Department of Physics and Astronomy

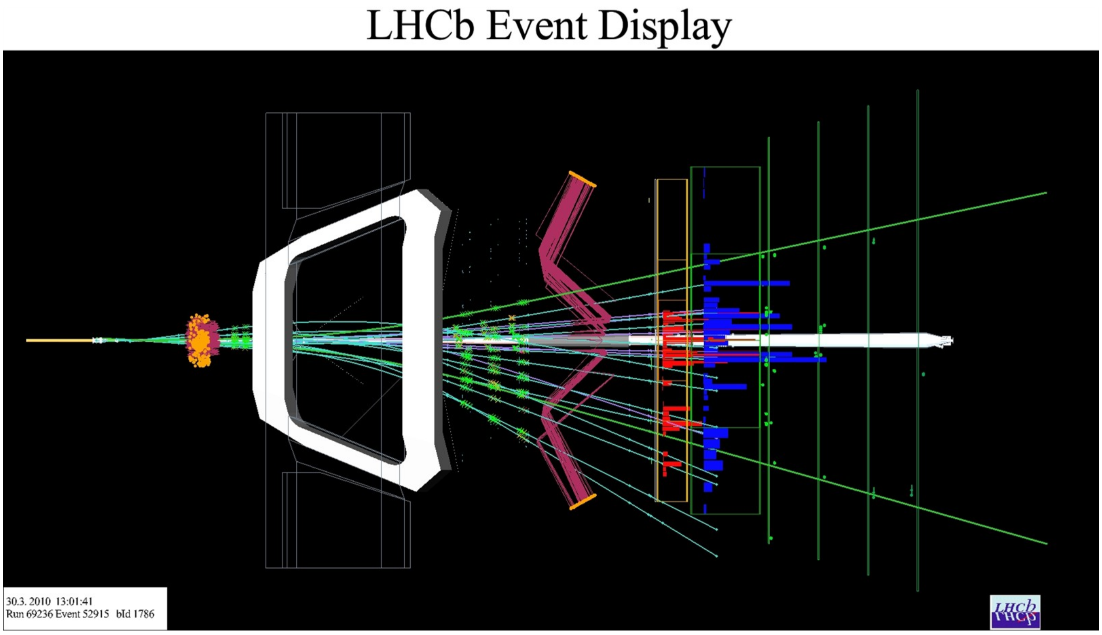
Heidelberg University

Diploma thesis
in Physics

submitted by
Sascha Stahl
born in Bad Kreuznach, Germany

June 2010

Reconstruction of displaced tracks and measurement of K_s production rate in proton-proton collisions at $\sqrt{s} = 900\,\mathrm{GeV}$ at the LHCb experiment



This diploma thesis has been carried out by Sascha Stahl
at the
Physikalisches Institut
under the supervision of
Prof. Dr. Stephanie Hansmann-Menzemer

Abstract

Since end of 2009 protons are colliding in the LHC at a center-of-mass energy that has never been reached before. The LHCb experiment, one of the four big experiments at the LHC, is dedicated to precision measurements in the B system. The first and most vital part in the reconstruction of an event is the track finding. In this work an algorithm which reconstructs the decay products of long lived particles which decay behind the first tracking station is described. Changes to the algorithm which improve its efficiency by 6 % without increasing the fraction of wrongly reconstructed tracks are discussed. Furthermore the algorithm which was developed on simulated data is tested on the first collisions at LHCb. In the second part of the work the first analysis of the data taken end of 2009 is presented. The K_s production cross-section in proton-proton collisions at a center-of-mass energy of 900 GeV is measured. The analysis is based on the before mentioned track finding algorithm.

Kurzfassung

Seit Ende 2009 kollidieren Protonen im LHC bei einer Schwerpunktsenergie, die noch nie vorher erreicht wurde. Das LHCb Experiment, eines der vier großen Experimente am LHC, widmet sich Präzessionsmessungen im B System. Der erste und wichtigste Teil der Ereignisrekonstruktion ist die Spurfindung. In dieser Arbeit wird ein Algorithmus beschrieben, der die Zerfallsprodukte langlebiger Teilchen, die nach dem ersten Spurfindungssystem zerfallen, rekonstruiert. Änderungen am Algorithmus, die die Effizienz um 6 % steigern ohne dabei die Anzahl fehlerhaft rekonstruierter Spuren zu erhöhen, werden diskutiert. Weiterhin wird der Algorithmus, der auf simulierten Daten entwickelt wurde, auf den ersten Kollisionen im LHCb Experiment getestet. Im zweiten Teil der Arbeit wird die erste Analyse der Daten, die Ende 2009 aufgenommen wurden, präsentiert. Dabei wird der Wirkungsquerschnitt für K_s Produktion in Proton-Proton Kollisionen bei einer Schwerpunktsenergie von 900 GeV bestimmt. Die Analyse basiert auf dem zuvor genannten Spurfindungsalgorithmus.

Contents

Introduction

In November 2009 protons were brought for the first time to collision in the largest particle accelerator of the world, the Large Hadron Collider located beneath the area around Geneva. In the first period of running (expected to last until end of 2011) proton-proton collisions will be observed at a center-of-mass energy of 7 TeV. After that the design energy of 14 TeV will be approached.

The world of particles and their interactions is described by the so called Standard Model, a theory which was introduced more than 30 years ago. Up to today no collider experiment has measured deviations from the Standard Model of particle physics. However, cosmological observations, like matter-antimatter asymmetry or Dark Matter, suggest that there must be a theory "beyond" the Standard Model, so called "New Physics". The LHC was built to test the Standard Model beyond the Electroweak Scale (100 GeV) at the TeV scale where "New Physics" is expected to manifest itself.

Extensions of the Standard Model often lead to new particles. There are two different approaches to detect new particles at the LHC. The multipurpose experiments ATLAS and CMS perform direct searches by searching for up to now undiscovered resonances. The Large Hadron Collider beauty experiment (LHCb) tries to find new particles indirectly. Additional particles appear virtually in new Feynman diagrams introducing quantum corrections which interfere with Standard Model processes. Branching ratios of rare decays and CP violating phases are observables which are especially sensitive to these corrections. Thus they are an excellent tool to probe the Standard Model. Due to the virtuality of the particles in the loop processes, very high energies can be probed.

The LHCb experiment was designed to perform these indirect searches in the B system. The b quark is the heaviest quark which can hadronize. About 10^{12} $b\bar{b}$ pairs are produced in a nominal year of data taking at a center-of-mass energy of 14 TeV at LHCb. One of the key measurements of the LHCb physics programme is the measurement of the CP violating phase in the decay $B_s \to J/\psi(\mu\mu)\ \phi(KK)$. The Standard Model makes theoretically very precise predictions of this phase. Thus a measurement of a small deviation would be a clear sign of New Physics.

However, before making any conclusive measurement of New Physics, the detector has to be understood and different well known reference channels have to be measured. One such channel is the decay $B^0 \to J/\psi\ K_s$ which is a good candidate to measure CP violation in the B^0 system. The particularity of this channel is the reconstruction of the K_s. The K_s has a lifetime which is rather long on the scale of the LHC. Thus often the K_s daughters cannot be reconstructed with the default tracking algorithm because they do not leave measurements in the first tracking detector. Then a special algorithm which reconstructs displaced particles has to be used.

For the understanding of the background to several of the LHCb measurements a simulation which describes properly the physics processes in proton-proton collisions is needed. For the description of proton-proton collisions a proper theory of the underlying QCD processes is required. The computations of these processes have large theoretical uncertainties, thus they need experimental input to improve the predictions. One example of an useful input is the cross-section for K_s production in proton-proton collisions. As LHCb is a forward spectrometer which covers an unique angular acceptance, no measurement of the production cross-section in this range exists so far.

Within this work, first the LHCb experiment and its physics programme is introduced. In the second chapter a short overview of the track reconstruction at the LHCb detector is given. Then an algorithm dedicated to the reconstruction of the decay products of long lived particles is described in Chapter 3. In Chapter 4 improvements of the algorithm are discussed. This is followed by a chapter in which the algorithm is tested for the first time on data. In the last chapter a measurement of the yield of K_s mesons in the data which were taken end of 2009 is presented. This results in a cross-section measurement of K_s production in proton-proton collisions at $\sqrt{s} = 900\,\mathrm{GeV}$.

Chapter 1

The LHCb experiment

The LHCb experiment (Large Hadron Collider beauty) is one of six experiments at the Large Hadron Collider (LHC). The detector is especially designed to reconstruct the decay products of B hadrons created in proton-proton collisions with a high precision. The design of the detector and the physics programme of the LHCb experiment is briefly explained in the following chapter.

1.1 The LHC

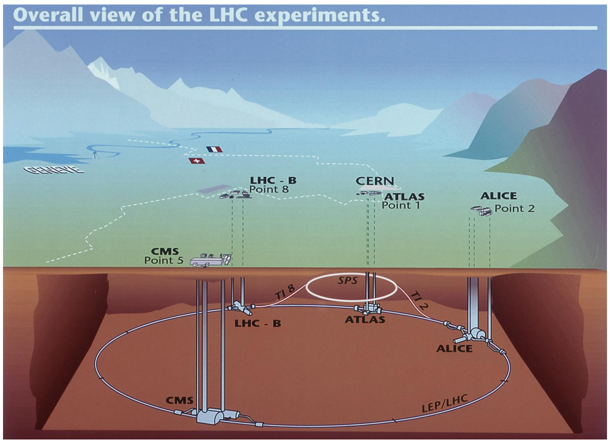


Figure 1.1: *Schematic view of the LHC. The four large experiments ATLAS, ALICE, CMS and LHCb are shown [1].*

With a circumference of 27 km the LHC is the world's largest particle accelerator. It is located about 100 m below the ground near Geneva, see Figure 1.1. The two proton beams collided for the first time in November 2009 at a center-of-mass energy of 900 GeV. In late December collisions at an energy of 2.36 TeV were recorded, outnumbering the Tevatron[1] as the collider with the highest center-of-

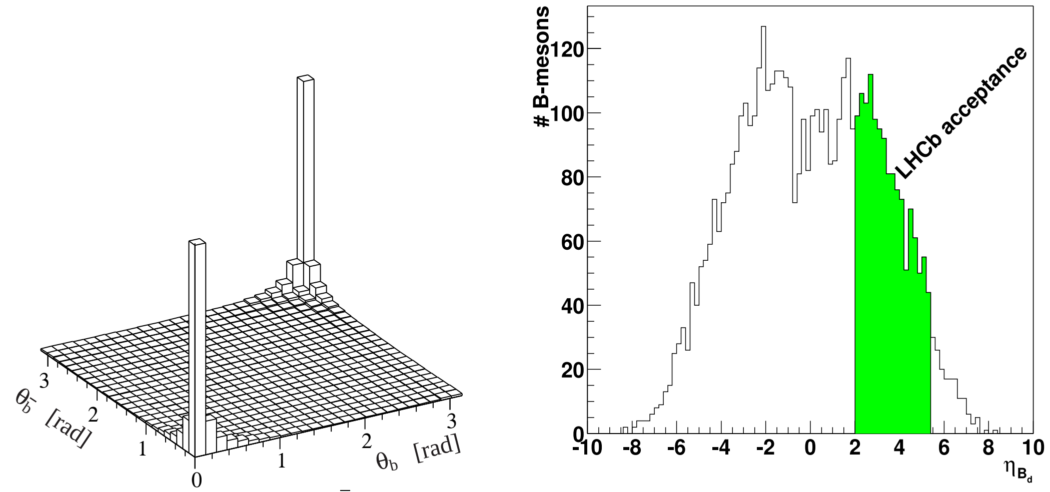
---

[1] $p\bar{p}$ collider at $\sqrt{s} = 1.98\,\text{TeV}$ at Fermilab/USA.

mass energy. After a short shutdown, in March 2010 protons were accelerated up to an energy of 3.5 TeV and brought to collision. These collisions at a center-of-mass energy of 7 TeV started the first period of B and charm physics. It is planned to run 18 - 24 months at a center-of-mass energy of 7 TeV. After this period it is foreseen to upgrade the accelerator to the original design energy of 14 TeV.

1.2 The LHCb detector

The LHCb detector was built to make precise measurements in the B hadron system. The dominant process to create b quarks at the LHC is gluon-gluon fusion. The calculation of the cross-section for $b\bar{b}$ production at the LHC is a theoretical challenge due to non-perturbative QCD effects. In [2] it is given as $\sigma_{b\bar{b}} = 698\,\text{mb}$ with an uncertainty of about 200 mb at a center-of-mass energy of 14 TeV. On the energy scale the LHC is operating B mesons are rather light particles, e.g. the B_s meson has a mass of 5.37 GeV. Thus it is possible to produce $b\bar{b}$ pairs in gluon-gluon fusion over a wide momentum range. Due to the gluon density distribution in the proton, it is likely that the gluons have different momenta and thus the $b\bar{b}$ pair is boosted along the beam axis, cf. Figure 1.2(a) . Therefore the detector was designed as a forward spectrometer.



(a) correlation between b and $\bar{b}$ polar angles

(b) η distribution of B mesons

Figure 1.2: *(a) Polar angle θ distribution of $b\bar{b}$ pairs produced at the LHC. (b) Pseudorapidity ($\eta = -log(tan\frac{\theta}{2})$) distribution of the B mesons [3].*

In Figure 1.3 the detector layout in the $y-z$ plane is shown. The dipole magnet in the middle is deflecting charged particles in the $x-z$ plane. Particles traversing the total magnetic field see an integrated field strength of about 4 Tm. The angular coverage of the detector is approximately from 10 to 300 mrad in

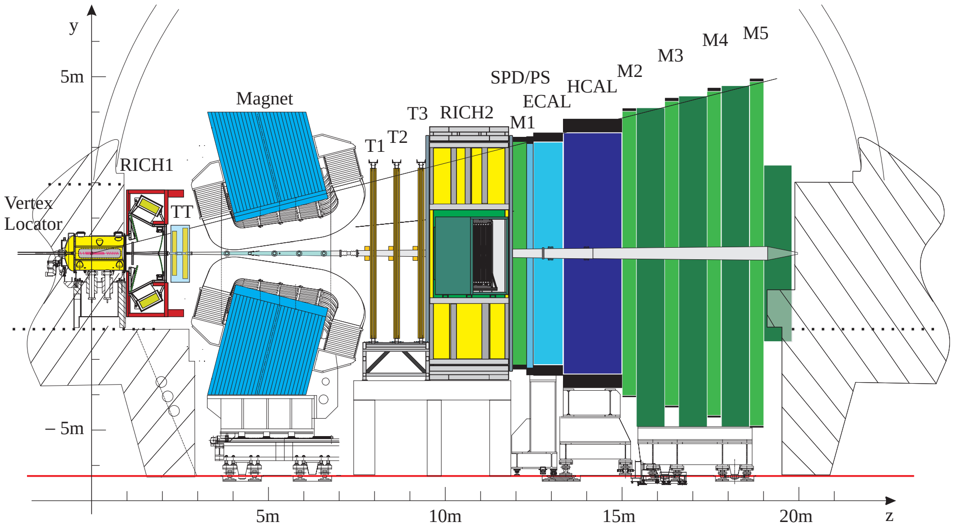


Figure 1.3: *Schematic view of the LHCb detector [4].*

the bending plane of the magnet and 10 to 250 mrad in the non-bending plane. About 25 % of the produced $b\bar{b}$ pairs are in the acceptance of the detector while it is only covering 8 % of the polar angle, cf. Figure 1.2(b).

To do precise measurements with B hadrons clean events are needed because the whole decay chain of the B hadron has to be reconstructed particle by particle. To achieve this, the beam is only weakly focused at the interaction point. This results in only one proton-proton collision per bunch crossing in average[2]. In comparison the multipurpose detectors, ATLAS and CMS, have in the order of 20 interactions per bunch crossing [5]. The advantage of the smaller number of interactions is that the occupancy of the different subdetectors is low and the events are simpler to analyse. Furthermore the radiation damage is reduced. The design luminosity at a center-of-mass energy of 14 TeV of the LHCb experiment is $L = 2\,\mathrm{fb}^{-1}$ per year with about 100.000 created $b\bar{b}$ pairs per second.

For the reconstruction of B events one needs two things: a good particle identification and a precise track reconstruction.
The track reconstruction, namely the precise spatial resolution of vertices and the momentum measurement are done by the tracking system consisting of

- the **Vertex Locator** around the interaction point,
- the **Trigger Tracker** in front of the magnet,
- the **T-Stations** behind the magnet

and the **dipole magnet** in the middle.
The particle identification is done with the following detectors:

---

[2] It is assumed that the number of interactions is Poisson distributed.

- **Ring Imaging Cherenkov Detectors:** RICH1 is located in front of the magnet to identify low momentum particles. RICH2 is behind the magnet to measure high momentum particles. Both together allow a separation of pions, kaons and protons up to a momentum of 100 GeV.
- **Calorimeters:** The Electromagnetic (ECAL) and Hadronic Calorimeters (HCAL) measure energies of electrons, photons and hadrons.
- **Muon stations:** Located at the end of the detector the muon stations (M1 - M5) identify muons.

This work concentrates on tracking in the LHCb detector, thus the tracking system is considered in more detail. For further details on the particle identification and the detector in general the reader is referred to [4].

1.3 Tracking system

The task of the tracking system is to reconstruct the trajectories of charged particles to measure their momentum and to do precise vertex reconstruction. The main component of the B-field of the LHCb dipole is in y direction. Thus charged particles are deflected in the $x - z$ plane. The integrated field for particles traversing the tracking detectors is about 4 Tm. In Figure 1.4 the field map is shown. While the Trigger Tracker and the T-Stations are inside the fringe field of the magnet, the B-field in the Vertex Locator is negligible. The measurement of the curvature of the trajectory by the tracking system determines then the particle's momentum.

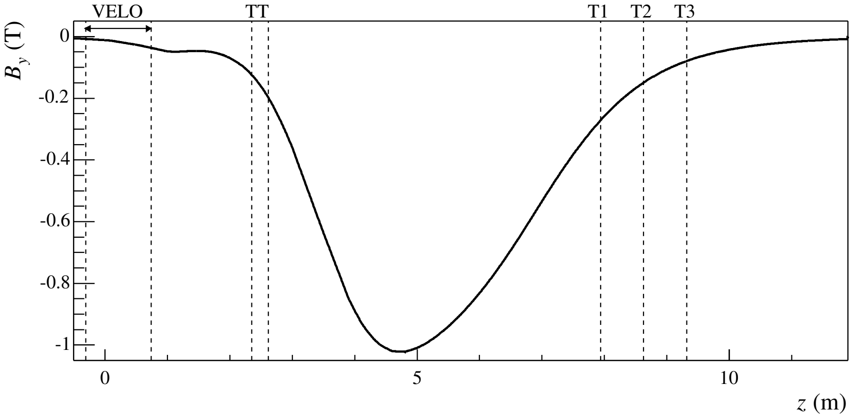


Figure 1.4: *The magnetic field component in y direction. The positions of the tracking detectors are also shown [4].*

1.3.1 Vertex Locator

The Vertex Locator (Velo) is located directly at the interaction point. Only a thin aluminium foil is separating the vacuum inside the Velo from the ultra high vacuum in the LHC beam pipe. As shown in Figure 1.5, the Velo consists out of 21 stations. Each station is composed out of two half-disc-shaped modules. These modules enclose the beam pipe. Each module has two types of silicon strip sensors. One type measures the radial distance to the beam line, the other type the azimuthal angle. The distance of the Velo modules to the beam is so small that they could be damaged during the injection of new protons into the LHC when the beam is defocused. Thus the two halves of the Velo can be moved away from the beam pipe during the injection phase and they are moved in once the beam is stable. The design of the Velo allows a reconstruction of the point of the proton-proton collision, the primary vertex, with a very high precision of 60 μm in the z direction and 10 μm in the $x-y$ plane.

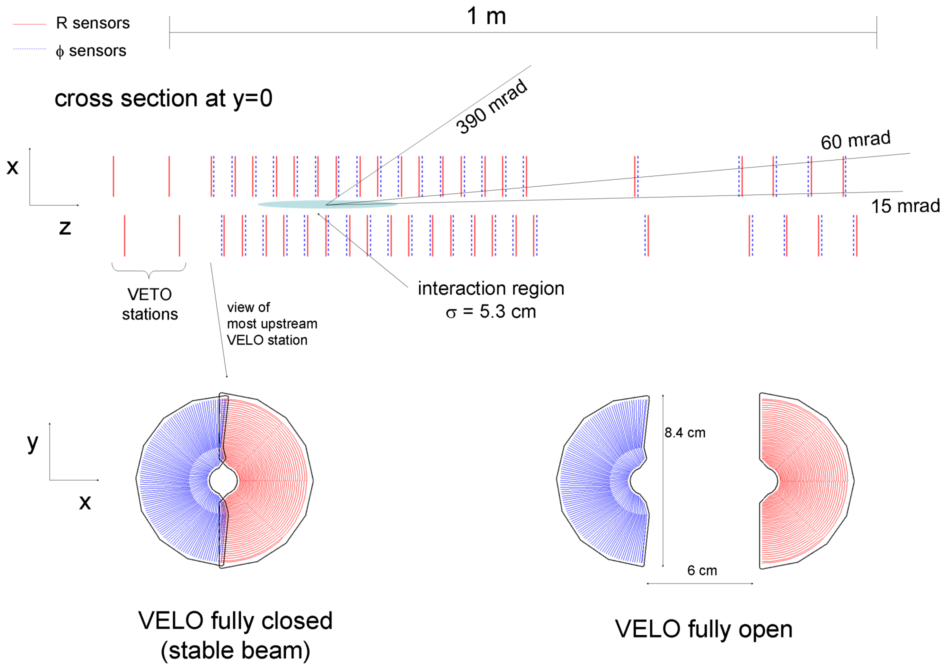


Figure 1.5: *Schematic view of the Vertex Locator [4].*

1.3.2 Trigger Tracker

The Trigger Tracker (TT) is located upstream of the LHCb dipole about 2 m behind the Velo. It is 150 cm wide and 130 cm high and covers the full acceptance of the detector. The TT is a silicon strip detector with a strip pitch of 183 μm.

This results in a single hit resolution of about 50 μm [4]. The TT consists of two stations which are 27 cm apart (TTa and TTb, cf. Figure 1.6). These modules are built out of two layers. The second layer of the first station is rotated by $-5°$ around the z axis with respect to the x axis and the first layer of the second station is rotated in the opposite direction by $+5°$. The rotated layers are called u and v layers. They are also referred as stereo layers. The layers with strips perpendicular to the x axis are called x layers. The $xuvx$-configuration allows a reconstruction of tracks in three dimensions with a good resolution in the main bending plane.

The TT is useful to reconstruct low momentum particles which are bent out of the detector by the magnet and long lived particles which decay behind the Velo. It also serves as an additional measurement in front of the magnet to improve the momentum estimate.

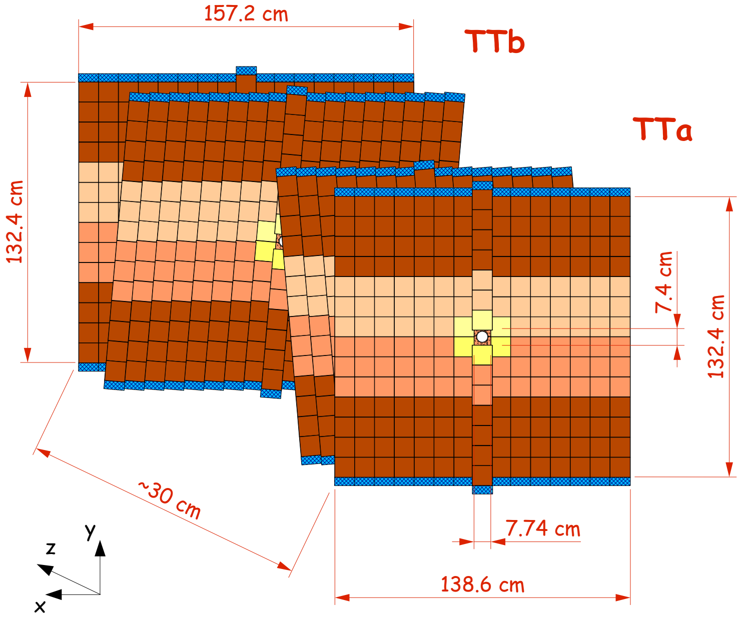


Figure 1.6: *Schematic view of the Trigger Tracker [4].*

1.3.3 T-Stations

The main tracking system (T-Stations) is located behind the LHCb dipole. It consists out of three stations which have a size of about 6 m × 5 m. Two different technologies are used to build up each station.

Around the beam pipe is the so called Inner Tracker (IT). It covers the region with the highest flux of charged particles. Like the TT, the Inner Tracker is a silicon strip detector and uses the same technology. Each station has also four layers with a *xuvx*-configuration. The layout of the x and stereo layers is shown in Figure 1.7.

The IT is surrounded by the Outer Tracker (OT). The OT is a straw-tube detector. Every x and stereo layer consists out of two layers of drift-tubes with inner diameters of 4.9 mm, cf. Figure 1.8. The target resolution in x direction is 200 μm. The OT covers an acceptance up to 300 mrad in x and 250 mrad in y direction. The total active area of one station is $5971 \times 4850\,\mathrm{mm}^2$, cf. Figure 1.8.

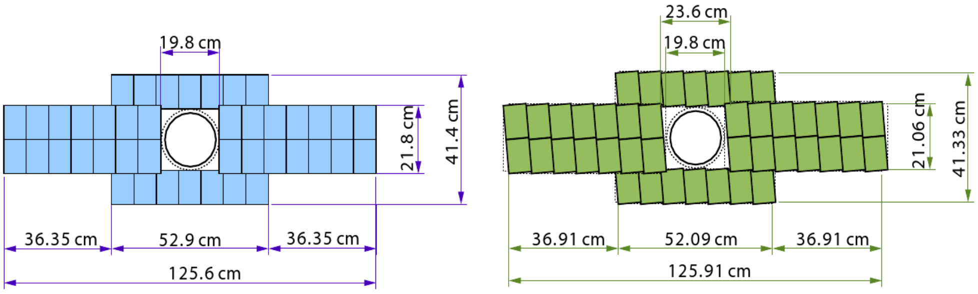


Figure 1.7: *Schematic view of the Inner Tracker. The left figure shows a x layer and the right figure a stereo layer.*

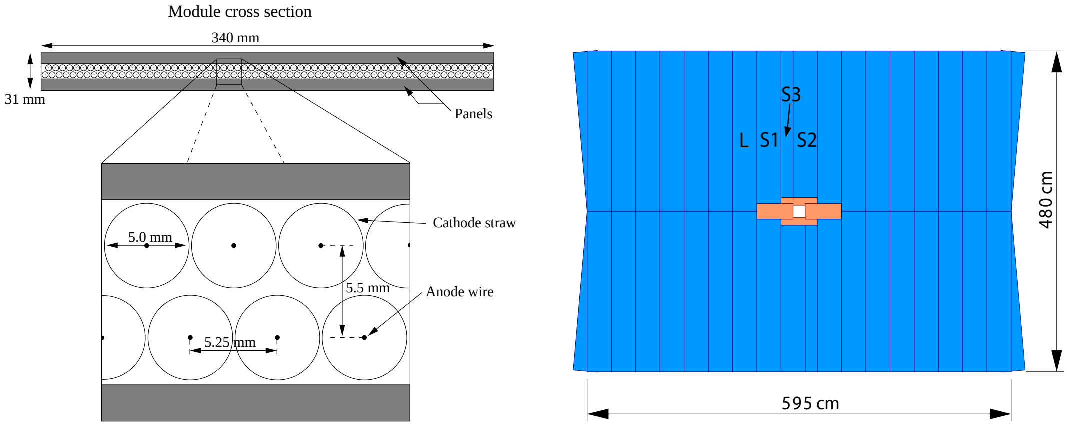


Figure 1.8: *Left: Top view of an OT layer showing the straw-tubes. Right: Front view of the Outer Tracker [4].*

1.4 Trigger system

In the LHCb detector proton bunches will collide every 25 ns which corresponds to a rate of 40 MHz. The task of the trigger is to reduce this vast number of events to a rate which can be stored. The crucial point is the selection of the events. Interesting events, e.g. events with a B meson, should be kept and events which are not in the focus of physics analyses should be discarded.

The LHCb trigger system consists of two levels and reduces the data rate to about 2 kHz which are then stored for physics analyses. The first level of the trigger is a hardware trigger called L0. It exploits that B and D decays produce particles with a high transverse momentum. Therefore the trigger looks for high transverse energy deposits in the calorimeters and high transverse momentum tracks in the muon stations. Additionally the L0 can veto events with a high multiplicity or multiple primary interactions. This is done with the two modules in the Velo which are located in front of the nominal interaction point. They are shown in Figure 1.5 and called pile up stations.

The second level is a software trigger, the High-Level Trigger (HLT). After further reduction of events the HLT can perform a full reconstruction of all tracks in an event. With this information selections for specific decay channels are performed and the rate is reduced to 2 kHz.

In the 2009 run the collision rate was so low that the HLT was not needed. Events were only triggered by the L0. Due to the lower center-of-mass energy compared to nominal running conditions, the L0 thresholds were lowered as well.

1.5 Software environment: Simulation and data processing

During the development of the detector and for the preparation of physics analyses tools are needed which simulate the physics processes of particle interactions and the detector response to them in great detail. This is done by the following two programs:

- **Gauss** is the simulation framework of the LHCb software. Monte Carlo generators like `PYTHIA` create proton-proton collisions. Then the particles are propagated through the LHCb detector. The material interactions are described with the `GEANT` package. The whole detector response is simulated and a collection of *associated hits* for every particle is created. The entire history of a particle is stored so that it can be accessed later.

- The digitisation of the event is then done by **Boole**. This means that the simulated hits are transformed into a format which is equivalent to the output of the detector from real data taking.

From this stage on, there is no difference in processing simulated events or events which have been recorded in proton-proton collisions. However, in the simulation it is possible to access the history of a particle to evaluate performance quantities like e.g. the tracking efficiency or the efficiency to select certain decays. To

distinguish between reconstructed and simulated particles the simulated particles will be referred as *Monte Carlo particles*.
The following two programs were used in this work:

- The full event reconstruction is done by **Brunel**. It performs the track finding and the particle identification. On simulated data *tracks are associated* to Monte Carlo particles to be able to test the performance of the track finding.
- After the reconstruction of the event, the analysis is done by **DaVinci**. In this phase a particle hypothesis is assigned to the reconstructed tracks. With this hypothesis tracks can be combined to reconstruct particle decays. To separate background from signal, informations from all subdetectors are combined to select interesting events, such as events containing a B hadron.

In process of time the LHCb software environment evolved. The first versions were developed before the construction of detector was finished and even with the first collisions it is still work in progress. Thus different simulations were necessary because the description of the geometry and the material budget of the detector was changed. Furthermore the B-field map of the LHCb dipole was measured under real conditions. All these things lead to a more realistic description of the LHCb experiment. Additionally the running conditions of the LHC were modified.

In the last years two major Monte Carlo productions were performed and used for performance studies, one in 2006 (DC06) and one in 2009 (MC09). They differ in a more realistic description of the detector setup and the center-of-mass energy was changed from 14 TeV to 10 TeV in 2009. Additionally the data taking of 2009 required a simulation of 900 GeV collisions.

1.6 Physics programme of the LHCb experiment

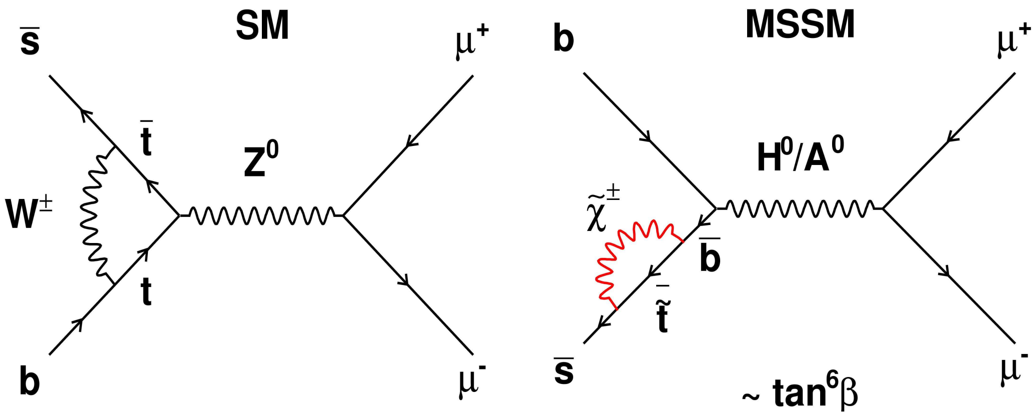
The Standard Model of particle physics (SM) presents todays knowledge of particle physics. It was developed in the 1960's by Glashow [6], Salam [7] and Weinberg [8] and up to today it is consistent with all measurements at collider experiments. However, it cannot explain observations like:

- **Gravitation:** The description of gravitation is not included in the SM.
- **Neutrino oscillations:** In the SM neutrinos are massless but the observation of neutrino oscillations requires massive neutrinos.
- **Dark Matter and Dark Energy:** Cosmological observations have shown that the amount of matter and energy in the Universe cannot be explained by the particles in the SM.
- **Baryon asymmetry:** The level of CP violation in the SM cannot explain the excess of matter over antimatter in the Universe.

There are a lot of theories, e.g. Supersymmetry (SUSY) or Extra Dimensions, which try to explain these problems. Most of the "Beyond the Standard Model" theories have in common that they lead to new particles. In SUSY models every SM particle has an additional partner. Fermions get a bosonic partner and bosons get a fermionic partner. As there has not been an observation of non-SM particles yet, these particles have to be heavier than the SM ones. One way to detect new particles is to search for peaks in invariant mass distributions. This is done by the multipurpose experiments ATLAS and CMS. A complementary approach is to search for new particles indirectly by observing loop processes. As the particles in loop processes are virtual, very high energies can be probed. SM observables which are sensitive to quantum corrections are e.g. branching ratios of rare decays or CP violating phases because the New Physics enters at the same order as the SM processes. Thus these are excellent processes to look for New Physics contributions in loop diagrams.

A selection of the key measurements of LHCb which are expected to give the first significant results before the shutdown in 2012 will be introduced in the following. A much more detailed description can be found in [9].

1.6.1 Rare decays



(a) SM process (b) MSSM process

Figure 1.9: *Examples of Feynman diagrams contributing to the decay $B_s \to \mu^+\mu^-$ (a) in the SM and (b) in a certain SUSY model (MSSM) (Figure from [9]).*

Due to the high number of B_s mesons produced at the LHCb also decays with a very low branching ratio in the SM can be measured. One particular example is the decay $B_s \to \mu^+\mu^-$ which is forbidden at tree level in the SM as it would require a flavour changing neutral current. Flavour changing neutral currents appear in the SM only in loop diagrams. The lowest order SM process is shown in Figure 1.9(a). The SM model prediction of the branching fraction is computed to be [10]

$$\mathcal{BR}(B_s \to \mu^+\mu^-) = (3.35 \pm 0.32) \times 10^{-9}. \tag{1.1}$$

The current upper limit is determined by the Tevatron [11]

$$\mathcal{BR}(B_s \to \mu^+\mu^-) < (3.6) \times 10^{-8}. \tag{1.2}$$

Figure 1.9(b) shows a process in the Minimal Supersymmetric Model (MSSM) where a SUSY particle contributes to the same order as the SM process. This would increase the branching fraction due to an additional flavour changing neutral current.

The decay $B_s \to \mu^+\mu^-$ has a very clear signature as it has two muons in the final state. Studies have shown that LHCb could already exclude many SUSY models with a luminosity[3] of $0.1\,\mathrm{fb}^{-1}$ [9]. A 3σ observation of the SM process could be reached with $3\,\mathrm{fb}^{-1}$.

1.6.2 Electroweak penguin decays

Another interesting channel with a flavour changing neutral current is the decay $B_0 \to K^{*0}\mu^+\mu^-$ with the transition $b \to sl^+l^-$. Figure 1.10 shows the lowest order penguin diagrams for this decay. New physics processes enter at the same level as the SM processes. Thus this channel is also sensitive to beyond the SM contributions.

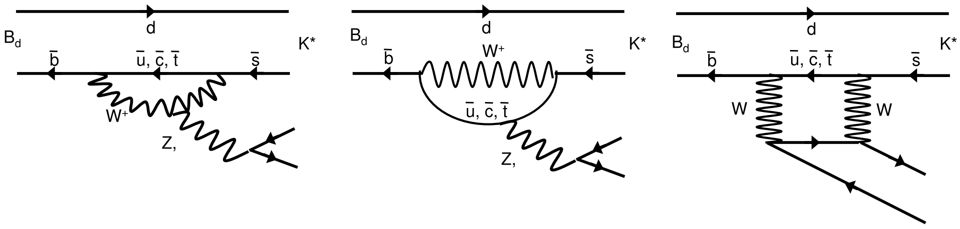


Figure 1.10: *Lowest order SM Feynman diagrams for the $B_0 \to K^{*0}\mu^+\mu^-$ decay. (Figure from [9])*

For this channel not only the high number of B mesons but also the good tracking resolution and the good muon identification of LHCb are needed. A large number of observables from the angular distributions of the decay can be constructed. One particularly interesting variable which can probe the SM in this decay is the forward backward asymmetry A_{FB} of the muons in the di-muon rest frame. Forward and backward direction are defined by the angle θ_L between the positive (positive) muon and the flight direction of the B_0 ($\bar{B_0}$) meson in the

---

[3] $0.1\,\mathrm{fb}^{-1}$ is expected by the end of 2010.

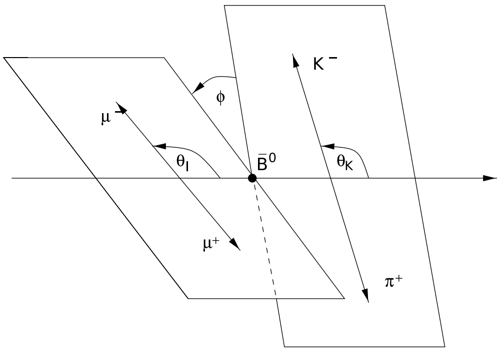


Figure 1.11: *Definition of the angles in the decay $B_0 \rightarrow K^{*0}\mu^+\mu^-$. θ_L is the angle between the μ^- and the $\bar{B}_0$ in the di-muon rest frame (Figure from [9]).*

di-muon rest frame, cf. Figure 1.11. A_{FB} depends on the di-muon invariant mass (q^2) and is then defined as

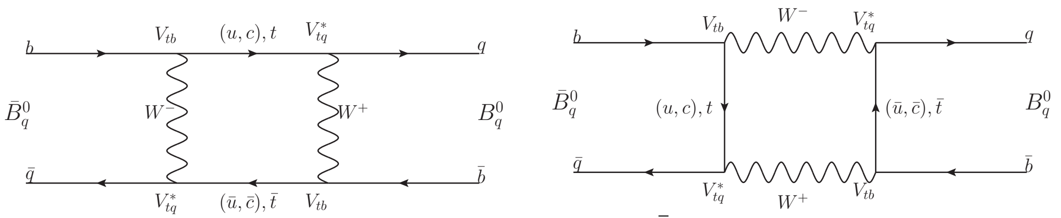
$$A_{FB}(q^2) = \frac{\#\text{forward}(q^2) - \#\text{backward}(q^2)}{\#\text{forward}(q^2) + \#\text{backward}(q^2)} \tag{1.3}$$

In the SM it is predicted that A_{FB} crosses the q^2 axis at $q^2 = 4.36^{+0.33}_{-0.31}$ GeV [12]. In this point the main theoretical uncertainties coming from QCD form factors cancel. Thus a deviation from the SM would be a clear sign of New Physics.

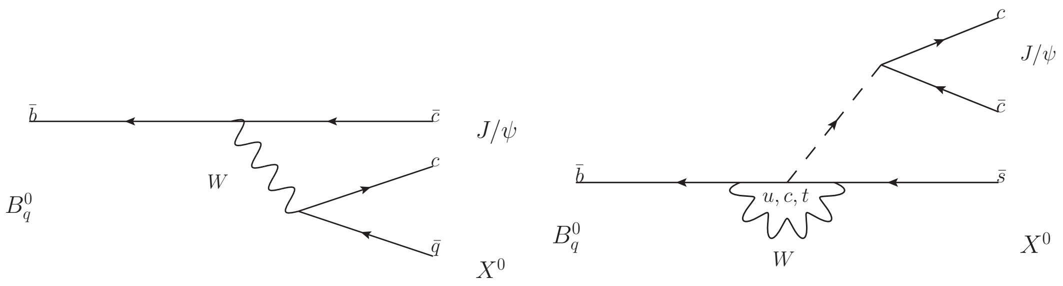
1.6.3 CP violation in the B meson system

One of the key measurements of LHCb is to measure the CP violating phase $\Phi_s^{J/\psi\ \phi}$ in the decay $B_s \rightarrow J/\psi\ \phi$. CP violation appears because the B_s meson can directly decay to $J/\psi\ \phi$ or it can first oscillate to a $\bar{B}_s$ meson, cf. Figure 1.12, and then decay to $J/\psi\ \phi$ which causes interference effects. In the SM $\Phi_s^{J/\psi\ \phi}$ depends only on the CKM Matrix elements and is predicted to be $-2\beta_s$. β_s is defined as $\beta_s = arg(-V_{ts}V_{tb}^*/V_{cs}V_{cb}^*)$. The phase $\Phi_s^{J/\psi\phi}$ has very low theoretical uncertainties within the SM and is thus a good candidate to search for New Physics contributions. The indirect determination via global fits to data gives $2\beta_s = 0.0360^{+0.0020}_{-0.0016}$ [13]. Beyond the SM particles could give an additional phase to the $B_s - \bar{B}_s$ box diagram.

One of the reference channels for $B_s \rightarrow J/\psi\ \phi$ is the so called “Golden Channel” $B^0 \rightarrow J/\psi\ K_s$. This channel has already been studied in detail by the B-factories, BaBar and Belle. Here the CP violation is also induced by the interference between direct decay and oscillation. The Feynman diagrams have the same topology as in the case of $B_s \rightarrow J/\psi\ \phi$, cf. Figure 1.12. The only relevant contribution to the CP violation is the term induced by mixing, $\mathcal{A}_{CP}^{mix}$. The time dependent CP



(a) The dominant diagrams for the $B_q - \bar{B}_q$ mixing in the SM.



(b) Feynman diagrams contributing to the decay $B^0_q \to J/\psi\ X^0$. in the SM

Figure 1.12: *Feynman diagrams responsible for CP violation in the B meson system with $q = d, s$ and $X^0 = K_s, \phi$.*

asymmetry is described by

$$\mathcal{A}_{J/\psi K_s} = \frac{\Gamma(B^0 \to J/\psi\ K_s) - \Gamma(\bar{B}^0 \to J/\psi\ K_s)}{\Gamma(B^0 \to J/\psi\ K_s) + \Gamma(\bar{B}^0 \to J/\psi\ K_s)} = sin(\Delta m_{B^0} t)\mathcal{A}^{mix}_{CP}, \qquad (1.4)$$

where Δm_{B^0} is the mass difference between the two B^0 mass eigenstates and $\mathcal{A}^{mix}_{CP} = sin(2\beta)$ with $\beta = arg(-V_{cd}V^*_{cb}/V_{td}V^*_{tb})$ in the SM. The current world averaged result is [14]

$$sin(2\beta) = 0.673 \pm 0.023. \qquad (1.5)$$

The statistical sensitivity of LHCb after $1\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 7\,\mathrm{TeV}$ is expected to be about 0.046, cf. [15].

This channel is particularly interesting for this work as the K_s has a long lifetime ($\tau = (8.953 \pm 0.005) \times 10^{-11}$) on the scale the LHC is operating. This means that about 3/4 of the K_s produced in B^0 decays decay after the Velo (1/4 even behind the TT). Therefore these particles have to be reconstructed by the TT and the T-Stations only. The algorithm to reconstruct K_s decay products has been optimised within this thesis presented here and will be discussed in detail in Chapter 3, 4 and 5.

Chapter 2

Tracking at LHCb

In this chapter a short overview over the tracking system of the LHCb experiment and its performance is given.

The particles passing the tracking system leave hits in the detector which are then used to reconstruct their trajectories. The tracking is the most vital part in the reconstruction of an event. No physics analysis can be performed without tracks. Besides providing precision vertex and momentum information, the reconstructed tracks are crucial for particle identification. Calorimeter cluster, muon station hits and Cherenkov rings need associated tracks to be properly reconstructed. The tracking consists of different stages which can be summarized in three steps:

- **Pattern recognition:** The pattern recognition is the first step in the tracking. Its task is to find hits that belong to the same particle. As there are a lot of hits in the detector, the pattern recognition has to deal with large combinatorics and has to be fast and efficient at the same time. E.g. in the software trigger (HLT) the whole event is reconstructed. There the timing is very crucial, compared to the reconstruction of the stored events where the focus is more on reconstruction efficiency and purity. The efficiency of the pattern recognition is of high importance because to reconstruct a B hadron decay several tracks have to be found. E.g. in the case $B_s \rightarrow J/\psi(\mu^+\mu^-)\ \Phi(K^+K^-)$ four charged particles have to be reconstructed. Thus in this case the decay reconstruction depends in the fourth order on the track reconstruction efficiency.
  The pattern recognition consists of several algorithms. Some algorithms are only looking at one subdector, other algorithms combine the informations from different subdetectors. Then again some pattern recognition algorithms are redundant. They find essentially the same tracks but still have different strengths. The optimal performance is then achieved by combining them. The different pattern recognition algorithms are introduced in Chapter 2.2.

- **Fitting:** The tracks that are found by the pattern recognition are fitted afterwards. The fitter uses a *Kalman filter* approach [16][17] to account for multiple scattering and energy loss in the detector. This part is the most time consuming due to the CPU intense propagation through the magnetic field and the computation of material interactions. The optimal parameter estimate of the track and the estimate of the according uncertainties is obtained in this step. The fitting procedure is described in [18] and [19].

- **Clone killing:** The last step is the *CloneKiller* [20]. In this stage tracks that have been found more than once are eliminated. This is mostly because of the redundant pattern recognition, but it can also happen that the same pattern recognition algorithm creates two tracks with similar hit content. The tracks passing the *CloneKiller* are merged in a collection which is then used for physics analyses afterwards.

2.1 Track types

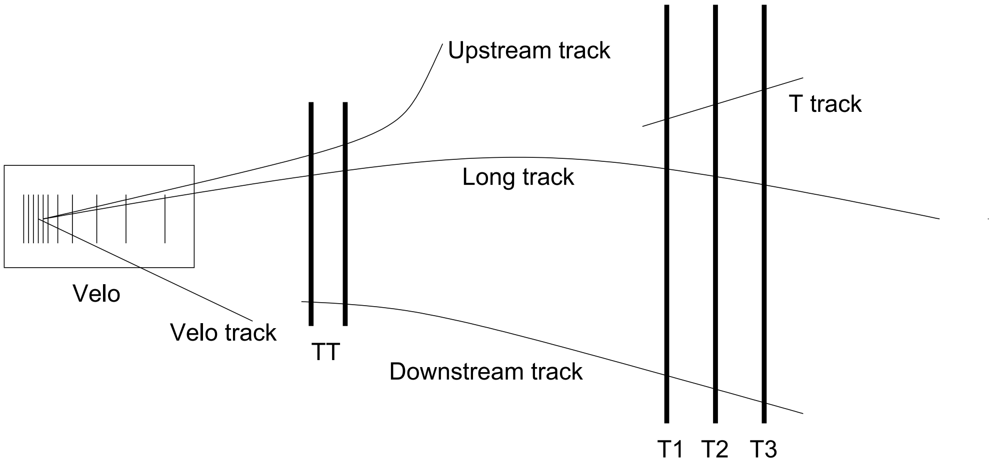


Figure 2.1: *Definition of track types at LHCb [21].*

Depending on the subdetectors that were used to reconstruct a track, one distinguishes between different types of tracks in the LHCb detector, cf. Figure 2.1.

- **Long tracks** have the best vertex and momentum resolution and are the most useful for physics analyses. They traverse the complete tracking system.
- **Upstream tracks** are reconstructed with the Velo and the TT. They are used to reconstruct low momentum particles which are bent out of the detector by the magnet.
- **Downstream tracks** are reconstructed with the TT and the T-Stations. They allow the reconstruction of decay products of long lived particles like K_s and Λ hadrons.

- **Velo tracks** have only hits in the Velo. They are useful for the reconstruction of the primary vertex. The primary vertex is the position of the proton-proton collision, and they are used as input to the Long and Upstream track reconstruction.

- **T-tracks** are reconstructed in the T-Stations only. They are used as input to the Long track and Downstream track reconstruction. Additionally they are useful by themselves for the alignment of the T-Stations.

2.2 Pattern recognition algorithms

For the different track types there are the following pattern recognition algorithms. They are described in the same order as they are executed in the default LHCb tracking sequence:

- **Velo tracking:** The Velo tracking is done in two stages. In the first stage tracks in the $r - z$ plane are searched. Afterwards hits in the ϕ sensors are added. The B-field in the Velo is sufficiently low that a straight line model for tracks can be used.

- **Forward tracking:** The Forward tracking searches for Long tracks. It starts with a track in the Velo (the seed) and uses a *Hough-transformation* approach to search for hits in the T-Stations. The Forward tracking is described in [22]. Hits in the TT are not used in the pattern recognition. They are added afterwards to improve the momentum resolution.

- **T-Seeding:** The T-Seeding searches for tracks in the T-Stations without using any information from other detectors. Due to the two different detectors in the T-Stations, the OT and the IT, the ambiguity of hits in the OT and the non-negligible fringe field, the pattern recognition is challenging. There are two different approaches to search for tracks in the T-Stations. The reader is referred to [23] and [24]. The name T-Seeding comes from the fact that T-Tracks are used as a seed for the following two algorithms.

- **Track matching:** The Track matching is the second algorithm that searches for Long Tracks. It uses seeds in the Velo and the T-Stations and tries to match them in the bending plane of the magnet. Again hits in the TT are added afterwards.

- **Downstream tracking:** The Downstream tracking uses hits in the TT and T-Tracks as input. The algorithm is explained in detail in Chapter 3.

2.3 Performance criteria

To measure the performance of a pattern recognition algorithm, it is necessary to define some quantities.

First of all the **efficiency** of an algorithm has to be defined. The definition has to take into account that in some cases the reconstruction of a particle is not possible because it leaves too less hits in the detector. Therefore the denominator of the efficiency is defined by **reconstructible** particles.
A Monte Carlo particle is called **reconstructible** [25]

- in the Velo, when it has at least three hits in the r and ϕ sensors.
- in the TT, when it has at least one hit in the first two planes (TTa) and one hit in the last two planes (TTb).
- in the T-Stations, when there is at least one x and one stereo hit in each of the three T-Stations.

A Monte Carlo particle is called **reconstructible**

- as a Long track, when it is reconstructible in the Velo and the T-Stations.
- as an Upstream track, when it is reconstructible in the Velo and the TT.
- as a Downstream track, when it is reconstructible in the TT and the T-Stations.

The definition of the numerator has to reflect that sometimes not all hits of a particle are found or that not all hits on a track are from the same particle, but there can still be enough information on a track to determine its parameters. Thus a Monte Carlo particle (MC particle) is called **reconstructed** or **truth-matched/associated** when at least 70 % of the hits on the track come from this particle. The hits that belong to a Monte Carlo particle are also referred as **true** or **associated** hits. The definition of efficiency is then

$$\epsilon = \frac{\text{number of }\textbf{reconstructed}\ \wedge\ \textbf{reconstructible}\text{ Monte Carlo particles}}{\text{number of }\textbf{reconstructible}\text{ Monte Carlo particles}}. \tag{2.1}$$

Other quality criteria of interest are:

- **Ghost rate:** A *ghost track* is a track which has no connection to a particle. The ghost rate is defined by the fraction of tracks which are not associated to a Monte Carlo particle.

$$\textbf{ghost rate} = \frac{\text{number of }\textbf{non-associated tracks}}{\text{number of }\textbf{all found tracks}} \tag{2.2}$$

- **Clone rate:** It could be that the same particle is reconstructed two or more times, either by the same or by different algorithms. Tracks are called clones if they are associated (ass.) to the same Monte Carlo particle.

$$\textbf{clone rate} = \frac{\text{number of } \textbf{MC particles with more than one associated track}}{\text{number of } \textbf{all reconstructed particles}} \tag{2.3}$$

- **Hit efficiency:** The hit efficiency measures if the pattern recognition collects all hits that the particle left in the detector.

$$\textbf{hit efficiency} = \sum_{tracks} \frac{\text{number of } \textbf{ass. hits from ass. MC particle on the track}}{\text{number of } \textbf{all associated hits of the MC particle}} \tag{2.4}$$

- **Hit purity:** The hit purity measures how many hits on a track do not belong to the associated particle and is defined by the fraction of associated hits on that track.

$$\textbf{hit purity} = \sum_{tracks} \frac{\text{number of } \textbf{ass. hits on track which are ass. to same MC particle}}{\text{number of } \textbf{all hits on the track}} \tag{2.5}$$

- **CPU time:** Also the running time of an algorithm is of interest. E.g. the pattern recognition is performed in the trigger where timing is very crucial.

The above mentioned definitions are called *track-averaged* quantities. Another method is to calculate the quantity first in each event and then the per-event results are averaged over all events. By doing this every event has the same weight. The quantity is then called an *event-averaged* quantity. In *track-averaged* quantities events with a high track multiplicity have a larger weight. Especially in the case of the ghost rate both definitions are used. The *event-averaged* result is always a few percent lower compared to the *track-averaged* result. However, only *track-averaged* quantities are quoted within this work.

2.4 Performance overview

In this section a short overview of the performance[1] of the pattern recognition algorithms is given.

Used samples

To test the efficiency, two different samples are used. One type is called a **minimum bias sample**. Minimum bias samples try to reproduce proton-proton

---

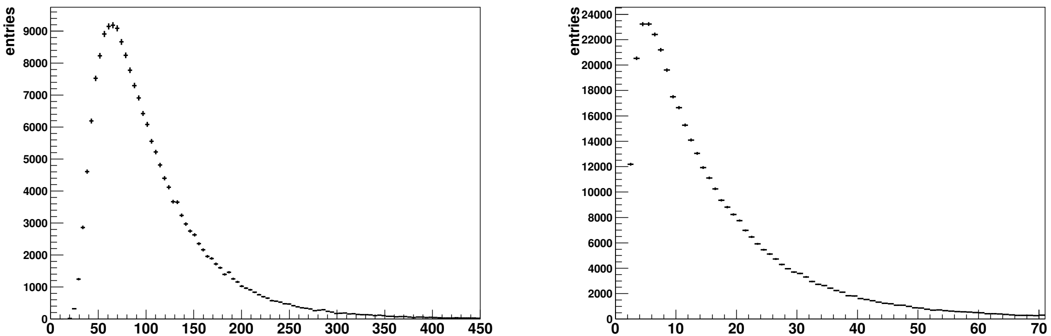
[1] Brunel v36r0p1 was used.

collisions as close to reality as possible. This means the current knowledge of the underlying physics processes is simulated. The other sample has at least one $B^0 \rightarrow J/\psi\ K_s$ decay in every event. This is necessary to get a decent amount of the events one wants to study. E.g. in only about 0.25 % of the proton-proton collisions a $b\bar{b}$ pair is produced and looking at a certain decay, millions or billions of proton-proton collisions have to be simulated to get a single event.

Here a $B^0 \rightarrow J/\psi\ K_s$ sample is chosen to show the tracking performance in B events. Both samples[2] are simulated with a center-of-mass energy of 10 TeV. For the measurement of the tracking performance only L0 triggered events are taken into account. These are the events the pattern recognition has to deal with, either in the software trigger or later in the processing of the stored data.

2.4.1 Pattern recognition efficiency

Figure 2.2(a) shows the typical momentum distribution of a B^0 which decays into a J/ψ and a K_s. The K_s further decays into two pions and the J/ψ into two muons. The momentum distribution of the pion tracks is shown in Figure 2.2(b). Most of them have a momentum greater than 5 GeV. Thus the efficiency measurement is split into all tracks and tracks with a momentum larger than 5 GeV where the later category is the most relevant for the reconstruction of B decays. The efficiency of B hadron daughter tracks is also listed separately.



(a) B^0 momentum distribution. (b) Pion momentum distribution.

Figure 2.2: *Momentum distribution of truthmatched $B^0 \rightarrow J/\psi\ K_s$.*

First the efficiency of the Velo tracking and the T-Seeding is shown because they are seeds for other types of tracks. For the Long tracks it is discussed how two redundant algorithms increase the overall performance. And in the end the efficiency of Downstream tracks is shown. This algorithm is described in detail in Chapter 3.

---

[2] The samples are from the MC09 production.

The minimum bias sample is taken for the overall efficiency of all tracks and the ghost rate. The $B^0 \to J/\psi\ K_s$ sample is used to show how efficient the tracking for particles from B decays is. Relatively B events are more busy than minimum bias events.

Velo tracking and T-Seeding

For Velo tracks only the efficiency for as Long track reconstructible particles is quoted. For T-Tracks the efficiency for particles which are reconstructible as Long tracks or as Downstream tracks, for completeness, is given.

The efficiency of the Velo tracking is 97.3 % for all tracks and the ghost rate is 5.7 %, cf. Table 2.1. The reconstruction efficiency of T-Tracks increases from 84.1 % for low momentum ($p < 5\,\mathrm{GeV}$) tracks to 95.7 % for tracks with a momentum larger than 5 GeV, cf. Table 2.2. The efficiency to reconstruct a track from a B hadron decay is higher as the average momentum is larger.

	Efficiency [%]
All tracks	$97.3 \pm < 0.1$
$p > 5\,\mathrm{GeV}$	$97.8 \pm < 0.1$
B decay tracks	$97.6 \pm < 0.1$
$p > 5\,\mathrm{GeV}$	$98.2 \pm < 0.1$
Ghost rate	5.7 %

Table 2.1: *Velo tracking performance.*

	Efficiency [%]	
	Long	Downstream
All tracks	$91.5 \pm < 0.1$	$91.1 \pm < 0.1$
$p > 5\,\mathrm{GeV}$	$95.7 \pm < 0.1$	$95.7 \pm < 0.1$
B decay tracks	94.3 ± 0.1	94.3 ± 0.1
$p > 5\,\mathrm{GeV}$	96.5 ± 0.1	96.7 ± 0.1
Ghost rate	4.7 %	

Table 2.2: *T-Seeding performance.*

The efficiency is defined with respect to as Long and Downstream track reconstructible particles.

Long track reconstruction

Table 2.3 summarizes the performance of the two Long track algorithms. The third column shows the efficiency after the output of the two is combined by the *CloneKiller* algorithm. The Forward tracking has a better efficiency than the Track Matching, 85.7 % compared to 82.9 %, but the ghost rate of the Forward tracking is higher, 14.4 % compared to 10 %. The combination of both increases the efficiency to 95.9 % for B daughter tracks with a momentum larger than 5 GeV. This number means that about $0.96^4 \approx 85$ % of decays with a topology like $B_s \to J/\psi(\mu^+\mu^-)\ \phi(K^+K^-)$ can be reconstructed by the tracking system[3].

The ghost rate of the combination is 17.2 %. This number might seem high but ghost tracks can still be discarded later. The first tool to discriminate real tracks from ghost tracks is the fitter after the pattern recognition. The χ^2/ndf of

---

[3] This is true if the particles are reconstructible as Long tracks, this means they are in the detector acceptance.

the fit result is in average lower for real tracks compared to ghost tracks. Later e.g. in the analysis of an event also informations from the particle identification can be used. In most cases the ghost track would not point to a cluster in the calorimeters so that it has no particle hypothesis. Or when ghost tracks are used together with other tracks to reconstruct a particle decay, the invariant mass distribution does not show a signal excess. The pattern recognition itself was optimised to a high efficiency.

	Efficiency [%]		
	Forward tracking	Match tracking	Combined
All tracks	85.7 ± < 0.1	82.9 ± < 0.1	90.2 ± < 0.1
$p > 5\,\mathrm{GeV}$	91.9 ± < 0.1	89.0 ± < 0.1	94.1 ± < 0.1
B decay tracks	89.5 ± 0.2	88.1 ± 0.2	93.3 ± 0.1
$p > 5\,\mathrm{GeV}$	93.7 ± 0.1	91.8 ± 0.2	95.9 ± 0.1
Ghost rate	14.4 %	10.0 %	17.2 %

Table 2.3: *Performance of Long track reconstruction for as Long track reconstructible particles.*

Downstream track reconstruction

As seen in Table 2.4 the tracking efficiency of the Downstream tracking[4] is a few percent lower compared to the Long track algorithms. But it is still more than 90 % for tracks from B hadron decays ($p > 5\,\mathrm{GeV}$). The ghost rate is much higher with 29.8 %. Still this is only in the output of the Downstream algorithm. After the clones of Long tracks have been removed, the remaining ghost rate is in the order of 50 %. In Chapter 4.6 a selection of K_s with only a few cuts is used and a clear peak can be seen, despite the high ghost rate. Thus the focus of this pattern recognition algorithm is as well more on efficiency than on ghost rate.

	Efficiency [%]
All tracks	78.1 ± 0.1
$p > 5\,\mathrm{GeV}$	85.8 ± 0.1
B decay tracks	86.3 ± 0.2
$p > 5\,\mathrm{GeV}$	91.3 ± 0.2
Ghost rate	29.8 %

Table 2.4: *Downstream tracking performance for as Downstream track reconstructible particles.*

2.4.2 Runtime

Table 2.5 summarizes the runtime of the different tracking stages and illustrates which part takes the longest time. As mentioned before the most time consuming

---

[4] The optimisation of the algorithm presented in Chapter 4 is already included.

part is the fitting stage after the pattern recognition. E.g. in the case of the Forward tracking it takes about 5 times longer to fit the tracks than to find them. The runtime of the Downstream tracking is compared to the other algorithms rather low, a fifth of the runtime of the Forward tracking. Here the fitting time is more than 20 times higher than the pattern recognition part.

	Event averaged runtime [ms]		
	Pattern recognition	Fitting	CloneKiller
Velo tracking	7.4	40	
Forward tracking	24.5	152	
T-Seeding	35.7	93	6.7
Track Matching	5.9	131	
Downstream tracking	4.4	118	

Table 2.5: *Runtime of the different tracking stages. The measurement was done on a minimum bias sample and about 6000 L0 triggered events. The runtime depends on the used computer. Thus only relative times should be considered.*

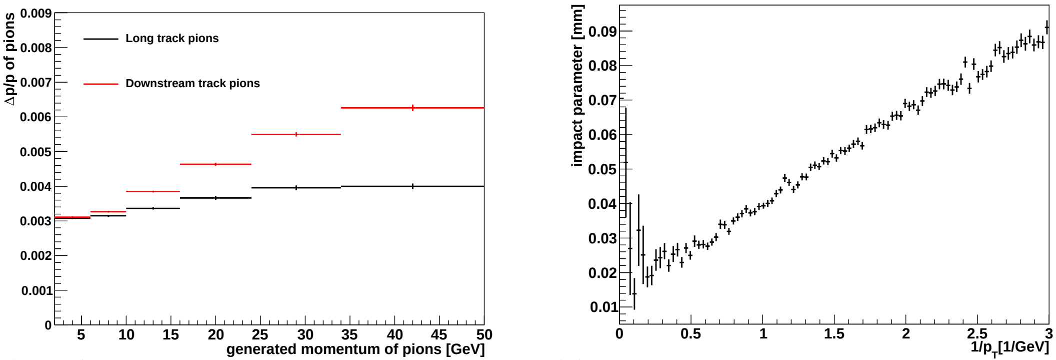
2.4.3 Momentum and mass resolution

The momentum and mass resolution of the detector will be discussed with the decay $B^0 \rightarrow J/\psi\ K_s$. A comparison between Long and Downstream tracks can be made because both are needed for the reconstruction. The K_s originating from the B^0 have a large flight distance. This means that only 25 % decay inside the Velo and 50 % decay between the Velo and the TT. The remaining particles decay after the TT and are therefore not reconstructible. The J/ψ is always reconstructed with two Long tracks as it decays almost immediately. Thus the sample of reconstructed decays consists of about 1/3 where the K_s is reconstructed with two Long tracks and 2/3 where two Downstream tracks are used. A combination of of both types is not taken into account.

The figures in the following chapter contain only truthmatched reconstructed particles. A reconstructed K_s candidate is truthmatched to a generated K_s when both daughter tracks are associated to the generated pions which originated in the K_s decay. Reconstructed $B^0 \rightarrow J/\psi\ K_s$ decays are associated accordingly.

Momentum and vertex resolution

Figure 2.3(a) shows the relative deviation of the generated momentum to the reconstructed momentum over the momentum of the Monte Carlo particle. The momentum resolution of Long track pions is always better compared to the resolution of Downstream track pions. The better momentum resolution of the Long tracks is mostly due to the longer lever arm. The momentum is essentially



(a) $\Delta p/p$ against p of truthmatched pions is shown for Downstream and Long tracks

(b) Reconstructed impact parameter of truthmatched particles (Long tracks) from the primary vertex against their inverse transverse momentum.

Figure 2.3: *Momentum and vertex resolution are shown.*

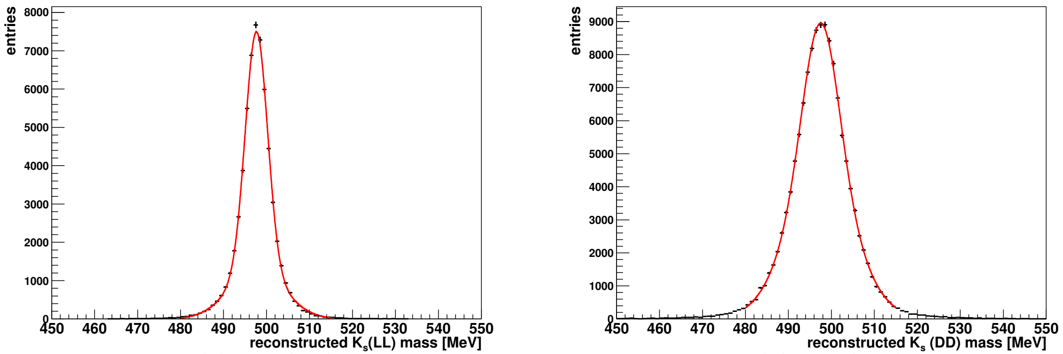
measured by determining the curvature in the B-field, resp. the slope of a track in front of and behind the magnet.

Figure 2.3(b) shows the average impact parameter of as Long track reconstructed tracks which originate from the primary vertex against their inverse transverse momentum. The impact parameter is the distance of closest approach of a track to the primary vertex. This figure shows the vertex resolution of Long tracks as it should be equal to 0 in the ideal case. For particles with a high transverse momentum the vertex resolution is about 25 μm and increases up to 90 μm for particles with a small transverse momentum.

Mass resolution

The mass resolution of the K_s is shown in Figure 2.4. The K_s reconstructed with two Downstream tracks have a mass resolution of 6.7 MeV which is 2.5 MeV worse compared to Long tracks. This deviation is due to the before mentioned worse momentum resolution but also due to the fact that the TT is inside the fringe field of the magnet. Thus the tracks of the two daughter particles have to be propagated through the B-field to the K_s decay vertex. This gives an additional uncertainty which increases with the length of the propagation as it is shown in Figure 2.5 where the mass resolution is plotted against the z coordinate of the K_s decay vertex.

After combining the K_s with a J/ψ the invariant mass of the B^0 is reconstructed, cf. Figure 2.6. The mass resolution for Long track combinations is (9.1 ± 0.2) MeV and for Downstream track combinations (9.6 ± 0.2) MeV. The combination of both samples has a resolution of (9.4 ± 0.2) MeV because the Long track sample contributes 1/3 and the Downstream track sample 2/3. The



(a) K_s (LL) (b) K_s (DD)

Figure 2.4: *The reconstructed mass distribution of truthmatched K_s which originate in B^0 decays is shown. A double Gaussian (Definition in Chapter 6) is fitted (red). LL (DD) stands for the combination of two Long (Downstream) tracks. The weighted width is* $(4.2 \pm 0.01)\,\mathrm{MeV}$ *resp.* $(6.7 \pm 0.01)\,\mathrm{MeV}$.

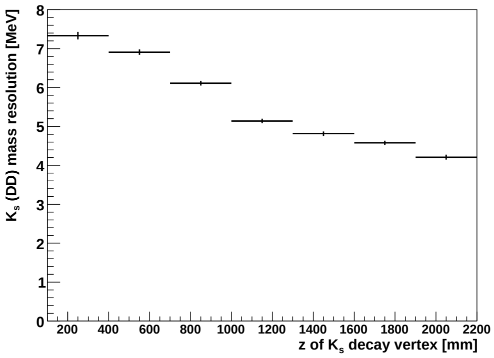
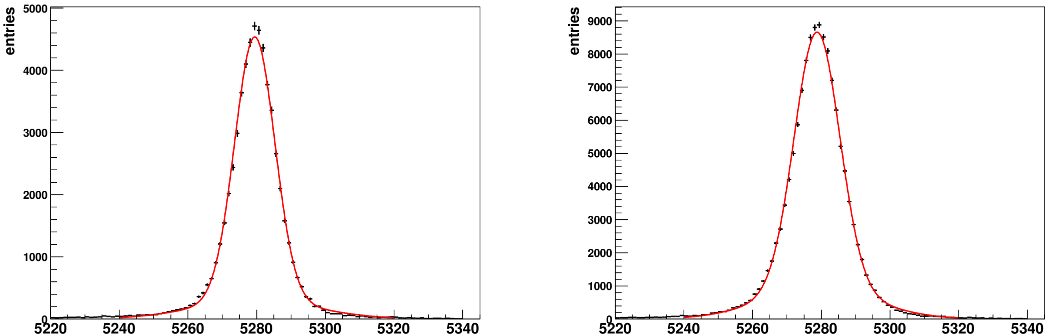
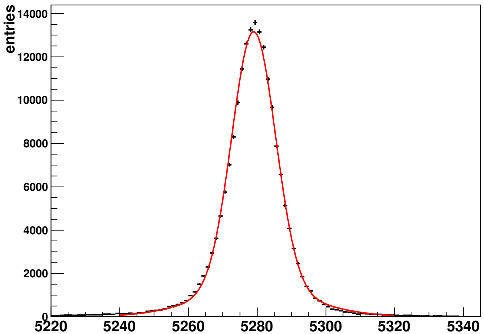


Figure 2.5: *The K_s mass resolution is shown against the z coordinate of the K_s decay vertex. The K_s candidates are reconstructed with two Downstream tracks (DD).*

difference is smaller compared to the difference in the mass resolution of the K_s. The B^0 mass resolution is mostly dominated by the resolution of the J/ψ which is about 12 MeV. The mass resolution of the J/ψ is worse because it is more determined by the momentum measurement as it is the case for the K_s. The difference between the sum of the masses of the two muons and the J/ψ mass is roughly $m_{J/\psi} - 2m_\mu \approx 2.9\,\mathrm{GeV}$. The mass of the muons is not reconstructed. It is taken as a constraint in the calculation. Thus about 90 % of the invariant mass of the J/ψ candidate are determined by the momentum measurement. For the K_s it is only about 50 %. The B^0 is then reconstructed with the 4-momenta of the K_s and the J/ψ.



(a) B^0 (LL) (b) B^0 (DD)



(c) B^0

Figure 2.6: *The reconstructed mass distribution of truthmatched* $B^0 \rightarrow J/\psi\ K_s$ *decays is shown. A double Gaussian (Definition in Chapter 6) is fitted (red). LL (DD) are* B^0 *candidates which are reconstructed with a* K_s *that is built of two Long (Downstream) tracks. The weighted widths are (a)* $(9.1 \pm 0.2)\,\mathrm{MeV}$*, (b)* $(9.6 \pm 0.2)\,\mathrm{MeV}$ *and (c)* $(9.4 \pm 0.2)\,\mathrm{MeV}$*.*

2.4.4 Event display

To give an impression of the whole tracking, an event display of one of the first 7 TeV collisions is shown in Figure 2.7. Hits in the tracking detectors are marked with green crosses. The active region of the Velo is so small compared to the dimension of the detector so that it is hardly seen. Hits in the TT are the green crosses directly in front of the magnet. The yellow dots are hits in the RICH1 detector. The T-Stations can be seen behind the magnet in front of the RICH2 detector. The two green tracks which extend into the Muon stations are Long tracks which are linked to hits in the Muon system. The red and blue bars are clusters in the calorimeters.

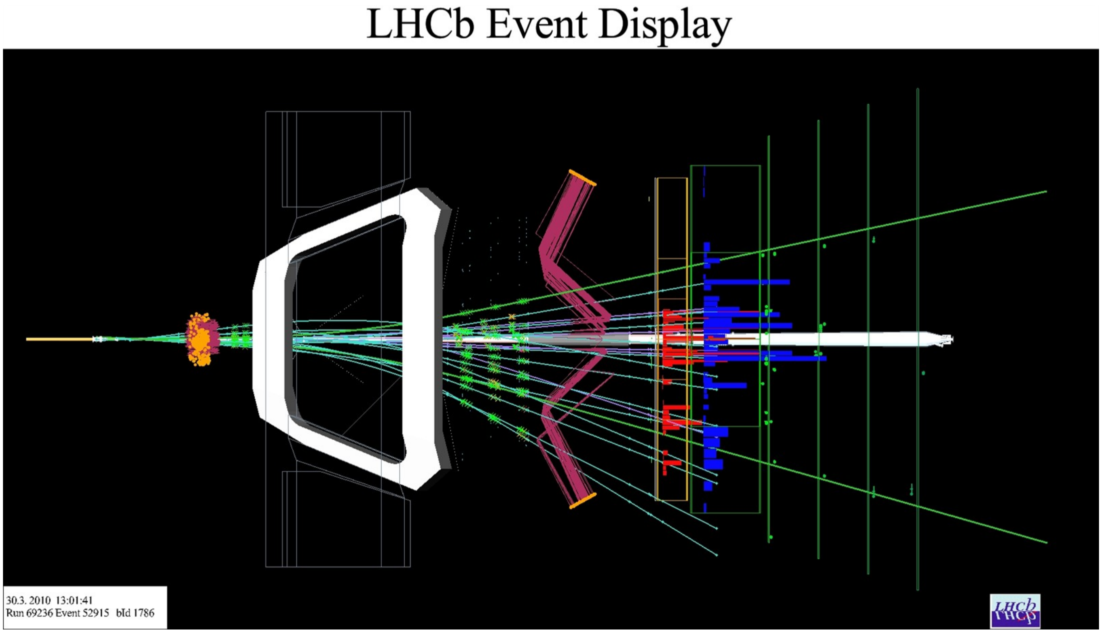

LHCb Event Display

Figure 2.7: *Event display of one of the first 7* TeV *collisions [1].*

Chapter 3

Downstream Tracking

In this chapter the Downstream algorithm which reconstructs displaced tracks is described in detail. Improvements to this algorithm are presented in Chapter 4.

In the previous chapter it was seen that Long tracks provide in general the most accurate track estimate. However, there are situations that require the use of Downstream tracks. The first point is that one needs an algorithm for the reconstruction of daughters of long lived particles such as K_s and Λ which often decay after the Velo. There are several B-decay channels which have long lived particles in the decay chain. One example is the aforementioned decay $B^0 \rightarrow J/\psi\ K_s$ which is the “Golden Channel” for CP violation measurements in the B^0 system. It has already been measured with high precision at other colliders, thus it is one of the reference channels which will be measured by LHCb. As seen in Chapter 2.4 the statistics of this analysis can be improved by a factor of three by using K_s which are reconstructed with two Downstream tracks. However, there are also other channels like $B^0 \rightarrow K_s\ K_s$ or $B_s \rightarrow J/\psi\ K_s$. Furthermore there are studies that measure the K_s or Λ production in proton-proton collisions, one will be presented in Chapter 6. Another feature of the Downstream tracking is the

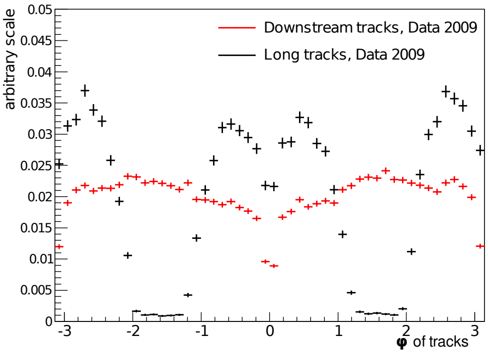


Figure 3.1: *ϕ distribution of Long and Downstream tracks in data from the 2009 run. The reduced acceptance of the Velo can be seen.*

possibility to reconstruct tracks independently of the Velo. As mentioned before the Velo can be moved away from the beam pipe. Normally it is only done during the injection phase but as the whole run in 2009 was at a beam energy of 450 GeV, the beams could not collide head-on but had a crossing angle towards each other. Furthermore they could not be focused to the nominal size. Thus the spread of the beams was increased and could potentially damage the Velo. The opening of the Velo lead to a significantly reduced acceptance of Long tracks and the majority of tracks were only reconstructed as Downstream tracks, cf. Figure 3.1.

3.1 Description of the algorithm

As mentioned before, the task of the pattern recognition is to find the hits that belong to a given particle. Thus algorithms have to be developed which recognise these patterns. The algorithms are contrived and studied on simulated data.

3.1.1 Optical model method: Ideal case

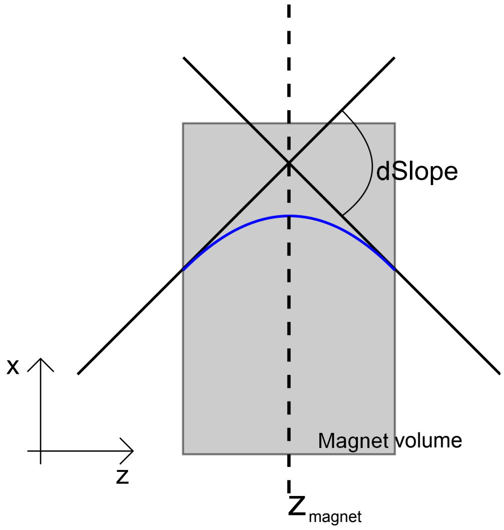


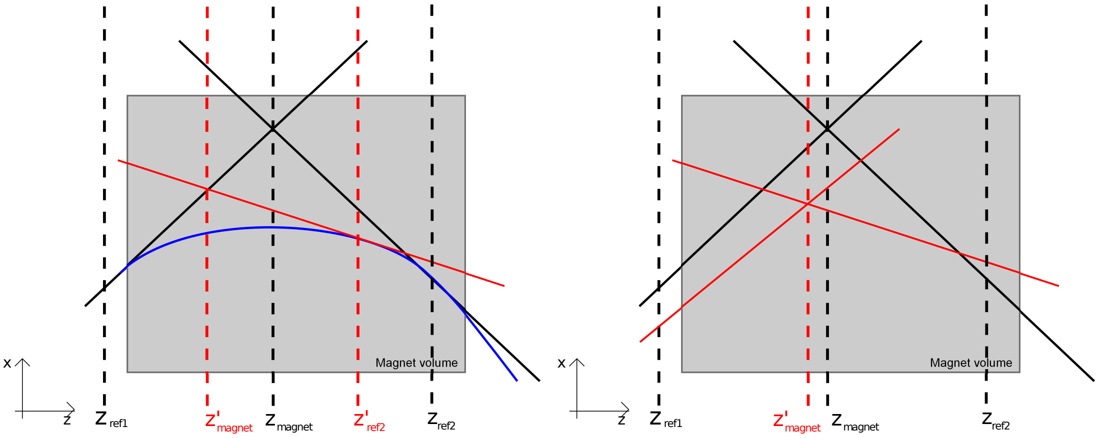
Figure 3.2: *Principle of the optical model.*

The Downstream tracking exploits that the tracking detectors of LHCb are in front of and behind the magnet. In an ideal case there is no multiple scattering and no energy loss. In this scenario the trajectory of a charged particle in a magnetic field is fully described if one phase space state and the field map is known. Then the particle can be propagated through the detector just by applying electrodynamics. The LHCb detector geometry suggests a parametrisation as a function of z. Five parameters are needed to describe the trajectory, those are x, y, $tx = dx/dz$, $ty = dy/dz$ and q/p at a given point z with q the measured charge and p the momentum. The LHCb dipole magnet has mainly a B_y component and thus it is bending tracks in the $x-z$ plane. In an ideal case where the tracking system is located outside the B-field the magnet can be considered as an optical lens which is only bending in the $x-z$ plane in the centre of the magnet. So neglecting fringe field effects one measures two straight lines, one before the magnet and one behind the magnet. The two lines match in the middle of the magnet. This is depicted in Figure 3.2. To reconstruct the track one has to find patterns in the TT and the T-Stations which correspond to this model. The momentum can be measured by determining the curvature of the trajectory in the B-field [26], exploiting the following relation:

$$p \propto \frac{4\,\mathrm{Tm}}{dSlope} \text{ with } dSlope = tx_{in\ front\ magnet} - tx_{behind\ magnet} \tag{3.1}$$

$dSlope$ is defined in Figure 3.2 and 4 Tm is roughly the integrated field for particles traversing the whole tracking system

3.1.2 Optical model method: Real case



(a) z_{magnet} is shifted due to a different choice of the second reference plane.

(b) z_{magnet} is shifted due to different track parameters.

Figure 3.3: *Principle of the optical model in the case where the tracking system is inside the fringe field of the magnet.*

In Chapter 3.1 the ideal case was discussed. In reality the particles are scattered by the material in the detector, the magnetic field has components in the y and z direction and the field extends into the tracking system. Thus the the optical model method has to be adapted. Two reference planes are defined, one in front of the magnet centre (z_{ref1}) and one behind (z_{ref2}), cf. Figure 3.3. In these planes the tangents to the tracks are obtained. One corresponds to the measurement of the track in the TT and the other to the measurement in the T-Stations. In the ideal case it does not matter where the tangents to the two tracks are taken, as long as the planes are outside the magnetic field. If the reference planes are inside the fringe field the intersection point depends on the position of the reference planes. Figure 3.3(a) is illustrating this. The z coordinate, z_{magnet}, of the point where these two lines intersect is defined as the magnet centre. The red lines are showing the effect of a different choice of the second reference plane. Additionally the z coordinate of the intersection point depends on the slopes of the track because the measurement is done while the particles are still being deflected, cf. Figure 3.3(b). Thus the centre of the magnet is not a fixed plane anymore as it was in the ideal case, but depends also on the choice of the reference planes and the parameters of the track.

The starting point of the algorithm will be tracks from the T-Stations. Thus the aim is now to find a parametrisation of z_{magnet} depending on the parameters of the track behind the magnet. In the simulated data the intersection point can be determined and thus the dependency on the track parameters can be predicted. In order to do this the associated hits in the TT and the T-Stations are fitted to two tracks and the intersection point - depending on the reference planes and the track parameters - is calculated. The exact fitting procedure is described in Chapter 4.1. The spread of the intersection point is shown in Figure 3.4 for the $x - z$ and $y - z$ plane. In the $y - z$ plane it has a huge spread because the tracks in front of and behind the magnet have almost the same slope in y direction, so a small kick leads to big shift of the intersection point. Therefore it is dominated by multiple scattering and the small components of the fringe field in x direction. In contrast to this, the spread in the $x - z$ plane is only in the order of 10 cm because of the large B-field component in y direction. Thus only the projection onto the $x - z$ plane is taken into account to derive the parametrisation of z_{magnet}.

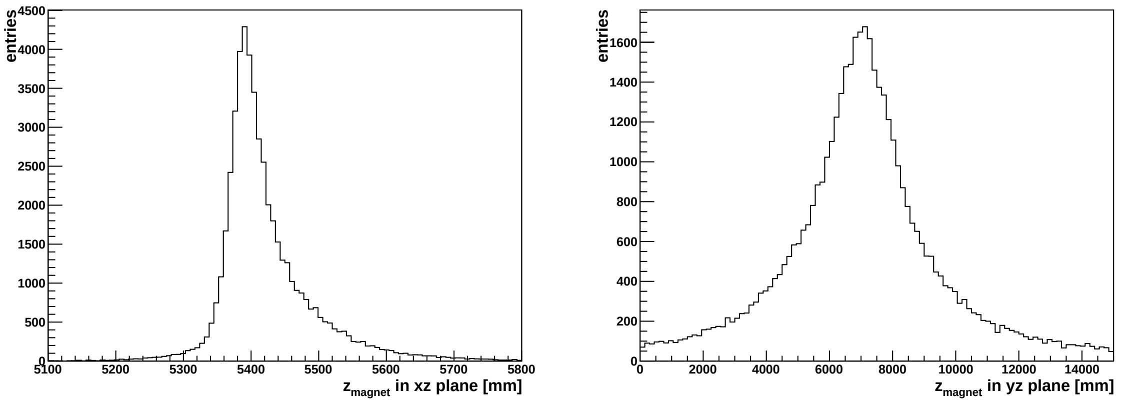


Figure 3.4: z_{magnet} *in the* $x - z$ *and* $y - z$ *plane. Note the different scales in the two histograms.*

In studies on simulated data a parametrisation of z_{magnet} from the parameters of the T-Track has been derived. The figures that display the dependencies are shown together with a more detailed description in Chapter 4.1. The parametrisation for z_{magnet} as a function of the parameters of the seed track (x_0, tx, y_0, ty) is given as

$$z_{magnet} = \alpha + \beta\ ty^2 + \gamma\ tx^2\ [\text{mm}] \tag{3.2}$$

The main idea of the Downstream algorithm is to start with a track in the T-Stations, the seed. Then the track is extended into the magnet to a coordinate $\vec{x}_{magnet}$ with $z = z_{magnet}(ty, tx)$.

$$\vec{x}_{magnet} = \begin{pmatrix} z_{magnet} \\ x_{magnet} \\ y_{magnet} \end{pmatrix} = \begin{pmatrix} \alpha + \beta\ ty^2 + \gamma\ tx^2\ [\text{mm}] \\ x_0 + (z - z_{magnet})\ tx\ [\text{mm}] \\ y_0 + (z - z_{magnet})\ ty\ [\text{mm}] \end{pmatrix}. \tag{3.3}$$

Afterwards hits in the TT which make a straight line with this estimated intersection point are searched. To reduce combinatorics the additional assumption that the tracks are coming close from the interaction point is made. A step by step description of the Downstream algorithm is given in the following section.

3.1.3 Description of the algorithm

In the following the Downstream algorithm is described as it was before this work [27]. Changes to the algorithm will be discussed later. The input are tracks from the T-Stations and hits in the TT. The tracks from the T-Stations are fully fitted to have a good parameter estimate. They will be referred as T-Seeds because one always starts with a track in the T-Stations.
The following steps are done for every T-Seed.

Step 1: First the z position of the magnet centre is computed by Formula 3.2. Then a straight line extrapolation of the state in the T-Stations to the magnet centre is made and thus the point $\vec{x}_{magnet}$ in the magnet plane is obtained, cf. Formula 3.3. Now the hypothesis is made that the particle is coming from the nominal interaction point $(0, 0, 0)$. The validity of this hypothesis is explained in Step 2. Thus one gets a first straight line track estimate from the point in the magnet $\vec{x}_{magnet}$ to $(0, 0, 0)$.

Step 2: Afterwards, the algorithm collects all interesting measurements in the TT around this track estimate. Interesting means that although the long lived particles decay far away from the interaction point, the decay products still point to the interaction point. This knowledge is used to define the size of a search window in the TT around the track estimate. A rough estimate for the size of this area can be made by the following equations which are depicted in Figure 3.5.

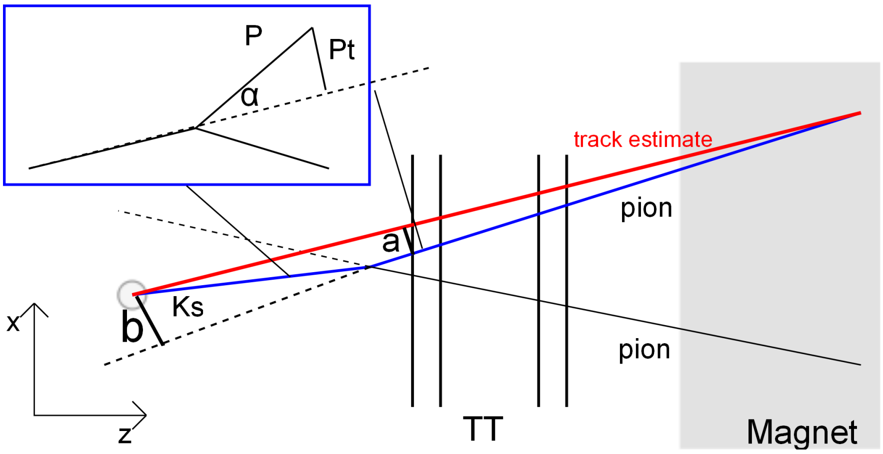


Figure 3.5: *Collection of interesting measurements in Step 2. Even if the K_s is coming from a B decay it is pointing to the interaction point.*

For example one takes a pion ($m_{\pi^\pm} \approx 139\,\mathrm{MeV}$) with a momentum of 2 GeV from a K_s ($m_{K_s} \approx 497\,\mathrm{MeV}$) originated close to the interaction point[1]. 2 GeV is the minimal momentum of a pion reconstructed in the T-Stations. Charged particles with a lower momentum are bent out of the detector by the magnetic field. The pion has a maximum transverse momentum of

$$p_{T,max} = m_{K_s} - 2m_{\pi^\pm} = 497\,\mathrm{MeV} - 2 \cdot 139\,\mathrm{MeV} = 209\,\mathrm{MeV} \tag{3.4}$$

in the K_s rest frame. This has to be taken into account to estimate the size of the search window. The opening angle between the pion trajectory and the line of flight of the K_s is about

$$\alpha = \frac{p_{T,max}}{p} = 209\,\mathrm{MeV}/2\,\mathrm{GeV} = 0.1\ rad \tag{3.5}$$

The first layer in the TT is at $z_{TT1} = 2.3\ m$, thus the impact parameter at $z = 0$ is at most

$$b = \alpha \cdot z_{TT1} = 0.1 \cdot 2.3\ m = 0.23\ m \tag{3.6}$$

The magnet centre is roughly at $z_{magnet} = 5.3\ m$. This means that the window size in TT should be around

$$a = \frac{b}{z_{magnet}}(z_{magnet} - z_{TT1}) = 0.23/5.3 \cdot 3.0\ m = 13\ cm \tag{3.7}$$

As seen in Formula 3.5, the window size depends inversely proportional on the momentum. The parametrisation of the window size in the algorithm reflects that. The parameters are tuned to a high efficiency and a low ghost rate [27]. Thus the window size is slightly smaller than the rough estimate given above::

$$Window\ size = \frac{150\,\mathrm{mm\,GeV}}{|p[\,\mathrm{GeV}]|} + \ 10.0\,\mathrm{mm} \tag{3.8}$$

---

[1] The average flight distance of B mesons is less than 1 cm.

For a 2 GeV pion this results in a tolerance of 8.5 cm. The momentum is here calculated by

$$p = \frac{a + b\ tx^2 + c\ ty^2}{dSlope}\ [\mathrm{MeV}]. \tag{3.9}$$

assuming that the track is originating from the interaction point. The parametrisation for this is also derived from studies on simulated data and corresponds to Formula 3.1.

The strip pitch in TT is 183 μm [4] so the size of the search window corresponds to roughly 1000 strips. Depending on the center-of-mass energy the overall occupancy in the TT is between 0.1 % and 1.0 %. In the innermost regions the occupancy goes up to about 3.5 % [4]. Thus the preselection will include up to 100 hits in total. Therefore several track candidates for each T-Seed will be contained in the preselection as each candidate has about four hits in the TT.

Figure 3.6 summarizes the first two steps.

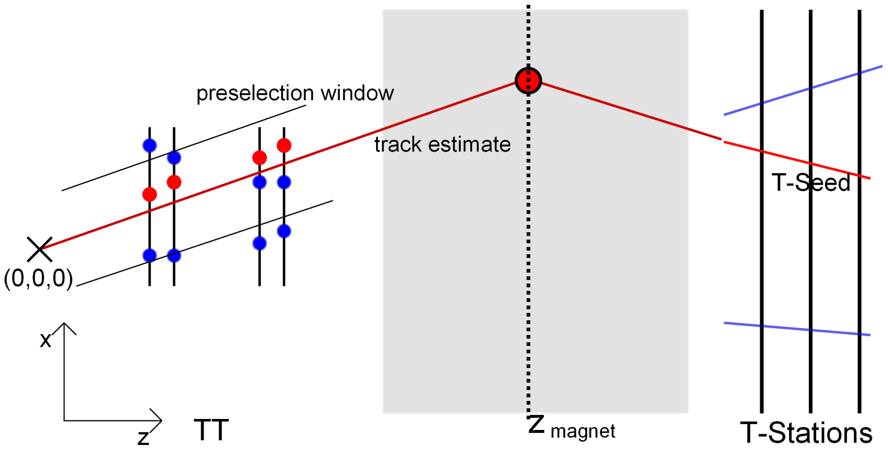


Figure 3.6: *Step 1 and 2 of the Downstream algorithm are depicted. The red hits are associated to the Monte Carlo particle associated to the red T-Seed.*

The remaining steps are done for every x hit in the preselection from Step 2.

Step 3: A new estimate is now built with a x hit and the point in the magnet $\vec{x}_{magnet}$ (Formula 3.2). Thus the assumption that the track is coming from $(0, 0, 0)$ is dropped, cf. Figure 3.7. Then all measurements in all x layers in a small window around the track are collected. The window size differs from layer to layer. In the same layer one only wants to collect other hits if the particle passed through an overlap region, thus the extrapolation distance is very small and the search window size accordingly. The search window in another layer is given by a constant offset and a term proportional to the slope in y direction:

$$Window\ size = \begin{cases} 0.1\,\mathrm{mm} & \text{same layer} \\ 5.0\,\mathrm{mm} \cdot ty + 2.5\,\mathrm{mm} & \text{different layer} \end{cases} \tag{3.10}$$

After the hits are picked up in all layers, the track is fitted in the $x - z$ plane. Then outliers are removed until the distance of every hit to the track is smaller than $0.1\,\mathrm{mm} = 100\,\mu\mathrm{m}$. This corresponds to the strip pitch in TT which is $183\,\mu\mathrm{m}$ [4] and therefore a hit resolution of about $50\,\mu\mathrm{m}$. The fit and the outlier removal procedure is described in Appendix A.2.

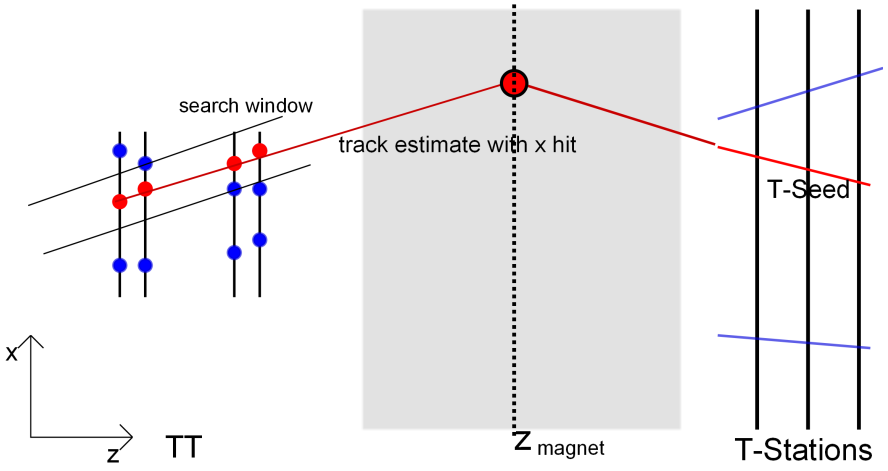


Figure 3.7: *Step 3 and 4 of the Downstream algorithm are depicted.*

Step 4: In this step also stereo measurements are taken into account. A measurement in the TT is not a single point but a line because the TT consists of silicon strips. With a given y position which is defined by the track parameters the x position of the stereo hit can be computed. One assumes that in first order the magnetic field does not change the trajectory in the $y - z$ plane. Thus only the distance in x is taken into account and all stereo measurements within 3.0 mm around the track are added to the track.

Again an outlier removal procedure is applied until all hits have a distance to the track which is smaller than 0.1 mm (See Appendix A.3 for further details.).

In the described procedure, several track candidates for every T-Seed are found, potentially one for every x hit in the preselection from Step 2.

Step 5: A new track candidate is accepted and replaces the previous one when it has:

- at least three measurements;
- two measurements with High Threshold bit[2];

---

[2] This is to remove combinations made with mainly/only spill-over measurements, cf. [27]. Spill-over means that two subsequent events overlay each other.

- a χ^2/ndf smaller than 10. Details of the fit can be found in Appendix A;
- The momentum computed with the parametrisation in Formula 3.9 should be compatible with the input from the T-Seed[3].

If there was already a selected candidate. The following criteria are applied sequently to select the better one:

- number of measurements in the TT
- smaller χ^2/ndf

If a candidate (cf. Figure 3.8) is accepted, its hits are marked as used and cannot be used again to build another track candidate for the same T-Seed.
Step 3 and 4 are repeated for every x hit in the preselection from Step 2.

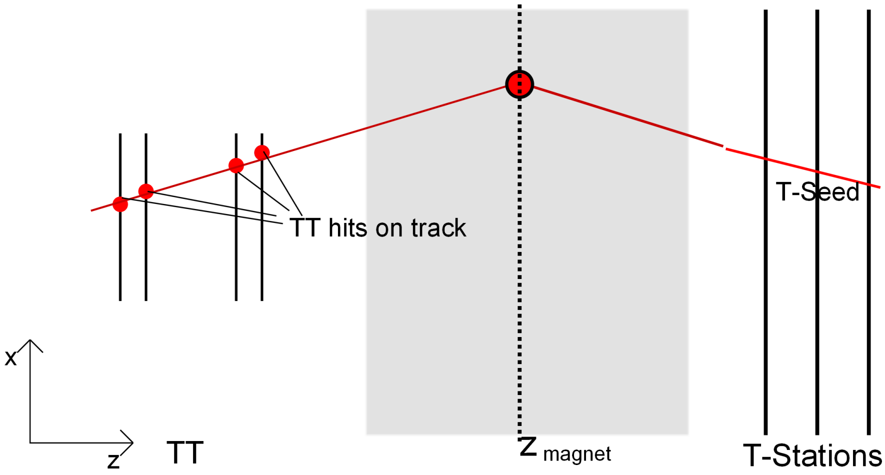


Figure 3.8: *Final track candidate of the Downstream algorithm.*

After this the algorithm goes to the next T-Seed. The before used hits in the TT are not marked as used for the next T-Seed. Thus the algorithm does not depend on the order in which the T-Seeds are used.

3.2 Performance of the original algorithm

Definition of Efficiency

As mentioned in Chapter 2.3, it is necessary to define criteria which measure the performance of an algorithm. The before mentioned criteria are shortly repeated

---

[3] The T-Seeds are Kalman fitted, thus they have a momentum estimate due to the residual field in the T-Stations.

and then modified to reflect the main goal in this chapter.
The efficiency is given by

$$\epsilon = \frac{\text{number of } \textbf{reconstructed} \wedge \textbf{reconstructible} \text{ particles}}{\text{number of } \textbf{reconstructible} \text{ particles}}. \tag{3.11}$$

The standard definition [25] at LHCb of **reconstructibility** for a Downstream track is:

- In the TT there is at least one cluster in the first two planes (TTa) and one cluster in the last two planes (TTb).
- In the T-Stations there is at least one x and one stereo cluster in each of the three T-Stations.

A Monte Carlo particle is matched to a track if

- at least 70% of the hits on this track are from this Monte Carlo particle.
- For a Downstream track there is the additional requirement of not more than one TT hit from another Monte Carlo particle on this track.

The main goal of this work was to improve the efficiency of reconstructing long lived particles which do not leave hits in the Velo. Therefore the definition of **reconstructibilty** above is adapted to this. Tracks that are already reconstructed as a Long track are not of interest. A Long track with the same hit content in the T-Stations is always preferred over the Downstream track in the analysis due to its better vertex and momentum resolution. Furthermore, the Downstream algorithm takes T-Seeds as input. Hence, if there is no T-Seed for the particle, there is no chance of reconstructing it as a Downstream track. Thus the additional requirements for the Monte Carlo particle are:

- There is no matched Long track.
- An associated T-Seed is reconstructed.

In the following only pions from K_s decays and separately pions from K_s from B decays are taken into account. The measurements do not include Λ decay products. However the efficiency is similar to K_s decay products.

Performance

Efficiencies are measured on a $B^0 \rightarrow J/\psi\ K_s$ sample[4] at $\sqrt{s} = 10\,\text{TeV}$ (10,000 events). This means that in every event at least one $B^0 \rightarrow J/\psi\ K_s$ decay is generated in the detector acceptance. Only L0 triggered events (8,470 events) are studied.

---

[4] From the MC09 Monte Carlo production.

The performance[5] of the original Downstream algorithm is shown in Table 3.1. The statistical errors are quoted. In the next chapter the impact of changes in the algorithm will be compared. Because the measurements are done on the same sample, they are highly correlated. Therefore the statistical error is only quoted here and will be dropped in the next chapter.

	Efficiency [%]
Pions from K_s	75.8 ± 0.4
$p > 5\,\mathrm{GeV}$	86.0 ± 0.4
Pions from K_s from B	82.3 ± 0.4
$p > 5\,\mathrm{GeV}$	89.9 ± 0.4
Ghost rate	31.6 %

Table 3.1: *Efficiencies of the original algorithm.*

---

[5] Brunel version v35r2 was used for this study.

Chapter 4

Improvements to the Downstream algorithm

In the previous chapter the algorithm was described as it was before this work [27]. In this chapter various modifications which improve the track reconstruction efficiency of the algorithm are discussed.

4.1 Retuning for the 2009 Monte Carlo production

The parametrisation to derive the z position of the magnet plane given in Formula 3.2 was with an old Monte Carlo simulation[1]. Since there are slight changes in the field map and a different center-of-mass energy (14 TeV resp. 10 TeV) in the 2009 Monte Carlo simulation[2], the parameters had to be retuned. The tuning was done on a minimum bias sample of roughly 170,000 events. Minimum bias samples simulate proton-proton collisions without forcing that there is a certain decay in it. In this sample there were about 80,000 K_s whose daughter tracks had enough hits in the TT and the T-Stations to be considered as reconstructible. Thus about 160,000 pions were used to extract the parameters in Formula 3.2.

The true z_{magnet} position is obtained by fitting the hits of a Monte Carlo particle with the following procedure. The bending of the track is more significant in the T-Stations because the B-Field is larger and the distance between the different stations is bigger. Thus the associated hits in the T-Stations are fitted with a cubic polynomial to build a track behind the magnet and the associated hits in the TT are fitted with a straight line to build a track in front of the magnet. Then the tangents in the TT ($z_{ref1} = 2469\,\mathrm{mm}$) and in the T-Stations ($z_{ref2} = 9410\,\mathrm{mm}$) are taken. The intersection of this two lines in the $x - z$ plane defines the true z_{magnet}, cf. Figure 3.3. Only the projection onto the $x - z$ plane

---

[1] This production is referred as DC06.

[2] This production is referred as MC09.

is taken into account because the intersection point in the $y - z$ plane has not much predictive power, as it was shown in Figure 3.4. The next step is to derive a parametrisation from the state parameters of the T-Seed for z_{magnet}. It will be called z_{pred} in this section.
In Figure 4.1 z_{magnet} is plotted against the slopes tx and ty in the T-Stations. A clear quadratical dependency in both parameters can be seen. This motivates the following ansatz

$$z_{pred} = \alpha + \beta\ ty^2 + \gamma\ tx^2. \tag{4.1}$$

The result $(z_{magnet} - z_{pred})$ of the minimisation can be seen in Figure 4.2 and 4.3.

In Figure 4.4 the pull $(z_{magnet} - z_{pred})$ is plotted against the momentum of the particle. To make the prediction of z_{magnet} better, an additional term was added. It corrects the small $1/p$ dependence which is seen in Figure 4.4. Thus z_{pred} was changed to:

$$z_{pred} = \alpha + \beta\ ty^2 + \gamma\ tx^2 + \delta/p. \tag{4.2}$$

The result of the minimisation is shown in Figure 4.5 and 4.6. The new parametrisation decreases the residual by about 10 %.

In the Downstream algorithm the state parameters (tx, ty, p) of the T-Seed are used to compute z_{pred}. The T-Stations are inside the fringe field of the magnet, thus the T-Seeds have a momentum estimate from the track fit. The parametrisation which is used in the algorithm and derived with the method above is

$$z_{pred} = 5376.8 - 3895.12\ ty^2_{\text{T-Seed}} + 309{,}877\ tx^2_{\text{T-Seed}} + 89{,}901/p_{\text{T-Seed}}. \tag{4.3}$$

With the new parametrisation there is an increase in efficiency[3] of about 1.5 % and the ghost rate is lowered by 1.6 % (Table 4.1).

Efficiency of	old tuning	new tuning	$+\delta/p$
Pions from K_s	75.8 %	77.0 %	77.3 %
$p > 5\,\text{GeV}$	86.0 %	86.5 %	86.5 %
Pions from K_s from B	82.3 %	83.3 %	83.6 %
$p > 5\,\text{GeV}$	89.9 %	90.6 %	90.4 %
Ghost rate	31.6 %	30.4 %	30.0 %

Table 4.1: *Efficiencies after retuning for the Monte Carlo production of 2009.*

---

[3] As mentioned in Chapter 3.2 no error will be given in this chapter. The measurements are highly correlated to the previous ones because they are derived on the same sample. Thus the statistical uncertainty should not be taken into account when comparing changes in the algorithm. Efficiencies are measured on a $B^0 \to J/\psi\ K_s$ sample.

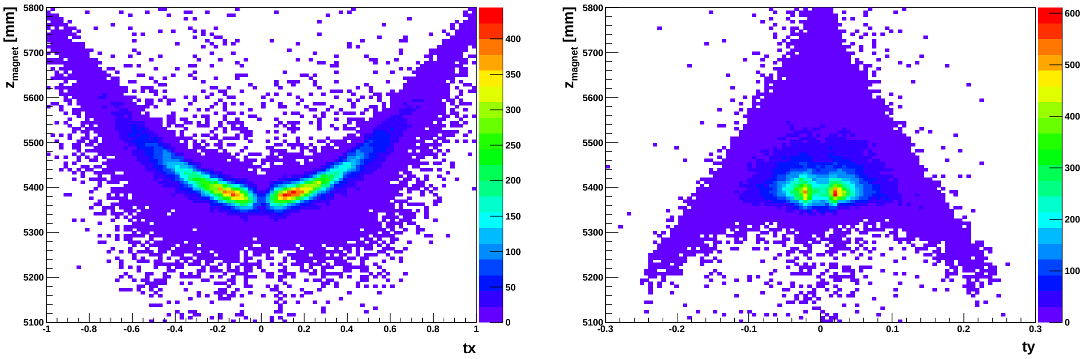


Figure 4.1: *z_{magnet} plotted against tx and ty.*

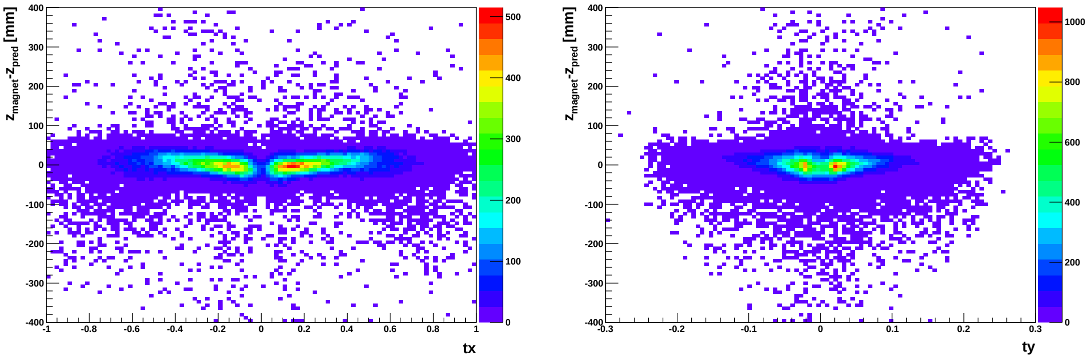


Figure 4.2: *$z_{magnet} - z_{pred}$ after correcting for tx and ty.*

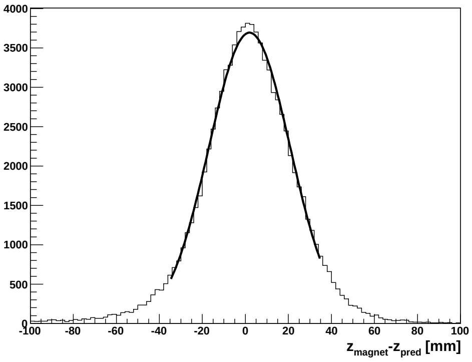


Figure 4.3: *Residual $z_{magnet} - z_{pred}$. A Gaussian with $\sigma = 18.9\,\mathrm{mm}$ and $\mu = 1.94\,\mathrm{mm}$ was fitted between $\pm 35\,\mathrm{mm}$.*

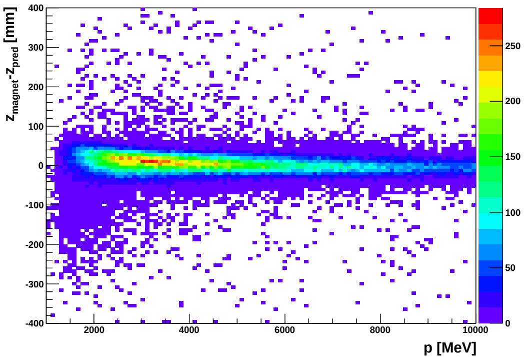


Figure 4.4: *$z_{magnet} - z_{pred}$ plotted against the momentum p.*

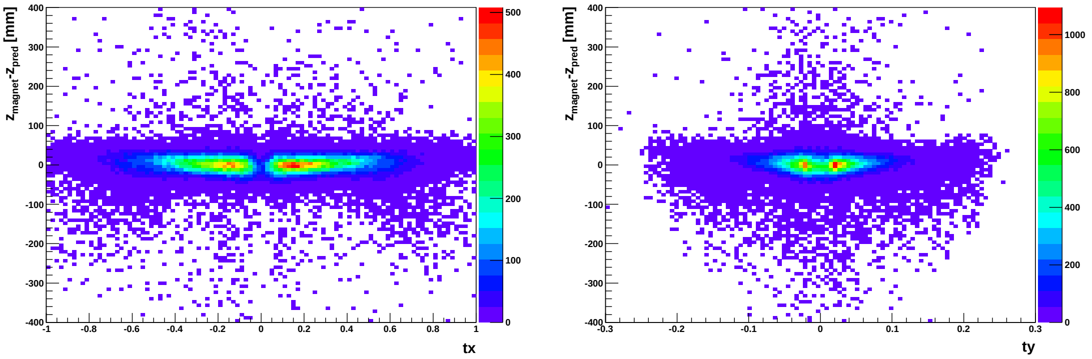


Figure 4.5: *$z_{magnet} - z_{pred}$ after correcting for tx, ty and p. The inner core is flatter with the additional momentum dependent term.*

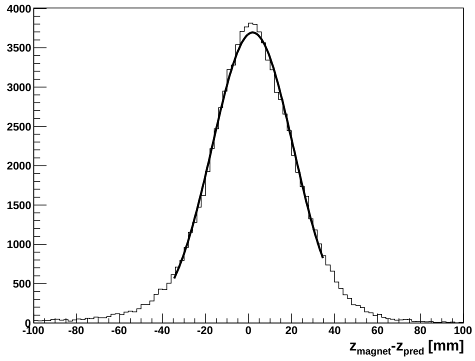


Figure 4.6: *After correcting for tx, ty and p. Residual $z_{magnet} - z_{pred}$. A Gaussian with $\sigma = 17.3\,\mathrm{mm}$ and $\mu = 2.12\,\mathrm{mm}$ was fitted between $\pm 35\,\mathrm{mm}$.*

4.2 New description for the search windows

There are several points were good track candidates are lost during the pattern recognition. One problem is that the true hits are outside the search windows in the different steps. Especially the searches in Step 2 and 3 are sensitive to the window size. If the tolerance is too small, a lot of true hits are missed. On the contrary, if the windows are too large, a lot of hits on the track have to be fitted and the ghost rate potentially increases.

A better description of the window size was investigated. For this the distance from a true hit to the track was plotted against different variables[4]. It turned out that the best separation power has the momentum estimate from the T-Seed as it is shown in Figure 4.7 and 4.9 for associated hits and in Figure 4.8 for non-associated hits. In comparison the original window description is shown. The effect of multiple scattering is taken into account by the momentum dependence of the new parametrisation. Low momentum particles are scattered more than high momentum particles. Therefore the window is wider for low momentum tracks and smaller for high momentum tracks. The hits above the black curves in the Figures 4.7, 4.8 and 4.9 are outside of the search windows. With the new descriptions less good hits lie above the lines. The window size was changed for the search of x hits and for the search of stereo hits in Step 3 and 4.

$$Window\ size = \begin{cases} 20,000\,\mathrm{mm\,MeV}/p_{\text{T-Seed}} \;+\; 0.5\,\mathrm{mm} & \text{for } x \text{ hits} \\ 20,000\,\mathrm{mm\,MeV}/p_{\text{T-Seed}} \;+\; 1.0\,\mathrm{mm} & \text{for stereo hits} \end{cases} \tag{4.4}$$

Additionally a *maximum window size* of 6 mm for both search windows is introduced. This does not affect the efficiency because hits that are further away than this tolerance are rejected in the outlier removal later in the algorithm anyway, but it leads to a reduction of processed hits and thus less CPU time.

The new description results in a higher efficiency of 1.7 % for low momentum tracks, cf.Table 4.2, and a 0.7 % higher efficiency for high momentum tracks. The ghost rate increases by 0.4 %. However, this is tolerable because the efficiency increase is more important.

Efficiency of	old cuts	new cuts
Pions from K_s	77.3 %	79.0 %
$p > 5\,\mathrm{GeV}$	86.5 %	87.2 %
Pions from K_s from B	83.6 %	85.1 %
$p > 5\,\mathrm{GeV}$	90.4 %	91.1 %
Ghost rate	30.0 %	30.4 %

Table 4.2: *Efficiencies with optimised search windows.*

---

[4] The $B^0 \to J/\psi\ K_s$ sample was used in this and the following sections.

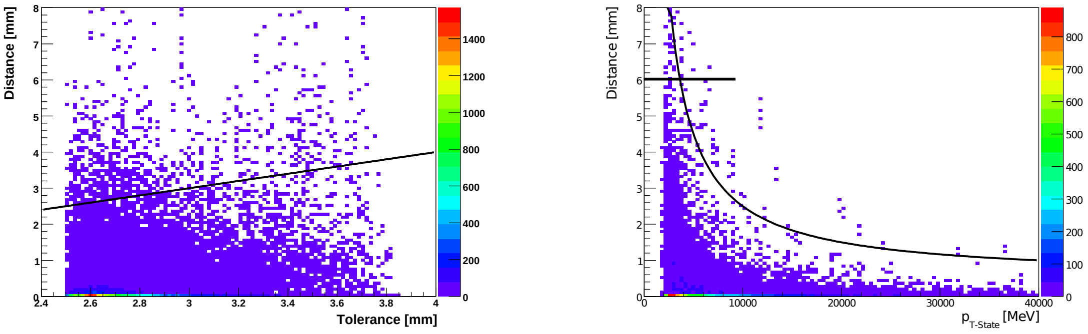


Figure 4.7: *Search for x hits. Distance from an associated hit to the track. The black line shows the parametrisation of the window size. The horizontal line is the maximum window size. Hits above the lines are not taken into account.*

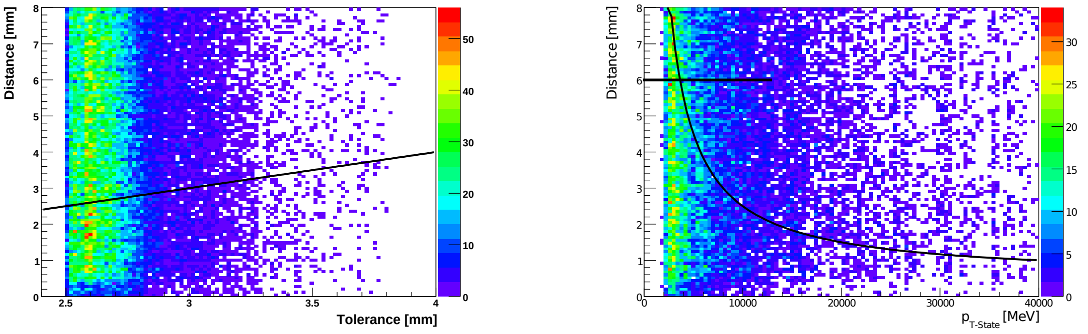


Figure 4.8: *Search for x hits. Distance from a non-associated hit to the track. The black line shows the parametrisation of the window size. The horizontal line is the maximum window size. Hits above the lines are not taken into account.*

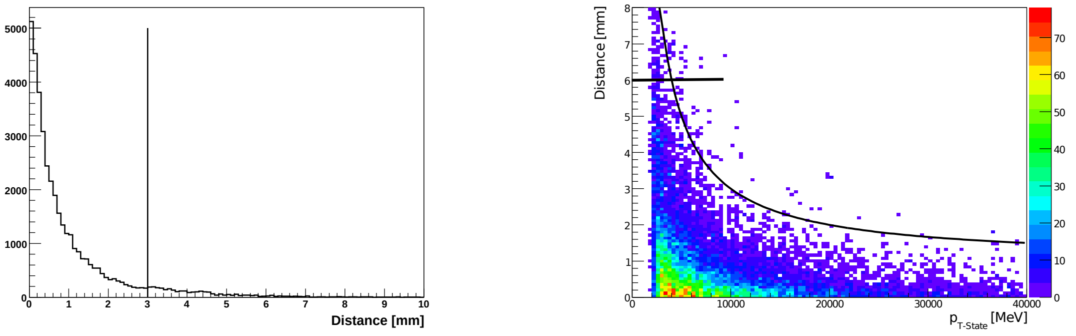


Figure 4.9: *Search for stereo measurements. Distance from an associated hit to the track. The black line shows the parametrisation of the window size. The horizontal line is the maximum window size. Hits above the lines are not taken into account.*

4.3 Change in the fit of the x projection

The next thing to check was if the fit of the track could be improved. The first fit is performed in the $x - z$ plane. Especially for low momentum tracks the number of hits in the search window is sometimes 5 or more, cf. Figure 4.10. Only two hits, one in each station, are expected in average. The point in the magnet derived from the T-Seed is only one additional constraint in the fitting procedure. If the number of hits is too high, it could happen that too many hits from other particles pull the fit into the wrong direction. Then an actual good hit has the greatest distance to the track estimate and is removed from the track in the process of the outlier removal. The idea was now to perform a fit with only two hits. Starting in Step 3 with one x hit as seed, all combinations with hits in the other x layer are built. Then all possibilities are fitted and the one with the smallest χ^2/ndf is taken to proceed with the pattern recognition. This procedure ensures that the above mentioned problem can not happen.

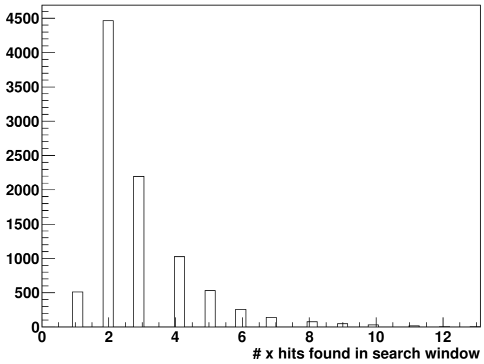


Figure 4.10: *Number of x hits found in the search window.*

With this way of fitting one has to care additionally about particles which have hits in overlap regions of the TT because only one x hit of each station enters into the fitting procedure. These hits are added after the best candidate is found by searching for nearby hits.

By applying the new fit method, the efficiencies get better by about 0.5 - 1.0 % and the ghost rate is lowered by 1.2 %, cf. Table 4.3.
Applying the same strategy to the space track fit after adding stereo measurements did not show any improvements.

4.4 Summary

Summing up all improvements presented in this chapter the track reconstruction efficiency of the Downstream algorithm was improved by relative 6.0 %, while the ghost rate was lowered by relative 10 %, cf. Table 4.4. The reconstruction

Efficiency of	new cuts	+ changed x fit
Pions from K_s	79.0 %	80.3 %
$p > 5\,\mathrm{GeV}$	87.2 %	87.8 %
Pions from K_s from B	85.1 %	85.8 %
$p > 5\,\mathrm{GeV}$	91.1 %	91.4 %
Ghost rate	30.4 %	29.2 %

Table 4.3: *Efficiencies with changed fit procedure.*

efficiency for tracks originating in B^0 decays was improved by relative about 4.0 % from 82.3 % to 85.8 % and for high momentum tracks the efficiency is above 90 %. The changes are integrated into the LHCb software environment.

Efficiency of	Original	Changed Algorithm
Pions from K_s	75.8 %	80.3 %
$p > 5\,\mathrm{GeV}$	86.0 %	87.8 %
Pions from K_s from B	82.3 %	85.8 %
$p > 5\,\mathrm{GeV}$	89.9 %	91.4 %
Ghost rate	31.6 %	29.2 %

Table 4.4: *Overall improvement of the Downstream algorithm.*

4.5 CPU performance

With increasing luminosity a vast number of events has to be processed for physics analyses, thus it is worth to study if the CPU time of the track reconstruction can be decreased.

4.5.1 Potential improvement in CPU time by restricting the input sample

The Downstream algorithm takes as input all hits in the TT and all T-Seeds. But many of them are already used to build Long tracks. Thus it is not necessary to reconstruct them again as Downstream tracks. In the end a Long track is always favoured in physics analyses because of its better momentum and vertex resolution. Hence an option was tested which allows to discard TT hits and T-Seeds which have already been used in the Long track reconstruction. If the Downstream algorithm is run on this restricted input sample, a drop in efficiency of about 4 % is seen. This is because about 17 % of the reconstructed Long tracks are ghosts tracks, cf. Chapter 2, and hits on them could belong to real Downstream tracks. The output of Downstream track candidates is reduced by factor of four which makes the following fit four times faster.

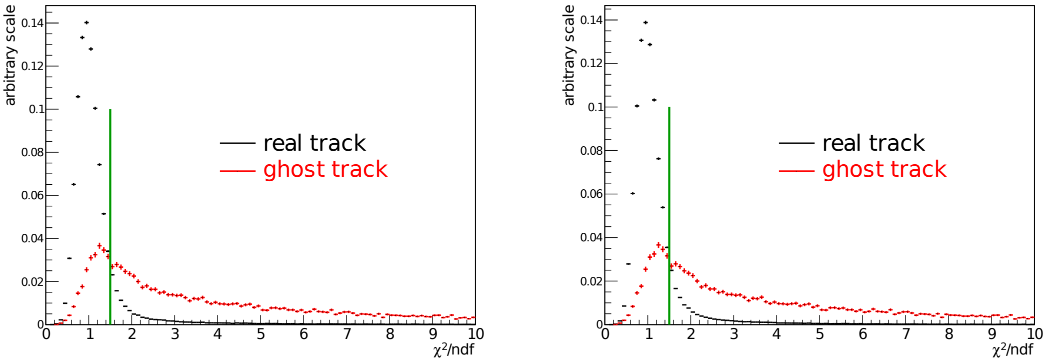


Figure 4.11: *χ^2/ndf of Long tracks. Left: Track Matching. Right: Forward tracking*

To avoid a loss in efficiency, one needs a criterion to distinguish between ghost and real Long tracks. One criterion is the χ^2/ndf of the track fit. In Figure 4.11 the separation power of this variable can be seen. The highest efficiency can be obtained with a χ^2/ndf of 1.5. Thus if only TT hits and T-Seeds which are used to build Long tracks with a $\chi^2/ndf < 1.5$ are removed, no efficiency is lost. The number of Downstream tracks is still reduced by a factor of three which results in a factor of three less CPU time spend in the fit because less tracks have to be processed. Also the number of ghosts among the tracks which are used for physics analyses is lowered by about 9.0 %, cf. Table 4.5.

Efficiency of	Changed Algorithm	+ $\chi^2/ndf < 1.5$ cut
Pions from K_s	80.3 %	81.0 %
$p > 5\,\text{GeV}$	87.8 %	88.0 %
Pions from K_s from B	85.8 %	86.4 %
$p > 5\,\text{GeV}$	91.4 %	91.5 %
Ghost rate before CK	29.2 %	51.1 %
Ghost rate after CK	63.0 %	54.0 %
# Downstream tracks before CK	36.8	13.5
# Downstream tracks after CK	11.2	9.3

Table 4.5: *CK is the abbreviation of CloneKiller. Efficiencies with only TT hits and T-Seeds as input which are not yet used to make a high quality track ($\chi^2/ndf < 1.5$).*

4.5.2 Timing measurements

In this section the impact of the changes on the runtime is studied. Applying all the improvements in efficiency to the algorithm, except the discussed restriction of the input sample, the algorithm spends about 2.3 ms (90 %) more time[5]. The changes in the runtime are summarized in Table 4.6. The increased runtime is because in the changed algorithm more hits are collected in the widened windows. Thus more candidates have to be processed. Also the changed way of fitting costs some CPU time. However, the runtime of the Downstream algorithm is not crucial because the fitting afterwards costs much more time compared to the pattern recognition, about 25 times more.

	Original	+ new search windows	+ new x fit
Downstream tracking	2.9 ms	3.9 ms	4.6 ms
Fit of Downstream tracks	125 ms	125 ms	125 ms
# Downstream tracks	37.1	37.3	37.5

Table 4.6: *CPU time consumption for different versions of the algorithm.*

In Chapter 4.5.1 a way was presented to decrease the output of tracks of the Downstream algorithm while keeping its efficiency. The impact on the CPU time is shown in Table 4.7. The pattern recognition itself is about 50 % faster because of a lower number of T-Seeds and TT-hits which have to be processed. The biggest win in CPU time is due to the by a factor of three lowered number of Downstream tracks which makes the fit of the Downstream tracks three times faster.

	New Algorithm	+ remove used
Downstream tracking	4.6 ms	2.4 ms
Fit of Downstream tracks	125 ms	40 ms
# Downstream tracks	37.5	13.7

Table 4.7: *Runtime with only TT hits and T-Seeds as input which are not yet used to make a high quality track ($\chi^2/ndf < 1.5$).*

In the beginning the CPU power is not the crucial bottleneck due to the reduced luminosity and one does not want the pattern recognition algorithms to influence each other. Thus this option is currently not used in the standard reconstruction sequence. However it will become important in the next months when the luminosity increases.

4.6 Reconstruction of K_s in minimum bias events and in a B_0 signal sample

The improved pattern recognition should also improve the reconstruction of decayed particles. If the reconstruction of pions coming from the same particle

---

[5] Timing measurements are done a certain machine. Thus only relative times are relevant. About 900 L0 triggered events were used and the uncertainty of the runtime measurement is in the order of 10 %.

is uncorrelated a 6.0 % better pattern recognition should result in 12.0 % more K_s. To study this, the reconstruction of K_s in minimum bias events and the reconstruction of $B^0 \rightarrow J/\psi\ K_s$ decays was analysed[6]. To be unbiased they differ to the ones that were used to tune the algorithm.

4.6.1 Event selection

Optimising the B^0 and K_s selections was not part of this work. Thus the selections were taken from the standard LHCb software. Only the difference in the yield with or without the new algorithm is of interest here, not the total number. The following cuts were used for this study:

- $p(\pi_1, \pi_2)\,[\mathrm{GeV}]$: Minimal momentum of daughter particles.
- χ^2/ndf of π *tracks*: Maximal χ^2/ndf of the daughter tracks.
- $\chi^2/ndf\ (K_s)$: Maximal χ^2/ndf of the vertex fit.
- $IP/\sigma(\pi_1, \pi_2)$: The impact parameter (IP) is the minimal distance between a track and the primary vertex. This is divided by the error. The error depends on the quality of the track fit and the primary vertex fit.
- $\Delta M(\pi_1 + \pi_2)\,[\mathrm{MeV}]$: The 4-momenta of the two tracks are combined to get the invariant mass of the mother particle and the difference to the PDG mass is taken. The tracks are not propagated through the B-field. Thus it is only a very loose preselection cut.
- $\Delta M(K_s)\,[\mathrm{MeV}]$: This is the difference of the invariant mass to the PDG mass after the vertex fit. The tracks have been properly propagated to the K_s vertex.

The cuts for K_s candidate selection are given in Table 4.8.

Cuts	Long track	Downstream track
$p(\pi_1, \pi_2)\,[\mathrm{GeV}]$	> 2.0	> 2.0
χ^2/ndf π *tracks*	< 20.0	< 20.0
$\chi^2/ndf\ (K_s)$	< 30.0	< 30.0
$IP/\sigma(\pi_1, \pi_2)$	> 3.0	> 2.0
$\Delta M(\pi_1 + \pi_2)\,[\mathrm{MeV}]$	< 50.0	< 80.0
$\Delta M(K_s)\,[\mathrm{MeV}]$	< 35.0	< 64.0

Table 4.8: *Cuts for K_s selection.*

---

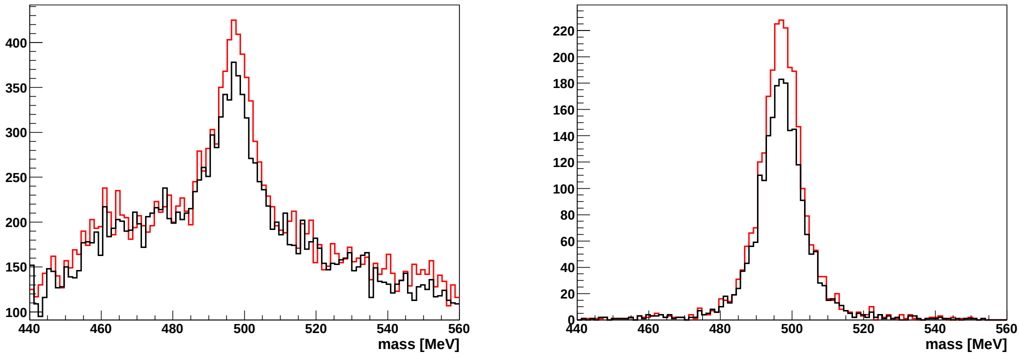
[6] The samples are taken from the MC09 production.

4.6.2 K_s in minimum bias events

The used sample consists of about 80,000 events. In this section no trigger requirements were applied. This has more technical reasons as it would take to long to reprocess a sufficient amount of minimum bias events which pass the L0 trigger. To select K_s the two above mentioned standard selections were used.

The stated increase in reconstruction efficiency for pions from K_s was 6.0 %[7]. However in the previous sections efficiencies were measured on a $B^0 \rightarrow J/\psi\ K_s$ sample. In minimum bias events K_s mesons tend to have a lower momentum because they are originated from the PV and not from a B decay. The changes of the algorithm have more impact on low momentum particles. Thus the efficiency depends on the momentum spectrum of the K_s mesons.

As one sees in Table 4.9 and Figure 4.12[8] the increase in K_s yield on a minimum bias sample is 20.0 %. Also the signal to background fraction is slightly improved. The Long track results are given for comparison.



(a) All K_s candidates (b) Truthmatched K_s candidates

Figure 4.12: *Mass of reconstructed K_s candidates which are built out of two Downstream tracks. Old algorithm (black), new algorithm (red).*

4.6.3 B_0 signal sample

The signal sample for $B^0 \rightarrow J/\psi\ K_s$ consists of 50,000 events. The selection is a standard selection taken from the LHCb software environment. $B^0 \rightarrow J/\psi\ K_s$ decays are reconstructed in two categories. One category uses two Downstream

---

[7] Note there was a bug in the original algorithm which resulted in a misidentification of x hits as stereo hits and vice versa. Due to this bug the efficiency was reduced by about 1 %. The bugfix is not included in the original sample.

[8] The background of Downstream tracks is not flat due to the combination cut ($\Delta M(\pi_1 + \pi_2)$) applied in the selection. The pions are not propagated through the B-field, just the invariant mass from the two 4-momenta in the TT is taken. The correct mass is obtained after the pions have been propagated to the K_s decay vertex. Thus the cut on the mass after the vertex fit is smeared out.

	original			new			increase
	signal	bg	S/B	signal	bg	S/B	of signal
Long tracks	3935	7827	0.50	3935	7827	0.50	0 %
Downstream tracks	2245	16980	0.13	2707	17781	0.15	20.6 %

Table 4.9: *Changes of signal and background in minimum bias events. Signal K_s candidates are matched to Monte Carlo K_s particles.*

tracks to reconstruct the K_s and the other two Long tracks. In the first category the number of reconstructed decays increases about 10 % with the new Downstream algorithm. This is the expected increase, because the efficiency increase on track level is about 5 % and one needs two tracks to reconstruct a K_s. This shows also that the reconstruction of the both pions is not correlated and the limitation in the decay is to reconstruct the K_s and not the J/ψ which is always found when the K_s has been reconstructed The J/ψ mesons have a higher average momentum and because of the short lifetime they are always reconstructed with two Long tracks. The performance of the other category stays the same thus the overall gain in signal is about 7 %, cf. Table 4.10.

	original	new	increase
	signal	signal	of signal
Long tracks	1690	1690	0 %
Downstream tracks	3645	4015	10 %
Overall	5338	5705	7 %

Table 4.10: *Changes of signal on a $B^0 \rightarrow J/\psi\ K_s$ signal sample. Signal events are matched to a Monte Carlo particle.*

Chapter 5

Tuning of the Downstream algorithm for first data

With the first collisions starting in late 2009 there was the opportunity to test the Downstream algorithm under real conditions. This is described in the following chapter. In the 2009 run the Vertex Locator was not at its nominal position but was moved away from the interaction point. Thus the acceptance was reduced and the Downstream tracking was the algorithm which provided the most reconstructed tracks useful for physics analyses, cf. Figure 3.1.

Of course there are differences in the data taken in the 2009 run and the simulated data[1] as the simulation does not describe the reality in full detail. E.g. the pattern recognition was tuned and tested on simulated samples which have every detector at its nominal position. But the mechanical precision during the installation of a detector is finite. This leads to a deviation between the nominal position and the real position. This is called misalignment. Therefore it was expected that the pattern recognition behaves differently on collision data than on simulated data and a retuning of the pattern recognition was needed to compensate the differences.

Furthermore the algorithm was developed on a sample which simulated proton-proton collisions at a center-of-mass energy of 10 TeV. On this sample about 1/3 of the Downstream tracks are ghost tracks. The first data were taken with a center-of-mass energy of 900 GeV. In Figure 5.1 it can be seen that the occupancy in the detector is much lower at lower energies. The occupancy in data is slightly higher than the simulated occupancy for the same energy which is related to the not yet optimal tuning of the Monte Carlo generator and of the detector response in the simulation. However the occupancy is still much lower in the $\sqrt{s} = 900\,\text{GeV}$ data than in the $\sqrt{s} = 10\,\text{TeV}$ Monte Carlo samples. Therefore the combinatorics is reduced and the chance to produce ghosts is smaller. This allows to open cuts in the pattern recognition while keeping the ghost rate still at a reasonable level.

---

[1] In the following chapters simulated samples will always be referred as simulated data. The data from the 2009 run are called data or collision data.

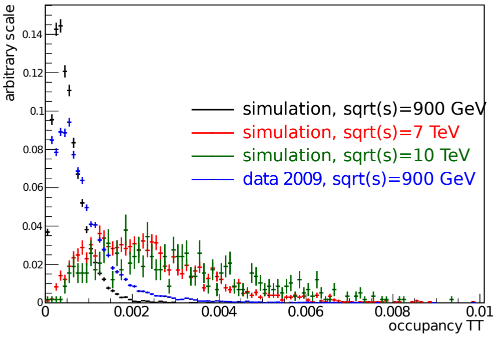


Figure 5.1: *Occupancy in the TT. The simulated samples contain minimum bias events which are triggered by the L0.*

5.1 Performance on simulated events at $\sqrt{s} = 900\,\mathrm{GeV}$

In the previous chapter the focus was to improve the efficiency of the pattern recognition in the case that the particle is not already reconstructed as a Long track. As mentioned before, the acceptance of the Velo was reduced in the 2009 run and the number of Downstream tracks was much higher than the number of Long tracks. Furthermore, an analysis of K_s in the first data is presented in the next chapter. These K_s are reconstructed with all Downstream tracks independent if they are clones of Long tracks or not. Thus the additional requirement of Chapter 3 that no Long track is already matched to the Monte Carlo particle is dropped in the study presented here. Therefore the efficiency is given by

$$\epsilon = \frac{\text{number of } \textbf{reconstructed} \wedge \textbf{reconstructible} \text{ MC particles with an ass. T-Seed}}{\text{number of } \textbf{reconstructible} \text{ MC particles with an ass. T-Seed}}. \tag{5.1}$$

The efficiency is given for all Monte Carlo particles and daughters of Monte Carlo K_s particles.

In Chapter 2, 3 and 4 the performance of the algorithm was tested on samples with a center-of-mass energy of 10 TeV. Here a minimum bias sample which simulates proton-proton collisions at $\sqrt{s} = 900\,\mathrm{GeV}$ is used. Additionally, there was also a tuning of the T-Seeding algorithm for the first data which is used here. The new tuning does not have an impact on the efficiency on simulated data but the number of T-Tracks on data is increased by 1.5 %. Table 5.1 summarizes the performance[2] of the Downstream algorithm on a $\sqrt{s} = 900\,\mathrm{GeV}$ minimum bias sample. One sees that the total efficiency is about 6 % higher as on 10 TeV. The efficiency if a T-Seed is found is 92.3 %.The hit purity is 99.8 % and the hit efficiency in the TT is 99.8 %. The ghost rate is 8.1 % which is only about a

---

[2] Brunel v36r0p1 was used.

third of the ghost rate in the $\sqrt{s} = 10\,\mathrm{TeV}$ minimum bias sample. The number of TT hits on the track is quoted twice. In one case only associated tracks enter the calculation and in the other case all tracks are used. The second number is quoted because it can be compared to collision data.

associated tracks	ϵ_{rec} all tracks	84.6 % / 92.3* %
	$p > 5\,\mathrm{GeV}$	92.8 % / 95.4* %
	ϵ_{rec} K_s daughter tracks	81.2 % / 90.4* %
	$p > 5\,\mathrm{GeV}$	92.2 % / 94.3* %
	hit purity in TT	99.8 %
	hit efficiency in TT	98.8 %
	# TT hits	4.05
all tracks	ghost rate	8.1 % / 6.8 %
	# TT hits	4.03
	# T-Seeds	99,548
	# Downstream tracks	58,193

Table 5.1: *Summary of the tracking performance on about 13,000 L0 triggered events at* $\sqrt{s} = 900\,\mathrm{GeV}$*. *A reconstructed seed is required in the efficiency denominator.*

5.2 Misalignment and pattern recognition

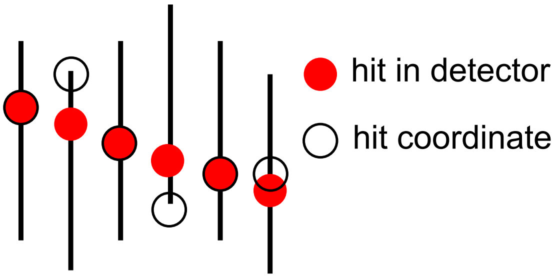


Figure 5.2: *Hit coordinates in a misaligned detector.*

One should mention that it was only possible to roughly align the whole detector before the start of the collisions due to the special forward geometry of LHCb. A multipurpose detector, like ATLAS or CMS, can be aligned with cosmic rays because the trajectory of most cosmic particles is perpendicular to the surface of the earth. But only a few cosmic rays fly parallel to the surface and such are needed to align the LHCb detector due to its special forward geometry. Furthermore the silicon trackers have a rather small diameter. For these reasons only a statistical very limited alignment of the whole detector with cosmic rays was possible.

If a charged particle traverses the detector it leaves hits in the various subdetectors. Figure 5.2 is depicting this for a particle on a straight trajectory. In reality the exact position of the individual subdectors is not known, but the pattern recognition assumes that every part of the detector is perfectly aligned. Thus the coordinate systems of reality and of the pattern recognition are shifted to one another due to the misalignment. Although the pattern recognition algorithms are designed to handle known misalignment, these shifts cause a loss in efficiency if they are larger than expected. The new tuning tries to compensate this effect and to make it more robust against misalignment.

The misalignment enters in various steps of the pattern recognition. First of all the misalignment in the T-Stations affects the quality of the T-Seeds. Thus the prediction of the point in the magnet is worse. Then there could be a global misalignment between the T-Stations, the TT and the magnet. And for Downstream tracking most important is the internal alignment of the TT.

The effect of misalignment is seen in Table 5.2. The performance of the Downstream algorithm is compared for an alignment that does only include the internal alignment of the TT[3] according to the survey when the TT was installed and an alignment that includes a first track based internal alignment of the TT[4]. The number of T-Seeds is the same for both but the number of Downstream tracks increases by 6 %. The average number of picked up hits for all tracks increases from 3.69 to 3.94. The invariant mass distributions of pion combinations are shown in Figure 5.3. The K_s signal yield[5] increases from about 2500 to about 3900 which is an increase of about 60 %, compared to only 6 % more reconstructed tracks. Thus the quality of the tracks must have been much worse before. Therefore many candidates did not pass selection cuts like the χ^2/ndf of the daughter tracks or the χ^2/ndf of the vertex fit. This shows that an alignment of the detector is very crucial. In the rest of this chapter only the geometry database including the first track based alignment of the TT will be used.

	no TT alignment	with TT alignment
# T-Tracks	153,432	153,432
# Downstream tracks	74,198	78,626
# TT hits	3.69	3.94

Table 5.2: *Number of T-Tracks and Downstream tracks in 36,000 events from the 2009 run for two different alignments.*

5.3 Tuned parameters

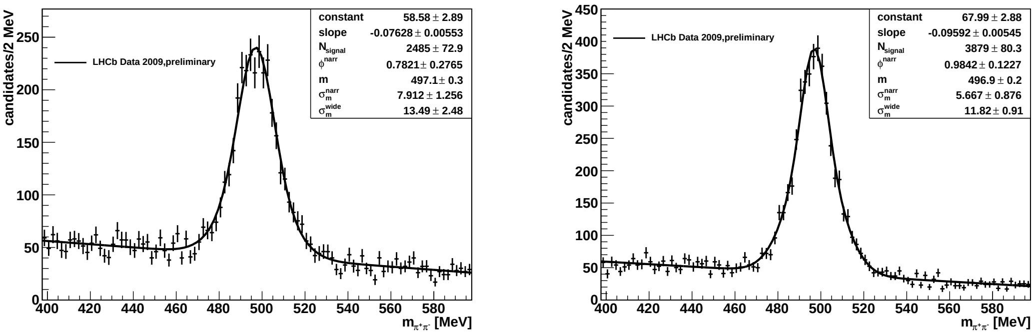
In Chapter 3.1.3 the algorithm was explained step by step. In almost every step one can change parameters that affect the pattern recognition. The parameters that were tuned are the following.

---

[3] Alignment database 1.0 of the LHCb alignment group.

[4] Alignment database 1.1 of the LHCb alignment group

[5] The selection will be discussed in Chapter 6.2.2.



(a) no TT alignment (b) with TT alignment

Figure 5.3: *Invariant mass distribution of reconstructed K_s candidates. A double Gaussian (Definition in Chapter 6.2.3) is fitted to the distributions. ϕ^{narr} is the fraction of the two widths.*

- **Preselection in Step 2:** The correctness of the preselection depends mainly on the global misalignment between the TT, the T-Stations and the magnet. Furthermore the preselection depends also on the the quality of the T-Track because the preselection starts with a point in the magnet which is calculated from the T-Track.

- **Search windows in Step 3 and 4:** In Step 3 and 4 hits are searched in relatively small windows. Thus the internal misalignment of the TT affects the search for hits.

- **Outlier distance in outlier removal procedure:** The outlier distance controls which hits are discarded in the fitting procedure of Step 3 and 4, cf. Figure 5.4.

- **Track quality(χ^2/ndf) in Step 5:** On data the hits tend to have a higher distance to the track and this will also increase the χ^2/ndf, 5.5. Additionally if the before mentioned parameters are loosened more hits with a larger distance to the track will be collected. Thus this parameter has to be changed as well in order to accept more tracks.

The search windows in Step 2, 3 and 4 have a parametrization like $\frac{constant}{momentum} + offset$. As mentioned in Chapter 4.2 the first part describes the effect of multiple scattering. This is not affected by the misalignment. Therefore only the offset is changed.

5.4 Impact on reconstructed tracks

In data there is no possibility to determine if a track is a real track or if it is a ghost track. Unless, if it is possible to combine it with another track to reconstruct

a particle. Unfortunately, this takes a lot of time because a lot of events are needed to get a sufficient number of reconstructed particles. E.g. in about 400,000 events of the 2009 run about 4000 K_s are reconstructed, cf. Chapter 6. Thus it is not feasible to test every change in the pattern recognition on the number of reconstructed particles.
Therefore criteria have to be defined on the level of reconstructed tracks:

- The first criterion is the number of found tracks. However, it is possible that additionally found tracks are ghost tracks. This can be estimated by comparing the effect of the tunings on simulation and on data. If the increase in tracks on data is higher than the increase in ghost tracks on simulated data, it is a hint that the additional tracks are good tracks.
- The second criterion is the number of TT hits on a track. This reflects how efficient the pattern recognition collects hits. This is evaluated by comparing the number of hits on simulated and on collision data.
- As a cross-check for the tuning the ratio between the number of T-Tracks and Downstream tracks is taken. By taking the ratio the efficiency of the T-Seeding is taken out. If the number of as Downstream track reconstructible particles is the same in the simulation and on data the ratio on data should be similar to the ratio in the simulation.

A data sample of 36,000 events was taken to evaluate the tracking performanc. The number of T-Tracks which enter the Downstream algorithm is about 153,000. The number of reconstructed Downstream tracks is without any tuning about 79,000. Thus the ratio is 0.51 while in the simulation it is 0.58. The simulated sample contains about 52,000 reconstructible Monte Carlo particles and depending on the tuning about 58,000 reconstructed Downstream tracks. First the cuts are tested by themselves leaving the other parameters unchanged and then the effect of their combination is evaluated.

Preselection

Table 5.3 shows how the key quantities of the pattern recognition change when the window for the preselection is opened from 1 cm to 2 cm resp. 3 cm. On the simulated data the efficiency goes up by 0.3 %. The ghost rate increases by 0.3 % resp. 0.6 % but on data the number of tracks is 2.2 % resp 3.3 % higher. This gives confidence that most of the additional tracks are non ghosts. The final tuning opens the window to 2 cm as this increase should compensate the misalignment.

Search windows

Also the smaller search windows in Step 3 and 4 were opened because the search for hits is affected by the internal misalignment of TT. Again the offset was

	preselection offset	10 mm	20 mm	30 mm
Simulation	ϵ_{rec} all tracks	92.3 %	92.5 %	92.5 %
	# TT hits	4.05	4.05	4.05
	TT hit efficiency	98.8 %	98.8 %	98.8%
	TT hit purity	98.8 %	99.8 %	99.8 %
	ghost rate	8.1 %	8.4 %	8.7 %
	# TT hits all tracks	4.03	4.03	4.02
Data	# tracks	78,626	80,350	81,450
	increase	0.0 %	2.2 %	3.5 %
	# TT hits	3.94	3.94	3.93

Table 5.3: *Evolution of the tracking performance for change of the preselection. The quoted efficiency is the efficiency for all as Downstream track reconstructible particles if a T-Seed has been found.*

increased by 1 mm resp. 3 mm (from 0.5 mm for x hits and from 1 mm for stereo hits), cf. Table 5.4. These windows were originally tuned to maximum efficiency. Thus changing them decreases the efficiency on Monte Carlo by 0.2 % resp. 0.6 %.

Changing only this parameter does not increase the number of tracks in data because the newly found hits are more far away and will probably be discarded in the outlier removal procedure. The number of tracks actually decreases because the newly found hits raise the χ^2/ndf and the track is not accepted as a good candidate. Both scenarios in combination with other loosened cuts are tested on the K_s yield in Chapter 5.5.

	search window offset	0.5 mm	1.5 mm	3.5 mm
Simulation	ϵ_{rec} all tracks	92.3 %	92.1 %	91.7 %
	# TT hits	4.05	4.06	4.06
	TT hit efficiency	98.8 %	98.8 %	98.8%
	TT hit purity	98.8 %	99.8 %	99.8 %
	ghost rate	8.1 %	8.3 %	8.3 %
	# TT hits all tracks	4.03	4.04	4.04
Data	# tracks	78,626	78,624	78,087
	increase	0.0 %	0.0 %	-0.6 %
	# TT hits	3.94	3.96	3.96

Table 5.4: *Evolution of the tracking performance for change of the search windows. The quoted efficiency is the efficiency for all particles if a T-Seed has been found.*

Outlier distance

The outlier distance determines which hits are added to a track. If it is too big, too many non-associated hits could be collected. If it is too small, associated hits are discarded. In Figure 5.4 the distance from the hits on a track to the track is shown. The distribution is shifted to the right for data compared to simulated

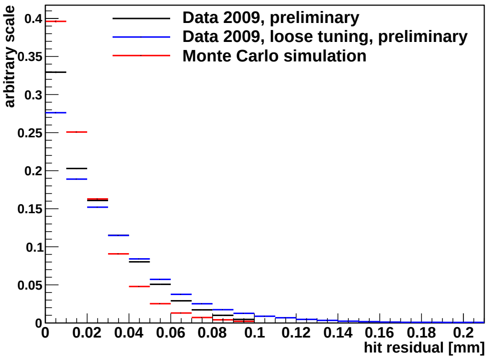


Figure 5.4: *The distance from a hit to the track is shown for Downstream tracks. Red shows simulated data, black collision data. The blue curve shows the distribution for tune1 in Table 5.7. The black and red distributions stop at 0.1* mm *because only hits with a smaller distance to the track are accepted in the original tuning.*

data. Thus the the maximum outlier distance was increased from 0.1 mm to 0.3 mm. In Table 5.5 it can be seen that the pattern recognition does not start to collect too many hits if the outlier distance is increased. The hit purity stays the same while the hit efficiency actually increases by 0.8 %. The number of TT hits on data (4.03) is close to the the standard tuning on simulated events(4.03), cf. Table 5.1.

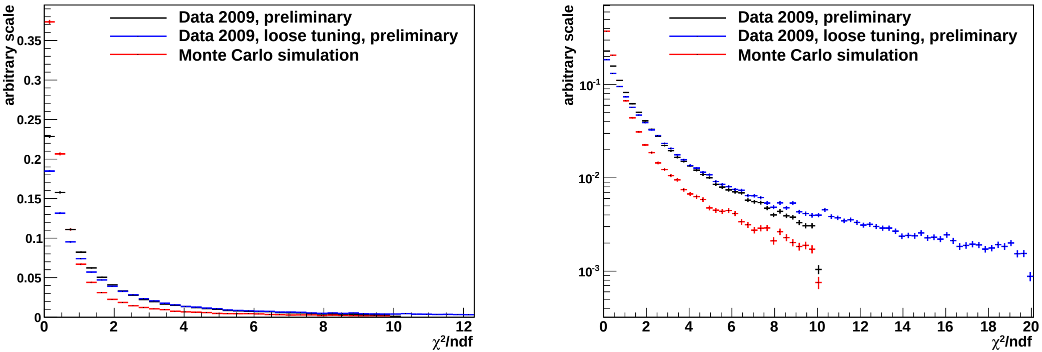
Although the hit efficiency increases the track reconstruction efficiency decreases when only this cut is loosened. If the pattern recognition finds hits that are more far away the χ^2/ndf of the internal fit (cf. Appendix A) increases. This can be seen in Figure 5.5 for the original tuning (data and simulation) and a loose tuning where the parameters of the pattern recognition were changed. The distribution of a loose tuning is shifted to higher values compared to the normal tuning. Then the χ^2/ndf could be above the threshold of 10 for more track candidates and the track candidate is not accepted. Thus more candidates do not pass the cut on the χ^2/ndf and the efficiency in the simulation and the number of tracks in data are lower if the outlier distance is increased. Table 5.5 shows that raising the tolerance to 0.5 mm significantly lowers the efficiency. Therefore the χ^2/ndf has to be increased as well.

Track quality (maximum χ^2/ndf)

Figure 5.5 shows the distribution of the χ^2/ndf in the pattern recognition on the simulation and on data. It is clearly seen that the distribution is shifted to higher values. As seen in Table 5.6, the maximum χ^2/ndf has the biggest impact on the ghost rate but also on the number of tracks. The change of the ghost rate on the simulated data is about 2 % smaller than the increase in the number of tracks on

	outlier distance	0.1 mm	0.3 mm	0.5 mm
Simulation	ϵ_{rec} all tracks	92.3 %	91.8 %	90.6 %
	# TT hits	4.05	4.09	4.09
	TT hit efficiency	98.8 %	99.6 %	99.6%
	TT hit purity	98.8 %	99.8 %	99.8 %
	ghost rate	8.1 %	7.7 %	7.7 %
	# TT hits all tracks	4.03	4.08	4.08
Data	# tracks	78,626	76,757	74,922
	increase	0.0 %	-2.4 %	-4.7 %
	# TT hits	3.94	4.03	4.05

Table 5.5: *Evolution of the tracking performance for change of the outlier distance. The quoted efficiency is the efficiency for all particles if a T-Seed has been found.*



(a) normal scale (b) log scale

Figure 5.5: *The χ^2/ndf of Downstream tracks is shown for simulated and collision data. The blue curve shows the distribution for tune1 in Table 5.7. The black and red distributions stop at 10 because only track candidates with a χ^2/ndf smaller than 10 are accepted in the original tuning.*

data when going from 10 to 20. The other quantities do not change.

Additionally all the parameters mentioned before lead to collecting hits that have a higher distance to the track than the hits that were collected before, cf. Figure 5.5. This also requires to open this cut. Thus the maximum χ^2/ndf was increased from 10 to 20.

Combination

Table 5.7 summarizes the parameter changes for two different tunings. They only differ in the size of the search windows. The impact of combining the changes leads to an increased ghost rate on Monte Carlo of 3.2 % for *tune1* (2.9 % for *tune2*) while keeping resp. increasing the overall efficiency of the original tuning, cf. Table 5.8. The ghost rate is still only about a third of the ghost rate at

	max χ^2/ndf	10	15	20
Simulation	ϵ_{rec} all tracks	92.3 %	92.8 %	93.0 %
	# TT hits	4.05	4.05	4.05
	TT hit efficiency	98.8 %	98.8 %	98.8%
	TT hit purity	98.8 %	99.8 %	99.8 %
	ghost rate	8.1 %	9.2 %	9.9 %
	# TT hits all tracks	4.03	4.02	4.01
Data	# tracks	78,626	81,015	82,289
	increase	0.0 %	3.0 %	4.7 %
	# TT hits	3.94	3.93	3.92

Table 5.6: *Evolution of the tracking performance for change of the χ^2/ndf. The quoted efficiency is the efficiency for all particles if a T-Seed has been found.*

higher energies. The track yield in data is 8.7 % (8.5 %) higher than before. The number of hits on a track is closer to the value of 4.03 (original simulation) and was increased from 3.94 to 4.05 (4.04). Also the fraction between T-Tracks and Downstream tracks increases from 0.51 to 0.58 (0.57) which is close to the expectation of 0.58. This gives confidence that most of the newly found tracks are real tracks and not ghost tracks. In the next section this will be tested on the number of reconstructed K_s signal candidates.

Parameter	original	tune1	tune2
Preselection offset	1 cm	2 cm	2 cm
Search window offset	+0 mm	+3 mm	+1 mm
Outlier distance	0.1 mm	0.3 mm	0.3 mm
χ^2/ndf	10	20	20

Table 5.7: *Three different tunings of the Downstream algorithm.*

	tuning	original	tune1	tune2
Simulation	ϵ_{rec} all tracks	92.3 %	92.8 %	93.2 %
	# TT hits	4.05	4.1	4.09
	TT hit efficiency	98.8 %	99.6 %	99.6%
	TT hit purity	98.8 %	99.8 %	99.8 %
	ghost rate	8.1 %	11.3 %	11.0 %
	# TT hits all tracks	4.03	4.09	4.08
Data	# tracks	78,626	85,536	85,280
	increase	0.0 %	8.7 %	8.5 %
	# TT hits	3.94	4.05	4.04

Table 5.8: *Tracking performance for different tunings, cf. 5.7. The quoted efficiency is the efficiency for all particles if a T-Seed has been found.*

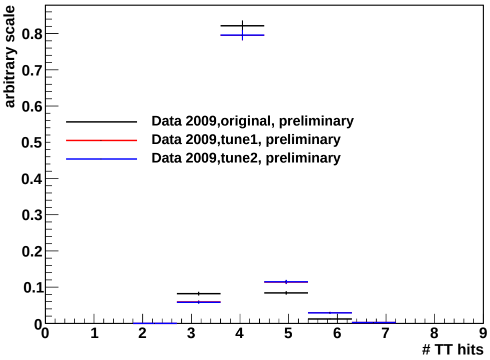


Figure 5.6: *Number of hits in TT on K_s daughter tracks. Black is the old tuning, red and blue new tunings. Note the blue curve lies on top of the red. The average number of hits increases from 4.0 to 4.1. The histograms are sideband subtracted to show the distribution of real K_s candidates (Further details in Chapter 6.6.1).*

5.5 Impact of tuning on K_s yield

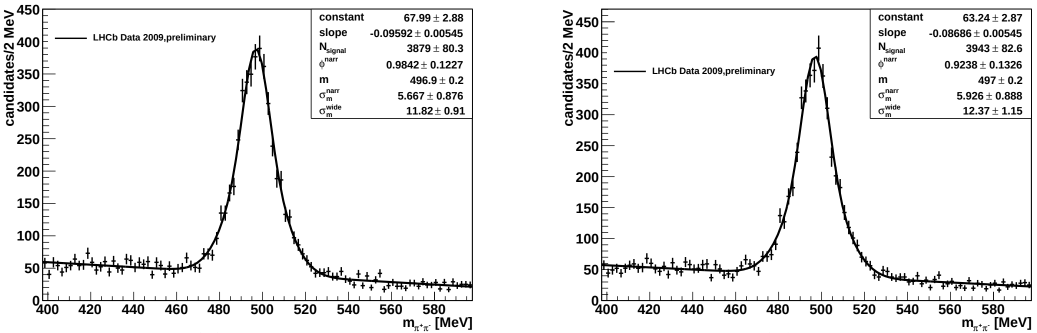
As mentioned before, on data the only way to know if a track is a good one, one has to check if it can be used to reconstruct a resonance. Therefore the different tunings were tested on the K_s resonance[6]. Figure 5.6 shows the number of hits on a daughter track for different tunings. The histograms are sideband subtracted. This means only signal events contribute to the histograms. The corresponding sideband subtraction is explained in detail in Chapter 6.6.1. A clear increase in the number of hits is seen for the two new tunings. Between the two new tunings there is no difference. The yield is shown in Figure 5.7. One sees an increase of about 1.6 % and 2.5 % in the number of reconstructed K_s mesons. The signal to background ratio stays approximately the same. Table 5.9 summarizes the results. The gain is higher for the tuning with the smaller window size (tune2). Thus this tuning was used for the reprocessing of the 2009 data.

Parameter	original	tune1	tune2
K_s yield	3870 ± 80	3943 ± 83	3980 ± 82
$\frac{signal}{background}$	3.17	3.25	3.25

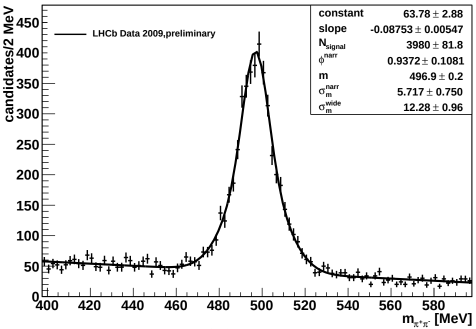
Table 5.9: *K_s yield for the three different tunings. The statistical uncertainties are correlated here because the same sample is used.*

---

[6] The selection is explained in Chapter 6.



(a) original (b) tune1



(c) tune2

Figure 5.7: *Mass distribution of reconstructed K_s candidates. A double Gaussian (Definition in Chapter 6.2.3) is fitted to the distributions. ϕ^{narr} is the fraction of the two widths. The signal yield is (a) 3879 ± 80, (b) 3943 ± 83 and (c) 3980 ± 82. The statistical uncertainties are correlated here because the identical sample is used.*

Chapter 6

Measurement of K_S production cross-section

In the following chapter the measurement of the K_s yield in the data taken end of 2009 is presented. Combined with the luminosity measurement it can be used to determine the cross-section for K_s production in $\sqrt{s} = 900\,\mathrm{GeV}$ proton-proton collisions in the rapidity range that is covered by LHCb. The rapidity is defined as

$$y = arctanh\frac{v}{c} = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}, \tag{6.1}$$

where $(E, \vec{p})$ is the four-momentum of the K_s. Since LHCb is a forward spectrometer, it covers a unique y range. In this region the theoretical uncertainties for QCD processes are rather high because no measurements exist so far. Thus the measured differential cross-section can later be used to tune theoretical models and therefore improve the input to Monte Carlo generators.

In the previous chapter it was mentioned that K_s can be reconstructed either using Long or Downstream tracks. At a center-of-mass energy of 900 GeV the two proton beams have to enter the Velo with a crossing angle as the focusing magnets are not optimised for this low energies. Furthermore the beam is more unstable and not that well focused. This results in a high risk to damage the Velo. Thus in the 2009 data taking period the Velo was moved away from the interaction point by 15 mm. The acceptance of the Velo was therefore highly reduced because particles with a small angle to the beam axis passed through the opening. Therefore they could not be reconstructed as Long tracks, cf. Figure 3.1. Besides this K_s mesons are rather long lived particles which often decay after the Velo. In these cases the daughters cannot be reconstructed as Long tracks and they can only be found as Downstream tracks. Thus the number of reconstructed K_s mesons with Downstream tracks is three to four times higher compared to a set of K_s mesons reconstructed with Long tracks. The disadvantage is that the resolution of Downstream tracks ($\sim 9.2\,\mathrm{MeV}$) is worse compared to Long tracks ($\sim 4\,\mathrm{MeV}$) and there is also more background due to less precise measurements and due to a higher fake rate.

The analysis presented here is solely based on K_s candidates reconstructed with Downstream tracks. To be completely independent of the Velo the selection of K_s candidates presented in Chapter 6.2.2 does not use any information from the Velo. Therefore this analysis is a valuable cross-check for other analyses which are based on Long tracks.

The cross-section in terms of measured quantities is defined as

$$\sigma_{K_s,y,p_T} = N_{K_s\rightarrow\pi^+\pi^-,y,p_T} \;\times\; \epsilon^{-1}_{reco,y,p_T} \;\times\; \epsilon^{-1}_{trigger,y,p_T} \;\times\; \mathcal{L}^{-1} \qquad (6.2)$$

First the yield of reconstructed K_s mesons ($N_{K_s\rightarrow\pi^+\pi^-,y,p_T}$) is measured in the data in different ranges of y and p_T. Then the reconstruction efficiency of K_s mesons (ϵ_{reco,y,p_T}) in these bins is determined on a simulated sample. A numerous of systematic studies are performed within this chapter to test the stability of the extraction of the signal yield in data and the determination of the reconstruction efficiency. The determination of the trigger efficiency ($\epsilon_{trigger,y,p_T}$) and the luminosity ($\mathcal{L}$) measurement were not part of this work and are taken from external sources.

6.1 Simulated and collision data

Not every recorded run in 2009 had the same conditions. For example the Velo was moving in some runs or the L0 thresholds were different. Thus several quality checks have been performed and the list of runs in Table 6.1 has been found to be taken under similar running conditions. They include about 400,000 events. The only difference are the L0 threshold settings used in the various runs. However, in the analysis presented here only events triggered by the calorimeter trigger have been used. As the settings for the calorimeter thresholds were identical in all runs, the used runs can be regarded as similar in all aspects.

The simulated sample consists of 10 million minimum bias events and uses the same reconstruction software as it was used for the processing of the data[1]. The effects of misalignment and detector inefficiencies have been studied and are included in the simulated sample. The K_s yield in this corrected sample is lowered by about 10 % with respect to an ideal Monte Carlo sample without any corrections. To do systematic studies two additional Monte Carlo samples were produced. One sample overestimates the corrections and one sample underestimates the corrections. A detailed description of the samples can be found in [28].

---

[1] Brunel v37r0 and DaVinci v25r2p3 were used to process the collision and the simulated data

run number	# physics events	TCK
63686	24391	0x1209
63687	15642	0x1209
63688	2169	0x1209
63690	20855	0x1209
63691	2074	0x1209
63713	14295	0x1209
63801	94112	0x1309
63807	75285	0x1309
63809	23465	0x1309
63813	71429	0x1309
63814	4629	0x1309
63815	11668	0x1309
63949	64179	0x1309

Table 6.1: *List of runs which are used in the analysis. The first column contains the run number, the second one the number of physics events in these runs and the last column the trigger configuration key (TCK).*

6.2 K_s reconstruction on 2009 data

In this section the measurement of the K_s signal yield in different ranges of y and p_T in the 2009 data is presented. Therefore a selection which is based on Downstream tracks only is developed and a fit model for the K_s mass distribution is introduced.

6.2.1 Beam-gas subtraction

The vacuum inside the beam pipe is not perfect. Therefore it is possible that the protons in the beams collide with the residual gas molecules. These so called beam-gas events are highly boosted into the forward direction along the beam pipe due to the high momentum difference of the proton and the gas molecule. However, because of the forward geometry of LHCb, these events can actually be observed. A collision of a proton beam with the residual gas can happen when only one beam is passing by the detector (beam-empty crossing) or when two beams cross each other (beam-beam crossing). Figure 6.1 shows the $x - z$ coordinate of the reconstructed primary vertices in beam-empty and beam-beam crossings. The luminous region of proton-proton collisions is about $\pm 15\,\text{cm}$ around the origin, cf. Figure 6.2. The beam-gas events contribute significantly to the observed events in beam-beam crossings. However, for the study presented here only a K_s candidate that was produced in proton-proton collisions is of interest. Therefore the beam-gas events are subtracted statistically from the beam-beam events to get a measurement of only proton-proton collisions. The

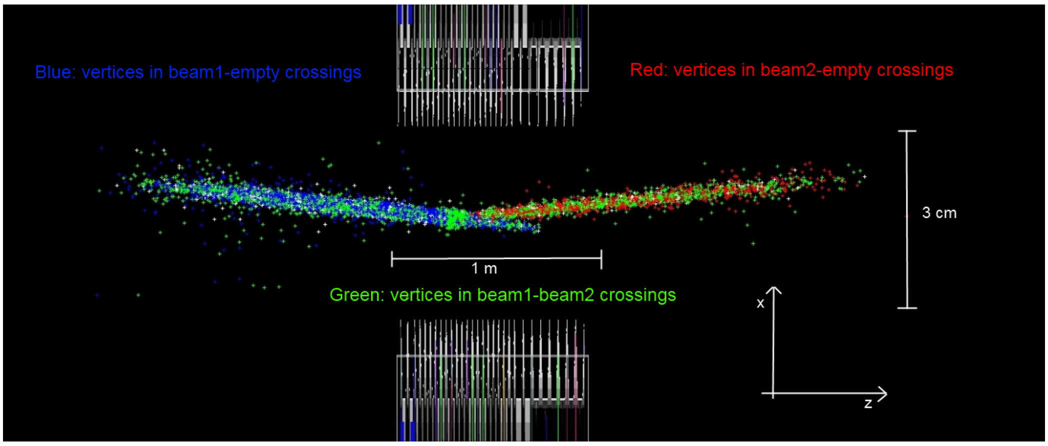


Figure 6.1: *Top view of the opened Velo. The colored points are reconstructed primary vertices. The crossing angle of the two beams can be seen.*

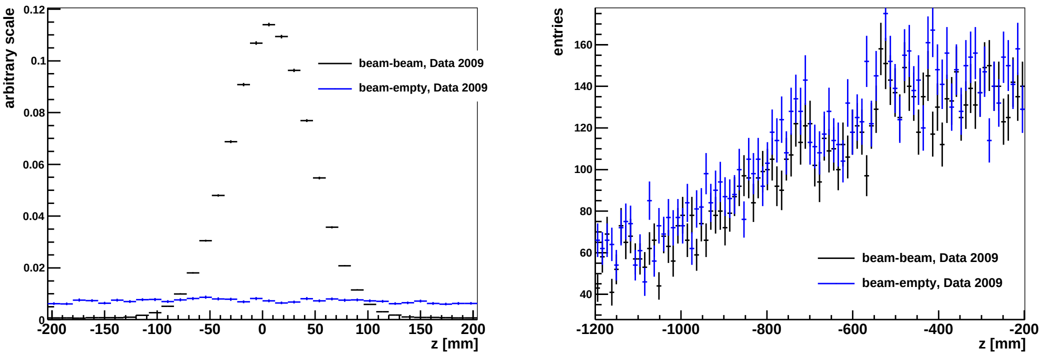


Figure 6.2: *z coordinate of the primary vertex. (left) Luminous region (Histograms are normalized). (right) Zoom into the region $[-1200\,\text{mm}, -200\,\text{mm}]$ (Histograms are not normalized).*

subtraction is done by

$$N_{proton-proton} = N_{beam-beam} - \beta\ N_{beam-empty} \tag{6.3}$$

where β represents the weighting factor because of different beam currents. And N represents an arbitrary observable one wants to measure, e.g. the number of reconstructed K_s or the track multiplicity. In every run from Table 6.1 the LHC was filled in such a way that the number of beam-empty and beam-beam passings was the same. Thus a β value not equal to unity indicates that the current was not the same in all injected beams.

β can be determined by counting reconstructed primary vertices. In Figure 6.2 the z coordinate of reconstructed primary vertices is shown for events which happened during a bunch crossing (black) and for events which happened when only one beam was passing through the detector (blue). Both histograms are normalized. A clear peak around 0 with a spread of about ±15 cm can be seen in the distribution of beam-beam events. This defines the luminous region. In the beam-empty events no peak can be seen. The right plot zooms into the region from $[-1200\,\mathrm{mm}, -200\,\mathrm{mm}]$ and it is not normalized. There the shape of the two histograms is compatible. Therefore events with a reconstructed primary vertex in this region in beam-beam events are actually beam-gas events. This can be used to determine β. The number of primary vertices outside the luminous region in both histograms is counted and their ratio is equal to β:

$$\beta = \frac{N_{beam-beam}}{N_{beam-empty}}. \tag{6.4}$$

β is determined to be 0.91 ± 0.01. By varying e.g. the range of the beam-gas region it has been studied that this value is stable. Thus only the statistical uncertainty is given.

Figure 6.3 shows K_s candidates[2] reconstructed in beam-empty events and in beam-beam events. Also in beam-empty events a clear peak can be seen. Figure 6.6 shows the result after the statistical subtraction of beam-empty events. About 2.1 % of all reconstructed K_s candidates are in beam-empty crossings. The reconstruction of beam-gas events in beam-beam crossings has an effect of about 1.3 % on the K_s yield. In the following only beam-gas subtracted quantities are shown.

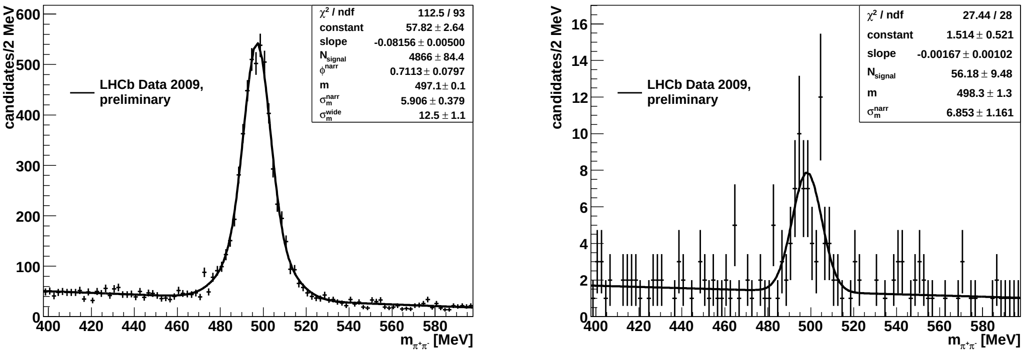


Figure 6.3: *K_s candidate in beam-beam events (left) and in beam-empty events (right). Selection explained in Chapter 6.2.2.*

6.2.2 K_s candidate selection

The selection of K_s candidates differs to the selection presented in Chapter 4.6. This selection is optimised to find particles originating from the primary vertex (prompt particles). Another point is that it does not use any information from the Velo. Thus a lot of quantities like impact parameter and lifetime are not accessible

---

[2] The selection is presented in the next section (6.2.2).

because they would require a reconstructed primary vertex which is associated to the particle. To bypass this disadvantage, the assumption that the primary vertex is close to the beam axis in the $x-y$ plane will be used. The distribution of the $x-y$ coordinate of reconstructed primary vertices is shown in Figure 6.4. The majority of primary vertices is close to the beam line. Thus prompt particles should point to the beam axis. Additionally only K_s candidates which were found in events which were triggered by the L0 calorimeter trigger are taken into account as most of the events were selected by the calorimeter trigger. The quantity $\frac{\sqrt{Signal+Background}}{Signal}$ is proportional to the statistical uncertainty of the measurement, thus it was used to optimise the selection. The selection was optimised on data but due to the high statistics it should not affect the measurement that the measurement and the optimisation were done on an identical sample.

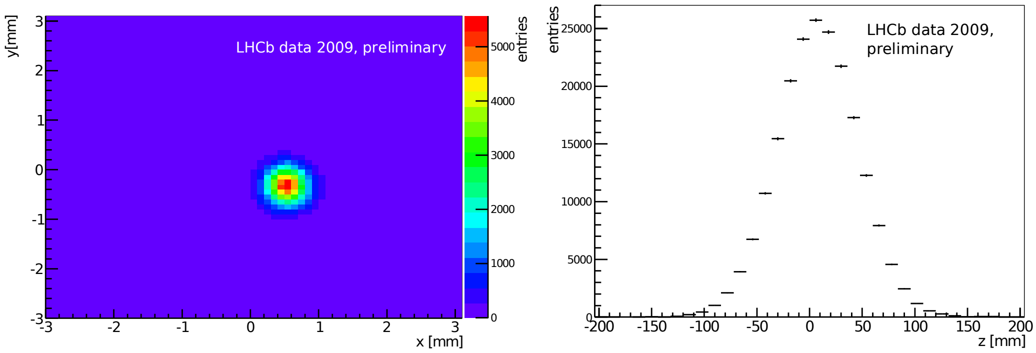


Figure 6.4: *(left) x and y coordinate of the primary vertex. (right) z coordinate of the primary vertex.*

The selection cuts are listed in Table 6.2. The first four cuts are concerning the daughter particles of the K_s candidate. First a minimal momentum p of 2 GeV and a minimal transverse momentum p_T of 50 GeV is required. The χ^2/ndf of the track fit has to be below 25 because ghost tracks tend to have a higher χ^2/ndf than tracks of real particles. Charged particles originating in the proton-proton collision are rejected by a cut on the distance of closest approach to the beam line, cf. Figure 6.5. Pions from a K_s decay should have a higher value due to the long lifetime of the K_s. A distance greater than 3 mm is required for both daughters.

To remove combinations of particles which do not originate from the same particle the χ^2 of the K_s vertex fit has to be below 25. Then the point of closest approach of the K_s candidate to the beam line is calculated. The z coordinate of this point is taken as a “pseudo” primary vertex with the coordinates $(0, 0, z)$. This assumption is justified because the interaction point is well defined in the $x - y$ plane but not so well along the z axis. The distributions of the primary vertex coordinates are seen in Figure 6.4. A small shift appears in the $x - y$ plane. However, it has no significant impact when doing the selection with these coordinates instead of $(0, 0, z)$. The point of closest approach is required to be

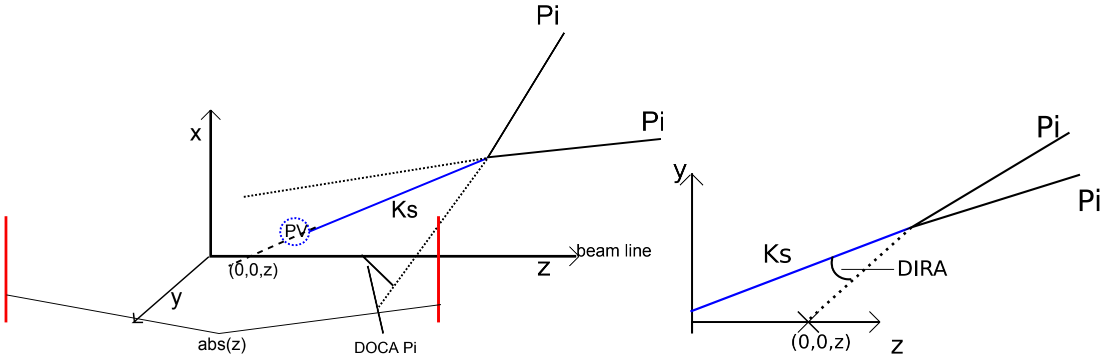


Figure 6.5: *The geometrical cuts in the selection are depicted.*

within the luminous region of $[-150\,\text{mm}, 150\,\text{mm}]$ around the origin. Then the direction angle (DIRA) is calculated with respect to the "pseudo" primary vertex. The direction angle is defined as the angle between the K_s momentum and the line between the K_s decay vertex and the "pseudo" primary vertex. With this variable (DIRA>0.99995) particles coming from the beam line are selected. The "pseudo" primary vertex is also taken to estimate the lifetime of the K_s candidate by taking the difference between the "pseudo" primary vertex and the decay vertex as the flight distance. The estimated lifetime (ct) has to be larger than 5 mm. As the first station in the TT is at 2200 mm the z coordinate of the decay vertex has to be smaller than 2200 mm.

Additionally the values of a preselection are given in Table 6.2. The preselection is used to reduce the amount of the sample on which the studies are performed. Furthermore the selection efficiency relative to the preselection will be determined as a systematic study in Chapter 6.6.

	value	
cut	preselection	selection
min π p	-	2 GeV
min π p_T	25 MeV	50 MeV
max track χ^2/ndf	35	25
min π IP to z axis	–	3 mm
max K_s vertex χ^2/ndf	35	25
max z of decay vertex	2200 mm	2200 mm
max $\lvert z \rvert$ of point of closest approach to z axis	500 mm	150 mm
min K_s DIRA to $(0, 0, z)$	0.995	0.99995
min K_s ct	–	5 mm

Table 6.2: *K_s preselection and selection cuts. The lifetime is calculated with the "pseudo" primary vertex.*

6.2.3 Fit model for the K_s mass distribution

In Figure 6.6 the distribution of all reconstructed and selected K_s candidates[3] is shown. The fit model that was chosen has one term for the background and one for the signal peak. For the background a linear model was taken:

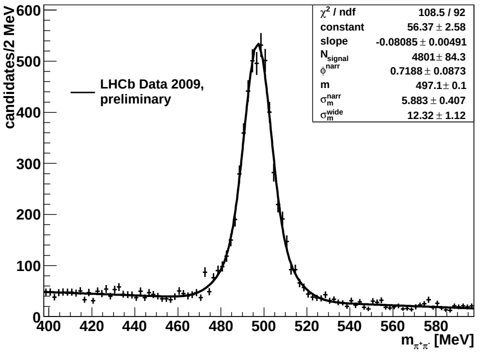


Figure 6.6: *K_s candidate mass distribution for 2009 data. The statistic box shows the χ^2/ndf and the parameters of the fit. They are explained in the text.*

$$f_B = constant + slope \times m_{\pi^+\pi^-} \tag{6.5}$$

The mass peak is best described by the sum of two Gaussians (double Gaussian):

$$f_S(N, \mu, \phi, \sigma_{narr}, \sigma_{wide}) = N\ (\cos^2\phi\ G(\mu, \sigma_{narr}) + \sin^2\phi\ G(\mu, \sigma_{wide})). \tag{6.6}$$

The angle ϕ describes the mixture of the narrow and the wide Gaussian in the mass peak.This parametrisation was chosen because it automatically includes the boundary condition that the fraction of the two width has to be between 0 and 1. Thus the fit is more stable. The total width is given by

$$\sigma_{tot} = \sqrt{\cos^2\phi\ \sigma^2_{narr} + \sin^2\phi\ \sigma^2_{wide}}. \tag{6.7}$$

The analysed data contain 4801 ± 84 K_s signal candidates. The weighted sum of the two Gaussians is 9.2 MeV.

The fit model describes the distribution very well ($\chi^2/ndf = 109/92$). However, in the following analysis the K_s candidates will be divided in intervals of rapidity y and transverse momentum p_T of the K_s candidate. There the statistics is much lower and the stability of the fit becomes crucial. In these cases the fit model is

---

[3] The quoted yield differs from the yield given in the previous chapter as a newer reconstruction software version was used. The new version includes a better alignment of the tracking detectors and a further tuning of the T-Seeding algorithm. However, the Downstream algorithm remained unchanged as it was already tuned to a higher efficiency on data. Brunel v37r0 and DaVinci v25r2p3 were used to process the collision and the simulated data.

	2009 data
N	4801 ± 84
σ_{narr} [MeV]	5.9 ± 0.4
σ_{wide} [MeV]	12.3 ± 1.2
m [MeV]	497.1 ± 0.1
ϕ^{narr}	0.72 ± 0.09
slope	-0.08 ± 0.01
a	56.3 ± 2.5
σ_{tot} [MeV]	9.2

Table 6.3: *Fit parameters of the mass distribution of K_s candidates in data. $\cos^2(\phi^{\mathrm{narr}})$ is the fraction of the Gaussian with width σ_{narr}.*

adjusted. Therefore a single Gaussian which has less parameters is fitted[4] as well and depending on the fit results either the single or the double Gaussian is chosen. The criteria for this decision are the following:

- If the statistics is too low a single Gaussian is chosen. The statistics is defined as too low when the estimated yield of the single Gaussian is below 100.
- If the double Gaussian is essentially a single Gaussian. This is the case when the difference of the two widths is smaller than 0.5 MeV or when the fraction of the Gaussians is close to 0 or 1.
- If the width of the wider Gaussian is greater than 15 MeV a single Gaussian is chosen to avoid that the second Gaussian describes the background.
- If the fit of the double Gaussian failed, the single Gaussian is chosen.
- In all other cases, the double Gaussian is taken.

The stability of the fit has been evaluated in different systematic studies which will be presented in Chapter 6.6.1.

6.3 Sideband subtraction

The histograms of K_s quantities like e.g. y, p or p_T should show the distribution of real K_s mesons but some fake K_s candidates pass the selection. Therefore fake K_s candidates in the selection are subtracted statistically. The assumption

---

[4] The fitter is based on a MINUIT χ^2 fit. For the estimated number of entries in each bin the integral over the fitting function is used instead of its value in the bin center. In regions with low statistics bins are combined in the fit until each combined bin has at least ten entries or five adjacent bins have been combined. This procedure ensures a proper error estimate for mass distributions with low statistics while retaining a fine binning in the case of high statistics.

is made that the fake K_s candidates in the background region have the same distributions as the fake K_s candidates in the signal region. Thus the K_s candidates in the background region can be subtracted from the candidates in the signal region. The mass region is defined as $(497.614 \pm 100)\,\text{MeV}$. The signal region is $(497.614 \pm 50)\,\text{MeV}$. The rest inside the mass region is the background region $([-100,\ -50],\ [+50,\ +100])$. The ranges are also depicted in Figure 6.7.

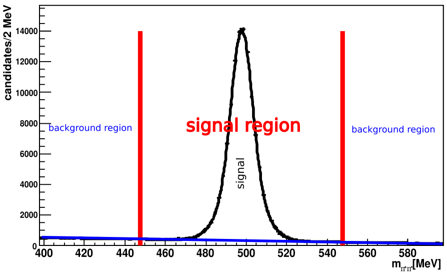


Figure 6.7: *Definition of signal and background region.*

For the sideband subtraction the background is fitted with a function f_{bg}. Then the fake K_s candidates in the background region get a negative weight which is determined by the fraction of the integrals of the background function in the background region and the signal region:

$$w = -\frac{\int_{signal\ region} f_{bg}}{\int_{background\ region} f_{bg}}. \tag{6.8}$$

As the background model is linear here, the fraction of the background inside the background region and inside the signal region is independent of the fitted parameters. With the ranges defined above the fraction is actually equal to one. Thus when filling histograms the candidates inside the signal region get the weight $+1$ and the candidates inside the background region get the weight -1. However, in the systematic studies later other background models are used as well.

Sideband subtraction can also be used to determine the signal yield. For this the number of entries in the signal region is counted. Then the number of entries in the background region is subtracted with the weight defined in Formula 6.8.

6.4 Measurement of K_s yield in y, p_T bins

The measurement of the total yield obtained in data has no conclusiveness as the LHCb detector only covers a very small solid angle region. Thus the measurement is done in bins of the transverse momentum p_T and the rapidity y of the K_s mesons. The transverse momentum p_T is given in the rest frame of the two proton beams. This is slightly different to the laboratory system because of the crossing

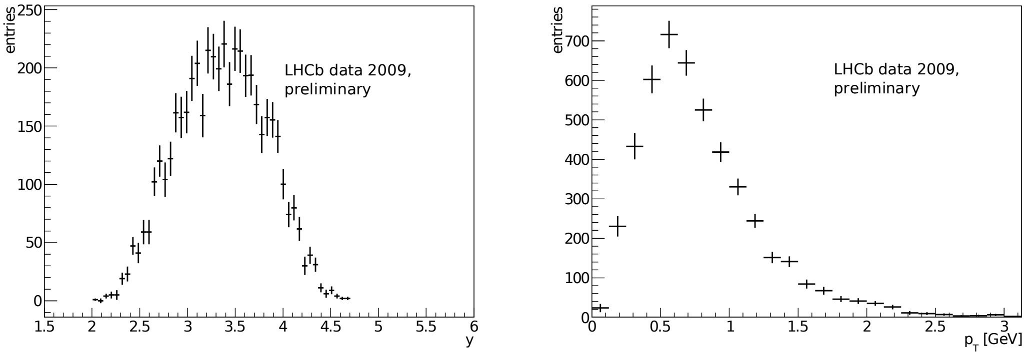


Figure 6.8: *y and p_T distributions of K_s signal candidates in data.*

angle of the two beams. The distributions of p_T and y are shown in Figure 6.8. The binning in p_T and y of the K_s candidates is a compromise between a fine granularity to use it as input for theory models and a sufficient statistics in every bin. The chosen bins in y are

- 2.5 - 3.0,
- 3.0 - 3.5,
- and 3.5 - 4.0,

and in p_T

- 200 – 400 MeV,
- 600 – 800 MeV,
- 800 – 1000 MeV,
- 1000 – 1200 MeV,
- 1200 – 1400 MeV,
- and 1400 – 1600 MeV.

The resulting mass distributions together with the fitted functions are shown in Figure 6.9, 6.10 and 6.11. Table 6.4 summarizes the obtained yields. They range from about 20 in the lowest bin to about 400 in the bin with the highest statistic. The chosen fit model can be seen when looking at the number of parameters in the plots. In the case of a single Gaussian only one width is quoted. The χ^2/ndf of the fit is reasonable in every bin, cf. Table 6.4. The number of degrees of freedom is not always the same because as explained earlier the fitter sums up bins with too low statistics.

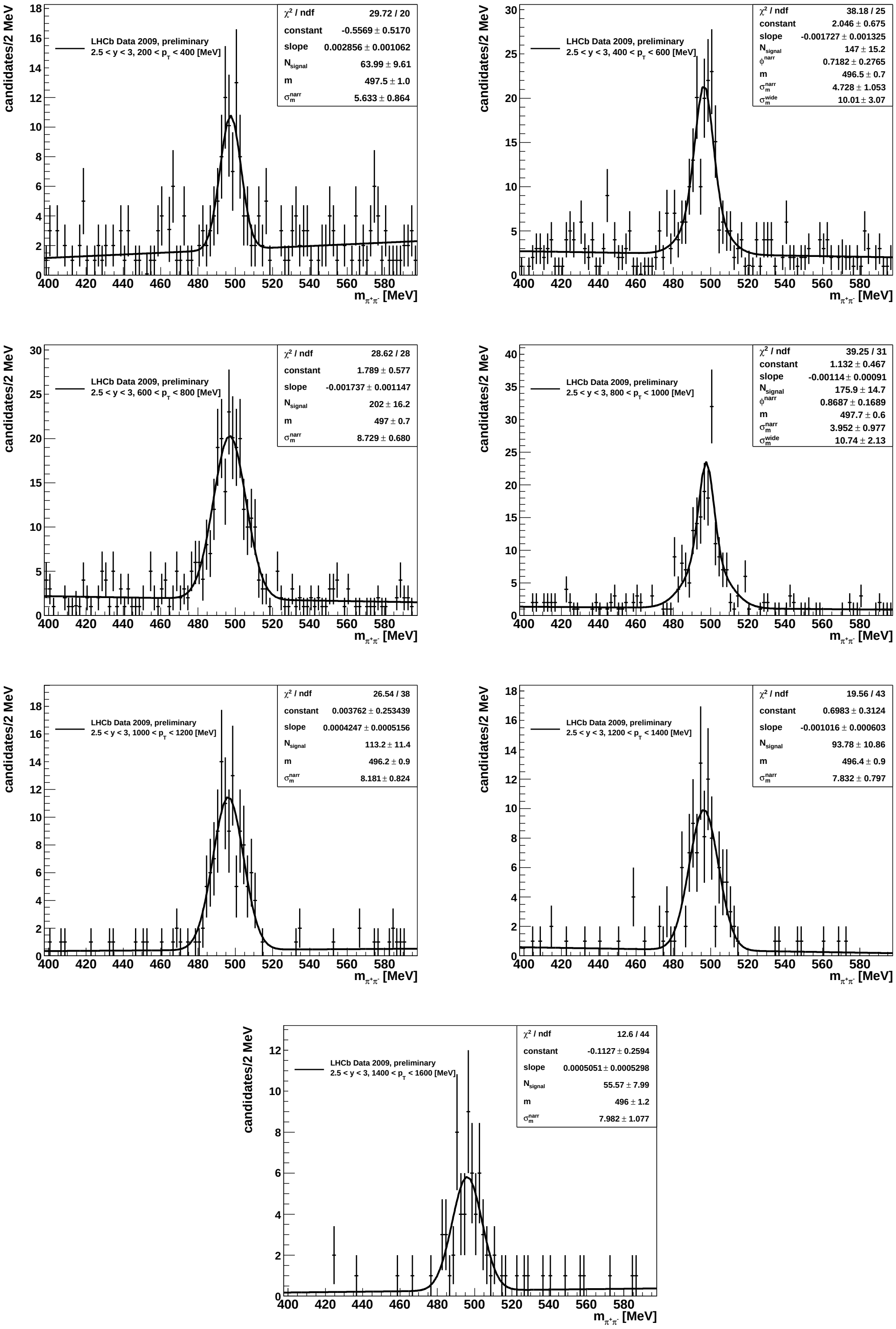


Figure 6.9: *K_s candidate mass distribution in data in different p_T bins with $2.5 < y < 3.0$.*

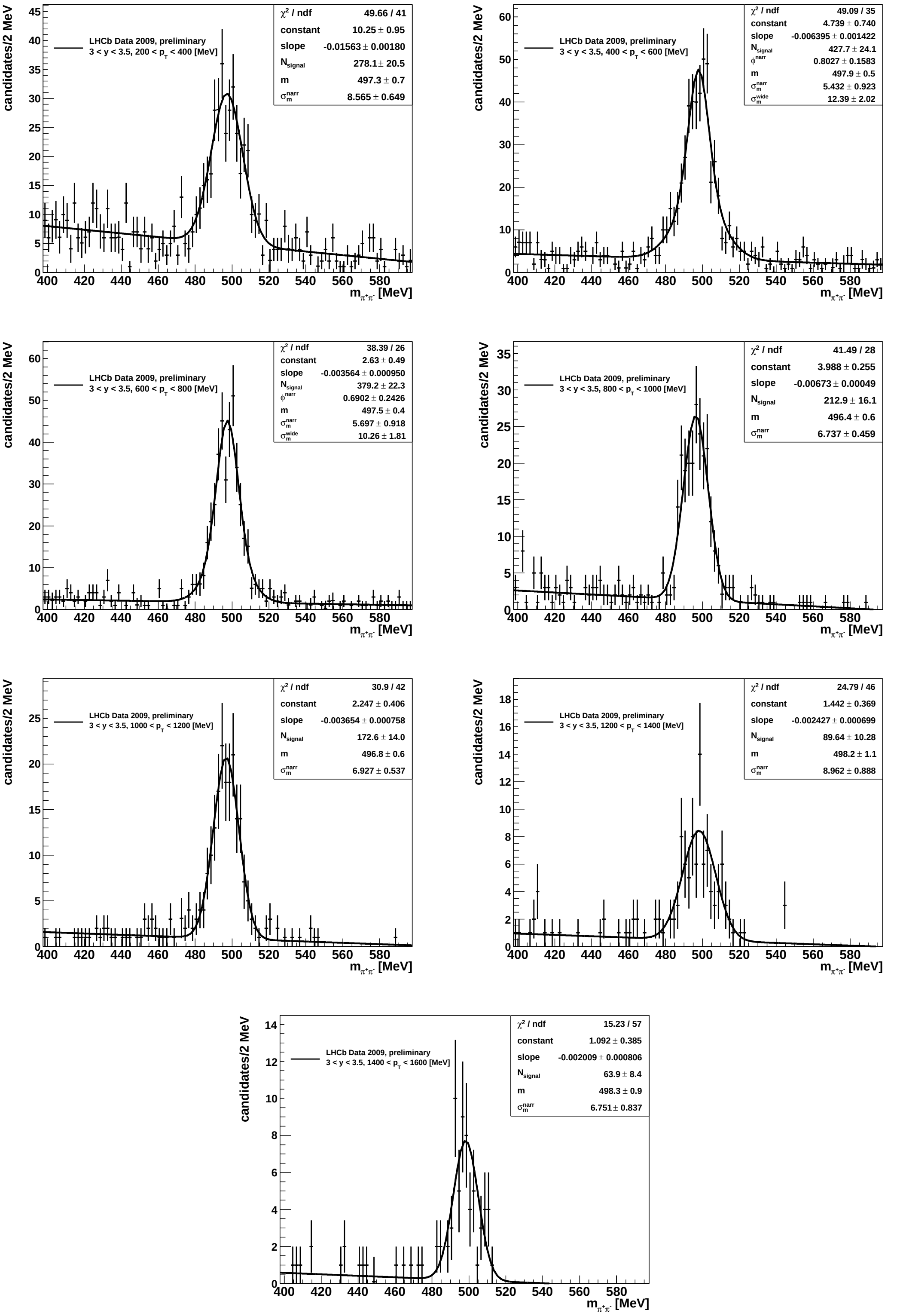


Figure 6.10: *K_s candidate mass distribution in data in different p_T bins with $3.0 < y < 3.5$.*

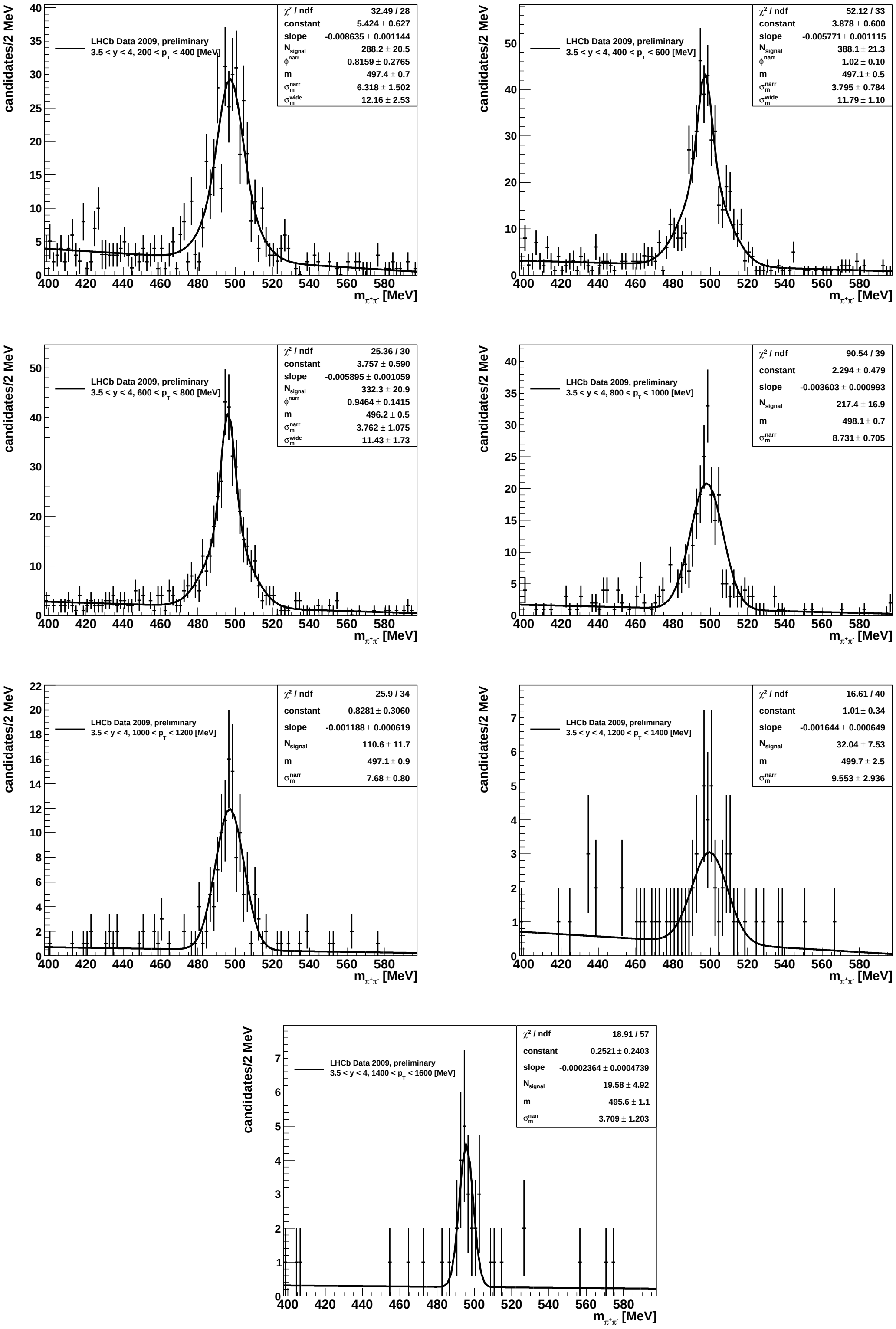


Figure 6.11: *K_s candidate mass distribution in data in different p_T bins with $3.5 < y < 4.0$.*

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	64 ± 10	278 ± 21	288 ± 21
400 − 600	147 ± 15	428 ± 24	388 ± 21
600 − 800	202 ± 16	379 ± 22	332 ± 21
800 −1000	176 ± 15	213 ± 16	217 ± 17
1000 −1200	113 ± 11	173 ± 14	111 ± 12
1200 −1400	94 ± 11	90 ± 10	32 ± 8
1400 −1600	56 ± 8	64 ± 8	20 ± 5

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	30/20	50/41	32/28
400 − 600	38/25	50/35	52/33
600 − 800	29/28	38/26	25/30
800 −1000	39/31	41/28	91/39
1000 −1200	27/38	31/42	26/34
1200 −1400	20/43	25/46	17/40
1400 −1600	13/44	15/57	19/57

Table 6.4: *Number of signal candidates in p_T and y bins and the corresponding fit quality (χ^2/ndf). The different number of degrees of freedom is related to the chosen fit model (single or double Gaussian) and the number of empty bins in the histogram.*

6.5 Determination of the reconstruction efficiency on simulated data

The efficiency to reconstruct a K_s meson is determined on the simulated sample mentioned in Chapter 6.1 where the number of generated K_s mesons is known. First a comparison between the data and the simulation is presented. Then the reconstruction efficiency is defined and determined.

6.5.1 Distributions of kinematic and selection variables in collision and simulated data

The mass distributions for collision and simulated data are shown in Figure 6.12 and the corresponding fit parameters are given in Table 6.5. The width of the mass peak in the simulation is smaller compared to the one obtained from data. This is expected due to the remaining misalignment in the data. The histograms

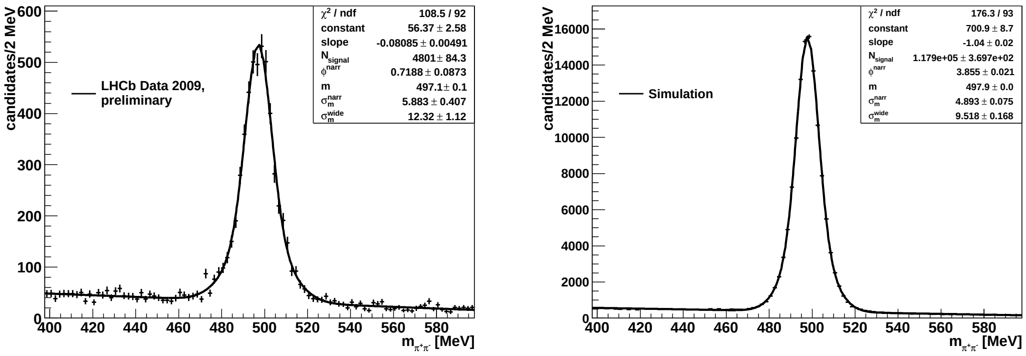


Figure 6.12: *K_s candidate mass distribution for data (left) and simulated data (right).*

	2009 data	MC
N	4801 ± 84	117910 ± 370
σ_{narr} [MeV]	5.9 ± 0.4	4.9 ± 0.1
σ_{wide} [MeV]	12.3 ± 1.1	9.5 ± 0.2
m [MeV]	497.1 ± 0.1	497.9 ± 0.1
ϕ^{narr}	0.72 ± 0.09	3.85 ± 0.02
slope	-0.08 ± 0.01	-1.04 ± 0.02
a	56.3 ± 2.6	700.9 ± 8.7
σ_{tot} [MeV]	9.2	7.2

Table 6.5: *Fit parameters to mass of K_s candidates in data and simulated events. ($\cos^2(\phi^{\mathrm{narr}})$ is the fraction of the Gaussian with width σ_{narr}.)*

of the K_s quantities are all sideband subtracted to show the distributions of real K_s candidates. Figure 6.13 shows a comparison of the y and p_T distributions. Whenever a comparison between collision and simulated data is performed, the simulated events are required to pass the L0 trigger. On data the K_s tend to have higher transverse momentum p_T compared to the simulation.

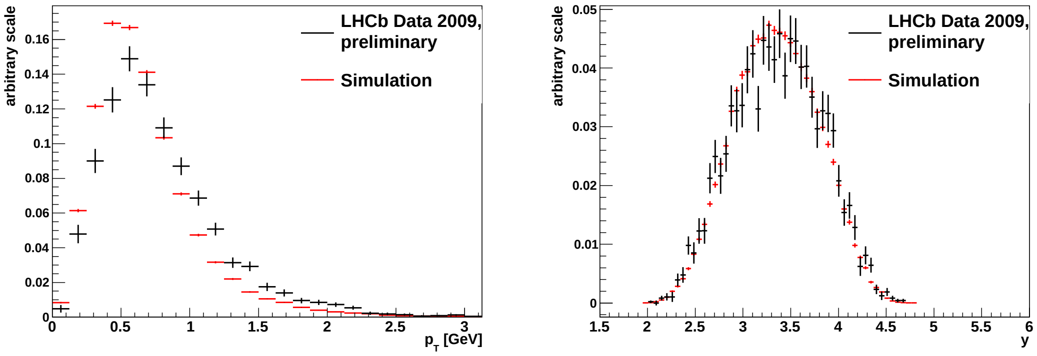


Figure 6.13: *p_T (left) and y (right) distribution of K_s for data (black) and simulated data (red).*

The histograms in Figure 6.14 show the distributions of the variables that are used in the selection. The plots are made after the preselection and are sideband subtracted. Differences can be seen especially in the geometrical variables like impact parameter to the beam axis of the daughter tracks or the lifetime of the K_s candidate. All these variables require a propagation of the tracks through the B-field. Thus they are affected by the misalignment in the tracking detectors and a potential wrong B-field estimate. Some tracks of K_s daughters have to be transported more than a distance of 2 m. Small inaccuracies in the track reconstruction are increased by the propagation through the B-field. In Chapter 6.6 the impact of these differences on the measurement of the total yield will be discussed.

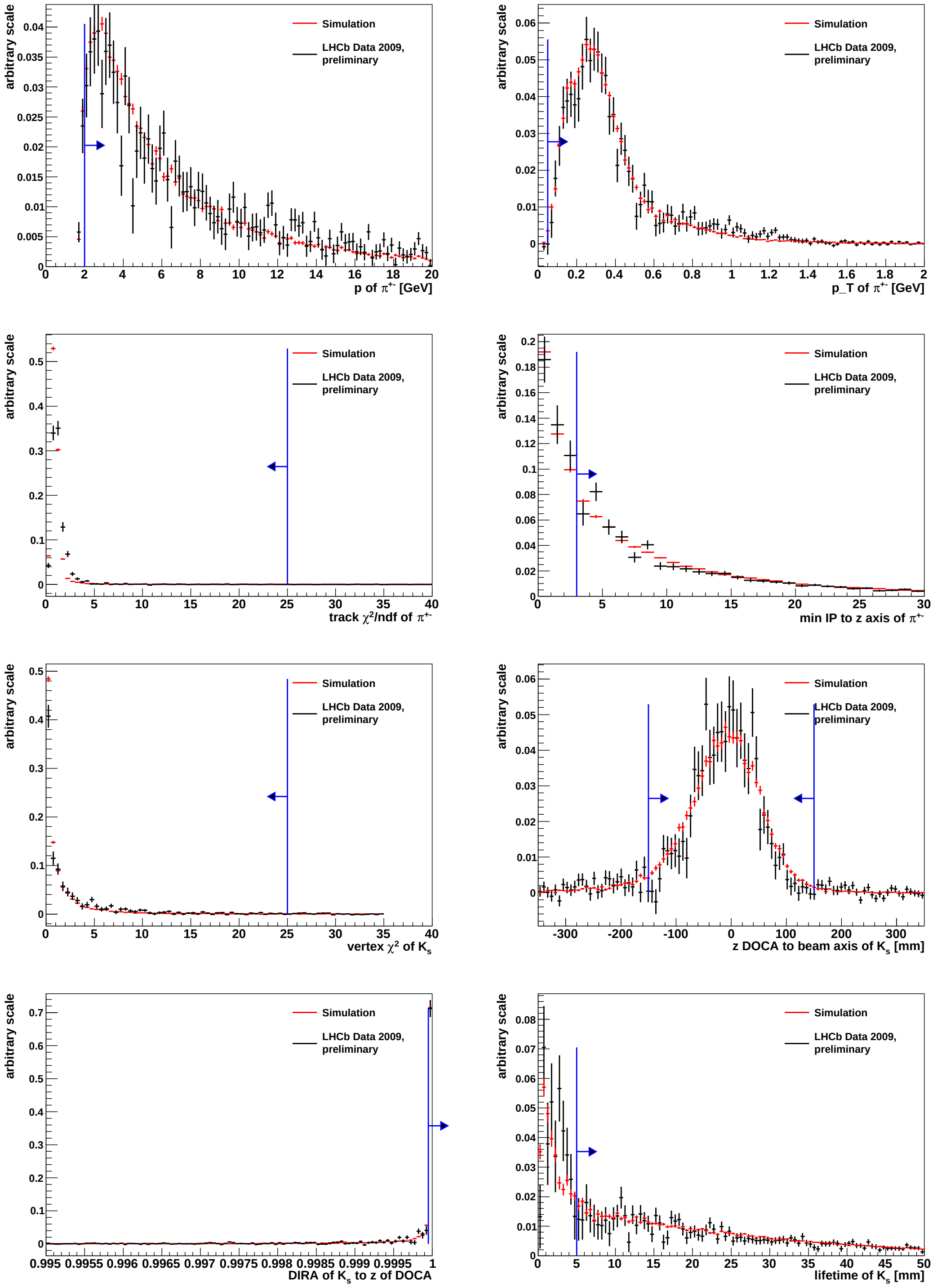


Figure 6.14: *Cut variable distributions for signal K_s candidates in data (black) and simulation (red) after the preselection. Differences in these quantities will be discussed in the section on systematic uncertainties. The blue lines indicate the cut values.*

6.5.2 Reconstruction efficiency in y, p_T bins

The reconstruction efficiency is taken from studies on simulated data.

Efficiency denominator

The efficiency denominator is defined by the number of generated K_s in proton-proton collision in each phase space bin. Particles originating from the primary vertex are called prompt particles. Although only K_s mesons which decay into charged pions are reconstructible, there is no requirement that the generated K_s decays into two charged pions. The branching fraction of

$$\mathcal{BR}(K_s \to \pi^+\pi^-) = 69.2 \pm 0.05\ \%. \tag{6.9}$$

is however included in the generation of the events and thus the reconstruction efficiency automatically includes the correction to the branching fraction $\mathcal{BR}(K_s \to \pi^+\pi^-)$. The advantage of this definition is that K_s mesons which are absorbed in the detector are included in the efficiency denominator.

Furthermore, for determining reconstruction efficiencies the total number of generated K_s is of interest. Thus for the efficiency determination no L0 requirement is requested. The efficiency to trigger an event will be applied afterwards.

Efficiency numerator and reconstruction efficiency

The efficiency numerator should include all reconstructed and prompt K_s Monte Carlo particles. However, there are different ways to define the efficiency numerator:

- The reconstructed particle is **associated** to its corresponding Monte Carlo particle. The p_T, y bin is defined by the **generated momentum**. Thus the efficiency is defined as

$$\epsilon_1 = \frac{\text{Number of associated prompt } K_s \text{ with generated } p_T \text{ and } y \text{ in a given bin}}{\text{Number of generated prompt } K_s \text{ in a given } p_T, y \text{ bin}} \tag{6.10}$$

- The reconstructed particle is **associated** to its corresponding Monte Carlo particle. The p_T, y bin is defined by the **reconstructed momentum**. Thus the efficiency is defined as

$$\epsilon_2 = \frac{\text{Number of associated prompt } K_s \text{ with reconstructed } p_T \text{ and } y \text{ in a given bin}}{\text{Number of generated prompt } K_s \text{ in a given } p_T, y \text{ bin}} \tag{6.11}$$

- The number of reconstructed particles is taken from the **fit** to the invariant mass distribution. The p_T, y bin is defined by the **reconstructed** quantities.

Thus the efficiency is defined as

$$\epsilon_3 = \frac{\text{Yield obtained by fit in given } p_T, y \text{ bin}}{\text{Number of generated prompt } K_s \text{ in a given} p_T, y \text{ bin}} \tag{6.12}$$

All three definitions are useful for different purposes. The difference of the first two efficiencies shows the impact of the detector resolution on the binning. E.g. if a particle has a generated momentum close to the lower bin limit and the reconstructed momentum is too low then the particle would contribute to the wrong bin, affecting the efficiency in both bins. Thus the difference (Table 6.6) is taken as a systematic uncertainty for the resolution effects on the measurement.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	< 0.1	< 0.1	< 0.1
400 − 600	< 0.1	< 0.1	0.1
600 − 800	< 0.1	< 0.1	< 0.1
800 −1000	< 0.1	< 0.1	< 0.1
1000 −1200	< 0.1	0.1	< 0.1
1200 −1400	< 0.1	< 0.1	0.2
1400 −1600	0.1	0.2	0.2

Table 6.6: *Difference in reconstruction efficiencies (in percent) for using the reconstructed p_T and y or the generated p_T and y for the K_s candidates which enter the numerator of the efficiency correction. As these numbers are highly correlated, no uncertainties are given. The observed difference will be added to the systematic uncertainties.*

The difference between the second and the third definition is later (Chapter 6.6) used for systematic studies of the fit model. In this measurement the third definition is chosen as default to determine efficiencies because the determination of the efficiency in the simulation and the determination of the yield in data should be done in the same way. On data prompt and non-prompt particles cannot be distinguished if the momentum of the K_s mesons points to the primary vertex. Thus it is correct to include non-prompt particles in the efficiency calculation as long as the fraction of prompt and non-prompt particles can be assumed to be the same in data and simulation. Thus non-prompt particles are included in the yield measurement on simulation and data. On simulated data about 0.6 % of the reconstructed particles are non-prompt. To account for the fact that the simulation might not reproduce the correct fraction of non-prompt K_s mesons a systematic uncertainty will be added later.

Furthermore the definition of associating a Monte Carlo particle to a reconstructed candidate is too strict in some cases. A K_s candidate is associated to a Monte Carlo particle when both daughters fulfil the criteria to be associated to a Monte Carlo particle. Sometimes the daughters tracks do not fulfil the definitions (cf. Chapter 2), but they can still have enough information to reconstruct a good K_s candidate. Figure 6.18 later shows the distribution of non-truthmatched K_s candidates, a clear peak can be seen. It contains about 3 % of the reconstructed K_s. These K_s candidates are automatically included in the third definition. Further details and systematic studies are given in Chapter 6.6.

Determined reconstruction efficiency

In Figure 6.15, 6.16 and 6.17 the fitted mass peaks are shown for the different bins. The obtained efficiency to reconstruct a K_s is given in Table 6.7. The statistics is much higher compared to the statistics on data. Thus in almost every bin a double Gaussian is fitted.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 – 400	1.2 ± 0.1	7.0 ± 0.1	7.1 ± 0.1
400 – 600	3.4 ± 0.1	11.3 ± 0.1	11.5 ± 0.1
600 – 800	7.1 ± 0.1	14.4 ± 0.2	14.3 ± 0.2
800 –1000	10.7 ± 0.2	16.4 ± 0.2	15.0 ± 0.3
1000 –1200	13.6 ± 0.3	17.7 ± 0.4	14.4 ± 0.4
1200 –1400	15.3 ± 0.4	18.1 ± 0.5	12.9 ± 0.6
1400 –1600	16.8 ± 0.6	18.1 ± 0.7	11.9 ± 0.7

Table 6.7: *Reconstruction efficiencies (in percent) for using the fit to extract the number of reconstructed K_s candidates. Contributions from non-prompt K_s are per construction included in the numerator, however not in the denominator of the efficiency.*

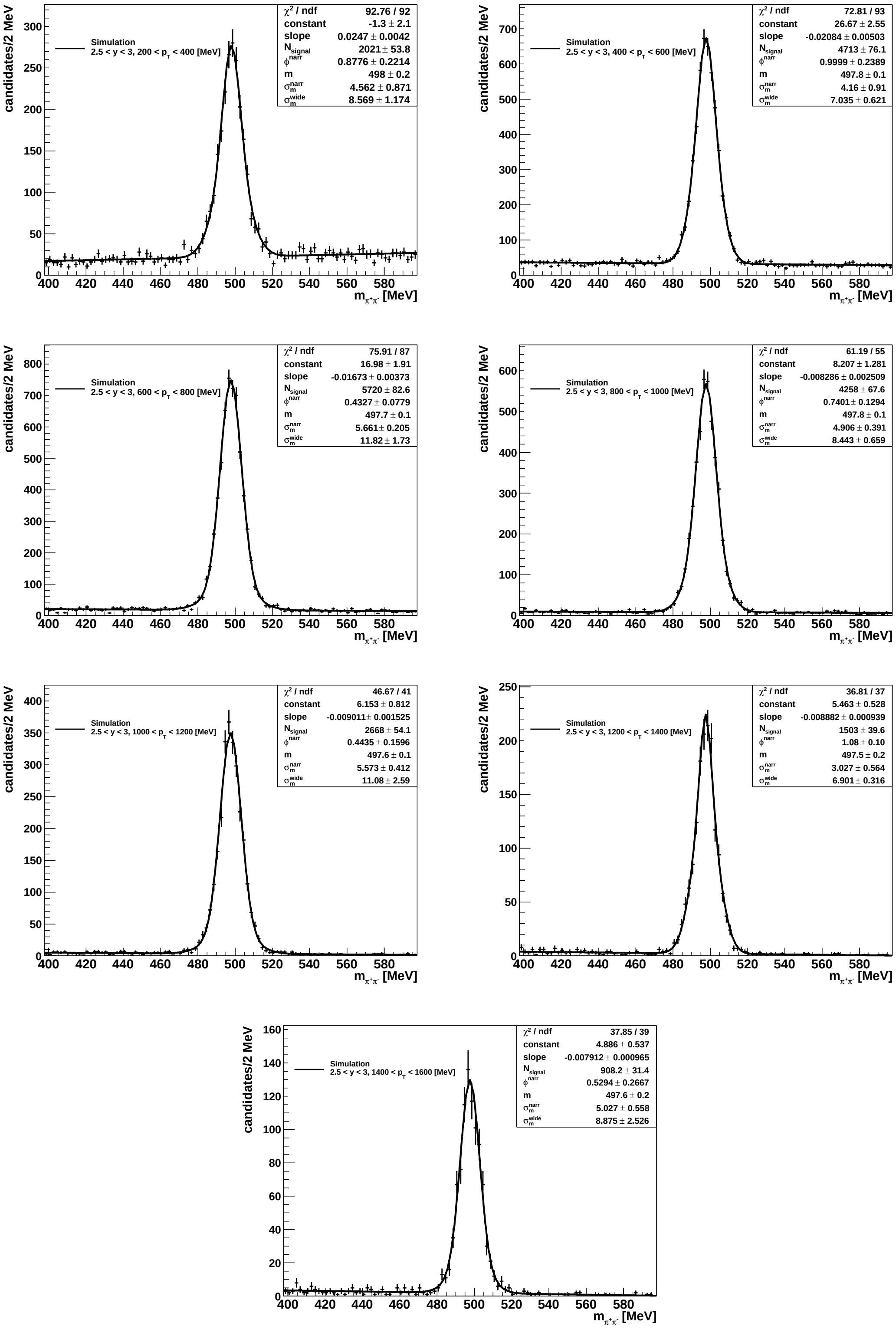


Figure 6.15: *Fits to the K_s candidates mass distribution in simulated data for different p_T bins with $2.5 < y < 3.0$.*

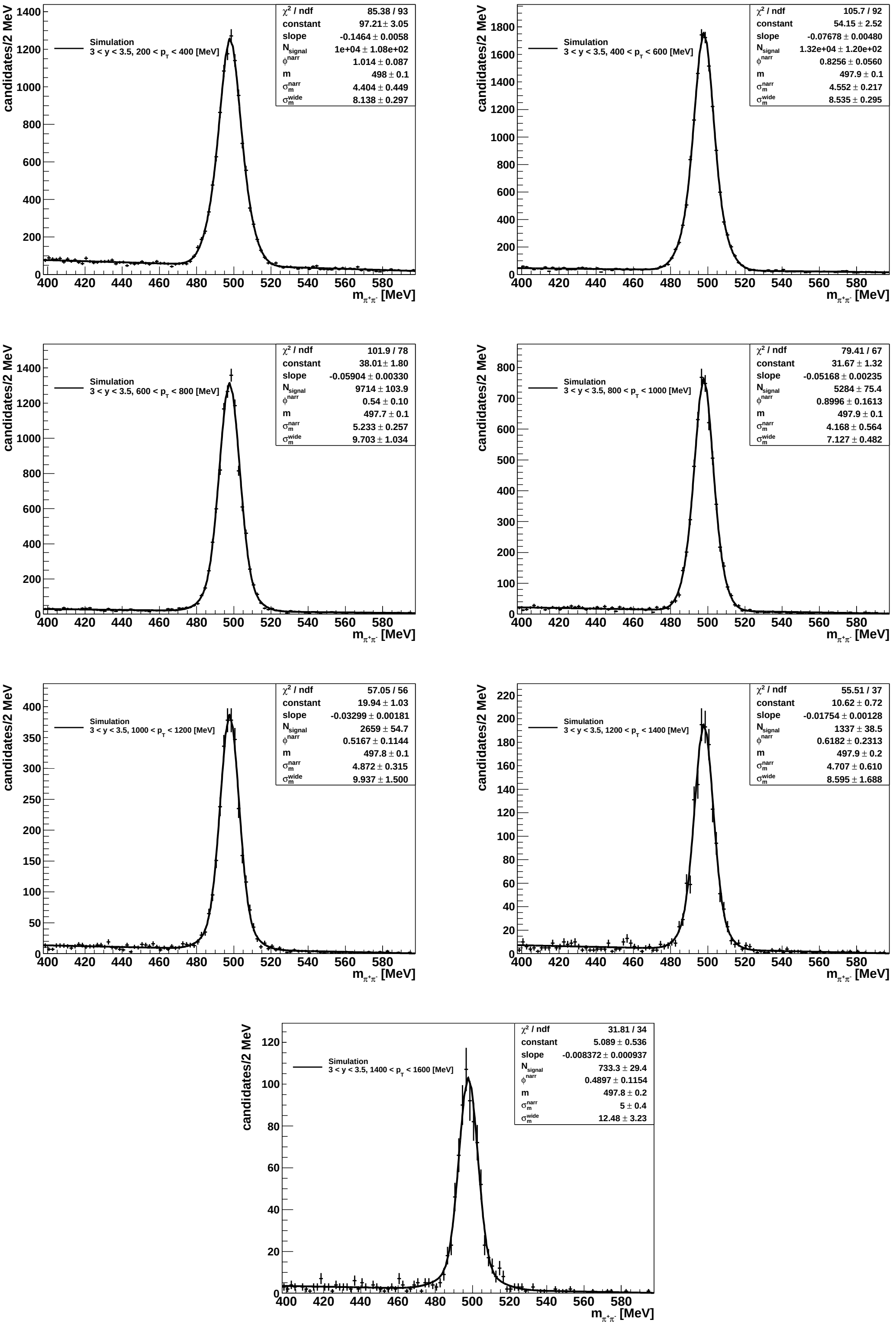


Figure 6.16: *Fits to the* K_s *candidates mass distribution in simulated data for different* p_T *bins with* $3.0 < y < 3.5$.

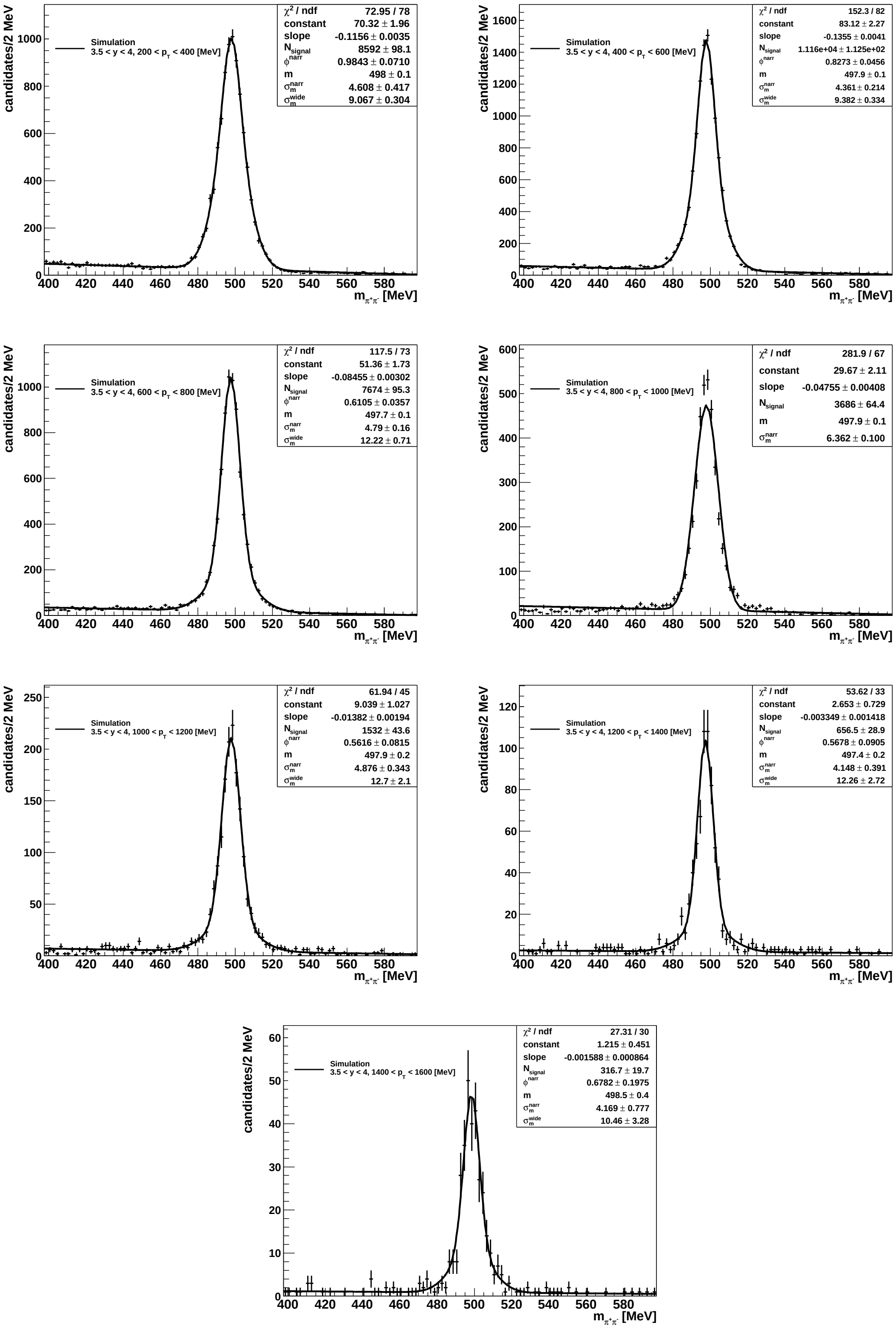


Figure 6.17: *Fits to the K_s candidates mass distribution in simulated data for different p_T bins with $3.5 < y < 4.0$.*

6.6 Systematic studies

The most important part of the analysis is to evaluate the systematics. Therefore a numerous of systematic studies are presented within this section to test the extraction of the signal yield in data and the determination of the reconstruction efficiency.

6.6.1 Stability of the fit model

The determination of the efficiency and the measurement of the yield depend on the fit model. Therefore systematic studies were performed to check if an appropriate fit model for the signal and the background was chosen.

Background model

Changing the background model could also result in a change of the yield on data and a change in the reconstruction efficiency. A linear function was chosen to describe the background. As a variation also an exponential function is considered. To determine the impact of an exponential background, the mass distribution was fitted with a double Gaussian and an exponential background. The shape of the signal peak (σ_{narr}, σ_{wide} and ϕ^{narr}) was fixed to the one obtained from the nominal fit. Only the parameters of the background and the total yield N were free parameters in the fit. The differences in yield in data are shown in Table 6.8 and the change of the reconstruction efficiency can be seen in Table 6.9. The differences between the two background models are only minor for the yield extraction and for the efficiency determination.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 – 400	1.4	1.9	10.0
400 – 600	1.2	6.8	11.4
600 – 800	2.2	6.1	2.4
800 – 1000	4.4	3.7	10.6
1000 – 1200	0.3	5.0	3.5
1200 – 1400	0.4	3.6	2.2
1400 – 1600	1.1	1.0	0.4

Table 6.8: *Difference in K_s signal yield in data assuming an exponential background model with respect to the default fit.*

Signal model

As seen before in Figure 6.12 a double Gaussian as a description for the signal seems to be an obvious choice. But in the bins with very low statistics sometimes a single Gaussian had to be chosen to describe the signal peak as described

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
$200 - 400$	< 0.1	< 0.1	0.1
$400 - 600$	< 0.1	< 0.1	< 0.1
$600 - 800$	< 0.1	0.1	< 0.1
$800 - 1000$	< 0.1	0.1	0.3
$1000 - 1200$	< 0.1	0.1	0.1
$1200 - 1400$	0.1	< 0.1	0.1
$1400 - 1600$	0.1	0.1	0.1

Table 6.9: *Difference in reconstruction efficiency (in percent) for K_s in Monte Carlo assuming an exponential background model with respect to the default fit.*

in Chapter 6.2.3. The second definition of efficiency ϵ_2 (Formula 6.11) in the simulation is based on counting the number of truthmatched particles (which is referred as *counting* in the following). This can be exploited to cross-check if the fit model is right as the number one wants to extract from the fit is known. However, two additions to the definition of the numerator have to be made. This is referred as *counting* $++$ in the following.

First of all the fit includes non-prompt particles because they cannot be distinguished in the mass fit from prompt particles. Thus the non-prompt truthmatched particles are added to the counted yield. About 0.6 % of the reconstructed K_s are non-prompt. Reconstructed K_s candidates that are not associated to a Monte Carlo particle are the second addition.

In Figure 6.18 the K_s mass distribution of associated (left) and non-associated (right) K_s candidates is shown. A clear peak which contains about 3 % of the reconstructed K_s is seen in the right plot. These are particles that did not fulfil the criteria to be associated to a Monte Carlo particle. A K_s candidate is associated to a Monte Carlo particle when the daughter tracks of the K_s candidate are truthmatched to the generated pions. The definition of the association of a reconstructed track to a Monte Carlo particle was given in Chapter 2.3. More than 70 % of the hits on the track are required to come from the same Monte Carlo particle. If e.g. only 65 % of the hits come from the same Monte Carlo particle, the track is not associated, although the track might still have enough information to reconstruct a K_s meson. Therefore an additional matching was performed. For every non-truthmatched K_s candidate in the mass window the momentum and the rapidity are compared to the momentum and rapidity of generated Monte Carlo particles in the same event. The distributions are shown for associated and non-associated particles in Figure 6.19. Deduced from these plots it was decided that if $\Delta p/p < 0.1$ and $\Delta y < 0.04$ of a reconstructed and a generated K_s, the K_s candidate is counted as successfully reconstructed and is thus contributing to the numerator of the efficiency.

In Chapter 6.2.3 it was explained in which cases a single or a double Gaussian

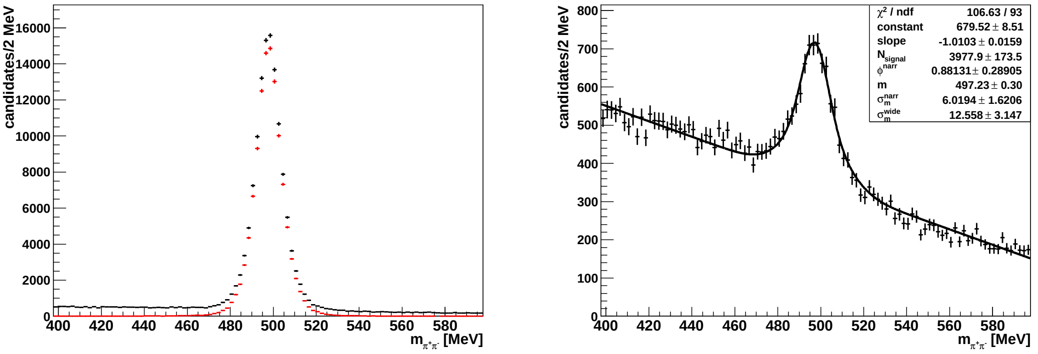


Figure 6.18: *K_s candidates in Monte Carlo (left). Black are all reconstructed candidates and red are all candidates which are associated to MC truth. K_s candidates in Monte Carlo which are not associated to MC truth (right); a clear peak is seen here as well.*

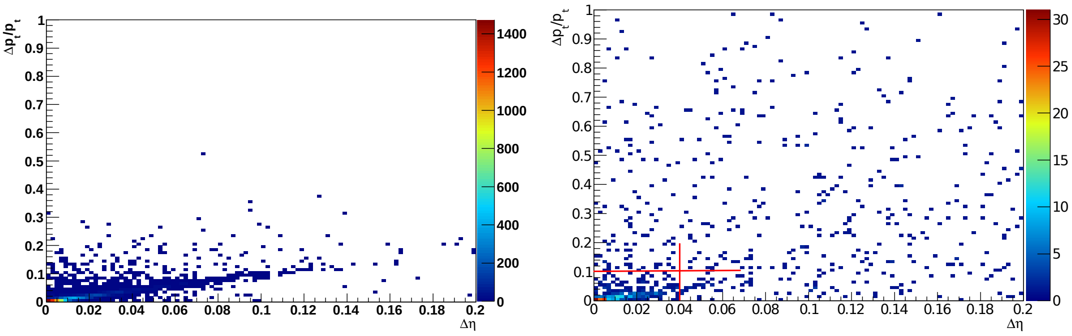


Figure 6.19: *$\Delta p/p$ and $\Delta\eta$ distributions between reconstructed and generated K_s candidates. On the left side truthmatched candidates are plotted, on the right side not-truthmatched candidates are compared to generated K_s in the event. The red lines indicate which candidates are counted as reconstructed.*

was chosen to describe the signal peak. Table 6.10 shows the determined yield by a single Gaussian and a double Gaussian in the cases a double Gaussian was chosen together with the difference to *counting* ++. The difference between all three methods gives a limit of the systematic uncertainties. The number of the single Gaussian is always lower than the number of the double Gaussian. Sometimes the fit is higher than the counted number and sometimes lower. No systematic pattern can be seen. Nevertheless, the difference of the chosen fit model to *counting*++ is taken as a systematic uncertainty both for the determination of the yield on data and the efficiency determination on the simulation. The systematic uncertainty is evaluated in each bin of p_T and y.

p_T[MeV]&y	counting	$count++$	single Gaussian	double Gaussian	Δ	$\frac{\Delta}{count++}$
200 – 400; 2.5–3.0	1900	2002	1984 ± 51	2021 ± 54	19	0.9 %
400 – 600; 2.5–3.0	4592	4845	4678 ± 74	4713 ± 76	-132	-2.7 %
600 – 800; 2.5–3.0	5410	5691	5605 ± 79	5720 ± 83	29	0.5 %
800 –1000; 2.5–3.0	4076	4285	4232 ± 67	4258 ± 68	-27	-0.6 %
1000 –1200; 2.5–3.0	2569	2689	2632 ± 53	2668 ± 54	-21	-0.8 %
1200 –1400; 2.5–3.0	1457	1534	1503 ± 40	1503 ± 40	-31	-2.0 %
1400 –1600; 2.5–3.0	886	936	901 ± 31	908 ± 31	-28	-3.0 %
200 – 400; 3.0–3.5	9787	10223	9928 ± 107	10000 ± 108	-223	-2.2 %
400 – 600; 3.0–3.5	12898	13434	13109 ± 120	13201 ± 120	-233	-1.7 %
600 – 800; 3.0–3.5	9552	9922	9594 ± 101	9714 ± 104	-208	-2.1 %
800 –1000; 3.0–3.5	5232	5442	5259 ± 75	5284 ± 75	-158	-2.9 %
1000 –1200; 2.5–3.0	2567	2678	2608 ± 53	2659 ± 55	-19	-0.7 %
1200 –1400; 2.5–3.0	1278	1331	1319 ± 38	1337 ± 39	6	0.4 %
1400 –1600; 2.5–3.0	678	704	713 ± 28	733 ± 29	29	4.2%
200 – 400; 3.5–4.0	8564	8886	8541 ± 98	8592 ± 98	-294	-3.3 %
400 – 600; 3.5–4.0	11031	11451	11028 ± 111	11156 ± 112	-295	-2.6 %
600 – 800; 3.5–4.0	7249	7504	7418 ± 92	7674 ± 95	170	2.3 %
800 –1000; 3.5–4.0	3549	3672	3686 ± 64	–	14	0.4 %
1000 –1200; 2.5–3.0	1429	1483	1473 ± 41	1532 ± 44	49	3.3 %
1200 –1400; 2.5–3.0	627	647	628 ± 27	657 ± 29	10	1.5 %
1400 –1600; 2.5–3.0	297	309	311 ± 19	317 ± 20	8	2.5 %

Table 6.10: *Different methods to determine the efficiency numerator, for detailed explanation see the text (count ++ = counting ++).*

Sideband subtraction

Furthermore the estimated yield in data and simulation when using sideband subtraction with a linear background is quoted, cf. Table 6.11. The difference to the yield obtained on data with a fit is within the error of the measurement. Thus no systematic uncertainty is added.

Fixed shape

To test the stability of the fit on data also another approach was studied. The idea is to take the shape (σ_{narr}, σ_{wide} and ϕ^{narr}) of the signal peak from the simulation. This means that first the fit is performed in every bin in the simulation. Then the shape is taken and used in the fits of the data. To take the different resolutions into account a global scaling factor is calculated. This scaling factor is the fraction of the weighted widths of the two mass distributions with full statistics, cf. Table 6.5:

$$s = \frac{\sigma_{all,\ data}}{\sigma_{all,\ sim}} = \frac{9.2\,\mathrm{MeV}}{7.2\,\mathrm{MeV}} = 1.28. \tag{6.13}$$

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 – 400	66 ± 16	261 ± 28	285 ± 24
400 – 600	147 ± 20	408 ± 28	382 ± 25
600 – 800	210 ± 20	361 ± 24	346 ± 23
800 –1000	178 ± 17	204 ± 19	229 ± 17
1000 –1200	108 ± 12	208 ± 16	120 ± 12
1200 –1400	102 ± 11	92 ± 11	42 ± 8
1400 –1600	59 ± 9	57 ± 9	25 ± 6

Table 6.11: *Number of signal candidates in data in p_T and y bins when using sideband subtraction to extract the signal yield.*

Then this scaling factor is applied to the width of every fit on data. In the case of a double Gaussian both widths are multiplied with the scaling factor:

$$\sigma_{p_T,y,\ data} = s\ \sigma_{p_T,y,\ sim}. \tag{6.14}$$

The parameters in the fit of the data are fixed to these values. Thus the only free parameter of the signal model is the yield. The background model parameters are not fixed as there is much more background in the data. Table 6.12 shows the absolute difference to the default fit and Table 6.13 the relative deviation $\Delta N/\sigma_N$. σ_N is the error of the default fit. The corresponding histograms can be found in Appendix B.1. The two methods agree well within the statistic uncertainties and no systematic pattern is visible. Thus no systematic uncertainty to the yield measurement is assigned related to this systematic study.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 – 400	-9.9	-7.7	-2.4
400 – 600	-2.5	8.2	7.8
600 – 800	-1.7	-0.5	-3.2
800 – 1000	-2.7	-4.0	5.8
1000 – 1200	-1.8	-7.6	-5.3
1200 – 1400	-1.5	0.8	0.3
1400 – 1600	-0.5	-3.6	-5.5

Table 6.12: *Absolute difference in K_s signal yield in data fixing the fit shape from simulation with respect to the default fit.*

6.6.2 Stability of selection cuts

In Chapter 6.5.1 the distributions of the selection variables were shown for K_s candidates. The histograms are sideband subtracted and the preselection mentioned in Table 6.2 is already applied. This is necessary to reject enough background to have a significant signal peak, cf. Figure 6.20.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
$200 - 400$	-1.0	0.4	-0.1
$400 - 600$	-0.2	<0.1	0.4
$600 - 800$	-0.1	<0.1	-0.2
$800 - 1000$	-0.2	-0.3	0.4
$1000 - 1200$	-0.1	-0.5	-0.5
$1200 - 1400$	-0.1	0.1	<0.1
$1400 - 1600$	-0.4	-0.4	-1.0

Table 6.13: *Relative difference $\Delta N/\sigma_N$ in K_s signal yield in data fixing the fit shape from simulated data with respect to the default fit. σ_N is the error of the default fit.*

Most of the distributions show a discrepancy. However, in the case of the momentum p and the transverse momentum p_T of the daughter particles, the reduced track χ^2/ndf and the vertex χ^2 of the K_s the selection does not cut hard into the distribution. Only a few signal K_s candidates do not pass these cuts on collision and simulated data. This is different for the impact parameter of the pions and the pointing angle, the lifetime and the z value of the point of closest approach of the K_s candidate. Here the selection cut differs between data and simulation. These are all variables where the reconstructed track of the pions has to be propagated through the B-field. Thus an error in the track reconstruction because of misalignment in the tracking detectors is transported through the B-field and the uncertainty increases accordingly. The distance between the TT and the nominal interaction point is more than 2 m. Also the B-field map could be slightly different in reality. The propagation of the tracks is done with the same B-field map which was used for the simulation. A detailed B-field calibration is still ongoing. A combination of these effects results in the above mentioned differences in the distributions of the selection variables and also the worse mass resolution in data compared to the simulation.

To estimate the efficiency of the selection, the signal yield after the preselection is compared with the signal yield after the full selection. In the simulated data the efficiency of the preselection can be determined by using the information of the generator. The preselection efficiency of associated K_s candidates is 89.6 ± 0.6 %. Furthermore all cuts applied in the preselection cut only into the tails of the distributions. Thus the assumption is made that there should be no significant difference between data and simulation in the efficiency of the preselection. Therefore the efficiency of the full selection is determined relative to the preselection.

The invariant mass distributions after the preselection for data and simulation are shown in Figure 6.20. The background is fitted with a quadratic function and the signal peak with a double Gaussian. The fit is performed in two steps. First the background is fitted in the background region. Then the parameters of the

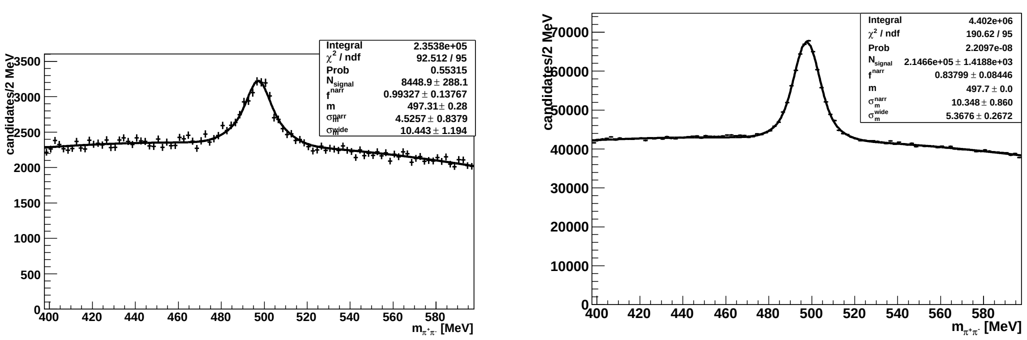


Figure 6.20: *K_s candidate mass distribution for collision data (left) and simulated data (right) after the preselection.*

background are fixed and the signal peak is fitted. Furthermore the signal yield was also determined by sideband subtraction where the fit of the signal peak is not necessary.

The relative difference between data and simulation is $\Delta\epsilon = \frac{\epsilon_{data}-\epsilon_{sim}}{\epsilon_{sim}} = (7 \pm 4)$ %, cf. Table 6.14. As the measurement of the yield is done in bins of p_T and y of the K_s candidates, it was checked that the difference of the selection efficiency does not depend on p_T and y. Therefore two bins in p_T and y respectively, cf. Table 6.14, with similar statistics were chosen. Within the errors, the difference between data and simulation is compatible with the deviation obtained on the whole sample. The determination of the signal yield and the selection efficiency by using sideband subtraction is shown in Table 6.15. The results are consistent with the previous mentioned results when using a fit.

The whole measurement is limited by the statistics of the data sample and the signal yield after the preselection. Nevertheless, a difference in the selection efficiency is seen. Thus the signal yield on data is corrected by 7 % and the statistical uncertainty of ±4 % is assigned as a systematic uncertainty on this correction.

Bin	$\epsilon_{fit.data}$	$\epsilon_{fit.sim}$	$\Delta\epsilon_{fit}$
whole sample	56 ± 2	53 ± 1	7 ± 4
$2.5 < y < 3.4$, $200 < p_T < 1600$	60 ± 7	53 ± 1	14 ± 13
$3.4 < y < 4.0$, $200 < p_T < 1600$	56 ± 7	58 ± 1	-4 ± 12
$2.5 < y < 4.0$, $200 < p_T < 600$	57 ± 4	52 ± 1	9 ± 7
$2.5 < y < 4.0$, $600 < p_T < 1600$	63 ± 3	58 ± 1	9 ± 5

Table 6.14: *Efficiency in percent of the selection cuts with respect to preselected K_s signal candidates on data and simulation. The efficiencies are obtained by fitting a double Gaussian with a quadratic background. Bin boundaries have been chosen to ensure an equal amount of K_s candidates in each bin.*

Bin	$\epsilon_{SB.data}$	$\epsilon_{SB.sim}$	$\Delta\epsilon_{SB}$
whole sample	58 ± 5	52 ± 1	11 ± 9
$2.5 < y < 3.4$, $200 < p_T < 1600$	64 ± 9	53 ± 1	22 ± 17
$3.4 < y < 4.0$, $200 < p_T < 1600$	55 ± 7	57 ± 1	-4 ± 12
$2.5 < y < 4.0$, $200 < p_T < 600$	57 ± 9	53 ± 1	7 ± 16
$2.5 < y < 4.0$, $600 < p_T < 1600$	63 ± 7	57 ± 1	11 ± 12

Table 6.15: *Efficiency in percent of the selection cuts with respect to preselected K_s signal candidates on data and simulation. The efficiencies are obtained by sideband subtraction with a quadratic background. Bin boundaries have been chosen to ensure an equal amount of K_s candidates in each bin.*

6.6.3 Variations of the simulated sample

As mentioned in Chapter 6.1 the simulated sample includes corrections for inefficiencies in the pattern recognition. The K_s reconstruction efficiency depends strongly on these corrections as the total yield is corrected by 10.3 % compared to the sample without corrections. In the sample where the corrections are underestimated the total yield is corrected by 4.9 % and in the overestimated sample 18.1 % compared to the nominal sample. Compared to the corrected sample it is +5.9 % and -8.8 %. The procedure of this correction and systematic studies for this are described in detail in [28]. In Table 6.16 and 6.17 the reconstruction efficiency in the different bins is given for the samples which under- and overestimate the corrections because of misalignment and detector inefficiencies. Table 6.18 summarizes the maximal differences to the corrected sample in each bin. The absolute value of the deviation is taken as the systematic uncertainty for the reconstruction efficiency per bin. In some bins the relative difference is up to 10 %.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	1.3 ± 0.1	7.2 ± 0.1	7.7 ± 0.1
400 − 600	3.6 ± 0.1	11.8 ± 0.1	12.4 ± 0.1
600 − 800	7.4 ± 0.1	15.2 ± 0.2	15.3 ± 0.2
800 −1000	11.1 ± 0.2	17.3 ± 0.2	16.0 ± 0.3
1000 −1200	14.1 ± 0.3	18.7 ± 0.4	15.5 ± 0.4
1200 −1400	15.9 ± 0.4	19.2 ± 0.5	14.0 ± 0.6
1400 −1600	17.6 ± 0.6	18.9 ± 0.7	12.4 ± 0.7

Table 6.16: *Reconstruction efficiencies (in percent) derived from the simulated sample which overestimates the corrections.*

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	1.2 ± 0.1	6.6 ± 0.1	6.4 ± 0.1
400 − 600	3.3 ± 0.1	10.5 ± 0.1	10.4 ± 0.1
600 − 800	6.7 ± 0.1	13.4 ± 0.2	12.8 ± 0.2
800 −1000	10.0 ± 0.2	15.2 ± 0.2	13.3 ± 0.3
1000 −1200	12.7 ± 0.3	16.4 ± 0.3	12.7 ± 0.4
1200 −1400	14.4 ± 0.4	17.3 ± 0.5	11.8 ± 0.5
1400 −1600	15.8 ± 0.6	16.2 ± 0.7	10.3 ± 0.7

Table 6.17: *Reconstruction efficiencies (in percent) derived from the sample which underestimates the corrections.*

p_T[MeV]/y	3 - 3.5	3.5 - 4	4 - 4.5
200 − 400	0.1	0.4	0.7
400 − 600	0.2	0.8	1.1
600 − 800	0.4	1.0	1.5
800 −1000	0.7	1.2	1.7
1000 −1200	0.9	1.3	1.7
1200 −1400	0.9	1.1	1.2
1400 −1600	1.0	1.9	1.6

Table 6.18: *Absolute systematic uncertainties to the reconstruction efficiency (in percent) assigned to the remaining data and simulation discrepancy.*

6.6.4 Variations of the reconstruction efficiency within a bin

To determine the corrected K_s yield in data the measured yield is divided by the reconstruction efficiency obtained from the simulation. The obtained reconstruction efficiency is the averaged efficiency over the y and p_T distributions within a bin. This could affect the measurement if the distributions of y and p_T are different in collision and simulated data and if there are large variations in the reconstruction efficiency within a bin.

To estimate this effect, a finer binning for the determination of the reconstruction efficiency was chosen. Then every K_s candidate in data gets a weight according to this reconstruction efficiency and the corrected K_s yield is determined on the weighted sample.

Table 6.19 summarizes the reconstruction efficiency in the finer bins. Especially in the smallest y, p_T bins there are large variations within a bin. The red highlighted bin shows a variation of a factor of more than 40.

The efficiency corrected yield is then determined with the normal procedure where the obtained yield (Table 6.11) is divided by the reconstruction efficiency (Table 6.7) and it is determined on the weighted sample. As it turned out, the fit model is not optimal to determine the yield on the weighted sample. Thus the

p_T[MeV]/y	2.5 - 2.75	2.75 - 3.0	3.0 - 3.25	3.25 - 3.5	3.5 - 3.75	3.75 - 4.0
200 − 300	0.07 ± 0.02	1.67 ± 0.08	4.70 ± 0.12	6.90 ± 0.15	6.89 ± 0.16	4.77 ± 0.14
300 − 400	0.35 ± 0.04	2.86 ± 0.09	6.80 ± 0.15	9.60 ± 0.18	9.38 ± 0.18	7.29 ± 0.17
400 − 500	0.89 ± 0.06	4.68 ± 0.12	9.05 ± 0.17	11.99 ± 0.21	11.45 ± 0.21	9.83 ± 0.21
500 − 600	2.12 ± 0.09	6.60 ± 0.16	11.36 ± 0.21	13.44 ± 0.24	13.32 ± 0.26	11.80 ± 0.26
600 − 700	3.60 ± 0.13	9.09 ± 0.22	13.19 ± 0.27	14.49 ± 0.29	14.40 ± 0.31	13.34 ± 0.33
700 − 800	5.54 ± 0.19	11.16 ± 0.27	14.89 ± 0.33	15.64 ± 0.35	15.89 ± 0.40	13.26 ± 0.39
800 − 900	7.81 ± 0.27	12.45 ± 0.34	15.87 ± 0.41	16.80 ± 0.45	17.07 ± 0.50	13.51 ± 0.49
900 −1000	8.78 ± 0.33	14.71 ± 0.45	16.57 ± 0.51	16.79 ± 0.55	16.43 ± 0.58	13.68 ± 0.58
1000 −1100	10.53 ± 0.43	15.99 ± 0.55	16.49 ± 0.61	17.12 ± 0.67	15.93 ± 0.74	13.81 ± 0.74
1100 −1200	12.38 ± 0.55	16.58 ± 0.68	18.47 ± 0.77	19.70 ± 0.88	14.62 ± 0.87	11.53 ± 0.82
1200 −1300	13.16 ± 0.67	17.06 ± 0.60	17.51 ± 0.91	18.65 ± 1.04	12.73 ± 0.95	12.10 ± 0.98
1300 −1400	13.79 ± 0.80	18.58 ± 1.01	16.66 ± 1.03	20.89 ± 1.29	13.56 ± 1.15	11.41 ± 1.16
1400 −1500	14.26 ± 0.98	18.42 ± 1.21	18.99 ± 1.29	16.93 ± 1.35	11.52 ± 1.22	10.36 ± 1.32
1500 −1600	15.85 ± 1.16	19.52 ± 1.39	18.41 ± 1.47	16.11 ± 1.48	12.22 ± 1.46	11.17 ± 1.58

Table 6.19: *Reconstruction efficiencies (in percent) derived from Monte Carlo in finer bins.*

yield is taken in both cases from sideband subtraction with a linear background model. The corresponding histograms can be found in Appendix B.2. Table 6.20 shows the absolute differences of the two procedures. Especially in the bins with the highest variations in the reconstruction efficiency the difference is large. In the other bins the differences are rather small. The absolute difference is added as a bin-wise systematic uncertainty to the efficiency corrected yield.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	-2373	109	-89
400 − 600	-410	45	35
600 − 800	-113	14	-13
800 − 1000	0	6	37
1000 − 1200	18	11	-9
1200 − 1400	-5	3	-15
1400 − 1600	-4	-5	-12

Table 6.20: *Difference in efficiency corrected K_s signal yield in data when applying a weight to each event according to its efficiency in fine bins with respect to the standard procedure. Signal yields have been extracted by sideband subtraction.*

6.6.5 Beam-gas subtraction

In Figure 6.3 the number of K_s found in events where a proton collided with the residual gas in the beam pipe was shown. These K_s candidates are statistical subtracted from the K_s candidates in beam-beam events to get a clean sample of

K_s candidates in proton-proton collisions. The factor $\beta = 0.91 \pm 0.01$ in Formula 6.3 has a relative error of about 2 %. As only about 50 K_s candidates are seen in the beam-empty events, the error of the beam-gas subtraction is negligible. Especially because the effect is added in quadrature to all other effects. Thus no systematic uncertainties are assigned.

6.6.6 Binning effects

Binning effects due to the finite resolution of the detector have already been discussed in Chapter 6.5.2 and the contribution to the systematic uncertainty is summarized in Table 6.21. The effects are small compared to other systematic uncertainties. The systematic uncertainty is added per bin.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 – 400	< 0.1	< 0.1	< 0.1
400 – 600	< 0.1	< 0.1	0.1
600 – 800	< 0.1	< 0.1	< 0.1
800 –1000	< 0.1	< 0.1	< 0.1
1000 –1200	< 0.1	0.1	< 0.1
1200 –1400	< 0.1	< 0.1	0.2
1400 –1600	0.1	0.2	0.2

Table 6.21: *Difference in reconstruction efficiencies (in percent) for using the reconstructed p_T and y or the generated p_T and y for the K_s candidates which enter the numerator of the efficiency correction. As these numbers are highly correlated, no uncertainties are given. The observed difference will be added to the systematic uncertainties.*

6.6.7 Correction for non-prompt Ks

Although the selection clearly prefers prompt K_s, about 0.6 % of the K_s are non-prompt in the simulated sample. The definition of efficiency takes this already into account because non-prompt K_s appear in the numerator of the efficiency and in the determination of the measured yield. However, the simulation might not reproduce the real fraction of non-prompt K_s. One reason could be that there is more material in the detector in reality and more K_s are produced in material interactions. Therefore a systematic uncertainty of ±0.5 % is assigned to the final K_s yield. This corresponds to a relative error of 83 % on the non-prompt K_s contribution.

6.6.8 Effects from hard scattering and material interactions

The amount of material could be different in reality and simulation. Particles traversing more material are scattered more often and thus the reconstruction efficiency is lowered.

Pions which do not suffer a material interaction have a reconstruction efficiency of about 82.5 %. The reconstruction efficiency of K_s Monte Carlo particles whose both daughters were not hard scattered is 79.1 %.

In the simulated sample 10 % of the reconstructible K_s daughters undergo a significant material interaction. Thus about 19 % of the K_s have at least one daughter which experiences a significant material interaction. The reconstruction efficiency of these pions is 79.1 % and the reconstruction efficiency of these K_s mesons is 65.3 %.

If the material budget is changed by ±10 % this would result in 10 % more or less hard scattered pions. Therefore the fraction of K_s mesons with at least one hard scattered daughter is 23.5 % resp. 14.4 %. If the above mentioned reconstruction efficiencies are assumed this results in a relative change of efficiency of less than ±0.2 %. This number is an added as a systematic uncertainty to the efficiency determination.

6.6.9 Summary of the systematic studies

A short summary of the systematic uncertainties and corrections to the K_s signal yield and the determination of the reconstruction efficiency is given. The following systematic uncertainties are assigned to the determination of the reconstruction efficiency on simulated data:

- **Tracking induced corrections:** Systematic uncertainties are added bin-wise. The relative uncertainty is up to **10 %**.
- **Stability of selection cuts:** The reconstruction efficiency is corrected by 7 % because of the difference in the selection efficiency on data and in the simulation and a systematic uncertainty of relative **4 %** is added because of the correction.
- **Fit model:** Systematic uncertainties are added bin-wise. The relative uncertainty is up to **3 %**.
- **Variation of the reconstruction efficiency within in a bin:** Systematic uncertainties are added bin-wise after the yield has been corrected by the reconstruction efficiency. The uncertainty is large in the first bin (relative **50 %**). In the other bins it is on the **per cent level**.
- **Binning effects:** Systematic uncertainties are added bin-wise but **negligible**.

- **Material interactions:** An overall uncertainty is added to every bin. However it is **negligible**.

Table 6.22 shows the final reconstruction efficiency together with statistical and systematic uncertainties.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	1.3 ± 0.1 ± 0.1	7.5 ± 0.1 ± 0.5	7.6 ± 0.1 ± 0.8
400 − 600	3.6 ± 0.1 ± 0.3	12.1 ± 0.1 ± 0.9	12.3 ± 0.1 ± 1.3
600 − 800	7.6 ± 0.1 ± 0.5	15.4 ± 0.2 ± 1.3	15.3 ± 0.2 ± 1.8
800 −1000	11.4 ± 0.2 ± 0.9	17.5 ± 0.2 ± 1.6	16.1 ± 0.3 ± 1.9
1000 −1200	14.6 ± 0.3 ± 1.1	18.9 ± 0.4 ± 1.6	15.4 ± 0.4 ± 2.0
1200 −1400	16.4 ± 0.4 ± 1.2	19.4 ± 0.5 ± 1.4	13.8 ± 0.6 ± 1.4
1400 −1600	18.0 ± 0.6 ± 1.4	19.4 ± 0.7 ± 2.3	12.7 ± 0.7 ± 1.8

Table 6.22: *Final result on the reconstruction efficiency given in percent including statistical and systematical uncertainties.*

The yield extraction has two sytematic uncertainties:

- **Fit model:** The systematic uncertainty because of the fit is added per bin and is on the **per cent level**.
- **Beam-gas subtraction:** The systematic uncertainty of the beam-gas subtraction is **negligible**.

The final yield is shown in Table 6.23 together with statistical and systematic uncertainties.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	64 ± 10 ± 1	278 ± 21 ± 6	288 ± 21 ± 10
400 − 600	147 ± 15 ± 4	428 ± 24 ± 7	388 ± 21 ± 10
600 − 800	202 ± 16 ± 1	379 ± 22 ± 8	332 ± 21 ± 8
800 −1000	176 ± 15 ± 1	213 ± 16 ± 6	217 ± 17 ± 1
1000 −1200	113 ± 11 ± 1	173 ± 14 ± 1	111 ± 12 ± 4
1200 −1400	94 ± 11 ± 2	90 ± 10 ± 0	32 ± 8 ± 0
1400 −1600	56 ± 8 ± 2	64 ± 8 ± 3	20 ± 5 ± 1

Table 6.23: *Final result on the measured yields including statistical and systematical uncertainties.*

6.7 Trigger efficiency and luminosity

The following things were not part of the author's work. But to make a measurement of the cross-section for K_s production in proton-proton collisions they are needed, cf. Formula 6.2.

- **Trigger efficiency:** The first step in the reconstruction of a K_s candidate is to record the event. In the 2009 run only the L0 trigger was used to select events. The interaction rate was low enough that a further reduction of the events was not needed. The efficiency has been determined in studies on simulated data but cross-checks with the data were made. They are described in [28]. The efficiency to trigger an event with a K_s in it is given in Table 6.24 for the different p_t, y bins. The first uncertainty is the statistical and the second the systematic uncertainty.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 − 400	96.5 ± 0.1 ± 2.7	97.0 ± 0.2 ± 2.7	97.6 ± 0.5 ± 2.8
400 − 600	97.2 ± 0.1 ± 2.8	97.5 ± 0.1 ± 2.6	97.2 ± 0.4 ± 2.6
600 − 800	97.5 ± 0.1 ± 2.7	97.6 ± 0.2 ± 2.6	98.0 ± 0.1 ± 2.6
800 −1000	97.7 ± 0.2 ± 2.6	97.7 ± 0.2 ± 2.6	98.4 ± 0.4 ± 2.6
1000 −1200	99.3 ± 0.1 ± 2.7	98.8 ± 0.2 ± 2.5	98.3 ± 0.6 ± 2.7
1200 −1400	99.0 ± 0.3 ± 2.5	97.6 ± 2.4 ± 2.5	98.8 ± 1.3 ± 2.5
1400 −1600	99.1 ± 0.3 ± 2.5	98.7 ± 0.3 ± 2.5	99.0 ± 1.0 ± 2.5

Table 6.24: *Final results on the trigger efficiencies given in percent including statistical and systematic uncertainties.*

- **Luminosity:** The last missing piece to determine the cross-section is the luminosity. The integrated luminosity for the used runs was determined to be $\mathcal{L}_{int} = 6.8 \pm 1.0\ \mu b^{-1}$. The uncertainty combines statistical and systematic uncertainty. The extraction of the luminosity is described in [29].

6.8 K_s production cross-section in proton-proton collisions at $\sqrt{s} = 900\,\mathrm{GeV}$

Combining the trigger efficiency, reconstruction efficiency and the measured yield leads to the final efficiency corrected yield ($N_{K_s,y,p_T,corrected}$) which is given by

$$N_{K_s,y,p_T,corrected} = N_{K_s \to \pi^+\pi^-,y,p_T} \times \epsilon^{-1}_{reco,y,p_T} \times \epsilon^{-1}_{trigger,y,p_T}. \tag{6.15}$$

The statistical uncertainty of the trigger efficiency and the reconstruction efficiency is only used once as they are highly correlated because the numerator of the reconstruction efficiency is the denominator of the trigger efficiency. The final efficiency corrected yield is given in Table 6.25. Especially in the lowest bin ($200 < p_T < 400\,\mathrm{MeV}$ and $2.5 < y < 3.0$) the measurement is dominated by the systematic uncertainty which is about 50 % of the measured yield. In most of the other bins statistical and systematic uncertainty are of the same order.

p_T[MeV]/y	2.5 - 3.0	3.0 - 3.5	3.5 - 4.0
200 – 400	5102 ± 888 ± 2496	3821 ± 293 ± 309	3883 ± 288 ± 454
400 – 600	4200 ± 444 ± 573	3628 ± 206 ± 296	3245 ± 178 ± 365
600 – 800	2726 ± 219 ± 227	2521 ± 150 ± 230	2194 ± 142 ± 284
800 –1000	1580 ± 138 ± 132	1246 ± 95 ± 124	1370 ± 108 ± 178
1000 –1200	779 ± 78 ± 65	926 ± 77 ± 83	733 ± 81 ± 75
1200 –1400	579 ± 69 ± 47	475 ± 54 ± 37	235 ± 60 ± 35
1400 –1600	314 ± 46 ± 28	334 ± 43 ± 44	159 ± 41 ± 13

Table 6.25: *Final efficiency corrected K_s yields including statistical and systematic uncertainties.*

Together with the luminosity the cross-section is calculated following Formula 6.2:

$$\sigma_{K_s,y,p_T} = N_{K_s\to\pi^+\pi^-,y,p_T} \times \epsilon^{-1}_{reco,y,p_T} \times \epsilon^{-1}_{trigger,y,p_T} \times \mathcal{L}^{-1}. \tag{6.16}$$

The cross-section for the different bins of p_T and y of the K_s meson are shown in Figure 6.21. Here the uncertainty of the luminosity determination which is about 15 % dominates the total uncertainty of the measurement in most of the bins. In the figures also theoretical predictions of two Monte Carlo generator tunings are shown. The black line shows the prediction of the theoretical model that was used in the Monte Carlo generator of the LHCb software. Especially in the higher p_T bins the data lie systematically above the theoretical predictions.
Thus the measurement presented in this chapter is a valuable input for the tuning of Monte Carlo generators so that they are consistent with data.

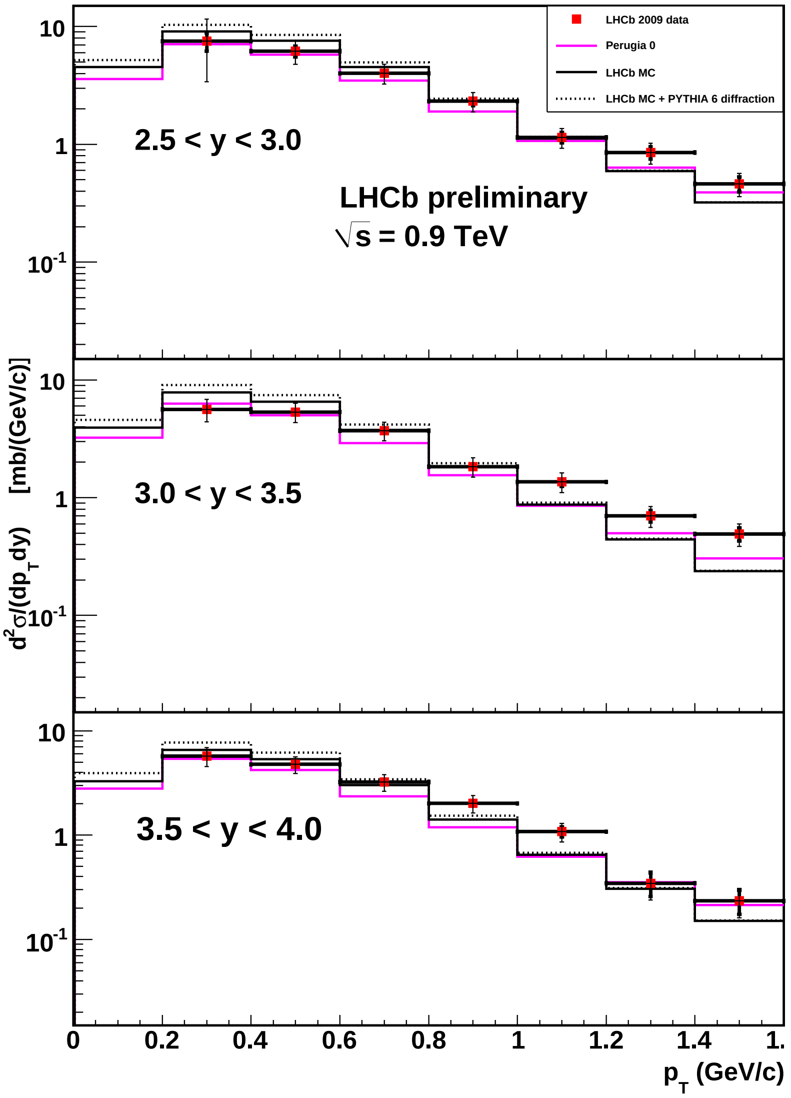


Figure 6.21: *K_s production cross-section in proton-proton collisions at $\sqrt{s} = 900\,\mathrm{GeV}$ in different bins of p_T and y. The thick error bar shows the statistical uncertainty. The thin error bars includes systematical uncertainties. The lines show the predictions of different theoretical models.*

Chapter 7

Summary and Conclusion

In this work a tracking algorithm dedicated to the reconstruction of very displaced tracks in the LHCb experiment has been studied. This algorithm is crucial for the reconstruction of long lived particles such as K_s or Λ hadrons which mainly decay outside the LHCb Vertex Locator. This special algorithm called Downstream tracking is only based on tracking detectors further down the LHCb spectrometer, thus do not require measurements in the Vertex Locator.

During the course of this work, the track reconstruction efficiency of the Downstream tracking was improved from 75.8 % to 80.3 %, while the rate of misreconstructed tracks was lowered by relative 10 %. These changes were integrated into the LHCb software and are used in the official reconstruction of LHCb data. Also an approach which reduces the runtime of the pattern recognition and the subsequent fitting by a factor of three was discussed. This will become vital once the LHCb experiment will take data with a higher average rate. The improvement of the Downstream algorithm resulted in an increase in the yield of the $B^0 \to J/\psi\ K_s$ decays in simulation by 7 %. The rate of with two Downstream tracks reconstructed K_s mesons in minimum bias events is improved by 20 %.

The Downstream tracking was successfully tested on the data taken end of 2009. Due to the special detector configuration with an open Vertex Locator used in this initial data taking period, this algorithm was the central pattern recognition strategy and thus the basis of the first physics signals observed in the LHCb experiment. The Downstream tracking was retuned for the not yet fully calibrated and aligned detector in the data taken end of 2009 which resulted in an increase of the K_s yield by additionally 2.5 %.

In the last chapter of this thesis the measurement of the K_s yield at LHCb at a center-of-mass energy of 900 GeV was presented. The analysis was performed in bins of rapidity y and transverse momentum p_T of the K_s. The signal yield was corrected by the reconstruction efficiency which was taken from studies on simulated data. A numerous of systematic studies were performed to test the stability of the extraction of the signal yield in data and the determination of

the reconstruction efficiency. Finally the corrected yield was combined with the luminosity measured on the 2009 data to quote the K_s production cross-section in proton-proton collisions at a center-of-mass energy of 900 GeV.
The analysis performed within this thesis is the first LHCb physics result which has been published.

Bibliography

[1] CERN multimedia gallery. *http://multimedia-gallery.web.cern.ch*, (2010).

[2] G. Altarelli and M.L. Mangano (editors). *Standard model physics (and more) at the LHC.* CERN-2000-004, (2000).

[3] R. Hierck. *Optimisation of the LHCb detector.* CERN-THESIS-2003-025, (2003).

[4] The LHCb Collaboration. *The LHCb Detector at the LHC.* 2008 JINST 3 S08005.

[5] ATLAS public web site. *http://www.atlas.ch/fact-sheets.html*, (2010).

[6] S.L. Glashow. *Nucl. Phys. 22* , 579 (1961).

[7] A. Salam and J.C. Ward. *Phys. Lett. 13* , 168 (1964).

[8] S. Weinberg. *Phys. Rev. Lett. 19* , 1264 (1967).

[9] LHCb Collaboration. *Roadmap for selected key measurements of LHCb.* arXiv:0912.4179v2 [hep-ex], (2010).

[10] M. Blanke, A.J. Buras, D. Guadagnoli, C. Tarantino. *Minimal Flavor Violation Waiting for Precise Measurements of Delta* M_s, $S_{\psi\phi}$, A^s_{SL}, $|V_u b|$, *gamma and* $B^0_{s,d} \rightarrow\ \mu^+ \mu$m. arXiv:hep-ph/0604057v5, (May 2006).

[11] T. Aaltonen et al[CDF collaboration]. Phys. Rev. Lett. **100** 101802, (2008).

[12] M. Beneke et al. *Exclusive radiative and electroweak* $b \rightarrow d$ *and* $b \rightarrow s$ *penguin decays at NLO.* Eur.Phys.J. C 41:173-188, (2005).

[13] CKMfitter group, J. Charles et al. *updated results and plots available at: http://ckmfitter.in2p3.fr.* Eur. Phys. J. C41, 1-131, hep-ph/0406184, (2009).

[14] E. Barberino et al. . *Heavy Flavour Averaging Group.* `http://www.slac.stanford.edu/xorg/hfag`.

[15] Amato, S., Calvi, M., Gandelman, M., Gbel, C., Mangiafave, N., Poss, S., and Rodrigues, F. Technical Report LHCb-2009-019. CERN-LHCb-2009-019, CERN, Geneva, Aug (2009).

[16] Kalman, R.E. *A new approach to linear filtering and prediction problems.* Trans. ASME J. Bas. Eng. **D82** (1960) 35.

[17] Frühwirth, R. *Application of Kalman Filtering to track and vertex fitting.* Nucl. Instrum. Meth. A **262** (1987) 444.

[18] J. Van Tilburg. *Track simulation and reconstruction in LHCb.* CERN-THESIS-2005-020, (2005).

[19] R. Van der Eijk. *Track reconstruction in the LHCb experiment.* CERN-THESIS-2002-032, (2005).

[20] Needham, M. *Clone Track Identification using the Kullback-Liebler Distance.* CERN-LHCb-2008-002, (2008).

[21] M. Schiller. *LHCb tracking Twiki page.* `https://twiki.cern.ch/twiki/view/LHCb/LHCbTrackingStrategies`, (2008).

[22] S. Hansmann-Menzemer, O. Callot. *The Forward Tracking: Algorithm and Performance Studies.* CERN-LHCb-2007-015, (2007).

[23] Callot, O., Schiller, M. *Pat Seeding: A Standalone Track Reconstruction Algorithm.* CERN-LHCb-2008-042, (2008).

[24] Forty, R., Needham, M. *Standalone Track Reconstruction in the T-Stations.* CERN-LHCb-2007-022, (2007).

[25] Rodrigues, E. *Tracking definitions.* Number LHCb-2007-006. CERN-LHCb-2007-006. Geneva, Feb (2007). revised version submitted on 2007-03-28 09:34:37.

[26] Benayoun, M. and Callot, O. *The forward tracking, an optical model method.* Number LHCb-2002-008. Geneva, Feb (2002).

[27] Callot, O. *Downstream Pattern Recognition.* Number LHCb-2007-026. CERN-LHCb-2007-026. Geneva, Mar (2007).

[28] Hansmann-Menzemer, S., Holubyev, K., Krocker, G., Linn, C., Schiller, M., Seyfert, P., Stahl, S., and Wandernoth, S. Technical Report LHCb-INT-2010-013. CERN-LHCb-INT-2010-013, CERN, Geneva, Mar (2010).

[29] V. Balagura. *Luminosity measurement at LHCb.* Recontres de Moriond QCD 2010, (2010).

[30] Callot, O. *News from the Pat packages.* LHCb tracking workshop in Heidelberg, 22-February 2007.

Appendix A

Track model and fit method in the Downstream algorithm

A.1 Track model

The pattern recognition needs a model that represents the trajectory of a particle. The Downstream tracking starts with a seed from the T-Stations. This seed is extrapolated to a point in the magnet called $\vec{x}_{magnet}$. How to get z_{magnet} was discussed in Chapter 3.1.2. x_{magnet} and y_{magnet} are obtained by a straight line extrapolation starting at the end of the T-Stations ($z = 9410\,\mathrm{mm}$), cf. Formula 3.3. The point $\vec{x}_{magnet}$ is later also used as a constraint in the fit. The trajectory of the particle in front of the magnet is described by straight lines in both planes with the parameters (x_0, tx) resp. (y_0, ty) .

$$\vec{x}_{tr}(z) = \begin{pmatrix} x_{tr}(z) \\ y_{tr}(z) \\ z \end{pmatrix} = \begin{pmatrix} x_0 + tx\ (z - z_{magnet}) \\ y_0 + ty\ (z - z_{magnet}) \\ z \end{pmatrix} \tag{A.1}$$

To compensate fringe field effects in TT, a small momentum dependent correction is added. The curvature c depends on the momentum resp. *dSlope* and was derived on simulated data [30]. In Figure A.1 the dependency is shown.

$$\begin{gathered} x_{tr,TT}(z) = x_0 + tx\ (z - z_{magnet}) + c\ (z - z_{TT})^2 \\ \text{with } c = 1.7\ 10^{-5}/\,\mathrm{mm} \cdot dSlope \text{ and } z_{TT} = 2469\,\mathrm{mm}. \end{gathered} \tag{A.2}$$

For a particle with $p = 3\,\mathrm{GeV}$ the correction compared to a straight line is about $200\,\mu\mathrm{m}$. The initial $\vec{x}_{magnet}$ is, as discussed before, calculated from the T-Seed.

A.2 Fit of the x projection

Having only x measurements in Step 3, the track is fitted in the $x - z$ plane. Thus the algorithm has to fit the two track parameters x_0 and tx. The Downstream

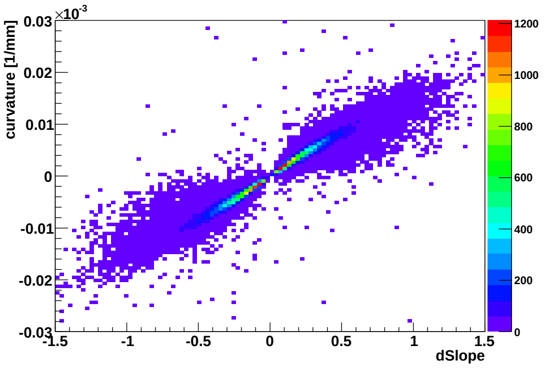


Figure A.1: *The curvature in the TT is drawn against dSlope.*

algorithm uses the least χ^2 method with the point in the magnet $\vec{x}_{magnet}$ as a constraint. Thus it is added to the χ^2:

$$\begin{aligned}
\chi^2 &= \sum_{x\ hits} \frac{(x_{hit} - x_{tr})^2}{\sigma_{hit}^2} + \frac{(x_{magnet} - x_{tr}(z_{magnet}))^2}{\Delta x_{magnet}^2} \\
&= \sum_i \frac{(x_i - x_{tr}(z_{magnet}))^2}{\sigma_i^2} \\
&= \sum_i \frac{(x_i - (x_0 + tx\ (z_i - z_{magnet})))^2}{\sigma_i^2} \\
&= \sum_i \frac{(x_i - (x_0 + tx\ dz_i))^2}{\sigma_i^2}
\end{aligned} \tag{A.3}$$

with the definition $z_i - z_{magnet} = dz_i$. The error of $\vec{x}_{magnet}$ is given by

$$\Delta x_{magnet} = \Delta z_{magnet}\ |dSlope| + 2.0\ [\text{mm}] \tag{A.4}$$

The default value for Δz_{magnet} is 30 mm.
Setting the partial derivatives to 0 and defining $\langle x \rangle := \sum_i \frac{x}{\sigma_i^2}$ one obtains:

$$\begin{aligned}
0 \overset{!}{=} \frac{\partial \chi^2}{\partial x_0} &= -2 \sum_i \frac{x_i - x_0 - tx\ dz_i}{\sigma_i^2} \Leftrightarrow 0 = \langle x \rangle - x_0 \langle 1 \rangle - tx \langle dz \rangle \\
0 \overset{!}{=} \frac{\partial \chi^2}{\partial tx} &= -2 \sum_i \frac{dz_i (x_i - x_0 - tx\ dz_i)}{\sigma_i^2} \Leftrightarrow 0 = \langle x\ dz \rangle - x_0 \langle dz \rangle - tx \langle dz^2 \rangle
\end{aligned} \tag{A.5}$$

This is equivalent to the matrix equation:

$$\begin{pmatrix} \langle 1 \rangle & \langle dz \rangle \\ \langle dz \rangle & \langle dz^2 \rangle \end{pmatrix} \begin{pmatrix} x_0 \\ tx \end{pmatrix} = \begin{pmatrix} \langle x \rangle \\ \langle x\ dz \rangle \end{pmatrix}. \tag{A.6}$$

After the fit the parameters of the track are updated with the result. Then the hit with the largest distance to the track is searched and if the distance is greater than 0.1 mm, the hit is removed. This procedure is repeated until the largest distance is smaller than 0.1 mm or there is only one hit left.

A.3 3D track fit

The TT has only two stereo layers which are rotated by ±5° with respect to the x layers. A full space track fit would require two additional parameters, leaving no degree of freedom for the fit. The bending of the magnet in the $y-z$ plane is only small. Therefore the slope in y is fixed to the one from the T-Seed with a small correction to compensate the B-field effects. The only additional parameter to fit is the displacement in y. A measurement in the TT is not a single point but a line because the TT is made out of silicon strips, cf. Chapter 1.3.2. Therefore one can shift the x and y coordinate of a hit along this line, cf. Figure A.2. Before the track is fitted, the x coordinate of every hit is updated with the track parameters.

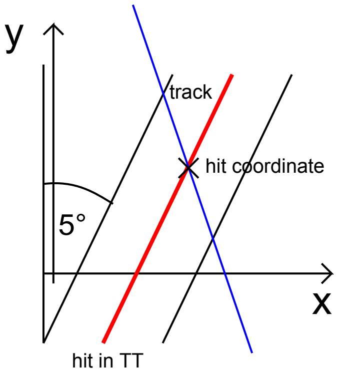


Figure A.2: *A measurement in the TT is a signal from a silicon strip. Thus the coordinate can be shifted along this line. In the algorithm the coordinates of the hit are adapted to match the track parameters.*

$$\begin{aligned} x_{hit} &= x_{hit}(y=0) + \left(\frac{dx}{dy}\right)_{hit} y_{tr}(z_{hit}) \\ &= x_{hit}(y=0) + \left(\frac{dx}{dy}\right)_{hit} (y_0 + ty\ dz) \end{aligned} \tag{A.7}$$

For x hits $\left(\frac{dx}{dy}\right)_{hit}$ is equal to 0. In equation A.7 y_0 is not a property of the measurement but a parameter of the fit. This has to be considered in the prediction of the trajectory, because a change in y results in a change in x.

$$\begin{aligned} x_{tr} = x_{tr}(dy=0) + \left(\frac{dx}{dy}\right)_{hit} dy \\ \text{with } dy_{magnet} = y_0 - y_{magnet} \end{aligned} \tag{A.8}$$

Thus the χ^2 is given by:

$$\chi^2 = \sum_i \frac{(x_i - (x_0 + tx\ dz_i + \left(\frac{dx}{dy}\right)_i\ dy_{magnet}))^2}{\sigma_i^2} \tag{A.9}$$

This results again in a linear system of equations for the three parameters x_0, tx and dy. As in the case of the 2D fit an outlier removal strategy is applied. During the outlier removal one has to make sure that the coordinates of the shifted hits are within the strip limits, that means the assumed hit coordinate lies on the strip. If this is not the case, the hit is removed from the track.

Appendix B

Addendum to systematic studies on K_S yield measurement

B.1 Fit model - fixed shape

In Figure B.1, B.2 and B.3 the K_s candidate mass distributions for data are shown. The shape of the mass peak (σ_{narr}, σ_{wide} and ϕ^{narr}) in the fit is taken from the corresponding bin in simulated data.

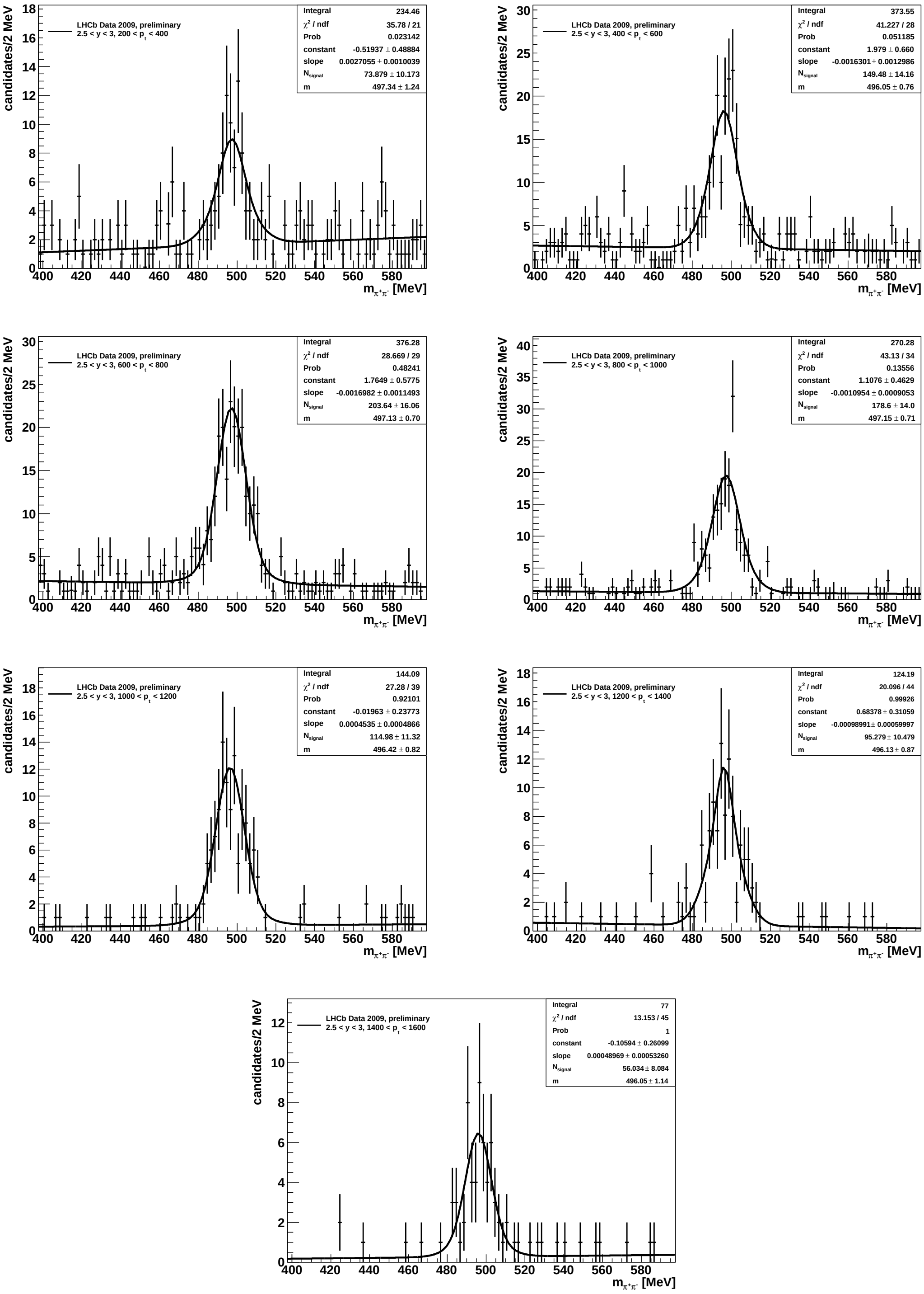


Figure B.1: *Fits to the K_s candidates mass distribution data for different p_T bins with $2.5 < y < 3.0$. The shape of the mass distribution is obtained from the simulation.*

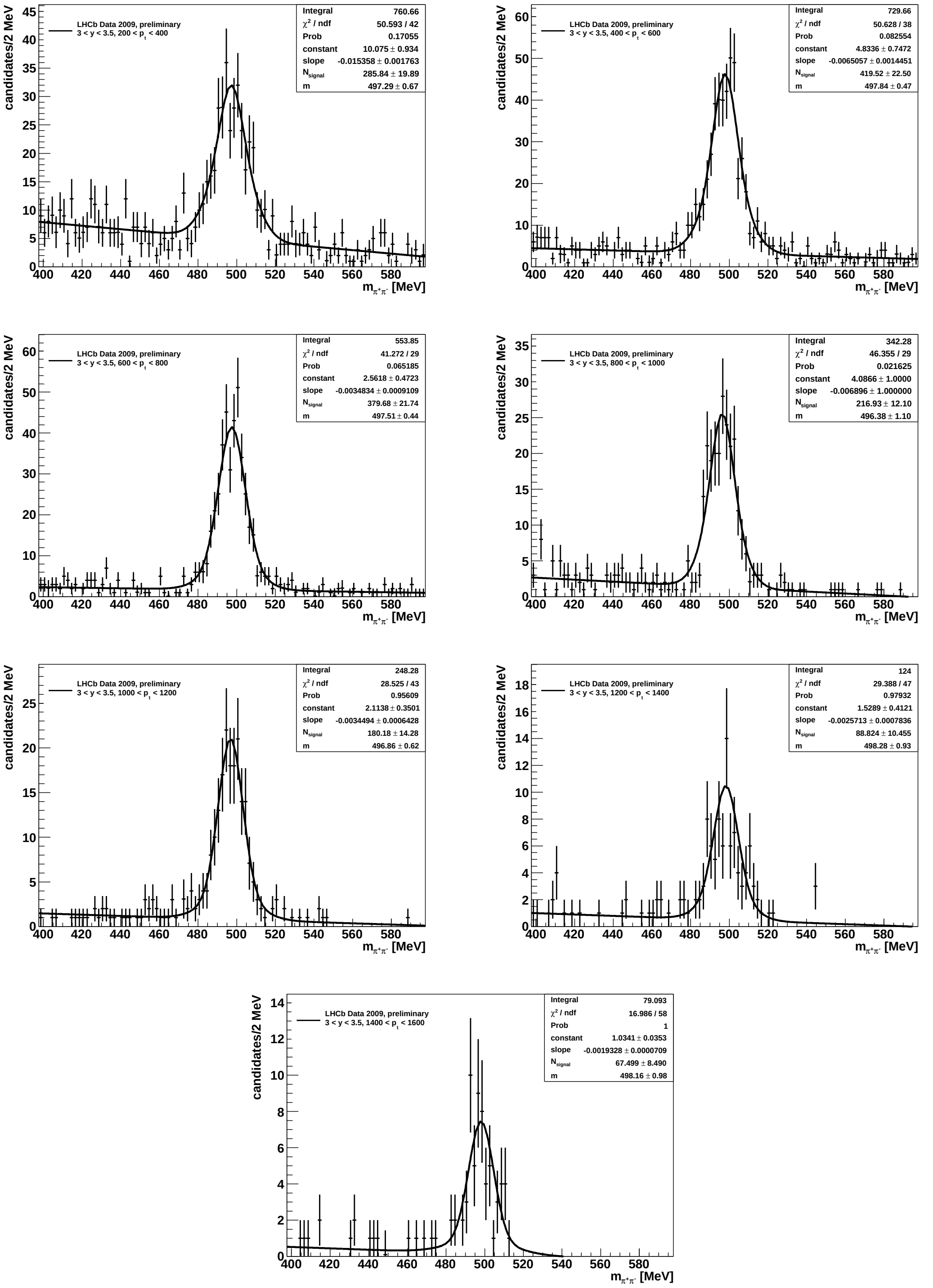


Figure B.2: *Fits to the K_s candidates mass distribution on data for different p_T bins with $3.0 < y < 3.5$. The shape of the mass distribution is obtained from the simulation.*

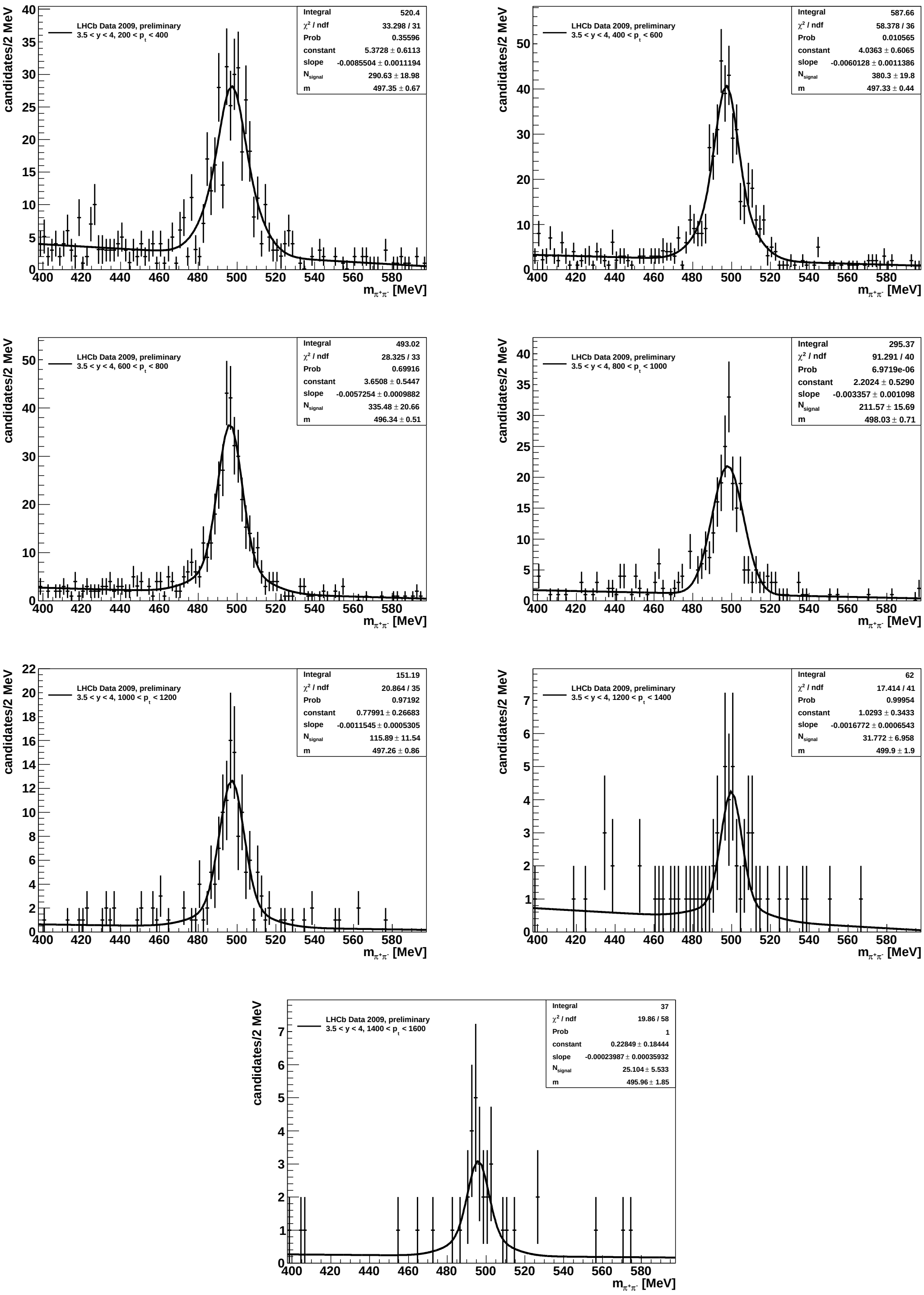


Figure B.3: *Fits to the K_s candidates mass distribution on data for different p_T ybins with $3.5 < y < 4.0$. The shape of the mass distribution is obtained from the simulation.*

B.2 Variations of the reconstruction efficiency within a bin

The efficiency weighted K_s candidates mass distribution on data are shown in Figure B.4, B.5 and B.6. For this systematic study the yield was determined by sideband substraction. Thus the fit of the mass peak is just for illustrative purposes.

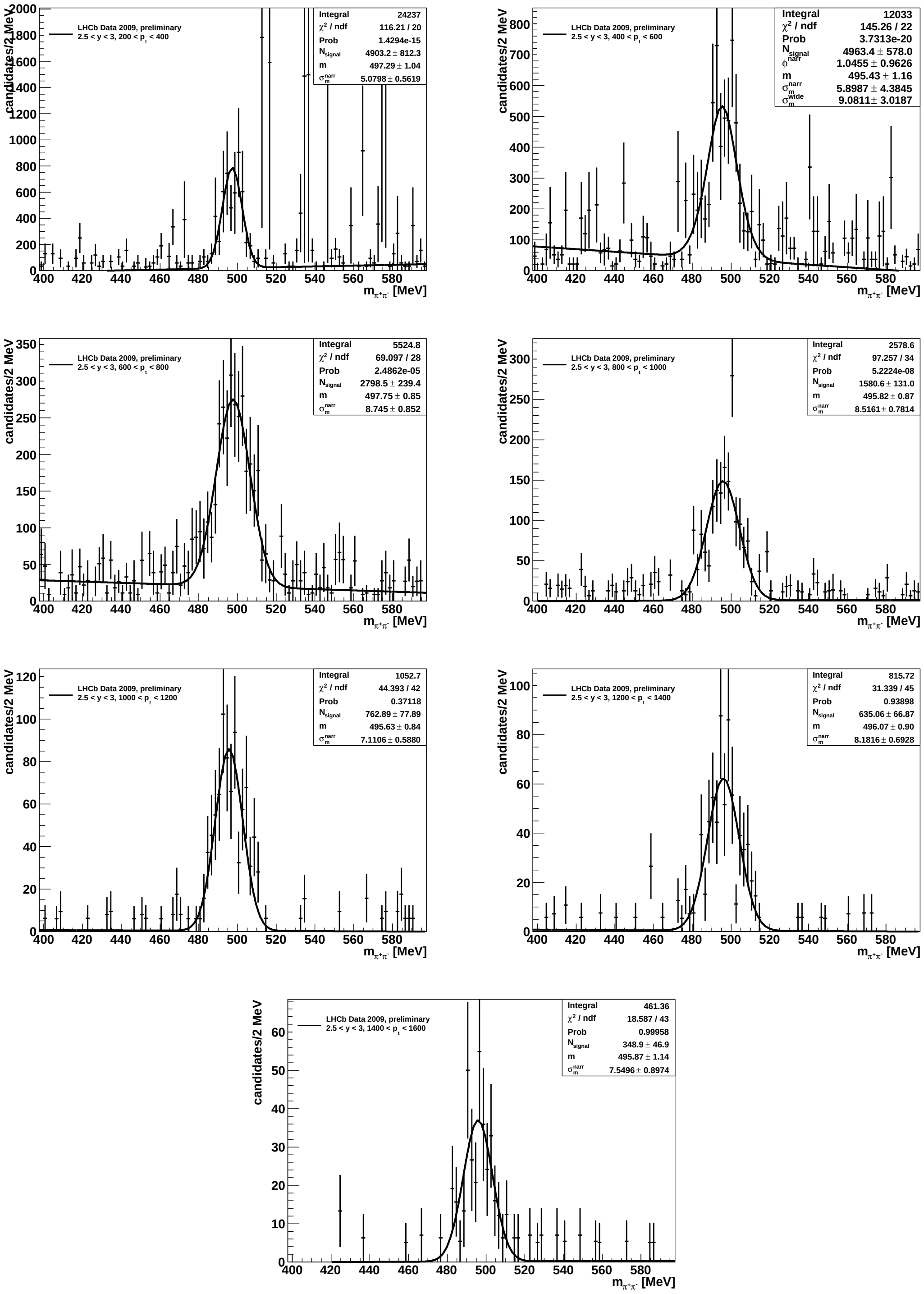


Figure B.4: *Fits to the efficiency weighted K_s candidates mass distribution on data for different p_T bins with* $2.5 < y < 3.0$.

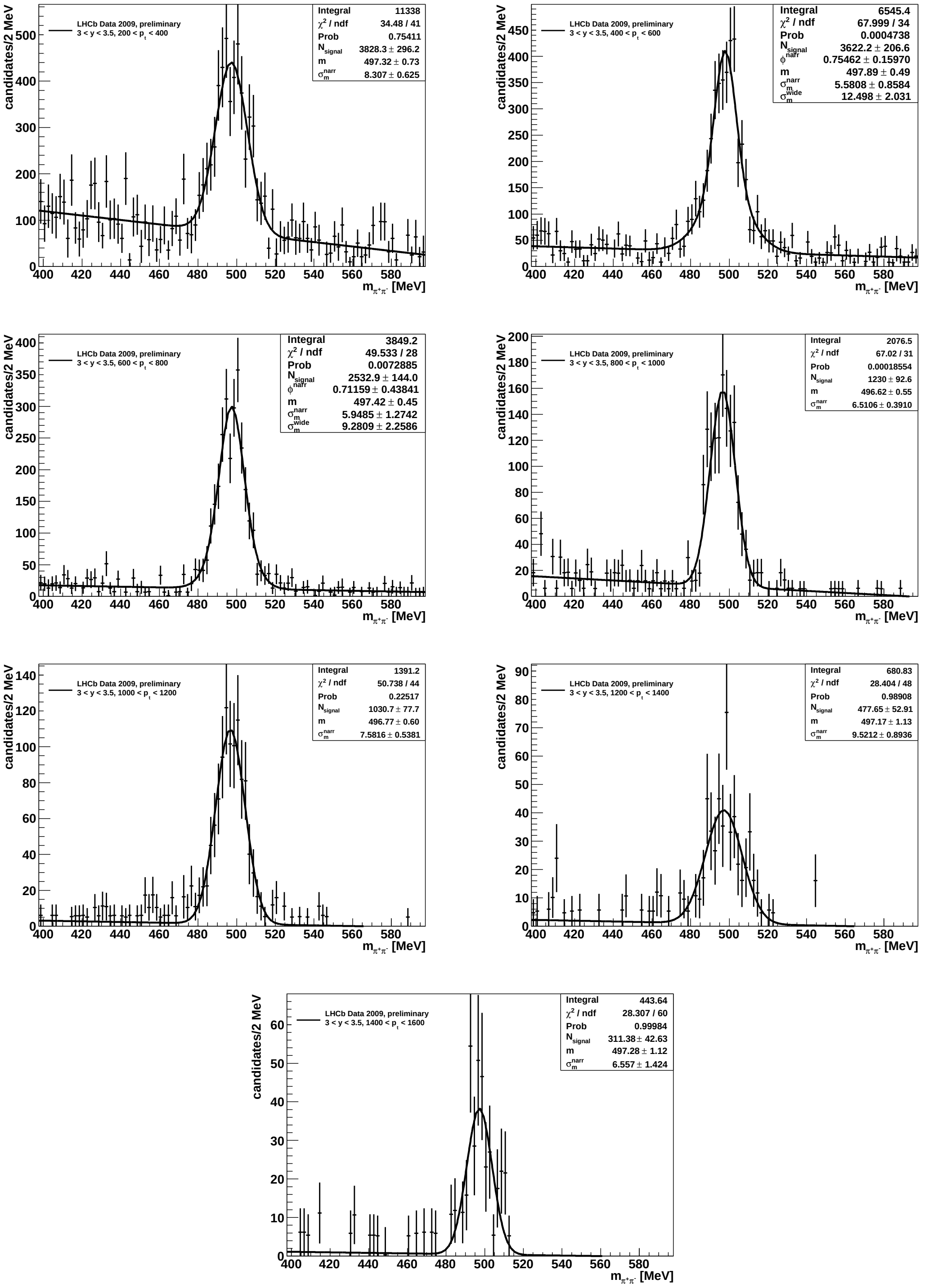


Figure B.5: *Fits to the efficiency weighted K_s candidates mass distribution on data for different p_T bins with* $3.0 < y < 3.5$.

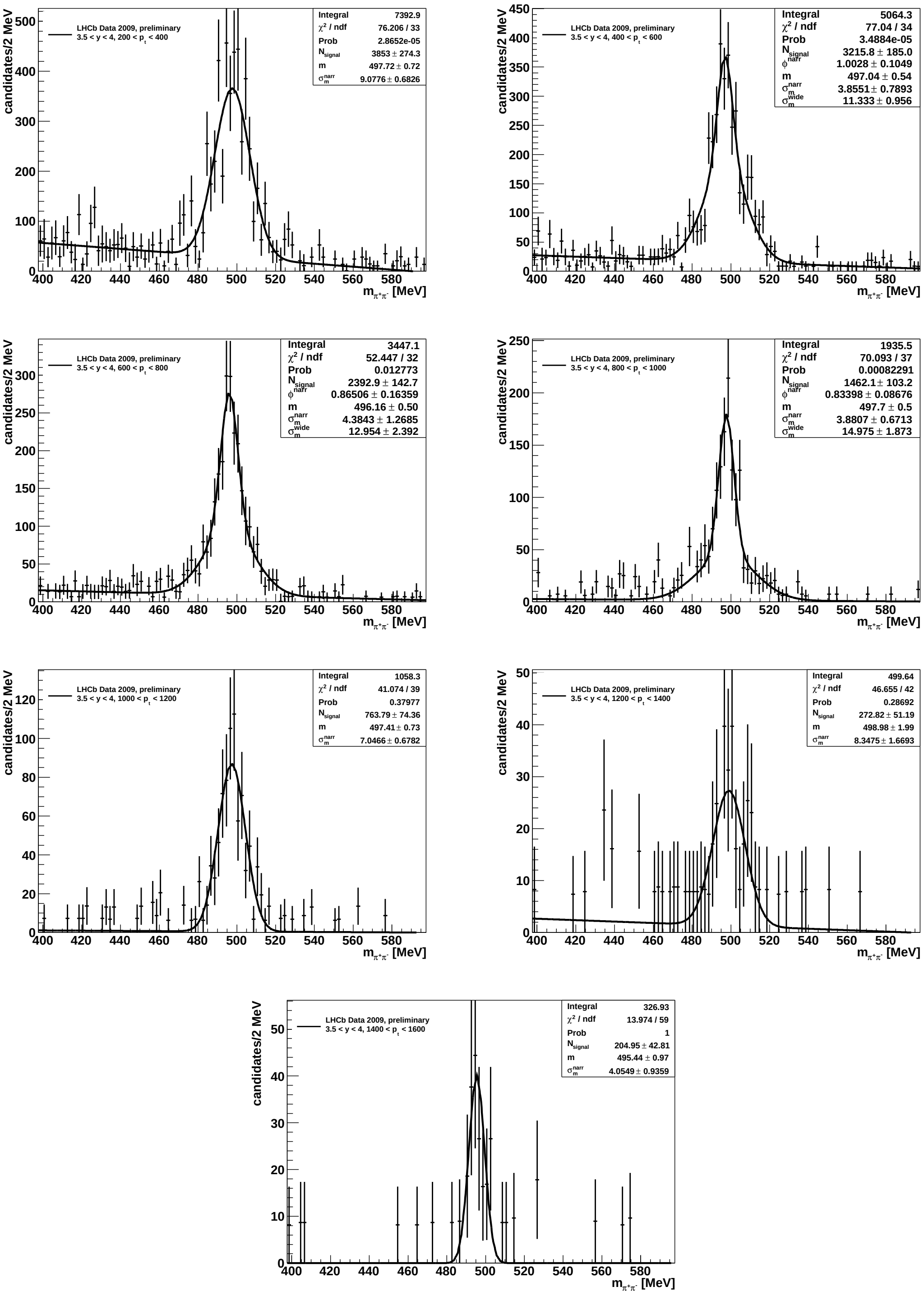


Figure B.6: *Fits to the efficiency weighted K_s candidates mass distribution on data for different p_T y bins with* $3.5 < y < 4.0$.

Erklärung

Ich versichere, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 11.Juni 2010