CDF/ANAL/BOTTOM/CDFR/8343 Version 2.3
March 26, 2007

Observation of orbitally excited (L=1) B_s mesons

Michael Feindt, Martin Heck[1], Michal Kreps[2]

Institut für Experimentelle Kernphysik, Universität Karlsruhe

Abstract

We observe the two narrow states of the orbitally excited ($L = 1$) B_s mesons using fully reconstructed $B^+ \to J/\psi K^+$, $J/\psi \to \mu\mu$ and $B^+ \to D^0\pi^+$, $D^0 \to K^+\pi^-$ decays. In addition to the previously observed B^*_{s2} state also the B_{s1} state is observed here. A fit to the Q distribution yields $Q(B_{s1}) = 10.73 \pm 0.21(stat) \pm 0.14(sys)\ MeV$ and $Q(B^*_{s2}) = 66.96 \pm 0.39(stat) \pm 0.14(sys)\ MeV$. The obtained result allows us for the time a unique interpretation and mass assigment of the two narrow B^{**}_s states. The assigments gives masses $m(B_{s1}) = 5829.41 \pm 0.21(stat) \pm 0.14(sys) \pm 0.6(PDG)\ MeV$ and $m(B^*_{s2}) = 5839.64 \pm 0.39(stat) \pm 0.14(sys) \pm 0.5(PDG)\ MeV$.

[1]heck@ekp.uni-karlsruhe.de
[2]kreps@ekp.uni-karlsruhe.de

Documentation of changes from version 2.0 to 2.1 of the note

- At the first occurance of the explaination of the Q value in chapter 2 page 6, it is mentioned, that the reconstructed mass of the decaying particles is used, not the PDG value.
- In the description of the MC production in chapter 3 page 9 is mentioned, that the decay of B_s^{**} mesons is done according to phase space.
- As well in the description of the MC production, page 9, it is mentioned, that the PID simulation is based on the parametric simulation provided by the BottomMods package.
- In chapter 4 pages 13, 17 the definitions of purity (at first occurance of figure 6) and efficiency (at first occurance of figure 13) are made clear. They refer to the training sample.
- In chapter 4 on page 20 it is made more clear, why we use for one of the B^+ neural networks a MC based method and in the other case a sideband subtrakting method. In the di-muon channel there was already a MC available and the special structure of the two track trigger data allows to cut away a big part of the background without cutting away much of the signal. This increases the signal to background level in a way that makes the used method possible.
- In chapter 4 the individual preprocessing for each variable is written in the lists on pages 12, 16, 21 and 24.
- The parameter values after the minimazation of the likelihood function are given in chapter 5 on page 42.
- The momentum distributions for checking of possible effects due to redecaying, the number of candidates distribution and the mass distributions for the data and MC samples as given to the B_s^{**} neural networks are given in the appendix, starting at page 58

Contents

1 Introduction

Studies of excited states of hadrons play a vital role in understanding Quantum Chromodynamics, the theory of strong interaction. The excited heavy mesons plays simillar role as hydrogen atom in Quantum electrodynamics. In this analogy, the heavy quark plays role of the atomic nucleus, while light quark takes role of the electron in the hydrogen atom. Lot of experimental results are obtained in recent years for the orbital excitations with $L = 1$ of D, D_s and B mesons (see for instance [1, 2] for CDF analysis). On the contrary for orbital excitations of the B_s meson, only one of the four states was observed up to now, first by the OPAL Collaboration [3] and later confirmed by DELPHI Collaboration [4]. Recently also the DØ experiment reported the observation of the same state [5]. In all three experiments, only one narrow state was observed and it is apriory unknown which one was observed. The DELPHI Collaboration interpreted the observed signal to stem more probably from the B^*_{s2} than from B_{s1}. This interpretation is based on the width of the seen state, which was more consistent with the B^*_{s2}.

In this note we report on the search and observation of orbital excitations with $L = 1$ of the B_s mesons, which are commonly named as B^{**}_s. Observed pattern of two narrow states allows us for the first time uniquely determine the masses of the narrow B^{**}_s states.

The analysis presented here is based on the data collected by the CDF RunII detector from February 2002 to February 2006, which corresponds to a total integrated luminostity of approximatelly 1 fb^{-1} of good runs with silicon detector switched on.

The note is organized as follows. In section 2 we described basic predictions for the B^{**}_s states and current experimental knowledge. Section 3 describes the datasets used, Monte Carlo samples and candidate reconstruction. This is followed in section 4 by details of the candidate selection. In section 5 we describe the unbinned maximum likelihood fit to extract Q values of the observed states. In section 6 systematic uncertainties on the extracted Q values are evaluated. Finally in section 7 results are given and the note is closed by a short summary.

2 Theory and previous experiments

2.1 Theory

Orbitally excited mesons are mesons with angular momentum $L = 1$. Heavy quark effective theory (HQET), which is a limit of QCD for $m_q \to \infty$, predicts four states. The basic characteristics of the four state are summarized in table 1. As one can see, there are two states decaying through s-wave and two decaying through d-wave. For the states decaying through s-wave large width is expected and are usually referred as "broad" states, while other two should have small width and are referred as "narrow" states. The common name for those four states is B^{**}_s. Decays to $B_s\pi$ are forbiden

by isospin conservation. In addition the decay of $B_{s1} \to BK$ is forbidden by angular momentum and parity conservation.

j_q	J^P	B_s^{**} state	decay mode	width
1/2	0^+	B_{s0}^*	BK	broad (S-wave)
1/2	1^+	B_{s1}	B^*K	broad (S-wave)
3/2	1^+	B_{s1}	B^*K	narrow (D-wave)
3/2	2^+	B_{s2}^*	BK, B^*K	narrow (D-wave)

Table 1: Properties of the orbitally excited ($L = 1$) B_s mesons

As the B_s^{**} are mostly unknown theory predictions for their properties are rare. Despite that, there are a few predictions for the masses of B_s^{**}, which are summarized in table 2. From the table one can see that all the calculations predict mass splitings for the narrow states on the level of $12 - 13\, MeV/c^2$.

Another possible prediction is on the ratio of the branching fractions of the two possible decays of the B_{s2}^*. The prediction is based on the observation for D_2^*, where following relation is defined

$$\frac{\mathcal{B}(D_2^* \to D\pi)}{\mathcal{B}(D_2^* \to D^*\pi)} = F_c \cdot \left(\frac{k_D}{k_D^*}\right)^5 \tag{1}$$

Taking the world average for the two branching fractions from PDG [10] one gets

$$\frac{\mathcal{B}(D_2^* \to D\pi)}{\mathcal{B}(D_2^* \to D^*\pi)} = 2.3 \pm 0.6 \tag{2}$$

Heavy quark effective theory sets $F_c = F_b$ [7] and taking same formula as for D_2^* decays and PDG masses one can predict for a B_{s2}^* of the mass $5840\, MeV/c^2$ ($Q = 67$ MeV) the following ratio of branching ratios

$$\frac{\mathcal{B}(B_{s2}^* \to BK)}{\mathcal{B}(B_{s2}^* \to B^*K)} = 12.0 \pm 3.5 \tag{3}$$

Taking all the information together, in the experiment one expects two or three peaks for narrow states. The number of peaks would depend on the mass of the B_{s2}^* state. If it is on the level of $5840\, MeV/c^2$, than it dominantly decays to the BK final

state/reference	[6]	[7]	[8]	[9]
$B_{s0}^*(1/2)$	5.841	5.83	-	-
$B_{s1}(1/2)$	5.859	5.86	-	-
$B_{s1}(3/2)$	5.831	5.86	5.834	5.886
$B_{s2}^*(3/2)$	5.844	5.88	5.846	5.899

Table 2: Predictions of the masses of B_s^{**}

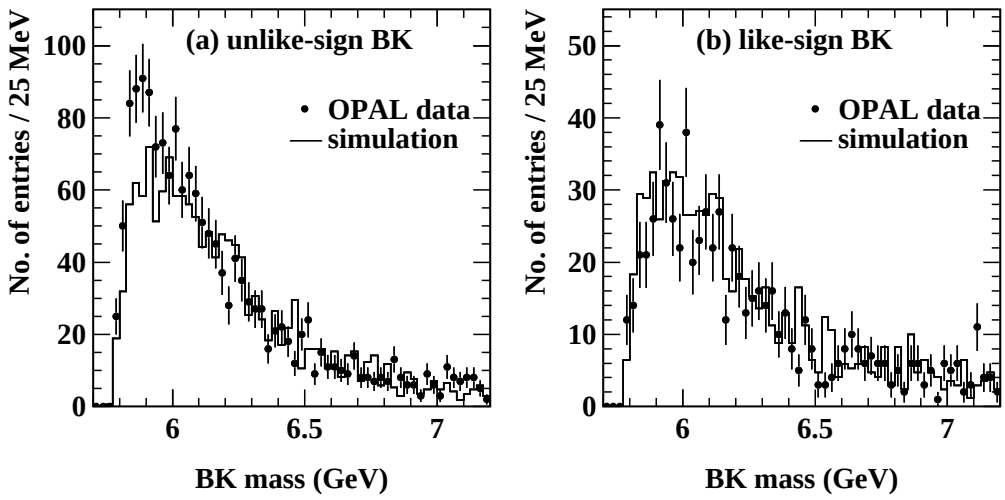

Figure 1: Invariant mass distribution for BK combinations from Opal experiment [3], a) for unlike-sign and b) for like-sign.

state. On the other hand if the mass is higher, the ratio of branching ratios of the two decay modes become smaller and the second peak corresponding to $B^{*+}K^-$ decay should appear in the BK mass distribution. The mass difference of the two peaks corresponding to the decay of B^*_{s2} will be the same as the mass difference between B and B^*, which is measured to be $\Delta M(B^* - B) = 45.78 \pm 0.35\, MeV/c^2$ [10]. The signal for B_{s1} is than expected $45.78 \pm 12\, MeV/c^2$ from peak of $B^*_{s2} \to BK$ depending which state is lighter.

2.2 Previous observations

There were several attempts to observe narrow B^{**}_s states out of which three were successful in the observation of one state.

The first observation was reported by Opal Collaboration. They look to the invariant mass of the B^+ and K^- pair using inclusively reconstructed B^+ mesons and observed an excess of 149 ± 30 events above Monte Carlo prediction of the background in the unlike-sign pair distribution [3]. The invariant mass distributions for both unlike-sign and like-sign BK combinations are shown by Figure 1. The extracted mass is $m = 5853 \pm 15\, MeV/c^2$, but without possibility to decide if the decay to BK or B^*K is seen, the measured mass is not uniquely defined.

The second experiment, which observed one of the narrow states, is the DELPHI experiment. They again use a sample of inclusively reconstructed B^+ mesons, which they combine with a charged kaon. The $Q = m(B^+K^-) - m(B^+) - m(K^-)$ distribution is shown in Figure 2 [4]. In the case of the Q value the reconstructed masses of decaying particles like the B^+ mesons is used and not the world average value. $m(B^+K^-)$ means

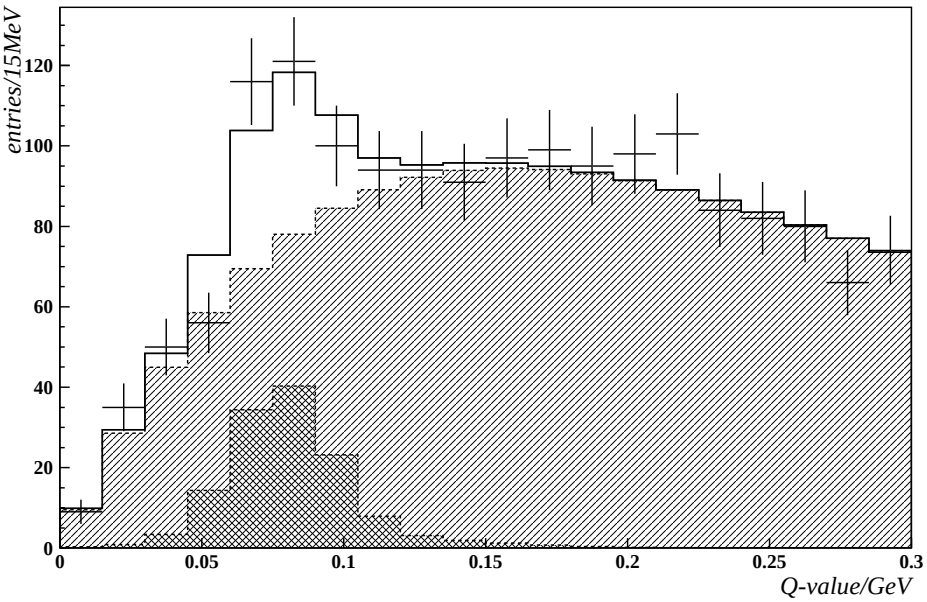

Figure 2: $Q = m(B^+K^-) - m(B^+) - m(K^-)$ distribution of the BK pair from Delphi collaboration [4].

the invariant mass of the B^+ and the K^- candidates which are combined to the B_s^{**}. A fit to the Q distribution yields a number of $134 \pm 32\,(stat) \pm 12\,(sys)$ signal events and $Q = 79.0 \pm 4.0\,(stat) \pm 2.0\,MeV/c^2$. Thanks to the improved resolution, the Delphi Collaboration comes with interpretation, that the seen signal comes from $B_{s2}^* \to BK$ decay. This would then translate to the mass of B_{s2}^* to be $5852 \pm 4\,MeV/c^2$. One should stress, that this interpretation was preferred based on the observed width in the data, but it is not excluded that the interpretation is not correct.

Recently also DØ reported an observation of the same narrow state[5] as seen by the two LEP experiments. They observe it in the B^+K^- channel using fully reconstructed B^+ mesons in the $J/\psi K^+$ channel. The Q distribution observed by the DØ experiment is shown in Figure 3. The result based on a dataset with integrated luminosity of $1\,fb^{-1}$ shows a peak in the Q distribution at $66\,MeV/c^2$. A fit with a third order polynomial for background and a Gaussian for signal yields $135 \pm 31(stat)$ signal events with $Q = 66.4 \pm 1.4(stat)\,MeV/c^2$ with resolution $\sigma = 4.7 \pm 1.5(stat)\,MeV/c^2$. The signal is interpreted as coming from $B_{s2}^* \to B^+K^-$ without deeper argumentation. This would then translate to the mass of the B_{s2}^* state being $m(B_{s2}^*) = 5839.1 \pm 1.4(stat)\,MeV/c^2$. In addition they argue that the signal from narrow the B_{s1} is not seen because it is about $20\,MeV/c^2$ lower than seen signal and therefore the decay to B^*K is not kinematically allowed.

Looking to this situation it is still not clear, which state was observed by previous experiments and what are the masses of the B_s^{**} states and mass splitting within the doublet. The CDF experiment can play vital role here with its large data sample of

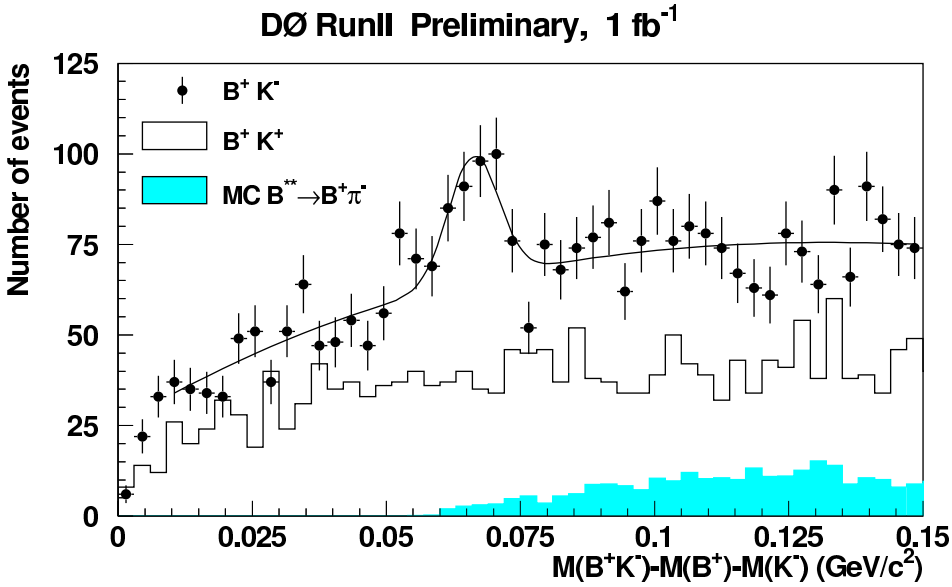

Figure 3: $Q = m(B^+K^-) - m(B^+) - m(K^-)$ distribution of the BK pair from DØ experiment [5].

fully reconstructed B^+ mesons together with excellent mass resolution and provide important information about B_s^{**} states.

3 Candidate reconstruction

3.1 Data samples

The present analysis uses events collected by the CDF RunII detector from March 2002 to February 2006. The analysis uses data from the J/ψ and Two Track triggers and selects runs based on the good run list provided by DQM group [11]. The final dataset corresponds to the total integrated luminosity of $\mathcal{L} = 1066 \pm 64\, pb^{-1}$ for J/ψ trigger and $\mathcal{L} = 976 \pm 59\, pb^{-1}$ for TTT data.

The result is based on two fully reconstructed B^+ channels. In this note charge conjugates are implied implicitly. The first decay channel is $B^+ \to J/\psi K^+$, reconstructed from the J/ψ trigger data. The second one is $B^+ \to D^0\pi^+$ with $D^0 \to K^+\pi^-$, reconstructed from two track trigger data. The corresponding datasets together with the version of the offline software used for event reconstruction are listed in the table 3. User analysis is done using BottomMods[12] package version *v6_1_1mitstn60* with version 6.1.2 of the CDF software for all the dataset.

One could possibly add also other decay channels for B^+ which are accessible in the two track trigger data, but those have at least two more pions, which means, that extracting a clean enough sample is harder and the statistics of such sample would be

J/ψ trigger		Two track trigger	
Dataset	SW version	Dataset	SW version
jpmm0d	5.3.1	xbhd0d	5.3.1
jpmm0h	6.1.1	xbhd0h	6.1.1
jpmm0i	6.1.1	xbhd0i	6.1.1

Table 3: Used datasets and CDF software version for the offline reconstruction.

much lower than statistics of the two main channels used here. Therefore, the expected gain by including them to the analysis is small.

3.2 Monte Carlo samples

In the analysis we use three different Monte Carlo sample. They are used to train a Neural Network for the selection of the B_s^{**} candidates and to choose cut on the Neural Network output. All three Monte Carlo samples are generated by ourselves and are accesible via SAM to the whole collaboration.[3] In all cases we use the EvtGen[13] module to decay B-hadrons and a full CDF detector simulation, trigger simulation and reconstruction is performed. Three samples are for $B^+ \to J/\psi K^+$, $B_s^{**} \to B^+K^-$ with $B^+ \to J/\psi K^+$ and $B_s^{**} \to B^+K^-$ with $B^+ \to D^0\pi^+$.

The first Monte Carlo sample is PYTHIA[14] Monte Carlo with MSEL=1 events for $B^+ \to J/\psi K^+$ decay. As an input we use existing MSEL=1 PYTHIA events in the sample nbot90, which we redecay by EvtGen. During redecay, one of the B^+ in the event is forced to decay to the selected channel, while all others are decayed according the default decay table. The sample is stored in the SAM under the name fzkbu1.

The samples for the B_s^{**} decays are generated using BGenerator[15] to generate B_s^{**}. For decay tables used, see appendix B and C. After generating them, the mass and the energy of generated B_s^{**} is changed to have uniform distribution in mass starting from B^+K^- threshold up to values larger than any expectations. This is done to have a sample, which allows us to analyse data and train Neural Networks in an unbiased way, as many quantities, which could discrimanate signal from background will depend on the exact mass of the B_s^{**} states. Generated B_s^{**} are than decayed to B^+K^- and $B^{*+}K^-$ with $B^{*+} \to B^+\gamma$ according to phase space. B^+ itself is decayed to the selected channel. In both cases appropriate trigger simulation is performed. In SAM, they are accessible under the names fzkbsds1, fzkbsds2, fzkbsds3.

For all Monte Carlo samples we use the parametric particle ID simulation implemented in the BottomMods package.

[3] Currently those samples are stored only at station cdf-fzkka, which is at GridKa Karlsruhe. This cluster is not in the DCAF system, so people cannot run over it on this cluster, which would mean copying it to other place. This should change in near future as we are working on GlideIn interface at GridKa.

3.3 Candidate reconstruction

The reconstruction of our candidates starts with refitting of tracks with Layer00 hits included. This is done separately for pion, kaon and muon mass hypothesis to properly take to account energy loss of a given particle. The scaling of the COT covariance matrix, measurement of the magnetic field and material description are done according to the measurement[16]. The reconstruction of data from J/ψ trigger was done by ourselves, while for two track trigger data we use BStntuples produced by the CCKMP group. As the decision to add $B^+ \to D^0\pi^+$ channel came latter, there are slight differences in the two reconstructions. Those differences are checked and have no strong effect on our result.

For all tracks in the J/ψ trigger we require to have 10/10 hits in the axial and stereo layers of the COT and at least 2 axial hits in the SVX II. This requirement is not used in TTT data and as we don't see difference when doing it in final steps of our analysis, we use all tracks without the requirement on the number of hits.

In the next step, the J/ψ and D^0 are reconstructed in corresponding datasets. For J/ψ all muon combinations of opposite charge are constrained to originate from a common vertex using the CTVMFT[17] vertex fitter. Candidates with mass between 2.9 and 3.3 GeV and χ^2 less than 30 are kept for further analysis. D^0 is formed from K and π of opposite charge again constrained to come from common vertex. Candidates with mass between 1.77 and 1.97 GeV, χ^2 less than 15 and a L_{xy} significance at least 3 are used in next steps.

J/ψ candidates are now combined with kaons and all three tracks are fitted to the common vertex to form a B^+. In the vertex fit, the J/ψ mass is constrained to the PDG value. All candidates with χ^2 less than 50 and mass between 4.6 and 6.8 GeV are stored for the next step. In TTT data, each D^0 candidate is combined with a pion and a vertex fit with the D^0 mass constraint is performed. B^+ candidates in the $D\pi$ channel are stored if the mass is between 4.4 and 6.6 GeV, the χ^2 of the fit is less than 25 and the L_{xy} significance is at least 4.

In the final step for the $B^+ \to J/\psi K^+$ channel, each B^+ candidate is combined with an additional kaon to form a B_s^{**} candidate. A vertex fit is performed and all candidates with $\chi^2 < 80$ and mass up to 10 GeV are stored on the Ntuple for further analysis. For the decay channel $B^+ \to D^0\pi^+$, this step is not done, but rather on the BStntuple level the existing B^+ candidates are combined with the another track from the pion collection, to which kaon mass is assigned and invariant mass is calculated from the four momenta of the track and B^+ candidate. The pion collection is used because on BStntuple all tracks are stored only in the pion collection. This difference in the reconstruction causes slightly worse resolution in the TTT data comparing to the J/ψ trigger data as will be shown later. This price has to be paid in exchange of speeding up the analysis as running over BSTntuples needs much less time than running over the full TTT datasets.

To get a feeling, what difference we can expect from the two ways of reconstructing B_s^{**} candidates, in Figures 4 and 5 we show the expected resolution derived from

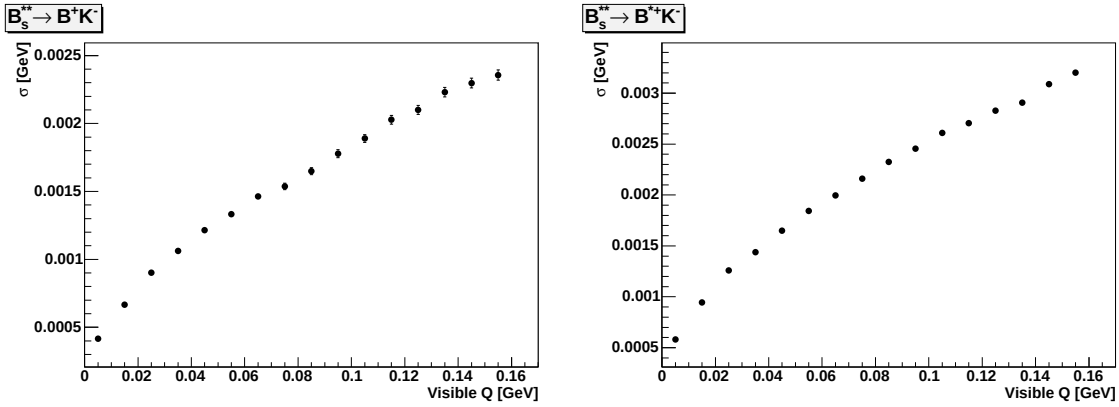

Figure 4: Expected resolution from Monte Carlo for the J/ψ trigger data. The two plots show the case where B_s^{**} decays directly to B^+ and the case, where decay goes through B^{*+}, where we miss a γ.

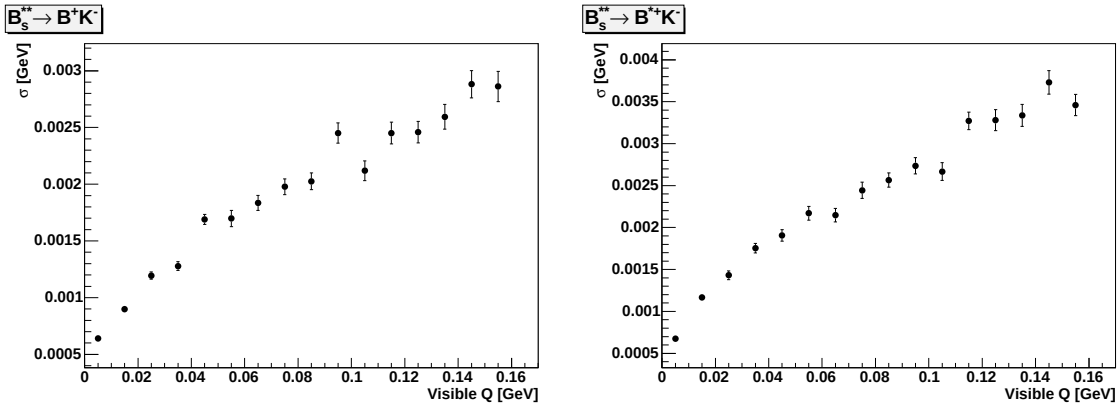

Figure 5: Expected resolution from Monte Carlo for the two track trigger data. The two plots show the case where B_s^{**} decays directly to B^+ and the case, where decay goes through B^{*+}, where we miss a γ.

Monte Carlo simulation. In both cases, we show the direct decay to B^+K^- and decay to $B^{*+}K^-$ as a function of Q value at which the decay would be reconstructed. The effect of the different reconstruction in the two track trigger data is clearly visible.

4 Candidate selection

In this section we describe the candidate selection. In order to select B_s^{**} candidates we use two quantities, which are the number of candidates n_{cands} and the output of the Neural Network trained to select B_s^{**} from background. The selection is done separately for the two B^+ channels. To preselect B^+ candidates, we use additional Neural Networks. In all cases we use $NeuroBayes^{©}$[18] package to build the Neural Networks. Some more information about usage of $NeuroBayes^{©}$ for the signal selection can be found in CDF Notes [19, 20].

First, we describe the four Neural Networks to select B^+ and B_s^{**} candidates, then we derive cuts to select B_s^{**} candidates. At the end of the section, we check, that peaks, which are visible after the selection are stable against variation of the selection and binning of the Q distribution.

4.1 Neural Networks for the J/ψ trigger data

4.1.1 Preselection of the $B^+ \to J/\psi K^+$ decay

For the preselection of the $B^+ \to J/\psi K^+$ decay a Neural Network is trainened on the combination of data and Monte Carlo events. In the training two classes of events are needed, one for the background and another one for the signal. The signal events are taken from the PYTHIA Monte Carlo with full CDF detector and trigger simulation. The simulation of the combinatorial background is generally a complicated task in the hadronic enviroment and therefore events from the data sidebands are used as training patterns 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 for the background patterns in the Neural Network training. In terms of the invariant mass resolution, those corresponds approximatelly to range from $3\,\sigma$ to $7\,\sigma$.

The variables in the list below are combined in the Neural Network for discriminating between B^+ events and combinatorial background. The number in brackets gives the preprocessing number and an eventually preprocessing parameter. The meaning of this is expalined in [21]

1. the impact parameter of the B^+ (14)
2. the fit-probability of the B^+ vertex fit (15)
3. the significance of transverse decay length of the B^+ (15)
4. the transverse momentum of the J/ψ , (14)

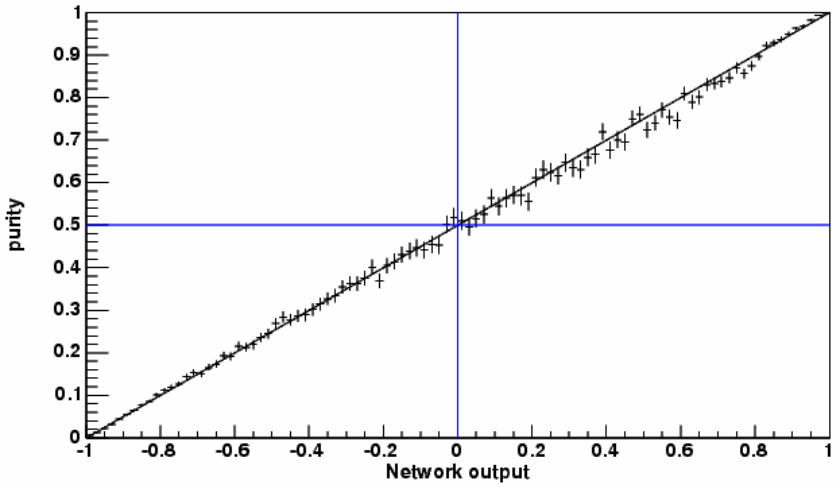

Figure 6: Purity as a function of the Neural Network output for the $B^+ \to J/\psi K^+$ Neural Network using the training sample.

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 kaon in center-of-mass frame of the B^+ relatively to momentum of 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 (14/9)

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.

From the Figure 6 where the purity as a function of the Neural Network output is shown, one sees, that the Neural Network is well trained as the points lie on the diagonal. The purity is defined here as the fraction of MC candidates of the training sample which is in the selected sample divided by the number of all candidates of the training sample, MC as well as data, in the selected sample. The Figure 7 shows

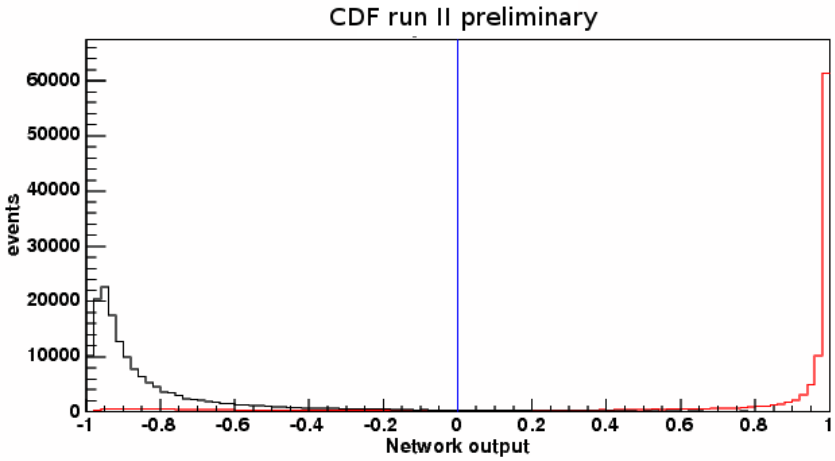

Figure 7: Neural Network output distribution for signal (red) and background (black) events in the $B^+ \to J/\psi K^+$ decay, using the training sample.

the Neural Network output distribution for the signal and background events using the training sample and the fact, that the Neural Network is able to distinguish well between signal and background events.

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 classification is shown in Figure 8. The number of signal events in the sample is ≈ 38700. After selected events with the Neural Network output larger than 0.5 we retain ≈ 31400 signal candidates while 98.5 % of the background events are removed. The resulting invariant mass distribution is shown in Figure 9. This cut is not further optimized as the main goal of this cut is to remove obvious background from next step while keeping most of the signal, which is achieved by this cut.

4.1.2 Selection of the B_s^{**} mesons

The selection of the B_s^{**} candidates starts first with soft precuts, which remove a large fraction of the background while keeping most of the signal. First precut is a cut at 0.5 on the Neural Network output of the B^+ preselection Neural Network. In addition only events with at most two right signed candidates and at most two wrong sign candidates are kept for the network training. The last precut selects events, where the kaon track from the B_s^{**} decay has no particle identification (PID) information or if it has PID information, that the corresponding likelihood ratio, that the track is a kaon has to be at least 0.2. This cut is designed to remove the candidates, where we are rather sure, that the track is not a kaon.

Again a combination of data and Monte Carlo simulation is used to train a Neural Network. Monte Carlo events are used as signal patterns while data events are used as background patterns. The events with a Q value from 0 to 200 MeV are used. As

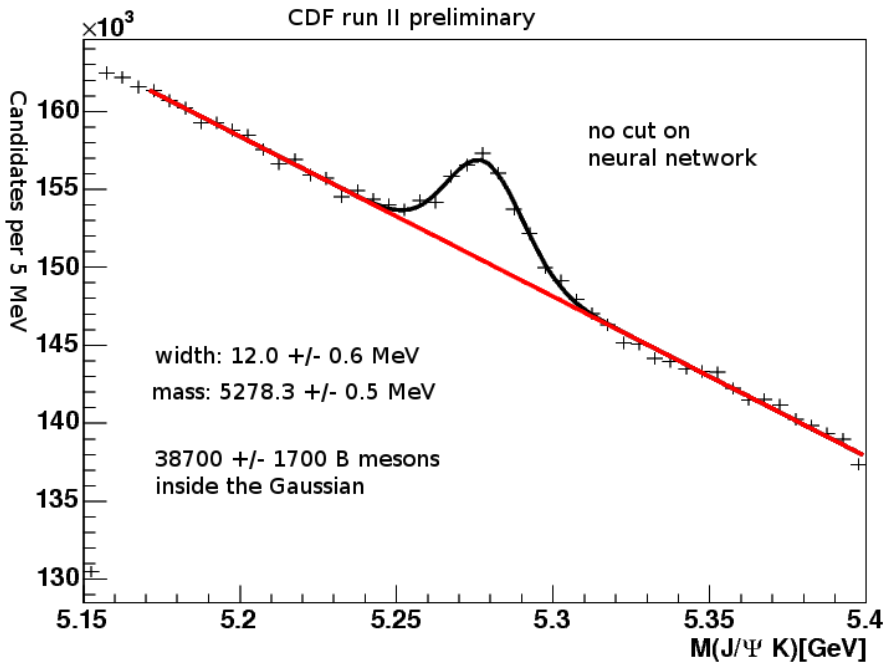

Figure 8: Invariant mass distribution of the B^+ candidates before preselection. The fit in the Figure has a single Gaussian for the signal and a first order polynomial for the background.

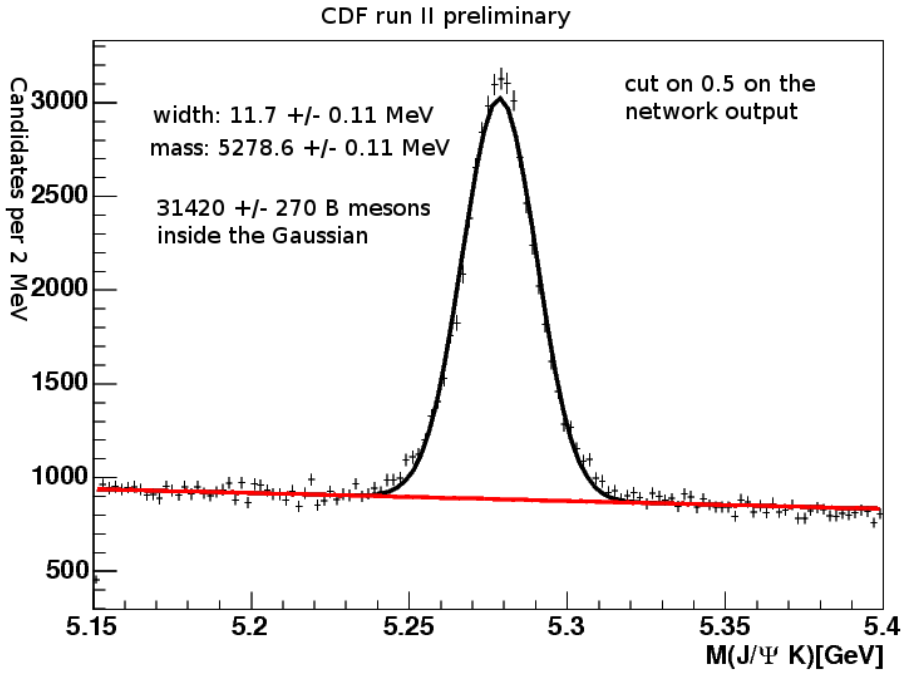

Figure 9: The invariant mass distribution of the B^+ candidates after a cut on the Neural Network output at 0.5. The fit in the Figure has a single Gaussian for the signal and a first order polynomial for the background.

the masses of the B_s^{**} states are unknown and discriminating variables depends on the mass, we generate the Monte Carlo with uniform mass distribution and reweight it for the training to the mass distribution of the background sample. In this way, we make sure, that the Neural Network cannot learn based on the mass of the candidate.

In the Neural Network the following variables are combined to discriminate B_s^{**} from background, where the number behind the variable means the used preprocessing and additional preprocessing parameters:

1. the impact parameter of the B^+ (14)
2. the fit-probability of the B^+ vertex fit (34)
3. the significance of the transverse decay length of B^+ (15)
4. the impact parameter of J/ψ (15)
5. the significance of transverse decay length of the J/ψ with respect to the beamline (15)
6. the transverse decay length of J/ψ with respect to B^+ decay vertex (15)
7. the transverse momentum of the kaon from the B^+ decay (15)

8. the distance of the mass of the B^+ from world avarage B^+ mass (5.279 GeV) (15)
9. the fit-probability of the B_s^{**} vertex fit (15)
10. the transverse momentum of the B_s^{**} (15)
11. the opening angle of the B^+ and the kaon from the B_s^{**} decay in ϕ in the laboratory frame (15)
12. the cosine of the angle of the kaon from the B_s^{**} decay in center-of-mass frame of the B_s^{**} relatively to momentum of B_s^{**} in laboratory frame (15/3)
13. the neural network output from the network for selecting B^+ mesons (15)
14. the transverse momentum of the kaon from the B_s^{**} decay (15)
15. the significance of the impact parameter of the kaon from the B_s^{**} decay (15)
16. the impact parameter of the kaon from the B_s^{**} decay (15)
17. the PID ratio of the Kaon from the B_s^{**} decay in the case there is TOF information (94)
18. the PID ratio of the Kaon from the B_s^{**} decay in the case there is no TOF information (94)
19. the absolute value of the pseudorapidity of the kaon from the B_s^{**} decay (15/7)
20. the mass of the B_s^{**} candidate (15)

The first seven variables are already used in the B^+ preselection Neural Network. Here we use them again to allow the Network to learn possible small differences between B^+ from B_s^{**} decay and from other sources. The mass of the B_s^{**} candidate is included as an input variable to get better handle on possible effects of the mass of the candidate. In the Neural Network, which we use, this variable is pruned away by the Network as an unimportant variable uncorrelated to the target. We also check, that this is not only due the information being available from the other variables. This can be seen from the color coded correlation matrix of the variables used in the Neural Network, which is shown in Figure 10.

In Figure 11 the purity as a function on the Neural Network output is shown. Again, the Network is well trained. The distributions of the Neural Network outputs for the signal and background events are shown in Figure 12. Here it is clear, that the Network doesn't have as large separation power as for the B^+ preselection. This statement is even more clear looking to figure 13. It shows the purity vs. the efficiency, which is defined by the number of candidates of the MC in a sample selected by a cut on the neural nerwork on the MC training sample devided by the number of candidates in the whole MC training sample. The Q distributions together with the determination of the working cut are described later in the section 4.3.

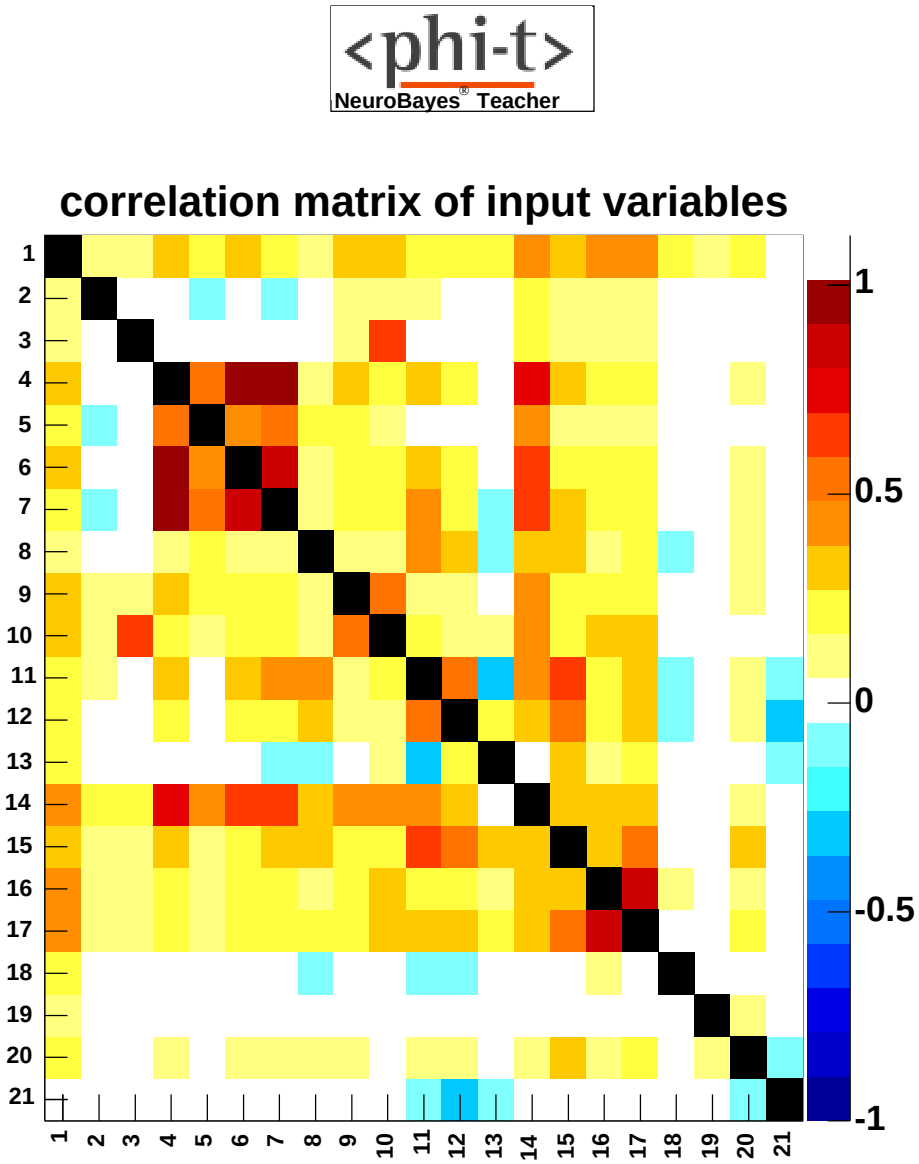

Figure 10: Correlation matrix of the Neural Network for selection of the B_s^{**} in the J/ψ trigger data. First row corresponds to the target (was it signal or background). The rest of the lines corresponds to the used variables. Here variable n from the variable list corresponds to the row $n+1$ in this figure.

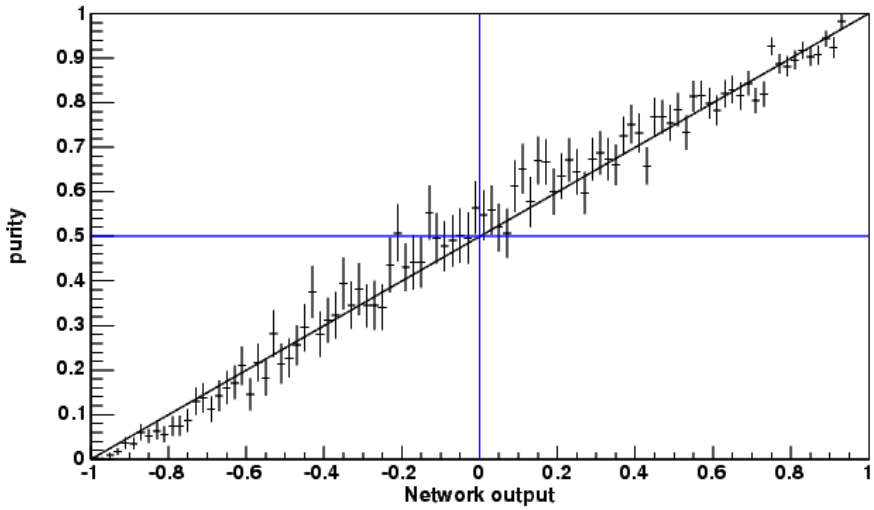

Figure 11: Purity as a function of the Neural Network output for the $B_s^{**} \to B^+K^- \to J/\psi K^+K^-$ Neural Network using the training sample.

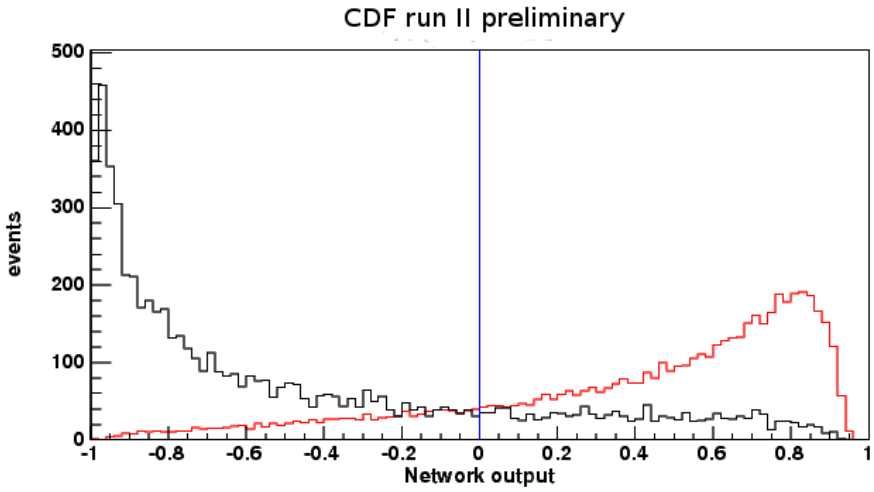

Figure 12: Neural Network output distribution for signal (red) and background (black) events in the $B_s^{**} \to B^+K^- \to J/\psi K^+K^-$ decay, using the training sample.

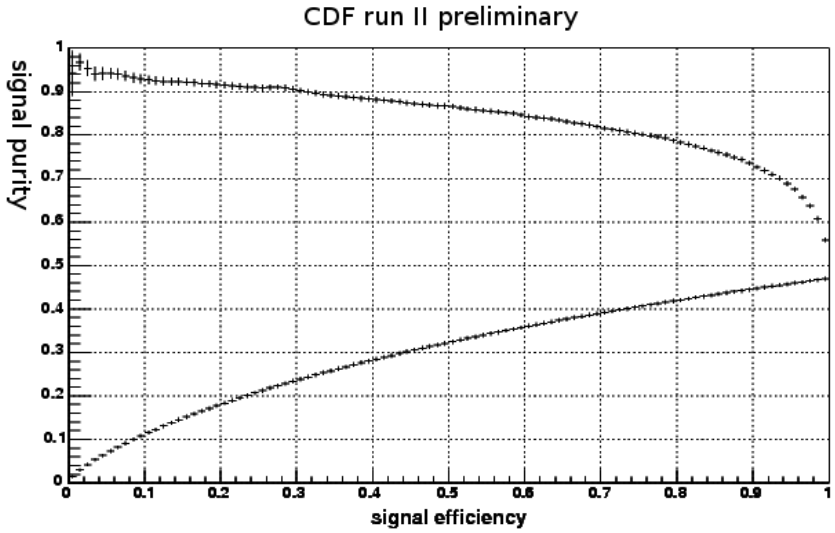

Figure 13: Purity versus efficiency dependence for the $B_s^{**} \to B^+K^- \to J/\psi K^+K^-$ Neural Network estimated from the training sample obtained by different cuts on the Neural Network output. The upper points in the graph corresponds to the events, which have Neural Network output larger than a given cut while lower ones correspond to events with the Neural Network output lower than the cut.

4.2 Selection in the two track trigger data

4.2.1 Preselection of the $B^+ \to \overline{D}\pi^+$ decay

To preselect the B^+ mesons in this decay channel we use a different strategy, since we do not have a MC sample. It is possible to train a neural network with sideband subtraktion, if the amount of background compared to signal is not too high and there is much statistics.
A high statistic is in both B^+ channels available, but the special structure of the two track trigger allows to remove a rather big part of the background without loosing much of the signal, since most B^+ mesons reconstructed in the two track trigger data have a secondary vertex distinguishable from the primary vertex. So we can come along without a MC sample.
The idea for the sideband subtrakted training originates from the fact, that the *NeuroBayes*© package allows a training of the Neural Network 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 patterns, the background region is used with weight -1[4] as signal patterns and the background

[4]The number is symbolical, the actual value is choosen such that it subtracts proper amount of background.

region is used once again with weight 1 as the source of background patterns.

In order to make background subtraction more stable it is usefull to supress background by some precuts, which would remove only marginal part of the signal. The precuts, which are done mostly comes from the context of the mixing analysis where they are used to skim BStntuples and are:

- 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 $\overline{D}$ has to be higher than 0.0025 cm.
- The significance of the transverse decay length of the $\overline{D}$ 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.
- The charge of the pion from the $\overline{D}$ decay has to be opposite to the charge of the pion from the B^+ decay. Striclty speaking, they can be same, but in the wrong charge sample, there is a huge background with negligable amount of signal.

The following variables are combined in the Neural Network:

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

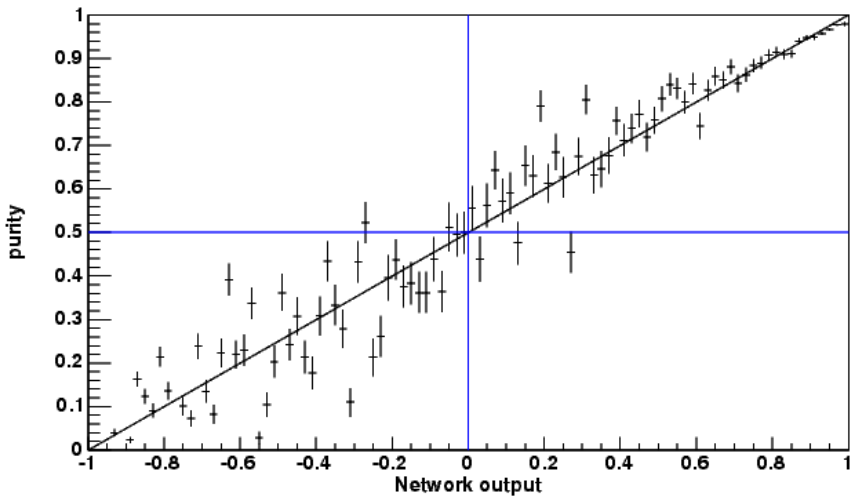

Figure 14: Purity as a function of the Neural Network output for the $B^+ \to \overline{D}\pi$ Neural Network using the training sample.

11. the cosine of the angle of the pion from the B^+ decay in the CMS 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 (4)

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 Figure 14 showing purity as a function of the Neural Network output tells us, that the Neural Network is reasonably well trained. From Figure 15 we could conclude, that the separation power of the Neural Network is good. The purity versus the efficiency is shown in Figure 16. In all three figures, the training sample is used. In Figure 17 the invariant mass distribution of the B^+ candidates before a cut on the Neural Network output is shown while in Figure 18 after a cut of -0.2 on the Neural Network output. From the simple fit with a single Gaussian for the signal and a linear background we found, that we keep ≈ 27000 signal events out of the ≈ 28000. On the other hand only 36 % of the background events survived this cut.

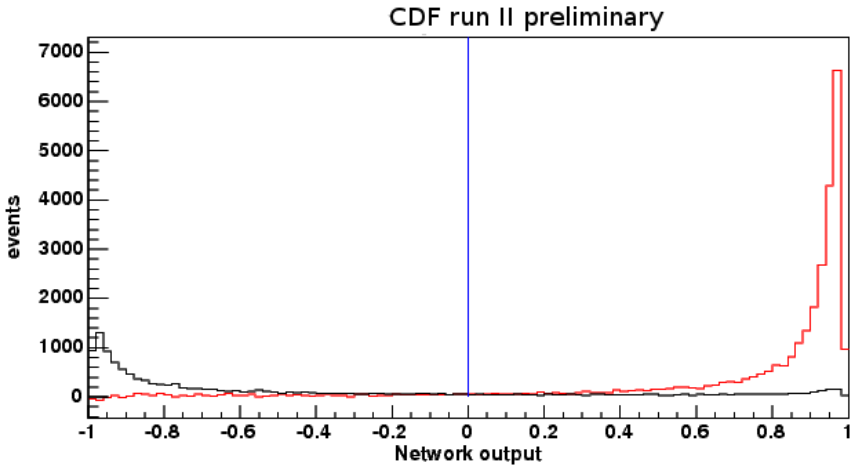

Figure 15: Neural Network output distribution for signal (red) and background (black) events in the $B^+ \to \overline{D}\pi$ decay, using the training sample.

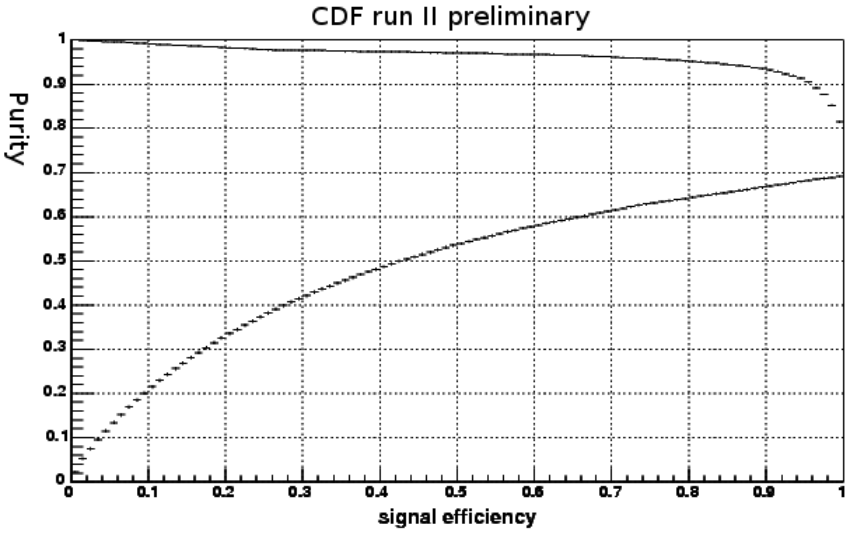

Figure 16: Purity versus efficiency dependence for the $B^+ \to \overline{D}\pi$ Neural Network estimated from the training sample obtained by different cuts on the Neural Network output. The upper points in the graph corresponds to the events, which have Neural Network output larger than a given cut while lower ones correspond to events with the Neural Network output lower than the cut.

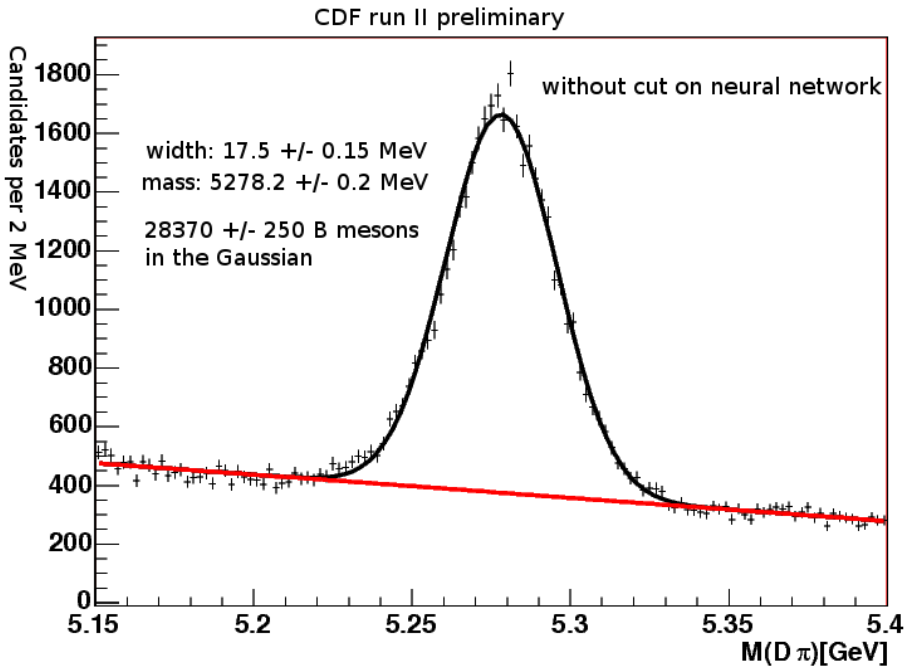

Figure 17: Invariant mass distribution of the $B^+ \to \overline{D}\pi$ candidates before preselection. The fit in the Figure has a single Gaussian for the signal and a first order polynomial for the background.

4.2.2 Selection of B_s^{**} mesons

The selection of the B_s^{**} in the decay to B^+K^- with $B^+ \to \overline{D}\pi$ first starts with some more preselection. The cuts made at this stage are:

- the transverse momentum of the pion from the B^+ decay has to be higher than 1.0 GeV
- the transverse momentum of the B^+ has to be higher than 5.0 GeV
- if the PID information is available for the kaon from the B_s^{**}, than the corresponding likelihood ratio has to be higher than 0.2
- the mass of the B^+ to be between 5.23 GeV and 5.34 GeV
- the output of the Neural Network trained for $B^+ \to \overline{D}\pi$ decay has to be higher than -0.2

After those precuts, we train the Neural Network to discriminate between B_s^{**} signal and background using Monte Carlo events as training patterns and data with a Q value in the range from 0 MeV to 200 MeV as background patterns. The Monte Carlo events are again reweighted to have the same invariant mass distribution as the events in the data to avoid bias in the Network. The variables combined in the Neural Network are:

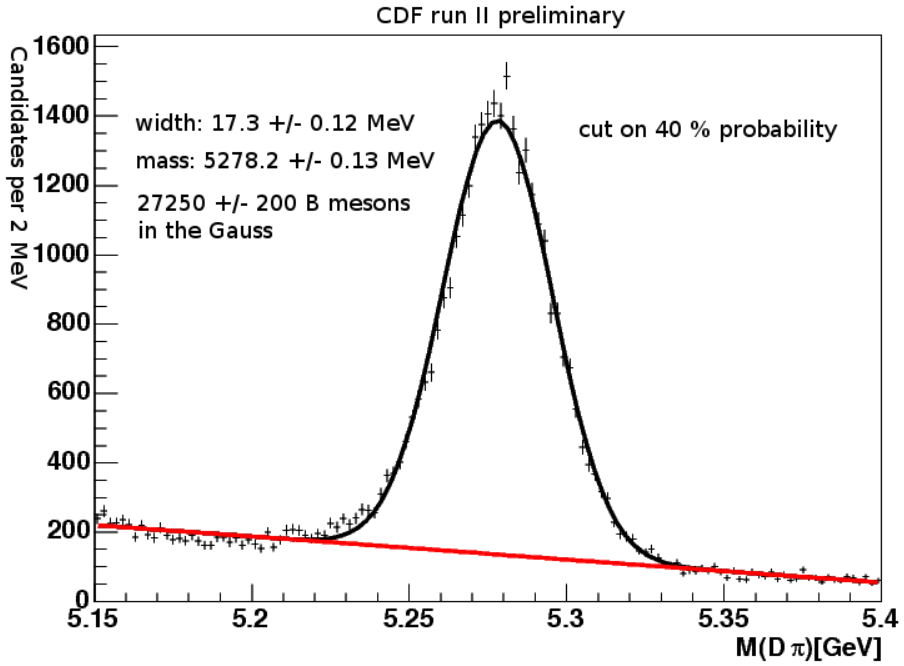

Figure 18: Invariant mass distribution of the $B^+ \to \overline{D}\pi$ candidates with the cut on Neural Network output of -0.2. The fit in the Figure has a single Gaussian for the signal and a first order polynomial for the background.

1. the absolute value of the impact parameter of the B^+ (14)
2. the fit probability of the B^+ vertex fit (34)
3. the significance of the transverse decay length of the B^+ (14)
4. the absolute value of the the impact parameter of the $\overline{D}$ (14)
5. the fit probability of the $\overline{D}$ vertex fit (34)
6. the significance of the transverse decay length of the D meson with respect to the primary vertex (15)
7. the significance of the transverse decay length of the $\overline{D}$ meson with respect to the decay vertex of the B^+ (14)
8. the transverse momentum of the pion from the B^+ decay (15)
9. the significance of the impact parameter of the pion from the B^+ decay with respect to the B^+ decay vertex (15)
10. the cosine of the angle of the pion from the B^+ decay in the cms of the B^+ relatively to the momentum of the B^+ in the laboratory frame (15)

11. the cosine of the helicity angle of pion from the D decay (15)
12. the transverse momentum of the kaon from the $\overline{D}$ decay in the case it is smaller than 2.0 GeV/c (95)
13. the transverse momentum of the kaon from the $\overline{D}$ decay in the case it is larger than 2.0 GeV/c (95)
14. the distance of the mass of the B^+ from world avarage B^+ mass (5.279 GeV) (14)
15. the transversal momentum of the B_s^{**} (15)
16. the opening angle of the B^+ and the kaon from the B_s^{**} decay in ϕ (14)
17. the cosine of the angle of the kaon from the B_s^{**} decay in the cms of the B_s^{**} relatively to the momentum of the B_s^{**} in the laboratory frame (15)
18. the neural network output from the network for selecting $B^+ \to \overline{D}\pi$ mesons (15)
19. the transverse momentum of the kaon from the B_s^{**} decay (15)
20. the impact parameter of the kaon from the B_s^{**} decay (15)
21. the impact parameter significance of the kaon from the B_s^{**} decay (14)
22. the PID likelihood ratio of the kaon from the B_s^{**} decay in the case there is time of flight information (95)
23. the PID likelihood ratio of the kaon from the B_s^{**} decay in the case there is no time of flight information (95)
24. the mass of the B_s^{**} (15)

The first thirteen variables are used already in the B^+ preselection Neural Network and here are used again to allow the Network to learn possible small differences between B^+ from B_s^{**} and from other sources. Also in this Neural Network we use the mass of the B_s^{**} to control, that the Neural Network doesn't learn anything about the mass. In the training process it is pruned away. Also from the correlation matrix shown in Figure 19, it is clear, that the mass of the B_s^{**} doesn't contain any information.

Again from Figure 20 it is clear that the Neural Network is well trained and from Figure 21 that the Neural Network has a reasonable separation power. The same comment as for the B_s^{**} selection Neural Network in the J/ψ trigger data is valid also here, namely, that the mass independence is much more important than the best possible separation. The Q distribution of the B_s^{**} selected candidates together with the cut derivation is the content of next section.

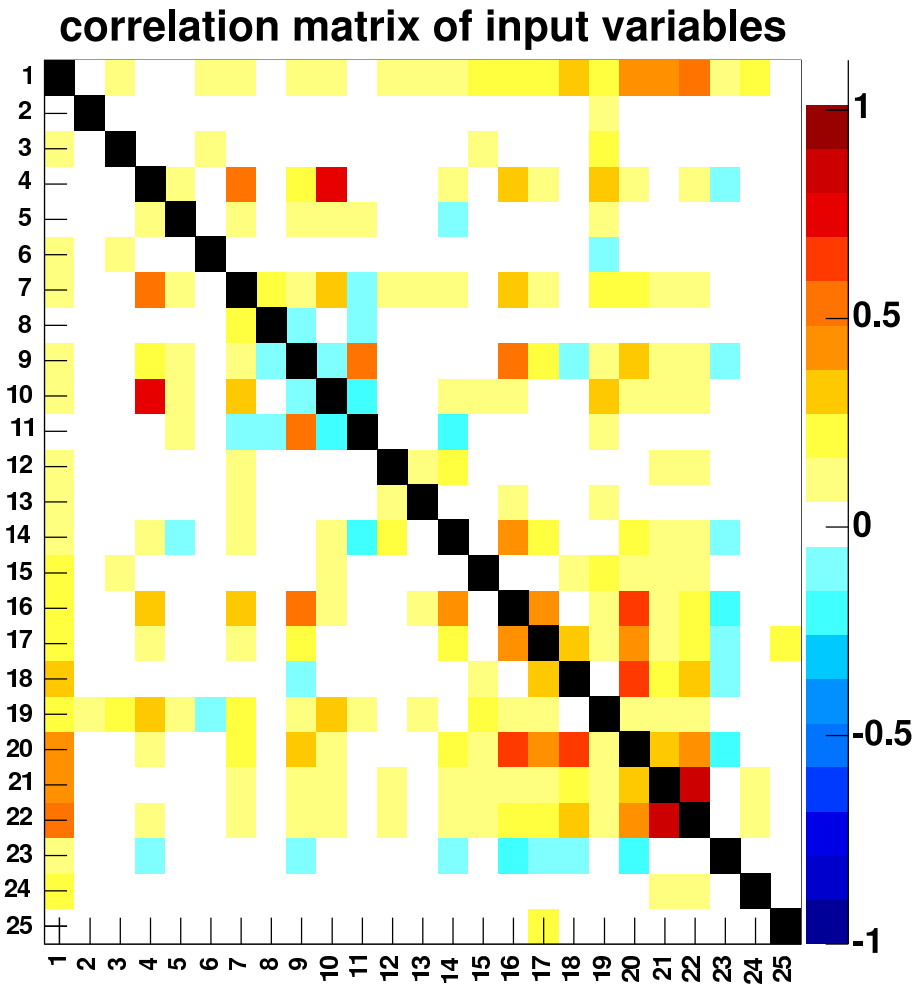

Figure 19: Correlation matrix of the Neural Network for selection of the B_s^{**} in the two track trigger data. First row corresponds to the taget (was it signal or background). The rest of the lines corresponds to the used variables. Here variable n from the variable list corresponds to the row $n+1$ in this figure.

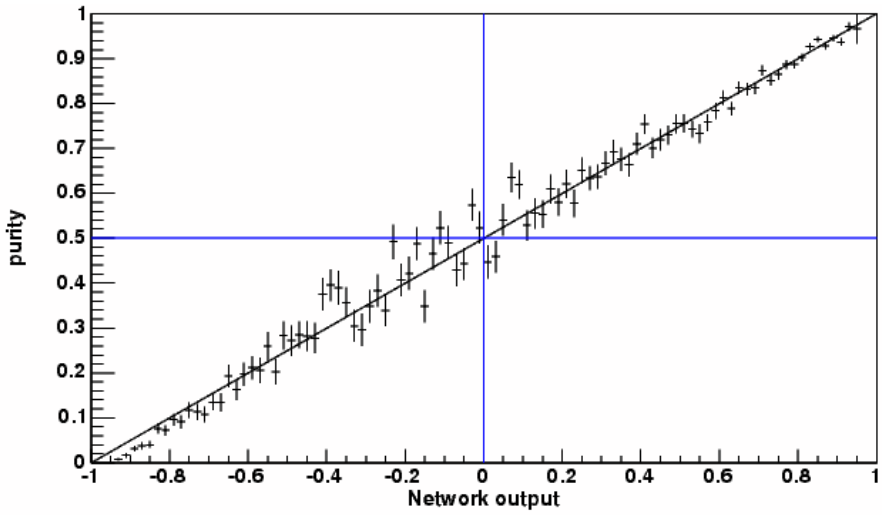

Figure 20: Purity as a function of the Neural Network output for the $B_s^{**} \to B^+K^- \to \overline{D}\pi$ Neural Network using the training sample.

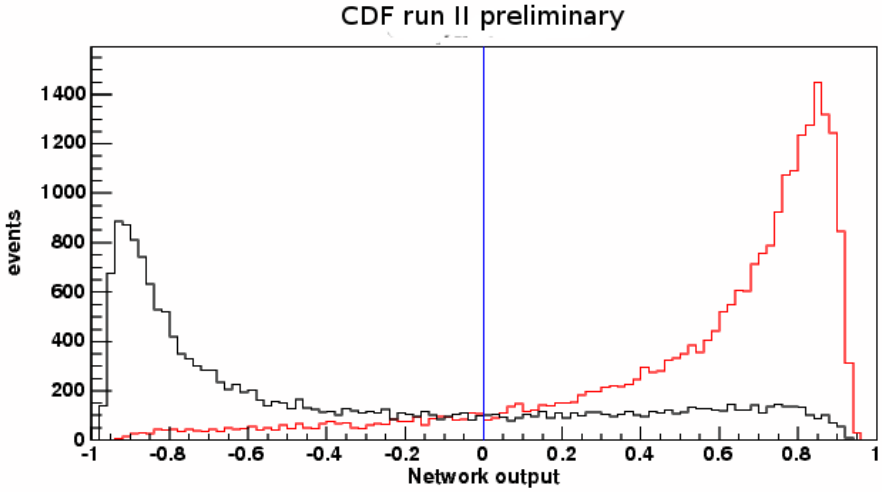

Figure 21: Neural Network output distribution for signal (red) and background (black) events in the $B_s^{**} \to B^+K^- \to \overline{D}\pi$ decay, using the training sample.

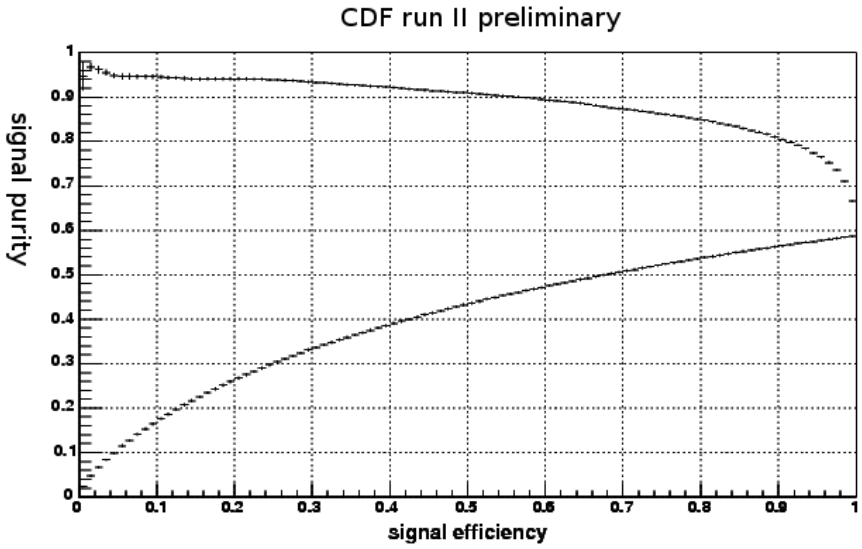

Figure 22: Purity versus efficiency dependence for the $B_s^{**} \to B^+K^- \to \overline{D}\pi$ Neural Network estimated from the training sample obtained by different cuts on the Neural Network output. The upper points in the graph corresponds to the events, which have Neural Network output larger than a given cut while lower ones correspond to events with the Neural Network output lower than the cut.

4.3 Cut optimization

As already mentioned in the introduction to this section, we cut on two quantities, the number of candidates in the event and the output of the Neural Network. The number of candidates in the event was taken out from the variables used in Neural Networks as it is challenging to model it properly. The main aim of using this quantity comes from the fact, that in most of the events at most one real B_s^{**} is produced. Therefore if we have a small number of candidates the signal to background ratio is relativily high. On the other hand if we have lot of candidates and only one comes from B_s^{**} we get a very poor signal to background ratio. Unfortunatelly as we use BGenerator Monte Carlo, we cannot model this quantity even in approximate way and have to decide where to cut without having the possibility to make an unbiased study. In order not to create bias in our selection we decide on the value at which we could cut empirically based on the arguments before rather than by looking to data. We decided to use only events in which there are less than four candidates. This decision is done before we extract any information from the Q distribution. Also as will be shown in the next subsection, varying this cut doesn't change the Q distribution significantly.

The cut on the Neural Network output can be choosen using a combination of our Monte Carlo samples and data. As the aim is to observe new states the best quantity on which one could make a decision is a quantity which has a direct connect to the

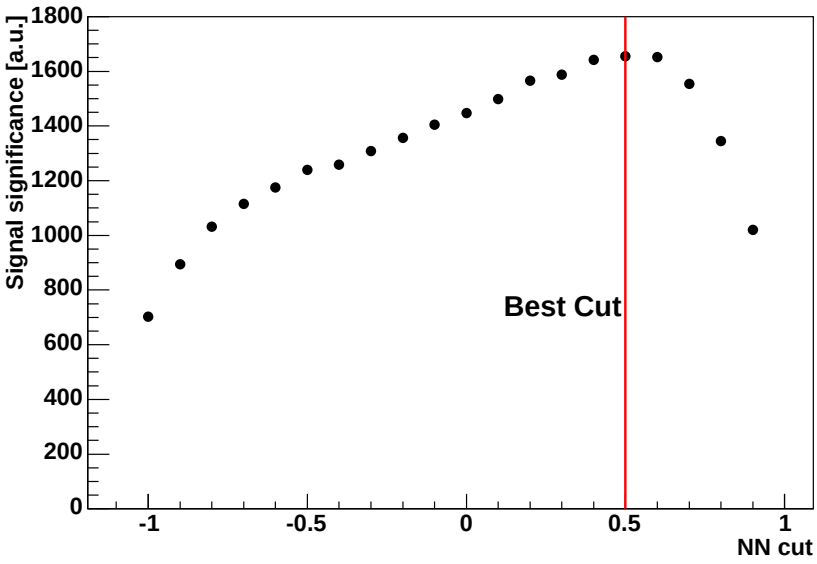

Figure 23: Scan of Σ over the cut on Neural Network output for J/ψ data.

significance. The quantity choosen is

$$\Sigma = N^{MC}_{cand}(nn)/\sqrt{N^{Data}_{cand}(nn)} \tag{4}$$

where $N^{MC}_{cand}(nn)$ is the number of candidates selected with a given cut in the Monte Carlo while $N^{Data}_{cand}(nn)$ is the number of candidates selected in the data. The number of candidates selected in the Monte Carlo is directly connected to the number of signal events while the number of candidates selected in the data gives sum of the real background and real signal independent of their ratio. The best cut on the Neural Network output is then the cut, which maximazes Σ. As the two Neural Networks for the two B^+ channels are independet optimization is performed independently for each channel. To perform this optimization we have to decide on the Q window in which events will be counted. The Q window selection is rather arbitrary and as the Neural Network efficiency is a function of the Q value we decided to use a window which would be somewhere in the middle of the search range. The search range can be defined roughly as a region starting from threshold ($Q = 0$) till around $Q = 150$ MeV. In order not to bias selection we decided to use a window between $Q = 60$ MeV and $Q = 70$ MeV, which containes the relatively established narrow state as this is the only place in the Q distribution where one knows what is inside and the definition of Σ has a very clear meaning there. The main issue coming from the choosen window is the fact, that we could bias the significance of the peak, which is located in the selected window. This is checked later in section 6.3 and no bias is found to be present.

In Figures 23 and 24 the scans of Σ over the Neural Network output are plotted for the J/ψ and two track trigger data respectively. From those scans we obtain the best

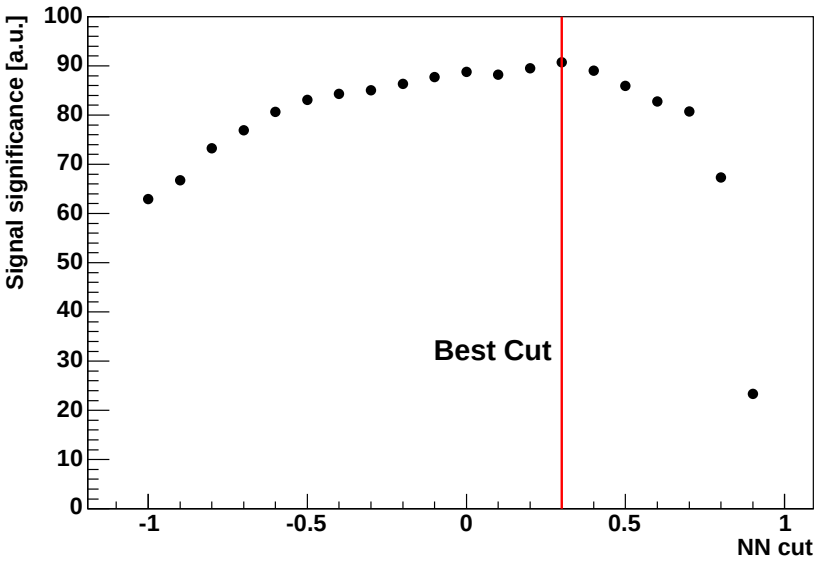

Figure 24: Scan of Σ over the cut on Neural Network output for two track trigegr data.

cut on the Neural Network output to be 0.5 for the J/ψ trigger data and 0.3 for the two track trigger data.

4.4 Selection stability

Here we make a check that the data are stable in a sense, that if there are some peaks visible they are not visible only at choosen cuts or only with a special binning.

In Figures 25 and 26 the Q distributions with different cuts on the Neural Network output are shown for J/ψ and two track trigger data. In both cases, in addition to the cut on the Neural Network output we demand that there are less than four candidates in the event. For the J/ψ trigger data in all shown cases two clean peaks are visible. For the two track trigger data, the situation is a little bit worse, mainly due to the worse Q value resolution, but again peaks are visible at the same Q values as in the case of J/ψ trigger data. The fact, that we see the same structure in the two independent decay channels with two independent triggers gives us first confidence, that we observe real signal and not only fluctuation.

In the next step we cross checks that what was seen in the previous figures is not just the artefact of the choosen binning. To do this, we plot Q distributions with different bin widths starting always from $Q = 0$. The three values for bin width were choosen to be 1.25 MeV, 1.5 MeV and 2 MeV. The Q distribution with different binnings for the best cut on the Neural Network output is shown on Figures 27 and 28 for the two triggers. Again the two peaks are visible with all three bin widths.

The last of the stability checks is to check the effect of the cut on the number of candidates in the event. To make this check, we examine Q distributions with the best

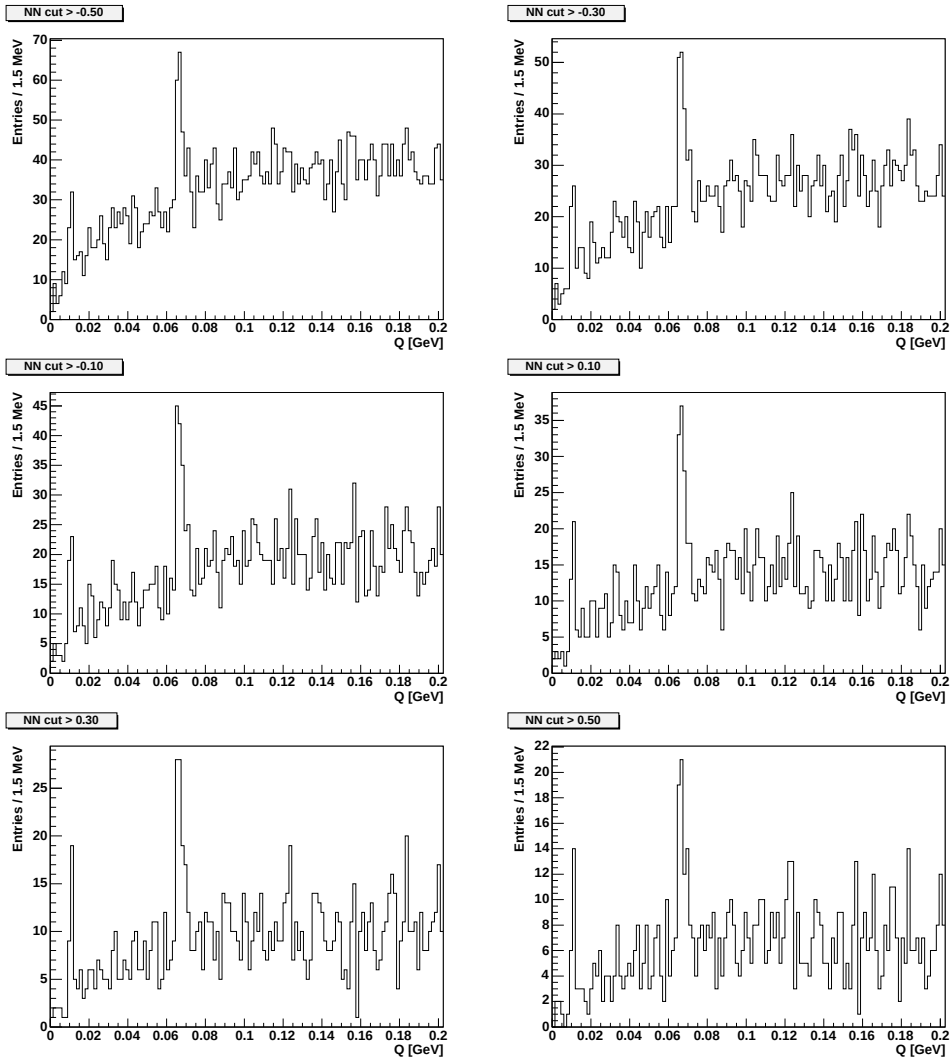

Figure 25: The Q distribution of the B_s^{**} candidates selected by different cuts on the Neural Network output in J/ψ trigger data. In all cases we demand, that less than four candidates are in the event. Starting from the top left corner the cut on the Neural Network output is -0.5. In the next plots we continue with steps of 0.2 in the Neural Network output.

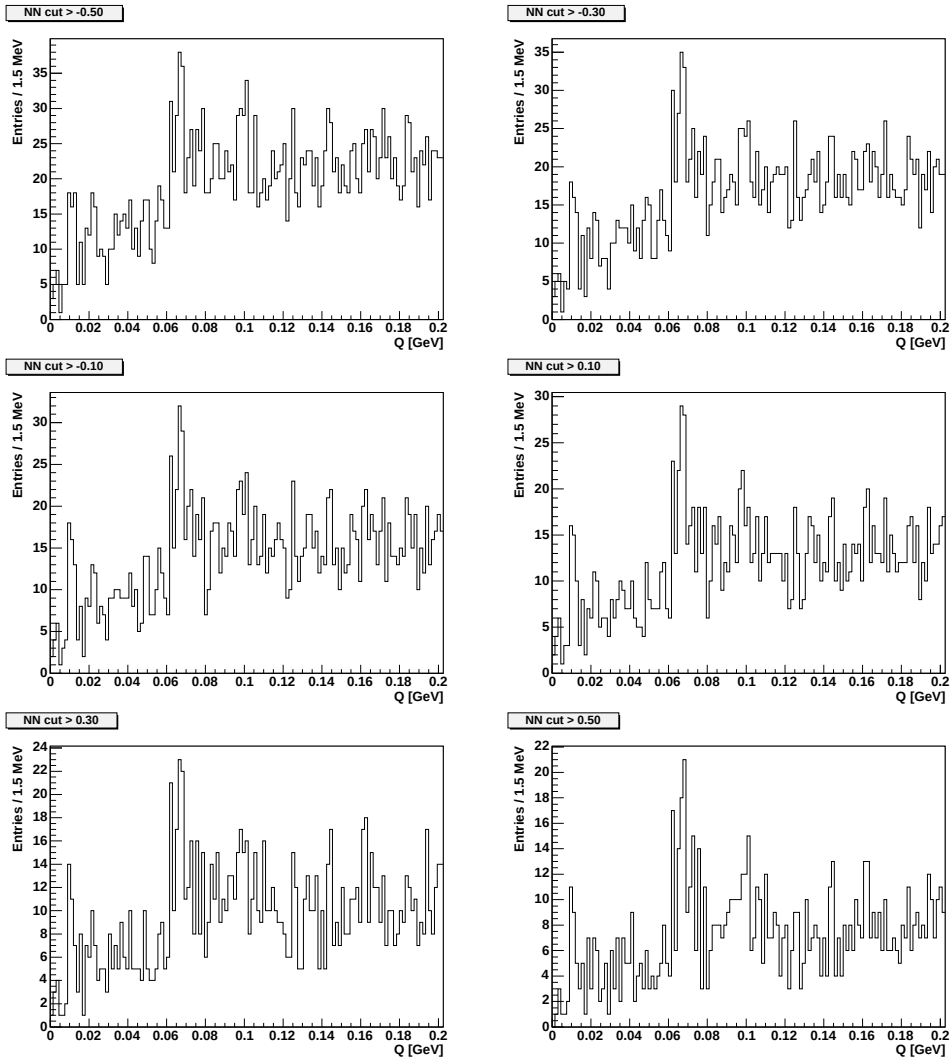

Figure 26: The Q distribution of the B_s^{**} candidates selected by different cuts on the Neural Network output in two track trigger data. In all cases we demand, that less than four candidates are in the event. Starting from the top left corner the cut on the Neural Network output is -0.5. In the next plots we continue with steps of 0.2 in the Neural Network output.

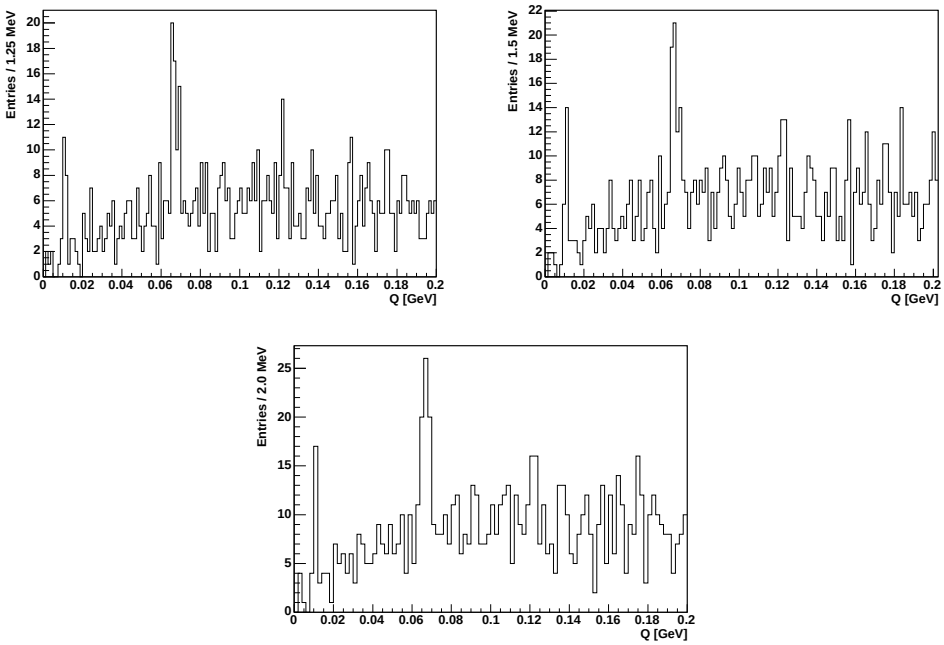

Figure 27: The Q distributions of the B_s^{**} candidates at the best Neural Network output cut with different bin widths for the J/ψ trigger data.

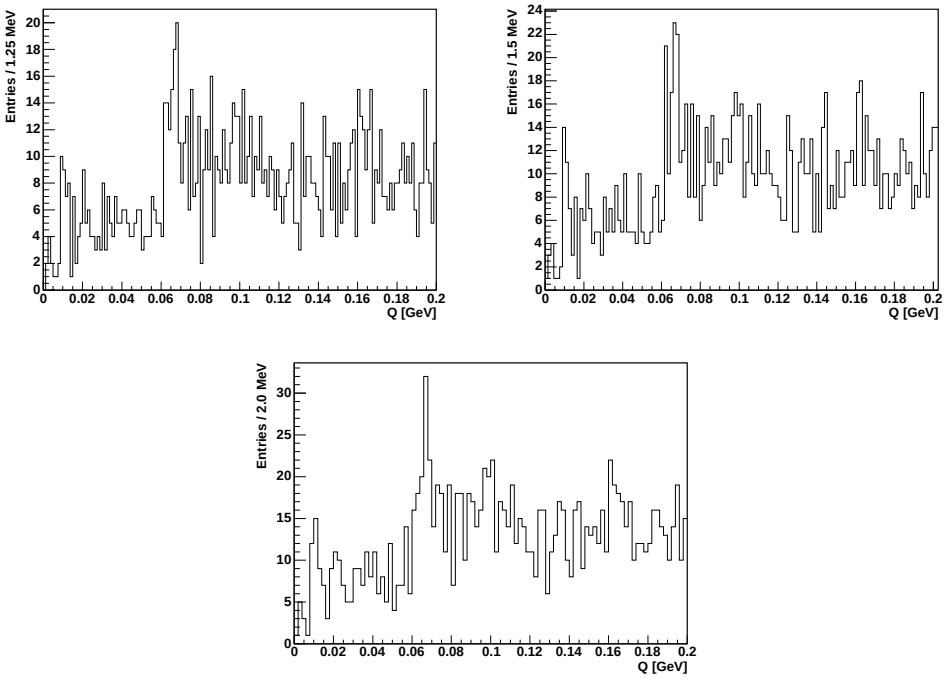

Figure 28: The Q distributions of the B_s^{**} candidates at the best Neural Network output cut with different bin widths for the two track trigger data.

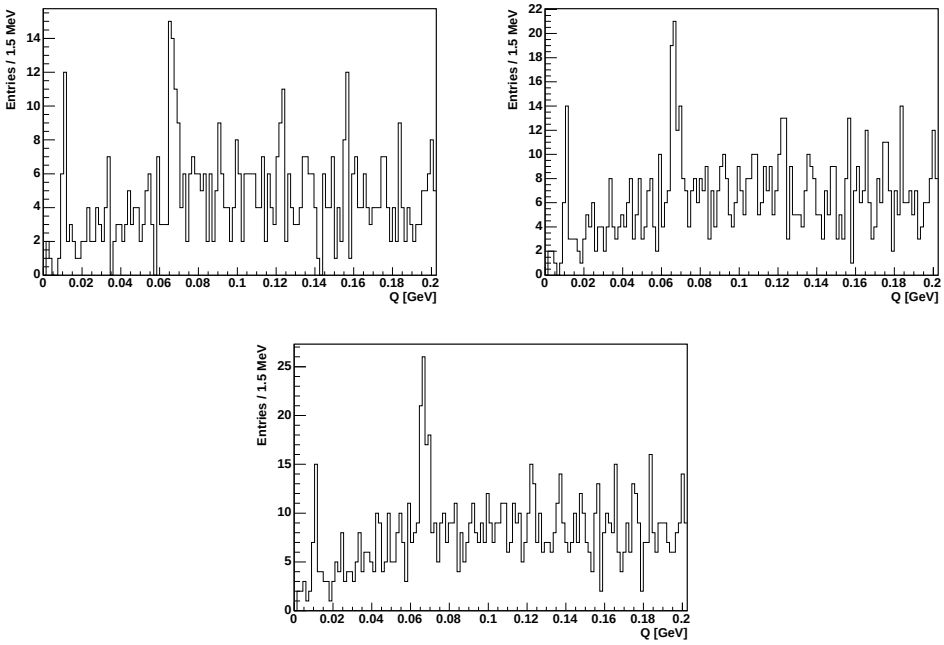

Figure 29: The Q distributions of the B_s^{**} candidates for different cuts on the number of candidates in the event for J/ψ trigger data. In top left $N_{cands} < 3$, top right $N_{cands} < 4$ (default) and bottom $N_{cands} < 5$ is shown.

cut on the Neural Network output and vary the cut on the number of candidates by ± 1 around the choosen value. Aim here is not to see, if the other possibility is better, the cut will stay fixed to less than four candidates in the further analysis, but to make sure again, that the seen signals are not just fluctuation at one single point in the selection space. In Figures 29 and 30 the Q distributions for J/ψ and two track trigger data with different cuts on the number of candidates are shown. In both cases in the top left distribution the number of candidates is required to be less than 3, the top right plot shows the default value (number of candidates less than four) and the bottom plot shows the Q distribution with the number of candidates less than five. Here again we can conclude, that the peaks seen on figures before are rather stable against the cut on the number of candidates.

To summarize this subsection, we scan possible selection space and different binnings to check, that we don't pick up some fluctuation with the final selection and the way of plotting Q distribution. After examination of distributions shown in this subsection, we can conclude, that this is indeed the case and that we see two real peaks, which can be interpreted as the two narrow states of the B_s^{**}.

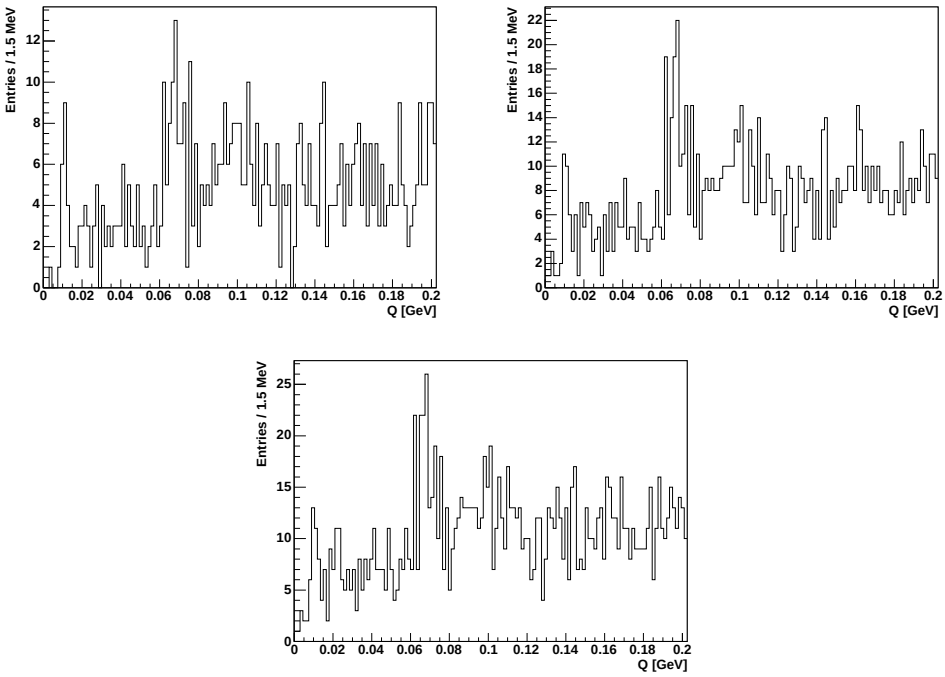

Figure 30: The Q distributions of the B_s^{**} candidates for different cuts on the number of candidates in the event for two track trigger data. In top left $N_{cands} < 3$, top right $N_{cands} < 4$ (default) and bottom $N_{cands} < 5$ is shown.

4.5 Wrong sign candidates

Another usual possibility how to check that the observed peaks are real is to examine distributions of the wrong sign candidates. In case of B_s^{**} we don't expect any peak in the wrong sign candidates.

One technical issue connected with it is, that we cut on the number of right sign candidates in the event to select our final sample. For wrong sign candidates, we can keep this cut same (means cutting on the number of right sign candidates) or cut on the number of wrong sign candidates in the event, which is closer to the logic of the selection cut for the right sign candidates. Both possibilities are checked and none of them show any significant structure. Figures 31 and 32 show Q distributions of the wrong sign candidates for different cuts on the Neural Network output with the requirement, that the number of right sign candidates is smaller than four. In Figures 33 and 34 we show same the except of the cut on the number of candidates, where we require the number of wrong sign candidates to be smaller than four.

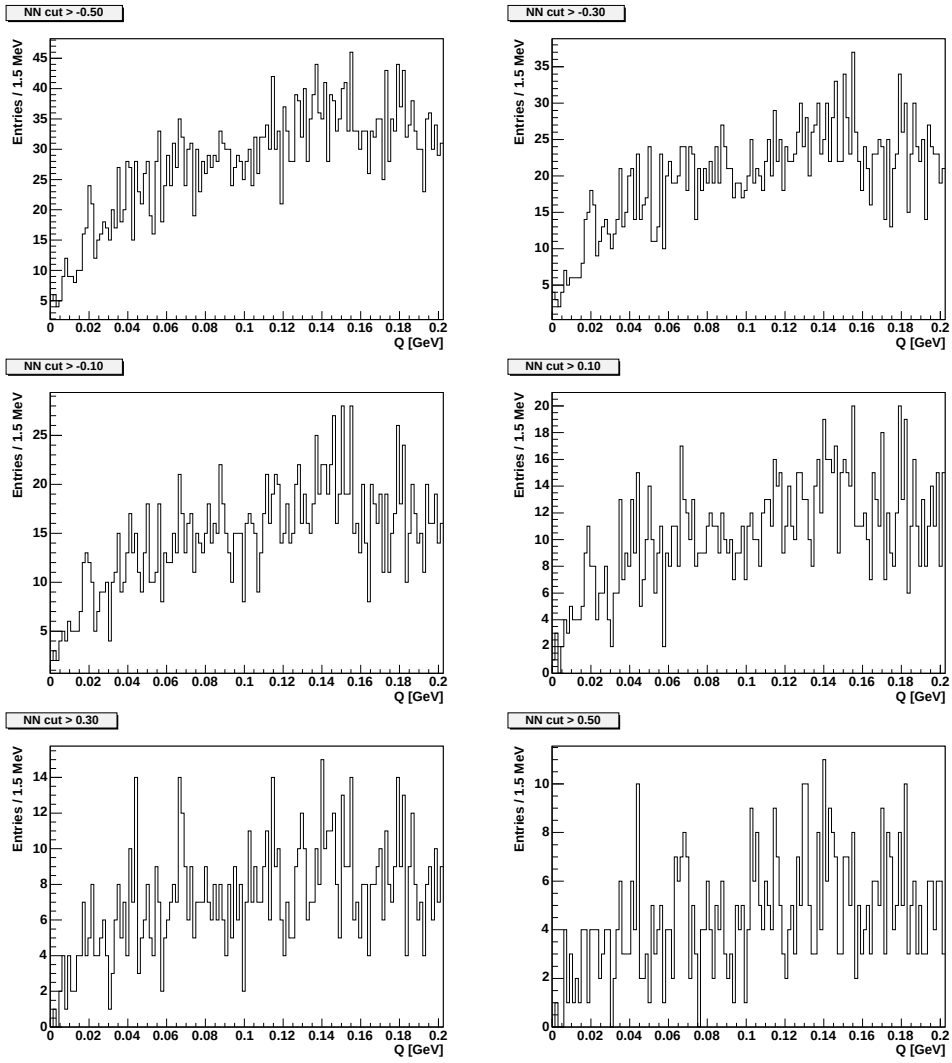

Figure 31: The Q distribution of the wrong sign B_s^{**} candidates selected by different cuts on the Neural Network output in J/ψ trigger data. In all cases we demand, that less than four right sign candidates are in the event. Starting from the top left corner the cut on the Neural Network output is -0.5. In the next plots we continue with steps of 0.2 in the Neural Network output.

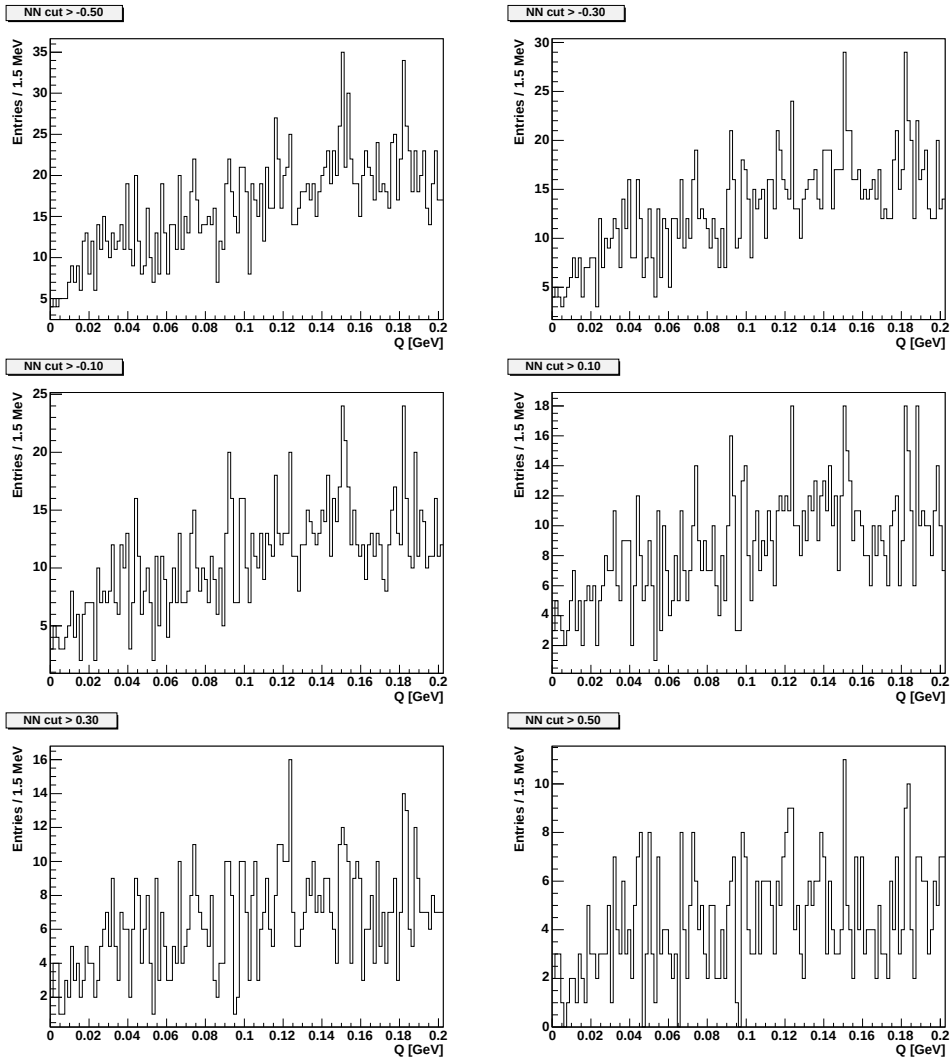

Figure 32: The Q distribution of the wrong sign B_s^{**} candidates selected by different cuts on the Neural Network output in two track trigger data. In all cases we demand, that less than four right sign candidates are in the event. Starting from the top left corner the cut on the Neural Network output is -0.5. In the next plots we continue with steps of 0.2 in the Neural Network output.

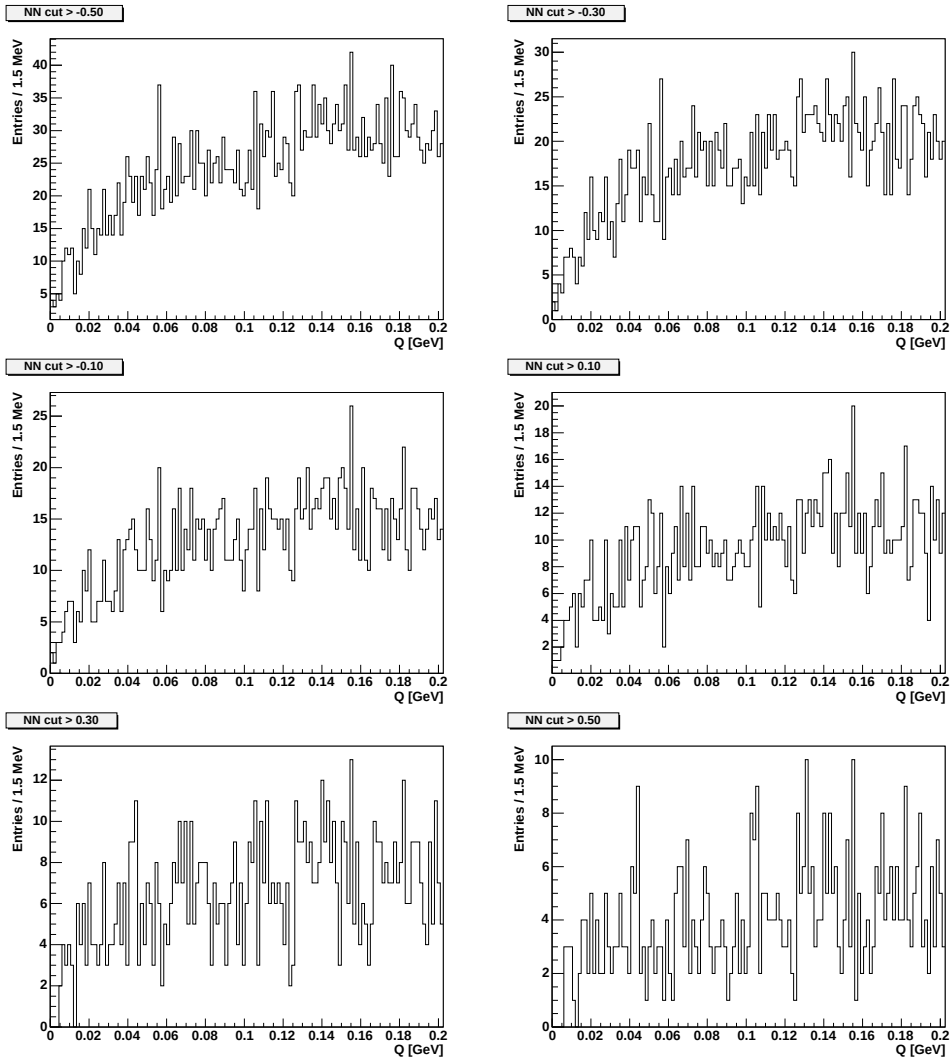

Figure 33: The Q distribution of the wrong sign B_s^{**} candidates selected by different cuts on the Neural Network output in J/ψ trigger data. In all cases we demand, that less than four wrong sign candidates are in the event. Starting from the top left corner the cut on the Neural Network output is -0.5. In the next plots we continue with steps of 0.2 in the Neural Network output.

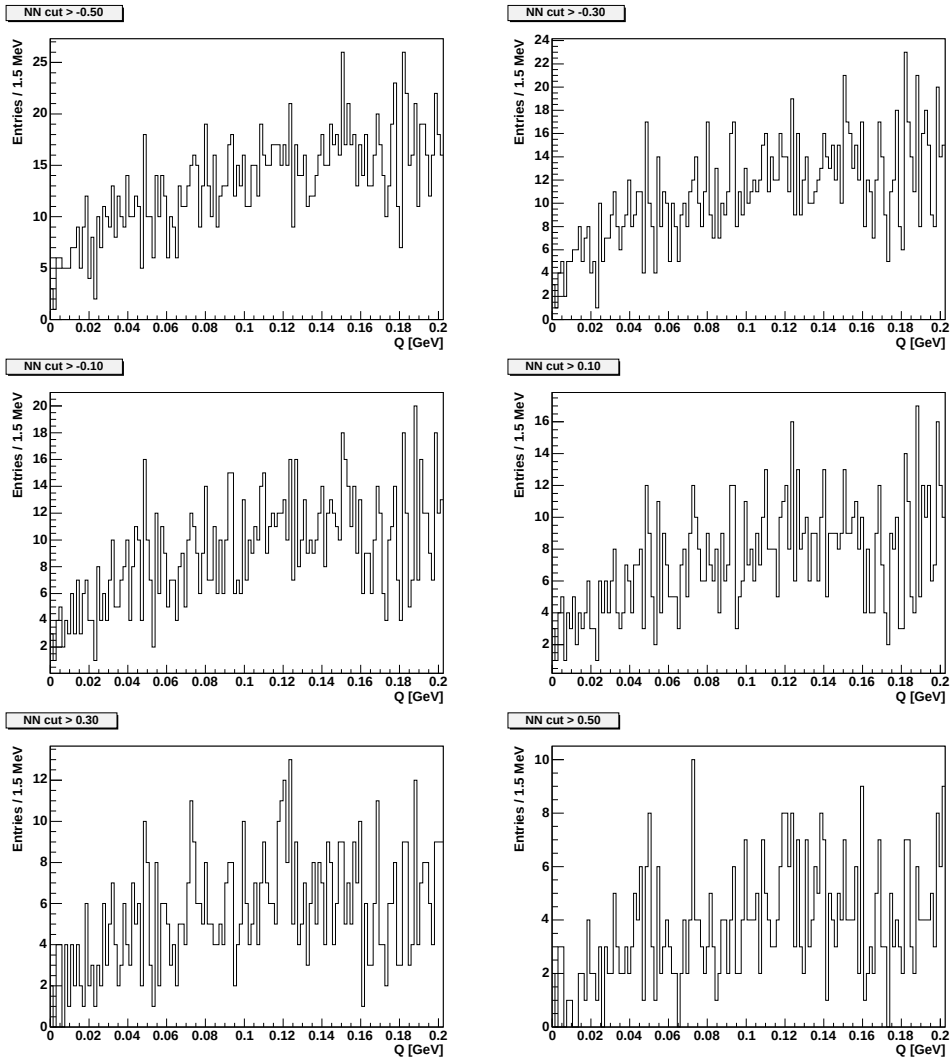

Figure 34: The Q distribution of the wrong sign B_s^{**} candidates selected by different cuts on the Neural Network output in two track trigger data. In all cases we demand, that less than four wrong sign candidates are in the event. Starting from the top left corner the cut on the Neural Network output is -0.5. In the next plots we continue with steps of 0.2 in the Neural Network output.

5 Unbinned Fit

In this section, the fit used to extract the Q values of the observed B_s^{**} states is described. As the statistics of resulting sample is relatively low, we use an unbinned maximum likelihood fit to perform the task. Two possibilities with different background parametrizations are implemented, in order to be able to check for the possible systematic effects.

5.1 Likelihood expression

We fit the Q value distribution, which has three contributions. Those are the two peaks for the signals seen and the background. In the fit itself we don't separate possible different contributions to the separate functions, but rather use one effective function, which describes the complete background. As the statistics is still small, we are neither sensitive nor interested to distinguish different types of the background. Highest contributions to the background are combinatorial background and possible contribution from the wide B_s^{**} states about which nothing is known.

The basic likelihood expression is

$$\mathcal{L} = \prod_i [f_s \cdot \mathcal{S}(Q_i) + (1 - f_s) \cdot \mathcal{B}(Q_i)] \tag{5}$$

where f_s is fraction of the two signals, $\mathcal{S}(Q_i)$ is the likelihood expression for the signals and $\mathcal{B}(Q_i)$ for the background. Index i runs over all events in the sample. Likelihood for the signals is composed from two Gaussian peaks in following way

$$\mathcal{S}(Q_i) = f_{B_1} \cdot \mathcal{G}(Q_1, \sigma \cdot R) + (1 - f_{B_1}) \cdot \mathcal{G}(Q_2, \sigma) \tag{6}$$

Here, f_{B_1} is fraction of the B_{s1} state from all signal events, Q_1 and Q_2 are Q-values of the two states, σ is the width of the B_{s2}^* state and R is ratio of the widths of the two states. For $\mathcal{B}(Q_i)$ the expression used is

$$\mathcal{B}(Q_i) = \alpha [Q_i(\beta - Q_i)]^\gamma \exp[-\gamma Q_i] \tag{7}$$

where α is a normalisation constant, and β and γ are free parameters. Alternatively we use instead of equation 7 a third order polynomial.

The actual fit is performed by minimizing $-2 \ln \mathcal{L}$ by MINUIT[22] provided by the ROOT[23] package. During the fit, all parameters in the likelihood are kept free, only two fractions were forced to be inside a physical reagion.
The result for the fit parameters of the fit for the two subsamples and the whole sample are shown in table 4.

Parameter	J/ψ	TTT	Full
Signal fractions	0.072 +- 0.012	0.060 +- 0.011	0.062 +- 0.009
Bs1 fraction	0.278 +- 0.069	0.270 +- 0.077	0.278 +- 0.056
Bs1 Q value	0.011 +- 0.000	0.011 +- 0.000	0.011 +- 0.000
Bs2 Q value	0.067 +- 0.000	0.067 +- 0.001	0.067 +- 0.000
Bs1/Bs2 width ratio	0.361 +- 0.114	0.410 +- 0.161	0.417 +- 0.134
Bs2 width	0.002 +- 0.000	0.003 +- 0.001	0.002 +- 0.001
beta	0.278 +- 0.022	0.310 +- 0.027	0.291 +- 0.021
gamma	0.655 +- 0.004	0.602 +- 0.073	0.634 +- 0.066

Table 4: The results for the parameters of the likelihood function after the minimization.

5.2 Fitter validation

In order to validate the fitter we use Toy Monte Carlo techniques. For each background parametrization, we generate 10000 Toy Monte Carlo according to the PDF's obtained from the fit to the data. Each Toy Monte Carlo has the statistics of the final dataset. On Figure 35 an example of the one fit on the Toy Monte Carlo is shown.

After fitting the Toy Monte Carlos, we examine the pulls of the important parameters. As the statistics in the fit is relatively small, some of the fits don't converge and those are excluded from the pulls shown here. On figure 36 the pulls for parameters f_s and f_{B_1} in a case of using background parametrization from equation 7 are shown. From the Gaussian fit to the pulls one can see that those two parameters are practically unbiased with reasonable error estimate. The parameters which are of real interest are the Q-values of the two signals, for which we show pulls in Figure 37. In the case of the Q-values pulls are consistent with mean zero and unit width, which means that the fitter code, returns unbiased values and proper error estimates for the important parameters.

The same exercise was done also in the case of background parametrization using third order polynomial. Pull distributions for four interesting parameters are shown on Figure 38. Again from the Gaussions fitted to the pull distributions we can conclude that the important parameters are unbiased.

6 Systematic uncertainties

In the analysis two quantities are to be considered as a final result, namely the Q-values for the two narrow B_s^{**} states, those are the only quantities for which we study systematic uncertainties. We consider two sources of systematic uncertainty and those are the systematic uncertainty due to the tracking system calibration and uncertainty due the fitting procedure.

In addition we perform a cross check that optimization of the cut on the Neural Network using data in a small window, where B_{s2}^* state resides doesn't bias the resulting

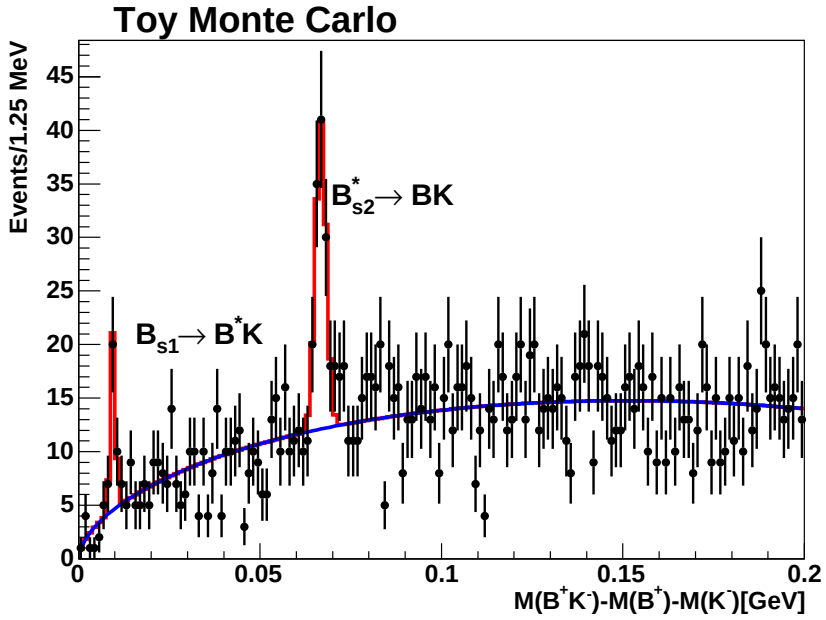

Figure 35: Example of the fit on Toy Monte Carlo sample, using background parametrization from equation 7.

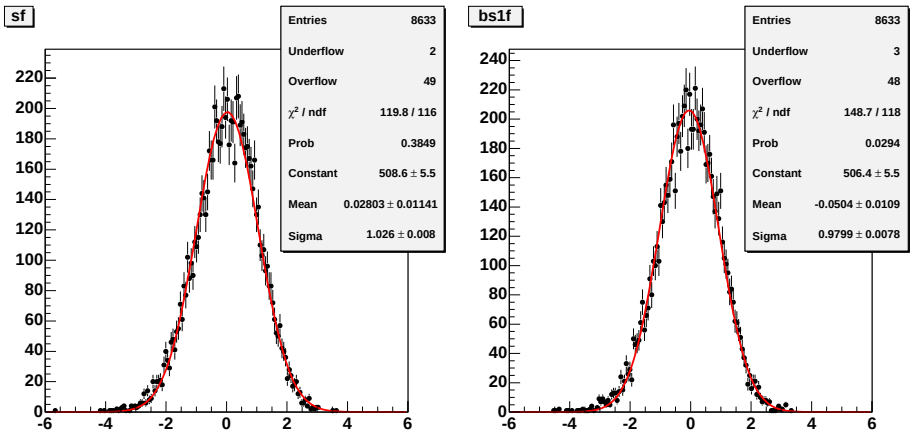

Figure 36: Pull distributions for the signal fraction and B_{s1} fraction from the Toy Monte Carlo experiments using background parametrization from equation 7.

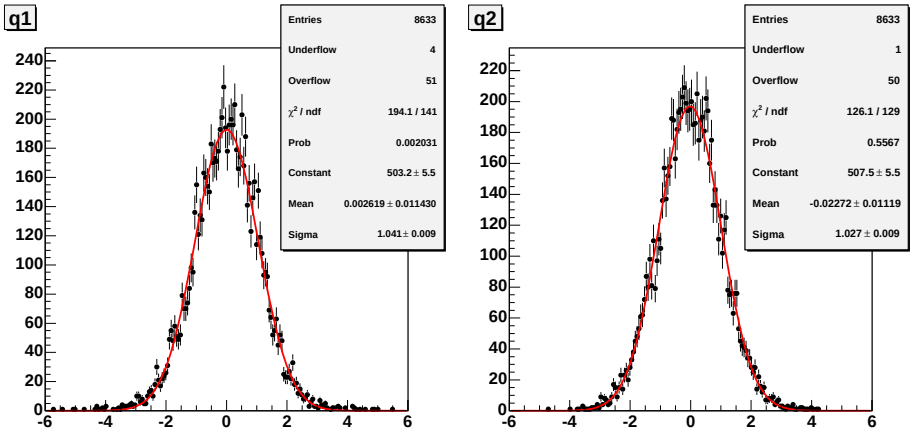

Figure 37: Pull distributions for the Q-values for the two signals from the Toy Monte Carlo experiments using background parametrization from equation 7.

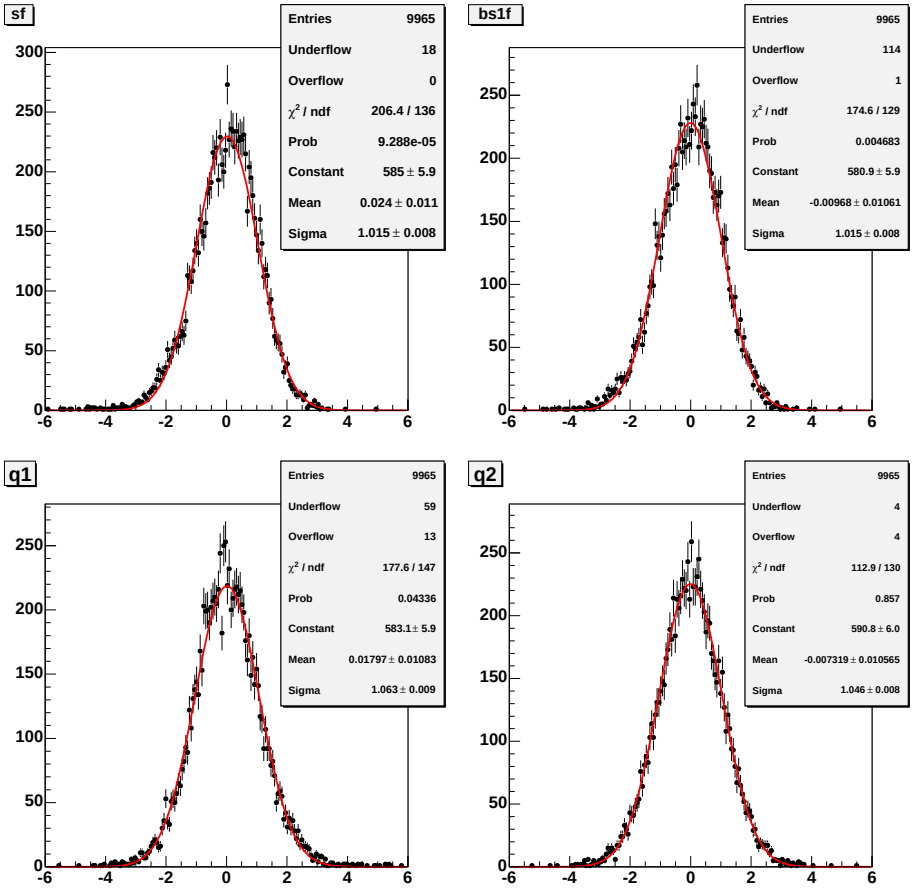

Figure 38: Pull distributions for the signal fraction and B_{s1} fraction and Q-values from the Toy Monte Carlo experiments using third order polynomial for background parametrization.

significance. This question was raised during one of our presentations and we address it here.

The last check performed deals with the fact, that we use two B^+ channels, which are reconstructed from different triggers, which can potentially lead to samples with different behaviour. Therefore we make a basic consistency check between two samples.

6.1 Tracking calibration

The main effect from tracking comes from the absolute calibration of the detector material description in the tracking, magnetic field and COT error matrix scaling. The systematic uncertainty comes from the error propagation of the uncertainties on the calibrations. The measurements of the $B^{**}[1]$ masses and $D^{**}[2]$ are analysis very simillar to this one, except of the fact, that Q value in the B_s^{**} decays is much smaller than in the two mentioned analysis. Therefore we consider it safe to use the error values quoted by those two analysis. They are found to be 0.1 MeV for the COT error matrix scaling and 0.1 MeV for tracking material and magnetic field calibration. Those values probably overestimate the actual uncertainty, but as the final uncertainty is anyhow dominated by the statistical one, we don't see neccessity to waste collaboration resources to narrow it down.

6.2 Fitting procedure

For the fitting procedure two possible sources of a systematic uncertainty are investigated. They are due the background model and signal resolution model as we use a single Gaussian for the signal parametrization in the fit, but the two subsamples from different triggers have slightly different resolution. Both effects are studied using Toy Monte Carlo simulation.

To find out the possible effect of the background modeling, we generate 10000 Toy Monte Carlo experiments using a third order polynomial as a PDF for generating background. The parameters of the polynomial are taken from the fit to the data. Those Toy Monte Carlo experiments are afterward fitted using the background parametrization from equation 7. The obtained pull distributions for the interesting parameters are shown on Figure 39. Examining the result of the fits with Gaussian function we find, that results of the fit to Toy Monte Carlo are unbiased and the obtained errors are correct. Therefore we don't assign any systematic uncertainty due the background modeling.

The second effect which we investigate is the approximation with single Gaussian for each peak. This is only an approximation as the resolution of the two subsamples are known to differ by about 20 % (see Figures 4 and 5). To estimate the effect of the approximation, we generate 10000 Toy Monte Carlo experiments with two Gaussion PDFs for each signal, where the width of the Gaussians were fixed to the expectation from Monte Carlo. The relative ratio between the two Gaussians was taken from the fit to the subsamples and is found to be 45%. The obtained pull distributions for the four

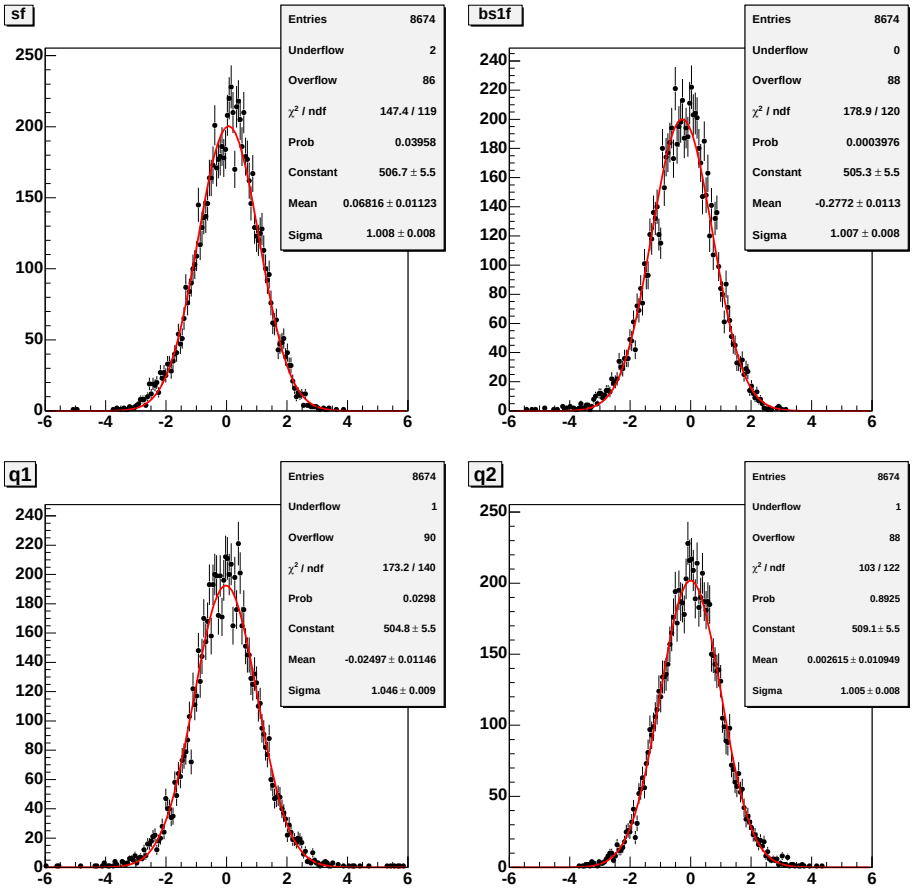

Figure 39: Pull distributions from fits to Toy Monte Carlo experiments generated with third order polynomial background and fitted fit background parametrization from equation 7.

important parameters are show in Figure 40. As one can see, they are consistent with the unbiased fit with proper error estimate and therefore no systematic uncertainty is assigned to this source.

6.3 Possible significance bias

As a cut on the Neural Network to select B_s^{**} candidates is optimazed using data in the narrow Q window from 60 to 70 MeV, where one of our signal is present, one can ask if the significance of that signal is unbiased. Looking to the Figures 12 and 21, which shows Neural Network Output distributions for signal and background in the two channels, we don't expect larger change if the cut on the Neural Network output is moved by 0.2. Anyhow, in order to check for it we make our optimization also in a wider Q window from 0 to 150 MeV and check if the two selections agrees in significance values. In Table 5 we list significance estimations using three different estimators for

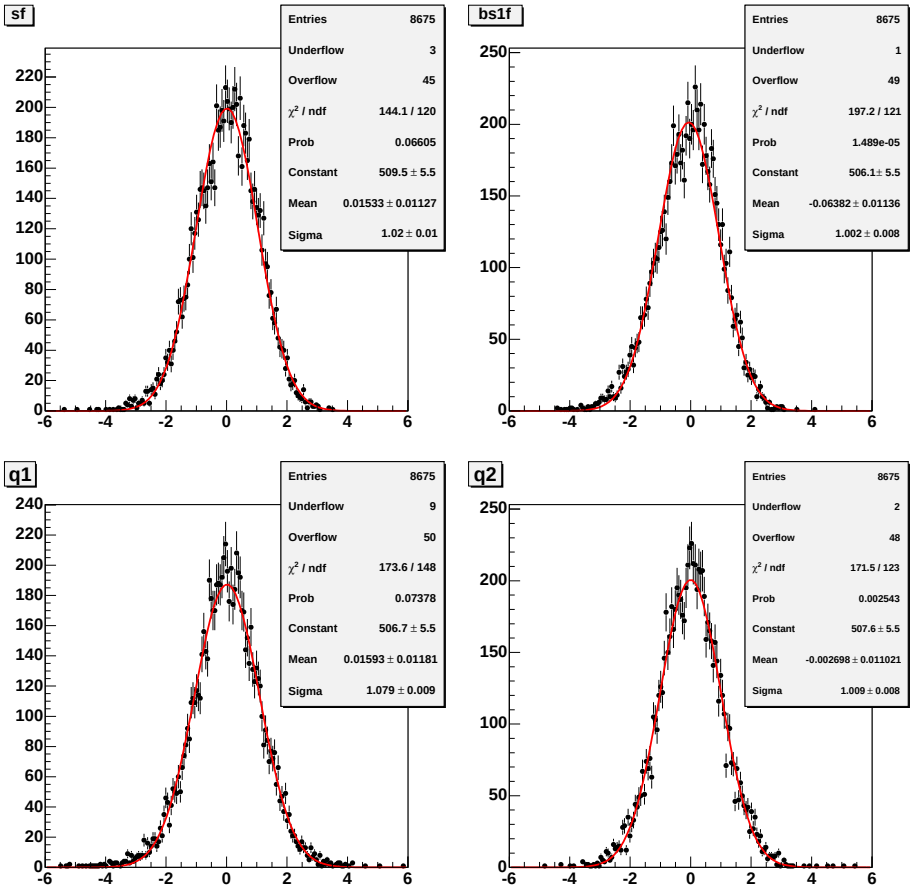

Figure 40: Pull distributions from fits to Toy Monte Carlo experiments generated with two Gaussians for each signal and fitted with one Gaussian for each signal.

the observed B_s^{**} states using two different Q windows for cut optimization. As can be seen from the table, the significance for the B_{s2}^* is not biased in a significant way.

6.4 Consistency of two B^+ decays

As last cross check we analyze the two subsamples from different B^+ decay channels separately in order to check if the two samples are consistent. Special attention is put to the position of the peaks and number of signal events in the peaks.

On Figures 41 and 42 fits with the default background model of the J/ψ and TTT subsamples are shown and Table 6 shows values of the interesting parameters in the two subsamples. Finally on Figure 43 the resulting background functions are plotted together with the background function obtained on the full sample. The functions for the subsamples are scaled to the area of the full sample in order to facilitate shape comparisons. From all these information one concludes, that the two subsamples are consistent with each other and therefore can be treated as one sample in the final fit.

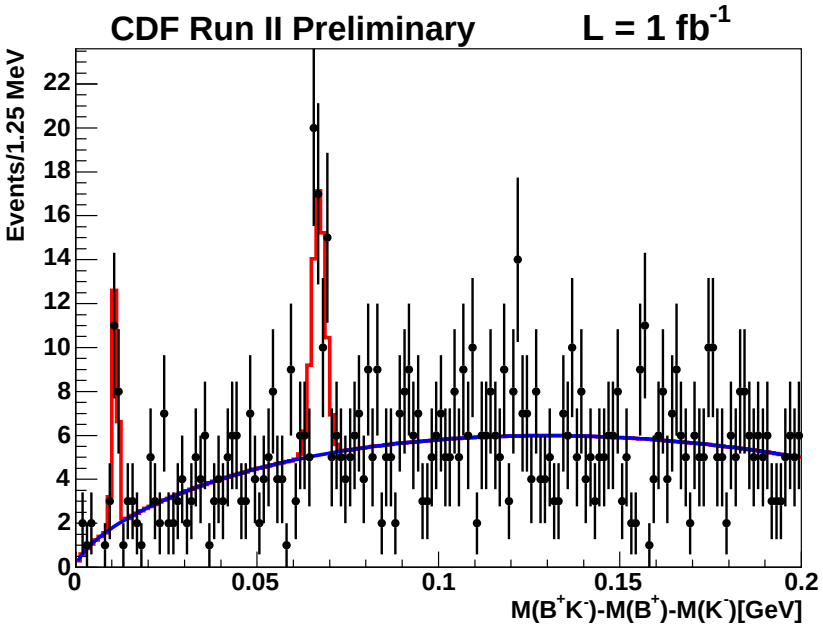

Figure 41: Unbinned maximum likelihood fit of the J/ψ trigger subsample. Values of the important parametersa re listed in table 6.

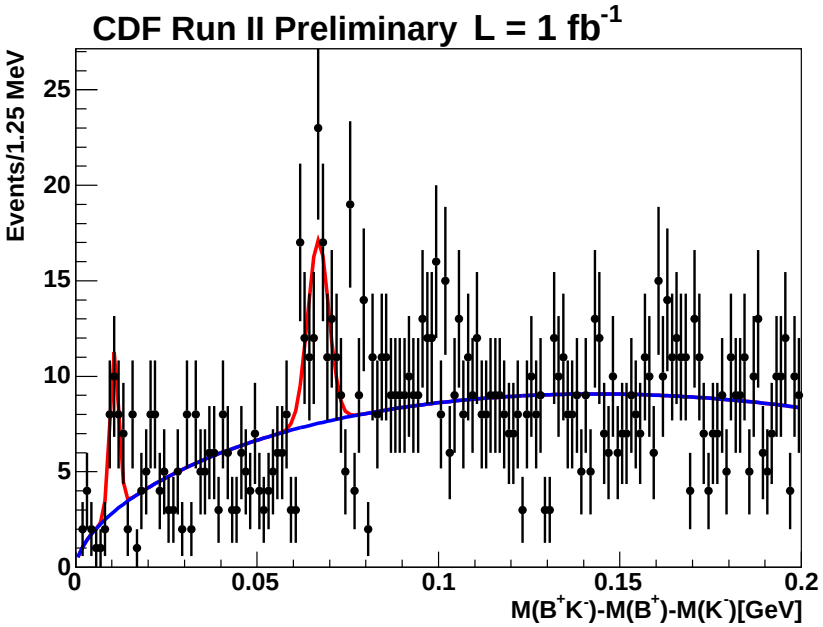

Figure 42: Unbinned maximum likelihood fit of the TTT trigger subsample. Values of the important parametersa re listed in table 6.

	60-70 MeV Window		0-150 MeV Window	
Estimator	B_{s1}	B^*_{s2}	B_{s1}	B^*_{s2}
$S/\sqrt{S+B}$	4.70	6.19	4.08	6.31
Poisson probability	7.75	> 8	7.24	> 8
$\mathcal{L}/\mathcal{L}_0$	6.29	7.74	4.98	8.2

Table 5: Estimation of the significance of the two observed B^{**}_s states using different Q windows for cut optimization.

	B_{s1}		B^*_{s2}	
	J/ψ data	TTT data	J/ψ data	TTT data
Q [MeV]	10.87 ± 0.19	10.68 ± 0.46	67.03 ± 0.44	66.85 ± 0.76
σ [MeV]	0.64 ± 0.25	1.18 ± 0.56	1.79 ± 0.42	2.88 ± 0.75
N	16.98 ± 5.14	20.66 ± 7.12	44.15 ± 13.36	55.74 ± 19.20

Table 6: Comparison of the unbinned maximum likelihood fits of the two data subsets coming from different triggers. Those two subsamples are independent and have roughly the same B^+ yield.

6.5 Systematic uncertainties summary

After investigating all possible important sources of systematic uncertainties we found, that only uncertainty due the tracking calibrations needs to be assigned. We didn't find any other source, which would need an assigment of the systematic uncertainty. Adding two contributions from the tracking calibrations in quadrature, the total systematic uncertainty on the Q value of both observed states is 0.14 MeV.

7 Results

As was shown in the previous section, the two subsamples with different B^+ decay coming from different triggers are consistent, we proceed with the analysis using it as one sample. The results of the unbinned maximum likelihood fit are listed in table 7. Fit projection is shown in Figure 44. Figure 45 zooms in to show better the signal peaks. Taking binned data on figure 44 we calculate the fit probability by merging neighbouring bins until each merged bin has at least 15 entries. Merged bins are then used to calculate χ^2. Resulting $\chi^2 = 86.01$ at 87 degrees of freedom, which can be translated to a fit probability 50.99%.

From the observed Q values the natural interpretation of the two signals is that the one at $Q = 10.73$ MeV is from decay $B_{s1} \to B^{*+}K^-$ and the one at $Q = 66.96$ MeV is from $B^*_{s2} \to B^+K^-$. In this case, one would also expects decay $B^*_{s2} \to B^{*+}K^-$, which would be roughly at 22 MeV. Indeed the data show one bin which is above background. If we include another peak with position fixed to be $m(B^*) - m(B)$ lower, than peak at 66.96 MeV, we get a result, which is consistent with the expectations. Unfortunately

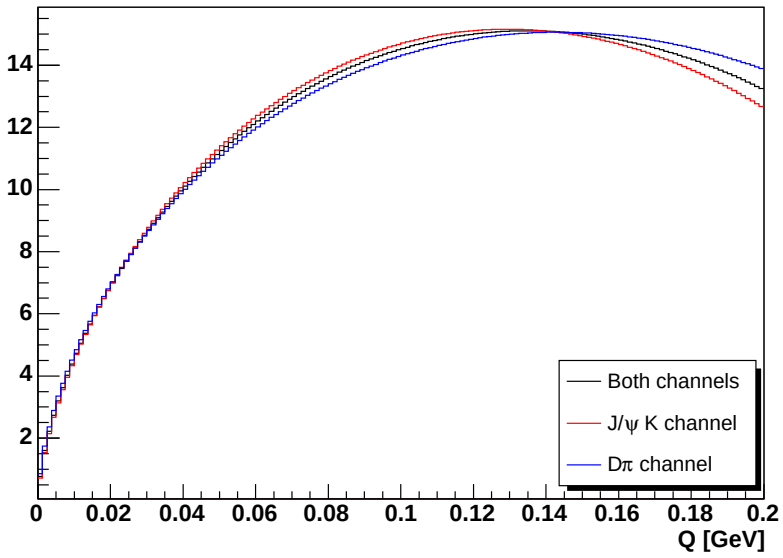

Figure 43: Comparison of the background shapes between two data subsamples. The functions for J/ψ trigger data and two track trigger data are scaled to same area as function from fit to the full sample to facilitate shape comparison.

	Default fit		Alternative fit	
	B_{s1}	B^*_{s2}	B_{s1}	B^*_{s2}
Q [MeV]	10.73 ± 0.21	66.96 ± 0.39	10.75 ± 0.21	66.95 ± 0.39
σ [MeV]	0.90 ± 0.36	2.16 ± 0.53	0.93 ± 0.41	2.04 ± 0.58
N	36.4 ± 9.0	94.8 ± 23.4	38.0 ± 9.5	90.6 ± 22.6

Table 7: Result of the fit to the full sample. Errors are statistical only. The default fit uses background parametrisation according to equation 7, while the alternative fit uses a third order polynomial.

statistics is too low to claim that we also see the decay $B^*_{s2} \to B^{*+}K^-$. This would need more statistics, which is beyond scope of the current analysis.[5]

To estimate the significance of the two observed signals, we repeat the fit with given peak switched off and form likelihood ratio $\mathcal{L}_0/\mathcal{L}$, where $\mathcal{L}$ is value at the minimum with both peaks included while $\mathcal{L}_0$ is value at the minimum if the corresponding peak is excluded. Then $-2\ln\mathcal{L}_0/\mathcal{L}$ is asymptotically distributed as a χ^2 distribution with the number of degrees of freedom equal to the difference in the number of parameters in two fits. We can therefore ask, what is a probability to obtain value of $-2\ln\mathcal{L}_0/\mathcal{L}$

[5]One can still add some B^+ channels, which however have larger background and smaller yield, than two used in the analysis. Another possibility is to use decays $B^{**}_s \to B^{(*)0}K^0$, which again have some limitations. Probably by inclusion of all possible channels with dataset on the level of $4\,fb^{-1}$ one could hope for seeing the third peak also.

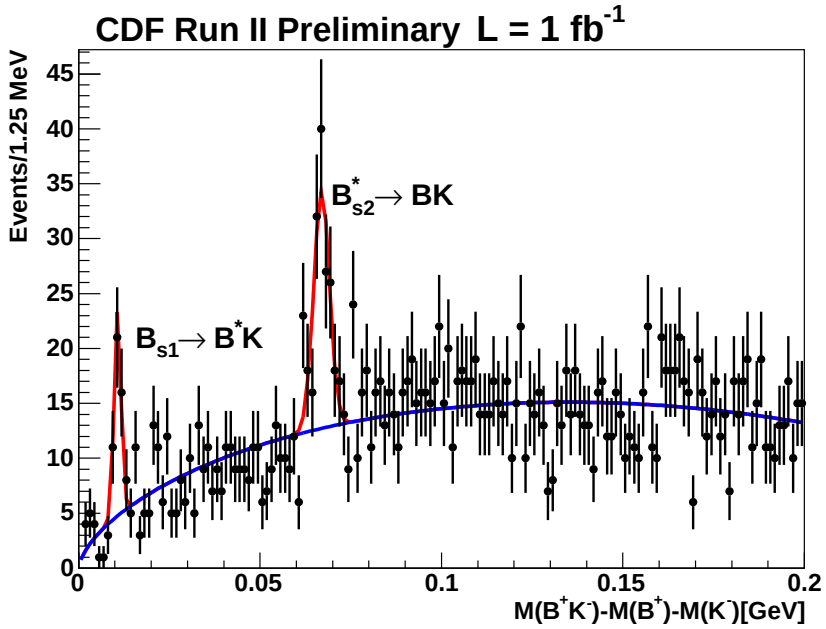

Figure 44: Resulting unbinned maximum likelihood fit to the full data sample.

same or larger than is observed in data and convert this probability to the significance. The advantage of the procedure is, that we can estimate significance completely from unbinned data, which is crucial at the current statistics. Doing this we obtain significances of the state at 10.73 MeV to be $6.3\,\sigma$ and state at 66.96 MeV to be $7.7\,\sigma$.

In order to get a better feeling about the significance of the two signals, we perform a significance scan over wider region. For every peak, we fit data without this peak and than fix width to the measured one and scan with fixed peak position and floating normalisation. The significance is than obtained from $-2\ln\mathcal{L}_0/\mathcal{L}$, which has now one degree of freedom. Results of those scans are on Figures 46 and 47 for B_{s1} and B^*_{s2} respectively. From the figures we conclude that there is no hint that the observed peaks would be just statistical fluctuations.

Up to now, we evaluated the significance under the assumption, that if there is a signal, it is in the place, where we see it. This is not the best assumption, especially for the B_{s1} state, which was not seen before. The B^*_{s2} state was seen already by three experiments and therefore, we don't need to consider a situation, that B^*_{s2} could be produced at a mass different from the one, where we see a signal. For the B_{s1} we therefore define a search window from $Q = 0\,\text{MeV}$ to $Q = 50\,\text{MeV}$ and calculate the probability to see a signal like in data from pure background fluctuation (p-value). As a figure of merit the $-2\ln\mathcal{L}_0/\mathcal{L}$ is taken and the p-value is defined as the probability to obtain a value equal or larger than one seen in data from pure background fluctuation. We evaluate it from Toy Monte Carlo experiments, where we generate only background and a peak corresponding to B^*_{s2} and perform a likelihood fit with and without the B_{s1} signal. In Figure 48 the distribution of the $-2\ln\mathcal{L}_0/\mathcal{L}$ for pure background

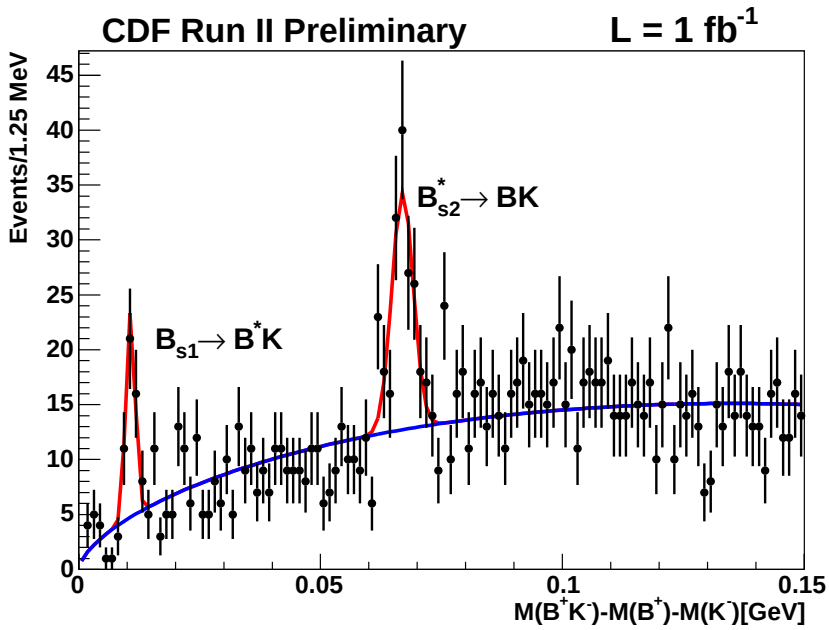

Figure 45: Zoom of the Figure 44 to the region from 0 to 150 MeV Q value.

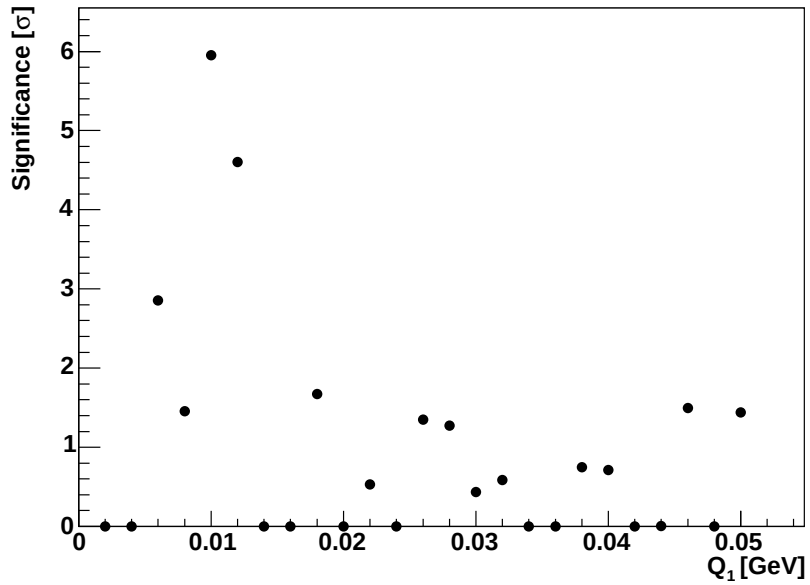

Figure 46: Significance scan for the B_{s1} signal over wider range of Q values. For details see text.

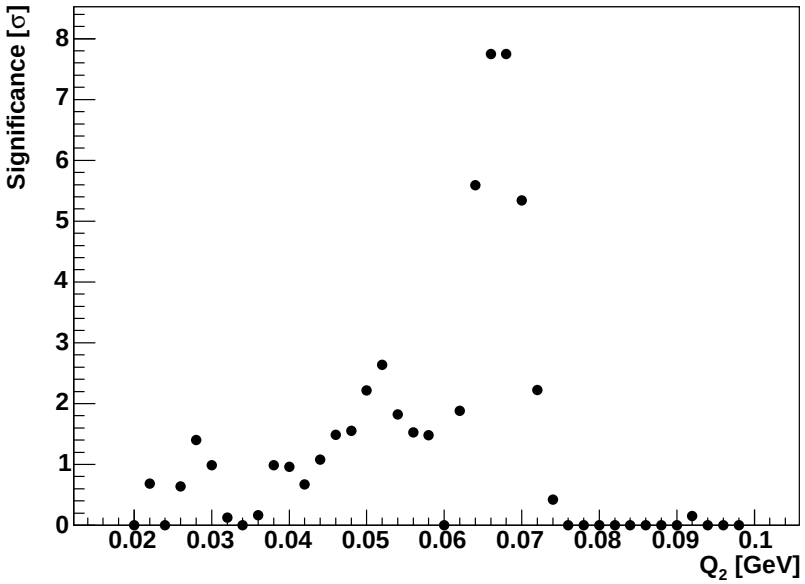

Figure 47: Significance scan for the B^*_{s2} signal over wider range of Q values. For details see text.

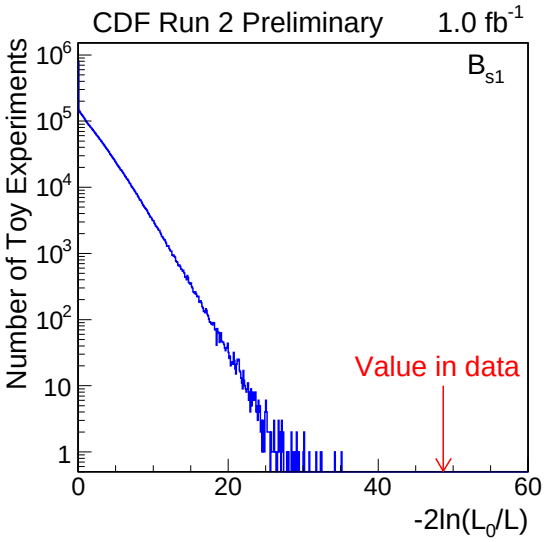

Figure 48: Distribution of the $-2\ln\mathcal{L}_0/\mathcal{L}$ for the peak corresponding to the B_{s1} state if it would stem pure background fluctuation. The search window is defined from $Q = 0\,\text{MeV}$ to $Q = 50\,\text{MeV}$.

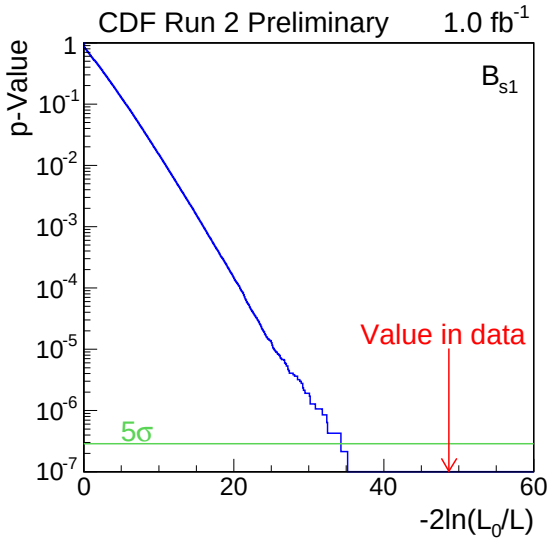

Figure 49: P-Value as a function of $-2\ln\mathcal{L}_0/\mathcal{L}$ for the peak corresponding to the B_{s1} state if it would stem pure background fluctuation. The search window is defined from $Q = 0\,\text{MeV}$ to $Q = 50\,\text{MeV}$.

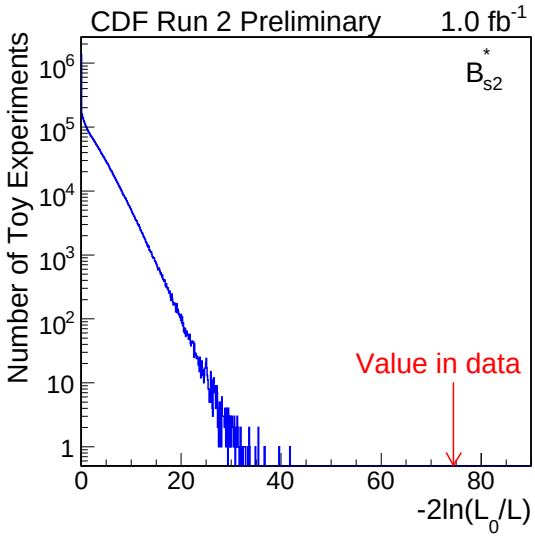

Figure 50: Distribution of the $-2\ln\mathcal{L}_0/\mathcal{L}$ for the peak corresponding to the B^*_{s2} state if it would stem pure background fluctuation. The search window is defined from $Q = 20\,\text{MeV}$ to $Q = 120\,\text{MeV}$.

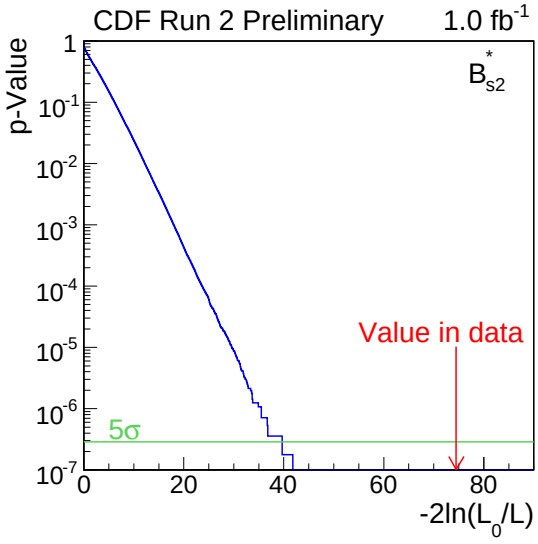

Figure 51: P-Value as a function of $-2\ln\mathcal{L}_0/\mathcal{L}$ for the peak corresponding to the B^*_{s2} state if it would stem pure background fluctuation. The search window is defined from $Q = 20\,\text{MeV}$ to $Q = 120\,\text{MeV}$.

fluctuation in the choosen Q window is shown. In total, in 4.7 million Toy Monte Carlo experiments both fits converged and none of the toy experiments yields value larger than $-2\ln\mathcal{L}_0/\mathcal{L} = 48.72$, which is seen in the data. In Figure 49 we integrate and normalise distribution from Figure 48 and plot directly the p-Value. Last non-zero bin gives p-Value $2.13 \cdot 10^{-7}$ at $-2\ln\mathcal{L}_0/\mathcal{L} \approx 35.15$, which is still far from the data value. This result gives additional support that the significance of the B_{s1} is above 5σ. As we would need much larger statistics to have reasonable sensitivity for p-Value at $-2\ln\mathcal{L}_0/\mathcal{L} \approx 48$ we don't continue the evaluation as it would not add new information at the end. One could extrapolate the distribution, but unfortunately there seems to be kink in the distribution at the place with low statistics and therefore the extrapolation wouldn't be reliable enough. Again much larger statistics would be needed to be able to conclude whether kink is there or not. Same exercise can be repeated for B^*_{s2} state. In this case we set the search window from 20 MeV to 120 MeV. Out of the 5.63 million Toy Monte Carlo experiments with both fits convereged we found largest observed value of $-2\ln\mathcal{L}_0/\mathcal{L} \approx 41.75$. Comparing it to the value $-2\ln\mathcal{L}_0/\mathcal{L} = 74.47$ obtained on data we conclude that also this state is observed with significance above 5σ level. Distribution of the $-2\ln\mathcal{L}_0/\mathcal{L}$ for pure background fluctuation in the choosen Q window is shown in Figure 50 and integral which gives directly p-Value is shown in Figure 51.

Figures 52 and 53 show projections of the likelihood around minimum of the fit for two Q values. The likelihood there behaves nicely without unwanted structures.

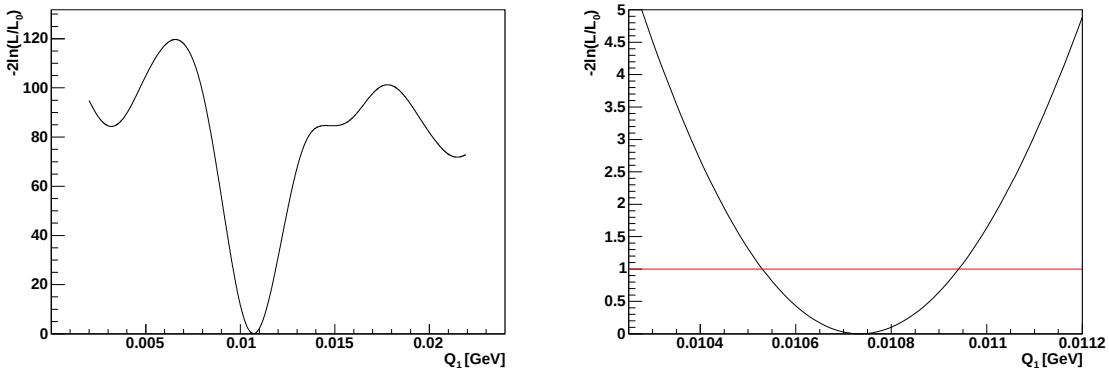

Figure 52: Projection of the likelihood for Q value of the B_{s1} state. On the left side the projection in wider region around minimum is shown, while on the right side, we zoom in around minimum.

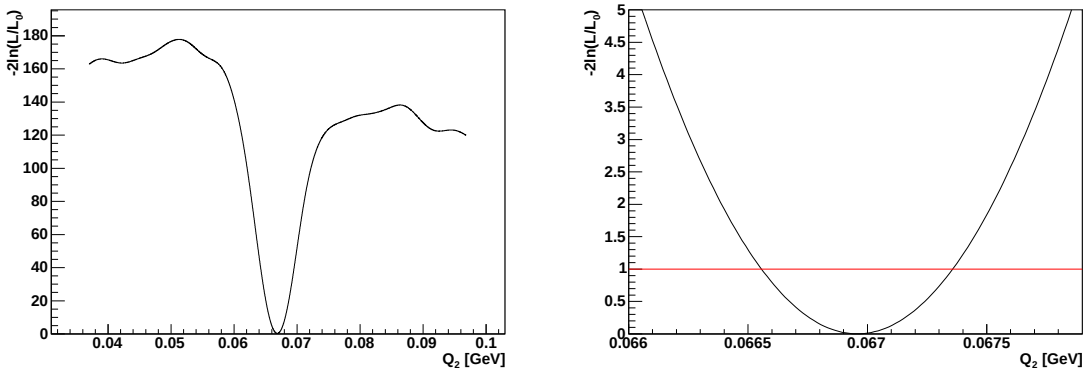

Figure 53: Projection of the likelihood for Q value of the B^*_{s2} state. On the the left side projection in wider region around minimum is shown, while on the right side, we zoom in around minimum.

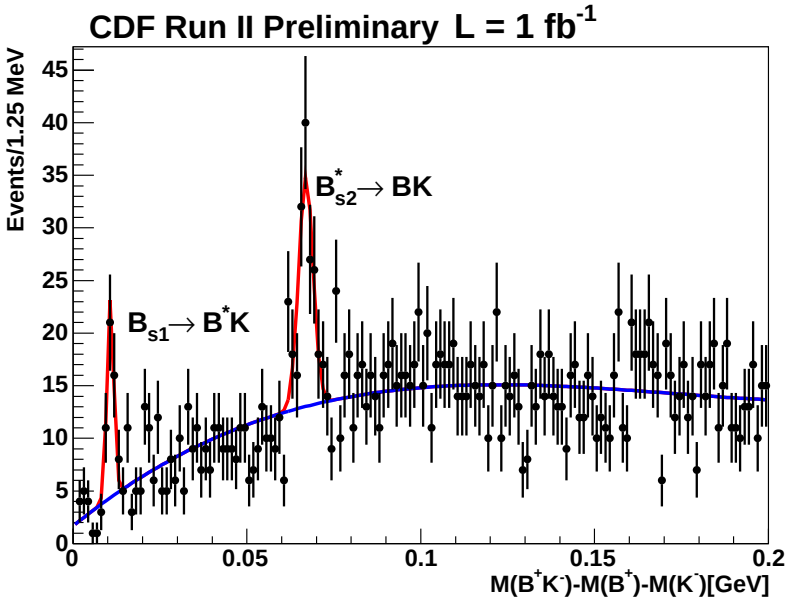

Figure 54: Fit to the data, using third order polynomial as an alternative for the background description. Values of important parameters are listed in table 7.

To make a cross check, we also use the alternative background description in our fit. Instead of default function, we use third order polynomial. Results of this fit are listed in table 7 and fit projection is in Figure 54. The Q values for the two peaks, which we want to quote as the results are exactly same.

8 Summary

In summary, we performed a search for the orbital excitations of the B_s mesons, commonly known as B_s^{**}. We observe two narrow peaks in the Q distribution of B^+K^- pairs. Those can be interpreted as due to the two narrow states decays $B_{s1} \to B^{+*}K^-$ and $B^*_{s2} \to B^+K^-$. From the unbinned maximum likelihood fit to the Q distribution we obtain

$$\begin{aligned} Q(B_{s1}) &= 10.73 \pm 0.21\,(stat) \pm 0.14\,(sys)\ \text{MeV} \\ Q(B^*_{s2}) &= 66.96 \pm 0.39\,(stat) \pm 0.14\,(sys)\ \text{MeV} \end{aligned}$$

The significance of the two observed peaks calculated using asymptotic behaviour of the $-2 \cdot \ln \mathcal{L}/\mathcal{L}_0$ of the unbinned maximum likelihood fit are $6.3\,\sigma$ and $7.7\,\sigma$ for the B_{s1} and B^*_{s2} respectively.

Using world avarage masses for K^-, B^+ and B^{*+} the Q values can be converted to the masses of the two states

$$\begin{aligned} m(B_{s1}) &= 5829.41 \pm 0.21(stat) \pm 0.14(sys) \pm 0.6(PDG)\ \text{MeV} \\ m(B^*_{s2}) &= 5839.64 \pm 0.39(stat) \pm 0.14(sys) \pm 0.5(PDG)\ \text{MeV} \end{aligned}$$

where the PDG error is due the error on the mass of the B^+ and the mass difference between B^+ and B^{*+}. Finally the mass difference of the two narrow B_s^{**} states is

$$\Delta m(B_{s2}^*, B_{s1}) \;=\; 10.20 \;\pm\; 0.44(stat) \;\pm\; 0.35(PDG) \text{ MeV}$$

which is close to the model expectations.

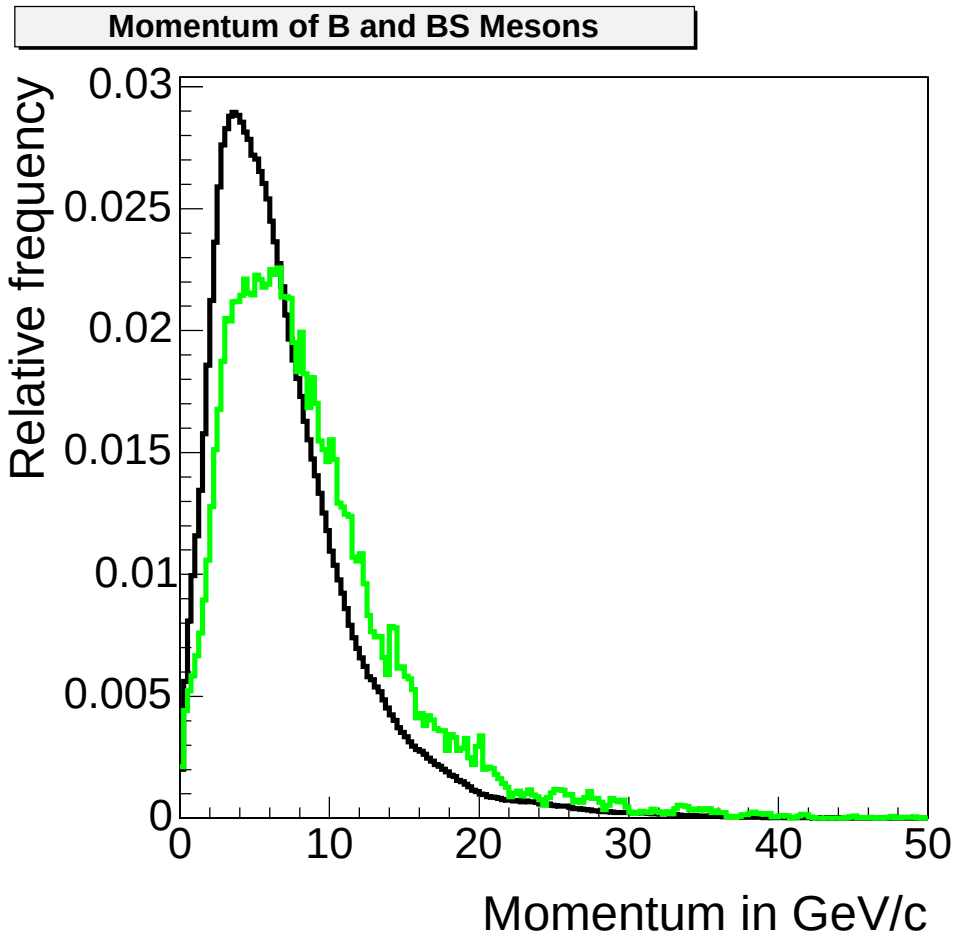

Figure 55: B^+ and B_s^{**} momentum distribution of the di-muon trigger sample.
In black: The B^+ momentum distribution of PYTHIA MC
In green: The B_s^{**} momentum distribution of BGen MC

A Additional information from Godparent committee

The following figures (55,56) show various momentum distributions for different MCs to make sure the the redecaying of the PYTHIA MC does not introduce new structures. All plots are made after the complete detector simulation and reconstruction.
Possible structures are due to this.
Figure 57 shows the number of candidates in both trigger samples.
Figure 58 shows for the di-muon trigger sample the mass distribution of the B_s^{**} candidates as given to the network. This shall demonstrate a correct reweighting of the MC to the data. The MC is reweighted to a smoothed data sample instead of the original data sample.
Figures 59 and 60 show the same for the TT trigger sample. Here are stronge fluctu-

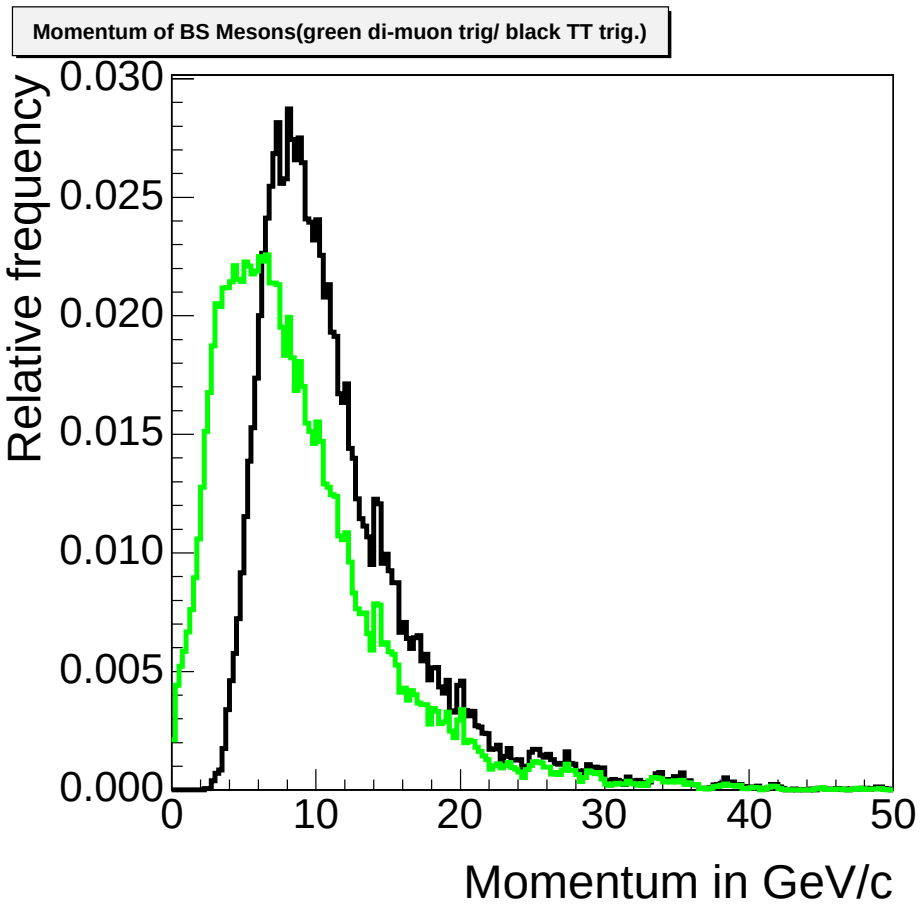

Figure 56: Momentum of B_s^{**} BGen MC mesons in both trigger samples
In green the B_s^{**} mesons of the di-muon trigger are shown, in black those from the TT trigger.

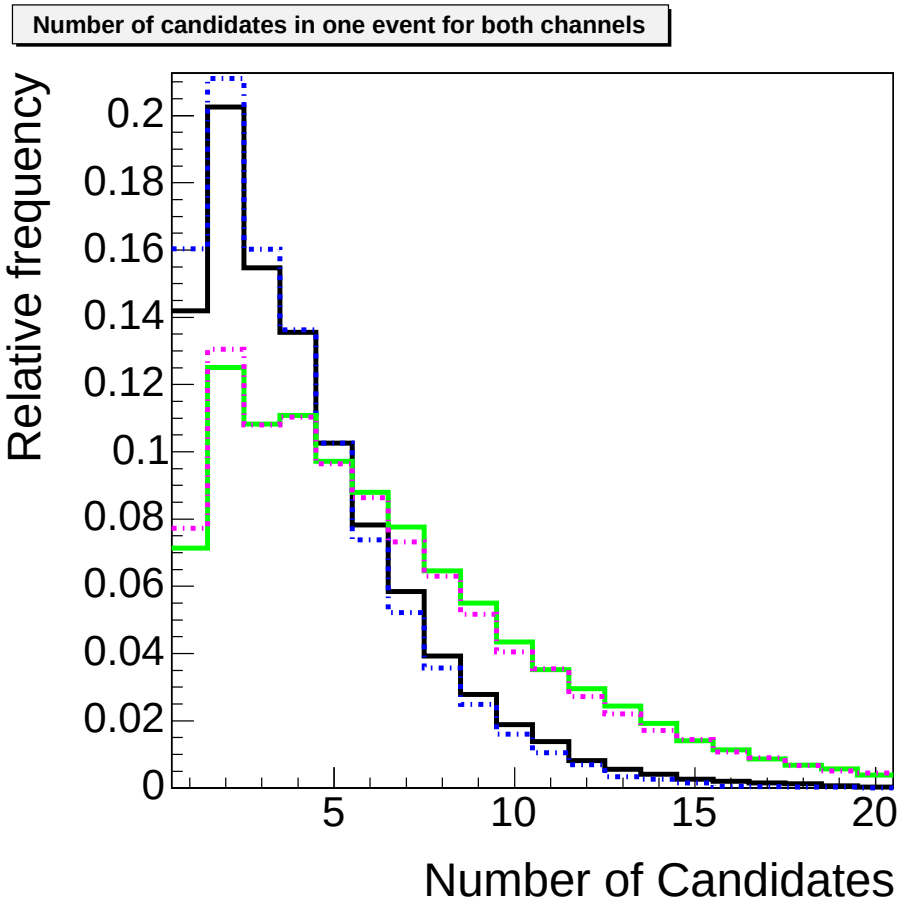

Figure 57: Distribution of the number of B_s^{**} candidates for both trigger samples and for right sign and wrong sign combinations
In black: number of candidates in the channel with J/ψ right sign combinations
In blue: number of candidates in the channel with J/ψ; wrong sign combinations
In green: number of candidates in the channel with D; right sign combinations
In pink: number of candidates in the channel with D; wrong sign combinations

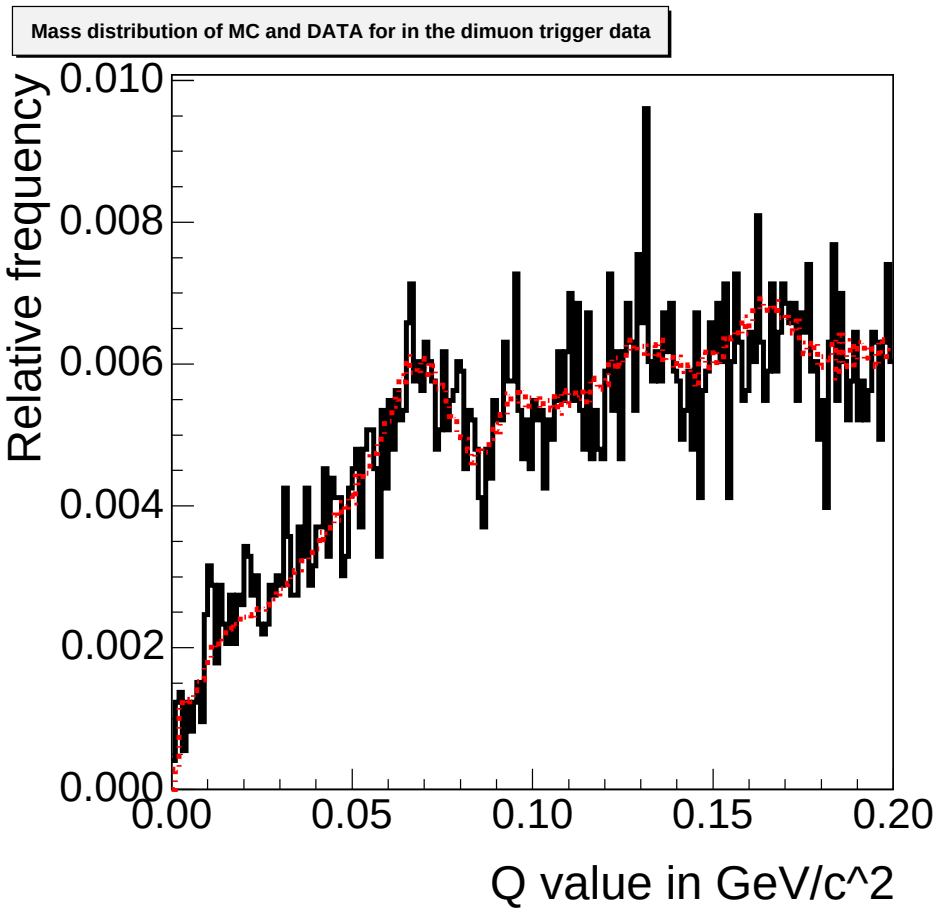

Figure 58: Mass distribution of the dimuon trigger sample B_s^{**} candidates as given to the net.
In red (dotted): The MC distribution after the reweighting to the smoothed data distribution.
In black (full line): The data after all precuts.

ations, so for a better understanding in the 60 there are error bars given to estimate the goddness of the reweighting.

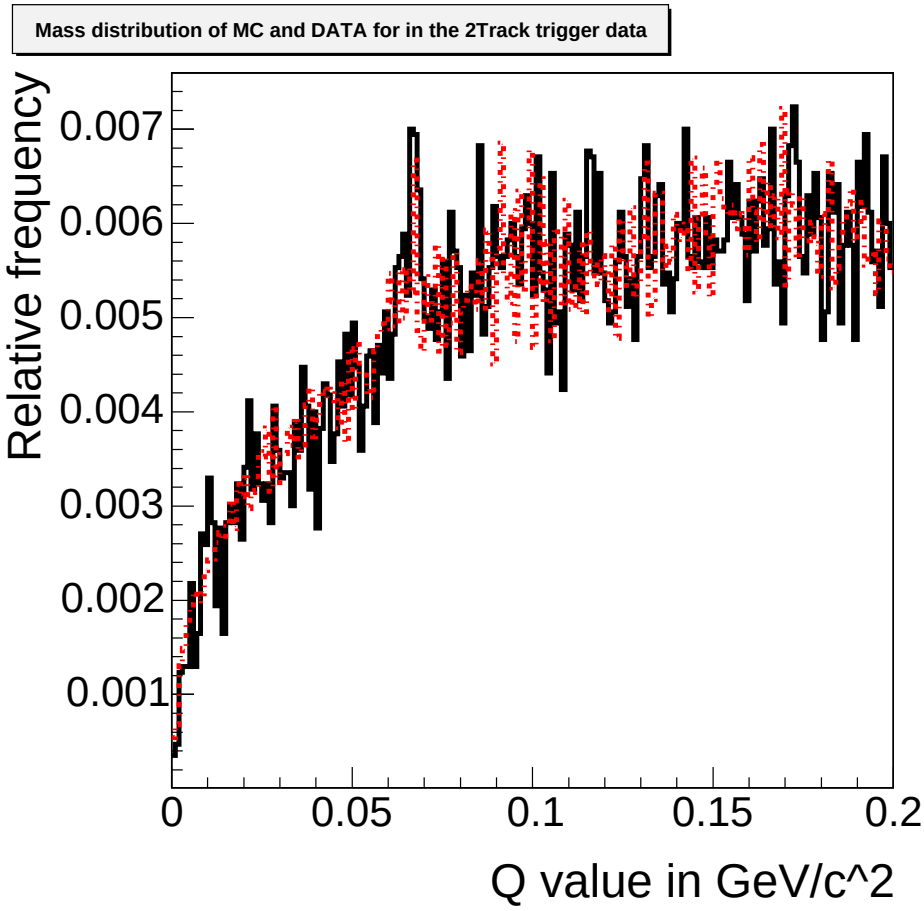

Figure 59: Mass distribution of the TT trigger sample B_s^{**} candidates as given to the net.
In red (dotted): The MC reweighted to the smoothed data distribution
In black (full line): The data after all precuts

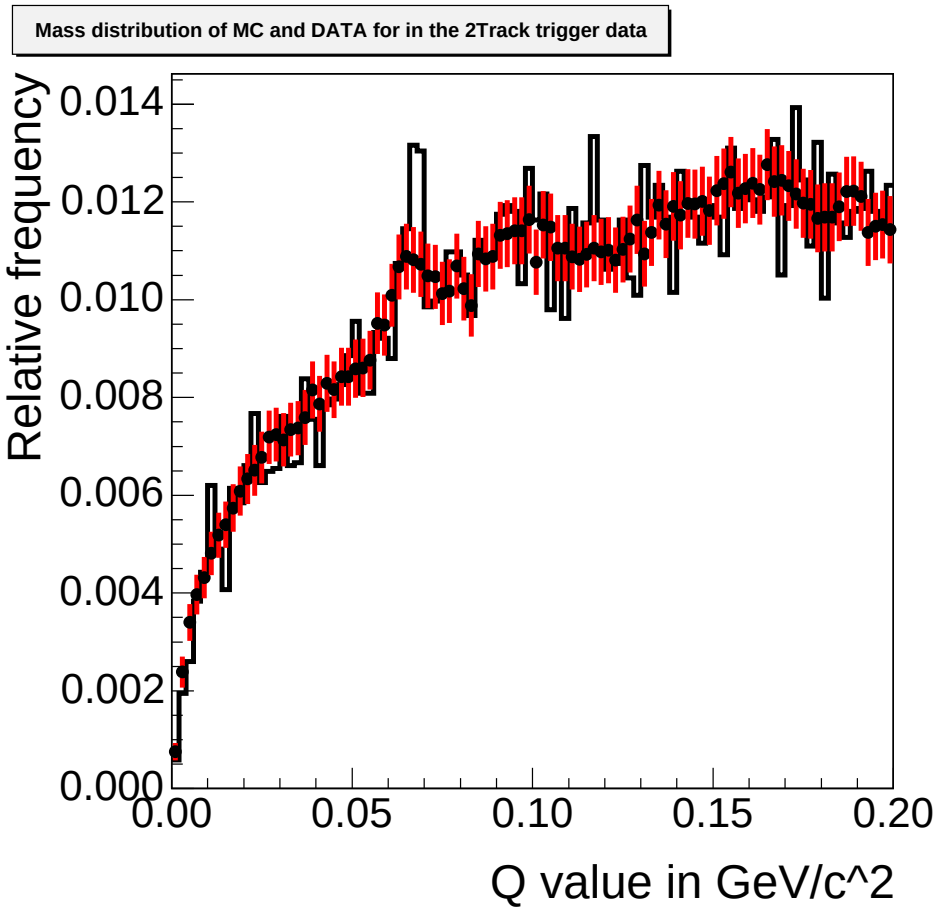

Figure 60: Mass distribution of the TT trigger sample B_s^{**} candidates as given to the net.
In red (dotted): The MC reweighted to the smoothed data sample. This figure includes error bars. The binning is increased to 2 MeV/c^2 for better visualisation
Black (full line): The data distribution after all precuts

B Decay table for J/ψ trigger

```
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
```

C Decay table for two track trigger

```
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
```

Acknowledgement

We would like to thanks to all people who attended our presentations for helpfull comments and discussion. This work was supported by the Bundesministerium für Bildung und Forschung, Germany and European Community's Human Potential Programme under contract HPRN-CT-2002-00292.

References

[1] F. Bedeschi et al., Observation of orbitally excited (L=1) B mesons. CDF note 7820, October 2005. 4, 45

[2] M. Campanelli, E. Gerchtein, Measurement of mass and width of the orbitally excited charm mesons $D_1(2420)^0$ and $D_2(2460)^0$. CDF Note 6918, February 2005. 4, 45

[3] OPAL Collab., R. Akers et al., Z. Phys. **C66**, 19 4, 6

[4] Delphi Collab., Z. Albrecht et al., DELPHI 2004-025 CONF 700; Z. Albrecht, *Analysis of Excited B-Mesons*, PhD-Thesis, IEKP-KA/03-16, Universität Karlsruhe, 2003; M. Moch, *Study of B^+-Mesons and of Excited b-Hadron Properties*, PhD-Thesis, IEKP-KA/2004-16, Universität Karlsruhe, 2004. Both thesis available at http://www-ekp.physik.uni-karlsruhe.de/pub/web/thesis. 4, 6, 7

[5] DØ Collab., *First Direct Observation of B_{s2}^{*0} meson*, DØnote 5027-Conf 4, 7, 8

[6] Ebert, Garkin, Faustov, Phys. Rev. **D57**, 5663; Ebert, Garkin, Faustov, hep-ph/9712318 5

[7] Godfrey, Kokoshi, Phys. Rev. **D43**, 1679 5

[8] Eichten, Hill, Quigg, Phys. Rev. Lett. **71**, 4116 5

[9] Falk, Mehen, Phys. Rev. **D53**, 231; hep-ph/9507311 5

[10] Particle Data Group, S. Eidelman et al., Phys. Lett. **B592**, 1 5, 6

[11] CDF DQM Group, http://www-cdf.fnal.gov/internal/dqm/goodrun/good.html 8

[12] C. Paus et al., see Cdf Code Browser http://cdfkits.fnal.gov/CdfCode/source/BottomMods/ 8

[13] D. J. Lange, The EvtGen particle decay simulation package. NIM **A462** (2001), 152 9

[14] T. Sjöstrand, P. Edn, C. Friberg, L. Lönnblad, G. Miu, S. Mrenna and E. Norrbin, Computer Phys. Commun. 135 (2001) 238 (LU TP 00-30, hep-ph/0010017) 9

[15] P. Murat, K. Anikeev, C. Paus, Description of BGeneratorII. CDF Note 5092 9

[16] M. Campanelli, E. Gerchtein, Calibration of the momentum scale for Kalman refitter using J/ψ events. CDF Note 6905 (March 25, 2004) 10

[17] J. Marriner, Secondary Vertex Fit with Mass and Pointing Constraints (CTVNFT), CDF Note 1996 (March 1993) 10

[18] M. Feindt, A Neural Bayesian Estimator for Conditional Probability Densities, physics/0402093, 2004 12

[19] B_s Mixing group, Signal Optimization for $B_d \to D\pi, D \to K\pi\pi$ using a neural network, 2006, CDF/DOC/BOTTOM/CDFR/8123 12

[20] Michael Feindt, Joachim Heuser, Ulrich Kerzel, Thomas Kuhr, Claudia Marino, Signal Selection of $X(3872) \to \mu^+\mu^-\pi^+\pi^-$ with NeuroBayes, CDF Note 8325 (June 2006) 12

[21] M. Feindt, The *NeuroBayes*© User's Guide, 2006 12

[22] F. James, MUNUIT, Function Minimization and Error Analysis, CERN Program Library Long Writeup D506, 1998 41

[23] Rene Brun, Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/ 41