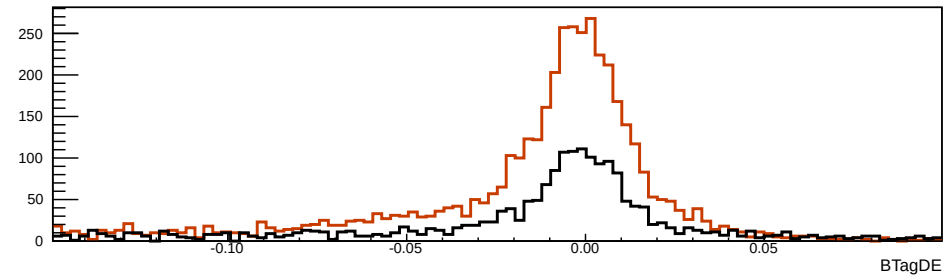


(h) Flat distribution

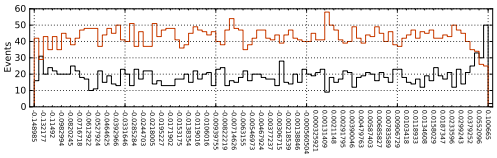


(i) Purity

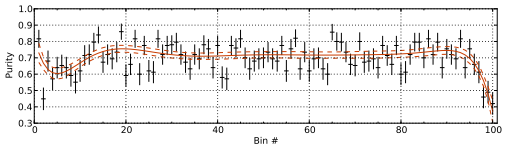
Figure 23.24. $K^{*0} \to K^+\pi^-$ mode



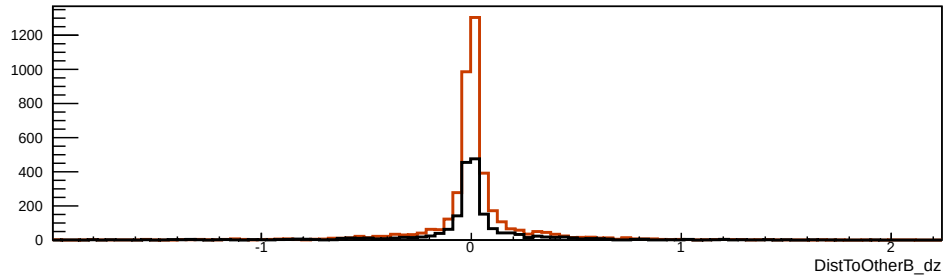
(a) BTagDE



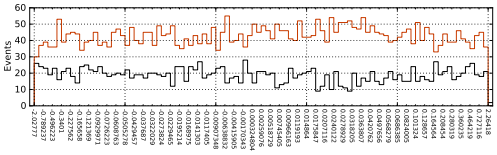
(b) Flat distribution



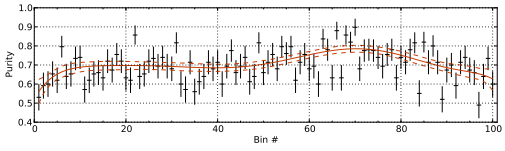
(c) Purity



(d) DistToOtherBdz



(e) Flat distribution



(f) Purity

Figure 23.25. $K^{*0} \to K^+\pi^-$ mode

List of Figures

Bibliography

[Aal+02] C. E. Aalseth et al. "Comment on 'Evidence for Neutrinoless Double Beta Decay'". In: *Mod. Phys. Lett.* A17 (2002), pp. 1475–1478. DOI: 10.1142/S0217732302007715. eprint: hep-ex/0202018.

[Aba+02] A. Abashian et al. "The Belle detector". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 479.1 (2002), pp. 117–232. ISSN: 0168-9002. DOI: DOI:10.1016/S0168-9002(01)02013-7. URL: http://www.sciencedirect.com/science/article/B6TJM-45BCM4R-2/2/905c8e450b275b0e557e7463d2228f12.

[Aba+95] S. Abachi et al. "Search for High Mass Top Quark Production in pp Collisions at $s = 1.8$ TeV". In: *Phys. Rev. Lett.* 74.13 (Mar. 1995), pp. 2422–2426. DOI: 10.1103/PhysRevLett.74.2422.

[Abe+01] K. Abe et al. "Observation of Large CP Violation in the Neutral B Meson System". In: *Phys. Rev. Lett.* 87.9 (Aug. 2001), p. 091802. DOI: 10.1103/PhysRevLett.87.091802.

[Abe+95] F. Abe et al. "Observation of Top Quark Production in pp Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74.14 (Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626.

[Alt+09] Wolfgang Altmannshofer et al. "New strategies for new physics search in $B \to K\nu\bar{\nu}$ and $B \to K^*\nu\bar{\nu}$ decays". In: *Journal of High Energy Physics* 2009.04 (2009), p. 022. URL: http://stacks.iop.org/1126-6708/2009/i=04/a=022.

[AS+10] P. del Amo Sanchez et al. "Search for the Rare Decay $B \to K\nu\bar{\nu}$". In: *Phys.Rev.* D82 (2010), p. 112002. DOI: 10.1103/PhysRevD.82.112002. eprint: 1009.1529.

[Aub+01] B. Aubert et al. "Observation of CP Violation in the B Meson System". In: *Phys. Rev. Lett.* 87.9 (Aug. 2001), p. 091801. DOI: 10.1103/PhysRevLett.87.091801.

[Aub+08] Bernard Aubert et al. "Search for $B \to K^*\nu\bar{\nu}$ decays". In: *Phys.Rev.* D78 (2008), p. 072007. DOI: 10.1103/PhysRevD.78.072007. eprint: 0808.1338.

[Aub+74] J. J. Aubert et al. "Experimental Observation of a Heavy Particle J". In: *Phys. Rev. Lett.* 33.23 (Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.

[Aug+74] J. E. Augustin et al. "Discovery of a Narrow Resonance in $e+e-$ Annihilation". In: *Phys. Rev. Lett.* 33.23 (Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.

[Bal+05a] Patricia Ball et al. "$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form factors from light-cone sum rules reexamined". In: *Phys. Rev. D* 71 (1 2005), p. 014029. DOI: 10.1103/PhysRevD.71.014029. URL: http://link.aps.org/doi/10.1103/PhysRevD.71.014029.

[Bal+05b] Patricia Ball et al. "New results on $B \to \pi, K, \eta$ decay form factors from light-cone sum rules". In: *Phys. Rev. D* 71 (1 2005), p. 014015. DOI: 10.1103/PhysRevD.71.014015. URL: http://link.aps.org/doi/10.1103/PhysRevD.71.014015.

[Bes+93] D Besson et al. "Upsilon Spectroscopy: Transitions in the Bottomonium System". In: *Annual Review of Nuclear and Particle Science* 43.1 (1993), pp. 333–378. DOI: 10.1146/annurev.ns.43.120193.002001. eprint: http://www.annualreviews.org/doi/pdf/10.1146/annurev.ns.43.120193.002001. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.ns.43.120193.002001.

[Bev11] Adrian J. Bevan. "The Physics of Heavy Flavours at SuperB". In: (2011). eprint: 1110.3901.

[Bhu] Bipul Bhuyan. "BN1165: High PT Tracking Efficiency Using Partially Reconstructed D^* Decays". Belle Note 1165.

[Bro+01] T.E. Browder et al. "A Search for $B \to \tau\nu$". In: *Phys.Rev.Lett.* 86 (2001), pp. 2950–2954. DOI: 10.1103/PhysRevLett.86.2950. eprint: hep-ex/0007057.

[Bru+97] Rene Brun et al. "ROOT - An Object Oriented Data Analysis Framework". In: *Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A* 389 (1997), pp. 81–86. URL: http://root.cern.ch/.

[Buc+00] Gerhard Buchalla et al. "Phenomenology of nonstandard Z couplings in exclusive semileptonic $b \to s$ transitions". In: *Phys. Rev. D* 63 (1 2000), p. 014015. DOI: 10.1103/PhysRevD.63.014015. URL: http://link.aps.org/doi/10.1103/PhysRevD.63.014015.

[Bur+10] Andrzej Buras et al. "Patterns of flavour violation in the presence of a fourth generation of quarks and leptons". In: *Journal of High Energy Physics* 2010 (9 2010). 10.1007/JHEP09(2010)106, pp. 1–74. ISSN: 1029-8479. URL: http://dx.doi.org/10.1007/JHEP09(2010)106.

[Cab63] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10.12 (June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[Che+05] K. F. Chen et al. "Time-dependent CP-violating asymmetries in $b \to s\bar{q}q$ transitions". In: *Phys. Rev.* D72 (2005), p. 012004. DOI: 10.1103/PhysRevD.72.012004. eprint: hep-ex/0504023.

[Che+07] K.-F. Chen et al. "Search for $B \to h^{(*)}\nu\nu$ Decays at Belle". In: *Phys. Rev. Lett.* 99.22 (Nov. 2007), p. 221802. DOI: 10.1103/PhysRevLett.99.221802.

[CKM05] CKMfitter Group. "CKMfitter Results". In: *Eur. Phys. J.* C41 (2005), pp. 1–131. URL: http://ckmfitter.in2p3.fr.

[Col+06] P. Colangelo et al. "Exclusive $B \to K^{(*)}\ell^+\ell^-$, $B \to K^{(*)}\nu\bar{\nu}$ and $B \to K^*\gamma$ transitions in a scenario with a single universal extra dimension". In: *Phys. Rev.* D73 (2006), p. 115006. DOI: 10.1103/PhysRevD.73.115006. eprint: hep-ph/0604029.

[Con+03] Jan Conrad et al. "Including systematic uncertainties in confidence interval construction for Poisson statistics". In: *Phys. Rev.* D67 (2003), p. 012002. DOI: 10.1103/PhysRevD.67.012002. eprint: hep-ex/0202013.

[Cow+56] C. L. Cowan Jr. et al. "Detection of the Free Neutrino: A Confirmation". In: *Science* 124 (July 1956), pp. 103–104. DOI: 10.1126/science.124.3212.103.

[Fei+06] M. Feindt et al. "The NeuroBayes neural network package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (2006). Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research - ACAT 05, pp. 190 –194. ISSN: 0168-9002. DOI: DOI:10.1016/j.nima.2005.11.166. URL: http://www.sciencedirect.com/science/article/pii/S0168900205022679.

[Fei+11] M. Feindt et al. "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (2011), pp. 432 –440. ISSN: 0168-9002. DOI: 10.1016/j.nima.2011.06.008. URL: http://www.sciencedirect.com/science/article/pii/S0168900211011193.

[Fel+98] Gary J. Feldman et al. "Unified approach to the classical statistical analysis of small signals". In: *Phys. Rev. D* 57.7 (1998), pp. 3873–3889. DOI: 10.1103/PhysRevD.57.3873.

[Fox+78] Geoffrey C. Fox et al. "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes". In: *Phys. Rev. Lett.* 41.23 (Dec. 1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581.

[Fuk+98] Y. Fukuda et al. "Evidence for Oscillation of Atmospheric Neutrinos". In: *Phys. Rev. Lett.* 81.8 (Aug. 1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562.

[Gag06] N. D. Gagunashvili. "Comparison of weighted and unweighted histograms". In: *ArXiv Physics e-prints* (May 2006). eprint: arXiv:physics/0605123.

[Gla+70] S. L. Glashow et al. "Weak Interactions with Lepton-Hadron Symmetry". In: *Phys. Rev. D* 2.7 (Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.

[GM64] M. Gell-Mann. "A schematic model of baryons and mesons". In: *Physics Letters* 8.3 (1964), pp. 214 –215. ISSN: 0031-9163. DOI: DOI:10.1016/S0031-9163(64)92001-3. URL: http://www.sciencedirect.com/science/article/pii/S003191636492001 3.

[Gro+73] D. J. Gross et al. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Physical Review Letters* 30 (June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[HAS70] W. K. HASTINGS. "Monte Carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97. eprint: http://biomet.oxfordjournals.org/content/57/1/97.full.pdf+html. URL: http://biomet.oxfordjournals.org/content/57/1/97.abstract.

[Her+77] S. W. Herb et al. "Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions". In: *Phys. Rev. Lett.* 39.5 (Aug. 1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252.

[Ika+06] K. Ikado et al. "Evidence of the Purely Leptonic Decay $B^- \to \tau^- \nu_\tau$". In: *Phys. Rev. Lett.* 97.25 (Dec. 2006), p. 251802. DOI: 10.1103/PhysRevLett.97.251802.

[Jef46] Harold Jeffreys. "An Invariant Form for the Prior Probability in Estimation Problems". English. In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 186.1007 (1946), pp. 453–461. ISSN: 00804630. URL: http://www.jstor.org/stable/97883.

[KK+06] H. V. Klapdor-Kleingrothaus et al. "The evidence for the observation of 0nu beta beta decay: The identification of 0nu beta beta events from the full spectra". In: *Mod. Phys. Lett.* A21 (2006), pp. 1547–1566. DOI: 10.1142/S0217732306020937.

[Kob+73] Makoto Kobayashi et al. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652. URL: http://ptp.ipap.jp/link?PTP/49/652/.

[Kur+03] S. Kurokawa et al. "Overview of the KEKB accelerators". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 499.1 (2003), pp. 1–7. ISSN: 0168-9002. DOI: DOI:10.1016/S0168-9002(02)01771-0. URL: http://www.sciencedirect.com/science/article/B6TJM-4753JSR-1/2/806377ce67ec41aacaaf720399be778b.

[Lan01] D.J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth.* A462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[Lee+03] S. H. Lee et al. "Evidence for $B^0 \to \pi^0\pi^0$". In: *Phys. Rev. Lett.* 91.26 (Dec. 2003), p. 261801. DOI: 10.1103/PhysRevLett.91.261801.

[Lin+] S.W. Lin et al. "Belle Note 901: Branching Ratio of B to hh". Belle Note 901.

[Liv+08] D. Liventsev et al. "Study of $B \to D^{**}l\nu$ with full reconstruction tagging". In: *Phys. Rev. D* 77.9 (May 2008), p. 091503. DOI: 10.1103/PhysRevD.77.091503.

[M.N+] M.Niiyama(RIKEN) et al. "Belle Note 779: Study of Kaon and Pion Identification Using Inclusive D^* Sample". Belle Note 779.

[Mon+09] Jose Montoya et al. "On a criticism of the profile likelihood function". In: *Statistical Papers* 50.1 (Jan. 2009), pp. 195–202–202. ISSN: 0932-5026. URL: http://dx.doi.org/10.1007/s00362-007-0056-5.

[Mur+00] S. A. Murphy et al. "On Profile Likelihood". English. In: *Journal of the American Statistical Association* 95.450 (2000), pp. 449–465. ISSN: 01621459. URL: http://www.jstor.org/stable/2669386.

[Nak+10] K Nakamura et al. "Review of Particle Physics". In: *Journal of Physics G: Nuclear and Particle Physics* 37.7A (2010), p. 075021. URL: http://stacks.iop.org/0954-3899/37/i=7A/a=075021.

[Ney37] J. Neyman. "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability". In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236.767 (1937), pp. 333 –380. DOI: 10.1098/rsta.1937.0005. URL: http://rsta.royalsocietypublishing.org/content/236/767/333.short.

[Per+75] M. L. Perl et al. "Evidence for anomalous lepton production in e^+e^- annihilation". In: *Physical Review Letters* 35 (1975), pp. 1489–1492. DOI: 10.1103/PhysRevLett.35.1489.

[Ruc+90] D.W. Ruck et al. "The multilayer perceptron as an approximation to a Bayes optimal discriminant function". In: *Neural Networks, IEEE Transactions on* 1.4 (Dec. 1990), pp. 296 –298. ISSN: 1045-9227. DOI: 10.1109/72.80266.

[Rum+86] David E. Rumelhart et al. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (Oct. 1986), pp. 533–536. URL: http://dx.doi.org/10.1038/323533a0.

[Whi] Eric White. "BN1207: Determination of K^0_S Efficiency and Systematic Uncertainty". BN1207.

[Wu+57] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105.4 (Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[Yon+07] Yongsheng et al. "Upper limit for Poisson variable incorporating systematic uncertainties by Bayesian approach". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 578.1 (2007), pp. 322 –328. ISSN: 0168-9002. DOI: 10.1016/j.nima.2007.05.116. URL: http://www.sciencedirect.com/science/article/pii/S0168900207009862.

[Yuk55] Hideki Yukawa. "On the Interaction of Elementary Particles. I *". In: *Progress of Theoretical Physics Supplement* 1 (1955), pp. 1–10. DOI: 10.1143/PTPS.1.1. URL: http://ptp.ipap.jp/link?PTPS/1/1/.

Danksagung

Ganz besonders möchte ich mich bei Prof. Dr. Michael Feindt für die erstklassige Betreuung dieser Dissertation bedanken. Die interessanten und lehrreichen Jahre werden mir immer in guter Erinnerung bleiben. Zudem bedanke ich mich herzlich bei Prof. Dr. Günter Quast für die Übernahme des Korreferats.

Die Anfertigung dieser Dissertation war nur möglich durch die hervorragende Betreuung durch Dr. Thomas Kuhr, Dr. Michal Kreps, Dr. Anže Zupanc und Dr. Martin Heck. Vielen Dank dafür.

Unterstützt wurde die Anfertigung dieser Dissertation durch ein Stipendium des Graduiertenkollegs Hochenergiephysik und Teilchenastrophysik der Deutschen Forschungsgemeinschaft, durch ein Abschlussstipendium des Karlsruhe House of Young Scientists (KHYS), durch die herausragenden Leistungen des Administratoren-Teams und durch die Hilfe der Sekretärinnen Frau Weissmann und Frau Bräunling.

Prof. Dr. Thomas Müller möchte ich in seiner Funktion als Institutsleiter dafür danken, dass er verschiedene Konferenzbesuche sowie diverse Forschungsaufenthalte ermöglichte.

Bei meinen Kollegen am Institut, insbesondere den langjährigen Kollegen Daniel Zander, Felix Wick, Simon Honc und allen *Mitbewohnern* aus Zimmer 9-2 möchte ich mich dafür bedanken, diese Zeit unvergesslich gemacht zu haben.

Bedanken möchte ich mich auch für die Unterstützung und den Rückhalt von meiner Familie, meinen Freunden und insbesondere auch von meiner Freundin Alexandra Schorer. Vielen Dank für die Kraft und auch das Verständnis dafür, wegen der Arbeit vielleicht nicht immer für euch da gewesen zu sein.