CDF/ANAL/BOTTOM/CDFR/9752
Version 1.2
June 1, 2009

Search for $B_s^{**} \to B_s\pi^+\pi^-$ decays

Michael Feindt, Andreas Gessler[1], Michal Kreps[2] Thomas Kuhr[3],

Institut für Experimentelle Kernphysik, Universität Karlsruhe

Abstract

We search neutral orbitally excited B_s^{**0} mesons in decays to $B_s\pi^+\pi^-$ and $B_s^*\pi^+\pi^-$. The B_s is reconstructed in six independent data samples $B_s \to D_s(3)\pi$ with $D_s \to K^*K$, $D_s \to \Phi\pi$ and $D_s \to 3\pi$. From the unbinned maximum likelihood fit to the mass difference $Q = m(B_s^{**}) - m(B_s) - 2\,m(\pi)$ we calculated a Bayesian limit of the branching ratio $\frac{\mathrm{BR}(B_s^{**0}\to B_s\pi\pi)}{\mathrm{BR}(B_s^{**0}\to B_s\pi\pi)+\mathrm{BR}(B_s^{**0}\to BK)}$ for the narrow B_{s1} and B_{s2}^* states.

[1]gessler@ekp.uni-karlsruhe.de
[2]kreps@ekp.uni-karlsruhe.de
[3]tkuhr@ekp.uni-karlsruhe.de

Documentation of the changes from version 1.0 to 1.1 of the note

- Line numbering is switched of since it was enabled mistakenly before.
- The neural network training plots are replaced by the right ones. The plots in the version 1.0 were not from the neural network training used for the selection.
- Added description of the Bayesian limit calculation.

Documentation of the changes from version 1.1 to 1.2 of the note

- Changed significance plot (figure 14) to be in the interval $[0.8, 1)$.
- Made labels more readable in figure 20 through 23.
- Added results from the frequentist limit calculation.

Contents

1 Introduction

The heavy mesons and their excited states play a similar role in quantum chromodynamics as the hydrogen atom in quantum electrodynamics. In this analogy the light quark plays the role of the electron which is bound to a heavy particle, the atomic nucleus or the heavy quark, respectively.

In the recent years a lot of effort has been done to find evidences for orbitally excited states in the D, D_s, B and B_s sector and to measure their masses. This effort includes also analyses made by CDF [1, 2]. Historically, the OPAL Collaboration [3] has first seen one excited B_s state which was later confirmed by the DELPHI Collaboration [4, 5, 6] and recently also observed by DØ.

So far, merely one B_s state has been observed until finally CDF could measure two separate excited states, the B^0_{s1} and B^{*0}_{s2}, in decays of $B^{**0}_s \to B^+K^-$ [7]. This states should in principle also decay to $B^{**0}_s \to B_s\pi^+\pi^-$. Looking for these decays is also intresting since they have a lower threshold than decays to BK. That is, if the broad B^{**0}_s states are below the BK threshold they could be seen in decays of $B^{**0}_s \to B_s\pi^+\pi^-$.

In this note we report the search of orbitally excited B_s mesons in decays to $B_s\pi^+\pi^-$ and $B^*_s\pi^+\pi^-$ with six independent B_s decay channels. The analysis is based on data samples collected with the CDF–II detector from February 2002 till April 2008 corresponding to an integrated luminosity of 2.8 fb^{-1}.

2 Reconstruction

In this analysis a search for the orbitally excited states B^0_{s1} and B^{*0}_{s2} into decays of $B^{**0}_s \to B_s\pi^+\pi^-$ is performed. Six different decay channels of the B_s are used to reconstruct the B^{**0}_s candidates. The different decay modes are

$$\begin{array}{lll} B^{**0}_s \to B_s\pi^+\pi^- & B_s \to D^-_s\pi^+ & D^-_s \to \bar{K}^*K^- \\ & & D^-_s \to \Phi\pi^- \\ & & D^-_s \to 3\pi^\pm \\ & B_s \to D^-_s 3\pi^\pm & D^-_s \to \bar{K}^*K^- \\ & & D^-_s \to \Phi\pi^- \\ & & D^-_s \to 3\pi^\pm \end{array}$$

A widely used variable in this analysis is the Q value defined as the mass difference of a particle and its decay particles

$$Q = m(B^{**0}_s) - m(B_s) - 2 \cdot m_\pi \tag{1}$$

In a previous CDF analysis [7] the masses of the B^0_{s1} and B^{*0}_{s2} states are well measured in decays of $B^{**0}_s \to B^+K^-$. Therefore, the exptected Q values in decays of $B^{**0}_s \to B_s\pi^+\pi^-$ are

$$\begin{aligned} Q(B^0_{s1}) &= 184 \text{ MeV/c}^2 \\ Q(B^{*0}_{s2}) &= 150 \text{ MeV/c}^2 \end{aligned}$$

Actually, the mass of the B^{*0}_{s2} state is slightly higher than the B^0_{s1} mass [7]. But the allowed decay for the B^{*0}_{s2} is $B^{*0}_{s2} \to B^*_s\pi^+\pi^-$ with $B^*_s \to B_s\gamma$ whose photon cannot be measured with the CDF-II detector. Compared to the Q value of the dominant $B^0_{s1} \to B_s\pi^+\pi^-$ decay the Q value of the B^{*0}_{s2} is therefore lowered by the photon energy being on the order of magnitude of $E(\gamma) = 47$ MeV/c^2 [8].

2.1 Data samples

The data samples used in this analysis are collected by the two track trigger. Based on that data private BStntuples were created applying a full vertex fit for the $B_s\pi\pi$ combinations. The BStntuples were skimmed afterwards to obtain small and handy flat ntuples for the analysis. For this samples the currently available data is used but only from runs listed on the good run list provided by the DQM group [9] for b-physics. Table 1 on the following page gives an overview of the used datasets corresponding to a total integrated luminosity of 2.8 fb^{-1}.

2.2 Monte Carlo samples

In this analysis two independent sets of Monte Carlo samples are used. One set comprises samples to train the neural networks for selecting the B_s candidates. The other

Dataset	Integrated luminosity [fb^{-1}]	
	per dataset	sum
0d	0.3	0.3
0h	0.4	0.7
0i	0.6	1.3
0j	1.0	2.3
0k	0.5	2.8

Table 1: Summary of the datasets used in this analysis. The integrated luminosity takes into account the good run list for b-physics.

set comprises samples dedicated to the different B_s^{**0} decays and is needed to train the neural network for selecting the B_s^{**0} candidates. Furthermore, this set is also taken for determining the signal efficiency and the detector resolutions of the different decay channels.

The Monte Carlo samples of the B_s decays are generated using the BGen package [10] except of the decays $B_s \to D_s^- \pi^+$ which are simulated by Pythia [11]. The reason for using different generator package is purely historical and has no physical motivation since only quantities from the signal B mesons are taken into account. A more detailed description of the B_s Monte Carlo samples is given in reference [12] where also the neural networks for the B_s selection are explained.

The B_s^{**0} Monte Carlo samples are exclusively generated for this analysis on the basis of the BGen event generator. For the generation a custom module was attached to the generator in order to have the B_s^{**0} mass flatly distributed over the whole mass range observed in this analysis. After the simulation the generated events are decayed using the EvtGen [13] package followed by a full detector and trigger simulation. Afterwards, the events are reconstructed by applying the standard reconstruction software including the BottomMods [14] package with the same `tcl` file as used for data reconstruction. The Monte Carlo samples contain decays into $B_s \pi^+ \pi^-$ as well as into $B_s^* \pi^+ \pi^-$. The complete decay tables are given in the appendix in section A on page 39.

Monte Carlo reweighting

The low statistics of the data samples in each B_s^{**0} decay mode makes it impossible to train independent neural networks for selecting the B_s^{**0} candidates on each decay channel. Therefore, both the data samples and the Monte Carlo samples are put together in each case and one single neural network is trained to select the B_s^{**0}. In order to avoid biasing the neural network training the Monte Carlo events are reweighting taking into account two physical conditions. Firstly, the Monte Carlo events are reweighted to the same Q value distribution as in data. This is necessary because of the flatly generated B_s^{**0} mass. Secondly, the right mixture of decay channels is required in the combined Monte Carlo sample. Hence, the individual weight w_i being applied to every Monte

Carlo event i having the Q value Q_i is

$$w_i(Q_i) = \frac{f_{\text{data}}(Q_i)}{f_{\text{MC}}(Q_i)} \cdot \frac{\text{BR}_{\text{PDG}}}{\text{BR}_{\text{MC}}} \tag{2}$$

Here, f_{data} and f_{MC} are functions describing the Q value distributions in data and in Monte Carlo, respectively. Their ratio corrects for the Q value distributions. These functions are fourth order polynomials whose parameters are obtained from different fits to the Q value in data and in the Monte Carlo samples. For each decay channel a distinct set of parameters is used. The second fraction in equation (2) corrects for the right channel mixture. Thereby BR_{PDG} is the world average branching ratio [8] of the decay in question and BR_{MC} is the associated branching ratio implemented in the decay tables which were used to generate the Monte Carlo samples. The decay tables are listed in the appendix in section A on page 39.

2.3 Candidate reconstruction

For this analysis private BStntuples are built containing the B_s^{**0} decays in question. In order to reconstruct the B_s^{**0} candidates tracks from the pion block of the BStntuples are combined with B_s candidates. The invariant mass of the so–formed B_s^{**0} meson is calculated by the four–momenta of the B_s candidate and the pion tracks. To keep the background level as low as possible on this stage a full vertex fit to the $B_s\pi\pi$ combinations is performed on the cost of a more time consuming reconstruction procedure.

3 Candidate selection

The candidate selection is based on a chain of two neural networks. The first neural network in this chain performs the selection of the B_s meson. For each B_s decay channel a distinct neural network is trained and used. These B_s networks have a high efficiency in separating signal and background and a soft cut is applied on their outputs to reject obvious background and keep most of the signal for the optimisation of the selection of B_s^{**0}. The second neural network is trained for selecting the B_s^{**0} candidates. For all six B_s^{**0} decay mode one common neural network is trained which reuses the output of the prior B_s networks among other input variables. Thus, the B_s^{**0} network acts on events preselected by the B_s networks.

All neural networks used in this analysis are taken from the NeuroBayes [15] package.

3.1 B_s selection

The selection of the B_s candidates is the same as already described in [12] and used in previous analyses. For completeness the important parts are copied over here.

The first step of the candidate selections is done be applying loose preselection cuts to the data samples. They cut away a significant part of background events with a marginal loss of signal events. Table 2 lists the precuts applied to each B_s decay sample.

preselection	$B_s \to D_s\pi$ modes			$B_s \to D_s3\pi$ modes		
cuts on	$D_s \to \phi\pi$	$D_s \to K^*K$	$D_s \to 3\pi$	$D_s \to \phi\pi$	$D_s \to K^*K$	$D_s \to 3\pi$
$\chi^2_{r\phi}(B_s)$	< 20	< 20	< 20	< 20	< 20	< 20
$\chi^2_{r\phi}(D_s)$	< 20	< 20	< 20	< 20	< 20	< 20
$L_{xy}/\sigma_{L_{xy}}(B_s)$	> 2	> 2	> 2	> 2	> 2	> 6
$\sigma_{L_{xy}}(B_s)$ [cm]	-	-	-	< 0.015	< 0.015	-
$L_{xy}/\sigma_{L_{xy}}(D_s)$	> 2	> 2	-	> 2	> 2	> 6
$\lvert d_0(B_s)\rvert$ [cm]	< 0.02	< 0.02	< 0.009	< 0.01	< 0.012	< 0.015
$p_t(B_s)$ [GeV]	> 5.5	> 5.5	> 5.5	> 5.5	> 5.0	> 4.0
$p_t(\pi_{B_s})$ [GeV]	> 1.0	-	> 0.4	-	-	-
$L_{xy}(B_s \leftarrow D_s)$ [cm]	-	-	$[-4, 4]$	-	> -0.05	> -0.1
$\lvert\Delta m(K\pi\pi, D^-)\rvert$ [GeV]	-	> 0.024	-	-	> 0.024	-
$q(K_{K^*}) \cdot q(K_{D_s})$	-	< 0	-	-	< 0	-
inv. mass $m_{3\pi_{B_s}}$[GeV]	-	-	-	-	-	< 2.5

Table 2: Summary of preselection cuts for all $B_s \to D_s(3)\pi$ decay modes.

For the final selection of the B_s candidates neural networks are trained for each B_s decay channel independently to separated the B_s candidates from background events. The setup of the neural network training and the candidate selection afterwards is done analogically for each decay. The neural networks are trained on Monte Carlo events as signal pattern and real data events as background pattern. The Monte Carlo

events for the decay $B_s \to D_s\pi$, $D_s \to \phi\pi$ are generated with PYTHIA, whereas the Monte Carlo events for the other B_s decays are generated with BGEN. All Monte Carlo samples are run through the complete CDF-II detector and trigger simulation. The data events for the background pattern are taken from the upper mass sideband in an arbitrary range within 5.55 GeV/c^2 and 6.55 GeV/c^2. Of course, it would be more satisfying to have the lower sideband as well in the background pattern but since it contains partially reconstructed B_s mesons it cannot be used for the network training. The particular sideband interval for each training is listed in table 3.

	$B_s \to D_s\pi$ modes			$B_s \to D_s3\pi$ modes		
	$D_s \to \phi\pi$	$D_s \to K^*K$	$D_s \to 3\pi$	$D_s \to \phi\pi$	$D_s \to K^*K$	$D_s \to 3\pi$
mass range [GeV/c^2]	$5.6 - 5.9$	$5.55 - 5.85$	$5.6 - 6.55$	$5.6 - 6.55$	$5.6 - 5.8$	$5.6 - 5.9$

Table 3: Mass ranges for the upper sideband used for the network trainings.

The variables put into the networks are supposed to contain reasonable information for discriminating background and signal. It is a rule of thumb to use as low as possible variables in the network. Table 4 on page 11 summarises the input variables used for the different B_s neural networks.

Figure 1 to 6 show the invariant mass distribution of the B_s of the different decay modes selected by the according neural network. The data shown in these plots comprises the official CDF datasets `xbhd0d`, `xbhd0h` and the first half of `xbhd0i` corresponding altogether to an integrated luminosity of $1\,\mathrm{fb}^{-1}$. In order to plot the mass distribution the selection cut on the network output was chosen in such a way that it maximises the significance σ defined as

$$\sigma = \frac{N_S\left(n_{\mathrm{cut}}\right)}{\sqrt{N_S\left(n_{\mathrm{cut}}\right) + N_B\left(n_{\mathrm{cut}}\right)}} \tag{3}$$

where N_S is the number of signal events and N_B is the number of background events at a given cut n_{cut} on the neural network output. The significance is calculated by fitting the mass of the B_s at a given cut on the neural net and the number of signal and background events are extracted in each case from the integral of either the background or signal component of the fit function over the mass range from 5.32 GeV/c^2 to 5.42 GeV/c^2.

As already mentioned above, the output of the B_s network is directly used as an input variable to the B_s^{**0} network. There is no hard cut applied on the output of the B_s network for the final B_s^{**0} selection except of a soft precut to prevent the data samples from being excessively large.

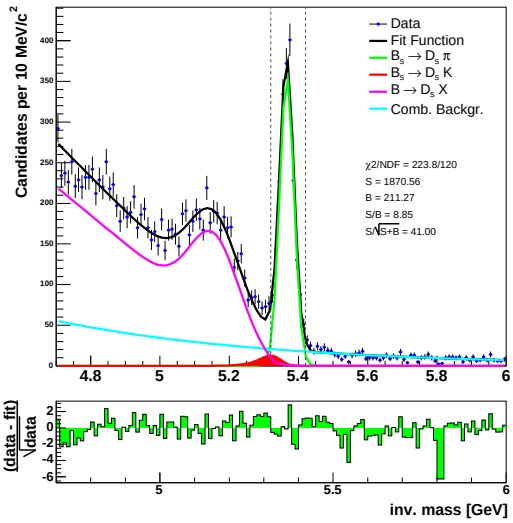

Figure 1: Invariant mass spectrum for the decay $B_s \to D_s\pi$, $D_s \to \phi\pi$ and the binned fit result for a cut on the neural network output, $n_{\mathrm{NN}} > 0.68$.

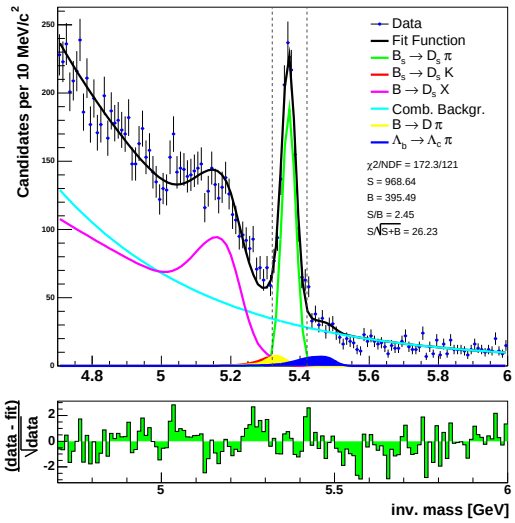

Figure 2: Invariant mass spectrum for the decay $B_s \to D_s\pi$, $D_s \to K^*K$ and the binned fit result for a cut on the neural network output, $n_{\mathrm{NN}} > 0.92$.

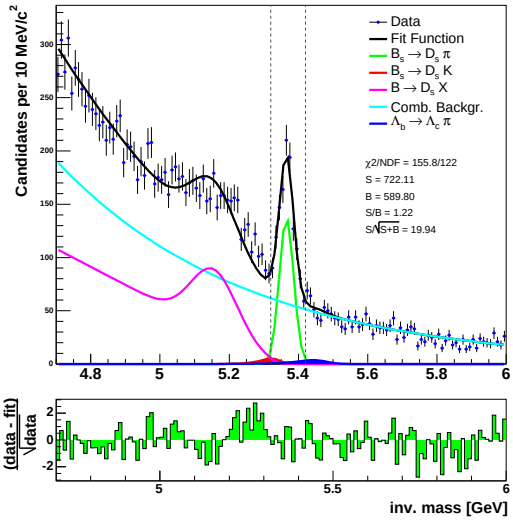

Figure 3: Invariant mass spectrum for the decay $B_s \to D_s\pi$, $D_s \to 3\pi$ and the binned fit result for a cut on the neural network output, $n_{\mathrm{NN}} > 0.48$.

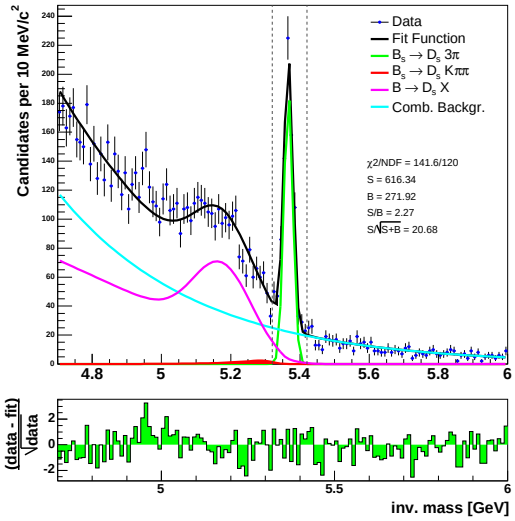

Figure 4: Invariant mass spectrum for the decay $B_s \to D_s\,3\pi$, $D_s \to \phi\pi$ and the binned fit result for a cut on the neural network output, $n_{\mathrm{NN}} > 0.90$.

	$B_s \to D_s\pi$ modes			$B_s \to D_s3\pi$ modes		
variable name	$D_s \to \phi\pi$	$D_s \to K^*K$	$D_s \to 3\pi$	$D_s \to \phi\pi$	$D_s \to K^*K$	$D_s \to 3\pi$
$L_{xy}/\sigma_{L_{xy}}(B_s)$	√	√	√	√	√	√
$\sigma_{L_{xy}}(B_s)$				√	√	√
$L_{xy}/\sigma_{L_{xy}}(D_s)$	√	√		√		
$L_{xy}(B_s \leftarrow D_s)$	√	√	√	√	√	√
$\chi^2_{3D}(B_s)$	√	√			√	√
$\chi^2_{3D}(D_s)$	√	√				√
$\chi^2_{r\phi}(B_s)$		√	√	√	√	√
$\chi^2_{r\phi}(D_s)$			√	√	√	√
$p_t(B_s)$			√	√		
$p_t(D_s)$		√			√	√
$p_t(\pi_{B_s})$	√	√	√			
$p_t(\pi^1_{B_s})$				√	√	
$p_t(\pi^2_{B_s})$				√		
$p_t(\pi^3_{B_s})$				√		
$p_t(K^1_\phi)$				√		
$p_t(K^2_\phi)$				√		
$\min p_t$	√	√			√	√
$\lvert d_0(B_s)\rvert$	√	√	√	√	√	√
$\lvert d_0(D_s)\rvert$	√				√	√
lts. $d_0/\sigma_{d_0}(K^1_\phi)$	√					
lts. $d_0/\sigma_{d_0}(\pi_{K^*})$		√				
lts. $d_0/\sigma_{d_0}(\pi_{B_s})$			√			
lts. $d_0/\sigma_{d_0}(\pi^n_{D_s})$			√			
$\min d_0/\sigma_{d_0}$		√			√	√
min lts. d_0/σ_{d0}					√	√
$\min d_0(\pi^1_{B_s}, \pi^2_{B_s}, \pi^3_{B_s})$				√		
$\max d_0(\pi^1_{B_s}, \pi^2_{B_s}, \pi^3_{B_s})$				√		
helicity angle K^1_ϕ	√					
helicity angle K_{K^*}		√			√	
CMS angle π_{B_s}			√			
$\theta(\vec{n}_{(\pi^1_{B_s},\pi^2_{B_s})}, \vec{p}(3\pi))$				√	√	√
$m_{\phi\to KK}$	√			√		
$\lvert m_{K^*} - m_{K^*(PDG)}\rvert$		√			√	
inv. mass $m^2_{\pi^1_{D_s}\pi^2_{D_s}}$			√			
inv. mass $m^2_{\pi^1_{D_s}\pi^3_{D_s}}$			√			
inv. mass $m^2_{\pi^2_{D_s}\pi^3_{D_s}}$			√			
inv. mass $m_{\pi^1_{D_s}\pi^2_{D_s}}$						√
inv. mass $m_{\pi^1_{D_s}\pi^3_{D_s}}$						√
inv. mass $m_{\pi^2_{D_s}\pi^3_{D_s}}$						√
inv. mass $m_{\pi^1_{B_s}\pi^2_{B_s}}$				√	√	√
inv. mass $m_{\pi^1_{B_s}\pi^3_{B_s}}$						√
inv. mass $m_{\pi^2_{B_s}\pi^3_{B_s}}$						√
$\min(m^{12}_{\pi_{B_s}}, m^{23}_{\pi_{B_s}})$				√	√	√
$\max(m^{12}_{\pi_{B_s}}, m^{23}_{\pi_{B_s}})$				√	√	√
$\min(m^{12}_{\pi_{D_s}}, m^{23}_{\pi_{D_s}})$						√
$\max(m^{12}_{\pi_{D_s}}, m^{23}_{\pi_{D_s}})$						√
inv. mass $m_{3\pi_{B_s}}$					√	√
m_{diff}				√		
$\Delta R(D_s, \pi_{B_s})$			√			

Table 4: Variables used in the B_s neural network trainings.

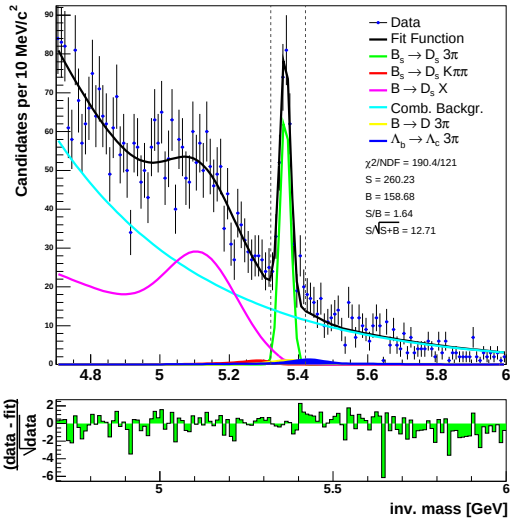

Figure 5: Invariant mass spectrum for the decay $B_s \to D_s 3\pi$, $D_s \to K^*K$ and the binned fit result for a cut on the neural network output, $n_{\mathrm{NN}} > 0.99$.

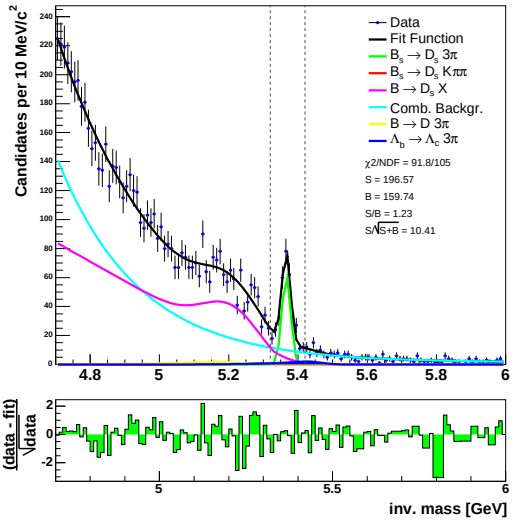

Figure 6: Invariant mass spectrum for the decay $B_s \to D_s\,3\pi$, $D_s \to 3\pi$ and the binned fit result for a cut on the neural network output, $n_{\mathrm{NN}} > 0.99$.

3.2 B_s^{**} selection

The selection of the B_s^{**0} candidates starts by applying soft precuts cutting away a large number of background events but only a marginal part of the signal events. These cuts are:

- The output of the neural network for selecting the B_s greater than 0 (78%).
- The Q value less than 0.5 GeV/c^2 (83%).
- The absolute value of the impact parameter of the B_s smaller than 0.008 cm (99%).
- The significance of the decay length of the B_s with respect to the primary vertex greater than 4 (98%).
- The significance of the decay length of the D_s with respect to the primary vertex greater than 2 (99%).
- The invariant mass of the B_s in the range from 5.2 GeV/c^2 to 5.5 GeV/c^2 (99%).
- Each pion from the B_s^{**0} decay must have hits in the silicon detectors (89%).

The number in parentheses are the distinct efficiencies of precuts showing that the cuts affect only a small part of the signal events except of the cut on the output of the B_s networks which is however necessary to reduce the background level to a reasonable extent.

The main part of the B_s^{**0} candidate selection is done by a neural network. For the neural network training Monte Carlo events are used as signal pattern and data in the Q value range from 0 GeV/c^2 to 0.25 GeV/c^2 are used as background pattern. As figure 7 on the next page shows, the data sample is still dominated by background events and can be well used for the neural network training.

To avoid biasing the neural network training the Monte Carlo events are reweighted to the same Q value distribution as in data. Furthermore, the Monte Carlo events are reweighted in order to correct for the mixture of the different decay channels in the combined Monte Carlo sample. The details of the Monte Carlo reweighting are described in section 2.2 on page 6. The variables which are used in the neural network are:

1. The absolute value of the impact parameter of the B_s^{**0} ($|d_0(B_s^{**0})|$).
2. The χ^2 of the two-dimensional vertex fit of the B_s^{**0} ($\chi^2_{r\phi}(B_s^{**0})$).
3. The significance of the transverse decay length of the B_s^{**0} ($\sigma_{L_{xy}}(B_s^{**0})$).
4. The transverse decay length of the B_s ($L_{xy}(B_s)$).
5. The significance of the transverse decay length of the B_s ($\sigma_{L_{xy}}(B_s)$).

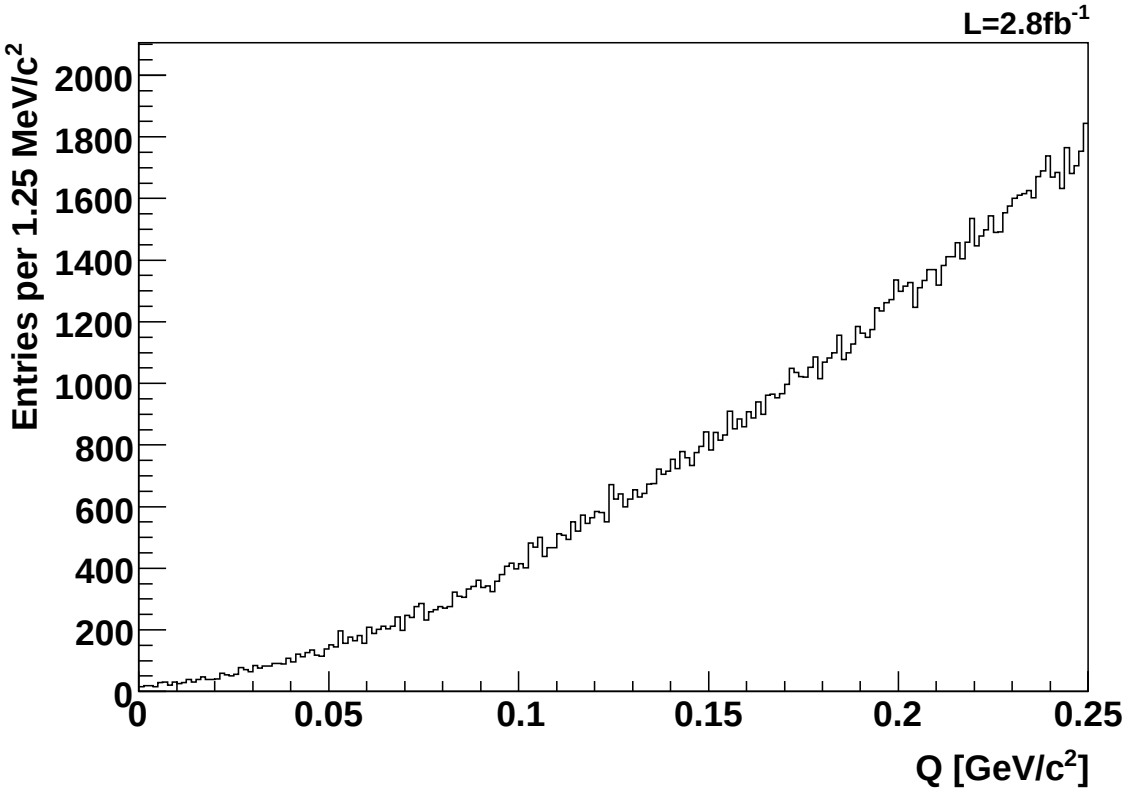

Figure 7: Q value distribution from the data sample being used for the neural network training as background pattern. After having applied the precuts a large ammount of background is still left without an evidence for any signal.

6. The invariant mass of the B_s ($m(B_s)$).
7. The transverse momentum of the D_s ($p_t(D_s)$).
8. The pseudo rapidity of the D_s ($\eta(D_s)$). [pruned]
9. The χ^2 of the two-dimensional vertex fit of the D_s ($\chi^2_{r\phi}(D_s)$). [pruned]
10. The fit probability of the D_s vertex fit ($P(D_s)$). [pruned]
11. The transverse decay length of the D_s with respect to the primary vertex ($L_{xy}(D_s)$).
12. The absolute value of the impact parameter of the D_s ($|d_0(D_s)|$).
13. The significance of the transverse decay length of the D_s with respect to the primary vertex ($\sigma_{L_{xy}}(D_s)$).
14. The transverse decay length of the D_s with respect to decay vertex of the B_s ($\ell_{xy}(D_s)$).

15. The cosine of the angle of the D_s from the B_s decay in the CMS of the B_s relatively to the momentum of the B_s in the laboratory frame ($\theta^*(D_s)$).

16. The life-time signed significance of the impact parameter of the pion from the decay $B_s \to D_s^- \pi^+$. For other decays it is set to -999 ($d_0^{lts}/\sigma_{d_0}(\pi_{B_s})$). [pruned]

17. The absolute value of the impact parameter of the pion from the decay $B_s \to D_s^- \pi^+$. For other decays it is set to -999 ($|d_0(\pi_{B_s})|$). [pruned]

18. The absolute value of the impact parameter of the first pion from the decay $B_s \to D_s^- 3\pi^\pm$. For other decays it is set to -999 ($\left|d_0(\pi_{B_s}^{(1)})\right|$). [pruned]

19. The absolute value of the impact parameter of the second pion from the decay $B_s \to D_s^- 3\pi^\pm$. For other decays it is set to -999 ($\left|d_0(\pi_{B_s}^{(2)})\right|$). [pruned]

20. The absolute value of the impact parameter of the third pion from the decay $B_s \to D_s^- 3\pi^\pm$. For other decays it is set to -999 ($\left|d_0(\pi_{B_s}^{(3)})\right|$). [pruned]

21. The minimum of the transverse momenta of the pions from the decay $B_s^{**0} \to B_s\pi^+\pi^-$ ($\min(p_t)$).

22. The maximum of the transverse momenta of the pions from the decay $B_s^{**0} \to B_s\pi^+\pi^-$ ($\max(p_t)$).

23. The minimum of the absolute value of the impact parameters of the pions from the decay $B_s^{**0} \to B_s\pi^+\pi^-$ ($\min(|d_0|)$).

24. The maximum of the absolute value of the impact parameters of the pions from the decay $B_s^{**0} \to B_s\pi^+\pi^-$ ($\max(|d_0|)$).

25. The ordinal number of the decay channel (n_{Dcy}).

26. The output of the neural network for selecting the B_s in the decay $B_s \to D_s^-\pi^+$, $D_s^- \to \bar{K}^*K^-$. For other decays it is set to -999 ($n_{\text{NN}}(B_s \to D_s^-\pi^+, D_s^- \to \bar{K}^*K^-)$).

27. The output of the neural network for selecting the B_s in the decay $B_s \to D_s^-\pi^+$, $D_s^- \to \Phi\pi^-$. For other decays it is set to -999 ($n_{\text{NN}}(B_s \to D_s^-\pi^+, D_s^- \to \Phi\pi^-)$).

28. The output of the neural network for selecting the B_s in the decay $B_s \to D_s^-\pi^+$, $D_s^- \to 3\pi^\pm$. For other decays it is set to -999 ($n_{\text{NN}}(B_s \to D_s^-\pi^+, D_s^- \to 3\pi^\pm)$).

29. The output of the neural network for selecting the B_s in the decay $B_s \to D_s^- 3\pi^\pm$, $D_s^- \to \bar{K}^*K^-$. For other decays it is set to -999 ($n_{\text{NN}}(B_s \to D_s^- 3\pi^\pm, D_s^- \to \bar{K}^*K^-)$).

30. The output of the neural network for selecting the B_s in the decay $B_s \to D_s^- 3\pi^\pm$, $D_s^- \to \Phi\pi^-$. For other decays it is set to -999 ($n_{\text{NN}}(B_s \to D_s^- 3\pi^\pm, D_s^- \to \Phi\pi^-)$).

31. The output of the neural network for selecting the B_s in the decay $B_s \to D_s^- 3\pi^\pm$, $D_s^- \to 3\pi^\pm$. For other decays it is set to -999 ($n_{\text{NN}}(B_s \to D_s^- 3\pi^\pm, D_s^- \to 3\pi^\pm)$).

32. The Q value (Q). [pruned]

Actually, not all of the above listed variables are kept in the neural network. The preprocessing algorithm of NeuroBayes keeps only the most significant variables. If the variable's significance is below 3σ it is pruned and not used in the neural network at all. The Q value is put into the network as crosschek to ensure the correctness of the Monte Carlo reweighting as the neural network cannot learn anything from the Q value distribution. Table 5 on the next page gives a list of the significances of the input variables.

The quality plots obtained by the neural network training can be evaluated in order to prove the success of the training. The network output for signal and background in figure 8 is well-separated and hence, the network is capable of distinguishing between signal and background. Figure 9 on page 18 demonstrates that the purity of the neural network is a linear function of the network output. Both plots indicate a successfully trained network. The last training plot seen in figure 10 on page 19 presents the correlation matrix of the input variables.

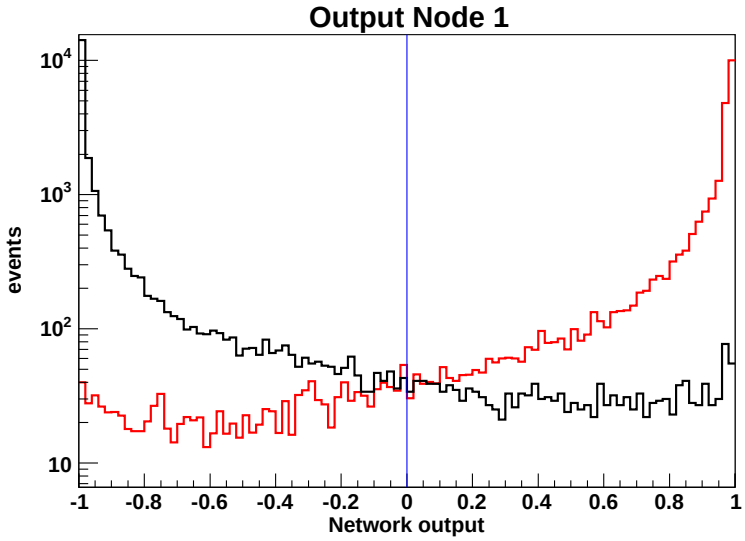

Figure 8: The distribution of the neural network output for signal (red) and background (black) events in the decay $B_s^{**0} \to B_s\pi^+\pi^-$ using the training sample.

Position	Variable	Significance [σ]
1	$m(B_s)$	146.9
2	$n_{\mathrm{NN}}(B_s \to D_s^- 3\pi^\pm, D_s^- \to 3\pi^\pm)$	40.5
3	$n_{\mathrm{NN}}(B_s \to D_s^- 3\pi^\pm, D_s^- \to \bar{K}^* K^-)$	44.9
4	$\max(\lvert d_0 \rvert)$	37.3
5	$n_{\mathrm{NN}}(B_s \to D_s^- 3\pi^\pm, D_s^- \to \Phi\pi^-)$	26.5
6	$n_{\mathrm{NN}}(B_s \to D_s^- \pi^+, D_s^- \to \bar{K}^* K^-)$	24.9
7	n_{Dcy}	25.3
8	$\sigma_{L_{xy}}(D_s)$	18.6
9	$p_t(D_s)$	15.1
10	$\sigma_{L_{xy}}(B_s^{**0})$	12.1
11	$\theta^*(D_s)$	9.9
12	$\sigma_{L_{xy}}(B_s)$	3.5
13	$L_{xy}(B_s)$	10.0
14	$n_{\mathrm{NN}}(B_s \to D_s^- \pi^+, D_s^- \to 3\pi^\pm)$	9.6
15	$\min(p_t)$	5.7
16	$\lvert d_0(D_s) \rvert$	5.1
17	$\min(\lvert d_0 \rvert)$	4.8
18	$\ell_{xy}(D_s)$	3.8
19	$L_{xy}(D_s)$	4.4
20	$\left\lvert d_0(B_s^{**0}) \right\rvert$	4.1
21	$n_{\mathrm{NN}}(B_s \to D_s^- \pi^+, D_s^- \to \Phi\pi^-)$	4.0
22	$\chi^2_{r\phi}(B_s^{**0})$	3.5
23	$\max(p_t)$	3.2
24	$\chi^2_{r\phi}(D_s)$	2.5
25	$P(D_s)$	2.9
26	$\eta(D_s)$	2.9
27	$\left\lvert d_0(\pi^{(3)}_{B_s}) \right\rvert$	2.3
28	$d_0^{lts}/\sigma_{d_0}(\pi_{B_s})$	2.4
29	$\lvert d_0(\pi_{B_s}) \rvert$	1.9
30	$\left\lvert d_0(\pi^{(2)}_{B_s}) \right\rvert$	1.8
31	$\left\lvert d_0(\pi^{(1)}_{B_s}) \right\rvert$	1.5
32	Q	0.6

Table 5: List of variables used in the neural network trained for selecting the B_s^{**0}. The variables are sorted by their significance for the neural network. Variables having a significance less than 3σ are pruned by the preprocessing algorithms of the neural network and they are not used in the neural net at all.

3.3 Cut optimisation

The final selection of the B_s^{**0} is done by cutting on the output of the neural network trained for selecting the B_s^{**0} candidates. The best cut on the network output is supposed to maximise the significance being a function of the cut on the network

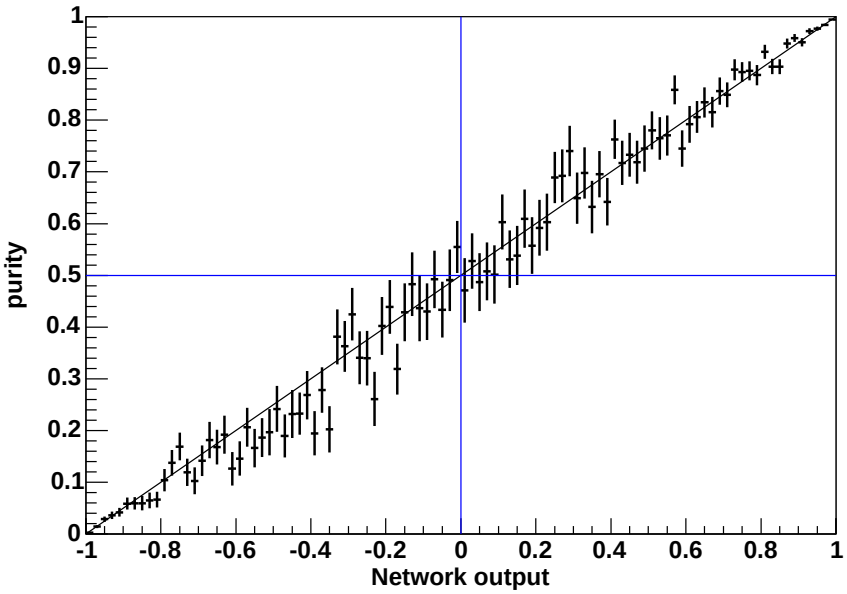

Figure 9: The purity as a function of the neural network output in the decay $B_s^{**0} \to B_s\pi^+\pi^-$ using the training sample.

output. The significance σ_{NN} is defined by

$$\sigma_{\mathrm{NN}} = \frac{N_{\mathrm{MC}}^{(\mathrm{cand})}\left(n_{\mathrm{NN}} > n_{\mathrm{cut}}\right)}{1.5 + \sqrt{N_{\mathrm{data}}^{(\mathrm{cand})}\left(n_{\mathrm{NN}} > n_{\mathrm{cut}}\right)}} \tag{4}$$

where $N_{\mathrm{MC}}^{(\mathrm{cand})}$ is the number of Monte Carlo events selected at a given cut n_{cut} on the network output n_{NN} and $N_{\mathrm{data}}^{(\mathrm{cand})}$ is the number of data events at the same cut on the network output. The number of Monte Carlo events is regarded to be proportional to the number of signal events whereas the number of data events is the sum of real background events and real signal events independently of the actual ratio. Since merely a low signal is expected, this formula for calculating the significance was chosen. The summand of 1.5 arises from the fact that the formula is optimised towards a 3σ observation [16].

In order to calculate the significance the Monte Carlo and data events are counted at a given cut on the network output in a Q value range where the signal peaks are expected. The search window in Q was therefore chosen from 0.134 GeV/c^2 to 0.2 GeV/c^2. Figure 13 and figure 14 on page 20 show the significance over the cut on the network output. Based on these plots the selection cut on the neural network output was set to 0.95. The Q value distribution of the B_s^{**0} candidates selected by this cut are given in figure 15 on page 21.

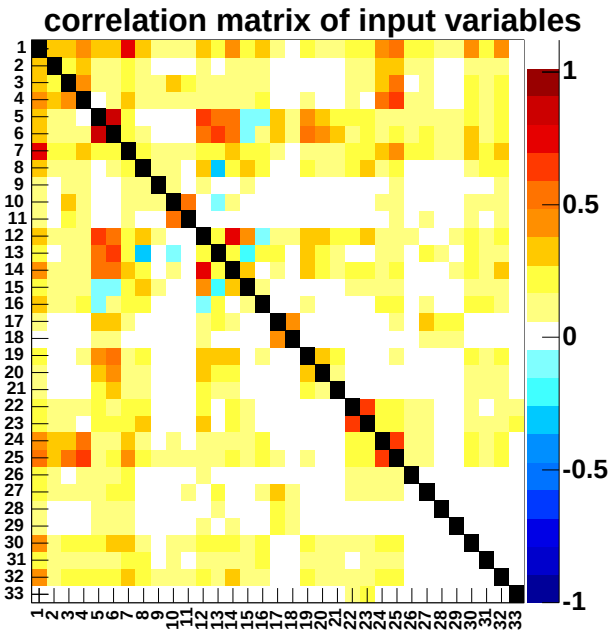

Figure 10: The correlation matrix of the input variables of the neural network for selecting the B_s^{**0}. The index i of each column or row in the correlation matrix corresponds to the variable $(i-1)$ in the variable list. Index 1 denotes the target variable.

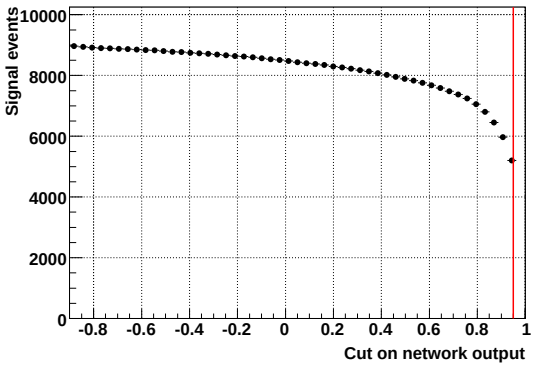

Figure 11: The number of signal events over the cut on the network output. The red line denotes where the selection cut is put.

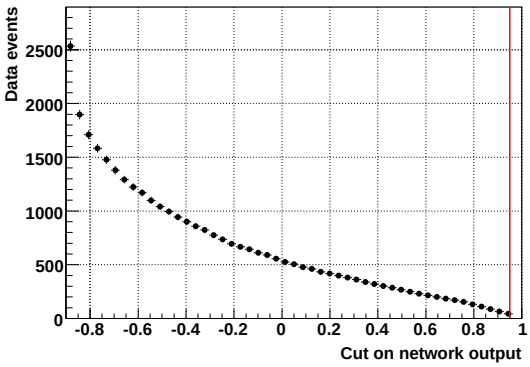

Figure 12: The number of data events over the cut on the network output. The red line denotes where the selection cut is put

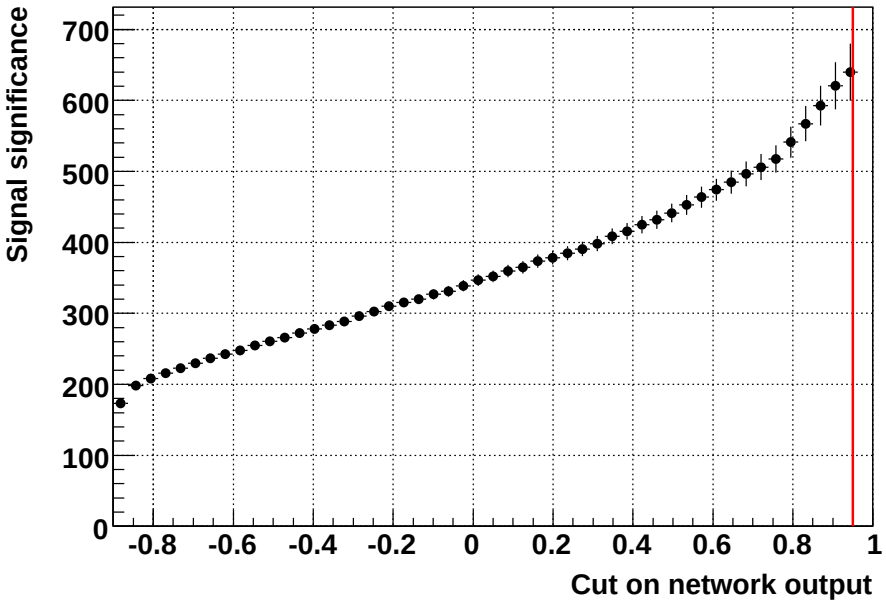

Figure 13: The significance σ_{NN} over the cut on the network output. The red line denotes where the selection cut is put.

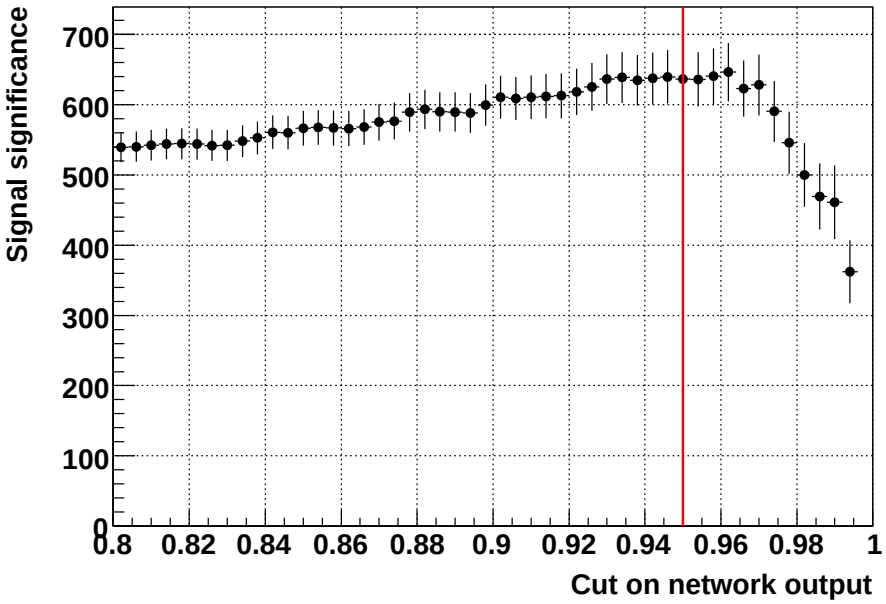

Figure 14: A more detailed plot of the significance σ_{NN} over the cut on the network output for cut values larger than 0.8. The red line denotes where the selection cut is put

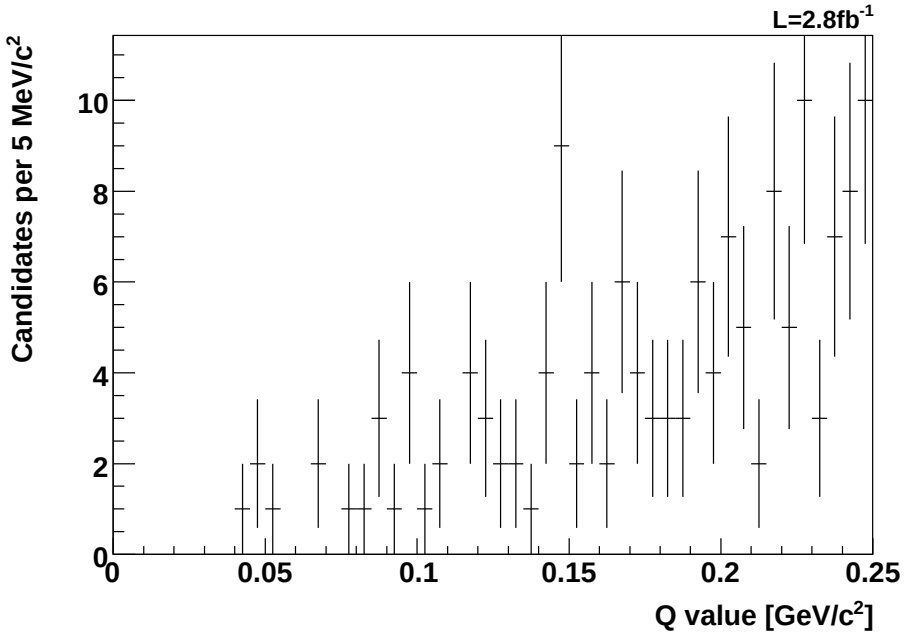

Figure 15: The Q value distribution of the B_s^{**0} candidates selected by a cut on the output of the neural network at 0.95.

4 Reconstruction efficiencies

Due to the lack of any significant signal in the used data samples a Baysian limit is set for the branching ratios of the B^0_{s1} and B^{*0}_{s2} states. The branching ratios are calculated based on a reference decay since the absolute production rates of the B_s mesons are unknown. Up to now, the B^{**0}_s mesons have only been measured in decays of $B^{**0}_s \to B^+K^-$. Therefore, the decay $B^{**0}_s \to B^+K^-$ with $B^+ \to \bar{D}^0\pi^+$ was chosen as reference decay for measuring the branching ratios of the B^0_{s1} and B^{*0}_{s2} states in decays of $B^{**0}_s \to B_s\pi^+\pi^-$.

4.1 Signal efficiency

The signal efficiency is determined by studying the same Monte Carlo samples used in the training of the neural network for selecting the B^{**0}_s candidates. The efficiency ϵ is defined as the ratio of generated ($N^{(\text{generated})}_{\text{MC}}$) and reconstructed ($N^{(\text{reconstructed})}_{\text{MC}}$) Monte Carlo events:

$$\epsilon = \frac{N^{(\text{reconstructed})}_{\text{MC}}}{N^{(\text{generated})}_{\text{MC}}} \tag{5}$$

The number of reconstructed Monte Carlo events is counted after applying the same selection procedure as for the data samples taking into account the proper branching ratios from the decay tables of the Monte Carlo generation. Table 6 and table 7 on the facing page shows the signal efficiencies of the different B_s decay channels. Due to the lack of any significant signal in real data a Baysian limit on the branching ratio is calculated with respected to a reference channel. The signal efficiencies of the reference decay $B^{**0}_s \to B^+K^-$ are listed in table 8 on the next page.

B^{**0}_s	Decay		Q [MeV/c^2]	Generated	Selected	Efficiency
B^0_{s1}	$B_s \to D^-_s\pi^+$	$D^-_s \to \bar{K}^*K^-$	176 ... 192	$1.276 \cdot 10^{+07}$	52	$1.284 \cdot 10^{-05}$
		$D^-_s \to \Phi\pi^-$			49	$1.210 \cdot 10^{-05}$
		$D^-_s \to 3\pi^\pm$			299	$7.384 \cdot 10^{-05}$
	$B_s \to D^-_s 3\pi^\pm$	$D^-_s \to \bar{K}^*K^-$	176 ... 192	$1.276 \cdot 10^{+07}$	351	$8.255 \cdot 10^{-05}$
		$D^-_s \to \Phi\pi^-$			635	$1.493 \cdot 10^{-04}$
		$D^-_s \to 3\pi^\pm$			605	$1.423 \cdot 10^{-04}$
B^{*0}_{s2}	$B_s \to D^-_s\pi^+$	$D^-_s \to \bar{K}^*K^-$	138 ... 162	$1.913 \cdot 10^{+07}$	60	$9.878 \cdot 10^{-06}$
		$D^-_s \to \Phi\pi^-$			61	$1.004 \cdot 10^{-05}$
		$D^-_s \to 3\pi^\pm$			372	$6.124 \cdot 10^{-05}$
	$B_s \to D^-_s 3\pi^\pm$	$D^-_s \to \bar{K}^*K^-$	138 ... 162	$1.913 \cdot 10^{+07}$	517	$8.106 \cdot 10^{-05}$
		$D^-_s \to \Phi\pi^-$			912	$1.430 \cdot 10^{-04}$
		$D^-_s \to 3\pi^\pm$			916	$1.436 \cdot 10^{-04}$

Table 6: Signal efficiencies of the B^0_{s1} and B^{*0}_{s2} in decays into $B^{**0}_s \to B_s\pi^+\pi^-$ estimated from the Monte Carlo samples.

	$B_s \to D_s^- \pi^+$		$B_s \to D_s^- 3\pi^\pm$	
	B_{s1}	B_{s2}^*	B_{s1}	B_{s2}^*
$D_s^- \to \bar{K}^* K^-$	$1.284 \cdot 10^{-05}$	$9.878 \cdot 10^{-06}$	$8.255 \cdot 10^{-05}$	$8.106 \cdot 10^{-05}$
$D_s^- \to \Phi\pi^-$	$1.210 \cdot 10^{-05}$	$1.004 \cdot 10^{-05}$	$1.493 \cdot 10^{-04}$	$1.430 \cdot 10^{-04}$
$D_s^- \to 3\pi^\pm$	$7.384 \cdot 10^{-05}$	$6.124 \cdot 10^{-05}$	$1.423 \cdot 10^{-04}$	$1.436 \cdot 10^{-04}$

Table 7: Summary of the signal efficiencies of the B_{s1}^0 and B_{s2}^{*0} in decays into $B_s^{**0} \to B_s\pi^+\pi^-$ estimated from the Monte Carlo samples.

	$B^+ \to \bar{D}^0 \pi^+$	
	B_{s1}	B_{s2}^*
$\bar{D}^0 \to K^+\pi^-$	$1.4329 \cdot 10^{-03}$	$8.8928 \cdot 10^{-04}$

Table 8: Signal efficiencies of the B_{s1}^0 and B_{s2}^{*0} in decays into $B_s^{**0} \to B^+K^-$ estimated from the Monte Carlo samples.

4.2 Efficiency uncertainty

For the decay $B_s^{**0} \to B^+K^-$ the efficiency uncertainty is estimated by studying the ratio of Monte Carlo to signal events in data at different selection cuts around the original selection cut:

$$\frac{N_{\mathrm{MC}}\left(n_{\mathrm{NN}} > n_{\mathrm{cut}}\right)}{N_{\mathrm{data}}\left(n_{\mathrm{NN}} > n_{\mathrm{cut}}\right)} \tag{6}$$

For the selection of these events different cuts n_{cut} on the output n_{NN} of the same neural network which was used in reference [7] are done. Moreover, the numbers of signal events obtained from the data sample are sideband subtracted. To study the efficiency for the B_{s1}^0 and B_{s2}^{*0} separately, the events are selected in the appropriate Q value ranges with accordingly different sideband regions.

Figure 16 on the following page and figure 17 on the next page show the ratio as a function of the cut on the output of the neural network. The major part of the values are within a band of 10 % uncertainty of the average value. Therefore, a relative uncertainty of 10 % is assumed for the efficiency of the B_{s1}^0 and B_{s2}^{*0} in the decay $B_s^{**0} \to B^+K^-$.

For lack of any clean signal the efficiency uncertainty of the B_{s1}^0 and B_{s2}^{*0} in the decay $B_s^{**0} \to B_s\pi^+\pi^-$ cannot be directly studied. Assuming the efficiency being dominated by the efficiency of the slow pions in this decay, the efficiency uncertainty might be estimated by studying a equivalent decay with similar final states and similar decay topology. Therefore, the decay $\Psi(2s) \to J/\psi\pi\pi$ is observed with the data samples and the selection procedure taken from reference [17]. The Monte Carlo to data ratio of the minimum of the transverse momenta of the two pions from the decay $\Psi(2s) \to J/\psi\pi\pi$ is plotted in bins of the transverse momentum shown in figure 18 on the following page. The distributions of the transverse momentum is obtained by applying the selection procedure from reference [17] and an additionally sideband subtraction for the data

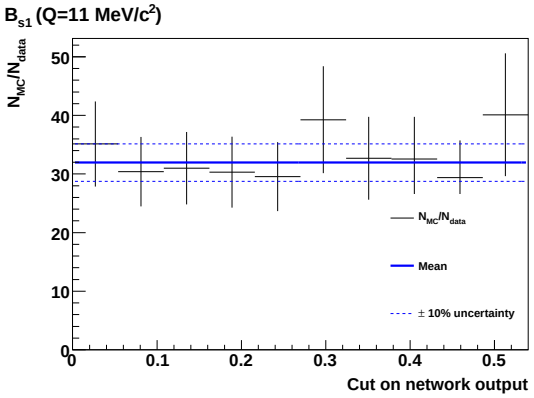

Figure 16: Ratio of the number of Monte Carlo to data events for the B^0_{s1} in the decay $B^{**0}_s \to B^+K^-$ at different cut on the output of the neural network.

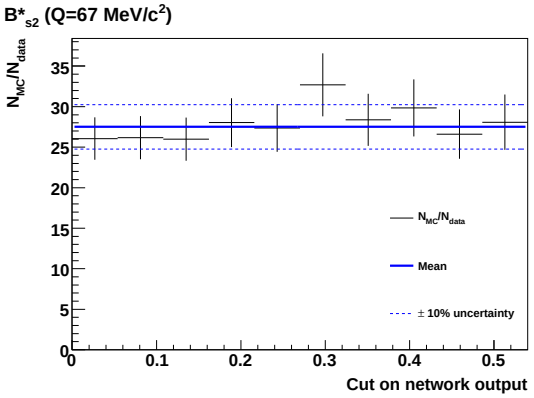

Figure 17: Ratio of the number of Monte Carlo to data events for the B^{*0}_{s2} in the decay $B^{**0}_s \to B^+K^-$ at different cut on the output of the neural network.

sample. Since most of these values are not more than 10 % away from the average value a relative efficiency uncertainty for the decay $B^{**0}_s \to B_s\pi^+\pi^-$ of 10 % is assumed.

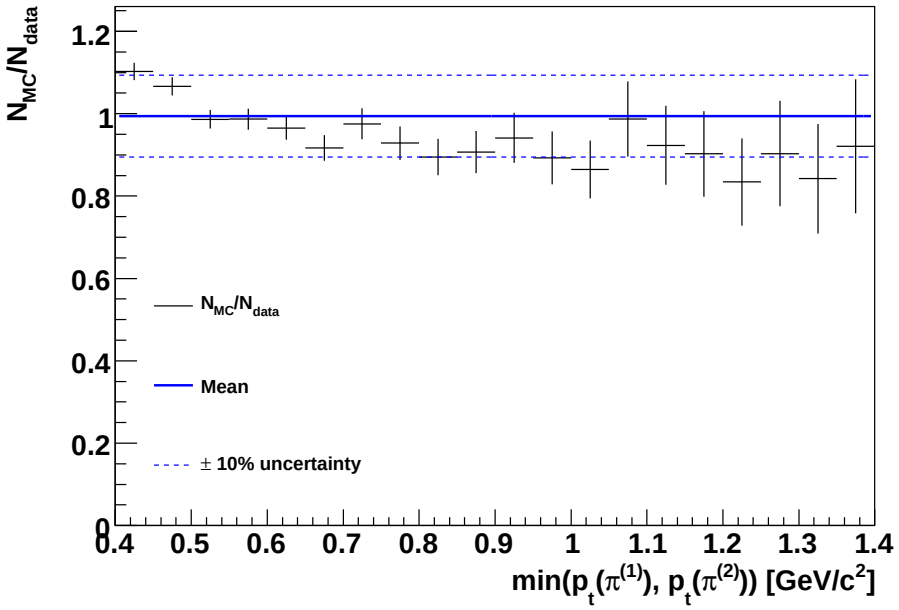

Figure 18: Ratio of the number of Monte Carlo to data events (normalized) as a function of the minimal pion transverse momentum which is the minimum of the two transverse momenta of the pions coming from the decay $\Psi\,(2s) \to J/\psi\pi\pi$.

5 Unbinned likelihood fit

5.1 Fit description

In order to measure the branching ratio of the B_{s1}^0 and B_{s2}^{*0} in decays of $B_s^{**0} \to B_s\pi^+\pi^-$ an extended unbinned likelihood fit to the Q value of the data sample is performed. The used fitter is built on MINUIT minimiser provided by the ROOT [18] package. The likelihood expression L of the fit function consists of a Gaussian signal component and a exponential background component:

$$L = \prod_i \left[\frac{2 \cdot \mathrm{BR}}{\frac{3}{2}(1-\mathrm{BR})} \cdot \widetilde{\mathrm{BR}} \cdot N'_{B_s^{**0}} \cdot f_{sig}(Q_i) + N_{bkg} \cdot f_{bkg}(Q_i) \right] \tag{7}$$

where i runs over all events in the data sample. The components of the likelihood functions in detail are:

- **Branching ratio BR**

  BR is the branching ratio in question of the B_s^{**0}. Based on this parameter a Baysian limit is set on the branching ratio for the B_{s1}^0 and B_{s2}^{*0} in decays of $B_s^{**0} \to B_s\pi^+\pi^-$. It is defined in the following way:

  $$\mathrm{BR} = \frac{\frac{3}{2}\,\mathrm{BR}(B_s^{**0} \to B_s\pi^+\pi^-)}{\frac{3}{2}\,\mathrm{BR}(B_s^{**0} \to B_s\pi^+\pi^-) + 2\,\mathrm{BR}(B_s^{**0} \to B^+K^-)} \tag{8}$$

  Under the assumption that $B_s^{**0} \to B_s\pi\pi$ and $B_s^{**0} \to BK$ are the only decay modes it is the absolute branching ratio of the decays $B_s^{**0} \to B_s\pi^+\pi^-$. The advantage of this definition is the restriction to the interval $[0,1]$ avoiding divergences in the limit calculation for a flat prior. The factors $\frac{3}{2}$ and 2 take into account the different ratios of decay probabilities into charged and uncharged particles considering the proper isospin selection rules:

  $$\frac{\mathrm{BR}(B_s^{**0} \to B_s\pi^+\pi^-)}{\mathrm{BR}(B_s^{**0} \to B_s\pi^+\pi^-) + \mathrm{BR}(B_s^{**0} \to B_s\pi^0\pi^0)} = \frac{2}{3} \tag{9}$$

  $$\frac{\mathrm{BR}(B_s^{**0} \to B^+K^-)}{\mathrm{BR}(B_s^{**0} \to B^+K^-) + \mathrm{BR}(B_s^{**0} \to B^0K^0)} = \frac{1}{2} \tag{10}$$

- **Branching ratio $\widetilde{\mathrm{BR}}$**

  $\widetilde{\mathrm{BR}}$ is the relative branching ratio of the B_s subdecays with respect to the reference decay taking into account the different efficiencies ϵ of the B_s and B decay modes. It is defined by

  $$\widetilde{\mathrm{BR}} = \frac{\sum_i \epsilon_i \mathrm{BR}_i}{\epsilon(B \to D\pi) \cdot \mathrm{BR}(B \to D\pi)} \tag{11}$$

  where i runs over the six different B_s subdecays whose branching ratios are denoted by the parameter BR_i. Section 4 on page 22 describes the efficiencies whereas the values for the branching ratios are taken from reference [8].

- **Signal events** $N'_{B_s^{**0}}$

  Parameter $N'_{B_s^{**0}}$ denotes the number of B_s^{**0} signal candidates in the reference decay $B_s^{**0} \rightarrow B^+K^-$. Section 5.3 on page 29 describes how this number is obtained.

- **Background events** N_{bkg}

  Number of background events in the $B_s^{**0} \rightarrow B_s\pi^+\pi^-$ data sample.

- **Signal component** f_{sig}

  The signal contribution of the likelihood expression is composed of a normalised Gaussian function

$$f_{sig}\left(Q_i\right) = \mathcal{G}\left(Q_i; \mu_{\mathrm{Q}}, \sigma_{\mathrm{Q}}\right) \tag{12}$$

  with mean μ_{Q} and width σ_{Q}. In the fit procedure the mean and the width of the signal component are kept at fixed values. For the signal mean the signal resolution described in section 5.2 on the next page is used. The position of the signal peaks is known from the reference channel $B_s^{**0} \rightarrow B^+K^-$. Therefore, the B_{s1}^0 signal mean is set to 184 MeV/c^2 and to 150 MeV/c^2 for the B_{s2}^{*0} when fitting the $B_s\pi^+\pi^-$ samples.

- **Background component** f_{bkg}

  Since the background is mainly combinatorial background and should be zero at $Q = 0$ due to physical reasons the background contribution is modelled with a linear term times an exponential function

$$f_{bkg}\left(Q_i\right) = \frac{Q_i \cdot \exp\left(\alpha \cdot Q_i\right)}{\int_{Q_{min}}^{Q_{max}} Q' \cdot \exp\left(\alpha \cdot Q'\right) \mathrm{d}Q'} \tag{13}$$

  where α is a free parameter. The background function is normalised within the fitting range from $Q_{min} = 0$ GeV/c^2 to $Q_{max} = 0.2$ GeV/c^2.

In the fit, not all parameters are free. The signal peaks for instance are kept at fixed positions where the signal is expected and parameters implying an a priori knowledge are put into the likelihood function with a Gaussian constraint. The main benefit of the Gaussian constraints is the automatic inclusion of systematic uncertainties into the fit. In the fitting procedure the negative logarithmic likelihood function $\mathcal{L}$ is minimised:

$$\begin{aligned}\mathcal{L} = &-2 \cdot \ln L \\ &+ 2 \cdot \frac{2 \cdot \mathrm{BR}}{\frac{3}{2}\left(1 - \mathrm{BR}\right)} \cdot \widetilde{\mathrm{BR}} \cdot N'_{B_s^{**0}} + 2 \cdot N_{bkg} \\ &+ \left(\frac{\widetilde{\mathrm{BR}} - \mu_{\widetilde{\mathrm{BR}}}}{\sigma_{\widetilde{\mathrm{BR}}}}\right)^2 + \left(\frac{N'_{B_s^{**0}} - \mu_{N'_{B_s^{**0}}}}{\sigma_{N'_{B_s^{**0}}}}\right)^2\end{aligned} \tag{14}$$

Here, the second line in equation (14) implies the extended part of the likelihood fit providing more reliable errors. As above mentioned the mean and width of the signal peak are kept fixed during the fit. The mean values of the signal peaks are taken from the CDF observation of orbitally excited B_s^{**0} mesons [7] whereas the widths of the signal peaks are set to the Q value resolution, described in section 5.2. The number of signal candidates of the reference decay and the branching ratio $\widetilde{\mathrm{BR}}$ are floating in the fit but confined by Gaussian constraints implemented in the third line of equation (14). The number of signal candidates in the reference decay is again taken from reference [7] and quoted for completeness in table 9 on page 30. The efficiency corrected branching ratio $\widetilde{\mathrm{BR}}$ is discussed in section 4.1 on page 22. Due to the Gaussian constraints the fit has to pay the penalty if it moves the floating parameters to far away from the values put initially into the fit.

5.2 Mass resolution of the B_s^{**0} signal

The mass resolution is put into the fit as the width of the signal peak. For estimating the mass resolution the Monte Carlo sample used in the neural network training is studied. Since the Monte Carlo samples were generated flat in the B_s^{**0} mass the dependence of the mass resolution on the the Q value can also be studied.

In order to estimated the resolution, Monte Carlo events from the Monte Carlo sample are selected by applying the same selection algorithm as for the data events as described in section 3 on page 8. The difference between the measured and generated Q value of the selected events is plotted in histograms with 40 MeV/c^2 bin width and each of the peaks of either the decay $B_s^{**0} \to B_s\pi^+\pi^-$ or $B_s^{**0} \to B_s^*\pi^+\pi^-$ is fitted using a double Gaussian function of the form

$$\mathcal{G}(Q) = \frac{N_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(Q-Q_0)^2}{2\sigma_1^2}} + \frac{N_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(Q-Q_0)^2}{2\sigma_2^2}} \tag{15}$$

where Q_0 is the common mean of the two Gaussian functions with width σ_1 and σ_2, respectively and N_1 and N_2 are the number of events of each single Gaussian component. For a better monotonous behaviour as a function of the measured Q value the ratio of the widths and numbers of events in each double Gaussian are fixed to the following values

$$N_2 = 0.33 \cdot N_1 \tag{16}$$

$$\sigma_2 = 2.63 \cdot \sigma_1 \tag{17}$$

The arbitrary factors of 0.33 and 2.63 are chosen since they describe the residual Q value distribution very well as figure 19 on the following page shows for the Monte Carlo events of the decay $B_s^{**0} \to B_s\pi^+\pi^-$ with $B_s \to D_s^-\pi^+$, $D_s^- \to \bar{K}^*K^-$, for instance.

Figure 20 and figure 21 on page 29 show the Q value resolution for the different B_s^{**0} decay channels. The Q value resolution at the different Q values where the B_{s1}^0

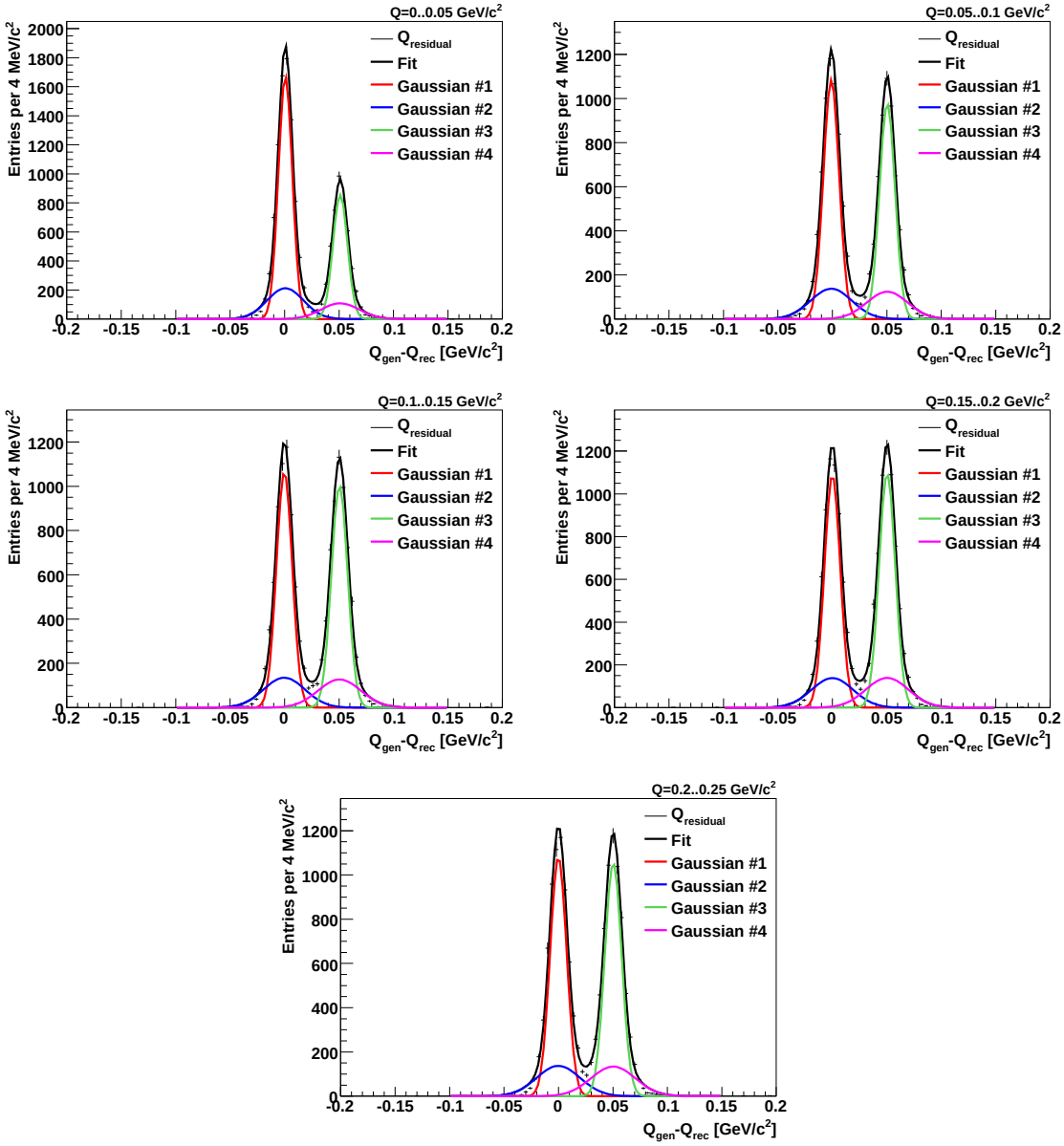

Figure 19: Residual Q value distribution in Monte Carlo of the decay $B_s^{**0} \to B_s\pi^+\pi^-$ with $B_s \to D_s^-\pi^+$, $D_s^- \to \bar{K}^*K^-$ for different Q value ranges. Having two peaks arises from the fact that the photon from the decay $B_s^* \to B_s\gamma$ is not resolved by the CDF-II detector.

and B_{s2}^{*0} is expected is shown in figure 22 and figure 23 on the facing page. Based on these plots the widths of the signal peaks were set to the average value of the Q value resolution in the unbinned likelihood fit. The values are

$$\sigma_Q(B_{s1}^0) = 7.46 \text{ MeV/c}^2 \quad (18)$$

$$\sigma_Q(B_{s2}^{*0}) = 7.14 \text{ MeV/c}^2 \quad (19)$$

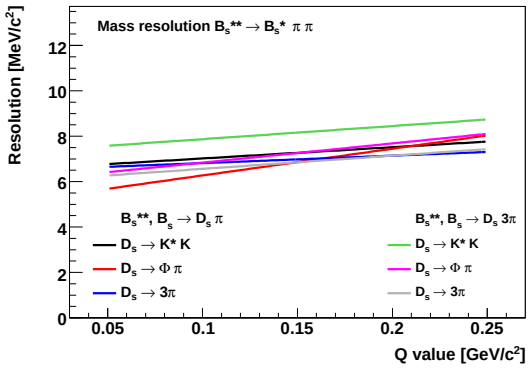

Figure 20: Q value resolution of the B_s^{**0} signal for the decay $B_s^{**0} \to B_s\pi^+\pi^-$ estimated from the Monte Carlo samples.

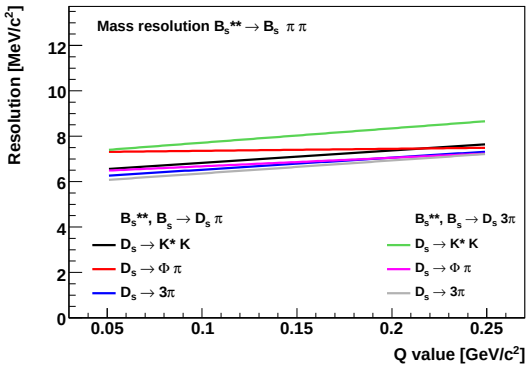

Figure 21: Q value resolution of the B_s^{**0} signal for the decay $B_s^{**0} \to B_s^*\pi^+\pi^-$ estimated from the Monte Carlo samples.

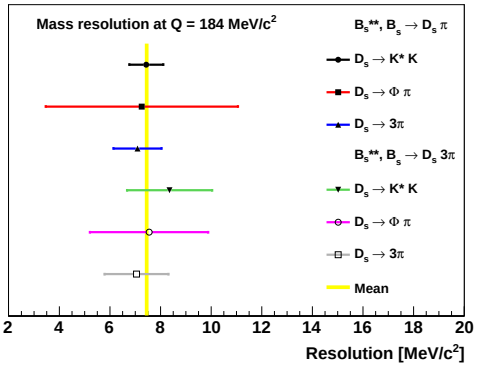

Figure 22: Q value resolution of the B_s^{**0} signal for the decay $B_s^{**0} \to B_s\pi^+\pi^-$ at the distinct Q value where the B_{s1}^0 ($Q = 184$ MeV/c^2) is expected. The mass resolutions are estimated from Monte Carlo samples.

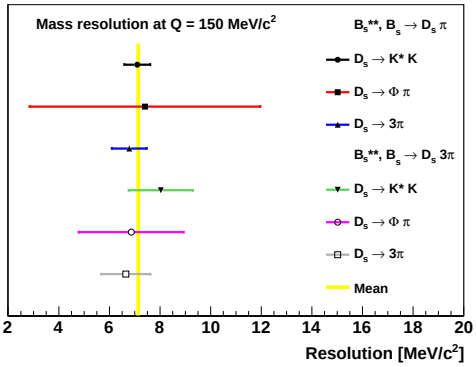

Figure 23: Q value resolution of the B_s^{**0} signal for the decay $B_s^{**0} \to B_s^*\pi^+\pi^-$ at the distinct Q value where the B_{s2}^{*0} ($Q = 150$ MeV/c^2) is expected. The mass resolutions are estimated from Monte Carlo samples.

5.3 Candidates in the reference channel

The number of the signal events in the reference channel $B_s^{**0} \to B^+K^-$ are taken from reference [7]. In the analysis of the reference decay a data sample of 1 fb^{-1} is used. Therefore, the measured event numbers and uncertainties are extrapolated to a data sample corresponding to 2.8 fb^{-1} as it is used for this analysis of the $B_s^{**0} \to B_s\pi^+\pi^-$ decays.

Assuming the number of B_s^{**0} mesons scales with the number of B_s mesons, the number of events given in the reference channel are multiplied by the ratio of the amount of B_s mesons in the full $B_s\pi^+\pi^-$ data sample of 2.8 fb^{-1} and below 1 fb^{-1} in this data sample. The number of B_s mesons in the $B_s\pi^+\pi^-$ data sample are obtained by fitting the B_s mass of events selected by applying cuts on the outputs of the neural networks trained for selection the B_s candidates. These cuts are decay mode dependent and they are same as described in reference [12]. Table 9 gives the numbers of signal events in the reference channel.

$B_s^{**0} \to B^+K^-$	measured in 1 fb^{-1}	extrapolated to 2.8 fb^{-1}
B_{s1}^0	20.66 ± 7.12	39.09 ± 13.47
B_{s2}^{*0}	55.74 ± 19.2	105.46 ± 36.33

Table 9: Number of signal candidates in the reference channel $B_s^{**0} \to B^+K^-$ with $B^+ \to \bar{D}^0\pi^+$, selected in 1 fb^{-1} and extrapolated to 2.8 fb^{-1}. The measured numbers are taken from reference [7].

5.4 Bayesian Limit

The Bayesian limit on the branching ratio of the narrow B_s^{**0} states is calculated according to the formalism described in section 32.3.1 of the Particle Physics Booklet [8]. In order to set the limit a fit to the data sample is performed as described in section 5.1 on page 25 which is referred to as the free fit. In addition to the free fit a set of consecutive fits is done where the parameter BR for the total branching ratio is fixed to values in the inteval $[0, 1)$. Each fit returns a negative logarithmic likelihood value $\mathcal{L}$ as defined in equation (14) on page 26. The negative logarithmic likelihood value is transformed back into the likelihood value L:

$$L = \exp\left(\frac{-\mathcal{L} + \mathcal{L}_0}{2}\right) \tag{20}$$

Here, $\mathcal{L}_0$ is the negative logarithmic likelihood value of the free fit. $\mathcal{L}_0$ is constant for the complete set of consecutive fits having a fixed value for the branching ratio. Thus, it scales L by a constant factor which has no effects on the result but avoids numerical problems with extreme values of L.

The Bayesian posterior probability density function is calculated from the likelihood value as defined by

$$p\left(\vec{a}\middle|\,\vec{x}\right) = \frac{L\left(\vec{x}\middle|\,\vec{a}\right)\pi\left(\vec{a}\right)}{\int L\left(\vec{x}\middle|\,\vec{a}'\right)\pi\left(\vec{a}'\right)\mathrm{d}\vec{a}'} \tag{21}$$

where $\pi\left(\vec{a}\right)$ is the prior probability density function which is assumed being flat:

$$\pi\left(\vec{a}\right) = 1 \qquad \text{for all } \vec{a} \tag{22}$$

The denominator in equation (21) on the facing page is merely intended to normalise the posterior probability density function. In the single parameter case a credible interval can be determined containing a given fraction $(1 - \beta)$ of the probability:

$$1 - \beta = \int_{a_{\mathrm{lo}}}^{a_{\mathrm{up}}} p\,(a|\,\vec{x})\,\mathrm{d}a \tag{23}$$

For the upper limit a_{up} on the branching ratios of the narrow B_s^{**0} states the single parameter a is the total branching ratio $a = \mathrm{BR}$ and $a_{\mathrm{lo}} = 0$. The credible level β is set to 5%.

6 Results

In this chapter, the results of the unbinned maximum likelihood fit to the Q value distribution are given. The results are obtained by applying the fit function described in section 5 on page 25 to the data samples selected as shown in section 3 on page 8.

Table 10 and table 11 give the fit parameter obtained from the free fit, where only the mean and width of the signal peaks are fixed. From the corresponding plots shown in figure 24 and figure 25 on the facing page of these fits it is obvious that no significant signal is observed. Therefore, a Bayesian limit on the branching ratios of these decays is set using a flat prior.

Parameter	Value	Error	
BR	$-1.787 \cdot 10^{+00}$	$4.615 \cdot 10^{+00}$	
$\widetilde{\mathrm{BR}}$	$1.834 \cdot 10^{-01}$	$5.415 \cdot 10^{-02}$	constrained
$N'_{B_s^{**0}}$	$3.910 \cdot 10^{+01}$	$1.349 \cdot 10^{+01}$	constrained
μ_{Q}	$1.840 \cdot 10^{-01}$	fixed	$\left[\mathrm{GeV/c^2}\right]$
σ_{Q}	$7.458 \cdot 10^{-03}$	fixed	$\left[\mathrm{GeV/c^2}\right]$
N_{bkg}	$1.521 \cdot 10^{+02}$	$1.355 \cdot 10^{+01}$	
α	$4.234 \cdot 10^{+00}$	$1.496 \cdot 10^{+00}$	

Table 10: Fit result parameters of the free fit for the B_{s1}^0.

Parameter	Value	Error	
BR	$1.550 \cdot 10^{-01}$	$1.074 \cdot 10^{-01}$	
$\widetilde{\mathrm{BR}}$	$2.817 \cdot 10^{-01}$	$8.424 \cdot 10^{-02}$	constrained
$N'_{B_s^{**0}}$	$1.055 \cdot 10^{+02}$	$3.649 \cdot 10^{+01}$	constrained
μ_{Q}	$1.500 \cdot 10^{-01}$	fixed	$\left[\mathrm{GeV/c^2}\right]$
σ_{Q}	$7.139 \cdot 10^{-03}$	fixed	$\left[\mathrm{GeV/c^2}\right]$
N_{bkg}	$1.387 \cdot 10^{+02}$	$1.247 \cdot 10^{+01}$	
α	$4.723 \cdot 10^{+00}$	$1.617 \cdot 10^{+00}$	

Table 11: Fit result parameters of the free fit for the B_{s2}^{*0}.

Figure 26 to 35 show the projection of the likelihood for all free parameters used in the likelihood function.

In order to calculate an upper limit on the branching ratios for either the B_{s1}^0 and B_{s2}^{*0} the Bayesian approach is used with a flat prior. The same data sample as used in the free fit is consecutively refitted with different but fixed values for the branching ratio BR. Based on these fits the fit posterior probability density as a function of the branching ratio BR is achieved, as shown in figure 36 and figure 37 on page 36

The integrated posterior probability function yields a direct value for the upper limit of the branching ratio for either the B_{s1}^0 and B_{s2}^{*0}, as seen in figure 38 and figure 39 on page 36. For the B_{s1}^0 a upper limit of 0.54 is obtained and 0.53 for B_{s2}^{*0}. Both limits are given for a credible level of 95%.

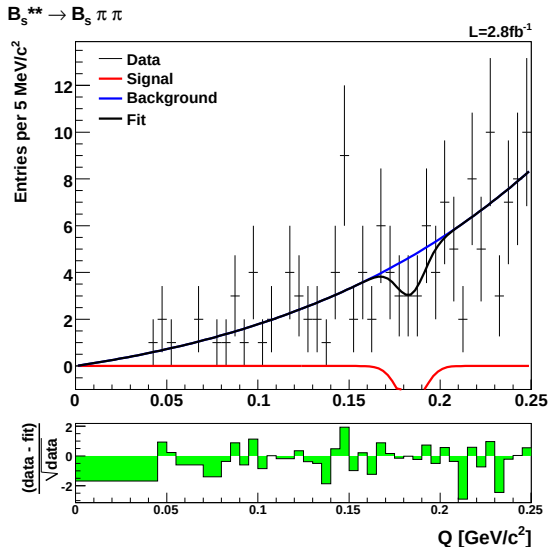

Figure 24: Result of the free fit for the B^0_{s1}.

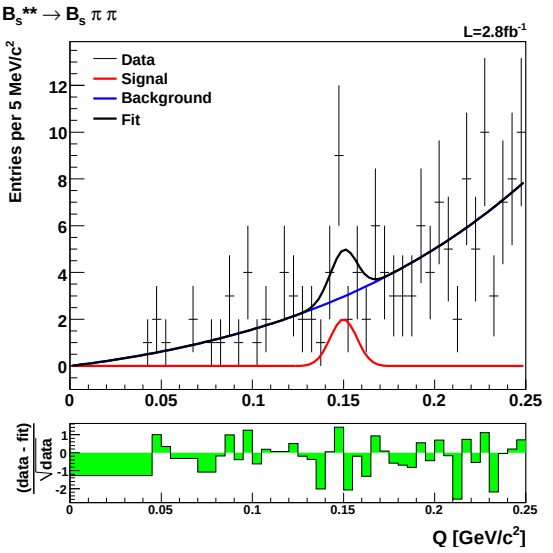

Figure 25: Result of the free fit for the B^{*0}_{s2}.

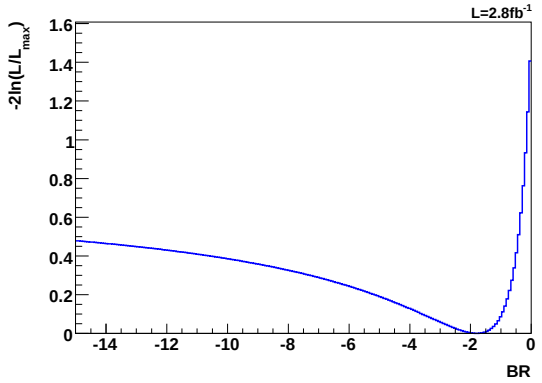

Figure 26: Negative log likelihood scan for the parameter BR in the free fit for the B^0_{s1}.

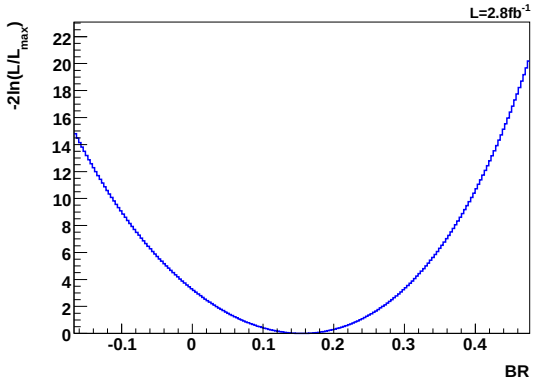

Figure 27: Negative log likelihood scan for the parameter BR in the free fit for the B^{*0}_{s2}.

The rather loose limit on the branching ratios of the B^0_{s1} and B^{*0}_{s2} can be understood in the context of the rather high uncertainties of values entering the fit function. Especially the large uncertainties of the branching ratios of the B_s subdecays and the quite high uncertainty of the number of candidates in the reference channel make it impossible giving a more stringent upper limit. Table 12 on the next page gives an overview about the quantities with the highest uncertainties entering the fit.

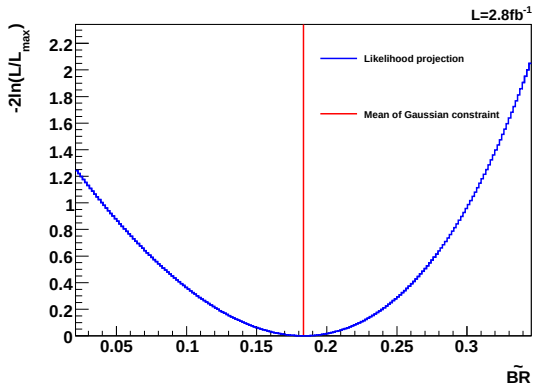

Figure 28: Negative log likelihood scan for the parameter $\widetilde{\mathrm{BR}}$ in the free fit for the B^0_{s1}.

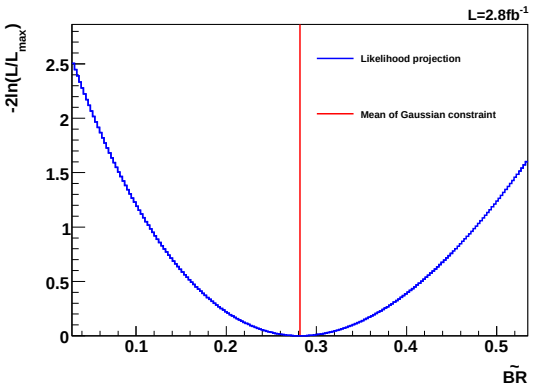

Figure 29: Negative log likelihood scan for the parameter $\widetilde{\mathrm{BR}}$ in the free fit for the B^{*0}_{s2}.

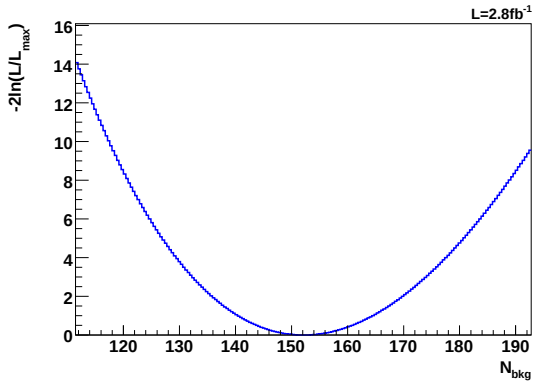

Figure 30: Negative log likelihood scan for the parameter N_{bkg} in the free fit for the B^0_{s1}.

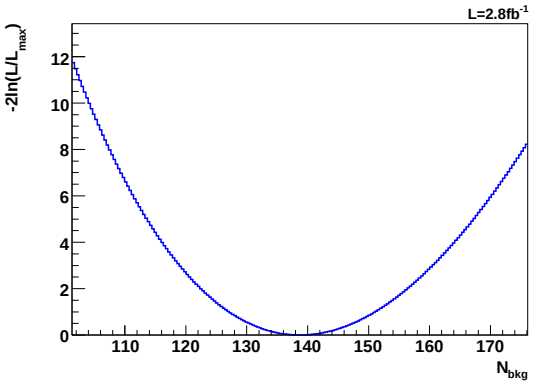

Figure 31: Negative log likelihood scan for the parameter N_{bkg} in the free fit for the B^{*0}_{s2}.

Quantity	Relative uncertainty	Affected parameter	Reference
$\mathrm{BR}(B_s \to D_s\pi)$	28%	$\widetilde{\mathrm{BR}}$	[8]
$\mathrm{BR}(B_s \to D_s 3\pi)$	39%	$\widetilde{\mathrm{BR}}$	[8]
$N(B_s^{**0} \to B^+K^-)$	31%	$N'_{B_s^{**0}}$	[7]

Table 12: Quantities entering the fit function with rather large relative uncertainties. It is also listed which fit function parameters they affect and where they are taken from.

6.1 Frequentist Limit

In order to set an upper limit on the branching ratios based on the frequentist approach Monte Carlo samples are generated. These samples are generated for a couple of

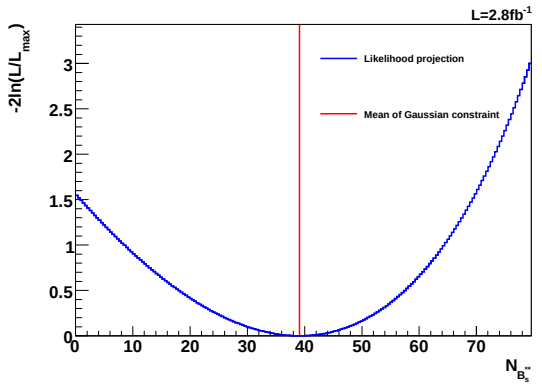

Figure 32: Negative log likelihood scan for the parameter $N'_{B_s^{**0}}$ in the free fit for the B_{s1}^0.

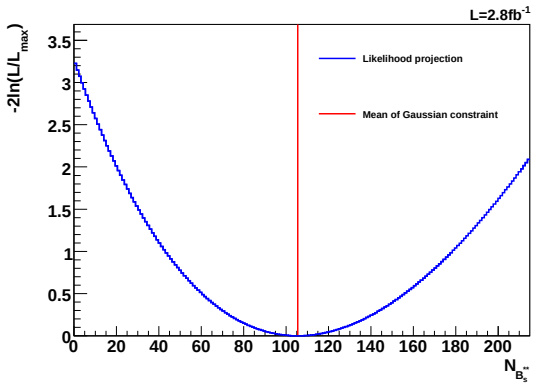

Figure 33: Negative log likelihood scan for the parameter $N'_{B_s^{**0}}$ in the free fit for the B_{s2}^{*0}.

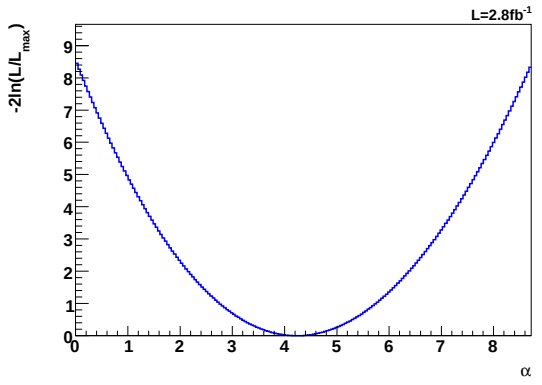

Figure 34: Negative log likelihood scan for the parameter α in the free fit for the B_{s1}^0.

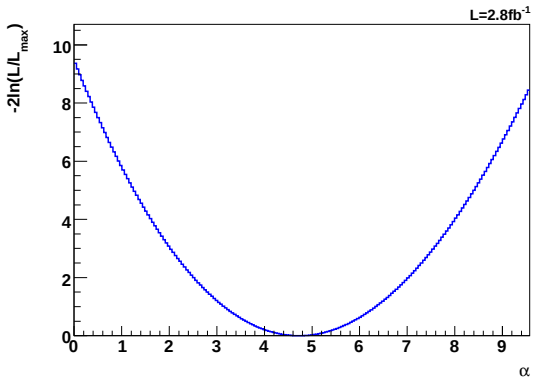

Figure 35: Negative log likelihood scan for the parameter α in the free fit for the B_{s2}^{*0}.

different branching ratios around the limit calculated by the Bayesian approach. From fits having fixed branching ratio to the data the parameters are obtained for generating the Monte Carlo samples. For each branching ratio 10000 samples are generated and fitted with two kind of fits. The first fit has a fixed branching ratio and in the second fit it is floating. From the distribution of the likelihood ratios the p value is calculated for each branching ratio.

Figure 40 and 41 show the p value for the B_{s1}^0 and B_{s2}^{*0} as a function of branching ratio. The points of the tested branching ratios are interpolated by an exponential function of the form:

$$f(x) = A\, e^{\alpha x} \tag{24}$$

where A and α are free nuisance parameters. From that interpolation the limit for a confidence level of 95% are determined to be 0.37 for the B_{s1}^0 and 0.48 for the B_{s2}^{*0}.

The frequentist limit of the B_{s2}^{*0} state is quite similar to the Bayesian limit, while

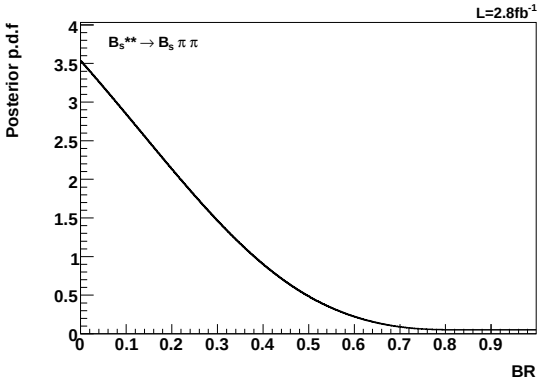

Figure 36: Posterior probability density function of the B^0_{s1} as a function of the branching ratio BR.

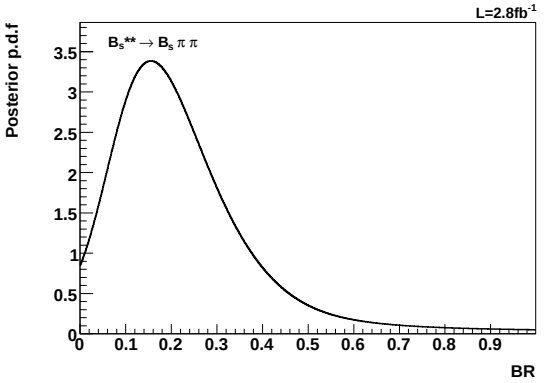

Figure 37: Posterior probability density function of the B^{*0}_{s2} as a function of the branching ratio BR.

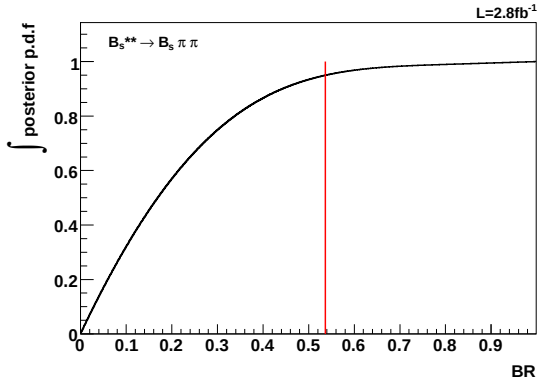

Figure 38: The integrated posterior probability density function of the B^0_{s1} as a function of the branching ratio BR.

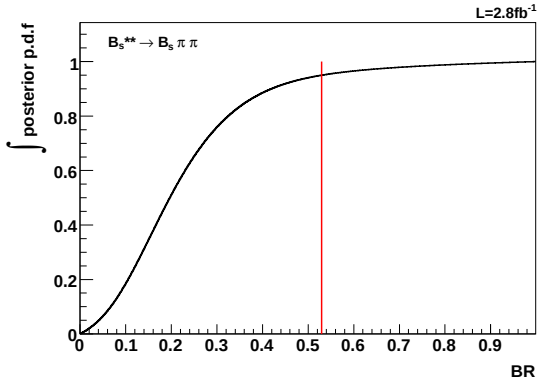

Figure 39: The integrated posterior probability density function of the B^{*0}_{s2} as a function of the branching ratio BR.

there is some difference for the B^0_{s1} state. This can be understood by looking at the free fit to the Q value in data as shown in figure 26 on page 33. The signal contribution dips below the background distribution yielding a negative number of signal candidates and hence a negative branching ratio. However, the Bayesian method does not include negative branching ratios since they are restricted to the range between 0 and 1.

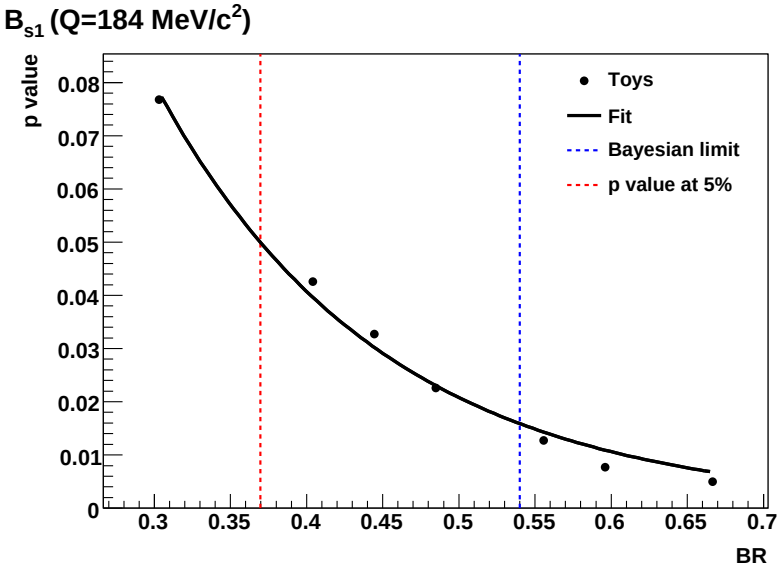

Figure 40: The p value of the B^0_{s1} as a function of the branching ratio.

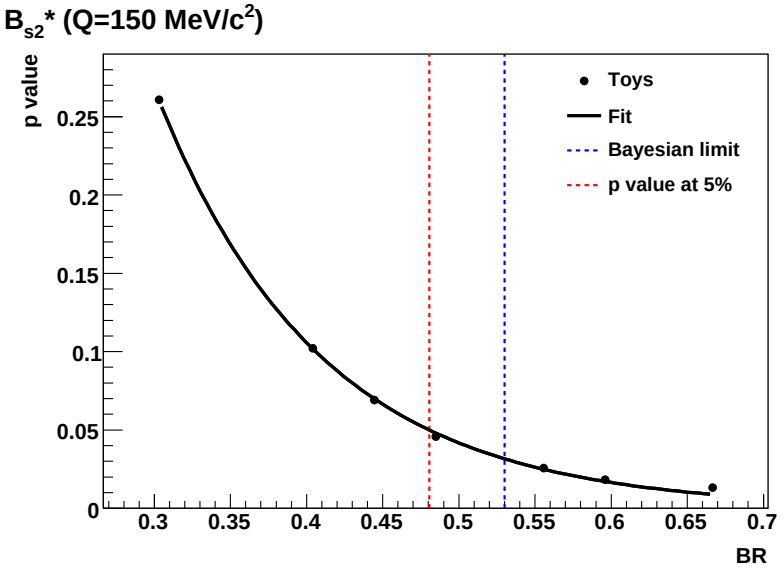

Figure 41: The p value of the B^{*0}_{s2} as a function of the branching ratio.

7 Summary

In summary, we performed a search for orbitally excited B_s mesons in decays of $B_s^{**0} \to B_s\pi^+\pi^-$. We calculated an upper limit of the branching ratios for either the B_{s1}^0 and B_{s2}^{*0} mesons with respect to the reference decay $B_s^{**0} \to B^+K^-$. From the unbinned maximum likelihood to the Q value distribution and the Bayesian approach for calculating an upper limit of the branching ratio we obtain

$$\frac{\mathrm{BR}(B_{s1}^0 \to B_s\pi\pi)}{\mathrm{BR}(B_{s1}^0 \to B_s\pi\pi) + \mathrm{BR}(B_s^{**0} \to BK)} < 0.54 \tag{25}$$

$$\frac{\mathrm{BR}(B_{s2}^{*0} \to B_s\pi\pi)}{\mathrm{BR}(B_{s2}^{*0} \to B_s\pi\pi) + \mathrm{BR}(B_s^{**0} \to BK)} < 0.53 \tag{26}$$

Both limits are given for a credible level of 95%.

A Monte Carlo decay tables

A.1 B_s^{**0} decays with $B_s \to D_s^- \pi^+$

```
Decay B_s2*0
.2        B_s*0   pi+ pi-                    PHSP;
.2        B_s0    pi+ pi-                    PHSP;
Enddecay
#
Decay anti-B_s2*0
.2        anti-B_s*0   pi- pi+                    PHSP;
.2        anti-B_s0    pi- pi+                    PHSP;
Enddecay
#
#
Decay B_s*0
.2        B_s0 gamma                         PHSP;
Enddecay
#
Decay anti-B_s*0
.2        anti-B_s0 gamma                         PHSP;
Enddecay
#
#
Decay B_s0
0.0100    D_s-          pi+                PHSP;
#0.0090    a_1+ D_s-                              SVS;
#0.0005    D_s-   rho0   pi+                    PHSP;
#0.0005    D_s-   pi-    pi+    pi+             PHSP;
0.0005    J/psi  phi                          PHSP;
Enddecay
#
Decay anti-B_s0
0.0100    D_s+          pi-                PHSP;
#0.0090    a_1- D_s+                              SVS;
#0.0005    D_s+   rho0   pi-                    PHSP;
#0.0005    D_s+   pi+    pi-    pi-             PHSP;
0.0005    J/psi  phi                          PHSP;
Enddecay
#
Decay D_s+
0.0247     phi    pi+                       SVS;
0.0247     anti-K*0    K+                         SVS;
0.0004     rho0    pi+                       SVS;
0.0180     f_0     pi+                       PHSP;
0.0023     f_2     pi+                       PHSP;
0.0040     pi+    pi-    pi+                 PHSP;
Enddecay
#
Decay D_s-
0.0247     phi    pi-                       SVS;
0.0247     K*0    K-                     SVS;
0.0004     rho0   pi-                       SVS;
0.0180     f_0     pi-                      PHSP;
0.0023     f_2     pi-                      PHSP;
0.0040     pi-    pi-    pi+                 PHSP;
Enddecay
#
Decay f_0
0.5200    pi+  pi-                          PHSP;
Enddecay
#
Decay f_2
0.5650    pi+  pi-                          TSS;
```

```
Enddecay
#
Decay phi
0.4910    K+    K-                              VSS;
Enddecay
#
Decay a_1+
0.6       rho0  pi+                         VVS_PWAVE 1.0 0.0 0.0 0.0 -0.1 0.0;
Enddecay
#
Decay a_1-
0.6       rho0  pi-                         VVS_PWAVE 1.0 0.0 0.0 0.0 -0.1 0.0;
Enddecay
#
Decay rho0
1.000     pi+ pi-                                 VSS;
Enddecay
#
Decay K*0
0.6657       K+   pi-                               VSS;
Enddecay
#
Decay anti-K*0
0.6657       K-   pi+                               VSS;
Enddecay
#
Decay J/psi
1   mu+ mu-    PHSP;
Enddecay
#
Decay phi
1   K+ K-    PHSP;
Enddecay
#
#
End
```

A.2 B_s^{**0} decays with $B_s \to D_s^- 3\pi^\pm$

```
Decay B_s2*0
.2        B_s*0   pi+ pi-                       PHSP;
.2        B_s0    pi+ pi-                       PHSP;
Enddecay
#
Decay anti-B_s2*0
.2        anti-B_s*0   pi- pi+                        PHSP;
.2        anti-B_s0    pi- pi+                        PHSP;
Enddecay
#
#
Decay B_s*0
.2        B_s0 gamma                            PHSP;
Enddecay
#
Decay anti-B_s*0
.2        anti-B_s0 gamma                             PHSP;
Enddecay
#
#
Decay B_s0
#0.0100    D_s-          pi+                  PHSP;
0.0090    a_1+ D_s-                                SVS;
0.0005    D_s-   rho0   pi+                        PHSP;
0.0005    D_s-   pi-    pi+    pi+                 PHSP;
Enddecay
```

```
#
Decay anti-B_s0
#0.0100    D_s+          pi-                PHSP;
0.0090    a_1-  D_s+                          SVS;
0.0005    D_s+   rho0   pi-                   PHSP;
0.0005    D_s+   pi+    pi-    pi-            PHSP;
Enddecay
#
Decay D_s+
0.0247     phi    pi+                        SVS;
0.0247     anti-K*0    K+                        SVS;
0.0004     rho0    pi+                        SVS;
0.0180     f_0     pi+                        PHSP;
0.0023     f_2     pi+                        PHSP;
0.0040     pi+    pi-    pi+                  PHSP;
Enddecay
#
Decay D_s-
0.0247     phi    pi-                        SVS;
0.0247     K*0    K-                      SVS;
0.0004     rho0   pi-                        SVS;
0.0180     f_0     pi-                       PHSP;
0.0023     f_2     pi-                       PHSP;
0.0040     pi-    pi-    pi+                  PHSP;
Enddecay
#
Decay f_0
0.5200    pi+  pi-                           PHSP;
Enddecay
#
Decay f_2
0.5650    pi+  pi-                           TSS;
Enddecay
#
Decay phi
0.4910    K+    K-                           VSS;
Enddecay
#
Decay a_1+
0.6     rho0 pi+                          VVS_PWAVE 1.0  0.0  0.0  0.0  -0.1  0.0;
Enddecay
#
Decay a_1-
0.6     rho0 pi-                          VVS_PWAVE 1.0  0.0  0.0  0.0  -0.1  0.0;
Enddecay
#
#
Decay rho0
1.000     pi+ pi-                             VSS;
Enddecay
#
Decay K*0
0.6657       K+  pi-                            VSS;
Enddecay
#
Decay anti-K*0
0.6657       K-  pi+                            VSS;
Enddecay
#
End
```

References

[1] F. Bedeschi et al. Observation of orbitally excited ($L = 1$) B mesons. *CDF/ANAL/BOTTOM/CDFR/7820*, 2005. 4

[2] E. Gerchtein M. Campanelli. Measurement of mass and width of the orbitally–excited charmed mesons D_1 (2420) and D_2 (2460). *CDF/PHYS/BOTTOM/CDFR/6918*, 2005. 4

[3] R. Akers et al. OPAL Collab. *Z. Phys.* ***C66****, 19.* 4

[4] Z. Albrecht et al. Delphi Collab. *DELPHI 2004-025 CONF 700.* 4

[5] Z. Albrecht. Analysis of Excited B–Mesons. *PhD-Thesis, IEKP-KA/03-16, Universität Karlsruhe, see* ***http://www-ekp.physik.uni-karlsruhe.de/***, 2003. 4

[6] M. Moch. Study of B^+–Mesons and of Excited b–Hadron Properties. *PhD-Thesis, IEKP-KA/2004-16, Universität Karlsruhe, see* ***http://www-ekp.physik.uni-karlsruhe.de/***, 2004. 4

[7] M. Kreps M. Feindt, M. Heck. Observation of orbitally excited (L=1) B_s mesons. *CDF/ANAL/BOTTOM/CDFR/8343*, 2007. 4, 5, 23, 27, 29, 30, 34

[8] C. Amsler *et al.* Review of Particle Physics. *Physics Letters B****667***, 2008. 5, 7, 25, 30, 34

[9] CDF DQM Group. Good run list. `http://www-cdf.fnal.gov/internal/dqm/goodrun/good.html`. 5

[10] C. Paus P. Murat, K. Anikeev. Description of BGeneratorII. *CDF Note 5092.* 6

[11] T. Sjöstrand et al. *Computer Phys. Commun. 135 (2001) 238 (LU TP 00-30, hep-ph/0010017).* 6

[12] CDF. Optimization of Signal Selection for $B_s \to D_s(3)\pi$ using a Neural Network. *CDF/ANAL/BOTTOM/CDFR/9751*, 2008. 6, 8, 30

[13] D. J. Lange. The EvtGen particle decay simulation package. *NIM* ***A462*** *(2001), 152*, 2008. 6

[14] C. Paus et al. see cdf code browser `http://cdfkits.fnal.gov/CdfCode/source/BottomMods/`. 6

[15] U. Kerzel. NeuroBayes – a sophisticated Neural Network package. Talk at BPAK meeting Dec. 10, 2004. 8

[16] Giovanni Punzi. Sensitivity of searches for new signals and its optimization. *arXiv:physics/0308063*, 2003. 18

[17] CDF. Measurement of the X(3872) Mass. *CDF/ANAL/BOTTOM/CDFR/9396*, 2009. 23

[18] Rene Brun and Fons Rademakers. ROOT – An Object Oriented Data Analysis Framework. *Nucl. Inst. & Meth. in Phys. Res. A*, 389:81–86, 1996. 25