ETH Institute for
Particle Physics

ETHZ-IPP Internal Report 2008-10
August 2008

The Search for $B_s^0 \to \mu\tau$ at CMS

DIPLOMA THESIS

presented by
Michel De Cian

under the supervision of
Prof. Dr. Urs Langenegger

ETH Zürich, Switzerland

Abstract

Using the technique called neutrino reconstruction, a method is presented to analyze the decay $B_s^0 \to \mu\tau$, where $\tau \to \pi\pi\pi\nu_\tau$, in the environment of the CMS detector at the LHC at CERN. Several models with physics beyond the Standard Model are discussed for a motivation of Lepton Flavor Violation within reach of current detectors. The actual event chain from the data production with Monte Carlo simulations to the final cut-based analysis is exposed in detail, considering some alternative approaches concerning the reconstruction procedure as well. Finally, a short overview of the calculation of the upper limit of the branching fraction and the value itself is given.

Contents

Preface: The Burden of Negative Knowledge 7

1 Towards a New Era at CERN 9
1.1 Fifty Years of Accelerated Physics . . . 9
1.2 The Large Hadron Collider . . . 10
1.3 The Compact Muon Solenoid . . . 11

2 Theoretical Motivation 13

3 Neutrino Reconstruction 17
3.1 Decay Topology . . . 17
3.2 Illustrative Understanding . . . 18
3.3 Full Calculation . . . 18
3.4 Reconstruction of the B_s^0 . . . 19

4 Data Production 21
4.1 CMSSW, Pythia, EvtGen and FAMOS . . . 21
4.1.1 CMSSW . . . 21
4.1.2 Pythia . . . 21
4.1.3 EvtGen . . . 22
4.1.4 FAMOS . . . 22
4.2 The Signal Sample . . . 22
4.3 The Background Sample . . . 22
4.4 Candidate Reconstruction . . . 23
4.5 Efficiencies of the Signal Sample . . . 24
4.6 Efficiencies of the Background Sample . . . 25
4.7 Pseudo-Code for Candidate Building . . . 25

5 Resolutions, the k-Factor, and All That 27
5.1 Resolution of the Tau-Vertex . . . 27
5.2 Reconstruction on a Line and on a Helix . . . 28
5.3 Neutrino Reconstruction vs. k-Factor Method . . . 29
5.4 Classification of the Background . . . 30

6 Analysis 35
6.1 Candidate Selection . . . 35
6.1.1 Variables and Candidate Selection Cuts . . . 35

6.1.2 Selection of a Single Candidate . . . 36
6.1.3 Efficiency of the Candidate Selection . . . 37
6.2 Neutrino Reconstruction . . . 38
6.2.1 Restriction to the Theory and Application . . . 38
6.2.2 Pseudo-Code for Neutrino Reconstruction . . . 40
6.3 Efficiencies . . . 41
6.3.1 Efficiencies for Signal . . . 41
6.3.2 Efficiencies for Background . . . 41
6.4 Cut-based Analysis . . . 42
6.4.1 Pre-Selection Cuts . . . 42
6.4.2 Factorizing Cuts . . . 43

7 Determination of the Upper Limit 47
7.1 Addition of Signal and Background Distribution . . . 47
7.2 Bayesian Approach for the Upper Limit . . . 48
7.2.1 Calculation of N($B_s^0 \to \mu\tau$)$_{90\%}$ . . . 49
7.2.2 Calculation of N($B_s^0 \to$ anything) . . . 49
7.2.3 Grid-Search . . . 50
7.3 The Upper Limit of $\mathcal{B}(B_s^0 \to \mu\tau)$ . . . 50
7.4 Discussion of the Upper Limit . . . 51

8 Summary & Outlook 53

A Additional Plots 55
A.1 Plots for Cut Variables in Mass Region $4.8\,\text{GeV} - 5.6\,\text{GeV}$ . . . 55
A.2 Generator Level Plots . . . 56

B Acknowledgments 61

Preface: The Burden of Negative Knowledge

This thesis is about the analysis $B_s^0 \to \mu\tau$, where $\tau \to \pi\pi\pi\nu_\tau$. It will be revealed in the following pages that the study of this decay is somewhat difficult, as far as its reconstruction and its appearance in an particle collisions event is concerned: It is predicted to be very rare or almost not existing. There will be no possibility to discover $B_s^0 \to \mu\tau$ with the presented analysis, given the various theoretical predictions, within the framework of the CMS detector. It is therefore one of the goals of this thesis to determine an upper limit on the branching fraction, which tells us that the frequency of appearance of the decay is lower than a certain value. Or, spoken more sloppily, it is just another expression for the certainty we know, that the decay «is not there where we're looking for it». This sort of knowledge is referred to as «negative knowledge», and is quite common in particle physics: «High Energy Physics [...] cultivates a kind of negative knowledge. Negative knowledge is not nonknowledge, but knowledge of the limits of knowing, of the mistakes we make in trying to know, of the things that interfere with our knowing, of what we are not interested and do not really want to know» [1].

But not only the calculation of the upper limit of the branching fraction is an example of this «liminal phenomenon», the whole analysis (as every other high energy physics analysis) is subject to the burden of negative knowledge, incorporated in various calculations of (in)efficiencies throughout the whole analysis chain. These efficiencies on the one hand account for the loss of physics events, that cannot be observed in the detector (due to limited acceptance), and on the other hand for the different cuts that are introduced to reduce the background with respect to the signal, counting the unwanted background rather than the signal.

Thus, the presented thesis is in large parts about what we do not *know* about the decay $B_s^0 \to \mu\tau$ and the attempt to figure this out, and about what we do not *want* to be part of the decay $B_s^0 \to \mu\tau$ in the final analysis. Even though this will not lead to a observation of Lepton Flavor Violating $B_s^0 \to \mu\tau$-decays, its discovery still remains the ultimate goal. In other words: «The point of interest here is that limit analyses are yet another way to mobilize negative knowledge and turn it into a vehicle for positive knowing. The limits experiment produce are the thresholds from which new experiments start their searches» [1].

1 Towards a New Era at CERN

1.1 Fifty Years of Accelerated Physics

The first step to the foundation of CERN began with a proposal of french physicist Louis de Broglie. At the European Cultural Conference in Lausanne in December 1949 he suggested the creation of an European atomic physics laboratory. Only two years later the «Conseil pour la Recherche Nucleaire» (CERN) was established. The provisional council, however, was not to last for a long time: On the 29. September 1954, it was replaced by the newly founded European Organization for Nuclear Research. The European particle physics laboratory as the organization we know it today, was launched – the name, however, was kept from its predecessor: The CERN was born.

A major date in the development of the still young organization was the 24. November 1959: This was the day when the first proton beams were accelerated in the Proton Synchrotron (PS), with an energy of 28 GeV – the highest achievable energy at that time (only to be surpassed a little later by the AGS in Brookhaven). The PS maybe was one of the best investments of CERN, still being used as a workhorse nowadays, mainly to pre-accelerate particles for succeeding colliders.

The next step in the development of high energy physics accelerators, particularly using hadrons, were the Intersecting Storage Rings (ISR), the first proton-proton collider ever built: The necessity of larger and larger centre of mass energies required the change from fixed target to collision experiments. The ISR, which began operation in winter 1970 / 1971, was an excellent test facility for the upcoming colliders, including the testing of a method called «stochastic cooling», which would be used later on in the collision of protons with antiprotons. The ISR actually was the first CERN facility really belonging to international ground: It was build in France, adjoining the older buildings on Swiss ground.

With the first protons accelerated in the newly built Super Proton Synchrotron (SPS), things became really huge: The machine had a circumference of 7 km and allowed a particle energy of 400 GeV, 100 GeV more than planned in the beginning. Due to its size, it was build subterraneous. This accelerator, besides being used for for instance to probe the inner structure of the proton, later became famous in the «Super Proton AntiProton Synchrotron» - mode, when the W- and Z-Bosons were discovered in the year 1983. The eighties were the decade of superlatives anyway, at least what CERN is concerned: Already two years after the discovery of the mediators of the weak interaction, the construction of LEP, the large electron-positron collider began: A 27 km circumference accelerator, situated up to 100 m under the surface. Till the year 2000, four different particle detectors measured the electron-positron collisions, corroborating the fact, that there are only three neutrino families, and that the mass of the long-searched Higgs boson is heavier than 114.4 GeV.

1.2 The Large Hadron Collider

The Large Hadron Collider (LHC) was already planned in the eighties and will start operation in summer 2008. It is constructed as a proton-proton collider, with the option for colliding heavy ions (e.g. lead atoms) as well. The LHC is situated in the old LEP-tunnel and will use the PS and the SPS as pre-accelerators (consider fig. 1.1). The most remarkable key-number coming with the new collider is its beam energy, which amounts to 7 TeV[1] for each proton beam – a factor of seven more compared to older accelerators. To hold the particles on their track in the circular LHC, superconducting dipole magnets with a field of 8.3 Tesla are used, which is on the order of 170'000 times the earth magnetic field in middle Europe. The design luminosity, not reached until 2010, is $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

Six experiments are placed on different locations in the ring of the LHC: ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), two «general purpose» detectors, have a rather universal design and are destined to find new physics and particles up to the TeV scale. LHCb (LHC beauty experiment) will be studying the properties of b-mesons (including CP-violation and rare decays) while ALICE (A Large Ion Colliding Experiment) is built for the study of heavy ion collisions. Additional experiments are LHCf (LHC forward), which is dedicated to cosmic ray physics and shares the interaction point with ATLAS, and TOTEM (TOTal cross section, Elastic scattering and diffraction dissociation Measurement) sharing the interaction point with CMS. It will, for instance, measure the LHC luminosity.

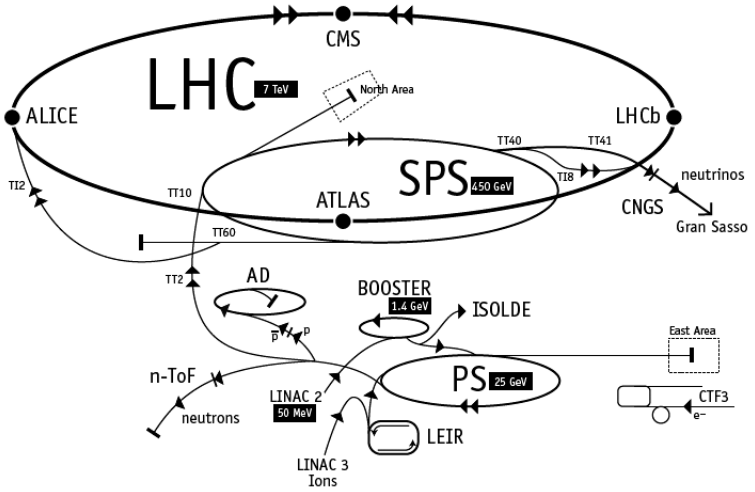

Figure 1.1: The accelerator facility at CERN [3].

[1] All the numbers and facts about the LHC and CMS are taken from [2]

1.3 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) detector is located at the LHC access point five, near the french town of Cessy. The five main parts of the detector were constructed on surface and later on assembled in the underground cavern. The Institute for Particles Physics of the ETH is part of the CMS-collaboration and mainly contributed to the inner tracking system and the electromagnetic calorimeter.

The most notable thing about CMS is its special design: Unlike many other detectors, the electromagnetic and the hadronic calorimeter (ECAL and HCAL) are within the superconducting coil, that generates a solenoidal field of 4 T. This allows a relatively compact design without making too strong restrictions to the strength of the magnetic field. The overall detector has a length of 21.6 m, a diameter of 15 m and weighs 12'500 tons.

The innermost parts of CMS are made of silica: It is the tracker consisting of **Pixel and Strip Detector**. The pixel detector consist of three layers of pixels, with a size of 100×150 μm^2 each, which are situated at 4.4 cm, 7.3 cm and 10.2 cm radial distance from the interaction point. In addition to this, there are two end disks placed on each side of CMS. The whole pixel detector comprises 66 million pixels, 48 million in the barrel and 18 million in the end disks. Next to the pixel detector is the silicon strip detector: It consists of four plus six layers of silicon strips with a length of 130 cm and 220 cm respectively. The end caps can be divided in Tracker End Caps (TEC) and Tracker Inner Disk (TID). While the TEC comprises nine disks, the TID consists of three smaller disks that fill the gap between the inner four layers of silicon strips and the TEC. The whole tracker can measure the position of charged particles up to a pseudorapidity of $|\eta| < 2.4$.

The next parts, situated further away from the interaction point, are the calorimeters: The **Electromagnetic Calorimeter (ECAL)** to measure the energy of electrons or photons (via electromagnetic showers) and the **Hadronic Calorimeter (HCAL)** to measure the energy of charged and neutral hadrons. The **ECAL** consists of 61'200 lead-tungstate crystals in the barrel and 7324 crystals in the end caps. These crystals have the property of being very radiation-hard and having a short radiation length, with a relatively low light yield, however. The scintillation light is read out using avalanche photo diodes for the crystals in the barrels and vacuum phototriodes in the end caps. The crystals are clustered together in supermodules in the barrel region, with a length of half the barrel length, while in the end caps they are put together in 5×5 units. The **HCAL** is built as a sampling calorimeter, with brass plates as the absorber material and scintillating tiles with embedded wavelength-shifting fibers to read out the signal. In addition to this, there is a so called hadron outer detecter, made of 1 cm thick scintillators covering the outside of the outer vacuum tank of the superconducting coil. This device is used to «catch the tails» of the energy distributions measured with the HCAL and therefore increases the effective thickness of the calorimeter. The HCAL is completed with the Hadron Forward Calorimeter which covers $3 < |\eta| < 5$, using a steel/quartz fiber and measures Cherenkov light emitted in the quartz fibers.

The **Muon System** is the outermost part of CMS: It is built as an «instrumented flux return», meaning that it intersects the iron yoke, made to return the magnetic flux of the solenoid, with muon chambers. They consist of drift tubes in the barrel region, cathode strip chambers in the two end caps and resistive plate chambers (RPC), which are used in both the barrel and the end cap (to allow for a better time resolution). The muon system as a whole can measure the properties of the

muon up to a pseudorapidity of $|\eta| = 2.4$, while the RPCs only cover a region up to $|\eta| = 1.6$.
At design performance, every 25 ns two bunches will cross and produce a particle shower, corresponding to a rate of 40 MHz. As only data from 100 crossings per second can be stored, the rate has to be reduced dramatically. For this purpose, several trigger system were developed. The first one is the **Level-1 Trigger (L1)**. This decision system is directly integrated in hardware processors and involves the muon-system and the calorimetry, as well as a combination of both. To trigger on photons and muons, a reconstruction with reduced granularity and resolution is used to make a decision, if the event shall pass or fail. The design value of the L1 trigger is about 100 kHz; at the startup of the LHC, however, only about 16 kHz will be reached.
In a second stage, the **High-Level Trigger (HLT)** comes into place. If the event has passed the L1 trigger, it is read out and transferred to a processor in a processor farm, running a high level trigger software. This reduces the rate from 100 kHz to 100 Hz, which is usable for storage.

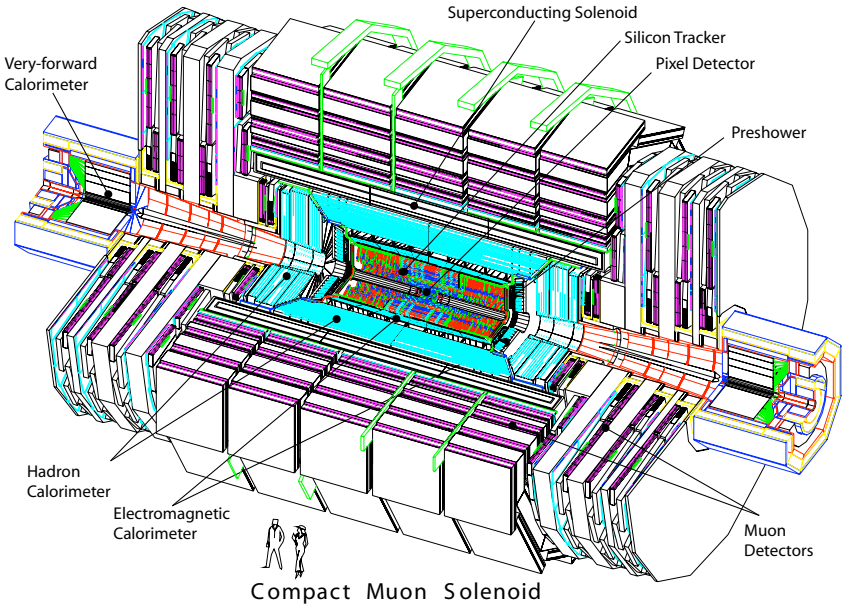

Figure 1.2: Scheme of the CMS detector [2].

2 Theoretical Motivation

The punch line of the search for the decay $B_s^0 \to \mu\tau$ is Lepton Flavor Violation: The Lepton Family Number in this decay is not conserved. In the Standard Model (SM), this is only possible with the contribution of neutrino oscillations (consider the Feynman diagram in fig. 2.1). However, as the neutrino masses are smaller than 2 eV [4] and taking the suppression with $\left(\frac{m(\nu)}{m(W)}\right)^2$ into account, the branching fraction of $B_s^0 \to \mu\tau$ would be on the order of 10^{-50}, far beyond the reach of present and future detectors.
The present upper limit for the similar decay $B^0 \to \mu\tau$ [5] is:

$$\mathcal{B}(B^0 \to \mu\tau) = 2.2 \cdot 10^{-5} \text{ (90 \% C.L.)}$$

Physics beyond the Standard Model
This starting position renders the decay an excellent probe for physics beyond the Standard Model. On the one hand, the decay is not «polluted» by background from known physics (i.e. if one observes the decay, it must be due to new physics), on the other hand, there are strong theoretical arguments that enhance the branching ratio by orders of magnitude to a limit that could be observable with detectors in the near future.

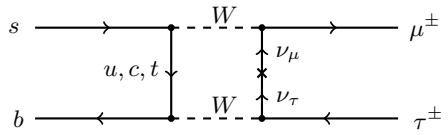

Figure 2.1: Feynman diagram for $B_s^0 \to \mu\tau$ in SM.

As there are many different models allowing for LFV, only two of them, giving rise to a branching ratio within a reasonable limit to detect, shall be briefly presented.
In [6], a model is described where the decays $B_{s,d} \to \mu\tau$ are mediated through a Higgs boson, using the supersymmetric seesaw mechanism. The seesaw mechanism leads to a low, but non-vanishing mass of the (SM) neutrino, introducing a super-heavy right-handed neutrino. Due to this new right-handed neutrino, one has to introduce new Yukawa couplings. They give rise to non-diagonal terms in the Yukawa matrix of the neutrinos, therefore leading to neutrino oscillations and hence flavor violation, similar as in the quark sector. This shall be shortly pointed out:
Suppose we have a matrix with Yukawa couplings f_l^{ij} for the charged leptons, f_ν^{ij} for the neutrinos and M_ν^{ij}, the mass matrix of the right-handed neutrinos (for details on this consult [7]). Now we can diagonalize the matrices for the charged leptons, $f_{li}\delta^{ij}$, and $M_{Ri}\delta^{ij}$. In a general case, however, this means that the Yukawa matrix f_ν^{ij} will not be diagonal. Now write $f_\nu^{ij} = U^{ik} f_{\nu k} V^{kj}$ with

U,V unitary (this is always possible). The seesaw mechanism then gives for the mass matrix of the neutrinos m_ν:

$$m_\nu = \frac{1}{M_R} V^T \begin{pmatrix} f_{\nu,1}^2 & & \\ & f_{\nu,2}^2 & \\ & & f_{\nu,3}^2 \end{pmatrix} V \cdot \frac{v^2}{2} \sin^2 \beta$$

with M_R the mass of the right-handed neutrino, $\frac{v^2}{2} = \langle h_1 \rangle^2 + \langle h_2 \rangle^2$ twice the expectation value of the Higgs doublet squared and $\tan \beta = \frac{\langle h_2 \rangle}{\langle h_1 \rangle}$. Thus, if $V \neq 1$, the neutrino mass matrix will not be diagonal and neutrino oscillations / LFV will occur. But as we know that the neutrino masses are very small, M_R needs to be very big ($\sim 10^{12}$ GeV or even bigger), which strongly suppresses lepton flavor violating processes, as their rates are $\propto \frac{1}{M_R}$.
In the context of supersymmetry (SUSY), mainly in the Minimal Supersymmetric Standard Model (MSSM), however it is possible to introduce a slepton mass matrix (with the slepton being the supersymmetric partner-particle of the lepton), that comprises off-diagonal terms as well, which themselves give rise to Lepton Flavor Violation at a higher rate:

$$(\Delta m_{\tilde{L}}^2)^{ij} \sim -\frac{L}{8\pi^2}(3m_0^2 + A_0^2) m_0^2 (f_\nu^\dagger f_\nu)^{ij}$$

with $(\Delta m_{\tilde{L}}^2)^{ij}$ being the difference squared between the slepton masses, where $L = \log(\frac{M_{GUT}}{M_R})\delta_{ij}$. M_{GUT} is denoting the mass scale of $\sim 2 \cdot 10^{18}$ GeV, m_0 is a universal scalar mass (all $(m_{\tilde{f}}^2)^{ij}$ take the same value m_0 at the M_{GUT}-scale). Furthermore we can convert f_ν^{ij}

$$f_\nu^{ij} = f_\nu^{ik\dagger} f_\nu^{jk} = V_k^{i\star} |f_{\nu k}|^2 V^{jk}.$$

As before in the Standard Model example, there will be off-diagonal terms in the general case when V is not the unit matrix, which lead to LFV.
In the discussed model, these terms represent the only source of Lepton Flavor Violation.
Assuming that the heavy neutrino masses are all equal and on the order of $\sim 10^{14}$ GeV, the supersymmetry breaking mass parameters are all equal at low scale and making restrictions to the Yukawa coupling matrices, we get an estimate of the upper limit of the branching fraction of

$$\mathcal{B}(B_s^0 \to \mu\tau) \sim 3.6 \cdot 10^{-7} \left(\frac{\tan\beta}{60}\right)^8 \left(\frac{100\,\text{GeV}}{M_A}\right)^4,$$

where M_A is the Higgs mass. As the name suggest, this branching fraction is largely dominated by the Higgs mass and $\tan\beta$. The Penguin diagram for this decay is shown in 2.2.

An alternative model that leads to observable Lepton Flavor Violation due to yet undiscovered particles is offered by a «Littlest Higgs Model with T-Parity» [8]. In this model, the predictions for $B_s^0 \to \mu\tau$ are:

$$\mathcal{B}(B_s^0 \to \mu\tau) = 2 \cdot 10^{-10} \quad (f = 1000\,\text{GeV}) \tag{2.1}$$
$$\mathcal{B}(B_s^0 \to \mu\tau) = 3 \cdot 10^{-9} \quad (f = 500\,\text{GeV}) \tag{2.2}$$

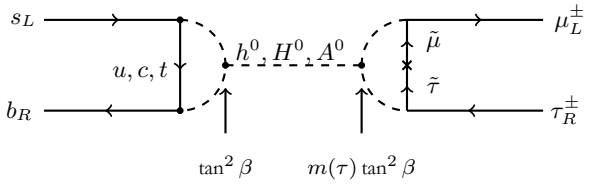

Figure 2.2: Feynman diagram for $B_s^0 \to \mu\tau$ in MSSM.

where f stands for the scale of New Physics. The Feynman diagram for these processes are shown in fig. 2.3 and 2.4. The new particles are (heavy) mirror particles introduced in the Little Higgs Model to solve the little hierarchy problem.

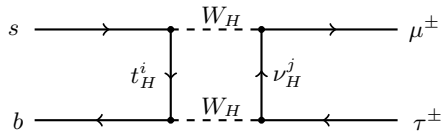

Figure 2.3: Feynman diagram for $B_s^0 \to \mu\tau$ in LHT.

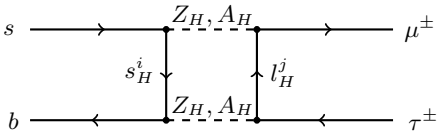

Figure 2.4: Feynman diagram for $B_s^0 \to \mu\tau$ in LHT.

Further Physical Aspects

In this thesis, only the decay $B_s^0 \to \mu\tau$ will be studied, the possible decays $B_s^0 \to e\mu$, $B_s^0 \to e\tau$, ... will be left away. As the decay is helicity-suppressed, the heavier the particles are, the larger the probability of them to be in the opposite helicity. The ratio $\frac{B_s^0 \to \mu e}{B_s^0 \to \mu\tau}$, for example, would be on the order of $\left(\frac{m(e)}{m(\tau)}\right)^2 \sim 10^{-8}$ [9]. It is therefore more likely to observe $B_s^0 \to \mu\tau$ than any of the other flavor violating decays of the B_s^0.

In an experimental environment, the study of $B_s^0 \to \mu\tau$ would be quite similar to the study of $B_d^0 \to \mu\tau$. However, the latter decay is suppressed with a factor of $\left|\frac{V_{td}}{V_{ts}}\right| \sim 0.2$ compared to the first one (consider fig. 2.5). Note that in this calculation only the top-quark has been considered, as it is dominant in the loop (due to its large mass). Further note that this explanation actually is only true, if CKM-like matrix-elements in models beyond the Standard Model behave in a similar way as they do in the SM. This, however, is expected from the non-observation of New Physics in K- and B-decays [10].

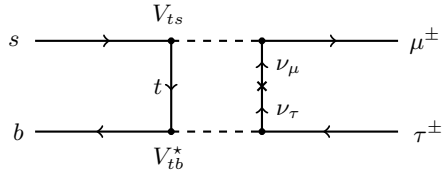

Figure 2.5: Feynman diagram with CKM-matrix elements in $B_s^0 \to \mu\tau$ (SM).

Decay of the Tau

As the tau is not a stable particle, a decay channel has to be chosen to study $B_s^0 \to \mu\tau$. In this thesis, $\tau \to \pi\pi\pi\nu_\tau$ was selected ($\tau \to hhh\nu_\tau$ where h is a hadron, is referred to as «three-prong»). The branching ratio [4] of this decay is

$$\mathcal{B}(\tau \to \pi\pi\pi\nu_\tau) = (9.02 \pm 0.08)\% \tag{2.3}$$

and the Feynman diagram is depicted in fig. 2.6. As in every (SM) decay of the tau a neutrino is produced, which cannot be observed in the detector, a decay channel was chosen that contains more than one observable particle. This will be become of great importance in the reconstruction of the tau (see chapter 3), as a (fitted) decay vertex of the tau is needed.

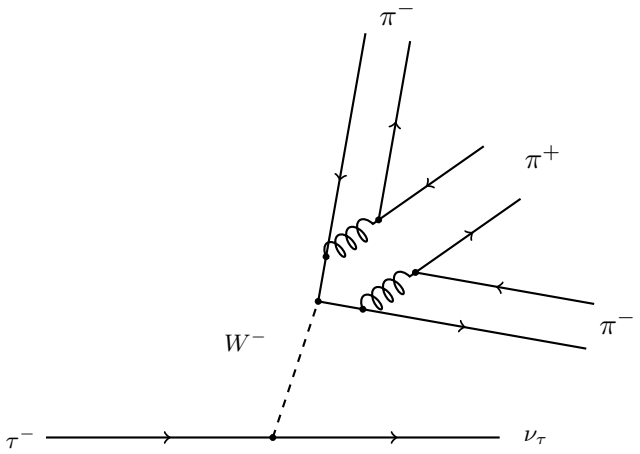

Figure 2.6: Feynman diagram for $\tau \to \pi\pi\pi\nu_\tau$.

3 Neutrino Reconstruction

To fully reconstruct the B_s^0, the whole decay chain $B_s^0 \to \mu\tau$, where $\tau \to \pi\pi\pi\nu_\tau$, has to be considered. In a general case, however, it is not possible to reconstruct a decay with a neutrino, as this particles leaves no tracks in the detector. Exceptions are given for instance if the missing energy of the neutrino can be determined precisely enough. As this is not the case in the CMS experiment, a different approach had to be chosen, namely using topological information and symmetry arguments, described in detail in [11].
In a first step the decay topology of $B_s^0 \to \mu\tau$, $\tau \to \pi\pi\pi\nu_\tau$ is presented. Then, an intuitive approach to the reconstruction of the neutrino is exposed, followed by the mathematical calculation of the four-momentum of the neutrino. At the end, it is shortly demonstrated, how the B_s^0 can be reconstructed in an ideal environment.

3.1 Decay Topology

The decay $B_s^0 \to \mu\tau$, $\tau \to \pi\pi\pi\nu_\tau$, the way it would present itself to an omniscient observer, is depicted in fig. 3.1. As only the long-lived particles (pions, kaons, muons, protons,...) even reach the detector and the neutrinos do not leave tracks in the detector, the topology of the reconstructible tracks greatly differs from the ideal case (consider fig. 3.2).

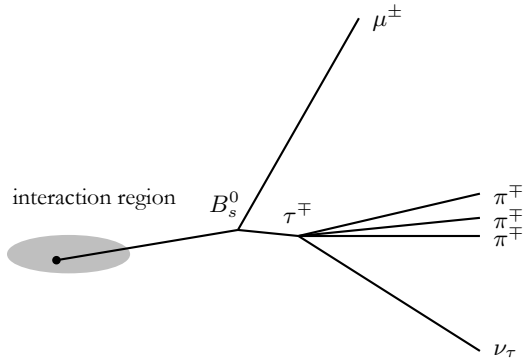

Figure 3.1: Decay topology of $B_s^0 \to \mu\tau, \tau \to \pi\pi\pi\nu_\tau$ for an omnisicient observer.

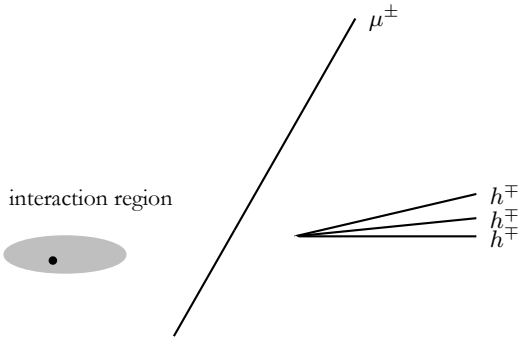

Figure 3.2: Decay topology of $B_s^0 \to \mu\tau$, $\tau \to \pi\pi\pi\nu_\tau$, reconstructed by the detector.

h stands for «hadron», as no particle-ID is incorporated in the CMS detector (e.g. pions and kaons cannot be distinguished). The muon on the other hand can be identified to a high efficiency and purity. Note that the straight tracks of the particles in these figures are an idealization for the sake of clarity. Certain aspects related to this topic are further discussed in 5.2.

3.2 Illustrative Understanding

To understand the process, how the neutrino four-momentum can be gained via topological information, it is easiest to consider a two-body decay of a particle at rest. Note that even though the tau decays in three pions and a neutrino, the four-momentum vectors of the three pions can be taken together to $\mathbf{p}_{3h}$ and therefore the three particles can be treated as a single one.
If a particle (a tau, in our case) decays at rest in two daugther particles, dau 1 and dau 2, the quantities $E_{\text{dau 1,2}}$ and $|\vec{p}_{\text{dau 1,2}}|$ are determined via the following formulae (see [4] and consider fig. 3.3).

$$\begin{aligned} E_{3h} = \frac{m_\tau^2 + m_{3h}^2}{2m_\tau} \quad &\Rightarrow \quad |\vec{p}_{3h}| = \sqrt{E_{3h}^2 - m_{3h}^2} \\ E_\nu = \frac{m_\tau^2 - m_{3h}^2}{2m_\tau} \quad &\Rightarrow \quad |\vec{p}_\nu| = E_\nu \\ &\text{and} \\ \vec{p}_\nu \quad &= \quad -\vec{p}_{3h} \end{aligned} \tag{3.1}$$

Now consider fig. 3.4, where we look at the tau rest frame. As we only know the flight direction of the tau rather than its exact momentum (resulting in a loss of symmetry), the condition in eq. 3.1 has to be replaced with the weaker one $|\vec{p}_\nu| = |\vec{p}_{3h}|$. Hence, the transverse component of the neutrino respective to the flight direction of the tau is minus the transverse component of the three-prong, $p_{\nu,\perp} = -p_{3h,\perp}$, whereas its component parallel to the tau flight direction is plus or minus the corresponding component of the three-prong. This gives rise to two solutions of the four-momentum of the neutrino: One solution accounts for the neutrino emitted in forward direction, one for the neutrino emitted in backward direction (with respect to the tau flight direction).

At last, look at the same situation, but this time in the lab rest frame. As the tau is boosted, the parallel component of the momentum of the three-prong and the neutrino are stretched in forward direction, but the transverse component remains the same as in fig. 3.4, because it is not subject to the Lorentz boost.

Therefore: If the four-momentum of the three-prong and the flight direction of the tau are known, the four-momentum of the neutrino can be reconstructed up to a twofold solution.

Note that the two solutions cannot be distinguished in a real physics case, meaning that only one solution is realized in a decay while the other one is just mathematically and physically possible, but does not occur.

3.3 Full Calculation

The calculation of the missing component $p_{\nu,x}$ is straightforward by using the relations of special relativity, but not as illustrative as the above explanation. As the decay can be completely described in two dimensions, only two vector components are assigned to the momentum of the particles. Without loss of generality, the flight direction of the tau is assumed to be in the x-direction.

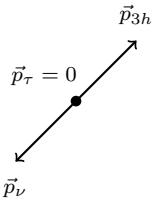

Figure 3.3: Two-body decay of τ, where $\vec{p}_\tau = 0$.

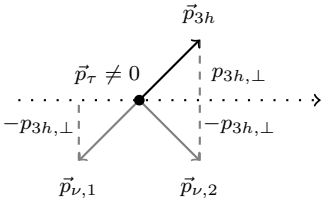

Figure 3.4: Two-body decay of τ, where $\vec{p}_\tau \neq 0$, τ rest frame.

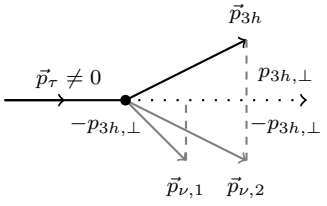

Figure 3.5: Two-body decay of τ, where $\vec{p}_\tau \neq 0$, lab rest frame.

$$
\begin{aligned}
p_{\tau,x} &= p_{\nu,x} + p_{3h,x} \\
E_\tau &= E_\nu + E_{3h} \\
\Rightarrow E_\tau &= \sqrt{p_{\nu,x}^2 + p_{\nu,y}^2} + E_{3h} \\
\Leftrightarrow E_\tau &= \sqrt{p_{\nu,x}^2 + p_{3h,y}^2} + E_{3h} \\
\Rightarrow m_\tau^2 &= E_\tau^2 - p_{\tau,x}^2 \\
&= (\sqrt{p_{\nu,x}^2 + p_{3h,y}^2} + E_{3h})^2 - (p_{\nu,x} + p_{3h,x})^2 \\
0 &= \underbrace{m_\tau^2 - p_{3h,y}^2 - E_{3h}^2}_{\alpha} + 2p_{\nu,x}p_{3h,x} - 2E_{3h}\sqrt{p_{\nu,x}^2 + p_{3h,y}^2} \\
\Rightarrow 0 &= \alpha^2 + 4p_{\nu,x}^2 p_{3h,x}^2 + 4\alpha p_{\nu,x} p_{3h,x} - 4E_{3h}^2 p_{\nu,x}^2 - 4E_{3h}^2 p_{3h,y}^2 \\
\Rightarrow 0 &= p_{\nu,x}^2 + \underbrace{\frac{\alpha p_{3h,x}}{p_{3h,x}^2 - E_{3h}^2}}_{a} p_{\nu,x} + \underbrace{\frac{\alpha^2 - 4E_{3h,h}^2 p_{3h,y}^2}{4(p_{3h,x}^2 - E_{3h}^2)}}_{b}
\end{aligned}
\tag{3.2}
$$

The calculation therefore leads to a quadratic equation, which accounts for the two solutions already mentioned in section 3.2.
The solution of this equation is (using a and b of eq. 3.2):

$$
p_{\nu,x,1,2} = \frac{-a \pm \sqrt{a^2 - 4b}}{2} \tag{3.3}
$$

The missing component $p_{\nu,x,1,2}$ therefore can be computed using values measurable with the detector.

3.4 Reconstruction of the B_s^0

As mentioned, the crucial point for reconstructing the neutrino is the knowledge about the flight direction of the tau. As only the decay vertex of the tau and the extrapolated muon trajectory (rather

than the muon production vertex) is known, additional information is needed. This is obtained by «choosing» a point on the muon trajectory. With this additional point, the flight direction and flight length of the B_s^0 as well as the flight direction and length of the tau are known. But as there is only one point corresponding to the physical decay vertex of the B_s^0, the muon trajectory has to be searched through to find the point for which

$$p_{\mu \perp B_s^0} = p_{\tau \perp B_s^0} \tag{3.4}$$

is valid, where $p_{\mu \perp B_s^0}$ is the perpendicular momentum of the muon with respect to the B_s^0 flight direction, and $p_{\tau \perp B_s^0}$ is the perpendicular momentum of the tau with respect to the B_s^0 flight direction. Note that this is an idealization, equation (3.4) will never be fulfilled in the environment of a detector. In section 6.2.1 it is explained how the basic idea of this technique is applied in the analysis.
The whole reconstruction is summarized in fig. 3.6.

Neutrino Reconstruction

- A point on the muon trajectory is chosen.
- The neutrino is reconstructed, balancing $p_{3h \perp \tau}$ and $p_{\nu \perp \tau}$.
- The four-momentum of the tau is calculated.
- The B_s^0 is reconstructed, balancing $p_{\mu \perp B_s^0}$ and $p_{\tau \perp B_s^0}$.

The implementation of this algorithm is discussed in section 6.2.1.

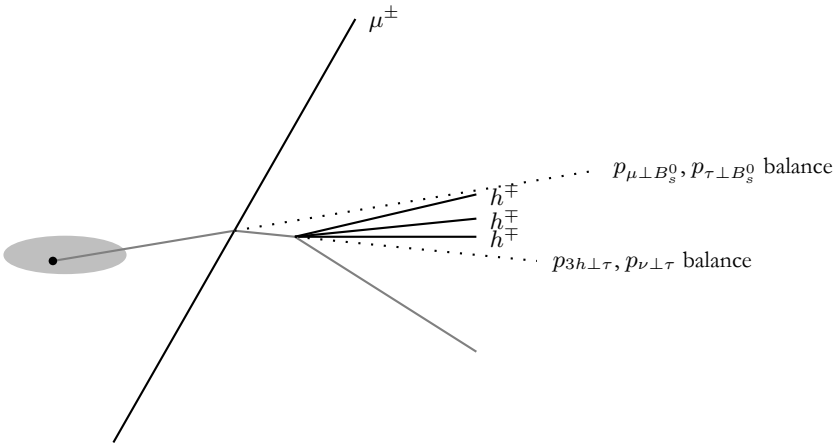

Figure 3.6: Scheme of the reconstruction of the B_s^0, using neutrino reconstruction and kinematical arguments.

4 Data Production

As no data has been available of the LHC and CMS yet, the collisions of the protons and the following decays of the particles had to be simulated using a Monte Carlo Simulation. The following chapter presents the production of the signal and background data sample, the constraints on the generated data to build a B_s^0-candidate and some details about the software used.

4.1 CMSSW, Pythia, EvtGen and FAMOS

4.1.1 CMSSW

CMSSW stands for CMS SoftWare and comprises the software needed to perform physics analysis with the CMS detector. The main idea in CMSSW is the *event*, a unit that contains all data (raw and reconstructed) of a real physics event. A feature of CMSSW is the fact, that it only has one *executable* (cmsRun), but many modules that can be put in a *path*. This allows the physicist to adapt the software to his needs. An event then is passed from one module to the next, till the end of the path is reached. The modules can be classified according to:

- *Source*: Reads in an event from a file (either with real CMS data or from a Monte Carlo Simulation) or on-the-fly from a Monte Carlo Simulation.
- *EDProducer*: Reads data from the event, produces something with it, and adds it to the event.
- *EDAnalyzer*: Analyzes data from the event and writes output, but does not change the event itself.
- *EDFilter*: Returns a boolean that may stop the execution of the path.
- *Output Module*: Stores the output to external media at the end of the path. Not used in this analysis.
- *EDLooper*: Allows to loop over an input's source data. Not used in this analysis.

Examples of the different types will be given in the following three sections and in section 4.7.

4.1.2 Pythia

Pythia is a high energy physics event generator, used in version 6.409 [12]. It serves as a *source* in CMSSW. Pythia handles a wide range of particle decays, but relies only on leading-order calculations and does not provide the full spectrum of B-decays and models (such as spin correlation).

4.1.3 EvtGen

EvtGen [13] originally was an event generator, dedicated to exact studies of B-decays, including complex sequential decays as well as CP-violation phenomena. However, in the used version 8.16 it is an *EDProducer*, that forces particles produced by Pythia to be stable and lets them decay itself (mostly B-mesons). In the CMSSW-version of EvtGen, Photos (see for example [14]) is incorporated as well, calculating the emission of bremsstrahlung.

4.1.4 FAMOS

While full detector simulations like GEANT 4 [15] are accurate but very CPU-time consuming, physics analysis often need a lot of statistics in order to give quantitative predictions. For this purpose, FAMOS (for Fast Monte Carlo Simulation) has been developed. This «detector simulation» smears the distribution of the properties of the particles rather than calculating every single particle interaction with the detector itself. This results in a much faster processing of the event, but in a slightly lower accuracy compared to the full simulation. In this analysis, only FAMOS was used to simulate the reaction of the detector.

4.2 The Signal Sample

The signal sample was produced using Pythia with MSEL=1 (i.e. every possible particle production out of the incoming protons was allowed), with the condition $\mathcal{B}(B_s^0 \to \mu\tau) = \mathcal{B}(\tau \to 3h\nu_\tau) = 1$. Over the GRID $7.8 \cdot 10^9$ events were generated, corresponding to a integrated luminosity of $\mathcal{L}_{prod} = 0.145\,\mathrm{pb}^{-1}$ (including the forced branching fractions).
As in this configuration not only B_s^0 were produced but every physically possible particle (as only the decay of the B_s^0 could be forced and not its production), two filters on generator level were introduced in order to select only events with the decay $B_s^0 \to \mu\tau$. The first one (Pythia Filter) made sure that the event comprised a muon and a tau from a B_s^0, where $|\eta_\mu| < 2.5$ was ensured. In a later stage, a second generator filter was used. This was mostly done to reduce the number of events that would not have been reconstructible in the detector later anyway due to kinematics. The following cuts on the transverse momentum of the pions and the muon were applied:

- $p_{\perp,\mu} > 2.5\,\mathrm{GeV}$
- $p_{\perp,\pi} > 1\,\mathrm{GeV}$.

Note that the condition for the transverse momentum of the pions had to be fulfilled for every single pion of the three-prong, not only for the three-prong overall perpendicular momentum.

4.3 The Background Sample

The Background Sample was produced using Pythia with MSEL=1 and EvtGen, where no constraints on the produced and decaying particles were set, i.e. the whole (known) background present at CMS was produced. Over the GRID $1.2 \cdot 10^{11}$ events were produced, corresponding to an integrated luminosity of 2.28 pb^{-1}. To speed up the simulation and to save disk space, a single

generator level filter was used, as in the production of the signal sample (and no Pythia Filter). It comprised the following conditions:

- If the event contained only one muon, it had to fullfill: $p_{\perp,\mu} > 4.5\,\text{GeV}$ (trigger simulation)
- If the event contained two or more muons, each of them had to fulfill: $p_{\perp,\mu} > 2.5\,\text{GeV}$ (trigger simulation)
- In a cone around one muon with $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < 1.7$ (with ϕ being the azimuthal angle and η being the pseudorapidity), there had to be at least three particles (h), satisfying $|\eta_h| < 2.5$ and $p_{\perp,h} > 1\,\text{GeV}$ each. If this was not the case, the next muon in the event was considered (if any).
- Three particles in the cone around the muon had to fulfill the condition $0.41\,\text{GeV} < m(hhh) < 1.85\,\text{GeV}$. The lower value is (approximately) the mass of three pions, the upper one (approximately) the mass of the tau.

If these conditions were not fulfilled, the event was discarded and the next one considered.
The whole path is again summarized in section 4.7, also including the following candidate reconstruction.

4.4 Candidate Reconstruction

After the reconstruction of the tracks with FAMOS, every event was searched for a muon: Events not including a muon track were discarded. This was mostly done to save disk space, as these events were of no use for the later analysis. To build a B_s^0-candidate in the remaining events, the following procedure was applied:

- If no primary vertex could be reconstructed, the event was discarded.
- If the event contained only one muon, it had to fulfill: $p_{\perp,\mu} > 4.5\,\text{GeV}$ (trigger simulation)
- If the event contained two or more muons, each of them had to fulfill: $p_{\perp,\mu} > 2.5\,\text{GeV}$ (trigger simulation)
- In a cone around every muon with $\Delta R < 1.7$ at least three particles (h) had to be located, with $|\eta_h| < 2.5$ and $p_{\perp,h} > 1\,\text{GeV}$ each. If this was not the case, the next muon in the event was considered (if any).
- Three particles in the cone around the muon had to fulfill the condition $0.41\,\text{GeV} < m(hhh) < 1.9\,\text{GeV}$. The lower value is (approximately) the mass of three pions, the upper one (approximately) the mass of the tau. Furthermore, the condition $Q(hhh) \neq Q(\mu)$ was set. If there were more than three particles satisfying these conditions, all of the possible combinations were considered.
- The three tracks of the mentioned particles were fitted to a (secondary) vertex using a Kalman Vertex Fitter. If fitting was not possible, the next combination of three tracks was considered.

- Using the fitted (secondary) vertex as an additional constraint, the three tracks were refitted.
- The four-momentum of the muon and of the three particles, as well as the position of the secondary vertex (plus some additional information related to these quantities), were stored and made up a B_s^0-candidate.

The position of the primary vertex and of the Point of Closest Approach (POCA) of the muon respective to the primary vertex were stored as well. The applied cuts of this procedure are again summarized in section 4.7.

Note that it was very likely to store more than one candidate, as the applied cuts were chosen rather loose (this will be further discussed in section 6.1.2) and combinatorics were very high. (If n particles passed the cuts, $\binom{n}{3}$ were stored as a B_s^0 - candidate, e.g. for n=5 there would already be 10 candidates).

4.5 Efficiencies of the Signal Sample

Due to kinematic restrictions (as the CMS detector has not a 100 % sensitivity in every direction and for every measurable quantity) and the application of cuts, not all the three-prongs and muons of a decaying B_s^0 could be reconstructed, leading to efficiency losses. The efficiencies corresponding to certain steps in the whole path and the modules are summarized in table 4.1. The initial number of B_s^0 was computed using the fact that

$$\#B_s^0 = \#\text{pp-coll.} \cdot \frac{\sigma_{b\bar{b}}}{\sigma_{pp}} \cdot f_s \cdot 2$$

where # pp-coll. stands for the number of proton-proton collisions, $\sigma_{b\bar{b}}$ is the $b\bar{b}$-cross-section (500 μb), σ_{pp} is the proton-proton cross-section (54.71 mb), and f_s is the hadronization probability for a b-quark to form a B_s^0 meson.

Table 4.1: Efficiency of signal sample.

Step in Candidate Building	**Fraction**
$B_s^0 \to \mu\tau$ (generator)	100 %
$B_s^0 \to \mu\tau$, $p_{\perp,\mu} > 2.5\,\text{GeV}$, $\lvert\eta_\mu\rvert < 2.5$ (generator)	47.6%
$B_s^0 \to \mu\tau$, $p_{\perp,\mu} > 2.5\,\text{GeV}$, $p_{\perp,\pi} > 1\,\text{GeV}$ (generator)	4.4%
μ-track reconstructed	2.8 %
primary vertex reconstructed	2.7 %
$1 \times \mu$ with $p_{\perp,\mu} > 4.5\,\text{GeV}$, or $\geq 2 \times \mu$ with $p_{\perp,\mu} > 2.5\,\text{GeV}$ reconstructed	2.2 %
B_s^0-candidate reconstructed	1.9 %

The numbers were computed using a signal sample with approximately $14.3 \cdot 10^6$ events containing a B_s^0. It is clearly visible that the steepest decrease in percentage occurs where three pions with

$p_{\perp,\pi} > 1\,\text{GeV}$ are demanded. This is mostly due to the fact that this condition has to be fulfilled for every pion, i.e. it comes with the power of three.

4.6 Efficiencies of the Background Sample

As for the signal sample, the efficiencies for the background sample were calculated (consider table 4.6). Because no decay channel was forced, the number of pp-collisions represent the 100 %. The steepest descent in the efficiency in the background sample is due to the low muon reconstruction efficiency.

Table 4.2: Efficiency of background sample.

Step in Candidate Building	**Fraction**
# pp-collisions	100 %
$\geq 1 \times \mu$ in event (generator)	1.685 %
$1 \times \mu$ with $p_{\perp,\mu} > 4.5\,\text{GeV}$, or $\geq 2 \times \mu$ with $p_{\perp,\mu} > 2.5\,\text{GeV}$ (generator)	1.685 %
$\geq 3\times$ particles with $\Delta R(\mu, h) < 1.7$, $\lvert\eta_h\rvert < 2.5$ and $p_{\perp,h} > 1\,\text{GeV}$ (generator)	1.152 %
$0.41\,\text{GeV} < m(hhh) < 1.85\,\text{GeV}$ (generator)	0.722 %
μ-track reconstructed	0.031 %
primary vertex reconstructed	0.031 %
$1 \times \mu$ with $p_{\perp,\mu} > 4.5\,\text{GeV}$, or $\geq 2 \times \mu$ with $p_{\perp,\mu} > 2.5\,\text{GeV}$ reconstructed	0.019 %
B_s^0-candidate reconstructed	0.010 %

For this analysis, a sample with $\approx 7.2 \cdot 10^9$ pp-collisions was used.

4.7 Pseudo-Code for Candidate Building

To give a better overview of the stage of the signal production, a pseudo-code for CMSSW (corresponding to the path) is provided.

```
1 Sig + BG: Generate Events with Pythia. [source]

2 BG: Decay particles with EvtGen. [EDProducer]

3 Sig: Pythia-filter: Event comprises μ and τ from B_s^0, |η_μ| < 2.5.
  [EDFilter]

4 Sig+BG: Dump generator information about produced particles.
  [EDAnalyzer]

5 Sig + BG: Generator filter: [EDFilter]
  Sig: B_s^0 → μτ in event, p_⊥,μ > 2.5 GeV, p_⊥,π > 1 GeV.
  BG: p_⊥,μ > 4.5 GeV (1 × μ), p_⊥,μ > 2.5 GeV (≥ 2 × μ),
  ΔR(μ, h) < 1.7, |η_h| < 2.5 and p_⊥,h > 1 GeV,
  0.41 GeV < m(hhh) < 1.85 GeV.
```

6 Sig + BG: FAMOS [*EDProducer*]

7 Sig + BG: Muon-Filter: Ensure that there are reconstructed μ tracks. [*EDFilter*]

8 Sig + BG: Dump track information. [*EDAnalyzer*]

9 Sig + BG: Build B_s^0-candidate. [*EDAnalyzer*]
Find primary vertex (abort path if failed).
$p_{\perp,\mu} > 4.5\,\text{GeV}\,(1 \times \mu)$, $p_{\perp,\mu} > 2.5\,\text{GeV}\,(\geq 2 \times \mu)$,
$\Delta R(\mu, h) < 1.7 (\geq 3\times)$,
$|\eta_h| < 2.5$,$p_{\perp,h} > 1\,\text{GeV}$,
$0.41\,\text{GeV} < m(hhh) < 1.9\,\text{GeV}$, $Q(hhh) \neq Q(\mu)$,
fit 3 h tracks to vertex, refit 3 h tracks using vertex as an additional point.
Store $\mathbf{p}_\mu, \mathbf{p}_{h,1-3}$, tau-vertex, primary vertex, POCA

10 Save all gathered information in path in ntuple. [*EDAnalyzer*]

5 Resolutions, the k-Factor, and All That

In this chapter, some introductory calculations and plots shall be presented, which do not directly contribute to the analysis of $B_s^0 \to \mu\tau$ or its branching fraction. First, the resolution of the tau-vertex is presented, which denotes a crucial number when secondary vertex reconstruction is needed. Then, a rather technical detail is exposed: The parameterization of the muon track with a straight line or with a helix, incorporating the magnetic field of CMS. The superiority of the neutrino-reconstruction, compared with a simpler method is discussed in 5.3. As a last point, a classification of the background of the decay $B_s^0 \to \mu\tau$ is made, that may contribute to the study of possible vetoes in a further analysis.

5.1 Resolution of the Tau-Vertex

To test the quality of the neutrino reconstruction (which will be fully described in section 6.2.1), the resolution of the tau-vertex was plotted, i.e. the fitted vertex was compared with the vertex-information of the generator. The distribution then was fitted with a double-gaussian (which is the sum of two gaussians, with different integral, sigma and mean), consider fig. 5.1 and 5.2. While the values of the x- and y-direction are comparable, the distribution in z-direction has much broader tails (also visible in table 5.1). This is mostly due to the larger pixel-size in z-direction (see section 1.3).

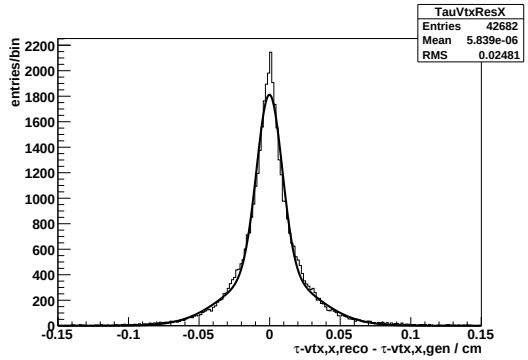

Figure 5.1: Resolution of τ-vertex., x-direction.

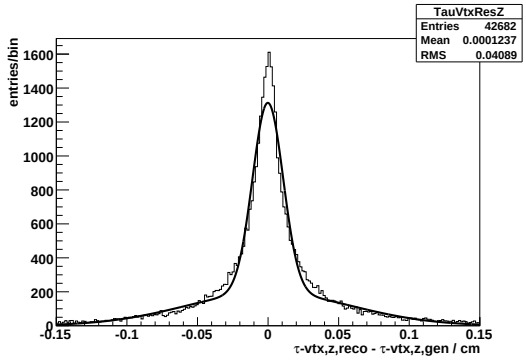

Figure 5.2: Resolution of τ-vertex., z-direction.

Table 5.1: Parameters of double gaussian for τ-vtx resolution. The errors given in this table correspond to the errors of the fitting.

x-Resolution		**z-Resolution**	
$\bar{x}_{wide}$	$(0 \pm 2)\ \mu$m	$\bar{x}_{wide}$	$(3 \pm 4)\ \mu$m
σ_{wide}	$(300 \pm 4)\ \mu$m	σ_{wide}	$(581 \pm 7)\ \mu$m
$\int_{wide}$	(410 ± 10) cm	$\int_{wide}$	$195 \pm 4)$ cm
$\bar{x}_{narrow}$	$(-1 \pm 1)\ \mu$m	$\bar{x}_{narrow}$	$(-2 \pm 1)\ \mu$m
σ_{narrow}	$(90 \pm 2)\ \mu$m	σ_{narrow}	$(110 \pm 1)\ \mu$m
$\int_{narrow}$	(1400 ± 20) cm	$\int_{narrow}$	(1120 ± 10) cm
$\frac{\int_{narrow}}{\int_{wide}}$	3.40 ± 0.1	$\frac{\int_{narrow}}{\int_{wide}}$	5.8 ± 0.1

5.2 Reconstruction on a Line and on a Helix

In a first approach the muon track was parametrized with a line, using the formula:

$$\vec{r}(x) = \vec{r}_{POCA} + x \cdot \hat{p}_\mu \tag{5.1}$$

where $\vec{r}_{POCA}$ is the vector from the primary vertex to the Point of Closest Approach, $\hat{p}_\mu$ is the normalized momentum vector of the muon, and x is the distance parameter on the line, with $x_{POCA} = 0$.
As the tracks of charged particles are bent in a magnetic field, hence in the solenoid field of CMS, their trajectory is a helix rather than a line. To account for this, in a second step, the muon track was parametrized with a helix (see [16]):

$$\vec{r}(x) = \vec{r}_{POCA} + \frac{\gamma}{Q}(\theta - \sin\theta) \cdot \hat{H} + \frac{\sin\theta}{Q} \cdot \hat{T}_0 + \frac{\alpha}{Q}(1 - \cos\theta) \cdot \hat{N}_0 \tag{5.2}$$

where $\vec{r}_{POCA}$ is defined as above, $\hat{H} = \frac{\vec{B}}{|\vec{B}|}$ is the normalized vector of the magnetic field, $\hat{T} = \frac{\vec{p}}{|\vec{p}|}$ is the normalized momentum vector of the muon in the Point of Closest Approach, $\hat{N}_0 = \frac{\vec{H}\times\vec{T}}{|\vec{H}\times\vec{T}|}$, $\alpha = \vec{H} \times \vec{T}$, $\gamma = \vec{H} \cdot \vec{T}$, $Q = -|\vec{B}| \cdot \frac{q}{|\vec{p}|}$ with q the charge of the muon, $\theta = Q \cdot x$ (with x the distance on the helix from the starting point) and $B = 0.3 \cdot 10^{-3} \times$ Field in units of kGauss, if p is given in units of GeV and x is given in units of cm.
To estimate the difference between the reconstruction done with the line and with the helix, in a first step, the distance between the computed decay vertex of the B_s^0 on the line and on the helix was calculated. The result is shown in fig. 5.3. It is clearly visible that the distance rarely exceeds 1 μm.
In a second step, the whole reconstruction of the B_s^0 was done with the line and with the helix simultaneously; the corresponding mass distributions (shown in fig. 5.4 and 5.5) are fitted with a double gaussian to account for the asymmetry. The corresponding means and sigmas of each of the two gaussians are shown in table 5.2.

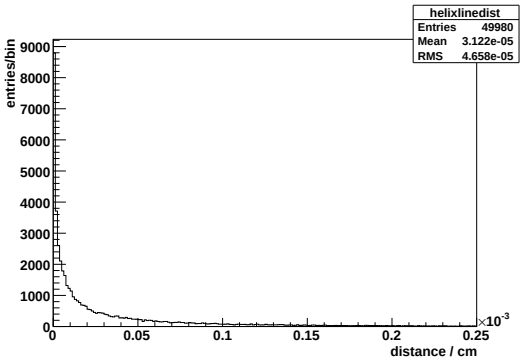

Figure 5.3: Distance between B_s^0-vertex., reconstructed on a line and on a helix.

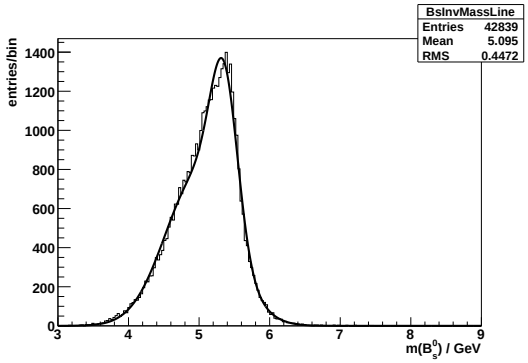

Figure 5.4: B_s^0-mass distribution, where μ track is a line.

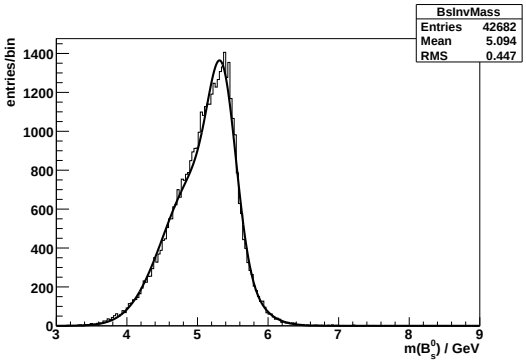

Figure 5.5: B_s^0-mass distribution, where μ track is a helix.

It can easily be seen that the difference between the reconstruction with a line and a helix is negligible. For consistency a helix was used further on in the analysis.

5.3 Neutrino Reconstruction vs. k-Factor Method

Instead of reconstructing the neutrino with the method described in chapter 3, one could use the so called k-factor method. The k-factor method assumes that the neutrino and the three-prong are emitted parallel to the flight direction of the tau (which of course is an approximation in any case). The k-factor itself is defined as:

$$m(\tau) = m(3h) \cdot (1 + k) \Rightarrow k = \frac{m(\tau)}{m(3h)} - 1,$$

where the tau mass is known and the mass of the three-prong is reconstructed out of the tracks and a mass hypothesis for the hadrons. This can then be applied to the four-momentum of the tau:

$$\mathbf{p}_\tau = \mathbf{p}_{3h} \cdot (1 + k).$$

Table 5.2: Parameters of double gauss for line and for helix.

Line		**Helix**	
$\bar{x}_{wide}$	4.980 GeV	$\bar{x}_{wide}$	4.978 GeV
σ_{wide}	0.469 GeV	σ_{wide}	0.467 GeV
$\bar{x}_{narrow}$	5.354 GeV	$\bar{x}_{narrow}$	5.355 GeV
σ_{narrow}	0.194 GeV	σ_{narrow}	0.194 GeV
entries	42839	entries	42682

Using the four-momentum of the muon, the B_s^0 can be reconstructed. The $m(B_s^0)$-distribution is shown in fig. 5.7, compared to the mass peak, obtained with the neutrino reconstruction (fig. 5.6).

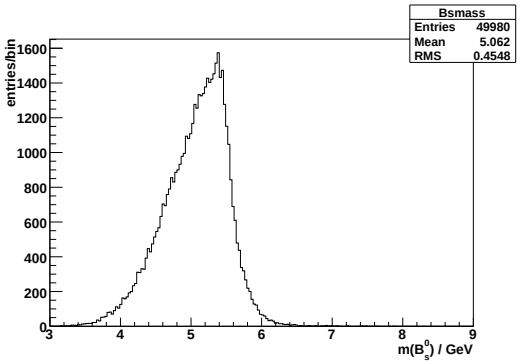

Figure 5.6: B_s^0-mass distribution, using neutrino reconstruction.

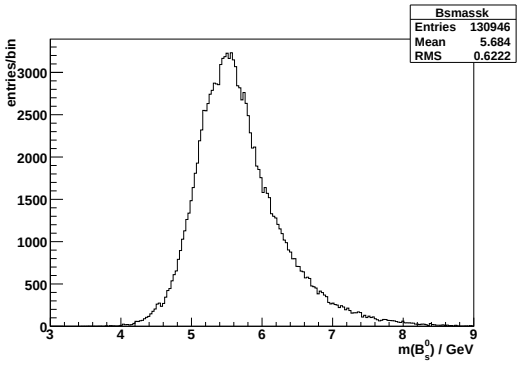

Figure 5.7: B_s^0-mass distribution, using the k-Factor method.

Note that while the number of entries is about three times as large with the k-factor method (this will be further exposed in section 6.3.1), the RMS is about 180 MeV smaller with the neutrino reconstruction. The angle between the three-prong momentum and the neutrino momentum (using generator information) is shown in fig. 5.8. Due to the fact that most of the neutrinos are emitted with an angle greater than 0.2 rad, the mass peak is slightly shifted to higher values.

5.4 Classification of the Background

To have an overview what particle decays the background comprises, the (real) particle ID of the particles of the selected three-prong candidates and of the muon candidate were checked using generator information. The results are shown in table 5.3.
For this task, 71467 muon candidates were analyzed in the background sample. Note that the identification for the muon is $\approx$ 100 %, whereas in the three-prong, kaons and protons make up a considerable amount of the background. This of course is the case, as the CMS detector provides no particle identification. Note that the baryons in the table appear most probably due to a missmatch of the truthmatching, as it is possible, but very unlikely, that the flight length of a baryon is big enough to reach the tracker. An important number in this context is the multiplicity of the pions

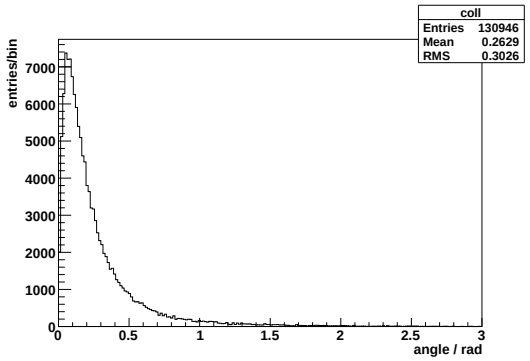

Figure 5.8: Angle between $\vec{p}_{3h}$ and $\vec{p}_{\nu}$.

Table 5.3: Particle ID of μ and $3h$-type (background).

	Particles		**Anti-Particles**	
μ	μ^-	49.7 %	μ^+	50.3 %
$3h$	π^+	35.3 %	π^-	35.6 %
	K^+	11.0 %	K^-	10.9 %
	p^+	2.2 %	p^-	2.1%
	μ^-	0.8 %	μ^+	0.7 %
	e^-	0.7 %	e^+	0.7 %
	$\Omega^-, \Xi^-, \Sigma^\pm$	< 0.1 %	$\Xi^+, \bar{\Sigma}^\pm$	< 0.1 %

in the three-prong, which is shown in fig. 5.9. Only in roughly one third of all the reconstructed decays the three-prong entirely consists of pions.
The muon candidates and the particles of the three-prong candidates were checked for their mother-particles as well (consider table 5.4 and 5.6). The mothers of the muons are mostly mesons containing a b- or a c-quark, whereas the hadrons composing the three-prong have a much wider spectrum of particle-mothers, from light mesons to charmed baryons.
To study the possible introduction of vetoes in a further analysis, the ID of the three-prong mother particles was studied, if all the particles of the three-prong derived from the same mother. The results are exposed in table 5.5, whereas the multiplicity of the mothers for a three-prong is shown in fig. 5.10. Only in roughly 4 % all the three-prong particles have the same mother.

Table 5.4: Particle ID of μ-mother (background).

Particles		**Anti-Particles**	
D^0	11.9 %	$\bar{D}^0$	12.1%
B^0	10.9 %	$\bar{B}^0$	10.5 %
B^+	9.6 %	B^-	9.5 %
D^+	7.8 %	D^-	7.6 %
D_s^+	3.4 %	D_s^-	3.3 %
B_s^0	3.0%	$\bar{B}_s^0$	2.9%
Λ_b^0	1.2 %	$\bar{\Lambda}_b^0$	3.3 %
Λ_c^+	0.6 %	Λ_c^-	0.6 %
D_s^+	3.4 %	D_s^-	3.3 %
η	0.4%		
$\omega(782)$	0.3 %		
$\phi(1020)$	0.3 %		
$\Xi_c^0, \Xi_c^+, \Xi_b^0, \Xi_b^-, \Omega_c^0, \Lambda_c^0$	< 0.1 %	$\bar{\Xi}_c^0, \Xi_c^-, \bar{\Xi}_b^0, \Xi_b^+, \bar{\Omega}_c^0, \bar{\Lambda}_c^0$	< 0.1 %

Table 5.5: Particle ID of h-mother, if all h have the same mother (background).

Particles		**Anti-Particles**	
D^0	23.3 %	$\bar{D}^0$	21.3 %
D^+	19.3 %	D^-	18.7 %
Lund-String/Jet	6.6 %		
D_s^-	2.7 %	D_s^+	2.7
Λ_c^-	1.0 %	Λ_c^+	1.0 %
τ^-	0.9 %	τ^+	0.8 %
B^0	0.4 %	$\bar{B}^0$	0.2 %
$\omega(782)$	0.2 %		
B_s^0	0.2 %		
B^+	0.1 %	B^-	0.1 %
other particles	< 0.1 %		

Table 5.6: Particle ID of h-mother (background).

Particles		**Anti-Particles**	
Lund-string/Jet	11.0 %		
$\rho(770)^0$	11.0 %		
D^0	7.7 %	$\bar{D}^0$	7.9 %
$\omega(782)$	7.5 %		
$K^\star(892)^0$	5.7 %	$\bar{K}^\star(892)^0$	5.7 %
K_s^0	5.1 %		
$\rho(770)^-$	5.0 %	$\rho(770)^+$	4.8 %
$\phi(1020)$	3.0 %		
$K^\star(892)^+$	3.0 %	$K^\star(892)^-$	2.9 %
η	2.6 %		
D^+	2.1 %	D^-	2.1 %
$D^\star(2010)^+$	1.7 %	$D^\star(2010)^+$	1.7 %
$\eta'(958)$	1.1 %		
Δ^{++}	0.6 %	Δ^{--}	0.6 %
D_s^+	0.5 %	D_s^-	0.5 %
B^0	0.5 %	$\bar{B}^0$	0.5 %
B^+	0.4 %	$\bar{B}^-$	0.4 %
Λ_c^+	0.4 %	Λ_c^-	0.4 %
Λ	0.3 %	$\bar{\Lambda}$	0.3 %
Δ^+	0.3 %	$\bar{\Delta}^+$	0.3 %
Δ^0	0.2 %	$\bar{\Delta}^0$	0.2 %
B_s^0	0.1 %	$\bar{B}_s^0$	0.1 %
Δ^-	0.1 %	$\bar{\Delta}^-$	0.1 %
Σ^+	0.1 %	$\bar{\Sigma}^+$	0.1 %
τ^-	0.1 %	τ^+	0.1 %
$J/\Psi(1S)$	0.1 %		
Λ_b^0	0.1 %	$\bar{\Lambda}_b^0$	0.1 %
other particles	< 0.1 %		

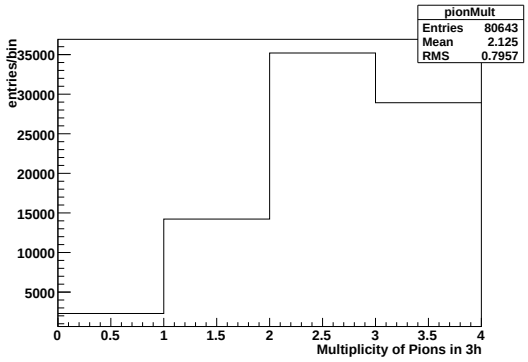

Figure 5.9: Multiplicity of pions in three-prong (background).

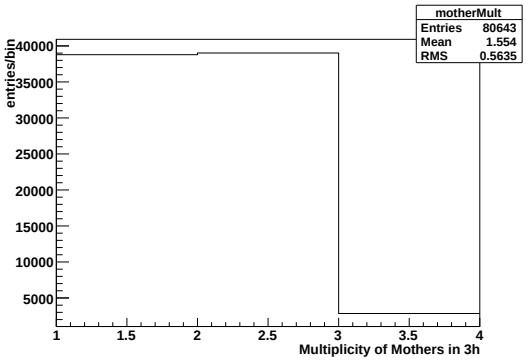

Figure 5.10: Multiplicity of mothers of particles of three-prong (background).

6 Analysis

The goal of the analysis was to reconstruct the B_s^0 out of the candidates and to develop methods to distinguish signal from background. The two main issues in the B_s^0 reconstruction are the candidate selection – to single out the candidates that do not correspond to the decay $B_s^0 \to \mu\tau$ – and the application of the neutrino reconstruction. The neutrino reconstruction as a method has already been introduced in an earlier chapter, but now a more thorough and technical discussion will be given. In the second half of the chapter, a cut-based analysis is presented, that was used to reduce the background.

6.1 Candidate Selection

As the cuts on CMSSW-level were rather loose, most of the time more than one B_s^0 candidate per event was stored. Because only one real $B_s^0 \to \mu\tau$ decay happened in every event in the signal sample (if more than one $B_s^0 \to \mu\tau$ decay was present, the event was discarded at the beginning), some criteria had to be developed to select the right one with the highest efficiency ε, where

$$\varepsilon = \frac{\text{selected good events}}{\text{all good events}}.$$

This was done in two steps: First the number of candidates per event was decreased, using cuts on variables that could be easily accessed (meaning that they did not need the full reconstruction of the B_s^0, namely the neutrino reconstruction could be omitted), in the following called «candidate selection cuts». In second step, among the remaining candidates, a single one was selected.

6.1.1 Variables and Candidate Selection Cuts

The used variables and their conditions were:

- $Q(3h) = -Q(\mu)$
- $\Delta R(\mu, 3h) < 1.2$
- $\Delta R(\tau\text{-vtx}, p_{\mu,3h}) < 0.3$
- $0.45\,\text{GeV} < m(3h) < 1.65\,\text{GeV}$

The distribution of the variables are shown in figs. 6.2, 6.3, 6.4 and 6.5. An illustration for $\Delta R(\tau\text{-vtx}, p_{\mu,3h})$ is shown in fig. 6.1. Fig. 6.10 and 6.11 then show how the application of the candidate selection cuts altered the number of candidates.
While the condition $Q(3h) = -Q(\mu)$ is just a consequence of the properties of the decay $B_s^0 \to \mu\tau$ (and hence represents no cut on (real) $B_s^0 \to \mu\tau$-decays), the other three conditions make

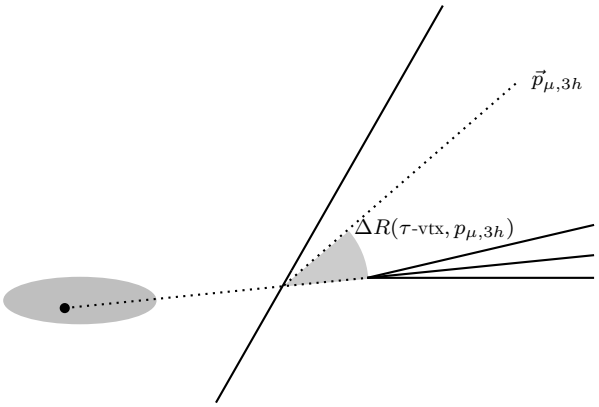

Figure 6.1: Illustration for $\Delta R(\tau\text{-vtx}, p_{\mu,3h})$.

constraints on the distribution of the variables, that also affect the (real) signal candidates. However, as these cuts are used to select a candidate rather than reducing the background (and therefore not cutting strongly into the signal distributions), the influence on the signal statistics with an efficiency of $\varepsilon = 90.4\%$ is small.

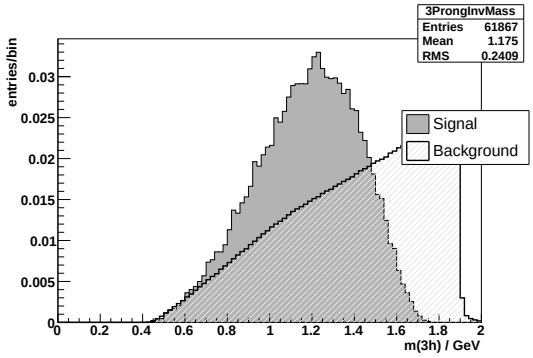

Figure 6.2: Distribution of $m(3h)$ for signal and background, used for candidate selection cuts.

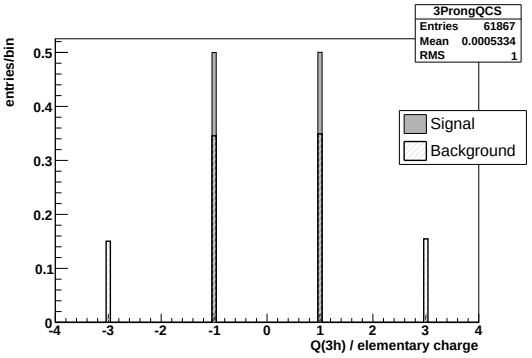

Figure 6.3: Distribution of $Q(3h)$ for signal and background, used for candidate selection cuts.

6.1.2 Selection of a Single Candidate

The best candidate was chosen to be the one with the largest s_{3d}, where

$$s_{3d} = \frac{l_{3d}}{\sigma_{3d}} = \frac{\text{dist(prim-vtx - } \tau\text{-vtx)}}{\text{error}} \tag{6.1}$$

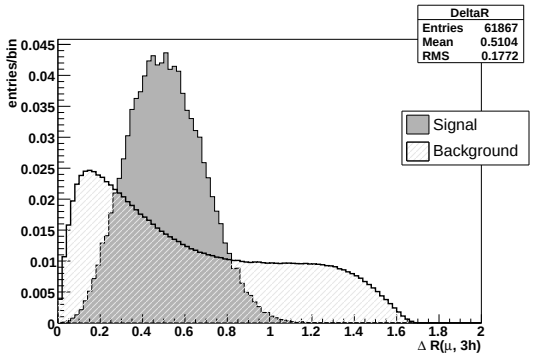

Figure 6.4: Distribution of $\Delta R(\mu, 3h)$ for signal and background, used for candidate selection cuts and the potential selection of a single candidate.

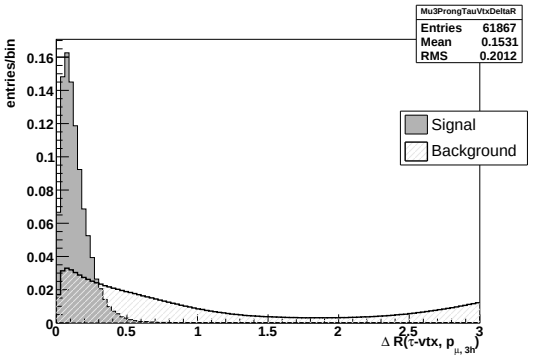

Figure 6.5: Distribution of $\Delta R(\tau\text{-vtx}, p_{\mu,3h})$ for signal and background, used for candidate selection cuts and the potential selection of a single candidate.

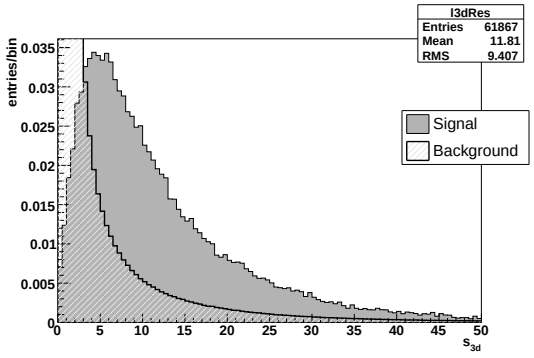

Figure 6.6: Distribution of s_{3d} for signal and background, used for the selection of a single candidate.

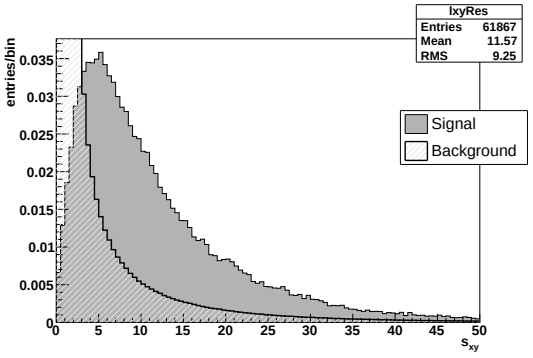

Figure 6.7: Distribution of s_{xy} for signal and background, used for the potential selection of a single candidate.

6.1.3 Efficiency of the Candidate Selection

The efficiency of the candidate selection with the described selection method (cuts and selection of one candidate) is 80.8 %. The achieved efficiencies with other selections variables are summarized in table 6.1. The second column takes *only* the selection of one candidate into account (meaning that 100 % is calculated after the application of the candidate selection cuts), whereas the third column represents the overall efficiency, incorporating candidate selection cuts and selection of the best candidate. To determine these efficiencies, a sample with 61'867 $B_s^0 \to \mu\tau$ decays was used. s_{xy} is defined similar to s_{3d} in (6.1), where the distance is measured only in the x-y-plane. The angle $\angle(\tau\text{-vtx}, p_{\mu,3h})$ is similar to $\Delta R(\tau\text{-vtx}, p_{\mu,3h})$, but the angle between the two vectors is computed

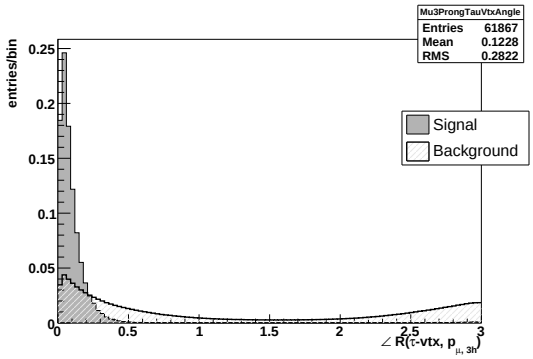

Figure 6.8: Distribution of $\angle R(\tau\text{-vtx}, p_{\mu,3h})$ for signal and background, used for the potential the selection of a single candidate.

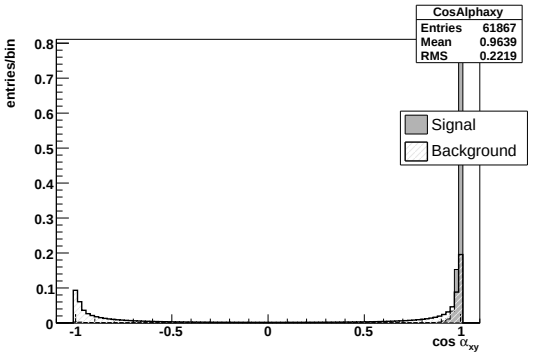

Figure 6.9: Distribution of $\cos\alpha_{xy}$ for signal and background, used for the potential selection of a single candidate.

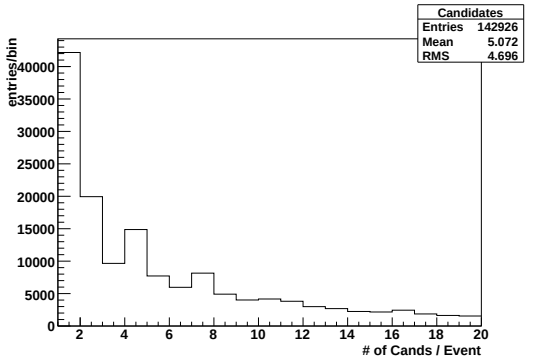

Figure 6.10: Distribution of number of candidates per event, before candidate selection cuts.

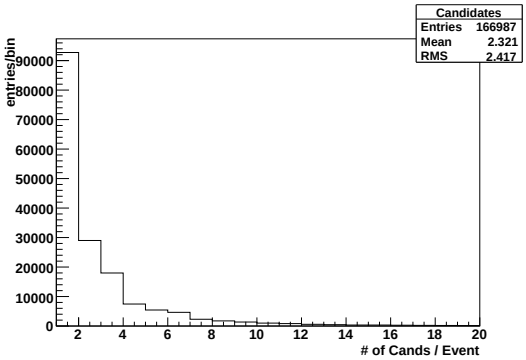

Figure 6.11: Distribution of number of candidates per event, after candidate selection cuts.

rather than ΔR. The variable α_{xy} is the projection of $\angle(\tau\text{-vtx}, p_{\mu,3h})$ in the x-y-plane.

6.2 Neutrino Reconstruction

6.2.1 Restriction to the Theory and Application

The concept of the neutrino reconstruction is explained in chapter 3. To be able to apply it in a physics analysis, several aspects have to be considered.

- $p_{\mu\perp B_s^0} = p_{\tau\perp B_s^0}$ will never be fulfilled in reality. On the one hand this is due to the detector

Table 6.1: Efficiency for different selection variables.

Variable to select one candidate	Fraction, only best Candidate	Overall Fraction
s_{3d}	89.3 %	80.8 %
s_{xy}	89.3 %	80.8 %
$\Delta R(\tau\text{-vtx}, p_{\mu,3h})$	76.3%	69.0 %
$\angle(\tau\text{-vtx}, p_{\mu,3h})$	76.1%	68.8 %
$\cos(\alpha_{xy})$	76.0 %	68.8 %
$\Delta R(\mu, 3h)$	64.1 %	58.0 %

resolution, on the other hand the precision is limited by numerics. Therefore the quantity $p_{\perp,bal} = |\vec{p}_{\mu\perp B_s^0} + \vec{p}_{\tau\perp B_s^0}|$ has to be minimized.

- The minimization was done with Migrad from TMinuit, which needs a starting point. Hence the muon track had to be segmented and a starting point had to be chosen in every segment.

- The segmentation started at the POCA of the muon track. As most of the muons are emitted with the projection of the flight direction of the B_s^0 on the muon track and not against it, the interval to be segmented was chosen asymmetrically, from -0.5 cm to 2.0 cm (where 0 cm is the muon-POCA and positive values are with the flight direction of the muon). Consider fig. 6.13.

- The segments were chosen to be of a length of 10 μm.

- As the minimization as a whole is slow, several methods were introduced to reduce the number of minimization procedures. First of all, a fast, but less precise Monte Carlo minimization was performed on the whole region of the segmented muon track. The $m(B_s^0)$ then was reconstructed (if the neutrino reconstruction was successful) and the event was discarded, if $m(B_s^0) > 8\,\text{GeV}$. This was considered safe as it is very unlikely for the mass to «wander» in the region of the peak with the more accurate Migrad minimization (less than 0.3 % of the events in the signal sample were discarded due to this constraint). The mass peak, obtained with only the Monte Carlo minimization is depicted in fig. 6.12.

- The minimization with Migrad only started, if the neutrino could be reconstructed in the starting point of the segment (with the starting point chosen to be in the middle of the segment).

- To terminate the minimization if the convergence is not good enough, a maximum number of convergence loops had to be chosen (500).

- As there are always two solutions for the neutrino reconstruction, the one with the lowest $p_{\perp,bal}$ was taken. This of course allows no information about the solution being the «real» one (i.e. the one realized in the event).

Two main restrictions occur when the neutrino reconstruction is applied in the environment of a detector:

- Not every neutrino can be reconstructed: This is mostly due to resolution issues, which leads to imaginary solutions of the square root in equation 3.3.
- Migrad sometimes may converge, even though the minimum is only a local and not a global one (because of the restriction in numerical precision).

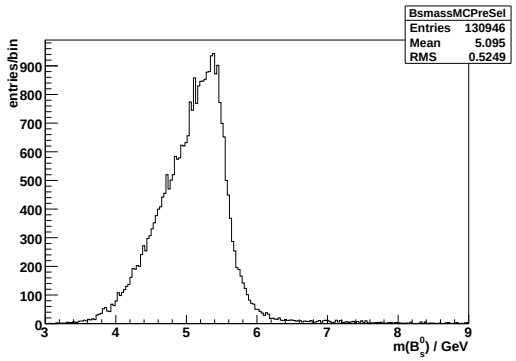

Figure 6.12: $m(B_s^0)$, minimization done with Monte Carlo minimization. Note that the RMS is larger compared with fig. 5.6, done with the full Migrad minimization.

6.2.2 Pseudo-Code for Neutrino Reconstruction

```
1  Segment muon track, define starting point in segment

2a Calculate first solution of neutrino reconstruction for start-
   ing point in segment. Fail: Go to 2b. Success:

3a Run Migrad for first solution of neutrino reconstruction,
   till it converges or reaches end of convergence loops. Cal-
   culate p⊥,bal for the convergence point. If p⊥,bal < p⊥,bal,best:
   Store convergence point, set p⊥,bal,best = p⊥,bal.

2b Calculate second solution of neutrino reconstruction for
   starting point in segment. Fail: Go to next segment. Suc-
   cess:

3b Run Migrad for the second solution of neutrino reconstruc-
   tion, till it converges or reaches end of convergence loops.
   Calculate p⊥,bal for the convergence point. If p⊥,bal < p⊥,bal,best:
   Store convergence point, set p⊥,bal,best = p⊥,bal.

4  Repeat procedure for all segments.
```

5 Result: Convergence point on muon track (= decay vertex of B_s^0), corresponding to smallest $p_{\perp,bal}$.

6.3 Efficiencies

The efficiencies for the signal and the background sample for the B_s^0-reconstruction are listed in table 6.2 and 6.3. For the signal, a sample with 61'867 (real) $B_s^0 \to \mu\tau$ events was used, while the background sample comprised 685'230 B_s^0 candidates.

6.3.1 Efficiencies for Signal

Table 6.2: Efficiency for signal, B_s^0-reconstruction.

Step in B_s^0 reconstruction	**Fraction**
Events that contain at least one B_s^0 cand	100%
Events that contain only one B_s^0	86.3%
Select one Candidate	
$p_{\perp,\mu} > 5\,\text{GeV}$ and $m(3h,\mu) < 8\,\text{GeV}$	61.3%
μ- and h-truthmatching successful	59.3%
$3h$ particles are πs, mother of πs is a τ	48.1 %
Successful ν-reconstruction	18.6 %

As the probability of observing a decay $B_s^0 \to \mu\tau$ more than once in a event is very small, the event was discarded if more than one B_s^0 meson was found (as B_s^0 was forced to $B_s^0 \to \mu\tau$).
This calculation takes the effect of the truthmatching into account, as only real $B_s^0 \to \mu\tau$ decays were desired for the reconstruction and the further analysis (and no decays of the background present in the signal sample).
Note that the efficiency for the neutrino reconstruction itself is 35.7 % (mostly because of complex solutions of the quadratic equation, as mentioned in 6.2.1). Further note that these numbers don't tell anything about the efficiency of the candidate selection.

6.3.2 Efficiencies for Background

As well as for the signal sample, the efficiencies in the background sample (concerning the reconstruction) were calculated, and are shown in table: 6.3. Note that the steps incorporating truthmatching are skipped, as no truthmatching was necessary for the background.
The efficiency for the neutrino reconstruction in the background sample is with a percentage of 40.2 % higher than in the signal-sample, whereas more than a third of the events is discarded due to a mass larger than 8 GeV.

Table 6.3: Efficiency for background, B_s^0-reconstruction.

Step in B_s^0 reconstruction	**Fraction**
Events that contain at least one B_s^0 cand	100%
Select one Candidate	
$p_{\perp,\mu} > 5\,\text{GeV}$ and $m(3h,\mu) < 8\,\text{GeV}$	25.2%
Successful ν-reconstruction	11.0 %

6.4 Cut-based Analysis

To reduce as much of the background as possible, while trying to maintain most of the signal entries, a cut-based analysis was used. It was splitted in two steps:
In a first step, four variables were chosen for the application of pre-selection cuts.
In a second step, three variables were used as factorizing cuts. The need for factorizing some of the cuts is the following: If all the cuts on all seven variables would be applied simultaneously, no background events would remain in the mass region of the B_s^0 of interest. As the background sample was generated with an integrated luminosity of $\mathcal{L}_{prod} = 2.28\,\text{pb}^{-1}$, but the analysis of real CMS data will be applied on a integrated luminosity of $\mathcal{L}_{prod} \approx 10\,\text{fb}^{-1}$, it would not have been possible to scale up the background distribution to the target luminosity.
However, to get a background distribution for $10\,\text{fb}^{-1}$ nevertheless, the following procedure was chosen: All the preselection cuts were applied simultaneously. This left enough statistics in the signal as in the background sample. Then a factorizing cut was applied, where the number of events in the mass region of $4.8\,\text{GeV}$ to $5.6\,\text{GeV}$ were counted before and after applying the cut. This was done for all three factorizing cuts. The factors corresponding to each factorizing cut were multiplied and hence the (non-integer) number of background entries in the mass region could be computed. How this procedure is applied to the calculation of the upper limit of the branching fraction is discussed in section 7.

6.4.1 Pre-Selection Cuts

The following variables were chosen for the pre-selection cuts:

Table 6.4: Pre-Selection cuts (hard).

Pre-Selection Cut-Variable	**Cut-Value**
$m(3h)$	$1\,\text{GeV} < m(3h) < 1.65\,\text{GeV}$
μ-DOCA	μ-DOCA > 0.2
xbest	0.3 < xbest < 1.5
χ^2/Ndof	χ^2/Ndof < 2

μ-DOCA and xbest are visualized in fig. 6.13, where DOCA stands for Distance Of Closest Approach (the distance between the POCA (Point of Closest Approach) and the primary vertex.) χ^2

Table 6.5: Pre-Selection cuts (loose).

Pre-Selection Cut-Variable	**Cut-Value**
$m(3h)$	$1\,\text{GeV} < m(3h) < 1.65\,\text{GeV}$
μ-DOCA	μ-DOCA > 0.1
xbest	0.1 < xbest < 1.5
χ^2/Ndof	χ^2/Ndof < 2

is taken from the tau vertex and Ndof are the number of degrees of freedom of the χ^2-fit.
The cuts were optimized by «eye», meaning that they were not searched optimizing a pre-defined quantity. Note that there are two versions of the pre-selection cuts, looser and harder ones. This is mainly a technical feature for calculating the upper limit, see section 7.1.
The variables with their corresponding cuts are shown in figs. 6.14 - 6.17.

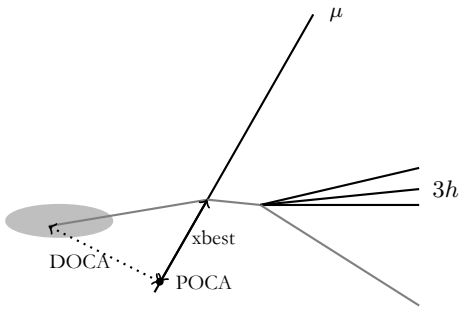

Figure 6.13: Illustration for POCA and xbest.

6.4.2 Factorizing Cuts

The following variables were chosen as factorizing cuts:

- $p_{\perp,3h}$
- s_{3d}
- IsolVar

As these cuts were altered during the optimization of the upper limit, only the variables themselves are given. The optimal cuts and the region they were searched for are described in the section «7.2.3 Grid-Search».
$p_{\perp,3h}$ stands for the transverse momentum of the three-prong, and s_{3d} is defined as in section 6.1.2. The isolation variable is defined as follows:

$$\text{IsolVar} = \frac{p_{B_s^0\perp}}{p_{B_s^0\perp} + \sum_{trk,trk\neq B_s^0} p_\perp} \tag{6.2}$$

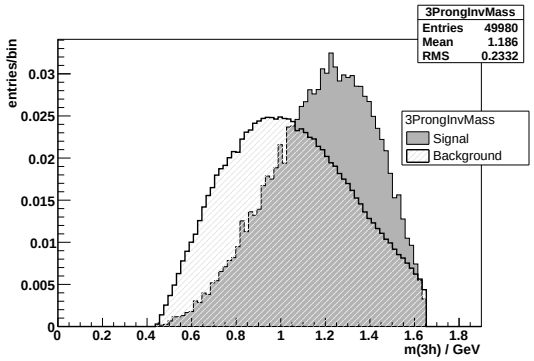

Figure 6.14: Distribution of $m(3h)$ for signal and background, used as a pre-selection cut. The steep decrease at 1.65 GeV is a result of the applied candidate selection cut.

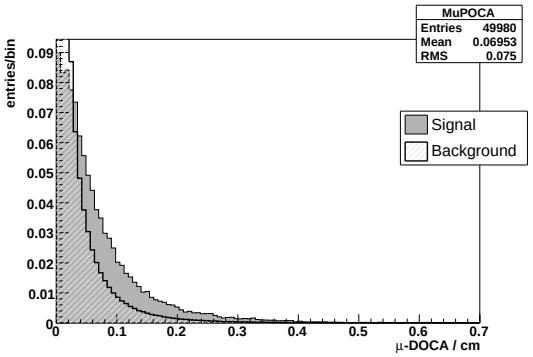

Figure 6.15: Distribution of μ-DOCA for signal and background, used as a pre-selection cut

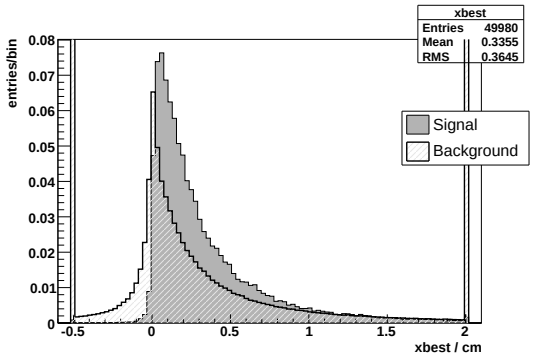

Figure 6.16: Distribution of xbest for signal and background, used as a pre-selection cut. The two «towers» in the background arise, if the minimum is situated at x < -0.5 cm or x > 2 cm, respectively.

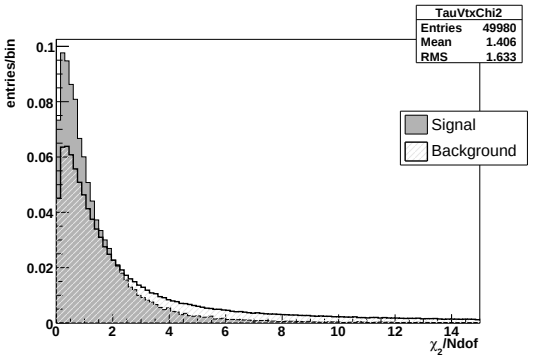

Figure 6.17: Distribution of χ^2 / Ndof for signal and background, used as a pre-selection Cut

where $trk \neq B_s^0$ means that the sum is not over the particles that are used to reconstruct the B_s^0, i.e. the muon and three-prong. Associated with the isolation variable is a cone width, in which the tracks are considered. In the analysis, four different cone widths with $\Delta R(B_s^0, \mu/\pi) = 0.8/1/1.2/1.4$ were considered, to have a greater flexibility to optimize the cuts.

An illustration for the isolation variable is given in fig. 6.18, the factorizing cut variables are shown in figs. 6.19 - 6.24.

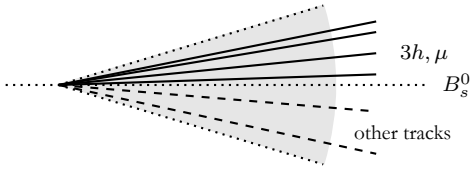

Figure 6.18: Illustration for the isolation variable.

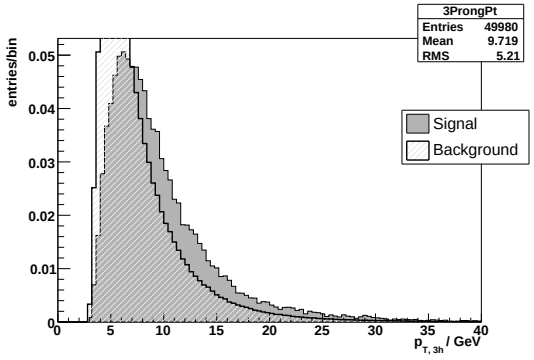

Figure 6.19: Distribution of $p_{\perp,3h}$ for signal and background, used as a factorizing cut.

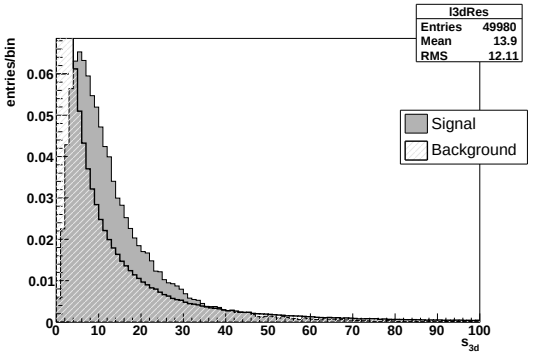

Figure 6.20: Distribution of s_{3d} for signal and background, used as a factorizing cut.

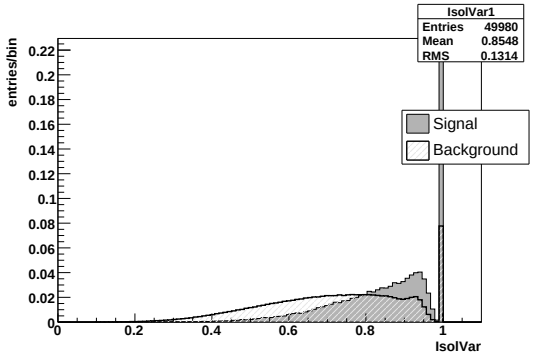

Figure 6.21: Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 0.8$ for signal and background, used as a factorizing cut.

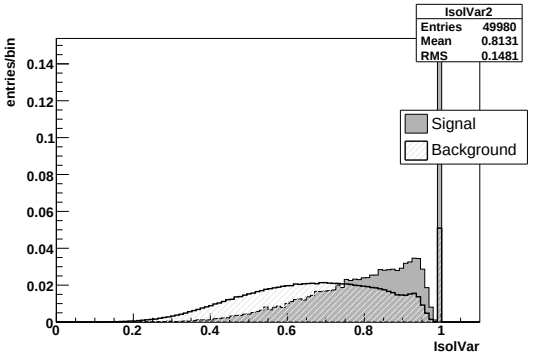

Figure 6.22: Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1$ for signal and background, used as a factorizing cut.

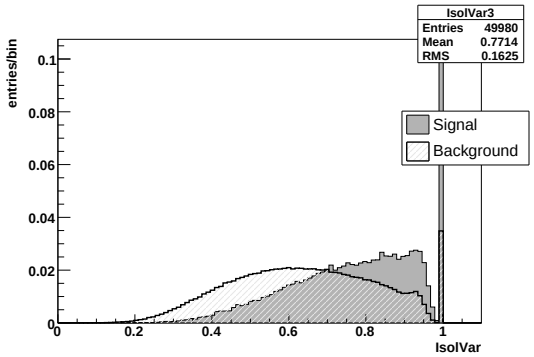

Figure 6.23: Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.2$ for signal and background, used as a factorizing cut.

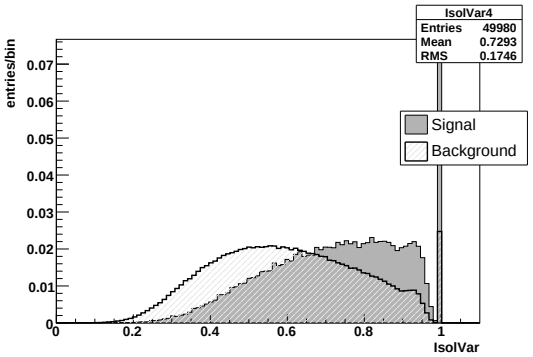

Figure 6.24: Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.4$ for signal and background, used as a factorizing cut.

7 Determination of the Upper Limit

The upper limit of a branching fraction, with an confidence level associated, is a means to specify the frequency of a particle decay, maybe without ever having observed it. In this chapter, a detailed explanation will be given, how the signal and background distributions were calculated and added, and how the upper limit of $B_s^0 \to \mu\tau$ was determined, using a Bayesian approach.

7.1 Addition of Signal and Background Distribution

As mentioned in section 6.4, cuts were applied to the signal and background sample to get rid of as much background events as possible while maintaining as much signal events as possible. As the background and signal events consisted of one separate sample each, they had to be merged together consistently. This was done in the following way:

- All the pre-selection cuts (loose constraints) and one factorizing cut were applied on the signal and background sample, the number of events in the mass-region of $4.8\,\text{GeV} - 5.6\,\text{GeV}$ with the pre-selection cuts and with the pre-selection cuts and the factorizing cut were counted and the ratio of these numbers was calculated. The loose constraints were introduced to avoid cutting away the whole background with only one factorizing cut applied. The assumption is that the shape of the background distribution is only altered insignificantly if, instead of the harder constraints, the looser ones are applied.

- This procedure was repeated for every factorizing cut, resulting in three ratios, which were multiplied.

- If this factor equaled zero (meaning that the loose pre-selection cuts and a single factorizing cut left no more events in the mentioned mass region), this combination of cuts was not considered any further and the process was aborted (see section 7.2.3 for details on this).

- The background distribution (filled in a histogram), obtained by only applying the harder pre-selection cuts was fitted with an exponential distribution from 4 GeV to 6 GeV. The signal-distribution (filled in a histogram), obtained by only applying the (harder) pre-selection cuts was fitted with a double gaussian distribution from 4 GeV to 6 GeV. The number of entries in these histograms from 4.8 GeV to 5.6 GeV were counted for the signal and the background sample. These numbers were multiplied by the multiplied ratios of the factorizing cuts (for signal and background, respectively). The resulting numbers now represented the number of signal and background events in the region of 4.8 GeV to 5.6 GeV with the luminosity of the corresponding samples.

- As these counted events in the background corresponded only to an integrated luminosity of $\mathcal{L}_{prod} = 2.28\,\text{pb}^{-1}$, but the analysis will be run on real data with an integrated luminosity of $\approx \mathcal{L}_{prod} = 10\,\text{fb}^{-1}$, the background had to be scaled up to the integrated target luminosity.

  First, the number of events in the mass-region of 4.8 GeV to 5.6 GeV, for $10\,\text{fb}^{-1}$, n_{BG} was calculated, using the target luminosity, the luminosity of the sample and the counted number of events in the mass region of the background sample. Second, random numbers distributed according to the fitted exponential function were generated (Toy Monte Carlo) and filled in a histogram, till the histogram comprised n_{BG} entries in the region of 4.8 GeV to 5.6 GeV. This histogram contained the background distribution for $\mathcal{L}_{prod} = 10\,\text{fb}^{-1}$, all the cuts and selection criteria applied.

- The same procedure was applied to the signal sample. For this purpose, a branching fraction of $\mathcal{B}(B_s^0 \to \mu\tau) = 10^{-5}$ was put in, and $\mathcal{B}(\tau \to \pi\pi\pi\nu_\tau) = 0.0952$ had to be considered as well. As in the background sample, a Toy Monte Carlo Simulation was run to simulate the signal distribution for $10\,\text{fb}^{-1}$. Note that depending on the applied cuts, it is possible to get zero signal events in the mass region.

- The two distributions were then added together. An example of an added signal and background distribution can be seen in fig. 7.3.

7.2 Bayesian Approach for the Upper Limit

The upper limit of a branching fraction for a confidence level of 90 % is defined as follows:

$$\mathcal{B}(B_s^0 \to \mu\tau)_{90\%} < \frac{N(B_s^0 \to \mu\tau)_{90\%}}{N(B_s^0 \to \text{anything})}$$

where $N(B_s^0 \to \mu\tau)_{90\%}$ signifies the upper limit of the 90%-confidence level for a certain distribution of the signal and background events. To calculate this number, a Bayesian approach including uncertainties in the signal efficiency was used (see [17] and [18].) The idea is the following: Suppose you know the number of signal events s and the acceptance of these signal events ε (i.e. accounting for the non-perfect 4π-detector geometry). Then the probability to observe n events follows a Poisson distribution with mean $s\varepsilon$.

$$P(n|s,\varepsilon) = \frac{(s\varepsilon)^n}{n!} \cdot e^{-s\varepsilon}$$

As we want to know $P(s|n)$ rather than $P(n|s,\varepsilon)$, we can use Bayes theorem, which is defined as

$$P(A|C) \cdot P(C) = P(C|A) \cdot P(A),$$

where P(A) is the probability that A is true and P(A | C) is the probability that A is true under the condition that C is true. For probability density functions (p.d.f.s) this becomes:

$$P(s|n) = \frac{P(n|s)\,\Pi(s)}{\int P(n|s)\,\Pi(s)\,\text{ds}}$$

where $\Pi(s)$ is called prior p.d.f for s.
This can now be used to calculate $P(s|n)$ out of $P(n|s,\varepsilon)$:

$$P(s|n) = \int_0^\infty P(s|\varepsilon,n)\,\mathrm{d}\varepsilon$$

and

$$P(s|\varepsilon,n) = \frac{P(n|s,\varepsilon)\,\Pi(s)\,\Lambda(\varepsilon)}{\iint P(n|s,\varepsilon)\,\Pi(s)\,\Lambda(\varepsilon)\,\mathrm{d}s\,\mathrm{d}\varepsilon}$$

where $\Pi(s)$ is the prior p.d.f for s and $\Lambda(\varepsilon)$ is the prior p.d.f for ε.
In our case $\Pi(s)$ is set equal 1. For $\Lambda(\varepsilon)$, the function $\frac{\kappa(\kappa\varepsilon)^{\mu-1}e^{-\kappa\varepsilon}}{\Gamma(\mu)}$ is used. This function is characterized with the mean $\varepsilon_0 = \frac{\mu}{\kappa}$ and the variance $\sigma_\varepsilon = \frac{\sqrt{\mu}}{\kappa}$, i.e. it is enough to specify these two values as an input.
The upper limit of signal events s_{up}, corresponding to a 90 % confidence level, is then defined by:

$$90\% = \int_0^{s_{up}} P(s|n)\,\mathrm{d}s$$

with $n = s + b$ (b = number of background events).

7.2.1 Calculation of $\mathrm{N}(B_s^0 \to \mu\tau)_{90\%}$

The starting point for the calculation of $\mathrm{N}(B_s^0 \to \mu\tau)$ is s_{up}, i.e. the maximum number of signal events one can observe on a 90 % confidence level with the given signal plus background distribution. To calculate the total number of signal events, one has to consider the efficiencies from the production of the $B_s^0 \to \mu\tau$ to the final number of entries in the signal histogram after having applied all the cuts and selection criteria. The calculation was done as follows:

$$N(B_s^0 \to \mu\tau)_{90\%} = \frac{s_{up}}{\varepsilon_{CMSSW} \cdot \varepsilon_{reco} \cdot \varepsilon_{cut}}.$$

ε_{CMSSW} denotes the efficiency of the B_s^0-candidate building on CMSSW-level (see chapter 4), ε_{reco} the corresponding efficiency on the level of the reconstruction (see chapter 6). ε_{cut} is the efficiency achieved by using the best pre-selection cuts and the factorizing cuts, which will be revealed in the next section. The respective values are summarized in table 7.1.

7.2.2 Calculation of $\mathrm{N}(B_s^0 \to \text{anything})$

As there is no restriction to the decay channels of the B_s^0, just the number of produced B_s^0 for $\mathcal{L}_{prod} = 10\,\mathrm{fb}^{-1}$ had to be counted.

$$N(B_s^0)_{10\,\mathrm{fb}^{-1}} = (\#\text{pp-coll.})_{\mathcal{L}_{prod}} \cdot \frac{\sigma_{b\bar{b}}}{\sigma_{pp}} \cdot f_s \cdot 2 \cdot \frac{10\,\mathrm{fb}^{-1}}{\mathcal{L}_{prod}}$$

Table 7.1: Efficiencies of different stages.

Specification	Efficiency
ε_{CMSSW}	1.9 %
ε_{reco}	18.6 %
ε_{cut}	2.3 %
ε_{total}	$7.9 \cdot 10^{-3}$ %

7.2.3 Grid-Search

To goal of the applied cuts is to produce the lowest upper limit. Two approaches for the cut optimization were performed: First, looping simultaneously through different combination of cuts (which is referred as «grid-search»), the value $\frac{\varepsilon_{sig}}{\sqrt{\varepsilon_{bg}}}$ was maximized, where ε_{sig} corresponds to the overall cut-efficiency of the signal and ε_{bg} to the overall cut-efficiency of the background.
Second, the upper limit directly was minimized using a grid-search. As the calculation of the upper limit with the given software package for the Bayesian approach was not possible for all combination of cuts (mainly when too many background events were present), the grid-search was performed using the class TRolke of ROOT, which computes confidence intervals with additional parameters (such as efficiency) in a fully frequentist way [19]. Note that the two methods do not yield exactly the same results, the difference for s_{up} however remains smaller than 10 %. As this change in software is mainly a technical feature and does not strongly affect the upper limit, it will not be discussed any further.
Cut combinations, that did not leave any background events in the mass region of $4.8\,\text{GeV} - 5.6\,\text{GeV}$ were not considered.
To speed up the grid-search, only one of the four different isolation variables was searched. After having obtained the best combination of cuts, the upper limit was calculated with the given cut values while altering the isolation variable. The best isolation variable was the one with cone width of $\Delta R(B_s^0, \mu/\pi) = 1.2$.
The ranges for the grid-search are given in table 7.2, while the cuts for the highest fraction $\frac{\varepsilon_{sig}}{\sqrt{\varepsilon_{bg}}}$ and the smallest upper limit are given in table 7.3. Note that the cuts are the same no matter which criterion out of the two was optimized.
In fig. 7.1 and 7.2 the fraction $\frac{\varepsilon_{sig}}{\sqrt{\varepsilon_{bg}}}$ and the upper limit for varying cuts are shown, respectively. These histograms were made with three nested loops (one for $p_{\perp,3h}$, s_{3d} and IsolVar), which account for the twofold periodic pattern of the histograms. The loop-variable s is the summation of all the three single loop-variables and has no physics meaning.

7.3 The Upper Limit of $\mathcal{B}(B_s^0 \to \mu\tau)$

With the cuts described in the last section, 68'836 background events and 4 signal events (using $\mathcal{B}(B_s^0 \to \mu\tau) = 10^{-6}$) remained in the mass region of $4.8\,\text{GeV} - 5.6\,\text{GeV}$. This corresponded to a signal efficiency of the cuts, ε_{sig}, of 0.023 and a background efficiency of the cuts, ε_{bg}, of $4.3 \cdot 10^{-4}$.

Table 7.2: Search regions for grid-search.

Factorizing Cut-Variable	Range	Step Size
$p_{\perp,3h}$	3 GeV - 42 GeV	1 GeV
s_{3d}	3 - 39	1
IsolVar	0.1 - 0.99	0.01

Table 7.3: Best cuts for upper limit calculation.

Factorizing Cut Variable	Cut Value
$p_{\perp,3h}$	5 GeV
s_{3d}	7
IsolVar	0.59

The upper limit of the branching fraction of $B_s^0 \to \mu\tau$ on a 90 % confidence level then is:

$$\mathcal{B}(B_s^0 \to \mu\tau)_{90\%} < 8.7 \cdot 10^{-6}.$$

The added signal and background distributions are shown in fig. 7.3.

7.4 Discussion of the Upper Limit

As mentioned in the previous section, 68'836 background events and 4 signal events remained in the mass region after having applied all the cuts, producing the best upper limit. However, it is somehow unusual in a determination of the upper limit to have that many (background) events that yield the best value (upper limit as well as $\frac{\varepsilon_{sig}}{\sqrt{\varepsilon_{bg}}}$). In most other cases, for instance in the search for $B_s^0 \to \mu\mu$, the best value is reached by cutting much harder against background. The problem that occurred in this analysis is the steep decline of the signal efficiency by introducing harder cuts, whereas the efficiency in the background does not decrease fast enough to account for this. The negative effect of loss in the signal efficiency therefore exceeds the positive effect of a smaller s_{up}. As an example of this effect, the upper limit was calculated using much harder cuts:

Table 7.4: Hard cuts for upper limit calculation.

Factorizing Cut Variable	Cut Value
$p_{\perp,3h}$	20 GeV
s_{3d}	25
IsolVar	0.98

This yielded $\varepsilon_{sig} = 1 \cdot 10^{-4}$ and $\varepsilon_{bg} = 1.4 \cdot 10^{-6}$ with 0 signal events and 217 background events in the mass region, resulting in an upper limit of $\mathcal{B}(B_s^0 \to \mu\tau)_{90\%} < 1.1 \cdot 10^{-4}$. This is considerably

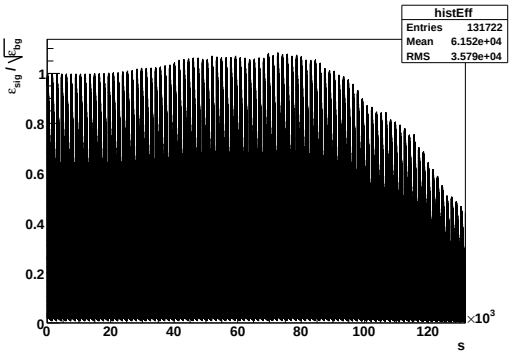

Figure 7.1: Visualization of $\frac{\varepsilon_{sig}}{\sqrt{\varepsilon_{bg}}}$ for varying cuts, where s is the loop-variable.

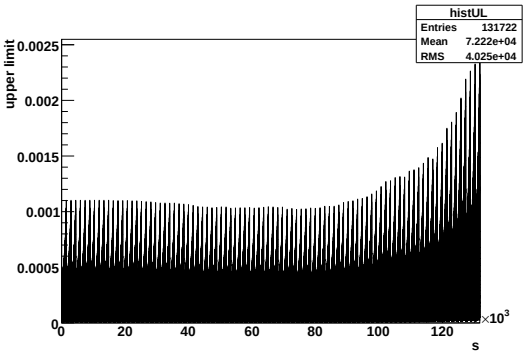

Figure 7.2: Visualization of the upper limit for varying cuts, where s is the loop-variable.

worse than the value presented in section 7.3. Hence the (consistent) method of calculating the upper limit mentioned in the sections before was applied.
The added signal and background distributions for these hard cuts is depicted in fig. 7.4.

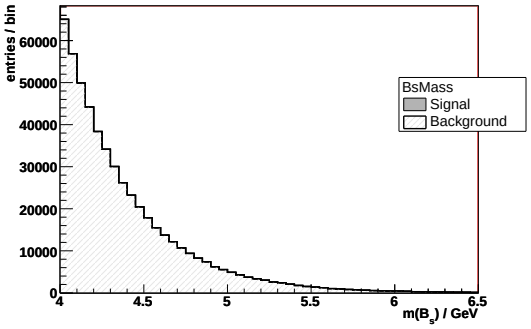

Figure 7.3: Added signal and background distribution for 10 fb^{-1}, generated with a Toy Monte Carlo, that produces best upper limit.

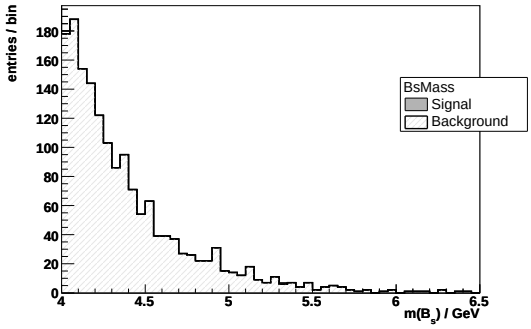

Figure 7.4: Added signal and background distribution for 10 fb^{-1}, generated with a Toy Monte Carlo, using hard cuts.

8 Summary & Outlook

In this thesis, a method to reconstruct the decay $B_s^0 \to \mu\tau$ with $\tau \to \pi\pi\pi\nu_\tau$ in the framework of the CMS detector has been presented, using the technique of neutrino reconstruction. An analysis then was performed in order to determine the upper limit of the branching fraction of $B_s^0 \to \mu\tau$, with a given integrated luminosity of $\mathcal{L} = 10\,\text{fb}^{-1}$. The achieved upper limit is:

$$\mathcal{B}(B_s^0 \to \mu\tau)_{90\%} < 8.7 \cdot 10^{-6}.$$

Several improvements could now be suggested:

- With real data, it will be possible to use sidebands as the background for the analysis. This allows much larger background samples and reduces the inaccuracy obtained with the Toy Monte Carlo.
- The trigger in this analysis on CMSSW-level was simulated manually and did not exactly correspond to the HLT incorporated in CMS. This, however, should not lead to a big difference.
- Studying the background with MC in more detail would possibly allow to use vetoes to exclude certain decay channels of certain particles, that contribute to the background of $B_s^0 \to \mu\tau$.
- The cut-based analysis could be replaced by a multivariate analysis, performed for example with maximum likelihood, or even more advanced, with a neural net.
- The number of variables used in the analysis (as pre-selection cuts and factorizing cuts) could be expanded to increase the discriminating power. However, many variables which were tried out show a strong correlation and therefore cannot be use concurrently.
- Instead of using the Bayesian Approach for the upper limit calculation, the CLs-method could be used (see f. ex. [20]).

A Additional Plots

A.1 Plots for Cut Variables in Mass Region $4.8\,\text{GeV} - 5.6\,\text{GeV}$

In the calculation of the upper limit, only the events lying in the mass region of $4.8\,\text{GeV} - 5.6\,\text{GeV}$ were used to compute the background and signal efficiency, and for the determination of s_{up}; the depicted histograms in section 6.4.1 and 6.4.2, however, were shown for the whole mass distribution. The histograms containing the distribution of the cut variables, incorporating the constraint $4.8\,\text{GeV} < m(B_s^0) < 5.6$ GeV, are plotted here for completeness.

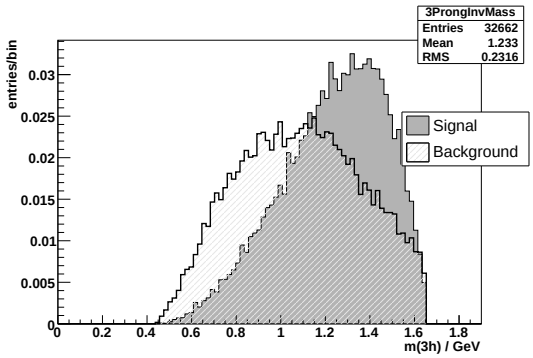

Figure A.1: $m(3h)$ for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

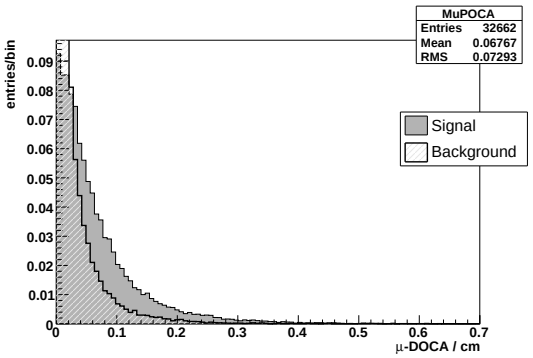

Figure A.2: μ-DOCA for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

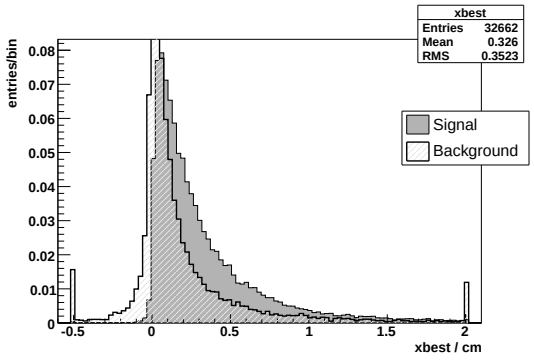

Figure A.3: xbest for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

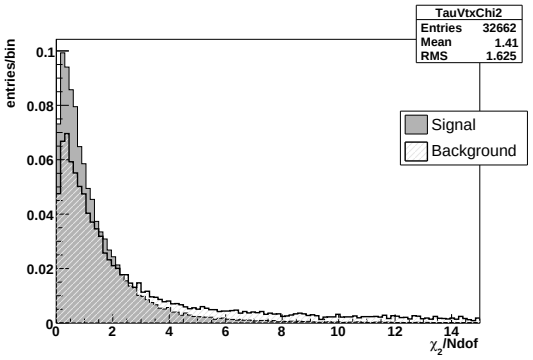

Figure A.4: χ^2 / Ndof for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

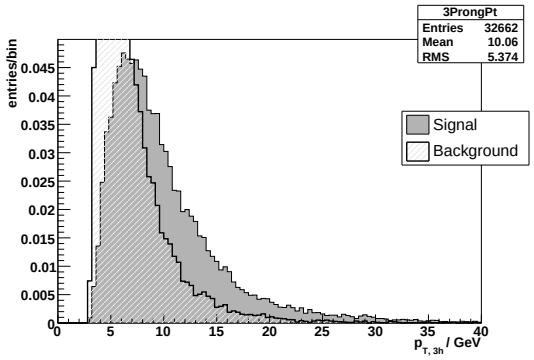

Figure A.5: $p_{\perp,3h}$ for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

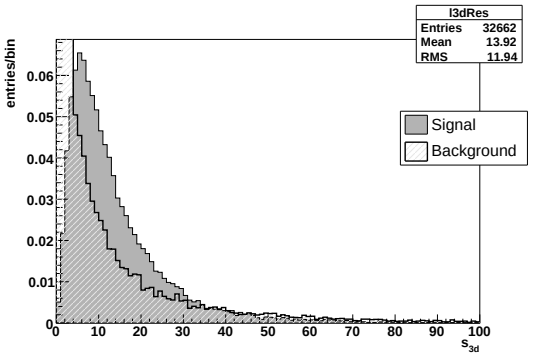

Figure A.6: s_{3d} for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

A.2 Generator Level Plots

The following figures depict plots of properties of the muons, the pions and the B_s^0, either from the signal sample or from the background sample, taken from the generator, with no cuts applied at all.
The plots corresponding to the signal all consist of pions and muons derived from a B_s^0, where the plots corresponding to the background comprise all the pions and muons in the event.
Note that a single distribution itself may not be representative, as many variables are correlated, e.g. $\Delta R(\mu, \pi)$ strongly correlates with $p_{\perp,\pi}$. The cuts used in the CMSSW candidate building (see section 4.4) can therefore not directly be deduced from these plots.

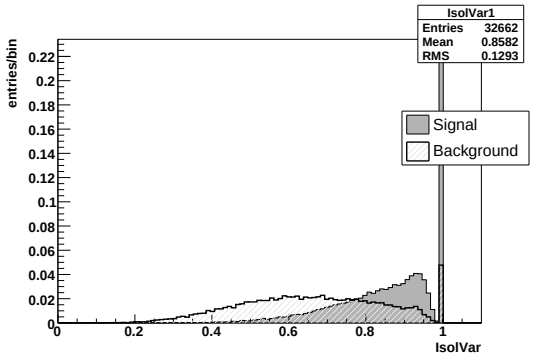

Figure A.7: IsolVar, $\Delta R(B_s^0, \mu/\pi) = 0.8$, for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

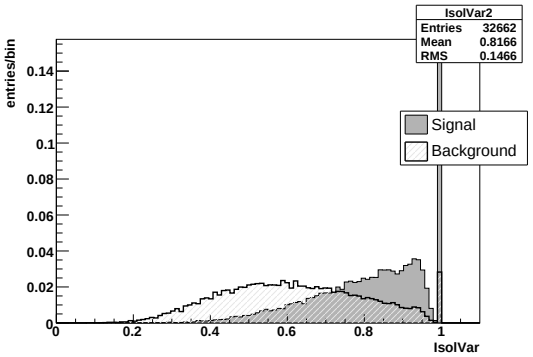

Figure A.8: IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1$ for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

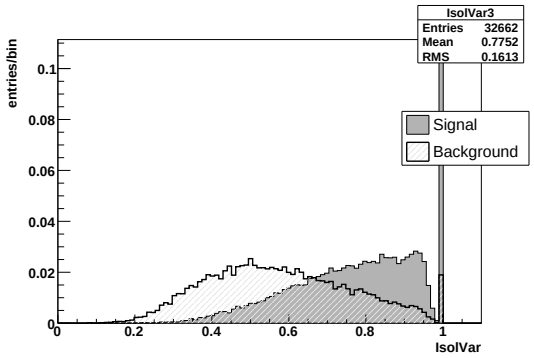

Figure A.9: IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.2$ for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

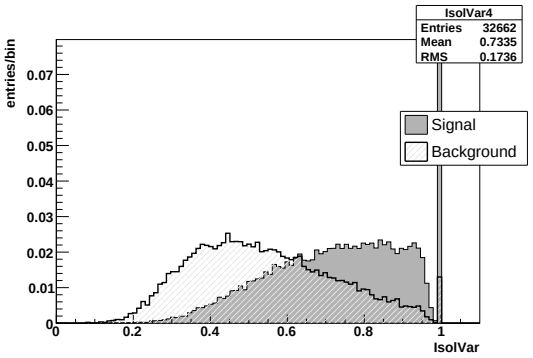

Figure A.10: IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.4$ for signal and background, with $4.8\,\text{GeV} < m(B_s^0) < 5.6\,\text{GeV}$.

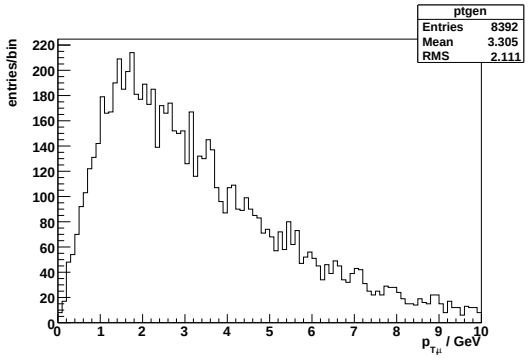

Figure A.11: Generator plot for $p_{\perp,\mu}$, signal.

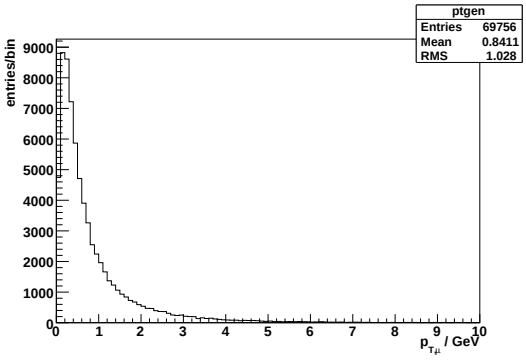

Figure A.12: Generator plot for $p_{\perp,\mu}$, background.

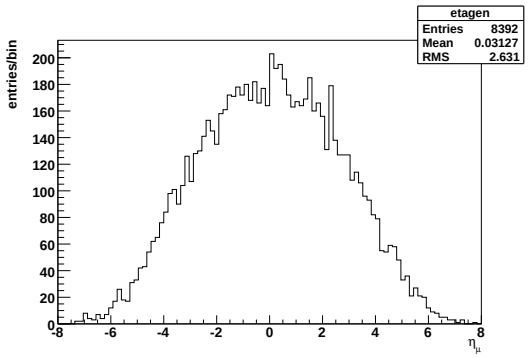

Figure A.13: Generator plot for η_μ, signal.

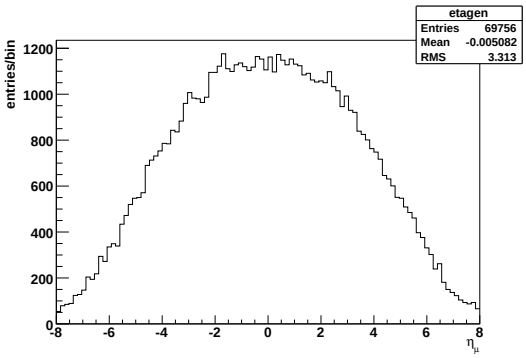

Figure A.14: Generator plot for η_μ, background.

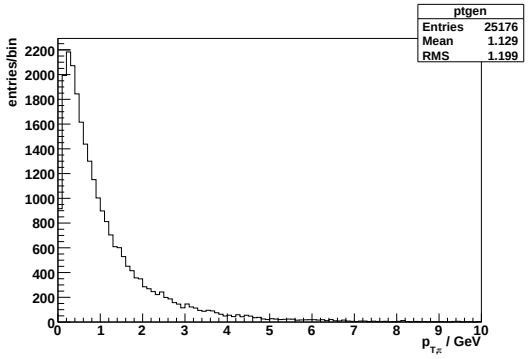

Figure A.15: Generator plot for $p_{\perp,\pi}$, signal.

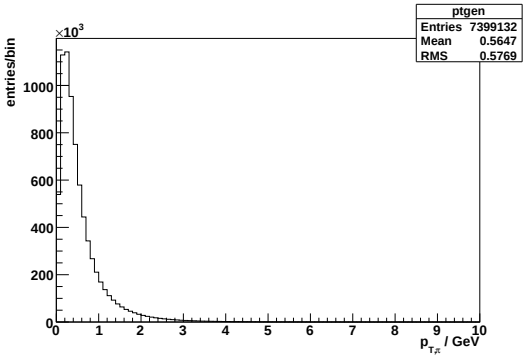

Figure A.16: Generator plot for $p_{\perp,\pi}$, background.

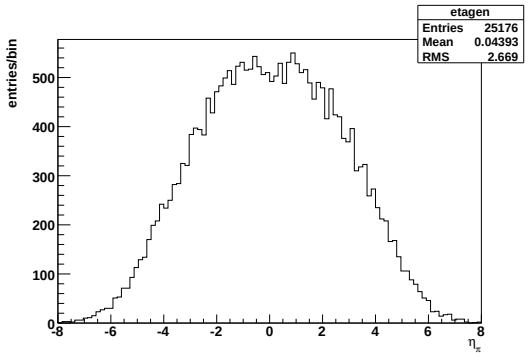

Figure A.17: Generator plot for η_π, background.

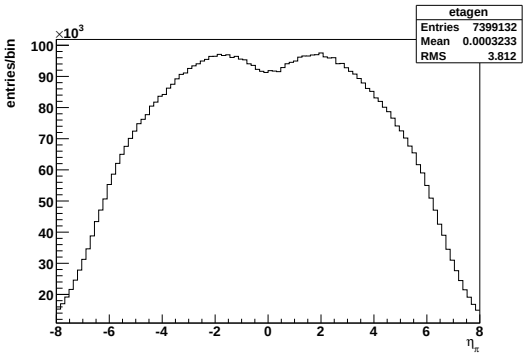

Figure A.18: Generator plot for η_π, background.

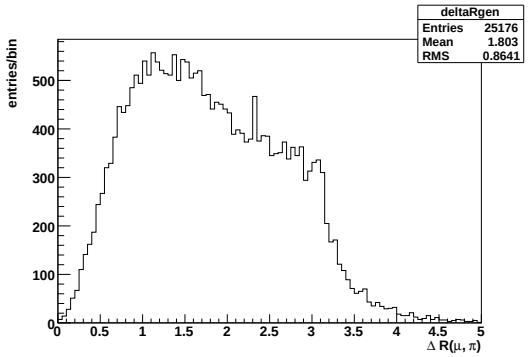

Figure A.19: Generator plot for $\Delta R(\mu, \pi)$, signal.

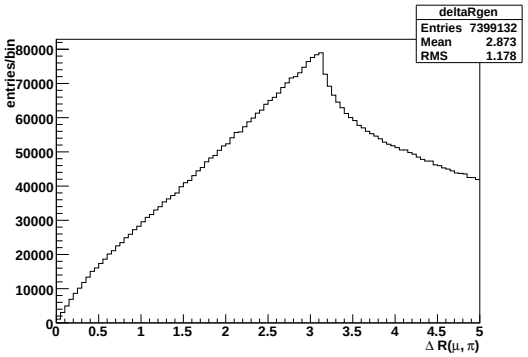

Figure A.20: Generator plot for $\Delta R(\mu, \pi)$, background.

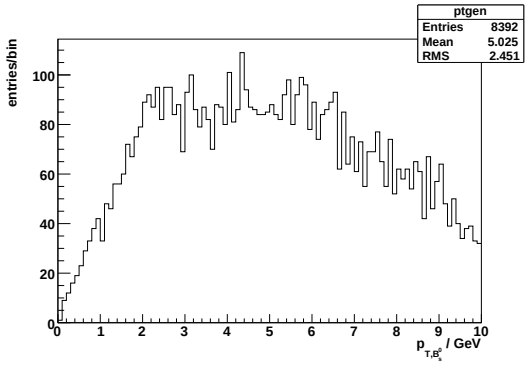

Figure A.21: Generator plot for $p_{\perp,B_s^0}$.

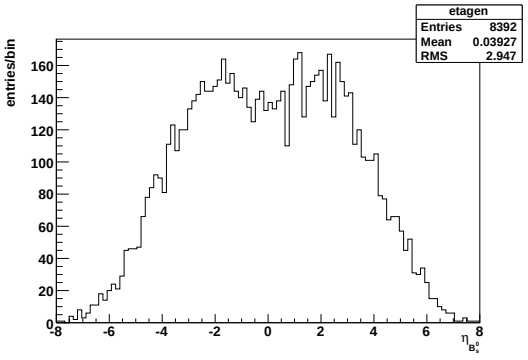

Figure A.22: Generator plot for $\eta_{B_s^0}$.

B Acknowledgments

Many thanks to Urs Langenegger for giving me the opportunity to perform my diploma thesis in his group and for his great commitment and enthusiasm. It was a pleasure to do a physics analysis in such an motivating environment.
Many thanks to Sarah Dambach and Christina Eggel for the friendly and relaxed atmosphere in the past five months and for the patience in answering all of my questions.
Thanks also to Marco Reinhard and Benjamin Stieger for sharing the office, providing help and for many discussions, regarding physics topics or anything else.
Finally I would like to thank my parents for their constant support during my time at the ETH. It is by no means a matter of course to have the chance to study such a «peculiar» science like physics.

This diploma thesis shall end with an excerpt of a poem of John Keats:

Then felt I like some watcher of the skies
When a new planet swims into his ken;
Or like stout Cortez, when with eagle eyes
He star'd at the Pacific – and all his men
Look'd at each other with a wild surmise –
Silent, upon a peak in Darien.

This document was typeset in Garamond, ETH-Font and Courier using LaTeX.

List of Figures

1.1 The accelerator facility at CERN [3]. 10
1.2 Scheme of the CMS detector [2]. 12

2.1 Feynman diagram for $B_s^0 \to \mu\tau$ in SM. 13
2.2 Feynman diagram for $B_s^0 \to \mu\tau$ in MSSM. 15
2.3 Feynman diagram for $B_s^0 \to \mu\tau$ in LHT. 15
2.4 Feynman diagram for $B_s^0 \to \mu\tau$ in LHT. 15
2.5 Feynman diagram with CKM-matrix elements in $B_s^0 \to \mu\tau$ (SM). 15
2.6 Feynman diagram for $\tau \to \pi\pi\pi\nu_\tau$. 16

3.1 Decay topology of $B_s^0 \to \mu\tau$, $\tau \to \pi\pi\pi\nu_\tau$ for an omnisicient observer. 17
3.2 Decay topology of $B_s^0 \to \mu\tau$, $\tau \to \pi\pi\pi\nu_\tau$, reconstructed by the detector. 17
3.3 Two-body decay of τ, where $\vec{p}_\tau = 0$. 19
3.4 Two-body decay of τ, where $\vec{p}_\tau \neq 0$, τ rest frame. 19
3.5 Two-body decay of τ, where $\vec{p}_\tau \neq 0$, lab rest frame. 19
3.6 Scheme of the reconstruction of the B_s^0, using neutrino reconstruction and kinematical arguments. 20

5.1 Resolution of τ-vertex., x-direction. 27
5.2 Resolution of τ-vertex., z-direction. 27
5.3 Distance between B_s^0-vertex., reconstructed on a line and on a helix. 29
5.4 B_s^0-mass distribution, where μ track is a line. 29
5.5 B_s^0-mass distribution, where μ track is a helix. 29
5.6 B_s^0-mass distribution, using neutrino reconstruction. 30
5.7 B_s^0-mass distribution, using the k-Factor method. 30
5.8 Angle between $\vec{p}_{3h}$ and $\vec{p}_\nu$. 31
5.9 Multiplicity of pions in three-prong (background). 34
5.10 Multiplicity of mothers of particles of three-prong (background). 34

6.1 Illustration for $\Delta R(\tau\text{-vtx}, p_{\mu,3h})$. 36
6.2 Distribution of $m(3h)$ for signal and background, used for candidate selection cuts. 36
6.3 Distribution of $Q(3h)$ for signal and background, used for candidate selection cuts. 36
6.4 Distribution of $\Delta R(\mu, 3h)$ for signal and background, used for candidate selection cuts and the potential selection of a single candidate. 37
6.5 Distribution of $\Delta R(\tau\text{-vtx}, p_{\mu,3h})$ for signal and background, used for candidate selection cuts and the potential selection of a single candidate. 37

6.6 Distribution of s_{3d} for signal and background, used for the selection of a single candidate. 37
6.7 Distribution of s_{xy} for signal and background, used for the potential selection of a single candidate. 37
6.8 Distribution of $\angle R(\tau\text{-vtx}, p_{\mu,3h})$ for signal and background, used for the potential the selection of a single candidate. 38
6.9 Distribution of $\cos \alpha_{xy}$ for signal and background, used for the potential selection of a single candidate. 38
6.10 Distribution of number of candidates per event, before candidate selection cuts. . 38
6.11 Distribution of number of candidates per event, after candidate selection cuts. . . 38
6.12 $m(B_s^0)$, minimization done with Monte Carlo minimization. Note that the RMS is larger compared with fig. 5.6, done with the full Migrad minimization. 40
6.13 Illustration for POCA and xbest. 43
6.14 Distribution of $m(3h)$ for signal and background, used as a pre-selection cut. The steep decrease at 1.65 GeV is a result of the applied candidate selection cut. 44
6.15 Distribution of μ-DOCA for signal and background, used as a pre-selection cut . 44
6.16 Distribution of xbest for signal and background, used as a pre-selection cut. The two «towers» in the background arise, if the minimum is situated at x < -0.5 cm or x > 2 cm, respectively. 44
6.17 Distribution of χ^2 / Ndof for signal and background, used as a pre-selection Cut 44
6.18 Illustration for the isolation variable. 45
6.19 Distribution of $p_{\perp,3h}$ for signal and background, used as a factorizing cut. 45
6.20 Distribution of s_{3d} for signal and background, used as a factorizing cut. 45
6.21 Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 0.8$ for signal and background, used as a factorizing cut. 45
6.22 Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1$ for signal and background, used as a factorizing cut. 45
6.23 Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.2$ for signal and background, used as a factorizing cut. 46
6.24 Distribution of IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.4$ for signal and background, used as a factorizing cut. 46

7.1 Visualization of $\frac{\varepsilon_{sig}}{\sqrt{\varepsilon_{bg}}}$ for varying cuts, where s is the loop-variable. 52
7.2 Visualization of the upper limit for varying cuts, where s is the loop-variable. . . 52
7.3 Added signal and background distribution for 10 fb^{-1}, generated with a Toy Monte Carlo, that produces best upper limit. 52
7.4 Added signal and background distribution for 10 fb^{-1}, generated with a Toy Monte Carlo, using hard cuts. 52

A.1 $m(3h)$ for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 55
A.2 μ-DOCA for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 55
A.3 xbest for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 56
A.4 χ^2 / Ndof for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 56
A.5 $p_{\perp,3h}$ for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 56

A.6 s_{3d} for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 56
A.7 IsolVar, $\Delta R(B_s^0, \mu/\pi) = 0.8$, for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 57
A.8 IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1$ for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 57
A.9 IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.2$ for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 57
A.10 IsolVar, $\Delta R(B_s^0, \mu/\pi) = 1.4$ for signal and background, with 4.8 GeV < $m(B_s^0)$ < 5.6 GeV. 57
A.11 Generator plot for $p_{\perp,\mu}$, signal. 58
A.12 Generator plot for $p_{\perp,\mu}$, background. 58
A.13 Generator plot for η_μ, signal. 58
A.14 Generator plot for η_μ, background. 58
A.15 Generator plot for $p_{\perp,\pi}$, signal. 58
A.16 Generator plot for $p_{\perp,\pi}$, background. 58
A.17 Generator plot for η_π, background. 59
A.18 Generator plot for η_π, background. 59
A.19 Generator plot for $\Delta R(\mu, \pi)$, signal. 59
A.20 Generator plot for $\Delta R(\mu, \pi)$, background. 59
A.21 Generator plot for $p_{\perp,B_s^0}$. 59
A.22 Generator plot for $\eta_{B_s^0}$. 59

List of Tables

4.1 Efficiency of signal sample. 24
4.2 Efficiency of background sample. 25

5.1 Parameters of double gaussian for τ-vtx resolution. The errors given in this table correspond to the errors of the fitting. 28
5.2 Parameters of double gauss for line and for helix. 30
5.3 Particle ID of μ and $3h$-type (background). 31
5.4 Particle ID of μ-mother (background). 32
5.5 Particle ID of h-mother, if all h have the same mother (background). 32
5.6 Particle ID of h-mother (background). 33

6.1 Efficiency for different selection variables. 39
6.2 Efficiency for signal, B_s^0-reconstruction. 41
6.3 Efficiency for background, B_s^0-reconstruction. 42
6.4 Pre-Selection cuts (hard). 42
6.5 Pre-Selection cuts (loose). 43

7.1 Efficiencies of different stages. 50
7.2 Search regions for grid-search. 51
7.3 Best cuts for upper limit calculation. 51
7.4 Hard cuts for upper limit calculation. 51

Bibliography

[1] K. Knorr Cetina, "Epistemic Cultures," *Harvard University Press* (2003) .

[2] D. Acosta *et al.*, "CMS Physics TDR: Volume I (PTDR1), Detector Performace and Software,".

[3] C. C. Group, "LHC The Guide," *CERN-Brochure-2008-001-Eng* (January 2008) .

[4] **Particle Data Group** Collaboration, W. M. Yao *et al.*, "Review of particle physics," *J. Phys.* **G33** (2006) 1–1232.

[5] **BABAR** Collaboration, B. Aubert *et al.*, "Searches for the decays B0 -> l+ tau- and B+ -> l+ nu (l=e,mu) using hadronic tag reconstruction," *Phys. Rev.* **D77** (2008) 091104, arXiv:0801.0697 [hep-ex].

[6] A. Dedes, J. R. Ellis, and M. Raidal, "Higgs mediated B/(s,d)0 –> mu tau, e tau and tau –> 3mu, e mu mu decays in supersymmetric seesaw models," *Phys. Lett.* **B549** (2002) 159–169, arXiv:hep-ph/0209207.

[7] J. Hisano, T. Moroi, K. Tobe, and M. Yamaguchi, "Lepton-Flavor Violation via Right-Handed Neutrino Yukawa Couplings in Supersymmetric Standard Model," *Phys. Rev.* **D53** (1996) 2442–2459, arXiv:hep-ph/9510309.

[8] N. Gaur, "Charged Lepton Flavour Violation in Littlest Higgs model with T-parity," *AIP Conf. Proc.* **981** (2008) 357–359, arXiv:0710.3998 [hep-ph].

[9] M. Artuso *et al.*, "B, D and K decays," arXiv:0801.1833 [hep-ph].

[10] G. Isidori, "Flavour Physics: Now and in the LHC era," arXiv:0801.3039 [hep-ph].

[11] S. Dambach, U. Langenegger, and A. Starodumov, "Neutrino reconstruction with topological information," *Nucl. Instrum. Meth.* **A569** (2006) 824–828, arXiv:hep-ph/0607294.

[12] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[13] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[14] Z. Was, P. Golonka, and G. Nanava, "PHOTOS Monte Carlo and its theoretical accuracy," arXiv:0807.2762 [hep-ph].

[15] J. Allison *et al.*, "Geant4 developments and applications," *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[16] A. Strandlie and W. Wittek, "Propagation of covariance matrices of track parameters in homogeneous magnetic fields in CMS,". CERN-CMS-NOTE-2006-001.

[17] J. Heinrich *et al.*, "Interval estimation in the presence of nuisance parameters. 1. Bayesian approach," `arXiv:physics/0409129`.

[18] J. Heinrich, "User Guide to Bayesian-Limit Software Package,".

[19] W. A. Rolke, A. M. Lopez, and J. Conrad, "Confidence Intervals with Frequentist Treatment of Statistical and Systematic Uncertainties," *Nucl. Instrum. Meth.* **A551** (2005) 493–503, `arXiv:physics/0403059`.

[20] A. L. Read, "Modified frequentist analysis of search results (The CL(s) method),". Prepared for Workshop on Confidence Limits, Geneva, Switzerland, 17-18 Jan 2000.