Physik-Institut

Masterthesis: analysis of $B \to K^* \mu \mu$ decay

Supervisors: Prof. Dr. Nicola Serra, Dr. Marcin Chzsaszcz

Author: Oliver Dahme

Here comes the abstract

1 $B \to K^* \mu \mu$ decay

The $B \to K^* \mu \mu$ decay has a quark $b \to s$ transition which is relevant

1.1 Kinematics

The Decay is a flavor changing neutral current (FCNC) with four charged particles in the final state: The K^+ and π^- from the K^* decay and two leptons from the loop or box digrams:
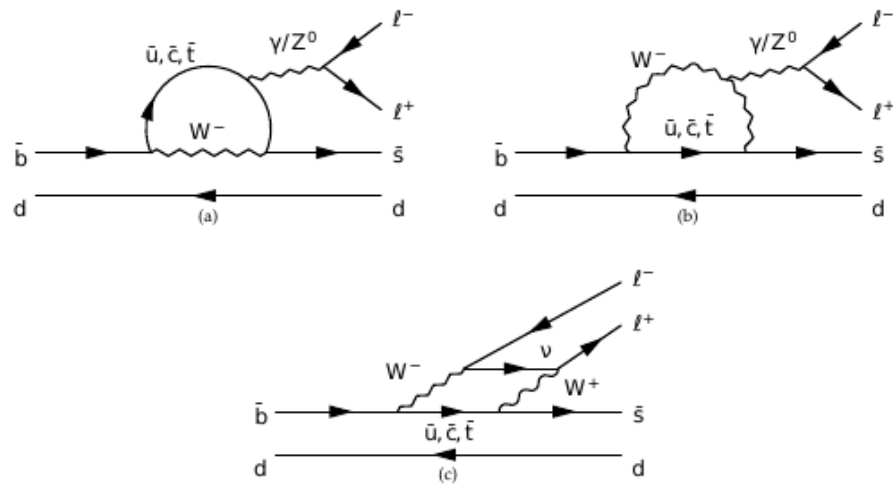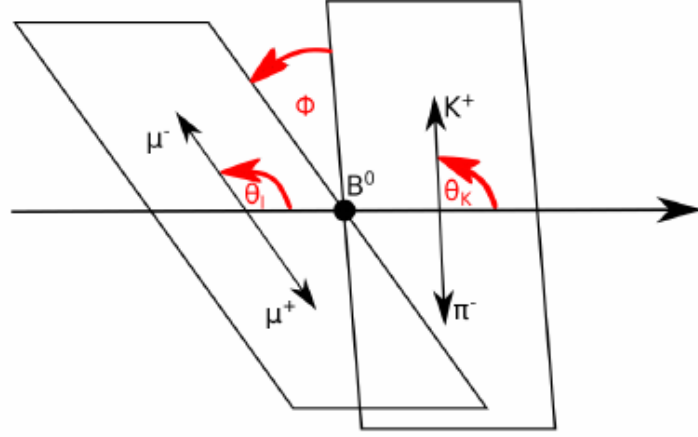

Figure 1: Feynman diagrams for decay $B_d \to \mu^+ \mu^-$ at lowest order

The kinematics of the decay are defined by the three angels θ_K, θ_L and ϕ:

Figure 2: kinematic variables of the decay $B^0 \to K^{*0} \, \mu \, \mu$

1.2 Operators for $B \to X_s l^+ l^-$ decays

The effective Lagrangian for $B \to X_s l^+ l^-$ decays has the form:

$$
\begin{aligned}
\mathcal{L}_{eff} = \; & \mathcal{L}_{QCD,QED}(u,d,s,c,b,e,\mu,\tau) \\
& + \frac{4G_F}{\sqrt{(2)}} \left[V_{us}^* V_{ub} (C_1^c P_1^u + C_2^c P_2^u) + V_{cs}^* V_{cb} (C_1^c P_1^c + C_2^c P_2^c) \right] \\
& + \frac{4G_F}{\sqrt{(2)}} \sum_{i=3}^{10} \left[(V_{us}^* V_{ub} + V_{cs}^* V_{cb}) C_i^c + V_{ts}^* V_{tb} C_i^t \right] P_i.
\end{aligned}
\tag{1}
$$

The first term in equation 1 contains the kintec terms of the light SM particles as well as their QCD and QED interactions. The remaining two terms consist of $\Delta B = -\Delta S = 1$ local operators of dimension ($d \leq 6$), which contain those light fields. The mass of the s quark can be negleted in comparisson with the b mass. One gets the following operators:

$$
\begin{aligned}
&\mathcal{O}_1^u = (\bar{s}_L \gamma_\mu T^a u_L)(\bar{u}_L \gamma^\mu T^a b_L), &&\mathcal{O}_6 = (\bar{s}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} T^a b_L) \sum_q (\bar{q} \gamma^{\mu_1} \gamma^{\mu_1} \gamma^{\mu_1} T^a q), \\
&\mathcal{O}_2^u = (\bar{s}_L \gamma_\mu u_L)(\bar{u}_L \gamma^\mu b_L), && \\
&\mathcal{O}_1^c = (\bar{s}_L \gamma_\mu T^a c_L)(\bar{c}_L \gamma^\mu T^a b_L), g &&\mathcal{O}_7 = \frac{e}{g^2} m_b (\bar{s}_L \sigma^{\mu\nu} b_R) F_{\mu\nu}, g \\
&\mathcal{O}_2^c = (\bar{s}_L \gamma_\mu c_L)(\bar{c}_L \gamma^\mu b_L), && \\
&\mathcal{O}_3 = (\bar{s}_L \gamma_\mu b_L) \sum_q (\bar{q} \gamma^\mu q), &&\mathcal{O}_8 = \frac{1}{g} m_b (\bar{s}_L \sigma^{\mu\nu} T^a b_R) G_{\mu\nu}^a, \\
&\mathcal{O}_4 = (\bar{s}_L \gamma_\mu T^a b_L) \sum_q (\bar{q} \gamma^\mu T^a q), &&\mathcal{O}_9 = \frac{e^2}{g^2} (\bar{s}_L \gamma_\mu b_L) \sum_l (\bar{l} \gamma^\mu l), \\
&\mathcal{O}_5 = (\bar{s}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} b_L) \sum_q (\bar{q} \gamma^{\mu_1} \gamma^{\mu_1} \gamma^{\mu_1} q), &&\mathcal{O}_{10} = \frac{e^2}{g^2} (\bar{s}_L \gamma_\mu b_L) \sum_l (\bar{l} \gamma^\mu \gamma_5 l),
\end{aligned}
\tag{2}
$$

where $\sum_q$ and $\sum_l$ denote the sums over light quarks and all leptons, respectivly.

1.3 Wilson coefficients

The calculation of decay amplitude of the $B \to K^* \mu \mu$ decay requires three different steps:

1. The speration of short-distance effects from long-sistance QCD in an effetive Hamiltonian H_{eff}
2. The calculation of matrix elements of local quark bilinear operators J of type $\langle K^*|J|B \rangle$
3. The claculation of effects of 4-quark operators in H_{eff} which give rise to non factorizable corrections and can be calculated using QCD factorization.

QCD factorization is only valid for small invariant dilepton mass $q^2 \sim \mathcal{O}$ (1 GeV^2), or large K^* energy $E \sim \mathcal{O}(m_B/2)$, which implies certain cuts on q^2 or E. The effective Hamiltonian for $b \to s\mu^+\mu^-$ transitions can be written as:

$$H_{eff} = -\frac{4G_F}{\sqrt{2}} \left(\lambda_t H_{eff}{}^{(t)} + \lambda_u H_{eff}^{(}u) \right) \tag{3}$$

The λ_i can be expressed with CKM combinations $\lambda_i = V_{ib}V_{is}^*$.

$$H_{eff}^{(t)} = C_1 \mathcal{O}_1^c + C_2 \mathcal{O}_2^C + \sum_{i=3}^{6} C_i \mathcal{O}_i + \sum_{i=7,8,9,10,P,S} (C_i \mathcal{O}_i + C_i' \mathcal{O}_i') \tag{4}$$

$$H_{eff}^{(u)} = C_1(\mathcal{O}_1^C - \mathcal{O}_1^u) + C_2(\mathcal{O}_2^C - \mathcal{O}_2^u). \tag{5}$$

The contribution of $H_{eff}^{(u)}$ has a double Cabibbo supression and is therfore usually dropped. It is kept here since it is sensitive to complex phases of decay amplitudes. The operators $P_{i \leq 6}$ are the same as for general $B \to X_s l^+ l^-$ decays, see equation 2. The remaining ones are given by:

$$
\begin{aligned}
\mathcal{O}_7 &= \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu}P_R b)F^{\mu\nu}, & \mathcal{O}_7' &= \frac{e}{g^2} m_b (\bar{s}\sigma_{\mu\nu}P_L b)F^{\mu\nu}, \\
\mathcal{O}_8 &= \frac{1}{g} m_b (\bar{s}\sigma_{\mu\nu}T^a P_R b)G^{\mu\nu a}, & \mathcal{O}_8' &= \frac{1}{g} m_b (\bar{s}\sigma_{\mu\nu}T^a P_L b)G^{\mu\nu a}, \\
\mathcal{O}_9 &= \frac{e^2}{g^2} (\bar{s}\sigma_{\mu}P_L b)(\bar{\mu}\gamma^{\mu}\mu), & \mathcal{O}_9' &= \frac{e^2}{g^2} (\bar{s}\gamma_{\mu}P_R b)(\bar{\mu}\gamma^{\mu}\mu), \\
\mathcal{O}_{10} &= \frac{e^2}{g^2} (\bar{s}\gamma_m u P_L b)(\bar{\mu}\gamma^{\mu}\gamma_5\mu), & \mathcal{O}_{10}' &= \frac{e^2}{g^2} (\bar{s}\gamma_{\mu}P_R b)(\bar{\mu}\gamma^{\mu}\gamma_5\mu), \\
\mathcal{O}_S &= \frac{e^2}{16\pi^2} m_b (\bar{s}P_R b)(\bar{\mu}\mu), & \mathcal{O}_S' &= \frac{e^2}{16\pi^2} m_b (\bar{s}P_L b)(\bar{\mu}\mu), \\
\mathcal{O}_P &= \frac{e^2}{16\pi^2} m_b (\bar{s}P_R b)(\bar{\mu}\gamma_5\mu), & \mathcal{O}_P' &= \frac{e^2}{16\pi^2} m_b (\bar{s}P_L b)(\bar{\mu}\gamma_5\mu),
\end{aligned}
\tag{6}
$$

where m_b denotes the running b mass in the $\overline{MS}$ scheme and g is the strong coupling constant and $P_{L,R} = (1 \pm \gamma_5)/2$. In the Standart Modell the primed Operators with opposite chirality to the unprimed operators vanish or are highly suppresd as are the $\mathcal{O}_S$ and $\mathcal{O}_P$. The contributions of $\mathcal{O}_{1,2,3,4,5,6}$ are neglected, since they are either heavily constrained or their impact turns out to be generically very small. For example in the left-right symmetric models or throughout gluino contributions in a general Minimal Supersymmetric Standard Model.

The C_i coefficients in the equations 4 and 5 are called Wilson coefficients. They encode short-distance physics and New Physics effects. For the calculation a matching scale $\mu = m_W$ is chosen, in a pertubative expansion in powers of $\alpha_s(m_W)$. Then the Wilson coefficents are evolved down to scales $\mu = m_b$ according to the solutions of the renomalization group equations. Contributions by New Physics enter

through $C_i(m_W)$, while the low scales are determined by the Standart Modell. To allow a more organized expansion of the Wilson coefficients in pertubation theory the factors $16\pi^2/g^2 = 4\pi/\alpha_S$ are included into the definitions of the operators $\mathcal{O}_{i \geq 7}$. All the C_i expand as:

$$C_i = C_i^{(0)} + \frac{\alpha_s}{4\pi} C_i^{(1)} + \left(\frac{\alpha_s}{4\pi}\right)^2 C_i^{(2)} + O(\alpha_s^3) \tag{7}$$

where $C_i^{(0)}$ is the tree-level contribution, which is quale to zero for all operators except $\mathcal{O}_2$ and $C_i^{(n)}$ denotes the n-loop contributions.

1.4 Differential Decay Distribution

In the experiments the observed decay is not $B \to K^*\mu\mu$, but $B \to K^*(K\pi)\mu\mu$. This provides addition information on the polarization of the K^* since the decay angle between K and π is senitiv to it. The matrix elemt of the effetive Hamiltonian (equation 3) for this decay can be written as:

$$
\begin{aligned}
\mathcal{M} = \frac{G_F \alpha}{\sqrt{2}\pi} V_{tb} V_{ts}^* \Bigg\{ &\bigg[\langle K\pi | \bar{s}\gamma^\mu (C_9^{eff} P_L + C_9^{eff} P_R) b | \bar{B} \rangle \\
&- \frac{2m_b}{q^2} \langle K\pi | \bar{s} i\sigma^{\mu\nu} q_\nu (C_7^{eff} P_L) b | \bar{b} \rangle \bigg] (\bar{\mu}\gamma_\mu \mu) \\
&+ \langle K\pi | \bar{s}\gamma^\mu (C_{10}^{eff} P_L + C_{10}^{eff} P_R) b \bar{B} \rangle (\bar{\mu}\gamma_5 \mu) \\
&+ \langle K\pi | \bar{s} (C_S P_R + C_S' P_L) b | \bar{B} \rangle (\bar{\mu}\mu) \\
&+ K\pi | \bar{s} (C_P P_R + C_P' P_L) b | \bar{B} (\bar{\mu}\gamma_5 \mu) \Bigg\}.
\end{aligned}
\tag{8}
$$

2 The LHCb Experiment

The Large Hadron Collider beauty experiment (LHCb) is one of four large experiments based at the CERN laboratory near Geneva in Switzerland. It is part of the Large Hadron Collider (LHC), a proton-proton accelerator and collider located in a vast unterground tunnel with 26.7 km circumference beneath the Swiss-French countryside. The other three experiments are CMS and ATLAS which are dedicated to a wide range of physics and have therefor very large detectors. ALICE investigates quark-gluon plasma and therefor needs geavy ion collisions, instead of proton collisions.

The protons in the LHC have a kinetic energy of 7 TeV, which allows a collision energy, in the LHCb detector, of 13 TeV. In the year 2016 the LHCb had a recorded luminosity of $1906\,\mathrm{pb}^{-1}$. For this thesis $2280\,\mathrm{pb}^{-1}$ of data, collected at LHCb during the years 2011 to 2016 are used. LHCb is dedicated to falvour physics. It therefor investigates rare decays and CP violation in beauty and charm hadrons.
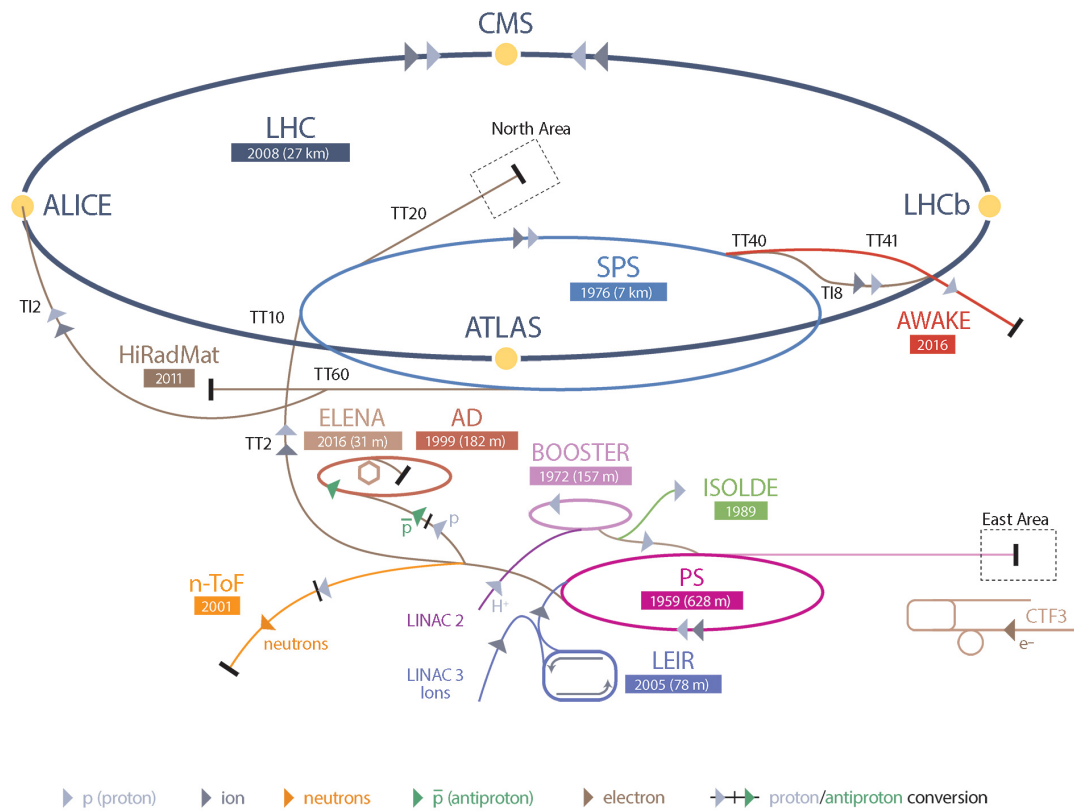
CERN's Accelerator Complex

Figure 3: CERN's Accelerator Complex.[1] The protons get injected in the lineare accelerator LINAC2. Then they get pre-accelerated in 3 synchrotons (BOOSTER,PS,SPS) where the protons reach a kinetic energy of 450 GeV. That is the entering energy of the LHC which accelerates them futher up to 7 TeV, before they collide at the four detectors: CMS, ATLAS, LHCb and ALICE.

2.0.1 The LHCb Detector

The LHCb Detector has a fix target geometry, because beauty hadrons are manily produced at small angles with respect to the beam pipe.
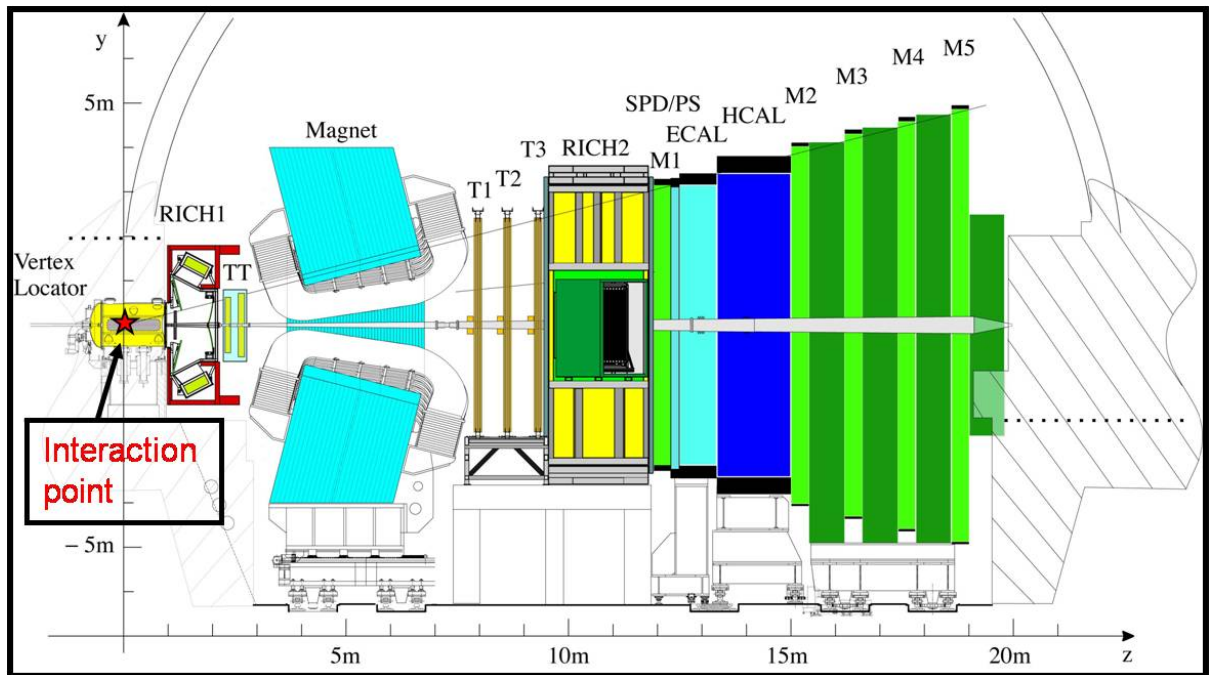
Figure 4: Basic layout of the LHCb detector [2]. The interaction point is inside the vertrex detector and the beam pipe passes through the center. The different subdetecors are the two Ring Imaging Cherenkov Detecors (RICH1 and RICH2), the tracking stations (TT and T1 to T3), the scintillator pad detector (SPD), the preshower electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the muon stations (M1-M5).

VErtex LOcator (VELO) [3] : Velo picks out B mesons from the multitude of other particles produced. This is a complex task since B mesons have very short livetimes spent colse to the beam. The VELO's silicon detecor elements must be placed at a distance of just five milimetres to the interaction point. To prevent damage to the detector during beam injection and stabilization it is mechanically moved to a safe distance. Velo measures B mesons indirectly be detecting its decay particles, nevertheless it has a resolution of 10 microns. .

Ring Imaging Cherenkov (RICH) detectors [4] : The RICH detectors meeasure the emission of Cherenkov radiation, which happens when a charged particle passes through a medium faster than light does. It is a similar effect like the sonic boom a aircraft produces by breaking the sound barrier. The shape of the light cone depends on the particle's velocity, enabling the detector to determine its speed.

Magnet [5] : The big magnet of the LHCb experiment weights 27 tonnes ands is mounted inside a 1,450 tonne steel frame. This powerful magnet forces all charged particels to change there trajectory. By examining the curvature of the path one can calculate its momentum.

Trackers [6] : The LHCb's tracking system consists of a series of four large rectangular stations, each covering an area of $40 \, \text{m}^2$. While flying through this area charged particles will leave a trace, therefor one can estimate the trajectory of a particle. The trajectory is used to link the signals left in other detecor elements to the corresponding particle. In LHCb two different tracker technologies are used: The silicon tracker placed close to the beam pipe, uses silicon microstripes. If a charged particle passes such a stripe it collides with the silicon atoms, liverating electrons and creating an electric current, which is then recorded. The outer tracker situated further from the beam pipe consists of gas-filled tubes. The gas ionizes when a charged particle hits a gas molecules, producing electrons. These reach an anode wire situated in the centre of each tube. The position of the track is found by timing how long it takes

electrons to reach it.

Calorimeters [7] : Calorimeters stop particles as they pass through, measuring the amount of energy lost. In LHCb there are two different types: The elctromagnetic calorimeter responsible for light particles like electrons and photons and the hadronic calorimeter responsible for heavier particles containing quarks. Both have a sandwich-like structure, with alternating layers of metal and plastic plates. If a particles hits a metal plate it produces a shower of secondary particles. These will excite polystyrenne molecules in the plastic plates, which then emit ultraviolet light. The energy lost by the particle in the metal plate is propoertional to the amount of UV light produced in the platic plates.

Muon System [8] : The muon system consistes of 5 rectangular stations, which cover an area of $435\,\mathrm{m}^2$. Each station has chambers filled with three gases: carbon dioxide, argon and tetrafluoromethane. Passing muons react with the mixture and electrodes detect the result.

2.0.2 The LHCb trigger system

The rate of events at the LHCb interaction point is $40\,\mathrm{MHz}$. But the rate to have a B meson contained in the detector is $15\,\mathrm{kHz}$. But the offline computing power just allows $2\,\mathrm{kHz}$ to be recorded. The LHCb trigger system aims to 'fill' this $2\,\mathrm{kHz}$ with intresting B decays and important control decays like J/ψ decays. The trigger has two levels:

The **Level Zero (L0)** trigger reduces the beginning $40\,\mathrm{MHz}$ to $1\,\mathrm{MHz}$. To get this high rate it can only rely on fast sub-detectors as the calorimeters and the muon system. The L0 trigger looks for events with high transverse momentum with respect to the patrticle beam axis (pT), because particles from a B decay have this attribute, since B Mesons are always produced almost parallel to the beam axis. In addition the L0 trigger performs a simplified vertex reconstruction with the signal of two silicon layers of the VELO to identifie events with multple proton-proton collisions. They are rejected because for this kind of events its much more difiicult to reconstruct B meson decays, since it is harder to distinguish primary and secondary vertex of the B decay.

The **High Level Trigger (HTL)** is an algorithm that runs on a farm of 1000 16-core computers. It has two stages: HLT1 which reduces the event rate to a few tens of kHz and HLT2 which reduces the rate to the $2\,\mathrm{kHz}$ which are recorded. HLT1 gets all the candidates of the L0 trigger and uses the full detector information on them to search for particles with a high impact parameter with respect the proton-proton collision. These particles are most likly decay products from B mesons, because of its relatively long life-time. They typically fly 1 cm away from the collision before decaying resulting in a high impact parameter for the decay products. HLT2 does a complete reconstruction of the events. It starts with the track of the VELO and connects them to the tracks in the other sub-detectors. Most important are displaced vertices, since they are strong indicator for B decays. The selection is devided into two parts. The inclusive selection searches for resonance decays like D^* or J/ψ. The exclusive selection is desigened to provide the highest possible efficiency to fully reconstruct B decays of interest. It therfor uses all information available such as mass and vertex quality and intermediate resonances.

3 Analysis

3.1 Classification

3.1.1 Introduction

To eliminate combinatorial background Machine Learning algorithms called classifiers are used. They can seperate data into two or more parts, depending on certain parameters.For example sperating colors depending on their share of magenta, cian and yellow. But first they need to be trained on so called labeled data. In a labeled data set it is known which entrie corresponds to which category, like colors. While training the algorithm will learn so seperate data into the pre defined categories. That is the reason it is called Machine learning. In this thesis the classfication should seperate signal from combinatorial background.

To do so Monte Carlo data which contains only $B \to K^* \, \mu \, \mu$ decays is labeled with probability 1 to be signal. Then it is merged with real data and used to train the classifiers. But since the classification becomes naturally biased if the data to classify is the same as the training data, a technique called K-folding is used. K-folding seperates the data Monte Carlo mix into several parts called folds. To classify one fold all the other folds are used for training. After iterating over all the folds one has a complety classified data set without any bias.

3.1.2 classifiers test

In this thesis the classifiers themselves are used as a black box and are not explained further. First the following list of classifers where tested and compared in terms of perfomance:

– Ada Boost
– uGB + knnAda
– uBoost
– uGB + Fl
– xgb
– sk_bdtg
– sk_bdt

The test was performed with 30000 events from the 2016 LHCb $B \to K^* \, \mu \, \mu$ data and 10000 events from the Monte Carlo simulation. The following list of parameters are used in brackets are the names of the parameters in the root files.

– Decay vertex location for reconstructed particles (ENDVERTEX)
– Primary vertex location (OWNPV)
– Impact parameter (IP_OWNPV)
– Flight distance (FD_OWNPV)
– The cosine of the angle between primary vertex and decay vertex and recorded momentum (DIRA_OWNPV)

To compare the different classifiers the ROC curves and the correlations to the kinematic variabels of the decay (see 2) are used. First lets inspect the ROC curve. Since the curves for the different folds all look alike only one is presented here 5. One can find the other nine curves in the appendix **??**. It turns

out that all the classifiers classify the data correctly with just some minor variances. The ROC curve is in that case not a good tool to compare the different classifiers.
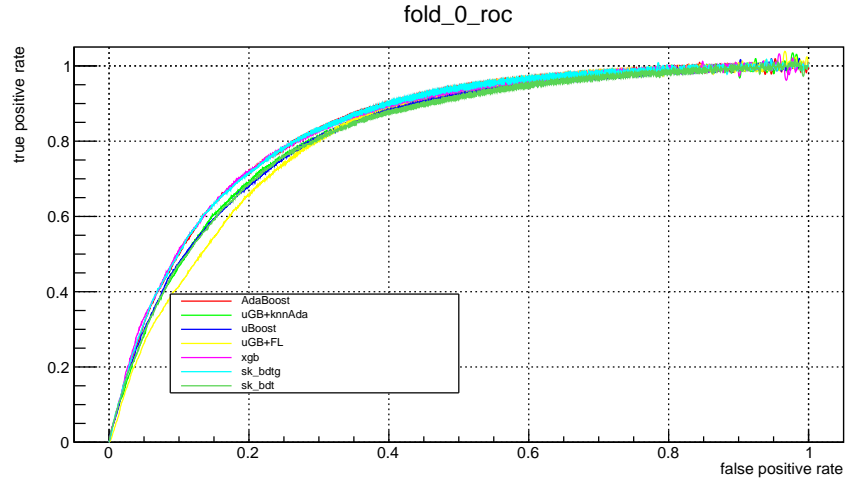

Figure 5: ROC curve of the first fold. One can see that all the classifiers are competetive in terms of classifing correctly

The next step is to check for unwanted correlations, only the correlations with the mass are shown here. The others can be found in appendix **??**.
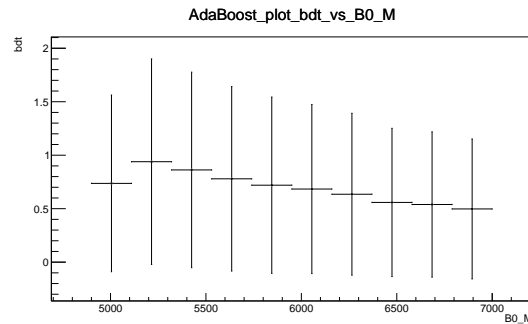

Figure 6: This plot shows the correltation between the bdt decision of the AdaBoost classifier and the B_0 mass.
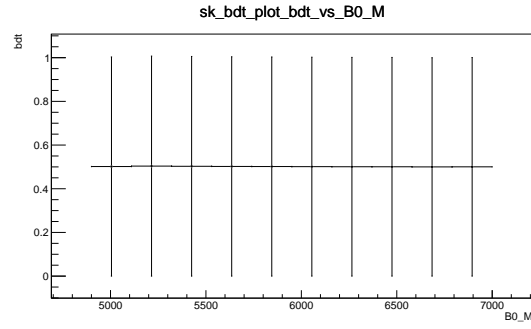
Figure 7: This plot shows the correltation between the bdt decision of the sk_bdt classifier and the B_0 mass.
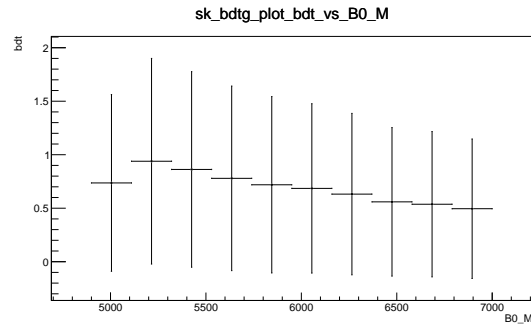
Figure 8: This plot shows the correltation between the bdt decision of the sk_bdtg classifier and the B_0 mass.
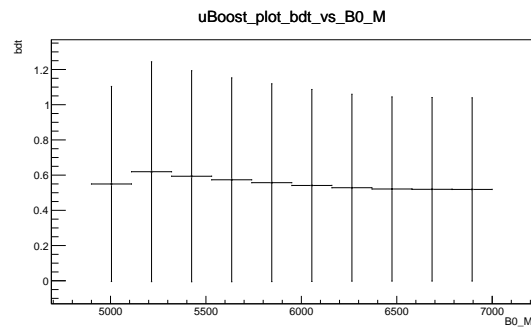
Figure 9: This plot shows the correltation between the bdt decision of the uBoost classifier and the B_0 mass.
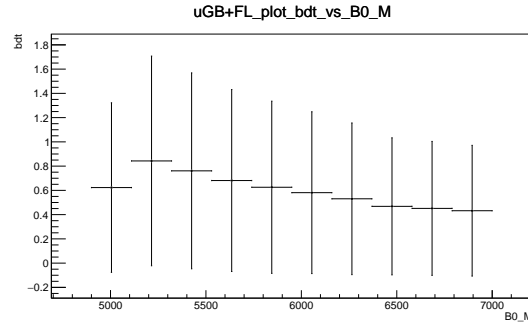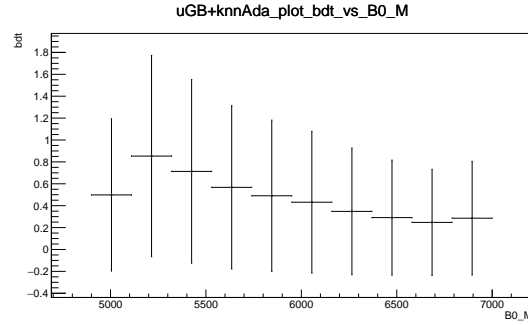
Figure 10: This plot shows the correltation between the bdt decision of the uGB+FL classifier and the B_0 mass.

Figure 11: This plot shows the correltation between the bdt decision of the uGB+knnAda classifier and the B_0 mass.
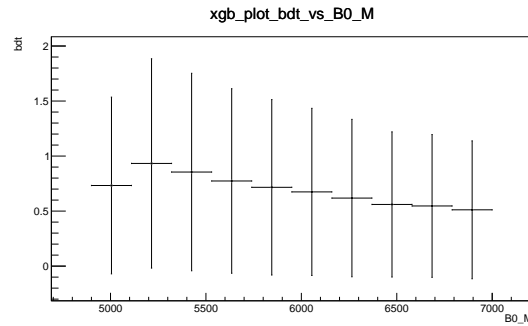

Figure 12: This plot shows the correltation between the bdt decision of the xgb classifier and the B_0 mass.

The two classifers with the least correlation are the sk_bdt and the uBoost classifers. Therfore they were used to do the seperation between signal and background in the $B \rightarrow K^* \mu \mu$ data.

3.2 SPlot

3.2.1 Likelihood method

Consider an analysis of a data sample, which consists of several types of events. These types represent signal components and background components, for example from different experiments. The log-liklihood of such a data sample is expressed as:

$$L = \sum_{i=1}^{N} \ln \left[\sum_{j=1}^{N_S} N_j f_j(y_i) \right] - \sum_{i=j}^{N_S} N_j \tag{9}$$

– N = total number of events

– N_S = number of types

– N_i = expected average number of events for type i

– y = set of diciminating variables

– f_j = PDF of the ith type

– $f_j(y_i)$ = value of PDF for event y_i

– x = control variable, not a part of L by construction

The yields N_i and the free parameters of the PDF are obtained by maximizing the above log-likelihood (eq 9).

3.2.2 $_{in}Plot$ technique

Consider a varibale x which can be expressed as a function of the dicriminating variables y used in the fit. Furthermore a fit has been performed to determine the yields N_i for all types. From the knowledge of the PDF and the values of N_i a naive weight can be defined as:

$$P_n(y_i) = \frac{N_n f_n(y_i)}{\sum_{k=1}^{N_s} N_k f_k(y_i)} \tag{10}$$

which will leed to the x-distribution $\tilde{M}_n$ defined by:

$$N_n \tilde{M}_n(\bar{x}) = \sum_{i \subset \delta x} P_n(y_i) \tag{11}$$

where sum $\sum_{i \subset \delta x}$ contains alle events for which x_i lies in the interval centered on $\bar{x}$ and of total width δx. Therefor $N_n \tilde{M}_n(\bar{x}) \delta x$ is the x-distribution of the histogrammed events, using the weights of eq 10. With this procedure one can on average reproduce the true distribution $\mathbf{M}_n(x)$. One can even replace the sum in eq 11 by an integral:

$$\left\langle \sum_{i \subset \delta x} \right\rangle \to \int dy \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) \delta x \tag{12}$$

Furthermore through identifying the number of events N_i from the fit one gets:

$$\langle N_n \rangle \tilde{M}_n(\bar{x}) = \int dy \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) P_n(y) \tag{13}$$

$$= \int dy \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) \cdot \frac{N_n f_n(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \tag{14}$$

$$= N_n \int dy \delta(x(y) - \bar{x}) f_n(y) \tag{15}$$

$$= N_n \mathbf{M}_n(\bar{x}) \tag{16}$$

One can see that the sum over enents of the daive weight P_n provides a direct estimate of the x-distribution for the nth type. But this procedure has a major drawback, since x is correlated to y, the PDFs of x enter implicity in the definition of the naive weight. Therfor the $\tilde{M}_n$ distributions are a bad estimate for the quality of the fit, since the these distributions are biased in a difficult way, when the PDFs $f_i(y)$ are not accurate.

Consider for example a data sample where one of the types has events on the tail of the x-distribution. Such events require the true ditribution to account for the tail. But since the events are averaged the weights on the tail are going to be very small missing those events in the extimated true distribution. Only the core of the x-distribution can be examined with $_{in}Plots$.

3.2.3 $_sPlot$ technique

In the previous section it was shown that if a variable x belongs to a set y of discriminating varibales, one can reconstruct the expected x distribution. Consider now two sets of variables x and y, where x does not belong to y and which are uncorrelated , hence the total PDFs $f_i(x, y)$ all factorize into products $\mathbf{M}_i(x) f_i(y)$. The equation 16 does not hold anymore because, when summing over the events the x-PDFs $\mathbf{M}_j(x)$ appear:

$$\langle N_n \rangle \tilde{M}_n(\bar{x}) = \int \int dy dx \sum_{j=1}^{N_s} N_j \mathbf{M}_j(x) f_j(y) \delta(x - \bar{x}) P_n \tag{17}$$

$$= \int dy \sum_{j=1}^{N_s} N_j \mathbf{M}_j(\bar{x}) f_j(y) \frac{N_n f_n(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \tag{18}$$

$$= N_n \sum_{j=1}^{N_s} \mathbf{M}_j(\bar{x}) \left(N_j \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \right) \tag{19}$$

$$\neq N_n \mathbf{M}_n(\bar{x}). \tag{20}$$

The correction term

$$N_j \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \tag{21}$$

is not identical to the kroenecker delta δ_{jn}. In fact the $N_n \tilde{M}_n$ distribution obtained by the naive weight is a linear combination of the true distribution $\mathbf{M}_j$.

To go forward one has to realize that the correction term is related to the inverse of the covariance matrix, given by the second derivatives of $-L$, after the minimization.

$$\mathbf{V}_{nj}^{-1} = \frac{\partial^2(-L)}{\partial N_n \partial N_j} = \sum_{i=1}^{N} \frac{f_n(y_i)f_j(y_i)}{\left(\sum_{k=1}^{N_s} N_k f_k(y_i)\right)^2} \tag{22}$$

If one averages and is replacing the sum over events by intergals (eq 12) the varaince matrix reads:

$$\langle \mathbf{V}_{nj}^{-1} \rangle = \int \int dy dx \sum_{e=1}^{N_s} N_e \mathbf{M}_e(x) f_e(y) \frac{f_n(y)f_j(y)}{\left(\sum_{k=1}^{N_s} N_k f_k(y)\right)^2} \tag{23}$$

$$= \int dy \sum_{e=1}^{N_s} N_e f_e(y) \frac{f_n(y)f_j(y)}{\left(\sum_{k=1}^{N_s} N_k f_k(y)\right)^2} \cdot \int dx \mathbf{M}_l(x) \tag{24}$$

$$= \int dy \frac{f_n(y)f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \tag{25}$$

Therefor equation 17 can be rewritten as:

$$\langle \tilde{M}_n(\bar{x}) \rangle = \sum_{j=1}^{N_s} \mathbf{M}_j(\bar{x}) N_j \langle \mathbf{V}_{nj}^{-1} \rangle. \tag{26}$$

To get the distribution of intrest one has to invert this matrix equation:

$$N_n \mathbf{M}_n(\bar{x}) = \sum_{j=1}^{N_s} \langle \mathbf{V}_{nj} \rangle \langle \tilde{M}_j(\bar{x}) \rangle \tag{27}$$

The true distribution of x can still be reconstructed using the naive weight (eq 10), through a linear combination of $_{in}Plots$. In other words: When x does not belong to the set y, the weights are not given by equation 10, they are given by a covariance-weighted quantity called sWeight defined by:

$$_s P_n(y_i) = \frac{\sum_{j=1}^{N_s} \mathbf{V}_{nj} f_j(y_i)}{\sum_{k=1}^{N_s} N_k f_k(y_i)} \tag{28}$$

With the sWeights on can obtain the distribution of the x variable by histogramming the $_s Plot$:

$$N_{ns} \tilde{M}_n(\bar{x}) \delta x = \sum_{i \subset \delta x} {}_s P_n(y_i) \tag{29}$$

On average it reproduced the true distribution:

$$\langle N_{ns} \tilde{M}_n(x) \rangle = N_n \mathbf{M}_n(x) \tag{30}$$

In the case were x is significantly correlated with y, the $_s Plots$ from equation 29 connt be compared with the pure distributions of the various types. To solve that problem one can perform a Monte Carlo simulation of the procedure and obtain the expected distributions to which the $_s Plots$ should be compared with.

For more information on $_s Plots$ consider reading [9].

3.3 Reweighting

The infomation needed in an analysis is not all contained inside the data. For example efficiencies can not be recorded by a detector. But they can be simulated by a Monte Carlo simulation of the particle

decay. The problem is that the simulation has to be the same as data to get the correct answers. Since resimulating everything with different parameters until one hits the data distribution would take too much resources. One can simply reweight the simulation to match the data. As an initial weight the sWeights see equation 28 are used. The parameters used to reweight the MC samples are applied in the following order: 'nTracks', 'B_0 p_T' and the quality of the $K\pi\mu\mu$ vertex. The new weights derived by the difference in data and simulation of these parameters are used to weight all the MC samples. For this analysis the MC and Data of the $K^* \rightarrow J/\Psi K^{*0}$ are used, because it is a very clean channel.

4 RooMCMarkovChain

RooMCMarkovChain is a new class for the ROOT Data Analysis Framework, which can be used to for fits and presents a alternative to the brodly used MINUIT algorithm. RooMCMarkovChain uses a metropolis algorithm as a minimizer for the negative log likelihood function. The metropolis algorithm is based ona Monte Carlo Markov Chain and can therfore easily be scaled to multidimensional parameter space and moreover, such kind of algorithms can easily be parallelized.

4.1 Monte Carlo Markov Chain (MCMC)

A Markov Chain is a random process which undergoes several states. From each state there is a probability ditribution to change into another state or to stay. Most important i the asumption that every next step just depends on the current state. The figure 13 illustrates the behavior of such a Markov Chain.
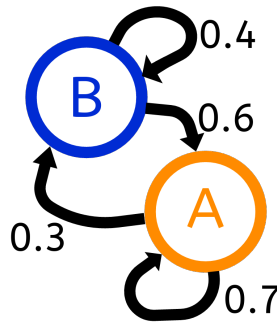

Figure 13: Illustration of states and the probability distribution to change or the stay in a state of a Monte Carlo Markov Chain. In state A there is a 0.3 probability to change into state B and a 0.7 probability to stay in state A.

4.2 Metropolis Hastings

Metropolis Hastings is a so called Monte Carlo updater. It updates the probabillity distribution when a change of states is proposed. That is very usefull in the cases where the states are not discrete like

in figure 13 but continous. Suppose that a specified distribution has unnormalized density h. The Metopolis Hastings update does the following:

1. The current state x is proposed to move to another state y with a conditional probability given x denoted as $q(x, \cdot)$.
2. The Hastings ratio is calculated:

$$r(x, y) = \frac{h(y) \cdot q(y, x)}{h(x) \cdot q(x, y)} \tag{31}$$

3. The proposed move to y is accepted with the probability $a(x, y)$:

$$a(x, y) = min(1, r(x, y)) \tag{32}$$

This principle is used in the RooMCMarkovChain class to minimize the negative log likelihood function. In detail a robust adaptive Metrpolis hastings process is implemented. It is defined as:

1. Compute next proposed state Y_n as a random change from the current state X_{n-1}:

$$Y_n \equiv X_{n-1} + S_{n-1} U_n$$

, where U_n is an independent random vector.
2. With the probability α_n the proposal is accepted and $X_n = Y_n$. Otherwise the proposal is rejected and $X_n = X_{n-1}$

$$\alpha_n \equiv min\{1, r(X_{n-1}, Y_n)\}$$

, where r is the hastings ratio see equation 31.
3. compute the lower-diagonal matrix S_n with positive diagonal elements satisfying the equation:

$$S_n S_n^T = S_{n-1} \left(I + \eta_n (\alpha_n - \alpha_*) \frac{U_n Y_n^T}{||U_n||^2} \right) S_{n-1}^T \tag{33}$$

,where $I \in R^{dxd}$ is the identity matrix.

4.3 Features

The features of the RooMCMarkovChain class will be shown by fitting the following probability density function (pdf):

$$g(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + f \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \tag{34}$$

It is so called double gaus pdf, which is just the sum of two gaussian pdfs with a fractional parameter f. The result of the fit is compared with the result of the Minuit algorithm:
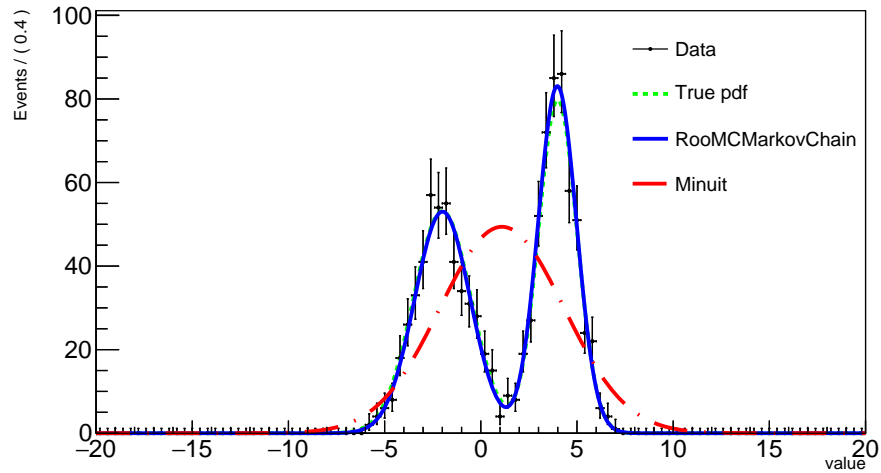
Figure 14: Fit of double gaus pdf (see equation 34) with RooMCMCarkovchain and with Minuit. The true pdf has the follwing values: $\mu_1 = 4$, $\sigma_1 = 1$, $\mu_2 = -2$, $\sigma_2 = 1.5$ and the fraction $f = 0.5$

The corresponding terminal output gives the estimated parameters with error interval and correlation coefficents of each parameter pair.

Figure 15: Terminal output of the RooMCMarkovChain fit.

This gives a good initial overview of results after the fit has finished. The error calculation can be set to assume gaussian or non-gaussian errors.

In addition several other properties of the parameters can be obtained:

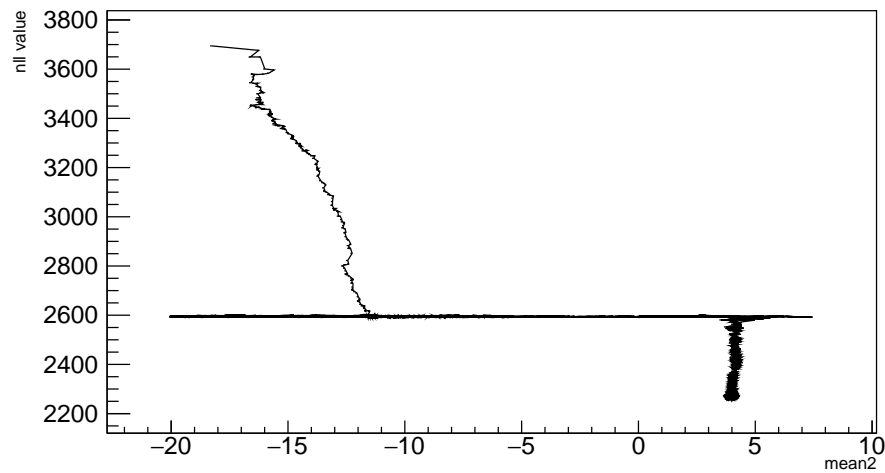1. The 1 dimensional profile of the negative log liklihood function for a given parameter.

Figure 16: Profile of the negative log lilihood function for μ_2 in equation 34. There is a local minimum at the nll value of 2600.

From this plot on can see how the algorithm reached the minimum of the negative log likelihood function and if there are local minima.

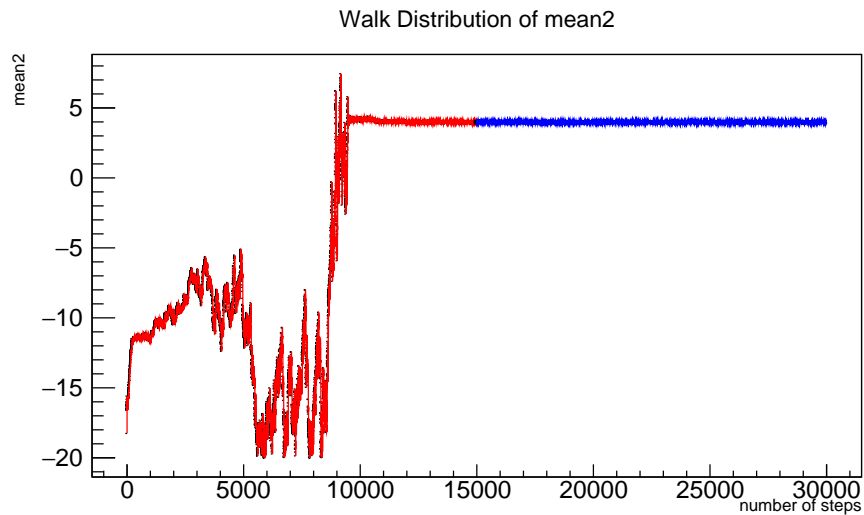2. The walk distribution of a given parameter.

Figure 17: Walk distribution of μ_2 in equation 34. The red points are not considered for the error calculation.

This plot is very important for handling the RooMCMarkovChain class. Since the error calculation is based on the variance of the walk distribution, cutting of points in the begginning greatly reduces this variance. The user has to choose how many points are cut off. Plotting the walk distribution of all parameters helps choosing the right amount.

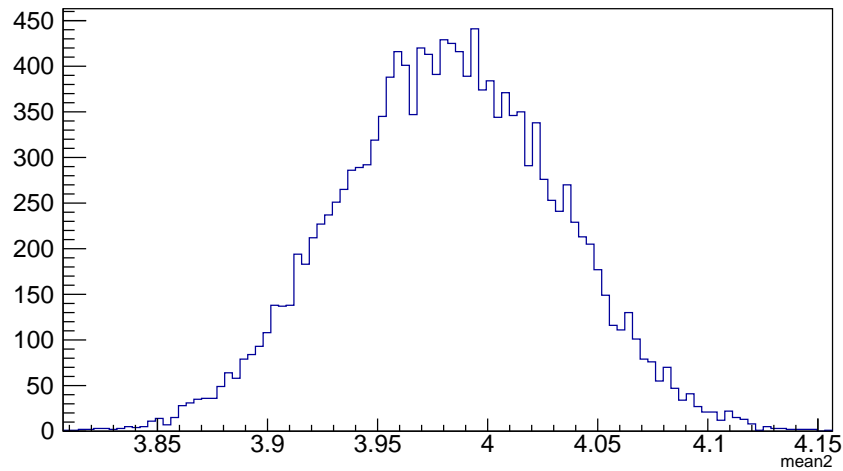3. The walk distribution of a given parameter as a histogramm.

Figure 18: Walk distribution of μ_2 in equation 34 as a histogramm.

This plot can be used to check, which error strategy should be used. If the walk distribution histogramm is gaussian, gaussian errors can be assumed.
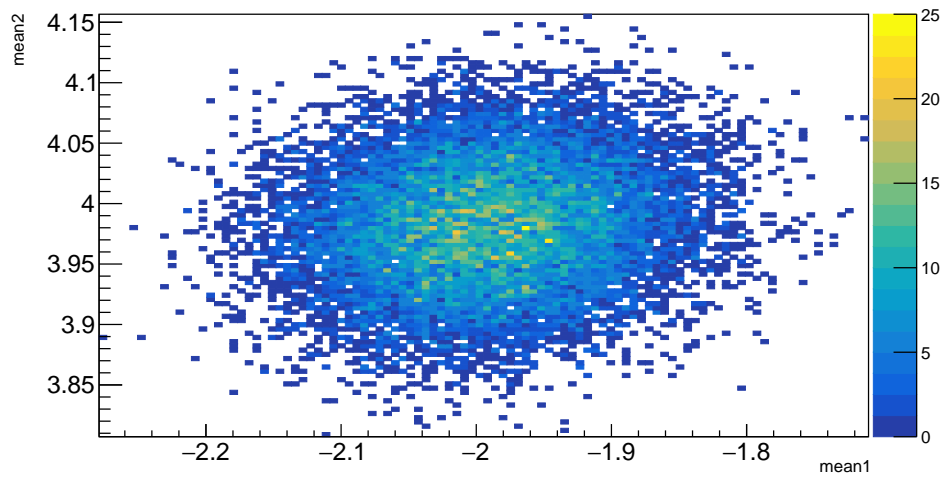
4. The scatterplot between two given parameters.

Figure 19: Scatterplot between μ_1 and μ_2 in equation 34.

This plot shows the correlation between two parameters graphically.

References

[1] CERN Accelerator Complex, *http://www.stfc.ac.uk/research/particle-physics-and-particle-astrophysics/ large-hadron-collider/cern-accelerator-complex/*

[2] Science and Technology Facilities Council article about LHCb , *https://www.ppd.stfc.ac.uk/Pages/LHCb.aspx*

[3] VELO description, *http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/VELO2-en.html*

[4] RICH description, *http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/RICH2-en.html*

[5] Magnet description, *http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Magnet2-en.html*

[6] Tracker description, *http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Trackers2-en.html*

[7] Calorimeters description, *http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Calorimeters2-en.html*

[8] Muon system description, *http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Muon2-en.html*

[9] sPlot: a statistical tool to unfold data distributions, arXiv:physics/0402083 [physics.data-an]