Spectroscopy of Orbitally Excited B_s Mesons with the CDF II Detector

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik der
Universität Karlsruhe (TH)

genehmigte

DISSERTATION

von

Dipl.-Phys. Martin Heck
aus Mainz

Tag der mündlichen Prüfung: 17.07.2009

Referent: Prof. Dr. M. Feindt, Institut für Experimentelle Kernphysik

Korreferent: Prof. Dr. W. de Boer, Institut für Experimentelle Kernphysik

Contents

Chapter 1

Introduction

The aim of particle physics is to understand the properties of matter on the most fundamental level. The two main questions are: 'What are the constituents of matter?' and 'How do those constituents interact with one another?'

One approach used throughout the experimental testing of theories of matter, that aim to answer these questions, is the study of the energy spectrum of excited states. Especially in cases of simple configurations as found in the hydrogenium atom, the potential energy related to an interaction can be studied.

In the Standard Model of particle physics, the interaction that glues quarks together to form hadrons, similar as the electromagnetic force binds electrons and nuclei to form atoms, is the strong interaction. While the electromagnetic force is described by Quantum Electro Dynamics [QED], the theory of the strong interaction is the Quantum Chromo Dynamics [QCD].
As the exchange particles in QCD carry charge themselves, which leads to a coupling constant in the order of one at smaller energies, calculations in QCD for precise predictions of the energy spectrum are very complex. There are various approaches to obtain approximate results. Many use the expansion in certain parameters such as the mass of the quarks forming the hadron. The b-quark is the heaviest quark that forms hadrons before it decays, and therefore its spectroscopy is of major interest. A third generation of fundamental particles was already suggested by Makoto Kobayashi and Toshihide Maskawa in 1973 [1] based on indirect effects. And only 4 years later in 1977 the first experimental discovery of b-quark containing hadrons was announced [2]. But despite of these early successes only few excited states of b-hadrons could be observed so far.
The main reason for the slow progress in the analysis of excited b-hadrons is the significant experimental challenge to produce them. To do so, large accelerators are used to attach enough kinetic energy to particles in order to produce new particles by transforming that energy to mass in accordance with Einstein's famous formula $m = E/c^2$. At the Tevatron sufficient energy for the creation of b-quarks and all kinds of their mesons and baryons, as well in excited states, is available. But higher energies accelerators and good detectors alone would not have enabled the studies carried out today. Various tools for data analysis such as multivariate analysis,

Monte Carlo simulations, and extended unbinned maximum likelihood fits have improved the possibilities to dig small signals out of large background contaminations.

In chapter 2 I describe basics of the Standard Model related to the study of orbitally excited mesons made of a b- and an s-quark [B_s], theoretical predictions, and results of previous experiments. In chapter 3 one can find a description of the experiment used to collect the data. In chapter 4 the important tools used in my analysis are explained. The following chapter 5 describes the first core of the analysis, the reconstruction and selection, before in chapter 6 the fits to the data, and the considered systematic uncertainties are presented.

Chapter 2

Orbital excitations of B_S mesons

In this chapter first some basic elements of the Standard Model are explained that are necessary to understand the context of orbitally excited B_s mesons. Then I discuss their decay and present some theoretical predictions of their properties. The chapter ends with a short description of experimental results known before the described analysis has been finished.

2.1 The Standard Model

The Standard Model is the current state of the art theory to describe matter and its interactions on the most fundamental level. The genuine matter particles are the fermions, that can be ordered in 3 families. All the fermions listed in table 2.1 are spin $\frac{1}{2}$ particles. Those fermions are divided further in quarks and leptons. Apart from the neutrinos, which show some special behaviour, only the members of the first generation, that can be found in the left column, are stable.
The major difference between the generations is the mass. With respect to their other properties the particles in one row are very similar and are addressed together as up-type quarks in case of the first row, or as down-type quarks for the particles in the second row.

up-quark	**c**harm-quark	**t**op-quark
down-quark	**s**trange-quark	**b**ottom-quark
electron	muon $[\mu]$	tau $[\tau]$
electron neutrino $[\nu_e]$	muon neutrino $[\nu_\mu]$	tau neutrino $[\nu_\tau]$

Table 2.1: The fermions of the Standard Model. The particles in the first two rows are called quarks. The particles in the last two rows are called leptons.

Additionally, there are the bosons that are the exchange particles for the interactions of the fermions. Gravitation is too weak to have measurable influence on experiments dealing with the microcosmos, and the Higgs boson is not associated with

one of the interactions, but with mass. Table 2.2 lists the interaction bosons and the correspondingly charged particles. The listed bosons all have spin 1.
The photon and gluons are massless, but due to their different coupling properties the yielded results are still very different. The exchange particles of the weak interaction are very heavy, which is the main reason for the weakness of their interaction at low energies. I want to discuss these three forces which emanate from these bosons a bit more.

Interaction	Bosonic carrier	Charged particles
Electromagnetic	photon [γ]	quarks and electron-like leptons
Weak	Z,$W^{\pm}$	all fermions
Strong	8 **g**luons	quarks and gluons themselves

Table 2.2: Forces, corresponding bosons, and charge carriers of the Standard Model

2.1.1 The Electromagnetic Interaction

As the electromagnetic interaction is the most widely known, it is often useful to understand the other interactions in analogy to it. However, for the subsumption of the electromagnetic interaction in the study of particle physics, one has to note, that the electromagnetic force is relatively strong and the decays happen almost instantaneous, if allowed at all.
The up-type quarks have a charge of $+\frac{2}{3}$, and the down-type quarks have a charge of $-\frac{1}{3}$ of the elementary charge of the electro-magnetic interaction. The electron-like leptons carry charge one, and the neutrinos are not charged electro-magnetically.
There is only one exchange boson for the electromagnetic force, the photon. It doesn't transport charge itself. However, there can be spontaneous creations of pairs of charged particles from vacuum. As these virtual particles have a dielectric effect, the electromagnetic force has some variation with the energy level of interactions. If a process happens at very high energies corresponding to very short distances, the interacting particles see less of that dielectric effect, and thus the electromagnetic coupling constant increases slightly with the energy scale.

2.1.2 The Weak Interaction

All quarks carry a quantum number called flavour that corresponds to its type, e.g. strangeness for the s-quark and beauty for the b-quark. The other interactions have to obey the conservation of these quantum numbers and can not lead to the decay e.g. of a b-quark. Only the annihilation with an anti-b-quark is possible for the strong and electromagentic force. The weak interaction can violate such quantum numbers and decay up-type and down-type quarks into each other within the weak quark doublet, an up-type and a down-type quark of the same family. The eigenstates of these doublets are not identical to the ones of spin and mass, but are

slightly rotated. This rotation is described by the CKM-matrix, and allows decays of the b-, and s-quark mass eigenstates into the mass eigenstates of c- and u-quarks, although the lifetime is increased compared with decays in one family, e.g. the decay of a c-quark to an s-quark.
The exchange particles of the weak interaction are massive. Therefore the propagator for this interaction is suppressed when the energy available for the decay is much lower than their mass. In the context of this thesis, this means that b-quarks have a lifetime, that allows them to fly a distance within the resolution of the measuring apparatus. Strange quarks can be treated as stable, as their lifetime most often will allow them, to hit the detector, before they decay.

2.1.3 The Strong Interaction

In the context of the study of excited B_s mesons especially important is the strong force, whose only charge carriers are the quarks and the exchange bosons, the gluons, themselves. The fact that gluons carry charge themselves has major consequences. It makes the strong interaction a short range force, despite gluons are massless like photons. As well it hampers perturbative calculations impossible at low energies, because the coupling constant is close to one. The strong interaction is invariant to parity and charge transformations separately, unlike the weak interaction.
As well the strong interaction can't distinguish between different flavours of any type. If the mass difference of two quarks is small compared to a fundamental parameter Λ_{QCD}, which one needs for the renormalisation of the theory and is around one $\mathrm{GeV/c^2}$, one can exchange one quark for another with small corrections proportional to the mass difference over Λ_{QCD}. This allows as well to construct a quantity called isospin which behaves like the spin. In this formalism u-quarks and $\overline{d}$-quarks are equal to spin up particles, d-quarks and $\overline{u}$-quarks equate spin down particles. Λ_{QCD} itself has to be determined by experiments and is as well connected to the scale, at which the non-linear effects of gluon-gluon interactions make it increasingly difficult to calculate them as perturbative corrections.
Generally the strong interaction is described as SU(3) group with the three charges called blue, green, and red. The gluons span the $3\otimes\overline{3}$ colour-anticolour space in form of 8 gluons, of which none is colour neutral. Despite the complexity of the strong interaction, there are some cases for which the general behaviour is predictable.

Asymptotic Freedom and Confinement

Two of these cases are, when two colour charge carriers are either very close to each other, or when one tries to separate them.
Despite there is a reduction in the dampening effect from vacuum fluctuations of charged particles at short distances like in QED, the effect of virtual gluons emaneted from the quark is dominating. As gluons can change the colour of the quark, when interacting, this effect reduces the seen charge. So the closer one gets, the smaller the effective colour charge is. This effect is called *Asymptotic Freedom*. Smaller

distances correspond to a higher energy, which is more often referred to as reference than space. At very high energies quarks can be assumed to be free particles that do not interact with other close particles by the strong interaction.
Hadronic particles consist not only of the valence quarks and anti-quarks but are surrounded by a cloud of virtual particles as a consequence of the selfinteraction of the gluons. If one tries to separate quarks, the virtual particles prevent a reduction of the attracting force, as they are themselves attracting the valence quarks as source of charge. We don't see long range interactions of strong force, as already after a separation of about 1 fm, the pair production of quarks allows the system to split into two seperate colour neutral systems. Due to that effect, it is not possible to have colour charged objects isolated much more than 1 fm from each other. This effect is called *Confinement*, and prevents the occurance of free quarks. Instead the quarks hadronise to colour neutral mesons and baryons, and potentially other colour neutral combinations.

Having discussed the cases when we have very high energies or close distances and the case of separation or low energies, I want to give a leading order term for the strong coupling constant:

$$\alpha_s(\mu) = \frac{4\pi}{(11 - \frac{2}{3}\mathrm{N}_F)\ln(\mu^2/\Lambda^2_{QCD})}$$

μ is an energy scale, that corresponds to a given distance, e.g. the distance between the quarks in a meson or the momentum transfer in a collision process, that one wants to calculate. N_F is the number of quark flavours with a mass lower than the energy scale μ.

Heavy quark symmetry

The heavy quark symmetry is exact in the limit of an infinite heavy quark Q, and is approximate true, if $\mathrm{m}(Q) >> \Lambda_{QCD}$. In this case the quark Q takes no notice from another light quark to which it is bound and can be exchanged by another infinite heavy quark of a different flavour. The heavy quark then can be seen as a static source of colour for the other quarks in a hadron and allows to study the potential of the QCD in a similar way, as the hydrogen atom can be used to study QED.
The exchange of one heavy quark with another is similar to the effect of taking different isotopes in atom physics. The decoupling of spin effects of the heavy quark from the interaction with the light quark correspond to the smallness of the hyperfine splitting in atom physics. The corrections in particle physics are much bigger though, since even the mass of the b-quark is only a few times Λ_{QCD}. The effective theory building up on this symmetry is called *Heavy Quark Effective Theory* [HQET] and is described in the literature, e.g. [3], [4], [5].
HQET allows as well to introduce quantum numbers for mesons describing the dynamics, not only for a whole meson, but as well for a heavy quark and a light

quark separately. In close analogy to the hydrogen atom one can define the total angular momentum J, the total angular momentum of the light quark j_q, the spin of the heavy quark S, the orbital momentum of the light quark with respect to the b-quark L, and the parity of the whole meson P.

2.2 The decay of orbitally excited B_s mesons

At a proton-antiproton collider the b-quarks are produced with considerable momentum, so that during the hadronisation process all kinds of mesons and baryons, including ones with orbital excitation are created. The particles, that are analysed in this thesis, are mesons consisting of an s-quark and an anti-b-quark, or the corresponding anti-mesons, with an orbital excitation of the s-quark relative to the b-quark L=1. Using heavy quark symmetry, the total angular momentum of the s-quark is composed of the orbital excitation and the spin of the s-quark. This leaves the possibilities $j_s = 1/2$ or $j_s = 3/2$, corresponding to the fine splitting in the hydrogen atom. Both s-quark states can be combined with the spin of the b-quark, corresponding to the hyperfine splitting in the hydrogen atom, which gives overall four combinations. We denote the states in the following way:

- $B^*_{s0}(1/2)$: This is the state with $j_s = 1/2$ and total angular momentum $J = 0$.
- $B_{s1}(1/2)$: This is the state with $j_s = 1/2$ and $J = 1$.
- $B_{s1}(3/2)$: This is the state with $j_s = 3/2$ and $J = 1$.
- $B^*_{s2}(3/2)$: This is the state with $j_s = 3/2$ and $J = 2$.

The usual denomination omits the angular momentum of the light state, but this leaves an ambiguity between two states, which turns out to be decisive. All four states together are referred to as B^{**}_s mesons.

The main decay mode of the B^{**}_s mesons is into a combination of B and K mesons. These can be the neutral mode $B^0_d K^0$, or $B^+_u K^-$. The decay proceeds by the strong interaction, which makes it very fast.
A kinematically more favourable decay into $B_s\pi$ is forbidden by isospin conservation of the strong interaction. The π has isospin one and the B_s isospin 0, adding up to one, while the original B^{**}_s has only isospin 0.
In the decay both, the parity and the total angular momentum, have to be conserved. This can be achieved by a combination of internal angular momenta of the decay products and additional angular momentum between them in the decay wave. Even the first excitation of a kaon has a much higher mass than its ground state, which forbids this kinematically. The B meson in the final state may be in its spin excited state B^*, which has angular momentum J=1. The B ground state and the spin excited state have parity -1 as well as the kaon, while the orbital excitation gives the B^{**}_s a parity of +1. The four B^{**}_s states and the possible B states, into

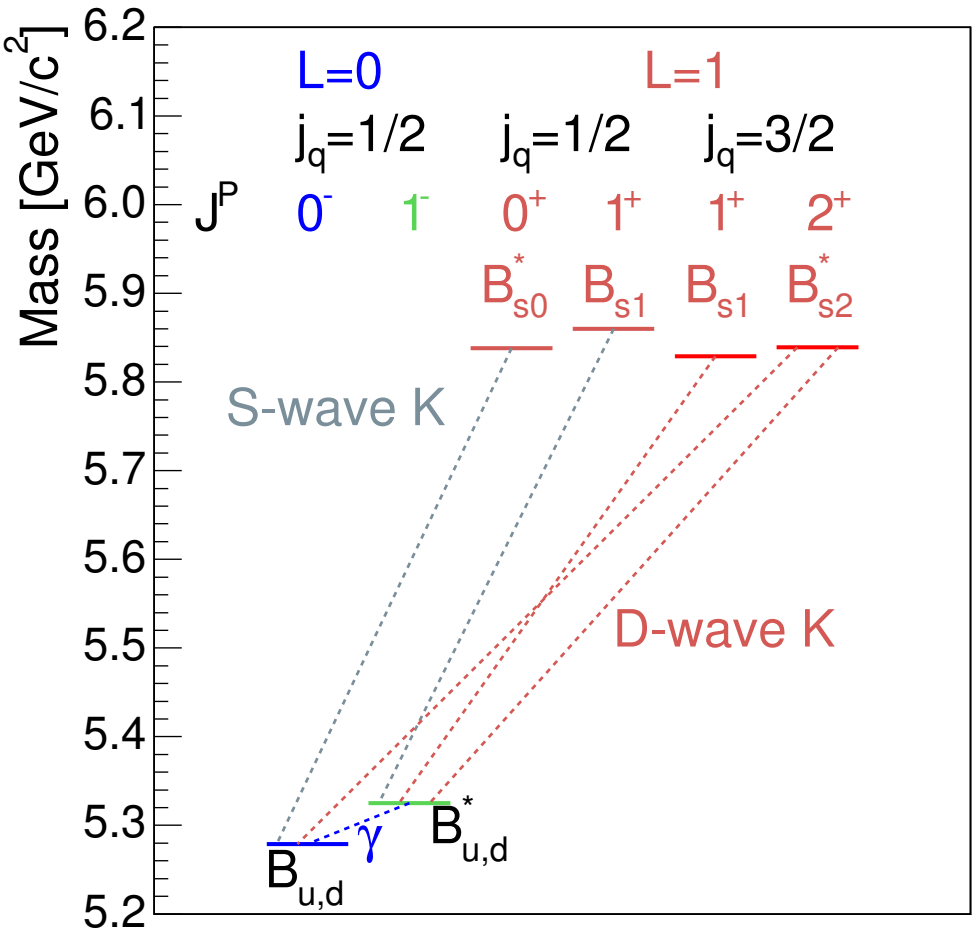

Figure 2.1: The spectrum of the B_s^{**} mesons and the dominant decay modes. The masses of the B_s^{**} mesons are taken from the results of this work in case of the $j_s = 3/2$ states and theory predictions for the $j_s = 1/2$ states.

which the B_s^{**} mesons may decay, are shown in figure 2.1.
As additional angular momentum between the final states reduces the overlap integral considerably, the mesons decay without it, if allowed in an S-wave, where S (P, D) denominates additional angular momentum of 0 (1, 2) as in atom physics. The $B_{s0}^*(1/2)$ decays into a kaon and a B in the ground state. The $B_{s1}(1/2)$ decays into a kaon and the spin excited state of the B. The $B_{s2}^*(3/2)$ with its total angular momentum of J=2 has no possibility to decay without additional angular momentum. A P-wave decay is as well not possible, as the parity changes with $(-1)^L$ for additional angular momentum L. As the kaon and the B have already $(-1) \cdot (-1) = +1$ as the B_s^{**}, a P-wave decay would violate parity conservation. The next possibility for the $B_{s2}^*(3/2)$ is a D-wave decay. The combination of the angular momentum of the D-wave with the angular momentum of the B_s^{**} allows both, the decay into BK and to B^*K.
Given only the rules named so far, the $B_{s1}(3/2)$ would decay as the $B_{s1}(1/2)$, but HQET tells us that the spin of the b-quark decouples from the total angular momentum of the s-quark. This means that the angular momentum of the s-quark system j_s has to be conserved separatly, which forbids the decay into B^*K by an S-wave. The argument against P-waves is valid for the $B_{s1}(3/2)$, too, so it only has the possibility to decay in a D-wave like the $B_{s2}^*(3/2)$.

2.3 Theory predictions for the B_s^{**}

There has been a lot of theoretical work to determine the mass of excited heavy mesons in general, including the B_s^{**}. Among the tools used for mass determination are HQET ([6], [7], [8]), chiral theory ([9], [10]), and potential models ([11]), that have lead to various predictions. An approach, that only recently became viable for predicting masses of excited mesons, is the use of lattice calculations ([12]). The resulting values for the mass in these calcualtions are given in table 2.3. Not all of the calculations try to estimate the uncertainty. The (hyperfine) mass splitting between the two narrow states, for which the calculations have less uncertainty, varies between 12 and 20 MeV/c^2.

Some of the calculation include predictions on the width of the states. The results are summarised in table 2.4. One has to keep in mind, however, that the width is not independent of the mass. If e.g. the $B_{s0}^*(1/2)$ is below the threshold to BK, the state becomes very narrow. As well the width of the $j_s = 3/2$ states depends on the energy that is available from the decay. When comparing results with this values, one therefore has to check the predicted width in the corresponding calculation as well.

The relative branching of $B_{s,2}^*$ in BK or B^*K is dominated by phase space. The actual decay rate is roughly proportional to $F \cdot p^{2L+1}$ with p the momentum of the kaon in the B_s^{**} restframe, in this case p^5, and F the ratio between the form factors. This ratio should be the same as in D^{**} decays, where it is 0.61 $\pm$ 0.16 So the relative branching varies very strongly over the mass range to which the various

State	Ref. [6]	Ref. [7]	Ref. [8]	Ref. [9]	Ref. [10]	Ref. [11]
$B^*_{s0}(1/2)$	5841	-	-	5718± 35	5710 ± 30	5830
$B_{s1}(1/2)$	5859	-	-	5765± 35	5770 ± 30	5860
$B_{s1}(3/2)$	5831	5886 ± 40	5834	-	5877 ± 3	5860
$B^*_{s2}(3/2)$	5844	5899 ± 40	5846	-	5893 ± 3	5880

State	Ref. [12]
$B^*_{s0}(1/2)$	5756 ± 31
$B_{s1}(1/2)$	5804 ± 31
$B_{s1}(3/2)$	5892 ± 52
$B^*_{s2}(3/2)$	5904 ± 52

Table 2.3: B^{**}_s mass spectra for various approaches. The masses are given in MeV/c^2.

$b\overline{s}$	Ref. [7]	Ref. [8]	Ref. [10]	Ref. [11]
$B^*_{s0}(1/2)$	-	-	-	630
$B_{s1}(1/2)$	-	-	-	-
$B_{s1}(3/2)$	2.8± 1.2	1 - 3	3.5 ± 1.0	-
$B^*_{s2}(3/2)$	7 ± 3	3 - 7	11.3 ± 2.6	1.9

Table 2.4: B^{**}_s width predictions for various approaches. The widths are given in MeV/c^2.

theories hint. With a $B^*_{s2}(3/2)$ mass at 5850 MeV/c^2, $\frac{B^*_{s2}\to B^*K}{B^*_{s2}\to BK}$ would be about 0.1, with a $B^*_{s2}(3/2)$ mass at 5900 MeV/c^2, $\frac{B^*_{s2}\to B^*K}{B^*_{s2}\to BK}$ at about 0.3.

Finally one can ask, if theory predicts something about the relative production and branching fractions. For these question I resort to very simple models, as there are no sophisticated predictions from theorists available.
The b-quark spin can end up either parallel or anti-parallel to the light quark's angular momentum, producing the two different states. The first model for the relative production ratio is state counting, which gives

$$B^*_{s,2} : B_{s,1} = 1 : 1.$$

But the anti-parallel state has spin 1 for the whole meson, providing 3 possible spin-positions, and the parallel state has spin 2, providing 5 possible spin-positions. So the other simple model is spin-counting, which leads to

$$B^*_{s,2} : B_{s,1} = 5 : 3$$

I will revisit all these discussed properties, when we interprete the experimental results.

2.4 Experimental status

Measurements existing before this work started were made by the LEP experiments OPAL and DELPHI. OPAL [13] measured a state with a mass of 5853 ± 15 MeV/c^2, which is attributed to B^{**}_s decays, but has been unable to determine which of the states or combination of B^{**}_s states produced their signal. DELPHI [14] improved the measurement to 5852 ± 4 MeV/c^2 and provided arguments that the observation made corresponds to the $B^*_{s,2}$ state. The plot of which this result has been produced is seen in 2.2 Parallel to the publishing of the first results of this thesis [15], the other Tevatron experiment DØ has published [16] a result with a resonance at 5839.6 ± 1.1 (stat) ± 0.7 (syst). The corresponding fit is shown in figure 2.3.

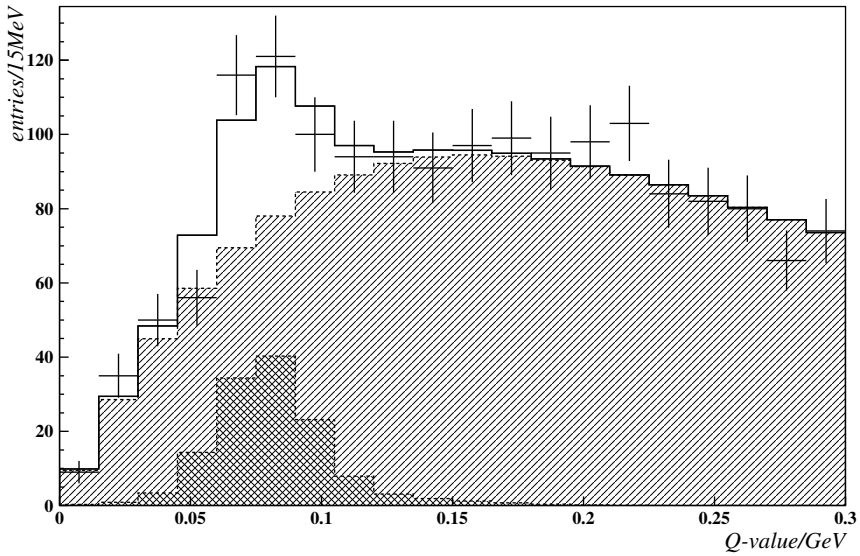

Figure 2.2: Observation of B_s^{**} by DELPHI. The signal contribution is the dark shaded area.

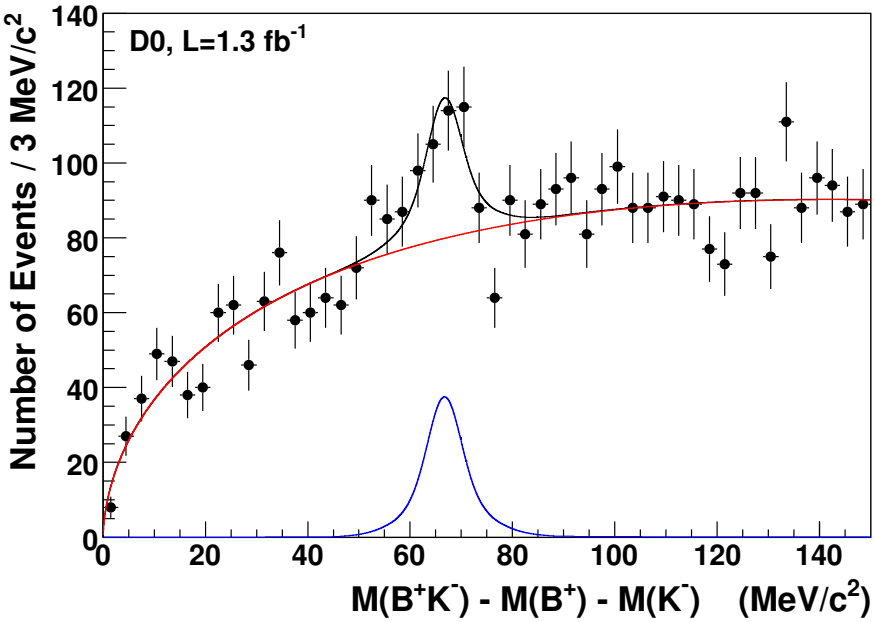

Figure 2.3: Mass measurement of one B_s^{**} state.

Chapter 3

The CDF II Experiment

In this chapter the CDF II experiment, that operates at the Fermi National Accelerator Laboratory (FNAL), a facility of the U.S. Department of Energy, is described. The collisions, that are recorded with the CDF II detector, are provided by the Tevatron accelerator.
The FNAL is located in Batavia, Illinois (USA), about 50 km west of Chicago. Figure 3.1 shows a main part of the FNAL from a bird's eye view. The Tevatron is located close to the maintenance road around the center of the picture, the CDF II detector is located in the orange building on the left side of the Tevatron.

3.1 The Tevatron and its Preaccelerators

The Tevatron has since 1995 been the accelarator with the highest center-of-mass energy in the world. At the Tevatron protons and antiprotons are collided. The name Tevatron derives from the center-of-mass energy being more than one TeV, therefore 'TeV-atron'. In a first phase this energy was $\sqrt{s}$= 1.8 TeV, while in a second phase, starting 2001 and commonly called 'Run II', the Tevatron operates at $\sqrt{s}$= 1.96 TeV. More important for this analysis and in b-quark related research than the improvement of the center-of-mass energy is the much higher number of interactions which have been recorded in Run II.

3.1.1 The Accelerator Chain

For technical reasons, it is impossible to accelerate the protons and antiprotons in just one step. Therefore a system of preaccelerators is built, as is shown in Figure 3.2.

In a first step, negatively ionised hydrogen atoms are produced in a Cockroft-Walton-type accelerator and accelerated to a kinetic energy of 750 keV. A linear accelerator ('Linac') increases this energy to 400 MeV, before the ions are transferred into the

Figure 3.1: The Tevatron from a bird's eye view. The Tevatron is located in the center of the picture under the upper of the two visible rings.

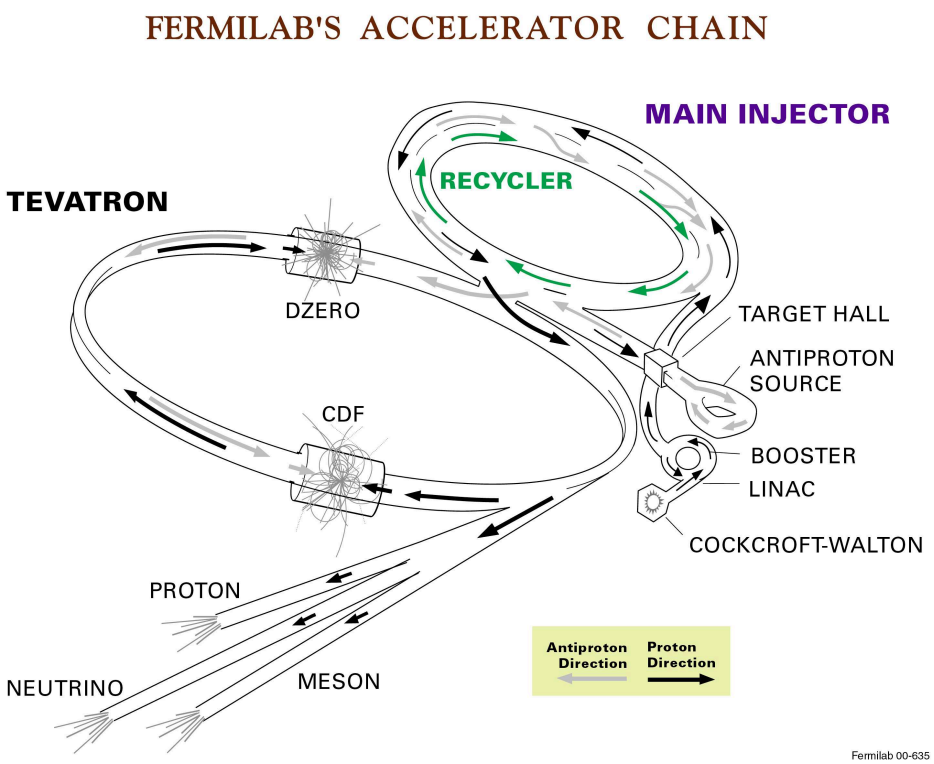

Figure 3.2: Schematic overview of the Tevatron Accelerator Chain.

'Booster', a synchrotron of 75 m radius, where the ions are stripped from the electrons and further accelerated to 8 GeV.
The last stage before the Tevatron is the Main Injector, which can as well be seen on Figure 3.1 in the lower third. The Main Injector accelerates the protons to an energy of 150 GeV, and injects part of them in a single bunch into the Tevatron. This process is repeated until 36 bunches are circulating in the Tevatron for a single 'shot setup'.
The Main Injector is as well used to create antiprotons, by beaming protons of 120 GeV on a nickel target. From the resulting cloud of various particles, antiprotons with an energy of roughly 8 GeV are collected and thermally cooled in the 'Debuncher' and the 'Accumulator'. From there the antiprotons are transferred into the 'Recycler', located in the same tunnel as the Main Injector. There the antiprotons are stored until they are transfered to the Main Injector to inject as well 36 bunches of antiprotons into the Tevatron.
All bunches together have about $\mathcal{O}(10^{13})$ protons and $\mathcal{O}(10^{12})$ antiprotons. The reason for using more protons than antiprotons is in the difficulty to produce the latter. Once they are loaded, both protons and antiprotons are accelerated to a kinectic energy of 980 GeV. The course of the particles is changed so to collide at two interaction points, one of those utilized by the CDF II experiment, with a center-of-mass energy of two times 980 GeV.
After some time, usually 16-36 hours, the number of protons and antiprotons is reduced so much by losses and collisions that the Tevatron has to be filled again. For that the remaining particles are dumped, and complete fresh particles are injected as described above. The data taken with one load of protons and antiprotons is called a 'store'.

3.1.2 Luminosity

A quantity for describing the performance of an accelerator beside the provided center-of-mass energy is the luminosity ($\mathcal{L}$), from which one can conclude how many interactions for a process with a given cross section can be expected in a unit of time. It can be parameterized depending on the number of bunches n, the frequency f with which the bunches circle, the number of protons (antiprotons) in an average bunch, the structure of the bunch described by the average transverse width for a proton (antiproton) bunch σ_P ($\sigma_{\overline{P}}$), and a form factor F. F would be 1 for a Gaussian distribution with a typical value of 70% at the Tevatron. The formula for the luminosity using this parametrisation is:

$$\mathcal{L} = n \cdot f \cdot \frac{N_p N_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} \cdot F.$$

As b-quarks are produced relatively rarely in proton antiproton collisions at the Tevatron energy, the analysis described in this thesis profits from a large time-integrated luminosity ($\mathcal{L}_{int}$), which is directly proportional to the number of b-quarks produced with the cross section as proportionality constant.

As cross sections have areas the dimension of an area, $\mathcal{L}_{int}$ has inverse area as dimension, typically expressed in inverse 'picobarn' (pb^{-1}), with 1 pb = $10^{-36}cm^2$. The integrated luminosity has made a pleasant development over the time of the Tevatron's operation. As can be seen in Figure 3.1.2 the curves for $\mathcal{L}_{int}$ increase more than linear, which means over the operational time more integrated luminosity can be delivered by the acceleretor before the Tevatron has to be refilled. This mainly comes from a better control of the beam, resulting in a lower likelihood for proton and antiproton losses per revolution, and from an increased number of antiprotons, due to improvements in their production and collection, leading to a higher initial luminosity.
The Tevatron was designed for an initial luminosity of $270 \cdot 10^{30} cm^{-2} s^{-1}$, which was reached the first time at the end of 2006. In the meantime even 20% higher initial luminosities could be achieved. There is a 'Delivered' and an 'On Tape' curve in Figure 3.1.2. 'Delivered' represents the amount of integrated luminosity, which was provided by the Tevatron accelerator. 'On Tape' represents that portion that was recorded. The CDF II experiment sometimes fails to record data, as it is a very complex device that suffers from technical issues. The data taking efficiency is typically 80-90 %. Until July 2009 about 5800 pb^{-1} have been recorded overall.

3.2 The CDF II Detector

3.2.1 General Overview

CDF II is the acronym for 'Collider Detector at Fermilab'. An extensive description can be found in ref. [17, 18, 19, 20].
CDF II has a cylindrical symmetry with respect to the beamline and a forward/backward symmetry with respect to the interaction point of the Tevatron.
For orientation one uses a right handed cylindrical coordinate system:

- z axis in direction of proton beam,
- polar angle θ measured from positive z axis,
- ϕ measured from the plane of the Tevatron,

Instead of θ one often uses the pseudorapidity η defined as $\eta = \ln[\tan(\theta/2)]$, since it is easy to handle under Lorentz transformations.
A schematic view of the CDF detector is shown in figure 3.4. The main componentes of CDF are:

- tracking system with
  - Silicon Detector
  - a drift chamber, called 'Central Outer Tracker' (COT)

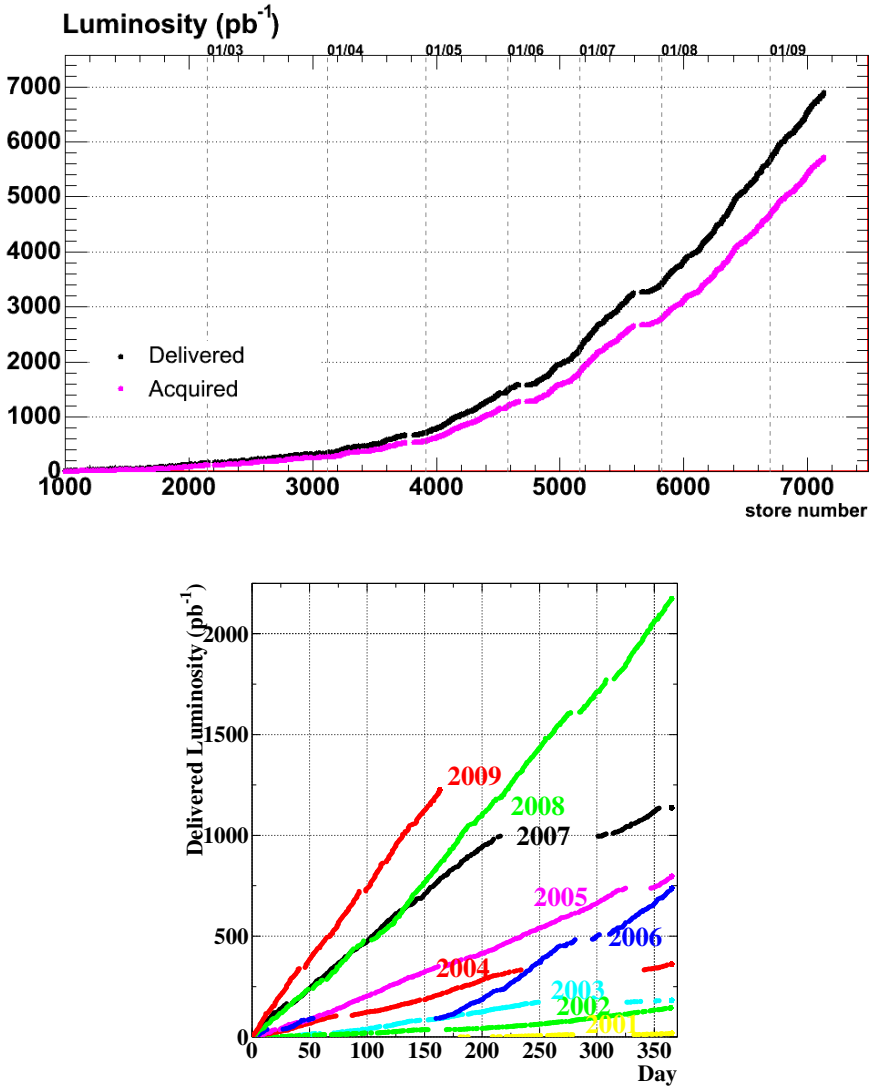

Figure 3.3: Evolution of the integrated luminosity as a function of store (left) and per year (right).

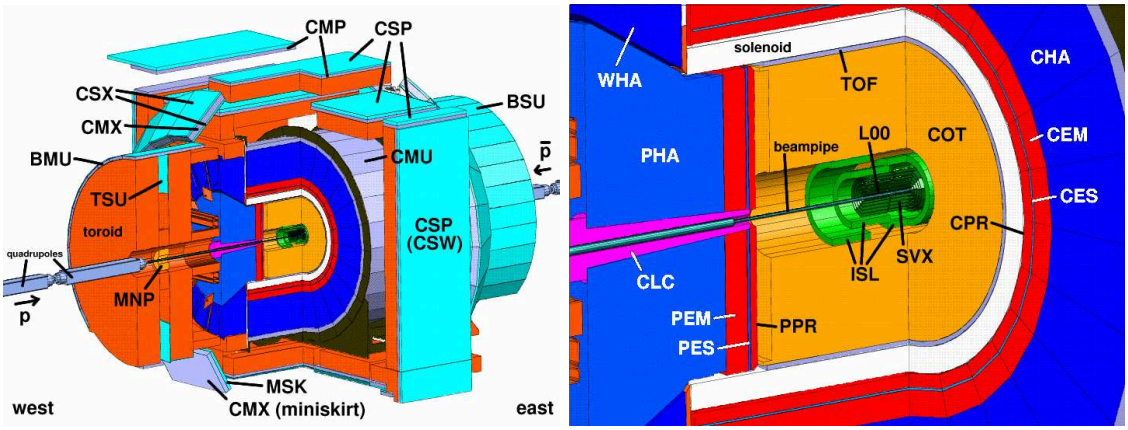

Figure 3.4: Schematic cut away view of the CDF II detector. On the left the whole detector with labeling of the muon system, that make up the outermost parts. On the right a zoomed in version with labels for the tracking and calorimeter systems is shown.

- Time of Flight Detector (TOF)
- Calorimetry with
  - Hadron Calorimeter
  - Electro-magnetic Calorimeter
- Muon system

3.2.2 Tracking System

The innermost part of CDF II, mounted directly at the beam pipe, is the tracking system, schematically shown in figure 3.5. It measures momentum and spatial information of charged particles near the interaction point. The tracking system consists of two main parts, the silicon detector, which is very good in reconstructing vertices with high precision and the drift chamber called COT (Central Outer Tracker), which is very good in measuring momentum of particles traversing the detector.

The silicon detector shown schematically in figure 3.6 has three components:

- the Silicon Vertex Detector (SVX II) [21]
  It is made of three cylinders, each with 5 layers made of 12 angular segments. It is located from 2.44 cm to 10.6 cm in radius and from -43.5 cm to 43.5 cm in z-direction, which translates to a coverage of $|\eta| < 2$.
  The segments have two sided silicon strip detectors, which allow for measurements in both $r - \phi$ and $r - z$ view, thus allowing for three dimensional track reconstruction. Two layers have a small stereo angle, while the other three have a 90 degree angle. The 90 degree angle layers give a better resolution,

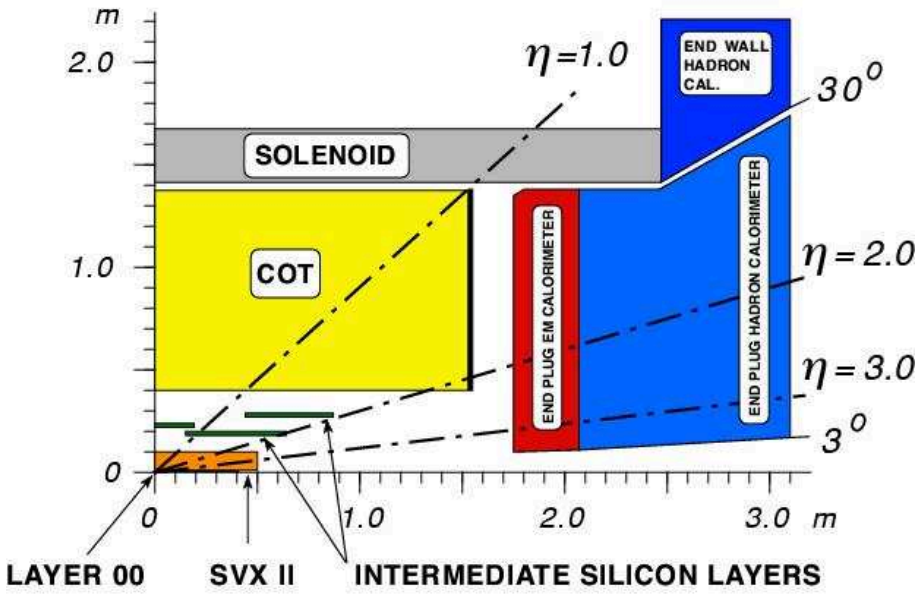

Figure 3.5: Cut through the $z - \theta$ plane of the CDF II tracking system.

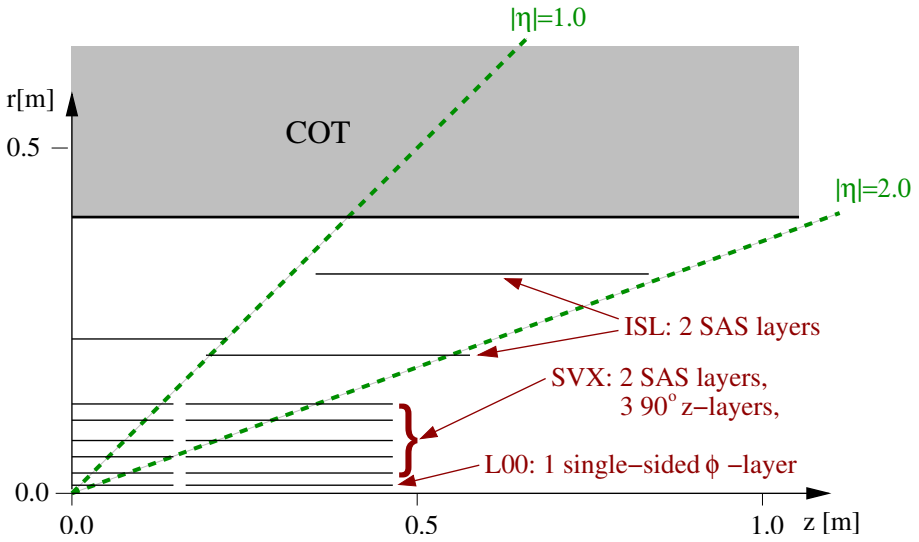

Figure 3.6: Silicon system of the CDF II detector.

while the small stereo angle layers have less combinatorial backgrond. The higher background in the 90 degree case originates from the response of the detector to several tracks through the same stripe.

- the Intermediate Silicon Layers (ISL) [22]
  consist of double-sided sensors with stereo angle stripes (SAS) with the same angles as in the SVX II. One has to distinguish between the central region with $|\eta| < 1$ with one layer and the outer region with $1 < |\eta| < 2$ with two layers.
  In the central region the ISL helps linking COT and SVX II track segments. In the outer region the ISL is used to improve forward/backward tracking.

- the Layer00 (L00) [23]
  which is mounted on the beam pipe is the innermost element of the *Silicon Detector* and is new in Run II. Due to its close distance to the beam pipe it improves the resolution in the $r-\phi$-plane. It uses single sided, radiation hard sensors which can sustain the radiation field close to the beam.

The Central Outer Tracker [24] is a cylindrical open cell drift chamber with coverage $|\eta| < 1$. Its radial range is 40 - 137 cm. 96 measuring layers are organized in eight superlayers. Alternating axial superlayers with 2 degrees stereo angle allow a three dimensional measurement of tracks through the COT.
The whole tracking system is inside a magnetic field with $B = 1.4$ Tesla, which allows momentum measurements for charged particles. It emanates from a solenoid outside the TOF detector. Due to the large radius coverage of the COT and therefore long tracks inside the tracking system the momentum resolution is quite good. The uncertainty of the transverse momentum of a charged stable particle with a transverse momentum of few GeV/c is e.g. about 0.5 MeV/c.
The silicon detector with its high precision and location at a small radius allows for a precise measurement of secondary vertices originating from the decays of long-lived particles.
From the COT with its 96 measuring layers one can get another important information, the energy loss per distance dE/dx used for particle identification. The energy loss is the result of ionisation described by the Bethe-Bloch formula. However, for relating the energy loss with the particle type, a parametrisation of the energy loss is derived from the experimental data.

3.2.3 Time of flight (TOF)

The TOF detector [25] measures the flight time of a particle. By combining it with the momentum of the particle one can calculate the mass of the particle, which can be used for particle identification.
The TOF detector is located directly outside the COT, at a radius of approximately 140 cm and consists of 216 bars of scintillator material arranged cylindrically around

the COT. With a time resolution of $\approx$ 100 ps, the TOF detector provides at least two standard deviations of statistical separation between charged pions and kaons for momenta less than 1.6 GeV/c.

3.2.4 Muon system

Due to their enormous penetration ability, muons can reach the outermost detector components, the muon chamber [26], which consists of 4 components (CMU, CMX, CMP and BMU) and the muon scintillators (CSP, CSW, CSX, MSX, BSU, TSU). All muon chambers are wire chambers operating in proportional mode. The components in detail:

- Central Muon Chambers (CMU)
  This muon chamber is used to detect muons in the central detector region and is located at the outside edge of CHA wedges covering $|\eta| < 0.6$.

- Central Muon Upgrade (CMP)
  The purpose of the CMP is to confirm the CMU tracks. The signal to background fraction for CMP is higher than for CMU, because there is more material between interaction point (IP) and CMP than between CMU and IP. The CMP is located along the walls, floor and top surface of CDF. It also covers an area of $|\eta| < 0.53$.

- Central Muon Extension (CMX)
  The CMX enlarges central muon chamber coverage to $|\eta| = 1.0$ by covering $0.6 < |\eta| < 1.0$. It is a truncated cone covering the area between BMU and CMP. The CMX is divided into two parts: the upper conical section, which covers the upper 270 degrees in ϕ, and the part of the lower 90 degrees in ϕ, which has a slightly different geometry due to the floor. The second part is called the 'miniskirt'.

- Barrel Muon Chambers (BMU)
  The BMU is used to detect muons in the forward region. It is located outside of the toroids and covers the area of $1.0 < |\eta| < 1.5$. The bottom 90 degrees of the barrel are not covered due to the support structures for the toroids. However, these part of the detector isn't used in the analysis.

  The muon scintillators consist of following subsystems.

- Central Scintillator Upgrade (CSP)
  The CSW is used for fast timing and trigger counters for CMU/CMP muons and is located outside of the CMP chambers.

- Central Scintillator Extension (CSX)
  The CSX serves the same purpose as the CSW (CSP), in this case for the

CMX muons (in the conical section). It is placed on both surfaces of the CMX chambers, but covers only the upper 270 degrees of the CMX.

- CMX Miniskirt Scintillators (MSX)
  The MSX has the same function as the CSX, in this case for the muons in the miniskirt section of CMX. It is placed on the inner surface of the CMX miniskirt chambers.

- Barrel Scintillator Upgrade (BSU)
  The BSU is used for the same purpose as the CSW, in this case for BMU muons. It is build on the outer surface of the BMU chambers.

- Toroid Scintillator Upgrade (TSU)
  The TSU provides additional triggering powers for forward muons. It is placed on the inner face of the toroids.

The combination BSU+TSU is often referred to as ISU (Intermediate Scintillator Upgrade), while BSU+TSU+BMU forms the IMU (Intermediate Muon Detector). Table 3.1 summarizes the η coverage and the minimal transverse momentum, which the muon needs to be detectable in some of the muon detectors.

	CMU	CMP/CSP	CMX/CSX	IMU
η coverage	$\lvert\eta\rvert < 0.6$	$\lvert\eta\rvert < 0.6$	$0.6 < \lvert\eta\rvert < 1.0$	$1.0 < \lvert\eta\rvert < 1.5$
Min $p_\perp$ of det. μ	1.4 GeV/c	2.2 GeV/c	1.4 GeV/c	1.4-2.0 GeV/c

Table 3.1: η coverage and the minimal p_T for a muon to be detected in the muon system detector.

3.2.5 Other detector systems

Calorimeters

Calorimeters [27, 28, 29] are used to measure the energy of particles. The sensitivity of calorimeters at CDF is not good enough to provide useful information for most b and c quark analyses. They are usually used in heavy flavor analyses as an absorber in front of the muon system. A description is included for completeness.
Two kinds of calorimeter systems are distinguished, the electromagnetical calorimeter and the hadronic calorimeter. The calorimeter systems are located outside magnetic field of the solenoid, first the electromagnetic calorimeter and then the hadronic calorimeter. This order is owned to the higher penetration ability of strongly interacting particles.
The calorimeters are designed to fully absorb the kinetic energy of the particles, except for muons or neutrinos. In CDF, they cover the range of $-3.6 < \eta < 3.6$ and $0 < \phi < 2\pi$. The whole system is built up in segments and consists of five

different calorimeter subsystems (CEM, PEM, CHA, WHA and PHA), where CEM and PEM are electromagnetic calorimeters and the other hadronic calorimeters:

- Central Electromagnetic Calorimeter (CEM)
  The purpose of CEM is to measure the energy of electromagnetic showers in the central detector. It is located outside the solenoid in the central part of the detector and covers the region of $|\eta| < 1.1$. CEM is a Pb/scintillator sampling calorimeter and 31 layers deep. The depth of this system is equivalent to 18 electromagnetic interaction lengths.

- Plug Electromagnetic Calorimeter (PEM)
  PEM measures the energy of electromagnetic showers in the plug region of the detector. It is placed outside the barrel end of the COT, one plug on each side, and covers the area of $1.1 < |\eta| < 3.6$. PEM is a Pb/scintillator sampling calorimeter, 23 layers deep and is new for Run II. The depth of this system is equivalent to 23.2 electromagnetic interaction lengths.

- Central Hadronic Calorimeter (CHA)
  CHA is used to measure the energy of hadronic showers in the central detector. It is placed in the central detector outside of CEM and covers the area of $|\eta| < 0.9$. CHA is a Fe/scintillator sampling calorimeter, 32 layers deep. The depths of this system is equivalent to 4.7 hadronic interaction lengths.

- Endwall Hadronic Calorimeter (WHA)
  WHA expands the hadronic calorimeter coverage to the endwall region. It is placed along the endwall outside of the plug and covers the region of $0.8 < |\eta| < 1.2$. WHA is also a Fe/scintillator sampling calorimeter, but only 15 layers deep. The depths of this system is equivalent to 4.7 hadronic interaction lengths.

- Plug Hadronic Calorimeter (PHA)
  PHA is used to measure hadronic showers in the plug region of the detector. It is placed beyond PEM and covers $1.2 < |\eta| < 3.6$. PHA is a Fe/scintillator sampling calorimeter, 23 layers deep and is new for Run II.

Luminosity counters

The Cherenkov Luminosity Counters or CLC provides a measurement of the luminosity. Located near the beamline next to the plug calorimeter, it uses inclusive elastic proton antiproton events to determine the luminosity either by counting 'empty bunch crossings', the ratio of bunch crossings without an interaction that creates a signal in the CLC, or by simply relating the number of measured hits in the CLC to the number of hits one would expect for exactly one such event. The accuracy of the luminosity measurement is about 5%.

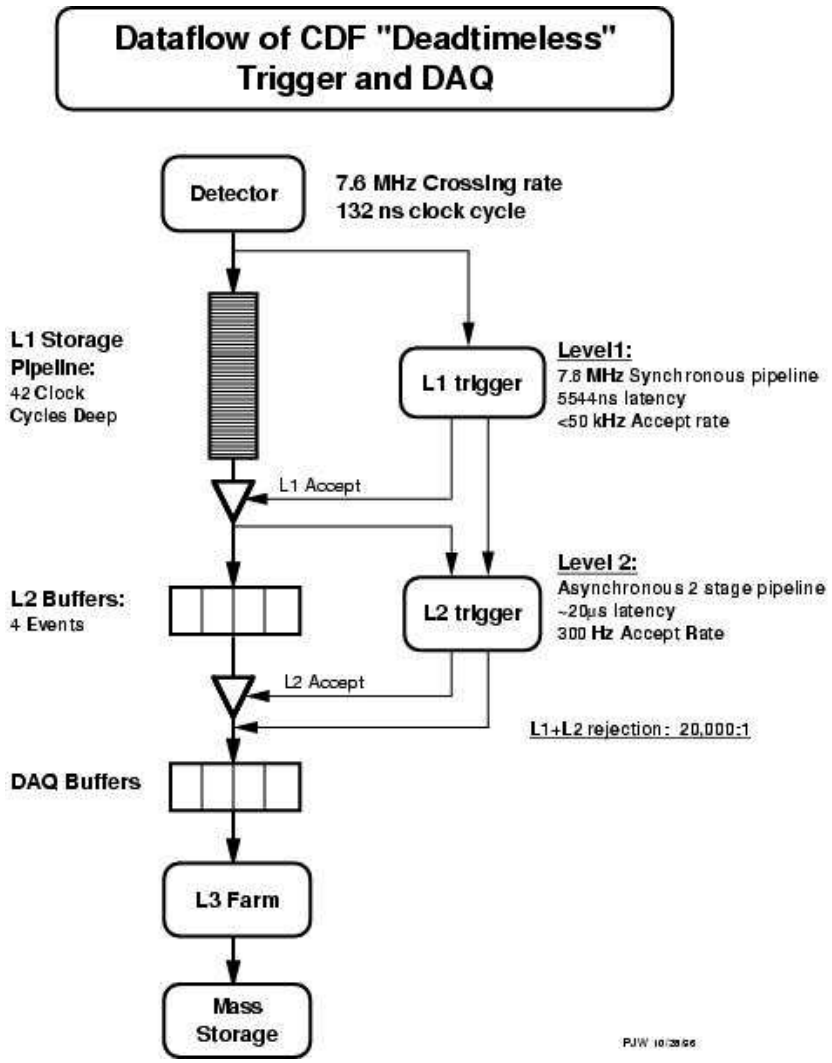

Figure 3.7: The figure shows a schematic representation of the CDF Trigger system.

3.3 Trigger-system

The cross-sections of interesting processes, especially those containing b-quarks, are small compared to cross-sections of other processes. To get enough events of the desired type, the luminosity has to be very high. It is impossible to provide the capacity to read out, process and store the information of every event. In CDF II reading out of every event would amount to about 250 GByte per second. Therefore one has to select carefully which events are worth to be stored and one focuses on events with special signatures, which are rare in general, but enhanced for the processes of interest.

At CDF this is done by a three level trigger system, a scheme of which is shown in figure 3.7. Level one is a pure hardware trigger. Level two consists of hardware and software components and level three consists just of software components. In the first two trigger steps, the rate is reduced in each level by a factor of about 100. In

the third step by another factor of five. The write out capacity is about 100 events per second.
At level one the trigger can only use muon information, calorimeter information and tracking information from the 'Extremely Fast Tracker' (XFT) [30], that finds tracks in the COT to make a selection. On level two information is added from the Silicon Vertex Trigger (SVT), which adds information from the silicon detector to the tracks found by the XFT. A detailed track reconstruction with a computer farm is performed at level 3. If the requirements for events allow too many events to pass through the trigger to be read out, which happens especially at high luminosities, a share of randomly determined events of some triggers - often b-quark physics triggers - is thrown away.

Chapter 4

Tools

In this chapter I describe some tools, that I use in the thesis. I start with the description of the NeuroBayes [31] package, which is a powerful implementation of a neural network in combination with various preprocessing methods for efficient classification. I will go on with the description of Monte Carlo simulations of the signal, that are utilised in this analysis. Finally the basics of an extended maximum likelihood fit are explained.

4.1 NeuroBayes

Only a very small fraction of the produced particles at CDF II are B_s^{**} mesons. As the detector observes only the particles living long enough to fly the macroscopic distances in which the detector is build before they decay, various backgrounds totally dominate any combination of detected particles, that might stem from a B_s^{**} signal. All of those possible combinations are called a candidate.
There are various ways to remove many candidates of the background without throwing away too much of the signal. One starts with defining various properties, that can be used for the description of a candidate, e.g. the momentum of a specific particle of the candidate combination. The simplest way to separate background and signal candidates is the application of cuts on those variables. For a more refined analysis, neural networks are useful, as they can easily take correlations between variables into account.

4.1.1 The NeuroBayes Neural Network

The basic idea of a neural network is to have nodes, that act similar as the nerve cells in the brain. Each node gets an input, does some calculation, and sends an output. NeuroBayes provides an easy framework to build a feed forward network with three layers of nodes. Feed forward declares that the nodes of the first layer

connect their output only to the nodes of the second layer, and the nodes of the second layer only to the nodes of the third layer. The variables describing the candidates are the inputs for the first layer. As in this analysis NeuroBayes is only used for classification, the third layer has just one node. The whole neural network represents a function from the variable input space $\mathcal{R}^{\mathcal{N}}$ to a single real number. In NeuroBayes this single number is between [-1,1], where minus one is the assignment as a clear background candidate and one stands for the assignment as a clear signal candidate.
The nodes process the incoming conncections by giving each connection a weight and adding them up, and transform the sum with a sigmoidal function. If the function represented by the neural network is not the desired one, the weights of the connections are changed. This is done in a training process using a sample, where the truth is known. The truth is often called target in the context of neural networks. Often Monte Carlo simulations, as described below, are used to get a signal sample. The weights are adjusted iteratively so that a loss function is minimised. The loss function used in this thesis is the entropic function E

$$\mathrm{E}(w) = |\Sigma_i \log\{\frac{1}{2}[1 + T_i o_i(w)]\}|,$$

where T_i is the target, which is one for a signal candidate and minus one for a background candidate, w is a set of weights describing the neural network and o_i is the output of the neural network for the candidate number i. If the neural network assigns all candidates correctly, E is equal to zero. If a candidate is assigned as a clear signal (background), despite being a background (signal) candidate, E is infinity.
The output o_i can be interpreted as a probability of $(1 + o_i)/2$ to be signal, if the share of signal in bins of the neural network output is a linear function of the neural network output. The interpretation for other samples, than the one used for training, is only allowed if the signal fraction in that sample is equal to that of the training sample.
As one wants to apply the neural network on a sample of candidates in which one does not know which candidates are signal, and which are background, the neural network needs to be sufficiently general. There are various methods applied in NeuroBayes, such as pruning and small random fluctuations, to prevent an event by event memorising of the training sample, and helping to learn only the general properties.

4.1.2 NeuroBayes Preprocessing

Before the values for the various variables enter the neural network, there are several preprocessing steps. The nodes of the neural network have value ranges, in which the sigmoidal activation function is rather sensitive; very high or very low inputs may just result in some form of over- or understimulation. The distribution of the variables is transformed, so that the values are in the range of high sensitivity. In this thesis the function I chose to use for that is a Gaussian with width 1 and mean

0.
As the size of the training samples is often not large enough to suppress statistical fluctuations completely, some methods to reduce their influence are part of the preprocessing. Most variable distributions used in particle physics are smooth, apart from effects like a failed measurement. With a spline fit to the signal fraction in bins of the variable in question, a regularisation can be used to suppress statistical fluctuations. If one knows, that for physical reasons the share of signal can only go monotonously up or down, this information can be used to restrict the freedom of the fit. Sometimes, e.g. when there are candidates that are missing a specific variable, one has a distinct class of events in addition to a smooth distribution. In those cases NeuroBayes allows to treat this specific class separately.
As a first step to take correlations into account, a linear decorrelation of the variables is performed analytically and linear uncorrelated combinations of the variables enter the actual neural network. Variables that contain only statistical noise and little or no information on the general structure of the samples just make it more difficult to filter out the relevant information. As the correlation of the variables to the target is calculated as well, it is possible to remove variables that have little correlation to the target.

To make it easy to give details of the individual preprocessing for each variable in the next chapter, I explain here, how the preprocessing can be denoted in a double digit number ij. j can be used to parse additional information to the fit. The numbers used in this analysis for j and their meanings are

- 4: Use result of regularised spline fit to mean values of target
- 5: Use result of monotonous regularised spline fit to mean values of target

The other digit i is used to declare what to do with the variable for special classes of candidates, e.g. when the variable is not filled for some candidates. Only one special class per variable is possible. The numbers used for i in this thesis and their meanings are

- 1: No special additional treatment, no special class is used.
- 3: Events in the special class are not considered in the fit and are treated according to the share of signal, that is contained in the special class
- 9: Events in the special class are not considered in the fit and the share of signal in the special class is ignored. For the events in this class, this preprocessing is equivalent to removing the variable from the training.

An additional preprocessing parameter allows to take the correlation to another specific variable into account for the regularisation fit.

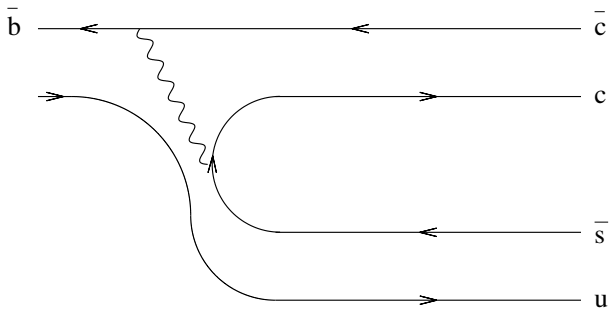

Figure 4.1: Decay of $B^+ \to J/\psi K^+$. The decay proceeds via a W^+, that decays to quarks of the second generation.

4.2 Monte Carlo Simulations

A Monte Carlo simulation provides a sample of candidates for which the values of the variables are distributed like in the experimental data including detector and trigger effects. For the purposes of this thesis, it is sufficient to have an accurate description of the B_s^{**} signal, as the experimental data can be used to get clear background candidates for the neural network training. The decay channels chosen for this thesis are

$$B_s^{**} \to B^{+(*)} K^-, \quad B^+ \to J/\psi K^+, \quad J/\psi \to \mu^+\mu^-, \tag{4.1}$$

$$B_s^{**} \to B^{+(*)} K^-, \quad B^+ \to \overline{D^0}\pi+, \quad \overline{D^0} \to K^+\pi^-, \tag{4.2}$$

$$B_s^{**} \to B^{+(*)} K^-, \quad B^+ \to \overline{D^0}\pi + \pi^+\pi^-, \quad \overline{D^0} \to K^+\pi^-. \tag{4.3}$$

The $B^{+(*)}$ denotes here either a B^+ or a B^{*+}, as the B_s^{**} mesons can decay in both of them depending on their angular momentum composition as described in chapter 2. A B^{*+} however, decays into a B^+ and a photon, that has an energy of 45.7 ± 0.9 MeV/c^2 [14]. In ref. [32] a smaller uncertainty is given, but ref. [14] measures exclusively the mass difference for the charged case, while ref. [32] gives flavour averaged measurements ($\overline{b}q$ for $q = u, d, s$). In all decay chains the B^+ decays via weak interaction.

In the decay chain of eq. 4.1 the W^+ decays into the second generation doublet, a c and a $\overline{s}$. The latter forms with the spectator u the kaon, while the first forms with the $\overline{c}$, that come out of the $\overline{b}$, a J/ψ. This decay is graphically represented in figure 4.1. The J/ψ decays by the electromagnetic interaction into two muons. As muons are rather rare, they can be used to trigger events containing a J/ψ. Overall the final state consists of K^- (from the B_s^{**} decay), K^+, and a $\mu^+\mu^-$ pair, which as charged particles can all be detected in the silicon detector and the COT.
In the decay chains of eq. 4.2 and 4.3, the W^+ decays into the first generation doublet u and $\overline{d}$. The $\overline{c}$ ends up with a u to form a D^0, while the other decay products form charged pions or short living resonances that decay immediately into two pions. The $\overline{D^0}$ decays to a pion and a kaon again through the weak interaction, so that the end state consists of K^- (from the B_s^{**} decay), π^+ or $\pi^+\pi^+\pi^-$, and

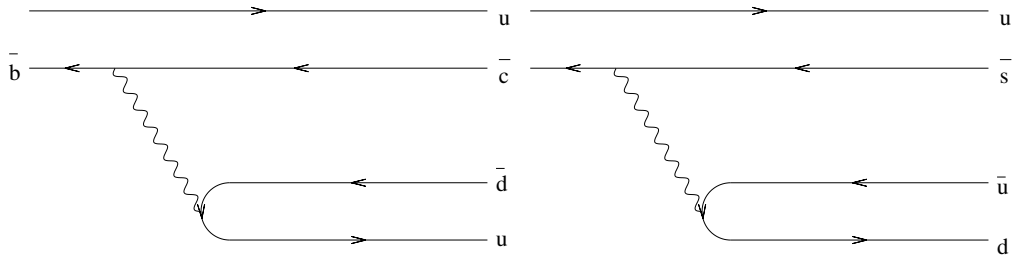

Figure 4.2: Decay of $B^+ \to \overline{D^0}\pi^+$ and $\overline{D^0} \to K^+\pi^-$. The decays proceed via a W^+ and a W^- bosons, that decay into quarks of the first generation.

the K^+ π^- from the $\overline{D^0}$ decay. This decay is presented in figure 4.2. The strong interaction can create additional particles for decay chain eq. 4.3, or the $B+$ decays into short living resonances, that decay further into pions. One possibility is an a_1^+, another possibility is that a pion and a ρ^0 are created, where the latter decays immediately into two pions. Again all particles are charged after the last decay. The trigger for the decay chains of eq. 4.2 and 4.3 is based on the lifetimes of the B^+ and $\overline{D^0}$ particles. Due to their lifetimes B^+ and $\overline{D^0}$ mesons can fly a distance within the resolution of the detector before they decay. Their decay products then can form a secondary vertex, which activates the trigger. The trigger may react either on the particles from the $\overline{D^0}$ decay or include one or two of the pions from the B^+ decay.

For this analysis four different Monte Carlo samples are used. One sample includes only candidates going through the decay chain:

$$B^+ \to J/\psi K^+, J/\psi \to \mu^+\mu^-$$

Charged kaons and muons live sufficiently long, so that most of them reach the detector, and are therefore considered as stable in this analysis. The Event Generator used for this sample is the PYTHIA [33] with parameter MSEL set to 1.
The complete decays chain 4.1, 4.2, and 4.3 are simulated using a proprietary generator called BGenerator [34], that has been developed in CDF.
Both PYTHIA and BGenerator are only used to create values for the first particle of the respective decay chain. The further decays of the generated particles at the beginning of the decay chain are done with the EvtGen [35] package, that was originially developed by the BaBar collaboration. EvtGen is wide spread, and improved or new measurements are included fast. It treats effects of the spin structure on the decay structure appropriately, which helps in the decay of the J/ψ. In the decay chain of eq. 4.3, the three pions have an addional substructure. In 90 % of the decays the three pions form an a_1^+, in 5% of the decays two of the pions form a ρ^0 meson, and in the last 5% of the decays, there is no resonant structure for the pions. The decay tables that were used to simulate the decay of eq. 4.1 can be found in appendix A. The decay table for the decay chains of eq. 4.2 and 4.3 can be found in appendix A
A realistic simulation of the CDF II detector effects based on the GEANT 3 [36] package then provides signal candidates in the same format as the raw data coming from the detector. The further treatment of this information for reconstruction is

exactly the same as for the data coming from the experiment, and is described in the next chapter.

4.2.1 Details of the PYTHIA based sample

In this case in the first step originally not only signal candidates created, but full events, that may not even contain a b-quark. However, only the events that do contain at least one b-quark, that hadronises into a B^+ are kept. This B^+ is then redecayed using EvtGen, forcing it into the selected decay modes. The other b-quark is redecayed as well with EvtGen, but this has no influence on this analysis. All other particles, that are part of a realistic proton anti-proton collision, are kept as PYTHIA performs their hadronisation and decay. However, those additional particles aren't used in this thesis. The reason to use such a full simulation is the availability of the Monte Carlo sample from a different analysis.

4.2.2 Details on BGenerator based sample

The BGenerator is utilised in the simulation of the B_s^{**} mesons. It doesn't simulate full events, but just the B_s^{**} signal. This is much faster than using PYTHIA and sufficient for the purpose of this analysis. BGenerator needs transversal momentum and rapidity spectra for the B_s^{**}, from which it is choosing randomly values for the generation. The output, that is needed to feed EvtGen, are the four-vectors of simulated B_s^{**} mesons.
In EvtGen, the mass and the energy of the B_s^{**} mesons are reset to a flat distribution between the kinematic threshold and values far larger than the expected signal before the decay. The unknown lifetime of the B_s^{**} is set to zero as one expects far lower lifetimes than measureable from a strong decay. The EvtGen simulation then enters the detector simulation as input.

4.3 Extended Maximum Likelihood Fit

If there is enough statistics, a χ^2 minimisation is a fast and accurate method to fit a distribution. However, one loses information due to the necessary binning. More over the χ^2 minimisation method is based on the idea that the uncertainty on the amount in a given bin is equal to the square root of this amount. This is justified in case of large numbers, but with little statistics, this assumption doesn't hold. A maximum likelihood fit doesn't need binning and can be performed accurately on a small statistic.

4.3.1 Parameter Determination

To perform a maximum likelihood fit, one needs a fit model in form of a function f(x|**p**s), that describes the set **x** of measurements x_i one wants to fit in dependence on a set **p** of parameters. The parameters are varied, so that the probability to yield the given set of measurements is maximised, if it is the result of a random process with the fit model as underlying probability distribution. First I want to look at fits, for which the total number N of measurements x_i is not the result of a random process, but fixed. In this case the parameters are only allowed to change the shape of the fit model. The fitted number of measurements is $A(\mathbf{p}) = \int_{x_0}^{x_1} f(x|\mathbf{p})dx$, where x_0 and x_1 limit the range in which the fit model is assumed to describe the data. For eliminating the dependency on this fitted amount A(**p**), the fit model is normalised by dividing it by A(**p**). The probability to measure x_i is $f(x_i|\mathbf{p})/A(\mathbf{p})$ dx. Under the assumption that the measurements are independent, the likelihood to get the set **x** of measurements is the product of the probability for each x_i. This product is called likelihood L. So this product has to be maximised, to maximise the likelihood. For technical reasons one takes the negative logarithm of this product. The calculation of products on computers is problematic, when very small numbers are involved, while sums are less affected by machine accuracy. Additionally the logarithm is multiplied by two. So now one has to minimise the following sum:

$$\mathcal{L} := -2\ln(L) = -2\Sigma_i \ln[f(x_i|\mathbf{p})/A(\mathbf{p})]. \tag{4.4}$$

Now I extend that formula to cases, when the total number of measurements N is the result of a random process. I assume, that the underlying distribution for that random process is a Poisson distribution. The mean value of the Poisson distribution is assumed to be N, but the fitted amount A(**p**) now varies, and can differ from N, according to the Poisson probability distribution of

$$P(A(p)) = A(\mathbf{p})^N e^{-A(\mathbf{p})}/N!. \tag{4.5}$$

This probability to have A(**p**) measurements is inserted as factor to the product. Taking the negative logarithm times two of the product gives

$$\mathcal{L} = -2\Sigma_i \ln[f(x_i|\mathbf{p})/A(\mathbf{p})] - N \cdot \ln[A(\mathbf{p})] + A(\mathbf{p}) + ln[N!]. \tag{4.6}$$

The last summand is a constant and can be neglected for minimisation. The sum in the first summand factors yields an own term $+N \cdot ln[A(\mathbf{p})]$, which cancels with the second summand. So the final term that has to be minimised is

$$\mathcal{L} = -2\Sigma_i \ln[f(x_i|\mathbf{p})] + A(\mathbf{p}). \tag{4.7}$$

As f(x|**p**) is not normalised any more, it can grow the logarithm, but as its integral grows along, the second summand with the opposite sign penalises arbitrary growth. Eq. 4.7 provides the final negative log likelihood $\mathcal{L}$, that is used for the minimisation, which is performed using the MINUIT package [37] provided by the ROOT [38] package.

4.4 Significane Determination

The maximum likelihood fit doesn't offer an absolute probability for the correctness of a fit. It is possible, though, to test one hypothesis against another, e.g. a fit with and without a signal contribution. Assuming that a priori both hypotheses are equally likely, one can do that in the following way. One calculates the likelihood for both hypotheses and divides the higher value by the lower value. The result is the a posteriori probability that the hypothesis with the higher likelihood is the correct one. If the hypotheses differ by more than the fixing of one parameter, this additional freedom has to be taken into account, but in this thesis I limit myself to test the significance of one parameter.
From the negative log likelihood $\mathcal{L}$ one can obtain the same, by taking the difference. This difference equals the square of the significance, as

$$\frac{\mathrm{L}_1}{\mathrm{L}_0} = \exp\{-\frac{1}{2}[-2\ln(\mathrm{L}_1) - (-2\ln(\mathrm{L}_0))]\} = \exp\{-\frac{1}{2}[\mathcal{L}_1 - \mathcal{L}_0]\}. \tag{4.8}$$

4.5 Limit Determination

If a measurement is not significant, e.g. a parameter p doesn't differ from zero by more than three standard deviations of its uncertainty, one may want to determine a limit. To do so, the Bayesian theorem can be utilised. First an a priori distribution $\pi(\mathrm{p})$ has to be defined. In this thesis I use

$$\pi(\mathrm{p}) = \begin{cases} 1 \text{ for } \mathrm{p}_0 < \mathrm{p} < \mathrm{p}_1 \\ 0 \text{ otherwise.} \end{cases} \tag{4.9}$$

Then one determines likelihood distribution $\mathrm{L}(\mathbf{x}|\mathbf{p})$, and calculates the posteriori probability density function as

$$\mathrm{p}(\mathbf{x}|\mathbf{p}) = \frac{\mathrm{L}(\mathbf{x}|\mathbf{p})\pi(\mathrm{p})}{\int \mathrm{L}(\mathbf{x}|\mathbf{p}')\pi(\mathrm{p}')\mathrm{dp}'}, \tag{4.10}$$

where $\mathbf{p}'$ is the set of parameters with parameter p fixed to p'. The denomintor of the fraction in eq. 4.10 is only a constant, to normalise the posteriori probability.
A limit with a credibility level of $1-\beta$ can be obtained now, varying the borders of the integral over the posteriori probability density function, so that it equals the fraction $1-\beta$. In this thesis, the error probability β is set to 5%.

Chapter 5

Data Selection

The data selection is done in several steps. In a first step candidates for B^+ and B_s^{**} mesons have to be reconstructed from the raw experimental data coming out of the CDF II detector. As most of the candidates are no real B_s^{**} mesons, statistical methods are used to enrich real mesons in the sample. The procedure to do so is again split into two steps. The first step is to enrich B^+ mesons and the second create a more refined B_s^{**} meson enriched sample.

5.1 Reconstruction

The present analysis uses collision events detected by the CDF II detector from March 2002 to April 2008. Two different triggers are used to collect the data samples, one is the dimuon trigger (with the $B^+ \to J/\psi K^+$ in its decay chain), the other is the displaced-vertex trigger (for the samples with $B^+ \to \overline{D^0}\pi^+$ and $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ in the decay chain).
The dimuon trigger [18] requires two tracks of opposite charge matched to track segments in the muon chambers, where the mass of the pair is consistent with the J/ψ mass. The transversal momentum, the momentum component perpendicular to the proton beam direction, of the muons that shall activate the trigger, has to be at least 1.5 GeV/c and a maximum opening angle of $\Delta\Phi < \frac{3}{4}\pi$ is enforced in order to prevent cosmic muons to activate the trigger. The mass window is between 2.7 GeV/c^2 and 4.0 GeV/c^2, which allows as well ψ(2S) mesons to activate the trigger, but they aren't used here.
The displaced-vertex trigger [39] requires two tracks with large impact parameters. Additionally, the intersection of the tracks has to be displaced from the interaction point and a minimum transverse momentum is required for each track. At level one the requirement is to have two tracks with a transverse momentum larger than 2 GeV/c each with an opening angle $\Delta\Phi < \frac{3}{4}\pi$ as in the dimuon trigger. At level two additional requirements, that have to be confirmed at level three, are, that each track has an impact parameter d_0 in the range 100 μm $< |d_0| <$ 1 mm, the $\chi^2 < 25$ for a fit to a common intersection point of the tracks, the intersection point has to

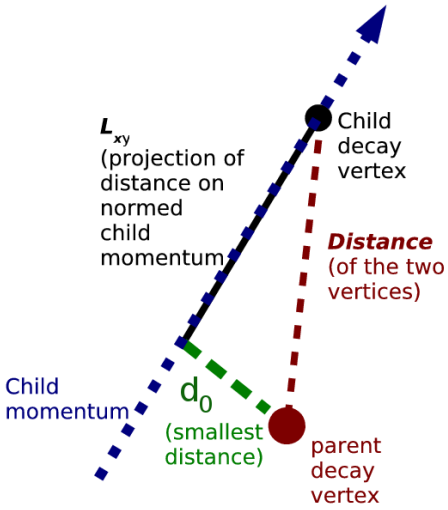

Figure 5.1: Explaination of L_{xy}, d_0, and the distance between its decay vertex and a parent vertex.

have a transverse decay length of at least 200 μm, and the opening angle is further constrained to $2° < \Delta\Phi < \frac{1}{4}\pi$. The reason to require as well a minimum opening angle is, to avoid the activation of the trigger by a single badly measured track, that is split into tracks by the tracking algorithm. Additional there is a requirement on the sum of the transversal momentum of these tracks, which varies between 4.0 GeV/c and 6.5 GeV/c, depending on the instantaneous luminosity.
The meaning of transverse decay length L_{xy} is explained in figure 5.1. The cartoon shows the quantities in the r-ϕ plane, because measurements in this plane are much more precise than orthogonal to it. The distance is simply the geometric distance between the place where a particle decays, e.g. a J/ψ meson, and the place where it orignated, e.g. the decay vertex of a mother particle, usually the primary vertex of the orginal collision, or if indicated, to a particle with lifetime like the B^+.
The impact parameter (d_0) is the smallest gap between the first vertex and a line through the secondary vertex with the direction of the momentum in this plane (p_t). L_{xy} is the projection of the distance on the direction of the particle's transverse momentum.

For the offline reconstruction, a good run list provided by the Data Quality Managment [DQM] group [40] ensures, that only experimental data of runs is used, for which the relevant detector parts worked well. The final data set corresponds to a total integrated luminosity of $L = 2.8$ fb^{-1}.
Tracks are refitted in the samples of these triggers separately for pion, kaon, and muon mass hypothesis to properly take into account the energy loss and multiple scattering of a given particle. The scaling of the COT covariance matrix, the value of the magnetic field and the material description, that one needs to fit tracks to the hits in the COT, are done as adviced by the internal CDF Note 6905. For all tracks in the dimuon trigger 10 hits in the axial and 10 hits in the stereo layers of the COT and at least 2 axial hits in the SVX II are required. This requirement is

not used in the displaced-vertex trigger.
Having now reconstruced candidates for particles that live long enough to hit the detector, in the next step we go up the decay chains, and reconstruct J/ψ and $\overline{D^0}$ candidates.
For the J/ψ mesons all muon combinations of opposite charge are refitted with the muon tracks constrained to originate from a common vertex. Candidates with a mass between 2.9 and 3.3 GeV/c^2 and a χ^2 of less than 30 of the fit are kept. $\overline{D^0}$ mesons are formed of a kaon and a pion of opposite charge again in a fit constrained to come from a common vertex. Kaons, pions, and protons can not be distinguished with great accuracy, so at this point every track matching above described quality standards is assumed to be a kaon or a pion with every possible combination. Candidates with a mass between 1.77 and 1.97 GeV/c^2, a χ^2 of the fit of less than 15 and a transverse decay length significance of at least 3 are kept in the displaced-vertex trigger sample. The significance is the value of a variable divided by its uncertainty, while here the event-by-event uncertainty is used.
J/ψ candidates are now combined with a kaon hypothesis refitted track and all three tracks (the two muons and the kaon) are fitted to the common vertex of a B^+ as the J/ψ has no considerable lifetime. The mass of the two muons in this fit is constraint to the J/ψ world average mass of 3096.9 MeV/c^2 [32]. All candidates with a χ^2 less than 50 and a mass between 4.6 and 6.8 GeV/c^2 are stored for the next step.
In the displaced vertex trigger sample, each $\overline{D^0}$ candidate is combined with a pion hypothesis fitted track and a vertex fit with the $\overline{D^0}$ mass constraint is performed. B^+ candidates in this sample are stored, if the mass is between 4.4 and 6.6 GeV/c^2, the χ^2 of the fit is less than 25 and the transverse decay length significance at least four. However, before using a neural network additional cuts are applied. For the second decay channel 3 pions are added, instead of just one, and again the additional cuts are applied, that are explained below.
To obtain B_s^{**} candidates, a track is added, which is not refitted under the kaon hypothesis, but for the calculation of its variables, the kaon hypothesis is used. The reason for this is, that the experimental raw data sample in the displaced-vertex trigger is very large, and the refit of the track would have been much effort for minimal improvement. No additional fit to constrain the particles to the primary vertex is applied, the kinematic for the B_s^{**} candidates is calculated from the four-momenta of the B^+ candidate and the kaon.

5.2 Explanation of selection variables

In this section I explain selection variables that have not been discussed so far.
For using the time of flight or the energy loss as particle identification [PID], one compares different hypotheses of having a pion, proton, kaon, electron, or muon. At the CDF II experiment the expected time of flight and energy loss are calibrated from data by using tracks from decays that are easy to identify. Knowing the distributions of time of flight and energy loss for the various hypotheses, one can construct a likelihood ratio. The relative frequency with which the various particles

are produced in general, are included as apriori information in that ration. Unfortunately, the calibration of energy loss for low transverse momentum particles has only been done up to 1.3 fb^{-1}. So the validity of the likelihood ratio from the general CDF software is only given for that part of the data. To have a PID variable for the whole data set, I construct a variable only from the time of flight information. With the calibrated central value and the resolution, one can create the pull, which is the difference to the central value divided by the resolution. To create a probability from the pulls I assume a Gaussian resolution, which gives the probability of a track to be a kaon as:

$$P_{kaon} = \frac{0.2e^{-\frac{(pT)_K^2}{2}}}{0.2e^{-\frac{(pT)_K^2}{2}} + 0.7e^{-\frac{(pT)_\pi^2}{2}} + 0.1e^{-\frac{(pT)_p^2}{2}}}.$$

pT is denominating the pull of the time of flight information under the assumption the particle is a kaon for the subindex K, pion for the subindex π, and a proton for the subindex p. As a priori share I have assumed 0.1 for protons, 0.2 for kaons, and 0.7 for pions.

Two other variables, that are used in this thesis are the Gottfried-Jackson [GJ] moment and the Collins-Soper [CS] moment. I define them only for the kaon from the B_s^{**} decay. The GJ moment is the angle between the momentum of this kaon, and the momentum of the orginal colliding proton in the B_s^{**} centre of mass system. The CS momentum is the angle between this kaon and the bisector of the momentum of the proton and the negative momentum of the anti-proton. In figure 5.2 the reference axes for both moments are explained graphically.

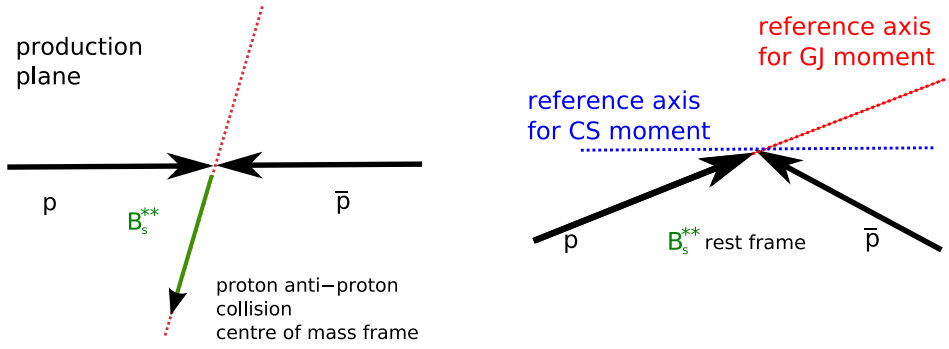

Figure 5.2: Collins-Soper and Gottfried-Jackson moment axes. To get them, one has to boost the proton and the anti-proton into the rest frame of the B_s^{**} meson. Then one takes either the proton momentum or the bisector between the proton momentum and the negative anti-proton momentum as reference axis.

The helicity angle exists only for particles that stem from a grandparent particle. It is an angle defined in the restframe of the parent particle between the directions of momentum of the grandparent particle and the particle itself.

5.3 Preselection of B^+ candidates

By far the most candidates that come out of the reconstruction are background and don't represent a real B_s^{**}. This makes a statistical analysis before further enriching signal candidates in the sample impossible. A combination of cuts and the utilisation of a sophisticated neural network allows such an enrichment.
The structure of the background is again splitted in two very different types. One type is a combination of a real B^+ meson with a random track, and the other consists just of a totally arbitrary combination of random tracks. The latter type makes up by far the largest part. Due to that, the use of the separation tools to distinguish directly any background from signal, may yield very little separation power between the the background containing a real B^+ and the signal of B_s^{**} mesons. Therefore in a first step, events that contain a real B^+, independent if it stems from the decay of a B_s^{**} or not, are enriched. In a second step, the B_s^{**} mesons are singled out from the remaining more challenging background.

5.3.1 Preselection in the $B^+ \to J/\psi K^+$ decay

For the preselection of the $B^+ \to J/\psi K^+$ decay a neural network is trained on the combination of experimental data and Monte Carlo candidates. In the training two classes of candidates are needed, one for the background and another one for the signal. The signal candidates are taken from the PYTHIA MC with full CDF detector simulation as described in the previous chapter. The simulation of the combinatorial background is generally a complicated task in the enviroment of hadron colliders and therefore events from the B^+ mass distribution sidebands are used as training pattern for the background. The regions of B^+ invariant mass from 5.190 GeV to 5.240 GeV and from 5.320 GeV to 5.370 GeV are used as the background pattern in the neural network training. In terms of the invariant mass resolution, those corresponds approximatelly to range from $3\,\sigma$ to $7\,\sigma$ away from the B^+ mass. The training is done only on the first 1.0 fb^{-1} of experimental data, but the patterns of background and signal don't vary too much over time, as the fundamental physical process doesn't change. A little variation comes from the variation in the trigger requirements, and the higher luminosities in the later runs, but these effects are small.

The variables in the list below are combined in the neural network for discriminating between real B^+ mesons and arbitrary combinatorial background. The number in brackets gives the preprocessing number. The meaning of them is explained in the previous chapter and in more detail in ref. [31]. The used variables are

1. the impact parameter of the B^+ (14),
2. the fit probability of the B^+ kinematic fit with vertex constraint (15),
3. the significance of transverse decay length of the B^+ (15),

4. the transverse momentum of the J/ψ (14),
5. the impact parameter of the J/ψ (14),
6. the significance of transverse decay length of the J/ψ with respect to the beamline (15),
7. the transverse decay length of J/ψ with respect to B^+ decay vertex (14),
8. the transverse momentum of the kaon (14),
9. the significance of the impact parameter of kaon (14),
10. the cosine of the angle of the kaon momentum in the center-of-mass frame of the B^+ relatively to momentum of the B^+ in laboratory frame (15),
11. the cosine of the helicity angle of the muon with the higher transverse momentum (14),
12. and the pseudorapidity of the kaon (14).

In order to avoid that neural network learns to calculate the mass and loose generalization power, the variables 4, 8, 10 and 12 are transformed to remove their dependence on the invariant mass of the candidate. This is done by fitting the profile of these variables with respect to the B^+ mass in a χ^2 fit with a linear function and dividing the variable by the result of the fit at the B^+ mass of the candidate.

From figure 5.3 where the purity as a function of the neural network output is

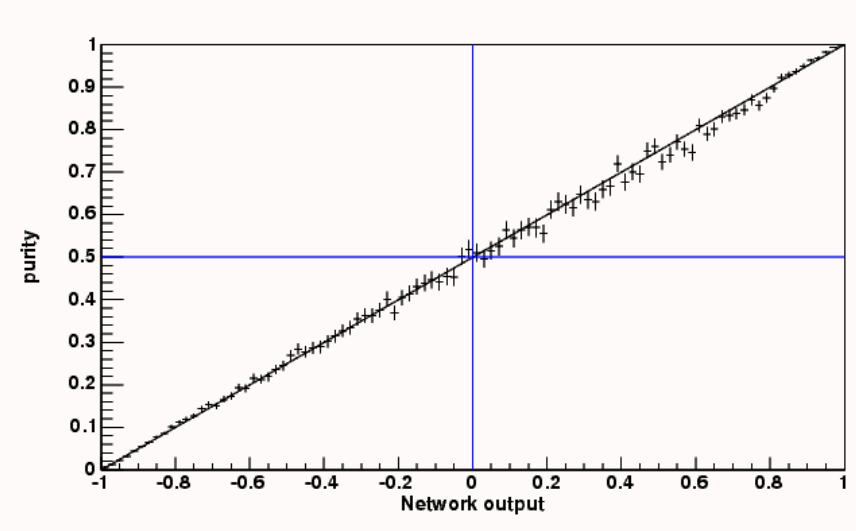

Figure 5.3: Purity as a function of the neural network output for the $B^+ \to J/\psi K^+$ neural network using the training sample.

shown, one sees, that the neural network is well trained as the points lie on the diagonal. The purity is defined as the signal share of the full training sample after

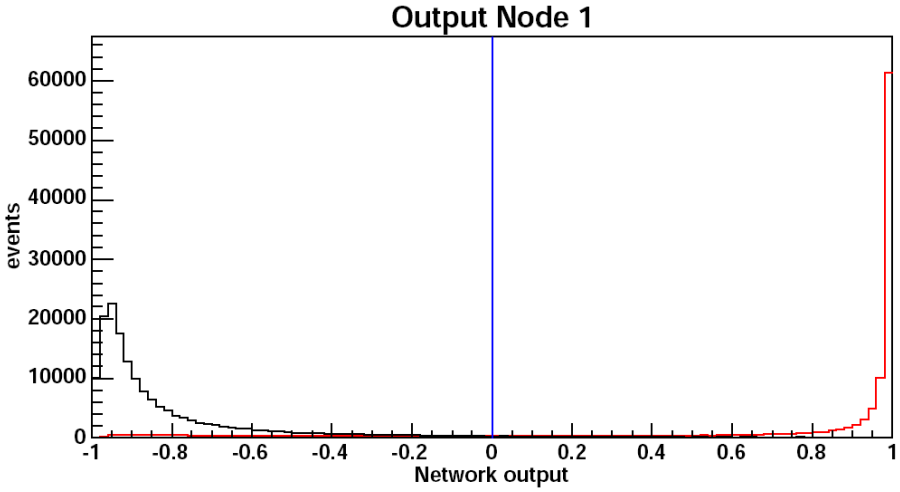

Figure 5.4: Neural network output distribution for signal (red) and background (black) events in the $B^+ \to J/\psi K^+$ decay, using the training sample.

applying a cut on the neural network output. Figure 5.4 shows the neural network output distribution for the signal and background candidates using the training sample. The neural network is able to distinguish well between signal and background candidates, as the two distributions are mostly separated.

Having a well trained neural network, the full sample of the $B^+ \to J/\psi K^+$ candidates can be classified. The invariant mass distribution before the neural network classification is shown in Figure 5.5. A χ^2 fit using a Gaussian for the signal, and a linear function for the background is performed. The number of signal events in the sample for the first 1.0 fm^{-1} is about 36000. After selecting candidates with a neural network output larger than 0.5 about 28000 signal candidates are retained while 98.5 % of the background candidates are removed. The resulting invariant mass distribution is shown in figure 5.6. The width of the Gaussian is fitted as 11.49 $\pm$ 0.12 MeV/c^2, and the mass of the B^+ is fitted as 5278.47 $\pm$ 0.12 MeV/c^2. The cut is not further optimised as the main goal of this cut is to remove obvious background before the next step, while keeping most of the signal.

5.3.2 Preselection in the $B^+ \to \overline{D^0}\pi^+$ decay chain sample

To preselect the B^+ mesons in the $B^+ \to \overline{D^0}\pi^+$ decay channel we use a different strategy. If the amount of background compared to signal is not too high and enough statistics available, it is possible to train a neural network with sideband subtraction.
A high statistics is in all B^+ channels available, but the special structure of the displaced-vertex trigger allows to remove a rather big part of the background without losing much of the signal, since most B^+ mesons reconstructed in the displaced-vertex trigger sample have a secondary vertex distinguishable from the primary vertex.

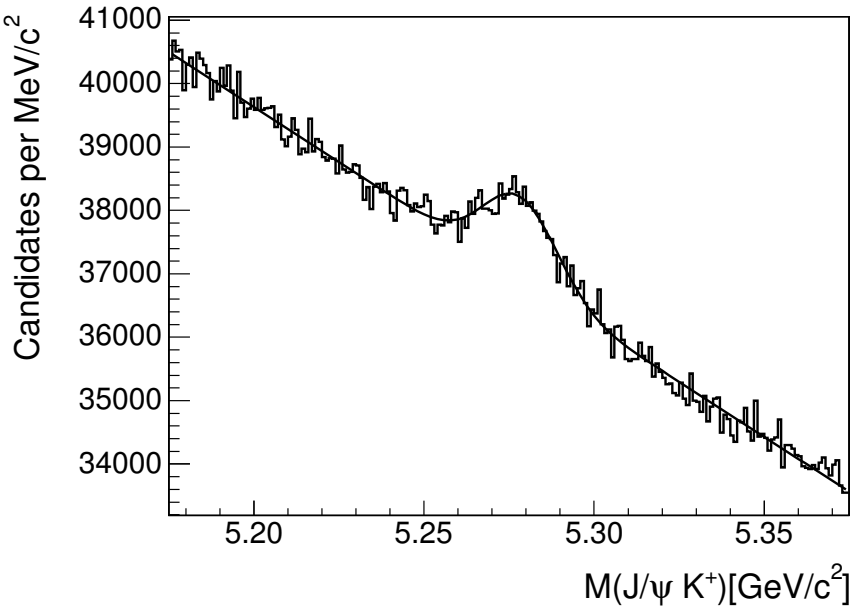

Figure 5.5: Invariant mass distribution of the B^+ candidates before neural network classification. Note that the zero is suppressed in the plot.

The NeuroBayes package allows a training using weights, which can be also negative. For the training two mass ranges from the B^+ invariant mass are used. First one starts from 5.24 GeV and ends at 5.31 GeV and is refered later as the signal region, while region from 5.325 GeV to 5.395 GeV is refered as the background region. Then to train the neural network, the signal region is used with weight 1 as signal pattern. The background region is used once with negative weight corresponding to the ratio of background in the signal region as signal pattern and once with weight 1 as the source of background pattern.

In order to make the background subtraction more stable it is useful to supress the background by some precuts, which remove only a marginal part of the signal. The following precuts are applied:

- The absolute value of the impact parameter of the B^+ has to be smaller than 0.0075 cm.
- The significance of the transverse decay length of the B^+ has to be higher than six.
- The absolute value of the impact parameter of the $\overline{D^0}$ has to be higher than 0.0025 cm.
- The significance of the transverse decay length of the $\overline{D^0}$ with respect to the B^+ decay vertex has to be larger than -4.
- The transverse momentum of the pion of the B^+ decay has to be higher than 800 MeV.

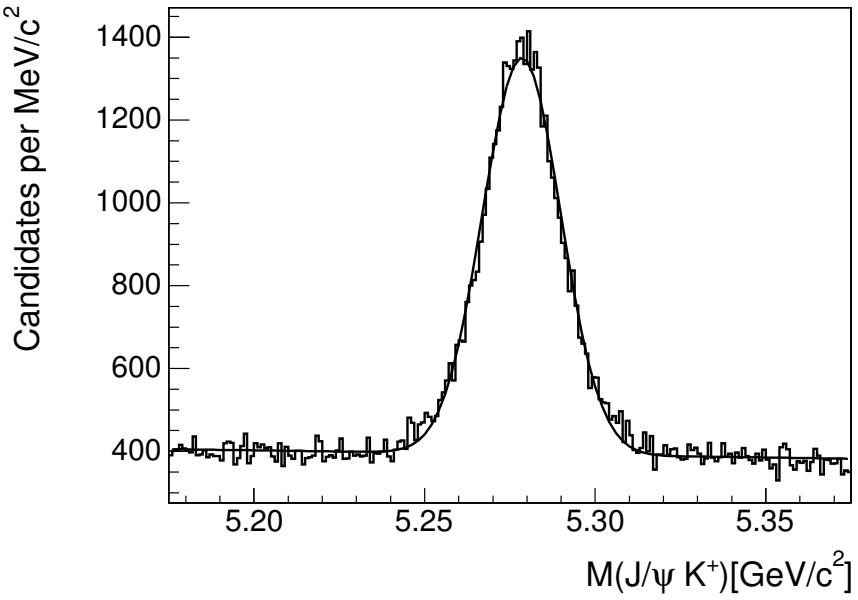

Figure 5.6: The invariant mass distribution of the B^+ candidates after a cut on the neural network output at 0.5. The zero is suppressed again.

- The charge of the pion from the $\overline{D^0}$ decay has to be opposite to the charge of the pion from the B^+ decay. I neglect doubly Cabbibo suppressed decays, because a negligible amount of signal is hidden in a similar amount of background as in the non suppressed channel.

After applying the cuts, the variables are filled into a neural network. The numbers in brackets are again the preprocessing numbers. The variables are

1. the absolute value of the impact parameter of the B^+ (15),
2. the fit probability of the B^+ kinematic fit with vertex constraint (35),
3. the significance of the transverse decay length of the B^+ (14),
4. the transverse momentum of the $\overline{D^0}$ (14),
5. the absolute value of the impact parameter of the $\overline{D^0}$ meson (15),
6. the fit probability of the $\overline{D^0}$ kinematic fit with vertex constraint (35),
7. the significance of the transverse decay length of the D^0 with respect to the primary vertex (14),
8. the significance of the transverse decay length of the D^0 meson with respect to the decay vertex of the B^+ (14),
9. the transverse momentum of the pion from the B^+ decay (14),

10. the significance of the impact parameter of the pion from the B^+ decay with respect to the B^+ decay vertex (14),

11. the cosine of the angle of the pion momentum from the B^+ decay in the centre-of-mass system of the B^+ relatively to the momentum of the B^+ in the laboratory frame (14),

12. the cosine of the helicity angle of pion from the $\overline{D}$ decay (14),

13. the transverse momentum of the kaon from the $\overline{D}$ decay in the case it is smaller than 2.0 GeV/c (95),

14. the transverse momentum of the kaon from the $\overline{D}$ decay in the case it is higher than 2.0 in GeV/c (94),

15. the PID likelihood ratio of the kaon in the case there is time of flight information (94),

16. the PID likelihood ratio of the kaon in the case there is no time of flight information (94).

The reason to split the transverse momentum of the kaon from the $\overline{D^0}$ decay into two variables is that it has a very sharp distribution at 2 GeV/c, as sometimes it activates the trigger, and sometimes the trigger is activated by other particles. The likelihood ration of the PID information looks different, when only constructed by energy loss information, so this variable is splitted as well. The neural network is trained on the first 1.0 fb^{-1}, where the energy loss information is calibrated. The neural network loses a bit of separation power on the part of the experimental data, where the calibration isn't done, but the effect is fairly small, as there are enough other variables to get a relatively clean $\overline{D^0}$. Figure 5.7 shows the purity as a function of the neural network output. The neural network is reasonably well trained, as the purity is about a linear function of the neural network output again. From figure 5.8 one can again conclude that the separation power of the neural network is good. The reason for the occurance of negative bin contents is the method, that was used. Statistical fluctuations lead to small differences between the subtracted background from the B^+ mass sideband, and the background under the signal. Due to this, it is possible that in a bin slightly more is subtracted than the positive contribution.
In figure 5.9 the invariant mass distribution of the B^+ candidates before a cut on the neural network output is shown while in Figure 5.10 after a cut of -0.2. The plots are made again with 1.0 fb^{-1}. on the neural network output. From the χ^2 fit with a single Gaussian for signal and a linear background I find, that we keep $\approx$ 27000 signal events out of the $\approx$ 28000. On the other hand only 36 % of the background events survived this cut. The width of the Gaussian is fitted as 17.38 $\pm$ 0.13 MeV/c^2, and the mass of the B^+ is fitted as 5278.20 $\pm$ 0.13 MeV/c^2.

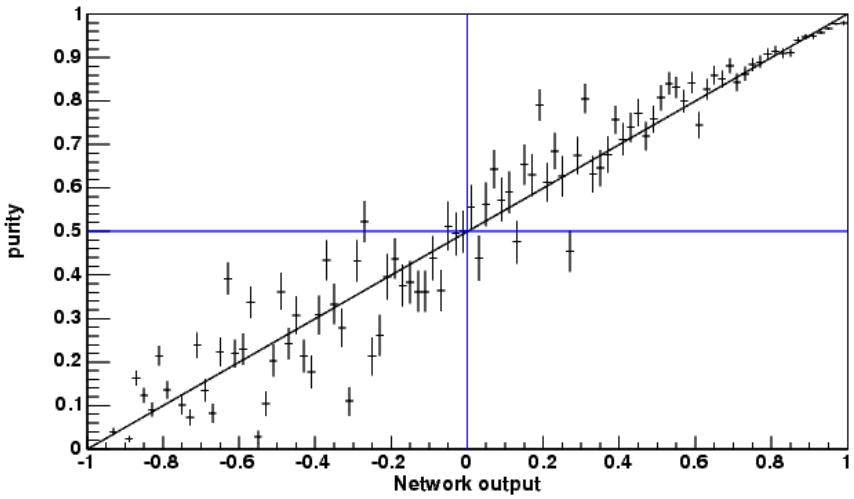

Figure 5.7: Purity as a function of the neural network output for the $B^+ \to \overline{D^0}\pi$ neural network using the training sample.

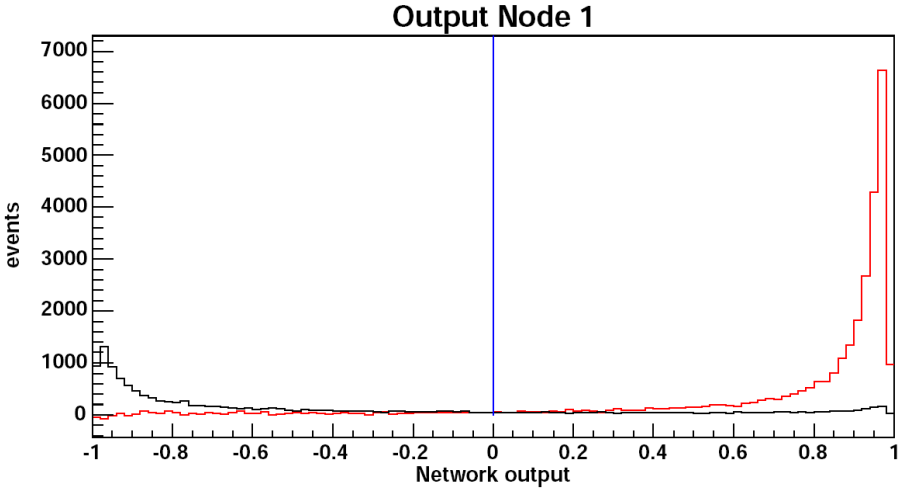

Figure 5.8: Neural network output distribution for signal (red) and background (black) events in the $B^+ \to \overline{D^0}\pi$ decay, using the training sample.

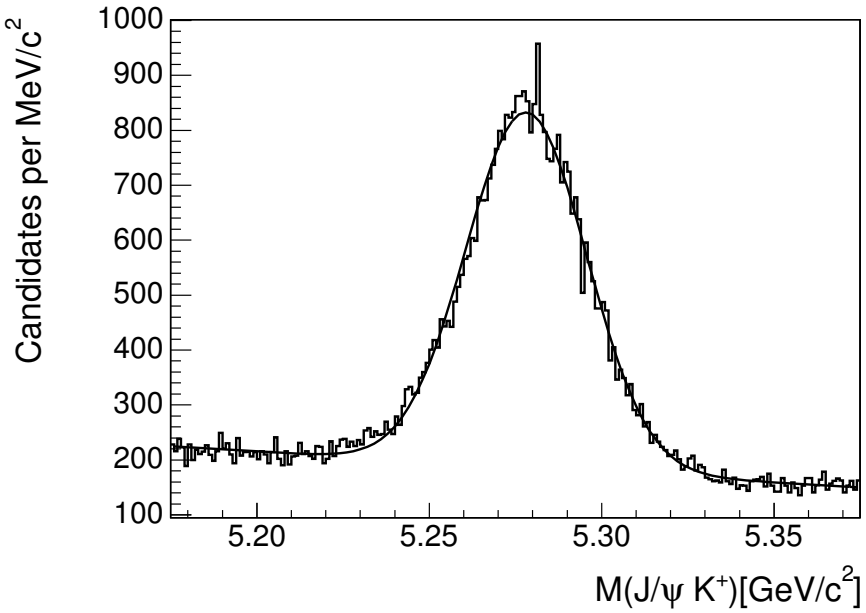

Figure 5.9: Invariant mass distribution of the $B^+ \to \overline{D^0}\pi$ candidates before neural network cut.

5.3.3 Preselection in the $B^+ \to \overline{D}\pi^+\pi^-\pi^+$ Decay Chain Sample

The preselection in this channel has been performed for the study of B_d^{**} mesons by Andreas Gessler. As in the $B^+ \to \overline{D^0}\pi^+$ channel, the preselection starts with cuts, that remove only a very small signal fraction.

- The signifcance of the transverse decay length of the B^+ has be larger than 7.5.
- The absolute value of the impact parameter of the B^+ has to be smaller than 0.0075 cm.
- The signifcance of the transverse decay length of the D0 with respect to the decay vertex of the B^+ has to be larger than minus four.
- The transverse momentum of the $\overline{D0}$ has to be larger than 1.5 GeV/c2.
- The charge of the pion from the $\overline{D^0}$ decay has to be opposite to the sum of charges of the pions from the B^+ decay. Doubly Cabbibo suppressed decays are neglected because a negligible amount of signal is hidden in a similar amount of background as in the non suppressed channel.
- The absolute value of the impact parameter of the $\overline{D^0}$ has to be larger than 0.002 cm.

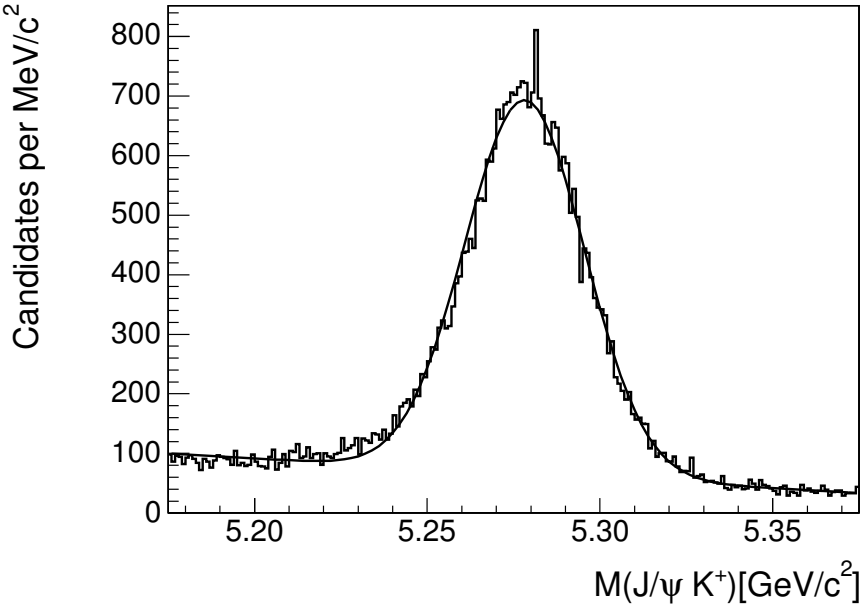

Figure 5.10: Invariant mass distribution of the $B^+ \to \overline{D^0}\pi$ candidates with a cut on neural network output on -0.2.

- The transverse momentum of the kaon of the $\overline{D^0}$ decay has to be larger than 400 MeV/c^2.

Then a neural network is trained. For the training a Monte Carlo simulation is used as signal pattern and experimental data from the upper B^+ mass sideband in the range from 5.325 GeV/c2 to 5.395 GeV/c2 is used as background pattern. More satisfying would be to use data from both the lower and upper sideband as background pattern, but since the lower sideband also contains partially reconstructed B^+ mesons, only the upper sideband can be considered containing rather pure background candidates. For the description of the variables in this decay channel, the three pions of the B^+ decay are ordered, so that the two pions of same charge come before the one with the opposite charge. Of the pions with the same charge, the one with the higher transverse momentum comes first. The variables which are used for the network training are

- the absolute value of the impact parameter of the B^+,
- the fit probability of the B^+ kinematic fit with vertex constraint,
- the significance of the transverse decay length of the B^+,
- the transverse momentum of the $\overline{D^0}$,
- the absolute value of the impact parameter of the $\overline{D^0}$,
- the fit probability of the $\overline{D^0}$ kinematic fit with vertex constraint,

- the significance of the transverse decay length of the $\overline{D^0}$ with respect to the primary vertex,
- the significance of the transverse decay length of the $\overline{D^0}$ with respect to the decay vertex of the B^+,
- the transverse momentum of the first pion from the B^+ decay,
- the transverse momentum of the second pion from the B^+ decay,
- the transverse momentum of the third pion from the B^+ decay,
- the significance of the impact parameter of the first pion from the B^+ decay with respect to the B^+ decay vertex,
- the significance of the impact parameter of the second pion from the B^+ decay with respect to the B+ decay vertex,
- the significance of the impact parameter of the third pion from the B^+ decay with respect to the B^+ decay vertex,
- the cosine of the angle of the first pion from the B^+ decay in the center of mass system of the B^+ relative to the momentum of the B^+ in the laboratory frame,
- the cosine of the angle of the second pion from the B^+ decay in the center of mass system of the B^+ relative to the momentum of the B^+ in the laboratory frame,
- the cosine of the angle of the third pion from the B^+ decay in the center of mass system of the B^+ relative to the momentum of the B^+ in the laboratory frame,
- the cosine of the helicity angle of the pion from the $\overline{D^0}$ decay,
- the transverse momentum of the kaon from the $\overline{D^0}$ decay in the case it is smaller than 2.0 GeV/c^2,
- the transverse momentum of the kaon from the $\overline{D^0}$ decay in the case it is higher than 2.0 GeV/c^2,
- the PID likelihood ratio of the kaon in the case there is time of flight information,
- the PID likelihood ratio of the kaon in the case there is no time of flight information,
- the invariant mass of the two pions from the B^+ decay having the same charge,
- the invariant mass of the combination of two pions with different charge signs from the B^+ decay having the smallest invariant mass,

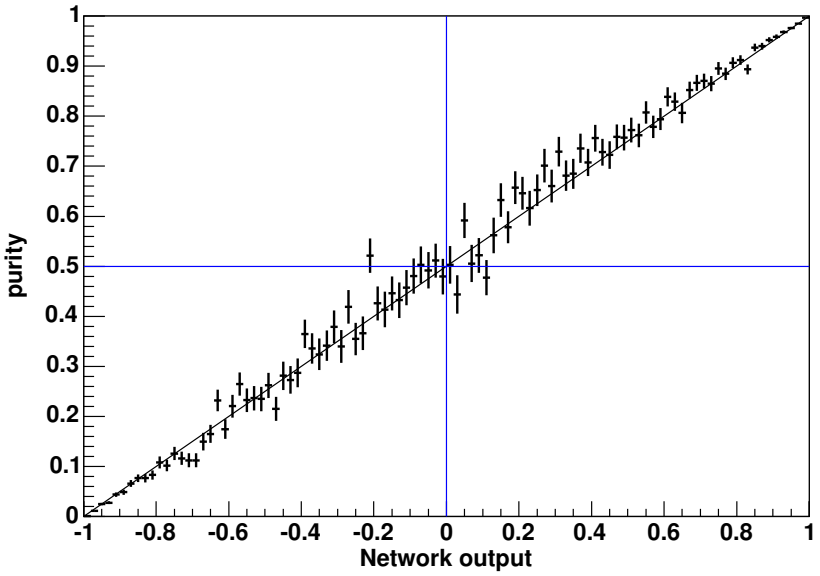

Figure 5.11: Purity as function of the neural network output for the $B^+ \rightarrow \overline{D^0}\pi^+\pi^+\pi^-$ channel using the training sample.

- the invariant mass of the combination of two pions with different charge signs from the B^+ decay having the highest invariant mass.
- the absolute value of the impact parameter of the pion from the B^+ decay having the smallest absolute value of the impact parameter.
- The absolute value of the impact parameter of the pion from the B^+ decay having the largest absolute value of the impact parameter.
- The angle between the total momentum of the three pions from the B^+ decay and the normal vector of the plane spanned by the momenta of the two pions with the same charge sign.
- The invariant mass of the three pions from the B^+ decay divided by the difference between the masses of the B^+ and the $\overline{D^0}$.

Figure 5.11 shows the purity as a function of the neural network output. The neural network is reasonably well trained as the purity is quite good a linear function of the neural network output. From figure 5.12 one can again conclude that the separation power of the neural network is good.

5.4 Selection of B_s^{**} mesons

After calculating the neural network output of the networks that separate B^+ from non-B^+ background for each candidate of all three decay chain samples, in the next

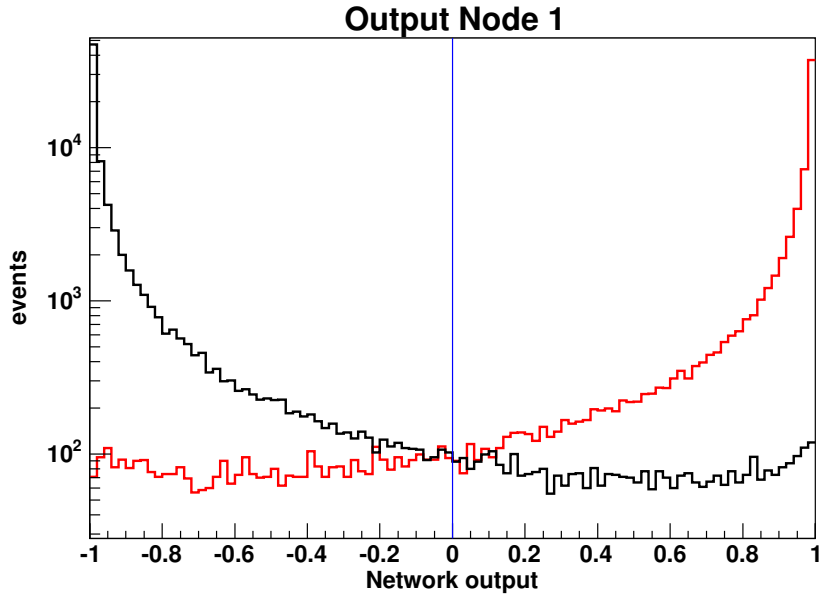

Figure 5.12: Neural network output distribution for signal (red) and background (black) events in the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel, using the training sample. For better visibility a logarithmic scale is chosen.

step neural networks to separate background with enriched B^+ mesons from B_s^{**} mesons are trained. In all three channels a neural network is trained after removing some obvious background with cuts. The trainings are done on a Monte Carlo sample as signal and an experimental data sample as background. The full 2.8 fb^{-1} are used for these trainings. As the mass of the various B_s^{**} states was considered unknown before the analysis, it is not possible to train on a sideband in the mass distribution, or Q-value distribution. The Q value is the defined in a decay as the mass of the parent particle minus the masses of all the daughter particles, e.g.

$$Q = M(B_s^{**}) - M(B^+) - M(K^-).$$

If the real decay was into a B^{*+}, this reduces the Q value for the decay, but as the photon from the B^{*+} decay has very little influence on the movement of the B^+, the calculation returns still the correct Q value. For the reverse calculation from the Q value to the mass, the photon energy has to be added.
To have a contribution of real B_s^{**} mesons in the experimental data, that is used for background, does not only reduce the separation power of the neural networks in the obvious way, that a real B_s^{**} meson then never could be counted as a clear signal event. There is as well the danger that a well trained neural network might learn small differences between Monte Carlo and the signal from experimental data. To avoid taking B_s^{**} mesons into the background sample for the neural network training, I use B^+K^+ and charged conjugated combinations as background. These combinations can not contain real B_s^{**} mesons due to electric charge conservation. Charge conservation can generate a different pattern for the background of these so

called wrong sign [WS] combinations. But it is reasonable to assume that the patterns of random B^+K^+ and B^+K^- combinations, that do not stem from the decay of a common mother particle, are similar enough to provide a valuable background description.
As the neural networks are more powerful when a wide range of variables is used and the signal that is searched for in this analysis is very close to the kinematic threshold, where the mass or Q value dependency is especially strong, there are variables that are correlated with the mass and Q value. To be able to use them and prevent the generation of artificial bumps, the Q value distribution of the Monte Carlo training sample is reweighted to the Q value distribution of the WS combinations used as background in the trainings. The mass window is up to 150 MeV/c^2, and candidates with a Q value of more than 75 MeV/c^2 get an additional factor of 0.1 as weight. This last measure ensures that the training is not completely dominated by the more numerous events at higher Q values, where the dependency of the structure of the background doesn't depend strongly on the Q value, and thus the neural network can provide good separation power at low Q values.

All trainings use the entropy loss function and BFGS algorithm as minimisation procedure. I chose the BFGS algorithm mostly because this algorithm stops iterating due to an internal mechanism. An additional advantage of this algorithm is, that it is very fast. With the global preprocessing I remove variables with less than 2 sigma significance in correlation to the target. I don't list the variables that are removed from the neural network in the description below. All three neural networks start with 15 nodes in the middle layer.

5.4.1 B_s^{**} Selection in the $B^+ \to J/\psi K^+$ Decay Chain Sample

Before the candidates enter the neural network, some cuts are applied again. In the $B^+ \to J/\psi K^+$ channel there are the following ones:

- The mass of the B^+ candidate has to be within $\pm$ 50 MeV/c^2 of the word average for B^+ mesons of 5.27915 GeV/c^2. In the preselection a sensitivity on the B^+ mass was not wished for, but now this becomes an important parameter. The cut range is about four times the resolution in both directions.

- If there is no PID information for the kaon stemming from the B_s^{**} decay, as neither a time of flight measurement, nor a energy loss measurement was possible, the candidate is removed.

- The neural network output used to enrich B^+ mesons has to be higher than 0.5.

- The significance of the transverse decay length of the B^+ candidate has to be higher than minus one. Of course negative values would be impossible with a perfect detector. The limited resolution makes such values possible, though.

As the B^+ has a lifetime, often the decay length is considerable positive, but still the highest probability to decay for a B^+ meson is immediately after it is created.

After these precuts the variables enter the neural network. Again the numbers in brackets indicate the chosen preprocessing. The variables for this channel are

- the impact parameter of the B^+ candidate (14),
- the fit probability of the B^+ kinematic fit with vertex constraint (35),
- the significance of the transverse decay length of the B^+ (15),
- the significance of the transverse decay length of the J/ψ with respect to primary vertex (15),
- the transverse momentum of the kaon of the B^+ decay (14),
- the significance of the impact parameter of kaon of the B^+ decay (14),
- the absolute mass difference of the B^+ candidate to the B^+ meson world average mass (15),
- the output of the neural network to enrich B^+ mesons (15),
- the cosine of the angle between the momentum of the kaon from the B_s^{**} decay in the centre-of-mass system of the B_s^{**} and the momentum of the B_s^{**} in the laboratory frame (14),
- the transverse momentum of the kaon from the B_s^{**} decay (14),
- the impact parameter of the kaon from the B_s^{**} decay (15),
- the significance of the the impact parameter of the kaon from the B_s^{**} decay (15),
- the pseudorapidity of the kaon from the B_s^{**} decay (15),
- the difference of the pseudorapidity of the kaon from the B_s^{**} decay and the B^+ (14),
- the CS moment of the kaon from the B_s^{**} decay (14),
- the pull of the time of flight measurement under the kaon hypothesis for the kaon from the B_s^{**} decay (94),
- the kaon probability from the time of flight measurement, as I described above (94),
- the PID likelihood ratio including energy loss information of the kaon from the B_s^{**} decay in case there is time of flight information (94),

- the PID likelihood ratio including energy loss information of the kaon from the B_s^{**} decay in case there is no time of flight information (94),

For the last two variables all events after the first 1.3 fb^{-1} are treated as not having energy loss information.

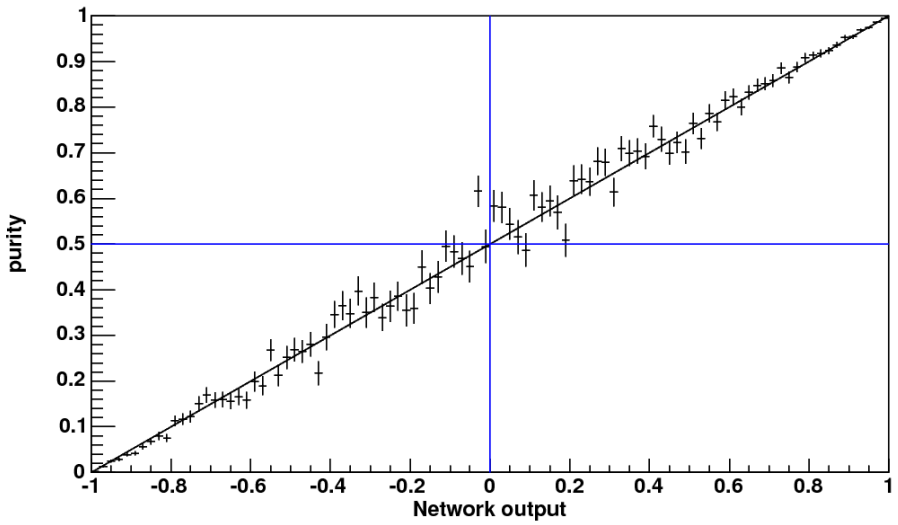

Figure 5.13: Quality plot for B_s^{**} neural network in the $B^+ \to J/\psi K^+$ decay chain. The plot shows an efficiency versus purity plot, based on the Monte Carlo signal and the WS combinations of experimental data used in the training. A well trained network shows a linear correlation.

Again I use the same set of plots to check if the neural network is reasonable trained, which is confirmed by figure 5.13. From figure 5.14 I see, that the separation power is very good as well. One has to keep in mind though, that the neural network is trained with a 50% signal fraction, while in the experimental data this fraction is very small.

5.4.2 B_s^{**} Selection in the $B^+ \to \overline{D^0}\pi^+$ Decay Chain Sample

Before the candidates enter the neural network, again cuts are applied. In this channel there are the following ones:

- The mass of the B^+ candidate has to be within $\pm$ 50 MeV/c^2 of the word average for B^+ mesons of 5.27915 GeV/c^2 [32].
- The transverse momentum of the pion from the B^+ decay has to be at least 1 GeV/c
- All precuts, that have been made for the neural network to enrich B^+ mesons in the $B^+ \to \overline{D\pi^+}$ sample are applied here again.

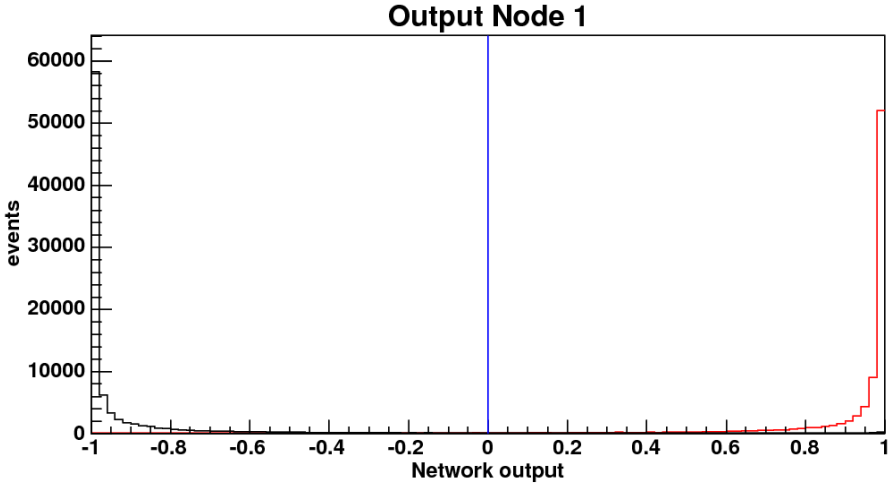

Figure 5.14: Separation for B_s^{**} neural network in the $B^+ \to J/\psi K^+$ decay chain. The plot shows the neural network output for the training samples of signal MC and WS combinations of experimental data used in the training.

- The output of the neural network for $B^+ \to \overline{D^0}\pi^+$ enrichement has to be higher than -0.2.
- If there is no PID information for the kaon stemming from the B_s^{**} decay the candidate is removed.
- The B^+ transverse momentum has to be at least 5 GeV/c.
- The transverse momentum of all kaons in the decay chain has to be at least 400 MeV/c. This cut is done, as the tracking is only for tracks stemming from particles with such a momentum reliable.

The variables that enter the neural network are

- the impact parameter of the B^+ (14),
- the fit probability of the B^+ kinematic fit with vertex constraint (34),
- the significance of the transverse decay length of the B^+ (14),
- the impact parameter of the $\overline{D^0}$ meson (14),
- the fit probability of the $\overline{D^0}$ meson vertex fit (34),
- the significance of the transverse decay length of the D^0 meson with respect to the primary vertex (14),
- the significance of the transverse decay length of the D^0 meson with respect to the decay vertex of the B^+ (14),
- the transverse momentum of the pion from the B^+ decay (14),

- the significance of the impact parameter of the pion from the B^+ decay (14),
- the transverse momentum of the kaon from the $\overline{D}$ decay in the case it is smaller than 2.0 GeV/c (95),
- the mass of the B^+ candidate (14),
- the cosine of the angle between momentum of the kaon from the B_s^{**} decay in the centre-of-mass system of the B_s^{**} and the momentum of the B_s^{**} in the laboratory frame (14)
- the output of the neural network for the enrichment of B^+ mesons in the corresponding channel,
- the impact parameter of the kaon from the B_s^{**} decay (14),
- the significance of the impact parameter of the kaon from the B_s^{**} decay (14),
- the pseudorapidity of the kaon from the B_s^{**} decay (14),
- the GJ moment of the kaon from the B_s^{**} decay (14),
- the pull of the time of flight measurement under the kaon hypothesis for the kaon from the B_s^{**} decay (94),
- the kaon probability from the time of flight measurement, as I described in 5.2 (94)
- the PID likelihood ratio including energy loss information of the kaon from the B_s^{**} decay in case there is no time of flight information (94).

Among other variables the PID likelihood with time of flight information didn't pass because of low significance. This is no problem and means just that the other variables, which use the time of flight information, have already almost all information that this PID likelihood would give.
In the figures 5.15 and 5.16 the quality of the training and the separation power can be checked. The purity is roughly a linear function of the neural network output. There is only relatively little statistics for values around a neural network output of zero. This leads to some deviation from the linear function, but this doesn't affect the analysis.
The separation of signal and background is good, but doesn't seem to be as powerful as in the $B^+ \to J/\psi K^+$ channel. This is not necessarily the case overall, as the precuts in this channel have already removed a lot of background.

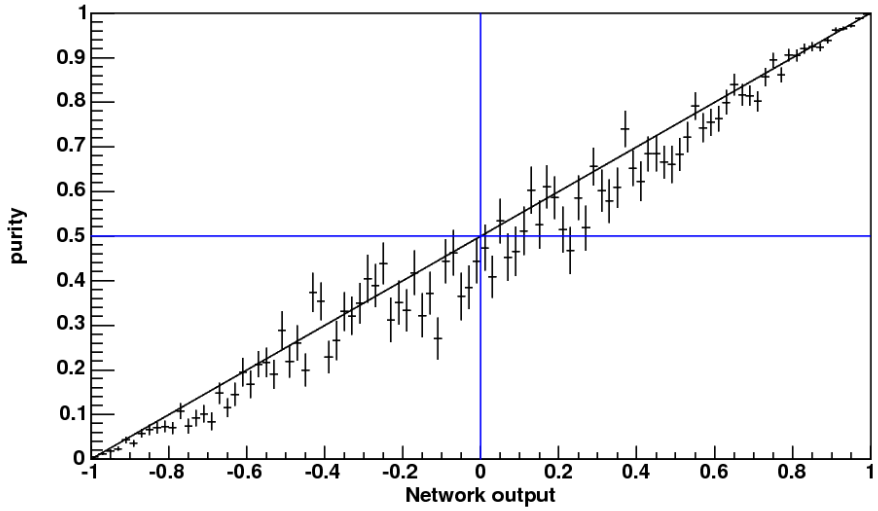

Figure 5.15: Quality plot for B_s^{**} neural network in the $B^+ \to \overline{D^0}\pi^+$ decay chain. The plot shows the purity as function of the output of the neural network, based on the MC signal and the WS combinations of experimental data used in the training.

5.4.3 B_s^{**} Selection in the $B^+ \to \overline{D^0}\pi^+\pi^-\pi^+$ Decay Chain Sample

The precuts in this channel are very similar as in the $B^+ \to \overline{D^0}\pi^+$ channel, as the sample is based on the same trigger:

- The mass of the B^+ candidate has to be within $\pm$ 50 MeV/c^2 of the word average for B^+ mesons of 5.27915 GeV/c^2 [32].
- The transverse momentum of the vector sum of the momenta of all 3 pions from the B^+ decay has to be at least 1 GeV/c.
- The significance of the transverse decay length of the B^+ candidate has to be higher than 7.5.
- The absolute value of the impact parameter d_0 of the B^+ has to be smaller than 0.0075 cm.
- The absolute value of the impact parameter of the $\overline{D^0}$ has to be higher than 0.0025 cm.
- The significance of the transverse decay length of the $\overline{D^0}$ with respect to the B^+ decay vertex has to be larger than minus four.
- The output of the neural network to enrich $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ has to be higher than 0.5.
- If there is no PID information for the kaon stemming from the B_s^{**} decay, the candidate is removed.

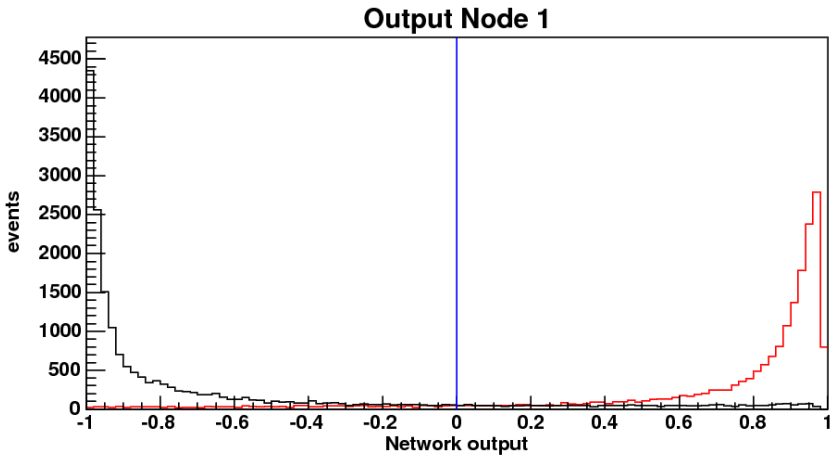

Figure 5.16: Separation for B_s^{**} neural network in the $B^+ \to \overline{D^0}\pi^+$ decay chain. The plot shows the neural network output for the training samples of signal MC and WS combinations of experimental data used in the training. A large fraction of the signal (red) is close to one and a large fraction of the background (black) close to minus one.

- The B^+ transverse momentum has to be at least 5 GeV/c.
- The transverse momentum of all kaons in the decay chain has to be at least 400 MeV/c.

As well the input variables are similar. They are

- the fit probability of the B^+ kinematic fit with vertex constraint (34),
- the fit probability of the $\overline{D^0}$ kinematic fit with vertex constraint (34),
- the significance of the transverse decay length of the $\overline{D^0}$ with respect to the primary vertex (14),
- the vector sum of the transverse momenta of the pions from the B^+ decay (15),
- the cosine of the angle of the $\overline{D^0}$ momentum in the centre-of-mass system of the B^+ relatively to the momentum of the B^+ in the laboratory frame (15),
- the transverse momentum of the kaon from the $\overline{D^0}$ decay in the case it is smaller than 2.0 GeV/c (95),
- the transverse momentum of the kaon from the $\overline{D^0}$ decay in the case it is higher than 2.0 GeV/c (95),
- the mass of the B^+ candidate (14),

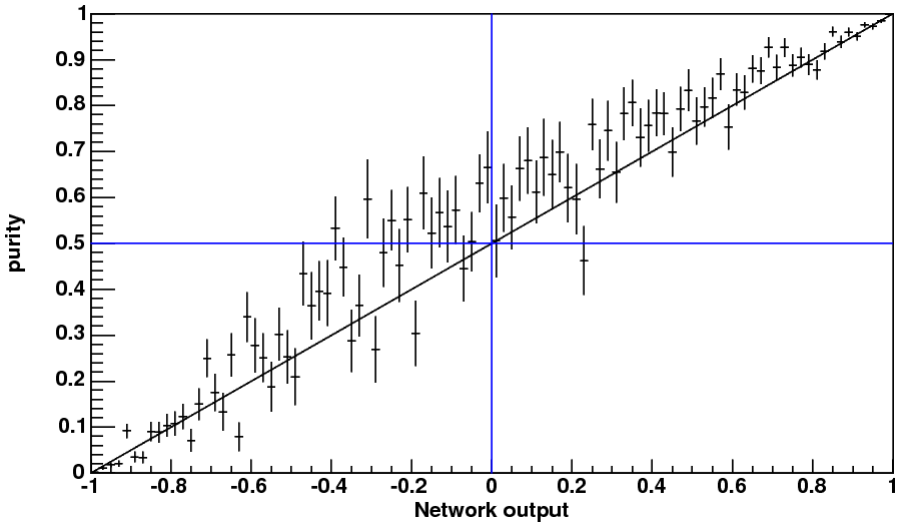

Figure 5.17: Quality plot for B_s^{**} neural network in the $B^+ \to \overline{D^0}\pi^+\pi^-\pi^+$ decay chain. The plot shows the purity as function of the neural network output, based on the Monte Carlo signal and the WS combinations of experimental data used in the training.

- the cosine of the angle between the momentum of the kaon from the B_s^{**} decay in the centre-of-mass system of the B_s^{**} and the momentum of the B_s^{**} in the laboratory frame (14),
- the output of the neural network for enrichment of B^+ in the corresponding channel,
- the transverse momentum of the kaon from the B_s^{**} decay (15),
- the impact parameter of the kaon from the B_s^{**} decay (14),
- the significance of the impact parameter of the kaon from the B_s^{**} decay (14),
- the pseudorapidity of the kaon from the B_s^{**} decay (14),
- the pull of the time of flight measurement under the kaon hypothesis for the kaon from the B_s^{**} decay (94),
- the kaon probability from the time of flight measurement, as I described in 5.2 (94)
- the PID likelihood ratio including energy loss information of the kaon from the B_s^{**} decay in case there is no time of flight information (94).

In the figures 5.17 and 5.18 the quality of the training and the separation power can be checked. There is less statistic, but within the available statistic, the results look reasonable.

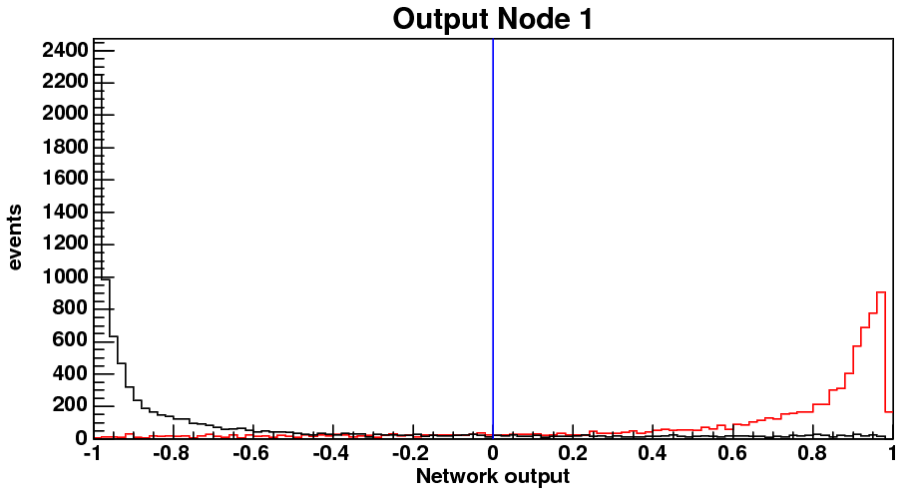

Figure 5.18: Separation for B_s^{**} neural network in the $B^+ \to \overline{D^0}\pi^+\pi^-\pi^+$ decay chain. The plot shows the neural network output for the training samples of signal Monte Carlo and WS combinations of experimental data used in the training. A large fraction of the signal (red) is close to one and a large fraction of the background (black) close to minus one.

5.5 Cut Optimisation for the B_s^{**} neural networks

There are various methods to optimise the cut on the final neural networks. Most important for the decision which to use, is not to bias the selection. A fairly often chosen path, that I follow here as well, is to maximise the signal over the square root of the sum of signal and background contributions. As signal I can use the Monte Carlo samples, background I have to take from the experimental data. As the position of one of the B_s^{**} states is already fairly well known, due to the efforts of previous experiments [14], I can define a region in Q value, where one of the expected signals can be found. In proximity to the value measured by DELPHI, although a little bit lower, a large peak is in the experimental data. I take a wide window around that peak in the Q value range of 62 MeV/c^2 to 72 MeV/c^2 and remain with

$$\frac{\text{Monte Carlo}}{\sqrt{\text{Exp. Data}(62 < \text{Q} < 72\ \text{MeV/c}^2)}}.$$

This optimisation is done for each of the three samples separately. The best result in the $B^+ \to J/\psi K^+$ channel is obtained for a cut at 0.99 as can be seen in figure 5.19. The absolute value of the significance in this plot is meaningless, as there has been no attempt to normalise the signal Monte Carlo to the expected signal in the experimental data.

In figure 5.20 the Q value distribution for the candidates in the $J\psi$ based sample are shown. It is useful not only to look to the best cut, but to look as well on some variations on the cut to see that irregularities are not just a random fluctuation, but contain a real signal. More over the optimal cut on the neural network can depend

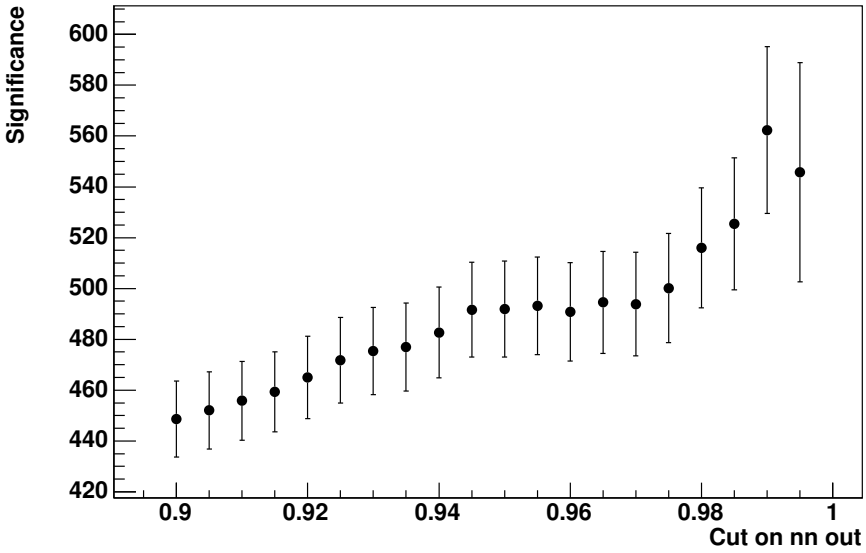

Figure 5.19: Significance optimisation for the $B^+ \to J/\psi K^+$ decay chain sample. The maximum is at 0.99 without an extended plateau.

on the Q value, so that signals at other places might be overlooked, if one doesn't consider some variation. Around 70 MeV/c^2 seems to be a very clear signal. Around 10 MeV/c^2 could be another. A third signal that would stem from a $B^*_{s2} \to B^{*+}K^-$ decay is not clearly visible.

In the next sample, that includes signal with the decay $B^+ \to \overline{D^0}\pi^+$, the maximisation procedure yields figure 5.21. The best cut is at 0.8. The reason, that a less hard cut is optimal in this sample is due to the greater purity of the signal in the beginning. All neural networks were trained with exactly 50% signal fraction, but in the experimental data, this channel includes less background.

In figure 5.22 one can see the Q value distribution for the sample including the $B^+ \to \overline{D^0}\pi^+$ decay chain. Again the two peaks mentioned before can be seen. A third peak for the alternative decay of the B^*_{s2} is much more pronounced.

In the last remaining sample the optimisation yields figure 5.23. I take 0.8 as the optimal cut value.

In figure 5.24 the Q value distribution for the last sample is shown. The clarity is not outstanding, but together with the other two channels, this is a useful contribution. In the next chapter I'm going to do a statistical analysis of the resulting samples with the optimal cuts applied. The interesting part seems to be in the area below 100 MeV/c^2, so I focus on that region to make the background description easier.

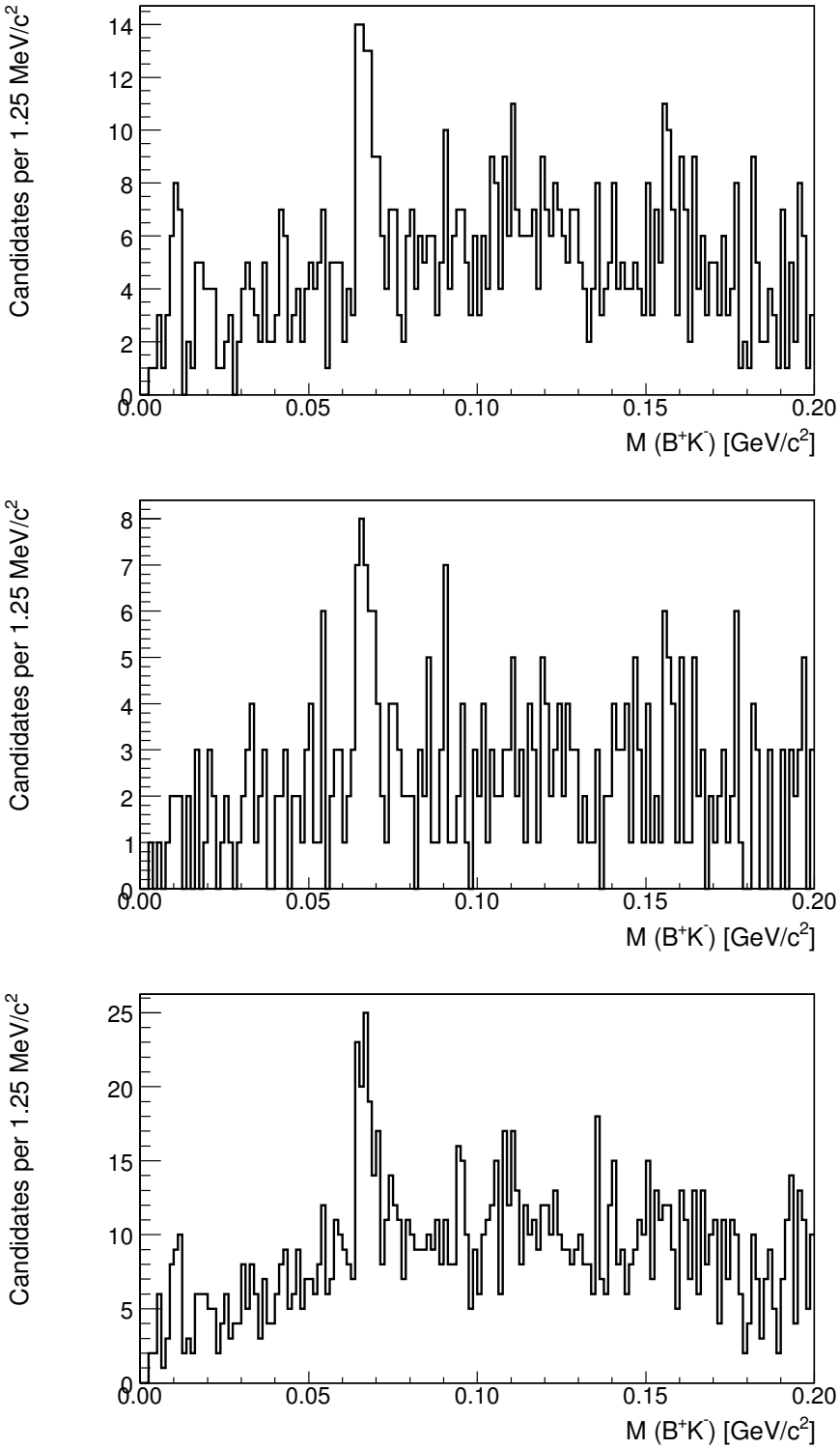

Figure 5.20: Q value distribution in the $B^+ \to J/\psi K^+$ decay chain sample. The upper plot is the result if one applies the optimal cut of 0.99 to the output of the neural network. The middle one is with an even harder cut of 0.995. The lower one is for a less hard cut at 0.98.

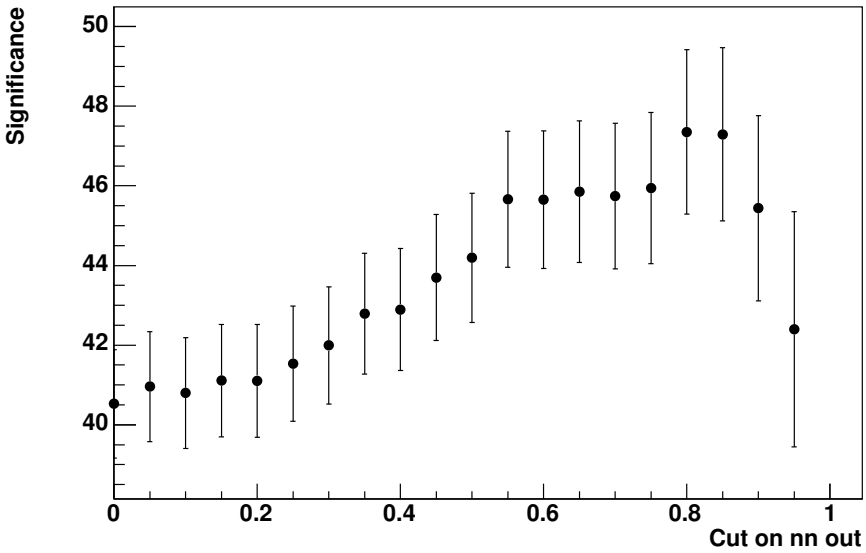

Figure 5.21: Significance optimisation in the $B^+ \to \overline{D^0}\pi^+$ decay chain sample. The maximal significance occurs at a cut on the neural network output of 0.8.

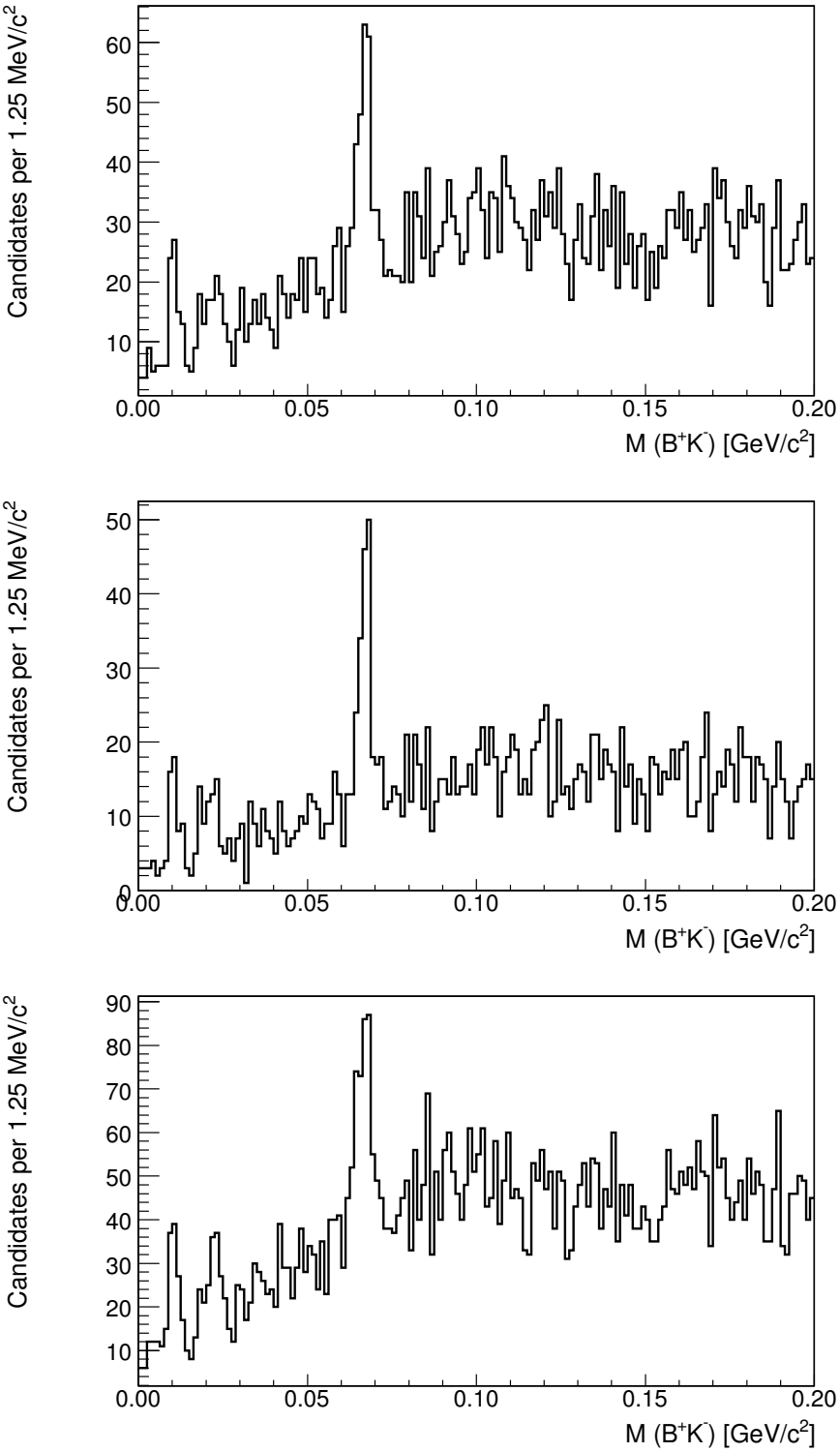

Figure 5.22: Q value distribution in the $B^+ \to \overline{D^0}\pi^+$ decay chain sample. The upper plot is the result if one applies the optimal cut of 0.8 to the output of the neural network. The middle one is with a harder cut of 0.9. The lower one is for a less hard cut at 0.6.

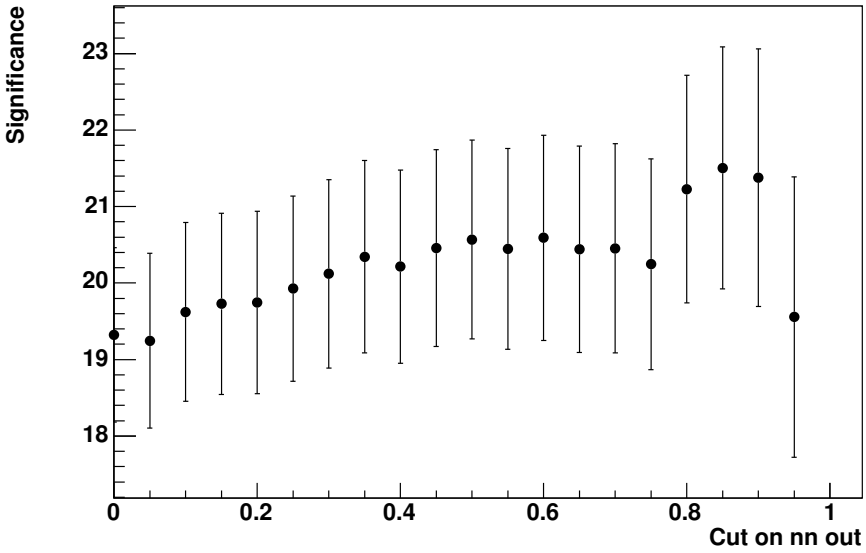

Figure 5.23: Significance optimisation in the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ decay chain sample.

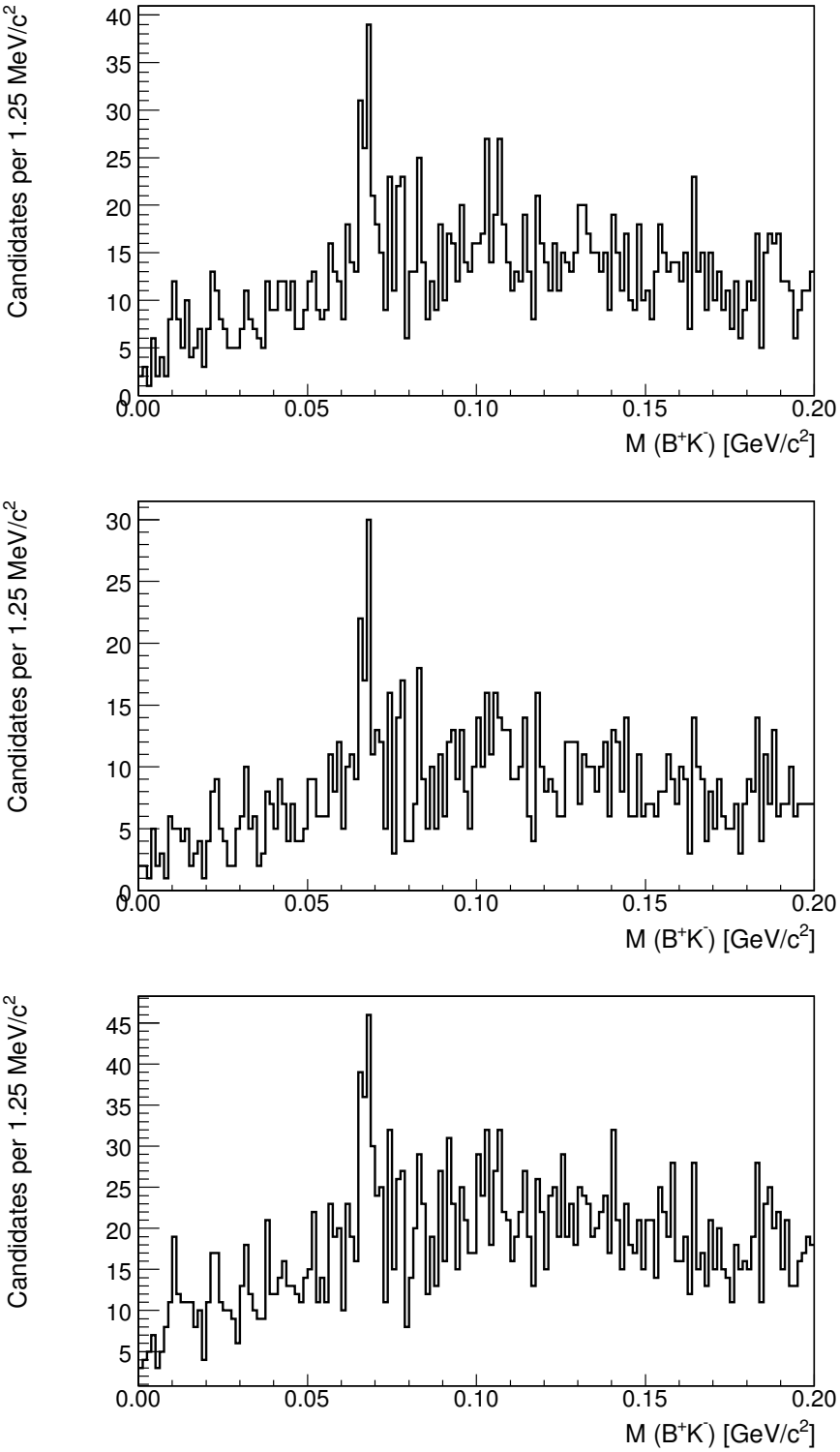

Figure 5.24: Q value distribution in the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ decay chain sample. The upper plot is the result if one applies the optimal cut of 0.8 to the output of the neural network. The middle one is with a harder cut of 0.9. The lower one is for a less hard cut at 0.6.

Chapter 6

Measurement of the B_s^{**} properties

In this chapter various statistical analyses are performed on the candidates that have been selected as described in the previous chapter. The first question is, what are the masses of the states that we see, including potential systematic effects. The next question is, if all the states that are assumed to be there, yield a significant amount of signal. Then I want to discuss the natural width Γ of the measured B_s^{**} states in more detail. The next part deals with the relative production ratio of the B_{s2}^* state and the B_{s1} state, as well as the relative branching fraction of the B_{s2}^* into B^+K^- and $B^{*+}K^-$. In the last part of this chapter an analysis of the B^+ - B^{*+} mass difference [Δm(B^{*+}, B^+)] is presented.

6.1 Fit Preparation

To measure the masses of the states, I will perform an extended unbinned maximum likelihood fit, whose basic principles are described in chapter 4. I start with the introduction of the PDFs for background and signal contributions. Then the determination of the detector resolution from Monte Carlo samples follows, and the validation of the fitter.

6.1.1 PDF for Background

As already mentioned, I focus on the Q value range up to 100 MeV/c^2, as in that region all the structures of interest can be analysed. In this region the background is rather smoothly increasing, so my first background PDF is a polynomial of first order. Especially when the statistics is low, the numerical convergence of approaches with more degrees of freedom is not guaranteed, while a linear function is very robust. The linear function is determined by the slope and the total amount of background.

$$N \cdot [a \cdot Q + \frac{1}{Q_{max} - Q_{min}} - \frac{a}{2}(Q_{max} + Q_{min})] \tag{6.1}$$

N is the parameter to regulate the amount of background. The slope is given by a. The rest of the expression is chosen in a way that the integral of the PDF is given by N. Q_{min} and Q_{max} are the borders of the fit region, in this case zero and 100 MeV/c^2.
For cross checks that test the influence of different background PDFs, I resort additionally to another function, that is based on a different idea. As the phase space at the Q value is zero, it should be possible to anchor the background PDF at Q = 0 to zero. Close to the threshold, a fast increase is expected, that fades away at higher Q values. A function that follows these ideas is

$$N \cdot \{\frac{1}{A_{Norm}} \cdot [\mathrm{Q} \cdot (\beta - \mathrm{Q})]^{\gamma} \cdot e^{-\gamma \mathrm{Q}}\}. \tag{6.2}$$

N is the parameter to regulate the amount of background. The expression in the braces is normalised to one by the factor A_{Norm}. The expression in the brackets makes sure that the function is zero at Q = 0, and is rising fast thereafter. The last factor guarantees that the function slowly fades out after the initial rapid increase.

6.1.2 PDF for signal

For the signal description I resort to the Voigt function. This function is a convolution of a Gaussian and a non-relativistic Breit-Wigner function. The former is an approximation for the imperfect detector resolution, the latter a description of the natural shape of a state if the width is much smaller than the energy available in the decay. As the states in the experimental data seem to be very narrow and the statistics is limited, the fit might not work well if both, the natural width and the resolution of the Gaussian component in the Voigt function, are free. To limit the freedom of the fit, I use the Monte Carlo samples to determine the resolution.

6.1.3 Resolution

If there is sufficient statistics, more subtle features of the shape of the detector resolution than just a simple Gaussian can be obtained from the Monte Carlo samples. In this analysis the amount of experimental data is still relatively small, so a single Gaussian to describe the resolution is sufficient.
There are three Monte Carlo samples for the B_s^{**}, one for the $B^+ \to J/\psi K^+$ decay channel, for the $B^+ \to \overline{D^0}\pi^+$, and for the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ decay channel. However, the Q value is more important than the further decay of the B^+ meson for the resolution, because the resolution is dependent on the fraction of kinetic energy from the total invariant mass. To take this effect into account, I determine the resolution in slices of the Q-value distribution. Clear peaks are visible in the Q value distribution at 10 and 67 MeV/c^2. If one interpretes the peak at 67 MeV/c^2 as stemming from the $B_{s2}^* \to B + K^-$ decay, and the one around 10 MeV/c^2 as stemming from the $B_{s1} \to B + K^-$ decay, a possible third peak has to be about 46 MeV/c^2 lower, as it would be the other decay channel of the B_{s2}^* to $B^{*+}K^-$. So the

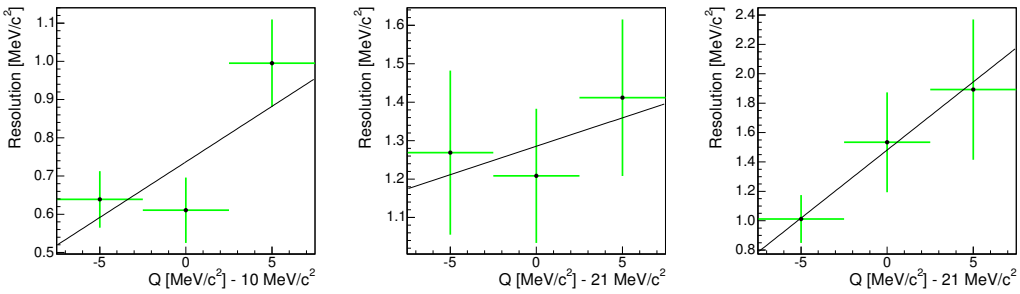

Figure 6.1: Summary of the resolutions for the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel. The Q value of 10 (21, 67) MeV/c^2 is set to zero in the left (middle, right) plot. Each plot has a χ^2 fit using a first order polynomial.

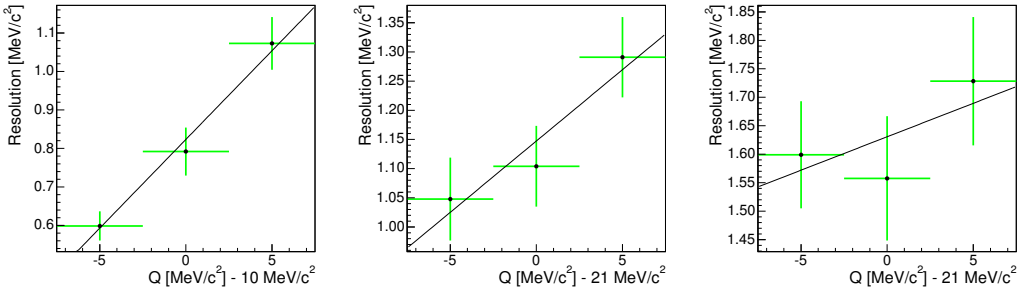

Figure 6.2: Summary of the resolutions for the $B^+ \to \overline{D^0}\pi^+$ channel. The Q value of 10 (21, 67) MeV/c^2 is set to zero in the left (middle, right) plot. Each plot has a χ^2 fit using a first order polynomial.

values, at which the resolution is of interest are at 10, 21, and 67 MeV/c^2. I take slices of 5 MeV/c^2 width around these values and adjacent to the one that has the relevant value at its center. This gives candidates in three regions between Q values of 2.5, 7.5, 12.5, and 17.5 MeV/c^2 for the lowest considered possible signal, between 13.5, 18.5, 23.5, and 28.5 MeV/c^2 for the middle possible signal, and 59.5, 64.5, 69.5, and 74.5 MeV/c^2 for the resolution of the large peak. I take into account, that at 67 MeV/c^2 the decay is likely a decay directly into a B^+, while the other two probably stem from decay chains, in which the B_s^{**} decays into a B^{*+}. All selection cuts that are applied to data, including the one on the neural network, are performed on the Monte Carlo samples, before they are filled into histograms. The resulting distributions of reconstructed minus generated Q values in each slice are fitted by a χ^2 fit with a single Gaussian as fit model. The results can be seen in figures B.1 to B.9.

To improve the resolution determination, I use a linear fit over the three slices in every region to determine the resolution at the considered Q values (figures 6.1 to 6.3). Additionally, in this way it is measured, how strongly the resolution changes in the proximity of the considered peaks with a variation in Q value. The results of that latter fits are summarised in table 6.1. The variation in resolution is less than 0.1 MeV/c^2 apart from one case in the $B^+ \to D\pi^+\pi^+\pi^-$ channel with a very high

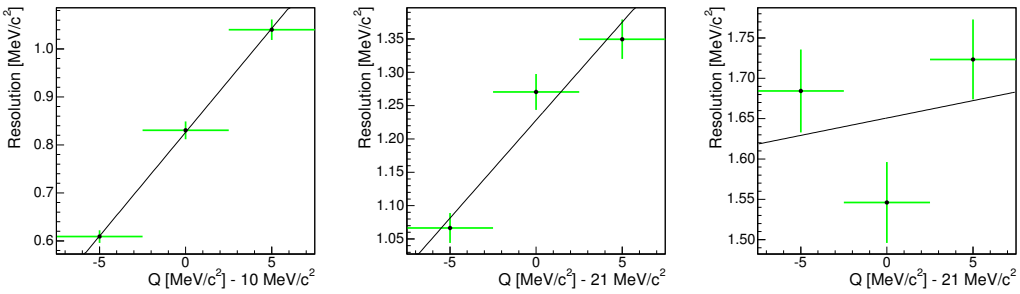

Figure 6.3: Summary of the resolutions for the $B^+ \to J/\psi K^+$ channel. The Q value of 10 (21, 67) MeV/c^2 is set to zero in the left (middle, right) plot. Each plot has a χ^2 minimising fit using a first order polynomial.

uncertainty. But in this case the statistics of the available Monte Carlo sample is not sufficient for further investigation. So I average both the central value and the slope for all channels.

	$B^+ \to D\pi^+\pi^+\pi^-$	$B^+ \to D\pi^+$	$B^+ \to J/\psi$	average
10 MeV/c^2				
Resolution	0.74± 0.05	0.82 ± 0.03	0.826 ± 0.010	0.822 ± 0.009
Slope	0.029 ± 0.013	0.046 ± 0.008	0.043 ± 0.002	0.0429± 0.0019
21 MeV/c^2				
Resolution	1.29± 0.11	1.15 ± 0.04	1.228 ± 0.015	1.220 ± 0.014
Slope	0.015 ± 0.029	0.024 ± 0.010	0.029 ± 0.004	0.028 ± 0.004
67 MeV/c^2				
Resolution	1.5± 0.2	1.63 ± 0.06	1.65 ± 0.03	1.64 ± 0.03
Slope	0.093 ± 0.044	0.012 ± 0.015	0.004 ± 0.007	0.006 ± 0.007

Table 6.1: Resolution for the various samples and averaged.

The results for the slopes at the different Q values confirm that the resolution is more sensitive to variations for low Q values. As this determination of resolution may be not exactly at the position of the respective signals, the slopes can be treated as uncertainties to chose a reasonable accuracy. I take 0.82 ± 0.04, 1.22 ± 0.03, and 1.64 ± 0.03 MeV/c^2 as resolutions for the three regions, when assuming the resolution was determined within a Q value of 1 MeV/c^2 of the signals.

6.1.4 Fitter Validation

To validate the fitter, I generate two ensembles of Monte Carlo using the PDFs used in the fit with 1000 candidate samples in each ensemble. The input parameters are listed in table 6.2. Set A has set the natural width to zero to have a better handling on the test of the other signal parameters. Set B includes natural widths. For the background generation I use in both cases the linear function. The exact amount of

the signal and background contribution is further randomised by a Poisson distribution, while the Q values are fixed. For the fits, the parameters are free, apart from

Variable	Toy Set A	Toy Set B
	[MeV/c^2]	[MeV/c^2]
Q(B_{s1})	10	10
Q(B^*_{s2})	69	69
Δm(B^{*+}, B^+)	48	48
$\Gamma(B_{s1})$	0	3.0
$\Gamma(B^*_{s2})$	0	5.0
	[Candidates]	
N(B_{s1})	100	120
N($B^*_{s2} \to B^+K^-$)	250	250
N($B^*_{s2} \to B^{*+}K^-$)	80	80
N(BKG)	1750	1750
Slope	100	110

Table 6.2: Input parameters for fitter validation Monte Carlo ensembles.

the natural width Γ, which is fixed to zero in sample A, and the difference between the $B^*_{s2} \to B^+K^-$, $B^*_{s2} \to B^{*+}K^-$ splitting, which has a Gaussian constraint of 1 MeV/c^2 width around 48 MeV/c^2 in sample B, to make the fit more stable. I check the fits, if they are generally reasonable, as sometimes a fit fails or the uncertainty matrix calculation fails, although the fit converged. In set A all fits meet this standard. In set B, 76 out of the 1000 fits fail and are removed before the next step.
For each variable I calculate the pull, which is the significance with which the parameter differs from the value with which the sample was generated. The pull distribution for the Q value validation of the B_{s1} ($B^*_{s2} \to B^+K^-$) decay can be found in figure C.1(C.2), the difference of the $B^*_{s2} \to B^+K^-$ and $B^*_{s2} \to B^{*+}K^-$ decay in figure C.3, the results for the yields of B_{s1} ($B^*_{s2} \to B^+K^-$, $B^*_{s2} \to B^{*+}K^-$) in figure C.4 (C.5, C.6), and the result for the natural widths in C.7.
The mean values and the widths of a binned likelihood fit with a Gaussian to the pull distributions can be found in table 6.3. Despite the mean values for the Q values of all three signals being always on the negative side, there is no clear divergence from zero within the uncertainty of the fit. The low width of the fit to the Q value difference of the B^*_{s2} decays is the result of the Gaussian constraint. The natural widths Γ show a very clear divergence to the negative side. The reason for this is that, if the signal contribution of the Toy Monte Carlo fluctuates to a narrow broadness, the broadness of the signal pdf is determined by the resolution, as the width of the Breit-Wigner can not give a negative contribution to the broadness of the signal pdf. As the resolution is fixed, the fit yields a very high certainty for a very low width. In the plot this is visible by the long tail on the side of negative values. On the positive side no such long tail is visible. Still 449 of the 921 fits yield a natural width that is larger than the generated value, which is no significant deviation from the expected 50%. It is even possible that some events are missing

because fits with very high widths tend to be unstable. This can as well have some effect on the yield, as fits with large width tend to have many signal candidates in the fit. The central value is then pulled a little bit to the negative side. I conclude that the fit is well implemented, and that if the width of a signal contribution can be sufficiently explained by the resolution, the width determination becomes unreliable, while a fit yielding a width much higher than the resolution can still be trusted, if the fit converges properly.

	Toy sample A		Toy Sample B	
	mean	width	mean	width
$Q(B_{s1})$	0.04±0.03	1.01±0.02	-0.02±0.03	1.02±0.02
$Q(B^*_{s2})$	-0.02±0.03	1.01±0.02	-0.02±0.03	1.01±0.02
$\Delta m(B^{*+}, B^+)$	-0.04±0.03	0.98±0.02	0.00±0.02	0.71±0.02
$\Gamma(B_{s1})$	-	-	-0.17±0.03	1.06±0.02
$\Gamma(B^*_{s2})$	-	-	-0.17±0.03	1.06±0.02
$N(B_{s1})$	-0.01±0.03	1.02±0.02	-0.02±0.03	0.99±0.02
$N(B^*_{s2} \to B^+K^-)$	0.02±0.03	1.04±0.02	-0.08±0.03	1.03±0.02
$N(B^*_{s2} \to B^{*+}K^-)$	0.07±0.03	0.98±0.02	-0.02±0.03	1.02±0.02

Table 6.3: Results for fitter validation Monte Carlo fits to the pull distributions.

6.2 Q Value Extraction

Now I can perform the extended maximum likelihood fit on the experimental data. The mass of the states is independent of the further decay chain of the B^+, so I start with a fit to the sum of all data samples. I use the default background function, fix the resolution, and use the measured [14] mass difference of B^+ and B^{*+} as Gaussian constraint for the Q value difference between the $B^*_{s2} \to B^+K^-$ decay and the $B^*_{s2} \to B^{*+}K^-$ decay. In the first attempt, the natural widths are free parameters of the fit to demonstrate that the resolution is sufficient to describe the data. The width of the B^*_{s2} is independent of its decay, so the Voigt functions for the signals around 21 and 67 MeV/c^2 have to have the same natural width.
The results of the fit are listed in table 6.4. A graphical representation of the result is shown in D.1.

The natural width is consistent with zero. A Breit-Wigner contribution fades only very slowly, so it is especially sensitive to small inaccuracies in the background description. It is therefore reasonable to assume, that the full width of the signal contribution is the result of just the detector resolution.
I redo the fit with the natural width set to zero, and postpone a discussion of a limit to the next section. The results of this fit can be found in table 6.5. In the following it will be considered as baseline scenario, and the results of this fit are the central values relative to which uncertainties are claimed. The fit projection is shown in fig. 6.4.

Variable	Value [MeV/c^2]	Constraint
Q(B_{s1})	10.32 ± 0.20	-
Q(B^*_{s2})	67.01 ± 0.21	-
Δm(B^{*+}, B^+)	45.2 ± 0.5	45.7 ± 0.9
$\Gamma(B_{s1})$	0.75 ± 0.55	-
$\Gamma(B^*_{s2})$	1.4 ± 0.7	-
	[Candidates]	
N(B_{s1})	83 ± 17	-
N($B^*_{s2} \to B^+K^-$)	309 ± 36	-
N($B^*_{s2} \to B^{*+}K^-$)	58 ± 17	-
N(BKG)	2506±67	-
Slope	148 ± 8	-

Table 6.4: Fit results for fit with free width. The table lists the results of a fit, in which all parameters that describe my fit model are left free or only have a Gaussian constraint.

Variable	Value [MeV/c^2]	Constraint
Q(B_{s1})	10.25 ± 0.18	-
Q(B^*_{s2})	67.00 ± 0.18	-
Δm(B^{*+}, B^+)	45.1 ± 0.5	45.7 ± 0.9
	[Candidates]	
N(B_{s1})	66 ± 11	-
N($B^*_{s2} \to B^+K^-$)	253 ± 22	-
N($B^*_{s2} \to B^{*+}K^-$)	38 ± 11	-
N(BKG)	2598±55	-
Slope	142 ± 6	-

Table 6.5: Results for baseline scenario fit. The natural width is set to zero.

As a cross check I fit as well the decay channels separately. The results are shown in table 6.6. The Q value related variables should be the same for the signals in all three decay channels. There are some differences, but nothing statistically significant. The fit projections are shown in figures D.2 to D.4.

6.3 Significance of Signal Contributions

The $B^*_{s2} \to B^+K^-$ and the B_{s1} decay are obviously significant. Their significance has already been tested on a subsample of the data with a slightly different selection by p-value calculations, described in ref. [15]. For both signals a significance of more than 5 σ has been proven. The $B^*_{s2} \to B^{*+}K^-$ decay yields a less clear signal, so I want to study its significance. To do so, I use the difference in the log likelihood as described in chapter 4. This approach depends on the number of additional

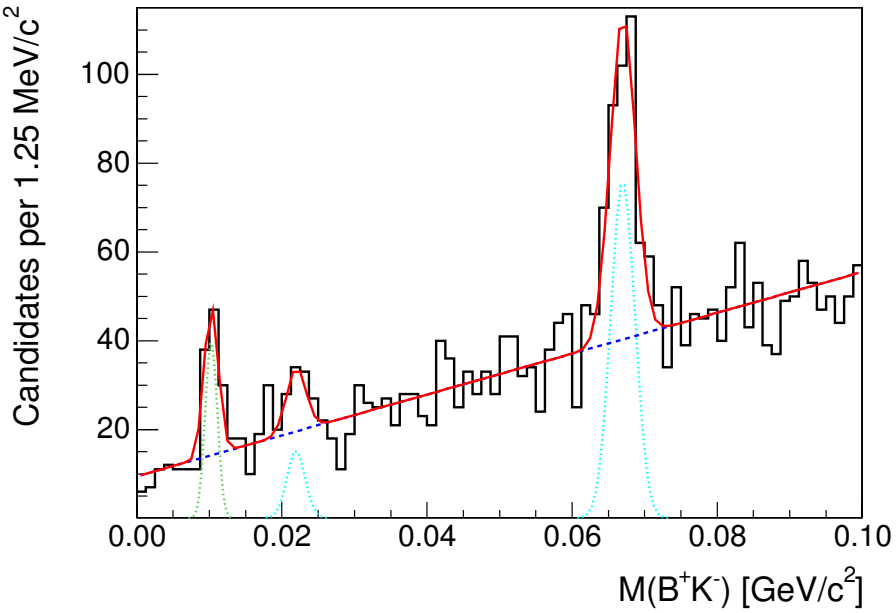

Figure 6.4: Result for baseline scenario fit.

Variable	$B^+ \to J\psi K^+$	$B^+ \to \overline{D^0}\pi^+$	$B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$
	Value [MeV/c^2]	Value[MeV/c^2]	
Q(B_{s1})	10.2 ± 0.4	10.1 ± 0.2	10.6 ± 0.4
Q(B^*_{s2})	66.4 ± 0.4	67.0 ± 0.2	67.4 ± 0.3
Δm(B^{*+}, B^+)	46.6 ± 0.8	44.7 ± 0.6	45.3 ± 0.6
	[Candidates]		
N(B_{s1})	15 ± 5	40 ± 8	13 ± 5
N($B^*_{s2} \to B^+K^-$)	40 ± 8	140 ± 17	74 ± 12
N($B^*_{s2} \to B^{*+}K^-$)	8 ± 5	25 ± 9	14 ± 6

Table 6.6: Results for fits on each channel separately.

parameters that are introduced via the signal function. So far the baseline scenario uses two free parameters for the description of the $B^*_{s2} \to B^{*+}K^-$ decay, but the one is constrained, as the mass difference between B^+ and B^{*+} is known quite well. Instead of throwing this information away, I repeat the baseline fit with a completely fixed mass difference, using again the value measured in the DELPHI experiment [14]. A fit projection can be seen in figure D.5.

The only free parameter for the description of the $B^*_{s2} \to B^{*+}K^-$ decay component is its yield with 36 ± 11 signal events. As the difference between it and the signal of the $B^*_{s2} \to B^+K^-$ decay is fixed, the large peak determines the position of the $B^*_{s2} \to B^{*+}K^-$ decay. The negative log likelihood value for this fit is: -2 ln($\mathcal{L}$) = -55635.7.
Now I repeat the identical fit, but fix the signal yield for the $B^*_{s2} \to B^{*+}K^-$ decay

to zero. The negative log likelihood in this case is -55622.0. The difference between the two fits is 13.7, which translates into a significance of 3.7 σ. The fit projection can be found in figure D.5.
As a cross check, I repeat the analysis for the B^{*+}, B^{+} mass difference $\pm$ one σ away from the central value of the DELPHI measurement. For the mass difference fixed to 45.7 + 0.9 MeV/c^2, I obtain a significance of 4 σ. For 45.7 - 0.9 MeV/c^2 I obtain a significance of 3.2 σ. This is reasonable and expected, as the fit with the variable mass difference returns a slightly lower difference than the central value of the DELPHI measurement. The cross check therefore is no matter of concern, and the significance remains 3.7 σ.

However, as the background shape is not known in principle, one has to consider also variations in the background shape. I redo the same procedure with the alternative background PDF of eq. 6.2. The fit is unstable as long as β is unconstrained. β runs to very high values, and the second derivative of the background PDF becomes zero. So I fix β to 200 GeV/c^2. The background PDF is now only dependent on γ and the amount of background.

Variable	Value [MeV/c^2]	Value[MeV/c^2]
Q(B_{s1})	10.22 $\pm$ 0.18	10.22$\pm$0.18
Q(B^*_{s2})	67.04 $\pm$ 0.19	67.00$\pm$0.19
	[Candidates]	
N(B_{s1})	65 $\pm$ 11	63 $\pm$ 11
N($B^*_{s2} \to B^+K^-$)	251 $\pm$ 22	250 $\pm$ 22
N($B^*_{s2} \to B^{*+}K^-$)	26 $\pm$ 11	0 (fixed)

Table 6.7: Results for alternative background scenario fit. Both fits have a fixed Q value difference between the decays of the B^*_{s2}. In the first fit the $B^*_{s2} \to B^{*+}K^-$ decay has a free amount, while in the second the amount is fixed to zero.

The results of the fit with the alternative background can be seen in table 6.7, and the fit projections in figure D.6 In this case, the negative loglikelihood yields -55631.8 with the additional signal and -55624.7 without it. The difference is only 7.1, which corresponds to a signal significance of 2.7 σ or a probability of 99.1 %.
Again I check the result, if I fix the mass difference one σ higher or lower. For the mass difference fixed to 45.7 + 0.9 MeV/c^2, I obtain a significance of 2.9 σ. For 45.7 - 0.9 MeV/c^2 I obtain a significance of 2.2 σ. The cross check yields no unexpected behaviour, so the significance in this scenario remains 2.7 σ.
As the alternative background function fits nearly as good as the linear one, I conservatively use the worse value of the scenarios for the significance, that is 2.7 σ.

6.4 Masses

The masses can be calculated from the Q values with parameters such as the B^+ mass and the K^+ mass. So most of this section deals with the Q values. The Q values from the baseline scenario can differ from the physical values, because the momentum calculation from the experimental data out of the detector is not precise, or because the fit model doesn't describe the data in an accurate way.

6.4.1 Uncertainties from fit model

Uncertainties that arise from the fit model can be estimated by using different descriptions. So far I have already listed the Q values of the B_{s1} decay and the $B_{s2}^* \to B^+K^-$ decay 4 times - the baseline scenario, when the signals have a floating natural width, the alternative background model, and the alternative background model without the $B_{s2}^* \to B^{*+}K^-$ decay of the B_{s2}^* state.
As uncertainty from the fit model, I take the highest discrepancy in the central value of any of these fits to the central value in the baseline scenario. In table 6.8 the various fit results are listed again. All the statistical uncertainties are in the same

	Q(B_{s1})	Q(B_{s2}^*)
Baseline	10.25 ± 0.18	67.00± 0.18
Floating width	10.32 ± 0.20	67.01± 0.21
Alt. backgound	10.22 ± 0.18	67.04± 0.19
Alt. bkg. w/o ind. decay	10.22 ± 0.18	67.00± 0.19
Max. discrepancy	0.07	0.04

Table 6.8: Q values in different fit models. The Max. discrepancy denotes the highest difference in the central value of any fit, to the central value of the baseline scenario.

order of magnitude, so there is no reason to dismiss one of the central values as irrelevant. The highest difference for the Q value of the B_{s1} to the baseline scenario stems from the fit with floating widths, and is 0.07 MeV/c^2. For the B_{s2}^* state, the highest difference is 0.04 MeV/c^2 and stems from the alternative background shape.

6.4.2 Uncertainties from the Track Reconstruction

The dominating systematic effects on the Q value from the tracking procedure stem from uncertainties in the material description and the strength of the magnetic field. The material in the detector leads to multiple scattering and energy loss in the detector, that has to be taken into account during the tracking procedure. Uncertainties in the magnetic field translate directly into the momentum calculation. A very extensive study about effects on the resolution has been performed to determine the size of these effects. The study is described in ref. [41], which deals with

the measurement of D^{**} mesons. In that study the material description is changed until the measured mass of J/ψ mesons in decays to $\mu^+\mu^-$ doesn't depend on the momentum of the J/ψ more than 0.006 ± 0.001 MeV/c^2 per GeV/c in momentum difference of the J/ψ mesons. The authors show, that by doing this for the J/ψ as well the mass measurements of $K_s^0 \to \pi^+\pi^-$, $B^\pm \to J/\psi K^\pm$, $\Upsilon \to \mu^+\mu^-$, and $D^0 \to K^-\pi^+$ do not depend on the momentum of the decaying particle within the uncertainty of the corresponding measurement. Given a range in the order of 10 GeV/c for the momentum of the the decaying particles, the systematic uncertainty arising from the uncertainty in the material description is of the order of 0.1 MeV/c^2 for the mass. This value has been chosen in [41] for a kinematically similar decay to the B_s^{**}. Note that despite the B_s^{**} mesons have higher masses, the Q value for the B_s^{**} decays is less than the 400 - 600 MeV/c^2 in the case of the D^{**} mesons. This means, that taking the same uncertainty is a conservative approach.
To determine the effect of the magnetic field, it is adjusted such, that the J/ψ mass is exactly the world average mass. The J/ψ can be produced resonantly at e^+e^- colliders, and is known extremely well. As well the other particles mentioned above yield results at the world average mass within the uncertainty of the measurement. From the remaining uncertainty in the mass measurements of these particles, the systematic uncertainty due to the magnetic field can be determined, and is found to be about 0.1 MeV/c^2. This is taken for the B_s^{**} mesons as conservative value, because measurements at a low Q value should be affected less than measurements at a higher Q value by the uncertainty of the strength of the magnetic field.

6.4.3 Mass calculation

For calculating the masses, first I summarise the Q values with the various uncertainties in table 6.9. With the central values of the baseline scenario, I obtain

	$\sigma(B_{s1})$ [MeV/c^2]	$\sigma(B_{s2}^*)$ [MeV/c^2]
Fit Model	0.07	0.04
Magnetic field	0.1	0.1
Material description	0.1	0.1
Total syst.	0.16	0.15
Statistical	0.18	0.18

Table 6.9: Summary of Q value uncertainties. In the upper part the three partial systematic uncertainties are listed. The total systematic uncertainty is obtained by adding the independent effects quadratically. The statistical uncertainty is taken from the baseline scenario.

$$\mathrm{Q}(B_{s1}) = 10.25 \pm 0.18\ (\mathrm{stat}) \pm 0.16\ (\mathrm{sys})\ \mathrm{MeV/c^2}\ \text{and} \tag{6.3}$$

$$\mathrm{Q}(B_{s2}^*) = 67.00 \pm 0.18\ (\mathrm{stat}) \pm 0.15\ (\mathrm{sys})\ \mathrm{MeV/c^2}. \tag{6.4}$$

I use the values from [32] for the mass of the B^+, which is 5279.17 ± 0.29 MeV/c^2, for the mass of the charged kaon, which is 493.677 ± 0.016 MeV/c^2, and for the mass of the B^{*+} meson, which I need for the $B^*_{s2} \to B^{*+}K^-$ decays, of 5325.1 ± 0.5 MeV/c^2. I obtain

$$\mathrm{M}(B_{s1}) = 5829.0 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{sys}) \pm 0.5\ (\mathrm{PDG})\ \mathrm{MeV/c^2}\ \text{and} \tag{6.5}$$

$$\mathrm{M}(B^*_{s2}) = 5839.9 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{sys}) \pm 0.3\ (\mathrm{PDG})\ \mathrm{MeV/c^2}. \tag{6.6}$$

6.5 The Natural Width

The analyses so far have not yielded a significant natural width. The detector resolution is sufficient to describe the form of the signal contributions. So it remains to set an upper limit on the width.
To do so, I use the method described in section 4.5. I consider the a priori distribution to be flat in the range from 0 to 10 MeV/c^2, and perform fits in 0.1 MeV/c^2 steps, beginning with 0.05 MeV/c^2. The fits are made using the baseline scenario for background and the Gaussian constraint in the difference between the decays of the B^*_{s2}. The results for the difference between each log likelihood and the best log likelihood can be found in figure E.1. The values of the log likelihood in the histogram are calculated for the value at the center of each bin. The posterior probability density function, that follows from the log likelihood scan, is shown in figure E.2. It is clear from these plots, that the a priori distribution covered a large enough range.
Then I integrate the probability density function from zero to Γ to obtain p(Γ), the probability that the real natural width is less or equal to Γ. The resulting p(Γ) is shown in figure E.3. The 95% credibility level limit is found to be 1.35 MeV/c^2 for the B_{s1} state, and 2.15 MeV/c^2 for the B^*_{s2} state.
If the resolution from Monte Carlo is not identical to the one in data, this can affect the limit on the width. However, the Monte Carlo tends to yield better resolutions than the one of the real detector. More over, high statistics samples, both of Monte Carlo and experimental data, show a resolution which can be described by one dominant Gaussian and minor broader Gaussian distributions. Those additional broader contributions could mimic a natural width. As I determine a limit, effects that make the signal broader just make the limit worse, not better. Therefore I neglect additional systematic uncertainties into the limit calculation.

As with the significance determination, one has to check for effects of alternative background PDFs. I redo the width analysis with the background PDF of e.q. 6.2. The plots for scan with the alternative background PDF are shown in figures E.4 to E.6. For the width of the B_{s1}, I get a limit of 1.75 MeV/c^2. For the width of the B^*_{s2} I get a limit of 1.65 MeV/c^2. The final limit is therefore the limit from the baseline scenario for the B^*_{s2}, and the limit from the fits with the alternative background PDF for the B_{s1}. The plots for the worse scenario in each case are shown in figures

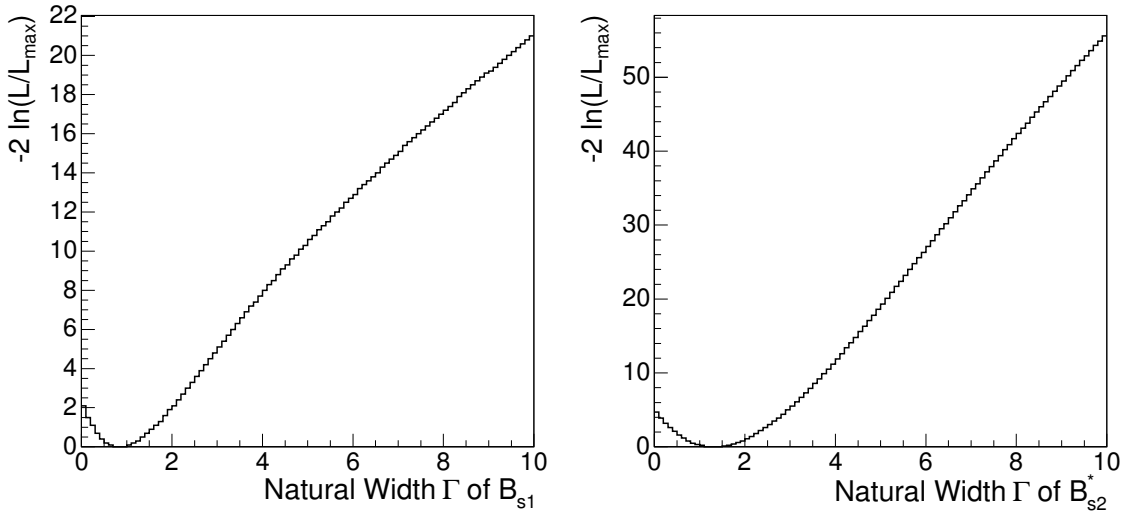

Figure 6.5: Log likelihood scan for the width determination. On the left side, the log likelihood distribution is shown for the B_{s1} state from the fits with the alternative background PDF, on the right side for the B^*_{s2} state from the baseline scenario.

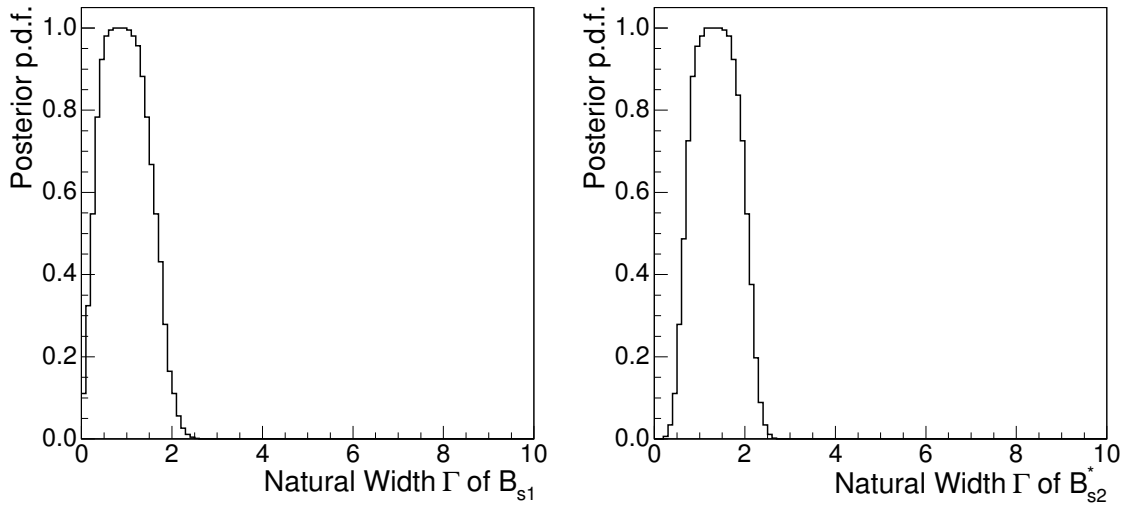

Figure 6.6: Probability density function for the width determination. On the left side, the probability density function is shown for the B_{s1} state from the fits with the alternative background PDF, on the right side for the B^*_{s2} state from the baseline scenario.

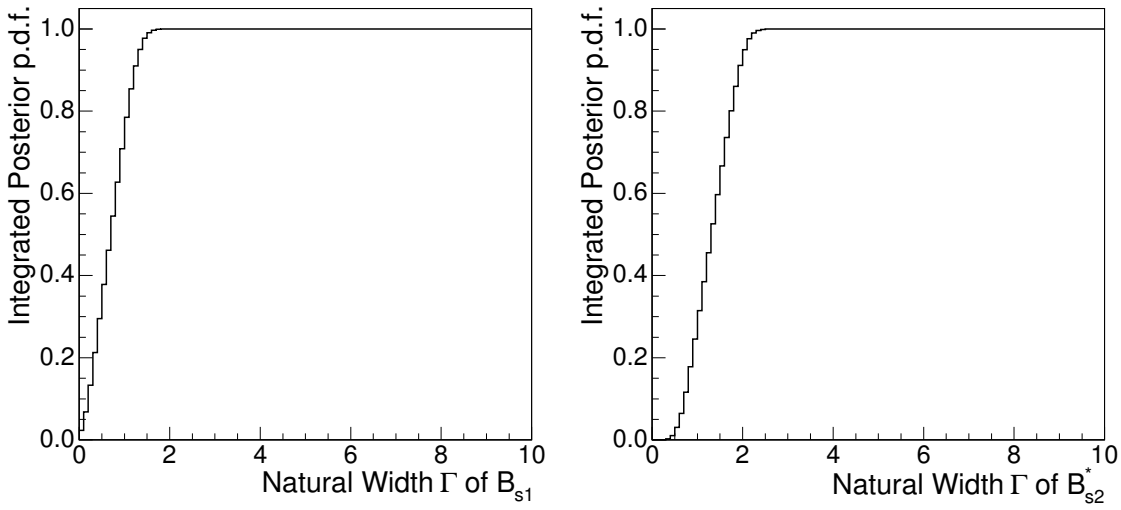

Figure 6.7: Integrated PDF for the width determination. On the left side, the integrated PDF is shown for the B_{s1} state from the fits with the alternative background PDF, on the right side for the B^*_{s2} state from the baseline scenario.

6.5 to 6.7. The final results are

$$\Gamma(B_{s1}) < 1.75 \text{ MeV/c}^2, \tag{6.7}$$

$$\Gamma(B^*_{s2}) < 2.15 \text{ MeV/c}^2. \tag{6.8}$$

6.6 Relative Production and Branching Ratios

To determine the relative production and branching ratios, it is necessary to know how the efficiency changes with the Q value. As this dependency can be different in the various channels, it is necessary to fit for the amount of signal in each channel separately.

6.6.1 Relative Efficiency Determination

I determine the efficiency from the Monte Carlo samples. To do so, I use a weighted Q value distribution of the Monte Carlos samples, with weights adjusted to make the generated mass of the B_s^{**} mesons flat in Q value. As I don't need the absolute efficiency, but only the relative efficiency for different Q values, I use normalised distributions in each channel. Then I perform a linear χ^2 fit for each sample, and each signal contribution. To fit for each signal contribution separately allows to anchor the constant term of the linear fit to the central value of the corresponding signal contribution, by using $a+b(Q-0.01025)$, $a+b(Q-0.06700)$, and $a+b(Q-0.0219)$ as fit functions. This minimises the uncertainty for the corresponding efficiency. As the efficiency drops very sharply at very low Q values, the fit is only performed in

the range from 5 MeV/c^2 to 100 MeV/c^2.
The graphical representation of the three fits in each channel doesn't change, so only one plot for each channel can be found in figures F.1 to F.3. The fit results can be found in table 6.10. The fit probability calculated from the χ^2 fit is 51.9% for the $B^+ \to J/\psi K^+$ channel, 6.6% in the $B^+ \to \overline{D^0}\pi^+$ channel, and 30.6% in the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel. The slope is small enough that the uncertainty in Q

	Constant	slope [c^2/GeV]
$B^+ \to J/\psi K^+$		
B_{s1}	1.049±0.009	-0.78 ±0.17
$B^*_{s2} \to B^+K^-$	1.005±0.005	-0.78 ±0.17
$B^*_{s2} \to B^{*+}K^-$	1.040±0.007	-0.78 ±0.17
$B^+ \to \overline{D^0}\pi^+$		
B_{s1}	0.998±0.017	0.1 ±0.4
$B^*_{s2} \to B^+K^-$	1.002±0.012	0.1 ±0.4
$B^*_{s2} \to B^{*+}K^-$	0.999±0.014	0.1 ±0.4
$B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$		
B_{s1}	0.92±0.03	1.3 ±0.7
$B^*_{s2} \to B^+K^-$	1.00±0.02	1.3 ±0.7
$B^*_{s2} \to B^{*+}K^-$	0.94±0.03	1.3 ±0.7

Table 6.10: Results of efficiency fits.

value for the signal contribution has no effect on the efficiency.
For getting the relative efficiency, I use the $B^*_{s2} \to B^+K^-$ decay as reference, and divide the efficiencies for the other decays by the efficiency of that one. The final relative efficiencies are shown in table 6.11. Systematic uncertainties in this determi-

Relative efficiency	for B_{s1} signal	for $B^*_{s2} \to B^{*+}K^-$ signal
$B^+ \to J/\psi K^+$	1.044± 0.014	1.035±0.012
$B^+ \to \overline{D^0}\pi^+$	1.00 ± 0.03	1.00 ± 0.03
$B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$	0.92±0.05	0.94 ± 0.05

Table 6.11: Results of efficiency determination. This table corresponds to figures F.1 to F.3.

nation are very small, as issues like track efficiency etc., that influence the absolute efficiency, cancel out of the relative efficiency. Small remaining effects from an imperfect description of the B_s^{**} momenta in the Monte Carlo simulation should be much smaller than the statistical uncertainty from the fits, as the additional energy for the kaon of the B_s^{**} decay from that decay is small compared to the energy, it takes from the boost of the B_s^{**} rest frame. Therefore I don't assign an extra systematic uncertainty on the relative efficiencies.

6.6.2 Simultaneous Fit

To gather the yields for the signal contributions in each decay channel separately, I perform a simultaneous fit. The Q values of the signals and the widths - if needed - have to be the same in all channels, the yields and the background description can differ in each channel.
The simultaneous fit is performed three times to get not only the statistical, but as well systematical uncertainties. First with the baseline scenario, second with free natural widths Γ, and at last with the alternative background description. The fit projection for the baseline scenario fit can be found in figure 6.8. The fit projection for free natural width and the alternative background are shown in figures G.1 and G.2.

Variable	Baseline Value [MeV/c^2]	Alternative background Value[MeV/c^2]	free Γ Value Γ[MeV/c^2]
Q(B_{s1})	10.23 ± 0.17	10.20±0.18	10.29±0.19
Q($B_{s2}^* \to B^{*+}K^-$)	66.94 ± 0.18	66.99±0.18	67.0±0.2
Δm(B^{*+}, B^+)	44.0± 0.5	43.9 ±0.5	45.0 ±0.7
$\Gamma(B_{s1})$	0	0	0.7±0.5
$\Gamma(B_{s2}^*)$	0	0	1.3±0.7
	[Candidates]		
$B^+ \to J/\psi K^+$			
N(B_{s1})	14 ± 5	14 ± 5	18 ± 6
N($B_{s2}^* \to B^+K^-$)	40 ± 8	40 ± 8	48 ± 10
N($B_{s2}^* \to B^{*+}K^-$)	-3 ± 3	-4 ± 3	3 ± 6
$B^+ \to \overline{D^0}\pi^+$			
N(B_{s1})	40 ± 8	40 ± 8	50 ± 11
N($B_{s2}^* \to B^+K^-$)	140 ± 17	139 ± 17	169 ± 24
N($B_{s2}^* \to B^{*+}K^-$)	25 ± 9	20 ± 8	37 ± 13
$B^+ \to \overline{D^0}\pi^+\pi^-\pi^+$			
N(B_{s1})	12 ± 5	12 ± 5	15 ± 6
N($B_{s2}^* \to B^+K^-$)	74 ± 12	73 ± 12	90 ± 16
N($B_{s2}^* \to B^{*+}K^-$)	14 ± 6	10 ± 7	17 ± 8

Table 6.12: Results for simultaneous fits.

The fit results are shown on table 6.12. To get the relative production and branching ratios, I divide the yields of the B_{s1} and the $B_{s2}^* \to B^{*+}K^-$ signal with the relative efficiency to the $B_{s2}^* \to B^+K^-$ signal. The result for this is shown in table 6.13, where I have taken the unrounded yields to multiply with the efficiency.

Next I calculate for each fit model and each channel the relative production of B_{s1} to B_{s2}^* times the branching ratio into $B^{(*)+}K^-$, and the relative branching ratio of the B_{s2}^* decays. I take the numbers with the higher relative uncertainty into the numerator, so that the uncertainty is more Gaussian. The results can be seen in table 6.14.

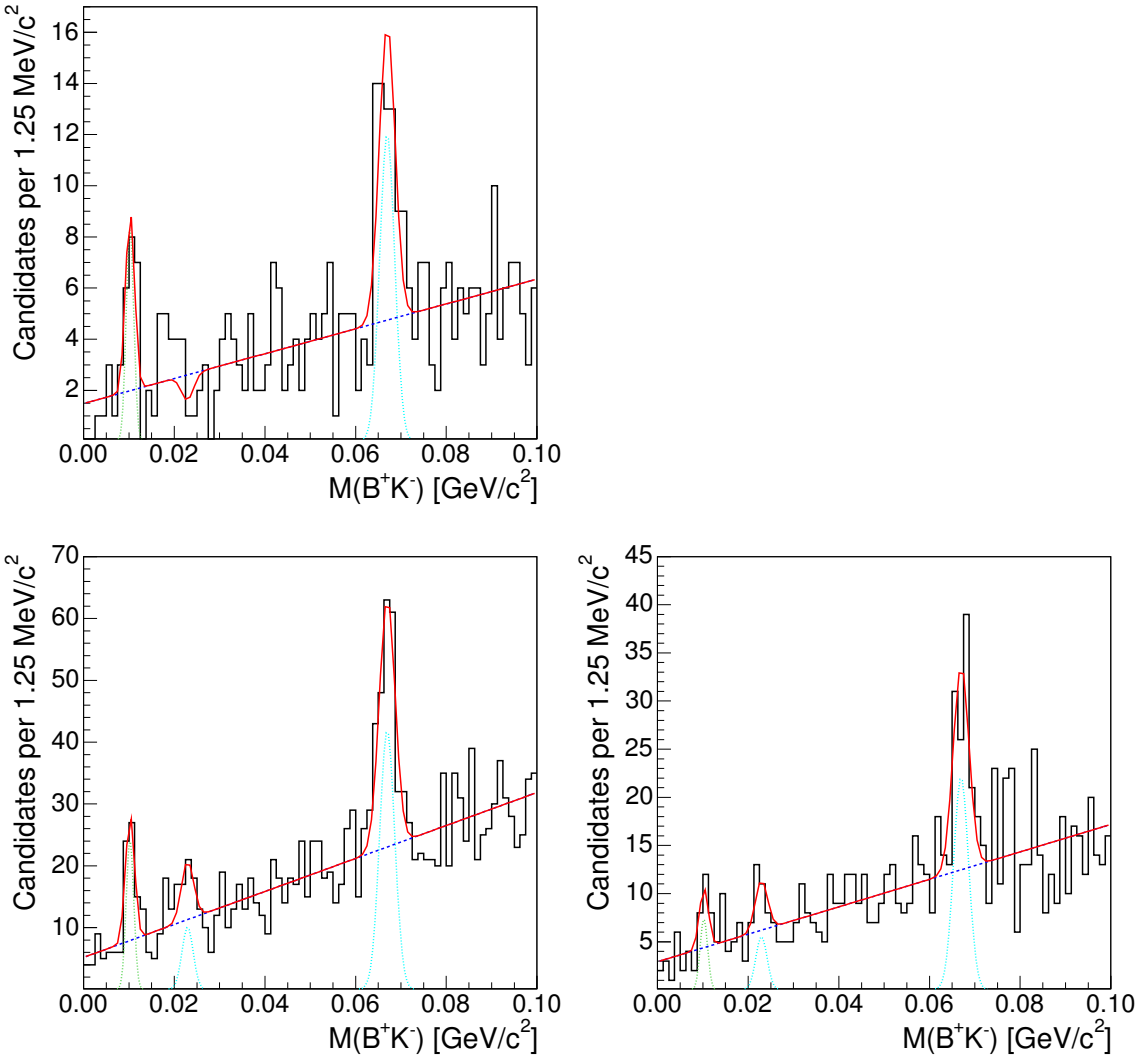

Figure 6.8: Simultaneous fits for baseline scenario. The upper plot is for the $B^+ \to J/\psi K^+$ channel. The lower left plot is for the $B^+ \to \overline{D^0}\pi^+$ channel, and the lower right plot for the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel.

	Baseline	Alternative background	free Γ
$B^+ \to J/\psi K^+$			
N(B_{s1})·ϵ	13 ± 5	13 ± 5	17 ± 6
N($B^*_{s2} \to B^+K^-$)·ϵ	40 ± 8	40 ± 8	48 ± 10
N($B^*_{s2} \to B^{*+}K^-$)·ϵ	-4 ± 3	-4 ± 3	3 ± 6
SUM B^*_{s2}	37 ± 9	36 ± 9	51 ± 12
$B^+ \to \overline{D^0}\pi^+$			
N(B_{s1})·ϵ	40 ± 8	40 ± 8	50 ± 11
N($B^*_{s2} \to B^+K^-$)·ϵ	140 ± 17	139 ± 17	169 ± 24
N($B^*_{s2} \to B^{*+}K^-$)·ϵ	25 ± 9	20 ± 8	37 ± 13
SUM B^*_{s2}	165 ± 19	159 ± 19	206 ± 27
$B^+ \to \overline{D^0}\pi^+\pi^-\pi^+$			
N(B_{s1})·ϵ	13 ± 6	13 ± 6	17 ± 7
N($B^*_{s2} \to B^+K^-$)·ϵ	74 ± 12	73 ± 12	90 ± 16
N($B^*_{s2} \to B^{*+}K^-$)·ϵ	15 ± 7	11 ± 7	18 ± 8
SUM B^*_{s2}	89 ± 14	84 ± 14	108 ± 18

Table 6.13: Yields times relative efficiency.

	Baseline	Alternative background	free Γ
$B^+ \to J/\psi K^+$			
Production	0.4±0.2	0.4 ± 0.2	0.34 ± 0.19
Branching	-0.06±0.09	-0.10 ± 0.09	0.07 ± 0.13
$B^+ \to \overline{D^0}\pi^+$			
Production	0.24 ± 0.08	0.25 ± 0.08	0.24 ± 0.09
Branching	0.18±0.08	0.14 ± 0.08	0.22 ± 0.11
$B^+ \to \overline{D^0}\pi^+\pi^-\pi^+$			
Production	0.15 ± 0.09	0.15 ± 0.09	0.15 ± 0.09
Branching	0.20±0.13	0.15 ± 0.12	0.20 ± 0.13
Averaged			
Production	0.21 ± 0.06	0.22 ± 0.06	0.21 ± 0.06
Branching	0.10 ± 0.05	0.06 ± 0.05	0.17 ±0.07

Table 6.14: Production and branching ratios for each decay channel. While the real branching can't be less than zero, a fluctuation to negative values has to be taken into account for not biasing the averaging.

I take the differences in the central value to the baseline scenario as systematic uncertainties. For the production fraction I obtain

$$\frac{\sigma(B_{s1}) \cdot BR(B_{s1} \to B^{*+}K^-)}{\sigma(B^*_{s2}) \cdot BR(B^*_{s2} \to B^{+(*)}K^-)} = 0.21 \pm 0.06 \text{ (stat)} \pm 0.01 \text{ (sys)}. \qquad (6.9)$$

For the relative branching ratio I obtain

$$\frac{BR(B^*_{s2} \to B^{*+}K^-)}{BR(B^*_{s2} \to B^+K^-)} = 0.10 \pm 0.05 \text{ (stat)} \quad {}^{+0.07}_{-0.04} \text{ (sys)}. \tag{6.10}$$

6.7 B^{*+} - B^+ mass difference

As the B^*_{s2} can decay into B^+ and B^{*+}, the difference in Q value of the direct and the indirect decay mode makes it possible to measure the mass difference between the two B^+ states, i.e. the B^+ hyperfine splitting. I perform the fits as in the baseline scenario, with the alternative background function, and with the linear background but free natural widths again, this time leaving the Q value difference between the direct and indirect decay of the B^*_{s2} completely free. The results for the parameters relevant for the indirect B^*_{s2} decay are listed in table 6.15. I vary as well the starting parameters in both fits, as the signal contribution is relatively small, but this doesn't affect the result. To determine the mass difference between B^+ and B^{*+}, I take the

Variable	Baseline Value [MeV/c^2]	Alternative background Value[MeV/c^2]	free Γ Value Γ[MeV/c^2]
Δm(B^{*+}, B^+)	44.7± 0.6	44.8±0.7	45.0 ±0.7
Γ	0	0	1.3±0.7
	[Candidates]		
N(B^*_{s2}, indirect)	39 ± 11	29 ± 11	55 ± 16

Table 6.15: Results for fit with free Q value difference between B^*_{s2} decays.

baseline scenario and use the difference between the central values of the fits as systematic uncertainty. The uncertainties from tracking and the magnetic field are negligible, as they have the same effect on both contributions and would just shift the whole Q value distribution up or down.
I check a potential influence of the missing photon in the $B^{*+} \to B^+$ decay with the Monte Carlo simulations. To do so, I subtract from the Q value, that one measures in the Monte Carlos sample, the generated B^{**}_s mass, and add the masses of the generated B^+ and kaon mass. I limit the study on Q values in the range from 5 to 25 MeV/c^2. The resulting distribution is shown in figure 6.9. Then I perform a χ^2 fit with a Gaussian. I obtain a mean value of minus 45.4981± 0.0010 MeV/c^2. The mass difference used in EvtGen is 45.5 MeV/c^2. The difference of 0.0019 MeV/c^2 is not significant.
As result I obtain

$$\text{M}(B^{*+}) - \text{M}(B^+) = 44.7 \pm 0.6 \text{ (stat)} \pm 0.3 \text{ (sys) MeV/c}^2. \tag{6.11}$$

This is a bit lower than the inclusive measurements of 45.78 ± 0.35 MeV/c^2 [32], but is less than two σ away from it. The present measurement is more precise than the exclusive DELPHI measurement of 45.7 ± 0.9 MeV/c^2 and as well compatible with that.

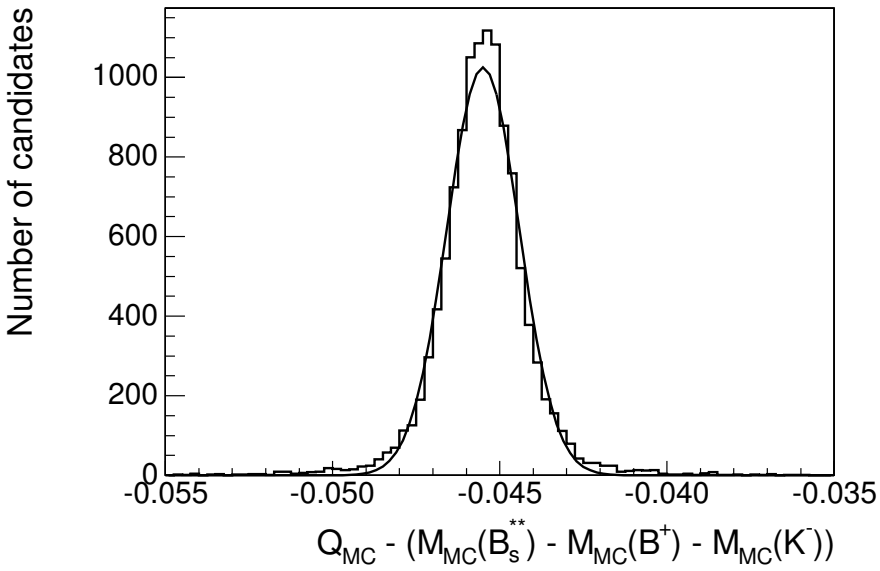

Figure 6.9: Check of influence of B^* photon. The difference between the true simulated Q value and the reconstructed Q value of the simulation peaks at the simulated mass difference of B^{*+} and B^+.

Chapter 7

Conclusions

In the presented analysis, I have searched for B_s^{**} mesons and aimed to determine their properties. I have reconstructed candidates for B_s^{**} mesons using data from the CDF II detector. A selection has been performed considering the decay of B_s^{**} mesons in the channels to B^+K^- and $B^{*+}K^-$ with $B^{*+} \to B^+\gamma$. For the B^+ the decay channels $B^+ \to J/\psi K^+$, $B^+ \to \overline{D^0}\pi^+$ and $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ have been considered. Neural networks build with the NeuroBayes package have been utilised in the selection to obtain the best possible result. The selection was optimised using Monte Carlo and experimental data around a peak in the Q value distribution, in whose proximity claims for B_s^{**} mesons by other experiments have been made.
Unbinned maximum likelihood fits have been performed to determine the properties of possible B_s^{**} states. Two clear signals have been found at Q values of 10.25 $\pm$ 0.24 MeV/c^2 and 67.00 $\pm$ 0.23 MeV/c^2. They have been interpreted to stem from the $B_{s1} \to B^{*+}K^-$ and $B_{s2}^* \to B^+K^-$ decay. The B_{s1} has been observed for the first time in the analysis for this thesis and has been published [15].
As the B_{s2}^* can decay to B^+ and B^{*+}, one expects another signal around a Q value of 67.00 MeV/c^2 minus the mass difference between B^+ and B^{*+} of 45.7 $\pm$ 0.9 MeV/c^2 [14]. A signal with a significance of 2.7σ has been found at the expected place. This confirms the interpretation for the signals at 10 and 67 MeV/c^2. The best value for the hyperfine splitting of the B^+ from the analysed data has been found to be 44.7 $\pm$ 0.7 MeV/c^2. The masses for the states have been determined as

$$\mathrm{M}(B_{s1}) = 5829.0 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{sys}) \pm 0.5\ (\mathrm{PDG})\ \mathrm{MeV/c^2}\ \text{and} \tag{7.1}$$

$$\mathrm{M}(B_{s2}^*) = 5839.9 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{sys}) \pm 0.3\ (\mathrm{PDG})\ \mathrm{MeV/c^2}. \tag{7.2}$$

These mass measurements are slightly below the lowest values given by theories. This leaves room for theorists, to improve calculations and use the presented measurement as calibration for the prediction of so far undiscovered states.
The width of the states could be limited at 95% C.L. to

$$\Gamma(B_{s1}) < 1.75\ \mathrm{MeV/c^2},\ \text{and} \tag{7.3}$$

$$\Gamma(B_{s2}^*) < 2.15\ \mathrm{MeV/c^2}. \tag{7.4}$$

Due to the small phase space in the decays of these states, these limits do not challenge the present theories.
For the relative production fraction of B_{s1} and B^*_{s2} multiplied by their branching fraction into the analysed decay channels, I have obtained

$$\frac{\sigma(B_{s1}) \cdot BR(B_{s1} \to B^{*+}K^-)}{\sigma(B^*_{s2}) \cdot BR(B^*_{s2} \to B^{(*)+}K^-)} = 0.21 \pm 0.06 \text{ (stat)} \pm 0.01 \text{ (sys)}. \tag{7.5}$$

This value is considerably below what I have expected from state or spin counting. The reason could be, that state or spin counting are not a good model for the production, or that the B_{s1} has another non-negligible decay channel. A decay e.g. to $B_s\pi^0$, which is isospin suppressed, may be relevant due to the very small phase space for the $B_{s1} \to B^+K^-$ decay.
For the branching fraction of the B^*_{s2} I have obtained

$$\frac{BR(B^*_{s2} \to B^{*+}K^-)}{BR(B^*_{s2} \to B^+K^-)} = 0.10 \pm 0.05 \text{ (stat)}^{+0.07}_{-0.04} \text{ (sys)}. \tag{7.6}$$

The expectation adjusted for the correct Q value is about 0.08. So the value is in good agreement with the expectation, and the interpretation of the two signals as stemming from a single state decaying into two different final states is reasonable.

Overall this analysis has yielded the first observation of the B_{s1} state, and delivered convincing evidence that the interpretation for the signal, that has been seen before, as B^*_{s2} state is correct. The results provide a good test case for models of excited heavy-light quark mesons.

Appendix A

EvtGen Decay tables

```
Alias myB+ B+
Alias myB- B-
Alias myB*+ B*+
Alias myB*- B*-
Alias myJ/psi J/psi


Decay B_s2*0
.2      myB*+  K-                        PHSP;
.2      myB+  K-                        PHSP;
Enddecay
#
Decay anti-B_s2*0
.2      myB*-  K+                        PHSP;
.2      myB-  K+                        PHSP;
Enddecay
#
#
Decay myB*+
.2      myB+ gamma                       PHSP;
Enddecay
#
Decay myB*-
.2      myB- gamma                       PHSP;
Enddecay
#
#
Decay myB+
.11   myJ/psi K+                         SVS;
Enddecay
#
```

```
Decay myB-
.11   myJ/psi K-                          SVS;
Enddecay
#
#
Decay myJ/psi
.11   mu+  mu-               PHOTOS  VLL;
Enddecay
#
End


Alias myB+ B+
Alias myB- B-
Alias myB*+ B*+
Alias myB*- B*-
Alias myJ/psi J/psi


Decay B_s2*0
.2      myB*+  K-                         PHSP;
.2      myB+  K-                          PHSP;
Enddecay
#
Decay anti-B_s2*0
.2      myB*-  K+                         PHSP;
.2      myB-  K+                          PHSP;
Enddecay
#
#
Decay myB*+
.2      myB+ gamma                        PHSP;
Enddecay
#
Decay myB*-
.2      myB- gamma                        PHSP;
Enddecay
#
#
Decay myB+
0.0100   anti-D0  pi+                         PHSP;
0.0090   a_1+ anti-D0                     SVS;
0.0005   anti-D0  rho0 pi+                PHSP;
0.0005   anti-D0  pi-  pi+  pi+           PHSP;
Enddecay
#
```

```
Decay myB-
0.0100   D0  pi-                        PHSP;
0.0090   a_1- D0                    SVS;
0.0005   D0  rho0 pi-               PHSP;
0.0005   D0  pi+  pi-  pi-          PHSP;
Enddecay
Decay D0
0.0383   K-  pi+                    PHSP;
Enddecay
#
Decay anti-D0
0.0383   K+  pi-                    PHSP;
Enddecay
#
Decay a_1+
0.4910   rho0 pi+                   VVS_PWAVE 0.9091 0.0 0.0 0.0 -0.0909
0.0;
Enddecay
#
Decay a_1-
0.4910   rho0 pi-                   VVS_PWAVE 0.9091 0.0 0.0 0.0 -0.0909
0.0;
Enddecay
#
Decay rho0
1.000    pi+ pi-                    VSS;
Enddecay

#
End
```

Appendix B

Resolution plots

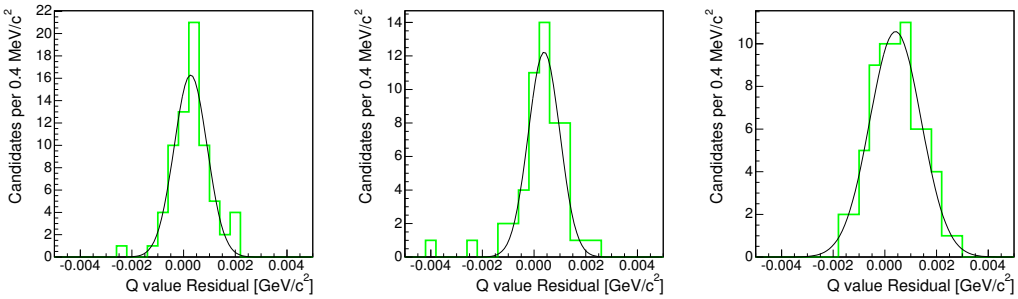

Figure B.1: Resolution fits for $\overline{D^0}\pi^+\pi^+\pi^-$ channel around 10 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 0.64 \pm 0.07$ in the left plot, $\sigma = 0.61 \pm 0.09$ in the middle plot, and $\sigma = 1.00 \pm 0.11$ in the right plot.

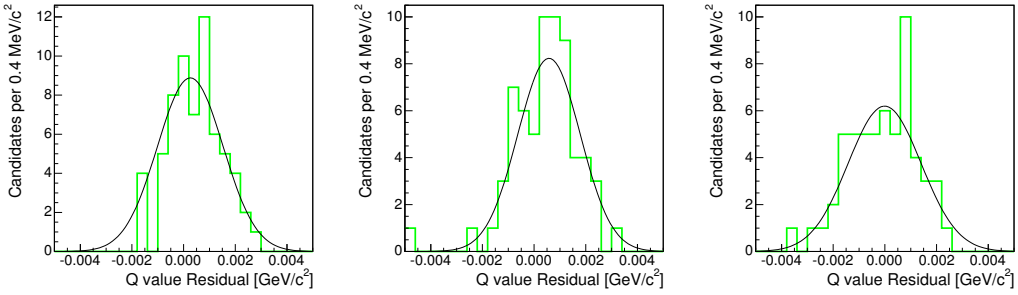

Figure B.2: Resolution fits for $\overline{D^0}\pi^+\pi^+\pi^-$ channel around 21 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 1.3 \pm 0.2$ in the left plot, $\sigma = 1.21 \pm 0.17$ in the middle plot, and $\sigma = 1.41 \pm 0.20$ in the right plot.

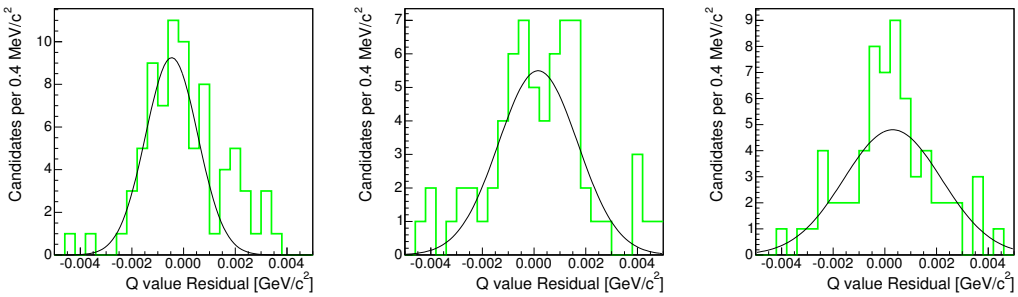

Figure B.3: Resolution fits for $\overline{D^0}\pi^+\pi^+\pi^-$ channel around 67 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 1.01 \pm 0.16$ in the left plot, $\sigma = 1.5 \pm 0.3$ in the middle plot, and $\sigma = 1.89 \pm 0.5$ in the right plot.

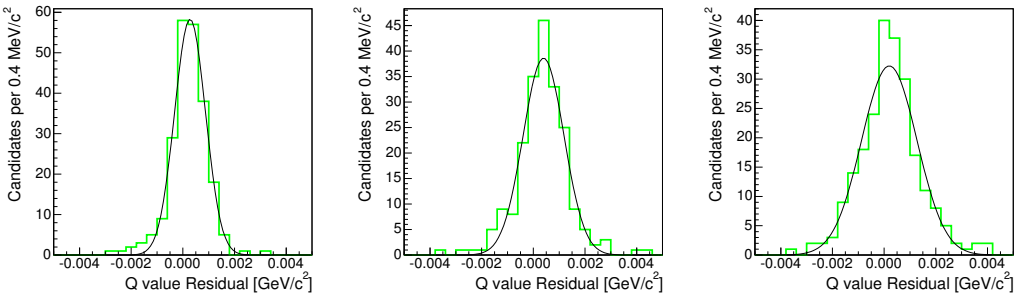

Figure B.4: Resolution fits for $\overline{D^0}\pi^+$ channel around 10 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 0.60 \pm 0.04$ in the left plot, $\sigma = 0.79 \pm 0.06$ in the middle plot, and $\sigma = 1.07 \pm 0.07$ in the right plot.

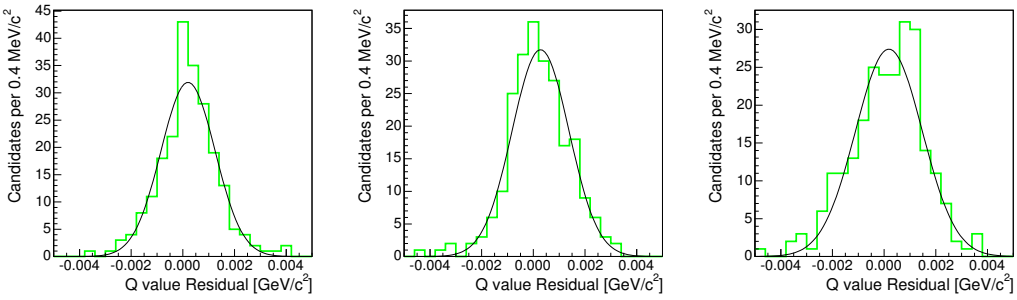

Figure B.5: Resolution fits for $\overline{D^0}\pi^+$ channel around 21 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 1.05 \pm 0.07$ in the left plot, $\sigma = 1.10 \pm 0.07$ in the middle plot, and $\sigma = 1.29 \pm 0.07$ in the right plot.

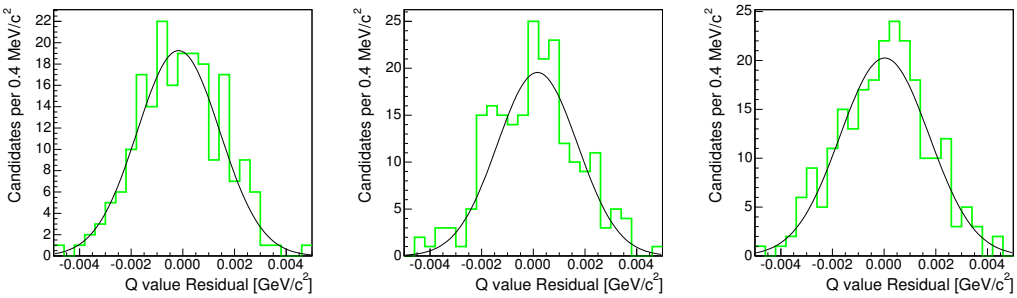

Figure B.6: Resolution fits for $\overline{D^0}\pi^+$ channel around 67 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 1.60 \pm 0.09$ in the left plot, $\sigma = 1.56 \pm 0.11$ in the middle plot, and $\sigma = 1.73 \pm 0.11$ in the right plot.

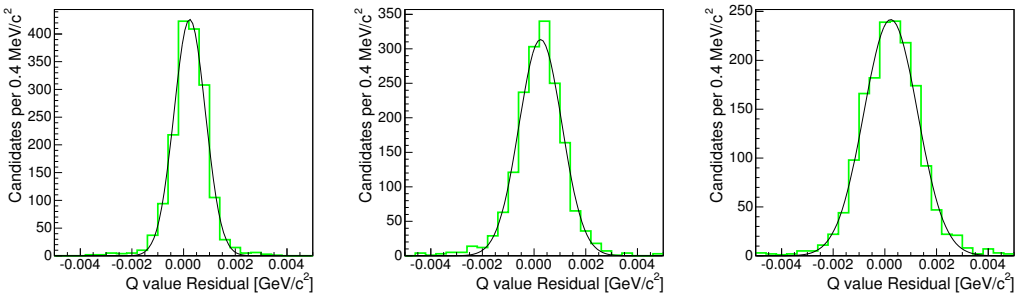

Figure B.7: Resolution fits for $J/\psi K^+$ channel around 10 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 0.609 \pm 0.013$ in the left plot, $\sigma = 0.830 \pm 0.019$ in the middle plot, and $\sigma = 1.04 \pm 0.02$ in the right plot.

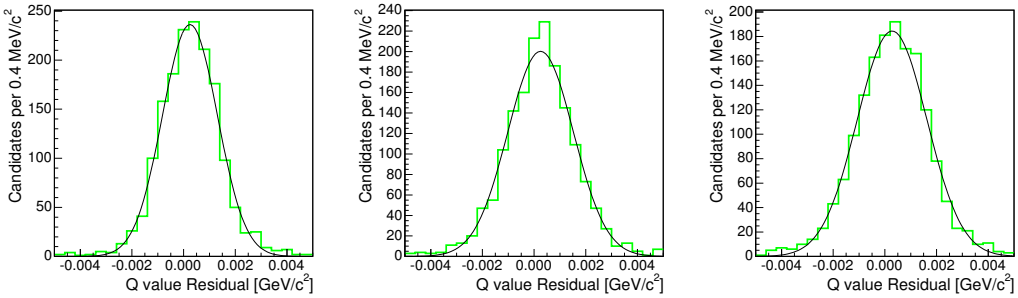

Figure B.8: Resolution fits for $J/\psi K^+$ channel around 21 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 1.07 \pm 0.02$ in the left plot, $\sigma = 1.27 \pm 0.03$ in the middle plot, and $\sigma = 1.35 \pm 0.03$ in the right plot.

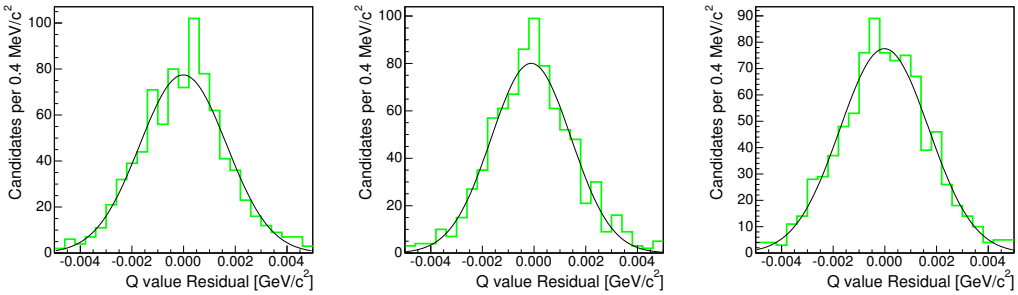

Figure B.9: Resolution fits for $J/\psi K^+$ channel around 67 MeV/c^2. Each plot has an χ^2 minimising fit using a Gaussian, yielding $\sigma = 1.68 \pm 0.05$ in the left plot, $\sigma = 1.55 \pm 0.05$ in the middle plot, and $\sigma = 1.72 \pm 0.05$ in the right plot.

Appendix C

Fitter Validation

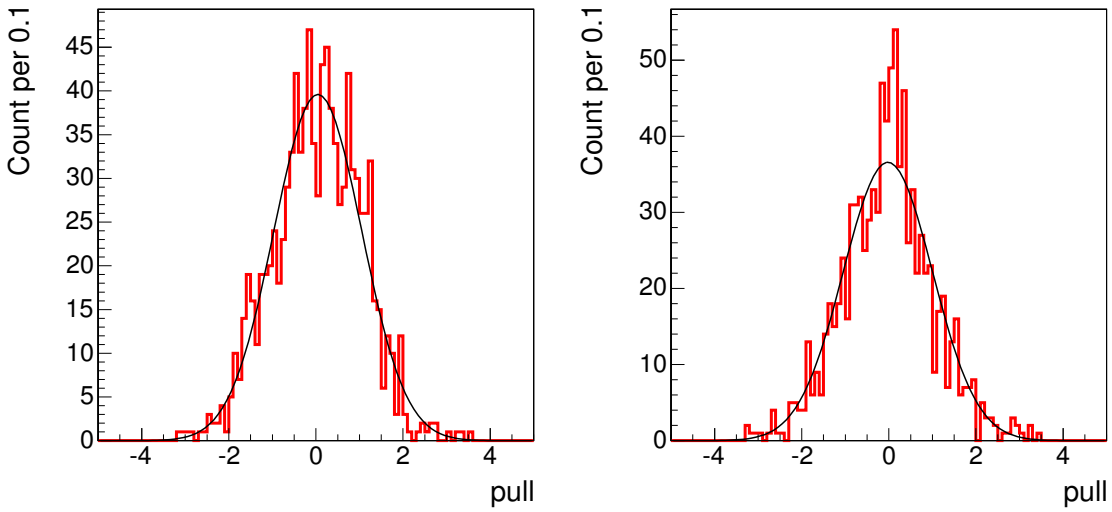

Figure C.1: Fitter validation for the Q value of the B_{s1} decay. The left plot is for the toy sample with natural width zero, the right one for the sample with finite width.

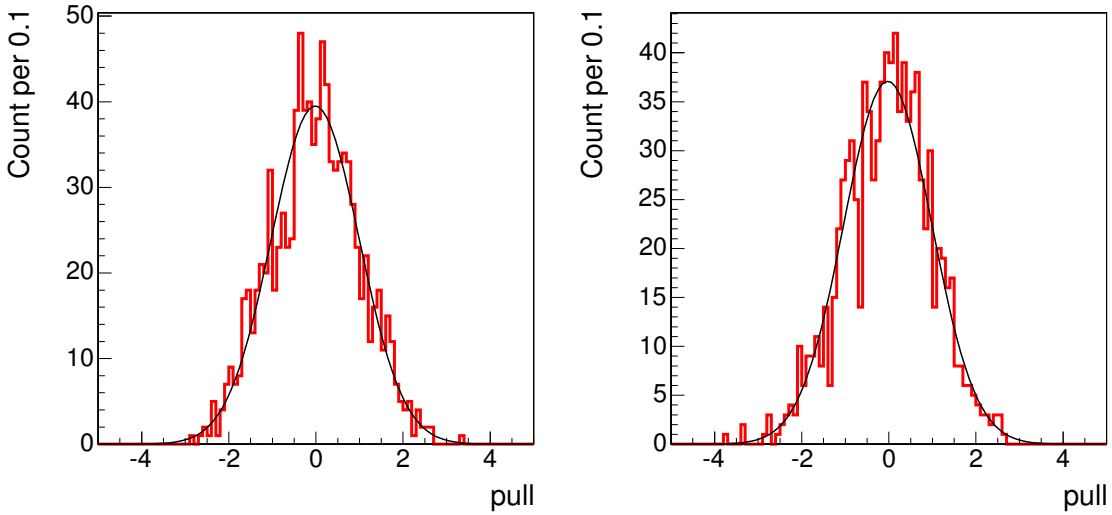

Figure C.2: Fitter validation for the Q value of the B^*_{s2} decay. The left plot is for the toy sample with natural width zero, the right one for the sample with finite width.

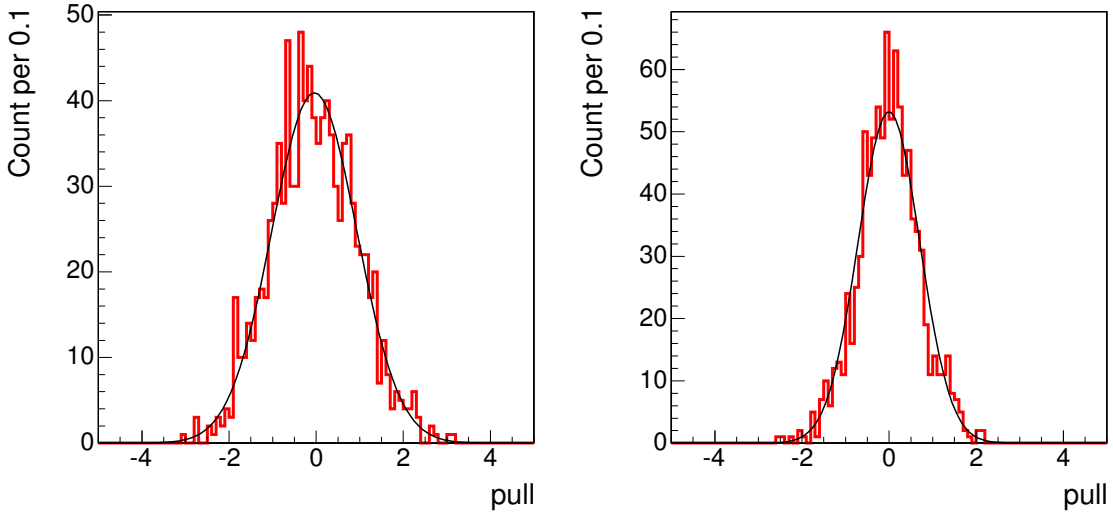

Figure C.3: Fitter validation for the Q value difference of the direct and indirect B^*_{s2} decay. The left plot is for the toy sample with natural width zero, the right one for the sample with finite width.

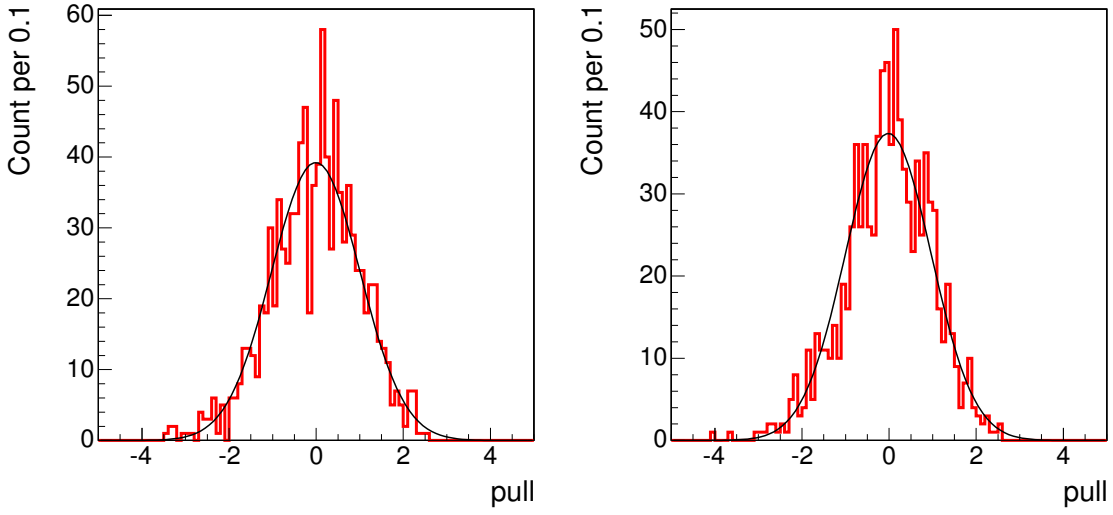

Figure C.4: Fitter validation for the amount of the B_{s1} signal. The left plot is for the toy sample with natural width zero, the right one for the sample with finite width.

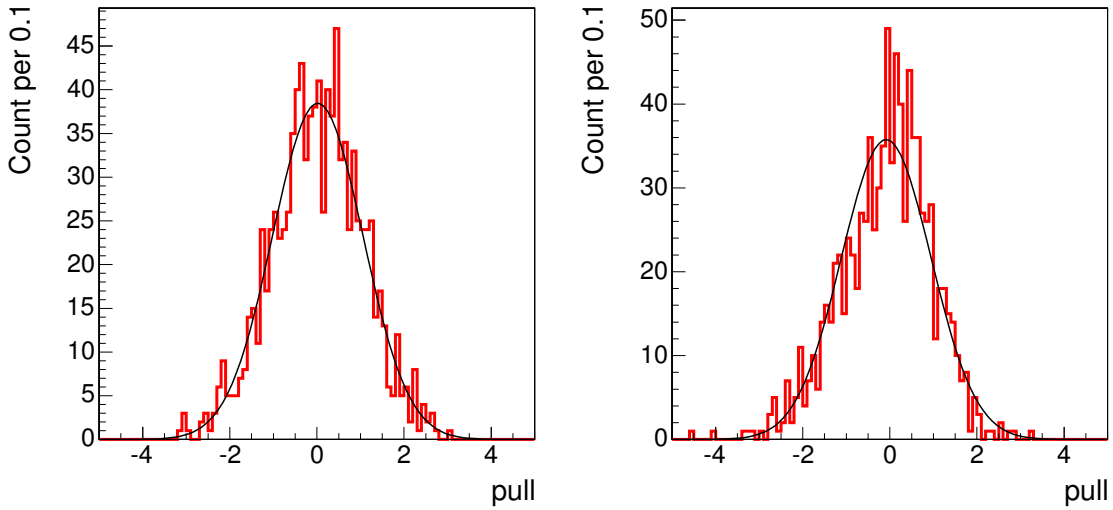

Figure C.5: Fitter validation for the amount of the B^*_{s2} signal in the direct decay. The left plot is for the toy sample with natural width zero, the right one for the sample with finite width.

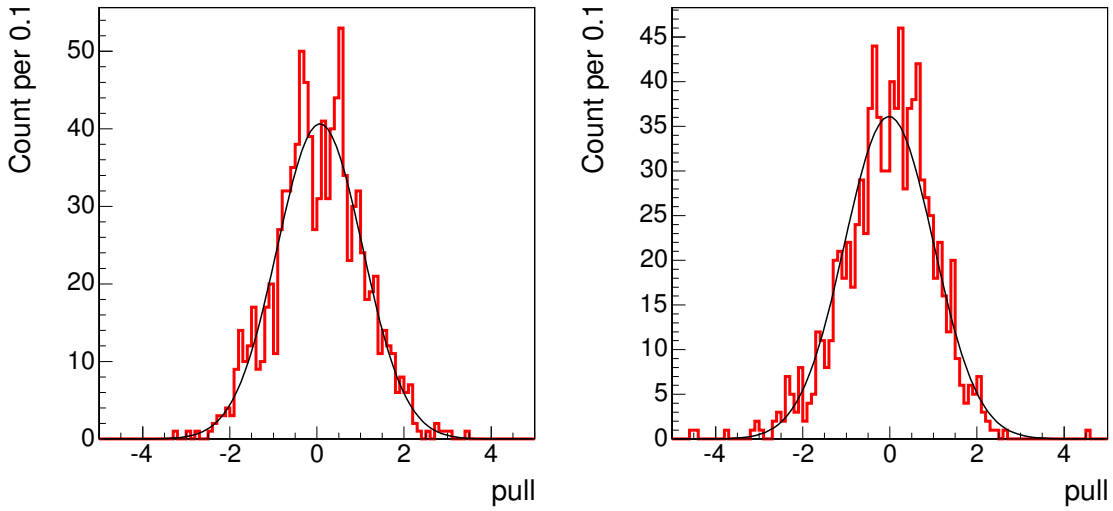

Figure C.6: Fitter validation for the amount of the B^*_{s2} signal in the indirect decay. The left plot is for the toy sample with natural width zero, the right one for the sample with finite width.

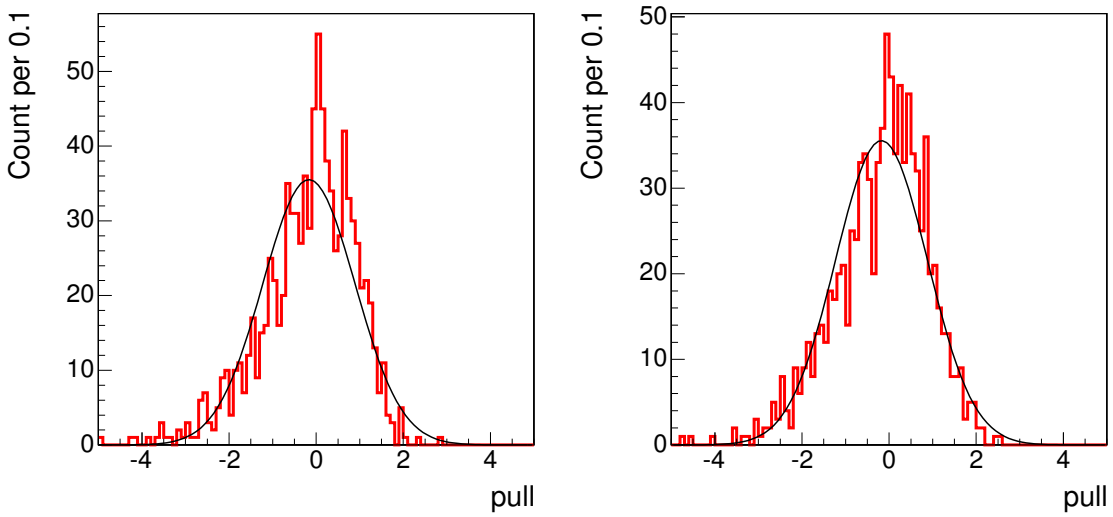

Figure C.7: Fitter validation for the Γ values. The left plot is for B_{s1}, the right for the B^*_{s2}.

Appendix D

Fits to Experimental Data, Sum and Separate Fits

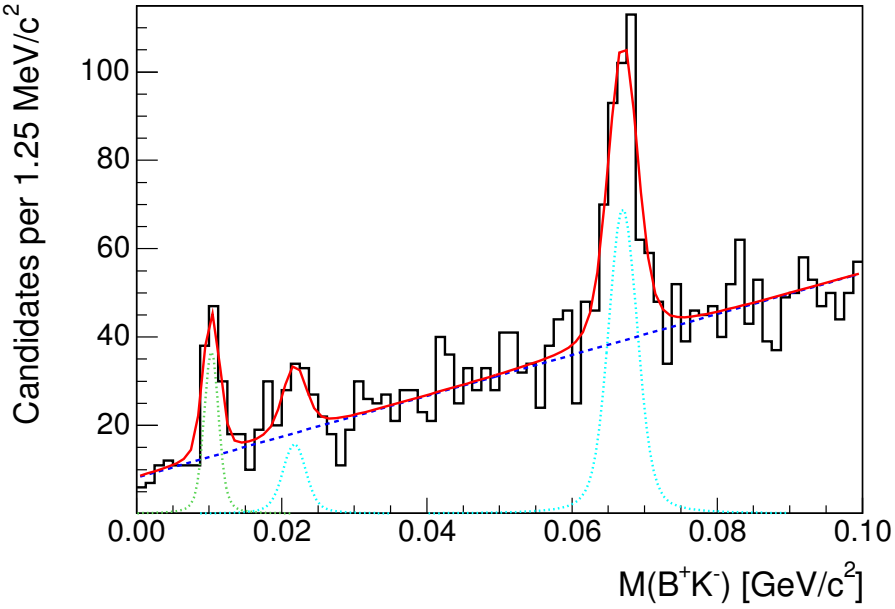

Figure D.1: Fit with free natural width.

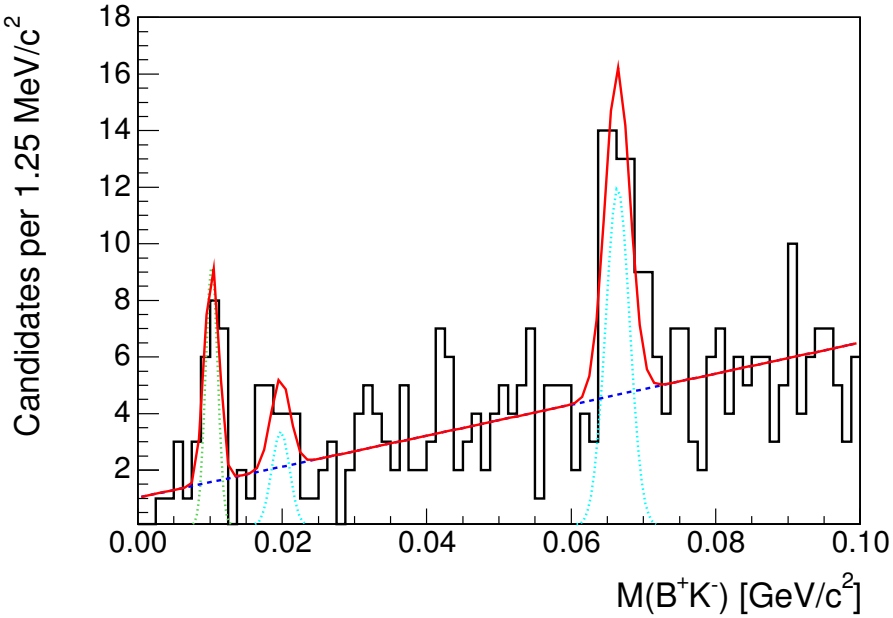

Figure D.2: Fit to $B^+ \to J/\psi K^+$ channel.

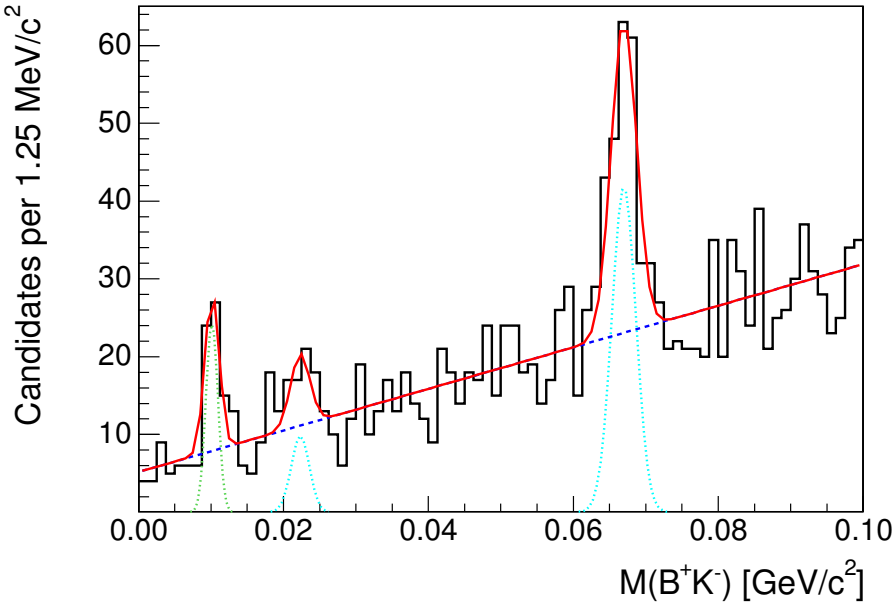

Figure D.3: Fit to $B^+ \to \overline{D^0}\pi^+$ channel.

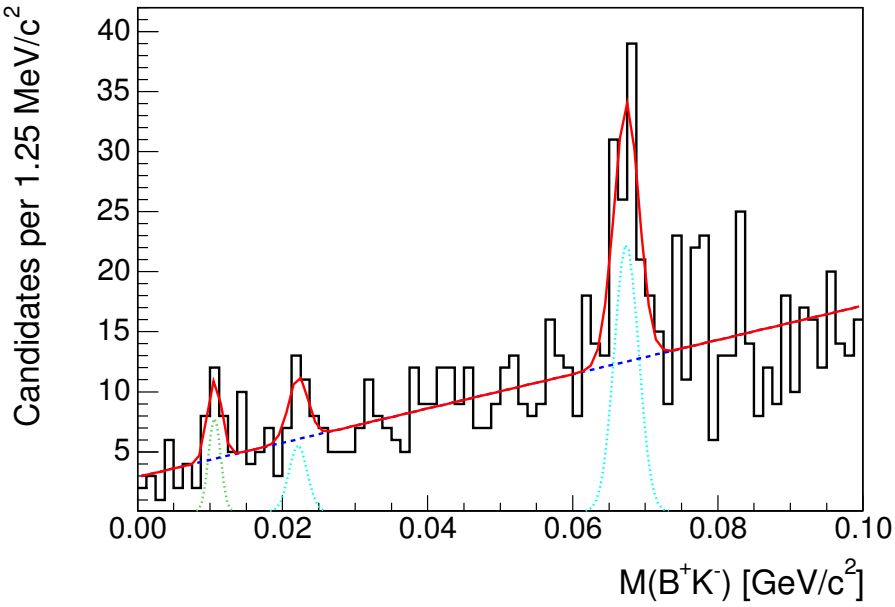

Figure D.4: Fit to $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel.

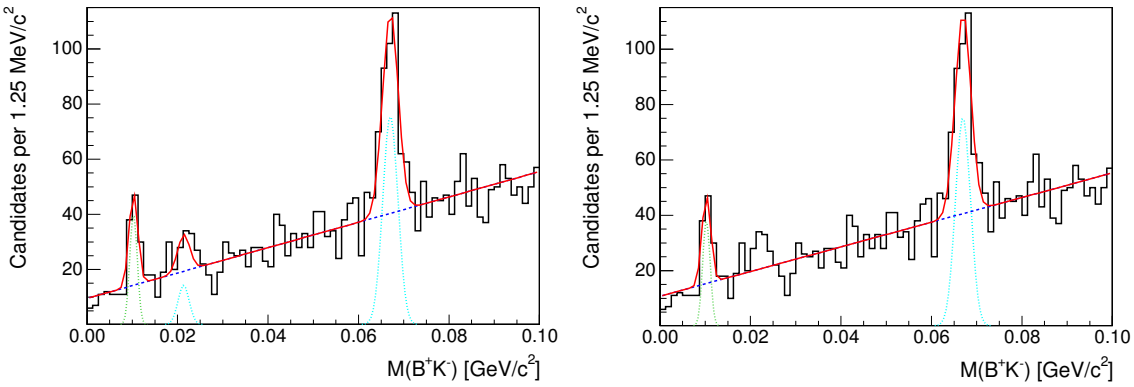

Figure D.5: Fit with fixed mass difference between B^+ and B^{*+} and without indirect B^*_{s2} decay signal. The loglikelihood difference indicates a significance of 3.7 σ for the omitted signal.

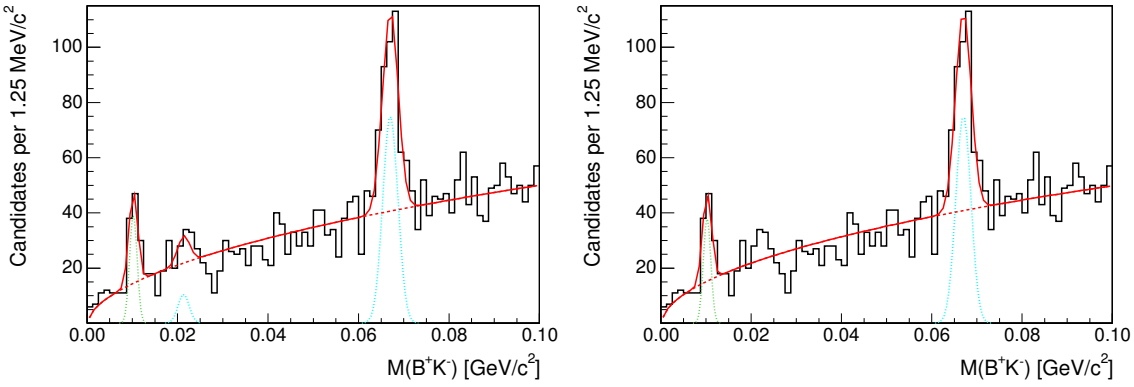

Figure D.6: Fit with alternative background description. The fit on the left side is for a fixed mass difference between B^+ and B^{*+}. On the right side the indirect B^*_{s2} decay signal is omitted. The loglikelihood difference indicates a significance of 2.7 σ for the omitted signal.

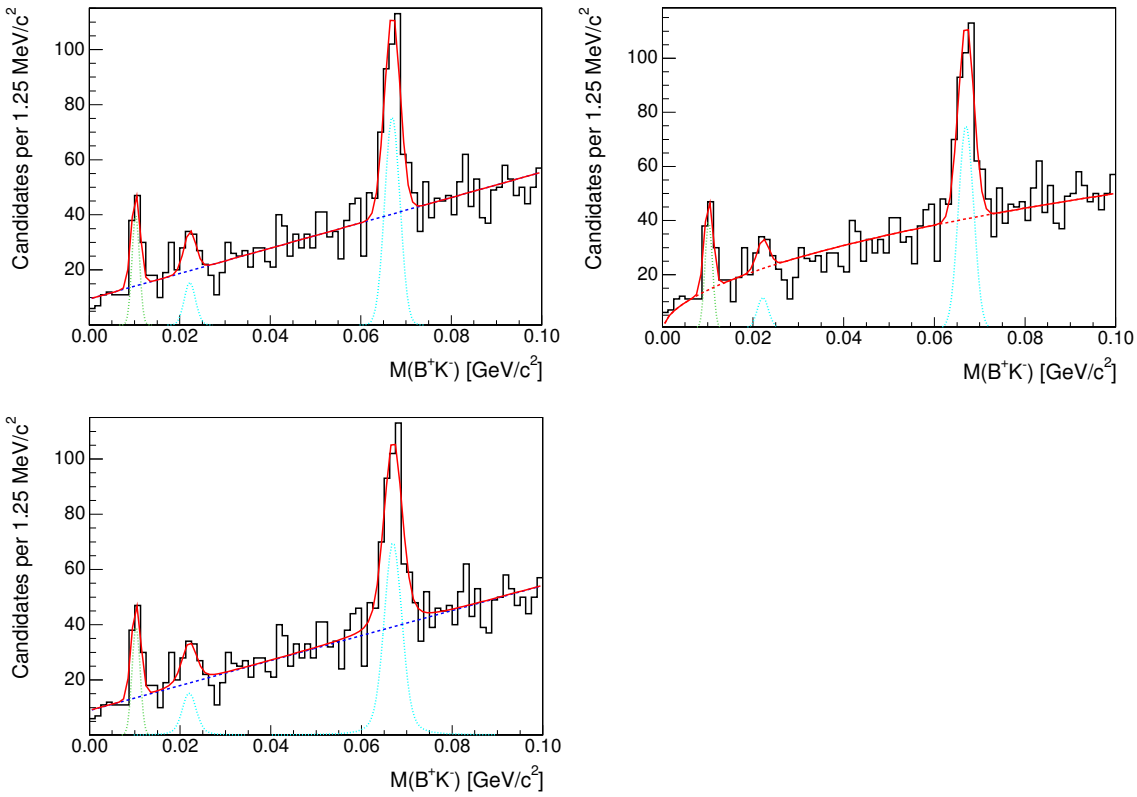

Figure D.7: Fits with free Q value difference between direct and indirect B^*_{s2} decays. The upper left plot is for the baseline scenario background, the upper right plot for the alternative background model. The lower plot is for free natural width Γ.

Appendix E

Plots for Limit on the Widths

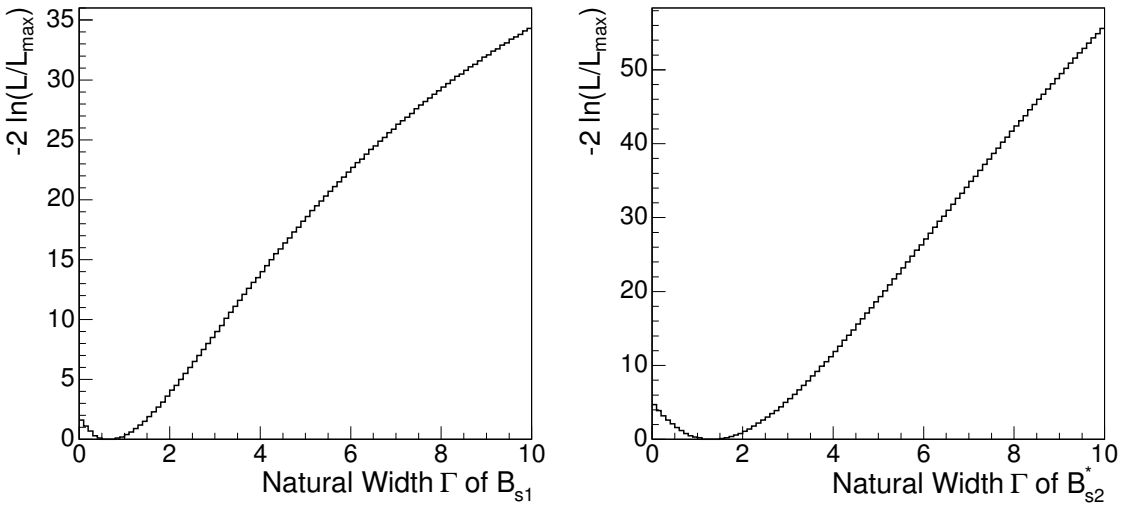

Figure E.1: Log likelihood scan for the width determination. On the left side, the log likelihood distribution is shown for the B_{s1} state, on the right side for the B^*_{s2} state.

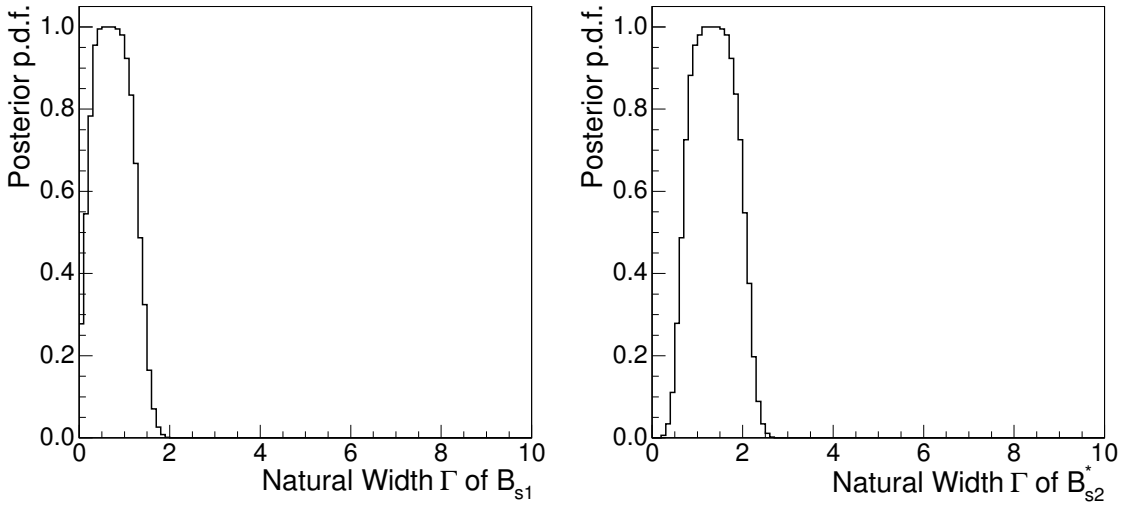

Figure E.2: Integrated PDF for the width determination. On the left side, the integrated PDF is shown for the B_{s1} state, on the right side for the B^*_{s2} state.

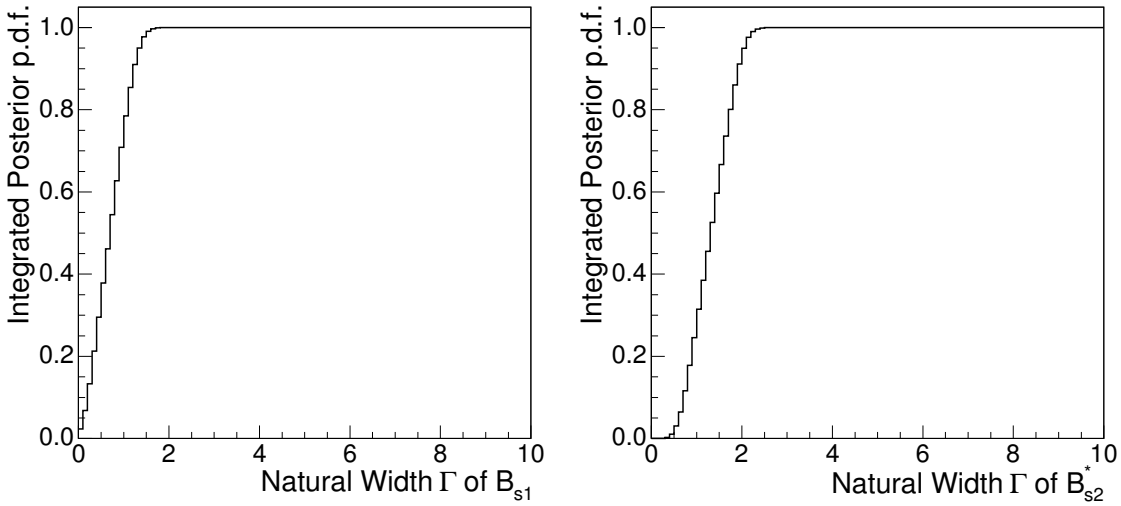

Figure E.3: Probability density function for the width determination. On the left side, the probability density function is shown for the B_{s1} state, on the right side for the B^*_{s2} state.

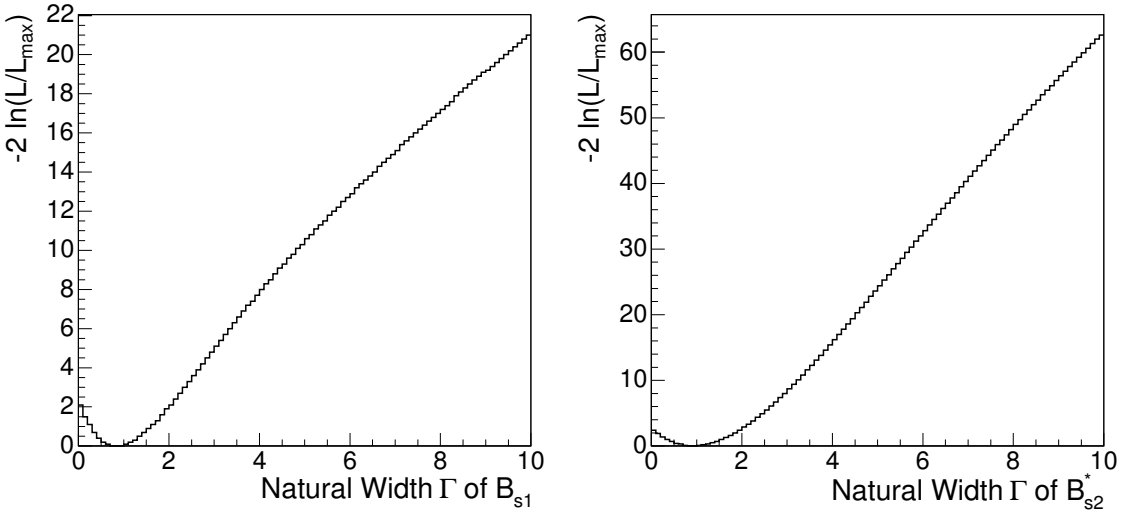

Figure E.4: Log likelihood scan for the width determination with alternative background PDF. On the left side, the log likelihood distribution is shown for the B_{s1} state, on the right side for the B^*_{s2} state.

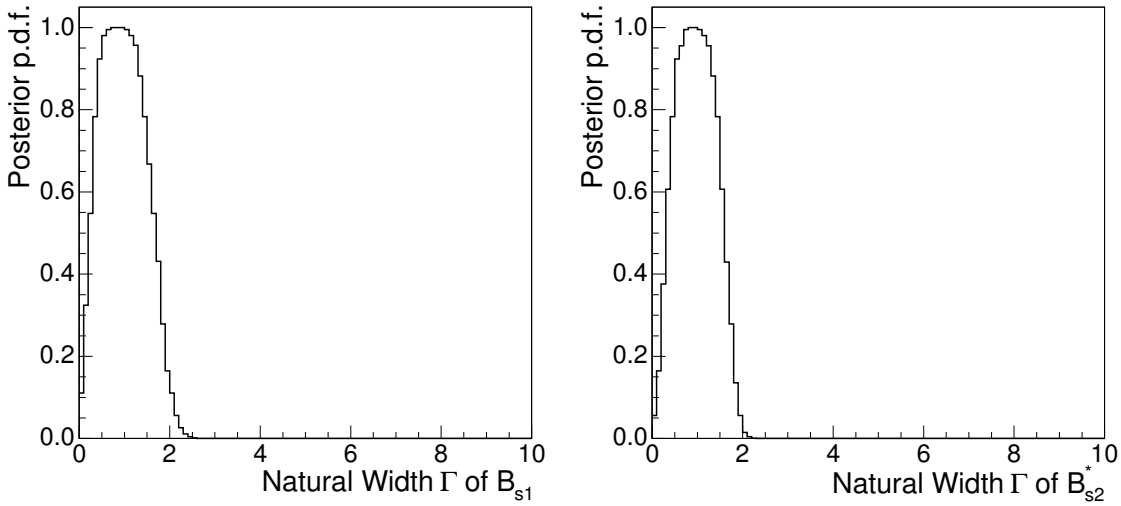

Figure E.5: PDF for the width determination with alternative background PDF. On the left side, the probability density function is shown for the B_{s1} state, on the right side for the B^*_{s2} state.

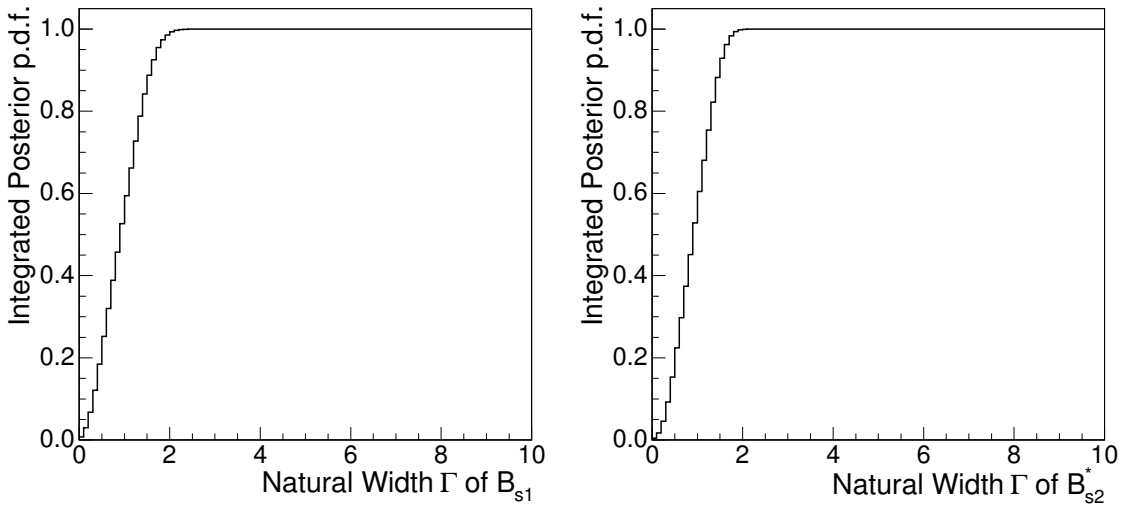

Figure E.6: Integrated PDF for the width determination with alternative background PDF. On the left side, the integrated PDF is shown for the B_{s1} state, on the right side for the B^*_{s2} state.

Appendix F

Plots for Efficiency Calculations

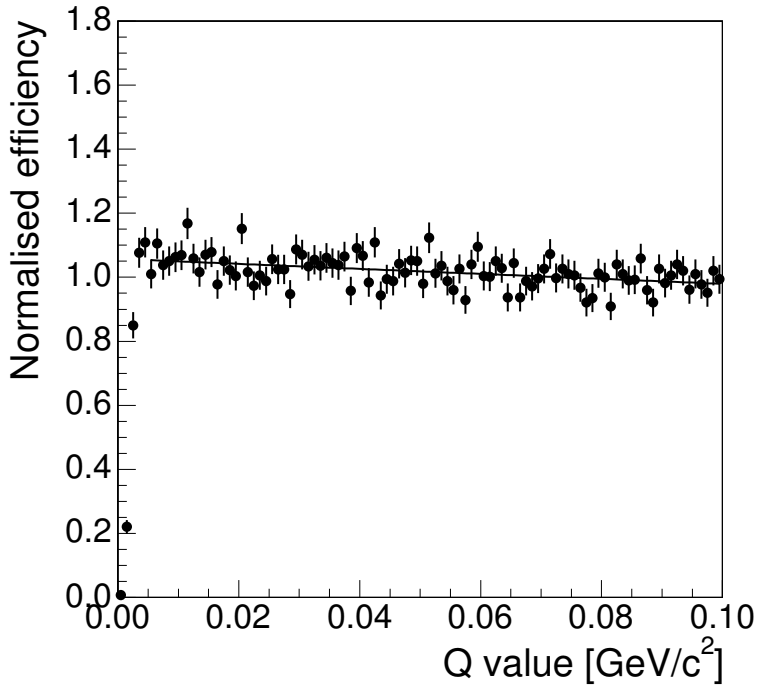

Figure F.1: Relative efficiency for the $B^+ \to J/\psi K^+$ channel.

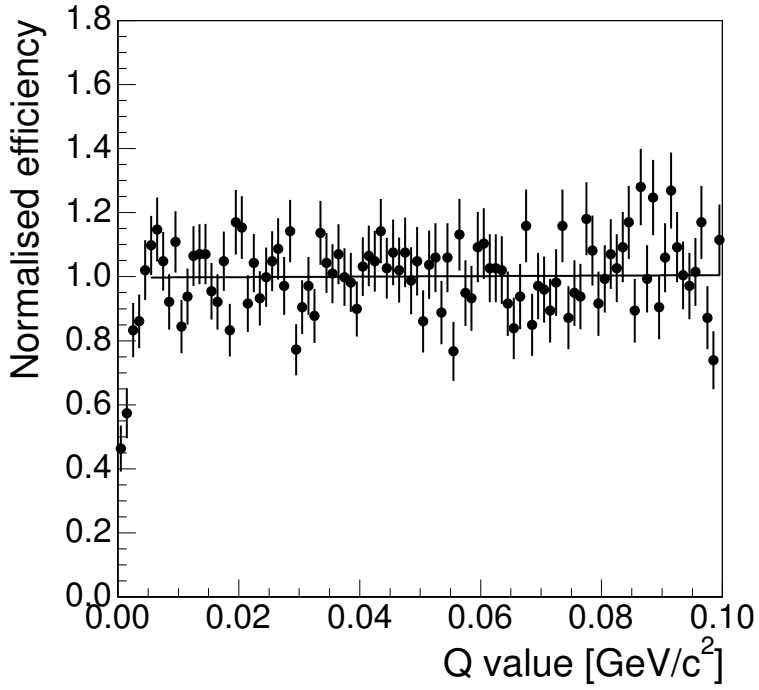

Figure F.2: Relative efficiency for the $B^+ \to \overline{D^0}\pi^+$ channel.

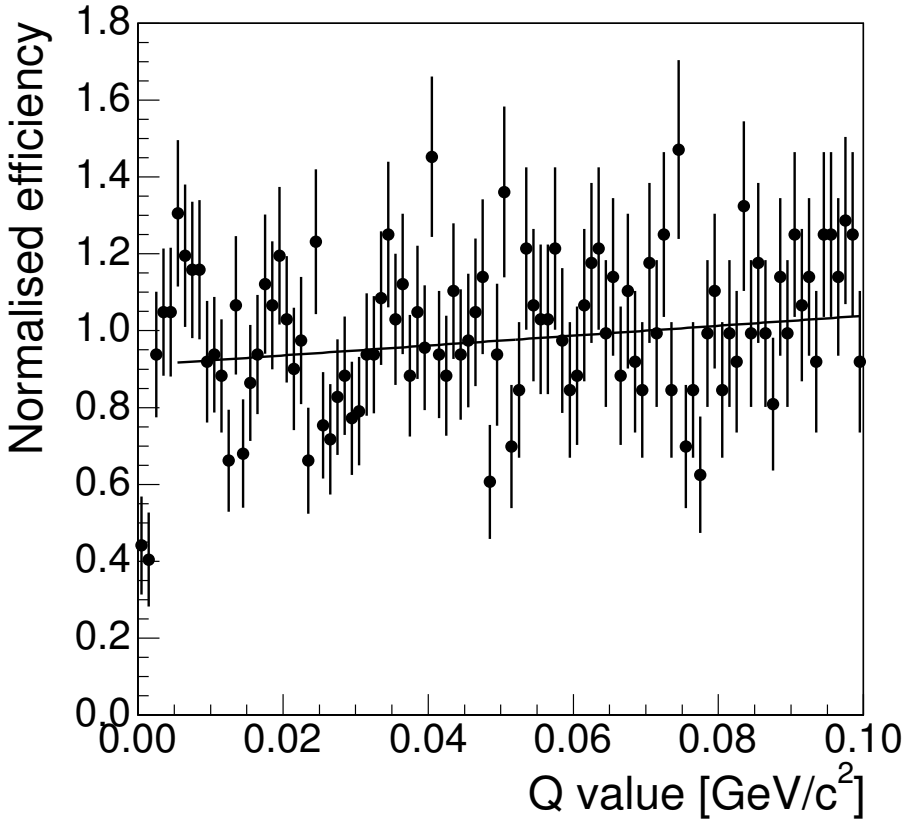

Figure F.3: Relative efficiency for the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel.

Appendix G

Simultaneous Fits

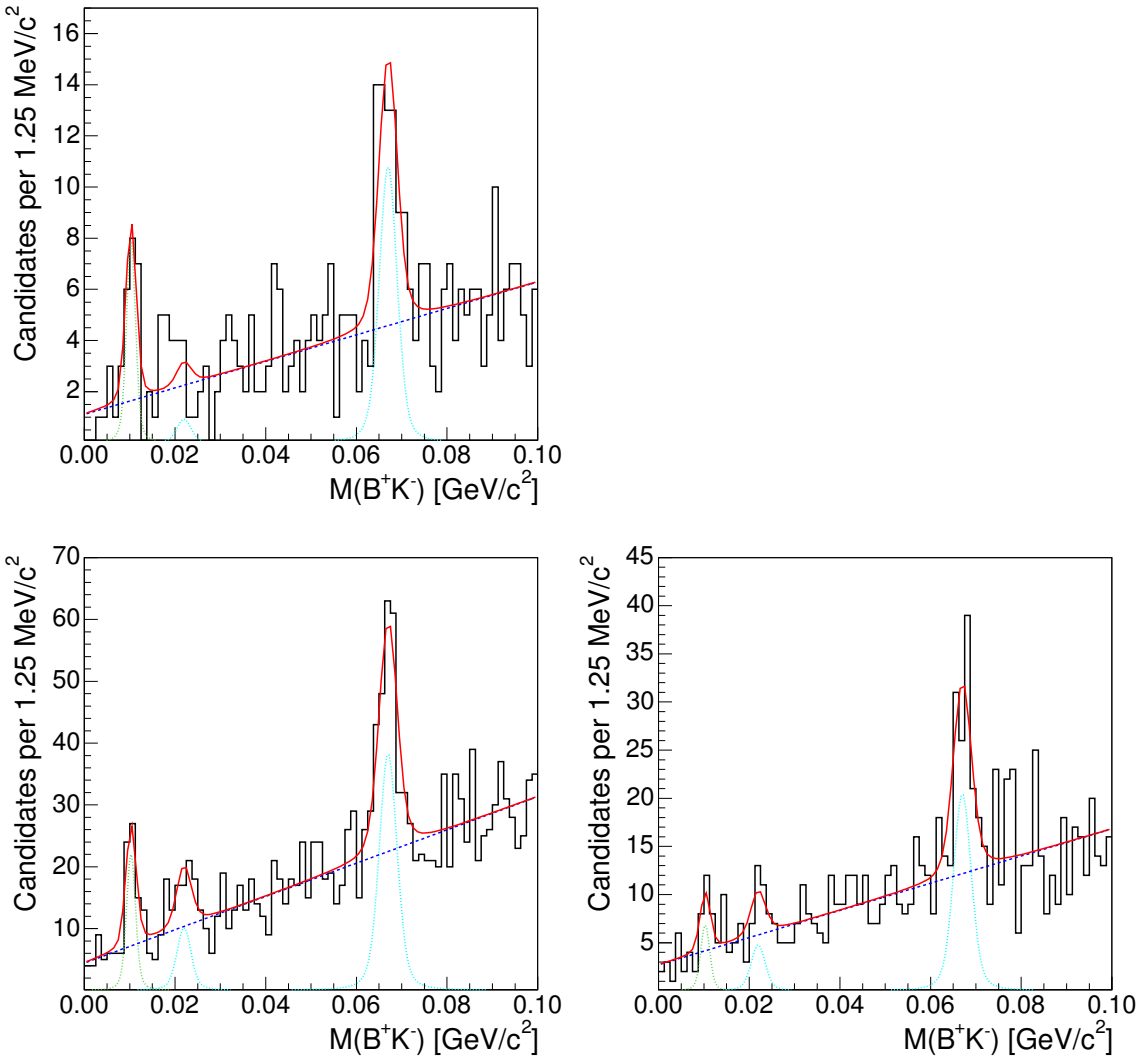

Figure G.1: Simultaneous fits for free natural widths. The upper plot is for the $B^+ \to J/\psi K^+$ channel. The lower left plot is for the $B^+ \to \overline{D^0}\pi^+$ channel, and the lower right plot for the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel.

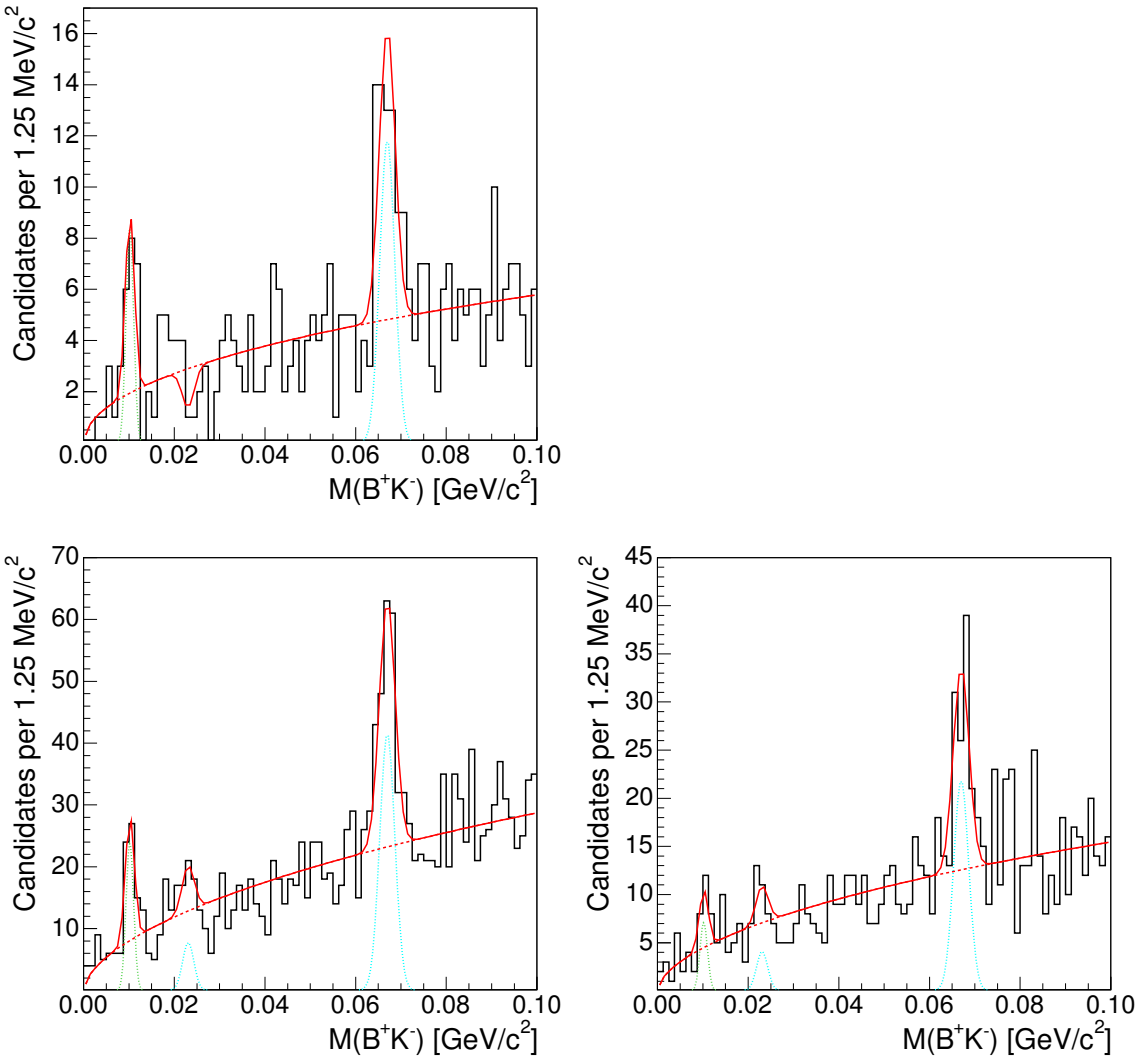

Figure G.2: Simultaneous fits for alternative background description. The upper plot is for the $B^+ \to J/\psi K^+$ channel. The lower left plot is for the $B^+ \to \overline{D^0}\pi^+$ channel, and the lower right plot for the $B^+ \to \overline{D^0}\pi^+\pi^+\pi^-$ channel.

List of Figures

List of Tables

Bibliography

[1] T. Maskawa M. Kobayashi. *Progress of Theoretical Physics*, 49-2:652–657.

[2] S. W. Herb et al. *Phys. Rev.*, D73:252–255, 1977.

[3] B. Grinstein. *Nucl. Phys.*, B339:253, 1990.

[4] H. Georgi. *Phys.Lett.*, B240:447, 1990.

[5] B. Grinstein A. Falk, H. Georgi and M. Wise. *Nucl. Phys.*, B343:1, 1990.

[6] V.O. Galkin D. Ebert and R.N. Faustov. *Phys. Rev.*, D 57:5663, 1998.

[7] A.F. Falk and T. Mehen. *Phys. Rev.*, D 53:231, 1996.

[8] C.T. Hill E.J. Eichten and C. Quigg. *Phys. Rev. Lett.*, 71:4116, 1993.

[9] E.J. Eichten W.A. Bardeen and C.T. Hill. *Phys. Rev.*

[10] F. De Fazio P. Colangelo and R. Ferrandes. *Nucl. Phys. (Proc. Suppl.)*, 163:177, 2007.

[11] S. Godfrey and R. Kokoski. *Phys. Rev.*, D 43:1679, 1991.

[12] A.M. Green et al. *Phys. Rev.*, D69:094505, 2004.

[13] OPAL Collaboration R. Akers et al. *Z. Phys. C*, C66:19, 1995.

[14] M. Moch. *Study of B^*-Meson and of Excited b-Hadron Properties*. PhD thesis, IEKP, Universität Karlsruhe, 2004.

[15] CDF Collaboration T. Aaltonen et. al. *Phys. Rev. Lett.*, 100:082001, 2008.

[16] D0 collaboration V. Abazov et. al. *Phys. Rev. Lett.*, 100:082002, 2008.

[17] R. Blair et al. FERMILAB-PUB-96-390-E.

[18] Darin E. Acosta et al. *Phys. Rev.*, D71:032001, 2005.

[19] Darin E. Acosta et al. *Phys. Rev.*, D71:052003, 2005.

[20] A. Abulencia et al. *J. Phys.*, G34:2457–2544, 2007.

[21] A. Sill. *Nucl. Instrum. Meth.*, A447:1–8, 2000.

[22] Anthony A. Affolder et al. *Nucl. Instrum. Meth.*, A453:84–88, 2000.

[23] Christopher S. Hill. *Nucl. Instrum. Meth.*, A530:1–6, 2004.

[24] Anthony Allen Affolder et al. *Nucl. Instrum. Meth.*, A526:249–299, 2004.

[25] D. Acosta et al. *Nucl. Instrum. Meth.*, A518:605–608, 2004.

[26] G. Ascoli et al. *Nucl. Instrum. Meth.*, A268:33, 1988.

[27] L. Balka et al. *Nucl. Instrum. Meth.*, A267:272, 1988.

[28] S. Bertolucci et al. *Nucl. Instrum. Meth.*, A267:301, 1988.

[29] M. G. Albrow et al. *Nucl. Instrum. Meth.*, A480:524–546, 2002.

[30] Evelyn J. Thomson et al. *IEEE Trans. Nucl. Sci.*, 49:1063–1070, 2002.

[31] Physics Information Technology GmbH. *www.phi-t.de.*

[32] C. Amsler et al. *Phys.Lett.*, B667:1, 2008.

[33] T. Sjostrand. *Comput. Phys. Commun.*, 82:74–90, 1994.

[34] C. Paus. *http://www-cdf.fnal.gov/cdfsim/generators/bgen.html.*

[35] P. Robbe. *http://lhcb-comp.web.cern.ch/lhcb-comp/Simulation/evtgen.htm.*

[36] CERN. *http* : *//wwwasd.web.cern.ch/wwwasd/geant/*, 1993.

[37] MINUIT F. James. *CERN Program Library Long Writeup*, D506, 1998.

[38] F. Rademakers R. Brun. *Nucl. Instrum. Methods*, A 389:81–86, 1996.

[39] CDF Collaboration D. Acosta et. al. *Phys. Rev.*, D68:072004, 2003.

[40] S. Eidelman et. al. CDF DQM Group. *http://www-cdf.fnal.gov/internal/dqm/goodrun/good.html.*

[41] A. Abulencia et al. *Phys. Rev. Lett*, 39:051104, 2006.