Using Neural Nets for Opposite Side Flavour Tagging

To cite this article: Milnik Michael and Cdf Collaboration 2006 J. Phys.: Conf. Ser. 53 431

View the article online for updates and enhancements.

Related content

- The application of neural network PID controller to control the light gasoline etherification
  Huanxin Cheng, Yimin Zhang, Lingling Kong et al.

- Application of neural network and Levenberg–Marquardt algorithm for electrical grid planning
  M Boopathi, S Senthil Kumar and G Kalaisarasan

- Application of Neural Network Optimized by Mind Evolutionary Computation in Building Energy Prediction
  Chen Song, Wu Zhong-Cheng and Lv Hong
Using Neural Nets for Opposite Side Flavour Tagging

Milnik Michael for CDF Collaboration
Institut für Experimentelle Kernphysik, Universität Karlsruhe, Wolfgang-Gaede-Str. 1,
D-76131 Karlsruhe, Germany
E-mail: milnik@ekp.physik.uni-karlsruhe.de

Abstract. This contribution to the proceedings will present different applications of neural networks as a powerful tool to improve the measurement of the $B^0_s$ oscillation frequency. Opposite side lepton tagger gain performance due to a better particle identification and a better candidate selection. The jet charge tagger uses neural networks to measure the probability of tracks originating from $B$ mesons and improve the identification of jets as $b$-jet candidates.

1. Introduction

One of the key aspects of the CDF II $b$-physics program is the measurement of the mixing frequency $\Delta m_s$ of the $B^0_s - \bar{B}^0_s$ meson system. The determination of the B meson flavour at production and decay time is of vital importance in this measurement. The flavour of a $B$ meson describes whether the particle contains a $b$ or $\bar{b}$ quark. Opposite side flavour tagging is used to determine the flavour of a neutral B meson at production time. To understand how this fits into the whole scheme we will shortly describe how to measure a particle-antiparticle oscillation frequency.

There are two kinds of neutral $B$ mesons, $b\bar{q}$, where $\bar{q} = d$ or $s$, depending on the second quark inside of the meson: $B^0_d$ and $B^0_s$. These mesons oscillate from particle to antiparticle due to second-order weak interactions. Today the $B^0_d - \bar{B}^0_d$ oscillation is measured with very high precision and the current world average value is $\Delta m_d = 0.505 \pm 0.005 \text{ps}^{-1}$ ([1], [2]). The oscillation frequency is often written as a mass difference, since this mass difference between the mass eigenstates is directly proportional to the oscillation frequency.

Measuring the $B^0_s - \bar{B}^0_s$ oscillation is in principle very similar to the $B^0_d$ system, but since its oscillation frequency is higher and the production rate of $B^0_s$ mesons is lower its measurement is far more challenging. So far, several limits of the value were published (e.g. see recent DØ result [3]), but recently the CDF II collaboration presented a measurement of the $B^0_s$ oscillation: $\Delta m_s = 17.33^{+2.32}_{-0.21}\text{stat.} \pm 0.07\text{syst.}\text{ps}^{-1}$ in the range $16.96 < \Delta m_s < 17.91\text{ps}^{-1}$ at 95% C.L. (see figure 1 and [4]).

To determine the mixing frequency $\Delta m_s$ the asymmetry of mixed and unmixed $B^0_s$ and $\bar{B}^0_s$ mesons at any time $t$ is given by:

$$A(t) = \frac{N_{unmix}(t) - N_{mix}(t)}{N_{unmix}(t) + N_{mix}(t)} = \cos(\Delta m_s t),$$

(1)

where

$$N_{mix/unmix} = \frac{N_0}{2} e^{-\Gamma t} (1 \mp \cos(\Delta m_s t))$$

(2)

© 2006 IOP Publishing Ltd
describes the number of mixed/unmixed $B_s$ mesons.

A meson has mixed when the produced $B$ meson changes the flavour during its lifetime into its own antiparticle when it decays. Therefore, it is necessary to measure the flavour at production and decay time. The latter is determined by reconstructing the decay products of the $B_s$ meson in flavour specific mode. To measure the flavour at production time, flavour taggers are used.

Different flavour tagger strategies exist. In general, they are divided into opposite and same side tagging. Same side in this context means the part of the detector where the $B_s$ was reconstructed exclusively. Figure 2 illustrates the typical event topology of a semi-leptonic $B$ decay. This reconstructed decay mode defines the same side in this event. The opposite side is then the direction of the other $b$ hadron. In general we will focus here on the opposite side tagging, but will discuss same side tagging shortly in the outlook.

The performance of a flavour tagger is usually measured in $\epsilon D^2$, where $\epsilon$ is the efficiency of the flavour tagger and $D$ the dilution. The efficiency is the fraction of tagged events and the dilution describes the probability of the flavour tagger to give the right decision: $D = \frac{N_{rs} - N_{ws}}{N_{rs} + N_{ws}}$, where $N_{rs}/N_{ws}$ is the number of correctly/wrongly tagged events. The dilution $D$ is related to the purity $P$ by: $P = 2D - 1$. A dilution of zero means that the tagger provides only random decisions and is therefore in 50% of the cases right. Since the significance of a $B^0_s$ mixing analysis is proportional to $\sqrt{\epsilon D^2}$, this factor is a measurement of the quality of the flavour tagger.

2. Opposite Side Flavour Tagging

To determine the flavour of the reconstructed $B$ meson at production time we concentrate in this chapter on the other side. The other $b$ quark fragments into a $b$ hadron which can decay in several ways. Opposite Side Lepton Tagging is a way to use the charge of the lepton in a semi-leptonic decay to determine the flavour of the $B$ meson. A different approach is the Jet Charge Tagger (JQT), which follows an inclusive approach to measure the net-charge in the opposite side of the event. Opposite side Kaon Tagging (OSKT) exploits kaon tagging to identify a decay chain of the $b$. These approaches will be explained in detail in the next sections.

In general, the idea of opposite side flavour tagging is to determine the $b$ quark flavour at
Figure 2. Left: Possible decay of a $B$ meson. The reconstructed semi-leptonic decay defines the same side. The decay of the other $b$ on the opposite side is not drawn.
Right: The principle of the Jet Charge Tagger. On the opposite side a jet, which comes from the $B$ decay, is found. Its total charge provides information of the flavour of the $B$ meson at production time.

production time using the other $b$ in the event. Since this is anti-correlated to the $b$ flavour on the same side at production time, together with the flavour at decay time it is possible to decide if the $B$ meson has mixed or not.

2.1. Jet Charge Tagger

A Jet Charge Tagger (JQT) does not try to reconstruct any special decay mode on the opposite side of the event. Since the $b$ and $\bar{b}$ have the opposite charge, the sum over the charge of the decay products of the $B$ meson will have a different net charge (see figure 2). Since the event consists of more than just the two $B$ mesons, e.g. particles from fragmentation, more sophisticated methods then just summing over all tracks on the opposite side are necessary. Therefore neural networks are used to improve the result.

The jet candidates are found by the Cone Clustering algorithm (see [5]). Each jet is then processed by the Secondary Vertex (SecVtx) algorithm to find displaced vertex inside the jet. A displaced, or secondary, vertex consists of tracks forming a vertex different from the primary vertex. The first neural network is used to estimate the probability of a track coming from a $B$ meson decay. This so called TrackNet is trained using simulated events. Figure 3(left) illustrates the performance of the TrackNet. The output of the TrackNet and related variables are a very important input for the next neural network.

Since there are several candidates a second neural network is trained to analyze each jet in an event and estimates the probability, that it is coming from a $B$ meson: $b$-JetNet. Figure 3(right) shows the performance of the neural network. The jet with the highest neural network output is chosen as the tagging candidate and the jet charge is calculated:

$$Q_{\text{Jet}} = \frac{\sum_i Q_i p_{T,i}(1 + t_i)}{\sum_i p_{T,i}(1 + t_i)},$$

(3)
where $t_i$ is the track probability that it is coming from a $B$, estimated by the TrackNet, $Q_i$ its charge and $p_{T,i}$ its transverse momentum.

The quality of the neural network based JQT was measured to be: $\epsilon D^2 = 0.927 \pm 0.014\%$, which is an improvement of about 30% compared to the cut based jet charge algorithm (see [6],[7]).

### 2.2. Opposite Side Lepton Tagging

Although the illustration in figure 2 shows a same side $B$ decay, this semi-leptonic decay can also occur on the opposite side. By identifying the lepton we can estimate the $b$ quark flavour of the $B$ meson at production time. Thus, a good lepton identification is essential for opposite side flavour tagging. There are different ways to achieve good lepton identification. The CDF II detector has different components which allow particle identification:

- Muon chambers are mounted in the outermost part of the CDF II detector, behind sufficient material to stop most hadrons. Mainly muons can penetrate the material and produce a signal in the muon chambers.
- The electromagnetic and hadronic calorimeter can separate between electrons and pion/kaons, but their separation power is better for high energetic particles.
- A Time of Flight (ToF) detector is used for particle identification in the low momenta region.
- The energy deposited ($dE/dx$) in the drift chamber: Central Outer Tracker (COT).
- Kinks in the track of electrons originating from Bremsstrahlung can be identified.

A traditional approach of applying cuts on the different variables can be used to select good lepton candidates, but this has several disadvantages: the correlation between variables are not taken into account, the determination of multidimensional cuts for different analysis requirements on efficiency and purity is very difficult and complicated treatment of missing variables, for example there might be no ToF measurement. There are more advanced techniques such as likelihood ratios or neural networks to combine all available information in a better way.

Likelihood based flavour taggers were developed at CDF and are used in the latest $B_s$ Mixing analysis at CDF. Muon and electron based opposite side flavour taggers use a likelihood based

#### Figure 3. Left: Neural network output of the TrackNet for data and simulation. Both distribution are very similar. The signal to background ratio increases for higher values of the neural network.

Right: Neural network output of the b-JetNet for data and simulation. The logarithmic plot shows a decent agreement of data and simulation over several orders of magnitude.
method to select a clean sample of leptons. The performance of both taggers are measured using data: \( \epsilon D^2 = 0.678 \pm 0.026\% \) for muons, \( \epsilon D^2 = 0.366 \pm 0.031(\text{stat})^{+0.065}_{-0.056}(\text{syst})\% \) for electrons ([8], [9]). Since the particle identification is one of the main component for a lepton based opposite side tagger, this should be as good as possible. Neural networks are used to improve the existing flavour taggers. The muon sample in general is already quite clean, but there is still background that can be reduced. For electrons not only the background can be reduced, it is also possible to significantly increase the efficiency. Both approaches, cut and likelihood based, rely on calorimeter information. Therefore, the efficiency is limited to electron candidates for which this information is available. Since electron identification is also possible by using dE/dx of the drift chamber and time-of-flight detector without calorimeter, the neural network identification also includes these candidates. Another large background for electrons are conversion electrons, which are identified using a further neural network. They have to be excluded by most physics analysis.

A good particle identification is essential, but it might be the case that there is more than a single lepton candidate. Thus the best candidate should be selected to achieve the best possible result. Therefore, for each candidate an estimation should be done which allows to chose the best one. One way to get an estimation of the best candidate is on a statistical basis. The dependence of the dilution of some variables like the transverse momentum relative to the jet \( (p_{t_{rel}}) \) can be estimated in many events. This dependence of the dilution from the variable can then be used to estimate the dilution for each candidate.

Another approach for the estimation of the best candidate is to use a neural network. Since there is a correlation to variables like \( p_{t_{rel}} \), it is possible to use a network to decide whether a particle originates from a \( B \) decay. For the opposite muon tagger this is less important, since in most cases there is only one muon found in the event. But for the electrons there may be, depending on the efficiency of the particle identification method used, many candidates and the best one has to be chosen. This is especially the case for the network based approach. Since it does not depend on the calorimeter information, its efficiency and the number of candidates is higher. In the case of electrons the candidate selection network alone will improve the estimated \( \epsilon D^2 \).

Overall, the estimated gain for the muon tagger is only a few percent, but the improvement of the electron tagger by using neural networks will be significantly higher.

2.3. Opposite Side Kaon Tagging

Another way to predict the \( b \) flavour on the opposite side is by looking at kaons from the decay chain \( b \rightarrow c \rightarrow s \). The Opposite Side Kaon Tagger (OSKT) mainly depends on a good particle separation of kaons from pions. At CDF II the information of dE/dx in the COT and time of flight measurement of the ToF is combined using a likelihood method. The separation power is higher for particles with lower transverse momentum, since the ToF works better in this region. The kaon candidate must be displaced from the primary vertex to distinguish it from the background kaons coming from fragmentation tracks. Further the candidates are categorized into different classes depending on the kind of track-based jets to estimating the opposite side \( b \) direction. The performance of the OSKT is measured to be: \( \epsilon D^2 = 0.229 \pm 0.016(\text{stat}) + 0.001(\text{syst})\% \) [10].

3. Outlook

Not only the opposite side but also the same side contains information about the flavour of the \( B \) meson at production time. By looking at the remaining fragmentation tracks around the \( B \) meson and its charge, it is possible to exploit its correlation to the flavour at production time. Since the \( B_s^0 \) meson consists also of an \( s \) or \( \bar{s} \) quark, the leading fragmentation track is also
expected to contain strangeness. The flavour tagger identifying the kaon candidate is therefore called Same Side Kaon Tagger (SSKT).

Similar to the OSKT, the SSKT depends on a good particle separation of kaons from pions. But using this particle identification, the SSKT is found to be much more powerful than the opposite side taggers (see [11]). In contrast to the opposite side flavour taggers, the dilution of the SSKT can not be measured on data without knowledge of the $\Delta m_s$, since the tagging characteristics depend on the particular $B$ meson in this event. Thus the performance must be evaluated on simulated data. The simulation at CDF II was improved to fulfill the necessary precision. The final tagging performance of the SSKT was estimated to be: $\epsilon D^2 = 4.0^{+0.9}_{-1.2}\%$.

So far there was a clear distinction between same side and opposite side flavour taggers. But the clear separation of the tracks associated with the two $B$ mesons depends on the momentum of the mesons. For high energetic $B$ mesons with several GeV energy, all decay products of one meson will be found in a jet. The size of the cone of the jet will be inverse proportional to the momentum of the meson. The probability for $B$ meson production drops exponentially for higher momentum. So most events are in the low energetic region at around a few GeV. The event topology in this energy region does not show a clear jet structure. Both "jets" are spread over major parts of the detector, since the $B$ mesons are nearly at rest. The distinction for low energetic events, which is the majority at CDF II, between opposite and same side is therefore misleading.

Another issue by using different flavour taggers arises from their independent decisions. For a given event different taggers might give contradicting results for the flavour of the $B$ meson at production time. Thus, there must be some kind of order or combination to treat these cases. At the moment at CDF II the different opposite side taggers have different priorities. If there is a muon candidate, the muon tagger decision is taken. The muon tagger provides a rare, but quite accurate decision and has therefore the highest priority for the opposite side taggers. If there is no muon, the decision of the electron tagger is taken. In case that there is no lepton candidate at all, the jet charge tagger is used, which most of the time provides a decision and hence has to be the last one. The same side and opposite side are combined using:

$$D = \frac{D_1 \pm D_2}{1 \pm D_1 D_2},$$

where $D_x$ is the dilution of each flavour taggers and $D_1 > D_2$. The plus corresponds to the case that the decisions of the OST and SSKT agree and the dilution of this event is increased, the minus to the case they disagree.

This ordering is probably not optimal in all cases. For example, the JQT must be the last tagger, since it provides a decision in nearly all cases. Even when its decision is true, there might be a lepton and its tagger has a higher priority. One way to deal with both problems, the lose distinction between opposite and same side and the combination of different flavour tagger decisions, is to use a neural network combining all information used for each single tagger.

4. Conclusion

Neural networks are very powerful tools, which help to measure $B_0^0$ oscillation in several ways. Opposite side lepton tagger gain performance due to a better particle identification and a better candidate selection. For electrons there is an additional network to find conversion electrons. The neural networks of the jet charge tagger help to identify tracks originating from $B$ mesons and improve the identification of the right $b$-jet. But not only can each flavour tagger be improved separately by using neural networks, the combination of different tagging strategies into a single flavour tagger using neural networks is also under current development. This will also include the same side information and will make the current ranking of different flavour taggers, in case of multiple or contradicting decisions, obsolete.
Besides of the usage for tagging, it is also possible to use neural networks for selection. This was not discussed here, but shows the flexible usage of neural networks. The number of events used for the measurement of the $B_s$ mixing is increased by using different decay modes for the reconstruction. Further, each of these decay modes can be improved using neural networks. The signal to background ratio can be significantly enhanced compared to cut based methods. This improved candidate selection provides a larger and cleaner sample of $B_s$ candidates [12].

Recent publications of the CDF II collaboration [4] showed already a good measurement of the mixing frequency in the $B^0_s$ system: $\Delta m_s = 17.33^{+0.42}_{-0.21}$ (stat.) $\pm 0.07$ (syst.) ps$^{-1}$, where the probability this measurement being a fluctuation is only 0.2%. Due to the ongoing data taking, the promising improvements by using neural networks for the selection and tagging and further development, CDF will measure the oscillation frequency with even higher significance in the future.

References

[1] Eidelman S. et al. Phys. Rev., B592, 2004.
[2] Abe K. et al. (Belle Collaboration). Phys. Rev. D, D71, 2005.
[3] The D0 Collaboration. First direct two-sided bound on the $b^0_s$ oscillation frequency. hep-ex/0603029, 2006.
[4] The CDF Collaboration. Measurement of the $b_0 - \bar{B}_s$ oscillation frequency. CDF Public Note 8204, 2006.
[5] Baur, U., Ellis, R. K., and Zeppenfeld, D. QCD and weak boson physics in run II. Prepared for Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting, Batavia, Illinois, 4-6 Nov 1999.
[6] Claudia Lecci. A Neural Jet Charge Tagger for the Measurement of the $B^0_s - \bar{B}^0_s$ Oscillation Frequency at CDF. PhD thesis, Karlsruhe, 2005.
[7] The CDF Collaboration. Neural network based jet charge tagger.
http://www-cdf.fnal.gov/physics/new/bottom/050303.blessed-jqtnn/.
[8] The CDF Collaboration. Muon b flavor tagging - a likelihood approach.
http://www-cdf.fnal.gov/physics/new/bottom/072204.blessed-like-mu-tag/.
[9] The CDF Collaboration. Likelihood based electron tagging.
http://www-cdf.fnal.gov/physics/new/bottom/081204.blessed-like-el-tag/.
[10] The CDF Collaboration. Opposite side kaon tagger.
http://www-cdf.fnal.gov/physics/new/bottom/060413.blessed-oskt/.
[11] B Mixing Group. Same side kaon tagging in fully reconstructed decays. CDF Note 8206, 2006.
[12] Christian Dör. In Progress. PhD thesis, Karlsruhe, 2006.